Comparison between Neighborhoods of Cairo, Istanbul and Marrakesh

Contents

• <u>Introduction:</u>

Many Business owners are targeting to increase their own business , specially if they have a succeeded opened business in some cities , and they would like to gain similar or better success in other suggested cities .

Our Study is about 3 Capitals for 3 different big countries which are:

1- Cairo (Egypt)
2- Istanbul (Turkey)
3- Marrakesh (Morocco)

We are targeting in this study to compare Neighborhoods for these 3 Capitals by doing some process (Fetching Data from Internet, getting their relevant neighborhoods for Foursquare, plotting the result to clear neighborhoods image and idea, working the created data sets and make some clustering/sorting, to highlight the shared clusters between all cities which we are going to study.

• <u>Data Sources:</u>

During this project, I used a lot of useful Python libraries to import/process/plot .

 These libraries were used in (web scraping the Wikipedia imported data (Beautifulsoup) , Work with Dataframes (Pandas) ,Geocode and getting neighborhoods data like coordinates (geopy) , plotting markers on the Maps(Folium) , Clustering (Sklearn) .

Below links were used to fetch the 3 Capitals data which later sorted in datasets):

https://en.wikipedia.org/wiki/Category:Districts_of_Cairo

https://en.wikipedia.org/wiki/List_of_districts_of_Istanbul

https://en.wikipedia.org/wiki/Subdivisions_of_Marrakesh

• <u>Importing Data:</u>

Importing data process passed through 3 Phases:

1- Importing list of Neighborhoods which related to each city from the 3 Capitals (from Wikipedia pages using Beautifulsoup Library, a dataframe was created (Per each city) before merging them in one dataframe which contains (Neighborhood, City name and Country name.

2- **Geolocation of the Neighborhood:** Getting location data of the neighborhoods and mapping them together (Geopy).

> Now in geopy library, Nominatim service is used. For using the free service of Nominatim, there is a restriction of 1call per sec to the service. To avoid 'timeout' error, there needs to be at least 1 sec gap in each call even if a for loop is used. To provide a sufficient gap to accommodate network delay, a gap of 2 sec is provided. The gap is provided by calling the sleep function.
>
> Timeout Errors
> Even after providing a 2 sec gap in calls, there are timeout errors. So, to handle these errors is simple. Simply call the Nominatim service again for these locations after checking for network connectivity.
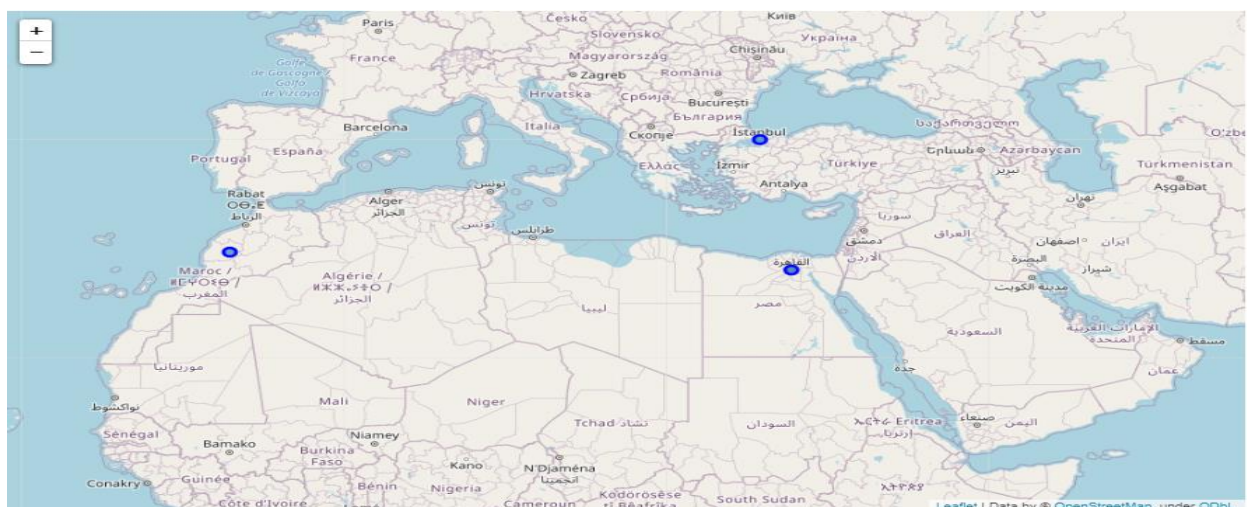>
> Missing/No Coordinates
> Some locations will not resolve into coordinates. This can happen because some locations may have different spellings. These can be rectified by using different spellings. Some locations will not resolve despite that. Then that data is procured manually searching on Google Maps.

