



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

ESTIMACIÓN Y CARACTERIZACIÓN DE DEMANDA DE CLIENTES DE CENTROS COMERCIALES USANDO DATOS DE MOVILIDAD

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

VIOLETA LEONTINA CORTÉS BRIONES

PROFESOR GUÍA:
MARCEL GOIC FIGUEROA

MIEMBROS DE LA COMISIÓN:
ALEJANDRA PUENTE CHANDÍA
JULIO COVARRUBIA CASTRO

Este trabajo ha sido parcialmente financiado por:
FONDECYT 1221711

SANTIAGO DE CHILE
2023

ESTIMACIÓN Y CARACTERIZACIÓN DE DEMANDA DE CLIENTES DE CENTROS COMERCIALES USANDO DATOS DE MOVILIDAD

Los establecimientos comerciales están envueltos en una industria con altos grados de dinamismo y competencia, es por esto que necesitan mejorar y actualizar sus modelos predictivos, con el fin de capturar nuevos factores relevantes, utilizando herramientas más complejas disponibles en la actualidad gracias al avance tecnológico. De este modo, este trabajo de memoria tuvo como objetivo desarrollar un modelo de predicción de demanda usando datos masivos de movilidad, capaz de describir la probabilidad de asistir a un centro comercial, con el fin de comprender el comportamiento y características del cliente, logrando ser un apoyo en la toma de decisiones.

En este contexto, se desarrollaron Modelos Gravitacionales, los cuales representan una perspectiva pionera en el ámbito del comercio minorista. Estos modelos se centran en la estimación de la probabilidad de elección de una alternativa específica, considerando factores como la atracción ejercida por una infraestructura comercial y los tiempos de viaje involucrados. Además, se han realizado Modelos de Regresión y Machine Learning, los cuales han sido calibrados utilizando datos de movilidad. Esta diversidad de enfoques y metodologías ha sido la principal motivación detrás de la realización de esta memoria, ya que se espera que aporte a la capacidad predictiva y robustez en la estimación de la demanda de clientes en el contexto de los centros comerciales.

Tras el desarrollo y aplicación de los modelos junto con la selección de variables pertinentes, se llevó a cabo un análisis comparativo. Los resultados revelaron que el modelo más efectivo corresponde a un algoritmo de random forest, el cual incorpora atributos tanto de origen, destino como de transporte. Esto resalta la importancia de utilizar características que abarquen el contexto general para obtener estimaciones más precisas y confiables. Con estas variables se explica el 51% de la varianza de probabilidad de elección, lo que es un resultado aceptable considerando que predice la probabilidad de todas las manzanas de la Región Metropolitana a la mayoría de centros comerciales ubicados en ella.

Se aplicó el modelo propuesto para estimar demanda, calcular ventas potenciales, realizar caracterización de clientes y estimar impactos frente a cambios del contexto comercial, demostrando las utilidades del modelo como herramienta para entregar información relevante en toma de decisiones.

Dentro de las conclusiones, se menciona que se logra evidenciar las bondades de utilizar datos de movilidad en comparación a fuentes de datos tradicionales como encuestas, teniendo una mejoría del estimada del 64%. Además, se confirma que variables como distancia y tiempo de viaje tienen un impacto negativo significativo en la disposición de viaje a un destino de compra.

Agradecimientos

El primer lugar quiero agradecer y dedicar este trabajo a mi familia, principalmente a mi mamá y papá, que siempre han confiado en mí y me han enseñado con sus actos a ser buena persona, hacer el bien y siempre ver lo positivo.

Agradezco al profesor Marcel Goic, por haberme escuchado siempre que solicité ayuda, por entregarme conocimientos y desafíos. Me siento muy agradecida de poder haber realizado este trabajo con un gran académico y excelente persona. También a la profesora Alejandra Puente, por estar siempre atenta a mis dudas y aportar en mi formación como futura profesional. Además agradecer a Gabriel y Diego, mis compañeros de proyecto por su buena disposición y amabilidad siempre.

También quiero agradecer a mis mejores amigas Mana y Nuñez, por estar conmigo siempre, ser mis compañeras de toda la vida y ser las personas en las que siempre confiaré. A mis amigos que conocí aquí en la universidad el primer día de inducción y nunca solté Cata, Miguel, Toño, Chris, Pipe y Bryan, los quiero mucho, seis años sufriendo y compartiendo con ustedes. A mis amigos de industria Caro y Felipe, mis compañeros de grupo en todos los cursos que se podía elegir, gracias a la vida que me los presentó, a personas tan amables, buenas de corazón, inteligentes y dedicadas.

Agradecer al deporte, música y yoga que alimentaron mi alma en todo este proceso.

A cada uno de ustedes muchas gracias de corazón.

Violeta.

Tabla de Contenido

1. Introducción	1
1.1. Antecedentes Generales	1
1.2. Descripción del proyecto y justificación	2
1.3. Objetivos	3
1.3.1. Objetivo general	3
1.3.2. Objetivos específicos	3
1.4. Alcances	4
2. Marco Conceptual	5
2.1. Encuestas de Selección de tienda	5
2.2. Modelos de elección y selección de ubicación de tienda	6
2.2.1. Modelos Gravitacionales	6
2.2.1.1. Modelos de Huff	7
2.2.1.2. Modelos MCI	8
2.2.2. Modelos Logit	10
2.2.3. Modelos de Regresión	11
2.2.4. Modelos de Machine Learning	11
2.2.4.1. Modelos Random Forest	12
2.3. Minería de datos (CRISP-DM)	13
2.4. Metodología	15
3. Desarrollo metodológico	16
3.1. Procesamiento y limpieza de datos	16
3.1.1. Datos de movilidad	17
3.1.2. Datos de infraestructura comercial	18
3.1.3. Datos de población	19
3.1.4. Datos de sistema de transporte	20
3.2. Selección de atributos a estudiar	21
3.2.1. Atributos de Destino	22
3.2.2. Atributos de Origen	22
3.2.3. Atributos de Transporte	23
3.3. Elección y creación de modelos a realizar	24
3.3.1. Desarrollo Modelos Gravitacionales	25
3.3.2. Desarrollo Modelos de Regresión	27
3.3.3. Desarrollo de Modelos Random Forest	28
3.4. Evaluación de modelos	29
3.4.1. Métricas de rendimiento de modelos	29

3.4.2.	Simulación de encuesta	30
3.4.3.	Resultados Modelos	31
4.	Aplicaciones del modelo	38
4.1.	Estimación de demanda	38
4.2.	Estimación de ventas potenciales	42
4.3.	Estimación variación de demanda	45
5.	Conclusiones y Trabajo Futuro	49
5.1.	Conclusiones	49
5.2.	Trabajo futuro	50
	Bibliografía	51
	Anexos	55
A.	Procesamiento y limpieza de datos	55
A.1.	Gráficos de Análisis Descriptivo	55
A.1.1.	Gráficos de Datos de movilidad	55
A.1.2.	Gráficos de Datos de infraestructura comercial	56
A.1.3.	Gráficos de Datos de población	60
A.1.4.	Gráficos de Datos de Sistema de Transporte	63
B.	Selección de Atributos	64
B.1.	Atributos Centros Comerciales	65
B.2.	Atributos Población	66
B.3.	Atributos Transporte	67
C.	Elección, desarrollo, resultados y evaluación de modelos	67
C.1.	Desarrollo, resultado y evaluación de modelos	67
C.2.	Aplicación de modelos	79
C.2.1.	Estimación de demanda	80
C.2.2.	Estimación de ventas potenciales	84

Índice de Tablas

3.1.	Stats Datos de movilidad	18
3.2.	Combinación de atributos utilizados	31
3.3.	Resumen de resultados	32
3.4.	Variables Importantes RF	33
3.5.	Comparación resultados: Costanera Center	35
3.6.	Visualización probabilidad estimada Costanera Center	36
4.1.	Ingreso de hogar por GSE	43
4.2.	Caso 1: Florida Center	45
4.3.	Caso 2: Parque Arauco	46
A.1.	Origen de Visitas Registradas	56
A.2.	Ubicación Centros Comerciales RM	57
A.3.	Sector Centros Comerciales	57
A.4.	Superficie Centros Comerciales	58
A.5.	Zonas RM. Fuente: Diario Financiero	61
A.6.	Nivel socioeconómico manzanas RM	61
B.1.	Variables atributos destino	65
B.2.	Variables Atributos origen	66
B.3.	Variables sistema de transporte	67
C.1.	Resultados de la simulación de la encuesta	67
C.2.	Resultados Huff	68
C.3.	Resultado modelo 1.2	68
C.4.	Resultados MCI origen destino sin transformación	69
C.5.	Resultados modelo 1.3	70
C.6.	Resultados modelo 1.4	71
C.7.	Resultados modelo 1.5	73
C.8.	Resultados modelo 2.1	74
C.9.	Resultados versiones modelo 2.2	74
C.10.	Resultados versiones modelo 2.3	75
C.11.	Resultados versiones modelo 2.4	76
C.12.	Resultados versiones Random Forest	77
C.13.	Métricas de error de probabilidad por comuna	78
C.14.	Métricas de error de probabilidad por centro comercial	79
C.15.	Estimación de demanda Modelo A	80
C.17.	Regresión Modelo A estimación de demanda	80
C.16.	Estimación de demanda Modelo B	81
C.18.	Regresión Modelo B estimación de demanda	82
C.19.	Tabla resumen resultados completo	83
C.20.	Estructura de gasto de Hogar. Fuente: INE. Elaboración propia	84

Índice de Ilustraciones

2.1.	Modelos de selección y ubicación de tienda. Elaboración propia	6
2.2.	CRISP-DM. Fuente: Health Data Miner	14
3.1.	Datos utilizados. Elaboración propia.	16
3.2.	Distribución espacial centros comerciales RM	18
3.3.	División del territorio. Elaboración propia.	19
3.4.	Grupo de variables por atributos	25
3.5.	Distribución MAE Random Forest	34
4.1.	Visitas estimadas y reales (Modelo A)	39
4.2.	Visitas estimadas y reales (Modelo B)	40
4.3.	Área comercial centros comerciales	42
4.4.	Ventas potenciales estimadas	44
4.5.	Estaciones línea 7	46
4.6.	Transporte a Parque Arauco. Fuente: Google Maps.	46
A.1.	Boxplot Probabilidad de Visita	55
A.2.	Cantidad de tiendas por Centro Comercial	59
A.3.	Correlación variables de destino	60
A.4.	Correlación variables de origen	62
A.5.	Área comercial de manzanas	62
A.6.	Cercanía a estación de metro	63
A.7.	Disponibilidad de recorrido de buses	64
A.8.	Correlación variables de movilidad	64

Capítulo 1

Introducción

1.1. Antecedentes Generales

En Chile, el *retail* moderno, comenzó a tomar forma a partir de la década de 1980, cuando surgieron las primeras cadenas de supermercados y tiendas por departamentos, como Almacenes París y Falabella (Ortega, 2009). Durante las décadas siguientes, el *retail* chileno se expandió y diversificó, siendo actualmente uno de los más desarrollados y competitivos de América Latina. La industria del *retail*, tiene una gran importancia económica en Chile, llegando a participar del 20% aproximadamente del PIB nacional (Suarez, 2022).

Según la Cámara de Centros Comerciales (2018) en Chile, existen 278 centros comerciales a lo largo del país, sumando un total de 4.4 millones de m^2 de infraestructura, dando lugar a 4.417 tiendas, sin considerar los 21 proyectos de destinos comerciales que se encuentran en construcción. La tarifa promedio anual de alquiler dentro de estas instalaciones es alta, alcanzando las 8,58 UF por metro cuadrado en el Parque Arauco (Mendez, 2020), uno de los principales actores junto a Mall Plaza y Cencosud. El alto valor de alquiler dentro de los centros comerciales se ve justificado por el gran flujo de demanda que atraen al ser destinos populares para el consumo, el entretenimiento y la interacción social, obteniendo un total de 757 millones de visitas anuales en nuestro país (Cámara de Centros Comerciales, 2018).

Con el propósito de garantizar que las tiendas dentro de un centro comercial generen ganancias suficientes para su sostenibilidad, resulta fundamental identificar formas de aumentar la demanda y asegurar un consumo continuo. Una estrategia eficaz implica analizar el comportamiento de compra de los clientes para desentrañar los factores que influyen en sus visitas. En el entorno actual, existen metodologías para evaluar y perfilar la demanda de los clientes en los centros comerciales, como las técnicas de minería de datos. Estas metodologías aprovechan una variedad de modelos, incluyendo análisis de series temporales y otros enfoques predictivos, para examinar y pronosticar patrones de demanda en función del tiempo y de factores contextuales. Es importante destacar que estos modelos tienen la capacidad de ajustar sus parámetros utilizando diversas fuentes de datos. Entre estas, algunas de las fuentes más comunes son:

- 1.- Encuestas a clientes: Las encuestas directas, recopilan datos mediante preguntas de comportamiento de compra. Son útiles para estimar la demanda total, identificar patrones de conducta, segmentación de clientes y obtener características demográficas.

2.- Registros de ventas: Las transacciones tienen información relevante como número de compras diarias y productos vendidos, con estos datos se pueden realizar estimaciones de demanda de períodos de tiempo para determinar por ejemplo, días de mayor consumo, productos más demandados y otros índices de comportamiento de compra por parte de los clientes que efectúan una compra.

Con esto, se obtienen estimaciones de demanda futura y predicciones de comportamiento de compra útiles para personalizar ofertas, planificar demanda e identificar segmentos de clientes. La combinación de estas y otras metodologías, permite tener una visión más completa de la demanda. En particular, para este trabajo, se hará énfasis al uso de datos de movilidad, los cuales permiten capturar de manera precisa los viajes realizados por los clientes a los centros comerciales.

1.2. Descripción del proyecto y justificación

Evaluar la demanda de los clientes y comprender sus características es esencial para orientar decisiones estratégicas en una organización. Por ende, contar con herramientas que respalden estas acciones adquiere gran relevancia. En la práctica, los centros comerciales recurren a encuestas y modelos para definir áreas comerciales y entender el comportamiento de los clientes. Sin embargo, estas fuentes de información enfrentan desafíos considerables: su implementación conlleva un alto costo económico debido a la complejidad logística, lo que resulta en limitaciones tales como la falta de actualización de datos y la incapacidad para reflejar las nuevas dinámicas de consumo.

La oportunidad de mejora identificada radica en enriquecer la comprensión espacial de la demanda de clientes y los atributos que impactan en sus visitas a centros comerciales, aprovechando la utilización de datos de movilidad. Esta fuente de información posibilita la obtención regular de datos a gran escala y a menor costo. Ampliar el conocimiento acerca de las características y comportamiento de los clientes resultaría beneficioso para los centros comerciales en diversas áreas, tales como la comprensión más precisa de la demanda. Al identificar los factores que influyen en las decisiones de los consumidores y al comprender sus distribuciones espaciales, los centros comerciales podrían segmentar de manera más precisa a su público.

La caracterización espacial más sólida de los clientes, unida a modelos de estimación de demanda, permitiría la detección de oportunidades de mejora al brindar una comprensión más detallada de los factores que influyen en las decisiones de los consumidores, lo que facilitaría la asignación estratégica de recursos en áreas cruciales para los clientes y la tienda. Adicionalmente, esta información permitiría calcular ventas potenciales basadas en características socioeconómicas como el ingreso y el presupuesto familiar.

Un aporte en los elementos mencionados anteriormente producen una toma de decisiones con mayor información relevante, lo que decanta en una mejor estrategia y un aumento de ingresos para los centros comerciales. Considerando la oportunidad de mejora detectada, se propone realizar una serie de modelos capaces de capturar las visitas a los centros comercia-

les. Para conseguirlo, se pretende replicar los modelos utilizados en la literatura del caso y desarrollar otros modelos que mejoren el entendimiento de la demanda y sus características. En particular, se elabora un estudio comparativo entre los modelos de (Huff, 1963), Multiplicative Competitive Interaction (MCI), Regresión y Random Forest, usando distintos tipos de variables, las cuales fueron agrupadas por destino, origen y sistema de transporte, con esto se espera:

1. Conocer virtudes de utilizar datos masivos de movilidad.
2. Comparar desempeño al desarrollar distintos tipos de modelos y agregar categorías de variables.
3. Levantar atributos relevantes de los centros comerciales, personas y transporte al momento de visitar una infraestructura comercial.
4. Entender comportamiento de clientes.
5. Aplicar modelo para estimar demanda de clientes.
6. Estimar ventas potenciales.
7. Estimar variaciones en la demanda al modificar infraestructura de centros comerciales y del sistema de transporte al abrir una estación de metro.

1.3. Objetivos

1.3.1. Objetivo general

Desarrollar un modelo usando datos de movilidad que describa la probabilidad de asistir a un centro comercial, con el fin de comprender el comportamiento y características del cliente y que sea un apoyo en la toma de decisiones.

1.3.2. Objetivos específicos

1. Estudiar el estado de arte actual del problema y distinguir los modelos que expliquen la elección de visita a tienda por parte del cliente, con el fin de conocer oportunidades de innovación e identificar conjunto de variables usadas para describir este tipo de viaje.
2. Determinar métricas que permita evaluar el desempeño de los modelos, para poder valorar, comparar y elegir los con mejores resultados.
3. Diseñar e implementar modelos de estimación de demanda que explique visita a centros comerciales para lograr capturar la naturaleza de los clientes de los diversos *malls*.
4. Evaluar desempeño de los modelos implementados según los resultados obtenidos.

5. Proponer aplicaciones útiles del modelo, evidenciando sus virtudes en toma de decisiones estratégicas.

1.4. Alcances

Los modelos se ajustarán utilizando datos de movilidad de clientes de una compañía telefónica que fueron efectuadas en octubre del 2021, por lo que el desplazamiento puede estar cargado de elementos vinculados a ese momento o segmento de clientes, no considerando posibles factores de estacionalidad, temporalidad o sucesos particulares. Se trabajará además con datos demográficos recolectados del CENSO 2017, debido que es la fuente de información más completa y actual a la fecha de la población estudiada a nivel de manzanas.

Finalmente, no se considerará la puesta en marcha del modelo en los centros comerciales, correspondiente a la etapa de implementación de CRISP-DM, ya que el modelo realizado no pretende reemplazar ningún modelo o herramienta actual dentro de la organización ni se está trabajando en forma colaborativa con algún centro comercial en particular.

Capítulo 2

Marco Conceptual

En este capítulo, se explora los elementos esenciales vinculados a la predicción de la demanda en centros comerciales. Igualmente, se proporcionará una comprensión del estado actual de conocimiento en esta área, identificando las limitaciones existentes, que aspectos o características se pueden replicar y cuáles otros pueden ser un espacio de mejora dentro de este mismo estudio.

2.1. Encuestas de Selección de tienda

Se define el área comercial de una tienda como un lugar delimitado con clientes potenciales que tienen una probabilidad mayor a cero de efectuar una compra (Huff, 1964). En cambio para (Ghosh & McLafferty, 1987) es el área geográfica donde existe una alta penetración de mercado y se encuentran la mayoría de clientes, como el área comercial central de (Applebaum, 1966). Applebaum recomienda subdividir en tres el área de estudio, donde el área comercial primaria de los supermercados metropolitanos sea el 60 % de las ventas, esta área generalmente presenta menor distancia a la tienda y tiene mayor densidad de clientes por población y ventas per cápita. La segunda y tercera área son el 20 % cada una.

Para delimitar y estimar con precisión estas áreas comerciales, se ha recurrido a la implementación de encuestas. Applebaum (1966) presenta un enfoque para determinar estas áreas, la penetración en el mercado de las tiendas existentes y un método para estimar ventas potenciales mediante la recopilación de información de encuestas y datos censales.

El proceso de estimación involucra la entrevista a los compradores para obtener sus patrones de compra y ubicación. A través de esta información, se cuantifican los clientes por zonas geográficas específicas. Además, se recopila la población de estas zonas y se calculan las ventas per cápita.

El análisis de estas herramientas de estudio han proporcionado valiosos resultados. En Huff (1964), se destacan hallazgos cruciales que subrayan la relación entre la distancia entre el origen y el área comercial, la variedad de mercancía ofrecida y la disposición de viajar de los consumidores. Asimismo, se evidencia cómo la proximidad a la competencia influye en la 'atracción' de un área comercial. Estos elementos entregan indicios sobre las dinámicas que rigen el comportamiento del consumidor y sus preferencias en relación con las ubicaciones de

compra.

2.2. Modelos de elección y selección de ubicación de tienda

Se han desarrollado diferentes modelos que ayudan a los minoristas a comprender las preferencias de los consumidores y seleccionar las ubicaciones más adecuadas para sus tiendas. En la documentación existente se reconocen varios enfoques de estimación que se diferencian en supuestos de comportamientos y datos utilizados. Según (Vandell & Carter, 1993) en la literatura consideran la ubicación de tiendas minoristas en dos niveles básicos:

- 1) Modelos de elección de tienda: analizan el proceso de elección de tienda del consumidor.
- 2) Modelos de ubicación de tienda: analizan la elección de locación, tamaño y características de tiendas que optimizan rendimiento.

Ambos niveles de análisis se complementan entre sí, lo que significa que la investigación en modelos de elección de tiendas puede contribuir al avance de la investigación en modelos de ubicación de tiendas, es por esto que se combinan los modelos de la siguiente forma:

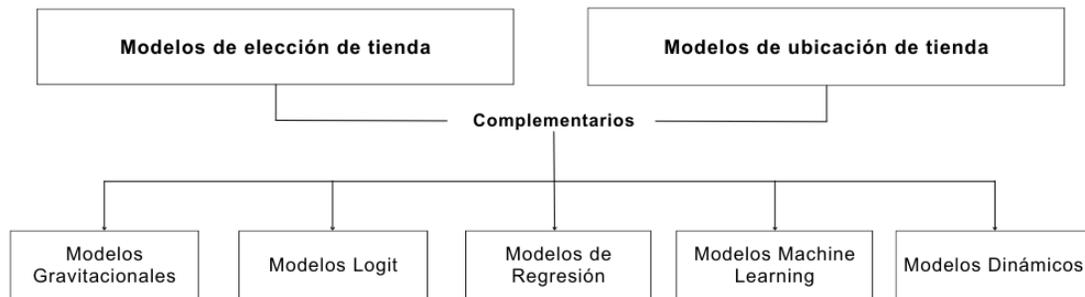


Figura 2.1: Modelos de selección y ubicación de tienda. Elaboración propia

2.2.1. Modelos Gravitacionales

Los modelos gravitacionales son una herramienta que permite la investigación de las interacciones entre distintas entidades en una variedad de campos, incluyendo el análisis de mercados, la geografía, la economía y la demografía. Estos se basan en la ecuación de gravedad planteada por Isaac Newton, la cual manifiesta que la fuerza de atracción entre dos cuerpos, es directamente proporcional al producto de sus masas y la constante de proporcionalidad (G) e inversamente proporcional al cuadrado de la distancia entre ellos.

Si bien, es un modelo que explica la atracción entre dos cuerpos, en la actualidad su uso es mucho más amplio. Para este caso de estudio, la literatura ha sido expandida para incluir diversas caracterizaciones de cómo los clientes deciden sus tiendas de destino. A continuación se describen algunas de las más relevantes.

2.2.1.1. Modelos de Huff

La teoría de la gravitación probabilística de Huff (1963) es una revisión de las teorías de la gravitación minorista de Reilly (1931) y Converse (1949) en una teoría aplicable al mundo real. El modelo desarrollado por Reilly establece que la atracción de las grandes ciudades sobre el poder adquisitivo de las ciudades más pequeñas cercanas está directamente relacionado con la población de las grandes ciudades y se ve afectado inversamente por el cuadrado de la distancia a cada una de ellas. Reilly llamó al modelo, dado a su gran capacidad predictiva como "la ley gravitacional del retail".

Además, determina un 'punto de quiebre' donde los clientes entre dos ciudades consideran ambas, sin distinción, como lugares adecuados. Sin embargo, este punto de quiebre da la impresión de que el área de comercio es fija y los clientes tienen un comportamiento homogéneo. Además, el exponente 2 está pensado para ciudades interurbanas aun cuando se ha estudiado que oscila entre 1.5 y 3 dependiendo del tipo de viaje y contexto geográfico en ciudades urbanas.

La ley original de gravitación en el ámbito del retail describía el poder de atracción sobre el poder adquisitivo de dos ciudades importantes. Posteriormente, experimentó modificaciones hechas por Converse (1949) para adecuarse a la descripción del poder de atracción de centros comerciales dentro de una ciudad. Estas adaptaciones comprendieron:

En primer lugar, se incorporó el tamaño de la población como un factor determinante de atracción, y la distancia recorrida se consideró como el componente que limita la propensión a realizar viajes de compra.

En segundo lugar, se observó que los viajes de compras tienden a disminuir a medida que la distancia entre la ubicación del consumidor y la tienda se incrementa. Esta relación entre distancia y frecuencia de viaje sigue una tendencia exponencial.

Finalmente, se añadió un componente adicional: la relación intrínseca entre los gastos de compra realizados por los consumidores y las ventas al por menor generadas por estas transacciones. Esta modificación profundizó en la comprensión de cómo el gasto influye en el dinamismo económico del comercio minorista.

Así, incorporando lo realizado por Reilly y Converse, el modelo de Huff considera a los consumidores como el elemento que influye en las áreas comerciales, donde aquellos que viven en áreas suburbanas, y por tanto, en lugares de venta donde los bienes son limitados, tienden a visitar tiendas de mayor tamaño y ubicadas en la ciudad más cercana. La expresión formal del modelo es:

$$P_{ij} = \frac{\frac{S_j}{T_{ij}^\alpha}}{\sum_{j=1}^n \frac{S_j}{T_{ij}^\alpha}} \quad (2.1)$$

Donde:

P_{ij} representa la probabilidad de que un consumidor de origen i viaje a un centro comercial j .

S_j : tamaño en metros cuadrados útiles del centro comercial j .

T_{ij} : es el tiempo de viaje desde el origen i a un destino j .

α : parámetro estimado empíricamente por categoría de artículo para reflejar el efecto del tiempo de viaje.

Con este modelo, es posible estimar el número de consumidores con la siguiente fórmula:

$$E_{ij} = P_{ij} \cdot C_i \quad (2.2)$$

Donde:

E_{ij} : número esperado de consumidores de cada punto de origen i que compran en un centro comercial particular j .

P_{ij} : probabilidad estimada de que un consumidor en i seleccione j para comprar.

C_i : número total de consumidores.

El modelo también permite estimar las ganancias por ventas anuales en base a presupuesto de las familias. La expresión formal es:

$$A_{ij} = P_{ij} \cdot C_i \cdot B_{ik} \quad (2.3)$$

Donde:

A_{ij} : ventas potenciales del centro comercial particular j para cada origen i .

B_{ik} : presupuesto del hogar de origen i en el artículo k .

En general, el modelo de Huff es una herramienta útil para el análisis de ubicación, estimación de demanda y ventas en el contexto de los centros comerciales. Dentro de sus ventajas se puede mencionar que es relativamente simple de entender y aplicar al estar en función del tamaño y el tiempo de viaje, reflejando además la importancia de la ubicación y la accesibilidad en la decisión de compra. Aunque el modelo tiene una base teórica, el valor del parámetro α se estima empíricamente a partir de datos de elección real, lo que lo hace adaptable a diversos contextos.

Sin embargo, es importante considerar sus limitaciones como que asume que la decisión de compra se basa únicamente en el tamaño del centro comercial y el tiempo de viaje, no considerando una variedad de factores económicos, psicológicos y sociales. Además, asume que las elecciones de los consumidores son independientes entre sí, desestimando la influencia en las elecciones de otros y situaciones de compra repetitiva.

2.2.1.2. Modelos MCI

Considerando las limitaciones del modelo de Huff, (Nakanishi & Cooper, 1974) extienden el modelo creando el Modelo de Interacción Competitiva Multiplicativa (MCI), el cual posibilita la incorporación de otras características de la infraestructura comercial. El modelo

tiene la siguiente expresión:

$$P_{ij} = \frac{\prod_{k=1}^K A_{kij}^{\beta_k} \xi_{ij}}{\sum_{j=1}^{m_i} \prod_{k=1}^K A_{kij}^{\beta_k} \xi_{ij}} \quad (2.4)$$

Siendo:

- P_{ij} : probabilidad de que el cliente que vive en el sitio i compre en el centro comercial j .
- A_{kij} : variable k -ésima del centro comercial j en una situación de elección i .
- m_i : número total de centros comerciales j disponibles a elección para clientes de origen i .
- K : número total de atributos de centros comerciales.
- β_k : parámetro que representa la sensibilidad del cliente ante la variable ($k = 1, \dots, K$).
- ξ_{ij} : término de error de especificación.

