Best Location for a Restaurant in Istanbul

Suleyman PAYASLI

April 7, 2021

1. Introduction: Business Problem

1.1 Background

Istanbul is the largest city in Turkey and the country's economic, cultural and historic center. The city, which is the heart of the country, straddles the Bosphorus strait, and lies in both Europe and Asia, with a population of over 15 million residents, comprising 19% of the population of Turkey. Istanbul had the eleventh-largest economy among the world's urban areas in 2018, and is responsible for 30 percent of Turkey's industrial output. The city houses international ports that link Europe and Asia. Istanbul is divided into 39 districts in total, more than any other province in Turkey. The city has numerous shopping centers, from the historic to the modern and is also well known for its historic seafood restaurants. Istanbul offers wide range of quality restaurants in its every districts and neighborhoods. Restaurants serves up world cuisine as Turkish, Italian, Mediterranean, French, Japan, Chinese and Indian foods.

1.2 Business Problem

In this project we will try to find an optimal location for a restaurant. In particular, this report will be targeted to stakeholders interested in opening new restaurant business in Istanbul, Turkey.

Since there are lots of restaurants in Istanbul we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no existing Turkish restaurants in vicinity. We would also prefer locations as close to the city center as possible, assuming that first two conditions are met. We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders. Objective of this study is to locate and recommend to those who wants to open a new restaurant in Istanbul which neighborhood will be best choice to start.

Consequently, the goal we want to reach with this project is to help a possible stakeholder to understand better the best possible boroughs (districts) for opening a restaurant with useful insights.

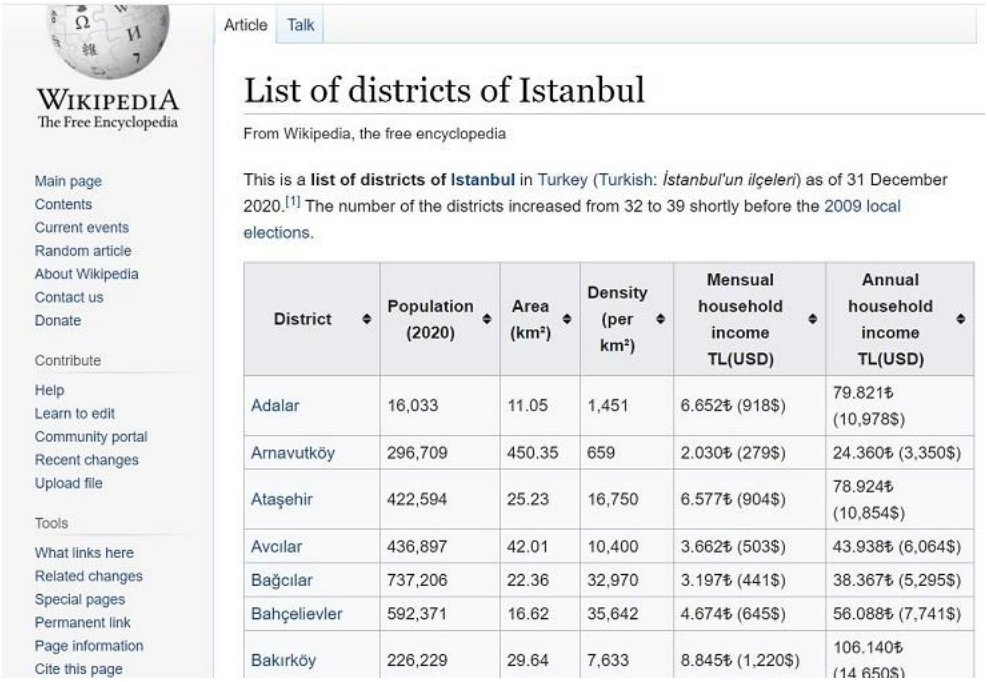
1.3 Target Audience

Entrepreneurs who are passionate about opening a restaurant in Istanbul would be very interested in this project. The project is also for business owners and stakeholders who want to expand their businesses and wonder how data science could be applied to the questions at hand. Another target audience would be foreign tourists and city visitors who wants to know where to find seafood or Turkish cuisine restaurants.

2. Data

Following data sources will be needed to extract or generate the required information:

Districts of Istanbul: the list of districts of Istanbul can be obtained from a Wikipedia page. We have to scrape the data. Here, you will find a HTML table which shows the 39 districts of Istanbul. The link as follows:



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Cite this page

Article Talk

List of districts of Istanbul

From Wikipedia, the free encyclopedia

This is a **list of districts of Istanbul** in Turkey (Turkish: *İstanbul'un ilçeleri*) as of 31 December 2020.[1] The number of the districts increased from 32 to 39 shortly before the 2009 local elections.

