

## Tarea Visualización Avanzada

Alumno: Paloma Holgado Pretell

### Airbnbs en Buenos Aires

Para el siguiente trabajo realizaré visualizaciones de datos sobre Airbnbs disponibles en la ciudad de Buenos Aires, Argentina. Todas las bases de datos utilizadas son datos abiertos, y los datos de airbnb fueron recomendados en la clase de Open Data. El ejercicio está basado en datos de la página [insideairbnb.com](https://data.insideairbnb.com/argentina/ciudad-autónoma-de-buenos-aires/buenos-aires/2023-09-22/visualisations/listings.csv), que pueden descargarse de este link: <https://data.insideairbnb.com/argentina/ciudad-autónoma-de-buenos-aires/buenos-aires/2023-09-22/visualisations/listings.csv>

Los datos de airbnb utilizados fueron los actualizados al 22/09/2023, serán adjuntados junto al trabajo por si al momento de la revisión se encuentra disponible sólo una versión nueva.

Luego, para construir los mapas, utilicé un geojason que facilita el gobierno de la Ciudad de Buenos Aires en este link: <https://data.buenosaires.gob.ar/dataset/barrios/resource/1c3d185b-fdc9-474b-b41b-9bd960a3806e>

Como primer paso voy a importar las librerías necesarias para llevar a cabo las visualizaciones deseadas.

```
In [1]: import pandas as pd
import seaborn as sns
import geopandas as gpd

import matplotlib.pyplot as plt
import plotly.express as px
import folium
from folium import Choropleth, Marker
from folium.plugins import HeatMap, MarkerCluster
import branca.colormap as cm
from branca.colormap import linear

import warnings
warnings.filterwarnings("ignore")
```

Ahora leo el fichero listings.csv que contiene los datos de airbnb y lo guardo como un dataframe llamado "datos".

```
In [2]: datos = pd.read_csv('listings.csv')
datos.head()
```

Out[2]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_o
0	11508	Condo in Buenos Aires · ★4.81 · 1 bedroom · 1 ...	42762	Candela	NaN	Palermo	-34.581840	-58.424150	Entire home/apt	23753		3
1	14222	Rental unit in Palermo/Buenos Aires · ★4.79 · ...	87710233	María	NaN	Palermo	-34.586170	-58.410360	Entire home/apt	11928		7
2	15074	Rental unit in Buenos Aires · 1 bedroom · 1 be...	59338	Monica	NaN	Nuñez	-34.538920	-58.465990	Private room	10501		29
3	16695	Loft in Buenos Aires · ★4.28 · 1 bedroom · 1 b...	64880	Elbio Mariano	NaN	Montserrat	-34.614390	-58.376110	Entire home/apt	17501		2
4	20062	Rental unit in Buenos Aires · ★4.93 · 2 bedroo...	75891	Sergio Damian	NaN	Palermo	-34.581848	-58.441605	Entire home/apt	35654		2

```
In [3]: datos.describe()
```

Out[3]:

	id	host_id	neighbourhood_group	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month
count	2.934600e+04	2.934600e+04	0.0	29346.000000	29346.000000	2.934600e+04	29346.000000	29346.000000	23841.000000
mean	4.871636e+17	1.761361e+08	NaN	-34.590884	-58.418264	2.478022e+04	6.090779	22.120425	1.463939
std	4.119683e+17	1.763116e+08	NaN	0.018267	0.030138	3.120830e+05	24.450390	38.742694	1.427855
min	1.150800e+04	1.342600e+04	NaN	-34.693700	-58.530890	2.200000e+02	1.000000	0.000000	0.010000
25%	3.600705e+07	2.169746e+07	NaN	-34.601380	-58.438053	1.000000e+04	2.000000	1.000000	0.410000
50%	6.834611e+17	1.069511e+08	NaN	-34.590140	-58.420150	1.365100e+04	3.000000	8.000000	1.000000
75%	8.672850e+17	2.971771e+08	NaN	-34.580511	-58.393820	2.025200e+04	4.000000	26.000000	2.100000
max	9.856615e+17	5.384406e+08	NaN	-34.534980	-58.355410	3.500256e+07	1000.000000	673.000000	19.110000

