Capstone Project - Housing Sales Prices & Venues Data Analysis of Casablanca-Morocco

Table of contents

Introduction: Business Problem

Casablanca is a port city on the Atlantic coast of western Morocco. It is Morocco's biggest city, principal port, and economic capital. The town of Casablanca was founded in 1515. It was an important strategic port during World War II and hosted the Anglo-American Summit in 1943. The city is served by Mohammed V International Airport. Casablanca covers the richest and most sophisticated aspects of Morocco life, style, and architecture. Casablanca boasts one of the most extensive and diverse displays of art deco architecture in the world.
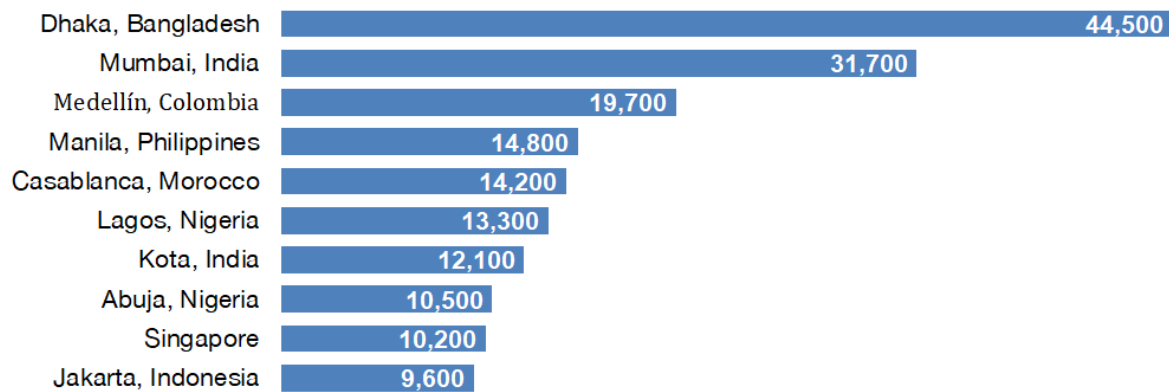
At the center of the former 'Great Casablanca' region (currently part of the Casablanca-Settat administrative region), this metropole is considered the locomotive of the development of the Moroccan economy. It attracts 32% of the country's production units and 56% of industrial labor. The region uses 30% of the national electricity production. The region contributes to 44% of the industrial production of the kingdom of Morocco. About 33% of national industrial exports, MAD 27 billion, comes from the Grand Casablanca; 30% of the Moroccan banking network is concentrated in Casablanca. [1]
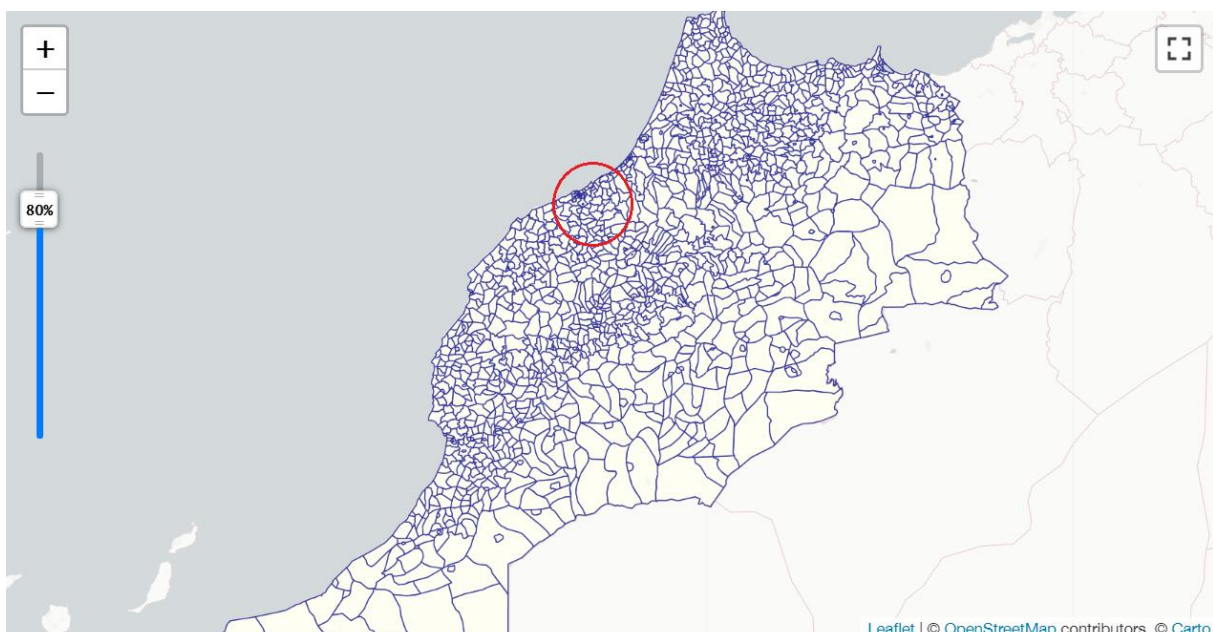
Based on a recent report released by the World Economic Forum (WEF), Casablanca is amongst the world's most crowded cities. With a population density of 14,200 inhabitants per square kilometer, Casablanca came fifth in the world and first in Africa [2].

