



Informe Final Capstone

Julio Delaunay

Introducción / Problema Comercial

Una cadena de hoteles de 4 estrellas en expansión tiene la intención de realizar la apertura de un nuevo hotel en la ciudad de Buenos Aires (Argentina).

El problema en este caso es la elección de la zona o barrio en dicha ciudad donde instalar el nuevo hotel, teniendo en cuenta que la cadena en su visión comercial tiene como parámetros para la nueva apertura :

- Cercanía de atracciones turísticas relevantes.
- Gastronomía de buen nivel y variedad de oferta de la misma.
- Acceso a centros comerciales importantes.
- Zona o barrio con buena proyección inmobiliaria.
- Densidad leve de hoteles de la misma categoría.

Nuestros objetivos entonces serán :

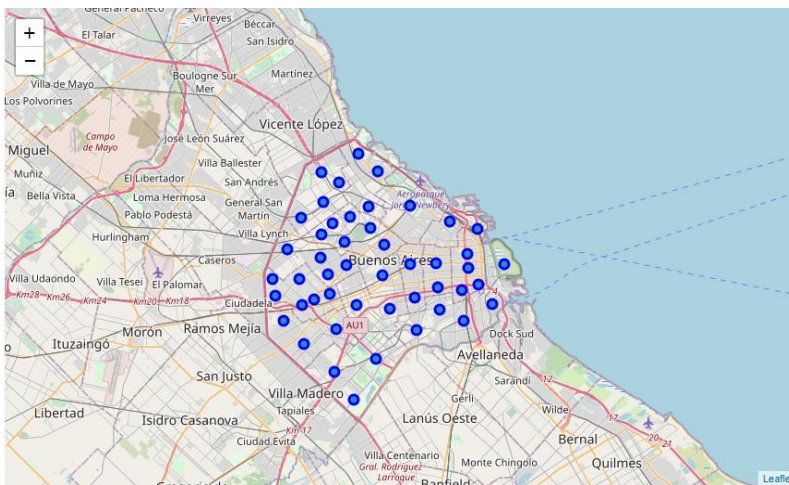
1. Agrupar los barrios con características parecidas para realizar una primera segmentación.
2. Elegido el grupo más relevante para nuestra misión , realizar dentro de ese grupo comparaciones más minuciosas de los barrios siguiendo los parámetros de lugares cercanos requeridos por la empresa.
3. Establecer la proyección inmobiliaria en base a datos históricos.
4. Verificar cantidad de hoteles 4 estrellas y relevar tarifas en las zonas segmentadas.

Datos

- Obtendremos los nombres y datos geográficos de los barrios mediante la pagina de Datasets del Gobierno de la Ciudad De Buenos Aires:
<https://cdn.buenosaires.gob.ar/datosabiertos/datasets/barrios/barrios.csv>
- Utilizaremos el API de Foursquare para la primera segmentación de barrios por similitud de categoria de lugares.
- Para la agrupación de barrios seleccionada, exploramos la gastronomía con los datos brindados por el Gobierno de la Ciudad, ya que son muy completos:
https://cdn.buenosaires.gob.ar/datosabiertos/datasets/oferta-gastronomica/oferta_gastronomica.csv
- Luego usaremos Foursquare para obtener lugares como museos, parques, centros comerciales , teatros, etc., y comparar los barrios agrupados.
- Para establecer la proyección inmobiliaria de cada zona, utilizaremos los datos de Mercado Inmobiliario del Gobierno de la Ciudad :
<https://cdn.buenosaires.gob.ar/datosabiertos/datasets/acceso-vivienda-y-mercado-inmobiliario/precio-venta-deptos.csv>
- Para segmentar los hoteles de 4 estrellas y explorar los precios utilizaremos :
https://www.estadisticaciudad.gob.ar/eyc/wp-content/uploads/2020/11/Eoh_1a%C3%B1o.xlsx

Metodologia

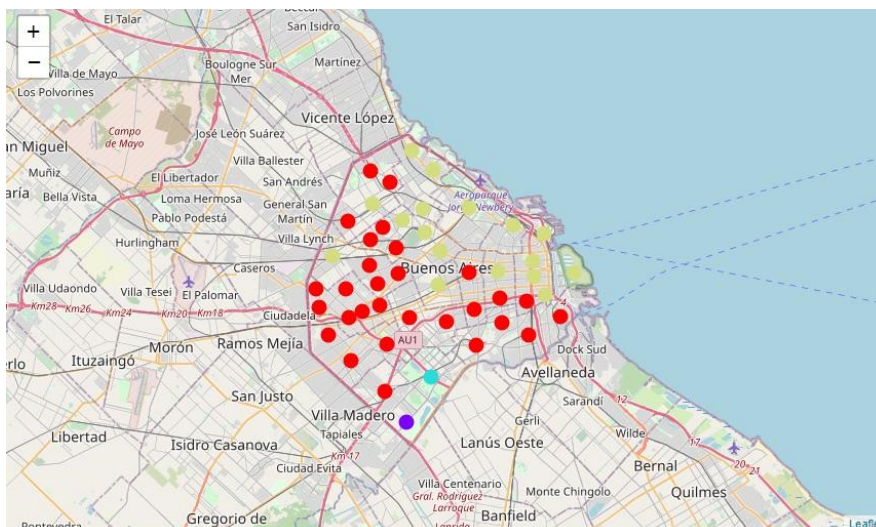
- **Primero usando Pandas se obtuvieron los datos de los barrios de la Ciudad de Bs As y se realizó limpieza y ordenamiento de los datos.**
- **Para un primer acercamiento se visualizó un mapa de la ciudad con sus correspondientes barrios usando Folium:**



- A través de la API de Foursquare, obtuvimos los lugares en un radio de 800 metros para cada barrio.
- Calculamos el promedio de ocurrencia de cada categoría por barrio y obtuvimos los 10 lugares más comunes por barrio.

