



Universidad del Desarrollo
Facultad de Ingeniería

HUMAN MOBILITY

XDR Movistar

POR: WLADIMIR NICOLÁS MATULIC LASTRA

Capstone Project presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Sra. LORETO BRAVO

Enero 2023
SANTIAGO

A Marianne por ser mi inspiración, a mis hijos por ser mi motivación, a mis padres y hermano por el constante empuje.

AGRADECIMIENTOS

Agradezco en primer lugar a todos los integrantes del cuerpo docente del Magister en Data Science por todos los conocimientos entregados durante el desarrollo del programa y en especial a la profesora Loreto Bravo quien guio el desarrollo y la culminación de mi Capstone Project.

También agradecer a SECTRA (Programa de Vialidad y Transporte Urbano) por su aporte y disposición al facilitar los datos del modelo Estraus vigente y depositar su confianza en este trabajo de investigación.

Por último, quiero agradecer a José Pedro Aldea por sus asesorías técnicas cuando aclaraba mis dudas e inquietudes sobre mi trabajo investigativo

TABLA DE CONTENIDO

RESUMEN	1
1. INTRODUCCIÓN	3
2. TRABAJOS RELACIONADOS	5
3. HIPÓTESIS Y OBJETIVOS	6
4. DATOS Y METODOLOGÍA	7
4.1. DATOS	7
4.1.1. MODELO DE TRANSPORTE DE PLANIFICACIÓN AÑO 2019 (ESTRAUS)	7
4.1.1.1. RESULTADOS PERÍODO PUNTA MAÑANA (07:00 – 09:00)	11
4.1.1.2. RESULTADOS PERÍODO FUERA DE PUNTA (10:00 – 11:00).....	12
4.1.1.3. RESULTADOS PERÍODO PUNTA TARDE (18:00 – 19:00)	13
4.1.1. DATOS EXTENDED DETAIL RECORDS (XDRs).....	14
4.2. METODOLOGÍA	16
4.2.1. RECOPIACIÓN DE LA INFORMACIÓN	17
4.2.2. LIMPIEZA Y PRE-PROCESAMIENTO	17
4.2.3. ANÁLISIS EXPLORATORIO	20
4.2.4. CONSTRUCCIÓN DE MATRICES	20
4.2.4.1. IDENTIFICACIÓN DE ESTADÍAS.....	22
4.2.4.2. IDENTIFICACIÓN DE VIAJES	24
4.2.5. GENERACIÓN DE INDICADORES DE DEMANDA DE TRANSPORTE	25
5. RESULTADOS	26
5.1. LIMPIEZA Y PREPROCESAMIENTO DE DATOS.....	26
5.2. CONSTRUCCIÓN DE MATRICES CON DATOS MÓVILES.....	44
5.3. VALIDACIÓN DE LOS RESULTADOS	52

6.	CONCLUSIONES	61
6.1.	VENTAJAS ENCONTRADAS EN EL DESARROLLO DEL CAPSTONE	61
6.2.	LIMITACIONES.....	62
6.3.	FUTURAS LÍNEAS DE TRABAJO.....	63
	BIBLIOGRAFÍA	65

Resumen

El Estado Chileno a través del Ministerio de Transporte y su Programa de Vialidad y Transporte Urbano: SECTRA, a contratado a través del tiempo a diferentes empresas de consultoría para desarrollar los estudios denominados “Actualización Diagnóstico del Sistema de Transporte Urbano”. Estos consisten en desarrollar encuestas de movilidad en una determinada área urbana, formular escenarios de usos de suelo, y finalmente, realizar la calibración de los modelos estratégico de transporte de la ciudad.

Las principales fuentes de información, que se recogen en terreno, para analizar y medir la movilidad del Sistema de Transporte Urbano son:

- Encuesta Origen Destino de Viajes en Hogares (EOD-H). Es una encuesta origen y destino de viajes realizada a nivel de hogares, con el propósito de recopilar información necesaria para caracterizar los viajes urbanos y sus requerimientos sobre el sistema de transporte de la ciudad.
- Conteos y Mediciones de Tasas de Ocupación a Usuarios. Es la medición de flujos vehiculares y conteos de tasa de ocupación a usuarios de transporte público (buses y taxis colectivos) y transporte privado (automóviles particulares y taxis básicos) en diferentes puntos del área de estudio. Estos conteos y mediciones se realizan con el propósito de validar la información de viajes obtenida de la EOD-H con la información de demanda obtenida en terreno.

- Mediciones de Tiempos de Viaje. Es la cuantificación de las condiciones de operación del sistema de transporte urbano, en particular los modos públicos y privados.

Esta información representa entre un 30% y un 60% del presupuesto total de los estudios, porque se requiere de mucho tiempo para recopilar y procesar toda la información, además es relevante porque *representa una observación estática de un fenómeno dinámico*.

Hoy en día, las nuevas tecnologías permiten recopilar esta información a un menor costo y analizar el fenómeno de manera dinámica, como son: las mediciones de flujo vehicular a través de grabaciones y la detección de objetos a través Computer Vision, las tasas de ocupación también utilizan el mismo principio y la EOD-H utilizando información de las conexiones de telefonía celular.

El presente estudio tiene por objetivo comparar las matrices de viajes obtenidas por conexiones móviles con los modelos de planificación en Transporte, dado que estos últimos manejan información más actualizada. Para ello se utiliza el tráfico de datos XDR que cubren la Región Metropolitana, y los datos del último modelo validado por SECTRA (año 2019).

En ambos casos se construyeron matrices de viajes a nivel de comunal representativas de los períodos que utilizan los modelos de planificación que son:

- Punta Mañana (07:00 – 09:00),
- Fuera de Punta (10:00 – 11:00) y
- Punta Tarde (18:00 – 19:00).

1. Introducción

Las Encuesta de viajes de los hogares (EOD-H) requieren un alto costo que involucra un diseño y control de muestreo cuidadoso, extensas entrevistas a los integrantes del hogar y una rigurosa limpieza y procesamiento de datos. A pesar de las mejoras con nuevos dispositivos electrónicos como tablets y aplicaciones que permiten un grado de autocompletado, el tiempo de finalización y eliminación de errores sigue siendo significativo. Además, hay que considerar el alto costo que generalmente cubren menos del 2% de la población de un área [10]. Una buena EOD-H genera información útil sobre los vehículos en el hogar, nivel de ingresos, las actividades y los patrones de viaje (viajes y recorridos) al menos para el día de la encuesta. Las EOD-H son una fuente clave de datos para estimar el modo y propósito del viaje [1].

Sin embargo, su pequeño tamaño de muestra no revela el amplio patrón de movimientos en forma de matrices de origen y destino, la pequeña muestra de viajes obtenida necesita ser aumentada a través de encuestas de interceptación [1].

Al final estos métodos convencionales de recopilación de datos pueden tomar una gran parte del presupuesto del estudio y una buena parte del tiempo disponible para completarlo. Los datos son caros y tienen una vida útil limitada, quedando obsoletos muy rápidamente. Los riesgos de sesgo y muestreo imperfecto son reales y requieren un tratamiento cuidadoso de control de calidad, detección y corrección.

El proceso de pasar de la recopilación de datos convencional al modelo válido no está exento de ajustes y eliminación de errores.

Durante los últimos años distintas tecnologías han apuntado a la recolección pasiva de datos, sin necesidad de interactuar con los usuarios. Esta característica genera nuevas posibilidades para recoger información de movilidad a un costo muy inferior al de los métodos tradicionales, lo que ha producido gran interés entre los modeladores de transporte. Entre las fuentes más interesantes son los datos de telefonía móvil, gracias a la posibilidad de obtener muestras de gran tamaño respecto de población, así como por proporcionar una resolución espacio-temporal suficientemente alta para la reconstrucción fiable de los patrones de movilidad de la población. En Chen et al. (2016) se presenta una extensa revisión de la literatura y se analiza la integración y posibles sinergias entre las disciplinas de análisis de datos y los estudios de transporte tradicionales con vistas a entender de forma detallada los patrones de movilidad de los individuos.

El análisis de los datos de telefonía móvil plantea varios desafíos, si se considera como referencia el modelo de transporte de cuatro etapas (Ortúzar y Willumsen, 2011), existen estudios que abordan el uso de datos de telefonía móvil para encontrar información acerca de las etapas de generación/atracción y de distribución de viajes. Sin embargo, muchos de ellos son abordados con los Call Detail Records (CDR) y validados con EOD-H extemporaneas.

Este estudio plantea el uso de los eXtended Detail Records (XDR) y modelos de transporte validados por SECTRA. Cabe señalar, en estudios recientes de modelos de inferencia de localización de vivienda, muestran mejores resultados al ocupar XDR en desmedro de CDR [9].

2. Trabajos Relacionados

La estimación de la matriz de viajes (OD), como se comentó anteriormente es producto de las Encuestas de Hogares y las de interceptación (Transporte Privado y Transporte Público) que permiten construir los modelos de Generación-Atracción, Distribución y Partición Modal de viajes (Actualización Metodología Análisis Sistema de Transporte de Ciudades de Gran Tamaño y Tamaño Medio “MESPE” 2008). También se han desarrollado otros métodos para la predicción de matrices OD mediante las tarjetas de pago para el transporte público (Tarjetas BIP), el análisis de transiciones de conexiones entre antenas celulares, que son la base de muchos métodos, que emplean técnicas que van desde la optimización y las reglas de asociación temporal hasta las reglas basadas en el transporte para aceptar o descartar transiciones. El enfoque desarrollado en este trabajo, también se basa en transiciones incorporando el concepto de estadía (tiempo de permanencia en un mismo lugar) que se transforma en el origen y/o destino del viaje. Es un enfoque busca determinar los viajes entre estadías y compararlas con el modelo de transporte.

Los “puntos de estadía” del viaje diario, permiten inferir lugares de residencia y/o de trabajo, dependiendo del horario en que se encuentran. Además, si se seleccionan los horarios en que desarrollan los modelos de transporte se pueden comparar con la predicción de ellos y verificar si las estructuras son similares.

3. Hipótesis y Objetivos

Hipótesis: Los eXtended Detail Records (XDRs) al ser más rápidos de obtener y poseer gran tamaño de información, permite analizar la variabilidad de los viajes y así construir matrices de viajes tan precisas como las calculadas y validadas en los modelos de planificación.

Objetivos:

1. Construir matrices de viajes a partir de los datos de telefonía móvil XDR.
2. Comparar los resultados de la telefonía móvil con las matrices de viajes empleadas en los modelos de planificación de transporte.
3. Incorporar cambios metodológicos en la construcción de los modelos de planificación, incluyendo el aporte de la información que entregan los teléfonos móviles y ayudar a corregirlos en caso de detectar importantes diferencias.

4. Datos y Metodología

4.1. Datos

Para este estudio se consideraron dos fuentes de datos, por un lado están los resultados del modelo vigente por SECTRA para la ciudad de Santiago año 2019 y los datos de telefonía del día laboral 03-03-2022 de la empresa Telefónica S.A.

Es importante señalar que los modelos de planificación en Transportes utilizan modelos que simulan un día Laboral normal (entre marzo y diciembre, para los días lunes, martes, miércoles, jueves y viernes) con niños y jóvenes yendo a clases, además siempre en los estudios de base se recomienda medir y observar entre martes y jueves. Por esta razón al disponer de datos de telefonía celular entre el martes 1ro de marzo y el jueves 3 de marzo, además se analizó que el inicio de clases para el año 2022 partió el 2 de marzo, el mejor conjunto de datos a utilizar en el Capstone Project es el 3 de marzo.

4.1.1. Modelo de Transporte de planificación año 2019 (Estraus)

ESTRAUS es un modelo de simulación del equilibrio entre oferta y demanda en el mercado de transporte urbano. Desarrollado por el Estado de Chile a partir de los años 80 y es una herramienta fundamental para la evaluación de planes y proyectos de transporte urbano en ciudades de gran tamaño.

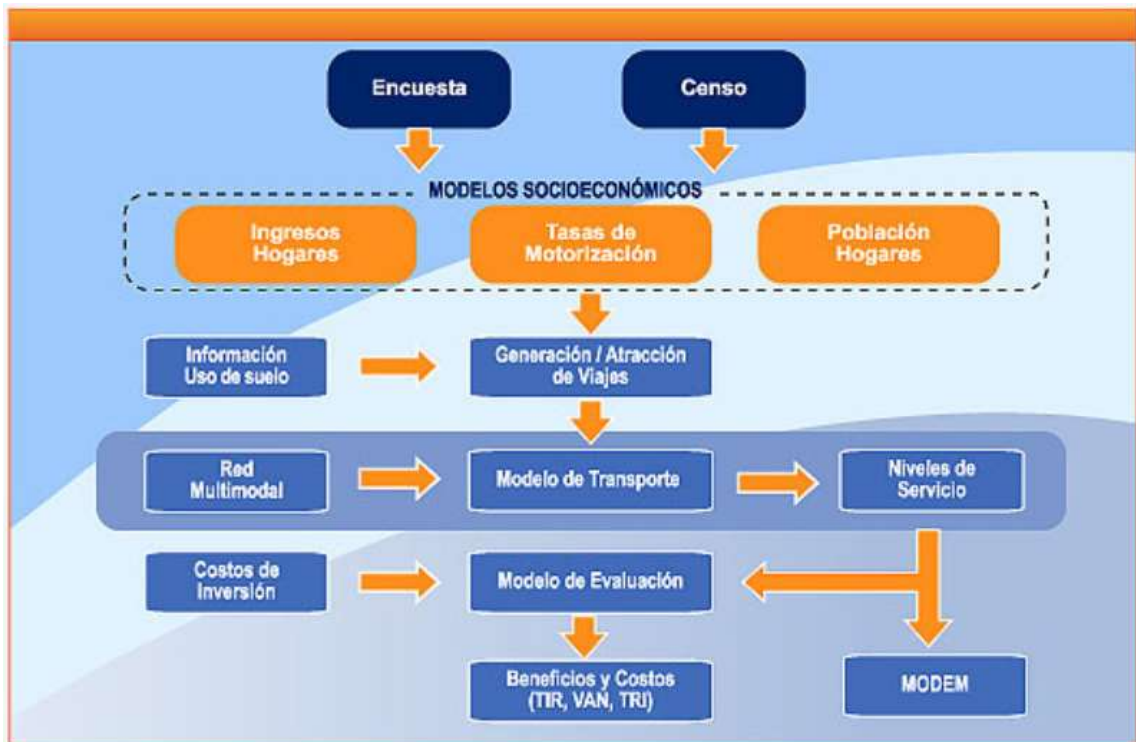
Diseñado inicialmente para simular la operación del sistema de transporte urbano (STU) de Santiago, su desarrollo incorporó las particularidades relevantes del sistema de transporte de esta ciudad: gran diversidad socioeconómica de los usuarios y viajes realizados mayoritariamente en transporte público. En la actualidad, sus fundamentos forman parte del cuerpo teórico más avanzado a nivel mundial en el área de modelación de redes de transporte.

Con él es posible evaluar los cambios en la operación del STU ocasionados por la implementación de planes y políticas de transporte, constituyéndose en la principal herramienta de apoyo a la toma de decisiones estratégicas por parte del Estado Chileno.

Los planes que permite evaluar son conjuntos de proyectos tales como: líneas de metro, autopistas, concesiones viales, ampliaciones de ejes viales, corredores de buses, políticas de tarificación vial, integración tarifaria en transporte público, aumento del precio de los combustibles, etc.

La principal ventaja respecto a otros modelos, es que permite predecir un equilibrio simultáneo entre oferta y demanda en un sistema de transporte. De esta forma, los resultados de las etapas de distribución de viajes, partición modal y asignación de tráfico del modelo clásico de transporte, son consistentes entre sí e internalizan los efectos de la congestión que se experimenta en las diferentes redes. Es decir, se considera una demanda elástica, en que el número de viajes entre pares origen-destino depende de los niveles de servicio existentes en las redes de los distintos modos de transporte que unen cada par.

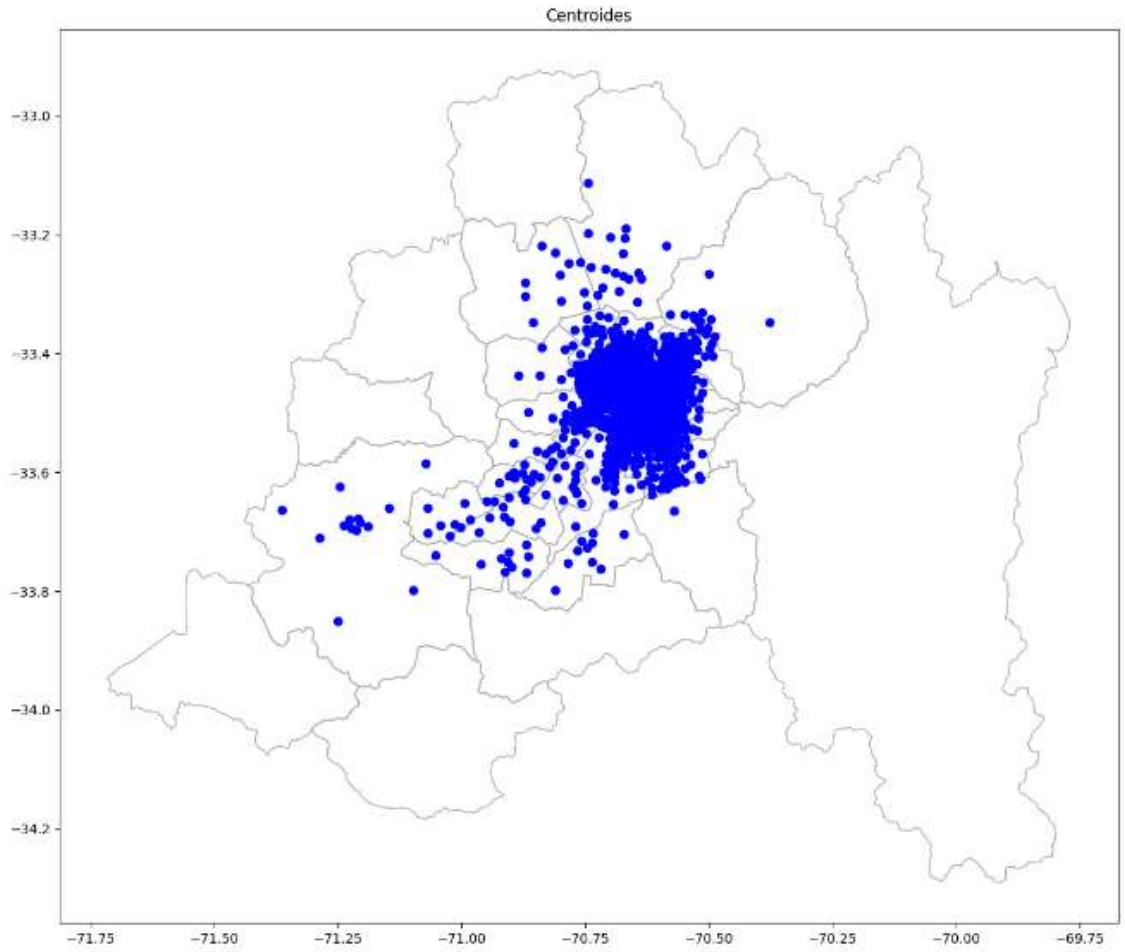
La siguiente figura muestra cómo se enmarca el modelo ESTRAUS en el contexto de la planificación estratégica del sistema de transporte de una ciudad.