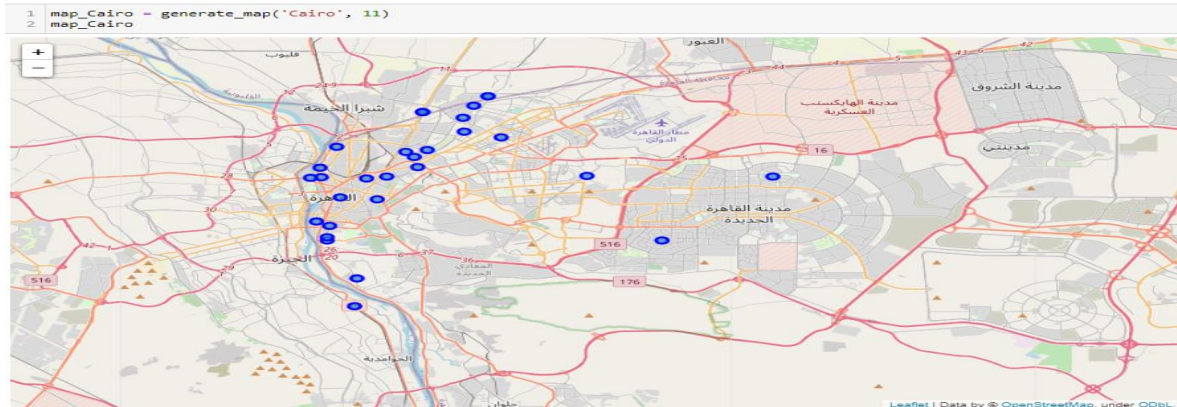## • <u>Drawing Neighborhoods/Cities Maps</u>

- One Global map was created including main 3 capitals to give an overview to distance which may be considered between 3 cities.
- A detailed map for every capital also were created which contains all their related neighborhoods.

<u>Global Map:</u>

Neighbourhoods Maps

Cairo Map:

```
1  map_Cairo = generate_map('Cairo', 11)
2  map_Cairo
```
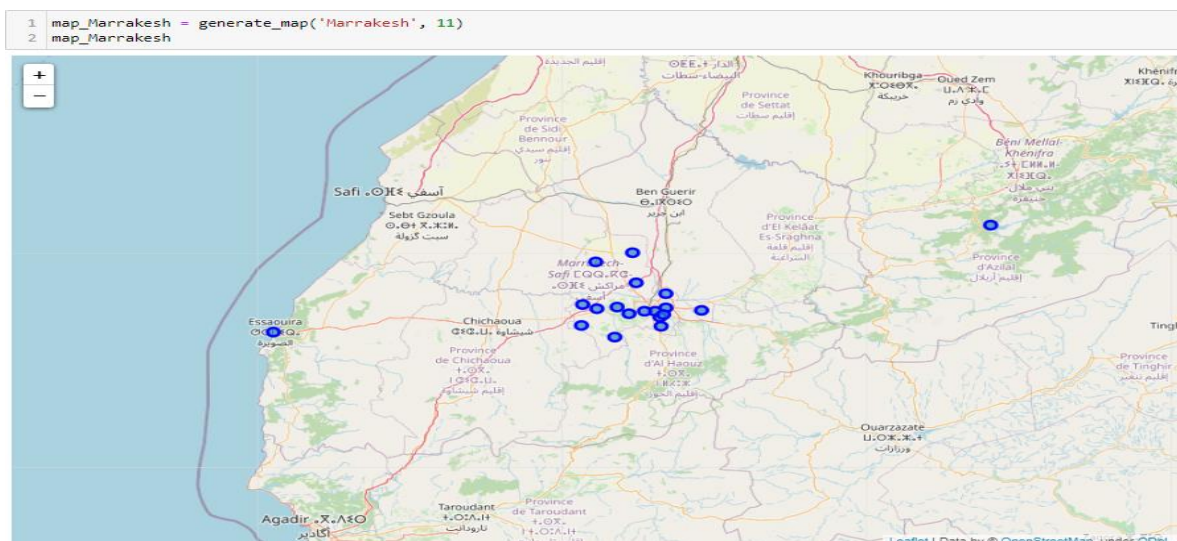