	BARRIO	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	AGRONOMIA	Soccer Field	Burger Joint	Pizza Place	BBQ Joint	Bus Stop	Café	Sports Club	Grocery Store	Gym / Fitness Center	Snack Place
1	ALMAGRO	Argentinian Restaurant	Ice Cream Shop	Café	Pizza Place	Indie Theater	Cheese Shop	Sandwich Place	Gym	Concert Hall	Bar
2	BALVANERA	Café	Pizza Place	Coffee Shop	Fast Food Restaurant	Japanese Restaurant	Electronics Store	Bakery	Gym	Theater	BBQ Joint
3	BARRACAS	Argentinian Restaurant	Clothing Store	Electronics Store	General Entertainment	Auto Workshop	Gym	Restaurant	Sporting Goods Shop	Plaza	Soccer Field
4	BELGRANO	Italian Restaurant	BBQ Joint	Pizza Place	Café	Grocery Store	Coffee Shop	Cocktail Bar	Bistro	Supermarket	Sushi Restaurant

- Ejecutamos el algoritmo de Aprendizaje No Supervisado K-means para agrupar los barrios por ocurrencia de lugares, y luego lo visualizamos:



- Creamos los dataframes correspondientes a cada cluster :

	barrio	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	AGRONOMIA	Soccer Field	Burger Joint	Pizza Place	BBQ Joint	Bus Stop	Café	Sports Club	Grocery Store	Gym / Fitness Center	Snack Place
1	ALMAGRO	Argentinian Restaurant	Ice Cream Shop	Café	Pizza Place	Indie Theater	Cheese Shop	Sandwich Place	Gym	Concert Hall	Bar
3	BARRACAS	Argentinian Restaurant	Clothing Store	Electronics Store	General Entertainment	Auto Workshop	Gym	Restaurant	Sporting Goods Shop	Plaza	Soccer Field
5	BOCA	Argentinian Restaurant	Art Gallery	Café	BBQ Joint	Pizza Place	Art Museum	Seafood Restaurant	Restaurant	Coffee Shop	Gift Shop
6	BOEDO	Argentinian Restaurant	Bakery	Pizza Place	Soccer Field	Pharmacy	Burger Joint	Cheese Shop	Brewery	Breakfast Spot	Stadium
9	COGHLAN	Argentinian Restaurant	Pizza Place	BBQ Joint	Plaza	Gym / Fitness Center	Ice Cream Shop	Bakery	Brewery	Supermarket	Event Space
11	CONSTITUCION	Café	Soccer Field	German Restaurant	Pizza Place	Ice Cream Shop	Residential Building (Apartment / Condo)	Metro Station	Bakery	Restaurant	BBQ Joint
12	FLORES	Pharmacy	Café	Ice Cream Shop	Pizza Place	Coffee Shop	Bakery	Gym	Dessert Shop	Department Store	Empanada Restaurant

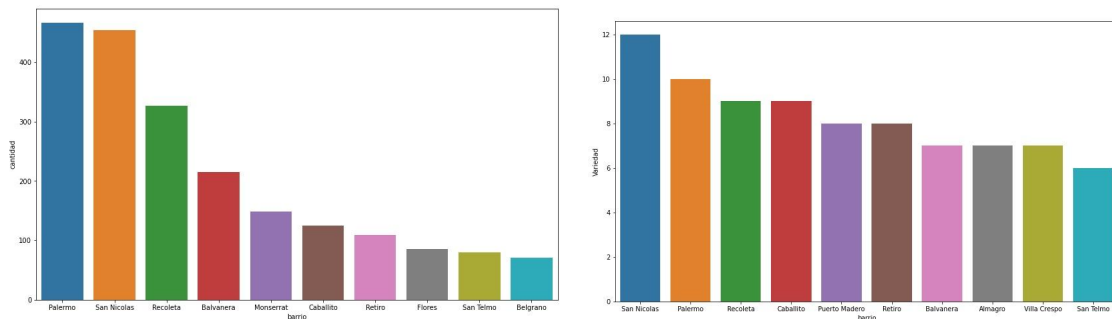
- Dado el cluster o grupo de barrios elegido por su promedio de ocurrencias para los requisitos solicitados, procedemos a la comparación de los barrios dentro de ese grupo.
 - Importamos el dataset con los datos de gastronomía por barrio, lo limpiamos de datos no deseados y lo ordenamos.
- Observamos la aparición de datos nulos en el dataframe:

categoria	cocina	barrio	comuna
0 RESTAURANTE	CUBANA	Puerto Madero	Comuna 1
1 VINERIA	NaN	San Cristobal	Comuna 3
2 RESTAURANTE	TRADICIONAL	Belgrano	Comuna 13
3 CAFE	MINUTAS	Villa Lugano	Comuna 6
4 RESTAURANTE	NaN	Palermo	Comuna 14