Planificación estratégica

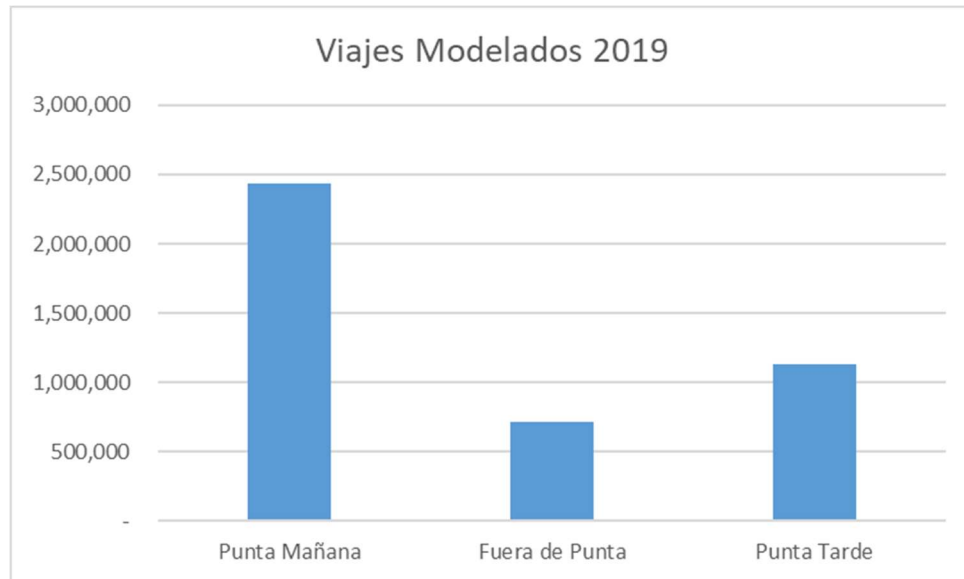
La información que entrega el modelo son matrices por modo de viaje (auto chofer, auto acompañante, caminata, bicicleta, bus, metro-bus, metro-taxicolectivo, metro y taxicolectivo), categoría de usuario (combinación entre nivel de ingreso y tasa de motorización) y horario, además está organizada de la siguiente manera:

- zona Estras de origen.
- Zona Estras de destino.
- N° de viajes



Ubicación espacial de los Centroides de las zonas modelo “Etraus”

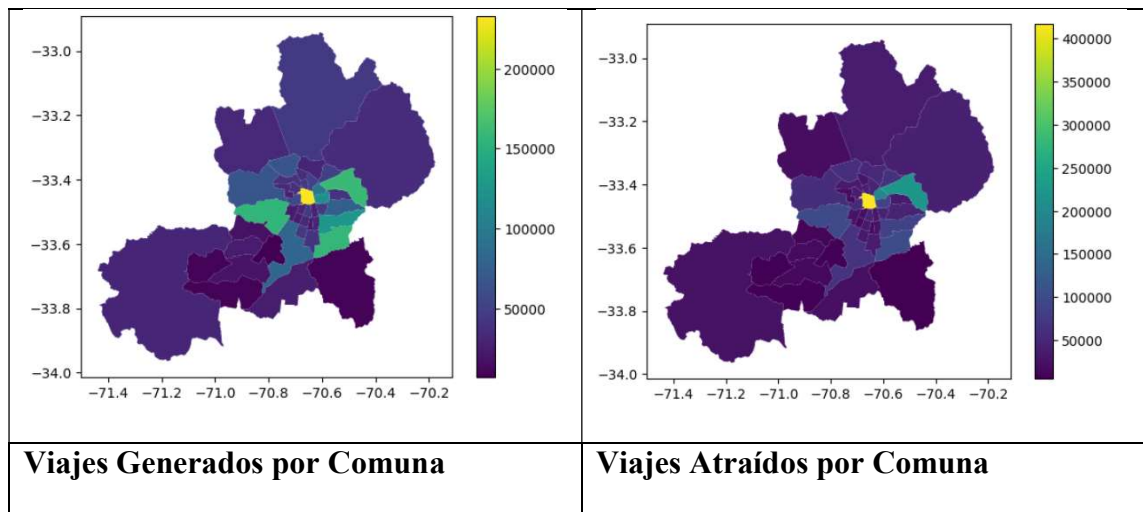
Finalmente, se sumaron todos los viajes (por modo y horario) para determinar a nivel global el período punta mañana con 2,431,413 viajes, 713,724 viajes en el fuera de punta y de 1,127,676 viajes para el punta tarde.



Viajes Modelados para el año 2019

4.1.1.1. Resultados período punta mañana (07:00 – 09:00)

A continuación, se presentan los resultados del período punta mañana a nivel de viajes originados y atraídos por comuna.

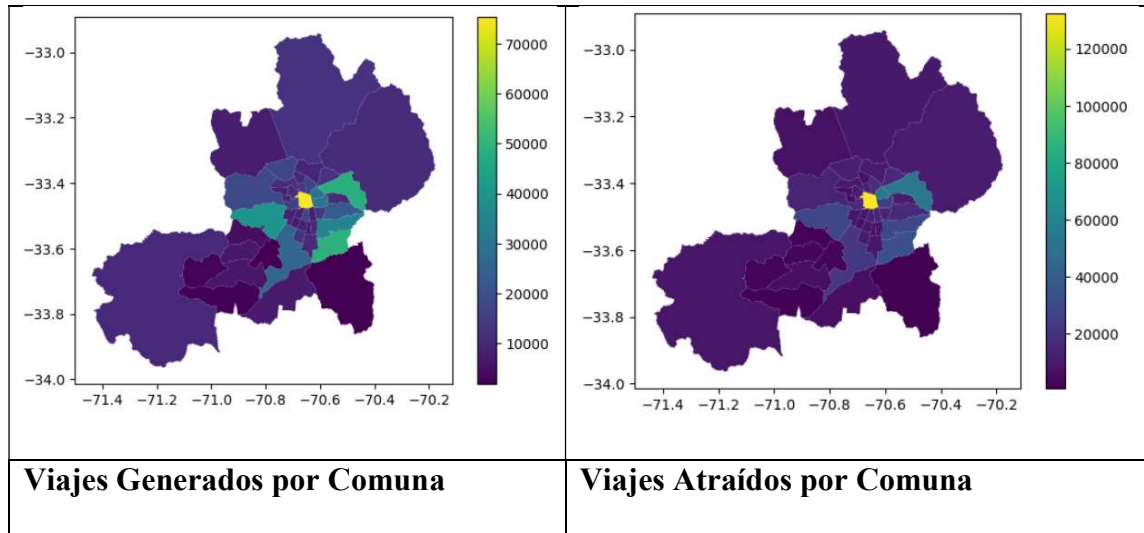


De las figuras anteriores se observa que las comunas con más viajes generados son: Santiago, Las Condes, Puente Alto, Maipú y La Florida, esto tiene relación con las comunas con mayor cantidad de habitantes donde se originan los viajes y de acuerdo al censo 2017 las comunas con mayor población: Puente Alto, Maipú, Santiago, La Florida, San Bernardo y Las Condes.

Para los viajes atraídos las comunas con mayor cantidad de viajes son: Santiago, Las Condes, Providencia, Puente Alto y Maipú, esto tiene relación con las comunas donde se realizan actividades como trabajo, estudio y otros. Las comunas de Santiago, Las Condes y Providencia son polos de actividades muy importantes como servicios, comercio y educación.

4.1.1.2. Resultados período fuera de punta (10:00 – 11:00)

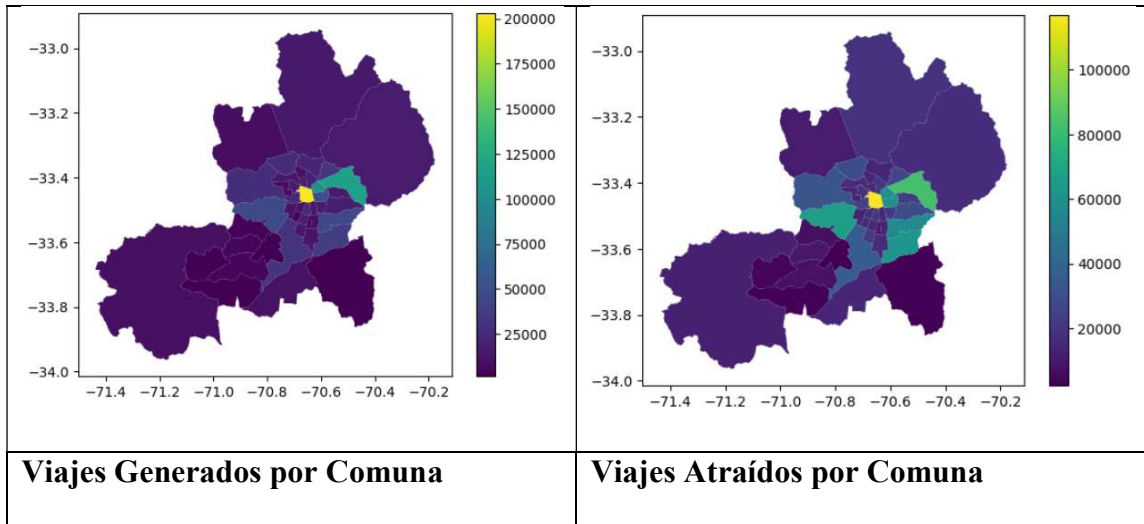
A continuación, se presentan los resultados del período fuera de punta a nivel de viajes originados y atraídos por comuna.



De las figuras se observa que los patrones se repiten al anterior, siendo las comunas con más viajes generados son: Santiago, Puente Alto, Las Condes, Maipú y La Florida. Para los viajes atraídos las comunas con mayor cantidad de viajes son: Santiago, Puente Alto, Las Condes, Maipú y La Florida.

4.1.1.3. Resultados período punta tarde (18:00 – 19:00)

A continuación, se presentan los resultados del período punta tarde a nivel de viajes originados y atraídos por comuna.



De las figuras se observa que las comunas con más viajes generados son: Santiago, Las Condes, Providencia, Ñuñoa y Maipú. Para los viajes atraídos las comunas con mayor cantidad de viajes son: Santiago, Las Condes, Providencia, Ñuñoa y Maipú.

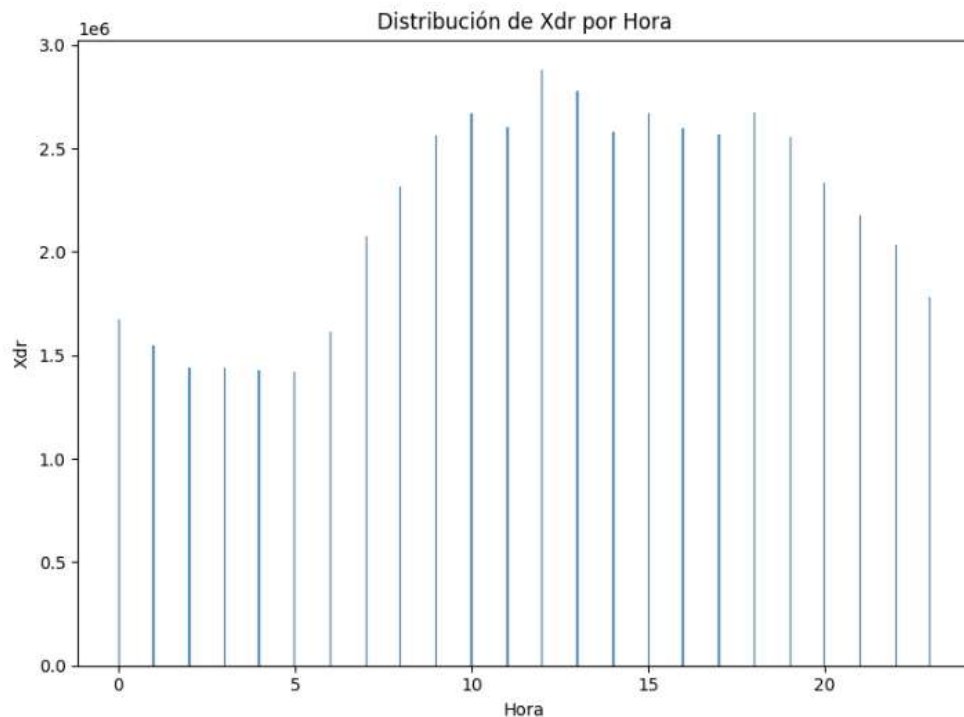
4.1.1. Datos eXtended Detail Records (XDRs)

La firma Telefónica S.A, según el Análisis de Tráfico Internet Enero-Abril 2020 de la Subsecretaría de Telecomunicaciones, tiene el 24,6% de participación en el mercado de telefonía móvil [14], proporcionó archivos de carácter privado en formato “parquet”, es decir, de valores separados por comas. El archivo a analizar es el “capstone_part000000000XXX.parquet”, el cual contiene registros constituidos por XDRs (eXtended Detail Records), que proporcionan información de un usuario cada vez que éste interactúa con la red a través de la carga y descarga de archivos.

Esta información se pre-procesó a fin de obtener los datos organizados para la fecha de 3 de marzo de 2022, con las siguientes columnas:

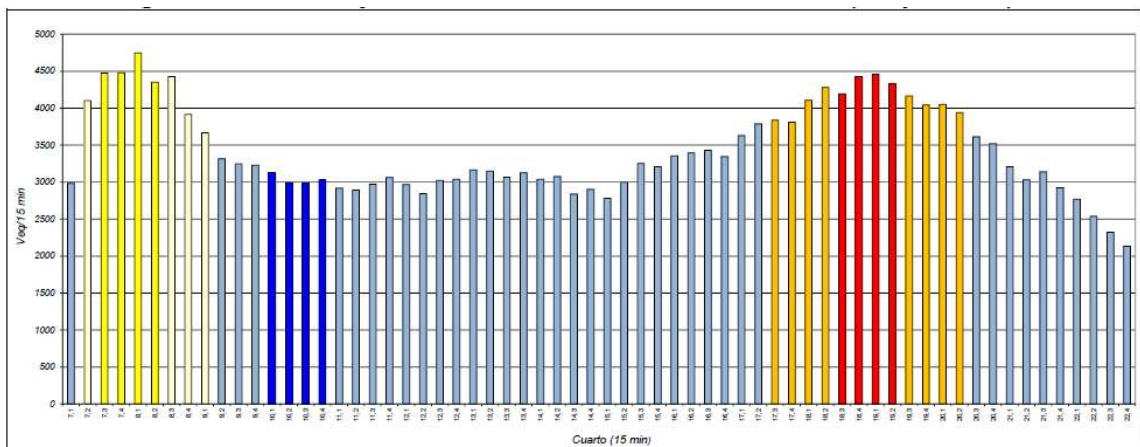
- device: dato de tipo string que corresponde al identificador del equipo móvil, codificado en una combinación alfanumérica. Cabe destacar que esta codificación resulta de un proceso de anonimización, por lo cual no es posible individualizar personas ni números telefónicos.
- datetime: dato de tipo timestamp que corresponde a la fecha y hora en que se realiza la transacción entre el dispositivo móvil y la torre.
- lat: corresponde a la latitud en que se ubica la torre, es un dato de tipo float redondeado a la diezmilésima.
- lon: corresponde a la longitud en que se ubica la torre, es un dato de tipo float redondeado a la diezmilésima.

A continuación, se presenta un perfil con las conexiones a lo largo del día.



Distribución horaria de las conexiones

El perfil muestra que las conexiones más altas se producen entre los horarios 09:00 y 19:00 horas. Cabe señalar, que este perfil no se asemeja a los tradicionales perfiles de viajes, donde se observan marcadas diferencias en la mañana (entre 07:00 y 09:00) y en la tarde (entre 17:00-20:00), sin embargo, se analizarán los horarios específicos para validar nuestra hipótesis.



Perfil de carga para comunas del sector sur de Santiago, año 2017

4.2. Metodología

La secuencia completa de los subprocesos para la obtención de las matrices origen-destino es la siguiente:

- i. Recopilación de la información
- ii. Limpieza y pre-procesamiento
- iii. Análisis Exploratorio
- iv. Construcción de Matrices Origen-Destino

v. Cálculo de Indicadores

Cada uno de estos subprocesos, así como los algoritmos asociados, se describen a continuación.

4.2.1. Recopilación de la información

La data que presenta los resultados del modelo “Estraus” fue proporcionada por SECTRA y la información ya se encontraba organizada en archivos dat con valores separados por espacio, lo único necesario fue sumar los modos y horarios para obtener los viajes por período y luego asignar a cada zona su comuna correspondiente, para llevarlos a esa unidad espacial.

De manera similar, los datos localización, fueron proporcionados por la empresa Telefónica S.A. a través del Instituto de Data Science de la Universidad del Desarrollo y estaban organizados en archivos “capstone_part000000000XXX.parquet”.

4.2.2. Limpieza y pre-procesamiento

El proceso de limpieza y pre-procesamiento se realizó para ambas bases de datos, la primera referida al modelo “Estraus” solo consistió en asociar a cada zona de origen y destino la comuna correspondiente en que se ubica, como se muestra en la siguiente tabla.

O	D	AM	FP	PT	Origen	Destino
0	1	1	0.000	0.000	0.000	Santiago Santiago
1	1	2	9.248	6.827	4.818	Santiago Santiago
2	1	3	27.751	28.552	10.898	Santiago Santiago
3	1	4	31.921	3.944	3.281	Santiago Santiago
4	1	5	32.833	7.737	5.155	Santiago Santiago
5	1	6	17.795	6.708	2.115	Santiago Santiago
6	1	7	46.648	10.278	4.283	Santiago Santiago
7	1	8	109.303	28.669	15.029	Santiago Santiago
8	1	9	124.938	31.339	13.436	Santiago Santiago
9	1	10	75.838	19.698	19.126	Santiago Santiago

Matriz origen-destino modelo Etraus

Para los datos móviles XDR el proceso fue más extenso, primero seleccionando la fecha jueves 03-03-2022, que representa un día laboral y representa un día de normal de desplazamientos en Santiago. Luego, se eliminaron los registros que no contenían información y los que estaban repetidos, finalmente de acuerdo a su ubicación espacial los que se encuentran dentro de la región metropolitana. Respecto a posibles errores de posicionamiento, se chequearon y no se encontraron.

Por otra parte, se filtró la data que su hora media del viaje estuviese contenida dentro de los horarios que utiliza el modelo Etraus, para ello se seleccionaron los siguientes horarios:

- En el horario punta mañana desde las 06:00 hasta las 10:00,
- Fuera de punta desde las 09:00-12:00 y
- En el punta tarde desde las 17:00 has las 20:00 horas,

Esto da una holgura para que viajes que hayan partido antes o durante el período, pero su hora media esté contenida dentro del período a analizar de la muestra.

Respecto de la metodología para considerar un viaje dentro de la muestra a partir de los dispositivos, consistió en lo siguiente:

- a) Ordenar los registros por dispositivo-día ordenados por hora.
- b) Eliminar a los usuarios que no presentaban movimientos a lo largo del día.
- c) Eliminar usuarios que registraban distancias recorridas sobre los 70 [km][15], puesto que estos usuarios podrían ser choferes de aplicaciones de UBER, CABIFY, taxis, etc...
- d) Para considerar un viaje dentro de la muestra deben existir como mínimo dos estadías, que representan la permanencia de un dispositivo en un mismo punto.

Esta estrategia busca proporcionar la mayor información posible para inferir los viajes diarios de manera fiable, evitando la inclusión de usuarios que realicen actividades y viajes imposibles de detectar que puedan por tanto afectar a la calidad de las matrices origen-destino. Es importante comentar también que existen usuarios que desaparecen temporalmente de la red como consecuencia, por ejemplo, la descarga del celular o pérdida de señal o un viaje fuera de la región, que alteren la construcción de la matriz con viajes poco habituales. Esta consideración representa un aspecto habitualmente ignorado en los estudios disponibles en la literatura y es muy difícil de comprobar.

4.2.3. Análisis Exploratorio

Finalizada la etapa anterior, se continuó con el análisis exploratorio. El objetivo de esta etapa fue identificar comportamientos de las variables, a través de tablas y visualizaciones como gráficos de barras, histogramas, entre otros. Este análisis permitió validar el comportamiento de los datos con los patrones de movilidad, mostrando distribuciones a lo largo de las horas del día, comunas con mayor cantidad de conexiones, distancias de viaje, distribuciones de viajes en los períodos de análisis.

Con el número de conexiones, se procedió a calcular los viajes totales a través de la representatividad de la muestra de estudio. Para ello se utilizó los datos de porcentaje de mercado de la empresa Telefónica que tiene el 24,6% de participación en telefonía móvil [14] y además suponiendo que toda persona tiene un celular, el cálculo de tamaño de población queda de la siguiente manera:

$$\text{Viajes Totales} = \frac{\text{Viajes de la muestra}}{\text{Participación de mercado}}$$

4.2.4. Construcción de Matrices

Cada dispositivo presenta una secuencia de registros del siguiente tipo:

	device	datetime	lat	lon	geometry	hh
0	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 06:22:10	-33.40226	-70.5778	POINT (-70.5778 -33.40226)	6
1	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 06:31:43	-33.40226	-70.5778	POINT (-70.5778 -33.40226)	6
2	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 06:52:10	-33.40226	-70.5778	POINT (-70.5778 -33.40226)	6
3	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 07:22:10	-33.40226	-70.5778	POINT (-70.5778 -33.40226)	7
4	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 07:31:43	-33.40226	-70.5778	POINT (-70.5778 -33.40226)	7

Registros de un dispositivo

Además, si uno visualiza esta información en un Sistema de Información Geográfica (SIG) puede observar un diagrama de desplazamiento del siguiente tipo.



