



**UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**DISEÑO DE UN SISTEMA DE PRONÓSTICO PARA UNA PLATAFORMA DE  
MARKETING MÓVIL GEOLOCALIZADO**

*PROYECTO DE GRADO PARA OPTAR AL GRADO DE MAGÍSTER EN  
INGENIERÍA DE NEGOCIOS CON TECNOLOGÍAS DE INFORMACIÓN*

*MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL*

**PAULINA TERESA SUÁREZ ALDUNATE**

**PROFESOR GUÍA:**  
SEBASTIÁN RÍOS PÉREZ

**MIEMBROS DE LA COMISIÓN:**  
NICOLÁS OZIMICA GACITUA  
JORGE RETAMALES MORENO

SANTIAGO DE CHILE  
2017

## RESUMEN

En base a la alianza entre el Centro de Investigación en Inteligencia de Negocios (CEINE) y una empresa de telecomunicaciones, nace un proyecto *spin off* con el objetivo de ser una empresa capaz de enviar la publicidad adecuada a una persona determinada, en el momento y lugar correctos. Entregando valor económico a la información georreferenciada de los centros de recolección de datos de antenas telefónicas y funcionando como una nueva línea de negocios para la TELCO. Cuyos clientes son otras empresas comercializadoras y los usuarios los clientes de la empresa de telecomunicaciones.

Para construir el proyecto se trabajó con dos líneas de foco: una línea de investigación, donde se creó un modelo capaz de detectar patrones móviles en base a la información de las antenas de telefonía, y con ello predecir posiciones relativas futuras; y una línea de negocios, en la cual se creó tanto la estructura de procesos (en base a la metodología de procesos de Oscar Barros [1], trabajando el macroproceso 1: Cadena de Valor) como los planes de operaciones y marketing adyacentes a un negocio que utilice los modelos creados en la línea de investigación.

Para el desarrollo del negocio se utiliza la metodología *Canvas* donde se definen los segmentos de clientes, propuesta de valor, canales de comunicación, tipo de relación con el cliente, fuentes de ingreso, recursos, actividades y socios clave, y estructura de costos. La línea de investigación se trabajó con la metodología *CRISP-DM* que consta de las etapas cíclicas de: comprensión del negocio, entendimiento y preparación de los datos, modelado, evaluación e instalación en producción. Gracias a la línea de investigación se realizó un *benchmark* de modelos de clasificación (OPTICS, DBSCAN y K-means, siendo el último el que entrega mejores resultados) y predicción (árboles de decisión, *logit*, o *support vector machine*. Donde los resultados usados son los del primero), los que generaron nuevos procesos de minería de datos que responden al problema de negocio de enviar los mensajes correctos en el momento y lugar correcto.

Como contexto de mercado, se tiene que la inversión en marketing móvil crece seis veces más rápido que la publicidad por página *web* en el mundo, mientras en Chile se posiciona como la tercera plataforma con mayor inversión del 2015 (USD \$161MM) [2]. Relacionado con lo anterior, la tendencia de la personalización se observa como factor de éxito en las acciones de marketing online, es más, el 74% de los clientes se frustra al recibir contenido publicitario irrelevante según sus intereses [3]. Por otro lado, las empresas han mejorado un 19% sus resultados gracias a la personalización del contenido.

El proyecto fue evaluado como rentable con un VAN efectivo de cerca de CLP\$700MM, evaluado en un horizonte de tiempo de 3 años.

## DEDICATORIA

*A mis padres, Alonso y Gemy, que me han llevado el camino del esfuerzo,  
mi hermana y gran amiga Claudia y mi mayor apoyo Mauricio.*

## **AGRADECIMIENTOS**

Esta no es sólo la culminación de una etapa de más de 7 años, sino el término de mi vida como estudiante, aunque quién sabe si vuelva a serlo en el futuro. Por esto, me gustaría agradecer a todas las personas que se cruzaron en mi camino.

A mis amigos de la infancia y profesores del Blas Cañas y Carmela que me motivaron a tener metas altas y me apoyaron en mis desafíos.

A mis amigas del Carmela Carvajal de Prat, con las que maduré y crecí como persona, con ellas me malacostumbré a tener buenas amigas y compañeras de trabajos.

A las chicas de aeróbica, deporte que marcó mi vida. Nos esforzamos mucho y a la vez compartimos grandes momentos que aún extraño. Gracias, en especial a las niñas con las que compartí coreografías y a la Coté y Claudia por enseñarme, corregirme, motivarme y entregarme el sentido de hacer deporte.

A mis amigas de la vida, Cho, Ale y Andrea. Las que siempre están ahí, esas amigas que aunque no veas por varias semanas, cuando las ves es como si no hubiese pasado el tiempo. ¡Las adoro! gracias por ser parte de mi vida. A mi amigo Yorch, te incluyo en mi etapa escolar porque ahí nos conocimos y soy muy feliz de que sigamos compartiendo los momentos importantes de la vida.

A los amigos que hice en la universidad, los de plan común Cata, Feña H., Seba, Benja, Armando, Sapu, Daniel, Feña B. y Samy. A las que conocí más adelante Blanca, Naty y Angie, por la alegría que siempre me entregaron. Y a mis buenos amigos Willy, Dani, Cony, Nacho, Mati, Javi y Mauri S., gracias por las conversaciones, las comidas y las risas.

Al CEINE y MBE. Al profe Seba por darme la oportunidad y confiar en mis capacidades. A la Coni por su ayuda y hacerme sentir en casa. A la Pabli, Laurita y Ana María por toda su ayuda.

A mis papás, Alonso y Gemy, los que me han guiado para ser una buena persona, que me han enseñado a valorar las cosas importantes de la vida, desde la comida en la mesa, hasta todas esas veces que me fueron a buscar a las 3 AM a casas de amigos, me compraban los materiales que les pedía a última hora, me llevaban la información que necesitaba cuando tenía que hacer trabajos, me entregaron su amor. Ellos me han levantado las veces que me he caído y son mi apoyo en las buenas y en las malas. No tengo palabras para decirles lo mucho que los amo y lo agradecida que estoy por tenerlos en mi vida.

A mi hermana, Claudia, que es la mejor. No sé si alguien tiene la tremenda suerte que tengo de tener a alguien como ella a su lado. Cariñosa, preocupada, ¡lindaa!, inteligente y esforzada, siempre me ayudaba a estudiar en el colegio, siempre me entregó todo su amor, siempre fue mi estrella a seguir. Ama a hermana.

Y por último, al mejor hombre que pude haber encontrado, Mauricio, mi compañero de vida hace casi 7 años. Sin ti, probablemente no estaría donde estoy ahora, te admiro y te amo, gracias por tooodaaa tu ayuda en esta tesis y en tantos ramos de la u. Pero también en tantas otras facetas de mi vida, donde fuiste la clave para seguir esforzándome, donde me entregaste tu amor y me hiciste feliz. Me haces mejor persona, soy inmensamente feliz a tu lado, te amo. Muchas gracias también a tu familia por su compañía y cariño.

Espero que nadie se me haya quedado en el tintero. Muchísimas gracias a todos, de todos he aprendido y con todos he crecido.

Besos y abrazos para todos.

Pauli.

# TABLA DE CONTENIDO

<b>Capítulo 1: Introducción y contexto</b> .....	<b>1</b>
1.1 Antecedentes de la industria .....	1
1.2 Oportunidad identificada.....	2
1.3 Objetivos y resultados esperados del proyecto .....	4
1.3.1 Objetivo general .....	4
1.3.2 Objetivos específicos .....	4
1.3.3 Resultados esperados .....	4
1.4 Alcance .....	5
1.5 Aspectos legales.....	6
1.5.1 Ley 19.496, ley de protección al consumidor .....	6
1.5.2 Ley 19.628, ley de protección de datos de carácter personal.....	8
1.6 Principales competidores.....	9
1.6.1 Servicios de aplicaciones móviles.....	9
1.6.2 Servicios de club de beneficios.....	11
1.6.3 Servicios de descuentos móviles .....	11
1.7 Dinámica competitiva.....	12
1.8 Riesgos potenciales.....	13
1.8.1 Precisión de la predicción según el tipo de data .....	13
1.8.2 Privacidad de la ubicación .....	13
1.8.3 Aspectos organizacionales .....	14
<b>Capítulo 2: Marco teórico</b> .....	<b>15</b>
2.1 Metodología de ingeniería de negocios.....	15
2.1.1 Metodología de procesos .....	15
2.1.2 Business model canvas .....	16
2.2 Lógica de negocios.....	17
2.2.1 K-means.....	18
2.2.2 BBSCAN (density-based spatial clustering of applications with noise) [11].....	19
2.2.3 OPTICS [12].....	21
2.2.4 Regresión logística [13].....	21
2.2.5 Support vector machine [16] .....	22
2.2.6 Árboles de decisión.....	23
2.2.7 Metodologías de validación de modelos .....	25
2.2.8 Medidas de desempeño de los modelos de clasificación .....	26
<b>Capítulo 3: Planteamiento estratégico y modelo de negocios</b> .....	<b>29</b>
3.1 Visión.....	29
3.2 Misión .....	29
3.3 Estructura básica de la empresa .....	29
3.4 Análisis foda .....	30
3.4.1 Fortalezas .....	30
3.4.2 Oportunidades.....	31
3.4.3 Debilidades .....	31
3.4.4 Amenazas .....	31
3.5 Posicionamiento estratégico.....	32

3.6	Balanced scorecard .....	33
3.7	Modelo de negocios.....	36
3.7.1	Segmentos de clientes.....	36
3.7.2	Propuesta de valor .....	38
3.7.3	Relación con clientes y usuarios .....	38
3.7.4	Canales .....	39
3.7.5	Socios clave .....	39
3.7.6	Actividades clave .....	39
3.7.7	Recursos clave.....	40
3.7.8	Estructura de costos .....	40
3.7.9	Fuentes de ingresos.....	40
<b>Capítulo 4:</b>	<b>Propuesta de diseño de procesos .....</b>	<b>41</b>
4.1	Arquitectura de procesos TO BE .....	41
4.2	Diseño detallado de procesos to be en ideo.....	44
4.2.1	Nivel a1: planificación del negocio: .....	44
4.2.2	Nivel a3: desarrollo de campañas de marketing: .....	47
4.2.3	Nivel a31: administración de la relación con el cliente y usuario .....	52
4.2.4	Nivel a311: análisis y campañas de marketing .....	56
4.2.5	Nivel a313: venta y atención al cliente.....	59
4.2.6	Nivel a3131: ventas.....	61
4.2.7	Nivel a3133: seguimiento.....	64
4.2.8	Nivel a32: administración de la relación con proveedores:.....	65
4.2.9	Nivel a34: producción y entrega.....	68
4.3	Diseño detallado de procesos to be de la cadena de valor en bpmn .....	69
4.3.1	Archivos y bases de datos relacionadas a las los procesos .....	69
4.3.2	Retroalimentación .....	71
4.3.3	Retroalimentación de modelos.....	72
4.3.4	Actualización de bd producción .....	74
4.3.5	Opt-in .....	75
4.3.6	Captación de clientes.....	76
4.3.7	Postventa .....	77
4.3.8	Producción .....	79
<b>Capítulo 5:</b>	<b>Diseño de lógica de negocios .....</b>	<b>85</b>
5.1	Entendimiento de los datos .....	85
5.2	Preparación de los datos: generación de un data warehouse.....	92
5.3	Clustering.....	97
5.3.1	DBSCAN .....	97
5.3.2	K-means.....	110
5.3.3	Replicaciones.....	127
5.4	Pronóstico.....	130
5.4.1	Generación de datos de entrenamiento y testeo: .....	130
5.4.2	Modelamiento:.....	131
<b>Capítulo 6:</b>	<b>Propuesta de apoyo tecnológico .....</b>	<b>136</b>
6.1	Especificación de requerimientos.....	136
6.1.1	Requerimientos funcionales.....	136

6.1.2	Requerimientos no funcionales+	137
6.2	Arquitectura tecnológica	143
6.3	Diseño de la aplicación	146
6.3.1	Casos de uso	146
<b>Capítulo 7:</b>	<b>Evaluación del proyecto</b>	<b>151</b>
7.1	Resultados de predicción obtenidos	151
7.2	Plan de marketing	155
7.2.1	Producto	156
7.2.2	Precio	156
7.2.3	Plaza	157
7.2.4	Promoción	157
7.3	Plan de operaciones	157
7.3.1	Implementación	157
7.3.2	Desarrollo	157
7.4	Definición de beneficios y costos	158
7.5	Flujo de caja	160
7.6	Análisis de sensibilidad	161
<b>Capítulo 8:</b>	<b>Conclusiones</b>	<b>163</b>
8.1	Conclusiones sobre la factibilidad del proyecto	163
8.2	Conclusiones sobre los riesgos del proyecto	163
8.3	Conclusiones sobre la investigación realizada	164
8.4	Trabajo futuro	165
<b>Capítulo 9:</b>	<b>Bibliografía</b>	<b>167</b>
<b>Capítulo 10:</b>	<b>Anexos</b>	<b>169</b>
10.1	anexo 1: KPI's según objetivos de negocios	169
10.2	anexo 2: Diseño de procesos de la cadena de valor en bpmn	173
10.2.1	Captación de usuarios	173
10.2.2	Reportes de avance	175
10.2.3	Registro de pagos	176
10.2.4	Retroalimentación de estrategia	177
10.2.5	Retroalimentación operacional	178
10.3	anexo 4: Estudio relacional para optics	179
10.4	anexo 5: Etras soluciones de k-means	181
10.5	anexo 6: Detalle del análisis para otros cg	182
10.5.1	Mall vespucio norte	183
10.5.2	Mall plaza oeste	186
10.5.3	Mall plaza maipú	189
10.5.4	Mall plaza egaña	192
10.5.5	Mall alto las condes	196
10.6	anexo 7: Flujo de caja sin proyecto	200

## INDICE DE TABLAS

Tabla 1. Resultados esperados y entregables .....	5
Tabla 2. Detalle de entradas y salidas. Nivel A1 Planificación del Negocio .....	41
Tabla 3. Detalle de entradas y salidas. Nivel A2 Desarrollo de Nuevas Capacidades .....	42
Tabla 4. Detalle de entradas y salidas. Nivel A3 Desarrollo de campañas publicitarias .....	43
Tabla 5. Detalle de entradas y salidas. Nivel A4 Procesos de apoyo.....	43
Tabla 6. Detalle de entradas y salidas. Nivel A11 Definir Concepto del negocio y visión .....	45
Tabla 7. Detalle de entradas y salidas. Nivel A12 Gestionar Desarrollo de estrategia .....	45
Tabla 8. Detalle de entradas y salidas. Nivel A13 Desarrollo de estrategia .....	46
Tabla 9. Detalle de entradas y salidas. Nivel A14 Mantenimiento de estado .....	46
Tabla 10. Detalle de entradas y salidas. Nivel A31 Administración de la relación con clientes y usuarios .....	48
Tabla 11. Detalle de entradas y salidas. Nivel A32 Administración de la relación con proveedores.....	49
Tabla 12. Detalle de entradas y salidas. Nivel A33 Gestión de producción y entrega .....	50
Tabla 13. Detalle de entradas y salidas. Nivel A34 Producción y entrega .....	51
Tabla 14. Detalle de entradas y salidas. Nivel A35 Mantenimiento de estado .....	52
Tabla 15. Detalle de entradas y salidas. Nivel A311 Análisis y campañas de marketing .....	53
Tabla 16. Detalle de entradas y salidas. Nivel A312 Atención a usuarios.....	54
Tabla 17. Detalle de entradas y salidas. Nivel A313 Venta y atención al cliente. ....	54
Tabla 18. Detalle de entradas y salidas. Nivel A314 Procesamiento de órdenes .....	55
Tabla 19. Detalle de entradas y salidas. Nivel A3111 Introducción de nuevos análisis y modelos .....	56
Tabla 20. Detalle de entradas y salidas. Nivel A3112: Análisis de la respuesta de usuarios .....	57
Tabla 21. Detalle de entradas y salidas. Nivel A311: Análisis del comportamiento de clientes..	57
Tabla 22. Detalle de entradas y salidas. Nivel A3113: Definir acciones de marketing.....	58
Tabla 23. Detalle de entradas y salidas. Nivel A3114: Planificar ventas .....	58



Tabla 24. Detalle de entradas y salidas. Nivel A3131 Ventas.....	59
Tabla 25. Detalle de entradas y salidas. Nivel A3132 Postventa.....	60
Tabla 26. Detalle de entradas y salidas. Nivel A3133 Seguimiento.....	60
Tabla 27. Detalle de entradas y salidas. Nivel A31311 Acercamiento a clientes .....	61
Tabla 28. Detalle de entradas y salidas. Nivel A31312 Evaluar oportunidad de cliente .....	62
Tabla 29. Detalle de entradas y salidas. Nivel A31313 Desarrollar contacto y generar propuestas .....	62
Tabla 30. Detalle de entradas y salidas. Nivel A31314 Cerrar venta.....	63
Tabla 31. Detalle de entradas y salidas. Nivel A31331 Evaluar desempeño de ventas .....	64
Tabla 32 Detalle de entradas y salidas. Nivel A31332 Generar acciones correctivas .....	65
Tabla 33. Detalle de entradas y salidas. Nivel A321 Determinar requerimientos.....	66
Tabla 34. Detalle de entradas y salidas. Nivel A322 Programar transacciones. ....	67
Tabla 35. Detalle de entradas y salidas. Nivel A323 Seguimiento de transacciones .....	67
Tabla 36. Detalle de entradas y salidas. Nivel A341: Producción de predicción y reportes.....	68
Tabla 37. Detalle de entradas y salidas. Nivel A342: Envío de mensajes y entrega de reportes. ....	69
Tabla 38. Especificación de atributos de base de datos inicial. ....	85
Tabla 39. Correlación de llamadas por día.....	89
Tabla 40. Correlación de visitas por día. ....	91
Tabla 41. Estructura del <i>Data Warehouse</i> creado.....	93
Tabla 42. Descripción de las variables añadidas a <i>Data Warehouse</i> .....	93
Tabla 43. Ejemplos de resultados para muestra A.....	98
Tabla 44. Configuraciones finales a probar en Muestra A .....	102
Tabla 45. Configuraciones finales a probar en Muestra B. ....	102
Tabla 46. Resultado óptimos por composición para muestras A y B.....	104
Tabla 47. Centroides para 4 <i>clusters</i> en Centro de Gravedad <i>Mall Plaza Vespucio</i> .....	118
Tabla 48. Correlación entre variables del modelo .....	119

Tabla 49. Comunas con menor participación según cluster .....	123
Tabla 50. Composición de tipo de visitas según cluster.....	125
Tabla 51. Descripción de <i>clusters</i> .....	126
Tabla 52. Variables generadas luego del <i>clustering</i> .....	127
Tabla 53. Proporción de datos .....	131
Tabla 54 . Tabla comparativa de desempeño .....	134
Tabla 55. Formula de precios.....	156
Tabla 56. Precios.....	156
Tabla 57. Estimación de usuarios iniciales .....	158
Tabla 58. Estimación de precios por campaña (CLP) .....	158
Tabla 59. Estimación de crecimiento del mercado en Chile .....	159
Tabla 60. Costos fijos anuales del proyecto (CLP).....	159
Tabla 61. Inversión Fija del proyecto (CLP).....	159
Tabla 62 . Flujo de caja puro anual en pesos chilenos (situación con proyecto) .....	161
Tabla 63 . Campañas necesarias con precio promedio .....	162
Tabla 64. Campañas necesarias con precio promedio de la competencia .....	162
Tabla 65 . KPI's según objetivos de negocios .....	169
Tabla 66. Centroides para de K-means similaridad coseno con una variable por día de semana. K=4.....	181
Tabla 67. Subclusterización K-means similitud coseno para días de semana. K=3 .....	182
Tabla 68. Información general de otros CG's .....	182
Tabla 69 . Flujo de caja sin proyecto en CLP .....	200
Tabla 70. Detalle de la inversión fija en estado sin proyecto (CLP).....	201

## INDICE DE FIGURAS

Figura 1. Proyecciones de inversión en publicidad digital para Estados Unidos.....	1
Figura 2. Gasto en Publicidad Chile (2015).....	2
Figura 3 . Homepages de aplicaciones segmentadoras .....	10
Figura 4. Homepages de aplicaciones no segmentadoras .....	11
Figura 5. Macroprocesos básicos .....	16
Figura 6. Lean Canvas .....	17
Figura 7. Función coseno.....	19
Figura 8. Validación por Holdout.....	25
Figura 9. Validación por Cross-Validation.....	26
Figura 10. Matriz de confusión .....	26
Figura 11. Curva ROC y criterio AUC .....	28
Figura 12. Organigrama funcional propuesto .....	29
Figura 13. Matriz FODA .....	30
Figura 14. Posicionamiento Estratégico según modelo Detal de Hax .....	33
Figura 15. Balanced Scorecard.....	36
Figura 16. Arquitectura básica general .....	41
Figura 17. Arquitectura de negocios A1 Planificación del negocio .....	44
Figura 18. Arquitectura de negocios A3 Desarrollo de campañas de marketing.....	48
Figura 19. Arquitectura de negocios A31: Administración de la relación con el cliente y usuario .....	53
Figura 20. Arquitectura de negocios A311: Análisis y campañas de marketing.....	56
Figura 21. Arquitectura de negocios A313: Venta y atención al cliente .....	59
Figura 22. Arquitectura de negocios A3131: Ventas .....	61
Figura 23. Arquitectura de negocios A3133: Seguimiento .....	64
Figura 24. Arquitectura de negocios A32: Administración de la relación con proveedores .....	66

Figura 25. Arquitectura de negocios A34: Producción y entrega .....	68
Figura 26. Proceso de Retroalimentación. En notación BPMN.....	71
Figura 27. Proceso de Retroalimentación de modelos. En notación BPMN.....	72
Figura 28. Proceso de Actualización de bases de producción. En notación BPMN.....	74
Figura 29. Proceso de <i>Opt-in</i> . En notación BPMN .....	75
Figura 30. Proceso de Captación de clientes. En notación BPMN .....	76
Figura 31. Proceso de Postventa. En notación BPMN.....	77
Figura 32 Proceso de Producción. En notación BPMN.....	79
Figura 33. Proceso de Ejecución del modelo de pronóstico. En notación BPMN .....	80
Figura 34. Proceso de Avances de campañas de marketing. En notación BPMN.....	81
Figura 35. Proceso de Finalización de campaña. En notación BPMN.....	83
Figura 36. Frecuencia de llamadas por mes .....	86
Figura 37.Frecuencia de llamadas por día. Mayo .....	87
Figura 38. Frecuencia de llamadas por día. Junio .....	87
Figura 39. Frecuencia de llamadas por día. Julio.....	87
Figura 40. Frecuencia de llamadas por día. Agosto.....	87
Figura 41. Frecuencia de llamadas por persona. Periodo mayo - agosto. ....	87
Figura 42. Muestra de tabla de llamadas.....	88
Figura 43. Frecuencia de llamadas por hora. ....	89
Figura 44. Frecuencia de visitas por hora.....	90
Figura 45. Frecuencia de visitas por persona, general. ....	90
Figura 46. Muestra A .....	92
Figura 47. Muestra B.....	92
Figura 48 . Muestra de base de datos DW_RegistroLlamadas.....	92
Figura 49. Relación de variables para muestra A. “EPS_cl” (abscisa), ruido (ordenada), “MinPts” (tamaño), valoresDU” (color).....	100

Figura 50. Relación de variables para muestra A. “MinPts” (abscisa), ruido (ordenada), “valoresDU” (tamaño), “EPS_cl” (color).....	100
Figura 51. Relación de variable para muestra A. “EPS_cl” (abscisa), ruido (ordenada), “MinPts” (color).....	101
Figura 52. Relación de variables para muestra A. “num_clusters” (abscisa), “valoresDU” (ordenada), “MinPts” (tamaño), “EPS_cl” (color).....	101
Figura 53. Relación de variables para muestra A. “EPS_cl” (abscisa), “MinPts” (ordenada), “num_clusters” (tamaño), “valoresDU” (color).....	101
Figura 54. Muestra A, histograma de “valoresDU” para composición 1.....	103
Figura 55. Muestra A, histograma de “valoresDU” para composición 2.....	103
Figura 56. Muestra A, histograma de “valoresDU” para composición 3.....	104
Figura 57. Muestra B, histograma de “valoresDU” para composición 1.....	104
Figura 58. Muestra B, histograma de “valoresDU” para composición 2.....	104
Figura 59. Muestra B, histograma de “valoresDU” para composición 3.....	104
Figura 60. Muestra A, Reachability Plot para OPTICS en composición 1.....	105
Figura 61. Muestra A, Reachability Plot para OPTICS en composición 2.....	105
Figura 62. Muestra A, Reachability Plot para OPTICS en composición 3.....	106
Figura 63. Muestra B, Reachability Plot para OPTICS en composición 1.....	106
Figura 64. Muestra B, Reachability Plot para OPTICS en composición 2.....	106
Figura 65. Muestra B, Reachability Plot para OPTICS en composición 3.....	106
Figura 66. Muestra A, DBSCAN para composición 1.....	107
Figura 67. Muestra A, DBSCAN para composición 2.....	107
Figura 68. Muestra A, DBSCAN para composición 3.....	108
Figura 69. Muestra B, DBSCAN para composición 1.....	108
Figura 70. Muestra B, DBSCAN para composición 2.....	109
Figura 71. Muestra B, DBSCAN para composición 3.....	109
Figura 72. K-means, distancia euclidiana K=3.....	111
Figura 73. K-means, similitud coseno K=3.....	111

Figura 74 . <i>Clustering</i> en espacio original (k=4). Variables vsab, “vvie” .....	112
Figura 75. <i>Clustering</i> en espacio transformado (k=4). Variables vsab, “vvie” .....	112
Figura 76. <i>Clustering</i> en espacio original (k=4). Variables “vfds”, “vsem” .....	114
Figura 77. <i>Clustering</i> en espacio transformado (k=4). Variables “vfds”, “vsem” .....	114
Figura 78. <i>Clustering</i> en espacio original (k=3). Variables vsab, “vsem” .....	114
Figura 79. <i>Clustering</i> en espacio transformado (k=3). Variables vsab, “vsem” .....	114
Figura 80. <i>Clustering</i> en espacio original (k=3). Variables “vdom”, “vsem” .....	115
Figura 81. <i>Clustering</i> en espacio transformado (k=3). Variables “vdom”, “vsem” .....	115
Figura 82. Configuración óptima de K-means con similitud coseno. ....	116
Figura 83. Configuración óptima de K-means con similitud coseno. ....	117
Figura 84. Configuración óptima de K-means con similitud coseno. ....	117
Figura 85. Configuración óptima de K-means con similitud coseno. ....	117
Figura 86. Método del codo, calculado con la mínima de los indicadores calculados para cada K. .....	118
Figura 87 . Comunas de residencia con alta frecuencia de visitas según <i>cluster</i> .....	121
Figura 88. Distribución geográfica de comunas con alta frecuencia de visitas para <i>cluster 1</i> ..	122
Figura 89. Distribución geográfica de comunas con alta frecuencia de visitas para <i>cluster 2</i> ..	122
Figura 90. Distribución geográfica de comunas con alta frecuencia de visitas para <i>cluster 3</i> ..	123
Figura 91. Distribución geográfica de comunas con alta frecuencia de visitas para <i>cluster 4</i> ..	124
Figura 92 . Ubicación de Centros de Gravedad para replicación .....	128
Figura 93. Óptimo K-means para <i>Mall Vespucio Norte</i> (1594 observaciones) K óptimo = 4 .....	129
Figura 94. Óptimo K-means para <i>Mall Plaza Oeste</i> (912 observaciones) K óptimo = 4 .....	129
Figura 95. Óptimo K-means para <i>Mall Arauco Maipú</i> (1077 observaciones) K óptimo = 4 .....	129
Figura 96. Óptimo K-means para <i>Mall Plaza Egaña</i> (165 observaciones) K óptimo = 4 .....	129
Figura 97. Óptimo K-means para <i>Mall Alto Las Condes</i> (1423 observaciones). K óptimo = 4	130
Figura 98. Óptimo K-means para unión de Centros de Gravedad (624.216 observaciones) K óptimo = 4 .....	130

Figura 99. Matriz de costos .....	132
Figura 100. Comparación curvas ROC .....	134
Figura 101. Mockup "Inicio de sesión" .....	138
Figura 102. Mockup "Ingresar datos" .....	138
Figura 103. Mockup "Ingresar preferencias" .....	138
Figura 104. Mockup "Home" .....	139
Figura 105. Reporte de ventas, composición de <i>clusters</i> por CG.....	140
Figura 106. Reporte de ventas, frecuencia de visitas por <i>cluster</i> /hora.....	140
Figura 107. Reporte de ventas, recomendación de horarios .....	141
Figura 108. Reporte de ventas, procedencia.....	141
Figura 109. Reporte de ventas, mapa de procedencia ( <i>cluster</i> 1) .....	142
Figura 110. Reporte de ventas, tipo de visitas <i>cluster</i> .....	142
Figura 111. Reporte final, desempeño de campaña .....	143
Figura 112. Diagrama de actores de aplicación computacional.....	143
Figura 113. Diagrama de despliegue y componentes combinado .....	145
Figura 114. Diagrama de casos de uso para Sistema 1: Sistema de Información de Usuarios ( <i>SIU</i> ).....	146
Figura 115. Diagrama de casos de uso para Sistema 2: Sistema de Clientes ( <i>SC</i> ) .....	147
Figura 116. Diagrama de casos de uso para Sistema 3: Sistema de Producción ( <i>SP</i> ) .....	147
Figura 117 . Matriz de confusión para prueba de concepto 1 .....	152
Figura 118 . Curva ROC para árbol de decisión en <i>cluster</i> 4.....	154
Figura 119. Matriz de confusión para prueba de concepto 2 .....	154
Figura 120. Posicionamiento de la empresa frente a usuarios .....	155
Figura 121. Proceso Captación de usuarios. En notación BPMN.....	173
Figura 122. Proceso de Reportes de avances. En notación BPMN .....	175
Figura 123. Proceso de Registro de pagos. En notación BPMN .....	176

Figura 124. Proceso de Retroalimentación de estrategia. En notación BPMN .....	177
Figura 125. Proceso de Retroalimentación operacional. En notación BPMN.....	178
Figura 126. Relación de variables para Muestra B. x = EPS_cl, y = ruido, size = num_clusters, color = valoresDU .....	179
Figura 127. Relación de variables para Muestra B. x = MinPts, y = ruido, size = valoresDU, color = EPS_cl. ....	180
Figura 128. Relación de variables para Muestra B. x = EPS_cl, y = ruido, color = MinPts. ....	180
Figura 129. Relación de variables para Muestra B. x = EPS_cl, y = MinPts, size = num_clusters, color = valoresDU. ....	181
Figura 130. Relación de variables para Muestra B. x = num_clusters, y = valoresDU, size = MinPts, color = EPS_cl. ....	181
Figura 131 . frecuencia de llamadas por mes VN.....	183
Figura 132. Frecuencia de llamadas por día de semana. Mayo. VN .....	184
Figura 133. Frecuencia de llamadas por día de semanao. Junio. VN .....	184
Figura 134. Frecuencia de llamadas por día de semana. Julio. VN .....	184
Figura 135. Frecuencia de llamadas por día de semana. Agosto. VN.....	184
Figura 136. Frecuencia de llamadas por persona VN.....	185
Figura 137. Frecuencia de llamadas por hora VN.....	185
Figura 138. Correlación de llamadas VN.....	185
Figura 139. Frecuencia de visitas por persona VN .....	186
Figura 140. Frecuencia de visitas por hora VN .....	186
Figura 141. Correlación de visitas VN .....	186
Figura 142. Frecuencia de llamadas por mes (PO).....	187
Figura 143. Frecuencia de llamadas por día de semana. Mayo. PO.....	187
Figura 144. Frecuencia de llamadas por día de semana. Junio. PO.....	187
Figura 145. Frecuencia de llamadas por día de semana. Julio. PO. ....	187
Figura 146. Frecuencia de llamadas por día de semana. Agosto. PO. ....	187
Figura 147. Frecuencia de llamadas por persona. PO.....	188



Figura 148. Frecuencia de llamadas por hora. PO.....	188
Figura 149. Correlación de llamadas PO.....	188
Figura 150. Frecuencia de visitas por persona. PO .....	189
Figura 151. Frecuencia de visitas por hora. PO .....	189
Figura 152. Correlación de visitas PO .....	189
Figura 153. Frecuencias de llamadas por mes PM.....	190
Figura 154. Frecuencia de llamadas por día. Mayo. PM.....	190
Figura 155. Frecuencia de llamadas por día. Junio. PM.....	190
Figura 156. Frecuencia de llamadas por día. Julio. PM .....	190
Figura 157. Frecuencia de llamadas por día. Agosto. PM .....	190
Figura 158. Frecuencia de llamadas por persona PM.....	191
Figura 159. Correlación de llamadas .....	191
Figura 160. Frecuencia de visitas por persona PM .....	192
Figura 161. Frecuencia de visitas por hora PM.....	192
Figura 162. Correlación de visitas PM .....	192
Figura 163. Frecuencia de llamadas por mes PE.....	193
Figura 164. Frecuencia de llamadas por día. Mayo. PE. ....	194
Figura 165. Frecuencia de llamadas por día. Junio. PE.....	194
Figura 166. Frecuencia de llamadas por día. Julio. PE.....	194
Figura 167. Frecuencia de llamadas por día. Agosto. PE.....	194
Figura 168. Frecuencia de llamadas por personas PE .....	195
Figura 169. Frecuencia de llamadas por hora PE .....	195
Figura 170. Correlación de llamadas PE .....	195
Figura 171. Frecuencia de visitas por persona PE.....	196
Figura 172. Frecuencia de visitas por hora PE.....	196
Figura 173. Correlación de visitas PE.....	196

Figura 174. Frecuencia de llamadas por mes ALC .....	197
Figura 175. Frecuencia de llamadas por día. Mayo. ALC. ....	198
Figura 176. Frecuencia de llamadas por día. Junio. ALC. ....	198
Figura 177. Frecuencia de llamadas por día. Julio. ALC.....	198
Figura 178. Frecuencia de llamadas por día. Agosto. ALC.....	198
Figura 179. Correlación de llamadas ALC.....	198
Figura 180. Frecuencia de llamadas por personas ALC .....	199
Figura 181. Frecuencia de llamadas por hora ALC.....	199
Figura 182. Frecuencia de visitas por personas ALC.....	199
Figura 183. Frecuencia de visitas por hora ALC .....	199
Figura 184. Correlación de visitas ALC .....	200

# CAPÍTULO 1: INTRODUCCIÓN Y CONTEXTO

## 1.1 Antecedentes de la Industria

El Marketing Móvil corresponde a una de las líneas del marketing en el cual se realizan promociones de productos y servicios a través de un dispositivo móvil personal, esto es publicidad y difusión por mensajería (SMS y MMS), internet móvil, voz y música, y aplicaciones.

Según un estudio de *eMarketer* [4], durante el 2015 la inversión en marketing móvil en Estados Unidos bordeó los USD\$30,45MM, lo cual representa un aumento del 59% respecto del año anterior, mientras que la inversión en publicidad de escritorio va a la baja, como se observa en la Figura 1. Esta inversión es una señal importante de que la industria digital se ha convertido en un medio clave de difusión y promoción para las empresas e indica que el marketing móvil es una industria sólida, aunque aún en evolución. Las empresas deberán entrar cada vez más al uso del marketing móvil, como complemento y sustituto del marketing tradicional, de hecho se espera que para fines de 2016 el gasto en marketing móvil haya alcanzado los USD\$42MM duplicando a la publicidad para computadores en 2017 con un gastos estimado de USD\$51MM, sin embargo esta forma de hacer marketing deberá ser tratada de tal forma de no generar un rechazo por los clientes, la clave es la personalización.

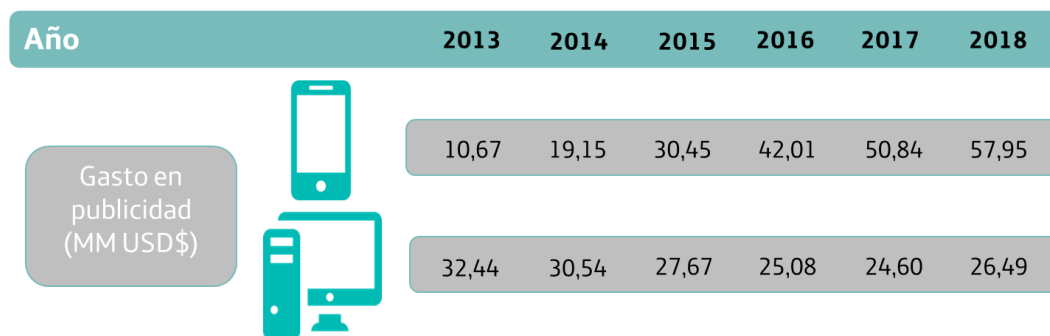


Figura 1. Proyecciones de inversión en publicidad digital para Estados Unidos

En Chile, la inversión publicitaria *online* (que incluye el gasto en publicidad móvil y para computadores) alcanzó los CLP\$105MM lo cual representa un incremento del 25% con respecto a 2014 [2]. Esto indica que Chile está alineado con la tendencia mundial de la inversión en marketing móvil, donde la publicidad online aumenta su participación en el total del gasto en publicidad, como se observa en la Figura 2, con un 13,7% de participación, siguiendo a la televisión (37,4%) y los diarios (23,8%).

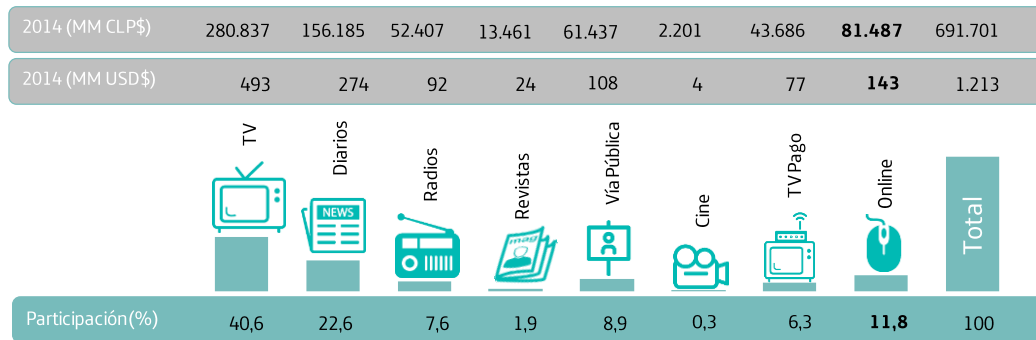


Figura 2. Gasto en Publicidad Chile (2015)

Según un estudio de *IDG Global Solution* [5], el 78% de los usuarios de teléfonos inteligentes en Asia Pacífico han realizado al menos una compra en el comercio móvil, lo siguen Norte América con el 70% de los usuarios, Medio Oriente y África (58%), Latino América (52%), Europa Oriental (52%) y Europa Occidental (48%), sin embargo estas cifras no están normalizadas, por lo que zonas con baja penetración de teléfonos inteligentes como Asia Pacífico muestran valores más altos. Considerando lo anterior se observa que América Latina destaca como la región donde más compras se realizan vía *m-commerce*. En Chile el 59% de los usuarios de *smartphone* dicen haber usado su teléfono para conseguir descuentos en productos mientras están en una tienda, esto lo realizan a través de aplicaciones que permiten comprar a través de dispositivos, en Chile las aplicaciones líderes son *Hellofood*, *Dafiti*, *Izit*, *HappyShop*, *Fainder*, entre otras.

Un 87% de las interacciones de los chilenos se dan a través de dispositivos móviles, donde los *smartphones* son la primera tendencia con un 36% de las interacciones (lo cual se traduce en un promedio de 3.5 horas diarias de navegación en estos aparatos), lo siguen los computadores personales (31%), televisión (26%) y *tablets* (4%) [6].

Se espera que en 2017 haya 35 millones de aparatos de cualquier tipo conectados a internet móvil en el país, de los cuales sólo el 28% serán teléfonos no inteligentes y el 48% *smartphones* [7]. El marketing móvil tiene, entonces, grandes oportunidades para llegar a potenciales clientes

## 1.2 Oportunidad Identificada

A continuación se describen las oportunidades de mercado para el desarrollo del proyecto

1. Aparición del Geobusiness o geomarketing: el *geomarketing* o *geobusiness* es una herramienta que integra manejo de gran cantidad de datos georreferenciados (analizar los datos de una región geográfica o de una locación física), los cuales entregan información útil para apoyar la toma de decisiones.

Hoy en día las compañías que tienen acceso a los datos de locación como las compañías telefónicas, proveedores de aplicaciones con acceso al *GPS* de los teléfonos inteligentes, entre otros tienen la oportunidad de promocionar sus productos mediante una nueva vía, más directa y relacionada con el estado real de los clientes.

2. Aumento de la inversión en publicidad móvil: el mercado del marketing móvil está creciendo entre 6 a 7 veces más rápido que el marketing para computadores [8]. En la línea a lo anterior y de lo comentado en el punto 1.1, se observa un aumento en el crecimiento del marketing móvil dando paso a la publicidad digital como tipo de marketing preferente tanto para los usuarios de teléfonos móviles como para las empresas, ya que este tipo de marketing muestra una mayor tasa de conversión en ventas que otros (sin ir más lejos, en España se observó que la tasa de conversión de la publicidad móvil duplica a la tasa de conversión de la publicidad de escritorio [9]).
3. Poca diversidad en la oferta actual: dado que el mercado del *geobusiness* y la tendencia de analizar los datos geo-referenciados de los clientes son técnicas que aún se encuentran en etapa de desarrollo, no se observan servicios que combinen estas dos disciplinas para realizar un proceso de compra. La oferta actual se centra en el uso de la tecnología 3G (y en algunos casos Realidad Aumentada) para encontrar la localización de sus usuarios, además las segmentaciones realizadas dependen de la disposición de los usuarios de entregar su información de preferencias.
4. Aumento de las funcionalidades e imprescindibilidad de teléfonos móviles: la tecnología de hoy en día ha hecho que el teléfono celular se vuelva un aparato imprescindible en la vida de sus usuarios [10], sin ir más lejos de acuerdo al estudio de mercado de *Flurry Analytics* y *Statista*, el uso de las aplicaciones móviles de los teléfonos inteligentes aumentaron en un 115% durante el año 2013, donde las aplicaciones de compras tienen un aumento de 78%.
5. Existencia de métodos de inferencia de la actividad de las personas: se ha demostrado la factibilidad de establecer sistemas para predecir la actividad de las personas utilizando como características base la ubicación y hora del día [11].

Se concluye la existencia de un ambiente apto para generar productos que aprovechen la información de posiciones relativas y uso de tiempo de las personas, para entregar beneficios de forma no invasiva, asertiva y sin requerir de la dependencia del uso constante de una aplicación para obtener regalías. Esto responde al avance de la industria 4.0, donde el *internet* de las cosas (IoT) se hace presente en el mercado de servicio y manufactura, usando modelos de predicción o captura de señales de eventos para actuar en menor tiempo, y con ello mejorar la productividad (disminuyendo costos operativos) y servicio a clientes (aumentando la personalización de los servicios).

Aún más importante, es la observación de un mercado de marketing móvil creciente, con poco entrantes, donde las ventajas competitivas se deben generar a partir de la comodidad y seguridad que se dan a los usuarios y que aumentarán su fidelidad en el uso de estos productos, aumentando con ello las conversiones de posibles clientes.

## **1.3 Objetivos y Resultados Esperados del Proyecto**

### **1.3.1 Objetivo General**

El objetivo de este trabajo es generar la capacidad de enviar promociones correctas a los usuarios correctos, en el momento y lugar correcto, con una sensibilidad del 60%, en un periodo de 1 año.

### **1.3.2 Objetivos Específicos**

Para lograr el objetivo general, se definen los siguientes objetivos específicos, con fecha esperada de cumplimiento noviembre de 2016:

1. Asegurar, en al menos un 60%, la predicción de visitas a una zona geográfica de alta densidad comercial, desde ahora denominada Centro de Gravedad comercial (CG).
2. Obtener predicciones ejecutables en no más de 1 día laboral.
3. Clasificar a usuarios según datos registrados desde cualquier Centro de Gravedad que contenga a un *mall*, extraídos de bases de datos que registran llamadas telefónicas, CDR por sus siglas en inglés (*Call Detail Record*) de la empresa de telefonía socia.
4. Desarrollar planes de negocios, para una empresa de envío de promociones personalizadas y geolocalizadas, en los ámbitos de marketing y operaciones basados en un estudio cuantificable del marketing móvil.
5. Desarrollar la estructura de procesos de la cadena de valor de una empresa de envío de promociones personalizadas y geolocalizadas.

### **1.3.3 Resultados Esperados**

Este proyecto consta del tratamiento de datos de llamadas telefónicas, para analizar comportamientos y con ello lograr la creación modelos de *clustering* y predicción o clasificación. A partir de estos modelos se tendrá un entendimiento de los tipos de visitas y se pronosticarán los comportamientos de estos grupos segmentados, el pronóstico de visitas será el *input* para sostener una plataforma de envío de promociones personalizadas y geolocalizadas.

Se esperan lograr tanto los modelos de *clustering* como los que permitan un pronóstico con una efectividad de 60%, donde se entiende como efectividad el pronóstico correcta de los usuarios que realmente visitan un Centro de Gravedad, este efecto se captura con la medida de desempeño *recall* explicada en el punto 2.2.8. Los modelos generarán la capacidad del envío de promociones en el momento y lugar correcto por usuario

Como parte de la generación de capacidades de negocios, se obtendrán los diseños de procesos de la cadena de valor y planes de marketing y operaciones de la empresa mencionada anteriormente.

Por último, se espera generar un prototipo reportería que muestre información relevante al resultado de campañas, con métricas de interés para los potenciales clientes de este negocio y el desarrollo a nivel de *mockups* de la plataforma *web* dedicada a usuarios.

Tabla 1. Resultados esperados y entregables

<b>Id</b>	<b>Resultado</b>	<b>Entregable</b>
1	Modelos de <i>clustering</i>	Códigos en <i>software</i> R
		Resultados de análisis de <i>clustering</i> y categorización de visitantes a un Centro de Gravedad
2	Modelos de pronóstico	Códigos en R
		Resultados de análisis de modelos de pronóstico
		Prueba de concepto con resultado de efectividad del modelo de pronóstico
3	Generación de capacidades de negocios	Procesos de la cadena de valor (macroprocesos y procesos de negocio)
		Plan de <i>Marketing</i>
		Plan de Operaciones
4	Prototipo	Reportes de resultados de campañas para clientes
		<i>Mockups</i> de aplicación <i>web</i> para usuarios

#### 1.4 Alcance

El proyecto abarcará:

- El desarrollo de un modelo de *clustering* que permita identificar a los clientes según las posiciones relativas que tienen durante el día.
- El desarrollo de modelos supervisados (predicción y/o clasificación) que permitan generar pronósticos de visitas de los usuarios de la aplicación.
- El desarrollo de un prototipo de interfaz usuaria móvil o *web* que reciba información de las preferencias de los usuarios (de ser necesario).
- El desarrollo de un prototipo de reportes de ventas y desempeño de una campaña de marketing móvil
- El diseño de los macroprocesos y procesos de negocios de una empresa que entregue el servicio de información de posibles clientes a empresas como cadenas de comida, tiendas de ropa, accesorios, bancos, etc. Se desarrollará el el macroproceso 1 en detalle (correspondiente a la cadena de valor de la empresa, “Desarrollo de campañas publicitarias”, y los procesos de negocios correspondientes a la cadena de valor de la empresa.
- El plan de negocios de una empresa como la descrita anteriormente.

Un CDR es un registro de los aspectos involucrados en una llamada, número de entrada, número de salida, antenas de conexión de estos aparatos, fecha y duración de la llamada. Para este trabajo se utiliza un base anonimizada, por lo que los campos número serán reemplazados por un identificador de usuario.

Los modelos que sustentan el proyecto se generan con datos de los meses de mayo, junio, julio y agosto de 2014, esto implica que los pronósticos estarán diseñados para predecir comportamientos basados en la historia de este periodo de tiempo, lo que podría generar un sesgo importante en el comportamiento real de los usuarios en el periodo actual de medición. La capacidad de extracción de datos en tiempo real de la empresa de telefonía se está aumentando, pero aún no es posible tener esta información para ajustar los modelos, por lo que el proyecto no será implementado ni testeado con envío de mensajes.

## **1.5 Aspectos legales**

A continuación se definen los aspectos legales que afectan la implementación de un proyecto que pronostica información georreferenciada según datos de telefonía para una empresa de marketing móvil.

### **1.5.1 Ley 19.496, Ley de Protección al Consumidor**

Esta ley define como consumidor o usuarios a las personas naturales o jurídicas que utilizan o disfrutan como destinatarios finales bienes o servicios y los distingue de los proveedores, personas naturales o jurídicas de carácter público o privado, que habitualmente presten desarrollen actividades de producción, fabricación, importación, construcción, distribución o comercialización de bienes o prestación de servicios a consumidores, por las que se cobre precio o tarifa.

Se establece la información básica comercial que debe suministrarse de forma obligatoria al público consumidor:

- Identificación del bien o servicio que se ofrece al consumidor: en relación al proyecto se detecta la información de tipo de envío, tipo de informe de desempeño, métricas a entregar, etc.
- Instructivos de uso y los términos de la garantía cuando procedan

La información comercial básica deberá ser suministrada al público por medios que aseguren un acceso claro, expedito y oportuno.

Se establecen los derechos y deberes del consumidor, de los cuales se desatacan:

- La libre elección del bien o servicio. El silencio no constituye aceptación en los actos de consumo.



- El derecho a una información veraz y oportuna sobre los bienes y servicios ofrecidos, su precio, condiciones de contratación y otras características relevantes de los mismos, y el deber de informarse responsablemente de ellos.
- El no ser discriminado arbitrariamente por parte de proveedores de bienes y servicios.

Estos se relacionan con el proceso de *Opt-in*, que se describe en el punto 4.3.5, y que permite el acceso consentido de los usuarios al servicio de promociones y ofertas móviles, con entrega de la información necesaria del funcionamiento del servicio y que abarca a todos los potenciales usuarios sin discriminación, simplemente segmentando el producto en base a un modelo de segmentación no arbitrario.

Además se indica que “(...) *en los contratos celebrados por medios electrónicos, y en aquéllos en que se aceptare una oferta realizada a través de catálogos, avisos o cualquiera otra forma de comunicación a distancia, el consentimiento no se entenderá formado si el consumidor no ha tenido previamente un acceso claro, comprensible e inequívoco de las condiciones generales del mismo y la posibilidad de almacenarlos o imprimirlos*”.

Por otro lado, se definen las obligaciones del proveedor, de los cuales se destacan:

- Respetar los términos, condiciones y modalidades conforme a las cuales se hubiere ofrecido o convenido con el consumidor la entrega del bien o la prestación del servicio.
- Enviar una confirmación de los contratos, una vez perfeccionados, por vía electrónica o cualquier medio que garantice el conocimiento oportuno del consumidor.
- Los correos electrónicos para la obtención de consumidores deben poseer un título y remitente informativos.
- Dar conocimiento público de los precios con excepción de los que por sus características deban regularse convencionalmente.
- Señalar por escrito en la boleta, recibo u otro documento, el plazo por el cual se hace responsable del servicio o reparación.

La ley decreta que la mensajería promocional no debe ser canal de información falsa además el artículo 43 señala que “ (...) *el proveedor que actúe como intermediario en la prestación de un servicio responderá directamente frente al consumidor por el incumplimiento de las obligaciones contractuales, sin perjuicio de su derecho a repetir contra el prestador de los servicios o terceros que resulten responsables*”.

Al respecto se tiene que es el cliente quien decide la promoción u oferta a enviar a los usuarios por lo cual es el responsable de hacer valer estos mensajes en sus tiendas

comerciales. Sin embargo, y dado que el negocio no se sustenta sin usuarios, se contempla un proceso de atención de reclamos que generará un input en la relación con los clientes.

Sobre las promociones y ofertas se indica que en todas se debe:

- Entregar las bases de las promociones y/u ofertas y el tiempo de duración de esta.
- *“Cuando se trate de promociones en que el incentivo consista en la participación en concursos o sorteos, el anunciante deberá informar al público sobre el monto o número de premios de aquéllos y el plazo en que se podrán reclamar. El anunciante estará obligado a difundir adecuadamente los resultados de los concursos o sorteos”.*

En relación a lo anterior, se observa que esta información debe ser entregada por el cliente y debe ser solicitada por la empresa como requisito para poder enviar los mensajes a los usuarios.

### **1.5.2 Ley 19.628, Ley de Protección de Datos de carácter personal**

Esta ley indica que el tratamiento de los datos personales solo puede efectuarse cuando la ley lo autorice o bien cuando el titular lo consienta expresamente, además la persona que autoriza debe ser informada del propósito del almacenamiento y posible comunicación de los datos al público. Esta autorización debe estar por escrito y puede ser revocada, lo cual también debe hacerse por escrito.

En efecto, el proyecto contempla que, en alianza con la TELCO, el usuario sea informado del fin del uso de datos, el período de uso y las condiciones de estas. Por otro lado, por los procesos de *Opt-in* y *Opt-out* los usuarios podrán expresar explícitamente tanto la suscripción al servicio (y correspondiente uso de sus datos) y la finalización de la relación con la empresa (eliminando sus datos del registro a ser usado para el producto).

No se requiere de la autorización del titular de los datos si la información se obtiene desde fuentes accesibles al público, de carácter económico, financiero, bancario o comercial, y cuando sea información básica utilizada para la comunicación comercial.

Aquellos organismos, ya sean públicos o privados, que trabajan con el tratamiento de datos, deben guardar secreto sobre los mismos, cuando provengan de fuentes no accesibles al público, asimismo con los datos y antecedentes relacionados con el banco de datos. Esta obligación no cesa al terminar las actividades en este campo. Además, en ningún caso se pueden comunicar los datos de personas identificadas o identificables.

Sobre la transmisión de los datos entre el aliado y la empresa se establece que debe estar regido por un proceso automatizado en el cual se deje constancia de: la individualización del requirente, el motivo del requerimiento y el tipo de datos que se transmiten. Además el receptor sólo podrá utilizar los datos para los fines motivacionales descritos en el proceso.

Sobre los derechos de los titulares de los datos se destaca:

- Toda persona tiene derecho a exigir que se indique de forma pública o privada el propósito del almacenamiento o la individualización de las personas u organismos a los cuales sus datos son transmitidos regularmente.
- El titular de los datos podrá exigir que se elimine el almacenamiento de sus datos, si este carece de fundamento legal o cuando estén caducos.
- Cuando el titular no desee continuar en el registro y aquellos datos fueron proporcionados voluntariamente o se utilizan para comunicaciones comerciales, estos podrán ser eliminados o bloqueados a petición del titular.
- La eliminación debe ser gratuita para el titular y debe entregarse una copia del registro de la eliminación.
- El artículo 13 indica que *“El derecho de las personas a la información, modificación, cancelación o bloqueo de sus datos personales no puede ser limitado por medio de ningún acto o convención”*.

## **1.6 Principales competidores**

Según los atributos del proyecto los posibles competidores son servicios que entregan información atractiva para los clientes, servicios geolocalizados móviles y/o que entregan algún tipo de almacenamiento de datos que permitan hacer segmentación de los clientes. En este sentido se observan los siguientes competidores:

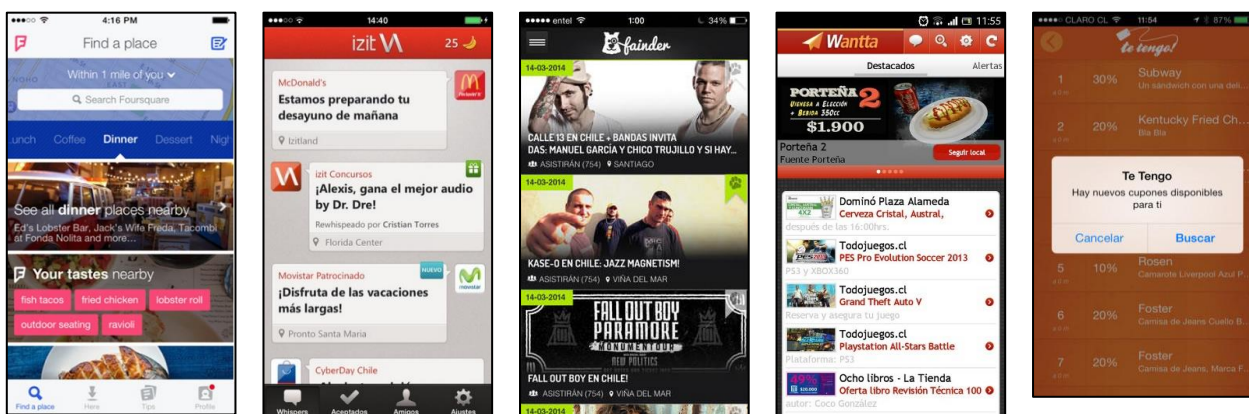
### **1.6.1 Servicios de aplicaciones móviles**

Este grupo representa los competidores principales, empresas que se dedican a entregar ofertas a quienes descargan y se registran en su aplicación Aquellas que perfilan a los usuarios por preferencias se les llamará segmentadoras.

- Foursquare: Aplicación creada en 2009, en la cual los usuarios pueden marcar los lugares donde se encuentran, a partir de esto la aplicación hace recomendaciones de lugares por visitar y entrega premios que sirven como incentivo para seguir usando la aplicación, funcionando como un motor de recomendaciones que aprende con la utilización de los usuarios.
- Izit: Aplicación chilena gratuita, lanzada en 2013, la cual utiliza una técnica de geoespaciamento in-store. Utiliza una segmentación en base a preferencias que

indica el usuario al momento de descargar la aplicación. Para que los descuentos sean enviados el usuario debe tener la aplicación abierta al entrar a una tienda.

- Fainder: Plataforma gratuita (sólo disponible en *AppStore*) en la cual los usuarios pueden buscar eventos y panoramas. La cartelera mostrada puede ser filtrada según las preferencias del usuario que indica sus intereses.
- Wantta: Aplicación gratuita que entrega ofertas según la ubicación geográfica del usuario utilizando la señal que el *GPS* del teléfono envía. Las ofertas son personalizadas según las preferencias establecidas por el usuario en la aplicación.
- Te Tengo: Aplicación móvil gratuita que hace interactuar al usuario en la búsqueda y obtención de cupones de descuento repartidos en un área geográfica acotada y determinada, mostrando descuento solo cuando el usuario se encuentra cerca una tienda, los descuentos son filtrados según los intereses del usuario y se entregan mediante la tecnología de realidad aumentada.



Foursquare

Izit

Fainder

Wantta

Te Tengo

Figura 3 . Homepages de aplicaciones segmentadoras

Dentro del grupo de competidores no segmentadores de usuarios se encuentran:

- Dongo: Plataforma que entrega beneficios de empresas a las que el usuario ya es socio, como clubes de lectura, compañías de teléfono, bancos, etc. El aviso llega según la geolocalización del usuario.
- Justter: Aplicación desarrollada por los creadores de *Mapcity*, aprovecha su información como participante principal del *geobusiness* en Chile para entregar ofertas geolocalizadas, la localización del usuario se consigue cuando este entra a la aplicación o bien escribiendo una dirección desde un computador. Las ofertas entregadas pueden ser filtradas por categorías y la aplicación es capaz de generar cupones de descuento de las ofertas mostradas.
- Around Me: Aplicación que permite buscar los servicios más cercanos a la ubicación del usuario como restaurantes, bancos, librerías, entre otros. Funciona gracias a la realidad aumentada utilizando la señal de *GPS* que entrega el teléfono.

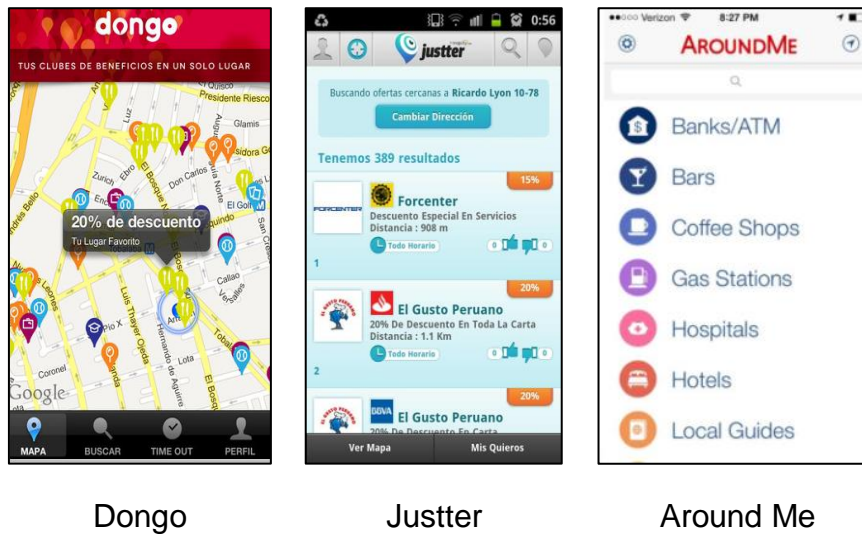


Figura 4. Homepages de aplicaciones no segmentadoras

### 1.6.2 Servicios de club de beneficios

Programas de beneficios para los afiliados a una empresa de manera de premiar la fidelidad y permanencia del cliente. Estos clubes de beneficios se subdividen según industria, en este aspecto se tienen los clubes de beneficios de bancos, diarios, *retail*, salud y telefonía.

Este grupo incluye también los descuentos entregados por Metro y beneficios que entrega *Lanpass*.

Estos beneficios no son personalizados, pero tienen el potencial de recopilar información de sus clientes para realizar segmentaciones y entrar como competidores.

### 1.6.3 Servicios de descuentos móviles

Estos servicios corresponden a plataformas que entregan cupones de descuento a las personas que se inscriben en su página. No realizan segmentación de sus clientes al menos en sus aplicaciones móviles, pero si se puede filtrar el contenido de los cupones según comunas.

Hoy en día se han expandido horizontalmente desde vender cupones de descuento para servicios de belleza, comida, cines, entre otros a vender productos online.

Empresas como *Groupon*, *Queremos Descuentos*, *Letbonus*, *Cuponatic*, entre otros son de esta categoría y no representan una competencia fuerte ya que no personalizan la gama de productos que muestra a sus clientes, pero podría hacerlo en el futuro. También podrían funcionar como socios clave para la nueva empresa.

## 1.7 Dinámica Competitiva

El mercado del marketing móvil va muy ligado al de *e-commerce* (comercio electrónico). En general los sitios de comercio y marketing electrónico tienen como característica principal que los sitios más grandes o ricos se enriquecen debido a que son la opción preferencial de los clientes, esto es el efecto "*winner-takes-all*". Por lo mismo es fácil que ocurra el fenómeno de monopolio, es decir las barreras de entrada son altas para los nuevos sitios.

Los sitios deben aumentar sus ingresos y disminuir sus pérdidas a través de la mejora de la calidad de su servicio y la competencia se produce en la publicidad y diferenciación de los sitios.

Dado que se trata de un mercado nuevo, existe mucha variabilidad en la dinámica competitiva de las empresas que lo componen. Sin embargo se observa que:

- Cuando hay menos clientes comunes la competencia es más violenta, es en estos casos cuando se produce el efecto "*winner-takes-all*".
- La competencia es débil cuando hay más clientes comunes.

En general, se observa que la competencia entre empresas de marketing móvil tiene un gran mercado común, debido a que se diversifican en los productos a ofrecer, del mismo modo, dado que la oferta actual es similar, se observa que sus recursos son comparables (uso de datos 3G, uso de información de suelo, ya sea específico para una(s) empresa(s) o generalizando los productos y servicios del mapa).

Las empresas tienen un mercado común, por lo que son interdependientes. Sin embargo, aún podría no haber un reconocimiento de este. Estas empresas tienen motivación a realizar acciones competitivas debido a la relación con las utilidades que se podrían percibir, dado que el mercado está en rápida expansión, es probable que la motivación a actuar competitivamente aumente en el tiempo, esta acción se verá suavizada debido a la gran interdependencia entre las firmas de este mercado.

Hasta el momento las empresas no poseen gran capacidad de ataque y contraataque ya que se trata de pequeñas empresas (en general) que no poseen gran capital para aventurarse en atacar, al mismo tiempo el riesgo de atacar y responder es menor, ya que son mercados muy diversificados donde el dinero se mueve en varias firmas. Por otro lado, al tener recursos similares se observa que las empresas tienen capacidad de ataque y contraataque semejante. Por lo tanto las acciones competitivas deben ser muy bien estudiadas antes de realizarlas, se observa que la competencia tiene tendencia a aumentar rápidamente.

Este mercado tenderá a ser de ciclo estándar a rápido, con ataques y respuesta en horizontes de tiempo menores a 5 meses, que se basa en los avances tecnológicos, esto provoca que las decisiones estratégicas sean complejas y sea difícil mantener las

ventajas competitivas durante mucho tiempo, por esto los beneficios de ser primer jugador podrían ser enormes, cruciales para el éxito, pero no absolutos.

Se espera que el mercado sea relativamente volátil, donde las innovaciones sean el motor del crecimiento de las empresas. Los precios deberían bajar a gran velocidad, por lo que las empresas deberán obtener rápidamente utilidades de las innovaciones de sus productos. Además, será relativamente fácil la imitación de muchos productos del mercado, por lo que la probabilidad de ataque y contraataque será relativamente alta.

## **1.8 Riesgos Potenciales**

### ***1.8.1 Precisión de la predicción según el tipo de data***

Los modelos a trabajar utilizan datos asociados a la antena a la cual se ha conectado el usuario para hacer/recibir una llamada o mensaje, sin embargo no muestran con precisión la posición de los usuarios, sino una ubicación relativa, en la cual el usuario podría estar en un radio de 750 metros (alrededor de 6 cuadras) de la antena a la que se conectó. Además, si la antena más cercana al usuario se encuentra sobredemandada, entonces la llamada se hará a través de la siguiente antena más cercana disponible, por lo tanto la bola de ubicación probable se verá afectada tanto por la antena de conexión como por el tráfico en antenas cercanas.

Por otro lado, se tienen datos de CDR's acotados en tiempo. Esto implica sesgos en la predicción de comportamientos en la actualidad, además se sabe que hoy en día el mapeo de antenas en la ciudad es distinto al existentes en 2014, por lo que podría haber bastante información perdida.

