Capstone Project - The Battle of Neighborhoods (Week 2)

BARCELONA RENTING HOUSE.

1. A full report consisting of all of the following components (15 marks):

- Conclusion section where you conclude the report.

1. Introduction where you discuss the business problem and who would be interested in this project.

Currently I am living in the UK. I am planning on moving back to Barcelona in a short future.
In this scenario, I really need machine learning tools in order to assist me to make a wise and effective decision about which neighborhood is the best to rent a house.

I am going to cluster Barcelona neighborhoods in order to analyse how many and what kind of venues and the current average price of real estate.

The neighborhood will be segmented according to amenities and essential facilities surrounding such i.e. grocery stores, restaurants, pubs..

Background

CBRE's Global Living report on the housing market in 35 of the world's most important cities asserts that house prices increased the most in Barcelona last year, up 16.9%, with Madrid not far behind in fourth place with an increase of 10.2%. This comes as a surprise to locals in the business.

The problem with all international housing rankings like this one from CBRE that compares Spain to other countries is that the source data is not very reliable when it comes to house prices in Spain. CBRE cites the Spanish Notaries' Association as the data source, but in my experience the Notaries' figures are highly volatile, and are revised significantly months later. The statistics provided by the notaries are user-unfriendly, which makes it difficult to delve into them and work out what's going on, but they never seem to match the reality described by property professionals.

No other source I can find thinks that Barcelona house prices rose by 16.9%. According to data from Barcelona City Hall, house prices in terms of €/m2 (built) were up 4.5% last year to 4,182€/m2, admittedly with new house prices up 17.7% to 4,619€/m2, but the much bigger resale market was only up 2.7% to 4,120€/m2.

Reports from agents at the coal face of the property market in Barcelona tell a similar story of low or stable Barcelona house prices last year. Alex Vaughan of Barcelona-based agents Lucas Fox reports an overall increase of 1.4% vs 2017, though the key Eixample district segment rose even less, but just 0.5% "That's closed prices," explains Alex. "Obviously the market was much slower last year, especially prime. This year has started very well, with the number of offers close to where they were in 2017 but I would say people are now willing to pay 10% less than they were before October 2017."

2. Data where you describe the data that will be used to solve the problem and the source of the data

The main resource has been
https://www.bcn.cat/estadistica/castella/dades/timm/ipreus/hab2mave/evo/t2mab.htm

Estadística i Difusió de Dades

Inicio > Cifras de la ciudad > Estadísticas urbanísticas > El mercado inmobiliario de Barcelona > Precio de oferta de las viviendas de segunda mano > Cifras evolutivas. 2001-2020

Seleccionar tabla: Precio medio de oferta en los barrios (€/m2). 2013-2019

1. Oferta de viviendas de segunda mano en venta en Barcelona.

3. Precio medio de oferta en los barrios (€/m2). 2013-2019

Dto.	Barrios	2013	2014	2015	2016	2017	2018	...	2019	2020
	BARCELONA	3.019	3.188	3.392	3.879	4.284	4.344		4.115	4.111
1	1. el Raval	2.614	2.404	2.775	3.251	4.029	4.034		4.591	3.719
1	2. el Barri Gòtic	3.811	3.791	4.236	4.813	4.884	4.660		3.811	4.707
1	3. la Barceloneta	4.212	4.168	4.043	4.683	5.165	4.815		4.849	4.906
1	4. Sant Pere, Santa Caterina i la Ribera	3.534	3.682	3.827	4.501	5.152	4.689		4.772	4.818
2	5. el Fort Pienc	3.038	3.022	3.228	4.012	4.107	4.500		4.250	4.250
2	6. la Sagrada Família	3.029	2.959	3.157	3.746	4.209	4.202		4.173	4.092
2	7. la Dreta de l'Eixample	4.296	4.528	4.961	5.949	6.332	6.128		5.514	5.726
2	8. l'Antiga Esquerra de l'Eixample	3.521	3.551	3.999	4.747	5.091	5.081		5.197	5.451
2	9. la Nova Esquerra de l'Eixample	3.158	3.292	3.340	4.085	4.465	4.797		4.634	4.688
2	10. Sant Antoni	2.926	3.000	3.369	3.817	4.591	4.530		4.412	4.355
3	11. el Poble Sec-AEI Parc Montjuïc	2.495	2.518	2.815	2.771	3.936	4.083		3.911	3.854
3	12. la Marina del Prat Vermell-AEI Zona Franca	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.		n.d.	1.905
3	13. la Marina de Port	2.152	2.080	2.174	2.348	2.723	2.879		2.819	2.920
3	14. la Font de la Guatlla	n.d.	2.580	2.582	n.d.	3.510	3.457		3.893	3.516
3	15. Hostafrancs	n.d.	2.719	2.742	2.970	3.912	3.398		3.915	3.697
3	16. la Bordeta	n.d.	2.323	2.361	2.829	3.171	3.153		3.114	3.239
3	17. Sants-Badal	2.575	2.392	2.607	3.127	3.469	3.429		3.183	3.307
3	18. Sants	2.633	2511	2.816	3.181	3.666	3.556		3.642	3.624
4	19. les Corts	3.597	3.712	3.825	4.469	4.821	4.650		4.469	4.647

