

Análisis de barrios donde es conveniente construir casas o edificios.

Christian Sebastian Russo

Buenos Aires, 2020

Índice general

1.. Introducción	1
2.. Descripción del problema	2
3.. Enfoque de la solución	3
4.. Análisis de los datos	4
5.. Cluster Neighborhood	8
6.. Análisis de un barrio en particular	12
7.. Trabajos futuros	14
8.. Enlaces externos	15

1. INTRODUCCIÓN

En Argentina, muchas empresas constructoras tienen problemas a la hora de encontrar el mejor barrio para construir casas o edificios.

Este se debe a que los lugares para construir tienen precios diferentes, pero no solo eso, sino que también depende del movimiento comercial del barrio, las empresas, las industrias, la cercanía al transporte, comercios, negocios, entre otras.

Por otro lado, no es lo mismo la construcción de una casa, como la de un edificio, el análisis comercial y de movimiento de una zona es diferente.

Es por este motivo que a las diferentes empresas tienen dificultad y se preguntan: ¿Cuál es el mejor lugar para comenzar un nuevo emprendimiento?.

2. DESCRIPCIÓN DEL PROBLEMA

En la provincia de **Buenos Aires** existen barrios donde una propiedad cuesta alrededor de USD 70.000, a su vez cerca de esta propiedad podemos encontrar una propiedad de similares características a USD 200.000, es decir , un precio de **285 %** mas elevado.

El precio de una propiedad esta dado entre otras cosas, por la proximidad a lugares de interés; por ejemplo, no tendrá el mismo precio una propiedad en cercanía de un supermercado/shopping que próxima a un cementerio.

Entre muchos de los factores que se tienen en cuenta a la hora de elegir *dónde* se considera más factible el emprendimiento, se encuentra la proximidad con diferentes puntos de interés, como son: comercios, supermercados, bares, estaciones de tren, estaciones de colectivo, avenidas, etc.

3. ENFOQUE DE LA SOLUCIÓN

En el presente trabajo, lo primero que haremos es obtener la información geográfica necesaria de los barrios de **Buenos Aires** (llamados correctamente departamentos) para poder trabajar con la API de *Foursquare* para así obtener los puntos de interés de cada barrio.

Una vez obtenida esta información, calcularemos, para cada barrio que proporción de cada categoría de lugares de interés cuenta, calculando finalmente un valor total (la suma de los porcentajes de cada categoría) y de esta forma poder medir que tan factible es construir en ese barrio.

Luego, utilizaremos el algoritmo de **Machine Learning** (llamado *k-means*) para poder determinar similitudes entre barrios, y con estas similitudes encontrar un barrio similar a otro con las características deseadas.

Las características deseadas que utilizaremos en estos ejemplos son:

- Estaciones de trenes
- Comercios
- Restaurantes
- Gimnasios
- Plazas
- Aeropuertos
- Cines
- Shoppings
- Supermercados
- Bares
- Clubs
- Etc.

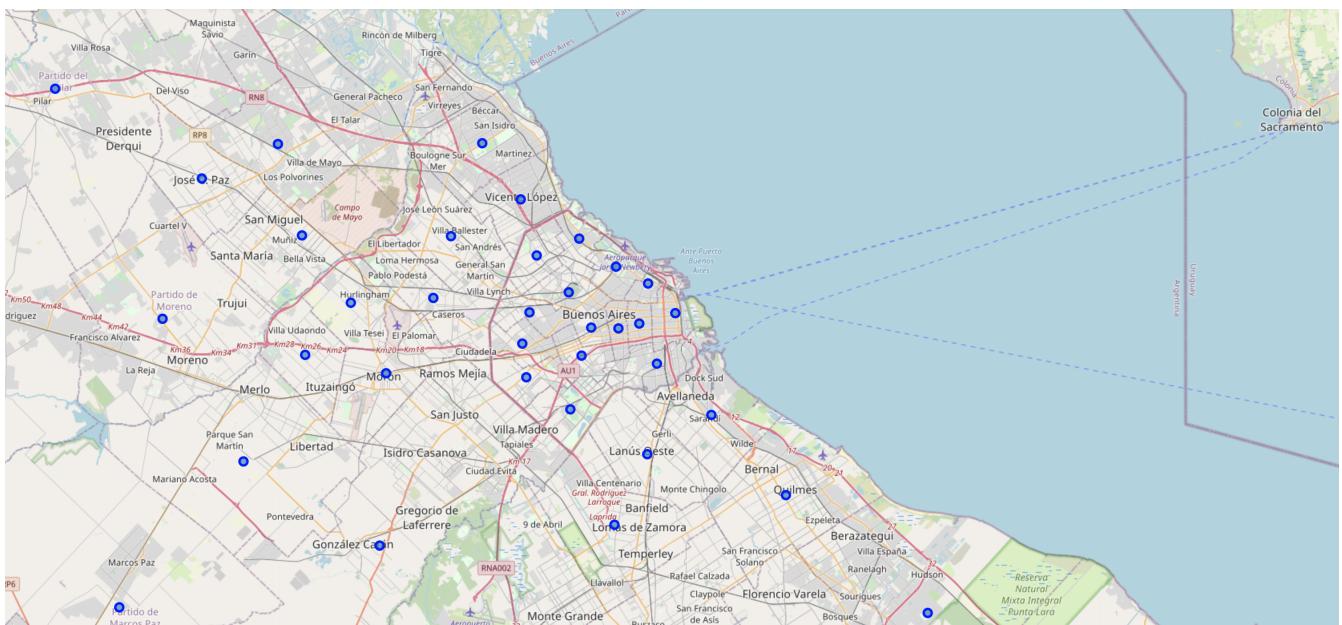
Por ultimo, analizaremos un barrio en particular, evaluando en qué zonas dentro de aquel, el movimiento comercial es mayor.