De las salidas de arriba se puede ver que el dataframe tiene información de 29.346 Airbnbs en la Ciudad de Buenos Aires y contiene datos de los nombres de los airbnbs en la plataforma, el nombre de cada uno de los dueños y el barrio en el que se encuentran, incluyendo la latitud y longitud (coordenadas) de su ubicación. Además hay datos de los precios (en precios argentinos), el tipo de alojamiento (piso entero, cuarto privado, habitación de hotel o habitación compartida) y la cantidad mínima de noches de estadía. Por último hay datos sobre las reseñas, cantidad de reseñas y la última vez que recibió una reseña.

Por lo que puedo ver en la descripción de los datos, hay outliers en los precios ya que el percentil 75 esta en 20.252 pesos argentinos y el valor más alto esta en 35.000.000. Esto puede darse por la conversión de pesos a dólares, dado que el peso se devalúa regularmente y su cotización cambia con frecuencia. Para evitar que estos datos contaminen las visualizaciones voy a filtrar el dataframe para aquellos airbnbs cuyos precios estén por debajo de 370.000 pesos argentinos (aproximadamente 1.000 dólares al tipo de cambio oficial) y trabajar con estos datos.

```
In [4]: datos_filtered = datos[datos['price'] <= 370000]
```

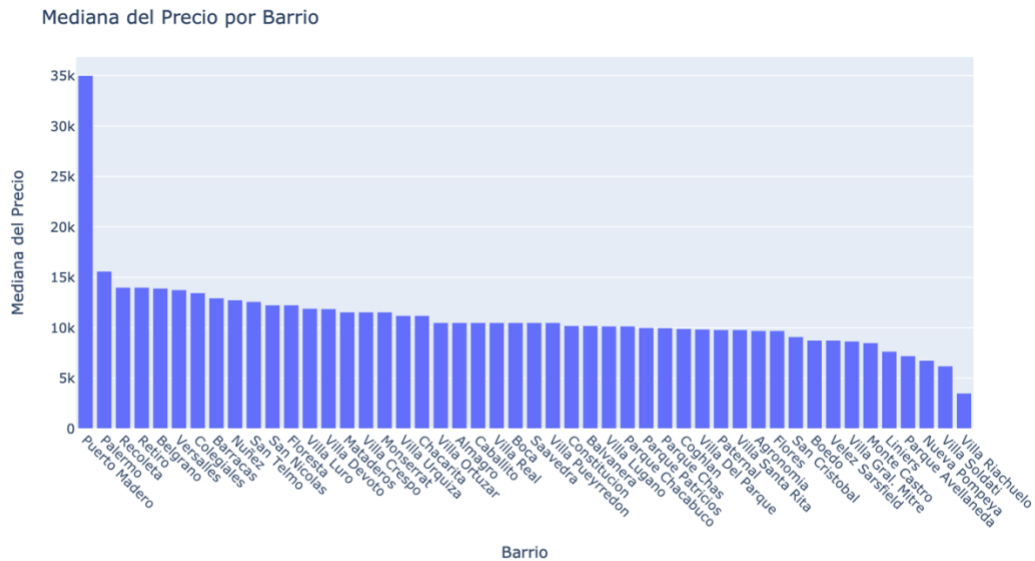
```
In [5]: datos_filtered.shape
```

```
Out[5]: (29267, 18)
```

## Gráficos

Primero voy a hacer gráficos simples para explorar los datos. Voy a empezar con un gráfico de barras para ver los barrios con los airbnbs más caros. Para hacer este gráfico voy a utilizar la mediana del precio dado que por el describe hecho arriba se puede ver que hay mucha diferencia entre los precios mínimos y máximos, para no tener problemas con estos outliers utilizo la mediana.

```
In [6]: mediana_precio_barrio = datos_filtered.groupby('neighbourhood')['price'].median().sort_values(ascending=False).reset_index()
fig = px.bar(mediana_precio_barrio, x='neighbourhood', y='price')
fig.update_layout(title='Mediana del Precio por Barrio', xaxis_title='Barrio', yaxis_title='Mediana del Precio')
fig.update_xaxes(tickangle=45)
fig.show()
```



Del gráfico de arriba podemos ver que Puerto Madero es el barrio con precios más altos, con una mediana del precio en 35.000 pesos la noche (95 dólares aproximadamente), lo cual tiene sentido porque es el barrio más moderno de Buenos Aires. Le siguen barrios como Palermo, Recoleta y Retiro que son los barrios más turísticos de la ciudad, con una mediana del precio en torno a los 15.000 pesos la noche.