- Para solucionarlo encontramos la mayor ocurrencia y reemplazamos los valore nulos por dichos valores :

```
: (<pandas.core.groupby.generic.DataFrameGroupBy)
  MINUTAS      0.299107
  INTERNACIONAL 0.181818
  TRADICIONAL  0.120536
  PIZZERIA     0.071429
  PARRILLA     0.064529
  SANDWICHERTA 0.051542
```

- Creamos los dataframes con los barrios agrupados por cantidad de locales gastronómicos y por variedad de gastronomía y visualizamos:



- Utilizamos nuevamente la API de Foursquare para consultar por categorías específicas (Museos,Cines,Teatros,Parques,Centros Comerciales) para los barrios del cluster,para ello utilizamos la ID que proporciona Foursquare para las categorías.

```
Museos=Lugares(names=BsAs_hotel['barrio'], latitudes=BsAs_hotel['Latitud'], longitudes=BsAs_hotel['Longitud'], radius=1000, categoryId='4bf58dd8d48988c181941735')
```

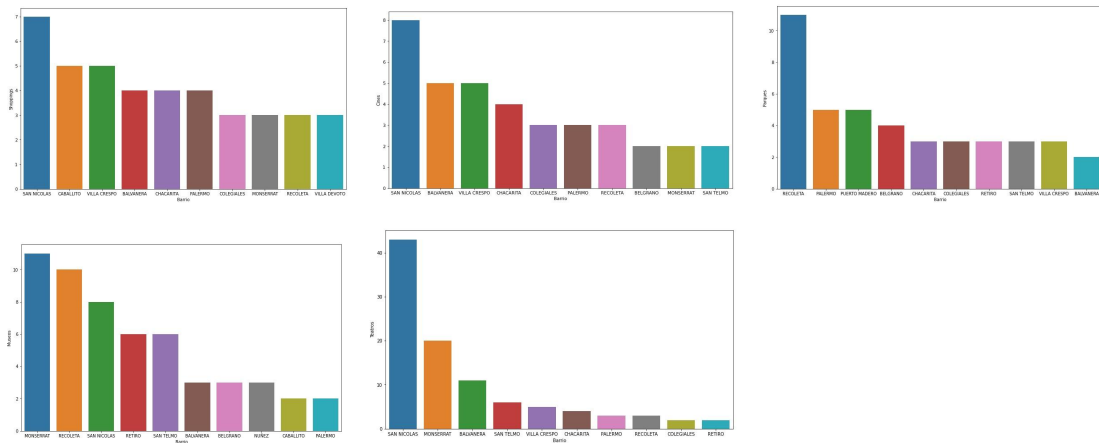
```
Parques=Lugares(names=BsAs_hotel['barrio'], latitudes=BsAs_hotel['Latitud'], longitudes=BsAs_hotel['Longitud'], radius=1000, categoryId='4bf58dd8d48988d163941735')
```

```
Cines=Lugares(names=BsAs_hotel['barrio'], latitudes=BsAs_hotel['Latitud'], longitudes=BsAs_hotel['Longitud'], radius=1000, categoryId='4bf58dd8d48988d17f941735')
```

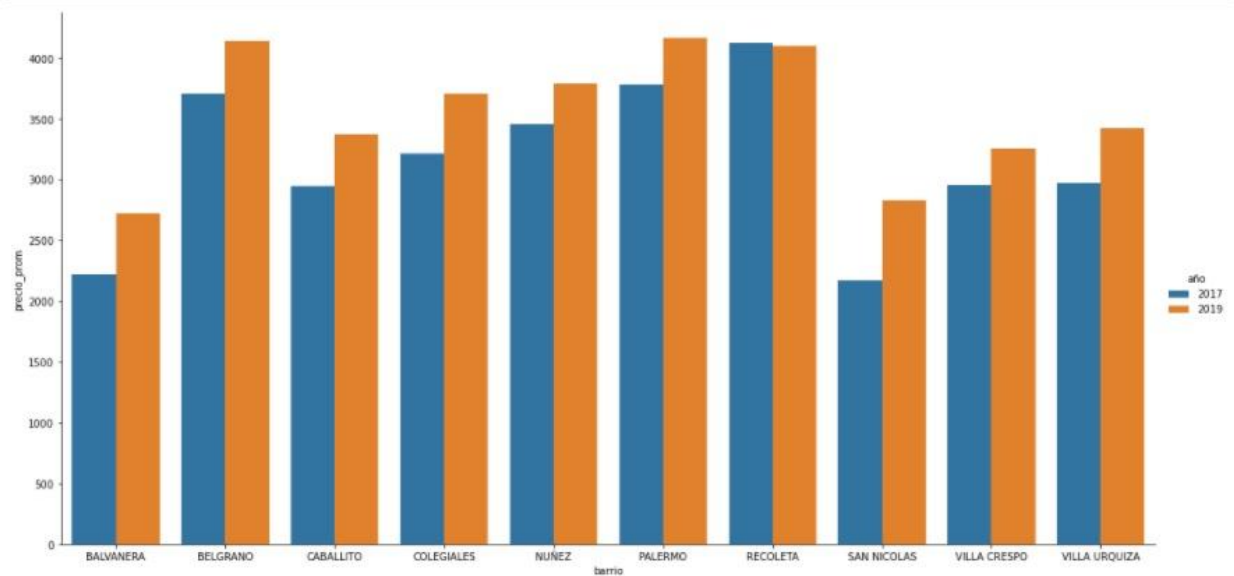
```
Teatros=Lugares(names=BsAs_hotel['barrio'], latitudes=BsAs_hotel['Latitud'], longitudes=BsAs_hotel['Longitud'], radius=1000, categoryId='4bf58dd8d48988d163941735')
```

```
Shoppings=Lugares(names=BsAs_hotel['barrio'], latitudes=BsAs_hotel['Latitud'], longitudes=BsAs_hotel['Longitud'], radius=1000, categoryId='4bf58dd8d48988d163941735')
```

