

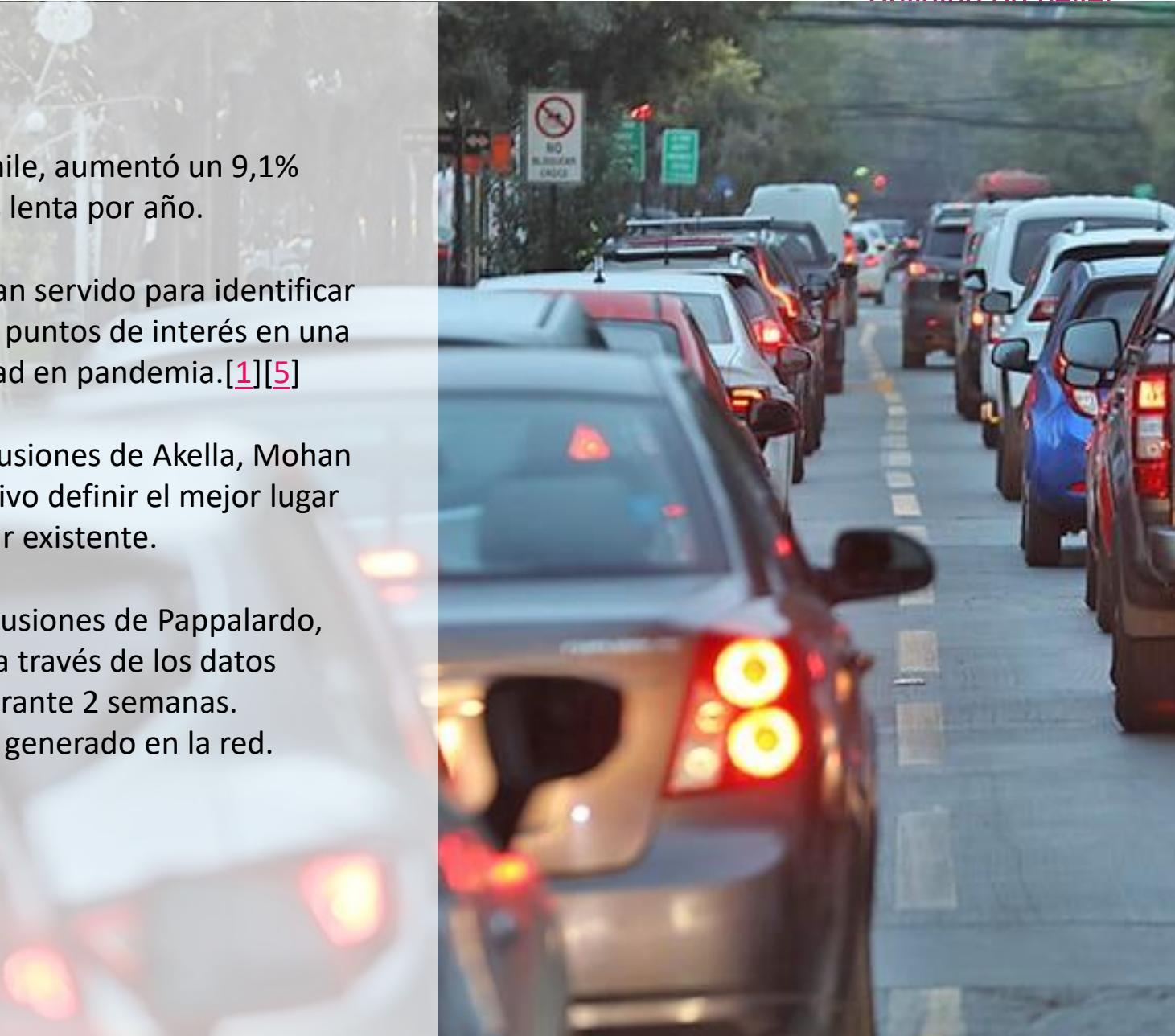
Aplicación de Geodatos y Machine Learning

Nicolas Klein Godoy
Ingeniero Comercial
Master en Data Science
Project Manager en TCS



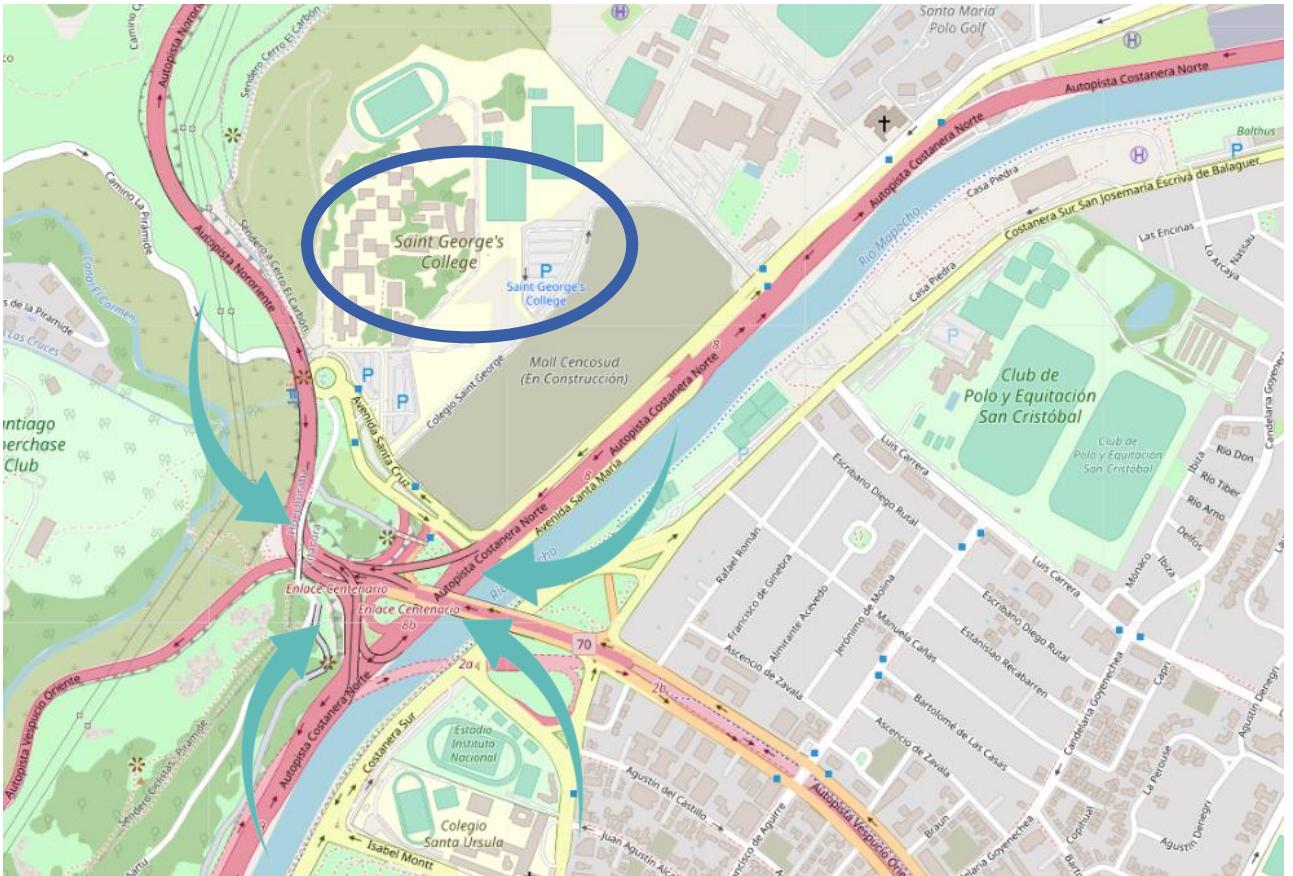
Introducción

- Según datos del INE, el parque automotriz en Chile, aumentó un 9,1% entre 2020 a 2021. Volviéndose 15 minutos mas lenta por año.
- Los datos generados por las antenas celulares han servido para identificar comportamiento de los usuarios, determinando puntos de interés en una ciudad, encuesta matriz origen-destino, movilidad en pandemia.[[1](#)][[5](#)]
- Para determinar las antenas utilizamos las conclusiones de Akella, Mohan [[2](#)] cuyo trabajo de investigación tiene por objetivo definir el mejor lugar para instalar una torre adicional a una red celular existente.
- Para determinar el domicilio utilizamos las conclusiones de 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.



Descripción del Problema

- El Saint George's College posee muy difíciles accesos, ya que muy cerca confluyen 4 autopistas.
- Esta situación se agudiza en horario peak de la mañana, al momento de dejar a los alumnos.
- Una forma abordar el problema es tener buses de acercamiento.
- De acuerdo a la ubicación de los hogares de los alumnos, se busca definir lugares de inicio del transporte.



A través de los datos capturados por antenas móviles (XDR), buscamos optimizar la ubicación de los buses de acercamiento



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.



Trabajos Relacionados

Akella, Mohan [2],

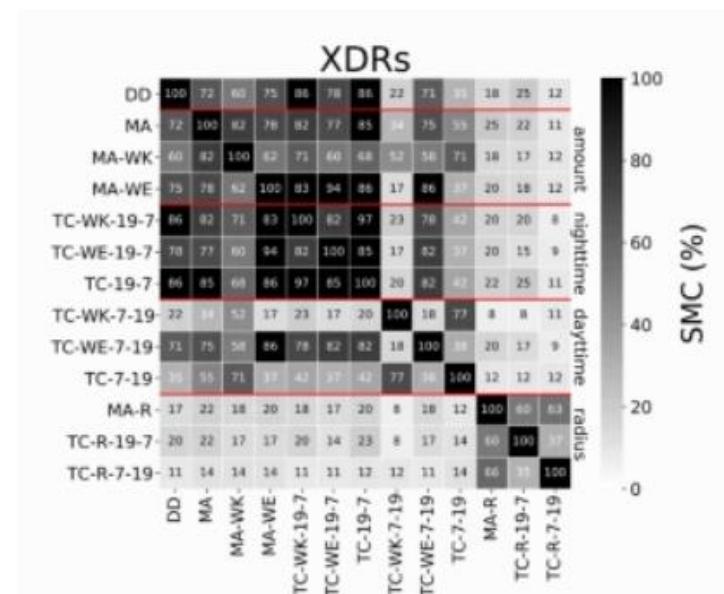
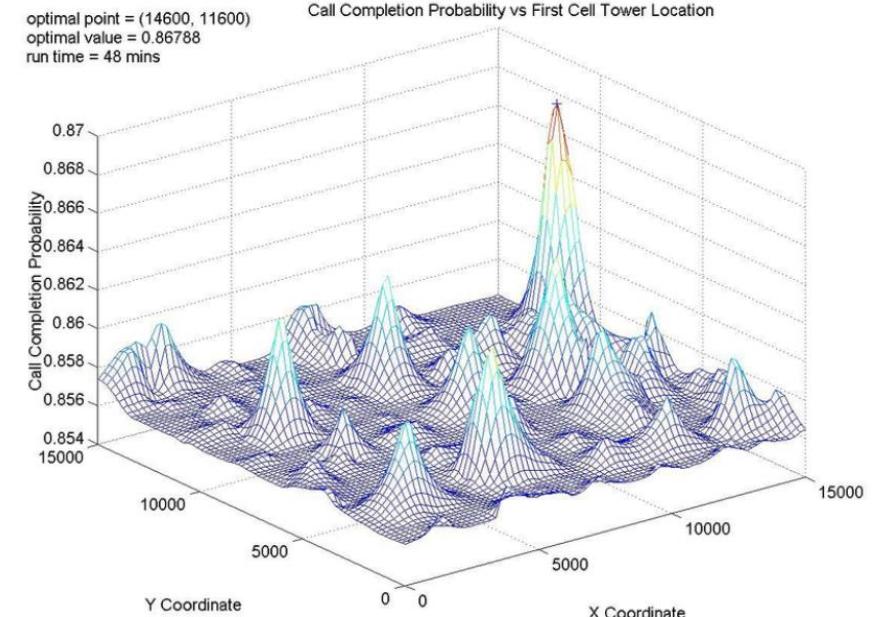
Objetivo calcular la posición optima de una nueva antena celular dentro de una red ya desplegada.