En el otro extremo se encuentra el barrio de Villa Riachuelo con los precios más bajos. Los últimos barrios son periféricos, quedan muy alejados del centro de la ciudad.

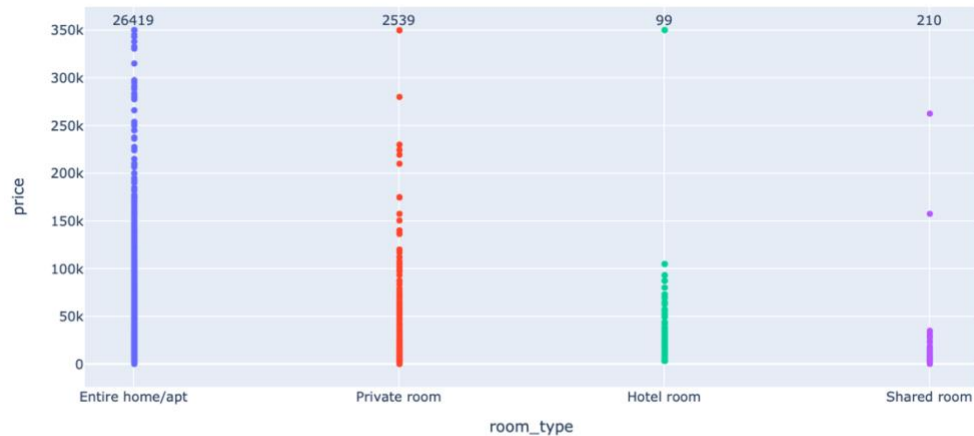
Además voy a mirar la distribución de los precios según cada tipo de alojamiento: piso entero, cuarto privado, habitación de hotel o habitación compartida.

```
In [7]: fig = px.scatter(
    datos_filtered,
    x='room_type',
    y='price',
    color='room_type',
    title='Distribución de Precios según Tipo de Alojamiento'
)

room_type_counts = datos_filtered['room_type'].value_counts()

for room_type, count in room_type_counts.items():
    fig.add_annotation(
        x=room_type,
        y=fig.data[0].y.max(),
        text=str(count),
        showarrow=False,
        font=dict(size=12),
        xanchor='center',
        yanchor='bottom',
    )
fig.update_layout(showlegend=False)
fig.show()
```

Distribución de Precios según Tipo de Alojamiento



Del scatter plot se puede ver que los precios máximos de las categorías de apartamentos, habitaciones privadas y habitaciones de hotel están en torno a los 350.000 pesos la noche (aproximadamente 1.000 dólares). En cuanto a las habitaciones privadas se puede ver que la mayoría de los alojamientos alcanzan un máximo de precio por debajo de los 175.000 pesos la noche y los hoteles por debajo de los 105.000 pesos, a medida que sube el precio a partir de esos puntos los alojamientos se hacen más escasos, siendo sus máximos outliers.

Los precios máximos de las habitaciones compartidas llegan en su mayoría hasta los 35.000 pesos, dado que este tipo de alojamiento suele ser en hostels y menos privado. Hay además algunos outliers que llegan hasta 260.000 pesos.

