



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL

SISTEMA INTEGRAL DE RECOMENDACIÓN DE PRODUCTOS PARA CANJE EN UN RETAIL FINANCIERO

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

CRISTIAN ALONSO AGUILERA ORTÚN

PROFESOR GUÍA:
ALEJANDRA PUENTE CHANDÍA

MIEMBROS DE LA COMISIÓN:
ERICK MÉNDEZ GUZMAN
MARCEL GOIC FIGUEROA

SANTIAGO DE CHILE
2018

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL
POR: CRISTIAN ALONSO AGUILERA ORTÚN
FECHA: 11/11/2018
PROF. GUIA: ALEJANDRA PUENTE

SISTEMA INTEGRAL DE RECOMENDACIÓN DE PRODUCTOS PARA CANJE EN UN RETAIL FINANCIERO

La industria de servicios financieros, forma parte de un mercado relevante, ya que, a febrero de 2018 en Chile, existían 21 millones de tarjetas de crédito activas, por lo que una persona posee en promedio más de una tarjeta, cuyo objetivo es ser la escogida a la hora del pago. Producto de esto nace la necesidad de diferenciación de los emisores de tarjetas de crédito, ya sea mediante alianzas para ofrecer descuentos, beneficios o a través de los planes de fidelización. Así, abordar un problema relacionado al plan de fidelización de una de las emisoras de tarjetas de crédito más grande del país resulta un desafío relevante.

Con el trabajo de título, se busca desarrollar y proponer una metodología integral de recomendación de productos para canje en un retail financiero, que reemplace la forma manual y no basada en datos con que hoy en día se seleccionan los clientes que reciben una comunicación de canje. Que considere, en primer lugar, la respuesta esperada de los clientes a un mail de canje, incluyendo a los que reaccionan positivamente y excluyendo a los que reaccionan negativamente a la comunicación. En segundo lugar, considerando la línea de productos de canje que el cliente es más propenso a escoger. En resumen, se pretende determinar a quién le envió un mail de canje y qué le envió.

Para llevar a cabo lo anterior, se realiza un diseño experimental que permita medir la respuesta esperada de los clientes frente a un mail genérico de canje, para luego modelar sus respuestas con un modelo Up-Lift, que ubica a los clientes en cuatro categorías, de las cuales las más importantes son los clientes “influenciables”, con respuesta positiva al mail de canje, y los “No molestar”, con respuesta negativa. Luego se calcula la propensión a canjear en una de las 7 líneas de canje definidas mediante 4 modelos distintos, que se comparan en base a accuracy, recall y ganancia de información para definir el mejor en cada caso.

Los resultados de la primera parte, relativa a la respuesta esperada a un mail de canje, indican que los clientes influenciados son, en general, personas más jóvenes, con menor gasto y días con compras con la tarjeta y que logran acumular relativamente pocos puntos, mientras que los no influenciados acumulan bastantes puntos y tienen mayor gasto, siendo clientes más cercanos al plan de fidelización. Para la segunda parte, se obtuvo que existen líneas de canje que presentan tipos de clientes más claros, por lo que muestran mejores resultados en los modelos a la hora de perfilarlos, mientras que hay otras que cuesta más encontrar patrones y que, por ende, sus resultados no son igual de robustos. Así, se propone combinar ambos modelamientos, considerando enviar comunicaciones de puntos a los clientes que sean categorizados como influenciados y comunicándole la línea de productos de canje que presente el mayor lift asociado.

AGRADECIMIENTOS

Con este trabajo de título termina un largo proceso de aprendizaje que, además de estar compuesto por muchas clases, conocimiento, evaluaciones, etcétera, estuvo compuesto de momentos y de personas que cumplieron un rol fundamental en cada etapa.

Así, agradezco enormemente el apoyo y constante comprensión de mi familia, mi mamá, Cristian, Maxi, Tía Pili, Ramiro, que a pesar de estar sumido en el vertiginoso mundo universitario, me entendieron y apoyaron siempre.

A mi madre por todo su esfuerzo a lo largo de estos años y los pasados, por cumplir una labor ejemplar tanto de educación académica como valórica, inculcándome siempre los mejores principios y dando el ejemplo con su intachable y solidario comportamiento.

A mi polola Javi por toda la comprensión, por todo el apoyo, pero por sobre todo por el reconocimiento entregado en cada éxito y el soporte entregado en cada fracaso, es todo lo que se puede esperar. Amor infinito para ti.

A mis amigos de la universidad, que más que un simple grupo de estudios fue una red de apoyo fundamental para cumplir con cada etapa de la mejor manera, y que además de compañeros se convirtieron en amigos de la vida.

A los profesores Alejandra, Erick y Carolina, que dieron los consejos precisos para que este largo y estresante proceso tuviera el resultado esperado.

Finalmente a todos los integrantes del área de Inteligencia de Clientes de la empresa, que siempre estuvieron dispuestos a ayudarme, enseñarme y apoyarme, son un tremendo equipo y unas tremendas personas.

“The people who are crazy
enough to think they can
change the world are the
ones who do”



Tabla de contenido

1	Introducción.....	8
1.1	Antecedentes de la industria.....	8
1.2	Caracterización de la empresa.....	10
1.3	Área de desarrollo del trabajo de título.....	11
1.4	Descripción del proyecto y justificación.....	12
1.4.1	Identificación del área.....	12
1.4.2	Descripción del problema.....	12
1.4.3	Justificación.....	14
1.4.4	Descripción del proyecto.....	18
1.5	Objetivos.....	19
1.5.1	Objetivo general.....	19
1.5.2	Objetivos específicos.....	19
1.5.3	Alcances.....	19
2	Marco teórico.....	21
2.1	Minería de datos.....	21
2.1.1	Metodología KDD.....	21
2.1.2	Preparación de los datos.....	22
2.1.3	Detección de datos fuera de rango.....	23
2.2	Experimentación.....	23
2.2.1	Diseño experimental.....	23
2.2.2	Tamaños de muestra.....	24
2.2.3	Evaluación de experimentos.....	25
2.3	Modelamiento.....	25
2.3.1	Up-Lift.....	25
2.3.2	Evaluación de modelos Up-Lift.....	27
2.3.3	Modelos de clasificación y propensión.....	28
2.3.4	Elección de modelos.....	29
3	Metodología.....	31
3.1	Comprensión del negocio.....	31
3.2	Selección de datos.....	31
3.3	Pre-procesamiento.....	31
3.4	Transformación de datos.....	31
3.5	Minería de datos.....	32
3.6	Análisis, interpretación y evaluación.....	32
4	Flujo metodológico.....	33
5	Resultados esperados.....	35
6	Desarrollo metodológico.....	36
6.1	¿A quién le envió un mail? Up-Lift.....	36
6.1.1	Tamaño muestral.....	36
6.1.2	Base potencial de clientes.....	36
6.1.3	Diseño experimental.....	37
6.1.4	Resultados experimento.....	37
6.1.5	Modelamiento.....	39

6.1.6	Perfilamiento	44
6.1.7	Descubrimientos relevantes	46
6.1.8	Conclusiones preliminares	49
6.2	¿Qué le envió?	50
6.2.1	Análisis exploratorio	50
6.2.2	Definición de líneas de productos	53
6.2.3	Exploratorio para líneas escogidas	55
6.2.4	Preparación de datos	58
6.2.5	Modelamiento	61
6.2.6	Resultados	62
6.2.7	Nodos más y menos propensos	70
6.2.8	Conclusiones preliminares	74
7	Justificación económica.....	75
8	Conclusiones.....	76
8.1	¿A quién le envió?.....	76
8.1.1	De la metodología	76
8.1.2	De los resultados	76
8.2	¿Qué le envió?.....	77
8.2.1	De la metodología	77
8.2.2	De los resultados	77
8.3	Globales.....	78
9	Recomendaciones	79
10	Limitaciones y trabajo futuro	81
11	Bibliografía.....	82
12	Anexos	84
12.1	Anexo 1: Valores y principios organizacionales	84
12.2	Anexo 2: Estructura organizacional	84
12.3	Anexo 3: Estructura del área	85
12.4	Anexo 4: Reglas de acumulación	85
12.5	Anexo 5: Tabla de equivalencia de puntos	85
12.6	Anexo 6: Evolución de cuentas abiertas.....	86
12.7	Anexo 7: Sucursales y funcionarios	86
12.8	Anexo 8: Evolución de la utilidad	86
12.9	Anexo 9: Trabajos anteriores.....	87
12.10	Anexo 10: Métricas de desempeño	89
12.11	Anexo 11: Flujo de definición de base analítica	89
12.12	Anexo 12: Cálculo tamaño mínimo de muestra	90
12.13	Anexo 13: Pieza enviada para el target 1	90
12.14	Anexo 14: Pieza enviada para el target 2	91
12.15	Anexo 15: Variables Up-Lift	92
12.16	Anexo 16: Variables finales Up-Lift.....	92
12.17	Anexo 17: Otros descriptivos líneas de canje	93
12.18	Anexo 18: Variables modelo líneas	95
12.19	Anexo 19: Productos más relevantes según frecuencia de canje por línea	96
12.20	Anexo 20: Matriz de correlación.....	98
12.21	Anexo 21: Crecimiento de datos móviles y capacidad	99
12.22	Anexo 22: Resumen del piloto experimental diseñado.....	100
12.23	Anexo 23: Nodos más y menos propensos a ser influenciables.....	100

12.24	Anexo 24: Distribución de canjes según género	101
12.25	Anexo 25: Distribución de canjes según drivers de engancho	101
12.26	Anexo 26: Distribución de canjes según actividad en tiendas del grupo	103
12.27	Anexo 27: Distribución de canje según ingreso web	104
12.28	Anexo 28: Distribución de edad y antigüedad de los clientes canjeadores	104

Índice de tablas

TABLA 1: PARTICIPANTES DE LA INDUSTRIA DE SERVICIOS FINANCIEROS CHILENA CON TARJETAS DE CRÉDITO ACTIVAS A FEBRERO DE 2018	8
TABLA 2: DIFERENCIA DE GASTO SEGÚN DECILES DE GASTO EN 2015	15
TABLA 3: RESULTADOS GENERALES DEL PILOTO EXPERIMENTAL ENVIADO	38
TABLA 4: COEFICIENTE DE QINI OBTENIDO PARA CADA DIVERGENCIA DEL MODELO AJUSTADO.....	42
TABLA 5: NODOS MÁS Y MENOS PROPENSOS.....	46
TABLA 6: DISTRIBUCIÓN DE CLIENTES INFLUENCIABLES Y NO MOLESTAR EN RELACIÓN A LA ZONA DE RESIDENCIA	47
TABLA 7: DISTRIBUCIÓN DE CLIENTES INFLUENCIABLES Y NO MOLESTAR EN RELACIÓN A LA EDAD	48
TABLA 8: DISTRIBUCIÓN DE CLIENTES INFLUENCIABLES Y NO MOLESTAR EN RELACIÓN A LOS PUNTOS ACUMULADOS	48
TABLA 9: DISTRIBUCIÓN DE CLIENTES INFLUENCIABLES Y NO MOLESTAR EN RELACIÓN AL NIVEL DE CANJE ANTERIOR	49
TABLA 10: PORCENTAJE DE CANJES Y PUNTOS CANJEADOS QUE REPRESENTA CADA LÍNEA SOBRE EL TOTAL SIN LAS LÍNEAS DE GIFT CARD, SEGUROS Y VIAJES	54
TABLA 11: DETALLES DE CANJES POR LÍNEA:.....	57
TABLA 12: MÉTRICAS DE EVALUACIÓN PARA LA LÍNEA DE HERRAMIENTAS	62
TABLA 13: MÉTRICAS DE EVALUACIÓN PARA LA LÍNEA DE ELECTRO	64
TABLA 14: MÉTRICAS DE EVALUACIÓN PARA LA LÍNEA DE BEBIDAS ALCOHÓLICAS.....	65
TABLA 15: MÉTRICAS DE EVALUACIÓN PARA LA LÍNEA DE ALIMENTOS	66
TABLA 16: MÉTRICAS DE EVALUACIÓN PARA LA LÍNEA DE TECNO	67
TABLA 17: MÉTRICAS DE EVALUACIÓN PARA LA LÍNEA DE DORMITORIO Y DECORACIÓN.....	68
TABLA 18: MÉTRICAS DE EVALUACIÓN PARA LA LÍNEA DE MENAJE	69
TABLA 19: ANÁLISIS DE ESCENARIOS ECONÓMICOS.....	75
TABLA 20: VARIABLES DISPONIBLES PARA EL MODELO UP-LIFT	92
TABLA 21: VARIABLES SELECCIONADAS PARA EL MODELAMIENTO UP-LIFT	92
TABLA 22: VARIABLES DISPONIBLES PARA EL MODELAMIENTO DE LÍNEAS DE CANJE	95
TABLA 23: PRODUCTOS SEGÚN FRECUENCIA DE CANJE PARA LA LÍNEA DE HERRAMIENTAS	96
TABLA 24: PRODUCTOS SEGÚN FRECUENCIA DE CANJE PARA LA LÍNEA DE ELECTRO	96
TABLA 25: PRODUCTOS SEGÚN FRECUENCIA DE CANJE PARA LA LÍNEA DE BEBIDAS ALCOHÓLICAS	96
TABLA 26: PRODUCTOS SEGÚN FRECUENCIA DE CANJE PARA LA LÍNEA DE ALIMENTOS.....	96
TABLA 27: PRODUCTOS SEGÚN FRECUENCIA DE CANJE PARA LA LÍNEA DE TECNO	97
TABLA 28: PRODUCTOS SEGÚN FRECUENCIA DE CANJE PARA LA LÍNEA DE DORMITORIO Y DECORACIÓN	97
TABLA 29: PRODUCTOS SEGÚN FRECUENCIA DE CANJE PARA LA LÍNEA DE MENAJE	97
TABLA 30: MATRIZ DE CORRELACIÓN.....	98

Índice de figuras

FIGURA 1: CANTIDAD DE TARJETAS DE CRÉDITO POR EMISOR A FEBRERO DE 2018.....	9
FIGURA 2: MONTO GASTADO ASOCIADO A TARJETAS DE CRÉDITO POR EMISOR A FEBRERO DE 2018.	9
FIGURA 3: DISTRIBUCIÓN DE LOS CLIENTES SEGÚN COMPORTAMIENTO DE CANJE.....	13
FIGURA 4: DISTRIBUCIÓN DE CLIENTES SEGÚN CANTIDAD DE CANJES EN SU VIDA COMO CLIENTE.	13
FIGURA 5: GASTO SEGÚN DRIVER DE ENGANCHE DEL CLIENTE.	14
FIGURA 6: CANTIDAD DE CLIENTES CONTACTABLES VÍA MAIL Y TELÉFONO.....	16
FIGURA 7: DISTRIBUCIÓN DE LA CANTIDAD DE CANJES EN LÍNEAS DE PREMIOS.....	18
FIGURA 8: TASAS DE RESPUESTA Y DIFERENCIAS POR HISTORIAL DE CANJE.	38
FIGURA 9: TASAS DE RESPUESTA Y DIFERENCIAS POR CATEGORÍA.	39
FIGURA 10: IMPORTANCIA RELATIVA DE LAS VARIABLES PARA EXPLICAR EL CANJE.	40
FIGURA 11: IMPORTANCIA RELATIVA DE LAS VARIABLES DEL MODELO CALIBRADO.....	41
FIGURA 12: RESULTADOS DEL MODELO UP-LIFT PARA EL MÉTODO DE INTERACCIONES.....	42
FIGURA 13: CURVA DE GANANCIA DE INFORMACIÓN DEL MODELO UP-LIFT PARA EL MÉTODO DE INTERACCIONES SOBRE EL CONJUNTO DE COMPROBACIÓN.....	43
FIGURA 14: CLIENTES DIVIDIDOS EN DECILES SEGÚN EL UP-LIFT PROMEDIO ORDENADO DE MAYOR A MENOR.....	43
FIGURA 15: UP-LIFTS POSITIVOS.....	44
FIGURA 16: IMPORTANCIA RELATIVA DE LAS VARIABLES DE UN ÁRBOL DE CLASIFICACIÓN PARA CLIENTES INFLUENCIABLES Y NO MOLESTAR.....	45
FIGURA 17: DIFERENCIAS EN MONTO GASTADO Y CANTIDAD DE COMPRAS DE LOS GRUPOS INFLUENCIABLES Y NO MOLESTAR.	47
FIGURA 18: CANJES Y PUNTOS CANJEADOS POR MES.....	50
FIGURA 19: CANJES Y PUNTOS CANJEADOS POR CATEGORÍA.	51
FIGURA 20: DISTRIBUCIÓN DE CANJES SEGÚN CATEGORÍA CON LIFT ASOCIADO.....	51
FIGURA 21: CANJES Y PUNTOS CANJEADOS POR GÉNERO.	52
FIGURA 22: CANJES Y PUNTOS CANJEADOS POR GSE.....	52
FIGURA 23: CANJES Y PUNTOS CANJEADOS POR LÍNEA.....	53
FIGURA 24: CANJES Y PUNTOS CANJEADOS POR LÍNEA RE-ESCALADO.....	54
FIGURA 25: CANTIDAD DE SKU'S DISTINTOS POR LÍNEA ESCOGIDA.	55
FIGURA 26: GRÁFICO DE CORRELACIÓN DE LAS VARIABLES CONTINUAS MÁS IMPORTANTES.	60
FIGURA 27: VARIABLES MÁS RELEVANTES PARA EL CANJE EN HERRAMIENTAS.....	62
FIGURA 28: GANANCIA DE INFORMACIÓN POR MODELO PARA EL CANJE EN HERRAMIENTAS.	63
FIGURA 29: VARIABLES MÁS RELEVANTES PARA EL CANJE EN ELECTRO.	63
FIGURA 30: GANANCIA DE INFORMACIÓN POR MODELO PARA EL CANJE EN ELECTRO.....	64
FIGURA 31: VARIABLES MÁS RELEVANTES PARA EL CANJE EN BEBIDAS ALCOHÓLICAS.	65
FIGURA 32: GANANCIA DE INFORMACIÓN POR MODELO PARA EL CANJE EN BEBIDAS ALCOHÓLICAS.....	65
FIGURA 33: VARIABLES MÁS RELEVANTES PARA EL CANJE EN ALIMENTOS.....	66
FIGURA 34: GANANCIA DE INFORMACIÓN POR MODELO PARA EL CANJE EN ALIMENTOS.....	66
FIGURA 35: VARIABLES MÁS RELEVANTES PARA EL CANJE EN TECNO.....	67
FIGURA 36: GANANCIA DE INFORMACIÓN POR MODELO PARA EL CANJE EN TECNO.	67
FIGURA 37: VARIABLES MÁS RELEVANTES PARA EL CANJE EN DORMITORIO Y DECORACIÓN.	68
FIGURA 38: GANANCIA DE INFORMACIÓN POR MODELO PARA EL CANJE EN DORMITORIO Y DECORACIÓN.....	69
FIGURA 39: VARIABLES MÁS RELEVANTES PARA EL CANJE EN MENAJE.....	69
FIGURA 40: GANANCIA DE INFORMACIÓN POR MODELO PARA EL CANJE EN MENAJE.	70
FIGURA 41: COMPILADO DE DESCRIPTIVOS PARA LÍNEAS DE CANJE.	94
FIGURA 42: CRECIMIENTO Y PROYECCIÓN DE LOS DATOS MÓVILES GLOBALES.	99
FIGURA 43: CAPACIDAD DE ALMACENAMIENTO DE INFORMACIÓN.....	99
FIGURA 44: DISTRIBUCIÓN DE LOS CANJES POR GÉNERO.	101
FIGURA 45: DISTRIBUCIÓN DE LOS CANJES POR DRIVER DE ENGANCHE PUNTOS.....	101
FIGURA 46: DISTRIBUCIÓN DE LOS CANJES POR DRIVER DE ENGANCHE TARJETA.....	102
FIGURA 47: DISTRIBUCIÓN DE LOS CANJES POR DRIVER DE ENGANCHE PROMOCIÓN.....	102
FIGURA 48: DISTRIBUCIÓN DE LOS CANJES POR SI ES ACTIVO EN LA TIENDA DE MEJORAMIENTO DEL HOGAR DEL GRUPO.....	103

FIGURA 49: DISTRIBUCIÓN DE LOS CANJES POR SI ES ACTIVO EN LA TIENDA POR DEPARTAMENTO DEL GRUPO.....	103
FIGURA 50: DISTRIBUCIÓN DE LOS CANJES POR SI TUVO INGRESO WEB EN LOS 3 MESES ANTES DEL CANJE.	104
FIGURA 51: DISTRIBUCIÓN DE LA EDAD DE LOS CLIENTES.	104
FIGURA 52: DISTRIBUCIÓN DE LA ANTIGÜEDAD DE LOS CLIENTES.	105

Índice de ilustraciones

ILUSTRACIÓN 1: RELACIÓN ENTRE LAS ÁREAS INVOLUCRADAS EN EL DESARROLLO DEL TRABAJO DE TÍTULO:.....	12
ILUSTRACIÓN 2: PROCESO KDD TÍPICO EN DETALLE.	22
ILUSTRACIÓN 3: RESULTADO ESPERADO DE UN MODELO UP-LIFT.	25
ILUSTRACIÓN 4: ESQUEMA DE MODELAMIENTO UP-LIFT.	26
ILUSTRACIÓN 5: DIAGRAMA DEL MODELAMIENTO INDIRECTO DEL UP-LIFT.....	26
ILUSTRACIÓN 6: DIAGRAMA DEL MODELAMIENTO DIRECTO DEL UP-LIFT.....	27
ILUSTRACIÓN 7: "CONTINGENCY TABLE" PARA CASOS BINARIOS.	29
ILUSTRACIÓN 8: FLUJO METODOLÓGICO FINAL DEL PROYECTO COMPLETO.....	33
ILUSTRACIÓN 9: FLUJO METODOLÓGICO FINAL DEL PROYECTO COMPLETO.....	79
ILUSTRACIÓN 10: ORGANIGRAMA DE LA EMPRESA.	84
ILUSTRACIÓN 11: ORGANIGRAMA GERENCIA DE INTELIGENCIA DE CLIENTES DE LA EMPRESA.	85
ILUSTRACIÓN 12: REGLAS DE ACUMULACIÓN DE PUNTOS.....	85
ILUSTRACIÓN 13: CONVERSIÓN DE PUNTOS A VALOR DE GIFT CARD.	85
ILUSTRACIÓN 14: EVOLUCIÓN DE CUENTAS ABIERTAS.....	86
ILUSTRACIÓN 15: CANTIDAD DE SUCURSALES Y FUNCIONARIOS DE LA ORGANIZACIÓN.	86
ILUSTRACIÓN 16: EVOLUCIÓN DE LA UTILIDAD DE LA ORGANIZACIÓN.	86
ILUSTRACIÓN 17: MÉTRICAS DE DESEMPEÑO	89
ILUSTRACIÓN 18: FLUJO DE DEFINICIÓN DE LA BASE ANALÍTICA PARA EL CÁLCULO DEL VALOR DE CANJEAR COMO JUSTIFICACIÓN... ..	89
ILUSTRACIÓN 19: CÁLCULO TAMAÑO MÍNIMO DE MUESTRA.	90
ILUSTRACIÓN 20: PIEZAS ENVIADAS AL TARGET 1.	90
ILUSTRACIÓN 21: PIEZAS ENVIADAS AL TARGET 2.	91
ILUSTRACIÓN 22: RESUMEN DEL DISEÑO EXPERIMENTAL.	100
ILUSTRACIÓN 23: INFOGRAFÍA DE NODOS MÁS Y MENOS PROPENSOS A SER INFLUENCIABLES	100

Índice de fórmulas

FÓRMULA 1: DISTANCIA DE MAHALANOBIS.	23
FÓRMULA 2: TAMAÑO MUESTRAL PARA DIFERENCIA DE PROPORCIONES.....	24
FÓRMULA 3: VALOR DE P EN LA ECUACIÓN DE TAMAÑO MUESTRAL.	24
FÓRMULA 4: FÓRMULA PARA CÁLCULO DEL LIFT DADA UNA ACCIÓN DE MARKETING	26
FÓRMULA 5: CÁLCULO DE LAS DISTINTAS DIVERGENCIAS PARA EL MODELAMIENTO UP-LIFT MEDIANTE ÁRBOLES ALEATORIOS.....	27
FÓRMULA 6: CÁLCULO DE DIVERGENCIA POR INTERACCIONES PARA EL MODELAMIENTO UP-LIFT POR ÁRBOLES ALEATORIOS.....	27
FÓRMULA 7: CÁLCULO DEL F-SCORE O F-MEASURE	30

1 Introducción

1.1 Antecedentes de la industria

El presente trabajo de título se realiza en una empresa que opera dentro de la industria de servicios financieros, que corresponde a una actividad comercial cuyo objetivo es prestar servicios de intermediación relativos a la generación de valor a través del dinero. Destacan las siguientes actividades dentro del rubro:

1. Bancos e instituciones financieras: Bancos nacionales, extranjeros, representaciones.
2. Oferta y contratación de seguros: Compañías de seguros generales, de vida, de crédito, corredores.
3. Administradoras de fondos: Generales, de pensiones, mutuos, para la vivienda, de inversión.
4. Mercado de valores: Bolsas de comercio y de valores, corredores.
5. Otros servicios financieros: Factoring, leasing, casas de cambio.
6. Servicios de apoyo: Administradoras de tarjetas de crédito, clasificadoras de riesgo, sociedades de cobranza, asesoría y consultoría financiera.

La organización se desempeña esencialmente en la oferta y operación de créditos, principalmente por medio de plástico crediticio. De hecho, corresponde a la mayor emisora de tarjetas de crédito de casa comercial de Sudamérica, que cuenta con operaciones en Argentina, Chile, Colombia y Perú.

La industria local en la que está inmersa la empresa cuenta con la participación de tres tipos de entidades:

Tabla 1: Participantes de la industria de servicios financieros chilena con tarjetas de crédito activas a febrero de 2018

Bancarias	Sociedades de apoyo al giro bancarias	No Bancarias
Banco Santander	Consortio Tarjetas de Crédito S.A.	Promotora CMR Falabella S.A.
Banco de Chile	CAR S.A. – Tarjeta Ripley	Presto S.A.
Banco del Estado de Chile	CAT Administradora de tarjetas S.A.	LP S.A.
Banco de Crédito e Inversiones		Crédito Organización y Finanzas S.A.
Banco Bilbao Vizcaya Argentaria		Inversiones y Tarjetas S.A.
Banco Itaú Corpbanca		Tricard S.A.
Banco Scotiabank		
Banco Security		
Banco BICE		
Banco Internacional		
Coopeuch		

Fuente: Elaboración propia basado en información de la Superintendencia de Bancos e Instituciones Financieras (SBIF)

El principal servicio que ofrecen dichas entidades es el crédito mediante sus plásticos crediticios, cuya participación se observa en el siguiente gráfico:

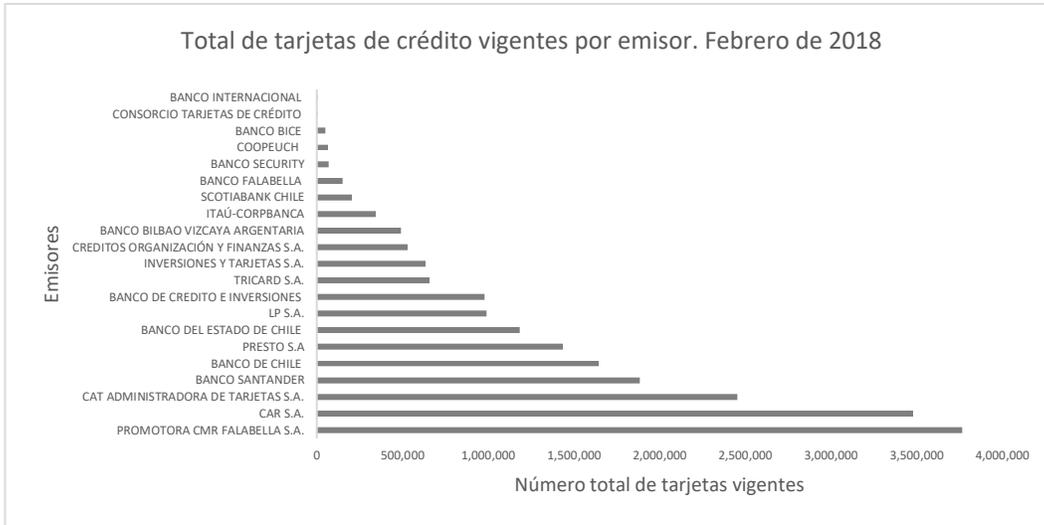


Figura 1: Cantidad de tarjetas de crédito por emisor a febrero de 2018.

Fuente: Elaboración propia basado en información de la Superintendencia de Bancos e Instituciones Financieras (SBIF)

Es además relevante estudiar la cantidad de dinero que se transa o se gasta con cada una de las tarjetas de crédito presentadas, que se muestra en el siguiente gráfico:

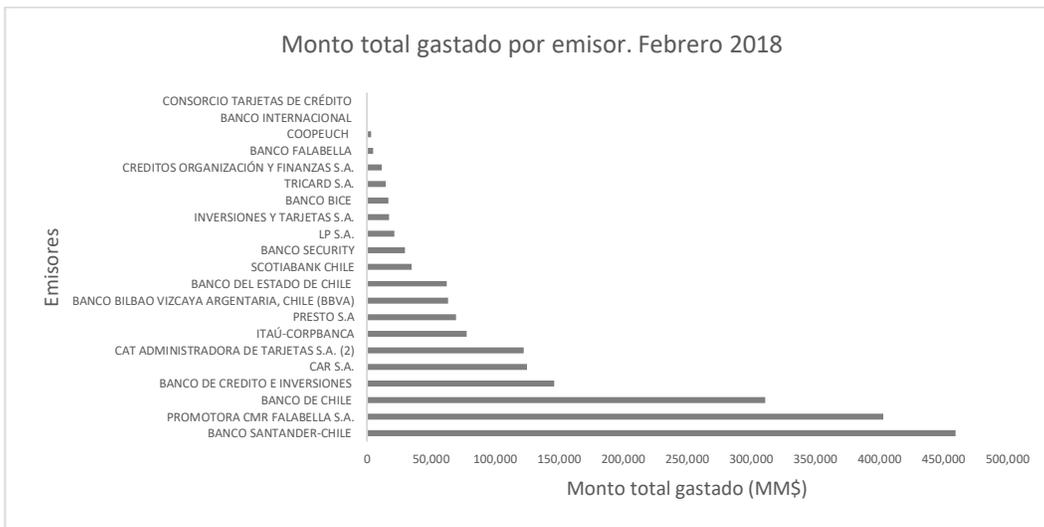


Figura 2: Monto gastado asociado a tarjetas de crédito por emisor a febrero de 2018.

Fuente: Elaboración propia basado en información de la Superintendencia de Bancos e Instituciones Financieras (SBIF)

En las dos figuras anteriores, se puede observar que, a pesar de que los emisores no bancarios o de apoyo al giro tienen en general una mayor cantidad de tarjetas vigentes, todavía están muy por debajo de la cantidad de dinero que mueven las entidades bancarias, por lo que su negocio se concentra en llegar a más clientes, pero de menor valor, lo que se condice con la propuesta de valor inicial de las tarjetas de crédito de casa comercial, que consistía en ofrecer acceso a crédito a los clientes que, por diversos motivos, no lo tenían en los bancos tradicionales.

De lo expuesto anteriormente, específicamente en la figura 1, se indica que existen más de 21 millones de tarjetas de crédito activas en Chile a febrero de 2018, lo que decanta en que las personas están teniendo más de una tarjeta de crédito, por lo que tienen distintas opciones a utilizar, de tal manera que las compañías buscan diferenciarse e incentivar a que sus clientes los prefieran por sobre las otras alternativas que poseen.

Así es como los programas de fidelización o de lealtad toman especial importancia, persiguiendo que, en igualdad de todas las condiciones adicionales, un cliente prefiera una empresa por sobre otra, debido a los beneficios extra que puede ofrecerle el plan de fidelización subyacente. En consecuencia, todos los emisores de tarjetas de crédito de Chile tienen, de alguna u otra manera, un club de fidelización, ya sea en forma de plan como CMR Puntos, millas LATAM con Santander, dólares premio del Banco de Chile, Todo Suma del Banco Estado, o simplemente con beneficios al utilizar la tarjeta propia en ciertas tiendas, tratos preferentes en los servicios asociados, etcétera.

1.2 Caracterización de la empresa

La empresa corresponde a una filial de una de las mayores compañías de comercio integradas de América Latina, que posee operaciones en Chile, Perú, Colombia, Argentina, Brasil, México y Uruguay.

Dicho holding tiene participación en 5 unidades de negocio:

- Tiendas por departamento
- Mejoramiento del hogar
- Supermercados
- Servicios financieros
- Inmobiliaria

Como se mencionó en el punto anterior, la compañía es parte del sector de servicios, ya que corresponde a una organización financiera que ofrece una tarjeta de crédito clásica, cerrada para las empresas del grupo y comercios asociados, la versión Visa en sus distintas modalidades, también de la tarjeta MasterCard, que va en retirada pero aún tiene presencia, además de productos como avances en efectivo y la posibilidad de acceder a productos y beneficios especiales mediante la acumulación y canje de los Puntos, que componen el programa de fidelización.

