Datatón

Tu dinero, tus datos.

Afluencia en el metro y la incidencia de casos de COVID-19 en la Ciudad de México durante 2020 - 2021.

Eje temático:

COVID-19 y reactivación económica

Equipo **Epi Mobility**

Daniela Valdez Hernández M

Alan García Zambrano

Objetivos del proyecto

Objetivo general

Describir la afluencia del metro por alcaldía y la incidencia de casos de COVID-19 en la ciudad de México durante el 01 de enero al 31 de agosto 2021.

Objetivos específicos

- Describir el comportamiento de los ingresos al Sistema de Transporte Colectivo Metro durante el periodo 2007 al 2021.
- Describir el porcentaje de cambio en los diferentes organismos de transporte en la Ciudad de México durante el periodo 01 de marzo 2020 al 31 de agosto 2021.
- Describir la opinión de los usuarios de Twitter acerca de las medidas sanitarias implementadas por las autoridades en el Sistema de Transporte Colectivo Metro durante el 01 de septiembre 2020 al 31 de septiembre 2021.

Desarrollo

Metodología

La limpieza, descripción, análisis y diseño de los gráficos, fueron creados en la versión 1.4 de RStudio y las librerías Tidyverse y Stringi en una PC con un procesador Intel core i5 con 8 Gb de RAM. Utilizamos la librería Leaflets y los shapes correspondientes a la Ciudad de México y sus 16 delegaciones, con la finalidad de establecer capas con los límites de cada delegación y geolocalizar las estaciones de metro con mayor concentración, así como las delegaciones con mayor densidad de casos de COVID-19. Se detalla a continuación las bases utilizadas, así como la periodicidad en la que fueron recolectadas.

- Base de datos COVID-19 DGE/SSA: Los datos analizados para casos y defunciones por COVID-19 en México se obtuvieron a partir de datos abiertos oficiales del Gobierno de México creados por el Sistema Nacional de Vigilancia Epidemiológica (SINAVE)(10). Utilizamos las variables fecha de inicio de síntomas, fecha de defunción y clasificación final para seleccionar los casos confirmados a SARS-CoV-2 ya sea por asociación epidemiológica, laboratorio o dictaminación.
  Actualmente, México es uno de los quince países con mayor número de casos (aproximadamente 3,600,00) y el cuarto en defunciones (aproximadamente 273,000) a nivel global, del 20 de febrero de 2020 al 15 de

octubre de 2021, según el portal de Datos Abiertos de la Dirección General de Epidemiología (10).

- Base Secretaría de Movilidad: Se utilizó la base de datos de “Afluencia diaria en el metro” de la Secretaría de Movilidad (SEMOVI)(11), con actualización del 1 de enero al 30 de septiembre de 2021. La base contiene 53,236 registros a partir de las 164 estaciones de servicio de cada una de las 12 líneas de metro, distribuidas en las 16 delegaciones en la Ciudad de México. La base de Afluencia se utilizó con la finalidad de conocer qué tipo de transporte es el más utilizado en la Ciudad de México y cuál es la afluencia de personas durante el periodo de estudio. Se le añadieron las variables latitud y longitud a cada una de las estaciones para crear un sistema de geolocalización en un mapa virtual.
- Base Ingresos del Sistema de Transporte Colectivo Metro: con actualización del 1 de enero de 2012 al 15 de octubre de 2021. La base contiene 10,523 registros y 14 variables (Fecha, tipo de ingreso, boleto, tarjeta y recarga de las doce líneas del metro). Esta base nos proporciona el valor de la cantidad de personas que ingresan al sistema, ya sea por boleto o recarga en cada una de las 12 líneas del metro. Mediante este análisis sabremos en qué líneas, y los días en que ocurre la mayor concentración de personas y durante el tiempo (12).
- Base Movilidad de Google Community Mobility Reports (CMR): con los datos de movilidad en los usuarios de la Ciudad de México. Se utilizaron 8 variables, de las cuales seis son indicadores de Movilidad activa (Lugares recreativos, Parques, Farmacias y tiendas, Estaciones de trasporte y lugares de trabajo) e inactiva (Residencial) a partir del 15 de febrero de 2020 al 15 de octubre de 2021, y miden los cambios porcentuales con respecto a la mediana de una semana atípica del mes de febrero 2020, antes iniciar las primeras medidas de restricción a la movilidad. La base se utilizó como una medida de referencia para correlacionar si el cambio en el desplazamiento de los usuarios es proporcional al de los sistemas de transporte público, con respecto a las variables de movilidad del Google CMR de la Ciudad de México. Estos datos son anonimizados y públicos, procedentes de usuarios que han activado el historial de ubicaciones.
- Twitter API: Parte importante en el desarrollo de este documento y propuesta de mejora, es utilizar la charla activa que se da en medios sociales, específicamente hablando de Twitter para descubrir la parte social, qué piensan las personas y sobre todo cuál es el sentimiento que genera la movilidad con respecto al tema de COVID- 19 en la Ciudad de México. Para ello se utilizó el acceso al Application Program Interface (API) de Twitter mediante el lenguaje de programación Python y la librería Tweepy, mediante una lista de palabras de búsqueda que incluyen: [“metro”, “cdmx”, “metrobus”, “covid”, “transporte”]. La información recolectada se almacenó en una base de datos (CSV) con tres variables principales: fecha, texto de