Como medidas de mitigación se considera que la TELCO se encuentra trabajando en la disminución del tiempo de desfase de los datos, con el objetivo de que este se reduzca a hora o minutos desde el próximo año. Al mismo tiempo, se ha convenido la obtención de los CDR de datos móviles los cuales poseen una alta frecuencia de registro a las antenas de internet, gracias a esto estimar la ubicación del usuario se verá facilitado ya que se tendrán para ajustar los modelo en el futuro.

Conjuntamente, se han definido los Centros de Gravedad Comercial (CG). Estas zonas de alta densidad comercial son abordados por al menos una antena telefónica por lo que el primer filtro se hará en base a la frecuencia en que las personas entran en el centro de gravedad comercial y no en base al seguimiento continuo de cada usuario, gracias a esto se eliminará el riesgo de la precisión de la ubicación y se reducirá el costo de procesamiento.

### ***1.8.2 Privacidad de la ubicación***

Los usuarios exigen privacidad, consideran la información sobre su ubicación como información privada y no desean que esta sea obtenida por otras personas o empresas.

La sensación de inseguridad podría ser aún mayor si sienten que su empresa de telefonía conoce todos sus movimientos.

El proceso de *Opt-in* y *Opt-out* y su correcta promoción serán factores clave para mantener conforme a los usuarios. Existen distintas formas de aplicarlos:

- Solicitar *Opt-in* a los clientes mediante comunicación previa con ellos en la que se presenten las condiciones del servicio y donde estos deban ser aceptados para poder suscribirse al sistema de promociones.
- Realizar campañas de suscripción constantemente dentro de los clientes de la TELCO, donde se permita optar a beneficios personalizados al hacer *Opt-in* (en el cual se solicitará información adicional de preferencias). Al hacer *Opt-out* el nuevo usuario dejará de recibir mensajes de las campañas de suscripción.

Además el sistema deberá tener claramente detallado el proceso de *Opt-out* y facilitarle al usuario su salida asegurando que sus datos no serán usados de ninguna manera una vez desvinculado del servicio.

### **1.8.3 Aspectos organizacionales**

A pesar de que los objetivos operacionales entre el área de investigación y desarrollo de la empresa de telefonía y el centro de investigación donde se realiza el proyecto se encuentran alineados, se observaron incongruencias entre los objetivos del área de I+D y la línea central de la TELCO, la cual busca generar rentabilidad en periodos de tiempo menores a los usados por proyectos que invierten en investigación. Este punto podría retrasar el avance del proyecto debido a que al no cumplir con las expectativas de tiempo, los recursos (monetarios, de personal, tiempo de trabajo, entre otros) entregados podrían verse reducidos.

Este riesgo no es directamente manejable desde el centro de investigación, sin embargo se realizan las gestiones con el área de investigación y desarrollo para dar preponderancia al proyecto. En este sentido, el área de I+D se encuentra bastante interesada y se encargará de mantener el flujo de recursos necesarios para el desarrollo e implementación del proyecto.



## **CAPÍTULO 2: MARCO TEÓRICO**

---

### **2.1 Metodología de Ingeniería de Negocios**

#### **2.1.1 Metodología de procesos**

Esta metodología establecida por Oscar Barros [1], entrega modelos conceptuales que hacen posible la integración de la gestión de empresas y las TI, para esto usa patrones de procesos que deben practicar las empresas.

Los patrones de procesos están orientados al negocio y son de carácter normativo, se basan en la experiencia y las mejores prácticas, y no son secuenciales, sino que demuestra las actividades que una empresa debe hacer para obtener resultados que cumplan efectivamente con los objetivos de negocio.

Se detectan cuatro patrones de procesos principales llamados macroprocesos:

1. El primero es el patrón Cadena de Valor (macro 1), demuestra el funcionamiento central de la o las líneas de negocio de la empresa y está compuesto por cinco procesos: “Administración de la relación con el cliente”, “Administración de la relación con proveedores”, “Gestión de producción y entrega”, “Producción y entrega” y “Mantenimiento de estado”.
2. El macroproceso 2, Desarrollo de Nuevas Capacidades, establece los elementos para desarrollar nuevos procesos y herramientas que ayuden al negocio a cumplir con sus objetivos estratégicos. Se forma de los procesos de “Evaluación de necesidad de nueva capacidad”, “Gestión, diseño y construcción de nueva capacidad”, “Diseño y construcción de nueva capacidad” y “Mantenimiento de estado”.
3. La Planificación del Negocio trabaja con la estrategia de la empresa, es aquí donde se exploran las oportunidades de negocios, se definen la misión, visión y objetivos, se proponen modelos de evaluación y se desarrollan y ejecutan los planes estratégicos que cumplen con los objetivos del negocio. Se subdivide en cuatro procesos: “Definir el concepto del negocio y visión”, “Gestionar el desarrollo de la estrategia”, “Desarrollo de la estrategia”, y “Mantenimiento de estado”.
4. Por último el cuarto macroproceso es el de Gestión de Recursos Habilitadores en el cual se obtienen recursos, se decide el manejo del recurso, se deciden la transferencia de estos a otros procesos y se maneja el ingreso y transferencia de estos recursos, cada una de estas actividades corresponden a procesos que conforman la macro 4.

Estos patrones sirven como modelos referenciales y pueden ser instanciados para cualquier empresa.

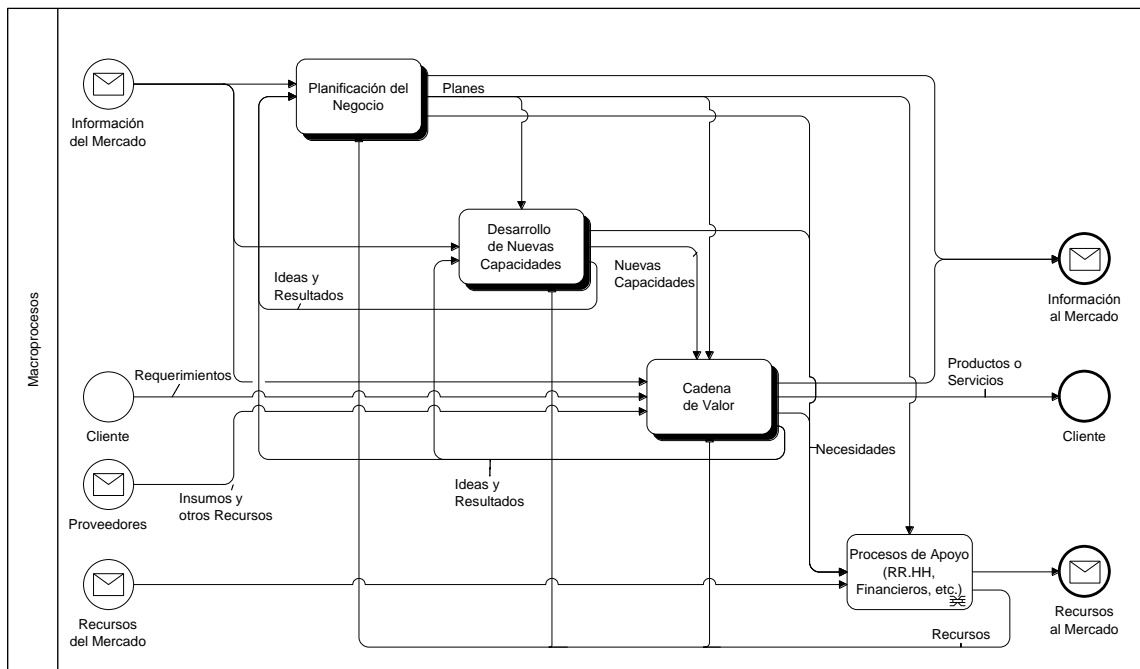


Figura 5. Macroprocesos básicos

### 2.1.2 Business Model Canvas

Este modelo creado por *Alexander Osterwalder*, busca plasmar la lógica del negocio y cómo esta crea, entrega y captura valor en nueve aspectos. Toma como concepto clave la propuesta de valor de una empresa, la cual debe ser entregada a los clientes mediante la generación de relaciones, para lo cual se usan canales. Además para lograr entregar la propuesta debe haber actividades claves, usar recursos y alianzas claves. Por último, se observan las estructuras de costos y formas de ingresos del negocio. A continuación se entrega una descripción en detalle de cada módulo del modelo de negocios.

- **Propuesta de valor:** describe qué problema o necesidad resuelve la empresa. El producto o servicio entregado debe aportar valor al segmento *target* y diferenciarse de la competencia.
- **Clientes:** define el o los grupos de personas o empresas a los cuales la empresa espera abordar. Se espera un reconocimiento geográfico, demográfico y social de los clientes.
- **Canales:** describe mediante qué medio la empresa planea conseguir que el producto llegue a los clientes.
- **Relación con clientes:** explicita tácticas para conseguir nuevos clientes, mantener a los que ya se tienen y hacerlos crecer en su ciclo de vida.
- **Modelo de ingresos:** indica cómo monetizar el producto o servicio entregado al cliente y qué medios de pago se usarán.

- Recursos clave: muestra las cosas más importantes que una empresa debe tener para que el modelo de negocio funcione, esto incluye elementos financieros, físicos, intelectuales e incluso humanos.
- Actividades clave: Describe las principales cosas que una empresa debe hacer para que el negocio funcione.
- Socios clave: indica cuáles son los principales socios y proveedores necesarios para que todo el sistema de negocios funcione.
- Estructura de costos: especifica cuáles son los costos relevantes para hacer funcionar al negocio. Se deben identificar los costos más altos, costos fijos, costos variables, economías de escala e impuestos.

Este análisis se resume en un lienzo como el siguiente:

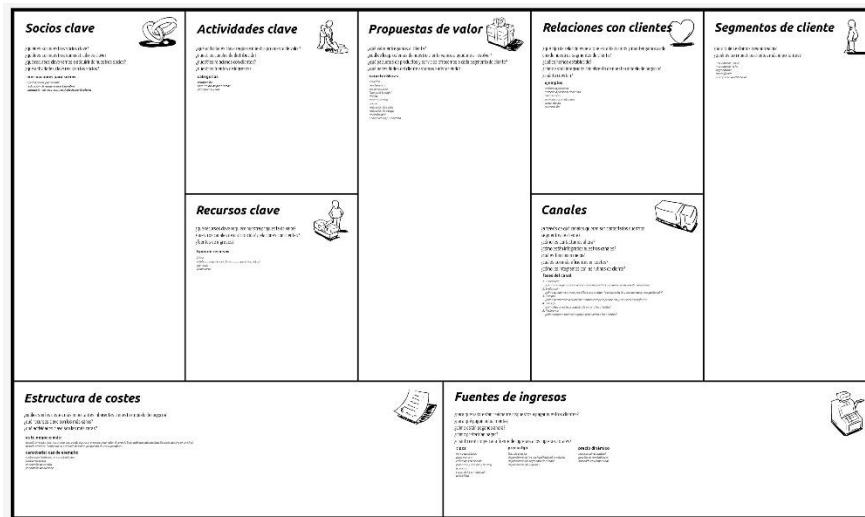


Figura 6. Lean Canvas

## 2.2 Lógica de Negocios

En el contexto del desarrollo de una empresa de marketing móvil que utiliza datos georreferenciados para decidir a quién y cuándo enviar una promoción u oferta se describen distintos tipos de métodos de *clustering*, clasificación y predicción.

Para la primera etapa del desarrollo se tiene la información de tráfico de antenas que cubren un Centro de Gravedad (CG). Un Centro de Gravedad es una zona de alta densidad comercial y estarán predeterminados. La matriz resultante registra las frecuencias de llamadas de un número determinado para ciertos espacios de tiempo, en la zona abarcada por el conjunto de antenas que cubren un Centro de Gravedad. En este sentido el primer grupo de métodos de *clustering* son aquellos métodos no supervisados que trabajan con vectores de información para encontrar patrones de comportamiento de usuarios alrededor de uno o más Centros de Gravedad.

### 2.2.1 K-means

Este es el Algoritmo aglomerativo por defecto, el cual busca minimizar una medida de distancia según los siguientes pasos:

1. Calcular k centroides aleatoriamente
2. Asignar cada observación al centroide con el cual minimice su distancia o disimilitud a la media
3. Recalcular los centroides como el punto central de cada cluster formado en el punto 2.
4. Volver al punto 2

El algoritmo se detiene cuando la diferencia entre centroides de 2 iteraciones contiguas es menor a un épsilon pequeño.

Algunas medidas de distancia factibles de utilizar con este modelo son:

1. **Euclidiana:** en el cual se generan k clusters (predeterminados), ubicando un punto p en un cluster con centro c, si la distancia euclidiana entre p y c es mínima, donde la distancia euclidiana entre un punto p y c es  $\sqrt{\sum_{i=1}^n (p_i - c_i)^2}$ , y valores bajo indican que los puntos p y c son cercanos.
2. **Coseno:** donde la similitud (o disimilitud) se calcula como una función del coseno ángulo generado por los vectores que generan un punto p y un centro c. La función utilizada es:

$$similitud(p, c) = \cos(\vec{p}, \vec{c}) = \frac{\vec{p} \cdot \vec{c}}{\|\vec{p}\| \cdot \|\vec{c}\|}$$

Luego, si  $similitud(p, c) = 1$ , entonces el ángulo formado entre los vectores que forman los puntos p y c es 0, y se considera que ambos puntos son 100% similares.

Por otro lado, si  $similitud(p, c) = 0$ , entonces el ángulo formado entre los vectores que forman los puntos p y c es  $\frac{\pi}{2}$ , y se considera que ambos puntos son 0% similares.

La similitud entre p y c es  $-1$ , cuando el ángulo formado entre los vectores que forman los puntos p y c es  $\frac{3\pi}{2}$ , y se considera que ambos puntos son 100% disimilares (completamente opuestos). Sin embargo, este caso no se da para la *clustering* calculada ya que tanto centroides como puntos se encuentran en el primer cuadrante formado por las variables estudiadas, luego todo ángulo entre un p y c cualquiera está entre 0 y  $\frac{\pi}{2}$ .

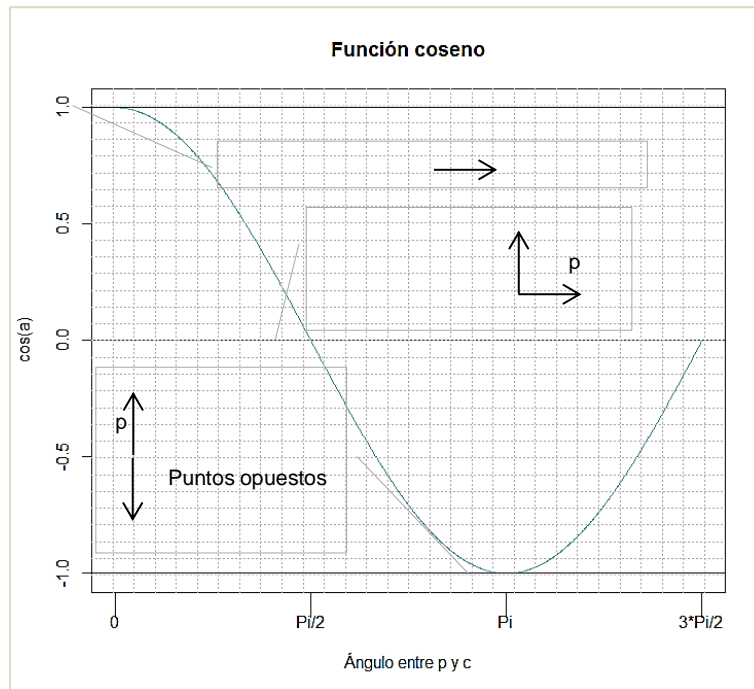


Figura 7. Función coseno

Este algoritmo entrega *clusters* de dimensión similar y con formas definidas

### 2.2.2 DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) [12]

Es reconocido como un algoritmo de *clustering* basado en densidad de alta calidad, permite identificar *clusters* de cualquier forma y el ruido de los datos.

Una visión general indica que un *cluster* es un área de alta densidad (con gran concentración de puntos), mientras que los puntos en zonas con baja densidad se consideran como ruido. Además, un punto en el espacio es considerado como el miembro de un *cluster* si hay un número suficiente de puntos dentro de una distancia, esta distancia se ajusta dependiendo de la aplicación.

*DBSCAN* descubre *clusters* en base a dos parámetros *Eps* y *MinPts* en dos pasos. Primero, elige un punto arbitrario que satisfaga la condición de punto central *p* y lo usa como semilla. Segundo, encuentra todos los puntos que son densidad alcanzables por *p*, obteniendo el *cluster* contenedor de *p*. Idealmente deben conocerse los parámetros apropiados *Eps* y *MinPts* de cada *cluster* y al menos un punto de cada *cluster*, luego pueden obtenerse todos los puntos que son densidad alcanzables por los puntos dados usando los parámetros correctos.

A continuación se presenta un detalle de las definiciones de este algoritmo:

- La vecindad-*Eps* de un punto  $p$  ( $N_{Eps}(p)$ ) se define como el set de puntos  $q$  en el *dataset*  $D$ , talque la distancia entre  $p$  y  $q$  es menor a *Eps*.
- Un punto  $p$  es definido como punto central si su  $N_{Eps}(p)$  contienen al menos *MinPts* (*MinPts*).
- Un punto  $p$  es considerado como densidad directamente alcanzable desde un punto  $q$  con respecto a *Eps* y *MinPts* si  $p$  pertenece a  $N_{Eps}(q)$  y  $q$  es un punto central
- Un punto  $p$  es densidad alcanzable desde un punto  $q$  con respecto a *Eps* y *MinPts* si existe una secuencia  $p_1, \dots, p_n$ , tal que  $p_1 = q, p_n = p \wedge p_{i+1}$  es directamente densidad alcanzable por  $p_i$ .
- Dos puntos de borde  $p$  y  $q$  están conectados en densidad si existe un punto  $o$  tal que,  $p$  y  $q$  son de densidad alcanzable por  $o$  con respecto a *Eps* y *MinPts*. Un punto central  $p$  está conectado en densidad consigo mismo.
- Un punto  $p$  es definido como punto borde si no es un punto central, pero es densidad alcanzable por algún punto central (pertenece a  $N_{Eps}(p)$ )
- $C(o)$  serán todos los puntos en  $D$  que son alcanzables por  $o$
- Un *cluster* se define como un set no vacío de todos los puntos en  $D$ , que son de densidad alcanzable por un punto central. Si  $p$  y  $q$  son puntos centrales, entonces pertenecerán al mismo *cluster* sí y sólo si  $C(p) = C(q)$ . Sin embargo un punto de borde puede pertenecer a más de un *cluster* al mismo tiempo.
- Todo punto del *dataset*  $D$  que no pertenece a ningún *cluster* es considerado ruido, es decir, todos los puntos en  $D$  que no son de densidad alcanzable por cualquier punto central.

Cada punto en un *cluster*  $C$  es de densidad alcanzable por cualquier punto central de  $C$ , es decir, el *cluster*  $C$  contiene exactamente los puntos que son densidad alcanzables por cualquier punto central de  $C$ .

Si dos *clusters*  $C1$  y  $C2$  son cercanos, entonces algún punto  $p$  podría pertenecer a ambos. Luego,  $p$  debe ser un punto de borde entre  $C1$  y  $C2$ , en este caso  $p$  será asignado al primer *cluster* descubierto.

Experimentos indican que el comportamiento del algoritmo no varía significativamente para valores de *MinPts* mayores a 4, luego puede eliminarse la búsqueda de este parámetro sustituyéndolo por 4 para todas las bases de datos de dos dimensiones.

### 2.2.3 OPTICS [13]

La idea principal de este algoritmo se basa en producir un orden aumentado de la base de datos que representa la estructura de los *clusters* basados en densidad. Busca usar parámetros variables de *Eps* y *MinPts*, por lo cual considera densidad variable para encontrar *clusters*.

*OPTICS* extiende *DBSCAN* realizando *clusters* basados en densidad con respecto a densidades diferentes simultáneamente. Al igual que *DBSCAN* recibe dos parámetros *Eps*, tal que se encuentre un objeto de densidad alcanzable con respecto al menor *Eps* que garantice que los *clusters* con respecto a la mayor densidad se encuentran primero. Y un parámetro *MinPts*, que indica el número de puntos requeridos para que un punto sea punto central.

A diferencia de *DBSCAN*, *OPTICS* considera una distancia central, que es la menor distancia *Eps* entre un punto *p* y un objeto en su vecindad-*Eps*, tal que *p* sea un punto central con respecto a *Eps*, en otro caso será indefinido.

$$\begin{aligned} & \text{Distancia central}_{Eps, minPts}(p) \\ &= \begin{cases} \text{Indefinido; si } |N_{Eps}(p)| < MinPts \\ \text{Distancia más cercana a MinPts de } N_{Eps}(p); \text{ otro caso} \end{cases} \end{aligned}$$

Además define la distancia de alcance de un punto *p* con respecto a un punto *o*, como la menor distancia tal que *p* es directamente densidad alcanzable por *o*, si *o* es un punto central.

$$\begin{aligned} & \text{Distancia de alcance}_{Eps, minPts}(p, o) \\ &= \begin{cases} \text{Indefinido; si } |N_{Eps}(o)| < MinPts \\ \max(\text{Distancia central}(o), \text{distancia}(o, p)); \text{ otro caso} \end{cases} \end{aligned}$$

Estas distancias serán indefinidas si no hay suficiente densidad para catalogar a un conjunto de puntos como *cluster*.

Con respecto a los modelos de clasificación y regresión, que buscan definir el tiempo de visita de cada usuario para un Centro de Gravedad particular se tiene:

### 2.2.4 Regresión logística [14]

Este método es, en muchos campos de investigación, el estándar cuando se trabaja con una variable objetivo categórica, que toma dos o más posibles valores, su objetivo es encontrar el modelo que ajuste mejor entre una variable dependiente y un conjunto de covariables. La diferencia de este modelo con una regresión lineal es que el resultado de la regresión logística es binario.

Las técnicas usadas en la regresión logística siguen los mismos criterios generales usados en los análisis de regresión lineal.

La transformación *logit*,  $g(x)$ , efectuada sobre la esperanza de  $Y$  dado  $x$ , permite mantener las propiedades de una regresión lineal: es lineal, puede ser continua y se mueve entre  $-\infty$  y  $+\infty$ , dependiendo del valor de  $x$ .

$$g(x) = \ln \left[ \frac{\pi(x)}{1-\pi(x)} \right] = \beta_0 + \beta_1 x, \text{ donde } \pi(x) = E(Y|x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

Una regresión logística se define como  $y = \pi(x) + \varepsilon$ , donde  $\varepsilon$  tiene una distribución con media cero y varianza igual a  $\pi(x)[1 - \pi(x)]$ . Esto implica que la variable dependiente sigue una distribución binomial, es decir deberá tomar valores entre 0 y 1, con probabilidad dada por la media condicional  $\pi(x)$ . Esta probabilidad disminuye progresivamente a medida que aumenta  $x$ .

Para ajustar un modelo logit es necesario estimar los valores  $\beta_0$  y  $\beta_1$ , lo cual se hace usando el método *máximum likelihood* que produce valores para los parámetros desconocidos maximizando la probabilidad de obtener el conjunto de datos observados. Esto se logra aplicando la ecuación  $l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}$ , donde  $l(\beta)$  es la función *likelihood* para el par  $(x_i, y_i)$  con observaciones independientes. Este análisis se facilita al aplicar la función logaritmo, obteniendo la expresión *log likelihood*.

El resultado es llamado *máximum likelihood* estimado de  $\pi(x_i)$ ,  $\hat{\beta}$ . Esta cantidad entrega una estimación de la probabilidad condicional de que  $Y$  sea igual a 1, dado que  $x = x_i$ .

### 2.2.5 Support Vector Machine [19]

Este modelo trata conjuntos de datos de clases distintas separándolos en hiperplanos paralelos mediante un hiperplano separador que maximiza la distancia entre los hiperplanos de cada clase y minimiza el error en el conjunto de entrenamiento. Los objetos que definen los hiperplanos paralelos son llamados *Support Vectors*, el espacio entre hiperplanos paralelos se conoce como margen.

El caso lineal separable es el caso en el cual todas las observaciones  $x_i$  de una clase permanecen en el mismo espacio.

El caso linealmente no separable, es el en cual algunos elementos de una de las clases se ubica en el espacio de la clase contraria, introduciendo un término de error a la función objetivo:

La función objetivo general (primal SVM) de un caso linealmente no separable se define como:



$$\min \left\{ \frac{\|w\|^2}{2} + C \sum \xi_i \right\}$$

Sujeto a:

$$\begin{aligned} y_i(x_i * w + b) - 1 + \xi_i &\geq 0 & \forall i \\ \xi_i &\geq 0 & \forall i \end{aligned}$$

Donde:

- w es el vector normal al hiperplano separador
- b es la distancia del origen al hiperplano separador
- $\xi_i$  es el error de clasificación de la observación i
- C es la penalización de los errores de clasificación

La formulación matemática (dual SVM) se define a partir del lagrangeano y la verificación de las condiciones de Karush Kuhn Tucker (KKT), resolver este problema es equivalente a resolver el problema primal:

$$\max \left\{ \sum \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_s y_i y_s x_i x_s \right\}$$

Sujeto a:

$$\begin{aligned} \alpha_i &\geq 0 & \forall i \\ \sum \alpha_i y_i &= 0 \end{aligned}$$

La ventaja del problema dual es la capacidad de transformar el espacio de origen a uno con mayor dimensionalidad, utilizando funciones de Kernel

$$K(x_i, x_s) = \Phi(x_i) \cdot \Phi(x_s)$$

Donde  $\Phi$  es una función que proyecta el espacio a una dimensionalidad mayor. Las funciones Kernel más conocidas son la geométrica y polinomial. La primera es la más usada en SVM y la segunda la más usada para procesamiento de lenguajes.

### 2.2.6 Árboles de decisión

Este algoritmo supervisado jerárquico utilizado para discernir sobre la clase de observaciones de un conjunto de datos, realiza un test sobre estos dividiendo el set de datos en nodos. Cada evento es asignado con una probabilidad y a cada una de las ramas se le determina un resultado.

En general se utiliza una medida de similitud entre observaciones como criterio de selección de variables. El algoritmo comienza eligiendo la variable con mejor criterio y ejecuta un test en base a una regla sobre esta variable. Todas las observaciones que cumplen con la regla se separan en una rama del árbol, mientras el resto de los datos

separa en otra. Iterativamente se prueba con distintas variables y distintas reglas hasta que todos o la mayoría de las observaciones ramificadas sean de la misma clase.

Algunos de los criterios de selección de variables son:

- *Gini Index*: Que muestra que tan “impuros” son los datos, es decir, qué tan probable es no obtener dos registros de la misma clase en un nodo. Se calcula de la siguiente forma:

$$G = 1 - \sum P(n, x)^2$$

Donde  $P(n, x)$  es la probabilidad de ocurrencia de la clase  $x$  en el nodo  $n$ . Mientras menor es este índice más puro es el Split, pues se tienen más observaciones de la misma clase en cada nodo.

- *Information gain*: Este criterio calcula la entropía de los atributos y comienza a ramificar por aquel que tiene un menor valor. La entropía se calcula como:

$$Entropía(n) = - \sum P(n, x) \log_2(P(n, x)^2)$$

Este método tiene un sesgo a seleccionar atributos con gran número de valores

- *Gain Ratio*: Este criterio es similar al anterior pero permite la amplitud y uniformidad de los valores de cada atributo.

Estos modelos se caracterizan por ser fáciles de construir e interpretar, funcionan tanto con datos categóricos como continuos, sin embargo estos últimos deben usarse con rangos de valores. Por otro lado, los modelos de árboles de decisión muestran problemas cuando los tamaños de las clases no están balanceadas y requieren una cantidad de datos representativa para poder ser entrenados

Se tienen diferentes modelos según el criterio de selección

<b>Modelo</b>	<b>Criterio</b>
<b>ID3</b>	Information Gain
<b>C4.5 y CART</b>	Gain ratio
<b>CHAID</b>	Chi cuadrado
<b>Ctree</b>	Test de permutación

Además existen otros parámetros usados para disminuir la probabilidad de sobreajuste del modelo:

- *Minimal size*: indica el número de observaciones mínima que debe tener un nodo para que sea ramificado.

- Minimal leaf size: indica la el número mínimo de observaciones que debe tener cada hoja del árbol
- Maximal deph: indica la cantidad de ramificaciones que habrán desde una hoja del árbol hasta su raíz. El valor de este indicador para la raíz del árbol es 0.
- Maximal height: indica la cantidad de ramificaciones que habrán desde un nodo a una hoja. El valor de este indicador para la raíz del árbol es 1, además el número máximo de nodos para maximal height (h) está dado por  $2^h - 1$ .

El análisis de un árbol de clasificación y regresión (CART) consiste en cuatro pasos básicos. El primero es construir el árbol utilizando criterios de Split como los mencionados anteriormente, cada resultado es asignado o asignado por probabilidad a una clase predeterminada, basada en la distribución de clases del data de aprendizaje y la matriz de costos. La segunda etapa consiste en detener el proceso de construcción del árbol, en este punto se construye un árbol máximo, el cual probablemente sobreajusta la información contenida en el conjunto de entrenamiento. El tercer paso es el de poda o *prunning*, el que crea una secuencia de árboles más simples tras la eliminación de nodos. El último paso es la selección de un árbol óptimo, en donde el árbol deja de estar sobreajustado a los datos de entrenamiento. [15]

Por otro lado los árboles condicionales (Ctree) usan múltiples tests de significancia en vez de seleccionar variables que optimicen un criterio de selección como las descritas antes. Estos test de significancia son conocidos como test de permutación, donde la distribución del estadístico bajo la hipótesis nula (la covariable es significativa con respecto a la variable objetivo) se obtiene calculando todos los posibles valores del estadístico bajo reordenamientos de las etiquetas de las observaciones. Con lo anterior busca disminuir el error de sobreajuste y el sesgo de selección de variables sin incurrir en la poda del árbol. El criterio de detención se da cuando ninguna covariable es significativa. [16]

### 2.2.7 Metodologías de validación de modelos

*Holdout*: Este es el método más sencillo de validación. Se crean dos subconjuntos, uno de entrenamiento y otro de testeo (con este se estima la tasa de error del modelo). Usualmente el primer conjunto de datos representa el 70% u 80% del total de datos, mientras que para el conjunto de testeo se utiliza el 30% o 20% restante.

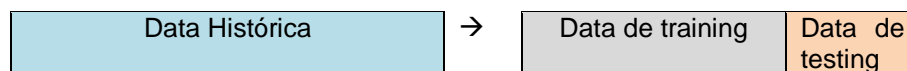


Figura 8. Validación por Holdout

*Cross-Validation*: Corresponde a una iteración de holdouts, donde se va cambiando el set de prueba y testeo, el número de iteraciones es definido por el data scientist. Luego se unifica el resultado con la media aritmética de los resultados anteriores

Ejemplo para una cross-validation de 3 iteraciones.

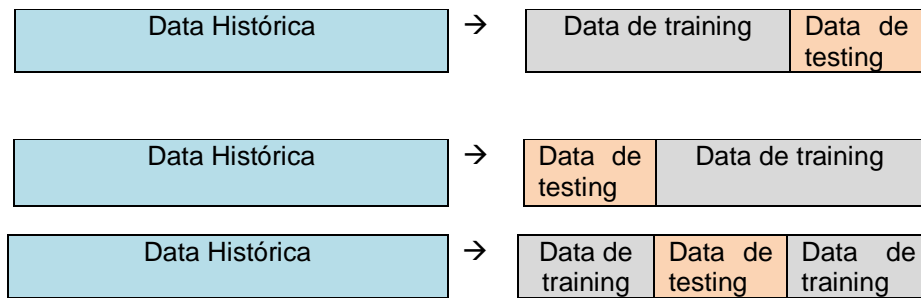


Figura 9. Validación por Cross-Validation

*Leave-one-out*. Se generan tantas iteraciones como datos tiene el set, en cada iteración se hace el conjunto de entrenamiento, dejando uno de los datos como conjunto de testing. Luego se unifica el resultado con la media aritmética de los resultados anteriores.

### 2.2.8 Medidas de desempeño de los modelos de clasificación

Sea una muestra a clasificar entre la clase +1 y -1, entonces se define un positivo verdadero como un evento positivo clasificado como positivo por el modelo y un falso positivo como un evento negativo clasificado como positivo por el modelo. De forma análoga un negativo verdadero será un evento negativo clasificado como negativo por el modelo y un falso negativo será un evento positivo clasificado como falso por el modelo.

	Respuesta del método	
	+	-
Clase +1	Positivo verdadero	Falso negativo Error tipo 1
Clase -1	Falso positivo Error tipo 2	Negativo verdadero

Figura 10. Matriz de confusión

Dado lo anterior se definen las siguientes medidas de desempeño:

#### Accuracy:

Se define como la proximidad de la media de los resultados encontrados por el modelo al verdadero valor de la muestra. En modelos de clasificación binaria es la proporción de resultados verdaderos en el número total de casos:

*Accuracy*

$$= \frac{(N^{\circ} \text{ positivos verdaderos} + N^{\circ} \text{ negativos verdaderos})}{(N^{\circ} \text{ positivos verdaderos} + N^{\circ} \text{ falsos positivos} + N^{\circ} \text{ negativos verdaderos} + N^{\circ} \text{ falsos negativos})}$$

Esta medida funciona mejor cuando los tamaños de las clases +1 y -1 son similares.

***Precision:***

Es la proporción de resultados positivos verdaderos y resultados positivos de un modelo de clasificación de patrones, determina la relación en la que métodos repetidos en condiciones diferentes generan los mismos resultados:

$$\begin{aligned} \textit{Precision} &= \frac{N^{\circ} \text{ positivos verdaderos}}{N^{\circ} \text{ positivos verdaderos} + N^{\circ} \text{ falsos positivos}} \\ &= \frac{N^{\circ} \text{ positivos verdaderos}}{\textit{total de resultados positivos}} \end{aligned}$$

Un resultado cercano a 1 (o 100%) indica que el modelo tiene una alta precisión de clasificación (comete baja cantidad de error tipo 2), por otro lado si la precisión es cercana a cero entonces se tienen muchos más resultados positivos que los clasificados por el modelo, por cual no es modelo recomendable de usar.

***Recall:***

Medida de la performance de un clasificador binario que mide la proporción de resultados positivos correctamente identificados y la tasa de falsos negativos. Un predictor perfecto tendrá un *recall* del 100% (en teoría, pues en la realidad todo predictor tendrá asociado un error bayesiano)

$$\begin{aligned} \textit{Recall} &= \frac{N^{\circ} \text{ positivos verdaderos}}{N^{\circ} \text{ positivos verdaderos} + N^{\circ} \text{ falsos negativos}} \\ &= \frac{TP}{\textit{total de observaciones de clase} + 1} \end{aligned}$$

Un resultado alto del indicador '*recall*' se interpreta como un modelo que tiene alta exactitud de predicción, comete pocos errores de tipo 1.

***Area Under the Curve (AUC):***

Antes de describir el criterio '*AUC*' (*Area Under the Curve*) es necesario, describir una herramienta de validación de modelos de clasificación en señales, el *Receiver Operating Characteristic (ROC)*, la cual describe la tasa de clasificaciones correctas contra la tasa de clasificaciones falsas. Se observa en la figura 1 el '*ROC*' en la curva azul en comparación con las tasas obtenidas al aplicar un clasificador aleatorio (línea roja).

El área total bajo la curva ROC (línea azul) es conocido como el estadístico 'AUC', un buen clasificador deberá tener un área cercana a 1, mientras que un clasificador aleatorio tiene un AUC de 0.5

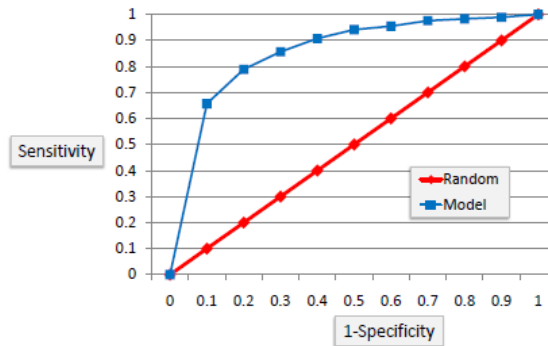


Figura 11. Curva ROC y criterio AUC

**F-Measure:**

Corresponde a la media armónica entre las medidas 'precision' y 'recall', podría considerarse como una alternativa al criterio AUC. Esta medida analiza la varianza de los datos mediante la comparación de dichas varianzas:

$$F_{\beta} = (1 + \beta^2) * \frac{precision * recall}{precision + \beta * recall}$$

Un valor de  $\beta$  mayor a 1 entrega un valor de F que da más peso a los errores de tipo 1, recogida por la medida *recall*. Por otro lado un valor menor a 1 dará más peso a los errores de tipo 2, capturados por la medida *precision*.

## CAPÍTULO 3: PLANTEAMIENTO ESTRATÉGICO Y MODELO DE NEGOCIOS

---

### 3.1 Visión

Un mundo donde la tecnología sea parte de la vida de las personas, donde seamos el medio de promoción geolocalizada en tiempo real más usado en Chile.

### 3.2 Misión

Ser pioneros en generar promociones personalizadas con geolocalización y en tiempo real, siendo parte de la revolución tecnológica que vivimos cada día, enviar los mensajes oportunos a las personas correctas en el momento preciso.

### 3.3 Estructura básica de la empresa

Se propone una estructura básica de 4 departamentos principales: Recursos Humanos (R.R.H.H.), Tecnologías de Información (TI), Administración y Estrategia.

El área de TI contendrá a una sub-área de Business Intelligence donde se crearán y recalibrarán modelos para ser luego llevados a producción. Por otro lado, se requiere de un área de soporte de la página y servidores asociados. Las áreas de operaciones, marketing y ventas y finanzas han sido agrupadas en un gran área de Administración, con el fin de alinear las acciones con el propósito de la cadena de valor. Dado que el mercado es creciente se requiere de un área de estrategia en el cual se analizarán los cambios del mercado de manera de anteponerse a ello y mantenerse activos en él.

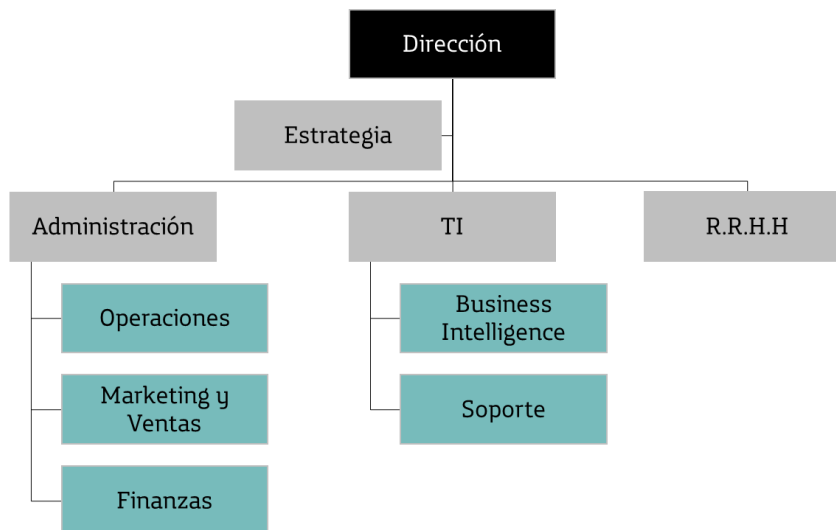


Figura 12. Organigrama funcional propuesto

### 3.4 Análisis FODA



Figura 13. Matriz FODA

#### 3.4.1 Fortalezas

- **Poseción de información de tráfico de datos, voz y SMS:** Los datos almacenados en los CDR corresponden al principal recurso de la empresa y marca la diferencia con la competencia actual, pues permite tener acceso a información geolocalizada sin la necesidad de que el usuario use una aplicación que para hacer un seguimiento de su comportamiento y adelantarse a sus necesidades. Además permite llegar tanto a usuarios de Smartphone como usuarios de teléfonos celulares no inteligentes.

Según el objetivo de negocio, la información de los CDR hará que se envíe el mensaje correcto, a la persona correcta, en el tiempo correcto.

- **Cultura de I+D+i tanto de los desarrolladores de los modelos como de los socios claves del proyecto:** la colaboración de CEINE y el área I+D de la TELCO se fortalece con su alineamiento en el desarrollo de modelos que sirvan a la sociedad y permitan realizar estudio que aumenten el conocimiento sobre la movilidad de las personas en Chile.
- **Mayor cobertura de red:** el Plan de Expansión Regional en Infraestructura implementado en 2013 posiciona al aliado estratégico, como la compañía de telecomunicaciones con mayor cobertura en redes del país, esto beneficia la alimentación de las bases de datos de tráfico de datos y permite ajustar la precisión de la posición de los usuarios.



- **Aliado líder en penetración de mercado de la telefonía móvil:** La TELCO concentra cerca del 38% de los usuarios del mercado móvil. Dado que los usuarios son los clientes activos del aliado, este es un factor importante para el desarrollo de la empresa.

### 3.4.2 Oportunidades

Las oportunidades de negocios fueron descritas en la sección 1.1:

- Aparición del *geobusiness* e importancia de la segmentación
- Fortalecimiento de la publicidad móvil
- Poca diversidad en la oferta actual
- Imprescindibilidad de teléfonos móviles

### 3.4.3 Debilidades

- **Es necesario completar perfiles de los usuarios:** sólo con la información de los CDR no es posible hacer *clusters* de usuarios confiables, ya que no se tiene información precisa de las preferencias del cliente, por lo que es necesario invertir en recursos para alimentar el modelo con este tipo de información. Probablemente pueda conseguirse con aliados estratégicos o bien con el proceso de *Opt-in*.
- **La percepción de marca de la TELCO no está bien valorada:** según el ranking de calidad del Ministerio de Transporte y Telecomunicaciones, se observa que a pesar de que el servicio es similar al de su mayor competencia, la compañía de telecomunicaciones aliada ha bajado en la encuesta de satisfacción de sus clientes. Esto podría afectar al proyecto que se lanzará bajo el alero de la marca.
- **Precisión de los datos:** como se menciona en la sección 1.8.1 existe baja precisión de los datos para detectar la posición exacta de los usuarios.

### 3.4.4 Amenazas

- **Bajas barreras de entrada para empresas con información geográfica:** estas empresas pueden integrarse con facilidad al mercado del *geomarketing*. La simplicidad en la creación de aplicaciones facilita la entrada de nuevos actores, en particular, una empresa con información de suelos puede agregar información de interés para cada usuario en el despliegue del mapa para formar parte del mercado.
- **Amplia gama de sustitutos en distintas funcionalidades:** a pesar de que la oferta del marketing móvil es poco diversa, existe gran cantidad de canales para entregar productos sustitutos, estos cumplen objetivos diferentes en los cuales las personas que obtienen promociones y descuentos son aquellas que se

“esfuerzan” accediendo al sitio para obtener beneficios. Estas empresas buscan crear la necesidad suficiente como para que el usuario invierta tiempo y recursos ingresando a sus sistemas para buscar beneficios. Al igual que en el punto anterior, el fácil acceso a las aplicaciones móviles disminuye las barreras de entrada al mercado para productos sustitutos.

- **Empresas de telecomunicaciones como futuros competidores:** Así como este spin-off nace desde una TELCO, otras empresas de comunicaciones cuentan con los recursos para entrar en el mercado del marketing móvil, utilizando una propuesta de valor similar basada en el uso de los datos móviles de sus clientes. Esto los convierte en potenciales competidores, lo que implicaría un aumento de la rivalidad por la exclusividad de los clientes, no así de los usuarios ya que no es posible acceder a los CDR de otras compañías.
- **Leyes asociadas a los derechos del consumidor y protección de datos:** Actualmente se trabaja bajo las leyes 19.628 y 19.496 que protegen la información personal de los usuarios. El riesgo aparece según las modificaciones que estas pueden sufrir o creación de nuevas leyes de esta línea en el futuro que puedan coartar el normal funcionamiento de la empresa o aumentar las barreras de entrada al mercado del marketing móvil.
- **Sensación de privacidad de los consumidores:** La información de localización es considerada como información personal por los usuarios y no desean que esta esté a disposición de otras personas u organizaciones.

### 3.5 Posicionamiento Estratégico

La empresa utiliza una estrategia con tendencia a la *solución integral al cliente*, más específicamente se trata de una estrategia de *integración con el cliente*, en la cual se transfieren capacidades, conocimientos y servicios, para potenciar el rendimiento del cliente. El producto está diseñado para fortalecer la cadena de valor de los compradores por lo que la estrategia gira en torno a la economía de ellos y no la del producto.

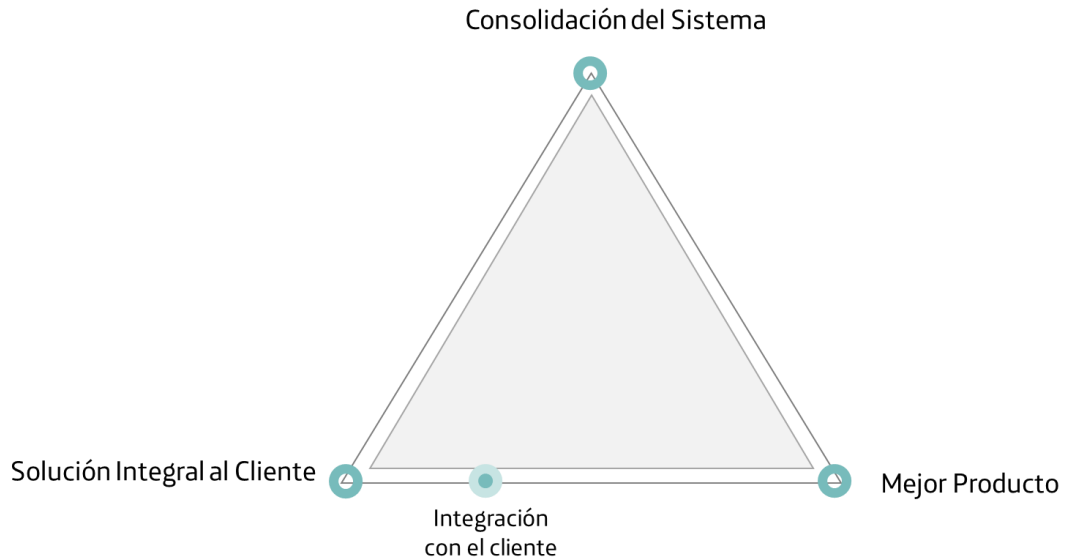


Figura 14. Posicionamiento Estratégico según modelo Detal de Hax

### 3.6 Balanced Scorecard

La estrategia de la empresa busca generar utilidad gracias al crecimiento de los ingresos, es decir, enfocándose en aumentar las ventas y no necesariamente aunando esfuerzos en disminuir los costos. Esto debido a que la estrategia no se enfoca en la rotación de productos y los costos de producción varían poco en la aplicación del modelo, sin embargo, esto no quiere decir que no se busque optimizar el uso de recursos.

En este sentido, la empresa busca:

- Primero. Mejorar el valor para los clientes, lo cual se alinea con su posición estratégica de *integración con el cliente*, ya que toda mejora de los procesos se enfocan en entregar un mejor producto.
- Segundo. Ampliar las oportunidades de negocio dentro de su única cadena de valor, generando nuevos tipos de análisis y reportes o metodologías de venta.

La empresa tiene propuestas de valor diferenciadas para clientes y usuarios como se muestra en la sección 3.7.2. En relación a esto debe cumplir tanto con las necesidades de los clientes (aumentar su conversión y el conocimiento de sus clientes), como las de los usuarios (obtener los beneficios correctos en el lugar correcto, sin vulnerar su privacidad).

Los procesos internos producen y entregan la propuesta de valor a los clientes y usuarios, de forma tal que se satisfagan las condiciones de mejorar el valor para los clientes y ampliar las oportunidades de mercado. Dentro de la cadena de valor se destacan los procesos:

- Producción: procesos que producen y entregan el servicio. Se relacionan tanto con la perspectiva del cliente como con la perspectiva financiera, por ejemplo, se observa que una mejora en los procesos de análisis de modelos, genera nuevos tipos de informes lo cual amplía la gama de productos y por ende de oportunidades de negocios. Al mismo tiempo, al ampliar la gama de productos se está entregando mayor valor al cliente.
  - generación de *clusters*
  - cruce con la información de los clientes
  - optimización del set receptor final
  - análisis de modelos
- Relación con los clientes: procesos que amplían y profundizan las relaciones con los clientes
  - venta, postventa y retroalimentación
- Relación con los usuarios: procesos que amplían y profundizan las relaciones con los usuarios. Se da particular importancia al servicio de consultas y reclamos puesto que son los que permiten la fidelización de clientes.
  - obtención de usuarios (*Opt-in* y *Opt-out*)
  - consultas y reclamos

Dentro de la perspectiva de innovación y desarrollo se contempla la capacitación del personal de *Business Intelligence* los cuales son fundamentales para el desarrollo de los productos y el constante trabajo en procesos de innovación, con enfoque en el desarrollo de nuevas metodologías y mejoramiento del producto. Además se busca contar con personal comercial competente para mantener la relación tanto con clientes como con usuarios.

Se reconocen dos ejes estratégicos. El primero es el eje comercial, en el cual se centralizan los objetivos de relación con el cliente y usuario. El segundo es el eje productivo, donde se busca mejorar la efectividad de los modelos y con ello cumplir con la consigna de entregar los mensajes correctos, a las personas correctas, en el momento correcto. En la figura se muestra el *Balanced Scorecard* de la organización destacando en color gris el eje comercial y en color celeste el eje productivo. En blanco se muestran objetivos que responden a ambos ejes estratégicos.

Además se han construido *KPI's* para medir el cumplimiento de los objetivos propuestos e iniciativas para lograrlo (ver .Anexo 1: *KPI's* según objetivos de negocios).

#### **Iniciativas:**

- Utilización de tecnologías de punta: Esta iniciativa busca identificar, monitorear y mejorar los procesos que asegurarán el desarrollo de modelos de *Data Mining* que son considerados de alta precisión por el cliente.

La compañía deberá hacer un esfuerzo sistemático para mejorar los procesos de I+D, entrenamiento del personal y equipamiento, y relación con el proveedor de tecnología para la ejecución de estos modelos.

Las investigaciones requerirán invertir constantemente en habilidades avanzadas de los empleados y mantener las licencias y arriendos de equipos avanzados y de softwares de última generación.

- Promociones de Satisfacción del Cliente: Programa de ventas y marketing que utiliza múltiples métodos *online* e impresos para comunicar y educar a los clientes con respecto a la propuesta de valor de la compañía. Los gerentes de cuentas y los vendedores apoyan los materiales de promoción mediante conversaciones uno a uno con clientes *target*. Se espera que la inclinación de los clientes a recomendar la compañía sea impactada significativamente por las campañas de marketing.
- Sistema de atención al cliente: Iniciativa que pretende generar canales de respuesta a consultas y reclamos de forma ágil y cercana los usuarios de la empresa. Para esto, y dada la línea de innovación de la empresa, se establecería un medio en línea para el registro de consultas y reclamos, información que luego podrá ser analizada en busca de tópicos similares, aplicando modelos de *text mining* [15], y con ello generar procedimientos de atención que aumenten la velocidad y calidad de respuestas a usuarios.

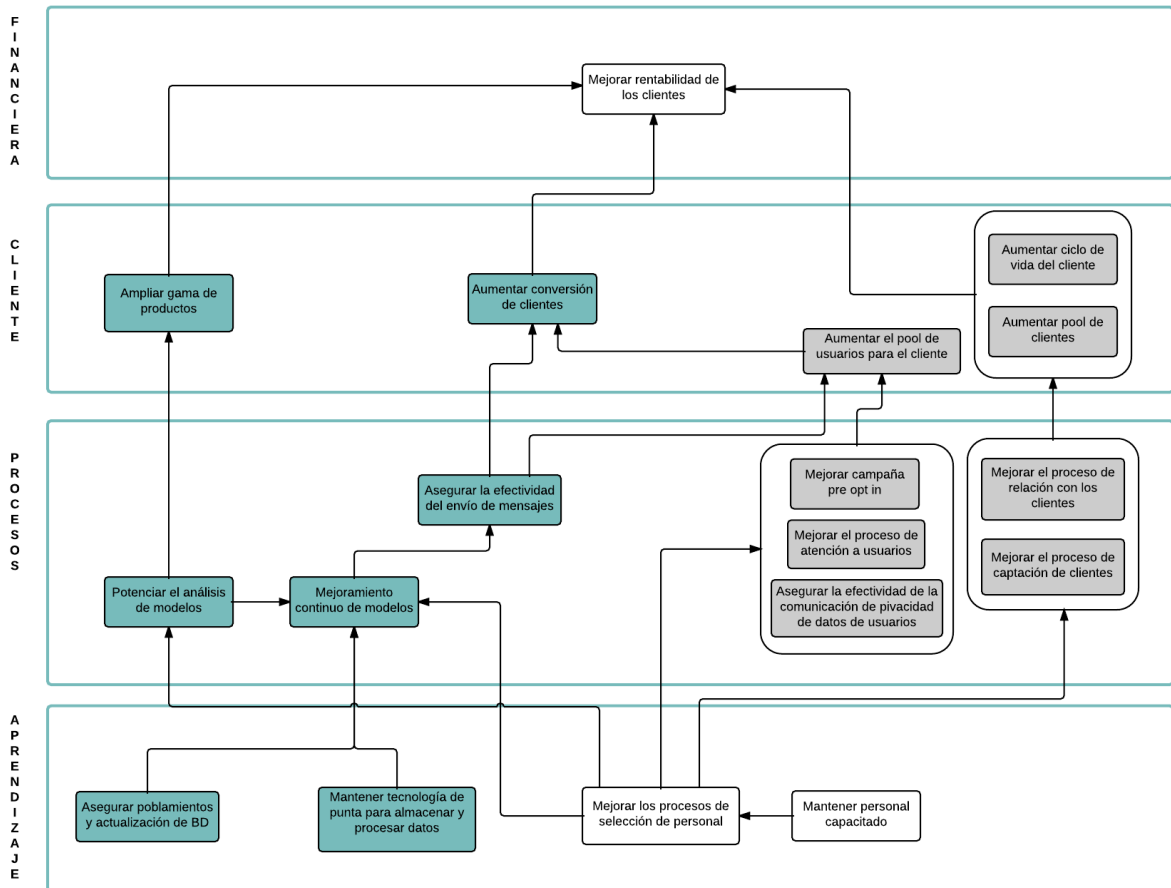


Figura 15. Balanced Scorecard

### 3.7 Modelo de Negocios

#### 3.7.1 Segmentos de clientes

- **Usuarios:** Los usuarios del servicio corresponden a actuales clientes de la compañía telefónica, con su marca de telefonía móvil. Será posible obtener el comportamiento rutinario de los usuarios a través del registro de sus llamadas, ya sean emisores o receptores de esta.

Una estimación de los posibles usuarios es realizada según información de la Subsecretaría de Telecomunicaciones, la cual indica que el total de abonados a telefonía móvil alcanza los 23,3 millones de personas a junio de 2015 y que, a la misma fecha, la TELCO aliada posee el 36,7% de la participación de mercado. Por otro lado, deben considerarse sólo los clientes con línea activa en la empresa, según datos históricos de ésta, los clientes con línea activa corresponden a aproximadamente el 60% de los clientes totales, por lo tanto se estima un total de 5.130.000 clientes aproximadamente. Esta será la cantidad máxima de usuarios, ya

que no es posible considerar como usuarios a clientes de otras compañías de teléfonos (pues no se pueden obtener sus datos de CDR). Además la empresa proyecta que a 2019 los clientes activos lleguen a los 7,3 millones de clientes.

- Cientes: Los clientes corresponden a empresas comercializadoras, que tengan como interés aumentar los puntos de comunicación con sus respectivos clientes.

La ubicación de las sucursales de estas empresas debe estar en territorio chileno, abarcado por antenas de telefonía del proveedor.

Estas empresas pueden estar en el rubro del comercio, entretención, banca y servicio financiero, turismo, salud y belleza, empresas estatales y universidades

Se han destacado los siguientes rubros del mercado con altos índices de inversión en publicidad digital:

- Educación
  - Universidades
  - Institutos
  - Preuniversitarios
- Comercio
  - Centros comerciales
  - *Retail*
  - Otras tiendas de cadena
- Mercado automotriz
- Entretención
  - Restaurantes
  - Bares y centros nocturnos
  - Cines
  - Centros de entretenimiento
- Banca y Servicios Financieros
- Turismo
  - Transporte aéreo
  - Hotelería
  - Agencias de viaje
- Salud y Belleza
  - Gimnasios
  - Clínicas
  - Centros de estética
  - Spa
- Empresas Estatales y Gobierno

### **3.7.2 Propuesta de valor**

La empresa presenta propuestas de valor tanto para sus clientes como para sus usuarios:

- A los clientes: *Acercar a nuestros clientes a sus consumidores como medio de comunicación geolocalizada de productos y/o servicios para usuarios segmentados.*

El servicio puede transmitir distintos tipos de comunicaciones: comerciales (promociones y ofertas), informativas y recomendaciones, mediante el envío de mensajes que contienen un hipervínculo al beneficio.

Cada campaña es específica para un segmento de usuarios y su éxito será medido en base a las transacciones registradas por el cliente derivadas del envío de algún mensaje a los usuarios.

- A los usuarios: *Entregar promociones que realmente quieren los usuarios, según su ubicación e intereses.*

Esta propuesta se logra obteniendo información relevante (¿Quién es?, ¿Qué lugares visita?, ¿Cuáles son sus intereses?) desde el uso de los teléfonos celulares.

De esta forma se puede predecir el momento en el cual el usuario podría considerar relevante el envío de publicidad o promociones, sin convertir el anuncio en un distractor dentro de la rutina del usuario.

El atractivo de la propuesta de valor es que, gracias al trabajo semántico, analítico y de retroalimentación se pueden inferir perfiles y comportamientos, el servicio será invisible, relevante y oportuno, ya que ocurrirá en el momento preciso en que el contenido sea relevante.

### **3.7.3 Relación con clientes y usuarios**

Las relaciones de la empresa se diferencian según clientes y usuarios. Para los primeros se tendrá contacto virtual en la captación de clientes. Luego de la captación se establecerá contacto físico para determinación del contenido y el target de usuarios receptores, se entregará un servicio de seguimiento online de la entrega de mensajes y el comportamiento de los usuarios luego de la recepción, acompañado de informe periódicos con control de las variables relevantes y KPI que se determinen con la empresa.

La relación con los usuarios se basa en la personalización de los contenidos enviados, de acuerdo a sus patrones de movilidad y preferencias. Para lograr esto de manera invisible y sin invadir la privacidad del usuario se utilizarán los procesos de *Opt-in* y *Opt-out* explicados en el punto 4.3.5, de esta manera se aumentará la confianza de los usuarios en el servicio entregado.



### **3.7.4 Canales**

Los canales para establecer las relaciones serán tanto virtuales como físicos. Para los clientes se tendrán canales virtuales de captación (página *web* o e-mails promocionales), físicos y virtuales para la captación del contenido relevante, para el cual ejecutivos podrán hacer visitas, llamadas telefónicas o contacto vía correo electrónico para atender consultas y nuevos requerimientos. El seguimiento de las campañas y la entrega de informes de estado podrán entregarse por medio de una página *web* con sesiones o envío mediante correo electrónico.

Por el lado de los clientes se usarán canales virtuales en todo el ciclo de vida de este. Para la captación (proceso de *Opt-in*) se utilizará una página *web* responsiva o bien una página completada con aplicaciones móviles. Este medio será relevante para el contacto permanente con el usuario, donde podrán informarse del servicio, sus funcionalidades y políticas de privacidad. Para la entrega del servicio se ha contemplado el uso de la red de mensajería de SMS, posiblemente complementado con una aplicación móvil.

### **3.7.5 Socios Clave**

Se identifican 3 socios clave:

- Área I+D de empresa de telefonía: Se comporta como socio y proveedor de recursos y servicios (Hardware, envío de mensajes, apoyo de personal, etc.)
- Proveedores de tecnología: Parte crucial del desarrollo de la empresa consiste en contar con soporte tecnológico que sustente el procesamiento y almacenamiento de gran cantidad de datos.

### **3.7.6 Actividades Clave**

- Población: Esta actividad se subdivide en fortalecer los procesos analíticos CRM (*Customer Relationship Management*) para potenciar el proyecto con los resultados obtenidos, definir las fuentes disponibles de información (internas o externas) y definir los mecanismos de poblamiento de datos para completar las bases de datos de los clientes.
- Modelado: Considera el pre-procesamiento segmentación y pronóstico de visitas de los usuarios (incluye la definición de los algoritmos) que puedan ser foco de campañas de marketing móvil geolocalizado.
- *Opt-in* y *Opt-out*: Esta actividad incluye tanto la definición de los mecanismos para captar efectivamente la autorización de los usuarios para el uso de sus datos de movilidad, como la ejecución de los procesos y el control de estado de estos.
- Envío de mensajes: Corresponde a la generación del *subset* de usuarios que recibirán los mensajes de campañas publicitarias.

### **3.7.7 Recursos Clave**

- Información de tráfico: Accesibilidad a la información del tráfico de redes de los clientes de la TELCO. Al respecto, es indispensable completar la mayor cantidad de información posible de las bases de datos existentes (CDR's) y por crear (base de preferencias obtenida desde el proceso de *Opt-in*), de manera de alimentar los modelos con suficiente información para el análisis de perfiles y pronóstico de visitas.
- Hardware/Software: Corresponde a la tecnología necesaria para el procesamiento y almacenamiento de los datos.
- RRHH: Es necesario tener dos grupos principales, un grupo analítico para la generación de modelos, levantamiento de requerimientos y determinación de grupos de personas a los cuales se enviarán mensajes. Por otro lado, se requiere de personal de ventas que atraigan a empresas dentro del perfil de clientes buscados y levanten las necesidades que tiene el perfil al cual se enviarán mensajes.

### **3.7.8 Estructura de costos**

Los costos asociados a los recursos y procesos necesarios para la entrega del servicio son fijos y variables:

- Inversión inicial en *hardware* y *software* que cumplan los requisitos de *Big Data* que tiene el proyecto; marketing, ya que se busca generar valor de marca, dar a conocer el servicio y captar usuarios y clientes
- Costos fijos: Corresponden a los costos por RRHH (equipo analítico y de ventas), mantención y desarrollo de modelos y softwares de contacto con usuarios y clientes.
- Costos variables: Asociados al personal a cargo de clientes y el costo del envío de mensajes de beneficios. La cantidad de personal varía según el crecimiento de la empresa y el costo de envío de mensajes es diferente para cada campaña ejercida.

### **3.7.9 Fuentes de ingresos**

Los ingresos provendrán de los clientes, mediante la venta de campañas publicitarias cuyos precios se determinan según una tarifa fija y una tarifa variable descritos en el punto 7.2.2. Además, no se descarta la venta de estudios de efectividad de campañas personalizadas y geolocalizadas de marketing móvil o comportamiento de los consumidores generados a partir del flujo de mensajería y las acciones ejecutadas por los usuarios.

# CAPÍTULO 4: PROPUESTA DE DISEÑO DE PROCESOS

## 4.1 Arquitectura de Procesos To Be

Esta empresa cuenta con un patrón de arquitectura básico con una cadena de valor.

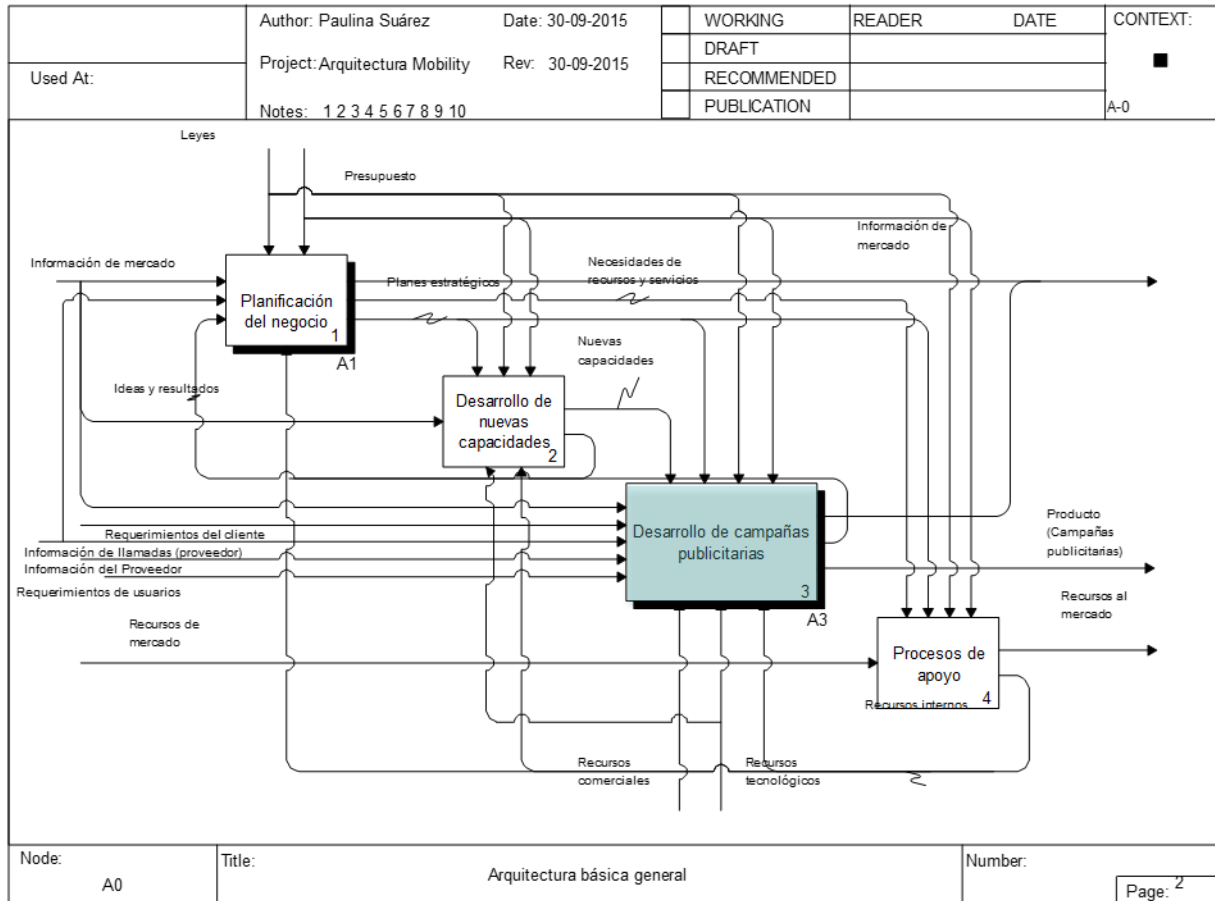


Figura 16. Arquitectura básica general

Tabla 2. Detalle de entradas y salidas. Nivel A1 Planificación del Negocio

Inputs:	<p>La información del mercado necesaria para realizar análisis de mercado (tanto externos como internos), identificar oportunidades, debilidades, fortalezas y amenazas; clarificar las ventajas competitivas que entrega la propuesta de valor.</p> <p>Además recibe como retroalimentación ideas y resultados del macroproceso de Desarrollo de Nuevas Capacidades, las cuales permiten actualizar la estrategia con nuevas metodologías para la operación.</p>
---------	---

	Para desarrollar la estrategia se requiere la información de los recursos entregados por los proveedores, en este caso la información de CDR y disponibilidad de máquinas, como parte de los recursos internos de la empresa.
Mecanismos:	Uso de recursos internos, ya sean personas, financieros y tecnológicos no asociados al trabajo comercial ni de modelado.
Controles:	Dos importantes restricciones son las de presupuesto y las leyes 19.496 (Derechos del Consumidor) y 19.628 (Protección de Datos).
Outputs:	A partir del macroproceso de planificación del negocio se generan los planes estratégicos (de ventas, recursos humanos, marketing y plan de operaciones) a utilizar en los siguientes 3 macroprocesos, se establecen necesidades de recursos y servicios que deberán ser considerados por Procesos de Apoyo en la generación de recursos internos y se entrega información al mercado sobre el valor que entrega la empresa y el posicionamiento que se busca.

