Capstone Project - Report

Introduction

This project aims to find potentially useful regions for opening new coffee shops in Baku city.

As a tradition, Azerbaijan people mostly drink tea, while westerners like coffee. So there are many tea houses called "chaykhana" where people go to drink tea and play board games (especially Dominos, Backgammon). However, the atmosphere of "chaykhana"s are noisy, and it is not suitable for study, work, reading. In contrary, coffee shops have a satisfactory atmosphere for such activities as reading. In recent years Baku people are looking for quiet places to study, to talk or to have some activities which are done better in quiet locations. Considering the concerns mentioned below, we can think that coffee shops seem to be a great business to do as there is a demand. Recently opened coffee shops are good examples to prove the feasibility of the idea.

In this project, I conducted research using data science techniques and find optimal places for opening a coffee shop that might be interesting to stakeholders. An optimal location means they meet specific criteria which may make the coffee shop more profitable.

Note: I used some methods from [https://cocl.us/coursera_capstone_notebook (https://cocl.us/coursera_capstone_notebook)](https://cocl.us/coursera_capstone_notebook) [1] and [https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/ (https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/)](https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/) [2].

Data

Usually, people in Baku meet near to "city centre" to have their business meetings. The city centre is the area which around "Sahil" subway station. The main reason I guess is ease of accessibility and presence of many ways to get in there using public transport. Hence, we choose that location as a central point. Additionally, other regions like:

- "Elmlar Akademiyasi" station;
- "28 May" station;
- "Khatai" station are also good candidates for such activities because they are very close to the city centre.

The existing coffee shops in those neighbourhoods have a good profit. The distance from "Sahil" to each of these places are not more than **3-4 kilometres**. Thus, I use a 4km radius to search for potentially profitable locations.

I used Geocoding API [3] from Google Cloud (paid) and Foursquare API [4] for getting information about points of interests. Additional tools and resources are used to do various manipulations on the data to improve quality.

To summarize:

- The centre point are coordinates of "Sahil" subway station
- The radius for discovery is 4 kilometres
- Geocoding API from Google Cloud is used for converting addresses to coordinates and vice-versa.
- I used Foursquare API for obtaining information about points of interest.

Data collection

Data collection process involves obtaining data from Foursquare using their API, generation of neighbourhoods around the city centre, and getting their addresses using Google Geocode API.

Generate neighbourhoods

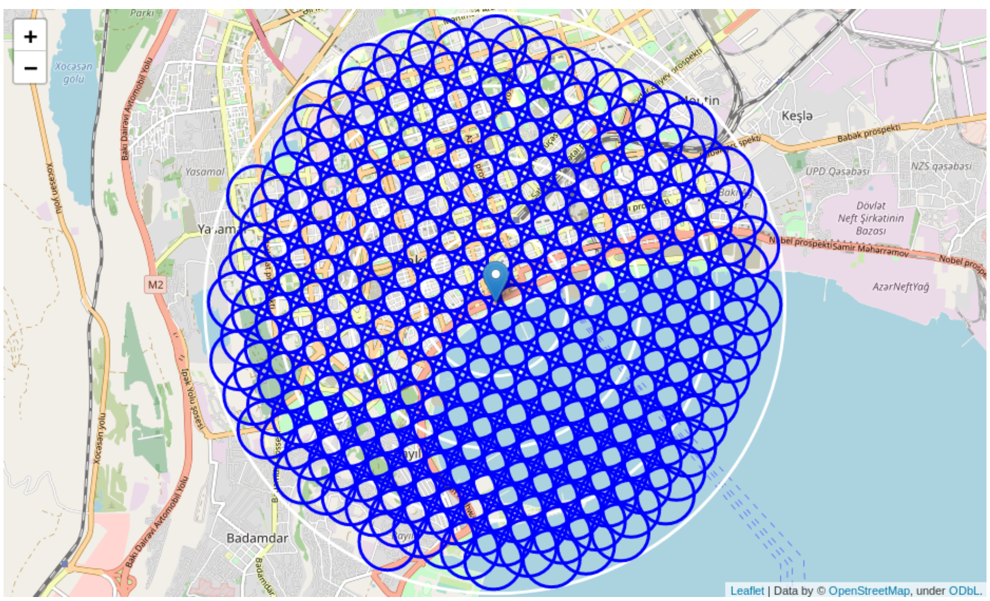
First, we need to define a central point and then generate candidate areas within a 4km radius around it. Let's find a central point for Baku. As we discussed in the previous section, we will consider "Sahil" subway station as a central point(city centre/downtown).