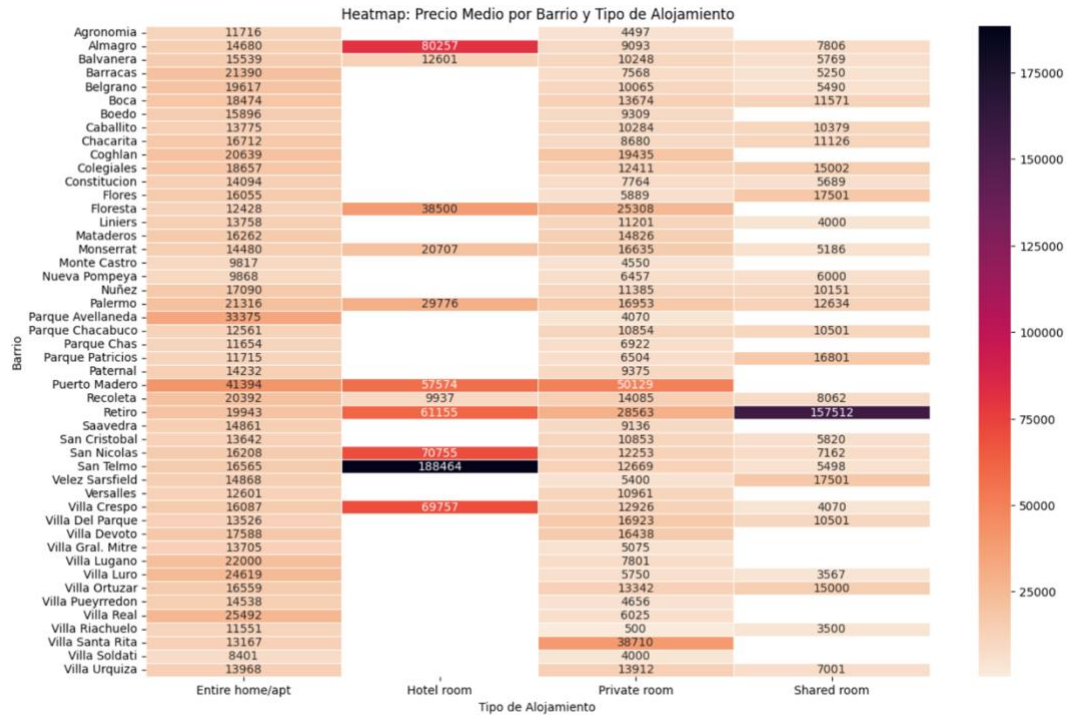
Por último se puede ver que los apartamentos son el alojamiento que más abunda, habiendo 26.419 apartamentos, y el que menos abunda son los hoteles con solo 99 disponibles en la plataforma de Airbnb.

A continuación voy a hacer un heatmap según los barrios y los tipos de alojamiento. Los colores del heatmap van a estar definidos por la media de los precios para cada combinación de barrio y tipo de alojamiento.

```
In [8]: heatmap_data = datos_filtered.pivot_table(index='neighbourhood', columns='room_type', values='price', aggfunc='mean')

plt.figure(figsize=(15, 10))
sns.heatmap(heatmap_data, cmap='rocket_r', annot=True, fmt=".0f", linewidths=.5)
plt.xlabel('Tipo de Alojamiento')
plt.ylabel('Barrio')
plt.title('Heatmap: Precio Medio por Barrio y Tipo de Alojamiento')

plt.xticks(rotation=0)
plt.show()
```



Del heatmap se puede ver que la categoría con la media de precio más alta es la de habitaciones de hotel en San Telmo, con un valor de 188.464 pesos argentinos la noche (510 dólares aproximadamente). San Telmo es un barrio antiguo en Buenos Aires, muy turístico, lleno de parrillas y atracciones por conocer, tiene sentido que haya mucha demanda por alojarse en hoteles en esta zona y que por ende los precios sean altos.

En segundo lugar están las habitaciones compartidas en Retiro, zona donde está la estación de autobuses de larga distancia y la estación central de trenes. Finalmente, a estas dos categorías le siguen en su mayoría las habitaciones de hotel en otras zonas, que por los amenities que vienen incluidos en la tarifa tiene sentido que sean la opción más cara.

Otra observación que se puede hacer a partir de este heatmap es que hay apartamentos y habitaciones privadas disponibles en todos los barrios mientras que no sucede lo mismo para las habitaciones compartidas ni para las habitaciones de hotel, estas últimas estando disponibles solo en 11 barrios.