Tabla 3. Detalle de entradas y salidas. Nivel A2 Desarrollo de Nuevas Capacidades

Inputs:	Para ser competentes en el mercado, esta macro requiere de la información del mercado como contexto de lo que se hace, lo que se podría hacer y lo que no se está haciendo.
Mecanismos:	Utiliza los recursos internos de personal, tecnología y presupuesto asignado. Además de los recursos tecnológicos utilizados para el desarrollo y utilización de modelos de segmentación y optimización.
Controles:	El presupuesto funciona también como una restricción para el trabajo de desarrollo de nuevas capacidades, así como las leyes 19.496 y 19.628; y los planes estratégicos entregados desde la Planificación del Negocio.
Outputs:	Entrega nuevas capacidades que servirán para la ejecución de la cadena de valor (como nuevos modelos de <i>clustering</i> , metodologías de procesamiento de datos, optimización de envío de mensajes, etc.) y generará nuevas ideas que deberán ser utilizadas en la revisión de la estrategia.

Tabla 4. Detalle de entradas y salidas. Nivel A3 Desarrollo de campañas publicitarias

Inputs:	El macroproceso de cadena de valor requiere la información de los recursos que recibirá de los proveedores, información del proveedor con el cual manejar la relación y requerimientos tanto de clientes como de usuarios para ajustar el producto a lo que ellos necesitan.
Mecanismos:	Este macroproceso requiere tanto de recursos internos de personal y presupuesto, como de recursos tecnológicos diferenciados según función. A los recursos destinados para la ejecución de modelos de segmentación y tratamiento de grandes cantidades se les llamó "Recursos Tecnológicos" ( <i>cluster</i> de servidores para la ejecución de modelos, computadores y cualquier artefacto tecnológico utilizado en el sistema de <i>clustering</i> ), mientras que los recursos tecnológicos utilizados para el procesos de ventas se denominan "Recursos comerciales" (recursos destinados a página <i>web</i> y contacto con clientes).
Controles:	La planificación y ejecución de la cadena de valor también debe realizarse en el contexto de las leyes mencionadas y el presupuesto disponible. Además debe seguir los planes estratégicos desarrollados en la Planificación del Negocio y regirse por la nuevas capacidades desarrolladas en el macroproceso anterior.
Outputs:	El <i>output</i> principal son campañas publicitarias móviles. Por otro lado, se genera retroalimentación del proceso para replanificar la estrategia y responder mejor al mercado.

Tabla 5. Detalle de entradas y salidas. Nivel A4 Procesos de apoyo

Inputs:	Los Procesos de apoyo utilizan los recursos provenientes del mercado (principalmente personal e ingresos de la empresa) para cumplir su objetivo de apoyo a los tres macroprocesos anteriores.
Controles:	Los mecanismos a seguir para estos procesos son tanto los planes estratégicos como las necesidades de recursos y servicios derivados desde la Planificación del Negocio, cualquier posible nueva capacidad desarrollada, el ajuste al presupuesto disponible y cumplimiento de las leyes 19.496 y 19.628.

Outputs:	Las salidas de este macroproceso son recursos internos y externos, lo primeros corresponden al personal necesario para la realización de ciertos macroprocesos, así como el presupuesto asignado desde la planta directiva y la tecnología no utilizada con fines comerciales ni de modelado.
----------	---

**4.2 Diseño Detallado de Procesos TO BE en IDEF0**

**4.2.1 Nivel A1: Planificación del negocio:**

El detalle en nivel 1 del macroproceso 3 “Planificación del negocio” indica los procesos necesarios para generar y ajustar la estrategia del negocio según las condiciones de mercado y la respuesta de clientes y usuarios. Se compone de 4 procesos principales:

- Definir el concepto del negocio y visión
- Gestionar desarrollo de estrategia
- Desarrollo de estrategia
- Mantención de estado

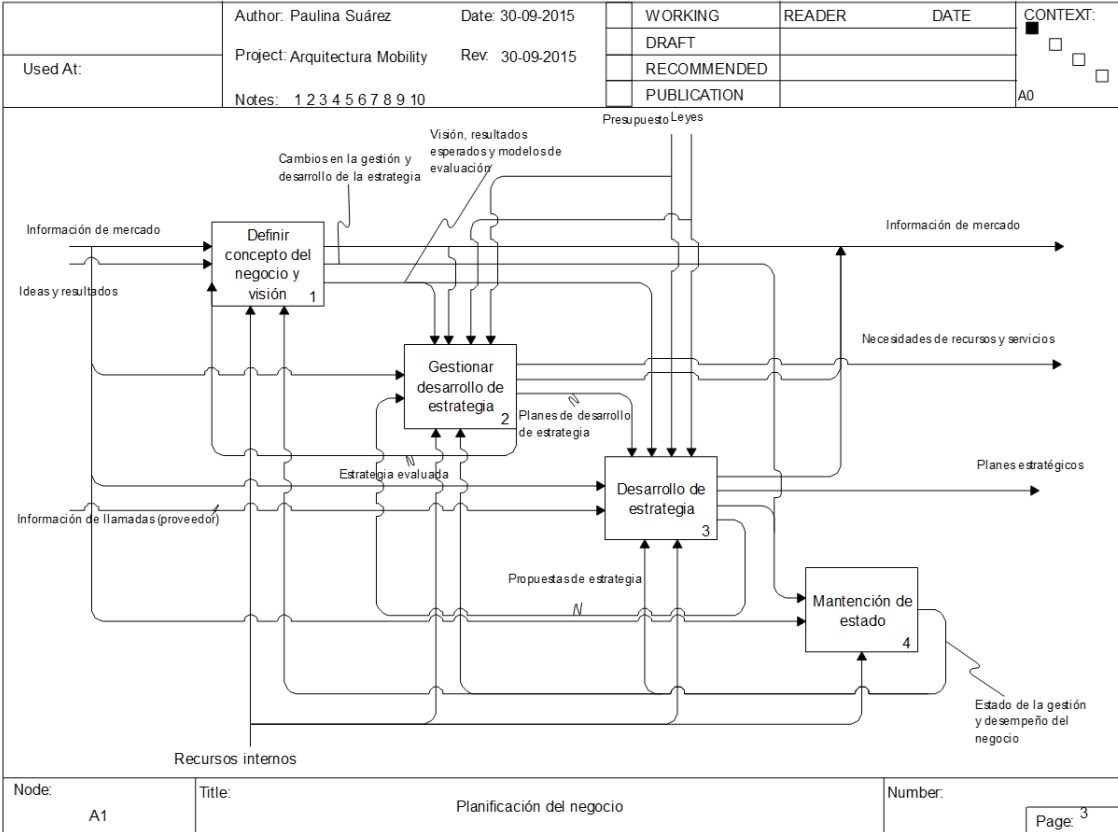


Figura 17. Arquitectura de negocios A1 Planificación del negocio

Tabla 6. Detalle de entradas y salidas. Nivel A11 Definir Concepto del negocio y visión

Inputs:	La información de mercado, planes de desarrollo de estrategia generadas en el Desarrollo de la Estrategia e ideas provenientes del Desarrollo de Nuevas Capacidades y/o Cadena de Valor son utilizadas para definir el concepto del negocio, la misión, visión, valores, generar el análisis estratégico, entre otras cosas.
Mecanismos:	Para lograr el objetivo de este proceso se utilizan recursos internos entregados por el macro proceso de nivel A0 Procesos de Apoyo, además de los recursos de retroalimentación entregados por la Mantenición de Estado, esto para controlar que el concepto a crear sea el correcto.
Outputs:	Como resultado se genera información de la imagen corporativa. Además esta información (visión, resultados esperados y modelos de evaluación) debe ser transmitida internamente para la Gestión del Desarrollo de la Estrategia. Por último, todos los cambios en la gestión y desarrollo de la estrategia son llevados como información de estado para el control de este macroproceso.

Tabla 7. Detalle de entradas y salidas. Nivel A12 Gestionar Desarrollo de estrategia

Inputs:	La Gestión del Desarrollo de la Estrategia se realiza a partir de la información del mercado, la que se utiliza como contexto externo. Además considera las propuestas de estrategias generadas en el Desarrollo de Estrategia, para trabajar según el medio interno.
Mecanismos:	Utiliza tanto los recursos internos entregados por los procesos de apoyo, como los recursos de retroalimentación generados por la Consolidación de Estado.
Controles:	Las leyes y presupuesto descritos en el nivel A0 son controles de este macroproceso, además debe estar alineado tanto con la visión y resultados esperados, como con la información de mercado generada para el entorno externo.
Outputs:	El proceso tiene por objetivo la creación de los planes de desarrollo de la estrategia, junto con información que entregará al mercado y un listado de necesidades de recursos y servicios que deberán ser

	cubiertos por los Procesos de Apoyo del nivel A0.
--	---

Tabla 8. Detalle de entradas y salidas. Nivel A13 Desarrollo de estrategia

Inputs:	Para desarrollar la estrategia es importante tener claro los recursos que se recibirán por parte del proveedor, en este caso la información de los CDR de llamadas y mensajería. Por otro lado, la estrategia requiere de la información de mercado, de modo que la estrategia permita a la empresa ser competitiva.
Mecanismos:	Al igual que en los dos procesos anteriores se utilizan tanto recursos internos de personal, financieros y otros, como recursos de retroalimentación entregados por la Mantención de Estado.
Controles:	Debe seguir la guía de los planes de desarrollo de estrategia entregados por el proceso de Gestión de Desarrollo de la Estrategia, acotarse a las leyes y presupuesto que rigen el proceso general y cumplir con la visión, resultados esperados (instanciados en objetivos, <i>KPI's</i> y metas) y modelos de evaluación determinados en la definición del concepto del negocio y visión.
Outputs:	Este proceso entrega como resultado los planes estratégicos de ventas, marketing, recursos humanos y operaciones alineados con los objetivos determinados en los procesos anteriores. La información de estado (cambios en la gestión y desarrollo de la estrategia) son enviados al proceso de Mantención de Estado para ejecutar los controles de implementación necesarios.  Por otro lado, genera información complementaria a la imagen corporativa entregada al entorno externo.

Tabla 9. Detalle de entradas y salidas. Nivel A14 Mantención de estado

Inputs:	La información de estado recibida corresponde a los cambios en la gestión y desarrollo de la estrategia en distintos niveles: a nivel de definición del negocio y a nivel de estrategia desarrollada.
Mecanismos:	Se usan los recursos internos de la empresa para generar el control de gestión.



Outputs:	El resultado son señales del estado de la gestión y desempeño del negocio, en particular se controla la implementación de la estrategia (cumplimiento de los requisitos propuestos, alcance de metas estratégicas, etc.) y se controla la estrategia (revisión, cuestionamiento y ajuste de esta de ser necesario).
----------	---

#### **4.2.2 Nivel A3: Desarrollo de campañas de marketing:**

El detalle en nivel 1 del macroproceso 1 “Desarrollo de campañas de marketing” indica los procesos necesarios para ejecutar la cadena de valor del negocio. Se compone de 4 procesos principales:

- Administración de la relación con el cliente y usuario: relacionado con el proceso de *Opt-in*.
- Administración de la relación con proveedores
- Gestión de producción y entrega
- Producción y entrega

Los 3 macroprocesos anteriores se relacionan en distintos niveles con el proceso de negocios de producción y entrega:

- Mantención de estado

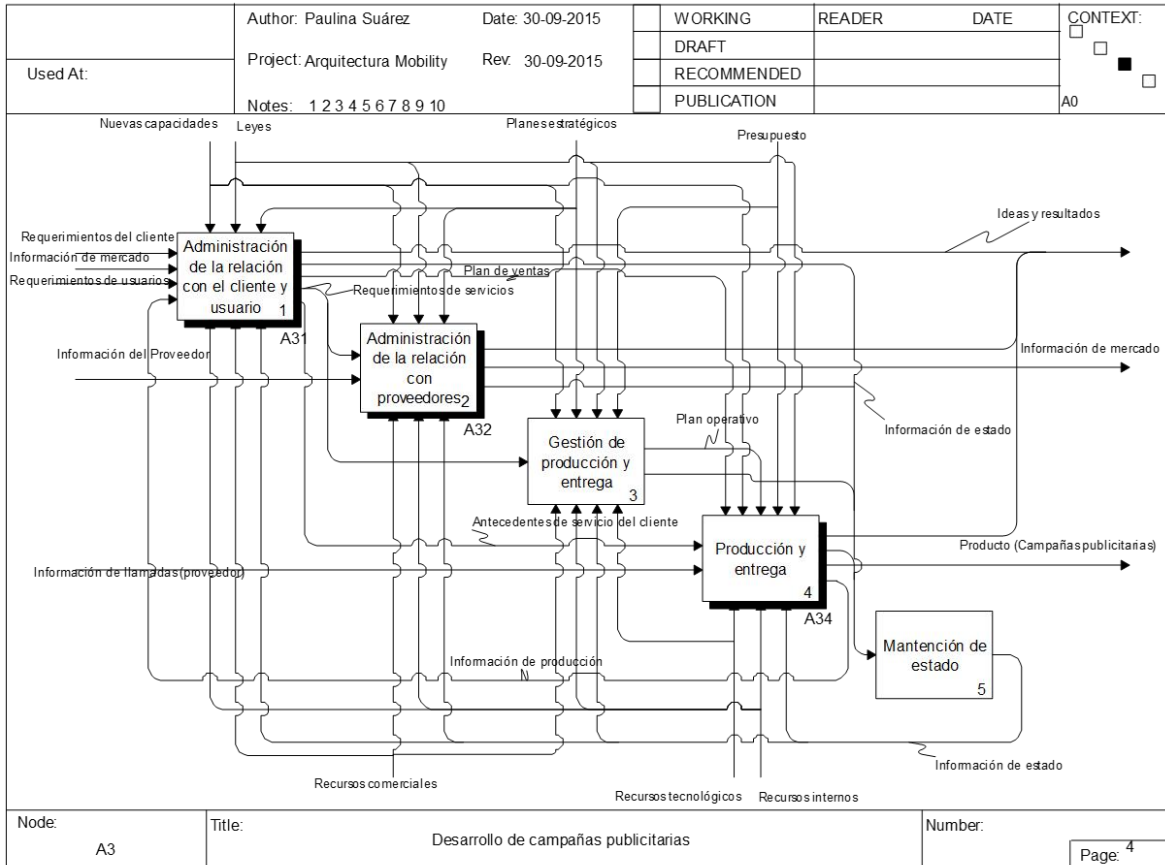


Figura 18. Arquitectura de negocios A3 Desarrollo de campañas de marketing

Tabla 10. Detalle de entradas y salidas. Nivel A31 Administración de la relación con clientes y usuarios

Inputs:	<p>Como insumos del proceso se tienen los requerimientos de los clientes (principalmente cuándo y de qué Centro de Gravedad y empres es el mensaje a enviar y qué segmentos desea cubrir la campaña) y los requerimientos de los usuarios (de poca invasividad, alta privacidad y efectividad del tipo de promociones enviadas). La relación con los usuarios se administra desde la empresa para los servicios de mensajes de beneficios. Por otro lado se utiliza la información del mercado como un referente en la relación con clientes y usuarios, y la información de retroalimentación del proceso de Producción y Entrega ejecutado (las necesidades creadas a partir de la ejecución que no fueron cubiertas en el proceso de Administración de la Relación con el Cliente anteriormente).</p>
Mecanismos:	<p>Se utilizan recursos comerciales y recursos internos, según la distinción realizada en el nivel A0. Además se utiliza la información de estado</p>

	como recurso retroalimentador al proceso, proveniente desde la Mantención de Estado.
Controles:	<p>La relación con clientes y sobre todo con los usuarios debe estar regida por las leyes de Protección al Consumidor y Protección a los Datos del Consumidor. Los planes estratégicos generados en el nivel A1 (Planificación del Negocio) deben ser utilizados como guía para encausar la relación al objetivo de negocio de la empresa.</p> <p>Cualquier nueva capacidad generada para la Administración de la Relación con el cliente y usuario debe ser utilizada como norma de relacionamiento.</p>
Outputs:	<p>Desde la relación podrán generarse nuevas ideas para la gestión de la comunicación con clientes y usuarios, los cuales podrían cambiar el enfoque estratégico utilizado en la Planificación del Negocio.</p> <p>Una vez establecido el contacto con los clientes se obtendrán requerimientos para el servicio a prestar (como tipo de mensaje, servicios de postventa esperados, cantidad de cobertura, entre otros) y antecedentes necesarios para la producción del servicio (localización y tiempo del segmento a enviar mensajes, información de retroalimentación del uso de promociones enviadas para clientes antiguos).</p> <p>Además se entregan los resultados como información de estado para su control posterior.</p>

Tabla 11. Detalle de entradas y salidas. Nivel A32 Administración de la relación con proveedores

Inputs:	<p>Para consolidar los requerimientos a proveedores es necesaria la información de requerimientos de servicios generada desde la Relación con Clientes y Usuarios, por otro lado es necesario conocer la capacidad y términos de trabajo del Proveedor (Información del Proveedor).</p> <p>Se habla de proveedores de forma genérica, ya que la empresa debe ser capaz de trabajar con cualquier empresa de telecomunicaciones capaz de entregar la información de sus CDR y enviar mensajes a un grupo de usuarios seleccionados.</p>
---------	--

Mecanismos:	Al igual que en el proceso de Administración de la Relación con el Cliente y Usuario se utilizan recursos comerciales y recursos internos, según la distinción realizada en el nivel A0. Además se utiliza la información de estado como recurso retroalimentador al proceso, proveniente desde la Mantención de Estado.
Controles:	La gestión de la Relación con Proveedores debe darse en el marco legal y financiero general de la empresa, en base a los planes estratégicos entregados por el proceso de nivel A1 y rigiéndose a las nuevas capacidades provistas por el macroproceso de Generación de Nuevas Capacidades.
Outputs:	Este proceso podría generar nuevas ideas para la gestión de la comunicación con proveedores, los cuales podrían cambiar el enfoque estratégico utilizado en la Planificación del Negocio.  Por otro lado, se entregará información al mercado, sobre las alianzas de la empresa con sus proveedores. Además se entregan los resultados como información de estado para su control posterior.

Tabla 12. Detalle de entradas y salidas. Nivel A33 Gestión de producción y entrega

Inputs:	Una vez obtenidos los requerimientos de servicios, se procederá a planear la producción que responderá a ellos.
Mecanismos:	Al igual que en el proceso de Administración de la Relación con el Cliente y Usuario se utilizan recursos comerciales y recursos internos, según la distinción realizada en el nivel A0. Además se utiliza la información de estado como recurso retroalimentador al proceso, proveniente desde la Mantención de Estado.
Controles:	La Gestión de Producción y Entrega debe considerar como restricciones el presupuesto disponible para producción, el no violar las leyes de consumidor descritas en el punto 2.3, el plan operativo a nivel estratégico, de ventas, recursos humanos y marketing (ya que estos fijan los objetivos y límites de la producción) y las posibles nuevas capacidades existentes.
Outputs:	A partir de lo anterior se genera un plan operativo a nivel operacional e

	información de estado para el control de los procesos.
--	--

Tabla 13. Detalle de entradas y salidas. Nivel A34 Producción y entrega

Inputs:	La producción requiere tanto de los datos de los CDR de llamadas y mensajería como de la posible información obtenida a través del proceso de <i>Opt-in</i> (preferencias de los usuarios). Lo anterior sumado a los antecedentes de servicio del cliente permitirá generar un <i>subset</i> de usuarios receptores de mensajería para Centros de Gravedad y tiempos determinados.
Mecanismos:	Se utilizan recursos tecnológicos especializados para la creación y uso de modelos, optimización y almacenamiento de datos. Igualmente, se requieren recursos internos de personal y financiamiento e información de estado recibida desde la Mantención de Estado.
Controles:	<p>Todo producto debe generarse en el marco legal correspondiente y sin superar las restricciones presupuestarias.</p> <p>Las nuevas capacidades para la producción (como nuevos algoritmos de modelado, procesamiento u optimización) deben ser el referente de la producción y entrega. Los planes estratégicos usados en la Gestión de Producción y Entrega, son también un marco regulatorio en este proceso, de manera de alinear el nivel operacional con el estratégico.</p>
Outputs:	<p>El producto principal entregado con campañas publicitarias a través de mensajes de texto (SMS) y/o información dada por una aplicación móvil y sus correspondientes reportes de respuesta. Este macroproceso se relaciona con el proceso de CRISP-DM.</p> <p>Los movimientos de producción y entrega generan información relevante para el mercado, según la participación de mercado, efectividad del servicio, opiniones de clientes y usuarios entre otras cosas.</p> <p>Por último se entrega información interna de la eficacia y eficiencia de la producción como información de estado para el control de este proceso.</p>

Tabla 14. Detalle de entradas y salidas. Nivel A35 Mantenición de estado

Inputs:	La información de estado recibida corresponde al rendimiento de la relación con clientes, la producción y entrega de productos, posibles cambiones en alguno de estos tópicos, fallas, cumplimiento de metas entre otras cosas.
Mecanismos:	Se usan los recursos internos de la empresa para generar el control de gestión.
Controles:	El resultado son señales del estado de la gestión y desempeño de la cadena de valor, en particular se controlan los objetivos especificados en los planes estratégicos como captación efectiva de clientes, captación efectiva de usuarios, nivel de respuesta de producción, nivel de conversión de los modelos, sensibilidad de los modelos según el comportamiento de los usuarios, etc.

#### **4.2.3 Nivel A31: Administración de la relación con el cliente y usuario**

La administración con el cliente se descompone en 4 procesos:

- Análisis de campañas de marketing
- Atención a usuarios
- Venta y atención al cliente
- Procesamiento de órdenes

Más detalle del proceso de administración con el cliente y usuarios se encuentra en la sección de anexos (ver 4.2.4).

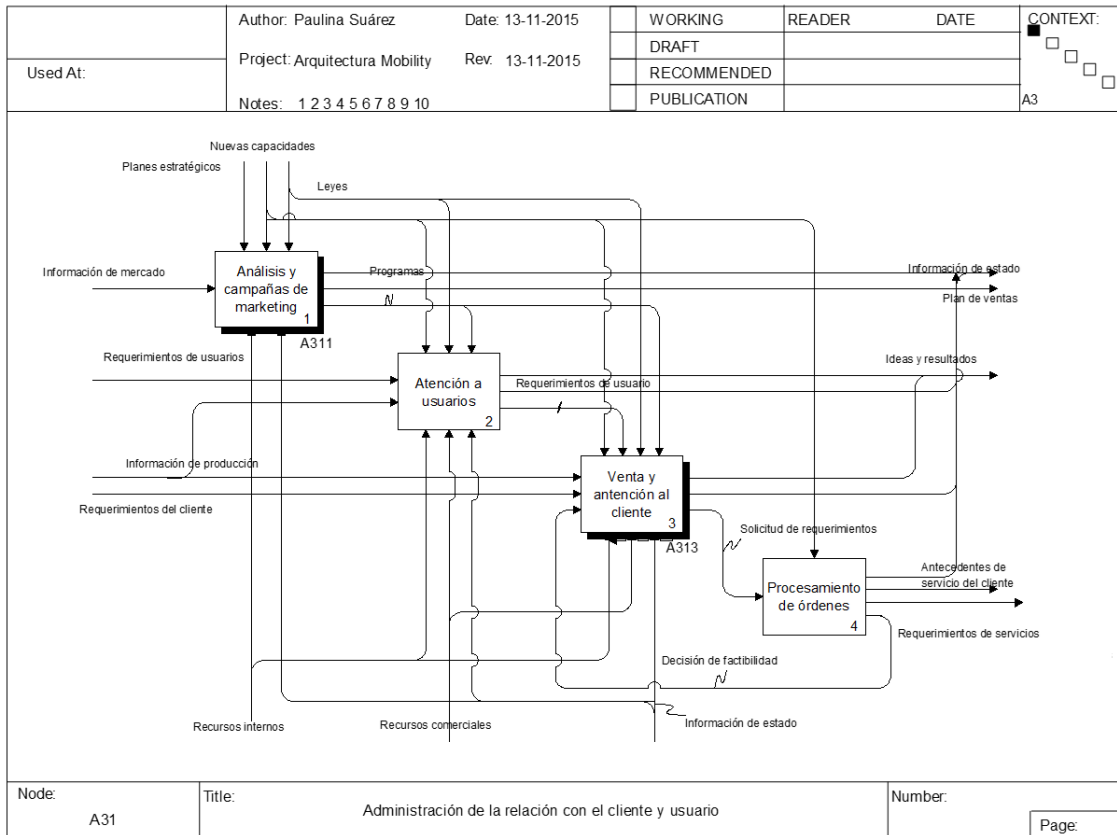


Figura 19. Arquitectura de negocios A31: Administración de la relación con el cliente y usuario

Tabla 15. Detalle de entradas y salidas. Nivel A311 Análisis y campañas de marketing

<b>Inputs:</b>	Para la generación de las campañas de marketing hacia los usuarios (para la venta de publicitarias a los usuarios) se necesitará una comparación con las campañas realizadas por la competencia y la aceptación o rechazo de estas por parte de los clientes.
<b>Mecanismos:</b>	Utiliza mecanismos internos de la empresa, es decir, tecnología no asociada a la creación de modelos y personal necesario, entregado por procesos de apoyo. Además requiere de la información de estado, donde se espera usar indicadores de efectividad de las campañas y recursos empleados en, entre otros.
<b>Controles:</b>	Al igual que otros procesos, deben regirse por las leyes comerciales y por las nuevas capacidades generadas en la empresa. Por otro lado una campaña debe cumplir con los objetivos de los planes estratégicos generados en el macroproceso de planificación del negocio.

Outputs:	Este proceso tiene como objetivo entregar programas que guíen las ventas de campañas publicitarias. Además debe generar la planificación de marketing del plan de ventas de la empresa y entregar la información de estado de proceso para mantenerlo en control.
----------	---

Tabla 16. Detalle de entradas y salidas. Nivel A312 Atención a usuarios

Inputs:	<p>La atención a usuarios recibe como información de entrada los requerimientos de estos con respecto al servicio prestado por la empresa y el comportamiento de los clientes para llevar a cumplimiento las promociones.</p> <p>Por otro lado, desde el proceso de producción y entrega se recibe información de retroalimentación sobre el funcionamiento de las campañas publicitarias y la reacción de los usuarios frente a estas.</p>
Mecanismos:	Se utilizarán tanto los recursos comerciales como los recursos internos de la empresa. La información de estado recibida desde el proceso de mantención de estado también funciona como mecanismo para la ejecución de la atención.
Controles:	Se rige por las leyes comerciales establecidas, los programas de marketing derivados del proceso de análisis y campañas de marketing y cualquier nueva capacidad que aplique a este proceso.
Outputs:	<p>Como resultado se genera información de retroalimentación: ideas de mejora y resultados del proceso los cuales luego se utilizarán en el proceso de e información de estado para el control del proceso.</p> <p>Igualmente se entregan nuevos requerimientos de usuarios necesarios como mecanismos de control del proceso de venta y atención a usuarios.</p>

Tabla 17. Detalle de entradas y salidas. Nivel A313 Venta y atención al cliente.

Inputs:	Al igual que el proceso anterior, la venta y atención al cliente requiere de la información de producción y entrega de las campañas publicitarias móviles, los cuales son el respaldo que la empresa entrega a sus clientes y con las cuales se pueden programar las órdenes de producción de próximas campañas.
---------	--



	Los requerimientos de los clientes son necesarios para establecer el tipo de venta que se realizará, incluso la viabilidad de esta. En este sentido se esperan tanto los requerimientos de performance esperada de las campañas, como la información que esperan obtener y el público al cual se desea cubrir con las campañas publicitarias.
Mecanismos:	Este proceso también utiliza la información de estado y recursos internos y comerciales de la empresa para operar.
Controles:	Su guía son los requerimientos de usuarios entregados por el proceso anterior, las leyes comerciales establecidas, los programas de marketing entregados por el proceso de análisis y campañas de marketing y cualquier nueva capacidad generada en el proceso de desarrollo de nuevas capacidades.
Outputs:	Se generará una solicitud de requerimientos hacia el procesamiento de órdenes, en el cual se destacan aspectos como, la zona geográfica a analizar, variables extraordinarias, resultados y análisis esperados, plazos, entre otros. Este macro proceso está relacionado con proceso de negocios de producción y entrega.  Por otro lado el proceso entrega información de retroalimentación: ideas, resultados e información de estado.

Tabla 18. Detalle de entradas y salidas. Nivel A314 Procesamiento de órdenes

Inputs:	Este proceso recibe las solicitudes de requerimientos generadas en el proceso de venta y atención al cliente.
Controles:	Deberá regirse a cualquier nueva capacidad generada en el proceso de desarrollo de nuevas capacidades.
Outputs:	Genera los requerimientos de servicios para la gestión de la producción.  Este macroproceso está relacionado con el proceso de negocios de captación de clientes

#### 4.2.4 Nivel A311: Análisis y campañas de marketing

Este proceso consta de las etapas de:

- Introducción de nuevos análisis y modelos
- Análisis de las respuestas de usuarios
- Análisis del comportamiento de clientes
- Definir acciones de marketing planificar ventas.

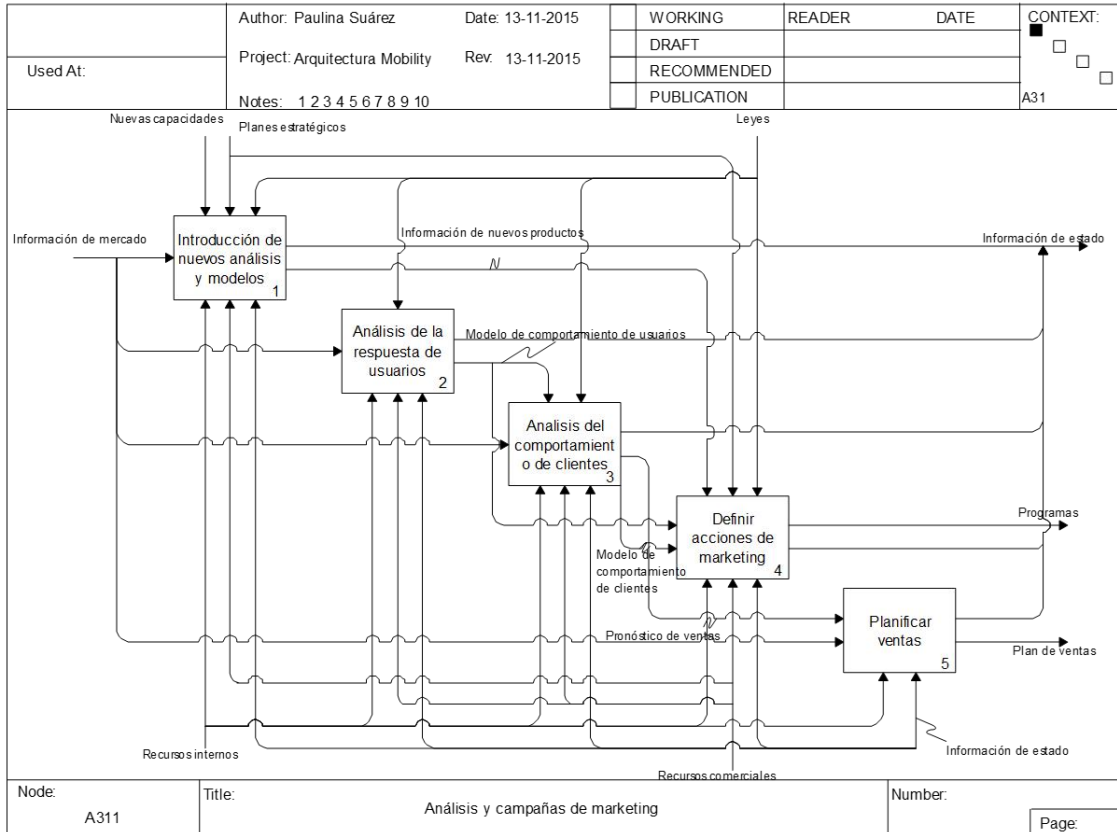


Figura 20. Arquitectura de negocios A311: Análisis y campañas de marketing

Tabla 19. Detalle de entradas y salidas. Nivel A3111 Introducción de nuevos análisis y modelos

<b>Inputs:</b>	Gracias a la información de mercado y conocimiento de nuevas aplicaciones de modelos, se podrá dar paso a introducir nuevos tipo de análisis y ejecución de modelos para la obtención de segmentos de usuarios (a nivel comercial).
<b>Mecanismos:</b>	Usa los recursos internos de la empresa entregados por los procesos de apoyo que esta posee. Además debe trabajar con mejora continua

	en base la información de estado
Controles:	Al igual que la mayoría de los procesos debe regirse por las leyes de privacidad de la información, las nuevas capacidades y los planes estratégicos generados en el proceso de gestión del desarrollo de estrategia.
Outputs:	El resultado de este proceso es un reporte de los nuevos productos de análisis y modelado de segmentos generados.

Tabla 20. Detalle de entradas y salidas. Nivel A3112: Análisis de la respuesta de usuarios

Inputs:	Las respuestas de los usuarios a las campañas de publicidad móvil se recibirán desde el proceso de producción y entrega. Por otro lado, es necesaria información comparativa del mercado para verificar que las metas puestas en estas campañas están bien fijadas y si se están o no cumpliendo.
Mecanismos:	En este proceso los recursos internos, de personal capacitado y tecnologías asociadas son fundamentales, siempre acompañadas de la información de estado provista por el proceso de mantención de estado.
Controles:	La respuesta a campañas debe analizarse usando las capacidades existentes y verificando el cumplimiento de las metas de los planes estratégicos.
Outputs:	El análisis entregará un informe con el modelo de comportamiento de usuarios.

Tabla 21. Detalle de entradas y salidas. Nivel A311: Análisis del comportamiento de clientes

Inputs:	De forma análoga al proceso anterior, el comportamiento de los clientes se analiza a partir de la información recibida desde el proceso de producción y entrega y se comparará con la información de mercado
Mecanismos:	Requiere de analistas entregados por los procesos de apoyo y la información de control generada en la mantención de estado.

Controles:	Al igual que el proceso anterior, debe seguir los planes estratégicos y utilizar las nuevas capacidades. Además debe cumplir con las expectativas entregadas en el modelo de comportamiento de usuarios
Outputs:	El análisis entregará un informe con el modelo de comportamiento de los clientes, un pronóstico de ventas y la información de estado necesaria para el control del proceso.

Tabla 22. Detalle de entradas y salidas. Nivel A3113: Definir acciones de marketing

Inputs:	Para definir los nuevos pasos a seguir se requiere tanto del modelo de comportamiento de los clientes como el modelo de comportamiento de los usuarios.
Mecanismos:	Al igual que los procesos anteriores, la definición de las acciones de marketing requiere a personal con conocimiento en marketing, tecnología asociada y la información de estado necesaria para la mejora continua.
Controles:	Las acciones de marketing nos deben transgredir las leyes de privacidad de los usuarios, deben cumplir con los objetivos de los planes estratégicos y utilizar cualquier nueva capacidad generada atingente con el proceso.
Outputs:	De acá nacen programas de marketing e información de la performance del proceso.

Tabla 23. Detalle de entradas y salidas. Nivel A3114: Planificar ventas

Inputs:	La planificación se desarrolla en base al pronóstico de ventas generado en el proceso de análisis del comportamiento de clientes y del contexto de mercado.
Mecanismos:	Utiliza los recursos internos entregados por los procesos de apoyo y la información del funcionamiento del proceso.
Controles:	Debe cumplir con los objetivos de los planes estratégicos

Outputs:	Genera el plan de ventas de la empresa y entrega información de estado para mantener el control y posibilitar mejoras en el proceso.
----------	--

**4.2.5 Nivel A313: Venta y atención al cliente**

La venta y atención al cliente se subdividen en venta, postventa y seguimiento. La venta está relacionada con el proceso de Producción y entrega, mientras que los dos últimos con el proceso de retroalimentación y postventa.

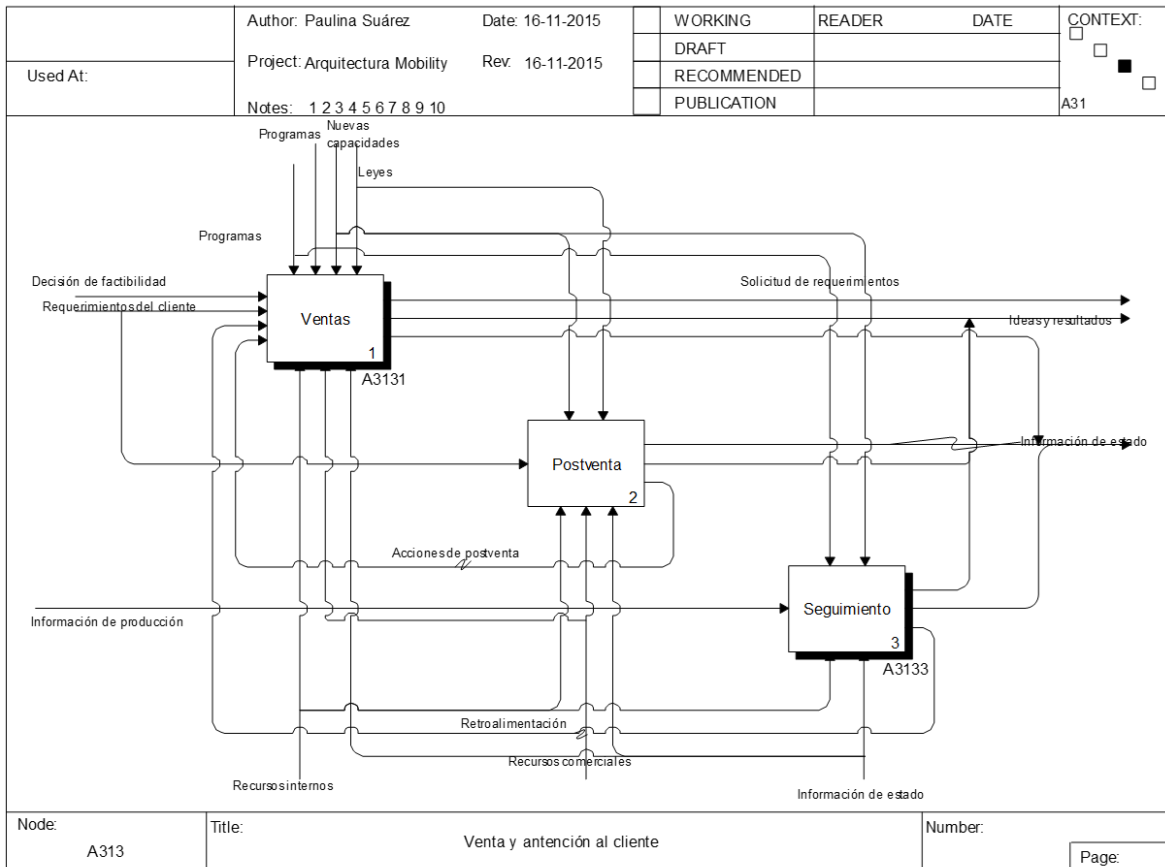


Figura 21. Arquitectura de negocios A313: Venta y atención al cliente

Tabla 24. Detalle de entradas y salidas. Nivel A3131 Ventas

Inputs:	El proceso de ventas requiere tanto de los requerimientos de compra del cliente como de la decisión de factibilidad entregada desde el procesamiento de órdenes.
Mecanismos:	Este proceso utiliza fuertemente los recursos comerciales de la

	empresa con los cuales mantienen la comunicación con los clientes. Por otro lado se necesitan los recursos internos entregados por los procesos de apoyo y la información de estado que permite mantener en control el proceso.
Controles:	Las ventas de campañas publicitarias no deben transgredir las leyes comerciales y los programas de ventas generados. Deben cumplir con las expectativas de los usuarios y utilizar cualquier nueva capacidad pertinente al proceso.
Outputs:	El proceso de ventas envía una solicitud de requerimientos al procesamiento de órdenes, así como también envía información de retroalimentación (ideas y resultados) e información de estado.

Tabla 25. Detalle de entradas y salidas. Nivel A3132 Postventa

Inputs:	El proceso de postventa recibe los requerimientos de postventa de clientes (informes de análisis y resultados esperados).
Mecanismos:	Utiliza los recursos comerciales como principal fuente de comunicación, además de recursos internos y la información de estado enviada desde el macroproceso de consolidación de estado.
Controles:	Debe usar las nuevas capacidades creadas para el proceso y regirse por las leyes comerciales vigentes.
Outputs:	Genera acciones de retroalimentación, las principales son las acciones de postventa que alimentarán el proceso de ventas. También entregan ideas y resultados necesarios para re-desarrollar la estrategia e información de estado, usadas en el control de los procesos.

Tabla 26. Detalle de entradas y salidas. Nivel A3133 Seguimiento

Inputs:	Desde el proceso de producción y entrega se recibe la información de producción necesaria para mantener un seguimiento de las ventas generadas.
Mecanismos:	Utiliza tanto los recursos internos de la empresa como la información

	de estado necesaria para mantener el proceso controlado.
Controles:	Debe utilizar las nuevas capacidades pertinentes al proceso generadas desde el desarrollo de nuevas capacidades y regirse a las leyes vigentes.
Outputs:	Entrega resultados e ideas relacionadas con la venta y el seguimiento de estas e información de estado al macroproceso de mantención de estado.

#### 4.2.6 Nivel A3131: Ventas

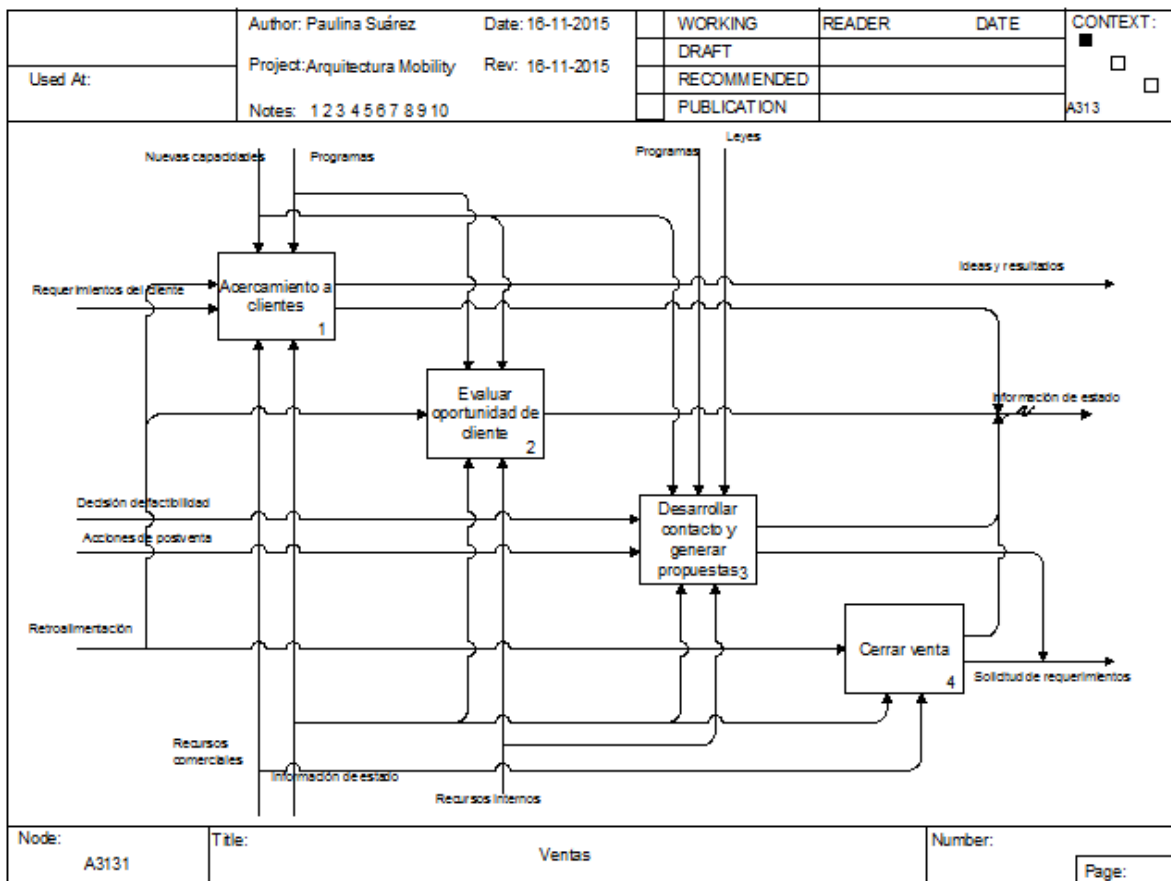


Figura 22. Arquitectura de negocios A3131: Ventas

Tabla 27. Detalle de entradas y salidas. Nivel A31311 Acercamiento a clientes

Inputs:	El acercamiento a clientes se logra a partir de la información recopilada sobre sus requerimientos e información de
---------	---

	retroalimentación del proceso de seguimiento de ventas.
Mecanismos:	Como mecanismos requiere tanto de los recursos comerciales de la empresa como de la información de estado obtenida de iteraciones anteriores de este proceso.
Controles:	Usa cualquier nueva capacidad y los programas obtenidos desde el macroproceso de análisis y campañas de marketing.
Outputs:	Genera información de estado, ideas y resultados referentes al proceso de acercamiento a clientes.

Tabla 28. Detalle de entradas y salidas. Nivel A31312 Evaluar oportunidad de cliente

Inputs:	La información de retroalimentación de ventas es el material principal para generar evaluaciones de oportunidades de clientes, obteniendo sus necesidades y recursos.
Mecanismos:	Utiliza recursos internos de la empresa (tecnológicos y de personas no destinados a la producción) e información de estado del proceso de evaluación de oportunidades de clientes.
Controles:	Se riega por los programas de marketing u utiliza cualquier nueva capacidad generada en el macroproceso de desarrollo de nuevas capacidades.
Outputs:	Entrega información de estado correspondiente a este proceso.

Tabla 29. Detalle de entradas y salidas. Nivel A31313 Desarrollar contacto y generar propuestas

Inputs:	Para el contacto de venta recibe la decisión de factibilidad desde el procesamiento de órdenes, mientras que para contactos de postventa utiliza las acciones de postventa generadas desde el proceso de postventa.
Mecanismos:	Utiliza recursos internos de la empresa (tecnológicos y de personas no destinados a la producción) e información de estado del proceso de evaluación de oportunidades de clientes.



Controles:	Responde a límites internos (programas de marketing y nuevas capacidades) y límites externos (leyes de protección al consumidor).
Outputs:	Genera un solicitud de requerimientos de venta (orden de trabajo) enviada al macroproceso de procesamiento de órdenes. Además de información de estado de este proceso.

Tabla 30. Detalle de entradas y salidas. Nivel A31314 Cerrar venta

Inputs:	El cierre de la venta, consiste en registrar una venta finalizada y se genera gracias a la información de retroalimentación enviadas desde el seguimiento de ventas (venta conforme).
Mecanismos:	Utiliza tanto recursos comerciales como la información de estado del proceso, en línea con la filosofía de mejora continua.
Outputs:	Produce solicitudes de requerimientos de cierre al macroproceso de procesamiento de órdenes.

## 4.2.7 Nivel A3133: Seguimiento

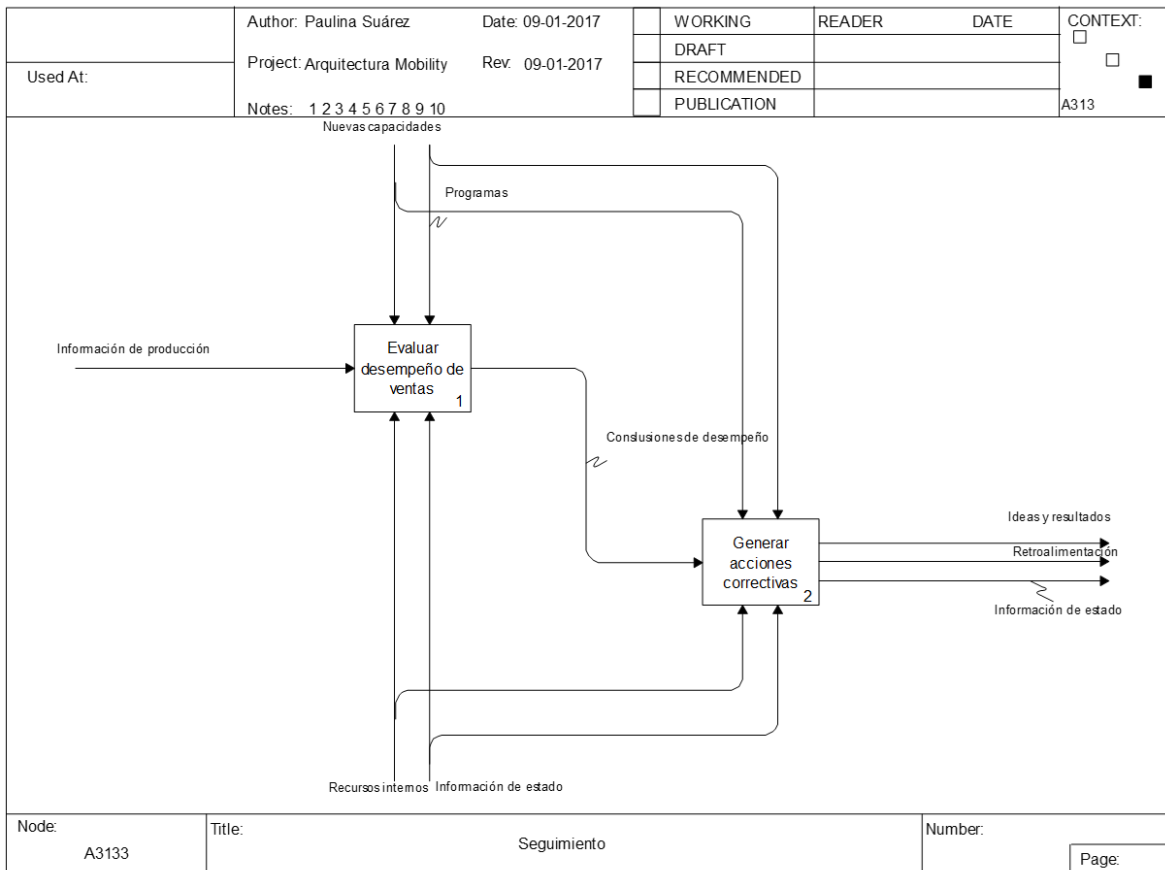


Figura 23. Arquitectura de negocios A3133: Seguimiento

Tabla 31. Detalle de entradas y salidas. Nivel A31331 Evaluar desempeño de ventas

Inputs:	Para la evaluación del desempeño se requieren los resultados de producción provenientes del macroproceso de producción y entrega.
Mecanismos:	Utiliza los recursos internos de la empresa e información de estado de iteraciones anteriores.
Controles:	Aplica las nuevas capacidades y programas generados en el macroproceso de atención a usuarios como mecanismos de control.
Outputs:	Genera un set de conclusiones de desempeño a partir del cálculo de métricas relevantes a las ventas de campañas de marketing.

Tabla 32 Detalle de entradas y salidas. Nivel A31332 Generar acciones correctivas

Inputs:	Las acciones correctivas se generan a partir de las conclusiones de desempeños del macroproceso anterior.
Mecanismos:	Utiliza los recursos internos de la empresa e información de estado de iteraciones anteriores.
Controles:	Aplica las nuevas capacidades y programas generados en el macroproceso de atención a usuarios como mecanismos de control.
Outputs:	Genera principalmente acciones de retroalimentación del proceso de ventas alineado con la filosofía de mejora continua. Por otro lado genera ideas y resultados e información de mantención de estado, del proceso de generación de acciones correctivas.

#### **4.2.8 Nivel A32: Administración de la relación con proveedores:**

La administración de la relación con proveedores consta de 3 subprocesos:

- Determinar requerimientos
- Programar transacciones
- Seguimiento de transacciones

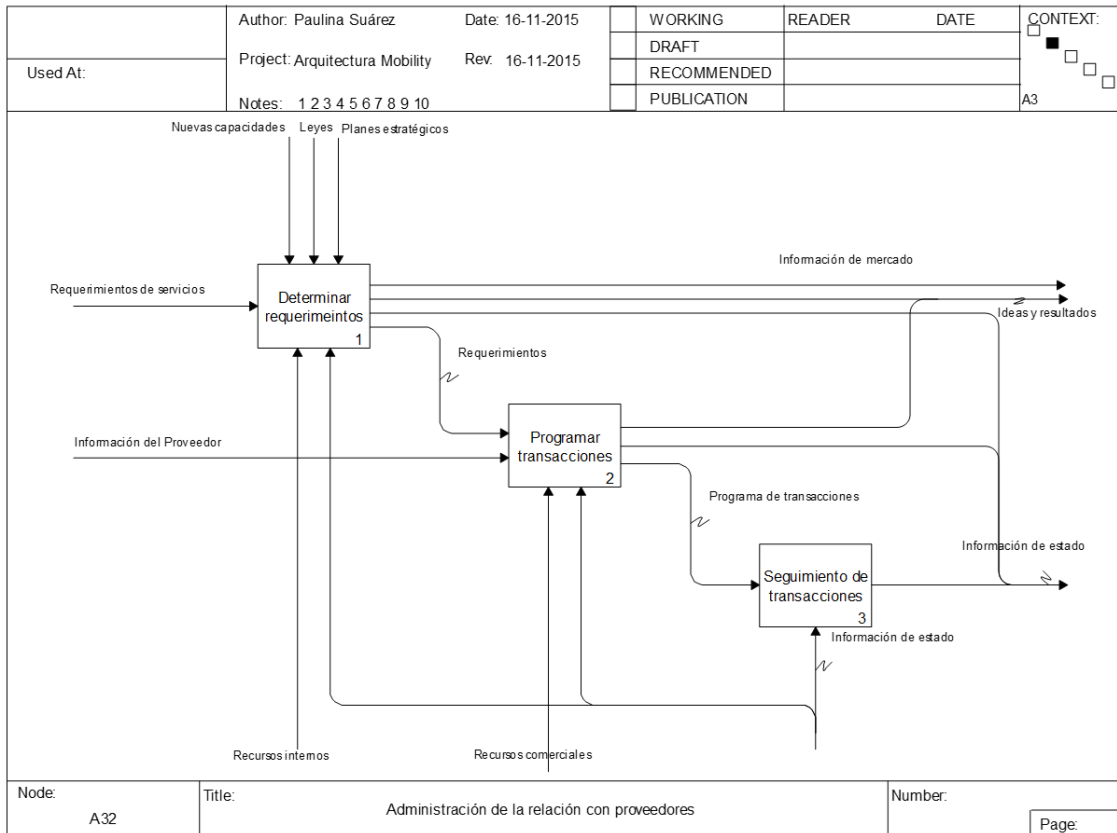


Figura 24. Arquitectura de negocios A32: Administración de la relación con proveedores

Tabla 33. Detalle de entradas y salidas. Nivel A321 Determinar requerimientos

<b>Inputs:</b>	Recibe los requerimientos de servicios desde el proceso de administración de la relación con el cliente y usuario.
<b>Mecanismos:</b>	Al igual que todos los procesos usa la información de estado desde el macroproceso de consolidación de estado. Por otro lado utiliza los recursos internos de personal y materiales entregados por los procesos de apoyo.
<b>Controles:</b>	Debe utilizar las nuevas capacidades generadas pertinentes al proceso y respetar las leyes y planes estratégicos de la empresa.
<b>Outputs:</b>	Entrega información al mercado de los requerimientos necesarios, ideas y resultados para mejorar la estrategia e información de estado. Además entrega requerimientos para la programación de transacciones

	(envío de mensajes y confirmación del envío) y uso de datos de CDR's.
--	---

Tabla 34. Detalle de entradas y salidas. Nivel A322 Programar transacciones.

Inputs:	Recibe los requerimientos para la programación de transacciones e información del proveedor.
Mecanismos:	Utiliza recursos comerciales para la comunicación con proveedores e información de estado para mantener el control.
Outputs:	Genera un programa de transacciones (que especifica, número de destino, fecha y hora de envío, y mensaje a enviar). Establece las reglas de actualización de datos móviles y permisos sobre estas bases de datos. Finalmente genera información de estado, e ideas y resultados del proceso.

Tabla 35. Detalle de entradas y salidas. Nivel A323 Seguimiento de transacciones

Inputs:	Requiere de un programa de transacciones que será seguido para asegurar su cumplimiento.
Mecanismos:	Utiliza recursos comerciales para la comunicación con proveedores e información de estado para mantener el control.
Outputs:	Genera un programa de transacciones (que especifica número de destino, mensaje a enviar, fecha y hora de envío), accesos a bases de datos (CDR's), información de estado, e ideas y resultados del proceso.

## 4.2.9 Nivel A34: Producción y entrega

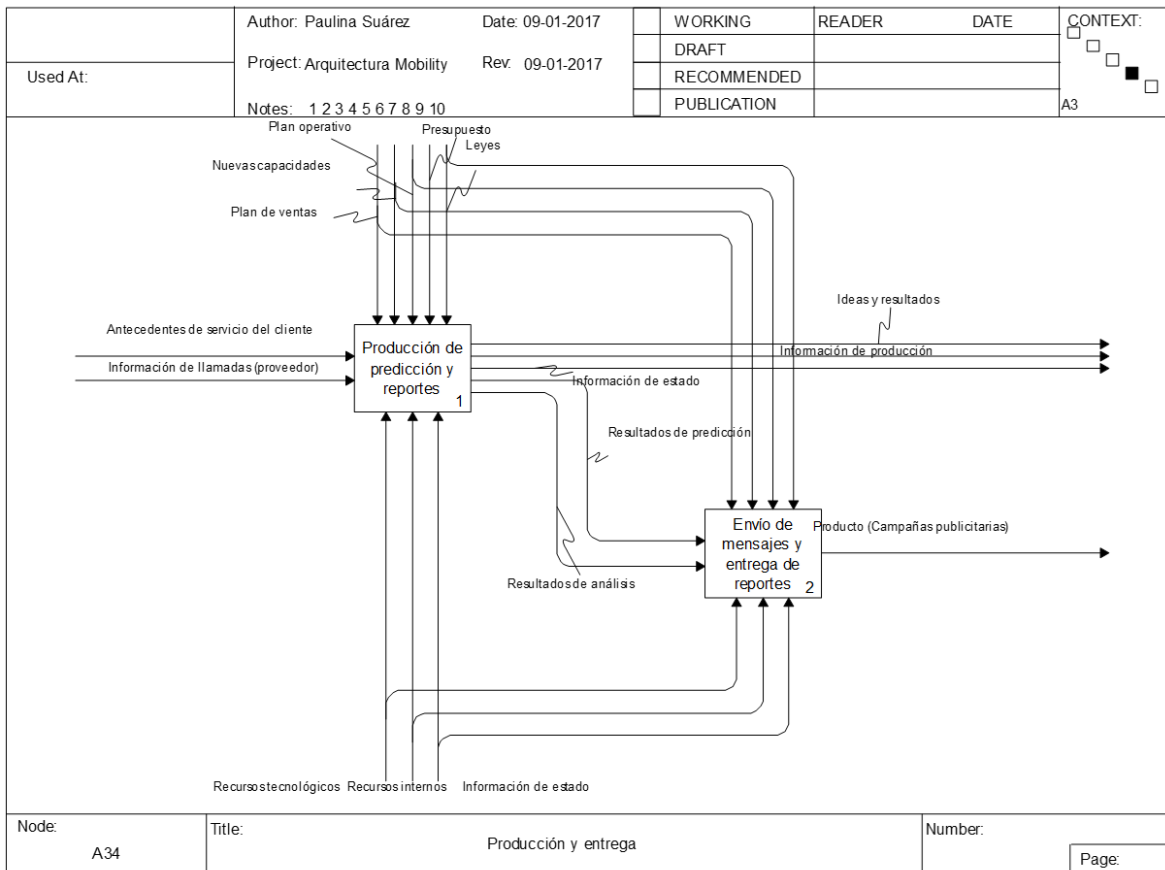


Figura 25. Arquitectura de negocios A34: Producción y entrega

Tabla 36. Detalle de entradas y salidas. Nivel A341: Producción de predicción y reportes

<b>Inputs:</b>	Recibe como input las bases de datos de registros de llamadas (CDR's) e información adicional de los usuarios obtenida a través del proceso de <i>Opt-in</i> .
<b>Mecanismos:</b>	Los recursos internos de personal y recursos tecnológicos asignados a la predicción son utilizados como recursos habilitadores. Por otro lado, la información de estado del proceso de producción de predicción y reportes se usa con el fin de mejorar su funcionamiento.
<b>Controles:</b>	Como mecanismos de control se tiene los planes de ventas, operativo, financieros (presupuesto), cualquier nueva capacidad generada en el macroproceso de desarrollo de nuevas capacidades y las leyes de


	protección al consumidor.
Outputs:	Los resultados de este macroproceso son tanto de predicción como de los análisis de las respuestas de usuarios (en coordinación con el cliente). La información de producción tanto de clientes, como de usuarios con entregados al macroproceso de seguimiento de ventas. Por último se entregan ideas, resultados e información de estado referentes al macroproceso de producción de predicción y reportes.


Tabla 37. Detalle de entradas y salidas. Nivel A342: Envío de mensajes y entrega de reportes.


Inputs:	El proceso recibe los resultados de predicción y análisis del macroproceso anterior.
Mecanismos:	Los recursos internos de personal y recursos tecnológicos asignados a la predicción son utilizados como recursos habilitadores. Por otro lado, la información de estado del proceso de producción de predicción y reportes se usa con el fin de mejorar su funcionamiento.
Controles:	Como mecanismos de control se tiene los planes de ventas, operativo, cualquier nueva capacidad generada en el macroproceso de desarrollo de nuevas capacidades y las leyes de protección al consumidor.
Outputs:	Los resultados son, por un lado el set de mensajes a enviar y en conjunto con la coordinación con el proveedor el envío de mensajes de beneficios al set generado. Además, genera reportes de campañas como producto a los clientes.


### 4.3 Diseño Detallado de Procesos TO BE de la cadena de valor en BPMN


#### 4.3.1 Archivos y bases de datos relacionadas a las los procesos


 BD *Opt-in*: corresponde a datos obtenidos desde el proceso de captación de usuarios y de la apertura de links que llevan a descuentos (declaración de interés por el beneficio).


 CDR: corresponde a la información de los CDR, entregadas por el aliado estratégico. Esta información, al menos, contiene los campos: número de teléfono del emisor (*integer*), latitud de antena emisora (*integer*), longitud de antena emisora (*integer*), fecha de llamada (*date*), hora de la llamada (*integer*), número de teléfono del receptor (*integer*), latitud de antena receptora (*integer*), longitud de antena receptora (*integer*).


 BD SMS: base de datos que registra la información de cada campaña en curso: *clusters*, usuarios por contactar, fechas para el contacto, mensajes enviados, fechas de envío de mensajes, respuesta de los usuarios, características de usuarios y características de la visita.


 DW desempeño: *Data Warehouse* histórico con datos de panel que considera las variables usadas en la operación de los modelos de *clustering* y pronóstico (ver 5.2), así como como las variables de declaración de interés, conversión y registro de visita de los identificadores asociados a las campañas.


 Plan operativo: corresponde a uno de los planes generados en el macroproceso de Planificación de la Estrategia (ver 4.2.2)

 Contrato: documento legal firmado entre el cliente y la empresa que establece los criterios de la relación entre las partes, como el tiempo de ejecución, cantidad de mensajes a enviar en la campaña, productos que la empresa debe entregar al cliente, multas por no cumplimiento, entre otras.


 Reporte de ventas: Reporte generado por el analista de producción con la información de *clusters* de cada Centro de Gravedad, generado con la finalidad de demostrar físicamente el producto final de las campañas de marketing móvil. Este reporte debe estar en línea disponible en todo momento para los ejecutivos de venta.

 Reporte de inicio: reporte de inicio de la campaña donde se detallan los mensajes a enviar, sus fechas de envío y sus *clusters* destinatarios.

 Reporte de avance: reporte que indica la cantidad de mensajes enviados, sus fechas de envío y *clusters* destinatarios, en contraste con lo entregado en el Reporte de inicio.

 Reporte final: reporte que muestra los resultados de la campaña y la descripción del tipo de visitas que tuvo el Centro de Gravedad de interés.



 Reporte de desempeño comercial: reporte generado por el área de marketing donde se muestra una comparación de resultados con la competencia y los registros de la relación con usuarios y clientes.

