

Accesibilidad al Ferrocaril Roca



Análisis de datos masivos
para la identificación de
corredores con potencial
de micromovilidad.

2021



Autores

Felipe Gonzalez
Sebastián Anapolsky

Editores

Julieta Abad
Cristian Moleres

**Catalogación en la fuente proporcionada
por la Biblioteca Felipe Herrera del Banco
Interamericano de Desarrollo**

González, Felipe.

Accesibilidad al Ferrocarril Roca: Análisis de datos masivos, para la identificación de corredores con potencial de micromovilidad / Felipe Gonzalez, Sebastián Anapolsky; editores, Julieta Abad, Cristian Moleres.

p. cm. — (Monografía del BID ; 951)

Incluye referencias bibliográficas.

Transportation-Argentina. 2. Railroads-Argentina. I. Anapolsky, Sebastian. II. Abad, Julieta, editora. III. Moleres, Cristian, editor. IV. Banco Interamericano de Desarrollo. División de Transporte. V. Título. VI. Serie.

IDB-MG-951

Código de JEL: L91, L92, 018

Palabras clave: Medio de Transporte, Área Metropolitana, Análisis de Datos, Micromovilidad

Copyright © 2021 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObras- Derivadas (CC-IGO 3.0 BY-NC-ND) (<https://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas. Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UN- CITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional. Note que el enlace URL incluye términos y condiciones adicionales de esta licencia. Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.

CRÉDITOS EDITORIALES

AUTORES

Felipe Gonzalez, Sebastián Anapolsky

EDITORES

Julieta Abad, Cristian Moleres

VISUALIZACIÓN CONCEPTUAL

--

DISEÑO GRÁFICO

Otros Pérez

CORRECCIÓN DE TEXTOS:

Cristian Moleres, Carlos Mojica,
Roberto Rodríguez

FOTOGRAFÍAS

Ministerio de Transporte de Argentina: Páginas:
6, 7, 18, 52, 63, 73, 76, 77, 80, 81, 88, 89
Wikimedia Commons: Portada y páginas 42, 64
Unsplash: Páginas 8, 11, 12, 15, 17, 25, 26, 31,
32, 41, 46, 51, 57, 90

Todas las imágenes y textos contenidos
en esta publicación fueron compendiados
y cedidos por el Banco Interamericano de
Desarrollo (BID)

Índice

6

Resumen ejecutivo

10

Análisis de viajes de
corta distancia en base
a datos de la tarjeta
del Sistema Único de
Boleto Electrónico

16

Accesibilidad a
las estaciones del
Ferrocarril Roca,
viajes cortos

| | | | |
|--|----|--|----|
| | 24 | | 50 |
| Uso de las tarjetas electrónicas de pago para el análisis del transporte | | Identificación de corredores con alta concentración de etapas sustituibles por micromovilidad | |
| | 30 | | 62 |
| Preprocesamiento: inferir destinos y recrear cadenas de viajes | | Caso de estudio: Estación Quilmes | |
| | | Etapas de viajes cortos sin agregación de datos | |
| | | Agregación de viajes en torno a estaciones | |
| | | Asignación a la red vial | |
| | | Demanda de micromovilidad | |
| | 40 | | 75 |
| Técnicas de cluster y su aplicación a clasificación de patrones de viajes | | Conclusiones | 80 |
| El método DBSCAN | | Anexo | |
| Algunas decisiones metodológicas y sus efectos | | | 90 |
| | | Bibliografía | |

Resumen



ejecutivo





Este trabajo utiliza los datos de la tarjeta del Sistema Único de Boleto Electrónico (SUBE), utilizada para el cobro del transporte público en la Región Metropolitana de Buenos Aires (RMBA), para identificar corredores de alta densidad de viajes cortos en torno a las estaciones del Ferrocarril Roca. El objetivo de analizar las etapas de viajes de corta distancia es identificar corredores con potencial de uso de bicicleta u otros modos de micromovilidad. A tal efecto se procesaron los datos de la tarjeta SUBE imputando destinos de viajes, calculando distancias y conformando una matriz de origen y destino de viajes. Como resultado, estimamos que de los 3,2 millones de etapas de viajes que tienen un origen o destino en el área de influencia del ferrocarril roca (un 28% del total de los viajes de la RMBA), 1,26 millones (un 40%) son menores de 5 kilómetros, una distancia factible para los sistemas de micromovilidad. Por último, utilizamos un algoritmo de aprendizaje no supervisado DBSCAN para clusterizar estas etapas de viajes sustituibles por micromovilidad y localizarlos en el espacio. Eso permitió determinar corredores de alta densidad de etapas de viajes en torno a las estaciones del Ferrocarril Roca. Los mismos fueron representados con diversos métodos de visualización de datos y utilizando técnicas de agregación espacial.

This work uses the data from the Unique Electronic Ticket System card (SUBE), used to charge public transport in the Metropolitan Region of Buenos Aires (RMBA), to identify high-density short-trip corridors around the Roca Railway stations. The objective of analyzing short-distance travel trips is to identify corridors with potential for cycling or other modes of micromobility. To this end, the SUBE card data was processed by imputing travel destinations, calculating distances, and creating a trips origin and destination matrix. As a result, we estimate that of the 3.2 million travel stages that have an origin or destination in the Roca railway's area of influence (28% of total RMBA trips), 1.26 million (40 %) are less than 5 kilometers, a feasible distance for micromobility systems. Finally, we use a DBSCAN unsupervised learning algorithm to cluster these micro-mobility substitutable travel stages and locate them in space. This made it possible to determine high-density corridors of public transport trips around the stations of the Roca Railway. They were represented using various data visualization methods, and spatial aggregation techniques.

Análisis de viajes de corta distancia en base a datos de la tarjeta del Sistema Único de Boleto Electrónico





El principal objetivo de este trabajo es la utilización de datos del Sistema Único de Boleto Electrónico (SUBE) para la identificación de corredores de alta densidad de viajes de corta distancia en la zona de influencia del ferrocarril Roca en la Región Metropolitana de Buenos Aires (RMBA)¹. Se van a explorar técnicas de minería de datos y visualización para informar decisiones sobre políticas o inversiones relacionadas con la movilidad sustentable y el uso de modos alternativos de transporte. Los datos de la tarjeta SUBE, por estar georeferenciados y por ser el único modo de pago del transporte público en la RMBA, brindan la posibilidad de identificar con gran detalle patrones de movilidad en transporte público y realizar este tipo de análisis más específicos.

En este trabajo se realizó un procesamiento de la tarjeta SUBE en donde se le imputaron los destinos de los viajes a cada transacción y se reconstruyeron las cadenas de viajes diarias para cada usuario. Esto permitió a su vez la construcción de una matriz de origen y destino de viajes en transporte público de un día completo para toda la RMBA. La imputación de los destinos, la georeferenciación de cada transacción y la posibilidad de calcular la distancia de cada etapa de viaje permite identificar aquellos viajes de corta distancia en la zona de influencia del ferrocarril Roca, permitiendo realizar un análisis específico de corredores con potencial de micromovilidad para acceder al ferrocarril.

El interés en analizar viajes de corta distancia en la zona de influencia de un corredor de transporte público masivo surge de poder identificar

1. Resolución 1903/2015. 25/09/2015. Ministerio de Transporte.

zonas de alta demanda de estos viajes. Se considera que los viajes cortos que permiten el acceso a corredores de transporte público masivo tienen alta potencialidad de uso de bicicleta o de sistemas de micromovilidad. Desde el punto de vista de la planificación del transporte, identificar destinos y corredores de viajes de corta distancia permite enfocar las políticas de promoción de modos alternativos de transporte, como el transporte activo (viajes caminando o en bicicleta) o la micromovilidad, así como planificar infraestructura específica que incentive un cambio en la partición modal existente.

El principal potencial de la micromovilidad en el contexto urbano radica en resolver, por un lado, los viajes de corta distancia y, por otro lado, lo que se denomina el problema de la primera y última milla, que está relacionado con el acceso a los modos masivos de transporte, como el ferrocarril. La micromovilidad puede contribuir a cambios en los patrones y comportamientos de movilidad, disminuyendo el transporte centrado en el automóvil². La micromovilidad, en combinación con el transporte público masivo, produce una sinergia producto de la complementariedad entre ambos modos. Mientras que la micromovilidad ofrece flexibilidad y accesibilidad puerta a puerta de modo eficiente, el transporte público se caracteriza por una mayor velocidad y cobertura espacial. Esta sinergia posibilita un nivel de accesibilidad, velocidad y comodidad que puede competir con los atributos de los modos privados motorizados³, contribuyendo en el desarrollo de ciudades más habitables, con menores niveles de congestión, contaminación atmosférica y acústica, un uso más eficiente del espacio, y menor cantidad de accidentes de tránsito⁴.

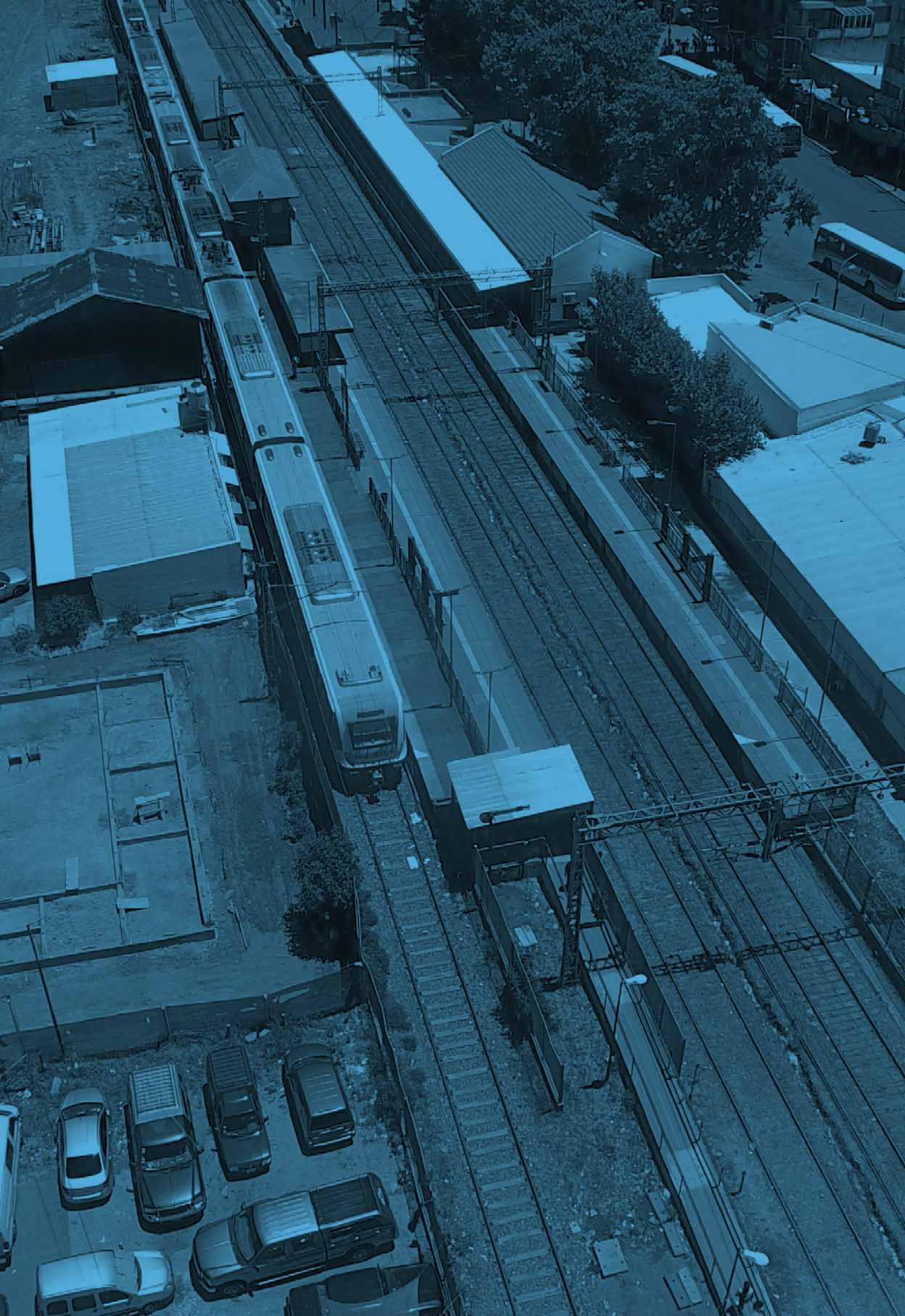
Se denomina micromovilidad al movimiento en micro-vehículos definidos como aquellos con una masa no superior a 350 kg y una velocidad de diseño no superior a 45 km/h. Estos micro-vehículos pueden ser autopropulsados (siendo la bicicleta el más paradigmático) como asistidos o propulsados por un motor eléctrico⁵. La promoción de la micromovilidad y la integración del transporte público es una política de movilidad que se puede lograr de diferentes maneras, dependiendo de la infraestructura y los servicios disponibles. Los sistemas de micro-vehículos compartidos han ganado cada vez más popularidad en los últimos años y están disponibles en ciudades de todo el mundo. Éstos, a su vez, se han vuelto cada vez más eficientes y fáciles de usar, lo que resulta en una alternativa competitiva y atractiva para los viajes de acceso y egreso del transporte público como así también para los viajes de distancias cortas como los que analizaremos en este trabajo.

