



Universidad del Desarrollo
Facultad de Ingeniería

OPTIMIZACIÓN DE LA UBICACIÓN DE BUSES DE ACCESO A COLEGIO EN
VITACURA

POR: CRISTIAN CAROCA Y NICOLAS KLEIN

Capstone Project presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

DRA. LORETO BRAVO CELEDON

Enero 2023
SANTIAGO

CCC: Dedicado a mi familia Paula, Simón y Lucas, por todo su apoyo.

NKG: Dedicado a mi familia Katherine mi compañera de vida, Ignacio y mi madre.

AGRADECIMIENTOS

A la Universidad que nos permitió realizar este magíster, a todos los profesores y en forma especial a Ismael Botti, nuestro profesor de Python de quien lamentamos su partida recordando su calidad humana y profesional.

A Pablo Rojas, por su extraordinario y desinteresado apoyo.

TABLA DE CONTENIDO

| | |
|------------------------------------------|-----------|
| RESUMEN | 1 |
| 1. INTRODUCCIÓN..... | 3 |
| 2. TRABAJO RELACIONADO | 6 |
| 3. HIPÓTESIS Y OBJETIVOS..... | 7 |
| 3.1. IDENTIFICACIÓN DE UN PROBLEMA | 7 |
| 3.2. HIPÓTESIS | 7 |
| 3.3. OBJETIVOS | 7 |
| 4. DATOS Y METODOLOGÍA..... | 9 |
| 4.1. DATOS..... | 9 |
| 4.2. METODOLOGÍA | 11 |
| 5. RESULTADOS..... | 20 |
| 6. CONCLUSIONES | 24 |
| 6.1. LIMITACIONES | 24 |
| 6.2. TRABAJO FUTURO | 25 |
| 7. BIBLIOGRAFÍA | 26 |

Resumen

El objetivo de este trabajo es establecer la ubicación de buses de acercamiento para un colegio ubicado en la comuna de Vitacura, a través del análisis del tráfico móvil, capturado por antenas móviles de una compañía de teléfonos. Se busca disminuir el tráfico en la zona de entrada del establecimiento, el cual además está ubicado en la intersección de 4 autopistas.

Este documento está dividido en 7 capítulos.

El primero es la introducción, la cual da cuenta de los estudios que se han realizado en torno a los datos recogidos desde las antenas de móviles, cómo estos sirven para determinar comportamiento de desplazamiento de las personas y cómo se puede estimar, con una alta probabilidad, el domicilio del individuo. Estos dos aspectos son fundamentales para plantear la forma de resolver el problema materia del documento.

El segundo capítulo describe los trabajos a partir de los cuales se apoya la hipótesis de este documento. El tercer capítulo plantea el origen del problema que queremos abordar, define la hipótesis que se requiere resolver y por último establece 2 objetivos específicos que también se busca obtener una aproximación a la solución.

El capítulo cuatro aborda la estructura de los datos que utilizamos en este trabajo, el origen y la conformación de estos, para luego profundizar en el tratamiento que se debió utilizar, cuales fueron los distintos procesos y filtros para lograr dar con la información que permitiera abordar el problema, también se realiza una breve descripción acerca de las librerías de Python que utilizamos en cada etapa. El quinto capítulo versa acerca de los resultados obtenidos, tanto gráficos como estadísticos. Se explica el detalle del uso de los

criterios definidos para modelar el problema y obtener un resultado que sea satisfactorio. En el sexto capítulo están las conclusiones del trabajo. También describe las limitaciones que tuvimos en el desarrollo de la hipótesis y plantea escenarios para la aplicación futura del análisis en que se basó el resultado encontrado. El séptimo capítulo enumera los trabajos sobre los cuales se definen los principales criterios para darle consistencia teórica a los supuestos y estimaciones que se utilizan para dar con una solución al problema.

1. Introducción

La ciudad de Santiago ha tenido un fuerte aumento en la congestión vehicular. Los factores principales son el aumento del parque vehicular (en 2021 se vendió un 61% más autos que el 2020) y el efecto pandemia (alza en el uso del auto particular por sobre transporte público).

El uso de los datos generados por las antenas móviles nos puede entregar información acerca de zonas con alto tránsito y como la gente se mueve dentro de la ciudad. Esto se ve validado por Han Sumin [1], quien identifica el uso de los datos generados por las torres celular LTE para establecer “Puntos de Interés” (PDI) dentro de la ciudad, incluso propone un esquema de explotación diaria (casi en tiempo real). Durante la pandemia de COVID-19, la información de movilidad de la población se obtuvo en gran medida por los grandes operadores móviles [5].

En este documento busca, a través de datos generados por las antenas móviles, definir la ubicación para buses de acercamiento de un colegio ubicado en la comuna de Vitacura (Región Metropolitana, Chile), evitando la alta congestión vehicular que se produce al momento de entrada de los alumnos. En primer lugar, determina las antenas cercanas al recinto educacional, dentro de una ventana de tiempo que rodea el horario de entrada de este, y captura de la información de los individuos conectados. Posteriormente establece el hogar de estos, para luego estimar las posibles trayectorias al colegio. Por último, contrastamos el tráfico, al momento de la mañana, con los domicilios y a partir de análisis visual, definimos el lugar de ubicación de los buses de acercamiento.

Para definir el subconjunto de apoderados, acotamos la información de las antenas cercanas al momento de ingreso al colegio. Para realizar este filtro nos basamos en Akella,

Mohan [2] cuyo trabajo de investigación tiene por objetivo definir el mejor lugar para instalar una torre adicional a una red celular existente. Dentro de los criterios para aumentar el rendimiento de la señal, tenemos la interferencia en la línea vista, ya sea por edificios, vegetación o elevaciones de terreno. Estos elementos hacen que la señal se atenúe o refleje en un alto grado. La altura del cerro La Pirámide supera por lejos el rango de altura que debería tener la antena para poder superar la elevación y lograr dar conectividad a los equipos que están al otro lado (en el colegio). El resto de las antenas poseen mucha distancia y/o a la vez no poseen una línea vista limpia hacia el establecimiento educacional.

Por otro lado, está el problema de definir el lugar de domicilio de los apoderados. Oosterlink [3] utiliza datos del registro de llamadas (CDRs – Call Detail Records) logrando una precisión (en el mejor caso) de un 61%. Si agrega la información de redes sociales aumenta a 72%. Esto no podemos aplicarlo ya que partimos de la premisa que no conocemos los individuos.

Pappalardo, Ferres, Sacasa, Bravo [4], abordan el problema, a través de los datos generados en la red móviles de 65 individuos durante 2 semanas. Realizan un benchmark respecto del mejor dato generado en la red, estos son: CDRs (Call Detail Records), XDRs (eXtended Detail Records) y CPRs (Control Plane Records). Posteriormente estos son analizados por medio de 37 algoritmos de detección de vivienda (HDA - Home Detection Algorithms), concluyendo que el escenario de mayor precisión se logra utilizando los datos XDR, midiéndolos en una ventana horaria de las 19:00 a las 07:00 hrs, durante días hábiles y tomando en consideración las 3 torres más cercanas, las cuales son definidas a partir de las que poseen mayor actividad de los usuarios. Este algoritmo logra una

precisión de un 69%. Este es el enfoque que se utiliza, sobre el cual se apoya la definición de los domicilios de los apoderados. Esta información nos permite establecer los posibles trayectos o zonas sobre las cuales podemos estimar lugares para la ubicación de buses de acercamiento.