Istanbul Map:

```
1  map_Istanbul = generate_map('Istanbul', 11)
2  map_Istanbul
```


Marakesh Map:

```
1  map_Marrakesh = generate_map('Marrakesh', 11)
2  map_Marrakesh
```
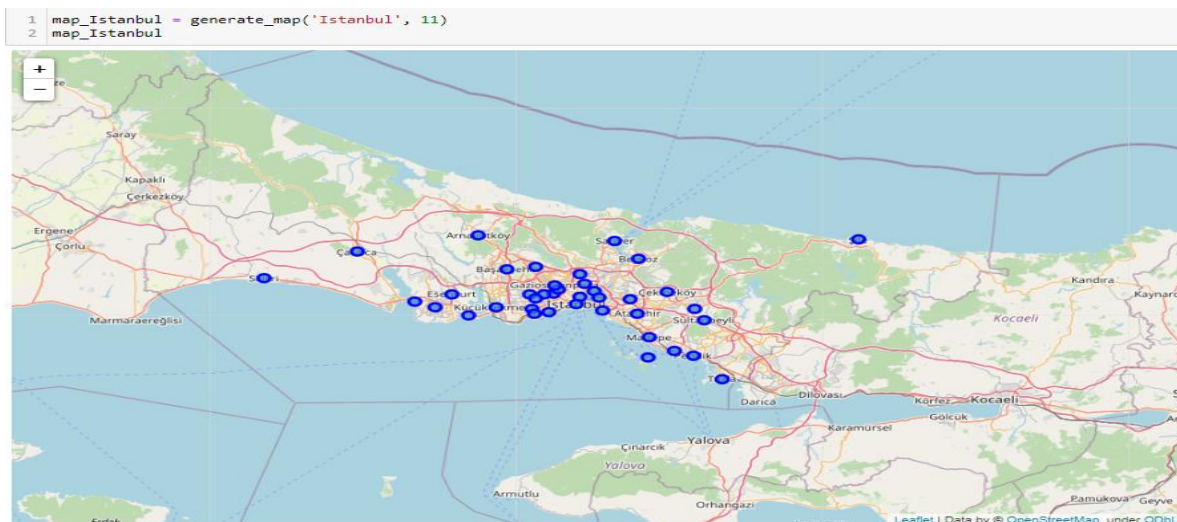
• <u>Venues:</u>

- Venues are places located in the neighborhoods like restaurants, hotels, cafes, parks etc. Foursquare API was used get the list of venues for a neighborhood. Since free version of the Foursquare API is used, a maximum of 50 venues can be retrieved as per (v3 API) which used in Foursquare venues.

City	No. of Neighborhoods	Avg Neighborhood Radius Considered (m)
Cairo	32	1000
Istanbul	38	1000
Marrakesh	33	1000

- The venues for each location are stored in separate dataframes.

To study the venues, the dataframes containing the venues are grouped by neighborhoods and summed up.

Number of venues per each neighborhood depends on the Capital Geographical, Taken radius.

As per above table: A 1000 m Radius was used for all 3 Capitals, which can be changed later on to enhance our results and get more venues for every Neighborhood.

For **Cairo** and **Istanbul,** we see many venues for ever neighborhood, but Marrakesh venues are less. Realistically this is not true, and this can be considered as Foursquare not having detailed venues for all neighborhoods. My assumption is that venues in Foursquare are more recognizable and an international restaurant franchisee will like to be in neighborhoods with more recognizable venues.

