Choosing a place to open a new mall in Cairo

IBM Data Science Capstone Project

Done by Mostafa El Far

Email: mostafaelfar96@gmail.com

Description of the project

Located on the banks of the Nile River, Cairo is Africa's largest city, as well as the largest city in the Arab world. In the course of its thousand-year history it has been the capital of the great Egyptian dynasties of the Middle Ages, a British colonial enclave, and a modern industrialized city. Today it is a teeming, vibrant national capital with one of the world's highest population densities per square kilometer. Even as the city struggles with the social and environmental effects of overcrowding, it dominates Egypt as well as the region politically, economically, and culturally and remains a prime tourist destination.

Cairo has set a shopping benchmark for the region, with over 24 huge malls, including ones of the biggest in the world and over 100 smaller shopping centers as well as dozens of street markets throughout the city. It also has many street markets which are locally called 'Souq' and tourists as well as locals can find anything, they desire in those shopping spots. One can find everything ranging from handmade clothes, ointments, spices to designer clothes, expensive jewelry and watches.

Cairo's market is ever growing. Egypt has remained the number one attractor of foreign investments in Africa with a direct foreign investment (DFI) of $7.9 Billion in 2018 alone, this number is just rising in 2019. This also affects shopping in Egypt; with the market expanding, opportunities of new malls are opening to accommodate the rise in demand. Many investors are already planning on building new malls in Egypt, for instance, Majid Al-Futtaim is planning to open 40 new properties in Egypt by the end of 2022.

Looking on Cairo, with a population of over 20 million, endless possibilities of opening new shopping centers are apparent.

Question to be answered

This project aims to tell investors who are interested in opening a mall in Cairo, which are the best neighborhoods to do so. Several factors can be included in finding out the optimum answer:

Population density of the neighborhood should be high enough to have enough traffic on the mall

Since Cairo is a relatively poor city, the mall should be opened in a relatively rich neighborhood

The mall should be opened close to other shopping malls since competition attracts customers, however it should not be too close in order to be able to thrive against competition.

Data collection

A list of all neighborhoods of Cairo and some of its demographics can be found in this website: https://www.citypopulation.de/php/egypt-greatercairo.php Location API from foursquare will be used to collect data on the attractions of each neighborhood, this data will be used to analyze the income of each neighborhood and which ones are richer than the others. Also, location data API will be used to find information about all malls and shopping centers in Cairo.

Methodology and Discussion

First using beautiful soup, data was scraped from the website stated above. A dataframe of all the neighborhoods was obtained. Also, the data for the population, area, longitude and latitude was obtained from the website. Population density is easy to calculate, it is done by dividing population over area. The final df obtained looked something like this.

	Name	Population	Area/ sqkm	Link	Latitude	Longitude	Population Density
0	15 Māyū [15th of May City]	96522.0	75.99	/en/egypt/greatercairo/0104__15_māyū/	29.833	31.384	1270.193447
1	'Ābidīn	41605.0	1.72	/en/egypt/greatercairo/0116__ābidīn/	30.044	31.243	24188.953488
2	Ad-Darb al-Aḥmar	60336.0	1.87	/en/egypt/greatercairo/0114__ad_darb_al_aḥmar/	30.041	31.258	32265.240642
3	Ad-Duqqī	73309.0	5.46	/en/egypt/greatercairo/2103__ad_duqqī/	30.039	31.205	13426.556777
4	'Ain Schams	633798.0	8.32	/en/egypt/greatercairo/0134__ain_schams/	30.122	31.327	76177.644231
5	Al-Ahrām	681383.0	17.96	/en/egypt/greatercairo/2108__al_ahrām/	29.994	31.138	37938.919822
6	Al-'Ajūzah	287818.0	7.35	/en/egypt/greatercairo/2102__al_ajūzah/	30.062	31.199	39158.911565
7	Al-Amīrīīah	157378.0	3.80	/en/egypt/greatercairo/0131__al_amīrīīah/	30.106	31.293	41415.263158
8	Al-Azbakiyah [Azbakeya]	20393.0	1.33	/en/egypt/greatercairo/0120__al_azbakiyah/	30.057	31.245	15333.082707
9	Al-Badrashayn [Badrshein]	563294.0	135.10	/en/egypt/greatercairo/2111__al_badrashayn/	29.823	31.258	4169.459660