Estas definiciones proporcionaron elementos suficientes para filtrar aquellas antenas que tomaban la conectividad de los apoderados a la hora de la entrada al colegio.

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.



Descripción de los Datos

Los datos corresponden a la información de todas las conexiones XDR (eXtended Detail Records) que se realizan entre los equipos móviles y las celdas o antenas móviles de una compañía de telecomunicaciones de todo Chile.

Marzo 2020						
	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
1	24	25	26	27	28	29
2	1	2	3	4	5	6
3	7	8				
4	9	10	11	12	13	14
5	15	16	17	18	19	20
6	21	22				
7	23	24	25	26	27	28
8	29	30	31	1	2	3
9	4	5	6	7	8	9
10	10	11	12	13	14	15
11	16	17	18	19	20	21
12	23	24	25	26	27	28
13	30	31	1	2	3	4
14	31	1	2	3	4	5

Marzo 2021						
	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
1	1	2	3	4	5	6
2	7	8	9	10	11	12
3	13	14	15	16	17	18
4	19	20	21	22	23	24
5	25	26	27	28	29	30
6	31	1	2	3	4	5
7	6	7	8	9	10	11
8	12	13	14	15	16	17
9	18	19	20	21	22	23
10	24	25	26	27	28	29
11	30	31	1	2	3	4
12	5	6	7	8	9	10
13	17	18	19	20	21	22
14	23	24	25	26	27	28
15	29	30	31	1	2	3

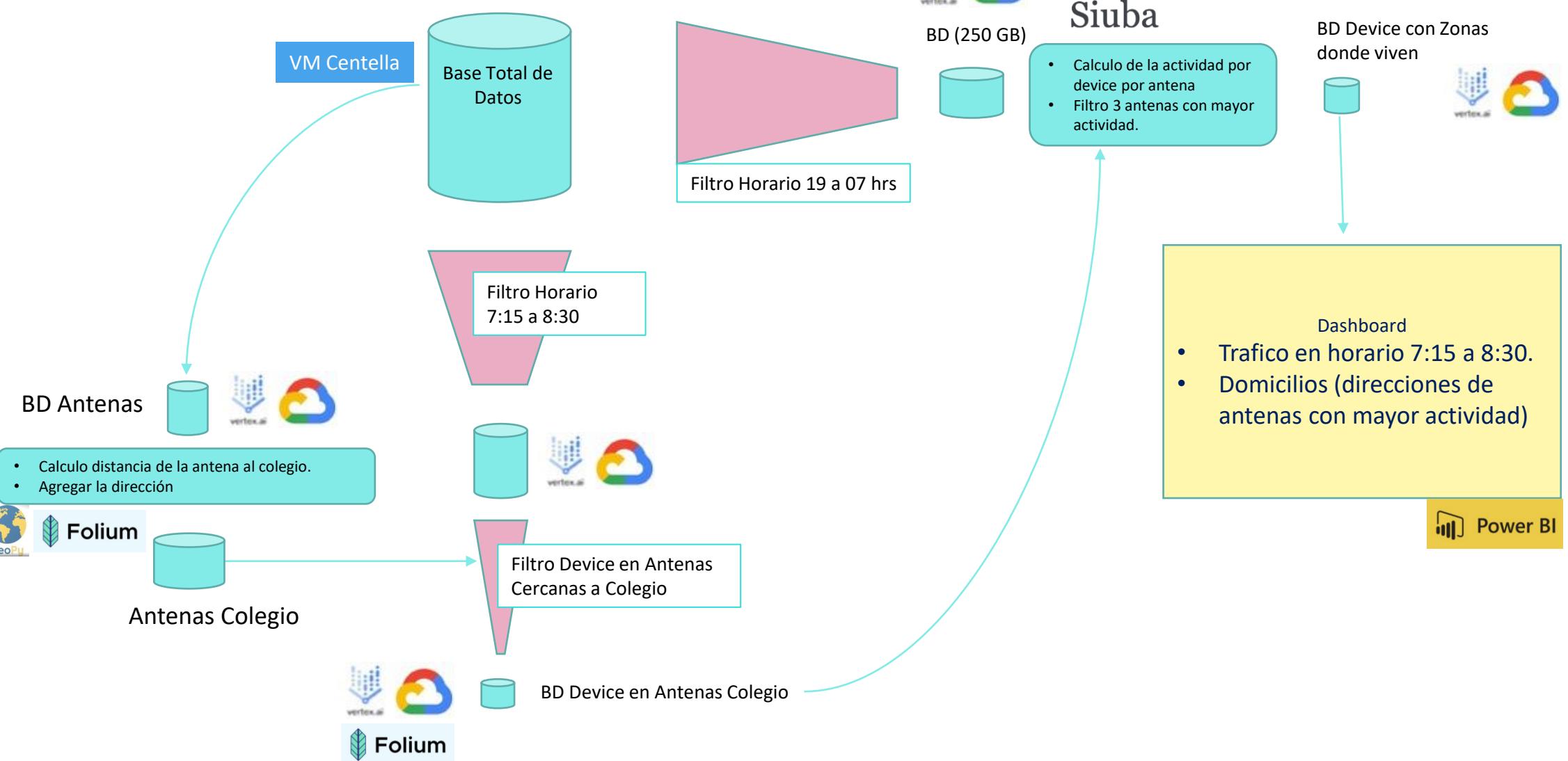
Marzo 2022						
	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
1	28	1	2	3	4	5
2	6	7	8	9	10	11
3	12	13	14	15	16	17
4	18	19	20	21	22	23
5	25	26	27	28	29	30
6	31	1	2	3	4	5
7	28	29	30	31	1	2
8	3	4	5	6	7	8
9	12	13	14	15	16	17
10	18	19	20	21	22	23
11	25	26	27	28	29	30
12	31	1	2	3	4	5

- 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):

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



Tratamiento de los Datos



Metodología

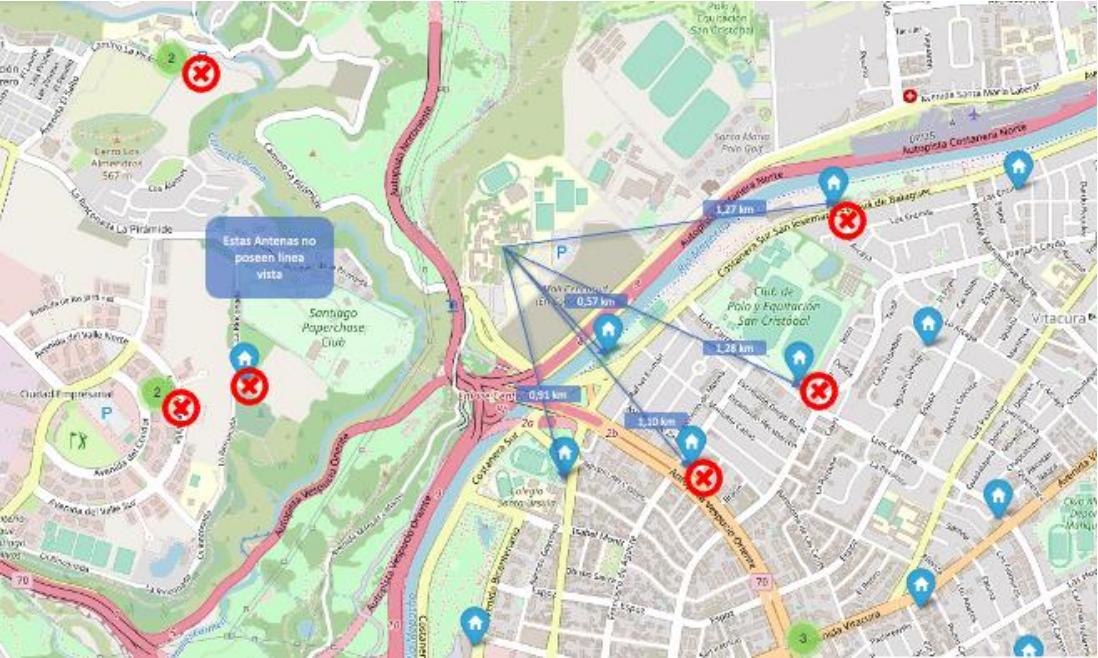
Información de apoderados

1. Fecha y horario de la muestra:

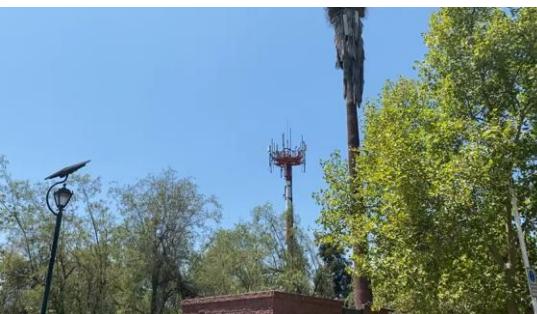
- ✓ Usamos los datos del año 2022, de los días 2 y 3 de Marzo.
- ✓ No usamos 2020 ya que los flujos de transito podrían ser menores (2 año es un 30% menos de tráfico).
- ✓ No usamos 2021 por efecto pandemia (menor transito)
- ✓ No usamos el 1 de marzo ya que las clases comenzaron el 2 de marzo.
- ✓ Filtro de horario: de las 07:15 a las 08:30 AM

2. Toma de la muestra:

- ✓ Calculamos la distancia de todas las antenas al colegio.
- ✓ Filtramos por aquellas que estaban a menos de 1km y que tuvieran línea vista (filtrado todo lo que estaba en Huechuraba) (*)