The coordinates of "Sahil" subway station are **[40.3702583, 49.8462667]**. Let's mark the candidate area. The candidate area is a circular area within **4 kilometres** from the city centre. Moreover, I divided the area into **400 meters** small neighbourhoods to make a grid.

To do so, I am creating a rectangular grid using two nested loops and checking if it lies inside a candidate area. I am using the equation of a circle to find if a point is inside a circle.

Another problem here is that the values for coordinates are in degrees. It is challenging to compute coordinates. Thus I convert them into a Cartesian coordinate system using World Geodetic System (WGS) [5] also known as WGS84 (latest revision).

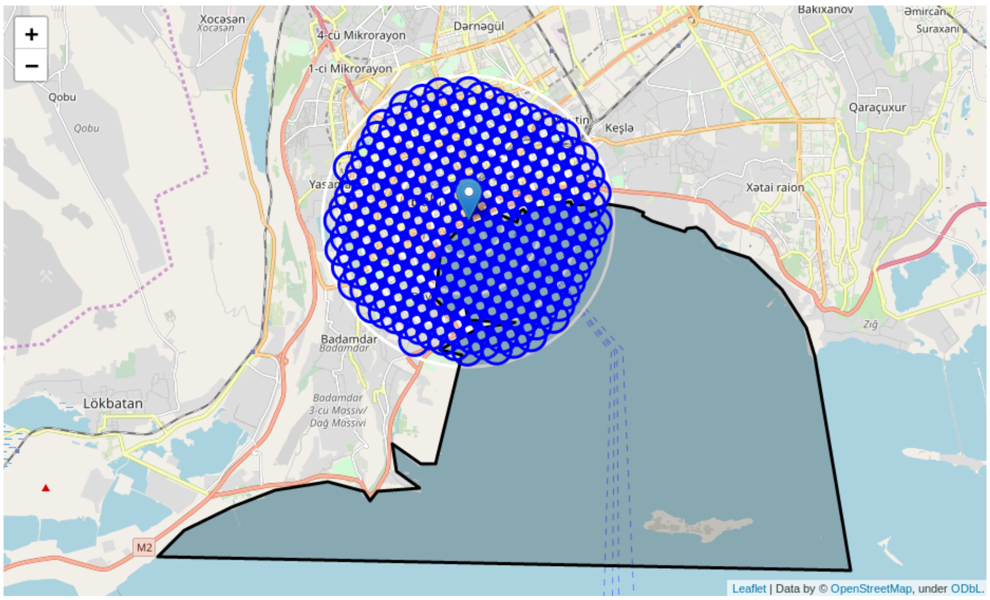
The generated are shown on the map.