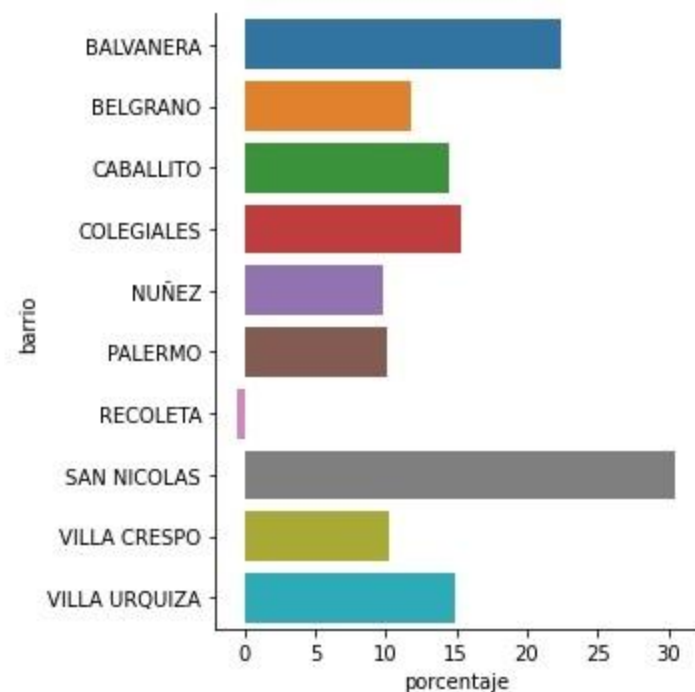
- Agrupamos por los 10 primeros barrios y visualizamos:



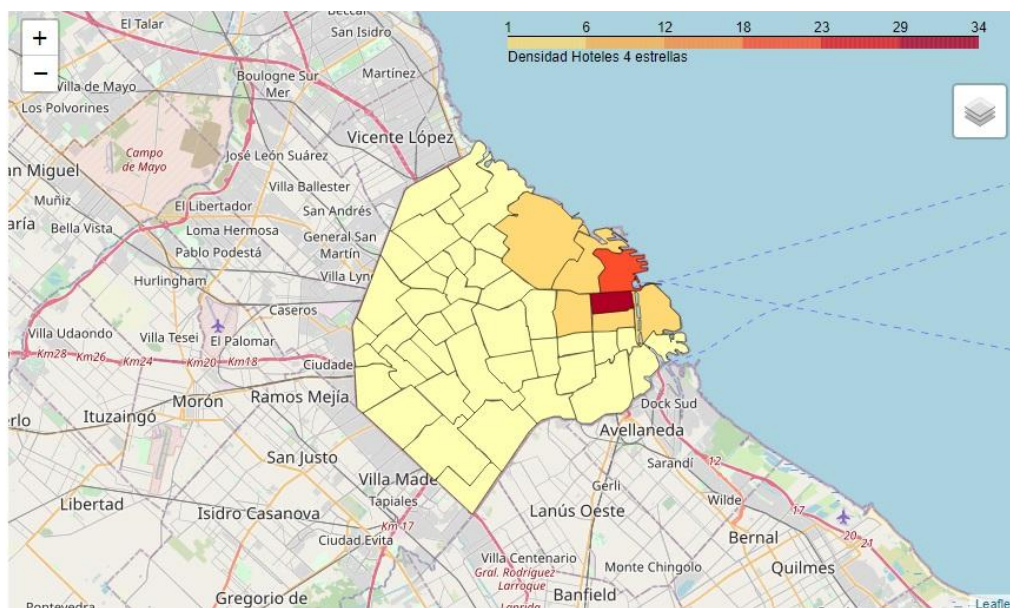
- Importamos los datos inmobiliarios provistos por el Gobierno de la Ciudad
- Ordenamos , limpiamos y filtramos los datos para los años 2017-2019 y solo para precio venta de departamentos a estrenar.
- Visualizamos:



- Reorganizamos y creamos dataframe con porcentaje de aumento entre 2017-2019 y visualizamos:



- Por último, obtuvimos los datos de hoteles en la ciudad , filtramos por la categoría de 4 estrellas para los barrios , y creamos un mapa coroplético con Folium:

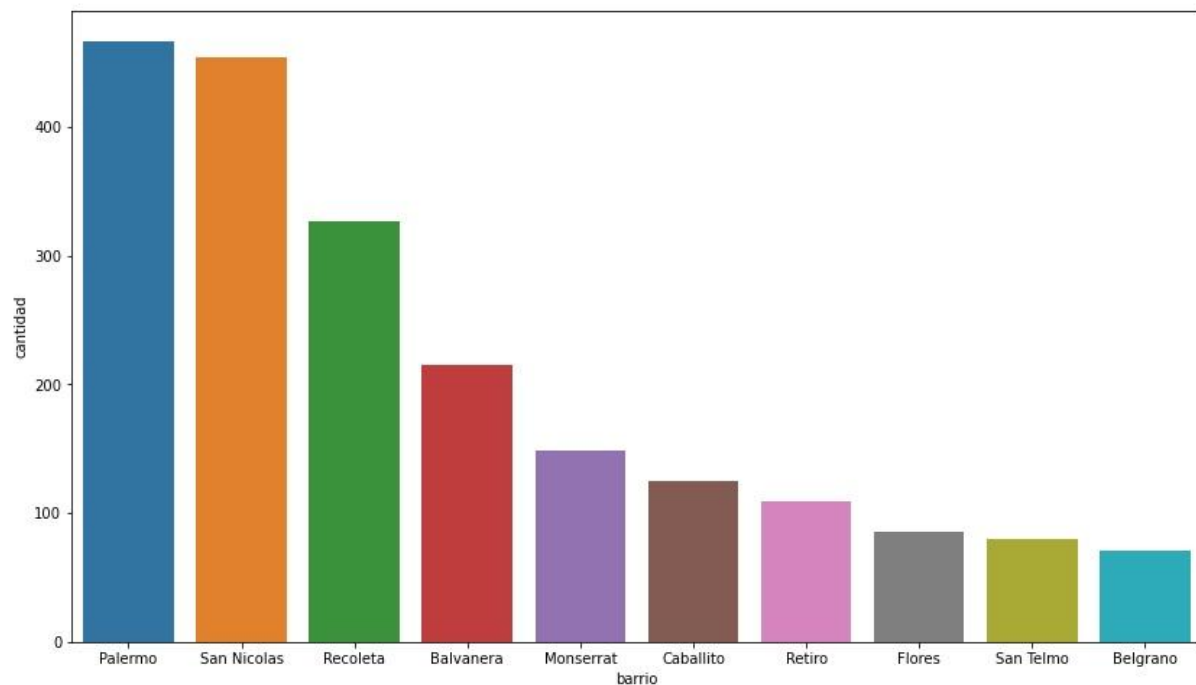


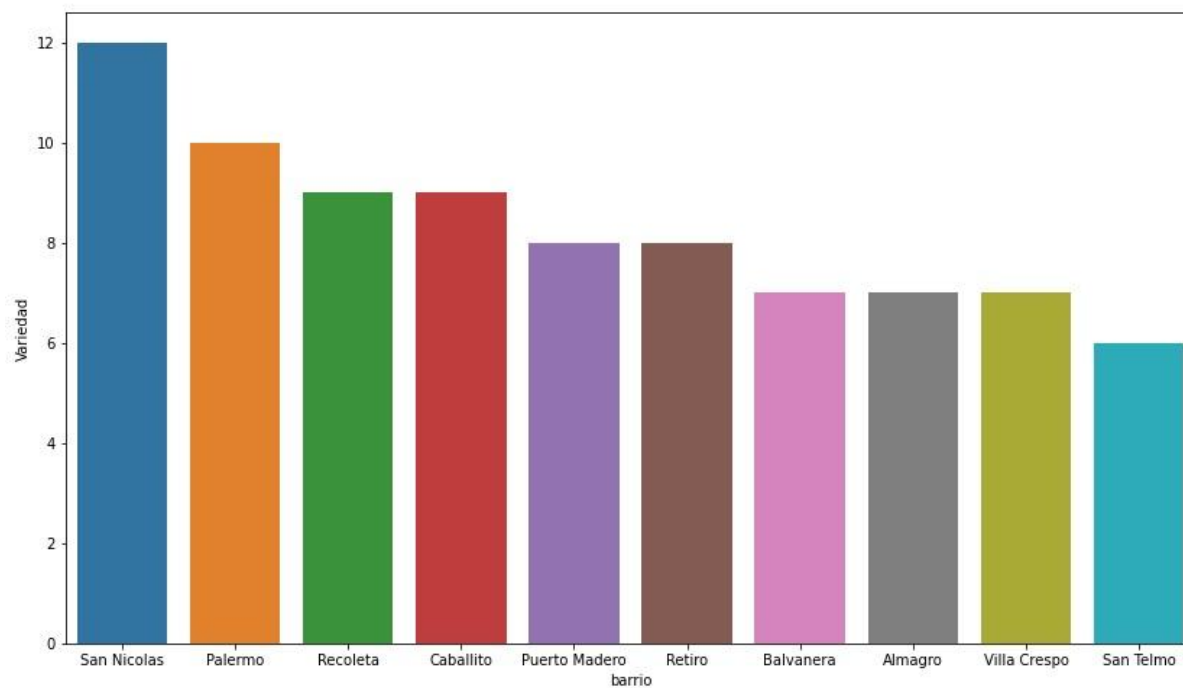
Resultados y Discusión

La primera impresión que obtuvimos es que la región noreste de la ciudad es la que comprende el mayor interés para nuestra búsqueda.

