

Construcción de una Matriz OD del Transporte Público en RMBA en base a datos SUBE

Felipe González

Sebastián Anapolsky

1 - Resumen ejecutivo	1
2 - Introducción	4
2.1 - Los datos SUBE	5
2.2 - Matrices Origen Destino	7
3. Metodología	8
3.1 - Antecedentes	8
3.2 - Metodología de elaboración de una matriz OD en base a datos SUBE	11
3.2.1 - Preprocesamiento de la base de datos	12
3.2.2 - Imputación de destinos	14
Validación customizada a RMBA con cartografía de paradas	18
Validación simplificada y generalizable	26
3.2.3 - Cálculo de distancias entre orígenes y destinos	27
3.3 - Datos contextuales	28
3.4 - Repositorio y modelo final de datos	29
4 - Principales Resultados	31
4.1 - Conversión de transacciones a viajes y etapas	31
4.2 - Orígenes y destinos en la RMBA	33
4.3 - Orígenes destinos por partidos y zonas	34
4.4 - Resultados de viajes por cantidad de etapas y modos	38
4.5 - Resultados de viajes por distancias	43
5 - Próximos pasos	45
6 - Bibliografía	47

1 - Resumen ejecutivo

- Elaborar Matrices Orígenes Destino en base a datos SUBE tiene algunas ventajas con respecto a encuestas tradicionales. Permite contar con matrices de mayor cobertura espacial y temporal del sistema de transporte público, mayor exactitud y granularidad geográfica con menores tiempos de elaboración, procesamiento y menores costos.
- Sus principales desventajas son la necesidad de reestructurar y preprocesar los datos administrativos existentes, la necesidad de validación exógena con conteos o encuestas preexistentes, los posibles sesgos o errores sistemáticos en los datos, la ausencia de modos privados y activos y, por último, la falta de datos sociodemográficos y otros datos como la declaración de actividades que motivan el viaje, la localización del hogar o la distancia de caminata a la red de transporte público.
- De las casi 14,9 millones de transacciones originales de un miércoles de noviembre de 2019, se consolidó una base de 9.2 millones de etapas (61,7%). El rubro que más explica este primer filtro de datos, además de no considerar las transacciones tipo *checkout* de tren como etapas, es la falta de un id tarjeta (14,9%), seguido por las transacciones que son únicas en el día y, por ende, imposible imputar un destino con los supuestos utilizados (9,0%) y, por último, problemas con el GPS o la cartografía de paradas (5,9%). Sobre las 9.3 millones de etapas, se logró un buen porcentaje de éxito en la imputación de destinos consistente con la literatura existente: 92,1% de destinos imputados en etapas y 89,1% en viajes.
- Los viajes internos en CABA explican el 30% del conjunto de viajes, mientras que los viajes internos en PBA explican el 48% de los viajes y los viajes entre CABA y PBA son el 11%. Los partidos con mayor participación en estos últimos son: La Matanza, Lomas de Zamora, Vicente López, Lanús, Avellaneda, Almirante Brown, Tres de Febrero, General San Martín, Quilmes, Merlo. Los partidos que explican la mayor cantidad de viajes intra partido (sin contar a CABA) son La Matanza, Almirante Brown, Lomas de Zamora y La Plata.
- Se obtiene un promedio de 2,15 viajes por tarjeta. El 68% de los viajes es de una etapa, el 28% de dos etapas y solo el 6% de tres o más. Esto cambia al analizarlo por origen y destino, dado que el grupo más frecuente es el de los viajes entre CABA y PBA son de 2 etapas (45%).
- En términos de reparto modal, el 73% de las etapas son en colectivo, el 15% en tren y el 12% en subte. Este reparto también se ve modificado al ver los viajes, que combinan

varias etapas y analizarlo por origen y destino del viaje. En los viajes entre CABA y PBA el tren cobra mucho más protagonismo, dado que casi el 40% de los mismos incluyen al menos alguna etapa en tren.

- Al considerar las distancias de los viajes, se puede observar que el promedio cambia con el modo. Mientras que en colectivo y subte son distancias en torno a 10,8 y 11,2 km¹ respectivamente, en tren alcanzan los 19,7 km promedio. Al analizarlo por origen y destino se observa que los viajes entre CABA y PBA son los de mayor distancia promedio (23,4 km) mientras que los realizados en CABA tienen un promedio de 6,6 km y los realizados internamente en PBA 9,6 km.

¹ Estos son viajes que pueden tener más de una etapa. Por ejemplo, un viaje que incluye subte puede tener una etapa previa en tren

Listado de acrónimos y abreviaturas utilizadas

AFC - Automated fare collection o Cobro automático del boleto

CABA - Ciudad Autónoma de Buenos Aires

ENMODO - Encuesta de Movilidad Domiciliaria

INTRUPUBA - Investigación de Transporte Urbano Público de Buenos Aires

PBA - Provincia de Buenos Aires

RMBA - Región Metropolitana Buenos Aires

SRE - Sistemas de Recaudo Electrónico

SUBE - Sistema Único de Boleto Electrónico

TAZ - Transportation Analysis Zones o Zonas de Análisis de Transporte

2 - Introducción

El objetivo principal de este trabajo es refinar y expandir lo realizado en el contexto de publicaciones anteriores (Gonzalez & Anapolsky, 2021 y Gonzalez & Anapolsky, 2022) sobre metodologías que permiten utilizar los datos de las transacciones de pago electrónico en el transporte público y procesar los datos de la tarjeta SUBE para imputar destinos y construir diferentes matrices de origen-destino de etapas y viajes para la Región Metropolitana de Buenos Aires (RMBA)².

Junto a este documento, se publica un set de datos compuesto principalmente por tablas de etapas y viajes realizados en transporte público para noviembre de 2019, con su correspondiente destino imputado. En base al mismo se elaboraron diferentes matrices de Origen-Destino (matrices OD), como una muestra de un caso de uso. Este set de datos y las matrices extraídas del mismo constituyen insumos importantes para analizar los patrones de movilidad de la RMBA previo al impacto de la pandemia COVID y permitirá nutrir trabajos posteriores. A su vez, se disponibilizan los scripts de procesamiento de datos para que todo el proceso sea reproducible y fiscalizable.

En el contexto de esta publicación de datos y scripts, este trabajo pretende oficiar como un documento metodológico citable sobre las principales decisiones tomadas en el proceso, respaldadas en la bibliografía especializada sobre el tema y presente un resumen con la descripción de las principales tendencias de movilidad en la RMBA.

A continuación en este apartado se ofrecen algunas precisiones conceptuales sobre el Sistema Único de Boleto Electrónico (SUBE), sus datos, así como sobre la cobertura espacial del trabajo y las matrices OD generadas. En el siguiente apartado se hace un repaso de la bibliografía específica sobre imputación de destinos y elaboración de matrices OD a partir de datos provenientes de sistemas de recaudo electrónico (SRE), como también la descripción en detalle de la metodología aplicada en el caso puntual de la RMBA y la ubicación de los set de datos resultantes y los scripts que constituyen la aplicación concreta de la metodología descripta. En el apartado final se ofrece un resumen de las principales tendencias de movilidad en el RMBA en términos de orígenes, destinos y distancias.

² Existen numerosos criterios para realizar un recorte del continuo urbano de la región de Buenos Aires, que comprende a la Ciudad Autónoma de Buenos Aires (CABA) y a diferentes municipios de la Provincia de Buenos Aires (PBA). El INDEC (INDEC,2003) ofrece diferentes precisiones y diferentes criterios (Aglomerado Gran Buenos Aires, Gran Buenos Aires, Conurbano). Sin embargo, en la medida en que constituye un análisis sobre transporte público, tomaremos el criterio de la Resolución ST No 1903/2015 que amplía el Decreto N° 656/94, dictado con el objeto de regir toda prestación de servicios de transporte por automotor de pasajeros de carácter urbano y suburbano que se desarrolle en el ámbito de la Jurisdicción Nacional (es decir, sobre los cuales tenga atribución regulatoria la Nación por cruzar límites provinciales), que en su Artículo 3 considera a la RMBA como los partidos de: CABA, Avellaneda, Lanús, Lomas de Zamora, La Matanza, Tres de Febrero, General San Martín, Vicente López, Quilmes, Almirante Brown, Esteban Echeverría, Morón, Hurlingham, Ituzaingó, San Miguel, José C. Paz, Malvinas Argentinas, San Fernando, San Isidro, Tigre, Berazategui, Florencio Varela, Presidente Perón, San Vicente, Ezeiza, Merlo, Moreno, Pilar, Escobar, Ensenada, Berisso, La Plata, Brandsen, Cañuelas, General Las Heras, Marcos Paz, General Rodríguez, Mercedes, Luján, Exaltación de la Cruz, Campana, Lobos, Navarro, Gral. Belgrano y Zárate.

2.1 - Los datos SUBE

El Sistema Único de Boleto Electrónico (SUBE) es un sistema de recaudo electrónico, SRE o ADC por sus siglas en inglés (*automated fare collection*), como los que se utilizan en ciudades en todo el mundo en virtud de sus beneficios para el transporte urbano (Pelletier, 2011). Utilizados originalmente con el fin de facilitar el recaudo de la tarifa, representan una fuente continua, precisa e íntegra de datos sobre la demanda de transporte. Estos datos se producen en forma masiva (ya que suelen ser el único o principal modo de pago para viajar en transporte público) y, dado que contienen un gran nivel de detalle en cada transacción, permiten ser utilizados para comprender mejor la movilidad en nuestras ciudades.

Aunque los usuarios que adquieren una tarjeta tienen la posibilidad de registrarse (donde solicita cargar datos personales y sociodemográficos) esto no es obligatorio en el caso de Argentina, por lo que no se dispone de información sociodemográfica, excepto para aquellos usuarios que utilizan algún tipo de tarifa diferencial y requieren estar registrados (ej. Estudiantes, jubilados y usuarios con tarifa social). Las transacciones son realizadas en el momento en que el usuario aproxima la tarjeta a un validador, dispositivo físico presente en las estaciones o dentro de los modos de transporte. El sistema almacena los datos básicos de cada transacción (fecha y horario donde se origina la transacción), sobre la tarjeta (usuario, número consecutivo de viaje de la tarjeta), sobre la tarifa aplicada (tarifa vigente, tipo de tarifa pagada, subsidio) y sobre el servicio (estación, modo, línea, ramal, sentido, vehículo, etc.), todo lo cual varía de acuerdo a la implementación del sistema (Tenenbaum de Olivera et al, 2015). A su vez, estas transacciones pueden ser asociadas a los datos de GPS que suelen tener estas máquinas validadoras lo que permite identificar para cada transacción la latitud y longitud donde se produjo. Las ventajas de estos sistemas son numerosas. Por un lado, facilitan el proceso de movilidad en sí mismo al acelerar los tiempos de pago, permiten aplicar tarifas diferenciales según zonas y relevan al conductor de realizar dicha tarea (Pelletier, 2011). Por otro lado, su implementación permite la recopilación de una masiva cantidad de datos administrativos de dichas transacciones cuya potencialidad puede ser explotada a partir de un proceso de inteligencia de datos para analizar la movilidad a partir de mejor entendimiento de patrones de viajes y operaciones sobre el sistema de transporte (Calabrese, Colonna, Lovisolo, Parata, & Ratti, 2011; Mondschein, 2014).

La tarjeta **SUBE**, utilizada en las principales ciudades de Argentina, es un sistema SRE con las características mencionadas que permite el pago de transporte urbano de la Región Metropolitana de Buenos Aires (RMBA) en sus diferentes modos (tren, subte-premetro, colectivo, etc.). La fuente de datos utilizada en este trabajo procede de un pedido de información pública a Nación Servicios SA³. La base de datos original obtenida para un día miércoles de noviembre de 2019⁴ contiene 14,9 millones de registros o transacciones. Luego de una limpieza de datos, que profundizaremos más adelante, quedan para utilizar 10,3

³ La información fue solicitada por Felipe González a través de un pedido de acceso a la información pública en las condiciones de la Ley N° 27.275, bajo el expediente EX - 2020 - 32945006 - DNAIP#AAIP

⁴ Se escogió ese día por no verse afectados los patrones de movilidad por los efectos de la pandemia de COVID

millones de registros que corresponden a transacciones de ingreso a un modo de transporte público con un id de tarjeta y sin problemas de geolocalización.

Los datos del pedido de información pública fueron entregados utilizando tres técnicas de anonimización para preservar el derecho de protección de datos personales:

- Mapeo de número de tarjeta: el *id_tarjeta* que figura en los datos no representa un número de tarjeta real, sino un número ficticio pero que garantiza que se mantenga en mismo número en todas las transacciones hechas por la misma tarjeta.
- Agregación espacial: para la geolocalización de cada transacción se redondearon los decimales de modo que cada transacción no quede georreferenciada con demasiada precisión a fin de poder vincularse con información contextual para desanonymizar al usuario. Esto suma un margen de error en el posicionamiento de la transacción de aproximadamente 111 metros con respecto a la posición original.
- Agregación temporal: con el mismo criterio se eliminaron los minutos de la transacción, dejando solo la hora. Si bien esto puede dificultar ordenar las transacciones en el tiempo, fue agregado un campo llamado *etapa_red_sube* que identifica si el viaje se considera que se realizó en etapas y que permite facilitar el orden de los mismos e identificar transferencias. Este atributo se incrementa cada vez que el usuario realiza una transacción dentro de la ventana de las dos horas con respecto a la primera transacción y se utiliza para el cobro diferenciado de tarifa de acuerdo a la política de trasbordos llamada Red SUBE.

Es importante distinguir entre las **transacciones** en el sistema SRE y los **viajes**. Una transacción es un registro administrativo propio del SRE consistente con un modelo de datos propio del sistema, sin necesidad de que responda a un marco conceptual propio de un análisis de transporte. Por otro lado, un **viaje** reviste un andamiaje conceptual y teórico: es un movimiento desde un punto de origen a un punto de destino (Ortúzar y Willumsen, 2011). El origen y el destino son lugares en que las necesidades de los usuarios son satisfechas por la realización de actividades (Munizaga et al., 2014). Cada **viaje** está compuesto por una o más **etapas** (en los casos donde los viajes se realizan con transferencias entre servicios de transporte), consideradas estas como un grupo de uno o más movimientos en un modo en particular. Los puntos de abordaje y descenso son las posiciones en que la etapa comienza y termina, respectivamente (Munizaga y Palma, 2012). Un viaje puede estar constituido por más de una etapa. Se puede caminar hasta la estación de origen del transporte público, descender en la estación de destino y caminar hasta el trabajo. También pueden tomarse más de un vehículo o modo. En este caso se consideran **trasbordos** para designar la continuidad entre las etapas en el contexto de un mismo viaje. Estos son consecuencia de la interacción entre la red de transporte y la satisfacción de la necesidad de los usuarios (Munizaga et al., 2014). Los sistemas SRE pueden no detectar todas las etapas del viaje, por ejemplo los desplazamientos peatonales, con lo cual no dan cuenta del conjunto del mismo. A su vez, un mismo viaje puede estar constituido por más de una transacción en el SRE en caso de transbordos entre servicios de transporte que utilicen dicho sistema. Por otro lado, numerosas transacciones en la tarjeta pueden no estar vinculadas a viajes (carga de saldo, intentos de pago fallidos que quedan registrados, etc).

El primer ascenso de un viaje puede ser llamado abordaje inicial; los abordajes siguientes pueden ser llamados abordajes intermedios (o transferencias entre modos o servicios de transporte). Similarmente, el descenso más cercano al destino del viaje será llamado descenso final y las bajadas anteriores, descensos intermedios. Durante el procesamiento, la detección acertada a partir de los datos de las transacciones entre abordajes iniciales o intermedios es altamente relevante para agrupar las etapas en viajes, lo que permite una caracterización de la demanda que sea representativa de la realidad (Tenenbaum de Olivera et al, 2015).

2.2 - Matrices Origen Destino

Una matriz origen-destino (**matriz O-D**) de viajes presenta en forma agregada la cantidad de viajes (o etapas de viajes) que se producen entre distintas zonas de una ciudad o región. Según el tipo de análisis que se quiera hacer se pueden realizar distintas zonificaciones, que pueden ser por ejemplo partidos, departamentos, comunas o barrios o alguna agregación geográfica que sea específica para el análisis de la movilidad. Esta matriz va a permitir comprender la demanda de movilidad entre las distintas zonas analizadas.

La matriz muestra la cantidad de viajes (o etapas) desde cada uno de los orígenes “*i*” hasta cada uno de los destinos “*j*”. En las filas se observan los viajes producidos o los orígenes, mientras que en las columnas los atractores o destinos. La matriz final tiene dimensiones $n \times n$, donde n = el número de zonas. En los estudios tradicionales de transporte, la generación de estas matrices corresponde con el segundo de los cuatro pasos del clásico modelo de transporte “modelo de 4 etapas” (*four-step algorithm*), posterior a la generación de viajes y anterior a las asignaciones de modos y rutas.