Resultado: 2.871 dispositivos



Video Antena 1

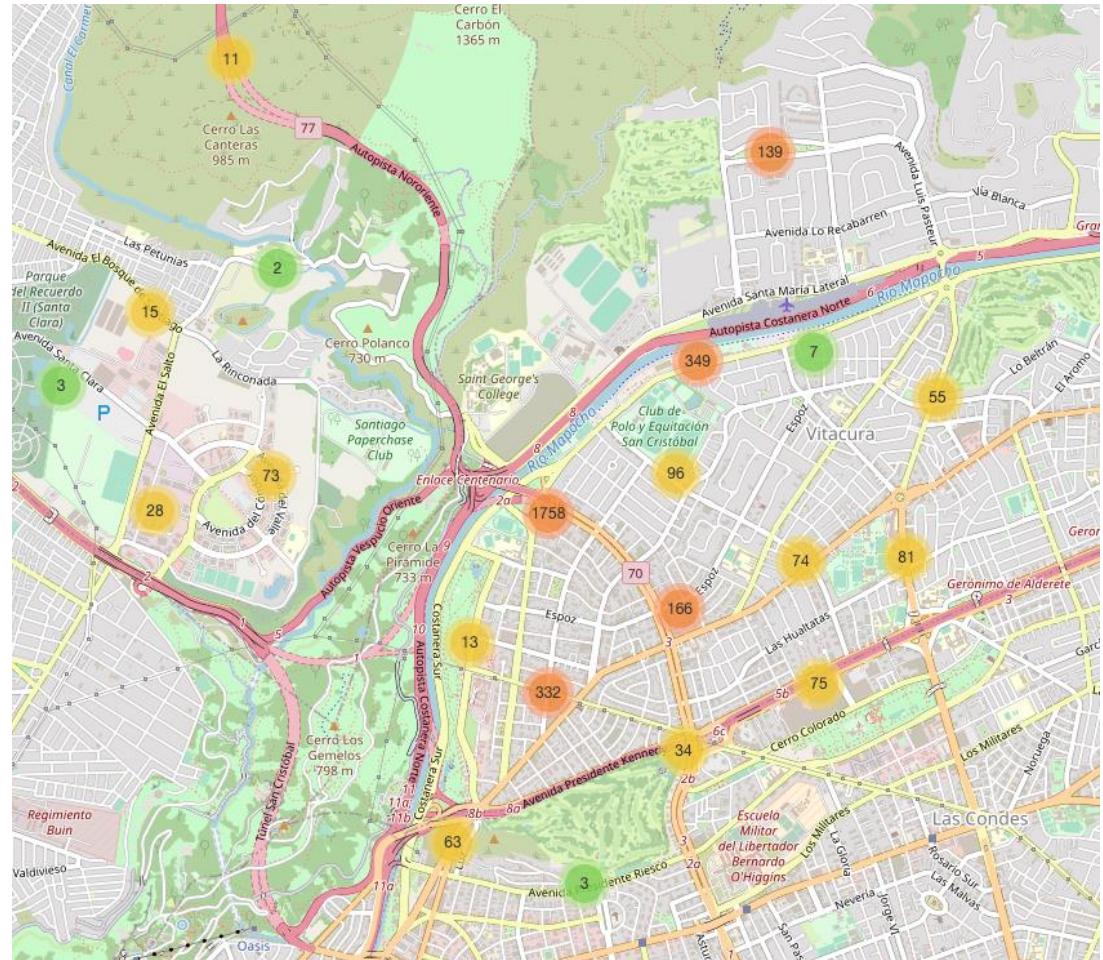
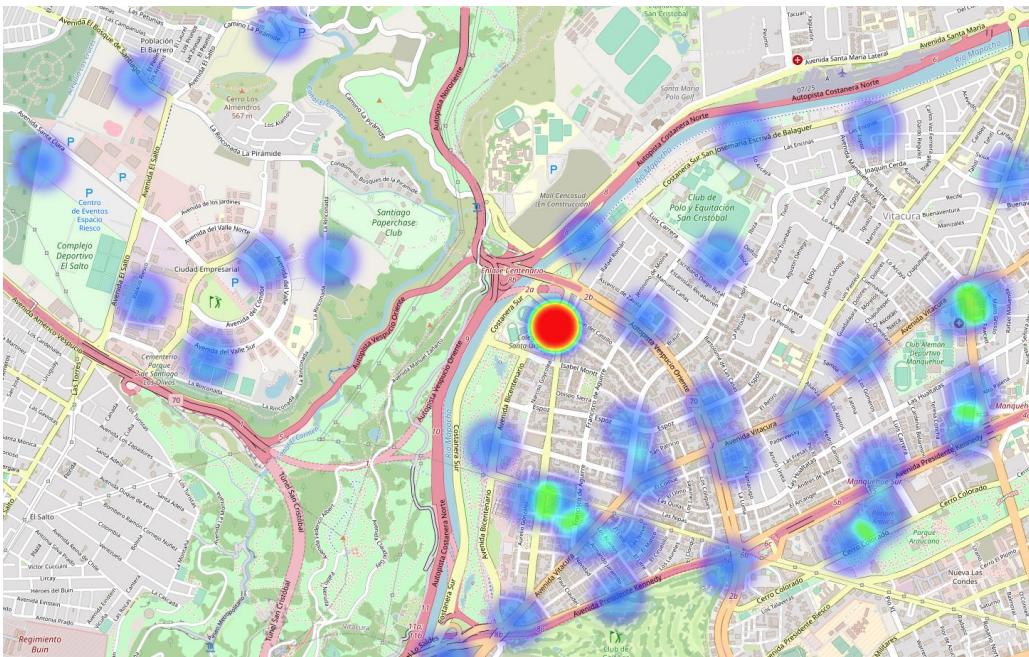


Video Antena 2

Metodología

Lugar de residencia de apoderados

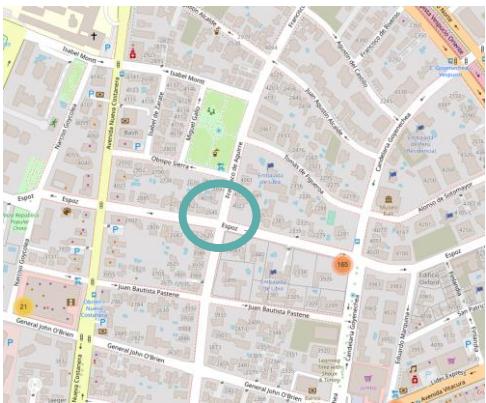
- Registros XDR
- Medidos en horario de las 19:00 a las 07:00 hrs
- Registro de días hábiles (2 y 3 de marzo de 2022)
- Ilustrando las 3 antenas que poseen mayor actividad



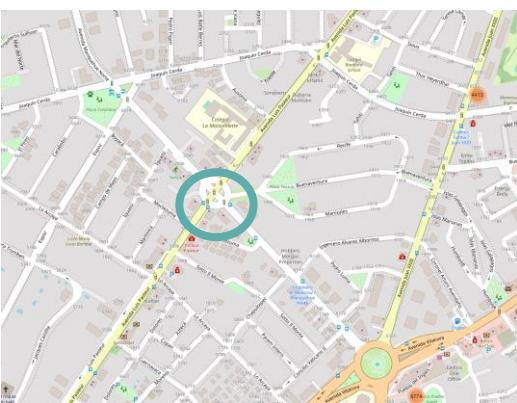
Resultados

Recomendación de Ubicación de Buses de Acercamiento

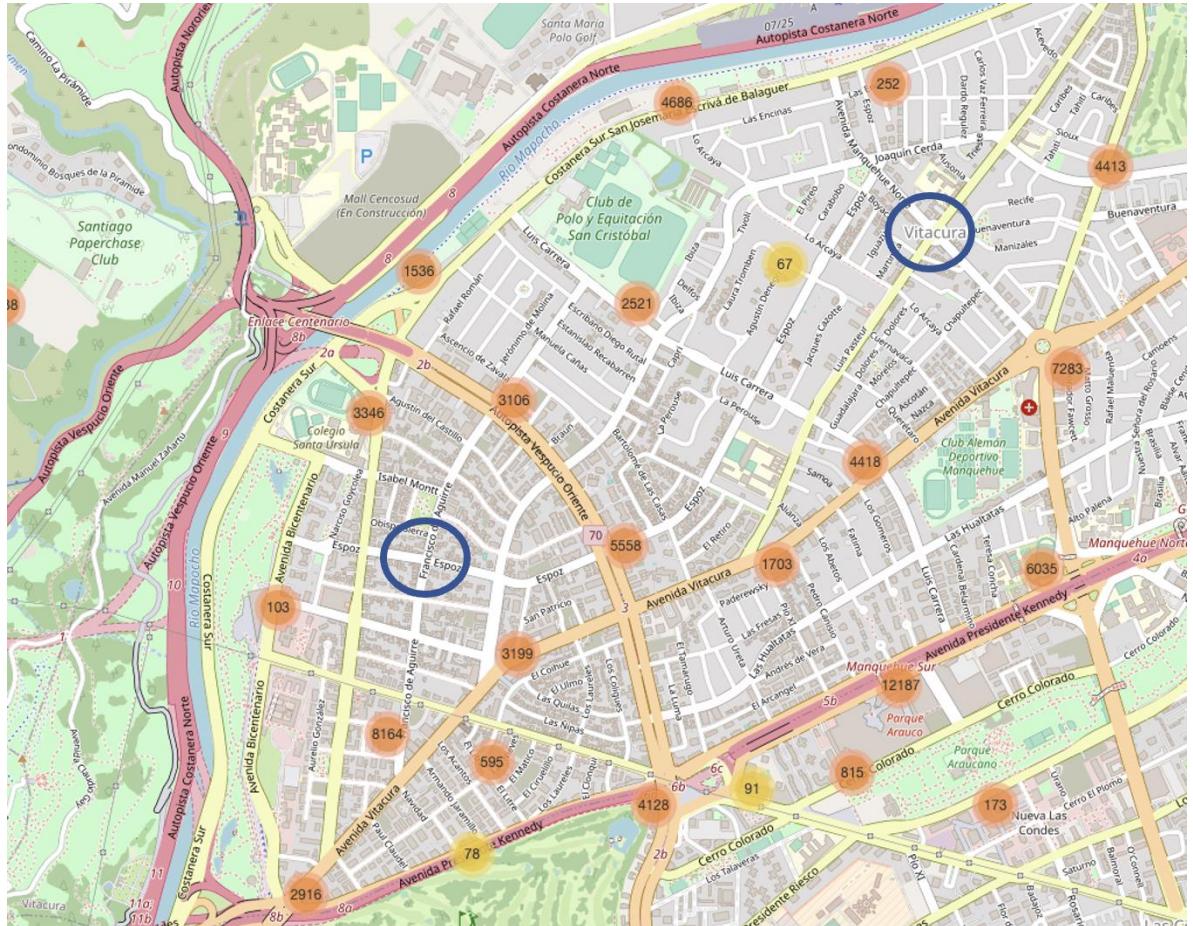
Dada la información anterior (ubicación de domicilios de apoderados) y la concentración de tráfico en horario peak, nuestra recomendación es que el punto en inicio del bus de acercamiento se encuentre :



Calles cercanas a la intersección de Espez 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)



Resultados

Dashboard

direcciones

Hora

7	8
---	---

Minutos

0	9	18	27
1	10	19	28
2	11	20	29
3	12	21	30
4	13	22	31
5	14	23	32
6	15	24	33
7	16	25	34
8	17	26	35

Rango de tiempo

0	59
---	----





- direcciones**
- Estadio Instituto Nac...
 - Club Deportivo Lo C...
 - 5802, Costanera Sur ...
 - 2481, José Ramón C...
 - 4469, Espoz, Vitacur...
 - Noi, 3736, Avenida ...
 - Club de Polo y Equit...