### 4.3.2 Retroalimentación

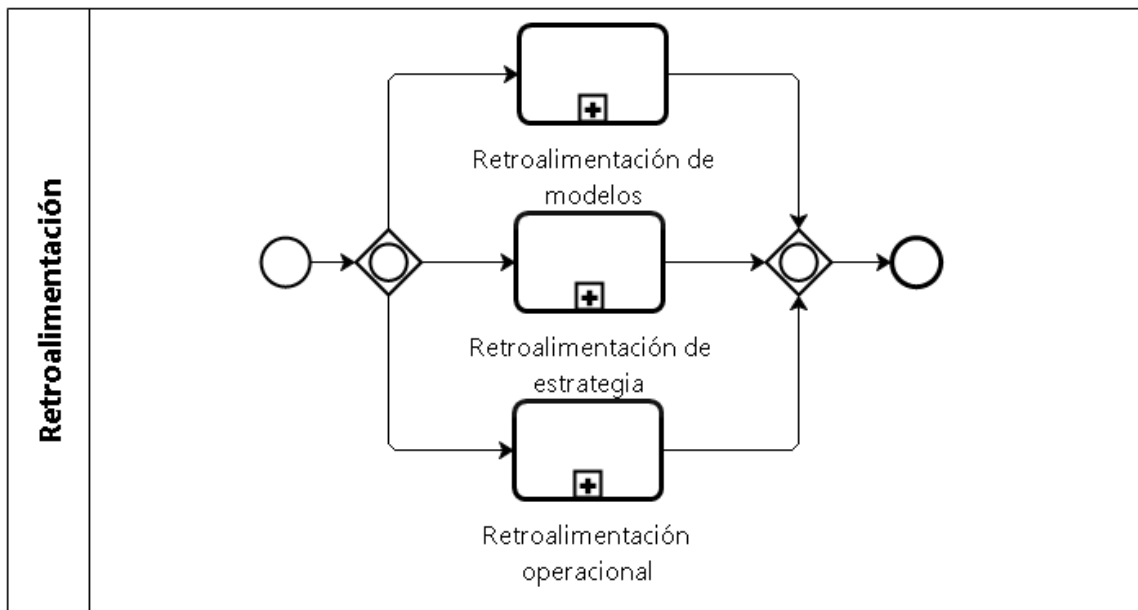





Figura 26. Proceso de Retroalimentación. En notación BPMN

 Evento de inicialización: la retroalimentación puede iniciarse luego de cada iteración de campañas, en cada mes, o semana, según el tipo de retroalimentación.

 Retroalimentación de modelos: corresponde al proceso de mejora de modelos de *clustering* y pronóstico utilizando la metodología *CRISP DM*.

 Retroalimentación de estrategia: corresponde al proceso de mejora de estrategia de negocios, donde se analizan los factores internos y externos del negocio.

 Retroalimentación operacional: corresponde al proceso de mejora de la operación de campañas de marketing, relación con actuales clientes y usuarios.

### 4.3.3 Retroalimentación de modelos

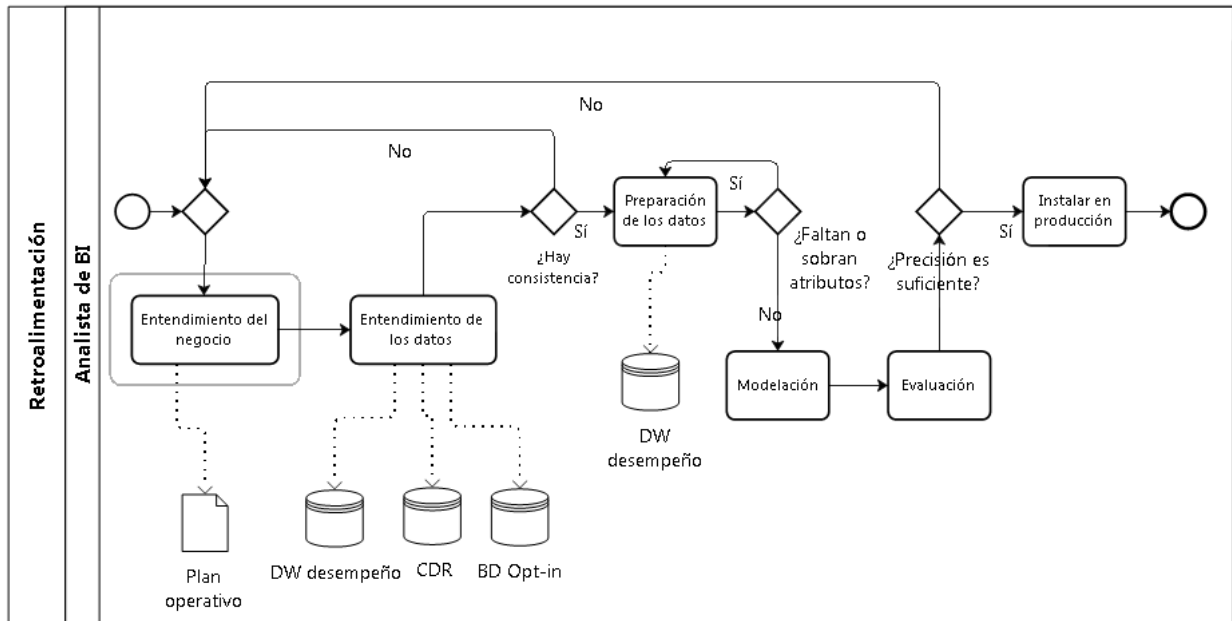







Figura 27. Proceso de Retroalimentación de modelos. En notación BPMN


 Evento de inicialización: el proceso se inicia una vez establecidos los objetivos estratégicos de la empresa y antes del lanzamiento del servicio. Luego tienen inicios iterativos para el mejoramiento continuo de los modelos.


 Entendimiento del negocio: el o los analista(s) de BI establecerá(n) de los objetivos y criterios de éxito de la aplicación en contexto con los objetivos del negocio. Realizará(n) una evaluación de la situación estableciendo un inventario de recursos, requerimientos, supuestos, terminologías propias del negocio y generarán el plan del proyecto, usando como restricción lo establecido en los planes operativos de la empresa.


 Entendimiento de los datos: la familiarización con los datos considera, la recopilación inicial, descripción, exploración y verificación de la calidad de los datos. En este caso se trabaja con, al menos, dos bases de datos: DW desempeño y Datos *Opt-in*.


 ¿Hay consistencia?: En esta etapa se decide si es necesaria o no una re-evaluación del negocio, de manera de tener datos que cumplan los requerimientos para lograr el objetivo propuesto.


 Preparación de los datos: si los datos son satisfactorios, se obtiene el *dataset* que se usará en los modelos, esto se logra con una serie de subprocessos de selección, limpieza, construcción, integración y formateo de los datos.


 ¿Faltan o sobran atributos?: Esta compuerta comprueba si el *dataset* generado es adecuado para ejecutar la tarea de modelado. Si no lo es, entonces se iterará en la preparación de datos.

 Modelado: Si el *dataset* es adecuado, entonces el *data scientist* hará una selección de la técnica de modelado. En caso de la *clustering* se aplicará alguna de las analizadas en el estado del arte (DBSCAN, TI\_DBSCAN, OPTICS, entre otras). En el caso de pronóstico será *logit*, árboles de decisión (y sus derivados), *support vector machine*, entre otros. Luego diseñará la evaluación, construirá el modelo y lo evaluará.

 Evaluación: Se genera una evaluación de los resultados de los modelos anteriores, se revisa el proceso y se establecen los pasos a seguir.

 ¿Precisión es suficiente?: si la precisión del modelo no es suficiente entonces, el proceso debe ser realizado nuevamente, desde el Entendimiento del negocio en adelante.

 Instalar en producción: si la precisión del modelo es suficiente, entonces el modelo puede ser instalado en producción. Esta tarea contempla la planificación del despliegue, de la monitorización, mantenimiento y la revisión del proyecto.

 Evento de finalización: el proceso termina una vez que el modelo ha sido instalado en producción. Sin embargo el proceso vuelve a comenzar debido al objetivo de mejoramiento continuo de modelos.

#### 4.3.4 Actualización de bd producción

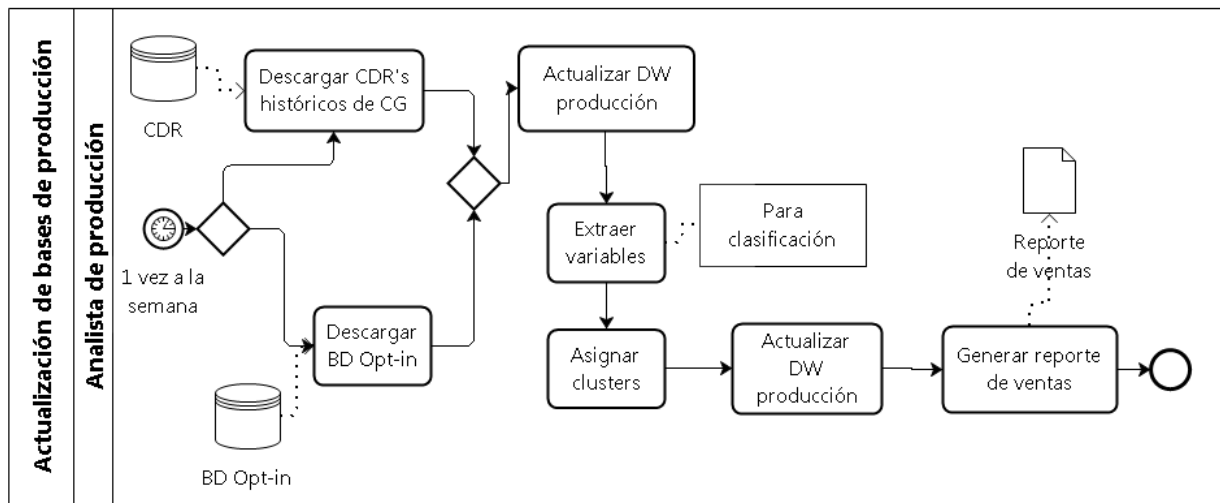



Figura 28. Proceso de Actualización de bases de producción. En notación BPMN

 1 vez a la semana: el proceso se inicializa todas las semanas.

Descargar CDR's históricos de CG: el analista de producción descargará todos los CDR's que contienen información de los Centros de Gravedad de interés a la fecha.

Descargar BD Opt-in: el analista de producción descargará la base de datos de Opt-in actualizada a la fecha

Actualizar DW producción: actualizará el *DW* de producción en sus variables de estado de suscripción según BD Opt-in.

Extraer variables: extraerá las variables necesarias para la clasificación de las personas de la base de datos dentro de algún *cluster* de interés

Asignar clusters: asignará los identificadores a algún *cluster* según los valores de sus variables descriptoras

Actualizar DW producción: actualizará el *DW* de producción, con la nueva asignación de clusters

Generar reporte de ventas: actualizará el reporte de ventas para cada Centro de Gravedad y publicará la información en el sistema interno de la empresa.

### 4.3.5 Opt-in

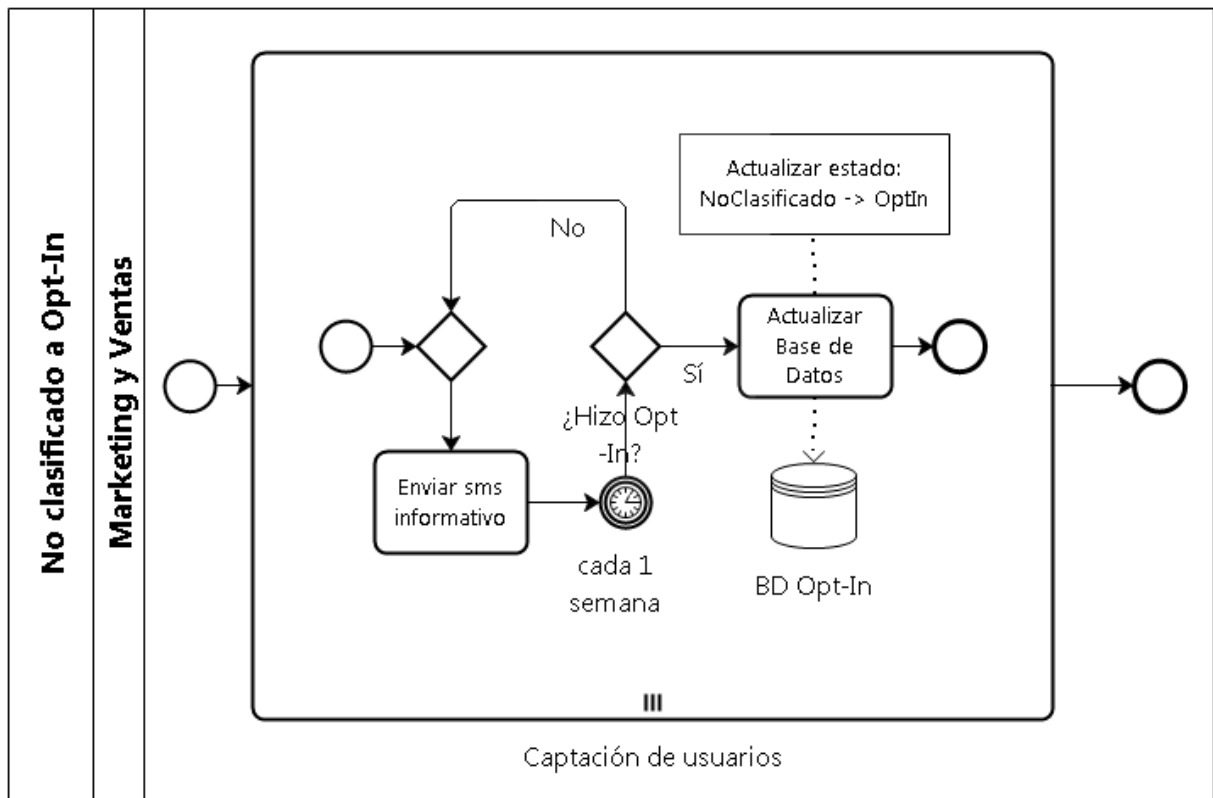


Figura 29. Proceso de *Opt-in*. En notación BPMN

● Evento de inicialización: este proceso se inicia con el inicio de campañas de obtención de usuarios.

□ Captación de usuarios: este subproceso se ejecuta de forma paralela para todos los potenciales usuarios del servicio que se encuentran inicialmente en un estado “No clasificado”. Este subproceso solo termina cuando todos los clientes con línea activa del aliado estratégico han pasado desde el estado “No clasificado” al estado “*Opt-in*”

● Evento de finalización: cuando se han captado todos los usuarios, según las bases de datos del aliado estratégico, termina el proceso, hasta que haya renovación de estas bases.

### 4.3.6 Captación de clientes

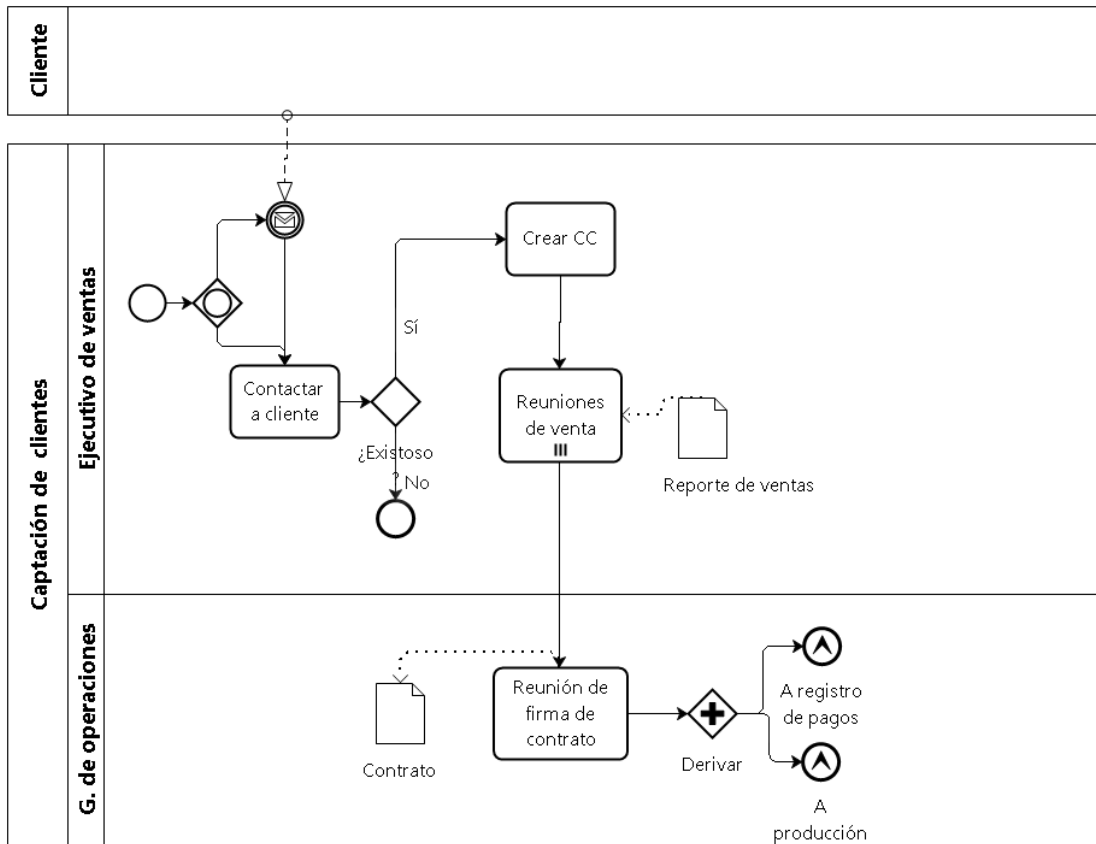


Figura 30. Proceso de Captación de clientes. En notación BPMN

Evento de inicialización: el proceso comienza cuando se presenta la necesidad de contacto a un potencial cliente ya sea por iniciativa del ejecutivo de ventas o luego de la solicitud de un cliente.

Contactar a cliente: el contacto con el cliente se realiza de forma virtual, ya sea telefónicamente o por medio de correos electrónicos. En este contacto se tiene como objetivo agendar una reunión con el cliente.

¿Exitoso?: en la compuerta se evalúa si el contacto a cumplido o no con su objetivo. En caso de no ser exitoso, se termina el proceso

Crear CC: cuando el contacto es exitoso se crea un Centro de Costo (CC), donde se asocia un correlativo al cliente/campaña. Este será usado tanto en los registros de producción como en los registros contables de la empresa.

☐ Reuniones de venta: se procede a una serie de reuniones de venta en las que se negocian los criterios de la relación entre el cliente y la empresa

☐ Reunión de firma de contrato: una vez establecidos los criterios se realiza una reunión final de firma de contrato donde participa el gerente de operaciones de la empresa.

⬢ Derivar: el contrato y campaña se derivan a los procesos de registro de pagos y producción

### 4.3.7 Postventa

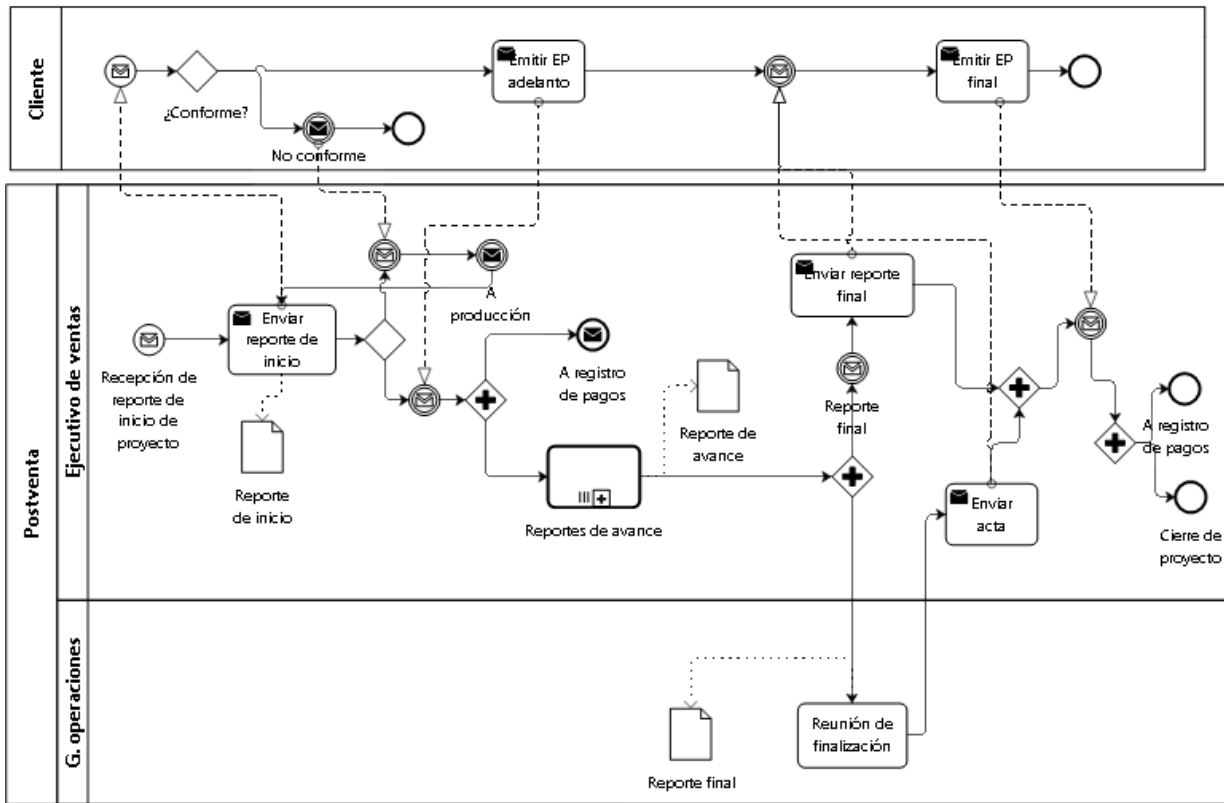




Figura 31. Proceso de Postventa. En notación BPMN


☑ Recepción de reporte de inicio de proyecto: el proceso se inicia cuando el ejecutivo de ventas recibe el Reporte de inicio de campaña desde el proceso de Producción.

☑ Enviar reporte de inicio: el ejecutivo de ventas envía el Reporte de inicio de campaña al cliente, como documento que indica la necesidad de la generación de un estado de pago por parte de cliente.


 A producción: cuando el cliente indica no estar conforme con lo estipulado en el reporte de inicio de campaña, por inconcordancias con lo estipulado en el contrato, el ejecutivo de ventas debe derivar el caso a producción para solucionar el problema. Luego reenviará el Reporte de inicio de campaña


 Recibir EP: si el cliente está conforme, entonces generará dos flujos:


 Reportes de avance: ejecución del subproceso de reportes de avance, donde se envían, aprueba y reciben estados de pago con la misma lógica descrita anteriormente (ver 10.2.2)

 A registro de pagos: se envía el estado de pago al proceso de registro de pagos.


 Luego de ejecutada la campaña se realizan dos tareas:


 Enviar reporte final: se envía al cliente el reporte final de desempeño de la campaña.

 Reunión de finalización: se genera una reunión de cierre de contrato, entre el gerente de operaciones, ejecutivo de ventas y cliente.

 Enviar acta: el ejecutivo de ventas envía el acta de la reunión para dar paso a la emisión del último estado de pago.

 Cuando se recibe el último estado de pago se ejecutan dos tareas:

 A registro de pagos: se envía el estado de pago al proceso de registro de pagos

 Cierre de proyecto: se realiza el cierre del proyecto



### 4.3.8 Producción

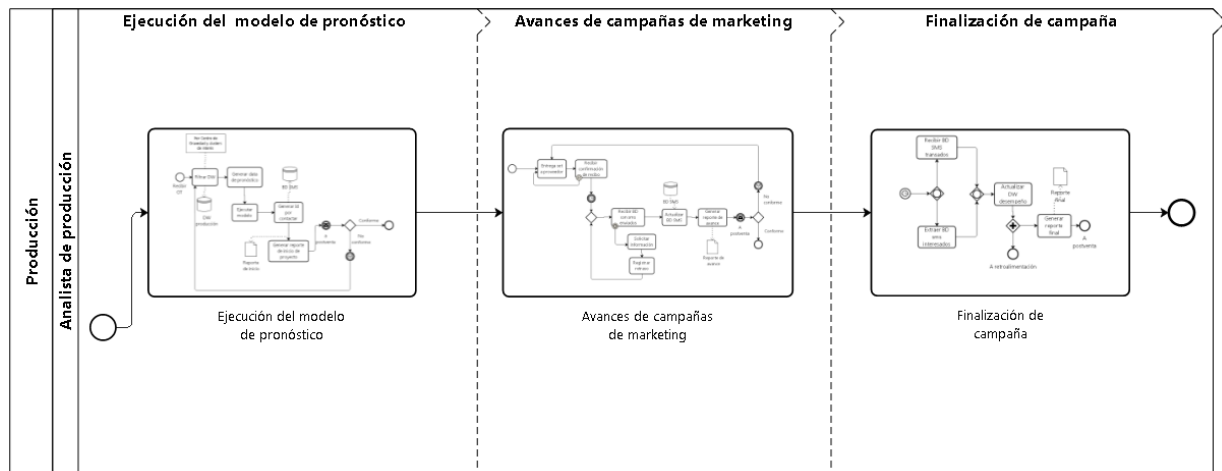


Figura 32 Proceso de Producción. En notación BPMN

● Evento de inicialización: el proceso de producción se inicia cada vez que se recibe un proyecto desde Captación de clientes.

El proceso se subdivide en 3 etapas:

⊕ Ejecución del modelo de pronóstico: etapa en la que se ejecutan los modelos de pronóstico y se entrega un set de personas por contactar y sus fechas de contacto.

⊕ Avances de campañas de marketing: etapa en la que se envían los mensajes planificados y se registra la respuesta a estos mensajes.

⊕ Finalización de campaña: etapa de análisis de campañas finalizadas.

## Ejecución del modelo de pronóstico

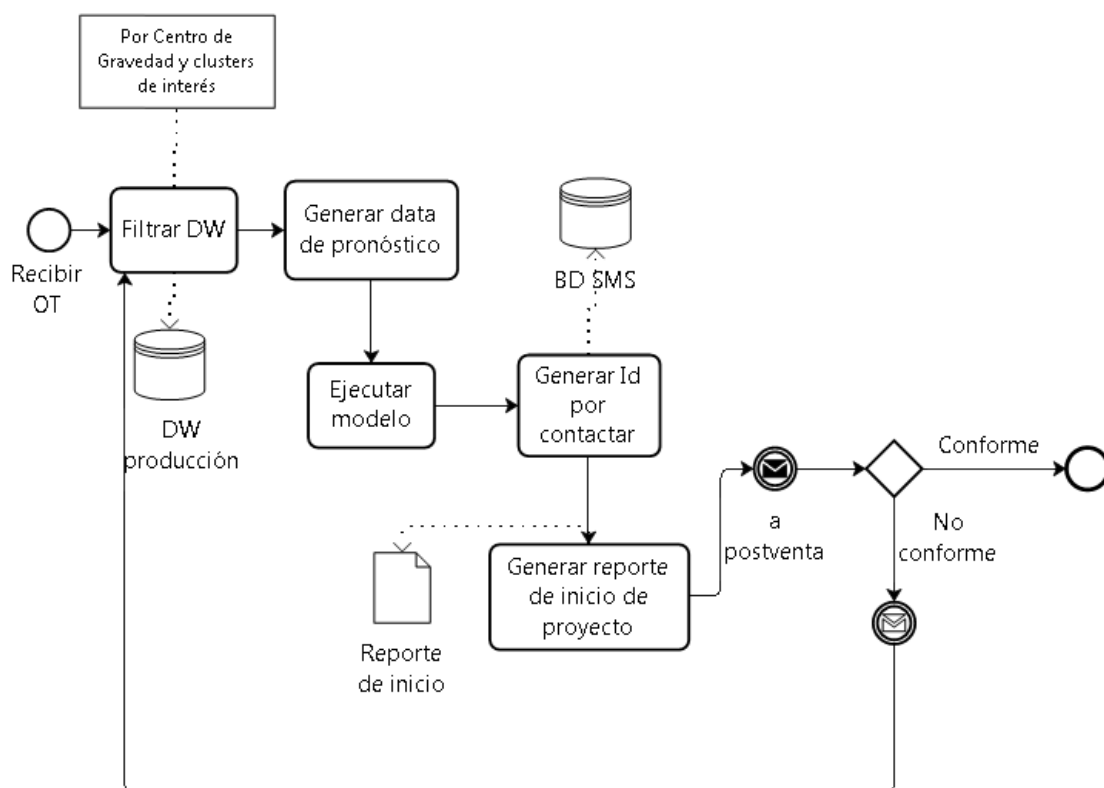


Figura 33. Proceso de Ejecución del modelo de pronóstico. En notación BPMN

- Recibir OT: la primera etapa del proceso se inicia al recibir una OT de trabajo desde el proceso de Captación de clientes
- Filtrar DW: el *DW* de producción será filtrado por los parámetros de campaña: *clusters* de interés, fechas de campaña, Centros de Gravedad de interés.
- Generar data de pronóstico: con los datos filtrados se generará la data a utilizar en la ejecución del modelo de pronóstico.
- Ejecutar modelo: el analista de producción ejecutará el último modelo entregado por el analista de BI.
- Generar Id por contactar: se generará un set de usuarios por contactar y fechas de contacto.

□ Generar reporte de inicio de proyecto: con los resultados del modelo se producirá el Reporte de inicio de proyecto.

✉ A Postventa: el Reporte de inicio será entregado al proceso de Postventa

✉ No conforme: cuando se recibe la notificación de no conformidad del cliente, por falta de mensajes a enviar, el proceso volverá a la tarea de filtrar DW para completar el set de usuarios por contactar.

En otro caso el proceso finaliza.

### Avances de campañas de marketing

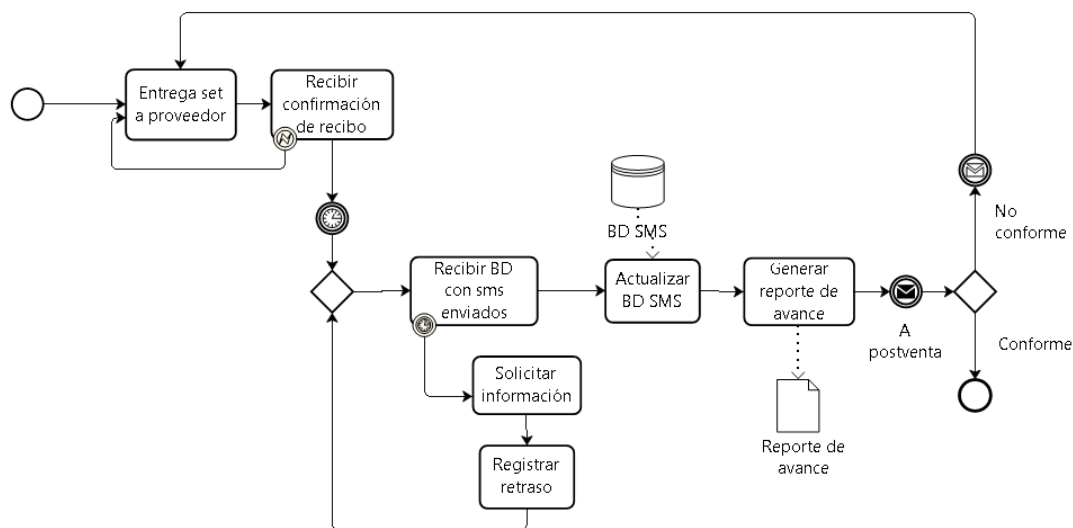



Figura 34. Proceso de Avances de campañas de marketing. En notación BPMN

● Evento de inicialización: el proceso se inicia una vez ejecutado el modelo de pronóstico y generado el set de usuarios por contactar.


□ Entrega set a proveedor: se entrega el set de usuarios por contactar y sus fechas de contacto al proveedor.

□ Recibir confirmación de recibo: se recibe la confirmación del recibo del listado de usuarios por contactar, por parte del proveedor.

✉ Si el proveedor no ha recepcionado el listado de usuarios por contactar, debe ser reenviado.

 El proceso espera hasta la ejecución del envío de mensajes por parte de proveedor para continuar.

Recibir BD con sms enviados: se recepciona, desde el proveedor, la lista de mensajes enviados y sus fechas de envío.

 Cuando se cumple el plazo máximo de entrega de resultados de envío de mensajes, estipulado con el proveedor, se desencadenan las siguientes dos tareas.


Solicitar información: solicitar la información de mensajes enviados al proveedor.

Registrar retraso: el retraso en la entrega del listado de mensajes enviados debe ser registrado para ajustar las relaciones con el proveedor.

Actualizar BD SMS: con los datos de mensajes enviados, se actualiza la base de datos de SMS del Centro de Costos, agregando las variables de mensajes enviados y fechas de envío.

Generar reporte de avance: se genera el Reporte de avance con los mensajes enviados, en contraste con los planificados a enviar, y sus fechas de envío.

 A postventa: se envía el reporte de avance al proceso de Postventa.

 Si se recibe la notificación de que el cliente no está conforme con los mensajes enviados, el proceso se reiniciará en la tarea de entregar un set de usuarios por contactar al proveedor.

En otro caso, el proceso finaliza.

## Finalización de campaña

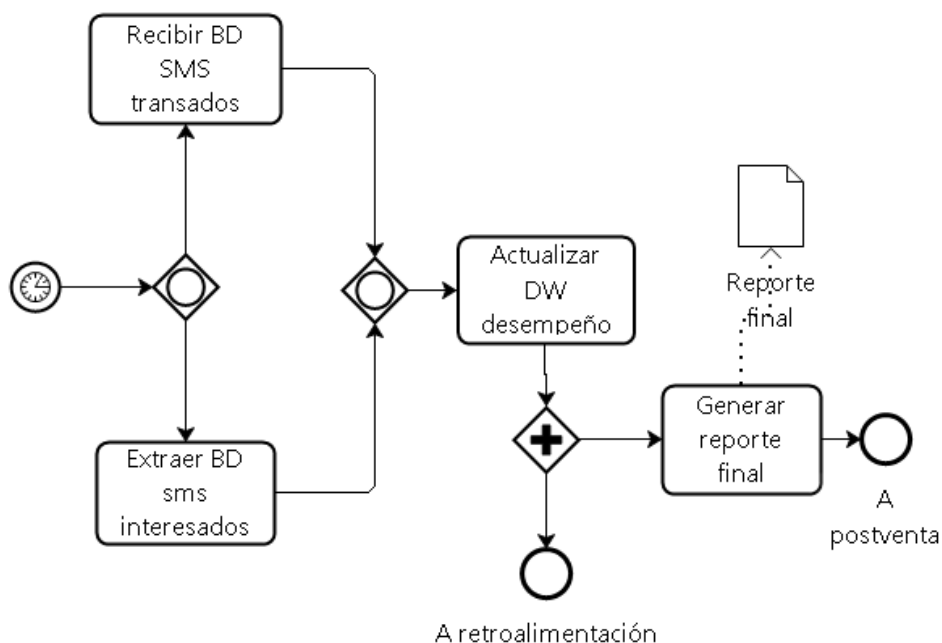






Figura 35. Proceso de Finalización de campaña. En notación BPMN

 El proceso se inicia, luego de enviados todos los mensajes de la campaña y transcurrido el tiempo de espera para que el usuario transe todos los beneficios recibidos.

 Recibir BD SMS transados: desde el Sistema de Cliente (ver Sistema de clientes (SC): ) se obtienen los códigos transados asociados al Centro de Costos. Esta tarea se ejecuta sólo si el cliente tiene la capacidad de registrar los beneficios transados.

 Extraer BD sms interesados: desde el Sistema de Información de Usuarios (ver Sistema de información de usuarios (SIU):), se extraen los datos de declaraciones de interés por beneficios enviados a usuarios.

 Actualizar DW desempeño: con la información anterior se actualiza el DW de desempeño, agregando la información de mensajes transados y declaraciones de interés.

Luego se ejecutan dos flujos:

 A retroalimentación: se envían el DW de desempeño actualizado al proceso de retroalimentación.

Generar reporte final: se produce el reporte final con el desempeño de la campaña de marketing móvil.

A postventa: los resultados son enviados al proceso de Postventa.

## CAPÍTULO 5: Diseño de Lógica de Negocios

---

En la siguiente sección se describen los resultados encontrados tras la aplicación de diferentes técnicas de minería de datos a información de telefonía para el perfilamiento de cliente según su comportamiento de visitas (días en los que se registra la concurrencia del usuario a un Centro de Gravedad). El análisis mencionado responde a la necesidad de identificar tipos de usuarios que, dadas sus características, recibirán beneficios en distintos periodos de tiempo y con connotaciones asociadas a ingresos per cápita de las comunas de residencia de los usuarios. Esta segmentación será refinada con la información de *Opt-in* que permitirá aumentar la personalización de las campañas.

Posteriormente se realiza una comparación de modelos de predicción y clasificación para obtener un pronóstico de visitas, donde la variable a pronosticar es la asistencia de un usuario a un Centro de Gravedad en una fecha particular. Lo anterior se realiza con el objetivo de lograr la precisión temporal en el set de mensajes a enviar de una empresa de campañas publicitarias móviles, personalizadas y geolocalizadas.

### 5.1 Entendimiento de los datos

#### Comportamiento de Llamadas

Para el siguiente análisis exploratorio se trabaja con la herramienta *Excel*, sobre una base de datos con 175.000 datos (para un total de 18 semanas) pertenecientes al CG que contiene al *Mall Plaza Vespucio*, la que contiene los siguientes atributos:

Tabla 38. Especificación de atributos de base de datos inicial.

<b>antena</b>	: Código de la antena al cual se conecta el teléfono emisor de la llamada
<b>persld</b>	: Id emisor de la llamada
<b>antena_entrada</b>	: Código de la antena al cual se conecta el teléfono receptor de la llamada
<b>receptld</b>	: Id receptor de la llamada
<b>fecha</b>	: Fecha de la llamada (aaaa-mm-dd)
<b>hora</b>	: Hora de inicio de la llamada (horario Chile)
<b>duración</b>	: Duración de la llamada en segundos

Para el posterior análisis se eliminan las variables *antena\_entrada* y *receptld*, y se crean las variables “día”, “mes” e “intervalo\_horario”, y se calculan las llamadas por mes. Se observa un comportamiento similar para los meses de mayo, junio y julio, con un promedio de llamadas cercano a las 48.000 llamadas, donde mayo es el mes con

mayor cantidad de registros, esto puede deberse al comportamiento de compra de los clientes de la empresa de telefonía por el día de la madre. Por otro lado, el mes de agosto registra 31.148 llamadas, lo cual se debe a que se registraron eventos de menos semanas de este mes.

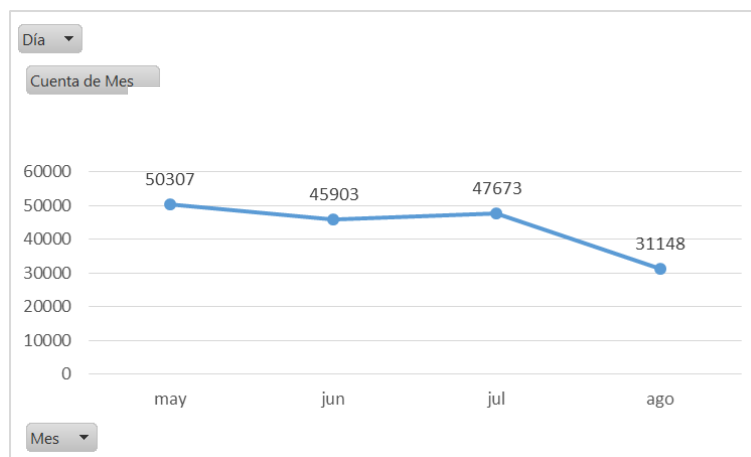


Figura 36. Frecuencia de llamadas por mes

Se nota un comportamiento diferente entre las llamadas en días de semana (lunes, a viernes) y las llamadas en días de fin de semana (sábado y domingo). Este comportamiento varía según meses, lo que se atribuye a estacionalidades y posibles errores de registros (especialmente en el mes de agosto).

Se identifica un comportamiento creciente de visitas para días de semanas y decreciente para días de fin de semana en los meses de mayo, junio y agosto. Julio posee una tendencia similar, con la diferencia de una baja de llamadas en día viernes.

Al entrar en detalle, se observa que el 99% de las personas realizan menos de 20 llamadas por semana, comportamiento que se mantiene para todos los meses analizados.



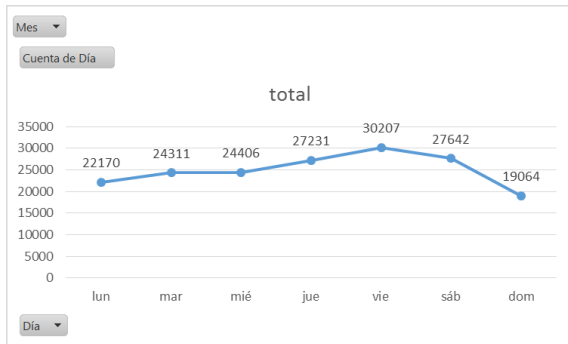


Figura 37. Frecuencia de llamadas por día. Mayo

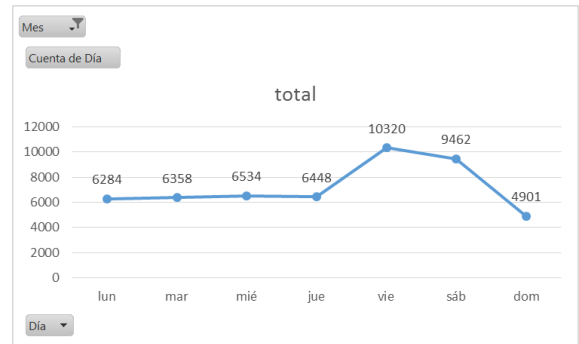


Figura 38. Frecuencia de llamadas por día. Junio

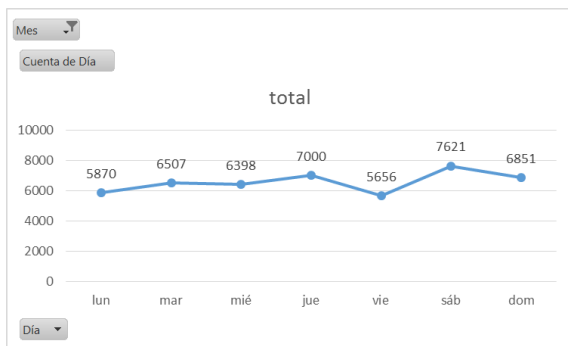


Figura 39. Frecuencia de llamadas por día. Julio

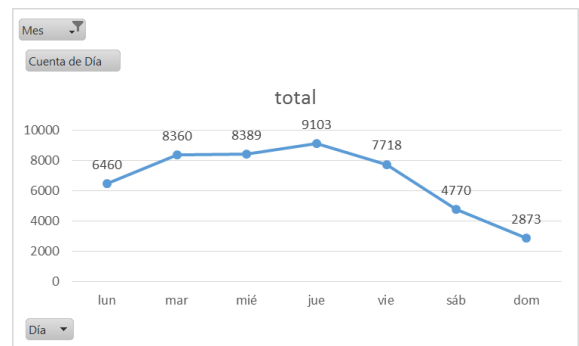


Figura 40. Frecuencia de llamadas por día. Agosto

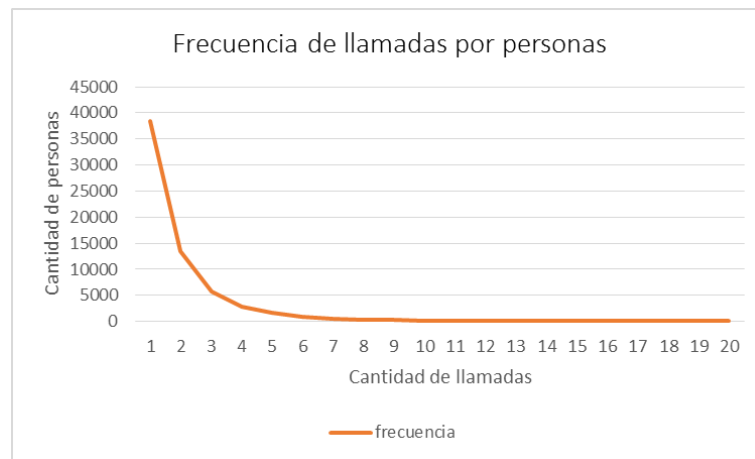


Figura 41. Frecuencia de llamadas por persona. Periodo mayo - agosto.

Luego la base original es cargada en el sistema R y las variables “dia”, “mes”, “intervalo\_hora” y semana son creadas, estas captan el día de la semana, mes, intervalo de hora y número de semana de cada llamada registrada. Además se transforman los datos en una nueva tabla que guarda la frecuencia de llamadas para cada día y cada número.

antena	persId	fecha	hora	duracion	dia	mes	intervalo_hora	semana
MPVSFU1	94314108	2014-05-02	17:31:10	61	vie	may	17	17
MPVSFU1	61464245	2014-05-02	17:32:20	142	vie	may	17	17
MPVSFU1	61422485	2014-05-02	17:33:34	2	vie	may	17	17
MPVSFU1	81238021	2014-05-02	17:33:32	12	vie	may	17	17
MPVSFU1	93355814	2014-05-02	11:59:22	12	vie	may	11	17
MPVSFU1	97630234	2014-05-02	17:33:35	26	vie	may	17	17
MPVSFU1	88611582	2014-05-02	11:55:52	144	vie	may	11	17
MPVSFU1	53392922	2014-05-02	17:34:45	1	vie	may	17	17
MPVSFU1	97331027	2014-05-02	17:38:10	12	vie	may	17	17
MPVSFU1	54726179	2014-05-02	17:38:12	5	vie	may	17	17

Figura 42. Muestra de tabla de llamadas

Son eliminados los días que registren una cantidad de llamadas fuera del intervalo de confianza al 95%, es decir, se elimina el registro del día  $d$  sí y solo sí:

$$\begin{aligned}
 \text{cuenta de llamadas}_d &< \text{promedio llamadas diarias} - 1,96 * \text{desviación estándar de llamadas diarias} \\
 &= 804,8735
 \end{aligned}$$

Lo cual elimina los días 01-05-2014 (jueves), 13-07-2014 (viernes), 17-08-2014 (domingo) y 18-08-2014 (lunes).

Se analiza el comportamiento de las llamadas por intervalo horario, concluyendo que la frecuencia de llamadas aumenta a las 9:00 horas con un *peak* de llamadas alrededor de las 12:00 horas, se produce un valle alrededor de las 15:00 horas, para aumentar hasta su segundo *peak* cerca de las 19:00 horas y disminuye drásticamente hasta las 22:00 horas. Lo anterior tiene sentido si se considera que la antena se encuentra en un centro con horario de atención de 10:00 a 21:30 horas, todos los días de la semana.

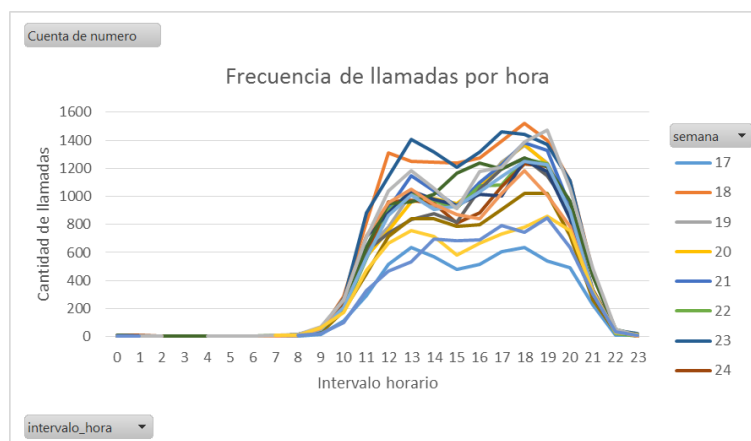


Figura 43. Frecuencia de llamadas por hora.

Se da formato a los datos de modo de obtener una tabla que registre la frecuencia de llamadas para cada número, a esta se le aplica un análisis de correlación entre las llamadas por día, donde se observa que existe una correlación fuerte entre la frecuencia de llamadas en días de semana, una correlación moderada entre frecuencias de llamada de días semana y fin de semana y correlación similar entre llamadas en días de fin de semana.

Con lo anterior se concluye con alta probabilidad una persona que llama en día de semana llamará el resto de los días de semana, mientras que esta probabilidad es menor en el caso de las interacciones con las llamadas en días de fin de semana.

	lun	mar	mie	jue	vie	sab	dom
lun	1.0000000						
mar	0.6507376	1.0000000					
mie	0.6511629	0.6415978	1.0000000				
jue	0.6476515	0.6643301	0.6989832	1.0000000			
vie	0.6342263	0.6695401	0.6741114	0.7054780	1.0000000		
sab	0.4723422	0.5085198	0.5226272	0.5433116	0.5505631	1.0000000	
dom	0.4359082	0.4601620	0.4845927	0.4723451	0.4553042	0.5429852	1.0000000

Tabla 39. Correlación de llamadas por día.

### Comportamiento de visitas

Bajo el supuesto de que el comportamiento de visitas a la zona cubierta por la antena está altamente correlacionado con las llamadas realizadas en la posición geográfica de la antena se genera una tabla de frecuencia de visitas que se forma a partir de la regla

$$\text{visita} = \begin{cases} 1; & \text{Si frecuencia de llamadas} > 0 \\ 0; & \text{Si frecuencia de llamadas} = 0 \end{cases}$$

Luego se analiza la frecuencia de visitas por intervalo horario, comprobando que el comportamiento, a pesar de verse suavizado, es análogo al comportamiento de frecuencia de llamadas por intervalo horario.

Además la conducta de visitas en general indica que el 99% de las personas visita el área cubierta por la antena menos de 13 veces en el total del periodo, es decir, en promedio visitan menos de 1 vez al mes la zona. El 1% restante es el grupo de personas con visitas regulares y que, por tanto, son interesantes como objeto de estudio.

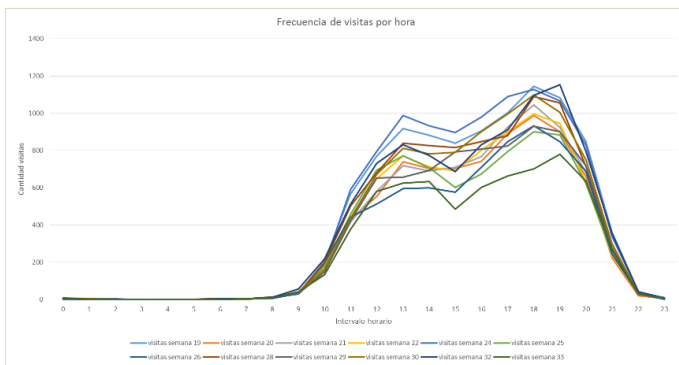


Figura 44. Frecuencia de visitas por hora.

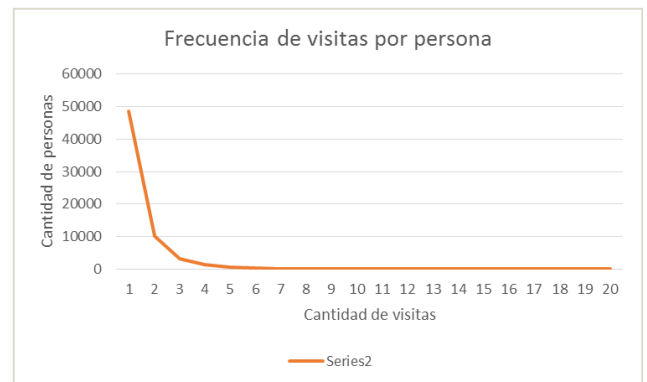


Figura 45. Frecuencia de visitas por persona, general.

Por otro lado, se aplica el análisis de correlación entre visitas, con el cual se concluye que el comportamiento de visitas entre semana es similar al comportamiento entre las llamadas registradas en el sector.

En este sentido, se aprecia mayor correlación entre dos días de semana y menor entre un día de semana y otro de fin de semana. Este resultado entrega niveles de correlación menor al estudiado anteriormente, sin embargo, se considera una mejor representación de los datos para el objetivo de negocios que motivó el trabajo realizado.

	vlun	vmar	vmie	vjue	vvie	vsab	vdom
vlun	1.0000000						
vmar	0.4812392	1.0000000					
vmie	0.4662228	0.4906340	1.0000000				
vjue	0.4720474	0.5013541	0.5184096	1.0000000			
vvie	0.4521317	0.4708944	0.4773248	0.4965622	1.0000000		
vsab	0.3325256	0.3479974	0.3521628	0.3652231	0.3578289	1.0000000	
vdom	0.2648753	0.2674592	0.2886704	0.2795954	0.2668351	0.3456961	1.0000000

Tabla 40. Correlación de visitas por día.

Finalmente, y como método de exploración, se generan dos muestras de datos, en base a reglas que permiten trabajar con el porcentaje de personas que tienen mayor cantidad de visitas, y por ende, mantienen mayor historia para generar futuras predicciones (que corresponde al 1% a 3% del total de número únicos):

- Muestra A:
  - Responde a la unión de datos que cumplen las condiciones: vistas en semana > 9 o visitas en fin de semana > 4
  - Este muestreo entrega un set de 947 observaciones.
- Muestra B:
  - Responde la intersección de datos que cumplen las condiciones: vistas en semana > 6 y visitas en fin de semana > 4
  - Este muestreo entrega un set de 411 observaciones.

En el análisis posterior, se comprueba la diferencia de información entregada al aplicar un algoritmo de *clustering* en una u otra muestra.

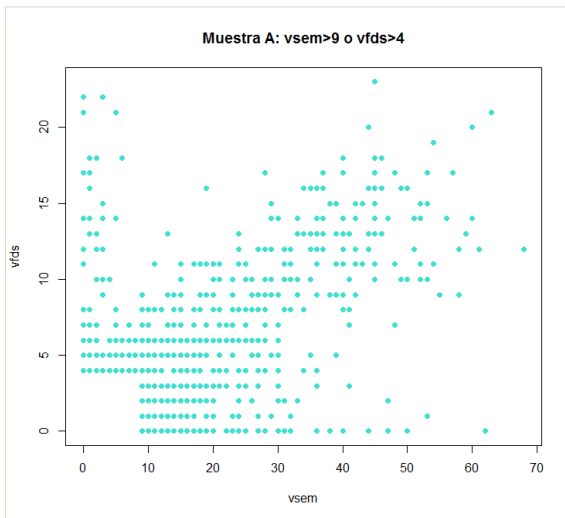


Figura 46. Muestra A

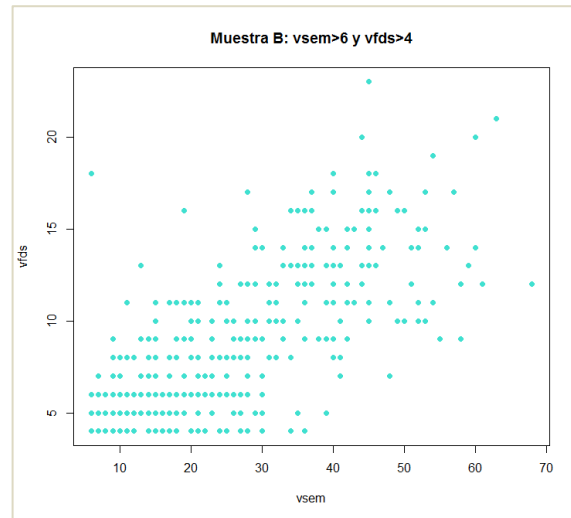


Figura 47. Muestra B

## 5.2 Preparación de los datos: Generación de un *Data Warehouse*

La generación del *Data Warehouse* se realiza a partir de la creación de diferentes servicios responsables de la producción de variables necesarias para los análisis posteriores. Estas variables son agregadas a la base de datos utilizando como llave primaria la combinación {persId (n) , fecha (f)} que, para cada visita, inscribe el id n y la fecha f del registro observado.

persId	fecha	casatrabajo	dia	mes	semana	semanaMes	objetivo	comunaAm1	comunaAm2	comunaPm1	comunaPm2
89455301	2014-06-27	casa	vie	jun	25	4	1	La Florida	La Florida	La Florida	Las Condes
85210683	2014-08-28	casa	jue	ago	34	4	1	Santiago	La Florida	La Florida	Santiago
97516553	2014-06-23	casa	lun	jun	25	4	1	Quinta Normal	La Florida	La Florida	Quinta Normal
97009054	2014-08-15	casa	vie	ago	32	3	1	sin información	sin información	sin información	sin información
97003490	2014-07-28	casa	lun	jul	30	4	1	La Florida	Santiago	La Florida	Santiago
93324546	2014-08-09	trabajo	sab	ago	31	2	1	La Florida	Quinta Normal	La Florida	Lo Prado
74790080	2014-05-31	casa	sab	may	21	5	1	La Florida	Renca	La Florida	Quinta Normal
95440931	2014-07-29	trabajo	mar	jul	30	5	1	Maipú	La Florida	Maipú	La Florida
68228441	2014-07-11	casa	vie	jul	27	2	1	Puente Alto	La Florida	La Florida	Puente Alto
95450764	2014-07-04	trabajo	vie	jul	26	1	1	La Florida	Santiago	La Florida	Macul

Figura 48 . Muestra de base de datos DW\_RegistroLlamadas

Se cargan los datos originales de cada CDR, se llama al servicio que elimina variables no utilizadas en este trabajo, crea las variables “dia”, “mes”, “semana” y “semanaMes” y transforma la base de llamadas a visitas. Este es el primer paso de una serie de tareas que generan una base final con 947 identificadores únicos, 82.059 observaciones y 55 variables, con una proporción de visitas y no visitas de 23% contra 77%. La base de

datos se estructura como datos de panel, donde las llaves primarias son el identificador del usuario y la fecha, habiendo o no registro de visita en esta (ver Tabla 41).

Tabla 41. Estructura del *Data Warehouse* creado

Identificador del usuario	fecha	Variable descriptora 1	...	Variable descriptora m
1	Fecha 1		...	
1	Fecha2		...	
2	Fecha 1		...	
2	Fecha 2		...	
...	...	...	...	...
n	Fecha1		...	
n	Fecha2		...	

Las variables generadas pueden ser descriptoras de la persona (“persId”), o bien de la fecha. A continuación, en la Tabla 42, se describen las variables generadas para el *Data Warehouse*.

Tabla 42. Descripción de las variables añadidas a *Data Warehouse*

Variable	Tipo	Descripción
<b>1. Llaves primarias y objetivo</b>		
"persId"	Numeric	Llave primaria que indica el ID del emisor de un visitante al Centro de Gravedad.
"fecha"	Date	Llave primaria que muestra la fecha en la que un visitante fue o no a un Centro de Gravedad.
"objetivo"	Factor	Esta es la variable objetivo para modelos de pronóstico. Indica 1 si “persId” n, visita en “fecha” f el Centro de Gravedad, y 0 en otro caso.

<b>2. Descriptores de "persld"</b>		
"casatrabajo"	Factor	Variable que categoriza a "persld" en las categorías "casa" o "trabajo" según el horario en el cual se han registrado llamadas en la antena ubicada en el Centro de Gravedad. <ul style="list-style-type: none"> <li>• "casa" si <math>21 \leq hora \leq 8</math></li> <li>• "trabajo" si <math>9 \leq hora \leq 20</math></li> </ul>
"comunaAm1"	Factor	Indica la comuna con mayor cantidad de llamadas hechas por cada usuario en "persld" en horario AM (si $21 \leq hora \leq 8$ )
"comunaAm2"	Factor	Indica la segunda comuna con mayor cantidad de llamadas en horario AM (si $21 \leq hora \leq 8$ )
"comunaPm1"	Factor	Indica la comuna con mayor cantidad de llamadas en horario PM (si $9 \leq hora \leq 20$ )
"comunaPm2"	Factor	Indica la segunda comuna con mayor cantidad de llamadas en horario PM (si $9 \leq hora \leq 20$ )
"cantComunaAm1", "cantComunaAm2", "cantComunaPm1", "cantComunaPm2",	Numeric	Contabiliza la cantidad de veces que se registra una llamada desde "comunaAm1", "comunaAm2", "comunaPm1" y "comunaPm2" respectivamente.
"cantidadComunasAM", "cantidadComunasPM"	Numeric	Cuenta la cantidad de comunas distintas desde donde se registran llamadas en horario AM y PM respectivamente.
"contieneLaFlorida"	Factor	La variable es 1 si la variable "comunaAm1" o "comunaAm2" contienen "La Florida" (comuna que contiene la antena en estudio).



"contieneLaFloridaPrimeraOpcion"	Factor	La variable es 1 si la variable "comunaAm1" contiene "La Florida" (comuna que contiene la antena en estudio).
"tipoVisita"	Factor	<p>Categoriza el tipo de visita en las categorías "Casa", "Visita" o "Trabajo", según la siguiente regla:</p> <ul style="list-style-type: none"> <li>• "Casa" si <math>"contieneLaFloridaPrimeraOpcion" = 1</math></li> <li>• "Trabajo" si <math>"contieneLaFloridaPrimeraOpcion" = 0</math> y <math>comunaPm2 = "La Florida"</math></li> <li>• "Visita" en cualquier otro caso</li> </ul>
"num_llamadas"	Numeric	Cantidad de llamadas que hizo "persId" n en "fecha" f.
"frecuencia"	Numeric	Frecuencia de visitas de "persId" n al Centro de Gravedad. Se calcula como el total de visitas contabilizadas en la fecha f sobre el total de días transcurridos del periodo evaluado.
"r_mayor_f"	Factor	<p>Indica si "persId" n lleva más tiempo que su frecuencia de visitas sin visitar el Centro de Gravedad. Se calcula según la siguiente regla:</p> <ul style="list-style-type: none"> <li>• 1 si <math>"dias_ultima_visita" &gt; "frecuencia"</math></li> <li>• 0 si <math>"dias_ultima_visita" &lt; "frecuencia"</math></li> </ul>
<b>3. Descriptores de "fecha"</b>		
"dia"	Factor	Indica el día de la semana de la "fecha" f.
"mes"	Factor	Indica el mes de la "fecha" f.

"semana"	Factor	Indica el número de semana del año de la variable "fecha".
"semanaMes"	Factor	Indica el número de semana del mes de la variable "fecha".
"lun", "mar", "mie", "jue", "vie", "sab", "dom"	Factor	Captura el día de la semana de la variable "fecha", mediante variables <i>dummy</i> .
"vfechaMenosUno", "vfechaMenosDos", "vfechaMenosTres", "vfechaMenosSiete"	Factor	Captura las visitas del número n uno, dos, tres y siete día antes de la "fecha" f.
"vsemanaMenosUno", "vsemanaMenosDos", "vsemanaMenosTres", "vsemanaMenosCuatro"	Factor	Captura las visitas del número n uno, dos, tres y cuatro semanas antes de la "fecha" f.
"minima_hora"	Character	Indica la hora del registro de llamadas más temprana de la "fecha" f para "persId" n.
"maxima_hora"	Character	Indica la hora del registro de llamadas más tardía de la "fecha" f para "persId" n.
"delta_horas"	Numeric	Calcula la diferencia entre "maxima_hora" y "minima_hora", con ello muestra un posible intervalo de tiempo en el que se encontró "persId" n en el Centro de Gravedad en la "fecha" f.
"vlun", "vmar", "vmie", "vjue", "vvie", "vsab", "vdom"	Factor	Calcula la cantidad de veces que "persId" n visitó el CG para cada día, en la "fecha" f.
"fecha_ultima_visita"	Date	Muestra la fecha de la última visita de "persId" n registrada, con respecto a la "fecha" f.

"dias_ultima_visita"	Numeric	Indica la cantidad de días que lleva "persld" n sin visitar el Centro de Gravedad.
"vtotal"	Numeric	Calcula la cantidad acumulada de veces que "persld" n visitó el CG hasta la "fecha" f.
"dias_atraso_persona"	Numeric	Indica hace cuántos días "persld" n debió haber visitado el Centro de Gravedad según el siguiente cálculo: <ul style="list-style-type: none"> <li>• 0 si "<math>fecha_{ultima\_visita} + frecuencia \geq fecha</math>" f</li> <li>• "<math>fecha - fecha\_ultima\_visita + frecuencia</math>", si no</li> </ul>
"vsem", "vfds"	Factor	Calcula la cantidad de veces que "persld" n visitó el CG para cada días de semana y fin de semana en la "fecha" f, respectivamente.

Cabe destacar que las variables asociadas a la comuna, se generan a partir de la exploración de la ubicación de las antenas asociadas a todas las llamadas realizadas por los 947 identificadores que visitan el Centro de Gravedad estudiado.

Para la realización del *clustering* se consideran los identificadores únicos de "persld" y las variables "vlun", "vmar", "vmie", "vjue", "vvie", vsab, "vdom", "vsem" y "vfds" de la última fecha del horizonte de tiempo registrado.

### 5.3 Clustering

El *clustering* se utiliza para tipificar los visitantes de un Centro de Gravedad. En este contexto es que se analizan dos técnicas de *clustering*: 1. *DBSCAN*; y 2. *K-means*.

#### 5.3.1 DBSCAN

Para este modelo, y como medida exploratoria, se trabaja con las variables "vsem" y "vfds". Esto pues, ha quedado comprobado que el comportamiento de visitas es diferenciado entre estas variables.

Debido a que encontrar los parámetros óptimos para el algoritmo DBSCAN es una tarea compleja que depende de la naturaleza de los datos, se trabaja con un tercer algoritmo, OPTICS, el cual, entrega un ordenamiento de los puntos de una base a su distancia de alcance (ver 2.2.3).

Como OPTICS y DBSCAN son algoritmos que no trabajan con centroides se genera un indicador para medir la compactibilidad de los puntos dentro de un *cluster* y separación entre *clusters*, a partir del índice de *Dunn* y las definiciones de distancia de alcance.

El índice de *Dunn* es un valor que permite encontrar la configuración de *clusters* que maximiza la distancia entre *clusters* y minimiza la distancia intra *clusters*, esta se encuentra a partir de una medida de similitud entre *clusters* y el diámetro intra *clusters*.

Similitud entre *clusters*  $c_i, c_j$ : 
$$Sim(c_i, c_j) = \min_{x \in c_i, y \in c_j} ||x - y||$$

Diámetro intra *cluster*  $c$ : 
$$diam(c) = \max_{x, y \in c} ||x - y||$$

Luego,

Índice de *Dunn*: 
$$DU_k = \min_{i=1, \dots, k} \left\{ \min_{j=i+1, \dots, k} \left( \frac{Sim(c_i, c_j)}{\max_{m=1, \dots, k} diam(c_m)} \right) \right\}$$

De lo anterior, se considera que, un punto entre *clusters* (que no pertenece a ningún *cluster*) es un punto de ruido, luego la distancia entre *clusters* (o similitud entre *clusters*) se homologa a la distancia de los puntos de ruido, mientras que el diámetro intra *clusters* corresponde a la distancia de alcance entre puntos de un mismo *cluster*, lo cual permite generar el índice “valoresDU”:

$$valoresDU = \frac{\min(distancia_{cluster=0})}{\max(\max(distancia_i)_{i = 1, \dots, k})}$$

Este índice muestra mejores configuraciones para valores altos, donde la distancia mínima de un punto entre *clusters* (considerado como *ruido* por el modelo) es alto y la distancia máxima entre los puntos de cada *cluster* es baja. Sin embargo, este índice no trabaja por sí solo, por lo que debe ser complementado con otra información como la cantidad de *clusters* generados (“*num\_clusters*”), el porcentaje de *ruido* de la configuración y los valores de los parámetros buscados (“*MinPts*” y “*EPS\_cl*”).

