



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

ANÁLISIS DE LA FRECUENCIA DE USO Y MIGRACIÓN DE USUARIOS DEL
TRANSPORTE PÚBLICO DE SANTIAGO

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA
INGENIERÍA, MENCIÓN TRANSPORTE

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

CAMILO ANDRÉS LENG OLIVARES

PROFESOR GUÍA:
MARCELA ADRIANA MUNIZAGA MUÑOZ

MIEMBROS DE LA COMISIÓN:
CRISTIAN ANGELO GUEVARA
RICHARD WEBER HAAS
MARTIN TRÉPANIER

SANTIAGO DE CHILE

2018

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN TRANSPORTE
Y AL TÍTULO DE INGENIERO CIVIL
POR: CAMILO ANDRÉS LENG OLIVARES
FECHA: ENERO 2018
PROF. GUÍA: DRA. MARCELA ADRIANA MUNIZAGA MUÑOZ

ANÁLISIS DE LA FRECUENCIA DE USO Y MIGRACIÓN DE USUARIOS DEL TRANSPORTE PÚBLICO DE SANTIAGO

¿En qué modo de transporte viajo hoy al trabajo? Por muy cotidiana que parezca esta pregunta, su respuesta no es para nada trivial. Esta está influenciada por factores propios de los modos de transporte y por factores externos de la situación de cada individuo y su entorno. En particular para el transporte público, se cree que el nivel de servicio que este le entregue a sus usuarios afecta su decisión de continuar viajando en él. Esta tesis plantea como hipótesis que existen factores del transporte público que tienen un impacto directo en el comportamiento de sus usuarios. Al verse estos modificados, la satisfacción percibida por los usuarios varía, lo que repercute en la frecuencia con que estos viajan en el sistema. Mediante el uso de datos pasivos del transporte público en Santiago, Chile y Gatineau, Canadá esta investigación analiza la incidencia de los principales atributos del sistema en la frecuencia de uso de este por parte de sus usuarios.

Esta tesis esta estructurada en cinco secciones. La primera contiene una revisión de la literatura relacionada con el uso de datos pasivos en el transporte público y la migración de usuarios. La siguiente sección está dedicada al análisis de los datos. Primero se definen criterios para identificar y eliminar observaciones erróneas o atípicas. Luego se lleva a cabo un análisis gráfico que permite observar los patrones temporales y espaciales de los usuarios. Gracias a este análisis se identifican los puntos críticos de la red, las diferencias en la demanda por zonas, la varianza en la cantidad de transacciones en el sistema, el impacto de la incorporación de un nuevo corredor de buses rápidos y otros aspectos propios de cada sistema. En la sección siguiente se modela la frecuencia con que un usuario viaja en transporte público según su experiencia pasada con el sistema. Dos cortes temporales son necesarios. El primero periodo explica la experiencia de viaje de cada usuario mediante atributos tales como el tiempo de viaje, espera y caminata, el modo de transporte preferido, la estimación de residencia, entre otros. El segundo periodo muestra con que frecuencia los usuarios viajan en transporte público. Dependiendo la cantidad de días y transacciones realizadas estos son clasificados en frecuentes, semi frecuentes, no frecuentes y ausentes. Construidas estas categorías, un modelo logit ordinal es utilizado para identificar los atributos que poseen un impacto significativo en el comportamiento de los usuarios. En la penúltima sección se replica la metodología propuesta en la ciudad de Gatineau, Canadá. Este nuevo escenario prueba la capacidad de reproducir la metodología con un sistema de transporte público diferente y permite encontrar factores comunes que afecten el comportamiento de los usuarios. Finalmente, en la última sección se discuten las contribuciones, limitaciones y trabajos futuros de la metodología propuesta, comparando los resultados de cada escenario y del global de la tesis.

A mis padres, Ricardo y Ximena.

Tabla de Contenido

Índice de Tablas	v
Índice de Ilustraciones	vi
1. Introducción	1
2. Revisión Bibliográfica	3
2.1. Principios del uso de datos pasivos en transporte público	3
2.2. Estimación de bajada y creación de la Matriz Origen–Destino	4
2.3. Enriquecimiento de datos pasivos	5
2.4. Regularidad en los patrones de viaje de los usuarios	7
2.5. Fidelidad y abandono del transporte público	8
2.6. Modelo Logit Ordinal	11
3. Análisis de las bases de datos	13
3.1. Descripción del caso de estudio	13
3.2. Pre procesamiento de datos	14
3.2.1. Estimación de bajada	15
3.2.2. Cálculo de indicadores de calidad de servicio	16
3.2.3. Estimación de residencia	17
3.3. Descripción de la base de datos	18
3.4. Limpieza de datos	19
3.5. Análisis de la base de datos	20
3.5.1. Primer Periodo: 18 - 25 Abril, 2015	21
3.5.2. Segundo Periodo: Noviembre, 2015	24
3.6. Indicadores de calidad de servicio	26
3.7. Variabilidad en el tiempo de viaje	27
3.8. Encuesta CASEN	29
4. Modelación	33
4.1. Clasificación de usuarios	34
4.2. Modelo Logit Ordinal	35
4.3. Supuestos y limitaciones del modelo	37
4.4. Agrupación por tarjeta y variables de modelación	37
4.4.1. Variable Dependiente	38
4.4.2. Variables Independientes	38
4.5. Resultados	41

5. Segundo escenario: Gatineau, Quebec, Canadá	45
5.1. Análisis de la base de datos	45
5.1.1. Descripción del escenario	45
5.1.2. Descripción de la base de datos	46
5.1.3. Primer periodo: Noviembre 2013	49
5.1.4. Censo de Canadá 2011	50
5.2. Estimación de residencia	51
5.3. Modelación	53
5.3.1. Clasificación de usuarios	53
5.3.2. Variables independientes y dependiente	55
5.3.3. Resultados	58
5.4. Síntesis	60
6. Conclusiones	61
6.1. Contribuciones	61
6.2. Limitaciones	63
6.3. Trabajos Futuros	64
7. Bibliografía	65
Anexos	67
A. Tablas del pre procesamiento de datos	68
B. Periodos definidos por Transantiago	70

Índice de Tablas

3.1. Descripción de las bases de datos.	19
3.2. Máxima cantidad de viajes y etapas permitidas por días viajados.	20
3.3. Quintiles de ingreso en la Región Metropolitana.	30
4.1. Distribución de la variable dependiente.	38
4.2. Distribución de las variables binarias.	41
4.3. Distribución de las variables enteras y continuas	42
4.4. Resultados Modelo Logit Ordinal en Santiago.	42
5.1. Descripción de las bases de datos en Gatineau, Quebec.	48
5.2. Distribución de la variable dependiente.	57
5.3. Distribución de las variables binarias.	57
5.4. Distribución de las variables enteras y continuas.	58
5.5. Resultados Modelo Logit Ordinal en Gatineau, Quebec.	59
A.1. Tabla de etapas	68
A.2. Tabla de viajes	69
B.1. Periodos Transantiago en días laborales	70
B.2. Periodos Transantiago en días sábado	70
B.3. Periodos Transantiago en días domingo	71

Índice de Ilustraciones

3.1. Ejemplo de una estimación de residencia exitosa.	17
3.2. Distribución de tarjetas por tipo de contrato.	18
3.3. Número de viajes por tipo de día.	21
3.4. Distribución temporal de las transacciones en Santiago.	22
3.5. Distribución espacial de las transacciones en la Línea 1 del metro.	23
3.6. Distribución temporal de las transacciones en la estación Universidad de Chile.	23
3.7. Distribución temporal de las transacciones en la estación Plaza de Puente Alto.	24
3.8. Distribución de la distancia viajada por etapa.	25
3.9. Mapa de calor de las transacciones en el mes de Noviembre, 2015.	26
3.10. Variabilidad en el número de transacciones de días laborales por hora.	26
3.11. Distribución de los Indicadores de Cumplimiento de Frecuencia y Regularidad.	28
3.12. Distribución de la variabilidad en el tiempo de expedición de los buses.	29
3.13. Distribución del ingreso en el área metropolitana de Santiago.	31
3.14. Distribución espacial del ingreso en el área metropolitana de Santiago.	31
3.15. Distribución de la tasa de motorización en el área metropolitana de Santiago.	32
4.1. Clasificación de usuarios según su grado de uso del sistema.	35
4.2. Evolución de la clasificación de usuarios según su grado de uso del sistema.	36
4.3. Distribución del promedio y la desviación estándar del tiempo de viaje.	39
4.4. Distribución del tiempo medio de espera y caminata.	40
5.1. Distribución temporal de las transacciones en Gatineau, Quebec.	47
5.2. Evolución mes a mes de la clasificación de los usuarios.	48
5.3. Mapa de calor de las transacciones del mes de Noviembre, 2013 en Gatineau.	49
5.4. Distribución temporal de las transacciones en Gatineau, Quebec.	50
5.5. Distribución del ingreso en Gatineau, Quebec.	51
5.6. Distribución espacial del ingreso en Gatineau, Quebec.	52
5.7. Partición modal en Gatineau, Quebec.	53
5.8. Ejemplo de nueva metodología de estimación de zona de residencia.	54
5.9. Clasificación de usuarios según su grado de uso del sistema.	55
5.10. Evolución de la clasificación de usuarios según su grado de uso del sistema.	56

Capítulo 1

Introducción

El área metropolitana de Santiago, la capital de Chile, posee alrededor de 6,6 millones de habitantes, los que en un día laboral normal realizan sobre 18 millones de viajes para llevar a cabo sus labores y satisfacer sus necesidades (Sectra, 2012). La movilidad urbana en una ciudad de estas características no es un tema sencillo y en Santiago existen dos tendencias preocupantes. Según las últimas dos encuestas Origen–Destino, la partición modal del transporte público ha disminuido de un 30,1 % en 2001 a un 25,1 % en 2012. En cambio, en los mismos años la partición modal del automóvil ha aumentado de un 21,0 % a un 26,1 % (Muñoz et al., 2016). Estas tendencias perjudican la movilidad urbana. El aumento de vehículos particulares en las calles genera mayor congestión, obligando a los buses a circular más lento. Esto deteriora la calidad de servicio que entrega el transporte público y aumenta sus costos operacionales. El aumento en los costos operacionales junto con la baja demanda por transporte público obliga a los operadores a disminuir sus frecuencias o aumentar las tarifas, perjudicando aun más el sistema. Este deterioro hace del transporte público un modo menos atractivo, aumentando la tasa de migración de pasajeros a modos alternativos y disminuyendo la tasa de atracción de nuevos usuarios. Las personas que dejan de utilizar el transporte público se ven tentadas a comprar un vehículo particular generándose así el llamado “Círculo Vicioso del Transporte Público”, donde el deterioro del sistema a causa del aumento de vehículos particulares, incita al uso de estos últimos agravando el problema. En la actualidad en Santiago el automóvil ya es un modo de transporte bien atractivo para una parte importante de la población. Sin embargo, si no se realizan esfuerzos por controlar este deterioro continuo del sistema, los problemas de movilidad urbana serán insostenibles en el largo plazo. Cuando esto suceda, la calidad de vida de todos los santiaguinos se verá afectada independiente del modo de transporte que utilicen.

La baja demanda de usuarios en el transporte público se puede enfrentar de dos formas; atrayendo nuevos usuarios o reteniendo los usuarios que ya utilizan el sistema. Lo primero generalmente se lleva a cabo invirtiendo en campañas publicitarias o impulsando proyectos de ley que restrinjan el uso del automóvil. En cambio, para la retención de usuarios es necesario mejorar la experiencia de viaje de los usuarios más propensos al abandono. Dado el tamaño del sistema, una inversión global de recursos podría generar pequeñas mejoras en la calidad de servicio de cada individuo. Pero, si se conociera a aquellos usuarios con la mayor predisposición a abandonar el sistema, la inversión podría realizarse de forma focalizada,

mejorando la experiencia de las personas que están más próximas a migrar a otros modos. De esta forma, la planificación y evaluación de proyectos se podría realizar dando prioridad a aquellas medidas que apunten a retener más usuarios.

El objetivo principal de esta investigación es entender cuales son los atributos del sistema de transporte público que influyen en la decisión del grado de uso o abandono del sistema. Para llevar esto a cabo, se hace un análisis descriptivo de los datos pasivos generados por la operación de Transantiago que luego serán utilizados para la creación de un modelo logit ordinal que explique la decisión de los usuarios sobre su frecuencia de uso del sistema. Como objetivos secundarios, se desea incorporar en la modelación la existencia de distintos estratos socio económicos muy marcados en la población de Santiago, incluir el efecto de la varianza en el tiempo de viaje de los servicios y mantener como fuente principal de información los datos pasivos generados por la operación del sistema. Esto último permite trabajar con una muestra de tarjetas muy superior a las normalmente reportadas por encuestas de viajes y también permite que la metodología sea reproducible en distintos cortes temporales y en otros escenarios con datos similares.

Esta tesis continúa la línea de trabajo propuesta por la memoria Calvo (2015). En esta oportunidad se busca realizar un trabajo más profundo en el tratamiento de los datos, corregir ciertos supuestos sobre el criterio para definir la migración, incorporar el efecto de nuevas variables y utilizar modelos más sofisticados.

La tesis está distribuida de la siguiente forma; el capítulo 2 presenta una completa revisión bibliográfica sobre el uso de datos pasivos en transporte público y migración de usuarios. El capítulo 3 describe el escenario de Santiago y los datos disponibles. En este capítulo se realiza un análisis gráfico de los datos del sistema, identificando patrones temporales y espaciales del comportamiento de la ciudad. El capítulo 4 presenta el modelo que será utilizado, como se construye su variable dependiente y las variables independientes que utiliza. El capítulo 5 explica los resultados encontrados. En el capítulo 6 se replica la metodología propuesta en un nuevo escenario. A lo largo de este capítulo se describirá el nuevo escenario, sus datos, el modelo y los resultados. Finalmente, el capítulo 7 presenta las conclusiones de toda la tesis.

Capítulo 2

Revisión Bibliográfica

En una constante búsqueda por mejorar los servicios de transporte público en las grandes metrópolis del mundo es que estos sistemas se han volcado cada vez más en verse apoyados por las herramientas tecnológicas que han emergido en el último tiempo. Junto con cada tecnología nueva que aparece, un sinfín de oportunidades surgen donde estas puedan ser aplicadas. En el transporte público varios han sido los cambios que se han apreciado; vehículos más eficientes, sistemas de pago automáticos, sistemas de localización automática de vehículos, trenes de metro conducidos por computador, etc. Los avances no solo se han manifestado en tecnología palpable, también grandes avances han surgido en la planificación de estos sistemas. Nuevos modelos son capaces de predecir el comportamiento de los usuarios con cada vez mayor exactitud, los tiempos de viaje son calculados de forma dinámica, matrices origen-destino son obtenidas a partir de la observación del uso del sistema en vez de mediante encuestas, entre otros. Toda esta información es canalizada y entregada a usuarios capaces de tomar decisiones más fundamentadas en el día a día. A continuación se presenta una revisión bibliográfica de las investigaciones más destacadas que se han llevado a cabo desde la aparición de los sistemas de pago automático con tarjetas inteligentes en el transporte público y otros trabajos complementarios en el área de transporte.

2.1. Principios del uso de datos pasivos en transporte público

Una vez que los sistemas de pago automático fueron incorporados en el transporte público, rápidamente se identificó que ellos eran una fuente de datos muy valiosa para la planificación y regulación del sistema. Bagchi y White (2005) describe los potenciales usos de estos datos llevando a cabo tres estudios en Gran Bretaña para estimar la tasa de recambio, tasa de viajes promedio en una semana y la proporción de viajes con transbordos. No obstante, ninguno de los casos analizados posee una tasa de penetración alta de la tarjeta inteligente, por lo tanto, los datos no son considerados representativos de la movilidad de la población. Utsunomiya et al. (2006) estudia la variabilidad en el origen del primer viaje del día en Chicago. Basándose en una muestra con información exacta de la dirección del hogar, analiza las diferencias por

zonas de residencia. Ambos estudios destacan la gran abundancia temporal de información en las bases de datos pasivos y el bajo costo que requieren comparados con las encuestas que se llevan a cabo normalmente. Además, advierten de posibles errores debido al mal funcionamiento de los sistemas de recolección de datos y por consiguiente la falta de datos o existencia de datos erróneos. Ambos concluyen que esta nueva fuente de información debe ser complementaria a las encuestas de viaje ya que las encuestas recopilan información valiosa sobre el usuario que no está presente en estos datos.

En adelante mucho es lo que se ha realizado. Pelletier et al. (2011) realiza una síntesis de los trabajos relacionados con el uso de datos provenientes de los sistemas de cobro automático en transporte público, describe las tecnologías implementadas alrededor del mundo, los sistemas de manejo de información necesarios, presenta los problemas de políticas de privacidad, compara los sistemas de cobro automático con otros tipos de sistemas de cobros desde el punto de vista de la operación, los usuarios y la regulación, presenta las ventajas y desventajas de los sistemas de tarjetas inteligentes y finalmente clasifica los estudios que se han llevado a cabo hasta la fecha en tres niveles; estratégicos, tácticos y operacionales.

En la actualidad existen distintos sistemas de cobro, cada uno con sus respectivas características que son capaces de adaptarse a las necesidades de cada ciudad. Cada uno de estos sistemas es único y por lo tanto también lo es la información que recopilan. En un principio estos sistemas son diseñados exclusivamente para la recaudación y redistribución de los ingresos, limitando la información capturada a lo esencialmente necesario para realizar esta tarea. Esto impide a los planificadores e investigadores en transporte gozar de información completa sobre el uso del sistema por parte de sus pasajeros.

2.2. Estimación de bajada y creación de la Matriz Origen–Destino

Muchos sistemas de transporte público en el mundo funcionan con tarifa plana, es decir, se realiza un pago por viaje. Estos sistemas generan el cobro al inicio del viaje, luego de la validación con una tarjeta inteligente o de prepago y permiten que el usuario viaje en ese servicio sin necesidad de registrar su bajada. En estos casos, la información registrada sobre los viajes de una persona incluye el inicio y no el final. Para solucionar esta falta de información se han desarrollado metodologías que observan la cadena de transacciones realizadas por una tarjeta y estiman el destino de los viajes por la ubicación del inicio del siguiente viaje.

En la ciudad de Gatineau, Quebec, Trépanier et al. (2007) propone una metodología para estimar el paradero de bajada de cada viaje minimizando la distancia entre las posibles paradas de bajada del servicio utilizado y la parada de subida de la siguiente validación. Una vez conocidas las paradas de bajada la metodología diferencia paradas de transferencia de paradas de actividades, identificando de esta forma las etapas dentro de un viaje. En caso de no existir información de la siguiente validación durante un día, la metodología es capaz de observar patrones históricos comparando variables espaciales y temporales para estimar la bajada.

La ciudad de Santiago de Chile presenta un sistema de transporte público similar. En este caso, Munizaga y Palma (2012) presenta una metodología capaz de estimar la parada de bajada de un 80 % de las transacciones del sistema. Para esto, minimiza una función de tiempo generalizado de viaje que considera el tiempo de viaje en el vehículo y el de caminata al siguiente paradero. Esta función objetivo permite corregir los problemas de recorridos que realizan retornos y circulan a una menor distancia de la ubicación de la siguiente transacción, pero requieren de un tiempo considerablemente mayor que el tiempo necesario para caminar esa distancia. Además, la metodología incorpora los paraderos con sistema de pre-pago, llamados zonas pagas, mediante el método de líneas comunes (Chriqui y Robillard, 1975). Mediante el uso de reglas simples las paradas de transferencia son diferenciadas de las paradas de actividades permitiendo así construir la estructura de viajes de cada usuario con sus respectivos viajes y etapas. Finalmente, la metodología construye una matriz Origen–Destino de los viajes observados en transporte público. Para contabilizar los viajes sin estimación de bajada se utilizan factores de expansión por servicio, parada y media hora.

Munizaga et al. (2014) continua la investigación validando la matriz Origen–Destino con tres fuentes: los mismos datos pasivos de forma endógena, una encuesta de viaje realizada por el Metro de Santiago y un pequeño grupo de voluntarios que accedieron a entregar su información. Los resultados fueron satisfactorios en todos los casos. Esta investigación aprovecha de implementar mejoras a la metodología y corregir errores. Las mejoras más relevantes incluyen la introducción de flexibilidad en la distancia de caminata máxima, la redefinición del corte entre un día y el siguiente según el momento de menor actividad en el sistema, la posibilidad de estimar la bajada de viajes a partir de información del día siguiente y una nueva metodología para identificar paradas de transferencias de paradas de actividades usando un criterio que compara la distancia en ruta con la distancia euclidiana de un viaje.

2.3. Enriquecimiento de datos pasivos

La mayor deficiencia de los datos pasivos frente a las encuestas de viajes es la falta de información complementaria al viaje como el propósito de viaje, información demográfica y socio-económica del portador de la tarjeta, indicadores de calidad de servicio u otros. En esta sección se presentarán aquellas investigaciones que han propuesto soluciones a esta problemática mediante el uso de reglas simples, modelos de clasificación y técnicas de minería de datos.

El propósito es quizás el atributo más relevante de un viaje. Este define el motivo por el cual el viaje se realiza, lo que permite caracterizar el uso del tiempo de las personas. Dada la gran importancia de este atributo se han encontrado tres metodologías para estimarlo en la literatura:

Devillaine (2012) propone una metodología que estima la ubicación, hora, duración y propósito de las actividades de los usuarios de transporte público. Utilizando reglas simples sobre el tipo de día, tipo de tarjeta, existencia de más viajes durante el día, duración entre una validación y la siguiente, ubicación de estas validaciones e información de uso de suelo. La metodología asigna los propósitos trabajo, casa, ocio u otros a las actividades de los

usuarios. Esta investigación se llevó a cabo con información de Santiago, Chile y Gatineau, Quebec, Canadá siendo posible observar diferencias en los patrones de uso de tiempo de los usuarios debido a variables sociológicas, culturales y geopolíticas. La comparación se realizó construyendo perfiles de uso de tiempo según propósito y tipo de día, observándose claras diferencias en la cantidad de transferencias realizadas, tiempos de viaje y duración de las actividades.

Haciendo uso de técnicas de minería de datos Kusakabe y Asakura (2014) crea una metodología de fusión de datos para estimar el propósito de viaje en Tokio, Japón mediante un clasificador bayesiano ingenuo calibrado previamente con las respuestas de una encuesta de viajes. El modelo es entrenado para distinguir entre viajes con propósito “viaje al trabajo”, “ocio o viajes con motivo de trabajo” y “vuelta a casa” a partir de la hora de llegada a un destino y la duración de la actividad en ese lugar. Una vez validado el modelo, se utiliza para estimar los propósitos de viaje de los usuarios del sistema de transporte público de Tokio. Esto permite observar el comportamiento de sus pasajeros en una ventana temporal de 20 meses.

Lee y Hickman (2014) estima el propósito de viaje en la zona metropolitana de Minneapolis, St. Paul utilizando un árbol de decisiones que se alimenta de tres fuentes: Información sobre los usuario, información temporal y espacial de las transacciones del sistema y la matriz Origen-Destino. Con esta información y el uso de reglas simples el árbol de decisiones asigna los propósitos viaje, trabajo, estudio y otros a los viajes de cada usuario. El árbol de decisiones es calibrado con una encuesta de viajes donde los propósitos son conocidos y posteriormente es aplicado a la base de datos del sistema de transporte público de Minneapolis. Los resultados permiten analizar los patrones de uso de tiempo de los usuarios en la ciudad. Además, el árbol de decisiones actualiza sus parámetros observando el comportamiento histórico de los usuarios. De esta forma el modelo de clasificación no queda obsoleto con el paso del tiempo.

Jovenes y adultos, hombres y mujeres o ricos y pobres no suelen comportarse de la misma forma al viajar. Los atributos demográficos y socio económicos definen las preferencias de los usuarios y las alternativas disponibles para viajar. Sin embargo, la privacidad de información de los viajes de las personas fuerza a que las tarjetas inteligentes sean anónimas, como en el caso de Santiago. Esta información es de mucho valor al realizar investigación para segmentar a la población y observar sus comportamientos por separado. Amaya (2013) propone una metodología para estimar la zona de residencia de los usuarios. La metodología asume que los usuarios pasan la noche en sus casas y, por lo tanto, su primera transacción del día se ubica cerca de su residencia. La metodología trabaja sólo con usuarios frecuentes, los que se definen como aquellos que utilizan el transporte público cuatro o más días a la semana. La estimación se hace observando la primera validación entre las 4:00AM y las 12:00PM de los días laborales de una semana. Si todas ocurren dentro de un radio razonable para caminar (500 metros), se asume que el usuario vive en esa zona. La ventana temporal fue definida en base a un análisis estadístico de los datos. El momento del día con menor uso de transporte público fue identificado entre las 3:15AM y las 4:00AM y por ende el término del ciclo natural de los viajes. De haberse considerado los viajes más tempranos se estaría observando viajes que corresponden al final de la estructura de viajes de ciertas personas. Por otra parte, los viajes después de las 12:00PM poseen una alta probabilidad de ocurrir

después de un viaje previo en la mañana en otro modo. Además, la investigación compara los perfiles de uso de tiempo según la zona de residencia estimada, propósito de viaje y tipo de día observándose resultados similares a los obtenidos por Olguín et al. (2009) con los datos de la Encuesta Origen–Destino 2001. Como una observación de la metodología, el parámetro de la distancia razonable de caminata debe ser calibrado según el nivel de agregación de la zonificación utilizada.

Una de las grandes ventajas que entregan los sistemas de transporte público con datos pasivos es que es posible calcular parámetros de calidad de servicio desagregados a muy bajo costo. Previo a los datos pasivos, para conocer el funcionamiento de un servicio era necesario realizar costosos estudios que midieran en terreno su rendimiento. Ahora, es posible registrar el funcionamiento de toda la flota de vehículos con un alto nivel de detalle casi de forma automática. Núñez (2015) hace uso de esta información para construir indicadores de calidad servicio con distintos alcances espaciales en la ciudad de Santiago. Previo al cálculo de los indicadores se estima el tiempo de transbordo y se identifican sus componentes (caminata y espera) por separado. El tiempo de caminata se determina a partir de la distancia euclidiana entre los paraderos de bajada y subida, mientras que el de espera como la resta entre el tiempo total de transbordo menos la caminata. A nivel de sistema presenta la evolución del número de etapas, la distancia viajada y el tiempo de viaje por periodos de media hora. A nivel comunal realiza un cruce entre la velocidad en ruta y la relación entre la distancia en ruta y euclidiana para todos los viajes realizados entre cada par de comunas, siendo posible identificar los pares de comunas con peor y mejor conectividad. Utilizando una zonificación desagregada de la ciudad, en cerca de 800 zonas, observa el número de etapas y el tiempo de viaje necesario para conectar una zona previamente seleccionada con el resto. Estos resultados permiten identificar los sectores con mayores problemas de accesibilidad al resto de la ciudad. Finalmente, a nivel de eje presenta las velocidades de automóviles y buses, observando en cuales ejes el funcionamiento del transporte público es mejor que el automóvil.

2.4. Regularidad en los patrones de viaje de los usuarios

La regularidad temporal y espacial de los usuarios al utilizar el sistema de transporte público ha sido siempre un tema de interés. Con la aparición de los datos pasivos surge la oportunidad de estudiar este fenómeno a mayor escala y a un menor costo. La precisión y abundancia temporal de los datos permite clasificar a los usuarios según su frecuencia de uso del sistema. Además, la georeferenciación de las validaciones y el registro de los servicios utilizados permite identificar patrones espaciales de los usuarios muy relevantes para la planificación del sistema. A continuación se presentan las investigaciones relacionadas con la clasificación de usuarios según su regularidad encontradas en la literatura.

En la ciudad de Gatineau, Quebec Morency et al. (2007) trabaja con 277 días de datos pasivos de la Société de Transport de l’Outaouais (STO) para medir la variabilidad en el uso del transporte público. En su investigación propone una metodología para obtener indicadores de regularidad espaciales y temporales a partir de las validaciones de subida diferenciando por tipo de tarjeta. El indicador espacial mide la frecuencia con que un usuario utiliza una parada en su primer viaje del día. El indicador temporal utiliza un algoritmo de clusterización

de k-medias para identificar los horarios donde los usuarios viajan más.