Tabla 2.2.1 Ejemplo de Matriz Origen Destino

Orígenes \ Destinos	1	2	3	j	n	P
1	T_{11}	T_{12}	T_{13}	T_{1j}	T_{1n}	P_1
2	T_{21}	T_{22}	T_{23}	T_{2j}	T_{2n}	P_2
3	T_{31}	T_{32}	T_{33}	T_{3j}	T_{3n}	P_3
i	T_{i1}	T_{i2}	T_{i3}	T_{ij}	T_{in}	P_i
m	T_{m1}	T_{m2}	T_{m3}	T_{mj}	T_{mn}	P_m
A	A_1	A_2	A_3	A_j	A_n	total viajes

Fuente: Ortuzar et al, 2011

Una matriz puede realizarse tanto para viajes como para etapas de viajes. Una matriz de viajes permite comprender el movimiento de las personas de acuerdo a sus necesidades de viajes (una persona realiza un viaje para acceder a una actividad determinada que justifica dicho desplazamiento) y una matriz de etapas permite entender el funcionamiento de la red de transporte público, ya que tiene en consideración no sólo el origen y destino de cada viaje sino el origen y destino de cada etapa de un viaje.

3. Metodología

3.1 - Antecedentes

Tradicionalmente, las ciudades suelen elaborar modelos de transporte (el más tradicional es el modelo de cuatro etapas) que permiten estimar estas matrices en forma general y para los distintos modos de transporte. Uno de los principales insumos para la realización de estos modelos son las Encuestas de Movilidad Domiciliaria que brindan información sobre la partición modal y preferencias de viajes de los ciudadanos. El área metropolitana de Buenos Aires cuenta con una Encuesta de Movilidad Domiciliaria (ENMODO)⁵ realizada en 2009–2010 por la Secretaría de Transporte del Gobierno Nacional (Movilidad en el Área Metropolitana de Buenos Aires). Las encuestas de movilidad domiciliarias se aplican a una muestra de hogares de todo un área y se consulta sobre los viajes realizados en un día de referencia (generalmente el encuestado responde sobre sus viajes del día anterior) y se entrevista a todos los integrantes del hogar.

Por otro lado, la Investigación de Transporte Urbano Público de Buenos Aires (INTRUPUBA) realizada por la Secretaría de Transporte del Ministerio de Planificación de la Nación durante los años 2006 y 2007 también elaboró una matriz O-D⁶. Este trabajo se basó en realizar encuestas de interceptación en el transporte público con métodos de recolección basados en muestras no probabilísticas.

Frente a los métodos tradicionales mencionados (encuestas domiciliarias y encuestas de interceptación en transporte público), la utilización de datos de los sistemas de pago electrónico de boletos en transporte público tienen algunas ventajas (Gordon et al, 2014; Pettelier et al, 2011; Tenenbaum de Olivera et al, 2015):

- Alcance: cobertura espacial y temporal extensiva del sistema de transporte público (en el caso de la SUBE es el único modo de realizar un pago en el sistema de transporte), en contraste con una encuesta de intercepción, que son relevamientos concentrados en períodos o recorridos específicos y basados en la elaboración de muestras no probabilísticas. Los datos de la SUBE nos permiten conseguir información de todos los días de la semana y en diferentes periodos temporales, lo que nos da la posibilidad de analizar diferentes días (por ejemplo fines de semana), temporadas de alto turismo,

⁵ <https://www.argentina.gob.ar/transporte/dgppse/publicaciones/encuestas>

⁶ <http://ondat.fra.utn.edu.ar/?p=1044>

grandes eventos o situaciones meteorológicas atípicas, líneas de transporte poco utilizadas u otras situaciones más específicas.

- Exactitud: los datos son recolectados automáticamente, siendo posible, por ejemplo, entrar en el detalle del lugar y de la hora/minuto de la realización de la transacción. En los métodos tradicionales, la información depende en gran parte de la percepción del usuario y suele ser difícil reconstruir cadenas de viajes completas con gran precisión.
- Costos: las encuestas son procedimientos de campo que suelen tener costos muy elevados y no suelen hacerse con mucha periodicidad. Los datos de transacciones están disponibles sin ningún costo adicional.
- Tiempo de procesamiento: con el soporte de herramientas de procesamiento y análisis, los datos pueden ser procesados rápidamente.
- Interferencia humana: dependencia menor del encuestador en campo o del usuario, lo que puede minimizar errores de declaración.
- Escala individual: métodos tradicionales realizan conteos agregados de pasajeros y se basan en procedimientos de muestreo.

En las desventajas puede citarse (Pettelier et al, 2011; Tenenbaum de Olivera et al, 2015; Munizaga y Palma, 2012; Munizaga et al, 2014); Chu y Chapleau, 2008).

- Estructura de datos: se requiere organizar y procesar los datos para su uso. Esto requiere contar con las herramientas computacionales y las metodologías apropiadas.
- Estos datos nos proveen información únicamente sobre el transporte público, mientras que las encuestas de Movilidad Domiciliaria nos brindan información sobre los patrones de viajes en modos privados y no motorizados, además de contar con información sociodemográfica y sobre los motivos de los viajes.
- Validación exógena: existe la necesidad de realizar encuestas para confirmar los análisis y el uso de las premisas adoptadas.
- Posibles sesgos: la evasión tarifaria no es captada por el SRE. Una misma persona utilizando diferentes tarjetas, más de una persona utilizando la misma tarjeta (viaje en grupo) y tarjetas transferibles pueden inducir sesgos en los datos. El poseedor de una tarjeta puede, ocasionalmente, utilizar otro medio de pago.
- Error sistemático: mal funcionamiento de un validador o de una transmisión puede generar problemas de captación y procesamiento de los datos;
- Eventos aleatorios y problemas de seguridad pública: estos tipos de eventos causan desvíos operativos que podrían confundir los análisis realizados.
- Ausencia de datos extra-transacción: no siempre existe información socioeconómica sobre el usuario. No hay información sobre el motivo del viaje. Transbordos entre las etapas del viaje y el destino final del viaje no son informados, aunque existen métodos para deducirlos, pudiendo ser evitado en “sistemas cerrados”, con control de entrada y salida.
- Posibilidad de acceso a los datos: Los gobiernos y empresas que gestionan estos servicios son renuentes a compartir la información y en muchos casos se requiere acudir a pedidos de acceso a información pública que llevan tiempo. A su vez, en muchos casos los datos se entregan incompletos y como resultado con información de baja calidad.

Uno de los principales problemas a los que nos enfrentamos a la hora de construir una matriz OD con datos de la SUBE es que contamos con información del origen de la transacción y no del destino. La SUBE es un sistema *tap-in* o *tap-on* únicamente, es decir, solamente activamos la tarjeta al subir al medio de transporte, no al descender del mismo. Si bien esto no es cierto para algunos trenes, sí lo es para los subtes y colectivos de la RMBA. Esto plantea un problema. Para cada viaje podemos saber el origen del mismo pero no su destino. Otros sistemas de tarjetas inteligentes utilizan el sistema *tap-in/tap-out* o *tap-on/tap-off* para los subtes también, como la Oyster card de Londres y SmarTrip de Washington DC. Sin embargo, no es así para los colectivos en ninguno de esos casos. Por otro lado, el sistema de subterráneos de la ciudad de Nueva York con la Metrocard tampoco implementa un sistema de ese tipo, debiendo utilizar la tarjeta solo al entrar al sistema.

La creciente adopción de sistemas automatizados de recolección de datos, sin embargo, está proporcionando a los operadores de tránsito grandes sets de datos desagregados, que pueden ser procesados para revelar información de viaje desde la perspectiva del pasajero. Sin embargo, esta destilación de información centrada en el pasajero no es trivial, ya que la mayoría de los sistemas automatizados de recolección de datos -como la recolección automatizada de tarifas (AFC por sus siglas en inglés *automated fare collection*) y la localización automatizada de vehículos (AVL por sus siglas en inglés *automated vehicle location*) han sido diseñados para otros fines (Gordon et al, 2014). Dada esta situación, se han planteado algunas metodologías para construir una matriz O-D a partir de tarjetas inteligentes en otros países. Dado que otros países tienen sistemas similares para buses (Oyster Card en Londres, SmartCard en Washington, Metrocard en Nueva York), existen algunos trabajos con soluciones propuestas para este problema.

El trabajo de Barry (2002) para la Metrocard en la ciudad de Nueva York sienta dos supuestos fundamentales para los trabajos subsiguientes:

- **Closest stop o Parada más cercana:** un alto porcentaje de pasajeros/as parten de la estación final de su viaje anterior para comenzar el siguiente.
- **Daily symmetry rules o Regla de la simetría diaria:** un alto porcentaje de pasajeros/as finalizan su último viaje del día en la estación donde comenzaron el primer viaje del día.

Estos supuestos fueron validados utilizando encuestas de movilidad en las respectivas ciudades donde realizaron estos trabajos. En la ENMODO 2009 se puede verificar que a más del 90% de los viajes en transporte público se accede caminando menos de 10 cuadras⁷.

Existen otros trabajos que han aplicado la misma metodología o con algunas modificaciones adicionales. Hanft y otros (2015) por su parte, utilizaron estas metodologías para la misma tarjeta en un análisis de las rutas de los buses vinculados al complejo urbanístico Co-op City en el barrio del Bronx. Gordon y otros (2013) utilizan estos supuestos para un análisis de la Oyster Card de Londres y agregaron datos de una aplicación llamada iBus que registra los tiempos de

⁷ Las cuadras en Buenos Aires tienen un promedio de 100 metros. Se usan cuadras porque así fue formulado en el cuestionario de la encuesta y es el dato al que se puede acceder.

llegada de los buses a las paradas. Estos autores a su vez agregan algunas complejidades adicionales al algoritmo como por ejemplo, eliminar casos cuyo registro siguiente en la tarjeta se encuentra en un lugar ubicado en la dirección contraria al sentido del viaje del bus inmediatamente anterior. También a la hora de decidir entre dos paradas de destino posibles por estar cerca de la parada donde se origina el viaje siguiente, calculan el tiempo de viaje a pie para cada una de las candidatas y eligen la que minimiza el tiempo de caminata. Finalmente, frente a criterios de corte absolutos en tiempo, plantea condiciones espaciales y temporales para determinar si un viaje constituye una mera etapa dentro de un viaje que incluye un transbordo, o si lo que sucede entre viajes es una actividad que generó y es el motivo detrás del viaje. En estudios de la región, se pueden mencionar los estudios de Farzin (2008) para San Pablo, Brasil y de Munizaga y Palma (2012) para Santiago de Chile.

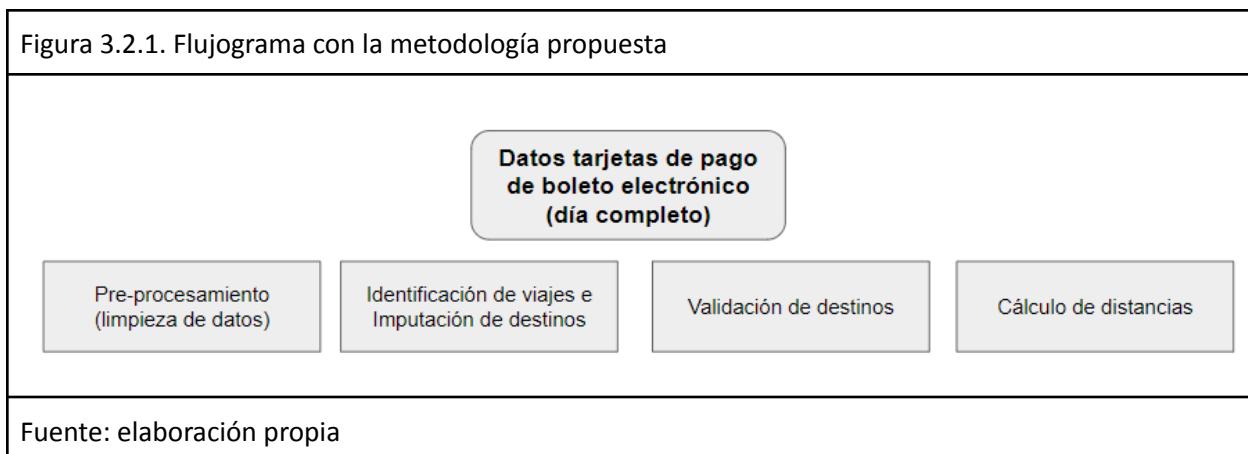
3.2 - Metodología de elaboración de una matriz OD en base a datos SUBE

En este apartado se detallan los pasos a seguir para poder construir la cadena de viajes de cada usuario del o los días completos que se están analizando. Con el fin de clarificar algunos conceptos, la tarjeta SUBE contiene registros de las transacciones de pago de los distintos sistemas de transporte público. Cada transacción no representa necesariamente una etapa de un viaje, dado que la misma tarjeta puede usarse para diferentes tipos de transacciones o el set de datos puede tener registrados diferentes tipos de transacciones (carga de saldo, errores de validación, otro tipo de compras, etc). En este trabajo nos enfocamos en las transacciones que refieren a las etapas de un viaje. En un viaje con transferencias, cada etapa representa un tramo que se realiza en un modo o línea de transporte distinto, ya que incluye alguna transferencia en algún momento del viaje. Como se verá más adelante, para cada etapa se imputará y validará un destino. Una serie de etapas de viaje es considerada que pertenece a un mismo viaje si el origen de la primera etapa y las subsiguientes está dentro de un rango temporal (en el caso de la RMBA es de 2 horas y se utiliza este rango para asignar una tarifa integrada⁸). Si se considera que un viaje fue realizado en distintas etapas de viaje, el viaje contiene el origen de la primera transacción y el destino imputado de la última transacción perteneciente a ese viaje. Una cadena de viajes es el conjunto de viajes realizados por un mismo usuario a lo largo de un día hábil.

En la siguiente figura se puede observar el fluograma de trabajo propuesto en este metodología:

⁸ Los viajes que se realizan en un rango de dos horas desde que se realiza la primera transacción pagan una tarifa integrada que tiene un valor mejor que la tarifa plena.

Figura 3.2.1. Flujograma con la metodología propuesta



3.2.1 - Preprocesamiento de la base de datos

El primer paso consiste en realizar una preprocesamiento de los datos donde se verifican los datos y se corrige la información que no es relevante o necesaria. Como se explicó al comienzo, no todo registro en la tabla de datos SUBE (transacción) conforma necesariamente una etapa de un viaje propiamente dicha, ni tampoco toda transacción que pueda considerarse una etapa puede ser utilizada para imputar un destino. En esta etapa se pueden identificar y eliminar los registros que no corresponden a etapas (como por ejemplo carga de saldo en la tarjeta, aunque el set de datos utilizado no contenía ese tipo de transacciones) y los registros que, pudiendo considerarse etapas tienen alguna falta de información, por ejemplo, registros con id de tarjetas nulos o incorrectos, problemas en la geolocalización: ya sea que dicha información no está presente, está fuera del área de estudio o demasiado alejada de una parada de la línea en la que dice realizar la transacción.

En la tabla de transacciones suele suceder que hay registros donde está duplicado el id de la tarjeta, el modo y línea utilizado y la hora de la transacción. Si bien no es posible confirmar el motivo, es probable que esto suceda cuando una misma tarjeta abona el boleto de 2 o más personas (por ejemplo, un mayor pagando el boleto de un menor). En estas situaciones existen dos alternativas, la primera es directamente eliminar los registros duplicados y no tener en consideración a los acompañantes que pagan con la misma tarjeta. Una segunda opción es asignar un nuevo id de usuario a los registros duplicados para intentar recuperar la cadena de viajes del acompañante. Para que se puedan validar las transacciones con el nuevo id de usuario asignado es necesario que acompañe al usuario principal durante toda la cadena del viaje.

En este punto también se estructuran las transacciones en etapas y viajes. Una serie de etapas de viaje es considerada que pertenece a un mismo viaje si la transacción de la primera etapa y las subsiguientes están dentro de un rango temporal (en el caso de la RMBA es de dos horas y

se utiliza este rango para asignar una tarifa integrada⁹, de acuerdo al tamaño de la ciudad y sus tiempos de viaje este rango temporal puede variar). La cadena de viajes se construye teniendo en consideración el origen del viaje y tomando como destino imputado el origen del viaje siguiente, o para el último viaje del día se utiliza el primer origen del día. Como resultado se va a obtener una cadena de viajes validada para cada usuario donde una de viajes es el conjunto de viajes realizados por un mismo usuario a lo largo de un día hábil.