Luego se itera el cálculo de OPTICS para varios valores de “*MinPts*” y “*EPS\_cl*” y se guardan los valores de “valoresDU”, “*MinPts*”, “*EPS\_cl*”, “*ruido*” y “*num\_clusters*” para cada iteración, generando una base para cada muestra (A y B) de alrededor de 16.000 y 19.800 observaciones respectivamente.

Tabla 43. Ejemplos de resultados para muestra A

“valoresDU”	“MinPts”	“EPS_cl”	<b><i>ruido</i></b>	<b>“<i>num_clusters</i>”</b>
1.30384048	3	3.6070465	0.00428571	3
1	10	0.00368819	0.76142857	12

1.11803399	10	2.00268532	0.11571429	1
0.4472136	2	0.00368819	0.25571429	124

La Tabla 43 muestra algunos resultados encontrados para la Muestra A, en ellos se observan configuraciones poco factibles debido a los valores de sus parámetros: “EPS\_cl” cercano a cero, baja o alta cantidad de *clusters*, altos niveles de *ruido*, valores menores a 0,8 para el indicador “valoresDU”, entre otros.

### Evaluación de los resultados

Se aplica un análisis relacional entre los resultados generados por las iteraciones del algoritmo OPTICS, tanto para la muestra A como para la muestra B, los resultados muestran que:

- A medida que aumenta el radio “EPS\_cl”, disminuye el porcentaje de *ruido*. En particular, para la Muestra A, el nivel de *ruido* se minimiza para radios mayores a 3. Sin embargo estas configuraciones entregaron resultados con pocos puntos mínimos por *cluster* (“MinPts”). Análogamente, para la Muestra B el nivel de *ruido* se minimiza para radios mayores a 3,125, entregando configuraciones con el mínimo de *clusters*.
- Se observa un efecto de disminución de “MinPts” junto con el *ruido* (ver Figura 50, Figura 127). Sin embargo, esta disminución se produce en base a valores de “EPS\_cl”, es decir, se produce una disminución según tramos de “EPS\_cl” (ver Figura 51, Figura 128). Para la Muestra A el efecto se estabiliza en valores de “EPS\_cl” superiores a 3.6.
- El indicador “valoresDU”, luego de un filtro, entrega valores entre 0.85 y 1.17, para la Muestra A y 0.87 a 1.17, para la muestra B. Se nota que los máximos están relacionados con baja cantidad de *clusters*, altos valores de “EPS\_cl” y por ende bajos valores de “MinPts” (ver Figura 53, Figura 129).
- Para la Muestra A se obtuvo que, para “EPS\_cl” $\geq$ 4 y “MinPts”={2,3}, se maximiza “valoresDU” y minimiza “num\_ *clusters*”, dadas las observaciones anteriores se espera menor *ruido* para estos valores. Por otro lado, para “EPS\_cl” $\leq$ 3y “MinPts”=3, se observan valores de “valoresDU” cercanos a 0.9 y mayor cantidad de *clusters* (ver Figura 52).
- Para la Muestra B, “valoresDU” se maximiza en “EPS\_cl” $\geq$ 3,15 y “MinPts”={2,3}, al mismo tiempo se minimiza el número de *clusters* de la configuración. Asimismo, para “EPS\_cl” $\leq$ 2y “MinPts”={4,5,6,7}, se distinguieron valores de “valoresDU” cercanos a 1 y mayor cantidad de *clusters* (Figura 129).

Otro grupo interesante, en esta muestra se encontró para  $2 \leq EPS\_cl \leq 2.25$  y “MinPts”=10, donde el comportamiento es similar al grupo anterior.

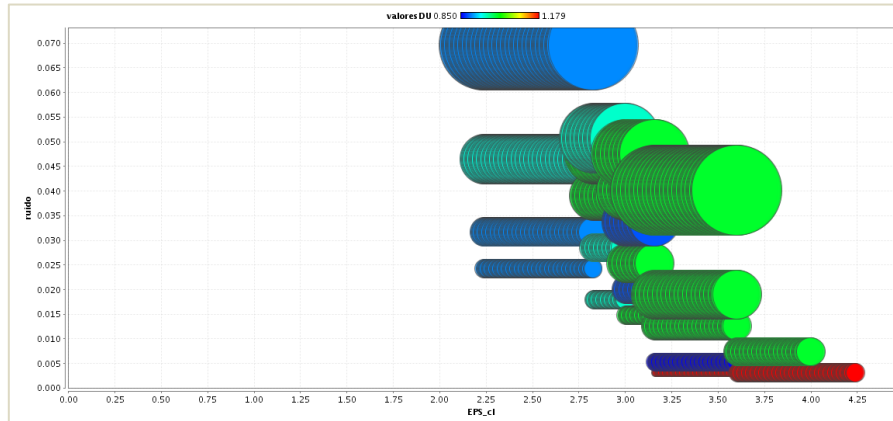


Figura 49. Relación de variables para muestra A. “EPS\_cl” (abscisa), ruido (ordenada), “MinPts” (tamaño), valoresDU” (color).

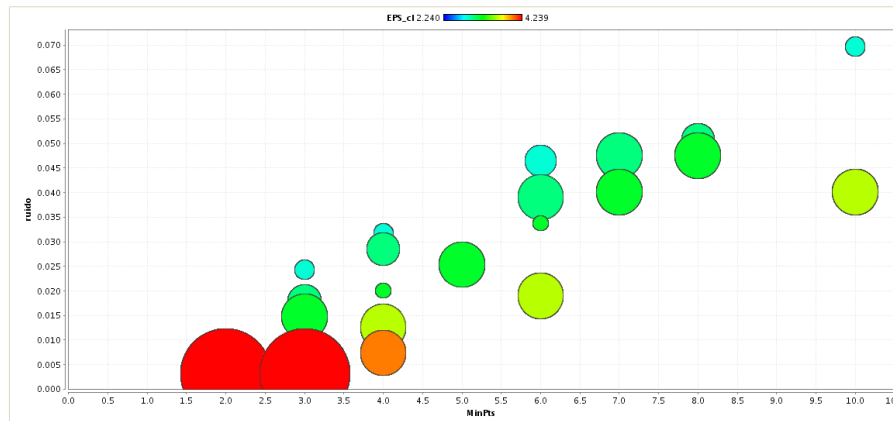


Figura 50. Relación de variables para muestra A. “MinPts” (abscisa), ruido (ordenada), “valoresDU” (tamaño), “EPS\_cl” (color).

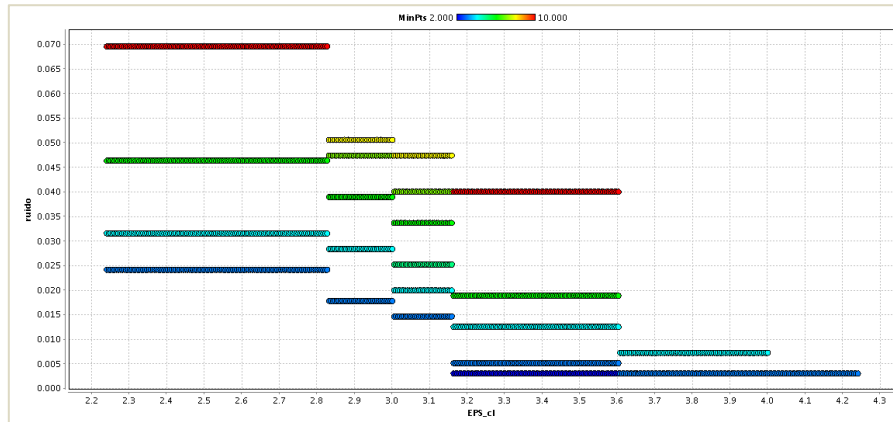


Figura 51. Relación de variable para muestra A. “EPS\_cl” (abscisa), ruido (ordenada), “MinPts” (color).

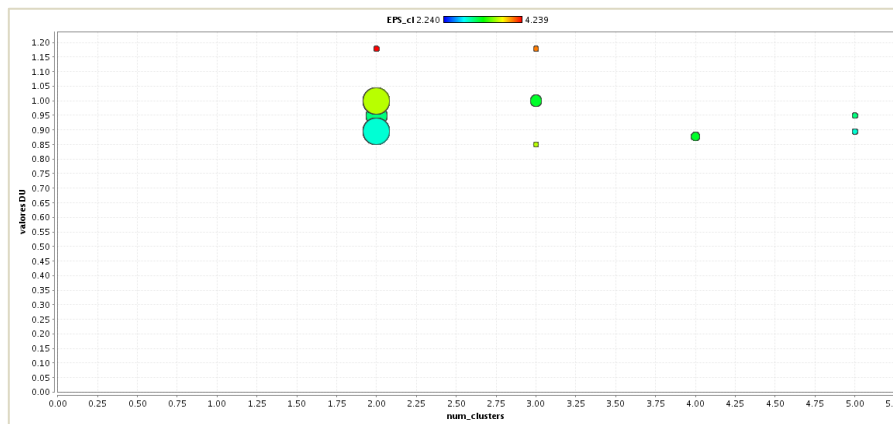


Figura 52. Relación de variables para muestra A. “num\_clusters” (abscisa), “valoresDU” (ordenada), “MinPts” (tamaño), “EPS\_cl” (color).

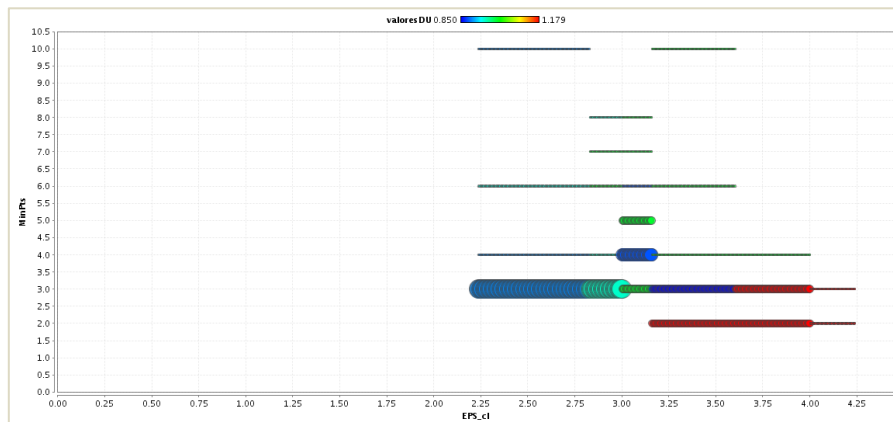


Figura 53. Relación de variables para muestra A. “EPS\_cl” (abscisa), “MinPts” (ordenada), “num\_clusters” (tamaño), “valoresDU” (color).

De lo anterior se encuentran dos grupos para la Muestra A y tres grupos para la Muestra B que podrían entregar buenos resultados de OPTICS, y por ende, para DBSCAN, dados los parámetros observados. Se ejecuta el algoritmo OPTICS de forma iterativa para las siguientes configuraciones:

Tabla 44. Configuraciones finales a probar en Muestra A

<i>Muestra A</i>		
<b>COMPOSICIÓN 1</b>	<b>COMPOSICIÓN 2</b>	<b>COMPOSICIÓN 3</b>
“EPS_cl” = [4, ... , 10] “MinPts” = {2, 3} Iteraciones =1675	“EPS_cl” = [2, ..., 3] “MinPts” = {3} Iteraciones =1000	“EPS_cl” = [3, ..., 4] “MinPts” = {2,3} Iteraciones =2000

Tabla 45. Configuraciones finales a probar en Muestra B.

<i>Muestra B</i>		
<b>COMPOSICIÓN 1</b>	<b>COMPOSICIÓN 2</b>	<b>COMPOSICIÓN 3</b>
“EPS_cl” = [3,14; ..., 9] “MinPts” = {2,3} Iteraciones = 1865	“EPS_cl” = [1,75; ..., 2] “MinPts” = {4, 5, 6, 7} Iteraciones =4000	“EPS_cl” = [2, ..., 2,25] “MinPts” = {10} Iteraciones =1000

Obteniendo las siguientes distribuciones para “valoresDU”. En color se observan las observaciones para las que “valoresDU” es mayor a 0.8



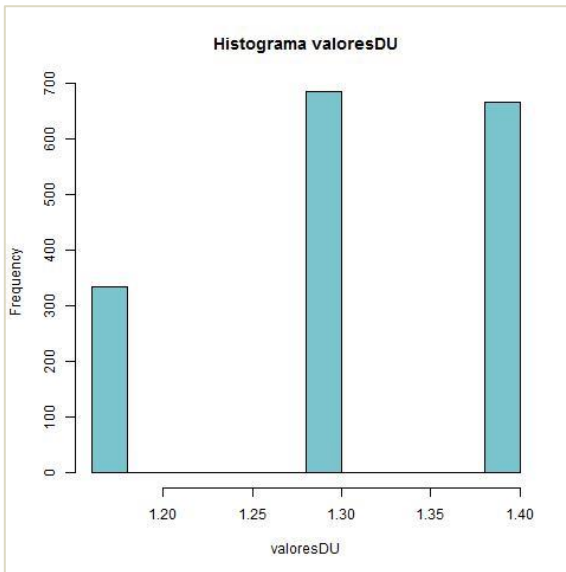


Figura 54. Muestra A, histograma de “valoresDU” para composición 1.

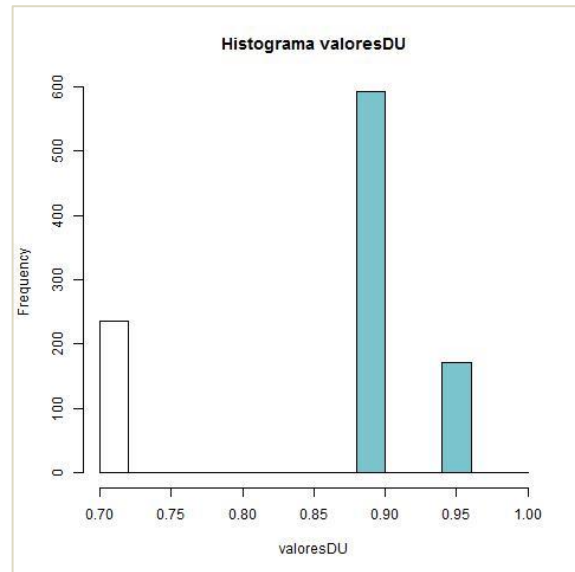


Figura 55. Muestra A, histograma de “valoresDU” para composición 2.

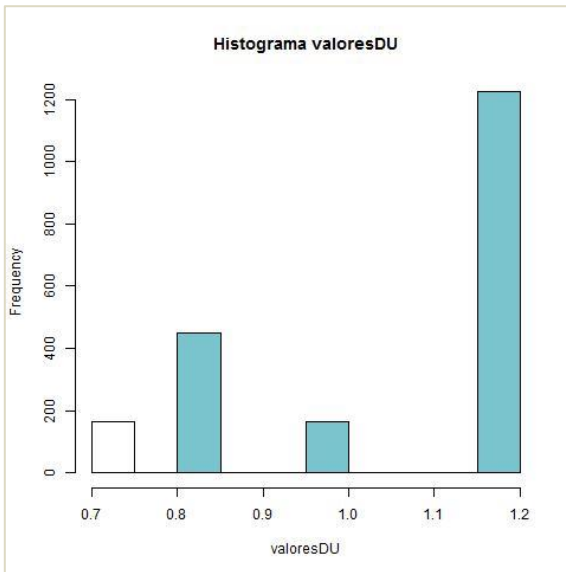


Figura 56. Muestra A, histograma de "valoresDU" para composición 3.

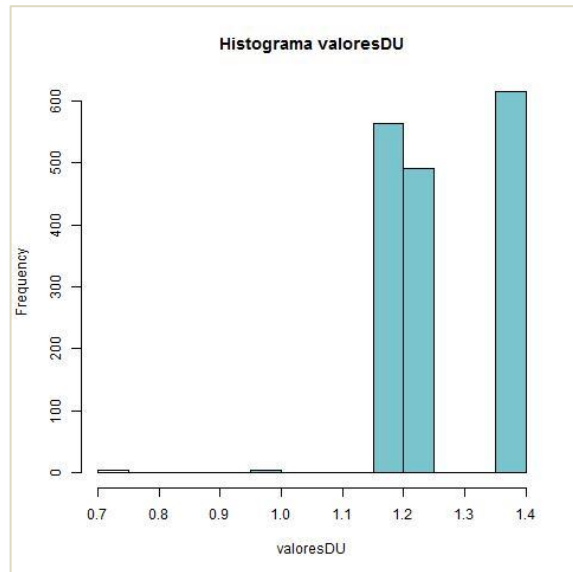


Figura 57. Muestra B, histograma de "valoresDU" para composición 1.

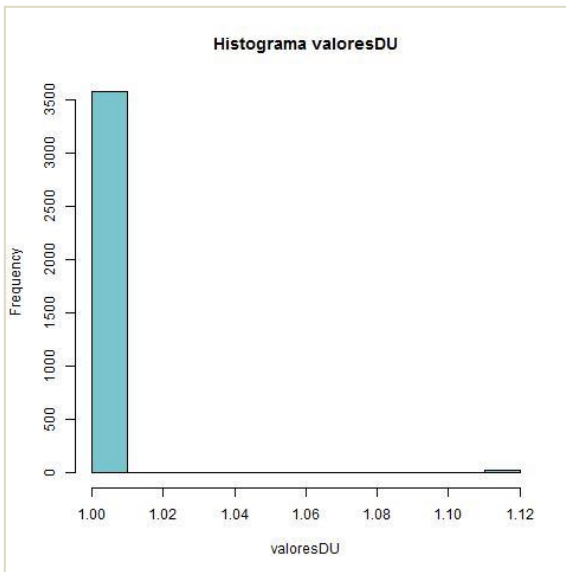


Figura 58. Muestra B, histograma de "valoresDU" para composición 2.

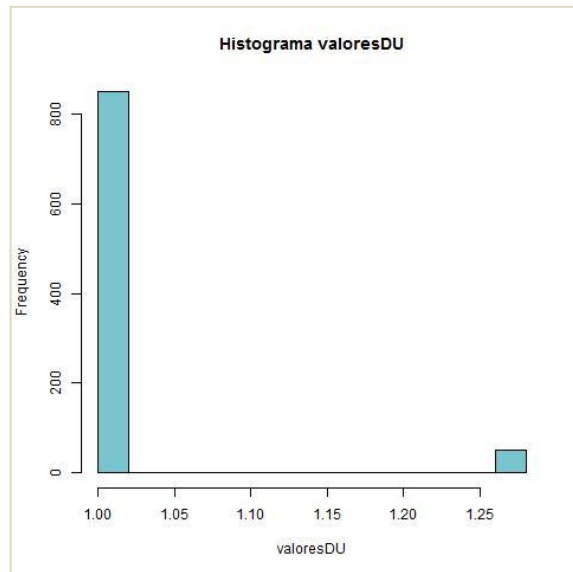


Figura 59. Muestra B, histograma de "valoresDU" para composición 3.

Se calculan los óptimos para cada configuración, con los cuales se analiza la respuesta de OPTICS y DBSCAN.

Tabla 46. Resultado óptimos por composición para muestras A y B.

**MUESTRA A**

**MUESTRA B**

<b>COMPOSICIÓN 1</b>	“MinPts” = 2	“MinPts” = 2
	“EPS_cl” = 5,003	“EPS_cl” = 5,005
	“valoresDU” = 1,4	“valoresDU” = 1,4
<b>COMPOSICIÓN 2</b>	“MinPts” = 3	“MinPts” = 4
	“EPS_cl” = 2,828	“EPS_cl” = 1,5
	“valoresDU” = 0,95	“valoresDU” =
<b>COMPOSICIÓN 3</b>	“MinPts” = 2	“MinPts” = 10
	“EPS_cl” = 3,163	“EPS_cl” = 2
	“valoresDU” = 1,179	“valoresDU” = 1

Luego, los *Reachability Plot* para las mejores configuraciones de las muestras A y B son:

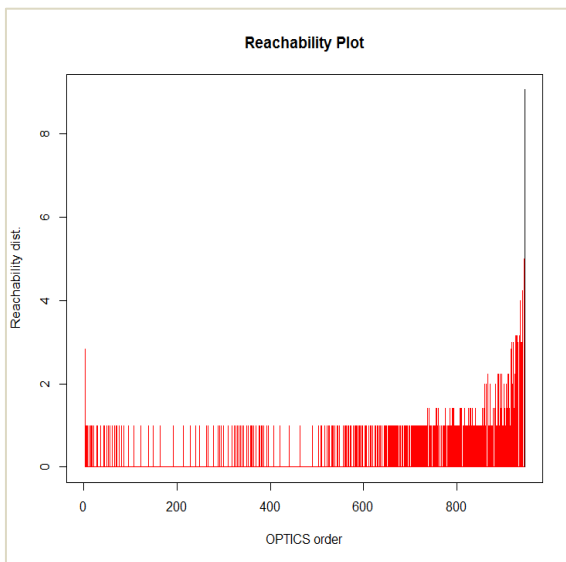


Figura 60. Muestra A, Reachability Plot para OPTICS en composición 1.

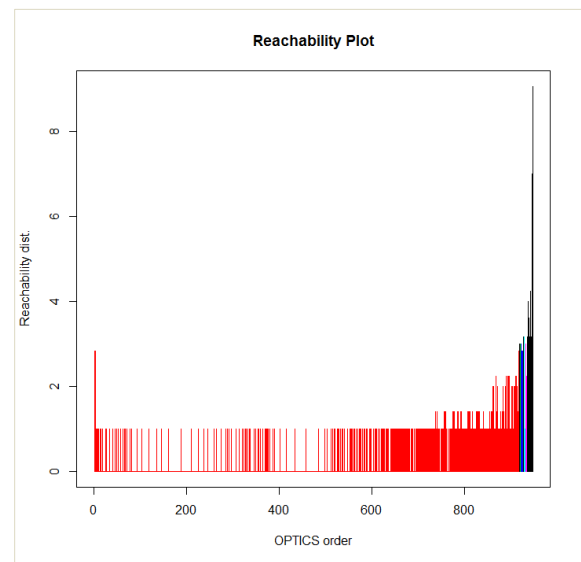


Figura 61. Muestra A, Reachability Plot para OPTICS en composición 2.

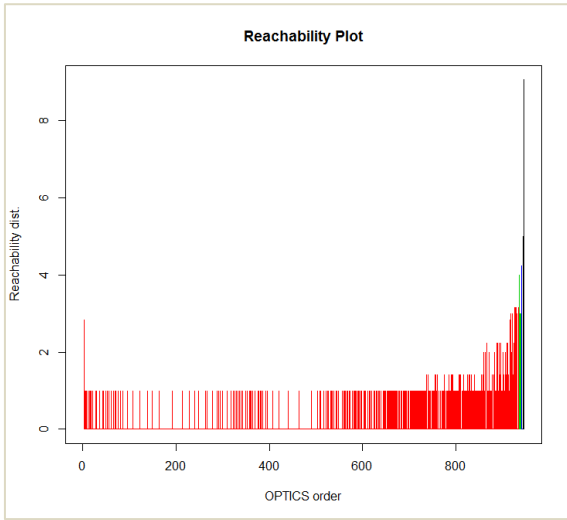


Figura 62. Muestra A, Reachability Plot para OPTICS en composición 3.

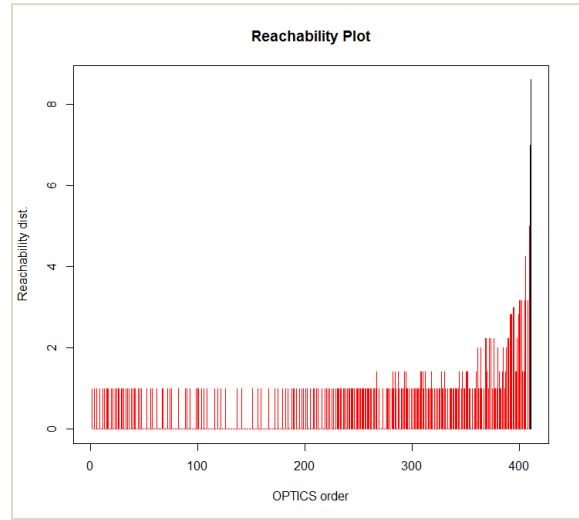


Figura 63. Muestra B, Reachability Plot para OPTICS en composición 1.

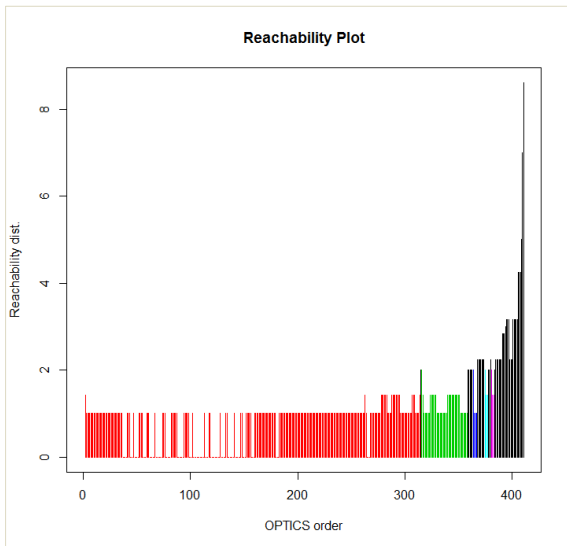


Figura 64. Muestra B, Reachability Plot para OPTICS en composición 2.

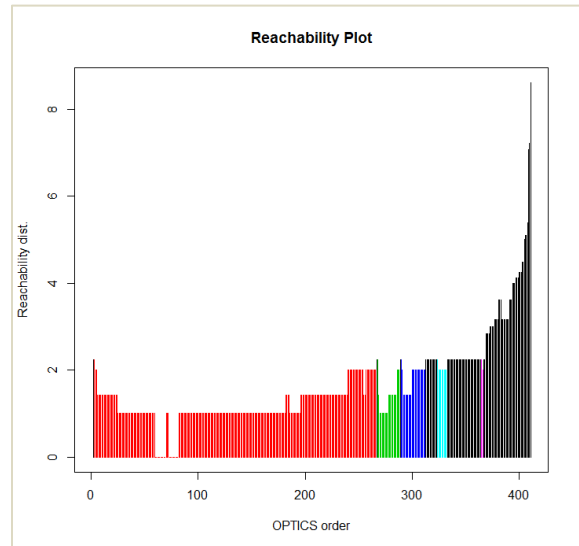


Figura 65. Muestra B, Reachability Plot para OPTICS en composición 3.

Los que indican que existen 1, 5 y 3 *clusters* para las respectivas configuraciones 1,2 y 3 de la Muestra A y 1, 5 y 5 *clusters* para todas las configuraciones de la muestra B. Por otro lado, las configuraciones de la Muestra A entregan resultados con bajo nivel de *ruido*. Este valor aumenta en las configuraciones de la Muestra B.

Finalmente se utiliza el software *RapidMiner* para aplicar el algoritmo DBSCAN con los valores de “EPS\_cl” y “MinPts” encontrados.

Observación: el *cluster* 0 corresponde al conjunto de puntos de *ruido*.

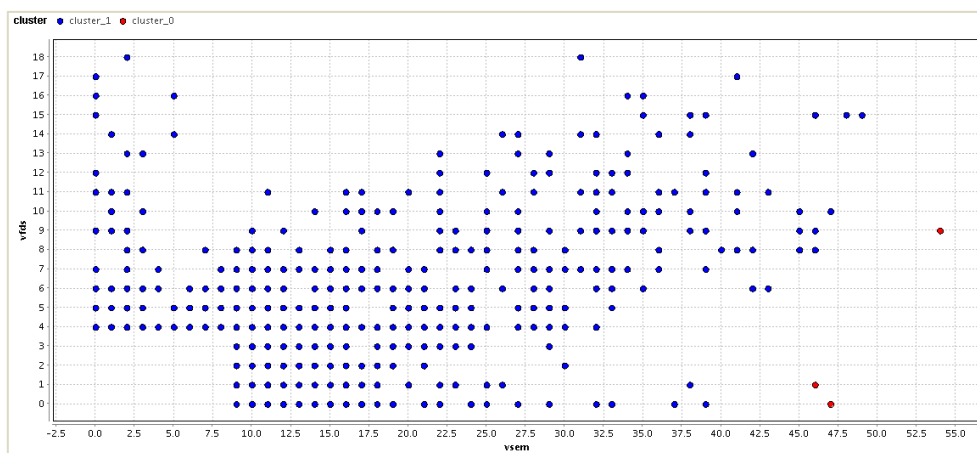


Figura 66. Muestra A, DBSCAN para composición 1.

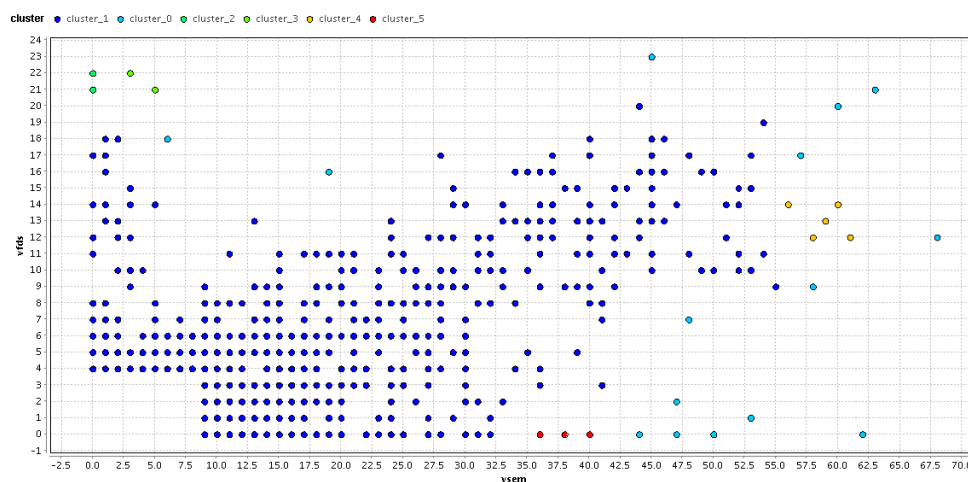


Figura 67. Muestra A, DBSCAN para composición 2.

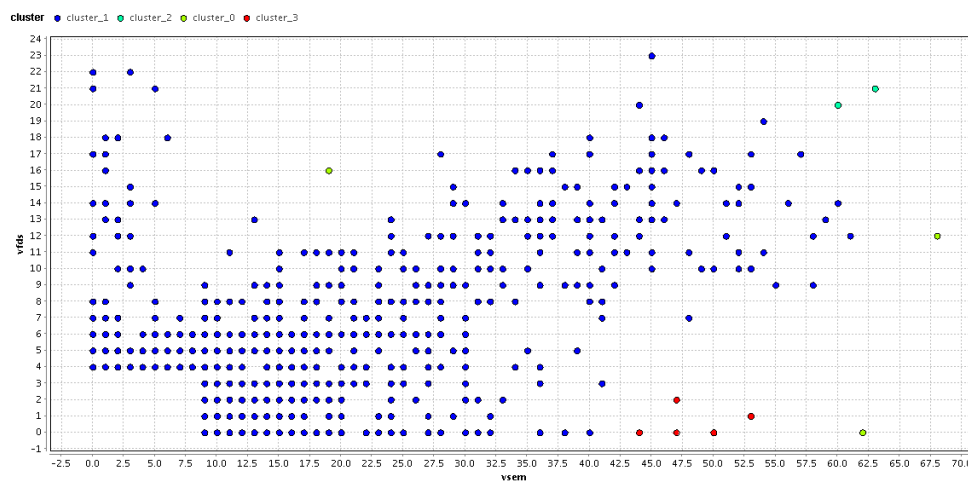


Figura 68. Muestra A, DBSCAN para composición 3.

Las agrupaciones de puntos para todas las configuraciones se diferencian para altos valores de “vsem” y/o “vfds”, sin embargo, estos no representan posibles comportamientos de visitas de personas a la zona cubierta por la antena telefónica.

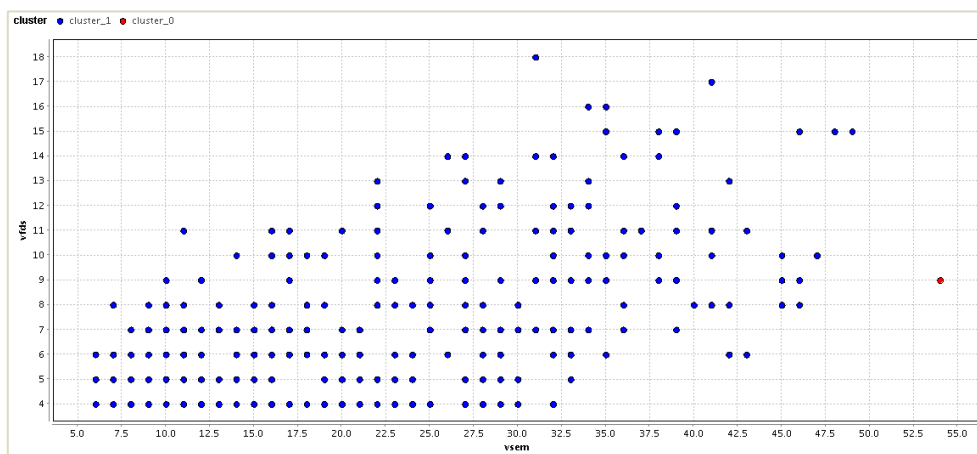


Figura 69. Muestra B, DBSCAN para composición 1.

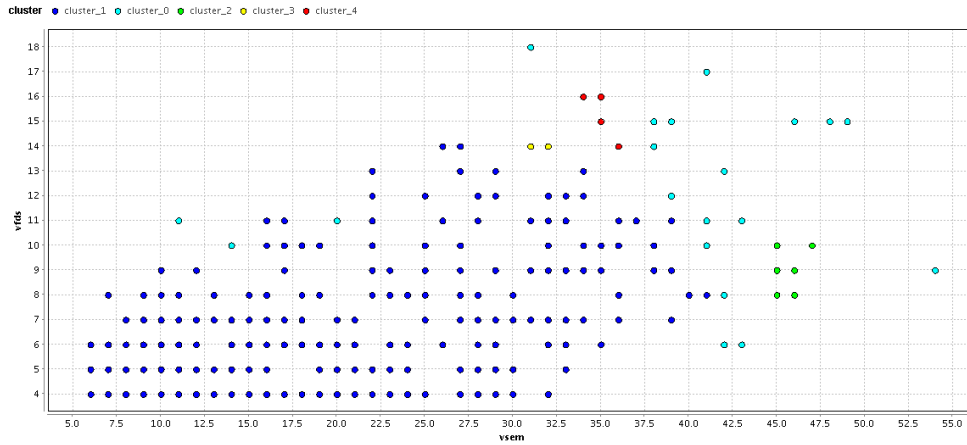


Figura 70. Muestra B, DBSCAN para composición 2.

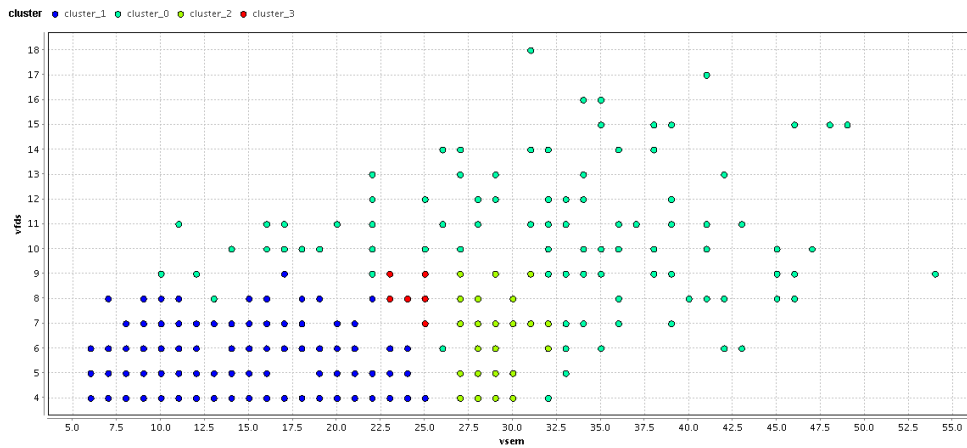


Figura 71. Muestra B, DBSCAN para composición 3.

El resultado de DBSCAN en este caso, no se ajusta a lo esperado por OPTICS en las configuraciones 2 y 3, donde se encuentran 4 y 3 *clusters* respectivamente. El *cluster* diferenciado de la masa de puntos (*cluster 2*), en la configuración 2, se interpreta como personas que tienen alta frecuencia de visitas semanales y media frecuencia de visitas en fin de semana, mientras que el *cluster 4* de la misma configuración se interpreta como personas con frecuencia medio alta en días de semana y alta en días de fin de semana. Por otro lado, la configuración 3, muestra en el *cluster 2* personas con frecuencia medio alta de visitas en día de semana y medio baja en días de fin de semana, se observa en el *cluster 3* un comportamiento similar al del *cluster 2*, donde el nivel de visitas en días de semana es menor y las visitas de fin de semana son acotadas a valores mayores a 7 (es decir, aproximadamente 1 visita al mes en días de fin de semana).

La configuración 2 muestra altos niveles de *ruido*. Sin embargo este disminuye en las configuraciones 3 y 1, en esta última casi no entregan puntos de este tipo.

Dada la forma en que se crea la Muestra B, se pierde el grupo de personas que visitan la zona en días de semana y no lo hacen en días de fin de semana.

### **Conclusiones del modelo**

Los resultados indican, que el algoritmo no es capaz de reconocer *clusters* como el de personas que visitan la zona abarcada por la antena en días de fin de semana, pero no en días de semana y viceversa, grupos que se esperan encontrar cerca de los ejes del gráfico.

Estos *clusters*, debido al problema de negocios, son considerados como requisito mínimo de una buena configuración de *clusters*, por lo que se decidió no utilizar DBSCAN como herramienta generadora de grupos de personas importantes para la cadena de valor del negocio.

En particular, la muestra B genera un subconjunto de datos que elimina el *cluster* de personas que visitan con alta frecuencia en días de fin de semana, y baja frecuencia en días de semana, así como los casos de alta frecuencia de visitas en días de semana y baja en días de fin de semana. Por lo que se califica como una muestra que no entrega información completa a los modelos de *clustering*.

Se presume que lo anterior se debe al tipo de datos con el que se trabaja, al parecer, sería necesario aplicar una transformación o *kernel* para diferenciar grupos, con separaciones de densidad relativamente constante. Sin embargo, la aplicación de estos métodos nos es trivial puesto que se deben definir ciertos parámetros de las transformaciones según los tipos de datos del problema de negocio.

Otra causa podría ser que los datos tengan distintas densidades, con lo cual, DBSCAN no es capaz de lidiar, por ejemplo, en la configuración 2 de la Muestra A, el algoritmo detecta dos *clusters* dados los parámetros  $ESP_{cl} = 3$  y “MinPts” = 3. Sin embargo, para el *cluster* 1, que contiene la mayoría de los puntos, al menos “EPS<sub>cl</sub>” debería disminuir para diferenciar grupos de datos.

Además, como se trata de frecuencias, si se tiene al menos un dato para cada tupla (“vsem”, “vsem”), entonces los parámetros “EPS<sub>cl</sub>” y “MinPts” que distinguen datos dentro del *cluster* 1, entregarían un resultado cercano a 1 dato es a 1 *cluster*, lo cual no es un resultado atractivo para resolver el problema de negocio.

### **5.3.2 K-means**

Dado lo anterior se probó con el algoritmo *K-means* en la muestra A, que soluciona el problema de generar *clusters*. En primera instancia, se realiza una exploración utilizando tanto la distancia euclidiana (ver Figura 72) como similitud coseno (ver Figura



73) como medidas de distancia entre los puntos y sus centroides. Asimismo, se prueba la generación de 3 *clusters*, según el comportamiento en las variables “vsem” y “vdfs”, ya que así se espera encontrar, a lo menos, los *clusters* de alta frecuencia de visitas en semana y baja en fin de semana, alta frecuencia de visitas en fin de semana y baja en semana, y visitas proporcionales y con cambio gradual.

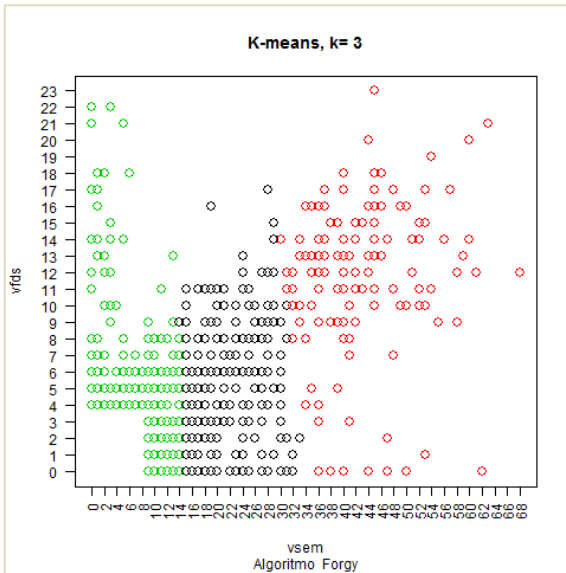


Figura 72. K-means, distancia euclidiana K=3

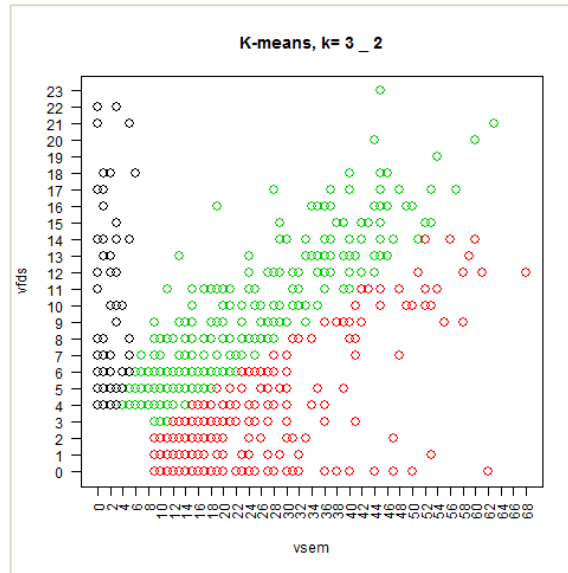


Figura 73. K-means, similitud coseno K=3

Al respecto, se observa que la similitud coseno entrega una mejor representación de los comportamientos en *clusters* para K-means, esto debido a que al aplicar distancia euclidiana el modelo segmenta según el valor de alguna de las variables analizadas, sin distinguir entre días de visitas. Esto se ve claramente en la Figura 72, donde no es factible obtener un segmento de personas caracterizado por bajas visitas en días de fin de semana y visitas regulares en día de semana. Por otro lado el *clustering* con similaridad coseno, captura comportamientos de visitas que se explican como combinaciones lineales entre las variables del modelo y, con ello, obtiene los *clusters* mínimos deseados.

Posteriormente se realizaron pruebas de *clustering*, utilizando similitud coseno. Cabe destacar que se ha trabajado con un paquete del software R capaz de realizar *clustering* con similaridad coseno. Este paquete, llamado “*akmeans*”, usa como fuente la función *norm.sim.ksc* para lograr su objetivo. Esta función requiere de datos normalizados L2 norm =1, lo cual genera lo siguiente:

- Para cada fila, obtiene la raíz cuadrada de la suma del cuadrado de todos los elementos  $x_{ij}$  de la observación, a este valor le llamaremos normalizador.

- Genera cada elemento  $x_{ij}'$  de la nueva matriz dividiendo el elemento original por el normalizador calculado antes. Con esto obtiene valores relativos a los demás atributos de cada observación.
- Aplica el *clustering* sobre el espacio transformado de valores

Debido a esto se realizó una comparación de los segmentos resultantes tanto en su espacio original (denormalizado), como en el espacio transformado. Debido al buen comportamiento inicial de este modelo, el análisis se realizó en datos con distintos niveles de agregación:

- Nivel 1: día de semana: lunes a domingo
- Nivel 2: tipo de día: semana, fin de semana
- Nivel 3: combinación de las dos primeras: tipo de día “semana”, sábado y domingo

Se concluye que con la denormalización de los valores obtenidos el resultado pierde relevancia algebraica. Este problema se ve acentuado en el nivel de agregación 1, donde se observa un resultado difuso tanto para resultados en espacio transformado como no transformado, es decir, los datos se encuentran demasiado desagregados como para generar diferencias significativas en el comportamiento. Lo anterior se demuestra en las Figura 74 y Figura 75, que representan las combinaciones de variables vsab y “vvie”.

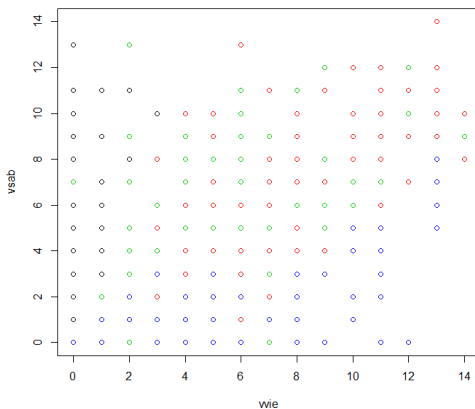


Figura 74 . *Clustering* en espacio original (k=4). Variables vsab, “vvie”

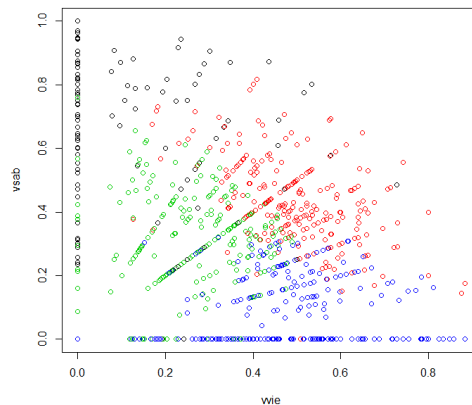


Figura 75. *Clustering* en espacio transformado (k=4). Variables vsab, “vvie”

La baja diferenciación entre *clusters* se mantiene en todos los niveles para resultados en espacios no transformados. Sin embargo, para resultados en espacios

transformados se observan resultados coherentes para los niveles de agregación 2 y 3. Ambos logran distinguir visitas de semana y fin de semana. En particular, el nivel 3 permite desagregar la información de visitas dentro de los días de fin de semana, entregando más información del cluster, por lo que se decide trabajar con las variables “vsem”, *vsab* y “vdom” (que guardan la información de visitas en días sábado y domingo).

Se presentan, desde la Figura 76 a la Figura 81 resultados parciales de *K-means* para espacios transformados y no transformados.

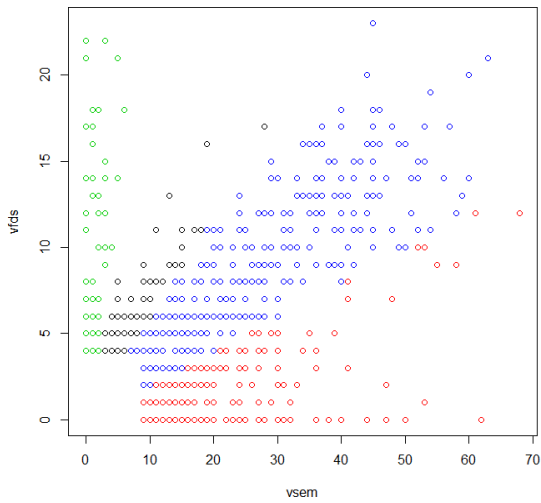


Figura 76. Clustering en espacio original (k=4). Variables "vdfs", "vsem"

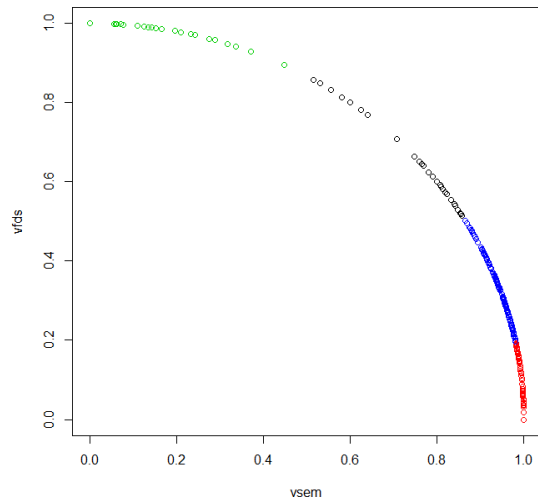


Figura 77. Clustering en espacio transformado (k=4). Variables "vdfs", "vsem"

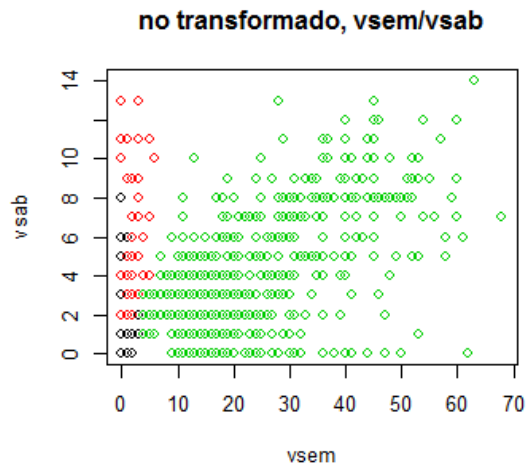


Figura 78. Clustering en espacio original (k=3). Variables vsab, "vsem"

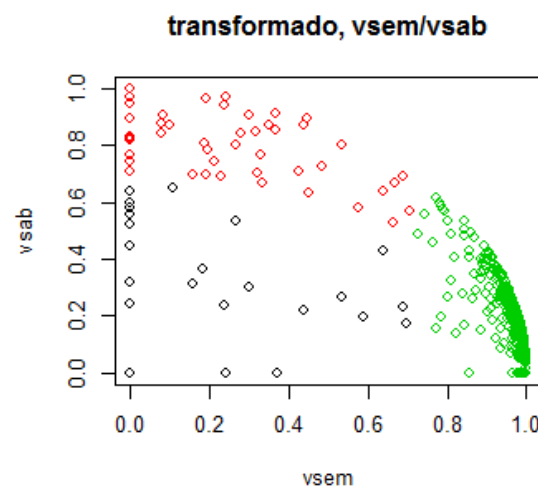


Figura 79. Clustering en espacio transformado (k=3). Variables vsab, "vsem"

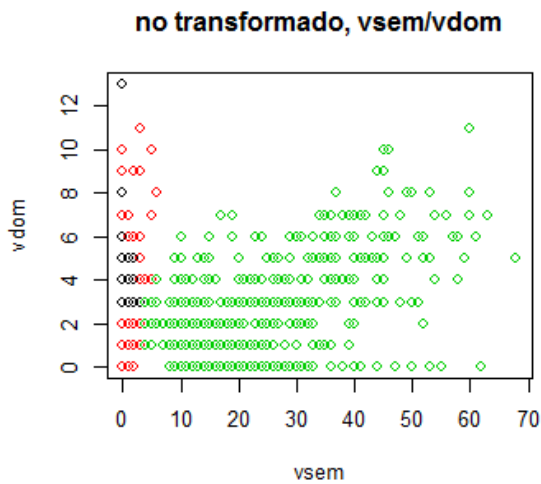


Figura 80. Clustering en espacio original (k=3). Variables “vdom”, “vsem”

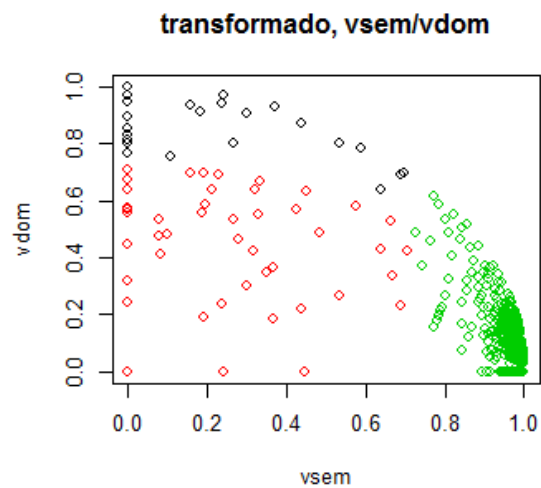


Figura 81. Clustering en espacio transformado (k=3). Variables “vdom”, “vsem”

Finalmente se realiza un algoritmo que calcula *K-means* para valores de *K* entre 2 y 30, con el fin de encontrar la mejor configuración de *clusters* y se sub-itera para cada *K* de manera de tener varias muestras del algoritmo, guardando los resultados en un directorio definido previamente. Se registró su medida de eficacia de cada iteración para luego calcular el método del codo.

La cantidad de *clusters* óptima *K*, se encuentra gracias al término de ganancia, que corresponde a la variación de la configuración para *K* y *K-1 clusters*, de la suma cuadrada de la distancia entre los puntos y sus centroides correspondientes para, la cual muestra la variación atribuida a la relación entre las posiciones de los puntos con sus centroides.

$$Ganancia = indicador_{K-1} - indicador_K \Delta K ;$$

$$indicador_K = F - value_K = \frac{\text{varianza entre clusters}}{\text{varianza intra clusters}} = \min \left( \frac{\frac{\sum_{i=1}^K \sum_{j=1}^{i-1} \text{sqrt}(c_i - c_j)^2}{K}}{\frac{\sum_i^N \text{sqrt}(p_{ji} - c_j)^2}{N - K}} \right) i \neq j$$

*C*, conjunto de centroides de tamaño *K*

*c<sub>i</sub>, c<sub>j</sub>*, centroides que pertenecen a *C*

*p<sub>ij</sub>* punto *i* perteneciente al cluster de centroide *c<sub>j</sub>*

## Evaluación de los resultados

Los resultados indican que la mejor configuración se da para  $K=4$ , en la octava iteración, que posee un indicador de 0,26 y una ganancia de 0,74 con respecto a la configuración con  $K=3$ . A continuación se muestran los *clusters* resultantes así como el resultado gráfico del cálculo del codo.

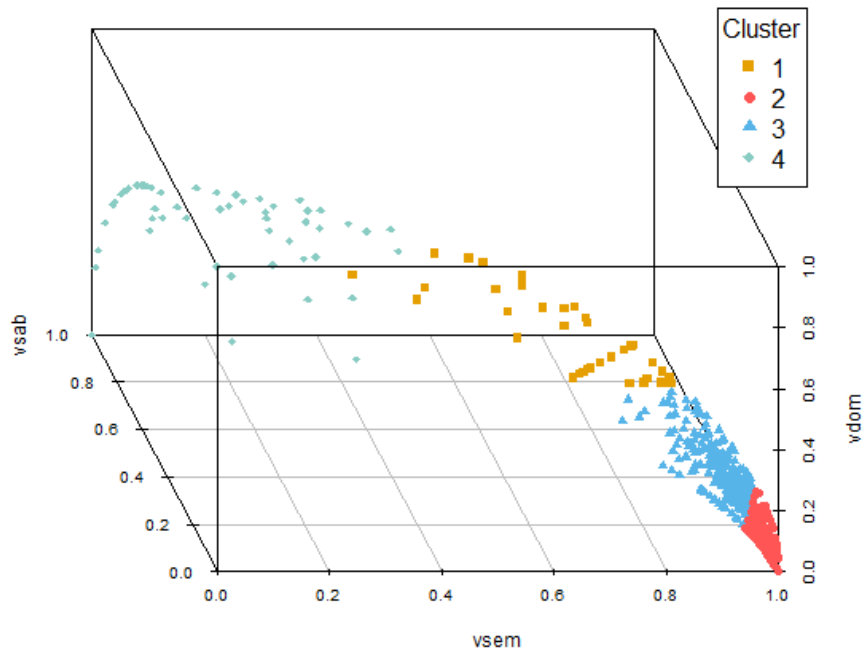


Figura 82. Configuración óptima de K-means con similitud coseno.

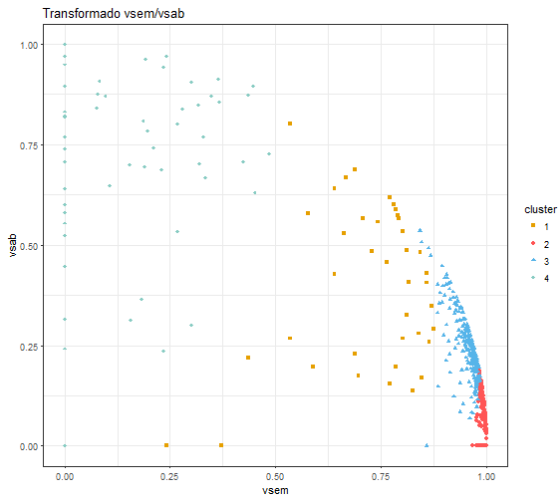


Figura 83. Configuración óptima de *K-means* con similitud coseno.

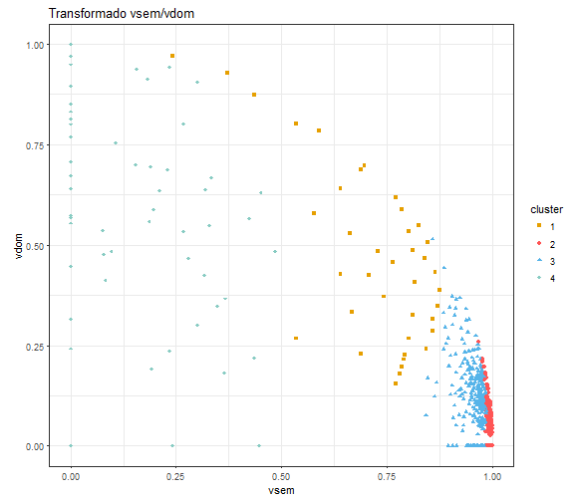


Figura 84. Configuración óptima de *K-means* con similitud coseno.



Figura 85. Configuración óptima de *K-means* con similitud coseno.

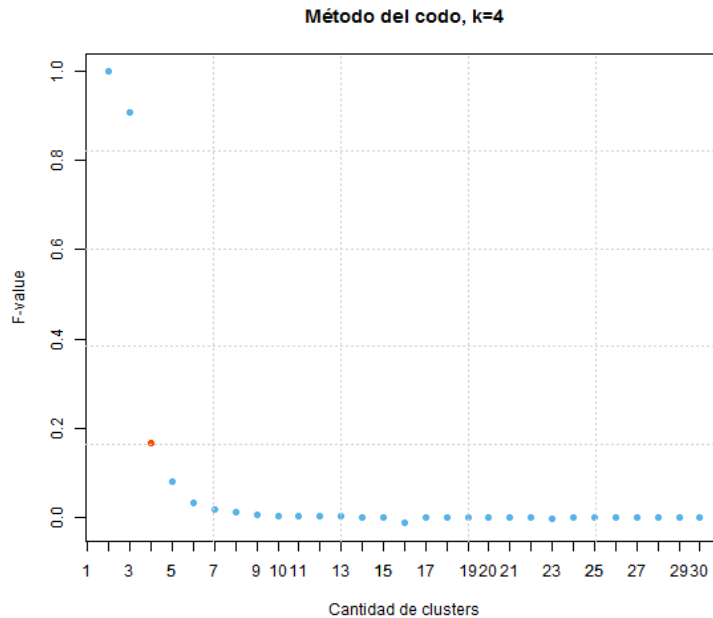


Figura 86. Método del codo, calculado con la mínima de los indicadores calculados para cada K.

La configuración para *K-means*,  $K=4$  y similitud coseno, identifica los *clusters* de clientes con frecuencias opuestas para días de semana y fin de semana. Por otro lado, se identifican *clusters* con cambio gradual en las frecuencias.

En la Tabla 47 se caracterizan los *clusters* según los valores de sus centroides, en esta se pueden identificar claramente en el *cluster 2* y *3* a segmentos con tendencias altas de visitas en días de semana y bajas en sábado y domingo, condición intensificada en el *cluster 2*. Por otro lado el *cluster 4* muestra el segmento de personas con altas visitas en días de fin de semana, principalmente en sábado, y bajas frecuencia de visitas en semana.

Tabla 47. Centroides para 4 *clusters* en Centro de Gravedad Mall Plaza Vespucio

<b>Cluster</b>	<b>Centroide "vsem"</b>	<b>Centroide vsab</b>	<b>Centroide "vdom"</b>	<b>Tamaño del cluster</b>
1	0.73448327	0.47707045	0.48263663	83
2	0.99726496	0.06142221	0.04110863	384
3	0.96232272	0.237719	0.13200249	363
4	0.16042088	0.7856055	0.59756937	117

En otro aspecto, se observa que el tamaño de los *clusters 2* y *3* es similar, concentrando la mayor cantidad de población (384 y 863 observaciones respectivamente), lo cual es esperable debido al filtrado de los datos mencionado anteriormente. Les siguen, en orden descendente, el *cluster 4* con 117 observaciones y el *cluster 1* con 83 observaciones.



Además los valores de frecuencias de visitas transformadas se comparan en base a regresiones lineales sin intercepto, explicando “vsem” por las variables “vsab” y “vdom”. Al respecto se obtiene que las covariables son estadísticamente significativas para explicar la variable “vsem” en todos los *clusters* y que es posible identificar características de cada *cluster* desde los valores obtenidos.

Por otro lado, se realiza un análisis de correlación ( $\rho$ ) de los *clusters* obtenidos y se define un límite de  $|0,7|$  para considerar una correlación importante entre dos variables. En la Tabla 48, se observan altas correlaciones negativas entre los días sábado y domingo en los *clusters* 1 y 4, mientras que los *clusters* 2 y 3 esta situación se da para las variables vsab y “vsem”, y “vdom” y “vsem”.

Tabla 48. Correlación entre variables del modelo

<b>Cluster</b>	$\rho(vsab, vsem)$	$\rho(vdom, vsem)$	$\rho(vsab, vdom)$
1	0,1088307	-0,6103528	-0,7582597
2	-0,6735822	-0,7223362	0,1294245
3	-0,7146842	-0,4452955	-0,2078770
4	-0,1392608	-0,2111503	-0,8854002

Luego se puede caracterizar según el tipo de visitas:

**Cluster 1** (83 observaciones): Personas que aumentan de forma constante la cantidad de visitas en semana con respecto a las visitas en sábado o domingo. Su comportamiento puede ser modelado según la siguiente función lineal:

$$vsem = 0,88 * vsab + 0,58 * vdom$$

De lo cual se entiende que el aumento en 1 visita en sábado genera el aumento en aproximadamente 0,88 visitas en días de semana. De forma análoga el aumento en 1 visita en domingo, genera el aumento en aproximadamente 0,58 visitas en días de semana.

Además del comportamiento de correlación se concluye que cuando una persona visita el Centro de Gravedad en sábado, es muy poco probable que lo visite en domingo, y viceversa.

**Cluster 2** (384 observaciones): Personas que aumentan intensamente la cantidad de visitas en semana con respecto a las visitas en sábado o domingo. Esta tendencia se acrecienta si se realiza una visita más en sábado y su comportamiento puede ser modelado según la siguiente función lineal:

$$vsem = 6,54 * vsab + 4,46 * vdom$$

De lo cual se entiende que el aumento en 1 visita en sábado genera el aumento en aproximadamente 6,54 visitas en días de semana. De forma análoga el aumento en 1

visita en domingo, genera el aumento en aproximadamente 4,46 visitas en días de semana.

Además del comportamiento de correlación se concluye que cuando una persona visita el Centro de Gravedad en sábado o domingo, es muy poco probable que lo visite días de semana, y viceversa.

**Cluster 3** (363 observaciones): Personas que aumentan la cantidad de visitas en semana con respecto a las visitas en sábado o domingo. El comportamiento es atenuado si la visita se realiza en día domingo y su comportamiento puede ser modelado según la siguiente función lineal:

$$vsem = 2,81 * vsab + 1,7 * vdom$$

De lo cual se entiende que el aumento en 1 visita en sábado genera el aumento en aproximadamente 2,81 visitas en días de semana. De forma análoga el aumento en 1 visita en domingo, genera el aumento en aproximadamente 1,7 visitas en días de semana.

Además del comportamiento de correlación se concluye que cuando una persona visita el Centro de Gravedad en sábado, es muy poco probable que lo visite en días de semana, y viceversa.

**Cluster 4** (117 observaciones): Personas que prácticamente no aumentan la cantidad de visitas en semana con respecto a las visitas en sábado o domingo. En particular las visitas en domingo no aumentan la probabilidad de generar visitas en un día de semana. Su comportamiento puede ser modelado según la siguiente función lineal:

$$vsem = 0,18 * vsab + 0,03 * vdom$$

De lo cual se entiende que el aumento en 1 visita en sábado genera el aumento en aproximadamente 0,18 visitas en días de semana. De forma análoga el aumento en 1 visita en domingo, genera el aumento en aproximadamente 0,03 visitas en días de semana.

Además del comportamiento de correlación se concluye que cuando una persona visita el Centro de Gravedad en sábado, es muy poco probable que lo visite en domingo, y viceversa.

### **Análisis de clusters por comuna**

A partir de la información generada al realizar el *Data Warehouse* del Centro de Gravedad es factible analizar las comunas de residencia de todos los identificadores “persId” pertenecientes a un *cluster*. Luego es posible analizar las comunas de procedencia con mayor frecuencia de visita al Centro de Gravedad por *cluster*, como se muestra en la Figura 87.

Se entiende por comuna de residencia a aquellas comunas registradas en la columna "comunaAm1".

Los registros de llamadas a los cuales no se les pudo determinar una comuna (debido a que la antena se encuentra en el límite entre 2 comunas, o bien la aplicación *Google Maps* las registra como parte de 2 comunas no colindantes) se han registrado como "indefinido", mientras que aquellos identificadores "persId" sin registro de llamadas en el horario AM se registran como "sin información".