Diagrama de desplazamiento

El proceso comienza con la construcción para cada dispositivo de los desplazamientos entre dos puntos consecutivos y determinar si existe o no permanencia ($\text{tiempo} > 0$ y $\text{dist} = 0$).

	device	datetime	geometry	dest_datetime	dest_geometry	tiempo	dist
0	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 06:22:10	POINT (353273.603 -3696995.319)	2022-03-03 06:31:43	POINT (353273.603 -3696995.319)	0 days 00:09:33	0.0
1	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 06:31:43	POINT (353273.603 -3696995.319)	2022-03-03 06:52:10	POINT (353273.603 -3696995.319)	0 days 00:20:27	0.0
2	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 06:52:10	POINT (353273.603 -3696995.319)	2022-03-03 07:22:10	POINT (353273.603 -3696995.319)	0 days 00:30:00	0.0
3	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 07:22:10	POINT (353273.603 -3696995.319)	2022-03-03 07:31:43	POINT (353273.603 -3696995.319)	0 days 00:09:33	0.0
4	00000000e3fc09803d408609f55cbf905120d3a41db564...	2022-03-03 07:31:43	POINT (353273.603 -3696995.319)	2022-03-03 07:46:43	POINT (353273.603 -3696995.319)	0 days 00:15:00	0.0

**Construcción de las variables tiempo (Tiempo final – Tiempo Inicial) y dist
(Posición final – Posición Inicial)**

El objetivo es generar para cada usuario la ubicación sus estadías y lugares de tránsito, para los modelos de planificación lo principal es la estadía que representa la actividad que una persona realiza en un lugar determinado y razón por cual este se desplaza, transformando esto en un diario de viajes.

La información asociada a cada viaje incluye:

- 1) Origen,
- 2) Destino,
- 3) Hora de inicio
- 4) Hora de término
- 5) Posición inicial
- 6) Posición final
- 7) Tiempo
- 8) Distancia recorrida.

Con el fin de describir en detalle los algoritmos utilizados para inferir esta información, es útil dividir dichos algoritmos en los siguientes grupos:

- 1) Identificación de estadías
- 2) Identificación de viajes.

4.2.4.1. Identificación de Estadías

El primer paso consiste en identificar las estadías de un usuario, es decir, su permanencia durante un tiempo en un lugar determinado. Para ello, se analizan los registros

consecutivos en una misma celda de telefonía y se identifican como estadías aquellas localizaciones en las que el individuo es detectado durante un tiempo superior a 0 minuto. A continuación, se muestran en este diagrama:

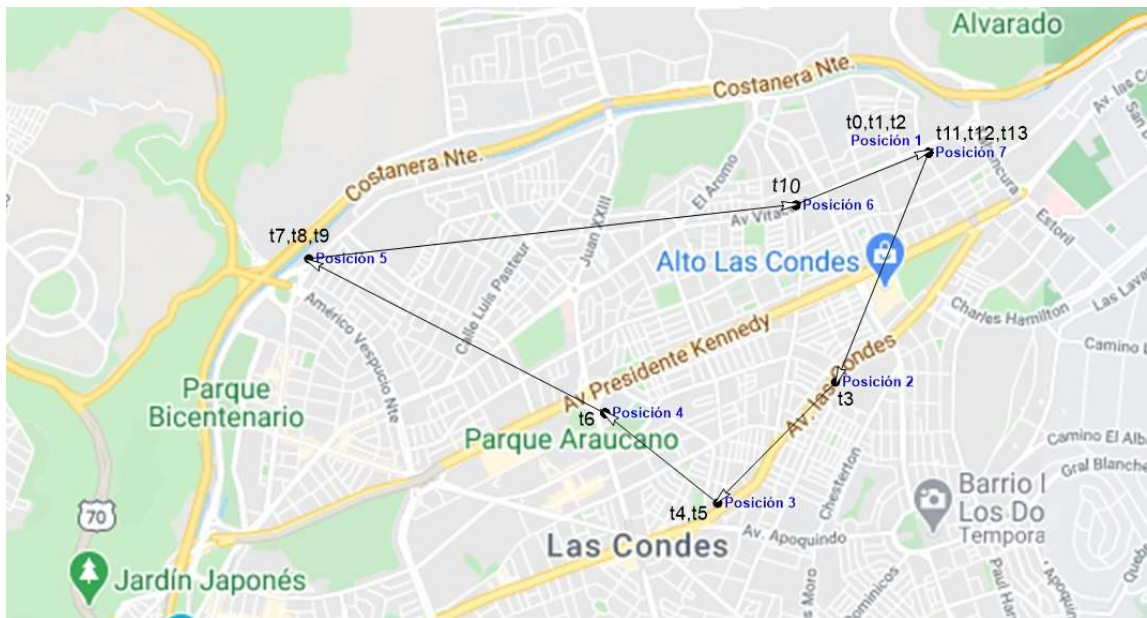


Diagrama de desplazamiento

En el diagrama se observa que las estadías se producen en la posición 1, posición 3, posición 5 y posición 7.

Se escogió la condición de estadía mayor que cero, en función del análisis a distintas EOD-H analizados en diferentes estudios, donde varios padres o madres van al colegio a dejar y buscar a sus hijos al colegio, donde la permanencia observada en cursos mayores es muy baja, otro ejemplo que justifica esta medida es el cambio modal que experimentan los usuarios al usar el metro (tiempo de espera baja).

4.2.4.2. Identificación de viajes

A partir de las estadías identificadas, se identifican los viajes realizados a partir de al menos dos estadías encontradas. A modo de ejemplo se presenta la siguiente figura.

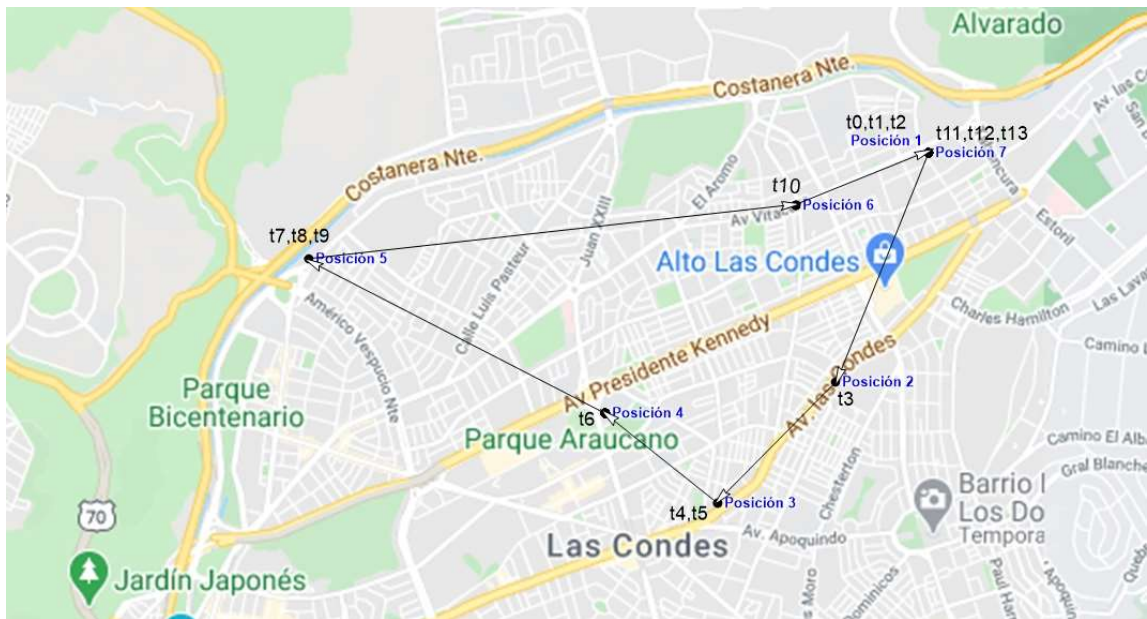


Diagrama de desplazamiento

Como se definió anteriormente, las estadías estaban en las posiciones 1, 3, 5 y 7, los viajes son Viaje N°1 {posición 1, posición 3}, Viaje N°2 {posición 3, posición 5} y Viaje N°3 {posición 5, posición 7}.

A cada viaje realizado se le determinó la hora de inicio, de finalización y una hora intermedia (permitirá asociar a los períodos de modelación, esto se realiza para asignar los viajes en las EOD-H). A continuación, se muestra una tabla con los resultados:

	device	ori_datetime	ori_geometry	dest_datetime	dest_geometry	tiempo	dist	tiempo_medio	hh
	00003946e20b7e48b4fb71b336dec0d48541b127a107f3...	2022-03-03 08:00:25	POINT (353340.952 -3707856.857)	2022-03-03 09:07:45	POINT (349883.715 -3701307.439)	0 days 00:33:40	7405.900882	2022-03-03 08:34:05.000	8
	00004acb9685adc7269de220981632fabf0d031a688b48...	2022-03-03 08:27:32	POINT (341654.139 -3713975.893)	2022-03-03 08:31:58	POINT (340053.231 -3714265.258)	0 days 00:02:13	1626.849024	2022-03-03 08:29:45.000	8
	00004acb9685adc7269de220981632fabf0d031a688b48...	2022-03-03 09:01:58	POINT (340053.231 -3714265.258)	2022-03-03 09:16:58	POINT (341654.139 -3713975.893)	0 days 00:07:30	1626.849024	2022-03-03 09:09:28.000	9
	000077095e0665725045e995b5b69602544b3f57db40d7...	2022-03-03 07:05:47	POINT (345468.152 -3702088.017)	2022-03-03 08:18:12	POINT (334048.203 -3700274.217)	0 days 00:36:12.500000	11563.092740	2022-03-03 07:41:59.500	7
	000081454861b0e26886ec9e51a2a719065c14f0ea7a8d...	2022-03-03 06:52:03	POINT (345846.188 -3716292.522)	2022-03-03 07:55:28	POINT (348128.660 -3700749.245)	0 days 00:31:42.500000	15709.969647	2022-03-03 07:23:45.500	7

Construcción de las variables tiempo_medio (Tiempo final + $\Delta T/2$) y hh (Hora de tiempo medio)

Luego en función de los horarios de los períodos se filtran los viajes que están fuera de rango y en función de las coordenadas se asocian las comunas.

	device	ori_datetime	dest_datetime	hh	geometry_ori	geometry_dest	Comuna_ori	Comuna_dest
0	00003946e20b7e48b4fb71b336dec0d48541b127a107f3...	2022-03-03 08:00:25	2022-03-03 09:07:45	8	POINT (-70.57885 -33.50020)	POINT (-70.61496 -33.44067)	Peñalolén	Providencia
1	00004acb9685adc7269de220981632fabf0d031a688b48...	2022-03-03 08:27:32	2022-03-03 08:31:58	8	POINT (-70.70570 -33.55370)	POINT (-70.72299 -33.55607)	San Bernardo	San Bernardo
2	000077095e0665725045e995b5b69602544b3f57db40d7...	2022-03-03 07:05:47	2022-03-03 08:18:12	7	POINT (-70.66258 -33.44708)	POINT (-70.78506 -33.42902)	Santiago	Pudahuel
3	000081454861b0e26886ec9e51a2a719065c14f0ea7a8d...	2022-03-03 06:52:03	2022-03-03 07:55:28	7	POINT (-70.66096 -33.57520)	POINT (-70.63374 -33.43539)	El Bosque	Providencia
4	0000cc32c34abb6cf8f85921540c0d41784f3854691a8f...	2022-03-03 08:03:48	2022-03-03 08:33:48	8	POINT (-70.55736 -33.43830)	POINT (-70.55528 -33.43028)	La Reina	Las Condes

4.2.5. Generación de indicadores de demanda de transporte

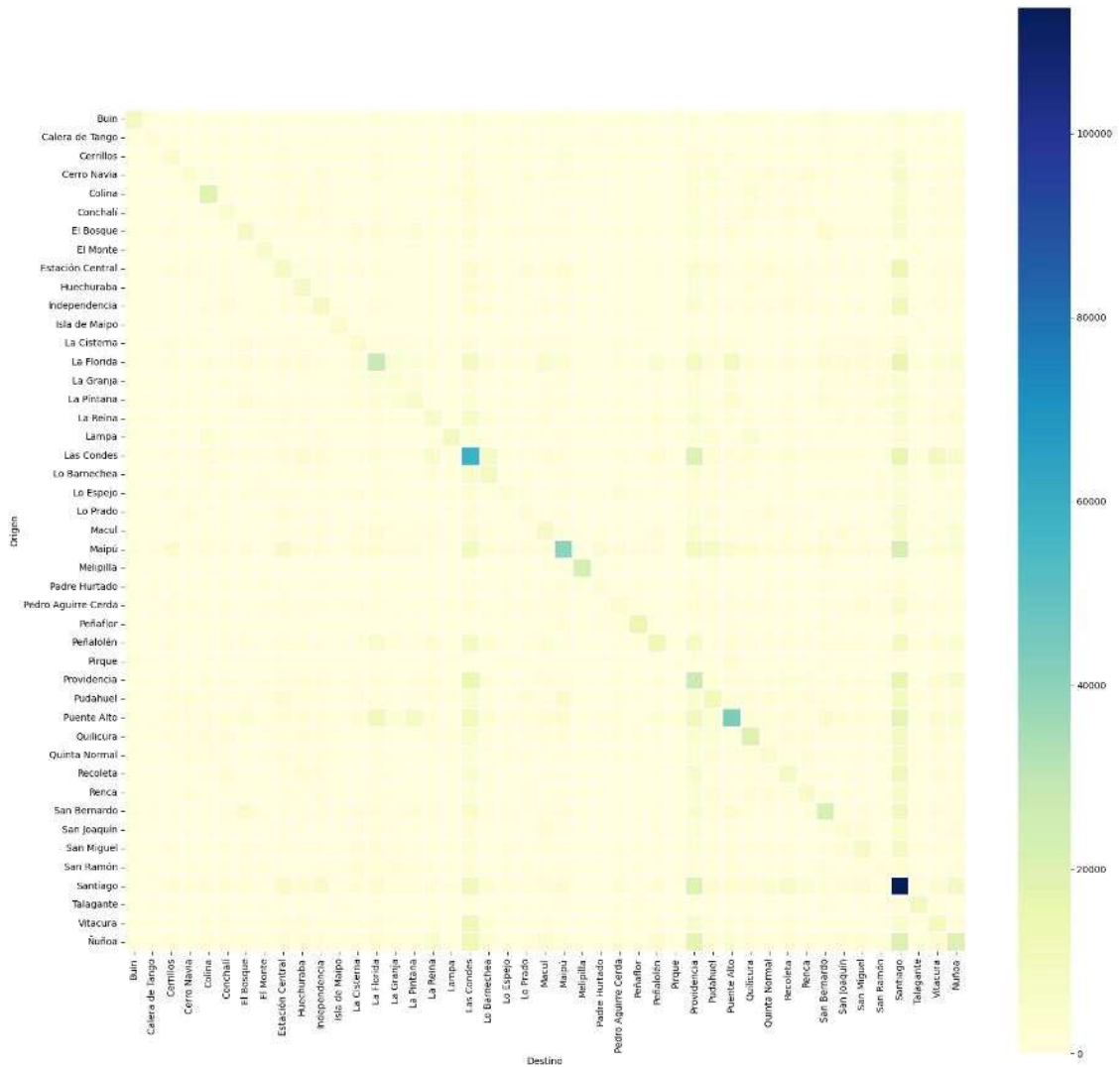
Una vez los resultados han sido ajustados, se generan las matrices origen-destino especificadas a nivel agregado por comuna y se compararán con los resultados del modelo Etraus.

5. Resultados

En la presente sección se presentan los hallazgos relevantes de cada etapa.

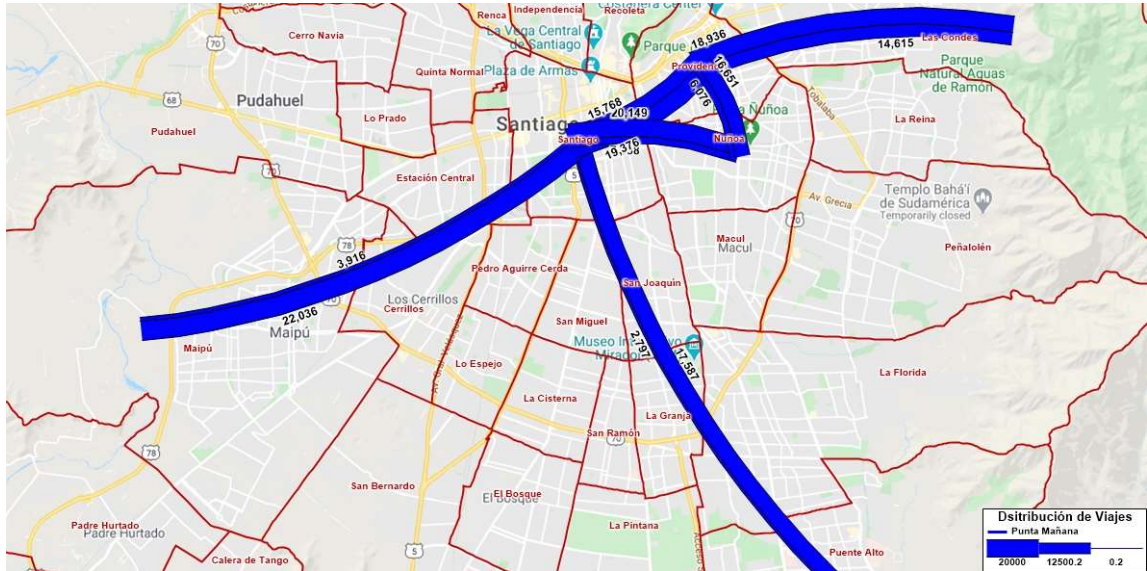
5.1. Limpieza y preprocesamiento de datos

Como se menciona en el punto anterior, a los datos del modelo Etraus se les agregó el campo comuna de origen y destino a las matrices de viajes, obteniendo la siguiente distribución de viajes.