The world's most crowded cities
Number of people per square kilometre, 2013

Dhaka, Bangladesh	44,500
Mumbai, India	31,700
Medellín, Colombia	19,700
Manila, Philippines	14,800
Casablanca, Morocco	14,200
Lagos, Nigeria	13,300
Kota, India	12,100
Abuja, Nigeria	10,500
Singapore	10,200
Jakarta, Indonesia	9,600

Source: UN UrbanData

The idea behind this project is to combine venue data with house sales pricing (apartment per square meter price) and try to analyze, index and cluster the results for better understanding of what could motivate a future home owner in Casablanca in his or her choice of district/neighborhood.

Data

In order to answer the main business problem, various data sources were used to collect relevant data, namely:

- Fourth level administrative divisions of Morocco from the spatial data repository of NYU [3]. The GeoJson file wasn't available for download, instead we used the shapefile that was converted later to a csv file and cleaned to reflect Boroughs of the region of 'Great Casablanca'.

- Foursquare API was used to get the most common venues of boroughs of 'Great Casablanca' [4].

- Per square meter housing average prices of boroughs of Casablanca as published by the Directorate-General for Taxation in Morocco (raw data was cleaned and processed to be integrated in our dataset) [5]. Data was wrangled to reflect our need for average per square meter prices of apartments in Casablanca.

Methodology

The master dataset used for this project is an aggregation of different boroughs/neighborhoods in the city of Casablanca with their longitude/latitude information as well as the average per square meter apartment prices (this data is in Moroccan dirhams, 1 MAD ≈ 0.10 USD).

	Country	Province	City	Borough	Neighborhood	Type	longitude	latitude	m_sq_price
0	Morocco	Grand Casablanca	Casablanca	Bouskoura	Bouskoura	Rural Commune	-7.621141	33.468276	10966
1	Morocco	Grand Casablanca	Casablanca	Bouskoura	Dar Bouazza	Rural Commune	-7.777721	33.512917	10966
2	Morocco	Grand Casablanca	Casablanca	Mediouna	El Majjatia Oulad Taleb	Rural Commune	-7.527317	33.461789	6416
3	Morocco	Grand Casablanca	Casablanca	Mediouna	Lahraouyine	Rural Commune	-7.523790	33.532434	6416
4	Morocco	Grand Casablanca	Casablanca	Ahl Laghlam	Ahl Laghlam	Urban Commune	-7.487686	33.574038	8216