2. Cardell, Martin, «How Micromobility Is Moving Cities into a Sustainable Future».
3. Kager, Bertolini, y Te Brömmelstroet, «Characterisation of and Reflections on the Synergy of Bicycles and Public Transport». centro de trasbordo de la RMBA.
4. Oeschger, Carroll, y Caulfield, «Micromobility and Public Transport Integration».
5. Edwina, «Safe Micromobility».



Accesibilidad a las estaciones del Ferrocarril Roca, viajes cortos





Para llevar adelante este análisis se realiza un procesamiento de las transacciones SUBE de un día hábil que permite estimar las cadenas de viajes de los usuarios de transporte público y calcular la distancia de cada uno de los viajes. A continuación, se identifican los viajes en la zona de influencia de las estaciones de ferrocarril Roca y se implementan una serie de técnicas de aprendizaje no supervisado para agrupar viajes, identificando corredores donde hay alta densidad de viajes de corta distancia. Se definen como viajes de corta distancia aquellos que tienen menos de cinco kilómetros ya que se considera una distancia apropiada para realizar viajes en bicicleta o con sistemas de micromovilidad.

Si bien en muchos casos se realizan análisis específicos para identificar viajes cortos de acceso al transporte público, por ejemplo, a través de encuestas u otro tipo de estudios⁶, la granularidad en los datos de las transacciones del sistema de pago electrónico de boletos permite identificar con gran nivel de detalle los corredores donde hay alta densidad de etapas de viaje de corta distancia. Si bien no se dispone de datos de viajes en otros modos, como viajes en automóvil, a pie o en bicicleta, dado que el transporte público es el principal modo utilizado en la RMBA (cerca del 60% de los viajes en modos motorizados se realiza en transporte público⁷), este análisis resulta relevante para la identificación de corredores con alto potencial de uso de bicicleta o de micromovilidad.

La tarjeta SUBE tampoco contiene información sobre el domicilio del usuario, con lo cual no se puede saber la distancia de caminata de

6. Mackett, «Policies to Attract Drivers out of Their Cars for Short Trips».

7. Partición modal obtenidos del análisis de las Encuestas de Movilidad Domiciliarias del año 2009 y 2014 para el Área Metropolitana de Buenos Aires.

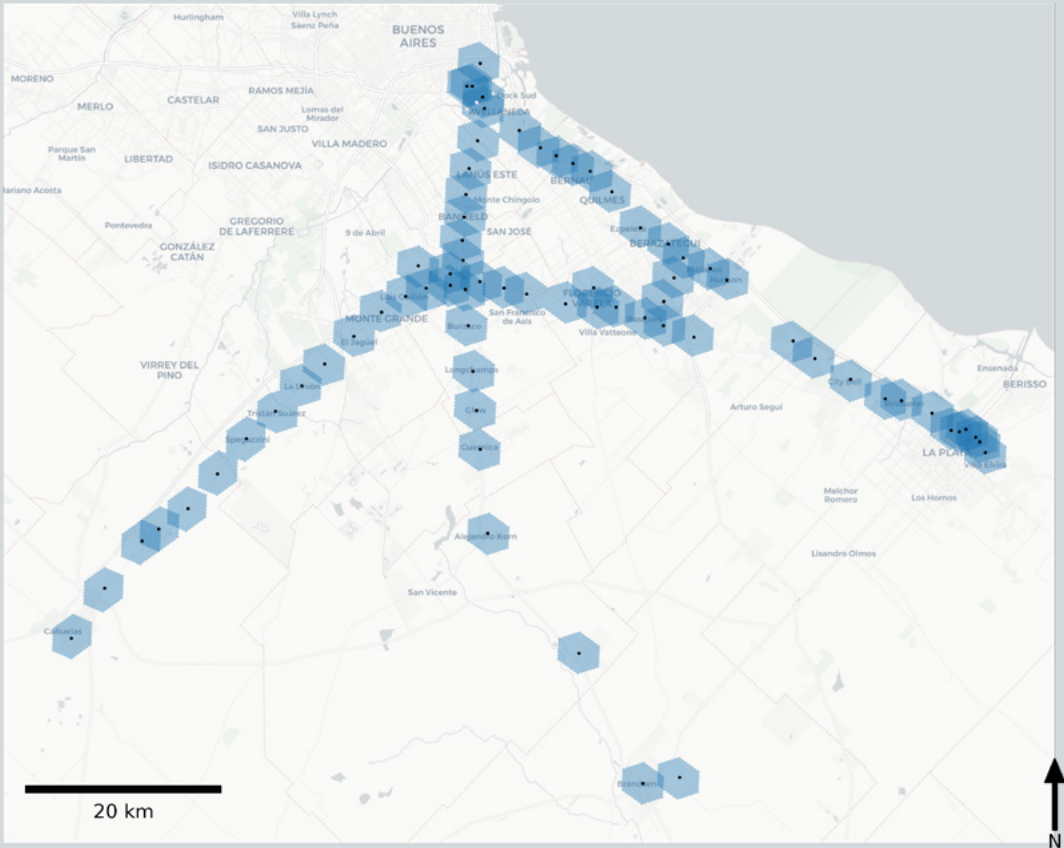
acceso a la red de transporte público en el primer o último viaje del día (como sí tiene en una encuesta de movilidad domiciliaria). Sin embargo, estas distancias tienden a ser cortas en la RMBA, dada la alta cobertura de la red, en especial la capilaridad de la red de colectivos, con lo cual no suman distancias significativas a las ya calculadas en los viajes. De todos modos, estas distancias de acceso a la red de transporte público no perjudican la probabilidad de sustitución por micromovilidad, sino todo lo contrario. El uso de la bicicleta permite que el viaje comience desde el momento en que se sale del hogar, recorriendo esas distancias de modo más directo (sin estar condicionado por la disposición de la red de transporte público) y a mayor velocidad promedio que la caminata.

El análisis se va a centrar en la zona de influencia del Ferrocarril Roca que es uno de los principales modos de transporte público masivo en la RMBA. Este ferrocarril permite conectar la ciudad de Buenos Aires con el Sur de la RMBA, conectando una serie de partidos de alta densidad poblacional y conectando a la ciudad de Buenos Aires con la ciudad de Plata, capital de la Provincia de Buenos Aires, localizada a 70 kilómetros. La red metropolitana de ferrocarril representa el 12% del total de viajes en transporte público, de los cuáles cerca del 41% son viajes realizados en los cuatro ramales del ferrocarril Roca.

Analizando los datos disponibles de la tarjeta SUBE, se observa que de los 11,2 millones⁸ de transacciones válidas identificadas para un día hábil, el 40% de las transacciones corresponden a etapas de viaje de menos de cinco kilómetros. De estas transacciones, en este trabajo se estudian en mayor detalle aquellas que tienen un origen o un destino en el área de influencia de las estaciones del Ferrocarril Roca (**Figura 1**)⁹.

8. La base de datos original obtenida para un día hábil de noviembre 2019 contiene 14.9 millones de registros, de los cuales, cerca de 13.6 millones corresponden a transacciones de ingreso a un modo de transporte público. Una vez depurada la base de datos con identificador de usuario y coordenadas de origen no válidos, se obtienen 11,2 millones de registros de transacciones.
9. El área de influencia está determinada como un área de alrededor de 1500 metros de cada estación del Ferrocarril Roca. Dado que las coordenadas recibidas en la base de datos de la tarjeta SUBE tiene un proceso de redondeo, el margen de error requiere un área de influencia amplia para captar estos viajes. El área de influencia no incluye el trayecto Haedo-Temperley.

Figura 1. Delimitación del área de influencia del Ferrocarril Roca



Fuente: Elaboración propia en base a datos de la SUBE y estaciones de Ferrocarril

En esta área se realiza cerca de 3,2 millones de etapas de viajes, que corresponden a un 28% del total de la RMBA, de las cuales 1,26 millones (un 40%) de éstas son de menos de cinco kilómetros y unas 560 mil (16%) son de menos de tres kilómetros. En cuanto al reparto modal en el área de influencia del ferrocarril, mientras que del total de las etapas de viajes cerca del 20% se realizan en ferrocarril, solo el 4 % usa este modo cuando los viajes son de menos de cinco kilómetros¹⁰.

10. Para la realización de este análisis se utilizaron los viajes de menos de cinco kilómetros en todos los modos de transporte, incluyendo colectivo, metro y ferrocarril. Este análisis podría efectuarse en algún modo en particular (ej. solo colectivo) de acuerdo a las características del lugar analizado o decisiones particulares de política.

Tabla 1. Transacciones SUBE según modo de transporte para un día hábil

| Modo | Transacciones SUBE | | | | | |
|-------------|--------------------|-----|-----------------------------------|-----|---------------|-----|
| | RMBA | | Origen o destino ferrocarril Roca | | | |
| | Total | | Total | | Menos de 5 KM | |
| Colectivo | 8.628.609 | 77% | 2.358.369 | 74% | 1.115.579 | 88% |
| Ferrocarril | 1.424.519 | 13% | 581.723 | 18% | 50.083 | 4% |
| Subte | 1.147.880 | 10% | 242.882 | 8% | 99.690 | 8% |
| Total | 11.210.008 | | 3.182.960 | | 1.265.352 | |

Fuente: Elaboración propia en base a Datos SUBE



El principal objetivo de este trabajo es la utilización de datos del Sistema Único de Boleto Electrónico (SUBE) para la identificación de corredores de alta densidad de viajes de corta distancia en la zona de influencia del ferrocarril Roca en la Región Metropolitana de Buenos Aires (RMBA). En el área de influencia de las estaciones del Ferrocarril Roca se realiza cerca de 3,2 millones de etapas de viajes, que corresponden a un 28% del total de la RMBA, de las cuales 1,26 millones (un 40%) de éstas son de menos de cinco kilómetros y unas 560 mil (16%) son de menos de tres kilómetros.

Uso de las tarjetas electrónicas de pago para el análisis del transporte





La información proveniente de las tarjetas electrónicas de pago de boletos en transporte público presenta información valiosa para el análisis del comportamiento de los usuarios y sus patrones de movilidad. Esta base de datos contiene los registros de todas las transacciones de pago de boletos en el sistema de transporte y la georeferenciación de donde se produjo el pago de boleto a través de los sistemas de GPS de los colectivos. Comparado con otras fuentes más tradicionales, como encuestas o diarios de viajes, los datos de las transacciones presentan información solamente de los viajes realizados en transporte público, pero con un nivel de cobertura mayor que no posible de obtener a través de otras fuentes.

Dado que las transacciones son traceables a través de un código identificador de tarjeta y que proveen información georeferenciada del origen de cada transacción¹¹, a través de una serie de metodologías para el procesamiento de datos, la información se puede utilizar para reconstruir cadenas de viaje a lo largo de un periodo de tiempo y permite comparar comportamientos en distintos periodos temporales, como por ejemplo viajes en un día hábil o fin de semana, o en determinadas zonas geográficas.

Sin embargo, el uso de datos de tarjetas electrónicas de pago para analizar la movilidad presenta algunos desafíos, como su gran volumen de información, la disponibilidad de información incompleta (solamente queda registrada la georeferenciación de los orígenes de los viajes, sin información sobre los destinos), no se encuentran disponibles los motivos de los viajes y es limitada la información sociodemográfica sobre los usuarios¹².

11. Mientras que, en algunos sistemas de pago electrónico, los viajes se validan tanto en el ingreso como el egreso del transporte, facilitando la ubicación del origen y el destino, en el caso de la tarjeta SUBE para el colectivo o el subterráneo se valida solamente al inicio del viaje. Solamente en el ferrocarril se registran ambas instancias.

12. En el caso de la tarjeta SUBE, se incluye información sobre el sexo y el atributo social de un conjunto de usuarios. Esta información podría ser utilizada para analizar patrones de viajes de este grupo de usuarios.

Dado que estos datos brindan información de alta calidad a bajo costo, en los últimos años se desarrollaron metodologías para procesar estos datos¹³. El procesamiento de esta información, basado en una serie de suposiciones y reglas de validación, permite obtener información sobre la utilización de los sistemas de transporte público. Estas metodologías buscan la reconstrucción de las cadenas de viajes y la elaboración de matrices de origen y destino para el transporte público, resultando útil para el análisis de problemáticas concretas relacionadas con la identificación de patrones de viajes, identificación de corredores y análisis de la demanda.

13. El Mahrsi et al., «Clustering Smart Card Data for Urban Mobility Analysis».
14. 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.
15. Munizaga et al., «Validating travel behavior estimated from smartcard data».

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 representa una etapa 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. Una serie de transacciones o etapas de viaje es considerada que pertenece a un mismo viaje si la transacción de la primera etapa y las subsiguientes está dentro de un rango temporal (en el caso de la RMBA es de dos horas y se utiliza este rango para asignar una tarifa integrada¹⁴). 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. A los efectos de este trabajo, se van a estar analizando principalmente las etapas de los viajes.