Un dato importante con relación a la participación de mercado de la organización donde se desarrolló el trabajo de título es que al año 2016, alcanzó un 13,3% de participación de mercado en las ventas del operador Transbank y llegó a tener más de 2,8 millones de cuentas abiertas, con casi dos millones de ellas correspondientes a la versión Visa.

Con respecto a su relación con el mercado y la evolución en el tiempo, la empresa cuenta con más de 37 años en el mercado chileno, logrando convertirse en el mayor emisor de tarjetas de crédito en el mercado nacional, persiguiendo el objetivo de otorgarle a sus clientes un sistema de pago más cómodo y flexible.

En el ámbito internacional, también opera en Argentina y está presente en Perú y Colombia a través del Banco del grupo. En el periodo 2008 a 2009, se estableció una alianza con Visa en Perú y Chile, abriendo la tarjeta y permitiendo su uso en locales fuera de las empresas del

holding. Más adelante, en 2013, se llegó a un acuerdo para formar alianza con MasterCard para Argentina y Colombia.¹

Es importante mencionar los elementos centrales que mueven a la empresa en su trabajo diario y a largo plazo son:

- Misión: *“Contribuir a mejorar el estándar de vida de los clientes, mediante la oferta integrada de servicios financieros más convenientes del mercado, potenciada por los beneficios del holding”*
- Visión: *“Estamos comprometidos con hacer realidad los sueños de cada uno de nuestros clientes, con un equipo humano entusiasta, que busca sorprender a través de servicios financieros simples y convenientes. En conjunto con nuestras tiendas, aumentamos el valor de nuestra cartera de clientes, haciendo de nuestra tarjeta el principal medio de pago y la alternativa preferida en financiamiento”*²

1.3 Área de desarrollo del trabajo de título

La organización se estructura mediante la definición de las 8 áreas señaladas en el organigrama (Ver Anexo 2: Estructura organizacional), cuyas principales funciones son:

- Gerencia de Riesgo y Cobranza: Se preocupa de velar porque los clientes cumplan las condiciones necesarias para obtener alguno de los productos financieros que se ofrecen, además de gestionar las acciones relativas a cobros y pagos.
- Gerencia de Recursos Humanos: Como en toda empresa, tiene que ver con el manejo de los trabajadores y colaboradores.
- Gerencia de Marketing y Productos: Área de la que dependen todas las actividades de marketing y gestión de productos de la compañía, en ésta se incluye el plan de fidelización, las relaciones con los distintos proveedores, la generación de campañas promocionales y publicitarias, etc.
- Gerencia de Administración y Finanzas: Dice relación con la parte financiera de la compañía.
- Gerencia de Canales: Esta área tiene como finalidad el velar porque los medios de contacto entre el cliente y la compañía funcionen correctamente, además de generar mejoras a las experiencias del usuario en todas las plataformas que ofrece.
- Gerencia de Operaciones y Sistemas: Se preocupa del correcto funcionamiento de todos los procesos que se deben llevar a cabo en la empresa.
- Gerencia Contraloría: Encargada de desarrollar e implementar planes de auditoría.
- Gerencia de Business Intelligence: Se dedica a analizar los datos que posee la compañía y encontrar oportunidades de negocio en ellos.

En lo que al trabajo de título respecta, se realiza en la Gerencia de inteligencia de negocios, pero como cliente se tiene a la Gerencia de Fidelización, que depende de la Gerencia de Marketing y Productos, al igual que:

- Gerencia de Productos.

¹ Servicios financieros, página 16, Memoria anual holding año 2016

² Administración, sitio web de la empresa – Quienes somos. Revisado el 26 de agosto de 2017

- Gerencia de Marketing.
- Gerencia de Retail.
- Gerencia de Planificación y Proyectos.

1.4 Descripción del proyecto y justificación

1.4.1 Identificación del área

Como se mencionó anteriormente, la memoria para optar al título de Ingeniero Civil Industrial se realiza en el área de Inteligencia de Clientes, dependiente de la Gerencia de Business Intelligence (Ver Anexo 3: Estructura del área), y que tiene como objetivo obtener, gestionar y analizar los datos generados por la compañía, de manera tal de poder proponer mejoras y encontrar patrones que indiquen oportunidades de negocios. Sin embargo, el tema tratado tiene como contraparte al área de Fidelización de Clientes, vale decir, el área que se preocupa de administrar todo lo relativo al plan de fidelización que posee el holding, que se traduce en los *puntos*. Así, el área de fidelización de clientes es la encargada de manejar la acumulación de *puntos*, canjes, campañas promocionales, categorías de clientes, beneficios, entre otras actividades. Esto de cara tanto a los clientes externos como a los internos, vale decir, las otras áreas del holding, como por ejemplo de cara a los ejecutivos de sucursales, a las tiendas que ofrecen sus productos para canjes, a las alianzas con otras empresas, etc.

A continuación, se presenta un diagrama que resume lo explicado anteriormente.

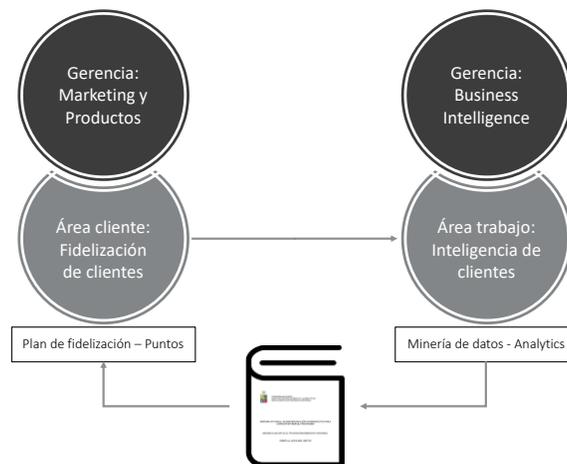


Ilustración 1: Relación entre las áreas involucradas en el desarrollo del trabajo de título:

Fuente: Elaboración propia

1.4.2 Descripción del problema

Actualmente en la organización se cuenta con 2,8 millones de clientes con cuenta activa, de los cuales 60% no ha canjeado nunca, de ese porcentaje, 13% tiene hoy en día más de 5.000 puntos, es decir, está más arriba del piso mínimo de canje³.

³ Basado en información de la empresa a junio de 2018

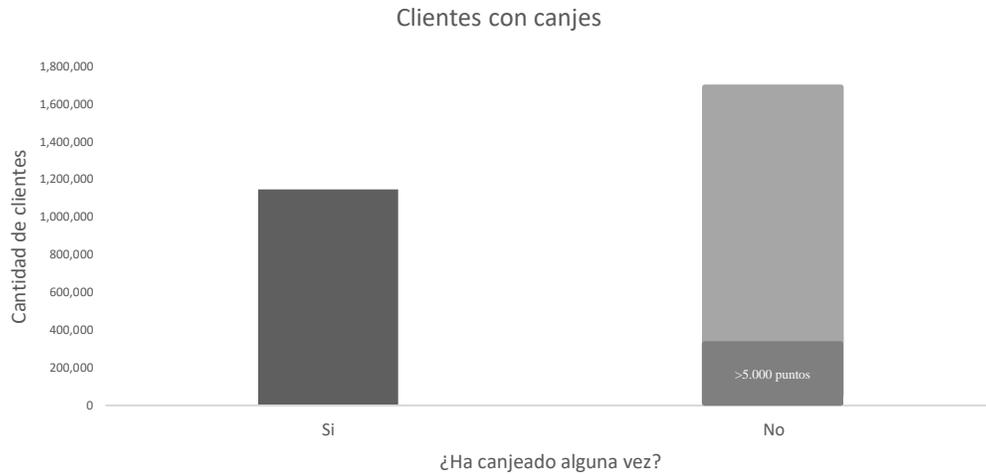


Figura 3: Distribución de los clientes según comportamiento de canje.

Fuente: Elaboración propia con datos de la empresa a junio de 2018

Es importante destacar también que, de los clientes que han canjeado, el mayor porcentaje se concentra en solo un canje, disminuyendo la frecuencia de ocurrencia a medida que aumenta la cantidad de canjes, como se puede observar en la Figura 4.

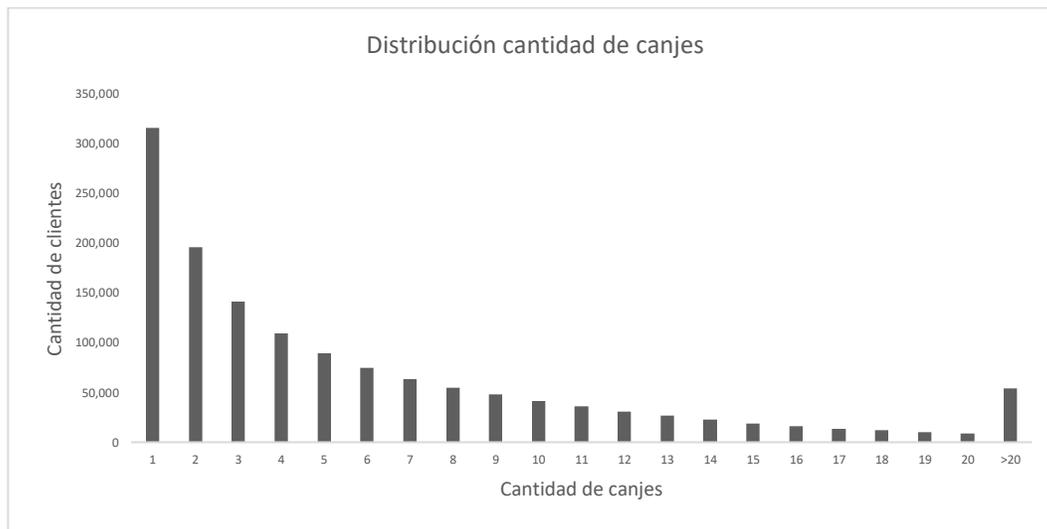


Figura 4: Distribución de clientes según cantidad de canjes en su vida como cliente.

Fuente: Elaboración propia con datos históricos de la empresa, desde 2010 a junio de 2018

Así, la organización tiene la necesidad de aumentar la relación de los clientes con su plan de fidelización, de manera tal de llegar a generar tanto primeros canjes en clientes que no tienen, como más canjes en los que han tenido muy pocos en su vida. Ya que, como se muestra más adelante, el canjear tiene efectos positivos en el comportamiento del cliente con la tarjeta de la empresa.

1.4.3 Justificación

1.4.3.1 Valor de canjear

En la organización se tiene como objetivo último el que sus tarjetas sean las principales de los clientes⁴, vale decir, que usen la tarjeta de crédito de la empresa para todo. Dicho esto, el plan de lealtad de la compañía busca contribuir a que la experiencia del cliente sea única, asociada al canje de productos mediante los *puntos*, pero también a incentivar el uso de la tarjeta en las tiendas del grupo, para el caso de tarjeta cerrada, además los distintos rubros asociados y cualquier otro uso que se le pueda otorgar a la tarjeta.

En este contexto, se afirma dentro de la organización que un cliente que ha canjeado al menos una vez es más “educado” como cliente que uno que no lo ha hecho nunca, ya que un cliente más educado tiende a usar más su tarjeta, conoce la operación y los beneficios, en definitiva, es un cliente más atractivo y se denomina un cliente “enganchado”.

Asimismo, en un estudio interno de la empresa, se dejó en evidencia que existen tres formas de que un cliente esté enganchado con la empresa:

- Enganchado por financiamiento
- Enganchado por promoción
- Enganchado por puntos

Concluyendo que el cliente enganchado por puntos tiene mayor cantidad de transacciones en promedio al año (más de 90 frente a 15 transacciones promedio de los clientes no enganchados), además de que su gasto en promedio es superior a \$3.000.000 al año, mientras que el de un cliente no enganchado es de \$500.000 (Gallardo, 2016).

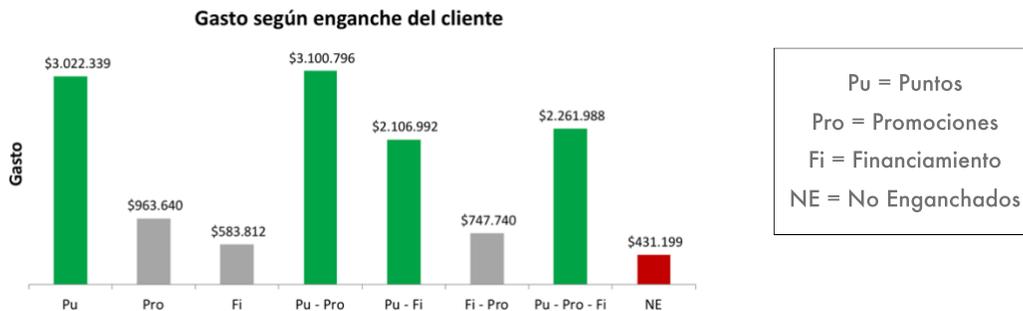


Figura 5: Gasto según driver de enganche del cliente.

Fuente: Gallardo, 2016

Con el fin de complementar el análisis existente, se realizó un estudio que pretendía determinar si existen diferencias entre los clientes que canjean y los que no canjean, a diferencia del anterior que se enfocaba en enganche por puntos. En este contexto, se procedió a definir una base analítica que consideraba a los clientes que cumplían con todas las características siguientes:

⁴ Principalidad: Porcentaje de veces que el cliente prefiere la tarjeta propia por sobre el total de veces que elige una tarjeta como medio de pago.

- Clientes que abrieron su cuenta durante el año 2014: Con el objetivo de contar con clientes que iniciaron su relación con la empresa en un periodo de tiempo similar y acotado.
- Clientes que no hayan canjeado en el año 2014 ni en el 2015: Esta restricción tiene por objetivo dejar un tiempo razonable para que el cliente acumule puntos, antes de estudiar su comportamiento de canje. Este tiempo razonable se estima que son 20 meses en promedio⁵.
- Clientes con su cuenta activa a diciembre de 2017: Para asegurar que el cliente se haya mantenido como tal sin bloqueos de tarjeta asociado a morosidad.

Ver diagrama de definición de la base en Anexo 11: Flujo de definición de base analítica.

Dicho esto, se separó la base en deciles según el gasto en 2015 y se marcó a los clientes que canjearon en el año 2016, para comparar su gasto con los clientes que no canjearon, pero pudieron hacerlo, es decir, llegaron a tener más de 5.000 puntos⁶. Los resultados del análisis se resumen en Tabla 2.

Tabla 2: Diferencia de gasto según deciles de gasto en 2015

Decil según gasto en 2015	Diferencia de gasto 2015	Diferencia de gasto 2016	Diferencia de gasto 2017
1	247.663	889.034	813.331
2	6.884	411.692	342.505
3	-1.615	200.971	170.563
4	-885	83.851	52.148
5	394	39.896	46.027
6	-1.411	-23.725	-14.127
7	1.152	-72.205	-77.376
8	-480	-23.735	-79.226
9	325	-13.827	-82.934
10	3.778	-57.178	-271.386
Diferencia total	255.805	1.434.776	899.525

Fuente: Elaboración propia según análisis propio realizado con datos de la empresa

Se observa que, tanto en el año 2016 como 2017, existe un aumento del gasto promedio de los clientes, lo que corresponde al gasto incremental que tienen los clientes que canjean en el periodo 2016 versus los que no lo hicieron, siendo que llegaron a niveles para hacerlo. Esta diferencia es, si la sumamos entre los deciles, de \$1.434.776 en el año 2016, y en promedio es \$143.477, mientras que para el año 2017 la diferencia promedio es de \$89.953, lo que puede ser un número más representativo que la suma de las diferencias en el gasto.

⁵ Análisis “Efecto puntos”, Inteligencia de Clientes de la empresa, junio de 2016

⁶ Nivel más bajo de canje, a pesar de que se puede canjear en niveles menores, el de 5.000 es el formalmente establecido más bajo

1.4.3.2 Gestión de clientes

Adicionalmente, en el área de gestión de clientes, se tiene una situación que aumenta la importancia del trabajo propuesto, ya que actualmente:

- Se selecciona el target de campañas de manera fundamentalmente manual y sin incorporar variables más allá de los puntos actuales de los clientes, cercanía a una tienda, entre otras marcas simples.
- No se proponen productos para que comercial los considere, sino que se espera que comercial diga qué productos se ofrecerán.
- No se realizan análisis para determinar si un cliente debería recibir un mail de canje o no.
- Se tiene una tasa de des-suscripción por los mails enviados de un 0,02%, que se traduce en 33.000 clientes que dejan de ser contactables al año⁷.

Si bien no hay un análisis formal de cuán mejor es un cliente contactable vía mail versus uno que no lo es, hay oportunidades de gestión que se pierden y que son relevantes para construir y/o mantener una buena relación con el cliente, ya que se pierde la posibilidad de sostener una comunicación fluida y eficiente. Además, como se observa en la Figura 6, un 40% de los clientes no es contactable actualmente, cifra que aumenta cada vez que un cliente se des-suscribe producto de un mail recibido.

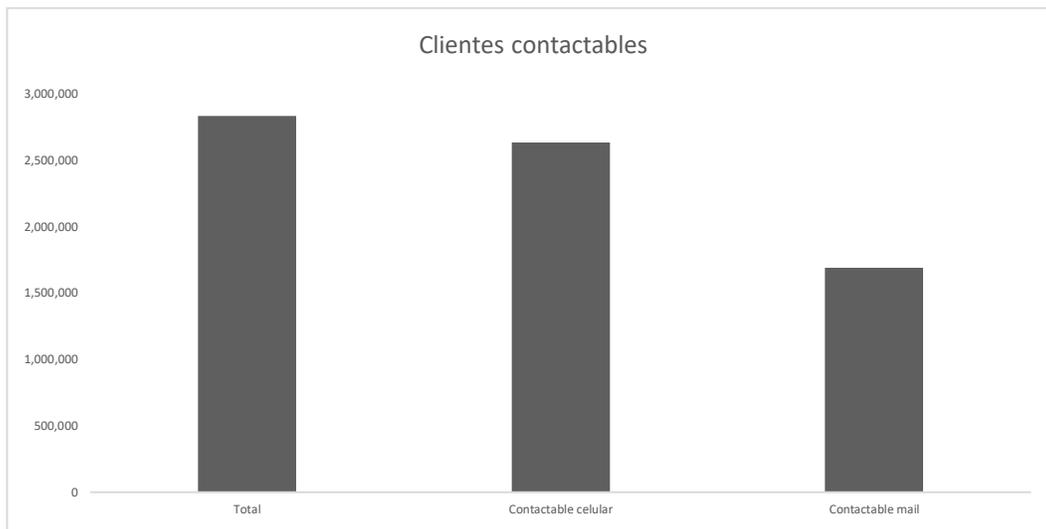


Figura 6: Cantidad de clientes contactables vía mail y teléfono.

Fuente: Elaboración propia con datos de la empresa a junio de 2018

1.4.3.3 Personalización del plan de puntos

El plan de puntos de la organización involucra distintos tipos de “premios” asociados a la acumulación de puntos, que se calcula mediante la relación: un punto cada \$150 gastados

⁷ Información obtenida en base a todos los envíos de marketing del año 2017 de la empresa.

con la tarjeta en empresas del grupo y un punto cada \$300 en otros comercios. Además de un punto cada \$300 gastados con la tarjeta del banco en empresas del grupo y uno cada \$600 en otros comercios (Detalles en Anexo 4: Reglas de acumulación).

Los puntos pueden ser canjeados por:

- *Gift Cards* (Ver tabla en Anexo 5: Tabla de equivalencia de puntos)
- Productos
- Experiencias
- Viajes

Como se puede observar en la Figura 7, el canje de *Gift Cards* representó en 2017 un 65% de los puntos canjeados en total aproximadamente y el 40% de los canjes, sin embargo, el canje de los otros elementos descritos es sumamente relevante en el funcionamiento del programa, dado que el canje de productos atractivos para el cliente, sin involucrar un gasto directo, provoca un lazo emocional con la empresa que se quiere mantener y potenciar⁸.

Así, la oportunidad que surge es la de reemplazar la forma de elección de los productos para canje vía mail, que hoy en día se realiza semanalmente (por parte del área comercial) de acuerdo a la cantidad de puntos actuales acumulados del cliente y un par de niveles hacia arriba, sin embargo, no hay un estudio acabado del comportamiento de los clientes frente al canje, de manera tal de que los mails que se están enviando no están generando la potencial tasa de respuesta que se podría alcanzar si se incorpora inteligencia de clientes al análisis. Posteriormente a la planificación del área comercial, ésta es enviada al área de inteligencia para que, según los lineamientos básicos dispuestos por comercial, se escoja la base de clientes adecuados para recibir el mail, de manera manual y sin un modelo de inteligencia de clientes o de minería de datos por detrás.

Un paso previo a los detallados es el de escoger los productos a ofrecer para ser negociados con la contraparte que corresponda, situación que puede tener éxito o no, por lo que es muy dependiente del criterio del área comercial, de los productos que parezca relevante negociar y del resultado de la negociación.

Cabe destacar que la organización invierte cerca del 10% de sus costos totales en el club de fidelización, específicamente 24 mil millones de pesos al año⁹, por lo que es bastante relevante observar y abordar temáticas relativas al plan de fidelización de la empresa.

Finalmente, como se mencionó anteriormente, *gift card* concentra el 64% de los puntos canjeados al año y casi el 40% de los canjes:

⁸ Según los lineamientos de la Gerencia de Fidelización de Clientes.

⁹ Estados financieros de la empresa año 2016.

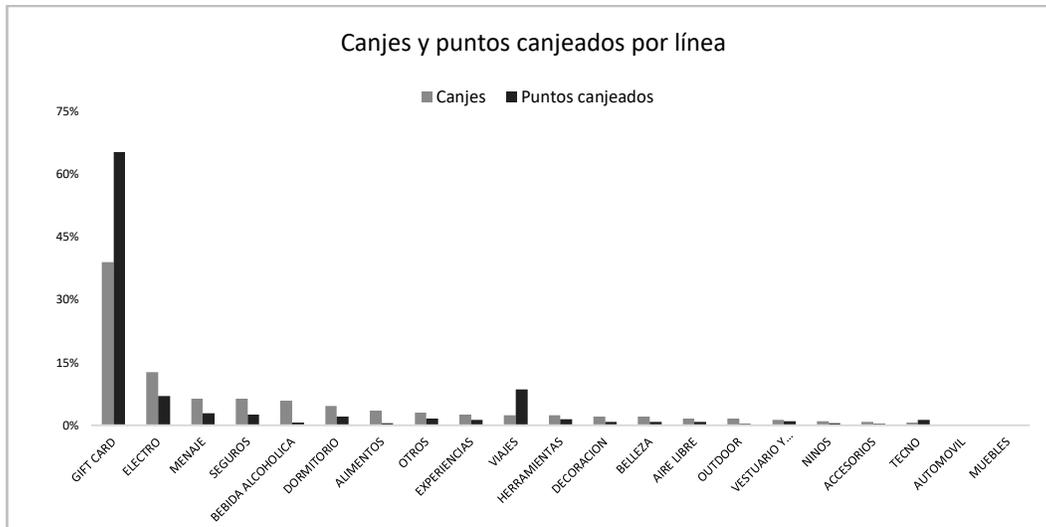


Figura 7: Distribución de la cantidad de canjes en líneas de premios.

Fuente: Elaboración propia con datos de la empresa de el año 2017

Sin embargo, según expresa la Gerencia de Fidelización de Clientes de la empresa, canjear un producto es más personal y genera mejores impactos en los clientes como plan de fidelización, además de entregar más rentabilidad a la empresa debido a las negociaciones que se realizan con los proveedores, en las que usualmente se logran mejores márgenes que en una *gift card* que puede ser utilizada en cualquier producto de las tiendas del holding. Por lo que es relevante potenciar el canje de productos, más aún si es en niveles bajos, ya que tiene un costo menor para la empresa en comparación con los niveles más altos de canje (Ver Anexo 5: Tabla de equivalencia de puntos).

1.4.4 Descripción del proyecto

Considerando los antecedentes antes expuestos, es que se propone un sistema integral de recomendaciones de productos para canje. Es integral porque no solamente recomienda un producto según la propensión a canjear, sino que lo hace de manera secuencial:

1. Se determina si un cliente debería recibir un mail de canje hoy en día, según su respuesta esperada al estímulo.
2. Dado que se espera una respuesta positiva al mail de canje, se establece qué línea de producto es más propenso a canjear.
3. Se envía un mail proponiendo canjear productos de la línea con mayor propensión de canje. (Ver: Ilustración 8: Flujo metodológico final del proyecto completo).

El proyecto a desarrollarse como trabajo de título busca resolver dos grandes preguntas:

- ¿A quién le envió un mail de canje?
- ¿Qué productos le envió?

Para dilucidar dichas preguntas, se propone la siguiente metodología:

- ¿A quién le envió un mail de canje?: Para determinar la sensibilidad de un cliente en particular a la recepción de un mail de canje, se desarrollará un modelo Up-Lift, que mide específicamente el efecto de una acción de marketing sobre un cliente. Dicho modelo, requiere de una experimentación previa, la que alimentará posteriormente al modelo, indicando finalmente a qué clientes se debería enviar un mail de canje.
- ¿Qué productos le envió?: Se estimará la propensión o probabilidad de un cliente de escoger cierta línea de productos, considerando como variables los productos canjeados, la acumulación, el tipo de tarjeta, la contactabilidad, demográficas, transaccionales, entre otras variables.

Se abordará de manera separada cada problemática, con el objetivo de generar una metodología que sea replicable a otros sub-negocios del holding que se comunican vía mail, pero que no necesariamente involucran el ofrecimiento de un producto para canje, como por ejemplo las alianzas. Además, en la organización actualmente no existe un modelo que entregue una categorización del cliente dada su respuesta esperada a un mail, a pesar que este es uno de los canales más explotados. Por esto, es útil generar el conocimiento de este tipo de modelos y sus aplicaciones a mailing, conocimiento que no se ganaría si se desarrolla un modelo en conjunto para recomendar productos a canjear.

1.5 Objetivos

1.5.1 Objetivo general

“Desarrollar una metodología que permita determinar a qué cliente enviar mail de canje en un retail financiero y qué línea de producto se le envía, de tal manera de aumentar la tasa de respuesta de los envíos”

1.5.2 Objetivos específicos

- Determinar qué variables inciden en la respuesta esperada a un mail de canje y en la probabilidad de canje de puntos en una línea.
- Diseñar e implementar un experimento que permita clasificar a los clientes según su respuesta a mail de canje.
- Desarrollar un modelo que determine si se debe enviar un mail de canje a un cliente particular, según su respuesta esperada.
- Elaborar un modelo de recomendación de líneas de productos de canje, según la propensión a canjear.
- Decidir criterios de entrega de mails de canje, considerando a quién y qué le envió.

1.5.3 Alcances

- Considerando las restricciones más macro, el análisis y las propuestas se realizarán para las operaciones de la organización en Chile.

- Para el caso del modelo Up-Lift, se trabajará con los resultados obtenidos del piloto experimental, con las restricciones de escalabilidad asociadas a la base de clientes escogida. Mientras que para el estudio de las líneas de canje, se trabajará con los datos de canje comprendidos entre el 01 de enero de 2017 y el 31 de diciembre de 2017.
- Como se trabajará con el programa de fidelización, solo se incluirá a los clientes que pertenezcan a éste, vale decir, que tengan puntos acumulados y que además cumplan con:
 - Ser cliente contactable de la empresa, ya que se busca realizar el modelo de recomendación para enviarle ofertas de productos vía mail, lo que no se puede lograr si el cliente no es contactable mediante este medio.
 - Ser titulares, ya que los clientes adicionales no pueden usar el plan de puntos.
 - Poseer datos suficientes para el análisis, vale decir, que esté correctamente identificado con las variables relevantes.
- Los modelos aplicados serán evaluados únicamente con una partición de la base destinada a la validación, es decir, no se implementarán experimentos posteriores para evaluar la efectividad de los algoritmos utilizados en cuanto a la recomendación de líneas de productos.
- Se considerará una cantidad limitada de líneas de productos, que explique la mayor cantidad de los canjes pero que sea manejable dentro del modelo, además de ser relevantes para la gerencia de fidelización y gestionables por la empresa.
- Solo se escogerán productos a ofrecer de manera distinta a la actualmente utilizada, no se modificará el plan de puntos, los niveles de canje, las formas de canjear, etc.
- No está dentro del alcance el diseño y la generación de las piezas gráficas a enviar si se desea hacer pilotos para testear el modelo a posteriori.

2 Marco teórico

2.1 Minería de datos

La minería de datos corresponde a una ciencia que persigue como objetivo primordial el extraer conocimiento útil a partir de bases de datos de gran volumen.

Dicha ciencia surgió como un campo interdisciplinario desde las ciencias de la computación y la estadística.

Actualmente la minería de datos se ha adaptado a distintas aplicaciones, como en el ámbito de la manufactura, de la ciencia, la ingeniería, problemas y desafíos gubernamentales, incluso en el área de los servicios, entre otros. Además de las numerosas aplicaciones que se han ido desarrollando y adquiriendo popularidad, el auge de la minería de datos se ha debido, entre otros factores relevantes, al aumento exponencial de los datos disponibles para trabajar en el último tiempo, el incremento en la capacidad de almacenamiento de éstos y la consecuente disminución de los costos de guardar la información (Tsoumakas, Katakis, & Vlahavas, 2010) (Ver: Anexo 21: Crecimiento de datos móviles y capacidad).

De lo anteriormente expuesto, se puede concluir que, fuertemente apoyada por la coyuntura en cuanto al desarrollo de las tecnologías relativas a la generación y almacenamiento de datos en la última década, la minería de datos ha tomado particular peso y relevancia dentro de las industrias en las que puede ser aplicada, sin embargo, de manera más práctica, la minería de datos es una etapa de un proceso analítico mayor.

2.1.1 Metodología KDD

La minería de datos como motor para generar conocimiento a partir de grandes bases de datos, está inmerso dentro de un proceso más grande y complejo denominado KDD, por sus siglas en inglés de Descubrimiento de Conocimiento en Bases de Datos, donde la minería de datos en sí se constituye como un paso dentro de una metodología que, dependiendo de la fuente y el objetivo del análisis, consta de mínimo cinco etapas para llevar a cabo el proceso completo.

Dicho esto, KDD es un proceso automático de análisis exploratorio de datos y generación de modelos a partir de bases de datos de gran volumetría. En otras palabras, es el proceso organizado, complejo y no trivial de identificar patrones previamente desconocidos, potencialmente útiles, válidos, nuevos y entendibles a partir de bases de datos grandes y complejas (Tsoumakas et al., 2010).

Los pasos que usualmente se recorren en un proyecto realizado mediante la metodología KDD se observan en la Ilustración 2.

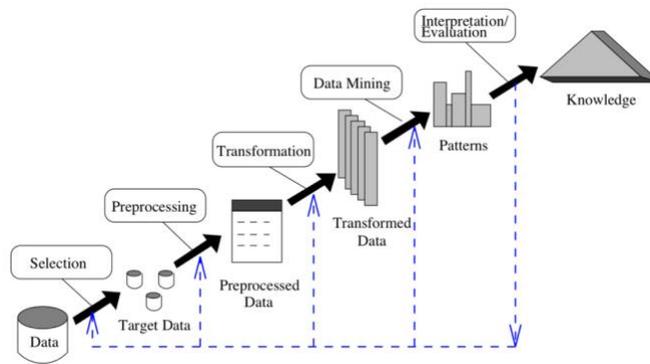


Ilustración 2: Proceso KDD típico en detalle.

Fuente: (Sumathi & Sivanandam, 2007)

Como se puede ver en la figura anterior, la etapa de minería de datos es el núcleo del proceso KDD, que además cuenta con pasos previos orientados a preparar la base de datos de la mejor manera posible, de tal forma de entregarle al modelo de minería una base ordenada, completa, coherente y con tipos de variables y escalas razonables y manejables, mientras que los pasos posteriores corresponden al análisis que se puede realizar de los patrones encontrados a partir de los modelos generados previamente, de donde se espera que se desprenda el conocimiento adquirido.

Es importante destacar que la metodología KDD tiene una marcada naturaleza interdisciplinaria, ya que nace de la intersección de campos de conocimiento tales como machine learning, identificación de patrones, bases de datos, estadística, inteligencia artificial, visualización de datos y computación de alto rendimiento (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

2.1.2 Preparación de los datos

A pesar de los importantes avances en minería de datos y sus técnicas evidenciados en los puntos anteriores, dicho progreso no se ha visto replicado en la preparación de los datos.

Este punto es muy relevante porque, además de ser el input de los modelos de minería que se desarrollarán, y por lo tanto datos de alta calidad llevan a obtener patrones de mayor calidad, es un proceso que usualmente toma más tiempo que el desarrollo mismo de los modelos y análisis (Zhang, Zhang, & Yang, 2010).

