Casablanca

Casablanca is the largest city of Morocco. Located in the central-western part of Morocco bordering the Atlantic Ocean, it is the largest city in the Maghreb region and the eighth-largest in the Arab world. Casablanca is Morocco's chief port and one of the largest financial centers in Africa. According to the 2019 population estimate, the city has a population of about 3.71 million in the urban area and over 4.27 million in the Greater Casablanca. Casablanca is considered the economic and business center of Morocco, although the national political capital is Rabat.

The leading Moroccan companies and many international corporations doing business in the country have their headquarters and main industrial facilities in Casablanca. Recent industrial statistics show Casablanca holds its recorded position as the primary industrial zone of the nation. The Port of Casablanca is one of the largest artificial ports in the world, and the second largest port of North Africa, after Tanger-Med 40 km (25 mi) east of Tangier.

In this exercise I will focus on clustering all the restaurants in this big economic city, because it's the most populated city in Morocco.

Introduction/Business problem

Casablanca, the city the author lives in, attracts a large number of investors, and less number of tourists, but it remains the biggest economic City where people can find a lot of job opportunities, that's why the people living in Casablanca are various. For foreign people especially, finding the right place to eat can be a challenge because the city is too big and Moroccan dishes may not be convenient for them.

Thus, the goal I want to reach with this exercise is to give a simple recommendation to people in Casablanca: in which district of the city will you find a large number or even concentration of which types of restaurants? Where to eat Mediterranean food, where to find German food, French food, Italian food, where to get fast food? The target audience are foreign people and also investors who want to open a new restaurant, so it's crucial for them to know how the existing restaurants are distributed in Casablanca's districts.

Description of the data

I will, as requested by the assignment task, use foursquare data about restaurants in Casablanca. Foursquare is a US tech company from New York focusing on location data. Their technology and data powers apps such as Apple's Maps, Uber, Twitter and many other household names. Here is an example of restaurant in Casablanca on foursquare: https://fr.foursquare.com/explore?mode=url&near=Al-Markaz%2C%20Casablanca&nearGeoId=10011708&q=Nourriture . I will use foursquare data such as the restaurant name, ID, location and category of food (vegetarian, Italian etc.).

Also, I will use the overview of districts/city parts of Casablanca from Wikipedia: https://en.wikipedia.org/wiki/Casablanca

https://sites.google.com/site/collectivitesaumaroc/regions/grand-casablanca/provinces-et-communes-du-grand-casablance

Here, you will find a table "Districts" which shows the nine city districts and its neighborhoods/city parts. I will use these districts and the data about restaurants in these districts from foursquare to show the density of restaurants in them.

Administrative divisions [edit]

Casablanca is a commune, part of the region of Casablanca-Settat. The commune is divided into eight districts or prefectures, which are themselves divided into 16 subdivisions or arrondissements and one municipality. The districts and their subdivisions are:[62]

1. **Aïn Chock** (عين الشق) – Aïn Chock (عين الشق)
2. **Aïn Sebaâ - Hay Mohammadi** (عين السبع الحي المحمدي) – Aïn Sebaâ (عين السبع), Hay Mohammadi (الحي المحمدي), Roches Noires (روش نوار).
3. **Anfa** (أنفا) – Anfa (أنفا), Maârif (المعاريف), Sidi Belyout (سيدي بليوط).
4. **Ben M'Sick** (بن مسيك) – Ben M'Sick (بن مسيك), Sbata (سباتة).
5. **Sidi Bernoussi** (سيدي برنوصي) – Sidi Bernoussi (سيدي برنوصي), Sidi Moumen (سيدي مومن).
6. **Al Fida - Mers Sultan** (الفداء – مرس السلطان) – Al Fida (الفداء), Mechouar (المشور) (municipality), Mers Sultan (مرس السلطان).
7. **Hay Hassani** (الحي الحسني) – Hay Hassani (الحي الحسني).
8. **Moulay Rachid** (مولاي رشيد) – Moulay Rachid (مولاي رشيد), Sidi Othmane (سيدي عثمان).