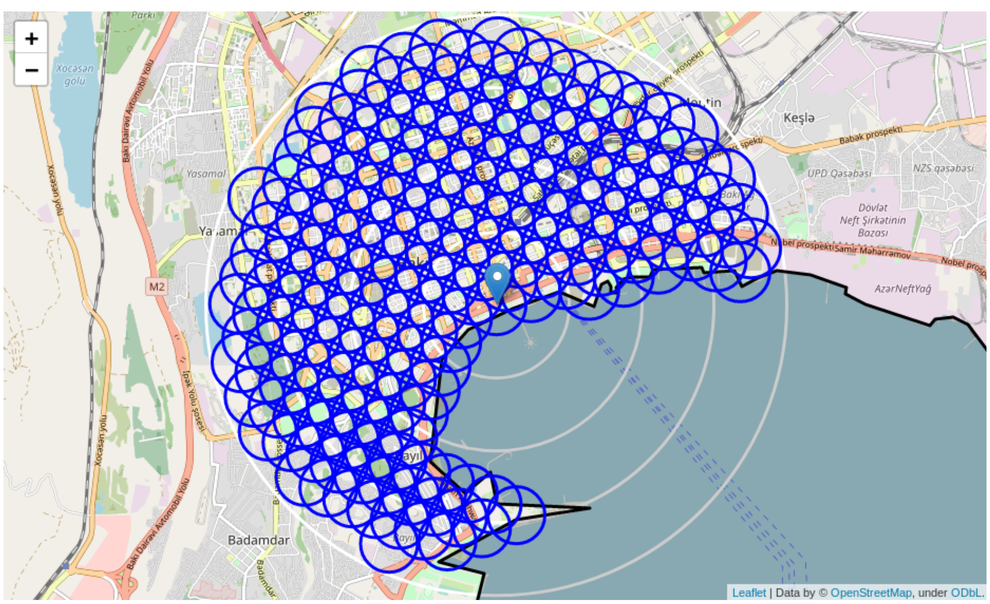
As seen in the map, we generated a grid of circular areas which are as candidate regions. Meanwhile, there is a problem here related to the Caspian Sea. Our grid contains regions that are in the sea area, which can not be used for commercial purposes.

First of all, we need to find a way to remove those regions because we can not propose them as a candidate area. To do so, we need to obtain polygon of the sea. It is okay to use shapefile of Caspian Sea [6]. The reason why I do not want to use it is that it is too vast and too detailed. Thus it would be difficult and slow to compute if the dots are inside it.

Instead, I used Google Maps [7] to make a shapefile, which contains an area in our candidate regions, while smaller enough in size. Google Maps allows us to download it in KML [8] file format. Then I convert it into GeoJSON [9] format using **mygeodata.cloud** tool [10].



The sea areas are excluded from map.

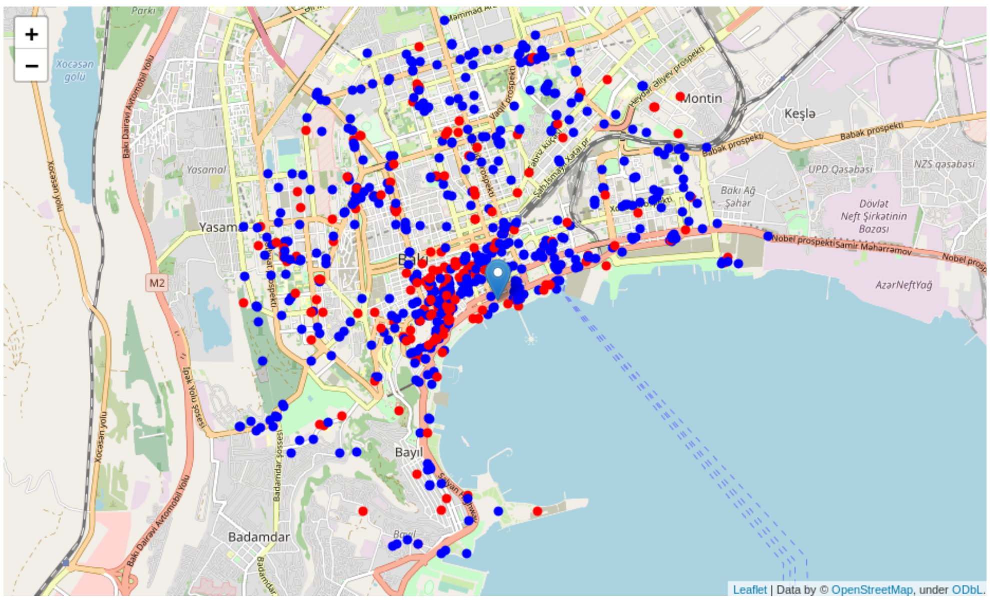


After having the candidate regions the dataframe with locations and addresses is made. The sample data is shown in the table:

	Addr	Lat	Lon	X	Y	DistFromCenter
0	47 Həsən bəy Zərdabi, Bakı, Azerbaijan	40.383597	49.807604	3.474268e+06	5.097769e+06	4000.000000
1	19 Abbas Mirzə Şərifzadə küçəsi, Bakı, Azerbaijan	40.370433	49.804498	3.474668e+06	5.096169e+06	3939.543121
2	12 Abbas Mirzə Şərifzadə küçəsi, Bakı, Azerbaijan	40.373391	49.806241	3.474668e+06	5.096569e+06	3794.733192
3	31/38, Mirali Seyidov, Baku, Azerbaijan	40.376348	49.807984	3.474668e+06	5.096969e+06	3687.817783
4	60 Matbuat avenue, Baku, Azerbaijan	40.379306	49.809728	3.474668e+06	5.097369e+06	3622.154055
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Using the data above we can obtain restaurants and coffee shops around them using Foursquare API. **746** restaurants and **200** coffee shops are found as shown on the map.

Finally, data collection process is done. Let's see our final data on the map.



Methodology

Once we have data, it is time to conduct an analysis. Our method uses to search for regions within 4 kilometres from the city centre. The main criteria for that are followings:

- low density of restaurants
- less number of coffee shops around the area

The next step is to create heatmaps to see the conditions visually. Then we choose the regions that meet the criteria below.

The last step is finding clusters of the locations that meet our criteria for candidate areas. We use k-means clustering for this purpose.

The table below includes the previous table by adding two additional columns:

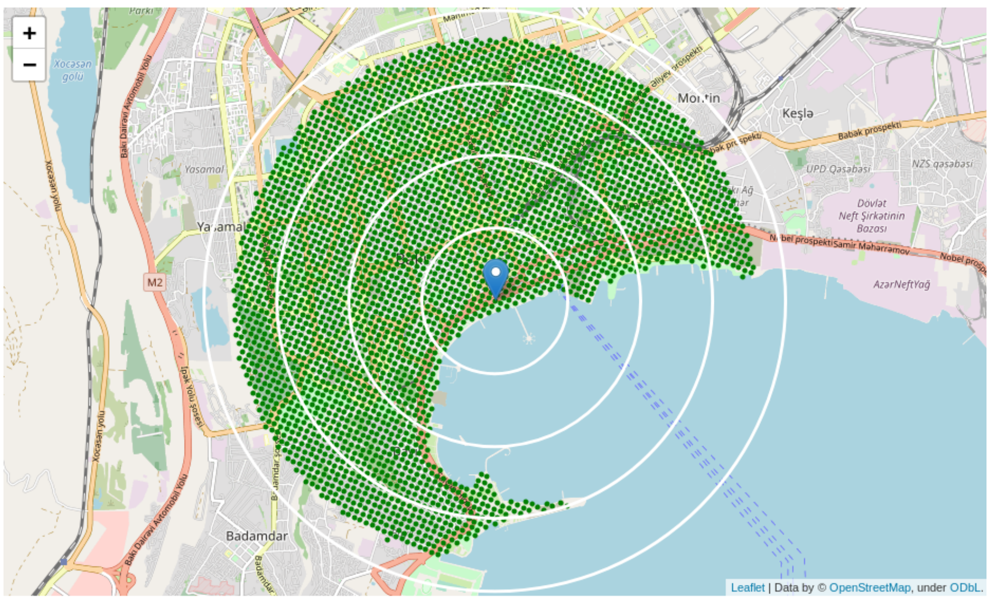
- Distance from the city center
- Number of restaurants within close distance