El proceso de preparación de los datos o preprocesamiento tiene distintas etapas que pueden aplicar o no para ciertos problemas específicos (Tsoumakas et al., 2010):

- Datos perdidos
- Extracción de atributos
- Selección de atributos
- Transformación de datos
- Detección de datos fuera de rango

2.1.3 Detección de datos fuera de rango

Existen métodos o herramientas muy poderosas para detectar datos fuera de rango (outliers) cuando los datos contienen solo un valor fuera de rango, poder que disminuye de manera drástica si existen más observaciones outlier en los datos.

Dicha pérdida o disminución del poder va de la mano con que no siempre se logra detectar los valores fuera de rango, debido a que los outliers son afectados por las mismas observaciones que supuestamente se deberían identificar.

Debido a lo anterior es que se propone la distancia de Mahalanobis, que corresponde a una medida de distancia entre cada uno de los datos en comparación con un estimador C relativo a una métrica de dispersión V . Cuando los valores de C y V corresponden a la media aritmética y a la matriz de covarianza de los datos, se da origen a la distancia de Mahalanobis:

$$MD_i = D_i(\bar{X}, S) = \sqrt{\{(x_i - \bar{X})^T S^{-1} (x_i - \bar{X})\}}, \quad i = 1, \dots, n.$$

Fórmula 1: Distancia de Mahalanobis.

Fuente: (Hadi, 1992)

Donde:

- \bar{X} = **Media aritmética**
- x_i = **Observación i**
- S = **Matriz de covarianza**

A partir de dicha distancia se realiza un proceso iterativo que culmina con un criterio de parada.

Finalmente, realizado todo el proceso, se obtiene un sub conjunto de datos que se presume son outliers, los que se testean mediante un test chi-cuadrado comúnmente utilizado al 1% (Hadi, 1992).

2.2 Experimentación

2.2.1 Diseño experimental

Remontándose a la década del 30, ya se sostenía que los experimentos eran muy importantes a la hora de obtener información sobre algún tema, de hecho quien es considerado el padre del diseño experimental moderno, Ronald Fisher, decía que los experimentos son “solo una experiencia cuidadosamente planificada con anticipación y diseñada para crear una base sólida de conocimiento”.

Hasta el día de hoy los experimentos son de las formas más aceptadas para generar conocimiento, ya sea sobre un tema poco explorado o ampliar el que se tiene de un tema conocido.

En este contexto, los experimentos corresponde a la manipulación de una o más variables independientes seguido de una observación y estudio sistemático de los efectos que tiene dicha manipulación en una o más variables dependientes.

El diseño experimental involucra el desarrollo de 5 actividades interrelacionadas:

1. Formulación de hipótesis.

2. Determinación de los niveles de tratamiento (Variables independientes), la métrica (Variable dependiente) y condiciones externas a ser controladas (Variables de ruido).
3. Determinación del tamaño de muestra.
4. Determinación del proceso que dará origen a los grupos tratamiento y control.
5. Definición del análisis estadístico que se llevará a cabo. (Kirk, 2015)

2.2.2 Tamaños de muestra

Cuando se realiza un estudio experimental de cualquier índole, siempre se tiene implícita la tarea de determinar el tamaño de muestra necesario para que la ejecución de este sea lo más perfecta posible. El hecho de no realizar este paso previo puede decantar en dos consecuencias (Fernández, 1996):

- En primer lugar, se puede llevar a cabo un experimento sin el número apropiado de sujetos para el análisis, de manera tal de que no se podrán estimar los parámetros ni encontrar diferencias significativas incluso existiendo.
- En segundo lugar, se puede desarrollar el experimento con un número innecesario de sujetos, malgastando tiempo y recursos que podrían haber sido ahorrados.

Específicamente, cuando se quiere comparar si las medias o proporciones de las muestras son distintas, se necesitan conocer los siguientes parámetros:

- Magnitud de la diferencia a detectar, en el caso del presente trabajo de título, las tasas de respuesta esperadas.
- Tener una idea aproximada de los parámetros de la variable a estudiar.
- Seguridad del estudio, es decir, el nivel de confianza, que representa el riesgo de cometer un error tipo I.
- Poder estadístico, que representa el riesgo de cometer un error tipo II.
- Definir si la hipótesis será unilateral (de una cola), es decir, exigir que sea en una dirección la diferencia, o bilateral (de dos colas), vale decir, que los parámetros difieran en alguna dirección no establecida previamente.

En el caso de estudiar las diferencias entre dos proporciones, el tamaño muestral se calcula como:

$$n = \frac{[Z_{\alpha} * \sqrt{2p(1-p)} + Z_{\beta} * \sqrt{p_{GT}(1-p_{GT}) + p_{GC}(1-p_{GC})}]^2}{p_{GT} - p_{GC}}$$

Fórmula 2: Tamaño muestral para diferencia de proporciones.

Fuente: (Fernández, 1996)

Donde:

$$p = \frac{p_{GT} + p_{GC}}{2}$$

Fórmula 3: Valor de p en la ecuación de tamaño muestral.

Fuente: (Fernández, 1996)

- n = Sujetos necesarios en cada muestra
- Z = Valor correspondiente al riesgo deseado, es decir, valor de z al nivel de confianza deseado y poder o potencia estadística
- P = Valor de la proporción en los grupos target y control

2.2.3 Evaluación de experimentos

2.3 Modelamiento

2.3.1 Up-Lift

Se utilizan para determinar la sensibilidad de los clientes a una acción de marketing, ya que son modelos incrementales, midiendo el impacto directo y puro de dicha acción.

Cabe destacar que, a diferencia de los modelos convencionales que estimarían la probabilidad de responder a la acción de marketing dentro de toda la población, los modelos Up-Lift buscan aislar un grupo aleatorio de la población total, denominado grupo de control, el que puede tener una respuesta positiva aún cuando no recibió la acción, de tal manera de entregar finalmente la probabilidad de responder del grupo tratado menos la probabilidad de responder del grupo no tratado, generando la final tasa incremental de respuesta atribuible a la acción de marketing (N. Radcliffe & Surry, 2011).

Los modelos Up-Lift entregan como resultado la siguiente clasificación:

		Control	
		Si	No
Tratamiento	No	No molestar ●	No canjeará ●
	Si	Canje seguro ●	Persuadible ●

Ilustración 3: Resultado esperado de un modelo Up-Lift.

Fuente: Elaboración propia

Donde cada uno representa:

- No molestar: Cualquier comunicación o acción de marketing genera un efecto negativo en este tipo de clientes, si un cliente es clasificado en esta categoría no debe recibir la intervención.
- No canjeará: En esta categoría están los clientes que no adoptarán la conducta deseada, en este caso el canje, independiente de si reciben o no la acción.
- Canje seguro: A diferencia de la anterior, los clientes pertenecientes a esta categoría van a canjear a pesar de recibir o no la intervención.
- Persuadible: Finalmente, se encuentran los clientes que tienen una respuesta positiva a la acción de marketing, vale decir, que su probabilidad de canjear se ve incrementada por la recepción del mail de canje.

Para hacerlo más didáctico, se debería enviar mail de canje a los clientes que poseen el círculo verde y por ningún motivo a los de círculo rojo. Con respecto a los de círculo amarillo, el

enviar o no la intervención no generará efectos negativos y probablemente tampoco positivos, por lo que hay que considerar otros aspectos como costos de envío. Gráficamente se puede representar:

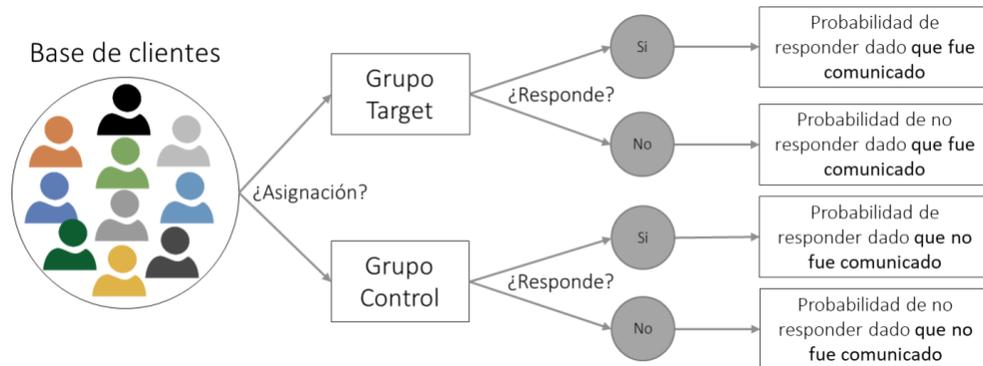


Ilustración 4: Esquema de modelamiento Up-Lift.

Fuente: Elaboración propia

Donde el lift se define como:

$$Lift = \mathbb{P}(Responder | Fue comunicado) - \mathbb{P}(Responder | No fue comunicado)$$

Fórmula 4: Fórmula para cálculo del lift dada una acción de marketing

En la literatura se presentan dos grandes formas de modelar un Up-Lift, dependiendo del enfoque con que se modela (Rzepakowski & Jaroszewicz, 2012b):

- Indirecto: Considera la construcción de dos modelos independientes, uno para la probabilidad de responder positivamente dada la recepción de la campaña de marketing por parte del grupo de tratamiento y otro para la probabilidad de responder positivamente dado que no se recibió la acción, es decir, por parte del grupo de control, finalmente, se restan los dos modelos para obtener el incremental generado por la acción de marketing. Gráficamente:

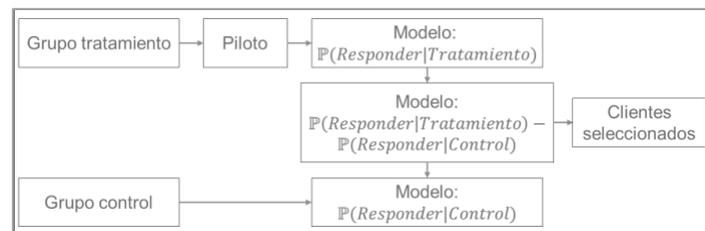


Ilustración 5: Diagrama del modelamiento indirecto del Up-Lift

- Directo: A diferencia del anterior, en vez de modelar cada probabilidad y restarlas, se modela directamente la diferencia de las probabilidades del grupo tratamiento y el control. Para este tipo de Up-Lift, se puede utilizar una adaptación de árboles de decisión, que utiliza distintos criterios de separación y los compara para entregar el que tenga un mejor desempeño. Gráficamente:

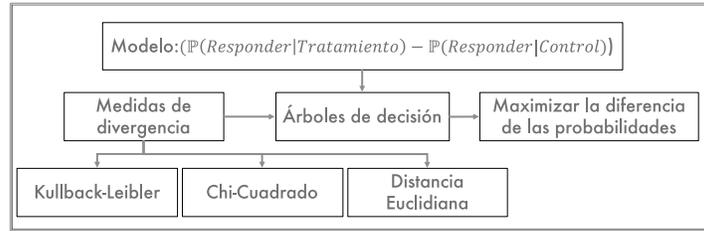


Ilustración 6: Diagrama del modelamiento directo del Up-Lift

Donde las distancias se definen de la siguiente manera:

$$\begin{aligned}
 \text{Kullback-Leibler} \quad KL(P : Q) &= \sum_i p_i \log \frac{p_i}{q_i}, \\
 \text{Euclidiana} \quad E(P : Q) &= \sum_i (p_i - q_i)^2, \\
 \text{Chi-Cuadrado} \quad \chi^2(P : Q) &= \sum_i \frac{(p_i - q_i)^2}{q_i}.
 \end{aligned}$$

Fórmula 5: Cálculo de las distintas divergencias para el modelamiento Up-Lift mediante árboles aleatorios

Adicionalmente se define como una alternativa a las divergencias antes mencionadas y se presenta como una cuarta opción al modelamiento en el software a utilizar, el método de interacciones o L1 ratio, como (Guelman, Guillén, & Pérez-Marín, 2015):

$$Int(P: Q) = \sum_i |p_i - q_i|$$

Fórmula 6: Cálculo de divergencia por interacciones para el modelamiento Up-Lift por árboles aleatorios

Es importante mencionar que, la mayoría de los enfoques de clasificación buscan lograr una alta precisión de predicción. Sin embargo, en la mayoría de los casos prácticos, algunas acciones como enviar una oferta vía mail o tratar a un paciente deben tomarse en los objetos clasificados, y no debemos modelar las probabilidades en sí, sino el cambio en las probabilidades causado por la acción. Lo anterior es la gran ventaja de utilizar modelos tipo Up-Lift.

2.3.2 Evaluación de modelos Up-Lift

Para analizar los resultados de un modelo Up-Lift generado mediante el desarrollo de árboles aleatorios, se utiliza el coeficiente de Qini, que corresponde a una generalización natural del coeficiente de Gini (el área bajo la curva Receiver-Operator Characteristic ROC) para el caso del Up-Lift. Gráficamente, la curva de Qini o Gráfico de ganancia para Up-Lift, corresponde en estricto rigor a un gráfico de ganancia de información convencional, pero con la salvedad de que en el eje vertical se observa el número incremental acumulado de las ventas, canjes u otro KPI que se esté midiendo, es decir, el Up-Lift generado (N. J. Radcliffe, 2007).

En estricto rigor, para evaluar el rendimiento de un modelo Up-Lift de clasificación binaria, se utiliza el coeficiente Q, que corresponde a la medida más general del Qini, y es definido

como el área entre la curva de ganancia incremental debido al modelo y una curva diagonal correspondiente a una asignación aleatoria del grupo tratamiento (N. Radcliffe & Surry, 2011).

2.3.3 Modelos de clasificación y propensión

Con el objetivo de establecer si un registro pertenece a una categoría o clase en particular, se utilizan modelos de clasificación que, según su naturaleza, pueden entregar una probabilidad de que un elemento pertenezca a cada una de las clases o una variable binaria que diga si pertenece o no a cierta categoría.

- Regresión logística multinomial: Consiste en una generalización de la regresión logística para problemas con más de dos clases posibles o para estimar propensiones. Específicamente, el modelo predice la probabilidad de diferentes resultados posibles de una variable dependiente categórica dado un conjunto de variables independientes que pueden ser de naturaleza variada, es decir, categóricas, binarias, continuas, etcétera (Starkweather & Moske, 2011).
- K – Nearest Neighbors: Llamado también KNN es un algoritmo que puede utilizarse tanto para problemas de clasificación o regresión predictiva, sin embargo, es comunmente utilizado para el primero de dichos fines. Su funcionamiento radica en el emparejamiento de cierto registro con sus vecinos más cercanos, donde el parámetro K indica los vecinos con los que se comparará para asignarlo a una clase en particular, dado una medida de similitud (Mucherino, Papajorgji, & Pardalos, 2009).
- Árbol de decisión: Es un tipo de algoritmo supervisado ampliamente utilizado en problemas de clasificación, que funciona tanto para variables categóricas como continuas. El árbol de decisión separa la muestra o población en dos o más particiones homogéneas de datos, basado en el separador o diferenciador más significativo de las variables (Kamiński, Jakubczyk, & Szufel, 2018) (Drummond et al., 2003). Existen distintos criterios de separación en un árbol, los más comunmente utilizados son (Rokach & Maimon, 2015):
 - Gini Index: Es una medida basada en impurezas, que calcula la divergencia entre las distribuciones de probabilidad de los atributos estudiados. En otras palabras indica, de un elemento aleatoriamente escogido, cuántas veces es clasificado de manera errónea.
 - Twoing Criterion: Corresponde a un criterio binario que se utiliza cuando el dominio a evaluar es relativamente amplio y el gini index presenta problemas. Árboles como los de tipo CART utilizan este criterio de separación.
 - Chi-Square: Realiza un test Chi-Cuadrado donde la hipótesis nula es que tanto el input como el objetivo de la separación son independientes. Es utilizada por árboles como el CHAID, que realiza distintos tests dependiendo del tipo de datos que ingresan, para atributos continuos se utiliza un F-test, para

nominales un Pearson Chi-Cuadrado y para ordinales un test de ratio de verosimilitud.

- Information Gain o gain ratio: Es un criterio basado en impurezas, similar al gini index, que mide el cambio relativo en entropía con respecto a las variables independientes. Es utilizada principalmente en árboles como el C4.5 y el C5.
- Random Forest o árboles aleatorios: Consiste en realizar varios árboles de clasificación, donde cada uno clasifica un registro en una clase, considerándose el resultado como una “votación” sobre la clase de ese registro, finalmente, cuando se corrieron todos los árboles, se determina la clase con mayor cantidad de votos, es decir, la clase más “popular”, que es donde finalmente queda el registro clasificado (Breiman, 2001). Cabe destacar que el algoritmo de árboles aleatorios se implementa a través del software SPSS Modeler, que realiza árboles tipo CART para su desarrollo.

2.3.4 Elección de modelos

El problema que se abordará tiene caminos metodológicos, para los cuales se debe definir un modelo a utilizar. En este contexto, se elaborarán los siguientes modelos:

- A quién le envió un mail de canje: Se desarrollará un modelo Up-Lift directo basado en árboles de decisiones, de manera tal de clasificar a los clientes en las 4 categorías presentadas en la ilustración 3. Para esto, se desarrollará un piloto experimental que mida el efecto que tiene el envío de un mail genérico de canje a los clientes pertenecientes al grupo tratamiento, sobre la respuesta natural que ejerzan los clientes en el grupo control.
- Qué le envió: Para este caso se pretenden comparar tres modelos de propensión, por una parte el desarrollo de una regresión logística, el desarrollo de un árbol de decisión y un random forest.

Para definir qué modelo es mejor, se considera la matriz de confusión (Powers, 2007):

		Predicho	
		0	1
Real	0	TN	FP
	1	FN	TP

Ilustración 7: "Contingency table" para casos binarios.

Donde:

- TN: “True Negative”, casos predichos como negativos que realmente son negativos.
- FN: “False Negative”, casos predichos como negativos que realmente eran positivos.
- FP: “False Positives”, casos predichos positivos que eran negativos.
- TP: “True Positives”, casos predichos positivos que realmente lo eran.

- N: Cantidad de registros clasificados totales.

Y en base a ello se definen las siguientes métricas (Powers, 2007):

- Accuracy: Es una métrica de los casos que fueron categorizados correctamente, sobre el total de registros totales, es decir, la división entre la suma de TP + TN sobre los casos totales N.
- Precision: Más comunmente llamado “confidence” en minería de datos, es la proporción de casos positivos predichos que son efectivamente reales positivos, lo que está correctamente alineado con el foco que tiene “Machine learning”, “Data mining” y también “Information retrieval”, pero que es totalmente ignorada en el análisis ROC. Precision puede ser denominada Precisión real positiva (True Positive Accuracy), siendo una medida de precisión de positivos predichos.
- Recall: También llamado “sensitivity”, es la proporción de casos reales positivos que son correctamente clasificados como tales. Esta medida se encarga de ver la cobertura de casos positivos reales por la regla +P (positivo anticipado). Su mejor característica es que refleja cuántos de los casos relevantes la regla +P selecciona. Cabe destacar que las métricas precision y recall se enfocan solo en los casos positivos de predicciones o clasificaciones, sin embargo, capturan algo de información referente a los ratios y tipos de error cometidos. A pesar de todo, ninguno de ellos entrega información sobre qué tan bien el modelo maneja los casos negativos. Se calcula como TP dividido en la suma de TP + FN.
- AUC: Es el área bajo la curva ROC (Receiver operating characteristic), curva que se crea graficando la tasa de verdaderos positivos versus la de falsos positivos. Dicha área es una medida de ranking de desempeño, que estima la probabilidad de que un positivo aleatorio sea clasificado antes de un negativo aleatorio, independiente del umbral de decisión. En otras palabras, AUC puede ser interpretada como la tasa esperada de verdaderos positivos, promediada sobre todas las tasas de falsos positivos.
- F-Measure: Esta medida es una combinación de dos de las vistas anteriormente, precision y recall, y se calcula como la media armónica entre ellas:

$$F = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Fórmula 7: Cálculo del F-Score o F-Measure

Esta métrica asocia los verdaderos positivos con la media aritmética de los positivos predichos y los reales positivos.

3 Metodología

El problema que se abordará con el trabajo de título (Ver: Ilustración 8: Flujo metodológico final del proyecto completo) requiere del estudio de las bases de datos con las que cuenta la empresa, de manera tal de extraer conocimiento de ellas que permitan tomar decisiones sobre la pertinencia de un mail de canje para los clientes, además de las líneas de productos óptimas a ofrecer para cada cliente según sus características y comportamiento frente a las variables definidas. En este contexto, la metodología a utilizar será KDD (“Knowledge Discovery in Databases”), que consiste en llevar a cabo diversos pasos lógicos para llegar a implementar un modelo de minería de datos óptimo y funcional, en estricto rigor, KDD es el proceso mediante el cual se buscan descubrir e identificar patrones en los datos, que permitan sostener o refutar una hipótesis que a la larga se puede transformar en decisiones importantes para la empresa u organización. Los pasos de la metodología a aplicar son:

3.1 Comprensión del negocio

En primer lugar, se debe entender el funcionamiento completo del negocio en general y todos los detalles del proceso involucrado en el trabajo de título, de tal manera de llegar a comprender claramente los objetivos de cada área de la organización. Adicionalmente, en esta etapa se realiza una extensa investigación sobre el estado del arte de los modelos, metodologías y conceptos aplicables a la memoria.

En paralelo, se definen los objetivos, tanto generales como específicos, en concordancia con los que persigue cada una de las áreas involucradas y, en consecuencia, el negocio completo.

3.2 Selección de datos

Luego, se lleva a cabo la selección de los datos, que comienza con la obtención de las bases de datos a utilizar y el posterior análisis exploratorio de las mismas, de manera tal de comprender cómo se componen las bases y poder determinar las variables que son de utilidad y aportan conocimiento al modelo, a diferencia de otras que se deban dejar de lado para no ensuciar los resultados y aumentar la dificultad del modelo.

3.3 Pre-procesamiento

Después de la selección de los datos, se debe estudiar la calidad de la base resultante, identificando valores faltantes (“missing values”), valores fuera de rango (“outliers”), datos no actualizados, datos redundantes entre muchas otras posibilidades. De esta forma, se llevan a cabo tratamientos de la información que permiten realizar de mejor manera los análisis y obtener los resultados esperados.

3.4 Transformación de datos

Corresponde al proceso de obtener variables nuevas a partir de las existentes y de modificar las variables que no puedan ser incorporadas al modelo tal y como están, por ejemplo, crear variables dummy a partir de variables binarias de texto, pasar de una fecha a una cantidad de días, entre otros.

3.5 Minería de datos

Hecho todo lo anterior, se procede a trabajar con herramientas de minería de datos, en el caso de la presente memoria, implementar los pilotos experimentales y desarrollar el modelo Up-Lift para determinar a qué clientes se le deben enviar mails de canje. Posteriormente se analizarán las variables y se determinará cuáles son óptimas para alimentar al modelo de recomendación de productos para canje, que finalmente será implementado y automatizado (Ver detalles en Flujo metodológico).

3.6 Análisis, interpretación y evaluación

Para culminar con la metodología, se realiza un análisis completo del trabajo realizado, además de una interpretación útil tanto para el negocio como para la universidad, vale decir, que sea relevante para la organización, pero también desde el punto de vista académico, proponiendo lineamientos claros de cómo proceder en cuanto a los modelos desarrollados.

4 Flujo metodológico

Para mayor detalle, se presenta el flujo metodológico que se desarrolla en la memoria, considerando los pasos lógicos explicitados anteriormente.

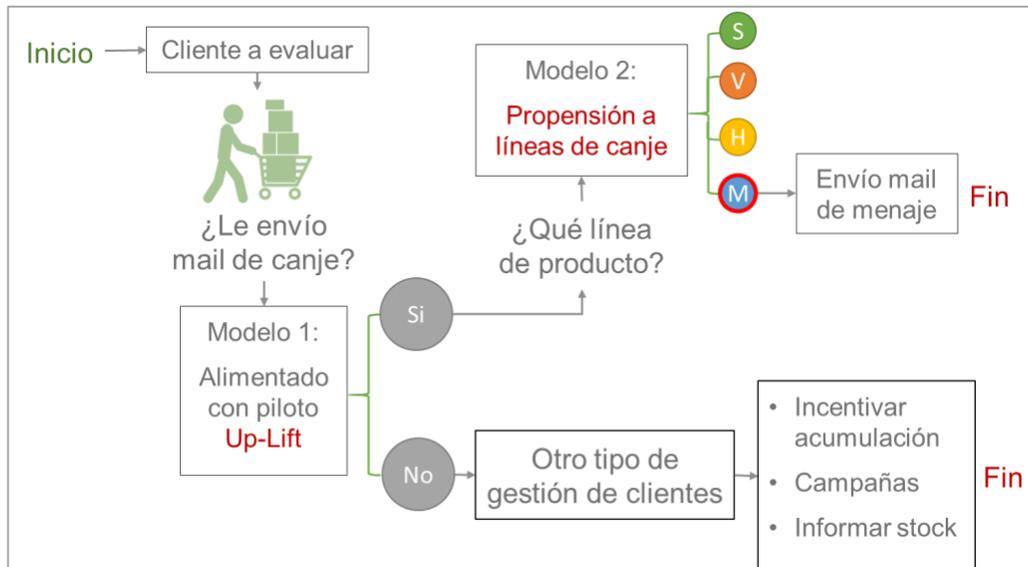


Ilustración 8: Flujo metodológico final del proyecto completo¹⁰

Fuente: Elaboración propia

Como se puede observar en la ilustración, en primer lugar, se pretende determinar si un cliente debería recibir un mail de canje, cuya metodología es la siguiente:

1. Realizar un diseño experimental que permita estudiar el comportamiento de un grupo de tratamiento frente a un envío genérico de mails.
2. Desarrollar un modelo Up-Lift que entregue la probabilidad incremental de canjear debido a la comunicación entregada.
3. Categorizar según el Up-Lift a los clientes en deciles, donde los deciles más altos representan a los clientes “Influenciables”, los del medio son los que no modifican su comportamiento y los más bajos son los clientes “No molestar” o “Sleeping dogs”.
4. Extraer las diferencias entre los clientes influenciados y los no molestar, para poder clasificar a nuevos clientes dentro de dichos grupos, esto mediante el desarrollo de un árbol de decisión que tiene como variable dependiente una binaria que toma por valor 1 si el cliente está en el grupo “influenciados” y 0 si es “no molestar”.

Luego, si un cliente no califica para recibir un mail de canje, se pueden proponer otro tipo de gestiones, como informar los beneficios y el funcionamiento del plan de puntos, incentivar acumulación, proponer alguna campaña para acelerar acumulación, informar la cantidad de puntos que posee actualmente y para qué le alcanza, entre otras gestiones. Cabe destacar que

¹⁰ Se muestran ejemplos de líneas en el flujo, no corresponden a la cantidad final de líneas.

proponer, diseñar e implementar dichas gestiones de clientes quedan fuera del alcance del presente trabajo de título.

Si un cliente califica para recibir un mail de canje, se procede a desarrollar un modelo de propensión clásico, que determina la propensión a canjear una de las líneas de productos para canje previamente definidas, de manera tal de comunicarle en el mail la que presente mayor propensión. Lo anterior se llevará a cabo mediante la implementación de tres modelos diferentes:

1. Un modelo de regresión logística.
2. Un árbol de decisión.
3. Un modelo de árboles aleatorios o *random forest*.

Para cada línea se definirá cuál es el mejor modelo de acuerdo a las métricas de comparación obtenidas, en particular, se evaluarán 3 métricas:

- Accuracy: Ya que considera el desempeño neto de clasificación del modelo.
- Recall: Como se quiere identificar clientes que canjeen en una línea, lo más relevante es maximizar la cantidad de casos positivos que son realmente categorizados como tales.
- Ganancia de información: Sirve para comparar el desempeño relativo a realizar la categorización de manera aleatoria.

Finalmente, luego de implementada la metodología completa, se tendrá una base de clientes que deberían recibir un mail de canje y sus propensiones a canjear en cada línea de productos, de manera tal de enviarles a cada uno un mail según sus mayores propensiones y los productos que son factibles según la disponibilidad que presenta el área comercial.

5 Resultados esperados

Los resultados esperados del trabajo de título van de la mano con los objetivos planteados en Objetivos del presente informe, vale decir, concretamente se espera obtener:

- Un set de variables relevantes que sean suficientes para explicar y luego predecir el comportamiento de los clientes frente a un mail relativo a puntos y frente al canje. Se entrega el modelo como código en SQL para ser replicado y las variables más relevantes en una tabla.
- Un modelo Up-Lift que permita determinar si a un cliente se le debe enviar o no una acción de marketing, específicamente un mail de canje. Dicho modelo se entrega como código en lenguaje SQL para la definición de las variables, en R para el desarrollo del modelo y en reglas de SQL para la clasificación expost de clientes distintos.
- Un modelo de recomendación de productos que entregue la línea de productos con mayor propensión a ser canjeada por un cliente. Este modelo se entrega en forma de reglas de SQL.
- Un criterio que permita determinar si a un cliente en cierto periodo se le enviará un mail de canje y con qué productos.

Todo lo anterior, en conjunto, forma un sistema integral de recomendación de productos para canje, que pretende aumentar la efectividad de los mails de canje enviados, cumpliendo con el objetivo general del trabajo de título.

6 Desarrollo metodológico

6.1 ¿A quién le envió un mail? Up-Lift

El objetivo de esta sección es clasificar a los clientes dentro de categorías según su respuesta esperada a un mail de canje. De manera tal de poder identificar a los clientes que deberían recibir un mail y a los que por ningún motivo se debería comunicar.

6.1.1 Tamaño muestral

Para determinar cuántos clientes se necesita para asegurar un poder estadístico y niveles de significancia importantes al enviar el piloto experimental, se desarrolló una estimación mediante una calculadora de poder estadístico¹¹, que efectúa el cálculo explicitado en el marco conceptual de manera automática, al que se le entregaron los siguientes parámetros:

- Tasa de respuesta esperada del grupo target: 0,76%
- Tasa de respuesta esperada del grupo de control: 0,37%¹²
- Nivel de significancia: 95%
- Poder estadístico: 95%

Obteniendo que la cantidad de clientes mínima en el grupo target debe ser de 10.000 clientes y en el grupo de control de 7.000 (Ver detalles en Anexo 12: Cálculo tamaño mínimo de muestra).

Cabe destacar que, considerando las diferencias globales entre las tasas de respuesta de grupo target y grupo de control, se obtuvo un poder estadístico menor al esperado, de un 49%.

6.1.2 Base potencial de clientes

Con el objetivo de estudiar si se cuenta con la cantidad necesaria de clientes para desarrollar el experimento, se creó una base de clientes potencialmente utilizable para el piloto, resultando un total de 576.694 clientes que cumplan con:

- Cuenta abierta: Que tiene la capacidad de comprar y canjear, sin bloqueos asociados a situaciones de deuda o morosidad.
- Más de 5.000 puntos: Primer nivel de canje formal, a pesar de que se puede canjear en niveles de hasta 1.000 puntos, pero en situaciones promocionales.
- Contactables vía mail¹³: De tal manera de poder comunicarle vía correo electrónico.

Por lo que se supera la cantidad mínima de clientes necesarios para el experimento, de forma tal de que se tiene un material inicial para proponer un diseño experimental robusto.

¹¹ Calculado con gpower calculator, disponible en <http://www.gpower.hhu.de>

¹² Tasas de respuesta esperadas según el promedio de tasas de respuesta de los envíos de puntos de la empresa entre el 1 de julio de 2017 y el 31 de marzo de 2018

¹³ Información al 5 de abril de 2018

6.1.3 Diseño experimental

El primer paso para desarrollar el modelo Up-Lift es diseñar un piloto experimental que permita medir las diferencias entre la probabilidad de canjear dado que se recibió el mail y la probabilidad de canjear dado que no se recibió el mail y clasificar a los clientes en alguna de las categorías anteriormente descritas.

Es importante destacar que, el piloto se requiere porque, para desarrollar un up-lift adecuado, la comunicación debe ser genérica respecto al mail de canje, a diferencia de los mails que actualmente envía la organización, que todos incluyen promociones de canjes particulares.

De esta manera, se diseñó un piloto cuyo objetivo consiste en determinar qué clientes son sensibles al envío de una comunicación de canje de puntos, considerando como medida de éxito el que los clientes canjeen.

Así, se diseñaron dos envíos diferentes, cuyos detalles son:

1. Grupo target 1: Mail de canje genérico, es decir, que contenga todas las formas de canjear (En tiendas del holding, en Gift Card para canjear en tiendas y en experiencias) y productos de cada uno de ellos.
2. Grupo target 2: Mail de canje enfocado solo en las posibilidades de canje en tiendas.
3. Grupo de control: No se le comunica nada en el piloto.

El periodo experimental fue de 10 días, que tuvieron lugar entre el 17 de mayo de 2018 y el 27 de mayo de 2018, donde tanto los clientes de ambos targets como del grupo de control fueron excluidos de cualquier otro envío de puntos durante la duración del experimento, para evitar que se presentaran variables de confusión que generaran ruido en los resultados y análisis. Ver Anexo 22: Resumen del piloto experimental diseñado.