Conclusiones

La georreferenciación utilizando los datos móviles abre múltiples escenarios de análisis.

Es muy importante tener una alta concentración de antenas para obtener conclusiones más ajustadas.

Limitaciones

1. Baja Concentración de Antenas: 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 buscar registros en antenas más alejadas del colegio, perdiendo la certeza respecto si se trata de un individuo dejando alguien en el colegio o simplemente pasando relativamente cerca (1 km)

2. Estimación de direcciones de los apoderados: Tener esta información y luego cruzarla con la información de tránsito del horario peack de la mañana, podría arrojarnos un resultado mucho más optimizado del lugar de localización de los buses de acercamiento.

Propuesta con Aeropuerto de Santiago

MOP anuncia que licitará proyectos por 625 millones de dólares para mejorar aeropuertos

Publicado por [Diego Vera](#)

La información es de [Joaquín Morales](#)

Martes 28 febrero de 2023 | 20:53

[Leer más tarde](#)



Relacionados

- MOP da a conocer plan en Aeropuerto de Santiago ante salida de 200 mil pasajeros en Semana Santa



MOP anuncia que licitará proyectos por 625 millones de dólares para mejorar aeropuertos

Publicado por [Diego Vera](#)
La información es de [Joaquín Morales](#)

Martes 28 febrero de 2023 | 20:53



Ministerio de Obras Públicas

PROYECTOS DE RED AEROPORTUARIA DE LA CARTERA DE CONCESIONES 2023

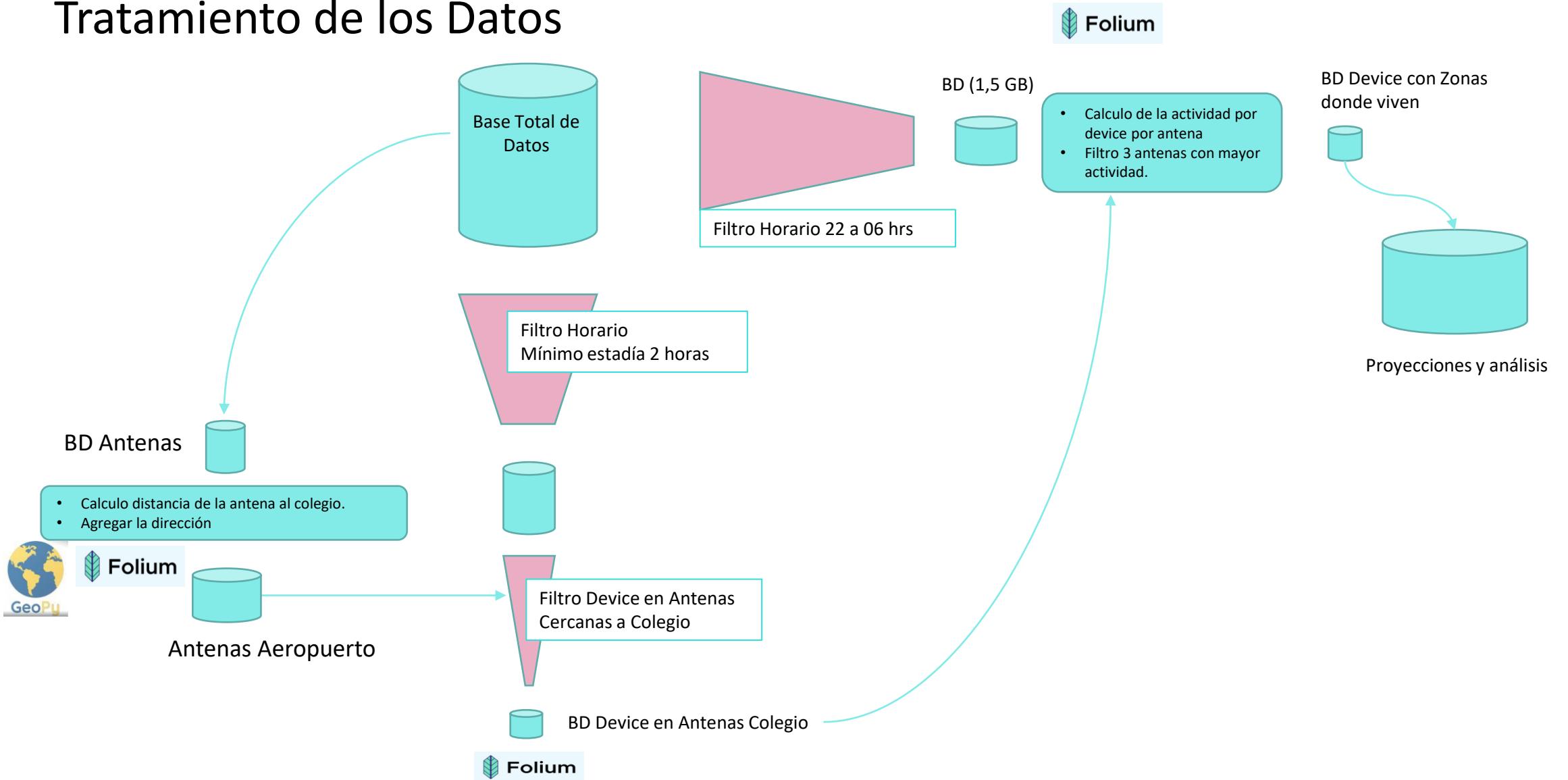


Red Aeroportuaria Norte:
3a concesión Andrés Sabella de Antofagasta
2^a concesión Desierto de Atacama
Presupuesto: MM USD 267

Segunda Concesión
Aeropuerto Región de la Araucanía
Presupuesto: MM USD 85

Red Aeroportuaria Sur:
Tepual (5a), Cañal Bajo, Mocopulli y Pichoy
Región de Los Ríos y Los Lagos
Presupuesto: MM USD 273

Tratamiento de los Datos



Metodología

Información de apoderados

1. Fecha y horario de la muestra:

Marzo 2020						
Lunes	Martes	Miercoles	Jueves	Viernes	Sábado	Domingo
24	25	26	27	28	29	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5

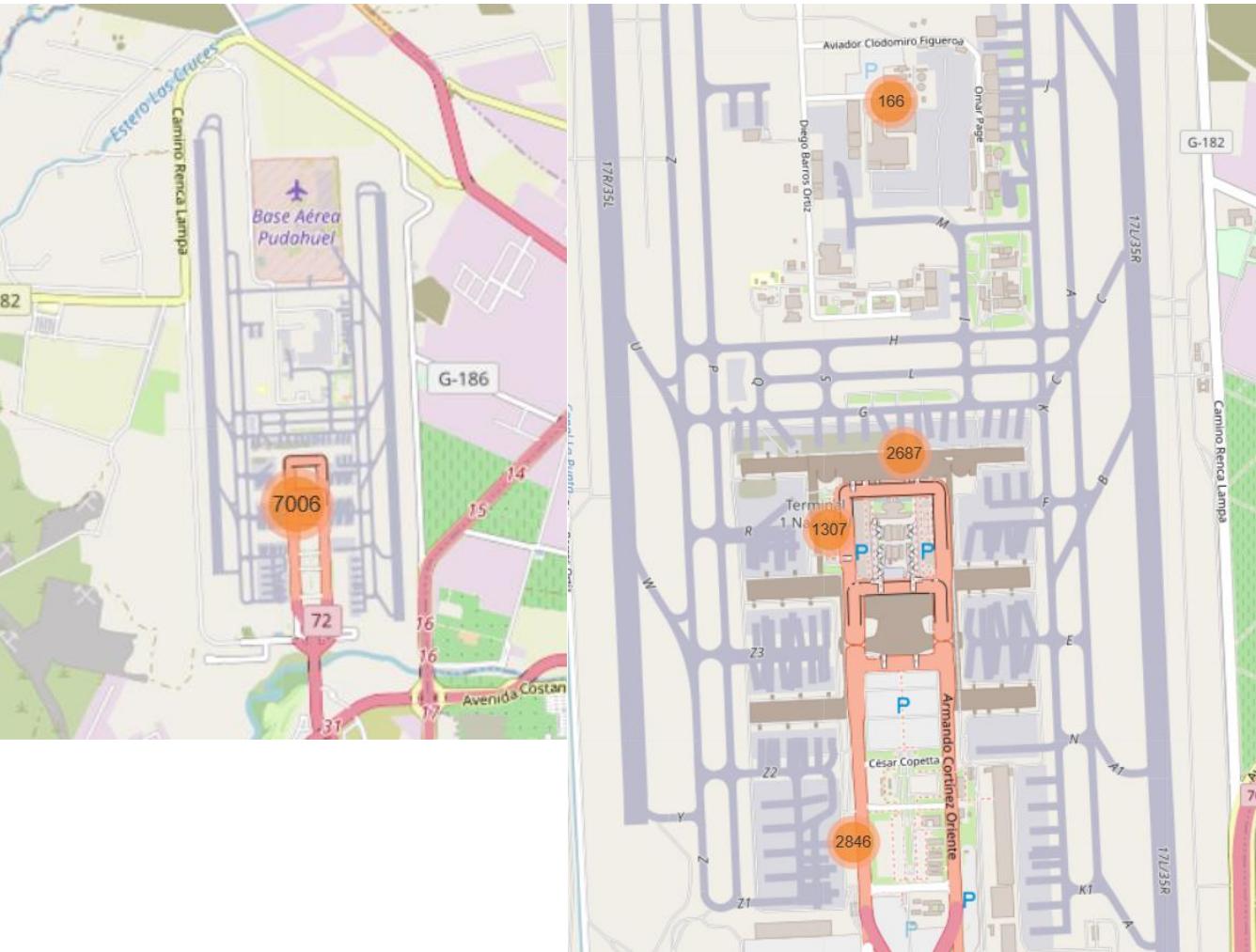
Marzo 2021						
Lunes	Martes	Miercoles	Jueves	Viernes	Sábado	Domingo
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

Marzo 2022						
Lunes	Martes	Miercoles	Jueves	Viernes	Sábado	Domingo
28	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	1	2	3