Una de las ventajas de este modelo es que puede ser calibrado mediante el método de Mínimos Cuadrados Ordinarios (OLS) y que los parámetros pueden ser estimados mediante la siguiente transformación log-centering:

$$\ln \left(\frac{p_{ij}}{\hat{p}_i} \right) = \sum_{k=1}^K \beta_k \ln \left(\frac{A_{kij}}{\hat{A}_{ki}} \right) + \ln \left(\frac{\xi_{kij}}{\hat{\xi}_{ki}} \right) \quad (2.5)$$

Donde:

Media geométrica de p_{ij} , A_{ij} y ξ_{ij} :

$$\hat{p}_i = \left(\prod_{j=1}^m p_{ij} \right)^{\frac{1}{m_i}}, \quad \hat{A}_{ki} = \left(\prod_{j=1}^m A_{kij} \right)^{\frac{1}{m_i}} \quad y \quad \hat{\xi}_{ij} = \left(\prod_{j=1}^m \xi_{ij} \right)^{\frac{1}{m_i}} \quad (2.6)$$

Los modelos MCI se han utilizado ampliamente para modelar la elección de tiendas y selección de locación. En (Albornoz, 2007) se modeló el comportamiento de clientes usando modelos MCI cuando se escoge una tienda supermercado, donde se muestra que para los consumidores la distancia a los supermercados es un factor más relevante que el tamaño de su superficie.

Como ventaja de este modelo, se menciona su consideración de la competencia, esto lo hace especialmente adecuado para situaciones en las que las decisiones de los consumidores están influenciadas por la presencia y las características propias de las otras opciones. Además, a diferencia de otros modelos gravitacionales basados en la superficie y distancia, el modelo MCI puede incorporar más variables del centro comercial que aporten al entendimiento del comportamiento de clientes como número de tiendas, variedad de productos, etc. Como limitación, está que no es adecuado para analizar comportamientos de elección individuales, por lo que requiere datos agregados, sin embargo, es la clase de datos disponibles.

2.2.2. Modelos Logit

Los modelos Logit de elección discreta son continuadores del modelo de Huff y están fundados en la Teoría de la Elección Discreta. Según (Chasco, 1997) este tipo de modelo es considerado parte de los más apropiados para expresar situaciones de elección. Existen varios modelos logit como el Logit Multinomial (MNL), Logit Anidado, Logit de Destino Competitivos y Mother Logit.

El Modelo Logit Multinomial (MNL), expuesto por McFadden (1974), para los modelos de elección de tienda, estima la probabilidad de que un individuo elija una alternativa en función de su utilidad observable, expresada como:

$$p_{ij} = \frac{\exp(v_{ij})}{\sum_{k=1}^m \exp(v_{ik})} \quad (2.7)$$

Donde v_{ij} representa utilidad observable de que un cliente de origen i visite el destino j , en función de factores de la tienda y características propias del cliente.

Mientras que el Modelo Logit Anidado, presentado también por McFadden (1977), es una extensión del modelo MNL que intenta solucionar los problemas mencionados anteriormente, incorporando como supuesto que los consumidores procesan la información de forma jerárquica y que existe una pertenencia de las alternativas a un cluster.

El Modelo de Logit de Destinos Competitivos, propuesto por (Fotheringham, A. S. & O'Kelly, M. E., 1989), aborda los problemas de los modelos Logit Multinomial y Anidado desde una perspectiva espacial. Añadiendo que la elección espacial es el resultado de una estrategia jerárquica de información-procesamiento en la que inicia seleccionando un cluster de alternativas y que la composición de los clusters espaciales percibidos por los consumidores, debido a su naturaleza continua, es usualmente desconocida, por lo que pondera la utilidad de cada alternativa por la probabilidad de que sea evaluada por un individuo. La dificultad de este modelo es definir la probabilidad de que una alternativa se encuentre en determinado cluster.

Finalmente, el Mother Logit es una generalización de los modelos logit multinomiales convencionales (McFadden, D., Tye, W., & Train, K., 1977). En este modelo, la utilidad de las alternativas de elección depende no solo de sus atributos, sino también incluye los atributos de otras alternativas en el conjunto de opciones. Esto se logra al incluir constantes adicionales y efectos de atributos en la función de utilidad.

En (Timmermans et al., 1991) implementan modelos de MNL y Mother Logit en la elección de 3 centros comerciales, mostrando un leve, pero significativamente mejor resultado en Mother Logit. Los resultados mostraron un impacto positivo en la apertura de tiendas anclas, estacionamientos y superficie, mientras que existe un impacto negativo en el aumento del costo del estacionamiento. El modelo Mother Logit logra proporcionar a los gerentes información más detallada sobre la estructura competitiva de los centros comerciales.

Cada modelo presenta ventajas y limitaciones en función de las características del problema

de elección y selección de ubicación. El Modelo Logit de Destinos Competitivos generaliza los otros dos modelos anteriores y tiene en cuenta la relación espacial entre las alternativas, pero requiere definir la probabilidad de que una alternativa pertenezca a un cluster, lo cual puede ser subjetivo y propenso a errores.

2.2.3. Modelos de Regresión

Los modelos de regresión permiten evaluar la relación entre una variable dependiente de interés respecto a otras variables independientes (Peláez, 2006). Existen varios tipos de modelos de regresión, tales como regresión lineal múltiple, regresión logística, entre otras.

Este tipo de modelo, son usados como método para explicar el comportamiento de clientes respecto a visitas de tiendas de retail (Das, 2014; Finn y Louviere, 1996; Carpenter y Moore, 2006), compra de artículos como papel, café y pasta de dientes (Aaker, D. A. & Jones, J. M., 1971), compras de bebidas alcohólicas transfronterizas (Asplund et al., 2007), entre otras. En la mayoría de los resultados muestran que los atributos del destino es relevante para la visita. Para tiendas de *retail*, en (Das, 2014), mediante análisis de regresión múltiple, muestran que todas las dimensiones de identidad de la tienda como estilo, variedad de producto, nombre de la tienda, publicidad y actitudes gerenciales tienen un impacto positivo en el consumidor.

Además de la distancia como variable relevante, en (Asplund et al., 2007) evidencian al precio como motivo de viaje, donde utilizando datos del monopolio minorista de bebidas alcohólicas del gobierno sueco estiman la sensibilidad de las ventas a los precios extranjeros y su dependencia de la distancia hacia la frontera, considerando el alto impuesto a este tipo de bebidas en el país de origen. Como resultado, se tiene que la sensibilidad de las ventas regionales con respecto a los precios extranjeros depende de la distancia a la frontera, especialmente en comunidades cercanas a la frontera danesa.

Utilizar regresiones para estimar la demanda de clientes en el contexto minorista ofrece ventajas como su interpretación directa y significativa de los resultados. Los coeficientes de la regresión proporcionan información clara sobre cómo cada variable impacta en la demanda y su relevancia en el caso. Sin embargo, es importante tener en cuenta las desventajas y limitaciones asociadas con el uso de regresiones en este contexto. Por un lado, estas técnicas están supeditadas a ciertas suposiciones críticas, como la linealidad y la independencia de errores. Además, las regresiones son sensibles a la presencia de valores atípicos en los datos. Finalmente, el problema más común con esta técnica es la multicolinealidad, que ocurre cuando el modelo tiene muchas variables independientes y algunas miden el mismo fenómeno.

2.2.4. Modelos de Machine Learning

Los modelos de Machine Learning aplicados a la industria minorista han demostrado ser efectivos para modelar patrones no lineales e irregulares. Entre estos modelos, las máquinas de vectores de soporte (SVM) y el bosque aleatorio (RF) son ampliamente utilizados (Jordan & Mitchell, 2015).

2.2.4.1. Modelos Random Forest

El modelo de bosque aleatorio, es un modelo de aprendizaje de conjunto que utiliza como algoritmo base árboles de decisión que buscan maximizar la diferencia entre grupos y minimizar la diferencia dentro de los grupos (Breiman, 2001). Esta técnica es muy versátil y se puede aplicar a diferentes tipos de problemas, como la clasificación y la regresión. Tiene como ventaja que asegura la aleatoriedad del proceso al crear diferentes modelos no correlacionados, disminuyendo el riesgo de sobreajuste y tiene la capacidad para lograr entregar variables con mayor influencia en predicciones.

Estos modelos pueden ser utilizados para predecir la atracción de los centros comerciales en función de variables ambientales, lo que resulta útil para estimar la popularidad de nuevos centros comerciales entre los residentes antes de su diseño. Además, debido a la naturaleza espacialmente heterogénea de estas variables, la relación entre ellas y la atracción no suele ser espacialmente estacionaria.

En el reciente estudio de (Huang et al., 2023) se modela la atracción de demanda a áreas comerciales utilizando datos de movilidad, comparando efectos de predicción entre modelos de regresión lineal, máquina de vectores de soporte y random forest añadiendo variables de destino, características de cobertura de suelo, densidad de población y transporte; obteniendo como resultado que la regresión lineal supera con diferencia a la máquina de vectores de soporte, mientras que con el método de aprendizaje de conjunto clásico, random forest obtiene mejores resultados de predicción.

Este modelo es criticado principalmente por el tiempo de entrenamiento, al construir y combinar múltiples árboles de decisión. A pesar de su capacidad para abordar el sobreajuste y brindar importancias de características, la interpretación de estos valores puede ser compleja, lo que dificulta la comprensión del impacto individual de las características.

La posibilidad de utilizar Random Forest en el proyecto, se justifica por su capacidad para manejar eficientemente grandes conjuntos de datos, evitando el sobreajuste y proporcionando resultados más precisos, su robustez ante valores atípicos y datos faltantes asegura la integridad de los resultados. La capacidad de identificar características clave y la posibilidad de paralelización en el entrenamiento también hacen que este tipo de modelo sea una elección estratégica para obtener información valiosa y procesar eficientemente la abundante información de los datos.

En resumen, existen varios modelos que describen la elección y selección de tienda en el rubro del retail que han sido abordados por la literatura. Un modelo pionero en la industria, es el modelo de Huff que presenta un enfoque clásico de elección de tienda que explica la probabilidad de que los clientes elijan un establecimiento en función de su atractivo y distancia. Se ha utilizado como complemento a otros modelos para explicar la selección de lugar de compra; como el Modelo MCI que permite agregar múltiples atributos del punto de destino y se puede agregar dinamismo al modelo de Huff. En (Liang et al., 2020) se desarrolla un modelo de Huff dinámico consciente del tiempo (T-Huff) para el análisis de patrones de visitas a tiendas de dos cadenas de supermercado y una de *retail*, usando datos de movilidad de 10 ciudades. Como resultado, se muestra que la distancia juega un papel negativo significativo tanto para una cadena de supermercado como para la de *retail*, pero no para la cadena de

supermercado restante. Se encuentra además diversidad de resultado en las ciudades, clientes de grandes ciudades con un sistema de tránsito bien desarrollado muestran menos sensibilidad a las visitas de larga distancia. Además, factores socioeconómicos como ingreso familiar promedio impulsan las visitas, mientras que la diversidad racial y étnica (medida de entropía) tiene una influencia positiva significativa para una tienda por departamento y una cadena supermercado y la edad media de las personas no tiene un papel importante, excepto en las tiendas de la cadena de *retail*, donde todos los factores son significativos.

En la literatura de elección de tienda, el Modelo Logit es muy utilizado. Estima la probabilidad de que un individuo elija una alternativa en función de su utilidad observable, puede ser utilizado para modelar la demanda y las preferencias de los clientes. Sin embargo, es difícil implementarlos en elecciones espaciales. Los modelos de Regresión permiten analizar la relación entre una variable dependiente (como la elección de tiendas) y variables independientes (como características de los puntos de venta y de los clientes). Los modelos Random Forest son métodos de aprendizaje automático que combinan múltiples árboles de decisión para realizar predicciones más precisas. Los Modelos de Localización-Asignación involucran la selección de ubicaciones y la asignación de demandas a esas ubicaciones para optimizar algún criterio específico.

Considerando el propósito de investigación, las ventajas y limitaciones de los modelos, desempeño en la literatura, simplicidad de aplicación e interpretación, sumado al tipo de datos disponibles (masivos, agregados y estáticos), se decide replicar el Modelo de Huff, Modelo Multiplicativo de Interacción Competitiva (MCI), Regresión y Random Forest. Estos modelos ofrecen enfoques complementarios para comprender y predecir la elección de tiendas en función de múltiples variables explicativas y se pueden adaptar a diferentes escenarios y conjuntos de datos para obtener resultados significativos.

Se espera aportar en la literatura realizando distintos modelos probabilísticos que expliquen los viaje de compras, realizando una comparación entre resultados de los modelos mencionados, distinguir beneficios de añadir variables explicativas de origen, destino y transporte, evidenciar virtudes de ocupar datos masivos de movilidad y, finalmente, exponer aplicaciones del modelo que guíen decisiones claves del negocio.

2.3. Minería de datos (CRISP-DM)

Las tres metodologías dominantes para el proceso de la minería de datos son: KDD, CRISP-DM y SEMMA.

En cuanto al nivel de detalle en las tareas de cada fase, los modelos KDD y SEMMA proponen pasos generales del proyecto de minería de datos, sin especificar puntualmente las tareas que deben llevarse a cabo en cada una de sus fases. En cambio, en CRISP-DM sí especifica con mayor detalle las actividades del proceso, por lo que podría ser considerado como una metodología, por el nivel de detalle con el que describen las tareas en cada fase según (Moine et al.,s.f.).

Además CRISP-DM, al tener un enfoque iterativo, como se muestra en la siguiente figura, permite realizar múltiples ciclos de desarrollo y refinamiento a lo largo del proyecto, lo cual

es útil para la mejora continua de los modelos y poder realizar adaptaciones mientras se adquiere conocimiento.

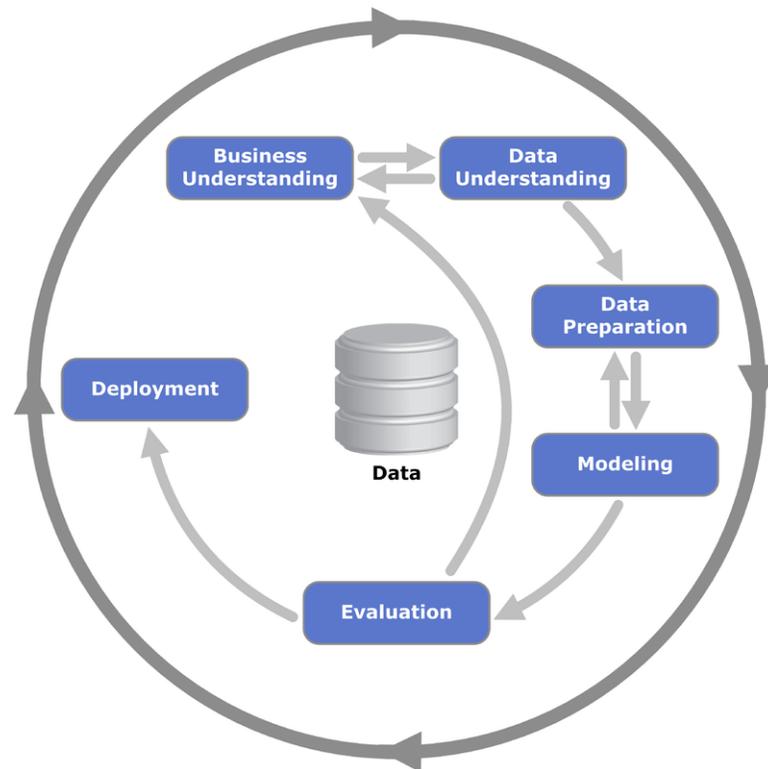


Figura 2.2: CRISP-DM. Fuente: Health Data Miner

Entonces, para realizar la minería de datos de este trabajo, se implementará la metodología (CRISP-DM, 2000), la cual es la guía de referencia más utilizada para el desarrollo de proyectos de Data Mining según (Gallardo, 2009). Esta herramienta contiene seis pasos:

1. Comprensión del negocio o problema: la primera fase busca lograr la comprensión de los objetivos y requisitos del proyecto, con el fin de obtener los objetivos técnicos. Se debe determinar el problema y entenderlo de forma completa. Posterior a esto, es necesario realizar una evaluación de la situación actual.

2. Comprensión de los datos: comprende la recolección inicial de los datos, descripción de los datos, su exploración y verificación de la calidad de los datos, para asegurar completitud y corrección de las bases de datos.

3. Preparación de datos: se seleccionan los datos a utilizar en base a la técnica de modelado, se realiza una limpieza, estructuración, integración y formateo de datos.

4. Modelado: se seleccionan las herramientas de modelado más adecuadas para el proyecto, considerando la disposición de datos, cumplimiento de requisitos del problema y tiempo de trabajo. Además, se elige el método de evaluación de los modelos elaborados, elaborando un plan de prueba de validez y calidad. Se desarrolla el o los modelos, los cuales se justifican y evalúan.

5. Evaluación: interpretación de los modelos en base a los resultados obtenidos, se revisa para identificar oportunidades de mejora. Si el modelo es válido, se procede a la explotación del modelo.

6. Implementación: se recomiendan acciones dentro de la empresa dado a lo observado según los resultados del modelo. Se asegura el mantenimiento de la aplicación y difusión de resultados. Se realiza un informe final y revisión del proyecto.

Para el caso de CRISP-DM es clave entender que es un proceso iterativo o cíclico en donde volver pasos atrás antes de avanzar es parte de la metodología.

2.4. Metodología

Para probar la hipótesis de investigación se propone la siguiente metodología inspirado en el marco conceptual de CRISP-DM. La metodología propuesta para este trabajo contiene 6 etapas, descritas a continuación.

1. Estudio de la situación actual.

Análisis de la literatura de modelos de retail, específicamente modelos de elección de tienda y selección de ubicación que estimen demanda, abarcando la primera etapa de la metodología CRISP-DM, es decir, comprensión del negocio.

2. Procesamiento y limpieza de datos.

Consolidación y creación de bases de datos necesarias para poder ejecutar y optimizar los modelos que se deseen realizar. Esto corresponde a las etapas de comprensión y procesamiento de datos del CRISP-DM.

3. Selección de atributos a estudiar.

Se escogen las características de personas, centros comerciales y otros factores que serán las variables independientes del modelo. Además, se decidirá si se implementarán niveles de agregación. Los atributos relevantes del caso, se extraerán de (Turhan, Akalın y Zehir, 2013) y (Pope, Lane y Stein, 2012).

4. Elección y creación de modelos a realizar.

Se elige cuáles son los modelos capaces de estimar y caracterizar la demanda en base a objetivos del caso y variables disponibles y posteriormente se desarrollan los modelos. Esta etapa corresponde al de modelado del CRISP-DM. Se replicarán modelos expuestos en la literatura como el modelo de Huff y modelos de interacción competitiva multiplicativa (MCI). Además, se desarrollarán modelos gravitacionales, modelos de regresión lineal y Machine Learning.

5. Evaluación de modelos.

Se evalúa el desempeño de los modelos realizados en cuanto a métricas de error preestablecidas. La evaluación corresponde a la penúltima etapa del CRISP-DM.

Capítulo 3

Desarrollo metodológico

En este capítulo, se presenta el desarrollo metodológico utilizado para llevar a cabo el estudio y alcanzar los objetivos planteados. La comprensión del negocio, fue abordada anteriormente en antecedentes generales y marco conceptual. Luego, se presenta el detalle de la recolección, procesamiento y limpieza de datos. Posteriormente, se describen los datos obtenidos y la selección de atributos a utilizar en los modelos escogidos, para luego analizar sus resultados.

3.1. Procesamiento y limpieza de datos

En esta etapa, se detalla el proceso de recopilación inicial y exploración de los datos consolidados. Se optó por emplear datos de movilidad para registrar las visitas a los centros comerciales, así como información sobre la infraestructura comercial, transporte y demografía. Esto permitió caracterizar la diversidad de ofertas, la accesibilidad al transporte y la composición demográfica de la población, proporcionando una visión completa del entorno urbano y su influencia en las visitas comerciales.

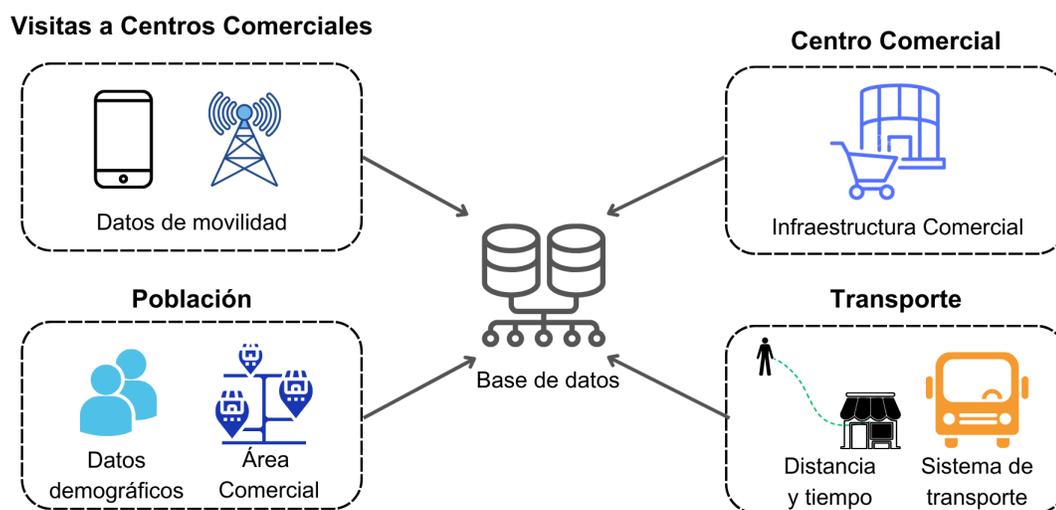


Figura 3.1: Datos utilizados. Elaboración propia.

3.1.1. Datos de movilidad

Los datos de movilidad utilizados en este estudio fueron proporcionados por una compañía telefónica con una participación significativa en el mercado (32%). Estos registros documentan los desplazamientos desde los hogares, representados por las manzanas, hacia los centros comerciales de la Región Metropolitana. La base de datos comprende información sobre las visitas a centros comerciales en 1.800 zonas, abarcando un total de 19.000 manzanas en la RM. El conjunto de datos proporciona detalles sobre la cantidad de visitas a cada centro comercial, visitas totales y la probabilidad asociada a cada manzana para asistir a cada centro comercial durante octubre de 2021.

En total, se registraron ocho millones de visitas. Entre las comunas con mayor flujo hacia las infraestructuras comerciales se encuentran Puente Alto, Las Condes, La Florida y Maipú, las cuales representan el 32% del total de visitas. Este patrón es coherente, dado que estas comunas son las más pobladas de la capital. La base de datos se compone de los siguientes elementos:

- ID de la manzana (ID_MANZENT): Número de identificación de la manzana o entidad.
- ID Point of Interest (POI_ID): Identificación del centro comercial.
- Nombre Point of Interest (POI_name): Nombre del centro comercial.
- Visitas totales (TOTAL_VISITS): Total de visitas realizadas a todos los centros comerciales por la manzana.
- Visitas únicas (UNIQUE_VISITS): Visitas realizadas al centro comercial por la manzana.
- Probabilidad de visita (PROBABILITY): Probabilidad de que una manzana visite un centro comercial específico.

La probabilidad de visita se calcula dividiendo las visitas únicas realizadas por la manzana a un centro comercial en específico por el total de visitas realizadas. La distribución de estas probabilidades se visualiza en la figura A.1, mostrando en general valores bajos para cada centro comercial. Esto puede explicarse por la amplia distribución de visitas entre 37 puntos. Se observa una mediana de probabilidad de visita más alta para el Mall Plaza Vespucio y una menor para el Espacio Urbano La Dehesa.

Se excluyeron las filas con probabilidad igual a 1, ya que esto indicaría que en esa manzana solo se realizaron visitas a un centro comercial. Al eliminarlos, se busca evitar posibles distorsiones en el análisis, ya que no reflejarían la diversidad de preferencias y comportamientos de los consumidores, las principales estadísticas están presentadas en la tabla 3.1. Finalmente, se incorporó a la base la información de las manzanas que no visitaron los centros comerciales, es decir, aquellas con cero visitas y, por lo tanto, cero probabilidad. Este proceso transformó la base desde 202 mil hasta 706 mil observaciones.

Tabla 3.1: Stats Datos de movilidad

Variable	Media	Mediana	Desviación Estándar	Mínimo	Máximo
PROBABILITY	0,081	0,035	0,112	7,28 e-05	0,892
UNIQUE_VISITS	39,491	6,569	175,574	1,826	12.586,871
TOTAL_VISITS	1.206,668	268,874	2.615,111	3,653	45.088,326
ID_MANZENT	13.153.468.294.786	13.119.181.001.006	102.194.192.035,158	13.101.011.001.005	13.605.031.007.021
POI_ID	20,275	21,000	9,569	1	37

3.1.2. Datos de infraestructura comercial

Se consolida información obtenida de forma manual investigando los sitios web de los centros comerciales y memorias anuales de empresas como Mall Plaza, Parque Arauco y Cencosud. Como factores del complejo comercial, se muestran características de su infraestructura. Los número de tiendas de las instalaciones es variada en cantidad, existen centros comerciales que superan las 300 tiendas como el Parque Arauco (411), Mall Plaza Vespucio (360), Mall Plaza Oeste (324), Costanera Center (312) y el Mall Plaza Egaña (307) y, a su vez, existen centros comerciales más pequeños con menos de 30 tiendas como el Mall Espacio Urbano Las Rejas (21) y Open Plaza Puente Alto (13), como se muestra en la figura A.2 del anexo.

Además, se incorporó las coordenadas de los centros comerciales obtenidas de Google Maps, identificando los 37 puntos de interés los cuales como se observa en la tabla A.3 y en la figura 3.2 casi la mitad de centros comerciales, se encuentran en el sector oriente, el más acomodado de la ciudad.

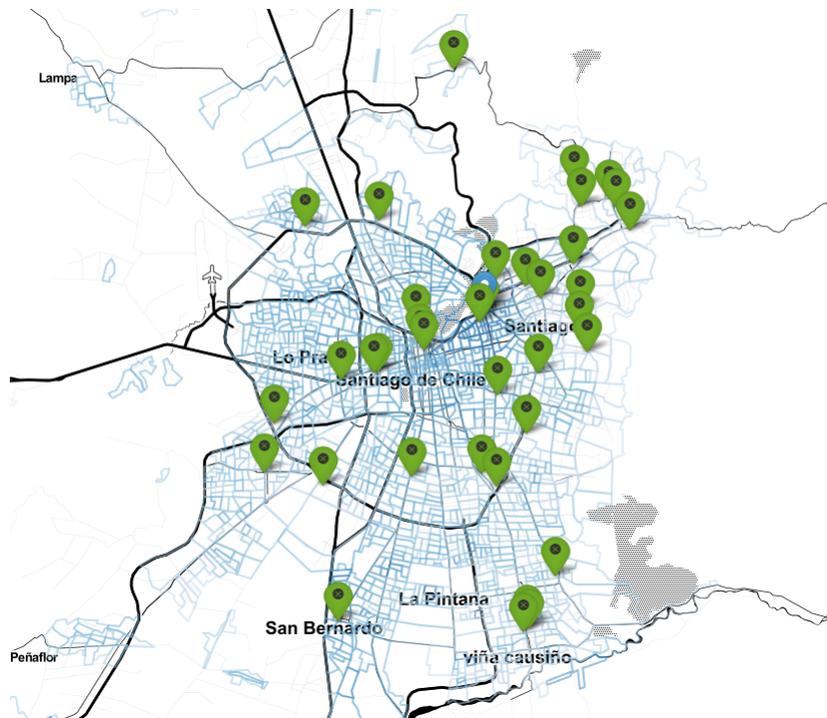


Figura 3.2: Distribución espacial centros comerciales RM

Todas las variables fueron incorporadas a la base mediante el ID del centro comercial.

La correlación entre los atributos obtenidos, expuestos en la tabla A.3 del anexo, indica que existe una alta correlación número de tiendas anclas y número de grandes tiendas, alcanzando 0.9; de la misma magnitud para número de tiendas anclas y grandes tiendas. Con la variable de interés (probabilidad), las correlaciones son bajas, pero positivas.

3.1.3. Datos de población

Se recopilan datos del CENSO realizado en 2017, disponibles en las descargas de su sitio web. Se utiliza la base de datos de manzanas/entidades para incorporar variables como el total de personas por sexo y edad, así como el total de viviendas y su materialidad, entre otros aspectos.