Haciendo uso de técnicas sofisticadas de minería de datos, Ma et al. (2013) propone el algoritmo “Density-Based Spatial Clustering of Application with Noise” para obtener resultados similares a los anteriores con datos provenientes de la ciudad de Beijing, China. El algoritmo, identifica las estructuras de viaje típicas en la red y luego clasifica a los usuarios en cinco niveles de regularidad dependiendo de la frecuencia con que sus patrones coinciden con los obtenidos en la red. El rendimiento del algoritmo propuesto es comparado con otros algoritmos de clasificación para demostrar su simplicidad y capacidad de ser duplicado en otros sistemas. Pese a que la comparación del rendimiento del algoritmo fue satisfactoria, más adelante Kieu et al. (2015) propondría modificaciones al algoritmo para mejorar su rendimiento sin realizar cambios en las salidas de la metodología.

En un enfoque más abstracto Langlois et al. (2016) utiliza datos del sistema de transporte público de Londres, Inglaterra para identificar estructuras de viaje típicas en la red. Las estructuras de viajes son inicialmente creadas observando las zonas de visita por cada usuario a lo largo de cuatro semanas. En un principio elabora la cadena de viajes de cada usuario, identificando las áreas de visita más frecuentes. Las áreas de visita son luego ordenadas según la duración de la actividad realizada en cada una. Una nueva clasificación agrupa a los usuarios por el tiempo destinado en cada área, lo que le permite observar once claros patrones de viajes. Observando la regularidad con que estos patrones visitan las áreas de interés durante un día, la semana y las cuatro semanas analizadas, el autor reconoce un perfil de usuario típico para cada patrón y asocia una posible actividad a cada zona. Finalmente, esto le permite clasificar a los usuarios de transporte público de Londres en once patrones de viaje y conocer sus perfiles de uso de tiempo.

2.5. Fidelidad y abandono del transporte público

Desde la introducción del nuevo sistema de transporte público, Transantiago, en la ciudad de Santiago, la fidelidad de los usuarios ha sido un tema de interés. Se cree que la frecuencia de buses, regularidad, niveles de hacinamiento, tiempos de espera en transbordos, entre otros factores, afectan la percepción del servicio de pasajeros. Cuando el transporte público de una ciudad no logra satisfacer las necesidades de sus usuarios, las tasas de migración aumentan. Los usuarios que reciben malos servicios buscan modos alternativos como la bicicleta, el taxi o el automóvil para realizar sus viajes. En esta sección repasaremos los artículos que estudian la fidelidad de los usuarios y la migración de estos.

Bass et al. (2011) realiza un completo análisis de los cambios en la partición modal de una pequeña muestra de usuarios de Santiago producto de la incorporación del Transantiago. Las particiones modales son observadas en cuatro cortes temporales; el caso base es dos meses antes de la puesta en funcionamiento del nuevo sistema, luego los cortes son llamados crisis, transición y estado estacionario a 3, 10 y 20 meses después de la implementación respectivamente. En el análisis se puede ver las diferencias en los tiempos de viaje, espera y caminata en bus y metro además de las particiones modales del transporte público, auto y otros. La información recopilada permite también entregar mediciones de indicadores de nivel

de servicio y de evasión por sectores. En una segunda parte, el estudio plantea un modelo de predicción de migración de usuarios utilizando un modelo multinomial anidado donde la primera decisión es si migrar o no y la segunda es a cual modo. El modelo permite concluir que las variables más relevantes para predecir la fuga son el ingreso, la edad, la calidad del servicio y los cambios en la disponibilidad de modos. En este caso se encontró que existe una mayor sensibilidad a la calidad del servicio que a los cambios en la tarifa por parte de los usuarios.

También en Santiago, Calvo (2015) estima la migración de usuarios utilizando un modelo de rezago a partir de dos cortes temporales semanales en abril y septiembre de un mismo año. La migración se define cuando una tarjeta observada en el primer periodo no realiza ninguna transacción en el segundo, lo que no es siempre correcto debido al extravío de tarjetas. A las tarjetas se les imputan variables socio económicas basadas en la estimación de sus zonas de residencia (Amaya, 2013) para describir a los usuarios. La muestra es separada por quintiles y luego por tipo de tarjeta (adultos y estudiantes) con el fin de encontrar comportamientos propios de cada estrato socio económico. De los resultados se puede confirmar varios supuestos generales del transporte público; el aumento en el tiempo de espera genera un aumento en la migración de usuarios, a mayor cantidad de transbordos los usuarios suelen abandonar más y que la tasa de motorización influye directamente con la tendencia a dejar el sistema. Además, se logran observar comportamientos más especiales como que los usuarios de comunas periféricas y los que realizan viajes cortos tienen una mayor probabilidad de migrar del sistema.

La fidelidad de los usuarios además de ser un buen indicador del rendimiento de un operador de transporte público, viene acompañada de otros beneficios. Los usuarios leales, independiente de las condiciones, tienden a ocupar el sistema en la mayoría de sus viajes y estos mismos suelen recomendar el uso del transporte público a otros potenciales usuarios. Imaz et al. (2015) utiliza información de encuestas de preferencias declaradas y preferencias reveladas realizadas en Toronto, Canadá para identificar los factores que afectan la fidelidad con el transporte público. Estas encuestas fueron diseñadas especialmente para poder incluir información psicométrica en el análisis. En el estudio se desarrollan dos modelos; el primero considera sólo los efectos fijos del comportamiento de los usuarios, en el segundo se incluyen los efectos aleatorios mediante los atributos socio económicos y psicométricos. En este último modelo, sólo un pequeño conjunto de las variables fueron significativas y sus signos eran contra intuitivos o la explicación de su significancia era problemática, obligando a desechar el modelo. Del primer modelo se concluye que al verse empeorados los atributos de calidad de servicio (tiempo de viaje, tiempo de espera, tiempo de transferencia, atrasos, dos o más transferencias) existe un impacto negativo en la fidelidad al sistema, que los usuarios son particularmente sensibles al nivel de hacinamiento, que tanto la puntualidad y la reducción de transferencias tienen un efecto positivo en la fidelidad, que existe una preferencia significativa para viajar al trabajo en metro y que la tarifa posee un impacto más moderado que las demás variables. Estas conclusiones permiten afirmar que no es necesario poseer una gran cantidad de información sobre los usuarios para construir buenos modelos. De hecho, las variables que se utilizaron en este trabajo son todas obtenibles en sistemas con recolección automática de datos.

Varias son las investigaciones que realizan estimación de fuga de clientes sin el uso de

encuestas. Bagchi y White (2005) entregó una definición del problema de la tasa de recambio y fue capaz de calcularla. Sus resultados no son muy confiables debido al bajo nivel de penetración de la tarjeta inteligente pero aun así marcó una línea de investigación importante. Trépanier y Morency (2010) trabaja con seis años de datos provenientes de la red de la Société de transport de l'Outaouais (STO) en un estudio a largo plazo del tiempo de vida de las tarjetas. Este trabajo presenta nuevos conceptos y métodos para calcular la fidelidad de sus usuarios. Estos nuevos conceptos son la tasa de migración, la tasa de retención, los perfiles de supervivencia y la tasa de utilización. La tasa de retención mide la proporción de usuarios que sigue activa en el sistema después de un determinado lapso de tiempo. Se pueden observar diferencias según tipo de tarjeta donde los estudiantes escolares y universitarios presentan las menores tasas teniendo una retención solo del 20 % dos años después de su creación. Las tarjetas adultas poseen una retención cercana al 70 % incluso después de cuatro años. La época del año en que las tarjetas son creadas también afecta en la retención, notándose una fuerte disminución en el grupo creado en enero comparado con los de septiembre. Los perfiles de supervivencia muestran el total de usuarios en el sistema por mes. En estos perfiles se puede apreciar como los usuarios que van dejando el sistema se ven suplidos por nuevos usuarios. La tendencia global del sistema es al alza, lo que según el autor depende de la época del año así como también del tipo de tarjeta. La tasa de utilización es calculada como la división entre la cantidad de meses donde se haya utilizado la tarjeta y la vida total de una tarjeta. Los resultados de la tasa de utilización muestran que existe una gran diferencia según el momento del año en que la tarjeta fue creada. Las tarjetas creadas en el primer cuarto del año poseen una tasa de utilización cercana al 60 % en cambio las tarjetas creadas en el último cuarto un 80 %. Las tarjetas de estudiantes son indiferentes a la época del año en que son creadas manteniendo una utilización del 85 % en todos los casos. La investigación también presenta un modelo de regresión múltiple que explica la vida de las tarjetas. Los resultados indican que las variables que poseen un mayor impacto en la capacidad predictiva del modelo son el mes de creación (mayor duración en agosto y enero), año de creación (gradualmente correlacionada), área de residencia (viajes más directos poseen mayor retención) y tipo de tarjeta (tarjetas con pago automático bancario viven más).

Continuando con la investigación, Trépanier et al. (2012) presenta un modelo proporcional de riesgo a partir de cinco años de la red de la STO. El concepto de utilizar un modelo proporcional de riesgo surge de que la probabilidad de seguir ocupando una tarjeta en un cierto periodo es condicional a la probabilidad de que esa tarjeta no sea cancelada en un periodo anterior. A cada tarjeta se le estima una zona de residencia y luego se le imputan atributos socio económicos agregados por comuna. El modelo permite concluir que la fidelidad de los usuarios adultos se ve afectada positivamente cuando existen áreas con alta densidad residencial y alta partición modal de transporte público, que los usuarios más jóvenes tienden a tener una mayor fidelidad, que en las comunas con altas tasas de desempleo habitan usuarios poco leales y que cuando existen modos alternativos la fidelidad al sistema disminuye.

En Chicago, Figler et al. (2011) utiliza información de la Encuesta de Satisfacción de Usuarios del 2008 realizada por la Autoridad de Tránsito de Chicago para analizar la lealtad de los usuarios con el sistema. Primero la investigación destaca los aspectos del sistema que necesitan ser mejorados con un análisis de cuadrantes. Luego buscan explicar la lealtad de los usuarios al servicio de buses mediante un análisis de regresión lineal multivariado. Los principales resultados muestran que los usuarios que tengan una mejor percepción del sistema

serán más leales, que los usuarios que hayan tenido algún tipo de problema o accidente con el servicio dejaran de utilizarlo y, por último, que la confiabilidad del sistema y la seguridad en el paradero y al interior del vehículo son buenas variables para explicar la lealtad de los usuarios.

Kim y Ulfarsson (2012) estudia el compromiso de los usuarios con un sistema de trenes ligeros en la región metropolitana de St. Louis, USA. Utilizando los resultados de una encuesta realizada entre 2005 y 2006 a los usuarios del sistema de trenes, desarrolla un modelo logit multinomial para explorar las diferencias entre los usuarios que declararon estar seguros de seguir utilizando el sistema en el futuro y los que presentaron dudas. Los resultados indican que los Afro-Americanos, las mujeres jóvenes y los pasajeros de un ingreso medio por hogar tienen una mayor probabilidad de expresar dudas sobre su futuro uso del sistema. En cambio, los usuarios que combinan el uso de trenes ligeros con vehículos particulares y los que expresan un alto nivel de satisfacción con respecto al sistema, poseen una alta probabilidad de expresar seguridad sobre su futuro uso del sistema.

2.6. Modelo Logit Ordinal

En el estudio de la modelación de elecciones discretas, Train (2002) presenta una forma muy conveniente de modelar una variable ordinal. Primero, explica que las variables ordinales pueden ser modeladas utilizando modelos logit anidados, logit mixtos o probits que consideren las similitudes y diferencias entre alternativas al especificar una mayor correlación entre las respuestas contiguas. Sin embargo, pese a que este tipo de especificaciones pueden entregar buenos resultados, en realidad no se ajustan a la estructura de los datos. En vez de asumir que existen distintos niveles de utilidades para cada posible respuesta, los modelos logit ordinales consideran una única utilidad que puede ser representada por distintos niveles. Por ejemplo, en el caso de la siguiente pregunta: ¿Cómo calificas el desempeño del presidente en el último año?

- Muy buen desempeño
- Buen desempeño
- Ni bueno ni malo
- Mal desempeño
- Muy mal desempeño

Se asume que el encuestado tiene una opinión sobre el desempeño del presidente. Esta opinión es representada por una variable (no observable) que llamamos U , donde a mayores valores de U la persona considera que el presidente ha tenido un mejor desempeño y a menores valores de U un peor desempeño. Para contestar la pregunta, a la persona se le pide que exprese su opinión en una de las cinco categorías mencionadas. La persona responde en base al valor de su U . Si U está por encima de un punto de corte, que llamamos θ_1 , la persona responde “Muy buen desempeño”. Si U está por debajo de θ_1 y por encima de un θ_2 , entonces contesta “Buen desempeño” y así sucesivamente. La respuesta queda representada como:

- “Muy buen desempeño” si $U > \theta_1$.

- “Buen desempeño” si $\theta_1 > U > \theta_2$.
- “Ni bueno ni malo” si $\theta_2 > U > \theta_3$.
- “Mal desempeño” si $\theta_3 > U > \theta_4$.
- “Muy mal desempeño” si $\theta_4 > U$.

Luego la opinión, U , de la persona es descompuesta en atributos observables y no observables: $U = \beta'x + \varepsilon$. Los atributos no observables son considerados aleatorios y por simplicidad se asume poseen una distribución logit. De esta forma, la función de distribución acumulada de ε queda descrita por $F(\varepsilon) = \exp(\varepsilon)/(1 + \exp(\varepsilon))$. Entonces, la probabilidad de que la persona conteste “Muy mal desempeño” es:

$$\begin{aligned}
\text{Prob}(\text{“Muy mal desempeño”}) &= \text{Prob}(U < \theta_4) \\
&= \text{Prob}(\beta'x + \varepsilon < \theta_4) \\
&= \text{Prob}(\varepsilon < \theta_4 - \beta'x) \\
&= \frac{e^{\theta_4 - \beta'x}}{1 + e^{\theta_4 - \beta'x}}
\end{aligned}$$

La probabilidad de “Mal desempeño” es:

$$\begin{aligned}
\text{Prob}(\text{“Mal desempeño”}) &= \text{Prob}(\theta_4 < U < \theta_3) \\
&= \text{Prob}(\theta_4 < \beta'x + \varepsilon < \theta_3) \\
&= \text{Prob}(\theta_4 - \beta'x < \varepsilon < \theta_3 - \beta'x) \\
&= \text{Prob}(\varepsilon < \theta_3 - \beta'x) - \text{Prob}(\varepsilon < \theta_4 - \beta'x) \\
&= \frac{e^{\theta_3 - \beta'x}}{1 + e^{\theta_3 - \beta'x}} - \frac{e^{\theta_4 - \beta'x}}{1 + e^{\theta_4 - \beta'x}}
\end{aligned}$$

El resto de las probabilidades son calculadas de forma análoga. Finalmente, las probabilidades son resueltas maximizando la función de verosimilitud, lo que estima los valores de β , que explica el impacto de las variables explicativas en la opinión de las personas sobre el desempeño del presidente, y los puntos de corte $\theta_1 \dots \theta_4$.

Capítulo 3

Análisis de las bases de datos

3.1. Descripción del caso de estudio

El sistema de transporte público Transantiago, disponible en Santiago desde principios del 2007, es un sistema multimodal integrado entre buses y metro que sirve una población cercana a los 6,6 millones de habitantes. En total existen 6.600 buses incorporados con tecnología GPS que sirven 378 servicios troncales y alimentadores. La operación de estos está asignada a siete empresas concesionarias que operan entre 400 y 1.200 buses cada una. La red de trenes es operada por la empresa estatal Metro de Santiago. La coordinación y supervisión del sistema está a cargo de un ente estatal llamado Directorio de Transporte Público Metropolitano (DTPM) que depende del Ministerio de Transporte y Telecomunicaciones. En términos de infraestructura el sistema cuenta con 72 *km* de pistas segregadas, 200 *km* de pistas solo bus, más de 11.000 paradas de buses y 125 estaciones con sistema de pre-pago. El metro está compuesto por 108 estaciones distribuidas en cinco líneas integradas. El sistema cuenta con 186 trenes que cubren una red de 104 *km*. El único sistema de pago aceptado en buses es la tarjeta inteligente “bip!”, lo que ha permitido que su tasa de penetración sea de un 97%. El sistema tarifario es plano por viajes, permitiendo realizar hasta tres transbordos en un lapso de dos horas. En metro existen pequeñas diferencias en la tarifa dependiendo el periodo del día. Esta diferencia se cobra al combinar entre modos cuando corresponda (DTPM, 2017).

La incorporación de las tarjetas inteligentes en el Transantiago ha permitido que la recolección de pagos se haga de forma automática por una empresa externa. Esto facilita la operación por parte de los conductores y disminuye los tiempos de carga de pasajeros. En Transantiago es necesario validar solamente al iniciar un viaje o etapa, lo que se puede dar de tres formas; en metro al ingresar al andén de la estación, en buses al abordar el vehículo o al acceder a los paraderos con sistema de pre-pago donde posteriormente no es necesario validar en el bus. A diferencia de otros sistemas de transporte, Transantiago no considera la distancia recorrida por el usuario para determinar la tarifa a cobrar, por lo tanto, no es necesario validar al salir del sistema. Esto permite conocer el origen del viaje pero no su destino.

En el sistema operan distintos tipos de contrato de tarjetas. Las tarjetas de adultos están

separadas en cuatro categorías: Valor y Contratos de Postpago 2, 3 y 9. Las más comunes son las denominadas Valor que representan a un adulto regular sin ningún beneficio. Estas tarjetas pueden ser compradas en cualquier punto de carga habilitado y cargadas con crédito para su uso inmediato. Las tarjetas de Postpago 2 y 3 corresponden a tarjetas CMR Falabella que permiten realizar hasta un máximo de ocho viajes diarios y deben ser canceladas a fin de mes. Las tarjetas de Postpago 9 corresponden a un tipo especial perteneciente al Administrador Financiero de Transantiago (AFT).

El sistema contempla un descuento de un 70 % o 100 % para los estudiantes según sea el caso. Los pases escolares están clasificados en siete categorías: Escolar Básica tienen un 100 % de descuento. Media Diurno, Superior Diurno, Escolar de Recambio Básica, Escolar de Recambio Media - Superior, Escolar Superior Extensión Nuevo y Estudiante en Práctica tienen un 70 % de descuento. Todos los estudiantes de establecimientos Municipales y Subvencionados tienen acceso a pases escolares de los tipos Escolar Básica y Media Diurno a partir de quinto básico dependiendo del nivel que estén cursando. Los estudiantes de colegios particulares pueden solicitar su pase escolar si su ingreso familiar pertenece a cualquiera de los primeros cuatro quintiles de ingreso y han sido previamente inscritos por su centro educacional en la Juneab (organismo encargado de las becas escolares). El resto de los estudiantes pertenecientes a colegios particulares no posee ningún tipo de beneficio y por lo tanto, deben usar una tarjeta de adulto. Los alumnos de establecimientos de educación superior como centros de formación técnica, institutos profesionales y universidades pueden acceder a pases escolares del tipo Superior Diurno. El resto de las categorías son entregadas a estudiantes en práctica o de forma temporal por extravío.

Al cumplir los 65 años de edad, las personas reciben beneficios de tercera edad y pueden solicitar una tarifa diferenciada para viajar en metro durante los horarios fuera de punta en días laborales y en cualquier horario los días sábado y domingo. En la actualidad estos beneficiarios reciben una tarjeta especial que les permite funcionar como el resto de los usuarios pero anteriormente estas validaciones eran realizadas con boletos, lo que impedía acceder a los beneficios de integración tarifaria con buses y también imposibilitaba observar estas validaciones en la base de datos.

3.2. Pre procesamiento de datos

Los datos disponibles para esta tesis provienen de la unión de tres fuentes. Los pulsos GPS integrados en los buses, las transacciones bip! realizadas en los validadores de cada modo e información complementaria sobre las rutas de los buses, la asignación de buses a las rutas, la ubicación de los paraderos y la ubicación de las estaciones de metro. Estos datos son procesados por un conjunto de metodologías que permiten enriquecer la base de datos. Inicialmente los pulsos GPS son asignados a la red de rutas de los servicios y a las transacciones se les asigna un paradero y hora basado en la interpolación de los pulsos GPS. Posteriormente se identifican las transacciones de transbordo de las de actividades, se estima la bajada de cada transacción, se calculan indicadores de calidad de servicio y se estima una zona de residencia a las tarjetas que viajan cuatro o más días a la semana. Esta sección de la tesis tiene como propósito dar a entender al lector como estas metodologías son capaces

de estimar los atributos recién mencionados.

Antes de continuar, es necesario definir los conceptos de viaje y etapa. Un viaje es la acción de cambiar de ubicación desde un origen a un destino donde una cierta necesidad será satisfecha al realizar una actividad. Cada transferencia realizada al interior de un viaje define el fin e inicio de una etapa. Así, un viaje puede estar conformado por una o más etapas que pueden ocurrir en los distintos modos disponibles. Mediante el uso de reglas simples, la metodología es capaz de identificar las etapas al interior de un viaje observando la ubicación, la hora, la duración de la actividad y el tipo de día. Esta información es fundamental para el uso de los datos, ya que permite entender la estructura de viajes de las tarjetas.

3.2.1. Estimación de bajada

La metodología de estimación de bajada está basada en el trabajo realizado por Munizaga y Palma (2012). La idea fundamental detrás de esta metodología es que el destino de un viaje y/o etapa está relacionado con el origen del siguiente. De esta forma lo que se busca es encontrar el lugar donde un usuario debiera bajarse del vehículo que abordó tal que se encuentre lo más cercano al origen de su próxima validación. En el caso de la última validación del día la metodología considera la primera validación del mismo día o la del día siguiente como el destino. El problema a resolver varía dependiendo del modo en que se hizo el viaje (bus, metro, zona paga) ya que la información disponible es distinta en cada caso. Cuando el problema no encuentra una solución, se asume que la persona utilizó otro modo de transporte para realizar un viaje intermedio.

En el caso del bus, la transacción se realiza en un validador asociado a una expedición única. Por lo tanto, se conoce el servicio, sentido, parada de subida y la hora de la validación. De esta forma se conocen el origen del viaje, el recorrido de la expedición y los horarios en que esta expedición circulará por cada uno de sus paraderos. Luego se minimiza el tiempo generalizado de viaje hasta la ubicación de su siguiente transacción según el problema presentado por la Ecuación 3.1.

$$\begin{aligned} \text{mín}_i &= t_i + f_w \cdot \left(\frac{d_{i+1}}{s_w} \right) \\ \text{s.a.} & \quad d_{i+1} \leq d_{max} \end{aligned} \tag{3.1}$$

Donde:

- t_i : Tiempo asociado a la posición i .
- f_w : Factor de penalización; desutilidad de caminar.
- d_{i+1} : Distancia geográfica euclidiana entre la posición i y el lugar de la siguiente validación.
- s_w : Velocidad media de caminata.
- d_{max} : Distancia euclidiana máxima de caminata.

En el caso de metro se conoce la ubicación de la estación de origen y el horario de la validación. Luego se busca aquella estación en la red que minimice la distancia hasta el origen de la siguiente transacción. Como los trenes no poseen dispositivos GPS, el tiempo que toma llegar a cada estación se calcula con el algoritmo de rutas mínimas de Dijkstra utilizando como

parámetros los tiempos de viaje, espera y caminata entre estaciones calculados previamente.

Cuando el usuario accede a zonas paga se conoce el horario en que la persona llegó al paradero y su respectiva ubicación, pero se desconoce el servicio y la hora en que lo aborda. Para estimar esta información se acude a la metodología de Líneas Comunes (Chriqui y Robillard, 1975) que busca el conjunto de líneas que minimizan la esperanza del tiempo de viaje generalizado hasta la posición de la siguiente transacción. De esta forma se asume que el usuario aborda el primer bus, después de su llegada, que pertenezca a alguna de las estrategias validas entregadas por la metodología de Líneas Comunes y posteriormente se resuelve el mismo problema de optimización que en el caso de los buses.

Una de las mayores ventajas de esta metodología es que minimiza una función del tiempo generalizado de viaje que considera el tiempo de viaje en el vehículo y el de caminata necesarios para llegar al destino. Esto permite que su capacidad predictiva sea mucho mejor que si solo se considerara la distancia o el tiempo en el vehículo. La metodología ha sido validada por tres fuente de información y es capaz de identificar exitosamente un 80 % de las bajadas de todas las transacciones (Munizaga et al., 2014).

3.2.2. Cálculo de indicadores de calidad de servicio

Una vez conocidas las etapas al interior de un viaje y sus respectivas bajadas, el cálculo de los atributos de calidad de servicio es directo. El tiempo de viaje es la diferencia entre el horario de bajada y el horario de la transacción, el tiempo de caminata se calcula a partir de la distancia entre la bajada y el origen de la siguiente transacción con una velocidad media de caminata y el tiempo de espera es el tiempo restante hasta que ocurra la siguiente transacción. El problema de calcular estos atributos de esta forma es que no se consideran las particularidades del metro y las zonas pagas. El problema está, en que la metodología considera que el tiempo de viaje comienza al momento de realizarse la transacción, como es normal en buses. En metro la validación ocurre en un torniquete desde donde aun falta caminar hasta el andén y luego esperar por un tren. En las zonas pagas la validación ocurre en el mismo paradero así que no es necesario agregar un tiempo de caminata pero si considerar la espera hasta la llegada del bus. La metodología descrita por Núñez (2015) considera estos problemas y dependiendo del tipo de transbordo corrige los tiempos de viaje, caminata y espera cuando corresponda. Tiempo de acceso se le denomina al tiempo necesario para caminar desde el torniquete de una estación de metro a su andén. El tiempo de egreso considera el tiempo de caminata entre el andén de una estación y el exterior de esta. Estos dos tiempos fueron calculados para cada estación de la red previamente y luego añadidos a la metodología cuando se realizan transferencias desde buses a metro o metro a buses respectivamente. Además, se agregó el tiempo de espera en andenes de metro y en paraderos de zonas pagas. Estas correcciones permiten eliminar observaciones con tiempos de espera negativos, tiempos de viaje sobrestimados y tiempos de caminata y espera subestimados. Lamentablemente, por problemas técnicos no fue posible aplicar estas correcciones a los datos disponibles en esta tesis, por lo tanto, los atributos de calidad de servicio son calculados en su forma original.

3.2.3. Estimación de residencia

Para estimar la residencia del portador de una tarjeta se utiliza la metodología propuesta por Amaya (2013). El supuesto detrás de esta metodología es que las personas pasan la noche en sus hogares y, por lo tanto, su primera validación del día ocurre en la cercanía de su residencia. La metodología considera solamente a los usuarios frecuentes, definidos como aquellos que viajan cuatro o más días a la semana. La estimación se hace observando la primera validación de cada día laboral entre las 4:00AM y las 12:00PM. Si estas validaciones se encuentran dentro de un radio razonable para caminar (500 metros), se asume que el portador de la tarjeta vive en esa zona. La ventana temporal fue definida posterior a un análisis estadístico de las transacciones en la ciudad de Santiago. El periodo entre las 3:15AM y las 4:00AM fue identificado como el de menor uso del transporte público y por lo tanto el fin del ciclo natural de las estructuras de viajes de las personas. De esta forma, al considerar viajes más tempranos, aumenta la probabilidad de que el origen de ese viaje no sea en la cercanía del hogar. De la misma forma, viajes después del medio día poseen una alta probabilidad de ocurrir después de un viaje previo en otro modo no observado en la base de datos como el taxi, el automóvil o la bicicleta. La distancia razonable de caminata depende de la zonificación utilizada, en el caso de Santiago se utiliza una zonificación comunal y se ha observado que al aumentar la distancia de 500 a 1.000 metros, la cantidad de nuevas estimaciones correctas es mayor a las estimaciones erróneas. La metodología fue validada con una encuesta de viajes de metro, obteniéndose un 70 % de estimaciones correctas. Los individuos que combinan con otros modos de transporte como el kiss-and-ride o el park-and-ride al abandonar sus hogares suelen tener estimaciones erróneas o no recibir una estimación.

En la Figura 3.1 se observa un ejemplo de una estimación exitosa. La tarjeta muestra cuatro validaciones en días laborales entre las 4:00AM y las 12:00PM. La primera validación de la semana (lunes) es utilizada como punto de referencia para delimitar el radio razonable de caminata. Como el resto de las transacciones se ubican al interior de este círculo, se asume que el portador de la tarjeta vive en esta comuna.