Figura 3.2.2 Identificación de la cadena de viajes

Datos SUBE					
id tarjeta	hora	etapa red sube	modo	lat	lon
1037034016	6	0	COL	-34,780	-58,830
1037034016	7	1	TRE	-34,664	-58,728
1037034016	7	2	COL	-34,650	-58,790
1037034016	10	0	COL	-34,436	-58,566
1037034016	10	1	COL	-34,432	-58,564
1037034016	19	0	TRE	-34,650	-58,790
1037034016	20	1	COL	-34,664	-58,728



Fuente: elaboración propia

Por cuestiones de privacidad de la información, la base de datos contiene una hora pero sin minutos y es necesario ordenar las transacciones¹⁰. La identificación del orden de las transacciones (por no contar con la hora/minuto) dentro de la misma hora fue hecha con el atributo *etapa_red_sube* que se utiliza para la tarifa integrada. Si bien para la mayoría de los casos esto es suficiente para ordenar las transacciones en el tiempo, existen algunos casos de

⁹ Los viajes que se realizan en un rango de dos horas desde que se realiza la primera transacción pagan una tarifa integrada que tiene un valor menor que la tarifa plena. En el caso de los datos de la tarjeta SUBE, las transacciones que pertenecen a un mismo viaje se encuentran ya identificadas a través de una variable ya incluida en la base de datos. (*etapa_red_sube*). Este campo es utilizado para la tarifa integrada

¹⁰ La base de datos de SUBE para el RMBA no contiene el timestamp completo y solo incluye la hora de la transacción sin los minutos.

borde donde se dificulta. Cuando una persona realiza 2 transacciones en la misma línea dentro de la misma hora, dado que no existe integración tarifaria en ese caso, esto hace que no quede registrado el orden en el atributo *etapa_red_sube*. Este caso es especialmente frecuente en el subte, que puede usarse para realizar una actividad de poca duración en el tiempo y regresar. Para eso, ordenamos las transacciones en el tiempo en base a la distancia con respecto a las transacciones posterior o anterior. Cuando haya dos orígenes de etapas en subte, en la misma hora y por ende sin valores diferentes en *etapa_red_sube*, se pondrá como primera etapa aquella que minimice la distancia con la etapa anterior a estas dos etapas en subte y se pondrá como segunda etapa aquella que minimice la distancia con la etapa posterior a estas dos transacciones en subte. Otro problema surge cuando en una misma hora una persona desciende del tren con un *checkout* y comienza otra etapa pero por fuera de la ventana de tiempo de dos horas, por lo tanto, esta etapa aparece consignada como un nuevo viaje. Si bien esto puede ocurrir con cualquier etapa que debería ser integrada pero no lo es por el corte arbitrario de red SUBE, en este caso tenemos en una misma hora dos transacciones que al ordenar en el tiempo por la variable *etapa_red_sube*, la transacción correspondiente al *checkout* aparece como una transacción posterior al tener un 1 en dicho atributo (perteneciente a la etapa del viaje anterior) mientras que la siguiente transacción aparece con un 0 (se reinicia la ventana de tiempo de red sube). De todos modos estos casos son casos extremos y no representan a la mayoría.

El procesamiento de la base de datos se realiza sin modificar la estructura original de la tabla de transacciones para mantener la compatibilidad entre el set de datos original y el procesado. El campo *id* permite mantener una trazabilidad de cada registro en las tablas del resultado final con la tabla de transacciones que fue el insumo primario. Esto puede servir, entre otras cosas, para construir ponderadores que permitan expandir la tabla de etapas para alcanzar los totales reflejados en la tabla transacciones.

3.2.2 - Imputación de destinos

El siguiente paso consiste en imputar el destino de cada etapa. Para ello se utilizaron los principales supuestos utilizados los estudios previamente mencionados (Barry, 2012; Trepanier et al., 2007; Munizaga & Palma, 2012; Marshi et al., 2017):

- Los pasajeros realizan sus viajes utilizando transporte público que se encuentra razonablemente cercano a sus orígenes y destinos como para realizar el acceso a las paradas/estaciones en forma peatonal (**Closest stop o Parada más cercana**).
- Además, realizan transferencias entre distintas líneas/modos para minimizar la distancia caminable desde/hasta estaciones o para optimizar los tiempos de viaje.
- Por otro lado, en la mayoría de los casos, las personas inician y terminan sus viajes diarios en el mismo lugar, presumiblemente el hogar (**Daily symmetry rules o Regla de la simetría diaria**).
- A su vez, se presupone que las tarjetas electrónicas de pago son utilizadas por un mismo usuario a lo largo de un día.

Tomando estas suposiciones como válidas, es posible inferir y validar datos faltantes para un conjunto importante de transacciones, como el destino de los viajes y las transferencias realizadas.

Utilizando el supuesto de **Closest stop**, para cada id de usuario se procedió a imputar como destino de cada etapa el origen de la etapa siguiente. Para la última etapa del día, dado que no existe una etapa posterior, se utilizó el supuesto **Daily symmetry rules** y se consideró a la primera etapa del día como si fuese el destino de la última etapa. Como consecuencia de este método, no se puede utilizar ninguna tarjeta que haya realizado una única transacción en el día y por eso fueron eliminadas en el proceso anterior.

Datos SUBE						Datos imputados			
id tarjeta	hora	etapa red sube	modo	lat	lon	nº viaje	lat destino	lon destino	distancia
1037034016	6	0	COL	-34,780	-58,830	1	-34,664	-58,728	16,4
1037034016	7	1	TRE	-34,664	-58,728	1	-34,650	-58,790	5,8
1037034016	7	2	COL	-34,650	-58,790	1	-34,436	-58,566	34,1
1037034016	10	0	COL	-34,436	-58,566	2	-34,432	-58,564	0,8
1037034016	10	1	COL	-34,432	-58,564	2	-34,650	-58,790	34,8
1037034016	19	0	TRE	-34,650	-58,790	3	-34,664	-58,728	5,8
1037034016	20	1	COL	-34,664	-58,728	3	-34,780	-58,8830	16,4

Diagrama que muestra la identificación de la cadena de viajes. Se detallan los pasos y destinos para tres viajes separados:

- Viaje 1:** Inicia en Tx1 - colectivo 6.30, pasa por Tx2 - tren 7.10 y Tx3 - colectivo 7.40, llega a Destino 1 (Tx4 - colectivo 10.30). Continúa con Tx5 - colectivo 10.40, llega a Destino 2 (Tx6 - tren 19.15) y finaliza en Destino 3 (Tx6 - colectivo 20.05).
- Viaje 2:** Inicia en Destino 1 (Tx4 - colectivo 10.30), pasa por Destino 2 (Tx5 - colectivo 10.40), llega a Destino 3 (Tx6 - tren 19.15) y finaliza en Tx1.
- Viaje 3:** Inicia en Destino 2 (Tx6 - tren 19.15), pasa por Destino 3 (Tx6 - colectivo 20.05), llega a Destino 1 (Tx1) y finaliza en Tx1.

Fuente: elaboración propia

Una vez inferidos los destinos de cada transacción bajo los supuestos establecidos se procede con la validación de los destinos imputados. La validación de los destinos se realiza sobre la cadena de etapas.

Para la validación se va a tener que considerar que exista en el destino imputado, en un cierto rango de distancia, una parada de la línea utilizada en el origen del viaje. Si la distancia es superior a la tolerada no se considera esa imputación del destino como válida, perdiendo así la

posibilidad de construir la cadena de etapas y en consecuencia la cadena de viajes. Se recuerda que una de las hipótesis de esta metodología era que el destino de un viaje está a una distancia razonablemente cercana del origen de un próximo viaje¹¹. Cuando esta premisa no se cumple, la imputación del destino es considerada no válida.

El margen de tolerancia adoptado fue de 2 km, que en principio puede parecer excesivo si se piensa como la distancia a caminar a una parada o estación de transporte público pero su significado es otro. Por ejemplo, un usuario puede caminar 10 cuadras entre la parada en la que se baja del tren o colectivo y el destino final. Al terminar la actividad que explica el viaje, puede caminar otras tantas cuadras hacia otra parada para emprender otro viaje. Lo que utilizamos aquí es la distancia entre esas dos paradas, la del destino del primer viaje y la del origen del segundo. A su vez, cabe recordar que todas las coordenadas de las transacciones SUBE en el pedido de información pública han sido agregadas hasta al tercer decimal, lo cual suma un margen de error entre la localización real de la parada y la informada en la transacción.

Existen diferentes metodologías para realizar esta validación. Una validación completa requiere elaborar una cartografía de las líneas y paradas de toda la red de transporte público y verificar para cada destino imputado la distancia a estas paradas del modo y línea de transporte utilizado en el origen del viaje. Se puede realizar una validación más simple que no requiere emplear una capa geográfica con todas las líneas y paradas de la red de transporte público. En este caso, se puede construir dicha cartografía de paradas en base a las transacciones registradas en la tabla de etapas. El inconveniente con utilizar una cartografía oficial de paradas es que no siempre los colectivos realizan los recorridos declarados, en cuyo caso, su recorrido real será diferente y no siempre tendremos paradas. Del otro lado, construir la cartografía de paradas en base a las transacciones efectivamente realizadas tiene el inconveniente de que los errores de GPS pueden pasar desapercibidos, estableciendo paradas donde no existen y validando destinos que no son correctos. Una mejor cartografía de paradas puede hacer que se pierdan menos casos en las validaciones e imputar un mayor porcentaje de destinos.

Para el procedimiento de imputación de destinos se utiliza un concepto ad-hoc de línea y ramal. En el caso de los colectivos, si bien una persona se sube a un ramal y cualquier transbordo con otro ramal o línea queda consignado en la tabla de transacciones de la tarjeta SUBE, conformando otra etapa, la información para el ramal y el sentido no es confiable (se ingresa manualmente por los choferes y hay un significativo error en la declaración). Es decir, que en la tabla puede decir que es el ramal A cuando en realidad el vehículo se encuentra realizando el recorrido del ramal B. Por este motivo, los datos obtenidos para colectivos no pueden ser analizados a nivel ramal.

En el caso del subte consideramos a todas las líneas del subte como una única línea, y cada una de ellas (A, B, C, D, E, H, Premetro) como un ramal de la línea subte. Esto sucede porque una persona, por ejemplo, puede subirse en una estación de la A y bajarse en la C sin que ese

¹¹ En un rango de distancia que puede realizarse a pie.

trasbordo quede identificado de algún modo en la tabla de transacciones de SUBE. De este modo el algoritmo de imputación de destinos funciona del mismo modo cuando consideramos al subte como una única línea, ya que considera como posibles destinos a todas las estaciones del subte e imputa aquella que minimice la distancia con respecto al origen de la etapa siguiente.

Con el tren se consideró cada una de las líneas (Roca, Sarmiento, Belgrano Sur, Belgrano Norte, San Martín) como una línea con cada uno de sus ramales propiamente dichos como un ramal en nuestro modelo de datos. Una persona que se suba en la línea Roca podría bajarse en cualquier estación de esa línea, sin importar el ramal, sin que el trasbordo quede identificado en la tabla de transacciones SUBE. Por ejemplo, una persona podría subirse en estación Temperley, viajar hasta Avellaneda y luego ir a estación Sarandí, realizando una transferencia entre ramales de la misma línea. De este modo, este modelo hace que el algoritmo de imputación de destinos pueda imputar un destino correctamente incluso aunque haya un cambio de ramal.

Figura 3.2.4 Ramales de la línea General Roca

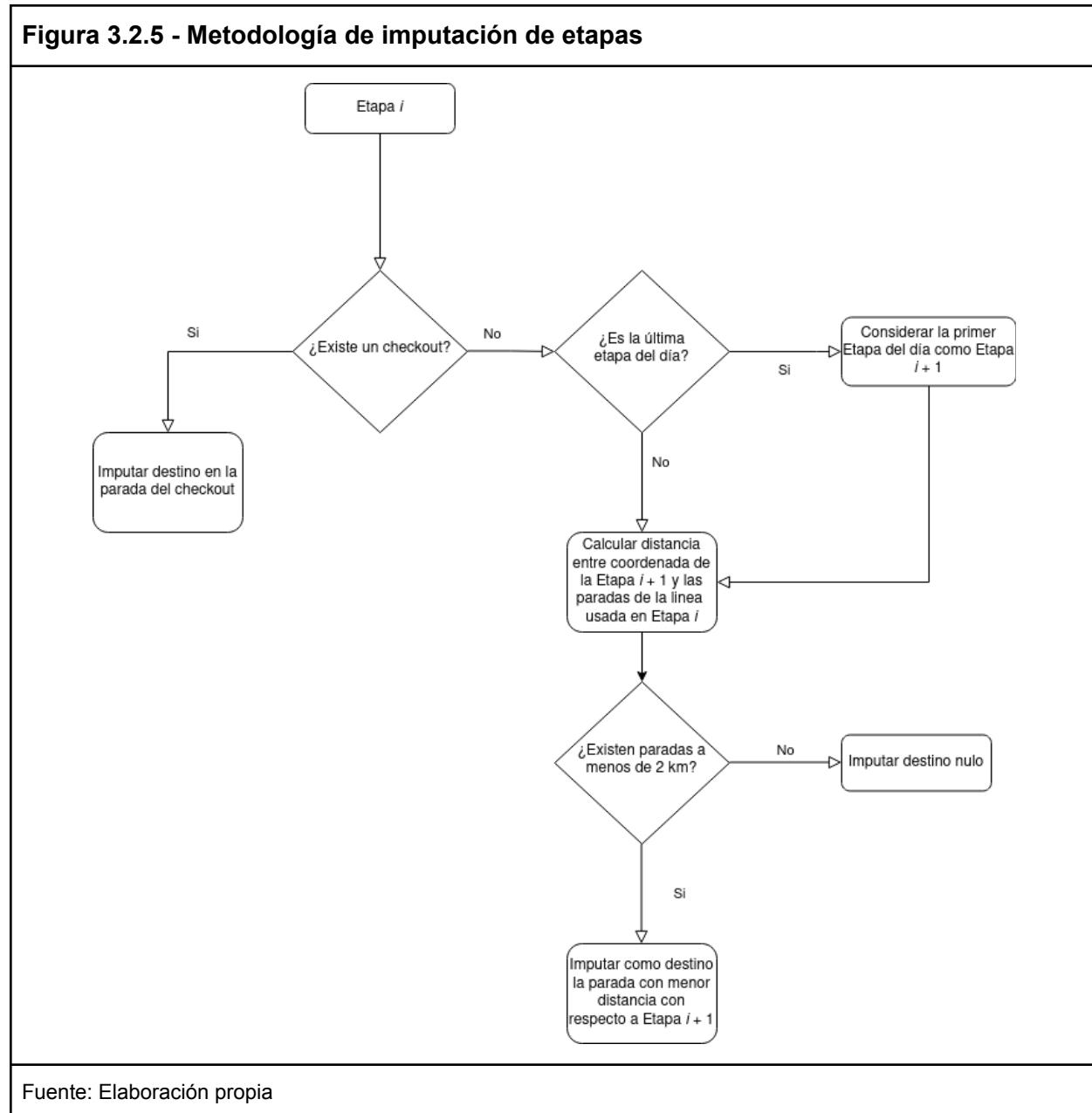


Fuente: Ministerio de Transporte de la Nación

A continuación se va a explicar con más detalle los diferentes métodos de validación y sus metodologías de implementación.

Validación customizada a RMBA con cartografía de paradas

La validación completa requiere tener una capa geográfica actualizada de todas las paradas de transporte público en el área de estudio. A continuación en la **Figura 3.2.4** se presenta un flujo de imputación de destino y un ejemplo práctico con un caso testigo.



En este procedimiento se tienen en consideración para el caso del ferrocarril las transacciones de destino que se encuentren registradas. Si existe una transacción de *checkout* se imputa la georeferenciación de esa transacción como destino de la transacción anterior. De no existir, se utiliza el método de imputación que supone que no existe transacción de *checkout*.

En la **Figura 3.2.6** se puede observar un caso testigo de la metodología de imputación. Dicha tarjeta realiza dos viajes entre Adrogué (PBA) y Colegiales (CABA), uno de ida y otro de vuelta. Cada viaje tiene tres etapas. El viaje de ida comienza a las 11 hs y el de vuelta comienza a las 20 hs. A cada etapa se le ha imputado una parada de origen y de destino en la línea correspondiente (salvo por las transacciones que en realidad constituyen un *checkout* del tren).