Entre los principales supuestos utilizados por estudios recientes¹⁵ y que serán adoptados en este trabajo se pueden enumerar los siguientes:

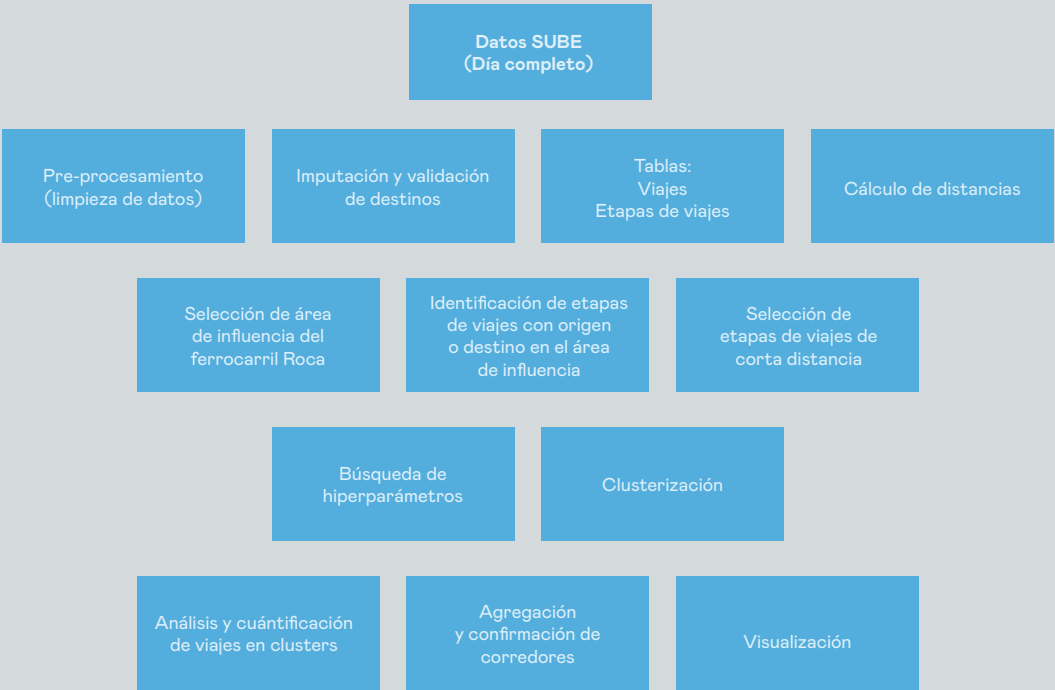
- ➔ Los pasajeros realizan sus viajes utilizando transporte público que se encuentra razonablemente cercano a sus orígenes y destinos como para realizar el acceso a las paradas/estaciones en forma peatonal.
- ➔ Además, realizan transferencias entre distintas líneas/modos para minimizar la distancia caminable 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.

➔ 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. Con esta información es posible calcular distancias, reconstruir cadenas de viajes y elaborar matrices de origen y destino para los viajes en transporte público con un gran nivel de detalle.

En la siguiente figura se puede observar el flujograma de trabajo propuesto en esta metodología:

Figura 2. Flujograma con la metodología propuesta



Preprocesamiento: inferir destinos y recrear cadenas de viajes





ESTACION Bs. As. - CONSTITUCION

AIRE ACONDICIONADO

¡MECLAMOS AL LINEA 59!
¡CNRT 105!

105

59

M.O.C.B.A.
S.A.

100

La tarjeta SUBE es un medio de pago gestionado por el Gobierno Nacional para el transporte público en la RMBA y otras ciudades del país. Esta tarjeta es el único modo posible de pagar los viajes en colectivos, subterráneos y trenes, y su base de datos contiene un registro de todas las transacciones de pago. Esta información puede ser utilizada para analizar patrones de movilidad. Sin embargo, los datos de la tarjeta SUBE disponibles de acceso público son limitados. Si bien el sitio de datos abiertos del Gobierno Nacional cuenta con información agregada de un día hábil y algunas agregaciones de datos por línea y modo de transporte, esta información resulta en ocasiones insuficiente para comprender en detalle la movilidad de la RMBA y realizar análisis de mayor profundidad. En este trabajo, la información se obtuvo a través de un pedido de acceso a la información pública y cuenta con transacciones georreferenciadas de un día hábil completo de noviembre de 2019 para toda la RMBA¹⁶.

Los pasajeros deben validar su tarjeta solo al abordar un autobús o ingresar a una estación de subterráneo (por lo tanto, no se recogen las ubicaciones del destino), y solamente en el caso del ferrocarril es necesario la validación de la tarjeta tanto al ingreso como al egreso de la estación. Estas transacciones quedan registradas en una base de datos que contiene una identificación de tarjeta anónima única, el modo de transporte relacionado a la transacción (colectivo, subterráneo, ferrocarril), la ubicación geográfica donde se registró la transacción, la hora, la línea de transporte utilizada y tarifa abonada. Dado que solamente las transacciones del ferrocarril validan el ingreso y el egreso de viaje, se desestimaron los egresos para lograr consistencia de datos entre los modos analizados.

16. 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 los expedientes EX-2020-33085207-APN-DGD#MTR, EX-2020-32945006-APN-DNAIP#AAIP-, NO-2020-33527820-APN-DIP#MTR, NO-2020-31201720-APN-SSPEYFT#MTR y NO-2020-33542724-APN-DIP#MTR.

Con esta información se realiza un preprocesamiento de los datos brutos para identificar transferencias y reconstruir las cadenas de viaje, infiriendo los destinos de cada transacción, y calcular la distancia de cada viaje. El destino de cada viaje se infiere asignando el origen del siguiente viaje y validando la existencia del modo/línea de transporte utilizado en ese destino. Se asume a su vez que el destino del último viaje del día corresponde a la vuelta al hogar (o primer viaje del día).

Siguiendo una metodología similar a otros estudios¹⁷, se reconstruye la cadena de viajes identificando primero las transacciones vinculadas (dos o más transacciones que corresponden a un mismo viaje) pudiendo identificar el inicio/fin de un viaje y sus transferencias. En el nivel más básico, esto se lleva a cabo utilizando un umbral de tiempo fijo: las validaciones que ocurren dentro de un período de tiempo determinado (dos horas en el caso de la RMBA) se consideran pertenecer a un mismo viaje¹⁸.

Un segundo paso es construir los viajes y armar las cadenas de los viajes para cada usuario (se identifica cada usuario por el ID de la tarjeta) en un día completo. En los casos que una misma tarjeta se utilice para pagar un mismo viaje dos o más veces (por ejemplo, un adulto que paga uno o más viajes de menores) se le asigna un nuevo identificador de usuario para no perder el segundo viaje. Para imputar los destinos de los viajes, es necesario realizar presunciones equivalentes a las que realizan estudios de estas características como Mahrssi et al¹⁹ o Munizaga y Palma²⁰:

- ➔ Supuesto de la parada más cercana: para una transacción determinada, el pasajero presumiblemente se baja en la parada o estación más cercana a donde tiene lugar la siguiente transacción.
- ➔ Bajo el supuesto de simetría diaria, se asume que, para la última transacción del día, el pasajero se baja en la parada o estación más cercana al lugar donde tuvo lugar su primera transacción del día (presumiblemente la parada más cercana al hogar).

Bajo estas presunciones, usando la base de datos de la tarjeta SUBE, se realizó un pre-proceso identificando las etapas que corresponden a un mismo viaje y asignando el destino correspondiente según las transacciones subsiguientes para cada tarjeta. El pre-proceso incluye también una verificación de las transacciones en los casos en que no es posible validar la georeferenciación de origen o hay datos faltantes.

17. El Mahrssi et al., «Clustering Smart Card Data for Urban Mobility Analysis».

18. 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.

19. El Mahrssi et al., «Clustering Smart Card Data for Urban Mobility Analysis».

20. Munizaga et al., «Validating travel behavior estimated from smartcard data».

Se eliminan también los casos donde solo puede identificarse un único viaje diario y no es posible reconstruir la cadena de viajes para ese usuario. Por último, se valida que en un rango de distancia caminable (alrededor de 1500 metros) al destino imputado se encuentre una parada del modo/línea de transporte utilizado en el viaje.

Una vez identificados los orígenes y destinos de cada transacción y diferenciadas las etapas y viajes para cada usuario, se calcula la distancia de cada uno de los viajes. Para llevar adelante este procedimiento se realizó un proceso para el cálculo de las distancias usando la red vial de Open Street Maps²¹. La **Tabla 2** muestra un ejemplo de los datos originales de las transacciones y la imputación de una serie de variables, como el número de viaje, la latitud de origen y destino y la distancia de cada tramo. Se observa en este ejemplo que una misma tarjeta que realiza siete transacciones, éstas corresponden a tres viajes.

Tabla 2. Base de datos SUBE original y datos imputados

| Datos SUBE | | | | | | Datos imputados | | | | |
|------------|------|----------------|------|---------|---------|-----------------|-------------|-------------|-----------|---------|
| id tarjeta | hora | etapa red sube | modo | lat | lon | n° viaje | lat destino | lon destino | distancia | |
| 1037034016 | 6 | 0 | COL | -34,780 | -58,830 | 1 | -34,664 | -58,728 | 16,4 | Viaje 1 |
| 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 | Viaje 2 |
| 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 | Viaje 3 |
| 1037034016 | 20 | 1 | COL | -34,664 | -58,728 | 3 | -34,780 | -58,830 | 16,4 | |

Fuente: Elaboración propia

21. Para el cálculo de distancias se utilizó una librería de Python llamada Pandana, desarrollada por UrbanSim: <https://udst.github.io/pandana/>. Noviembre 2020.

Una vez realizado este primer preprocesamiento, se pudo construir una base de datos de viajes, identificando para cada viaje su origen y destino, como se puede observar en la **Figura 3**. En esta etapa del proceso se calcula la distancia de los viajes usando el origen y destino de cada viaje sin tener en consideración las transferencias realizadas. Como resultado, se obtiene una cadena de viajes diaria para cada usuario.

Tabla 3. Identificación de viajes con imputación de orígenes y destinos

| id tarjeta | n° viaje | hora | lat origen | lon origen | lat destino | lon destino | COL | TRE | SUB | distancia |
|------------|----------|------|------------|------------|-------------|-------------|-----|-----|-----|-----------|
| 1037034016 | 1 | 6 | -34,780 | -58,830 | -34,436 | -58,566 | 2 | 1 | 0 | 46,65 |
| 1037034016 | 2 | 10 | -34,436 | -58,566 | -34,650 | -58,790 | 2 | 0 | 0 | 34,07 |
| 1037034016 | 3 | 19 | -34,650 | -58,790 | -34,780 | -58,830 | 1 | 1 | 0 | 17,44 |

Fuente: Elaboración propia

Para validar la imputación de los destinos de cada transacción es necesario verificar que en la ubicación de destino es posible acceder al modo y línea de transporte utilizado en origen. 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 viaje²². Cuando esta premisa no se cumple, la imputación del destino es considerada no válida.

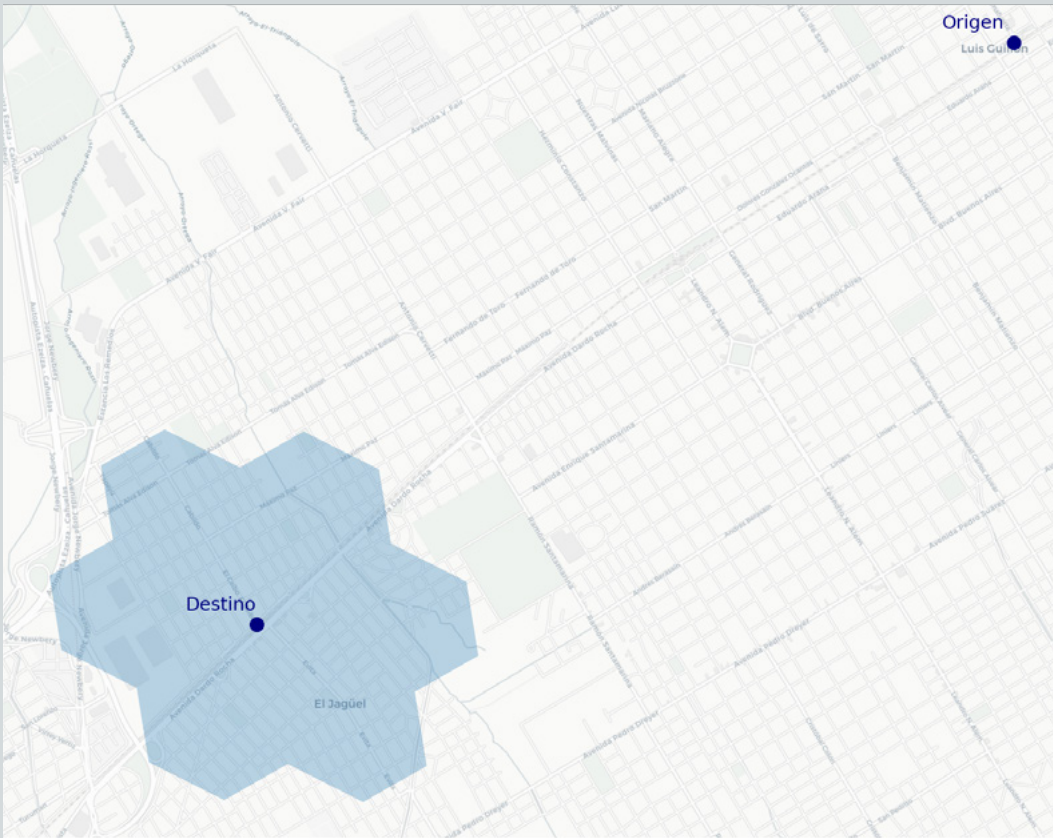
Existen diferentes metodologías para realizar esta validación. Una validación completa requiere identificar en una capa geográfica las líneas y paradas de toda la red de transporte público y verificar para cada destino imputado la distancia a estas paradas del modo y línea de transporte utilizado en el origen del viaje.