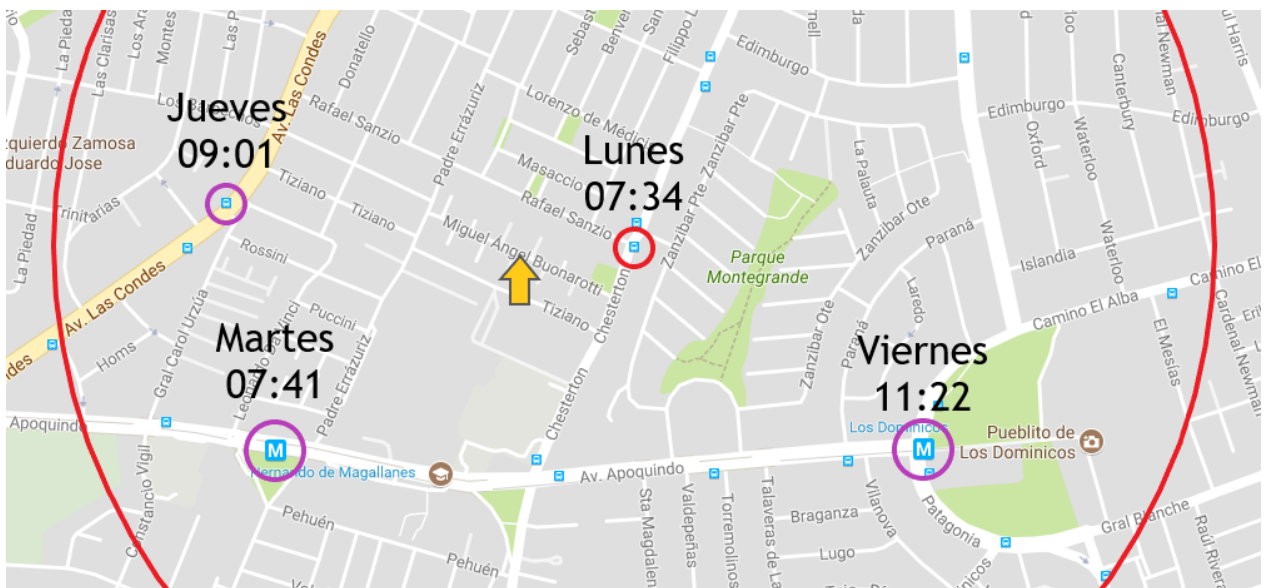


Figura 3.1: Ejemplo de una estimación de residencia exitosa.

3.3. Descripción de la base de datos

El objetivo de esta tesis es analizar el grado de uso del sistema y la posible migración de usuarios. Para llevar a cabo esta tarea se utiliza un modelo de rezago que basa la decisión de los usuarios en el periodo actual, en la experiencia de viaje percibida en un periodo anterior. El primer periodo es de ocho días consecutivos entre el 18 y el 25 de abril del 2015, donde solo se repite el día sábado. El segundo periodo es el mes de noviembre del mismo año. Los datos del primer periodo han sido previamente procesados por las metodologías mencionadas, en cambio, el segundo periodo solo consiste de el registro de las transacciones. Las diferencias en la extensión y tipos de datos entre los periodos es a propósito y la razón de esto es el objetivo de cada periodo para la modelación. El primer periodo busca describir la experiencia de viaje percibida por los usuarios, en cambio, el segundo periodo solo observar la frecuencia de uso del sistema. De esta forma, las variables explicativas serán construidas a partir de los datos del primer periodo y la variable a explicar del segundo.

Para homogeneizar la muestra se trabaja solamente con las tarjetas de Adulto Regular. El resto de las tarjetas poseen un comportamiento particular de cada categoría y de ser incluidas podrían generar ruido en los resultados de los modelos. La Figura 3.2 muestra la distribución de tarjetas observadas en el primer periodo. Las tarjetas de Adulto Regular son las más comunes con 2.671.062 tarjetas, lo que equivale a un 77,6 % del total. La siguen las tarjetas de estudiantes de Educación Superior, Educación Media y Educación Básica con 431.305 (12,5 %), 164.860 (4,8 %) y 105.022 (3,1 %) respectivamente. Las tarjetas de Contratos de Postpago y el resto de las tarjetas de estudiantes poseen 40.601 (1,2 %) y 28.454 (0,8 %).

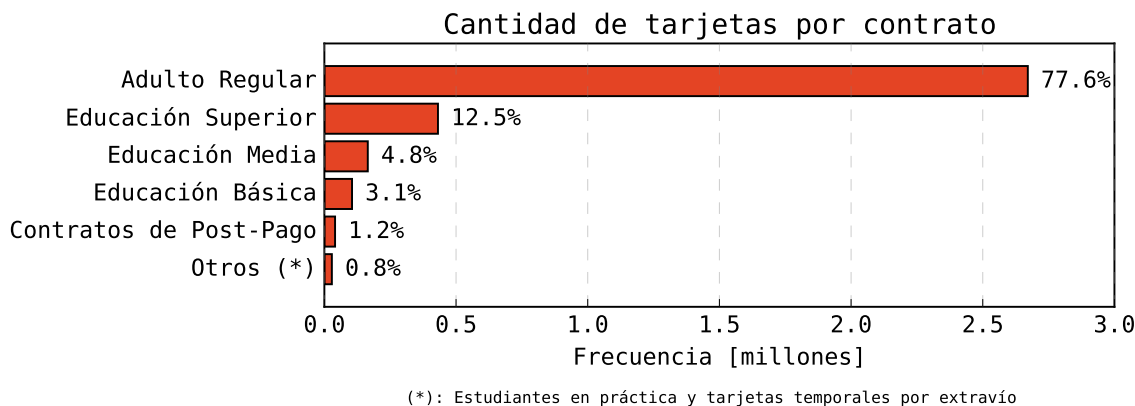


Figura 3.2: Distribución de tarjetas por tipo de contrato.

El resultado del procesamiento de los datos pasivos del Transantiago son dos tablas agrupadas por etapas y viajes. En la primera tabla cada observación representa una transacción única, en cambio en la segunda una observación puede considerar hasta cuatro transacciones (etapas) que en conjunto representan un viaje. Estas tablas se utilizan en el primer periodo de análisis y poseen los atributos necesarios para describir la experiencia de un usuario en el transporte público. El detalle de los atributos de cada una se puede ver en los Anexos A.1 y A.2. A continuación, se presentan los atributos más importantes de cada tabla:

- Identificador de tarjeta.

- Identificador de la etapa.
- Identificador del viaje.
- Tipo de contrato de la tarjeta.
- Coordenadas X e Y de subida y bajada.
- Paradas de subida y bajada.
- Fecha y hora de subida y bajada.
- Servicio y sentido utilizado.
- Estimación de tiempos de viaje, caminata y espera.

Además, varias otras variables son construidas a partir de estas como el tipo de transporte, el tipo de día, el periodo del día de subida y bajada, la comuna de subida y bajada, etc. Un 20 % de las transacciones no poseen una estimación de bajada y por lo tanto ninguno de los atributos relacionados con el destino del viaje o los indicadores de calidad de servicio.

El segundo periodo es más extenso pero no posee tanta información. Los únicos atributos disponibles en este son:

- Identificador de tarjeta.
- Fecha y hora de subida.
- Tipo de contrato de la tarjeta.

En la Tabla 3.1 se pueden observar los tamaños de las tablas en los dos periodos del análisis. El segundo periodo posee más observaciones y tarjetas únicas dado que la ventana temporal es más extensa. La cantidad de viajes no se puede conocer en el segundo periodo ya que para esto es necesario procesar los datos.

Tabla 3.1: Descripción de las bases de datos.

	18–25 Abril 2015	Noviembre 2015
Número de Etapas	24,4 millones	95,8 millones
Número de Viajes	17,3 millones	-
Tarjetas Adulto Regular	2,7 millones	4,1 millones

3.4. Limpieza de datos

Como la mayoría de las bases de datos, esta no está exenta de errores y falta de información. Las causas más comunes de estos errores son el mal funcionamiento de los instrumentos utilizados como los pulsos GPS o los validadores de tarjetas y los errores humanos como la mala asignación de un servicio a su validador. Además, existen comportamientos especiales de los usuarios que las metodologías no son capaces de explicar, como el caso de una persona validando dos veces en un mismo validador con la misma tarjeta, con el fin de pagar por el viaje de dos personas. Debido a todas estas fuentes de errores es que es importante realizar una limpieza de los datos.

En la base de datos existen tarjetas con una cantidad de viajes y etapas sobre lo esperado. Es posible que estas tarjetas sean utilizadas con propósitos especiales en el transporte público y no representen el comportamiento de un usuario. Para estos casos se creó una metodología de detección de observaciones atípicas. En un principio, no basta con contabilizar los viajes de cada tarjeta y eliminar aquellas consideradas outliers por algún criterio, ya que es necesario considerar el tiempo que le tomó al usuario realizar los viajes. Por lo tanto, las tarjetas son agrupadas por la cantidad de días viajados y luego al interior de cada grupo se eliminan aquellas tarjetas que posean más viajes/etapas que la media más tres veces la desviación estándar del conjunto. La Tabla 3.2 muestra la cantidad máxima de viajes y etapas permitida por días viajados.

Tabla 3.2: Máxima cantidad de viajes y etapas permitidas por días viajados.

Días Viajados	1	2	3	4	5	6	7	8
Viajes Máximos	4,5	7,0	10,5	13,0	18,3	21,4	24,6	26,5
Etapas Máximas	5,4	10,4	16,2	20,7	30,3	39,2	46,4	50,2

En el caso de las transacciones en buses las coordenadas de estas son estimadas según la interpolación de la ubicación y la hora de los dos pulsos GPS más cercanos a la hora de la transacción. Cuando los pulsos GPS entregan información errónea, la ubicación de las transacciones no es correctamente estimada y tampoco su asignación a un paradero. En los datos se han observado casos donde las coordenadas de las transacciones ocurren fuera de los límites de la Región Metropolitana. Para evitar estos casos se ha delimitado un rectángulo que contiene a todos los paraderos establecidos por DTPM en su interior con un margen de error de 100 metros. Cualquier validación con coordenadas fuera de este rectángulo es eliminada de la base de datos.

Los tiempos de viaje, espera y caminata estimados por las metodologías previamente mencionadas también pueden contener errores. Estos pueden ocurrir por errores en la estimación del paradero de bajada o por comportamientos no esperados como correr o usar modos de transporte alternativos tales como la bicicleta durante una transferencia. Cuando esto ocurre, los tiempos de espera compensan a los sobrestimados tiempos de caminata, pudiendo llegar a tener valores negativos. Las etapas con tiempos de espera negativos fueron reemplazadas por “sin observación” para no perder información relevante sobre el resto del viaje. Además, existe un 20 % de las etapas que no poseen estimación de bajada. En estos casos no es posible calcular los tiempos de viaje, espera y caminata. Para atender estos problemas, después de agrupar la base de datos por tarjeta (para su posterior modelación), aquellas que tengan algún campo sin información son eliminadas del análisis.

3.5. Análisis de la base de datos

En esta sección se presenta un análisis de las dos bases de datos en su estado desagregado por viajes y etapas. El propósito de este análisis es dar a conocer al lector los alcances de los datos disponibles. En el caso del primer periodo, se priorizará la desagregación de las validaciones por los atributos obtenidos del pre procesamiento de los datos. En el segundo

periodo, al disponer de un periodo más extenso, se hará un análisis más estadístico de la cantidad de transacciones observadas durante el mes.

3.5.1. Primer Periodo: 18 - 25 Abril, 2015

El análisis comienza en su forma más agregada, observando la cantidad de viajes que se realizan en un día laboral promedio (5 observaciones), un sábado promedio (2 observaciones) y un domingo. En la Figura 3.3 se muestra la cantidad de viajes por tipo de día y número de etapas. Claramente el día laboral es el que concentra la mayor cantidad de viajes con sobre 5 millones. El día sábado posee un poco más de la mitad de los viajes de un día laboral con aproximadamente 2.9 millones y finalmente los días domingo el sistema funciona con menor demanda, cerca de 1.75 millones de viajes. Además, se puede observar la cantidad de viajes que constan de una, dos y tres etapas. La relación entre estos viajes permanece constante entre los tres tipos de día. Aproximadamente 64 % de los viajes son de una etapa, 29 % de dos etapas y 6 % de tres etapas. Existe un 1 % de viajes que la metodología asignó con cuatro o más etapas. Estos no fueron incluidos en la figura debido a su pequeña parte del total.

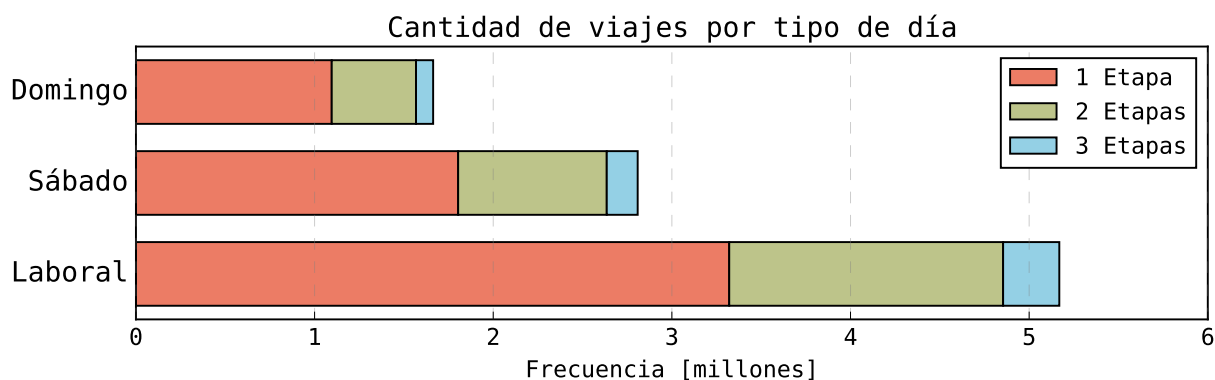


Figura 3.3: Número de viajes por tipo de día.

La abundancia y precisión con que se registra la hora de las transacciones en el sistema hacen de los datos una fuente de información ideal para observar la estructura de viajes en la ciudad. En la Figura 3.4 se presenta la cantidad de transacciones realizadas cada quince minutos diferenciando por día. Lo primero que se observa en la figura es la existencia de tres estructuras de viaje distintas; día laboral, sábado y domingo. En un día laboral la actividad comienza a las 5:00AM aumentando rápidamente hasta llegar a la punta de la mañana a las 8:00AM, donde hay sobre 100.000 transacciones cada quince minutos. Dos horas después el uso del transporte público se regulariza y fluctúa cerca de las 40.000 transacciones cada quince minutos con una leve punta a la hora de almuerzo (1:30PM). A las 5:00PM vuelve a aumentar el número de transacciones a causa del término de la jornada laboral y escolar. La punta tarde se alcanza alrededor de las 6:30PM y esta es levemente menor que la de la mañana. Por último, el uso del sistema decae al ir terminando el día. El día viernes presenta un comportamiento distinto. La punta tarde comienza a las 3:30PM, 1 hora y media antes, y su punta no es tan elevada como el resto de los días laborales, superando las 80.000 transacciones cada quince minutos. Los días sábados (18 y 25 de abril) la estructura de viaje comienza más

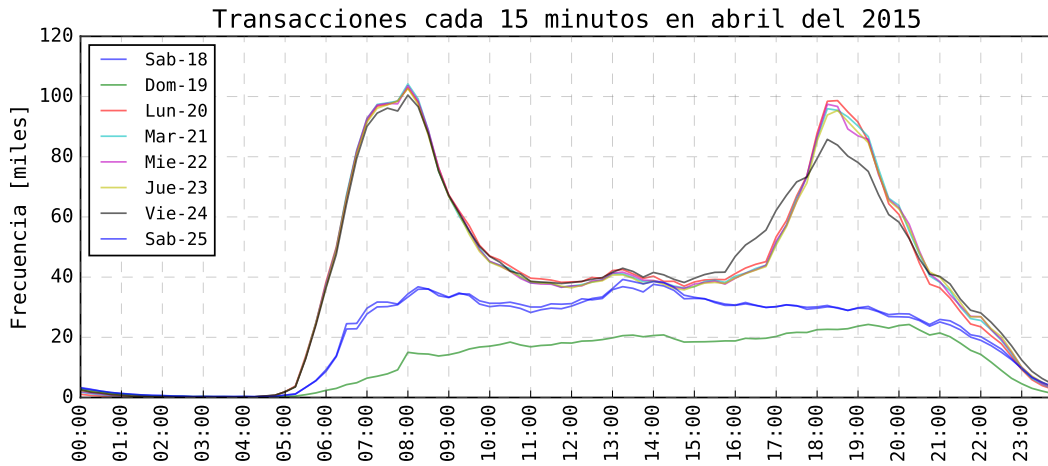


Figura 3.4: Distribución temporal de las transacciones en Santiago.

tarde y no posee punta en la mañana ni en la tarde. La cantidad de transacciones se mantiene entre las 30.000 y 40.000 cada quince minutos con un leve incremento después del medio día a causa de viajes con propósitos relacionados al almuerzo (1:00PM - 2:45PM). El domingo es el día de menor uso del transporte público. Su uso parte cerca de las 6:30AM y tiene un lento aumento durante el día hasta llegar a las 21.000 transacciones a las 8:00PM. Luego decae hasta llegar a las cero transacciones pasada las 12 de la noche.

Al filtrar por servicio y modo de transporte es posible realizar un análisis más desagregado de las transacciones en la ciudad. En la Figura 3.5 se observa el total de transacciones en cada estación de la Línea 1 del metro en un día laboral promedio. Este gráfico permite identificar tres sectores de alta demanda por transporte público. De oriente a poniente, el primer sector se ubica en la comuna de Las Condes en las estaciones de Manquehue y Escuela Militar, donde se efectúan en conjunto cerca de 80 mil transacciones. Luego, en la comuna de Providencia, la demanda se distribuye de forma homogénea entre las estaciones de Tobalaba, Los Leones y Pedro de Valdivia. El último sector se encuentra en la comuna de Santiago entre las estaciones de Santa Lucía, Universidad de Chile y Moneda, siendo Universidad de Chile la principal estación de las tres con 15.000 transacciones más que las otras dos al final del día.

Estos tres focos de alta de demanda se ubican en sectores conocidos por la abundancia de altos edificios de oficinas que concentran una gran cantidad de puestos de trabajo, por lo tanto, es de esperarse que la mayoría de las transacciones vistas en el gráfico ocurran en el periodo de la tarde cuando las personas regresan a sus hogares. En la Figura 3.6 se puede observar este comportamiento, donde se muestra la distribución de las transacciones en la estación de Universidad de Chile en días laborales. A diferencia de la Figura 3.4 donde se muestra el total del sistema, en esta estación la punta mañana no concentra tantas validaciones, la punta del medio día es distinguible por sobre el resto del día y la punta tarde es el periodo de mayor demanda con sobre 2.000 transacciones cada 15 minutos. Además, la temprana salida del trabajo los días viernes es claramente visible en esta figura.

De forma opuesta, las estaciones de metro que se ubican lejos del centro de la ciudad

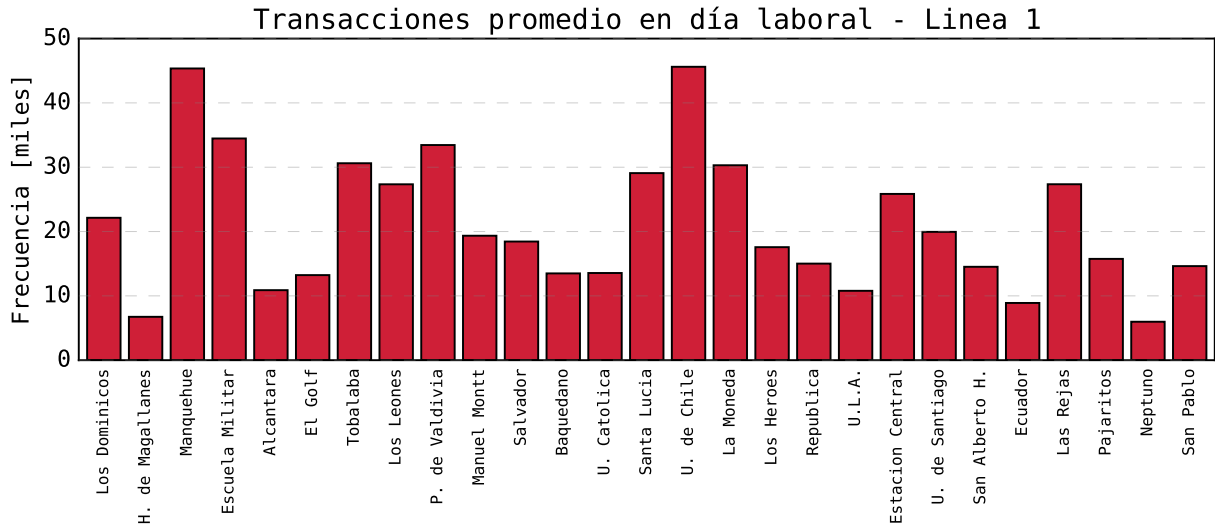


Figura 3.5: Distribución espacial de las transacciones en la Línea 1 del metro.

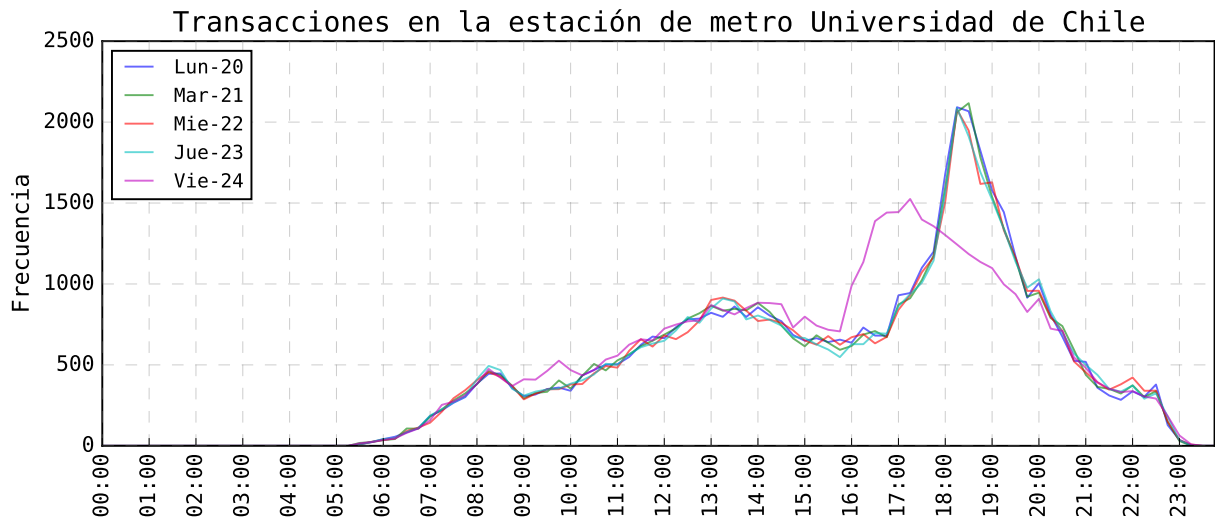


Figura 3.6: Distribución temporal de las transacciones en la estación Universidad de Chile.

y en áreas residenciales, concentran la mayoría de sus transacciones en la punta mañana cuando las personas se dirigen a sus trabajos. La Figura 3.7 presenta la distribución de transacciones en la estación Plaza de Puente Alto. El gráfico confirma el supuesto, se puede apreciar que la mayoría de la demanda ocurre en el periodo de la mañana, alcanzándose las 1.600 transacciones cada 15 minutos en cuatro de los cinco días laborales analizados. El resto del día se realizan entre 200 y 300 validaciones cada 15 minutos con un pequeño aumento a las 18:30 horas. El día viernes no destaca del resto de los días laborales. Los datos disponibles permiten realizar este mismo análisis en las cinco líneas de metro y en todas sus estaciones. De forma similar, se pueden analizar los servicios de buses y la distribución de las transacciones en cada una de sus paradas.

La Figura 3.8 presenta otro de los atributos que se obtienen por el pre procesamiento de los

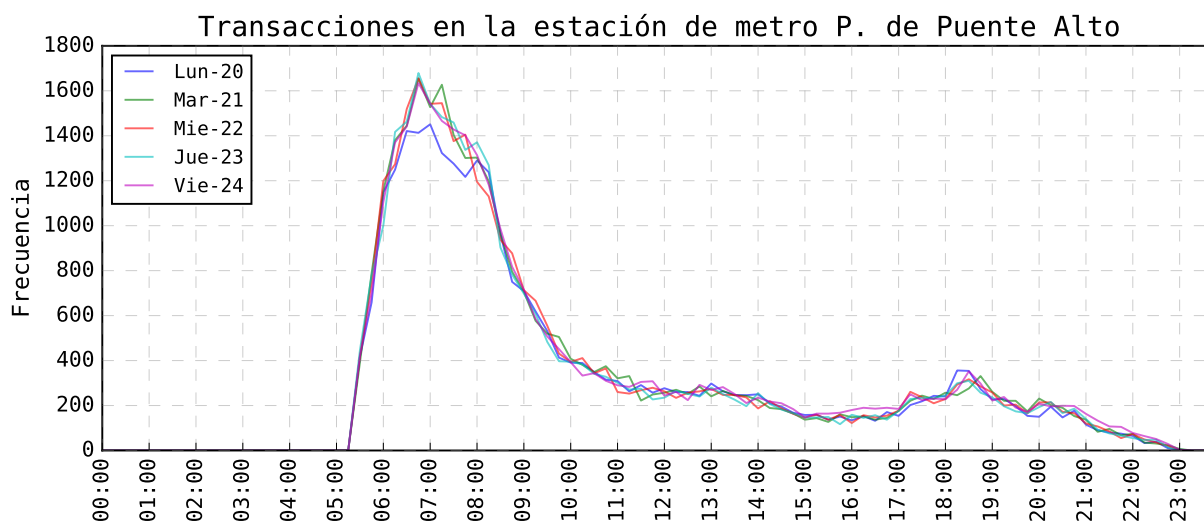


Figura 3.7: Distribución temporal de las transacciones en la estación Plaza de Puente Alto.

datos. En el gráfico se presenta la distribución de la distancia de las etapas calculada de forma euclidiana y sobre la ruta. Como es de esperarse, la distancia euclidiana es menor o igual que la calculada sobre la ruta. La forma de los histogramas da a entender que esta restricción se confirma, ya que la distribución de la distancia euclidiana es levemente más angosta y posee un máximo más alto. Observando cualquiera de las dos curvas se infiere que la mayoría de las etapas de viaje son de entre 2 a 3 *km* de distancia y que casi no existen etapas de más de 25 *km*. El cálculo de la distancia de los viajes es más susceptible a errores que el de las etapas, ya que para la distancia sobre la ruta es necesario que todas las etapas al interior del viaje tengan estimación de bajada y para la distancia euclidiana solo se necesita que la última etapa tenga estimación de bajada. La metodología para calcular la distancia sobre la ruta de las etapas posee dos errores que no se logran apreciar en la figura; la existencia de distancias negativas y distancias sobre los 30 *km* (ningún servicio permite viajar tan lejos en una sola etapa). Estos errores solo constituyen un 0.08 % de la base de datos y por lo tanto, no son preocupantes.

El análisis de los atributos de calidad de servicio como el tiempo de viaje, espera y caminata se hará una vez que los datos hayan sido agrupados por tarjeta. De esta forma se evitará mostrar gráficos con observaciones atípicas u posible errores que serán eliminados a la hora de preparar los datos para su posterior modelación.