Cabe destacar que, como condición de la empresa, la base de clientes debía componerse en su 50% por clientes que no tuvieran canjes en toda su vida, ya que son los clientes que se pretenden gestionar por la Gerencia de Fidelización para que generen un primer canje, por todo lo explicado en la sección de justificación del trabajo de título, mientras que el otro 50% debían ser clientes que no tuvieran canjes en el último año móvil, de manera tal de excluir a clientes que hubiesen experimentado hace poco tiempo el canje. Dicha estratificación se mantuvo dentro de los 3 grupos de clientes.

Adicionalmente, se aplicó una estratificación relativa a la categoría de los clientes, de tal manera que los grupos de tratamiento y control tuvieran la misma distribución de clientes normales, premium y elite que la base potencial completa.

El detalle de las gráficas enviadas se encuentra en Anexo 13: Pieza enviada para el target 1 y Anexo 14: Pieza enviada para el target 2.

6.1.4 Resultados experimento

Concluido el periodo de duración del experimento, se extrajeron los resultados generales, obteniendo los siguientes:

Tabla 3: Resultados generales del piloto experimental enviado

Asignación	Base total	Clientes con canjes	Canjes totales	Puntos canjeados	Media puntos canjeados	TR	TR incremental
Target 1	21.030	620	736	18.606.000	25.280	2,95%	0,22 PP ¹⁴
Target 2	21.030	623	742	15.847.000	21.357	2,96%	0,23 PP ¹⁵
Control	18.060	494	591	13.726.000	23.225	2,73%	-

Fuente: Elaboración propia basado en datos extraídos de las bases de la empresa

El KPI definido para la evaluación del experimento es la cantidad de clientes con canje, dado que, lo que se quiere lograr es aumentar la cantidad de canjes y los canjes nuevos. Lo anterior debido a que es el objetivo de la Gerencia de Fidelización y va de la mano con el objetivo general propuesto en el presente trabajo, de aumentar la efectividad de los mails.

Se puede observar que, a diferencia de los puntos canjeados, en la cantidad de clientes con canjes, que da origen a la tasa de respuesta, se observa una diferencia positiva entre ambos grupos target y el grupo de control, sin embargo, entre grupos target no existe una diferencia significativa, por lo que, para efectos del análisis y modelamiento posterior, se considerará como un solo grupo target y un grupo de control.

Al abrir el análisis por las variables que dieron origen a la estratificación de la base de clientes utilizados para el experimento, se pueden concluir cosas más interesantes que del análisis de los resultados generales.

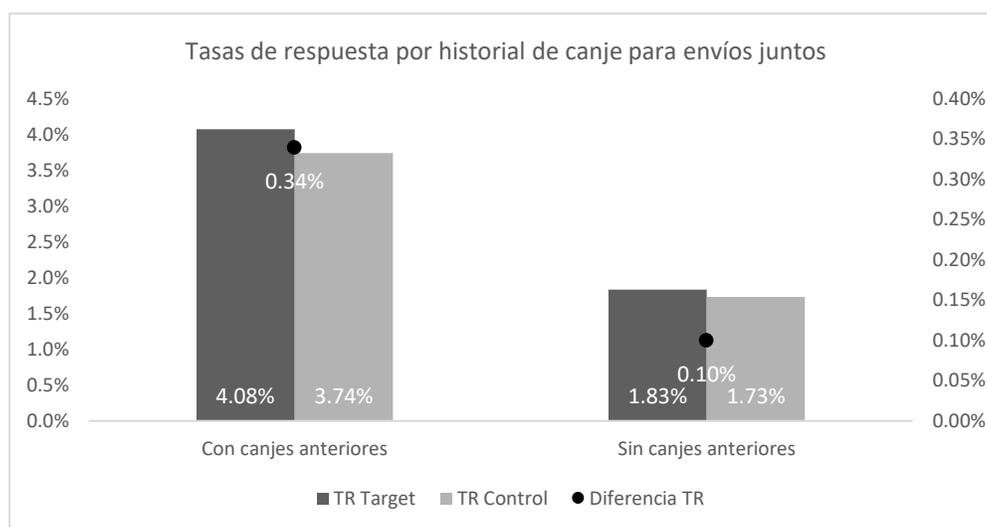


Figura 8: Tasas de respuesta y diferencias por historial de canje.

Fuente: Elaboración propia

El gráfico anterior muestra los resultados al abrir según historial de canje de los clientes, que dividía la base en un 50% para clientes con canjes anteriores pero sin canjes en el último año móvil y 50% de clientes que nunca habían canjeado. Al realizar un test de proporciones, cuya hipótesis nula es que las tasas de respuesta de grupo target y grupo control son iguales, se puede observar que, solo en el caso de los clientes con canjes anteriores, se puede rechazar

¹⁴ Puntos porcentuales, significativo al 78% con un valor z de 1,30

¹⁵ Puntos porcentuales, significativo al 81% con un valor z de 1,36

dicha hipótesis al 90% de confianza. Lo anterior nos indica que, dado el diseño del experimento y las condiciones establecidas, se generó un mayor impacto en los clientes que ya estaban relacionados o familiarizados con el plan de fidelización.

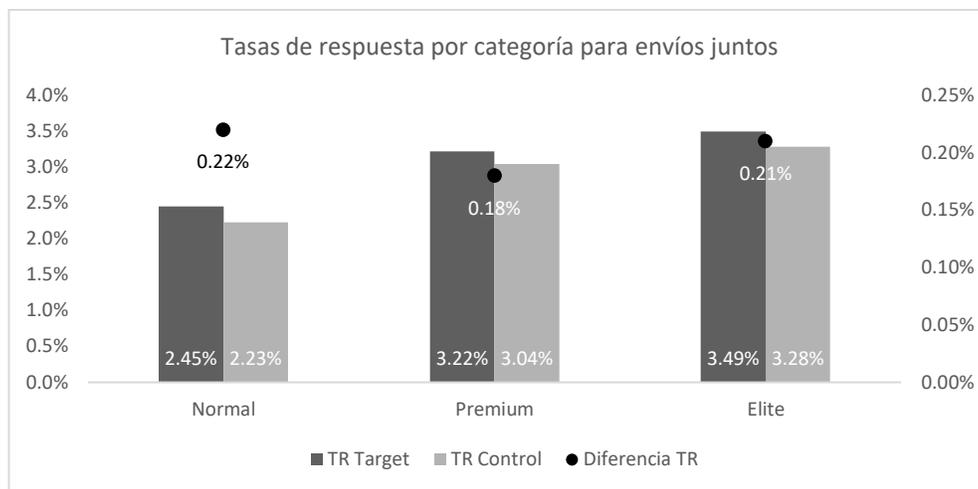


Figura 9: Tasas de respuesta y diferencias por categoría.

Fuente: Elaboración propia

En el caso de la categoría de los clientes, se puede rechazar la misma hipótesis nula al 80% de confianza, en el caso de los clientes Normales, mientras que para las dos categorías más altas, Premium y Elite, no existen diferencias significativas. Lo anterior tiene sentido, dado que los clientes Normales es intuitivo que respondan mejor a la comunicación entregada, por lo tanto que sean más influenciables, debido a que tienen puntos para canjear, pero no tantos como para llegar a niveles muy altos, a diferencia de los Premium y Elite que en la mayoría de los casos esperan promociones en niveles mayores de canje para utilizar sus puntos con un retorno de pesos por puntos mayor, mientras que los clientes normales tienen menor conocimiento del plan y menor capacidad de acumular puntos.

6.1.5 Modelamiento

Cabe destacar que, para efectos del modelamiento Up-Lift, se balanceó la base haciendo un oversampling aleatorio del 50% de los casos de éxito del experimento, de manera tal de amplificar por 1,5 la tasa de respuesta original del piloto, que era de 2,9%.

6.1.5.1 Variables

Para desarrollar el modelo Up-Lift que categoriza a los clientes en Influenciables, no molestar, seguro si y seguro no, se deben extraer las variables relevantes para el modelamiento, de manera tal de poder explicar la mayor cantidad de la varianza de la adopción o no de un canje en el periodo experimental. De esta manera, se consideraron 93 variables (Detalles en Anexo 15: Variables Up-Lift) dentro de las categorías:

- Demográficas: Género, zona de residencia, edad, entre otras.

- Transaccionales: Monto gastado con la tarjeta, cantidad de usos de la tarjeta, rubros en los que compra, entre otras.
- De contrato: Tipo de tarjeta, cupo, potencial de gasto, entre otras,
- De fidelización: Stock actual de puntos, cantidad de canjes, nivel de canje anterior, distancia al próximo nivel de canje, tasa de canje, entre otras.
- De evolución: Porcentaje adicional de las variables entre un mes y el siguiente, para variables como días con compras, stock de puntos, monto gastado, entre otras.
- De canales: Si es contactable vía mail, contactable vía teléfono, cantidad de ingresos a la web, entre otras.

Para seleccionar las variables a utilizar, se realizaron dos métodos de extracción iterativos, el primero consistió en la elaboración de un árbol de decisión tipo CHAID, que intentara explicar con todas las variables disponibles el canje para la base de datos del experimento, de manera tal de obtener las más relevantes según la importancia relativa de los predictores. El resultado del primer árbol de decisión se presenta a continuación, del cual se extraen las variables más importantes.

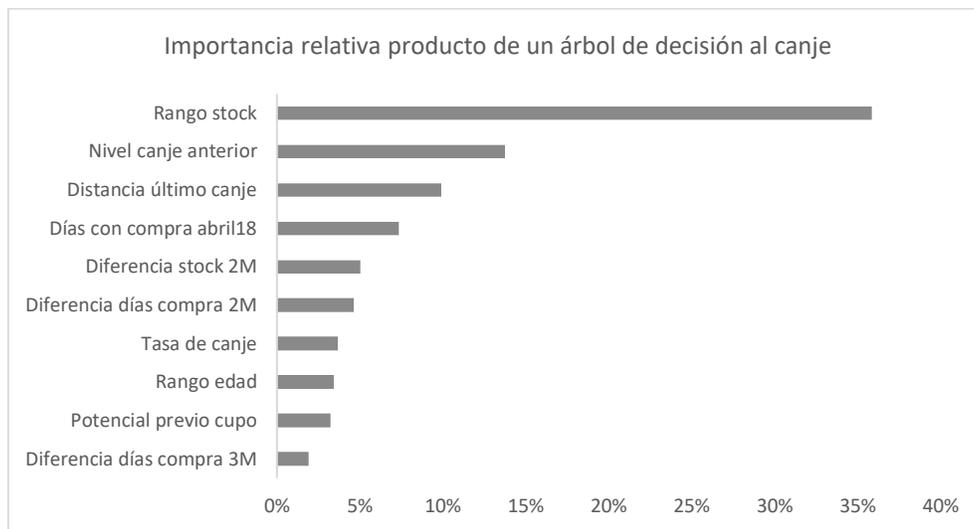


Figura 10: Importancia relativa de las variables para explicar el canje.

Fuente: Elaboración propia en SPSS Modeler

El segundo método fue realizar el modelo Up-Lift mediante Random Forests en el software estadístico R, y analizar los resultados con el comando VarImportance, que arroja la importancia relativa de las variables para el modelo realizado en particular.

Ambos métodos se realizaron alternadamente en varias iteraciones, eliminando en cada etapa las variables que aportaran menos información, obteniendo finalmente 23 variables (Detalle en Anexo 16: Variables finales Up-Lift), destacándose según su importancia relativa:

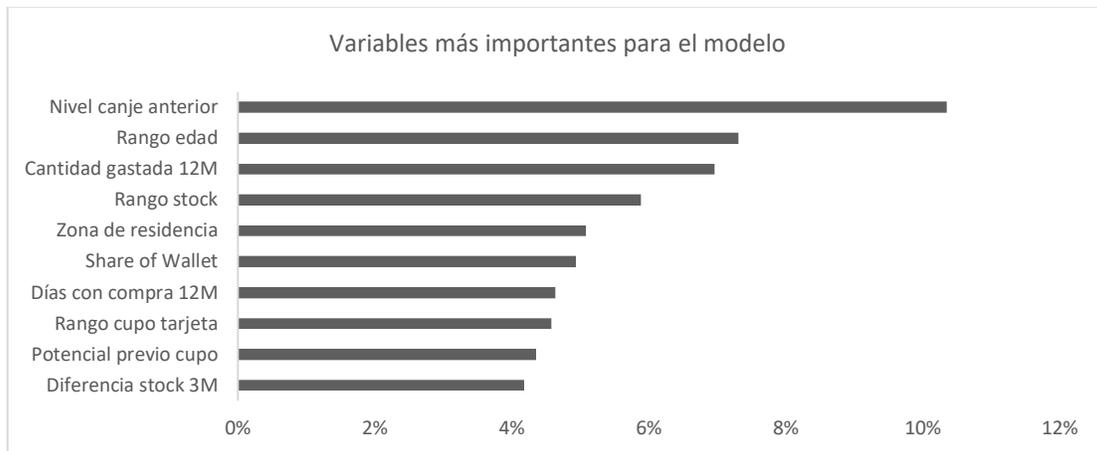


Figura 11: Importancia relativa de las variables del modelo calibrado.

Fuente: Elaboración propia

Así, se desarrolló un método de determinación de variables basado no solo en los resultados del modelo Up-Lift, sino que en el desarrollo iterativo de dos modelos independientes, que permiten incorporar variables que podrían aportar al modelo final pero que, de utilizarse solo un método en una sola ejecución, se pueden perder por hiperparámetros definidos, cortes aleatorios u otro elemento que haga variar los resultados, que en este caso eran los cortes realizados por los árboles.

6.1.5.2 Elección de parámetros y divergencia

Al modelar con el algoritmo Up-Lift mediante Random Forests en el software R, se deben escoger los siguientes hiperparámetros:

- Cantidad de variables que prueba por nodo
- Cantidad de árboles que realiza
- Método de corte o divergencia utilizado en cada nodo de cada árbol
- El número mínimo de observaciones que deben existir en un nodo para probarlo

Se calibró el modelo y, mediante iteraciones del mismo modelo evaluando tiempo de procesamiento y calidad de resultados, se obtuvo que los hiperparámetros adecuados fueron:

- Cantidad de variables que prueba por nodo: 23 (Todas las disponibles)
- Cantidad de árboles que realiza: 200
- Método de corte o divergencia utilizado en cada nodo de cada árbol: Se determina más adelante
- El número mínimo de observaciones que deben existir en un nodo para probarlo: 200

Una vez escogidos los parámetros, se comparan los cuatro métodos de divergencia, obteniendo los siguientes valores de ganancia de información:

Tabla 4: Coeficiente de Qini obtenido para cada divergencia del modelo ajustado

Tipo divergencia	Interacciones	Kullback Leiblel	Euclideana	Chisq
Qini	1,19%	1,06%	1,18%	0,92%

Fuente: Elaboración propia basado en los resultados obtenidos de R

De donde se desprende que el modelo a utilizar será el que considera el método de interacciones como medida de divergencia, ya que presenta mayor coeficiente de Qini, es decir, entrega el mayor porcentaje extra de ganancia de información en comparación a una asignación aleatoria de mails.

6.1.5.3 Resultados

Como se mencionó anteriormente, se decidió como definitivo el modelamiento que tiene como inputs:

- Cantidad de variables que prueba por nodo: 23
- Cantidad de árboles que realiza: 200
- Método de corte o divergencia utilizado en cada nodo de cada árbol: Interacciones
- El número mínimo de observaciones que deben existir en un nodo para probarlo: 200

Obteniendo un coeficiente de qini de 1,19% y las siguientes curvas elaboradas sobre los datos de evaluación:

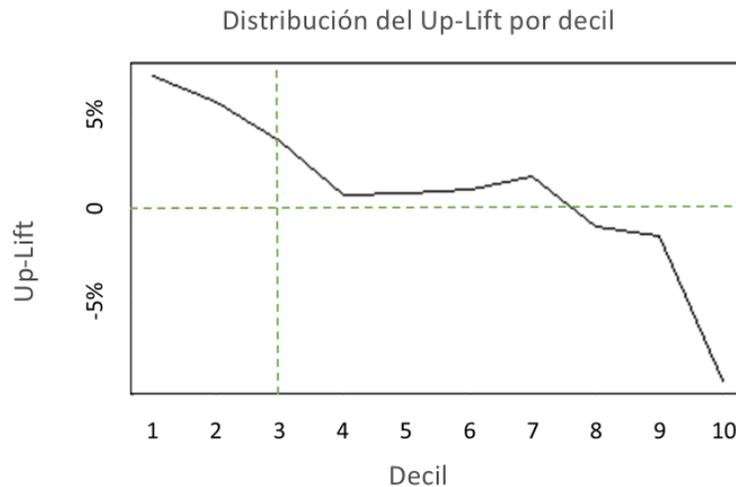


Figura 12: Resultados del modelo Up-Lift para el método de interacciones.

Fuente: Elaboración propia en R

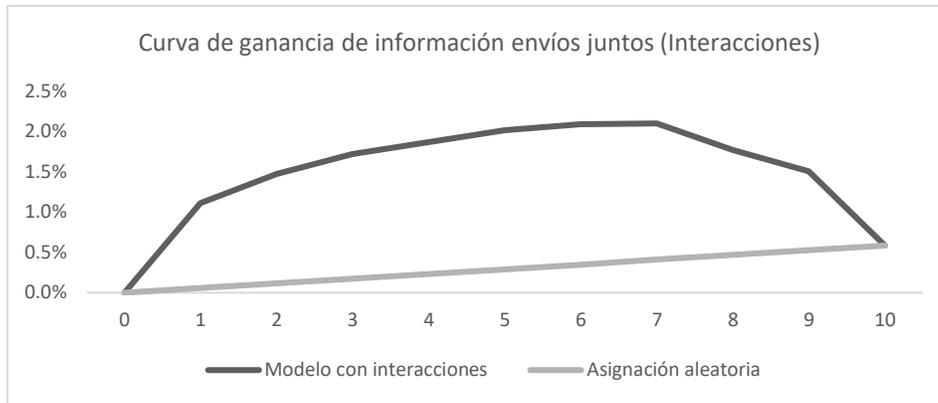


Figura 13: Curva de ganancia de información del modelo Up-Lift para el método de interacciones sobre el conjunto de comprobación.

Fuente: Elaboración propia con los resultados del modelo calibrado

A partir de los gráficos anteriores se puede concluir que se observa el comportamiento esperado de los primeros deciles según el Up-Lift pronosticado, donde se aprecia claramente un valor positivo, mientras que decae a medida que se avanza en los deciles, terminando con un Up-Lift bastante bajo al tocar a todos los clientes, ya que se incorporan a los clientes “No molestar”, que presentan una respuesta esperada negativa a la comunicación.

Además, se observa una ganancia de información importante respecto de la generación de un modelo de mailing aleatorio, obteniendo una ganancia muy pronunciada al tocar a los mejores clientes, decayendo a medida que se acerca a los clientes de los deciles más altos.

6.1.5.4 Categorización de los clientes

Para una mejor clasificación de los clientes en las cuatro categorías, se clasifican en deciles según el Up-Lift que le entrega el modelo a cada cliente, obteniendo:

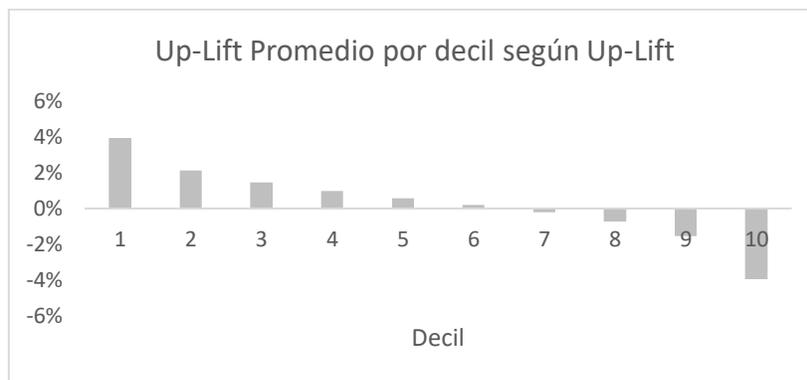


Figura 14: Clientes divididos en deciles según el Up-Lift promedio ordenado de mayor a menor.

Fuente: Elaboración propia con los resultados del modelo calibrado

En trabajos de título anteriores como (Aguirre, 2017), se consideran a los primeros dos deciles como los clientes influenciables y a los últimos dos como los no molestar, mientras que los de en medio constituyen las otras dos categorías que, para efectos del presente trabajo de título, no tienen relevancia. Lo anterior dado que las acciones que se esperan realizar son,

en primer lugar, comunicarle principalmente a los clientes que se espera tengan una respuesta positiva al mail de canje (Influenciables), y en segundo lugar, evitar enviar la comunicación a los clientes que tienen una respuesta esperada negativa (No molestar o “*Sleeping dogs*”). No obstante lo anterior, para tomar una decisión más precisa del corte a realizar para determinar a los clientes influenciables, se grafican los Up-Lift positivos solamente:

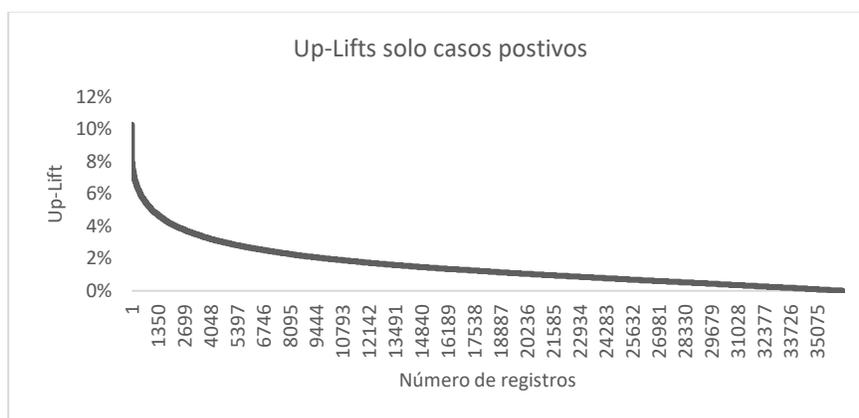


Figura 15: Up-Lifts positivos.

Fuente: Elaboración propia con los resultados del modelo calibrado

Se puede observar que existe un cambio relativamente significativo en el 2% de Up-Lift promedio, de manera tal de que, por regla del codo, se puede escoger ese punto como el punto de corte para los clientes influenciables, que representa a aproximadamente 10.000 registros, mientras que en los dos primeros deciles hay 12.000, por lo que el criterio del codo ratifica la elección de los dos primeros deciles como clientes influenciables y los dos últimos como los clientes no molestar.

6.1.6 Perfilamiento

Una vez realizada la clasificación en las categorías de interés producto del modelo Up-Lift, se debe desarrollar un método que permita determinar si un cliente ajeno al experimento, es decir, que no estuvo ni en el grupo de control ni el target, pertenece a alguna de las categorías extremas (“Influenciables” y “No molestar”).

Lo anterior se lleva a cabo de la siguiente manera:

- Según la respuesta al experimento, mediante el Up-Lift, se clasificó a cada cliente en un decil.
- Los primeros dos deciles conforman el grupo “Influenciables”, ya que presentan mayor lift, mientras que los deciles 9 y 10 corresponden a los clientes “No molestar” o “*Sleeping dogs*”.
- Se desarrolla un árbol de decisión cuya variable a predecir es una variable binaria que toma el valor 1 si el cliente pertenece al grupo “Influenciables” y 0 si pertenece al grupo “No molestar”.
- Se determinan las variables más relevantes.
- Se perfilan a los dos grupos según las variables más importantes.

- Finalmente se tiene la capacidad de clasificar a un cliente en alguno de los dos grupos, que determina su respuesta esperada a un mail de canje.

Los primeros dos puntos fueron realizados en la sección anterior, datos con los cuáles se desarrolló un árbol de decisión tipo CHAID¹⁶ para clasificar a los clientes en los dos grupos descritos, obteniendo:

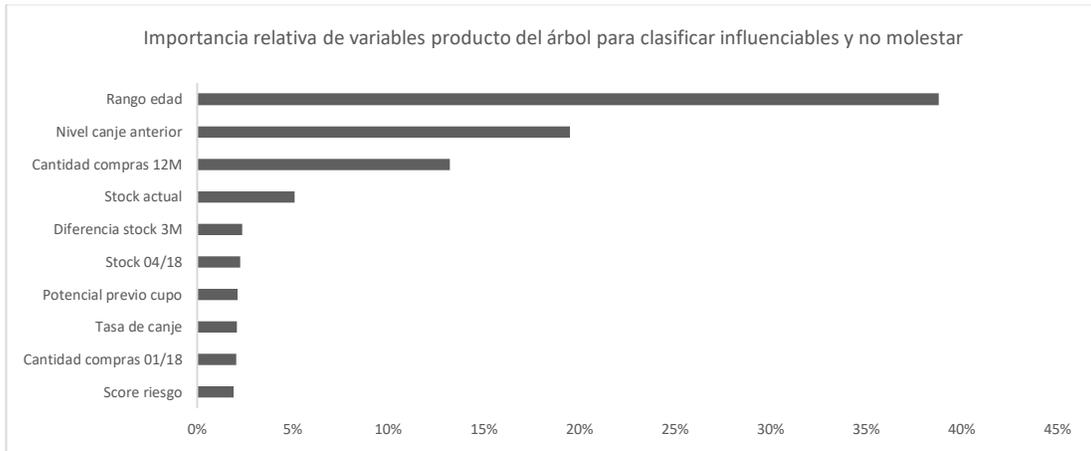


Figura 16: Importancia relativa de las variables de un árbol de clasificación para clientes influenciados y no influenciados.

Fuente: Elaboración propia

Donde se desprende que las variables más relevantes para la categorización según respuesta esperada al mail de canje tienen relación con:

- La edad del cliente, que es un indicador que podría aproximar a un cliente al uso más intenso del canal email.
- Comportamiento de canje y acumulación, donde se destaca el nivel de canje anterior del cliente, su stock actual de puntos y una medida de aceleración de su acumulación con respecto a 3 meses atrás.
- Comportamiento de compra con la tarjeta, con principal énfasis en la cantidad de compras del último año móvil.

Así, los perfiles más y menos propensos se presentan en la siguiente tabla, ver también Anexo 23: Nodos más y menos propensos a ser influenciados:

¹⁶ Método que tuvo mejor desempeño al compararlo con un árbol C5

Tabla 5: Nodos más y menos propensos

Tipo de nodo	Cantidad	Porcentaje del total	Variables
Más propenso	975	5%	Stock: Menor a 19.000 puntos Score riesgo: [425,705] Edad: [26,30]
Más propenso	1050	5,4%	Diferencia stock 3M: > 0 Nivel de canje anterior: < 5.000 Edad: [31,35] o [41,45]
Menos propenso	1151	5,9%	Edad: > 65
Menos propenso	1855	9,5%	Stock: < 35.000 puntos Edad: [51,60]

Fuente: Elaboración propia

Donde se puede observar que, en consecuencia con lo mostrado anteriormente en las variables más relevantes, la edad y relación con el plan de fidelización de los clientes son las variables que tienen más impacto en la propensión a pertenecer al grupo de influenciables, mostrando una tendencia a clientes más jóvenes y con menor cantidad de puntos acumulados, mientras que los clientes menos propensos a ser influenciables son clientes que acumulan gran cantidad de puntos pero no tan alta como para llegar a los niveles más altos de canje y son principalmente clientes de mayor edad.

Lo anterior hace sentido con la descripción de un cliente más digital y que no cuenta con gran cantidad de puntos como para tener una preferencia u objetivo claro de canje.

6.1.7 Descubrimientos relevantes

Del piloto experimental y el posterior modelamiento del Up-Lift, se pueden desprender análisis relevantes que sirvan para transmitir el conocimiento extraído a la contraparte comercial del trabajo de título y para explicitar las diferencias que existen entre los clientes categorizados como influenciables y los que responden negativamente a una comunicación de canje vía mail.

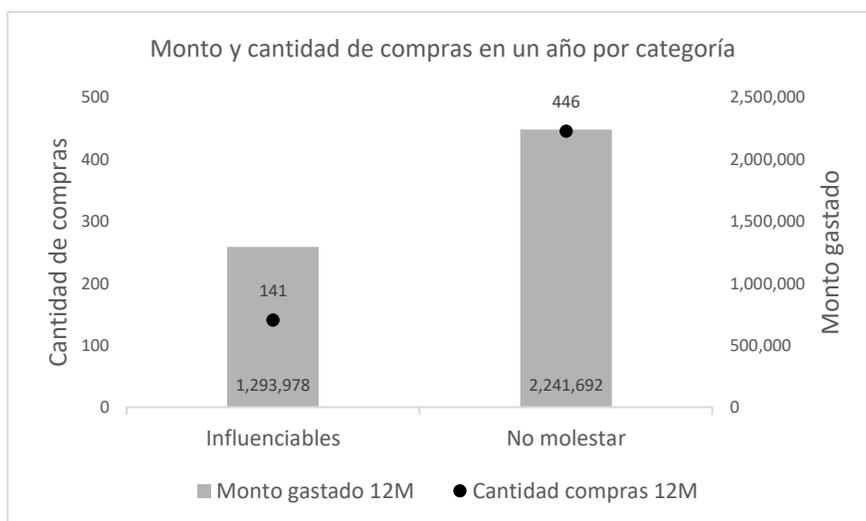


Figura 17: Diferencias en monto gastado y cantidad de compras de los grupos influenciables y no molestar.

Fuente: Elaboración propia con los resultados del Up-Lift

Los clientes influenciables presentan menor cantidad de compras y menor monto gastado con la tarjeta en el último año móvil, lo que permite inferir que son clientes que no son tan principales en el uso de la tarjeta y que no tienen tanta capacidad de acumular alta cantidad de puntos.

Tabla 6: Distribución de clientes influenciables y no molestar en relación a la zona de residencia

Zona\Categoría	Influenciables	No molestar
NORTE	24%	75%
STGO. CENTRO	71%	28%
STGO. EXTERIOR	44%	55%
STGO. NORTE	58%	41%
STGO. ORIENTE	21%	78%
STGO. PONIENTE	60%	39%
STGO. SUR	74%	25%
SUR	58%	41%

Fuente: Elaboración propia

En relación a la distribución geográfica de los clientes, cabe destacar que los influenciables se concentran en sectores donde predominan las clases socioeconómicas de clase media para abajo, comprendiendo a Santiago Sur, con comunas como Puente Alto y La Florida o Santiago Centro, que considera a comunas como Santiago y Estación central, mientras que los clientes no molestar se concentran principalmente en Santiago Oriente, con comunas como Las Condes y Vitacura, y en el Norte de Chile, apalancado por comunas de alto poder adquisitivo proveniente de la minería como Antofagasta y Calama.

Tabla 7: Distribución de clientes influenciables y no molestar en relación a la edad

Rango edad	Influenciables	No molestar
18-25	26%	74%
26-30	82%	18%
31-35	68%	32%
36-40	14%	86%
41-45	70%	30%
46-50	58%	42%
51-55	35%	65%
56-60	37%	63%
61-65	52%	48%
>65	10%	90%

Fuente: Elaboración propia

Con respecto a la edad, no hay un patrón muy claro y continuo que distinga a los clientes influenciables, sin embargo, si se puede sostener que hay una tendencia a ser influenciable en la medida que se es un “adulto joven”, dado que a medida que se sube de los 50 años comienza a decaer la influenciabilidad del mail, lo que tiene sentido si consideramos que para tomar la decisión dada una comunicación digital, se pensaría que las personas más jóvenes tienen la mayor respuesta.

Cabe destacar que pasan cosas contraintuitivas en los extremos de edad, ya que los más jóvenes tienen respuesta negativa al mail de canje, esto puede ser porque, dado su menor gasto promedio, no pueden llegar a niveles de canjes deseados y por ende la comunicación de puntos “frustra” su relación con el plan de fidelización.

Tabla 8: Distribución de clientes influenciables y no molestar en relación a los puntos acumulados

Rango stock	Influenciables	No molestar
< 5.000	8%	92%
5.000 - 17.999	51%	49%
18.000 - 47.999	54%	46%
48.000 - 89.999	67%	33%
> 90.000	13%	87%

Fuente: Elaboración propia

En concordancia con lo anterior, los clientes influenciables están principalmente en los niveles medios de canje, ni muy bajos, ya que no pueden acceder a muchos productos y pueden decepcionarse por esto, ni muy altos porque son clientes que conocen el plan de puntos y usualmente canjean en los niveles más altos, porque ganan una mejor conversión¹⁷, y en productos ya conocidos¹⁸.

¹⁷ Según el diseño del plan de puntos, el valor del punto aumenta a medida que aumenta el nivel de canje

¹⁸ Los viajes y gift cards más altas son los productos preferidos por quienes logran acumular gran cantidad de puntos

Tabla 9: Distribución de clientes influenciables y no molestar en relación al nivel de canje anterior

Canje anterior	Influenciables	No molestar
< 5.000	63%	37%
5.000 - 17.999	66%	34%
18.000 - 47.999	37%	63%
48.000 - 89.999	45%	55%
> 90.000	24%	76%
SIN CANJE ANT	47%	53%

Fuente: Elaboración propia

Finalmente y muy relacionado al resultado anterior, los clientes que antes han canjeado en niveles bajos son los más influenciables, debido probablemente a que saben que no acceden a muchos puntos entonces son más propensos a canjear gracias a la comunicación, mientras que los niveles altos de canje ya conocen el plan y tienen sus preferencias claras y no quieren ser comunicados de cualquier producto.