Para este trabajo se utilizó una versión más simple que no requiere emplear una capa geográfica con todas las líneas y paradas de la red de transporte público. Utilizando los orígenes de todas las operaciones SUBE y su georeferenciación, se construyó una matriz utilizando una capa geográfica de hexágonos, donde los hexágonos conforman un espacio que rodea cada destino imputado de viaje (buffer) de

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

aproximadamente 1500 metros (ver **Figura 2**). 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 2. Ejemplo de cómo se valida el destino imputado



Como resultado del proceso de pre-procesamiento y validación, de los 14,9 millones de registros de la base SUBE, una vez depurada la base de casos con información errónea o faltante, 11,2 millones de transacciones de la RMBA fueron procesadas con la metodología propuesta en este trabajo.

De estas, a 8,7 millones de transacciones (78% del total) se les pudo imputar y validar el destino. Una vez identificadas las etapas de viaje con transferencia, se obtuvieron cerca de ocho millones de viajes, de los cuales 5,7 millones (71%) corresponden a cadenas de viajes completas. Estos viajes corresponden a 4,4 millones de usuarios de transporte público, de los cuales 2,85 millones (65%) tienen viajes validados con origen y destinos imputados. Teniendo en cuenta las metodologías analizadas, los porcentajes de validación e imputación de datos faltantes son apropiados para poder caracterizar el comportamiento del sistema de transporte en la RMBA.

En términos de representatividad, los datos de la tarjeta SUBE tienen una cobertura censal, dado que todos los modos de transporte público de la RMBA la utilizan (excluyendo sólo los taxis y el sistema de bicicletas públicas compartidas). Si bien en el preprocesamiento de datos hay una pérdida de datos (22% de los registros) que resulta de la limpieza de la base de datos, se considera que no implica un sesgo sobre una determinada población. La no imputación de los destinos se debe mayormente a errores de captura en la georeferenciación de los sistemas GPS de los colectivos o tarjetas que realizan un único viaje y no es posible imputarles un destino. Este tipo de problemas en el equipamiento de los sistemas GPS tiene una naturaleza aleatoria y no introduce sesgos en la muestra. A su vez, la pérdida de casos es consistente con otros trabajos similares. Zhao et al (2007) obtuvieron 71% de viajes con destinos mientras que Trepanier et al. (2007) obtuvieron el 66%.



Para llevar adelante este análisis se realiza un procesamiento de las transacciones SUBE de un día hábil que permite estimar las cadenas de viajes de los usuarios de transporte público y calcular la distancia de cada uno de los viajes. De los 14,9 millones de registros de la base SUBE, quedaron 11,2 millones de transacciones de la RMBA luego de depurar la información errónea o faltante. De estas, a 8,7 millones de transacciones (78% del total) se les pudo imputar y validar un destino.

Técnicas de cluster y su aplicación a clasificación de patrones de viajes



Línea Roca

SEÑAL DE
EMERGENCIA

SEÑAL DE
EMERGENCIA
SEÑAL DE
EMERGENCIA

SEÑAL DE
EMERGENCIA

La Plata

SEÑAL DE
EMERGENCIA

SEÑAL DE
EMERGENCIA

SEÑAL DE
EMERGENCIA

En base al trabajo de otros autores que implementan métodos que permiten identificar patrones de movilidad a través del análisis de datos masivos de transacciones de los boletos electrónicos²³, se utilizan en este trabajo técnicas de clusterización para identificar corredores de alta densidad de viajes de corta distancia.

La generación de clústeres (agrupaciones) es una técnica de aprendizaje automático no supervisado que consiste en aplicar algoritmos para clasificar información en grupos de forma homogénea de acuerdo a las características de las variables. Como resultado, se obtiene una clasificación para cada observación constituyendo grupos de observaciones. Las observaciones que están en el mismo grupo tienen propiedades o características similares, mientras que las observaciones en diferentes grupos tienen propiedades o características diferentes. Los métodos de clustering son unas técnicas comunes para el análisis de datos estadísticos que se utilizan en muchos campos de conocimiento y es utilizada también para el análisis espacial.

En este caso, cada registro o punto de datos está conformado por un viaje o etapa entre un origen y un destino. Cada origen y destino está definido por un par de coordenadas (latitud y longitud). Por lo tanto, para cada punto (viaje o etapa) existe un vector de cuatro elementos (dos pares de coordenadas en el espacio).

23. Kadir et al., «Clustering of public transport operation using K-means».

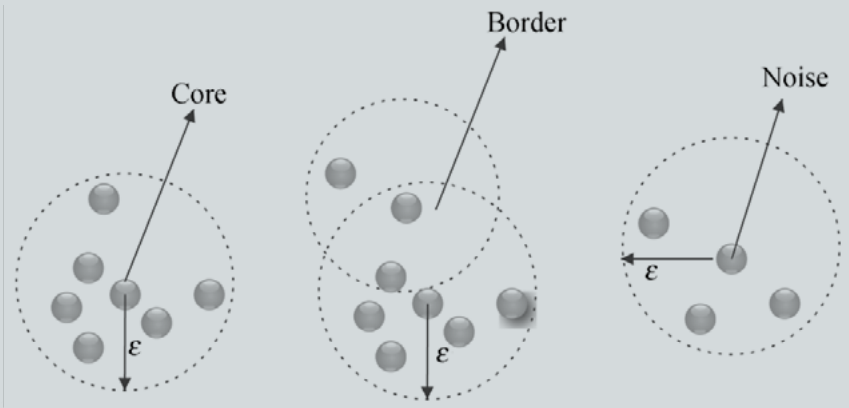
El método DBSCAN

Se procedió a utilizar principalmente la técnica de aprendizaje no supervisado de clusterización llamada DBSCAN. Esta técnica fue desarrollada en 1996²⁴ como una forma de mejorar los algoritmos de clusterización de Robert F. Ling llamado K medias. Este es un algoritmo de agrupamiento basado en densidad que juega un papel vital en la búsqueda de estructuras de formas no lineales basadas en la densidad y que es de los más utilizados para el análisis espacial. El algoritmo se basa en la detección de patrones, identificando aquellas áreas donde hay alta concentración de observaciones y conformando grupos con esos datos. Asimismo, identifica observaciones que no pertenecen a ninguna agrupación y son clasificados como ruido.

Este método toma las coordenadas geográficas de cada punto y su distancia con respecto a otros. Esto permite identificar puntos como miembros de un grupo, si cada uno de estos puntos está dentro de una distancia entre ellos establecida como parámetro (parámetro *épsilon* o *eps*). A su vez, cada grupo debe contener al menos un número mínimo de puntos, también establecido como parámetro (parámetro *min_samples*). En la Figura 3, los puntos deben estar cerca entre sí (el radio de los círculos es el parámetro *epsilon* ϵ) y para que ese grupo sea considerado válido debe tener al menos cierta cantidad de puntos igual a *min_samples*. DBSCAN por lo tanto toma estos únicos parámetros y devuelve los clusters o grupos. La cantidad de grupos es uno de los productos de esta técnica. Cuando hay observaciones que no cumplen con las condiciones para pertenecer a un grupo, éstas quedan sin agrupar (el caso que se observa en la Figura 3 por fuera de la zona Border y es catalogado como ruido).

24. Ester, Kriegel, y Xu, «A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise».

| Figura 3. Esquema de DBSCAN



Fuente: Amini et al (2014)

Algunas decisiones metodológicas y sus efectos

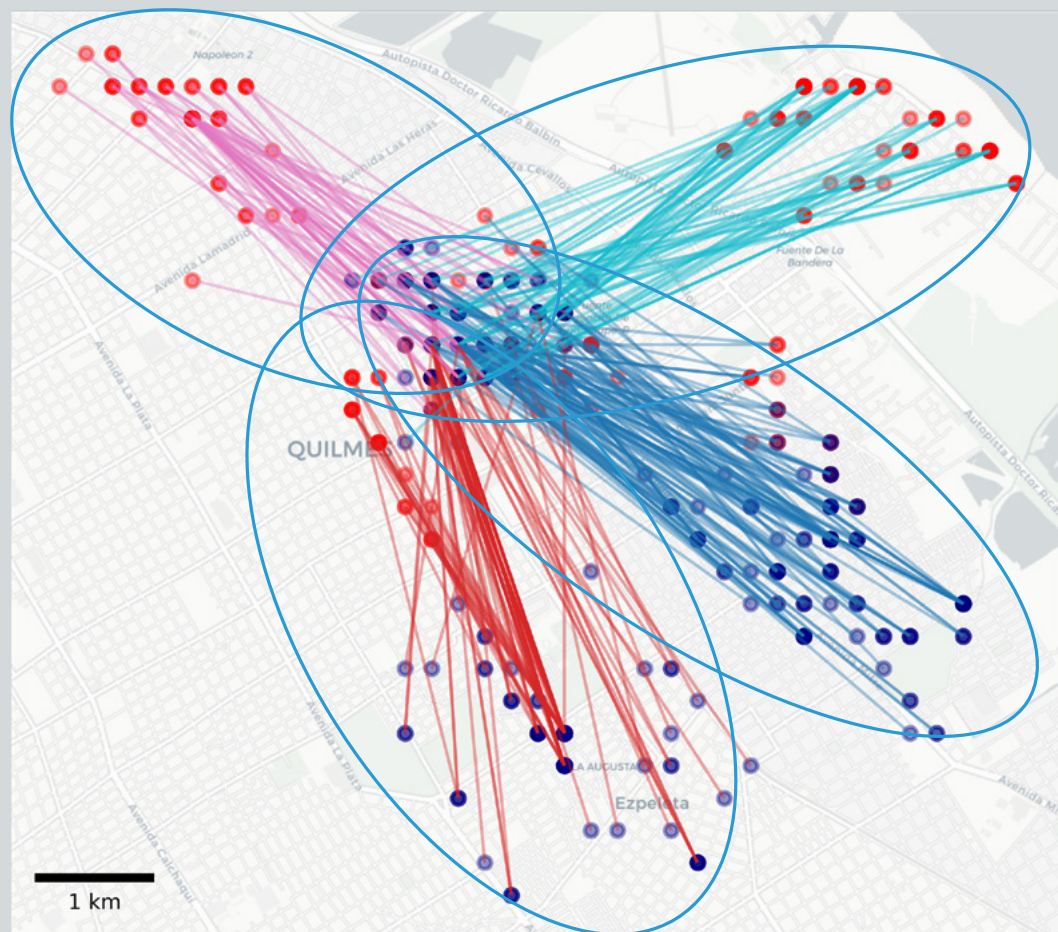
Se le va a aplicar esta técnica de clusterización a las etapas de viaje que se les pudo imputar y validar un destino y que además cumplen con las condiciones de cubrir una distancia de menos de cinco kilómetros y que el origen o el destino se encuentra en el área de influencia del ferrocarril Roca. La información utilizada por este algoritmo son las coordenadas de origen y de destino de cada etapa de viaje.

La técnica de DBSCAN puede ser aplicada tanto a un espacio de dos o más dimensiones. Como se observó previamente, un viaje o etapa puede ser definido por las dos dimensiones de su origen (latitud/longitud de origen) o por las dos dimensiones de su destino (latitud/longitud de destino). Geométricamente, un origen o destino se encuentra representado por un punto. Sin embargo, DBSCAN puede también tomar las cuatro dimensiones de una etapa o viaje (las dos coordenadas del origen y las dos de los destinos). El siguiente gráfico muestra un ejemplo²⁵ donde coordenadas de orígenes y destinos de etapas de viajes que acceden a una estación de ferrocarril se clasifican en cuatro clusters teniendo en cuenta la ubicación geográfica y densidad de estas coordenadas de origen y destino.

25. Para este ejemplo se utilizó una muestra de datos de la estación Quilmes y es solo a fines ilustrativos de cómo funciona el método DBSCAN para agrupar etapas de viajes.



Figura 4. Ejemplo de generación de clusters con coordenadas de origen y de destino



Fuente: Elaboración propia, datos SUBE

Al usar las coordenadas de origen y de destino con la técnica de DBSCAN se debe tener en cuenta que la distancia para considerar “cercanas” a las observaciones (el parámetro *epsilon*) suma a las distancias en 4 dimensiones (latitud y longitud tanto en origen y como en destino). Es decir, tanto el punto de origen como el de destino deben ser “cercanos”, y la distancia entre los puntos de origen se suma a la distancia entre los puntos de destino y es esa distancia acumulada la que considera el parámetro *epsilon*.

Los resultados de cada técnica de clusterización son muy sensibles a los parámetros que cada una requiere (en el caso de DBSCAN los parámetros epsilon y cantidad mínima de casos). Existe una técnica denominada Silhouette Score²⁶ que intenta representar la coherencia interna de cada cluster y mide cuán cercana es cada observación en un cluster de los puntos pertenecientes a los otros clusters. El coeficiente Silhouette varía entre -1 y 1. Por un lado, 1 indica mucha distancia con respecto a los otros clusters vecinos (y por ende más coherencia interna del clusters). Por otro, indica que esa observación está cerca de la línea de decisión sobre si pertenece a este cluster o a otro. Por último -1 indica que puede haberse clasificado en el cluster equivocado. Esta técnica permite evaluar agrupamientos obtenidos con diferentes parámetros y seleccionar cuál es la combinación de parámetros más adecuada.

26. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html



Se utilizan en este trabajo técnicas de clusterización para identificar corredores de alta densidad de viajes de corta distancia. La técnica de DBSCAN puede ser aplicada tanto a un espacio de dos o más dimensiones. Clasificamos cada viaje o etapa por las dos dimensiones de su origen (latitud/longitud de origen) y por las dos dimensiones de su destino (latitud/longitud de destino). Es decir, tanto el punto de origen como el de destino deben ser “cercanos” para pertenecer al mismo cluster.