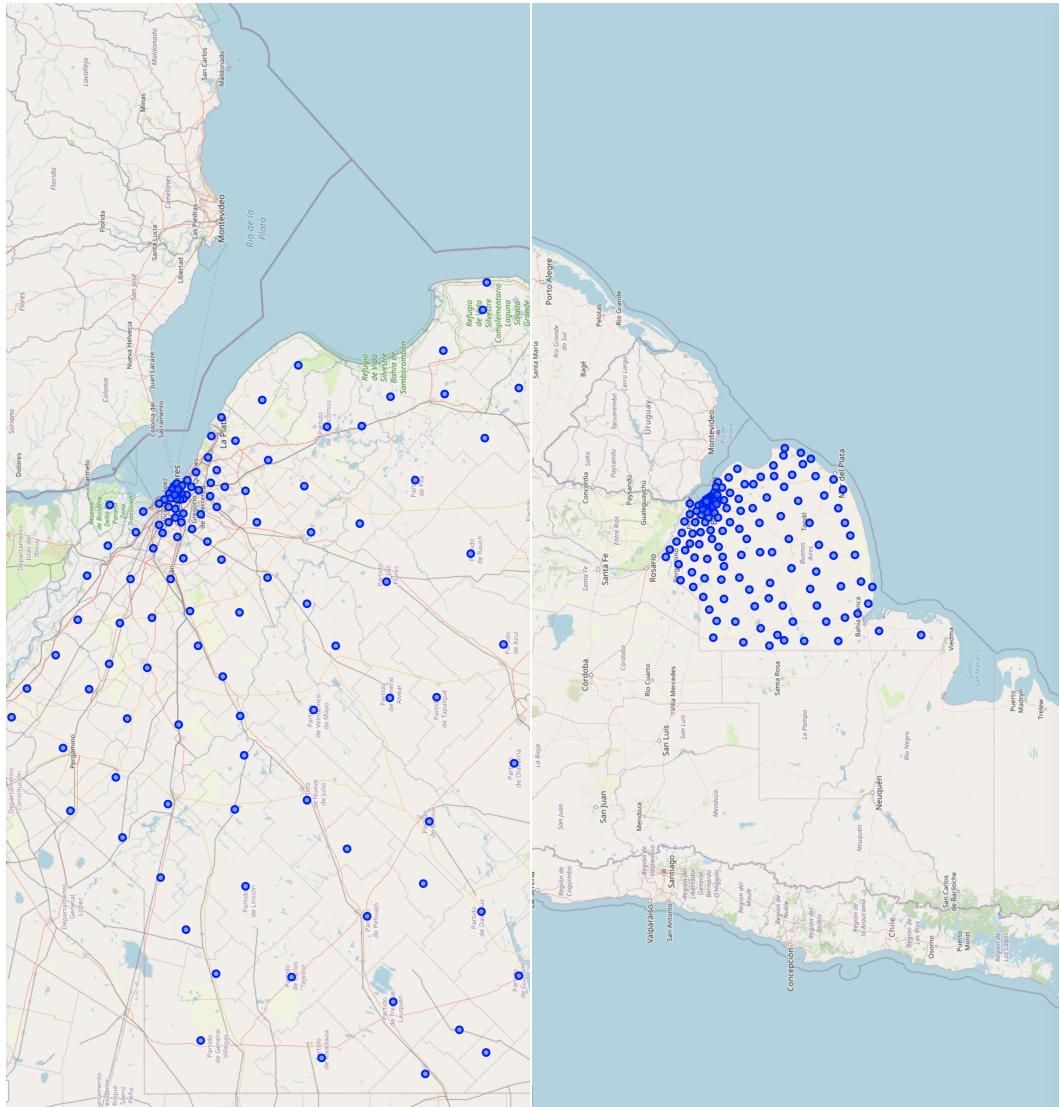
4. ANÁLISIS DE LOS DATOS

Lo primero que haremos será analizar los datos obtenidos de <https://www.datos.gob.ar/>. Una vez ya filtrados los datos en Buenos Aires y Capital Federal se mostrará el resultado de aquellos datos de los barrios de Argentina.