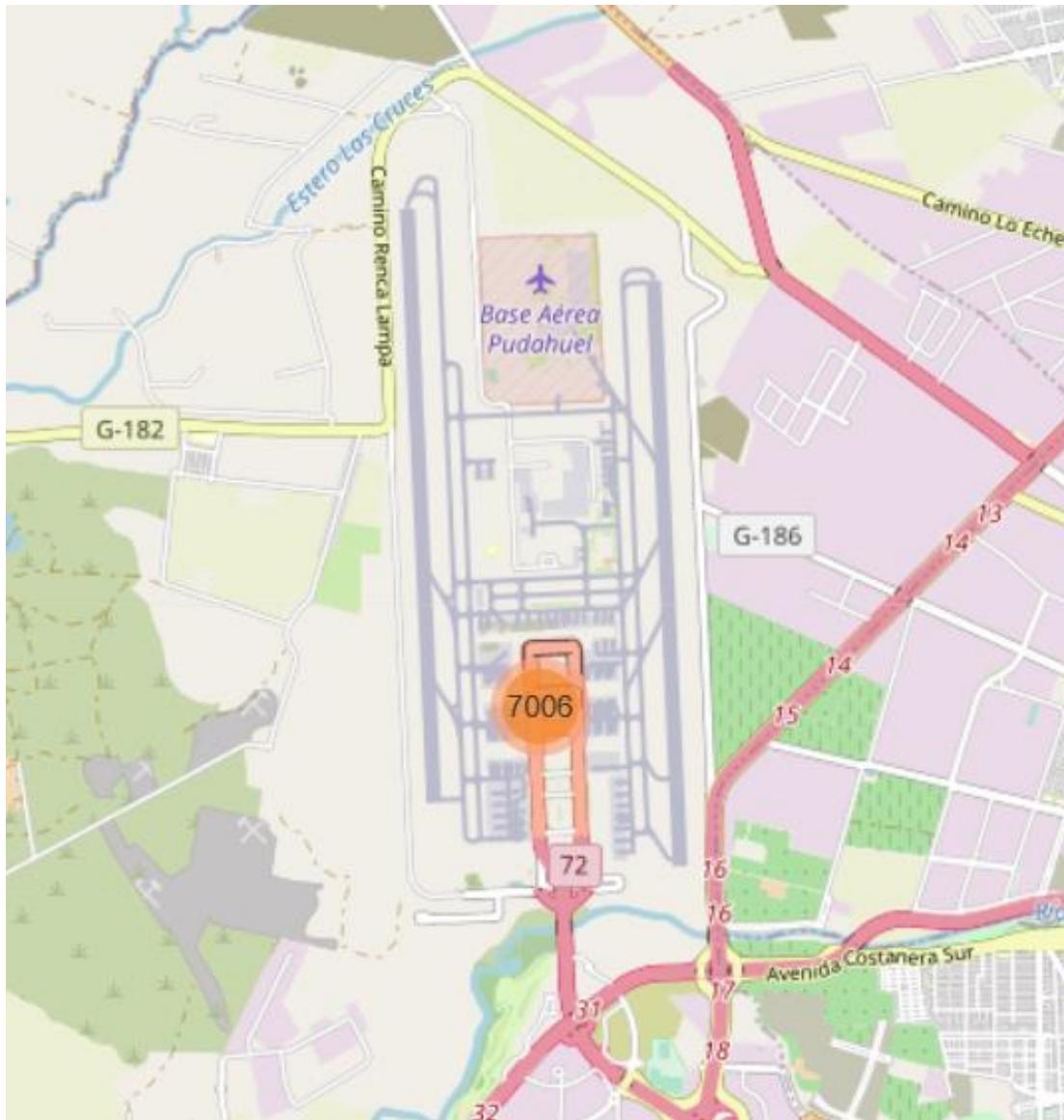
Filtro de horario: mínimo 2 horas de estadía en el aeropuerto

2. Toma de la muestra:

- ✓ Calculo la distancia de todas las antenas al aeropuerto
- ✓ Filtro de todas las antenas que se encuentran en el aeropuerto



2020



2021

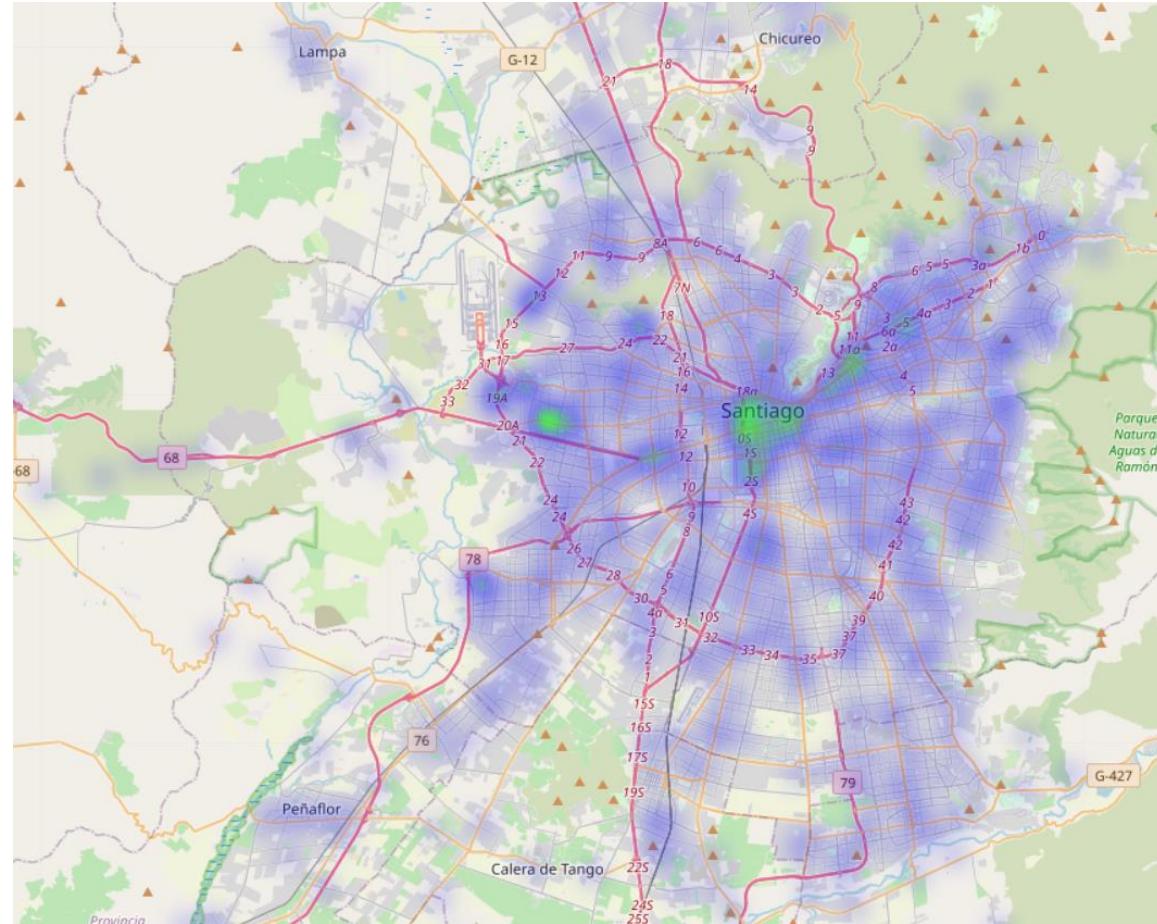
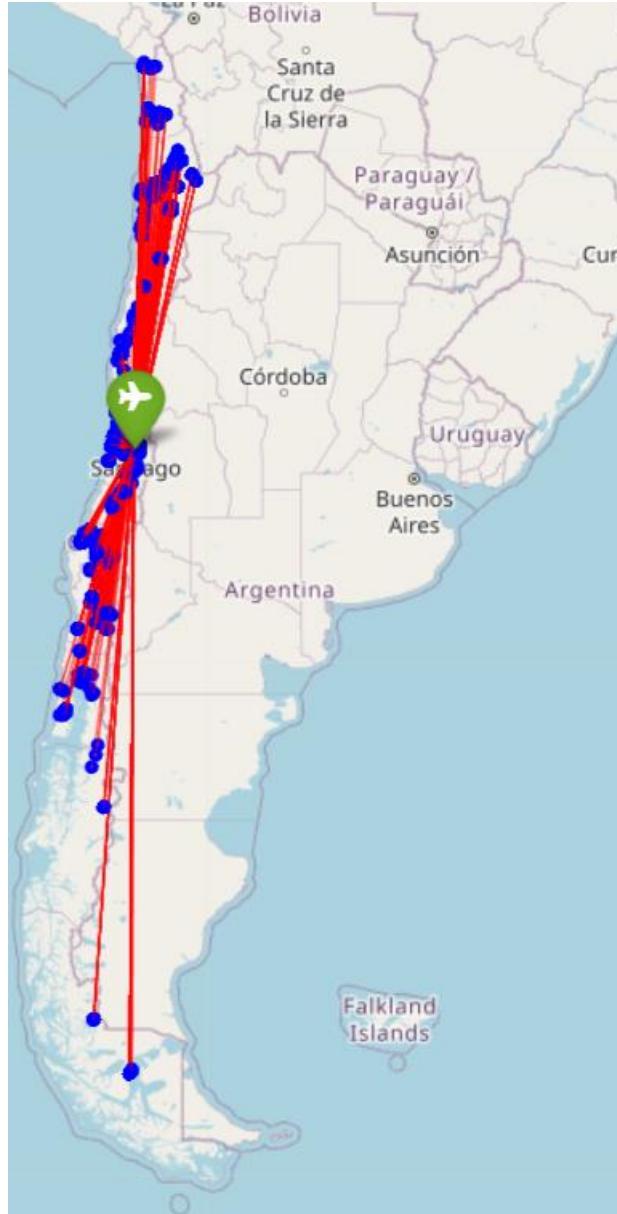