Gastronomía

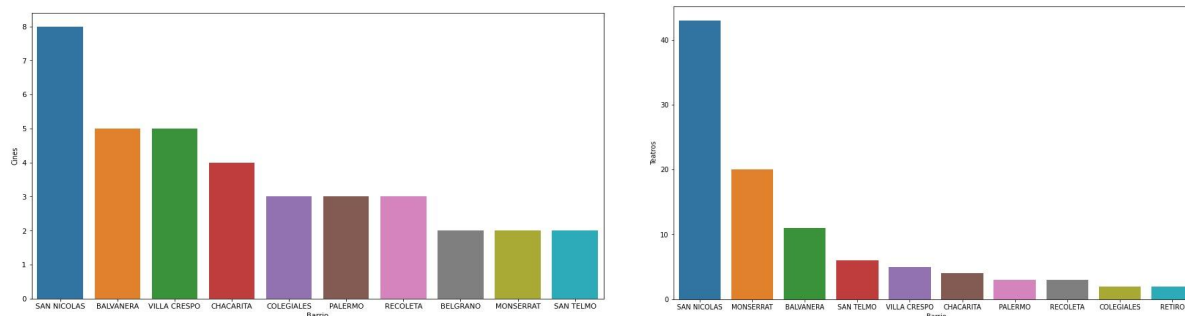
Los barrios de San Nicolás , Palermo y Recoleta resaltan en cuanto a cantidad de locales gastronómicos, siendo la variedad de categorías de cocina más uniforme en todos los barrios.



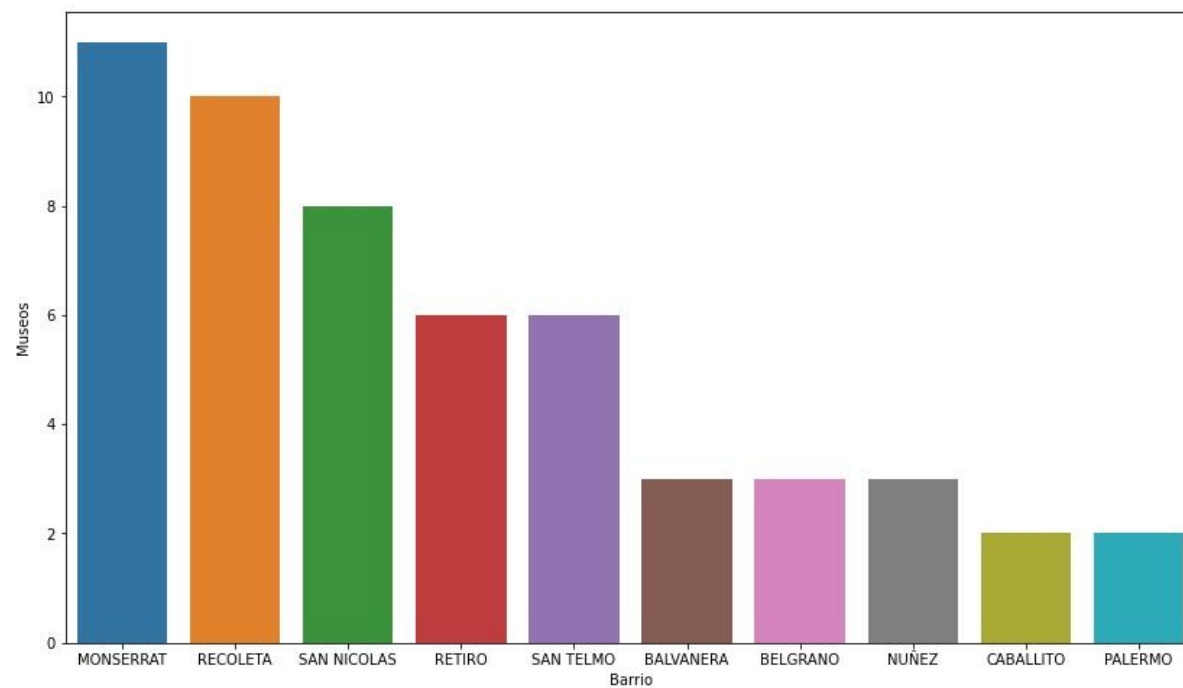
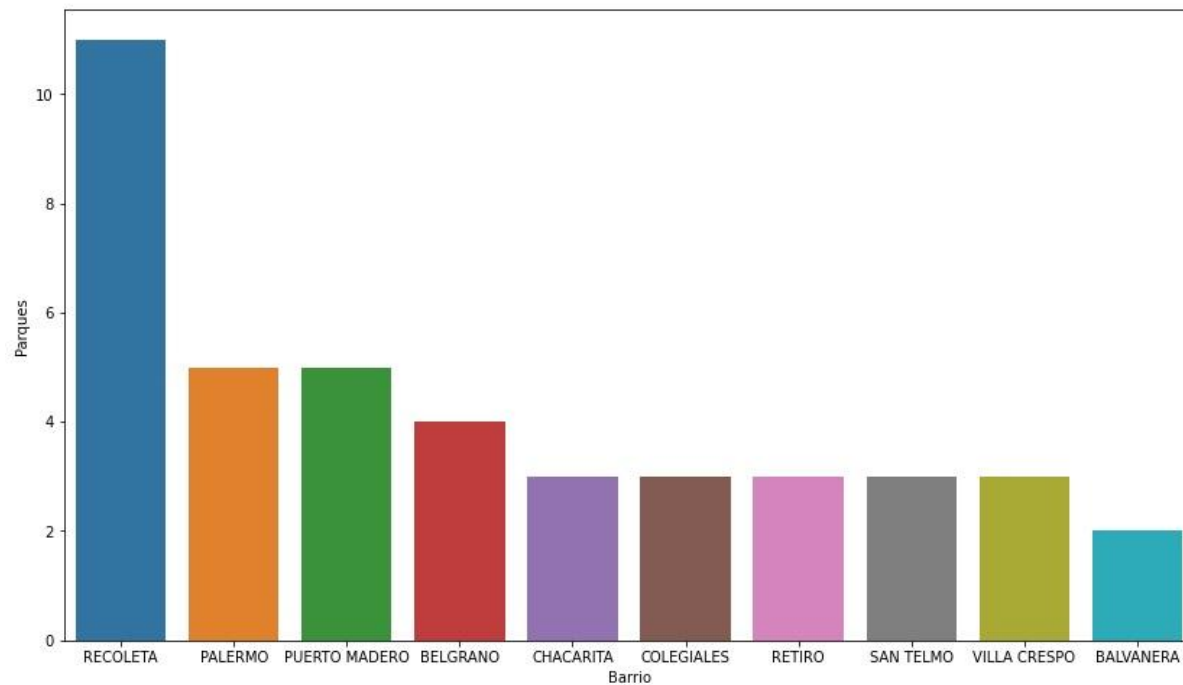


