

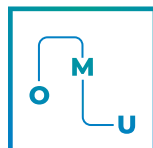
Metodología para el estudio de la movilidad con datos de Facebook

GENERACIÓN DE MATRICES ORIGEN-DESTINO
EN CIUDADES DE AMÉRICA LATINA Y ANÁLISIS
PARA BUENOS AIRES



AUTORES

MOLERES, CRISTIAN JOAQUIN
FULPONI, JUAN IGNACIO



Una iniciativa de

CAF BANCO DE DESARROLLO
DE AMÉRICA LATINA

BID
Banco Interamericano
de Desarrollo

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

Fulponi, Juan.

Metodología para el estudio de la movilidad con datos de Facebook: generación de matrices origen-destino en ciudades de América Latina y análisis para Buenos Aires / Juan Ignacio Fulponi, Cristian Moleres.

p. cm. – (Monografía del BID ; 1058)

Incluye referencias bibliográficas.

1. Sustainable transportation-Research-Latin America. 2. Sustainable transportation-Research-Argentina. 3. Data sets-Latin America. 4. Data sets-Latin America-Argentina. 5. Freedom of information-Latin America. 6. Freedom of information-Argentina. 7. Global Positioning System. I. Moleres, Cristian. II. Banco Interamericano de Desarrollo. División de Transporte. III. Título. IV. Series.

IDB-MG-1058

Códigos JEL: L91, C8, O18

Palabras clave: movilidad urbana, redes sociales, Facebook, datos, origen, destino, ciudades

Coordinación

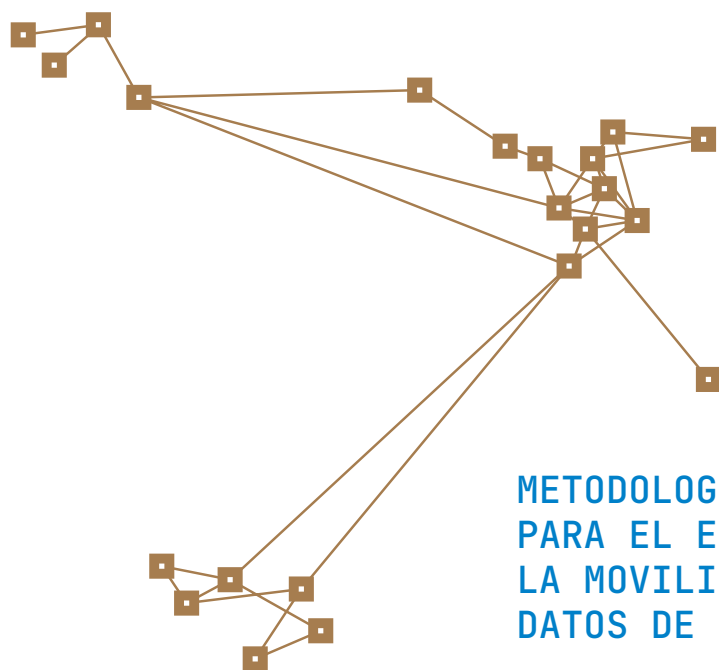
Abad, Julieta

Diseño gráfico

El Gato y La Caja

Copyright © 2022 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.





METODOLOGÍA
PARA EL ESTUDIO DE
LA MOVILIDAD CON
DATOS DE FACEBOOK

Resumen ejecutivo

El conocimiento de las características y patrones de desplazamiento de las personas en las ciudades es un insumo clave para la definición de políticas públicas de movilidad urbana, especialmente en las grandes áreas metropolitanas. Desafortunadamente, la disponibilidad de datos actualizados y confiables en muchas de las ciudades de América Latina y el Caribe (ALC) es baja, y su generación por métodos de estudio tradicionales requiere grandes esfuerzos por parte de los gobiernos locales para su recolección y procesamiento. Por este motivo, investigar nuevas fuentes de información que aprovechen los datos masivos generados por aplicaciones de dispositivos móviles se convierte en un objetivo clave para resolver esta problemática. El Observatorio de Movilidad Urbana de América Latina (OMU)^[01] busca dar respuesta a esta necesidad de información sólida, confiable y actualizada sobre el transporte y la movilidad urbana en la región.

En este trabajo se utilizaron bases de datos de la aplicación Facebook, obtenidas a través del programa *Data for Good*^[02] en colaboración con el *Development Data Partnership*^[03] y el Departamento de Trans-

[01]

Bien público regional desarrollado por el Banco de Desarrollo de América Latina —CAF— y el Banco Interamericano de Desarrollo —BID— junto con las ciudades de la región.

[02]

Iniciativa de Meta Inc. que permite a instituciones de investigación y organismos oficiales acceder en forma controlada a datos generados por los usuarios de sus aplicaciones móviles. <https://dataforgood.facebook.com/>

[03]

Asociación entre organizaciones internacionales y empresas creada para facilitar el uso de datos de terceros en investigación y desarrollo internacional. <https://datapartnership.org/>

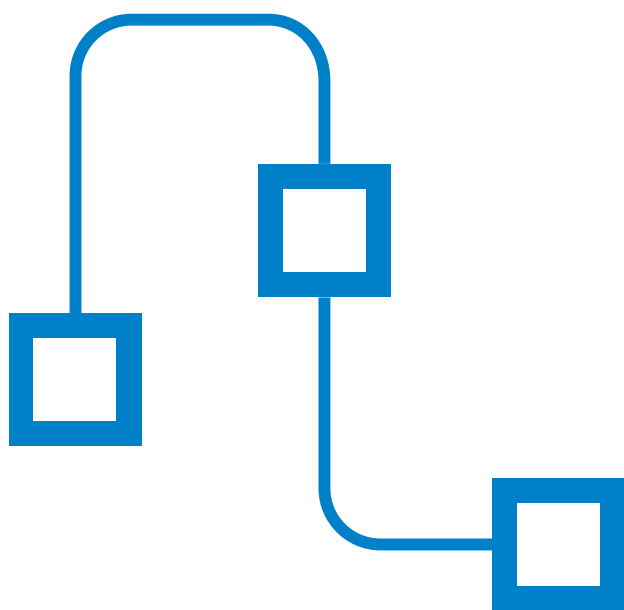
porte de la Facultad de Ingeniería de la Universidad de Buenos Aires, para elaborar una metodología que genere matrices origen-destino en las grandes áreas metropolitanas de ALC que forman parte del OMU. Estas matrices representan los desplazamientos que realiza la población entre sus residencias y sus actividades principales (trabajo, estudio, etc.) durante las horas de la mañana en abril-mayo del año 2022. Además, se realizó un análisis detallado de los resultados obtenidos en la Región Metropolitana de Buenos Aires (RMBA). Se estudió la relación entre las distancias de desplazamiento y variables geográficas y socioeconómicas de la metrópolis, y se compararon los resultados obtenidos con otros datos y estudios disponibles. Los resultados de este estudio muestran el potencial del uso de big data para el desarrollo de políticas de transporte que apoyen la implementación de sistemas de movilidad asequibles, eficientes y sostenibles, contribuyendo así a un mejor acceso de la población de bajos ingresos a oportunidades económicas y sociales. Finalmente, los resultados de este trabajo permiten identificar las características de movilidad de los

grupos de población con mayores niveles de vulnerabilidad económica y delinear potenciales cambios en la oferta de servicios de transporte público para responder de manera eficiente a las demandas de movilidad de estos grupos.



LISTADO DE ACRÓNIMOS Y ABREVIATURAS UTILIZADAS

CABA	CIUDAD AUTÓNOMA DE BUENOS AIRES
ENMODO	ENCUESTA DE MOVILIDAD DOMICILIARIA
GCBA	GOBIERNO DE LA CIUDAD DE BUENOS AIRES
HRDM	HIGH RESOLUTION DENSITY MAP
INDEC	INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS
NSE	NIVEL SOCIOECONÓMICO
OMU	OBSERVATORIO DE MOVILIDAD URBANA
RMBA	REGIÓN METROPOLITANA DE BUENOS AIRES
PBA	PROVINCIA DE BUENOS AIRES
SUBE	SISTEMA ÚNICO DE BOLETO ELECTRÓNICO
UTC	TIEMPO COORDINADO UNIVERSAL



METODOLOGÍA

PARTE 01

1.1 INTRODUCCIÓN	13
1.2 SOBRE LOS DATOS UTILIZADOS	14
1.3 METODOLOGÍA DE PROCESAMIENTO	18
1.4 LIMITACIONES METODOLÓGICAS	24

APLICACIÓN

PARTE 02

Y ANÁLISIS DE LA METODOLOGÍA PARA BUENOS AIRES

2.1.1 ANÁLISIS DE LOS FACTORES DE EXPANSIÓN	30
2.1.2 GENERACIÓN DE VIAJES	31
2.1.3 DISTANCIAS MEDIAS	32
2.1.4 ORÍGENES Y DESTINOS DE LOS VIAJES	42
2.1.5 COMPARACIÓN ENTRE LOS RESULTADOS OBTENIDOS Y OTRAS FUENTES DE INFORMACIÓN	44

CONCLUSIONES

PARTE 03

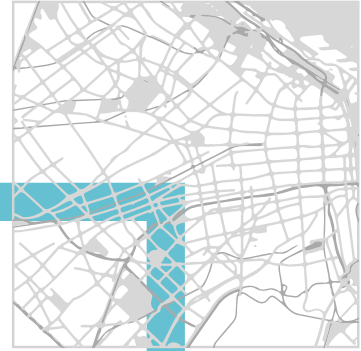
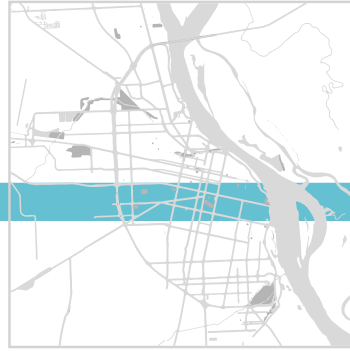
3.1 CONCLUSIONES Y PRÓXIMOS PASOS	53
-----------------------------------	----

1.2 SOBRE LOS
DATOS UTILIZADOS

1.1
INTRODUCCIÓN

1.3 METODOLOGÍA
DE PROCESAMIENTO

1.4 LIMITACIONES
METODOLÓGICAS



PARTE 01

METODOLOGÍA >>

- Desarrollo de una metodología para la identificación de los patrones de movilidad urbana. Validación metodológica. Datos utilizados: datasets, privacidad y sesgos. Procesamiento de los datos. Limitaciones metodológicas.

METODOLOGÍA

1.1 Introducción 13

1.2 Sobre los datos utilizados 14

Tabla 1. Dataset Movement between Tiles 15

Tabla 2. Dataset Movement between 17

Administrative Regions

Tabla 3. Ciudades seleccionadas 17

1.3 Metodología de procesamiento 18

Fig 1. Distribución poblacional HDRM 20

Tabla 4. Distancias medias 21

Fig 2. Flujograma de ejemplo de procesamiento de información 22

Fig 3. Distancias medias para la RMBA 23

1.4 Limitaciones metodológicas 24



SECCIÓN 1.1

Introducción

El objetivo general del trabajo consistió en generar una metodología para identificar y caracterizar los patrones de movilidad a través de datos de la aplicación Facebook en 17 regiones metropolitanas que forman parte del OMU.^[04] Para esto, se construyeron las matrices de origen y destino para cada una de las regiones administrativas de las ciudades analizadas utilizando datos de abril y mayo de 2022. Obtenidas estas matrices, se calcularon las distancias medias de los viajes de cada una de las regiones y de los distintos municipios que las conforman. La información obtenida y procesada fue recopilada de manera analítica y ordenada mediante un instrumento de visualización en un entorno Shiny. Shiny es una extensión del lenguaje de programación libre R que permite ejecutar algoritmos para crear visualizaciones de datos enriquecidas y dinámicas, y resulta de enorme utilidad a la hora de generar mapas y gráficos para distintas ciudades en un mismo entorno web.

Para realizar una comparación de los resultados obtenidos con metodologías de estudio tradicionales, se estudiaron en detalle los patrones de desplazamiento de la Región Metropolitana de Buenos Aires (RMBA)^[05] y se validaron los resultados con otras fuentes de información disponibles, aplicando modelos estadísticos para explicar y describir los flujos de personas en el conglomerado urbano más poblado de la Argentina.

Más allá de estos objetivos metodológicos y de procesamiento de datos, la meta final que se busca con este estudio es brindar herramientas a los gobiernos para establecer políticas públicas basadas en datos certeros y dinámicos como, por ejemplo, el fomento a ciertos modos de transporte dependiendo de las distancias medias observadas en las ciudades o el incentivo a la construcción de alternativas de transporte público en las zonas donde se registren grandes flujos de personas.

[04]

Se estudiaron aquellas ciudades con información publicada en la plataforma *Data for Good*.

[05]

Se eligió realizar esta validación en la RMBA por la disponibilidad de información y estudios que poseían los autores del trabajo. En el marco de los trabajos del OMU se continuará profundizando esta clase de análisis en las otras ciudades de ALC.

Sobre los datos utilizados

Para realizar este estudio, se utilizaron datos provenientes de Facebook mediante un *data partnership* entre el Banco Interamericano de Desarrollo y Meta Platforms a través de la iniciativa *Development Data Partnership*. Más específicamente, se trabajó con datos publicados en el portal de *Data for Good*. Este portal surgió como una necesidad de recolectar información en tiempo real sobre ciertos eventos que potencialmente puedan poner en peligro a parte de la población y que requieran fuentes de datos confiables, actualizadas y homogéneamente estructuradas. Después de un desastre natural, las organizaciones humanitarias y los *policy makers* precisan saber dónde se encuentran las personas afectadas y qué recursos se necesitan. Esta información es extremadamente difícil de capturar a través de métodos convencionales de recopilación de datos de forma rápida y eficiente.

Cuando las personas usan la aplicación de Facebook con los servicios de ubicación habilitados, los servidores reciben y guardan una aproximación de la latitud y de la longitud del dispositivo a intervalos regulares. La información de ubicación se usa de varias maneras, por ejemplo, para proporcionar dentro de la aplicación las funciones o el contenido más relevante para el usuario. Los *datasets* de Facebook Disaster Maps (incluidos en el portal de Facebook for Good) proporcionan información sobre dónde se encuentran las personas y cómo se mueven de forma agregada. Esto preserva la privacidad mediante algunos procesos como *privacy masking* y *random noise*, a la vez que mantiene la usabilidad y la interpretabilidad de los datos.

A medida que más personas se conectan a la aplicación de Facebook, los datos son más certeros y el volumen de información permite realizar procesos de análisis estadístico y geoespacial con más seguridad. La información se presenta en dos niveles: por mosaico o *tiles* (cuadrados geográficos) o por región administrativa (como los censos oficiales de población); y en tres definiciones horarias: 8 a. m., 4 p. m. y a medianoche (en huso horario del Tiempo Coordinado Universal [UTC]).

En todos los casos, para compilar la información, Facebook parte de *tiles* de aproximadamente 2 km por lado (cambian ligeramente entre ciudades debido a la proyección utilizada), detecta las coordenadas de las personas y las clasifica según el tile en el que han estado más tiempo en las ocho horas previas al momento de la hora mencionada en el *dataset*. Por ejemplo, si el conjunto de datos se refiere a las 8 a. m. y la persona estaba en su casa, Facebook clasificará a esa persona en el *tile* correspondiente y se lo asignará como un “origen”. Si la persona pasó la mayor parte de las siguientes ocho horas en otro cuadrado, esa persona aparecerá en los datos agregados del segundo *tile*, que se considerará como “destino”. Si aparece en el mismo tile, el origen y el destino será el mismo. A partir de esta información, se crea el *dataset* de *Movement between Tiles*. A continuación, se muestra una parte de estas tablas, en este caso, para la ciudad de Cali.

quadkey <chr>	length_km <dbl>	n_crisis <dbl>	m_baseline <dbl>	lat <dbl>	lon <dbl>
03223203300320 03223203300300	4.83	80	80	3.37 -76.5	3.41 -76.5
03223203300212 03223203300211	3.45	NA	NA	3.39 -76.0	3.41 -76.6
03223203300320 03223203300300	4.83	NA	NA	3.37 -76.5	3.41 -76.5
03223203300320 03223203300300	4.83	80	80	3.37 -76.5	3.41 -76.5
03223203300320 03223203300300	4.83	80	80	3.37 -76.5	3.41 -76.5
03223203300320 03223203300300	4.83	NA	NA	3.37 -76.5	3.41 -76.5
03223203300320 03223203300300	4.83	NA	NA	3.37 -76.5	3.41 -76.5



Fuente: Facebook Disaster Maps.

En esta versión resumida del *dataset* se encuentran variables geográficas como los *quadkeys* (identificación única para cada uno de los tiles), la distancia entre los centroides de cada cuadrado (*length_km*), el número de viajes correspondientes a la fecha seleccionada (*n_crisis*), el número de viajes considerando una media móvil de tres semanas (*n_baseline*), la ubicación de los centroides de origen y destino, y un *timestamp*. Cabe destacar que este *dataset* ya contiene el efecto del *privacy masking*, por lo que orígenes y destinos con escaso flujo de personas figurarán con *missing values* (NA) en las correspondientes columnas. Esto se debe a que Facebook se asegura de que no se puedan identificar personas o flujos individuales mediante la localización específica de los tiles con movimientos muy poco significativos.

Facebook parte de la información en grillas para elaborar el *dataset* de *Movement between Administrative Regions*. Sin embargo, en este caso, al no individualizar cada uno de los cuadrados de manera espacial, el efecto del *privacy mask* es menor. Es decir, para la elaboración del *dataset* dividido en Regiones Administrativas, Facebook no parte de los datasets públicos en formato tile, sino que parte de su propia información, eliminando el efecto de *privacy mask* para los tiles con pocas personas involucradas en los viajes. El efecto es menor ya que es mucho más probable un flujo de personas significativo entre dos regiones administrativas que entre dos *tiles*, ya que tienen un área considerablemente mayor y, por ende, mayor cobertura espacial.

La Tabla 2 es un ejemplo para la ciudad de Santiago de Chile de algunas de las columnas del *dataset Movement between Administrative Regions*.

[06] Para este *dataset*, la metodología para discriminar orígenes y destinos es la misma que para *Movement between tiles*, con la diferencia de que en la tabla cada polígono representa una región administrativa con sus límites oficiales provenientes del GADM (*Database of Global Administrative Areas*) en lugar de la grilla simétrica de Bing, con sus respectivos *códigos quadkey* para identificar los tiles. [06]

La metodología completa se encuentra disponible en <https://research.facebook.com/blog/2017/6/facebook-disaster-maps-methodology>.

polygon_name	trips	length_km	city
Santiago San Ramón	52	10.42	Santiago
Santiago Buin	47	32.49	Santiago
La Florida San Bernardo	1509	10.26	Santiago
San Joaquín Puente Alto	163	11.83	Santiago
San Joaquín Recoleta	90	10.00	Santiago
Santiago Vitacura	878	11.00	Santiago

T.
02

MUESTRA DEL
DATASET DE
MOVEMENT
BETWEEN
ADMINISTRATIVE
REGIONS.

Fuente: Facebook Disaster Maps

Las 17 ciudades analizadas seleccionadas en base a la disponibilidad de la información previamente descripta para el marco temporal deseado son las siguientes:



País	Ciudad
ARG	ROSARIO
BRA	BELO HORIZONTE - BRASILIA - FLORIANOPOLIS PORTO ALEGRE - RIO DE JANEIRO - SAO PAULO
CHI	SANTIAGO
COL	BARRANQUILLA - BOGOTÁ - MEDELLÍN - CALI
MEX	MÉXICO
PAN	PANAMÁ
PER	LIMA
URU	MONTEVIDEO

T.
03

CIUDADES
SELECCIONADAS
PARA EL
ESTUDIO.