categoria	centroide_lat	centroide_lon	fuente	id	nombre	nombre_completo	
0	Partido	-34.511876	-58.777671	ARBA - Gerencia de Servicios Catastrales	6412	José C. Paz	Partido de José C. Paz
3	Partido	-37.964225	-60.249078	ARBA - Gerencia de Servicios Catastrales	6014	Adolfo Gonzales Chaves	Partido de Adolfo Gonzales Chaves
4	Partido	-37.153214	-57.230787	ARBA - Gerencia de Servicios Catastrales	6315	General Juan Madariaga	Partido de General Juan Madariaga
5	Partido	-37.335655	-59.181806	ARBA - Gerencia de Servicios Catastrales	6791	Tandil	Partido de Tandil
13	Partido	-38.147744	-61.264659	ARBA - Gerencia de Servicios Catastrales	6196	Coronel Pringles	Partido de Coronel Pringles

Utilizando las coordenadas geográficas de cada barrio (latitud y longitud) dibujaremos un mapa para poder verificar la información obtenida:





Como se puede ver en los tres mapas, todos los barrios de la provincia de buenos aires son representados con círculos azules.

Luego, utilizando la API de Foursquare (<https://api.foursquare.com/v2/venues/explore>), calcularemos para cada barrio sus puntos de interés.

Notar, que para el uso de la API necesitamos configurar un límite y radio de resultados para cada barrio. Para esta instancia configuraremos el radio en 8000 metros y el límite de resultados en 500 unidades.

Una vez utilizada la API, encontramos 3904 puntos de interés, y para cada uno de estos su categoría, latitud, longitud, nombre, tipo, etc.

Finalmente, lo que haremos, utilizando la técnica de One Hot Encoding y agrupando datos será generar una tabla que para cada barrio podamos ver cuántos puntos de interés tiene separado por categoría.

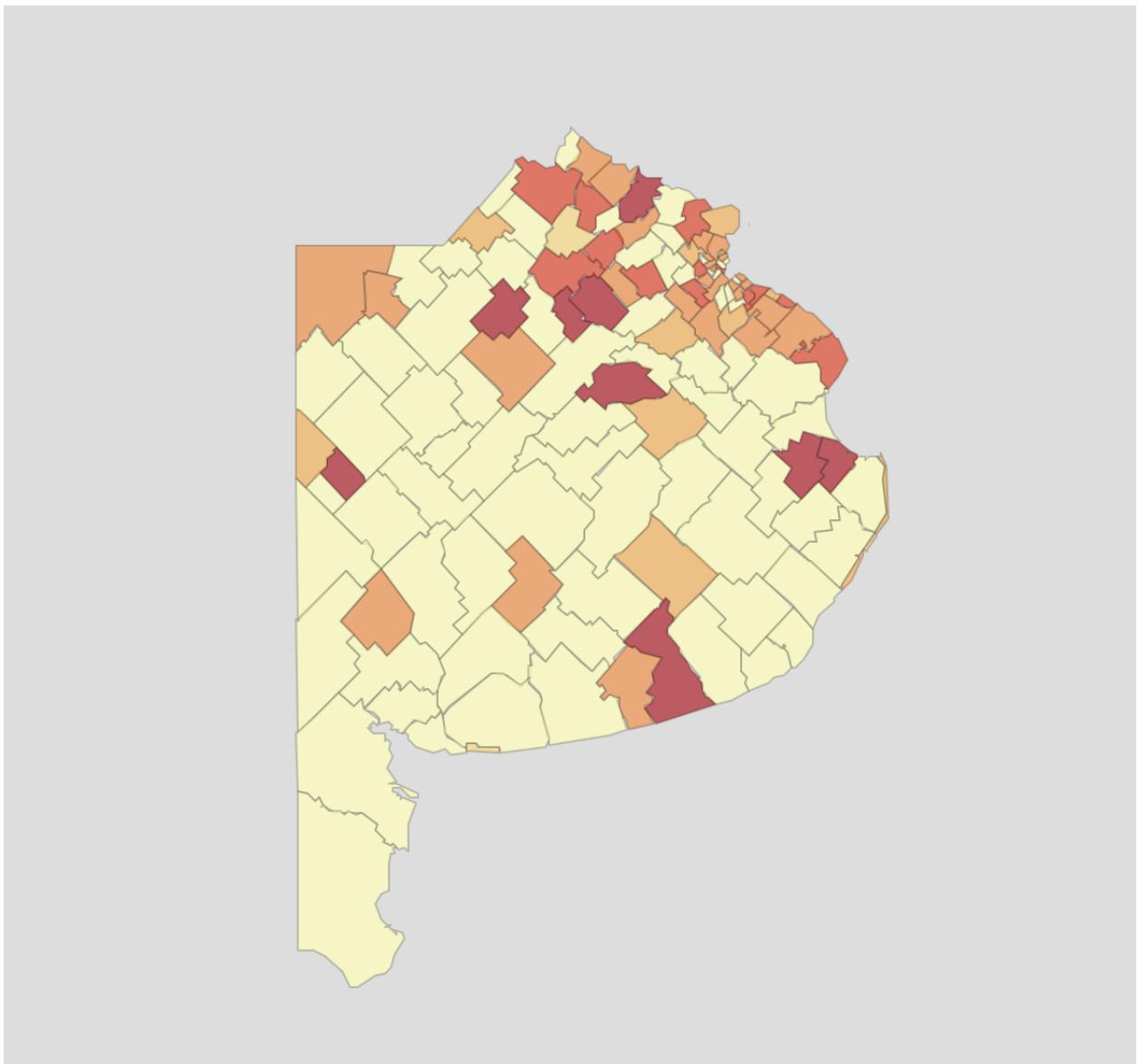
Veamos una parte de la tabla:

	Neighborhood	Zoo	Accessories Store	Airport	Airport Lounge	Airport Terminal	American Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop
0	9 de Julio	0.0	0.00	0.0	0.0	0.5	0.000	0.500000	0.00	0.00	0.000000	0.000000	0.0	0.0
1	Alberti	0.0	0.00	0.0	0.0	0.0	0.000	0.000000	0.00	0.00	0.000000	0.000000	0.0	0.0
2	Almirante Brown	0.0	0.00	0.0	0.0	0.0	0.000	0.032258	0.00	0.00	0.016129	0.032258	0.0	0.0
3	Arrecifes	0.0	0.00	0.0	0.0	0.0	0.125	0.250000	0.00	0.00	0.000000	0.000000	0.0	0.0
4	Avellaneda	0.0	0.00	0.0	0.0	0.0	0.000	0.130000	0.02	0.01	0.000000	0.000000	0.0	0.0
...
106	Tres Lomas	0.0	0.00	0.0	0.0	0.0	0.000	0.000000	0.00	0.00	0.000000	0.000000	0.0	0.0
107	Tres de Febrero	0.0	0.00	0.0	0.0	0.0	0.020	0.050000	0.00	0.00	0.000000	0.020000	0.0	0.0
108	Vicente López	0.0	0.01	0.0	0.0	0.0	0.020	0.040000	0.00	0.00	0.000000	0.020000	0.0	0.0
109	Villa Gesell	0.0	0.00	0.0	0.0	0.0	0.000	0.075000	0.00	0.00	0.000000	0.000000	0.0	0.0
110	Zárate	0.0	0.00	0.0	0.0	0.0	0.000	0.666667	0.00	0.00	0.000000	0.000000	0.0	0.0

Como se puede ver en la tabla, tenemos un valor promedio en cada categoría agrupada por barrio.

Por ejemplo, podemos ver que para 9 de Julio, el 0.5 de los resultados obtenidos son Restaurantes Argentinos.

Luego de esto, sumaremos para todos los barrios la información de la tabla, para obtener un valor por barrio y poder dibujarlo obteniendo el siguiente gráfico. Es importante remarcar que no se sumaron todas las categoría, sino solo las categoría que deseadas:



Asimismo en el mapa verá de rojo más fuerte los barrios en donde las categorías deseadas son mas frecuentes y por lo tanto lo consideramos un barrio con mas movimiento según nuestros deseos.

Es interesante remarcar que con esta información podríamos, dado un barrio, saber que categorías de los puntos de interés es más frecuente, por ejemplo veamos los resultados en el barrio Hurlingham:

Objectivo	Valor
Ice Cream Shop	0.13
Argentinian Restaurant	0.12
Cafe	0.07
Plaza	0.04
Bar	0.03