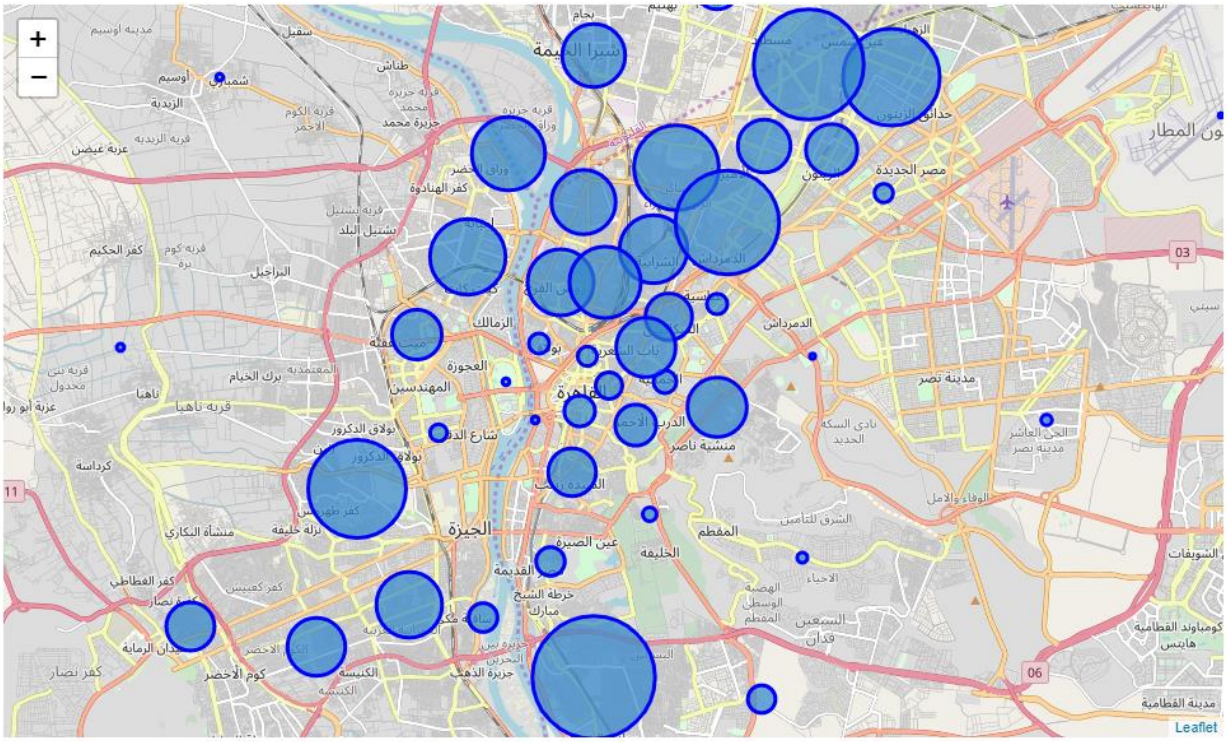
Then, using foursquare API search queries, numbers of banks and malls in each neighborhood was obtained and the density of banks was calculated.

	Name	Population	Area/ sqkm	Link	Latitude	Longitude	Population Density	Number of Banks	Banks Density
0	15 Māyū [15th of May City]	96522.0	75.99	/en/egypt/greatercairo/0104__15_māyū/	29.833	31.384	1270.193447	7	0.000073
1	'Ābidīn	41605.0	1.72	/en/egypt/greatercairo/0116__ābidīn/	30.044	31.243	24188.953488	23	0.000553
2	Ad-Darb al-Aḥmar	60336.0	1.87	/en/egypt/greatercairo/0114__ad_darb_al_aḥmar/	30.041	31.258	32265.240642	2	0.000033
3	Ad-Duqqī	73309.0	5.46	/en/egypt/greatercairo/2103__ad_duqqī/	30.039	31.205	13426.556777	45	0.000614
4	'Ain Schams	633798.0	8.32	/en/egypt/greatercairo/0134__ain_schams/	30.122	31.327	76177.644231	6	0.000009