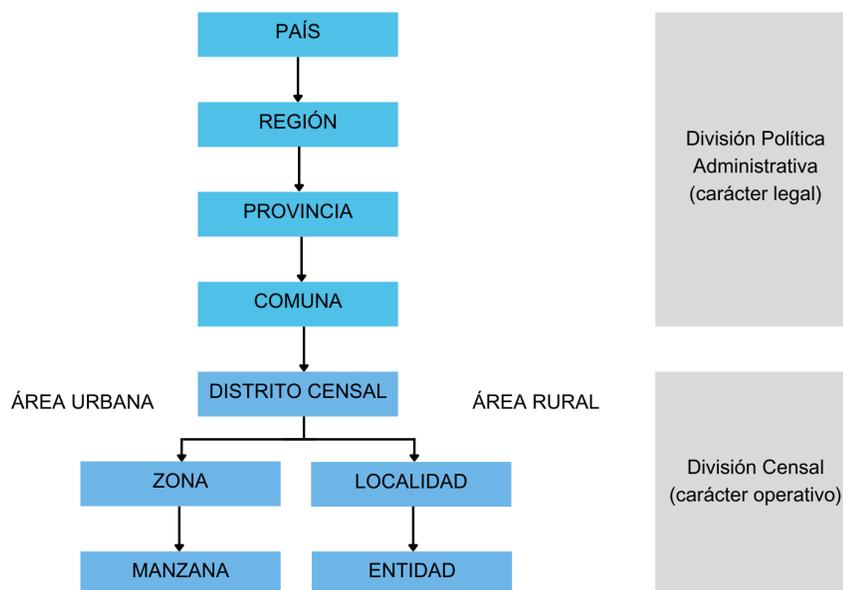


Figura 3.3: División del territorio. Elaboración propia.

La división geográfica censal se deriva del territorio comunal, dividiéndose en distritos urbanos o rurales. A su vez, en el área urbana se reconocen zonas censales y en el área rural, localidades. Para simplificar el análisis y facilitar la comprensión, se utilizarán exclusivamente términos de zonas y manzanas. La base se filtra para incluir solo datos de la Región Metropolitana. También se emplean datos geospaciales del CENSO para obtener coordenadas de manzanas y zonas censales.

Dado que no existen variables que reflejen el nivel socioeconómico de las manzanas, se agregan variables socioeconómicas como el Indicador Socio Material Territorial, elaborado por el IDE Observatorio de Ciudades UC a partir de los datos del CENSO, y la variable Grupo Socioeconómico (GSE). En Chile, los grupos socioeconómicos se clasifican en siete categorías según Acuña (2021): AB (clase alta), C1a (clase media acomodada), C1b (clase media emergente), C2 (clase media típica), C3 (clase media baja), D (clase media vulnerable) y E (pobres). La mayoría de las manzanas estudiadas en la Región Metropolitana pertenecen al nivel socioeconómico D, seguido por C3 y C2.

Además, se obtienen datos sobre el entorno residencial de las personas, específicamente en relación a las tiendas y servicios disponibles. Estas variables se capturan a través de `openstreetmapz` la librería `osmdata`, registrando la cantidad de supermercados, farmacias, centros de belleza y restaurantes cercanos al centroide de cada zona censal (aproximadamente a 1.5-2 km de radio) en la Región Metropolitana. Todas las variables demográficas se incorporan a la base mediante el ID de la manzana.

En cuanto a las correlaciones entre variables, como se muestra en la figura A.4, se observa una alta correlación entre el número de farmacias de la zona y el número de bancos, lo cual podría ser propio de un área atractivo de tiendas y servicios. Por otro lado, existe una correlación negativa perfecta entre hombres y mujeres debido a la proporción de género en cada manzana. En variables demográficas como el número de personas, el índice de hogar ISMT y el número de inmigrantes, no hay problemas de correlación. En relación a la variable de interés, en general, la correlación con las otras variables es nula o cercana a cero negativo.

3.1.4. Datos de sistema de transporte

Se recopilan datos relacionados con el sistema de transporte de la ciudad, así como variables que registran los tiempos y distancias entre la población y los centros comerciales. Para obtener estos datos, primero se determinan las coordenadas geográficas de las manzanas, los centros comerciales y las estaciones de metro cercanas a ambos. Luego, se calculan las distancias utilizando la librería `geosphere`, considerando las distancias entre centros comerciales y estaciones de metro, entre diferentes estaciones y entre estaciones de metro y las manzanas. Esto permite obtener variables que representan estas distancias totales, además de implementar una metodología para estimar los tiempos de viaje. En esta estimación se considera que, dependiendo de las distancias involucradas, una persona puede desplazarse caminando, tomar un bus o combinar bus y metro para llegar al centro comercial.

Sin embargo, durante la validación de los tiempos de viaje con datos reales de Google Maps, se detectaron discrepancias significativas en algunas manzanas. Por lo tanto, se tomó la decisión de recurrir a la API de Google Maps, específicamente la Distance Matrix API, para obtener datos precisos acerca de los tiempos y distancias de viaje en transporte público y auto. Esta medida garantiza una mayor exactitud en la información relacionada con la movilidad de la población. Dada la amplitud de la solicitud de datos, se optó por obtener la información desde la zona censal hasta el centro comercial, considerada lo suficientemente robusta para el caso.

Además, se incorporó atributos vinculados a la disponibilidad de buses cercanos al centro comercial utilizando la aplicación Moovit. Estos indicios revelan qué tan accesible es llegar en transporte público. Los recorridos de buses que transitan cerca de los centros comerciales van desde el cero hasta nueve, siendo el Espacio Urbano Laguna Piedra Roja el único que no tiene un paradero cercano y además, junto con el Mall Paseo San Bernardo, son los destinos con mayor distancia al metro más cercano, superando ambos los 4,5 km. Por el contrario, Mall Pumay, Vivo Imperio, Arauco Maipú, Alto Las Condes, Paseo Los Trapenses tienen nueve recorridos y el Mall Arauco Estación, Espacio Urbano Gran Avenida, Apumanque y Vivo Centro están ubicados a menos de 100 metros de una estación.

Los atributos fueron incorporados a la data mediante Id de la manzana y del centro comercial. En la tabla A.8, se muestra que existe una correlación entre tiempo y distancia con el mismo medio de transporte y una correlación negativa con la probabilidad con la mayoría de atributos de transporte incorporados.

3.2. Selección de atributos a estudiar

En (Turhan et al., 2013) se presentan resultados de una revisión de la literatura que busca comprender los criterios de selección de la ubicación de la tienda en el contexto minorista. Los criterios que se han relacionado con la ubicación son características de tienda, estructura de la población, rendimiento y factores económicos. Entre las características de la tienda se incluyen elementos como la accesibilidad a través de vías y estacionamientos, así como los aspectos económicos como los costos de construcción, alquiler, compra del espacio y renovación física de la tienda. En relación con las características demográficas, aspectos como los ingresos, edad, género, tamaño y densidad de población, crecimiento demográfico, estado civil, tiempo de viaje, preferencias políticas, clases sociales y características culturales desempeñan un papel crucial. El rendimiento de la tienda, medido por el volumen de ventas, margen de ganancias, cuota de mercado, patrocinio minorista y elasticidad de precio, también afecta la elección de ubicación. Además, factores económicos como los ingresos del hogar, distribución de ingresos, movilidad en términos de posesión de automóviles por hogar, disposición a gastar en la tienda, fuentes de ingresos, alquileres y el tipo y precio de la vivienda en el área circundante tienen un impacto en la decisión de ubicación.

En concordancia con estos aspectos, en un estudio realizado por (Pope et al., 2012) se mencionan atributos relevantes del modelo tales como metros cuadrados arrendables, tiendas anclas, tipo de centro comercial, ventas del centro comercial por pie cuadrado, ubicación de la tienda en el centro comercial, actitud de los comerciantes en el centro comercial, perspectivas para las tiendas agrupadas, importancia del desarrollador, población comercial, otros centros comerciales en el área, competencia dentro del centro comercial, área de competencia fuera del centro comercial, ingreso promedio en el área de comercio y motivo de defensa.

Luego de conocer las variables relevantes para el caso expuestas por (Turhan et al., 2013) y (Pope et al., 2012), se seleccionaron aquellas que están disponibles en la base y que tuvieran completitud e importancia. De las 99 variables disponibles originalmente, se escogieron 29 y se clasificaron en 3 categorías: destino, origen y transporte. Esta categorización en tres grupos distintos, resulta de suma importancia para el análisis y comprensión efectiva de las variables. Además, se logra un enfoque más estructurado y facilita la identificación de patrones significativos, permitiendo una interpretación más clara y profunda de los datos, ya que se destacan las relaciones y tendencias específicas en función de la ubicación, el destino final y el medio de transporte utilizado. En el Anexo B se encuentran con mayor detalle todas las variables utilizadas .

3.2.1. Atributos de Destino

Corresponde a las variables propias de la infraestructura y oferta ofrecida de centros comerciales, las cuales están presentes en la tabla B.1 del anexo.

Las variables son:

1. SUPERFICIE_M2: superficie del centro comercial medida en metros cuadrados.
2. NUM_TIENDAS: cantidad de tiendas que tiene el centro comercial considerando tiendas de retail, entretención, comida y servicios.
3. NUM_TIENDAS_ANCLAS: cantidad de tiendas capaces de atraer gran cantidad de público como supermercados, grandes tiendas, cine, tiendas de construcción.
4. SUPERMERCADO: cantidad de supermercados dentro del mall tales como Lider, Santa Isabel, Jumbo y Tottus.
5. GRANDES.TIENDAS: cantidad de grandes tiendas dentro del mall como Falabella, Paris y Ripley.
6. CONSTRUCCIÓN: cantidad de tiendas de construcción grandes dentro del mall como Easy, Ikea y Sodimac.
7. CINE: cantidad de cines en el centro comercial como Cineplanet, Cine Hoyts (Cinépolis) y Cinemark.

La variable SUPERFICIE_M2 proporciona información sobre el tamaño y la amplitud del centro comercial, lo que puede ser un factor determinante en términos de la cantidad de tiendas, diversidad de ofertas y la comodidad para los visitantes para transitar en las instalaciones. De la misma forma, NUM_TIENDAS y NUM_TIENDAS_ANCLAS ayudan a comprender la variedad y el tipo de tiendas presentes en el centro comercial. En particular, las tiendas anclas son capaces de atraer un gran flujo de público, y la cantidad total de tiendas en general influyen en la oferta comercial y, por ende, en la atracción de visitantes. Las variables SUPERMERCADO, GRANDES.TIENDAS, CONSTRUCCIÓN y CINE son indicadores específicos de la diversidad de servicios y entretenimiento ofrecidos en el centro comercial. La presencia de supermercados, grandes tiendas, tiendas de construcción y cines puede influir en las preferencias de los consumidores, ya que estos factores añaden valor a la experiencia de compra y recreación.

En conjunto, estas variables proporcionan una representación más completa y detallada de la composición del centro comercial. Esto es especialmente importante al analizar la influencia de estas características en la toma de decisiones de los clientes en cuanto a la elección de visitar un centro comercial en particular, lo que tiene implicaciones significativas para la planificación y el diseño de estrategias comerciales efectivas, ya que estas variables pueden ser controladas por la administración del centro comercial al momento de construcción del lugar o en discusiones de ampliación-remodelación.

3.2.2. Atributos de Origen

Corresponde a variables provenientes de las características de las manzanas divididas en características demográficas y socioeconómicas de las personas y atractivo comercial de la zona, las cuales se encuentran detalladas en la tabla B.2 del anexo.

Las variables son:

1. PERSONAS: corresponde al total de personas efectivamente censadas en la manzana.
2. MUJERES: corresponde al total de mujeres efectivamente censadas.
3. HOMBRES: corresponde al total de hombres efectivamente censadas.
4. INDICE_HOGAR_ismt: Indicador Socio Material Territorial, calculado a partir de variables censales.
5. GSE_N: Grupo Socioeconómico
6. AVE_GSE: Grupo Socioeconómico Agregado.
7. N_FARMACIA_ZONA: cantidad total de farmacias en el área comercial de origen (zona).
8. N_SUPER_ZONA: cantidad de supermercados de grande cadenas en el área comercial de origen (zona).
9. N_FARMACIA: cantidad de grandes farmacias en el área comercial de origen (zona).
10. N_SUPER: cantidad de tiendas de supermercados de todos los tamaños del área comercial de origen.
11. N_BELLEZA: cantidad de salones de Belleza en zona.
12. N_RESTAURANTES: cantidad de restaurantes en el área comercial de origen.
13. NOM_COMUNA: nombre de la comuna de la manzana de origen.
14. NOM_ZONA: zona de la manzana de origen.

VARIABLES COMO PERSONAS, MUJERES, HOMBRES, INDICE_HOGAR_ismt, GSE_N y AVE_GSE capturan aspectos socioeconómicos y demográficos esenciales de la población, los cuales podrían ser valiosos para planes estratégicos y campañas de marketing. A su vez, indicadores como N_FARMACIA_ZONA, N_SUPER_ZONA, N_FARMACIA, N_SUPER, N_BELLEZA y N_RESTAURANTES revelan la oferta comercial y de servicios en el área de origen de los clientes que podrían capturar la retención efectuada sobre ellos.

La inclusión de estas variables en el análisis es esencial para comprender de manera integral las influencias demográficas, socioeconómicas y comerciales propias del cliente o del lugar donde viven que afectan la dinámica de los centros comerciales.

3.2.3. Atributos de Transporte

Corresponde a características relacionadas a la composición del sistema de transporte como metro, bus, distancias y tiempos de viaje, disponibles con mayor detalle en la tabla B.3 del anexo.

Las variables son:

1. METRO_CERCANO_MANZANA: nombre del metro más cercano a la manzana.
2. METRO_CERCANO_MALL: nombre del metro más cercano al centro comercial.
3. DIST_MALL_METRO: distancia entre el metro más cercano al centro comercial y centro comercial.
4. TIEMPO_T.PUBLICO: tiempo de viaje en minutos desde la zona hasta el centro comercial usando transporte público.
5. DISTANCIA_T.PUBLICO: distancia en kilómetros recorrida desde la zona del cliente hasta el centro comercial en transporte público.
6. TIEMPO_AUTO: tiempo de viaje en minutos desde la zona hasta el centro comercial usando vehículo particular.
7. DISTANCIA_AUTO: distancia en kilómetros recorrida desde la zona del cliente hasta el

centro comercial en vehículo particular.

8. NUM_RECORRIDOS_BUS: cantidad de recorridos de buses de los paraderos más cercanos al centro comercial.

Las variables METRO_CERCANO_MANZANA y METRO_CERCANO_MALL identifican la proximidad tanto de las manzanas como del centro comercial al sistema de metro, esto refleja la importancia del transporte público en la elección del modo de desplazamiento. Por su parte, las mediciones de distancias y tiempos de viaje en distintos modos, como transporte público (TIEMPO_T.PUBLICO, DISTANCIA_T.PUBLICO) y vehículo particular (TIEMPO_AUTO, DISTANCIA_AUTO), revelan cómo la proximidad afecta las preferencias de viaje de los clientes. Esto es crucial para evaluar no solo la distancia geográfica, sino también la conveniencia y eficiencia de los distintos medios de transporte. En particular, tiempo y distancia son variables ampliamente utilizadas en la literatura investigada. La variable NUM_RECORRIDOS_BUS agrega un componente adicional, al proporcionar información sobre la disponibilidad y frecuencia de servicios de autobuses en las cercanías. Esto puede influir significativamente en la elección de los clientes que dependen del transporte público.

La inclusión de estas variables son vitales para la realización del trabajo, específicamente en el análisis de las dinámicas de movilidad y accesibilidad que modelan el comportamiento del consumidor en un contexto de transporte como el de Santiago, compuesto principalmente por metro, bus y auto.

3.3. Elección y creación de modelos a realizar

Se optó por implementar una variedad de modelos, incluyendo la replicación de los modelos de Huff, el desarrollo de los modelos MCI, Regresión y Random Forest. Para ello, fue esencial realizar diversas combinaciones de las variables disponibles. Esto se debió a las notables correlaciones y dependencias que existen entre ciertas variables, como por ejemplo, la suma de CONSTRUCCIÓN, CINE, SUPERMERCADO y GRANDES.TIENDAS, es equivalente al total de NUM_TIENDAS_ANCLAS. Además, se observó una marcada correlación entre las distancias y los tiempos utilizando un mismo medio de transporte, así como entre tiendas y servicios del área comercial como NUM_FARMACIAS_ZONA y N_FARMACIAS. Las diferentes combinaciones fueron evaluadas a través de grupos visualizados en la figura 3.4. A continuación, se presentan las versiones con mejores resultados de los modelos representados por el primer grupo dentro de los atributos, mientras que el resto se detallan en el Anexo C.

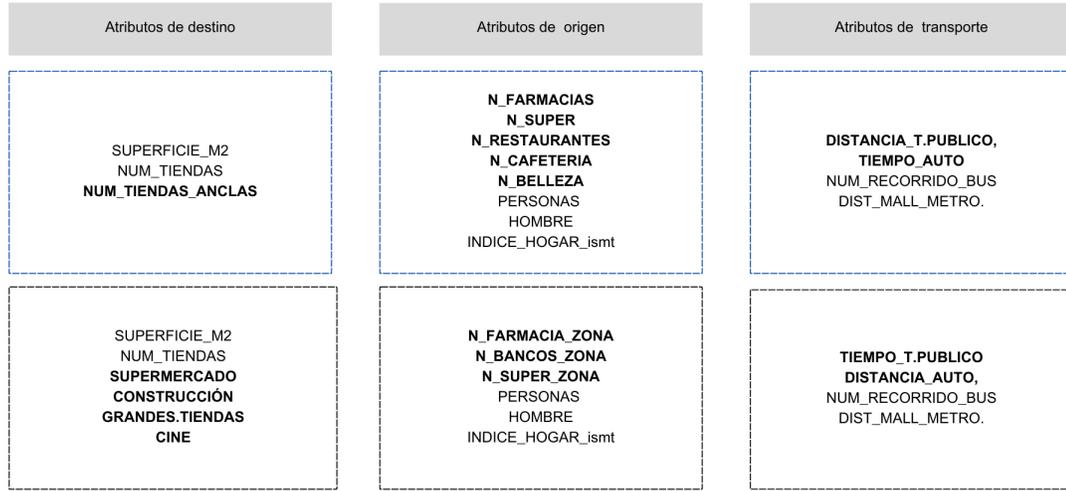


Figura 3.4: Grupo de variables por atributos

3.3.1. Desarrollo Modelos Gravitacionales

Se replica el modelo de Huff utilizando tiempo de viaje en transporte público, el mejor de las variables de tiempo disponible en la sección C.2.

Modelo 1.1:

$$P_{ij} = \frac{\frac{S_j}{T_{ij}^\alpha}}{\sum_{j=1}^n \frac{S_j}{T_{ij}^\alpha}} \quad (3.1)$$

Para explicar la probabilidad de que cliente de origen i viaje a destino j (P_{ij}), se escoge SUPERFICIE_M2 (S_j) como medida de tamaño del destino y TIEMPO_T.PUBLICO (T_{ij}) como tiempo de viaje desde origen i hasta el centro comercial j .

Modelo 1.2:

Para la implementación de los modelos MCI, se requiere aplicar ciertas transformaciones. Dado que la media geométrica opera como denominador en este contexto, se enfrenta a una indeterminación cuando se encuentra con valores de cero. Este escenario se presenta en centros comerciales que carecen de tiendas anclas, así como en otras variables que cumplen con esta condición. Se introduce, entonces, la media geométrica transformada para tiendas anclas y otras variables con esta particularidad (ver figura A.5), dada por la fórmula:

$$\hat{A}_{ki} = \left(\prod_{j=1}^m A_{kij} + 1 \right)^{\frac{1}{m_i}} \quad (3.2)$$

Por otro lado, al detectar varias manzanas con una probabilidad nula de visitar ciertos centros comerciales, se opta por reemplazar este valor por la mitad de la probabilidad

mínima registrada, correspondiente al valor de 7.284×10^{-5} . Posterior a estas modificaciones, se desarrolla el primer modelo MCI usando atributos de destino como superficie del centro comercial medida en metros cuadrados (SUPERFICIE_M2), número de tiendas del centro comercial j (NUM_TIENDA) y número de tiendas anclas del centro comercial j (NUM_TIENDAS_ANCLAS).

$$\ln \left(\frac{p_{ij}}{\hat{p}_i} \right) = \sum_{k=1}^3 \beta_k \ln \left(\frac{A_{kij}}{\hat{A}_{ki}} \right) + \ln \left(\frac{\xi_{kij}}{\hat{\xi}_{ki}} \right) \quad (3.3)$$

Donde:

$$A_{ij} = (\text{SUPERFICIE_M2}, \text{NUM_TIENDAS} \text{ y } 1 + \text{NUM_TIENDAS_ANCLAS})$$

Modelo 1.3:

Para el modelo MCI, se plantea una transformación que permite incorporar las variables demográficas, que son características de las personas y no de los centros comerciales, en la ecuación. Esta transformación consiste en la multiplicación de una variable demográfica por una variable asociada al centro comercial, lo que posibilita que las características demográficas tengan un impacto en la elección del centro comercial. En esencia, esta modificación busca integrar de manera efectiva los factores demográficos en el modelo MCI, permitiendo que influyan en la decisión de los clientes al elegir un centro comercial específico. Si la transformación no se lleva a cabo, los resultados se ven alterados como se muestra en la tabla C.4 del anexo.

Con las modificaciones mencionadas, es posible incorporar al modelo anterior atributos de origen C_{in} como características demográficas y atractivo comercial. Las variables demográficas consideradas fueron el Indicador Socio Material Territorial (INDICE_HOGAR_ismt), total de personas censadas en la manzana (PERSONAS), proporción de hombres efectivamente censados (HOMBRES); mientras que del atractivo comercial de origen se considera cantidad de cafeterías (N_CAFETERIA), número de restaurantes (N_RESTAURANTES), supermercados (N_SUPER), farmacias (N_FARMACIAS) y tiendas que brinden servicio de belleza (N_BELLEZA).

$$\ln \left(\frac{p_{ij}}{\hat{p}_i} \right) = \beta_0 + (\beta_1 + \beta_{2,\dots,n-1} \cdot C_{in}) \ln \left(\frac{A_{ijk}}{\hat{A}_{ki}} \right) + \ln \left(\frac{\xi_{kij}}{\hat{\xi}_{ki}} \right) \quad (3.4)$$

Donde:

$$A_{ij} = (\text{SUPERFICIE_M2}, \text{NUM_TIENDAS} \text{ y } 1 + \text{NUM_TIENDAS_ANCLAS})$$

$$C_{in} = (\text{INDICE_HOGAR_ismt}, 1 + \text{N_CAFETERIA}, \text{N_SUPER}, 1 + \text{N_RESTAURANTES}, 1 + \text{N_FARMACIAS}, 1 + \text{N_BELLEZA}, \text{HOMBRES} \text{ y } \text{PERSONAS})$$

Modelo 1.4:

Considerando los atributos de destino del modelo 1.2 se agregan variables de transporte como tiempo de viaje desde origen i hasta el centro comercial j usando vehículo particular (TIEMPO_AUTO), distancia entre origen i hasta centro comercial j en transporte público (DISTANCIA_T.PUBLICO), distancia entre el centro comercial j y la estación de metro

más cercano (DIST_MALL_METRO) y número de recorridos de buses cercanos del centro comercial (NUM_RECORRIDOS_BUS).

$$\ln\left(\frac{p_{ij}}{\hat{p}_i}\right) = \sum_{k=1}^7 \beta_k \ln\left(\frac{A_{kij}}{\hat{A}_{ki}}\right) + \ln\left(\frac{\xi_{kij}}{\hat{\xi}_{ki}}\right) \quad (3.5)$$

Donde:

A_{ij} = (SUPERFICIE_M2, NUM_TIENDAS, 1+NUM_TIENDAS_ANCLAS, TIEMPO_AUTO, DISTANCIA_T.PUBLICO, DIST_MALL_METRO y 1+NUM_RECORRIDOS_BUS)

Modelo 1.5:

Se consideran atributos de origen, destino y transporte anteriormente mencionados.

$$\ln\left(\frac{p_{ij}}{\hat{p}_i}\right) = \beta_0 + (\beta_1 + \beta_{2,\dots,n-1} \cdot C_{in}) \ln\left(\frac{A_{ijk}}{\hat{A}_{ki}}\right) + \ln\left(\frac{\xi_{kij}}{\hat{\xi}_{ki}}\right) \quad (3.6)$$

Donde:

A_{ij} = (SUPERFICIE_M2, NUM_TIENDAS, 1+NUM_TIENDAS_ANCLAS, TIEMPO_AUTO, DISTANCIA_T.PUBLICO, DIST_MALL_METRO y 1+NUM_RECORRIDOS_BUS)

C_{in} = (INDICE_HOGAR_ismt, 1+N_CAFETERIA, N_SUPER, 1+N_RESTAURANTES, 1+N_FARMACIAS, 1+N_BELLEZA, HOMBRES y PERSONAS)

3.3.2. Desarrollo Modelos de Regresión

Se aplicaron diversos modelos de regresión con el propósito de entender la probabilidad de que un cliente se dirija a un centro comercial. En todos estos modelos, se optó por utilizar el logaritmo de la probabilidad como variable independiente. Esta elección se justifica por el hecho de que la transformación logarítmica puede resultar beneficios en situaciones donde la variable dependiente tiene una ocurrencia extremadamente baja. Esto se debe a que puede suavizar los cambios abruptos en el modelo, proporcionando una representación más estabilizada. Debido a la transformación logarítmica, para aquellas manzanas con una probabilidad nula de visitar ciertos centros comerciales, se reemplaza este valor por la mitad de la probabilidad mínima registrada, es decir, 7.284×10^{-5}

Dado que hay numerosas manzanas que carecen de tiendas de servicio (ver figura A.5), se aplicó una corrección utilizando la expresión 1+N_TIENDA_SERVICIO, según corresponda, siguiendo la misma lógica empleada en los modelos MCI. Asimismo, para mantener el rango de la probabilidad predicha dentro del intervalo (0,1), la variable independiente se ajusta mediante la transformación $\ln\left(\frac{P_{ij}}{(1-P_{ij})}\right)$. Este enfoque asegura que los resultados se mantengan dentro de los límites realistas de probabilidad.

Modelo 2.1:

$$\ln\left(\frac{P_{ij}}{(1-P_{ij})}\right) = \alpha_0 + \alpha_1 \ln(SUPERFICIE_M2_j) + \alpha_2 \ln(NUM_TIENDAS_j) + \alpha_3 \ln(1+NUM_TIENDAS_ANCLAS_j) + \xi_i \quad (3.7)$$

Modelo 2.2:

$$\ln\left(\frac{P_{ij}}{(1-P_{ij})}\right) = \alpha_0 + \alpha_1 \ln(SUPERFICIE_M2_j) + \alpha_2 \ln(NUM_TIENDAS_j) + \alpha_3 \ln(1 + NUM_TIENDAS_ANCLAS_j) + \alpha_4 \ln(1 + N_FARMACIAS_i) + \alpha_5 \ln(1 + N_SUPER_i) + \alpha_6 \ln(1 + N_CAFETERIA_i) + \alpha_7 \ln(1 + N_RESTAURANTES_i) + \alpha_8 \ln(1 + N_BELLEZA_i) + \alpha_9 \ln(PERSONAS_i) + \alpha_{10} \ln(HOMBRES_i) + \alpha_{11} \ln(INDICE_HOGAR_ismt_i) + \xi_{ij} \quad (3.8)$$

Modelo 2.3:

$$\ln\left(\frac{P_{ij}}{(1-P_{ij})}\right) = \alpha_0 + \alpha_1 \ln(SUPERFICIE_M2_j) + \alpha_2 \ln(NUM_TIENDAS_j) + \alpha_3 \ln(1 + NUM_TIENDAS_ANCLAS_j) + \alpha_4 \ln(DISTANCIA_T.PUBLICO_{ij}) + \alpha_5 \ln(TIEMPO_AUTO_{ij}) + \alpha_6 \ln(1 + NUM_RECORRIDOS_BUS_{ij}) + \alpha_7 \ln(DIST_MALL_METRO_{ij}) + \xi_{ij} \quad (3.9)$$

Modelo 2.4:

$$\ln\left(\frac{P_{ij}}{(1-P_{ij})}\right) = \alpha_0 + \alpha_1 \ln(SUPERFICIE_M2_j) + \alpha_2 \ln(NUM_TIENDAS_j) + \alpha_3 \ln(1 + NUM_TIENDAS_ANCLAS_j) + \alpha_4 \ln(1 + N_FARMACIAS_i) + \alpha_5 \ln(1 + N_SUPER_i) + \alpha_6 \ln(1 + N_CAFETERIA_i) + \alpha_7 \ln(1 + N_RESTAURANTES_i) + \alpha_8 \ln(1 + N_BELLEZA_i) + \alpha_9 \ln(PERSONAS_i) + \alpha_{10} \ln(HOMBRES_i) + \alpha_{11} \ln(INDICE_HOGAR_ismt_i) + \alpha_4 \ln(DISTANCIA_T.PUBLICO_{ij}) + \alpha_5 \ln(TIEMPO_AUTO_{ij}) + \alpha_6 \ln(1 + NUM_RECORRIDOS_BUS_{ij}) + \alpha_7 \ln(DIST_MALL_METRO_{ij}) + \xi_{ij} \quad (3.10)$$

3.3.3. Desarrollo de Modelos Random Forest

Se empleó la técnica de Random Forest para modelar la probabilidad de demanda de clientes en centros comerciales. Dada la gran cantidad de datos procesados, se optó por utilizar la librería ranger de Rstudio y se configuró el modelo con un total de 300 árboles para garantizar un análisis robusto. Para el desarrollo de estos modelos, no fue necesario aplicar ninguna transformación a probabilidades ni a variables de tiendas.