	Unnamed: 0	Addr	Lat	Lon	X	Y	DistFromCenter	Restaurants in area
0	0	47 Həsən bəy Zərdabi, Bakı, Azerbaijan	40.383597	49.807604	3.474268e+06	5.097769e+06	4000.000000	0
1	1	19 Abbas Mirzə Şərifzadə küçəsi, Bakı, Azerbaijan	40.370433	49.804498	3.474668e+06	5.096169e+06	3939.543121	3
2	2	12 Abbas Mirzə Şərifzadə küçəsi, Bakı, Azerbaijan	40.373391	49.806241	3.474668e+06	5.096569e+06	3794.733192	1
3	3	31/38, Mirali Seyidov, Baku, Azerbaijan	40.376348	49.807984	3.474668e+06	5.096969e+06	3687.817783	4
4	4	60 Matbuat avenue, Baku, Azerbaijan	40.379306	49.809728	3.474668e+06	5.097369e+06	3622.154055	8
5	5	52 Zahid Xəlilov Küçəsi, Bakı, Azerbaijan	40.382263	49.811471	3.474668e+06	5.097769e+06	3600.000000	5
6	6	5B/2 Əhməd Cəmil Küçəsi, Bakı, Azerbaijan	40.385221	49.813215	3.474668e+06	5.098169e+06	3622.154055	3
7	7	273c Şəfayət Mehdiyev Küçəsi, Bakı, Azerbaijan	40.388178	49.814959	3.474668e+06	5.098569e+06	3687.817783	2
8	8	95 Shafayat Mehdiyev, Bakı, Azerbaijan	40.391135	49.816704	3.474668e+06	5.098969e+06	3794.733192	3
9	9	41 Mosvka prospekti, Bakı, Azerbaijan	40.394092	49.818448	3.474668e+06	5.099369e+06	3939.543121	6

The table below shows minimal distance to a nearest coffee shop from every candidate area.

	Unnamed: 0	Addr	Lat	Lon	X	Y	DistFromCenter	Restaurants in area	Distance to Coffee shop
0	0	47 Həsən bəy Zərdabi, Bakı, Azerbaijan	40.383597	49.807604	3.474268e+06	5.097769e+06	4000.000000	0	500.166878
1	1	19 Abbas Mirzə Şərifzadə küçəsi, Bakı, Azerbaijan	40.370433	49.804498	3.474668e+06	5.096169e+06	3939.543121	3	73.926962
2	2	12 Abbas Mirzə Şərifzadə küçəsi, Bakı, Azerbaijan	40.373391	49.806241	3.474668e+06	5.096569e+06	3794.733192	1	414.980535
3	3	31/38, Mirali Seyidov, Baku, Azerbaijan	40.376348	49.807984	3.474668e+06	5.096969e+06	3687.817783	4	115.452467
4	4	60 Matbuat avenue, Baku, Azerbaijan	40.379306	49.809728	3.474668e+06	5.097369e+06	3622.154055	8	178.100199

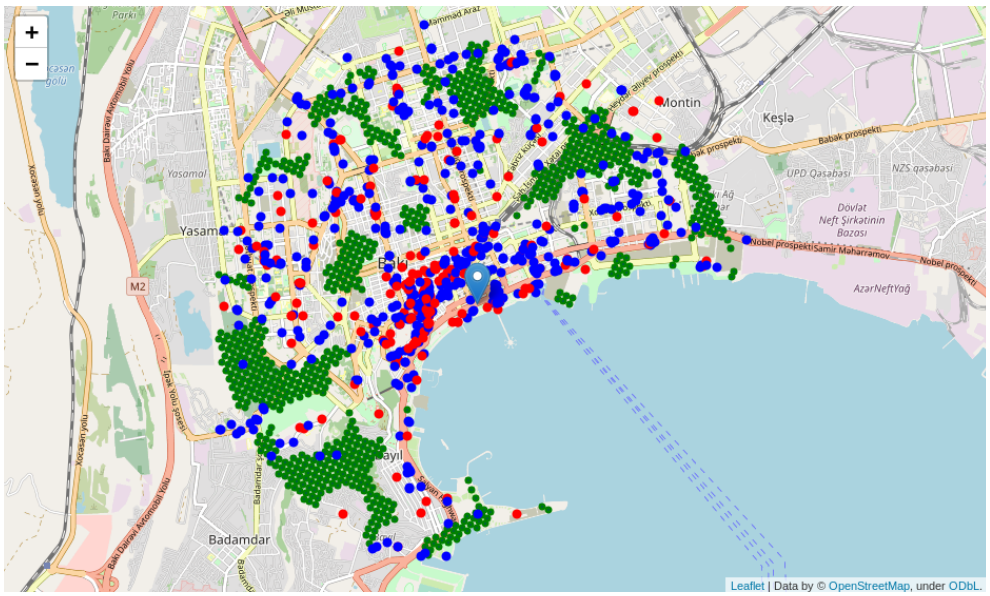
The same way as we did in data collection section, let's generate candidate areas again. The only difference this time is we use smaller regions, to have them in large quantity. These points are used to analyse if they are good candidates for coffee shops. "Bad" ones are removed. Then we perform cluster analysis with remaining regions.