Fuente: Elaboración propia.

Metodología de procesamiento

Una vez obtenidos los datos del portal Facebook for Good, se indagó sobre la conveniencia de utilizar todas las fechas y horarios disponibles para el desarrollo del estudio. La primera decisión que se tomó fue acotar el trabajo a cuatro semanas “normales”, es decir, las más cercanas a niveles de movilidad previos a la pandemia de COVID-19. Si bien los datos en el portal comienzan a mediados de marzo del 2020 (ya que en un primer momento se disponibilizaron para hacer análisis de políticas públicas referentes a la pandemia de COVID-19 y a las medidas de confinamiento), el trabajo se realizó sobre el tramo final de los datos (desde mitad de abril hasta mitad de mayo de 2022). Esto se debe a la necesidad de generar un estudio que no se viera afectado por las condiciones particulares que se vivieron durante los períodos de confinamiento y restricción de circulación, para obtener *insights* sobre un período de tiempo en el que la movilidad fuera relativamente similar con respecto a los niveles pre y post pandemia.

En segundo lugar, se filtró el rango horario de los datasets. Para tener una noción más cabal de los movimientos de las personas, se tomó el dataset de las 8 a. m. UTC, equivalente a las 5 a. m. para Argentina o las 3 a. m. para México, por ejemplo. Siguiendo la metodología adoptada por Facebook, se puede tomar el supuesto de que a la madrugada la mayoría de las personas duerme, y que en las siguientes 8 horas, están en otro punto geográfico realizando sus actividades diarias cotidianas. Esto permite identificar a las personas en su lugar de origen la mayor parte del tiempo entre medianoche y las 8 a. m. UTC (entre las 9 p. m. y las 5 a. m. para Argentina y las 7 p. m. y las 3 a. m. para México), y en su lugar de destino entre las 8 a. m. y las 4 p. m. UTC, (o entre las 5 a. m. y la 1 p. m. para Argentina y entre las 3 a. m. y las 11 a. m. para México).

Esta metodología se basa en la construcción de los supuestos del trabajo de Anapolsky y González (2021), los cuales se centran en la publicación de El Mahrsi et al. (2017), Clustering Smart Card Data for Urban Mobility Analysis. El concepto fundamental tomado de estos trabajos es el de *simetría diaria*, que describe que los movimientos de las personas se comportan como un espejo. Es decir, que las personas van hacia un destino determinado y luego retornan al punto de origen de manera simétrica.

Cabe destacar que en este estudio no se analizan modos de transporte, por lo que no es posible identificar recorridos de manera más precisa. La información brindada por Facebook no permite inferir cómo se movieron las personas, ya que está presentada de manera agregada entre pares origen y destino sin datos de velocidades ni tiempos de viaje (solo distancias medias).

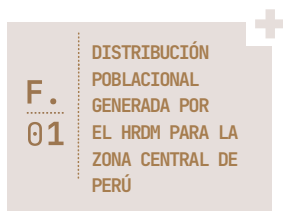
Avanzando con el procesamiento de la información, es para destacar el posible sesgo que pueden tener las bases de datos debido a la variabilidad en la distribución del uso de las aplicaciones de Facebook y/o el nivel de aceptación a activar la geolocalización en los dispositivos móviles a lo largo de cada una de las regiones administrativas a estudiar. Para intentar quitar este sesgo, se introdujo el uso de un tercer dataset proveniente de Facebook, el *High Resolution Density Map* (HRDM).

Los *datasets* de HRDM toman información oficial de los censos y proyecciones de población de cada país y ubican geoespacialmente a la población en función de su densidad y cantidad a través de mecanismos de *machine learning*. A esta información se le realiza una unión espacial con los polígonos de los municipios, y de esta forma, se obtiene la población de cada uno de ellos. Debido a que existe la posibilidad de clasificar la cantidad de personas en cada municipio según definiciones como “población general”, “mayores de 60 años” y “menores de 5 años”, se procede a restar de la población total esos extremos del rango etario, que no se corresponden con el universo de usuarios de Facebook.

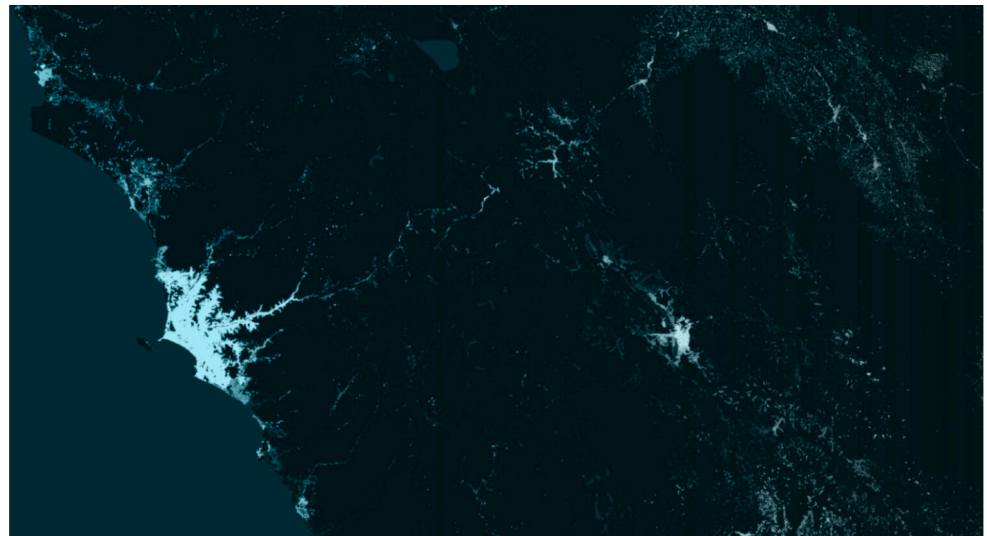
De esta forma, se obtienen los resultados de población de entre 5 y 60 años, con su correspondiente proyección oficial hasta el año 2020. A partir de esto, se busca eliminar el sesgo de uso tomando la suma de todos los viajes de cada uno de los orígenes y

comparándolos con la población correspondiente de cada región. De allí se extraen factores de expansión o pesos que aplicamos a los viajes de cada región administrativa para tener una noción más real de los flujos de personas entre municipios.

En el siguiente gráfico se encuentran georreferenciados cada uno de los puntos del *dataset* de HRDM, en este caso, para una de las zonas urbanas de Perú. Mientras más distanciados estén los puntos (y se generen zonas más oscuras), la densidad poblacional será indefectiblemente menor. En las zonas donde los puntos están más cercanos (gráficamente las partes más claras y definidas), la densidad poblacional es mayor. Debido al nivel de granularidad de los datos y la homogeneidad de la estructura del *dataset* a lo largo de las distintas ciudades, es más útil y práctico utilizar esta fuente de información geoespacial que obtener la misma información de cada uno de los institutos de estadística nacionales de cada país, sin necesidad de reparar en la existencia o no de la información procesada espacialmente.



Fuente: Data for Good at Meta (2019), Peru: High Resolution Density Maps.



Otra salvedad metodológica adoptada a lo largo del trabajo es que no se va a hablar de *viajes*, sino que los flujos que se determinan son personas que viajan. Esto tiene que ver con la ventana temporal utilizada y con la imposibilidad de segregar en etapas los movimientos de las personas. En otras palabras, es imposible detectar si la persona que viajó de un *tile* a otro lo hizo mediante un solo viaje, así como tampoco se puede identificar si a lo largo del día fue a más destinos que a este.

Como se mencionó anteriormente, los tiles son cuadrados de 2 km por lado. Esto genera una limitación en el trabajo: por ejemplo, los viajes menores a 2 km dentro del mismo tile van a figurar como “viajes cortos” con distancia media 0 (inicio y llegada en el mismo tile), mientras que otros viajes de menos de 2 km, pero con origen y destino en tiles vecinos y significativamente alejados del centroide del tile van a figurar de manera agregada dentro del grupo de viajes que poseen distancia media de 2 o 2,82 km, dependiendo de si el tile vecino está limítrofe a los lados o en diagonal, respectivamente.

Para evitar que este sesgo se propague cuando se analizan regiones administrativas y distancias medias, se decidió eliminar del análisis los viajes que tienen origen y destino en el mismo *tile*. Esto se materializó de la siguiente manera: se tomaron las personas que tienen como origen y como destino la misma región administrativa pero distinto *tile*. A este dato se le aplicaron dos factores de expansión: el primero, de acuerdo con la cantidad de personas que viajan dentro del límite administrativo que figura en el dataset de *Movement between Administrative Regions*; el segundo, según la población (como se explicó previamente). Este resultado se sumó a la cantidad de personas que viajan entre distintas regiones administrativas. Este flujo de trabajo puede representarse mediante el flujograma de la página 22.

Una vez consolidada la metodología del estudio, se obtiene una serie de indicadores necesarios para comprender patrones básicos de movilidad de cualquier ciudad. En primer lugar, se obtienen los flujos de personas entre cada una de las regiones administrativas. Esto posibilita conocer los principales focos de concentración y de movimiento en los grandes conglomerados urbanos, y permite definir las regiones con mayor nivel de generación de viajes y los nodos agregados de orígenes y destinos. Estos nodos se representan con matrices origen-destino, las cuales grafican la cantidad de personas que se mueven de un municipio “A” a un municipio “B” en base a los supuestos introducidos en el modelo.

El segundo indicador que se analiza en el presente trabajo es la distancia media. Este número refleja el promedio ponderado del recorrido de todos los viajes de cada región administrativa. Es decir, se calcula la media de las distancias recorridas por las personas teniendo en cuenta la cantidad de personas que viajan en cada distrito, como se ilustra en la tabla 4. [67]

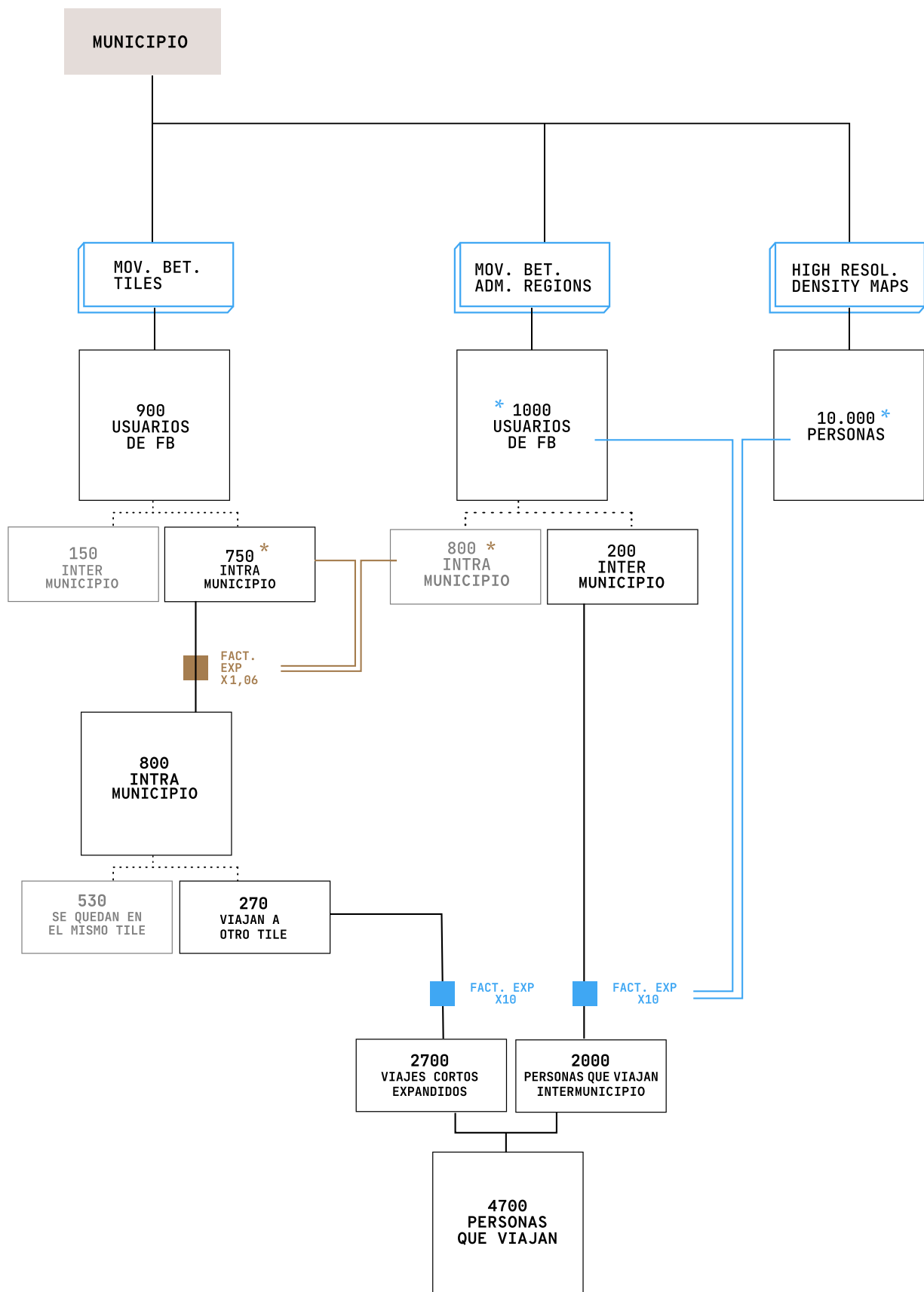
municipio origen	municipio destino	# de personas que viajan	distancia
A	A	400	0.1
A	B	100	10
A	C	300	5
A	D	200	9
Distancia Media Ponderada (km, D)			4,43

T.
04

EJEMPLO DE
TABLA DE
DISTANCIAS
MEDIAS

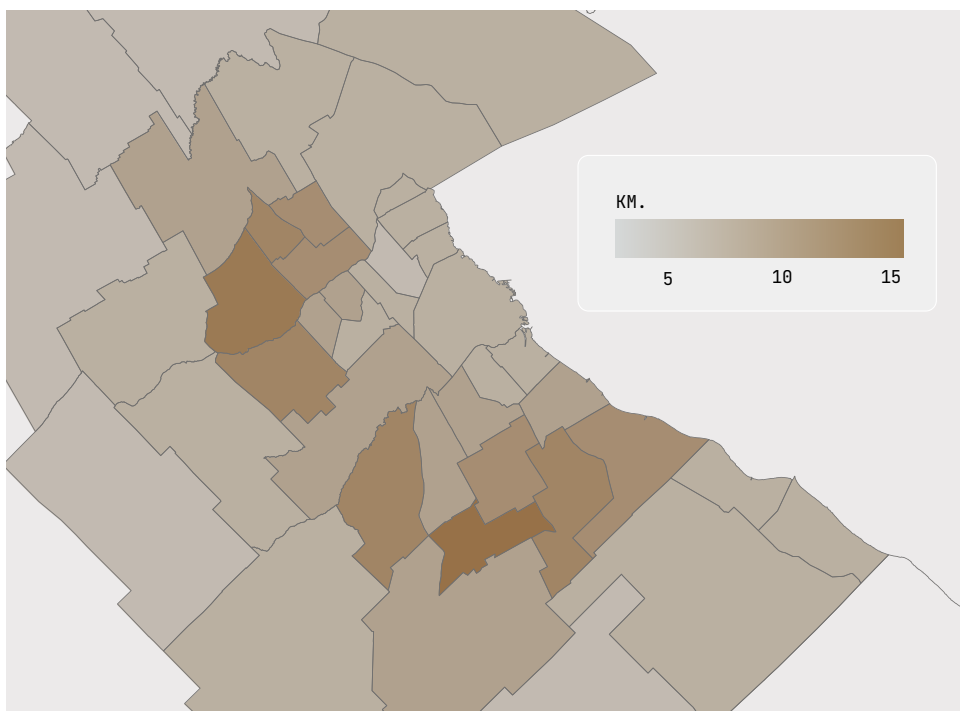
[67]

Para considerar solamente a aquellos usuarios que se desplazaron dentro del cálculo de la distancia media, se dejaron fuera del análisis a aquellos usuarios que tenían igual tile de origen y de destino.



Las distancias medias permiten tener una noción de las necesidades de viaje que poseen las personas en cada municipio. Las distancias medias más largas podrían responder a una lejanía relativa mayor con respecto al centro de actividad económica y mayor disponibilidad de medios de transporte, mientras que distancias medias menores muestran que las personas realizan sus tareas de trabajo, educación, salud o esparcimiento en zonas más cercanas a su hogar y que no necesitan o no pueden trasladarse a otras regiones.

A lo largo del trabajo, se relevaron distancias medias para las 17 ciudades mencionadas. Para ejemplificar, se toma un día hábil típico para la Región Metropolitana de Buenos Aires y se proyectan los datos como se observan en el dashboard creado en Shiny. En el anexo ubicado al final del trabajo se muestra la misma representación de datos para las otras 16 ciudades contempladas en el estudio.



Fuente: Elaboración propia en base a Facebook Disaster Maps.

En el gráfico anterior se observan las regiones administrativas de la RMBA con un color de relleno de cada polígono que representa las distancias medias recorridas por las personas que viajan y que tienen como lugar de origen cada uno de los municipios.

Limitaciones metodológicas

Resulta conveniente, antes de encarar el desarrollo del caso particular de la aplicación de la metodología para la RMBA, resumir las limitaciones metodológicas que se presentaron a lo largo del planteo de la manera de procesar y analizar los datos.

En primer lugar, se debe destacar la limitación horaria que se tuvo a la hora de procesar los datasets. Estas ventanas temporales predefinidas por Facebook obligaron a adoptar la hipótesis de la simetría de viajes descripta anteriormente, e impidieron medir los desplazamientos de personas que trabajan por la noche y descansan de día, o de aquellas con horarios laborales menos frecuentes. Tampoco es posible identificar viajes en cadena o intermedios debido a la agregación temporal y espacial de la información. Además, los horarios de los datasets están definidos en UTC, por lo que difieren según el huso horario de cada región metropolitana; esto dificulta una lectura homogénea de las conclusiones del estudio.

Con respecto a la toma primaria de datos, es para remarcar el hecho de que Facebook no puede identificar movimientos de usuarios que no tengan la aplicación instalada en el teléfono móvil ni los permisos de geolocalización activados, lo que puede generar distintos sesgos. Estos se intentan reducir mediante los factores de expansión generados en base a la población real de cada municipio.

Sobre el registro de la movilidad de las personas, una limitación del estudio es la grilla homogénea de 2 km por lado debido al filtro de privacidad aplicado. Para impedir la identificación de patrones individuales de movilidad, los orígenes y destinos con escaso nivel de flujo de personas son filtrados en el dataset de Movement between Tiles. En el dataset de Movement between Administrative regions el efecto es menor por el tamaño de los polígonos utilizados. Por otro lado, el tamaño de las grillas no permite visualizar desplazamientos que se realicen dentro del mismo tile, ya sean de escaso trayecto (100 m, por ejemplo) o viajes de hasta 2,83 km aproximadamente (la diagonal del cuadrado de las tiles).

Por último, con el acceso a los datos del HRDM, sólo se pudieron quitar del análisis a los menores de 5 años y a los mayores de 60 (únicas divisiones disponibles) como forma de aproximarse a la población activa. Sin embargo, este filtro puede estar obviando un sesgo, ya que cualquier mayor de 13 años puede registrarse en Facebook, y de este modo formar parte de la muestra.