tweet y localidad. Se recolectaron 494 tweets del 01 de septiembre de 2020 al 19 de septiembre del 2021 en la Ciudad de México, se realizó una limpieza identificando únicamente temas relacionados con el STC, manteniendo el 21.2% (n=105) del total. La finalidad es descubrir qué palabras ocurren con mayor frecuencia de los usuarios hacia el sistema de transporte colectivo de la Ciudad de México, durante el periodo de estudio. Los datos de Twitter son públicos y cualquier persona puede acceder a ellos.

Hallazgos

Durante el periodo de estudio (20 de febrero de 2020 al 30 de septiembre de 2021) se registraron aproximadamente 3,470,239 casos y 274,086 defunciones a nivel nacional. De los cuales el 26% (n = 954,920) y el 14% (n=39,385) respectivamente, son casos y defunciones asociados a SARS-CoV-2, con una tasa de incidencia de 10,368 casos acumulados y 427 defunciones por cada cien mil habitantes pertenecientes a la Ciudad de México. Las cinco delegaciones con mayor número de casos acumulados son: Iztapalapa, Álvaro Obregón, Gustavo A. Madero, Tlalpan y Coyoacán, respectivamente (Ver anexo 1 y Tabla 1). La afluencia en el transporte pública en la Ciudad de México disminuyó considerablemente a inicio de la pandemia por COVID-19. A partir del 11 de marzo de 2020, cuando COVID-19 fue declarado pandemia, se ha visto una disminución en la afluencia; se observa que el 01 de junio de 2020 cuando inició la "nueva normalidad", la afluencia empieza a ir en aumento. El transporte con mayor afluencia es el STC Metro. (Ver anexo 2).

En la base de Ingresos al Metro comprobamos que, al inicio del periodo de estudio, a pesar de la drástica caída de los ingresos por boletos y recargas, se ha mantenido y aumentado la interacción de los usuarios con los sistemas de recarga. Esto probablemente se debe a las medidas de distanciamiento y al aumento de la infraestructura tecnológica con los nuevos sistemas de recarga de créditos. Durante el 11 de marzo al 30 de mayo de 2020, que coincide con la jornada de sana distancia, fue la peor época para el STC Metro en décadas (Ver anexo 3 y 4).

La base de Google CMR nos muestra el cambio porcentual entre los diferentes tipos de movilidad. Observamos que a partir del inicio de la pandemia disminuyo casi al 75% con respecto a una semana atípica de febrero de 2020, en algunas actividades relacionadas con el transporte en el Metro (Ver anexo 5). Aunque aún no se ha recuperado en su totalidad, el comportamiento de algunas variables como recreación, tránsito y lugares de trabajo, tienen relación con los Ingresos al Metro. Se comprobó que existe una correlación positiva superior a 0.9 y un valor [$p < 0.05$] por método de "spearman", durante el periodo de estudio. Esto nos ayudó a comprobar que los datos generados por el STC Metro (afluencia diaria e ingresos), tienen relación con los datos proporcionados por Google CMR y que representan a la proporción de usuarios con dispositivos Android que utilizan el metro como medio de transporte en estaciones de transporte público, actividades recreativas y lugares de trabajo (Ver anexo 6).

