



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

CONSTRUCCIÓN Y CARACTERIZACIÓN DE PERFILES DE CLIENTES EN BASE A
SU MOVILIDAD

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

CONSTANZA ISABEL RICHARDSON CORVALÁN

PROFESOR GUÍA:
MARCEL GOIC FIGUEROA

MIEMBROS DE LA COMISIÓN:
MARCEL GOIC FIGUEROA
SEBASTIÁN RÍOS PÉREZ
ÁLVARO OVANDO VÁSQUEZ

SANTIAGO DE CHILE
2014

Resumen

Este trabajo de título se enmarca en el desarrollo de un proyecto de movilidad en tiempo real de una empresa de telecomunicaciones en cooperación con el Centro de Investigación en Inteligencia de Negocios (CEINE).

La empresa posee los registros de telefonía móvil de sus clientes, a partir de los cuales es posible establecer con cierto margen de error la posición de los usuarios en el tiempo. Este conocimiento permite inferir distintas características de los clientes, desde la descripción del movimiento de las personas hasta la identificación de las actividades que desarrolla, en los lugares y tiempos en que lo hace. Esta información presenta gran potencial para establecer una comunicación efectiva con los clientes, que llegue a ellos en los lugares y momentos más apropiados.

Motivado por el alto potencial del conocimiento de la movilidad de los usuarios, este trabajo tiene por objetivo construir y caracterizar perfiles de clientes basándose en su movilidad. Para esto se analizó la información de ubicación temporal de los usuarios y se identificaron métricas que resumieron sus patrones de movilidad.

Una vez generada la información, se evaluó su utilidad para poder diferenciar entre clientes con distinto valor para la compañía, específicamente determinando su poder explicativo de la variable de ingresos promedio por cliente. De aquí se determinó que las métricas de movilidad obtenidas presentan una baja relación con el valor de los clientes, lo que no significa que no representen un aporte al proyecto.

Se aplicó un algoritmo de clusterización para encontrar conjuntos de clientes con características similares entre sí, obteniéndose 10 grupos diferentes, los que fueron perfilados tomando como base la movilidad de los individuos en ellos contenidos y enriqueciéndolo con información demográfica y transaccional.

A partir de los resultados obtenidos se concluye que la información de movilidad puede aportar conocimiento relevante para caracterizar a los clientes y dirigir ciertas comunicaciones a ellos. Sin embargo, quedan amplios espacios para buscar otras métricas y aplicaciones para explotar esta información.

Agradecimientos

En este momento en que culmina un proceso tan importante, intenso y feliz como es el haber estudiado esta carrera, no puedo dejar de agradecer a quienes formaron parte de él también. En primer lugar destaco a mi familia: Mamá, Papá y Cele, durante toda mi vida me han enseñado, aconsejado y apoyado, han sufrido con mis caídas y han disfrutado con mis logros, han confiado en mí y han dejado que tome mis propias decisiones. Estoy orgullosa y feliz de que sean mi familia y deseo hacerlos sentir igual durante el resto de la vida juntos.

Quiero agradecer también a Nicolás, mi pololo, mejor amigo y parte de mi familia. Pasamos toda esta etapa juntos y espero que pasemos muchas más de la misma forma, felices y riendo en los mejores momentos y apoyándonos cuando no eran tan buenos. Has sido fundamental no solo para mis estudios sino que en todo mi desarrollo y crecimiento como persona en estos más de 6 años. Te amo y te estoy eternamente agradecida por todo lo que haces por mí día a día.

Muy importantes han sido también todas las amistades que he hecho. Kol, nuestra amistad viene de mucho antes y sé que se mantendrá por siempre, nos hemos tenido paciencia y mucho cariño por más de doce años y eres ya parte de mi familia también. Pasita, eres mi persona, hemos permanecido juntas desde el día uno y lo seguiremos haciendo con toda seguridad. Nico, Dani M., Consu, Gilla, Naty, Dani A. y Mila las quiero a todas hasta el infinito y les agradezco por apoyarme siempre. Por último, la pandilla es el grupo de amigos que todo el mundo debiese tener en la universidad para aprender que no es solo estudio y pasarlo bien... les agradezco a todos ustedes que la conforman por eso.

Gracias al CEINE por su apoyo en este trabajo. Gracias a mis múltiples tutores: Ricardo, Camilo e Iván por ser tan abiertos con su conocimiento y pacientes para transmitirlos. Gracias especiales a Iván que además de mi tutor te convertiste en un amigo y en un motivo de risa permanente. Lo mismo para Gonzalo, que sufriste conmigo el día a día y entremedio nos hicimos amigos. Gracias a Lautaro y Raúl por su compañía y apoyo cada día.

Agradezco a mis profesores Marcel Goic y Sebastián Riós por su apoyo y ayuda y por responder siempre a mis mil y una dudas. Agradezco a Álvaro Ovando en representación de todo el equipo de la empresa que participó en este proyecto por su disposición y confianza.

Tabla de Contenido

1. Introducción	1
1.1. Antecedentes Generales	1
1.2. Descripción del Proyecto y Justificación	2
1.3. Objetivos	3
1.3.1. Objetivo General	3
1.3.2. Objetivos Específicos	3
1.4. Resultados Esperados	3
1.5. Alcance	3
1.6. Metodología	4
1.7. Estructura del Trabajo	4
2. Marco Conceptual	6
2.1. Movilidad	6
2.1.1. Trabajo Previo	6
2.2. Trabajo previo en proyecto de movilidad en tiempo real	9
2.2.1. Posiciones Relevantes de un Usuario	9
2.2.2. Caracterización Tempo-Espacial de la ciudad	10
2.3. Métodos de Clusterización	12
2.3.1. K-means	13
2.3.2. Evaluación	13
2.4. Regresión Lineal	14
2.4.1. Método de los Mínimos Cuadrados	14
2.4.2. Calidad Global del Modelo	15
3. Metodología	16
3.1. Definición del problema	16
3.1.1. Lugares relevantes y rutina diaria	16
3.1.2. Actividades Esporádicas	18
3.1.3. Métricas de Movimiento	18
3.1.4. Patrones Grupales de Movimiento	20
3.2. Análisis y Selección de Datos	21
3.2.1. Información Disponible	21
3.2.2. Análisis de Datos	21
3.2.3. Selección de Datos	24
3.3. Preprocesamiento de Datos	25
3.3.1. Tratamiento de Valores Fuera de Rango	25

3.3.2. Tratamiento de Valores Perdidos	27
3.4. Transformación de Variables	30
3.5. Análisis de Resultados	31
3.5.1. Identificación y Caracterización del Hogar y Trabajo	31
3.5.2. Actividades Esporádicas	37
3.5.3. Cálculo de Métricas de Movimiento	38
3.5.4. Identificación de Patrones de Movimientos Grupales	43
4. Perfilamiento	45
4.1. Identificación de Perfiles	45
4.2. Perfiles de clientes	45
5. Evaluación	52
6. Conclusiones	56
6.1. Conclusiones del Trabajo	56
6.2. Trabajo Futuro	57
Bibliografía	63
A. Análisis estadístico de variables	64
B. Análisis de correlaciones	67
C. Tratamiento de outliers	69
D. Análisis poblacional	71
E. Tratamiento de outliers en variables generadas	73
F. Resultados regresiones lineales	74
G. Medias de perfiles	79

Índice de tablas

2.1. Extracto tabla residencial	11
2.2. Extracto tabla comercio y servicios	11
2.3. Ejemplo tabla caracterización de uso de suelo	12
2.4. Correspondencias entre uso de suelo y de tiempo	12
2.5. Algoritmo K-means	13
3.1. Información disponible por categoría	22
3.2. Descripción información registros telefónicos	22
3.3. Estadísticos tras tratamientos de valores perdidos y fuera de rango	31
3.4. Transformación GSE	31
3.5. Distribución hogar en sectores	34
3.6. Distribución trabajo en sectores	34
3.7. Proporción de población por sector	35
3.8. Proporción de población por sector	35
3.9. Ejemplo caracterización para hogar y trabajo	36
3.10. Estadísticos caracterización actividades hogar y trabajo	36
3.11. Horarios actividades esporádicas	37
3.12. Estadísticos actividades esporádicas	37
3.13. Estadísticos métricas de movimiento	42
3.14. Medias clusters de movimientos similares	43
3.15. Tamaños relativos clusters de movimientos similares	44
4.1. Tamaño relativo perfiles	46
5.1. Grupos de variables	52
5.2. Resultados regresión lineal métricas de movimiento	53
5.3. Resultados regresión lineal variables demográficas	53
5.4. Resultados regresión lineal sectores hogar	54
5.5. Resultados 15 mejores regresiones lineales - ARPU	54
5.6. Coeficientes regresión lineal definitiva	55
A.1. Estadísticos Edad	64
A.2. Estadísticos Género	64
A.3. Estadísticos GSE	64
A.4. Estadísticos antigüedad cliente	65
A.5. Estadísticos tráfico de voz y mensajería	65
A.6. Estadísticos tráfico de datos	65

A.7. Estadísticos equipo	65
A.8. Proporción usuarios con equipos touch y smartphone	65
A.9. Estadísticos descriptivos ARPU medido en pesos	65
A.10. Estadísticos Mora	66
A.11. Estadísticos registros de llamadas	66
A.12. Estadísticos descriptivos Posiciones Relevantes independientes del tiempo . .	66
A.13. Estadísticos descriptivos Posiciones Relevantes cada una hora	66
B.1. Resultados correlaciones 1	67
B.2. Resultados correlaciones 2	67
B.3. Resultados correlaciones 3	68
C.1. Análisis desviaciones tráfico voz	69
C.2. Análisis desviaciones tráfico de datos	69
C.3. Análisis desviaciones tráfico mensajería	69
C.4. Análisis desviaciones ARPU	70
C.5. Análisis desviaciones número de recambios	70
C.6. Análisis desviaciones antigüedad equipo	70
D.1. Asignación Comuna a Sector - Parte 1	71
D.2. Asignación Comuna a Sector - Parte 2	72
E.1. Análisis desviaciones posiciones	73
E.2. Análisis desviaciones distancia recorrida	73
E.3. Análisis desviaciones radio de giro	73
F.1. Resultados regresión lineal - ARPU. Parte 1	74
F.2. Resultados regresión lineal - ARPU. Parte 2	75
F.3. Resultados regresión lineal - ARPU. Parte 3	76
F.4. Resultados regresión lineal - ARPU. Parte 4	77
F.5. Resultados regresiones lineales - log(ARPU)	77
F.6. Resultados regresión lineal - Log(ARPU). Edad transformada.	78
F.7. Resultados regresiones lineales - log(ARPU). Edad y métricas transformadas.	78
G.1. Medias Perfiles. Parte 1.	79
G.2. Medias perfiles. Parte 2.	80

Índice de figuras

2.1. Ejemplo posiciones relevantes	9
2.2. Mapa discretizado de la Región Metropolitana. Fuente: [35]	9
2.3. Ejemplo posiciones relevantes en ventanas de tiempo	10
2.4. Tópicos obtenidos - Distribución en una semana completa. Fuente: [28]	11
3.1. Ritmo diario de las personas. Fuente: [11]	17
3.2. División RM en 6 sectores	17
3.3. Distancia recorrida para distintos usuarios	19
3.4. Diferencia de número de posiciones entre usuarios	19
3.5. Diferencia de radio de giro entre usuarios	20
3.6. Histograma de llamadas por usuario	23
3.7. Llamadas agregadas por bloque horario en una semana	23
3.8. Histogramas edad	25
3.9. Histogramas tráfico de voz	26
3.10. Histogramas tráfico de mensajería	28
3.11. Histogramas tráfico de datos	29
3.12. Histogramas ARPU	29
3.13. Histograma recambios	30
3.14. Histogramas antigüedad equipo	30
3.15. Llamadas en horario y fuera de horario para el Hogar	32
3.16. Llamadas en horario y fuera de horario para el Trabajo	32
3.17. Histogramas distancia recorrida	39
3.18. Histogramas número posiciones	40
3.19. Histogramas radio giro	41
3.20. Histogramas Grado de Retorno	42
3.21. Índices de Dunn y Davies Bouldin para patrones grupales	43
4.1. Índices DB y Dunn para perfiles	45
A.1. Histograma de número de llamadas por posiciones no dependientes del tiempo	66
A.2. Histograma de número de llamadas por posiciones dependientes del tiempo .	66

Capítulo 1

Introducción

1.1. Antecedentes Generales

Los perfiles y segmentaciones de clientes hasta el día de hoy se han basado principalmente en información demográfica y/o transaccional disponible (por ejemplo [2], [44]), pero conforme avanza el tiempo, nuevos tipos de datos han aparecido, permitiendo enriquecer el conocimiento que se tiene de los clientes.

Actualmente es posible conocer la ubicación de los clientes a lo largo del tiempo, o por lo menos una aproximación de ella, ya sea a través de dispositivos GPS o a partir del registro de llamadas de los usuarios de telefonía móvil. Con este tipo de información se pueden conocer los movimientos usuales de una persona e inferir respecto a las actividades desarrolladas por ésta y su comportamiento.

Conocer lo que hace un cliente y el momento en que esto sucede abre amplias posibilidades de negocios, pues la ubicación juega un rol fundamental en determinar el tipo y la naturaleza de la actividad humana, proveyendo información que puede contribuir a determinar las necesidades de información o sus elecciones de productos o servicios [32].

La información de localización está comenzando a ser explotada por el *Mobile Marketing*, que corresponde a la comunicación, promoción y oferta entre una empresa y sus clientes utilizando un medio móvil [39]. La rápida proliferación de aparatos móviles ha derivado en un amplio campo de investigación en esta área[42].

El *Mobile Marketing* está aprovechando el conocimiento referente a la movilidad de las personas para entregar la información o promoción adecuada en el momento oportuno. En *Mobile Advertising*, una de las ramas del ya mencionado *Mobile Marketing*, se espera un amplio crecimiento en los ingresos globales, pasando de 2,8 billones de dólares en 2010 a 10,3 billones en 2015 [41], siendo las compañías de telecomunicaciones quienes principalmente han explotado esta forma de publicidad.

1.2. Descripción del Proyecto y Justificación

Este trabajo se enmarca en el desarrollo de un proyecto de movilidad en tiempo real de una compañía de telecomunicaciones, el cual busca ofrecer un canal para que las empresas envíen contenido a potenciales clientes de acuerdo a su información georeferenciada. De este modo la compañía lograría levantar una nueva área de negocios, aprovechando la información que tiene hoy en día disponible.

El proyecto descrito surge de la cooperación entre el Centro de Investigación en Inteligencia de Negocios (CEINE) y la compañía. De él participan otros dos trabajos de título: [30] y [35]. En el primero se desarrolló el plan de negocios, evaluando el mercado y la factibilidad y definiendo las líneas estratégicas a seguir. El segundo desarrolla la construcción de modelos para la construcción de patrones de movilidad y es desde donde proviene gran parte de la información que se utiliza en este trabajo. Además existe una arista de investigación desde el punto de vista de la arquitectura de *Hardware* y *Software*, desarrollada por miembros del equipo del CEINE.

El proyecto parte del reconocimiento de la existencia de información proveniente de los registros de telefonía móvil de la compañía, a partir de los cuales es posible obtener la ubicación de los usuarios en el tiempo. Se desea utilizar dicha información para hacer inferencia respecto a la movilidad de los clientes y con ella enriquecer lo que se obtiene a partir de la información demográfica y transaccional, generando perfiles de usuarios que incorporen todas estas dimensiones.

La movilidad de los clientes aporta información valiosa para el proyecto de la empresa, pues a través de ella se puede hacer inferencia sobre cuáles son las actividades que las personas desarrollan, el momento en que lo hacen, los lugares que visitan, la regularidad con que lo hacen, las zonas que abarcan en sus desplazamientos y mucho más. A partir de esta información se puede, por ejemplo, seleccionar las personas que se mueven habitualmente cerca de determinado comercio para que éste pueda enviar promociones o publicidad a dichas personas, ofrecer información a clientes que hayan participado de algún tipo de evento masivo en particular sobre otros relacionados y a la vez reportar sobre el número y característica de los asistentes a los organizadores de los mismos, también se puede determinar aquellos usuarios que con mayor probabilidad atenderían a cierta promoción en base a la información de lugares de características similares visitados por los usuarios, entre muchas otras posibilidades.

Se desea inferir lo que las personas realizan en sus puntos de actividades diarias, además de identificar formas de movimiento tanto de manera individual como grupal, entre otras cosas. La definición de la información a generar y el porqué de la misma se detallará en próximas secciones.

El proyecto y la información generada para él permiten la creación de otros negocios, como por ejemplo la venta de estudios demográficos de clientes a otras entidades interesadas en grupos de personas de acuerdo a los lugares que visitan. Por otro lado, no solo tienen aplicaciones comerciales, sino que pueden incluso colaborar en temas de planificación urbana, desarrollo de nuevos recorridos de transporte público, creación de nuevas líneas de tren subterráneo, nuevos tipos de análisis demográficos y otros, generando un mayor conocimiento

sobre la ciudad y sus habitantes tanto para autoridades como comercios y ciudadanos.

1.3. Objetivos

1.3.1. Objetivo General

Identificar y caracterizar perfiles de clientes de telefonía móvil basados en variables que expliquen su movilidad.

1.3.2. Objetivos Específicos

1. Definir y construir métricas de movilidad de los usuarios para incorporarlas en el análisis.
2. Construir y caracterizar perfiles de clientes en base a las métricas de movilidad generadas, enriquecidos con información demográfica y transaccional.
3. Evaluar la relevancia de las métricas generadas para discriminar entre los clientes de acuerdo al valor que representan para la compañía.

1.4. Resultados Esperados

Se espera obtener como resultados lo siguiente:

1. Métricas individuales de movilidad de los usuarios para aportar al entendimiento y descripción de sus movimientos.
2. Un conjunto de variables que ayuden a discriminar entre clientes de distinto valor para la compañía.
3. Perfiles de clientes construidos a partir de la información de movilidad generada y la demográfica y transaccional previamente disponible.

1.5. Alcance

Se utilizarán datos correspondientes a clientes de la Región Metropolitana con planes de voz y post-pago pertenecientes solo a la categoría personas y que se hayan mantenido activos durante el periodo de análisis. Con esta reducción el número potencial de clientes se aproxima a 670.000, de los cuales se trabajará con una muestra cercana a 350.000 clientes.

No se realizarán recomendaciones formales para el uso de la información generada, ésta será entregada a la empresa para que en base a su conocimiento experto, tomen las decisiones que les parezcan pertinentes.

1.6. Metodología

La metodología a aplicar será basada en el proceso *Knowledge Discovery in Data Bases* (KDD) [14]. Para el desarrollo de este trabajo, la metodología a seguir es la siguiente:

1. Identificación del problema y objetivos

Se identifica el objetivo de este trabajo en el proyecto general y se definen los aspectos de la movilidad de los usuarios relevantes para medir e incorporar en el análisis.

2. Selección

Se escogen las variables a utilizar en el desarrollo metodológico completo. Se debe crear un conjunto de datos coherentes, alineados con los objetivos definidos y de calidad apropiada para su tratamiento.

3. Preprocesamiento

Se realiza el análisis y limpieza de datos, lo que incluye el tratamiento de valores faltantes y fuera de rango.

4. Transformación

Se generan las nuevas variables que describen las características deseadas para el análisis. Se utilizan distintas técnicas que van desde la transformación de tipos de variables a la aplicación de funciones.

5. Procesamiento

Se aplica análisis de datos y se desarrollan procedimientos para construir la información de movilidad deseada y de este modo cumplir con los objetivos definidos.

6. Evaluación de la información generada

Se evalúa la relevancia de la información de movilidad generada. Se considera relevante aquella que logre diferenciar entre usuarios según su valor para la compañía.

7. Perfilamiento de clientes

Se aplica un algoritmo de clusterización sobre las variables generadas para obtener los perfiles y se realiza una caracterización de los mismos.

1.7. Estructura del Trabajo

En el siguiente capítulo se explican los conceptos necesarios para el entendimiento del presente trabajo de título, entre ellos se muestra el trabajo realizado previamente tanto en movilidad de forma general, como en el proyecto de la compañía. Luego se expone una breve reseña de las herramientas utilizadas en el desarrollo del trabajo.

En el Capítulo 3 se explica el desarrollo metodológico del trabajo, partiendo por una descripción de los datos y su posterior análisis para la selección, luego se expone el preprocesamiento de los mismos, incluyendo el tratamiento de valores faltantes y fuera de rango y las transformaciones que se aplicaron a las distintas variables seleccionadas. Posteriormente se desarrolla el procesamiento de la información que culmina con la construcción de las variables

de movilidad deseadas. Estas variables son evaluadas posteriormente en el Capítulo 4 para determinar su capacidad de discriminar entre clientes.

En el Capítulo 5 se construyen los perfiles de usuarios y se realiza la caracterización de los mismos en base a su movilidad e información demográfica y transaccional y por último , en el Capítulo 6 se presentan las conclusiones obtenidas a partir del trabajo realizado junto a algunas consideraciones y recomendaciones para el trabajo futuro.

Capítulo 2

Marco Conceptual

2.1. Movilidad

El objetivo de esta sección es exponer trabajos previos relacionados al trabajo de título desarrollado. Adicionalmente, se establecen definiciones básicas para entender el desarrollo metodológico expuesto en el Capítulo 3.

2.1.1. Trabajo Previo

Se encontró en la literatura una base importante para el desarrollo de este trabajo. Se pueden ordenar las investigaciones de acuerdo a sus principales conclusiones y como éstas aportan a la presente memoria.

Lugares Relevantes y Rutina Diaria

Estudios plantean que tanto la existencia de ciertos lugares más relevantes para un usuario, como la rutina diaria que éstos siguen son determinantes para explicar el movimiento de las personas.

Entre los trabajos que llegan a la conclusión anterior se encuentra [16], donde se busca una distribución que refleje el movimiento de las personas, utilizando tanto registros telefónicos como la posición exacta de los usuarios captada cada dos horas a través de sistemas GPS. En sus resultados se muestra que la trayectoria humana presenta una gran regularidad espacial y temporal, con una marcada tendencia a dirigirse a ciertos lugares más relevantes.

Otro trabajo que resalta la importancia de la rutina diaria es [8], donde con información proveniente de registros telefónicos y de redes sociales basadas en posición, se explican los movimientos individuales de las personas, mostrando que éste puede dividirse en dos componentes: un 70 % de los movimientos provienen de la rutina diaria de las personas -donde

citan como ejemplo el traslado entre hogar y trabajo-, mientras que el otro 30 % podría ser explicado a través de las relaciones sociales obtenidas a partir de la red de llamados telefónicos. A partir de estos resultados el autor es capaz de obtener un modelo predictivo de las locaciones futuras de los usuarios.

Un modelo de radiación es construido en [40], el que es comparado a un modelo gravitacional -basado en la expresión matemática de la ley de gravitación universal de Newton-, aplicado en este caso para la atracción de personas a diferentes lugares. El modelo de radiación se basa en el transporte diario entre hogar y trabajo para establecer cómo una persona escoge este último. Para esto asumen la tasa de disponibilidad de empleo en base a la densidad poblacional de los lugares, siendo esta la única información necesaria para la aplicación del modelo. En contraste, el modelo gravitacional debe ser calibrado a través de datos de tráfico, los que pueden no estar disponible en todas las locaciones. El resultado obtenido es que el modelo de radiación refleja mucho mejor lo ocurrido en la realidad que el modelo gravitacional. Cabe destacar que en esta investigación los autores utilizan el hogar y el trabajo -lugares relevantes para el usuario- como las principales actividades que motivan el movimiento de las personas.

De estas investigaciones se concluye que es de gran importancia para el proyecto la identificación de los señalados lugares relevantes, pues se presentan como los principales promotores del movimiento habitual de los usuarios, los que representan una importante proporción del mismo.

Construcción de perfiles basados en movilidad

Consecuentemente con los resultados de los trabajos descritos anteriormente, en [7] se modelan perfiles de movilidad como trazas de los lugares frecuentemente visitados por los usuarios - regiones de interés- a través del uso de información obtenida con dispositivos GPS. A partir de esto se mide la similitud entre los distintos perfiles individuales con el objetivo de aportar en la generación de recomendaciones a través de redes sociales basadas en ubicación.

Otro trabajo dedicado a encontrar perfiles en base a la movilidad de las personas es [22], donde a través de una encuesta de viajes se obtiene información referente a los lugares que visitan los individuos y la(s) actividad(es) que ahí realizan. Dichas actividades se categorizan dentro de 9 grupos entre los que se cuentan hogar, trabajo, entretenimiento, encargos y otros. Posteriormente, a través del uso de k-medias como método de clusterización, se obtienen grupos de los cuales se rescatan sus principales actividades para caracterizarlos en base a las mismas. La combinación de esta información con la de tipo demográfico es lo que termina por definir los perfiles formados.

Si bien estos trabajos se desarrollaron con el mismo objetivo que el presente, se diferencian en el tipo de información y cantidad de datos disponibles. Cabe destacar que en el caso de [7] la finalidad de desarrollar perfiles basados en movilidad es la misma que en este trabajo de título.