2.1.1.1 ANÁLISIS DE LOS FACTORES DE EXPANSIÓN

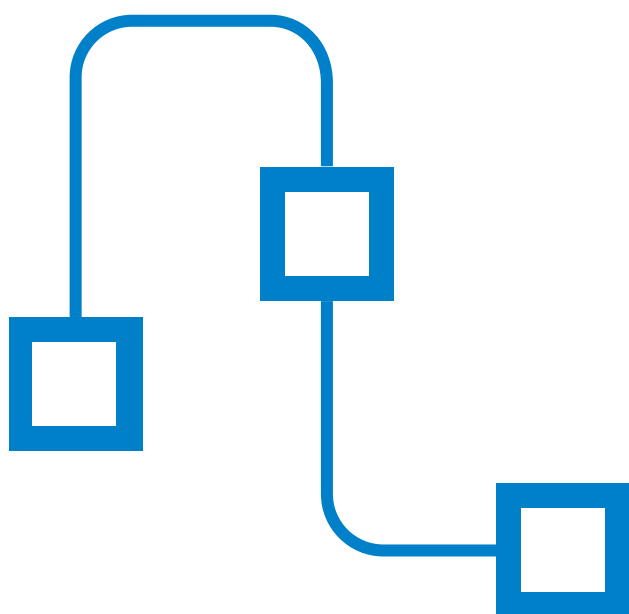
2.1.2 GENERACIÓN DE VIAJES

2.1.4 ORÍGENES Y DESTINOS DE LOS VIAJES

2.1.3 DISTANCIAS MEDIAS

2.1.5 COMPARACIÓN ENTRE LOS RESULTADOS OBTENIDOS Y OTRAS FUENTES DE INFORMACIÓN





APLI CACIÓN

PARTE 02



- Análisis del caso de la Región Metropolitana de Buenos Aires. Movimientos dentro y fuera de los distritos. Distribución de las distancias medias según zona de origen. Condiciones de posibilidad para los movimientos hacia centros económicos. Desde dónde y hacia dónde viajan las personas.

APLICACIÓN Y ANÁLISIS DE LA METODOLOGÍA EN BUENOS AIRES.

2.1.1 Análisis de los factores de expansión 30

Fig 4. Factores de expansión RMBA 30

2.1.2 Generación de viajes 31

Fig 5. Personas que viajan según municipio de destino (RMBA) 31

2.1.3 Distancias medias 32

Fig 6. División de zonas de la RMBA 32

Fig 7. Distribución de las distancias medias según zona 33

Fig 8. Proporción de distancias medias en días no hábiles sobre hábiles 34

Fig 9. Ferrocarriles urbanos de RMBA 36

Fig 10. Cajeros automáticos por metro² 36

Fig 11. Red de Buses en la RMBA 36

Fig 12. Densidad poblacional y punto de comparación 37

Fig 13. Distancias medias recorridas vs. distancia al centro 38

Fig 14. Distancias medias recorridas vs. NSE 39

Fig 15. Distancias medias recorridas vs. misma variable con lag espacial 40

Fig 16. Estadístico "Gi" local 41

2.1.4 Orígenes y destinos de los viajes 42

Fig 17. Matriz origen-destino 43

2.1.5 Comparación entre los resultados obtenidos y otras fuentes de información 44

Fig 18. Personas que viajan de PBA a CABA para trabajar según decil 44

Fig 19. Densidad de distancias medias (con datos de SUBE y datos de Facebook) 45

Fig 20. Principales flujos de personas en la RMBA 48

SECCIÓN 2.1

Aplicación y análisis de la metodología

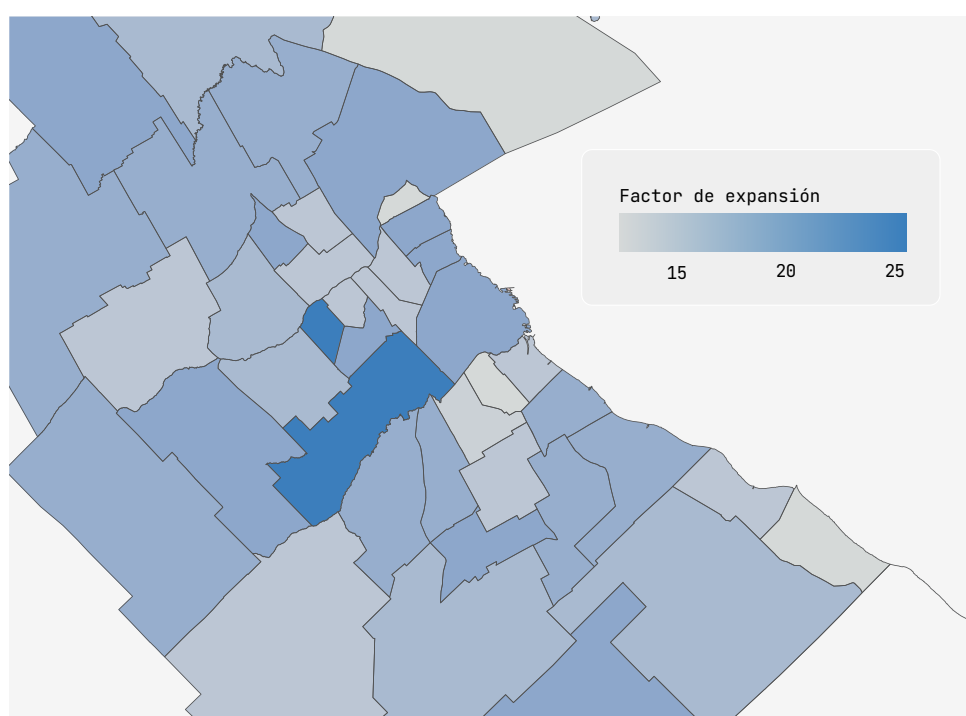
Luego de realizado el procesamiento de la información de acuerdo con la metodología previamente explicada, se procedió a analizar los resultados para la RMBA. Esta región comprende a la Ciudad Autónoma de Buenos Aires (CABA) y 40 municipios de la Provincia de Buenos Aires (PBA). Aquí habita aproximadamente el 35% de la población total de Argentina y se alcanza el 48% del Producto Bruto Interno (PBI) de este país, según cifras oficiales del Instituto Nacional de Estadísticas y Censos (INDEC).

La CABA forma parte del OMU desde su lanzamiento en 2007 y es una de las ciudades donde se han probado las nuevas herramientas para la actualización 2022. Además, con la colaboración del Departamento de Transporte de la Facultad de Ingeniería de la Universidad de Buenos Aires, se han podido recopilar y comparar diversas métricas relativas a la población y la movilidad, que se presentan más adelante.

Análisis de los factores de expansión

2.1.1

Para comenzar a estudiar la dinámica del transporte en esta región, es pertinente describir el desempeño de la muestra estadística de Facebook. A tal fin, se pueden descomponer los factores de expansión utilizados para calcular el total de personas que viajan en función de la población real de cada distrito. La media de los factores de expansión arroja un resultado de 16, y se distribuye geográficamente de la siguiente manera:



Fuente: Elaboración propia en base a Facebook Disaster Maps y High Resolution Density Maps.

De esta forma, el municipio con el factor de expansión más alto fue La Matanza, con 25; seguido por Ituzaingó, con 24,9; mientras que los distritos con menos necesidad de expansión fueron Lanús y San Fernando, ambos con 13. Este fenómeno puede darse por diversas causas, como, por ejemplo, diferencias en los usos de teléfonos celulares o de Facebook; o la cantidad de viajes filtrados por privacidad (cuando el número de personas que viajan a puntos lejanos queda oculto por el masking metodológico de privacidad proveniente de las bases de datos originales).

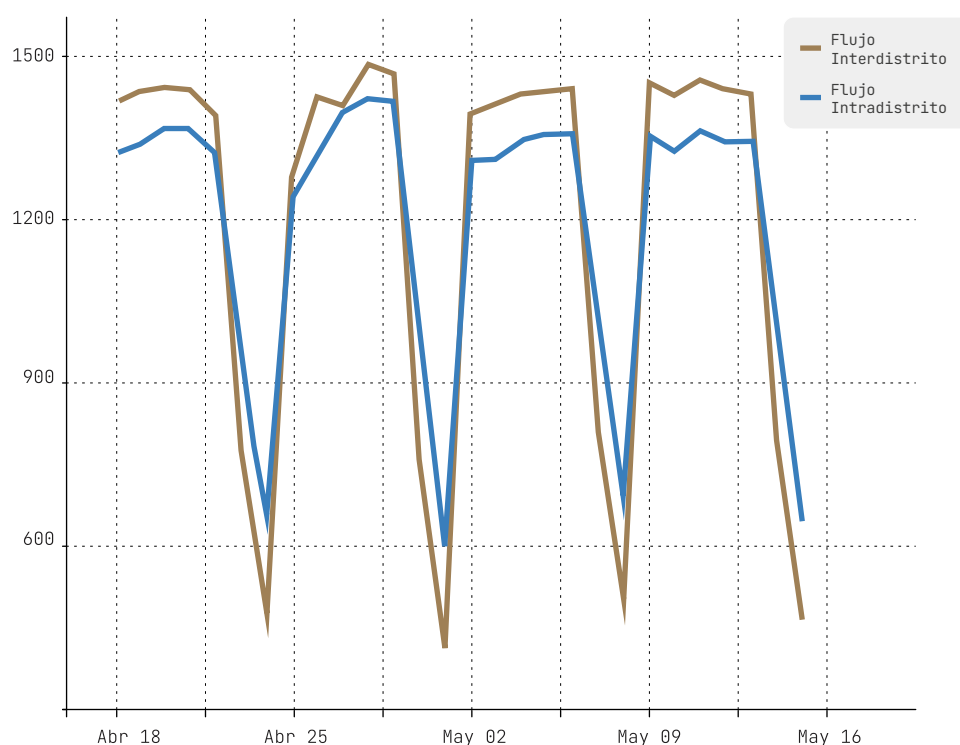
Generación de viajes

2.1.2

En el período analizado, la cantidad de personas que viajan^[88] en la región se calcula en torno a los 2,8 millones de personas diarias en día hábil, 1,7 millones los sábados y 1,2 los domingos, aproximadamente. Es interesante, en este punto, analizar en términos generales si los viajes se realizan mayoritariamente dentro del mismo municipio o si las personas en general viajan a otros. El resultado es el siguiente:

[88]

Considerando el período de la mañana como se detalló en la metodología y no teniendo en cuenta a las personas que permanecen en el mismo tile.



PERSONAS QUE VIAJAN SEGÚN MUNICIPIO DE DESTINO (RMBA) F. 05

Fuente: Elaboración propia en base a Facebook Disaster Maps y High Resolution Density Maps.

Mientras que en los días hábiles el flujo interdistrito es mayor al intradistrito, en los fines de semana se observa que la situación se revierte. Esto significa que, en los días no hábiles, las personas generalmente tienen una preferencia a realizar viajes dentro del mismo distrito en lugar de trasladarse hacia otro. La situación puede responder a una estructura de trabajo y de actividades educativas más alejadas de la vivienda, mientras que es posible que en cada distrito haya elementos de recreación y ocio suficientes como para no tener que recorrer más allá de los límites jurisdiccionales durante los fines de semana.

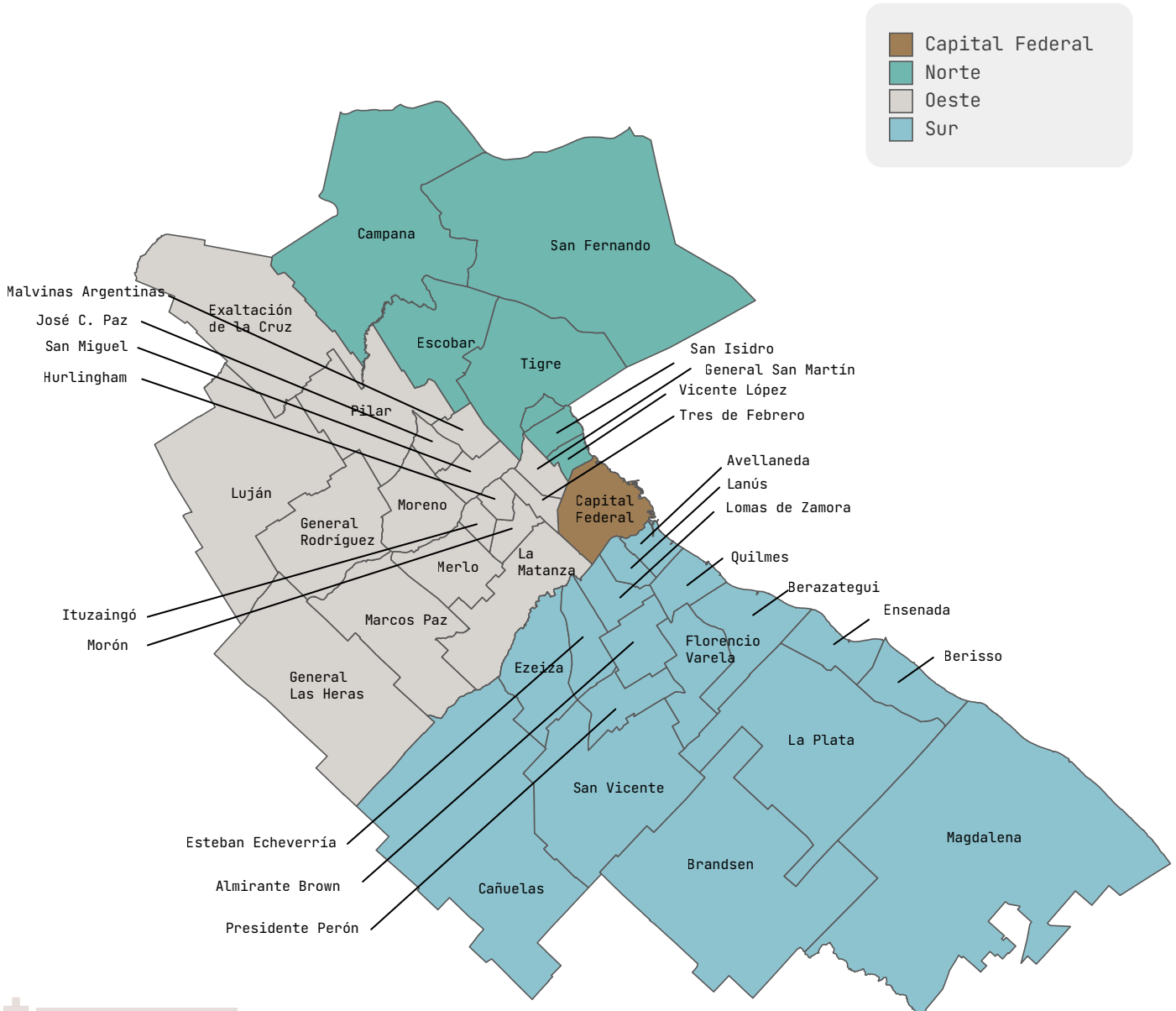
Distancias medias

2.1.3

[09]

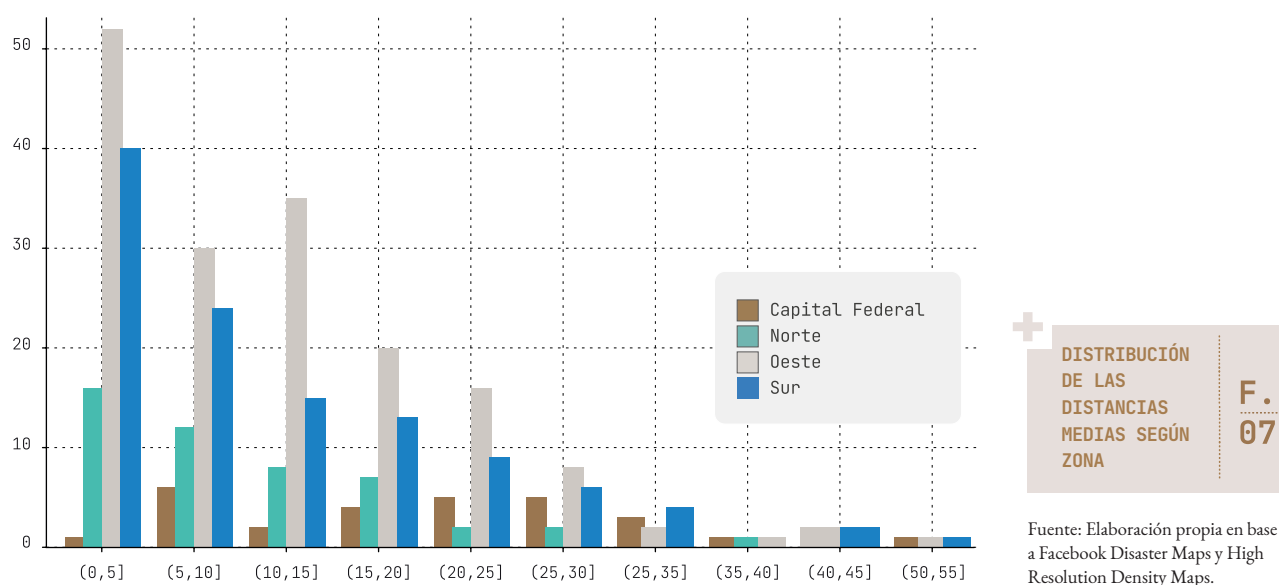
En el *dataset* la CABA aparece como "Capital Federal".

El fenómeno antes mencionado puede reforzarse en el análisis de las distancias medias por municipio. Para una mejor visualización, se dividió la región según la denominación más común de las zonas de la RMBA (CABA, [09] Zona Norte, Zona Oeste y Zona Sur).



Fuente:Elaboración propia en base a Gobierno de la Ciudad de Buenos Aires (GCBA) e Instituto Nacional de Estadísticas y Censos (INDEC)

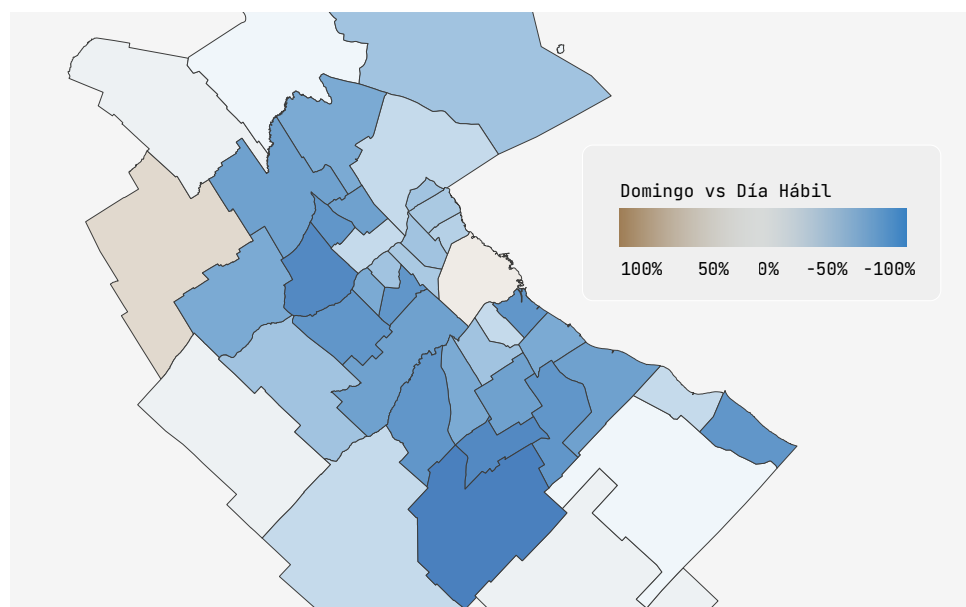
El procesamiento de datos permite obtener la distancia media de los viajes desde cada uno de los municipios hacia otro, y al mismo tiempo se puede obtener un histograma de las distribuciones de las densidades de ocurrencia de cada una de las distancias medias. Es decir, se incorporan a un histograma cada una de las distancias recorridas desde un municipio de origen hacia otro ponderadas por la cantidad de personas que realizaron ese tramo.