Del total de los datos registrados en la base de Afluencia en el metro de la SEMOVI, identificamos que la variable "afluencia" presenta corrupción en el 13.8% de la información total (n=7,348). El 6% (n =3,205) de los registros presenta un carácter

agregado (coma) y no permite operaciones aritméticas, mientras que el 7.8% (n= 4,143) son valores nulos, por lo que se retiraron estos registros, los cuales no representaban valores numéricos. Se geolocalizaron los puntos de latitud y longitud de cada estación y mediante un mapa de calor se visualiza que la mayor concentración en afluencia de personas corresponde a las terminales Pantitlán, Indios Verdes, Constitución de 1910, Tacubaya, Cuatro Caminos, Tasqueña, Buenavista y Observatorio. Se observa además que en la zona céntrica perteneciente a las delegaciones Cuauhtémoc, Benito Juárez y Álvaro Obregón, es en donde se acumula mayor número de personas (Ver anexo 7). Posteriormente se utilizó una capa de shapes con los polígonos para distribuir por delegación los casos de COVID-19 registrados. Encontramos que en la zona centro (Cuauhtémoc y Benito Juárez) es en donde se concentra mayor cantidad de casos, al igual que Iztapalapa, Álvaro Obregón y Gustavo

A. Madero (Ver anexo 8). Por último, encontramos entre los comentarios de Twitter algunas opiniones de rechazo relacionadas con uso incorrecto del cubrebocas por parte de usuarios y mismo personal del metro, así como el retraso en los trenes, que implica aglomeraciones de personas. Estos comentarios aparecían con mayor frecuencia durante el periodo de la segunda ola de coronavirus, aproximadamente del 20 diciembre 2020 a febrero de 2021 (Ver anexos 9 y 10).

Propuesta

Debido a nuestros descubrimientos en el análisis de esta información se propone:

Concluir un tablero con Shiny de RStudio que incluirá todos los gráficos que se desarrollaron de forma dinámica e interactiva, así como mapas de calor y el desarrollo de polígonos de datos y mapas de calor para cada una de las estaciones al interior del metro con la finalidad de reforzar y priorizar las acciones sanitarias en esas estaciones del metro.

(dependiendo de la mejoría de la información) mantener los tableros alimentándose de forma automática y constante, con una periodicidad diaria.

Un acercamiento con los coordinadores de información de los sistemas de transporte colectivo, como un esfuerzo para desarrollar nuevas bases que representen información de mejor calidad, como el análisis por hora, estación, torniquete e identificar y mejorar otras áreas de oportunidad como los retrasos en trenes, zonas de mayor concentración de personas, y dónde surgen quejas en tiempo real por medio de charlas sociales como Twitter (análisis de sentimiento), con una visión hacia el análisis predictivo.

Conclusiones

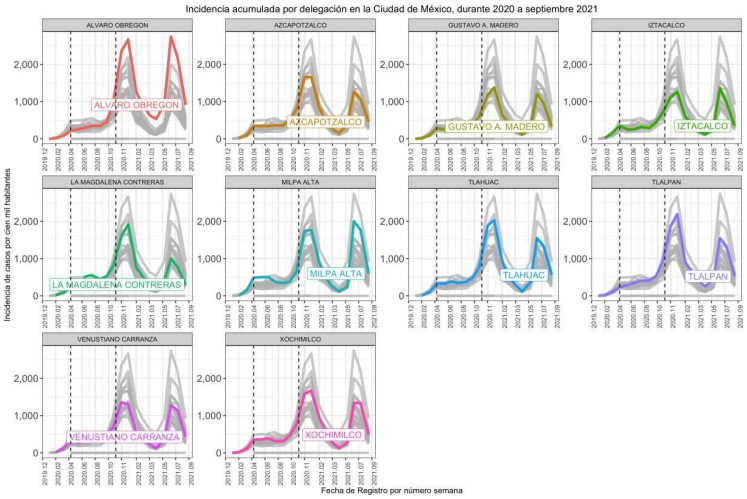
La pandemia de COVID-19 ha tenido muchas repercusiones en el país y la movilidad se vio afectada desde el inicio de esta. La decisión de realizar un confinamiento tuvo muchos beneficios con respecto a salud y a su vez afectaciones en todos los sectores, y debido a la disminución de la afluencia, las

pérdidas económicas han sido importantes. Durante el análisis de este proyecto nos dimos cuenta que fue necesario realizar un esfuerzo extra en la limpieza de la información, sin embargo, y dado el tiempo de desarrollo decidimos enriquecer el análisis con información adicional como la charla en medios sociales.