In [22]:
```python
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head()
```

Out[22]:

	name	categories	lat	lng
0	Chulapio	Cocktail Bar	41.379264	2.165905
1	La Robadora	Gastropub	41.379500	2.170463
2	Arume	Spanish Restaurant	41.378953	2.166008
3	A Tu Bola	Tapas Restaurant	41.380096	2.169054
4	La Monroe	Spanish Restaurant	41.378795	2.170692

In [12]:
```python
df_coor=[]

for index, item in df.iterrows():
    address='Barcelona '+np_df[index][0]
    #print(address)

    geolocator = Nominatim(user_agent="ny_explorer")
    location = geolocator.geocode(address)
    latitude = location.latitude
    longitude = location.longitude
    df_coor.append({'Barri':np_df[index][0],'Latitude':latitude,'Longitude':longitude})
    print('The geograpical coordinate of {} are {}, {}.'.format(np_df[index][0],latitude, longitude)
df_coordinates=pd.DataFrame(df_coor)
```

```
The geograpical coordinate of  . el Raval are 41.3795176, 2.1683678.
The geograpical coordinate of  . el Barri Gòtic are 41.3833947, 2.1769119.
The geograpical coordinate of  . la Barceloneta are 41.3806533, 2.1899274.
The geograpical coordinate of  . Sant Pere, Santa Caterina i la Ribera are 41.372251, 2.1775315.
The geograpical coordinate of  . el Fort Pienc are 41.3959246, 2.1823245.
The geograpical coordinate of  . la Sagrada Família are 41.4034789, 2.1744103330097055.
The geograpical coordinate of  . la Dreta de l'Eixample are 41.39412395, 2.166470697643847.
The geograpical coordinate of  . l'Antiga Esquerra de l'Eixample are 41.38876465, 2.156597362161013.
The geograpical coordinate of  . la Nova Esquerra de l'Eixample are 41.3828159, 2.1499663437362098.
The geograpical coordinate of  . Sant Antoni are 41.3784116, 2.1617677.
The geograpical coordinate of  . el Poble Sec-AEI Parc Montjuïc are 41.3687898, 2.1631845.
The geograpical coordinate of  . la Marina de Port are 41.3602964, 2.1375842.
The geograpical coordinate of  . la Font de la Guatlla are 41.3707824, 2.1446756.
The geograpical coordinate of  . Hostafrancs are 41.3750877, 2.1429334.
```

Web barcelona ajuntament:
Filters applied and snapp

To explore and target recommended locations across different venues according to the presence of amenities and essential facilities, we will access data through FourSquare API interface and arrange them as a dataframe for visualization. By merging data on Barcelona properties and the relative price paid data from Ajuntament de Barcelona and data on amenities and essential facilities surrounding such properties from FourSquare API interface, we will be able to recommend profitable real estate investments.

3. Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

The objective is to build a table with the different neighborhoods, adding coordinates and the revenues. Statistical testing will share the city into 5 different clusters related with the price of renting.

a) Next table is merged neighborhoods and coordinates

In [14]:
```
df_merge = df.join(df_coordinates.set_index('Barri'), on='Barri')
df_merge
```

Out[14]:

	Barri	Year_2019	Latitude	Longitude
0	. el Raval	4.591	41.379518	2.168368
1	. el Barri Gòtic	3.811	41.383395	2.176912
2	. la Barceloneta	4.849	41.380653	2.189927
3	. Sant Pere, Santa Caterina i la Ribera	4.772	41.372251	2.177532
4	. el Fort Pienc	4.250	41.395925	2.182325
5	. la Sagrada Família	4.173	41.403479	2.174410
6	. la Dreta de l'Eixample	5.514	41.394124	2.166471
7	. l'Antiga Esquerra de l'Eixample	5.197	41.388765	2.156597
8	. la Nova Esquerra de l'Eixample	4.634	41.382816	2.149966
9	. Sant Antoni	4.412	41.378412	2.161768
10	. el Poble Sec-AEI Parc Montjuïc	3.911	41.368790	2.163184
11	. la Marina de Port	2.819	41.360296	2.137584

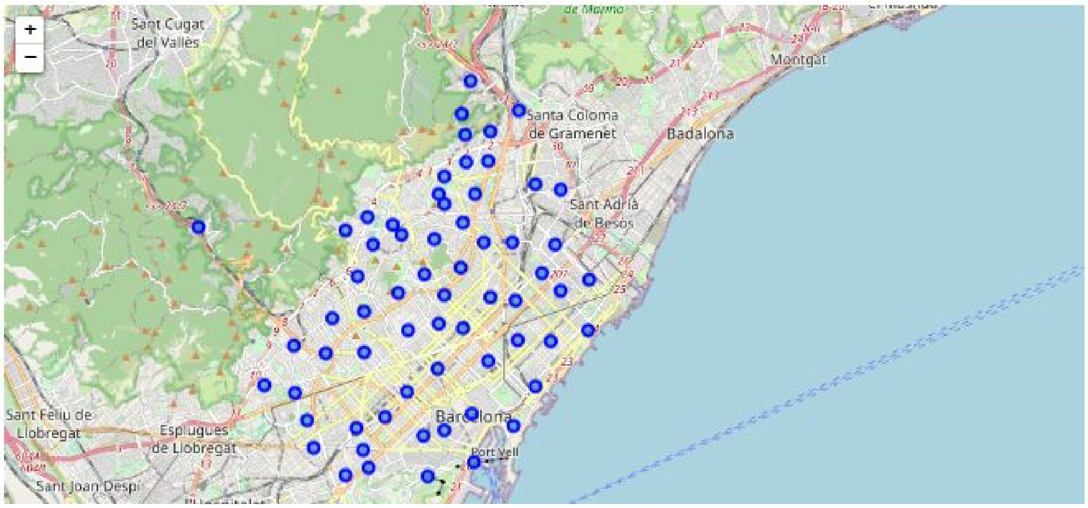
b) Using folium to create a map of Barcelona with the different neighborhoods.

```python
# create map of Toronto using latitude and longitude values
map_barcelona = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, Barri in zip(df_merge['Latitude'], df_merge['Longitude'], df_merge['Barri']):
    label = '{}'.format(Barri)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_barcelona)
map_barcelona
```

```
pip is /opt/conda/envs/Python36/bin/pip
pip is /opt/conda/bin/pip
```

Out[16]:



c) Using foursquare to find out the different venues.