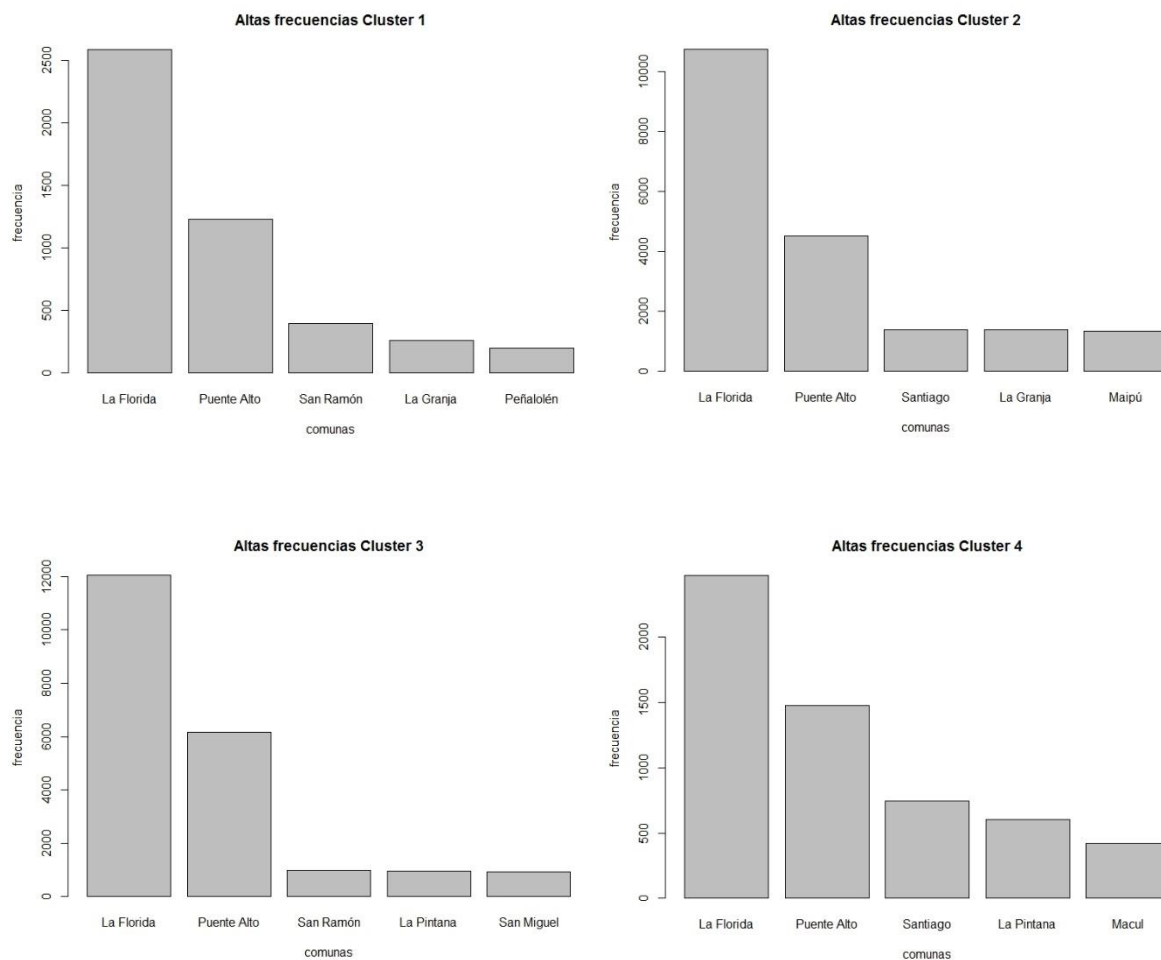


Figura 87 . Comunas de residencia con alta frecuencia de visitas según *cluster*

En la Figura 88, Figura 89, Figura 90 y Figura 91 se entrega una visualización en el mapa permite apreciar la distribución geográfica de las comunas de procedencias con mayor frecuencia de los visitantes del Centro de Gravedad que contiene al *Mall Plaza Vespucio*

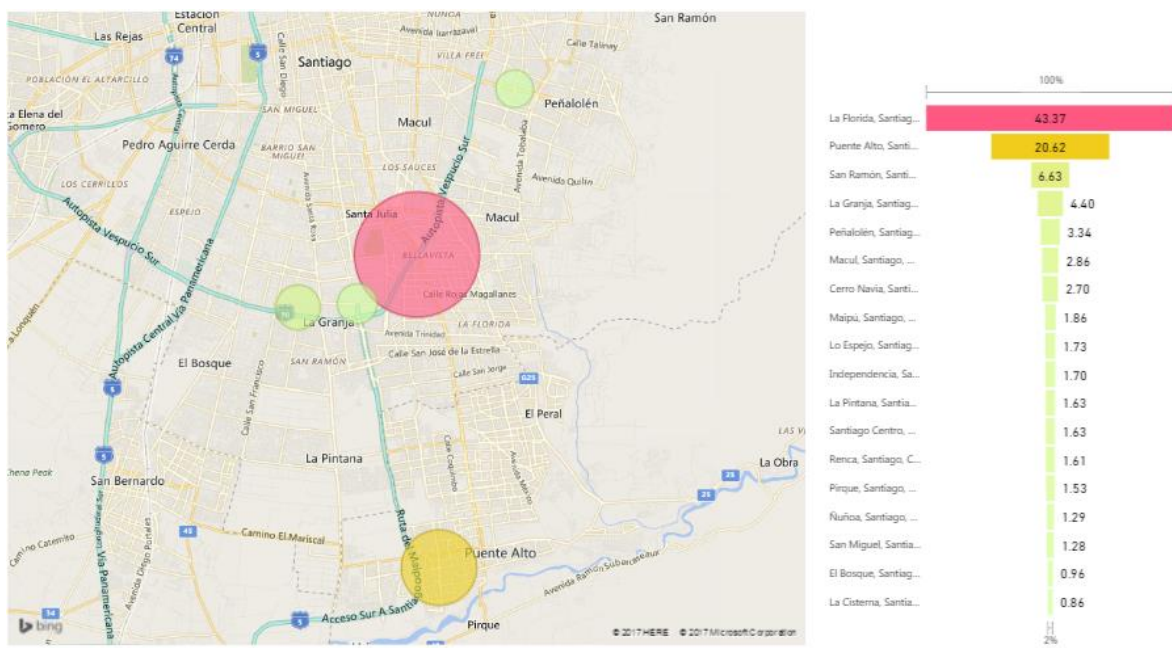


Figura 88. Distribución geográfica de comunas con alta frecuencia de visitas para *cluster 1*

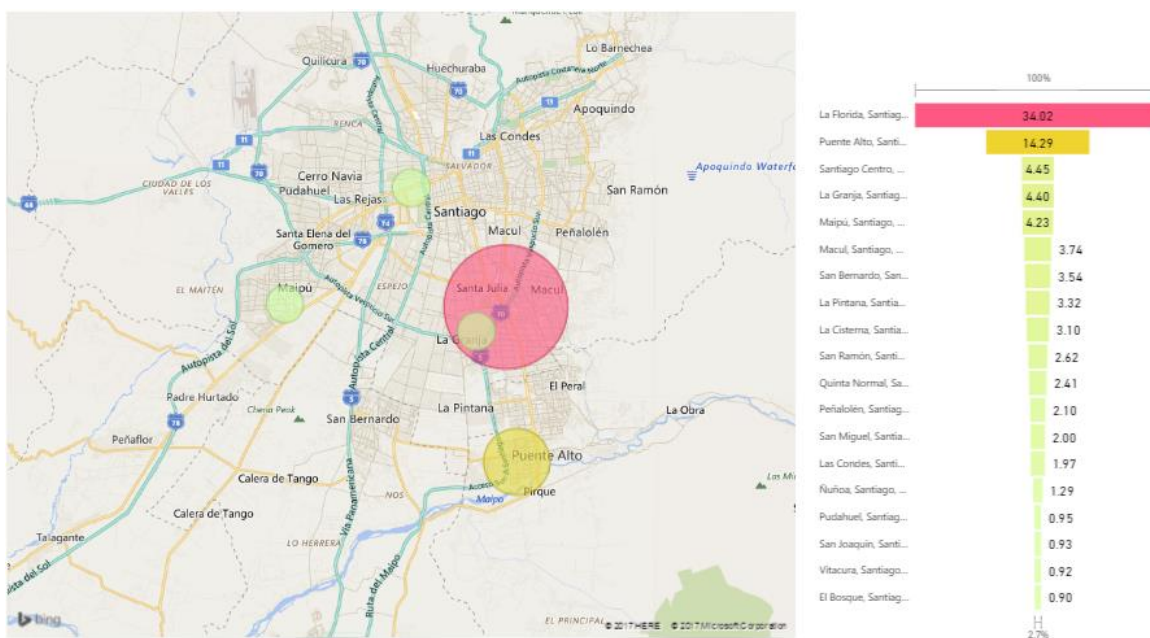


Figura 89. Distribución geográfica de comunas con alta frecuencia de visitas para *cluster 2*

Se observa de forma constante en todos los *clusters* que las comunas de La Florida y Puente Alto corresponden a las comunas de mayor frecuencia de llamadas en horario AM, para todos los *clusters*, llegando hasta al 40% y 20% de las comunas respectivamente, mientras el resto de las comunas representan menos del 15% de las llamadas realizadas.

Las diferenciaciones se encuentran en las comunas con menor participación como se indica en la Tabla 49.

Tabla 49. Comunas con menor participación según cluster

<b>Cluster 1</b>	San Ramón	La Granja	Peñalolén
<b>Cluster 2</b>	Santiago	La Granja	Maipú
<b>Cluster 3</b>	San Ramón	La Pintana	San Miguel
<b>Cluster 4</b>	Santiago	La Pintana	Macul

Los *clusters* 2, 3 y 4 poseen visitas de identificadores “persId” de comunas que no pertenecen a la Región Metropolitana.

- *cluster* 2 posee 0.3% visitas de Antofagasta
- *cluster* 3 posee 0.33% de Arica y 0.19% Curanilahue
- *cluster* 4 posee 0.61% de visitas desde Chañaral

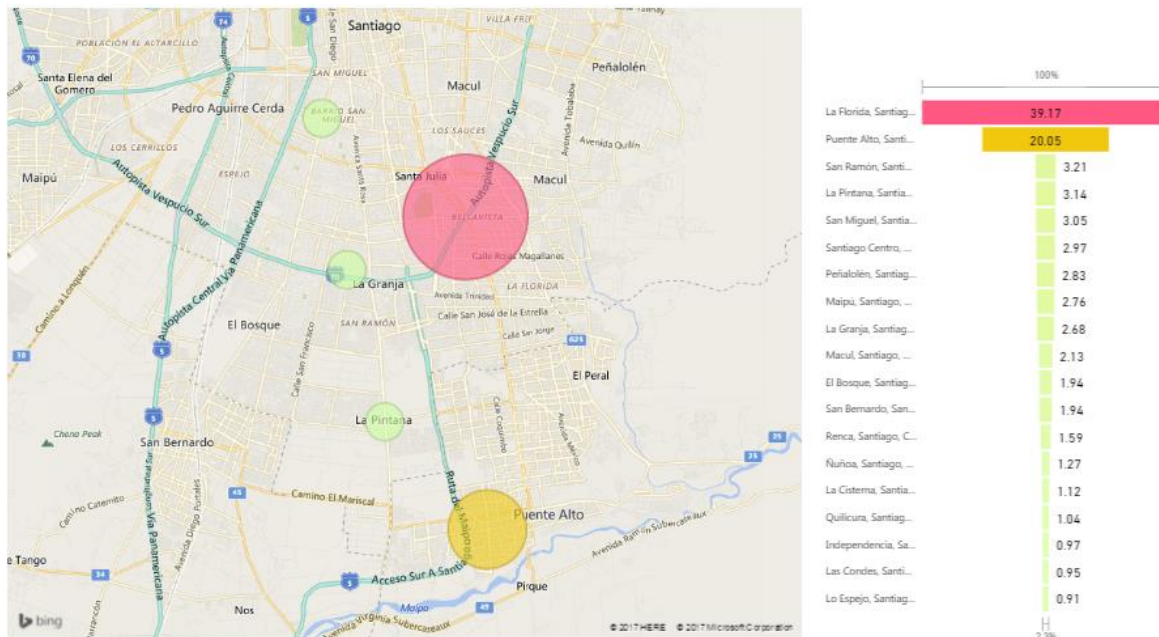


Figura 90. Distribución geográfica de comunas con alta frecuencia de visitas para *cluster* 3



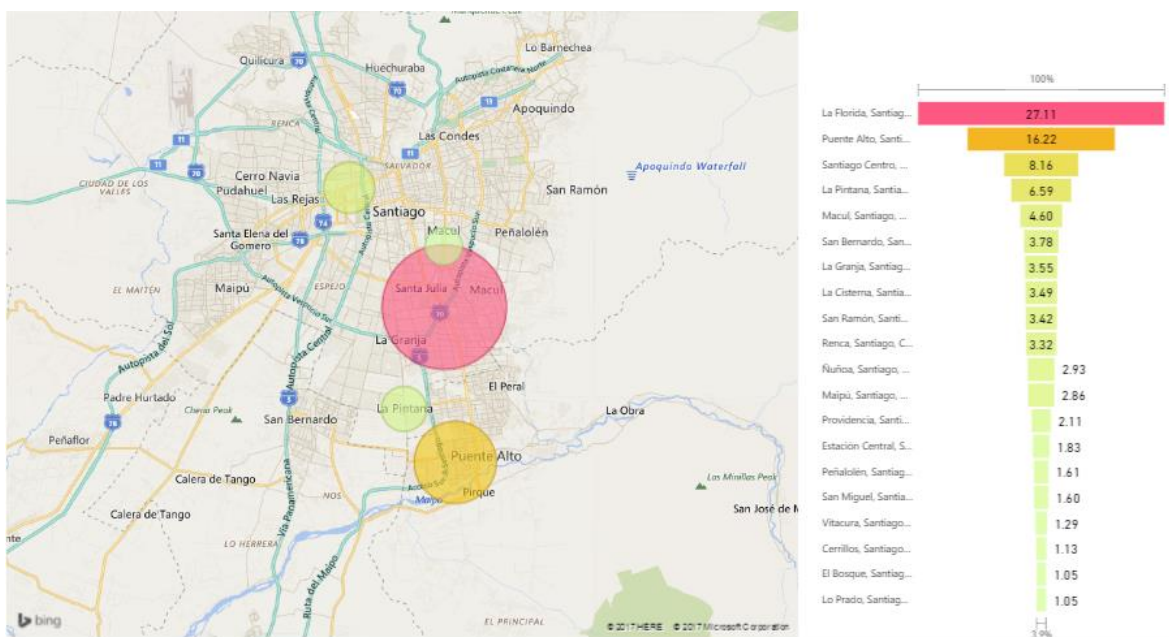


Figura 91. Distribución geográfica de comunas con alta frecuencia de visitas para cluster 4

En adición a los análisis anteriores se genera un estudio económico de los *clusters* a partir del cálculo ponderado de ingresos de cada *cluster*, en base al ingreso promedio por comuna registrado en la encuesta CASEN 2013 [16] y la intensidad porcentual de visitas procedentes de diversas comunas de Santiago (para este cálculo fueron eliminadas las comunas que no pertenecen a la Región Metropolitana).

En este sentido los resultados indican (ingreso en CLP):

- ingreso ponderado *cluster* 1 = \$909.517
- ingreso ponderado *cluster* 2 = \$999.914
- ingreso ponderado *cluster* 3 = \$950.883
- ingreso ponderado *cluster* 4 = \$1.013.994

Lo cual indica que el *cluster* con mayor ingreso corresponde al cluster 4, mientras el con menor ingreso es el *cluster* 1. La diferencia máxima de ingreso es de CLP\$104.477

Con respecto a la composición del tipo de visita según *cluster* se nota en la Tabla 50 la diferenciación del *cluster* 4 compuesto principalmente por registros de tipo “Visita, a diferencia de los demás que tienen como primer tipo la visita de tipo “Trabajo”, además se aprecian altos porcentajes de visitas tipo “Trabajo” en los *clusters* 2 y 3.

Para todos los *clusters* el tipo de visita “Casa” se ubica en el centro de las frecuencias

Tabla 50. Composición de tipo de visitas según cluster

Tipo de Visita	Cluster 1 (%)	Cluster 2 (%)	Cluster 3 (%)	Cluster 4 (%)
Casa	40.55955	31.55265	38.13256	26.37882
Trabajo	15.57154	39.31291	44.28593	16.12564
Visita	43.86890	29.13444	17.58151	57.49554

Uniendo lo anterior con la información de visitas se concluye que:

- El *cluster 4* representa al grupo de personas que visitan la zona con intenciones de compra principalmente en fines de semana y poseen un mayor ingreso promedio.
- Los *clusters 2 y 3* corresponden a visitas compuestas principalmente por registros de tipo Trabajo, con alta frecuencia en días de semana y bajas en fin de semana.
- El *cluster 1* presenta mayor composición de visitantes que de trabajadores, sin embargo sus registros de tipo Visita tienen una tendencia menos marcada que en el *cluster 4*.

En complemento con la información de comunas se concluye:

- Con alta probabilidad quienes trabajan (*clusters 2 y 3*) pertenecen, principalmente, a la comuna del Centro de Gravedad o bien comunas cercanas (Santiago, Puente Alto, La Granja, La Pintana, San Ramón, San Miguel), con excepción de la comuna de Maipú.
- San Miguel sólo aparece en el *cluster 3* por lo que podrían tratarse de visitas de tipo Trabajo con tendencia a concurrir al Centro de Gravedad en días de semana.
- Es probable que, dada la composición del tipo de visitas y las comunas de alta frecuencia del *cluster 1*, que aquellos registros de San Ramón y La Granja sean de tipo "Trabajo", pues estas comunas ya tienen una tendencia marcada de visitas en los *clusters 2 y 3*, mientras las visitas de Peñalolén sean de tipo "Visita", con una distribución de registros uniforme entre días de semana, sábado y domingo.
- Se induce que las visitas desde Macul aumentan en sábados y domingos y podrían tratarse de visitas de tipo "Visita" (pues están presentes con alta frecuencia únicamente en el *cluster 4*).
- Las visitas desde Santiago y La Pintana tienen tendencia menos definida, ya que a pesar de estar presentes en el *cluster 4*, también se registran con alta frecuencia en los *clusters 2 y 3*.

Finalmente se presenta en la Tabla 51 un resumen de la información de los *clusters* encontrados

Tabla 51. Descripción de *clusters*

	<b>Cluster 1</b>	<b>Cluster 2</b>	<b>Cluster 3</b>	<b>Cluster 4</b>
<b>Nombre del cluster</b>	Visitantes neutros	Trabajadores de semana o vecinos de CG	Trabajadores de semana	Compradores y trabajadores de fin de semana
<b>Tamaño del cluster</b>	83	384	363	117
<b>Tendencia de días de visita</b>	Neutra	Semana	Semana	Sábados y Domingos
<b>Comunas de procedencia (en orden decreciente)</b>	La Florida	La Florida	La Florida	La Florida
	Puente Alto	Puente Alto	Puente Alto	Puente Alto
	San Ramón	Santiago	San Ramón	Santiago
	La Granja	La Granja	La Pintana	La Pintana
	Peñalolén	Maipú	San Miguel	Macul
<b>Ingreso promedio</b>	\$ 909.517	\$ 999.914	\$ 950.883	\$ 1.013.994
<b>Tipo de visita principal</b>	Visita	Trabajo	Trabajo	Visita

El resultado del *clustering* genera nuevas variables descriptoras de “persId” (ver Tabla 52) que se agregan al *Data Warehouse*, el que finalmente contará con 61 variables.



Tabla 52. Variables generadas luego del *clustering*

Variable	Tipo	Descripción
"cluster"	Factor	Captura el <i>cluster</i> al que pertenece "persId" n.
"centroidesem"	Factor	Captura el valor de total de visitas en semana del centroide más cercano de "persId" n
"centroidesab"	Factor	Captura el valor de total de visitas en sábado del centroide más cercano de "persId" n
"centroidedom"	Factor	Captura el valor de total de visitas en domingo del centroide más cercano de "persId" n
"cluster_1", "cluster_2", "cluster_3", "cluster_4"	Factor	Indica la pertenencia o no de "persId" n al <i>cluster</i> c.

### Conclusiones del modelo

Las conclusiones de la realización de *K-means* nos indican que este algoritmo es capaz de representar grupos de personas por frecuencias de visitas y que es viable agrupar *clusters* con comportamiento similar de la configuración óptima para facilitar el análisis orientado al negocio.

Por último, se advierte que es posible predecir la cantidad de visitas en días de semana de una persona, al tener la información del *cluster* al que pertenece, la cantidad de visitas en sábados y domingo, considerando que estos podrían representarse como una regresión lineal.

### 5.3.3 Replicaciones

El análisis exploratorio se replica para datos de 5 Centros de Gravedad que concentran los siguientes centros comerciales: *Mall* Vespucio Norte, *Mall* Alto Las Condes, *Mall* Plaza Egaña, *Mall* Arauco Maipú y *Mall* Plaza Oeste. Las cuales contiene las mismas variables iniciales: "antena", "persId", "antena\_entrada", "receptId", "fecha", "hora", "duración". La información se obtiene a partir de los resultados del estudio de uso de suelo según datos de telefonía de R. Muñoz, S. Ríos, R. Roda y C. Contreras. [20].

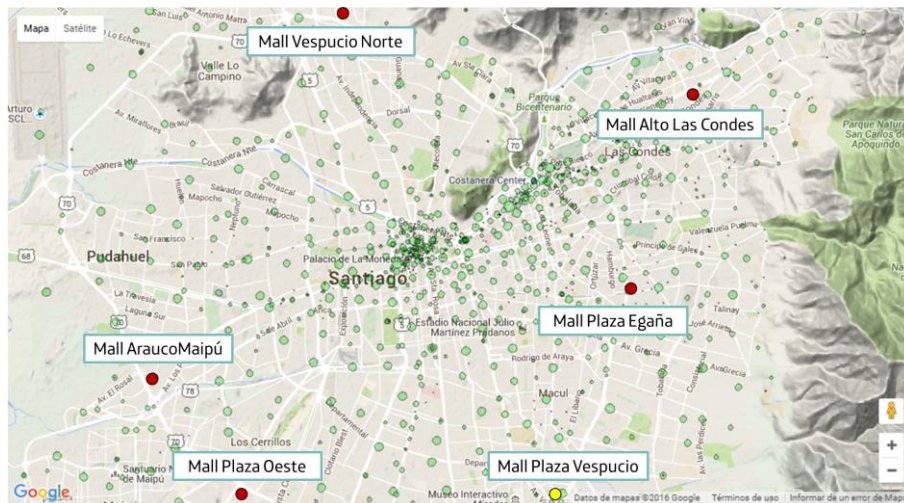


Figura 92 . Ubicación de Centros de Gravedad para replicación

Los resultados muestran que se mantiene la tendencia decreciente de compras entre mayo y agosto. A nivel semanal la tendencia general es de llamadas estables durante la semana y aumento de estas en los días de fin de semana. Además se observa que existen Centros de Gravedad que comparten tendencia con mayor fuerza que otros. Como ejemplo se tiene comportamientos muy similares entre los Centros de Gravedad que contienen a los *malls* Alto Las Condes y Plaza Egaña, así como los que contienen a Arauco Maipú y Vespucio Norte.

Por otro lado el comportamiento exponencial negativo de la frecuencia de llamadas por persona se mantiene en todos los Centros de Gravedad, donde en promedio el 99% de las personas llama menos de 25 veces y visita menos de 16 veces el Centro de Gravedad, en el periodo de estudio. Se observa que tanto las llamadas como visitas mantienen dos *peaks* alrededor de las 12 y 19 horas y un valle cerca de las 15 horas. El comportamiento se ve suavizado en el caso de las visitas debido a que más de una llamadas de un número en el mismo intervalo de tiempo sólo se considera como una visita.

Se observa que el comportamiento en todos los Centros de Gravedad es similar y, por ende, se espera que los resultados de la segmentación sean similares al analizado en la sección 5.3.2, lo cual se corrobora al replicar el modelo para cada Centro de Gravedad. Además, y en base a lo analizado anteriormente, se plantea la posibilidad de un análisis global que permita caracterizar a personas que visitan algún centro de gravedad. Al respecto se observa que, en general el 99% de las personas visita menos de 21 veces algún Centro de Gravedad, lo que corresponde a 4.488 número únicos.

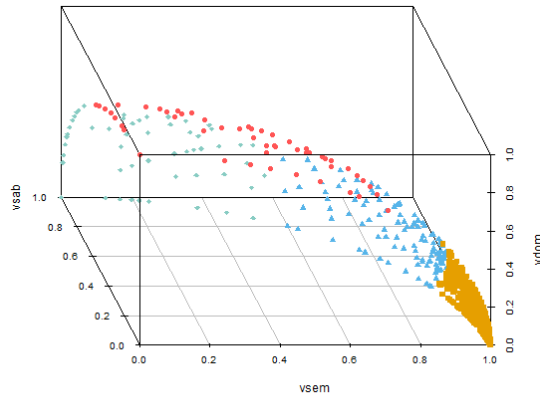


Figura 93. Óptimo K-means para *Mall Vespucio Norte*(1594 observaciones)  
K óptimo = 4

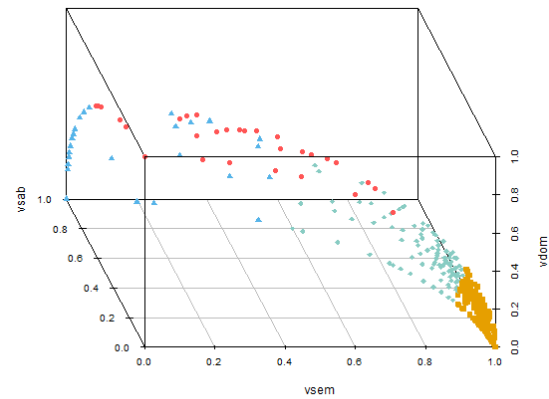


Figura 94. Óptimo K-means para *Mall Plaza Oeste* (912 observaciones)  
K óptimo = 4

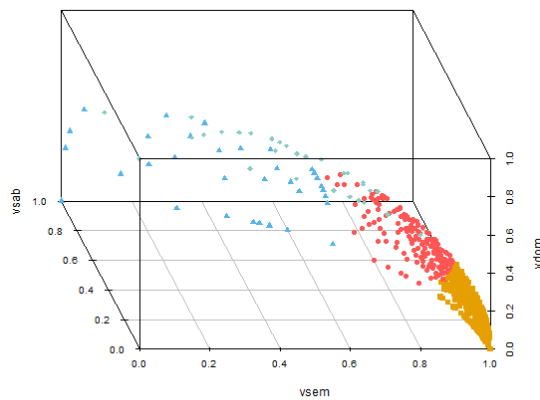


Figura 95. Óptimo K-means para *Mall Arauco Maipú* (1077 observaciones)  
K óptimo = 4

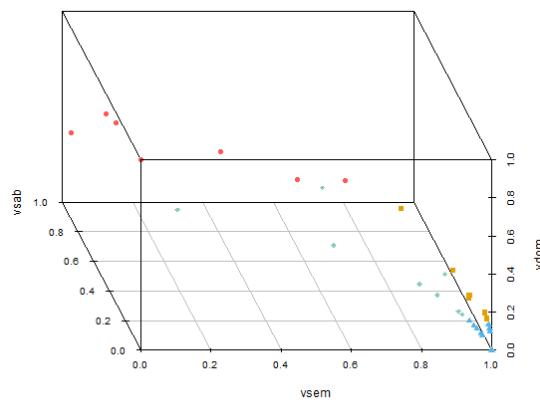


Figura 96. Óptimo K-means para *Mall Plaza Egaña* (165 observaciones)  
K óptimo = 4

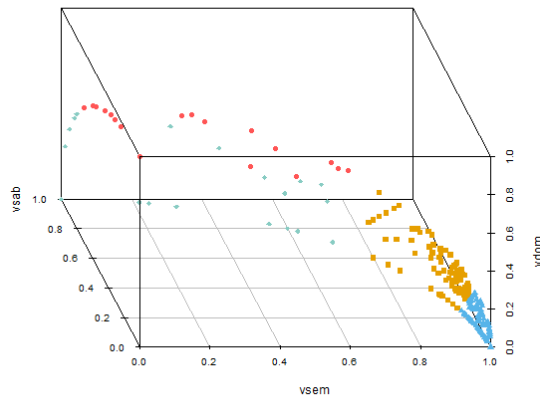


Figura 97. Óptimo K-means para Mall Alto Las Condes (1423 observaciones).  
K óptimo = 4

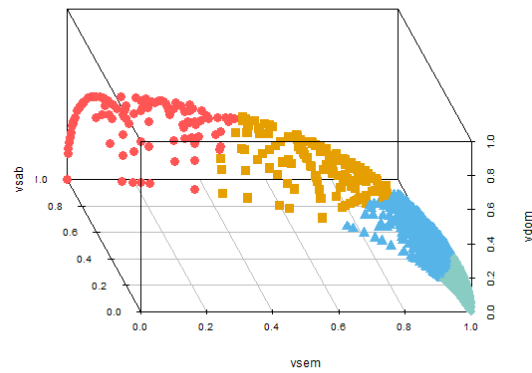


Figura 98. Óptimo K-means para unión de Centros de Gravedad (624.216 observaciones).  
K óptimo = 4

La cantidad de *clusters* óptima es igual para todos los Centros de Gravedad analizados, con valor  $K = 4$ , donde se destaca la existencia de los grupos de personas con altas visitas en fin de semana pero baja y en sábado o domingo, un grupo con altas visitas en sábado y/o domingo y baja en días de semana, y un grupo intermedio.

Más detalle de los resultados del análisis exploratorio y modelados se describen en la sección de anexos (ver 10.5).

## 5.4 Pronóstico

Los modelos de pronóstico se trabajaron con los datos del CG que contiene al centro comercial Mall Plaza Vespucio.

### 5.4.1 Generación de datos de entrenamiento y testeo:

Se genera un Split no estratificado de los datos que entrega una base de entrenamiento de 70.267 observaciones y una base de testeo de 30.115 observaciones. Tanto la base de entrenamiento como la de testeo conservan la proporción de visitas y no visitas (20% versus 80%) de la base original.

La selección de variables usada para la ejecución de los siguientes modelos itera una regresión lineal generalizada con el método *stepwise regression* para detectar aquellas variables que entregan mayor valor al modelo, mediante la minimización del indicador *Akaike Information Criterion (AIC)*, el cual penaliza el ajuste por la cantidad de variables del modelo.

Las variables finalmente contempladas son “*vfechaMenosUno*”, “*vfechaMenosTres*”, “*vsem*”*anaMenosUno*”, “*vsemanaMenosDos*”, “*jue*”, “*vie*”, “*sab*”, “*dom*”, “*casatrabajo*”, “*semanaMes*”, “*mes*”, “*vlun*”, “*vmar*”, “*vmie*”, “*vjue*”, “*vvie*”, “*vsab*”, “*vdom*”, “*vsem*”, “*vfds*”, “*cluster*”.

Dada la proporción de datos se aplica una combinación de *under* y *over sampling* sobre los datos de entrenamiento, con el objetivo de balancear las clases y obtener una mejor respuesta de los modelos de pronóstico. El *under* y *over sampling* eliminan observaciones de la clase mayoría y agrega observaciones de la clase minoritaria respectivamente.

Tabla 53. Proporción de datos

objetivo	0	1
original	0,7967325	0,2032675
<i>Under sampling</i>	0,5	0,5
<i>Over sampling</i>	0,5	0,5
<i>Under-over sampling</i>	0,5017149	0,4982851

#### 5.4.2 Modelado:

Se definen los costos de cometer errores tipo 1 y tipo 2 para mejorar el ajuste de los modelos:

El error tipo 1 corresponde al falso positivo (FP), que se produce cuando el modelo clasifica como visita (clase +1) una observación donde no se produce visita (clase -1). Esto implica el envío de un mensaje a una persona que no lo ocupará, pues no visitará el Centro de Gravedad, este costo es de alrededor de CLP\$40.

Por otro lado el error tipo 2 corresponde a un falso negativo (FN), donde el modelo clasifica como no visita (clase -1) una observación que efectivamente visita el CG (clase +1). En este caso el costo corresponde al costo de oportunidad de ganar utilidad con el envío y cobro de un mensaje. Este costo es de aproximadamente CLP\$240 y se calcula a partir del valor presente del flujo operacional sobre la cantidad de campañas estimadas por la cantidad promedio de mensajes enviados, el flujo operacional considera la tasa de conversión de mensajes dentro del precio de la campaña (alrededor de un 2%).

El costo del nivel de error del modelo se calcula como  $CLP\$40 * FP + CLP\$240 * FN$ .

Respuesta del método	
1	0

Clase 1	0	CLP \$240
Clase 0	CLP \$40	0

Figura 99. Matriz de costos

Los modelos son ajustados según la base de entrenamiento y su resultado es optimizado con el objetivo de minimizar el costo por error, utilizando la base de testeo.

Para esto se fija la siguiente regla:

$$valor_{predicción} < umbral. \text{ Entonces, } valor_{predicción} = 0$$

$$valor_{predicción} \geq umbral. \text{ Entonces, } valor_{predicción} = 1$$

Luego se itera el umbral entre 0 y 1, se calculan las medidas de desempeño y costo de modelo para cada configuración.

## Logit

Los resultados reflejan que el modelo predice mejor que una asignación aleatoria, esto puede apreciarse tanto en la curva *ROC* como en el valor de la medida de desempeño *AUC* (0,7). Por otro lado la configuración que minimiza el costo del modelo entrega valores de *precision*, *recall* y *accuracy* de 59%, 39% y 74% respectivamente, lo que se da para un umbral de 0.1, es decir, si el modelo predice una probabilidad menor a 0.1 entonces se clasifica como 0 (no visita), sino clasifica como 1 (visita)

## Conditional Tree

Este tipo de árbol particiona de forma recursiva, basándose en una inferencia condicional. Se fija un *split* mínimo de 30 observaciones por nodo y se itera el umbral de predicción entre 0 y 1. Los resultados muestran que este modelo es similar al árbol de decisión con *gini index* como criterio de selección, lo cual se observa en el valor del área bajo la curva *ROC* (0,79).

Por otro lado el costo del modelo se minimiza para un umbral de 0.4 y genera las siguientes medidas de desempeño:

- *Recall* = 87%
- *Precision* = 31%
- *Accuracy* = 59%
- $F_2 = 0,58$

## Árbol de decisión: CART

Se prueba un árbol de decisión que usa el criterio *Gini index* como seleccionador de variables. Los resultados indican que el árbol de decisión tienen un poder de predicción similar al del modelo *logit* ( $AUC = 0,68$ ), para un umbral de 0,14. Sin embargo, los valores de *precision*, *recall* y *accuracy* que minimizan el costo de los errores difieren bastante de los modelos anteriores

- *Recall*: 92%
- *Precision*: 36%
- *Accuracy*: 65 %

Además se calcula el estadístico  $F_2$  (0,83) que indica que el desempeño del modelo explicado por las varianzas de *precision* y *recall* es efectivo el 83% de las veces, dándole mayor peso a los errores de tipo 1.

## Support Vector Machine

Dada la complejidad del algoritmo se hace un re-muestreo de 15.000 datos reduciendo el tiempo de ejecución a unos 10 minutos (sobre alrededor de 2 horas, corriendo el modelo en un computador con Intel core i3 y 6Gb de memoria RAM).

En general, se observa que el modelo responde mejor que una predicción aleatoria con un criterio AUC de 0,69. Para minimizar el costo del modelo se requiere de un umbral de 0,26, lo cual entrega medidas de desempeño de 51% de *accuracy*, 85% de *recall*, 27% de *precision* y  $F_2$  de 0,6.

## Conclusiones:

Se presentan las curvas ROC de cada uno de los algoritmos probados y una tabla de comparación de sus medidas de desempeño.

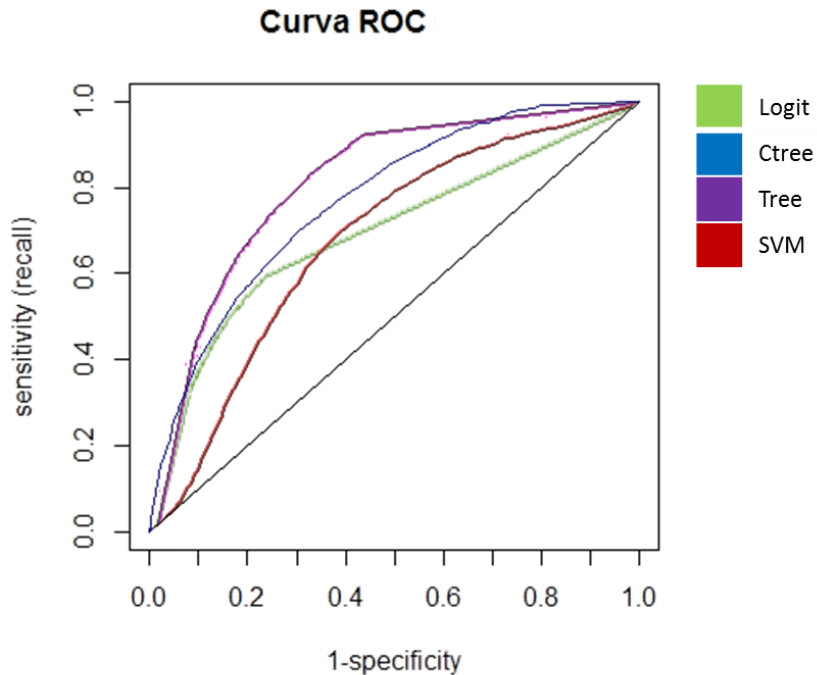


Figura 100. Comparación curvas ROC

Tabla 54 . Tabla comparativa de desempeño

	<b>Logit</b>	<b>SVM</b>	<b>Tree (CART)</b>	<b>Ctree</b>
<i>Recall</i>	59%	85%	92%	87%
<i>Precision</i>	39%	27%	36%	31%
<i>Accuracy</i>	74%	51%	65%	59%
Umbral	0.1	0.26	0.14	0.4
$F_2$	54%	60%	83%	58%
<i>AUC</i>	70%	69%	68%	79%
Costo	CLP\$822.680	CLP\$780.160	CLP\$520.880	CLP\$650.720

A pesar de que el comportamiento, según las medidas de desempeño, es similar en todos los casos, se observa que el modelo que minimiza el costo por errores es el árbol de decisión con criterio de selección *gini index*, el cual genera 3600 hojas resultantes y toma 30 segundos de ejecución. La diferencia de costos por error es de alrededor de CLP\$130.000 con el siguiente modelo más económico y de cerca de CLP\$300.000 con el modelo más caro.

Es destacable que el modelo *logit* genere las predicciones más costosas pues se esperaba que este algoritmo respondiera mejor al tipo de datos de panel presentados.



Además se destaca que los árboles de decisión, tanto *CART* como *Ctree*, entreguen los mejores resultados, trabajando con reglas de decisión según la correlación de las variables independientes con la variable objetivo. Esto hace pensar que un algoritmo *Random Forest* podría funcionar aún mejor al encontrar el mejor resultado de una comparación con gran cantidad de árboles.

Sobre las medidas de desempeño se observa que el criterio *accuracy* no parece reflejar la efectividad del modelo, esto se observa en el modelo *logit* donde a pesar de tener el mayor valor de esta medida de desempeño, también es el más costoso. Sin embargo, al observar el ratio entre precisión y recall se observa que este modelo tiene una razón 2 veces mayor que el resto de los algoritmos. En relación a lo anterior, se observa que el árbol de decisión *CART* entrega menores costos ya que aumentan los valores de *recall* y *precision*.

La forma de la curva ROC también parece entregar información importante, tanto en el árbol *CART* como en el *Ctree* la curva es pronunciada y se esperan altos valores para *recall*. Por otro lado, los modelos *logit* y *SVM*; se ve claramente que el máximo desempeño se encontrará para un recall serán cercano al 60% y 80% respectivamente.

Con respecto al comportamiento de los modelos en relación a casos extremos de pronóstico se tiene que:

- Un modelo donde todas las predicciones sean iguales a la clase mayoritaria (clase 0) genera un costo aproximado de CLP\$1.469.000. Por otro lado el pronosticar todos los eventos como 0, implica suponer que ningún usuario visitará el centro de gravedad, por lo que la ejecución de las campañas perdería sentido.
- De forma análoga, pronosticar que todos los usuarios visitarán el Centro de Gravedad, produce campañas con un costo por error aproximado de CLP\$960.000. En este caso los mensajes de beneficios se enviarían a todos los usuarios suscritos, perdiendo la personalización temporal del servicio.

Una campaña basada en modelos de pronóstico genera, en el mejor de los casos, campañas con costo 2 y 3 veces menor a los casos donde todas las predicciones son “visita” (clase 1) y todas las predicciones son “no visita” (clase 0), respectivamente. Además la aplicación de un modelo de pronóstico permite al servicio acercarse al objetivo de beneficios en el momento correcto, utilizando información registrada en antenas telefónicas.

## CAPÍTULO 6: PROPUESTA DE APOYO TECNOLÓGICO

---

### 6.1 Especificación de Requerimientos

En el siguiente capítulo se describen los sistemas tecnológicos necesarios para el funcionamiento de una empresa de marketing móvil geolocalizado.

#### 6.1.1 *Requerimientos Funcionales*

##### 1. Sistema de información de usuarios (*SIU*):

Requiere seguir un patrón de Modelo Vista Controlador (*MVC*). En este patrón la administración del comportamiento y estado de los datos de la aplicación son llevados por el Modelo; la interacción con el usuario final se realiza en la Vista; y la lógica de cambios en el Modelo o interacciones entre la Vista y el Modelo son llevadas por el Controlador. Los datos deberán ser guardados en una base de datos, por lo que el Modelo debe tener todas las funciones necesarias para acceder a las tablas y hacer las operaciones de datos de crear, leer, actualizar y borrar (*CRUD*) sobre esta base.

Las vistas deberán seguir los estándares de seguridad *web* del sistema de cifrado *Hypertext Transfer Protocol Secure (HTTPS)*. Además de los siguientes estándares y tecnologías establecidas por el *World Wide Web Consortium (W3C)* para *WebApps*:

- Uso de *Hyper Text Markup Language (HTML)*: Lenguaje para describir la estructura de las páginas, gracias a este es posible incluir encabezados, tablas, textos planos, entre otros recursos. Además permite obtener información de links, diseñar formas de conducir transacciones con servicios remotos, etc.
- Uso de *Cascading Style Sheets (CSS)*: Lenguaje para describir la presentación de páginas *web*, incluyendo colores, *layout* y fuentes. Este lenguaje permite adaptar la presentación de diferentes tipos de servicios, es independiente de *HTML*, lo cual hace que los sitios sean más fáciles de mantener.
- Uso de *JavaScript*: lenguaje de programación interpretado, el cual es orientado a objetos y que entrega dinamismo a las páginas. Las formas más básicas de *scripting* desarrolladas por el *W3C* es el *Document Object Model (DOM)*, el cual permite a los programas y scripts acceder y actualizar dinámicamente el contenido, estructura y estilo de documentos.

Además:

- El sistema debe permitir la ejecución del proceso de *Opt-in* y *Opt-out*
- El sistema debe permitir el registro sólo de clientes de la TELCO aliada
- El sistema debe permitir la desuscripción al servicio
- El sistema debe tener un inicio de sesión

- Cada sesión se asigna a un cliente de la TELCO asociada
  - El sistema debe desplegar el código QR del mensaje enviado
  - El sistema debe permitir la revisión de los mensajes recibidos de cada usuario
  - El sistema debe permitir editar la información de cada usuario
2. Sistema de clientes (SC):
- El cliente debe poder verificar en el sistema los códigos QR entregados por un usuario
  - Debe permitir el despliegue de informes de avances, así como de un informe final de desempeño.
  - El sistema debe permitir el registro del uso de códigos QR
3. Sistema de producción (SP):
- La información de segmentos generados en la etapa de modelado deben poder ser cargados en el sistema, cubriendo los siguientes campos: Número de teléfono, segmento, id del centro de gravedad estudiado. Donde la llave de cada registro es una combinación de los 3 atributos.
  - El sistema debe permitir la carga de datos de contrato como: segmentos buscados, cantidad de mensajes a enviar, centro de gravedad estudiado, etc.
  - El sistema debe permitir la carga de la información de los mensajes enviados en campañas anteriores, cubriendo los siguientes atributos: Número de teléfono, mensaje enviado, fecha enviada, hora enviada, id de la campaña (idCampaña\_idMensaje)
  - El sistema debe ser capaz de generar un listado de mensajes por enviar, cumpliendo las restricciones de capacidad y preferencias del cliente.
  - El listado generado debe incluir las siguientes categorías: Número telefónico, mensaje a enviar, fecha de envío, hora de envío, id de la campaña (idCampaña\_idMensaje)

### **6.1.2 Requerimientos No Funcionales+**

[1] Sistema de información de usuarios:

- La interfaz de una sesión iniciada debe tener un menú en el cual se encuentren los módulos de información de mensajes recibidos y un módulo de desuscripción
- La interfaz debe mostrar un botón de desuscripción en la vista “home”
- La interfaz debe mostrar un botón de cierre de sesión en la vista “home”
- La aplicación debe ser dirigida a un usuario que lea y entienda español
- El sistema debe ser *responsive* a distintos tipos de pantalla, con una resolución de pantalla entre 320x568 *pxls* y 1136x640 *pxls*.
- Los datos de un usuario deben ser privados para el resto de los usuarios

## Mockups del Sistema de Información de Usuarios:

Mockup of the login page for 'EMPRESA'. The browser address bar shows 'http://direccion.empresa.com'. The page features a central form with a placeholder for the company name, fields for 'Teléfono' and 'Contraseña', and an 'INGRESAR' button. A link for '¿Olvidaste tu contraseña?' is located below the button. The footer contains a 'Log In TERCERO' button and links for 'Contáctanos', 'FAQs', and '¿Quiénes somos?'.

Figura 101. Mockup "Inicio de sesión"

Mockup of the 'TUS DATOS' registration page. The browser address bar shows 'http://nombre.empresa.com/Datos'. The page includes a vertical scrollbar on the right. The form contains fields for 'Número de teléfono', 'Nombre o Nick', 'Mail de Seguridad' (with a note 'Te enviaremos un correo de confirmación' and a help icon), 'Fecha de nacimiento' (with a calendar icon), gender selection ('Hombre' selected, 'Mujer'), 'Contraseña', and 'Repite tu contraseña'. 'CANCELAR' and 'CONFIRMAR' buttons are at the bottom.

Figura 102. Mockup "Ingresar datos"

Mockup of the 'TUS PREFERENCIAS' registration page. The browser address bar shows 'http://nombre.empresa.com/Preferencias'. The page includes a vertical scrollbar on the right. It features a list of categories with checkboxes: 'Alimentos y Restaurantes' (checked), 'Bancos', 'Belleza', 'Decohogar', 'Deportes' (checked), 'Entretención' (checked), 'Libros y Artes', 'Música' (checked), 'Tecnología' (checked), 'Vestuario' (checked), and 'Viajes' (checked). 'CANCELAR', 'VOLVER A MIS DATOS', and 'CONFIRMAR' buttons are at the bottom.

Figura 103. Mockup "Ingresar preferencias"

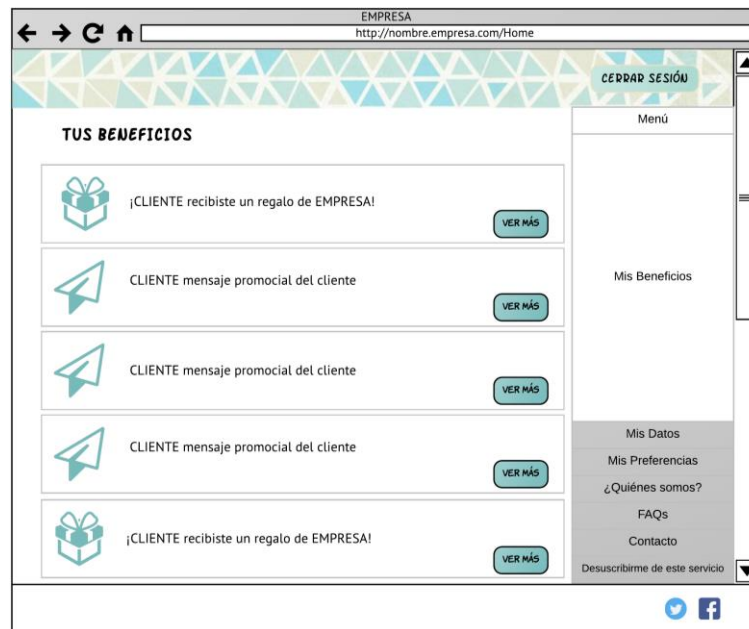


Figura 104. Mockup "Home"

[2] Sistema de clientes:

- El tiempo de generación del código QR, desde que se recibe la petición, no debe superar los 3 segundos.

### Reportes de desempeño de campaña para clientes

El sistema de reportes se trabaja en el sistema Power BI de Microsoft, el cual permite una conexión en tiempo real con las bases de datos resultantes de la aplicación de modelos y ejecución de campañas de marketing móvil.

La información de campañas es visible desde cualquier dispositivo móvil, permite generar *dashboards* en formato PDF y exportar reportes en formato PPTX.

## ANÁLISIS DE SEGMENTOS DE CLIENTES

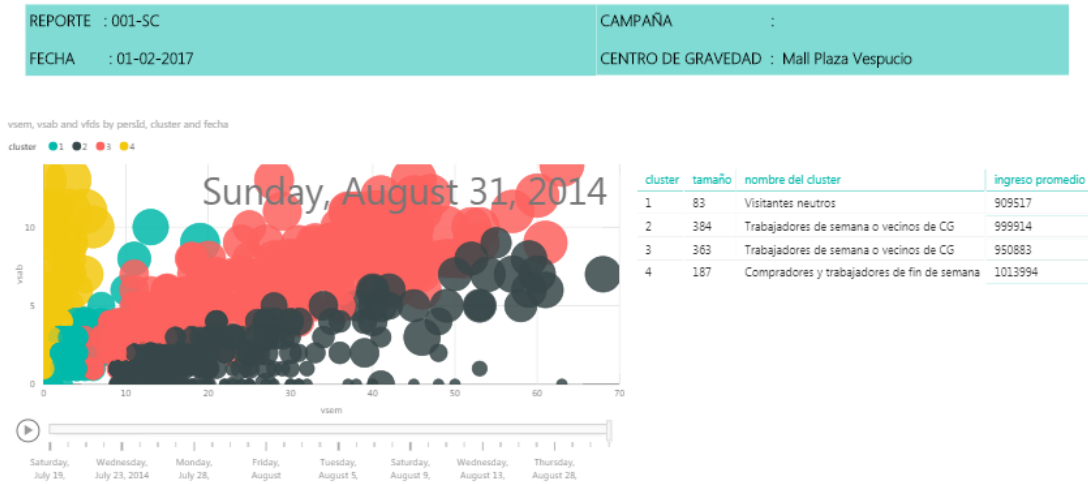


Figura 105. Reporte de ventas, composición de *clusters* por CG

## FRECUENCIA DE VISITAS POR HORA/DIA



Figura 106. Reporte de ventas, frecuencia de visitas por *cluster*/hora

## FRECUENCIA DE VISITAS POR HORA/DIA

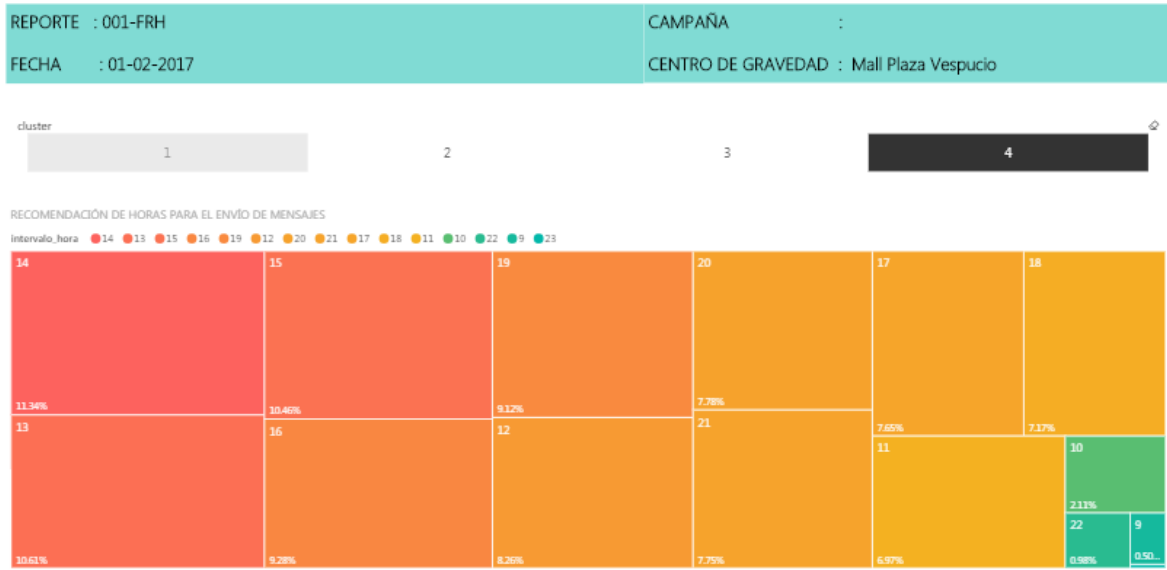


Figura 107. Reporte de ventas, recomendación de horarios

## ANÁLISIS DE PROCEDENCIA POR CLUSTER

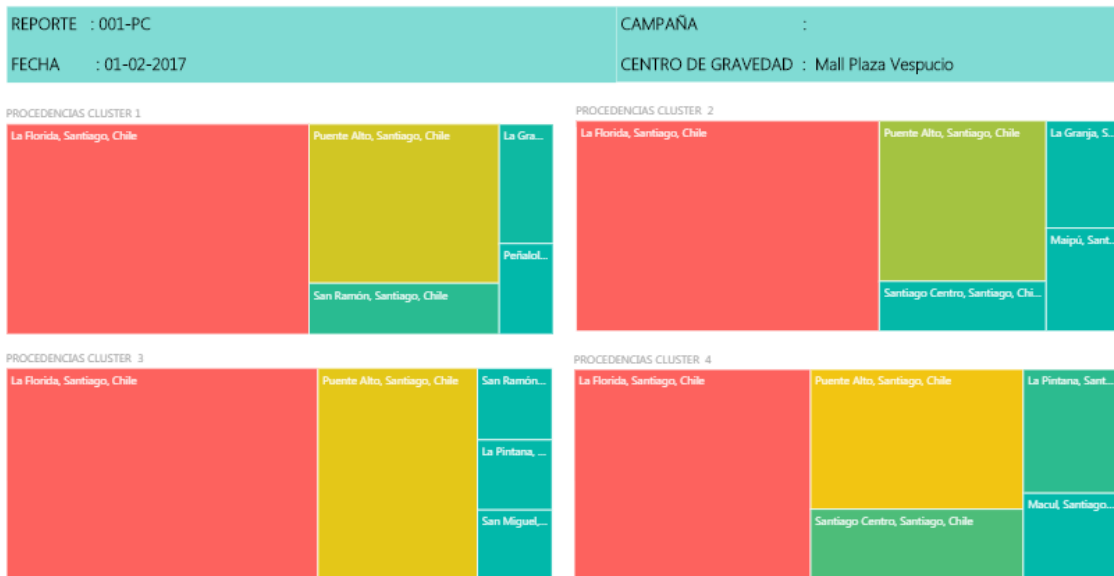


Figura 108. Reporte de ventas, procedencia

## ANÁLISIS DE PROCEDENCIA POR CLUSTER, CLUSTER 1

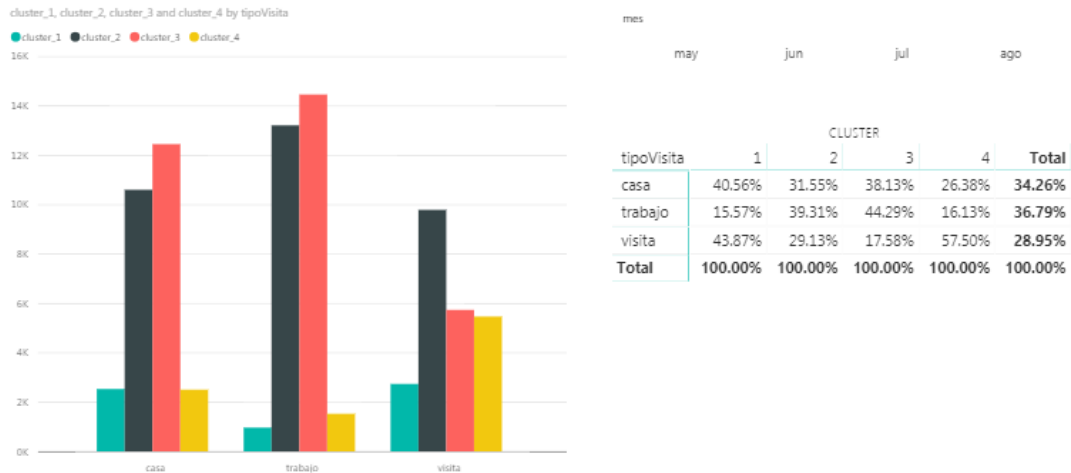
**REPORTE :** 002-PC **CAMPAÑA :**  
**FECHA :** 01-02-2017 **CENTRO DE GRAVEDAD :** Mall Plaza Vespucio



Figura 109. Reporte de ventas, mapa de procedencia (*cluster 1*)

## COMPOSICIÓN DE TIPO DE VISITAS

**REPORTE :** 001-TV **CAMPAÑA :**  
**FECHA :** 01-02-2017 **CENTRO DE GRAVEDAD :** Mall Plaza Vespucio



mes	CLUSTER				Total
	1	2	3	4	
may					
jun					
jul					
ago					
tipoVisita					
casa	40.56%	31.55%	38.13%	26.38%	34.26%
trabajo	15.57%	39.31%	44.29%	16.13%	36.79%
visita	43.87%	29.13%	17.58%	57.50%	28.95%
<b>Total</b>	<b>100.00%</b>	<b>100.00%</b>	<b>100.00%</b>	<b>100.00%</b>	<b>100.00%</b>

Figura 110. Reporte de ventas, tipo de visitas *cluster*



## RESUMEN DE RESULTADOS DE CAMPAÑA

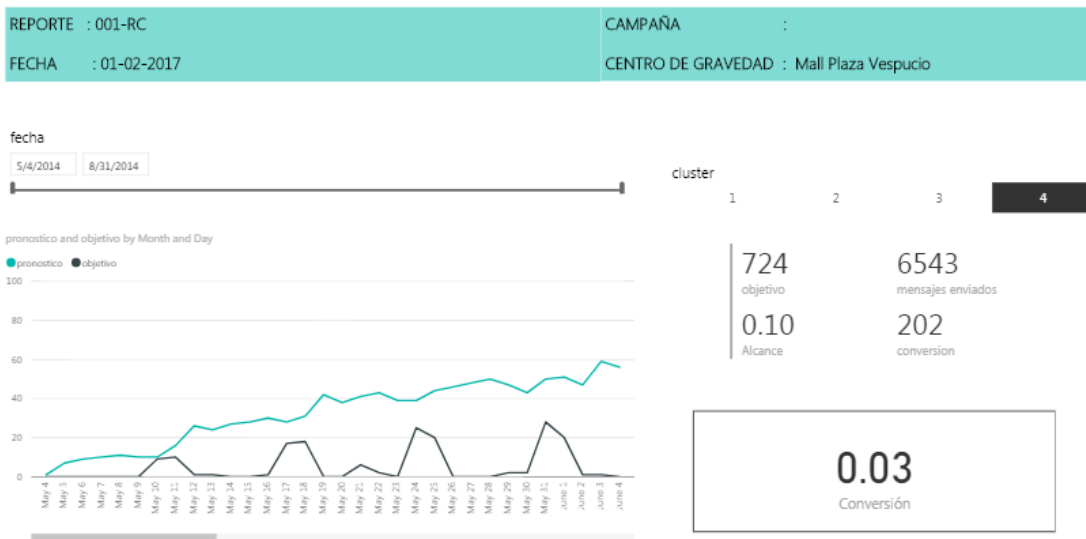


Figura 111. Reporte final, desempeño de campaña

## 6.2 Arquitectura Tecnológica

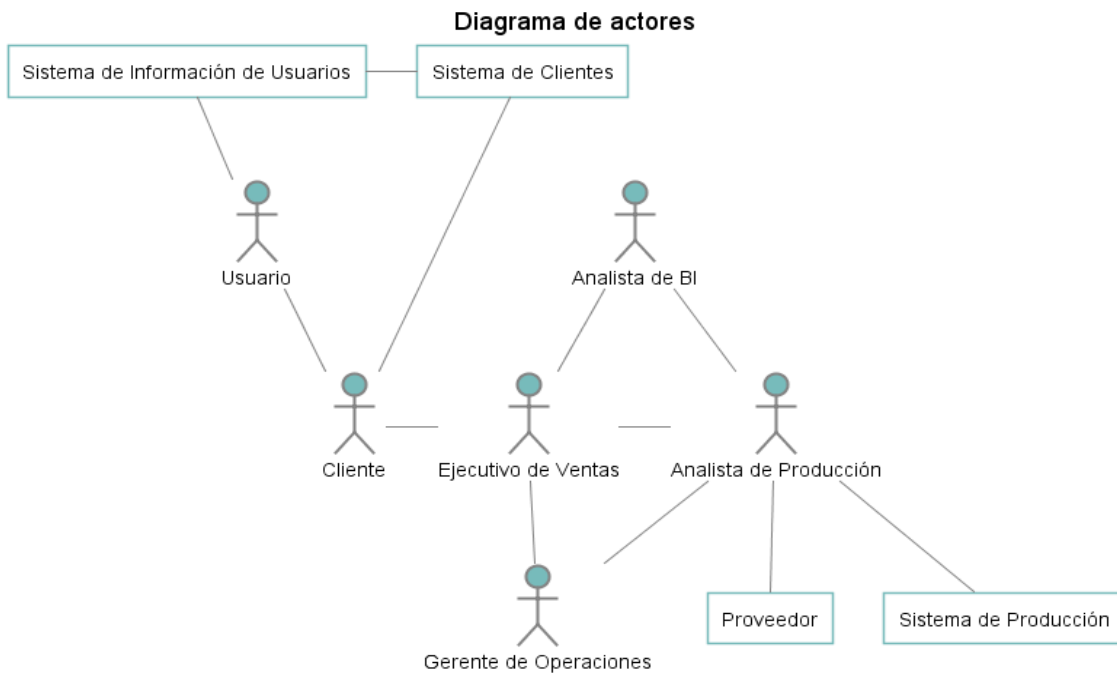


Figura 112. Diagrama de actores de aplicación computacional

El diagrama de actores muestra las siguientes interacciones:

El usuario se relaciona con el *SIU*, mediante el proceso de *Opt-in*, la edición de datos y el módulo de mensajes enviados. A su vez se relaciona con el cliente al generar compras con los códigos recibidos en los mensajes de descuento

El *SC* debe conectarse con el *SIU* para validar la existencia y propiedad de los códigos QR, esta interacción la inicia el cliente al entrar al *SC* por los procesos de validación de y registro de códigos QR.

Para llevar a cabo los contratos de servicios de campañas publicitarias, el cliente debe tener contacto con el ejecutivo de ventas. El ejecutivo de ventas deberá obtener aprobaciones tanto del analista de *BI* (que verifica si es factible la generación de segmentos que quiere el cliente con los atributos disponibles) y del analista de producción (que deberá verificar que haya capacidad para enviar los mensajes a un set de personas específicos).

El gerente de operaciones dará el arranque a la campaña, con una autorización final y solicitará los datos de contrato al ejecutivo de ventas, para que luego el analista de producción los cargue al sistema de producción, así como la información de los modelos y del conteo actualizado de mensajes enviados hasta la última campaña. El listado final de empleados se obtiene desde el sistema de producción y se envía al proveedor, quien luego entregará un listado de mensajes enviados.

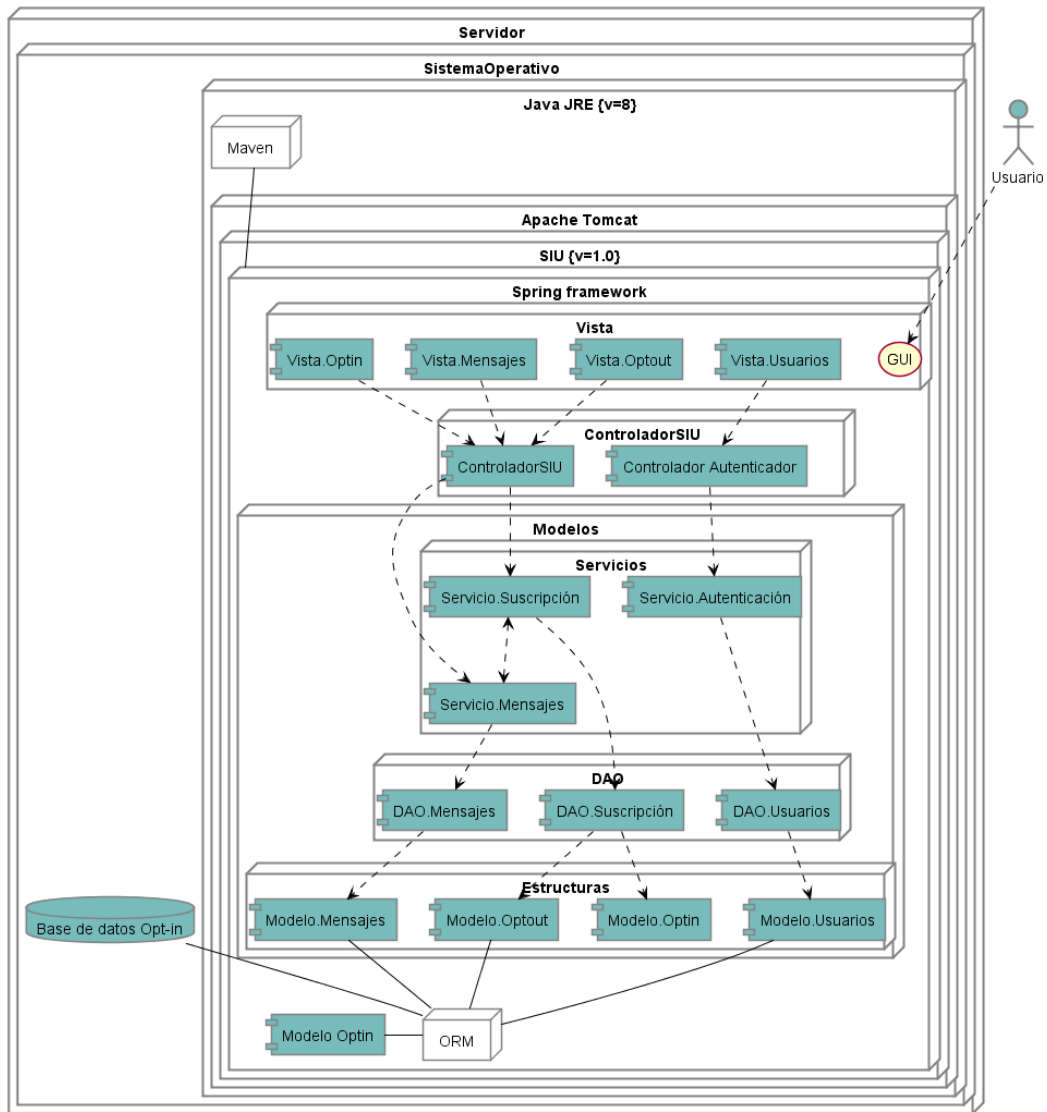


Figura 113. Diagrama de despliegue y componentes combinado

El diagrama de despliegue y componentes combinados muestra la relación de los softwares involucrados con la aplicación y la relación de sus componentes internas.