Figura 3.2.6 - Ejemplo de datos correspondientes a un mismo ID de tarjeta										
id_viaje	id_etapa	tipo_trx_tren	hora	modo	etapa_red_sube	id_linea	id_ramal	parada_id_o	parada_id_d	
1.0	1.0	None	11.0	COL	0.0	287.0	287.0	58727.0	58650.0	
1.0	2.0	CHECK IN	11.0	TRE	1.0	16.0	16.0	71123.0	71143.0	
1.0	2.0	CHECK OUT	12.0	TRE	1.0	16.0	16.0	71143.0	NaN	
1.0	3.0	None	12.0	SUB	2.0	32.0	269.0	71315.0	71333.0	
2.0	4.0	None	20.0	COL	0.0	223.0	223.0	12674.0	12614.0	
2.0	5.0	CHECK IN	21.0	TRE	1.0	16.0	16.0	71143.0	71130.0	
2.0	5.0	CHECK OUT	22.0	TRE	1.0	16.0	16.0	71130.0	NaN	
2.0	6.0	None	22.0	COL	2.0	287.0	287.0	58406.0	58632.0	

Fuente: Elaboración propia en base a datos SUBE

A continuación se resume en los siguientes mapas algunos de esos tramos para ilustrar el método de imputación. El círculo rojo indica el origen de la etapa y la cruz roja el destino imputado. El origen de la siguiente etapa aparece, cuando no está superpuesto totalmente con el destino, con un círculo verde. Finalmente los puntos son las diferentes paradas o estaciones de la línea utilizada en la etapa en cuestión.

Figura 3.2.7 - Origen y destino de la primera etapa del primer viaje de la tarjeta ejemplo

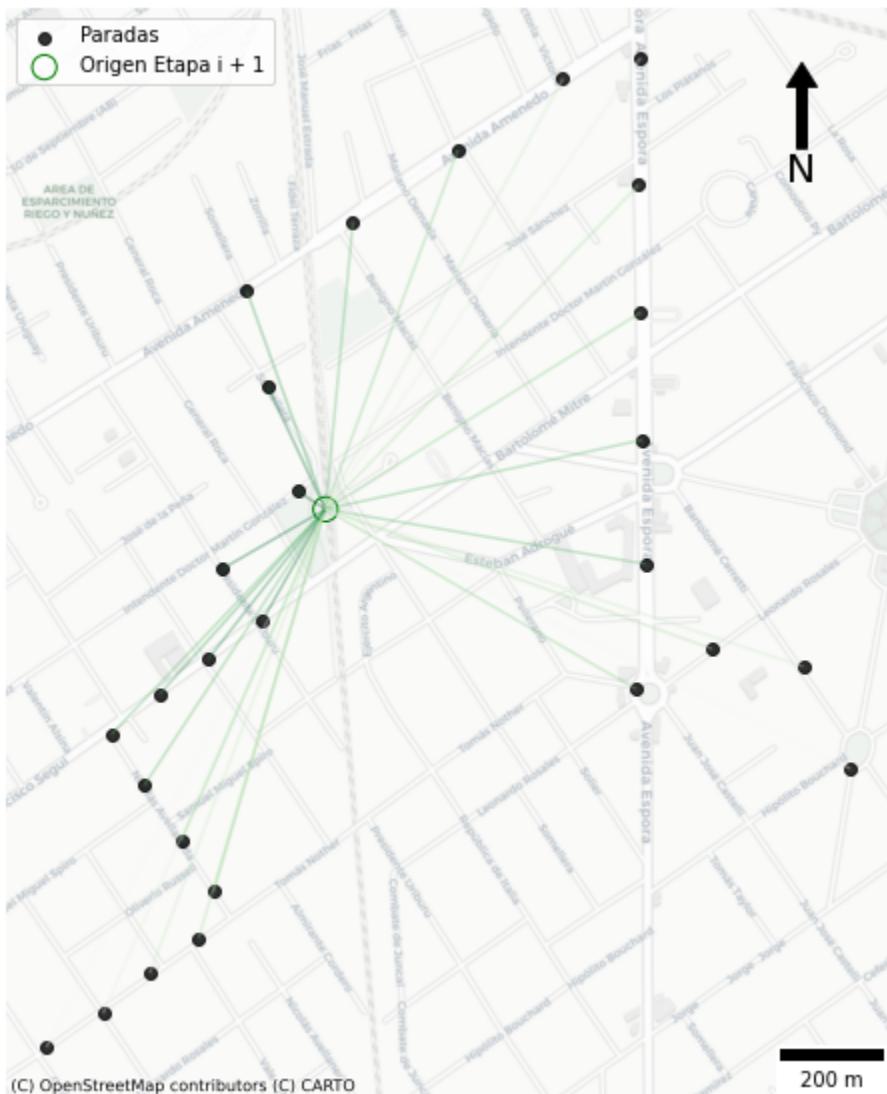


Fuente: Elaboración propia en base a datos SUBE

Se puede observar con un punto rojo el origen de la etapa en el barrio Adrogué. El origen de la siguiente etapa (en círculo verde) se ubica en la estación Adrogué del FFCC Roca. Por lo tanto, se imputó como destino de la primera etapa (con la equis roja) la parada de la línea de colectivo más cercana a la estación. La línea de colectivo que se toma en cuestión es la provincial 318, y el mapa muestra todas las paradas de todos los ramales ya que al tomar esa línea se asume que puede bajarse en cualquiera de las paradas de cualquiera de esos ramales.

En el mapa siguiente se puede observar todas las paradas cercanas a la siguiente transacción (la estación Adrogué) y las distancias respectivas.

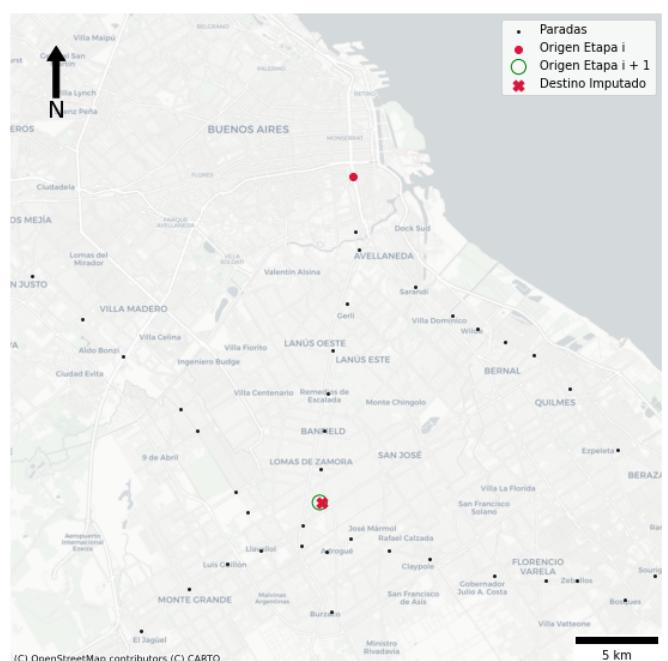
Figura 3.2.8 - Distancias de cada parada de la línea con respecto a la siguiente transacción SUBE

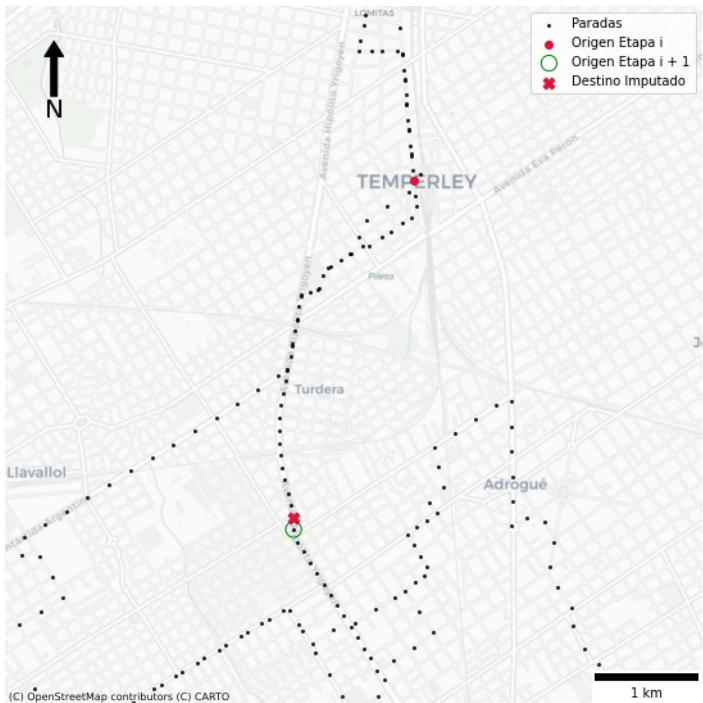


Fuente: Elaboración propia en base a datos SUBE

La segunda etapa la realiza en tren, en el FFCC Roca. Al registrar una transacción de tipo *checkout* en la estación Constitución, se imputó ese destino. Para el resto de las etapas se procedió con este mismo método para imputar los destinos.

Figura 3.2.9 - Origen y destino del resto de las etapas de la tarjeta ejemplo.





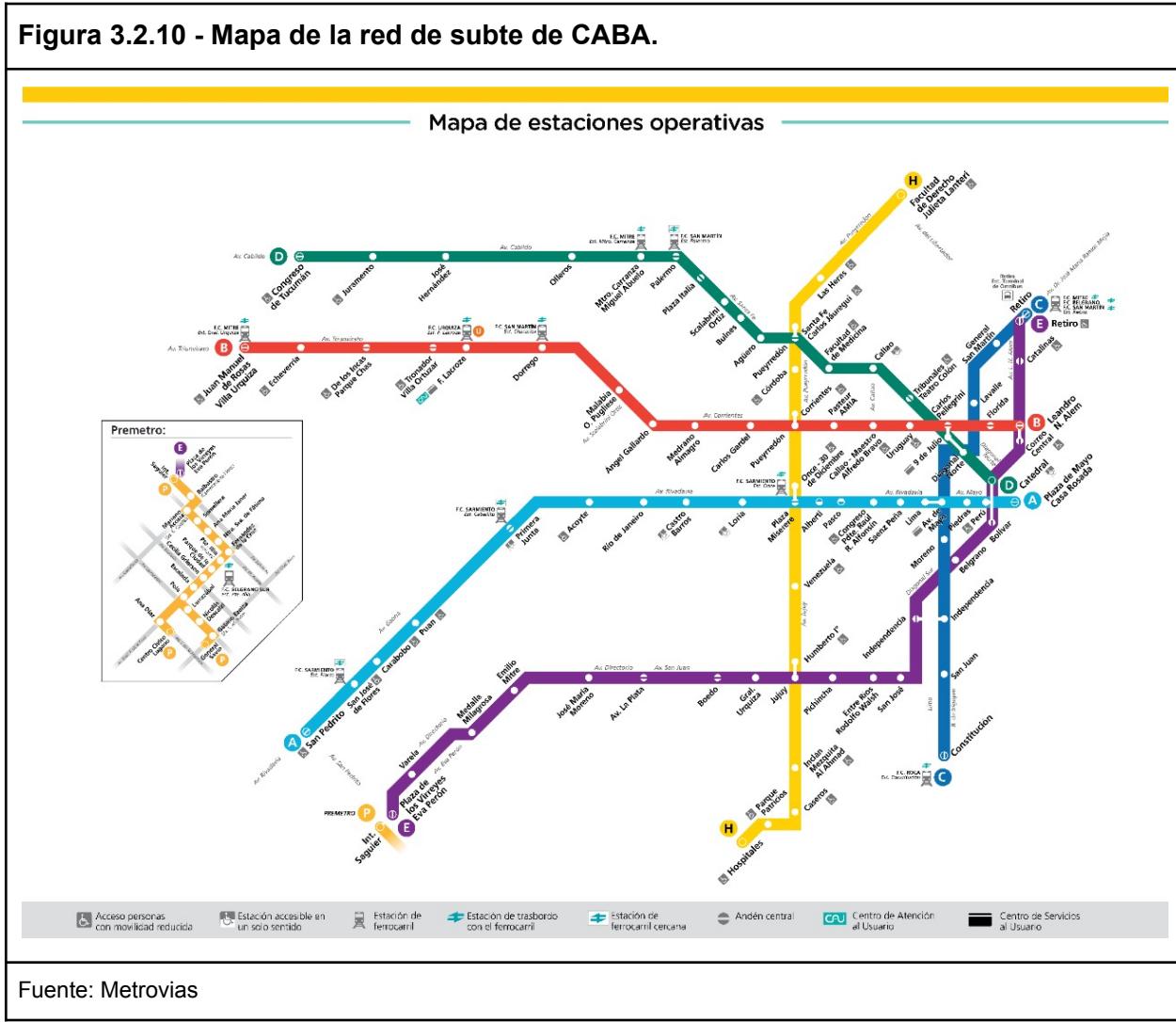
Fuente: Elaboración propia en base a datos SUBE

Como su siguiente etapa (la etapa tres) comienza en la estación Constitución de la línea C de subte también podría haberse imputado el destino en la estación Constitución del FFCC Roca teniendo como criterio que es la más cercana. La última etapa del primer viaje se hace en la Línea C de subte en combinación con la línea D entre las estaciones Constitución y Olleros. El destino de esta etapa está dado por la estación de subte (considerada como una única línea con varios “ramales”) más cercana al origen de la siguiente transacción (el origen de la primera etapa del viaje de vuelta). Esta primera etapa del viaje de vuelta se realiza en la línea de colectivo 59 hasta Constitución. El viaje de vuelta continúa con una etapa en el FC Roca de Constitución a Temperley, cuyo destino está asignado en base a la transacción de *checkout*. Finalmente la etapa final que cierra el viaje de vuelta se realiza en la línea provincial 318. Como esta etapa no tiene una transacción posterior, utilizamos la primera del día (supuesto de simetría). En este caso la línea provincial 318 tiene una parada dentro de la distancia tolerada con respecto a ese primer origen de la primera etapa del día, en el barrio de Adrogué.

Podemos listar algunas limitaciones de este método. En primer lugar, cualquier persona que tome un modo de transporte que no tenga SUBE (bicicleta, taxi, caminando, etc), tendrá desplazamientos que no registramos de modo que no siempre podremos imputar sus destinos correctamente. En el ejemplo de la última etapa del viaje hacia Belgrano que observamos (la etapa tres), la persona podría haberse bajado antes de Olleros, realizado alguna actividad y posteriormente caminado hasta la zona cercana a Olleros. De ese modo, hay una actividad en el medio que justificaría considerar que se trata de dos viajes diferentes. Esto no podría ser registrado por este método.

Por otro lado, puede suceder, por ejemplo con el subte, que haya dos estaciones a distancias relativamente cerca de la siguiente transacción, pero de diferentes ramales. Imaginemos una persona que toma la Línea B de Subte (ramal B para nuestro modelo) hacia el centro y su siguiente transacción SUBE se localiza en una zona cerca a las estaciones Florida (ramal B) y Lavalle (ramal C), estando esta última ligeramente más cerca. Puede suceder que esta persona decida bajarse en Florida y caminar un poco más, antes que perder tiempo en un trasbordo por solo una estación. Sin embargo este algoritmo imputará su destino en la estación Lavalle.

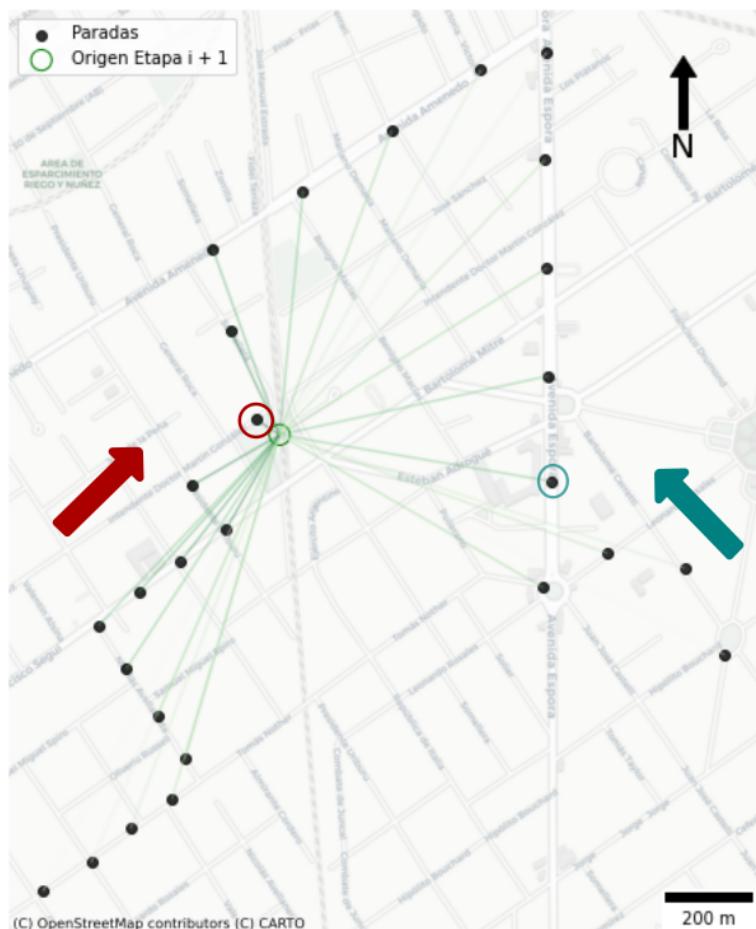
Figura 3.2.10 - Mapa de la red de subte de CABA.