2. Trabajo Relacionado

Este documento se apoya fundamentalmente en 2 trabajos. Akella, Mohan [2], el cual tiene por objetivo calcular la posición óptima de una nueva antena celular dentro de una red ya desplegada. Se aborda esta problemática ya que el mercado de las telecomunicaciones en USA, en lugares remotos o ciudades pequeñas, las grandes empresas de telecomunicaciones no llegan con redes propias, sino que utilizan un modelo de roaming con proveedores locales. El modelo de negocio consiste en que se paga al proveedor local en la medida que logre el mayor número de usuarios conectados utilizando su servicio. No existe una tarifa plana o un servicio de postpago que garantice al proveedor local un ingreso asegurado. Se debe extremar recursos para lograr el máximo de usuarios conectados. Es por esto que la definición de una nueva antena implica un estudio detallado de los usuarios potenciales, de acuerdo con la cobertura que se logra a través de la red ya existente. Para establecer la nueva torre se definen criterios de línea vista de las torres, concluyendo que la interposición de edificios, vegetación o desniveles del terreno inciden directamente en la degradación del servicio. Estas definiciones proporcionaron elementos suficientes para filtrar aquellas antenas que tomaban la conectividad de los apoderados a la hora de la entrada al colegio.

El paper de Pappalardo, Ferres, Sacasa, Bravo [4], establece criterios de selección de antenas móviles para definir el lugar de domicilio de un individuo. Esto permitió avanzar rápidamente en la forma de establecer el punto de inicio del trayecto hacia el colegio. Los datos que se utilizan coinciden con el HDA que obtiene el mayor porcentaje de asertividad a la hora de definir el domicilio. Estos datos son los XDR, los cuales nacen a partir de la utilización de aplicaciones desde el celular.

3. Hipótesis y Objetivos

3.1. Identificación de un problema

Todos los días se genera una alta concentración vehicular en la entrada (y en las calles aledañas) al Saint George's College (Santa Cruz 5400, Vitacura, 7660103, Región Metropolitana) durante el arribo de los estudiantes (07:55 hrs AM). El objetivo es conocer la ubicación de los hogares de los individuos que transitan hacia el colegio en horario de entrada a clases o en sus cercanías, de manera de determinar recomendación de lugares donde establecer buses de acercamiento. Buscando reducir el tránsito en el lugar, junto con disminuir el peligro de accidentes.

3.2. Hipótesis

La hipótesis es validar si, a través del uso de los datos XDR, podemos estimar los puntos de concentración de tránsito, en las zonas cercanas al colegio, en el horario de entrada a clases (desde las 07:15 a las 8:30 hrs AM). Así también, establecer la ubicación de los hogares de los individuos que circulan a esa hora. Por último, con toda esta información determinar puntos de acercamiento al colegio, evitando que los apoderados lleguen a la entrada.

3.3. Objetivos

El objetivo general es determinar los lugares donde ubicar buses de acercamiento al colegio.

Los objetivos específicos son:

1. Determinar las zonas con mayor tráfico a la hora de entrada y establecer el hogar de los apoderados para estimar el recorrido desde sus hogares al colegio.
2. Visualizar la información en un formato que permita la manipulación por parte de personas responsables del colegio o la comuna, que puedan hacer uso de la información procesada y obtener conclusiones adicionales al objetivo general de este trabajo.

4. Datos y Metodología

4.1. Datos

Los datos corresponden a la información de todas las conexiones XDR (eXtended Detail Records) ¹ que se realizan entre los equipos móviles y las celdas o antenas móviles de una compañía de telecomunicaciones. Este tipo de registro es generado en parte por el usuario (al momento de usar una aplicación que requiere conectarse a la red) y en parte por el dispositivo del teléfono ². La muestra consiste en 3 días consecutivos del mes de Marzo, de los años 2020, 2021 y 2022. Los días de cada año son:



Los datos se encuentran distribuidos en forma aleatoria en 500 archivos, de 530 MB (app) cada uno. Los cuales contiene, en total, cerca de 3.500.000.000 de registros.

Cada uno de los registros posee 4 elementos (argumentos): device, datetime, lat y lon.

| Nombre | Descripción | Tipo de Dato |
|----------|---------------------------------------------------------------------------------------------------|--------------|
| device | Identificación anonimizada del dispositivo móvil del usuario | object |
| datetime | El momento en que se realiza la conexión, identificado por: día, mes, año, hora, minuto y segundo | datetime64 |
| lat | Latitud (coordenada) de la antena a la cual se realiza la conexión | float64 |
| lon | Longitud (coordenada) de la antena a la cual se realiza la conexión | float64 |

¹ <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-021-00284-9>

² <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-021-00284-9>

```
part_499.head(5)
```

| | device | datetime | lat | lon |
|---|---------------------------------------------------|---------------------|-----------|-----------|
| 0 | 61424f08b0d2c26b13080b740edb41e5f2f1c0944d0c23... | 2020-03-04 17:15:36 | -33.57610 | -70.70188 |
| 1 | e4673e895272c9f50f78973d4a9e217f8c90aed9a6f60f... | 2020-03-03 23:14:41 | -23.48974 | -70.11490 |
| 2 | 791160e1f3bf4d93cc06602f5a591efd644cea68d418ed... | 2020-03-05 22:03:41 | -41.24914 | -73.63945 |
| 3 | 4142f8e0003e51bedeb201faa01468a00a255a2d65057f... | 2020-03-03 07:53:23 | -41.92497 | -73.51266 |
| 4 | ed8ff0081e3796a314d07c91143c13799c95711df47cc2... | 2020-03-05 17:19:16 | -38.87402 | -72.62003 |

Imagen 2: df 499 del dataset.