6.1.8 Conclusiones preliminares

Del cierre de la primera parte de la metodología se pueden concluir algunas cosas previamente a las conclusiones finales del trabajo de título, en los ámbitos:

- Perfiles de clientes: Los clientes no influenciables son clientes que compran y gastan más con la tarjeta, constituyéndose en clientes más “educados” en cuanto al plan de puntos y que son poco influenciables porque saben lo que quieren y a lo que pueden acceder con los puntos que logran acumular. Por otra parte, los clientes influenciables se pueden dirigir u orientar más fácilmente al canje mediante la comunicación, ya que son clientes que logran acumular menos puntos en promedio, por lo que pueden reaccionar mejor al comunicarle productos que pueden canjear para evitar perder sus puntos debido al vencimiento.
- Respuestas esperadas: Como se presumía, mayor edad implica menor “influenciabilidad” vía mail, que aunque no es el canal más digital que existe, sigue siendo digital. Además, en los extremos de edad ocurren cosas distintas, probablemente asociadas a preferencias más claras por parte de los clientes o totalmente inexistentes. Finalmente, se espera una respuesta positiva a mails de canje principalmente en clientes premium y normales del primer quintil.
- Utilización en el área: Gracias al modelo desarrollado, ahora se podrían identificar clientes que deban recibir un mail de canje debido a su respuesta esperada al mismo, evitando enviarles a quienes se espera que tengan una respuesta negativa al mail de canje.
- Beneficios potenciales: Se podrían esperar dos efectos asociados al punto anterior, el primero tiene relación con el aumento en la tasa de efectividad de los mails de canje, al orientarlos a clientes influenciables, mientras que el segundo tiene que ver con la

disminución de la tasa de des-suscripción, al evitar tocar a clientes que responderán negativamente.

6.2 ¿Qué le envió?

Una vez realizado lo anterior, se obtienen los clientes a los que se les va a comunicar un mail de canje según su respuesta esperada. Así, con la segunda parte, se pretende determinar qué línea de productos es más propenso a canjear el cliente, de forma tal de enviarle un mail a aquellos influenciables, pero con los productos que es más probable que escoja a la hora de convertir sus puntos. Todo esto con el objetivo final de enviar mails más efectivos, orientados a clientes influenciables y evitando llegar a los que tienen respuestas negativas a los mails de canje.

6.2.1 Análisis exploratorio

6.2.1.1 Comportamiento de los canjes y puntos canjeados

En primer lugar, se extrajo una base de datos que contenía los canjes en el año 2017, correspondientes a 1.376.543 canjes y a 24.958 Ítems distintos, considerando a 770.000 clientes canjeadores aproximadamente. A continuación, se presentan algunos gráficos de distribución de canjes y puntos canjeados según distintas variables, cabe destacar que es un análisis exploratorio, no conlleva un análisis estadístico profundo ni de relevancia de las variables escogidas.

- Por mes:

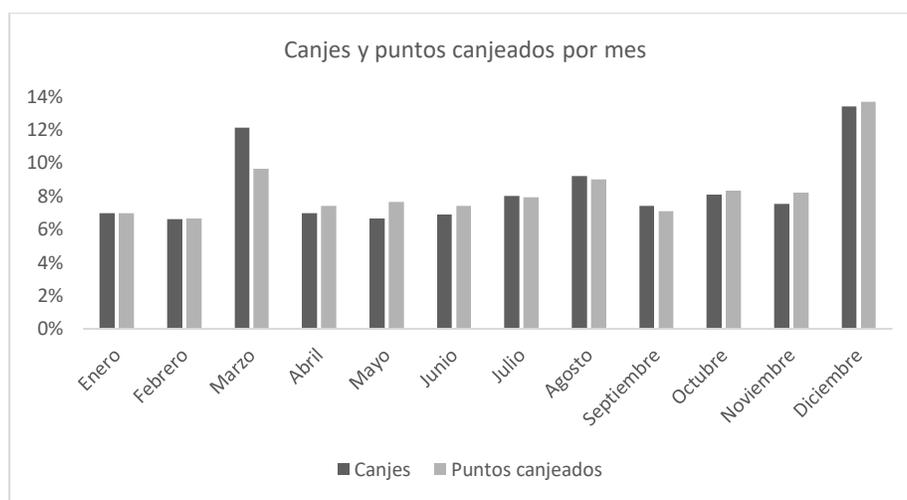


Figura 18: Canjes y puntos canjeados por mes.

Fuente: Elaboración propia con datos de la empresa

Se observa que marzo y diciembre concentran la mayor cantidad de canjes del año, lo que coincide con el ofrecimiento de canje de seguros automotrices y las fiestas de fin de año respectivamente.

- Por categoría de clientes (Donde Elite es la categoría más alta, Premium es la categoría intermedia y Normal es la categoría más baja, que concentra la gran mayoría de los clientes, por lo que dicha categoría se separa en quintiles de gasto con la tarjeta Q1 ... Q5):

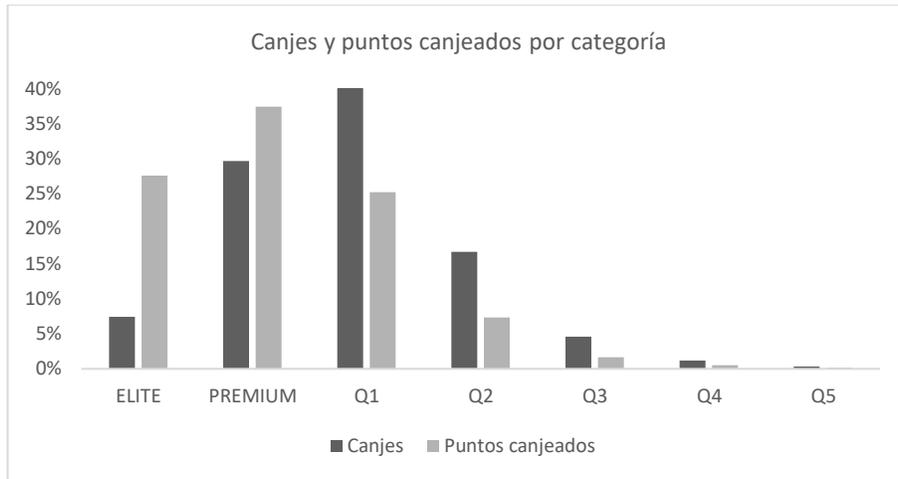


Figura 19: Canjes y puntos canjeados por categoría.

Fuente: Elaboración propia con datos de la empresa:

A pesar de que los clientes Premium (12,3% de la cartera) y Elite (3,7% de la cartera) son significativamente menos que los normales (84% de la cartera), tienen una cantidad importante de los canjes y puntos canjeados totales.

Dado esto, se estudia el porcentaje de la cartera completa que canjea abierto por categoría.

- Por distribución de canjes de los clientes por categoría con el lift asociado:

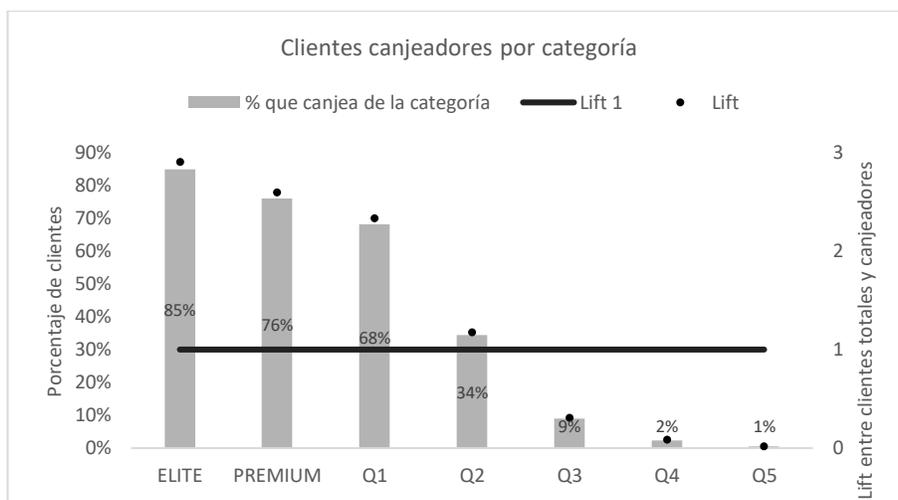


Figura 20: Distribución de canjes según categoría con lift asociado.

Fuente: Elaboración propia con datos de la empresa

De la figura anterior se desprende que la mayoría de los clientes Elite y Premium canjean, sin embargo, hay una gran oportunidad de incentivar a los clientes de categorías más bajas a que vivan la experiencia del canje, principalmente en niveles bajos, que además son los niveles más rentables para la empresa.

- Por género:

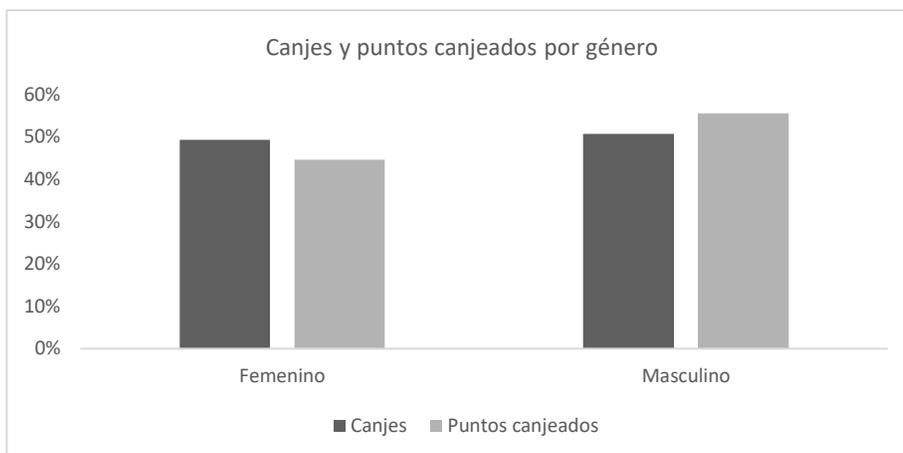


Figura 21: Canjes y puntos canjeados por género.

Fuente: Elaboración propia con datos de la empresa

Se puede observar que la cantidad de canjes es esencialmente igual según el género, sin embargo, los hombres canjean en promedio en niveles más altos, lo que se traduce en mayor cantidad de puntos canjeados.

- Por nivel socioeconómico:

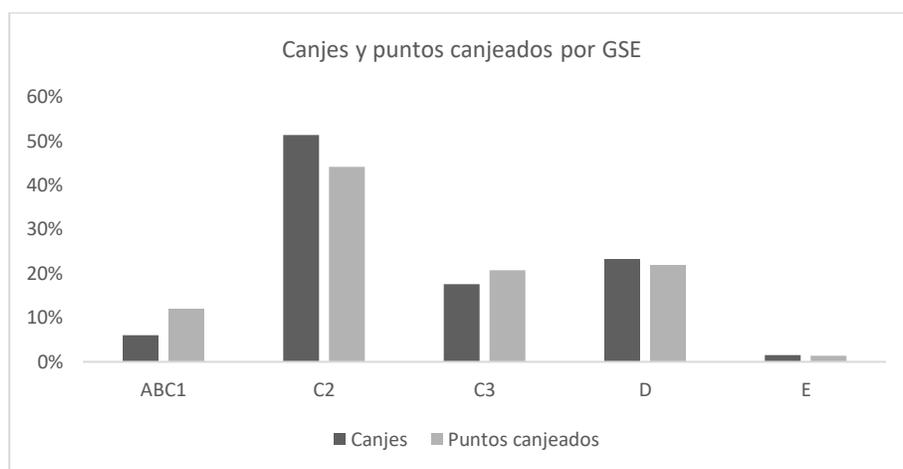


Figura 22: Canjes y puntos canjeados por GSE.

Fuente: Elaboración propia con datos de la empresa

En cuanto a la clasificación socioeconómica, la categoría que concentra más canjes y puntos canjeados es la C2, sin embargo, para las categorías ABC1 y C3, la cantidad de puntos canjeados superan a los canjes, lo que habla de canjes en mayores niveles y por lo tanto una disposición a gastar sus puntos mayor.

6.2.2 Definición de líneas de productos

El objetivo de esta parte es clasificar los productos canjeados en líneas que permitan manejar de mejor manera los datos, dado que por la granularidad de los canjes, que se presentan a nivel de SKU, no es factible llegar a un modelo tan preciso, por lo que se requiere de una agregación de los datos.

En la organización no se contaba con una clasificación completa y exhaustiva de los productos canjeados en algún elemento más agregado, por lo que se tuvo que manejar la base y agregarle campos de interés desde otras tablas existentes en la empresa.

Así, de la base analítica definida anteriormente que contaba con todos los canjes del año 2017, se realizaron transformaciones mediante minería de texto sobre las descripciones de los productos canjeados, para que pudiera ser rápidamente clasificado un producto canjeado en una línea de canje particular. Dichas transformaciones se realizaron mediante comandos de minería de texto que permiten extraer palabras claves de la descripción del producto y, según el contenido de la extracción, clasificarlos en algunas de las líneas de canje.

De esta forma, se llegó a:

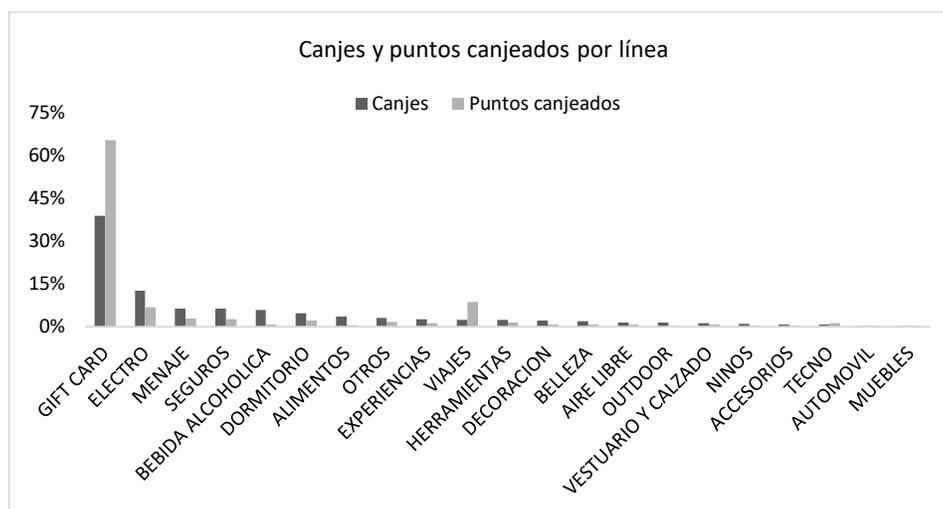


Figura 23: Canjes y puntos canjeados por línea.

Fuente: Elaboración propia con datos de la empresa

Al validar las líneas de productos de canje con la Gerencia de Fidelización, se acordó que las líneas de Gift Card, Seguros y Viajes no son gestionables, ya que tienen campañas periódicas definidas y gestionadas por un área distinta a la de fidelización, por lo que se eliminan del gráfico y se reescala al 35%, de manera tal de tener un mejor entendimiento del comportamiento de las otras líneas, obteniendo:



Figura 24: Canjes y puntos canjeados por línea re-escalado.

Fuente: Elaboración propia con datos de la empresa

Así, se procedió a escoger con qué líneas se trabajará finalmente, para lo que se consideraron los criterios:

- Representar al menos con una línea a cada negocio del holding (Tiendas por departamento, mejoramiento del hogar y supermercados)
- Concentrar la mayor cantidad de canjes y puntos canjeados posible
- Ser importantes para la Gerencia de Fidelización de Clientes¹⁹

Resultando:

Tabla 10: Porcentaje de canjes y puntos canjeados que representa cada línea sobre el total sin las líneas de gift card, seguros y viajes

Línea	Canjes	Puntos canjeados
Electro	24,2%	29,4%
Dormitorio y decoración	13,0%	12,3%
Menaje	12,3%	12,3%
Bebidas alcohólicas	11,3%	2,9%
Alimentos	6,8%	2,0%
Herramientas	4,4%	6,4%
Tecno	1,3%	5,8%
Total	73,2%	71,1%

Fuente: Elaboración propia

¹⁹ Según productos exitosos, campañas vigentes o periódicas, capacidad de negociación entre otras consideraciones

Cabe destacar que la tabla anterior muestra el porcentaje que representan del total de las líneas sin las 3 eliminadas anteriormente, sin embargo, los modelos finales serán desarrollados con toda la base analítica, las líneas de gift card, seguros y viajes fueron eliminadas de este análisis para otorgar un mejor entendimiento de la distribución de las otras líneas, pero si se eliminaran del modelamiento se tendría un modelo sesgado y no representativo de la realidad.

Las líneas de canje consideran:

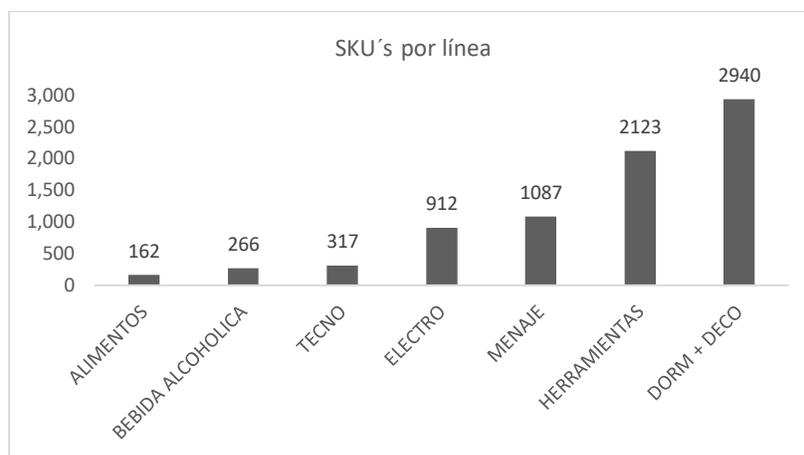


Figura 25: Cantidad de SKU's distintos por línea escogida.

Fuente: Elaboración propia con datos de la empresa

De lo anterior se desprende que, a priori, los clientes que canjean en las líneas de dormitorio y decoración, menaje y electro deben ser los más difíciles de perfilar, debido a la alta cantidad de productos distintos que considera y la consiguiente heterogeneidad que existe entre los clientes que canjean sus puntos en dichas líneas de productos.

Se deja fuera de este pseudo análisis previo a la línea de herramientas, ya que a pesar de tener numerosos sku's asociados, se tienen características particulares muy definidas de los clientes que canjean en esta línea, debido a que los productos que están en esta línea tienen muy poco margen de caer en otras líneas, son muy marcadas sus pertenencias.

6.2.3 Exploratorio para líneas escogidas

Considerando las líneas anteriormente expuestas, se procede a estudiar distintas variables que podrían ser relevantes para el desarrollo del modelo de propensión de líneas de productos.

El principal objetivo del análisis exploratorio de esta etapa del trabajo de título es la obtención de insights que permitan inferir características particulares de los clientes canjeadores de ciertas líneas en particular, teniendo una primera impresión de cuáles deberían ser las variables relevantes a la hora de desarrollar los modelos, sin embargo, hay un segundo objetivo, que consiste en la realización de un primer filtro de variables que no deberían ser utilizadas en el modelo, que posteriormente se complementa con un análisis de correlación.

En el Anexo 24: Distribución de canjes según género, se observa que para las líneas de canje de herramientas, tecnología y bebidas alcohólicas, destaca la participación de hombres,

mientras que las mujeres no predominan en ninguna línea en particular, sin embargo, para la línea de alimentos se observa un mayor porcentaje de canjes realizados por mujeres.

En el Anexo 25: Distribución de canjes según drivers de enganche, se presentan las distribuciones de los canjes según los drivers de enganche más importantes definidos en la organización. De los que se puede concluir que no existen diferencias significativas entre los canjes en relación a las variables de enganche, es decir, se mantienen relativamente similares a lo largo de las líneas, por lo que dichas variables no deberían impactar en los modelos posteriores para discriminar entre líneas de canje.

En el Anexo 26: Distribución de canjes según actividad en tiendas del grupo, se observan los canjes distribuidos según si el cliente es activo en las tiendas más importantes del grupo al que pertenece la empresa, estas son, mejoramiento del hogar y tiendas por departamento, donde activo quiere decir que tuvo compras con su tarjeta en la tienda en el último año móvil antes del canje. De donde se obtiene que la variable activo en mejoramiento del hogar tiene un impacto muy superior al promedio en la línea de herramientas y un poco menos marcado en la línea de tecnología. Mientras que la variable activo en tiendas por departamento tiene relevancia en la línea de electro, donde predominan los activos, y en las de herramientas y alimentos, donde predominan los activos pero hay mayor proporción de no activos en relación a las otras líneas.

En el Anexo 27: Distribución de canje según ingreso web, para la línea de tecnología se observa la mayor proporción de clientes con ingreso a la página web, mientras que para las líneas de dormitorio y menaje se tiene la menor, lo que habla de perfiles más o menos digitales en cada caso.

Finalmente, en el Anexo 28: Distribución de edad y antigüedad de los clientes canjeadores, se observan las distribuciones etáreas y de antigüedad de los clientes por línea, con el objetivo de determinar si existen diferencias a priori en la edad de los canjeadores de ciertas líneas y si el tiempo del cliente con la empresa refleja algún patrón de comportamiento frente al canje en las distintas líneas.

La conclusión principal de los gráficos anteriores tiene que ver con que la antigüedad no presenta diferencias significativas entre las líneas analizadas, sin embargo con respecto a la edad si se pueden extraer algunas conclusiones interesantes:

- Para la línea de herramientas se observan clientes principalmente jóvenes, donde el 60% de los canjes se concentran el tramo de 25 a 45 años.
- En el caso de electro, en el tramo de 25 a 60 años concentra el 80% de los canjes.
- Las líneas de alimentos y bebidas alcohólicas concentran gran porcentaje de sus canjes en clientes más jóvenes.
- Dormitorio y decoración, la línea de menaje y la de electro presentan clientes de mayor edad.

6.2.3.1 Composición de las líneas

Para comprender el peso de las líneas y los productos que corresponden a cada una, se presentan las características generales de los canjes por línea:

Tabla 11: Detalles de canjes por línea:

Línea	Cantidad de canjes	Porcentaje del total	Promedio de puntos	Productos más canjeados
Herramientas	31.973	2,32%	6.584	Set de herramientas (4.466 pp) Destornillador (3.177 pp)
Electro	174.342	12,67%	5.533	Minipimer (4.838 pp) Hervidor (4.647 pp)
Bebidas alcohólicas	81.324	5,91%	1.173	Pisco (1.000 pp) Cervezas (1.000 pp)
Alimentos	48.624	3,53%	1.364	Chocolates pascua (1.188 pp) Pollo asado (2.000 pp)
Tecno	9.355	0,68%	20.394	Audífonos (8.820 pp) Teléfonos (79.780 pp)
Dormitorio y deco	93.307	6,78%	4.311	Sábanas (4.692 pp) Toallas (4.711 pp)
Menaje	88.224	6,41%	4.564	Cuchillos (4.546 pp) Batería cocina (7.787 pp)

Fuente: Elaboración propia

Los detalles se explicitan en Anexo 19: Productos más relevantes según frecuencia de canje por línea.

De los datos anteriores se desprende que hay líneas que son fundamentalmente de canjes en niveles bajos, por ejemplo alimentos y bebidas alcohólicas, mientras que hay otras como tecno, que consideran una inversión de puntos mucho mayor por parte del cliente.

Por otra parte, se puede inferir que existen que son canjeadas principalmente porque son los puntos que alcanzó a juntar el cliente y los canjea para que no le venganzan y los pierda, como por ejemplo la línea de electro o dormitorio y decoración, que serían líneas para “liquidar” sus puntos, que considera productos que difícilmente un cliente desee canjearlos por preferencias intrínsecas, mientras que hay líneas como tecno y herramientas que involucran productos que el cliente muy probablemente estaba buscando obtener, ya sea comprando o canjeando, como por ejemplo un teléfono o una sierra caladora, siendo más de “destino”.

Dicho esto, sería razonable pensar que, al momento de desarrollar los modelos de propensión, la líneas definidas como de “destino” deberían considerar clientes más fáciles de perfilar, dado que sus preferencias son las que lo llevan a tomar la decisión de canjear en esa línea, mientras que las de “liquidar” pueden contener clientes más diversos, que son similares en su acumulación de puntos pero podría ser difícil encontrar un perfil claro, ya que hacen el papel de “comodín” del canje.

6.2.3.2 Variables

En el Anexo 18: Variables modelo líneas se detallan todas las variables que se construyeron para comenzar a trabajar con el modelo de canje en líneas de productos.

6.2.4 Preparación de datos

En esta sección se detallan los pasos previos al modelamiento que se desarrollaron para obtener una base analítica de datos lo más limpia y coherente posible.

6.2.4.1 Definición de la base analítica

En primer lugar y como se mencionó anteriormente, la base analítica para obtener la propensión a canjear en las líneas de canje corresponde a la información completa relativa a los canjes del año 2017, donde cada fila corresponde a un canje llevado a cabo por un cliente, al que se le agregaron datos dentro de las categorías:

- Demográficas: Género, estado civil, zona de residencia, etc.
- De contrato: Tipo de tarjeta que tiene, fecha de apertura de la tarjeta, cupo, disponible, etc.
- Transaccionales: Cantidad de compras en un periodo, monto gastado, gasto en los rubros definidos por la organización, días con compra, meses con compra, etc.
- De fidelización: Puntos acumulados, canjes anteriores, nivel de canje anterior, distancia al próximo nivel de canje, etc.
- De evolución: Que indican la variación de la variable desde periodos atrás al momento del canje, de puntos acumulados, compras, monto gastado, etc.

Además, para la clasificación de los productos en líneas, se desarrolló un código de minería de texto que permitió automáticamente clasificar los productos canjeados en una de las líneas definidas anteriormente, ya que en la empresa no se cuenta con una categorización de este estilo respecto de los productos relacionados al canje.

Finalmente la cardinalidad de la base resultante fue de 1.376.543 filas (canjes) por 150 columnas (variables relativas al canje mismo o características del cliente que lo llevó a cabo) aproximadamente.

6.2.4.2 Datos perdidos

La base analítica se construyó insertando campos provenientes de distintas bases de datos de la empresa, sin embargo, una de ellas dependía de la situación con respecto a la mora del cliente en el periodo de la consulta, por lo que se tenía una falta de datos que correspondía al 0,1% de los casos.

Se podría haber imputado los datos, sin embargo, como era la tabla que proveía más campos a la base analítica y la cantidad de datos faltantes era muy baja, se optó por eliminar los datos nulos.

6.2.4.3 Extracción y selección

La consolidación de los datos involucró un proceso de validación de los mismos, mediante la obtención del valor desde distintas fuentes, de manera tal de obtener los campos lo más precisos posible.

La primera extracción de datos consistió en eliminar los campos duplicados, manteniendo el que hiciera más sentido a los integrantes del área de Inteligencia de Clientes de la organización y al autor.

Luego, como ya se tenía categorizado los productos en líneas, se eliminaron campos relativos al producto canjeado que no se utilizarían posteriormente en el modelo, como la descripción del producto canjeado, el sku asociado, entre otros campos.

Posteriormente se llevó a cabo una metodología de filtro de pertinencia de las variables según el análisis exploratorio:

- Filtro a priori: Eliminar variables que por su naturaleza no afectan al explicar el canje de productos, como por ejemplo variables que indican si un cliente es activo o no en rubros como vivienda, combustibles, seguros, comunicaciones, entre otros.
- Filtro por descriptivos: Eliminar variables según los aprendizajes del análisis previo, como por la variable de antigüedad del cliente, dado que no aportaba información extra a la que aporta la edad, que es mucho más completa, la variable de estado civil y una que indica si compra o no en un rubro, dado que la variable de gasto en el rubro resume esa información y aporta más.
- Filtro por construcción: Eliminar las variables relacionadas a los drivers de enganche con la empresa, dado que no aportaban información relevante y además se construyen a partir de modelos que utilizan las mismas variables que se están ocupando para modelar.
- Filtro por correlación: Se realizó un análisis de correlación y se eliminaron las variables que tuvieran más de 0,6 de correlación, donde principalmente se desecharon variables que estaban calculadas a distintos horizontes de tiempo, como monto y cantidad de compras, manteniéndose las que consideran un periodo más largo de evaluación. También se eliminó la variable del monto gastado en 12 meses, por su alta correlación con el gasto en la tienda por departamento del holding, los puntos canjeados por su alta correlación con la cantidad de canjes y el disponible por su correlación con el cupo. A continuación se presenta un gráfico de correlación de las variables continuas más relevantes que se mantuvieron en el modelo.

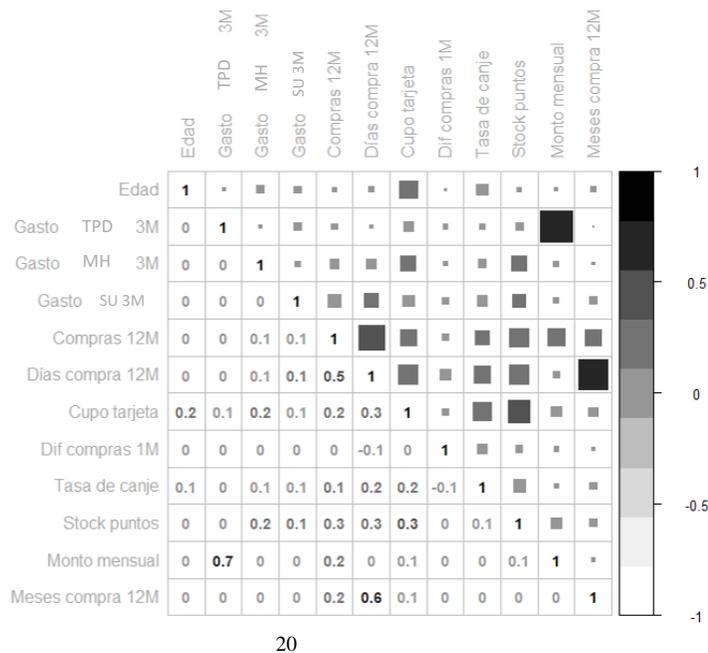


Figura 26: Gráfico de correlación de las variables continuas más importantes.

Fuente: Elaboración propia

Cabe destacar que las variables monto mensual y gasto en tiendas por departamento se mantuvieron ambas pero se evitó usar las mismas en el mismo modelo, dado que son variables muy importantes para explicar el canje en algunas líneas.

6.2.4.4 Transformación de datos

Para mayor interpretabilidad y facilidad en el modelamiento, se realizaron algunas transformaciones de datos:

- Datos continuos a categóricos: Las variables como recency de canje, tasa de lectura, cantidad de ingresos web, nivel de canje anterior, se categorizaron en rangos, sin embargo otras como la edad, los puntos acumulados, entre otras se mantuvieron como continuas.
- Variables binarias: Para lograr desarrollar de manera exitosa los modelos, se definieron variables binarias que aportaban a la identificación de ciertas conductas, como por ejemplo una variable dummy por cada línea, que indicara con un 1 si el cliente canjeó o no en esa ocasión en dicha línea.

²⁰ TDP: Tienda por departamento del grupo MH: Tienda de mejoramiento del hogar del grupo SU: Supermercado del grupo

6.2.4.5 Detección de datos fuera de rango

La identificación de outliers se realizó mediante la definición de la distancia de Mahalanobis, que es una métrica comúnmente utilizada para identificar valores fuera de rango en datos multivariados, implementada en el software R.

Se consideró el criterio de corte mayor a 0,99 para determinar que un valor está fuera de rango, obteniéndose 2,4% de los datos totales como outlier, los que fueron eliminados de la base analítica final para el modelamiento.

6.2.5 Modelamiento

Para determinar la propensión a canjear en una línea, se desarrolló la metodología explicada en la sección de flujo metodológico, pero con algunos pasos previos que se detallan a continuación.

- Preparación de los datos: Pulir la base analítica para evitar trabajar con datos faltantes o sesgos por construcción de variables.
- Filtro de correlación: Eliminar variables altamente relacionadas para evitar problemas de sobreajuste.
- Balanceo de la base: Como se modela la probabilidad de canjear en una línea en particular, se cuenta con una base muy desbalanceada, por lo que se procedió a realizar un “undersampling” de la base, que consiste en eliminar aleatoriamente registros de los casos de fracaso hasta obtener una base con 50% de casos de éxito (canje en la línea) y 50% de casos de fracaso.
- Determinación de variables relevantes: Para hacer comparables y manjeables los modelos que se desarrollarán en el siguiente punto, se realiza un paso previo para extraer las variables más importantes de cada caso, mediante la implementación de un árbol de decisión.
- Desarrollo de los modelos: Obtenidas las variables relevantes, se separa la base en 80% de los datos para entrenar los modelos y 20% para comprobar el funcionamiento y se implementan 4 modelos de propensión, debido a que, dentro de los modelos de minería de datos, son los que tienen un mejor ratio entre desempeño e interpretabilidad de resultados:
 - Árbol de decisión CHAID
 - Árbol de decisión C5
 - Árboles aleatorios
 - Regresión logística
- Comparación de modelos: Con el objetivo de implementar para cada caso el mejor modelo, se compararán los 4 mediante 3 métricas: Recall, Accuracy y ganancia de información, escogiendo el que tenga mejor desempeño en todas las métricas, en caso de no haber un resultado absoluto, se prioriza la ganancia de información.