Métricas de movilidad

En la revisión de la literatura aparecen conceptos y métricas que permiten caracterizar la movilidad de los usuarios. De estos se ha identificado el siguiente conjunto a utilizarse en el desarrollo de este trabajo, ya sea directamente o como base para otros conceptos.

- **Regiones de Interés (ROIs¹):** Utilizadas en distintos trabajos ([7][15]) y obtenidas a través de la fusión de puntos espaciales densos en regiones, simplificando de esta forma el problema al pasar de grandes números de puntos representados por pares (latitud,longitud) a una menor cantidad de sectores de interés.
- **Radio de Giro:** Utilizado en [16], se define como la distancia recorrida característica de cada usuario. La fórmula asociada a esta medida de movimiento es:

$$r_g^a(t) = \sqrt{\frac{1}{n_c^a(t)} \sum_{i=1}^{n_c^a} (r_i^a - r_{cm}^a)^2}$$

, donde $r_g^a(t)$ es el radio de giro del usuario a en el tiempo t , n_c^a es el número de posiciones distintas del usuario, r_i^a es la posición i y r_{cm}^a es el centro de masa. Es decir, puede interpretarse como una especie de desviación típica respecto al centro de masa de las posiciones del usuario.

- **Probabilidad de Retorno:** Es definida como la probabilidad de que un determinado usuario vuelva a la posición en la que fue observado por primera vez después de t horas [16]. A través de esta medida se puede establecer si es que el usuario repite periódicamente sus conductas de movimiento o no.

Otros Trabajos

Estos trabajos, si bien no tienen una aplicación directa en este trabajo de título, presentan líneas de investigación que pueden ser incorporadas al proyecto de movilidad en el futuro.

Usando los registros telefónicos de 6 millones de usuarios, en [43] se encuentra una correlación entre los movimientos de las personas y la cercanía con el resto en su red social de llamados, llegando de esta forma a desarrollar un método de predicción de nuevas conexiones a formarse en base a los movimientos y la red social.

A partir de información de localización obtenida a través de *Radio-frequency identification* (RFID), en [19] se contrasta la trayectoria óptima (de acuerdo al Problema del Vendedor Viajero) con la de los clientes dentro de un supermercado, teniendo en cuenta la canasta de compras del mencionado cliente. Así se hace inferencia sobre la forma en que las personas escogen su ruta dentro de un lugar, midiendo la desviación del comportamiento del cliente respecto al óptimo según desviación causada por orden en el recorrido y desviación causada por distancia recorrida.

¹Por sus siglas en inglés provenientes de *Regions of Interest*

2.2. Trabajo previo en proyecto de movilidad en tiempo real

A continuación se mencionará parte del trabajo previo realizado por los distintos participantes del proyecto y que fue utilizado en la presente memoria.

2.2.1. Posiciones Relevantes de un Usuario

En [35] se desarrolló la identificación de Posiciones Relevantes² para los usuarios de la compañía de telecomunicaciones a partir de sus registros de llamadas. Para esto se realizó un *benchmark* con algoritmos basados en densidad, del cual el autor concluye que TI-DBSCAN tiene mejor performance general que algoritmos como DBSCAN u Optics.

Aplicando TI-DBSCAN se obtuvieron *clusters* que representan los lugares en que más frecuentemente se realizan o reciben llamadas para un usuario en específico y que pretende reflejar las zonas de mayor actividad de una persona. Los *clusters* están representados por un par (latitud, longitud) más el radio de giro del mismo, el que fue calculado a partir de todos los registros que forman el cluster o Posición Relevante.

En la Figura 2.2.1 se muestra un ejemplo de Posiciones Relevantes para un usuario cualquiera. En ella se observa que a partir de la *clusterización* de la ubicación de las llamadas con el algoritmo ya mencionado, se obtienen 2 áreas identificadas por la localización del centroide del *cluster* y el radio de giro calculado sobre las llamadas que lo integran. A estas áreas es que se nombra como Posiciones Relevantes, pues dada su alta actividad se presume son relevantes para el desarrollo de las actividades cotidianas del usuario.

En el mismo trabajo se realizó una discretización de la región a través del diagrama de Voronoi, donde cada celda corresponde a una antena. El mapa discretizado se puede ver en la Figura 2.2. Con esto, se cuenta también con la información de en qué celda del diagrama se encuentra ubicada cada Posición Relevante, ayudando a la interpretación de éstas y disminuyendo la intensidad de cálculos que se deberán hacer en adelante.

El trabajo ya descrito utilizó dos enfoques distintos para encontrar las regiones en que había mayor actividad:

- **Independientes del tiempo**

Se consideran todos los registros que figuran para el usuario, independiente del momento en que las llamadas fueron realizadas. El objetivo aquí es obtener aquellas posiciones más recurrentes de forma general y no para intervalos de tiempo específicos.

- **En ventanas de tiempo**

Se agrupan las llamadas en intervalos de una hora. Se utiliza el supuesto de que el comportamiento de los usuarios tiene una periodicidad de una semana, es decir, en

²Posiciones relevantes es el equivalente a las Regiones de Interés definidas en la sección 2.1.2, pero se decidió conservar el nombre utilizado por quien las generó para este proyecto.



Figura 2.1: Ejemplo posiciones relevantes

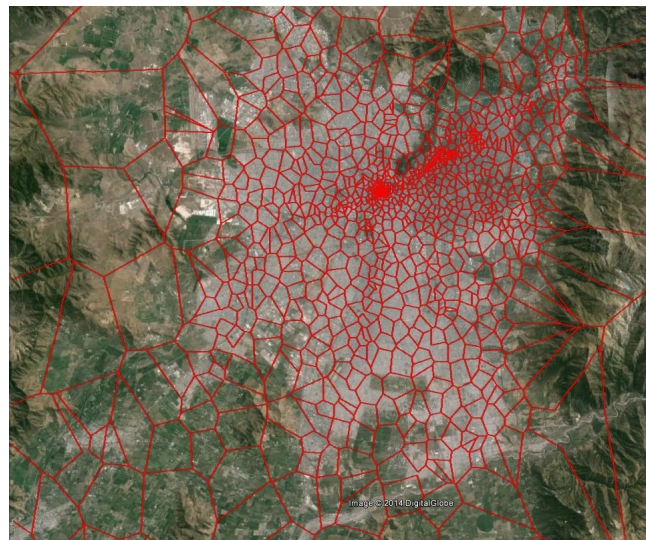


Figura 2.2: Mapa discretizado de la Región Metropolitana. Fuente: [35]

semanas distintas en un mismo día el usuario se comporta de forma similar, mientras que dentro de una semana, diferentes días no son necesariamente similares en cuanto a comportamiento. De esta forma se puede ubicar al usuario en sus posiciones habituales en el tiempo.

En la Figura 2.3 se muestran las posiciones de las llamadas de un usuario en el período de análisis en 6 ventanas de tiempo de un día específico de la semana. Se observa que en cada ventana se forman distintas Posiciones Relevantes de acuerdo a los registros que se tiene. En el caso del bloque horario $t = 4$ no se formaron *clusters* debido a la falta de registros, mientras que en la ventana $t = 5$ se forman 2 Posiciones Relevantes distintas. En este último caso, se asigna además a cada grupo de llamadas una probabilidad correspondiente a la proporción de llamadas que integra cada uno sobre el total de

registros que forman ambos.

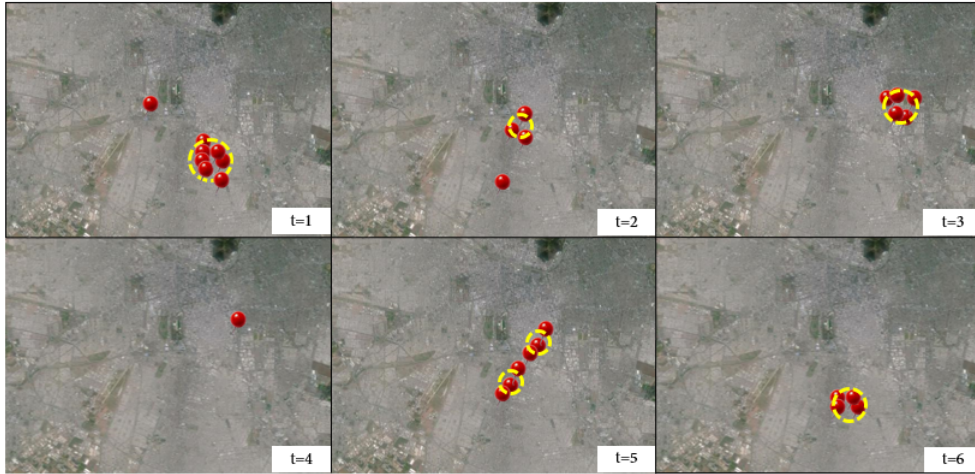


Figura 2.3: Ejemplo posiciones relevantes en ventanas de tiempo

Las pruebas realizadas concluyeron que el error de las posiciones encontradas, considerando la diferencia entre el centroide del cluster y la posición en que la persona desarrolla efectivamente sus actividades, corresponde a 400 metros en promedio. Esto es coherente con el hecho de que la distancia promedio entre antenas es de alrededor de 800 metros.

2.2.2. Caracterización Tempo-Espacial de la ciudad

Una caracterización temporal de las antenas -o celdas de Voronoi- de la Región Metropolitana es realizada en [28]. Utilizando Latent Dirichlet Allocation (LDA) [6] se obtuvieron 4 tópicos característicos de la ciudad. Para entender esto se puede revisar también [34], donde se aplica el algoritmo de manera similar, pero a redes sociales.

Los tópicos obtenidos fueron nombrados por los autores como Residencial, Trabajo, Comercio y Entretenimiento y Hora Punta. Éstos se muestran en la Figura 2.4.

A través de una organización que se dedica al desarrollo de mapas digitales para la aplicación en negocios, se obtuvo una base de datos cuyo contenido son todos los edificios o construcciones según su uso de los que tenían registro hasta ese momento, equivalentes a más de 6,27 millones de registros. Esta base está dividida en dos, una con construcciones de tipo residencial y otra con las de tipo comercial y de servicios.

La tabla residencial caracteriza cada construcción a través de un ID, su dirección, información de localización (región, comuna, calle, número, latitud y longitud), el tipo de edificación (casa, edificio o condominio) y el número de pisos. En caso de edificios y condominios, para cada residencia dentro de éstos existe un registro con diferente ID, pero con las mismas características. Esta base incluye además, locales comerciales ubicados en los edificios de tipo residencial y son identificados de la misma forma que el resto, con la diferencia de que su tipo de edificación corresponde a local comercial. Una muestra de esta base se presenta en la Tabla 2.1.



Figura 2.4: Tópicos obtenidos - Distribución en una semana completa. Fuente: [28]

Tabla 2.1: Extracto tabla residencial

Id_direccion	Region	Comuna	Calle	N°	Latitud	Longitud	Tipo	Pisos
629866	RM	San Miguel	Pedro Mira	530	-33,499	-70,640	Casa	1
863476	RM	Est. Central	Sta. Teresa	832	-33,462	-70,692	Casa	1
306246	RM	Providencia	El comendaros	1957	-33,419	-70,617	Casa	1
157079	RM	Providencia	Av. Italia	1350	-33,447	-70,625	Casa	1
435694	RM	La Florida	La higerilla	8216	-33,531	-70,601	Casa	1
809886	RM	La Florida	Salar de Tacolte	10294	-33,552	-70,604	Casa	1

La tabla de tipo comercial y de servicios caracteriza cada una de sus construcciones a través de un ID de servicio, un ID de dirección, la misma información de localización que en el caso anterior, ID de una categoría definida por la empresa y una breve descripción del rubro de la construcción en base a la que se hace la categorización. Un extracto de esta base se muestra en la Tabla 2.2.

Tabla 2.2: Extracto tabla comercio y servicios

Id_servicio	Id_direccion	Region	Comuna	Calle	N°	Latitud	Longitud	Categoria	Descripcion
786208	621045	RM	Maipu	Pdte de Gaulle	47	-33,484	-70,748	141	Panaderia
1191622	301966	RM	Pudahuel	El Anillo	1281	-33,434	-70,748	36	Iglesia
993567	816913	RM	Pudahuel	San Daniel	1203	-33,432	-70,747	136	Botilleria
646208	3662036	RM	Pudahuel	Jorge Pratt	473	-33,456	-70,747	142	Paradero
970527	5023677	RM	Renca	Vicuna Mackenna	0	-33,397	-70,747	123	Gimnasio
687667	2504393	RM	Renca	Compania de Tacna	1649	-33,399	-70,747	137	Almacen

Utilizando la información de los edificios de la ciudad, se categorizaron las distintas construcciones en 6 tipos diferentes:

- Residencial
- Trabajo Oficina
- Trabajo Industrial
- Educación
- Entretenimiento
- Comercio

Una vez establecido a qué categoría pertenecía cada uno de los edificios de la RM, se procedió a localizarlos dentro de la región de Voronoi correspondiente según su posición. Luego se calculó una proporción de cada uno de los grupos dentro de cada región para obtener así una caracterización del uso de suelo de cada celda de Voronoi. Un ejemplo de los resultados de este proceso se encuentra en la Tabla 2.3.

Combinando la información de caracterización de suelo y la de tiempo de las regiones de Voronoi, [35] construyó una caracterización tempo-espacial de las mismas. El proceso consistió en que por cada región de Voronoi, por cada bloque horario se multiplicó la probabilidad de pertenecer a cierto tópico por su uso de suelo relacionado y luego se normalizó. Las correspondencias utilizadas fueron las presentadas en la Tabla 2.4.

Tabla 2.3: Ejemplo tabla caracterización de uso de suelo

Zona	Actividad	Proporción
1254	residencial	0,6
1254	comercial	0,1
1254	entretenimiento	0,1
1254	educacion	0
1254	oficina	0,2
1254	industrial	0

Tabla 2.4: Correspondencias entre uso de suelo y de tiempo

Tópico tiempo	Categoría suelo
Comercio-Entretenimiento	Comercio
Comercio-Entretenimiento	Entretenimiento
Trabajo	Oficina
Trabajo	Industrial
Trabajo	Educación
Residencial	Residencial

2.3. Métodos de Clusterización

Los métodos de *clusterización* son técnicas no supervisadas de clasificación cuyo objetivo es separar datos en grupos o categorías homogéneas en su interior, pero distanciadas entre sí [45]. De manera similar, los *clusters* se definen como regiones contiguas de alta densidad de puntos separados por regiones de baja densidad de puntos [20]. A diferencia de otras técnicas de clasificación, no existe un *target* por el cual diferenciar los objetos entre sí, sino que las distinciones se hacen a partir de los resultados del agrupamiento.

Se tienen variados ejemplos de aplicaciones de los distintos métodos de clusterización, entre ellos se cuentan reconocimiento de imágenes [3], inteligencia de negocios [1], recuperación de información[25], biología [5] y seguridad [31].

2.3.1. K-means

El método de clusterización K-means [18] tiene por objetivo dividir un conjunto de datos en k *clusters* distintos, minimizando la suma de las distancias entre los puntos de un mismo grupo. Este método utiliza típicamente la distancia Euclidiana como medida de similitud y requiere que el número de *clusters* a buscar sea entregado como parámetro.

Al inicializar, K-means utiliza k elementos aleatorios como los centroides iniciales de los *clusters*, calculando así la distancia entre los distintos objetos a asignar y estos puntos generados de forma aleatoria para integrar cada elemento en el grupo más cercano a él. En cada iteración el centroide es actualizado y son reasignados los distintos objetos en caso de ser necesario. Las iteraciones continúan hasta que no hayan cambios.

En la Tabla 2.5 se resume el algoritmo K-means.

Algoritmo K-means	
Input:	
	k: el número de clusters
	D: el conjunto de datos que contiene los n objetos
Output:	
	Un conjunto de k clusters
Método:	
1	Se escoge arbitrariamente k objetos pertenecientes a D como los centroides iniciales de los clusters
2	Repetir:
3	Asignar cada objeto al cluster al que el objeto es más similar, basándose en el centro del cluster
4	Actualizar los centroides de los clusters, es decir, calcular la media de cada valor de los objetos de cada cluster
5	Hasta que no haya cambio

Tabla 2.5: Algoritmo K-means

2.3.2. Evaluación

Existen distintos criterios para evaluar la calidad de los *clusters* obtenidos, entre ellos se cuentan los Índices de Davies Bouldin [9], de Dunn [12], de Jaccard [17], el Coeficiente de Silhouette [36] y otros. A continuación se presentan los índices utilizados para llevar a cabo este trabajo.

Índice de Davies Bouldin

El índice de Davies Bouldin (DB) se utiliza para inferir qué tan apropiadas son las particiones obtenidas sobre un conjunto de datos y compararlas entre sí. No depende del número de grupos obtenidos ni del método utilizado para clusterizar.

Corresponde a la razón entre la distancia intra-*clusters* y la distancia inter-*clusters* y, considerando el objetivo de separar observaciones heterogéneas y agrupar las homogéneas entre sí, lo que se desea es que el valor obtenido sea lo más bajo posible.

Índice de Dunn

El índice de Dunn corresponde al cociente entre la mínima distancia inter-*clusters* y la máxima distancia intra-*clusters*.

Al igual que el índice de Davies Bouldin es una métrica de evaluación para *clusters* que apunta a encontrar grupos compactos a partir de los datos mismos que forman dichos grupos. En este caso se pretende encontrar la configuración de *clusters* que maximice el valor del índice.

2.4. Regresión Lineal

Una regresión lineal es tal que, existiendo una variable Y , la relaciona con variables explicativas X_1, \dots, X_k a través de una función lineal del tipo:

$$Y = B_0 + B_1X_1 + \dots + B_kX_k$$

, donde B_0 se conoce como el intercepto de la función y B_i $i = 1, \dots, k$ son los parámetros asociados a las variables explicativas.

Para estimar los valores de los parámetros y poder así construir la función lineal se puede utilizar el método de los mínimos cuadrados.

2.4.1. Método de los Mínimos Cuadrados

Sean $(y_i, x_{i1}, x_{i2}, \dots, x_{ip}) | i = 1, \dots, n$ los valores de las observaciones obtenidas para las $p+1$ variables de una muestra de tamaño n . Se plantea el modelo:

$$Y_i = B_0 + B_1X_{i1} + \dots + B_pX_{ip} \forall i$$

, donde B_0, \dots, B_p son las constantes desconocidas o parámetros del modelo. En realidad, dado que es difícil obtener constantes que cumplan la relación exacta, la relación debe escribirse:

$$Y_i = B_0 + B_1X_{i1} + \dots + B_pX_{ip} + e_i \forall i$$

, donde e_i es el error para la observación i debido al modelo.

El método de mínimos cuadrados busca minimizar la función de los errores cuadrados,

$$\sum e_i^2$$

a través de los parámetros del modelo. En el caso de una regresión lineal simple $\hat{y}_i = \hat{B}_0 + \hat{B}_1x_i$ el error sería $\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{B}_0 - \hat{B}_1x_i$. Con esto

$$\sum e_i^2 = \sum (y_i - \hat{B}_0 - \hat{B}_1x_i)^2$$

Resolviendo esto se obtiene que

$$\hat{B}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{B}_0 = \bar{y} - \hat{B}_1\bar{x}$$

2.4.2. Calidad Global del Modelo

Un índice interesante para medir la calidad global del modelo es el coeficiente de determinación R^2 :

$$R^2 = \frac{\sum_{i=0}^n (\hat{y}_i - \bar{y})}{\sum_{i=0}^n (y_i - \bar{y})}$$

es decir, R^2 es la razón entre la varianza explicada por el modelo y la total, tomando valores entre 0 y 1. Cuando toma valor 0 significa que las variables no explican nada en el modelo, mientras que si toma el valor 1 significa que el modelo permite describir las observaciones de manera exacta a partir de las variables explicativas.

Otra forma de medir la precisión de un modelo de regresión lineal es a través del *Mean Average Percentage Error* (MAPE). Esto es una métrica expresada en forma porcentual y cuya fórmula es:

$$M \frac{100\%}{n} \sum_{t=0}^n \frac{A_t - F_t}{A_t}$$

, donde n es el número de observaciones, A_t es el valor real de la variable objetivo y F_t es el valor obtenido para esta variable a través de la regresión.

La principal limitación que presenta esta métrica ocurre cuando la variable objetivo toma valor cero, lo que hace que se indefina el MAPE. De todas formas, es una medida fácil de aplicar y de comprender.

Capítulo 3

Metodología

3.1. Definición del problema

Enmarcado en el proyecto de movilidad en tiempo real, el objetivo de este trabajo es construir perfiles de clientes basados en su movilidad, enriquecidos con información demográfica y transaccional y evaluar su valor. La finalidad de los perfiles es lograr una comunicación más efectiva con los clientes que en el caso en que solo fuesen construidos con la información tradicional.

Para lograr construir los perfiles descritos, se debe primero tener una idea de qué es lo que se pretende conocer de los usuarios desde el punto de vista de la movilidad, lo que se aborda a continuación.

3.1.1. Lugares relevantes y rutina diaria

Como se menciona en el Capítulo 2, diversos estudios concluyen que los lugares relevantes para un usuario y la rutina diaria que éste desarrolla son la base para poder entender y describir gran parte de sus movimientos. Por lo tanto, estas dimensiones deben ser abarcadas en este trabajo.

Los lugares relevantes para los usuarios fueron identificados en [35], tal como se explica en la sección 2.2. Se pretende identificar las principales actividades habituales -o que conforman la rutina diaria- que un usuario cualquiera realiza para encontrar en cuáles de estos lugares identificados las desarrolla.

De acuerdo a [11] y tal como se muestra en la Figura 3.1, entre las principales actividades desarrolladas en la rutina diaria de las personas se cuentan el trabajo remunerado, los estudios, el trabajo no remunerado (tareas del hogar, cuidado de miembros del hogar y otros), las necesidades fisiológicas y el uso de medios de comunicación masivos. Se puede pensar que estas actividades son desarrolladas por los usuarios principalmente en su hogar o lugar de

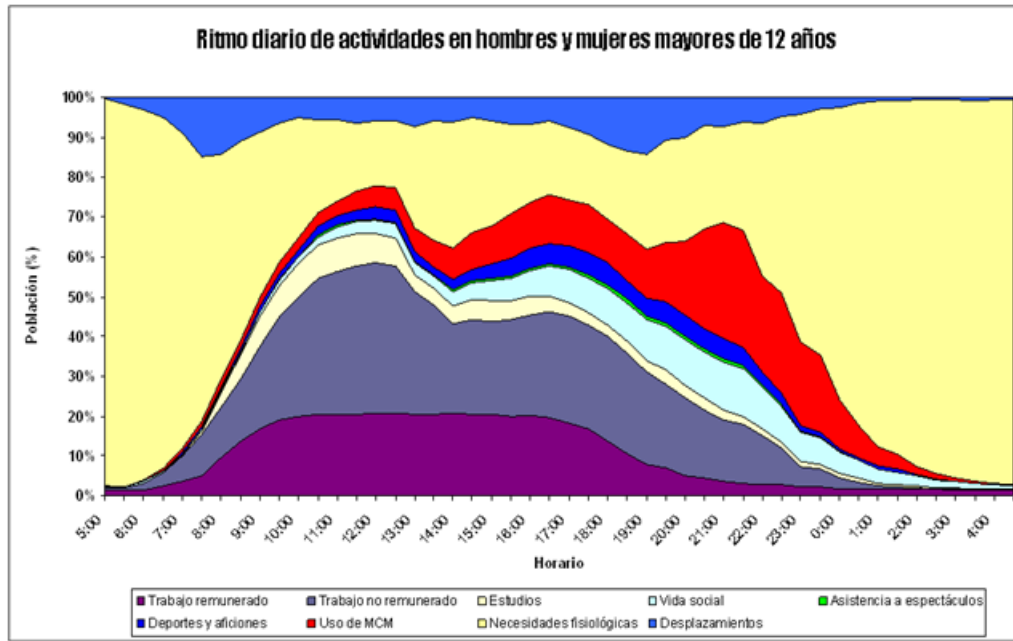


Figura 3.1: Ritmo diario de las personas. Fuente: [11]

trabajo o estudio, lo que coincide con el hecho de que para la mayoría de los clientes se hayan encontrado 2 PRI. A partir de esto se decide concentrar el estudio de los lugares relevantes y la rutina diaria de las personas en identificar y caracterizar las PRI correspondientes a sus hogares y trabajos.

Las PRI en promedio tienen un radio de cobertura de 1,2 km, con lo que las que sean identificadas como hogar y trabajo comprenderán una mayor cantidad de actividades posibles a desarrollar en el lugar que únicamente las relacionadas a la actividad principal del usuario. Se piensa que la composición de posibles actividades de una zona puede guardar relación con el mayor o menor desplazamiento de los usuarios en la ciudad. Por ejemplo, en el caso de alguien que viva en una zona totalmente residencial, tendrá que desplazarse fuera del área comprendida por su PRI definida como hogar para desarrollar actividades comerciales, de estudio, de trabajo, etc., mientras quien viva en una zona con un mix más amplio podría encajonar gran parte de sus actividades en la misma PRI a la que se atribuye se ubica el hogar del usuario. Se decide a partir de lo explicado, caracterizar las PRI en función de su uso de suelo y tiempo para captar las actividades que en esas áreas se desarrollan.