Concluimos que la finalidad de este documento es proponer la recuperación del metro como un espacio de transitar sin riesgo al contagio o transmisión de la enfermedad. Esto propiciará que los ingresos de los capitalinos regresen a su normalidad lo más pronto y seguro posible.

Anexos

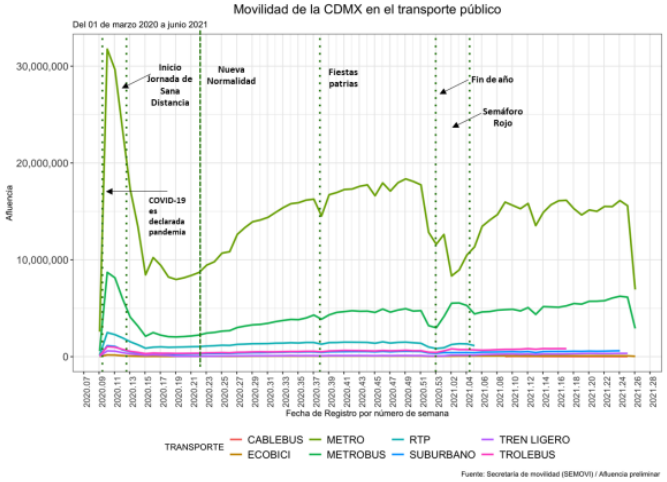
Se incluye el código en R al final de la descripción y en el siguiente enlace de Github https://github.com/AllanZamb/Dataton



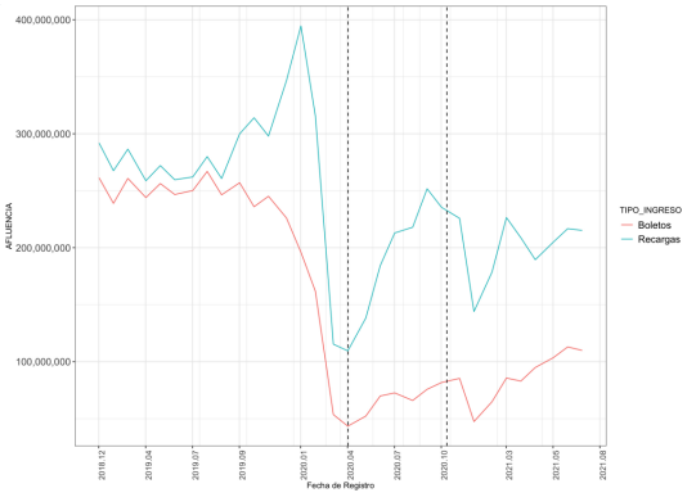
Anexo 1 Se visualizan las diez delegaciones con mayor incidencia de casos por COVID-19 en las delegaciones de la ciudad de México por cien mil habitantes.

Tabla 1 Diez principales delegaciones con número de casos acumulados por COVID-19 en la Ciudad de México durante el 20 de febrero de 2020 al 15 de octubre de 2021.

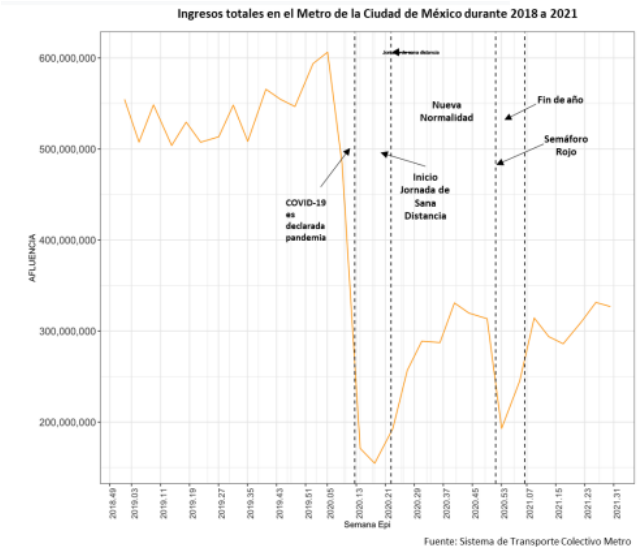
Delegación	Casos acumulados confirmados a SARS CoV-2	Incidencia de casos confirmados a SARS CoV-2 por 100,000 habitantes.
Iztapalapa	150,759	8,213
Álvaro Obregón	137,775	18,148
Gustavo A. Madero	108,370	9,236
Tlalpan	93,333	13,344
Coyoacán	51,813	8,432
Xochimilco	51,701	11,692
Tláhuac	49,714	12,672
Cuauhtémoc	49,073	8,989
Azcapotzalco	49,068	11,352
Venustiano Carranza	45,035	10,149