Así, se obtendrá un modelo ganador para cada línea, que será el que se aplicará para comparar la propensión a canjear. Finalmente, se comparan los lift a canjear en cada línea, priorizando la que presente mayor lift.

6.2.6 Resultados

6.2.6.1 Canje en herramientas

Para la línea de canje herramientas, se desarrolló el árbol de decisión, obteniendo como resultado las siguientes variables relevantes:



Figura 27: Variables más relevantes para el canje en herramientas.

Fuente: Elaboración propia a partir de los resultados de un árbol de decisión CHAID

Como se presumía, las variables más relevantes para predecir el canje en herramientas tienen relación con el género de los clientes, el gasto de los mismos en la tienda de mejoramiento del hogar del holding y la edad, por lo que tiene todo el sentido esperado según los análisis exploratorios y el conocimiento del negocio.

Posteriormente, se corren los 4 modelos a comparar, obteniendo los siguientes resultados en las métricas relevantes:

Tabla 12: Métricas de evaluación para la línea de herramientas

Modelo\Métrica	Recall	Accuracy
Árbol C5	67,96%	63,71%
Árboles aleatorios	71,30%	62,60%
Logit	67,57%	66,53%
Árbol CHAID	65,12%	64,57%

Fuente: Elaboración propia

Donde el modelo de árboles aleatorios es el que tiene mejor recall, es decir, el ratio de casos que se indican positivos sobre todos los casos positivos es mayor que en los otros casos, mientras que para el caso de accuracy, que indica el ratio entre los casos que se categorizan

bien, ya sean positivos o negativos, sobre el total de casos, el que tiene mejor desempeño es la regresión logística.

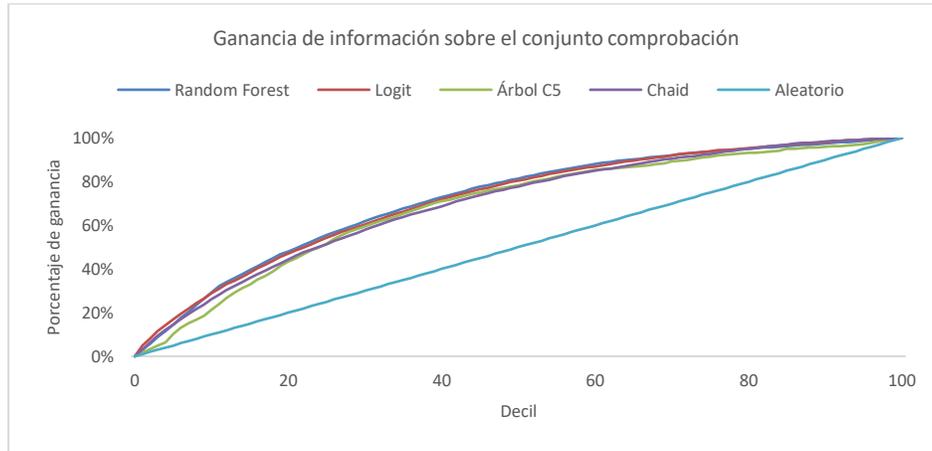


Figura 28: Ganancia de información por modelo para el canje en herramientas.

Fuente: Elaboración propia

La ganancia de información indica un mejor desempeño del modelo de árboles aleatorios, obteniendo un 48% de ganancia de información con solo el 20% de los casos totales, mientras que al tener un 40% de los casos totales se obtiene una ganancia de 73%.

Dado que el modelo de árboles aleatorios se antepone a los otros modelos en dos de las tres métricas, es el modelo escogido para la línea de herramientas.

6.2.6.2 Canje en electro

En el caso de la línea que comprende productos como hervidores, minipimers, entre otros, se obtuvieron como variables relevantes:

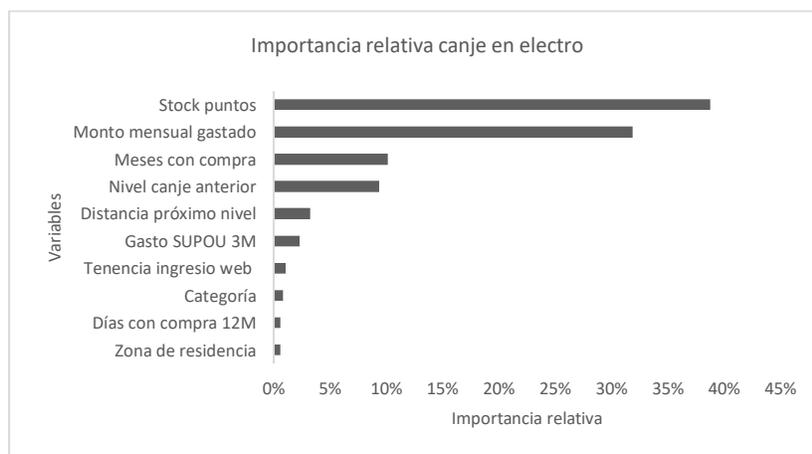


Figura 29: Variables más relevantes para el canje en electro.

Fuente: Elaboración propia a partir de los resultados de un árbol de decisión CHAID

Las variables anteriores fueron utilizadas para correr los modelos finales, obteniendo:

Tabla 13: Métricas de evaluación para la línea de electro

Modelo\Métrica	Recall	Accuracy
Árbol C5	66,39%	59,06%
Árboles aleatorios	72,20%	56,09%
Logit	62,22%	58,46%
Árbol CHAID	61,96%	58,48%

Fuente: Elaboración propia

Donde se puede observar que, a pesar de que el accuracy de los modelos es menor que en las otras líneas, es decir, es difícil para el modelo entregar un resultado certero, ya sea positivo o negativo, el recall es bastante alto para el caso del modelo de árboles aleatorios, lo que indica una buena categorización de los casos positivos.

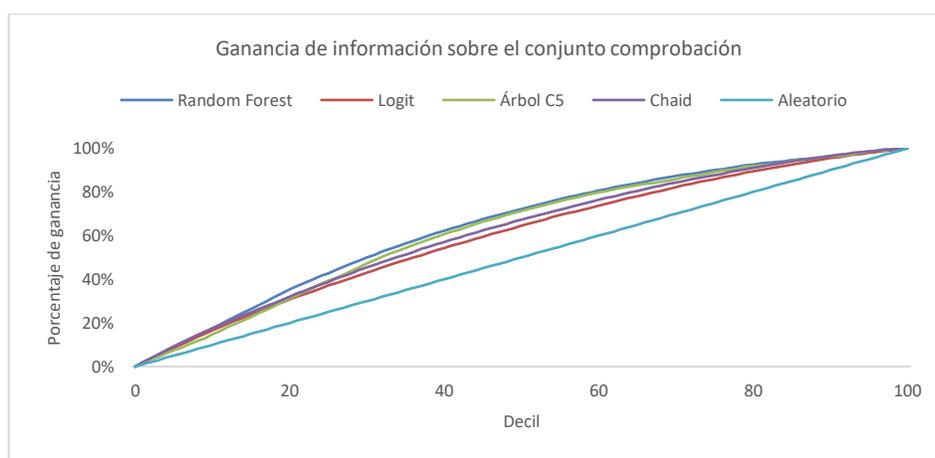


Figura 30: Ganancia de información por modelo para el canje en electro.

Fuente: Elaboración propia

Con respecto a la ganancia de información, el modelo de árboles aleatorios nuevamente es el que tiene un mejor desempeño, obteniendo un 35% y un 62% de ganancia de información al alcanzar el 20% y 40% de la población respectivamente.

6.2.6.3 Canje en bebidas alcohólicas

Como era de esperarse, las variables más relevantes del canje en beidas alcohólicas tienen relación con el comportamiento transaccional de los clientes en el supermercado del holding, lugar donde se concentra el canje de bebidas alcohólicas, además del stock de puntos, que tiene sentido porque, como se detalló en la sección anterior, los canjes son fundamentalmente con promociones especiales en niveles de mil puntos.



Figura 31: Variables más relevantes para el canje en bebidas alcohólicas.

Fuente: Elaboración propia a partir de los resultados de un árbol de decisión CHAID

Tabla 14: Métricas de evaluación para la línea de bebidas alcohólicas

Modelo\Métrica	Recall	Accuracy
Árbol C5	88,25%	79,12%
Árboles aleatorios	81,00%	76,10%
Logit	76,59%	73,87%
Árbol CHAID	72,47%	74,46%

Fuente: Elaboración propia

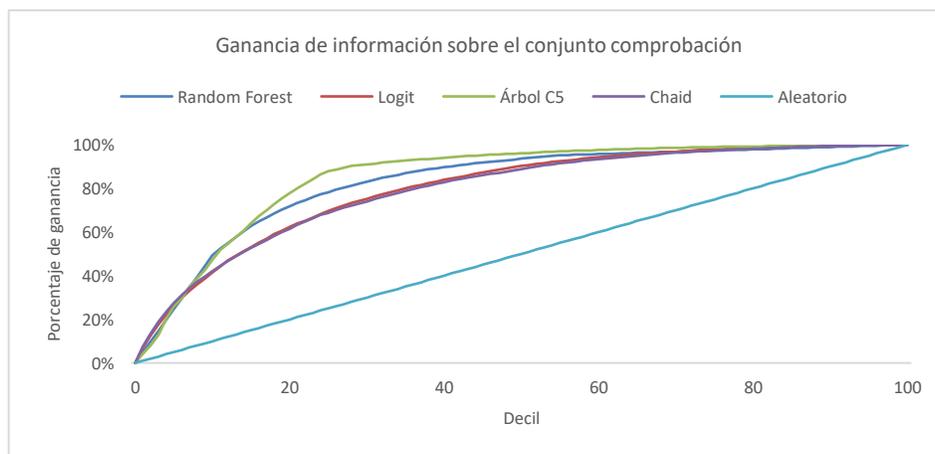


Figura 32: Ganancia de información por modelo para el canje en bebidas alcohólicas.

Fuente: Elaboración propia

Tanto en las primeras métricas como en la ganancia de información, el árbol C5 obtiene los mejores resultados, alcanzando una ganancia de información casi máxima, de un 94% al considerar el 40% mejor de la población. Cabe destacar que el, contrario a la intuición, un modelo de árboles aleatorios pierde frente un árbol simple solo porque son distintos árboles los que se realizan en ambos casos.

6.2.6.4 Canje en alimentos

Al igual que lo que ocurre en bebidas alcohólicas, en alimentos la variable más importante es si el cliente es activo o no en el supermercado del grupo, seguido de su comportamiento de gasto y de acumulación de puntos, ya que en alimentos también se concentran muchos canjes en niveles bajos, de mil y dos mil puntos.

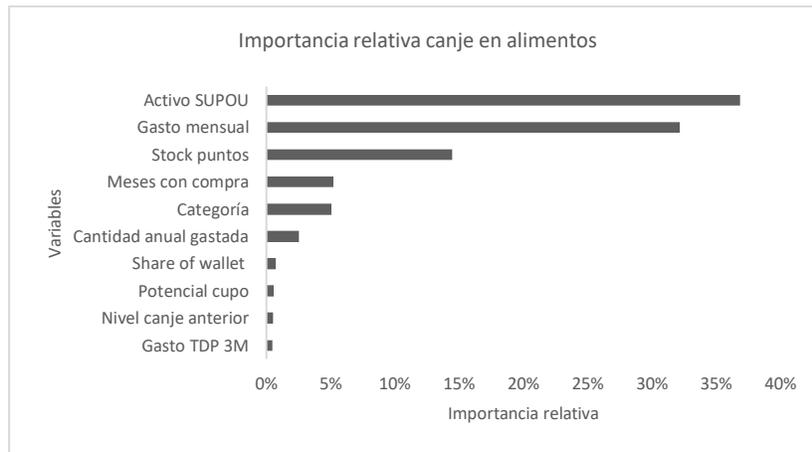


Figura 33: Variables más relevantes para el canje en alimentos.

Fuente: Elaboración propia a partir de los resultados de un árbol de decisión CHAID

Tabla 15: Métricas de evaluación para la línea de alimentos

Modelo\Métrica	Recall	Accuracy
Árbol C5	84,95%	83,49%
Árboles aleatorios	85,56%	81,09%
Logit	78,09%	76,17%
Árbol CHAID	78,28%	78,10%

Fuente: Elaboración propia

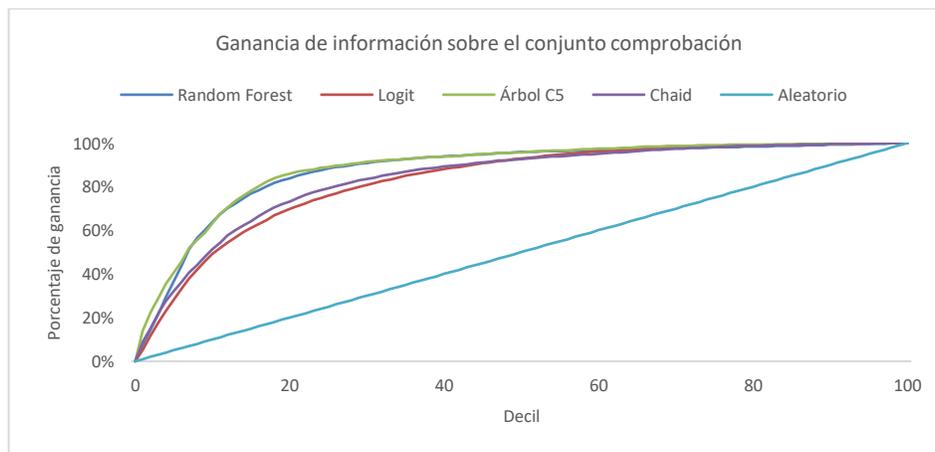


Figura 34: Ganancia de información por modelo para el canje en alimentos.

Fuente: Elaboración propia

En este caso, las métricas son muy parecidas entre el árbol C5 y árboles aleatorios, sin embargo, este último presenta mejor ganancia de información, con un 94% al alcanzar el 40% de la población, por lo que super a los demás en dos de tres métricas de desempeño.

6.2.6.5 Canje en tecno

Canjes en tecnología implican mayor cantidad de puntos promedio que el resto de líneas, por lo que la variable stock es la más relevante, seguida de demográficas como edad y género.



Figura 35: Variables más relevantes para el canje en tecno.

Fuente: Elaboración propia a partir de los resultados de un árbol de decisión CHAID

Tabla 16: Métricas de evaluación para la línea de tecno

Modelo\Métrica	Recall	Accuracy
Árbol C5	63,57%	67,51%
Árboles aleatorios	67,69%	65,11%
Logit	59,88%	63,51%
Árbol CHAID	64,69%	65,22%

Fuente: Elaboración propia

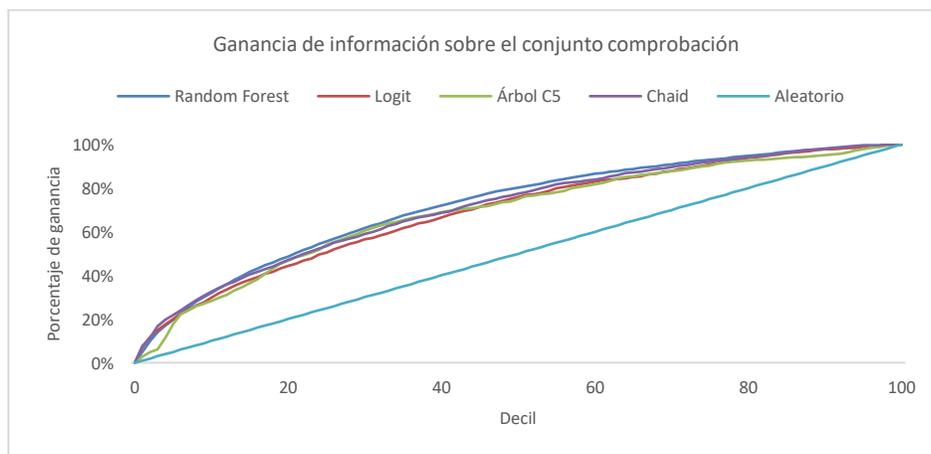


Figura 36: Ganancia de información por modelo para el canje en tecno.

Fuente: Elaboración propia

Nuevamente el mejor desempeño en las métricas estudiadas corresponde al modelo de árboles aleatorios, que presenta un 72% de ganancia de información al llegar al 40% de los casos totales y un recall superior al 65%.

6.2.6.6 Canje en dormitorio y decoración

El modelo para la línea de dormitorio y decoración otorgó como variables más relevantes al nivel de canje anterior, variable categórica que considera también a los clientes que no han tenido canjes en su vida, además de variables de comportamiento digital, como la tasa de lectura de mails y una variable de si el cliente ingresó o no a la web en los tres meses anteriores al canje.

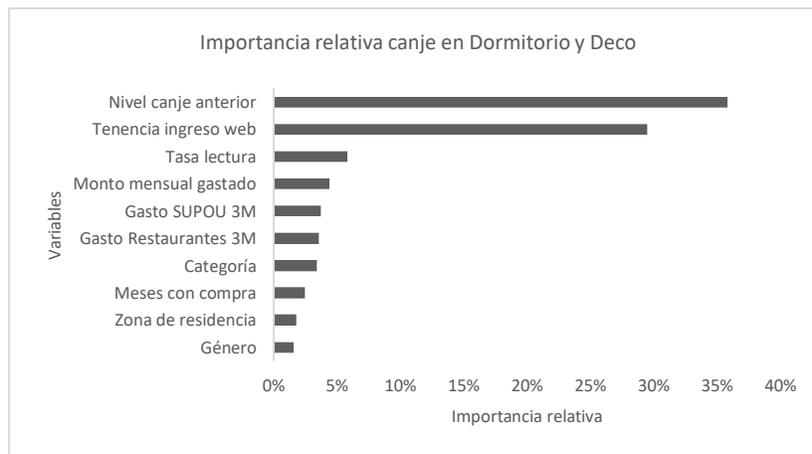


Figura 37: Variables más relevantes para el canje en dormitorio y decoración.

Fuente: Elaboración propia a partir de los resultados de un árbol de decisión CHAID

Tabla 17: Métricas de evaluación para la línea de dormitorio y decoración

Modelo\Métrica	Recall	Accuracy
Árbol C5	67,04%	59,46%
Árboles aleatorios	65,26%	60,68%
Logit	61,89%	61,12%
Árbol CHAID	63,55%	57,88%

Fuente: Elaboración propia

Para la presente línea se obtienen mejores resultados para el árbol C5, a pesar de lo similares que son con los resultados del modelo de árboles aleatorios, ya que el C5 obtiene un mejor desempeño en ganancia de información, obteniendo un 64% al considerar el 40% de la población.

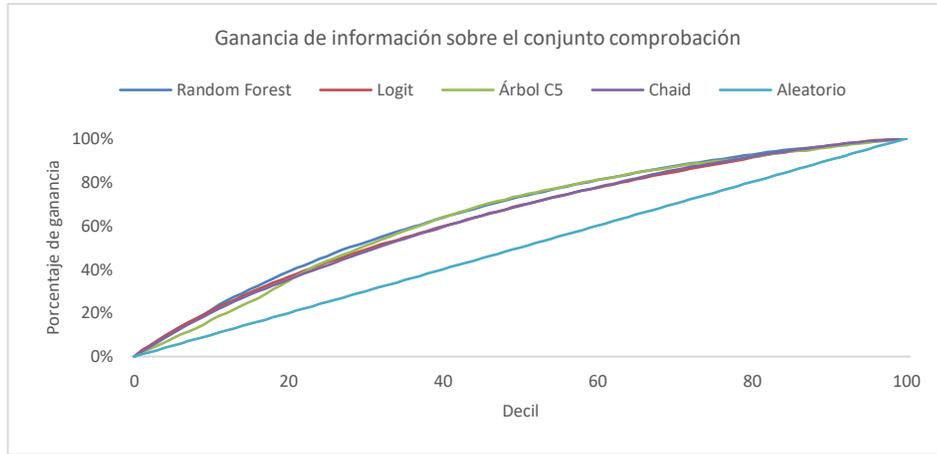


Figura 38: Ganancia de información por modelo para el canje en dormitorio y decoración.

Fuente: Elaboración propia

6.2.6.7 Canje en menaje



Figura 39: Variables más relevantes para el canje en menaje.

Fuente: Elaboración propia a partir de los resultados de un árbol de decisión CHAID

El canje en menaje implica como variables más relevantes a los puntos acumulados, seguidos del monto gastado y utilización de la tarjeta a lo largo del último año móvil, entregando como resultado que el mejor modelo es el de árboles aleatorios, seguido muy de cerca por el árbol CHAID, pero que no logra superarlo en ninguna métrica.

Tabla 18: Métricas de evaluación para la línea de menaje

Modelo\Métrica	Recall	Accuracy
Árbol C5	69,80%	57,94%
Árboles aleatorios	71,98%	56,77%
Logit	64,25%	59,87%
Árbol CHAID	71,70%	52,21%

Fuente: Elaboración propia

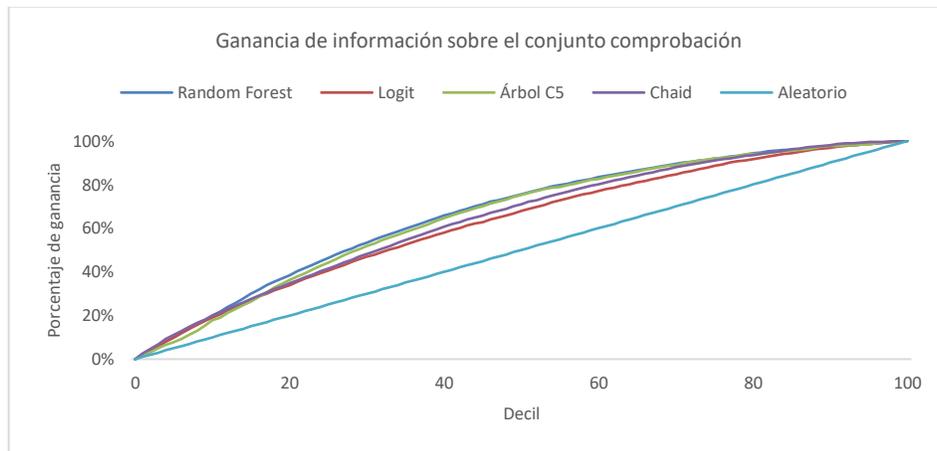


Figura 40: Ganancia de información por modelo para el canje en menaje.

Fuente: Elaboración propia

6.2.7 Nodos más y menos propensos

Si bien las métricas y resultados mostrados anteriormente son potentes a la hora de determinar el desempeño de un modelo y compararlo con bases sólidas frente a otros, no sirven para obtener insights y conocimiento relevante a la hora de gestionar clientes y escoger qué enviarle a un cliente en particular, es por esto que mediante los nodos más y menos propensos se pueden comprender los perfiles que tienen los clientes que escogen una línea de canje particular, indicando una mirada previa de las reglas de clasificación que se utilizarán y el tipo de clientes que son más cercanos o lejanos a cada línea.

6.2.7.1 Herramientas

El nodo más propenso a canjear en herramientas representa el 3,6% de los casos con la base balanceada, correspondiente a 1.832 casos, son clientes que cumplen con:

- Género: Hombres.
- Edad: Jóvenes, menores a 34 años.
- Relación con la tienda de mejoramiento del hogar: Activos en la tienda del grupo en un año móvil.
- Días con compras: Tienen entre 30 y 100 días con compras al año.
- Relación con otras tiendas del grupo: Inactivos en el supermercado del grupo en un año móvil.

El lift del nodo más propenso a canjear en herramientas es de $78/50 = 1,56$.

El nodo menos propenso a canjear en herramientas posee 1.006 casos, que representan el 2% del total y cumplen con las siguientes condiciones:

- Género: Mujer.
- Relación con la tienda de mejoramiento del hogar: Sin gasto en la tienda en el último año móvil.

- Relación con otras tiendas del grupo: Con gasto en la tienda por departamento del grupo mayor a \$70.000 los 3 meses anteriores al canje.
- Edad: Mayor a 50 años.

El lift del nodo menos propenso a canjear en herramientas tiene un valor de $90/50 = 1,8$.

6.2.7.2 Electro

En el caso de electro, los nodos más y menos propensos corresponden respectivamente a clientes que tienen las siguientes características:

- Stock de puntos: Posee al mes anterior al canje en torno a los 5.000 puntos.
- Monto mensual gastado: Gasta mensualmente menos de \$130.000 con la tarjeta.
- Meses con compras: Utiliza su tarjeta entre 6 y 10 meses al año.

Representa el 1,1% de los canjes con 3.063 casos y un lift de $70/50 = 1,4$.

- Stock de puntos: Posee más de 24.000 puntos al momento antes del canje.
- Meses con compras: Utiliza la tarjeta los 12 meses del año.
- Historial de canje: Ha canjeado antes en tramos altos, mayores a 12.000 puntos.

Corresponde a 2.507 casos que representan el 0,9% del total y presenta un lift de $85/50 = 1,7$.

6.2.7.3 Bebidas alcohólicas

El nodo más propenso a canjear en bebidas alcohólicas presenta los siguientes atributos:

- Relación con el supermercado del holding: Activo en el supermercado en el último año móvil.
- Stock de puntos: Tiene al momento anterior al canje menos que el primer nivel de canje formal que es de 5.000 puntos.
- Meses con compra: En promedio utiliza su tarjeta 10 meses en el año.
- Género: Hombre.

Concentra un 2,1% de los casos totales, con 2.644 y un lift de $94/50 = 1,88$.

El nodo menos propenso para esta línea cumple:

- Relación con el supermercado del holding: No activo en el supermercado en el año móvil anterior al canje.
- Stock de puntos: Posee más de 22.000 puntos al momento anterior al canje.
- Historial de canje: Tiene canjes anteriores en niveles mayores a 9.000 puntos.

Corresponden a 2.180 casos que son el 1,7% del total y tiene un lift de $96/50 = 1,92$.

6.2.7.4 Alimentos

Los nodos más y menos propensos cumplen respectivamente con las siguientes condiciones:

- Relación con el supermercado del holding: Activo en el supermercado en el último año móvil.
- Stock de puntos: Que acumula menos de 3.000 puntos al momento antes del canje.
- Relación con la tienda de mejoramiento del hogar del holding: Tiene poco gasto en dicha tienda con la tarjeta en los últimos 3 meses.
- Cupo en la tarjeta: Tiene un cupo menor a los \$500.000.

Representa el 3,8% de los casos, con 2.987 de veces, y un lift de $90/50 = 1,8$.

- Relación con el supermercado del holding: Inactivo en el supermercado en el último año móvil.
- Meses con compras: Con más de 11 meses con compras en el año anterior al canje.
- Monto mensual gastado: Gasta en torno a los \$300.000 mensualmente con la tarjeta.
- Categoría: Principalmente son de categorías Premium y Elite.

Son 1.666 casos que representan el 2,1% y tiene un lift de $97/50 = 1,94$.

6.2.7.5 Tecno

En el caso de la línea tecno, hay dos nodos más propensos con características bastante disímiles, por lo que se presentan ambos:

- Stock de puntos: Tiene muchos puntos al momento anterior al canje, más de 45.000.
- Meses con compra: Utiliza su tarjeta todos los meses del año.
- Tasa de lectura: Tiene una muy buena tasa de lectura de los mails recibidos, más del 50%²¹.

Como esta línea tiene considerablemente menos canjes que las demás, considera 375 casos que representan 2,7% con un lift de $92/50 = 1,84$.

Nodo muy propenso especial:

- Stock de puntos: Tiene pocos puntos al momento anterior al canje, menos de 4.000.
- Monto mensual gastado: Tiene un gasto mensual menor a \$90.000.
- Relación con el supermercado del holding: Es activo en el supermercado en el último año móvil.

Este nodo más propenso “especial” representa un 1,3% de los casos con 175 ocurrencias y un lift de $79/50 = 1,58$. Cabe destacar que este nodo propenso se generó a partir de un canje promocional vigente en 2017 que se ofreció en el supermercado del holding correspondiente a un artículo de audio en un nivel de puntos menor a 5.000.

El nodo menos propenso a canjear en tecno corresponde a:

²¹ Alto frente al promedio de los envíos de puntos, que corresponde a un 40% promedio según datos del área de gestión de clientes de la empresa del año 2018 hasta el 31 de agosto de 2018.

- Stock de puntos: Posee entre 5.000 y 12.000 puntos al momento anterior al canje.
- Edad: Tiene más de 55 años.
- Género: Mujer.

Concentra un 3,7% de las ocurrencias con 518 casos y un lift de $80/50 = 1,6$.

6.2.7.6 Dormitorio y decoración

El nodo más propenso a canjear en la línea de dormitorio y decoración contempla clientes con las siguientes características:

- Login en la página de la empresa: No tiene ingreso a la web en los 3 meses anteriores al canje.
- Historial de canje: Ha canjeado antes en tramos bajos, menores o iguales a 5.000 puntos.
- Relación con el supermercado del holding: Tiene poco o nulo gasto en el supermercado del grupo en los 3 meses anteriores.

Corresponde al 5,2% de los casos con 7.739 ocurrencias y un lift de $70/50 = 1,4$.

El nodo menos propenso a canjear en la línea ocurre 2.477 veces, representando el 1,7% de los casos totales, con un lift de $82/50 = 1,64$ y cuyas características son:

- Login en la página de la empresa: Tiene ingreso web en los últimos meses antes del canje.
- Historial de canje: Ha canjeado antes en niveles altos, es decir, mayor a 12.000 puntos.
- Monto mensual gastado: Tiene un gasto importante mensualmente con su tarjeta, pero menor a \$1,5 MM.

6.2.7.7 Menaje

El nodo más propenso a canjear en menaje representa el 2,2% de los casos con 3.039 ocurrencias, con un lift de $69/50 = 1,38$ y que contempla clientes que cumplen con:

- Stock de puntos: Posee entre 3.500 y 6.000 puntos al momento anterior al canje.
- Meses con compras: Utiliza su tarjeta en promedio 8 meses en el año móvil anterior.
- Relación con el supermercado del holding: No es activo en el supermercado del grupo.

En el caso del nodo menos propenso, ocurre 2.428 veces que corresponden al 1,7% del total y tiene las siguientes condiciones sobre las variables relevantes:

- Stock de puntos: Tiene más de 22.000 puntos al momento anterior al canje.
- Monto mensual gastado: Tiene un gasto mensual con la tarjeta menor a los \$500.000.
- Categoría: A pesar de no gastar tanto mensualmente, pertenece a categorías altas, Premium o Elite.

El lift del nodo menos propenso es de $90/50 = 1,8$.

6.2.8 Conclusiones preliminares

Al finalizar los modelos completos de canje en líneas de productos, se pudo confirmar que:

- Existen líneas de productos que consideran muchos productos distintos pero que tienen perfiles claros en cuanto a los clientes que las prefieren, mientras que hay otras que son muy heterogéneas y dificultan el proceso de clasificación de clientes en las mismas.
- Una explicación plausible es el rol que cumple la línea de canje, ya sea una línea de “destino” a la que los clientes recurren porque buscan y gustan de ese tipo de productos, o una línea “comodín” o “liquidar”, que es principalmente utilizada para evitar el vencimiento de los puntos.
- En el caso de los canjes en bebidas alcohólicas, alimentos y el caso especial de tecno, una variable muy importante es que el cliente sea activo en el supermercado del grupo, esto debido a que dichos canjes se realizan ahí, entonces se podrían evaluar estrategias para comunicarle a clientes que no compran regularmente en dicho supermercado, para que acceda de igual forma a esos canjes y así aumentar el espectro de canjeadores de esas líneas.
- Hay líneas en las que se debe ser particularmente cuidadoso de no comunicarle a clientes que no tienen los puntos necesarios para canjear en ellas, como por ejemplo tecnología cuando se ofrecen teléfonos celulares o computadores, ya que puede generar un efecto adverso en el cliente.

7 Justificación económica

Como el canje de puntos no está directamente asociado con un ingreso para la empresa, de hecho, implica una inversión por parte de la misma. Lo anterior debido a que los puntos entregados a los clientes corresponden a una cuenta por cobrar, que puede ser canjeada en cualquier momento y por lo tanto se debe incurrir en un gasto para cubrirlo. Es difícil pensar en un beneficio directo, como si fuera la venta de un producto. No obstante, a largo plazo se pueden observar cambios en el comportamiento de los clientes que han canjeado en comparación con quienes no lo han hecho, tal como se presentó en la sección de justificación. Dicho esto, se puede argumentar económicamente mediante el estudio de distintos escenarios, teniendo en consideración los siguientes datos extraídos de la empresa:

- Anualmente canjean aproximadamente 800.000 clientes.
- Actualmente en la organización se cuenta con aproximadamente 2,8 millones de clientes con cuenta activa.
- 60% de los clientes con cuenta activa no ha canjeado nunca, es decir, más de un millón y medio de clientes.
- Del porcentaje de clientes que nunca ha canjeado, 13% tiene hoy en día más de 5.000 puntos.