## Mapas

Voy a hacer un mapa de Buenos Aires que marque cada uno de los alojamientos según sus coordenadas. Además voy a asignar a cada punto en el mapa un color según el tipo de alojamiento que son: los apartamentos en rojo, los hoteles en amarillo, las habitaciones privadas en verde y las habitaciones compartidas en azul. Le agrego un tooltip a cada uno de los markers que diga el nombre del alojamiento y un pop-up que al hacer click en el punto despliegue el nombre, el tipo, el precio y el barrio del alojamiento.



```
In [9]: mapa = folium.Map(location=[datos_filtered['latitude'].mean(), datos_filtered['longitude'].mean()], tiles="cartodbpositron")

room_type_colors = {
    'Entire home/apt': '#FF0000',
    'Hotel room': '#FFFF00',
    'Private room': '#0000FF',
    'Shared room': '#00FF00'
}

title_html = """
<h3 align="left" style="font-size:16px"><b>Mapa de Tipos de Alojamientos en Buenos Aires</b></h3>
"""
mapa.get_root().html.add_child(folium.Element(title_html))

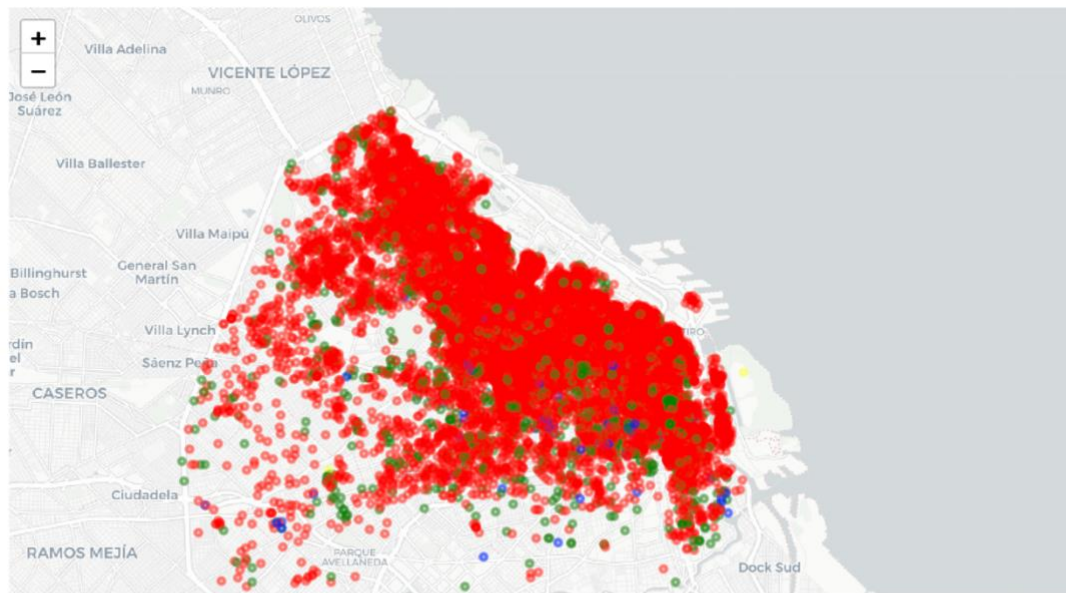
for index, row in datos_filtered.iterrows():
    circle_marker = folium.CircleMarker(
        location=[row['latitude'], row['longitude']],
        radius=3,
        color=room_type_colors.get(row['room_type'], '#808080'),
        fill=True,
        fill_color=room_type_colors.get(row['room_type'], '#808080'),
        fill_opacity=0.6,
        tooltip=row['name']
    )

    listing_info = f"Nombre: {row['name']}<br>Tipo: {row['room_type']}<br>Precio: {row['price']}<br>Barrio: {row['neighborhood']}"
    popup = folium.Popup(listing_info, max_width=300)
    circle_marker.add_child(popup)
    circle_marker.add_to(mapa)

mapa.save('listings_map.html')

mapa
```

Mapa de Tipos de Alojamientos en Buenos Aires



Del mapa de arriba se puede apreciar que la mayoría de los alojamientos son apartamentos y que la mayor concentración de alojamientos se encuentra hacia el este de la Ciudad de Buenos Aires, en torno al Río de la Plata. Más adelante veremos en más detalle la cantidad de alojamientos en cada barrio.