I used python folium library to visualize geographic details of Casablanca and its boroughs and superimposed on top. I used latitude and longitude values to get the visual as shown in the map below:
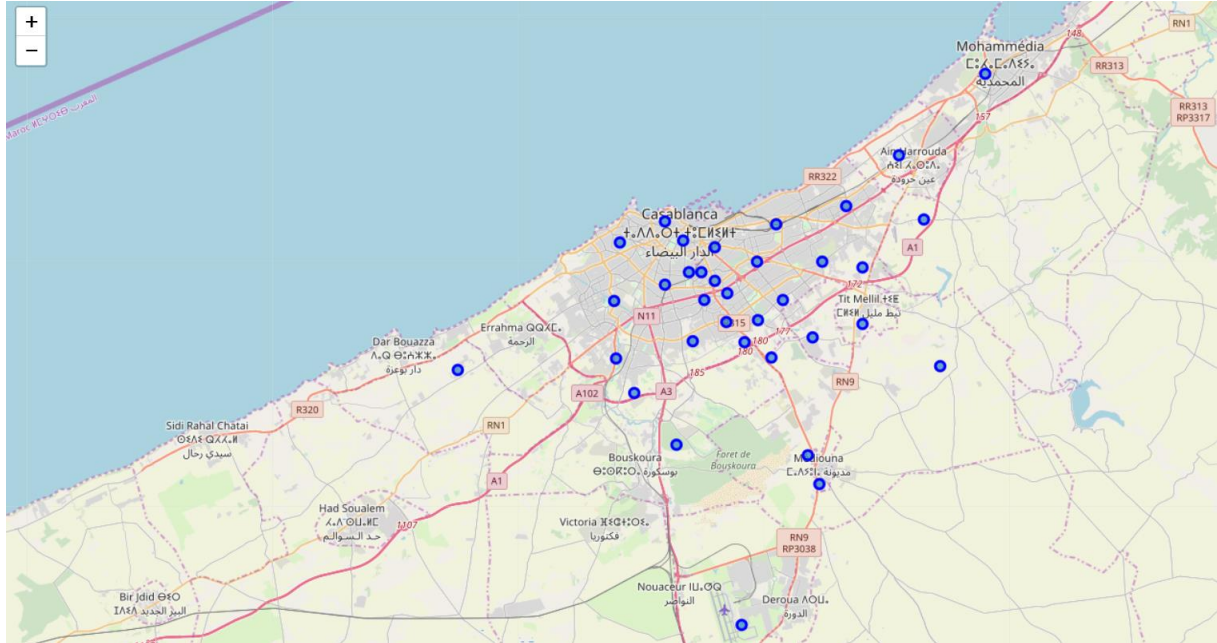
Analysis
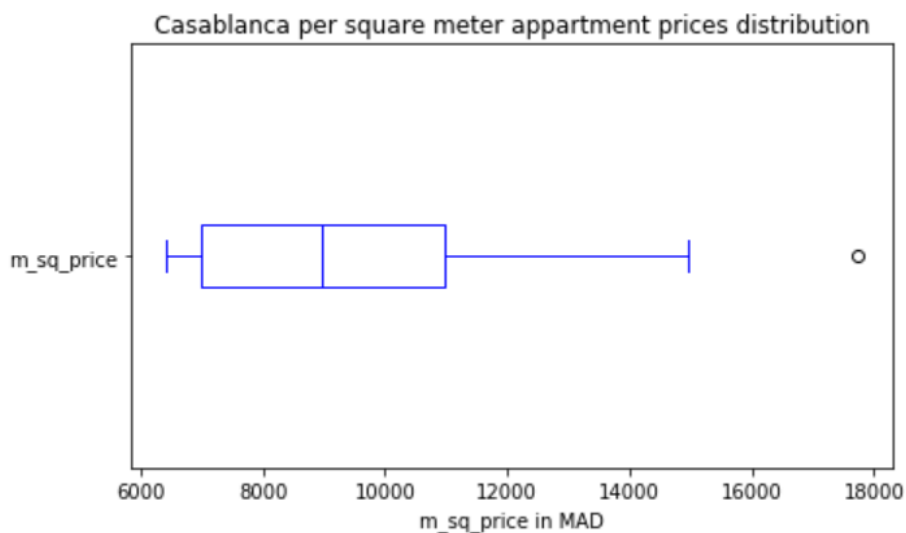
We can immediately make a few key observations from the plot above:

- The minimum per square meter price of an apartment in Casablanca is around 6416 MAD, the maximum price is 17738 MAD, and the median price of our sample is around 8954 MAD.
- 25% of Casablanca's neighborhoods have a per-square-meter of less 6979 (First quartile).

- 75 % of Casablanca's neighborhoods have a per-square-meter of less 10966 (Third quartile).
- The maximum price per square meter in Casablanca is in the neighborhood of Anfa (which represents a discrepancy compared to the rest of the sample).