Como el objetivo del trabajo de título es orientar mejor la comunicación de canje a los clientes y por lo tanto aumentar la tasa de respuesta, es decir, los canjes, se realiza un análisis de escenarios considerando aumentar la base de clientes canjeadores en distintos porcentajes, tomando en cuenta el gasto extra de \$143.477 que genera el canje en el mismo año del canje y \$89.953 en el año siguiente al canje, que fue determinado con el análisis llevado a cabo en la sección de justificación del presente informe y cuyos detalles son entregados allí.

Así, los resultados son los siguientes:

Tabla 19: Análisis de escenarios económicos

Ítem \ Porcentaje de clientes extra	0,1%	0,2%	0,3%
Clientes canjeadores extra	1.680	3.360	5.040
Gasto extra con la tarjeta en 2019	\$241.041.360	\$482.082.720	\$723.124.080
Gasto extra con la tarjeta en 2020	\$151.121.040	\$302.242.080	\$453.363.120
Gasto extra con la tarjeta total	\$392.162.400	\$784.324.800	\$1.176.487.200

Fuente: Elaboración propia con datos de la empresa y del análisis de justificación

De esta manera, se podría esperar un aumento en el gasto con la tarjeta, generado por estos nuevos clientes canjeadores. No obstante, pueden haber más factores a considerar, como la rentabilidad mayor que generan los canjes en niveles bajos, por lo que la comunicación se podría orientar a ese tipo de productos y a clientes que no pueden acceder a niveles más altos. Logrando que, con la comunicación adecuada de un producto atractivo y el estudio de si su respuesta esperada a un mail de canje es positiva, éstos pueden tomar la decisión de canjear y por ende elevar aún más el gasto con la tarjeta y de paso engancharse con el plan de puntos de la compañía.

8 Conclusiones

8.1 ¿A quién le envió?

8.1.1 De la metodología

La metodología Up-Lift es un campo aún no tan explorado por los académicos y la literatura en general en cuanto a problemas de marketing se trata (Rzepakowski & Jaroszewicz, 2012a). No obstante, es una herramienta potente para determinar el efecto incremental de una acción, mientras que con los modelos clásicos, se evalúa un conjunto de factores, sin aislar los efectos de la comunicación propiamente tal.

Sin embargo, como requiere de un piloto experimental previo para su desarrollo, aumenta su costo y dificultad de implementación, además de la posibilidad de incurrir en algún sesgo asociado a un mal diseño del experimento.

En este contexto, para el presente trabajo de título, se considera una metodología lo suficientemente robusta e interpretable para ser aplicada y replicada en la organización. Teniendo en cuenta que, para ser utilizada, se debería replicar el experimento sin limitar a la base de clientes a ser 50% sin canje anterior y 50% sin canje anterior en el último año móvil, además de poder extender el tiempo de evaluación del mismo y tener una base de clientes aún más grande que permita realizar comunicaciones separadas por tipos de clientes, debido a que la comunicación genérica de canje que se entregó en el experimento, tenía ejemplos de productos que podrían ser de más interés de algunos clientes que de otros.

8.1.2 De los resultados

Los resultados del modelo Up-Lift implementado se consideran exitosos, teniendo en cuenta las condiciones en las que se desarrolló la metodología expuestas anteriormente. Ya que se logró determinar un perfil de clientes cuya respuesta esperada a una comunicación vía email relacionada con el plan de fidelización de la organización es positiva, por lo tanto hay que comunicarles de todas maneras, además de determinar un perfil para los clientes cuya respuesta esperada es negativa, por lo que hay que evitar enviarles comunicaciones de puntos. Dicho perfilamiento hasta el día de hoy no está estudiado en la organización, por lo que no se realiza análisis alguno de los clientes que deberían recibir o no un mail de canje debido a su respuesta esperada al mismo.

Hay algunas hipótesis que pueden explicar el comportamiento de los clientes:

- Como el periodo de estudio del canje era acotado (dos semanas), puede que haya sido poco para que el cliente canjeara, dado que puede haber tomado la decisión pero esperar un tiempo para llevarlo a cabo, ya que no eran productos que estuvieran por un tiempo muy limitado dentro de la oferta de canje.
- Puede que el mail, dadas sus características informativas, le haya abierto el panorama frente al canje y haya pospuesto su canje al siguiente nivel, por lo que su respuesta esperada sería negativa.

A pesar de lo novedosos y relevantes que son los resultados expuestos, se requiere una nueva realización del piloto y posterior modelamiento para poder generalizar los resultados obtenidos a toda la cartera de clientes, ya que hoy no es del todo correcto hacerlo, debido a las condiciones especiales que se tuvieron que cumplir para la determinación de la base de clientes que se consideraría para el experimento.

8.2 ¿Qué le envió?

8.2.1 De la metodología

En consecuencia con lo difícil que resulta entender el fenómeno del canje en una empresa de retail financiero como en la que se desarrolló el trabajo de título, es que se han hecho muy pocas cosas por dicho tópico, dejando de lado los estudios del comportamiento de los clientes frente a algo que resulta tan importante como la determinación de qué producto va a escoger para utilizar sus puntos.

En este contexto, la metodología aplicada tiene un sustento importante desde la vereda relativamente opuesta pero a la vez muy similar, en la que se sitúa el comportamiento de compra de los clientes, situación que ha sido más ampliamente estudiada en la literatura, y donde se utilizan modelos de minería de datos como los aplicados en el presente trabajo.

El desarrollo metodológico implicó algunas consideraciones que podrían ser mejoradas en el futuro, como la definición de la base analítica, que consideró los canjes de un periodo de tiempo específico, dejando fuera del análisis a los clientes que no habían canjeado en el periodo. Lo anterior debido a que esta es la segunda parte de una metodología mayor, en la que se supone que se determina primero si el cliente va a reaccionar bien a un mail de canje, es decir, va a canjear si le comunico, por lo que luego solo habría que determinar, de entre las posibilidades que tiene, qué producto va a escoger. También se enfrentaron dificultades como el desbalanceo de la base, que concentra actualmente el 40% de los canjes en una sola línea, gift card.

Así, a pesar de las limitaciones propias de la metodología, se considera una buena aproximación para aumentar la tasa de canje de los clientes u orientar la comunicación de productos que se quieren potenciar.

Cabe destacar que el proceso metodológico es fácilmente testeable con un desarrollo experimental, que queda fuera del alcance del trabajo de título.

8.2.2 De los resultados

Los resultados obtenidos del estudio de los canjes de los clientes en las distintas líneas, confirman algunas de las hipótesis que se tenían previamente, como por ejemplo que hay líneas que son más fáciles de perfilar de acuerdo a los tipos de productos que se encuentran clasificados en ellas, como lo son las líneas de herramientas, bebidas alcohólicas y alimentos, mientras que hay otras que son más difícilmente representables, ya sea porque hay mucha diversidad de productos en ellas o porque son líneas “comodín” que básicamente son

utilizadas como una forma de no perder los puntos debido al vencimiento, como lo son las líneas de electro, menaje y dormitorio y decoración.

También hay que considerar que algunas líneas poseen resultados de los modelos más robustos, mientras que hay otras que, por las mismas razones expuestas antes, son menos potentes.

En definitiva son resultados dentro de lo esperado, que en algunos casos se pueden considerar muy robustos pero que en otros requieren de una segunda implementación de modelos o del desarrollo de algún experimento que permita medir el funcionamiento más allá del 20% de comprobación que se utilizó en estos casos.

8.3 Globales

Finalmente se obtuvo un desarrollo metodológico que se hace cargo tanto de la respuesta esperada de los clientes frente a una acción de marketing, como de sus preferencias de cara a los productos que pueden escoger para convertir sus puntos.

En su conjunto, el trabajo de título logra modelar el comportamiento de los clientes dentro del proceso íntegro de mailing relativo al canje. Problemática que, cuando es abordada, se hace de manera miope, enfocándose exclusivamente en variables de estado de los clientes y aplicando reglas simples y vagamente justificadas.

En este contexto, la metodología logra resolver un problema complejo, incluyendo un modelo exclusivo para estudiar el impacto de los mails de canje, que combinado al de propensión a líneas de productos, logran entregarle la comunicación precisamente a los clientes que quieren recibirla y con los productos que son de su preferencia, aumentando la efectividad esperada de los mails de canje.

9 Recomendaciones

Cabe destacar que actualmente en la organización no hay análisis desarrollados mediante minería de datos para determinar si un cliente va a tener una respuesta esperada positiva o negativa a un mail de canje, ni para determinar la línea que es más propenso a canjear, sino que se realizan filtros en algunas de las variables disponibles según criterios que se definen en base a la historia de los clientes, por ejemplo, para enviar un mail de canje en tecnología, se escogen como target a los clientes que han canjeado en tecnología antes y que cumple con la cantidad de puntos que involucra el canje comunicado.

En este contexto, los modelos propuestos son un avance en cuanto al funcionamiento actual del área, ya que permite incorporar inteligencia a la definición de las bases y orientar la comunicación a resultados más efectivos.

Así, se propone el sistema integral de recomendación de productos para canje cuyo funcionamiento es el siguiente:

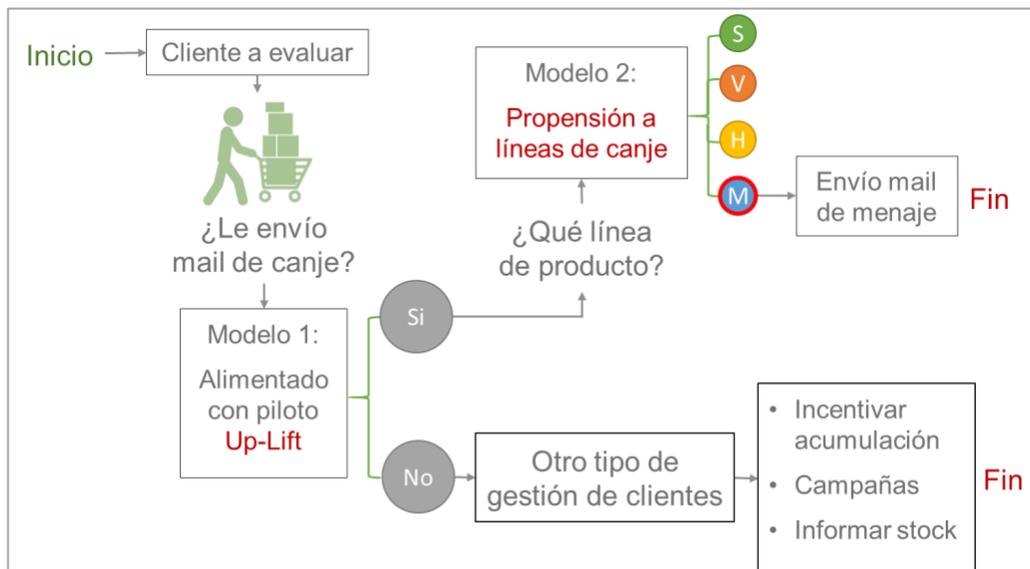


Ilustración 9: Flujo metodológico final del proyecto completo

Fuente: Elaboración propia

1. Con una frecuencia razonable, semanalmente por ejemplo, se debe aplicar el primer modelo a todos los clientes mediante reglas en lenguaje SQL, cuyo resultado será si un cliente es clasificado en la categoría “Influenciable” o en la “No molestar”. De esta manera se debería excluir (hasta que se actualice la marca) a los clientes categorizados en las bases de datos como “No molestar”, mientras que los “Influenciables” deberían ser la prioridad a la hora de enviar una comunicación de puntos.

2. Se debe recordar que los clientes no molestar reaccionan negativamente a una comunicación de canje, sin embargo, se deberían gestionar desde otras miradas, como por ejemplo ofrecerle alguna campaña de meta de compra con regalo de puntos, informar cuántos puntos ha acumulado y cuando vencen, entre otras acciones que podrían acercar al cliente al plan de fidelización sin comunicarle directamente un canje.
3. A los clientes influenciables se le aplica el modelo de líneas, también a través de reglas de SQL, que puede ser enfocado de dos maneras:
 - a. Si se tiene definida una campaña en cierta línea, se estudia el lift de los clientes a canjear en esa línea, y se envía a los que tengan un lift mayor estricto que 1, incluso se le puede exigir más al modelo y enviar a quienes tengan un lift superior, dependiendo la cantidad de clientes que se quiere incluir.
 - b. Si no se tiene una campaña en particular o se tiene más de una, se estudia el lift de los clientes influenciables a canjear en esas líneas y se escoge para un cliente la que presente mayor lift, siempre exigiendo que sea mayor que 1.

Así, se tiene un sistema integral de recomendación de productos para canje, que es integral dado que combina la respuesta esperada generada únicamente por la comunicación de puntos además de la probabilidad de canjear en una línea de productos, evitando problemas relacionados con impacto negativo de la comunicación, al no tocar a clientes “No molestar”.

10 Limitaciones y trabajo futuro

Para complementar y mejorar lo realizado, se propone trabajar en los siguientes aspectos:

Para la primera parte (Up-Lift):

- Replicar el diseño experimental considerando más clientes, más tiempo para analizar y variando en los ejemplos de productos que se muestran según el nivel de puntos que posee el cliente.
- Comparar otra forma de extraer las variables más importantes a la hora de categorizar a los clientes influenciados y no molestar, en este caso se utilizó un árbol de decisión pero podría existir una metodología que se ajuste mejor al caso particular.
- Proponer líneas de acción para los clientes que no caen en las categorías influenciados ni no molestar, ya que en el presente trabajo fueron obviadas, pero puede generar algo de valor al atacarlas.
- Utilizando como base la metodología propuesta, se puede replicar para otras comunicaciones existentes en la empresa, ya que el canje no es la única acción de marketing que se lleva a cabo, hay muchas otras líneas del negocio que envían comunicaciones que podrían manejarse de manera más eficientes y eficaz.

Para la segunda parte (Propensión a líneas de canje):

- Desagregar un poco más algunas líneas muy heterogéneas como electro, menaje y dormitorio y decoración para lograr resultados más precisos y orientados a cada caso.
- Incorporar variables que indiquen en qué se gastó una gift card canjeada, ya que son tantos canjes que sería interesante incorporarlo. Cabe destacar que no se hizo por limitaciones operacionales, actualmente no se puede distinguir entre una gift card canjeada y una comprada.
- Incorporar nuevas variables que aporten más información, como transaccionales dentro de las líneas estudiadas, de respuesta a campañas de mails pasadas, entre otras.
- Finalmente, se podría desarrollar un piloto experimental que mida la efectividad de los modelos desarrollados.

11 Bibliografía

- Aguirre, V. (2017). Perfilamiento de clientes influenciables en campañas de productos financieros en una empresa de retail financiero.
- Alliance, D.-P. (2015). Data-Pop Alliance White Papers Series Official Statistics , Big Data , and Human Development, (March).
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Drummond, C., Holte, R. C., Chawla, N. V., Sheng, V. S., Gu, B., Fang, W., & Wu, J. (2003). Exploiting the cost (in)sensitivity of decision tree splitting criteria. *International Conference on Machine Learning*, 66(1), 239–246. <https://doi.org/10.1.1.126.1087>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37. <https://doi.org/10.1609/aimag.v17i3.1230>
- Fernández, P. (1996). Investigación: Determinación del tamaño muestral. *Cad Aten Primaria Actualización*, 303(06), 138–14.
- Gallardo, C. (2016). Identificación de clientes con patrones de alta interacción con los drivers de una tarjeta de crédito.
- Guelman, L., Guillén, M., & Pérez-Marín, A. M. (2015). Uplift random forests. *Cybernetics and Systems*, 46(3–4), 230–248. <https://doi.org/10.1080/01969722.2015.1012892>
- Hadi, A. (1992). Identifying Multiple Outliers in Multivariate Data. Retrieved from <https://www.jstor.org/stable/2345856>
- Hilbert, M. (2017). Information Quantity, 1–4.
- Kamiński, B., Jakubczyk, M., & Szufel, P. (2018). A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research*, 26(1), 135–159. <https://doi.org/10.1007/s10100-017-0479-6>
- Kirk, R. E. (2015). Experimental Design. *The Blackwell Encyclopedia of Sociology*. <https://doi.org/10.1002/9781405165518.wbeose087.pub2>
- Letelier, I. (2017). Evaluación experimental de sistema de recomendación para campañas de email marketing.
- Mucherino, A., Papajorgji, P. J., & Pardalos, P. M. (2009). Data Mining in Agriculture, 34. <https://doi.org/10.1007/978-0-387-88615-2>
- Nalda, J. A. (2010). Sistema secuencial de recomendaciones personalizadas.
- Powers, D. M. W. (2007). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation, (December).

- Radcliffe, N. J. (2007). Using control groups to target on predicted lift: Building and assessing uplift model. *Direct Marketing Analytics Journal*, (3), 14–21.
- Radcliffe, N., & Surry, P. (2011). Real-world uplift modelling with significance-based uplift trees. *White Paper TR-2011-1, Stochastic ...*, (section 6), 1–33. Retrieved from <http://www.stochasticsolutions.com/pdf/sig-based-up-trees.pdf>
- Rokach, L., & Maimon, O. (2015). *Data mining with decisions trees. Theory and applications*. (H. Bunke & P. Wang, Eds.) (Second edi).
- Rzepakowski, P., & Jaroszewicz, S. (2012a). Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32(2), 303–327. <https://doi.org/10.1007/s10115-011-0434-0>
- Rzepakowski, P., & Jaroszewicz, S. (2012b). Uplift modeling in direct marketing. *Journal of Telecommunications and Information Technology*, 2012(2), 43–50.
- Starkweather, J., & Moske, A. K. (2011). Multinomial logistic regression. *Multinomial Logistic Regression*, 51(6), 404–410. <https://doi.org/10.1097/00006199-200211000-00009>
- Sumathi, S., & Sivanandam, S. N. (2007). Data mining tasks, techniques, and applications. *Studies in Computational Intelligence*, 29, 195–216. https://doi.org/10.1007/978-3-540-34351-6_7
- Troncoso, I. (2016). Estudio del efecto de un programa de fidelización en el comportamiento de compra de sus clientes.
- Tsoumakas, G., Katakis, I., & Vlahavas, I. (2010). *Data Mining and Knowledge Discovery Handbook. Journal of Chemical Information and Modeling*. <https://doi.org/10.1017/CBO9781107415324.004>
- Zhang, S., Zhang, C., & Yang, Q. (2010). Data preparation for data mining. *Applied Artificial Intelligence*, 17, 2003. <https://doi.org/10.1080/08839510390219264>

12 Anexos

12.1 Anexo 1: Valores y principios organizacionales

- Valores: La empresa basa su trabajo en 3 ejes fundamentales:
 - Transparencia: *“Para generar relaciones de confianza a largo plazo, sobre la base de la honestidad y claridad”*
 - Conveniencia: *“Orientada al beneficio mutuo, a partir de la cercanía, la flexibilidad y la propuesta de valor”*
 - Simplicidad: *“En la aproximación y promesas hacia nuestros clientes internos y externos, haciendo eficientes y ágiles nuestros procesos, con miras a la competitividad y, por lo tanto, la sostenibilidad de nuestras actividades”*
- Principios: Adicional a los valores, tiene cinco pilares que intentan definir el porqué de su trabajo diario y cómo deben desarrollarlo para cumplir con los objetivos últimos de la organización, estos principios son:
 - Nos apasionamos por los clientes
 - Jugamos en equipo
 - Somos protagonistas
 - Desarrollamos a las personas
 - Hacemos que las cosas pasen

12.2 Anexo 2: Estructura organizacional

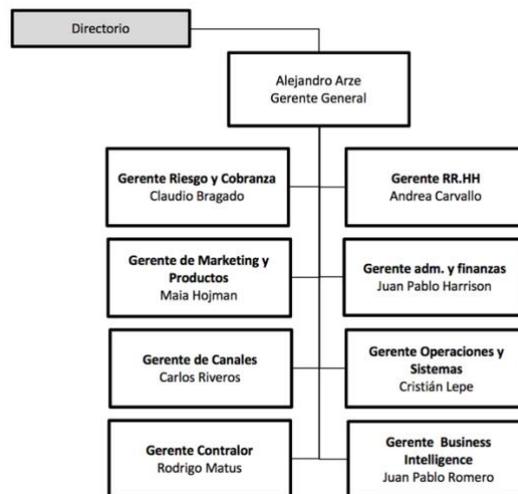


Ilustración 10: Organigrama de la empresa.

Fuente: Elaboración propia con datos del sitio web de la empresa actualizado a septiembre de 2017

12.3 Anexo 3: Estructura del área

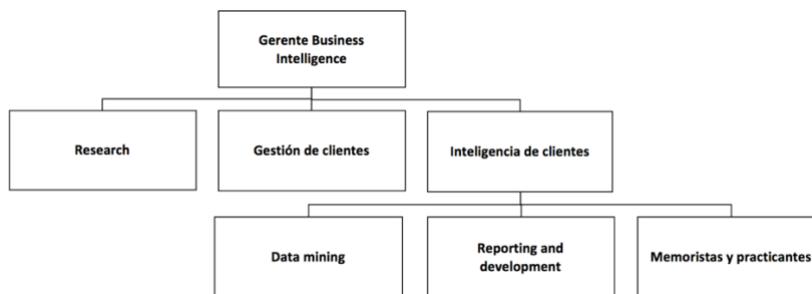


Ilustración 11: Organigrama gerencia de inteligencia de clientes de la empresa.

Fuente: Elaboración propia

12.4 Anexo 4: Reglas de acumulación



Ilustración 12: Reglas de acumulación de puntos.

Fuente: Sitio web de la empresa actualizado a septiembre de 2017

12.5 Anexo 5: Tabla de equivalencia de puntos

puntos	Valor
5.000 puntos	\$ 8.000
9.000 puntos	\$ 15.000
12.000 puntos	\$ 20.000
18.000 puntos	\$ 35.000
24.000 puntos	\$ 70.000
36.000 puntos	\$ 120.000
48.000 puntos	\$ 200.000
60.000 puntos	\$ 250.000
90.000 puntos	\$ 450.000
120.000 puntos	\$ 700.000
240.000 puntos	\$ 1.500.000

Ilustración 13: Conversión de puntos a valor de gift card.

Fuente: Sitio web de la empresa actualizado a septiembre de 2017

12.6 Anexo 6: Evolución de cuentas abiertas



Ilustración 14: Evolución de cuentas abiertas.

Fuente: Memoria anual 2016 de la empresa

12.7 Anexo 7: Sucursales y funcionarios



Ilustración 15: Cantidad de sucursales y funcionarios de la organización.

Fuente: Memoria anual 2016 de la empresa

12.8 Anexo 8: Evolución de la utilidad

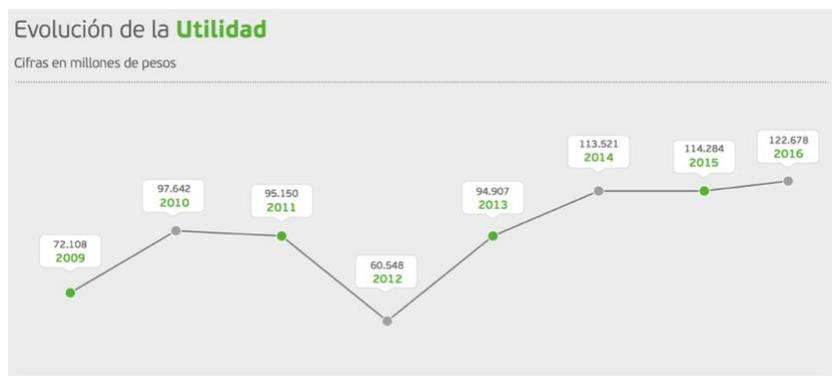


Ilustración 16: Evolución de la utilidad de la organización.

Fuente: Memoria anual 2016 de la empresa

12.9 Anexo 9: Trabajos anteriores

A continuación, se detallan memorias anteriores que utilizaron algunos de los modelos explicados en el marco conceptual, con el objetivo de establecer similitudes, obtener ideas de ellas y establecer parámetros de comparación.

Sistema secuencial de recomendaciones personalizadas en una empresa de home improvement (Nalda, 2010):

Este trabajo de título se enfocó en el estudio de los proyectos de una tienda de mejoramiento del hogar, mediante la aplicación de una metodología dividida en tres dimensiones:

- Caracterización de los clientes de la empresa.
- Perfilamiento de los tipos de proyectos.
- Combinación de las dos segmentaciones anteriores para realizar un análisis secuencial de canastas, mediante reglas de asociación.

Estudio del efecto de un programa de fidelización en el comportamiento de compra de sus clientes (Troncoso, 2016):

Más centrado en el programa de lealtad mismo, Isamar desarrolló la memoria cuyo objetivo consistió en estudiar el comportamiento de los consumidores en un programa de recompensas en relación con las decisiones de compra y canje que toman durante su trayectoria a la obtención de un premio. Objetivo que se abordó mediante el desarrollo de tres modelos distintos para resolver:

- Comportamiento durante la carrera, es decir, el comportamiento del cliente frente a la acumulación de puntos y el canje de los mismos por premios, considerando la aceleración en la acumulación dado un premio cercano, el efecto posterior al canje, entre otros. Dicho comportamiento se modeló mediante un modelo Tobit II, que permitiera medir cambios en los patrones de frecuencia y monto de compra causados por estímulos generados entorno al programa de lealtad.
- Preferencia por tipo de producto, dice relación con la oferta que tiene disponible la empresa frente a los canjes de los clientes, particularmente, se estudió el efecto de productos utilitarios (que resuelven una necesidad o prestan alguna utilidad) y hedonistas (aspiracionales, que entregan placer o estatus) dependiendo del total de puntos. Para estimar la elección discreta de un producto de canje del tipo hedonista, se utilizó un modelo Logit binario.
- Efecto del primer canje, se pretende evaluar si un cliente que ha canjeado, tiene un comportamiento distinto que uno que no lo ha hecho, mediante la aplicación de algoritmos de matching, específicamente vecinos más cercanos 1:1.

Identificación de clientes con patrones de alta interacción con los drivers de una tarjeta de crédito (Gallardo, 2016):

Esta memoria busca identificar a los clientes más propensos a tener una alta interacción con los drivers de la tarjeta de crédito. Esto se logró mediante la aplicación de modelos de regresión logística y árboles de decisión para estimar los efectos de cada uno de los drivers de uso de la tarjeta de crédito propuestos.

Perfilamiento de clientes influenciables en campañas de productos financieros en una empresa de retail financiero (Aguirre, 2017):

Este reciente trabajo de título se enfocó en el perfilamiento de clientes influenciables en campañas de productos financieros en un retail financiero. A grandes rasgos, se desarrollaron dos líneas de modelamiento, en la primera se confeccionaron experimentos que pretendían probar si es que el sugerir un producto mejoraba los resultados de una campaña y qué gancho es mejor para las diferentes categorías de clientes. En la segunda se buscó identificar a los clientes que efectivamente se pudieran influenciar, mediante la aplicación de un modelo Up-Lift directo basado en árboles de decisiones.

Evaluación experimental de sistema de recomendación para campañas de email marketing (Letelier, 2017):

Consistió en la realización de sistemas de recomendación de productos a comunicarse vía campañas por correo electrónico, en las que se compararon tres metodologías:

- Filtros colaborativos, relacionando un perfil de compra determinado con un producto, en función de la vecindad de clientes con perfil de compra similar. Esto se desarrolló mediante 4 pasos:
 - Cálculo del perfil de cada cliente
 - Matriz intermedia de perfiles agregados
 - Cómputo de los vecinos más cercanos
 - Cálculo de puntajes
- Reglas de asociación, mediante el algoritmo a priori del software SPSS.
- Asignación genérica, vale decir, sin incorporar un algoritmo, donde solo se invitaba al cliente a navegar por la página web de la organización.

Ésto se llevó a cabo con el objetivo de evaluar métodos para generar campañas de e-mail marketing personalizadas con una estrategia de ventas cruzadas en empresa de tiendas por departamento.

12.10 Anexo 10: Métricas de desempeño



Ilustración 17: Métricas de desempeño

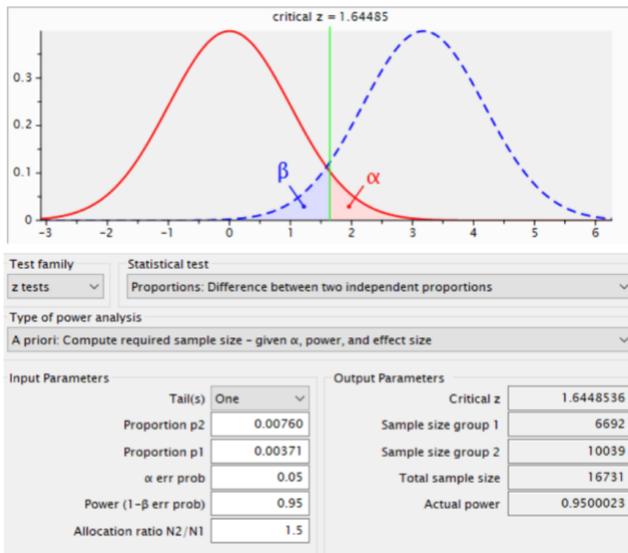
12.11 Anexo 11: Flujo de definición de base analítica



Ilustración 18: Flujo de definición de la base analítica para el cálculo del valor de canjea como justificación.

Fuente: Elaboración propia

12.12 Anexo 12: Cálculo tamaño mínimo de muestra



Con proporción p2 y proporción p1 las tasas de respuesta de grupos target y control respectivamente, de los envíos de Puntos entre el 1 de julio de 2017 y el 31 de marzo de 2018

Ilustración 19: Cálculo tamaño mínimo de muestra.

Fuente: Elaboración propia a través de gpowercalculator

12.13 Anexo 13: Pieza enviada para el target 1

¡Descubre todas nuestras opciones de cargo!

Cargos CMR Cards y veales para lo que quieras

Descubre a cuánto equivalen

5.000	€8.000
9.000	€15.000
12.000	€20.000
18.000	€35.000

Usa tus CMR Puntos y carga experiencias

1.000	3.000	10.000
-------	-------	--------

Domino's | Telepizza | SushiBlues

Ilustración 20: Piezas enviadas al target 1.

Fuente: Diseño propio, elaborado por marketing de la empresa

12.14 Anexo 14: Pieza enviada para el target 2



Ilustración 21: Piezas enviadas al target 2.

Fuente: Diseño propio, elaborado por marketing de la empresa

12.15 Anexo 15: Variables Up-Lift

Tabla 20: Variables disponibles para el modelo Up-Lift

Variable	Tipo	Descripción
ID_SEXO	Nominal	Género
EDAD, RANGO_EDAD	Numérica, Ordinal	Cantidad de años al mes anterior al canje
GSE	Ordinal	Clasificación socioeconómica en la actualidad
N_HIJOS, TIENE_HIJOS	Numérica, Nominal	Número de hijos en la actualidad, Binaria si es >0
N_VEHICULOS, TIENE_VEHICULO	Numérica, Nominal	Número de autos en la actualidad, Binaria si es >0
CATEGORÍA	Nominal	Elite, Premium y normal dividido en quintiles al mes anterior
CONTACTABLE_CELULAR, MAIL	Nominal	Binaria si es contactable por celular, Mismo por mail
ENGANCHADO_PUNTOS, FIN_PROM	Nominal	Binaria de si está enganchado con puntos, financiamiento, promoción
ACTIVO_RUBROS (16)	Nominal	Binarias de si utilizó su tarjeta en los 16 rubros existentes
POTENCIAL_GASTO	Numérica	Gasto potencial de un cliente
RENTABILIDAD, RANGO_RENT	Numérica, Ordinal	Rentabilidad anual que entrega el cliente, Rango
PUNTAJE_SCORE, RANGO_SCORE	Numérica, Ordinal	Rango según clasificación de riesgo al mes anterior
T1, T2T8, T3, T6, T7	Nominal	Binarias de si tiene o no dicha tarjeta al mes anterior
MONTO_CUPO, RANGO_CUPO	Numérica, Ordinal	Suma del cupo de los contratos al mes anterior, Rango
N_ADICIONALES, RANGO_ADICIONALES, TIENE_ADICIONAL	Numérica, Ordinal, Nominal	Cantidad de adicionales al mes anterior, Rango, Binaria si es >0
TIENE_AVANCE - SUPERAVANCE	Nominal	Binaria si tiene el producto al mes anterior
STOCK_PUNTOS, RANGO_STOCK	Numérica, Ordinal	Cantidad de puntos al mes anterior al canje, Rango
NIV_CANJE_ANT	Numérica	Nivel en el que realizó el canje anterior, 0 si no tuvo canjes
RANGO_RECENCY_CANJE	Ordinal	Rango de los meses que ha pasado de su último canje
DCC	Numérica	Días con compra en el último año

Fuente: Elaboración propia

12.16 Anexo 16: Variables finales Up-Lift

Tabla 21: Variables seleccionadas para el modelamiento Up-Lift

Categoría	VARIABLES
Demográficas	Edad, Zona residencia
De canje	Nivel canje anterior, Stock, Tasa de canje, Distancia al último canje
Transaccionales	Monto gastado 12 meses, DCC 6 y 12 meses, Cantidad gastada 3 y 12 meses
Contrato	Cupo, Potencial, Disponible
Del cliente	Share of wallet, Score
De evolución	Aceleración de stock del mes actual con respecto al anterior

Fuente: Elaboración propia

12.17 Anexo 17: Otros descriptivos líneas de canje





Figura 41: Compilado de descriptivos para líneas de canje.