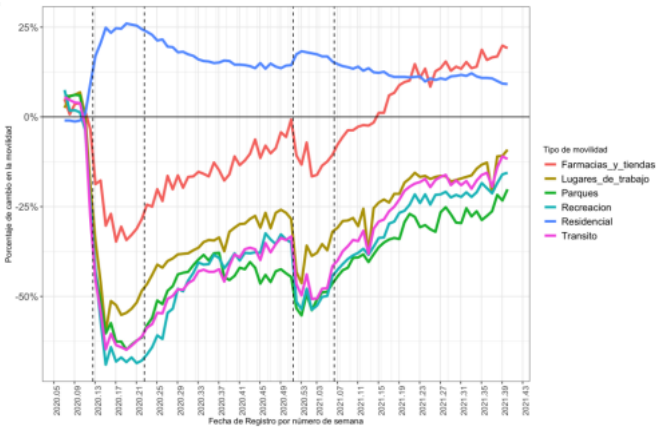
Anexo 2 Se Ilustra la gráfica de movilidad para los diferentes organismos de transporte público en la Ciudad de México durante 01 de marzo a junio de 2021.



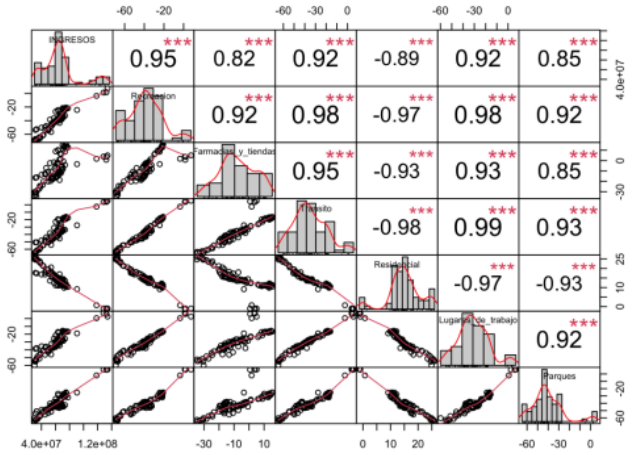
Anexo 3 Se Ilustran los ingresos al STC Metro mediante boletos y recargas. Se puede visualizar que las caídas en los ingresos dependen de las medidas del gobierno para reducir el número de contagios y esto afecta directamente a la economía del transporte. Por otra parte, podemos observar que los ingresos en recargas se han recuperado mucho más rápido que los boletos.



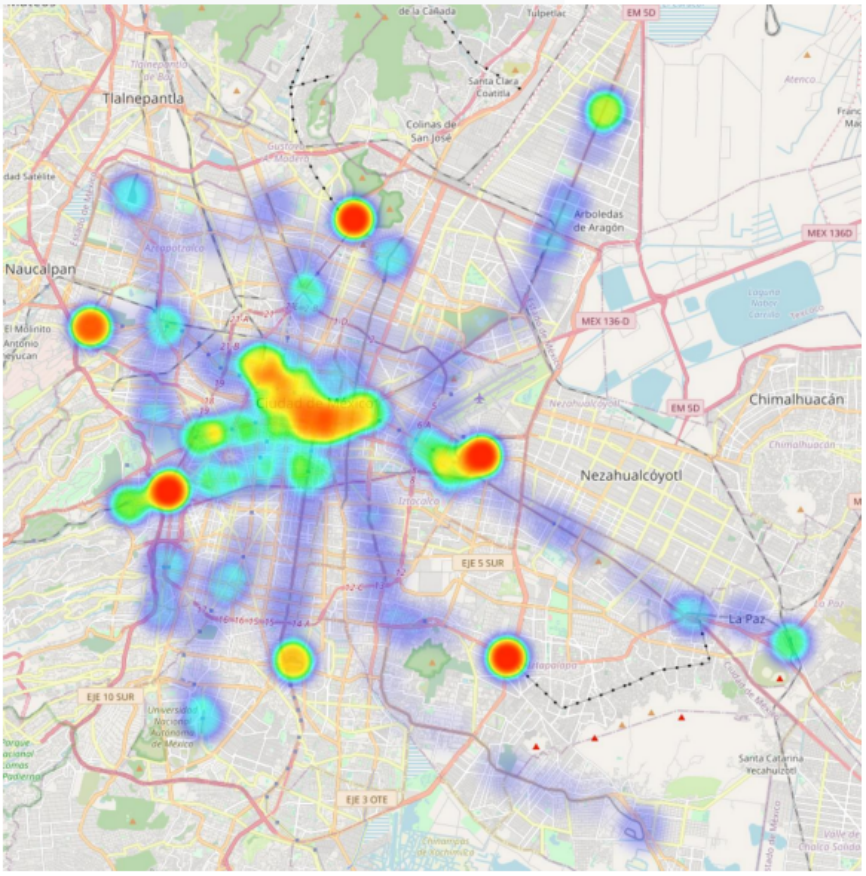
Anexo 4 Se Ilustran los ingresos acumulados al STC Metro mediante boletos y recargas. Se puede visualizar que las caídas en los ingresos dependen de las medidas del gobierno para reducir el número de contagios y esto afecta directamente a la economía del transporte. Por otra parte, podemos observar que la afluencia del metro aún no se ha recuperado ni la mitad a una semana atípica antes de la pandemia.