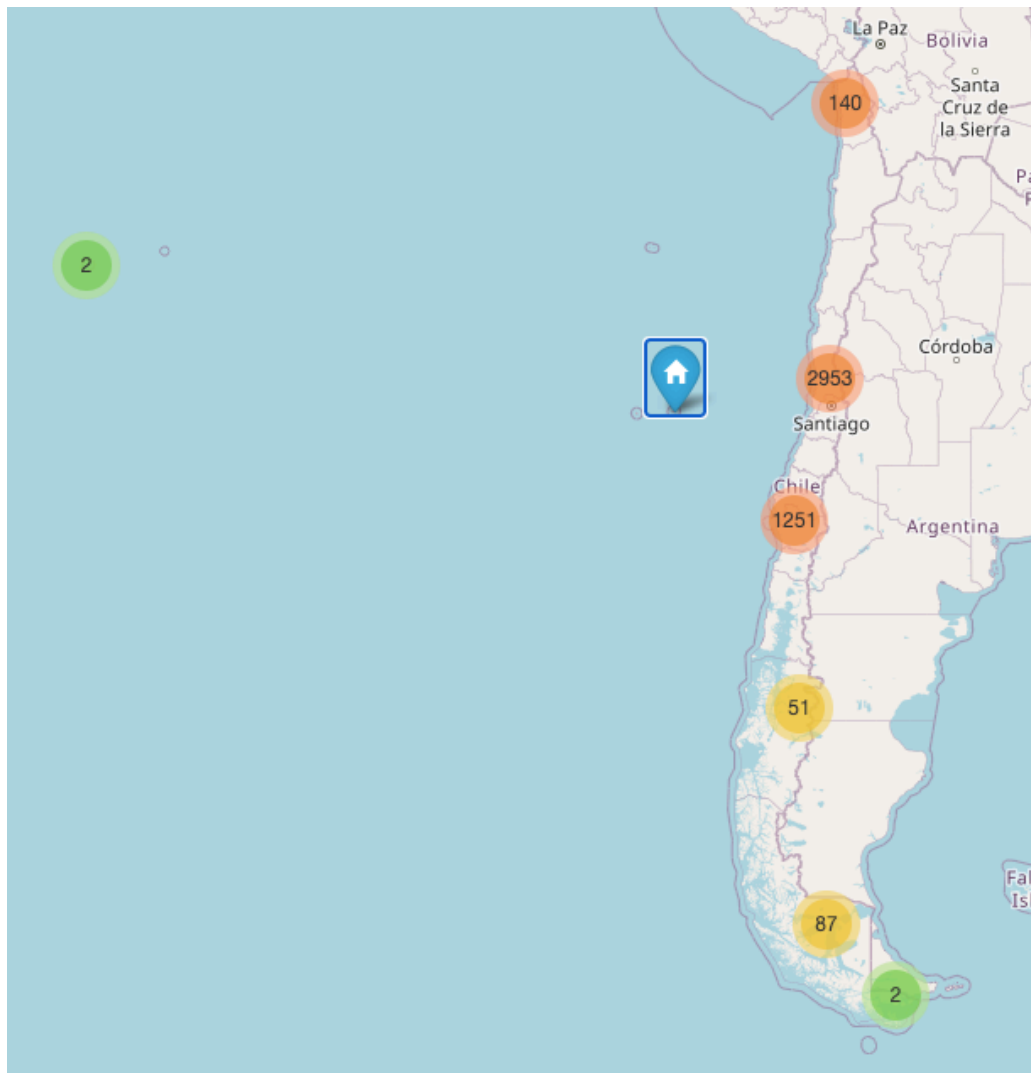


Imagen 3: Distribución de la antenas en reflejadas en el dataset (imagen utilizando librería

Folium)

El otro dato importante es la dirección del Saint George's College, esta es Santa Cruz 5400, Vitacura, 7660103, Región Metropolitana.

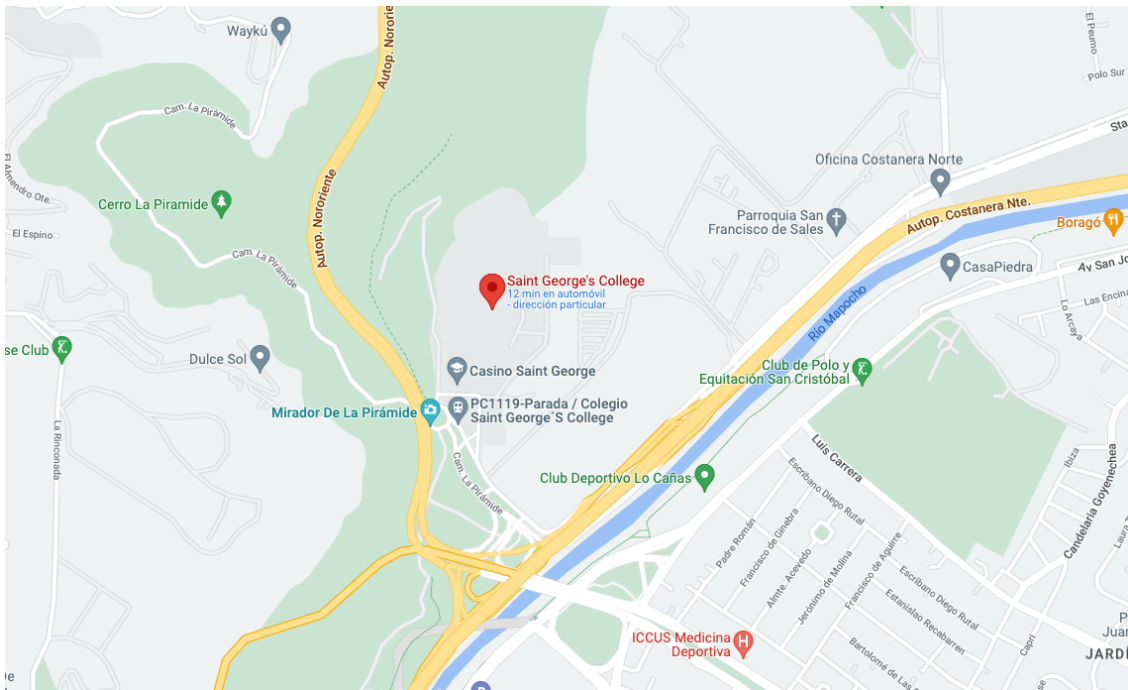


Imagen 4: Ubicación Saint George's College (-33.38398, -70.60001)³, Google Map.

4.2. Metodología

Del punto de vista del procesamiento, debido al alto volumen de datos (más de 250 GB y casi 3.500 millones de registros), trabajamos sobre una máquina virtual en Google Cloud Platform (GCP).

Bajamos la información desde el servidor virtual provisto por la Universidad y a través del uso de la herramienta SCP y la cargamos en un bucket en GCP. El proceso de bajada de datos tomó más de 10 horas.

La máquina virtual configurada en GCP fue la siguiente:

³ Sistema de Coordenadas WGS84 (<https://mappinggis.com/2022/02/diferencias-entre-los-sistemas-de-coordenadas-geograficas-y-proyectadas/>)

| INFORMACIÓN BÁSICA | ESTADO | SUPERVISIÓN | REGISTROS |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|-----------|
| Región | us-west1 (Oregon) | | |
| Zona | us-west1-b | | |
| Entorno ? | TensorFlow Enterprise 2.8 (with LTS and Intel® MKL-DNN/MKL) | | |
| Versión del entorno | M102 | | |
| Tipo de máquina ? | c2d-highmem-8 (8 CPU virtuales, 64 GB de RAM) | | |
| GPU ? | Ninguno | | |
| Disco de arranque | 100 GB disk | | |
| Disco de datos | 1500 GB disk | | |
| Fecha de creación | 5 ene 2023 02:06:31 | | |
| Última modificación | 21 ene 2023 00:29:07 | | |
| Copia de seguridad | Sin especificar | | |
| Subred | default | | |
| Cuenta de servicio | 242322464483-compute@developer.gserviceaccount.com | | |
| Modo de permiso | Service account | | |
| Acceso sudo | Habilitada | | |
| IP externa ? | Habilitada | | |
| Descargas de archivos | Habilitada | | |
| nbconvert | Habilitada | | |
| Colaboración en tiempo real | Inhabilitada | | |
| Estado del notebook | El informe de estado del sistema está habilitado Los informes de métricas personalizadas no están habilitados El agente de Cloud Monitoring no está instalado | | |
| VM protegida | El inicio seguro no está habilitado vTPM está habilitado La supervisión de integridad está habilitada | | |

Imagen 5: Configuración de VM en GCP

Se asignó un disco grande de 1.500 GB ya que inicialmente no se conoce cuántas combinaciones de datos se realizarían en el procesamiento y limpieza. A pesar que la máquina no era pequeña, muchos procesos superaron las 8 horas.

Selección del año 2022 para realizar el análisis

Debido que los datos pertenecen a muestras anuales, es que definimos que la información más reciente es la que representa de mejor forma el modelo que queremos resolver. Es decir, definimos trabajar solo con los datos del año 2022. Apoya también esta decisión el hecho que el año 2021 está fuertemente influenciado por el efecto pandemia, por lo que la muestra podría no ser representativa. En el caso del año 2020, si bien no tiene el efecto pandemia (el gobierno suspende las clases el 15 de Marzo ⁴), si nos presenta un escenario con menor cantidad de autos circulando por la calle.