En la Región Metropolitana existen importantes diferencias entre quienes habitan en un lugar u otro. En [21] se muestra por ejemplo que quienes residen en el sector oriente de la región presentan un ingreso promedio por hogar equivalente a casi 4 veces el de quienes habitan en el sector de menos ingresos y son a la vez quienes menos utilizan el transporte público y más autos por hogar poseen. Se muestra también que el sector centro de la región tiene la menor proporción de actividades residenciales y educacionales y la menor cantidad de personas por hogar, lo que contrasta con el sector suroriental que presenta la mayor proporción de área destinada a actividades residenciales y educacionales y un número más alto de residentes por hogar.

Considerando las diferencias entre los distintos sectores de la región mostrados en la Figura 3.2, los que fueron particionados de manera similar a lo visto en [21] para aprovechar las características que en dicho trabajo se señalan, se establece que es relevante para describir el hogar y trabajo de un individuo el conocer en cuál de estos sectores está ubicado.

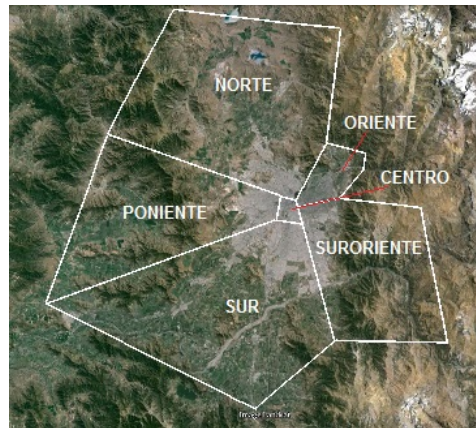


Figura 3.2: División RM en 6 sectores

3.1.2. Actividades Esporádicas

Si las actividades habituales de un usuario explican gran parte de sus movimientos, entonces las esporádicas debiesen explicar el resto. Por lo tanto, se debe incorporar una métrica que describa el desarrollo de este tipo de actividades para cada cliente.

Se debe desarrollar un procedimiento que capture aquellos momentos en que las personas salen de su rutina diaria, pero para tener una aplicación en el proyecto de movilidad en tiempo real se debe tener además una interpretación para las actividades esporádicas que el usuario realiza en dichas ocasiones.

Pueden existir diversas aplicaciones derivadas de conocer e interpretar las actividades no habituales de un cliente reflejadas en sus registros telefónicos. En primer lugar el conocer qué tan propenso es el usuario a salir de su rutina y en qué momentos lo hace se puede convertir en una fuente de información importante a la hora de comunicar sobre una promoción o evento en especial si es que se piensa que aquellos con mayor tendencia a escapar de su cotidianidad tienen mayor probabilidad de hacer caso a dicha comunicación.

Comunicaciones enfocadas a lugares en específico pueden ser también abordadas conociendo la frecuencia con que el usuario visita este lugar que no pertenece a las actividades habituales que realiza. Si se quisiera atraer a personas a un evento masivo realizado en un punto de la región, como un centro comercial, puede ser útil enfocar una campaña de comunicación a aquellos que han asistido al lugar en ocasiones anteriores, pero que no transitan diariamente por él, pues estas últimas con alta probabilidad se enterarán o asistirán como parte de su rutina diaria.

Las hipótesis anteriormente planteadas para la aplicación de la información que entrega el conocer las actividades esporádicas de un usuario pueden ser ciertas o no y debe ser testeado.

Lo claro en esto es que en las ocasiones que los clientes salen de su habitualidad existe un potencial para generar comunicaciones efectivas que aporten al proyecto.

3.1.3. Métricas de Movimiento

Teniendo cubierto lo que explica el movimiento de los usuarios, se debe pasar a desarrollar la forma en que éste se debe medir y describir como tal. Para esto se utiliza como base lo desarrollado en [16], donde se calculan métricas para medir distintas dimensiones del movimiento de cada persona y que en conjunto son capaces de dar una idea general de su magnitud y forma, haciendo posible la comparación entre usuarios.

Las métricas a utilizar y el porqué fueron escogidas se presenta a continuación.

Distancia Recorrida por el Usuario

Esta métrica refleja cuánto se mueve una persona, independiente de cualquier otra característica que su movilidad pueda tener.

Si se compara el movimiento de una persona que se traslada incesantemente entre dos puntos específicos ubicados cerca el uno del otro, con el de un taxista que se desplaza por toda la ciudad, se puede decir que ambos tienen un alto nivel de movilidad dado que en un período de tiempo recorren altas distancias, pero la forma en que se mueven tienen características diferentes. Es decir, un usuario que recorre altas distancias, es un usuario que tiene alta movilidad, independiente de las características de su desplazamiento.

En la Figura 3.1.3 se muestran 3 casos de usuarios con distinta movilidad. En el primero la distancia recorrida es de 2 en total, mientras que para los usuarios 2 y 3 la distancia recorrida es la misma y equivalente a 14, pero ambos tienen formas de moverse muy diferentes. El usuario 1 tiene una baja movilidad en comparación a los otros 2, pero a pesar de que estos últimos tienen el mismo valor para esta métrica, reflejan que para describir el movimiento de un usuario se necesita más información.

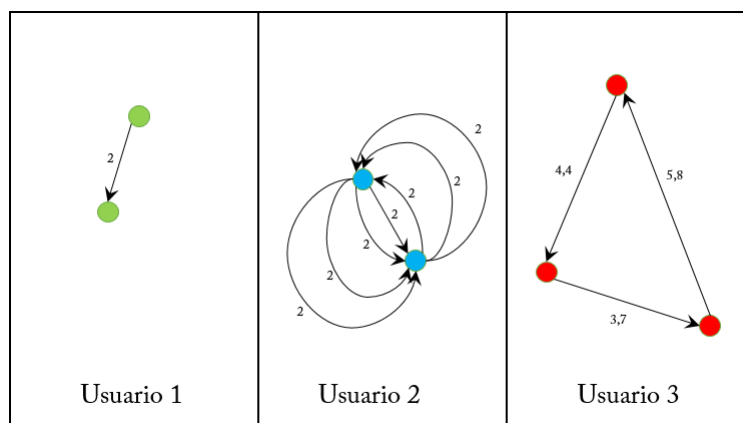


Figura 3.3: Distancia recorrida para distintos usuarios

Número de Locaciones Diferentes

Se define esta métrica para tener una idea de cuántos lugares son los que motivan el movimiento de un usuario. Distintas personas pueden recorrer la misma distancia en un cierto período, pero existe diferencia si es que en ese tiempo acude a una baja cantidad de posiciones, ya sean estas distantes entre sí o cercanas pero visitadas con una mayor periodicidad, versus un número mucho mayor de locaciones, tal como se muestra en la Figura 3.4.

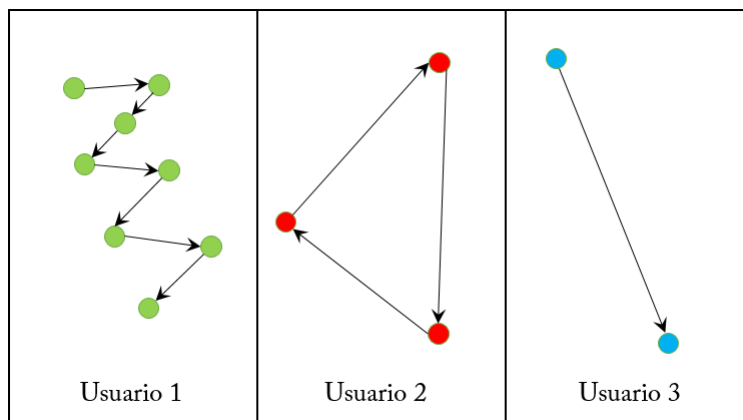


Figura 3.4: Diferencia de número de posiciones entre usuarios

Radio de Giro

Se pueden examinar casos de usuarios que recorren una gran distancia entre muchas posiciones diferentes, pero abarcando un área pequeña dado que las locaciones que visita se encuentran cercanas entre sí, mientras es posible pensar en otro caso de quien, con la misma distancia y número de posiciones, se desplaza con una alta dispersión por la ciudad al encontrarse sus puntos de actividad lejanos entre sí. Así como estas situaciones, se pueden encontrar muchas otras en que para continuar describiendo los movimientos de los clientes es necesario tener una medida de dispersión de los movimientos que permita dimensionar las áreas que abarca cada usuario.

El Radio de Giro se puede interpretar como una medida de dispersión de los movimientos, pues es una medida de desviación desde las diferentes posiciones del usuario con respecto al centro de masa de las mismas. En la Figura 3.5 se observan los casos descritos anteriormente y cómo el radio de giro ayuda a dimensionar cómo las distancias que recorre el usuario y las posiciones que visita están emplazadas en la región. Observar que el área al interior del círculo con el radio obtenido centrado en el centro de masa de las posiciones puede incluso nunca ser visitada por el usuario, pero de todas formas esta métrica permite establecer diferencias en el movimiento de las personas.

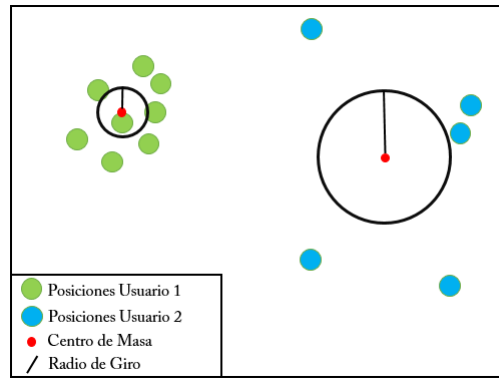


Figura 3.5: Diferencia de radio de giro entre usuarios

Grado de Retorno (GR)

Habiendo descrito el movimiento de un usuario con las métricas anteriores, es deseable una que incluya el desplazamiento en el tiempo para un conocimiento más íntegro de la movilidad de los clientes. Es importante la diferencia entre quienes tengan los mismos valores para las métricas anteriores, pero una visite sus distintas locaciones en un orden establecido día a día versus una con un comportamiento más errático.

El Grado de retorno fue diseñado como una medida simple de la regularidad temporal, de los movimientos de las personas o, más bien, de la regularidad con que visita los distintos lugares que registra.

3.1.4. Patrones Grupales de Movimiento

Una vez identificada la forma en que se mueven individualmente los clientes, parece interesante para el proyecto conocer cómo esto se relaciona con las formas grupales de movimiento para poder por ejemplo comparar entre las direcciones en que las personas se mueven o los sectores por los que pasan al desplazarse de un lado a otro.

Es valioso para el proyecto conocer por ejemplo cuáles son las tendencias en los movimientos de las personas, las rutas más utilizadas, combinaciones de lugares usuales, etc. De esta forma se pueden enfocar campañas masivas dependiendo de características o patrones grupales de movimiento en vez de individuales.

Para aprovechar la información que entrega la división de la región en los 6 sectores de la región y conocer las direcciones de movimiento y los lugares por los que las personas se desplazan, se decide analizar las transiciones entre sectores que realizan los usuarios durante sus movimientos habituales. De esta forma se puede establecer por ejemplo que a personas que se muevan siempre entre sectores Norte y Centro y no visiten otros no se les envíen campañas que obliguen su traslado a un sector distinto de la ciudad.

Las celdas de Voronoi son en total 1.200 en la región estudiada, por lo que los resultados que se obtendrían al realizar el cálculo de transiciones entre ellas serían difíciles de interpretar.

3.2. Análisis y Selección de Datos

3.2.1. Información Disponible

Información Demográfica y Transaccional

La información demográfica y transaccional proporcionada por la empresa respecto a sus clientes corresponde a los datos de Abril 2013 y Mayo 2013. Ésta contiene información demográfica, de tráfico y comercio. La información de esta fuente, agrupada por categorías, se presenta en la Tabla 3.1.

Registros de Llamadas Telefónicas

Como se menciona en la sección 1.2, se cuenta con los registros de llamadas (*Call Detail Records* - CDR) de una empresa de telecomunicaciones, correspondientes a todas las llamadas de los clientes de telefonía móvil bajo régimen contractual realizadas entre los días 18 de Abril de 2013 y 3 de Julio de 2013. Esta información constituye un total por sobre los 800 millones registros de cerca de 2,8 millones de números telefónicos distintos. Cabe señalar que los datos utilizados se han anonimizado por la empresa y no existe manera de identificar a un usuario determinado. Se ordenan los datos de forma que cada registro incluya lo que se muestra en la Tabla 3.2.

Cabe hacer notar que por cada llamada realizada, en caso de que ambos números involucrados pertenezcan a la compañía, existen 2 registros como los mostrados en la Tabla 3.2. En caso de que solo uno de los números sea cliente de la compañía, entonces solo se encuentra un registro de la llamada.

Además de lo que aquí se menciona, se encuentra a disposición lo presentado en la sección 2.2, es decir, las Posiciones Relevantes independientes (PRI) y dependientes (PRD) del tiempo y la caracterización tempo-espacial de la ciudad.

3.2.2. Análisis de Datos

Información Demográfica y Transaccional

De los cerca de 670.000 potenciales usuarios que cumplen con tener contrato, ser de la categoría personas y pertenecer a la Región Metropolitana, se escoge una muestra aleatoria de 350.000 para el análisis. Dicha muestra es bastante amplia, pero se utiliza dada la necesidad

Tabla 3.1: Información disponible por categoría

Tipo de Información	Nombre	Descripción
Identificación del Cliente	Código Cliente	Identificación del cliente
	ID Telefónico	Código único generado por la empresa
	Código Abonado	ID del usuario
Características del Cliente	Segmento	Personas, Pyme, Empresa, Negocios, ...
	Antigüedad cliente	Meses del cliente en la compañía
	Valor del cliente	Tradicional o Alto
Información socio-demográfica	Región	Región de residencia declarada
	Comuna	Comuna de residencia declarada
	Grupo Socio Económico	
	Edad	
	Género	
Información del Servicio	Tipo de contrato	Híbrido o Contrato Puro
	Tipo de plan	
	Servicio	Voz, Banda Ancha,...
	Antigüedad plan	Meses con el mismo plan
	Recambios de plan	Veces que ha cambiado de plan
	Fecha último recambio	
Características del móvil	Modelo del equipo	
	Smartphone	Si el equipo es smartphone o no
	Pantalla Táctil	Si el equipo tiene pantalla táctil o no
	Recambios aparato	Nº veces de cambio aparato
	Antigüedad móvil	
Tráfico de Voz	Minutos de salida	Nº de minutos correspondientes a llamadas realizadas
	Minutos de entrada	Nº de minutos correspondientes a llamadas recibidas
	Llamadas de salida	Nº de llamadas realizadas
	Llamadas de entrada	Nº de llamadas recibidas
Tráfico de Mensajería	SMS salida	Nº de mensajes de texto enviados
	SMS entrada	Nº de mensajes de texto recibidos
	MMS salida	Nº de mensajes multimedia enviados
	MMS entrada	Nº de mensajes multimedia recibidos
Navegación	Tráfico de datos (KB)	
	Tráfico de datos (Minutos)	
	3G	Si tiene o no tecnología 3G contratada
Facturación	ARPU	Ingreso promedio por cliente
Cobranza	Mora 6	Veces en mora en últimos 6 meses

Tabla 3.2: Descripción información registros telefónicos

Dato	Descripción
ID Telefónico	ID dado a quien realiza o recibe la llamada
Fecha	Fecha exacta de la llamada
Hora	Hora exacta de la llamada
Id Antena	Identificación de la antena que captó la llamada
Latitud	Latitud de la antena
Longitud	Longitud de la antena
Azimut	ángulo de orientación de la antena
Ancho	Ancho de cobertura de la antena
Distancia	Radio de cobertura antena

de poblar los movimientos que se realizan en toda la ciudad y poder obtener patrones de esto.

Se establece que las variables de interés son aquellas asociadas a información sociodemográfica, tráfico de voz y mensajería, navegación, características del móvil, mora y antigüedad del cliente.

El GSE de los clientes a analizar se concentra en los grupos ABC1 y C2. La distribución entre hombres y mujeres en la muestra es pareja y existe una gran cantidad de casos -sobre 180 mil del total, más de 100 mil de la muestra- cuyo género es desconocido para la empresa.

Respecto a la información asociada al tráfico de voz y de mensajería, se excluye del análisis a los mensajes multimedia (MMS) debido al bajo uso que le dan los usuarios (promedio mensual menor a 0,04 por usuario). Si bien se podría pensar hacer lo mismo para el caso de los mensajes de texto (SMS), se debe destacar que en promedio los usuarios envían alrededor de 200 veces más SMS, por lo que se considera mucho más relevante y se incluye en esta etapa del análisis.

En el Anexo A se muestra estadística descriptiva de todas las variables demográficas y transaccionales analizadas.

Registro de llamadas Telefónicas

En promedio, cada usuario registra 487,3 llamadas en el período de análisis, es decir, 6,7 llamadas diarias. En la Figura 3.6 se puede apreciar que, a pesar de que el máximo de registros de llamadas es superior a las 11.000, casi el 100 % de los usuarios presenta menos de 3.000 en el período, es decir, menos de 41 registros diarios. Incluso cerca de un 80 % de los usuarios a analizar posee menos de 1.000 registros en total en el período completo, equivalente a menos de 14 llamadas -recibidas o realizadas- diarias.

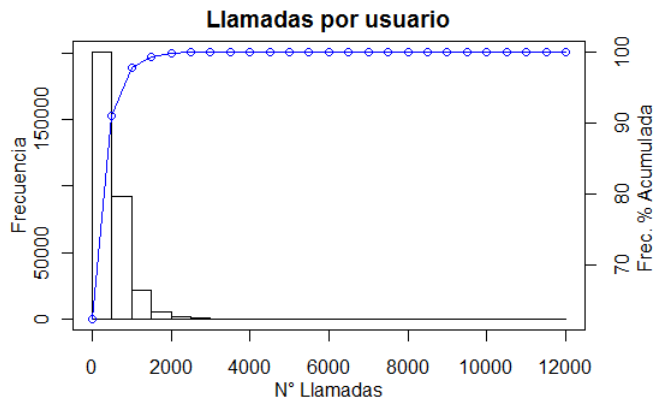


Figura 3.6: Histograma de llamadas por usuario

En la Figura 3.7 se muestran las llamadas agrupadas por bloque horario en una semana completa. Se observa que la cantidad de llamadas en el tiempo es similar entre Lunes y Jueves, donde en las madrugadas la actividad decae, mientras que los fines de semana no tienen una

disminución tan marcada en los mismos horarios. Esto es consecuente con el entretenimiento nocturno que se desarrolla Viernes y Sábados por la noche.

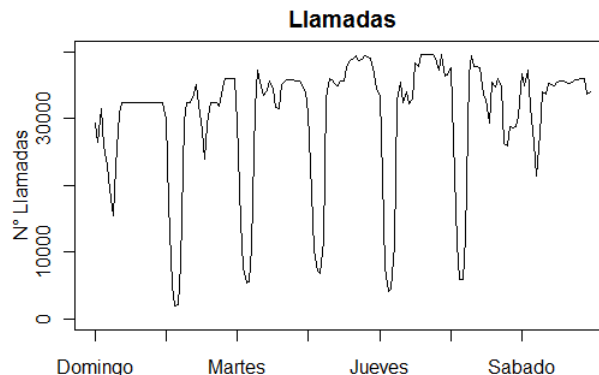


Figura 3.7: Llamadas agregadas por bloque horario en una semana

Posiciones Relevantes

Se analizan las Posiciones Relevantes obtenidas en [35] y explicadas en el Capítulo 2.

Se cuenta con 5,2 Posiciones Relevantes no dependientes del tiempo (PRI) por usuario en estudio. Lo anterior quiere decir que en promedio para cada usuario se identificaron 5,2 zonas intensas en actividad telefónica y que se presumen relevantes para su desarrollo habitual. Es importante mencionar que la moda de este dato es de 2 posiciones por cliente.

En el caso de las Posiciones Relevantes en ventanas de tiempo de una hora (PRD), el promedio por usuario es de 83,7 con una moda de 80, es decir, en ambos casos menos de una posición por bloque horario de una semana promedio¹.

3.2.3. Selección de Datos

Se realiza un análisis de correlación para descubrir los datos que presentan una mayor relación y que pueden estar replicando información. Las tablas con estos resultados se encuentran en el Anexo B.

Las mayores correlaciones se presentan entre la información de los dos meses de análisis en cuanto a tráfico de mensajería, voz y datos. Esto también se observa en la interacción entre número de llamadas y minutos tanto entrantes como salientes.

Se decide no incluir en el análisis el número de llamadas, pues la información que entrega está muy relacionada a la que entrega la cantidad de minutos de un usuario. Además, minutos es una medida que se usa tradicionalmente para medir la intensidad de tráfico de voz en la empresa.

¹Existen un total de 168 bloques horarios

Si bien en el caso de mensajería, voz y datos, los meses a analizar se encuentran estrechamente relacionados, se decide incluir la información de ambos períodos en una sola variable. Esto se explica con mayor detalle en secciones posteriores relacionadas a transformación de variables.

La información del tráfico de datos medido en kilobytes y segundos es redundante aunque no se refleje en el análisis de correlaciones, pues se mide el mismo hecho de distintas formas. Se decide conservar la medida en kilobytes dado que se utiliza comúnmente en la empresa y, en general, tiene menores coeficientes de correlación con el resto de las variables.

Para el caso del 3G, ésta variable se encuentra altamente correlacionada con el tráfico de datos, lo que se suma al hecho de que prontamente será eliminada la posibilidad de navegar con tecnología 2G, por lo que no se considera relevante para el análisis.

3.3. Preprocesamiento de Datos

3.3.1. Tratamiento de Valores Fuera de Rango

Al fijarse en la edad de los clientes se observa en la Figura 3.8(a) que existen valores negativos y otros muy altos. Para solucionar este problema se considera como un rango normal el equivalente a dos desviaciones respecto a la media, es decir, edades de entre 12 y 85 años. Por otro lado, la empresa no trabaja con usuarios menores de edad, por lo que el rango final es de entre 18 y 85 años. Se aprecia el resultado de los cambios en la Figura 3.8(b).

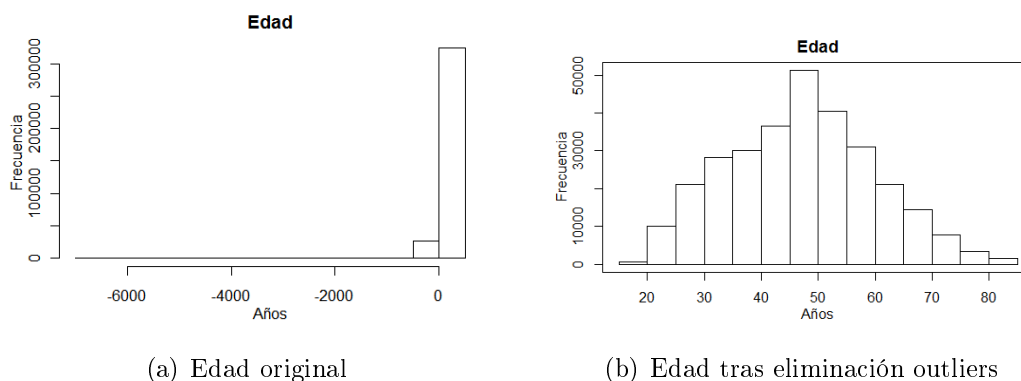
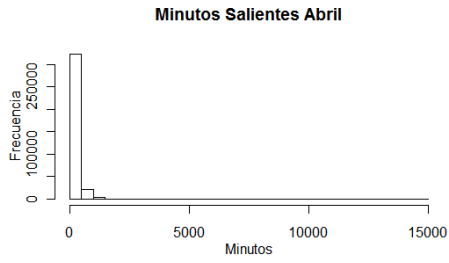


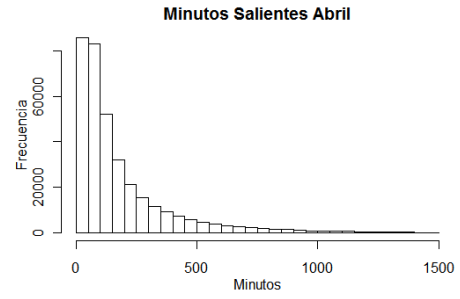
Figura 3.8: Histogramas edad

Respecto al tráfico de voz, se observa en la Figura 3.9 que los valores originales son en su mayoría menores a 15.000 minutos en el caso del tráfico saliente y a 2.00 en el caso del entrante, pero existen datos que se escapan de dichos rangos. Se toma como valor máximo el de la media más 5 desviaciones estándar. Esto para eliminar aquellos valores que se escapan del rango, pero conservar la mayor cantidad de observaciones para continuar el análisis².