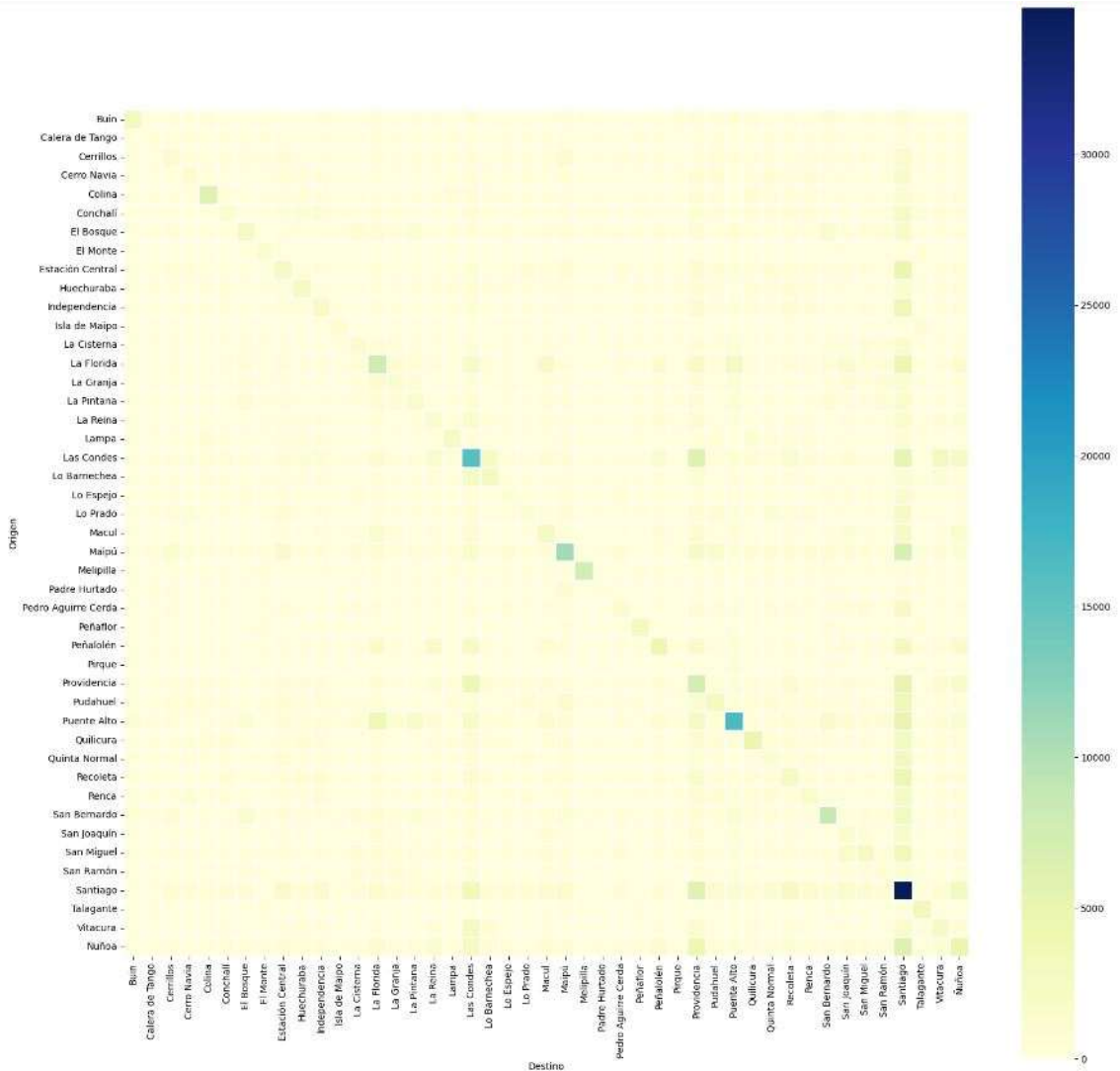


Resultados Modelación Estras período punta mañana, año 2019

Del cuadro se observa que los viajes más predominantes corresponden a los viajes intracomunales de Santiago, Las Condes, Puente Alto, Maipú, La Florida, Providencia, Melipilla y San Bernardo. Mientras que los viajes intercomunales destacan Maipú-Santiago, Ñuñoa-Santiago, Santiago-Providencia, Las Condes-Providencia, Puente Alto-Santiago y Ñuñoa-Providencia. A continuación, se muestra una gráfica.

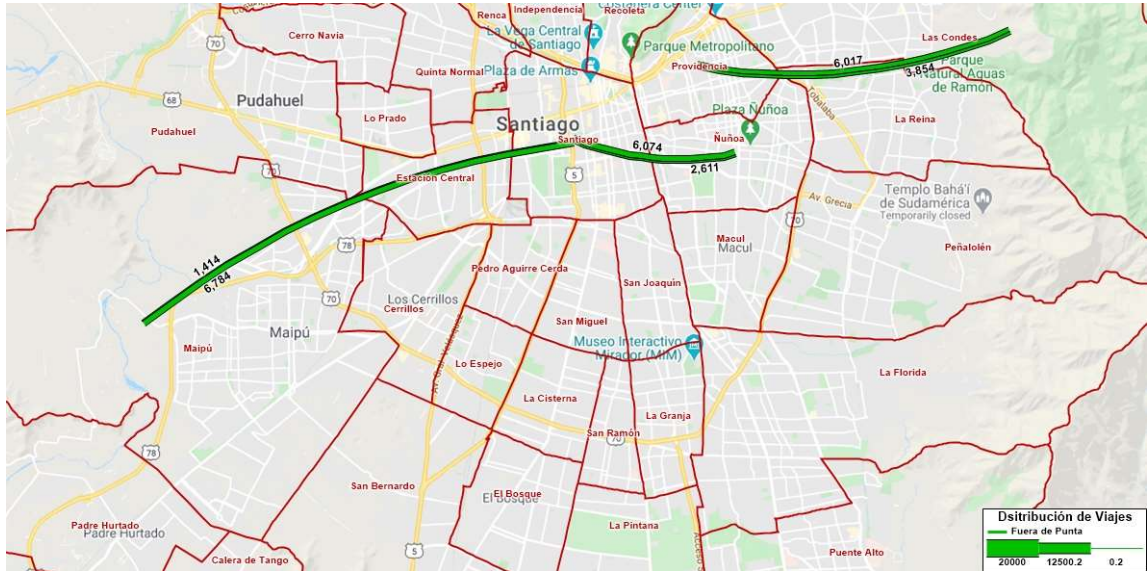


Viajes Intercomunales más importantes, período punta mañana

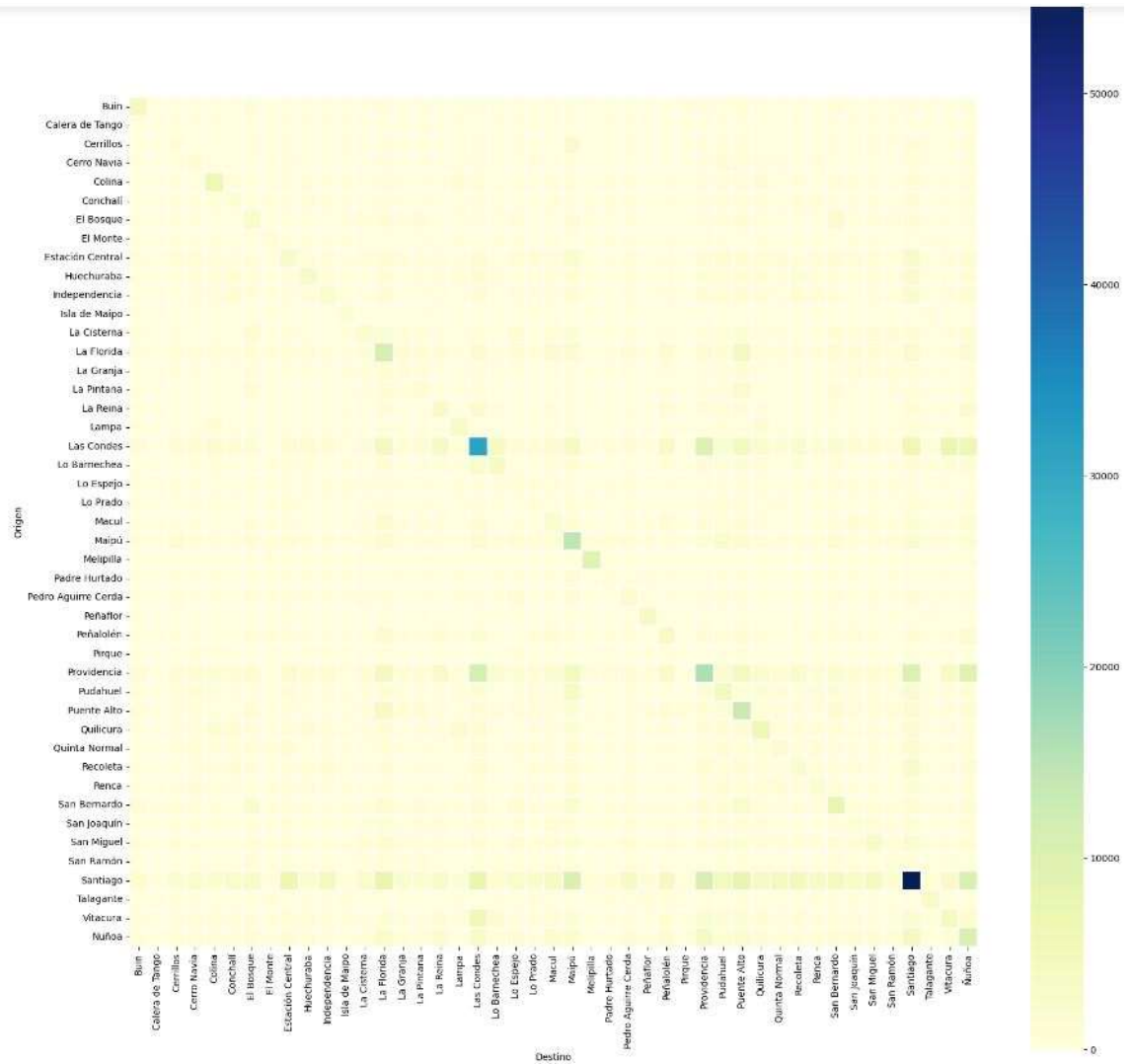


Resultados Modelación Estras período fuera de punta, año 2019

Del cuadro se observan ciertos patrones similares al punta mañana, los viajes más predominantes corresponden a los viajes intracomunales de Santiago, Puente Alto, Las Condes y Maipú. Mientras que los viajes intercomunales destacan Maipú-Santiago, Ñuñoa-Santiago, Las Condes-Providencia, Santiago-Providencia, Providencia-Santiago y Las Condes-Santiago.

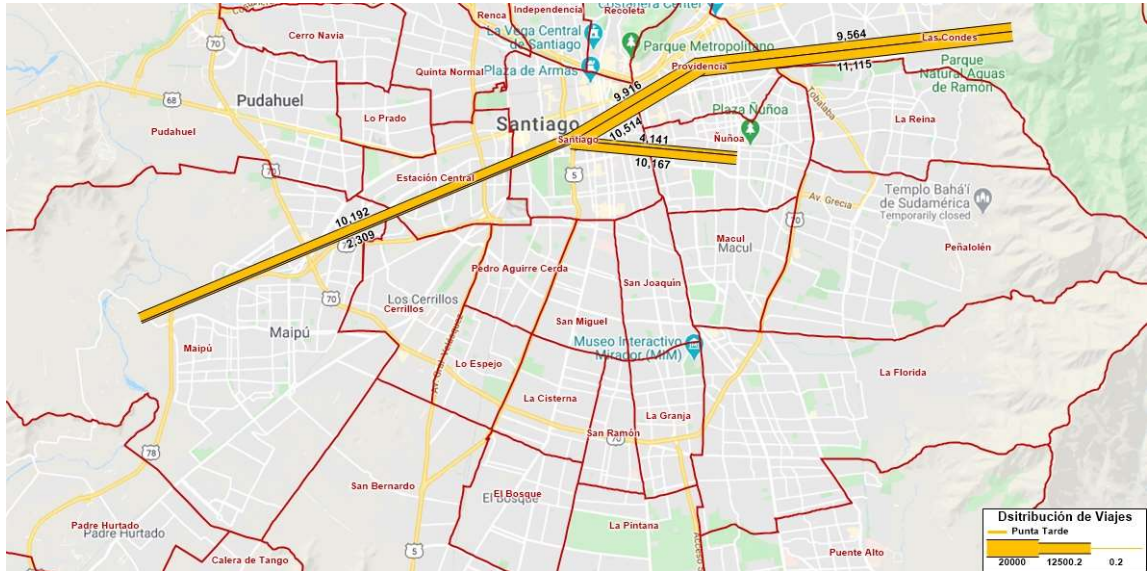


Viajes Intercomunales más importantes, período fuera de punta



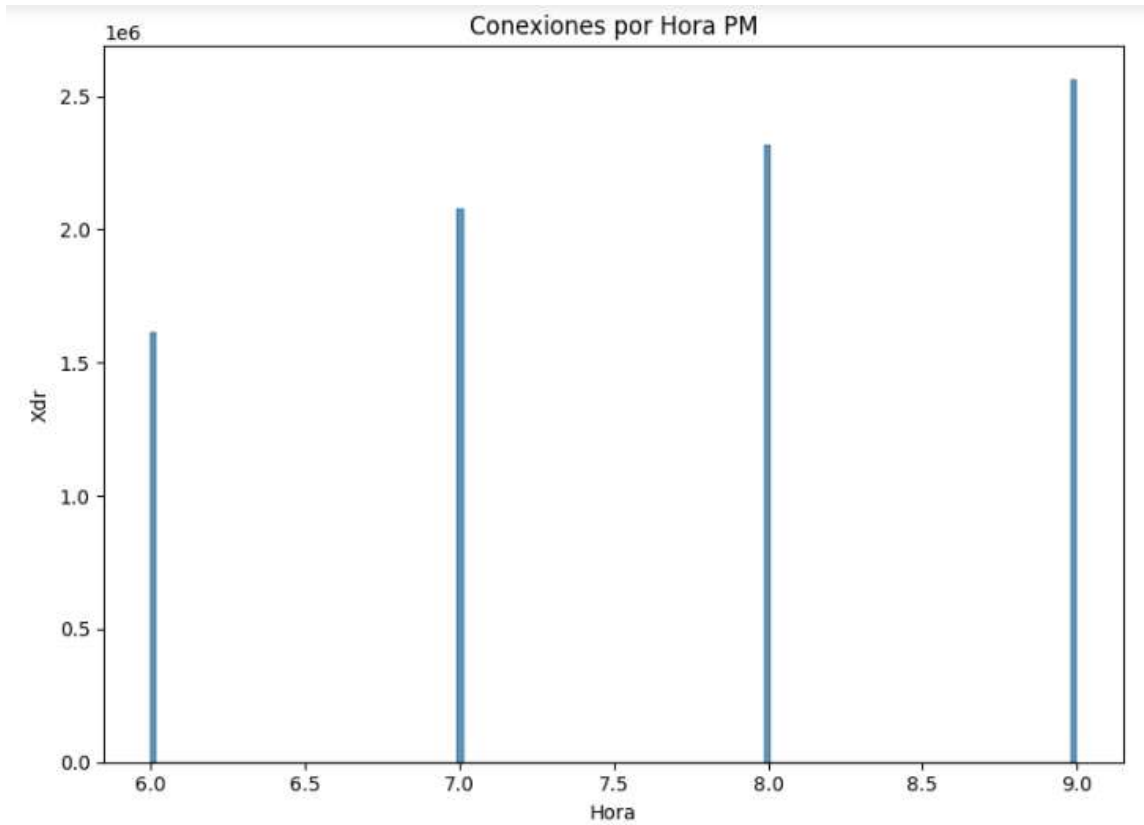
Resultados Modelación Estraus período punta tarde, año 2019

Del cuadro se observan que los viajes más predominantes corresponden a los viajes intracomunales de Santiago, Las Condes, Providencia, Maipú, Puente Alto, La Florida y Ñuñoa. Mientras que los viajes intercomunales destacan Providencia-Las Condes, Santiago-Maipú, Santiago-Providencia y Santiago-Ñuñoa.



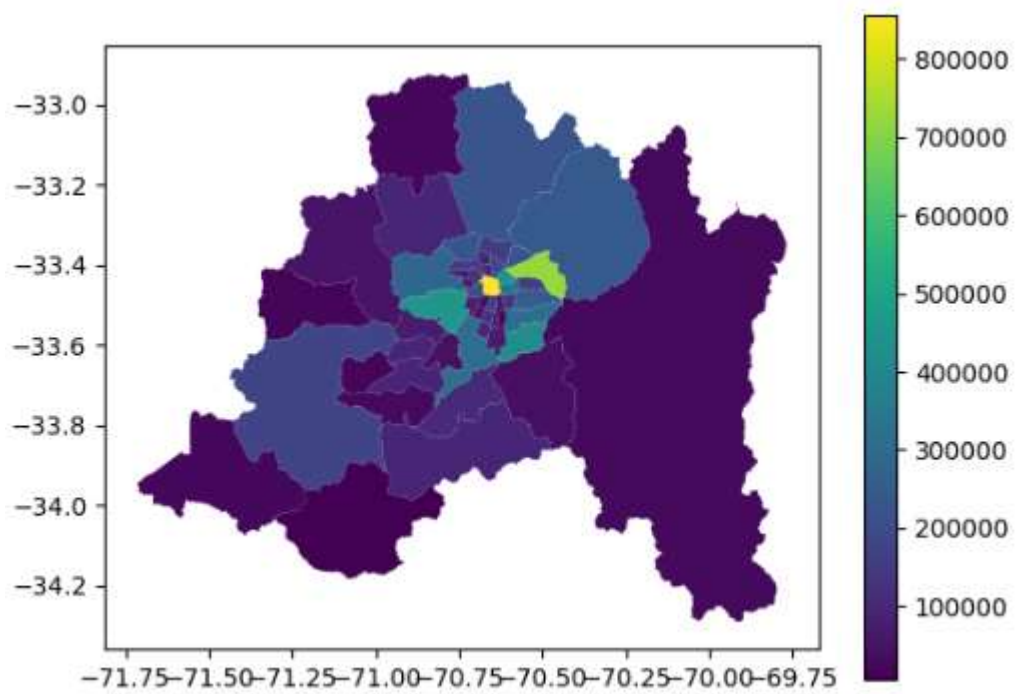
Viajes Intercomunales más importantes, período punta tarde

Respecto de la información contenida en los datos móviles y realizando los filtros comentados anteriormente (Día seleccionado, datos vacíos, Región Metropolitana y horarios definidos), se obtienen los siguientes resultados:



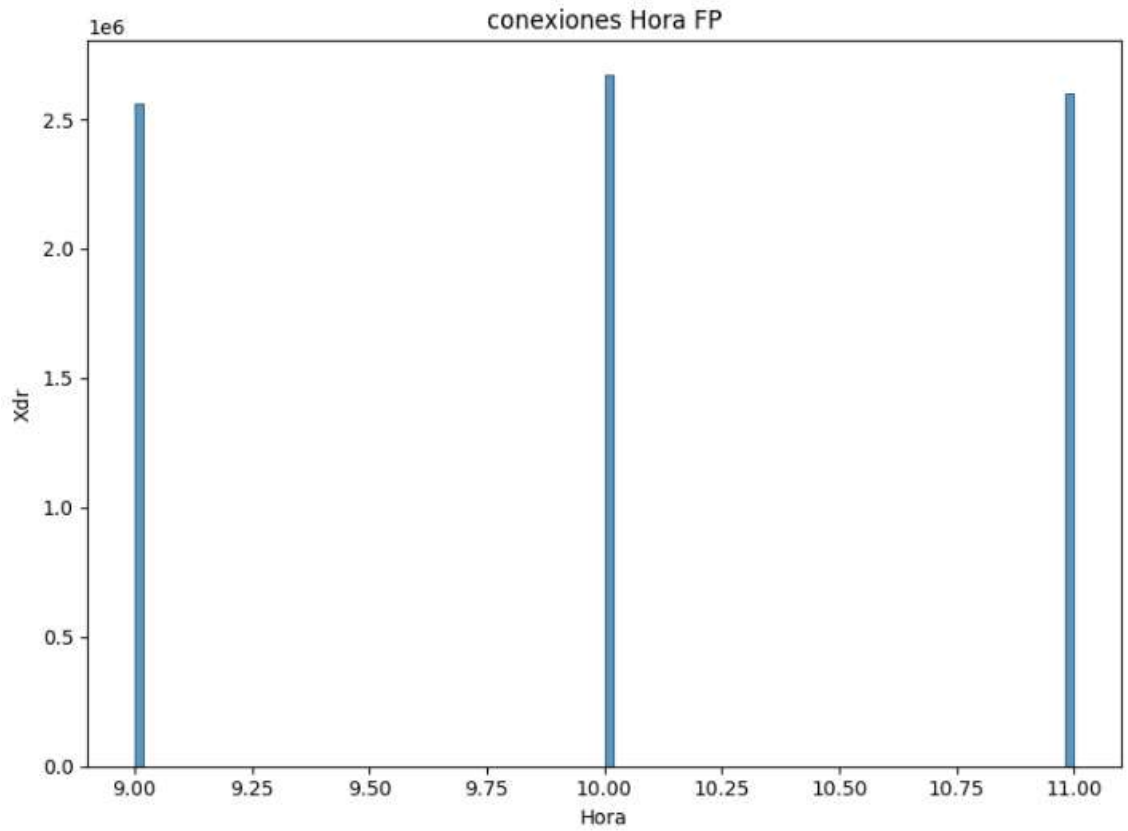
Número de Conexiones entre las 06:00-09:00

A medida que aumenta el tiempo existe un crecimiento del número de conexiones, situación distinta al comportamiento de viajes donde se observa un descenso después de la 9:00.