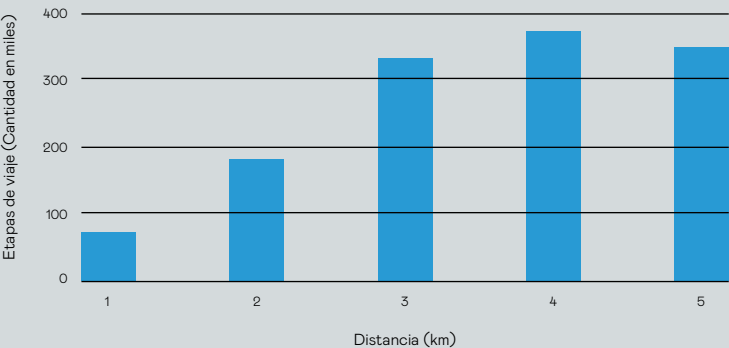
Identificación de corredores con alta concentración de etapas sustituibles por micromovilidad





Hasta este punto, se habló de manera indistinta de etapas y viajes. Sin embargo, es importante una distinción entre ambos a la hora de evaluar la sustitución de cualquiera de ellos por modos propios de la micromovilidad (de los cuales la bicicleta es hoy el más característico). En primer lugar, la distancia recorrida es la principal variable de interés. Se asume que etapas o viajes muy largos no son plausibles de ser realizados en bicicleta o en sistemas de micromovilidad, por lo tanto, se define ese umbral en cinco kilómetros²⁷. La **figura 5** muestra cómo es la distribución de las etapas de viajes cortos según la distancia a las estaciones del ferrocarril Roca.

Figura 5. Etapas de viaje según distancia
a la estación del ferrocarril roca



Fuente: Elaboración propia,
datos SUBE

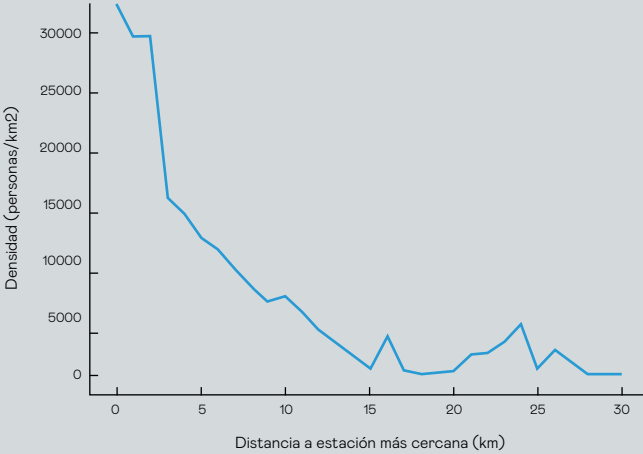
*Viajes de menos de cinco kilómetros con origen o
destino en el área de influencia del ferrocarril Roca

27. Se eligió como distancia de corte en cinco kilómetros por representar un viaje de corta-media distancia que puede ser realizado en bicicleta en unos 20 minutos por personas adultas en condiciones físicas normales. Este parámetro podría ajustarse de acuerdo a las características propias del lugar analizado.

Viajes de corta distancia o alguna de las etapas de corta distancia en un viaje multimodal (por ejemplo, un viaje en colectivo en una primera etapa y en tren en una etapa posterior) se pueden reemplazar por una opción de movilidad activa o micromovilidad. Este reemplazo se puede dar a través del uso de movilidad propia (una bicicleta) o un sistema de uso compartido (bicicletas públicas o sistemas de micromovilidad eléctrica compartida). Las decisiones de los usuarios respecto al uso de su bicicleta propia para una etapa del viaje son afectadas por la disponibilidad de estacionamiento seguro en el lugar de transferencia o la posibilidad de llevar la bicicleta dentro del vehículo. La movilidad compartida brinda mayor flexibilidad, ya que permite reemplazar etapas sin contar con vehículo propio. Por ejemplo, el caso de una persona que aborda un colectivo a la mañana para ir a una estación de tren luego sube a ese tren y finalmente para su destino final hace otra etapa corta en otro modo. Las etapas cortas de ese viaje son sustituibles por bicicleta privada solo en la medida en que esa persona pueda llevar consigo la bicicleta en todo momento o estacionarla en la estación de ferrocarril de origen. En este último caso, la etapa corta final sería sustituible sólo en la medida en que en la estación de ferrocarril de destino de la segunda etapa haya un sistema de bicicletas públicas compartidas para completar su etapa final a destino. Las políticas que consideren la sustitución de viajes deben tener estos factores en consideración.

Este trabajo se enfoca principalmente en las etapas de los viajes que tienen un origen o un destino en el área de influencia del Ferrocarril Roca. Esto incluye a los viajes de una etapa única que tienen por origen o destino la zona de influencia de la estación pero que no necesariamente utilizan luego el ferrocarril como parte del viaje. Las estaciones de ferrocarril Roca en la zona sur de la RMBA son centros atractores de viajes en sí mismos por diversos motivos. Por un lado, constituyen centros comerciales y administrativos, y, al mismo tiempo, suelen ser las zonas de mayor densidad poblacional como muestra la [Figura 6](#).

**Figura 6. Densidad poblacional promedio por radio
censal según distancia a la estación de FC más cercana**



Fuente: Elaboración propia,
datos CENSO 2010 y
Ministerio de Transporte

Por último, y lo más importante para este trabajo, las estaciones son centros de trasbordo importantes, ya sea porque conectan con el ferrocarril o porque conectan con otras líneas de colectivos. Dado el interés en la movilidad activa y la micromovilidad, el foco estará puesto en las etapas cortas de menos de cinco kilómetros y que tengan un origen o un destino en la zona de influencia de una estación del ferrocarril Roca.

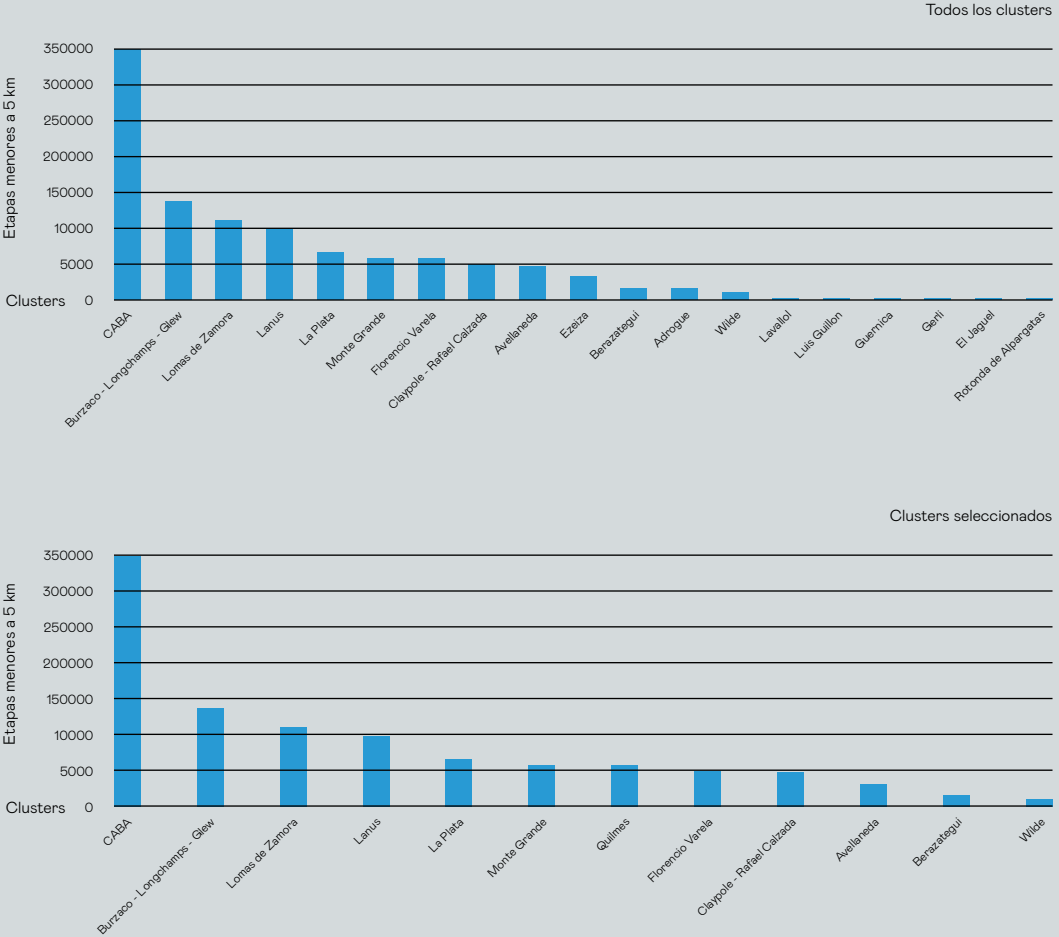
Una vez identificadas las etapas de viajes cortos, se les aplicó la técnica de clustering DBSCAN para identificar y cuantificar los principales corredores con alta densidad de viajes de corta distancia. La **Figura 7** muestra la demanda actual por cluster y la **Figura 8** muestra los clusters obtenidos en el mapa. Como parámetros fueron utilizados 1000 metros para el área de influencia (parámetro *eps*) y 1500 para la cantidad mínima de casos (etapas) dentro de un cluster. Un análisis posterior de cada cluster va a permitir identificar y cuantificar corredores con alta densidad de viajes cortos. En una primera visualización se puede observar que el algoritmo detectó una gran cantidad de clusters. Posteriormente se procedió a un análisis más detallado de los clusters, el acceso a las estaciones y se incluye como caso de estudio la estación Quilmes.

Para comprender la magnitud se muestra a continuación un gráfico con la cantidad de viajes cortos en transporte público para cada cluster. El

cluster ubicado en la Ciudad de Buenos Aires (CABA) es el de mayor importancia en término de corredores de alta densidad de viajes cortos, principalmente por la Estación Constitución²⁸ (el ferrocarril Roca cuenta con otra estación en CABA, Hipólito Yrigoyen, de menor caudal). Se pueden observar también corredores importantes dentro del corredor Burzaco, Longchamps y Glew (unificados por la técnica de clusters en un único grupo), como así también en torno a los centros urbanos de Lomas de Zamora, Lanús, La Plata, Monte Grande y Quilmes.

28. La estación Constitución es el principal centro de trasbordo de la Ciudad de Buenos Aires ya que confluyen todos los ramales del ferrocarril Roca, la línea C de subterráneo y una gran cantidad de líneas de colectivo que conectan con la ciudad y el área metropolitana.

Figura 7. Clusters con etapas en el área de influencia del corredor Roca (etapas de menos de cinco kilómetros)

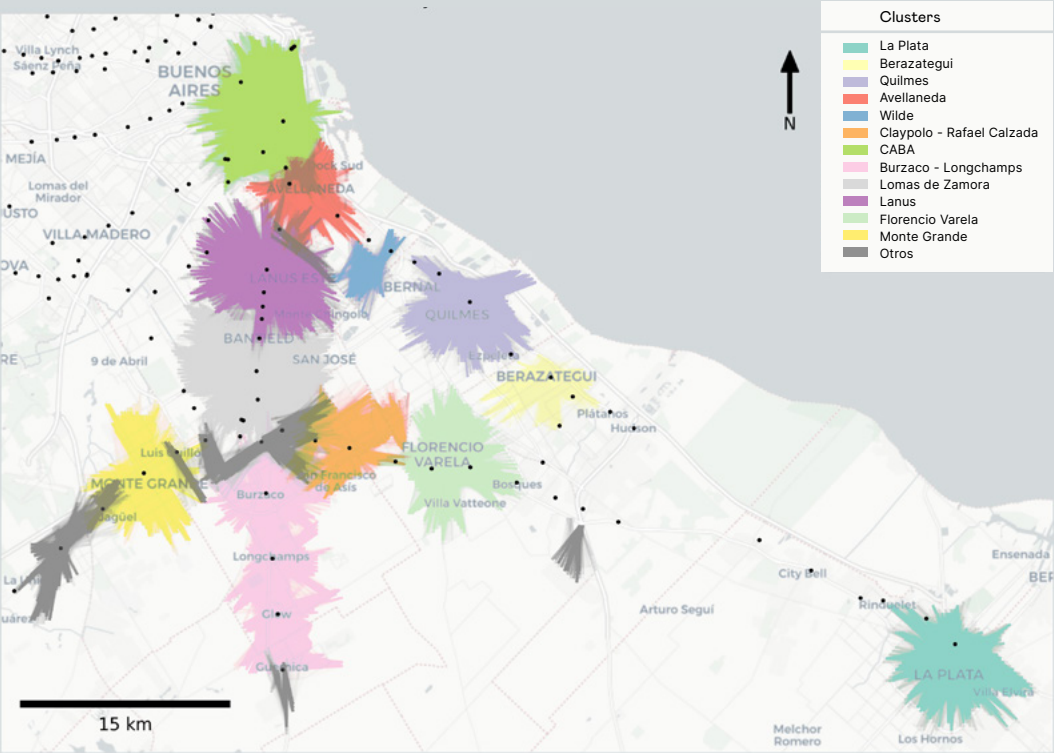


Fuente: Elaboración propia, datos SUBE



En el siguiente mapa se observa la ubicación en el espacio de estos clusters. Se agruparon los de magnitudes menos significativas con el objetivo de simplificar la visualización.