3.5.2. Segundo Periodo: Noviembre, 2015

El segundo periodo sólo posee el registro de la hora y fecha de las transacciones, por lo tanto no es posible observar la estructura de viajes de las personas (las etapas dentro de un viaje no están identificadas), ni desagregar los datos por servicio o ubicación. Tampoco es posible observar variables de calidad de servicio relacionadas con la bajada de las etapas. No obstante, este periodo es más extenso, lo que permite analizar patrones en el mediano plazo. Al igual que en el primer periodo, si se agrupan las transacciones cada 15 minutos se puede

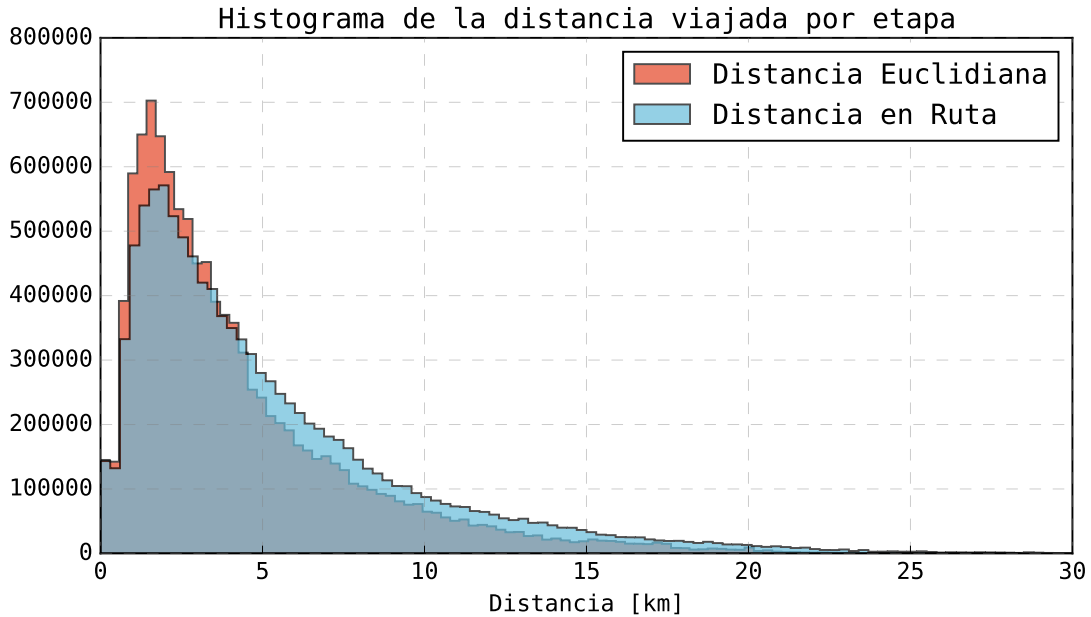


Figura 3.8: Distribución de la distancia viajada por etapa.

observar los periodos de alta y baja demanda durante el mes de noviembre. La Figura 3.9 presenta un mapa de calor donde en el eje de las ordenadas se encuentran los días del mes y en el eje de las abscisas las horas del día. El gráfico confirma los resultados previos. La punta mañana ocurre los días laborales entre las 7:00AM y las 8:30AM casi sin variación. La punta tarde también ocurre solo en días laborales y es entre las 6:00PM y las 7:30PM. Los días viernes la punta tarde tiene una mayor dispersión dado que comienza 1 hora y media antes y, por lo tanto, su máximo no es tan elevado como el resto de los días laborales. El número de transacciones al medio día de los días laborales se mantiene estable y presenta un leve aumento, generando una punta casi imperceptible. El número de transacciones los días sábado y domingo es menor y entre ellos no varían mucho. Además, el mapa de calor permite identificar dos periodos anormales durante el mes. El día jueves 12 entre las 8:00PM y las 10:00PM existe un uso característicamente bajo del sistema y lo mismo ocurre el día martes 17 de forma más prolongada hasta las 10:45PM. Estas anomalías son debido a la fuerte influencia que posee la Selección Nacional de Fútbol sobre una parte importante de la población. En estas dos fechas, Chile jugó con Colombia y Uruguay por las clasificatorias al Mundial 2018 respectivamente.

En la Figura 3.10 se utilizan diagramas de caja para observar la variabilidad en el número de transacciones por hora en días laborales (21 observaciones). La figura permite observar la baja variabilidad, entre días de la semana, presente en el sistema de transporte público de la ciudad de Santiago. Durante la mañana las cajas son bien angostas y no existen observaciones atípicas. Las primeras observaciones atípicas ocurren entre las 9:00AM y el medio día donde cada caja presenta una observación por sobre los bigotes. Entre las 12:00PM y las 2:00PM la variabilidad vuelve a disminuir para luego llegar al periodo con mayor variabilidad. La punta tarde posee entre 2 y 3 observaciones atípicas sobre el diagrama entre las 3:00PM y las 6:00PM y luego por debajo del diagrama entre las 6:00PM y las 10:00PM. Este aumento en

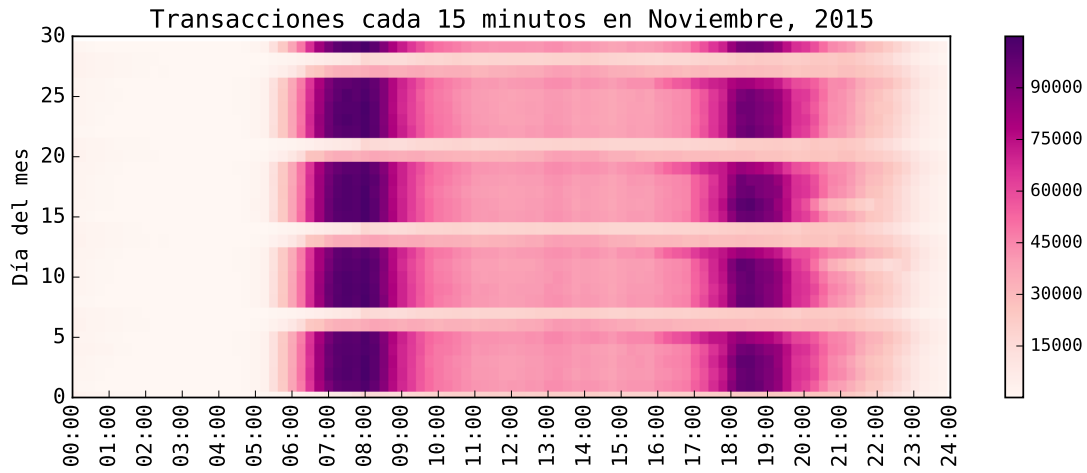


Figura 3.9: Mapa de calor de las transacciones en el mes de Noviembre, 2015.

la variabilidad en la punta tarde ocurre por la temprana salida del trabajo de las personas los días viernes, como se ha visto en otras figuras anteriormente. Pese a la variabilidad en la punta tarde, se puede apreciar que el sistema es bastante estable en cuanto al número de transacciones.

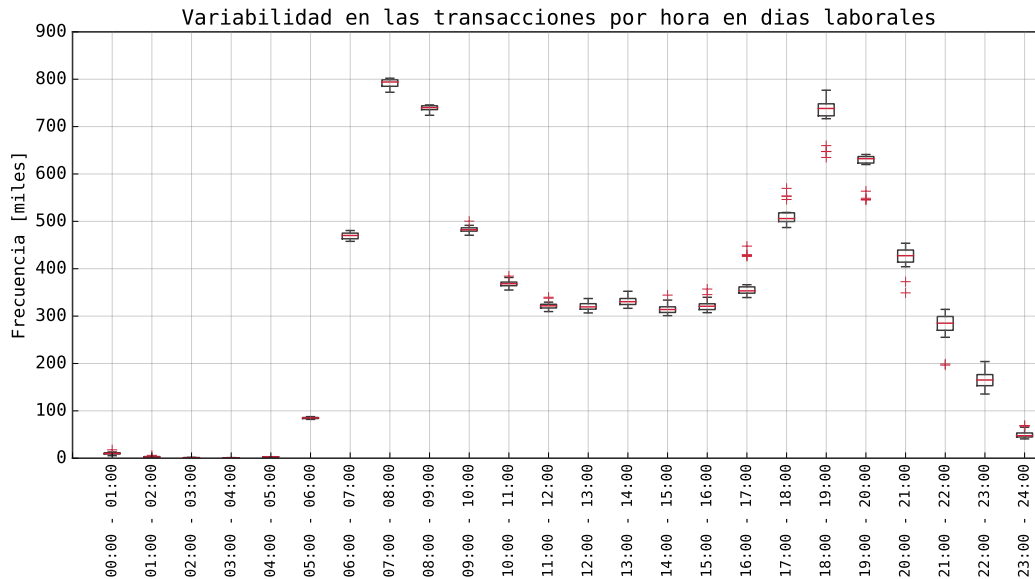


Figura 3.10: Variabilidad en el número de transacciones de días laborales por hora.

3.6. Indicadores de calidad de servicio

El Directorio de Transporte Público Metropolitano (DTPM) es la entidad que regula, controla y supervisa el sistema de transporte público de Santiago. Para llevar a cabo su labor, este calcula cinco indicadores que miden el desempeño de los operadores:

- Indicador de Calidad de Atención al Usuario (ICA).
- Indicador de Calidad de Vehículos (ICV).
- Indicador de Capacidad de Transporte (ICT).
- Indicador de Cumplimiento de Frecuencia (ICF) y Regularidad (ICR).

El Indicador de Calidad de Atención al Usuario se mide en una escala del 0 al 1 con la técnica del “cliente incognito”. Un inspector sube a los buses de forma incógnita y toma nota de ciertos atributos cualitativos mientras el bus viaja. El Indicador de Calidad de Vehículos mide el estado de los vehículos a través de atributos determinados por inspectores en los terminales preferentemente. Ambos indicadores se calculan sin detalle, solamente entregan un gran promedio de la empresa cada mes. El Indicador de Capacidad de Transporte pretende medir si la empresa ofrece la cantidad de “plazas” acordadas en el Plan Operacional. O sea, si la empresa dispuso de la cantidad de buses (frecuencia) y del tamaño comprometido. Estos indicadores entregan resultados muy agregados para el nivel de detalle con el que se desea trabajar en el resto de la investigación y por esto, no son considerados en el análisis. Los indicadores de Cumplimiento de Frecuencia y Regularidad en cambio, se miden con una desagregación de Servicio - Sentido - Periodo Transantiago - Día, lo que permite que sean imputados en la base de datos con el nivel de detalle deseado.

El Indicador de Cumplimiento de Frecuencia mide la cantidad de buses que han sido dispuestos en servicio en una determinada ventana de tiempo. Este indicador se calcula como el cociente entre la cantidad de buses que el operador puso en servicio y la cantidad acordada en el plan operacional en una escala de 0 al 1.

El Indicador de Cumplimiento de Regularidad mide la regularidad con que circulan los buses de un determinado servicio. Dado un plan operacional los intervalos de pasada de los buses son conocidos. El indicador se calcula como el cociente entre el número de intervalos cuya duración es igual o menor a lo acordado y el número total de intervalos en ese periodo de tiempo, nuevamente en una escala de 0 al 1.

La Figura 3.11 presenta la distribución de ambos indicadores mediante un gráfico de dispersión y dos histogramas. Al observar el gráfico de dispersión los resultados dan la sensación de que existe una gran cantidad de casos donde no se cumple el plan operacional para la frecuencia y la regularidad. No obstante, los puntos en el gráfico de dispersión están superpuestos y no permiten observar bien su concentración. En cambio, al observar los histogramas se aprecia que ambos indicadores toman el valor 1 en la mayoría de los casos. El Indicador de Cumplimiento de Regularidad posee una mayor dispersión, pero aún así su varianza es pequeña. Las líneas en azul señalan los puntos de corte bajo los cuales el Directorio de Transporte Público Metropolitano cobra multas a los operadores por su bajo rendimiento.

3.7. Variabilidad en el tiempo de viaje

La variabilidad en el tiempo de viaje de un usuario es un atributo muy relevante a la hora de medir la calidad de viaje entregada por un servicio. Cuando la variabilidad es alta, los usuarios no pueden planificar su estructura de viajes con certeza y suelen desperdiciar valioso

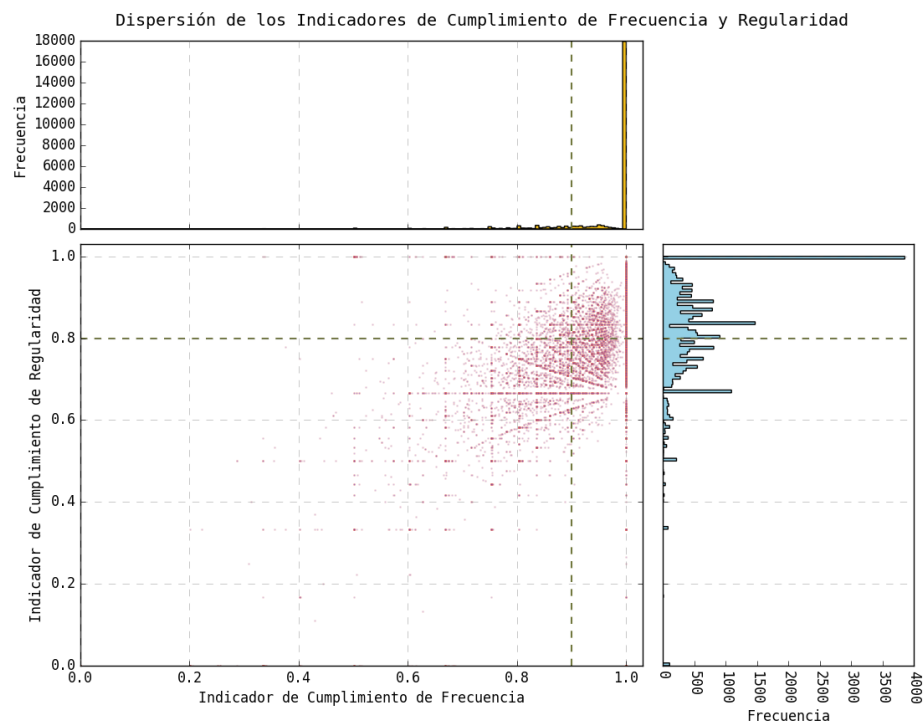


Figura 3.11: Distribución de los Indicadores de Cumplimiento de Frecuencia y Regularidad.

tiempo en esperas innecesarias o llegar tarde a sus destinos. En Santiago Durán (2016) estudió el efecto de la media de los tiempos de viaje, espera y caminata en los diferentes modos de transporte público sobre la desviación estándar del tiempo de viaje. Sus resultados muestran que existe una relación positiva entre estos tiempos y la desviación estándar del tiempo de viaje de las personas. Como en este caso el alcance de la investigación es mayor y no se dispone de información detallada de los viajes entre cada par origen–destino de la red, se calculará la variabilidad en el tiempo de viaje de los buses y no de las personas. Cada expedición de bus en el sistema registra su horario de inicio y fin en los cabezales de su ruta. Lo que permite conocer el tiempo que le toma al bus realizar una expedición completa. Luego, las expediciones son agrupadas temporalmente y al interior de cada conjunto se calcula la variabilidad en el tiempo de viaje como su desviación estándar. Las distintas agregaciones que se utilicen en un caso como este pueden definir los resultados que se obtengan. Una agregación más fina permite observar con mayor detalle los cambios en el comportamiento de los buses, sin embargo, posee menos observaciones por agrupación para calcular la variabilidad y, por ende, es más susceptible a outliers. En esta investigación se utilizaron las siguientes agregaciones:

1. Tipo de día - Servicio - Sentido - Media hora.
2. Servicio - Sentido - Periodos Transantiago.

Es importante mencionar que los “Periodos Transantiago” distinguen entre tipos de día, por lo tanto, la diferencia de las agregaciones es que una agrupa las expediciones cada media hora y la otra lo hace por periodos que varían entre una hora y 5 horas y media. Los periodos Transantiago separan los días laborales en 12 ventanas, los días sábado en 9 y los domingos en 8. En cambio, la agregación por medias horas divide un día en 48 periodos de igual duración.

Así, la agregación por medias horas es capaz de observar diferencias en el comportamiento de los buses cada 30 minutos y la agregación por periodos de Transantiago lo hace en promedio cada dos horas y media. No obstante, los periodos definidos por DTPM están basados en el comportamiento de la demanda y reflejan los cambios en el plan operacional, por lo tanto al interior de cada periodo la oferta debería comportarse de forma homogénea. El detalle de los periodos definidos por DTPM se puede encontrar en los Anexos B.1, B.2 y B.3. Los grupos resultantes de las agregaciones deben tener por lo menos cinco observaciones para ser considerados válidos, sino son eliminados del análisis. Finalmente, se imputa la variabilidad de cada grupo a las expediciones observadas en el periodo del 18 al 25 de abril. En resumen, a las expediciones de buses se les imputa un puntaje que mide la variabilidad en el tiempo que le toma a ese servicio en ese periodo de ese tipo de día en realizar su expedición completa. Este puntaje está en minutos y permite ordenar los servicios por su regularidad. En la Figura 3.12 se puede observar la distribución de la variabilidad encontrada para las dos agregaciones. En ambos casos, el valor más frecuente esta entre los tres y los cuatro minutos. Es importante mencionar que la variabilidad en el tiempo de viaje de los buses está fuertemente influenciada por la infraestructura por la que circulan (pistas segregadas, pistas solo bus y tráfico mixto) y los horarios del día. La metodología propuesta no captura diferencias por tramos, sólo calcula un valor general de la expedición completa.

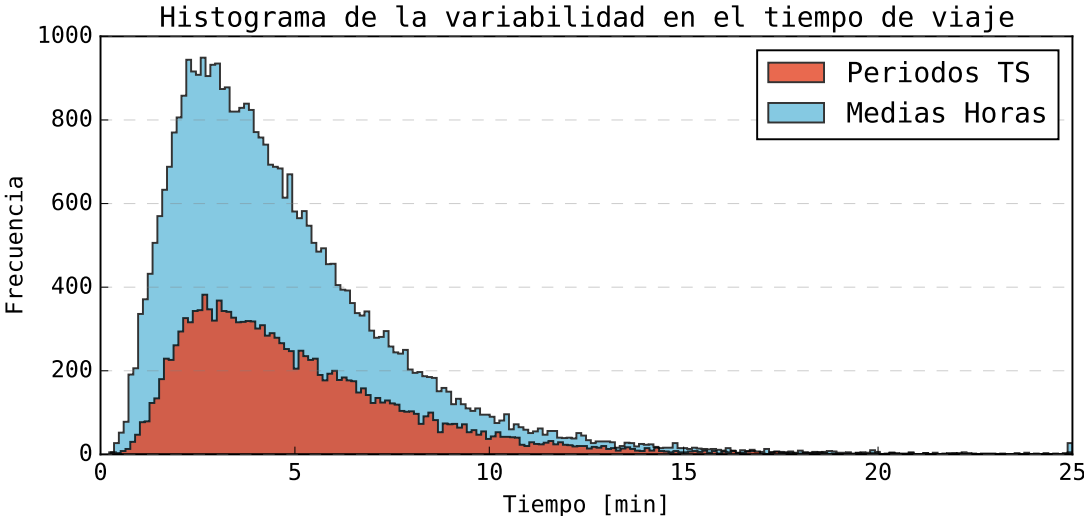


Figura 3.12: Distribución de la variabilidad en el tiempo de expedición de los buses.

3.8. Encuesta CASEN

La Encuesta de Caracterización Socio-Económica Nacional (CASEN) 2015 tiene por objetivo conocer la situación socio económica de los hogares y de la población que habita en viviendas particulares en el país. La encuesta recoge información sobre aspectos tales como la composición de los hogares y familias, educación, salud, vivienda, trabajo e ingreso de las personas de cada hogar encuestado. El levantamiento de la encuesta se inició el 2 de noviembre del 2015 y finalizó el 31 de enero del 2016. Los datos obtenidos por los 83.887 hogares encuestados son representativos a nivel país, zonal (rural y urbano), regional y comunal en las

15 regiones de todo el país. Del conjunto de atributos medidos por la encuesta, para el trabajo de esta tesis se utilizarán solamente el ingreso total de hogar y número de vehículos particulares y laborales por hogar. El ingreso será utilizado principalmente para segmentar la muestra de tarjetas por quintiles (observando solamente los hogares de la Región Metropolitana). En la tabla 3.3 se pueden observar los puntos de corte de los cinco quintiles.

Tabla 3.3: Quintiles de ingreso en la Región Metropolitana.

Quintil	Mínimo	Máximo
I	-	\$ 484.560
II	\$ 484.560	\$ 730.938
III	\$ 730.938	\$ 1.074.284
IV	\$ 1.074.284	\$ 1.790.000
V	\$ 1.790.000	\$ 49.750.000

Posteriormente, el ingreso y el número de vehículos por hogar serán utilizados para calcular el ingreso medio y la tasa de motorización media por comuna en las 34 comunas donde opera el sistema Transantiago. La Figura 3.13 muestra los ingresos medios totales por hogar y su varianza obtenidos para cada comuna ordenados de forma descendente. La figura permite observar que el área metropolitana de Santiago posee una distribución desigual del ingreso. La comuna más rica (Vitacura) posee casi seis veces el ingreso medio de la comuna más pobre (Lo Espejo). Las comunas han sido agrupadas por los quintiles calculados previamente. En el quinto quintil se agrupa a 6 comunas, en el cuarto 8, en el tercero 19 y finalmente en el segundo solo la comuna de Lo Espejo. Ninguna comuna posee un ingreso medio bajo los \$484.560 para ser catalogada en el primer quintil. Otra observación importante del gráfico es que la desigualdad al interior de cada comuna aumenta con el ingreso de esta. La distribución espacial del ingreso también es un tema importante. Las comunas más adineradas de la ciudad se encuentran concentradas en el sector nor-oriente como se puede apreciar en la Figura 3.14. Esto hace que una gran cantidad de actividades se concentren en esta zona ya que es donde se encuentran los centros comerciales más grandes, los edificios de oficina más importantes, entre otros.

La Figura 3.15 muestra la tasa de motorización por comuna ordenada según su ingreso medio de forma descendente, al igual que en la Figura 3.13. El comportamiento de la curva es similar a la de ingreso, como es de esperarse. Las comunas de Providencia y Santiago destacan por sus bajas tasas de motorización dado su nivel de ingreso. Es posible que esto se debe a dos factores; su privilegiada ubicación en la zona central de la ciudad lo que les permite tener una alta accesibilidad al resto de la ciudad y en particular al transporte público y al alto costo de los estacionamientos en estas zonas. Si se observa la curva de vehículos laborales, se puede apreciar que esta no sufre grandes cambios, en cambio la curva de vehículos particulares es la que explica las variaciones en la tasa de motorización total.

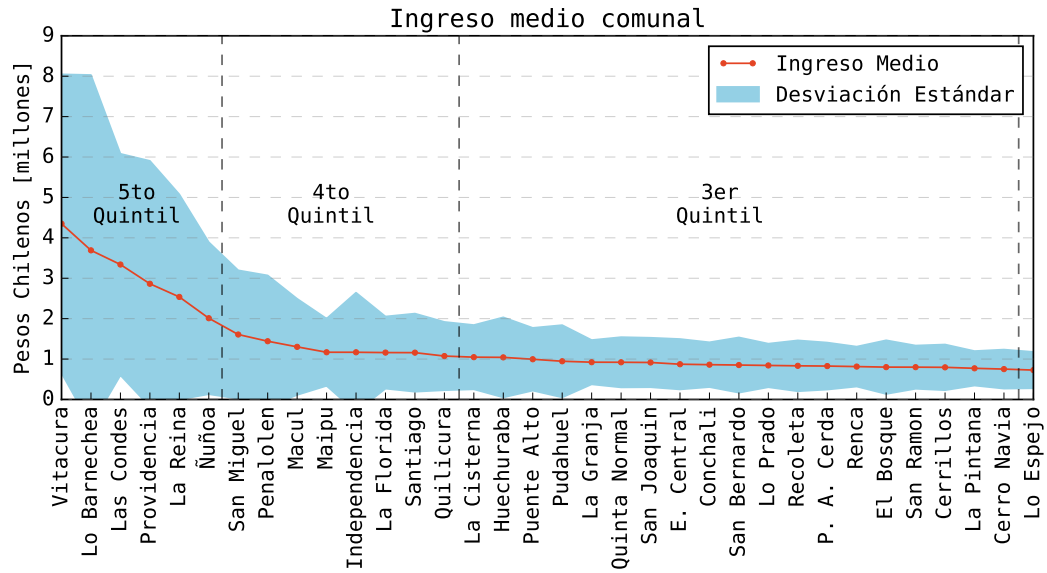


Figura 3.13: Distribución del ingreso en el área metropolitana de Santiago.

Distribución espacial del ingreso en la ciudad de Santiago

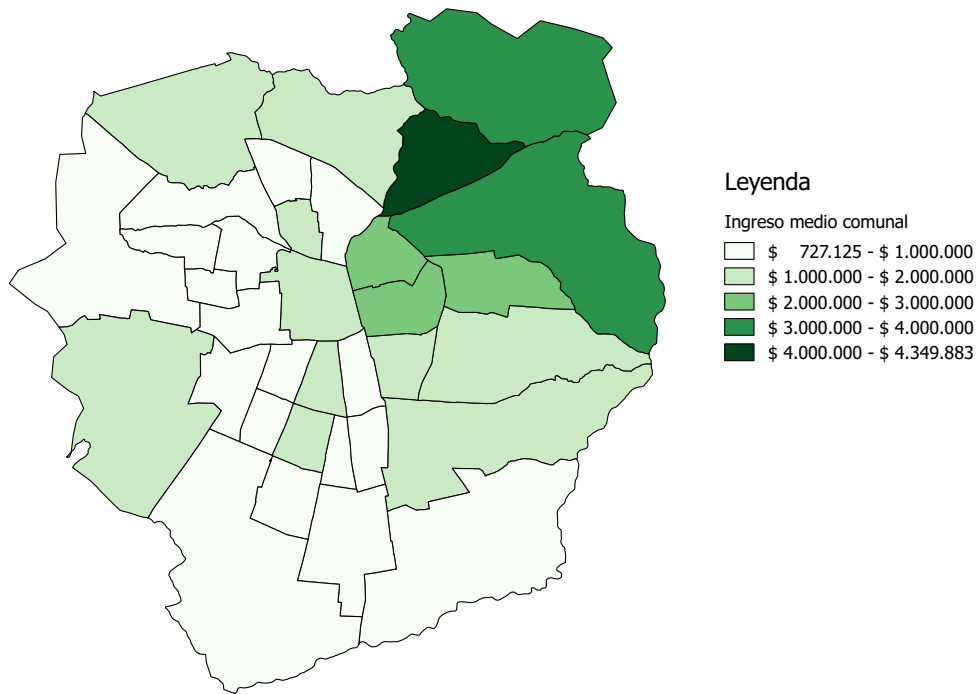


Figura 3.14: Distribución espacial del ingreso en el área metropolitana de Santiago.

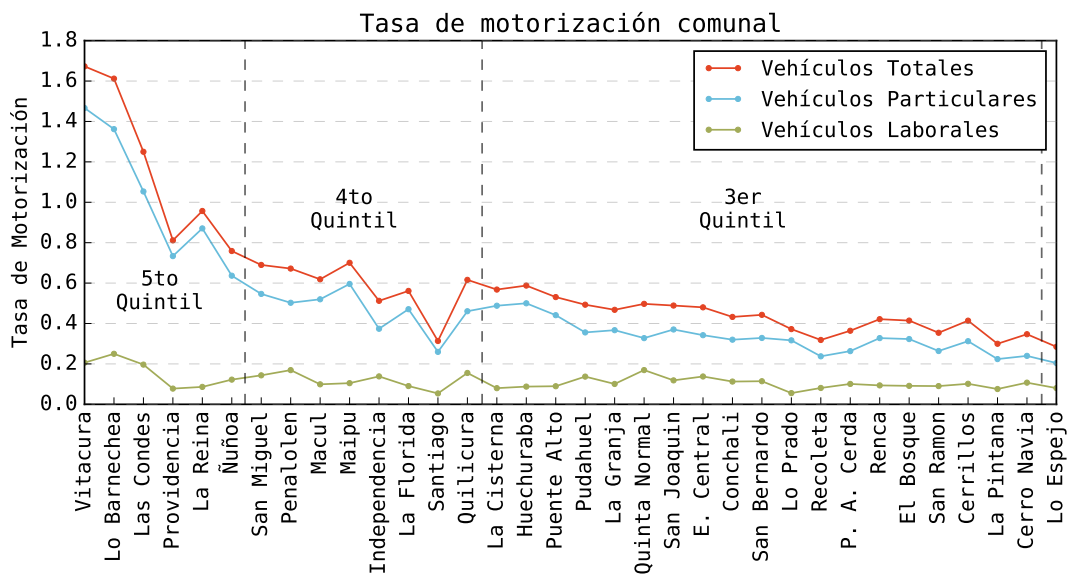


Figura 3.15: Distribución de la tasa de motorización en el área metropolitana de Santiago.