The information describing our Dataset is as follows:

count	35
mean	9358
std	2793
min	6416
25%	6979
50%	8954
75%	10966
max	17738

I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venue and the radius 1500 meter for each borough from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude for the Anfa neighborhood:

	name	categories	lat	lng
0	Royal Golf D'Anfa	Golf Course	33.588384	-7.658432
1	Starbucks	Coffee Shop	33.589764	-7.649899
2	Om Yoga	Yoga Studio	33.593753	-7.668892
3	Roosevelt Café	Café	33.594718	-7.667885
4	Frédéric Cassel	Café	33.589944	-7.650890

In total, Foursquare returned 109 venues in Casablanca, and compiled them by venue category:

```
[26]: print(casablanca_venues.shape)
      casablanca_venues.head()
      (109, 7)
```

[26]:		Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
	0	Ain Chock	33.529698	-7.609077	Horizon California	33.528067	-7.612827	Café
	1	Ain Chock	33.529698	-7.609077	Pharmacie Les Anges	33.533525	-7.608327	Pharmacy
	2	Ain Chock	33.529698	-7.609077	Cafe 7even	33.533360	-7.606924	Beer Garden
	3	Ain Chock	33.529698	-7.609077	Salle Imane	33.533775	-7.609749	Gym
	4	Ain Chock	33.529698	-7.609077	Segafredo Bv al Qods	33.533761	-7.606956	Café

The number of unique venue categories returned by Foursquare is 56 (categories range from Airport Service to Coffee shops, Diners, Fast foods, Hotels, etc.)

```
[81]: print('There are {} uniques categories.'.format(len(casablanca_venues['Venue Category'].unique())))
      There are 56 uniques categories.
```

Then, I analyzed all neighborhoods and grouped them with the top 5 most common venues:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Ain Chock	Beer Garden	Pharmacy	Café	Tram Station	Dessert Shop
1	Ain Sebaa	Theater	Burrito Place	Fast Food Restaurant	Tram Station	Department Store
2	Al Fida	Department Store	Supermarket	Soccer Field	Fast Food Restaurant	Tram Station
3	Al Idrissia	Café	Astrologer	Tram Station	Dessert Shop	Hotel
4	Anfa	Racetrack	Art Gallery	Golf Course	Concert Hall	Tram Station

After that, I run k-means to cluster the neighborhood into 3 clusters, the resulting map is shown below:
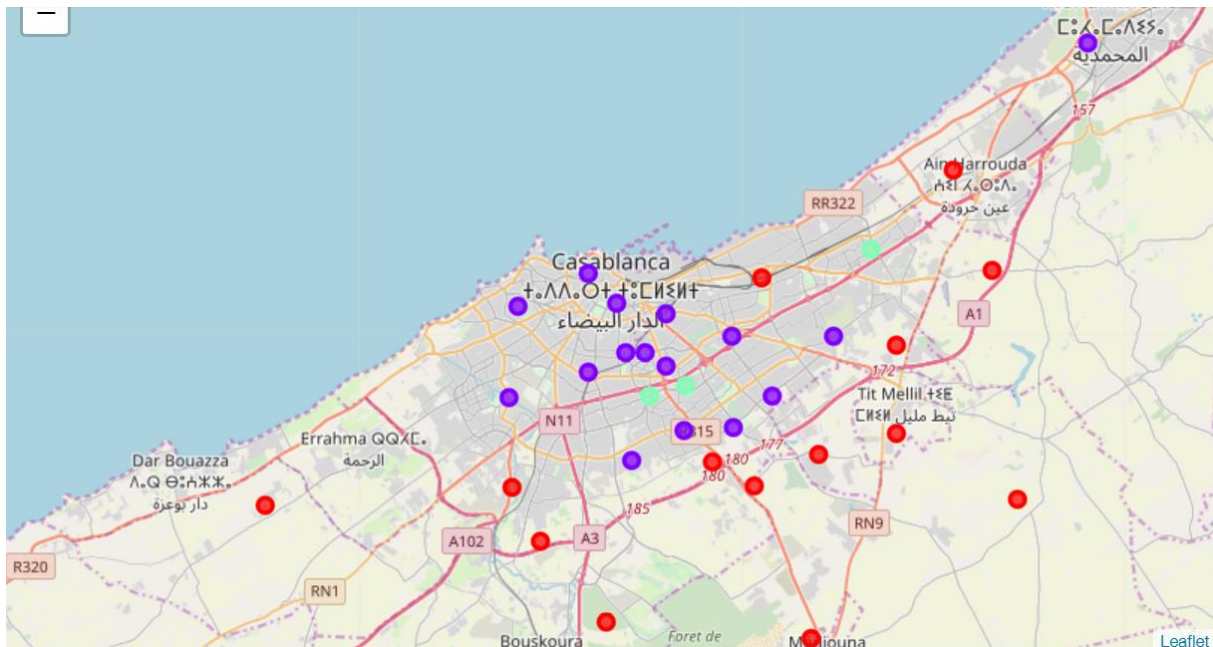
Cluster in green dots:

		Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
[53]:	9	Al Idrissia	2	Café	Tram Station	Doner Restaurant	Hotel	Gym
	13	Bou Chentouf	2	Café	Fast Food Restaurant	Tram Station	Doner Restaurant	Hotel
	27	Sidi Bernoussi	2	Café	Sports Club	Fish & Chips Shop	Tram Station	Doner Restaurant

Cluster in red dots:

[51]:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bouskoura	0	NaN	NaN	NaN	NaN	NaN
1	Dar Bouazza	0	NaN	NaN	NaN	NaN	NaN
2	El Majjatia Oulad Taleb	0	NaN	NaN	NaN	NaN	NaN
3	Lahraouyine	0	NaN	NaN	NaN	NaN	NaN
4	Ahl Laghlam	0	NaN	NaN	NaN	NaN	NaN
6	Ain Harrouda	0	NaN	NaN	NaN	NaN	NaN
7	Ain Sebaa	0	Burrito Place	Fast Food Restaurant	Tram Station	Doner Restaurant	Hotel
17	Lissasfa	0	Fast Food Restaurant	Tram Station	Doner Restaurant	Hotel	Gym
19	Mediouna	0	NaN	NaN	NaN	NaN	NaN
24	Salmia	0	NaN	NaN	NaN	NaN	NaN
25	Sbata	0	NaN	NaN	NaN	NaN	NaN
28	Sidi Maarouf	0	NaN	NaN	NaN	NaN	NaN
31	Tit Mellil	0	NaN	NaN	NaN	NaN	NaN
32	Echchallalate	0	NaN	NaN	NaN	NaN	NaN
33	Sidi Hajjaj Ouad Hassar	0	NaN	NaN	NaN	NaN	NaN

Cluster in blue dots:

[51]:

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bouskoura	0	NaN	NaN	NaN	NaN	NaN
1	Dar Bouazza	0	NaN	NaN	NaN	NaN	NaN
2	El Majjatia Oulad Taleb	0	NaN	NaN	NaN	NaN	NaN
3	Lahraouyine	0	NaN	NaN	NaN	NaN	NaN
4	Ahl Laghlam	0	NaN	NaN	NaN	NaN	NaN
6	Ain Harrouda	0	NaN	NaN	NaN	NaN	NaN
7	Ain Sebaa	0	Burrito Place	Fast Food Restaurant	Tram Station	Doner Restaurant	Hotel
17	Lissasfa	0	Fast Food Restaurant	Tram Station	Doner Restaurant	Hotel	Gym
19	Mediouna	0	NaN	NaN	NaN	NaN	NaN
24	Salmia	0	NaN	NaN	NaN	NaN	NaN
25	Sbata	0	NaN	NaN	NaN	NaN	NaN
28	Sidi Maarouf	0	NaN	NaN	NaN	NaN	NaN
31	Tit Mellil	0	NaN	NaN	NaN	NaN	NaN
32	Echchallalate	0	NaN	NaN	NaN	NaN	NaN
33	Sidi Hajjaj Ouad Hassar	0	NaN	NaN	NaN	NaN	NaN

Discussion and Conclusion:

Casablanca is a big city with a high population density in a narrow area. The total number of measurements and population densities of the 35 neighborhoods is disproportionate. Various approaches could be considered in clustering and classifying these neighborhoods.

Moreover, it is obvious that not every classification method can yield the same high quality results. The idea behind this little project is to show that a household choice of a neighborhood to live in, though correlated to the price per-square-meter (which reflects to some degree the quality of the

neighborhood), could be influenced by different factors related to venues proximity, public services availability, etc.

References

[1] https://en.wikipedia.org/wiki/Economy_of_Casablanca

[2] https://www.weforum.org/agenda/2017/05/these-are-the-world-s-most-crowded-cities/

[3] https://geo.nyu.edu/catalog/stanford-fd597jf1799

[4] https://developer.foursquare.com/

[5] https://portail.tax.gov.ma/wps/portal/DGI/Referentiels-des-prix-de-l_immobilier?classic=1