Adicional a lo anterior, también filtraremos los datos del día 1 de Marzo, dado que las clases se iniciaron el 2 de Marzo.



Imagen 5: Extracto informativo del colegio.

Identificar las antenas cercanas al colegio:

Luego de procesar la información, lo primero fue identificar las antenas con sus respectivas coordenadas. Los datos poseen información de todo Chile, contabilizando 4.487 antenas. Definimos como criterio que las antenas cercanas al colegio están a un radio igual o menor a 1 km. Adicional a eso descartaremos las antenas que no tengan línea

⁴ <https://www.pauta.cl/nacional/cronologia-primer-ano-pandemia-chile>

vista (ubicadas en la comuna de Huechuraba principalmente) ⁵. Para determinar la distancia desde las antenas al colegio utilizamos la formula de distancia Haversine.

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right)$$

Imagen 6: Formula de Haversine

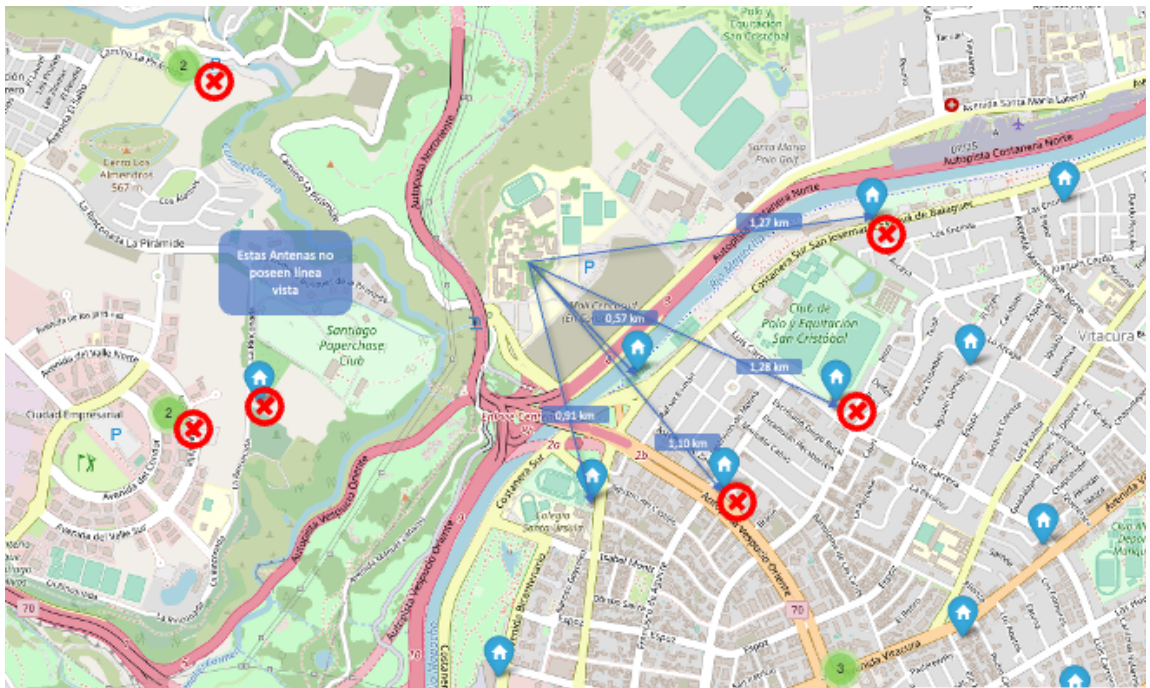


Imagen 7: Grafico de Antenas, que ilustra la ubicación de las 2 antenas que usamos en nuestro análisis.

Determinar los apoderados que transportan al colegio a sus pupilos

A la base total de datos le aplicamos los siguientes filtros:

⁵

https://www.researchgate.net/publication/346128774_Discovering_Daily_POI_Exploitation_Using_LTE_Cell_Tower_Access_Traces_in_Urban_Environment

- Horario de las 7:15 a las 8:30 AM: Ventana de horario que los apoderados trasladan al colegio sus pupilos.
- Antenas cercanas al colegio (2 antenas)
- Días 2 y 3 de Marzo de 2022.

Obtenemos una base de 2.871 dispositivos.

Para poder graficar los datos utilizamos 2 librerías:

- Geopy: Esta librería nos permitió agregar la dirección a los datos teniendo como origen las coordenadas.



- Folium: Nos sirvió para levantar el mapa y graficar, a través de tamaño de la burbuja y mapas de calor. El cual facilitó en gran medida la visualización de los datos.



Determinar hogar de los apoderados

Para determinar el lugar de residencia (hogar) de los apoderados utilizamos el criterio definido por el paper “Evaluation of home detection algorithms on mobile phone data using individual-level ground truth”⁶, el cual determina que el resultado con mayor probabilidad de tratarse de su hogar es usar datos del tipo XDR y midiendo la ubicación

⁶ <https://rdcu.be/c3t5y>

en el intervalo de las 19:00 y hasta las 07:00 hrs y utilizando la información de las 3 antenas con mayor actividad de cada dispositivo.

El procedimiento lo llevamos a cabo en 2 partes:

- Primero: Medir la actividad de todos los device obtenidos en el proceso anterior (2.871). Esto arroja 22.073 registros

In [22]: `count_v2`

Out [22]:

| | device | activity | direcciones | date_format | hora | minutos |
|-------|---------------------------------------------------|----------|---------------------------------------------------|-------------|------|---------|
| 0 | 1fc45336093943b81c13c7b4bc7309219652a3d21b0d74... | 9 | Nutrikelef, 0170, Lo Marcoleta, Santa Luisa, Q... | 03-02-2022 | 20 | 18 |
| 1 | 1fc45336093943b81c13c7b4bc7309219652a3d21b0d74... | 9 | Nutrikelef, 0170, Lo Marcoleta, Santa Luisa, Q... | 03-02-2022 | 21 | 27 |
| 2 | 1fc45336093943b81c13c7b4bc7309219652a3d21b0d74... | 9 | Nutrikelef, 0170, Lo Marcoleta, Santa Luisa, Q... | 03-02-2022 | 23 | 29 |
| 3 | 1fc45336093943b81c13c7b4bc7309219652a3d21b0d74... | 9 | Nutrikelef, 0170, Lo Marcoleta, Santa Luisa, Q... | 03-02-2022 | 0 | 45 |
| 4 | 1fc45336093943b81c13c7b4bc7309219652a3d21b0d74... | 9 | Nutrikelef, 0170, Lo Marcoleta, Santa Luisa, Q... | 03-02-2022 | 1 | 45 |
| ... | ... | ... | ... | ... | ... | ... |
| 22068 | ecd8c8a313583b918c84770b56f49792cbc689ca9a7304... | 1 | SP Digital, 356, Padre Mariano, Providencia, P... | 03-03-2022 | 6 | 43 |
| 22069 | b321c68a4f36d67e2b3263e5783f1a181eaf218d73151... | 1 | Estadio Instituto Nacional, Avenida Nueva Cost... | 03-02-2022 | 7 | 59 |
| 22070 | 1ea44dde1130cceb5c975df4afc1bf0329150e329b9020... | 1 | Estadio Instituto Nacional, Avenida Nueva Cost... | 03-02-2022 | 7 | 38 |
| 22071 | dfe05846aa11cc834f3b38e88a2d933477351b2ef78c5b... | 1 | Estadio Instituto Nacional, Avenida Nueva Cost... | 03-02-2022 | 7 | 9 |
| 22072 | c94b5daa02cddcbf88e42e0fd5f6593aa1128f0e467ba6... | 1 | 2281, Espoz, Vitacura, Provincia de Santiago, ... | 03-02-2022 | 7 | 26 |

22073 rows x 6 columns

- Segundo: Filtrar las 3 antenas con mayor actividad por dispositivo.