Capítulo 4

Modelación

En esta investigación no sólo se modela la decisión de los usuarios de migrar, sino también el grado de uso del sistema de transporte público. Para hacer esto, los usuarios son clasificados en cuatro categorías: Frecuente, Semi Frecuente, No Frecuente y Ausente. Las primeras tres categorías son el resultado de una clusterización de k-medias basada en la cantidad de transacciones y de días viajados de una tarjeta. La última categoría corresponde a la migración. La que sucede cuando una tarjeta presente en el primer periodo, no lo esta en el segundo. Así, se entiende por grado de uso del sistema como la frecuencia de transacciones y días viajados con las que una tarjeta aparece en las bases de datos, considerando incluso la ausencia de esta. Esta definición permite al modelo identificar los factores que afectan la migración de usuarios y también la tendencia de estos a disminuir su uso del sistema.

El modelo propuesto es un modelo de rezago donde las variables explicativas son observadas en un primer periodo y la variable a explicar en un periodo posterior. El primer periodo corresponde a ocho días de datos procesados por las metodologías previamente mencionadas. El tiempo de procesamiento necesario para aplicar las metodologías y para la posterior construcción de las variables de modelación impide que este periodo sea más extenso. En cambio, el algoritmo para la clasificación de usuarios utilizado en el segundo periodo no es tan complejo ni tampoco requiere que los datos sean previamente procesados, haciendo posible observar una ventana temporal mayor. De esta forma, el primer periodo comienza el día 18 y termina el 25 de abril del 2015 y el segundo es el mes completo de noviembre del 2015. Los periodos fueron cuidadosamente seleccionados para que no existan eventos especiales que puedan influenciar el comportamiento de los usuarios como manifestaciones, feriados, fiestas, etc. Además, analizar la migración de usuarios en un periodo más extenso permite disminuir la probabilidad de mal clasificar una tarjeta en la categoría “Ausente” cuando esta no estuvo presente por razones de fuerza mayor como enfermedades, viajes, trabajo u otros.

El modelo asume que la utilidad que le reportan los modos alternativos de transporte al usuario permanecen constantes entre un periodo y el siguiente. Esto es fundamental para poder explicar la migración de usuarios por atributos propios del sistema y no por razones externas no observables como la construcción de nueva infraestructura para algún modo de transporte en particular (ciclovías o autopistas).

Otra característica importante del modelo es que este describe el comportamiento de tarjetas y no el comportamiento de personas. Esto es importante dado que un usuario de transporte público puede utilizar más de una tarjeta o una tarjeta puede ser usada por más de una persona. Además, el modelo solo es capaz de observar los viajes asociados al uso de la tarjeta inteligente. El resto de los viajes que un usuario realice en otros modos no existen para la modelación. La frecuencia con que estos comportamientos suceden se desconoce y se espera que al trabajar con una muestra grande de tarjetas estos errores causen un ruido menor en el modelo.

4.1. Clasificación de usuarios

Para clasificar a los usuarios se utiliza el algoritmo de clusterización k-medias. Este es un método de agrupamiento que tiene como objetivo particionar un conjunto de N observaciones en k grupos a fin de minimizar la suma de los cuadrados de las diferencias al interior de cada grupo. El algoritmo asigna cada observación al grupo cuyo valor medio este más cercano midiendo la distancia de forma euclidiana. La clasificación de las tarjetas se hace según el número de transacciones y la cantidad de días viajados en el periodo analizado.

Los resultados de la clusterización en el mes de noviembre del 2015 se presentan en la Figura 4.1. El tamaño de los círculos indica la cantidad de tarjetas que poseen la misma cantidad de transacciones y días viajados y los colores la pertenencia a las tres categorías. Las tarjetas no frecuentes están concentradas en la esquina inferior izquierda del gráfico. Por el tamaño de los círculos se entiende que existe un gran conjunto de tarjetas que realiza pocas transacciones y viaja pocos días en el mes. La categoría Semi Frecuente es la más dispersa en cuanto al número de días viajados, teniendo tarjetas que viajan desde los 5 hasta los 28 días. En la categoría Frecuente hay tarjetas que viajan entre 12 días y todo el mes. Sobre el número de transacciones la categoría Frecuente es la de mayor varianza con tarjetas que validan desde 30 hasta incluso más de 200 veces. Los centroides definen el valor medio de cada categoría, estos están representados por tres cruces en el gráfico y permiten entender el comportamiento de un usuario promedio de cada grupo. Así, durante un mes una tarjeta promedio de la categoría No Frecuente realiza 6,9 transacciones y viaja en 3,1 días, de la Semi Frecuente 34 transacciones y 13,9 días y de la Frecuente 81,6 transacciones y 22,9 días. Al observar la cantidad de tarjetas que pertenecen a cada conjunto y el total de transacciones que realizan, se concluye que pese a que las tarjetas no frecuentes son un 63,7 % del total, estas solo realizan un 18,8 % de todas las transacciones. En cambio, las tarjetas frecuentes y semi frecuentes son un 13,9 % y 22,4 % del total y realizan un 48,6 % y 32,5 % de todas las transacciones respectivamente. Estos valores justifican que la investigación se realice solamente sobre las categorías Frecuente y Semi Frecuente, las que en conjunto son responsables del 80 % de todas las transacciones del sistema.

Si se realiza esta misma clasificación de usuarios en los dos periodos, normalizando por el número de días, se puede observar la evolución del uso del sistema de las tarjetas de un periodo al siguiente. En la Figura 4.2 se presenta la evolución de las tarjetas clasificadas como frecuentes y semi frecuentes en el primer periodo. Al analizar la categoría Frecuente de abril se observa que un 36 % de las tarjetas vuelven a ser clasificadas como frecuentes en noviembre,

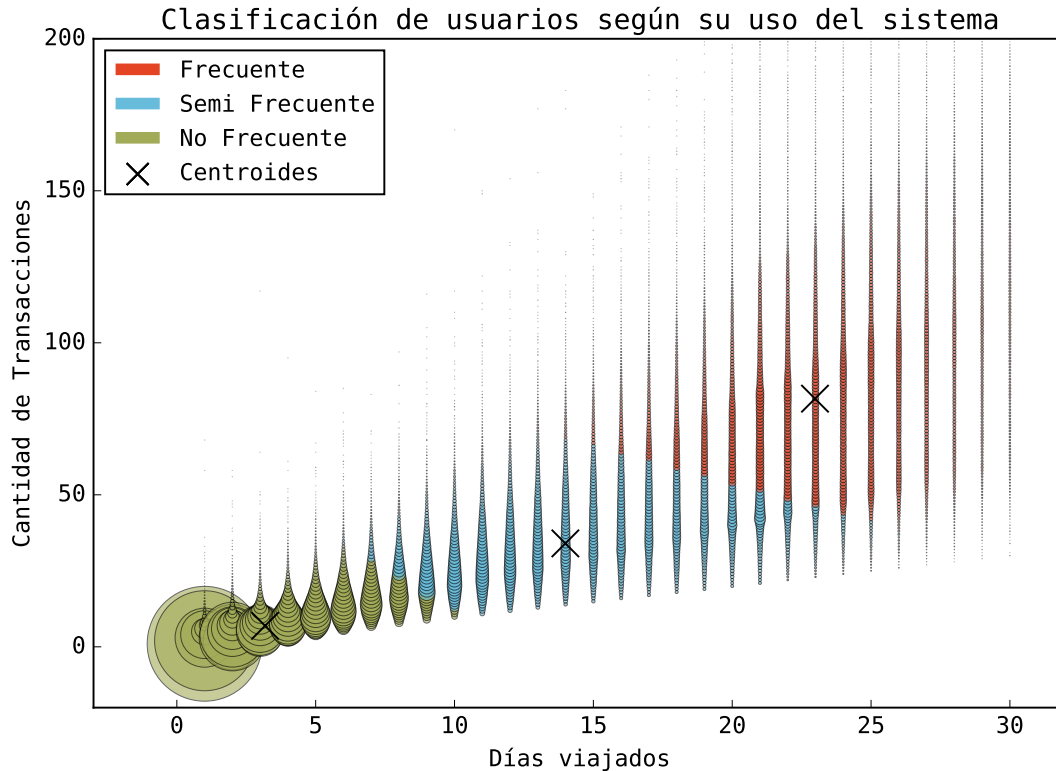


Figura 4.1: Clasificación de usuarios según su grado de uso del sistema.

un 15 % son clasificadas como semi frecuentes, un 14 % como no frecuentes y un 36 % no están presentes en el segundo periodo y por lo tanto son clasificadas como ausentes. En total un 29 % del conjunto disminuyó su uso del sistema sin abandonarlo y un 64 % disminuyó su uso del sistema o lo abandonó. Si se analiza el grupo de las tarjetas semi frecuentes de abril, se observa que un 23 % de ellas aumentó su uso del sistema y son catalogadas como frecuentes en noviembre, un 21 % vuelven a ser catalogadas como semi frecuentes, un 18 % disminuyó su uso del sistema sin abandonarlo y finalmente un 37 % migró. Por último, si se observa la tendencia global de la muestra vemos que la tasa de abandono es en promedio de 36,5 % y la porción de usuarios que disminuye su uso o abandona el sistema es cercana al 60 %.

4.2. Modelo Logit Ordinal

El orden natural de las categorías creadas para la variable dependiente (grado de uso del sistema) hace muy conveniente el uso del Modelo Logit Ordinal (Train, 2002) para modelar el comportamiento de las tarjetas. Este modelo asume que cada individuo percibe una utilidad al utilizar el transporte público que depende de variables observables como el tiempo de viaje o el tipo de transporte y otras no observables como el nivel de hacinamiento o la sensación térmica al interior del vehículo. Luego, el nivel de utilidad percibido por cada usuario en el primer periodo de análisis determinará con que frecuencia este utilizará el sistema en el futuro. De esta forma, usuarios que perciban un nivel de utilidad alto serán más propensos a

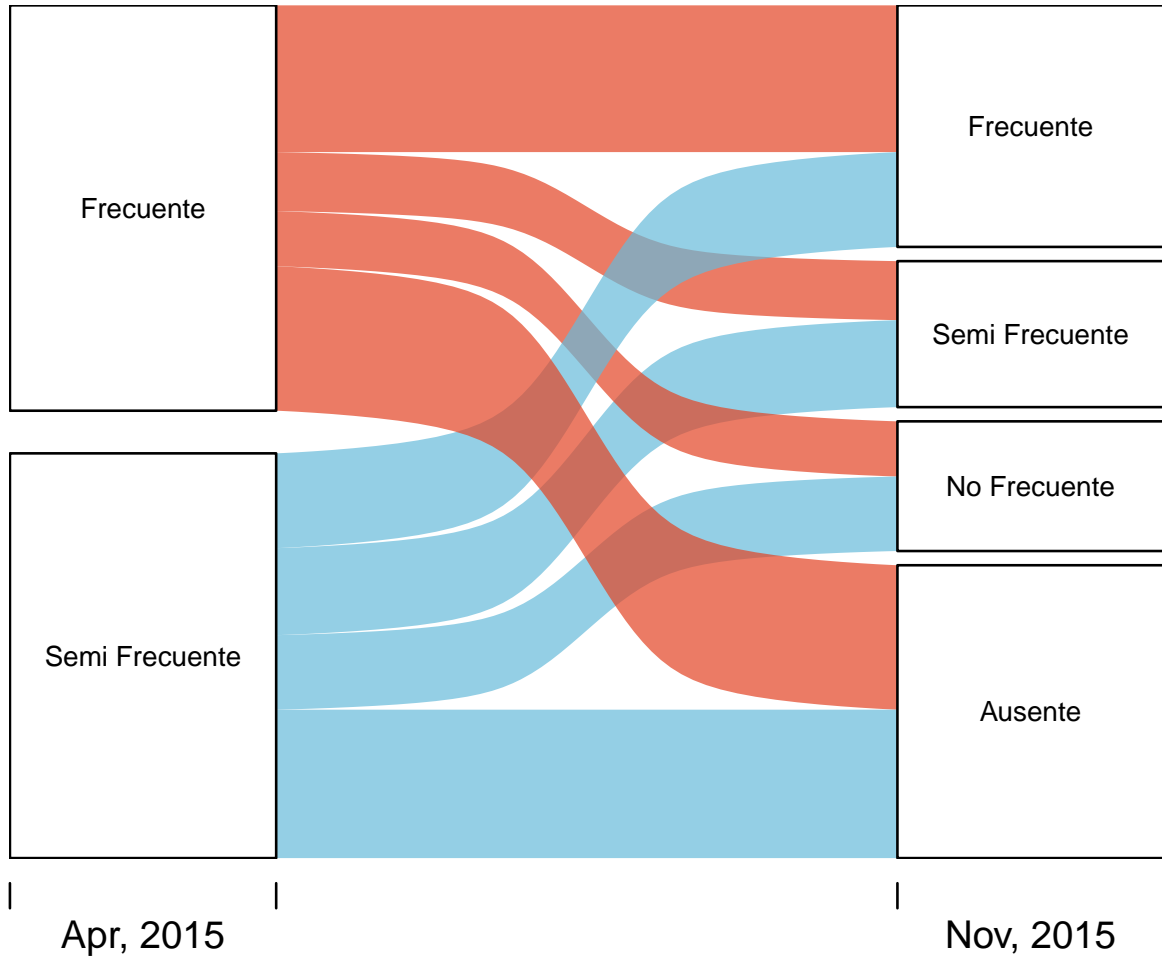


Figura 4.2: Evolución de la clasificación de usuarios según su grado de uso del sistema.

viajar con mayor frecuencia en transporte público (y por consiguiente ser catalogados como Frecuentes o Semi Frecuentes), en cambio aquellos que perciban una baja utilidad disminuirán su frecuencia de uso, llegando al extremo de migrar del sistema.

Este modelo se expresa como en la Ecuación 4.1, donde j va desde 1 hasta el número de categorías menos 1 e i va desde 1 hasta el número de variables. El vector β contiene los coeficientes que definen la utilidad marginal de cada variable y es independiente de las categorías. Los valores de θ_j definen los puntos de corte entre las categorías. El signo negativo en la ecuación permite que la comprensión de los resultados sea más intuitiva. Un parámetro de valor positivo, implica una utilidad marginal positiva, por lo cual un aumento en la variable asociada produce una disminución de los puntos de corte θ_j disminuyendo la probabilidad de caer en las categorías iniciales (Ausente o No Frecuente) y aumentando la probabilidad de caer en las últimas (Semi Frecuente o Frecuente).

$$\ln(\theta_j) = \alpha_j - \beta \cdot X_i \quad (4.1)$$

En resumen, las variables asociadas a coeficiente positivos aumentan la probabilidad de

ser catalogado como frecuente y las variables asociadas a coeficientes negativos explican la migración y la disminución del uso del sistema por parte de los usuarios.

4.3. Supuestos y limitaciones del modelo

La especificación del modelo propuesta implica realizar varios supuestos y estar sujeto a ciertas limitaciones que se discutirán en esta sección. Primero, al trabajar con un modelo de rezago se asume que la decisión de los usuarios sobre su grado de uso del sistema esta basada en la experiencia pasada utilizando este mismo sistema. En específico, esta basada en la experiencia durante una semana hace seis meses atrás. Lamentablemente, el tamaño del sistema y costo computacional que requiere obtener estos datos imposibilita observar más cortes temporales intermedios. Por otra parte, durante estos seis meses se asume que el resto de los modos de transporte permanecen iguales. De lo contrario, el comportamiento de los usuarios se podría ver influenciado por factores externos como por ejemplo la construcción de una nueva ciclovía. Durante este periodo también se asume que la estimación de residencia de las tarjetas, sus trabajos y lugares de estudio no varían. Cualquiera de estos cambios altera la utilidad percibida por los modos de transporte y por consiguiente la frecuencia con que un usuario viaja en el transporte público. Los dos cortes temporales han sido escogidos especialmente dentro de un mismo año calendario y a una distancia razonable para minimizar estos efectos.

Otros de los supuestos del modelo están relacionados con el uso de las tarjetas. Esta investigación modela el comportamiento de tarjetas y no de personas. Esto debido a que una tarjeta puede ser utilizada por más de una persona, así como también, una persona puede utilizar más de una tarjeta. Dado que las tarjetas de adultos regulares no son personalizadas se espera que existan estos comportamientos. Sin embargo, estos efectos son considerados como ruido para el modelo gracias a que se trabaja con grandes cantidades de observaciones. El extravío de tarjetas es también una fuente de posibles errores en la modelación. El modelo asume que las tarjetas que no realizan ninguna transacción durante el segundo periodo han migrado del sistema. No obstante, existen varias razones por las cuales una tarjeta se podría ausentar como viajes de trabajo, enfermedad, vacaciones, extravío o deterioro de las tarjetas, etc. Esta mal clasificación de las tarjetas se busca minimizar el observar una ventana temporal más extensa en el segundo periodo, pero aun así la posibilidad de que ocurran siempre esta presente. Finalmente, las tarjetas son clasificadas según un valor agregado del ingreso por comuna proveniente de la Encuesta CASEN 2015. La metodología asume que estos valores son representativos de todos los usuarios al interior de estas comunas pese a que en la Figura 3.13 se puede observar que existe una alta variabilidad al interior de cada comuna.

4.4. Agrupación por tarjeta y variables de modelación

Para modelar el comportamiento de las tarjetas es necesario agrupar el registro de los viajes y las etapas de tal forma que cada observación en la nueva base de datos describa la experiencia vivida por una tarjeta en el transporte público durante los ocho días del

primer periodo. Al agrupar la base de datos, se construyen nuevas variables a partir de los atributos creados anteriormente por el pre procesamiento de los datos brutos. Las nuevas variables buscan describir aspectos tales como la calidad de servicio entregada por el sistema, la preferencia de los usuarios sobre los modos de transporte y los periodos en los que viajan, las diferencias socio económicas entre las comunas de la ciudad, entre otros. En esta sección se presenta el conjunto de variables creadas para la etapa de modelación.

4.4.1. Variable Dependiente

La variable dependiente es el grado de uso del sistema en el mes de noviembre y esta se construye a partir de la clasificación de usuarios descrita anteriormente. La variable posee cuatro categorías con un orden natural: Frecuente, Semi Frecuente, No Frecuente y Ausente. La proporción de estas categorías se puede apreciar en la Tabla 4.1. La clasificación muestra que el comportamiento más común entre la muestra es la migración (Ausente) con un 36 % del total. No obstante, es seguido por las tarjetas frecuentes con un 30 %. Las tarjetas semi frecuentes son un 18 % y las no frecuentes un 16 %. En la tabla se aprovecha de mostrar el total de observaciones disponibles para la modelación. La base de datos de modelación contiene 436.920 tarjetas todas con una estimación de zona de residencia y sus atributos de calidad de servicio correctamente calculados.

Tabla 4.1: Distribución de la variable dependiente.

Clasificación en Noviembre, 2015		
Frecuente	130.294	(30 %)
Semi Frecuente	78.740	(18 %)
No Frecuente	69.939	(16 %)
Ausente	157.950	(36 %)
Total	436.920	(100 %)

4.4.2. Variables Independientes

Quizás las variables más relevantes para describir la experiencia de viaje de las personas son las relacionadas a la calidad de servicio. A continuación se presentan las seis variables asociadas a este aspecto que se han incluido en la modelación:

- Tiempo de Viaje Corto.
- Tiempo de Viaje Largo.
- Tiempo de Espera.
- Tiempo de Caminata.
- Velocidad de Viaje.
- Variabilidad del Tiempo de Viaje.

Las variables Tiempo de Viaje Corto y Tiempo de Viaje Largo son variables binarias que toman el valor 1 cuando el tiempo medio de viaje de una tarjeta es menor al percentil 25

o mayor al percentil 75 del tiempo medio de viaje de toda la muestra respectivamente. Los tiempos de espera y caminata, en minutos, son calculados directamente como el promedio de estos tiempos en los viajes de cada tarjeta. De la misma forma, la velocidad de viaje, medida en metros por segundo, es el promedio de los cocientes entre la distancia euclidiana y el tiempo total de los viajes de cada tarjeta. La variabilidad en el tiempo de viaje, en minutos, se calcula observando los servicios de buses en los que valida cada tarjeta y luego promediando el valor de la variabilidad del tiempo total de expedición asignado a ese servicio en ese periodo. Esta última puede ser calculada a partir de dos agregaciones temporales, como se vio en la Sección 3.7. La agregación que entregó mejores resultados y, por lo tanto la que se ocupa es: Tipo día - Servicio - Sentido - Media Hora.

En la Figura 4.3 se presenta el tiempo medio de viaje y la desviación estándar de los tiempos de viaje por tarjeta. De la figura se concluye que en la mayoría de los casos el tiempo medio de viaje de una tarjeta es de 24 minutos y que la desviación estándar más frecuente es de 10 minutos. Esta figura también presenta los percentiles 25 y 75 del tiempo medio de viaje mediante dos líneas verticales punteadas. Tarjetas con tiempos medios de viaje menores a la primera línea llevan un 1 en la variable Tiempo de Viaje Corto y aquellas con tiempos superiores a la segunda línea llevan un 1 en la variable Tiempo de Viaje Largo. En la figura 4.4 se muestran las distribuciones del tiempo medio de espera y caminata en transbordos. Al comparar estas curvas se deduce que el tiempo de espera posee una mayor varianza que el de caminata. También, que la moda del tiempo de espera es cercana a los 6 minutos, en cambio, el tiempo de caminata concentra la mayoría de sus observaciones entre los 1 y los 3 minutos.

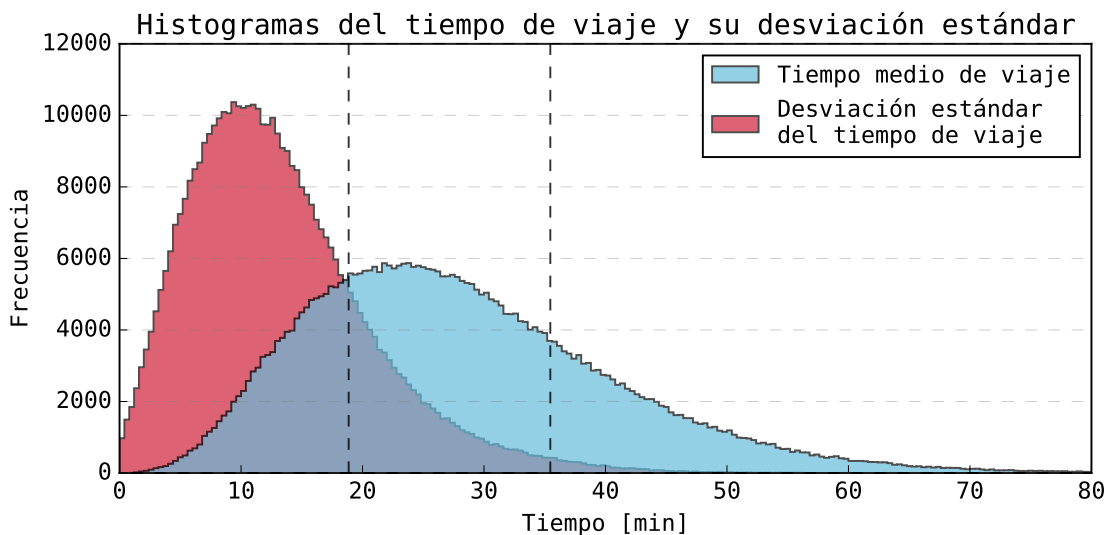


Figura 4.3: Distribución del promedio y la desviación estándar del tiempo de viaje.

La preferencia de un usuario por utilizar un modo de transporte por sobre el resto fue modelada de dos formas: Mediante el uso de variables binarias que indican cuando una tarjeta ha validado por lo menos una vez en alguno de los tres modos de transporte distinguibles en la base de datos (Metro, Bus, Zona Paga) y mediante una variable continua que mide el porcentaje del tiempo total que una tarjeta viaja en metro. Dado que el 99 % de la muestra ha validado por lo menos una vez en bus, la variable fue eliminada por su poca varianza. También, se quiso incluir variables binarias que identifiquen las tarjetas que solo validan

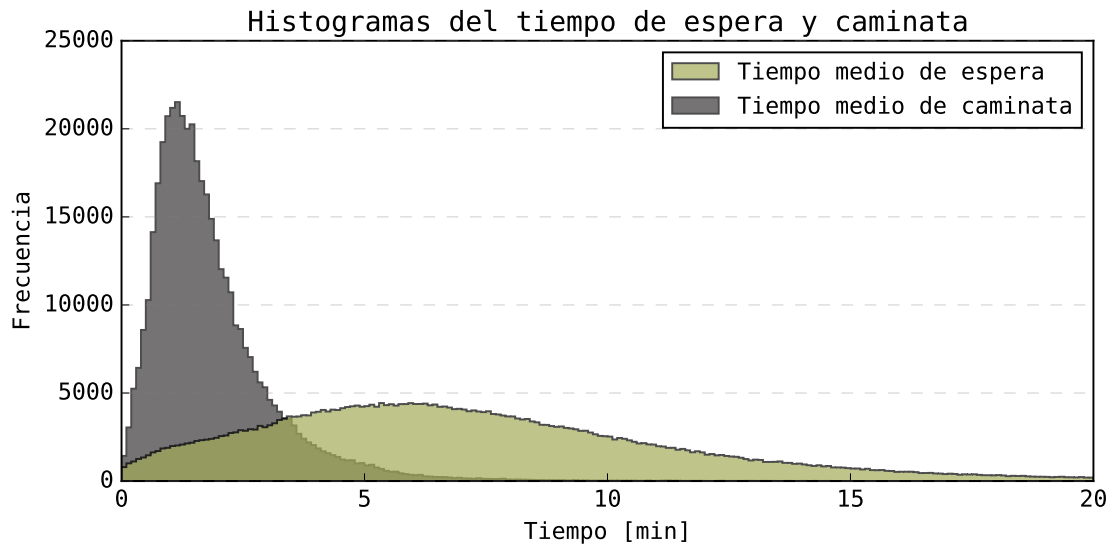


Figura 4.4: Distribución del tiempo medio de espera y caminata.

en un modo de transporte, pero en los casos de Metro y Zonas Pagas no hubieron tarjetas que hagan uso exclusivo de estos modos, por lo tanto, solo se incluyó para buses. Además, se incorporó el uso de los servicios de buses expresos con una variable binaria que indica cuando una tarjeta ha validado por lo menos una vez en estos servicios. Así, las variables que describen la preferencia del usuario por un modo de transporte son las siguientes:

- Usa Metro.
- Usa Zona Paga.
- Solo Bus.
- Porcentaje del Tiempo de Viaje en Metro.
- Usa Bus Expreso.

Los patrones temporales y espaciales de cada tarjeta son capturados por tres atributos. Una variable binaria que se activa cuando una tarjeta valida solo en días laborales (no viaja durante los fines de semana), una variable binaria que indica cuando una tarjeta valida solo en horarios punta (mañana o tarde) y una variable entera que mide la variabilidad espacial del usuario según la cantidad de servicios distintos en los que valida la tarjeta. Esta última variable se basa en la idea de que un usuario que usa más servicios distintos, satisface más necesidades viajando en transporte público, las que probablemente se encuentran ubicadas en distintos lugares y, por lo tanto, posee una variabilidad espacial mayor. Estas variables llevan los siguientes nombres:

- Solo Laboral.
- Solo Periodo Punta.
- Servicios Únicos.

Los Indicadores de Cumplimiento de Frecuencia y Regularidad fueron incluidos observando los servicios de buses en los que cada tarjeta valida y luego promediando los valores asignados previamente por el DTPM dependiendo del periodo y el día. A estas variables se les llama

“ICF” e “ICR” respectivamente. La clasificación del grado de uso del sistema realizada en el primer periodo que permite eliminar a las tarjetas no frecuentes del análisis y además identificar a las tarjetas semi frecuentes y las frecuentes, fue introducida como una variable binaria que toma el valor de 1 cuando una tarjeta fue catalogada como frecuente en el periodo de abril y 0 cuando fue catalogada como semi frecuente. Esta variable lleva el nombre de “Frecuente”.

Los datos socio económicos de la Encuesta CASEN 2015 fueron utilizados de dos formas: la tasa de motorización fue imputada directamente a cada tarjeta como el promedio comunal de su estimación de residencia y el ingreso medio comunal fue utilizado para identificar la pertenencia de las tarjetas a comunas de distintos estratos socio económicos en la ciudad. En particular el ingreso se utilizó para identificar a las tarjetas cuya estimación de residencia pertenece a las comunas de ingreso medio por hogar del quinto quintil. Esto dado que es del cuarto al quinto quintil donde se marca la mayor diferencia en la distribución del ingreso en la ciudad. El detalle de los puntos de corte de los quintiles y las comunas que pertenecen a cada uno de ellos se encuentra en la Sección 3.8.