Por último, un caso como el mostrado en el **Mapa 3.2.11** puede servir para ilustrar una potencial mala asignación de destino. La parada que minimice la distancia a la siguiente transacción en la estación de tren Adrogué, siempre será la misma sin importar desde dónde viene el viaje. Esto no simula el comportamiento más probable de un usuario. Alguien cuyo viaje viene desde el Sur-Este en el sentido de la flecha celeste, probablemente se baje en la

parada indicada con el círculo del mismo color. Quizás tenga que caminar más, pero se ahorrará tiempo. De este modo, seguramente y al igual que en el caso del subte, el usuario esté minimizando tiempos y no distancias.

Figura 3.2.11 - Distancias de cada parada de la línea con respecto a la siguiente transacción SUBE



Fuente: Elaboración propia en base a datos SUBE

Existen algoritmos que realizan esta minimización de tiempos, pero se necesita contar con la posición de cada unidad de todos modos y saber en qué unidad viaja cada usuario. Sin embargo, esto no es posible de realizar con estos datos SUBE ya que solo contamos con el interno de los colectivos y aún así no contamos con los datos de posicionamiento GPS de estos. Un procedimiento basado en minimización de tiempos fue realizado por Munizaga & Palma (2012) para Santiago de Chile.

Validación simplificada y generalizable

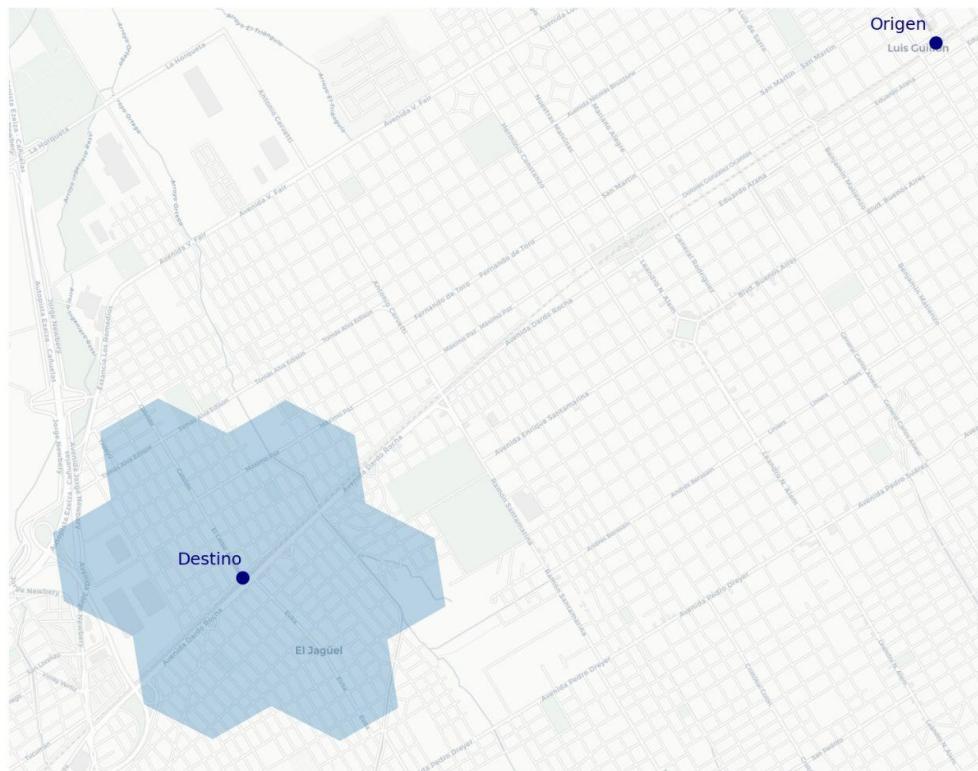
En los trabajos previos (Gonzalez & Anapolsky, 2021 y Gonzalez & Anapolsky, 2022) se utilizó una versión más simple de esta metodología que no requería emplear una capa geográfica con todos los recorridos y estaciones de la red de transporte público, por lo cual es plausible de ser generalizable a otras áreas urbanas con mayor facilidad. Este procedimiento produce igualmente muy buenos resultados de validación teniendo en cuenta la falta de información detallada de las paradas de cada línea para todos los modos.

El procedimiento se basa en generar un *buffer* en cada uno de los destinos y verificar si en ese *buffer* algún otro usuario utilizó un modo de transporte de la misma línea/ramal que el imputado en destino. Este método se basa en suponer que en todas las paradas imputadas como destino, al menos un usuario utilizó un servicio de esa línea/ramal. Es probable que este supuesto se cumpla en las zonas más densas de la ciudad donde el volumen de viajes es muy alto, pero puede que algunas zonas periféricas algunos destinos queden no validados por la falta de información. Una forma de mejorar la calidad de esta validación es utilizar información de varios días para la construcción de la matriz, lo que aumenta las probabilidades que en todas las paradas del sistema se haya registrado alguna transacción.

Las transacciones de tren fueron tratadas como cualquier otra transacción, sin utilizar los *checkout* para imputar el destino. En los trabajos citados, a las transacciones duplicadas se le asignó un nuevo identificador de usuario para no perder el segundo viaje.

La forma de implementación utiliza los orígenes de todas las operaciones SUBE y su georeferenciación para construir una matriz utilizando una capa geográfica de hexágonos, donde los hexágonos conforman un espacio que rodea cada destino imputado de viaje (*buffer*) de aproximadamente 1.500 metros (ver Figura 3.2.10). En cada *buffer* quedan establecidos todos los modos y líneas de transporte que tienen una parada en ese espacio. Esto permite validar cada uno de los destinos imputados, verificando que el modo y línea de transporte utilizado en origen exista en el área de influencia del destino imputado.

Figura 3.2.12: Ejemplo de cómo se valida el destino imputado



Fuente: Elaboración propia

3.2.3 - Cálculo de distancias entre orígenes y destinos

Para realizar un análisis de movilidad urbana es importante conocer las distancias de cada viaje. Esto es posible de calcular una vez que se pudo inferir y validar el destino de cada viaje. Existen distintas opciones para calcular estas distancias.

La opción más simple es calcular la distancia euclíadiana (a vuelo de pájaro) entre los orígenes y destinos de cada viaje. Este es un cálculo simple y rápido que se puede hacer a través de métodos que ya vienen establecidos en la librería H3¹² que se utiliza para trabajar con la cartografía de hexágonos.

Otra opción es utilizar librerías específicas que permiten realizar el cálculo de distancias entre puntos teniendo en cuenta la red vial existente. Estas librerías (como por ejemplo osmnx,

¹² <https://eng.uber.com/h3/>

Pandana o *Google Maps*) permiten calcular distancias entre orígenes y destinos teniendo en cuenta la red vial y los sentidos de circulación de acuerdo al modo de transporte que se utilice (ejemplo, automóvil o caminata). A su vez, se podría calcular la distancia sobre la red efectiva de dicho modo (tomando la red de trenes, subtes y los recorridos de colectivos). Pero esto implica un trabajo adicional sobre la cartografía de paradas que puede resultar muy complejo en el caso de los colectivos, dada la complejidad de sus recorridos en la RMBA.

3.3 - Datos contextuales

Para la validación del caso de la RMBA que requiere datos detallados de la red de transporte público para realizar la validación se necesito de información específica para complementar la información de las transacciones de la tarjeta SUBE.

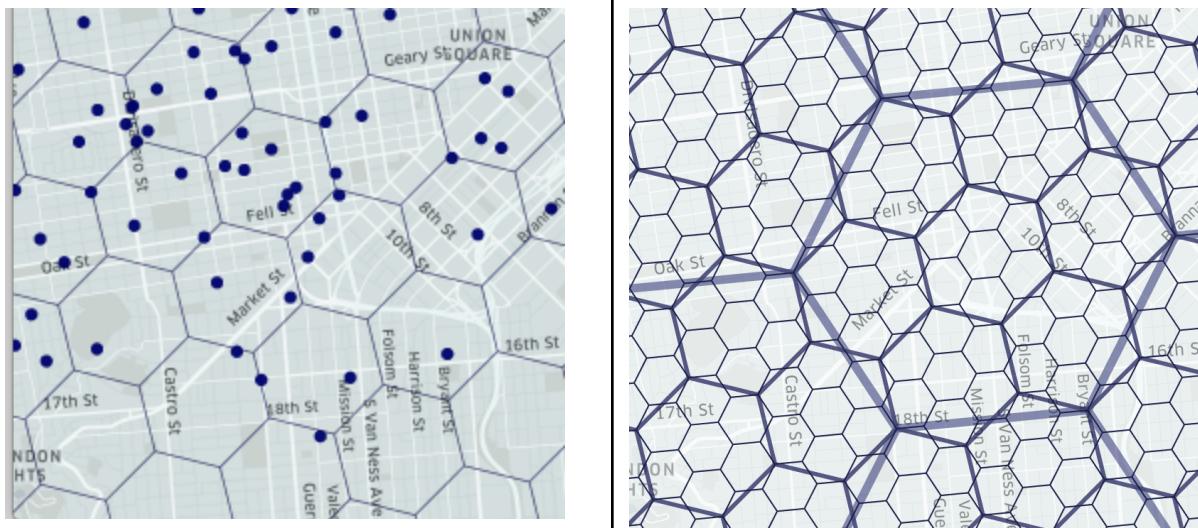
En primer lugar, se utilizó una tabla de las empresas de transporte público con las líneas y ramales que prestan para todos los modos (tren, subte y colectivo) para las tres jurisdicciones (Nacionales, Provinciales de Buenos Aires y Municipales), con IDs únicos correspondientes a las líneas y ramales, que coincidan con los IDs utilizados en la tabla de transacciones.

En segundo lugar, se procedió a obtener la información oficial de paradas de transporte público para los tres modos, consignando como línea y ramal lo especificado previamente. Para los buses se obtuvo el recorrido lineal (de acuerdo a los permisos concedidos por la autoridad regulatoria; no el efectivamente recorrido) de las líneas nacionales, provinciales y municipales. Para obtener paradas se interpolan paradas cada 400 metros¹³ sobre el recorrido lineal.

Por último, se procedió a georreferenciar cada par de coordenadas en el índice espacial H3 desarrollado por Uber con una resolución nivel 10 (un hexágono cuyo lado mide 66 metros). Este sistema de celdas hexagonales anidadas tiene la ventaja de poder realizar operaciones espaciales computacionalmente costosas (fundamentalmente distancias euclidianas) de un modo mucho más eficiente. Al mismo tiempo, una vez que cada origen y destino de las etapas y viajes se encuentra georeferenciado en H3, se pueden usar los diferentes niveles de agregación y resolución de las grillas como diferentes unidades espaciales de agregación para la matriz origen destino. Por último, cualquier información contextual que quiera vincularse con las etapas y viajes puede hacerse a través de los índices H3. Por ejemplo, para relacionar esta información con datos censales identificamos a qué radio censal (la unidad espacial censal más desagregada) pertenece cada celda H3 de resolución 10. A partir de esa identificación se puede relacionar cualquier información sociodemográfica que provea el censo.

¹³ Si bien las distancias entre paradas de colectivo en la RMBA se aproximan a los 200 m utilizamos 400 m porque de otro modo la cantidad de paradas a evaluar como potencial destino por el algoritmo se habrían duplicado, lo cual afectaría significativamente el costo computacional del mismo. A su vez, usar una distancia menor entre paradas tendría como efecto positivo una mejora en la precisión espacial de la imputación del destino, pero que no sería aprovechada dado que los datos utilizados cuentan con sus coordenadas redondeadas, lo cual garantiza mayor anonimidad de la información persona pero también causa una pérdida de precisión espacial.

Figura 3.3.1 - Ejemplo de esquema de celdas hexagonales anidadas H3.



Fuente: Uber

3.4 - Repositorio y modelo final de datos

Con el fin de garantizar la reproducibilidad del trabajo y su fiscalización por la comunidad de investigadores, se ha publicado tanto el set de datos original, como los procedimientos que llegan a los resultados aquí publicados. Los datos se encuentran publicados en el repositorio de Github del Banco Interamericano de Desarrollo¹⁴, tanto el insumo principal (la tabla de transacciones original de SUBE) como las tablas de etapas y viajes producto del resultado de imputación de destinos. Los procedimientos para elaborar las matrices finales a partir de los datos SUBE se encuentran en forma de queries de PostgreSQL en el repositorio.

El modelo final de datos cuenta con las siguientes tablas:

- **paradas:** identifica cada estación del sistema de transporte público de la RMBA con su ID de línea y ramal (entendiendo por línea y ramal lo descrito en el apartado previo).
- **lineas_ramaes:** identifica las líneas y ramales que operan en la RMBA.
- **etapas:** identifica las etapas (clasificadas en viajes) con el ramal, línea y modo, como así también el id de la parada de su origen y destino (cuando haya sido posible imputar).
- **viajes:** identifica los viajes, sus orígenes y destinos, la cantidad de etapas, los modos utilizados y si alguna de las etapas que lo componen tiene algún destino sin imputar.

¹⁴ <https://github.com/EL-BID/Matriz-Origen-Destino-Transporte-Publico>

- **tarjetas:** identifica todas las tarjetas con el índice H3 de su primera transacción del día (que puede ser usado como proxy de la ubicación del hogar de la tarjeta) y el sexo del usuario cuando la tarjeta contenga dicha información.
- **indices_h3:** identifica el índice H3 de resolución 10 y el radio censal al que pertenece, el partido y la provincia. Para esta tabla se utilizaron sólo los partidos pertenecientes a la RMBA. En la tabla **etapas** hay transacciones que pueden no pertenecer a ninguno de estos partidos, ya que no pertenecen a la RMBA. Se dejaron en la tabla para que quede a criterio de cada usuario si utilizarlos o no.

Figura 3.4.1. Modelo de datos



Fuente: Elaboración propia

4 - Principales Resultados

4.1 - Conversión de transacciones a viajes y etapas

Como se especificó al principio, no toda transacción del sistema SUBE consiste en lo que conceptualmente puede definirse como viaje o etapa. Es necesario acomodar las transacciones, producto de una necesidad administrativa de cobro, en un modelo conceptual de viajes y etapas de viaje más propio de un análisis de transporte.

Una vez que se realiza la limpieza, el preprocesamiento y la asignación de destinos en las etapas de viajes, se procede la generación de las cadenas de viajes. A continuación se muestran los resultados del proceso en sus distintas etapas.

En primer lugar se procede a eliminar todas las transacciones que no pueden utilizarse en el proceso de imputación de destinos. El principal problema es la falta de un identificador único para las tarjetas (14,7% de las transacciones originales se pierden por este problema), seguido de tarjetas con una única transacción en todo el día (9,0 %) y por último transacciones que dicen hacerse sobre una línea de transporte público pero se encuentran demasiado lejos de alguna de las estaciones de dicha línea (4,8%). Finalmente, el proceso termina con poco más de 10 millones de transacciones que pueden utilizarse para imputar destinos, de las cuales 1,3 millones son checkouts.

Tabla 4.1.1 Depuración de transacciones	
Concepto	% transacciones eliminadas
Transacciones originales	100 % (14.922.654)
Sin id tarjeta	14,7 %
Problemas de geolocalización	1,1 %
Transacciones duplicadas	0,1 %
Transacciones únicas por tarjeta	9,0 %
Check outs in check in	0,4 %
Transacciones demasiado lejos de línea de origen	4,8 %
Transacciones sin problemas (incluyendo check outs)	69.9% (10.435.957)
Fuente: Elaboración propia en base a datos SUBE	

Una vez creada la tabla de etapas, se procedió a imputar destinos de las mismas. Dentro de esas transacciones finales se encuentran también la transacción correspondiente al *checkout* del tren. Al hablar de etapas no consideraremos estas transacciones, ya que solo se utilizan como un indicador de destino de la etapa, por lo cual no se consideran una etapa en sí. La **tabla 4.1.2** muestra el porcentaje de éxito al imputar un destino para cada etapa, viaje y usuarios. Como se puede observar, se pudo imputar con éxito un destino al 92 % de las etapas. Cuando agrupamos las etapas en viajes, obtenemos 6,5 millones de viajes, para los cuales pudimos reconstruir el total de los destinos en un 89,1%, dado que si al menos una de las etapas del viaje tiene un destino sin imputar, no podemos considerar válido el viaje en su conjunto. Finalmente, de las 3 millones de tarjetas únicas, para el 73,4% se puede reconstruir la totalidad de las etapas y viajes. Es decir, todas las etapas de esa misma tarjeta tienen un destino correctamente imputado. Otros trabajos similares obtuvieron porcentaje de éxito de imputación entre el 71% (Zhao et al, 2007) y el 66% (Trepanier et al., 2007) de los viajes.