In [48]: `resultado`

Out [48]:

| | device | closest | second_closest | third_closest |
|------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|
| 0 | 0027775f0301d8e4f3b2033740f55651f8a04243e7a7e... | Estadio Instituto Nacional, Avenida Nueva Cost... | Jardin La Dehesa, Lo Barnechea, Provincia de S... | 19, San Crescente, El Golf, Las Condes, Provih... |
| 1 | 00345e51bf20363e25c3b471cda79365592feb5535b22a... | Montserrat, Salvador Sanfuentes, Los Quillayes... | 2356, Roque Esteban Scarpa, Villa Universidad ... | 75, Cruz del Sur, Las Condes, Provincia de San... |
| 2 | 0048b13311bce2730d3c294c2f1598c23a52443a5a3d... | Nueva Carvajal / San Francisco, Avenida San Fr... | 4434, Nevería, Las Condes, Provincia de Santia... | Estadio Instituto Nacional, Avenida Nueva Cost... |
| 3 | 00750a313ae29d538ca8ae766891d0681a9754631153a7... | 4434, Nevería, Las Condes, Provincia de Santia... | 622, Soria, Las Condes, Provincia de Santiago,... | 75, Cruz del Sur, Las Condes, Provincia de San... |
| 4 | 00bd4c1effbe715c90b77089df0be72a5bfa99081f0356... | 585, Marín, Providencia, Provincia de Santiago... | Gimnasio Pacific, Avenida Rancagua, Providenci... | 7489, Avenida Presidente Kennedy, Las Condes, ... |
| ... | ... | ... | ... | ... |
| 1227 | ff2d54911ee2a3e93671f15e64aaf33318f886ab2d983e... | Club Deportivo Lo Cañas, 5000, Costanera Sur S... | Estadio Instituto Nacional, Avenida Nueva Cost... | 4469, Espoz, Vitacura, Provincia de Santiago, ... |
| 1228 | ff471e2c5e85e23835729dda3ed830e93395a1d21bceb... | 2219, Avenida Pedro de Valdivia, Providencia, ... | 3592, Avenida Vitacura, Vitacura, Provincia de... | Club de Polo y Equitación San Cristóbal, Casa ... |
| 1229 | ff7a626762aac4ee2ced76c2973c33c21ba2d2236cc514... | 3592, Avenida Vitacura, Vitacura, Provincia de... | Camino La Paz - Collico Alto, Collico Alto, Lo... | Estadio Instituto Nacional, Avenida Nueva Cost... |
| 1230 | ff9feec4964b1d2e47a2d0596181f0b7bfc6c336fc6... | Club Deportivo Lo Cañas, 5000, Costanera Sur S... | 8594, Avenida Vitacura, El Dorado, Vitacura, P... | Estadio Instituto Nacional, Avenida Nueva Cost... |
| 1231 | fff7859bf04fcedfd1097bc7277e781468ac8bb0c46b6... | 115, Ernesto Pinto Lagarrigue, Bellavista de S... | 265, Avenida Vicuña Mackenna, Barrio Lastarria... | 818, Santo Domingo, Barrio Lastarria, Santiago... |

1232 rows x 4 columns

Dado que no todos los device poseen actividad en 3 antenas, se nos reduce el ámbito de análisis de 2.871 a 1.232 dispositivos.

Para ilustrar las 3 antenas con mayor actividad utilizamos la librería Siuba.

Siuba

Dado que en el origen hacemos una estimación de los datos que constituyen el subconjunto de apoderados, se define un último criterio para acotar los equipos que son del colegio. Este criterio es que su hogar debe estar a una distancia inferior a 10 km del colegio.

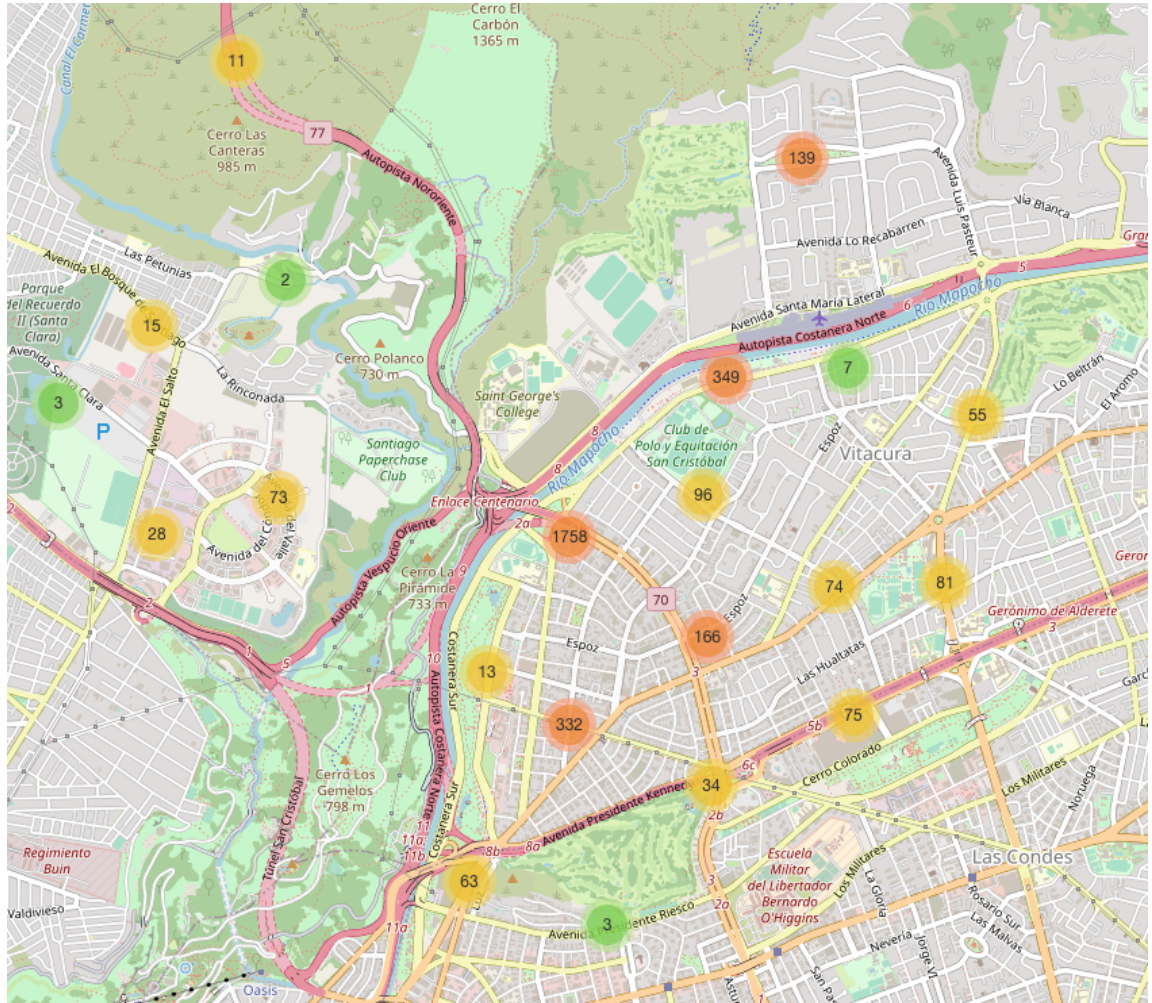


Imagen 8: Imagen en Folium con el domicilio de los apoderados con filtro “distancia menor a 10km del colegio”