Now lets make a dataframe of the points we generated. Then add columns to indicate distance to closes coffee shop and number of restaurants.

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to coffee shop
0	40.383597	49.807604	3.474268e+06	5.097769e+06	0	500.166878
1	40.377348	49.805084	3.474368e+06	5.096969e+06	1	229.107563
2	40.378088	49.805520	3.474368e+06	5.097069e+06	3	216.347223
3	40.378827	49.805956	3.474368e+06	5.097169e+06	4	200.633073
4	40.379567	49.806391	3.474368e+06	5.097269e+06	3	138.974544

Let's exclude "bad" candidates from our dataframe.



In the above map, we can see regions of good locations using green dots, restaurants and coffee shops with blue and red dots respectively.

It is time to run a cluster analysis using the K-means algorithm.

Let's obtain addresses and make a dataset using our final data.

	Address	Distance from center
0	Mikayıl Müşfiq, Bakı, Azerbaijan	3246.581876
1	K.Səfərəliyeva 27, Azerbaijan	2061.212492
2	Sabail square, Bakı, Azerbaijan	3615.058854
3	5D Akim Abbasov Küçəsi, Bakı, Azerbaijan	3716.901936
4	Bakı Ağ Şəhər Ofis Binası, 25 Nobel Prospekti,...	3744.705898
5	108 Azadlıq prospekti, Bakı 1005, Azerbaijan	3453.558730
6	Gülbala Əliyev küçəsi, Bakı, Azerbaijan	3656.744731
7	32 Fətəli Xan Xoyski, Bakı, Azerbaijan	3804.749985
8	120 Zulfu Adigozalov, Bakı 1009, Azerbaijan	2029.205167
9	10b Babək Prospekti, Bakı 1025, Azerbaijan	3100.678547
10	Buxta küçəsi, Bakı, Azerbaijan	3544.381469
11	Yasamal, Bakı, Azerbaijan	3627.672853
12	4 Hənifə Ələsgərov, Bakı, Azerbaijan	2941.323526
13	62 Neftçilər Prospekti, Bakı 1010, Azerbaijan	1364.734406
14	Baku City Main Police Department, Bakı, Azerba...	2725.445937
15	59 Mərdanov Qardaşları, Bakı, Azerbaijan	1633.986264
16	ул. Нахчивани 15, кв. 257a, Baku, Azerbaijan	3715.676251
17	Yuksak Inshaat MTK, Bakı, Azerbaijan	3853.014812
18	Aydın Məmmədov, Bakı, Azerbaijan	2779.516665
19	Alley of Honor, Parlament Pros, Bakı, Azerbaijan	2593.147136
20	Süleyman Vəzirov küçəsi, Bakı, Azerbaijan	3861.141950
21	Unnamed Road, Bakı, Azerbaijan	3737.002851
22	2 Həsən Əliyev Küçəsi, Bakı, Azerbaijan	3752.532138
23	Fətəlixan Xoyski 75,, Baku, Azerbaijan	3586.975091
24	Unnamed Road, Bakı, Azerbaijan	3182.309132
25	Atatürk prospekti, Bakı, Azerbaijan	3048.419821
26	White City Boulevard, Bakı, Azerbaijan	2295.844071
27	1 Heydər Əliyev prospekti, Bakı 1000, Azerbaijan	3234.510095
28	Unnamed Road, Bakı, Azerbaijan	3207.181882
29	127 Nəriman Nərimanov Prospekti, Bakı 1009, Az...	2097.712320