En el gráfico anterior se observan las distribuciones promedio según la zona. La CABA posee una distribución más achatada que las zonas del conurbano bonaerense, generando distancias medias levemente más uniformes. En cambio, en la región perteneciente a la PBA, las distancias medias están más concentradas en valores más bajos, pero con distribuciones más inestables a lo largo de la serie analizada. El análisis anterior corresponde al promedio de los días hábiles. Para entender mejor la dinámica de los patrones de movilidad de la región, se procede a calcular la diferencia que existe entre las distancias medias desarrolladas en días hábiles y las desarrolladas un domingo.

+ PROPORCIÓN DE DISTANCIAS MEDIAS EN DÍAS NO HÁBILES SOBRE HÁBILES. F. 08

Fuente: Elaboración propia en base a Facebook Disaster Maps y High Resolution Density Maps.



Del análisis anterior se desprende que en sólo dos distritos las distancias medias del domingo aumentan en proporción respecto a los días hábiles: Luján (+10%) y la CABA (+4%). Los dos municipios que presentan mayor caída de distancias los domingos son San Vicente (-80%) y Moreno (-72%). Lo más llamativo de la representación geoespacial de las distancias medias es el efecto “corona”, por el cual más allá de los límites de la CABA el promedio de distancias medias va aumentando progresivamente hasta cierto punto donde comienza a descender.

Para abordar este fenómeno, es necesario considerar la accesibilidad al transporte público y la distancia a los centros económicos de cada una de las regiones (esto es cierto no sólo para la RMBA, sino también para la mayor parte de las ciudades latinoamericanas comprendidas en el trabajo). En el siguiente gráfico se presentan los ramales de ferrocarril eléctrico y los ramales diésel que poseen la demanda más importante de la región. A partir de algunas estaciones terminales e intermedias de los ramales mostrados, parten líneas secundarias con poca oferta y muy escasa frecuencia.

Desde este punto de vista, se supone que las personas van a moverse a las zonas más densas en cuanto a actividad económica y empleo mientras exista una red de transporte público que posibilite cada uno de estos viajes de manera eficiente. Un proxy interesante para graficar la actividad económica es la cantidad de cajeros automáticos que cada distrito tiene ajustada por su área.

Este proxy tiene un comportamiento muy similar al mapa de las distancias medias por cada uno de los distritos. Más allá de la cuestión de la distribución poblacional en la región, un tema extremadamente relevante para el análisis es, como se mencionó antes, la disponibilidad de buenas alternativas de transporte público para acceder a puestos de trabajo, educación y esparcimiento.

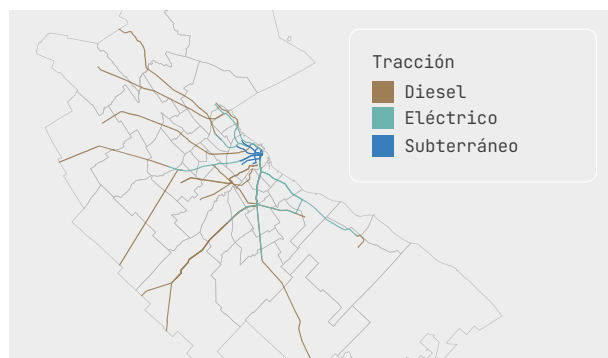
Si bien toda la RMBA tiene alcance con ferrocarriles y buses, la realidad es que a mayor distancia de los límites administrativos de la CABA, el sistema de transporte público se vuelve más desintegrado y menos eficiente. En primer lugar, los servicios de buses y ferrocarriles empiezan a distanciarse cada vez más entre sí debido a la escasa densidad poblacional de algunas de las zonas del conurbano bonaerense en comparación a la CABA, lo que limita en cierta medida la decisión de trasladarse en alguno de estos medios de transporte.

Otro factor no menos importante es el de la oferta de transporte. Si bien al graficarla en un mapa, la región parece sumamente conectada, en gran parte del área las frecuencias comienzan a condicionar negativamente el uso del transporte público.

La RMBA tiene una heterogeneidad muy marcada en relación a la calidad y la frecuencia del transporte público. Por ejemplo, en la Zona Oeste, los ramales eléctricos del ferrocarril llegan sólo hasta Moreno. A partir de Moreno, la zona se conecta con ramales diésel, pasando de una frecuencia de 6 trenes por hora, a menos de un tren por hora. La situación de los buses es análoga a la de los ferrocarriles. Mientras que en la CABA la red de buses cuenta con buena frecuencia y densidad de líneas, los recorridos que cubren los municipios de la provincia en general están más separados entre sí y tienen peores frecuencias.

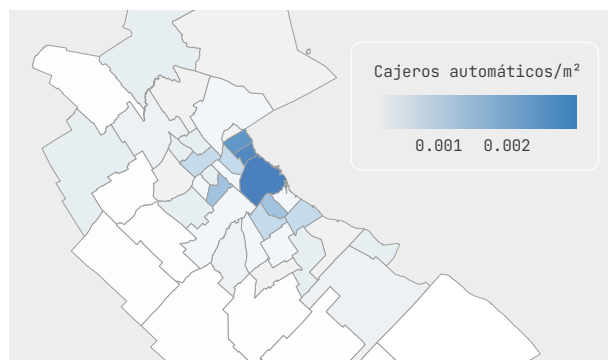
+ FERROCARRILES URBANOS CON MÁS OFERTA DE LA RMBA **F. 09**

Fuente: Elaboración propia en base a GCBA, INDEC y Open Street Maps (OSM).



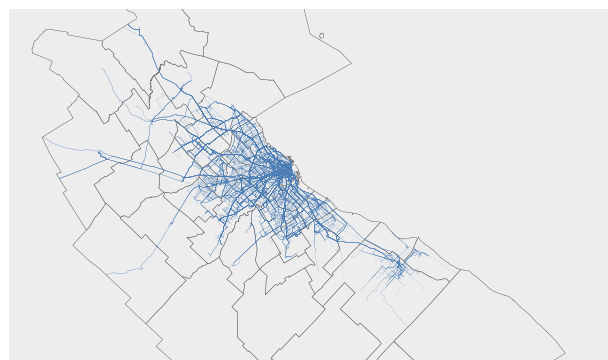
+ CANTIDAD DE CAJEROS AUTOMÁTICOS POR METRO CUADRADO **F. 10**

Fuente: Elaboración propia en base a OSM.



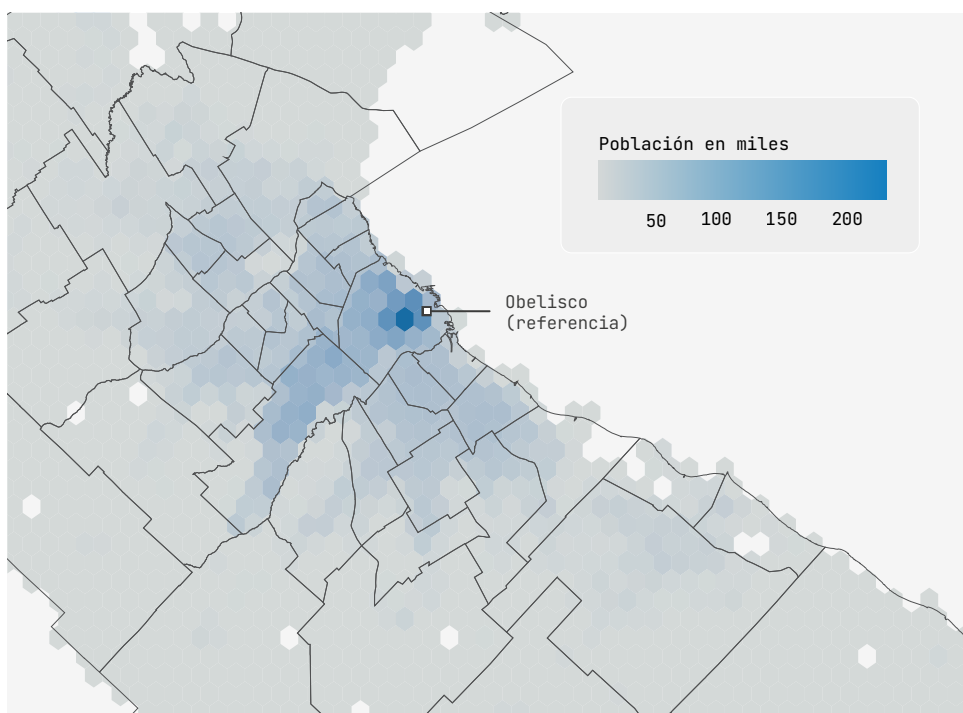
+ RED DE BUSES EN LA RMBA SEGÚN JURISDICCIÓN **F. 11**

Fuente: Elaboración propia en base a GCBA y OSM.



Para mostrar este comportamiento, se toman las distancias medias de cada uno de los distritos según si es el promedio de los días hábiles o si es un domingo, y se calcula la distancia con respecto a un centro relativamente arbitrario de la actividad económica (en este caso, se eligió la intersección donde se encuentra el Obelisco, Av. 9 de Julio y Av. Corrientes).

En la figura 12, se representa y contextualiza esta elección. En esta se observa la densidad poblacional para la RMBA y la ubicación exacta del punto elegido para la comparación de las distancias medias. Se evidencia una fuerte concentración de los hexágonos con mayor densidad poblacional dentro del entorno de la CABA, aunque desde la conformación y consolidación del Estado Nacional a finales del siglo XIX, la expansión de la mancha urbana ha ido más allá de los límites de la CABA. Esto ha impulsado el desarrollo de modos de transporte masivos que permitieron, cada vez con mayor facilidad, alcanzar el centro administrativo del país y acceder a los puestos de trabajo de otra ciudad de manera menos costosa, tanto en términos económicos como también físicos (tiempos de viaje). Sin embargo, a partir de una determinada distancia a la CABA, la oferta del transporte (sobre todo público) deja de encontrarse en los estándares del centro urbano.

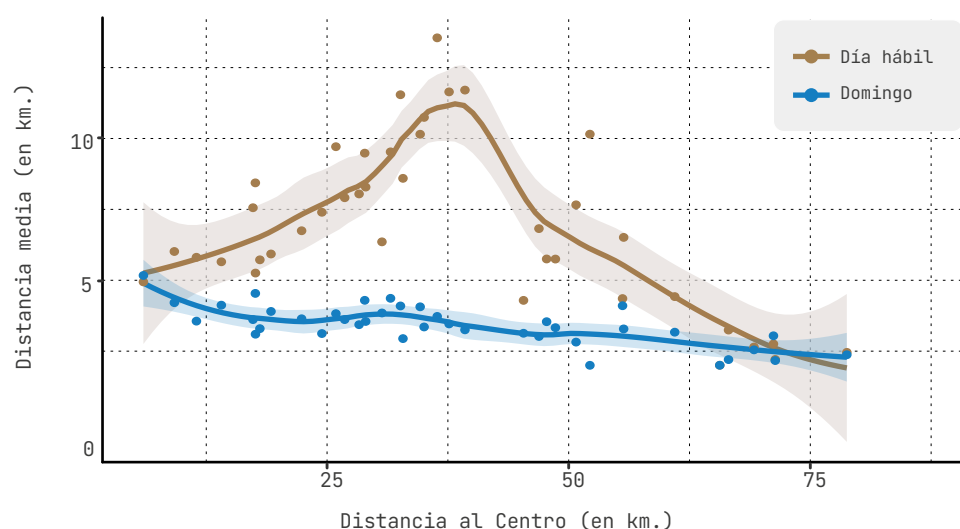


Fuente: Elaboración propia en base a GCBA, INDEC y OSM.

Volviendo a la cuestión fundamental de las distancias recorridas comparadas con la distancia al centro, una vez procesada la información se obtienen los siguientes resultados:

+ DISTANCIAS MEDIAS RECORRIDAS VS. DISTANCIA AL CENTRO F. 13

Fuente: Elaboración propia en base a Facebook Disaster Maps.

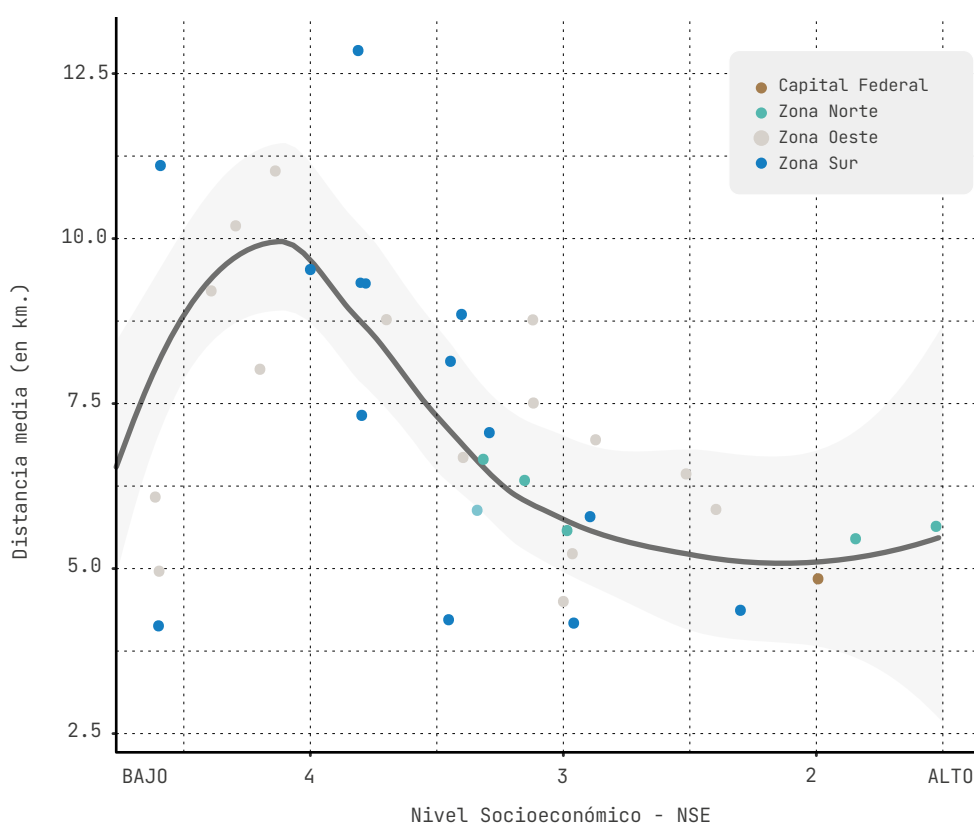


Otro aspecto a considerar respecto a las distancias medias observadas es el Nivel Socioeconómico (NSE) existente en las áreas geográficas que se mencionan. Este indicador sintetiza la medición de varios subindicadores. A los efectos de presentar relaciones entre las distancias medias y esta importante perspectiva de análisis, se utilizarán datos de Anapolsky y González (2022).

Dentro de las variables que se utilizan a la hora de elaborar el indicador de NSE, se encuentran la ocupación, el acceso a la educación y la disponibilidad de uso de los servicios públicos, todo estudiado desde el punto de vista geoespacial. Cabe aclarar que el máximo NSE es 1 (Alto) y el más bajo es 5 (Bajo). Agregando esta información a nivel distrito, se puede obtener la relación que se grafica en la figura 14.

La distribución observada es interesante ya que denota que, a partir de determinado nivel de ingreso, las distancias medias de cada uno de los municipios de la RMBA aumentan hasta el NSE ubicado en niveles “medios-bajos”. Mientras el NSE mejora a partir de este punto, la tendencia de las distancias medias observa un sostenido descenso hasta su estancamiento alrededor de los 5 km.

A partir de este punto, es conveniente analizar desde una perspectiva estadística la relación de las distancias medias con la distribución espacial. Para esto, se realizarán algunos tests geoestadísticos. En primer lugar, para preparar los datos, se toman los promedios de las distancias medias a lo largo del período para cada uno de los distritos de la zona. Luego, se computará el índice I de Moran.



+
 DISTANCIAS
 MEDIAS
 RECORRIDAS
 F.
 14

Fuente: Elaboración propia en base a Facebook Disaster Maps y Anapolsky y González (2022).

El índice I de Moran es un número que refleja el nivel de autocorrelación espacial que posee el fenómeno que se analiza. Así como en estadística de series de tiempo, los factores de autocorrelación representan el nivel de relación entre los valores presentes de una serie con los valores pasados, la autocorrelación espacial mide si existe algún tipo de agrupamiento entre valores geográficamente cercanos. En otras palabras, si en una serie de datos, estos están autocorrelacionados espacialmente, la ubicación geográfica es un determinante de los valores que se toman, sólo por el hecho de estar en una zona en particular.

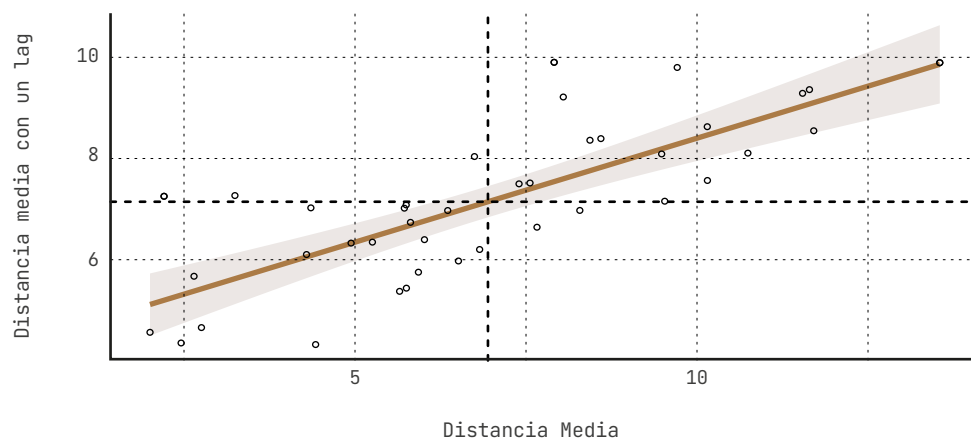
El indicador I puede tomar valores entre -1 y 1. Los valores más cercanos a -1 representan “repulsión” de los valores, es decir, que la serie no se clusteriza en zonas, sino que cada polígono o distrito se comporta opuesto a su vecino, algo asimilable a un esquema de un campo de ajedrez si fueran todos cuadrados. Los más cercanos a 1 evidencian un comportamiento de clusterización (los valores dependen de la zona donde están ubicados).