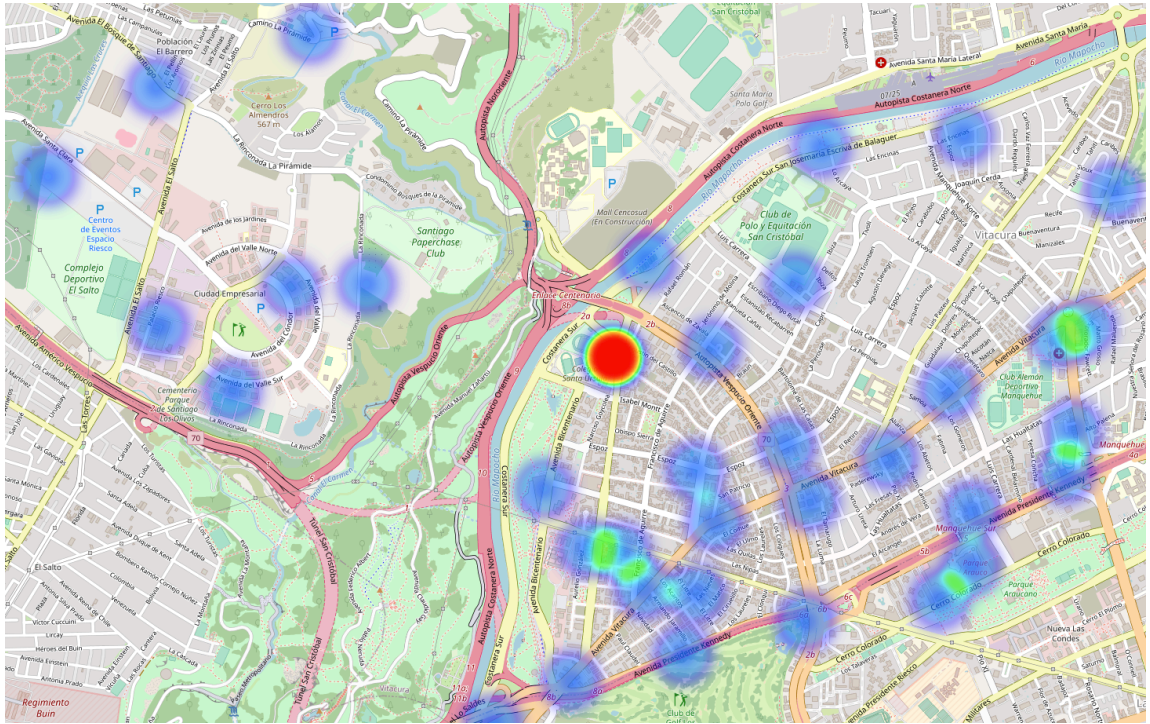


Imagen 9: Heatmap con domicilio de los apoderados con filtro distancia 10km.

Cruce de información desde el hogar hasta el colegio sobre el transito normal de la mañana.

Este cruce de información busca establecer las zonas para la ubicación de buses de acercamiento.

Tratamiento de los Datos

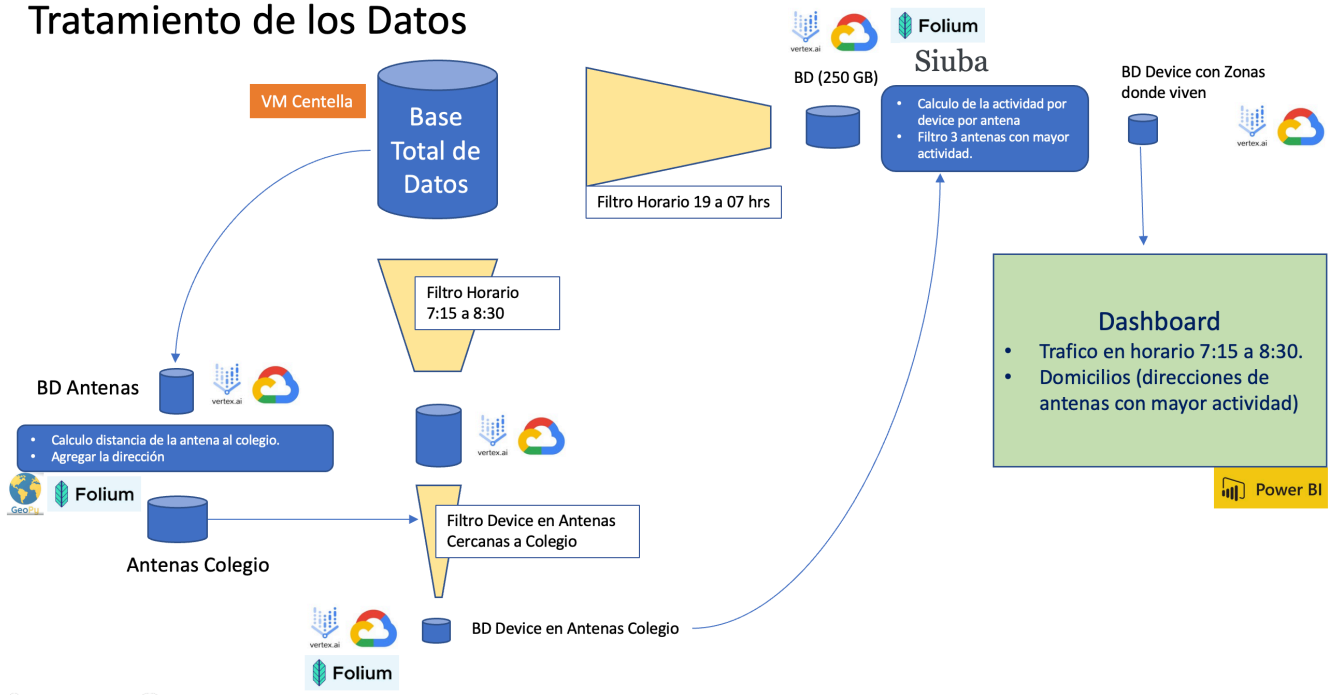


Imagen 10: Ilustración de la limpieza de datos.

5. Resultados

Del capítulo anterior ya ilustramos la estimación de zonas de apoderados del colegio, obteniendo una fuerte concentración en la misma comuna de Vitacura.

Ahora graficamos el escenario de congestión en la mañana.