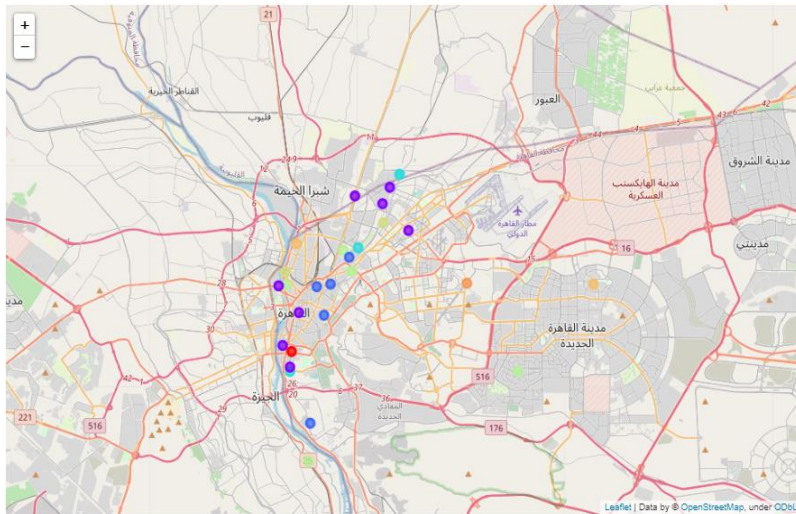## • <u>Clustering:</u>

Two Clustering Types where in the final results of this study:

<u>Individual Clustering</u> will help understand how the individual locations can be clustered. To be consistent with all the individual location clustering and the complete clustering, there are going to be 8 clusters.

It must be noted that cluster labels are not the same across different locations.

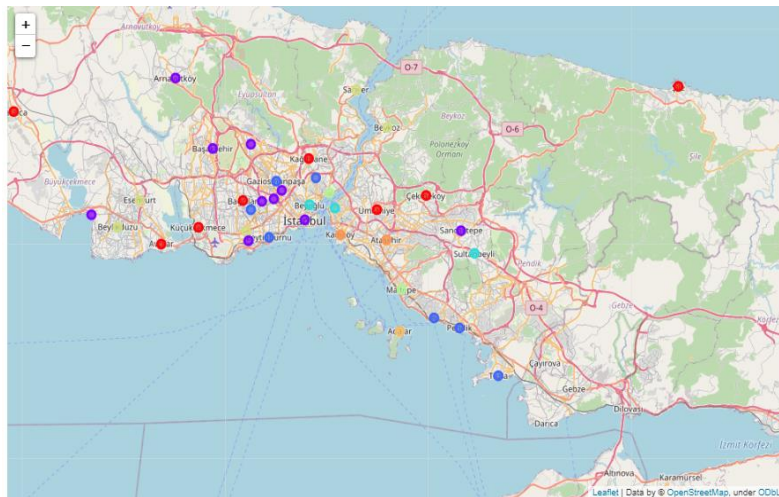Below are the summary of 3 Capital created (Individual Clustering):

Cairo Individual Clustering:

Cluster Labels	Neighbourhood
0	1
1	9
2	5
3	2
4	3
5	1
6	4
7	6

8 rows × 25 columns

Istanbul Individual Clustering:

Cluster Labels	Neighbourhood
0	9
1	10
2	7
3	5
4	1
5	2
6	3
7	2

8 rows × 25 columns

Marakesh Individual Clustering:
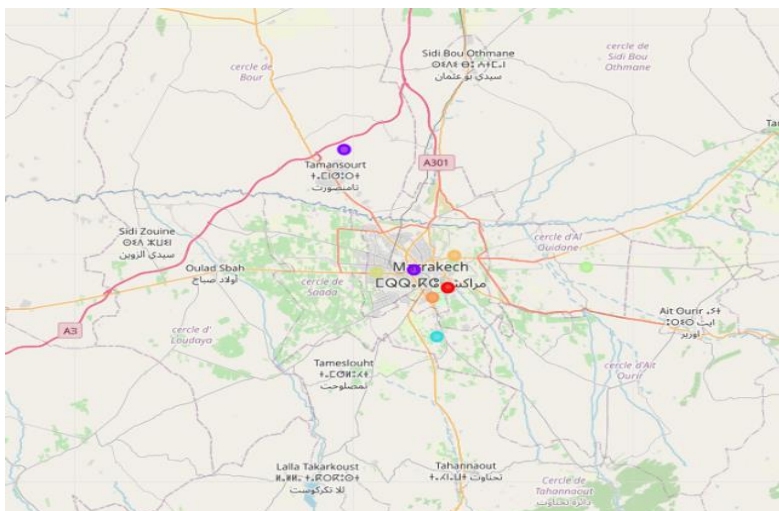
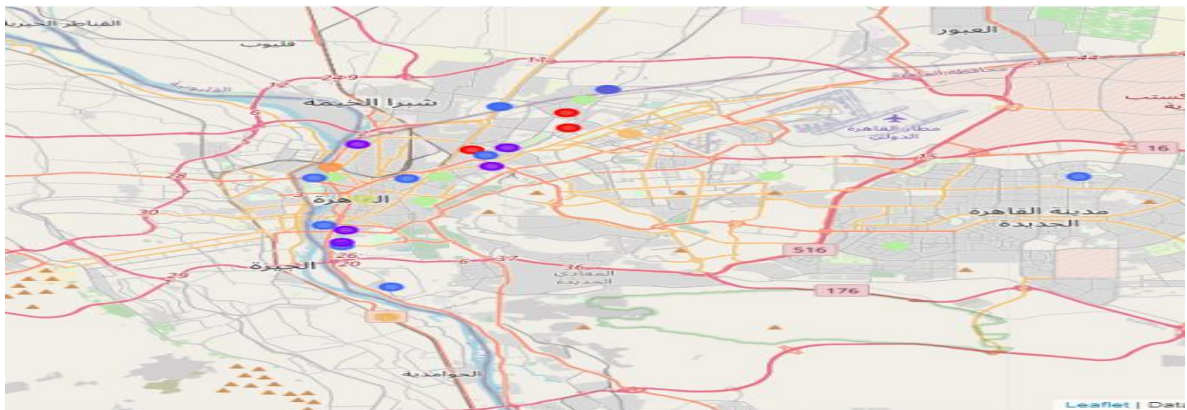| Cluster Labels	Neighbourhood
0	1
1	2
2	1
3	1
4	1
5	1
6	1
7	1

8 rows × 25 columns

Individual Clustering Summary:

I. In Cairo: There are 30 Clusters, (1,2,7) having   most of the neighborhoods.
II. In Istanbul: There are 39 Clusters, (0,1,2) having most of the neighborhoods.
III. In Marrakesh and as mentioned in the clustering map, very few neighborhoods available, and it's difficult to judge if cluster with 1 neighbor is outlier since most on them having 1 or 2 neighbors.

Note: for Marrakesh (10 out of 19 neighborhoods returned zero values by foursquare) and were not appended in the dataframe, its recommended to regenerate the data of neighborhoods again to get more venues, and that can be done by increasing the radius as mentioned above also by correcting some correcting names of neighborhoods which were rejected by foursquare .
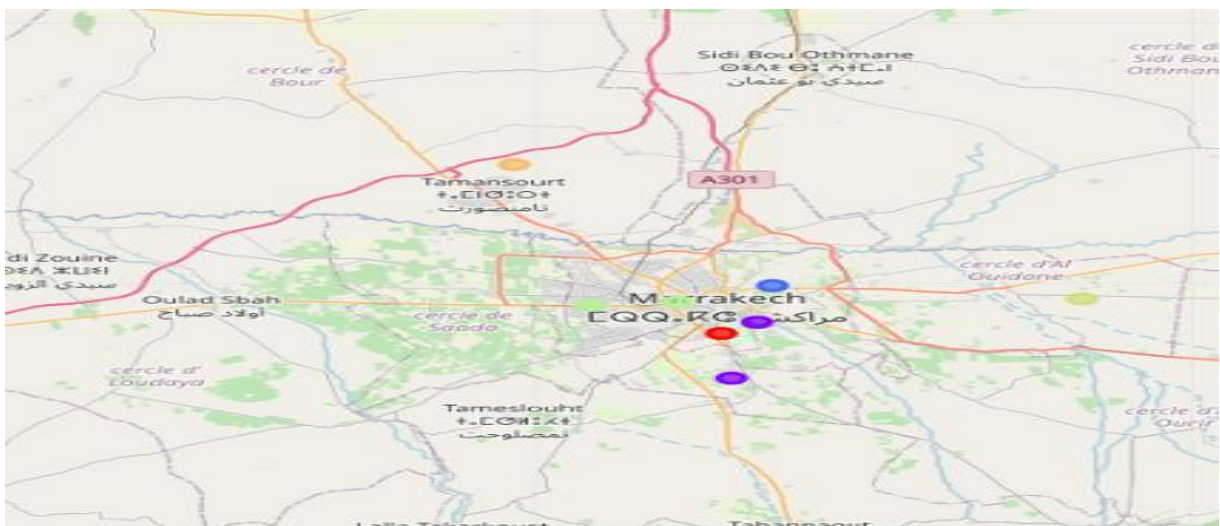
Complete Clustering Results:

Cairo Complete Clustering:

Istanbul Complete Clustering:
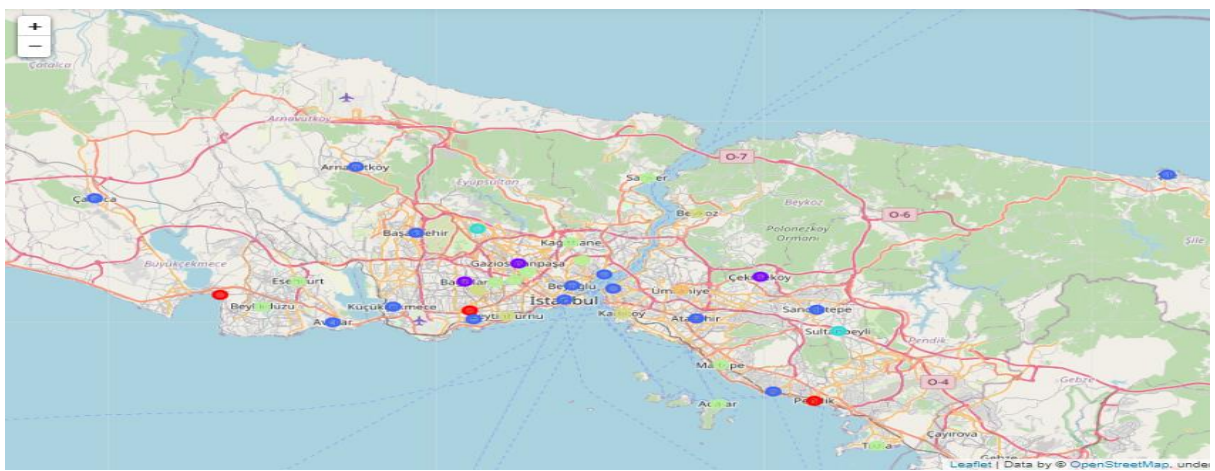
Marakesh Complete Clustering:

Cluster Labels	City	Neighbourhood
0	Cairo	4
	Istanbul	3
	Marrakesh	1
1	Cairo	5
	Istanbul	3
	Marrakesh	2
2	Cairo	12
	Istanbul	14
	Marrakesh	1
3	Cairo	1
	Istanbul	4
	Marrakesh	1
4	Cairo	2
	Istanbul	1
	Marrakesh	1
5	Cairo	1
6	Istanbul	3
7	Cairo	7
	Istanbul	11
	Marrakesh	3

20 rows × 25 columns

- Cluster 2 and 7 having highest Neighboring number for both Cairo and Istanbul

The big takeaway from this is that there are three clusters with Neighbourhoods from all the locations (2, 5, 6). So, these Neighbourhoods can be considered similar based on the venues present in them.

Discussion

The objective of this analysis was that if there is a restaurant in both Cairo and Istanbul and we want to open Marrakesh then in which Neighbourhoods of the cities they should open. Based on Complete Clustering Neighbourhoods in clusters 1,2 and 7 are similar Neighbourhoods. So, if the restaurant in the Neighbourhoods of these clusters in Cairo and Istanbul then a new franchise can be opened in the Neighbourhoods of the same clusters in Marrakesh and . Since majority of the Neighbourhoods of all the locations are in these three clusters then there is a good probability of finding a match.

Thank you

Ihssan Alfaqeah