+

**DISTANCIAS
MEDIAS
RECORRIDAS VS.
MISMA VARIABLE
CON LAG
ESPACIAL**

F. 15

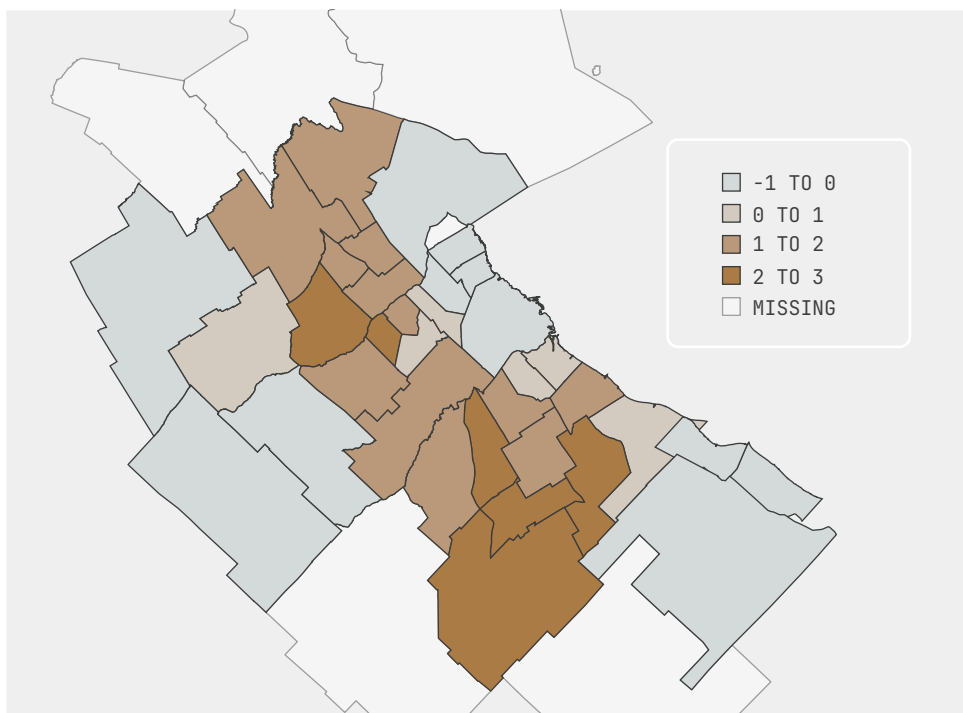
Fuente: Elaboración propia en base a Facebook Disaster Maps.



La hipótesis nula del test establece que el atributo que se analiza está distribuido en forma aleatoria entre los distritos o polígonos del área de estudio; es decir, los procesos espaciales que promueven el patrón de valores observado se constituyen de manera totalmente aleatoria. Con un valor p casi nulo, en este caso se puede rechazar la hipótesis nula y afirmar que hay evidencia estadística para demostrar la autocorrelación espacial entre las distancias medias, como se puede demostrar en el gráfico de dispersión donde se representan distancias medias con sus respectivos valores de los lags espaciales.

Además, se puede computar, de manera local, el estadístico G_i de Getis Ord. A diferencia del índice I de Moran, el estadístico G_i se calcula específicamente para cada observación dentro del contexto de sus polígonos vecinos. Se busca de esta manera determinar si existen patrones de atracción o de repulsión entre cada uno de los valores de la serie desde un punto de vista geoespacial.

El valor G_i puede ser negativo, positivo o nulo. Valores más alejados al 0 representan mayor clusterización (en valores altos si es positivo y en valores bajos si es negativo); y si el valor es nulo, no existe evidencia estadística para demostrar un patrón no aleatorio en el comportamiento de la serie. Del gráfico anterior se desprende una conclusión que ya había sido esbozada. Los municipios más cercanos a la CABA (sobre todo en el Oeste y en el Sur) tienen comportamientos muy similares en cuanto a las distancias medias, mientras que al alejarnos más de la CABA se pierde esa relación y comienzan a observarse efectos de “repulsión” de valores de distancias medias entre municipios vecinos (con valores G_i negativos).



+ ESTADÍSTICO
 GI LOCAL
 F.
16

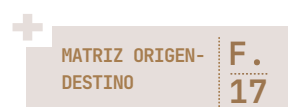
Fuente: Elaboración propia en base a Facebook Disaster Maps.

Orígenes y destinos de los viajes

2.1.4

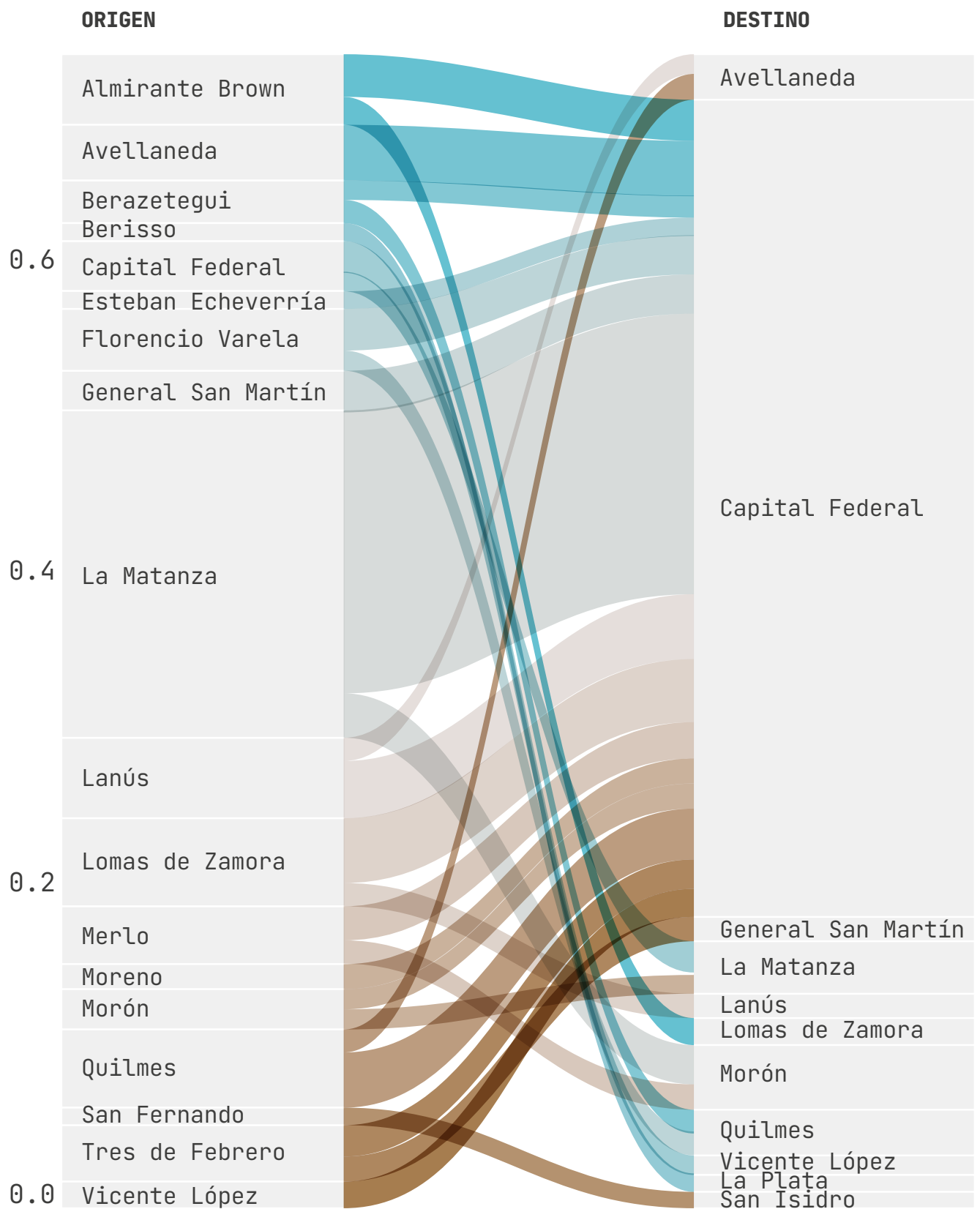
A partir de este punto, se analizarán los flujos de personas que viajan en la RMBA. En primer lugar, se construye la matriz origen-destino. Esta matriz representa la cantidad de personas que viaja de un municipio a otro. Para este caso, se utilizaron los viajes expandidos según la metodología explicada anteriormente. La cantidad total de personas que viajan en la región se estima en aproximadamente 2,8 millones. A fines de un mejor entendimiento del gráfico, se filtraron los orígenes y destinos con menos de 15.000 ocurrencias y se filtraron los flujos que comienzan y terminan en un mismo municipio.

La primera conclusión que se obtiene al observar la matriz origen-destino es que existe una alta concentración de destinos en la CABA, que recibe personas de casi todos los municipios de origen de la región metropolitana (sobre todo, de los municipios más cercanos). También se evidencian ciertos flujos interdistrito de los municipios más grandes del Conurbano bonaerense, sobre todo entre municipios vecinos (como Morón, que recibe personas de La Matanza y de Merlo).



Fuente: Elaboración propia en base a Facebook Disaster Maps.

PERSONAS QUE VIAJAN (EN MILLONES)



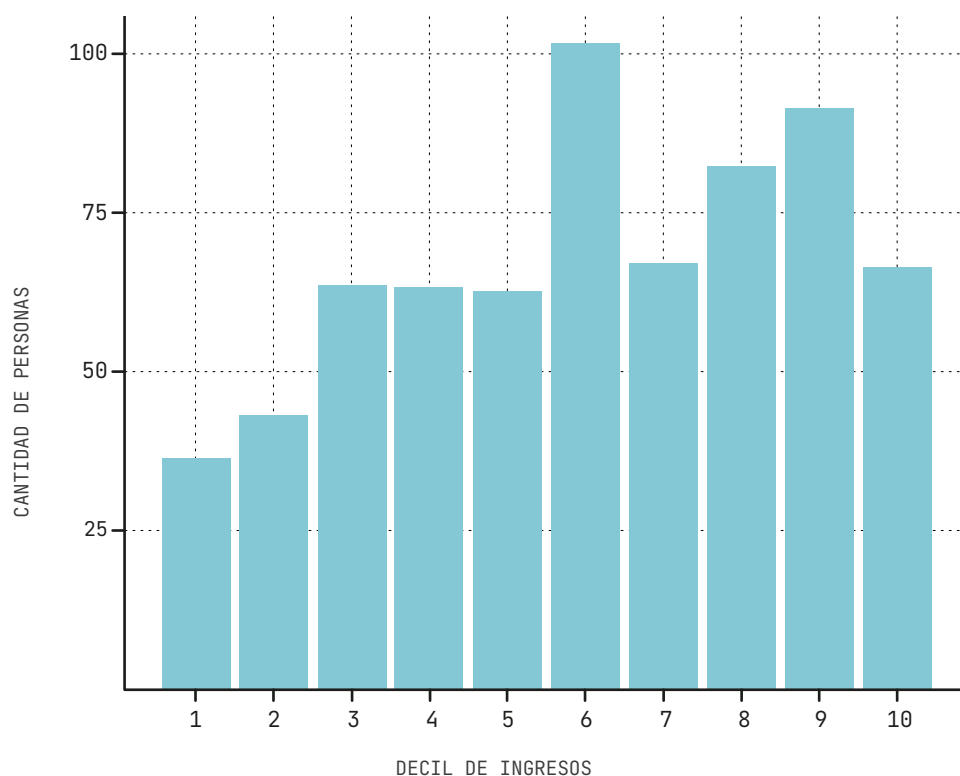
Comparación entre los resultados obtenidos y otras fuentes de información

2.1.5

La Encuesta Permanente de Hogares (EPH), elaborada trimestralmente por el INDEC, permite conocer de manera aproximada la caracterización de las personas que viven en la PBA y viajan a la CABA a trabajar todos los días. En total, según el INDEC al cuarto trimestre del año 2021, realizan este movimiento 819.938 personas, versus las aproximadamente 600.000 estimadas en este estudio. Cabe recordar en este punto que, como se mencionó en la metodología, se está estudiando una franja horaria determinada con el concepto de simetría diaria, por lo que existen personas que viajan en otros horarios no tan habituales ni mayoritarios que quedan fuera del análisis.

La concentración de personas que viajan es más alta en los deciles de mayores ingresos, mientras que las personas ubicadas en los deciles de menores ingresos viajan menos en proporción. Esto tiene que ver con el concepto de *inmovilidad relativa*. Es poco probable que una persona de bajos ingresos pueda costear el viaje hacia los grandes centros económicos si no vive en él y se encuentra lejos, ya que el costo del viaje aumenta mientras más alejada esté (ya sea por tarifas o por tiempo de viaje).

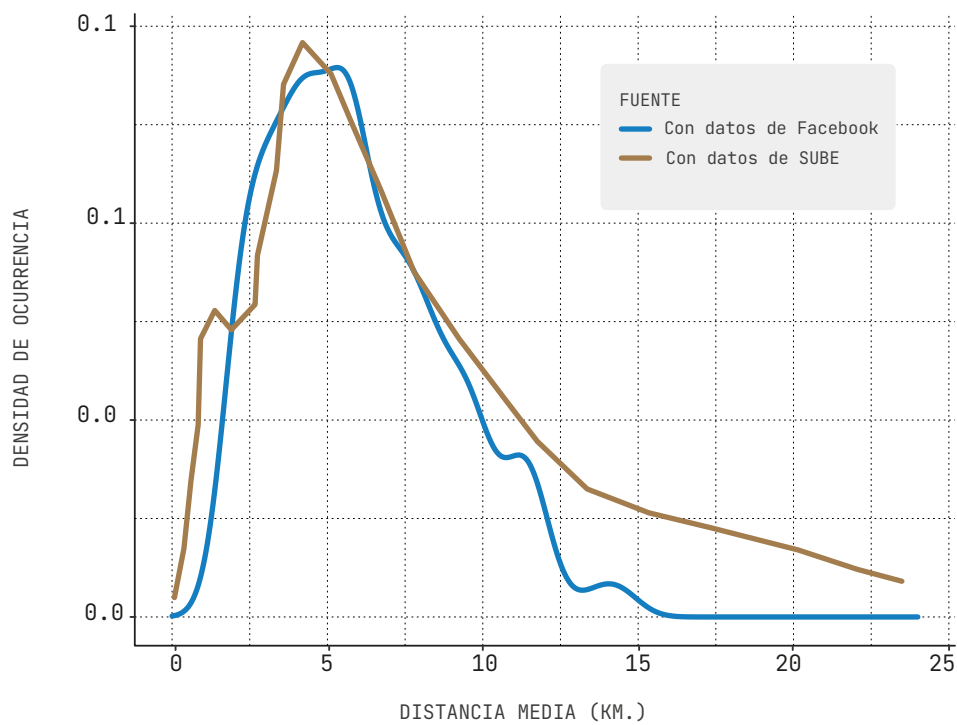
+ CANTIDAD DE PERSONAS QUE VIAJAN DE PBA A CABA PARA TRABAJAR SEGÚN DECIL DE INGRESOS F. 18



Fuente: Elaboración propia en base a INDEC.

Para continuar con la comparación de los datos obtenidos en el estudio con respecto a otras fuentes oficiales o trabajos elaborados, es necesario remarcar algunas cuestiones que se mencionaron en los aspectos metodológicos. Por ejemplo, está el problema del privacy mask que aplica la propia metodología de recolección de datos de Facebook desde el origen. Esto genera que mientras menos flujo de personas exista entre dos polígonos, mayor es la posibilidad de que esos movimientos se filtren en el proceso. La consecuencia directa de este mecanismo es la imposibilidad de estudiar los flujos pequeños, que suelen darse en las zonas más alejadas al centro.

Para ver cómo funciona este sesgo en la práctica, se propone comparar los resultados de este trabajo y los obtenidos por Anapolsky (2022) con la información del Sistema Único de Boleto Electrónico (SUBE).



+

DENSIDAD DE DISTANCIAS MEDIAS

F. 19

Fuente: Anapolsky y González (2022).

Con esta comparación, se evidencia que la distribución es muy similar en ambos casos hasta pasados los 12 km de distancia media aproximadamente. A partir de esa distancia, los viajes según los datos de Facebook caen bruscamente, y llegan a valores casi nulos en los 20 km, cuando las ocurrencias en los datos de SUBE se mantienen relativamente significativas. Se puede concluir, entonces, que los datos de Facebook predicen muy bien las distribuciones de las distancias medias hasta cierto rango o umbral.

Otro problema que surge a la hora de comparar con otras fuentes oficiales no tiene que ver con la metodología de los datos, sino con los distintos períodos que estos datos representan. Los datos de Facebook surgieron como una necesidad de poseer datos en tiempo real en las diversas crisis surgidas por la pandemia de COVID-19. Por esto, la información es presentada a partir de abril del año 2020. Sin embargo, la última Encuesta de Movilidad Domiciliaria (ENMODO) publicada de manera oficial para la RMBA data del año 2010, lo que significa que los datos son incomparables por muchísimos motivos. Aun sin contar el contexto pandémico, los patrones de movilidad en 12 años no se mantienen estáticos, sobre todo en una región tan dinámica como la de Buenos Aires y con todos los cambios en las formas de moverse que se han dado en los últimos tiempos (desde el aumento de la oferta del transporte público hasta esquemas tarifarios totalmente disímiles con respecto al ingreso). Esto se profundizó con los cambios en las decisiones modales y de demanda de transporte en la región por la pandemia. Parte de los trabajos y los métodos de estudio dejaron de ser presenciales de manera permanente. Esto generó un flujo de personas sensiblemente menor sobre todo en áreas céntricas de la CABA y en los grandes conglomerados del Conurbano e intensificó la presencia de personas en municipios más alejados del centro, gracias a la posibilidad de trabajar y/o estudiar de manera remota.

Un aspecto metodológico que es necesario remarcar a la hora de comparar datos entre los datasets de Facebook y la ENMODO es la diferencia entre viajes y personas que viajan. Mientras que en las encuestas de movilidad los datos se pueden establecer en viajes (con cada una de sus respectivas etapas), los datos de Facebook brindan la posibilidad (y la limitación al mismo tiempo) de analizar los flujos de las personas. Esto implica que con los datos de Facebook se puede analizar, de

manera agregada, el destino de una serie de personas agrupadas en cada polígono o región administrativa (con todos los supuestos del modelo), sin tener en cuenta la cantidad de viajes que la persona realiza (pueden ser más que el viaje de ida y el viaje de vuelta incluso).

A pesar de lo antes mencionado, algunas cuestiones centrales se pueden seguir comparando y analizando. Como se explicó en el desarrollo del trabajo y paralelamente se evidencia en la ENMODO, la CABA atrae gran parte de los viajes de la RMBA, sobre todo de lo que se considera el primer cordón del conurbano bonaerense; esta tendencia se reduce progresivamente más allá de esta zona. También hay viajes diarios entre el segundo cordón y la CABA, aunque de menor intensidad. En varios casos la cantidad de personas que viajan intradistrito se ve superada por este movimiento (Lomas de Zamora, Merlo, Lanús, Vicente López).