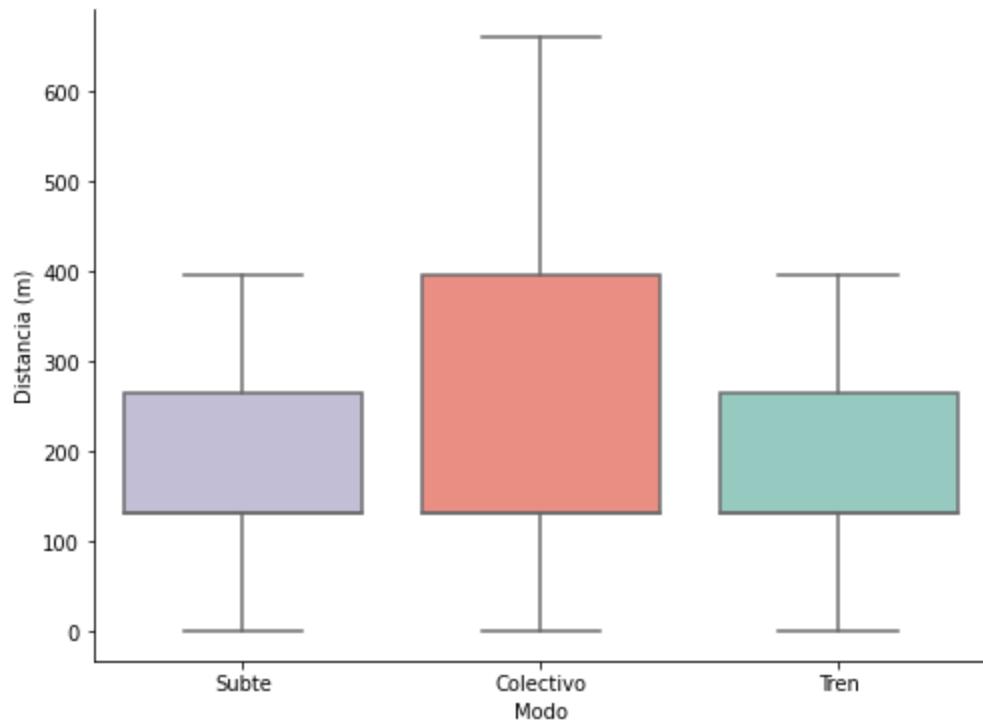
Tabla 4.1.2. Porcentaje de éxito en la imputación de destinos		
Tabla	Cantidad	Validados
Transacciones (con <i>check out</i>)	10.435.957	
Etapas (transacciones preprocesadas, sin <i>checkout</i>)	9.173.945	92,1 %
Viajes	6.509.176	89,1 %
Usuarios	3.022.985	73,4 % ¹⁵

Fuente: Elaboración propia en base a datos SUBE

Al observar para todas las etapas la distancia entre la parada a imputar como destino y la siguiente transacción del día (antes de eliminar aquellas que se encuentran más allá de la tolerancia de 2 km), vemos que los que superan los 2 kilómetros son verdaderos “outliers”. Incluso, se puede observar que la mayoría de los casos se encuentra dentro de los 400 metros. También se puede observar, cuando se desagrega por modo, que los colectivos tienen un rango mayor dado que su estructura de paradas es más irregular y a su vez pueden estar realizando recorridos no declarados y, por ende, no presentes en la cartografía oficial.

¹⁵ con cadenas de viajes completas

Figura 4.1.3. Distancia de la parada de destino imputada a la siguiente transacción por modo



Fuente: Elaboración propia en base a datos SUBE

4.2 - Orígenes y destinos en la RMBA

Como primer criterio de validación se procedió a comparar la matriz origen destino a nivel provincial con la obtenida en la encuesta de movilidad más reciente para el área (ENMODO 2009). Como la cobertura espacial de ENMODO 2009 no coincide con la de nuestro análisis, en los cuadros a continuación se presentan para la comparación la matriz OD considerando solo aquellos municipios que coinciden con los utilizados por la ENMODO.

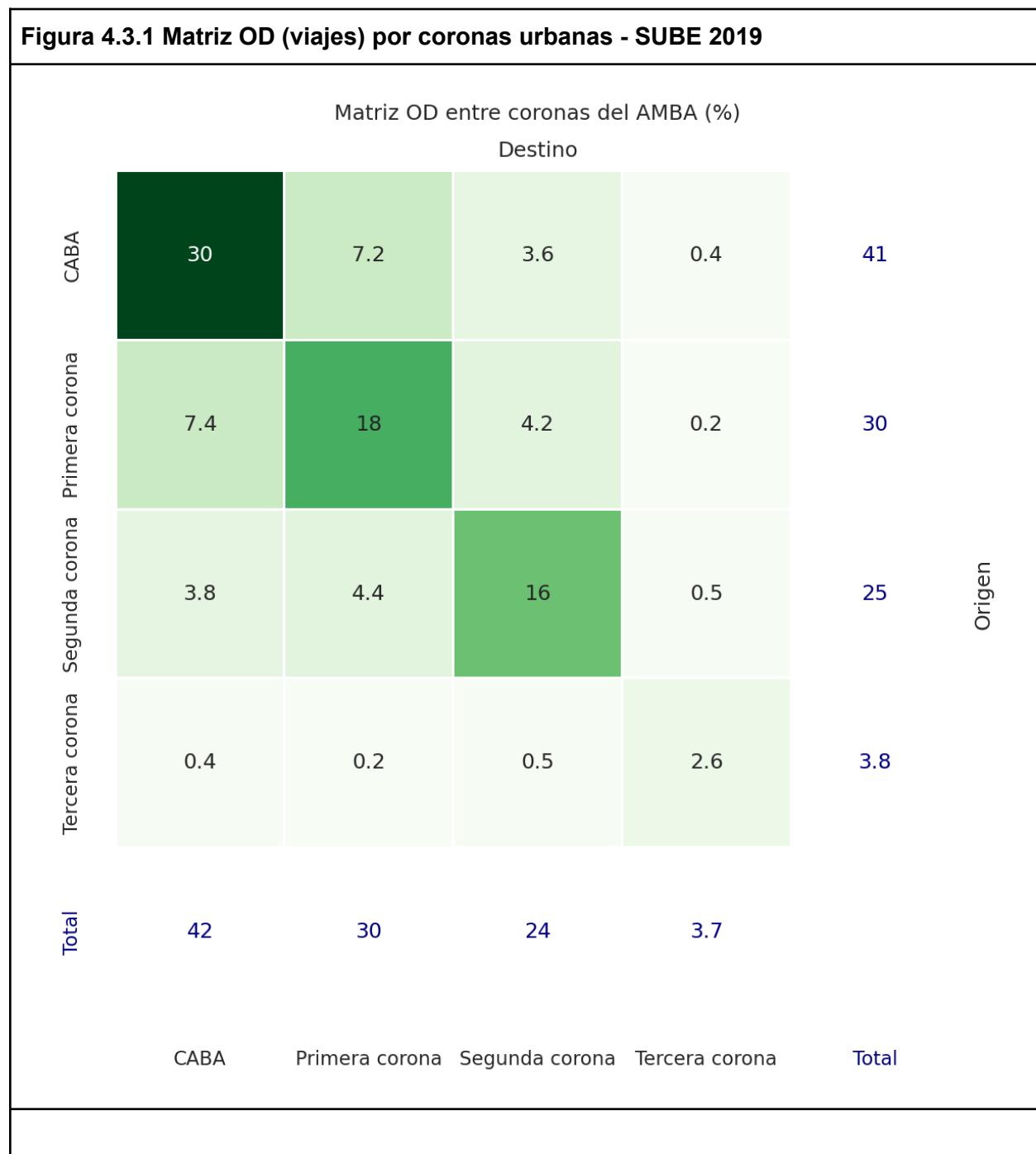
Viajes (%) por distrito de origen y destino - ENMODO 2009			Viajes (%) por distrito de origen y destino - SUBE		
Matriz OD entre CABA-PBA (%) - ENMODO			Matriz OD entre CABA-PBA (%) - SUBE		
	Destino			Destino	
CABA	26.2	11.1	37.3	CABA	31.3
PBA	11.3	51.4	62.7	PBA	11.5
Total	37.5	62.5		Total	42.8
	CABA	PBA	Total		CABA
					PBA
					Total
Fuente: Elaboración propia en base a datos ENMODO 2009 y datos SUBE 2019 *Incluye solo los municipios incluídos en la ENMODO 2009					

Como se puede observar, a pesar de los 10 años que separan la ENMODO (2009) de la fecha de los viajes de SUBE de nuestro estudio (2019), existe un patrón de movilidad similar en el área en comparación. Se puede mencionar como la principal diferencia una mayor concentración en los viajes dentro de CABA en 2019 (de 26,2% a 31,3%) en detrimento de los viajes dentro de PBA. En el contexto de este trabajo clasificaremos a los viajes por tipo del siguiente modo: por un lado los que suceden en el interior de alguna de las jurisdicciones (CABA o PBA) y por otro lado aquellos que tengan origen en una y destino en la otra.

4.3 - Orígenes destinos por partidos y zonas

Para lograr una mayor granularidad en el análisis construimos una matriz OD utilizando los cordones o coronas de la RMBA, una división no formal o administrativa pero que suele ser utilizada para el análisis. Estas coronas reflejan una continuidad geográfica en relación a CABA, donde la primera incluye a los partidos lindantes a CABA, la segunda el anillo sucesivo más alejado y la tercera es la más alejada. Se considera que estas coronas presentan cierta homogeneidad respecto de algunas características, como condiciones habitacionales, densidad poblacional, infraestructura, entre otras (Mariana Tallarico, 2020).

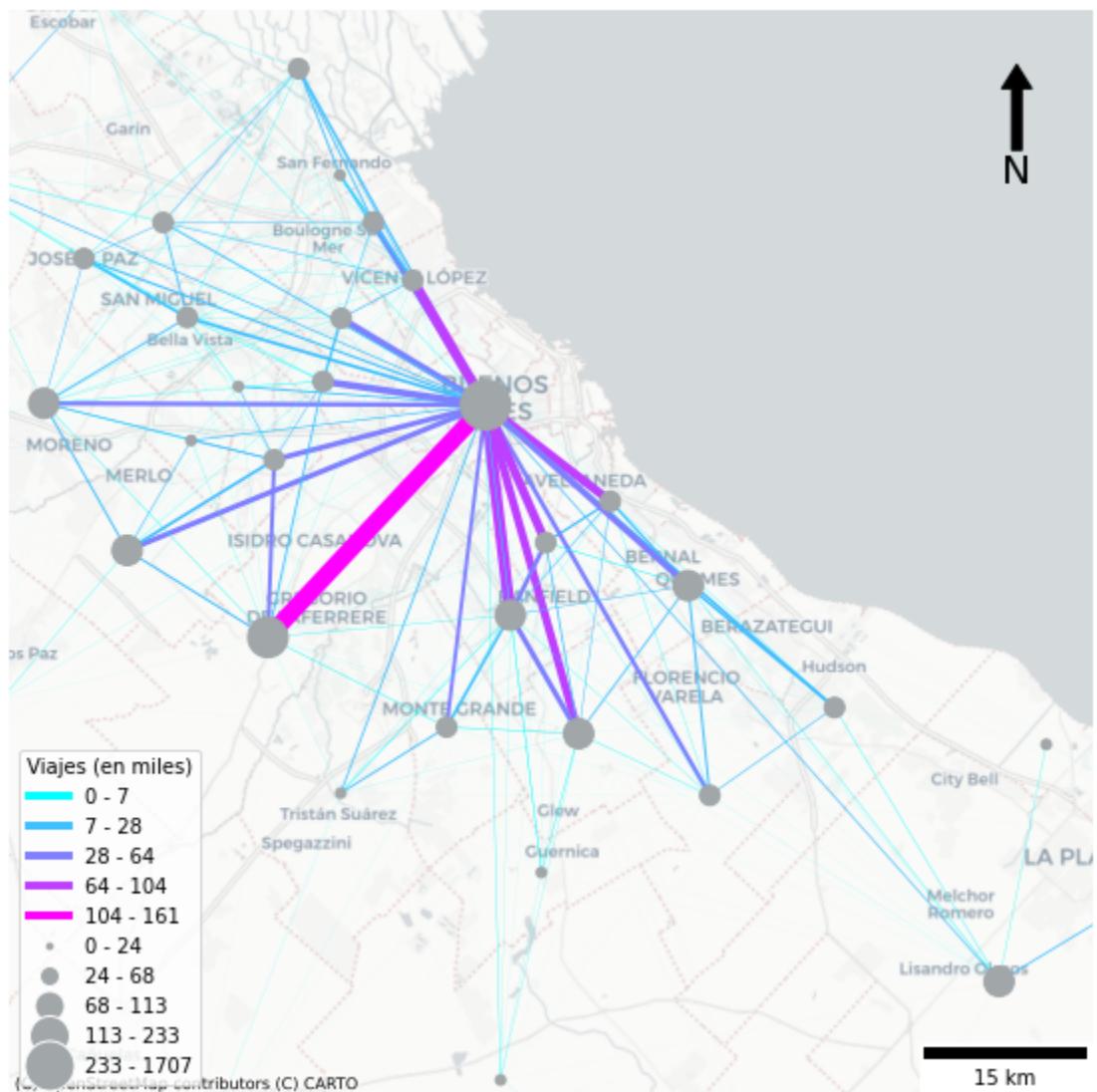
Como se puede ver en la siguiente matriz, alrededor del 40% del total de viajes tiene destino CABA y el 60% son viajes en los partidos del área metropolitana, con un 30% del total de viajes con destino en la primera corona, cerca del 25% en la segunda y cerca del 4% en la tercera.¹⁶



¹⁶ Incluye CABA y todos los partidos pertenecientes a la primera, segunda y tercera corona.

A su vez, podemos tomar los partidos de la RMBA como zonas de agregación geográfica. En ese caso. Se puede elaborar una visualización de estos patrones de viajes a partir de un mapa con líneas reflejando los viajes entre partidos (considerando a la CABA como un único partido) y puntos para considerar los viajes dentro de un mismo partido. En la **Figura 4.3.2** vemos que los principales corredores de viajes entre partidos se observan en aquellos distritos adyacentes a la CABA y que tienen a esta como origen o destino. Los principales partidos que ofician como generadores en este caso son La Matanza, Lomas de Zamora, Vicente López, Lanús, Avellaneda, Almirante Brown, Tres de Febrero, General San Martín, Quilmes y Merlo. También se puede observar que CABA, La Matanza, Almirante Brown, Lomas de Zamora y Quilmes son los distritos con mayor cantidad de viajes con origen y destino dentro del mismo partido. Es importante prestar atención a la leyenda del mapa y ver las magnitudes que cada símbolo representa. Mientras que el grupo de viajes entre partidos incluye hasta 7.000 viajes, el más pequeño de los viajes intra partidos llega hasta 24.000 viajes. A su vez, el grupo más grande de viajes intra partidos parte de 233.000 viajes, magnitud que no existe en los viajes entre partidos.