In [17]:
```python
CLIENT_ID = 'NAYWBRRIY3P4BBK1CARNPQB2ERTWJACBIABJO1EH4E0UCSAI' # your Foursquare ID
CLIENT_SECRET = 'LYFXV1YWTUMG02US1MDQ0AJ2DVF4JPKV3UZTDUD1EMZZ2OQX' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: NAYWBRRIY3P4BBK1CARNPQB2ERTWJACBIABJO1EH4E0UCSAI
CLIENT_SECRET:LYFXV1YWTUMG02US1MDQ0AJ2DVF4JPKV3UZTDUD1EMZZ2OQX
```

In [18]:
```python
neighborhood_latitude = df_merge.loc[0, 'Latitude'] # neighborhood latitude value
neighborhood_longitude = df_merge.loc[0, 'Longitude'] # neighborhood longitude value

neighborhood_name = df_merge.loc[0, 'Barri'] # neighborhood name

print('Latitude and longitude values of {} are {}, {}.'.format(neighborhood_name,
                                                               neighborhood_latitude,
                                                               neighborhood_longitude))
```

```
Latitude and longitude values of  . el Raval are 41.3795176, 2.1683678.
```

c) The definitive table, neighborhood, coordinates, and venues with their coordinates

In [26]:
```
print(barcelona_venues.shape)
barcelona_venues.head()
```

```
(2877, 7)
```

Out[26]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	. el Raval	41.379518	2.168368	Chulapio	41.379264	2.165905	Cocktail Bar
1	. el Raval	41.379518	2.168368	La Robadora	41.379500	2.170463	Gastropub
2	. el Raval	41.379518	2.168368	Arume	41.378953	2.166008	Spanish Restaurant
3	. el Raval	41.379518	2.168368	A Tu Bola	41.380096	2.169054	Tapas Restaurant
4	. el Raval	41.379518	2.168368	La Monroe	41.378795	2.170692	Spanish Restaurant

In [27]:
```
barcelona_venues.groupby('Neighborhood').count()
```

Out[27]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	V Cat
Neighborhood						
. Can Baró	26	26	26	26	26	26
. Can Peguera ()	11	11	11	11	11	11

d) Using statistical tool to cluster the city of Barcelona in 5 differents clusters.

```
n [35]: # set number of clusters
kclusters = 5

barcelona_grouped_clustering = barcelona_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(barcelona_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
ut[35]: array([1, 3, 2, 3, 1, 1, 1, 1, 1, 3], dtype=int32)
```

```
n [36]: # add clustering labels

neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
barcelona_merged = df_merge

# merge barcelona_grouped with barcelona_data to add latitude/longitude for each neighborhood
barcelona_merged = df_merge.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Barri')
```

e) Using folium to create a map of Barcelona with the 5 different clusters

```
In [37]: !type -a pip
import folium
from geopy.geocoders import Nominatim
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(barcelona_merged['Latitude'], barcelona_merged['Longitude
ona_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color='red',
        fill=True,
        fill_color='blue',
        fill_opacity=0.7).add_to(map_clusters)
```

f) Next snipping is an example of two clusters.

Cluster 1

In [38]:
```
barcelona_merged.loc[barcelona_merged['Cluster Labels'] == 0, barcelona_merged.columns[[0] + list(range(5,barcelona_merged.shape
[1]))]]
```

Out[38]:

	Barri	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
20	. Vallvidrera, el Tibidabo i les Planes	Train Station	BBQ Joint	Restaurant	Women's Store	Falafel Restaurant	Electronics Store	Empanada Restaurant	Escape Room	Ethiopian Restaurant	Fabric Shop

Cluster 2

In [39]:
```
barcelona_merged.loc[barcelona_merged['Cluster Labels'] == 1, barcelona_merged.columns[[0] + list(range(5,barcelona_merged.shape
[1]))]]
```

Out[39]:

	Barri	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
			Mediterranean	Tapas		Spanish				

- Results section where you discuss the results.

The statistical results reflect the difference between expensive and cheap neighborhoods. There is a relationship between the quality of venues and the price of renting.

The results split Barcelona mainly into two areas, the area near Tibidabo, Bonanova and the new neighborhoods near the sea.

L'eixample , both dreta and esquerra are in the same cluster, keeping the same offer in venues and similar renting prices.

Some areas like Parallel reflect a weak offer of venues and low prices.

- **Conclusion section where you conclude the report.**

The algorithm is telling me:

1) There is exclusive and expensive renting with two options, mountain or see neighborhoods.
2) There is affordable renting mainly in dreta and esquerra eixample areas.
3) There is some cheaper renting with poor offer of venues in some areas like parallel.