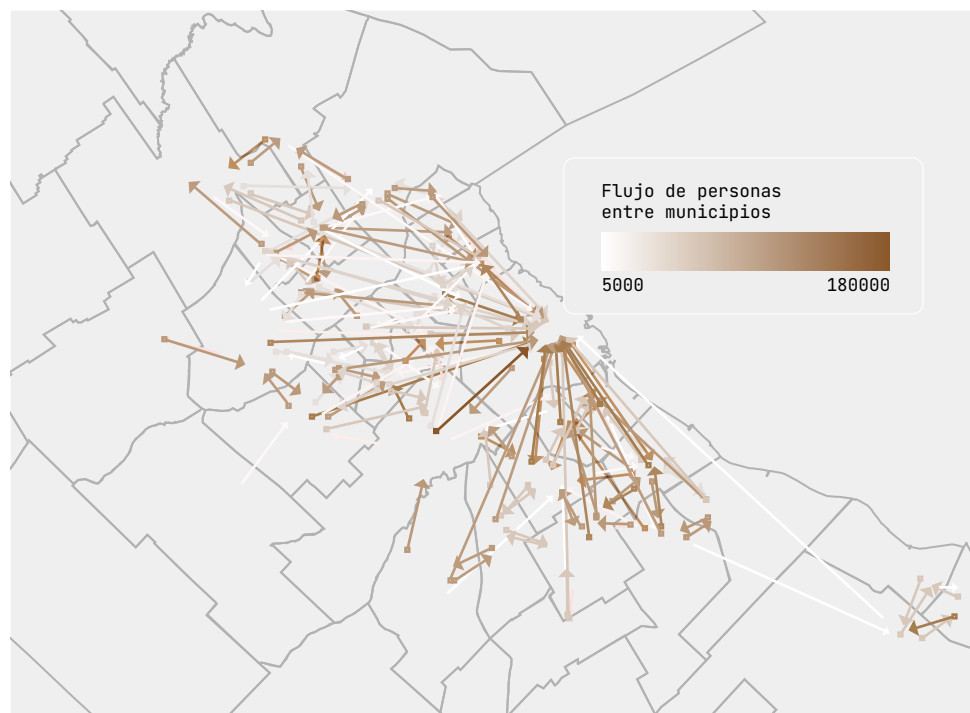
Con datos de la ENMODO y Facebook se pueden identificar flujos de menor magnitud pero que revelan un sistema de flujos por corredores, entre municipios vecinos, con sentido dominante periferia-centro, con dominancia de la estructura de las zonas cardinales (es decir, formando corredores en la zona norte, zona sur y zona oeste).

Para mostrar cómo se generan estos flujos, es conveniente visualizarlo mediante un mapa. En este caso, los flujos serán representados por flechas, las cuales poseen un atributo de opacidad según la cantidad de personas que viajan en ese par origen y destino. Las flechas más transparentes representan un menor flujo de personas, mientras que las flechas con un color más fuerte evidencian una mayor cantidad de personas que viajan entre esos municipios. Para una mejor comprensión visual y para evitar el overplotting, los flujos que representen menos de 1000 personas que viajan son eliminados.

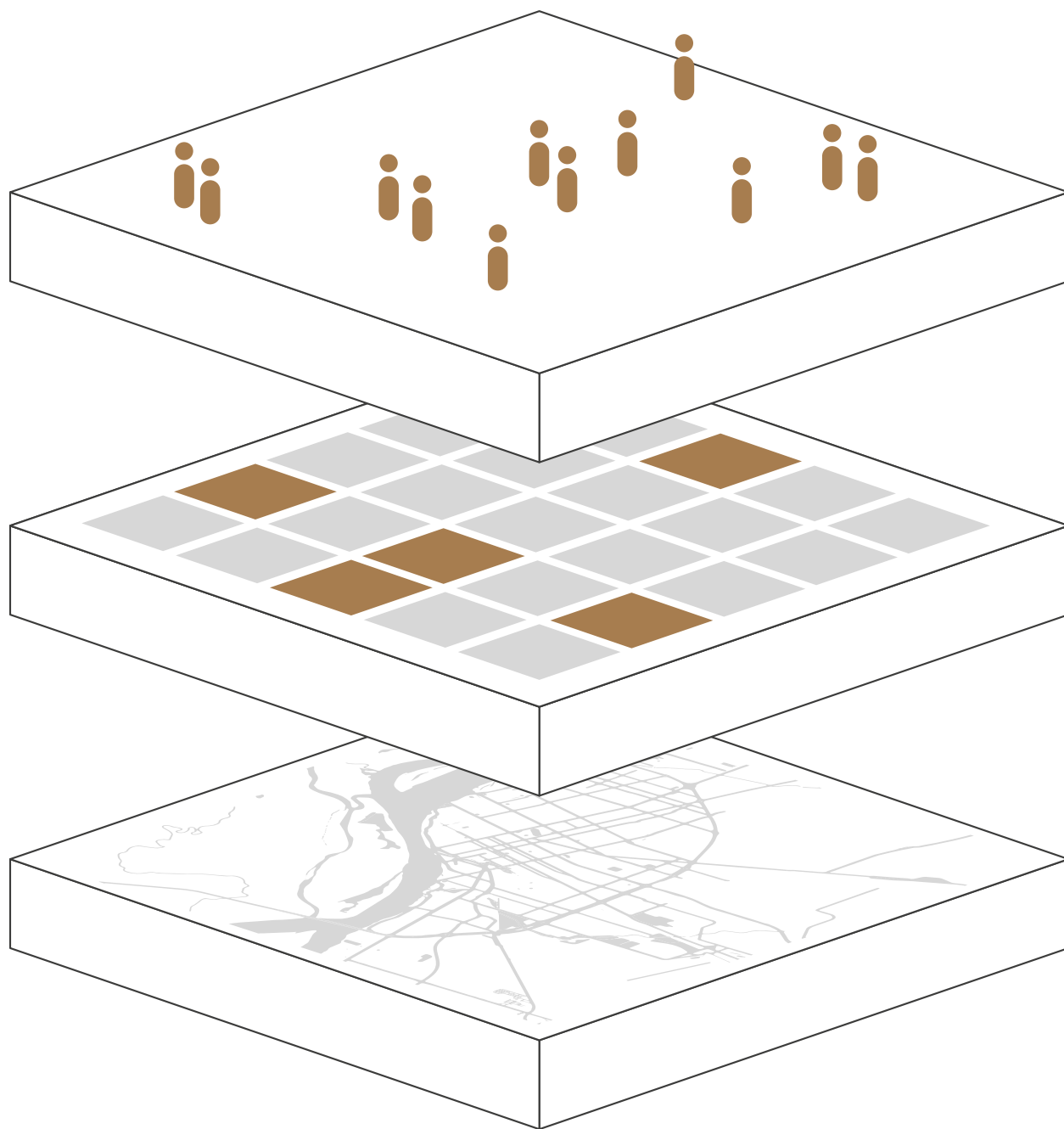
PRINCIPALES
FLUJOS DE
PERSONAS EN
LA RMBA

F.
20

Fuente: Elaboración propia en base a
Facebook Disaster Maps.



En este diagrama, los distritos del primer cordón juegan un papel atractivo en comparación con las regiones administrativas de segunda línea, lo que crea una serie de enlaces entre municipios en el corredor, hasta progresivamente poseer sólo flujos significativos dentro de los mismos municipios en los distritos más alejados de la CABA. Según los datos de Facebook, la cantidad de personas que viajan intradistrito representan aproximadamente el 54% del total, mientras que según la ENMODO este tipo de viajes está más cercano al 68%, y la cantidad de personas que ingresan a la CABA representan el 21% y el 19% del total respectivamente.



3.1

CONCLUSIONES Y PRÓXIMOS PASOS



- Potencialidad de la metodología propuesta: beneficios y próximos pasos.

P. SECCIÓN
.....
03

CONCLUSIONES Y PRÓXIMOS PASOS



Conclusiones y próximos pasos

En este trabajo, se demostró la gran potencialidad que tiene la metodología propuesta para la caracterización de los patrones de desplazamiento a partir de datos de la aplicación Facebook. La posibilidad de generar matrices origen-destino de los grandes desplazamientos diarios, estudiar cómo varía el número de viajes a lo largo del tiempo y/o conocer la distancia que recorren en promedio las personas para llegar a sus actividades principales son insumos de gran utilidad para la elaboración de políticas públicas. Contar con datos de aplicaciones móviles permite un ahorro significativo de costos y formas de diagnóstico de problemas mucho más ágiles, dinámicas y responsivas a los cambios de contexto.

Este trabajo representa un primer acercamiento a los datos, la metodología de análisis y la validación de los resultados. Quedan planteadas a futuro nuevas líneas de trabajo como la profundización de la comparación de los datos de Facebook con otras fuen-

tes de este tipo (registros de otras aplicaciones móviles o de antenas telefónicas) o datos oficiales tradicionales (encuestas, censos, registros públicos). También es necesario cruzar los datos obtenidos con otras variables geoespaciales que permitan explicar el fenómeno de la concentración de flujos de viajes y distancias, sobre todo con aspectos que tengan que ver con usos del suelo y equidad del espacio público. Sería interesante aplicar mecanismos de machine learning para clusterizar flujos a nivel polígono o incluso para generar alguna aproximación de probabilidades de particiones modales en base a las distancias medias y la distribución geográfica, y en base a estas conclusiones generar recomendaciones de políticas públicas al respecto.

Es muy importante continuar con el análisis de los resultados aprovechando la disponibilidad de fuentes tradicionales para validación en ciudades de mayor tamaño, como Buenos Aires. De esta forma, se podrá luego expandir su aplicación en ciudades de menor dimensión y con recursos más limitados, donde la utilización de metodologías tradicionales fue siempre un desafío.

Referencias

Anapolsky, S. y González, F. (2022). *Identificando la desigualdad en los patrones de movilidad en transporte público*. BID.

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

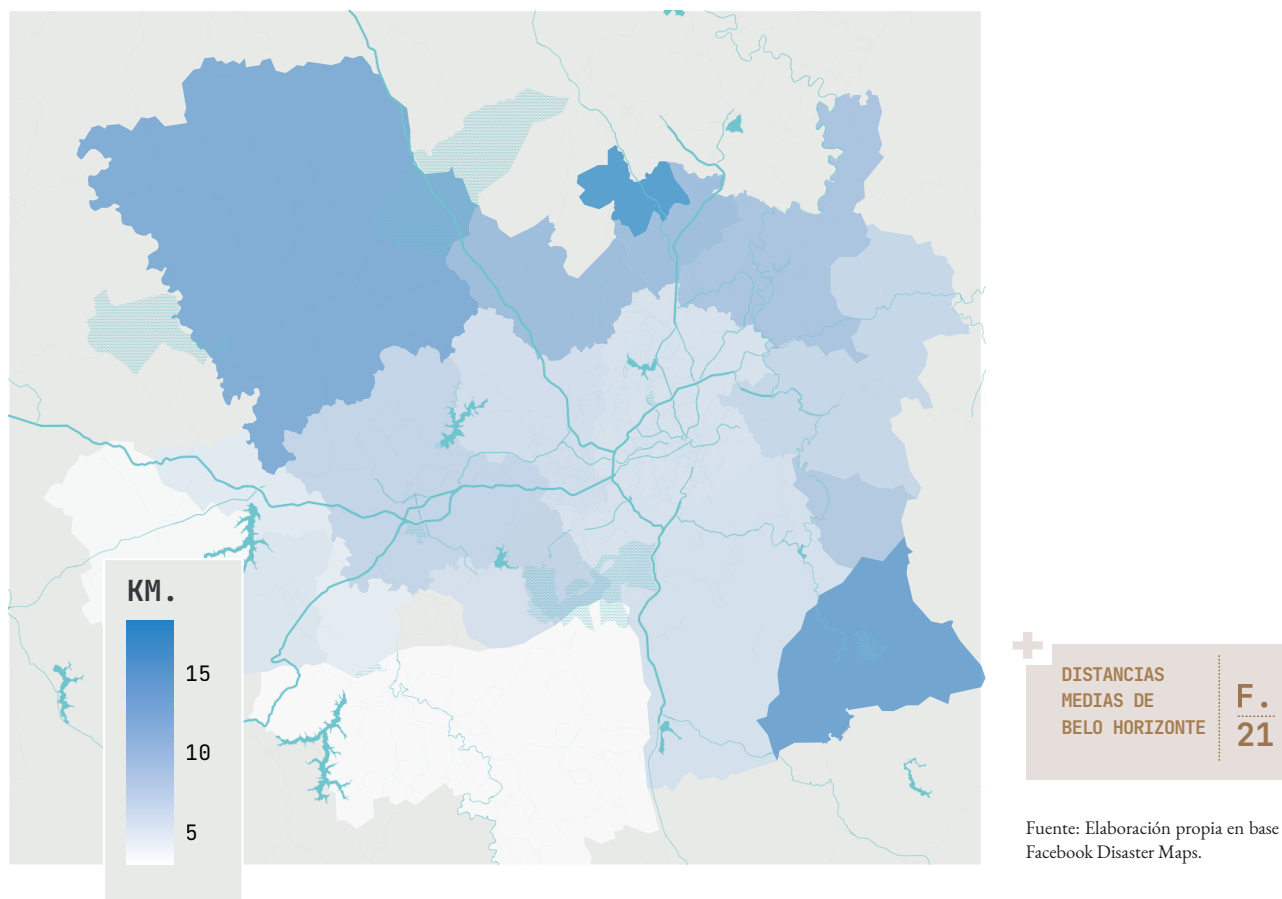
El Mahrsi, M. K.; Côme, E.; Oukhellou, L. y Verleysen, M. (2017). Clustering Smart Card Data for Urban Mobility Analysis. *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 712-728, doi: 10.1109/TITS.2016.2600515.

Gutierrez, A. (2021). *Atlas ENMODO - Encuesta Movilidad Domiciliaria. Cartografías de movilidad del AMBA 2009-2010*. El Guion Ediciones.

Lovelace, R.; Nowosad, J. y Münchow Jannes. (2020). *Geocomputation with R*. CRC Press.

Nusslé, S. (2005). *Spatial autocorrelation: A false enemy in data modelling?* University of Neuchâtel, Faculté des sciences économiques et sociales.

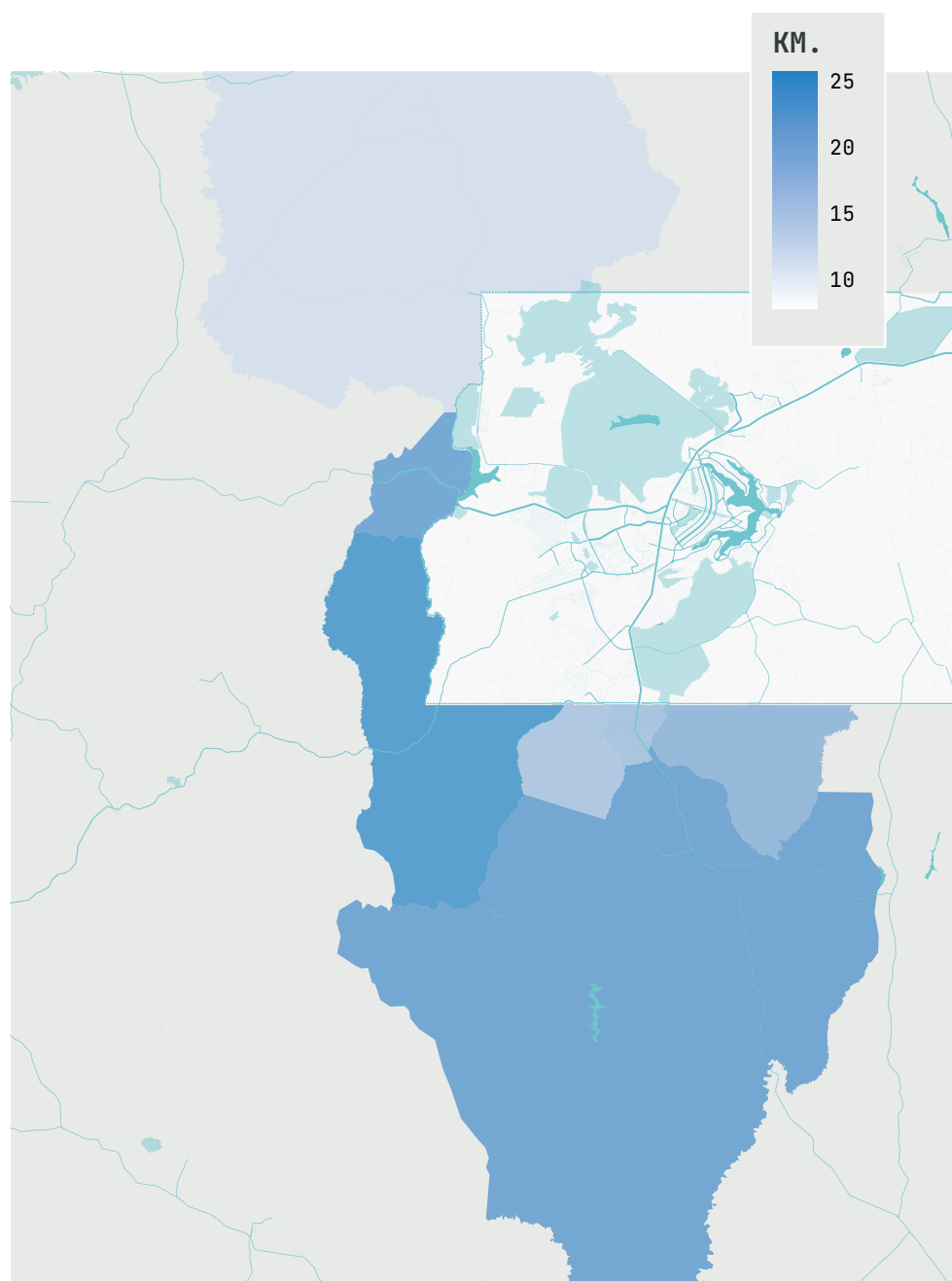
Anexo

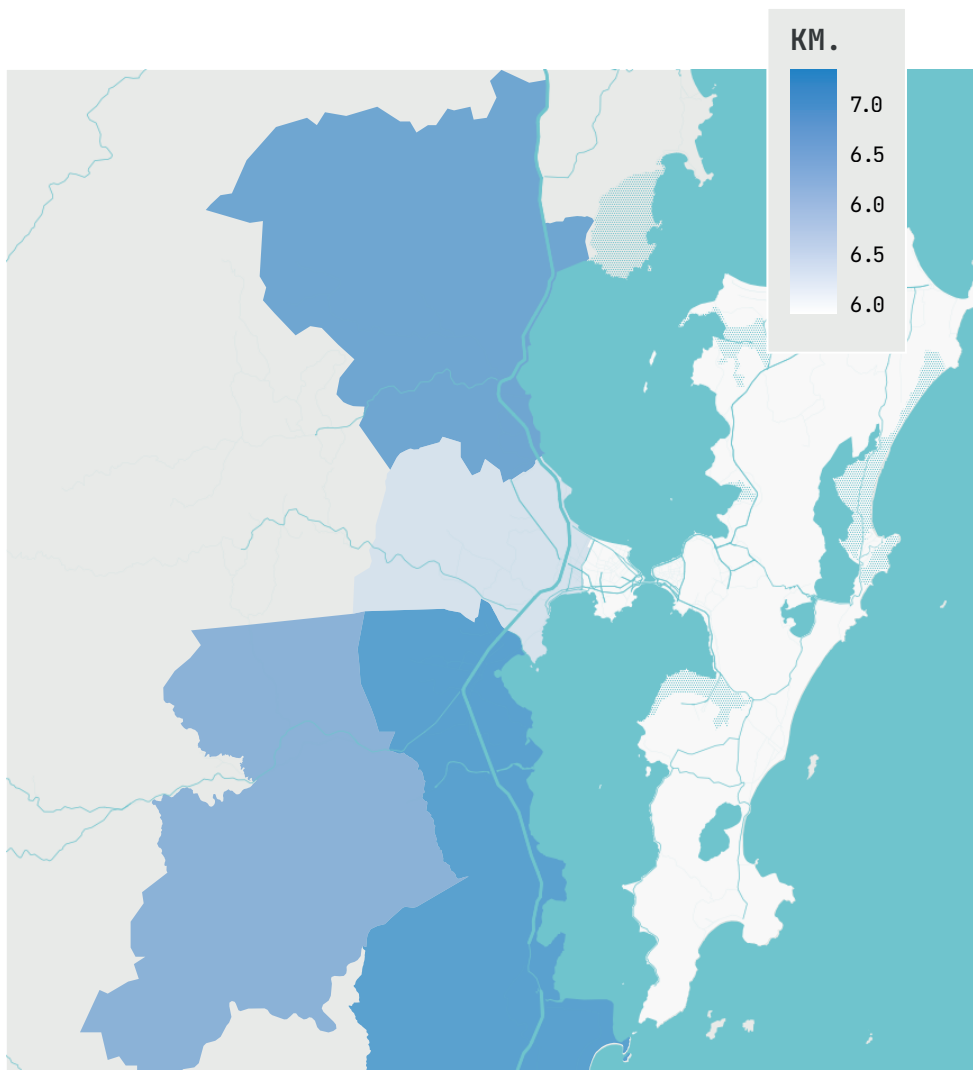


**DISTANCIAS
MEDIAS DE
BRASILIA**

**F.
22**

Fuente: Elaboración propia en base a
Facebook Disaster Maps.