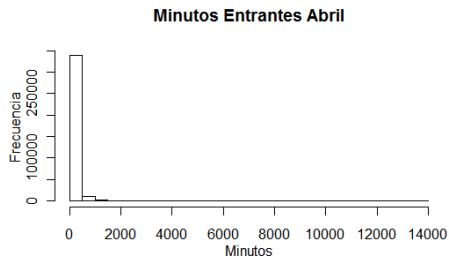
²El análisis de desviaciones se encuentra en el Anexo C



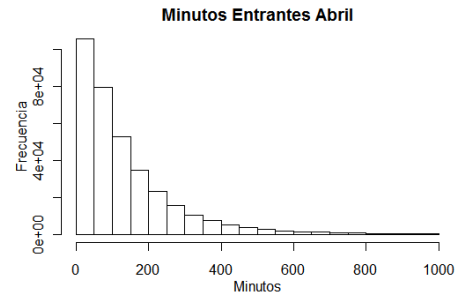
(a) Minutos Salientes Abril original



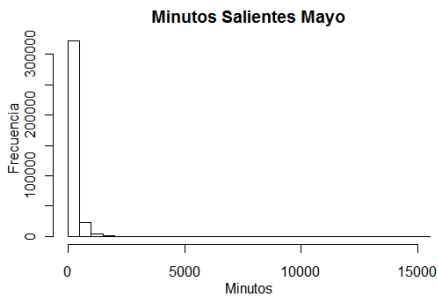
(b) Minutos Salientes Abril tras eliminación outliers



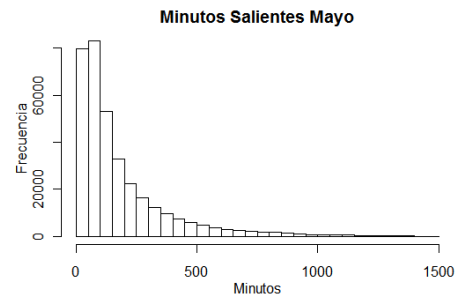
(c) Minutos Entrantes Abril original



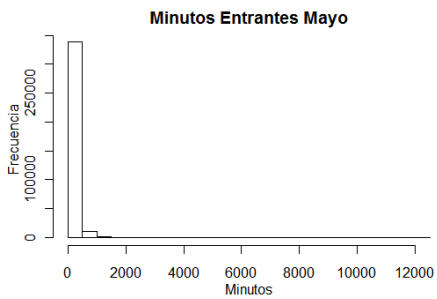
(d) Minutos Entrantes Abril tras eliminación outliers



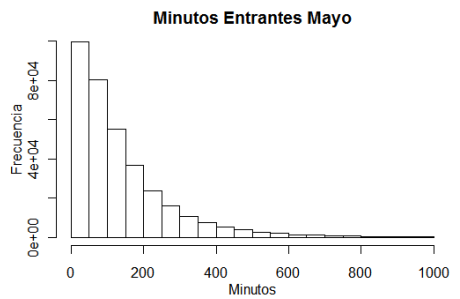
(e) Minutos Salientes Mayo original



(f) Minutos Salientes Mayo tras eliminación outliers



(g) Minutos Entrantes Mayo original



(h) Minutos Entrantes Mayo tras eliminación outliers

Figura 3.9: Histogramas tráfico de voz

Tanto para el tráfico de datos como de mensajería existen datos que escapan del rango en que se concentran sus valores. Para ambos se impone como valor máximo la media más 5 desviaciones estándar. Los histogramas previos y resultantes se muestran en las Figuras 3.10 y 3.11.

El ARPU cuenta con la presencia de valores fuera de rango tanto muy superiores como muy inferiores a lo que se esperaría, tal como lo refleja la Figura 3.12. Se decide en primera instancia considerar solo aquellos valores mayores o iguales a cero y luego imponer como valor máximo 5 desviaciones a partir de la media, al igual que para las variables anteriormente analizadas.

Para la variable que refleja la cantidad de recambios del equipo se observa que el máximo corresponde a 35 veces, pero su media es de 2,5 con una dispersión de 2,1, con lo que se ve que cabe realizar algún análisis respecto a los potenciales *outliers* que pudiera tener esta información. Tras el análisis, se decide limitar el número de recambios a 13 veces. El histograma original y como queda tras la aplicación del tratamiento se observan en la Figura 3.13.

Para el caso de la antigüedad de los equipos, tras el análisis correspondiente, se considera que el valor máximo aceptable es de 97 meses, equivalente a 4 desviaciones por sobre la media. El histograma original y como queda tras la aplicación del tratamiento se observan en la Figura 3.14.

La mora 6 se define como el número de veces en que el usuario cayó en mora en los últimos 6 meses, por lo tanto que su valor máximo sea de 9 veces no es consistente. En este caso se eliminan directamente aquellos valores mayores a 6.

Tras todos estos cambios, el número de clientes a analizar es ligeramente superior a los 335.000.

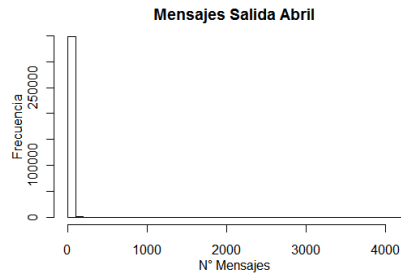
3.3.2. Tratamiento de Valores Perdidos

El primer paso es analizar cuántos de los usuarios en análisis no registran posiciones relevantes, ya sea porque no cumplen con los criterios del algoritmo para formarlos o por falta de información. De aquí resulta que 313.025 sí cuentan con esta información, es decir, un 93,4% de la muestra original, el resto de los clientes se excluyen del proceso.

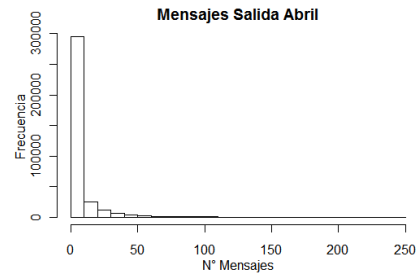
La variable edad presenta un 7,3% de datos perdidos. Se decide usar una regresión lineal para relacionar el resto de la información demográfica y transaccional con la faltante y así obtener valores coherentes con el resto de las observaciones para la edad.

En el caso del grupo socio-económico la proporción de datos perdidos es cercana al 3%. Se usa el método de la regresión lineal al igual que en el caso anterior.

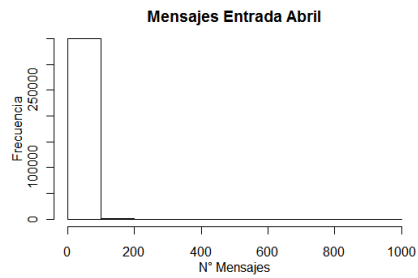
Tras los cambios descritos, el número de observaciones disponibles para continuar el análisis es 313.025. En la Tabla 3.3 se muestra un resumen de los estadísticos descriptivos de las variables tras los tratamientos aplicados.



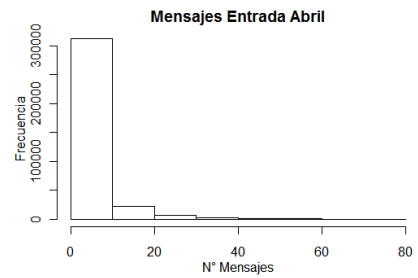
(a) Mensajes Salida Abril original



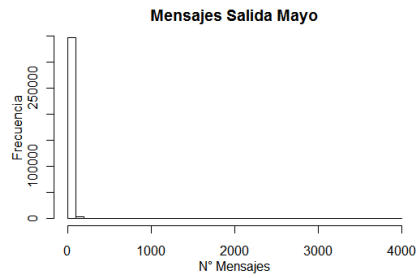
(b) Mensajes Salida Abril tras eliminación outliers



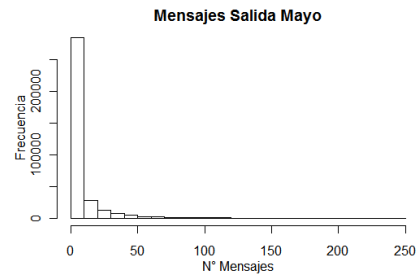
(c) Mensajes Entrada Abril original



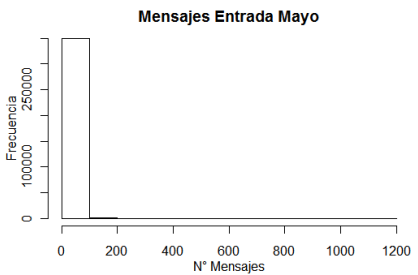
(d) Mensajes Entrada Abril tras eliminación outliers



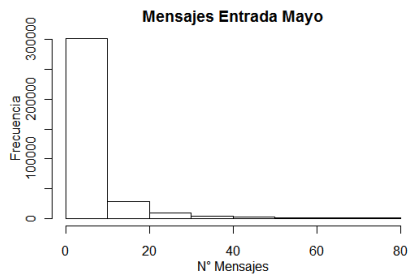
(e) Mensajes Salida Mayo original



(f) Mensajes Salida Mayo tras eliminación outliers

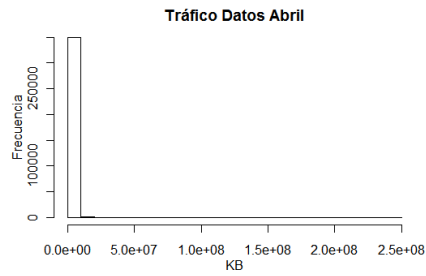


(g) Mensajes Entrada Mayo original

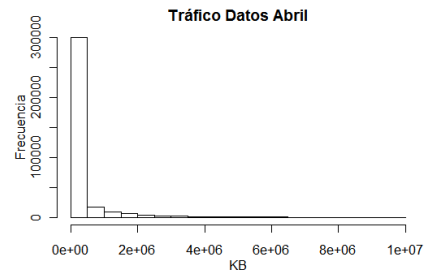


(h) Mensajes Entrada Mayo tras eliminación outliers

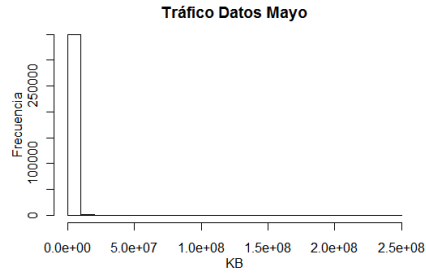
Figura 3.10: Histogramas tráfico de mensajería



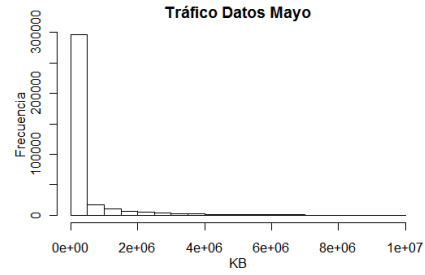
(a) Tráfico Datos Abril original



(b) Tráfico Datos Abril tras eliminación outliers

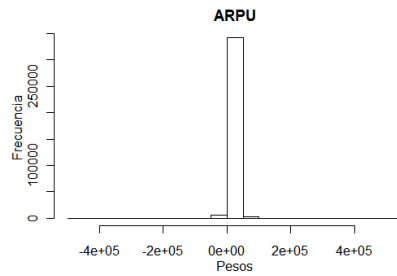


(c) Tráfico Datos Mayo original

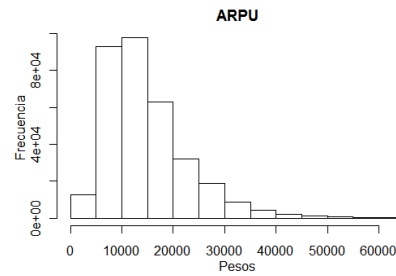


(d) Tráfico Datos Mayo tras eliminación outliers

Figura 3.11: Histogramas tráfico de datos

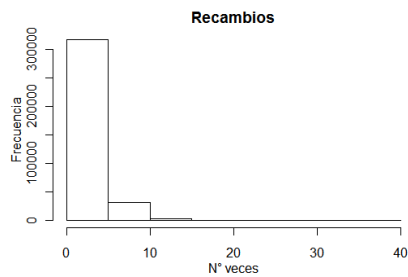


(a) ARPU original

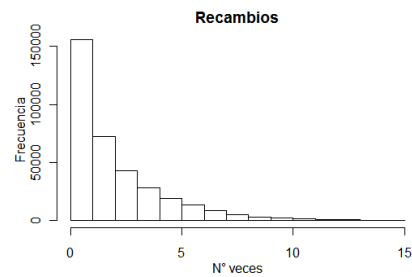


(b) ARPU tras eliminación outliers

Figura 3.12: Histogramas ARPU

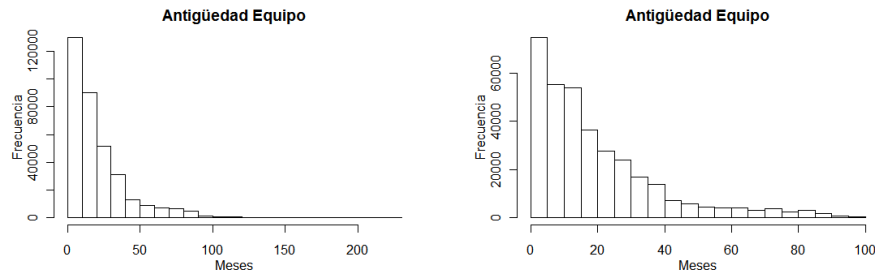


(a) Recambios original



(b) Recambios tras eliminación outliers

Figura 3.13: Histograma recambios



(a) Antigüedad equipo original (b) Antigüedad equipo tras eliminación outliers

Figura 3.14: Histogramas antigüedad equipo

Tabla 3.3: Estadísticos tras tratamientos de valores perdidos y fuera de rango

	Media	Desviación	Mínimo	Máximo
Grupo Socio Económico (GSE)	2,8	1,0	1,0	5,0
Tráfico Datos Abril	294,1	812,4	0,0	7.337,4
Minutos Salida Abril	167,0	179,2	0,0	1.394,0
Minutos Entrada Abril	132,5	127,2	0,0	979,0
SMS Salida Abril	5,9	13,3	0,0	214,0
SMS Entrada Abril	4,1	7,3	0,0	65,0
Antigüedad Cliente	50,1	43,6	0,0	288,0
Edad	47,6	12,9	18,0	85,0
Antigüedad Equipo	19,8	19,1	0,0	224,0
Tráfico Datos Mayo	335,7	915,9	0,0	8.231,4
Minutos Salida Mayo	172,1	179,8	0,0	1.408,0
Minutos Entrada Mayo	135,5	126,3	0,0	982,0
Mora en últimos 6 meses	2,2	2,0	0,0	6,0
SMS Salida Mayo	7,4	16,1	0,0	232,0
SMS Entrada Mayo	5,1	8,7	0,0	78,0
Número recambios equipo	2,4	2,0	0,0	13,0
ARPU (<i>Average Revenue Per User</i>)	14.100,1	7.958,8	0,0	62.946,0

3.4. Transformación de Variables

Para la variable GSE existen 5 categorías: ABC1, C2, C3, D y E. Se transforman a variables ordinales como se muestra en la Tabla 3.4.

La transformación no se hace con variables binarias, ya que para el caso de GSE se puede decir que existe un orden lógico entre los distintos valores que toma.

Para el caso de la variable género existen 3 categorías, femenino, masculino y desconocido. Como no existe un orden entre estas variables se crean dos variables *dummy*. Una que toma valor 1 si el usuario era hombre y otra que toma el mismo valor en caso de ser mujer. Ambas son 0 en caso de ser de género desconocido.

En el caso de tráfico de mensajería, datos y voz se integran al análisis ambos meses promediados de forma separada por variables entrantes y salientes. El tráfico de datos se transforma de Kilobytes a Megabytes dividiendo por 1.024.

Tabla 3.4: Transformación GSE

GSE	Transformación
ABC1	1
C2	2
C3	3
D	4
E	5

3.5. Análisis de Resultados

3.5.1. Identificación y Caracterización del Hogar y Trabajo

De manera similar a lo realizado en [27], para identificar las PRI correspondientes a hogar y trabajo, se decide utilizar el siguiente procedimiento:

1. Definir horarios por defecto para las actividades residenciales y de trabajo para todos los usuarios.
2. Listar las PRI de cada usuario
3. Para cada una de las PRI, contar la cantidad de días que la persona realiza llamadas desde esa zona en el horario indicado, ya sea residencial o de trabajo.
4. Seleccionar como hogar aquella PRI con mayor cantidad de días con llamadas en horario residencial y como trabajo aquella que cumpla con el mismo criterio, pero en horario de trabajo.

Se utiliza el número de días en vez de llamadas debido al efecto que puede causar el que una persona en un determinado momento tenga la necesidad de realizar -o recibir- muchas llamadas desde un lugar no relacionado a su comportamiento habitual. Además, muchos usuarios presentan una gran cantidad de llamadas en una misma posición versus el resto de ellas debido a la forma en que utilizan su teléfono móvil, lo que hace que dicha posición sea con alta probabilidad como ambas actividades buscadas.

El horario en que es más común estar desarrollando actividades en el trabajo en Chile es desde las 9 a.m. hasta las 6 p.m.[11]. En el caso de las actividades relativas al hogar, el horario correspondiente es desde las 7 p.m. hasta las 8 a.m..

Tras obtener los resultados de aplicar el procedimiento planteado, se constata que incluso utilizando los días en vez de las llamadas, para varios usuarios se identifica el hogar como la misma PRI que identifica como trabajo. Esto se debe a que en algunos casos las personas realizan llamadas en su hogar en horario de trabajo. Para solucionar esto, se modifica el procedimiento para que el trabajo sea identificado como aquella PRI con mayor cantidad de días con llamados en horario laboral sujeto a que no sea la PRI previamente identificada como hogar.

En las Figuras 3.15 y 3.16 se muestran los histogramas de llamadas de un número cualquiera para sus lugares identificados como hogar y trabajo tras la modificación al procedimiento,

diferenciando entre los registros en el horario definido de la actividad y los que se realizaron fuera de dicho horario. Se observa una gran cantidad de llamadas fuera de horario en el hogar, lo que como se explicó anteriormente, lleva a que dicho lugar sea inicialmente identificado como ambas actividades diarias. Por otro lado, se ve que en el trabajo las llamadas fuera de horario disminuyen ostensiblemente y que el total de llamadas registradas en esta posición corresponde en general a la mitad de las asociadas a la otra.

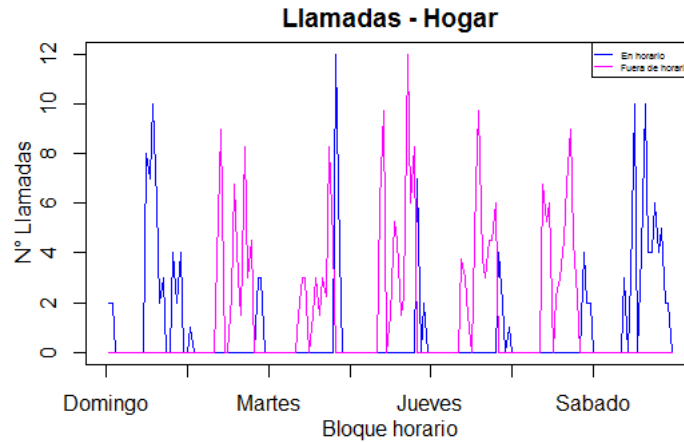


Figura 3.15: Llamadas en horario y fuera de horario para el Hogar

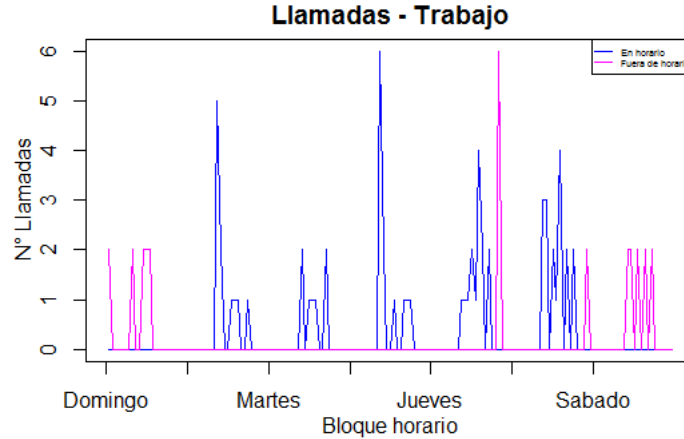


Figura 3.16: Llamadas en horario y fuera de horario para el Trabajo

Para disminuir el error al identificar una cierta posición como una actividad específica, se agrega una métrica de desviación que captura la proporción de registros en un lugar identificado como una actividad que fueron realizados en un horario diferente al correspondiente a dicha actividad, es decir, la desviación del hogar corresponde a la proporción de llamadas de las que se tiene registro en esa posición en horario de trabajo sobre el total de llamadas registradas en dicha posición. Para el caso visto en las Figuras 3.15 y 3.16, la desviación hogar correspondería a 0,33 y en el caso del trabajo sería de 0,18.

Una vez procesados los datos con el procedimiento establecido, se obtiene una media de desviación para el hogar (DH) de 51,3 % y 48,1 % para el trabajo (DT). A partir de esto, se analizan los casos extremos para llegar a crear las siguientes reglas de corrección:

1. Si $DH > 0,5$ y $DT > 0,5$, entonces se debe identificar primero el trabajo y luego el hogar, ya sea porque son suarios que muestran un comportamiento que parece indicar un horario de trabajo distinto al regular o porque registraban una baja actividad de llamadas en la posición que parecía ser su hogar, forzando a la PRI que debía asociarse al trabajo a ser identificada como hogar al ser este el primero en etiquetarse.
2. Si $DH < 0,3$ y $DT > 0,6$, entonces se debe dejar que el trabajo sea el mismo que el hogar en caso de que así lo determine el procedimiento, ya que estas personas mostraban un comportamiento que podría reflejar que no trabajan o que trabajan desde su residencia.
3. Si $DH > 0,6$ y $DT < 0,3$, entonces se aplica la misma regla que en el primer caso, pues estas personas parecían tener un horario de trabajo distinto al normal, pero muy variado entre ellos, haciendo difícil determinar un horario laboral para ellos.

Con estos cambios, el procedimiento queda como sigue:

1. Listar PRI de cada usuario.
2. Para cada PRI, contar la cantidad de días que la persona realiza llamadas desde esa zona en el horario indicado, ya sea residencial o de trabajo.
3. Seleccionar como hogar aquella PRI con mayor cantidad de días con llamadas en horario residencial.
4. Seleccionar como trabajo aquella PRI con mayor cantidad de días con llamadas en horario laboral, pero que es distinto al hogar.
5. Calcular DH y DT para ambas posiciones.
6. Si $DH > 0,5$ y $DT > 0,5$.
 - (a) Seleccionar trabajo como aquella PRI con mayor cantidad de días con llamadas en horario laboral.
 - (b) Seleccionar hogar como aquella PRI con mayor cantidad de días con llamadas en horario residencial, pero que es distinta al trabajo.
7. Si $DH > 0,3$ y $DT < 0,6$
 - (a) Seleccionar hogar como aquella PRI con mayor cantidad de días con llamadas en horario residencial.
 - (b) Seleccionar trabajo como aquella PRI con mayor cantidad de días con llamadas en horario laboral.
8. Si $DH > 0,6$ y $DT > 0,3$
 - (a) Seleccionar trabajo como aquella PRI con mayor cantidad de días con llamadas en horario laboral.
 - (b) Seleccionar hogar como aquella PRI con mayor cantidad de días con llamadas en horario residencial, pero que es distinta al trabajo.

Habiendo aplicado las reglas señaladas se obtiene una disminución de las desviaciones, pasando la DH a un valor de 38,13 % y la DT a un 34,87 %.

Para incluir en el perfilamiento la ubicación de las actividades diarias en los 6 sectores de la región, se calcula la pertenencia de las PRI a cada uno de ellos y se almacena esta información como variables *dummy*. La distribución para hogares y lugares de trabajo se muestra en las Tablas 3.5 y 3.6.

Tabla 3.5: Distribución hogar en sectores

Sector Hogar	Proporción
Oriente	19,1 %
Suroriente	19,7 %
Centro	6,0 %
Norte	12,2 %
Sur	22,6 %
Poniente	20,4 %

Tabla 3.6: Distribución trabajo en sectores

Sector Trabajo	Proporción
Oriente	24,1 %
Suroriente	14,5 %
Centro	13,8 %
Norte	12,1 %
Sur	18,6 %
Poniente	16,9 %

De acuerdo a lo obtenido la mayoría de los clientes viven en los sectores Sur y Poniente, mientras que una mínima parte vive en el sector Centro. Por otro lado, la mayor parte de los usuarios trabaja en el sector Oriente.

A partir de la información de densidad poblacional por comuna encontrada en [10] se calcula una aproximación de la proporción de la población que habita en cada uno de los sectores³. Los resultados de estos cálculos son presentados en la Tabla 3.7 y se utilizan para comparar con respecto a lo obtenido en este trabajo. Se observa que las diferencias más altas se encuentran en los sectores Sur y Oriente, lo que puede ser atribuido a los ingresos promedios de los sectores, pues el sector Oriente presenta el ingreso por hogar más alto, lo que puede impactar en que exista una mayor concentración de población con contrato en dicho lugar, mientras el sector Sur tiene un ingreso muy inferior, provocando el efecto contrario a pesar de ser el lugar más poblado de la región.