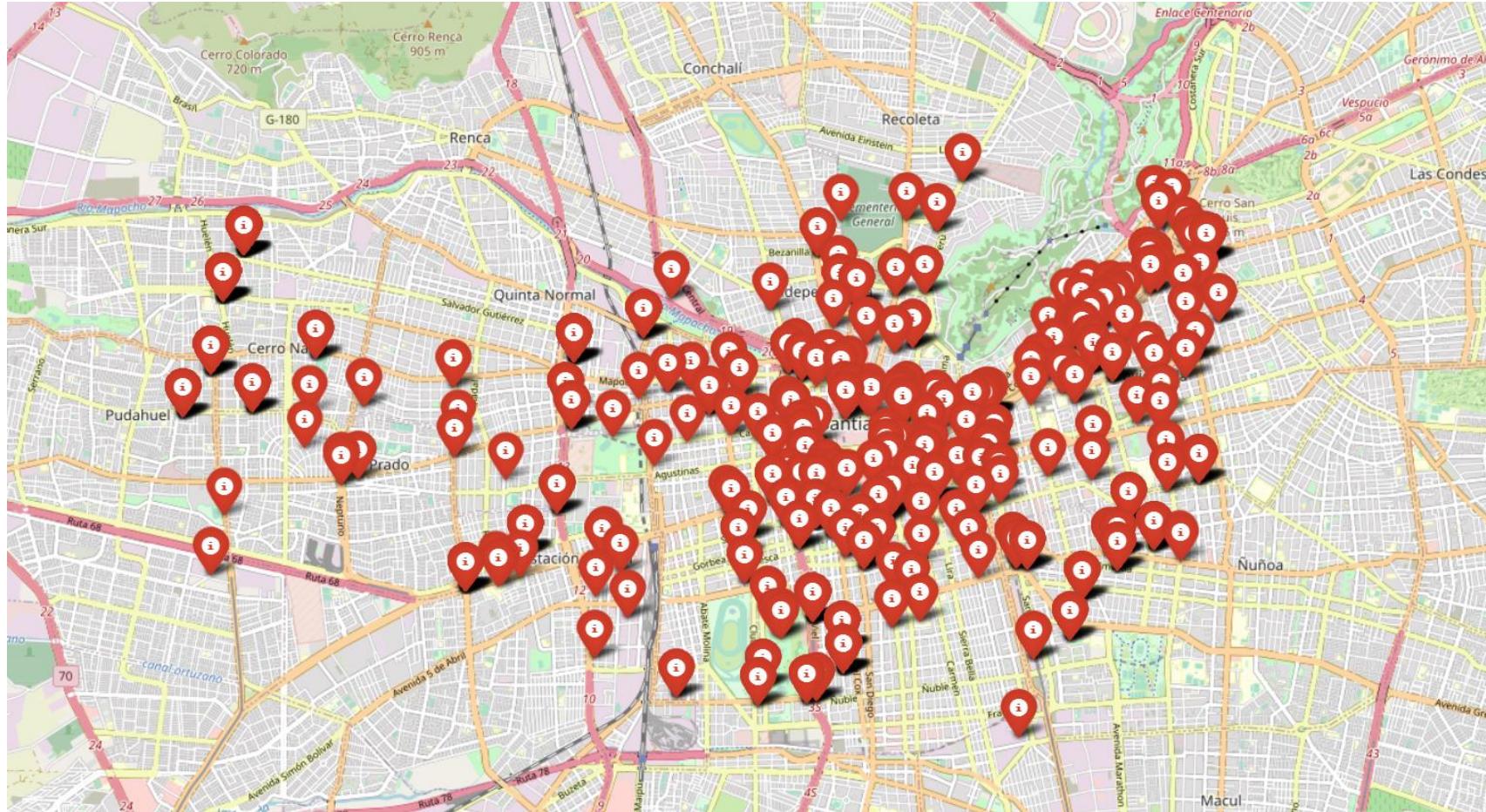
2022



Distancia del aeropuerto y densidad de devices identificados en el aeropuerto

Building on belief

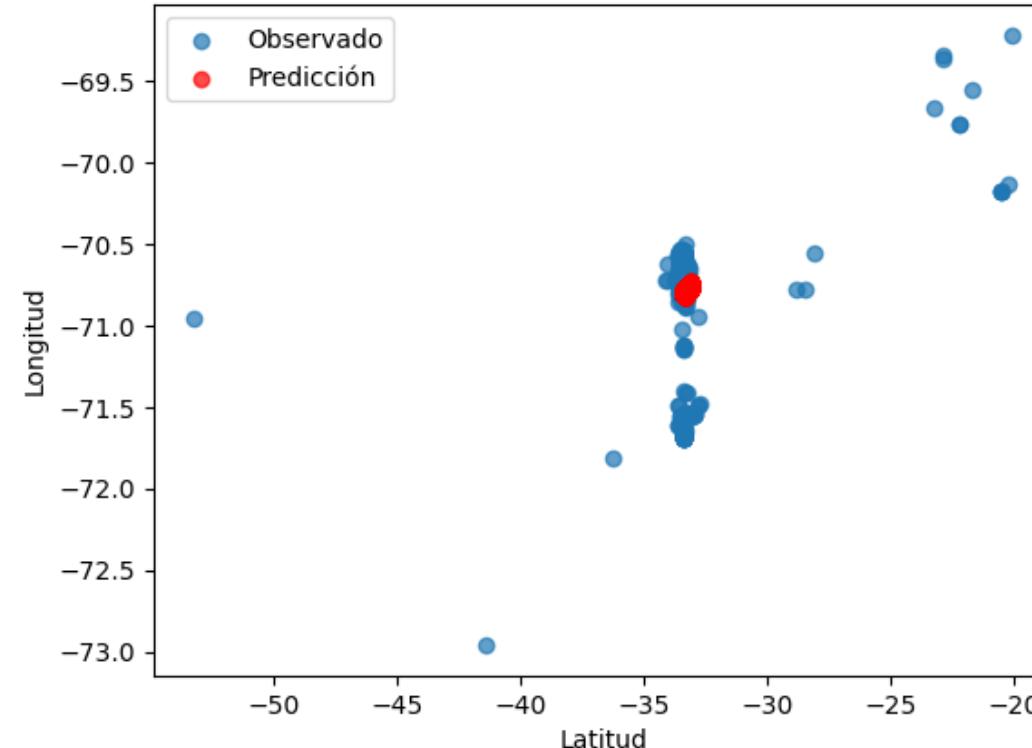




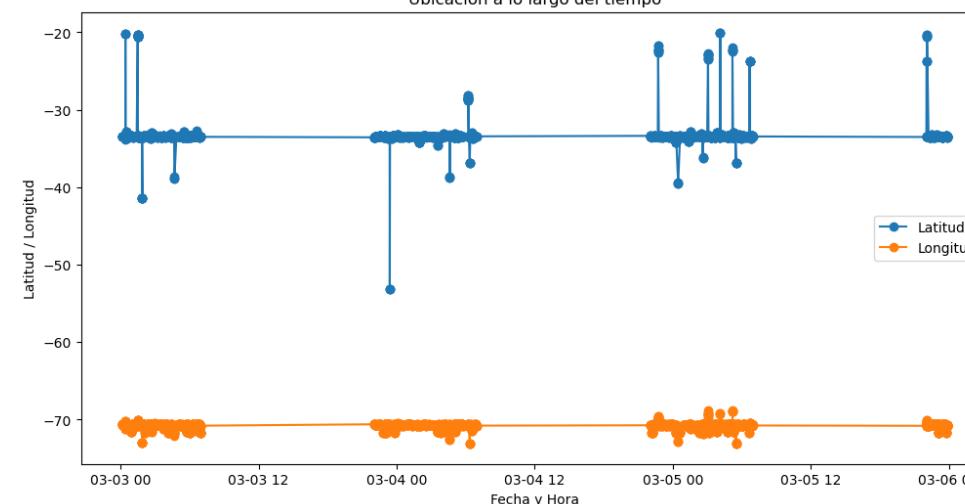
Distribución proporcional en el país y movilidad