+
DISTANCIAS
MEDIAS DE
FLORIANÓPOLIS

F.
23

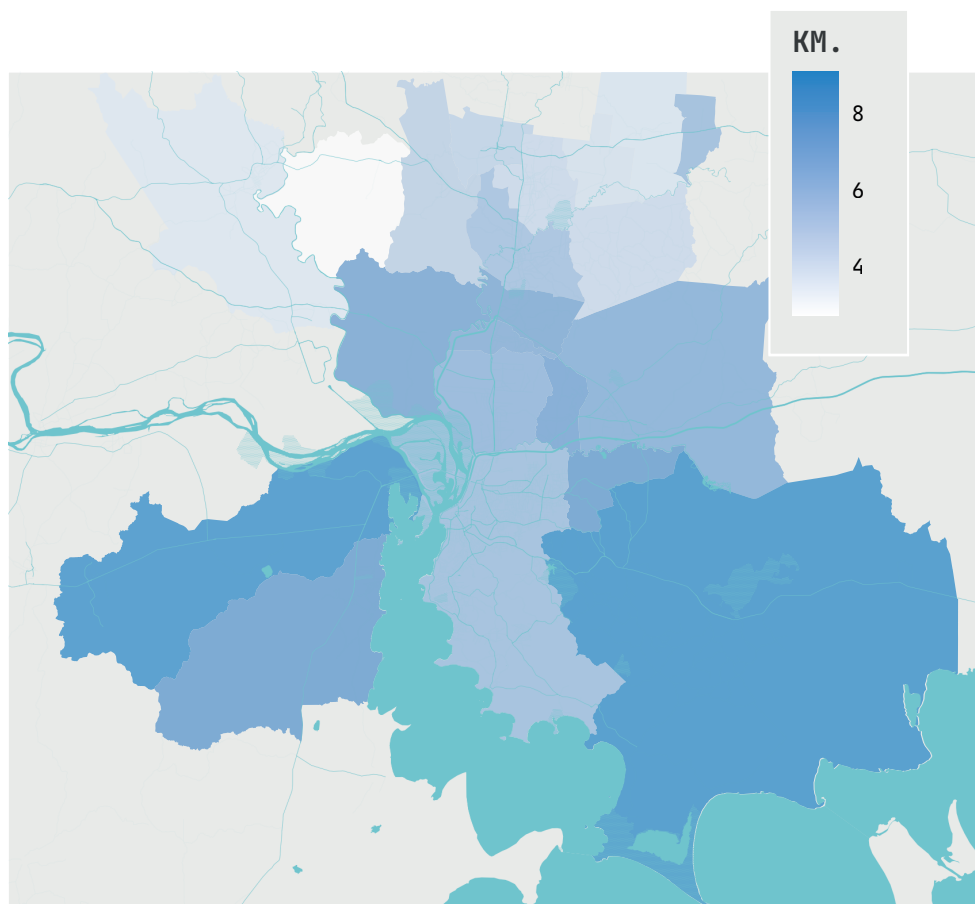
Fuente: Elaboración propia en base a
Facebook Disaster Maps.

+

DISTANCIAS
MEDIAS DE
PORTO ALEGRE

F.
24

Fuente: Elaboración propia en base a
Facebook Disaster Maps.

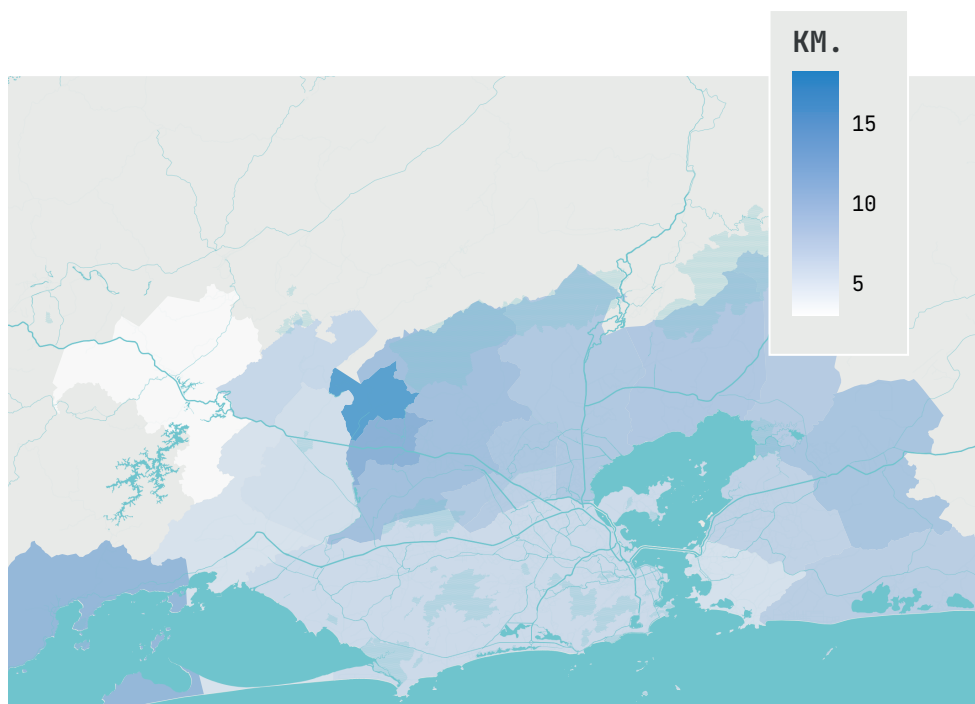


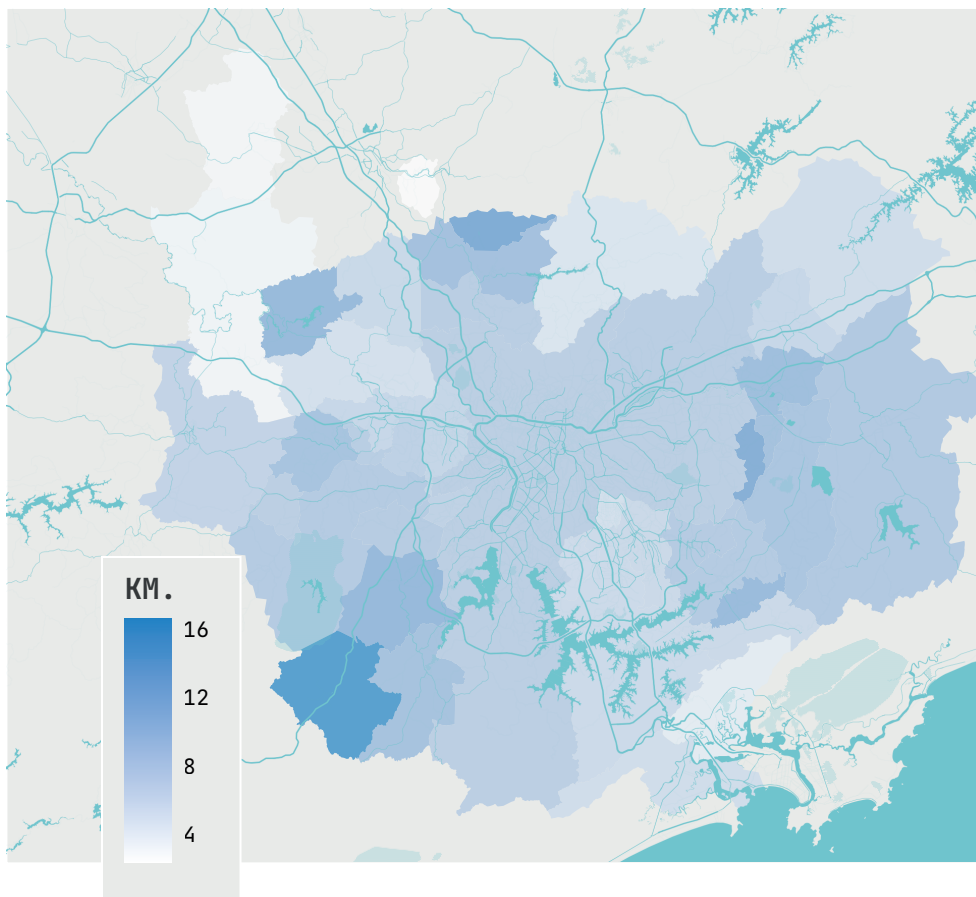
+

DISTANCIAS
MEDIAS DE
RÍO DE JANEIRO

F.
25

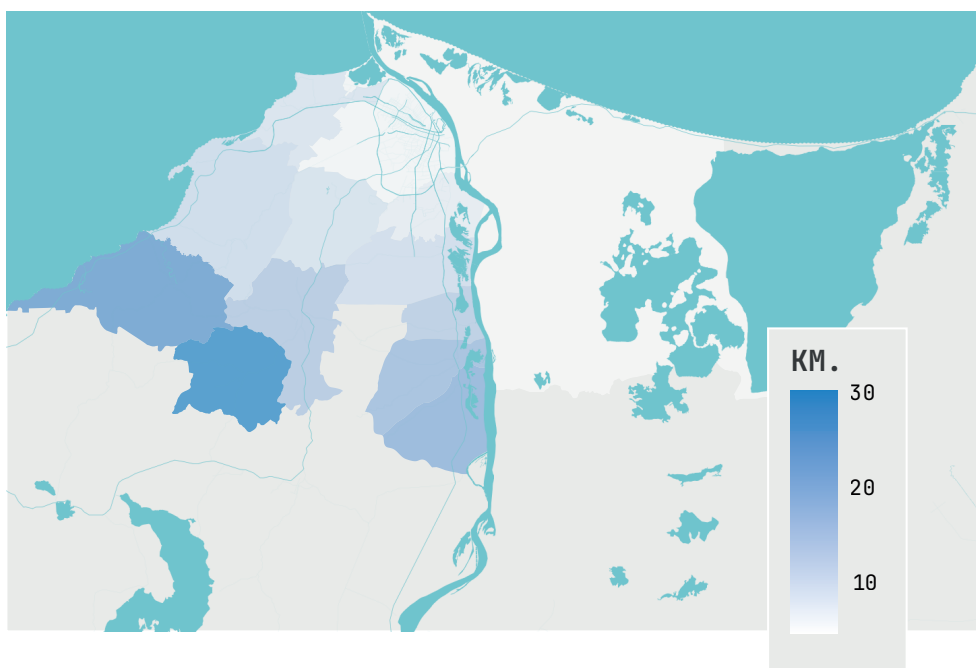
Fuente: Elaboración propia en base a
Facebook Disaster Maps.





DISTANCIAS MEDIAS DE SAN PABLO	F. 26
--------------------------------------	----------

Fuente: Elaboración propia en base a Facebook Disaster Maps.



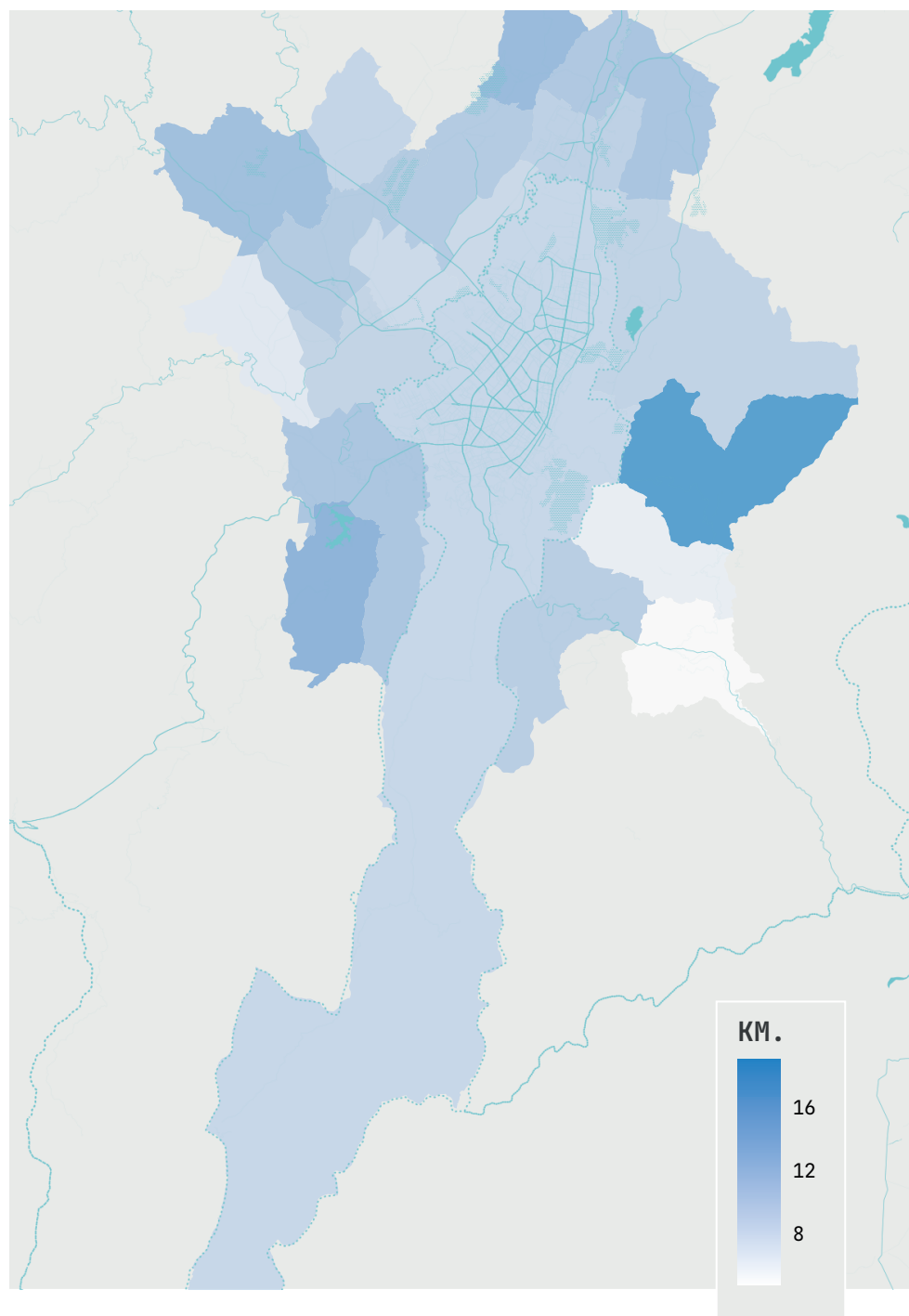
DISTANCIAS MEDIAS DE BARRANQUILLA	F. 27
---	----------

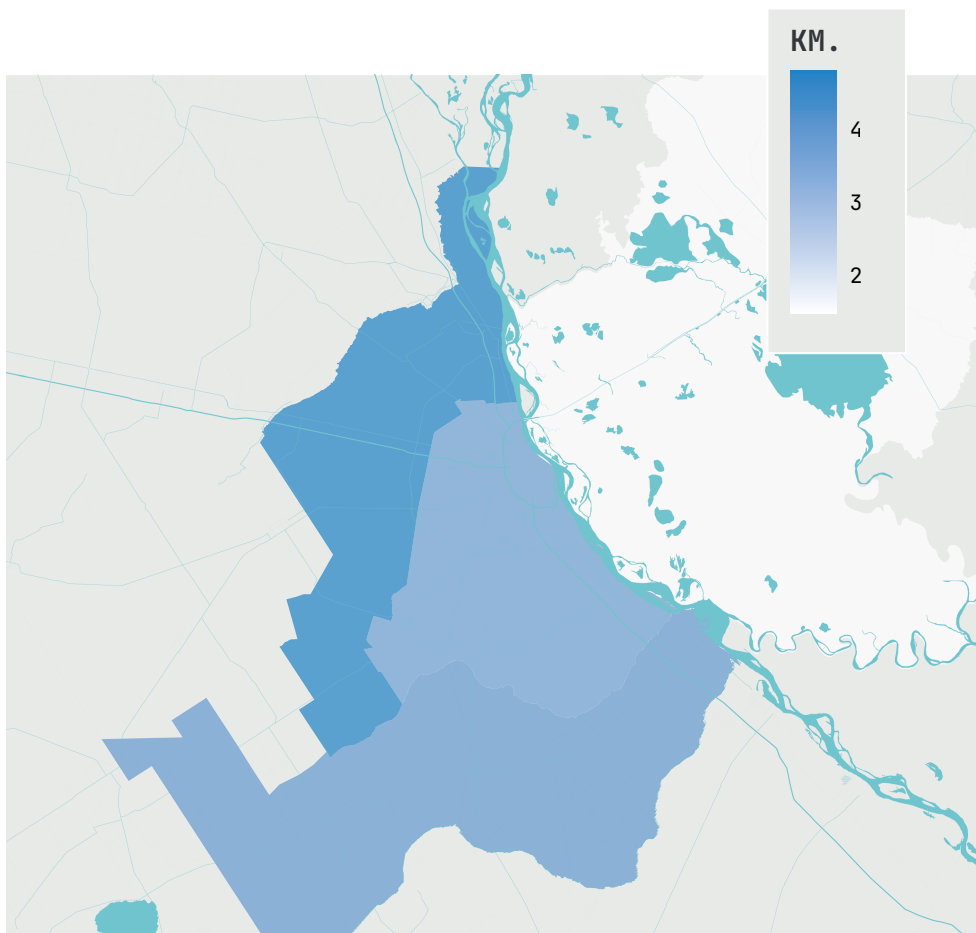
Fuente: Elaboración propia en base a Facebook Disaster Maps.


**DISTANCIAS
MEDIAS DE
BOGOTÁ**

F.
28

Fuente: Elaboración propia en base a Facebook Disaster Maps.