³Se hace a través de la asignación de cada comuna de la región a un sector. El detalle se puede ver en el Anexo D

Tabla 3.7: Proporción de población por sector

Sector	Proporción
Oriente	9,5 %
Norte	13,4 %
Centro	3,3 %
Poniente	22,3 %
Suroriente	23,3 %
Sur	28,1 %

Se puede comparar también lo obtenido con la base de lugares residenciales disponible. De dicha base se extrae lo que se presenta en la Tabla 3.8, que refleja una proporción ligeramente mayor de residencias en el sector Suroriente, seguido del sector Sur. Esto apunta al número de lugares residenciales y no personas que habitan en ellos, por lo que se puede entender la pequeña diferencia obtenida.

Tabla 3.8: Proporción de población por sector

Sector	Proporción
Oriente	13,8 %
Norte	13,1 %
Centro	9,2 %
Poniente	19 %
Suroriente	22,7 %
Sur	22,2 %

La caracterización de las actividades desarrolladas en las zonas de hogar y trabajo de los usuarios se realiza a través de la caracterización tempo-espacial de las celdas que las componen. En primer lugar se identifican las celdas de Voronoi que componen la PRI a caracterizar, ya sea el hogar o el trabajo. Luego, para cada una de las celdas se obtiene el promedio de la caracterización tempo-espacial en el horario deseado completo, es decir, por cada categoría se calcula el promedio en el horario completo y luego se normalizan. De la misma forma, la caracterización de la PRI completa corresponde al promedio normalizado de las zonas que la componen.

Cabe destacar que para caracterizar el lugar identificado como trabajo se elimina la categoría Residencial dada su alta presencia en la mayoría de las zonas, lo que crea una distorsión en la interpretación de los resultados.

En la Tabla 3.9 se ejemplifica como queda la información descrita para un usuario cualquiera. En el caso del hogar, la categoría de actividad predominante en el horario establecido es Residencial, pues representa en promedio un 64 % de las actividades desarrolladas en la zona en el bloque temporal, mientras que para el trabajo la categoría predominante es Oficina con un 54 %. Los valores que se muestran en la tabla se pueden interpretar como que existe una probabilidad de un 54 % de que el usuario tenga un trabajo que se desempeña en oficina y solo un 25 % de que sea en comercio. De todas formas, se debe recordar que

esta caracterización es un atributo de la posición y no del usuario e indica la proporción de actividades de cada categoría en el interior de la zona de acuerdo al uso de tiempo y suelo de la misma.

Tabla 3.9: Ejemplo caracterización para hogar y trabajo

Categoría	Hogar	Trabajo
Residencial	0,64	–
Entretenimiento	0,01	0,1
Comercio	0,1	0,25
Educación	0,2	0,03
Oficina	0,03	0,54
Industrial	0,02	0,08

En promedio los hogares y trabajos se distribuyen en actividades como se muestra en la Tabla 3.10. Se aprecia que los hogares presentan una clara caracterización residencial, mientras el trabajo resulta un poco más dividido, siendo sus categorías principales el trabajo de oficina, el comercio y educación.

Tabla 3.10: Estadísticos caracterización actividades hogar y trabajo

		Media	Desviación	Mínimo	Máximo
Hogar	Residencial	0,958	0,127	0	1
	Entretenimiento	0,002	0,009	0	0,379
	Comercio	0,015	0,029	0	0,654
	Educación	0,002	0,008	0	0,238
	Oficina	0,007	0,030	0	1
	Industrial	0,001	0,009	0	0,333
Trabajo	Entretenimiento	0,030	0,094	0	1
	Comercio	0,152	0,249	0	1
	Educación	0,151	0,225	0	1
	Oficina	0,485	0,286	0	1
	Industrial	0,073	0,106	0	0,333

Limitaciones

Una limitación se presenta en el caso de los usuarios con horarios muy distintos a los comunes, por ejemplo, para personas que trabajan durante la noche o el fin de semana probablemente se identificarán erróneamente sus actividades diarias.

Importante es destacar que, si bien gran parte de las construcciones pueden ser fácilmente catalogadas dentro de las categorías escogidas para este trabajo, algunas pueden ser ambiguas. Un ejemplo de esto es el caso en que existen edificios de oficinas en los que se atienden personas, pues esto puede señalarse como una actividad de trabajo de oficina o una actividad de comercio si fuera el caso.

3.5.2. Actividades Esporádicas

Se desea construir una métrica que refleje qué tan frecuente es para los usuarios participar de actividades que escapen de las habituales en determinados horarios. Se decide enfocar el estudio en horarios en que se cree las personas son más proclives a romper su cotidianidad.

El enfoque seleccionado tiene la ventaja de que no es necesario conocer el objetivo con que una persona sale de su habitualidad ni aplicarlo a una gran cantidad de lugares u horarios distintos para poder llegar a una conclusión respecto a la tendencia de una persona a salir de la rutina.

Los horarios escogidos para realizar el análisis son los que se consideraron ligados al almuerzo, ya sea en la semana o durante el fin de semana, el entretenimiento nocturno en los fines de semana y las salidas durante el fin de semana (quitando los horarios destinados a las actividades antes explicadas). Los horarios se muestran en la Tabla 3.11

Tabla 3.11: Horarios actividades esporádicas

Actividad	Hora Inicio	Hora Fin	Días Considerados
Almuerzo	13	16	Lunes -Viernes
Almuerzo	13	16	Sábado - Domingo
Entretenimiento Nocturno	21	6	Lunes -Viernes
Entretenimiento Nocturno	21	6	Sábado - Domingo
Salida Fin de Semana	6	13	Sábado - Domingo
	16	21	Sábado - Domingo

Se establece que la métrica a constuir debe medir la proporción de días en que la persona presenta entre sus registros un comportamiento distinto al habitual en el horario en que se mide, es decir, se toma la proporción de días en que una persona -durante el horario seleccionado- registró llamadas fuera de las celdas de Voronoi que conforman las Posiciones Relevantes identificadas como su Hogar o Trabajo, sobre el total de días analizados⁴.

Los estadísticos básicos de las métricas de actividades esporádicas obtenidas se presentan en la Tabla 3.12

Tabla 3.12: Estadísticos actividades esporádicas

Estadístico	Nocturno Sem	Nocturno FDS	Almuerzo Sem	Almuerzo FDS	Fin de Semana
Media	0,076	0,109	0,141	0,133	0,247
Desviación	0,106	0,141	0,148	0,130	0,190
Máximo	0,981	0,958	0,906	0,792	0,833
Mínimo	0	0	0	0	0

Se observa que de acuerdo a lo obtenido, el horario de fin de semana es el más utilizado por los usuarios para realizar actividades no habituales, mientras el horario de entretenimiento nocturno durante la semana registra una proporción mucho menor.

⁴En caso de que la actividad sea realizada solo durante fin de semana, entonces el total de días analizados es el total de días de fin de semana del período

Limitaciones

Una primera limitación de este procedimiento recae en la definición de las actividades a medir, pues si bien se puede considerar que existen horarios en que comúnmente se realizan ciertas acciones, esto no es cierto para la totalidad de la población, por lo que se puede estar midiendo algo distinto a lo que se pretende.

Un ejemplo de la limitación expuesta es el caso de las personas que se realizan compras durante los fines de semana en la noche. Al mirar su medida de entretenimiento nocturno se podría pensar que es alguien que pasa gran parte de su tiempo en actividades como fiestas, pubs u otros, mientras se encuentra en realidad en una acción distinta. Por esto, es necesario interpretar correctamente la métrica y no como que el individuo estuviese realizando una actividad específica asociada al horario.

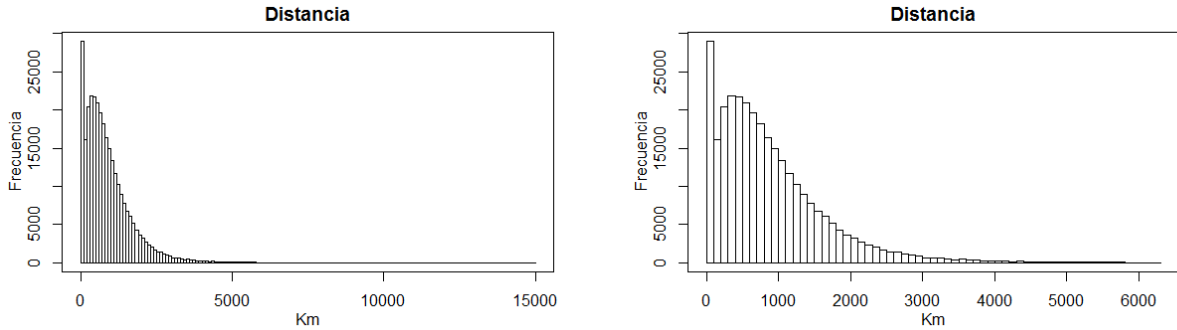
Cabe destacar que lo que aquí se presenta respecto a las actividades esporádicas es lo que se puede obtener a través de los registros telónicos, es decir, si una persona realiza con alta frecuencia actividades fuera de su rutina en uno de los horarios seleccionados, pero no es asidua a utilizar su teléfono móvil en dichos momentos, entonces erróneamente aparece como alguien que se apega a su rutina diaria durante los bloques analizados.

3.5.3. Cálculo de Métricas de Movimiento

Distancia Recorrida por el Usuario

Se calcula esta métrica con el total de registros de llamadas en el período de análisis, considerando el orden temporal y asumiendo que los movimientos de un punto a otro se realizan en línea recta. No se hace en base a Posiciones Relevantes de ningún tipo, dado que en un bloque horario se puede encontrar más de una de ellas para un usuario, que abarque más de una posición específica (o antena), lo que agregaría un mayor sesgo a los cálculos.

Una vez obtenida la variable, se extrae que en la mayoría de los casos toma un valor inferior a 6.000 kms. recorridos por el usuario en el total el período -equivalentes a 79 kms. diarios-, pero existen datos cercanos hasta a los 15.000 kms (197 kms diarios). Se determina que el valor máximo aceptado correspondería a la media más 7 desviaciones estándar, es decir, 6.422 kms (85 kms. diarios). Los histogramas de la variable original y tras la aplicación del tratamiento se muestran en la Figura 3.17.



(a) Distancia recorrida original

(b) Distancia recorrida tras eliminación outliers

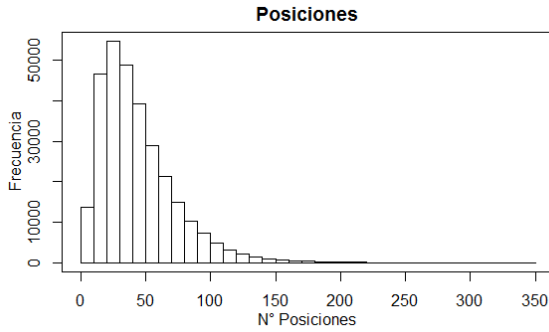
Figura 3.17: Histogramas distancia recorrida

Número de Locaciones Diferentes

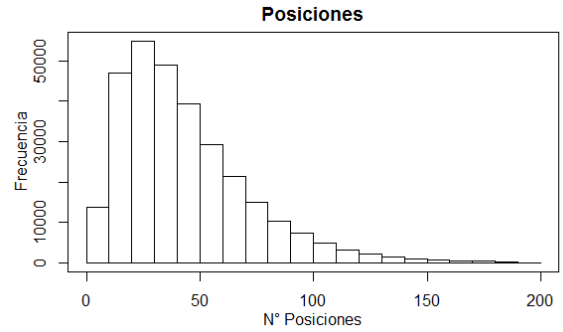
Se calcula como el número de antenas -o regiones de Voronoi- distintas desde donde se tiene registro de llamadas del usuario durante el período. Esto tiene una importante limitación en el caso de quienes mientras se desplazan realizan o reciben llamadas, pues a pesar de no encontrarse en un punto de relevancia o en el que el usuario desarrolla una actividad, se registra como una posición distinta.

Si el cálculo se realiza con las Posiciones Relevantes es posible que se omita información, pues éstas tienen un radio de cobertura bastante amplio, por lo que una Posición Relevante puede abarcar más de un punto que el usuario frecuente. Un ejemplo de esto se da en personas cuyo hogar se ubica cerca de un lugar en que realizan una actividad recurrente, tales como casas de familiares, comercio, gimnasio y otros. Ambas locaciones pueden corresponder a celdas de Voronoi distintas y por lo tanto que sus llamadas con alta probabilidad sean registradas por las antenas de forma diferenciada, pero de todos modos pertenecer a la misma Posición Relevante.

Una vez generada la métrica, se observa que la mayoría de los usuarios presentan hasta cerca de 200 posiciones, pero existen algunos que las superan llegando hasta sobre las 300, tal como se muestra en la Figura 3.18(a). Se realiza el mismo análisis de desviaciones mostrado en la sección 3.2.1 para concluir que se debían eliminar aquellos usuarios con un número de posiciones mayor a la media más 5 desviaciones, es decir, 190 locaciones. Aplicando esta regla se obtiene una distribución de la métrica como se muestra en la Figura 3.18(b).



(a) Número posiciones original



(b) Número posiciones tras eliminación outliers

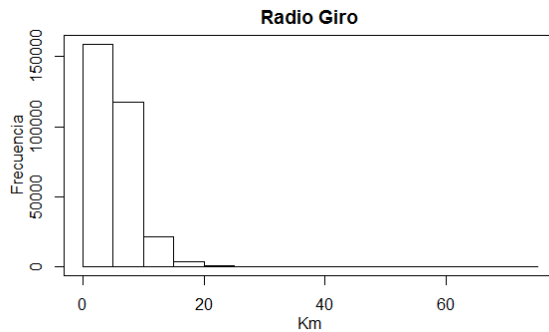
Figura 3.18: Histogramas número posiciones

Radio de Giro

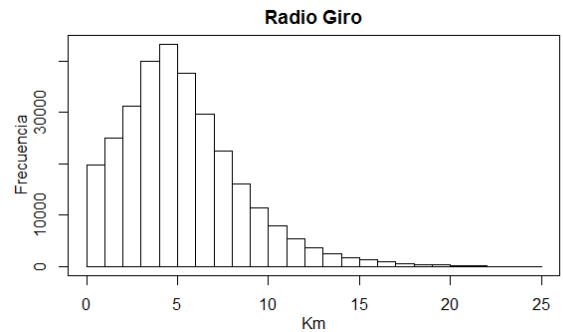
Tras el cálculo de la variable se observa que los datos se concentran en valores menores a 25 kms, mientras el máximo registrado es ligeramente mayor a los 70 kms. Se determina utilizar como límite superior 22,62 kms, equivalente a 5 desviaciones por sobre la media.

Existen personas cuyas llamadas son captadas por solo una antena, por lo que su radio de giro toma el valor 0, pero esto representaría que la persona se encuentra en la posición de la antena y que no tiene movimientos alrededor. Se decide por lo tanto reemplazar los radios de giro nulos por el valor mínimo de 0,8 kms, ya que este es el radio de cobertura promedio de las antenas.

En la Figura 3.19 se muestra el histograma de la variable original y cómo queda este tras aplicar los tratamientos para valores fuera de rango y perdidos mencionados.



(a) Radio de giro original



(b) Radio de giro tras eliminación outliers

Figura 3.19: Histogramas radio giro

Grado de Retorno

Esta métrica se calcula como la proporción de días promedio distintos en que el usuario, en la misma ventana de tiempo, pero días diferentes, se encuentra en el mismo lugar, diferenciando entre días de la semana y fin de semana debido a la diferencia de comportamiento que puede existir entre estos. Es decir, en la medición para la semana se sigue el procedimiento a continuación⁵:

1. $\forall i \in \{0, \dots, 23\}$ se cuentan las apariciones de cada PRD registradas en los bloques horarios $i, i + 24, i + 48, i + 72$ e $i + 96$.
2. Para cada i se toma el mayor valor de los calculados en el paso anterior, obteniéndose un GR para cada bloque de un día promedio de la semana (GR_i).
3. Se promedian los GR_i y se divide por la cantidad de días de la semana (5).

Los histogramas de lo obtenido para el Grado de Retorno en la semana y fin de semana se muestran en la Figura 3.20.

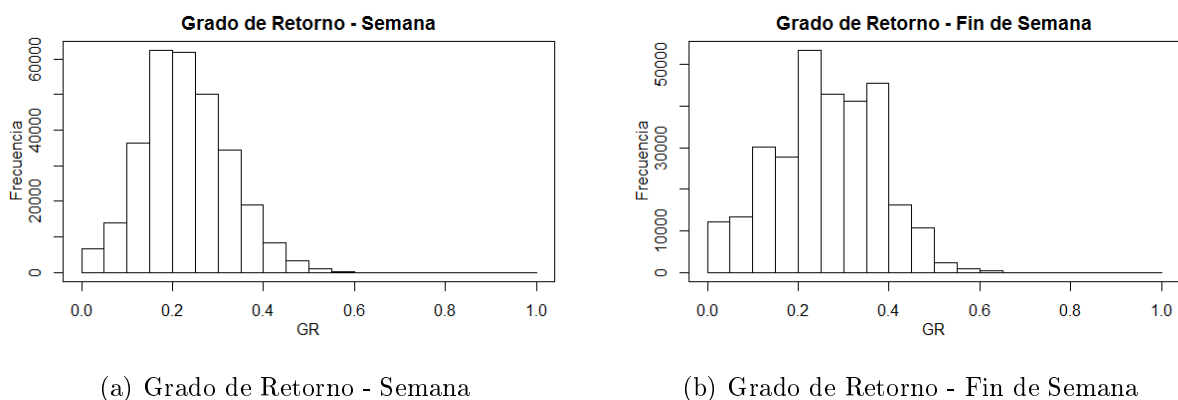


Figura 3.20: Histogramas Grado de Retorno

Cabe destacar que existe una amplia gama de otras posibles métricas que serían interesantes de incorporar en el futuro al análisis. Ejemplos de esto pueden ser métricas que expresen los momentos del día o la semana que reflejan mayor movimiento por parte del usuario, o que midan cuánto tiempo pasa en promedio una persona en los lugares que visita, o que asocien la forma de movimiento a figuras geométricas como elipses, círculos u otros (por ejemplo medir la excentricidad). Se propone que esto se considere para próximos trabajos, pues en esta etapa se prefirió tener primero claridad del aporte de medir el movimiento como tal al proyecto y al conocimiento del usuario.

En la Tabla 3.13 se muestran estadísticos descriptivos de la métricas obtenidas. En ella se observa una alta dispersión entre los usuarios para las métricas de número de posiciones, distancia recorrida y radio de giro, mientras que una mucho menor en los grados de retorno.

⁵Análogo para los fines de semana

Estadístico	Posiciones	Radio Giro	Distancia	GR Semana	GR FDS
Media	43,8	5,3	877	0,23	0,27
Desviación	28,2	3,2	760,4	0,10	0,12
Máximo	190	22,6	6.420	0,69	0,83
Mínimo	1	0,8	0	0	0

Tabla 3.13: Estadísticos métricas de movimiento

La distancia recorrida se presenta como total en el período, pero cabe notar que llevándola a un equivalente diario el promedio es de 10,5 kms. y el máximo alcanza los 190 kms.

Tiene sentido que el mínimo de posiciones registradas no sea 0, ya que dicho caso significaría que la persona no realiza ni recibió ninguna llamada durante el período y dichas personas fueron excluidas del análisis.

3.5.4. Identificación de Patrones de Movimientos Grupales

Utilizando las Posiciones Relevantes obtenidas en ventanas de tiempo de los usuarios se construye una matriz de transición simétrica⁶ entre los 6 sectores de la RM, posteriormente se aplica el algoritmo de clusterización k-medias sobre la matriz normalizada a través de max-min⁷.

Para encontrar el número apropiado de grupos de personas se obtuvieron los indicadores de Dunn y Davies Bouldin, con los que se determina utilizar 12 clusters. Los valores de los índices por número de grupos se presentan en la Figura 3.21.

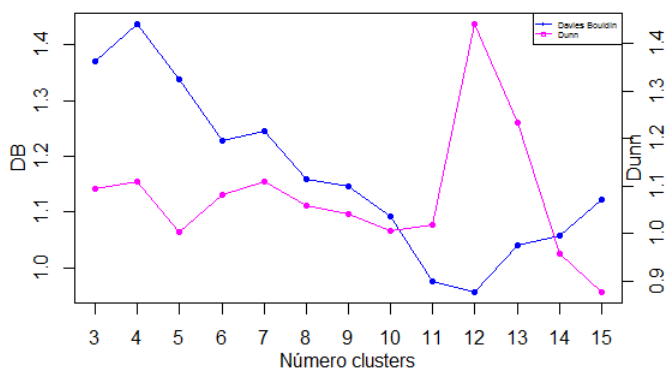


Figura 3.21: Índices de Dunn y Davies Bouldin para patrones grupales

En la Tabla 3.14 se pueden ver las medias de los grupos, mientras en la Tabla 3.15 se muestra el tamaño relativo de cada cluster.

⁶Se probaron dos tipos de matrices de transición, una simétrica y otra asimétrica, obteniéndose resultados muy similares

⁷ $\frac{x_i - \min(x)}{\max(x) - \min(x)}$

Tabla 3.14: Medias clusters de movimientos similares

Transición	Gr.1	Gr.2	Gr.3	Gr.4	Gr.5	Gr.6	Gr.7	Gr.8	Gr.9	Gr.10	Gr.11	Gr.12
Oriente-Suroriente	0,02	10,26	0,62	0,57	0,02	0,29	0,01	0,37	0,22	0,43	1,34	0,01
Oriente-Norte	0,92	0,19	0,39	0,15	0,10	8,07	0,42	0,25	0,22	0,00	0,04	0,01
Oriente-Sur	0,01	0,23	0,25	7,43	0,02	0,11	0,55	0,30	0,13	0,40	0,03	0,27
Oriente-Centro	0,53	0,52	8,29	0,58	0,39	0,85	0,06	0,30	0,86	0,03	0,82	0,03
Oriente-Poniente	0,02	0,17	0,33	0,23	0,58	0,28	0,02	0,26	8,01	0,01	0,02	0,21
Suroriente-Norte	0,22	0,22	0,01	0,01	0,02	0,20	0,16	0,43	0,01	0,14	0,29	0,00
Suroriente-Sur	0,01	0,51	0,02	0,79	0,01	0,02	0,48	0,41	0,00	9,00	0,99	0,27
Suroriente-Centro	0,08	0,59	0,29	0,03	0,09	0,01	0,01	0,23	0,02	0,34	7,04	0,01
Suroriente-Poniente	0,01	0,30	0,02	0,01	0,22	0,01	0,01	0,56	0,22	0,26	0,41	0,13
Norte-Sur	0,32	0,00	0,01	0,22	0,06	0,18	6,35	0,20	0,01	0,15	0,02	0,24
Norte-Centro	7,56	0,03	0,29	0,02	0,42	1,02	0,82	0,24	0,02	0,01	0,20	0,02
Norte-Poniente	1,02	0,01	0,03	0,00	3,10	0,41	0,64	0,34	0,36	0,01	0,01	0,21
Sur-Centro	0,12	0,02	0,26	0,80	0,18	0,02	0,73	1,17	0,02	0,46	0,40	0,60
Sur-Poniente	0,02	0,02	0,02	0,55	0,47	0,01	0,72	0,42	0,37	0,35	0,03	8,23
Centro-Poniente	0,36	0,01	0,32	0,02	5,52	0,04	0,05	0,26	1,08	0,01	0,21	0,46

Los resultados muestran que cada grupo está principalmente asociado a una dirección de movimiento, a excepción del número 5 asociado a 2 direcciones y del grupo 8 que, si bien sus transiciones se presentan más altas para la combinación Sur-Centro, el valor de las mismas es bajo en comparación a lo visto en el resto de los grupos. Al analizar este último cluster se observa que además de aquellos con movimientos Sur-Centro, en este grupo se incluye una gran cantidad de usuarios que no presentan transiciones, lo que se debe probablemente por el bajo promedio de movimientos entre las zonas Sur-Centro.

Tabla 3.15: Tamaños relativos clusters de movimientos similares

Grupo	Transición Principal	Proporción
Gr. 1	Norte - Centro	4,5 %
Gr. 2	Oriente - Suroriente	9,2 %
Gr. 3	Oriente - Centro	5,5 %
Gr. 4	Oriente - Sur	10,4 %
Gr. 5	Norte - Poniente , Centro - Poniente	4,4 %
Gr. 6	Oriente - Norte	2,5 %
Gr. 7	Norte - Sur	4,9 %
Gr. 8	Sur - Centro o Sin Transición	35,6 %
Gr. 9	Oriente - Poniente	6,0 %
Gr. 10	Suroriente - Sur	3,4 %
Gr. 11	Suroriente - Centro	6,2 %
Gr. 12	Sur - Poniente	7,2 %

A partir de la información de densidad poblacional mostrada en la Tabla 3.7, se sabe que el sector más habitado es el Sur, lo que junto a la incorporación de casos que no reportan movimiento entre sectores ayudan a explicar la alta proporción del grupo 8 (Sur-Centro).