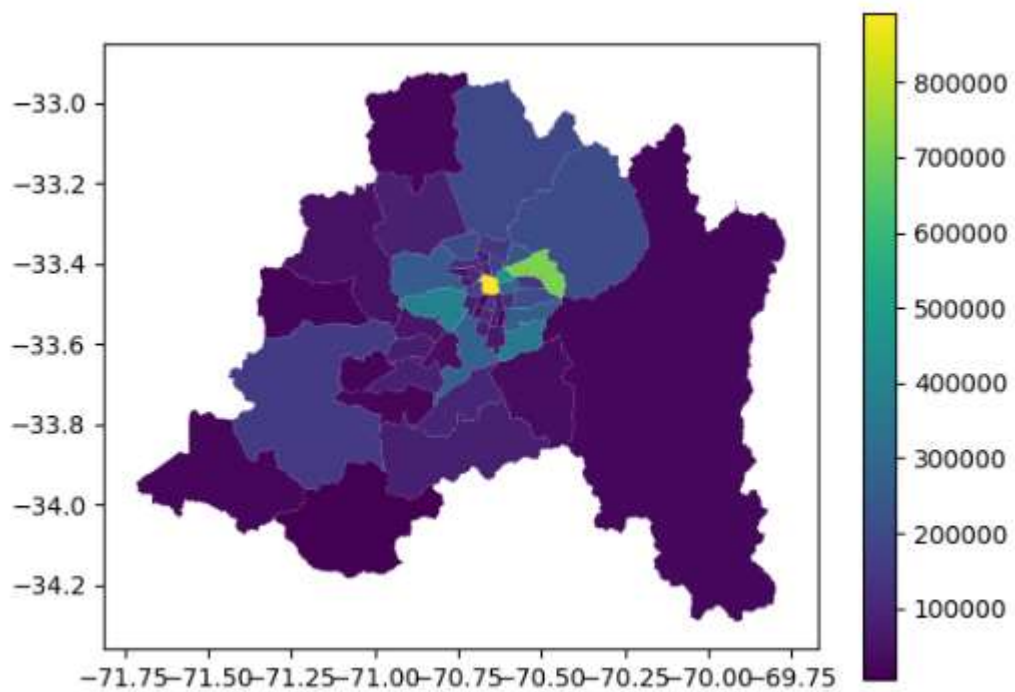
Distribución comunal de las Conexiones entre las 06:00-09:00

El mayor número de conexiones se concentra en las comunas de Santiago, Las Condes, Maipú y Puente Alto.



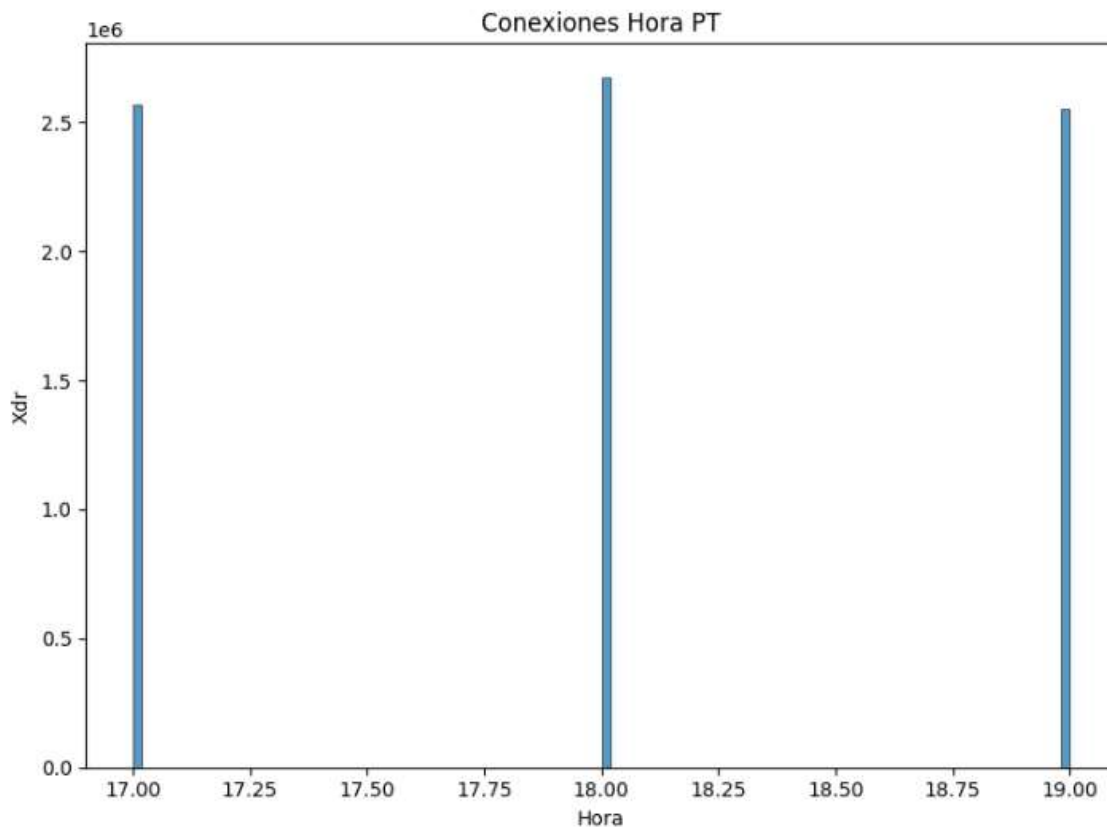
Número de Conexiones entre las 09:00-11:00

A medida que aumenta el tiempo se observa un pequeño peak a las 10:00.



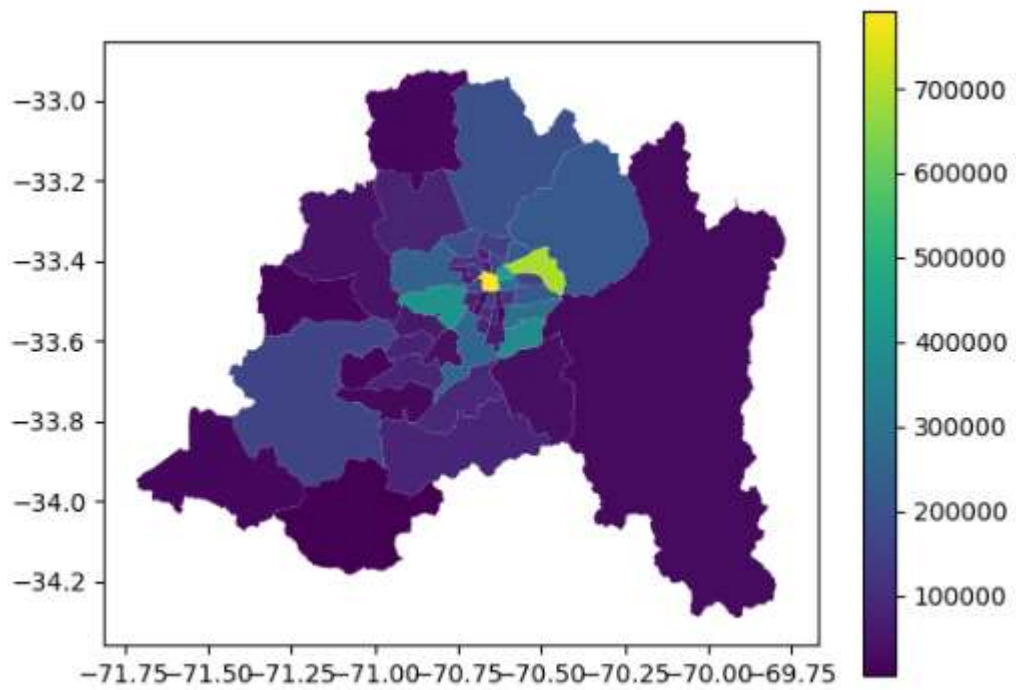
Distribución comunal de las Conexiones entre las 09:00-11:00

El mayor número de conexiones se concentra en las mismas comunas del punta mañana, como son: Santiago, Las Condes, Maipú y Puente Alto.



Número de Conexiones entre las 17:00-19:00

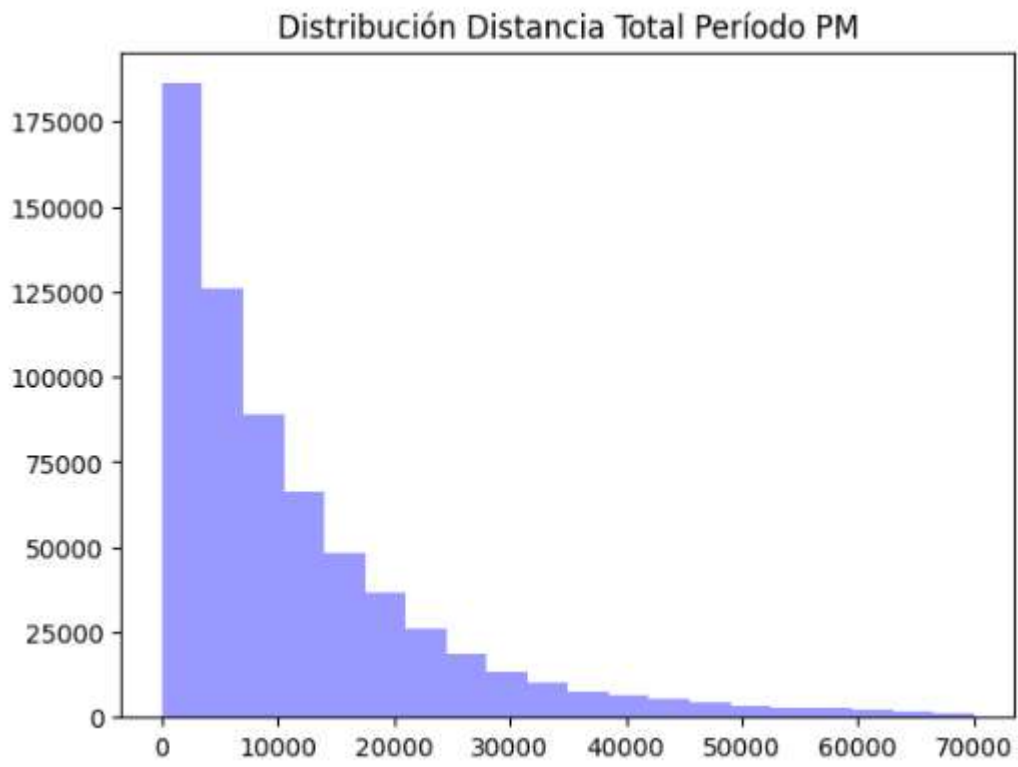
A medida que aumenta el tiempo se observa un pequeño peak a las 18:00, situación parecida al fuera de punta.



Distribución comunal de las Conexiones entre las 17:00-19:00

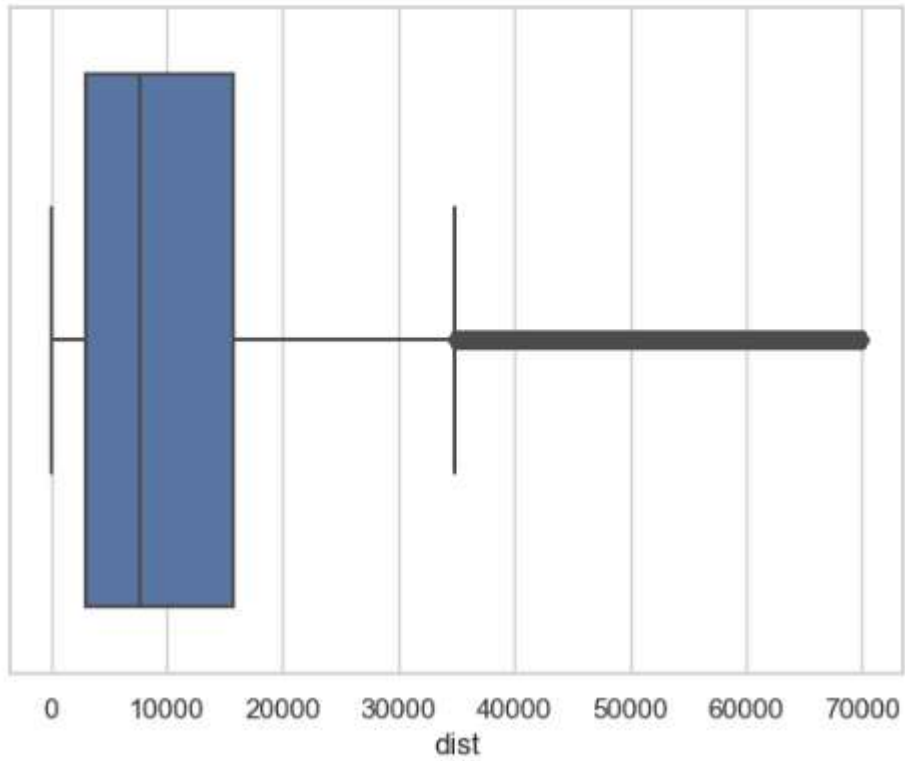
El mayor número de conexiones se concentra en las mismas comunas del punta mañana y fuera de punta.

Incorporando filtros por distancia, eliminando dispositivos con distancia nula y superior a 70 [km], se observan los siguientes perfiles de distancia.

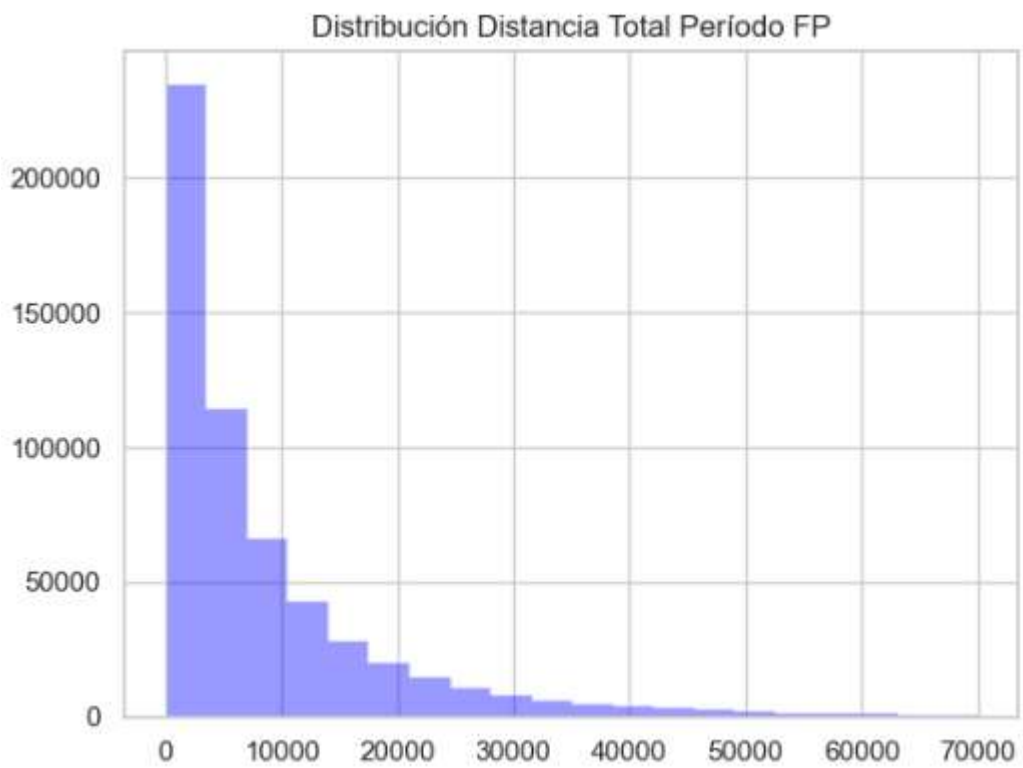


Distribución de distancias para el punta mañana

IQR..... 12741.309225839417
Q1..... 3045.3794360507873
Q2..... 7606.881344934556
Q3..... 15786.688661890204
Bigote Máximo..... 34898.65250064933
Bigote Mínimo..... -16066.584402708339
La Media..... 11583.138216475189

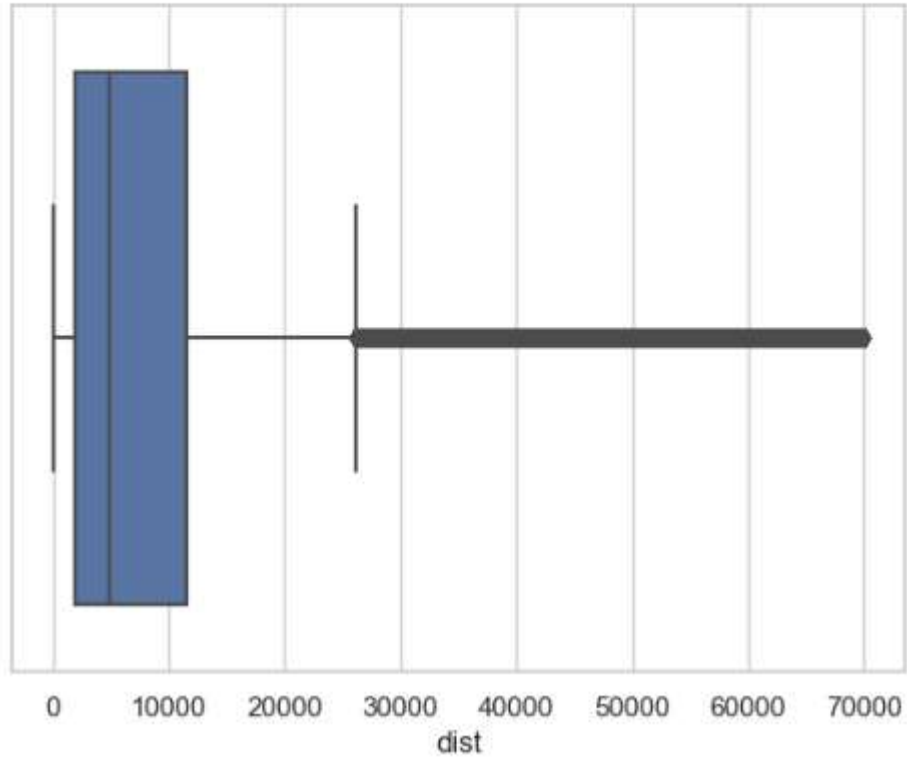


Indicadores para el punta mañana

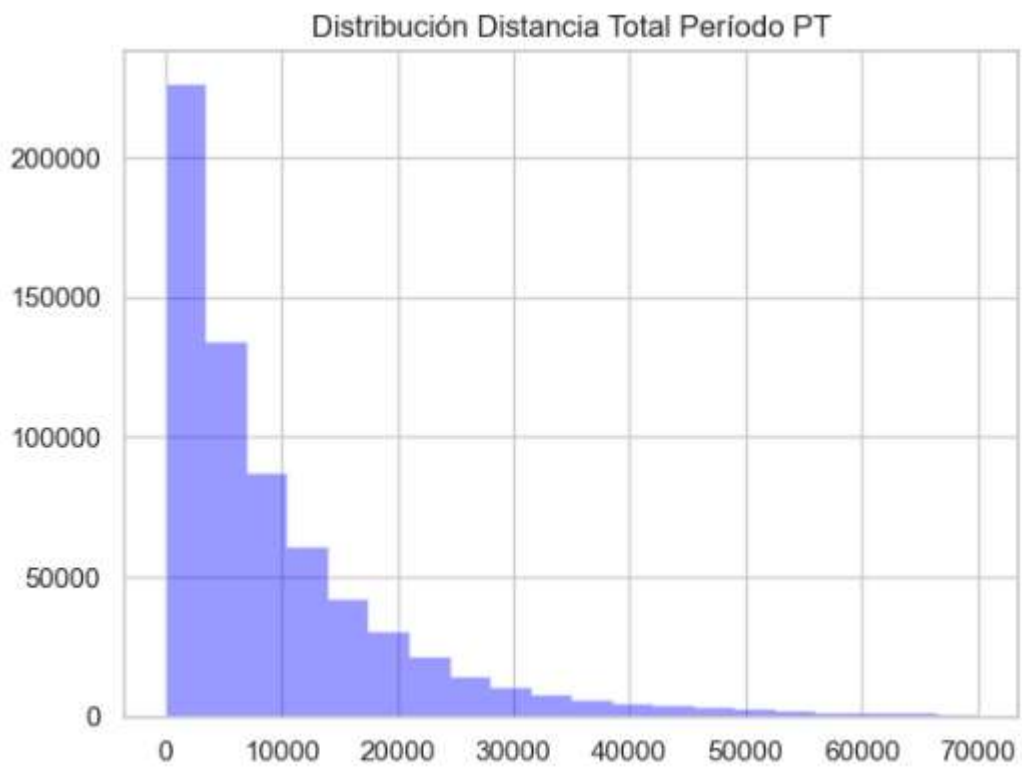


Distribución de distancias para el fuera de punta

IQR..... 9711.448744910118
Q1..... 1917.5232910110371
Q2..... 4838.73404830793
Q3..... 11628.972035921155
Bigote Máximo..... 26196.145153286332
Bigote Mínimo..... -12649.64982635414
La Media..... 9200.91283719072