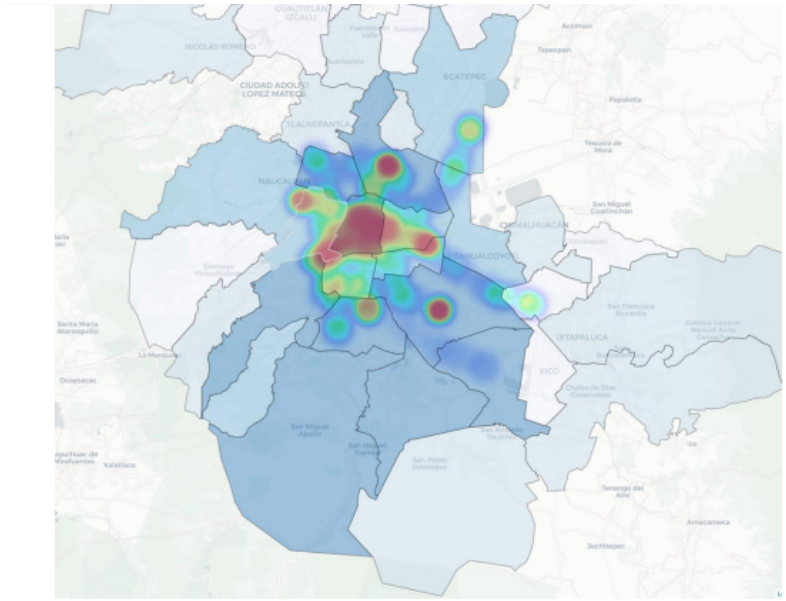
Anexo 5 La información de Movilidad de Google CMR, nos proporciona información actualizada sobre cómo son los comportamientos sociales. Las líneas verticales señalan la jornada de sana distancia y el semáforo rojo de diciembre de 2020. Es importante prestar atención a estos indicadores ya que permiten inferir qué movimientos realizan los ciudadanos.



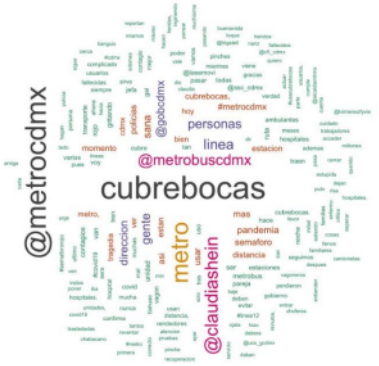
Anexo 6 Se muestran las correlaciones que se encontraron en las bases de Google CMR e Ingresos en el STC. Como se puede notar, la fuerza de correlación aumenta de forma similar en Recreación, Transporte y lugares de trabajo contra Ingresos.