Se espera que los resultados aquí obtenidos guarden estrecha relación con la ubicación de hogar y trabajo de los clientes analizados, pues como se ha mencionado en repetidas ocasiones estas son las principales actividades promotoras de movimiento y se puede ver en esta sección que los usuarios realizan una transición en particular con mayor frecuencia que cualquier otra, la que debiese estar ligada al desplazamiento entre dichas actividades. Se analiza esta

coherencia en la descripción de los perfiles que se obtienen en secciones posteriores.

Los patrones obtenidos son una variable descriptiva del movimiento de los usuarios a agregar en el análisis final que se realiza, pero pueden ser utilizados además por sí solos para el proyecto de movilidad, pues señalan los movimientos más comunes de un cliente, por lo que se puede enfocar de mejor manera la comunicación que se envíe.

Capítulo 4

Perfilamiento

4.1. Identificación de Perfiles

Se aplica el algoritmo k-medias sobre las variables construidas para identificar perfiles de usuarios. Para encontrar el número apropiado de grupos se obtienen los índices de Dunn y Davies Bouldin. El gráfico de éstos se muestra en la Figura 4.1, donde se puede ver que se un total de 10 grupos obtiene la combinación entre el mayor valor de Dunn y el menor de Davies Bouldin. También existe la opción de continuar con 5 perfiles, pues para este número de grupos los índices marcan *peaks* interesantes, pero tras las revisión de ambas posibilidades, se decidió que al continuar con 10 perfiles existe un mayor potencial desde el punto de vista de negocios para el proyecto.

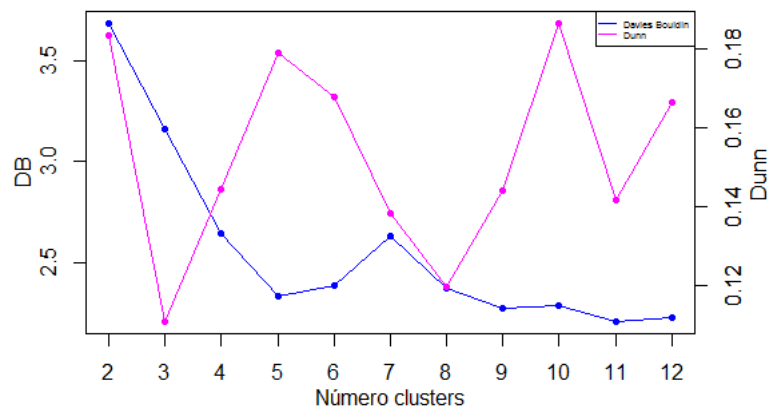


Figura 4.1: Índices DB y Dunn para perfiles

Los tamaños relativos de los perfiles obtenidos se muestran en la Tabla 4.1 y las medias de las variables por cada perfil se encuentran en el Anexo G.

Tabla 4.1: Tamaño relativo perfiles

Perfil	Proporción
1	6,2 %
2	13,4 %
3	6,6 %
4	10,7 %
5	12,4 %
6	9,9 %
7	3,7 %
8	15,6 %
9	13,8 %
10	9,1 %

4.2. Perfiles de clientes

En base a los resultados, se construyen las descripciones de los perfiles que se presentan a continuación.

De manera general, los grupos obtenidos son integrados por personas de toda edad, género y grupo socioeconómico, a excepción de aquellos en que se mencione alguna característica especial para estos atributos. Lo mismo sucede con la caracterización tempo-espacial del hogar y el trabajo, pues en su mayoría el primero es primordialmente residencial y el segundo cuenta con alta presencia de trabajos de oficina (sobre 45 %)¹.

Grupo 1

Presentan uno de los más altos tráficos de voz, datos y mensajería, un número de recambios y una antigüedad de los equipos entorno a la media y un 37% de presencia de smartphones, lo que lleva a un ARPU superior a los \$14.700, es decir, también por encima de la media.

Su radio de giro promedio es el segundo más alto (5,9 kms) acompañado del más alto número de posiciones registradas (53) con una distancia recorrida (12,9 kms. diarios) superior al promedio. Por el contrario, su GR en la semana está ubicado claramente bajo la media de la variable.

Son personas cuyo trabajo se encuentra en el sector Oriente (98 %) de la región, mientras los lugares de residencia están divididos entre los sectores Suroriente (58 %), Norte (21 %) y Centro, lo que explica su pertenencia a los patrones grupales de movimiento que reflejan transiciones entre Oriente y Suroriente(53 %), Oriente y Norte (20 %) y Oriente y Centro (17%).

Sus actividades esporádicas registradas se encuentran por lo menos en un 15% sobre la media en todos los horarios a excepción del almuerzo en los días de semana.

¹Cualquier característica que no sea mencionada de manera explícita en un perfil debe ser considerada como promedio

Grupo 2

Presentan un tráfico de datos y voz ligeramente por sobre el promedio que los lleva a tener un ARPU con las mismas características equivalente a \$14.200 aproximadamente.

Su radio de giro se encuentra levemente bajo la media al igual que el número de posiciones que registran en promedio. La distancia recorrida diariamente por los miembros de este perfil es de 11,8 kms., es decir, a pesar de tener un área de desplazamiento habitual menor se mueven bastante dentro de ella.

Trabajan en el sector Norte de la región (98 %) y residen principalmente en él también (71 %). Otra porción de las viviendas está ubicada en los sectores Oriente (11 %) y Suroriente (9 %) en mayor medida. Los patrones de movimiento a los que pertenecen reflejan que no realizan transiciones (45 %) o lo hacen entre el Norte y el Centro (24 %) o el Norte y el Oriente (20 %). La porción de personas que se trasladan hacia el Centro es mayor que la que este trabajo puede explicar a través de las actividades diarias identificadas.

Grupo 3

Tráfico de datos y de voz ligeramente sobre el promedio y un ARPU promedio sobre los \$14.200.

La distancia que recorren diariamente es en promedio de 9,5 kms, su radio de giro de 4,65 kms y registran 38 posiciones, todas ellas las terceras más bajas entre los perfiles. Sus grados de retorno se encuentran entorno a las medias.

El 99 % de los usuarios con este perfil trabajan en el sector Suroriente de la región y un 78 % reside también en él, mientras un 14 % lo hace en el sector Oriente. Un 51 % pertenece al patrón que incluye a quienes no realizan transiciones y un 27 % al de quienes se mueven entre el Oriente y el Suroriente, es decir, existen una proporción de personas que viven en el sector Suroriente que se trasladan hasta el Oriente con motivos desconocidos para este trabajo.

Las actividades esporádicas se presentan superiores a la media en todos los horarios, especialmente en el caso del entretenimiento nocturno los fines de semana (13 % sobre) y el almuerzo durante la semana (17 % sobre).

Este es el único grupo en que la categoría "trabajo de oficina" no se impone con sobre un 45 % de las actividades de la zona de trabajo, si no que tiene además una alta presencia de actividades de comercio (26 %) y educación (25 %).

Grupo 4

Son personas cuyo tráfico de datos es bastante superior a la media, a pesar de no ser el más alto entre los perfiles. Su tráfico de voz en cambio está ubicado justo en la media y son quienes más mensajes de texto envían y reciben, aunque sigue siendo números bajos -8 de salida y 5 de entrada al mes-. Son el segundo grupo más antiguo en la empresa, poseen equipos más nuevos que la media y han realizado más recambios. Su proporción de smartphones es la segunda más alta con un 39 %. El ARPU de este grupo se ubica sobre la media, llegando

a superar los \$14.650.

Desde el punto de vista de su movilidad, presentan un radio de giro (5,23 kms.) y una distancia recorrida (10,5 km. diarios) en el promedio, pero un alto número de posiciones registradas que alcanza las 52 en promedio. Esto último es la mayor densidad de posiciones en el área en la que normalmente se desenvuelve entre todos los grupos.

Un 100 % trabaja en el sector Centro y sus residencias se reparten entre prácticamente todos los sectores de la siguiente forma:

- 28 % Oeste
- 21 % Oriente
- 20 % Suroriente
- 15 % Norte
- 14 % Centro
- 2 % Sur

De la misma forma su pertenencia a los distintos patrones grupales de movimiento se encuentra dividida como sigue:

- 27 % Norte - Oeste y Centro - Oeste
- 20 % Oriente - Centro
- 17 % Suroriente - Centro
- 15 % Sur - Centro o sin transiciones
- 14 % Norte - Centro
- 7 % Otros

Sus actividades esporádicas se muestran en general de la media hacia arriba a excepción del almuerzo durante la semana, cuyo valor (0,107) es el menor entre todos los perfiles.

Grupo 5

Este grupo está integrando en un 100 % por mujeres de toda edad y grupo socioeconómico.

Presentan el menor tráfico de datos entre los perfiles, pero un tráfico de voz ligeramente superior a la media. Esto está acompañado de una de las mayores antigüedades promedio de los equipos y una proporción de teléfonos inteligentes de 29 %, la segunda más baja. Su ARPU es ligeramente superior a los \$13.400, es decir, menor al promedio.

Su movilidad no se escapa de los rangos promedio a excepción de la distancia recorrida que alcanza los 13,2 kms. diarios, superior a la media de esta variable.

Un 85 % de los clientes que integran este grupo reside en el sector Sur de la región y un 59 % trabaja también en esta zona. El resto de los integrantes trabaja principalmente en los sectores Oriente (12 %) y Centro (11 %). Esta información es coherente con que un 44 % de los miembros pertenece al grupo de personas que transita mayormente entre los sectores Sur y Centro o que no realiza transiciones y un 20 % lo hace entre los sectores Sur y Oriente.

Grupo 6

Usuarios pertenecientes principalmente a los segmentos ABC1 (47 %) y C2 (37 %) cuyo promedio de edad es el mayor entre todos los grupos obtenidos, sobrepasando los 50 años. Si bien son personas de todo género, presentan una proporción mayor de mujeres, la que alcanza un 46 % del grupo.

Desde el punto de vista transaccional, tienen la mayor cantidad de minutos registrados tanto de salida como de entrada con una gran diferencia con respecto al grupo que les sigue, son quienes presentan el mayor número de recambios de equipo y un 49 % de los usuarios cuenta con smartphone. Su ARPU alcanza un promedio por sobre los \$17.000 y presentan la mayor antigüedad como clientes de la compañía (sobre los 5 años).

Respecto a las características de su movilidad, son quienes recorren menores distancias a cerca de 8,5 kms. diarios en promedio, poseen el segundo menor radio de giro promedio y registran una de las mayores cantidades de posiciones diferentes. Esto es coherente con el ARPU del grupo, pues quienes mayores posiciones registran recorriendo menores distancias en un radio pequeño presentan mayores ingresos en promedio para la compañía. Por el contrario, sus grados de retorno promedio son los menores entre los grupos obtenidos, pero al moverse esta variable en rangos pequeños no afecta en gran medida.

Un 69 % de los integrantes pertenece al grupo 8 de los patrones de movimiento, es decir, realizan transiciones especialmente entre los sectores Sur-Centro o no las hacen. Esto último es lo que tiene mayor coherencia con el hecho de que 93 % de los miembros viven en el sector Oriente y 95 % trabajan en el mismo. Todas estas características tienen una relación positiva con el ARPU de acuerdo a lo obtenido en el Capítulo anterior.

Grupo 7

Personas de género masculino (56 %) y desconocido (44 %). Se puede pensar a partir de esto que aquellos usuarios de género desconocido que integren este grupo son con alta probabilidad hombres, dado su comportamiento transaccional y de movilidad. Tienen una ligera tendencia hacia estratos socioeconómicos más bajos.

Su tráfico de voz es el menor entre todos los perfiles y su nivel de navegación es uno de los más bajos. Presentan el menor número de recambios en promedio y la menor proporción de teléfonos inteligentes (7 %). Todo esto deriva en que el ARPU promedio en este grupo sea cercano a los \$12.400, es decir, el más bajo.

Su distancia recorrida diaria es cercana a los 10 kms., es decir, si bien no es la más baja es bastante menor al promedio. Presenta el menor radio de giro entre los perfiles ligeramente superior a los 4 kms.- y un número de posiciones totales de 28, es decir, también la menor. Esta baja movilidad está ligada a su pertenencia en un 100 % al patrón de movimiento que integra a quienes no realizan transiciones entre sectores o a quienes se mueven entre el Sur y el Centro de la región, lo que es coherente con el hecho de que 92 % de los miembros del grupo residen en el sector Sur, un 76 % trabaja en el mismo sector y cerca de un 20 % se traslada al Centro a trabajar.

Presentan el más bajo registro de actividades no habituales en horarios nocturnos y en fin de semana completo. En el caso de los horarios de almuerzo este índice es también uno de los más bajos.

Grupo 8

Integrado por personas de género masculino (65 %) y desconocido (35 %). Al igual que para el grupo anterior, se puede pensar basados en el comportamiento transaccional y de movilidad de los integrantes, que aquellos usuarios de género desconocido son con alta probabilidad hombres.

Este grupo presenta un tráfico promedio de manera general y un ARPU entorno a los \$13.500, es decir, menor a la media de la variable.

Sus métricas de movilidad se encuentran entorno al promedio, a excepción del número de posiciones registradas que alcanza las 42 en promedio.

Viven principalmente en el sector Oeste (81 %) y gran parte trabaja también en él (72 %). El resto de los integrantes trabaja en el sector Oriente (10 %) o en el resto de la región (18 %) a excepción del sector Centro. Coherentemente con esto, el 36 % del grupo pertenece al patrón que recoge a aquellos usuarios que no realizan transiciones entre sectores y un 18 % se traslada principalmente entre Oeste y Oriente, pero existe un 30 % con movimientos Oeste-Norte y Centro - Oeste, proporción mucho mayor que lo explicado por las actividades integradas en el análisis.

El registro de actividades esporádicas de los miembros de este perfil es de la media hacia arriba para todos los horarios estudiados.

Grupo 9

Un 89 % de los usuarios de este grupo son mujeres y no se conoce el género del 11 % restante. Basados en su comportamiento transaccional y de movilidad se puede pensar que estos últimos son mujeres también.

Este grupo presenta un comportamiento transaccional promedio de manera general, destacando la mayor antigüedad de sus equipos. Su ARPU es cercano a los \$13.400.

Sus métricas de movilidad se encuentran entorno al promedio, a excepción del número de posiciones registradas que es superior y alcanza las 42. Un 85 % de los miembros vive en el sector Oeste de la región y los trabajos se reparten entre los sectores Oeste (57 %) y Oriente (27 %). Esto guarda relación con los patrones grupales a los que pertenecen, siendo éstos los que capturan transiciones Oeste-Oriente (34 %), Sur-Centro o son transiciones (31 %) y Norte-Oeste y Centro-Oeste (23 %). Este último patrón está presente a pesar de que los trabajos en sectores Norte y Centro sumados no sobrepasan el 7 %, por lo que se piensa existen razones que nos e explican en este trabajo para que las personas se movilizan hacia dichos lugares.

Grupo 10

Formado en un 59 % por hombres y un 41 % por personas de género desconocido. Basados en su comportamiento transaccional y de movilidad se puede pensar que los de género desconocido son hombres también.

Presenta el mayor tráfico de datos, pero su proporción de smartphones es de 33 %, lo que no destaca del promedio. Sus minutos de salida se encuentran por sobre el promedio, mientras los de entrada son menores. Son los clientes de menor antigüedad en la compañía. Su ARPU supera los \$14.200.

Su radio de giro y distancia recorrida son las más altas de los perfiles, siendo estas métricas de 7 kms. y 17 kms. diarios en promedio respectivamente. Su regularidad a la hora de visitar locaciones es la más alta para los fines de semana y una de las mayores en la semana.

La mayoría vive en el sector Sur (80 %), pero sus trabajos se encuentran menos concentrados, estando un 42 % ubicado en el mismo sector, un 23 % en el Oriente de la región y el otro 25 % dividido en el resto. De la misma forma su pertenencia a un determinado patrón de movimiento no es clara, pues un 30 % realiza transiciones entre sectores Sur y Oriente de forma predominante, un 29 % lo hace entre Sur y Suroriente y un 26 % entre Sur y Oeste. Cabe destacar que no hay quienes pertenezcan al patrón que captura a aquellos que no realizan transiciones a pesar de que se puede pensar que gran cantidad de miembros viven y trabajan en el mismo sector. De ser este último el caso, entonces dichas personas realizan movimientos hacia otros lugares por motivos que no son captados en este trabajo.

Los miembros con este perfil presentan los mayores registros de actividades fuera de las habituales en todos los horarios definidos, siendo especialmente altos en almuerzo en la semana y fin de semana completo.

Capítulo 5

Evaluación

Se desea que los perfiles obtenidos en este trabajo tengan una aplicación general para el proyecto de movilidad en tiempo real, pero para realizar comunicaciones específicas a los clientes posiblemente sea necesario complementar lo aquí hecho con otra información relevante diferenciada para cada ocasión. Por ejemplo, en el caso de que se desee comunicar sobre una promoción puntual en un centro comercial a un segmento con mayor probabilidad de tomar dicha promoción, se puede complementar la información que los perfiles otorgan con una métrica de asistencia a dicho centro comercial en función de los registros que hayan sido captados por antenas aledañas y, de esta manera, comunicar la promoción a quienes hayan visitado el lugar con cierta frecuencia en ocasiones anteriores.

Se piensa que para la evaluación de este trabajo es de mayor relevancia analizar cómo las métricas construidas aportan valor al proyecto. Se determina por lo tanto, medir cómo los distintos grupos de variables generados ayudan a explicar el valor que los clientes tienen para la compañía, lo que se realiza a través de regresiones lineales donde la variable objetivo es ARPU y los grupos de variables de movilidad las variables explicativas. Estos grupos se resumen en la Tabla 5.1.

Tabla 5.1: Grupos de variables

Grupo	Variables
Demográficas	Edad, GSE, Género
Métricas	Radio de giro, Número de posiciones, Distancia recorrida, Grado de retorno
Patrones Grupales	Pertenencia a cada cluster (11 variables dummy)
Sectores Hogar	Sectores de residencia (5 variables dummy)
Sectores Trabajo	Sectores de trabajo (5 variables dummy)
Caracterización	Caracterización tempo-espacial del lugar de residencia y trabajo
Esporádicas	Medidas de almuerzo, entretenimiento nocturno y salidas de fin de semana

Se probaron todas las combinaciones posibles de estos 7 grupos -127 en total-. Para evaluar las regresiones se utiliza R^2 y MAPE sobre el total de la muestra.

Se ejecutan las regresiones directamente sobre las 127 combinaciones de grupos de variables. Se obtiene como resultados que el mayor R^2 es de 0,2 y el menor MAPE es de 87,5%. Estos resultados y los que siguen se pueden ver en el Anexo F.

A partir de las regresiones sobre grupos de manera individual se puede ver que los que mayor poder explicativo tienen son las métricas de movilidad, las variables demográficas y el sector de residencia. Cabe destacar que las métricas de movilidad son capaces de explicar mejor el ARPU que cualquier otro grupo de variables, incluyendo las demográficas que consideran GSE.

Las métricas de movilidad con mayores coeficientes asociados son los grados de retorno en la semana y fin de semana, tal como se muestra en la Tabla 5.2, pero se debe recordar que estas variables se mueven entre 0 y 1, mientras el resto de las métricas toman valores mucho más altos, por lo que diferencias en el GR entre usuarios no tendrán un efecto tan grande en el ARPU. Cabe destacar que el efecto del radio de giro en la variable objetivo es negativo, mostrando que bajo igualdad en el resto de las características, aquellos usuarios cuyo desplazamiento abarca grandes zonas de la región presentan un ingreso menor para la compañía.

Tabla 5.2: Resultados regresión lineal métricas de movimiento

Variable	Coefficiente	Error Std.
Intercepto	8.111,6	51,2
Posiciones	112,9	0,7
Distancia	0,4	0,0
Radio de Giro	-226,1	7,0
Grado Retorno Sem	2.172,7	203,3
Grado Retorno FDS	5.898,4	170,3

En la Tabla 5.3 se muestran los coeficientes asociados a cada variable en la regresión aplicada al grupo demográfico. Se observa que el ser hombre tiene asociado un coeficiente bastante mayor al del atributo mujer. Para el caso del grupo socioeconómico, mientras más bajo sea el valor que toma la variable mayor será el ARPU, es decir, a mayor poder adquisitivo mayor gasto en telefonía. Además, se muestra que usuarios de mayor edad tienen un ARPU menor.

Tabla 5.3: Resultados regresión lineal variables demográficas

Variable	Coefficiente	Error Std.
Intercepto	19.701,4	80,9
Mujer	792,1	41,1
Hombre	1.335,2	41,4
Edad	-63,6	1,2
GSE	-1.151,9	15,6

Los resultados de la regresión hecha sobre la ubicación del hogar, mostrados en la Tabla 5.4, reflejan que vivir en los sectores Oriente y Centro tiene una alta relación con mayores valores de ARPU y que el sector Sur es el único en este caso que tiene una relación negativa con la variable objetivo.

Se escogen las 15 combinaciones cuyas regresiones presentan mejores resultados para continuar el análisis e intentar mejorar el ajuste de los modelos. Dichas combinaciones se muestran

Tabla 5.4: Resultados regresión lineal sectores hogar

Variable	Coefficiente	Error Std.
Intercepto	13.474,9	33,7
Oriente	3.724,4	51,3
Suroriente	193,6	48,5
Norte	468,2	56,2
Sur	-91,1	47,8
Centro	2.054,7	70,4
Poniente	0,0	0,0

en la Tabla 5.5. Se debe destacar que en todos los casos se encuentran presentes los grupos de variables demográficas, métricas de movimiento y sector de residencia, las que como se mencionó tuvieron el mejor ajuste de manera individual también. Es más, si se elimina el grupo de métricas de movimiento del mejor modelo obtenido hasta el momento, los valores obtenidos para los índices son 0,1 en el caso de R^2 y 94 % para el MAPE, lo que refleja nuevamente la importancia de este grupo de variables por sobre los otros para describir el ARPU.

Tabla 5.5: Resultados 15 mejores regresiones lineales - ARPU

Combinación de Variables							R2	MAPE
Demográficas	Métricas	Hogar					0,200	87,53
Demográficas	Métricas	Hogar	Trabajo				0,200	87,52
Demográficas	Métricas	Hogar	Caracterización				0,201	87,53
Demográficas	Métricas	Patrones	Hogar				0,201	87,65
Demográficas	Métricas	Patrones	Hogar	Trabajo			0,202	87,63
Demográficas	Métricas	Patrones	Hogar	Caracterización			0,202	87,65
Demográficas	Métricas	Hogar	Esporádicas				0,203	87,38
Demográficas	Métricas	Patrones	Caracterización	Hogar	Trabajo		0,203	87,62
Demográficas	Métricas	Hogar	Trabajo	Esporádicas			0,203	87,37
Demográficas	Métricas	Hogar	Caracterización	Esporádicas			0,204	87,36
Demográficas	Métricas	Hogar	Trabajo	Caracterización	Esporádicas		0,204	87,36
Demográficas	Métricas	Patrones	Hogar	Esporádicas			0,204	87,51
Demográficas	Métricas	Patrones	Hogar	Trabajo	Esporádicas		0,205	87,49
Demográficas	Métricas	Patrones	Caracterización	Hogar	Esporádicas		0,205	87,49
Demográficas	Métricas	Patrones	Caracterización	Hogar	Trabajo	Esporádicas	0,206	87,47

Tras varias pruebas sobre las 15 combinaciones elegidas, cuyos resultados se presentan en el Anexo F, se determina utilizar los 7 grupos de variables y transformar ARPU y edad a través de una función logarítmica. Además se prueban interacciones entre los grupos demográficos y de métricas de movimiento obteniéndose que mejoran los índices estudiados las que se se *alanacontinua* in :

- Edad/Distancia
- GSE/Distancia
- GSE/Posiciones

Con lo presentado se obtiene finalmente que la mejor regresión presenta un R^2 de 0,18 y un MAPE de 3,4 %. Los coeficientes asociados a dicha regresión se presentan en la Tabla 5.6. De estos valores destaca que la relación entre las variables demográficas se mantiene igual

que cuando se analiza a nivel individual, que la distancia recorrida pasa a ser un factor que disminuye la variable objetivo cuando de manera individual era prácticamente neutra y que, si bien todos los sectores de residencia presentan una relación positiva o neutra con el ARPU, quienes residen en los sectores Oriente y Centro presentan mayores valores para esta variable.