Port of Casablanca

Methodology

In this section, I will describe the data analysis and how I used the data to yield the results.

Starting out, I scraped data from Wikipedia to create a dataframe with the city Boroughs of Casablanca: https://sites.google.com/site/collectivitesaumaroc/regions/grand-casablanca/provinces-et-communes-du-grand-casablance. For this, I used the pandas read function. I had to clean the resulting data frame in terms of unnecessary information or data that could not be handled in a data frame, such as picture data of the coat of arms of each district. The result is a nice data frame:

	Borough
2	Casablanca
3	Aîn-chock
4	Aîn-sebaâ
5	Anfa
6	El-fida casablanca
7	Assoukhour assawda
8	Ben m'sick
9	Maarif casablanca
10	Hay-hassani
11	Hay mohammadi
12	Mers-sultan

Then, I enabled geopy functions by installing the conda-forge geopy package. I used the nominatim function to add geospatial data to the data frame that is the latitude and the longitude seen on the right side of the following table.

	Borough	Latitude	Longitude
2	Casablanca	33.595063	-7.618777
3	Aîn-chock	33.536215	-7.615698
4	Aîn-sebaâ	33.608591	-7.524144
5	Anfa	33.592632	-7.672177
6	El-fida casablanca	33.566522	-7.607288
7	Assoukhour assawda	33.586519	-7.588056
8	Ben m'sick	33.554261	-7.581366
9	Maarif casablanca	33.576130	-7.636660
10	Hay-hassani	33.567046	-7.678489
11	Hay mohammadi	33.584144	-7.556956
12	Mers-sultan	33.575768	-7.599929
13	Moulay rachid	33.567545	-7.545244
14	Sbata	33.880766	-5.566873

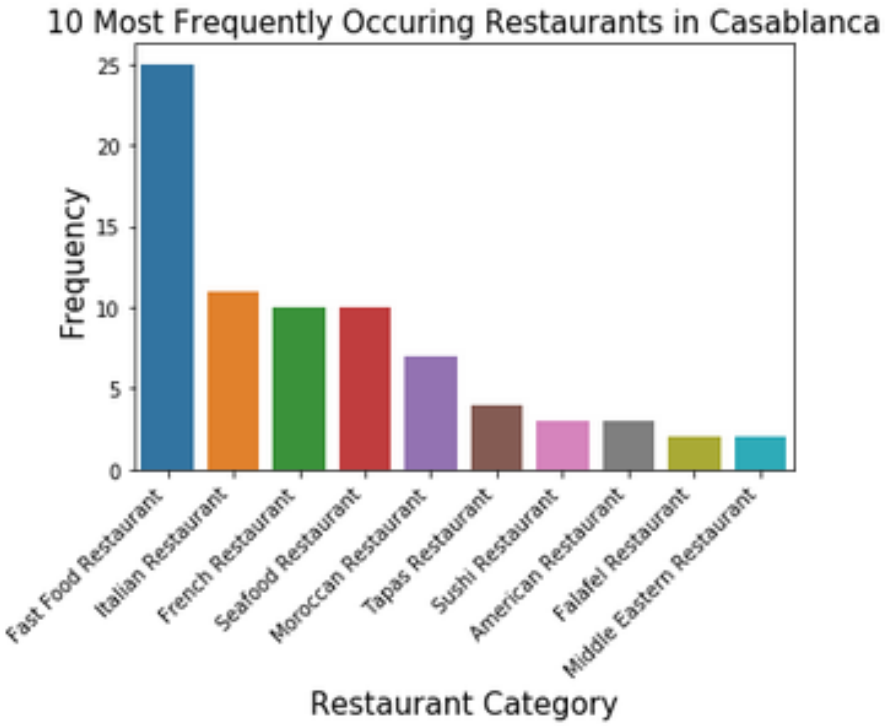
Using the folium package and my data frame, I then created a map with on it the Casablanca's Boroughs.