Sitios Turísticos

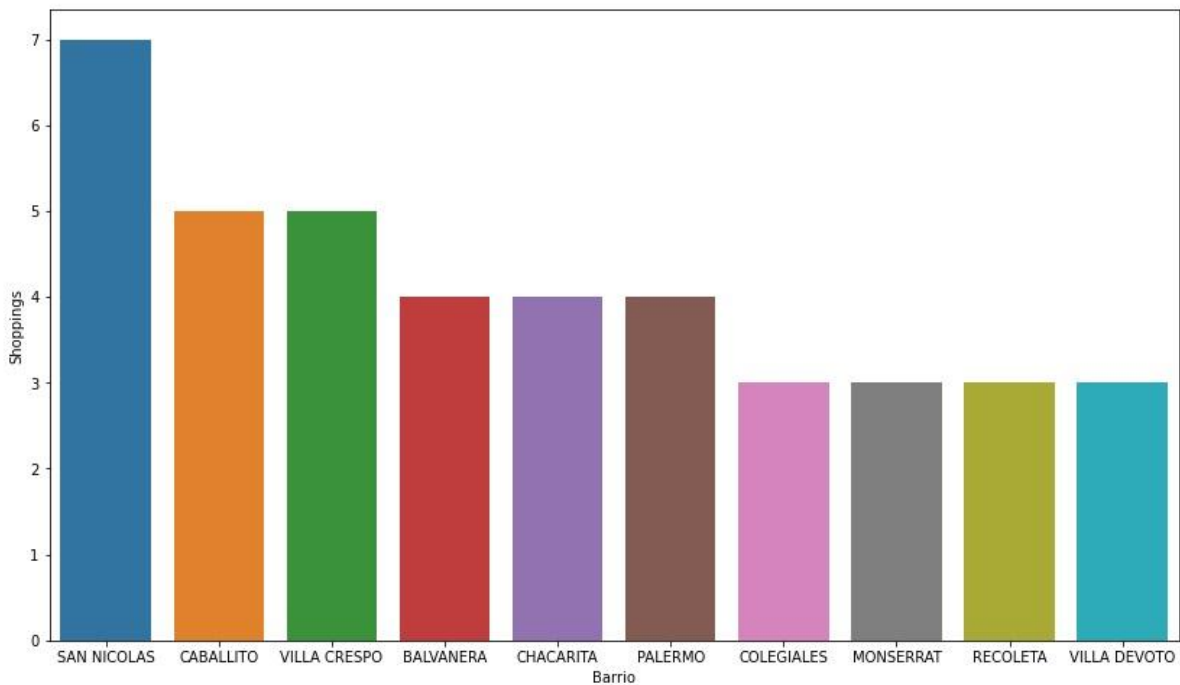
En cuanto a cines y teatros observamos una clara predominancia de el barrio San Nicolás, lo que nos hace presumir la existencia de alta actividad nocturna.