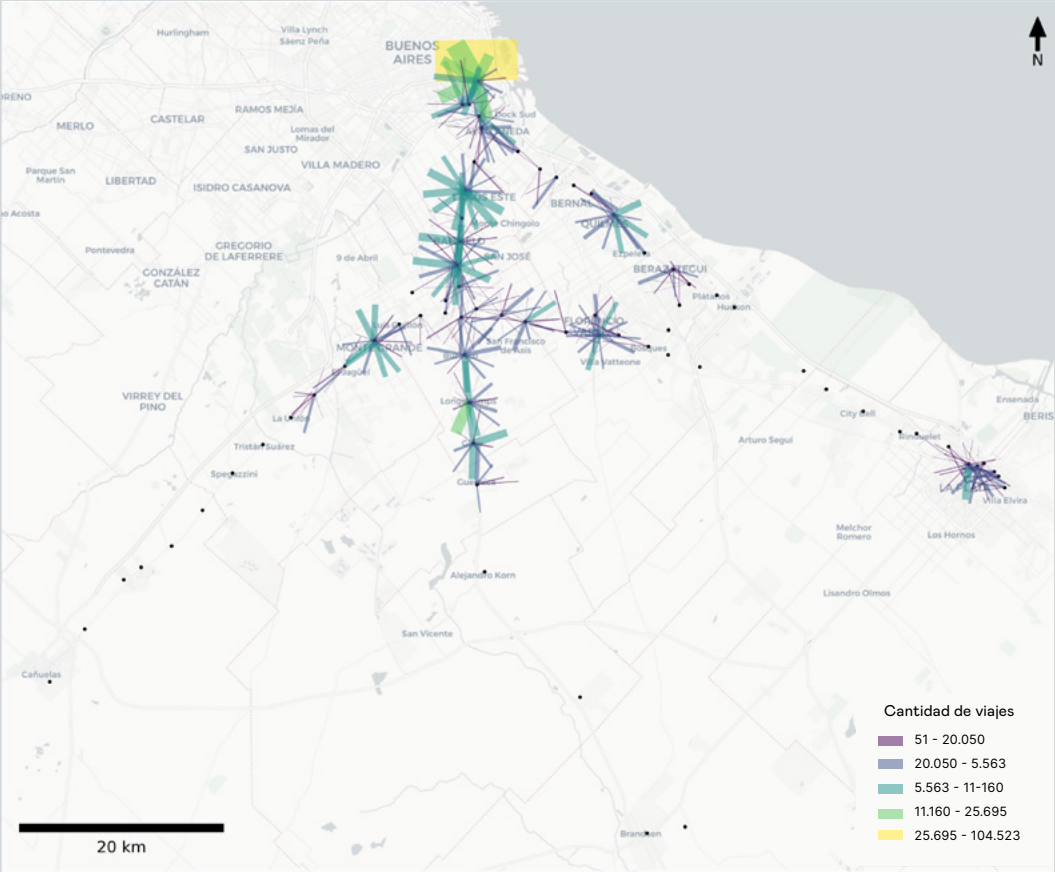
Figura 8. Clusters con etapas en el área de influencia del corredor Roca (etapas de menos de cinco kilómetros)



Fuente: Elaboración propia, datos SUBE

Dado a que el mapa anterior muestra la totalidad de las etapas de viajes, se procedió a utilizar técnicas de reducción y agregación de información con el objeto de minimizar la saturación y el ruido. De este modo podemos obtener información significativa a partir de los datos. El siguiente mapa nos permite identificar mejor los corredores agregando en torno a orígenes y destinos comunes y representando las líneas de deseo de los viajes de corta distancia en el área de influencia del ferrocarril Roca.

Figura 9. Líneas de deseo (clusters en etapas de viajes de menos de cinco kilómetros)



Fuente: Elaboración propia, datos SUBE

Un análisis con mayor detalle de los clusters y la accesibilidad a cada estación permite cuantificar las líneas de deseo (o cantidad de viajes) por cada corredor de acceso a cada estación. En el siguiente gráfico se muestra una selección de estaciones. En el anexo se pueden ver los resultados para el conjunto de estaciones del ferrocarril Roca.

Figura 10. Corredores agregados según cluster por estación (estaciones seleccionadas)



Fuente: Elaboración propia, datos SUBE



Identificamos los corredores agregando en torno a orígenes y destinos comunes y representando las líneas de deseo de los viajes de corta distancia en el área de influencia del ferrocarril Roca. Esto permitió cuantificar las líneas de deseo (o cantidad de viajes) por cada corredor de acceso a cada estación.

Caso de estudio: Estación Quilmes



Prohibido transitar
por zona de vías 3

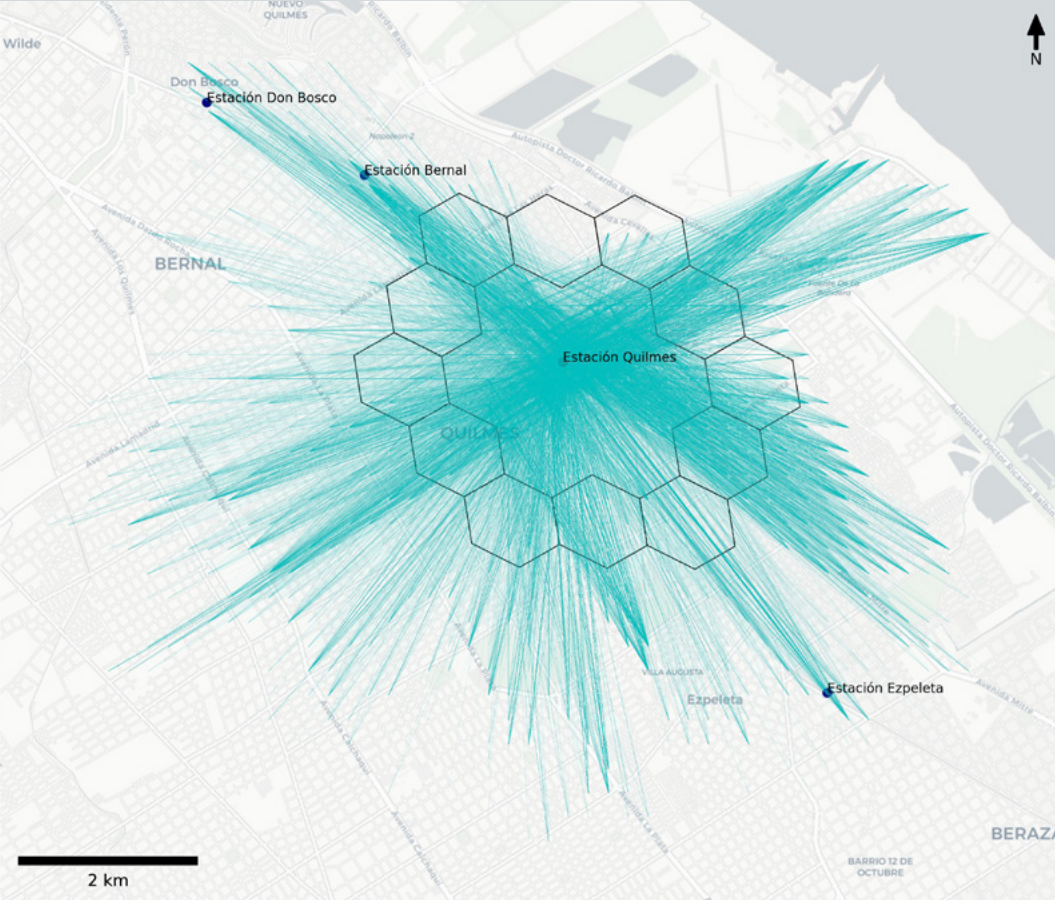


Para analizar con mayor profundidad el uso y el alcance que pueda tener esta herramienta para identificar corredores de viajes utilizando datos de la SUBE, se analiza con mayor detalle el caso de las etapas cortas que acceden a la estación Quilmes y se exploran diferentes técnicas de reducción y agregación de datos para producir información relevante.

Etapas de viajes cortos sin agregación de datos

La **Figura 11** muestra todas las etapas de viajes de menos de cinco kilómetros que tienen como origen o destino la Estación Quilmes. Si bien es posible identificar algunos patrones, la gran cantidad de información no permite individualizar corredores según su importancia relativa. Para poder agregar esta información, se va a particionar el espacio en celdas hexagonales de igual tamaño (en este caso son hexágonos con aristas de aproximadamente 500 metros) y se toma un anillo de hexágonos que bordea el área de influencia de la estación. Esta técnica va a permitir agregar (como se va a observar más adelante) a las etapas de viajes en distintos corredores y cuantificar su cantidad de viajes.

Figura 11. Etapas de viajes de menos de cinco kilómetros.
Estación Quilmes.



Fuente: Elaboración propia,
datos SUBE

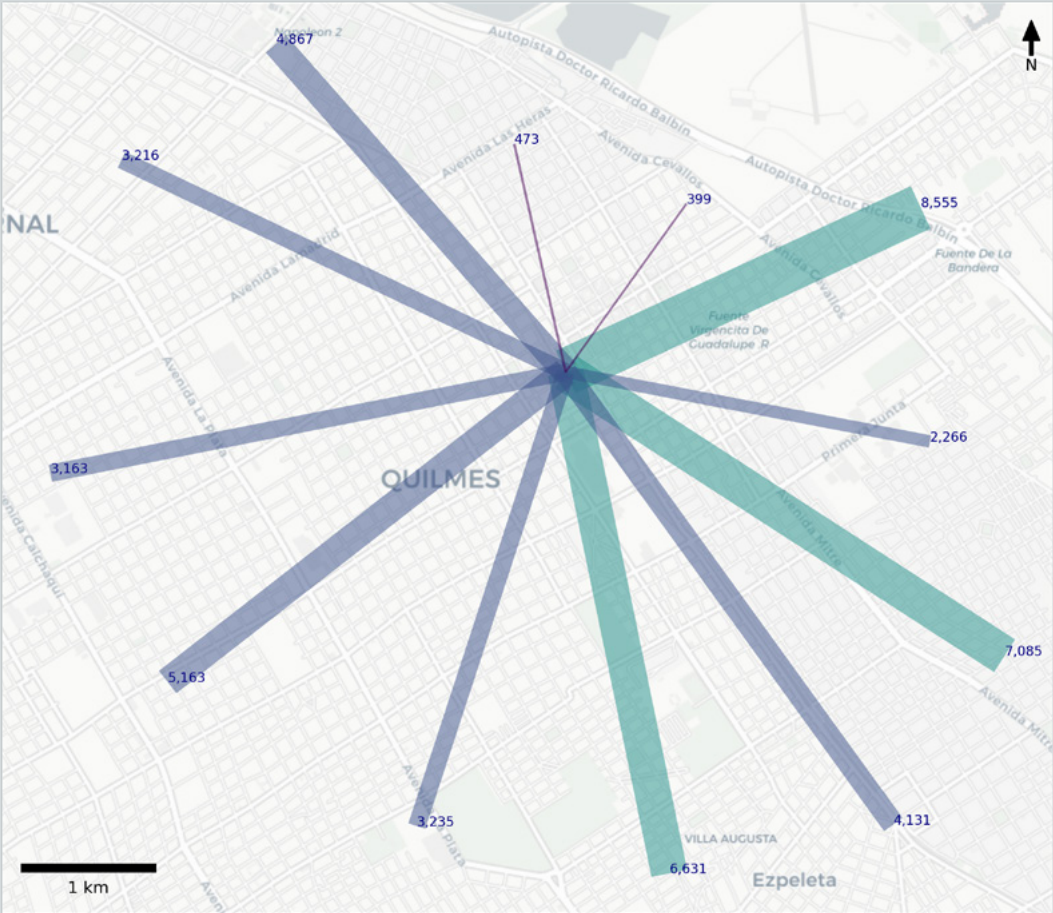
En este primer mapa, donde no se les aplica a los datos ninguna técnica de agregación, es más complicado identificar los corredores más significativos. De cualquier manera, se puede ver el rol estructurante del ferrocarril. En el caso de la estación Quilmes, se puede observar que es un polo principal de atracción de viajes, y que la estación Bernal se encuentra fuertemente vinculada a la estación Quilmes demarcando claramente un corredor de los más voluminosos (que incluso podría incluir a la estación Don Bosco). El otro gran corredor

según importancia por su magnitud es el que vincula estación Quilmes con el barrio Villa Augusta al sur más cercano a la estación Ezpeleta. Otros corredores que destacan por su volumen son, hacia el este, el que vincula la estación con el propio barrio de Quilmes y con la Ribera. El resto de los corredores de menor magnitud son los que vinculan la zona Sureste del partido con la estación.

Agregación de viajes en torno a estaciones

La **Figura 12** busca identificar y cuantificar corredores de viajes, quitando ruido a la presentación de los datos sin ningún tipo de agregación (como se ve en el mapa anterior). El anillo de hexágonos que rodea el área de influencia de la estación permite identificar a los corredores, visualizar mejor el ángulo de cada línea de deseo y asignar cantidad de viajes en cada corredor. A su vez, se normalizan los destinos hacia la estación (corrigiendo para que todos los viajes que se encuentran en el área de influencia tengan destino en la estación de Ferrocarril). El siguiente mapa muestra el resultado de esta agrupación basado en las etapas de viajes de menos de cinco kilómetros que corresponden al Cluster de la estación Quilmes. El ancho del corredor está dado por el volumen de viajes dentro de cada agrupamiento. La coordenada de origen y de destino de cada corredor agrupado se calcula como un promedio de las coordenadas de todos los viajes de esa agrupación. Este mapa incluye también la cuantificación de etapas de viajes cortos en transporte público para cada corredor.

Figura 12. Etapas de viajes de menos de cinco kilómetros.
Estación Quilmes.