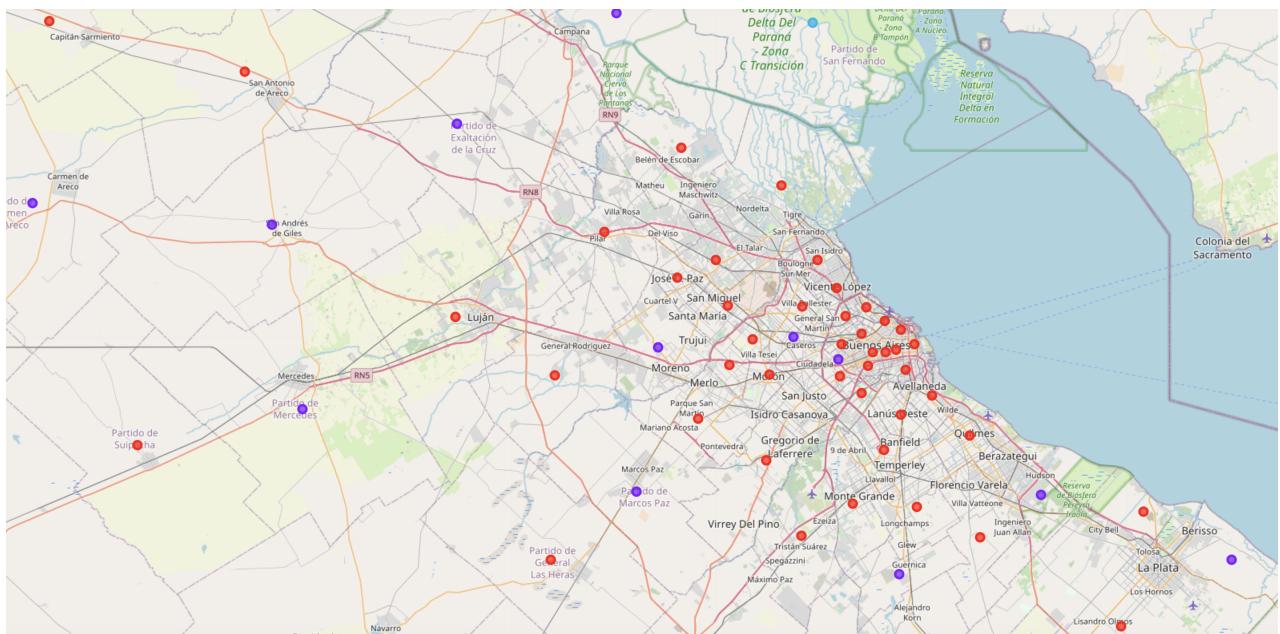
5. CLUSTER NEIGHBORHOOD

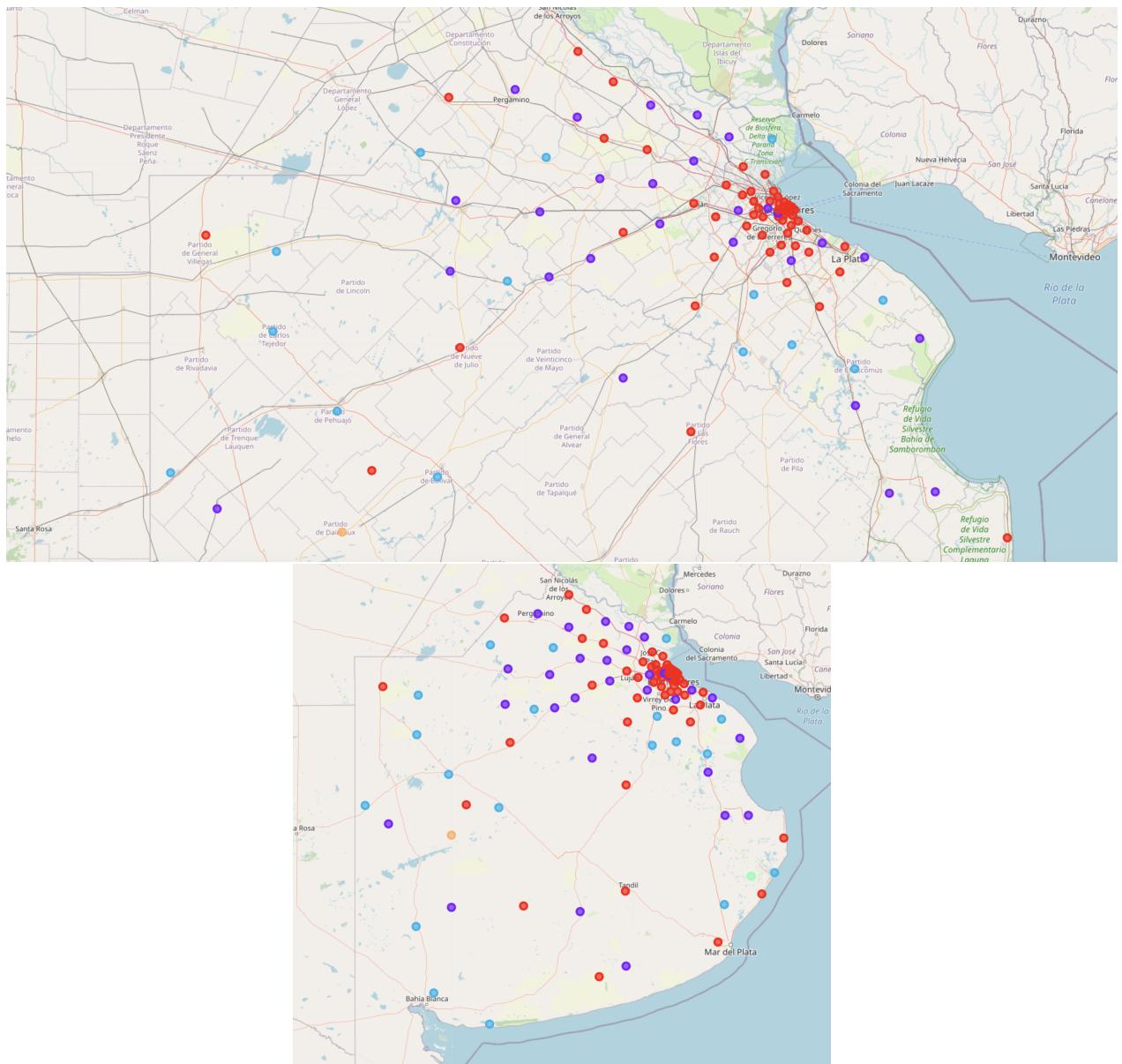
Con el algoritmo de Machine Learning llamado KMeans, generaremos 5 clusters para poder encontrar similitud entre diferentes barrios, esto podría ayudar a las empresas constructoras en su análisis y así determinar donde les es conveniente invertir.