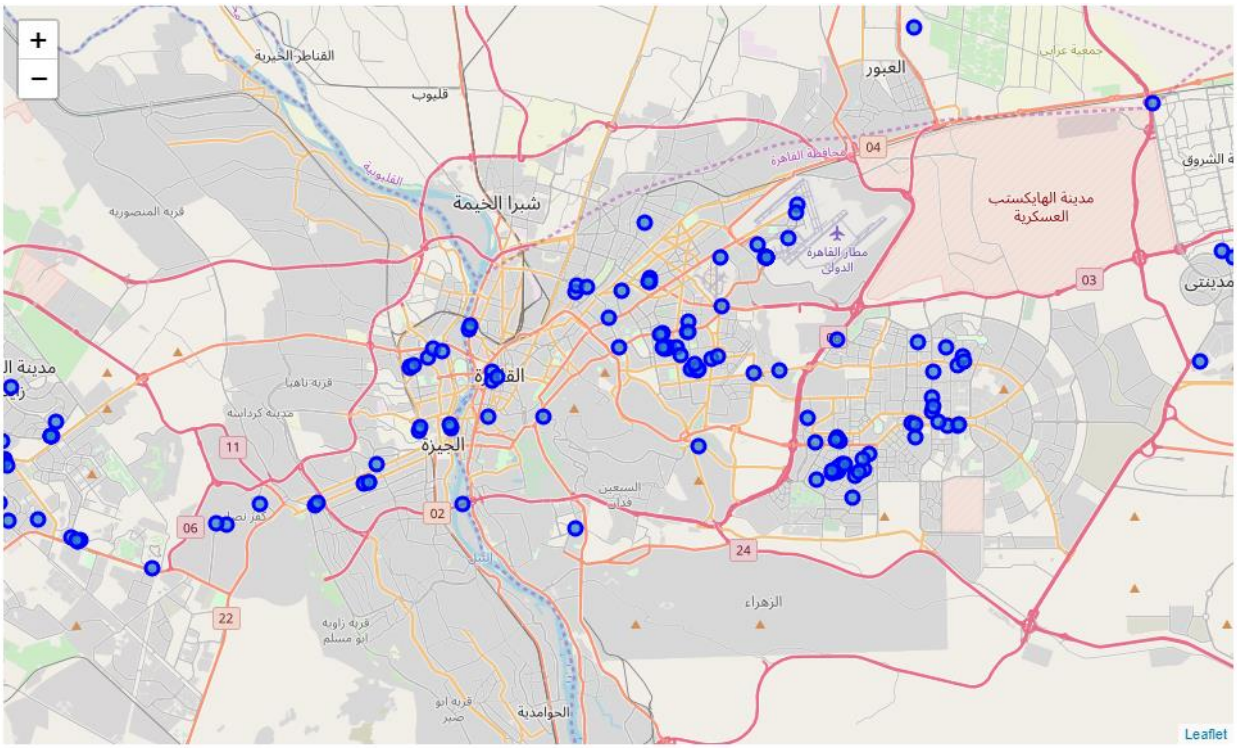
Also, another dataframe with the location of all malls was obtained.

To visualize the data, maps were built with the locations of neighborhoods superimposed on it. Each neighborhood was showed as a circle with its radius proportional to the population density. This map was