Fuente: Elaboración propia,
datos SUBE

Como se puede observar, la técnica presentada en este mapa logra agregar de manera efectiva etapas similares dentro de una misma línea de deseo. Por ejemplo, las que vinculan la Estación Quilmes con el Este (la Ribera de Quilmes) ahora quedan conformadas por un único corredor. Algo similar ocurre con las provenientes del Sureste, ahora agregadas en solo dos corredores. Por otro lado, las etapas que vinculan las estaciones de Don Bosco y Bernal con la estación Quilmes se mantuvieron como corredores separados.

Asignación a la red vial

Por último, se procedió a agregar estas etapas ya no por sus orígenes o destinos, sino por sus recorridos, o, mejor dicho, por los segmentos de calle que los desplazamientos de estas etapas atraviesan. La información de la tarjeta SUBE disponible no contiene información sobre recorridos, por lo que se simuló, a partir de los orígenes y destinos, un ruteo teniendo en cuenta la vialidad del Partido de de Quilmes²⁹, uniendo los pares orígenes y destinos a través del camino más corto. Si bien es cierto que los viajes en transporte público no necesariamente siguen esta lógica, los viajes en bicicleta o con micromovilidad, al tener más grados de libertad, tienden a aproximar los recorridos al camino más corto y por lo tanto se considera que esta aproximación es apropiada.

Para realizar esta asignación, primero se dividió el recorrido simulado de cada etapa en los segmentos de calle que atraviesa. Luego, se procedió a obtener un agregado de cantidad de etapas por cada segmento de calle. Como resultado, la **Figura 13** nos muestra las etapas de viajes en cada segmento de calle, generando corredores troncales que parten de la estación y se ramifican en forma radial, para el conjunto del cluster Quilmes. En esta técnica se trabaja con todas las etapas de viajes pertenecientes al cluster (similar a la **Figura 11**) y no solo con las etapas que tienen origen o destino en la estación Quilmes (como en la **Figura 12**).

29. La librería `osmnx` en Python permite simular viajes con un origen y destino utilizando la red vial de Open Street Maps.

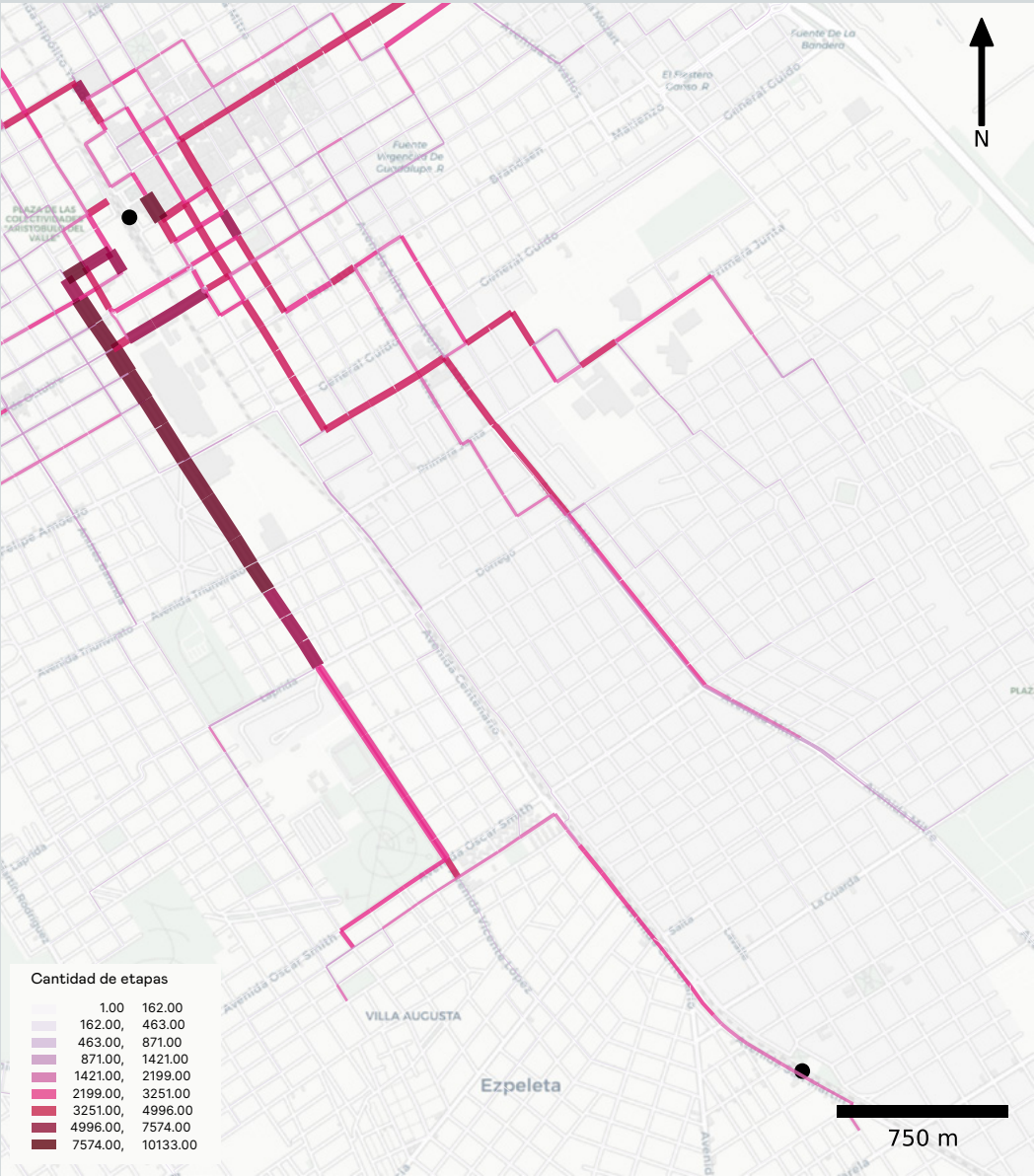
Figura 13. Ruteo de etapas en el Cluster Quilmes
(etapas de menos de cinco kilómetros)



Fuente: Elaboración propia, datos SUBE

En la **Figura 14** se ofrece acercamientos con mayor nivel de resolución y detalle a los corredores con mayor caudal, principalmente los que discurren desde la Estación Quilmes hacia el Sur.

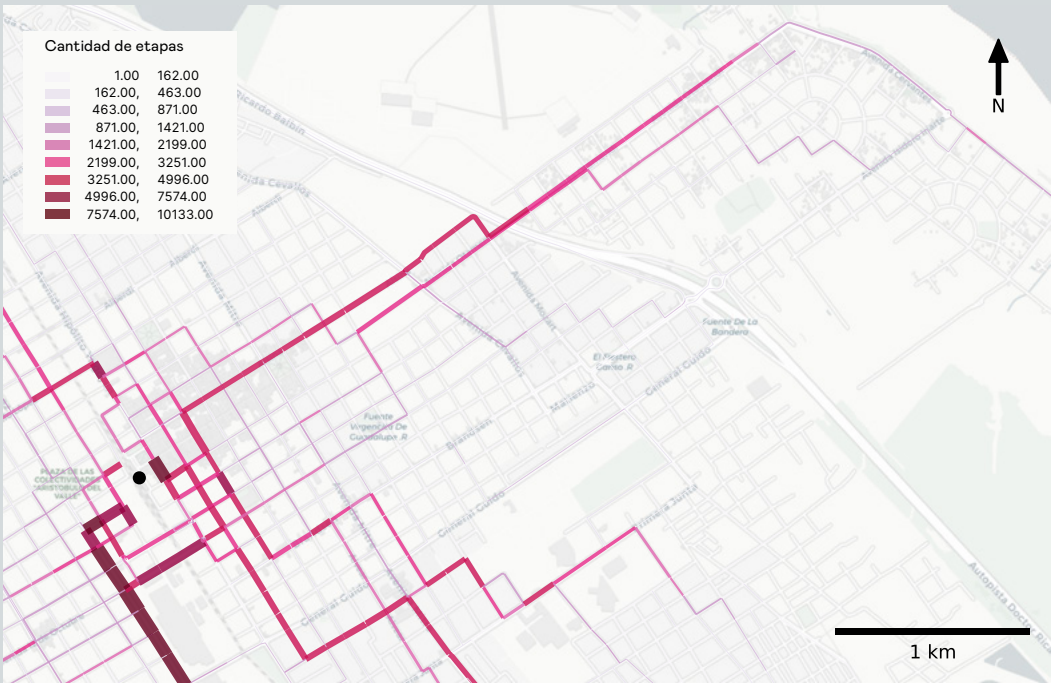
Figura 14. Ruteo de etapas en el Cluster Quilmes
(etapas de menos de 5) - Detalle Sur



Fuente: Elaboración propia, datos SUBE

Por último, la **Figura 15** muestra los corredores de mayor caudal que se alejan hacia el Este desde la misma Estación.

Figura 15. Ruteo de etapas en el Cluster Quilmes
(etapas de menos de 5) - Detalle Oeste



Fuente: Elaboración propia,
datos SUBE

Por lo que se puede observar, esta técnica de agregación genera recorridos comunes para aquellas etapas que vinculan el centro de Quilmes con la Ribera al Este, como así también con el Sureste. Esta técnica no evita que existan pequeñas diversificaciones que podrían unificarse, como puede visualizarse en algunos recorridos que podrían tomar caminos comunes pero tienen pequeñas desviaciones.

En términos de magnitud, el corredor más voluminoso es el que se encuentra sobre la Avenida Vicente Lopez dado que recoge todas las etapas de viajes provenientes de la zona de la Estación Ezpeleta y

de la zona del barrio Villa Augusta, como así también otras etapas de menor magnitud provenientes del Sur. Por último, se puede observar que los segmentos de calle de la zona periférica a la Estación Quilmes se caracterizan por un alto volumen de etapas de viajes menores a cinco kilómetros.

30. CROW, «Manual Diseño Tráfico Bicicletas».

Demanda de micromovilidad

Las estimaciones de demanda potencial y corredores OD a partir de información de SUBE pueden ser insumos útiles para planificar el desarrollo de infraestructura ciclista³⁰ que es el uso que analizamos en este trabajo. Se puede considerar que los viajes en transporte público son un proxy de cuáles son los patrones de movilidad de la población y cuáles viajes cortos son potencialmente realizables en modos como la bicicleta o alternativas de micromovilidad. En la situación actual, existe una demanda de acceso al ferrocarril en bicicleta en la RMBA que puede ser potenciada con infraestructura de mayor calidad.

Figura 16. Estaciones de ferrocarril y uso de la bicicleta en el AMBA



Estaciones de ferrocarril en el AMBA. Acceso al tren (izquierda), guarda bicicletas en estación (derecha). Fotos Sebastián Anapolsky

Como explicita el objetivo del trabajo, se procura desarrollar una metodología para determinar corredores de alta densidad de viajes plausibles de ser realizados en vehículos propios de la micromovilidad. Este tipo de información puede ser un insumo importante para el diseño y planificación de otras políticas complementarias que busquen incentivar los modos activos y la micromovilidad. Sin embargo, el diseño de detalle de la red ciclista, estacionamientos, políticas de operación de los ferrocarriles, entre otras, escapan a los objetivos de este trabajo. Dicho esto, el análisis de los datos masivos constituye una herramienta que puede dar información sobre la demanda potencial, que deberá ser complementada por otras técnicas una vez que se decida avanzar con etapas más avanzadas del diseño y ejecución. Los manuales y guías existentes establecen algunos lineamientos generales útiles para una planificación preliminar de la infraestructura ciclista en base a la demanda potencial estimada (Figura 17).

Figura 17. Esquema de opciones para secciones viales en zonas urbanas

| Categoría de la vía | Velocidad máxima del tráfico motorizado (km/h) | | Volumen del tráfico motorizado (vm/día) | Categoría de red ciclovial | | |
|---------------------|--|---------------|---|--|---|--|
| | | | | red básica (I _{bicicleta} > 750/día) | cicloruta (I _{bicicleta} 500-2.500/día) | cicloruta principal (I _{bicicleta} >2.000/día) |
| | n/a | | 0 | C I C L O V Í A A P A R T A D A | | |
| Calle de servicio | Velocidad peatón o 30 km/h | | 1 - 2.500 | tráfico mixto | | ciclocalle (con preferencia) |
| | | | 2.000 - 5.000 | | | ciclovía o ciclobanda (con preferencia) |
| | | | <4.000 | ciclobanda o ciclovía | | |
| Vía recolectora | 50 km/h | 2x1 pistas | irrelevante | <div></div> <div>ciclobanda o calle paralela</div> | | |
| | | 2x2 pistas | | | | |
| | | 70 km/h | | | ciclovía, ciclomotor / ciclobanda o calle paralela | |

Vm: vehículos motorizados
2x1: dos pistas unidireccionales. uno de cada lado de la calle
2x2: una vía bidireccionales en cada lado de la calle

Fuente: CROW



Las estimaciones de demanda potencial y corredores OD a partir de información de SUBE pueden ser insumos útiles para planificar el desarrollo de infraestructura ciclista (...) En la situación actual, existe una demanda de acceso al ferrocarril en bicicleta en la RMBA que puede ser potenciada con infraestructura de mayor calidad.

Conclusiones





La utilización de datos de la tarjeta SUBE y las técnicas de análisis de datos implementadas permitió construir cadenas de viajes e identificar los principales corredores en el área de influencia del Ferrocarril Roca, en particular, analizando viajes de corta distancia. La utilización técnicas de machine learning nos permitio detectar patrones espaciales de viajes para localizar y cuantificar corredores de transporte con alta densidad de viajes de corta distancia para planificar e informar políticas de transporte que promuevan la movilidad sustentable.