Modelo 3.1:

Para este, se consideraron variables de atributos de destino como número de tiendas (NUM_TIENDAS), superficie en metros cuadrados (SUPERFICIE_M2) y número de tiendas anclas (NUM_TIENDAS_ANCLAS).

Modelo 3.2:

Se le añaden a las variables antes mencionadas factores del origen como número total de farmacias en la zona (N_FARMACIAS), totalidad de supermercados (N_SUPER), número de cafeterías (N_CAFETERIA), cantidad de restaurantes (N_RESTAURANTES), número de tiendas de servicios vinculadas a la belleza (N_BELLEZA), número de personas censadas en la manzana de origen (PERSONAS), proporción de hombres (HOMBRES) e índice de hogar socioeconómico (INDICE_HOGAR_ismt).

Modelo 3.4:

Para el siguiente se consideraron las variables de destino anteriormente mencionadas como (NUM_TIENDAS), (SUPERFICIE_M2) y (NUM_TIENDAS_ANCLAS) y variables del sistema de transporte como la distancia entre el centro comercial y su estación de metro más cercana (DIST_MALL_METRO), el número de recorridos de buses que están ubicadas cercanas al centro comercial (NUM_RECORRIDOS_BUS), tiempo de viaje en auto (TIEMPO_AUTO) y distancia recorrida en transporte público (DISTANCIA_T.PUBLICO).

Modelo 3.3:

Para el último modelo, se unieron todas las variables de las categorías destino, origen y transporte anteriormente mencionadas (NUM_TIENDAS, SUPERFICIE_M2, NUM_TIENDAS_ANCLAS, N_FARMACIAS, N_SUPER, N_CAFETERIA, N_RESTAURANTES, N_BELLEZA, DIST_MALL_METRO, NUM_RECORRIDOS_BUS, TIEMPO_AUTO y DISTANCIA_T.PUBLICO).

3.4. Evaluación de modelos

En esta sección se evalúa el desempeño de los modelos realizados en cuanto a las siguientes métricas de error expuestas. A continuación, se exponen los resultados e interpretación de los modelos realizados. Primero se presentan los resultados obtenidos al utilizar datos de movilidad en distintos modelos presentados por categoría de atributos. Luego, se expone una comparación con resultados obtenidos simulando una encuesta.

3.4.1. Métricas de rendimiento de modelos

Se implementarán las siguientes herramientas para la evaluación de modelos, utilizando las métricas de desempeño descritas a continuación:

1. *MAE* (Mean Absolute Error): medida de error que determina la diferencia absoluta entre los valores reales y los pronosticados.

$$MAE = \frac{1}{n} \sum_{j=1}^n |Real_n - Pron_n| \quad (3.11)$$

2. *MAPE* (Mean Absolute Percentage Error): es una medida de error relativo que indica el porcentaje de error promedio en relación a los valores observados del caso.

$$MAPE = \frac{1}{n} \sum_{j=1}^n \left| \frac{Real_n - Pron_n}{Real_n} \right| 100 \quad (3.12)$$

3. *RSME* (Root Mean Square Error): medida comúnmente utilizada para evaluar la precisión de un modelo de regresión o pronóstico.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.13)$$

Donde n es el número total de observaciones o muestras, y_i : representa el valor observado de la variable dependiente y \hat{y}_i representa el valor predicho por el modelo para la variable dependiente.

4. R^2 : mide la proporción de la variabilidad de la variable dependiente que puede ser explicada por la variable independiente en un modelo de regresión, usada normalmente para evaluar bondad de ajuste de modelo y la capacidad explicativa de las variables que lo componen.

5. $CORR$: medida de desempeño normalmente utilizada para evaluar modelos de Random Forest, mide la relación lineal entre las predicciones de un modelo y los valores reales de la variable a predecir en el conjunto de prueba. Un Corr cercano a 1 indica una fuerte correlación positiva entre los valores estimados y reales, esto indica un buen desempeño. Por el contrario, un valor cercano a 0 indica baja correlación y bajo desempeño.

6. $Verosimilitud$: busca encontrar los valores de los parámetros que hacen que los datos observados sean más probables de acuerdo con el modelo.

7.- AIC (Criterio de Información de Akaike): busca un equilibrio entre el ajuste del modelo a los datos y la complejidad del modelo. Cuanto más bajo sea el valor de AIC, mejor se considera el modelo en términos de ajuste y complejidad. Es una herramienta útil en la selección de modelos.

$$AIC = -2\log(\text{máx_verosimilitud}) + 2N\text{parámetros} \quad (3.14)$$

3.4.2. Simulación de encuesta

Con el objetivo de estimar la probabilidad de visita utilizando el modelo de Huff y posteriormente evaluar la idoneidad de los datos móviles, se lleva a cabo una simulación. Para ello, primero, se selecciona el centro comercial Costanera Center como punto focal de la simulación. A continuación, se determina el número de encuestas a realizar. Este cálculo se basa en el número de manzanas y las visitas acumuladas por manzana. La fórmula utilizada para determinar el tamaño de la muestra para cada manzana se rige por la siguiente ecuación:

$$n = \frac{(Z^2 \cdot p \cdot (1 - p))}{E^2} \quad (3.15)$$

Donde:

n : es el tamaño de muestra requerido para la manzana.

Z : es el valor crítico de la distribución normal estándar correspondiente al nivel de confianza deseado.

p : proporción de visitas en la manzana en relación con el total de visitas acumuladas.

E : margen de error especificado.

Considerando un nivel de confianza del 0.95, se obtiene que es necesario realizar 1064 encuestas. Posteriormente, se procede a realizar dos preguntas clave durante la encuesta: ¿Dónde vives? y ¿Qué otros centros comerciales frecuentas?. Estas respuestas proporcionan el punto de origen y permiten la obtención de las distancias pertinentes.

Seguidamente, se replica el Modelo de Huff, tomando en consideración tanto la superficie como el tiempo de viaje en automóvil. Para este propósito, se emplea un parámetro α establecido en -0.847, el cual se obtuvo a través de una encuesta realizada a 93 individuos (Stanley & Sewall, 1976). Finalmente, se procede a evaluar el desempeño del modelo utilizando diversas métricas de precisión.

3.4.3. Resultados Modelos

A continuación, se presentan los resultados obtenidos a través de los modelos desarrollados más significativos. Estos modelos fueron calibrados utilizando el 70 % de los datos y evaluados con el 30 % restante con el conjunto de prueba. Inicialmente, se desglosan los resultados y análisis según la categoría de atributos y el tipo de modelo aplicado. Posteriormente, se identifica el modelo sobresaliente, y finalmente se exponen los hallazgos derivados de la implementación del modelo de Huff, ajustando sus parámetros a través de la simulación de encuestas y comparándolos con el uso de datos de movilidad.

Con el propósito de evaluar el rendimiento de distintos tipos de modelos utilizando diversas categorías de variables, se implementó una clasificación de atributos en función a si pertenece al origen (población), destino (centro comercial) o transporte. A continuación, se presenta el conjunto de variables que arrojó los mejores resultados en cada categoría. Estos atributos fueron utilizados en los modelos MCI, de regresión y random forest.

Tabla 3.2: Combinación de atributos utilizados

Atributos	destino	destino+origen	destino+transporte	destino+origen+transporte
Variables	SUPERFICIE_M2 NUM_TIENDAS NUM_TIENDAS_ANCLAS	SUPERFICIE_M2 NUM_TIENDAS NUM_TIENDAS_ANCLAS N_FARMACIAS N_SUPER N_CAFETERIA N_RESTAURANTES N_BELLEZA PERSONAS HOMBRES INDICE_HOGAR_ismt	SUPERFICIE_M2 NUM_TIENDAS NUM_TIENDAS_ANCLAS DISTANCIA_T.PUBLICICO TIEMPO_AUTO NUM_RECORRIDOS_BUS DIST_MALL_METRO	SUPERFICIE_M2 NUM_TIENDAS NUM_TIENDAS_ANCLAS N_FARMACIAS N_SUPER N_CAFETERIA N_RESTAURANTES N_BELLEZA PERSONAS HOMBRES INDICE_HOGAR_ismt NUM_TIENDAS_ANCLAS DISTANCIA_T.PUBLICICO TIEMPO_AUTO NUM_RECORRIDOS_BUS DIST_MALL_METRO

Para los modelos, se optó por utilizar el Error Absoluto Medio (MAE) y el Coeficiente de Determinación (R-cuadrado) como métricas de evaluación por diversas razones. El MAE ofrece una interpretación directa y sencilla al representar el error promedio entre las pre-

dicciones y los valores reales, lo que facilita su comprensión y comparación entre modelos. Por otro lado, el R-cuadrado proporciona una medida de cuánta variación en la variable de respuesta es explicada por el modelo, lo que lo convierte en un indicador valioso del ajuste del modelo. En la tabla C.19 del anexo se encuentran el resto de indicadores de los modelos como MAPE, AIC, CORR y RSME. Los resultados obtenidos son:

Tabla 3.3: Resumen de resultados

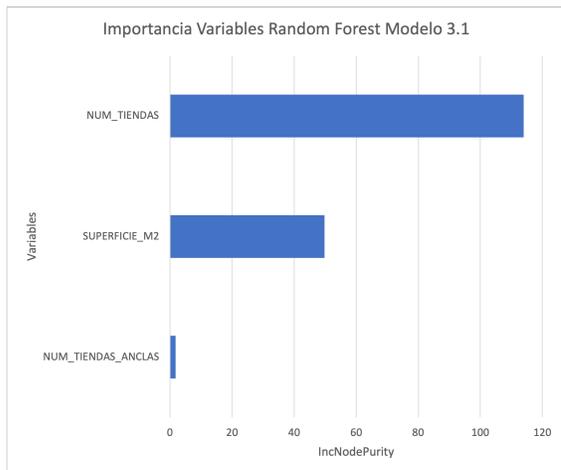
Atributos	Tipo de Modelo	MAE	R2
Destino	MCI (1.3)	0.117	0.094
	RL (2.1)	0.027	0.071
	RF (3.1)	0.037	0.060
Destino+ origen	MCI (1.4)	0.115	0.122
	RL (2.2)	0.032	0.118
	RF (3.2)	0.030	0.404
Destino + transporte	MCI (1.5)	0.108	0.297
	RL (2.3)	0.027	0.212
	RF(3.3)	0.026	0.416
Destino+ origen+ transporte	MCI (1.6)	0.106	0.344
	RL (2.4)	0.030	0.282
	RF (3.4)	0.025	0.510

Los resultados expuestos muestran, con respecto a las categorías de variables implementadas, que los modelos con atributos solo de destino son los que peor rendimiento tienen, sobre todo para modelos MCI. Un modelo que incluye solo variables de infraestructura pueden estimar atracción en base a su tamaño, sin embargo, no considera el contexto de los clientes, por lo que no cumple el propósito de obtener información sobre características espaciales de la demanda.

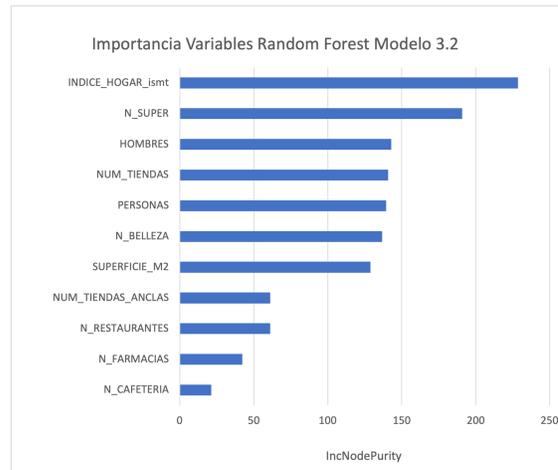
Siguiendo con los modelos con atributos de destino y origen, estos obtienen mejores resultados en todos los tipos de modelos en comparación al discutido anteriormente, a excepción del MAE para los modelos de regresión. Luego, al incorporar variables de transporte a las de destino, aun siendo menor en cantidad (4) en comparación a los de origen (8), estos modelos tienen mejor ajuste. Esto se debe a que la relevancia de distancia y tiempo de viaje es considerablemente superior, evidenciado en los modelos random forest como se muestra en el gráfico c), donde número de tiendas disminuye notoriamente su importancia. Si bien en el gráfico b) se muestra que la importancia de número de tiendas disminuye, no lo hace tan drásticamente como al incorporar atributos de transporte. El resto de atributos de sistema de transporte como número de recorridos disponibles y distancia entre el centro comercial y la estación de metro, no presentan mayor relevancia en el gráfico c). Sin embargo, esto puede deberse a una posible relación con la variable distancia utilizando transporte público, dado

que si hay mejor conectividad de metro y bus, el viaje en transporte público es más directo.

Tabla 3.4: Variables Importantes RF

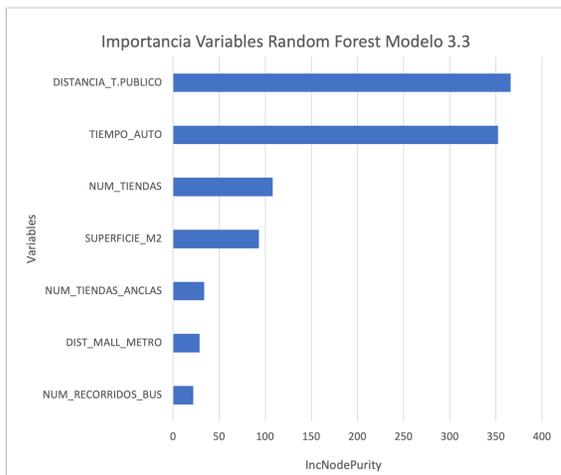


(a) RF Modelo 3.1

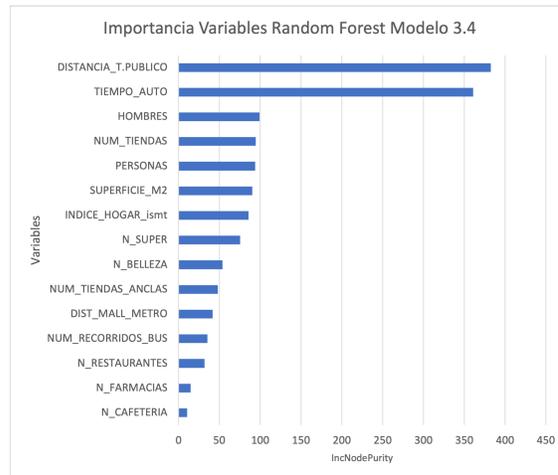


(b) RF Modelo 3.2

Variables Importantes RF



(c) RF Modelo 3.3



(d) RF Modelo 3.4

A medida que se introducen nuevas variables, se observa una mejora en todas las métricas, con la excepción del MAE en los modelos de regresión, el cual experimenta un incremento de 0.03 en comparación con los modelos 2.1 y 2.3. En el caso de los modelos random forest, el modelo 3.4 exhibe un rendimiento superior en todas las métricas en comparación con el modelo 3.1. Es interesante notar que las cinco variables más relevantes para explicar la probabilidad de visita en los modelos random forest provienen de las tres categorías disponibles, incluyendo distancia, tiempo, proporción de hombres, número de tiendas y cantidad de población. Las mejoras de predicción y la importancia mencionada de variables de destino, origen y transporte, evidencian las virtudes de agregar distintos tipos de atributos que componen el contexto de elección de tiendas del cliente.

Con respecto a las variables de origen, en el gráfico d), se observa que las variables menos influyentes son el número de restaurantes, farmacias y cafeterías. Esto sugiere que la presencia o ausencia de estas tiendas en la zona de residencia no ejerce un fuerte impacto en la decisión de dirigirse a un centro comercial. Es posible que estas sean tiendas de productos y servicios que normalmente se utilizan de forma complementaria dentro de los centros comerciales. En contraste, la variable que parece tener un mayor peso es la presencia de supermercados, lo cual puede indicar que es un tipo de tienda que se visita con mayor frecuencia y que influye significativamente en la elección de viajar hacia un centro comercial, de hecho, es considerada como tienda ancla.

Al examinar la distribución de error absoluto en los modelos random forest de la figura 3.5, se confirma que el modelo con mejor predicción es aquel que incorpora todas las categorías de variables. En este sentido, la categoría de transporte demuestra ser ligeramente más relevante en términos de MAE por sobre a la categoría de origen. Sin embargo, esto sucede principalmente al inicio de la distribución y luego los modelos 3.2, 3.3 y 3.4 tienen una distribución casi similar.

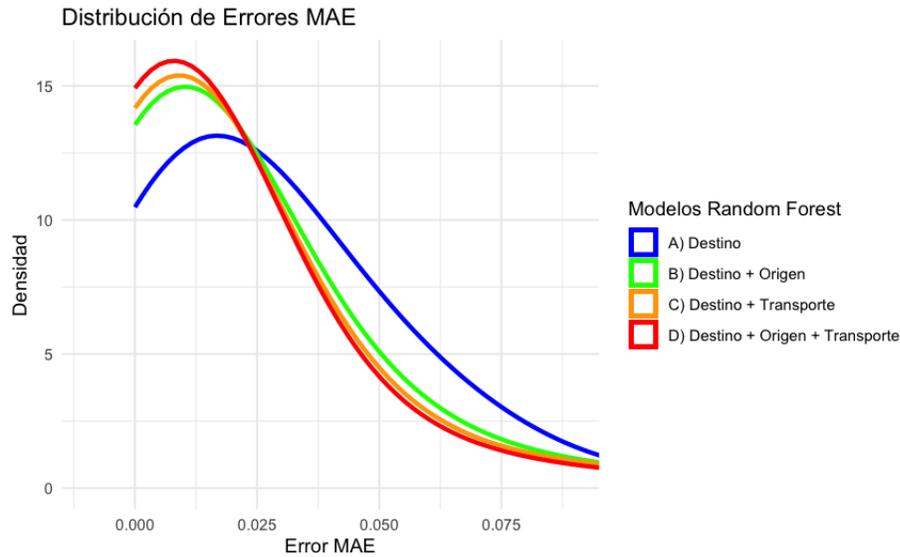


Figura 3.5: Distribución MAE Random Forest

Entre los modelos analizados, se observa que el modelo MCI exhibe un desempeño inferior en términos de la métrica MAE en comparación con los modelos de regresión y random forest. Sin embargo, destaca por presentar un coeficiente de determinación R^2 más elevado que los de regresión y el mejor utilizando solo atributos de destino. Esta diferencia en el rendimiento puede atribuirse a la particular configuración del modelo MCI, donde los atributos se multiplican entre todos los centros comerciales que el cliente considera para tomar la decisión de visita. Al expandir la base de datos para incluir centros comerciales con cero visitas y aplicar la transformación realizada a los valores nulos, siendo modificados a la mitad de la probabilidad más baja registrada, es posible que se haya visto reflejado en los resultados. Es importante tener en cuenta que el modelo MCI no predice una probabilidad de viaje de cero, lo que podría haber contribuido a este fenómeno. Asimismo, al introducir una transformación para incorporar atributos de origen, se añadieron más parámetros al modelo, lo que

potencialmente condujo a un coeficiente de determinación más elevado.

Los modelos de regresión fueron los con segundo mejor desempeño, los cuales al igual que en los modelos MCI, al ser de nivel logarítmico, no acepta probabilidad cero. Las observaciones con probabilidad cero corresponden al 70 % de los datos, por lo que la regresión se ve impactada por esto. Lo anterior se refleja en el gran valor negativo en los interceptos de las versiones de los modelos de regresión disponibles en las tablas C.8, C.9, C.10 y C.11 del anexo.

El modelo que mejor describe la probabilidad de viaje en términos de MAE y R^2 es el modelo 3.4, el cual corresponde a un modelo de random forest con atributos de origen, destino y transporte. Presenta el menor MAE de 0.025, lo cual es un buen resultado para un modelo que mide probabilidad, ya que quiere decir que si el modelo predice una probabilidad de asistencia de esa manzana de un 20 %, esta en realidad puede ser entre un 17.5 % a un 22.5 %. En términos generales, un R^2 más alto indica que el modelo puede explicar mejor la variabilidad en los datos y que las predicciones se ajustan bien a los valores reales, el resultado indica que aproximadamente el 51 % de la variabilidad en la probabilidad de asistencia a un centro comercial es explicada por el modelo, lo que puede ser considerado como aceptable considerando que predice la probabilidad de todas las manzanas de la Región Metropolitana a la mayoría de centros comerciales ubicados en ella.

En cuanto al rendimiento por centro comercial (ver tabla C.14) existe un mejor ajuste en probabilidad para el Costanera Center, Parque Arauco y Mall Plaza Oeste, los cuales ocupan tercer, sexto y segundo lugar en probabilidad media real dispuesta en la figura A.1. Por el contrario, el Espacio Urbano La Dehesa es el destino comercial con peor ajuste de MAPE y es también el con menos probabilidad real, por lo que se ajusta mejor a aquellos centros comerciales con probabilidad mayor. Con respecto al desempeño por comunas, el modelo se ajusta mejor a las comunas de Pirque, Lo Barnechea, Lampa y Macul, las cuales son diversas entre ellas, dado a que están distribuidas en toda la región.

Para poder hacer una comparación entre los datos móviles y datos tradicionales, se simuló una encuesta. Luego, se le aplicó el modelo de Huff, el cual es un modelo emblemático dentro del rubro. Finalmente, se aplicó el mejor modelo desarrollado con el fin de hacer una comparación más completa.

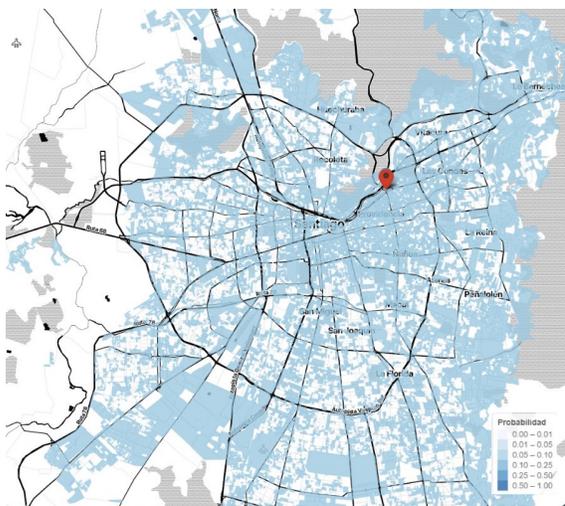
Tabla 3.5: Comparación resultados: Costanera Center

Modelo	Modelo Encuesta	Modelo Huff	Modelo RF 3.4
MAE	0.120	0.071	0.043

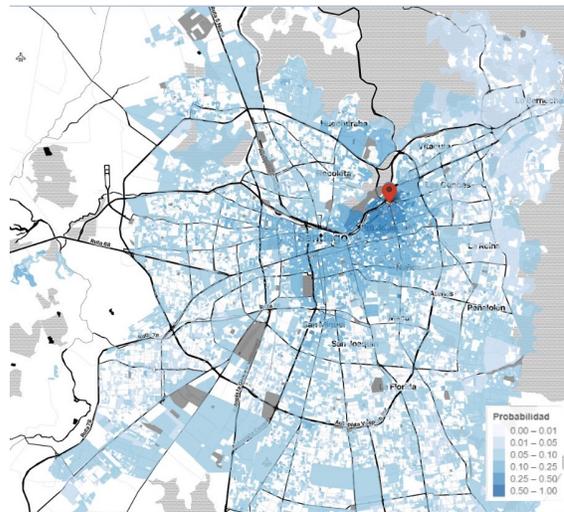
Se puede apreciar en la tabla 3.5 que en el modelo de Huff, utilizando datos de movilidad, se obtiene un mejor resultado de predicción en comparación a datos de encuesta simulada, evidenciando las bondades de utilizar este tipo de datos en este y el resto de modelos. Existe mejor métricas de MAE al aplicar datos móviles en el modelo desarrollado, disminuyendo el MAE en 0.077, lo que significa una mejoría del 64 % . Además, los datos de movilidad son más robustos, dado que al aplicar los distintos modelos, es posible ajustar los parámetros con alrededor de 15 mil observaciones en comparación a las mil encuestas simuladas. Incluso, con los datos móviles es posible obtener información demográfica y socioeconómica de

los clientes, al estar vinculada a la manzana de origen. Esto resulta útil para conocer, por ejemplo, el presupuesto de los clientes y poder aplicarlo para calcular ventas potenciales. La implementación de datos móviles, también permite tener una visualización espacial de las probabilidades de visitas reales y estimadas.

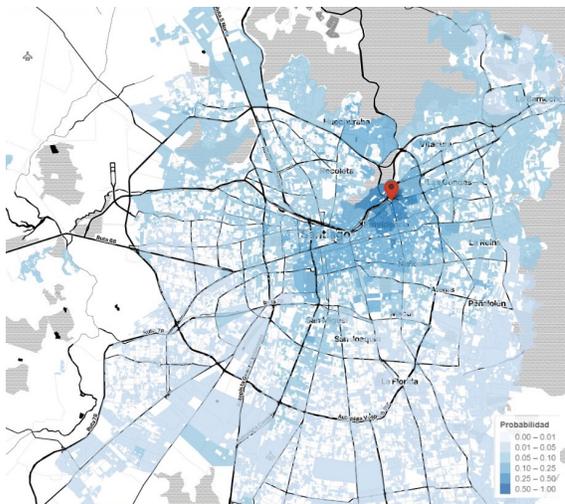
Tabla 3.6: Visualización probabilidad estimada Costanera Center



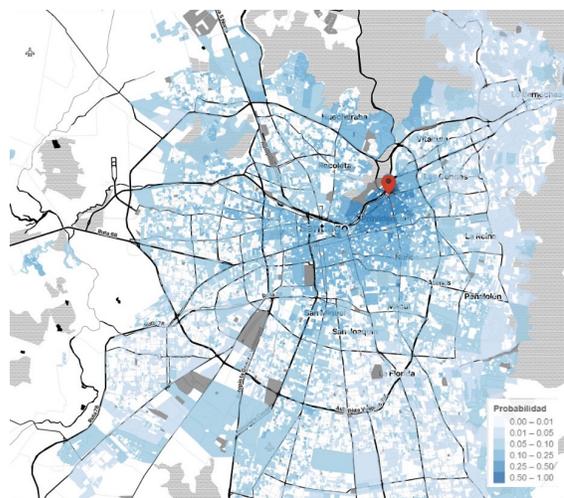
(e) Probabilidad estimada Modelo 3.1



(f) Probabilidad estimada Modelo 3.2



(g) Probabilidad estimada Modelo 3.3



(h) Probabilidad estimada Modelo 3.4

La visualización espacial de la probabilidad permite comprender de mejor forma la dinámica de atracción ejercida, permitiendo distinguir la distribución de manzanas estudiadas, dado que con los datos se conoce el ID, su zona y comuna. El mapa e), al utilizar solo variables de destino, muestra una probabilidad homogénea para toda la RM, lo cual no es útil ni real. En los gráficos f), g) y h) es posible notar como la atracción del centro comercial se hace principalmente a las manzanas más cercanas al centro comercial, lo que quiere decir que atri-

butos de origen pueden capturar esta atracción. A simple vista, el mapa f) con atributos de destino y origen es similar al que incorpora atributos de destino y transporte. Sin embargo, en el g) se forma una esfera más densa cercana al centro comercial, la cual a medida que se aumenta la distancia, disminuye la probabilidad de forma proporcional.

Capítulo 4

Aplicaciones del modelo

En este capítulo se pretende exponer algunas utilidades prácticas que se puede utilizar el modelo realizado. En particular, se aplica el modelo 3.4 para estimar demanda, ventas potenciales y variaciones de demanda al alterar la dinámica de infraestructura comercial y sistema de transporte. Estas simulaciones son fundamentales para informar y respaldar decisiones estratégicas en la gestión de centros comerciales. El modelo proporciona una plataforma sólida para proyectar escenarios, lo que facilita la formulación de estrategias que pueden marcar la diferencia en el éxito y la sostenibilidad de los centros comerciales en un entorno dinámico.

4.1. Estimación de demanda

En esta sección, se despliegan los resultados del Modelo 3.4 para anticipar las visitas realizadas. Dado que si bien es útil conocer las probabilidades de visitas de cada manzana, es importante estimar a cuántas personas se traduce, sobre todo cuando las probabilidades reales y estimadas son bajas y difíciles de interpretar de una manera directa. Se proponen dos versiones, donde se modifica la fórmula de Huff (ecuación 4.3) para estimación de clientes.