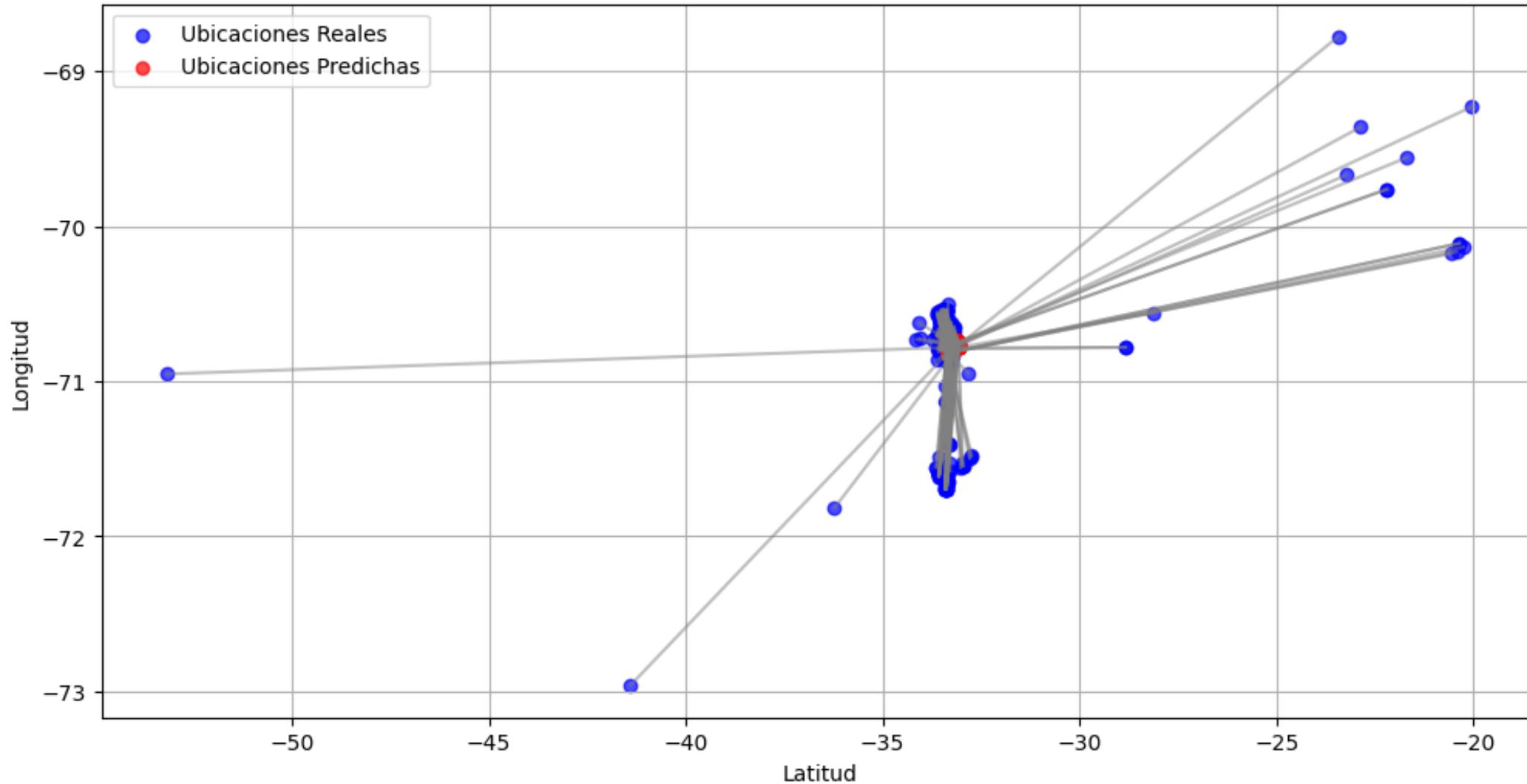


Predicción de Ubicaciones

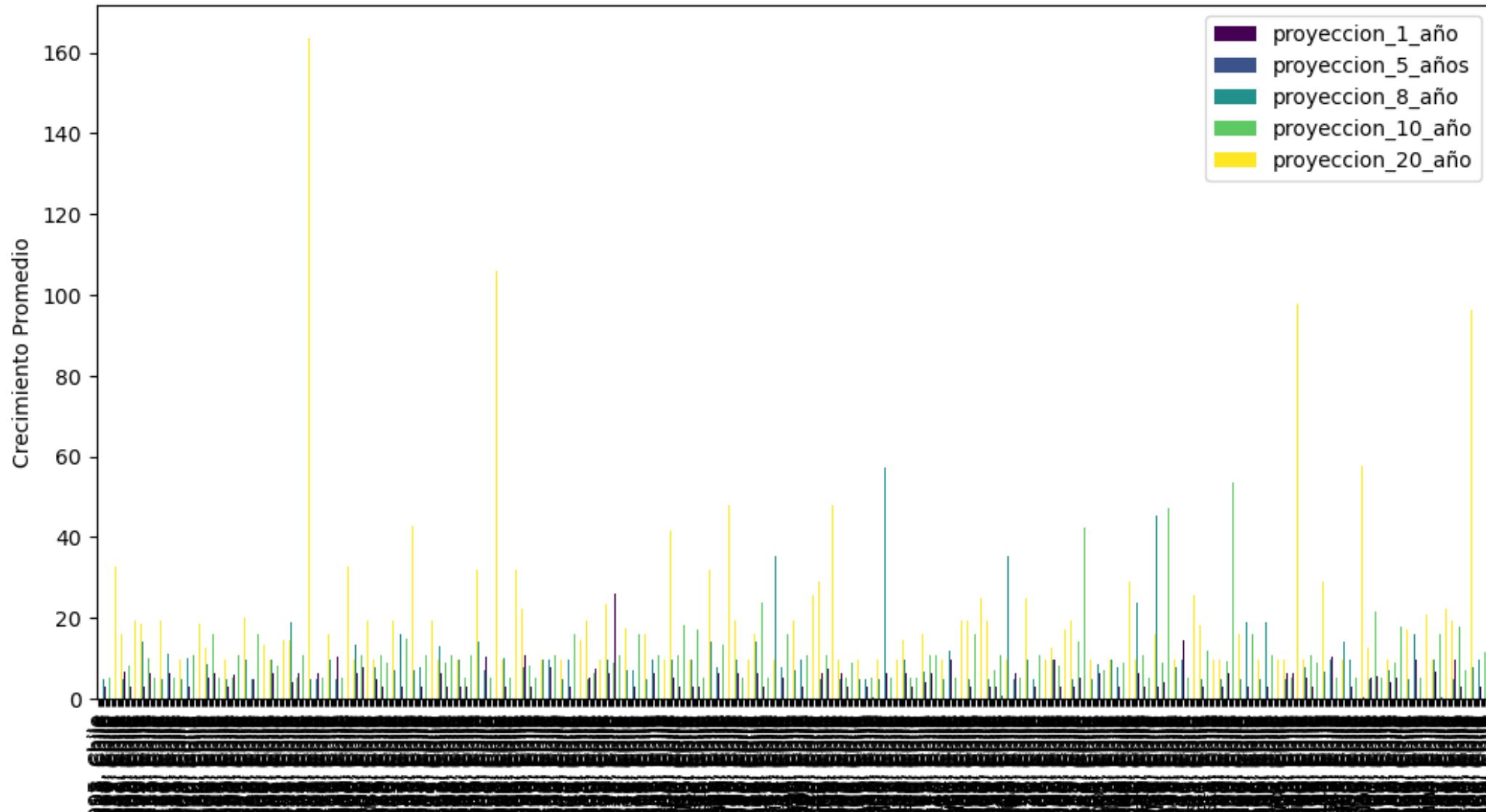


Ubicación a lo largo del tiempo

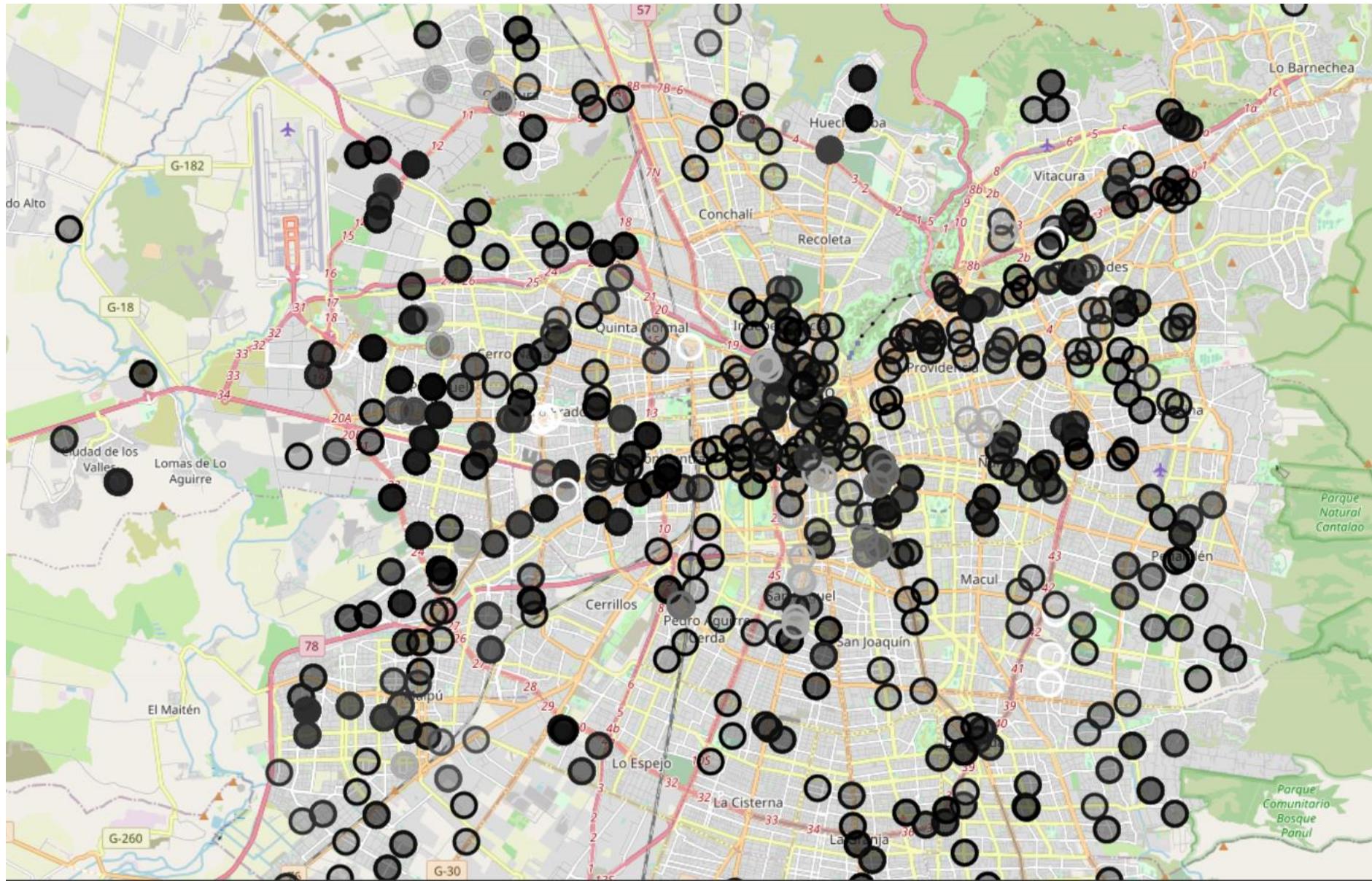


Ubicaciones Reales vs. Predichas

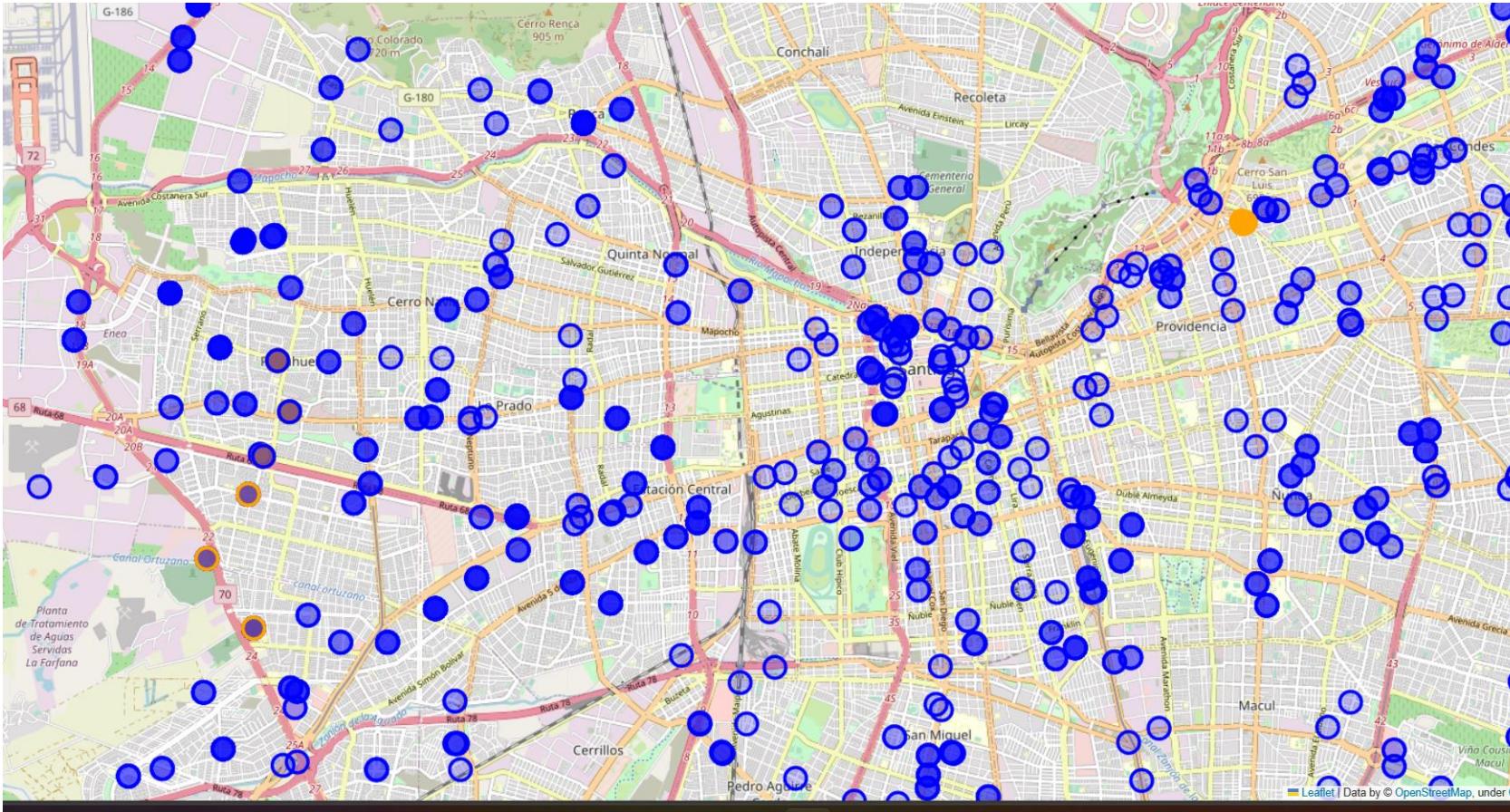