Los datos obtenidos de las transacciones de los sistemas de pago electrónico del transporte público y los registros de los GPS a bordo de los vehículos de transporte brindan información valiosa para analizar una gran variedad de problemáticas relacionadas con la planificación del transporte. Si bien estos datos son recolectados por cuestiones administrativas relacionadas con la gestión del sistema, un preprocesamiento de esta información utilizando la georeferenciación de cada transacción ofrece oportunidades para el análisis del comportamiento del sistema y el análisis de patrones de movilidad en transporte público.

En los últimos años se desarrollaron distintas metodologías que justifican la utilización de estos datos y se desarrollaron técnicas para la imputación de datos faltantes, como la inferencia y validación de los destinos de cada transacción y la identificación de las cadenas de viajes para cada usuario, sus etapas de viaje y transferencias. Esto da la posibilidad de inferir matrices de origen y destino de viajes con un gran nivel de detalle.

A su vez, las técnicas de aprendizaje automático (machine learning) ofrecen herramientas nove-

das para el análisis de información espacial. La utilización de estas técnicas nos permite identificar patrones de viajes para planificar e informar políticas de transporte. Este trabajo se enfoca en cuantificar corredores de transporte con alta densidad de viajes de corta distancia, información de especial relevancia para planificar políticas de movilidad sustentable.

La utilización de datos de la tarjeta SUBE y las técnicas de análisis de datos implementadas permitió construir cadenas de viajes e identificar los principales corredores en el área de influencia del Ferrocarril Roca, en particular, analizando viajes de corta distancia (menores a los cinco kilómetros). Sin embargo, la utilización de estas técnicas puede servir también para analizar una gran variedad de problemáticas, como comprender patrones de viajes a nivel metropolitano, analizar transferencias, elaborar matrices de origen y destino de viajes para diferentes rangos horarios, días de la semana o épocas del año, identificar demanda en corredores específicos, o analizar el comportamiento de grupos de usuarios teniendo en cuenta datos asociados a la tarjeta de pago, como el pago por atributo social o el género.

Anexo





Corredores agregados según cluster por estación

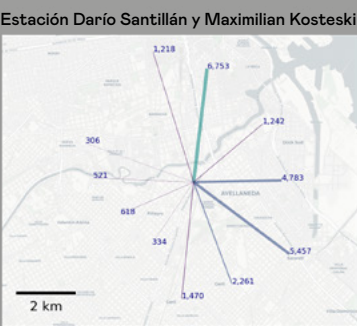
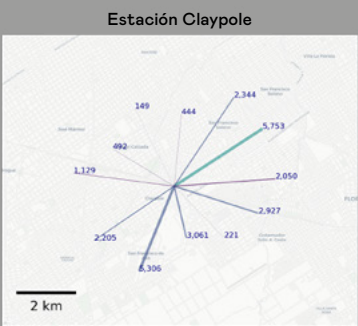
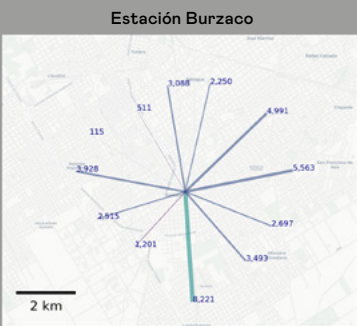
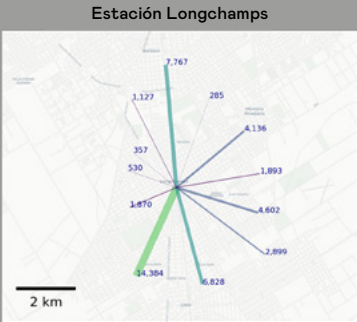
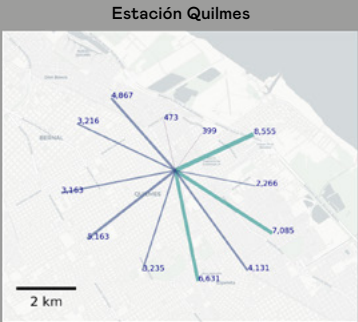
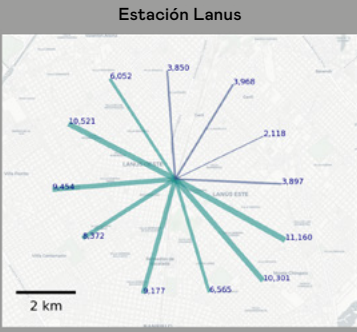
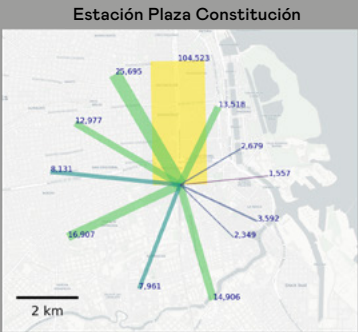
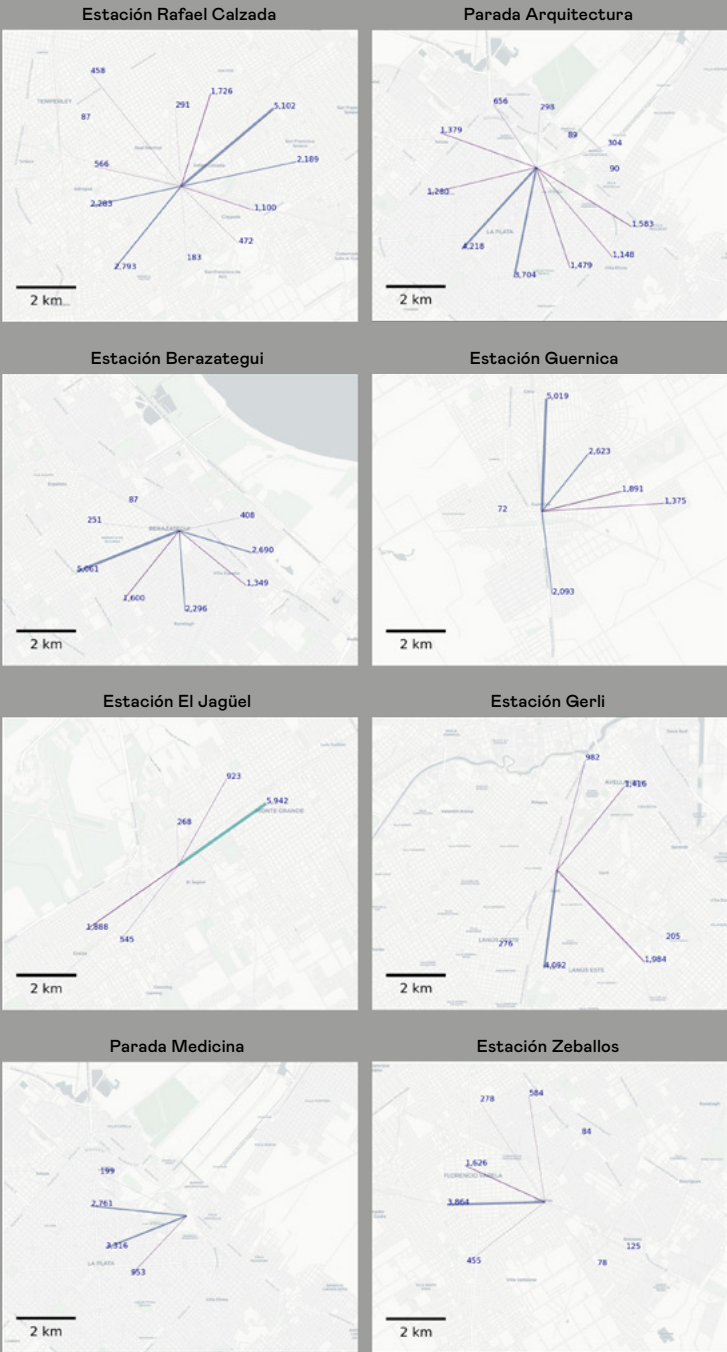


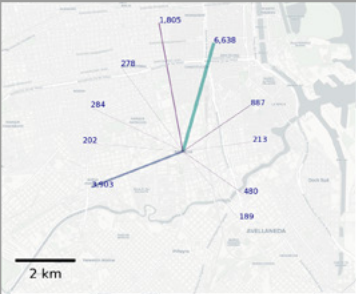
Figura 18. Líneas de deseo por estación (1)

Fuente: Elaboración propia, datos SUBE

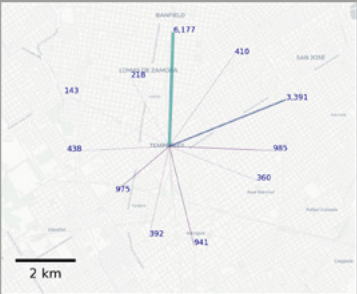




Estación Sola



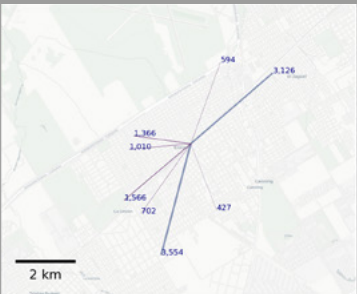
Estación Temperley



Estación Remedios de Escalada



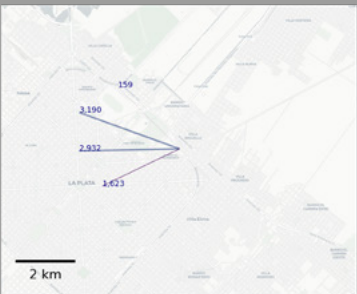
Estación Ezeiza



Estación Ardigó



Parada Periodismo



Estación Gobernador Monteverde

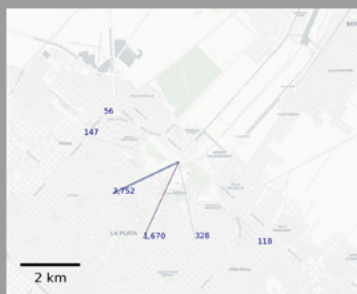


Parada Diagonal 73





Estación Informática



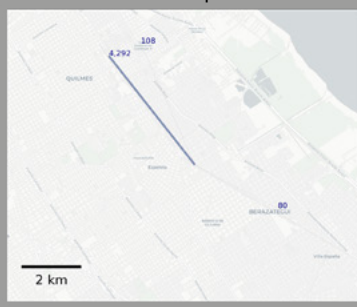
Estación Wilde



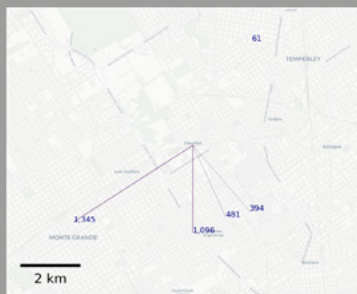
Estación Hospital Español



Estación Ezpeleta



Estación Llavallol



Estación Villa Domínico



Estación Don Bosco



Estación Ranelagh



Bibliografía







- Ester, Martin; Kriegel, Hans-Peter; Sander, Jörg; Xu, Xiaowei (1996). Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M. (eds.). *A density-based algorithm for discovering clusters in large spatial databases with noise*. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231.
- Cardell, Martin, Møller Holm, Thomas. «How Micromobility Is Moving Cities into a Sustainable Future». Accedido 3 de mayo de 2021. https://www.ey.com/en_gl/automotive-transportation/how-micromobility-is-moving-cities-into-a-sustainable-future.
- «Cities Safer By Design». Accedido 19 de marzo de 2021. <https://publications.wri.org/citiessafer/>.
- CROW. «Manual Diseño Tráfico Bicicletas». www.crow.nl, s. f.
- Edwina. «Safe Micromobility». Text. ITF, 9 de enero de 2020. <https://www.itf-oecd.org/safe-micromobility>.
- Ester, Martin, Hans-Peter Kriegel, y Xiaowei Xu. «A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise», s. f., 6.
- Kadir, R. A., Y. Shima, R. Sulaiman, y F. Ali. «Clustering of public transport operation using K-means». En *2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, 427–32, 2018. <https://doi.org/10.1109/ICBDA.2018.8367721>.
- Kager, R., L. Bertolini, y M. Te Brömmelstroet. «Characterisation of and Reflections on the Synergy of Bicycles and Public Transport». *Transportation Research Part A: Policy and Practice* 85 (1 de marzo de 2016): 208–19. <https://doi.org/10.1016/j.tra.2016.01.015>.
- Mackett, R. L. «Policies to Attract Drivers out of Their Cars for Short Trips». *Transport Policy* 8, n.o 4 (1 de octubre de 2001): 295–306. [https://doi.org/10.1016/S0967-070X\(01\)00025-7](https://doi.org/10.1016/S0967-070X(01)00025-7).
- Munizaga, Marcela, Flavio Devillaine, Claudio Navarrete, y Diego Silva. «Validating travel behavior estimated from smartcard data». *Transportation Research Part C: Emerging Technologies* 44 (julio de 2014): 70–79. <https://doi.org/10.1016/j.trc.2014.03.008>.
- Oeschger, Giulia, Páraic Carroll, y Brian Caulfield. «Micromobility and Public Transport Integration: The Current State of Knowledge». *Transportation Research Part D: Transport and Environment* 89 (1 de diciembre de 2020): 102628. <https://doi.org/10.1016/j.trd.2020.102628>.
- Transport for London. «London Cycling Design Standards», Edition de 2014.
- 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, Nº1, 2007, pp. 1–14.
- 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, Nº5, 2007, pp. 376–387.