Then, retrieved the foursquare data for all venues on foursquare with a distance of less than 900 meters from each center of borough, as indicated as blue dots in the map above. The result was a list of 757 venues all over Cologne city. Out of these 414 venues, 89 where restaurants. These 89 restaurants come from 19 unique restaurant categories, such as Italian, Moroccan or French.

```
(414, 7)
```

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Casablanca	33.595063	-7.618777	Six PM	33.595940	-7.618684	Hotel Bar
1	Casablanca	33.595063	-7.618777	Le Riad Restaurant	33.593936	-7.614676	Moroccan Restaurant
2	Casablanca	33.595063	-7.618777	Hyatt Regency Casablanca	33.596195	-7.618708	Hotel
3	Casablanca	33.595063	-7.618777	Casa Jose	33.597823	-7.615341	Tapas Restaurant
4	Casablanca	33.595063	-7.618777	Le Rouget de l'Isle	33.592591	-7.622857	French Restaurant

I plotted a bar chart with the frequency of the 10 most frequently occuring restaurants in the whole city, using seaborn/matplotlib packages. We can see that Fast Food, Italien, French, Seafood and Moroccan restaurants are the most frequently occuring restaurants in Casablanca, which seems pretty reasonable.



To find clusters of restaurant types in the different Borough districts, I first transformed the data frame with the restaurant, associated to Boroughs, by one-hot encoding (0/1), as seen in the picture below.

Borough	American Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Italian Restaurant	Japanese Restaurant	Kebab Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Moroccan Restaurant	R
Casablanca	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
Casablanca	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Casablanca	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
Casablanca	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
Casablanca	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Next, I used grouping to show the frequency of each category of restaurants in each borough.

	Borough	American Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Italian Restaurant	Japanese Restaurant	Kebab Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Moroccan Restaurant	Seafood Restaurant
0	Ain harrouda	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
1	Anfa	0.000000	0.000000	0.000000	0.000000	0.090909	0.000000	0.272727	0.000000	0.272727	0.000000	0.000000	0.090909	0.090909	0.090909	0.000000
2	Assoukhour assawda	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Aîn-chock	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.750000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Aîn-sebaâ	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	Casablanca	0.037037	0.000000	0.000000	0.037037	0.037037	0.000000	0.222222	0.185185	0.037037	0.000000	0.037037	0.037037	0.000000	0.111111	0.185185
6	Dar bouazza	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.666667
7	El-fida casablanca	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.600000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	Hay mohammadi	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

I used this information to create a data frame in which you can see the most common restaurant venue types for each borough.

	Borough	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
0	Ain harrouda	Moroccan Restaurant	Tapas Restaurant	French Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Restaurant	Fast Food Restaurant	Italian Restaurant
1	Anfa	Italian Restaurant	Fast Food Restaurant	Middle Eastern Restaurant	Doner Restaurant	Mediterranean Restaurant	Tapas Restaurant	Moroccan Restaurant	Kebab Restaurant	Japanese Restaurant	Sushi Restaurant
2	Assoukhour assawda	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Restaurant	French Restaurant	Italian Restaurant
3	Aîn-chock	Fast Food Restaurant	Sushi Restaurant	Tapas Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Restaurant	French Restaurant	Italian Restaurant
4	Aîn-sebaâ	Italian Restaurant	Sushi Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Tapas Restaurant
5	Casablanca	Fast Food Restaurant	French Restaurant	Seafood Restaurant	Moroccan Restaurant	Tapas Restaurant	Chinese Restaurant	Doner Restaurant	Italian Restaurant	Kebab Restaurant	Mediterranean Restaurant
6	Dar bouazza	Seafood Restaurant	Japanese Restaurant	Tapas Restaurant	French Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Restaurant	Fast Food Restaurant
7	El-fida casablanca	Fast Food Restaurant	Italian Restaurant	Falafel Restaurant	Sushi Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	French Restaurant	Tapas Restaurant
8	Hay mohammadi	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Restaurant	French Restaurant	Italian Restaurant
9	Maarif casablanca	Italian Restaurant	Seafood Restaurant	Fast Food Restaurant	American Restaurant	Moroccan Restaurant	Middle Eastern Restaurant	Mediterranean Restaurant	Kebab Restaurant	Japanese Restaurant	Sushi Restaurant
10	Mechouar de casablanca	Moroccan Restaurant	Tapas Restaurant	French Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Restaurant	Fast Food Restaurant	Italian Restaurant

Now, with all this data, I could finally run an unsupervised machine learning algorithm, more specifically, a k-means clustering algorithm from the scikit-learn package. One could use the ellbow method to systematically define the k value, but I simply chose k to be 4, having been inspired by one of the coursera courses to do so.