Modelo A de estimación de demanda:

$$E_{ij} = \beta_0 + \beta_1 \text{Prob}_{ij} \cdot \text{PERSONAS}_i + e \quad (4.1)$$

La probabilidad estimada se mantiene, pero se adapta la variable de consumidores a *PERSONAS*, en virtud de la disponibilidad de información. Además, se incorpora el parámetro adicional β_0 . Los parámetros son ajustados por medio de un proceso de regresión para que los clientes esperados concuerden con *UNIQUE_VISITS*, que representa el total de visitas efectuadas en el mes al centro comercial. Los resultados obtenidos para la estimación de la demanda revelan que, en el caso de visitas acumuladas, el modelo tiende a sobreestimar la demanda mensual, tal como se ilustra en la figura 4.1.

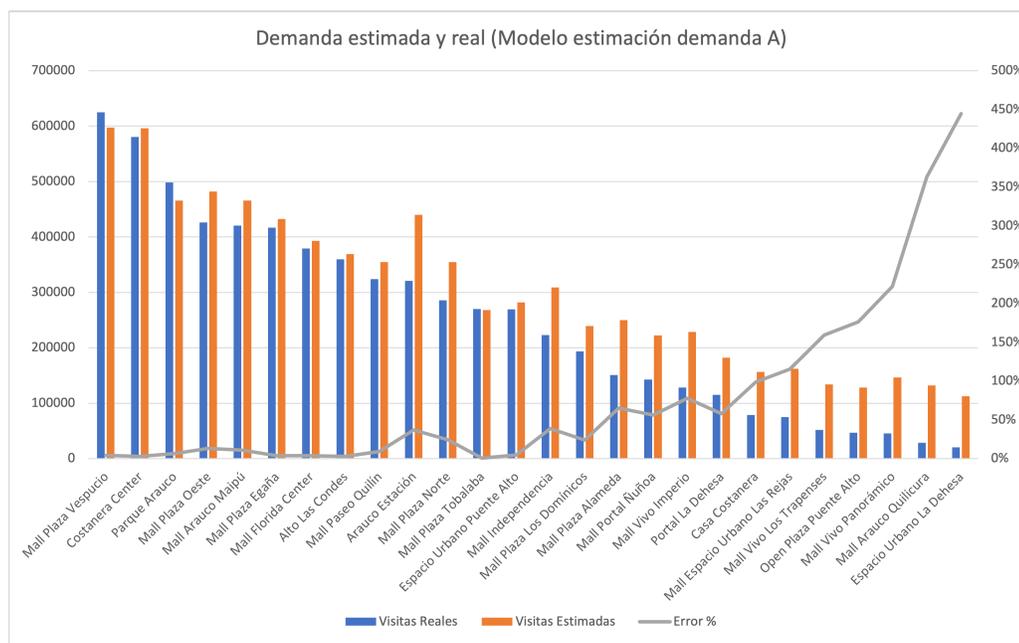


Figura 4.1: Visitas estimadas y reales (Modelo A)

El desempeño en la estimación de la demanda a nivel de manzanas arroja un MAE de prueba general de 14 personas y un MAPE del 677 %, lo cual indica que, en promedio, las predicciones del modelo tienen un error del 677 % en relación con las visitas reales. Esto sugiere que las estimaciones del modelo tienen una discrepancia significativa con los datos observados. Sin embargo, al considerar la estimación a nivel de centro comercial y la demanda total, se evidencia un MAPE del 78 %, como se muestra en detalle en la tabla C.15 del anexo. Esto indica que, aunque el modelo de demanda A no se ajusta de manera precisa a la demanda por manzana a centro comercial, sí presenta un buen nivel de ajuste en su demanda agregada. Esto representa un beneficio destacado de la capacidad del modelo para identificar posiciones y niveles de demanda, lo cual puede ser de gran utilidad para comprender la participación de mercado y la dinámica competitiva en el sector.

Cabe resaltar que el modelo muestra un desempeño notable para aquellos centros comerciales con un alto volumen de visitas, como Mall Plaza Vespucio, Costanera Center y Parque Arauco, con errores porcentuales de solo 4 %, 3 % y 7 %, respectivamente. No obstante, para centros comerciales con un menor número de visitas registradas, como Mall Vivo Panorámico, Mall Arauco Quilicura y Espacio Urbano La Dehesa, el error porcentual es más elevado, alcanzando cifras de 222 %, 363 % y 444 %. Esto se atribuye a la notable cantidad de manzanas que no registraron ninguna visita a estos destinos en la fecha estudiada. Esto se puede deber a la locación alejada de los últimos dos centros comerciales, ubicados en comunas distantes del centro de la ciudad, sumado a su baja cantidad de tiendas y conectividad. En este contexto, incluso si el modelo 3.4 emitió una probabilidad relativamente baja, al multiplicarla por la población correspondiente, esta situación ocasiona un deterioro significativo en sus métricas de desempeño, considerando que $\beta_0 = 6,276$ y $\beta_1 = 1,758$, por lo que centros comerciales con bajas visitas se ven más afectadas.

Modelo B de estimación de demanda:

La segunda versión del modelo de estimación de demanda, incluye un efecto fijo β_j por centro comercial para considerar lo detectado anteriormente, es decir, que para ciertos centros comerciales, la demanda tiene un nivel con base por sus características. La expresión del modelo es:

$$E_{ij} = \beta_0 + \beta_j + \beta_1 \text{Prob}_{ij} \cdot \text{PERSONAS}_i + e \quad (4.2)$$

Se anticipa que para centros comerciales como el Mall Arauco Quilicura y Espacio Urbano La Dehesa, el parámetro β_j será más negativo en comparación con el resto de los destinos comerciales. Esto se debe a que, aunque en el modelo A la demanda fue sobreestimada en la mayoría de los casos, los errores fueron más prominentes en aquellos con menor demanda registrada.

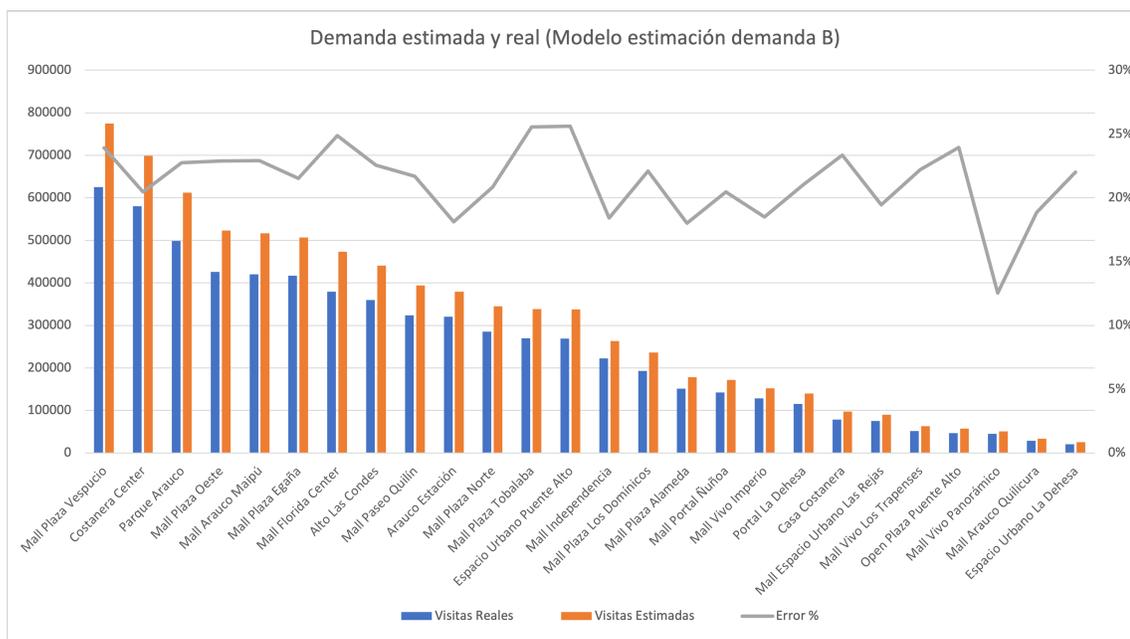


Figura 4.2: Visitas estimadas y reales (Modelo B)

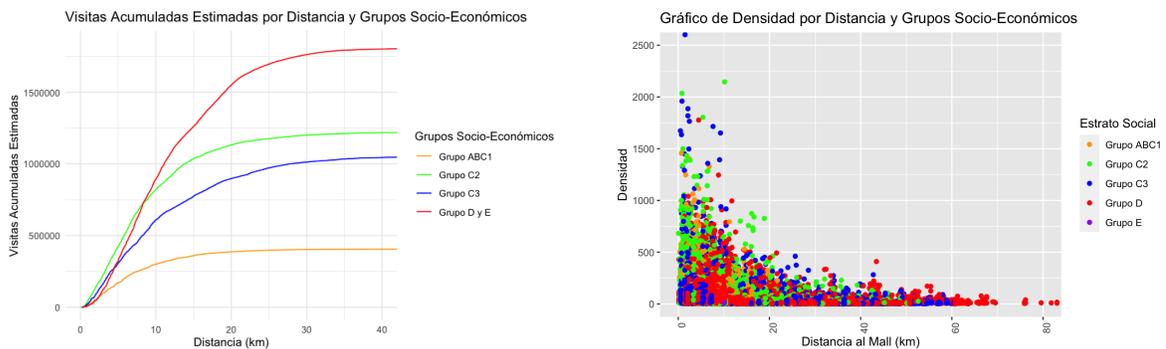
El segundo modelo demuestra una mejora sustancial en la estimación de demanda, particularmente a nivel de agregación de manzanas y destinos comerciales. Como se puede observar en la figura 4.2 y en la tabla C.16 detallada en el anexo, los errores porcentuales muestran una mayor uniformidad, oscilando entre el 13% y el 35%. Como era de esperarse, los parámetros de los centros comerciales menos visitados resultan negativos y significativos, siendo más notorios en Mall Arauco Quilicura, Espacio Urbano La Dehesa y Mall Vivo Panorámico, que se encuentran al final de ambos gráficos.

La introducción del efecto fijo por centro comercial en el modelo permite una estimación de demanda mucho más precisa para todos los centros comerciales, lo que aporta robustez y relevancia al entendimiento del contexto de cada uno. A pesar de que se mantiene la tendencia real de participación de infraestructuras comerciales, en destinos como Mall Plaza Vespucio y Costanera Center, el error porcentual se incrementó, pasando del 4% al 24% en el caso de Mall Plaza Vespucio y del 3% al 20% en el caso de Costanera Center. Es relevante

señalar que ambos coeficientes obtenidos son positivos, aunque el de Costanera Center no alcanza significancia estadística. En comparación, para la demanda acumulada a nivel de infraestructura comercial, el MAPE disminuyó significativamente, pasando de un 78 % a un 21 %. Esto posiciona al modelo B como la mejor alternativa para estimar la demanda con los datos disponibles.

Con información de la demanda, se puede conocer su relación con distintas variables mencionadas con anterioridad. A modo ejemplo, considerando visitas acumuladas por grupo socioeconómicos (GSE) en función de la distancia, se visualiza su comportamiento en la figura i). Donde se puede apreciar que existe mayor cantidad de visitas del grupo D y menor del ABC1, lo que coincide con la composición de la RM. Además, en el grupo D y E, al inicio de la curva, se ve una pendiente menos empinada, lo que indica que se desplazan más para asistir a un centro comercial, posiblemente al tener que viajar una mayor distancia que otros grupos para el destino comercial más cercano. En la figura (j), se confirma que la distancia es un factor relevante, dado que hay una concentración mayor a una menor distancia y que el grupo ABC1 es el menos dispuesto a viajar.

Visitas estimadas por GSE



(i) Visitas acumuladas por GSE

(j) Visitas estimadas por distancia

Asimismo, se puede establecer la relación entre la demanda y la distancia, captando la extensión de influencia de los centros comerciales. Mediante el modelo propuesto, se logra calcular la demanda total y obtener el porcentaje que representa el 70 % de dicha demanda. Esto proporciona una visión de la concentración de la demanda en un radio específico. Con esta información, es posible determinar hasta qué punto los clientes del centro comercial están dispuestos a desplazarse para acceder a él. Además, se puede agregar características como tamaño y ubicación. En particular, se decidió escoger 6 centros comerciales, agrupados según su tamaño en número de tiendas y ubicación para visibilizar relaciones entre variables.

	TAMAÑO GRANDE	TAMAÑO MEDIANO	TAMAÑO PEQUEÑO
SECTOR ORIENTE	 <p>Parque Arauco 20km</p>	 <p>Casa Costanera 16,78km</p>	 <p>Mall Vivo Panorámico 7,04km</p>
SECTOR OTRO	 <p>Mall Plaza Vespucio 14,25km</p>	 <p>Mall Plaza Alameda 13,29 km</p>	 <p>Mall Espacio Urbano Las Rejas 12,57 km</p>

Figura 4.3: Área comercial centros comerciales

Como resultado, se observa que el radio del área comercial guarda relación con el tamaño de las instalaciones; es decir, a mayor tamaño, mayor es la extensión del territorio que abarca, siendo el Parque Arauco con más extensión con 20 km de radio. Además, se puede constatar que existe una mayor disposición a viajar a centros comerciales ubicados en el sector oriente en relación a sus mismos pares de categoría de tamaño. Incluso aunque el Casa Costanera es más pequeño que el Mall Plaza Vespucio, este último tiene un radio menor, lo que sugiere que su clientela está compuesta principalmente por vecinos del centro comercial y personas de áreas cercanas. Por otro lado, las personas de sectores más distantes muestran menos disposición a desplazarse, a pesar de que el Mall Plaza Vespucio es el segundo centro comercial más grande de la región.

4.2. Estimación de ventas potenciales

La estimación de ventas potenciales es una actividad crucial para las organizaciones comerciales, ya que proporciona información de su rendimiento. A diferencia de la simple estimación de la demanda y la probabilidad de visita, conocer las ventas potenciales permite evaluar el verdadero alcance del mercado, capturar las oportunidades de crecimiento y levantar el valor económico que un centro comercial puede captar. Al tener noción de los posibles ingresos, los administradores y planificadores están mejor equipados para tomar decisiones estratégicas vinculadas al rubro. Es por esto que utilizando la ecuación 2.3, se propone una ecuación que estima las ventas potenciales de los destinos comerciales considerando clientes estimados, su grupo social-económico y presupuesto.

Esta estimación se basa en la estructura de gasto de los hogares, presentada por el Instituto Nacional de Estadística (2018) disponibles en la tabla C.20 del anexo. De acuerdo a las categorías, se estima que un 19,65 % podría ser gastado en tiendas y servicios que brinda

el centro comercial, tales como prendas de vestir, calzado, restaurantes, muebles, artículos de hogar, bienes y servicios diversos. De forma adicional, con información de ANDA (2018), se estiman los ingresos mensuales promedio per cápita categorizado por GSE utilizando el ingreso y miembros promedios de los hogares por grupo.

Tabla 4.1: Ingreso de hogar por GSE

GSE	Ingreso promedio del hogar	Miembros del hogar	Ingreso promedio del hogar pp
AB	\$6452000	2,2	\$2932727
C1a	\$2739000	3,2	\$855937
C1b	\$1986000	3,2	\$620625
C2	\$1360000	3,1	\$428710
C3	\$899000	3,2	\$280938
D	\$562000	3,6	\$156111
E	\$324000	4,2	\$77143

Así, se obtiene la información necesaria para crear la variable $PRESUPUESTO_i$ del hogar de origen i , la cual es resultado de la multiplicación entre ingreso per cápita y 19% por. Se utiliza en conjunto la estimación de demanda (E_{ij}) para obtener ventas potenciales de un centro comercial por manzana. Esta es obtenida multiplicando la demanda por el presupuesto per cápita.

$$V_{ij} = DEMANDA_{ij}PRESUPUESTO_i \quad (4.3)$$

Con las ventas potenciales de los centros comerciales para cada manzana, es posible obtener las ventas potenciales totales. Sumando las ventas por centro comercial, se obtiene que el Parque Arauco, Costanera Center y Alto Las Condes son los centros comerciales con mayor ventas estimadas. Sin embargo ¿por qué a pesar de que el Mall Plaza Vespucio es el centro comercial con mayor visitas estimadas (y reales) obtiene menos ventas potenciales? Para responder esta pregunta es necesario estudiar la composición de los clientes de los centros comerciales.

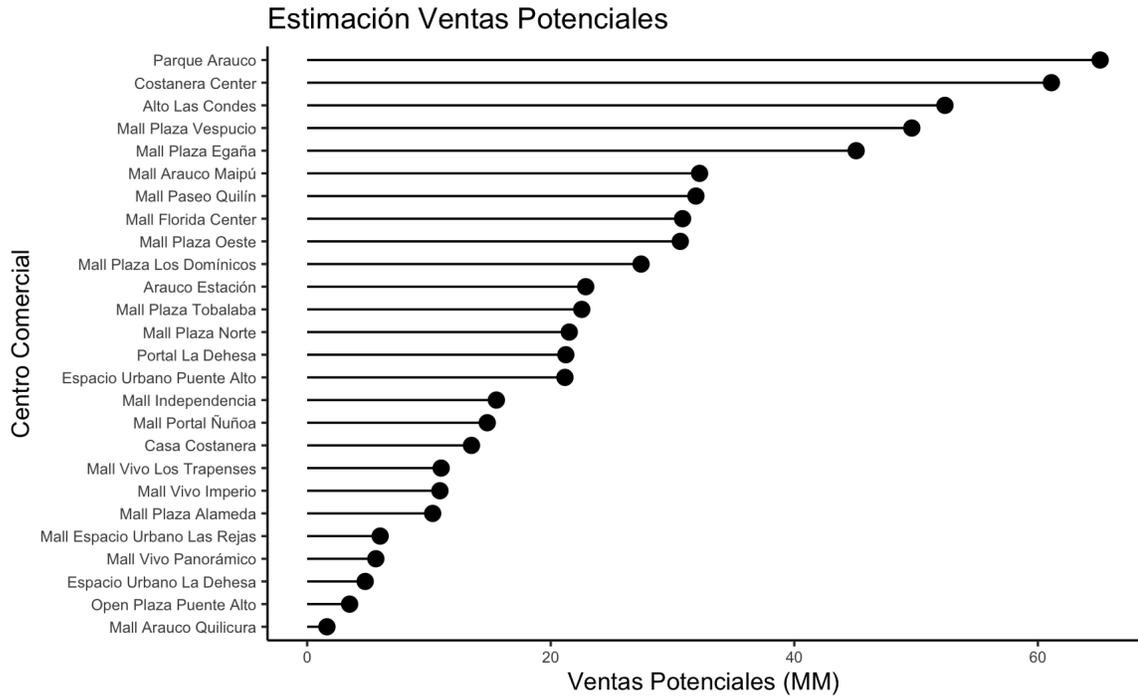
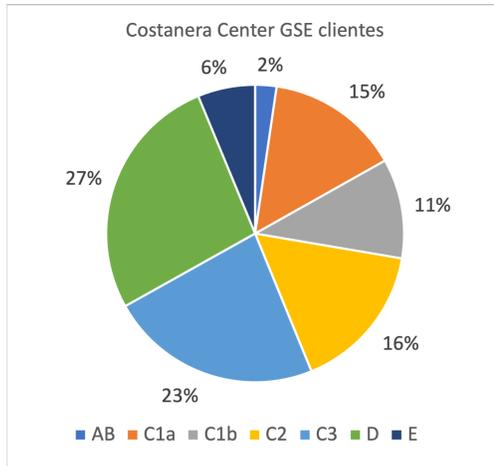


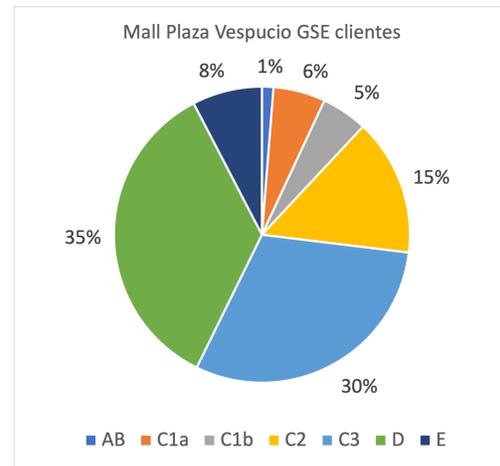
Figura 4.4: Ventas potenciales estimadas

El análisis de las figuras e) y f) revela que el Costanera Center atrae a una mayor cantidad de clientes pertenecientes al segmento ABC1 en comparación con el Mall Plaza Vespucio, donde predominan los grupos D y C3. Esto indica que el presupuesto mensual de los visitantes del Costanera Center tiende a ser superior al de los visitantes del Mall Plaza Vespucio. Es importante recordar que, como se ha demostrado, la distancia juega un rol crucial en la elección de un destino comercial y la ubicación de estos dos centros es considerablemente diferente. El Mall Plaza Vespucio se encuentra en la comuna de La Florida, ubicada en la zona sur de la ciudad, mientras que el Costanera Center está en la comuna de Providencia, en el sector oriente, donde se concentra la población de mayores recursos. Ambos centros comerciales atraen a una gran cantidad de personas debido a su tamaño considerable, su accesibilidad y la densidad de población en sus alrededores. Sin embargo, la composición demográfica de su clientela los diferencia en términos de público objetivo y, por ende, en los resultados de ventas que obtienen.

Composición de clientes GSE



(e) GSE Mall Costanera Center



(f) GSE Mall Plaza Vespucio

4.3. Estimación variación de demanda

A continuación, se presenta la aplicación del modelo para evaluar la dinámica de los centros comerciales en situaciones reales. Para esto, se presentan dos casos que cambian el contexto de competencia. Esto pretende demostrar que el modelo no solo es útil para analizar el estado actual de los centros, sino también para medir y comprender los efectos de variaciones en el entorno comercial. Este enfoque proporciona una valiosa herramienta para anticipar y responder a cambios en la dinámica de los centros comerciales, lo que es fundamental en la toma de decisiones estratégicas.

Caso 1: Cambio infraestructura comercial

Según (Diario Financiero, 2023) Cencosud Shopping anunció para el centro comercial Florida Center una ampliación de 23 mil metros cuadrados con 30 tiendas. El escenario antes y después de la ampliación es la siguiente:

Tabla 4.2: Caso 1: Florida Center

Florida Center	Antes	Después
NUM_TIENDAS	185	215
SUPERFICIE_M2	110000	133000

Caso 2: Cambio infraestructura de transporte:

En febrero del 2022, el presidente de turno anunció el inicio de obras de la próxima Línea 7 del Metro de Santiago, la cual contará con 19 estaciones y unirá las comunas de Renca y Vitacura (Hermosilla, 2022). Las nuevas estaciones son las siguientes:



Figura 4.5: Estaciones línea 7

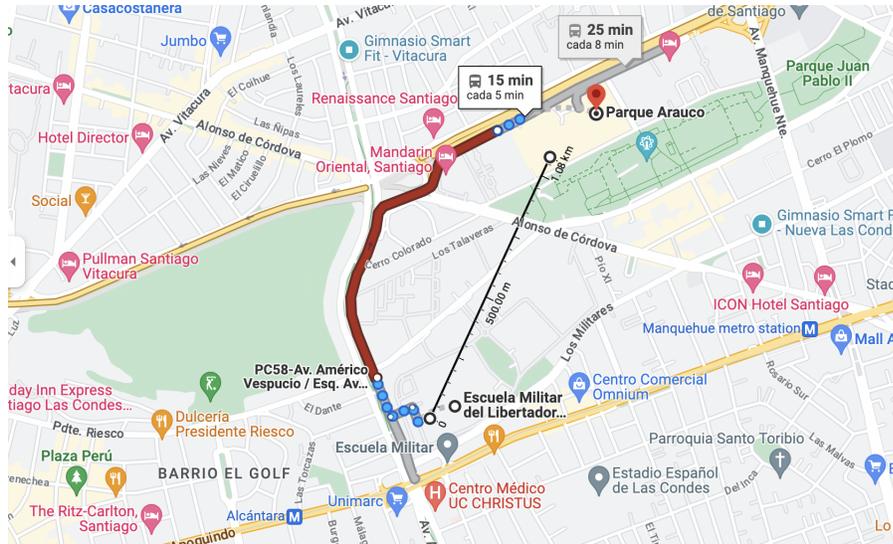


Figura 4.6: Transporte a Parque Arauco. Fuente: Google Maps.

La apertura de la nueva estación Parque Araucano representa un avance significativo en el acceso al centro comercial. Actualmente, a partir de la estación de metro más cercana (Escuela Militar), los usuarios deben realizar una combinación con un bus que cubre un trayecto de 1.9 km hasta el centro comercial. Sin embargo, con la nueva estación, los usuarios de las líneas 1, 2, 3, 5 y 6 pueden llegar directamente sin necesidad de hacer este transbordo. Para aquellas personas que no llegan caminando al Parque Arauco (cuando la distancia al centro comercial es mayor a 1 km), se considera una reducción de un kilómetro en la variable $DISTANCIA_T.PUBLICO$. Esta suposición se debe a que, actualmente, la distancia directa entre la estación de metro y el centro comercial es de 1 km como se muestra en la imagen 4.6 y el bus que recorre cerca de 2 km. La nueva estación se ubicará a tan solo 40 metros del destino final, lo que implica un considerable ahorro en tiempo y comodidad para los visitantes. El escenario antes y después de la apertura es el siguiente:

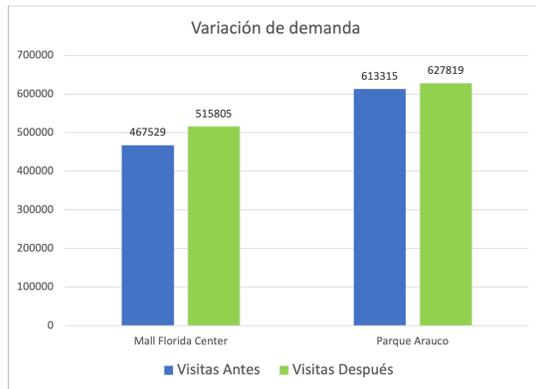
Tabla 4.3: Caso 2: Parque Arauco

Parque Arauco	Antes	Después
$DIST_MALL_METRO$	0.788	0.040
$DISTANCIA_T.PUBLICO$	=	-1km

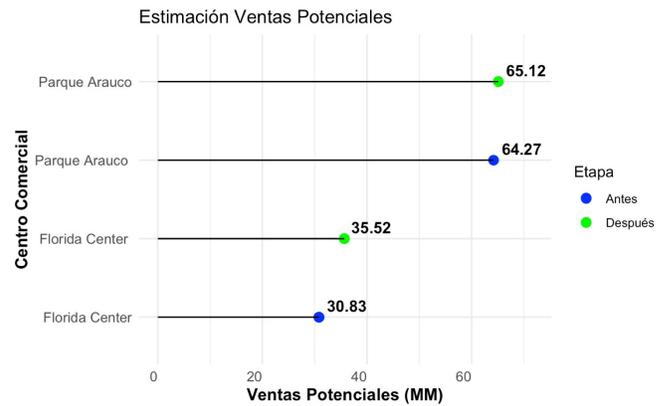
¿Cómo impactaría en la demanda y ventas los cambios mencionados? Aplicando las mo-

dificaciones en las variables, se utiliza el modelo 3.4 para obtener la nueva probabilidad y las fórmulas propuestas, obteniendo los siguientes resultados:

Estimación Caso 1 y Caso 2



(g) Impacto en Demanda



(h) Impacto en Ventas Potenciales

Para ambos casos, este cambio de infraestructura provocó un aumento en las visitas y ventas estimadas. En particular, en el Florida Center un aumento del 10 % de clientes significa un aumento de 4.69 MM considerando la naturaleza de sus clientes. Mientras que para el Parque Arauco, la apertura del metro aumenta un 2 % de visitas, incrementando 0,85 MM sus ventas potenciales. Con esta información los administradores del Florida Center pueden estimar en cuánto tiempo su inversión tendrá retorno. Mientras que los del Parque Arauco pueden invertir ese dinero extra en proyectos de interés para recibir a los nuevos clientes. Además, esta estimación es útil para la competencia, al tener conocimiento de este dinamismo, puede ejercer acciones para combatir este efecto.

Como se ha demostrado, el modelo desarrollado, puede ayudar potencialmente a la toma de decisiones estratégicas en un destino comercial. Se pueden mencionar otras como:

Ubicación locación centro comercial: Si bien el objetivo del modelo no es optimizar el lugar de ubicación, este permite comparar la estimación de demanda y ventas potenciales dependiendo de variables relevantes las cuales los centros comerciales tienen el control y otras que no. Como variables relevantes se menciona distancia a la estación de metro más cercana, es por esto la importancia de que al construir exista una estación de metro cerca, de lo contrario, se pueden tomar medidas como buses de acercamiento del metro al destino, como es el caso del Mall Plaza Los Domínicos, el cual se encuentra en el puesto 21 con respecto a la cercanía al metro con 0.548 km y dispone de buses en el metro Los Dominicos que se dirigen de forma gratuita hasta sus instalaciones.