El detalle de las variables independientes se presenta en las Tablas 4.2 y 4.3. La primera, contiene todas las variables binarias de la base de datos de modelación con sus respectivas proporciones y la segunda, contiene las variables enteras y continuas con sus principales estadísticos.

Tabla 4.2: Distribución de las variables binarias.

	1	0
Tiempo de Viaje Corto	109.905 (25 %)	329.715
Tiempo de Viaje Largo	109.905 (25 %)	329.715
Usa Metro	367.036 (84 %)	69.884
Usa Zona Paga	199.165 (46 %)	237.755
Solo Bus	47.657 (11 %)	389.263
Usa Bus Expreso	117.888 (27 %)	319.032
Solo Laboral	125.521 (29 %)	311.399
Solo Periodo Punta	16.687 (4 %)	420.233
Quintil V	51.217 (12 %)	385.703
Frecuente	218.623 (50 %)	218.297

4.5. Resultados

A continuación, se describe en detalle los resultados obtenidos del Modelo Logit Ordinal para la ciudad de Santiago. El modelo presentado en la Tabla 4.4 es el mejor resultado de varios modelos anteriores que buscaron explicar el comportamiento de los usuarios con menor éxito. Anteriormente se intentó modelar a las tarjetas por separado según el quintil de ingreso de su comuna de residencia. Siendo necesario construir tres modelos para los tres quintiles más altos. Este enfoque resultó muy confuso a la hora de comparar los resultados y muchas variables presentaron los mismos efectos sobre la decisión de los usuarios, insinuando que

Tabla 4.3: Distribución de las variables enteras y continuas

	Unid.	Prom.	Des. Est.	Min.	25 %	50 %	75 %	Max.
Tiempo de Viaje	<i>min</i>	28,0	12,6	0,63	18,8	26,3	35,4	122,1
Tiempo de Espera	<i>min</i>	7,52	4,79	0,00	4,19	6,73	9,85	77,9
Tiempo de Caminata	<i>min</i>	1,83	1,30	0,02	1,00	1,53	2,29	29,1
Velocidad de Viaje	<i>m/s</i>	5,31	11,79	0,00	4,00	5,00	6,00	1.704
Variabilidad del TV	<i>min</i>	6,04	1,89	0,36	5,00	6,00	7,00	51,0
Porcentaje del TV Metro	%	0,38	0,32	0,00	0,07	0,35	0,64	1,00
ICF		0,95	0,05	0,20	0,93	0,96	0,98	1,00
ICR		0,81	0,05	0,00	0,78	0,81	0,84	1,00
Tasa Motorización	%	0,56	0,23	0,30	0,42	0,53	0,67	1,67
Servicios Únicos		9,20	3,95	1,00	6,00	9,00	12,0	42,0

la segmentación de la muestra no era necesaria. Finalmente, se decidió incluir una variable binaria que identifique a las tarjetas con zona de residencia en las comunas del quinto quintil (mayor ingreso), ya que es este el que posee la mayor diferencia con el resto de la muestra. A continuación se describen los coeficientes, sus valores y signos del modelo de migración y grado de uso del sistema de transporte público.

Tabla 4.4: Resultados Modelo Logit Ordinal en Santiago.

Coefficientes	Unid.	Valor	Error Est.	t-valor
Tiempo de Viaje Corto		-0,093	0,007	-12,694
Tiempo de Viaje Largo		0,161	0,006	24,906
Tiempo de Espera	<i>min</i>	-0,003	0,006	-5,689
Tiempo de Caminata	<i>min</i>	-0,017	0,002	-7,926
Variabilidad del TV.	<i>min</i>	-0,006	0,001	-4,374
Porcentaje del TV. en Metro	%	0,053	0,010	5,450
Usa Bus Expreso		-0,016	0,007	-2,514
Solo Laboral		0,239	0,007	32,767
Solo Periodos Punta		0,146	0,015	9,515
Servicios Únicos		0,010	0,008	11,284
Frecuente		0,372	0,007	54,806
Q5 * Velocidad [m/s]	<i>m/s</i>	0,003	0,001	2,862
Q5 * Usa Zonas Pagas		-0,049	0,018	-2,661
Q5 * Tasa de Motorización	%	-0,044	0,010	-4,542
Interceptos		Valor	Error Est.	t-valor
Ausente No Frec.		-0,291	0,015	-20,015
No Frec. Semi Frec.		0,367	0,015	25,334
Semi Frec. Frecuente		1,144	0,015	78,303
Desviación Residual:			1.142.830,24	
AIC:			1.142.864,24	

En un principio, los coeficientes de las variables Tiempo de Viaje Corto y Tiempo de Viaje Largo parecieran tener un efecto contra intuitivo en el comportamiento de las tarjetas. Esto ocurre por la fuerte correlación entre la distancia y el tiempo de viaje. Lo que las variables

posiblemente explican es que un viaje corto en transporte público puede ser fácilmente reemplazado por otros modos como la caminata, la bicicleta o el taxi. En cambio, un viaje largo normalmente no es posible reemplazarlo por un modo de transporte no-motorizado, y para ser reemplazado por un modo motorizado como el taxi o el auto es necesario incurrir en un alto costo. De esta forma, los usuarios que realizan viajes cortos tienen una mayor probabilidad de disminuir su uso del sistema o abandonarlo y los que realizan viajes largos son más fieles y frecuentes. Otro aspecto importante que revelan estas variables es la existencia de usuarios cautivos, que viven en los suburbios de la ciudad y realizan largos viajes para acceder a sus trabajos. Estos usuarios no poseen otra alternativa de movilización que el transporte público haciéndolos extremadamente fieles y frecuentes. Al comparar el valor de los coeficientes de estas dos variables se deduce que el efecto de la variable Tiempo de Viaje Largo es casi el doble sobre el comportamiento de los usuarios que el de la variable Tiempo de Viaje Corto.

Los coeficientes de los Tiempos de Espera y Caminata son negativos, como era de esperarse, aumentando la probabilidad de que los usuarios migren o disminuyan su uso del sistema a medida que aumentan. Marginalmente se puede apreciar que un minuto más de caminata tiene un efecto mucho mayor en la decisión de los usuarios que uno de espera.

Como se explicó anteriormente, la Variabilidad del Tiempo de Viaje asigna un puntaje a los servicios según la varianza en el tiempo que le toma al servicio en completar su expedición. Que esta esté asociada a un coeficiente de signo negativo significa que los usuarios que validan en servicios con un alto puntaje (variabilidad) perciben una baja calidad de servicio y serán propensos a abandonar el sistema o disminuir su uso.

La variable que mide el porcentaje del tiempo de viaje que una persona viaja en metro posee un coeficiente de signo positivo. Esto significa que los usuarios que más tiempo pasan viajando en metro, más fieles son con el sistema. Esto puede deberse al buen servicio que entrega metro con respecto a los buses. Pese a que existen altos niveles de hacinamiento en la red de trenes en ciertas estaciones y horarios, el servicio de trenes es confiable, veloz y cómodo la mayoría del tiempo.

El uso de los buses expresos posee un coeficiente de signo negativo, lo que es a priori contra intuitivo. No obstante, existe una correlación entre la ubicación de estos servicios y los pares origen-destino más críticos de la red. Los buses expresos son una medida que busca mejorar la calidad de servicio en pares altamente demandados que inicialmente ya reciben un mal servicio. Es por esto que pese a que estos servicios buscan mejorar la calidad de servicio, los usuarios que los utilizan son más propensos a la fuga que el resto de todas formas.

El uso de los buses expresos posee un coeficiente de signo negativo. Esto es a priori contra intuitivo. Los buses expresos buscan mejorar la calidad de servicio en pares de alta demanda que se encuentren muy alejados restringiendo el número de paradas a lo largo de la expedición. Como estos servicios especiales no son asignados de forma aleatoria, se cree que el problema ocurre por endogeneidad, donde pese a que esta medida ayuda a mejorar el servicio entregado, los usuarios de estos servicios

Los usuarios que validan solo en días laborales o los que validan solo en horarios punta tienen una alta probabilidad de ser usuarios frecuentes. Esto se puede deber a que el Transantiago fue diseñado para satisfacer la demanda de los horarios punta (buena conexión

entre zonas de hogares y trabajo) y no para satisfacer viajes en horarios valle o con otros propósitos, haciéndolo poco atractivo para viajar durante fines de semana o fuera de punta. Pese a que los horarios punta son un subconjunto de los días laborales, los coeficientes de las variables indican que las tarjetas que validan durante días laborales tienen una mayor probabilidad de ser frecuentes. Esto puede ser un indicio de que los usuarios que solo validan en horarios punta son usuarios cautivos y lo hacen no porque el transporte público les reporte algún grado de felicidad sino que por obligación. La variable Servicios Únicos que se utiliza como un proxy de la variabilidad espacial de los usuarios posee un coeficiente de valor positivo, indicando que a medida que un usuarios ocupa más servicios distintos, más alta es su probabilidad de ser un usuario frecuente.

La variable binaria Frecuente, que indica cuando una tarjeta fue clasificada como frecuente en el primer periodo, posee el mayor coeficiente de todo el modelo y es positivo. Esto muestra que los usuarios frecuentes del sistema no migrarán de él fácilmente y que su probabilidad de seguir utilizando frecuentemente el sistema es alta.

Se identificaron tres variables significativas para el conjunto de tarjetas con una estimación de residencia en las comunas del quinto quintil de Santiago. La velocidad de viaje, con un signo positivo, muestra que los usuarios de alto ingreso consideran importante el tiempo que les toma realizar sus viajes. Esto puede deberse a que estos usuarios poseen un mayor valor del tiempo y a que estos disponen de más alternativas de transporte gracias a su mayor ingreso. El uso de zonas pagas es significativo con un coeficiente de valor negativo, lo que indica que a estos usuarios no les agradan estos paraderos. Este tipo de paraderos permiten disminuir los costos del sistema al disminuir los tiempos de carga de pasajeros. Sin embargo, estas paradas operan en las zonas y servicios de mayor demanda del sistema, por lo tanto, su nivel de hacinamiento es muy alto en periodos punta. Al igual que con los servicios expresos estas variables sufren de un problema de endogeneidad por su correlación con los servicios más críticos del sistema. Se cree que esta es la razón por la cual los usuarios del quinto quintil se ven desalentados a usar el transporte público después de usar un paradero de Zona Paga. Por último, en comunas donde la tasa de motorización es más alta (Vitacura y Lo Barnechea) la probabilidad de abandonar o disminuir el uso del sistema es mayor por parte de sus residentes.

Finalmente, los tres puntos de corte o interceptos estimados por el modelo poseen valores en orden creciente y un t-valor mayor a 1,96 en su valor absoluto, lo que permite corroborar que las categorías se encuentran bien definidas.

Capítulo 5

Segundo escenario: Gatineau, Quebec, Canadá

A lo largo de este capítulo se describe la aplicación de la metodología propuesta para la ciudad de Santiago en un nuevo escenario. Este escenario es la ciudad de Gatineau en la región de Quebec, Canadá. El propósito de esto es poner a prueba la capacidad de aplicar la metodología para el caso de un sistema de transporte público diferente. Además, se busca confirmar los resultados obtenidos en Santiago y encontrar nuevos factores propios de la cultura y la ciudad que afecten la migración de usuarios.

5.1. Análisis de la base de datos

5.1.1. Descripción del escenario

El sistema de transporte público de la ciudad de Gatineau, llamado Sociedad de Transporte de Ottawa (STO) en español, es un sistema de mediano tamaño con 310 buses sirviendo cerca de 290.000 habitantes. Gatineau está ubicada junto a la ciudad de Ottawa, capital de Canadá, Por consiguiente, muchas de las líneas de buses en Gatineau también sirven la ciudad de Ottawa, donde viven cerca de 883.000 habitantes y se llevan a cabo muchas actividades relacionadas al trabajo. Gatineau introdujo el sistema de pago con tarjeta inteligente en el año 2001. El sistema de cobro de la STO funciona con tarifa plana por viaje donde los pasajeros solo deben validar al ingresar a los buses y tienen la opción de realizar un transbordo a otros buses en las siguientes dos horas. En Gatineau existen múltiples formas para pagar por el transporte público. Entre ellas se encuentran los abonos (mensuales o anuales), los pases (diarios o por tres días), los tickets (viaje único), el pago con efectivo y el pago con la tarjeta inteligente previa carga de dinero en esta. En el presente, sobre un 90 % de las transacciones en el sistema se realizan con tarjetas inteligentes (abonos, pases y uso regular de la tarjeta). En términos de infraestructura, el sistema cuenta con 37 *km* de vías exclusivas y más de 2.100 paraderos. En noviembre del 2013 se inauguró un sistema de corredor de buses de alto estándar BRT. El corredor BRT tiene diez estaciones en los sectores más demandados de la

ciudad. (Briand et al., 2017; STO, 2014).

5.1.2. Descripción de la base de datos

A diferencia de Santiago, donde el tamaño del sistema hace muy difícil trabajar con periodos prolongados, la cantidad de datos pasivos generados por el uso del sistema de transporte público en Gatineau es bastante menor. Esto facilita analizar periodos más extensos. Los datos disponibles comienzan el primero de septiembre del 2012 y terminan el 30 de octubre del 2014, comprendiendo un periodo de dos años. Durante este tiempo se llevaron a cabo 23,8 millones de transacciones en el sistema. En esta ocasión los datos se encuentran en su estado bruto y no han sido procesados anteriormente. Por lo tanto, los atributos disponibles no son capaces de caracterizar con tanto detalle la experiencia de viaje de los usuarios como en Santiago. Los principales atributos de la base de datos son:

- Identificador de la tarjeta.
- Tipo de tarjeta.
- Fecha y hora de la transacción.
- Identificador de la etapa.
- Servicio utilizado.
- Dirección del servicio.
- Paradero de la transacción.

A diferencia del primer escenario, en este no existe una estimación de la bajada de los viajes, por lo tanto, todos los atributos mencionados están relacionados a la subida del viaje. Es por esto, que no existe ningún atributo de calidad de servicio disponible.

A lo largo de estos dos años es posible observar hasta 70 tipos de tarjeta distintos. Esto se debe a un proceso de modernización y actualización de estas que se llevó a cabo durante este periodo. Al comienzo se observan sobre 50 tipos de tarjeta por mes, en cambio al final, estas no superan los 25 tipos. Los principales tipos de tarjeta son: Adulto, Estudiante, Tercera Edad y Niño. Luego existen sub categorías dependiendo del tipo de pago de la tarjeta o el tipo de estudiante que amplían el conjunto posible. Pese a existir un número tan grande de tarjetas, un 90 % de todas las transacciones se realizan solo por 16 de estas.

Al disponer de un periodo más extenso de datos, es posible realizar un análisis en el largo plazo del uso del sistema. En la Figura 5.1 se presenta la cantidad de transacciones agrupadas por semanas durante los dos años. Este análisis permite observar el fuerte efecto de las vacaciones de invierno en la cantidad de transacciones en el sistema. Durante las dos primeras semanas de enero el uso del sistema cae en un 75 % con respecto a sus semanas anteriores. En cambio, en verano (21 Junio - 21 Septiembre) la baja en el número de transacciones es más suave y se mantiene por más tiempo. Además, la figura permite observar el aumento en la cantidad de transacciones después de la inauguración del sistema de BRT. Tras su incorporación el uso del sistema aumenta en un 15-20 % con respecto a los mismos meses del año anterior.

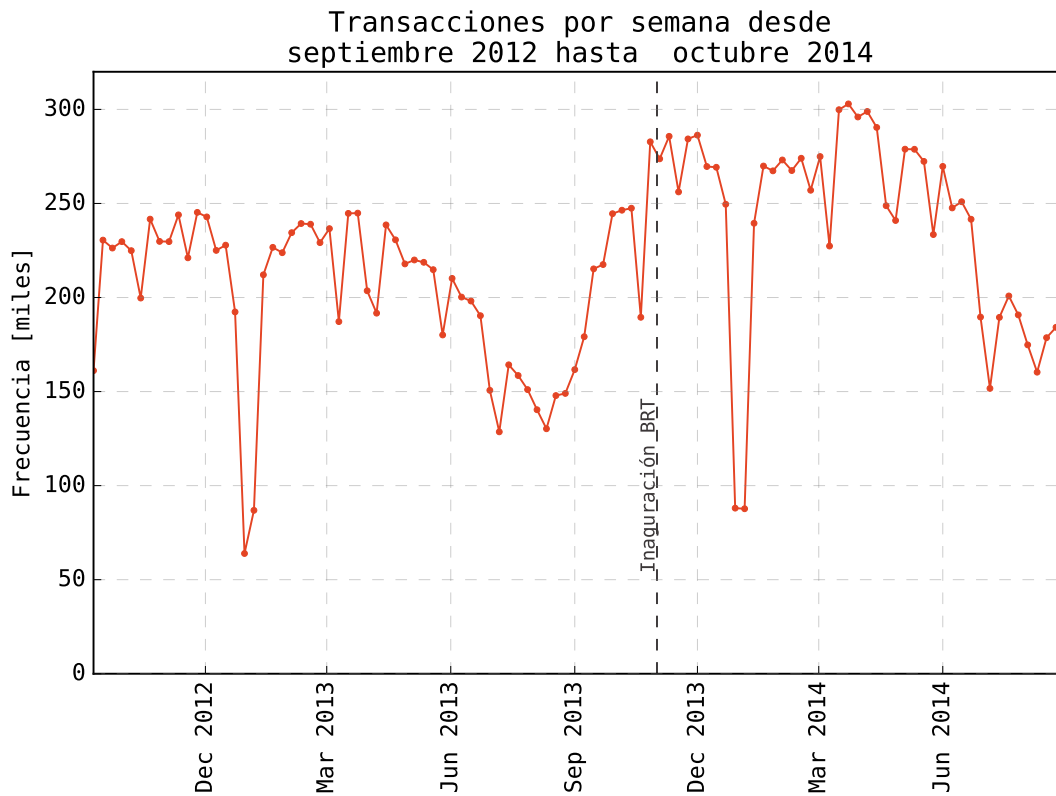


Figura 5.1: Distribución temporal de las transacciones en Gatineau, Quebec.

Utilizando la clasificación de usuarios mencionada en la Sección 4.1, es posible observar la evolución mes a mes de las tarjetas. Dependiendo de la clasificación basada en el grado de uso del sistema (Frecuente, Semi Frecuente y No Frecuente) en un mes y el mes anterior, las tarjetas fueron catalogadas en las siguientes cinco categorías:

- Mantiene: Tarjetas clasificadas en la misma categoría en ambos meses.
- Aumenta: Tarjetas que aumentan su uso del sistema con respecto al mes anterior.
- Disminuye: Tarjetas que disminuyen su uso del sistema con respecto al mes anterior.
- Migra: Tarjetas que están presentes el mes anterior pero no el actual.
- Nueva Tarjeta: Tarjetas que no estuvieron presentes el mes anterior y si en el actual.

En la Figura 5.2 se muestra el resultado de esta clasificación de las tarjetas según su evolución mes a mes. La figura muestra que los meses de mayor migración ocurren durante el verano (Junio y Julio) y que estos vienen seguidos de los meses con mayor cantidad de nuevas tarjetas (Agosto y Septiembre). Los meses de noviembre y diciembre muestran un aumento en la cantidad de tarjetas que disminuyen su uso del sistema y nuevamente vienen seguidos de meses donde aumenta la cantidad de tarjetas que utilizan más el sistema (enero y febrero). El mes de mayo de 2014 presenta un considerable aumento en las tarjetas que disminuyeron su uso del sistema. Este efecto no se aprecia el año anterior, por lo que pudo deberse a un evento particular.

Con el fin de replicar la metodología realizada en Santiago, es necesario escoger dos cortes

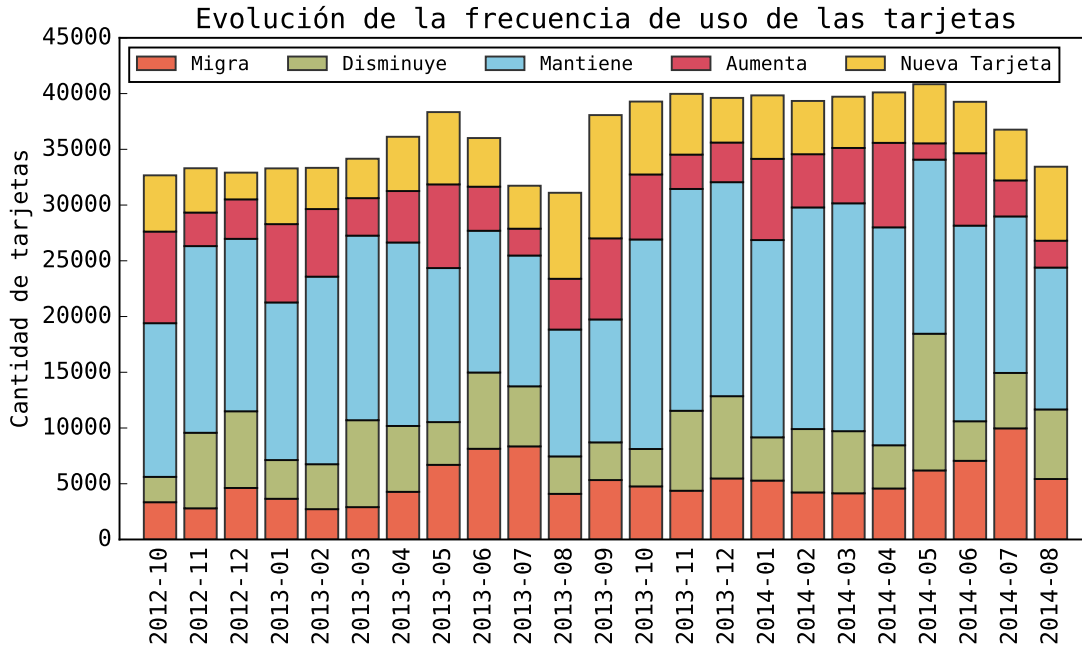


Figura 5.2: Evolución mes a mes de la clasificación de los usuarios.

temporales donde analizar el uso y migración del sistema. Con este fin se eligen dos cortes temporales que describen un uso habitual del sistema de transporte, que no incluyan periodos de vacaciones y que pertenezcan al mismo año escolar de Canadá. Esto último es para minimizar la probabilidad de que los usuarios cambien de residencia. El análisis de la cantidad de transacciones por semana y la evolución de la clasificación de los usuarios mes a mes han sido de gran ayuda para esta tarea, ya que permiten observar los periodos del año donde el sistema de transporte público está influenciado por efectos relacionados a la temporada o por cambios estructurales del mismo.

Los periodos escogidos son los meses de noviembre del 2013 y abril del 2014. En este caso, ambos periodos tendrán una extensión de un mes y, al igual que en Santiago, se analizará solo el comportamiento de las tarjetas de adultos. Los periodos están separados por cuatro meses y al interior del mismo año escolar en Canadá, como es deseado. Ambos periodos son posteriores a la inauguración del corredor BRT, por lo tanto, se asume que la utilidad que le reportan los modos de transporte a los usuarios permanece constante. Los datos de Gatineau son revisados en busca de observaciones atípicas o erróneas según los mismos criterios definidos para el primer escenario. Una vez seleccionadas las tarjetas que serán analizadas y revisadas las observaciones, los datos resultantes para el primer y segundo periodo son los descritos en la Tabla 5.1.

Tabla 5.1: Descripción de las bases de datos en Gatineau, Quebec.

	Noviembre 2013	Abril 2014
Número de Etapas	726.344	744.985
Número de Viajes	532.855	547.862
Tarjetas de Adulto	19.268	18.972

5.1.3. Primer periodo: Noviembre 2013

El comportamiento de los usuarios en un sistema de transporte público depende de muchos factores tales como la cultura, las jornadas laborales, el clima, etc. Es por esto, que es de esperarse que los patrones temporales de los ciudadanos de Gatineau no sean iguales a los de Santiago. Para esto, se analizará la distribución de las transacciones durante el mes de noviembre 2013. Como ambos periodos poseen la misma información y tienen la misma extensión, el análisis se centrará en el primer periodo, de donde se busca entender la experiencia de viaje percibida por los usuarios y, el segundo periodo será analizado posteriormente cuando se presente la clasificación de tarjetas por frecuencia de uso del sistema.

La Figura 5.3 muestra la cantidad de transacciones cada 15 minutos durante todo el mes mediante un mapa de calor. Esta figura permite observar la baja en la cantidad de transacciones por el “Día de la Remembranza” (11 de Noviembre), celebrado como un feriado en Canadá. Sin considerar este feriado, el mes escogido para analizar la experiencia de viaje de los usuarios es bastante regular, con una baja presencia de viajes durante los fines de semana y dos puntas muy marcadas en los días laborales.

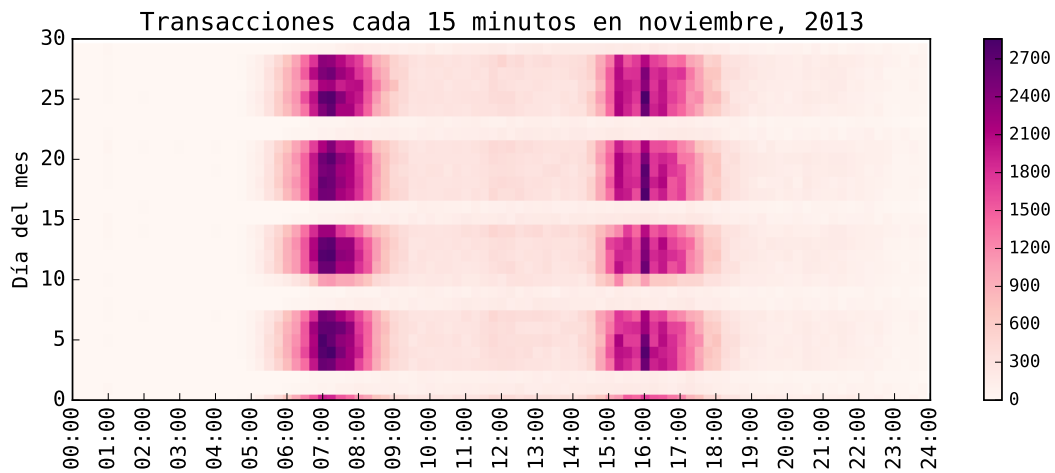


Figura 5.3: Mapa de calor de las transacciones del mes de Noviembre, 2013 en Gatineau.

La Figura 5.4 permite entender mejor el patrón temporal de los usuarios de la STO. En esta figura se presenta una semana promedio de noviembre del 2013 sin considerar el día feriado. El sistema comienza su operación a las 4:45 AM y rápidamente llega a la punta mañana (7:00 - 7:15 AM), donde se efectúan sobre 2.500 transacciones cada 15 minutos. Luego la cantidad de validaciones disminuye y se mantiene al rededor de las 400 validaciones cada 15 minutos entre las 9:30 AM y las 2:15 PM con una leve punta al medio día. La punta tarde comienza a las 2:15 PM y dura hasta las 7:00 PM. Esta punta posee tres picos a las 3:15, 4:00 y 4:30 PM respectivamente. Estos picos están relacionados a la salidas diferenciadas de estudiantes y trabajadores. Posteriormente, la cantidad de transacciones cada 15 minutos se mantiene entre las 200 y 300 hasta las 12:00 PM. Debido a la pequeña cantidad casi no son observables en la figura, pero existen validaciones a las 1:00, 2:00, 3:00 y 4:00 AM debido a la operación de servicios nocturnos. A diferencia de Santiago, la punta tarde del día viernes no comienza antes, sino que a lo largo de todo el día se observan menos transacciones. El

día sábado nunca supera las 500 transacciones y al igual que en los días laborales existe una punta en la mañana y en la tarde. Finalmente, el día domingo es bastante homogéneo, manteniéndose cerca de las 250 transacciones por cada 15 minutos durante todo el día.