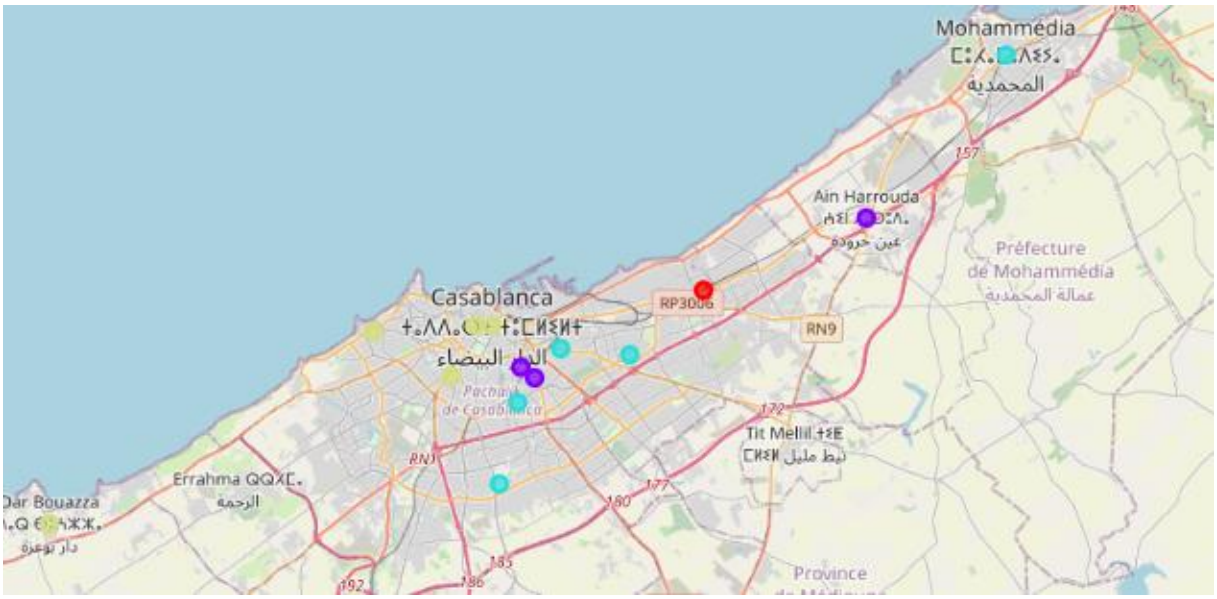
Results

And here already comes the result:

	Borough	Latitude	Longitude	Cluster_labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Cor Resta
2	Casablanca	33.595063	-7.618777	3	Fast Food Restaurant	French Restaurant	Seafood Restaurant	Moroccan Restaurant	Tapas Restaurant	Chinese Restaurant	Doner Restaurant	Italian Resta
5	Anfa	33.592632	-7.672177	3	Italian Restaurant	Fast Food Restaurant	Middle Eastern Restaurant	Doner Restaurant	Mediterranean Restaurant	Tapas Restaurant	Moroccan Restaurant	Kebab Resta
9	Maarif casablanca	33.576130	-7.636660	3	Italian Restaurant	Seafood Restaurant	Fast Food Restaurant	American Restaurant	Moroccan Restaurant	Middle Eastern Restaurant	Mediterranean Restaurant	Kebab Resta
15	Sidi belyout	33.594943	-7.626314	3	French Restaurant	Italian Restaurant	Fast Food Restaurant	Sushi Restaurant	Tapas Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Seafood Resta
36	Dar bouazza	33.521583	-7.816437	3	Seafood Restaurant	Japanese Restaurant	Tapas Restaurant	French Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Resta
3	Aîn-chock	33.536215	-7.615698	2	Fast Food Restaurant	Sushi Restaurant	Tapas Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Resta
6	El-fida casablanca	33.566522	-7.607288	2	Fast Food Restaurant	Italian Restaurant	Falafel Restaurant	Sushi Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Resta
7	Assoukhour assawda	33.586519	-7.588056	2	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Resta
11	Hay mohammadi	33.584144	-7.556956	2	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Resta
14	Sbata	33.880766	-5.566873	2	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Resta
26	Mohammedia	33.695838	-7.389329	2	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Resta
12	Mers-sultan	33.575768	-7.599929	1	Moroccan Restaurant	Fast Food Restaurant	Tapas Restaurant	French Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Resta
19	Mechouar de casablanca	33.579529	-7.605685	1	Moroccan Restaurant	Tapas Restaurant	French Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Doner Restaurant	Falafel Resta