Si existe una planificación de construcción de estación de metro, como se demostró antes, es posible que lleguen más clientes. Es por esto que pueden ser justificadas decisiones de construcción. Andrés Torrealba, CEO de Parque Arauco Chile, declara en (Diario Financiero, 2021), con respecto a la apertura del metro y las ampliaciones a futuro del centro comercial que "Hay espacio para seguir ampliando este centro comercial, muy asociado a la nueva estación de metro. Tenemos factibilidad para hacer oficinas, hotel y residencial. Lo que hay que ver es cuál de esos usos es el más atractivo. Pero, efectivamente, queda espacio para crecer y

lo más probable es que sea en altura".

Publicidad: Al conocer la probabilidad de visita y características socio-económicas de las manzanas, los centros comerciales pueden potenciar campañas publicitarias en las áreas de interés para cumplir con sus objetivos estratégicos. Esto les permite focalizar sus esfuerzos en áreas de mayor interés, maximizando así el impacto de sus iniciativas. Al comprender los perfiles de los potenciales visitantes, se abre la puerta a estrategias de marketing más efectivas y, en última instancia, a un aumento en la participación y la rentabilidad del centro comercial.

Para finalizar, el modelo desarrollado se muestra como una herramienta útil para la toma de decisiones estratégicas en el contexto de centros comerciales. Permite una comprensión más profunda y precisa del comportamiento de los clientes, arrojando luz sobre sus preferencias y elementos que afectan en su visita. Además, proporciona una valiosa caracterización espacial, identificando las zonas de mayor atractivo y sus características demográficas. El modelo tiene la capacidad de anticipar la respuesta de los clientes ante cambios en el entorno, ya sean voluntarios o involuntarios por parte de los centros comerciales, lo que otorga una ventaja competitiva. Al estar respaldado por datos de movilidad, el modelo no solo refleja la realidad actual, sino que también puede adaptarse a futuras transformaciones, asegurando su relevancia a largo plazo en la toma de decisiones estratégicas.

Capítulo 5

Conclusiones y Trabajo Futuro

En este capítulo, primero se presentan las conclusiones derivadas del estudio y desarrollo de modelos de probabilidad de visita a centros comerciales utilizando datos móviles. Luego, considerando el dinámico entorno de la industria minorista y la constante evolución de los comportamientos de los consumidores, existe un amplio espacio para futuras investigaciones y mejoras en la metodología empleada.

5.1. Conclusiones

Respecto al trabajo realizado, se puede mencionar que:

- Se pudo comprobar los beneficios de implementar datos de movilidad en comparación de utilizar fuentes de datos tradicionales como las encuestas, evidenciando una mejora estimada del 64 % de ajuste. Los datos de movilidad permiten registrar de manera masiva las visitas realizadas, capturando el origen del cliente, del cual se extrae información relevante útil para su caracterización y entendimiento.
- Como indica la literatura, la distancia tiene un impacto negativo significativo en la elección de centro comercial visitado, destacando como la variable más relevante junto al tiempo de viaje. Además, factores demográficos y económicos de las personas y el área comercial como cantidad de supermercado en la zona, porcentaje de hombres, cantidad de personas e índice de hogar fueron significativas y afectan en la visita a centros comerciales, al igual que la infraestructura de estos mismos.
- Se logró comprender el estado del arte del caso, en donde se pudo identificar los modelos que explican el comportamiento de visita del clientes. Algunos autores dividen los modelos de retail entre modelos de elección de tienda y selección de ubicación, pero en la práctica son complementarios.
- Se determinaron métricas de evaluación las cuales muestran que los modelos random forest obtuvieron mejor resultado en comparación al modelo tradicional de Huff, MCI y regresión. Además, la incorporación de variables de destino, origen y transporte son un aporte para el ajuste del modelo, en especial aquellas de transporte.

- Se proponen aplicaciones útiles del modelo como estimación de demanda, caracterización de clientes, cálculo de ventas potenciales e impacto de cambios internos y externos del contexto comercial.
- La calibración del modelo se ve limitada por la naturaleza de los datos, que están agregados a nivel de manzana y representan el total de visitas en un mes particular. A pesar de esta limitación, la fortaleza y estabilidad de los datos móviles confiere al modelo una base sólida para proyecciones futuras y vigencia a largo plazo.

5.2. Trabajo futuro

Para futuras líneas de investigación se sugiere:

- Añadir atributos de la oferta ofrecida por los centros comerciales como la variedad de tiendas. Además, se sugiere implementar más variables de transporte como factores de tráfico, metro tren y colectivo.
- Trabajar con variables de tiempo como en (Liang et al., 2020) donde implementan un modelo Huff dinámico consciente del tiempo que incorpora la variabilidad temporal horaria de las visitas a las tienda que obtuvo mejores resultados que el modelo estático. Agregando este tipo de variables se podría detectar variabilidad del comportamiento de la demanda ante acontecimientos específicos como vacaciones, festivos, eventos, etc.
- Solicitar base de datos desagregada por cliente, tipo panel, que registre frecuencia de visitas por cliente, con el fin de obtener más precisión de resultados, como se menciona en (Aaker, D. A., & Jones, J. M., 1971). Además de poder ejecutar modelos tipo Modelo Logit Multinomial. También, realizar modelos Mother Logit, que incluya los atributos de otros centros comerciales en su decisión de compra con el fin de detectar factores de competencia (Timmermans et al., 1991).
- Emplear el modelo con nuevos datos, buscando establecer la frecuencia más adecuada para mantener resultados precisos, así como determinar el momento óptimo para recalibrar el modelo con información actualizada.

Bibliografía

- [1] Acuña, G. (2021, noviembre 24). Los grupos socioeconómicos en Chile. Percepciones Económicas. [1](#)
- [2] Albornoz, M. (2007). Modelamiento de la Elección de Supermercados en el Gran Santiago. Memoria para Optar al Título de Ingeniero Civil Industrial. Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
- [3] Arcgis. (s/f). Huff Model (Business Analyst). Arcgis.com. Recuperado el 15 de julio de 2023. [3](#)
- [4] Applebaum, W. (1966). Methods for determining store trade areas, market penetration, and potential sales. *JMR, Journal of marketing research*, 3(2), 127. [4](#)
- [5] Asociación Nacional de Avisadores de Chile. (2019). Estilos de vida de los grupos socioeconómicos de Chile. *Anda.cl*. [5](#)
- [6] Baviera, M. A., Buitrago, J. M., & Rodríguez, J. E. (2012). Análisis metodológico del modelo de interacción espacial MCI. Documentos de Trabajo de la Cátedra Fundación Ramón Areces de Distribución Comercial (DOCFRADIS), 1–22.
- [7] Asplund, M., Friberg, R., & Wilander, F. (2007). Demand and distance: Evidence on cross-border shopping. *Journal of Public Economics*, 91(1–2), 141–157. [7](#)
- [8] Breiman, L. (2001). *Machine learning*, 45(1), 5–32. [8](#)
- [9] Cámara de Centro Comerciales Chile. (2018, mayo 17). Cámara de Centro Comerciales Chile; Cámara de Centros Comerciales. [9](#)
- [10] Carpenter, J. M., & Moore, M. (2006). Consumer demographics, store attributes, and retail format choice in the US grocery market. *International Journal of Retail & Distribution Management*, 34(6), 434–452. [10](#)
- [11] Catalán, J. (2020). Valoración de empresa Parque Arauco. *Uchile.cl*. [11](#)
- [12] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. Technical report, 9(13), 1-73.
- [13] Chasco, M. (1997). Modelos de determinación de áreas de mercado del comercio al por menor. *www.uv.es*. [13](#)
- [14] Chatfield, C. (1986). Exploratory data analysis. *European Journal of Operational Research*, 23(1), 5–13. [14](#)
- [15] Cohen, S., & Applebaum, W. (1960). Evaluating Store Sites and Determining Store Rents. *Economic Geography*. [15](#)
- [16] Converse, P. D. (1946). “Retail Trade Areas in Illinois”. Urbana, Illinois: University of

- Illinois Bulletin, Business Studies No. 4.
- [17] Das, G. (2014). Store personality and consumer store choice behaviour: an empirical examination. *Marketing Intelligence & Planning*, 32(3), 375–394. [17](#)
- [18] Diario Financiero. (2021). Parque Arauco Kennedy proyecta nuevos edificios y cuadruplica tamaño desde su inicio. [www.df.cl](#). [18](#)
- [19] Diario Financiero. (2023). Cencosud Shopping anuncia construcción de nuevo mall en Vitacura y ampliación del Florida Center. [www.df.cl](#). [19](#)
- [20] El Definido - Noticias para cambiar el mundo. (s/f). [Eldefinido.cl](#). Recuperado el 20 de marzo de 2023 [20](#)
- [21] Finn, A., & Louviere, J. J. (1996). Shopping center image, consideration, and choice: Anchor store contribution. *Journal of Business Research*, 35(3), 241–251. [21](#)
- [22] Fotheringham, A. S. & O’Kelly, M. E. (1989). *Spatial Interaction Models: Formulations and Applications*. Dordrecht: Kluwer Academic.
- [23] Fotheringham, H. A. (1985). *Gravity and Spatial Interaction Models*.
- [24] Galetovic, A., Poduje, I., & Sanhueza, R. (s/f). Malles en Santiago: de centros comerciales a centros urbanos. [Uandes.cl](#). Recuperado el 3 de diciembre de 2022 [24](#)
- [25] Gobierno Regional Metropolitano de Santiago. (s/f). Datos Geográficos. [Gobiernosantiago.cl](#). Recuperado el 22 de mayo de 2023 [25](#)
- [26] Gosh, A., & McLafferty, S. L. (1987). *Estrategias de Localización para Retail y Empresas de Servicios*. Lexington Books, Reading, Mass. Ghosh, A., Rushton, G.,. Lexington Books, Reading, Mass. Ghosh, A., Rushton, G.
- [27] Hermosilla, I. (2022, febrero 18). Línea 7 del Metro: estas son las 19 estaciones que unirán Renca con Vitacura. [BioBioChile](#). [27](#)
- [28] Huang, Z., Yin, G., Peng, X., Zhou, X., & Dong, Q. (2023). Quantifying the environmental characteristics influencing the attractiveness of commercial agglomerations with big geo-data. *Environment and Planning. B, Urban Analytics and City Science* [28](#)
- [29] Huff, D. (1963). A Probability Analysis of Shopping Center Trade Areas. *Land Economics*, 81-90.
- [30] Huff, D. (1964). Defining and estimating a trading area. *Journal of Marketing*, 28, 34-8.
- [31] Instituto Nacional de Estadística. (2018). Síntesis de Resultados VIII Encuesta de Presupuestos Familiares. [Gob.cl](#). [31](#)
- [32] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science (New York, N.Y.)*, 349(6245), 255–260. [32](#)
- [33] Liang, Y., Gao, S., Cai, Y., Foutz, N. Z., & Wu, L. (2020). Calibrating the dynamic Huff model for business analysis using location big data. *Transactions in GIS: TG*, 24(3), 681–703. [33](#)
- [34] McFadden, D. (1974). Conditional Logit Analysis of Quantitative Choice Behaviour. P. Zarembkar (ed), *Frontieres in Economics*, New York, Academic Press.34
- [35] McFadden, D. (1977). *Econometric Models of Probabilistic Choice*. *Econometric Models of Probabilistic Choice*”. C.F. Manski y D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, Mass.: MIT Press.

- [36] Marañón, M. (s/f). Reportes Sectoriales. Diario Financiero. Recuperado el 22 de mayo de 2023 [34](#)
- [37] McFadden, D., Tye, W., & Train, K.,. (1977). An Application of Diagnostic Tests for the Irrelevant Alternatives Property of the Multinomial Logit Model. *Transportation Research Record* 637 39-46.
- [38] Méndez, J., & Jara, M. (2020). Valoración de empresa Parque Arauco: por flujo de caja descontados. *Uchile.cl*. [38](#)
- [39] Moine, J., Gordillo, S., & Silvia, A. (2011). Análisis comparativo de metodologías para la gestión de proyectos de minería de datos. *Edu.ar*.[39](#)
- [40] Nakanishi, M., & Cooper, L. G. (1974). Parameter estimation for a multiplicative competitive interaction model: Least squares approach. *JMR, Journal of marketing research*, 11(3), 303–311. [40](#)
- [41] Niebuhr, A. (2004). Market Access and Regional Disparities. *Researchgate.net*. [41](#)
- [42] Peláez, I. M. (2006). Modelos de regresión: lineal simple y regresión logística. *Revista-seden.org*. [42](#)
- [43] Pope, J. A., Lane, W. R., & Stein, J. (2012). *Southern Business Review Southern Business Review*. *Core.ac.uk*. [43](#)
- [44] Quan, Z. X., & Youn, M.-K. (2016). Analysis on Preceding Study of Consumer’s Store-Choice Model: Focusing on Commercial Sphere Analysis Theories. *The Journal of Industrial Distribution & Business*, 7(4), 11–16. [44](#)
- [45] Quiroga, O., & Soledad, P. (2009). Retail en Chile: comportamiento de sus clientes en la relación precio procedencia de sus productos.[45](#)
- [46] Reilly, W. J. (1931). *The Law of Retail Gravitation*. W.J. Reilly.
- [47] Rueda, J. F. V. (2019, noviembre 4). CRISP-DM: una metodología para minería de datos en salud. *healthdataminer.com*. [47](#)
- [48] Sirmans, C. F.,& Guidry, K. A. (1993). The determinants of shopping center rents. *The Journal of real estate research*, 8(1), 107–115. [48](#)
- [49] Stanley, T. J., & Sewall, M. A. (1976). Image inputs to a probabilistic model: Predicting retail potential. *Journal of marketing*, 40(3), 48.. [49](#)
- [50] Suarez, M. (2022, mayo 25). Una mirada a la industria del retail en Chile. *Clase Ejecutiva UC*. [50](#)
- [51] Sud, D. F. (s/f). El top de los retailers de América Latina: México, Brasil y Chile los únicos en el ranking. *DF SUD*. Recuperado el 15 de mayo de 2023. [51](#)
- [52] Timmermans, H., Borgers, A., & van der Waerden, P. (1991). Mother logit analysis of substitution effects in consumer shopping destination choice. *Journal of Business Research*, 23(4), 311–323. [52](#)
- [53] Turhan, G., Akalın, M., & Zehir, C. (2013). Literature review on selection criteria of store location based on performance measures. *Procedia, Social and Behavioral Sciences*, 99, 391–402. [53](#)
- [54] Vandell, K., & Carter, C. (1994). Retail store location and market analysis: A review of the research. *Journal of Real Estate Literature*, 2(2), 13–45. [54](#)

[55] (Cuadrado-Roura, Fernández, y Santos, 2018). Flujos migratorios, características y consecuencias: un modelo gravitacional. 55

Anexos

Anexo A. Procesamiento y limpieza de datos

A.1. Gráficos de Análisis Descriptivo

A.1.1. Gráficos de Datos de movilidad

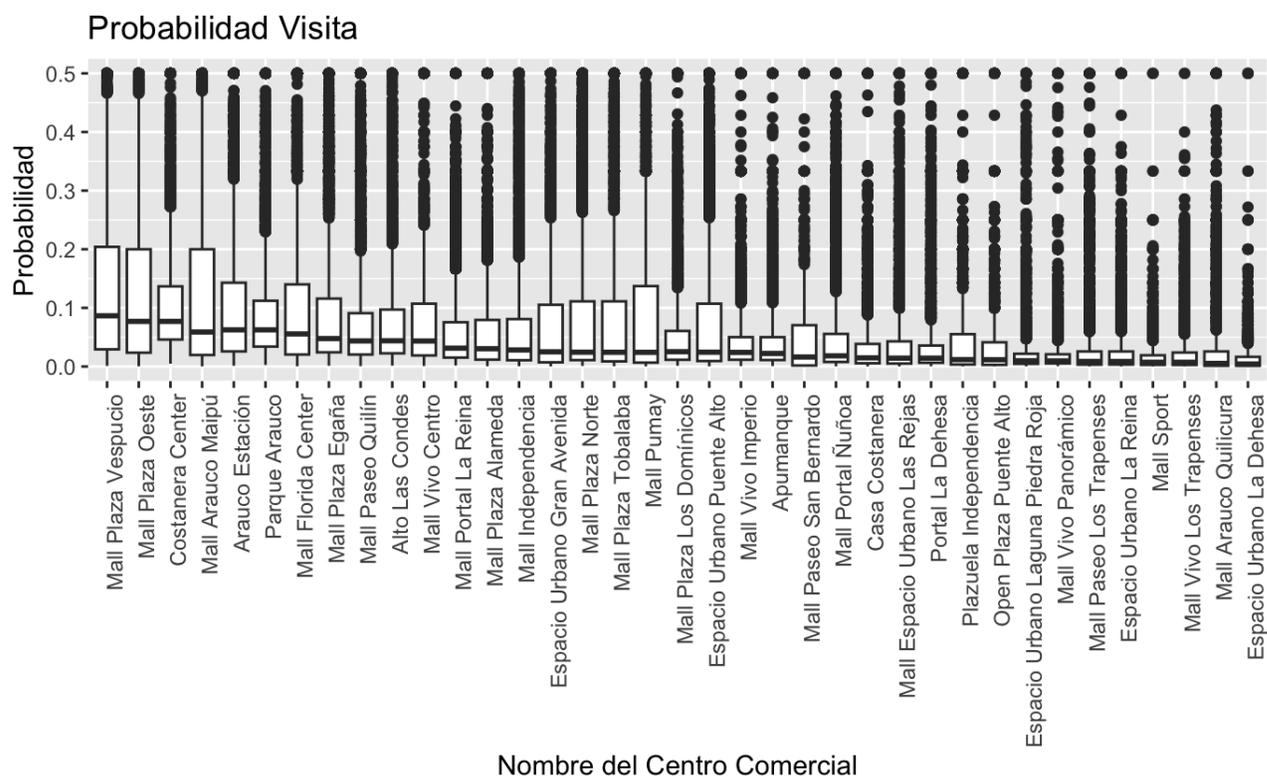


Figura A.1: Boxplot Probabilidad de Visita

n°	Nombre Comuna	Número de Visitas	Proporción
1	PUENTE ALTO	879939	11 %
2	LAS CONDES	567933	7 %
3	LA FLORIDA	559108	7 %
4	MAIPÚ	558255	7 %
5	SANTIAGO	517023	6 %

6	ÑUÑO A	379774	5 %
7	PEÑALOLÉN	345693	4 %
8	ESTACIÓN CENTRAL	248567	3 %
9	LO BARNECHEA	225909	3 %
10	MACUL	215043	3 %

Tabla A.1: Origen de Visitas Registradas

A.1.2. Gráficos de Datos de infraestructura comercial

Centro Comercial	Comuna	Sector
Mall Vivo Centro	SANTIAGO	CENTRAL
Mall Vivo Imperio	SANTIAGO	CENTRAL
Espacio Urbano La Reina	LA REINA	ORIENTE
Mall Plaza Egaña	LA REINA	ORIENTE
Alto Las Condes	LAS CONDES	ORIENTE
Apumanque	LAS CONDES	ORIENTE
Mall Plaza Los Domínicos	LAS CONDES	ORIENTE
Mall Portal La Reina	LAS CONDES	ORIENTE
Espacio Urbano La Dehesa	LO BARNECHEA	ORIENTE
Mall Paseo Los Trapenses	LO BARNECHEA	ORIENTE
Mall Sport	LO BARNECHEA	ORIENTE
Mall Vivo Los Trapenses	LO BARNECHEA	ORIENTE
Portal La Dehesa	LO BARNECHEA	ORIENTE
Mall Portal Ñuñoa	NUNOA	ORIENTE
Costanera Center	PROVIDENCIA	ORIENTE
Mall Vivo Panorámico	PROVIDENCIA	ORIENTE
Casa Costanera	VITACURA	ORIENTE
Parque Arauco	VITACURA	ORIENTE
Mall Plaza Norte	HUECHURABA	NORTE
Mall Independencia	INDEPENDENCIA	NORTE
Mall Arauco Quilicura	QUILICURA	NORTE
Espacio Urbano Laguna Piedra Roja	COLINA	PERIURBANA
Mall Paseo San Bernardo	SAN BERNARDO	SUR
Espacio Urbano Gran Avenida	SAN MIGUEL	SUR
Mall Florida Center	LA FLORIDA	SUR
Mall Plaza Vespucio	LA FLORIDA	SUR
Mall Paseo Quilín	PENANOLEN	ORIENTE
Espacio Urbano Puente Alto	PUENTE ALTO	SUR
Mall Plaza Tobalaba	PUENTE ALTO	SUR

Open Plaza Puente Alto	PUENTE ALTO	SUR
Plazuela Independencia	PUENTE ALTO	SUR
Mall Plaza Oeste	CERRILLOS	PONIENTE
Mall Espacio Urbano Las Rejas	ESTACION CENTRAL	PONIENTE
Mall Plaza Alameda	ESTACION CENTRAL	PONIENTE
Mall Arauco Maipú	MAIPU	PONIENTE
Mall Pumay	MAIPU	PONIENTE
Arauco Estación	ESTACION CENTRAL	PONIENTE

Tabla A.2: Ubicación Centros Comerciales RM

Sector	Cantidad	Porcentaje
ORIENTE	17	46 %
SUR	8	22 %
PONIENTE	6	16 %
NORTE	3	8 %
CENTRAL	2	5 %
PERIURBANA	1	3 %
Total	37	100 %

Tabla A.3: Sector Centros Comerciales

Tabla A.4: Superficie Centros Comerciales

Centro Comercial	SUPERFICIE M2
Alto Las Condes	115258
Apumanque	N/A
Arauco Estación	68500
Casa Costanera	28000
Costanera Center	173000
Espacio Urbano Gran Avenida	N/A
Espacio Urbano La Dehesa	17300
Espacio Urbano La Reina	N/A
Espacio Urbano Laguna Piedra Roja	N/A
Espacio Urbano Puente Alto	40000
Mall Arauco Maipú	66000
Mall Arauco Quilicura	32000
Mall Espacio Urbano Las Rejas	9400
Mall Florida Center	110000
Mall Independencia	106000
Mall Paseo Los Trapenses	N/A
Mall Paseo San Bernardo	N/A
Mall Paseo Quilín	80000
Mall Plaza Alameda	57000
Mall Plaza Egaña	88000
Mall Plaza Los Domínicos	90000
Mall Plaza Norte	100000
Mall Plaza Oeste	145000
Mall Plaza Tobalaba	54000
Mall Plaza Vespucio	131000
Mall Portal La Reina	N/A
Mall Pumay	N/A
Mall Portal Ñuñoa	35000
Mall Sport	N/A
Mall Vivo Centro	N/A
Mall Vivo Imperio	30348
Mall Vivo Los Trapenses	26679
Mall Vivo Panorámico	6953
Open Plaza Puente Alto	13600
Parque Arauco	138000
Plazuela Independencia	N/A
Portal La Dehesa	143468
TOTAL	1904506

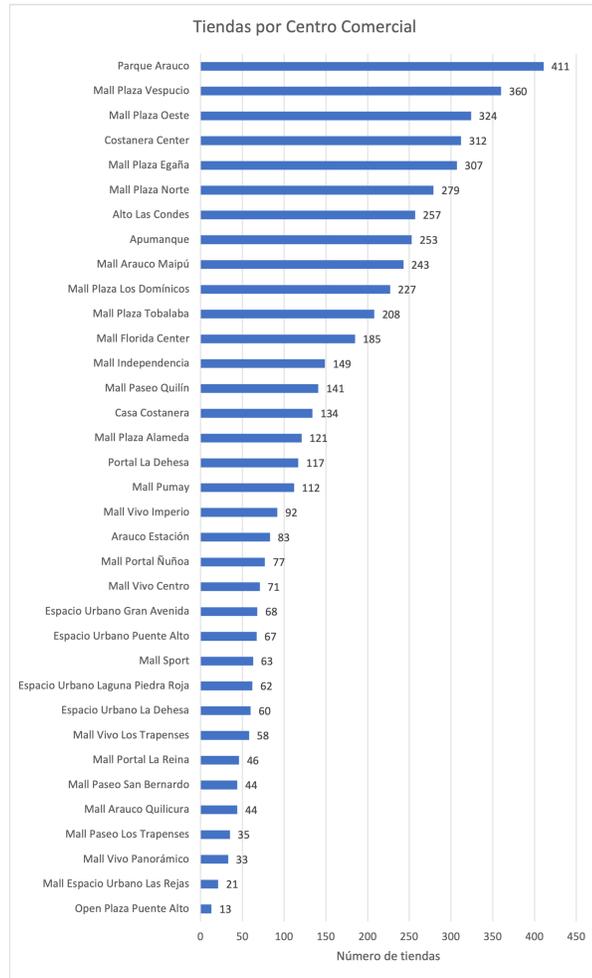


Figura A.2: Cantidad de tiendas por Centro Comercial

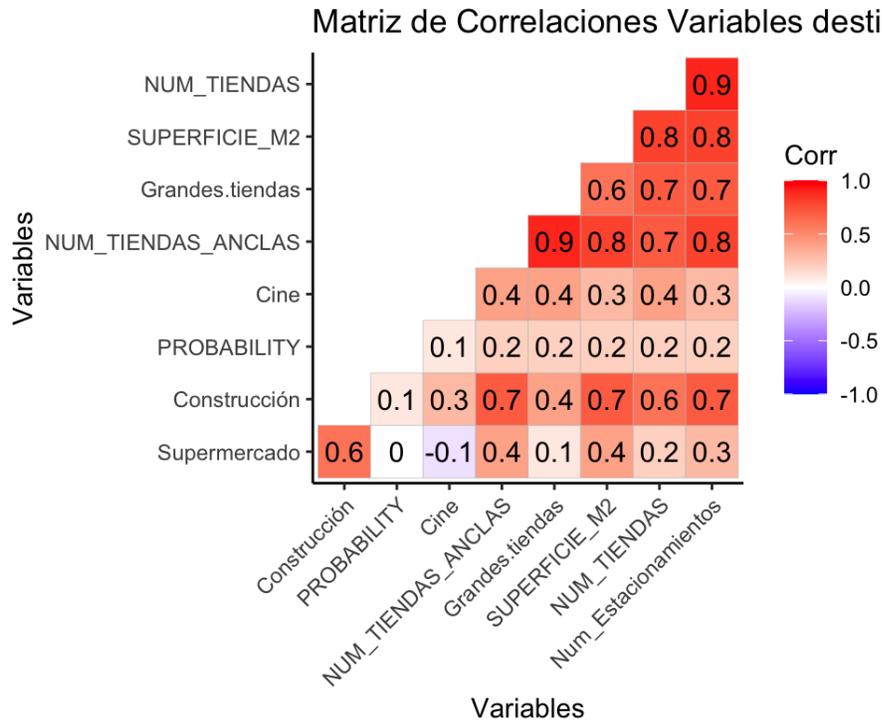


Figura A.3: Correlación variables de destino

A.1.3. Gráficos de Datos de población

Zona Norte	Zona Oriente	Zona Centro	Zona Sur	Zona Poniente	Periurbana
Conchalí Huechuraba Independencia Recoleta Quilicura Renca	Lo Barnechea Vitacura Las Condes Providencia Ñuñoa La Reina Macul Peñalolén	Santiago	San Joaquín San Miguel La Cisterna La Granja El Bosque San Ramón La Florida San Bernardo La Pintana Puente Alto	Cerro Navia Quinta Normal Lo Prado Estación Central Cerrillos Pedro Aguirre Cerda Lo Espejo Maipú Pudahuel	Colina Peñaflor Buin Lampa Melipilla Talagante El Monte Paine Isla de Maipo Calera de Tango Til til Curacaví Alhué María Pinto Padre Hurtado Pirque

Tabla A.5: Zonas RM. Fuente: Diario Financiero

Nivel Socioeconómico	Porcentaje manzanas
ABC1	6 %
C2	17 %
C3	24 %
D	53 %
E	0 %

Tabla A.6: Nivel socioeconómico manzanas RM

El grupo ABC1 es la suma de los grupos AB, C1a y C1b.