Agregaremos a las tablas los resultados de nuestro algoritmo para así tener, dado un id de barrio, a que cluster pertenece, como vemos en la siguiente tabla:

Cluster Labels	
0	0
1	1
2	0
3	1
4	0

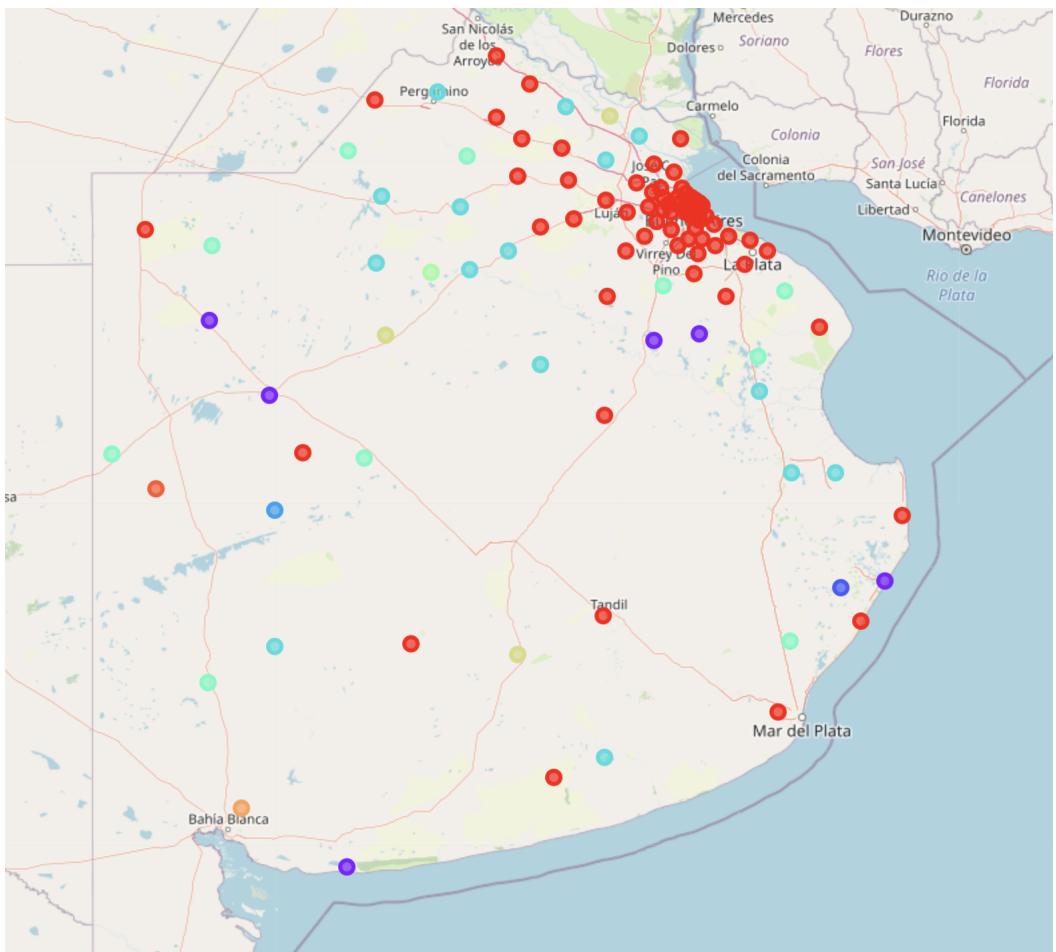
Para ver mejor los resultados y ver los clusters de nuestro algoritmo lo dibujamos en el mapa:



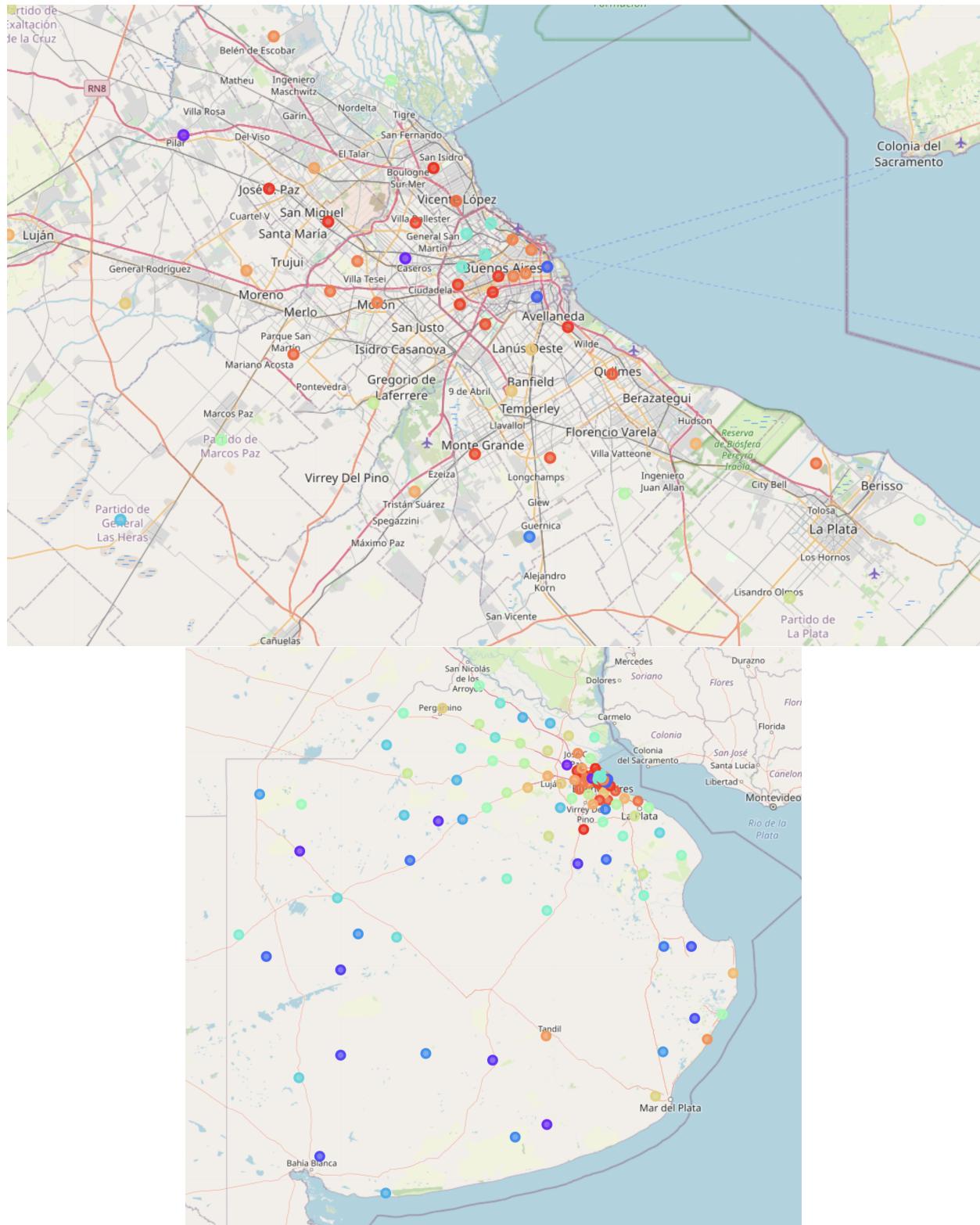


En los mapas, observamos que todos los puntos del mismo color, son los barrios que consideramos que son similares en el movimiento de los puntos de interés seleccionados.

Ahora, analicemos el mismo algoritmo pero con 10 clusters y veamos los resultados obtenidos:



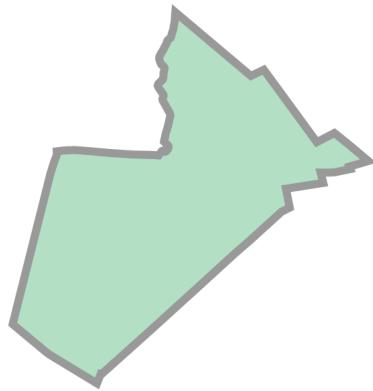
Por ultimo, evaluemos con cantidad de clusters igual a 100.



6. ANÁLISIS DE UN BARRIO EN PARTICULAR

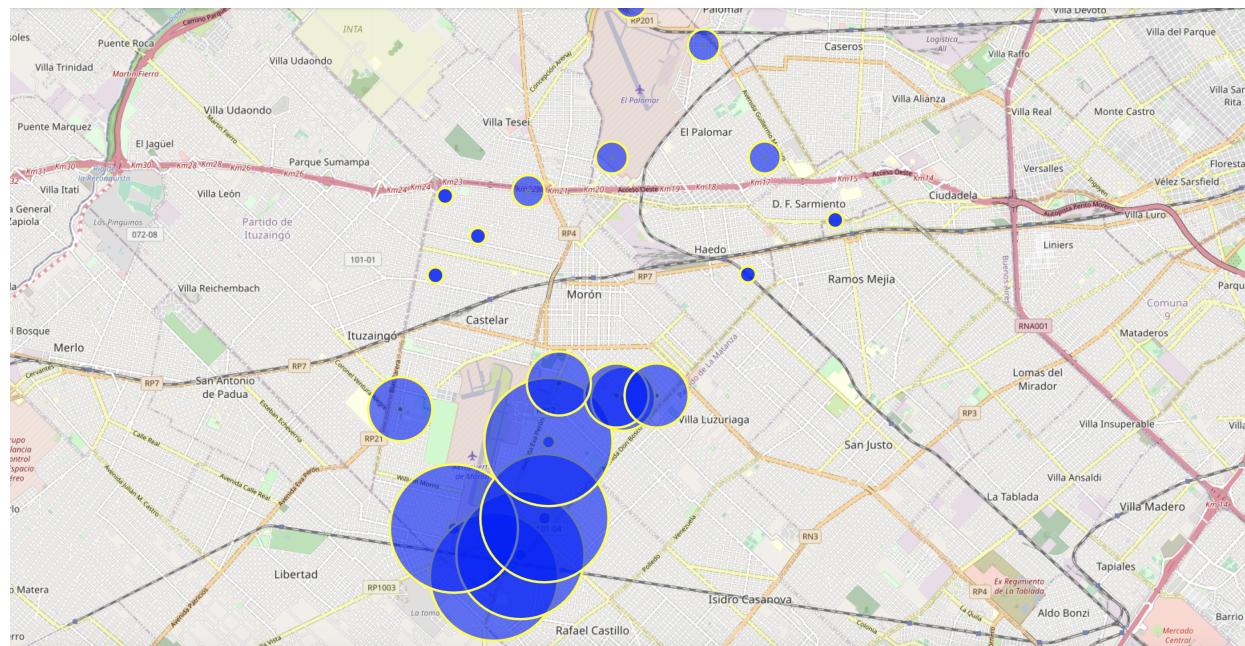
En este capítulo evaluaremos la misma lógica utilizada en el capítulo anterior, pero en vez de en toda la provincia de Buenos Aires, solamente tomaremos el barrio llamado **Morón**.