Figura 4.3.2 Matriz OD (viajes) por partido - SUBE 2019

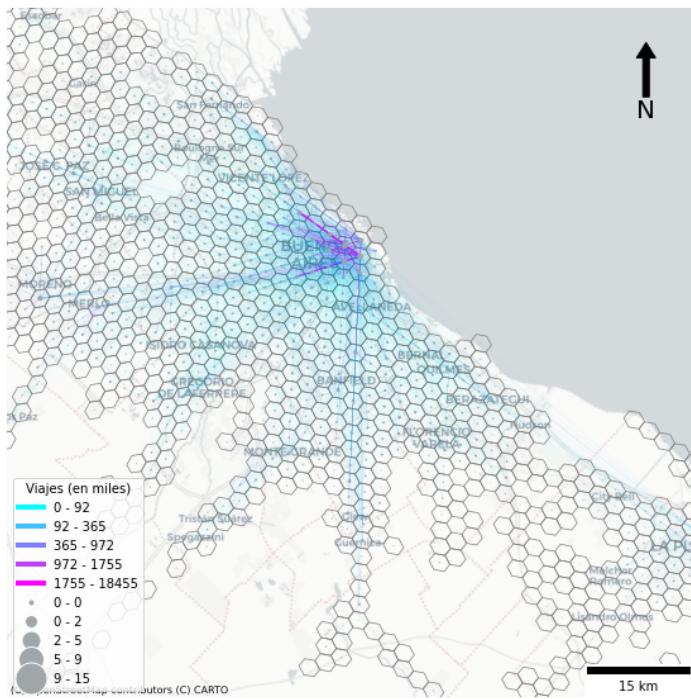
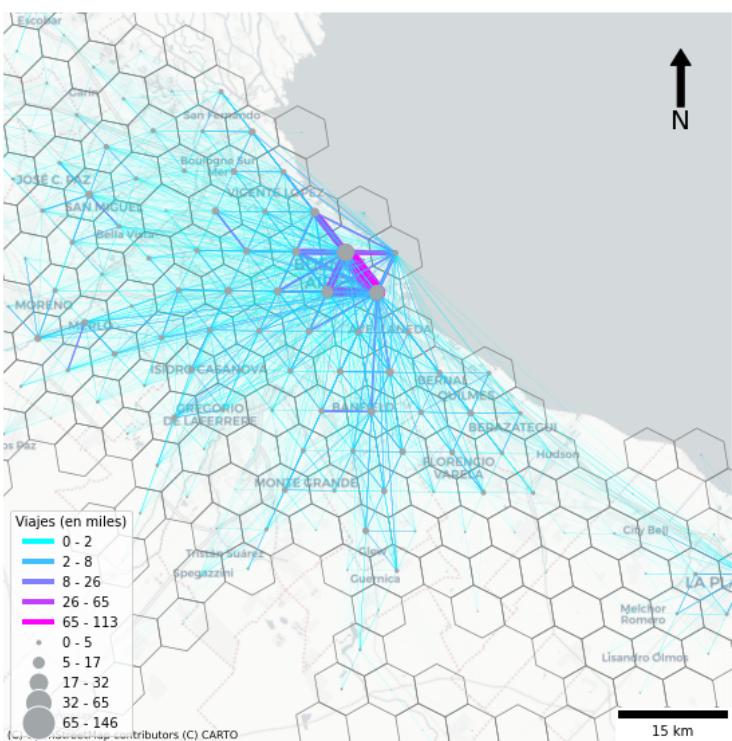
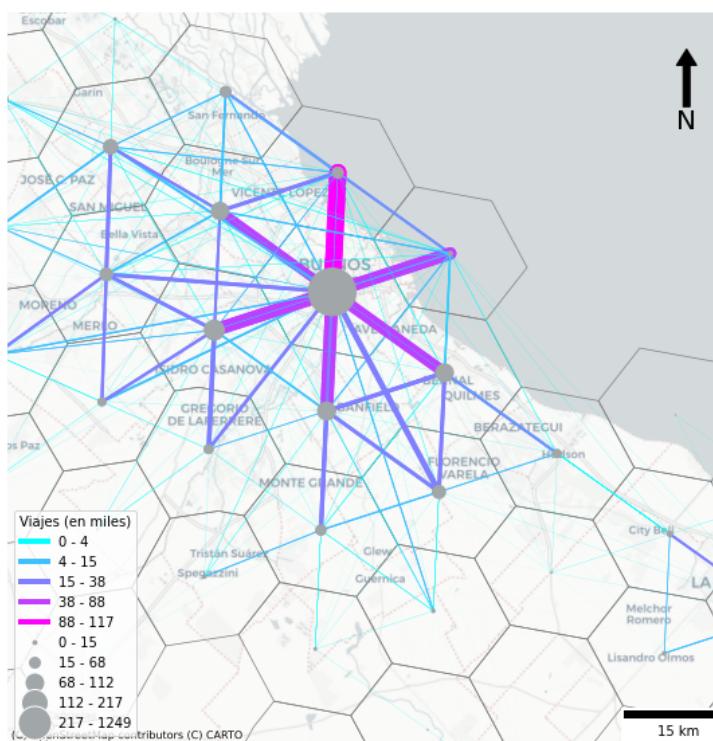


Fuente: Elaboración propia en base a datos SUBE

Como ventaja adicional de este método, con respecto a la ENMODO, es que se puede obtener una matriz de Origen Destino para unidades geográficas de mayor granularidad¹⁷.

¹⁷ Dado que se toma el punto medio de cada grilla hexagonal, algunos de los mismos pueden caer en espacios imposibles de generar viajes en el sistema de transporte público, como puede ser el Río de La Plata.

Figura 4.3.3. Orígenes destinos para diferentes niveles de agregación



Fuente: Elaboración propia en base a datos SUBE

4.4 - Resultados de viajes por cantidad de etapas y modos

En primer lugar, de la Matriz OD SUBE se puede obtener que de 3,1 millones de tarjetas se realizan 6,3 millones de viajes, lo que da un promedio de dos viajes por tarjeta. Estos viajes se realizan, en su mayoría, en una etapa (el 68%), mientras que poco más de un cuarto se realiza en dos etapas (26%), con apenas el 6% restante en tres etapas o más.

Figura 4.4.1. Cantidad de etapas por viaje

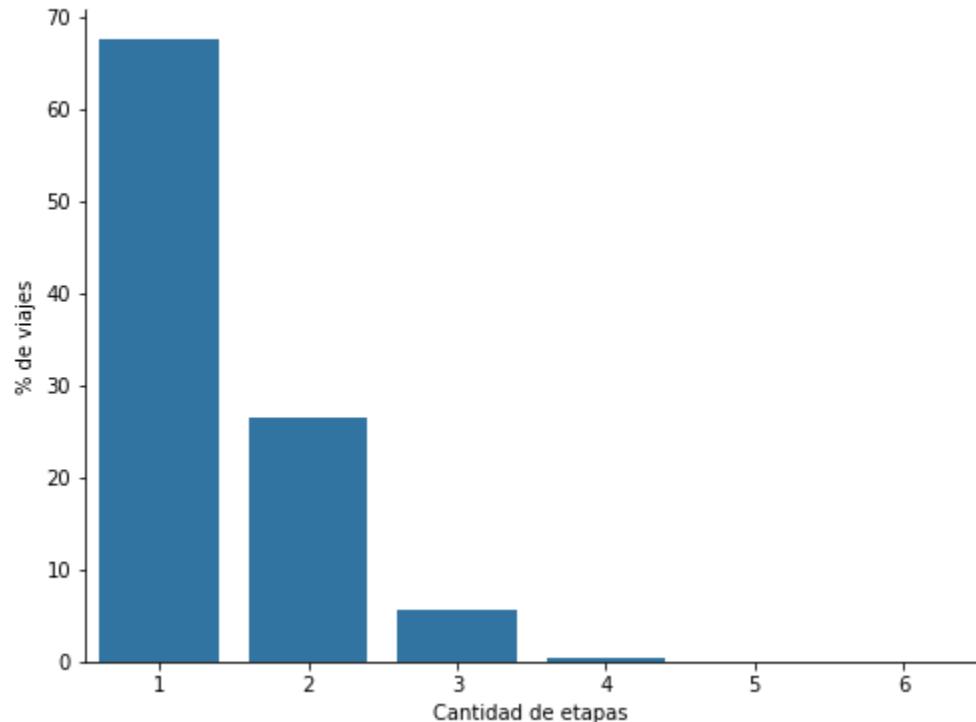
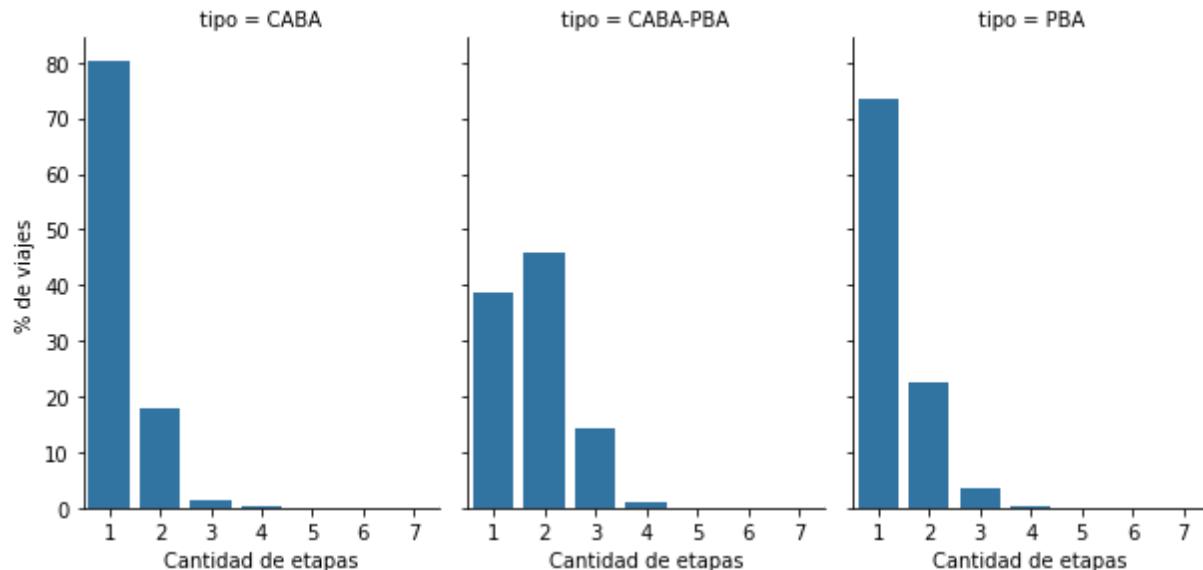


Figura 4.4.1. Cantidad de etapas por viaje

Cuando uno observa la cantidad de etapas por tipo de viaje, se observa que mientras la mayoría de los viajes que son dentro de CABA o de PBA se hacen en una sola etapa, al analizar los que vinculan ambas jurisdicciones vemos que más de la mitad se hace en dos etapas o más.

Figura 4.4.2 Cantidad de etapas por tipo de viaje



Fuente: Elaboración propia en base a datos SUBE

Otra forma de ver esto es por el modo utilizado. Cabe mencionar una aclaración sobre cómo nos referimos al modo de cada viaje. En primer lugar se toman viajes completos, con todas las etapas involucradas con destino imputado. Es decir, que si un viaje de dos etapas donde se combina colectivo y tren, ese mismo viaje contará como multimodal. Pero si un viaje de dos etapas donde se combinan dos líneas de colectivo o subte, ese viaje se clasificará como colectivo o subte respectivamente. Esto es sólo válido para los viajes, dado que cada etapa se realiza en un único modo.

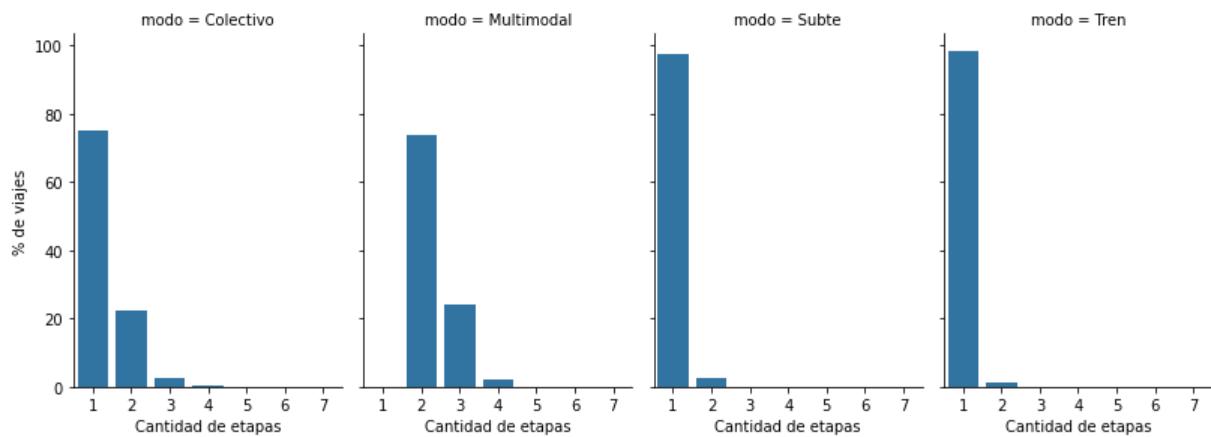
Tabla 4.4.3 Transacciones, Etapas y Viajes por modos utilizados

Modo	Etapas (%)	Viajes (%)
Colectivo	73,4	65,9
Subte	11,3	9,2
Tren	15,3	8,0
Multimodal	-	16,9

Fuente: Elaboración propia en base a datos SUBE

Cuando vemos la cantidad de etapas de los viajes de acuerdo a los modos utilizados, vemos que la mayoría de los viajes de más de una etapa corresponden a los viajes multimodales. Los viajes que utilizaron el colectivo son mayormente de una sola etapa (casi 80%). En el caso del subte y tren, dado que cuando una persona combina dos líneas de subte o tren no hace otra transacción con la tarjeta SUBE, no podemos detectarlo. Por lo cual prácticamente el total de los viajes es de una sola etapa. Existe un pequeño grupo de dos etapas pero que corresponde a un viaje de una persona que sale del subte y luego vuelve a ingresar al subte, con lo cual más que dos etapas probablemente se trate de dos viajes mediados por una actividad en el medio.

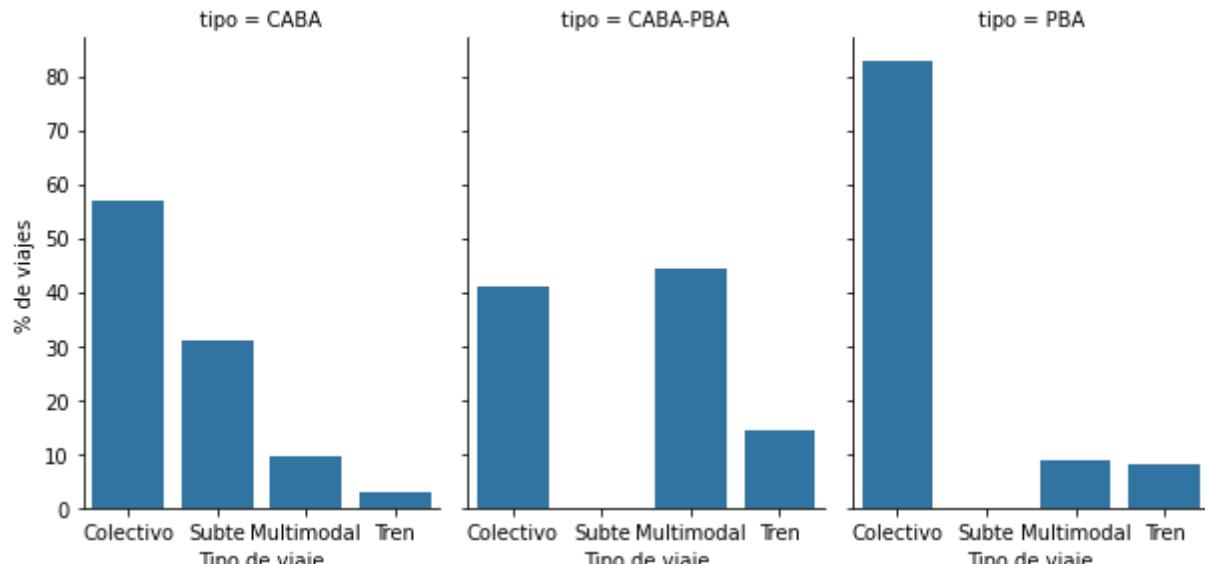
Figura 4.4.4 Cantidad de etapas por modo utilizado



Fuente: Elaboración propia en base a datos SUBE

Cuando vemos el reparto modal hacia dentro de cada tipo de viaje, vemos que obviamente los viajes dentro de la PBA una enorme mayoría incluye el colectivo (84%) con poca participación de los viajes multimodales y en tren. Cuando se observan los viajes dentro de CABA también se observa una mayoría en colectivo (casi 60%) con el subte en segundo lugar (casi 40%) y dejando a multimodal y tren con poca participación (menor a 10%). Cuando vemos los viajes entre CABA y PBA hay un protagonismo compartido entre el colectivo y los multimodales (ambos en torno al 40%). Dentro de estos últimos, el 80 % incluye al menos una etapa en tren.

Figura 4.4.5 Tipo de viaje según modo utilizado



Fuente: Elaboración propia en base a datos SUBE

Otra forma de afirmarlo es verlo por modo. De los viajes hecho en colectivo, el 60% se realiza dentro de la PBA, mientras que al ver los viajes hechos en tren la mitad se realiza dentro de la PBA y el 40% para viajes entre CABA y PBA. De los viajes multimodales, casi el 60% es entre CABA y PBA. Obviamente el subte aparece como exclusivo para viajes dentro de CABA por tener toda su red dentro de CABA (para hacer viajes entre CABA y PBA es necesario complementar el subte con otro modo y es catalogado como Multimodal). De hecho, el subte se puede observar que constituye un modo muy importante en los viajes Multimodales que vinculan CABA y PBA. Si uno toma todos los viajes que tienen al menos una etapa en subte, el 27% son viajes que unen CABA y PBA.