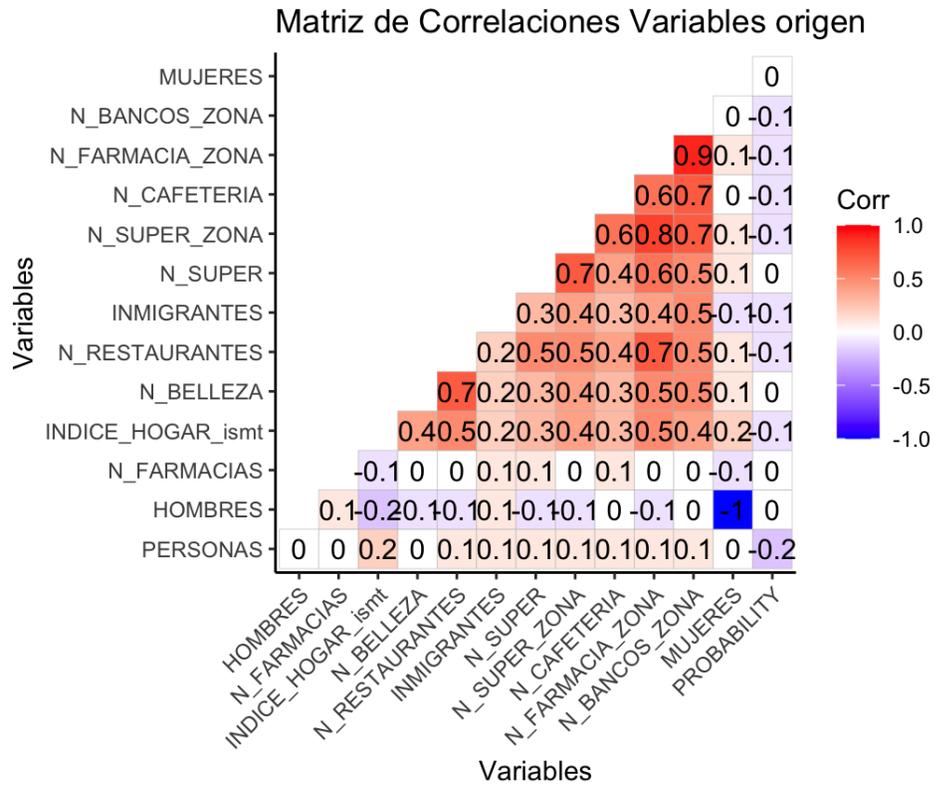


Figura A.4: Correlación variables de origen

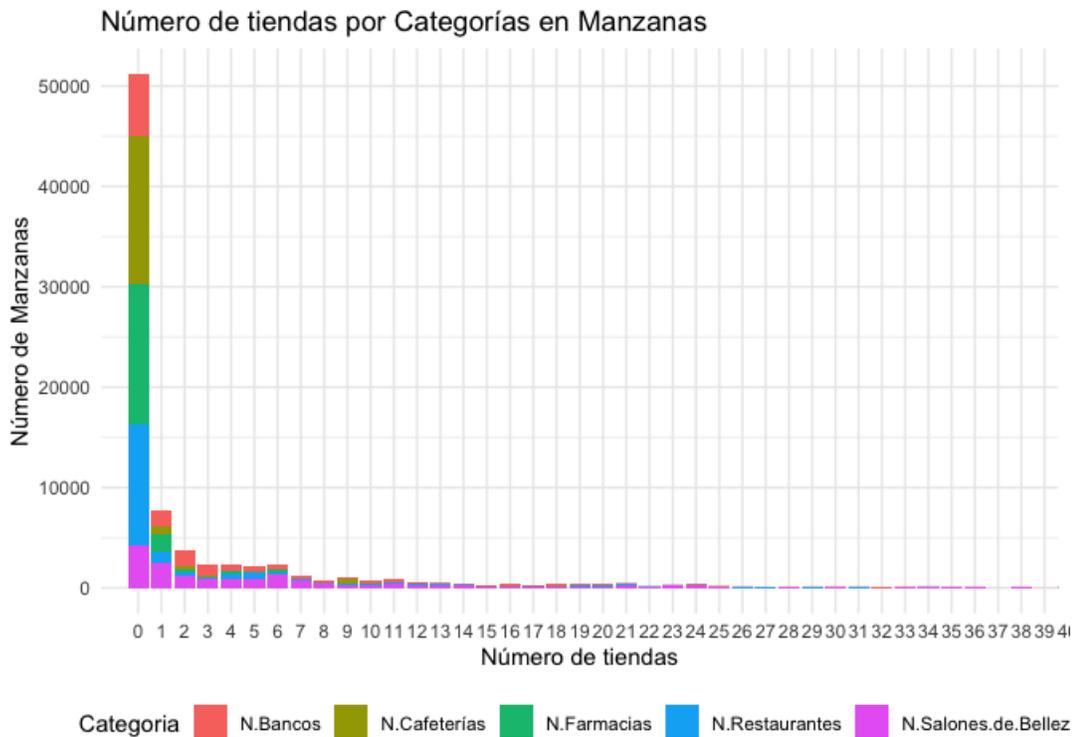


Figura A.5: Área comercial de manzanas

A.1.4. Gráficos de Datos de Sistema de Transporte

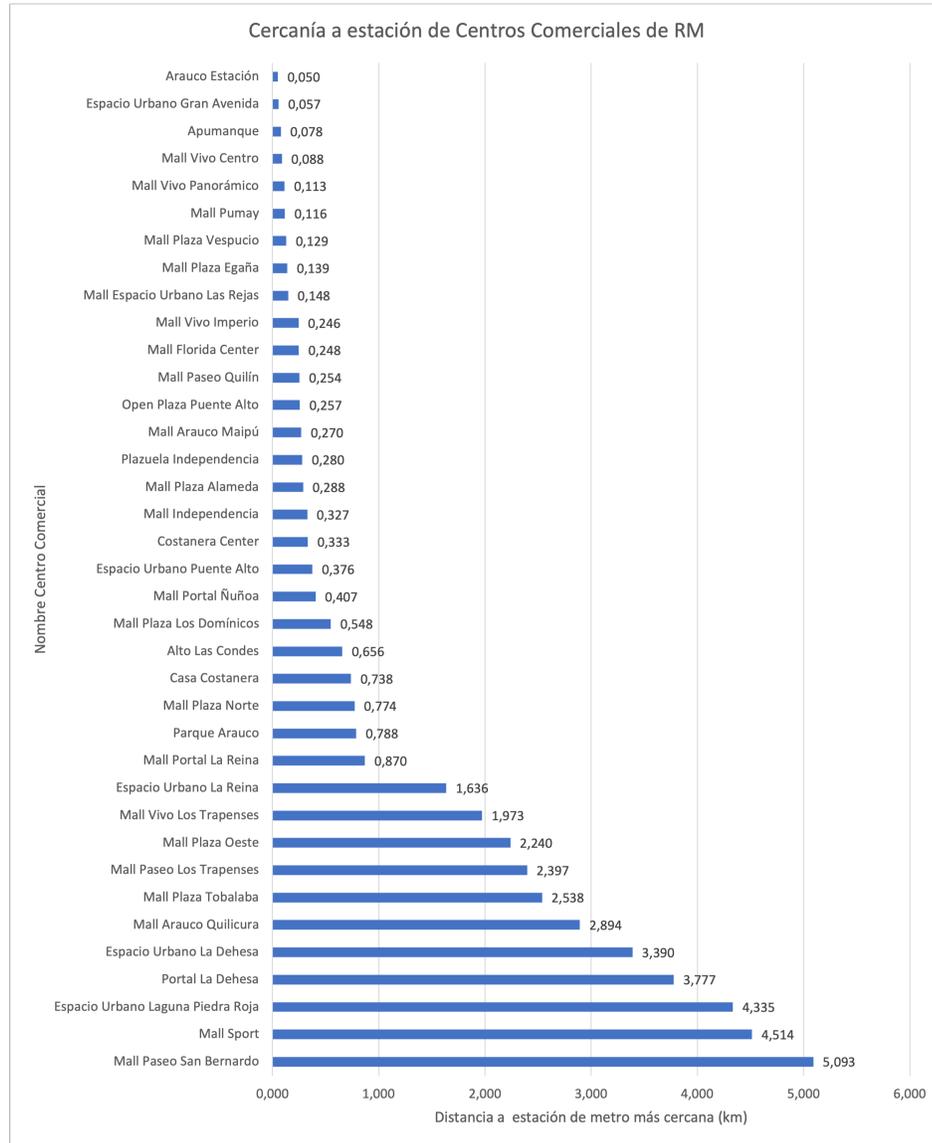


Figura A.6: Cercanía a estación de metro

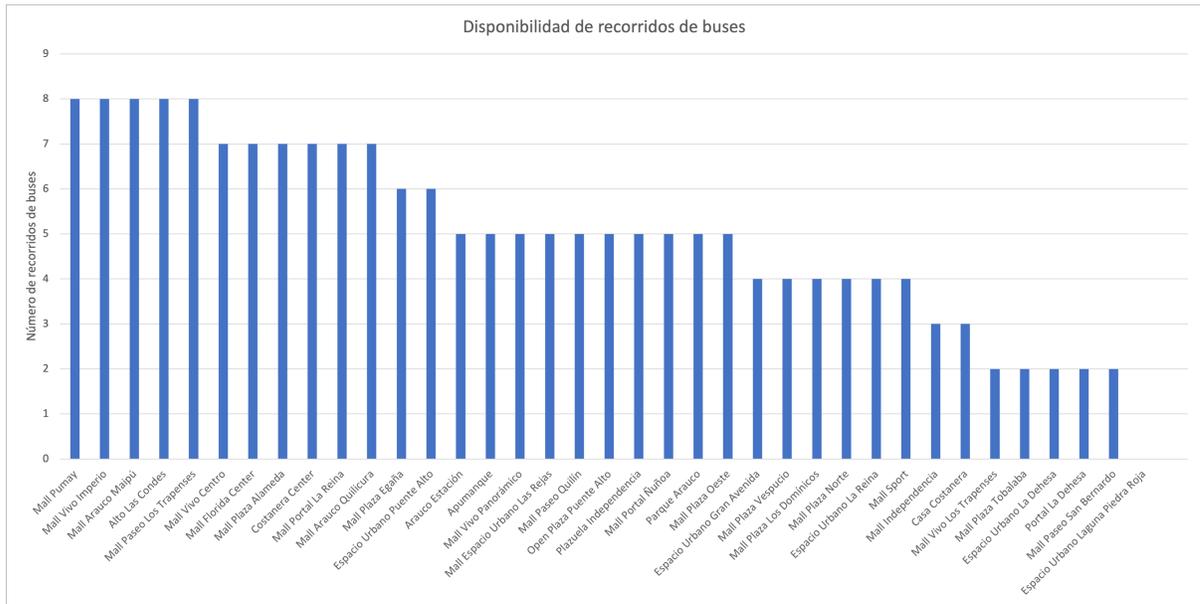


Figura A.7: Disponibilidad de recorrido de buses

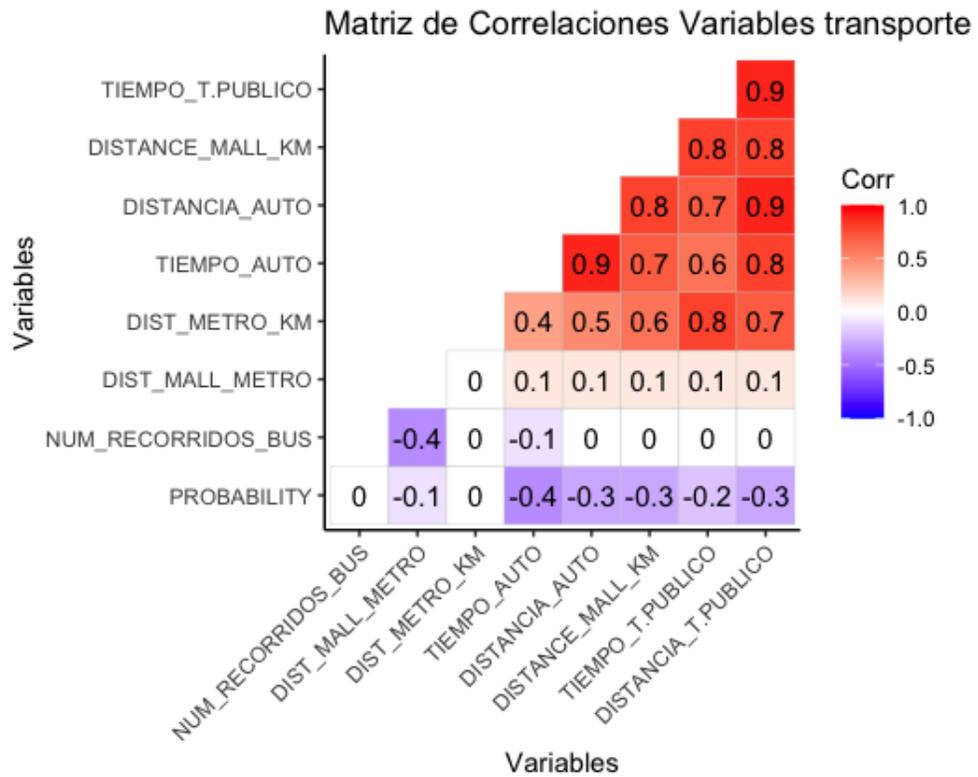


Figura A.8: Correlación variables de movilidad

Anexo B. Selección de Atributos

B.1. Atributos Centros Comerciales

Tabla B.1: Variables atributos destino

Nombre Variable	Etiqueta	Descripción	min	max	median	desv
SUPERFICIE_M2	Superficie en M2	Superficie del centro comercial medida en metros cuadrados	6230	173000	88000	45943.09
Num_Estacionamientos	Número de estacionamientos	Cantidad de estacionamientos en el centro comercial	0	5000	3000	1803.027
NUM_TIENDAS	Número de tiendas	Cantidad de tiendas que tiene el centro comercial considerando tiendas de retail, entretención, comida y servicios.	13	411	182	114.842
NUM_TIENDAS_Anclas	Número de tiendas anclas	Cantidad de tiendas capaces de atraer gran cantidad de público (Supermercados, grandes tiendas, cine, tiendas de construcción)	0	8	5	2.240
Supermercado	Número de supermercados	Cantidad de supermercados dentro del mall tales como Lider, Santa Isabel, Jumbo y Tottus	1	2	1	0.483
Grandes.tiendas	Número de grandes tiendas	Cantidad de grandes tiendas dentro del mall como Falabella, Paris, Ripley.	0	5	2	1.482
Construcción	Número tiendas de construcción	Cantidad de tiendas de construcción grandes dentro del mall como Easy, Ikea y Sodimac.	0	1	0	0.499
Cine	Número de cines	Cantidad de cines en el centro comercial como Cineplanet, Cine Hoyts (Cinépolis) y Cinemark.	0	1	1	0.421

B.2. Atributos Población

Tabla B.2: Variables Atributos origen

Nombre Variable	Etiqueta	Descripción	min	max	desv
PERSONAS	Número total de personas	Corresponde al total de personas efectivamente censadas en la manzana	0	9197	430.865
MUJERES	Total de mujeres	Corresponde al total de mujeres efectivamente censadas	0	1	0.051
HOMBRES	Número total de hombres	cantidad de hombres efectivamente censadas en la manzana	0	1	0.051
INMIGRANTES	Total de personas migrantes	Corresponde al total de personas no nacidas en el país efectivamente censadas en la manzana que residen habitualmente en el territorio nacional	0	0.88	0.114
INDICE_HOGAR_ismt	Indicador Socio Material Territorial (ISMT)	Indicador Socio Material Territorial, calculado a partir de variables censales (2017) correspondientes a escolaridad, hacinamiento, allegamiento y materialidad de la vivienda. Más alto el indicador, mejor puntaje.	0.606	0.920	0.056
GSE_N	Grupo Socioeconómico	Categorías de la clasificación a la población en base a factores económicos y sociales	character		
AVE_GSE	Grupo Socioeconómico Agregado	Categorías de la clasificación a la población en base a factores económicos y sociales agrupados (el grupo ABC1 es la suma de los grupos AB, C1a y C1b.)	character		
N_SUPER	Número Total de Supermercados en Zona	Cantidad de tiendas de supermercados de todos los tamaños del área comercial de origen (zona)	0	168	36.340
N_FARMACIAS	Número Total de Farmacias en Zona	Cantidad total de farmacias en el área comercial de origen (zona)	0	6	
N_BELLEZA	Cantidad de salones de Belleza en Zona	Corresponde a la cantidad de salones de belleza ubicados en la zona como peluquería, barbería, salón de manicura, etc.	0	65	10.404
N_CAFETERIA	Número de Cafeterías en Zona	Cantidad de cafeterías en el área comercial de origen (zona)	0	10	1.919
N_RESTAURANTES	Número de Restaurantes en Zona	Cantidad de restaurantes en el área comercial de origen (zona)	0	43	6.407
N_FARMACIA_ZONA	Número de Farmacias en Zona	Cantidad de grandes farmacias en el área comercial de origen (zona)	0	57	11.264
N_BANCOS_ZONA	Número de Bancos en Zona	Cantidad de bancos en el área comercial de origen (zona)	0	132	21.166
N_SUPER_ZONA	Número de Supermercados grandes en Zona	Cantidad de supermercados de grande cadenas en el área comercial de origen (zona)	0	17	3.258
NOM_COMUNA	Nombre Comuna	Nombre de la comuna de la manzana de origen	character		

B.3. Atributos Transporte

Tabla B.3: Variables sistema de transporte

Nombre Variable	Etiqueta	Descripción	min	max	median	desv
METRO_CERCANO_MANZANA	Nombre del metro manzana	Nombre del metro más cercano a la manzana			character	
METRO_CARCANO_MALL	Nombre del metro mall	Nombre del metro más cercano al centro comercial			character	
DISTANCE_MALL_KM	Distancia manzana - mall	Distancia en línea recta entre manzana y mall en km	0.048	81.355	7.452	8.243
DIST_METRO_KM	Distancia metro - manzana	Distancia entre metro más cercano a la manzana y manzana	0.016	53.91	2.210	4.861
DIST_MALL_METRO	Distancia metro-mall	Distancia entre el metro más cercano al centro comercial y centro comercial	0.0502	5.092	0.327	1.089
TIEMPO T.PUBLICO_{ij}	Tiempo de viaje en transporte público	Tiempo de viaje en minutos desde la zona hasta el centro comercial usando transporte público	0.217	737.483	50.317	45.063
DISTANCIA_T.PUBLICO_{ij}	Distancia en transporte público	Distancia en kilómetros recorrida desde la zona del cliente hasta el centro comercial en transporte público	0.017	148.664	13.782	11.462
TIEMPO_AUTO_{ij}	Tiempo de viaje en auto	Tiempo de viaje en minutos desde la zona hasta el centro comercial usando vehículo particular	0.10	141.53	23.10	10.542
DISTANCIA_AUTO_{ij}	Distancia en auto	Distancia en kilómetros recorrida desde la zona del cliente hasta el centro comercial en vehículo particular	0.017	158.859	15.414	12.549
NUM_RECORRIDOS_BUS_{j}	Número de recorridos de buses	Cantidad de recorridos de buses de los paraderos más cercanos al centro comercial j	0.00	9.00	5.00	19.399

Anexo C. Elección, desarrollo, resultados y evaluación de modelos

Variabes	α	MAE	MAPE	RMSE
SUPERFICIE_M2 TIEMPO_AUTO	-0.847	0.120	100 %	0.138

Tabla C.1: Resultados de la simulación de la encuesta

C.1. Desarrollo, resultado y evaluación de modelos

- Modelo de Huff

Se aplica el Modelo de Huff para el Costanera Center utilizando variables de tiempo en transporte público y tiempo en auto.

Modelo	Modelo 1.1-v1	Modelo 1.1-v2
Variabes	SUPERFICIE_M2 TIEMPO_T.PUBLICO	SUPERFICIE_M2 TIEMPO_AUTO
α	273915	-0.055
MAE	0.071	0.070
RSME	0.101	0.100
MAPE	71 %	72 %

AIC	18263	19586
-----	-------	-------

Tabla C.2: Resultados Huff

El mejor modelo de Huff según AIC es el modelo 1.1-v1, usando tiempo en transporte público, por lo que se utiliza ese como replicación de modelo de Huff para datos móviles.

- Modelo MCI

Primer modelo MCI:

$$\ln\left(\frac{P_{ij}}{P_i}\right) = cte + \beta_1 \ln\left(\frac{SUPERFICIE_M2_j}{SUPERFICIE_M2_j}\right) + \beta_2 \ln\left(\frac{NUM_TIENDAS_j}{NUM_TIENDAS_j}\right) + \beta_3 \ln\left(\frac{NUM_TIENDAS_ANCLAS_j}{NUM_TIENDAS_ANCLAS_j}\right) \quad (C.1)$$

Tabla C.3: Resultado modelo 1.2

	log(PROBABILIDAD)	
Predictors	Estimates	std. Error
(Intercept)	-0.300 ***	0.028
SUPERFICIE Mt2 [log]	0.372 ***	0.014
Num de tiendas [log]	0.558 ***	0.011
1 + Num de tiendas anclas [log]	0.415 ***	0.037
Observations	345370	
R2 / R2 adjusted	0.094 / 0.094	
* p<0.05 ** p<0.01 *** p<0.001		

Si MAPE fuera partido por PROBABILITY real y no por PROBABILITY estimada, sería 218269 % en vez de 96 %

Los resultados del modelo con atributos de destino:

Al incorporar variables de origen:

$$\ln\left(\frac{P_{ij}}{P_i}\right) = cte + \beta_1 \ln\left(\frac{SUPERFICIE_M2_j}{SUPERFICIE_M2_j}\right) + \beta_2 \ln\left(\frac{NUM_TIENDAS_j}{NUM_TIENDAS_j}\right) + \beta_3 \ln\left(\frac{NUM_TIENDAS_ANCLAS_j}{NUM_TIENDAS_ANCLAS_j}\right) + \beta_4 \ln\left(\frac{PERSONAS_i}{PERSONAS_i}\right) + \beta_5 \ln\left(\frac{HOMBRES_i}{HOMBRES_i}\right) + \beta_6 \ln\left(\frac{INDICE_HOGAR_ismt_i}{INDICE_HOGAR_ismt_i}\right) + \beta_7 \ln\left(\frac{N_SUPER_ZONA_i}{N_SUPER_ZONA_i}\right) + \beta_8 \ln\left(\frac{N_FARMACIA_ZONA_i}{N_FARMACIA_ZONA_i}\right) \quad (C.2)$$

Sin embargo, al replicar el mismo atributo de origen a los distintos centros comerciales, los resultados se ven alterados, obteniendo los siguientes parámetros:

log(PROBABILIDAD)		
Predictors	Estimates	std. Error
(Intercept)	-0.000	0.003
SUPERFICIE M2 [log]	0.248 ***	0.009
NUM TIENDAS[log]	0.248 ***	0.006
NUM TIENDAS ANCLAS [log]	0.003	0.010
PERSONAS[log]	-3.703	13563769744015.547
HOMBRES [log]	-8.176	334039025151929.438
N FARMACIAS ZONA [log]	4.198	35083410269653.855
N SUPERMERCADO ZONA [log]	0.744	37198444929960.305
Observations	155676	
R2 / R2 adjusted	0.119 / 0.118	
p<0.05	** p<0.01	*** p<0.001

Tabla C.4: Resultados MCI origen destino sin transformación

Fórmula extendida con transformaciones:

$$\begin{aligned}
\ln\left(\frac{P_{ij}}{\hat{P}_i}\right) = & \alpha_0 + \ln\left(\frac{SUPERFICIE_M2_j}{SUPERFICIE_M2_j}\right) (\alpha_1 + \alpha_2 PERSONAS_i + \alpha_3 HOMBRES_i + \alpha_4 INDICE_HOGAR_ismt_i + \\
& \alpha_5 N_FARMACIAS_i + \alpha_6 N_SUPER_i + \alpha_7 N_BELLEZA_i + \alpha_8 N_RESTAURANTES_i + \alpha_9 N_CAFETERIAS_i) + \\
& \ln\left(\frac{NUM_TIENDAS_j}{NUM_TIENDAS_j}\right) (\alpha_{10} + \alpha_{11} PERSONAS_i + \alpha_{12} HOMBRES_i + \alpha_{13} INDICE_HOGAR_ismt_i + \alpha_{14} N_FARMACIAS_i + \\
& \alpha_{15} N_SUPER_i + \alpha_{16} N_BELLEZA_i + \alpha_{17} N_RESTAURANTES_i + \alpha_{18} N_CAFETERIAS_i) + \\
& \alpha_{19} \ln\left(\frac{1+NUM_TIENDAS_ANCLAS_j}{NUM_TIENDAS_ANCLAS_j}\right) (\alpha_{20} + \alpha_{21} PERSONAS_i + \alpha_{22} HOMBRES_i + \alpha_{23} INDICE_HOGAR_ismt_i + \\
& \alpha_{24} N_FARMACIAS_i + \alpha_{25} N_SUPER_i + \alpha_{26} N_BELLEZA_i + \alpha_{27} N_RESTAURANTES_i + \alpha_{28} N_CAFETERIAS_i)
\end{aligned} \tag{C.3}$$

Tabla C.5: Resultados modelo 1.3

	log(PROBABILIDAD)	
Predictors	Estimates	std. Error
(Intercept)	-0.003	0.005
SUPERFICIE Mt2 [log]	-4.120 ***	0.371
Num de tiendas [log]	-1.743 ***	0.267
Num de tiendas anclas [log]	14.003 ***	0.554
SUPERFICIE Mt2 [log] × HOMBRES	-0.546	0.350
SUPERFICIE Mt2 [log] × INDICE HOGAR ismt	6.710 ***	0.383
SUPERFICIE Mt2 [log] × PERSONAS	0.000	0.000
SUPERFICIE Mt2 [log] × NBELLEZA	0.020 ***	0.002
SUPERFICIE Mt2 [log] × NCAFETERIA	0.025 *	0.012
SUPERFICIE Mt2 [log] × NRESTAURANTES	0.023 ***	0.004
SUPERFICIE Mt2 [log] × NFARMACIAS	-0.199 ***	0.016
SUPERFICIE Mt2 [log] × NSUPER	-0.010 ***	0.001
HOMBRES × Num de tiendas [log]	-0.228	0.251
INDICE HOGAR ismt × Num de tiendas [log]	2.831 ***	0.277
PERSONAS × Num de tiendas [log]	0.000 ***	0.000
N BELLEZA × Num de tiendas [log]	-0.007 ***	0.002
N CAFETERIA × Num de tiendas [log]	-0.015	0.009
N RESTAURANTES × Num de tiendas [log]	0.015 ***	0.003
N FARMACIAS × Num de tiendas [log]	0.018	0.011
N SUPER × Num de tiendas [log]	0.003 ***	0.000
HOMBRES × Num de tiendas anclas [log]	2.576 ***	0.523
INDICE HOGAR ismt × Num de tiendas anclas [log]	-19.801 ***	0.573
PERSONAS × Num de tiendas anclas [log]	0.000 ***	0.000
N BELLEZA × Num de tiendas anclas [log]	-0.042 ***	0.003
N CAFETERIA × Num de tiendas anclas [log]	0.004	0.018
N RESTAURANTES × Num de tiendas anclas [log]	-0.069 ***	0.007
N FARMACIAS × Num de tiendas anclas [log]	0.425 ***	0.024
N SUPER × Num de tiendas anclas [log]	0.015 ***	0.001
Observations	297070	
R2 / R2 adjusted	0.122 / 0.121	
* p<0.05 ** p<0.01 *** p<0.001		