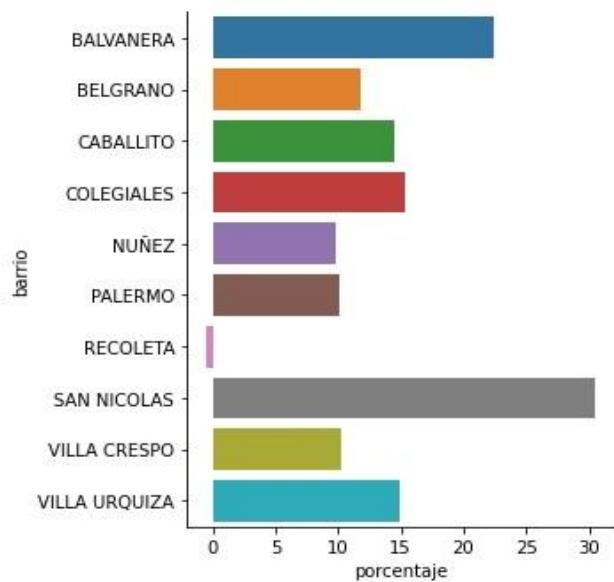
Observamos en el barrio de Recoleta una muy buena relación de parques y museos , lo que lo hace atractivo en este sentido.



Por último se observa nuevamente una predominancia de el barrio San Nicolas en cuanto a los centros comerciales se refiere :

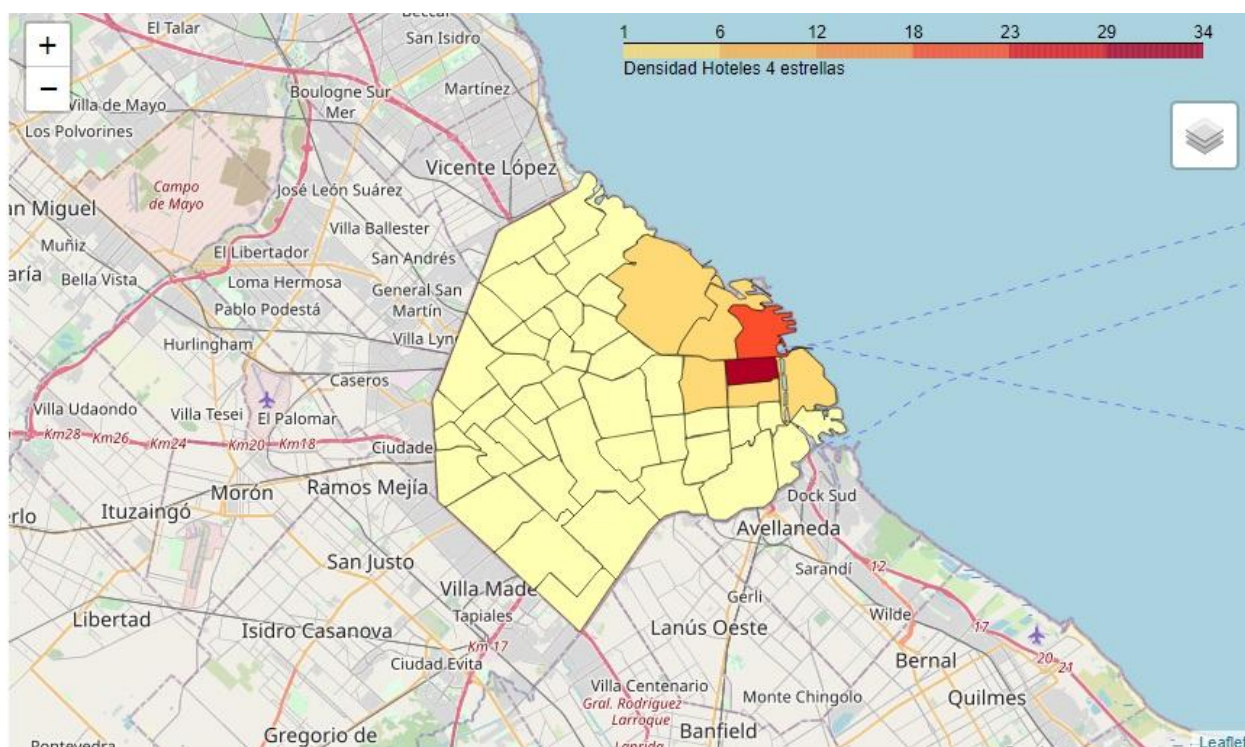


Proyeccion Inmobiliaria



Los barrios de San Nicolás y Balvanera muestran un aumento más pronunciado entre los años de 2017 y 2019 , observamos una anomalía en el barrio de Recoleta lo que podría significar un error en el origen de los datos, factor a revisar.

Densidad hoteles 4 estrellas



Se observa una muy alta densidad de hoteles 4 estrellas en el barrio de San Nicolás , seguido del barrio de Retiro , con el resto de los barrios con baja densidad.

Conclusion

En primera instancia, el barrio de San Nicolás aparece como el más indicado para la instalación del nuevo hotel, ya que muestra excelente performance en todos los indicadores propuestos, excepto en lo que refiere a densidad de hoteles de su misma categoría, allí es donde la decisión se torna más difícil.

Tomando en cuenta esto, el barrio de Recoleta en su conjunto es más homogéneo y



con buenos indicadores , suficiente razón para incluirlo en la discusión.

Dicho esto, se podría ampliar el rango de requisitos inicialmente propuesto para reafirmar lo observado.