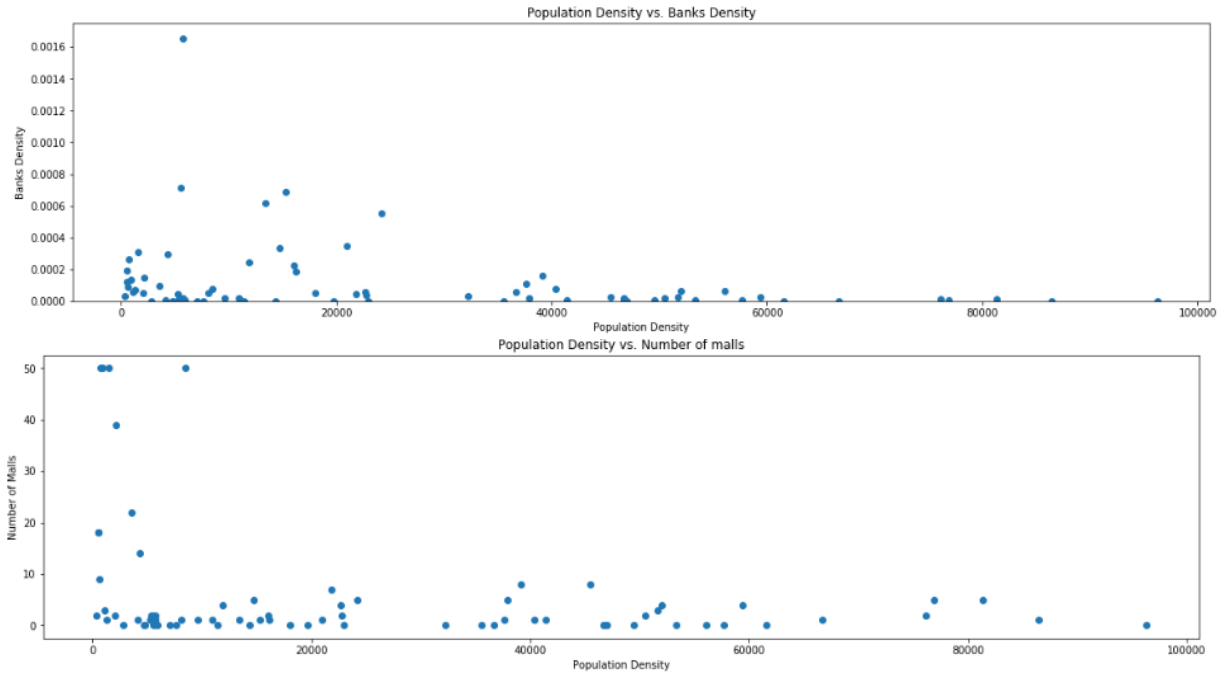
used to understand the distribution of the population in Cairo.



Another map was built with all the malls superimposed on it.



To be able to analyze the data, relationships between the variables were checked and looking for trends. Number of malls and Banks density was plotted against population density to check for trends.