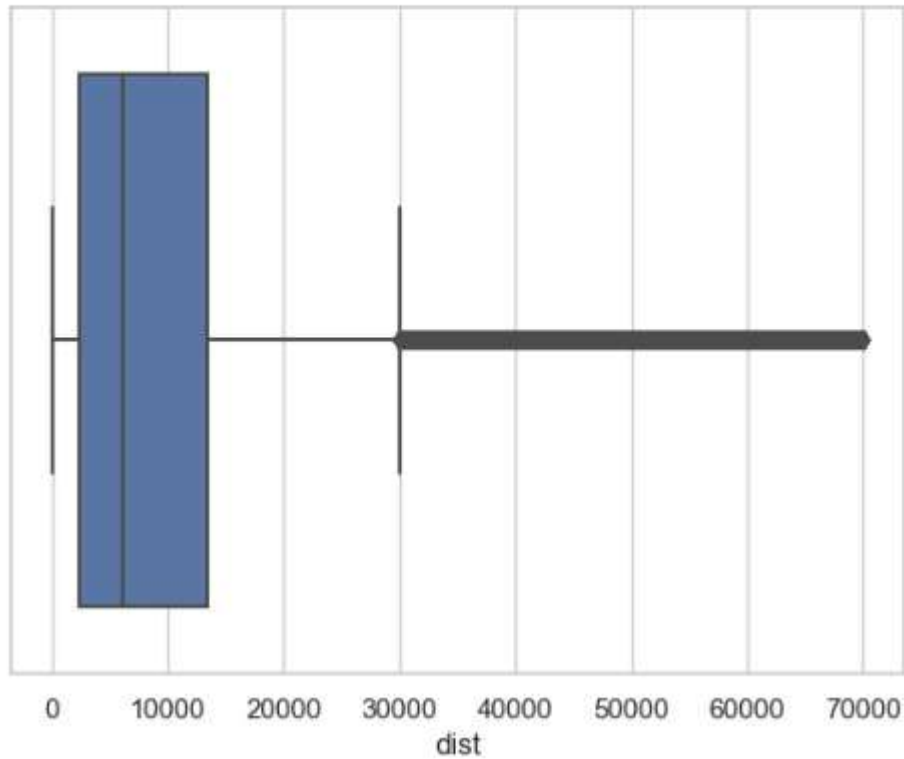


Indicadores para el fuera de punta



Distribución de distancias para el punta tarde

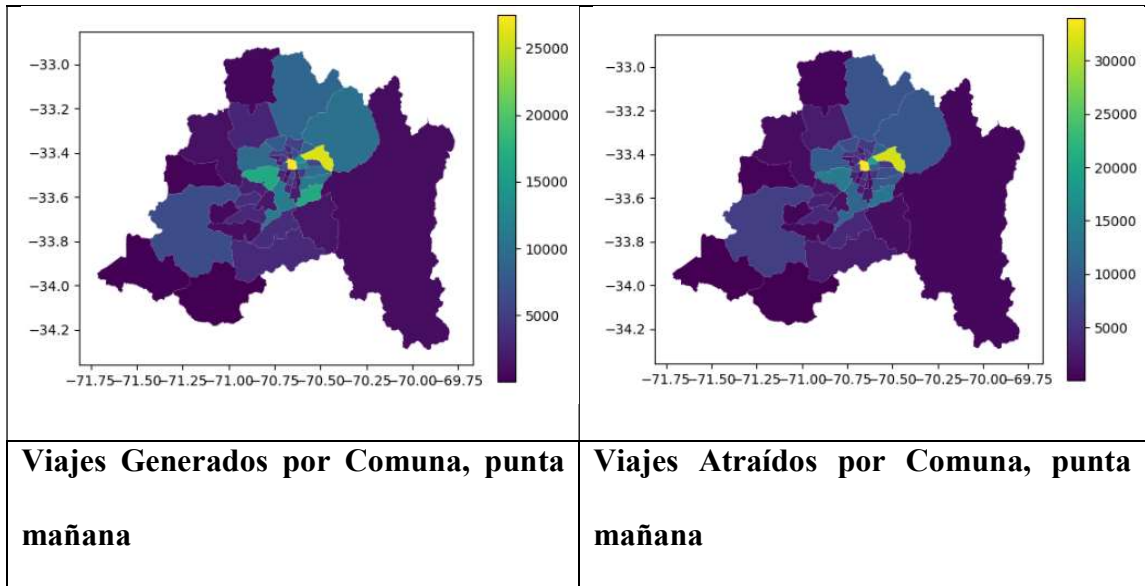
```
IQR..... 11043.717938917247
Q1..... 2451.0737286773565
Q2..... 6159.178003048774
Q3..... 13494.791667594603
Bigote Máximo..... 30060.368575970475
Bigote Mínimo..... -14114.503179698515
La Media..... 10139.657190006576
```



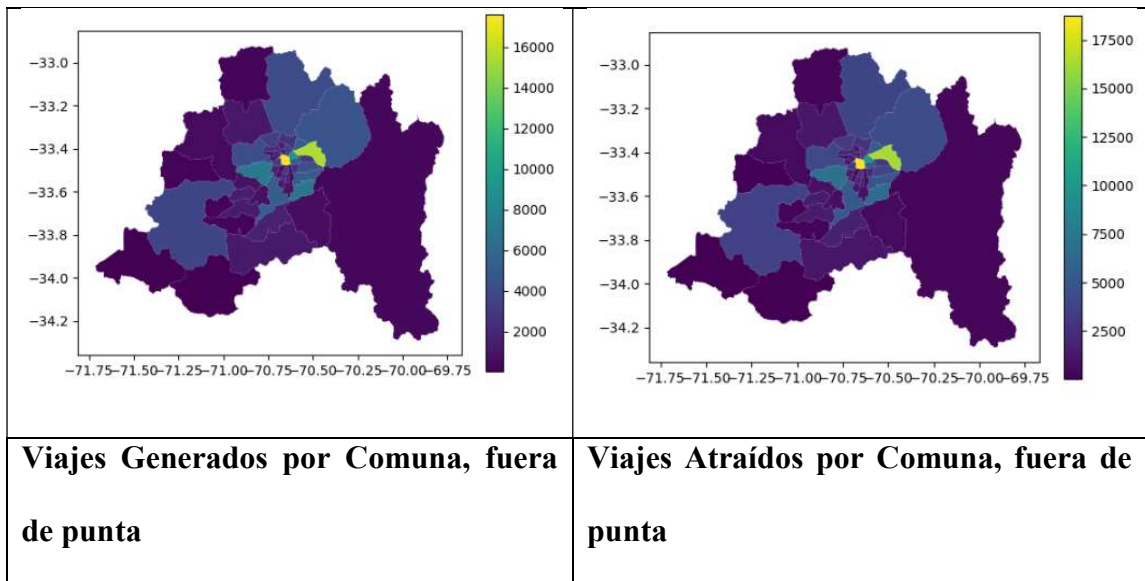
Indicadores para el punta tarde

5.2. Construcción de Matrices con datos móviles

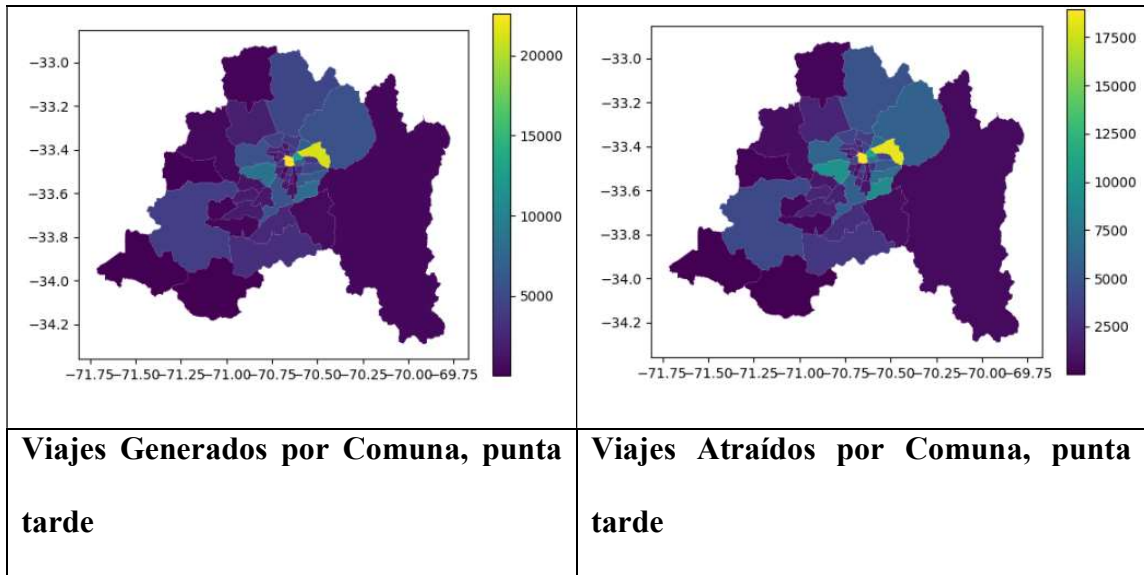
En este punto se presentan los resultados de la construcción de matrices de viajes, que incluyen los períodos punta mañana, fuera de punta y punta tarde. A modo ilustrativo, se presentan los resultados de la generación, atracción y distribución de viajes por períodos.



Las comunas con mayor generación de viajes son: Santiago, Las Condes, Maipú, Puente Alto, Providencia y Ñuñoa. Mientras que las comunas con mayor atracción de viajes son: Santiago, Las Condes, Providencia, Maipú y Puente Alto.

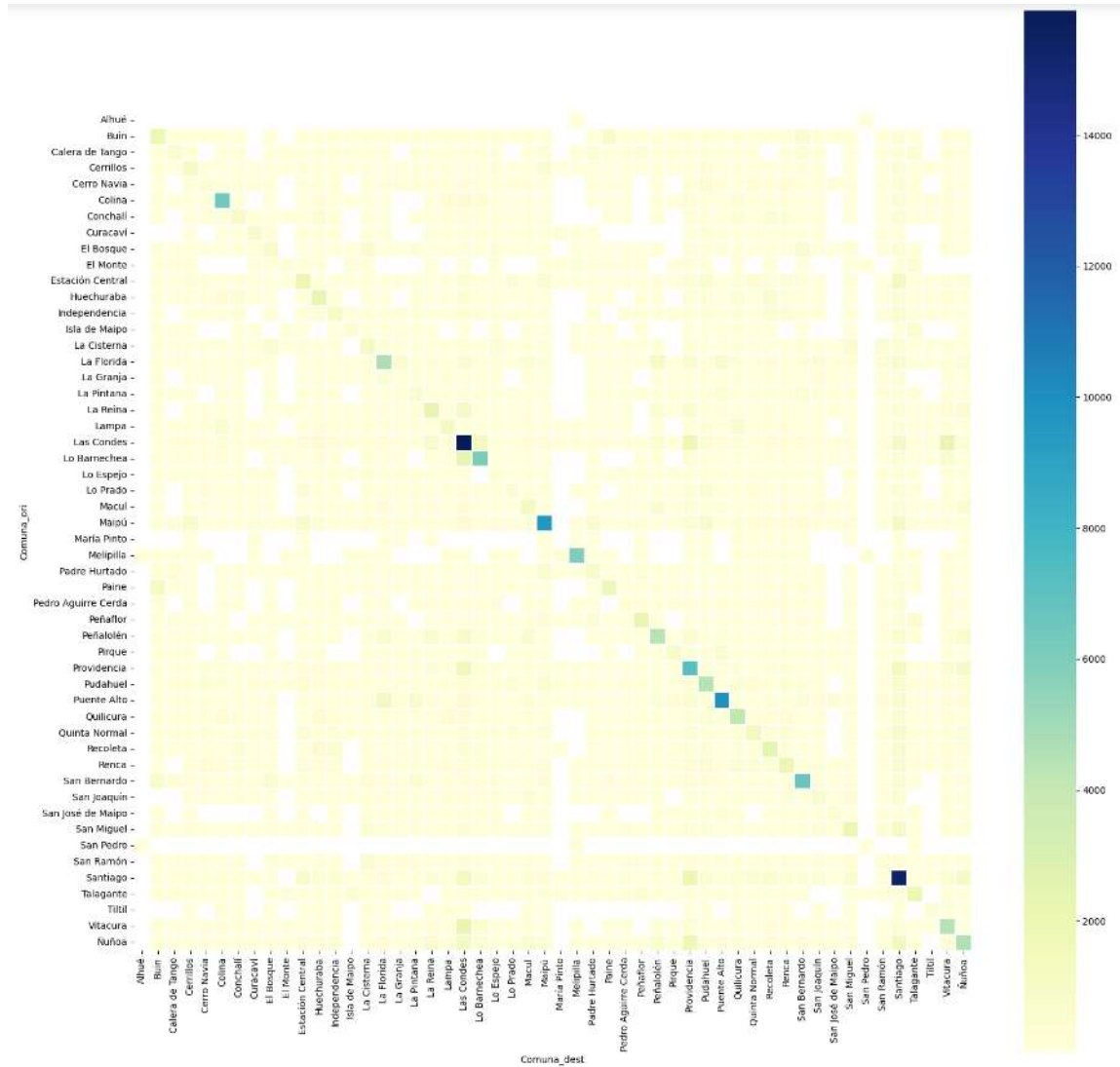


Las comunas con mayor generación de viajes son: Santiago, Las Condes, Providencia, Maipú, Puente Alto y Ñuñoa. Mientras que las comunas con mayor atracción de viajes son: Santiago, Las Condes, Providencia, Maipú, Puente Alto y Vitacura.



Las comunas con mayor generación de viajes son: Santiago, Las Condes, Providencia, Maipú, Puente Alto y Ñuñoa. Las comunas con mayor atracción de viajes son las mismas que la generación.

A nivel de distribución de viajes se observan los siguientes resultados.

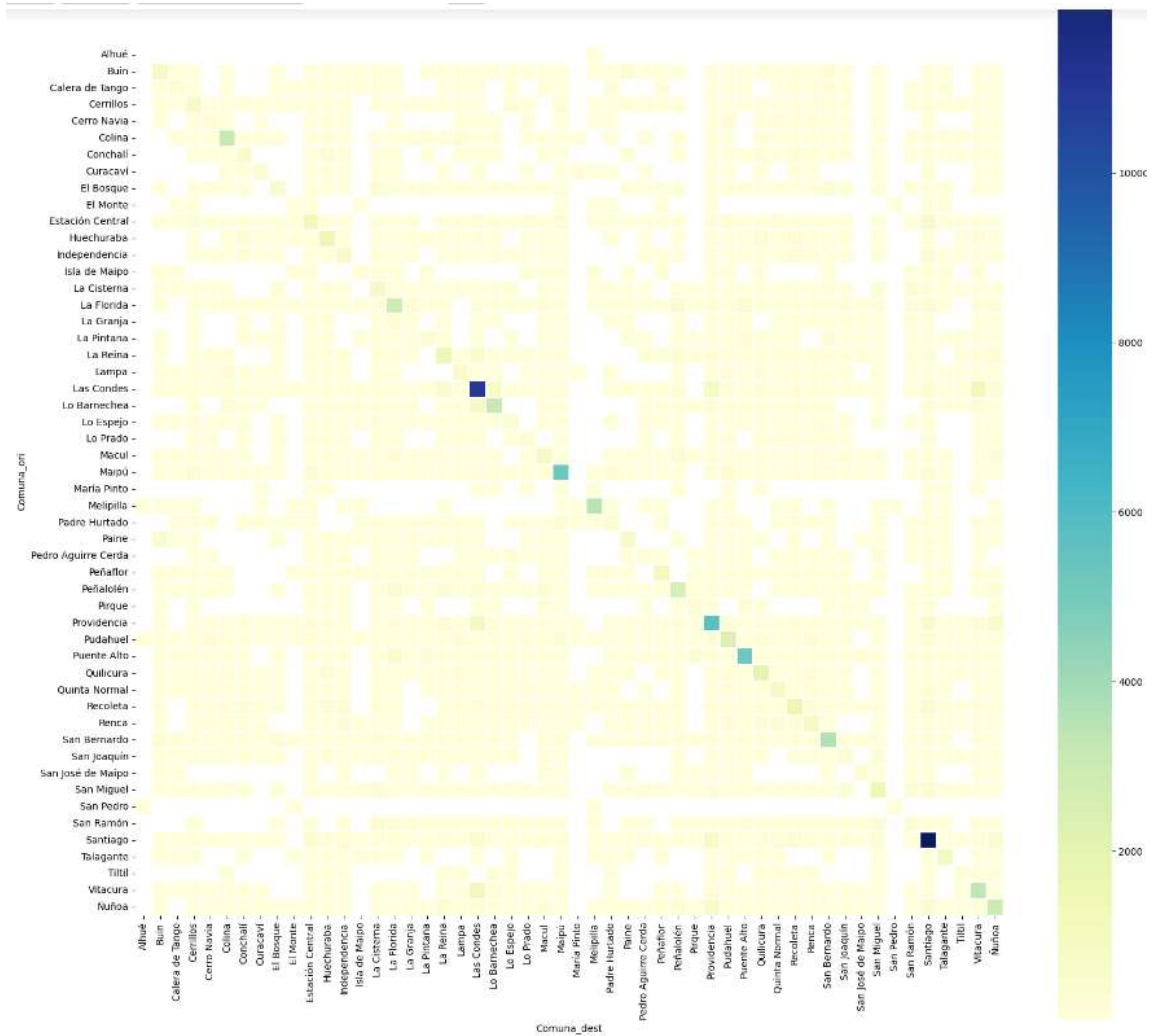


Resultados Telefonía Móvil período punta mañana, año 2022

Del cuadro se observa que los viajes más predominantes corresponden a los viajes intracomunales de Las Condes, Santiago, Puente Alto, Maipú, Providencia, San Bernardo y Colina. Mientras que los viajes intercomunales destacan Las Condes-Providencia, Las Condes-Vitacura, Lo Barnechea-Las Condes, Santiago-Providencia, Vitacura-Las Condes y Ñuñoa-Providencia. A continuación, se muestra una gráfica.