Figura 4.4.6 Tipo de viaje según el modo utilizado - En porcentajes

Tipo de viaje	Colectivo	Multimodal	Subte	Tren
PBA	60,4	25,2	-	49,4
CABA	25,8	16,6	100,0	10,5
CABA-PBA	13,8	58,2	-	40,2
Total	100,0	100,0	100,0	100,0

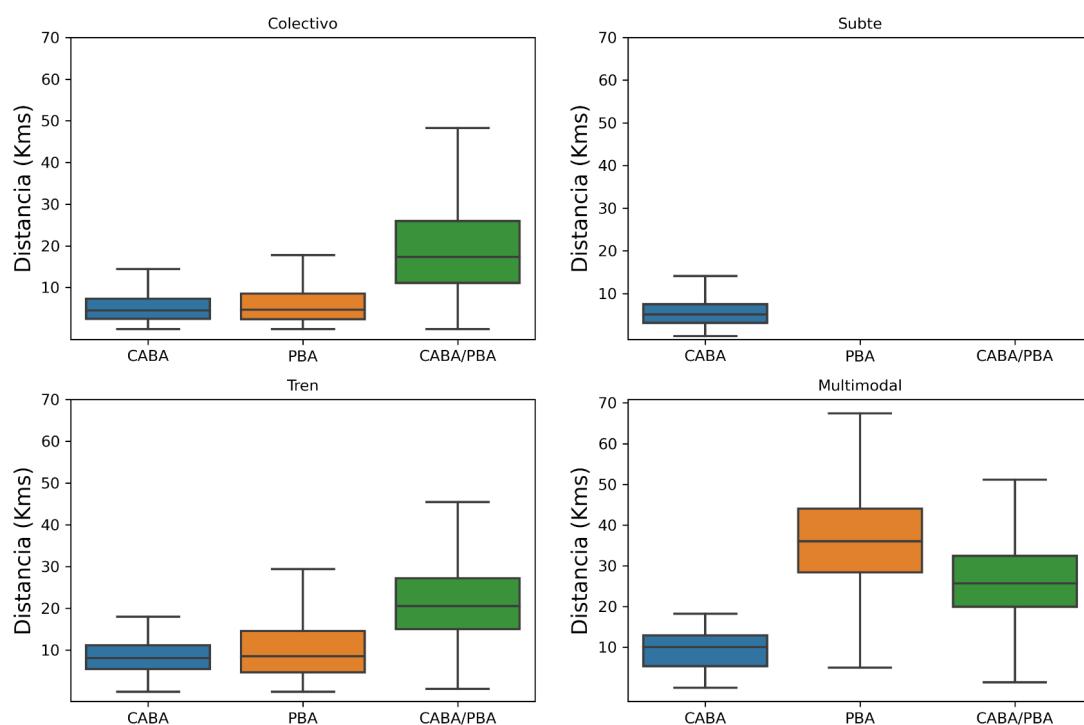
Fuente: Elaboración propia en base a datos SUBE

4.5 - Resultados de viajes por distancias

Se calcularon para cada etapa las distancias aproximadas entre el punto de origen y el de destino tomando el siguiente criterio. Para los modos guiados ferroviarios (subte y tren) dado que su trazado sigue una línea recta, se utilizó una distancia euclíadiana. Esto puede ocurrir en un error cuando se realiza combinación entre líneas, e idealmente lo mejor sería calcular distancias a lo largo de la red existente para todos los modos. En ese sentido esta variable es una variable proxy. Para los colectivos, dado que se mueven en superficie y siguen el callejero, se utilizó la distancia Manhattan¹⁸. Una vez calculada la distancia para cada etapa, se agregó al nivel de viajes.

Como se puede observar, los viajes que utilizan el tren tienen una distancia mediana mayor (18,4 km) mientras que tanto colectivo como subte es del orden de los 7 km.

Figura 4.5.1 - Distancia por modo

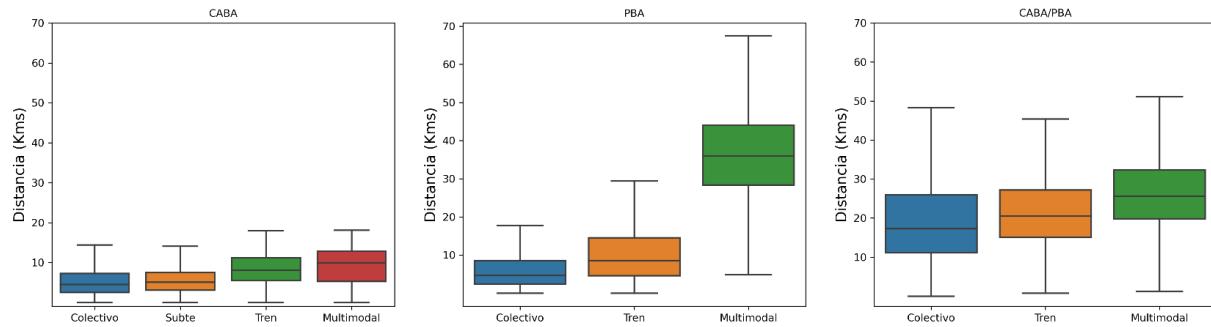


Fuente: Elaboración propia en base a datos SUBE

¹⁸ La distancia Manhattan es una forma de distancia en la que la métrica usual (euclíadiana) es reemplazada por la suma de las diferencias (absolutas) de sus coordenadas. Es decir, en lugar de calcular el largo de la hipotenusa, suma los catetos. Intenta reproducir la trayectoria real en una ciudad cuyo callejero sigue un plano damero.

Al analizar por tipo de viaje, lógicamente los viajes en CABA tienen distancias más cortas que los realizados en PBA o los realizados entre CABA y PBA

Figura 4.5.2 - Distancia por tipo de viaje



Fuente: Elaboración propia en base a datos SUBE

5 - Próximos pasos

Tal y como se explicita al comienzo, el objetivo de este trabajo fue refinar y expandir lo realizado en el contexto de publicaciones anteriores (Gonzalez & Anapolsky, 2021 y Gonzalez & Anapolsky, 2022) para imputar destinos y construir diferentes matrices de origen-destino de etapas y viajes para la Región Metropolitana de Buenos Aires (RMBA) en base a datos de la tarjeta SUBE. Este trabajo utiliza criterios específicos a la RMBA como el tratamiento diferenciado de las transacciones tren con sus *checkouts* y una cartografía de paradas de todos los modos en la región, entre otros. Estas decisiones específicas al RMBA pueden permitir obtener datos de mejor calidad para esta región, pero hace más dificultoso su procesamiento y actualización en el tiempo.

La metodología utilizada en los trabajos anteriores ofrece un esquema de trabajo plausible de generalización y capaz de ser aplicado a cualquier ciudad de Argentina siempre y cuando existan datos SUBE en la misma y que tengan el mismo modelo de datos. Sin embargo, siempre será necesario un proceso de customización a las realidades de cada ciudad. A su vez, si bien cada modelo de datos de los sistemas SRE es específico a su implementación en cada ciudad, tienen elementos estructurantes en común. Estos podrían traducirse en un modelo de datos común que pueda ser analizado de modo sistemático, sin importar la ciudad, el país o el tipo de sistema de SRE. Es decir, todos tendrán un conjunto de transacciones identificadas en tiempo y espacio y un identificador de la tarjeta.

En ese sentido, una agenda de trabajo futura puede ser la elaboración de una librería de software libre que permita tomar un set de datos de transacciones con estas características básicas, imputar los destinos y elaborar las matrices de origen destino. En una primera instancia, se haría caso omiso de los modos, dando a las transacciones de todos los modos el mismo tratamiento, imputando su destino en base a su siguiente transacción independientemente de si alguno de esos modos tiene un *checkout*. Tampoco sería necesaria una cartografía de paradas o estaciones, dado que se elaboraría una matriz de validación de los destinos imputados en base a las transacciones, utilizando conteos para cada celda hexagonal del sistema H3. A su vez, el uso de una infraestructura de geolocalización como H3, independiente de los límites jurisdiccionales de cada país o ciudad, permite escalar este trabajo más allá de las zonas específicas de cada región y utilizar las celdas hexagonales de diferente nivel de resolución como zonas de análisis de tránsito (o TAZ por las siglas en inglés de *traffic analysis zone*)¹⁹.

Esta librería permitiría automatizar una porción importante del trabajo de elaboración de matrices OD en base a datos SRE para cualquier ciudad y país, con, a lo sumo, un pequeño trabajo de preprocesamiento para adaptar la información de cada ciudad al modelo de datos fundamental necesario. Esto permitiría construir un insumo de datos indispensable para la planificación del transporte público y su continua actualización en tiempos de muchos cambios

¹⁹ Una zona de análisis de tráfico es una unidad espacial de territorio con características similares en sus atributos, que simplifica los procesos en la modelación de la demanda. Estas áreas son delimitadas por planificadores que contienen información relacionada con orígenes y destinos de viajes.

por la agenda climática y de salud pública. Fundamentalmente, para la red de buses, que es la que ofrece la mayor versatilidad para cambiar su oferta, con lo cual un análisis de la demanda plausible de estar actualizados en cortas ventanas de tiempo es un recurso muy útil.

6 - Bibliografía

Alsger, A., Mesbah, M., Ferreira, L., Safi, H. (2015) Public Transport Origin-Destination Estimating Using Smart Card Fare Data. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.

Amini, Amineh & Wah, Teh & Saboohi, Hadi. (2014). On Density-Based Data Streams Clustering Algorithms: A Survey. *Journal of Computer Science and Technology*. 29. 116-141. 10.1007/s11390-013-1416-3.

Banks J. (1998) *Introduction to Transportation Engineering*. Mc Graw-Hill

Barry, J.J., Newhouser, R., Rahbee, A., Sayeda, S. (2002) Origin and Destination Estimation in New York City with Automated Fare System Data. *Transportation Research Record: Journal of the Transportation Research Board* №1817, 183–187

Barry, J. J., R. Freimer, and H. L. Slavin. (2009) Use of Entry-Only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City. In *Transportation Research Record: Journal of the Transportation Research Board*, №2112, *Transportation Research Board of the National Academies*, Washington, D.C., 2009, pp. 53–61.

Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., & Ratti, C. (2011). Real-time urban monitoring using cell phones: A case study in Rome. *Intelligent Transportation Systems, IEEE Transactions on*, 12(1), 141–151.

Chu, K. K. A., and R. Chapleau (2010). Augmenting Transit Trip Characterization and Travel Behavior Comprehension: Multiday Location-Stamped Smart Card Transactions. In *Transportation Research Record: Journal of the Transportation Research Board*, №2183, *Transportation Research Board of the National Academies*, Washington, D.C., 2010, pp. 29–40.
<https://trid.trb.org/view.aspx?id=910384>

Cui, A. (2006) Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems. Tesis de Maestría. Massachusetts Institute of Technology, USA.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining. Portland, OR, pp. 226–231.

Farzin, J. M. (2008) Constructing an Automated Bus Origin–Destination Matrix Using Farecard and Global Positioning System Data in São Paulo, Brazil. In *Transportation Research Record: Journal of the Transportation Research Board*, №2072, *Transportation Research Board of the National Academies*, Washington, D.C., 2008, pp. 30–37.

<http://trjournalonline.trb.org/doi/abs/10.3141/2072-04>

Gonzalez, F. & Anapolsky, S. (2021) Accesibilidad al Ferrocarril Roca: análisis de datos masivos para la identificación de corredores con potencial de micromovilidad.

<http://dx.doi.org/10.18235/0003456>. Banco Interamericano de Desarrollo. Disponible en:
<https://publications.iadb.org/es/accesibilidad-al-ferrocarril-roca-analisis-de-datos-masivos-para-a-identificacion-de-corredores>

Gonzalez, F. & Anapolsky, S. (2022) Identificando la desigualdad en los patrones de movilidad en transporte público. Banco Interamericano de Desarrollo. Disponible en:

<https://publications.iadb.org/publications/spanish/document/Identificando-la-desigualdad-en-los-patrones-de-movilidad-en-transporte-publico.pdf>

Gordon, J.; Koutsopoulos, H.; Wilson, N. & Attanucci, J. (2013) Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 2343, ISSN 0361-1981, <http://dx.doi.org/10.3141/2343-03>

Hanft, J.; Lyer, S.; Levine, B. & Reddy, A. (2015) A novel approach for transforming bus service planning using 2 integrated electronic AFC & AVL data at MTA in New York City Presented at 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016.

Holm Møller, T., Simlett, J., & Mugnier, E. (2019). Micromobility: Moving cities into a sustainable future - EY Report.

Instituto Nacional de Estadísticas y Censos, Ministerio de Economía de la Nación (INDEC), (2003), ¿Qué es el Gran Buenos Aires?. Buenos Aires, Argentina.

ITF. (2020). Safe Micromobility. Report by the International Transport Forum OECD/ITF.

Kager, R., Bertolini, L., Te Brömmelstroet, M., (2016). Characterisation of and reflections on the synergy of bicycles and public transport. *Transport. Res. Part A: Policy Practice* 85, 208–219. <https://doi.org/10.1016/j.tra.2016.01.015>.

Lu, A., & Reddy, A.V. (2011). Algorithm to Measure Daily Bus Passenger Miles Using Electronic Farebox Data for National Transit Database Section 15 Reporting. In *Transportation Research Record: Journal of the Transportation Research Board* 2216, pp. 19–32 (2011).

El Mahrsi, M. K.; Côme, E.; Oukhellou, L. & Verleysen, M. (2017) "Clustering Smart Card Data for Urban Mobility Analysis," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 712-728, March 2017, doi: 10.1109/TITS.2016.2600515.

Mariana Tallarico, V. (2020) "Entre la autonomía y la reproducción". Disponible en:
<https://www.teseopress.com/violenciacon-tralasmujeres/>

Mondschein, A. (2014) Re-Programming Mobility. Literature Review. Rudin Center for Transportation Policy and Management. NYU. United States. Disponible en:

<http://reprogrammingmobility.org/wp-content/uploads/2014/09/ReProgrammingMobility-LitReview-08.04.2014.pdf>

Munizaga, M., & C. Palma (2012) Estimation of a Disaggregate Multimodal Public Transport Origin-Destination Matrix from Passive Smartcard Data from Santiago, Chile. *Transport Research Part C: Emerging Technologies*, Vol. 24, 2012, pp. 9–18.

<http://www.sciencedirect.com/science/article/pii/S0968090X12000095?via%3Dihub>

Munizaga, M., Devillaine, F., Navarrete, C., Silva, D. (2014) Validating travel behavior estimated from smartcard data. *Transportation Research Part C Emerging Technologies* 44, 70–79.

<http://www.sciencedirect.com/science/article/pii/S0968090X14000801?via%3Dihub>

Navick, D. S., and P. G. Furth (2002). Estimating Passenger Miles, Origin-Destination Patterns, and Loads with Location-Stamped Farebox Data. In *Transportation Research Record: Journal of the Transportation Research Board*, №1799, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 107–113.

Oeschger, G., Carroll P., Caulfield B. (2020) Micromobility and public transport integration: The current state of knowledge. *Transportation Research Part D: Transport and Environment* <https://doi.org/10.1016/j.trd.2020.102628>

Ortuzar, J. D. D., & Willumsen, L. G. (2011). *Modelling transport*. John Wiley & Sons.

Pelletier, M.P., Trépanier, M. & Morency C. (2011), Smart card data use in public transit: A literature review, *Transportation Research Part C: Emerging Technologies*, Volume 19, Issue 4, 2011, Pages 557–568, ISSN 0968–090X, <http://dx.doi.org/10.1016/j.trc.2010.12.003>.
(<http://www.sciencedirect.com/science/article/pii/S0968090X1000166X>)

Reddy, Alla V., A. Lu, S. Kumar, V. Bashmakov, and S. Rudenko (2009). Entry-Only Automated Fare-Collection System Data Used to Infer Ridership, Rider Destinations, Unlinked Trips, and Passenger Miles. In *Transportation Research Record: Journal of the Transportation Research Board*, №2110, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 128–136.

Resolución 160, Alcaldía de Bogotá, Bogotá, Colombia, 2 de Febrero de 2017

Resolución 1903/2015. Boletín Oficial de la República Argentina, Buenos Aires, 24 de septiembre de 2015. Disponible en:
<https://www.argentina.gob.ar/normativa/nacional/resoluci%C3%B3n-1903-2015-252390/texto>

Trépanier, M., Chapleau, R. (2006) Destination estimation from public transport smart card data. In: The 12th IFAC symposium on Information Control Problems in Manufacturing, Saint-Etienne, France

Trépanier, M., N. Tranchant, and R. Chapleau (2007). Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System. *Journal of Intelligent Transportation Systems*, Vol. 11, №1, 2007, pp. 1–14.

Tenenbaum de Olivera, G. & Nassi, C.D. (2015) Estimación de matriz origen-destino a partir de datos de tarjetas inteligentes: una revisión bibliográfica. *Ingeniería de transporte*, Volumen 19, Nº 01, 73–86. Universidad Federal de Rio de Janeiro / Programa de Engenharia de Transportes / COPPE, Brasil Disponible en:

<http://www.ingenieriadetransporte.org/index.php/sochitran/article/view/144>. Consultado el 12/09/2017

Zhao, J., A. Rahbee, and N. H. M. Wilson (2007). Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 22, №5, 2007, pp. 376–387.