Ahora voy a utilizar el archivo geojason de Buenos Aires para poder hacer mapas que contengan la separación de los barrios. Primero importo el file y lo nombro "geo".

```
In [10]: geo = gpd.read_file(r'barrios.geojson')
```

```
In [11]: geo.head()
```

```
Out[11]:
```

	BARRIO	COMUNA	PERIMETRO	AREA	OBJETO	geometry
0	CHACARITA	15.0	7724.852955	3.115707e+06	BARRIO	MULTIPOLYGON (((-58.45282 -34.59599, -58.45366...
1	PATERNAL	15.0	7087.513295	2.229829e+06	BARRIO	MULTIPOLYGON (((-58.46558 -34.59656, -58.46562...
2	VILLA CRESPO	15.0	8131.857075	3.615978e+06	BARRIO	MULTIPOLYGON (((-58.42375 -34.59783, -58.42495...
3	VILLA DEL PARQUE	11.0	7705.389797	3.399596e+06	BARRIO	MULTIPOLYGON (((-58.49461 -34.61487, -58.49479...
4	ALMAGRO	5.0	8537.901368	4.050752e+06	BARRIO	MULTIPOLYGON (((-58.41287 -34.61412, -58.41282...

El primer mapa que voy a llevar a cabo va a ilustrar la mediana de los precios por barrios, pintando de un color más oscuro aquellos barrios con precios más altos y más claro para aquellos con precios más bajos. Esta visualización pretende ilustrar la información del primer gráfico de barras llevado a cabo.

El mapa va a incluir marcadores con tooltips que desplieguen el nombre del barrio y la mediana de los precios de los airbnbs en cada uno.

```
In [12]: geo = gpd.read_file('barrios.geojson')

datos_filtered['neighbourhood'] = datos_filtered['neighbourhood'].str.upper()

merged_data = geo.merge(datos_filtered, how='left', left_on='BARRIO', right_on='neighbourhood')

median_prices = merged_data.groupby('BARRIO')['price'].median().reset_index()

barrios_geo = geo.merge(median_prices, how='left', left_on='BARRIO', right_on='BARRIO')

mapa = folium.Map(location=[barrios_geo.geometry.centroid.y.mean(), barrios_geo.geometry.centroid.x.mean()], zoom_start=15)

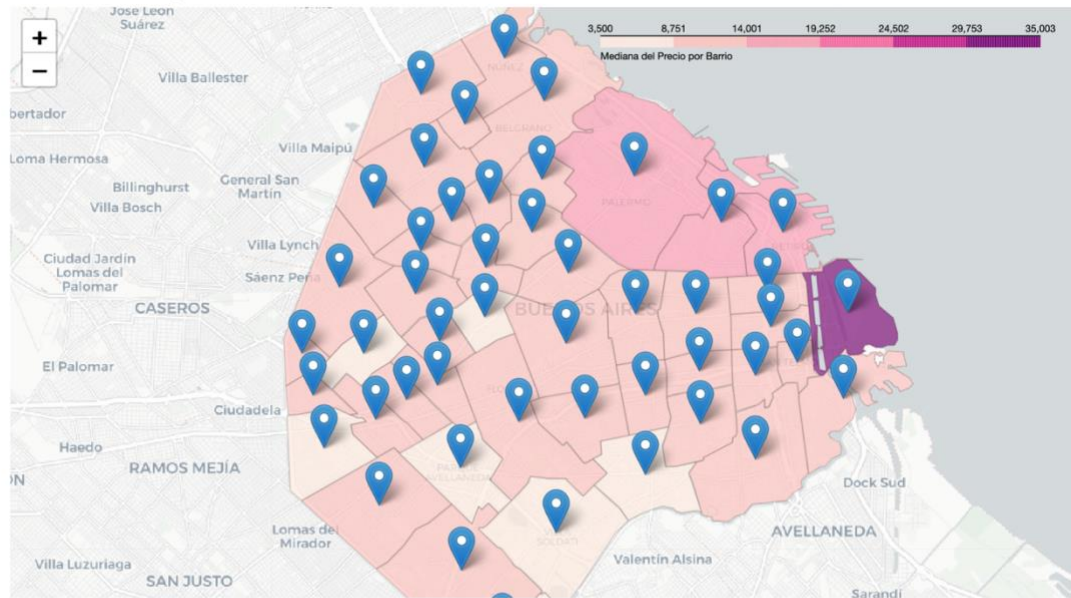
Choropleth(
    geo_data=geo,
    data=median_prices,
    columns=['BARRIO', 'price'],
    key_on='feature.properties.BARRIO',
    fill_color='RdPu',
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='Mediana del Precio por Barrio'
).add_to(mapa)

for idx, row in barrios_geo.iterrows():
    tooltip_text = f"Barrio: {row['BARRIO'].capitalize()}<br>Mediana del Precio: ${row['price']:.2f}"
    Marker([row.geometry.centroid.y, row.geometry.centroid.x], tooltip=tooltip_text).add_to(mapa)

title_html = """
<h3 align="left" style="font-size:16px"><b>Mediana del Precio por Barrio</b></h3>
"""
mapa.get_root().html.add_child(folium.Element(title_html))

mapa.save('map_with_choropleth.html')
mapa
```