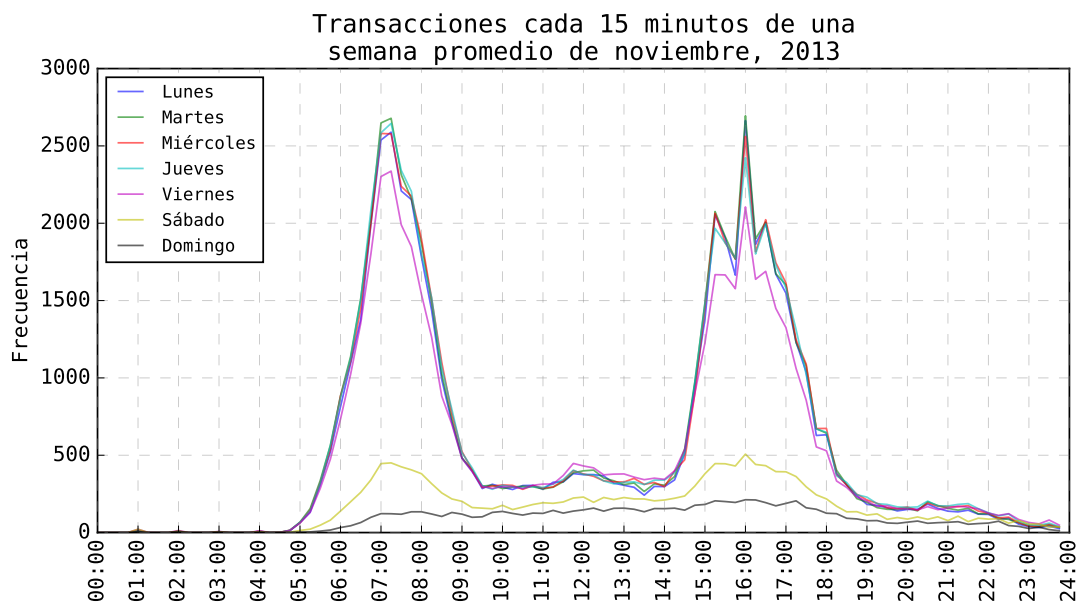


Figura 5.4: Distribución temporal de las transacciones en Gatineau, Quebec.

5.1.4. Censo de Canadá 2011

Los datos del Censo del Canadá del año 2011 (STATCAN, 2011) son utilizados para imputar atributos socio económicos a las tarjetas. Este censo se llevó a cabo por una agencia del gobierno de Canadá llamada “Statistics Canada” en mayo del 2011. De la información disponible, se han seleccionado un conjunto de variables para conocer los siguientes atributos de la zonificación de Gatineau: Ingreso medio por hogar, porcentaje de inmigrantes y partición modal. Tal como en el caso de Santiago, el ingreso medio por hogar se utiliza para identificar a las tarjetas que residen en zonas de alto o bajo ingreso. En la Figura 5.5 se puede apreciar el ingreso medio por hogar en la zonificación de Gatineau. Esta zonificación tiene solo fines estadísticos y de planificación urbana, por lo tanto las zonas no llevan nombres. En la figura se observa que la distribución del ingreso no es tan desigual como en Santiago, pero de todas formas, la diferencia entre las zonas más pobres y las más ricas es abrumadora. Esto fomenta que el quintil más pobre y el más rico sean identificados para posteriormente analizar si existen comportamientos particulares de las tarjetas que habitan en estas zonas.

La distribución espacial del ingreso en la ciudad de Gatineau es presentada en la Figura 5.6. En la figura se presentan las paradas de la STO como puntos naranjos y las comunas están clasificadas por tonalidades del color azul dependiendo del quintil al que pertenezcan. Los colores más claros están relacionados a las zonas más pobres y los más oscuros a las zonas más ricas. De la figura se desprende que los sectores más céntricos de la ciudad son los de menor ingreso, en cambio, en los suburbios el ingreso medio por hogar aumenta, sobretodo hacia el poniente de la ciudad.

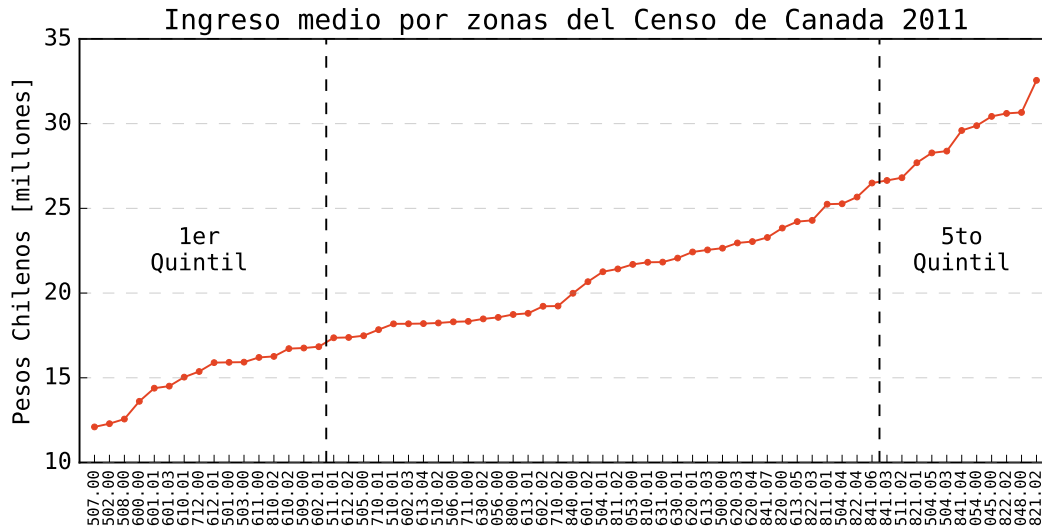


Figura 5.5: Distribución del ingreso en Gatineau, Quebec.

La partición modal por zona se presenta en la Figura 5.7, donde se puede apreciar que el vehículo particular, ya sea como conductor o como pasajero, es el modo de transporte más común en toda la región. Tanto es el uso del vehículo particular que en algunas zonas llega a tener un 80 % o más del total de viajes. El transporte público posee valores entre el 3 y el 30 % dependiendo de la zona. Finalmente los modos alternativos como la caminata, bicicleta y otros no son tan relevantes para la población de Gatineau.

5.2. Estimación de residencia

La metodología de estimación de residencia descrita por Amaya (2013) está diseñada para trabajar con periodos cortos de datos (1 semana). Como en este caso los periodos son de un mes, la probabilidad de que un usuario valide lejos de su residencia en un día laboral durante la mañana es más alta. Es por esto que la metodología fue adaptada a los datos disponibles. Ya que se mantiene el supuesto de que las personas pasan la noche en su residencia, el conjunto de observaciones que se selecciona es el mismo. Posteriormente, la metodología difiere para eliminar las observaciones más distantes del centro geográfico de las transacciones y así, solo observar la zona donde se concentra la mayoría de las validaciones. El procedimiento se describe a continuación:

1. Seleccionar la primera transacción entre las 4:00AM y las 12:00PM de todos los días laborales.
2. Calcular el centro geográfico (C.G.) de las transacciones seleccionadas.
3. Calcular la distancia euclidiana entre el C.G. y las transacciones seleccionadas.
4. Calcular la mediana y la desviación estándar de las distancias calculadas.
5. Eliminar las observaciones con una distancia mayor a la mediana más una desviación estándar.

Distribución espacial del ingreso en la ciudad de Gatineau

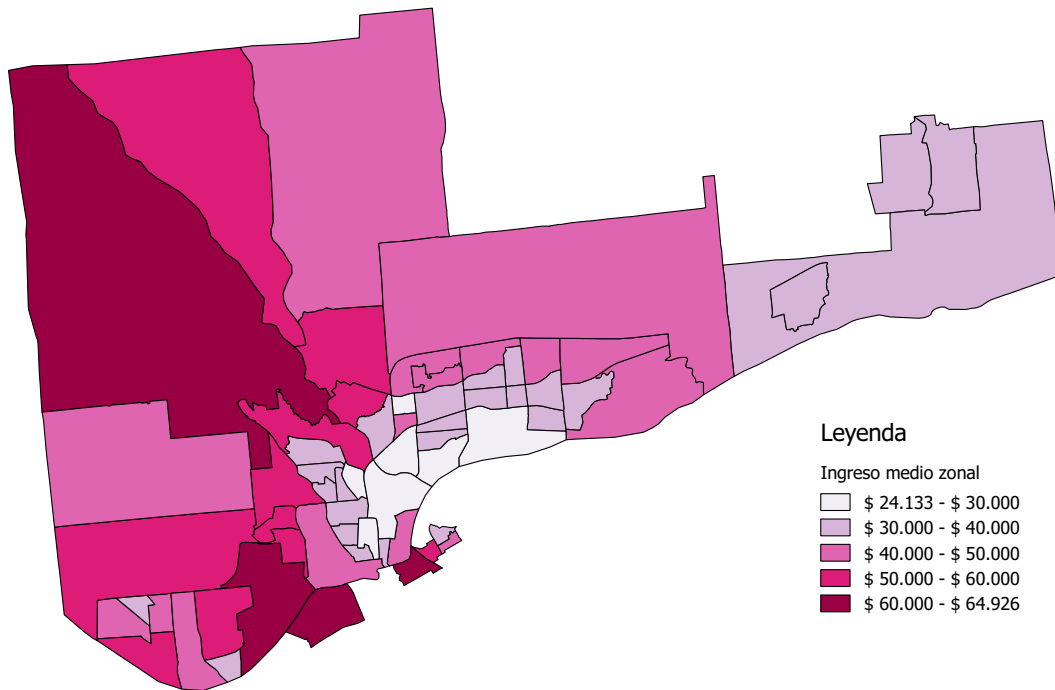
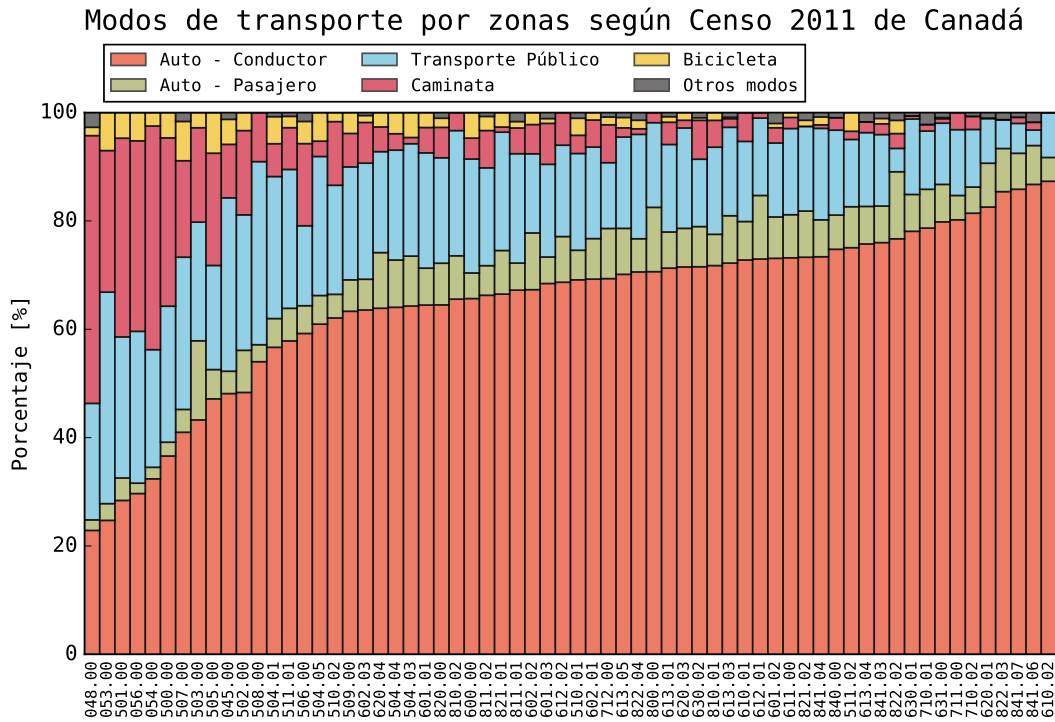


Figura 5.6: Distribución espacial del ingreso en Gatineau, Quebec.

6. Calcular el nuevo centro geográfico (N.C.G.) de las transacciones restantes.
7. Calcular la distancia euclidiana entre el N.C.G. y las transacciones restantes.
8. Eliminar aquellas tarjetas cuya máxima distancia euclidiana entre el N.C.G. y las transacciones sea mayor a 1.000 metros.
9. Escoger la zona de residencia como la zona más frecuente entre las transacciones finales.

En la Figura 5.8 se ejemplifica la estimación de zona de residencia de una tarjeta mediante la nueva metodología. En la imagen, los puntos (celestes y rojos) representan la ubicación de la primera transacción entre las 4:00AM y las 12:00PM de todos los días laborales del mes. La cruz roja es el centro geográfico de todas las transacciones mencionadas y el círculo rojo el rango definido por la mediana más una desviación estándar de las distancias entre el centro geográfico (cruz roja) y todas las transacciones. Los puntos (rojos) por afuera de este círculo son outliers y no son considerados para definir la zona de residencia de la tarjeta. Posteriormente se calcula el centro geográfico (cruz celeste) de las transacciones restantes (puntos celestes) y se verifica que estén todas dentro de un radio de 1.000 metros.



5.3. Modelación

El modelo propuesto busca explicar no solo la migración de usuarios sino también el grado de uso del sistema. Esto se lleva a cabo clasificando a los usuarios en cuatro categorías: Frecuente, Semi Frecuente, No Frecuente y Ausente. Las primeras tres categorías son el resultado de un algoritmo de clusterización y la última ocurre cuando una tarjeta deja de utilizar el sistema. El modelo es un modelo de rezago donde las variables explicativas son observadas en un primer periodo y la variable a explicar en un periodo posterior. El primer periodo es el mes de noviembre de 2013 y en este se busca capturar la experiencia de viaje percibida por los usuarios. El segundo periodo es el mes de abril de 2014 y en este se observa la decisión sobre el grado de uso o abandono del sistema por parte de los usuarios. Dado el orden natural de las categorías de la variable dependiente se utiliza un Modelo Logit Ordinal para calcular la probabilidad de pertenencia a las cuatro posibles categorías.

5.3.1. Clasificación de usuarios

El grado de uso del sistema se mide clasificando a los usuarios en tres categorías según la cantidad de transacciones y días viajados en el mes. La clasificación de estos se hace utilizando el algoritmo de clusterización k-medias. En la Figura 5.9 se presentan los resultados obtenidos de la clusterización. El tamaño de los círculos en la figura indica la cantidad de tarjetas que poseen la misma cantidad de transacciones y días viajados, los colores la pertenencia a las tres categorías y las cruces representan los centroides de cada clase. El tamaño de

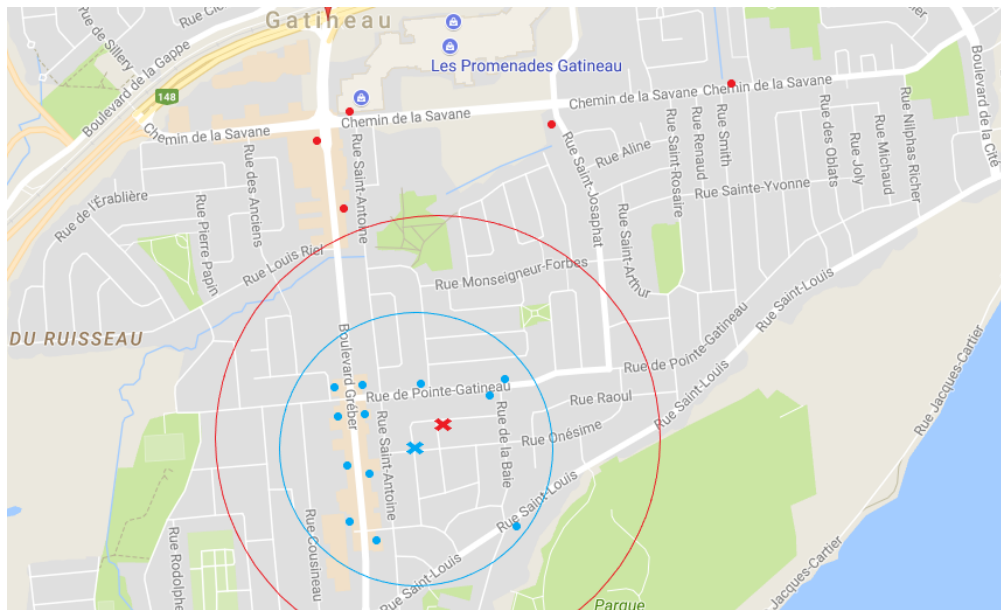


Figura 5.8: Ejemplo de nueva metodología de estimación de zona de residencia.

los círculos revela que existe un conjunto importante de tarjetas que viajan uno o dos días al mes y realizan menos de cinco transacciones. En el global, en este caso los grupos son más compactos y no existen tantas observaciones atípicas como en Santiago. El grupo de las tarjetas frecuentes es el de mayor dispersión con respecto al número de transacciones, siendo su máximo una tarjeta con 213 transacciones en 30 días. Observando los centroides se concluye que una tarjeta promedio de la categoría No Frecuente realiza 7,9 transacciones en 4,6 días, de la categoría Semi Frecuente 31,9 transacciones en 16,3 días y de la categoría Frecuente 63,7 transacciones en 21,3 días.

Al analizar el tamaño de los grupos y la cantidad de transacciones que realizan, vemos que un 50,6 % de las tarjetas son catalogadas como semi frecuentes. Estas son las más comunes en el sistema y también las que realizan más transacciones con 49,6 % del total. En cambio, el conjunto de las tarjetas frecuentes es el más pequeño con 22,4 % de tarjetas, no obstante, es responsable de un 43,9 % de todas las transacciones. Finalmente, las tarjetas no frecuentes son un 27,0 % y solo realizan 6,5 % de las transacciones. El análisis de la migración y grado de uso del sistema de los usuarios se realizará sobre las tarjetas frecuentes y semi frecuentes, las que en conjunto son un 73 % de todas las tarjetas y realizan un 93,5 % de todas las transacciones.

Si se analiza la evolución de la clasificación de los usuarios entre los dos periodos analizados, es posible construir un diagrama alluvial estático como el de la Figura 5.10. Observando las tarjetas frecuentes de noviembre del 2013, vemos que un 49,2 % de estas mantienen su uso del sistema y vuelven a ser clasificadas como frecuentes en abril del 2014. Un 33,0 % de este mismo grupo disminuye su uso del sistema y son catalogadas como semi frecuentes (29,4 %) y no frecuentes (3,6 %). El 17,8 % restante paso de ser catalogadas como frecuentes a estar ausentes en el sistema. En cambio, si se analizan las tarjetas semi frecuentes de noviembre del 2013, vemos que un 64,6 % de las tarjetas mantiene su uso del sistema. Las que aumentan su uso son un 6,2 % y las que lo disminuyen son un 9,3 %. Por último, un 19,9 % de este grupo

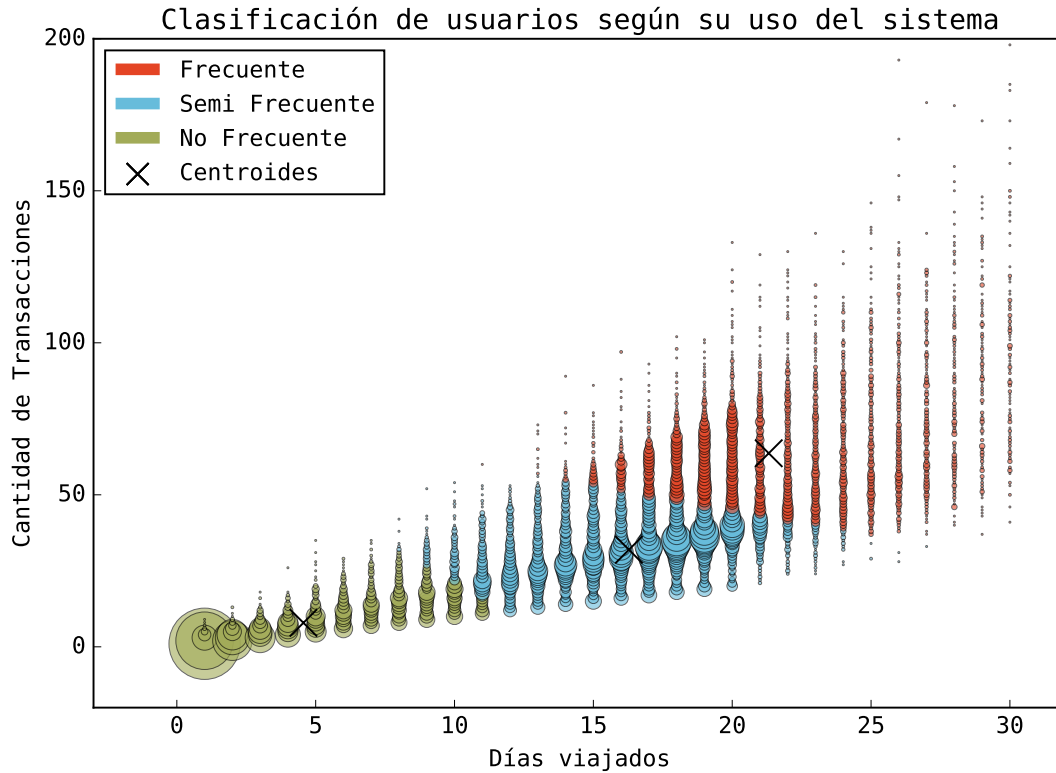


Figura 5.9: Clasificación de usuarios según su grado de uso del sistema.

migró del sistema. En el global, un 19,2% de las tarjetas catalogadas como frecuentes o semi frecuentes migró del sistema, un 59,5% mantuvo su grado de uso, ya sean frecuentes o semi frecuentes, y un 17,2% disminuyó su uso del sistema sin abandonarlo.

5.3.2. Variables independientes y dependiente

El registro de las transacciones realizadas en el sistema de Gatineau junto con el censo de Canadá del 2011 son utilizados para construir la base de datos de modelación. Esta base contiene una observación por tarjeta describiendo la experiencia de viaje del usuario, aspectos socio económicos sobre su estimación de residencia y su decisión del grado de uso del sistema en el segundo periodo. Al agrupar la información se construyen las variables independientes que luego serán introducidas en el modelo.

Variable dependiente

La distribución de la variable dependiente se puede observar en la Tabla 5.2. La categoría más común es la Semi Frecuente con 6.211 tarjetas, lo que equivale a un 52,8%. Luego la siguen las categorías Frecuente y No Frecuente con 2.423 (20,6%) y 869 (7,4%) respectivamente. Por último, la cantidad de tarjetas que no aparecen en el segundo periodo es de 2.252

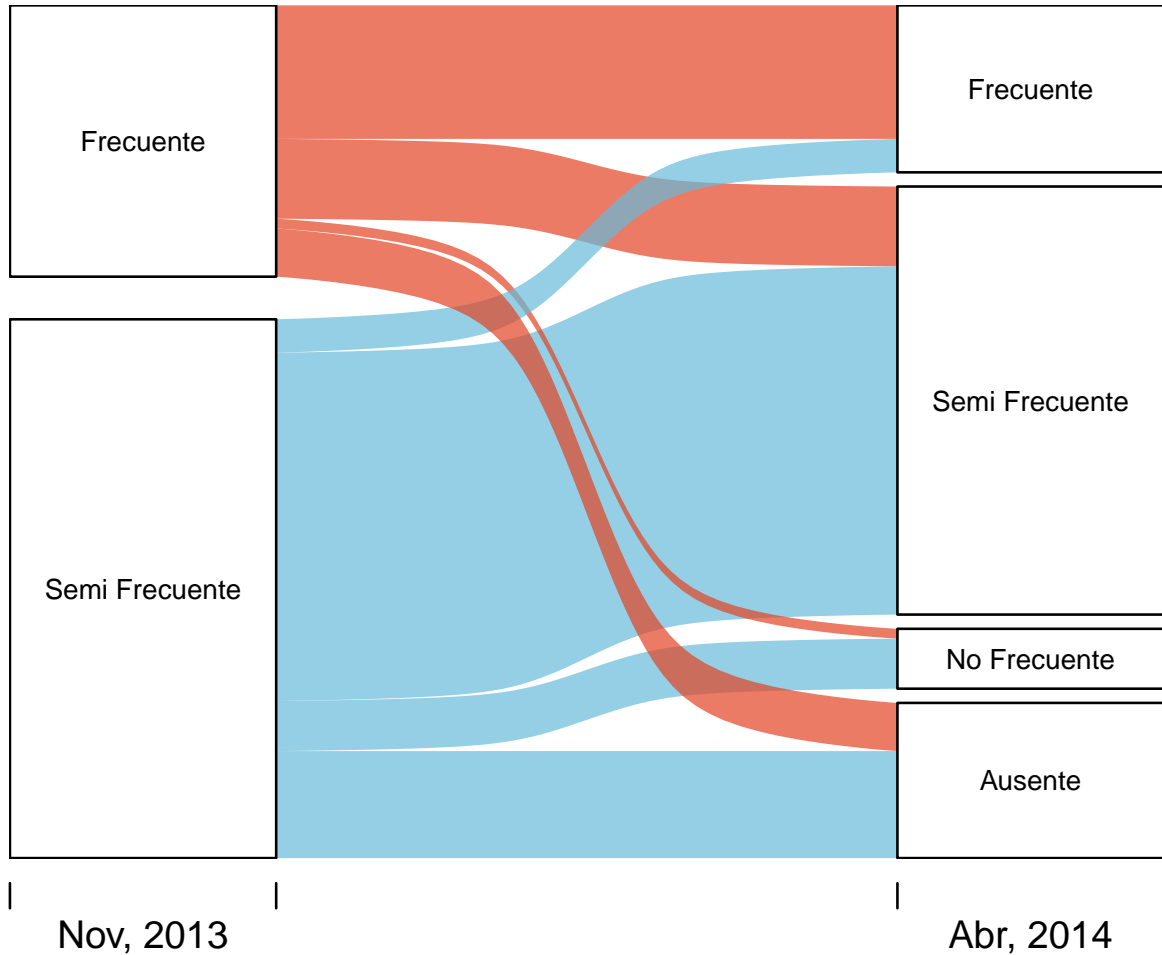


Figura 5.10: Evolución de la clasificación de usuarios según su grado de uso del sistema.

tarjetas, un 19,1% de la muestra. Este valor es bastante menor al de Santiago, donde las tarjetas ausentes eran las más comunes con un 36% del total.

Variables binarias

El patrón de viaje de los ciudadanos de Gatineau hacen que el sistema deba funcionar bajo dos estados muy marcados. El primero sirve los periodos punta, donde la mayoría de las personas viaja de sus hogares a sus puestos de trabajo y viceversa. El segundo son los periodos valle, donde se genera una cantidad de viajes muy inferior a los periodos punta. Para cubrir estas demandas el sistema posee servicios de buses que circulan todo el día y otros que circulan solo en horarios punta. El uso de estos últimos es identificado mediante dos variables binarias. La variable “Usa Servicio Punta” identifica a las 10.873 (92%) tarjetas que viajaron por lo menos una vez con uno de estos servicios y la variable “Solo Servicio Punta” hace lo mismo con las 744 (6%) tarjetas que solo utilizaron estos servicios durante el mes. Las proporciones de estas variables dan a entender que estos servicios son muy requeridos por la población, sin embargo, no suficientes para satisfacer todos los viajes de las personas.

Tabla 5.2: Distribución de la variable dependiente.

Clasificación en Noviembre		
Frecuente	2.423	(20,6 %)
Semi Frecuente	6.211	(52,8 %)
No Frecuente	869	(7,4 %)
Ausente	2.252	(19,1 %)
Total	11.755	(100,0 %)

El corredor BRT es utilizado por lo menos una vez en el mes por 6.652 tarjetas, lo que equivale a un 56 % del total. Después de su introducción en octubre de 2013, el sistema de transporte de Gatineau ha aumentado su cantidad de transacciones considerablemente. Es por esto que se espera que el servicio de BRT tenga un efecto en la decisión de las personas.

Por otra parte, los patrones temporales son observados mediante dos variables binarias que indican el uso del sistema solo durante días laborales y solo durante los periodos punta. La primera se llama “Solo Laboral” y toma el valor 1 en 8.715 (74 %) tarjetas. La segunda se llama “Solo Periodo Punta” y ocurre en 3.553 (30 %) tarjetas.

La clasificación que haya obtenido la tarjeta durante el primer periodo también es incorporada con una variable binaria llamada “Frecuente”. Esta toma el valor de 1 cuando la tarjeta fue catalogada como frecuente y 0 cuando fue catalogada como semi frecuente. La cantidad de tarjetas frecuentes y semi frecuentes es de 3.937 (34 %) y 7.818 (66 %) respectivamente.

El resumen de las proporciones de las variables binarias se presenta en la Tabla 5.3.

Tabla 5.3: Distribución de las variables binarias.