Results

The result of our analysis shows that there are good neighbourhoods to open new coffee shops in Baku. At first it seemed that too many restaurants and coffee shops are there; however, after analysis, we found out that we still can open new ones not far from the city centre. Close distance to city centre guarantees good sales.

Discussion

We found 30 regions using the criteria like the low density of restaurants and less number of coffee shops around. However, we can imporve our analysis by eliminating some areas such as national parks, the places that are reserved for future construction areas.

Additionally, some areas which are a little far from the city centre also can be good candidates for opening coffee shops because there are many companies, universities around that. Criteria like rental fee also could be useful.

As we are doing this report for educational purposes, I did not go that much deeper.

Conclusion

My project aimed to find suitable areas for opening coffee shops. Suitable area means low restaurant density and less number of coffee shops around the area.

This kind of analysis can be beneficial for stakeholders who want to open a coffee shop and looking for a suitable address.

We used Google Geocode API, Foursquare API to obtain interesting information. Besides that, we used data science techniques to generate final data. Clustering was performed to find significant centres and their addresses.

Reference

1. https://cocl.us/coursera_capstone_notebook (https://cocl.us/coursera_capstone_notebook)
2. Sercan Yıldız, Housing Sales Prices & Venues Data Analysis of Istanbul (https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/ (https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/))
3. Google Geocoding API (https://developers.google.com/maps/documentation/geocoding/start (https://developers.google.com/maps/documentation/geocoding/start))
4. Foursquare API (https://developer.foursquare.com/places-api (https://developer.foursquare.com/places-api))
5. Wikipedia Contributors. (2019, June 5). World Geodetic System. Retrieved from Wikipedia website: https://en.wikipedia.org/wiki/World_Geodetic_System (https://en.wikipedia.org/wiki/World_Geodetic_System)
6. Nyu.edu. (2015). Boundary, Caspian Sea, 2015 - NYU Spatial Data Repository. [online] Available at: https://geo.nyu.edu/catalog/stanford-zb452vm0926 (https://geo.nyu.edu/catalog/stanford-zb452vm0926).
7. Google Maps (2019). Google Maps. [online] Google Maps. Available at: https://maps.google.com (https://maps.google.com) [Accessed 23 Jun. 2019].
8. Wikipedia Contributors (2019). Keyhole Markup Language. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Keyhole_Markup_Language (https://en.wikipedia.org/wiki/Keyhole_Markup_Language) [Accessed 23 Jun. 2019].
9. Wikipedia Contributors (2019). GeoJSON. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/GeoJSON (https://en.wikipedia.org/wiki/GeoJSON) [Accessed 23 Jun. 2019].
10. Mygeodata.cloud. (2018). KML to GeoJSON Converter Online - MyGeodata Cloud. [online] Available at: https://mygeodata.cloud/converter/kml-to-geojson (https://mygeodata.cloud/converter/kml-to-geojson) [Accessed 23 Jun. 2019].
11. Readthedocs.io. (2018). The Shapely User Manual — Shapely 1.6 documentation. [online] Available at: https://shapely.readthedocs.io/en/stable/manual.html (https://shapely.readthedocs.io/en/stable/manual.html) [Accessed 23 Jun. 2019].
12. Scikit-learn.org. (2010). 2.3. Clustering — scikit-learn 0.20.3 documentation. [online] Available at: https://scikit-learn.org/stable/modules/clustering.html#k-means (https://scikit-learn.org/stable/modules/clustering.html#k-means).

In [ ]:

```
```