Tabla 5.6: Coeficientes regresión lineal definitiva

Variable	Coefficiente	Variable	Coefficiente
Intercepto	11,10	Hogar Oficina	-0,13
Mujer	0,06	Hogar Industrial	-0,08
Hombre	0,07	Trabajo Entretenimiento	-0,04
Log(Edad)	-0,86	Trabajo Comercio	-0,06
GSE	-0,04	Trabajo Educación	-0,02
Log(Posiciones)	0,22	Trabajo Oficina	-0,07
Log(Distancia)	-0,03	Trabajo Industrial	-0,04
Log(Radio de Giro)	-0,03	Hogar Oriente	0,18
GR Semana	0,35	Hogar Suroriente	0,02
GR Fin de Semana	0,61	Hogar Norte	0,03
Norte - Centro	-0,02	Hogar Sur	0,01
Oriente - Suroriente	-0,02	Hogar Centro	0,13
Oriente - Centro	0,02	Trabajo Oriente	0,00
Oriente - Sur	-0,03	Trabajo Suroriente	-0,01
Norte - Poniente, Centro - Poniente	-0,01	Trabajo Norte	0,01
Oriente - Norte	0,00	Trabajo Suroriente	0,02
Norte - Sur	-0,02	Trabajo Centro	-0,02
Sur - Centro o Sin Transición	0,03	Ent. Nocturno Sem.	0,28
Oriente - Poniente	-0,03	Ent. Nocturno FDS	-0,02
Suroriente - Sur	-0,01	Almuerzo Sem.	0,03
Suroriente - Centro	-0,01	Almuerzo FDS	0,10
Hogar Residencial	-0,07	Fin de Semana	-0,11
Hogar Entretenimiento	-0,11	Edad/Distancia	0,00
Hogar Comercio	-0,28	GSE/Distancia	0,00
Hogar Educación	0,09	GSE/Posiciones	0,00

Se concluye que las variables generadas que mayor poder explicativo sobre el ARPU presentan son las demográficas, las métricas de movimiento y el sector de residencia, que en conjunto obtienen un R^2 de 0,168, mientras utilizando el total de variables -menos las interacciones- el R^2 llega a 0,171. Cabe destacar que esto no implica necesariamente que el resto de los grupos tengan un menor valor para el proyecto.

Si bien el valor de R^2 es bastante bajo, tiene lógica al pensar que no se incluye en este análisis la información transaccional que da origen a la variable objetivo. Si se aplica una regresión lineal a las variables de tráfico transformadas a través de una función logaritmo para explicar el $\log(\text{ARPU})$ se obtiene un R^2 de 0,42, lo que mejora a 0,49 al agregar la información de movilidad.

Capítulo 6

Conclusiones

6.1. Conclusiones del Trabajo

El objetivo general de este trabajo de título era construir y caracterizar perfiles de clientes de telefonía móvil en base a su movilidad, siendo ésta identificada a partir del análisis del registro de sus llamadas. Además, existían una serie de objetivos específicos relacionados a la definición y generación de información de movilidad, además de la validación de la relevancia de éstos a la hora de explicar el valor del cliente para la compañía, específicamente el ARPU.

Considerando lo rescatado de la literatura relacionada se definen variables relevantes para explicar la movilidad de los usuarios, construyendo y adaptando métricas que permitan captar dicha información en el contexto en que este trabajo de título se envuelve. Las formas de cálculo de las mismas son simples para de esta forma poder ser replicadas en el tiempo de manera rápida y efectiva para su inclusión en el proyecto de movilidad de la empresa.

La información de movilidad generada es de distintos tipos para capturar diferentes ámbitos de los movimientos de los usuarios, lo que incluye la ubicación y caracterización tempoespacial de sus actividades diarias, una medición de la recurrencia de actividades esporádicas en determinados horarios, métricas de movimiento individuales y patrones de movilidad grupales.

Tras aplicar el algoritmo *K-means* sobre la información construida, se obtienen 10 perfiles de clientes diferenciados principalmente por los sectores de residencia y trabajo y por los patrones de movimientos grupales a los que pertenecen. Esto tiene sentido al tener en conocimiento que los sectores tienen características demográficas asociadas que los diferencian entre sí. También se distinguen los perfiles a partir del resto de las variables, pero en menor magnitud.

La información generada se testea a través de regresiones lineales para conocer su poder descriptivo sobre una variable comercial, en este caso el ARPU o ingreso promedio por cliente para la compañía. A partir de esto se concluye que las que aportan más a este objetivo es la combinación de las métricas de movimiento, las variables demográficas y la ubicación del

hogar. Cabe destacar que esto no significa que el resto de la información generada no sea útil para el proyecto de movilidad que envuelve este trabajo.

A nivel de grupos individuales se destaca que las métricas de movimiento obtienen mejores índices que todos los otros grupos, incluso que las variables demográficas que incluyen el GSE.

Se considera que los perfiles pueden ser utilizados para realizar acciones de comunicación a los clientes, pero parece importante complementarlos con otro tipo de información que podría continuar enriqueciendo el conocimiento de los clientes, como por ejemplo la tendencia de un usuario a asistir a un cierto lugar o realizar un tipo de actividad. De todas formas cumplen con el objetivo de este trabajo, pues se basan en la movilidad de los usuarios, siendo la información generada la que explica en mayor medida los perfiles formados.

Dentro de las limitaciones identificadas a lo largo del trabajo destaca se repite como causa la información incompleta con la que se cuenta, pues solo se registran los momentos en que existen llamadas de un usuario, mientras que el comportamiento del cliente en el resto del tiempo continúa siendo una incógnita. Para solucionar esto de forma parcial se pueden utilizar otras fuentes de información, desde agregar a los datos otras conexiones realizadas a las antenas, por ejemplo cuando una persona navega o, mejor aún, el registro que mantiene de la comunicación que mantiene todo el tiempo el móvil con la antena.

6.2. Trabajo Futuro

Para continuar con la línea de desarrollo, se identifican distintos caminos por los que seguir avanzando:

Incorporación de nuevas métricas y mejora de las desarrolladas

En este trabajo se abordan métricas que permiten conocer la forma en que los usuarios se mueven, pero a la vez queda espacio para incorporar muchas medidas que podrían enriquecer la información que se obtuvo.

Ejemplos de posibles medidas son aquellas que aporten mayor información temporal a la descripción del movimiento, pero además se puede agregar lo generado en otras etapas de este trabajo para la creación de medidas que interactúen con el resto de las componentes identificadas, por ejemplo la distancia entre hogar y trabajo puede ser determinante en la forma en que alguien se mueve.

Se diferencia entre semana y fin de semana para los grados de retorno debido a que parecía muy clara la diferencia entre ellos. Lo mismo sucede con las actividades esporádicas. Se piensa que el comportamiento obtenido podría variar también en el caso de las otras métricas escogidas para el análisis y para otras que se pudieran desarrollar en el futuro.

Identificar las actividades desarrolladas en otras PRI no abordadas

Se abordan como actividades diarias un hogar único y un trabajo único, pero esto no es necesariamente cierto para todos los usuarios. Pueden existir clientes con un segundo hogar, como aquellos que realizan visitas de forma cotidiana a familiares o parejas. También se deben considerar los casos de personas que trabajan en más de un lugar, ya sea en el mismo o diferentes días.

Un análisis de este tipo ayudaría a identificar una mayor cantidad de posiciones relevantes de los usuarios, ya que a pesar de que la moda era de 2 posiciones, existen bastantes clientes con una mayor cantidad. Esto llevaría a una mejor explicación del comportamiento y a menores sesgos en lo que se refiere a actividades esporádicas. Además contribuiría a explicar porqué en algunos perfiles se dan movimientos entre sectores que no están relacionados a las actividades que en este trabajo se identifican.

Mejorar el análisis de horarios establecidos para hogar y trabajo

Si bien se construye una métrica de desviación de llamadas en hogar y trabajo para identificar casos diferentes a los normales, se piensa que existe un amplio trabajo que hacer en esta materia.

Se debiese buscar una mejor forma de identificar aquellos usuarios que desarrollan sus actividades cotidianas en un horario distinto al que lo hace la mayoría de los habitantes de la ciudad. Por ejemplo, en este trabajo no se identificaron casos de trabajadores nocturnos y, a pesar de que se definieron reglas que apuntaban a identificar personas que no trabajasen, esto no pudo ser individualizado para incluirlo en la construcción y análisis de los perfiles.

Identificar asistencia a lugares en específico para crear perfiles enfocados a productos o servicios

Para poder medir la participación de usuarios en actividades esporádicas se construyó un algoritmo simple, el que cuenta las veces que una persona se encuentra fuera de su hogar o trabajo en ciertos horarios determinados.

Se piensa que esto podría tener un potencial comercial mucho más alto si es que se orientasen las actividades esporádicas a productos o servicios. Si bien el almuerzo o el entretenimiento nocturno pueden ser considerados de esta forma, se considera se debe abordar esto desde lugares específicos. Es decir, identificar zonas de potenciales negocios como centros de eventos masivos, sectores gastronómicos o de entretención nocturna, para luego analizar quiénes son los usuarios que frecuentan dichos lugares o que desarrollan actividades entorno a estos.

Si esto fuese considerado, los perfiles podrían quizás ser más específicos y se podría orientar la gestión de estos hacia un negocio de *Mobile Marketing* de mejor forma.

Evaluación de los perfiles y las variables generadas

Para este trabajo se evalúa la información generada de acuerdo a su capacidad para explicar una variable comercial determinada y los perfiles obtenidos no son evaluados.

Si bien el ARPU representa una característica importante de los clientes para la empresa, pues es la forma en que miden los ingresos obtenidos por cada usuario en el tiempo, es conveniente para el caso particular de este trabajo medir su aplicabilidad en el proyecto de movilidad del cual es parte. Esta evaluación es posible hacerla a través de pruebas en que se envíen comunicaciones a clientes de distintos perfiles o características y medir si existen diferencias relevantes entre las reacciones de ellos.

En conclusión, se piensa que los perfiles cumplen el objetivo establecido, pero que existen diversas formas de enriquecerlos y de avanzar hacia el punto en que tengan un mayor valor comercial. Este trabajo sienta precedentes para una línea investigativa que mejore los resultados expuestos.

Bibliografía

- [1] Ajith Abraham. Business intelligence from web usage mining. *Journal of Information & Knowledge Management*, 2(04):375–390, 2003.
- [2] Gediminas Adomavicius and Alexander Tuzhilin. Using data mining methods to build customer profiles. *Computer*, 34(2):74–82, 2001.
- [3] Sameer Antani, Rangachar Kasturi, and Ramesh Jain. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern recognition*, 35(4):945–965, 2002.
- [4] F. Barrientos Inostroza. Diseño e implementación de una metodología de predicción de fuga de clientes en una compañía de telecomunicaciones.
- [5] Amir Ben-Dor, Ron Shamir, and Zohar Yakhini. Clustering gene expression patterns. *Journal of computational biology*, 6(3-4):281–297, 1999.
- [6] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.
- [7] Xihui Chen, Jun Pang, and Ran Xue. Constructing and comparing user mobility profiles for location-based services. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 261–266. ACM, 2013.
- [8] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.
- [9] David L Davies and Donald W Bouldin. A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (2):224–227, 1979.
- [10] Instituto Nacional de Estadísticas. División político administrativa y censal. Technical report, Instituto Nacional de Estadísticas, INE, 2007.
- [11] Instituto Nacional de Estadísticas. Encuesta experimental sobre uso de tiempo en el gran santiago, antecedentes metodológicos y principales resultados. Technical report, Instituto Nacional de Estadísticas, INE, 2009.

- [12] Joseph C Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, 4(1):95–104, 1974.
- [13] Alireza Farhangfar, Lukasz Kurgan, and Jennifer Dy. Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, 41(12):3692–3705, 2008.
- [14] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37, 1996.
- [15] Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 330–339. ACM, 2007.
- [16] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
- [17] Lieve Hamers, Yves Hemeryck, Guido Herweyers, Marc Janssen, Hans Keters, Ronald Rousseau, and André Vanhoutte. Similarity measures in scientometric research: the jaccard index versus salton’s cosine formula. *Information Processing & Management*, 25(3):315–318, 1989.
- [18] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979.
- [19] Sam K Hui, Peter S Fader, and Eric T Bradlow. Research note-the traveling salesman goes shopping: The systematic deviations of grocery paths from tsp optimality. *Marketing Science*, 28(3):566–572, 2009.
- [20] Philipp K Janert. *Data analysis with open source tools*. O’Reilly Media, Inc., 2010.
- [21] Sergio Jara-Díaz, Marcela Munizaga, and Javiera Olguín. The role of gender, age and location in the values of work behind time use patterns in santiago, chile*. *Papers in Regional Science*, 92(1):87–102, 2013.
- [22] Shan Jiang, Joseph Ferreira, and Marta C González. Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, 25(3):478–510, 2012.
- [23] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.
- [24] Marzena Kryszkiewicz and Piotr Lasek. Ti-dbscan: Clustering with dbscan by means of the triangle inequality. In *Rough Sets and Current Trends in Computing*, pages 60–69. Springer, 2010.
- [25] Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge, 2008.
- [26] Girish Maskeri, Santonu Sarkar, and Kenneth Heafield. Mining business topics in source

- code using latent dirichlet allocation. In *Proceedings of the 1st India software engineering conference*, pages 113–120. ACM, 2008.
- [27] D. Miranda Valenzuela. Técnicas de imputación para viajes con información incompleta a partir de datos transaccionales de transantiago, 2011.
- [28] Sebastián & Rodas Raúl & Contreras Constanza Muñoz, Ricardo & Ríos. Land use detection with cell phone data using topic models: Case santiago, chile. *unpublished*.
- [29] Nikhil R Pal and James C Bezdek. On cluster validity for the fuzzy c-means model. *Fuzzy Systems, IEEE Transactions on*, 3(3):370–379, 1995.
- [30] D. Palma Alarcon. Diseno de un plan de negocios para proyecto de movilidad en tiempo real.
- [31] Leonid Portnoy. *Intrusion detection with unlabeled data using clustering*. PhD thesis, Computer Science, Columbia University, 2000.
- [32] Bharat Rao and Louis Minakakis. Evolution of mobile location-based services. *Communications of the ACM*, 46(12):61–65, 2003.
- [33] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, Seong Joon Kim, and Song Chong. On the levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking (TON)*, 19(3):630–643, 2011.
- [34] Sebastián A Ríos and Ricardo Muñoz. Content patterns in topic-based overlapping communities. *The Scientific World Journal*, 2014, 2014.
- [35] Raúl Rodas. Tesis para optar al grado de magíster en gestión de operaciones.
- [36] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [37] Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.
- [38] Jörg Sander, Martin Ester, Hans-Peter Kriegel, and Xiaowei Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Mining and Knowledge Discovery*, 2(2):169–194, 1998.
- [39] Venkatesh Shankar and Sridhar Balasubramanian. Mobile marketing: a synthesis and prognosis. *Journal of Interactive Marketing*, 23(2):118–129, 2009.
- [40] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.
- [41] Jt Sythoff and J Morrison. Location-based services: Market forecast, 2011-2015. *Pyramid Research*, 2011.
- [42] Kaan Varnali and Aysegül Toker. Mobile marketing research: The-state-of-the-art. *International Journal of Information Management*, 30(2):144–151, 2010.

- [43] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1100–1108. ACM, 2011.
- [44] Michel Wedel. *Market segmentation: Conceptual and methodological foundations*. Springer, 2000.
- [45] Rui Xu, Donald Wunsch, et al. Survey of clustering algorithms. *Neural Networks, IEEE Transactions on*, 16(3):645–678, 2005.

Anexo A

Análisis estadístico de variables

Tabla A.1: Estadísticos Edad

Estadístico	Edad	Proporción
Media	44,72	
Desviación	21,85	
Moda	51	
Mínimo	-6875	
Máximo	113	
Registros	325.168	92,7 %
Missing	25.689	7,3 %

Tabla A.2: Estadísticos Género

Género	N Usuarios	Proporción
M	130.245	37,1 %
F	132.495	37,8 %
H	313	0,1 %
Registros	263.053	75,0 %
Missing	87.804	25,0 %

Tabla A.3: Estadísticos GSE

GSE	N Usuarios	Proporción
ABC1	48.348	13,8 %
C2	107.579	30,4 %
C3	83.109	23,8 %
D	98.607	28,3 %
E	1.880	0,5 %
Registros	339.551	96,7 %
Missing	11.306	3,3 %

Tabla A.4: Estadísticos antigüedad cliente

Estadístico	Valor (en meses)
Media	53
Desviación	46,16
Mínimo	0
Máximo	289

Tabla A.5: Estadísticos tráfico de voz y mensajería

		Minutos entrada	Minutos salida	Llamadas entrada	Llamadas salida	SMS entrada	SMS salida
ABRIL	Media	140,63	184,26	81,87	93,32	4,60	7,87
	Desviación	167,50	242,89	71,10	93,14	12,21	43,37
	Máximo	13.814	14.091	3.111	7.152	810	4.118
	Mínimo	0	0	0	0	0	0
MAYO	Media	143,52	189,04	85,03	97,06	5,71	8,69
	Desviación	167,79	243,97	71,11	94,66	14,62	44,77
	Máximo	12.122	15.027	1.691	7.437	1.053	3.398
	Mínimo	0	0	0	0	0	0

Tabla A.6: Estadísticos tráfico de datos

	ABRIL		MAYO	
	Kilobytes	Minutos	Kilobytes	Minutos
Media	344.695,5	6.697,3	390.760,9	7.252,8
Desviación	1.434.046,5	12.967,3	1.608.625,5	13.720,0
Mínimo	0	0	0	0
Máximo	205.865.313,0	85.443,0	238.287.030,0	88.908,0

Tabla A.7: Estadísticos equipo

Estadístico	Antigüedad (en meses)	Recambios
Media	20,13	2,50
Desviación	19,27	2,09
Moda	1	1
Mínimo	0	0
Máximo	224	35

Tabla A.8: Proporción usuarios con equipos touch y smartphone

	SI	NO
Touch	55,6 %	44,3 %
Smartphone	37,1 %	62,8 %

Tabla A.9: Estadísticos descriptivos ARPU medido en pesos

Estadístico	Abril	Mayo
Media	13.942,5	14.316,6
Desviación	10.170,6	9.729,8
Moda	0	4.266,0
Mínimo	- 865.242,0	- 496.165,0
Máximo	470.717,0	502.675,0

Tabla A.10: Estadísticos Mora

Estadístico	Valor (en meses)
Media	2,2
Desviación	1,97
Mínimo	0
Máximo	9

Tabla A.11: Estadísticos registros de llamadas

Estadístico	Valor
Media	487,3
Desviación	388,3
Mínimo	1,0
Máximo	11.004,00

Tabla A.12: Estadísticos descriptivos Posiciones Relevantes independientes del tiempo

Estadístico	Valor
Media	5,2
Desviación	4,4
Moda	2
Mínimo	98,0
Máximo	1,0

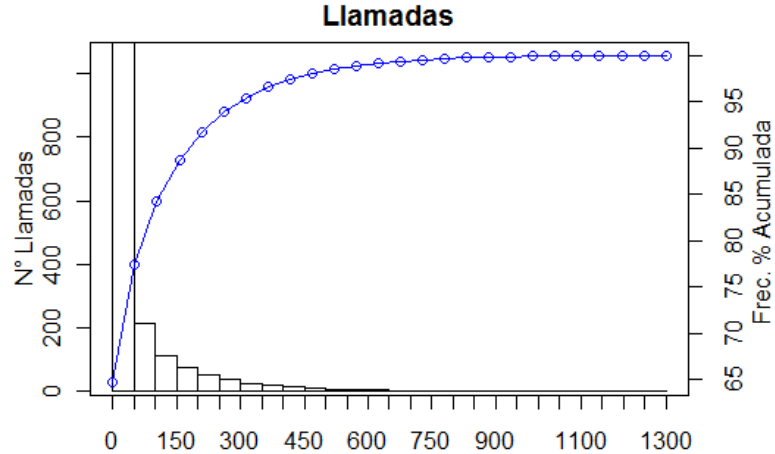


Figura A.1: Histograma de número de llamadas por posiciones no dependientes del tiempo

Tabla A.13: Estadísticos descriptivos Posiciones Relevantes cada una hora

Estadístico	Valor
Media	83,7
Desviación	46,1
Moda	80
Mínimo	312,0
Máximo	1

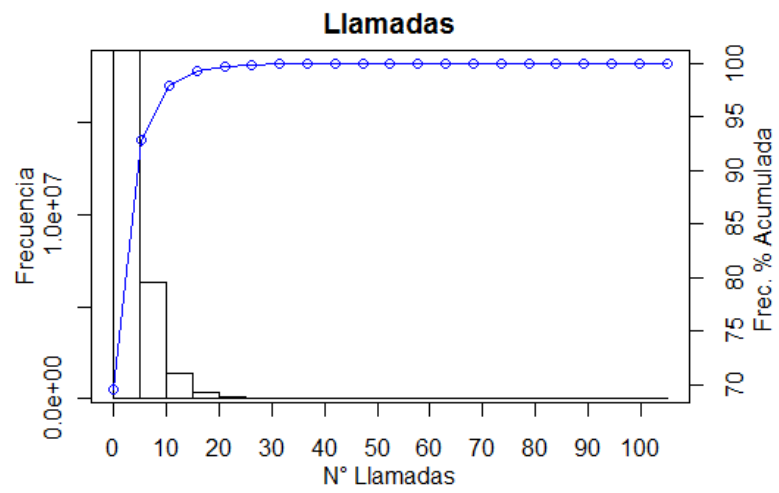


Figura A.2: Histograma de número de llamadas por posiciones dependientes del tiempo

Anexo B

Análisis de correlaciones

Tabla B.1: Resultados correlaciones 1

Variables	Traf. Datos Abr.(KB)	Traf. Datos Abr.(S)	Min. Sal. Abr.	Llam. Sal. Abr.	Min. Ent. Abr.	Llam. Ent. Abr.	SMS Ent. Abr.
Traf. Datos Abr. (KB)	1						
Traf. Datos Abr.(S)	0.39	1					
Min. Sal. Abr.	0.06	0.11	1				
Llam. Sal. Abr.	0.08	0.16	0,67	1			
Min. Ent. Abr.	0.03	0.13	0.38	0.36	1		
Llam. Ent. Abr.	0.06	0.18	0.37	0,63	0,67	1	
SMS Ent. Abr.	0.04	0.07	0.05	0.06	0.05	0.06	1
SMS Sal. Abr.	0.06	0.12	0.09	0.19	0.14	0.24	0.28
Antig. Línea	-0.10	-0.11	0.09	0.06	0.05	0.05	-0.03
Edad	-0.07	-0.11	0.00	-0.01	-0.02	-0.03	-0.03
Traf. Datos May.(KB)	0,79	0.37	0.04	0.07	0.03	0.05	0.03
Traf. Datos May.(S)	0.38	0,94	0.10	0.14	0.12	0.16	0.07
Min. Sal. May.	0.06	0.11	0,89	0,60	0.34	0.33	0.04
Llam. Sal. May.	0.09	0.15	0.58	0.87	0.31	0.55	0.05
Min. Ent. May.	0.03	0.12	0.34	0.32	0,84	0.58	0.05
Llam. Ent. May.	0.06	0.17	0.34	0.58	0.59	0,88	0.05
Mora 6	-0.01	-0.05	0.04	0.11	0.06	0.13	0.00
SMS Sal. May.	0.04	0.06	0.05	0.06	0.06	0.06	0,90
SMS Ent. May.	0.06	0.11	0.08	0.18	0.12	0.21	0.22
Recambios	-0.03	0.03	0.17	0.16	0.11	0.14	-0.01
ARPU	0.18	0.40	0.48	0.53	0.30	0.38	0.10

Tabla B.2: Resultados correlaciones 2

Variables	Min. Ent. May.	Llam. Ent. May.	Mora 6	SMS Ent. May.	SMS Sal. May.	Recambios	Arpu
Traf. Datos Abr. (KB)							
Traf. Datos Abr. (S)							
Min. Sal. Abr.							
Llam. Sal. Abr.							
Min. Ent. Abr.							
Llam. Ent. Abr.							
SMS Ent. Abr.							
SMS Sal. Abr.	1						
Antig. Cliente	-0,03	1					
Edad	-0,05	0,17	1				
Traf. Datos May. (KB)	0,06	-0,11	-0,08	1			
Traf. Datos May. (S)	0,12	-0,13	-0,12	0,40	1		
Min. Sal. May.	0,08	0,07	0,00	0,06	0,12	1	
Llam. Sal. May.	0,17	0,02	-0,03	0,10	0,17	0,64	1
Min. Ent. May.	0,13	0,03	-0,03	0,04	0,13	0,36	0,33
Llam. Ent. May.	0,22	0,03	-0,05	0,06	0,18	0,36	0,61
Mora 6	0,02	-0,04	-0,01	-0,02	-0,06	0,02	0,08
SMS Ent. May.	0,25	-0,04	-0,03	0,04	0,07	0,05	0,07
SMS Sal. May.	0,72	-0,04	-0,06	0,07	0,12	0,09	0,20
Recambios	0,02	0,76	0,15	-0,04	0,02	0,16	0,13
Arpu	0,21	0,09	0,00	0,17	0,39	0,47	0,51