### Mediana del Precio por Barrio



Para el siguiente mapa pretendo ilustrar los barrios con mayor cantidad de alojamientos. En este caso voy a utilizar un heatmap que marque en rojo las áreas con mayor concentración de airbnbs y en azul aquellas con menor concentración. Además el mapa incluye tooltips en cada barrio que describen el nombre del barrio y la cantidad de alojamientos que hay en él al apoyar el cursor sobre cada una de las áreas.

```
In [13]: geo = gpd.read_file('barrios.geojson')

datos_filtered['neighbourhood'] = datos_filtered['neighbourhood'].str.upper()

merged_data = geo.merge(datos_filtered, how='left', left_on='BARRIO', right_on='neighbourhood')

count_listings = merged_data.groupby('BARRIO')['id'].count().reset_index()
count_listings.columns = ['BARRIO', 'count_listings']

barrios_geo = geo.merge(count_listings, how='left', left_on='BARRIO', right_on='BARRIO')

map_center = [merged_data['latitude'].mean(), merged_data['longitude'].mean()]
mapa = folium.Map(location=map_center, tiles="cartodbpositron", zoom_start=12)

heat_data = [[row['latitude'], row['longitude']] for index, row in merged_data.iterrows()]

feature = folium.features.GeoJson(
    barrios_geo,
    tooltip=folium.GeoJsonTooltip(fields=['BARRIO', 'count_listings'],
                                   aliases=['Barrio', 'Cantidad de Alojamientos'],
                                   localize=True,
                                   sticky=True,
                                   labels=True),
    style_function=lambda x: {
        'color': 'grey',
        'weight': 0.8,
        'fillOpacity': 0.0
    }
)

mapa.add_child(feature)
mapa.keep_in_front(feature)

HeatMap(heat_data, radius=15).add_to(mapa)

title_html = '''
    <h3 align="left" style="font-size:16px"><b>Heatmap: Concentración de Airbnbs en Buenos Aires</b></h3>
'''

mapa.get_root().html.add_child(folium.Element(title_html))

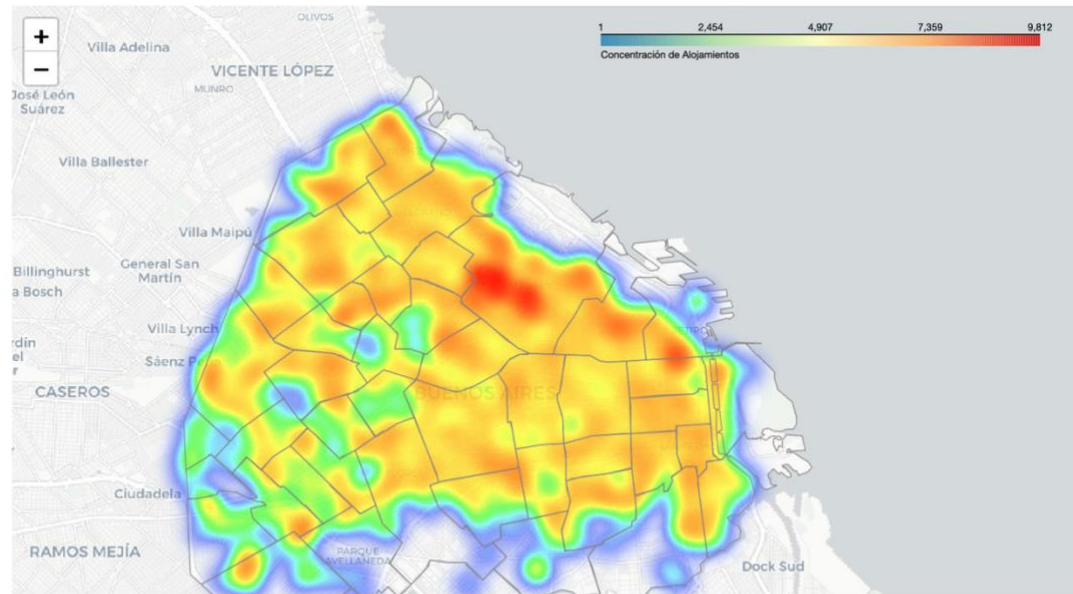
min_value = 1
max_value = barrios_geo['count_listings'].max()

colormap = cm.linear.Spectral_05.scale(min_value, max_value)
colormap.colors = list(reversed(colormap.colors))
colormap.caption = "Concentración de Alojamientos"

mapa.add_child(colormap)

mapa.save('airbnb_heatmap.html')
mapa
```