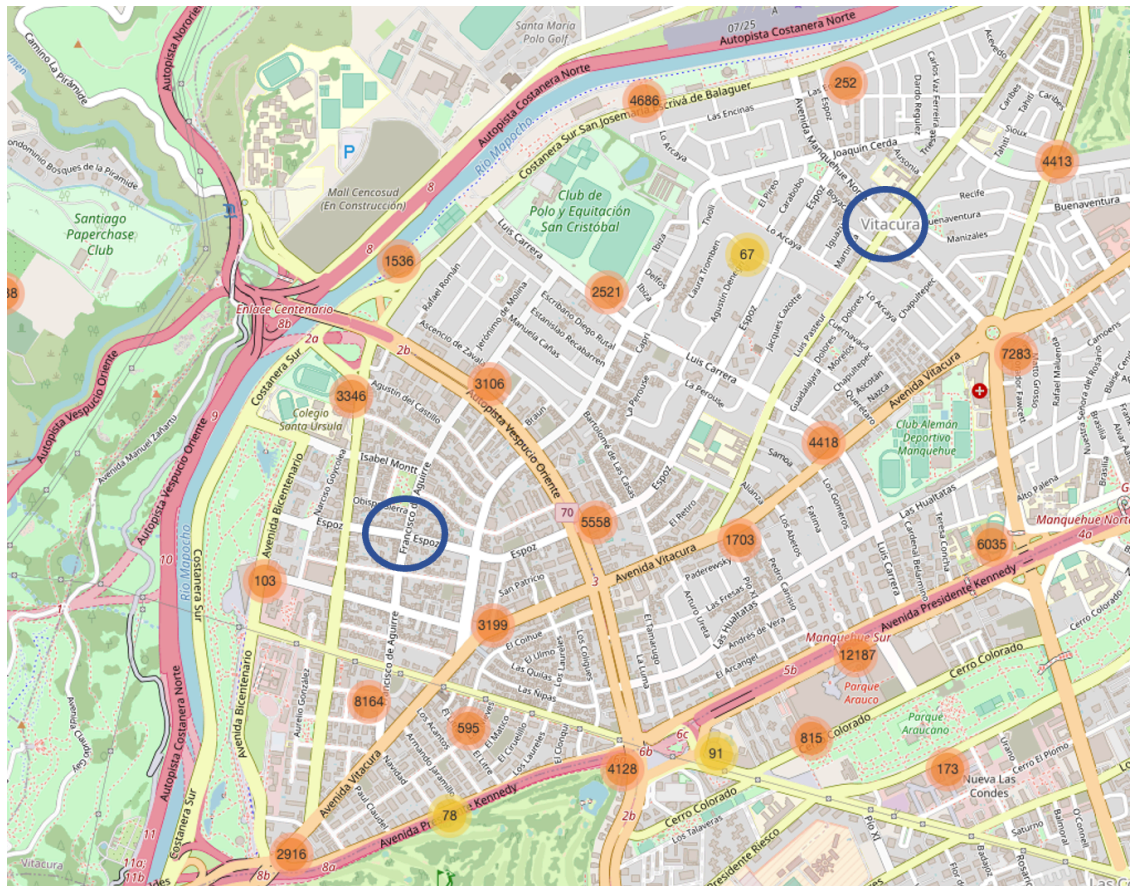


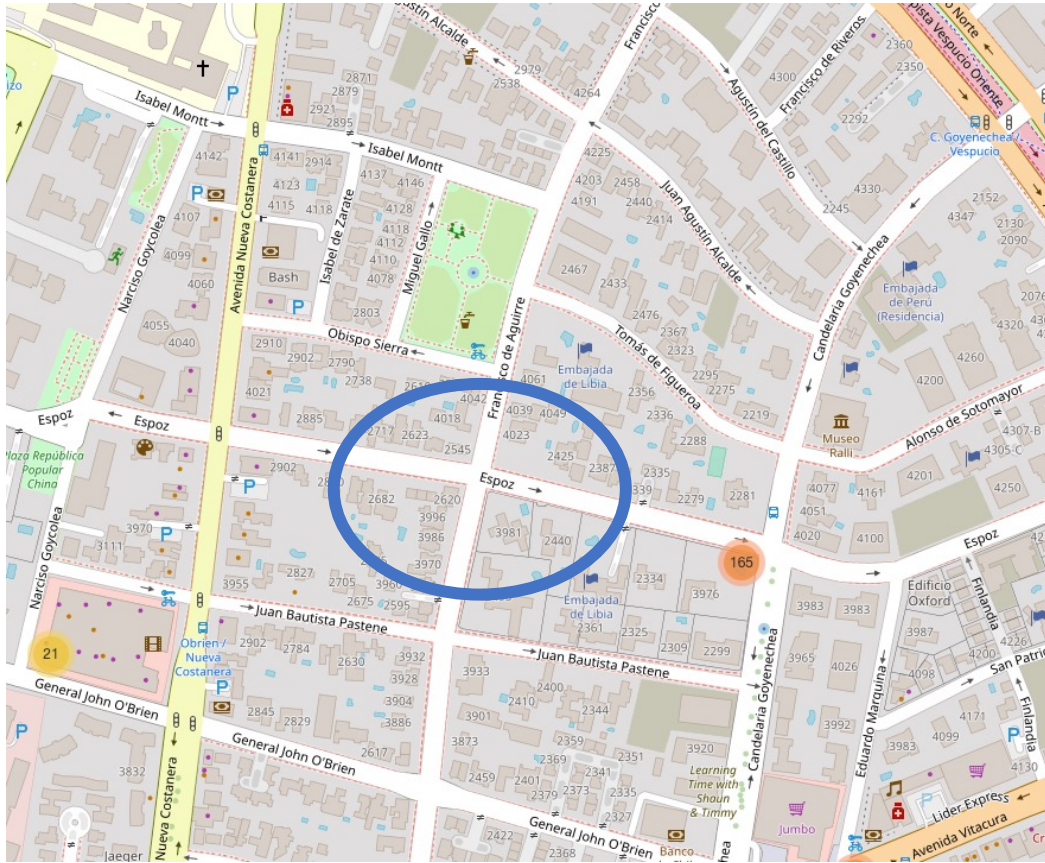
Imagen 11: Concentración en horario peak y recomendación para buses de acercamiento.



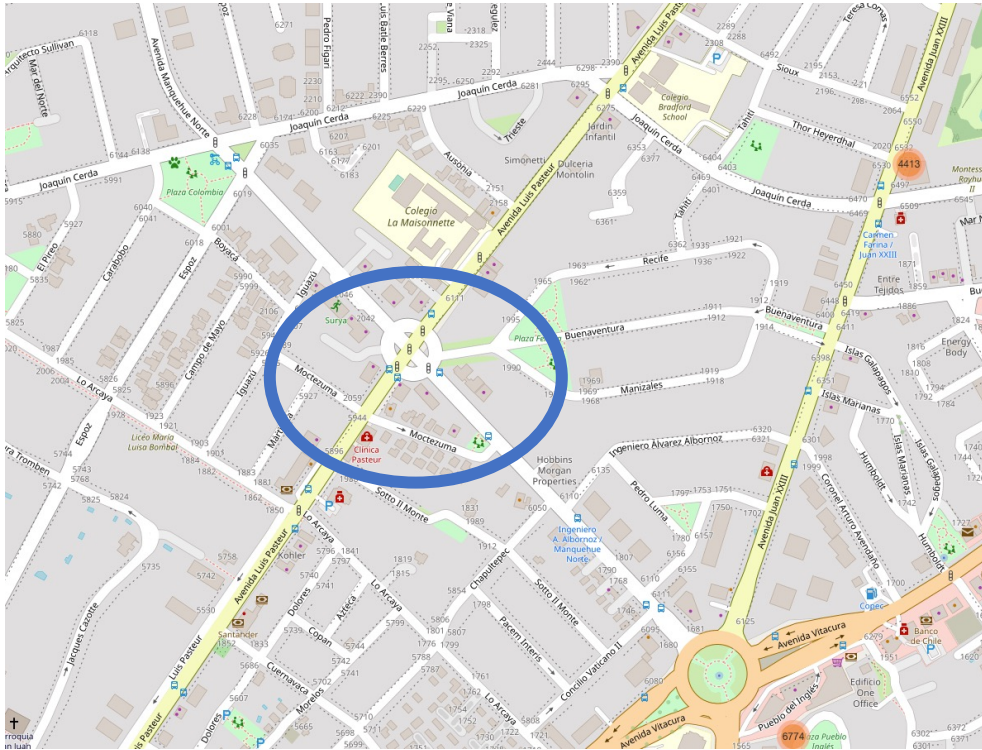
Imagen 12: Heatmap Concentración en horario peak