What we see in the table are the city districts and their most common venues, and they now have been assigned five different cluster labels from 0 to 3.

We can now use the cluster labels to show the city districts marked with a cluster-specific color on a map, again using folium:



You will see the 15 bubbles for the boroughs after removine the none clustred ones, with four different colors for the four different clusters.

Cluster 1 – the French, Italian & Seafood Cluster

	Borough	Latitude	Longitude	Cluster_labels	1st Most Common Restaurant	2nd Most Common Restaurant
2	Casablanca	33.595063	-7.618777	3	Fast Food Restaurant	French Restaurant
5	Anfa	33.592632	-7.672177	3	Italian Restaurant	Fast Food Restaurant
9	Maarif casablanca	33.576130	-7.636660	3	Italian Restaurant	Seafood Restaurant
15	Sidi belyout	33.594943	-7.626314	3	French Restaurant	Italian Restaurant
36	Dar bouazza	33.521583	-7.816437	3	Seafood Restaurant	Japanese Restaurant

Cluster 2 - the Fast Food Cluster

3	Aîn-chock	33.536215	-7.615698	2	Fast Food Restaurant	Sushi Restaurant	Tapas Restaurant
6	El-fida casablanca	33.566522	-7.607288	2	Fast Food Restaurant	Italian Restaurant	Falafel Restaurant
7	Assoukhour assawda	33.586519	-7.588056	2	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant
11	Hay mohammadi	33.584144	-7.556956	2	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant
14	Sbata	33.880766	-5.566873	2	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant
26	Mohammedia	33.695838	-7.389329	2	Fast Food Restaurant	Tapas Restaurant	Sushi Restaurant

Cluster 3 - the Moroccan Food Cluster

12	Mers-sultan	33.575768	-7.599929	1	Moroccan Restaurant	Fast Food Restaurant	Tapas Restaurant
19	Mechouar de casablanca	33.579529	-7.605685	1	Moroccan Restaurant	Tapas Restaurant	French Restaurant
27	Ain harrouda	33.635372	-7.451398	1	Moroccan Restaurant	Tapas Restaurant	French Restaurant

Cluster 4 - the sushi and Italian Food

4	Aîn-sebaâ	33.608591	-7.524144	0	Italian Restaurant	Sushi Restaurant

Discussion

If I reflect the work necessary to create these results, what comes to my mind is that for typical ways of scraping, cleaning, handling, transforming and visualizing data, all the tools are simply there. We just have to get to know the available open source packages and learn how to use them. What I find fantastic is that nearly all of them are free of charge. Also, a simple notebook computer is enough: in my case, I used a ThinkPad T430, more than three years old. All the rest is concentrated, creative, interesting, sometimes hard work and searching for hints, tips, examples, explanations etc. in the web. With these tools, many exciting data science use cases can be created, for all kinds of useful purposes.

Conclusion

The biggest constraint I encored in this projects is the lack of restaurant venues data this can be visible in the classified clusters, the same to be logic, except the fourth one.

We achieved the goal presented at the outset of this report: tourists can see in the results which city districts best match their food desires. This is just one example of fantastic data science uses cases one can realize applying technology which is available for free today! What a time to be alive.