El entorno que rodea la aplicación está compuesto por un servidor, una base de datos, el ambiente de desarrollo *Java Runtime Environment* (JRE), el cual permite ejecutar aplicaciones en *JAVA*. Además se utiliza el operador *Maven* para la administración de dependencias y *Apache Tomcat* para ejecutar la aplicación.

El sistema se desarrollará en el *framework Spring*, el cual provee un módulo de seguridad para la autenticación de los usuarios (*Spring Security*).

La arquitectura contiene una capa de vistas, controladores y modelos. Para las vistas se utilizará la tecnología *JavaScript*, como es mencionado en las restricciones de diseño.

La capa controlador contiene dos controladores, el primero es un orquestador central (ControladorSIU) encargado de sirve de enlace entre las solicitudes de las vistas de mensajes, *Opt-in* y *Opt-out*. El segundo es un controlador encargado de la autenticación para el proceso de *Login*.

La capa modelos se subdivide en tres capas, las capas de Servicio, *Data Access Object* (DAO) y Estructura. La capa de estructura contiene la lógica estructural de las tablas de la base de datos, en este caso existe un modelo por cada vista. La capa DAO maneja genera las consultas necesarias para acceder a los datos sobre un modelo y la capa de servicios complementa entrega el ambiente transaccional que permite efectuar consultas a la base de datos y la conexión de los modelos con los controladores.

Esta arquitectura muestra una composición entre el servicio de mensajes y el servicio de suscripción, ya que se requiere acceso a la información de usuarios suscritos para ejecutar ciertas consultas de la tabla de mensajes, por ejemplo, extraer la lista de mensajes con capacidad de envío (que estén suscritos y no hayan recibido más de 10 mensajes de beneficio en el trimestre).

La base de datos del sistema, será una base de datos relacional y estará fuera de la aplicación.

### 6.3 Diseño de la Aplicación

#### 6.3.1 Casos de Uso

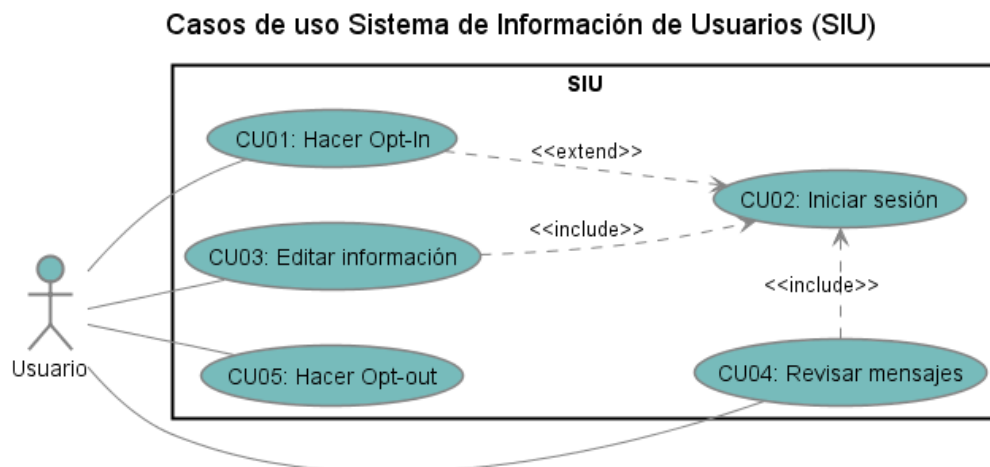


Figura 114. Diagrama de casos de uso para Sistema 1: Sistema de Información de Usuarios (SIU)

El diagrama de casos de usos para el Sistema de Información de Usuarios muestra las interacciones del usuario con el sistema en el cual podrá, hacer *Opt-in*, iniciar sesión, editar su información (Nombre, fecha de nacimiento, intereses, entre otros), revisar sus mensajes de beneficios y hacer *Opt-out*.

#### Casos de uso Sistema de Clientes (SC)

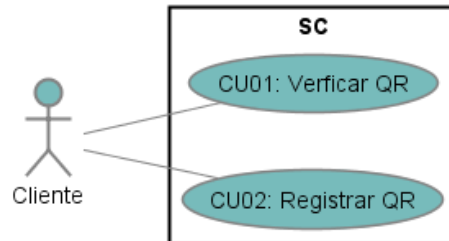


Figura 115. Diagrama de casos de uso para Sistema 2: Sistema de Clientes (SC)

#### Casos de uso Sistema de Producción (SP)

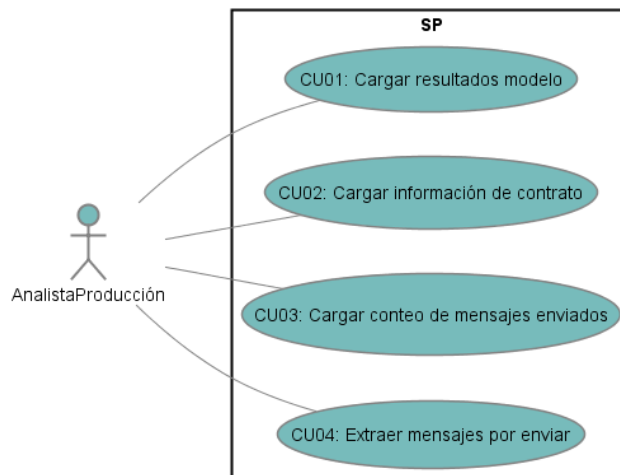


Figura 116. Diagrama de casos de uso para Sistema 3: Sistema de Producción (SP)

El Sistema de Información de Usuarios cuenta con las siguientes reglas de negocios:

- Sólo los usuarios de la TELCO asociada pueden hacer *Opt-in*
- Un usuario no puede ver la información de otro usuario
- Los datos de usuario solo pueden ser modificados por el usuario dueño de los datos
- Si el usuario cambia de teléfono deberá hacer un nuevo *Opt-in*

- Solo el analista puede extraer la información de *Opt-in*, por medio de un proceso *Extract, Transform and Load (ETL)*.
- Sólo se pueden eliminar usuarios por 2 vías: que el usuario haga *Opt-out*, que el administrador del sistema elimine al usuario.

Luego, se han detallado las siguientes lógicas de negocios según los casos de uso.

### ***Opt-in***

Se requiere de una validación de los datos ingresados en el formulario, las validaciones se hacen tanto en la vista como en el controlador. Cualquier error en la vista generará una notificación de error bajo el campo correspondiente, mientras que los errores encontrados en el controlador devolverán al usuario a la vista de *Opt-in* mostrando una notificación de error bajo el campo corrupto.

Las validaciones positivas en la vista se muestran un ticket en el campo.

- Teléfono: solicita el número de teléfono que recibirá los mensajes de beneficio. Deben ser escritos con el formato 1234567 (solo *integers*) y los precódigos +56 9 se encontrarán por defecto.
- Nombre: solicita ingresar el nombre del usuario. Puede no ser único y debe poseer al menos 2 y máximo 10 letras (string).
- Contraseña: debe contener entre 8 y 16 caracteres (char), permite símbolos como @#\$&%, pero no ^\*()[]!.
- Confirmación de contraseña: debe contener los mismos caracteres que el campo "Contraseña".
- Género: se tendrá dos *radio button* con las opciones "Hombre" y "Mujer". Debe marcarse una y sólo una.
- Fecha de nacimiento: las fechas deben seguir el patrón dd/mm/yyyy, ningún campo puede quedar vacío y no pueden haber fechas del tipo 30 de febrero.

El rango de años permitidos a ingresar debe mantener siempre a personas menores de 120 años y mayores de 14.

- Intereses: se mostrará un listado con *check boxes*. Requiere que al menos se marque 1 preferencia.

Una vez cliqueado el botón de envío del formulario, se realiza una segunda validación en el controlador, la cual acepta o rechaza el formulario. Si se rechaza entonces devolverá al usuario al formulario indicando el error y la zona del error. Si se acepta entonces, cambiará el estado del teléfono en la base de datos de *Opt-in* de

“NoClasificado” a “*Opt-in*”, encriptará la contraseña y guardará en la misma base de datos un listado de preferencias, edad, nombre y contraseña encriptada. Luego enviará al usuario a la vista de *Home*.

### ***Login***

Requiere del ingreso de teléfono y contraseña (establecidos en *Opt-in*). Se valida que la llave teléfono-contraseña sean correctos, lo cual se desarrolla en el controlador al hacer clic en el botón de *Login*. En el procedimiento validador se encripta la contraseña enviada en el *Login* y se compara con la contraseña encriptada de la base de datos de *Opt-in*. En caso de no estar correcta la llave, se despliega la página principal junto con un mensaje de error.

Si la llave es correcta, se verifica el estado del usuario. Si el usuario tiene el estado “*Opt-out*” (un usuario con estado “NoClasificado” no tiene llave teléfono-contraseña, pues esta se establece al hacer *Opt-in*), entonces se despliega vista principal con un mensaje que indique que el usuario no se encuentra suscrito. Si el usuario tiene el estado “*Opt-in*” entonces, se despliega la ventana de *Home* de usuarios, que corresponde a un listado con los mensajes de beneficios recibidos.

Una sesión con inactividad de 30 minutos es cerrada automáticamente.

### **Edición de perfil**

Se corrobora que haya una sesión iniciada. Si no hay sesión iniciada, entonces envía al usuario a la página principal de *Login*, con un mensaje de sesión caducada. Si la sesión está iniciada, se envía al usuario a la vista de edición, que corresponde a la misma de *Opt-in*.

Se pueden editar los campos mencionados en la lógica de *Opt-in* y se aplicarán las mismas validaciones de vista y controlador.

### ***Opt-out***

Se corrobora que haya una sesión iniciada. Si no hay sesión iniciada, entonces envía al usuario a la página principal de *Login* con un mensaje de sesión caducada. Si la sesión está iniciada, se envía al usuario a la vista de *Opt-out*.

En la vista de *Opt-out*, se pregunta si el usuario está seguro de desuscribirse del servicio de beneficios. Si el usuario declara estar seguro de hacer *Opt-out*, el controlador cambia el estado de la base de datos de *Opt-in* de “*Opt-in*” a “*Optout*” y guarda la base. Luego envía al usuario a la vista principal con un mensaje que confirme el *Opt-out*.

## **Extracción de datos**

Todos los días, durante la noche, se ejecuta una *ETL* de manera automática (*cron job*), la cual hace una copia de la base de datos a un servidor de desarrollo.



## CAPÍTULO 7: EVALUACIÓN DEL PROYECTO

---

Para el análisis retrospectivo se utiliza el modelo de clasificación que muestra mejor desempeño en el análisis general descrito en el punto 0. Sin embargo, este se entrenará con 14 de las 18 semanas de datos (base de entrenamiento y testeo), para lo cual se modifican los valores de las variables que describen al usuario “*vlun*”, “*vmar*”, “*vmie*”, “*vjue*”, “*vvie*”, *vsab*, “*vdom*”, “*vsem*” y “*vfds*”, de manera que muestren la información acumulada las primeras 14 semanas de los datos y se testeará como predictor de las semanas 15, 16, 17 y 18 de la base (base de prueba de concepto), estas corresponden a las semanas del mes de agosto de 2014.

Ambas bases contienen las variables mencionadas y descritas en el punto 5.2. El *split* de datos de testeo y entrenamiento sigue la mismas reglas especificadas en el punto 5.4.1. Las proporciones de observaciones de clase 1 (visita) y 0 (no visita) se mantienen para la base de entrenamiento, testeo y prueba de concepto.

Además se generaron los planes de negocios (plan de marketing y operaciones) para una empresa de envío de promociones móviles personalizadas y geolocalizadas, soportada por el modelo descrito anteriormente

### 7.1 Resultados de predicción obtenidos

El modelo se re-entrena con los datos de las primeras 14 semanas de la base de datos, que contiene 84.107 observaciones y 947 números únicos. Estos datos son balanceados con el método de *under sampling*, lo que entrega un base final que mantiene los números únicos, proporciona equitativamente las clases de la variable objetivo entregando un total de 24.410 observaciones.

El árbol de decisión (CART) selecciona las variables que describen al número “*vsem*”, “*vfds*”, “*vlun*”, “*vmar*”, “*vmie*”, “*vjue*”, “*vvie*”, *vsab*, “*vdom*”, “*vfechaMenosUno*”, “*vfechaMenosTres*”, “*vfechaMenosSiete*”, “*vsem*”*anaMenosDos*”, “*vsem*”*anaMenosTres*”, “*vsem*”*anaMenosCuatro*”, “*casatrabajo*”, “*cluster*”; y las variables que describen la fecha, “*lun*”, “*mar*”, “*mie*”, “*jue*”, “*vie*”, “*sab*”, “*dom*”, “*semanaMes*”, “*mes*” como predictores. Los resultados de entrenamiento son los siguientes:

- *Recall*: 87%
- *Precision*: 31%
- *Accuracy*: 58%
- $F_2$ : 0,64
- AUC: 0,77
- Costo mínimo: CLP\$561.360
- Umbral que minimiza el costo: 0,35

La prueba de concepto se realiza fijando el umbral en 0,35 y probando el modelo anterior para datos del mes de agosto de 2014, esta base contiene 11.492 observaciones y 676 números únicos en la base de prueba de concepto.

. Los resultados muestran un costo de CLP\$296.080

- Recall: 85%
- Precision: 36%
- Accuracy: 56%
- $F_2$ : 0,66
- AUC: 0,725

La matriz de confusión muestra la cantidad de aciertos para cada valor de la variable objetivo

	Respuesta del método	
	1	0
Clase 1 (visita)	2.562	<b>461</b>
Clase 0 (no visita)	<b>4.636</b>	3.833

Figura 117 . Matriz de confusión para prueba de concepto 1

En este caso, al igual que en las pruebas de modelos de predicción y clasificación mostradas en 5.4, se observa que la generación de campañas basadas en modelos de pronóstico produce menores costos por error y se alinea a los objetivos estratégicos de la empresa. En este sentido se destacan dos conclusiones:

- Un modelo que clasifica todas las observaciones como la clase mayoritaria (Clase 0), pierde sentido de negocios, pues considera que ningún usuario visitará el Centro de Gravedad, luego ninguna campaña se llevaría a cabo. Por otro lado genera costos por error cercanos a CLP\$725.000, lo cual es casi 2,5 veces más que el costo por error del modelo propuesto
- Un modelo que clasifica todas las observaciones como la clase minoritaria (clase 1), no aporta a la ejecución de campañas personalizadas en el tiempo de visitas a Centros de Gravedad. Por otro lado entrega bajos valores en sus medidas de desempeño y un costo por error de predicción aproximado de CLP\$340.000, es decir, 1,15 veces más costoso que la generación de campañas basadas en el modelo propuesto.

Además se analizó la respuesta de una selección aleatoria la que genera:

- Recall: 53%
- Precision: 52%

- Accuracy: 62%
- Costo: CLP\$754.440

Un modelo aleatorio entrega peores resultados que el modelo propuesto, por lo que aleja a la empresa de su objetivos de entrega de beneficios en el momento correcto. Por otro lado genera un costo por error mayor en 2,53 veces con respecto al costo de un modelo CART.

Dado lo anterior, se comprueba que esta prueba de concepto aporta valor al desarrollo de campañas geolocalizadas.

Un análisis detallado se realiza en una segunda prueba de concepto, considerando sólo las observaciones de los números que pertenecen al *cluster* 4 (personas que trabajan o son compradores de fin de semana en el CG), contiene 21.894 observaciones, 117 números únicos en la base de entrenamiento y testeo, 3.332 observaciones y 105 números únicos en la base de prueba de concepto.

### Entrenamiento

Los datos son balanceados con una combinación de *under* y *over sampling*, entregando un total de 58.874 observaciones.

El modelo selecciona las variables que describen al número “vsem”, “vfds”, “vlun”, “vmar”, “vmie”, “vjue”, “vvie”, vsab, “vdom”, “vfechaMenosUno”, “vfechaMenosDos”, “vfechaMenosTres”, “vfechaMenosSiete”, “vsemanaMenosDos”, “vsemanaMenosTres”, “vsemanaMenosCuatro”, “casatrabajo” y “cluster”; y las variables que describen la fecha “lun”, “mar”, “mie”, “jue”, “vie”, “sab”, “dom”, “mes” y “semanaMes”. Los resultados obtenidos son mejores comparados con las medidas de desempeño de la prueba anterior (con todos los segmentos):

- *Recall*: 93%
- *Precision*: 39%
- *Accuracy*: 52 %
- $F_2$ : 0,73
- *AUC*: 0,747
- Costo mínimo: CLP\$154.640
- Umbral que minimiza el costo: 0,35

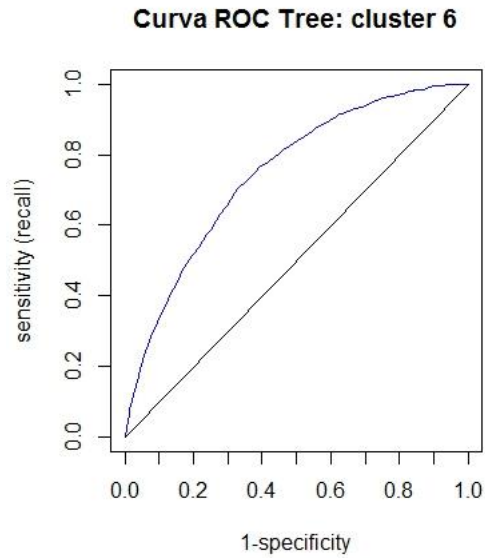


Figura 118 . Curva ROC para árbol de decisión en *cluster 4*

Aplicación de la prueba de concepto

Los resultados de la segunda prueba muestran un costo por error de predicción de CLP\$95.120, *recall* del 89% y *precision* de 40%. Otros resultados son

- *Accuracy*: 49%
- $F_2$ : 0,71
- AUC: 0,657

Además la matriz de confusión resultante sigue la tendencia de lo obtenido en la prueba de concepto 1.

	Respuesta del método	
	1	0
Clase 1 (visita)	1.053	<b>135</b>
Clase 0 (no visita)	<b>1.568</b>	576

Figura 119. Matriz de confusión para prueba de concepto 2

Se observa que los resultados para una predicción sobre el cluster de compradores y trabajadores de fin de semana no tienen grandes variaciones a la predicción general de visitas.

Se comprueba que, dado los costos calculados (CLP\$40 por error tipo 1 y CLP\$240 por error tipo 2), el modelo da preferencia a evitar los errores de tipo 2, en los cuales no se envía mensaje a personas que sí visitan el Centro de Gravedad. El modelo comete 10 veces más errores tipo 1 que errores tipo 2, mientras el costo del error tipo 1 es 6 veces mayor al costo del error tipo 2. Esto indica que los resultados cubren las necesidades de disminución de costos, lo cual es lo esperado para el negocio y lo preferible dado el balance de la variable objetivo.

## 7.2 Plan de marketing

El plan de marketing se desarrolla a partir de un análisis de los segmentos de clientes, búsqueda del segmento objetivo (mencionados en la sección 3.7.1) y posicionamiento frente a ello. Frente a sus usuarios esta empresa debe perfilarse como un servicio con alto nivel de geolocalización y personalización y baja dependencia del uso de la aplicación para la obtención de beneficios. Estos tres puntos se logran gracias al análisis y uso de Centros de Gravedad, la creación de un sistema que guarde la preferencia personales, un modelo que pronostique el comportamiento de visitas a Centros de Gravedad y el envío de promociones basadas en estos modelos. Este posicionamiento permite a la empresa desenvolverse en una rama del marketing móvil aún no explotada por la competencia.

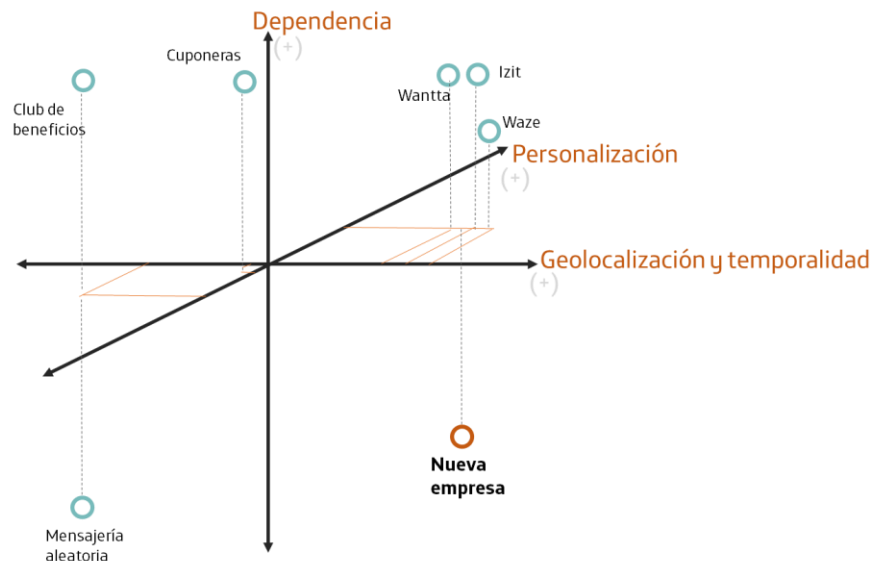


Figura 120. Posicionamiento de la empresa frente a usuarios

El marketing mix se desarrolla como una forma de concretar el posicionamiento deseado

### 7.2.1 Producto

La estrategia para entregar el producto de campañas móviles de beneficios personalizados y geolocalizados consta de un sistema gradual de captura de usuarios que recibirán beneficio mediante el envío de promociones por mensajería de texto, accesibles a través de una aplicación *web*, con sesiones personales, que despliega un código QR correspondiente al beneficio enviado.

Para el cliente, se considera la entrega de reportes de avance y desempeño de campañas, con el objetivo de visualizar la llegada a usuarios con las campañas efectuadas.

### 7.2.2 Precio

Se fijan precios base de envíos de mensajes, diferenciados según el tipo de usuario receptor y precios de conversión. Se entiende por conversión, el uso de la promoción enviada por parte del usuario en tiendas del cliente.

Tabla 55. Formula de precios

<b>Campaña</b>	<b>Fórmula de precio</b>
General	$(N^{\circ} \text{ mensajes enviados} * \text{PrecioBase}_{\text{base}}) + (N^{\circ} \text{ conversiones} * \text{Precio}_{\text{conversión}})$
Segmentado	$(N^{\circ} \text{ mensajes enviados} * \text{PrecioBase}_{\text{segmentado}}) + (N^{\circ} \text{ conversiones} * \text{Precio}_{\text{conversión}})$

Abordar un Centro de Gravedad ya entrega una segmentación de los usuarios, pues se sabe que son visitantes de la zona abarcadas por la(s) antena(s) que conforman un Centro de Gravedad, sin embargo una segmentación más fina aumenta el costo de procesamiento y por ende el precio de envío base de mensajes. Por otro lado, la componente variable del precio de una campaña depende de la cantidad de conversiones por campaña.

Los precios mencionados fueron establecidos según consultas a un experto en el tema.

Tabla 56. Precios

<b>Precio de envío base (PEB)</b>	CLP \$160
<b>Precio de envío segmentado (PES)</b>	CLP \$240
<b>Precio conversión (PC)</b>	CLP \$300

### **7.2.3 Plaza**

La venta de campañas de publicidad móvil personalizada y geolocalizada se realiza de forma presencial por ejecutivos de ventas que presentan el proyecto a empresas clientes. Una vez comprometido el servicio los contactos con el cliente pueden mantenerse de forma presencial o virtual.

### **7.2.4 Promoción**

La promoción dirigida a clientes se realiza a través de correos electrónicos, llamadas telefónicas y promociones en la página *web* de la empresa. Sin embargo también se contempla la captura de clientes de forma pasiva, en la cual los clientes interesados tendrán canales para solicitar los servicios de la empresa.

También es importante destacar la estrategia de promoción a usuarios, la cual se realiza con el proceso de *Opt-in*, donde se envían mensajes de texto con publicidad del servicio y opción a ingresar a la página de la empresa donde los usuarios podrán registrarse.

## **7.3 Plan de operaciones**

El plan de operaciones se realiza considerando las etapas de implementación y producción del proyecto desarrollado en esta memoria.

### **7.3.1 Implementación**

Considera una estrategia de inversión de 3 meses para las cuales se requieren recursos que permitan implementar los modelos creados, estos son:

- Un servidor de bases de datos con capacidad de 40 Tb que asegura el almacenamiento de la historia de visitas de los usuarios del proyecto
- Un servidor de producción capaz de soportar la ejecución de los modelos de *clustering* y pronóstico para datos de 1 o más Centros de Gravedad
- Computadores y licencias para todos los trabajadores asociados al proyecto
- Un dominio *web* asociado a la aplicación *web* del proyecto
- Una impresora dedicada al proyecto
- Horas de trabajo de un desarrollador *web*, un diseñador *web*, un desarrollador de modelos y un administrador de sistemas para el desarrollo de modelos

Por otro lado se ejecutarán las campañas de captación usuarios y captación de clientes 3 veces cada 6 meses.

### **7.3.2 Desarrollo**

La etapa de desarrollo considera las estrategias de producción, conexión y aumento de clientes.

La estrategia de producción requiere de operadores de producción, los cuales se encargarán de ejecutar los modelos para generar pronósticos y estarán a cargo de un jefe de operaciones. Un operador de producción tiene capacidad de ejecutar y controlar 4 campañas. Además requiere un administrador de sistema a cargo de la aplicación *web* y los sistemas internos que permiten la producción.

La estrategia de conexión requiere de la inversión en un servidor *web* que aloje la aplicación *web* del proyecto y el pago mensual de un correo electrónico institucional para la comunicación con clientes y usuarios.

Por último la estrategia de aumento de clientes requiere de ejecutivos de venta que gestionen los procesos de captación (tanto pasiva como activa) de clientes, los cuales estarán a cargo de un jefe de ventas. Se considera que 1 ejecutivo de ventas puede hacerse cargo de 4 campañas al mismo tiempo.

#### 7.4 Definición de Beneficios Y Costos

Los beneficios percibidos gracias al proyecto se calculan según el establecimiento de precios, la cantidad estimada de usuarios y el crecimiento estimado del mercado del marketing móvil en Chile.

Se calcula el promedio de la cantidad de usuarios hábiles, el ratio medio de usuarios de semana y fin de semana (como grupos segmentados tipo) y se estableció un ratio de conversión según información de expertos en el tema. Un usuario hábil es aquel que pertenece al porcentaje de personas que visita en promedio más de 16 veces un Centro de Gravedad, lo cual comprende entre el 1% a 3% de los usuarios activos de la TELCO asociada.

Tabla 57. Estimación de usuarios iniciales

<b>Cantidad de usuarios hábiles</b>	1191
<b>Ratio de usuarios de fin de semana</b>	20%
<b>Ratio de usuarios de semana</b>	16%
<b>Ratio de conversión (r)</b>	2%

Como se menciona en el punto 7.2, el modelo establecido para precios de campañas es una función con una tarifa fija, en base a los precios de envío de mensajes, y una tarifa flexible, según la cantidad de conversiones. Lo cual entrega los siguientes precios estimados por tipo de campaña:

Tabla 58. Estimación de precios por campaña (CLP)

<b>Campaña</b>	<b>Precio final aproximado (CLP)</b>
General	\$ 2.300.000
Fin de semana	\$ 2.000.000
Semana	\$ 1.500.000



Además, se realizó una estimación del crecimiento del mercado del marketing móvil en base a los ratios de crecimientos históricos chilenos y de estadounidenses

Tabla 59. Estimación de crecimiento del mercado en Chile

<b>2016</b>	24%
<b>2017</b>	21%
<b>2018</b>	19%

Luego los ingresos esperados, considerando una puesta en marcha que captura al primer cliente en el cuarto mes son de aproximadamente CLP \$412MM un horizonte de tiempo de 3 años.

Por otro lado, se obtienen los costos de la inversión fija y costos fijos y variables obtenidos a partir del desarrollo del plan de operaciones, que se traducen en costos fijos totales de CLP\$163MM (a 3 años) e inversión fija de CLP\$40MM.

Tabla 60. Costos fijos anuales del proyecto (CLP)

<b>Costos Fijos (CLP)</b>	<b>\$ 67.865.952</b>
Horas Hombre jefe de ventas (CLP)	\$ 21.600.000
Horas Hombre jefe de operaciones (CLP)	\$ 21.600.000
Horas Hombre Administrador de sistema (CLP)	\$ 24.000.000
Arriendo correo electrónico (CLP)	\$ 504.672
Arriendo servidor <i>web</i> (CLP)	\$ 161.280

Tabla 61. Inversión Fija del proyecto (CLP)

<b>Inversión Fija (CLP)</b>	<b>\$ 41.284.970</b>
Servidor de DB (CLP)	\$ 1.076.800
Servidor de producción (CLP)	\$ 4.038.000
Dominio (CLP)	\$ 20.170
Horas desarrollador <i>web</i> (CLP)	\$ 10.000.000
Horas diseñador <i>web</i> (freelance) (CLP)	\$ 3.000.000
Computadores (CLP)	\$ 8.000.000
Impresora (CLP)	\$ 350.000
Horas de desarrollador de modelos (CLP)	\$ 8.000.000
Envío de mensajes (CLP)	\$ 7.000.000
Campaña en medios a clientes (CLP)	\$ 4.000.000
Administrador de sistema para desarrollo (CLP)	\$ 7.000.000

Los costos variables consideran el aumento de operadores de producción y ejecutivos de ventas según el crecimiento en ventas de campañas del proyecto y los costos de envío de mensajes por campaña (mensajes operacionales), el valor presente de estos costos en un horizonte de 3 años es de alrededor de CLP\$140MM.

## 7.5 Flujo de Caja

Para el análisis de flujo de caja de la situación con y sin proyecto se utilizó una tasa de descuento del 12%, tasa usada para evaluar todos los proyectos de la compañía telefónica.

Para la situación sin proyecto se consideró la implementación del proyecto sin el análisis de frecuencia de visitas, con lo que se pierde la información de los usuarios hábiles para hacer predicciones según su historia, Esto incrementa los costos en inversión fija, ya que se realizarían campañas de *Opt-in* para el total de líneas activas de la TELCO (alrededor de 5.000.000 de líneas activas) y costos variables, pues los mensajes operacionales se envían a todos los usuarios de la TELCO, aumentando también los costos de inversión en servidores capaces de almacenar la información del total de líneas activas. Por otro lado se eliminan los costos de realización de modelos y se cancela la inversión variable en operadores de los modelos de producción. Además se contempla el uso de un crédito de CLP\$18MM pagado en 3 cuotas fijas con tasa de interés de 16,2%.

El detalle del flujo de caja sin proyecto se encuentra en la sección de anexos (ver 10.6).

Tanto para el estado con y sin proyecto se contempla la obtención del primer pago en el mes 4 desde la implementación y luego una tasa de crecimiento de un cliente nuevo cada 2 meses en el primer año. Para los años 2 y 3 se aplica la tasa de crecimiento del mercado chileno, estimada en la sección 7.4, al promedio de ventas de los últimos 6 meses de operación del año, donde se espera una estabilización de las ventas. Además se contempla el uso de un crédito de CLP\$18MM pagado en 2 cuotas fijas con tasa de interés de 10% anual. No se considera el uso de capital de trabajo debido a que el primer año de operación tiene un flujo de caja privado negativo.

Como supuesto, para el estado con proyecto, se considera que cada cliente compra en promedio dos campañas con un valor total aproximado de CLP\$3,5MM.

El flujo de caja efectivo es de aproximadamente CLP\$700MM, el signo de los flujos efectivos fluctúa entre valores positivos para el periodo 0, negativos para el año 1 y 2 positivo el año 3. Esto ocurre debido a que en el caso sin proyecto se incurre en una inversión inicial mucho mayor que en el caso con proyecto. En consecuencia existen 2 valores *TIR*, por lo que este indicador no se usa para validar el proyecto económicamente.

Tabla 62 . Flujo de caja puro anual en pesos chilenos (situación con proyecto)

<b>Año</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>Ingresos por ventas</b>	\$-	<b>\$73,500,000</b>	<b>\$210,000,000</b>	<b>\$252,000,000</b>
Ganancias/Pérdidas de Capital	\$-	\$-	\$-	\$-
<b>Costos Fijos</b>	\$-	<b>\$67,865,952</b>	<b>\$67,865,952</b>	<b>\$67,865,952</b>
<b>Costos Variables</b>	\$-	<b>\$35,640,360</b>	<b>\$71,589,600</b>	<b>\$71,795,520</b>
Depreciaciones legales	\$-	\$2,302,467	\$2,302,467	\$2,302,467
Pérdidas del ejercicio anterior	\$-	\$34,579,590	\$48,632,422	\$-
<b>Utilidad antes de impuestos</b>	\$-	<b>-\$66,888,368</b>	<b>\$19,609,560</b>	<b>\$110,036,061</b>
Impuesto de primera categoría	\$-	\$-	\$4,706,294	\$27,509,015
<b>Utilidad después de impuestos</b>	\$-	<b>-\$66,888,368</b>	<b>\$14,903,265</b>	<b>\$82,527,046</b>
<b>Flujo de Caja Operacional</b>	\$-	<b>-\$30,006,312</b>	<b>\$65,838,154</b>	<b>\$84,829,513</b>
<b>Inversión Fija</b>	<b>\$52,484,970</b>	<b>\$8,309,200</b>	<b>\$10,309,200</b>	<b>\$20,170</b>
Valor residual de los Activos	\$-	\$-	\$-	\$-
Capital de Trabajo	\$-	\$-	\$-	\$-
Recuperación del Capital de Trabajo	\$-	\$-	\$-	\$-
Préstamos	\$17,905,380	\$-	\$-	\$-
Amortizaciones	\$-	\$10,316,910	\$10,316,910	\$-
<b>Flujo de Capitales</b>	<b>-\$34,579,590</b>	<b>-\$18,626,110</b>	<b>-\$20,626,110</b>	<b>-\$20,170</b>
<b>Flujo de Caja Privado</b>	<b>-\$34,579,590</b>	<b>-\$48,632,422</b>	<b>\$45,212,044</b>	<b>\$84,809,343</b>
<b>Delta valor presente de los flujos</b>	<b>\$167,742,200</b>	<b>\$87,790,440</b>	<b>\$268,554,906</b>	<b>\$349,975,465</b>

## 7.6 Análisis de Sensibilidad

El análisis de sensibilidad se realiza modificando las variables precio y demanda del modelo económico expuesto anteriormente.

El primer análisis busca encontrar el precio mínimo a cobrar, considerando la misma demanda del flujo de caja (42, 120 y 144 campañas anuales para cada año) con el objetivo de hacer el VAN efectivo igual a cero. El resultado indica que el precio mínimo aproximado por campaña es de CLP\$844.000

El segundo análisis considera un precio promedio por campaña de CLP\$1.750.000, luego para obtener VAN efectivo igual a 0 son necesarias 46 campañas promedio anuales promedio, es decir 4 campañas mensuales. En particular para cada año se requieren:

Tabla 63 . Campañas necesarias con precio promedio

<b>Año</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>Total de campañas</b>	24	48	57

Por último, tercer análisis busca fijar la cantidad mínima de campañas a vender dado el precio promedio de la competencia, el cual es de CLP\$3.000.000. El resultado que anula el VAN efectivo es de 26 campañas promedio anuales, unas 2 campañas mensuales.

Tabla 64. Campañas necesarias con precio promedio de la competencia

<b>Año</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>Total de campañas</b>	26	24	29

Dado lo anterior se aprecia que es factible acercarse al precio de competencia para aumentar la rentabilidad del proyecto, esto es efectivo si se aumentan los precios de envío de mensajes (tanto base como segmentado).

## **CAPÍTULO 8: CONCLUSIONES**

---

### **8.1 Conclusiones sobre la factibilidad del proyecto**

Se observa que el mercado está encaminado a trabajar cada vez más con la publicidad online en medios móviles. La inversión en estos medios nos demuestra una tendencia clara que se ve corroborada con los movimientos de las empresas en el mundo [4]. En particular, Chile se muestra como un país favorable para el desarrollo del mercado, sus niveles de inversión en publicidad online van en aumento sostenido a pesar de que las inversiones totales del país han disminuido [2], la población aumenta el uso de teléfonos inteligentes con tasas crecientes y se perfila como uno de los países líderes en el uso de estos dispositivos.

La tecnología que permite soportar el manejo de datos existe, las formas de solventar el procesamiento por medios de paralelismos muestran que es posible entrar en el mercado del marketing móvil.

Las personas buscan publicidad dirigida y son cada vez más exigentes en los contenidos que reciben, consideran sus dispositivos móviles como artículos personales que llevan con ellos todo el día. El estudio de su movimiento podría no solo entregar patrones de movilidad, sino que ser la fuente de información necesaria para segmentar a la población y entregarles contenido relevante.

La evaluación económica muestra al proyecto como una fuente de ingresos creciente, de costos operacionales bajos e inversión media con reinversión en años posteriores. El beneficio del uso del análisis y modelado de datos, entrega valor y disminuye los valores de la inversión de forma importante.

Dado lo anterior se corrobora la factibilidad de la aplicación de un negocio de marketing móvil georreferenciado que utilice como insumo esencial de la cadena de valor, modelos de pronóstico del comportamiento de visitas a Centros de Gravedad.

### **8.2 Conclusiones sobre los riesgos del proyecto**

A pesar de lo anterior, se observan dos tipos de usuarios, las personas de adopción temprana a las nuevas tecnologías y aquellos que pueden sentirse inseguros con una aplicación que funciona en todo momento y en cualquier lugar. El último grupo es el más grande en la población y al que hay que atender con especial cuidado al momento de ofrecer un servicio que utiliza su información relativa como input

Aún existe el riesgo de entregar un servicio que dañe la sensación de privacidad de los usuarios, debido al uso de su ubicación relativa, por lo que el tratamiento de los datos debe ser extremadamente cuidadoso, controlado y ligado a las leyes que rigen a nuestro país. Además, la comunicación es fundamental, el establecimiento de las reglas del negocio que involucran la información de usuario deben ser comunicadas para entregar la sensación de seguridad

Los riesgos del tipo de datos mencionados en 1.8.1 son mitigables y permiten enfocar el negocio a centros de gravedad comercial, donde la predicción indicará si los usuarios llegan o no a la zona de atracción, permitiendo a los clientes de la empresa aumentar la fidelidad y obtener clientes de la competencia.

El aspecto organizacional es importante para una empresa que depende de la interacción de organizaciones. La visión estratégica de cada una de las partes debe estar alineada para generar acciones que permitan el desarrollo del *spin off*. El establecimiento de objetivos comunes y el entendimiento de los ciclos de negocio debe ser claro para las partes involucradas de forma de generar compromiso y armonía en la operación.

### 8.3 Conclusiones sobre la investigación realizada

La realización de un *benchmark* de algoritmos de *clustering* ha generado nuevo conocimiento. *DBSCAN* no es capaz de reconocer clusters en dos dimensiones para el tipo de datos entregados (datos de frecuencias), esto debido a que se tiene densidad variable en los datos. Por otro lado *k-means* entrega resultados efectivos tanto utilizando distancia euclidiana como similitud coseno. Esta última se ajusta mejor al tipo de datos debido a que transforma la información en coordenadas polares y aplica el algoritmo a los datos transformados, lo que permite generar *clusters* según el ángulo de similitud de dos puntos distribuidos en el espacio de observación. No se descarta que aplicar una nueva segmentación con otro criterio de distancia sobre los clusters generados, pueda explicar con más detalle el comportamiento de visitas de las personas.

Además se notó que el resultado óptimo entregado por el método del codo no necesariamente es aplicable al negocio, debido a la capacidad de manejo de una gran cantidad de *clusters*, y se identificaron 3 *clusters* como básicos dentro del estudio: el de visitas en semana, el de visitas en fin de semana, y el de visitas constante. La relación entre cantidad de visitas en días de semana versus cantidad de visitas en fin de semana es explicable de forma lineal.

Se observó la efectividad de la segmentación previa en la respuesta de los métodos de predicción, ya que la variable resultó lo suficientemente significativa para todos los modelos probados.

Los modelos de predicción entregan resultados similares para ciertos indicadores, sin embargo el costo de estos es alto para aquellos que cometen más errores de tipo 2. Lo anterior no se ve reflejado en la medida de *accuracy* por sí sola, a pesar de haber balanceado las clases de las muestras trabajadas. Se concluye que  $F_2$  entrega mejores resultados globales para métodos aplicados a los datos del *data warehouse* construido, y la observación *Recall* y *Precision* acercan al investigador a la evidencia de la efectividad de los modelos.

Se notó que la relación de estos últimos indicadores ponderados por *accuracy* ( $precision * (1 - accuracy) / recall$ ) es clave para explicar los costos de error de los modelos, donde valores más altos se relacionan con mayores costos del modelo.

Se comprobó la factibilidad de replicación tanto de los modelos de *clustering*, como de la generación de *data warehouses* y métodos de pronóstico (clasificación y/o predicción), ejecutables en tiempos razonables para la producción (menores a 8 horas en total).

Se confirma la posibilidad de pronóstico de comportamientos de visitas que soporten un negocio de campañas promocionales móviles, geolocalizadas.

#### **8.4 Trabajo futuro**

Nuevos estudios pueden realizarse a partir de lo presentado en esta tesis, obteniendo más información de los usuarios, lo que puede realizarse con varios métodos no excluyentes:

1. Completando la historia del comportamiento de llamadas
2. Trabajando con registro del uso de datos móviles
3. Agregando información del proceso de suscripción al servicio, lo que requiere la implementación del servicio en fase de marcha blanca.
4. Utilizando la información de transacciones realizadas

Las dos primeras permiten mejorar el rastro de la huella de los usuarios de la aplicación y con ello los resultados predictivos de los modelos presentados. El tercero y cuarto mejorarán la creación de *clusters* y personalización de los beneficiarios del servicio de marketing móvil, permitiendo predicciones con mejor segmentación, la generación de nuevas estrategia de promoción de productos y servicios, y con ello la transformación de esta empresa en una cada vez más interesante para posibles clientes.

Se espera, en primera instancia, que la predicción mejore para trabajadores de cada Centro de Gravedad debido a que la huella de visitas será mucho más clara para ellos. Paralelamente, para visitantes, se cree que la huella mejorará en la medida en que el entrenamiento con uso de datos móviles sea factible. Dado lo anterior, no es descartable generar modelo de predicción diferenciados para trabajadores y visitantes, debido a que los parámetros podrían ser diferenciados en cada caso.

Además será posible el análisis de comportamientos en feriados y de periodicidad de compras en fechas importantes como día de la madre, fiestas patrias, navidad o fechas recurrentes de cada usuario como cumpleaños, aniversarios, entre otros.

Por otro lado, ya sea con los datos actuales, o con información extraordinaria, se proponen estudios de conglomerados de Centros de Gravedad, para lograr una predicción más fina de las visitas a Centros de Gravedad, los que abarcan dos aspectos importantes:

- El aumento de datos históricos de comportamientos frecuencia de visitas a zonas de compra
- La caracterización geográfica de la movilidad del usuario, la que puede ser analizada temporalmente según días y horarios

En estos casos se espera encontrar comportamientos de visitas altas en Centros de Gravedad cercanos a sectores de trabajo en días de semana, bajando durante los fines de semana momentos en los cuales las visitas deberían aumentar en zonas relativamente cercanas a los hogares de los usuarios.

Otras alternativas de mejora vienen de la optimización del set de beneficios a entregar, gracias a la información de transacciones de clientes tipo, con la cual es factible agregar sistemas de recomendación de nuevos productos y categorías de productos a los usuarios de la empresa. Lo anterior es factible gracias a la generación de comunidades de productos [21]

Finalmente se propone el análisis de *clusters* donde usuarios puedan pertenecer a más de un segmento, en base a características como categorías de productos de interés, cupones usados, zonas geográficas frecuentemente visitas, entre otras. Con ello realizar análisis de comunidades como las efectuadas en redes sociales, donde se enfrenta el problema de comunidades superpuestas, considerando el punto de vista de que las personas tienen múltiples intereses y que las comunidades deben reflejarlos [22].



## CAPÍTULO 9: BIBLIOGRAFÍA

---

- [1] Barros, O. Rediseño de Procesos de Negocios mediante el Uso de Patrones, Comunicaciones Noreste Ltda., 2003.
- [2] Asociación Chile de Agencias Publicitarias (2016) . Inversión Publicitaria en Medios 2015  
Recuperado de [http://www.achap.cl/wp-content/uploads/2016/05/Informe\\_Inversion\\_Publicitaria2015.pdf](http://www.achap.cl/wp-content/uploads/2016/05/Informe_Inversion_Publicitaria2015.pdf)  
Consultado en julio de 2016.
- [3] Marketing Directo. (). Recuperado de <http://www.marketingdirecto.com/especiales/mobile-marketing-blog/la-inversion-en-publicidad-movil-duplicara-la-desktop-en-2017/>  
Consultado en septiembre de 2015.
- [4] US Digital Display Ad Spending to Surpass Search Ad Spending in 2016 - eMarketer. (n.d.).  
Recuperado de <http://www.emarketer.com/Article/US-Digital-Display-Ad-Spending-Surpass-Search-Ad-Spending-2016/1013442>.  
Consultado en julio de 2016.
- [5] Starterdaily. (2014/08/18).<http://starterdaily.com/estudios/2014/08/18/latinoamerica-es-la-region-en-que-mas-compras-se-realizan-via-mobile/>  
Consultado en septiembre de 2015.
- [6] Think with Google. (2014/07). <https://www.thinkwithgoogle.com/intl/es-419/research-studies/the-new-multi-screen-world-chile-study-2014.html>  
Consultado en septiembre de 2015.
- [7] Starterdaily. (2014/04/07). Recuperado de <http://starterdaily.com/mobile/2014/04/07/el-m-commerce-esta-cambiando-el-habito-de-consumo-de-los-chilenos/>  
Consultado en septiembre de 2015.
- [8] Bigdata. (2014/10/13). <http://www.bigdata.cl/tendencias-de-marketing-digital-para-2015/>  
Consultado en octubre de 2015.
- [9] Puromarketing. (2013). Recuperado de <http://www.puromarketing.com/21/15555/publicidad-movil-gana-goleada-desktop.html>  
Consultado en octubre de 2015.
- [10] Emol. (2014/01/15). Recuperado de <http://www.emol.com/noticias/tecnologia/2014/01/15/639815/uso-de-aplicaciones-moviles-aumento-un-115-durante-el-ano-pasado.html>  
Consultado en agosto de 2015.

- [11] Partridge, K., Golle, P., 2008. On using existing time-use study data for ubiquitous computing applications. In: Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp 08), pp. 144–153.
- [12] Martin Ester, H.-P. K. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.
- [13] Mihael Ankerst, M. M.-P. (s.f.). OPTICS: Ordering Points To Identify the Clustering Structure.
- [14] Hosmer, D. W. and Lemeshow, S. (2000) Introduction to the Logistic Regression Model, in Applied Logistic Regression, Second Edition, John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0471722146.ch1.
- [15] Lewis, R. (2000). An Introduction to Classification and Regression Tree (CART) Analysis.
- [16] Gobierno Regional Metropolitano de Santiago. Recuperado de <https://www.gobiernosantiago.cl/.../INDICADORES-COMUNALES-CASEN-RMS-2013>. Consultado en mayo de 2017.
- [17] Contreras-Piña, C., & Ríos, S. A. (2016). An empirical comparison of latent semantic models for applications in industry. *Neurocomputing*, 179, 176-185.
- [18] Hothorn, T., Hornik, K., & Zeileis, A.(2006). Unbiased Recursive Partitioning: A Conditional Inference Framework.
- [19] Chih-Chung, C., Chinh-Jen L., & Chih-Wei H. (2003). A Practical Guide to Support Vector Classification. Department of Computer Science and Information Engineering, National Taiwan University, Taiwan.
- [20] Muñoz, R., Ríos, S., Rodas, R., & Contreras, C.Land Use Detection with cell phone data using Topic Models: Case Santiago, Chile. Department of Industrial Engineering, 4 Universidad de Chile.
- [21] Ríos, S. A., & Videla–Cavieres, I. F. (2014). Generating groups of products using graph mining techniques. *Procedia Computer Science*, 35, 730-738.
- [22] Ríos, S. A., & Muñoz, R. (2014). Content patterns in topic-based overlapping communities. *The Scientific World Journal*, 2014.

## CAPÍTULO 10: ANEXOS

### 10.1 Anexo 1: KPI's según objetivos de negocios

Tabla 65 . KPI's según objetivos de negocios

Perspectiva	Objetivo	KPI	Unidad	Meta
Financiera	Mejorar rentabilidad de los clientes	Rentabilidad por cliente	Ingreso por cliente	3*costo
		Frecuencia de compra del cliente	$\frac{\text{Número campañas por semes}}{\text{clientes}}$	$\geq 6$
Clientes	Ampliar gama de productos	Número de nuevos productos atractivos anuales	$\frac{\text{Número productos}}{\text{año}}$	$\geq 2$
	Aumentar la conversión de los clientes	Número de transacciones mensuales derivadas de una campaña de marketing	$\frac{\text{Número códigos canjeado}}{\text{total códigos enviados}}$	$\geq 10\%$
	Aumentar el pool de usuarios para el cliente	% de clientes que se les ofrece más de 3 segmentos de usuarios	%	$\geq 85\%$
		% de segmentos de usuarios ofrecidos a clientes con más de 100 usuarios	%	90%
	Aumentar ciclo de vida del cliente	Número promedio de meses al año	$\frac{\text{Número meses}}{\text{cliente * año}}$	6

		que se está realizando una campaña por cliente		
		Porcentaje de clientes activos que han realizado alguna campaña en un periodo mayor a un año	%	70%
	Aumentar pool de clientes	% de aumento del número de clientes trimestrales	%	>=10 %
Procesos	Asegurar la efectividad del envío de mensajes	$\frac{\text{Número de links}}{\text{total de mensajes}}$	%	>= 5%
	Mejoramiento continuo de los modelos	Precisión del modelo	%	70%
		Recall del modelo	%	70%
	Potenciar el análisis de modelos	Número de nuevos tipos de análisis anuales	Número tipos de nuevos análisis	>=2
	mejorar campaña pre <i>Opt-in</i>	% de aumento del número de usuarios trimestrales	%	>=10 %
	Mejorar el proceso de atención a	Número de reclamos debido a la	Número reclamos/trimestre	<=10

	usuarios	atención a usuarios por trimestre		
		Porcentaje de reclamos de usuarios que se demoran más de un día en ser resueltos	%	<=15
	Asegurar la efectividad de la comunicación de privacidad de datos de usuarios	% de usuarios encuestados que indican conocer las privacidad de sus datos	%	>=50
		Número máximo de <i>clics</i> en sitio <i>web</i> para llegar a información de privacidad de datos	Número de <i>clics</i>	<=1
	Mejorar el proceso de relación con los clientes	Número de reclamos debido a la relación con los clientes por trimestre	$100 * \frac{N^{\circ} \text{ reclamos de clientes por trimestre}}{\text{Total de reclamos en el trimestre}}$	<=10%
		% de reclamos de relación con el cliente respondidos en menos de 1 semana	%	>=80%
	Mejorar el proceso de captación de	% de contactos efectivos	%	>=30%

	clientes	(Activa)		
		% de tiempo menor a 13d que pasa entre el contacto y firma del contrato	%	>= 95%
		% de solicitudes que son respondidas en menos de 2 días (Pasiva)	%	>= 95%
		% de aumento de solicitudes con respecto al trimestre anterior	%	>= 5%
Aprendizaje	Asegurar poblamiento y actualización de base de datos	Lead time de actualización de datos	días hábiles	<= 5dh
	mantener tecnología de punta para almacenar y procesar datos	Número de contingencias por trimestre debido a falta de tecnología de punta	Número	<= 7
	Mantener personal capacitado	% de trabajadores que recibieron capacitación en el último año en su área de trabajo	%	>= 60%

	Mejorar los procesos de selección del personal	Tasa de rotación de personal	%	
		Porcentaje de cargos que son cubiertos en menos de 3 meses	%	>= 80%
		Mantener perfiles actualizados	% de perfiles de cargo actualizados	>=95 %

## 10.2 Anexo 2: Diseño de procesos de la cadena de valor en BPMN

### 10.2.1 Captación de usuarios

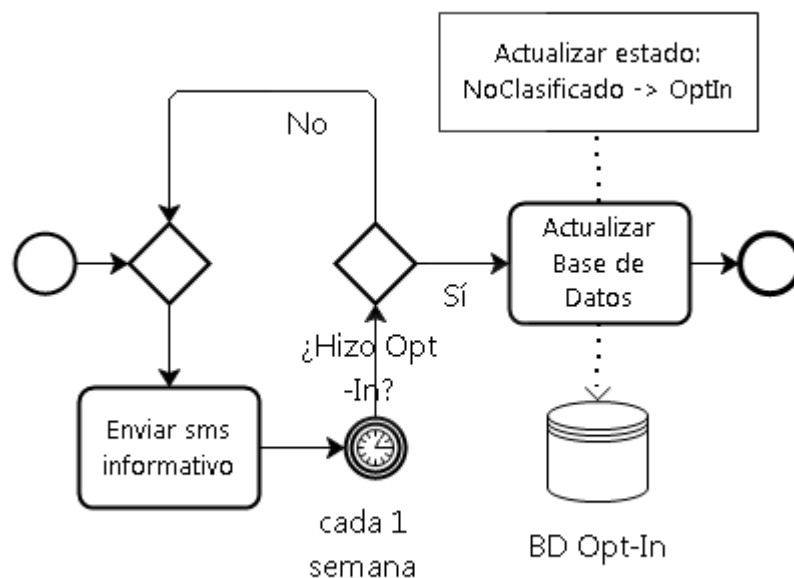







Figura 121. Proceso Captación de usuarios. En notación BPMN


 Evento de inicialización: el subproceso comienza junto con el inicio de campañas de *Opt-in* (captación de usuarios)

 Enviar SMS informativo: el mensaje de texto invitará a los clientes con línea activa del aliado estratégico a cambiar de estado “No clasificado” a “*Opt-in*”.

 Cada 1 semana: cada semana se chequeará el estado de cada cliente con línea activa del aliado estratégico.

 ¿Hizo *Opt-in*?: si el cliente con línea activa del aliado estratégico no ha pasado al estado “*Opt-in*”, entonces se volverán a enviar mensajes por diferentes canales

 Actualizar Base de Datos: si el cliente con línea activa del aliado estratégico pasó al estado “*Opt-in*”, entonces se actualizará la base de datos *Opt-in*.

 Evento de finalización: luego el proceso termina para el nuevo usuario registrado.



## 10.2.2 Reportes de avance

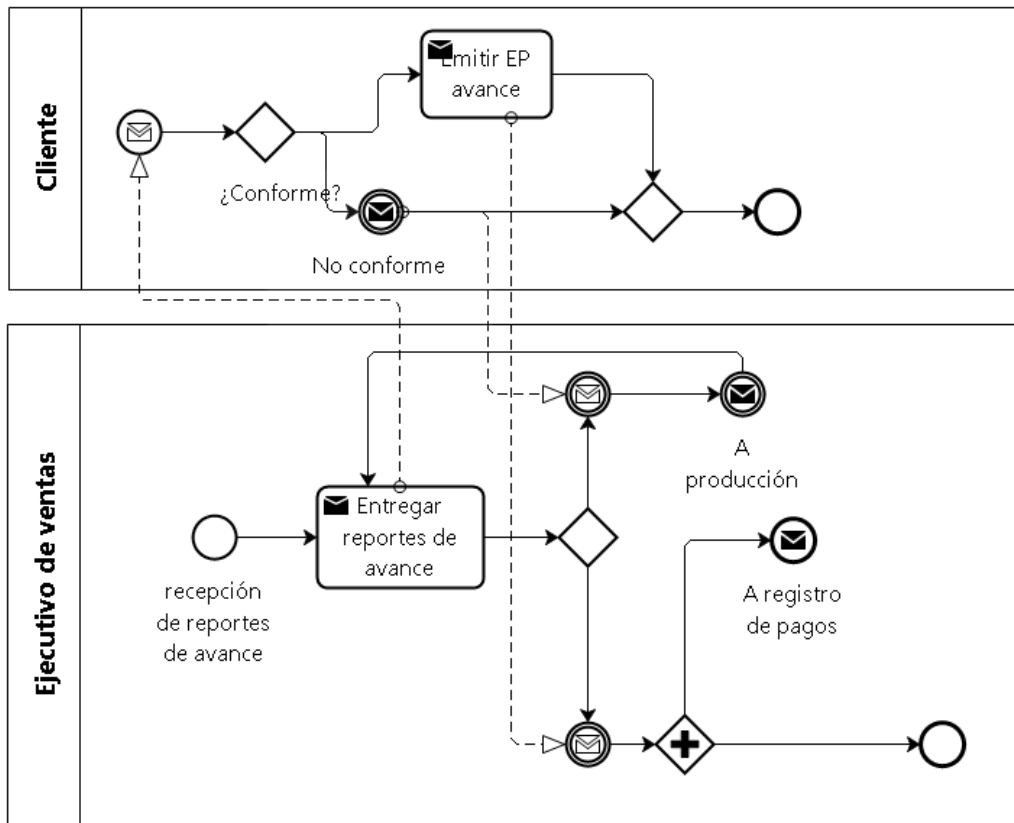


Figura 122. Proceso de Reportes de avances. En notación BPMN

● recepción de reportes de avance: el proceso se inicia cada vez que se recibe un Reporte de avance, con sms enviados.

✉ Entregar reportes de avance: el ejecutivo de ventas entrega los Reportes de avance al cliente.

✉ Si se recibe el estado de pago desde el cliente se desencadenan dos flujos:

✉ A registro de pagos: se entrega el estado de pago al proceso de registro de pagos.

● Finaliza el proceso para volver a ejecutarse o bien continuar con su proceso madre

 A producción: si el cliente no está conforme con lo entregado en el Reporte de avance, en contraste con el contrato y Reporte de inicio de proyecto, se deriva el caso a producción

### 10.2.3 Registro de pagos

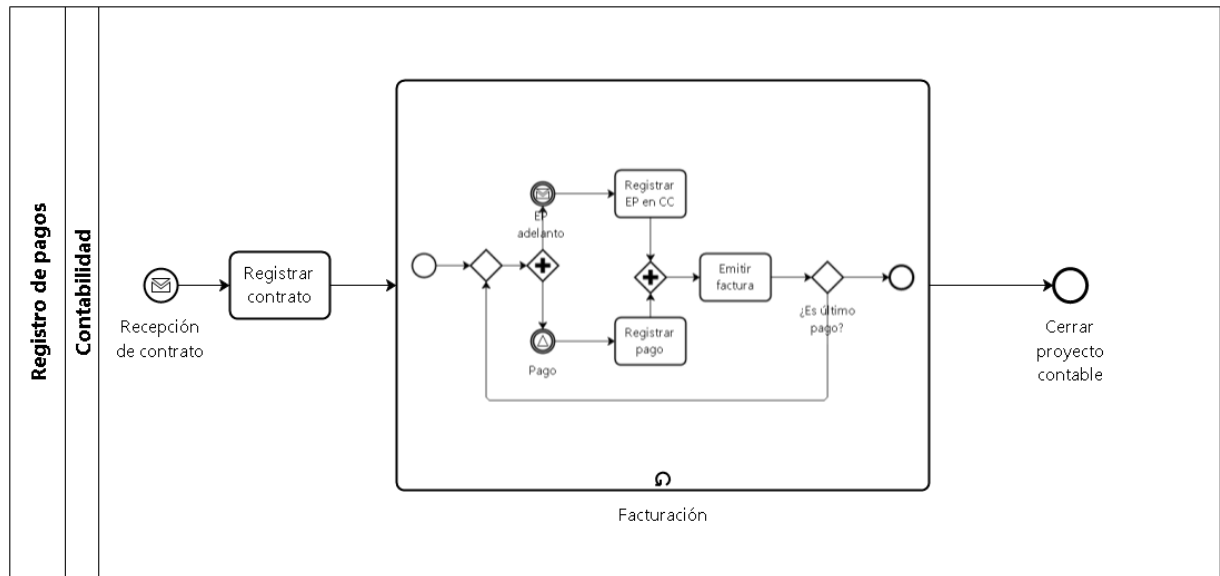






Figura 123. Proceso de Registro de pagos. En notación BPMN

 Recepción de contrato: el proceso se inicia con la recepción de contrato en contabilidad desde el proceso de Captación de clientes

 Registrar contrato: el contrato se registra en el sistema contable de la empresa asociado a su Centro de Costo.

 Facturación: se ejecuta el proceso de facturación para estado de pago, hasta que se recibe el estado de pago final.

 Cerrar proyecto contable: luego de la facturación final se cierra contablemente el proyecto

## 10.2.4 Retroalimentación de estrategia

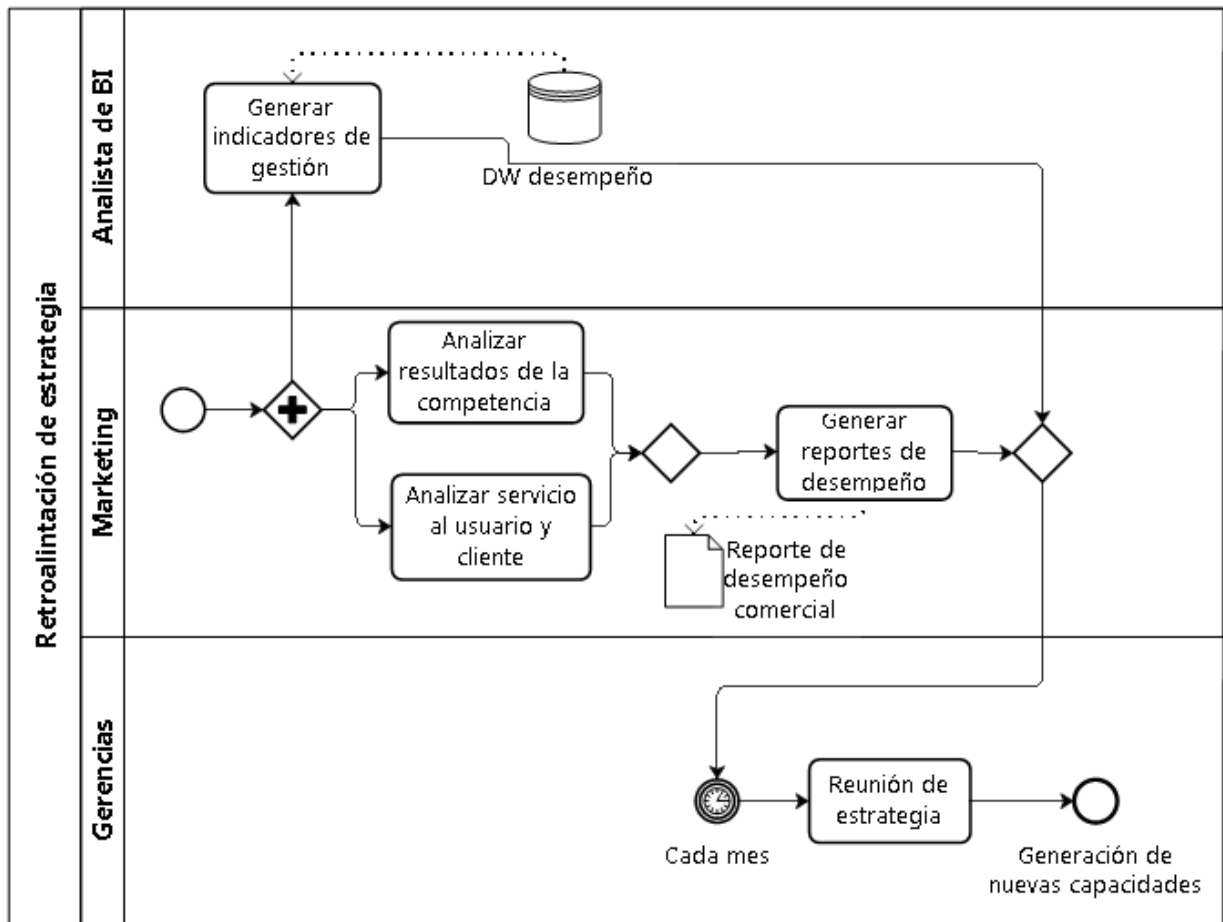


Figura 124. Proceso de Retroalimentación de estrategia. En notación BPMN

- Evento de inicialización: Este proceso se inicia una vez al mes
- Generar indicadores de gestión: el analista de BI, usando los datos del *DW* desempeño, podrá obtener indicadores de los resultados históricos de las campañas, tasas de suscripción y desuscripción de clientes, y cantidad de campañas generadas en el periodo de evaluación.
- Analizar resultados de la competencia: el área de Marketing deberá analizar los resultados comerciales de la competencia, poniendo atención en la cantidad de clientes, usuarios, *pricing* y satisfacción de clientes y usuarios.
- Analizar servicio al usuario y cliente: se deberá hacer un análisis interno de la satisfacción de usuarios y clientes, promesas cumplidas, calidad de la atención, cantidad de reclamos, entre otros.

□ Generar reportes de desempeño: con la información obtenida en las tareas anteriores, se desarrollará un reporte de desempeño comercial.

🕒 □ Reunión de estrategia: Con los reportes generados por el área de BI y Marketing, se realizará una reunión entre gerencias, donde se verificará el cumplimiento de los objetivos estratégicos, se modificarán los objetivos obsoletos y se crearán iniciativas para lograr los aún no resueltos.

🔴 Generación de nuevas capacidades: lo anteriores debe ser instaurado en la organización para la generación de nuevas capacidades.