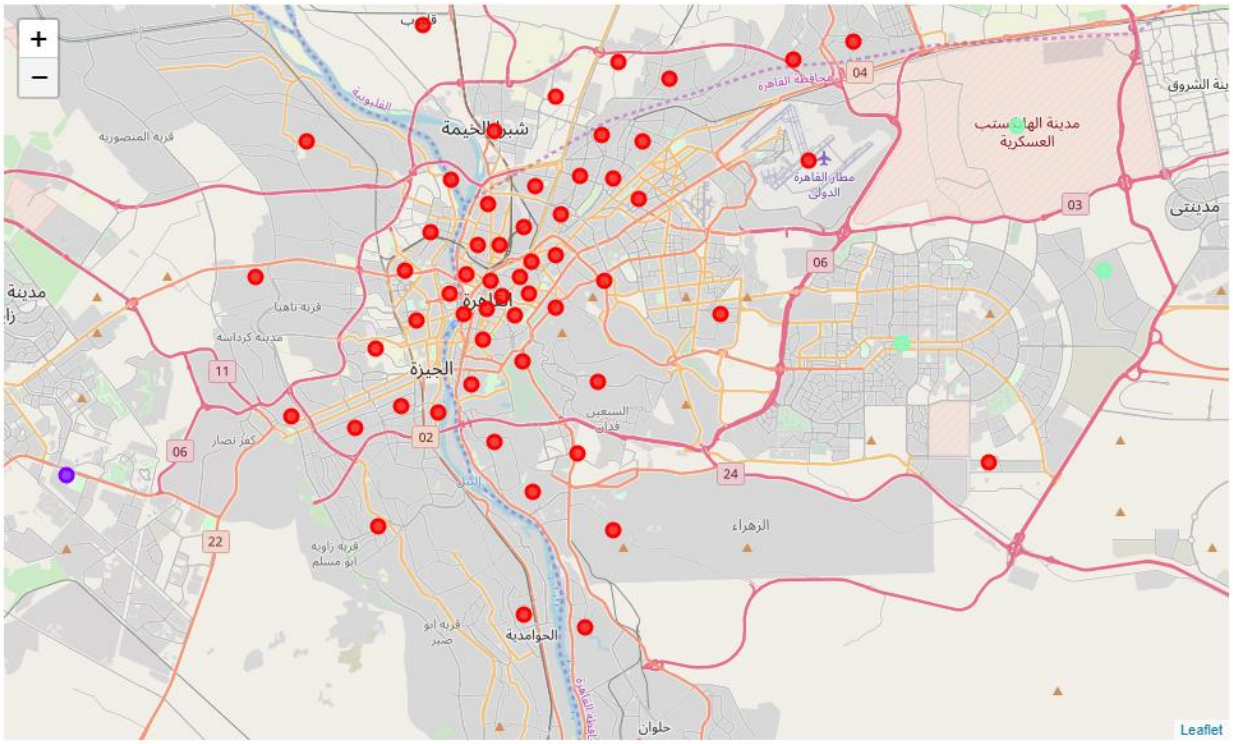


When the population density is low, banks density and number of malls increase. This is due to several factors. When population density decreases, places are easier to be accessible and people can reach malls faster. Also, richer people tend to live in bigger places, i.e. big houses, that is why there are many banks where the population density is low. Also, malls tend to open closer to richer people.

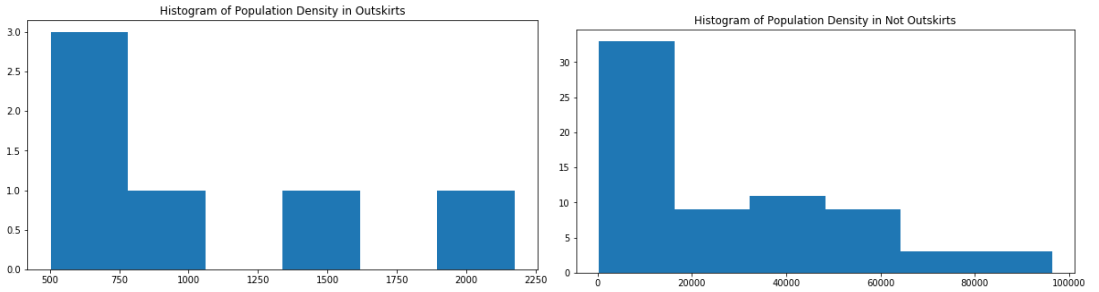
Thus, the pearson correlation was calculated. Correlation between population density and number of malls is -0.2911016155620632 while correlation between population density and banks density is -0.24923281466862976.

Then, to distribute the data, k-means clustering algorithm was used. It produced three clusters. Clusters 1 and 2 where the newer parts of Cairo while cluster 0 are the older parts. This was visualized using a map where the red circles are cluster 0 while the other circles were clusters 1 and 2.

To understand how the clustering algorithm worked the means of the data was calculated. The mean population density of cluster 2 and 3 is 1061.2161632523605 while the mean population density of cluster 0 is 27582.16722571647. The mean banks density of cluster 2 and 3 is 0.00019400665533550032 while the mean banks density of cluster 0 is 0.00010882332013127793. Outskirts have a much less population density, of just 1061 /sq. km compared to a whopping 27,580 in places inside the city. Also, the bank density is double than that of the places inside the city. Outskirts are generally less crowded and more expensive.



The data was divided into 2 groups, outskirts and not outskirts, where not outskirts are cluster 0 while outskirts are clusters 1 and 2. The distribution of the data in the 2 groups was analyzed by analyzing the histogram of them.

