

Mapa de Movilidad de las Personas

Metodología

Resumen

El Mapa de Movilidad de las Personas presenta información acerca de la movilidad por áreas geográficas por día. Esta información corresponde a la reducción en el porcentaje de personas que recorrieron más de un kilómetro en un día. Esta reducción se expresa como el porcentaje de caída relativo a una semana de referencia pre-coronavirus (del 5 al 11 de marzo).

Se presenta información para áreas geográficas definidas como el primer nivel de desagregación geográfica de un país. Por ejemplo, esta desagregación corresponde al nivel de provincias en el caso de Argentina y a regiones en el caso de Perú.

La actual versión de la metodología es la 1.0. Esta versión comenzó a ser utilizada el 17-abr-20. La metodología, y este documento, se irá actualizando continuamente.

Procedimiento

A continuación, se describe el procedimiento llevado a cabo para realizar el análisis de cada país incluido en el Mapa de Movilidad de las Personas. Los pasos 1 a 3 involucran obtener la base inicial, definir los usuarios que se incluirán en el análisis para todo un mes (por ejemplo, marzo del 2020) y el lugar de residencia. Los pasos siguientes se realizan para cada día del período de análisis en un área geográfica.

- 1) *Obtención de la base.* Se obtiene una base de datos en donde la unidad de observación es un “ping”. Un ping es una medición de la latitud y longitud de un celular en un momento determinado. Esta base contiene las siguientes variables relevantes:
 - Identificación anonimizada del celular
 - Día y hora (incluyendo minutos y segundos)
 - Latitud
 - Longitud

- 2) *Selección de la muestra de usuarios incluidos en el análisis.* Se procede a restringir la muestra a usuarios con alta frecuencia de pings. Este paso está motivado en el hecho de que algunos usuarios presentan un alto número de pings para la mayoría de días, mientras que otros usuarios pueden tener pings solamente para unos pocos días del mes. Al restringir el análisis a los usuarios con una alta frecuencia de pings, se puede mantener un panel más balanceado a lo largo del tiempo porque estos usuarios serán observados continuamente.

Un criterio de inclusión que requiera una mayor frecuencia de pings de un individuo para ser incluido en el análisis, tendría el beneficio de que la muestra de usuarios que se

observa a lo largo del tiempo se mantendría más estable. Sin embargo, tendría un costo asociado a enfocar el análisis a una muestra potencialmente más selecta (y quizás menos representativa).

En nuestro análisis principal, consideramos como usuarios activos a aquellos que tienen más de 4 pings entre las 10 pm y 6 am en más de 8 días durante el mes de marzo. Utilizamos este criterio de inclusión en base a los pings que ocurren durante la noche por el siguiente motivo. Existen reportes que indican que la cantidad de pings generados por un celular estarían asociados con la movilidad de las personas.¹ Por ende, si restringiéramos la muestra en base a un alto número de pings generados durante todo el día, podríamos estar seleccionando una muestra que estaría sesgada a individuos con un alto nivel de movilidad. De esta forma, al seleccionar los usuarios activos en base al número de pings generados por la noche, evitamos este potencial sesgo.

- 3) *Definición del área de residencia.* Para cada usuario se define el lugar de residencia en base a la moda de la latitud y la longitud (en ambos casos, con tres decimales) de todos los pings registrados en el mes entre las 10 pm y las 6 am. Luego, se identifica en que área geográfica (como provincia para Argentina o región para Perú) se ubica esta residencia.

Los siguientes pasos se realizan para cada día y cada área geográfica.

- 4) *Selección de la muestra de usuarios incluidos en un día.* El siguiente paso involucra definir qué usuarios se incluirán para calcular el porcentaje de usuarios que se mueven más de un kilómetro en un día. En nuestro análisis principal incluimos los usuarios que tienen al menos dos pings (con menos de dos pings no es posible calcular la distancia).

Al definir un criterio mínimo para incluir a un usuario en un día se tienen que balancear dos objetivos. Por un lado, sería bueno enfocar el análisis en los usuarios que en el día bajo análisis tienen un alto número de pings (y bien distribuidos durante el día) ya que de esta forma el cálculo de la distancia recorrida por ese usuario será más preciso. Sin embargo, si una mayor movilidad genera más pings, la muestra quedaría sesgada hacia aquellos usuarios más móviles.

- 5) *Cálculo de la distancia recorrida en un día.* Para realizar este cálculo, se mide la distancia en línea recta entre el primer y segundo ping del día, luego entre el segundo y el tercer ping del día, y así sucesivamente. Luego, se suman estas distancias para aproximar la distancia total recorrida durante el día.

¹ El sitio web Unacast realiza un análisis de movilidad de personas para Estados Unidos. En la descripción de la metodología utilizada, se menciona que el número de pings está relacionado con el movimiento de las personas ([link](#), ver sección “Identifying the Strongest Signals”).

- 6) *Porcentaje de usuarios que recorrieron más de un kilómetro.* Se calcula el porcentaje de los usuarios que recorrieron más de un kilómetro entre los usuarios para los cuales se calculó la distancia recorrida durante el día.
- 7) *Cálculo de la reducción porcentual respecto a la semana de referencia.* Se calcula la reducción porcentual en el indicador estimado en el punto anterior (porcentaje de usuarios que recorrieron más de un kilómetro) en comparación con la semana de referencia, del 5 al 11 de marzo. Los días de lunes a viernes se comparan contra el promedio del valor del indicador durante los días de semana de referencia (jueves 5, viernes 6, lunes 9, martes 10 y miércoles 11). Para los sábados, se calcula la reducción porcentual en comparación al sábado 7 de marzo. Y, para los domingos, se calcula la reducción porcentual en comparación al domingo 8 de marzo.
- 8) *Suavizado de la serie en base a un promedio móvil de tres días.* La serie se suaviza realizando un promedio móvil de tres días del indicador calculado en el paso anterior. Por ejemplo, el indicador para cierta área geográfica para el 20 de marzo del 2020 corresponde al promedio de los valores del 19 al 21 de marzo del 2020 del indicador descrito en el paso anterior.
- 9) *Se eliminan los días en donde se utilizaron menos de 100 observaciones para calcular el valor.* Para asegurar la privacidad de los usuarios, se eliminan los valores de días en donde se utilizaron menos de 100 usuarios para realizar su cálculo.