Lo primero que haremos será determinar los límites geográficos del barrio y obtener el polígono del mismo.



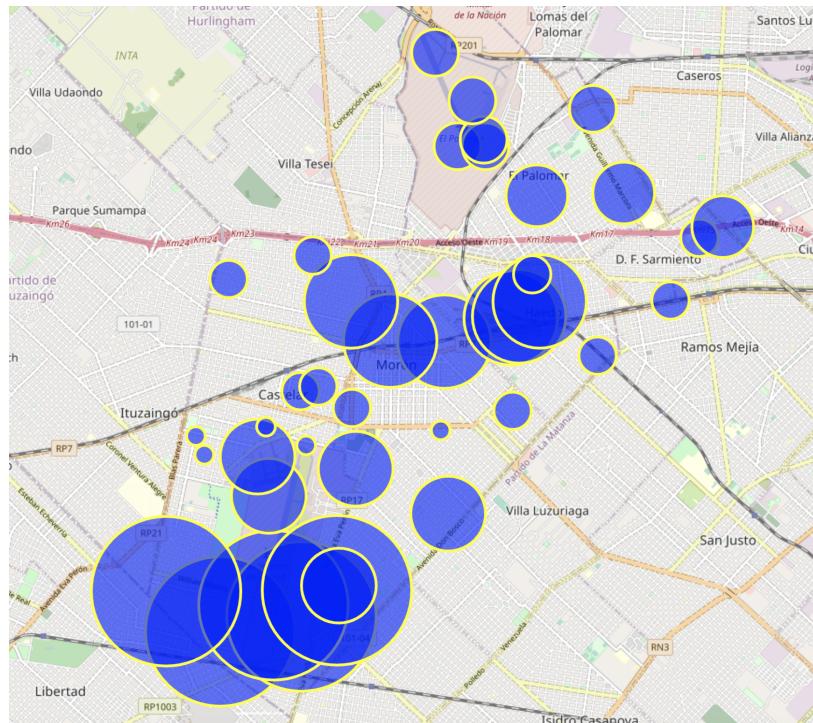
Luego generaremos puntos aleatorios dentro de las coordenadas límites del barrio. Para este ejemplo ejecutaremos solamente 20 puntos.

Repetimos el proceso anterior, es decir, generamos una tabla con One Hot Encoder y agrupamos por un identificador único (en este caso utilizamos como identificador la tupla compuesta por la latitud y la longitud). Calculamos el total de la suma de los porcentajes de movimiento en cada punto de interés y realizamos en un mapa.



En el mapa, se pueden ver los puntos marcados y el radio es equivalente al porcentaje de movimiento del punto. Es decir, cuanto más radio tiene el círculo centrado en el punto, es porque ese punto, tiene más movimiento de los puntos de interés deseado.

Probemos ahora con 40 puntos aleatorios nuevos:



7. TRABAJOS FUTUROS

Queda por fuera de este informe, el análisis de los hiperparámetros, es decir, en los algoritmos de Machine Learning, se podría evaluar el valor de K (cantidad de clusters) es mas adecuado.

Por otro lado, mejorar la utilización de la API de Foursquare, dado que recordemos que utilizamos límites y radios que fueron configurados muy pequeños para obtener de forma mas rápida los resultados.

También, a la hora de calcular el porcentaje de movilidad de cada barrio, se podría estudiar una mejor heurística o ecuación para obtener resultados mas generales.

Por ultimo, se podría sumar al estudio un análisis del precio por metro cuadrado en cada barrio.

8. ENLACES EXTERNOS

- <https://www.datos.gob.ar/>
- <https://es.foursquare.com/>