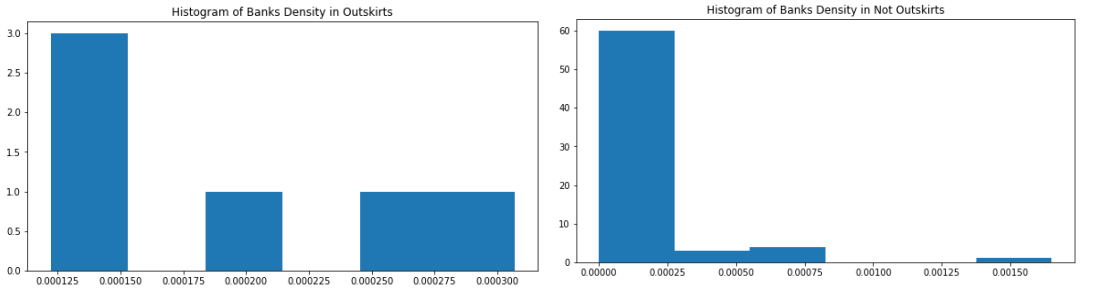



```
plt.figure(figsize = (10,5))
plt.title('Histogram of Banks Density in Outskirts')
plt.hist(neighbourhoodsOut['Banks Density'], bins = 6)
```

```
(array([3., 0., 1., 0., 1., 1.]),
 array([0.00012218, 0.00015296, 0.00018375, 0.00021453, 0.00024532,
        0.0002761 , 0.00030689]),
 <a list of 6 Patch objects>)
```

```
plt.figure(figsize = (10,5))
plt.title('Histogram of Banks Density in Not Outskirts')
plt.hist(neighbourhoodsNOut['Banks Density'], bins = 6)
```

```
(array([60.,  3.,  4.,  0.,  0.,  1.]),
 array([0.        , 0.0002752 , 0.00055041, 0.00082561, 0.00110082,
        0.00137602, 0.00165122]),
 <a list of 6 Patch objects>)
```



Again, from the histograms, outskirts are much less crowded and have a higher banks density. Not outskirts histogram of banks density shows that some areas have very low banks density while some still have high. Let's try to cluster not outskirts again. Not outskirts data was clustered into 3 clusters again; 0,1 and 2. The histograms were analyzed. Cluster 0 has very high population density as well as higher banks density than cluster 1 and 2. Cluster 1 and 2 have very low banks density. This means that these 3 clusters will not be suitable places for a mall. They are either very crowded and there will not be easy access to the mall or do not have high salaries, which will make the malls not profitable.

Results

All this analysis show that the only possible places to open mall are the outskirts that were collected in the first cluster. They have relatively low or moderate population density that will allow the newly built mall to be easily accessible in the crowds of Cairo. Also, these places are richer than other places that are not crowded. They also tend to have a higher number of malls which indicates that the shopping in these areas is actively growing.

	Name	Population	Area/ sqkm	Link	Latitude	Longitude	Population Density	Number of Banks	Banks Density	Number of Malls
25	Al-Qāhirah al-Jadīdah 1 [New Cairo]	140117.0	91.37	/en/egypt/greatercairo/0142__al_qāhirah_al_jad...	30.028	31.466	1533.512094	43	0.000307	50
26	Al-Qāhirah al-Jadīdah 2 [New Cairo]	93534.0	186.00	/en/egypt/greatercairo/0143__al_qāhirah_al_jad...	30.062	31.575	502.870968	18	0.000192	18
35	Ash-Shaykh Zāyid [Sheikh Zayed City]	93742.0	43.12	/en/egypt/greatercairo/2119__ash_shaykh_zāyid/	30.047	30.980	2173.979592	14	0.000149	39
36	Ash-Shurūq	90033.0	175.00	/en/egypt/greatercairo/0145__ash_shurūq/	30.129	31.528	514.474286	11	0.000122	18
61	Madīnat Sittah Uktūbar 2 [6th of October City]	202945.0	227.80	/en/egypt/greatercairo/2121__madīnat_sittah_uk...	29.931	30.886	890.891133	27	0.000133	50
62	Madīnat Sittah Uktūbar 1 & 3 [6th of October C...	157604.0	209.70	/en/egypt/greatercairo/2120__madīnat_sittah_uk...	29.967	31.017	751.568908	41	0.000260	50

Conclusion

To conclude this report. Data science was used to explore the possible neighborhoods that would be suitable to open a new mall. Data was collected from web scraping citypopulation.de and from the Foursquare API. The data collected included information like neighborhood names, population, population density, banks and malls. Several data analysis techniques were used like clustering. Data

visualization tools was also used to completely understand the data, such as heat maps, maps and histograms. In the end, a few neighborhoods were selected to be the most suitable for opening up a mall.