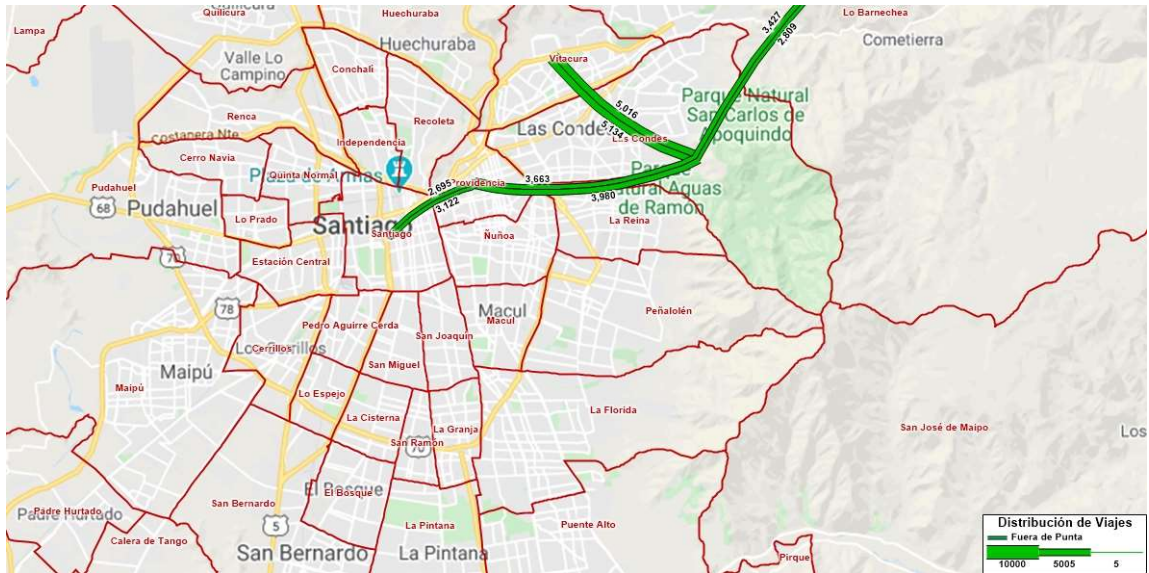


Viajes Intercomunales más importantes, período punta mañana

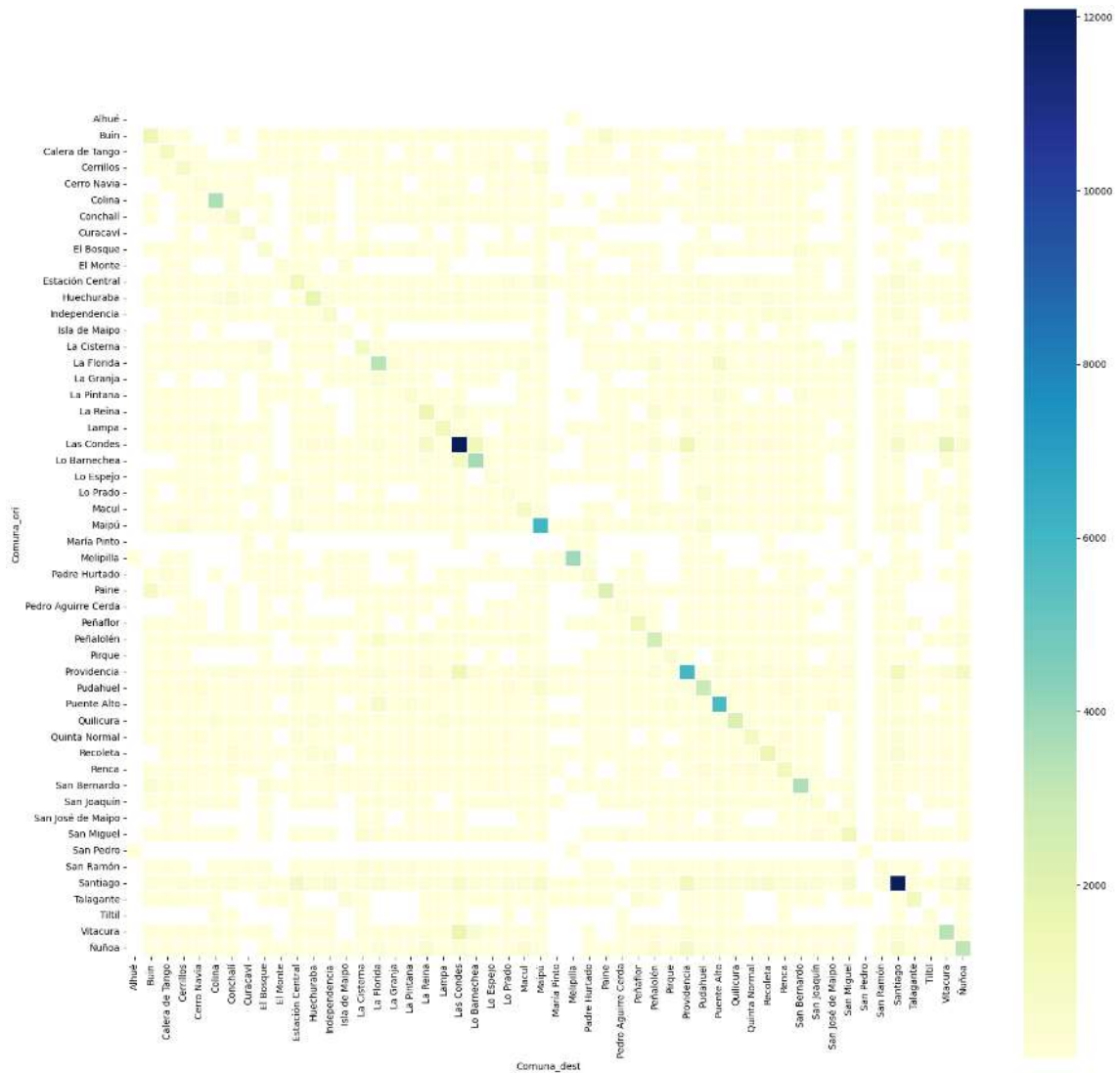


Resultados Telefonía Móvil período fuera de punta, año 2022

Del cuadro se observa que los viajes más predominantes corresponden a los viajes intracomunales de Santiago, Las Condes, Providencia, Maipú y Puente Alto. Mientras que los viajes intercomunales destacan Las Condes-Providencia, Las Condes-Vitacura, Lo Barnechea-Las Condes, Santiago-Providencia, Vitacura-Las Condes y Ñuñoa-Providencia. A continuación, se muestra una gráfica.

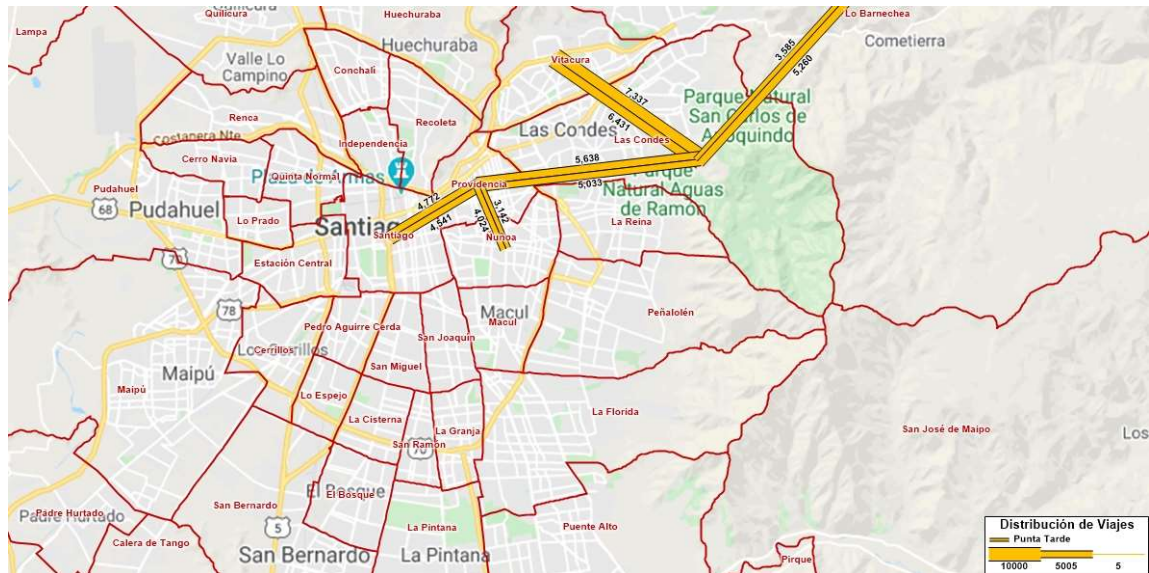


Viajes Intercomunales más importantes, período fuera de punta.



Resultados Telefonía Móvil período punta tarde, año 2022

Del cuadro se observan que los viajes más predominantes corresponden a los viajes intracomunales de Las Condes, Santiago, Providencia, Maipú, Providencia y Puente Alto. Mientras que los viajes intercomunales destacan Las Condes-Lo Barnechea, Las Condes-Providencia, Las Condes-Vitacura, Providencia-Las Condes, Providencia-Santiago, Providencia-Ñuñoa, Santiago-Providencia y Vitacura-Las Condes.



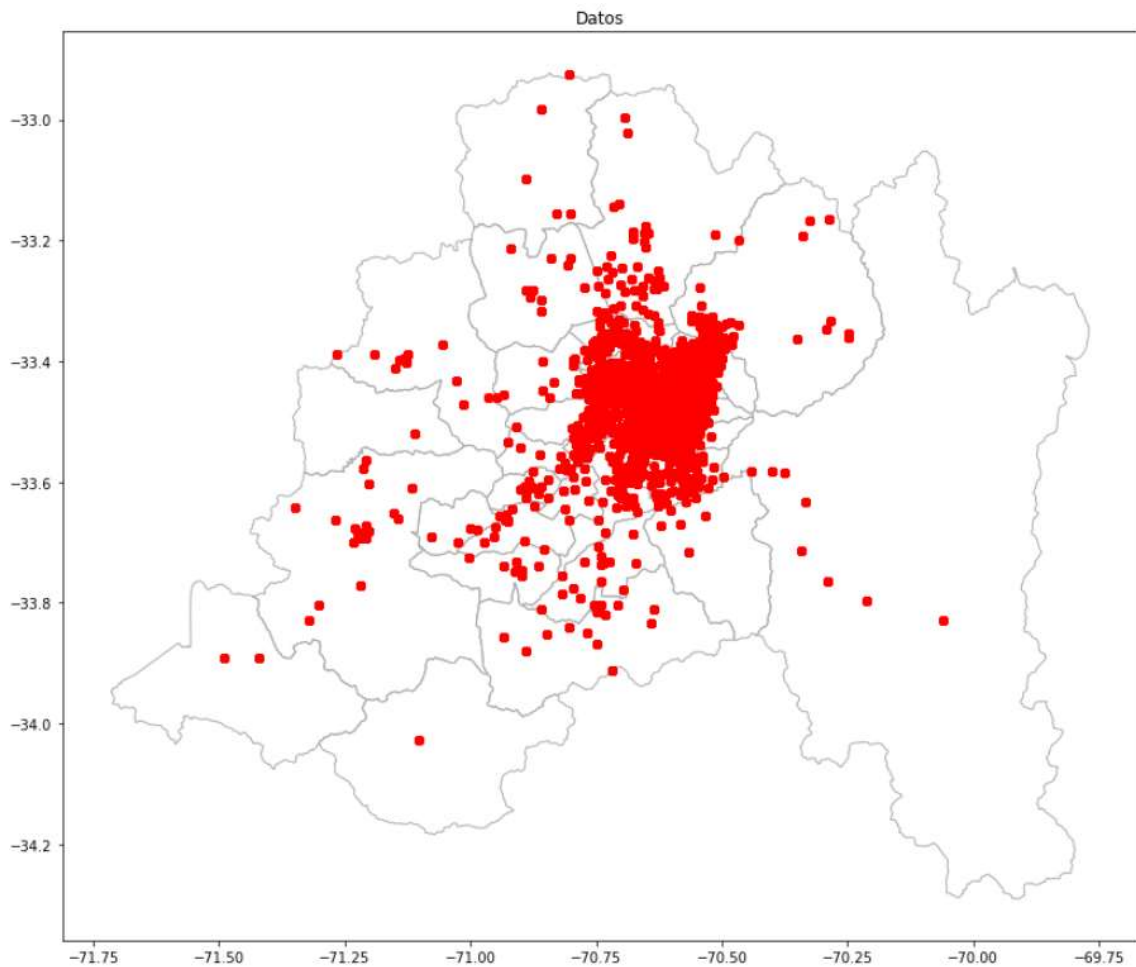
Viajes Intercomunales más importantes, período punta tarde.

5.3. Validación de los Resultados

La validación de este estudio consiste en comparar las matrices del modelo Etraus 2019 con los datos de telefonía del 3 de marzo del año 2022, dada su cercanía laboral. Sin embargo, hay que tener en consideración que:

- 1) La procedencia del modelo Etraus utiliza una encuesta del año 2012 y ha sido ajustado con los planes de desarrollo urbano, proyectando m^2 de comercio, m^2 servicios, número de hogares y otras variables de usos de suelo, para determinar los viajes al 2019. Por ende, no considera el efecto crisis social y pandemia sufridos entre el 2020 y 2021.

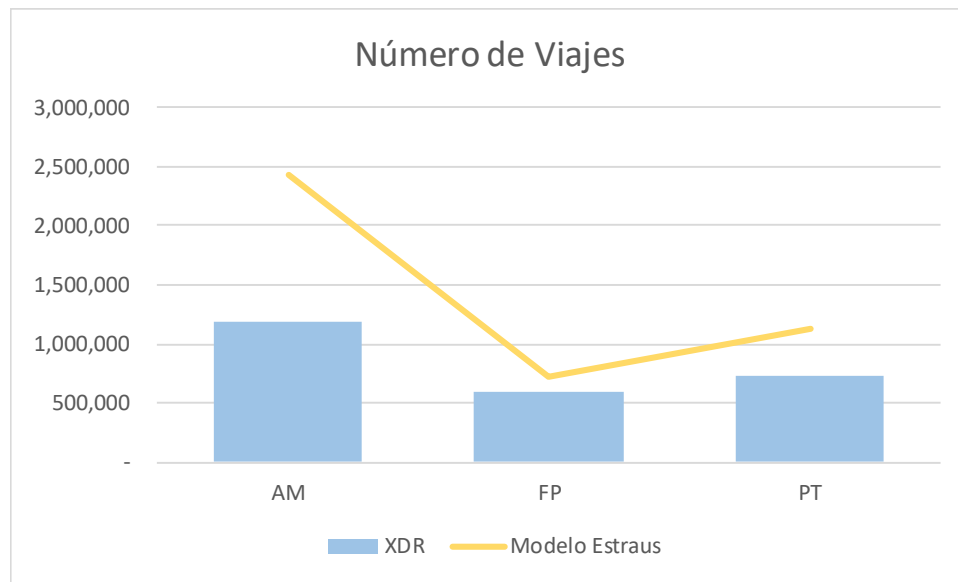
- 2) Existe una marcada tendencia de antenas ubicadas en las comunas céntricas y sector oriente en desmedro de las comunas periféricas o de estratos bajos, como: Cerro Navia, Colina, Conchalí, El Bosque, entre otras.



Ubicación de las antenas

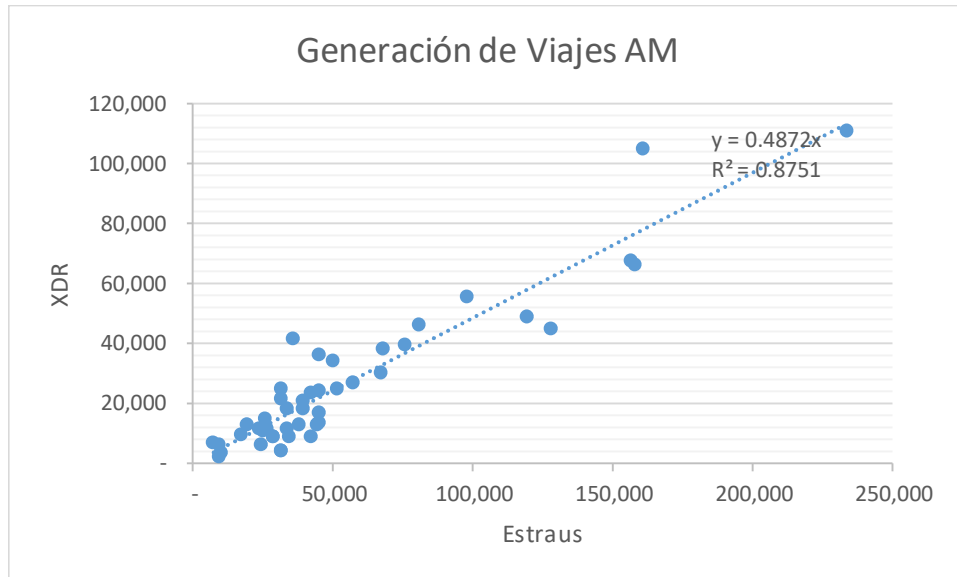
- 3) El modelo Estras no considera el efecto del teletrabajo, que según cifras del INE un 10.9% mantiene labores a distancia o teletrabajo en diciembre de 2021 [17].

Para el período punta mañana los viajes detectados con datos móviles fueron de 1,192,203, para el fuera de punta de 586,110 y punta tarde de 737,142. Esto significa que los datos del 2022, registran una caída respecto de Etraus entre 18% y 51%, siendo el fuera de punta el período más parecido entre lo modelado y medido.

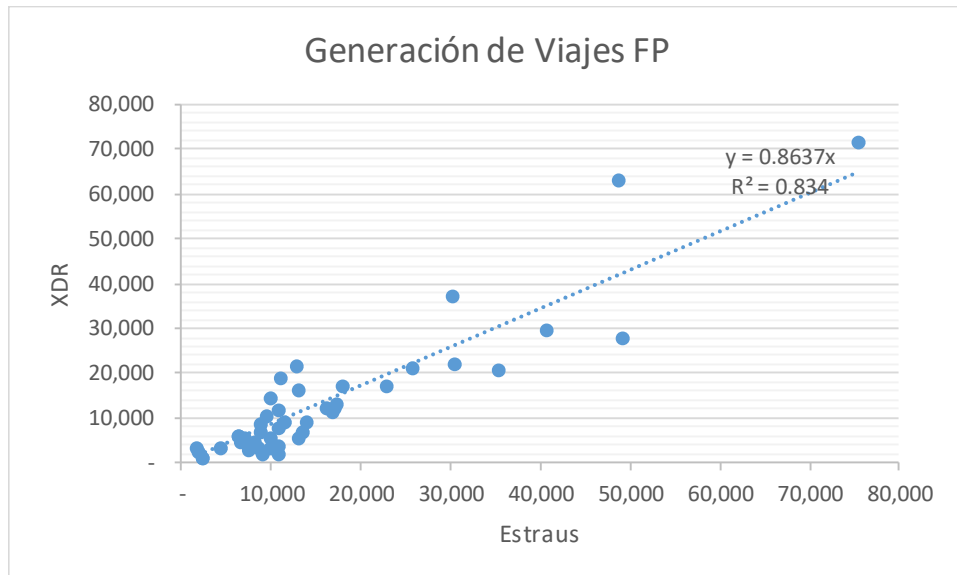


Número de Viajes según datos XDR y modelo Etraus

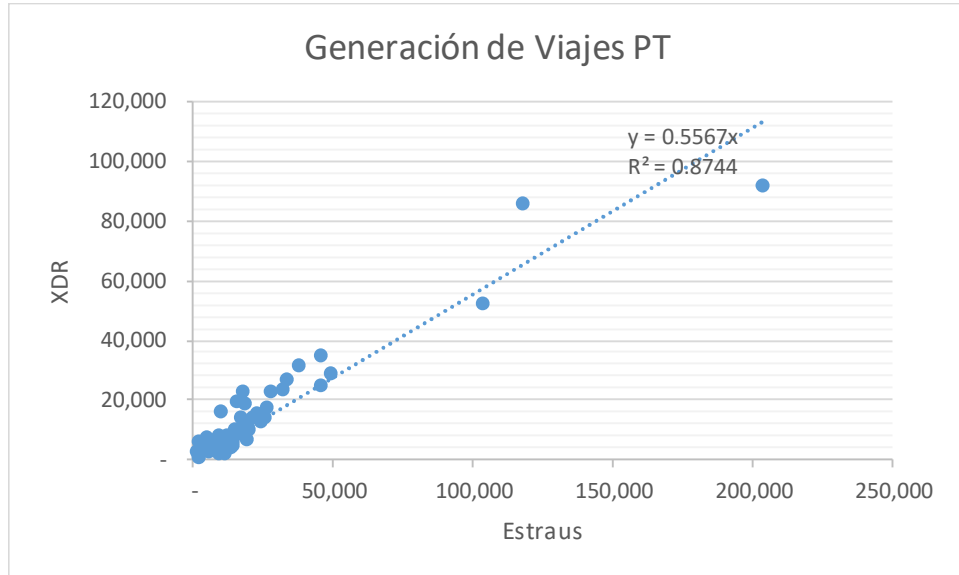
Si se analiza el comportamiento a nivel de viajes generados y atraídos se observan, que los datos XDR se ajustan bastante bien a la tendencia de los viajes del modelo, observando ajustes sobre el 80%



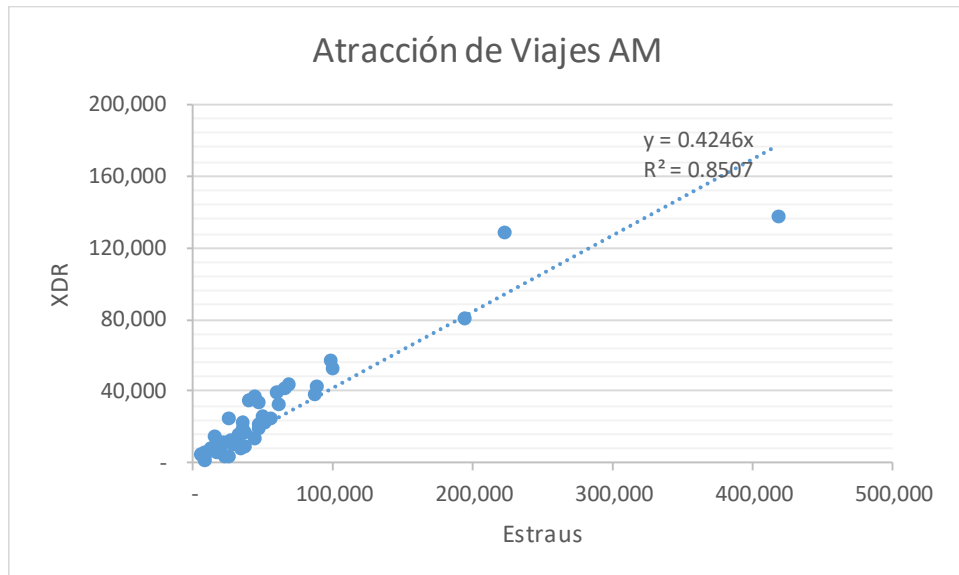
Tendencia de los viajes generados por comuna, período punta mañana.