### 10.2.5 Retroalimentación operacional

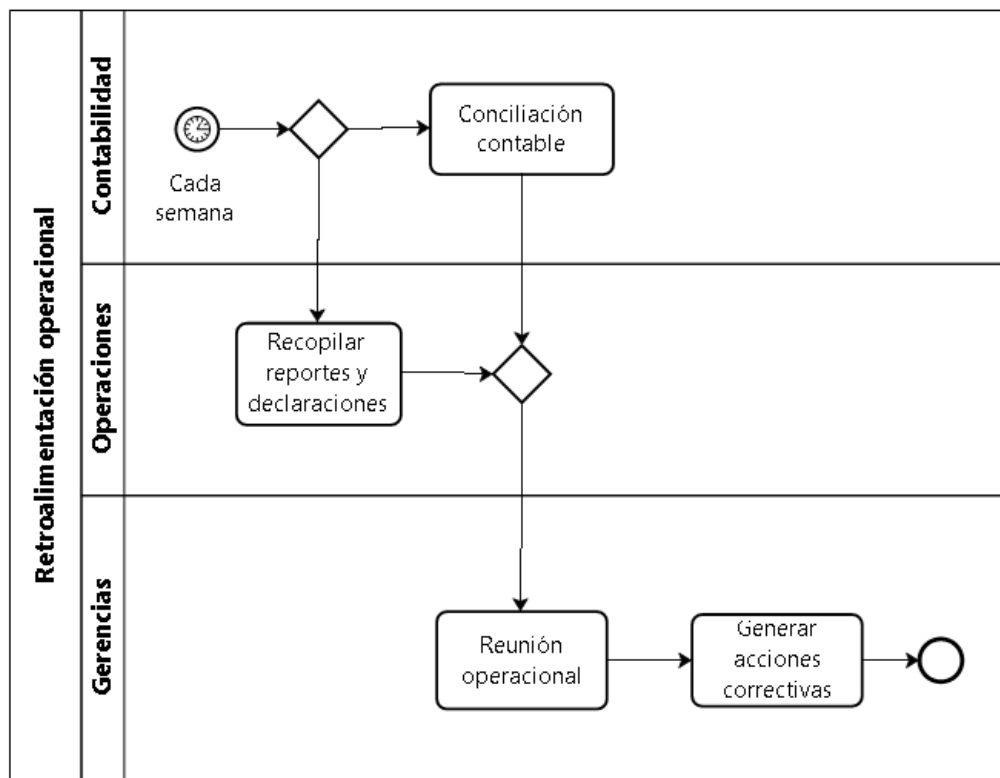


Figura 125. Proceso de Retroalimentación operacional. En notación BPMN

🕒 Cada semana: el evento se inicia semanalmente

□ Conciliación contable: el área de contabilidad deberá conciliar los estados de pago recibidos y las facturas emitidas, en concordancia con lo estipulado en los contratos de campañas con los clientes

Recopilar reportes y declaraciones de conformidad: el área de operaciones recopilará los reportes de avance, cantidad de reportes rechazados, motivos, estados de avances de las campañas, entre otros

Reunión operacional: las gerencias realizarán una reunión del desempeño operacional de la empresa

Generar acciones correctivas: establecerán acciones correctivas de avance, pagos, cobros, relación con clientes y usuarios, entre otros.

### 10.3 Anexo 4: Estudio relacional para *OPTICS*

Gráficos del estudio relacional de las variables *EPS\_cl*, *MinPts*, *ruido*, *num\_clusters*, *valoresDU*, para la Muestra B.

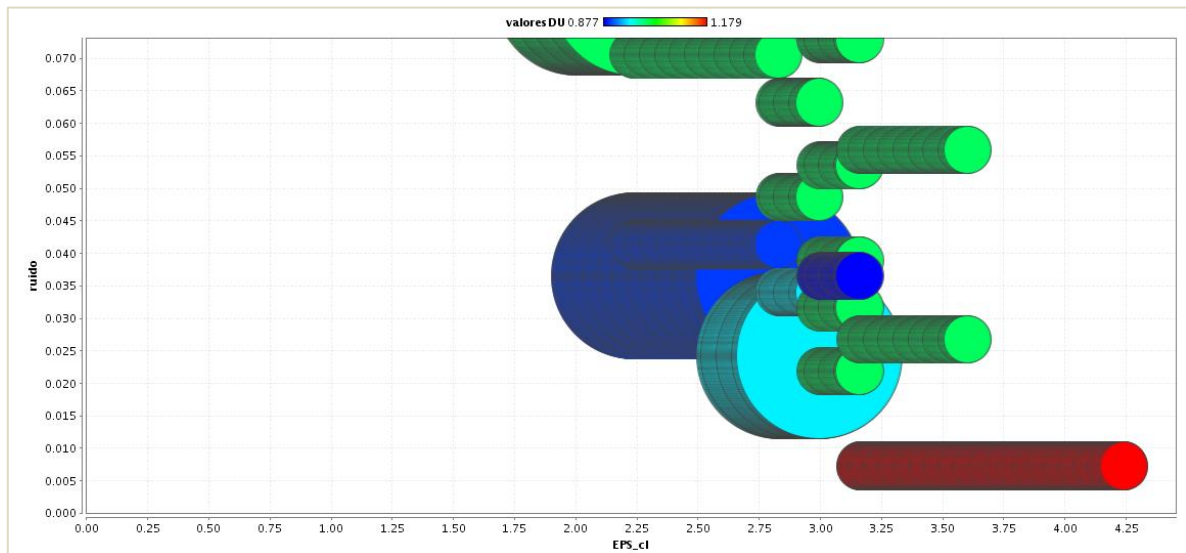


Figura 126. Relación de variables para Muestra B. x = *EPS\_cl*, y = *ruido*, size = *num\_clusters*, color = *valoresDU*

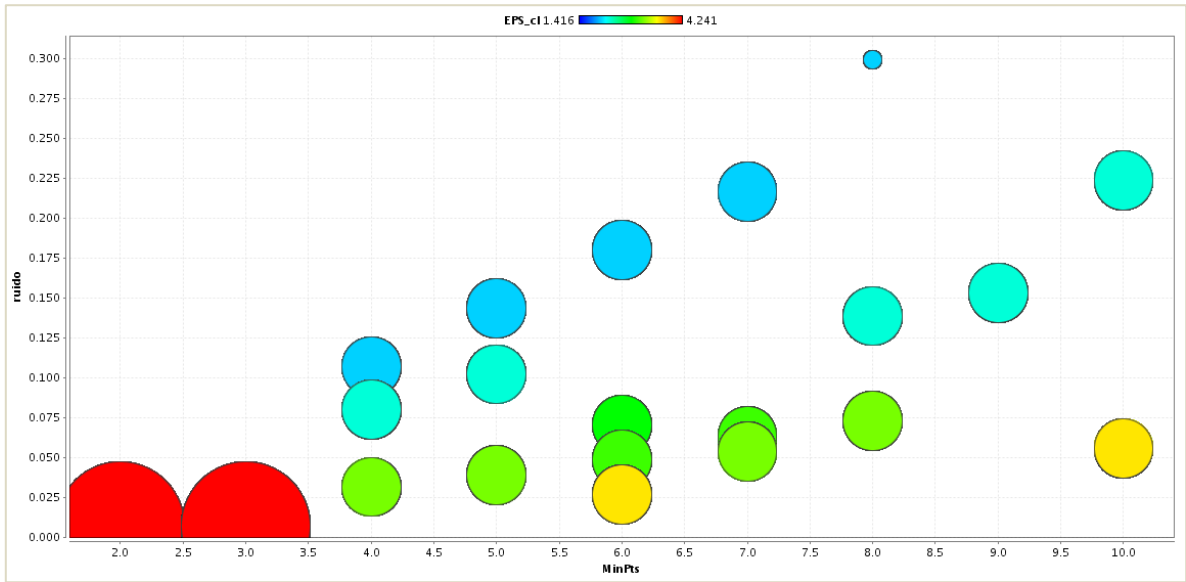


Figura 127. Relación de variables para Muestra B. x = MinPts, y = ruido, size = valoresDU, color = EPS\_cl.

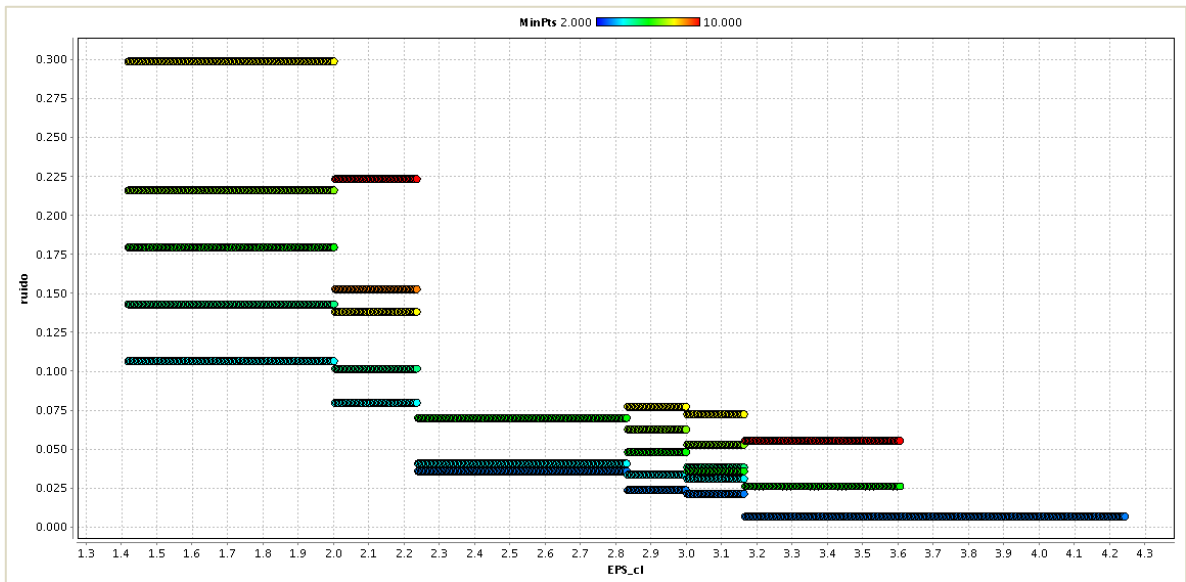


Figura 128. Relación de variables para Muestra B. x = EPS\_cl, y = ruido, color = MinPts.

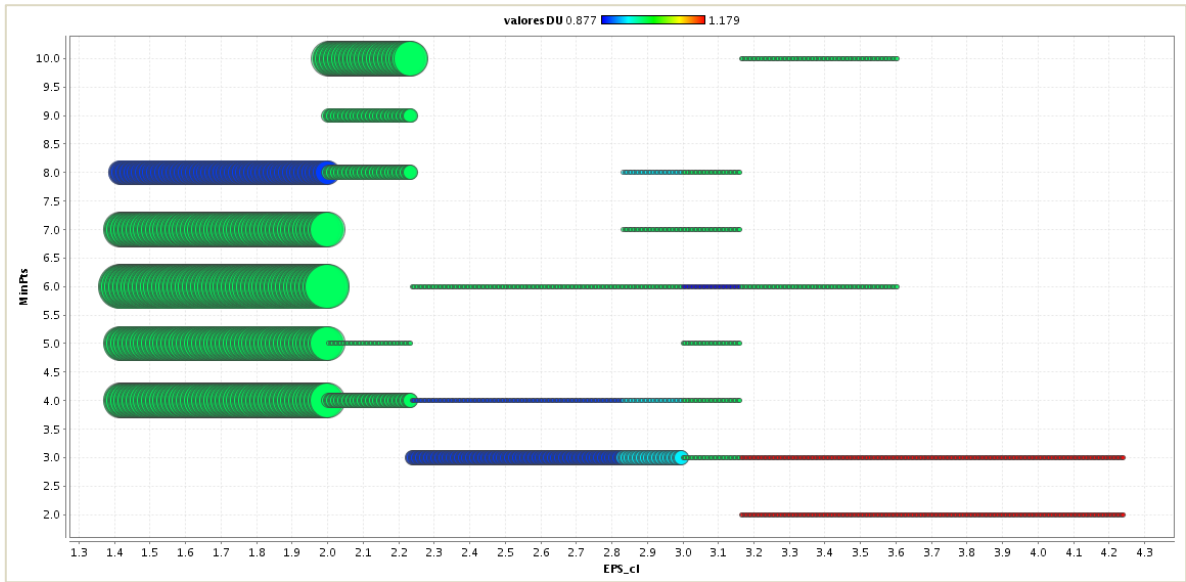


Figura 129. Relación de variables para Muestra B. x = EPS\_cl, y = MinPts, size = num\_clusters, color = valoresDU.

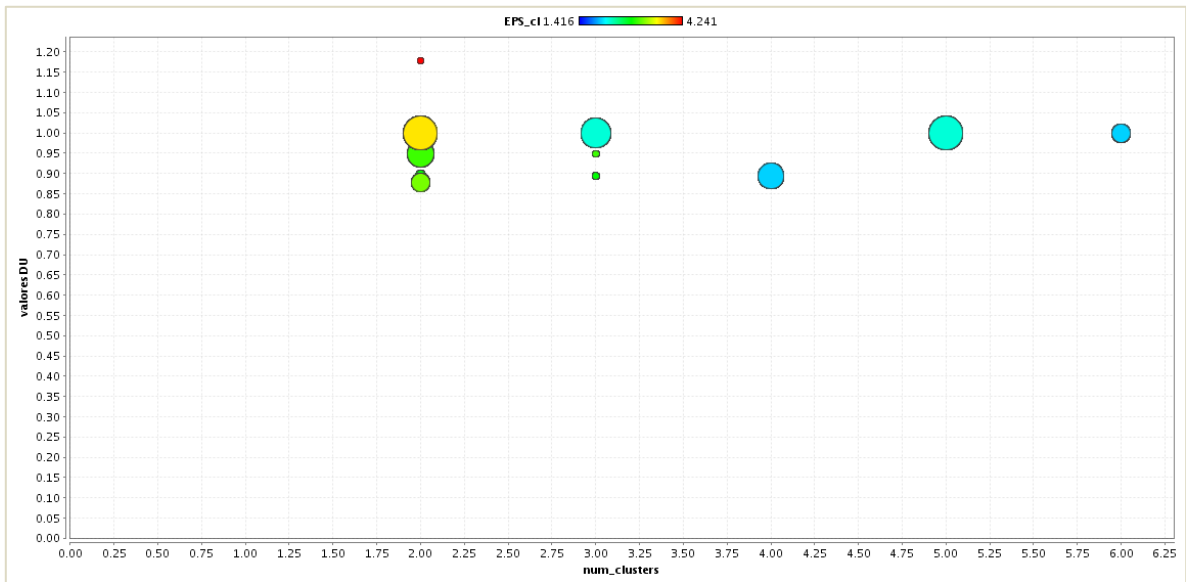


Figura 130. Relación de variables para Muestra B. x = num\_clusters, y = valoresDU, size = MinPts, color = EPS\_cl.

## 10.4 Anexo 5: Otras soluciones de *K-means*

Tabla 66. Centroides para de K-means similaridad coseno con una variable por día de semana. K=4

Cluster	vlun	vmar	vmie	vjue	vvie	vsab	vdom	tamaño
1	0.42634321	0.52434045	0.4566407	0.4290034	0.3598855	0.1122783	0.09278948	227

2	0.07485713	0.05492928	0.0937475	0.1028624	0.1263801	0.7539403	0.62258493	189
3	0.35144383	0.41303763	0.3564935	0.3646445	0.4088476	0.4533313	0.27049638	326
4	0.24254083	0.25302952	0.3774274	0.5362017	0.6131736	0.2408234	0.11492831	205

Los resultados de la Tabla 66 no muestran mayor distinción a los encontrados en 5.3.2 , pues no es posible rescatar información adicional de los *clusters* 1 y 4 , correspondientes a los visitantes con alta frecuencia en semana, ni *cluster* 3, visitantes constantes.

De todas formas, se analiza el comportamiento de aquellos usuarios que registran visitas en días de semana. El resultado (ver Tabla 67) muestra 2 grupos con tendencia de visitas hacia el final de la semana, y otro con visitas constantes, lo cual no se considera como información relevante para este estudio.

Tabla 67. Subclusterización K-means similitud coseno para días de semana. K=3

<i>Cluster</i>	<i>vlun</i>	<i>vmar</i>	<i>vmie</i>	<i>vjue</i>	<i>vvie</i>	tamaño
1	0.5783956	0.4141243	0.3598177	0.4226241	0.4311376	269
2	0.280161	0.5807786	0.5100764	0.4267571	0.3767048	261
3	0.2382864	0.279966	0.3889461	0.533962	0.6545564	253

## 10.5 Anexo 6: Detalle del análisis para otros CG

A continuación se presenta un tabla con información general de las bases de datos de los Centros de Gravedad que por sí solos contemplan los siguientes centros comerciales: Mall Vespucio Norte, Mall Plaza Oeste, Mall Plaza Maipú, Mall Plaza Egaña y Mall Alto Las Condes.

Tabla 68. Información general de otros CG's

	<b>Mall Vespucio Norte</b>	<b>Mall Plaza Oeste</b>	<b>Mall Plaza Maipú</b>	<b>Mall Plaza Egaña</b>	<b>Mall Alto Las Condes</b>
<b>Observaciones de la base saliente</b>	239444	98220	66915	8654	35952
<b>Números únicos</b>	69425	38977	13317	5288	11933
<b>días por mes (filtrado a 95% de confianza)</b>					
<b>Mayo</b>	28	30	31	31	31
<b>Junio</b>	28	28	27	27	28
<b>Julio</b>	29	29	28	28	25
<b>Agosto</b>	28	31	31	30	20



### 10.5.1 Mall Vespucio Norte

A continuación se presentan gráficas de comportamiento de llamadas para los meses de mayo, junio, julio y agosto en la antena que cubre el área del Mall Vespucio Norte (VN).

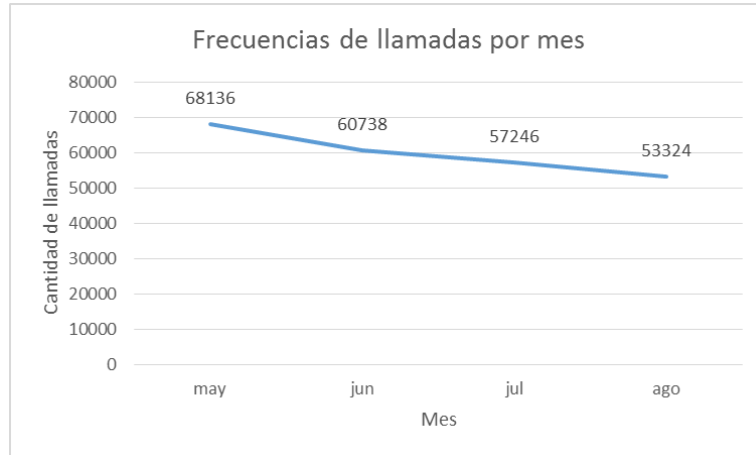


Figura 131 . frecuencia de llamadas por mes VN

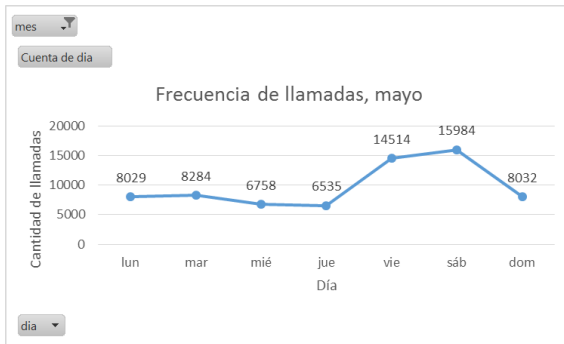


Figura 132. Frecuencia de llamadas por día de semana. Mayo. VN

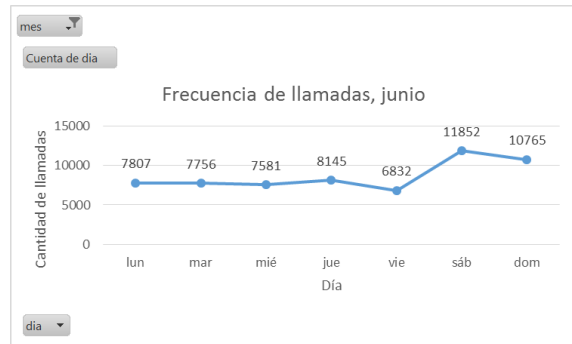


Figura 133. Frecuencia de llamadas por día de semana. Junio. VN

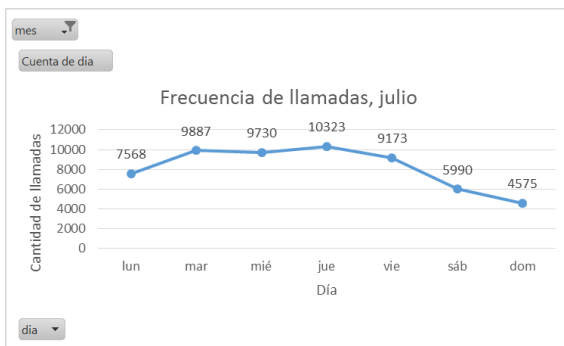


Figura 134. Frecuencia de llamadas por día de semana. Julio. VN

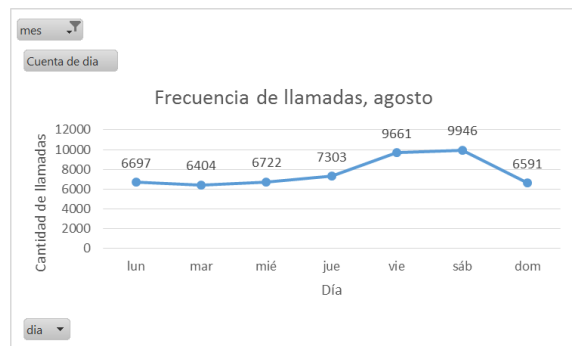


Figura 135. Frecuencia de llamadas por día de semana. Agosto. VN

El 99% de las personas llama menos de 25 veces en el periodo de tiempo analizado. Además se tienen *peaks* de llamadas luego de las 12:00 horas y entre las 18:00 y 19:00 horas. En este CG se observa que las llamadas aumentan en la segunda mitad del día. Este comportamiento es más marcado para los días viernes, sábado y domingo.

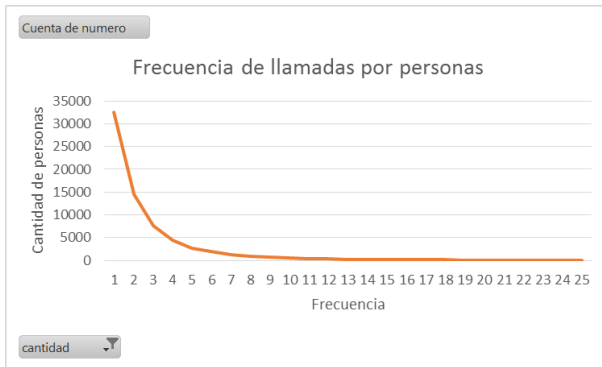


Figura 136. Frecuencia de llamadas por persona VN

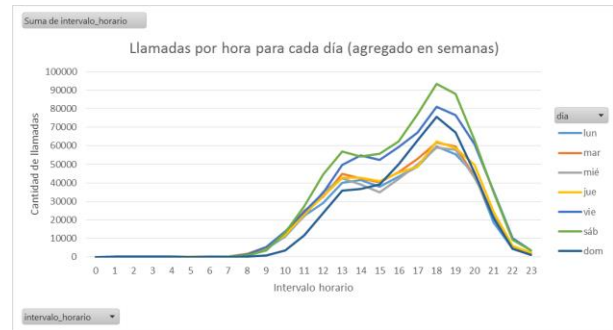


Figura 137. Frecuencia de llamadas por hora VN

Se observa alta correlación de las llamadas entre días de semana, la cual disminuye en días de fin de semana.

	vie	sab	dom	lun	mar	mie	jue
vie	1.0000000	0.5875879	0.5079220	0.7198418	0.7328574	0.7101287	0.7247493
sab	0.5875879	1.0000000	0.5972173	0.6343493	0.6414244	0.5701202	0.6014657
dom	0.5079220	0.5972173	1.0000000	0.5477238	0.5292633	0.5197885	0.5393572
lun	0.7198418	0.6343493	0.5477238	1.0000000	0.8326444	0.7325291	0.7928820
mar	0.7328574	0.6414244	0.5292633	0.8326444	1.0000000	0.7335133	0.7809183
mie	0.7101287	0.5701202	0.5197885	0.7325291	0.7335133	1.0000000	0.7349957
jue	0.7247493	0.6014657	0.5393572	0.7928820	0.7809183	0.7349957	1.0000000

Figura 138. Correlación de llamadas VN

El análisis de las visitas por hora, muestra un suavizamiento en el comportamiento, lo que indica que las personas que visitan en la segunda mitad del día tienen un comportamiento de llamadas más activo. Por otro lado, el análisis de visitas por número, indica que el 99% va menos de 17 veces en el periodo de tiempo, lo que implica que visitan alrededor de 1 vez por semana el CG (se tienen 18 semanas de datos). El 1% restante es el objetivo de estudio para obtenerlo se fijan umbrales de cantidad de visitas por tipo de día ( $ssem, sfd$ ) = (9,4). Se obtiene 1594 observaciones que corresponden al 0.67% de la población.



Figura 139. Frecuencia de visitas por persona VN

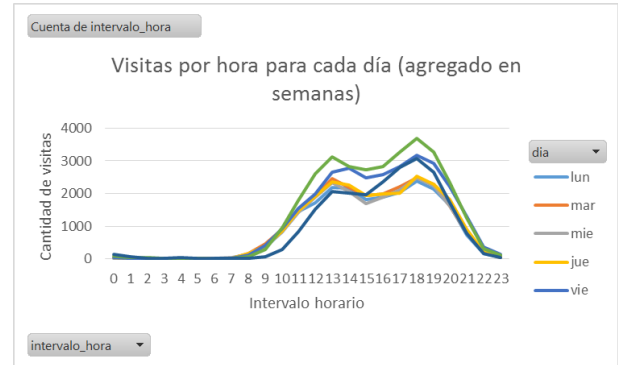


Figura 140. Frecuencia de visitas por hora VN

	vlun	vmar	vmie	vjue	vvie	vsab	vdom
vlun	1.0000000	0.4962779	0.4822301	0.4905946	0.4701404	0.2725943	0.2079414
vmar	0.4962779	1.0000000	0.4673048	0.4769780	0.4466752	0.2509026	0.1894244
vmie	0.4822301	0.4673048	1.0000000	0.4733155	0.4420158	0.2751194	0.2058073
vjue	0.4905946	0.4769780	0.4733155	1.0000000	0.4674298	0.2685253	0.2056138
vvie	0.4701404	0.4466752	0.4420158	0.4674298	1.0000000	0.2876090	0.2031826
vsab	0.2725943	0.2509026	0.2751194	0.2685253	0.2876090	1.0000000	0.2609156
vdom	0.2079414	0.1894244	0.2058073	0.2056138	0.2031826	0.2609156	1.0000000

Figura 141. Correlación de visitas VN

### 10.5.2 Mall Plaza Oeste

Para el Centro de Gravedad que contiene al Mall Plaza Oeste (PO) se observa que la mayoría de las llamadas se producen en el mes de mayo, con descenso en junio y aumento en julio y estabilización en agosto.

Además la tendencia de las llamadas para días de semana y fin de semana es similar al Centro de Gravedad anteriormente analizado, con estabilidad en la semana y aumento en los días de fin de semana.

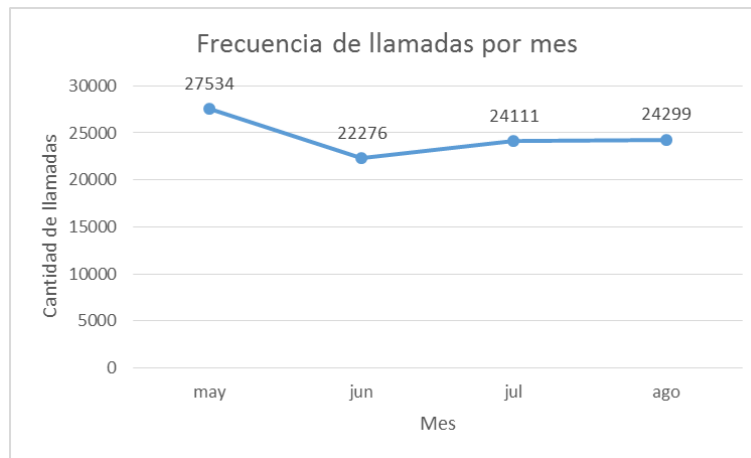


Figura 142. Frecuencia de llamadas por mes (PO)

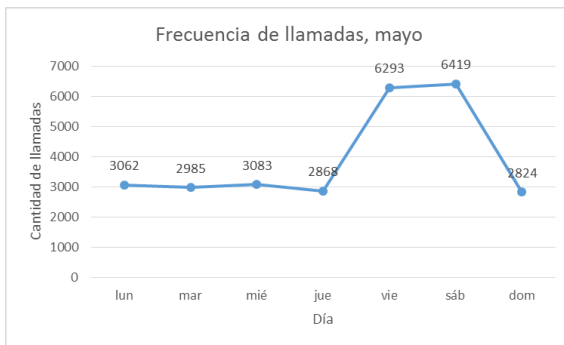


Figura 143. Frecuencia de llamadas por día de semana. Mayo. PO.

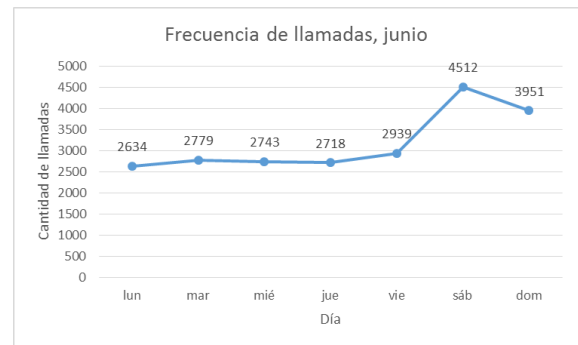


Figura 144. Frecuencia de llamadas por día de semana. Junio. PO.

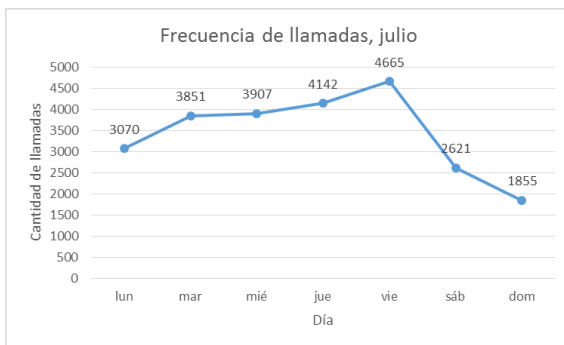


Figura 145. Frecuencia de llamadas por día de semana. Julio. PO.

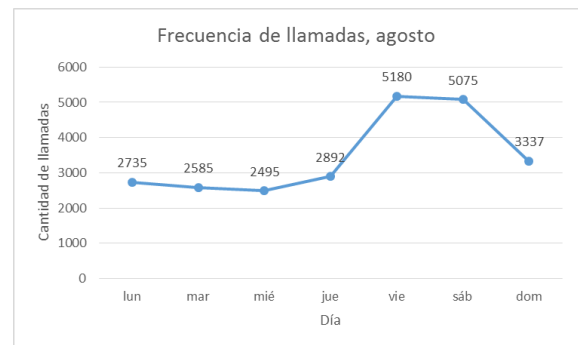


Figura 146. Frecuencia de llamadas por día de semana. Agosto. PO.

El análisis de frecuencia de llamadas por persona muestra que el 99% de las personas llama menos de 19 veces en el horizonte estudiado. Por otro lado se observan *peaks* de llamadas a las 13:00 y 19:00 horas.

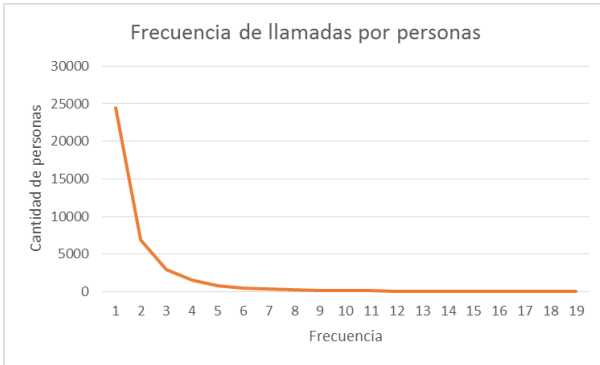


Figura 147. Frecuencia de llamadas por persona. PO.

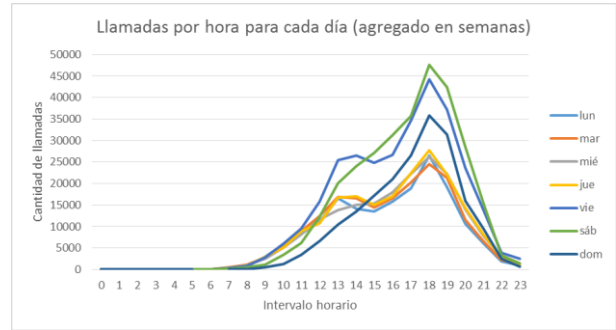


Figura 148. Frecuencia de llamadas por hora. PO.

Se observa alta correlación entre llamadas en día de semana y fin de semana.

	vie	sab	dom	lun	mar	mié	jue
vie	1.0000000	0.8332931	0.7888448	0.8529653	0.8381217	0.8030606	0.8676852
sab	0.8332931	1.0000000	0.8219729	0.8119307	0.7887556	0.7612722	0.7903950
dom	0.7888448	0.8219729	1.0000000	0.7860755	0.7628301	0.7439447	0.7499630
lun	0.8529653	0.8119307	0.7860755	1.0000000	0.8829953	0.8436882	0.8409199
mar	0.8381217	0.7887556	0.7628301	0.8829953	1.0000000	0.8507384	0.8445520
mié	0.8030606	0.7612722	0.7439447	0.8436882	0.8507384	1.0000000	0.8157401
jue	0.8676852	0.7903950	0.7499630	0.8409199	0.8445520	0.8157401	1.0000000

Figura 149. Correlación de llamadas PO

El análisis de visitas por hora y personas, muestra una manteniación atenuada de los *peaks* de llamadas y que el 99% de las personas va menos de 14 veces al Centro de Gravedad estudiado. Dado esto se fija el límite ( $ssem, sfd_s$ ) = 8,3, con lo cual se obtiene 912 observaciones correspondientes al 0.93% de los datos.



Figura 150. Frecuencia de visitas por persona.  
PO

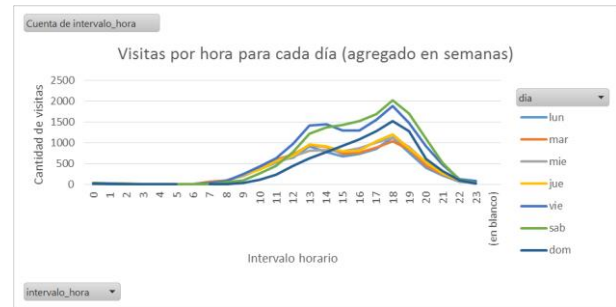


Figura 151. Frecuencia de visitas por hora.  
PO

	vlun	vmar	vmie	vjue	vvie	vsab	vdom
vlun	1.0000000	0.4787010	0.4347962	0.4606894	0.4046201	0.2647279	0.2382408
vmar	0.4787010	1.0000000	0.4419290	0.4731394	0.4059957	0.2585803	0.2312603
vmie	0.4347962	0.4419290	1.0000000	0.4565229	0.3932349	0.2460214	0.2308353
vjue	0.4606894	0.4731394	0.4565229	1.0000000	0.4243560	0.2753870	0.2279417
vvie	0.4046201	0.4059957	0.3932349	0.4243560	1.0000000	0.2468025	0.2131581
vsab	0.2647279	0.2585803	0.2460214	0.2753870	0.2468025	1.0000000	0.2349389
vdom	0.2382408	0.2312603	0.2308353	0.2279417	0.2131581	0.2349389	1.0000000

Figura 152. Correlación de visitas PO

### 10.5.3 Mall Plaza Maipú

En este caso, llamadas en el Centro de Gravedad que contiene al Mall Plaza Maipú (PM) muestra el máximo de llamadas en el mes de agosto (a diferencia de los CG anteriores). Sin embargo, la tendencia a la baja del mes de junio coincide con los CG de VN y PO

Por otro lado, el comportamiento de llamadas por días de semana es similar para los meses de mayo, jun y agosto (con aumento de llamadas en fin de semana). El mes de julio es particular, pues la tendencia es a la baja en días sábados y domingo.

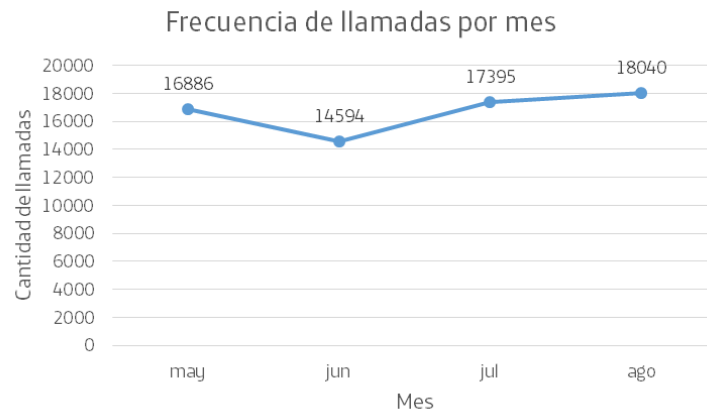


Figura 153. Frecuencias de llamadas por mes PM

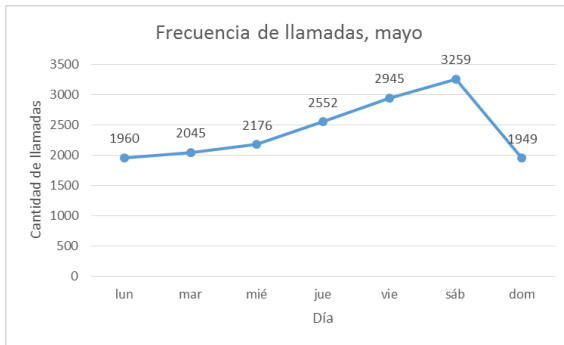


Figura 154. Frecuencia de llamadas por día. Mayo. PM

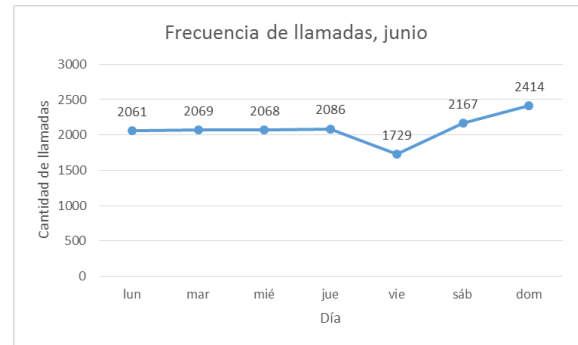


Figura 155. Frecuencia de llamadas por día. Junio. PM

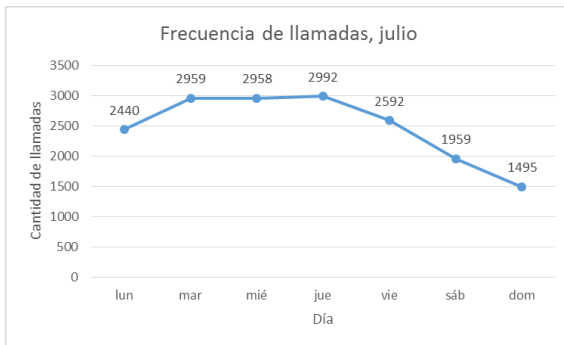


Figura 156. Frecuencia de llamadas por día. Julio. PM

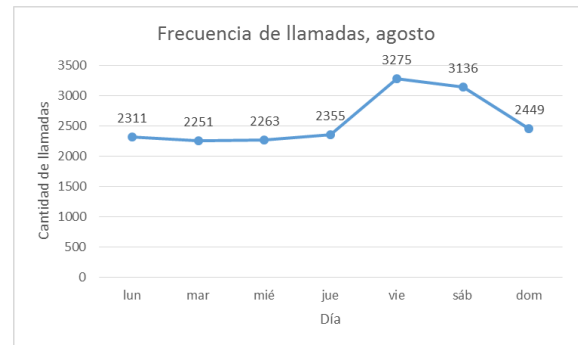


Figura 157. Frecuencia de llamadas por día. Agosto. PM

El análisis de frecuencia de llamadas por persona muestra que el 99% de las personas llama menos de 84 veces en el CG estudiado, mientras que el 96% llama menos de 23 veces.



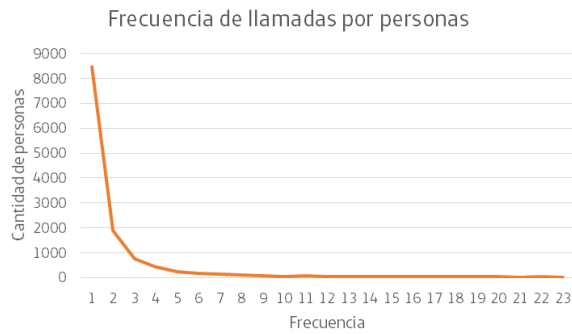


Figura 158. Frecuencia de llamadas por persona PM

	vie	jue	sab	dom	lun	mar	mie
vie	1.0000000	0.8486189	0.7539519	0.7225182	0.8469758	0.8305085	0.8558139
jue	0.8486189	1.0000000	0.7236946	0.6809928	0.8521155	0.8500209	0.8505457
sab	0.7539519	0.7236946	1.0000000	0.7903874	0.6941563	0.6832841	0.7206506
dom	0.7225182	0.6809928	0.7903874	1.0000000	0.6787981	0.7027970	0.7289156
lun	0.8469758	0.8521155	0.6941563	0.6787981	1.0000000	0.8504784	0.8578308
mar	0.8305085	0.8500209	0.6832841	0.7027970	0.8504784	1.0000000	0.8711590
mie	0.8558139	0.8505457	0.7206506	0.7289156	0.8578308	0.8711590	1.0000000

Figura 159. Correlación de llamadas

Se observa un comportamiento similar entre días. El día sábado aumentan las visitas en la primera mitad del día, mientras que los viernes aumentan las visitas en la segunda mitad del día. El análisis por personas indica que el 95% va menos de 16 veces, es decir alrededor de una vez a la semana. Se obtiene al 1.62% de las personas fijando el umbral ( $ssem, sfds$ ) = 8,3.

. La correlación tanto para llamadas como para visitas es alta para todos los días.

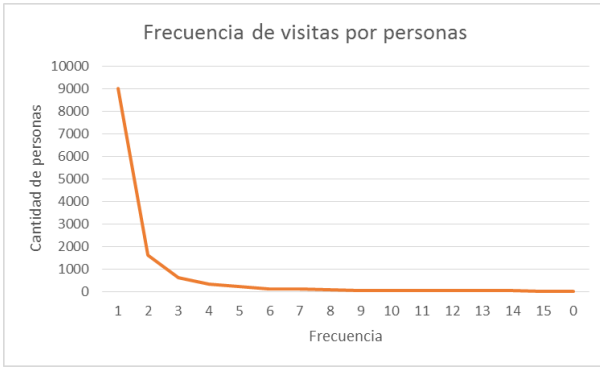


Figura 160. Frecuencia de visitas por persona PM

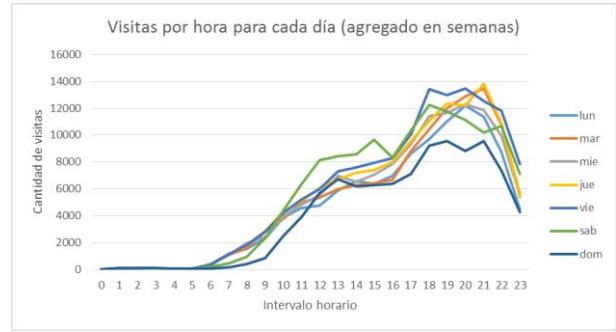


Figura 161. Frecuencia de visitas por hora PM

	vlun	vmar	vmie	vjue	vvie	vsab	vdom
vlun	1.0000000	0.7889651	0.7701654	0.7735200	0.7654367	0.6824123	0.6624397
vmar	0.7889651	1.0000000	0.7923195	0.7922713	0.7686075	0.6813002	0.6602236
vmie	0.7701654	0.7923195	1.0000000	0.7829385	0.7705704	0.7053772	0.6852156
vjue	0.7735200	0.7922713	0.7829385	1.0000000	0.7798218	0.6974808	0.6667888
vvie	0.7654367	0.7686075	0.7705704	0.7798218	1.0000000	0.7216992	0.6860429
vsab	0.6824123	0.6813002	0.7053772	0.6974808	0.7216992	1.0000000	0.7366607
vdom	0.6624397	0.6602236	0.6852156	0.6667888	0.6860429	0.7366607	1.0000000

Figura 162. Correlación de visitas PM

### 10.5.4 Mall Plaza Egaña

En este caso el comportamiento sale de la norma, lo que se explica por la fecha en la cual se capturan los datos, periodo en el cual este Mall no estaba completamente activo.

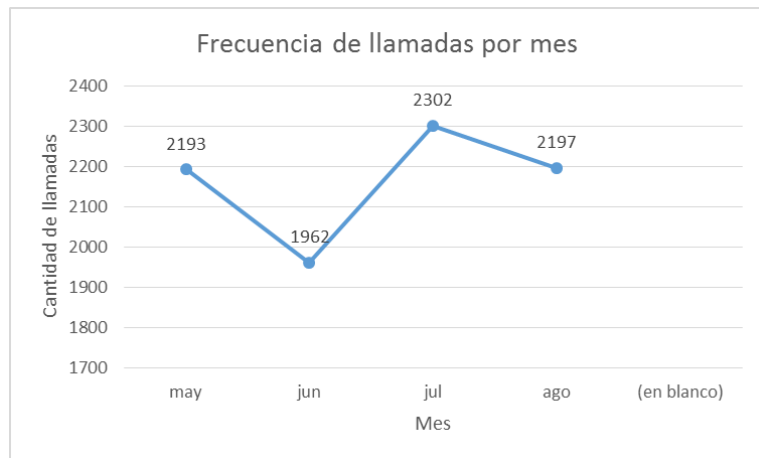


Figura 163. Frecuencia de llamadas por mes PE

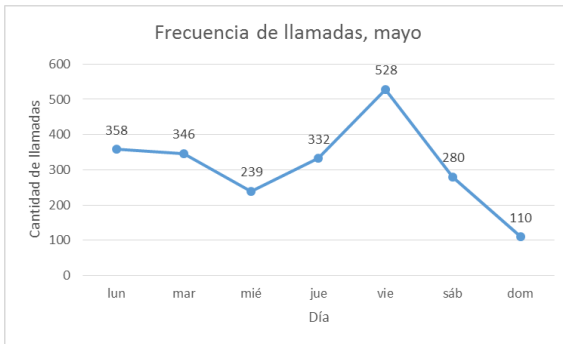


Figura 164. Frecuencia de llamadas por día. Mayo. PE.

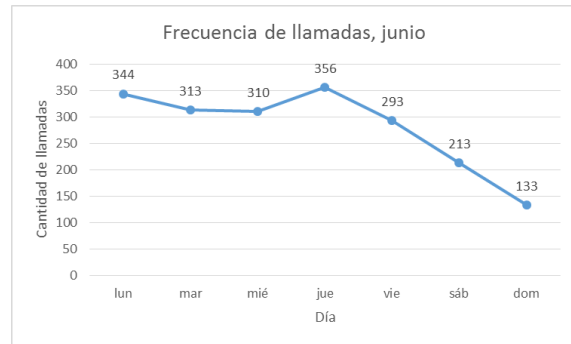


Figura 165. Frecuencia de llamadas por día. Junio. PE.

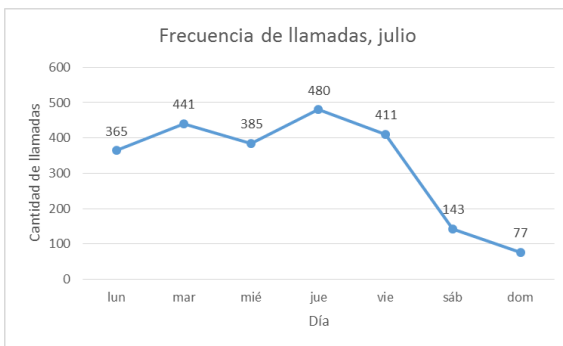


Figura 166. Frecuencia de llamadas por día. Julio. PE.

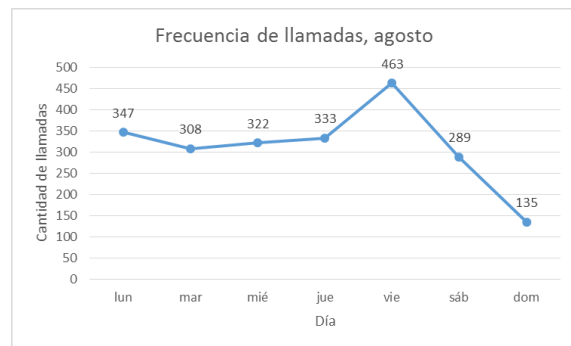


Figura 167. Frecuencia de llamadas por día. Agosto. PE.

Se observa que el 99% de las personas llama menos de 10 veces y que el *peak* de llamadas a las 19:00 horas se conserva sólo para los días de semana.

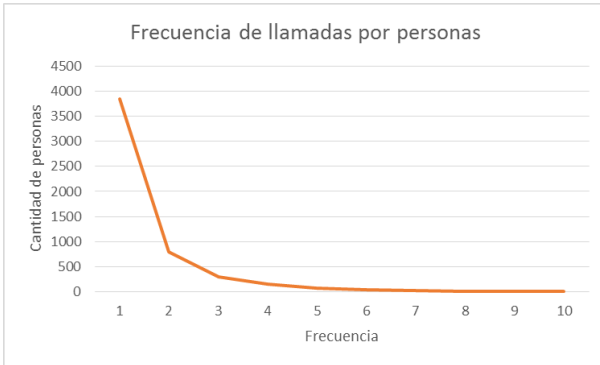


Figura 168. Frecuencia de llamadas por personas PE

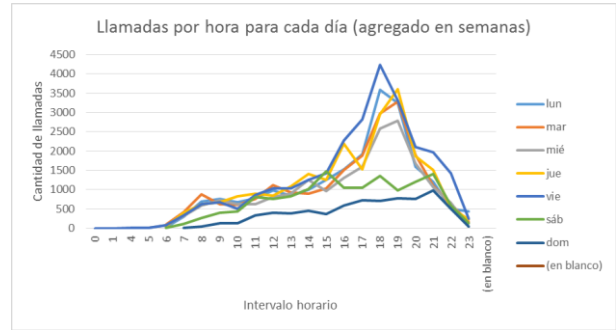


Figura 169. Frecuencia de llamadas por hora PE

La correlación entre llamadas, muestra un comportamiento anómalo, en comparación con el resto de los CG. Los días de fin de semana tienen correlación negativa con los días de semana, mientras que la correlación entre semana es baja.

	jue	vie	sab	dom	lun	mar	mié
jue	1.000000000	0.09388716	-0.004233947	-0.02658023	0.10151056	0.14361027	0.09191673
vie	0.093887160	1.00000000	-0.006918720	-0.01170882	0.09453834	0.19938205	0.13279466
sab	-0.004233947	-0.00691872	1.000000000	0.01029587	-0.02902198	-0.02194563	0.01003783
dom	-0.026580235	-0.01170882	0.010295873	1.00000000	-0.01988603	-0.02741711	-0.02001341
lun	0.101510557	0.09453834	-0.029021976	-0.01988603	1.00000000	0.11966147	0.11244438
mar	0.143610275	0.19938205	-0.021945627	-0.02741711	0.11966147	1.00000000	0.13330864
mié	0.091916726	0.13279466	0.010037834	-0.02001341	0.11244438	0.13330864	1.00000000

Figura 170. Correlación de llamadas PE

Las visitas mantienen el comportamiento de las llamadas por hora, tanto para días de semana como para días de fin de semana. Además 99% va menos de 8 veces en el periodo de tiempo observado, por lo que se fijan los límites *ssem* y *sfds* para personas que van al menos 1 vez al mes en días de semana o en 3 de 4 fin de semanas. Lo anterior filtra de 5.256 a 165 observaciones.



Figura 171. Frecuencia de visitas por persona PE

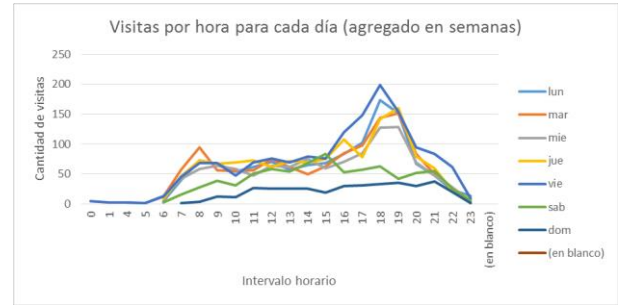


Figura 172. Frecuencia de visitas por hora PE

	vlun	vmar	vmie	vjue	vvie	vsab	vdom
vlun	1.00000000	0.08726791	0.09271662	0.09582295	0.08839647	-0.05736956	-0.03433207
vmar	0.08726791	1.00000000	0.10905624	0.09905513	0.20075594	-0.04937706	-0.03348700
vmie	0.09271662	0.10905624	1.00000000	0.07267645	0.10194798	-0.03605910	-0.01457690
vjue	0.09582295	0.09905513	0.07267645	1.00000000	0.07703911	-0.04420414	-0.03451989
vvie	0.08839647	0.20075594	0.10194798	0.07703911	1.00000000	-0.03736744	-0.01500508
vsab	-0.05736956	-0.04937706	-0.03605910	-0.04420414	-0.03736744	1.00000000	0.02246840
vdom	-0.03433207	-0.03348700	-0.01457690	-0.03451989	-0.01500508	0.02246840	1.00000000

Figura 173. Correlación de visitas PE

### 10.5.5 Mall Alto Las Condes

Para este CG (ALC) se registra una baja de llamadas en el mes de agosto, lo cual podría explicarse por la falta de registros en varios días de este mes (sólo se registran 20 días).

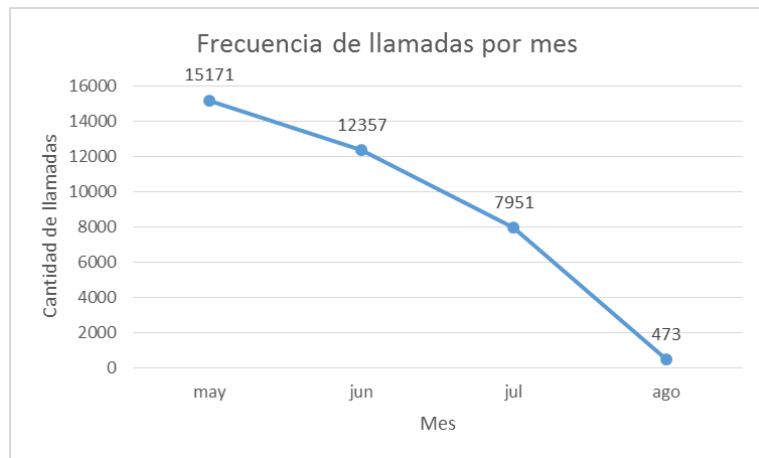


Figura 174. Frecuencia de llamadas por mes ALC

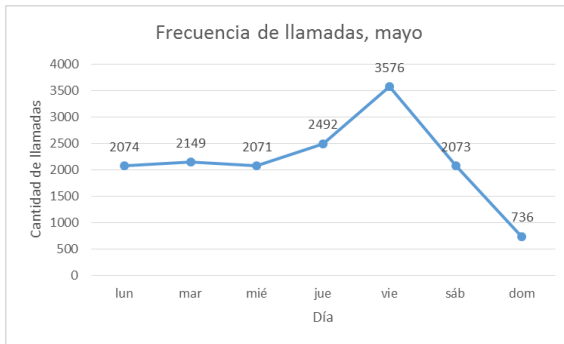


Figura 175. Frecuencia de llamadas por día. Mayo. ALC.

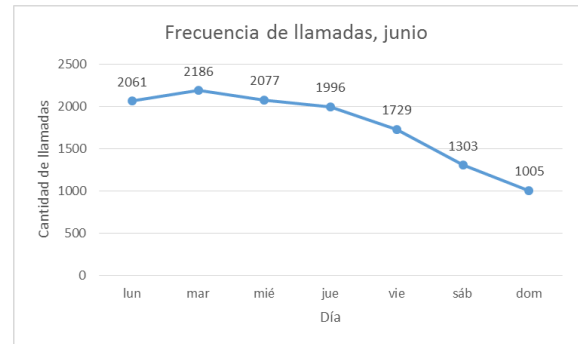


Figura 176. Frecuencia de llamadas por día. Junio. ALC.

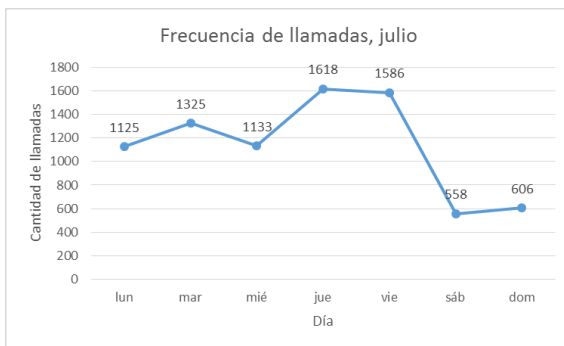


Figura 177. Frecuencia de llamadas por día. Julio. ALC.

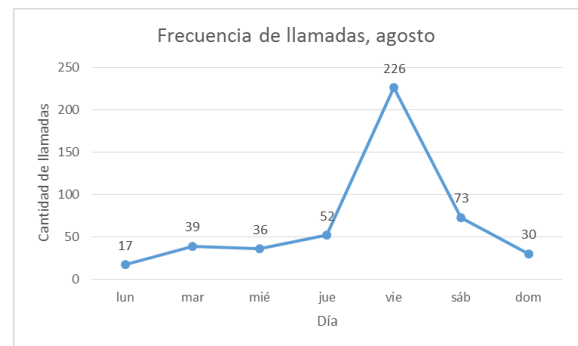


Figura 178. Frecuencia de llamadas por día. Agosto. ALC.

	vie	sab	dom	lun	mar	mie	jue
vie	1.0000000	0.5796686	0.4064239	0.6994860	0.7094050	0.7164371	0.7301972
sab	0.5796686	1.0000000	0.5031115	0.4942922	0.4729954	0.5217090	0.5252365
dom	0.4064239	0.5031115	1.0000000	0.3723610	0.4230982	0.4407793	0.4532761
lun	0.6994860	0.4942922	0.3723610	1.0000000	0.7164961	0.6810439	0.6865506
mar	0.7094050	0.4729954	0.4230982	0.7164961	1.0000000	0.7355532	0.7440641
mie	0.7164371	0.5217090	0.4407793	0.6810439	0.7355532	1.0000000	0.7936061
jue	0.7301972	0.5252365	0.4532761	0.6865506	0.7440641	0.7936061	1.0000000

Figura 179. Correlación de llamadas ALC

Se nota que el 99% de las personas llama menos de 32 veces en el horizonte de tiempo estudiado, mientras que el 97% lo hace menos de 14 veces.



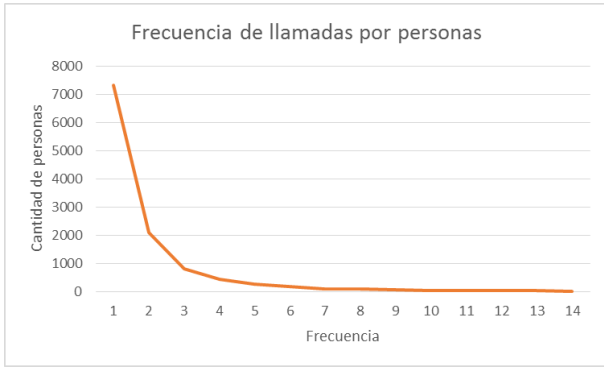


Figura 180. Frecuencia de llamadas por personas ALC

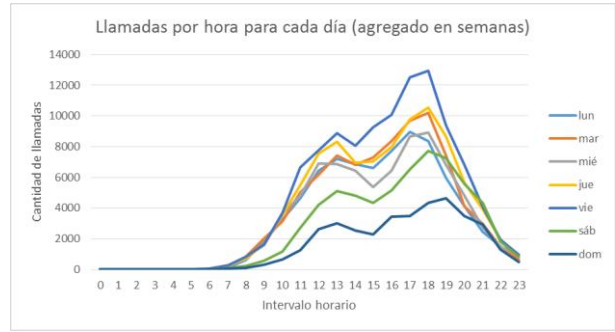


Figura 181. Frecuencia de llamadas por hora ALC

Finalmente, para las visitas se observan los *peaks* a las 13:00 y 19:00 horas, con menor tendencia para los días sábado y domingo. Además se nota que el 99% de las personas va menos de 22 veces en 4 meses y el 98% menos de 14 veces al mes. Se obtiene, con  $(ssem, sfd_s) = 8,3$ , un total de 1423 observaciones que corresponden al 0.82% de las personas.

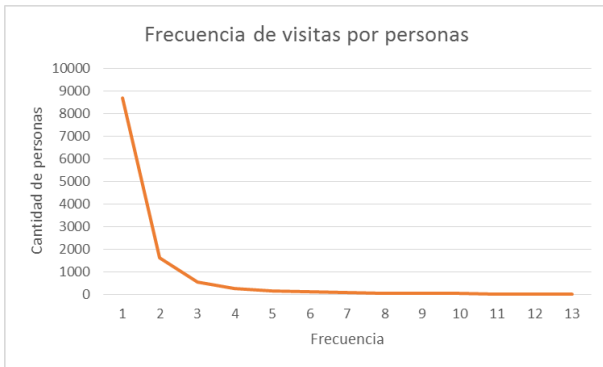


Figura 182. Frecuencia de visitas por personas ALC

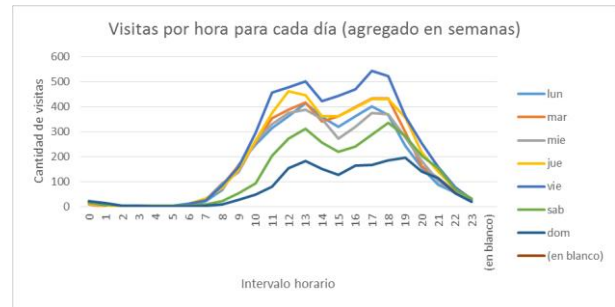


Figura 183. Frecuencia de visitas por hora ALC

	vlun ↕	vmar ↕	vmie ↕	vjue ↕	vvie ↕	vsab ↕	vdom ↕
vlun	1.0000000	0.5092675	0.4835853	0.4783262	0.4854376	0.2939333	0.2350646
vmar	0.5092675	1.0000000	0.4864244	0.4934182	0.4868020	0.3010528	0.2338114
vmie	0.4835853	0.4864244	1.0000000	0.5023701	0.4789255	0.3223665	0.2642066
vjue	0.4783262	0.4934182	0.5023701	1.0000000	0.4997028	0.3322670	0.2665793
vvie	0.4854376	0.4868020	0.4789255	0.4997028	1.0000000	0.3501509	0.2586471
vsab	0.2939333	0.3010528	0.3223665	0.3322670	0.3501509	1.0000000	0.3557602
vdom	0.2350646	0.2338114	0.2642066	0.2665793	0.2586471	0.3557602	1.0000000

Figura 184. Correlación de visitas ALC

## 10.6 Anexo 7: Flujo de caja sin proyecto

Tabla 69 . Flujo de caja sin proyecto en CLP

Año	0	1	2	3
<b>Ingresos por ventas</b>	\$-	<b>\$63.000.000</b>	<b>\$180.000.000</b>	<b>\$180.000.000</b>
Ganancias/Pérdidas de Capital	\$-	\$-	\$-	\$-
<b>Costos Fijos</b>	\$-	<b>\$67.865.952</b>	<b>\$67.865.952</b>	<b>\$67.865.952</b>
<b>Costos Variables</b>	\$-	<b>\$117.240.000</b>	<b>\$320.160.000</b>	<b>\$377.280.000</b>
Depreciaciones legales	\$-	\$2.459.500	\$2.626.167	\$2.626.167
Pérdidas del ejercicio anterior	\$-	\$204.021.790	\$136.422.862	\$223.342.862
<b>Utilidad antes de impuestos</b>	\$-	<b>-\$328.587.242</b>	<b>-\$347.074.980</b>	<b>-\$491.114.980</b>
Impuesto de primera categoría	\$-	\$-	\$-	\$-
<b>Utilidad después de impuestos</b>	\$-	<b>-\$328.587.242</b>	<b>-\$347.074.980</b>	<b>-\$491.114.980</b>
<b>Flujo de Caja Operacional</b>	\$-	<b>-\$122.105.952</b>	<b>-\$208.025.952</b>	<b>-\$265.145.952</b>
<b>Inversión Fija</b>	<b>\$221.927.170</b>	<b>\$4.000.000</b>	<b>\$5.000.000</b>	<b>\$20.170</b>
Valor residual de los Activos	\$-	\$-	\$-	\$-
Capital de Trabajo	\$-	\$-	\$-	\$-
Recuperación del Capital de Trabajo	\$-	\$-	\$-	\$-
Préstamos	\$17.905.380	\$-	\$-	\$-
Amortizaciones	\$-	\$10.316.910	\$10.316.910	\$-
<b>Flujo de Capitales</b>	<b>-\$204.021.790</b>	<b>-\$14.316.910</b>	<b>-\$15.316.910</b>	<b>-\$20.170</b>
<b>Flujo de Caja Privado</b>	<b>-\$204.021.790</b>	<b>-\$136.422.862</b>	<b>-\$223.342.862</b>	<b>-\$265.166.122</b>

La inversión en costos fijos se mantienen para el estado con y sin proyecto, sin embargo se incurre en un costo fuertemente mayor en la inversión inicial debido a que el envío de mensajes de la campaña de *Opt-in* se realiza a todas las líneas activas. Sin embargo no se costean las horas hombre de desarrollo de modelos. Los costos de maquinaria necesaria para el almacenamiento y procesamiento de datos también varían.

Tabla 70. Detalle de la inversión fija en estado sin proyecto (CLP).

<b>Inversión Fija (CLP)</b>	<b>\$</b>	<b>221.927.170</b>
Servidor de <i>DB</i> (CLP)	\$	2.019.000
Servidor de producción (CLP)	\$	4.038.000
Dominio (CLP)	\$	20.170
Horas desarrollador <i>web</i> (CLP)	\$	2.000.000
Horas diseñador <i>web freelance</i> (CLP)	\$	1.500.000
Computadores y licencias (CLP)	\$	8.000.000
Impresora (CLP)	\$	350.000
Envío de mensajes (CLP)	\$	200.000.000
Campaña en medios a clientes (CLP)	\$	4.000.000