	1	0
Usa BRT	6.652 (56,6 %)	5.103
Usa Servicios Punta	10.873 (92,5 %)	882
Solo Servicios Punta	744 (6,3 %)	11.011
Solo Laboral	8.715 (74,1 %)	3.040
Solo Periodo Punta	3.553 (30,2 %)	8.202
Quintil I	2.118 (18,0 %)	9.637
Quintil V	2.696 (22,9 %)	9.059
Frecuente	3.937 (34,5 %)	7.818

VARIABLES CONTINUAS

Disponer de datos históricos del sistema permitió calcular la edad de las tarjetas. Esta se calculó como la cantidad de días desde su primera validación a partir del primero de septiembre del 2012. Esta variable se llama “Edad Tarjeta” y toma valores entre 0 y 425 días.

Observando la salida de los buses se construyó una variable que mida la frecuencia promedio de los servicios de buses. Esta calcula para cada servicio la cantidad promedio de

expediciones por hora. Pese a que en Gatineau los servicios funcionan con una programación por paradero y que los usuarios frecuentes suelen conocer el horario de sus buses. Disponer de un servicio más frecuente permite organizarse de mejor forma al tener más horarios de partida y llegada de los orígenes y destinos de los viajes de una persona. La variable se llama “Expediciones Promedio” y está en unidades de $[veh/h]$.

Para medir la variabilidad espacial de un usuario se utilizó el número de paradas distintas donde haya validado durante el mes. Esta variable asume que los usuarios que más actividades satisfacen viajando con el transporte público deben usar una mayor cantidad de paraderos distintos. Por lo tanto, un usuario que visita más paraderos distintos (mayor variabilidad espacial) durante el mes, se espera que sea más frecuente. Esta variable toma valores entre 1 y 60 y su media es de 8,9 paradas distintas en un mes por tarjeta.

Del Censo de Canadá se utilizaron el porcentaje de inmigrantes por comuna y la partición modal del auto y transporte público por comuna. Estas variables fueron imputadas de acuerdo a la estimación de zona de residencia de cada tarjeta. Además, se utilizó el ingreso para identificar las tarjetas que residen en las comunas del primer y quinto quintil de ingreso de la ciudad.

Los principales indicadores estadísticos de todas las variables continuas y enteras recién mencionadas se presentan en la Tabla 5.4.

Tabla 5.4: Distribución de las variables enteras y continuas.

	Unid.	Prom.	Des. Est.	Min.	25 %	50 %	75 %	Max.
Expediciones Promedio	veh/dia	4,87	0,96	1,00	4,21	4,86	5,48	8,02
Paradas Únicas		8,99	5,10	1,00	6,00	8,00	11,0	60,0
Edad Tarjeta	$dias$	222,4	120,9	0,00	150,0	192,0	332,0	425,0
% Inmigrantes	%	0,12	0,06	0,01	0,07	0,12	0,18	0,29
Partición Modal Auto	%	0,74	0,10	0,25	0,69	0,75	0,82	0,94
Partición Modal TPU	%	0,18	0,04	0,03	0,16	0,18	0,21	0,34

5.3.3. Resultados

Los resultados del Modelo Logit Ordinal para la ciudad de Gatineau son presentados en la Tabla 5.5. Debido a que los datos no eran tan ricos en información como en Santiago, el modelo posee menos variables propias del sistema. Sin embargo, los signos y significancia de las variables son satisfactorias. A continuación se describen los coeficientes, signos y valores de las variables del modelo resultante.

Lo primero que reportan los resultados, es que el uso del corredor de buses BRT tiene un efecto significativo sobre la decisión de los usuarios. Su signo es positivo, lo que significa que los usuarios de este modo poseen una mayor probabilidad de ser frecuentes que el resto. En cambio, las variables asociadas al uso de los servicios en horarios punta no fueron significativas. Esto puede deberse a la baja variabilidad de las variables.

En cuanto a los patrones temporales, los usuarios que utilizan el sistema solo durante

Tabla 5.5: Resultados Modelo Logit Ordinal en Gatineau, Quebec.

Variablen	Unid.	Valor	Error Std.	t-valor
Usa BRT		0,111	0,045	2,430
Solo Periodo Punta		0,134	0,040	3,302
Paradas Únicas		0,020	0,004	4,247
Partición Modal TPU	%	0,426	0,204	2,084
% Inmigrantes	%	-1,755	0,243	-7,220
Edad Tarjeta	dias	0,001	0,000	10,321
Frecuente		1,421	0,050	28,212
Q1 * Partición Modal Auto	%	-0,331	0,065	-5,095
Q5 * Expediciones Promedio	veh/dia	0,036	0,009	3,897
Interceptos		Valor	Error Std.	t-valor
Ausente No Frecuente		-0,712	0,096	-7,407
No Frec. Semi Frec.		-0,284	0,095	-2,970
Semi Frec. Frecuente		2,404	0,098	24,301
Desviación Residual:			25.850,00	
AIC:			25.874,00	

los periodos punta (mañana o tarde) poseen una mayor probabilidad de ser frecuentes. Esto puede deberse a la baja oferta de buses durante los periodos fuera de punta, lo que genera una disconformidad en los usuarios incluso cuando los buses cumplen un programa conocido. Por otra parte, pese a que un 74 % de los usuarios viaja solo en días laborales, la variable asociada a este comportamiento no fue significativa.

La variabilidad espacial entregó el resultado esperado. A mayor cantidad de paraderos distintos utilizados, mayor es la probabilidad del usuario de ser frecuente. Esto concuerda con el supuesto de que los usuarios que más actividades realizan en transporte público, son los más frecuentes y, para hacer esto deben usar más paraderos.

Gracias a la encuesta del Censo de Canadá, se observó que las tarjetas que residen en las zonas con mayor partición modal del transporte público, poseen una mayor probabilidad de ser usuarios frecuentes. También que las tarjetas que residen en comunas donde la población de inmigrantes es mayor porcentualmente, tendrán una mayor probabilidad de disminuir su uso del sistema o abandonarlo.

La edad de las tarjetas es significativa con un coeficiente de signo positivo, o sea, entre más antigua sea la tarjeta, mayor es la probabilidad de que su usuario sea frecuente. Este resultado concuerda con lo observado por Trépanier y Morency (2010). Es importante mencionar que las tarjetas incluidas en esta investigación no están siendo actualizadas durante los periodos analizados. Sino la migración se podría haber visto influenciada por este motivo.

Los usuarios catalogados como frecuentes en noviembre del 2013 tienen una probabilidad más alta de ser frecuentes en abril del 2014 que los que fueron semi frecuentes en el primer periodo.

Finalmente, se encontraron dos efectos particulares en las comunas pertenecientes al pri-

mer y quinto quintil de ingreso. Entre las tarjetas que residen en las zonas más pobre de Gatineau, el aumento en la partición modal del auto aumenta la probabilidad de fuga de los usuarios, como es de esperarse. Por otra parte, las tarjetas asociadas a las zonas más ricas de la ciudad poseen una probabilidad más alta de ser frecuentes cuando viajan en servicios con mayor frecuencia.

5.4. Síntesis

Resumiendo, en este capítulo se ha aplicado la metodología desarrollada en esta tesis en un nuevo escenario. Primero, se ha realizado un análisis gráfico de las transacciones en Gatineau, observándose comportamientos de la demanda propios de su cultura y el efecto en la demanda de la incorporación de un corredor de buses de alto estándar. También, se han utilizado los resultados del Censo de Canadá 2011 para caracterizar las tarjetas según su estimación de zona de residencia. La caracterización ha sido según su nivel de ingreso, identificándose el quintil más alto y más bajo de la población. La metodología de estimación de zona de residencia ha sido adaptada a los datos disponibles de la STO, lo que ha permitido eliminar transacciones outliers que bajo la metodología anterior impedirían asignar una zona de residencia a muchas tarjetas. Finalmente, se ha modelado la frecuencia de uso y migración de los usuarios del transporte público encontrándose factores que afectan de forma similar en Gatineau y Santiago y otros propios de la ciudad de Canadá.

Capítulo 6

Conclusiones

Esta tesis presenta una metodología para evaluar y explicar los cambios en el comportamiento de los usuarios del transporte público. Para llevar a cabo esta tarea, la investigación ha realizado un análisis gráfico y estadístico de los datos pasivos del Transantiago. Este análisis ha permitido observar el comportamiento de la ciudad y caracterizar a la muestra de tarjetas según su frecuencia de uso del sistema. En este capítulo se presentaran las contribuciones, limitaciones y posibles trabajos futuros de la metodología propuesta.

6.1. Contribuciones

Primero, la clasificación de usuarios según la frecuencia de uso del sistema ha demostrado ser una herramienta muy útil a la hora de analizar la composición de usuarios que viajan en el sistema. Gracias a esta herramienta es posible identificar la porción de usuarios responsables de la mayoría de los viajes e ingresos de toda la red. Además, permite conocer la existencia de un conjunto importante de tarjetas que están prácticamente en desuso. Poder identificar estas diferencias entre las tarjetas permite enfocar los recursos del sistema de mejor forma. Realizar este análisis en dos escenarios ha permitido advertir de los beneficios de tener tarjetas personalizadas. Es posible que la razón de que exista una gran cantidad de tarjetas en desuso en Santiago sea que los usuarios usan más de una tarjeta. Se cree que esta práctica disminuiría con la incorporación de tarjetas personalizadas.

Además, analizar la evolución en el tiempo de la clasificación de los usuarios permite ver la tendencia de cada grupo. En el caso de Santiago, se ha podido observar que la migración ocurre por partes iguales de las tarjetas frecuentes y semi frecuentes. En cambio en Gatineau, por cada tres tarjetas semi frecuentes que abandonan el sistema, una tarjeta frecuente lo hace. Este análisis es mucho más valioso cuando se dispone de un periodo más extenso de datos como en el segundo escenario. En este se analizó la evolución de las tarjetas a lo largo de dos años, identificándose patrones que se repiten de un año al siguiente.

Por otra parte, el resultado de esta investigación ha sido en parte gracias al esfuerzo de varios autores que anteriormente han contribuido en el procesamiento de los datos pasivos del

Transantiago. El gran valor de estos trabajos se puede apreciar al trabajar con datos de otros sistemas de transporte público que no han dedicado tantos recursos en enriquecerlos. En Gatineau, al disponer solo de datos brutos se restringe el abanico de posibilidades para enfrentar cualquier problema. No obstante, los datos han sido muy valiosos y han permitido analizar en detalle el comportamiento de las ciudades. Gracias al análisis gráfico de las transacciones en ambos escenarios ha sido posible identificar los periodos punta y valle durante un día, las diferencias entre los tipos de día de una semana, los días feriados en un mes, la respuesta de los usuarios con respecto a cambios en el sistema (la incorporación de los servicios BRT en Gatineau) o incluso han sido suficientes para identificar eventos más pequeños que hayan alterado la cantidad de transacciones con respecto a sus patrones temporales regulares. La georeferenciación de las validaciones ha permitido descubrir espacial y temporalmente los puntos críticos la red. En Santiago, se analizó la cantidad de transacciones en la Línea 1 del metro, identificándose las estaciones de metro más demandadas. Posteriormente se han observado las distribuciones temporales de estaciones ubicadas en zonas céntricas y periféricas de la línea, encontrándose patrones opuestos de la demanda.

Lo más importante de los datos, es que su potencial no yace en la información que tienen disponible hoy en día, sino en el valor que pueda obtener uno de ellos. Esta tesis aprovechó el registro de las salidas de buses en el Transantiago para crear un nuevo indicador de calidad de servicio relacionado a la variabilidad en los tiempos de viaje de las expediciones de buses. Este indicador permite ordenar los servicios según su desempeño y analizar el efecto que tienen en el comportamiento de los usuarios. No era el objetivo de esta tesis, sin embargo, este indicador puede servir para encontrar falencias sistemáticas del sistema en ciertos lugares o horarios que aumentan la variabilidad en el tiempo de viaje de los buses. En Gatineau, la falta de atributos que describan la calidad de servicio motivó la construcción de la frecuencia real de buses por hora. Se le llama real porque esta es construida a partir de las salidas de buses observadas y no del plan operacional. Esta variable permite controlar y corroborar el cumplimiento de los planes operacionales por parte de las empresas de buses. Así, existen un sin fin de utilidades para estos datos que aun no han sido descubiertos.

En cuanto a los resultados del modelo, ha sido posible corroborar los resultados encontrados por otros autores y también se han encontrado particularidades propias de cada escenario. Primero, se ha corroborado que al verse empeoradas las variables de calidad de servicio, los usuarios disminuirán su uso del sistema o lo abandonarán y que en las comunas donde mayor es la tasa de motorización, mayor será la fuga de usuarios. En Santiago, se concluye que cuanto mayor sea el uso del metro, mayor será la probabilidad de un usuario de ser frecuente. En cambio en Gatineau, los usuarios más frecuentes son los que viajan en el nuevo corredor de buses BRT. En ambos escenarios se ha encontrado que los usuarios con alta probabilidad de ser frecuentes son aquellos que viajan sólo durante días laborales, los que poseen mayor variabilidad espacial y los que han sido clasificados como frecuentes anteriormente. Además, en ambas ciudades se ha observado que existen variables relevantes solo para ciertos estratos socio económicos. En Santiago, la decisión de los usuarios pertenecientes a las comunas más ricas se encuentra afectada por la velocidad media de viaje, los paraderos con servicios de zonas pagas y la tasa de motorización de la comuna de residencia. En cambio en Gatineau, son los usuarios pertenecientes a las comunas más pobre los que ven afectada su decisión del grado de uso del sistema por la tasa de motorización y los de las comunas más ricas por la frecuencia de buses.

Al observar la distribución de las variables dependientes, se ha corroborado la tendencia de movilidad urbana en Santiago. Existe un porcentaje importante de usuarios que está viajando menos en transporte público y probablemente más en auto. En cambio en Gatineau, el porcentaje de usuarios que migran es muy inferior al de Santiago, demostrando que esa ciudad está pasando por un momento totalmente distinto en temas de movilidad urbana.

Como una última contribución, esta tesis tuvo que adaptar la estimación de residencia propuesta por Amaya et al. (2017) debido a la diferencia de los datos disponibles en Gatineau. La nueva metodología propuesta no ha sido aun validada. No obstante, es un punto de partida para estimar la residencia de usuarios cuando se dispone de periodos más extensos de datos.

6.2. Limitaciones

Tal vez una de las limitaciones más importantes de esta metodología es el criterio para definir una tarjeta ausente. Este criterio no es capaz de saber si una tarjeta esta ausente por la decisión del usuario de dejar de usar el sistema o por razones de fuerza mayor como enfermedades, viajes de trabajo, extravío de la tarjeta u otros. Este problema se busca solucionar observando una ventana temporal lo más extensa posible en el segundo periodo pero aun así, esta medida está lejos de ser perfecta.

Otro supuesto que no es siempre correcto, es que cada tarjeta es utilizada por un usuario único y que cada usuario posee una sola tarjeta. En un sistema donde las tarjetas son anónimas, es normal observar que las tarjetas son compartidas entre amigos y familiares dificultando la tarea de entender el patrón de viajes de una persona.

También, el modelo asume que los modos de transporte alternativos permanecen constantes entre los periodos. Sin embargo, pueden haber existido cambios en la infraestructura para vehículos particulares o de bicicleta, cambios en los precios de la bencina o incluso cambios en el mismo Transantiago que afecten la utilidad que le reporta cada alternativa a los usuarios.

Las variables socio económicas utilizadas son limitadas y están restringidas a describir las zonas de residencias de los usuarios. Ideal sería poseer información demográfica de los portadores de las tarjetas para la modelación.

Otro factor importante que afecta la decisión de los usuarios es el clima. El primer periodo ocurre en abril y el segundo en noviembre. Las temperaturas más agradables del verano y el cese de las lluvias inducen a los usuarios a utilizar modos alternativos de transporte como la bicicleta o la caminata.

Por último, la base de datos está sujeta a errores causados por los instrumentos de medición, por humanos o por las metodologías de procesamiento. Estos errores son, en lo posible, eliminados o corregidos, sin embargo, nunca se tiene la certeza de no estar sujeto a errores no observados.

6.3. Trabajos Futuros

Como trabajos futuros, esta investigación puede ser complementada en varias aristas. A continuación se presentarán las más relevantes para el autor.

- Disponer de un periodo más extenso de datos procesados para la creación de las variables explicativas, donde se pueda capturar de forma más amplia el comportamiento de los usuarios y la experiencia que reciben por parte del transportes público.
- Implementar las mejoras descritas por Munizaga et al. (2017) a la metodología que calcula los tiempos de viaje, espera y caminata. De esta forma se considerarían los tiempos de ingreso y egreso a las estaciones de metro y se eliminarían otros errores.
- Incorporar la evasión del pago de la tarifa en el estudio. Corrigiendo el número de transacciones por sector y/o paradero.
- Expandir la cantidad de cortes temporales a observar. Esto permitiría hacer un seguimiento más detallado de los usuarios y el grado de uso del sistema. Además, sería posible observar tarjetas que migran y luego vuelven al sistema.
- Mejorar el algoritmo de clusterización utilizado para clasificar a los usuarios. En particular, buscar un algoritmo que clasifique a los usuarios en conjuntos más marcados y distintos entre si.
- Implementar un algoritmo que sea capaz de solucionar el problema de las tarjetas extraviadas observando los patrones temporales y espaciales de estas, emparejando aquellas que posean una alta probabilidad de pertenecer al mismo dueño.

Capítulo 7

Bibliografía

- Amaya, M. (2013). Análisis de patrones de viaje utilizando datos masivos de transporte público. Tesis para calificar al título de Ingeniero Civil, Universidad de Chile.
- Amaya, M., R. Cruzat y M. Munizaga (2017). Estimating the residence zone of frequent public transport users to make travel pattern and time analysis. Working Paper, Universidad de Chile.
- Bagchi, M. y P. White (2005). The potential of public transport smart card data. *Transport Policy* 12 (5), 464–474.
- Bass, P., P. Donoso y M. Munizaga (2011). A model to assess public transport demand stability. *Transportation Research Part A: Policy and Practice* 45 (8), 755–764.
- Briand, A.-S., E. Côme, M. Trépanier y L. Oukhellou (2017). Analyzing year-to-year changes in public transport passenger behaviour using smart card data. *Transportation Research Part C: Emerging Technologies* 79, 274 – 289.
- Calvo, M. F. (2015). Análisis desagregado del comportamiento de usuarios de transporte público utilizando datos masivos. Tesis para calificar al título de Ingeniero Civil, Universidad de Chile.
- Chriqui, C. y P. Robillard (1975). Common bus lines. *Transportation science* 9 (2), 115–121.
- Devillaine, F. (2012). Estimación de viajes y actividades en base a sistemas tecnológicos de transporte público. Tesis para calificar al título de Ingeniero Civil, Universidad de Chile.
- DTPM (2017). Informe de gestión 2015-2016. Disponible en http://www.dtpm.cl/archivos/INFORME_GESTION_DTPM_15_16_2.pdf.
- Figler, S., P. Sriraj, E. Welch y N. Yavuz (2011). Customer loyalty and Chicago, Illinois, transit authority buses: Results from 2008 customer satisfaction survey. *Transportation Research Record: Journal of the Transportation Research Board* (2216), 148–156.
- Imaz, A., K. Habib, A. Shalaby y A. O. Idris (2015). Investigating the factors affecting transit

- user loyalty. *Public Transport* 7 (1), 39–60.
- Kieu, L.-M., A. Bhaskar y E. Chung (2015). A modified Density-Based Scanning Algorithm with Noise for spatial travel pattern analysis from smart card AFC data. *Transportation Research Part C: Emerging Technologies* 58, 193–207.
- Kim, S. y G. Ulfarsson (2012). Commitment to light rail transit patronage: Case study for St. Louis Metrolink. *Journal of Urban Planning and Development* 138 (3), 227–234.
- Kusakabe, T. y Y. Asakura (2014). Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C: Emerging Technologies* 46, 179–191.
- Langlois, G., H. Koutsopoulos y J. Zhao (2016). Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C: Emerging Technologies* 64, 1–16.
- Lee, S. G. y M. Hickman (2014). Trip purpose inference using automated fare collection data. *Public Transport* 6 (1-2), 1–20.
- Ma, X., Y.-J. Wu, Y. Wang, F. Chen y J. Liu (2013). Mining smart card data for transit rider’s travel patterns. *Transportation Research Part C: Emerging Technologies* 36, 1–12.
- Morency, C., M. Trépanier y B. Agard (2007). Measuring transit use variability with smart-card data. *Transport Policy* 14 (3), 193–203.
- Munizaga, M., F. Devillaine, C. Navarrete y D. Silva (2014). Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies* 44, 70–79.
- Munizaga, M., C. Núñez y A. Gschwender (2017). Smart card data for wider system evaluation. En: Kurauchi, J.-D., F. y Schmöcker (Ed.), *Public Transport Planning with Smart Card Data*. CRC Press, Cap. 9, pp. 163–179.
- Munizaga, M. A. y C. Palma (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies* 24, 9–18.
- Muñoz, V., A. Thomas, C. Navarrete y R. Contreras (2016). Encuesta Origen–Destino de Santiago 2012: Resultados y validaciones. *Ingeniería de Transporte* 19 (1).
- Núñez, C. (2015). Cálculo de indicadores de calidad de servicio para el sistema de transporte público de Santiago a partir de datos pasivos. Tesis para calificar al título de Ingeniero Civil, Universidad de Chile.
- Olguín, J., S. Jara-Díaz y M. Munizaga (2009). Análisis de patrones de actividades a partir de la EOD 2001. *Ingeniería de Transporte* 13 (4).
- Pelletier, M.-P., M. Trépanier y C. Morency (2011). Smart card data use in public transit: A

- literature review. *Transportation Research Part C: Emerging Technologies* 19 (4), 557–568.
- Sectra (2012). Encuesta Origen–Destino Santiago 2012. Disponible en <http://www.sectra.gob.cl/biblioteca/detalle1.asp?mf=3253>.
- STATCAN (2011). Censo de Canadá. Disponible en <http://www12.statcan.gc.ca/census-recensement/2011/index-eng.cfm>.
- STO (2014). Rapport annuel 2014. Disponible en <http://www.sto.ca/index.php?id=rapporannuel>.
- Train, K. (2002). Ordered responses. En: *Discrete Choice Methods with Simulation*. Cambridge University Press, Cap. 7, pp. 182–186.
- Trépanier, M., K. Habib y C. Morency (2012). Are transit users loyal? Revelations from a hazard model based on smart card data. *Canadian Journal of Civil Engineering* 39 (6), 610–618.
- Trépanier, M. y C. Morency (2010). Assessing transit loyalty with smart card data. En: *12th World Conference on Transport Research*, Lisbon, Portugal.
- Trépanier, M., N. Tranchant y R. Chapleau (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems* 11 (1), 1–14.
- Utsunomiya, M., J. Attanucci y N. Wilson (2006). Potential uses of transit smart card registration and transaction data to improve transit planning. *Transportation Research Record: Journal of the Transportation Research Board* (1971), 119–126.

Anexo A

Tablas del pre procesamiento de datos

Las Tablas A.1 y A.2 presentan los atributos y una corta descripción de su contenido para las tablas resultantes del pre procesamiento de las transacciones al ser agrupadas por etapas y viajes respectivamente.

Tabla A.1: Tabla de etapas

Variable	Descripción
id	Número identificador de la tarjeta
nviaje	Correlativo asociado al viaje de la tarjeta
netapa	Correlativo asociado a la etapa dentro del viaje
tipo_trasporte	Bus, Metro o Zona Paga
serv_un_zp2	Código que especifica el servicio y su sentido
tipo_dia	Laboral, Sábado o Domingo
mediahora	Periodos de 30 min al que pertenece la hora de la transacción
tiempo_subida	Fecha y hora de la transacción
tiempo_bajada	Fecha y hora de la bajada estimada
x_subida	Coordenada X de la transacción
y_subida	Coordenada Y de la transacción
par_subida	Paradero o estación de metro de inicio de la etapa
par_bajada	Paradero o estación de metro de termino estimado de la etapa
comuna_subida	Comuna donde se ubica el origen de la etapa
comuna_bajada	Comuna donde se ubica el destino estimado de la etapa
tiempo_etapa	Tiempo total entre la una transacción y la siguiente
tiempo_trasbordo	Tiempo entre la bajada estimada y la siguiente transacción del mismo viaje
tiempo_caminata	Tiempo necesario caminando entre la bajada estimada y el paradero de la siguiente etapa del mismo viaje
tiempo_espera	Tiempo de transbordo menos el tiempo de caminata
disteuclid	Distancia euclidiana entre el paradero de origen y el paradero estimado de destino
distonroute	Distancia sobre la ruta entre el paradero de origen y el paradero estimado de destino

Tabla A.2: Tabla de viajes

Variable	Descripción
id	Número identificador de la tarjeta inteligente
nviaje	Correlativo asociado al viaje de la tarjeta
netapa	Correlativo asociada a la etapa dentro del viaje
netapassinbajada	Cantidad de etapas sin una estimación de bajada
ultimaetapaconbajada	1: En caso que la última etapa posea una estimación de bajada. 0: Si no.
etapas	Resumen de los servicios utilizados en cada etapa del viaje
contrato	Tipo de contrato de la tarjeta
tipodia	Laboral, Sábado o Domingo
mediahora	Periodos de 30 minutos durante un día al que pertenece la hora de la transacción
paraderosubida	Paradero o estación de metro de inicio del viaje
paraderobajada	Paradero o estación de metro del destino final estimado del viaje
comunasubida	Comuna donde se realizó la transacción de inicio del viaje
comunabajada	Comuna donde se estimo el fin del viaje
tiemposubida	Fecha y hora de la transacción de inicio del viaje
tiempobajada	Fecha y hora estimada del fin del viaje
periodosubida	Periodo definido por Transantiago al que pertenece la hora de la transacción de inicio del viaje
periodobajada	Periodo definido por Transantiago al que pertenece la hora estimada del fin del viaje
dviajeeuclidiana	Distancia euclidiana del total del viaje
dviajeenruta	Distancia sobre la ruta del total del viaje
tviaje_min	Tiempo total del viaje
t_X_etapa	Tiempo total de la etapa X
d_X_etapa	Distancia euclidiana de la etapa X
ttrasbordo_X_etapa	Tiempo de transbordo de la etapa X
tcaminata_X_etapa	Tiempo de caminata de la etapa X
tespera_X_etapa	Tiempo de espera de la etapa X
serv_X_etapa	Servicio y sentido de la etapa X
tipotransporte_X	Tipo de transporte de la etapa X

Anexo B

Periodos definidos por Transantiago

Las Tablas B.1, B.2 y B.3 presentan los periodos definidos por Transantiago para caracterizar la oferta y la demanda por transporte público en Santiago para los días laborales, sábados y domingos respectivamente.

Tabla B.1: Periodos Transantiago en días laborales

Horario	Periodo
00:00 - 01:00	01 - Pre Nocturno
01:00 - 05:30	02 - Nocturno
05:30 - 06:30	03 - Transición Nocturno
06:30 - 08:30	04 - Punta Mañana
08:30 - 09:30	05 - Transición Punta Mañana
09:30 - 12:30	06 - Fuera de Punta Mañana
12:30 - 14:00	07 - Punta Mediodía
14:00 - 17:30	08 - Fuera de Punta Tarde
17:30 - 20:30	09 - Punta Tarde
20:30 - 21:30	10 - Transición Punta Tarde
21:30 - 23:00	11 - Fuera de Punta Nocturno
23:00 - 00:00	12 - Pre Nocturno

Tabla B.2: Periodos Transantiago en días sábado

Horario	Periodo
00:00 - 01:00	01 - Pre Nocturno Sábado
01:00 - 05:30	02 - Nocturno Sábado
05:30 - 06:30	03 - Transición Nocturno Sábado
06:30 - 11:00	04 - Punta Mañana Sábado
11:00 - 13:30	05 - Mañana Sábado
13:30 - 17:30	06 - Punta Mediodía Sábado
17:30 - 20:30	07 - Tarde Sábado
20:30 - 23:00	08 - Transición Sábado Nocturno
23:00 - 00:00	09 - Pre Nocturno Sábado

Tabla B.3: Periodos Transantiago en días domingo

Horario	Periodo
00:00 - 01:00	01 - Pre Nocturno Domingo
01:00 - 05:30	02 - Nocturno Domingo
05:30 - 09:30	03 - Transición Domingo Mañana
09:30 - 13:30	04 - Mañana Domingo
13:30 - 17:30	05 - Mediodía Domingo
17:30 - 21:00	06 - Tarde Domingo
21:00 - 23:00	07 - Transición Domingo Nocturno
23:00 - 00:00	08 - Pre Nocturno Domingo