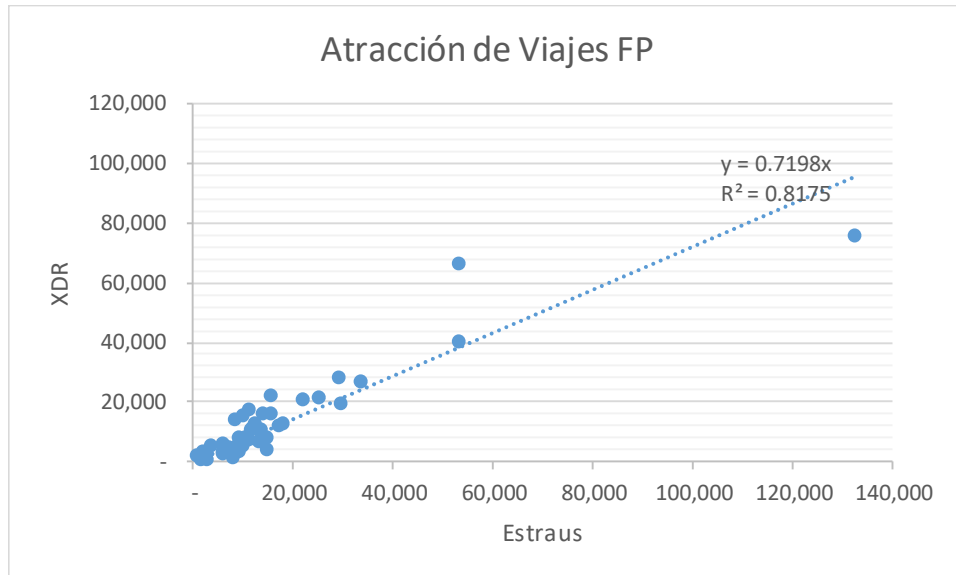
Tendencia de los viajes generados por comuna, período fuera de punta.



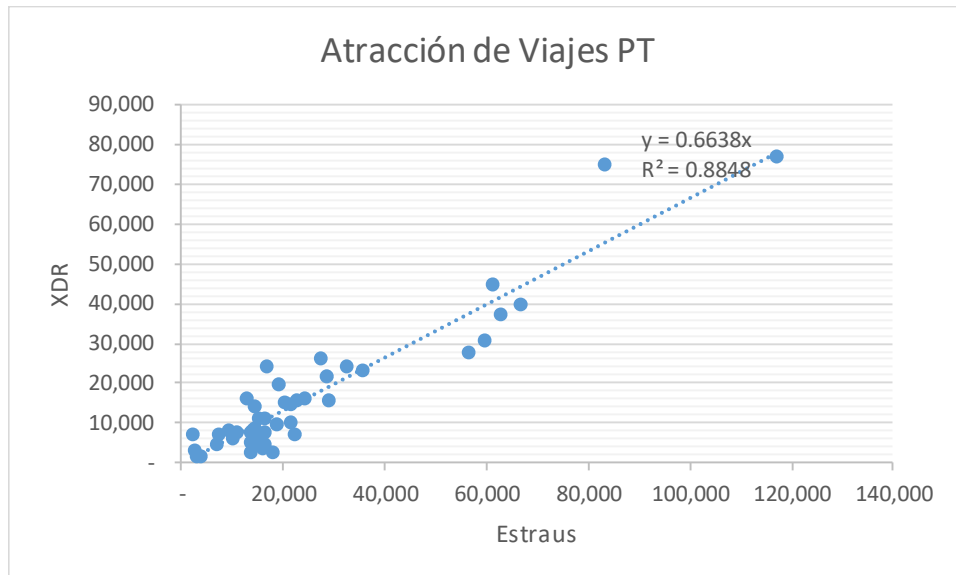
Tendencia de los viajes generados por comuna, período punta tarde.



Tendencia de los viajes atraídos por comuna, período punta mañana.



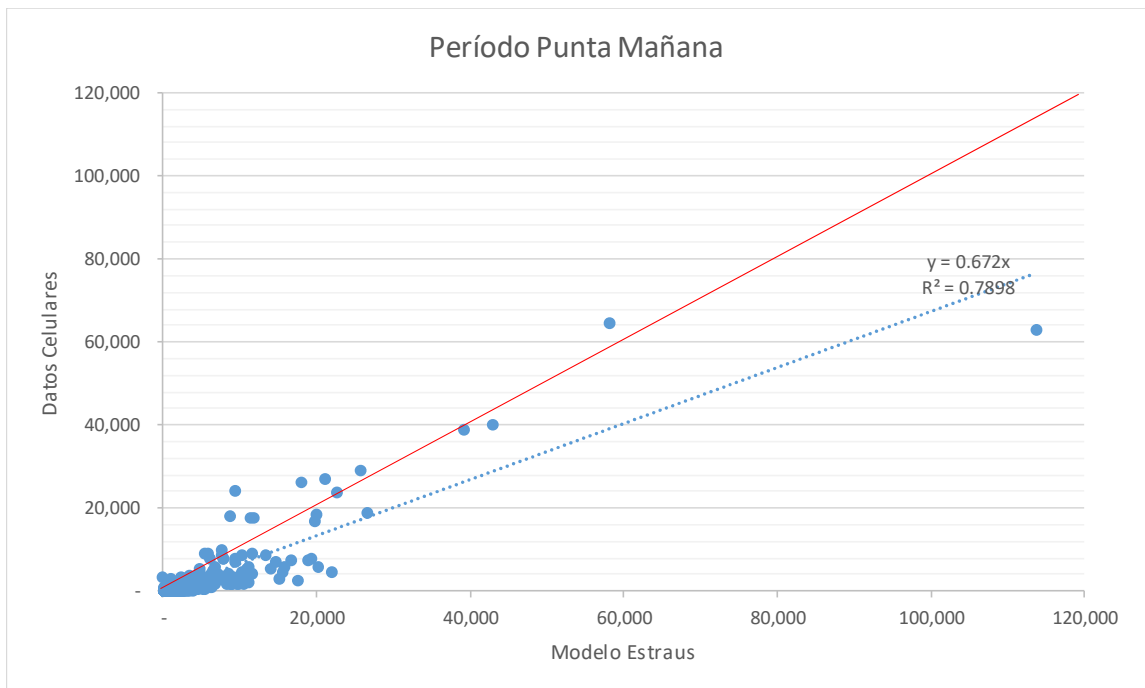
Tendencia de los viajes atraídos por comuna, período fuera de punta.



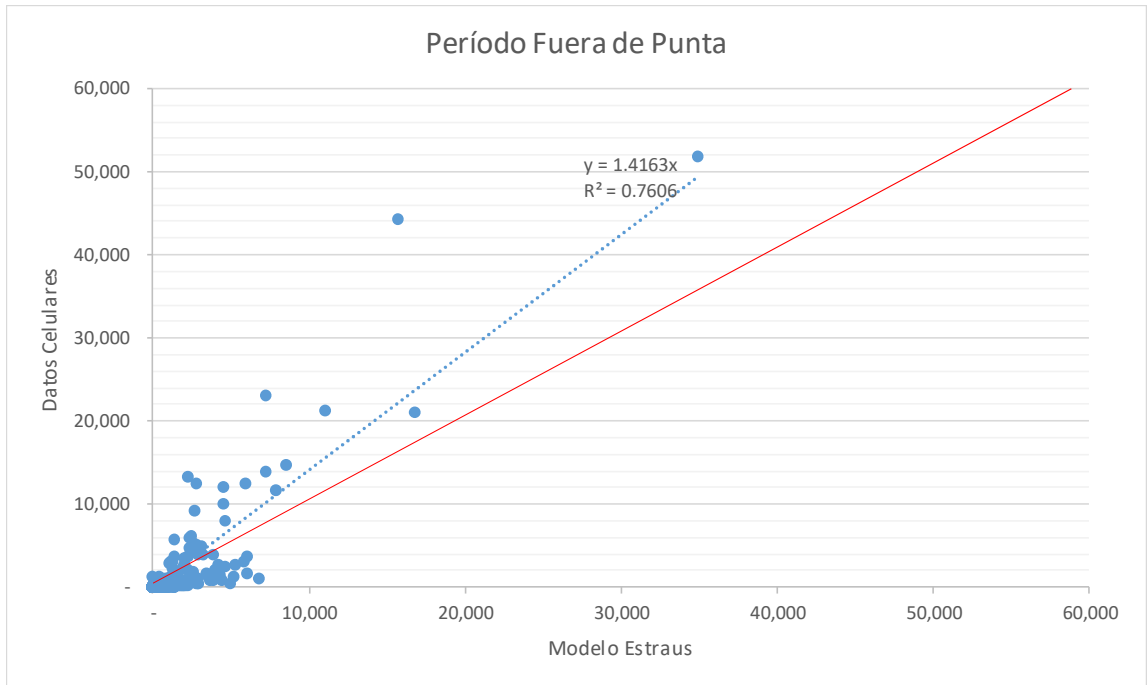
Tendencia de los viajes atraídos por comuna, período punta tarde.

Un resultado muy importante es la importancia relativa que se muestra en cada comuna respecto de los resultados analizados, tanto en la generación de viajes como atracción, existe un ajuste bastante considerable.

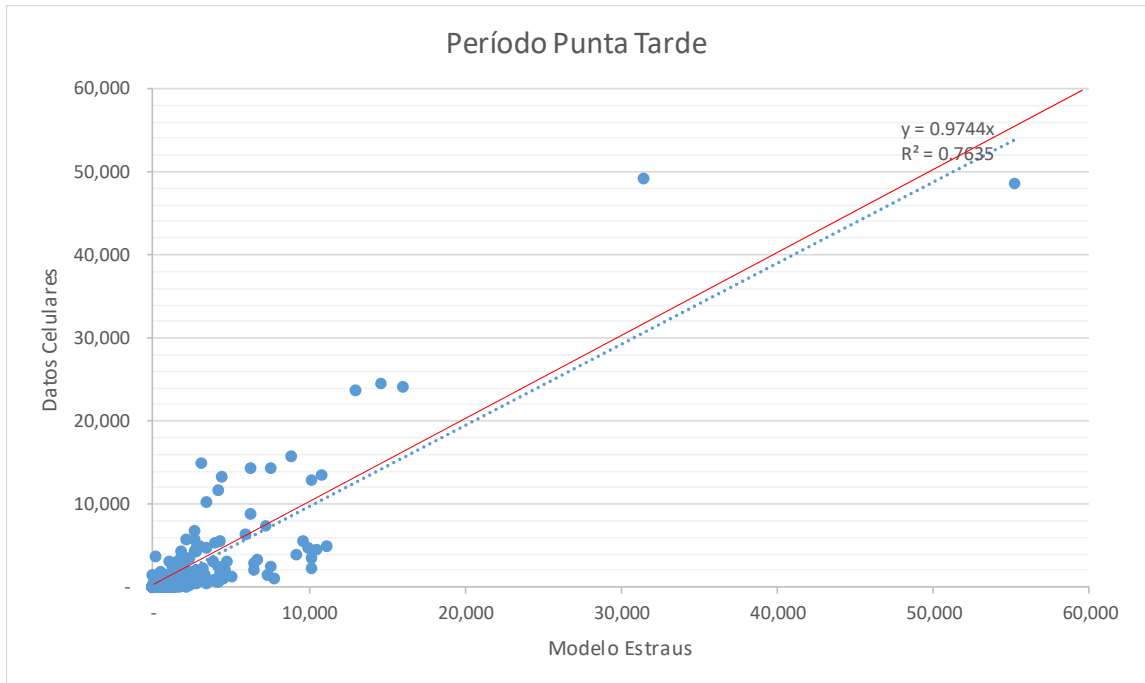
A continuación, se analiza los viajes a nivel de pares origen-destino por comuna para cada período de análisis.



Tendencia de los viajes entre pares orígenes-destino por comuna, período punta mañana.



Tendencia de los viajes entre pares orígenes-destino por comuna, período fuera de punta.



Tendencia de los viajes entre pares orígenes-destino por comuna, período punta tarde.

Los ajustes a nivel de viajes origen –destino mantienen un alto nivel grado de ajuste, son superiores al 70%, sin embargo la dispersión es más alta como era de esperar dado las consideraciones que se indicaron en el inicio y las diferencias en magnitudes totales observadas.

6. Conclusiones

6.1. Ventajas encontradas en el desarrollo del Capstone

Las principales conclusiones encontradas en el estudio son:

- Los datos de telefonía móvil permiten obtener grandes tamaños muestrales y son muy superiores a los que se utilizan las tradicionales EOD-H u otras encuestas.
- Las diferencias obtenidas en los niveles globales de viajes, pueden ser explicadas principalmente por las consideraciones de pandemia y teletrabajo, no recogidos en la modelación Etraus.
- Los datos de telefonía móvil podrían ayudar a corregir futuras predicciones de los modelos de transporte cuando surgen contingencias como pandemia o cambios importantes en la movilidad ejemplo teletrabajo.
- El mayor tamaño muestral permite entender de mejor manera la estructura de viajes, por esta razón se obtuvieron altos niveles de ajuste para los viajes generados y atraídos por comuna.
- Los niveles de ajuste para los pares orígenes-destino a nivel comunal, de igual manera se observan ajustes superiores al 70%. Por ende, se puede afirmar que la información de demanda de transporte generada a partir de los datos de telefonía móvil es de gran calidad.
- Los costos y plazos de ejecución del Capstone llevado a cabo han sido muy inferiores a los que se hubieran requerido emplear otras metodologías como las

basadas en encuestas, trabajos de toma de datos en campo y calibración de modelos de transporte.

- Los modelos tradicionales de transporte nunca son validados a través de los años y sin duda la utilización de datos de telefonía móvil podría ser una buena herramienta para ello, esto permitiría que el Estado tuviera un modelo actualizado sin grandes costos y ni largos plazos de ejecución. Esto ayudaría a que las priorizaciones de proyectos de transporte tuvieran un antecedente más dentro del análisis

6.2. Limitaciones

Las limitaciones encontradas son:

- Sesgo por el tipo de usuario, a pesar de indicar la participación de mercado que ofrece la empresa de telefonía, esta podría estar asociada a un tipo de cliente o grupo homogéneo de la muestra.
- Error muestral, a pesar de disponer una gran muestra de datos, de hecho muy superior a los métodos basados en encuestas, es importante tener en cuenta que el análisis se realiza sobre una muestra de usuarios de la población y no sobre el total, por lo que, al igual que en una encuesta, los resultados presentan un error muestral intrínseco a este tipo de aproximaciones de inferencia poblacional. La metodología de ajuste con otras fuentes de datos ayuda a resolver posibles errores cometidos en la inferencia.

- En este estudio no se tenía la información sociodemográfica de los usuarios de dispositivos móviles.

6.3. Futuras líneas de trabajo

En base a los resultados y conclusiones extraídos del estudio, se han identificado las siguientes futuras líneas de trabajo:

- El cálculo de distancias empleado, usa el criterio de la distancia más corta entre dos puntos y no sobre un trazado de red vial, podría ayudar a calcular velocidades de manera certera podría ayudar a corregir la estructura de los viajes.
- Utilizar otras fuentes de información que validen tasas de expansión diferenciadas, por ejemplo utilizando información socioeconómica de los clientes.
- Incluir datos de las tarjetas Bip que ayuden a validar los resultados.
- Incluir datos de recorridos y paraderos, que validen puntos de inicio y término de viajes, también sectores de estacionamientos.
- Trabajar con otras unidades espaciales, como zonificaciones más desagregadas.
- Aunque una parte importante del tiempo utilizado para el desarrollo del Capstone ha sido en mejorar la calidad de la información, aún existe margen de mejora para depurar esta información. Por ejemplo, utilizar datos de localización de importantes lugares de generación y/o atracción de viajes como por ejemplo grandes Industrias, Centros Comerciales, Hospitales, Colegios, etc.. Esto permitiría afinar los orígenes y destino de viajes.

- Utilizar más días de análisis que permitan validar la estructura de viajes y que no sea un día puntual y particular.

Bibliografía

1. Metodología de Análisis de Sistemas de Transporte de Ciudades de Tamaño Medio y Grande (MESPE, 2008)
2. A Day of Your Days: Estimating Individual Daily Journeys Using Mobile Data to Understand Urban Flow (Eduardo Graells-Garrido, Diego Saez-Trumper)
3. Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area (Francesco Calabrese, Member, IEEE, Giusy Di Lorenzo, Liang Liu, Carlo Ratti)
4. Use of Big Data in Transport Modelling Discussion Paper 186
5. Development of Origin-Destination Matrices Using Mobile Phone Call Data (Md. Shahadat Iqbal, Charisma F. Choudhury, Pu Wang, Marta C. González).
6. Origin–destination trips by purpose and time of day inferred from mobile phone data (Lauren Alexander, Shan Jiang, Mikel Murga, Marta C. González)
7. Inferring Dynamic Origin-Destination Flows by Transport Mode using Mobile Phone Data (Danya Bachir, Ghazaleh Khodabandelou, Vincent Gauthierc, Mounim El Yacoubi, Jakob Puchinger)
8. Experimental study design and grant writing in eight steps and 28 questions (Georges Bordage, Beth Dawson)
9. Evaluation of home detection algorithms on mobile phone data using individual-level ground truth (Luca Pappalardo, Leo Ferres, Manuel Sacasa, Ciro Cattuto, Loreto Bravo)

10. International Transport Forum: Use of Big Data in Transport Modelling
Discussion Paper (186)
11. Estudio de la Movilidad Interprovincial en España mediante la fusión de datos de telefonía móvil con otras fuentes de datos. (Javier Torres, Carlos Olivos, Miguel Picornell, Oliva García Cantú, Ricardo Herranz Nommon Solutions and Technologies),
12. Chen, J., & Bierlaire, M. (2015). Probabilistic multimodal map matching with rich Smartphone data. *Journal of Intelligent Transportation Systems*, 19(2), 134-148.
13. Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68, 285-299.
14. Especial Análisis Tráfico Internet Enero-Abril 2020 (2020). Subsecretaría de Telecomunicaciones. [Web; accedido el 03-09-2020].
https://www.subtel.gob.cl/wp-content/uploads/2020/06/PPT_Series_MARZO_2020_V1.pdf
15. Estudio determina el comportamiento de los usuarios de taxis vía smartphones en Chile <https://www.latercera.com/noticia/estudio-determina-el-comportamiento-de-los-usuarios-de-taxis-via-smartphones-en-chile/>
16. 25. Torres, M., Paz, K., & Salazar, F. (2006). Tamaño de una muestra para una investigación de mercado. *Boletín electrónico*, 2, 1-13.

17. INE entrega información de personas realizando teletrabajo

www.ine.gob.cl/prensa/2022/02/22/un-10-9-de-personas-trabajadoras-realiz%C3%B3-labores-a-distancia-o-teletrabajo-en-diciembre-de-2021