### Heatmap: Concentración de Airbnbs en Buenos Aires



En el mapa se puede ver que el barrio con más alojamientos es Palermo, duplicando la cantidad de alojamientos que hay en Recoleta que se queda con el segundo puesto. Estos son los barrios donde se encuentran la mayoría que las atracciones turísticas de Buenos Aires. Además, se puede ver que a medida que nos alejamos del centro bajan la cantidad de alojamientos, lo cual es lógico porque los turistas van a buscar quedarse lo más cerca del centro posible y la oferta acompaña la demanda.

### Conclusiones

Hay registro de 29.346 airbnbs en la Ciudad de Buenos Aires. De los 29.267 alojamientos analizados luego de filtrar por aquellos con precios por debajo de los 1000 dólares se puede concluir lo siguiente:

- El barrio más caro para alojarse es Puerto Madero, con una mediana de precio de 35.000 pesos argentinos por noche. A Puerto Madero le siguen Palermo, Recoleta y Retiro, los barrios más turísticos de la ciudad.
- Los barrios más baratos para alojarse como Villa Riachuelo son barrios periféricos, alejados del centro de la ciudad.
- Los alojamientos más caros y los que más abundan son los apartamentos, habiendo 26.419 disponibles en la plataforma de Airbnb.
- Las habitaciones compartidas son los alojamientos con los precios máximos más bajos, llegando en su mayoría hasta los 35.000 pesos argentinos por noche.
- La categoría con la media de precio más alta es la de habitaciones de hotel en San Telmo, con un valor de 188.464 pesos argentinos la noche. Le siguen las habitaciones compartidas en la zona de Retiro.
- Hay apartamentos y habitaciones privadas disponibles en todos los barrios mientras que no sucede lo mismo para las habitaciones compartidas ni para las habitaciones de hotel, estas últimas estando disponibles solo en 11 barrios.
- En el mapa se puede ver que el barrio con más alojamientos es Palermo, duplicando la cantidad de alojamientos que hay en Recoleta que se queda con el segundo puesto.