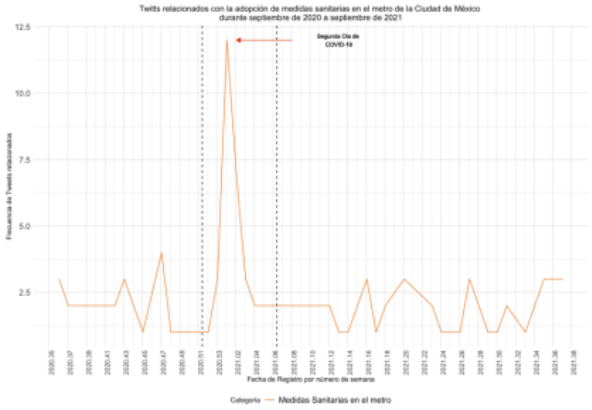
Anexo 7 Se visualiza un mapa de calor con la concentración de afluencia en cada una de las estaciones del metro de la Ciudad de México. Se puede observar que la afluencia se concentra en terminales, pero sobre todo es mayor en la zona centro de la ciudad.



Anexo 8 Se visualiza el mapa de calor con la mayor concentración de afluencia en la ciudad de México, y las partes azules son zonas en donde se registró mayor número de contagios por COVID-19 durante el periodo de estudio.



Anexo 9 La nube de palabras nos permite conocer qué palabras aparecen con mucho mayor frecuencia dentro de un conjunto de palabras. La palabra con mayor frecuencia es cubrebocas, semáforo, pandemia, distancia, etc. Esto representa qué tantas personas utilizan esa palabra para referirse al metro de la Ciudad de México.



Anexo 10 Durante el periodo de tiempo en el que se utilizó la información de Twitter de septiembre 2020 a Septiembre 2021, podemos observar que en diciembre de 2020, aparece un pico en la cantidad de tweets registrados. Esto tiene que ver con los casos que se registraron durante el periodo conocido como "segunda ola".

Bibliografía y fuentes consultadas

- Archivado: Cronología de la OMS - COVID-19 [Internet]. [cited 2021 Oct 5]. Available from: https://www.who.int/news/item/27-04-2020-who-timeline---covid-19
- Conferencia 11 de marzo – Coronavirus [Internet]. [cited 2021 Oct 11]. Available from: https://coronavirus.gob.mx/2020/03/11/conferencia-11-de-marzo/
- Comunicados Técnicos Diarios COVID 19 | Secretaría de Salud | Gobierno | gob.mx [Internet]. [cited 2021 Oct 15]. Available from: https://www.gob.mx/salud/documentos/comunicados-tecnicos-diarios-covid19
- DOF - Diario Oficial de la Federación [Internet]. [cited 2021 Oct 13]. Available from: https://www.dof.gob.mx/nota_detalle.php?codigo=5590339&fecha=24/03/2020
- Movilidad durante la emergencia sanitaria COVID-19 [Internet]. [cited 2021 Oct 5]. Available from: https://www.semovi.cdmx.gob.mx/tramites-y-

servicios/transparencia/preguntas-frecuentes/preguntas-frecuentes-covid-19/movilidad-durante-la-emergencia-sanitaria-covid-19

- NUEVA NORMALIDAD | COVID-19 Medidas Economicas | Gobierno | gob.mx [Internet]. [cited 2021 Oct 13]. Available from: https://www.gob.mx/covid19medidaseconomicas/acciones-y-programas/nueva- normalidad-244196
- Donde Vamos H. ANATOMÍA DE LA MOVILIDAD EN MÉXICO.
- afluencia por tipo de acceso 2019 [Internet]. [cited 2021 Oct 14]. Available from: https://metro.cdmx.gob.mx/afluencia-por-tipo-de-acceso-2019
- afluenciadeestacionporlinea2020 [Internet]. [cited 2021 Oct 14]. Available from: https://metro.cdmx.gob.mx/afluenciadeestacionporlinea2020
- Datos Abiertos Dirección General de Epidemiología | Secretaría de Salud | Gobierno | gob.mx [Internet]. [cited 2021 Oct 15]. Available from: https://www.gob.mx/salud/documentos/datos-abiertos-152127
- Afluencia diaria del Metro CDMX - Conjunto de datos - Portal de Datos Abiertos de la CDMX [Internet]. [cited 2021 Oct 15]. Available from: https://datos.cdmx.gob.mx/dataset/afluencia-diaria-del-metro-cdmx
- Ingresos del Sistema de Transporte Colectivo Metro - Conjunto de datos - Portal de Datos Abiertos de la CDMX [Internet]. [cited 2021 Oct 15]. Available from: https://datos.cdmx.gob.mx/dataset/ingresos-del-sistema-de-transporte-colectivo- metro