Fuente: Elaboración propia

12.18 Anexo 18: Variables modelo líneas

Tabla 22: Variables disponibles para el modelamiento de líneas de canje

Variable	Tipo	Descripción
ID_SEXO	Nominal	Género
EDAD, RANGO_EDAD	Numérica, Ordinal	Cantidad de años al mes anterior al canje
GSE	Ordinal	Clasificación socioeconómica en la actualidad
N_HIJOS, TIENE_HIJOS	Numérica, Nominal	Número de hijos en la actualidad, Binaria si es >0
N_VEHICULOS, TIENE_VEHICULO	Numérica, Nominal	Número de autos en la actualidad, Binaria si es >0
CATEGORÍA	Nominal	Elite, Premium y normal dividido en quintiles al mes anterior
CONTACTABLE_CELULAR, MAIL	Nominal	Binaria si es contactable por celular, Mismo por mail
ENGANCHADO_PUNTOS, FIN_PROM	Nominal	Binaria de si está enganchado con puntos, financiamiento, promoción
ACTIVO_RUBROS (16)	Nominal	Binarias de si utilizó su tarjeta en los 16 rubros existentes
POTENCIAL_GASTO	Numérica	Gasto potencial de un cliente
RENTABILIDAD, RANGO_RENT	Numérica, Ordinal	Rentabilidad anual que entrega el cliente, Rango
PUNTAJE_SCORE, RANGO_SCORE	Numérica, Ordinal	Rango según clasificación de riesgo al mes anterior
T1, T2T8, T3, T6, T7	Nominal	Binarias de si tiene o no dicha tarjeta al mes anterior
MONTO_CUPO, RANGO_CUPO	Numérica, Ordinal	Suma del cupo de los contratos al mes anterior, Rango
N_ADICIONALES, RANGO_ADICIONALES, TIENE_ADICIONAL	Numérica, Ordinal, Nominal	Cantidad de adicionales al mes anterior, Rango, Binaria si es >0
TIENE_AVANCE - SUPERAVANCE	Nominal	Binaria si tiene el producto al mes anterior
STOCK_PUNTOS, RANGO_STOCK	Numérica, Ordinal	Cantidad de puntos al mes anterior al canje, Rango
NIV_CANJE_ANT	Categoría	Nivel en el que realizó el canje anterior, considera una categoría si no tuvo canjes
RANGO_RECENCY_CANJE	Ordinal	Rango de los meses que ha pasado de su último canje
DCC	Numérica	Días con compra en el último año
TASA_CANJE	Numérica	Indica porcentaje de puntos que son canjeados sobre el total acumulado
DISTANCIA_PROX_NIVEL	Numérica	Distancia en porcentaje al próximo nivel de canje
GASTO_RUBRO_3M (16)	Numérica	Gasto con la tarjeta en los 16 rubros categorizados en los 3 meses anteriores al canje
MCC	Numérica	Meses con compra en el último año móvil
MONTO_MENSUAL	Numérica	Monto mensual gastado con la tarjeta
N_LOGIN, TIENE_LOGIN	Numérica, Nominal	Cantidad de ingresos a la página web de la empresa en los últimos 3 meses, binaria si tuvo o no
TASA_LECTURA, RANGO_TL	Numérica, Ordinal	Porcentaje de correos que abre sobre los que recibe, Rango

Fuente: Elaboración propia

12.19 Anexo 19: Productos más relevantes según frecuencia de canje por línea

Tabla 23: Productos según frecuencia de canje para la línea de herramientas

Producto	Canjes	Puntos totales	Puntos promedio
SET HERRAMIENTAS	10.604	47.361.000	4.466
DESTORNILLADOR	5.125	16.280.000	3.177
TALADRO	4.466	44.178.000	9.892
SIERRA CALADORA	4.332	32.340.000	7.465
KIT TALADRO	2.308	23.219.000	10.060

Fuente: Elaboración propia

Tabla 24: Productos según frecuencia de canje para la línea de electro

Producto	Canjes	Puntos totales	Puntos promedio
MINIPIMER	36.142	174.866.000	4.838
HERVIDOR	26.043	121.022.000	4.647
LICUADORA	22.405	171.268.000	7.644
SECADOR DE PELO	22.027	105.408.000	4.785
ARROCERA	21.076	100.418.000	4.765

Fuente: Elaboración propia

Tabla 25: Productos según frecuencia de canje para la línea de bebidas alcohólicas

Producto	Canjes	Puntos totales	Puntos promedio
CAMPANARIO	7.943	7.943.000	1.000
CERVEZA HEINEKEN LATA 350	7.672	7.672.000	1.000
PISCO MISTRAL 750 CC	4.296	4.448.000	1.035
PISCO ALTO DEL CARMEN	4.110	4.206.000	1.023
PISCO MISTRAL 1 LITRO	3.799	3.799.000	1.000

Fuente: Elaboración propia

Tabla 26: Productos según frecuencia de canje para la línea de alimentos

Producto	Canjes	Puntos totales	Puntos promedio
CHOCOLATES PASCUA	16.471	19.562.000	1.188
POLLO ASADO	5.912	11.824.000	2.000
BOMBONES	4.874	5.713.000	1.172
CHOCOLATE	4.576	4.576.000	1.000
DULCES CARAMELOS	3.419	3.419.000	1.000

Fuente: Elaboración propia

Tabla 27: Productos según frecuencia de canje para la línea de tecno

Producto	Canjes	Puntos totales	Puntos promedio
AUDIFONOS	1.242	10.955.000	8.820
TELEFONO	971	77.466.000	79.780
PARLANTE	968	9.205.000	9.509
TARJETA MICRO SD	924	2.436.000	2.636
SOPORTE CELULAR	897	1.796.000	2.002

Fuente: Elaboración propia

Tabla 28: Productos según frecuencia de canje para la línea de dormitorio y decoración

Producto	Canjes	Puntos totales	Puntos promedio
SABANAS	37.659	176.714.000	4.692
TOALLA	19.340	91.102.000	4.711
DETERGENTE	7.863	24.064.000	3.060
MANTA	4.382	14.480.000	3.304
LAMPARA	4.341	16.694.000	3.846

Fuente: Elaboración propia

Tabla 29: Productos según frecuencia de canje para la línea de menaje

Producto	Canjes	Puntos totales	Puntos promedio
CUCHILLO/AFILADOR	53.674	244.001.000	4.546
BATERIA COCINA	9.028	70.301.000	7.787
TAZONES	4.175	5.027.000	1.204
JUEGO VAJILLA	3.587	18.060.000	5.035
OLLA	2.814	17.006.000	6.043

Fuente: Elaboración propia

12.20 Anexo 20: Matriz de correlación

Tabla 30: Matriz de correlación

Variable 1	Variable 2	Correlación	Variable 1	Variable 2	Correlación
MONTO_3M	MONTO_12M	0.9871857	MONTO_3M	MONTO_12M	0.9871857
MONTO_6M	MONTO_12M	0.9940449	MONTO_6M	MONTO_12M	0.9940449
GASTO_TPD_OU	MONTO_12M	0.6372469	GASTO_TPD_OU	MONTO_12M	0.6372469
GASTO_TPD_OU_3M	MONTO_12M	0.7434863	GASTO_TPD_OU_3M	MONTO_12M	0.7434863
MONTO_MENSUAL	MONTO_12M	0.9936495	MONTO_MENSUAL	MONTO_12M	0.9936495
MONTO_1M	GASTO_TPD_OU	0.7735572	MONTO_1M	GASTO_TPD_OU	0.7735572
MONTO_2M	GASTO_TPD_OU	0.7899397	MONTO_2M	GASTO_TPD_OU	0.7899397
MONTO_3M	GASTO_TPD_OU	0.6976006	MONTO_3M	GASTO_TPD_OU	0.6976006
MONTO_6M	GASTO_TPD_OU	0.6875995	MONTO_6M	GASTO_TPD_OU	0.6875995
MONTO_12M	GASTO_TPD_OU	0.6372469	MONTO_12M	GASTO_TPD_OU	0.6372469
GASTO_TPD_OT	GASTO_TPD_OU	0.6679373	GASTO_TPD_OT	GASTO_TPD_OU	0.6679373
GASTO_TPD_OU_3M	GASTO_TPD_OU	0.8949292	GASTO_TPD_OU_3M	GASTO_TPD_OU	0.8949292
MONTO_MENSUAL	GASTO_TPD_OU	0.6283774	MONTO_MENSUAL	GASTO_TPD_OU	0.6283774
MONTO_2M	GASTO_TPD_OT	0.6163807	MONTO_2M	GASTO_TPD_OT	0.6163807
GASTO_TPD_OU	GASTO_TPD_OT	0.6679373	GASTO_TPD_OU	GASTO_TPD_OT	0.6679373
GASTO_TPD_OT_3M	GASTO_TPD_OT	0.7779578	GASTO_TPD_OT_3M	GASTO_TPD_OT	0.7779578
GASTO_COMBUSTIBLE_3M	GASTO_COMBUSTIBLE	0.6340337	GASTO_COMBUSTIBLE_3M	GASTO_COMBUSTIBLE	0.6340337
GASTO_MH_OT_3M	GASTO_MH_OT	0.7861134	GASTO_MH_OT_3M	GASTO_MH_OT	0.7861134
GASTO_VIAJES_OT_3M	GASTO_VIAJES_OT	0.6886837	GASTO_VIAJES_OT_3M	GASTO_VIAJES_OT	0.6886837
GASTO_SUPERMERCADOS_OT_3M	GASTO_SUPERMERCADOS_OT	0.6572083	GASTO_SUPERMERCADOS_OT_3M	GASTO_SUPERMERCADOS_OT	0.6572083
GASTO_TRANSPORTE_3M	GASTO_TRANSPORTE	0.6375936	GASTO_TRANSPORTE_3M	GASTO_TRANSPORTE	0.6375936
GASTO_AUTOMOTRIZ_3M	GASTO_AUTOMOTRIZ	0.8789110	GASTO_AUTOMOTRIZ_3M	GASTO_AUTOMOTRIZ	0.8789110
GASTO_OTROS_OU_3M	GASTO_OTROS_OU	0.7539482	GASTO_OTROS_OU_3M	GASTO_OTROS_OU	0.7539482
GASTO_OTROS_OT_3M	GASTO_OTROS_OT	0.7066890	GASTO_OTROS_OT_3M	GASTO_OTROS_OT	0.7066890
GASTO_VIVIENDA_3M	GASTO_VIVIENDA	0.8551777	GASTO_VIVIENDA_3M	GASTO_VIVIENDA	0.8551777
GASTO_RECAUDACION_3M	GASTO_RECAUDACION	0.6454519	GASTO_RECAUDACION_3M	GASTO_RECAUDACION	0.6454519
GASTO_SEGUROS_OT_3M	GASTO_SEGUROS_OT	0.7744683	GASTO_SEGUROS_OT_3M	GASTO_SEGUROS_OT	0.7744683
MONTO_1M	GASTO_TPD_OU_3M	0.8318085	MONTO_1M	GASTO_TPD_OU_3M	0.8318085
MONTO_2M	GASTO_TPD_OU_3M	0.8136223	MONTO_2M	GASTO_TPD_OU_3M	0.8136223
MONTO_3M	GASTO_TPD_OU_3M	0.7848130	MONTO_3M	GASTO_TPD_OU_3M	0.7848130
MONTO_6M	GASTO_TPD_OU_3M	0.7806287	MONTO_6M	GASTO_TPD_OU_3M	0.7806287
MONTO_12M	GASTO_TPD_OU_3M	0.7434863	MONTO_12M	GASTO_TPD_OU_3M	0.7434863

Fuente: Elaboración propia

12.21 Anexo 21: Crecimiento de datos móviles y capacidad

Global Mobile Data - Traffic growth & forecast (terabytes per month)

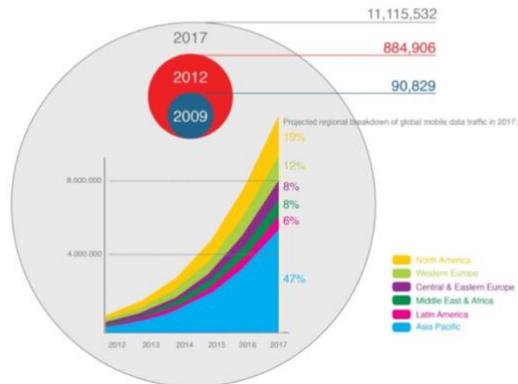


Figura 42: Crecimiento y proyección de los datos móviles globales.

Fuente:(Alliance, 2015)

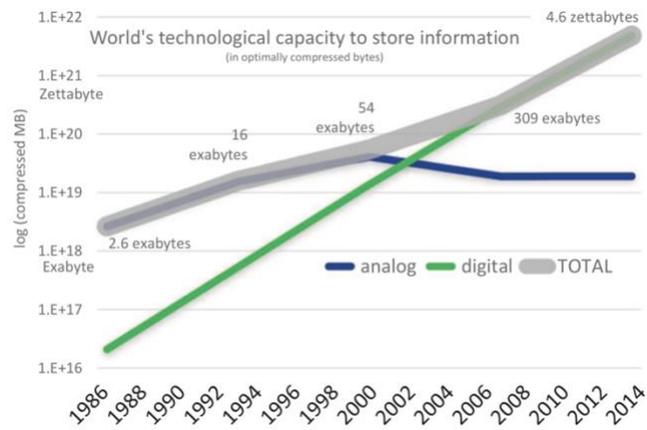


Figura 43: Capacidad de almacenamiento de información.

Fuente: (Hilbert, 2017)

12.22 Anexo 22: Resumen del piloto experimental diseñado

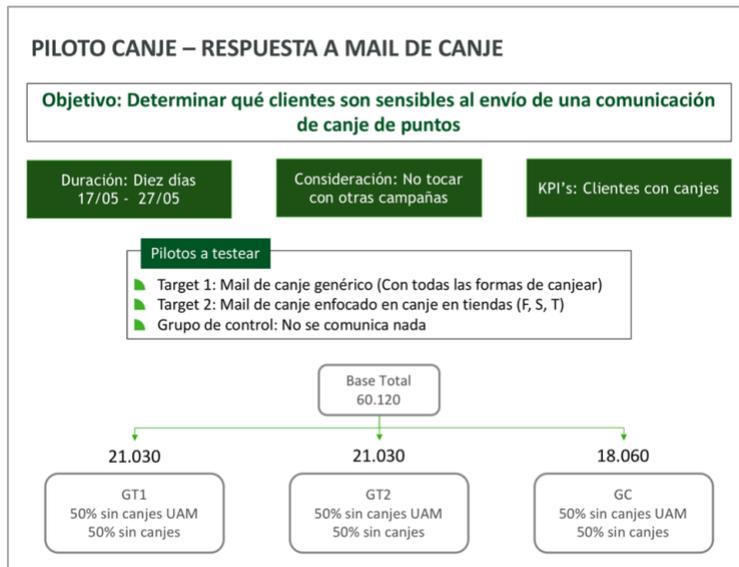


Ilustración 22: Resumen del diseño experimental.

Fuente: Elaboración propia

12.23 Anexo 23: Nodos más y menos propensos a ser influenciados

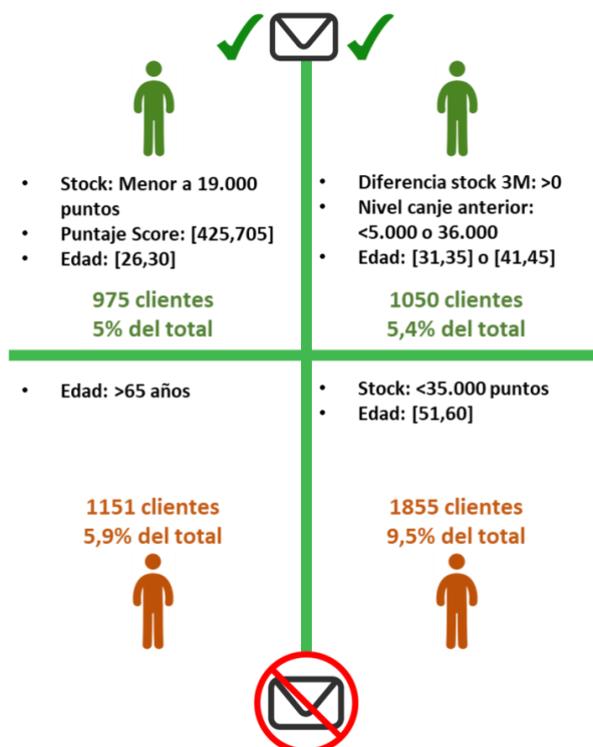


Ilustración 23: Infografía de nodos más y menos propensos a ser influenciados

Fuente: Elaboración propia.

12.24 Anexo 24: Distribución de canjes según género

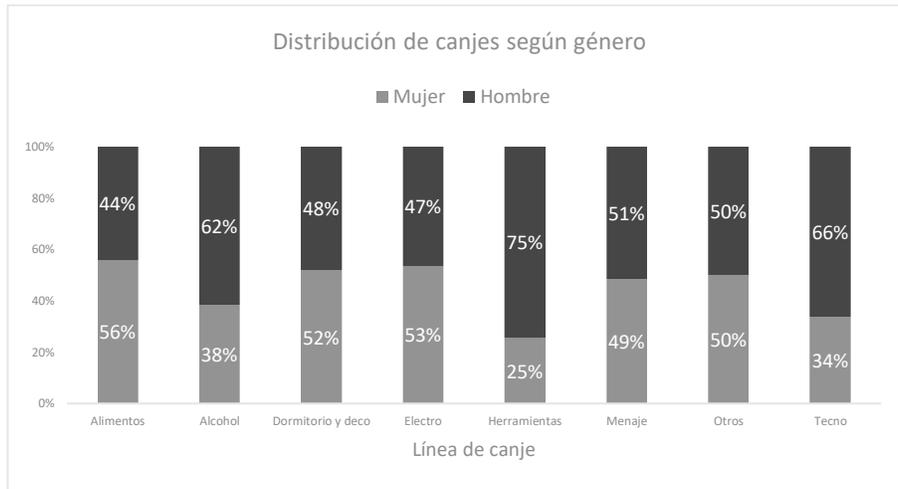


Figura 44: Distribución de los canjes por género.

Fuente: Elaboración propia con datos de la empresa relativos a los canjes de 2017

12.25 Anexo 25: Distribución de canjes según drivers de enganche

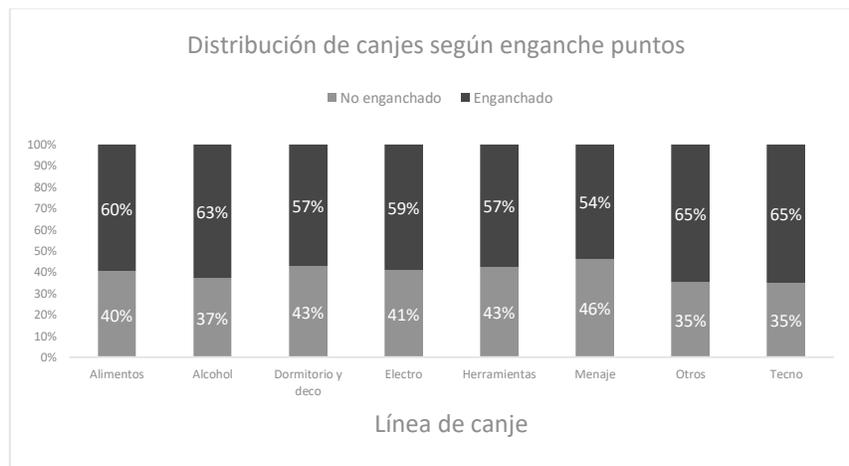


Figura 45: Distribución de los canjes por driver de enganche puntos.

Fuente: Elaboración propia con datos de la empresa relativos a los canjes de 2017

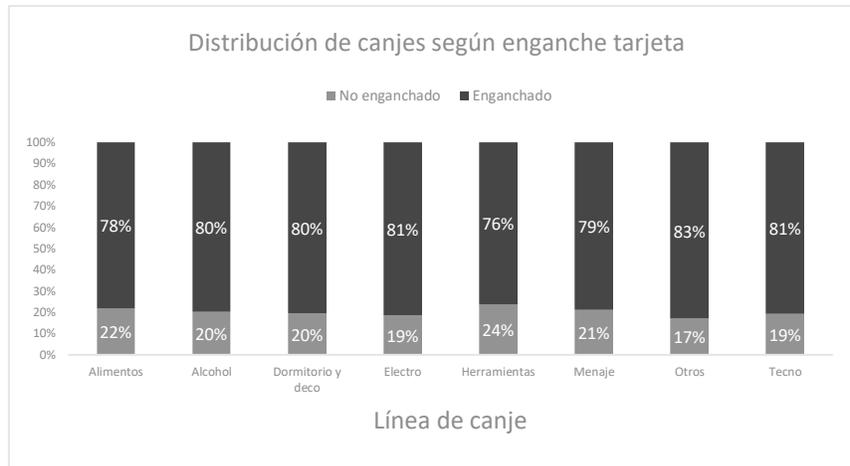


Figura 46: Distribución de los canjes por driver de enganche tarjeta.

Fuente: Elaboración propia con datos de la empresa relativos a los canjes de 2017

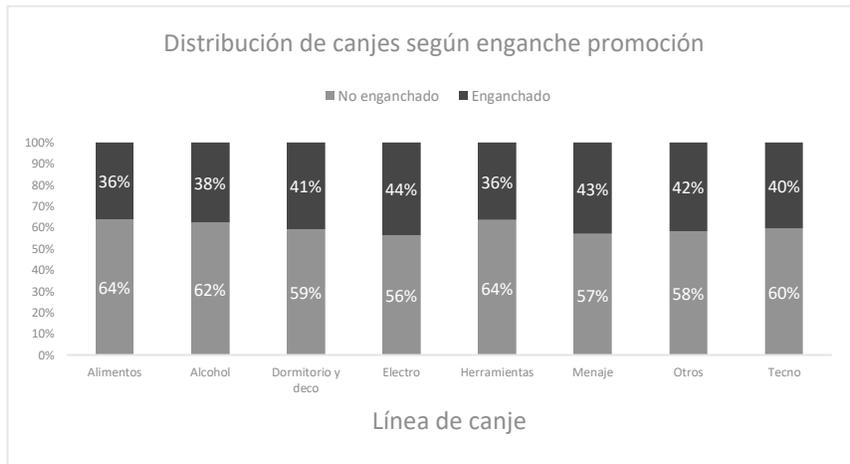


Figura 47: Distribución de los canjes por driver de enganche promoción.

Fuente: Elaboración propia con datos de la empresa relativos a los canjes de 2017

12.26 Anexo 26: Distribución de canjes según actividad en tiendas del grupo

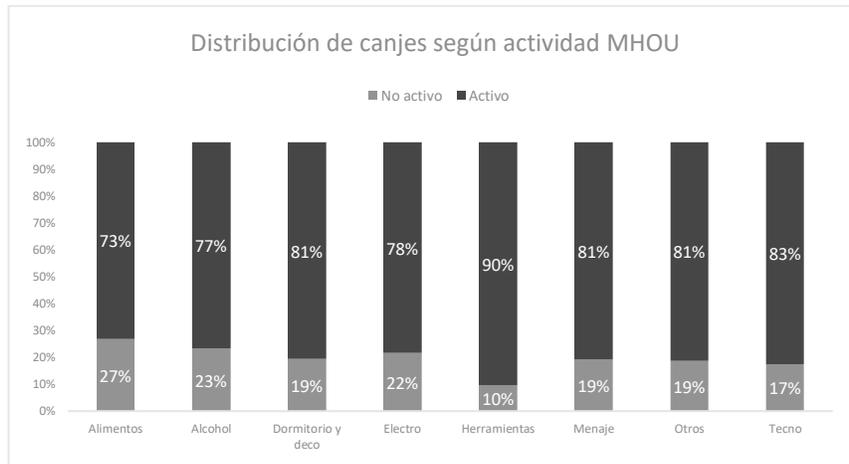


Figura 48: Distribución de los canjes por si es activo en la tienda de Mejoramiento del Hogar del grupo.

Fuente: Elaboración propia con datos de la empresa relativos a los canjes de 2017

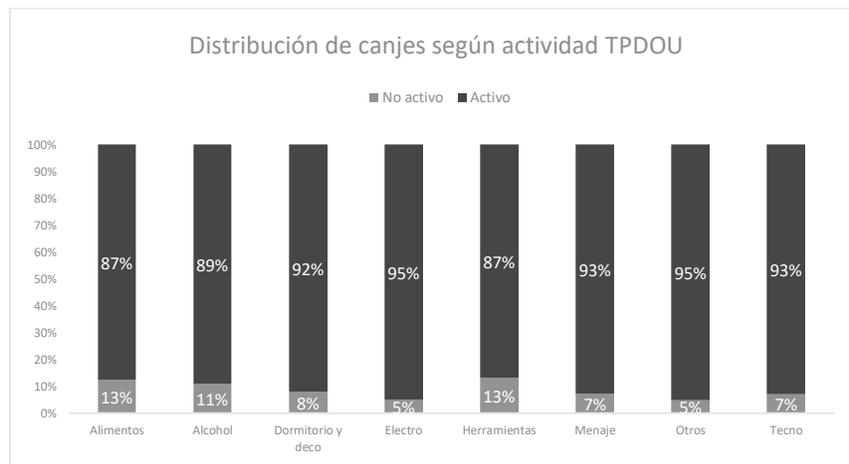


Figura 49: Distribución de los canjes por si es activo en la Tienda por Departamento del grupo.

Fuente: Elaboración propia con datos de la empresa relativos a los canjes de 2017

12.27 Anexo 27: Distribución de canje según ingreso web

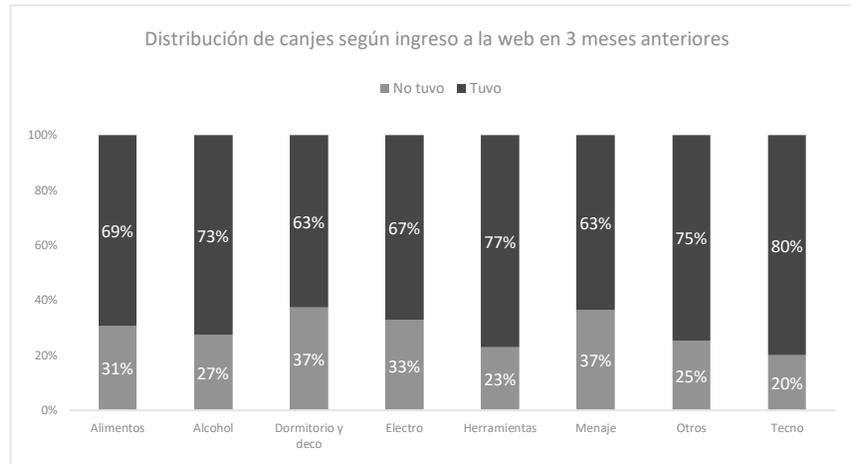


Figura 50: Distribución de los canjes por si tuvo ingreso web en los 3 meses antes del canje.

Fuente: Elaboración propia con datos de la empresa relativos a los canjes de 2017

12.28 Anexo 28: Distribución de edad y antigüedad de los clientes canjeadores

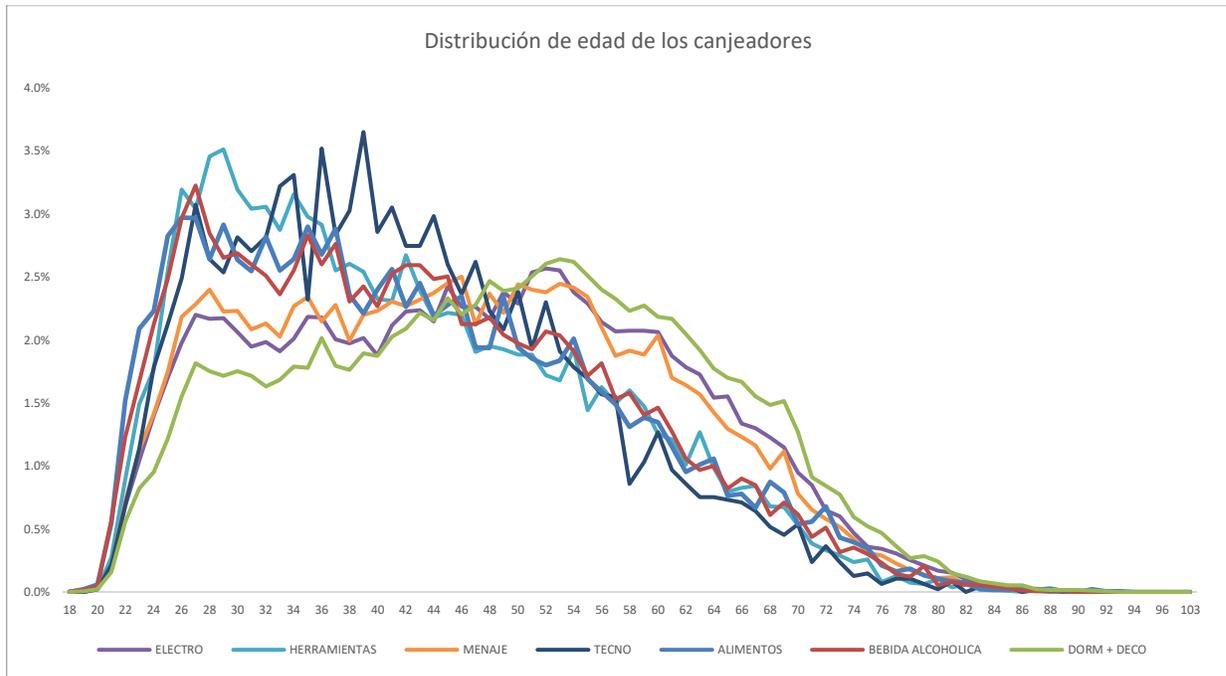


Figura 51: Distribución de la edad de los clientes.

Fuente: Elaboración propia con datos de la empresa relativos a los canjes de 2017

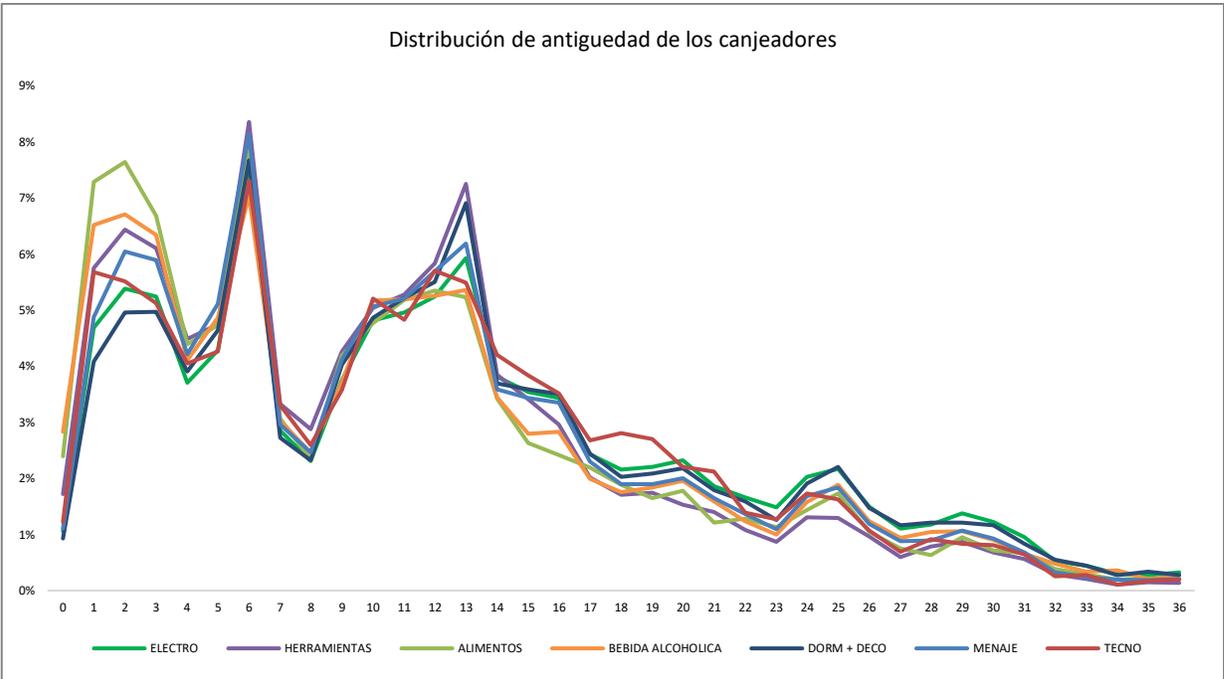


Figura 52: Distribución de la antigüedad de los clientes.

Fuente: Elaboración propia con datos de la empresa relativos a los canjes de 2017