Crecimiento Promedio por Dirección



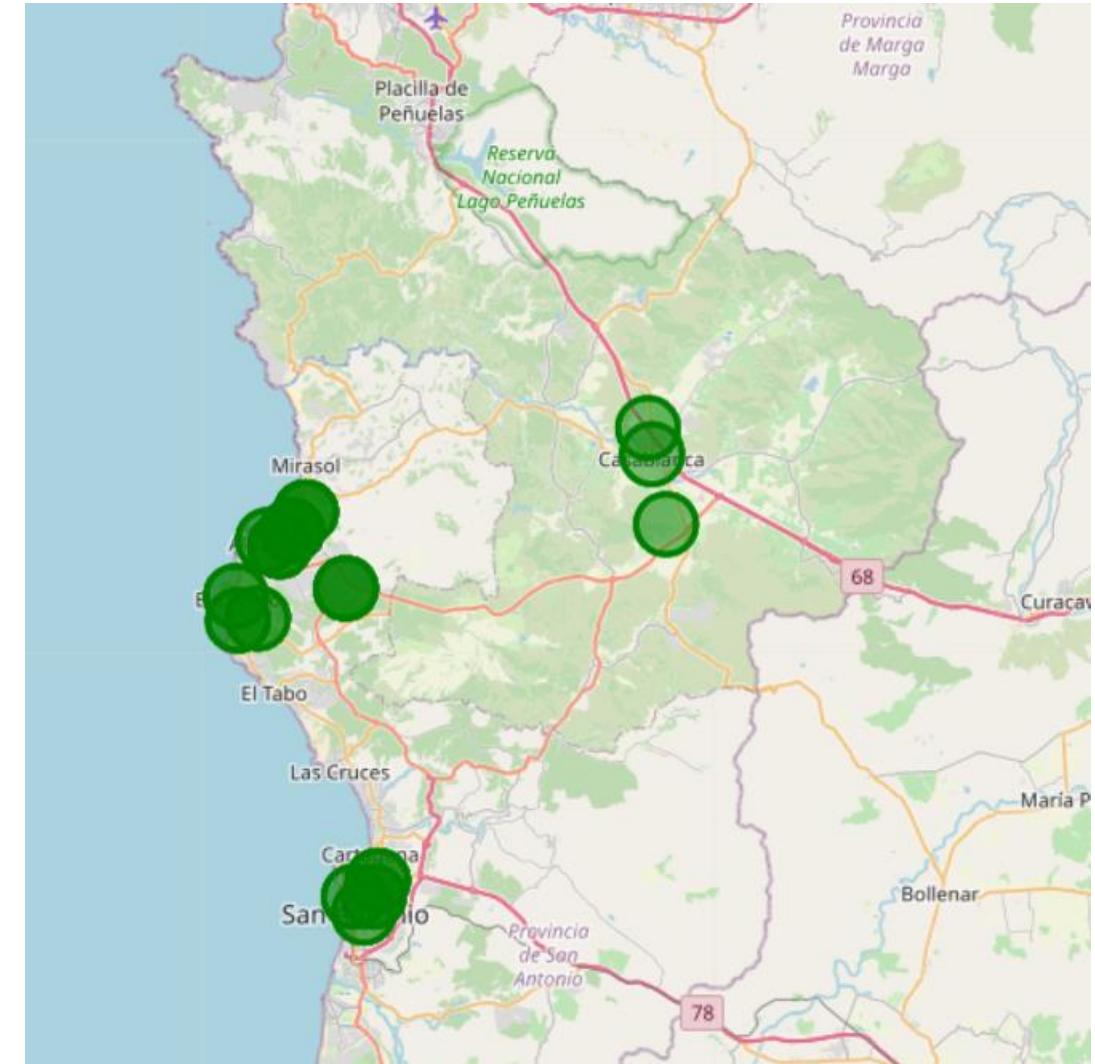
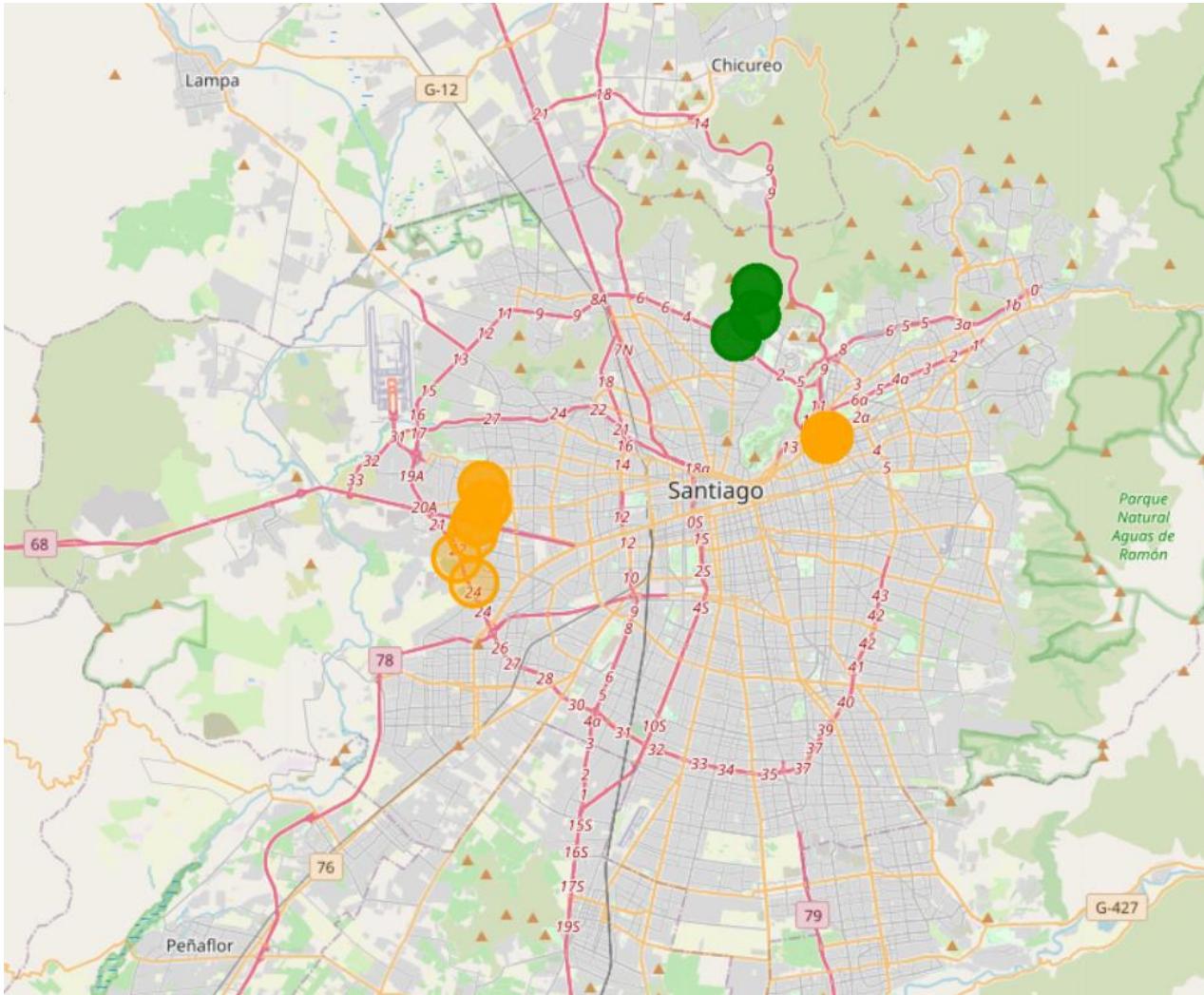
Proyección de densidad



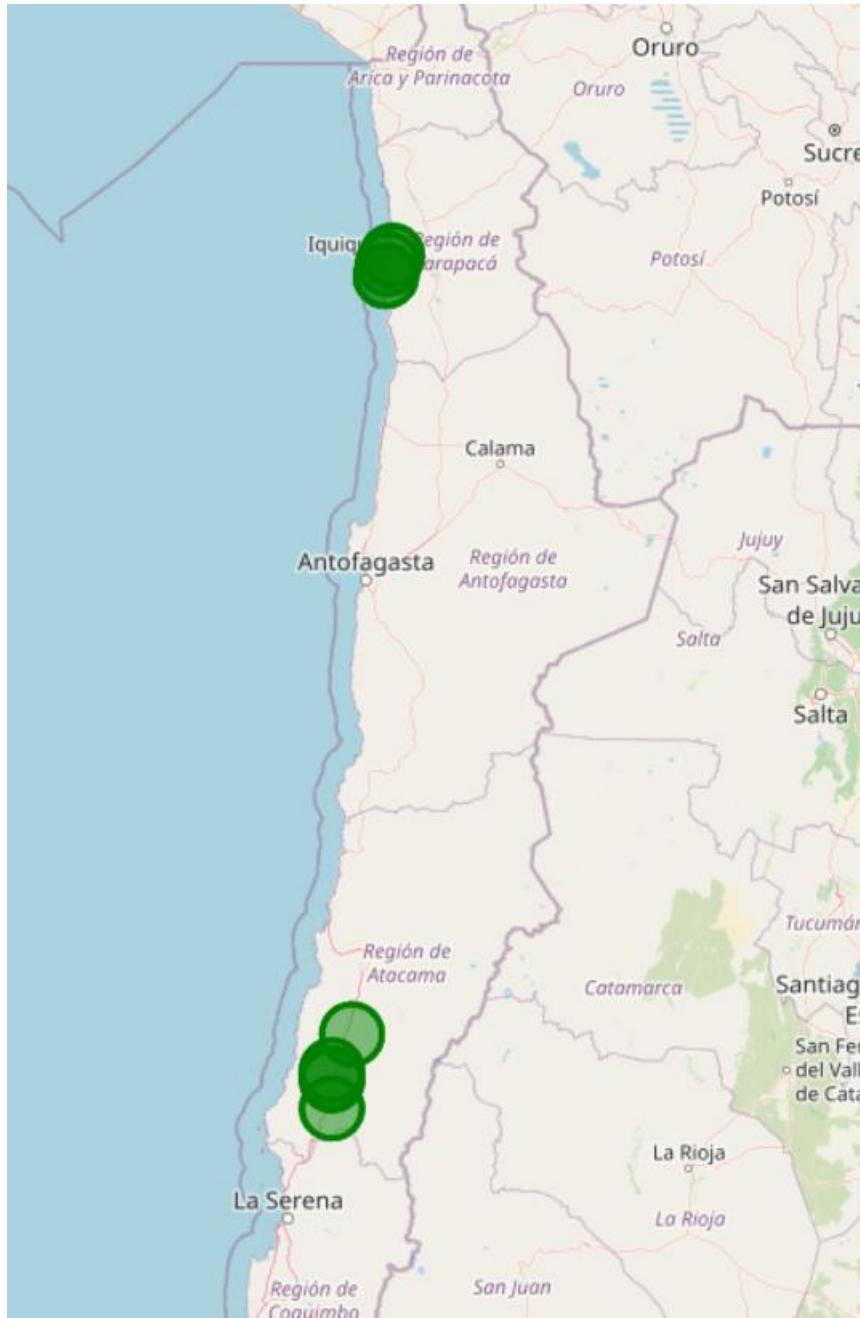
Proyección de crecimiento



Proyección de crecimiento



Proyección de crecimiento



Building on belief