Tabla C.6: Resultados modelo 1.4

	log(PROBABILIDAD)	
Predictors	Estimates	std. Error
(Intercept)	-0.129 ***	0.006
SUPERFICIE Mt2 [log]	1.004 ***	0.014
Num de tiendas [log]	0.291 ***	0.010
Num de tiendas anclas [log]	-0.575 ***	0.016
DISTANCIA TPUBLICO [log]	-1.557 ***	0.018
TIEMPO T AUTO [log]	-1.245 ***	0.026
N recorridos buses [log]	0.149 ***	0.014
DIST METRO MALL [log]	-0.122 ***	0.005
Observations	332826	
R2 / R2 adjusted	0.297 / 0.297	
* p<0.05 ** p<0.01 *** p<0.001		

$$\begin{aligned}
 \ln\left(\frac{P_{ij}}{P_i}\right) = & \alpha_0 + \ln\left(\frac{SUPERFICIE_M2_j}{SUPERFICIE_M2_j}\right) (\alpha_1 + \alpha_2 PERSONAS_i + \alpha_3 HOMBRES_i + \alpha_4 INDICE_HOGAR_ismt_i + \alpha_5 N_FARMA- \\
 & CIAS_i + \alpha_6 N_SUPER_i + \alpha_7 N_BELLEZA_i + \alpha_8 N_RESTAURANTES_i + \alpha_9 N_CAFETERIAS_i) + \ln\left(\frac{NUM_TIENDAS_j}{NUM_TIENDAS_j}\right) \\
 & (\alpha_{10} + \alpha_{11} PERSONAS_i + \alpha_{12} HOMBRES_i + \alpha_{13} INDICE_HOGAR_ismt_i + \alpha_{14} N_FARMACIAS_i + \alpha_{15} N_SUPER_i + \\
 & \alpha_{16} N_BELLEZA_i + \alpha_{17} N_RESTAURANTES_i + \alpha_{18} N_CAFETERIAS_i) + \ln\left(\frac{1 + NUM_TIENDAS_ANCLAS_j}{NUM_TIENDAS_ANCLAS_j}\right) (\alpha_{19} + \\
 & \alpha_{20} PERSONAS_i + \alpha_{21} HOMBRES_i + \alpha_{22} INDICE_HOGAR_ismt_i + \alpha_{23} N_FARMACIAS_i + \alpha_{24} N_SUPER_i + \\
 & \alpha_{25} N_BELLEZA_i + \alpha_{26} N_RESTAURANTES_i + \alpha_{27} N_CAFETERIAS_i) \ln\left(\frac{DISTANCIA_T.PUBLICO_{ij}}{DISTANCIA_T.PUBLICO_{ij}}\right) (\alpha_{28} + \\
 & \alpha_{29} PERSONAS_i + \alpha_{30} HOMBRES_i + \alpha_{31} INDICE_HOGAR_ismt_i + \alpha_{32} N_FARMACIAS_i + \alpha_{33} N_SUPER_i + \\
 & \alpha_{34} N_BELLEZA_i + \alpha_{35} N_RESTAURANTES_i + \alpha_{36} N_CAFETERIAS_i) + \ln\left(\frac{TIEMPO_AUTO_{ij}}{TIEMPO_AUTO_{ij}}\right) (\alpha_{37} + \alpha_{38} PERSO- \\
 & NAS_i + \alpha_{39} HOMBRES_i + \alpha_{40} INDICE_HOGAR_ismt_i + \alpha_{41} N_FARMACIAS_i + \alpha_{42} N_SUPER_i + \alpha_{43} N_BELLEZA_i + \\
 & \alpha_{44} N_RESTAURANTES_i + \alpha_{45} N_CAFETERIAS_i) + \ln\left(\frac{NUM_RECORRIDOS_BUS_j}{NUM_RECORRIDOS_BUS_j}\right) (\alpha_{46} + \alpha_{47} PERSONAS_i + \\
 & \alpha_{48} HOMBRES_i + \alpha_{49} INDICE_HOGAR_ismt_i + \alpha_{50} N_FARMACIAS_i + \alpha_{51} N_SUPER_i + \alpha_{52} N_BELLEZA_i + \\
 & \alpha_{53} N_RESTAURANTES_i + \alpha_{54} N_CAFETERIAS_i) + \ln\left(\frac{DIST_MALL_METRO_j}{DIST_MALL_METRO_j}\right) (\alpha_{55} + \alpha_{56} PERSONAS_i + \\
 & \alpha_{57} HOMBRES_i + \alpha_{58} INDICE_HOGAR_ismt_i + \alpha_{59} N_FARMACIAS_i + \alpha_{60} N_SUPER_i + \alpha_{61} N_BELLEZA_i + \\
 & \alpha_{62} N_RESTAURANTES_i + \alpha_{63} N_CAFETERIAS_i)
 \end{aligned}
 \tag{C.4}$$

	log(PROBABILIDAD)	
Predictors	Estimates	std. Error
(Intercept)	-0.151 ***	0.006
SUPERFICIE Mt2 [log]	-0.201	0.302

Num de tiendas [log]	-0.440	0.237
Num de tiendas anclas[log]	1.860 ***	0.301
DISTANCIA TPUBLICO [log]	-3.854 ***	0.435
TIEMPO T AUTO [log]	-1.884 **	0.622
N recorridos buses [log]	-1.902 ***	0.340
DIST METRO MALL [log]	-0.623 ***	0.111
SUPERFICIE Mt2 [log] × HOMBRES	0.478	0.283
SUPERFICIE Mt2 [log] × INDICE HOGAR ismt	1.516 ***	0.314
SUPERFICIE Mt2 [log] × PERSONAS	0.000 *	0.000
SUPERFICIE Mt2 [log] × NBELLEZA	-0.002	0.002
SUPERFICIE Mt2 [log] × NCAFETERIA	0.020 *	0.010
SUPERFICIE Mt2 [log] × NRESTAURANTES	0.029 ***	0.004
SUPERFICIE Mt2 [log] × NFARMACIAS	-0.013	0.013
SUPERFICIE Mt2 [log] × NSUPER	-0.004 ***	0.000
HOMBRES × Num de tiendas[log]	-0.112	0.221
INDICE HOGAR ismt × Num de tiendas [log]	0.914 ***	0.248
PERSONAS × Num de tiendaslog]	0.000 ***	0.000
N BELLEZA × Num de tiendas [log]	-0.003 *	0.001
N CAFETERIA × Num de tiendas [log]	0.011	0.008
N RESTAURANTES × Num de tiendas [log]	-0.011 ***	0.003
N FARMACIAS × Num de tiendas [log]	0.024 *	0.010
N SUPER × Num de tiendas [log]	0.001 ***	0.000
HOMBRES × Num de tiendas anclas [log]	0.295	0.281
INDICE HOGAR ismt × Num de tiendas anclas [log]	-3.429 ***	0.314
PERSONAS × Num de tiendas anclas [log]	0.000	0.000
N BELLEZA × Num de tiendas anclas [log]	0.001	0.002
N CAFETERIA × Num de tiendas anclas [log]	0.004	0.010
N RESTAURANTES × Num de tiendas anclas [log]	-0.018 ***	0.004
N FARMACIAS × Num de tiendas anclas [log]	0.032 *	0.013
N SUPER × Num de tiendas anclas [log]	0.001 *	0.000
HOMBRES × DISTANCIA TPUBLICO [log]	0.491	0.393
INDICE HOGAR ismt × DISTANCIA TPUBLICO [log]	2.302 ***	0.457
PERSONAS × DISTANCIA TPUBLICO [log]	0.000 ***	0.000
N BELLEZA × DISTANCIA TPUBLICO [log]	-0.008 **	0.002
N CAFETERIA × DISTANCIA TPUBLICO [log]	0.027 *	0.013
N RESTAURANTES × DISTANCIA TPUBLICO [log]	0.011 *	0.005
N FARMACIAS × DISTANCIA TPUBLICO [log]	0.006	0.019
N SUPER × DISTANCIA TPUBLICO [log]	-0.001 *	0.001
HOMBRES × TIEMPO T AUTO [log]	-0.227	0.578
INDICE HOGAR ismt × TIEMPO T AUTO [log]	-0.106	0.641

PERSONAS × TIEMPO T AUTO [log]	-0.000	0.000
N BELLEZA × TIEMPO T AUTO [log]	0.021 ***	0.004
N CAFETERIA × TIEMPO T AUTO [log]	0.098 ***	0.021
N RESTAURANTES × TIEMPO T AUTO [log]	-0.004	0.008
N FARMACIAS × TIEMPO T AUTO [log]	0.095 ***	0.029
N SUPER × TIEMPO T AUTO [log]	0.004 ***	0.001
HOMBRES × N recorridos buses [log]	-0.514	0.319
INDICE HOGAR ismt × N recorridos buses [log]	2.571 ***	0.354
PERSONAS × N recorridos buses [log]	0.000	0.000
N BELLEZA × N recorridos buses [log]	-0.013 ***	0.002
N CAFETERIA × N recorridos buses [log]	0.033 **	0.011
N RESTAURANTES × N recorridos buses [log]	0.006	0.004
N FARMACIAS × N recorridos buses [log]	0.087 ***	0.014
N SUPER × N recorridos buses [log]	0.003 ***	0.000
HOMBRES × DIST METRO MALL [log]	-0.258 *	0.104
INDICE HOGAR ismt × DIST METRO MALL [log]	0.804 ***	0.115
PERSONAS × DIST METRO MALL [log]	-0.000 ***	0.000
N BELLEZA × DIST METRO MALL [log]	0.001	0.001
N CAFETERIA × DIST METRO MALL [log]	-0.026 ***	0.004
N RESTAURANTES × DIST METRO MALL [log]	-0.002	0.001
N FARMACIAS × DIST METRO MALL [log]	-0.014 **	0.005
N SUPER × DIST METRO MALL [log]	0.001 ***	0.000
Observations	286486	
R2 / R2 adjusted	0.344 / 0.344	
* p<0.05 ** p<0.01 *** p<0.001		

Tabla C.7: Resultados modelo 1.5

- Regresión Atributos Destino

Debido a que varias manzanas tienen cero tiendas y servicios en su área comercial como se refleja en la figura A.5, se realiza la transformación $\ln(1+NUM_TIENDA_SERVICIO)$. Además, para que mantener límites de probabilidad $p_{ij} \in (0, 1)$

$$\ln\left(\frac{P_{ij}}{(1-P_i)}\right) = \beta_0 + \beta_1 \ln(1+NUM_TIENDA_SERVICIO_{ij}) + \beta_2 \ln(A_{ijk}) \quad (C.5)$$

Para el modelo 1. se consideran variables de destino como número de tiendas, número de tiendas anclas y superficie.

Tabla C.8: Resultados modelo 2.1

log(PROBABILITY/(1-PROBABILIDAD))		
Predictors	Estimates	std. Error
(Intercept)	-15.264***	0.136
NUM TIENDAS [log]	0.573 ***	0.013
SUPERFICIE M2 [log]	0.395 ***	0.018
1 + NUM TIENDAS ANCLAS[log]	0.164 ***	0.021
Observations	347402	
R2 / R2 adjusted	0.071 / 0.071	

- Regresión Atributos Destino-Origen

Para los atributos de destino-origen, se añaden 2 versiones, dado que existen variables del atractivo de la zona de origen los cuales son obtenidas de distintas fuentes.

Tabla C.9: Resultados versiones modelo 2.2

log(PROBABILITY/(1-PROBABILIDAD))				
	Modelo 2.2-v1		Modelo 2.2-v2	
Predictors	Estimates	std. Error	Estimates	std. Error
(Intercept)	-17.333***	0.163	-17.300***	0.164
NUM TIENDAS [log]	0.671 ***	0.014	0.665 ***	0.014
SUPERFICIE M2 [log]	0.423 ***	0.020	0.448 ***	0.020
1 + NUM TIENDAS ANCLAS[log]	0.212 ***	0.022	0.173***	0.022
1 + N FARMACIA ZONA [log]	-0.052 ***	0.012		
1 + N SUPER ZONA [log]	0.017 ***	0.013		
1 + N BANCOS ZONA [log]	0.141 ***	0.008		
PERSONAS [log]	0.566 ***	0.007	0.244 ***	0.007
HOMBRES [log]	0.837 ***	0.061	0.338 ***	0.062
INDICE HOGAR ismt [log]	3.319 ***	0.116	1.185 ***	0.114
1 + N FARMACIAS [log]			0.135 ***	0.008
1 + N SUPER [log]			-0.059 ***	0.004
1 + N RESTAURANTES [log]			0.104 ***	0.006
1 + N CAFETERIA [log]			0.132 ***	0.009
1 + N BELLEZA [log]			-0.015 *	0.004
Observations	298771		298771	
R2 / R2 adjusted	0.117 / 0.117		0.118 / 0.118	

- Regresión Atributos Destino-Transporte

Se agregan variables de transporte de metro y bus, distancia y tiempo a los modelos (por separado), evitando las altas correlaciones entre tiempo y distancia de viaje del mismo medio de transporte.

Tabla C.10: Resultados versiones modelo 2.3

log(PROBABILITY)/(1-PROBABILITY)				
	Modelo 2.3-v1		Modelo 2.3-v2	
Predictors	Estimates	std. Error	Estimates	std. Error
(Intercept)	-12.370 ***	0.084	-15.370 ***	0.084
NUM TIENDAS [log]	0.204 ***	0.007	0.204 ***	0.007
SUPERFICIE M2 [log]	0.593 ***	0.007	0.593 ***	0.007
1 + NUM TIENDAS ANCLAS[log]	-0.397 ***	0.011	-0.397 ***	0.011
TIEMPO AUTO [log]	-1.323 ***	0.014		
TIEMPO T PUBLICO [log]			0.271 ***	0.013
DISTANCIA AUTO [log]			-1.263 ***	0.013
DISTANCIA T PUBLICO [log]	-1.160 ***	0.028		
1 + NUM RECORRIDOS BUS [log]	0.187 ***	0.010	0.135 ***	0.010
DIST MALL METRO [log]	-0.165 ***	0.003	-0.134 ***	0.003
Observations	332826		332826	
R2 / R2 adjusted	0.216 / 0.216		0.214 / 0.214	

- Regresión Atributos Destino-Origen-Transporte

Tabla C.11: Resultados versiones modelo 2.4

log(PROBABILITY / (1-PROBABILIDAD)				
Predictors	Modelo 2.4-v1		Modelo 2.4-v2	
	Estimates	std. Error	Estimates	std. Error
(Intercept)	-10.370 ***	0.081	-8.130 ***	0.090
NUM TIENDAS [log]	0.361 ***	0.007	0.379 ***	0.007
SUPERFICIE M2 [log]	0.581 ***	0.010	0.610 ***	0.010
1 + NUM TIENDAS ANCLAS[log]	-0.264 ***	0.011	-0.356 ***	0.012
1+ N FARMACIA[log]	0.102 ***	0.007	0.149 ***	0.007
1+ N SUPER[log]	-0.083 ***	0.003	-0.082 ***	0.003
1+ N RESTAURANTES[log]	-0.111 ***	0.004	-0.109 ***	0.004
1+ N CAFETERIA[log]	-0.006	0.007	-0.009	0.007
1+ N BELLEZA[log]	-0.079 ***	0.003	-0.058 ***	0.003
PERSONAS [log]	-0.193 ***	0.003	-0.197 ***	0.003
HOMBRES [log]	-0.228 ***	0.028	-0.209 ***	0.028
INDICE HOGAR ismt [log]	-3.297 ***	0.052	-3.806 ***	0.052
DISTANCIA AUTO [log]			-1.179 ***	0.008
TIEMPO AUTO [log]	-0.984 ***	0.013		
TIEMPO T PUBLICO [log]			-0.054 ***	0.012
DISTANCIA T PUBLICO [log]	-0.689 ***	0.009		
1 + NUM RECORRIDOS BUS [log]	0.162 ***	0.010	0.232 ***	0.010
DIST MALL METRO [log]	-0.121 ***	0.003	-0.128 ***	0.003
Observations	286432		286432	
R2 / R2 adjusted	0.282 / 0.282		0.280 / 0.280	
MAE test	0.062		0.056	
RSME test	0.138		0.119	
AIC test	299846		302000	

- Random Forest

Tabla C.12: Resultados versiones Random Forest

Atributos	Destino Modelo 3.1	Destino + origen Modelo 3.2	Destino+ transporte Modelo 3.3	Destino+origen+transporte Modelo 3.4
				SUPERFICIE_M2
		SUPERFICIE_M2		NUM_TIENDAS
		NUM_TIENDAS		NUM_TIENDAS_ANCLAS
		NUM_TIENDAS_ANCLAS	SUPERFICIE_M2	N_FARMACIAS
		N_FARMACIAS	NUM_TIENDAS	N_SUPER
Variables	SUPERFICIE_M2	N_SUPER	NUM_TIENDAS_ANCLAS	N_CAFETERIA
	NUM_TIENDAS	N_CAFETERIA	DISTANCIA_T.PUBLICO	N_RESTAURANTES
	NUM_TIENDAS_ANCLAS	N_RESTAURANTES	TIEMPO_AUTO	N_BELLEZA
		N_BELLEZA	NUM_RECORRIDOS_BUS	PERSONAS
		PERSONAS	DIST_MALL_METRO	HOMBRES
		HOMBRES		INDICE_HOGAR_ismt
		INDICE_HOGAR_ismt		DISTANCIA_T.PUBLICO
				TIEMPO_AUTO
				NUM_RECORRIDOS_BUS
				DIST_MALL_METRO
MAE test	0.037	0.030	0.026	0.025
MAPE test	144 %	142 %	146 %	159 %
RSME test	0.073	0.062	0.057	0.056
Corr test	0.263	0.640	0.645	0.710
R2 test	0.06	0.404	0.416	0.510

- Mejor Modelo - Modelo 3.4

En el contexto de un modelo de predicción, un MAPE del 159 % se consideraría bastante alto y sugiere que el modelo tiene un rendimiento deficiente en términos de precisión, ya que indica que las predicciones del modelo están alejadas del valor real en un 159 % en promedio, lo que indica una falta de ajuste preciso del modelo. Sin embargo, es importante considerar que el MAPE puede ser sensible a valores cercanos a cero o a predicciones muy bajas, si la probabilidad de asistencia es generalmente baja o cercana a cero, el MAPE podría aumentar incluso con pequeñas desviaciones, lo que sí ocurre con los datos reales, como se muestra en A.1. Como se añadieron 500.000 filas incluyendo aquellas filas con cero probabilidad, para poder aplicar MAE, no se dividió por el valor de probabilidad real, sino que por el estimado.

En el contexto de un modelo de predicción de probabilidad de asistencia a un centro comercial, un RMSE de 0.077 se consideraría bastante bajo y sugiere que el modelo tiene un buen ajuste y precisión en las predicciones.

Comuna	MAPE	MAE
PIRQUE	73 %	0,040
LO BARNECHEA	84 %	0,013
LAMPA	90 %	0,032
MACUL	94 %	0,021
LAS CONDES	95 %	0,013
CERRILLOS	96 %	0,018
PROVIDENCIA	97 %	0,016
ESTACION CENTRAL	97 %	0,024

SAN MIGUEL	97 %	0,019
SAN JOAQUIN	98 %	0,022
SANTIAGO	99 %	0,020
PADRE HURTADO	99 %	0,019
INDEPENDENCIA	99 %	0,019
NUNOA	99 %	0,018
QUINTA NORMAL	101 %	0,024
LA FLORIDA	102 %	0,021
MELIPILLA	102 %	0,032
ALHUE	102 %	0,073
PUENTE ALTO	103 %	0,022
LA CISTERNA	103 %	0,020
HUECHURABA	104 %	0,020
LA REINA	104 %	0,015
VITACURA	104 %	0,015
PAINE	106 %	0,028
CONCHALI	107 %	0,019
PENALOEN	108 %	0,021
RENCA	108 %	0,024
CERRO NAVIA	109 %	0,024
BUIN	109 %	0,026
LA GRANJA	110 %	0,021
PEDRO AGUIRRE CERDA	112 %	0,024
RECOLETA	112 %	0,021
MAIPU	112 %	0,017
QUILICURA	113 %	0,023
PUDAHUEL	114 %	0,022
LO ESPEJO	118 %	0,021
LO PRADO	119 %	0,027
LA PINTANA	120 %	0,027
SAN BERNARDO	120 %	0,025
EL BOSQUE	120 %	0,026
SAN RAMON	135 %	0,023
CURACAVI	139 %	0,032

Tabla C.13: Métricas de error de probabilidad por comuna

Centro Comercial	MAPE	MAE
Costanera Center	76 %	0,047

Mall Plaza Vespucio	77 %	0,046
Mall Plaza Oeste	82 %	0,042
Parque Arauco	89 %	0,034
Arauco Estación	89 %	0,041
Mall Arauco Maipú	90 %	0,035
Mall Plaza Egaña	90 %	0,032
Mall Paseo Quilín	92 %	0,030
Mall Florida Center	94 %	0,032
Alto Las Condes	96 %	0,023
Mall Plaza Norte	102 %	0,020
Mall Independencia	105 %	0,019
Mall Vivo Imperio	105 %	0,018
Mall Plaza Tobalaba	108 %	0,018
Mall Plaza Alameda	109 %	0,020
Espacio Urbano Puente Alto	111 %	0,019
Mall Vivo Panorámico	116 %	0,007
Mall Plaza Los Domínicos	116 %	0,013
Open Plaza Puente Alto	117 %	0,005
Mall Portal Ñuñoa	120 %	0,012
Mall Arauco Quilicura	122 %	0,003
Portal La Dehesa	123 %	0,007
Mall Espacio Urbano Las Rejas	123 %	0,009
Mall Vivo Los Trapenses	129 %	0,003
Casa Costanera	134 %	0,005
Espacio Urbano La Dehesa	137 %	0,001

Tabla C.14: Métricas de error de probabilidad por centro comercial

C.2. Aplicación de modelos

C.2.1. Estimación de demanda

Tabla C.15: Estimación de demanda Modelo A

Centro Comercial	Visitas Reales	Visitas Estimadas	Error %	Error
Mall Plaza Vespucio	625166	597216	4 %	27950
Costanera Center	580237	596039	3 %	-15802
Parque Arauco	498461	465464	7 %	32997
Mall Plaza Oeste	426036	481802	13 %	-55766
Mall Arauco Maipú	420266	465848	11 %	-45582
Mall Plaza Egaña	416900	432536	4 %	-15636
Mall Florida Center	379179	392831	4 %	-13652
Alto Las Condes	359506	369243	3 %	-9737
Mall Paseo Quilín	324144	354523	9 %	-30379
Arauco Estación	320852	439868	37 %	-119016
Mall Plaza Norte	285324	354903	24 %	-69579
Mall Plaza Tobalaba	269679	267869	1 %	1810
Espacio Urbano Puente Alto	269088	281615	5 %	-12527
Mall Independencia	222614	308947	39 %	-86333
Mall Plaza Los Dominicos	193400	239432	24 %	-46032
Mall Plaza Alameda	151062	249798	65 %	-98736
Mall Portal Ñuñoa	142419	222303	56 %	-79884
Mall Vivo Imperio	128457	228271	78 %	-99814
Portal La Dehesa	115340	182428	58 %	-67088
Casa Costanera	78672	156401	99 %	-77729
Mall Espacio Urbano Las Rejas	75219	162061	115 %	-86842
Mall Vivo Los Trapenses	51719	133967	159 %	-82248
Open Plaza Puente Alto	46442	128202	176 %	-81760
Mall Vivo Panorámico	45412	146187	222 %	-100775
Mall Arauco Quilicura	28506	132021	363 %	-103515
Espacio Urbano La Dehesa	20676	112577	444 %	-91901
Total Reales/ Total estimado/ MAPE/ MAE	6474776	7902352	78 %	59734,2308

Tabla C.17: Regresión Modelo A estimación de demanda

Predictors	UNIQUE_VISITS	
	Estimates	std. Error
(Intercept)	6.274 ***	0.252
ProbabilidadEstimada PERSONAS	1.758 ***	0.009
Observations	291789	
R2 / R2 adjusted	0.115 / 0.115	
* p<0.05 ** p<0.01 *** p<0.001		

Tabla C.16: Estimación de demanda Modelo B

Centro Comercial	Visitas Reales	Visitas Estimadas	Error %	Error
Mall Plaza Vespucio	625166	774701	24 %	-149535
Costanera Center	580237	698918	20 %	-118681
Parque Arauco	498461	611829	23 %	-113368
Mall Plaza Oeste	426036	523508	23 %	-97472
Mall Arauco Maipú	420266	516609	23 %	-96343
Mall Plaza Egaña	416900	506665	22 %	-89765
Mall Florida Center	379179	473470	25 %	-94291
Alto Las Condes	359506	440595	23 %	-81089
Mall Paseo Quilín	324144	394404	22 %	-70260
Arauco Estación	320852	378987	18 %	-58135
Mall Plaza Norte	285324	344829	21 %	-59505
Mall Plaza Tobalaba	269679	338621	26 %	-68942
Espacio Urbano Puente Alto	269088	337974	26 %	-68886
Mall Independencia	222614	263610	18 %	-40996
Mall Plaza Los Domínicos	193400	236130	22 %	-42730
Mall Plaza Alameda	151062	178283	18 %	-27221
Mall Portal Ñuñoa	142419	171548	20 %	-29129
Mall Vivo Imperio	128457	152224	19 %	-23767
Portal La Dehesa	115340	139608	21 %	-24268
Casa Costanera	78672	97037	23 %	-18365
Mall Espacio Urbano Las Rejas	75219	89854	19 %	-14635
Mall Vivo Los Trapenses	51719	63199	22 %	-11480
Open Plaza Puente Alto	46442	57566	24 %	-11124
Mall Vivo Panorámico	45412	51101	13 %	-5689
Mall Arauco Quilicura	28506	33880	19 %	-5374
Espacio Urbano La Dehesa	20676	25226	22 %	-4550
Total Reales/ Total estimado/ MAPE/ MAE	6474776	7900376	21 %	54830,7692

Tabla C.18: Regresión Modelo B estimación de demanda

Predictors	UNIQUE_VISITS	
	Estimates	std. Error
(Intercept)	11.146 ***	1.229
MALL NAME [Arauco Estación]	-8.232 ***	1.733
MALL NAME [Casa Costanera]	-8.421 ***	1.742
MALL NAME [Costanera Center]	2.404	1.726
MALL NAME [Espacio Urbano La Dehesa]	-10.359 ***	1.761
MALL NAME [Espacio Urbano Puente Alto]	-0.855	1.736
MALL NAME [Mall Arauco Maipú]	-1.039	1.734
MALL NAME [Mall Arauco Quilicura]	-11.059 ***	1.749
MALL NAME [Mall Espacio Urbano Las Rejas]	-9.142 ***	1.746
MALL NAME [Mall Florida Center]	0.222	1.729
MALL NAME [Mall Independencia]	-7.678 ***	1.734
MALL NAME [Mall Paseo Quilín]	-1.687	1.737
MALL NAME [Mall Plaza Alameda]	-9.064 ***	1.732
MALL NAME [Mall Plaza Egaña]	0.503	1.729
MALL NAME [Mall Plaza Los Domínicos]	-4.784 **	1.733
MALL NAME [Mall Plaza Norte]	-5.437 **	1.735
MALL NAME [Mall Plaza Oeste]	-1.960	1.735
MALL NAME [Mall Plaza Tobalaba]	-0.074	1.734
MALL NAME [Mall Plaza Vespucio]	6.505 ***	1.726
MALL NAME [Mall Portal Ñuñoa]	-7.730 ***	1.741
MALL NAME [Mall Vivo Imperio]	-9.590 ***	1.741
MALL NAME [Mall Vivo Los Trapenses]	-9.058 ***	1.739
MALL NAME [Mall Vivo Panorámico]	-10.669 ***	1.743
MALL NAME [Open Plaza Puente Alto]	-9.279 ***	1.759
MALL NAME [Parque Arauco]	4.872 **	1.725
MALL NAME [Portal LaDehesa]	-7.518 ***	1.739
ProbabilidadEstimada PERSONAS	1.711 ***	0.009
Observations	291789	
R2 / R2 adjusted	0.115 / 0.115	
* p<0.05 ** p<0.01 *** p<0.001		

La tabla con todos los datos:

Tabla C.19: Tabla resumen resultados completo

Atributos	Destino	Destino+origen	Destino-transporte	Destino+origen+transporte
Variables	SUPERFICIE_M2 NUM_TIENDAS NUM_TIENDAS_ANCLAS	SUPERFICIE_M2 NUM_TIENDAS NUM_TIENDAS_ANCLAS N_FARMACIAS N_SUPER N_CAFETERIA N_RESTAURANTES N_BELLEZA PERSONAS HOMBRES INDICE_HOGAR_ismt	SUPERFICIE_M2 NUM_TIENDAS NUM_TIENDAS_ANCLAS DISTANCIA_T.PUBLICO TIEMPO_AUTO NUM_RECORRIDOS_BUS DIST_MALL_METRO	SUPERFICIE_M2 NUM_TIENDAS NUM_TIENDAS_ANCLAS N_FARMACIAS N_SUPER N_CAFETERIA N_RESTAURANTES N_BELLEZA PERSONAS HOMBRES INDICE_HOGAR_ismt DISTANCIA_T.PUBLICO TIEMPO_AUTO NUM_RECORRIDOS_BUS DIST_MALL_METRO
MCI	Modelo 1.3	Modelo 1.4	Modelo 1.5	Modelo 1.6
MAE test	0.117	0.115	0.108	0.106
MAPE test	96 %	102 %	120 %	181 %
RSME test	0.151	0.156	0.190	0.197
AIC	1676983	1472753	1531768	1338035
R2 ajustado	0.094	0.122	0.297	0.344
RL	Modelo 2.1	Modelo 2.2-v2	Modelo 2.3	Modelo 2.4-v2
MAE test	0.027	0.032	0.027	0.030
MAPE test	758 %	1061 %	1468 %	1714 %
RSME	0.081	0.087	0.080	0.099
AIC	1811558	1562839	1680372	1440668
R2 ajustado	0.071	0.118	0.212	0.282
RF	Modelo 3.1	Modelo 3.2	Modelo 3.3	Modelo 3.4
MAE test	0.037	0.030	0.026	0.025
MAPE test	144 %	142 %	146 %	159 %
RSME test	0.073	0.062	0.057	0.056
R2	0.06	0.404	0.416	0.51
Corr	0.263	0.64	0.645	0.710

C.2.2. Estimación de ventas potenciales

Tabla C.20: Estructura de gasto de Hogar. Fuente: INE. Elaboración propia

Divisiones de Clasificación de Consumo Individual por Finalidades	Participación en el gasto
Alimentos y bebidas no alcohólicas	18,7 %
Transporte	15,2 %
Vivienda y gastos en servicios básicos	14,3 %
Bienes y servicios diversos	7,7 %
Salud	7,6 %
Recreación y cultura	6,8 %
Educación	6,5 %
Restaurantes y hoteles	6,5 %
Muebles y artículos para el hogar	6,3 %
Comunicaciones	5,3 %
Prendas de vestir y calzado	3,4 %
Bebidas alcohólicas, tabaco y estupefacientes	1,8 %