Recomendación de Ubicación de Buses de Acercamiento

Dada la información anterior (ubicación de domicilios de apoderados) y la concentración de tráfico en horario peak, nuestra recomendación es que los puntos de inicio de los buses de acercamiento se encuentre en:

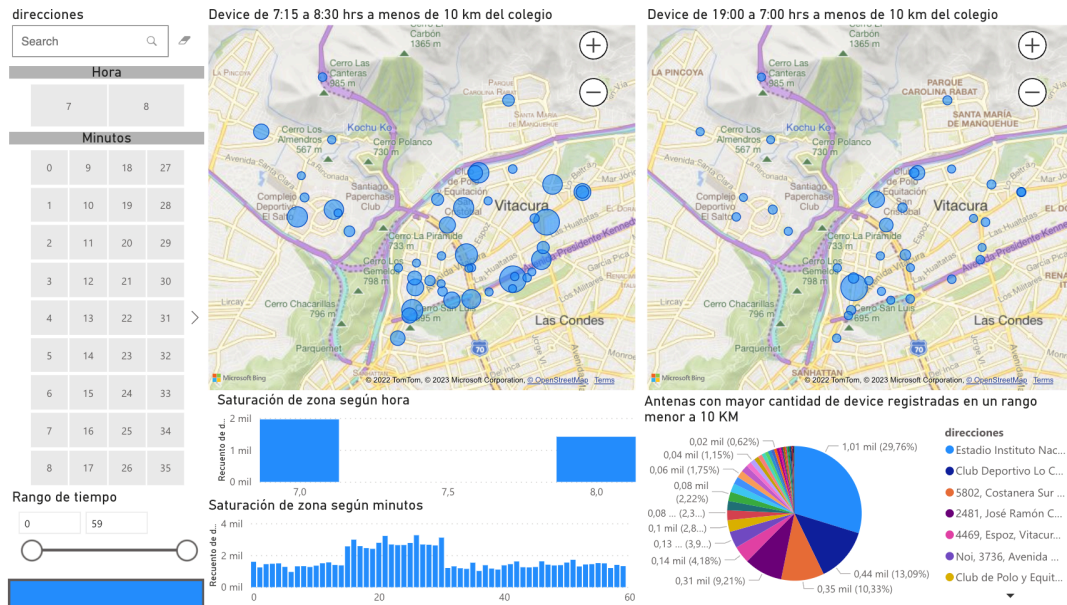


- Calles cercanas a la intersección de Espoz con Francisco de Aguirre
(coordenadas: -33.397179,-70.595271)



- Calles cercanas a la intersección de Manquehue Norte con Luis Pasteur (coordenadas: -33.387101,-70.576440)

DashBoard en Power BI



6. Conclusiones

6.1. Limitaciones

Una de las principales limitaciones del trabajo es hacer una estimación (a través de supuestos) de los apoderados pertenecientes al colegio con su percepción óptima de tiempo de traslado. De obtener estos datos y luego cruzarlos con la información de tránsito del horario definido en la mañana (07:15 – 08:30 hrs), podría arrojar un resultado mucho más optimizado del lugar de localización de los buses de acercamiento.

Dado que el colegio se encuentra en la falda del cerro La Pirámide, existe una baja concentración de antenas móviles. Este detalle implica que debemos acotar a pocas antenas (2) la certeza de los apoderados del colegio. Se podría tratar de un individuo dejando alguien en el colegio o simplemente pasando relativamente cerca (aproximadamente a 1 km).

Profundizando más en la conclusión del problema y futuras aplicaciones es necesario mencionar que si bien se trabajó en base a la fórmula de distancia de Haversine, la cual determina la distancia entre dos puntos de una esfera dadas las longitudes y latitudes, nos vimos limitados por el mismo criterio utilizado para la determinación de hogares de los dispositivos registrados, lo que está determinado por el tiempo de muestra que obtuvimos, los cuales son solo dos días de inicio de clases después de la pandemia en el año 2022. El colegio inicio clases el día 2 de marzo, con un horario reducido las primeras dos semanas para facilitar el proceso de adaptación, por lo cual solo obtuvimos 1.232 dispositivos como punto de origen con un índice de confianza aceptable según la literatura consultada y nombrada con anterioridad en el documento, esto nos entrega información muy relevante para conocer los hogares de las personas que transitan la

zona, pero nos limita por la cantidad de días y de dispositivos registrados para generar una data más robusta y obtener información estadística relevante y por consiguiente la aplicación de alguna red neuronal artificial que nos diera un valor optimo según un entrenamiento de un modelo.

6.2. Trabajo Futuro

Contar con la información personal de los apoderados, cruzándolo con los datos de congestión de las antenas, podríamos establecer en forma más precisa la ubicación de los buses de acercamiento, lo cual podríamos extrapolar a otros colegios, universidades o cualquier entidad o institución a la cual se tenga acceso a data de origen de traslado.

Con la disposición de una muestra típica con un mayor número de datos (más días de muestra) y adhiriendo variantes logísticas al modelo, podríamos generar un prototipo con un alcance distinto, de orden logístico, con la dirección de tránsito de las calles, rutas de autobuses de transporte público y sus frecuencias para generar un recorrido optimo (no solo con buses de acercamiento) y no solo el punto de origen optimo según el análisis descrito en el documento.

Adicionalmente, a través de una API como la de Google Maps, nos podría entregar data de otros orígenes además de lo obtenido y determinar un producto más certero con una muestra que se realice en días hábiles sin el efecto de distorsión de la pandemia del COVID 19, generando incluso información relevante para centros comerciales, determinación del impacto ambiental de construcción de conjuntos habitacionales y clusterización de datos asociados a otras fuentes como informes de planificación municipal, densidad de población e información censal, junto con la proyección en el tiempo dada por la tasa de crecimiento poblacional.

7. Bibliografía

- [1] Han, Sumin & Park, Kinam & Lee, Dongman. (2020). Discovering Daily POI Exploitation Using LTE Cell Tower Access Traces in Urban Environment. 10.1007/978-3-030-60975-7_7.
- [2] Akella, Mohan & Delmelle, Eric & Batta, Rajan & Rogerson, Peter & Blatt, Alan. (2010). Adaptive Cell Tower Location Using Geostatistics. Geographical Analysis. 42. 10.1111/j.1538-4632.2010.00791.x.
- [3] Oosterlinck, Dieter & Baecke, Philippe & Benoit, Dries. (2020). Home location prediction with telecom data: Benchmarking heuristics with a predictive modelling approach.. Expert Systems with Applications. 170. 114507. 10.1016/j.eswa.2020.114507.
- [4] Pappalardo, L., Ferres, L., Sacasa, M. et al. Evaluation of home detection algorithms on mobile phone data using individual-level ground truth. EPJ Data Sci. 10, 29 (2021). <https://doi.org/10.1140/epjds/s13688-021-00284-9>
- [5] Nicola Perra, Non-pharmaceutical interventions during the COVID-19 pandemic: a review, <https://doi.org/10.1016/j.physrep.2021.02.001>.