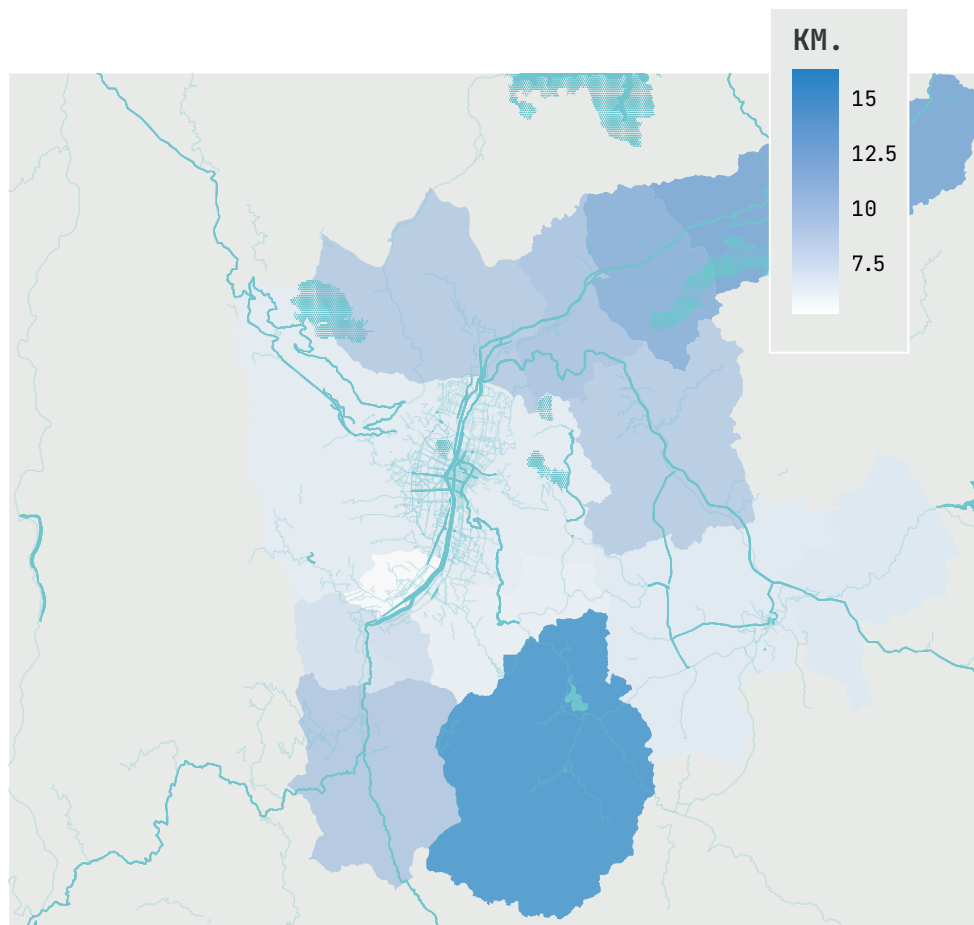
+
DISTANCIAS
MEDIAS DE
ROSARIO
F.
29

Fuente: Elaboración propia en base a
Facebook Disaster Maps.

**DISTANCIAS
MEDIAS DE
MEDELLÍN**

**F.
30**

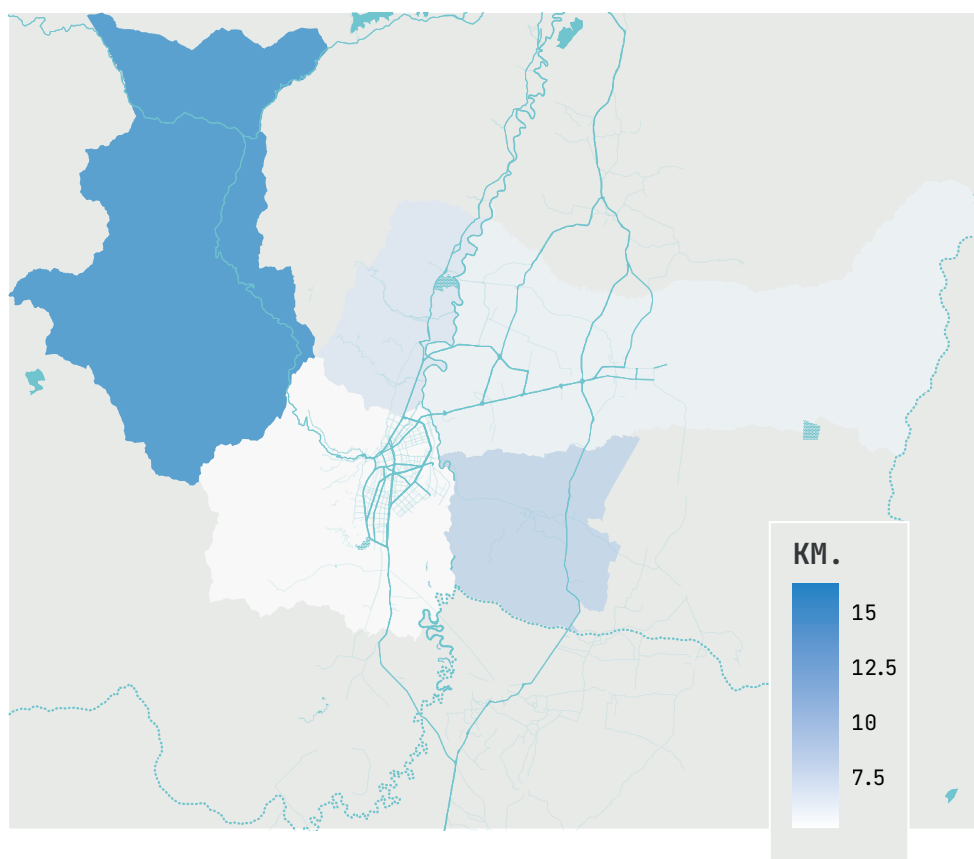
Fuente: Elaboración propia en base a Facebook Disaster Maps.

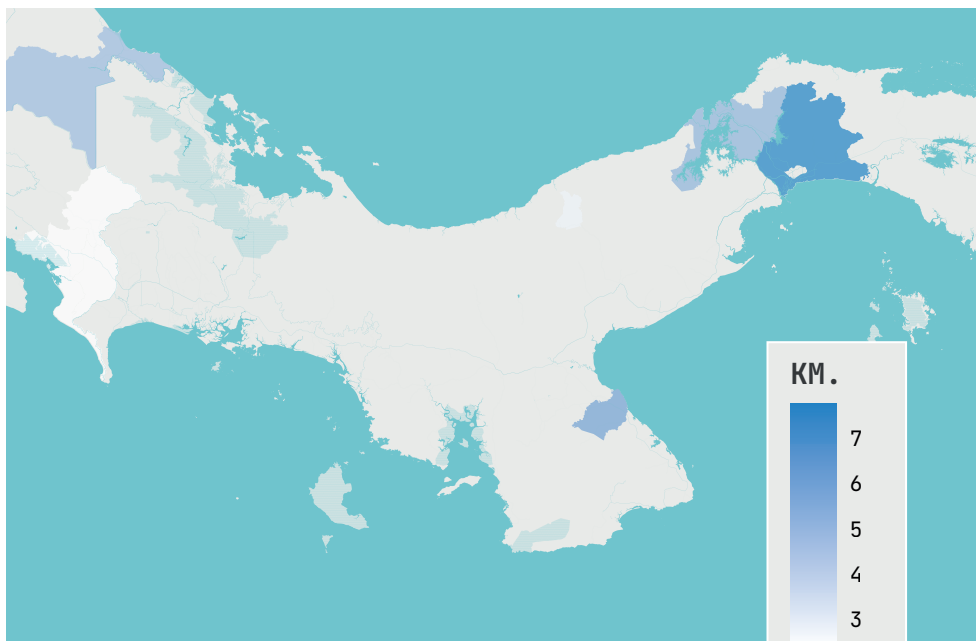


**DISTANCIAS
MEDIAS DE
CALI**

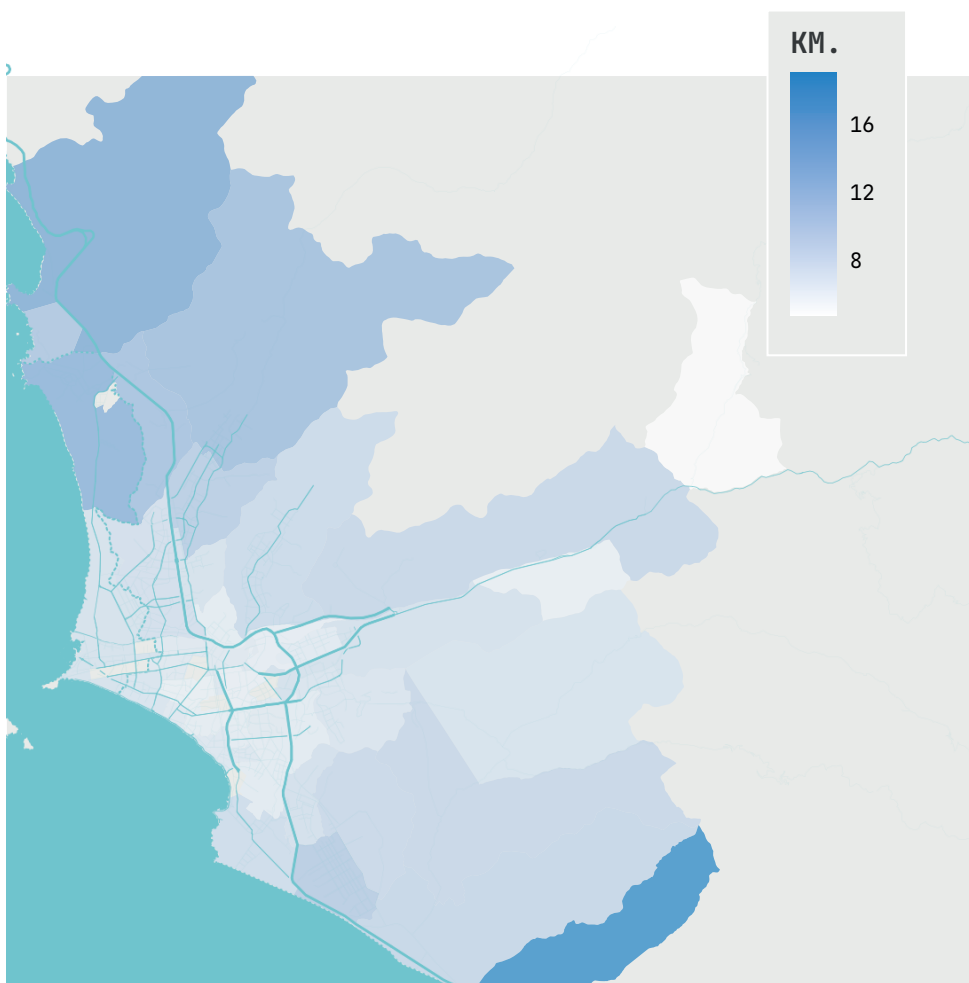
**F.
31**

Fuente: Elaboración propia en base a Facebook Disaster Maps.





Fuente: Elaboración propia en base a Facebook Disaster Maps.

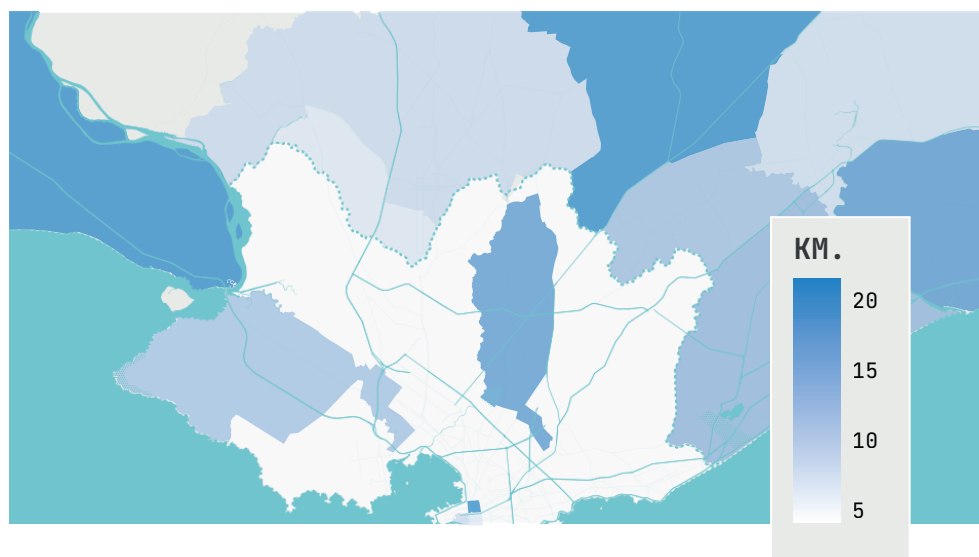


Fuente: Elaboración propia en base a Facebook Disaster Maps.

**DISTANCIAS
MEDIAS DE
MONTEVIDEO**

34

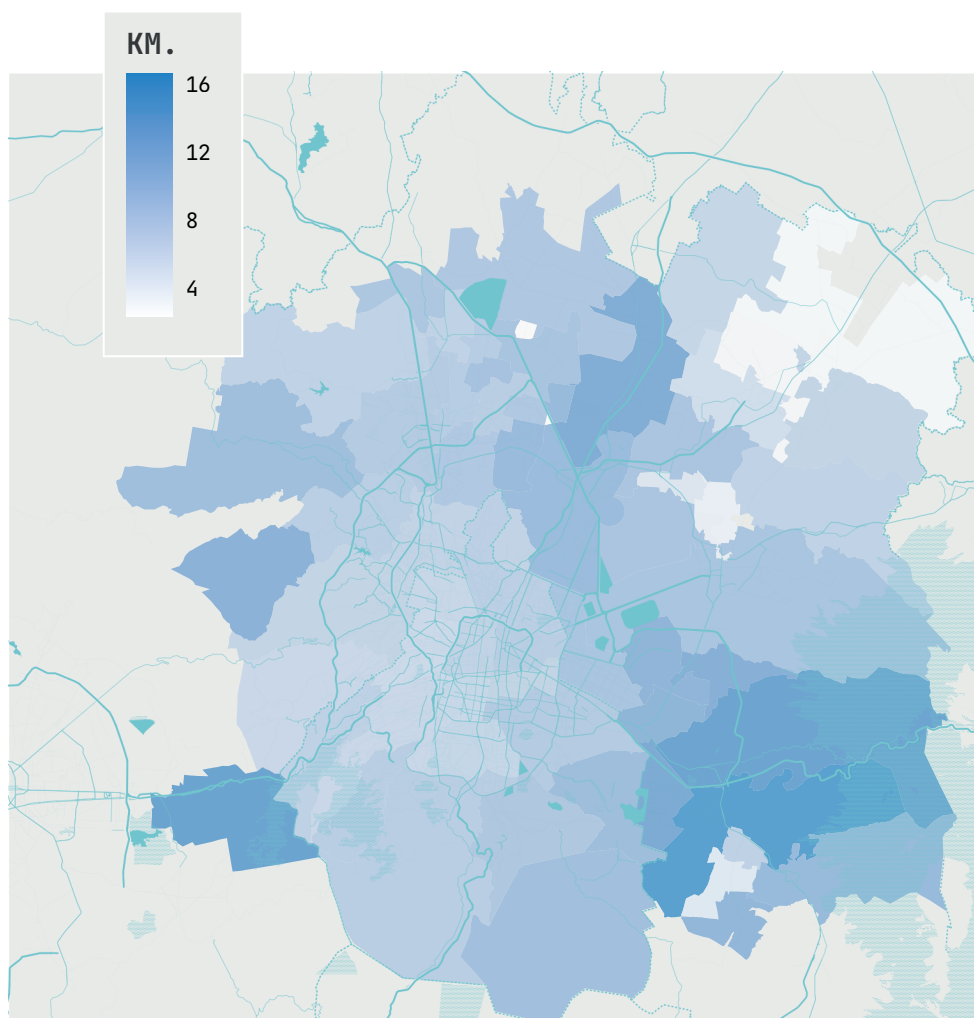
Fuente: Elaboración propia en base a Facebook Disaster Maps.

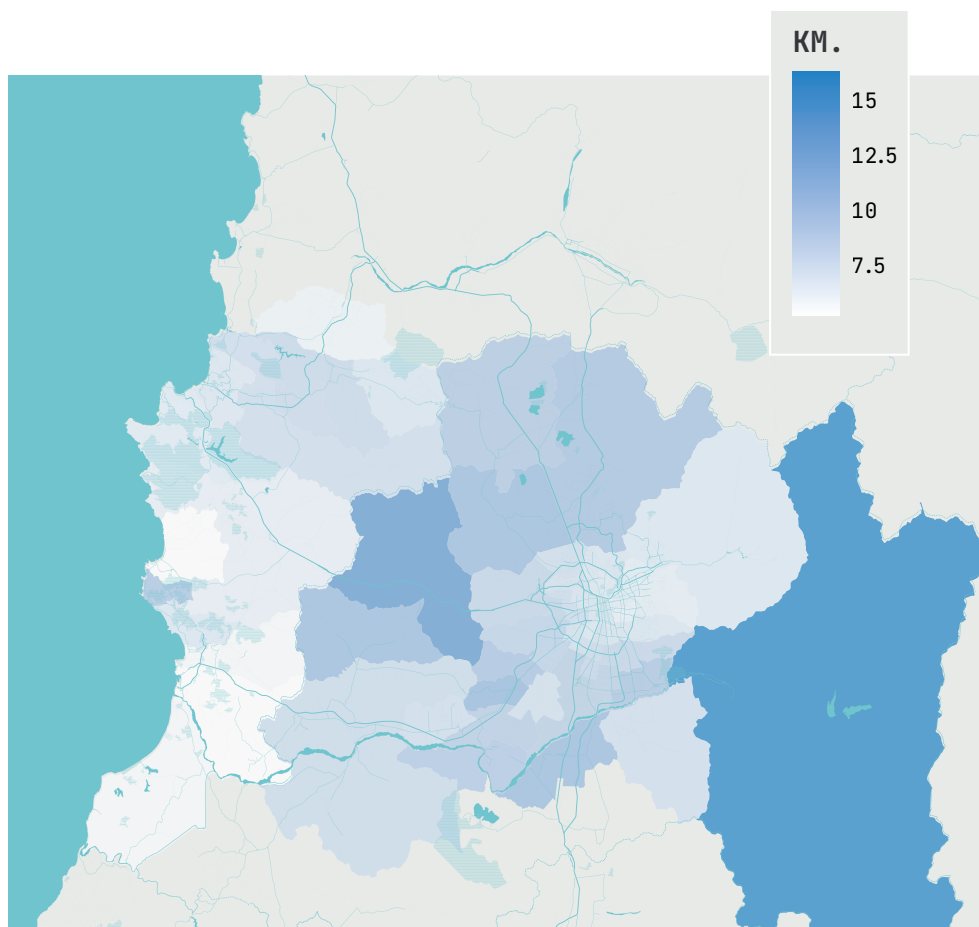


**DISTANCIAS
MEDIAS DE
CIUDAD
DE MÉXICO**

35

Fuente: Elaboración propia en base a Facebook Disaster Maps.





+

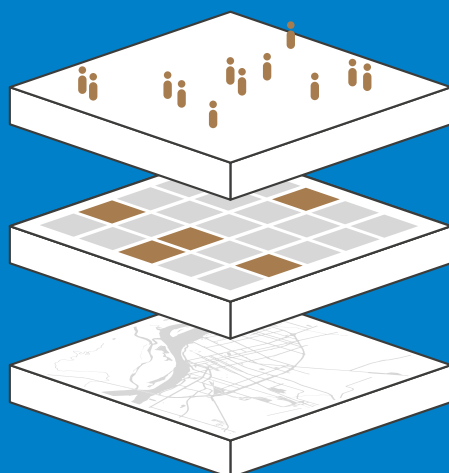
DISTANCIAS
MEDIAS DE
SANTIAGO
DE CHILE

36

Fuente: Elaboración propia en base a Facebook Disaster Maps.



OBSERVATORIO
DE MOVILIDAD
URBANA



ESTE REPORTE
SE PUBLICÓ
EN ARGENTINA.
NOV 2022