Tabla B.3: Resultados correlaciones 3

Variables	Min. Ent. May.	Llam. Ent. May.	Mora 6	SMS Ent. May.	SMS Sal. May.	Recambios	Arpu
Traf. Datos Abr. (KB)							
Traf. Datos Abr. (S)							
Min. Sal. Abr.							
Llam. Sal. Abr.							
Min. Ent. Abr.							
Llam. Ent. Abr.							
SMS Ent. Abr.							
SMS Sal. Abr.							
Antig. Cliente							
Edad							
Traf. Datos May. (KB)							
Traf. Datos May. (S)							
Min. Sal. May.							
Llam. Sal. May.							
Min. Ent. May.	1						
Llam. Ent. May.	0,66	1					
Mora 6	0,04	0,11	1				
SMS Ent. May.	0,07	0,07	0,00	1			
SMS Sal. May.	0,14	0,24	0,01	0,33	1		
Recambios	0,09	0,12	-0,03	-0,01	0,01	1	
Arpu	0,30	0,38	0,02	0,11	0,21	0,20	1

Anexo C

Tratamiento de outliers

Tabla C.1: Análisis desviaciones tráfico voz

N Desv.	Abril Min. Entrada		Abril Min. Salida		Mayo Min. Entrada		Mayo Min. Salida	
	Valor Máx	% Sup	Valor Máx	% Sup	Valor Máx	% Sup	Valor Máx	% Sup
2	475,6	4,2	668,1	4,9	479,1	4,4	677,0	4,9
3	643,1	1,9	910,0	2,2	646,9	1,8	921,0	2,1
4	810,6	0,9	1.151,9	1,0	814,7	0,9	1.164,9	1,0
5	978,1	0,5	1.393,8	0,5	982,5	0,5	1.408,9	0,5

Tabla C.2: Análisis desviaciones tráfico de datos

N Desv.	Abr. Traf. Datos		May. Traf. Datos	
	Valos Máx	% Superiores	Valor Máx	% Superiores
2	3.137,5	2,9	3.523,4	2,9
3	4.537,9	1,5	5.094,4	1,5
4	5.938,3	0,9	6.665,3	0,9
5	7.338,8	0,6	8.236,2	0,5

Tabla C.3: Análisis desviaciones tráfico mensajería

N Desv.	Abr. SMS Entrada		Abr. SMS Salida		May. SMS Entrada		May. SMS Salida	
	Valor Máx	% Superiores	Valor Máx	% Superiores	Valor Máx	% Superiores	Valor Máx	% Superiores
2	29,0	2,8	90,0	0,9	35,0	3,0	98,2	1,2
3	41,2	1,6	131,5	0,4	49,6	1,6	143,0	0,5
4	53,4	1,0	173,0	0,2	64,2	1,0	187,8	0,3
5	65,6	0,7	214,5	0,1	78,8	0,6	232,5	0,2

Tabla C.4: Análisis desviaciones ARPU

ARPU		
N Desv.	Valor Máx.	% Superiores
2	33.776,2	4,0
3	43.505,9	2,1
4	53.235,7	0,7
5	62.965,4	0,3

Tabla C.5: Análisis desviaciones número de recambios

RECAMBIOS		
N Desv	Valor Máx.	% Superiores
2	6,7	6,4
3	8,8	2,2
4	10,9	0,7
5	12,9	0,3

Tabla C.6: Análisis desviaciones antigüedad equipo

EQ. ANTIGÜEDAD		
N Desv	Valor Máx.	% Superiores
2	58,7	6,8
3	77,9	2,5
4	97,2	0,4
5	116,5	0,1

Anexo D

Análisis poblacional

Tabla D.1: Asignación Comuna a Sector - Parte 1

Sector	Comuna	Población
Centro	Santiago	200.792
Poniente	Cerro Navia	148.312
	Estación Central	130.394
	Lo Prado	104.316
	Maipú	468.390
	Pudahuel	195.653
	Quinta Normal	104.012
	Melipilla	95.540
	Curacaví	24.298
	María Pinto	10.343
	Padre Hurtado	38.768
	1/2 - Peñaflores	33.310
Oriente	Las Condes	248.893
	Lo Barnechea	74.749
	Providencia	120.874
	Vitacura	81.499
	1/2 - La Reina	48.381
Norte	Conchalí	133.256
	Huechuraba	74.070
	Independencia	65.479
	Quilicura	126.518
	Recoleta	148.220
	Renca	133.518
	Colina	77.815
	Lampa	40.228
Til Til	14.755	

Tabla D.2: Asignación Comuna a Sector - Parte 2

Sector	Comuna	Población
Suroriente	La Florida	365.674
	1/2 - La Reina	48.381
	Macul	112.535
	Ñuñoa	163.511
	Peñalolén	216.060
	Puente Alto	492.915
	San José de Maipo	13.376
Sur	El Bosque	175.594
	La Cisterna	85.118
	La Granja	132.520
	La Pintana	190.085
	Lo Espejo	112.800
	Pedro Aguirre Cerda	114.560
	San Joaquín	97.625
	San Miguel	78.872
	San Ramón	94.906
	Pirque	16.565
	San Bernardo	246.762
	Buín	63.419
	Calera de Tango	18.235
	Paine	50.028
	Alhué	4.435
	San Pedro	7.549
	Talagante	59.805
	El Monte	26.459
	Isla de Maipo	25.798
1/2 Peñaflo	33.310	
Cerrilos	71.906	

Anexo E

Tratamiento de outliers en variables generadas

Tabla E.1: Análisis desviaciones posiciones

N Desv	Valor Máx	% Superiores
2	102,391	4,4
3	131,49	1,4
4	160,61	0,5
5	189,72	0,2

Tabla E.2: Análisis desviaciones distancia recorrida

N Desv	Valor Máx	% Superiores
2	2.469,7	4,1
3	3.260,1	1,5
4	4.050,5	0,7
5	4.840,9	0,1
6	5.631,3	0,01
7	6.421,7	0,01

Tabla E.3: Análisis desviaciones radio de giro

N Desv	Valor Máx	% Superiores
2	12,21	3,5
3	15,68	1,1
4	19,15	0,3
5	22,62	0,01

Anexo F

Resultados regresiones lineales

Tabla F.1: Resultados regresión lineal - ARPU. Parte 1

Grupos de Variables		R2	MAPE
Demográficas		0,04	100,24
Métricas		0,17	88,04
Patrones		0,02	98,15
Hogar		0,03	98,74
Trabajo		0,01	99,42
Caracterización		0,01	99,61
Esporádicas		0,05	95,82
Demográficas	Métricas	0,19	88,37
Demográficas	Patrones	0,05	98,61
Demográficas	Caracterización	0,02	100,06
Demográficas	Hogar	0,05	99,22
Demográficas	Trabajo	0,02	99,89
Métricas	Patrones	0,18	87,94
Métricas	Hogar	0,19	86,98
Métricas	Trabajo	0,18	87,92
Métricas	Caracterización	0,18	87,92
Métricas	Esporádicas	0,18	87,91
Patrones	Hogar	0,04	97,28
Patrones	Trabajo	0,02	98,07
Patrones	Esporádicas	0,07	94,59
Patrones	Caracterización	0,03	98,01
Hogar	Esporádicas	0,09	93,92
Hogar	Trabajo	0,03	98,57
Trabajo	Esporádicas	0,06	95,12
Caracterización	Hogar	0,03	98,58
Caracterización	Esporádicas	0,06	95,33
Caracterización	Trabajo	0,02	99,25
Demográficas	Esporádicas	0,06	96,35

Tabla F.2: Resultados regresión lineal - ARPU. Parte 2

Grupos de Variables			R2	MAPE
Demográficas	Patrones	Hogar	0,06	97,93
Demográficas	Métricas	Patrones	0,19	88,31
Demográficas	Métricas	Caracterización	0,19	88,28
Demográficas	Métricas	Esporádicas	0,19	88,21
Demográficas	Métricas	Hogar	0,20	87,53
Demográficas	Métricas	Trabajo	0,19	88,27
Demográficas	Patrones	Caracterización	0,05	98,52
Demográficas	Patrones	Esporádicas	0,09	95,08
Demográficas	Patrones	Trabajo	0,05	98,54
Demográficas	Caracterización	Esporádicas	0,07	95,87
Demográficas	Hogar	Caracterización	0,05	99,07
Demográficas	Hogar	Esporádicas	0,10	94,51
Demográficas	Hogar	Trabajo	0,05	99,17
Demográficas	Trabajo	Caracterización	0,03	99,73
Demográficas	Trabajo	Esporádicas	0,07	95,67
Métricas	Patrones	Hogar	0,19	87,11
Métricas	Patrones	Trabajo	0,18	87,88
Métricas	Patrones	Esporádicas	0,18	87,83
Métricas	Patrones	Caracterización	0,18	87,89
Métricas	Hogar	Trabajo	0,19	86,97
Métricas	Hogar	Esporádicas	0,20	86,84
Métricas	Trabajo	Esporádicas	0,18	87,80
Métricas	Caracterización	Esporádicas	0,19	87,11
Métricas	Caracterización	Hogar	0,19	86,97
Métricas	Caracterización	Trabajo	0,18	87,83
Patrones	Hogar	Esporádicas	0,10	93,37
Patrones	Hogar	Trabajo	0,04	97,26
Patrones	Caracterización	Trabajo	0,03	97,97
Patrones	Caracterización	Esporádicas	0,08	94,38
Patrones	Caracterización	Hogar	0,05	97,19
Patrones	Trabajo	Esporádicas	0,07	94,39
Trabajo	Esporádicas	Caracterización	0,07	94,83
Hogar	Esporádicas	Caracterización	0,09	93,81
Caracterización	Hogar	Trabajo	0,04	98,46
Esporádicas	Hogar	Trabajo	0,09	93,71

Tabla F.3: Resultados regresión lineal - ARPU. Parte 3

Grupos de Variables				R2	MAPE
Demográficas	Métricas	Esporádicas	Patrones	0,20	88,16
Demográficas	Métricas	Esporádicas	Caracterización	0,19	88,13
Demográficas	Métricas	Patrones	Caracterización	0,19	88,27
Demográficas	Métricas	Patrones	Esporádicas	0,10	94,93
Demográficas	Métricas	Patrones	Hogar	0,20	87,65
Demográficas	Métricas	Patrones	Trabajo	0,19	88,24
Demográficas	Métricas	Hogar	Caracterización	0,20	87,53
Demográficas	Métricas	Hogar	Esporádicas	0,20	87,38
Demográficas	Métricas	Hogar	Trabajo	0,20	87,52
Demográficas	Métricas	Trabajo	Caracterización	0,19	88,20
Demográficas	Métricas	Trabajo	Esporádicas	0,10	94,91
Demográficas	Métricas	Trabajo	Esporádicas	0,19	88,12
Demográficas	Patrones	Hogar	Caracterización	0,07	97,86
Demográficas	Patrones	Hogar	Esporádicas	0,11	94,10
Demográficas	Patrones	Hogar	Trabajo	0,06	97,90
Demográficas	Patrones	Trabajo	Caracterización	0,05	98,46
Demográficas	Hogar	Caracterización	Esporádicas	0,10	94,40
Demográficas	Hogar	Trabajo	Caracterización	0,11	94,44
Demográficas	Hogar	Trabajo	Esporádicas	0,11	94,44
Demográficas	Trabajo	Caracterización	Esporádicas	0,08	95,38
Métricas	Patrones	Caracterización	Esporádicas	0,19	87,78
Métricas	Patrones	Hogar	Esporádicas	0,20	86,99
Métricas	Patrones	Hogar	Trabajo	0,19	87,09
Métricas	Patrones	Hogar	Caracterización	0,20	87,11
Métricas	Patrones	Trabajo	Caracterización	0,18	87,84
Métricas	Patrones	Trabajo	Esporádicas	0,18	87,76
Métricas	Hogar	Trabajo	Caracterización	0,19	86,96
Métricas	Hogar	Caracterización	Esporádicas	0,20	86,83
Métricas	Hogar	Trabajo	Esporádicas	0,20	86,84
Métricas	Trabajo	Caracterización	Esporádicas	0,18	87,72
Patrones	Hogar	Caracterización	Esporádicas	0,10	93,30
Patrones	Hogar	Trabajo	Caracterización	0,05	97,17
Patrones	Hogar	Trabajo	Esporádicas	0,10	93,32
Patrones	Trabajo	Caracterización	Esporádicas	0,08	94,24
Hogar	Trabajo	Caracterización	Esporádicas	0,10	93,64

Tabla F.4: Resultados regresión lineal - ARPU. Parte 4

Grupos de Variables							R2	MAPE
Demográficas	Métricas	Patrones	Caracterización	Esporádicas			0,20	88,13
Demográficas	Métricas	Patrones	Hogar	Caracterización			0,20	87,65
Demográficas	Métricas	Patrones	Hogar	Esporádicas			0,20	87,51
Demográficas	Métricas	Patrones	Hogar	Trabajo			0,20	87,63
Demográficas	Métricas	Patrones	Trabajo	Caracterización			0,20	88,22
Demográficas	Métricas	Patrones	Trabajo	Esporádicas			0,20	88,10
Demográficas	Métricas	Hogar	Caracterización	Esporádicas			0,20	87,36
Demográficas	Métricas	Trabajo	Caracterización	Esporádicas			0,20	88,06
Demográficas	Patrones	Hogar	Caracterización	Esporádicas			0,11	94,05
Demográficas	Patrones	Hogar	Trabajo	Caracterización			0,07	97,83
Demográficas	Patrones	Hogar	Trabajo	Esporádicas			0,11	94,38
Demográficas	Patrones	Trabajo	Caracterización	Esporádicas			0,10	94,80
Demográficas	Métricas	Hogar	Trabajo	Esporádicas			0,20	87,37
Demográficas	Hogar	Trabajo	Caracterización	Esporádicas			0,11	94,38
Métricas	Patrones	Hogar	Caracterización	Esporádicas			0,20	86,98
Métricas	Patrones	Hogar	Trabajo	Caracterización			0,20	87,09
Métricas	Patrones	Hogar	Trabajo	Esporádicas			0,20	86,98
Métricas	Patrones	Trabajo	Caracterización	Esporádicas			0,19	87,73
Métricas	Hogar	Trabajo	Caracterización	Esporádicas			0,20	86,83
Patrones	Hogar	Trabajo	Caracterización	Esporádicas			0,10	93,26
Demográficas	Métricas	Patrones	Hogar	Trabajo	Esporádicas		0,20	87,49
Demográficas	Métricas	Patrones	Hogar	Caracterización	Esporádicas		0,21	87,49
Demográficas	Métricas	Patrones	Hogar	Trabajo	Caracterización		0,20	87,62
Demográficas	Patrones	Caracterización	Hogar	Trabajo	Esporádicas		0,11	94,00
Demográficas	Patrones	Caracterización	Trabajo	Esporádicas	Métricas		0,20	88,08
Demográficas	Caracterización	Hogar	Trabajo	Esporádicas	Métricas		0,20	87,36
Patrones	Caracterización	Hogar	Trabajo	Esporádicas	Métricas		0,20	86,96
Demográficas	Patrones	Caracterización	Hogar	Trabajo	Esporádicas	Métricas	0,21	87,47

Tabla F.5: Resultados regresiones lineales - log(ARPU)

Combinación de Variables							R2	MAPE
Demográficas	Métricas	Hogar					0,167	4,47
Demográficas	Métricas	Hogar	Trabajo				0,167	4,47
Demográficas	Métricas	Patrones	Hogar				0,168	4,47
Demográficas	Métricas	Patrones	Hogar	Trabajo			0,168	4,47
Demográficas	Métricas	Hogar	Caracterización				0,168	4,46
Demográficas	Métricas	Patrones	Hogar	Caracterización			0,168	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo	Caracterización		0,169	4,46
Demográficas	Métricas	Hogar	Esporádicas				0,169	4,46
Demográficas	Métricas	Hogar	Trabajo	Esporádicas			0,169	4,46
Demográficas	Métricas	Patrones	Hogar	Esporádicas			0,169	4,46
Demográficas	Métricas	Hogar	Caracterización	Esporádicas			0,170	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo	Esporádicas		0,170	4,46
Demográficas	Métricas	Hogar	Trabajo	Caracterización	Esporádicas		0,170	4,46
Demográficas	Métricas	Patrones	Hogar	Caracterización	Esporádicas		0,170	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo	Caracterización	Esporádicas	0,170	4,46

Tabla F.6: Resultados regresión lineal - Log(ARPU). Edad transformada.

Combinación de Variables							R2	MAPE
Demográficas	Métricas	Hogar					0,168	4,47
Demográficas	Métricas	Hogar	Trabajo				0,168	4,46
Demográficas	Métricas	Patrones	Hogar				0,168	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo			0,168	4,46
Demográficas	Métricas	Hogar	Caracterización				0,168	4,46
Demográficas	Métricas	Patrones	Hogar	Caracterización			0,169	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo	Caracterización		0,169	4,46
Demográficas	Métricas	Hogar	Esporádicas				0,169	4,46
Demográficas	Métricas	Hogar	Trabajo	Esporádicas			0,169	4,46
Demográficas	Métricas	Patrones	Hogar	Esporádicas			0,170	4,46
Demográficas	Métricas	Hogar	Caracterización	Esporádicas			0,170	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo	Esporádicas		0,170	4,46
Demográficas	Métricas	Hogar	Trabajo	Caracterización	Esporádicas		0,170	4,46
Demográficas	Métricas	Patrones	Hogar	Caracterización	Esporádicas		0,170	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo	Caracterización	Esporádicas	0,171	4,46

Tabla F.7: Resultados regresiones lineales - log(ARPU). Edad y métricas transformadas.

Combinación de Variables							R2	MAPE
Demográficas	Métricas	Hogar					0,163	4,47
Demográficas	Métricas	Hogar	Trabajo				0,163	4,47
Demográficas	Métricas	Hogar	Caracterización				0,164	4,47
Demográficas	Métricas	Patrones	Hogar				0,164	4,47
Demográficas	Métricas	Patrones	Hogar	Trabajo			0,164	4,47
Demográficas	Métricas	Hogar	Esporádicas				0,165	4,46
Demográficas	Métricas	Patrones	Hogar	Caracterización			0,165	4,46
Demográficas	Métricas	Hogar	Trabajo	Esporádicas			0,166	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo	Caracterización		0,166	4,46
Demográficas	Métricas	Hogar	Caracterización	Esporádicas			0,167	4,46
Demográficas	Métricas	Patrones	Hogar	Esporádicas			0,167	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo	Esporádicas		0,167	4,46
Demográficas	Métricas	Patrones	Hogar	Caracterización	Esporádicas		0,168	4,45
Demográficas	Métricas	Hogar	Trabajo	Caracterización	Esporádicas		0,168	4,46
Demográficas	Métricas	Patrones	Hogar	Trabajo	Caracterización	Esporádicas	0,169	4,45

Anexo G

Medias de perfiles

Tabla G.1: Medias Perfiles. Parte 1.

	Perfil 1	Perfil 2	Perfil 3	Perfil 4	Perfil 5	Perfil 6	Perfil 7	Perfil 8	Perfil 9	Perfil 10
Mujer	0,4	0,3	0,4	0,4	1,0	0,5	-	-	0,9	-
Hombre	0,4	0,4	0,4	0,4	-	0,4	0,6	0,7	-	0,6
GSE	2,7	2,9	2,6	2,6	3,0	1,8	3,2	2,9	2,9	3,1
Edad	46,4	46,8	48,5	47,6	47,7	51,1	48,3	47,3	47,1	46,8
Tráfico Datos	353,5	324,4	269,1	343,0	243,0	322,6	250,7	299,2	250,0	394,7
Min. Salida	179,0	173,7	160,4	167,5	170,9	194,4	145,7	170,2	174,5	179,2
Min. Entrada	141,5	134,4	129,3	132,8	138,6	164,3	112,3	122,1	140,0	129,2
SMS Salida	7,4	6,1	6,4	7,8	6,6	7,1	5,2	5,6	6,6	6,4
SMS Entrada	5,2	4,5	4,5	5,3	4,4	5,0	3,5	4,0	4,4	4,6
Ant. Cliente	50,8	49,1	51,4	55,3	54,1	69,9	45,3	47,8	51,8	44,7
Ant. Equipo	19,2	20,3	19,4	19,8	21,6	18,7	20,2	20,0	21,7	18,9
Smartphone	0,4	0,3	0,3	0,4	0,3	0,5	0,3	0,3	0,3	0,3
Recambios	2,5	2,4	2,5	2,6	2,5	3,3	2,1	2,3	2,4	2,2
ARPU	14.718	14.227	13.437	14.687	13.436	17.142	12.385	13.539	13.431	14.206
Mora	2,1	2,3	2,2	2,1	2,3	1,7	2,3	2,2	2,3	2,3
Radio de Giro	5,9	5,1	4,7	5,2	5,3	4,4	4,2	5,5	5,4	6,9
GR Semana	0,2	0,3	0,2	0,2	0,3	0,2	0,2	0,2	0,2	0,2
GR FDS	0,3	0,3	0,3	0,3	0,3	0,2	0,3	0,3	0,3	0,3
Distancia Recorrida	982,9	895,3	718,9	802,0	1.004,2	639,2	781,0	964,9	953,8	1.298,7
Posiciones	52,9	41,6	39,0	51,2	38,0	50,1	27,8	42,0	42,0	47,9
Norte-Centro	0,02	0,24	0,00	0,14	0,00	0,00	-	0,01	0,01	0,00
Oriente-Suroriente	0,53	0,01	0,27	0,02	0,00	0,08	-	0,00	0,00	0,00
Oriente-Centro	0,17	0,01	0,01	0,20	0,01	0,10	-	0,01	0,01	0,01
Oriente-Sur	0,01	0,00	0,00	0,00	0,20	0,04	-	0,00	0,00	0,30
Norte-Oeste y Centro-Oeste	0,01	0,04	0,00	0,27	0,00	0,00	-	0,30	0,23	0,00
Oriente-Norte	0,20	0,20	0,00	0,01	0,00	0,05	-	0,00	0,01	0,00
Norte-Sur	0,00	0,04	0,00	0,00	0,06	-	-	0,00	0,00	0,13
Sur-Centro	0,04	0,45	0,51	0,15	0,44	0,69	1,00	0,36	0,31	-
Oriente-Oeste	0,01	0,00	0,00	0,02	0,00	0,04	-	0,18	0,34	0,00
Suroriente-Sur	0,01	0,01	0,09	0,01	0,16	-	-	0,00	0,00	0,29
Suroriente-Centro	0,01	0,01	0,11	0,17	0,00	0,00	-	0,00	0,00	0,00
Sur-Oeste	0,00	0,00	0,00	0,01	0,12	0,00	-	0,12	0,09	0,24

Tabla G.2: Medias perfiles. Parte 2.

	Perfil 1	Perfil 2	Perfil 3	Perfil 4	Perfil 5	Perfil 6	Perfil 7	Perfil 8	Perfil 9	Perfil 10
H. Oriente	0,00	0,11	0,14	0,21	0,03	0,93	0,01	0,04	0,04	0,04
H. Suroriente	0,58	0,09	0,78	0,20	0,07	0,00	0,01	0,04	0,03	0,11
H. Norte	0,21	0,71	0,02	0,15	0,01	0,01	0,00	0,05	0,04	0,03
H. Sur	0,00	0,00	0,00	0,02	0,85	0,00	0,92	0,00	0,00	0,80
H. Centro	0,18	0,08	0,04	0,14	0,03	0,02	0,03	0,05	0,04	0,02
H. Oeste	0,01	0,01	0,01	0,28	-	0,00	0,01	0,81	0,85	0,00
T. Oriente	0,98	0,00	-	-	0,12	0,95	0,01	0,10	0,27	0,24
T. Suroriente	-	0,01	0,99	-	0,06	0,00	0,01	0,02	0,02	0,10
T. Norte	0,01	0,98	0,00	-	0,03	0,00	0,01	0,08	0,06	0,07
T. Sur	0,00	0,00	0,00	-	0,59	0,00	0,76	0,08	0,05	0,41
T. Centro	-	-	-	1,00	0,11	0,00	0,18	0,00	0,01	0,05
T. Oeste	0,00	0,00	0,00	-	0,06	0,01	0,02	0,72	0,57	0,13
Alm. Semana	0,1	0,2	0,2	0,1	0,2	0,1	0,1	0,2	0,2	0,2
Alm. FDS	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,2	0,2
Ent. Noc. Sem.	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1
Ent. Noc. FDS	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1
Fin de Semana	0,3	0,3	0,3	0,2	0,3	0,2	0,2	0,3	0,3	0,3
H. Residencial	0,96	0,97	0,97	0,96	0,98	0,94	0,98	0,98	0,98	0,99
H. Entretenimiento	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
H. Comercio	0,02	0,02	0,02	0,02	0,01	0,02	0,01	0,01	0,01	0,01
H. Educación	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
H. Oficina	0,01	0,01	0,00	0,01	0,00	0,01	0,01	0,01	0,00	0,00
H. Industrial	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
T. Entretenimiento	0,02	0,03	0,04	0,02	0,03	0,03	0,03	0,04	0,04	0,03
T. Comercio	0,11	0,16	0,26	0,13	0,14	0,12	0,14	0,16	0,16	0,13
T. Educación	0,17	0,05	0,25	0,11	0,18	0,18	0,17	0,10	0,13	0,17
T. Oficina	0,50	0,55	0,33	0,57	0,46	0,47	0,45	0,52	0,50	0,48
T. Industrial	0,02	0,14	0,05	0,03	0,08	0,02	0,09	0,12	0,09	0,10