District	Population (2020)	Area (km²)	Density (per km²)	Mensual household income TL(USD)	Annual household income TL(USD)
Adalar	16,033	11.05	1,451	6.652₺ (918$)	79.821₺ (10,978$)
Arnavutköy	296,709	450.35	659	2.030₺ (279$)	24.360₺ (3,350$)
Ataşehir	422,594	25.23	16,750	6.577₺ (904$)	78.924₺ (10,854$)
Avcılar	436,897	42.01	10,400	3.662₺ (503$)	43.938₺ (6,064$)
Bağcılar	737,206	22.36	32,970	3.197₺ (441$)	38.367₺ (5,295$)
Bahçelievler	592,371	16.62	35,642	4.674₺ (645$)	56.088₺ (7,741$)
Bakırköy	226,229	29.64	7,633	8.845₺ (1,220$)	106.140₺ (14,650$)

Location Data: The location data of districts will be obtained from Geocoder API. This API provides geographic coordinates as latitude and longitude, which we can use to place markers on a map.

Foursquare API: We will look for the most common venues list, number of restaurants, their type, and location in every neighborhood will be obtained from Foursquare API. This allows us to make calls to the API to explore a region depending on the latitude and longitude, categories, radius and limit when calling the API.

With the datasets mentioned above, we will answer the following questions:

- What are the most popular venues?
- What neighborhoods have more restaurants?
- What is best the location in Istanbul to open a restaurant?
- What areas have the potential to open a restaurant?

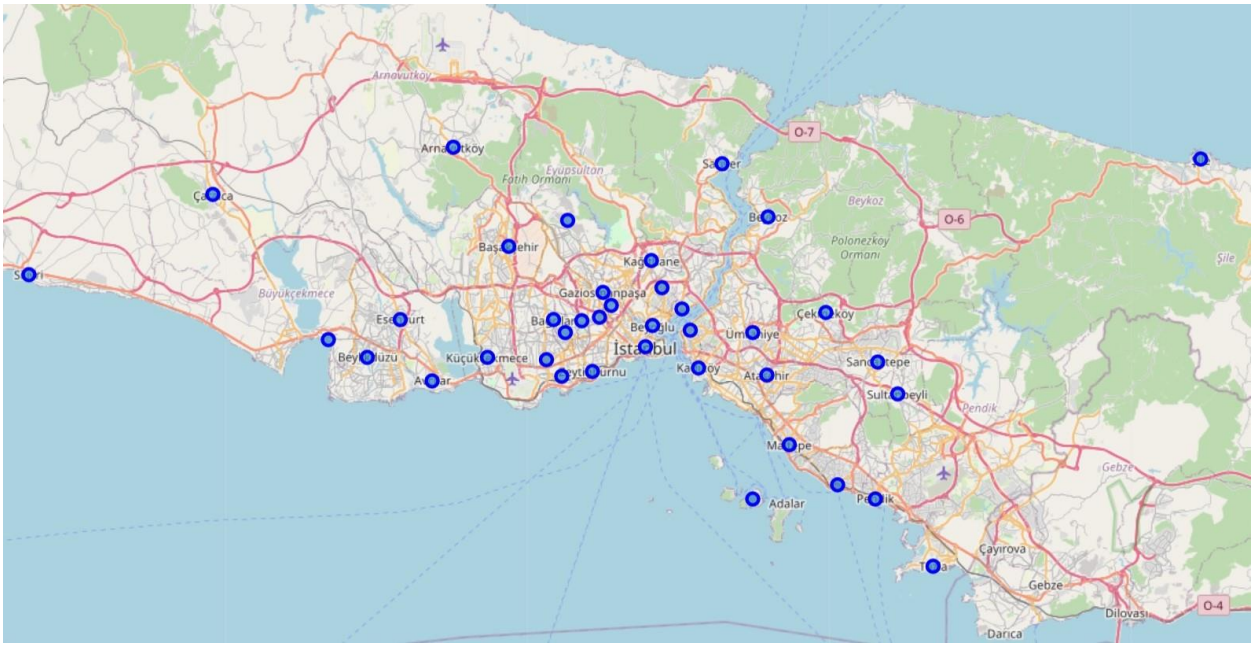
Then we will combine all these data sources and we will provide the distribution and density of the restaurant categories. After that we will try to classify the districts into clusters in order to see if we are getting some patterns. Finally we will try to find out some interesting insights which might be useful to investors, tourists or to people with business interests. So, we will get a glimpse of the Restaurants in Istanbul and hopefully visualize the results on map.

3. Methodology

Firstly, I obtained dataset containing list of districts of Istanbul from the wiki page as mentioned above “list of districts of Istanbul” using pandas to pull necessary data through the webpage. It is a quick and convenient way to turn an HTML table into a pandas data frame. In result of this data 39 districts were returned.

Secondly, I obtained geographical coordinates of Istanbul using the geocoder package. I fetched geographical coordinates of districts as latitude and longitude.

Thirdly, I visualized the location of districts of Istanbul to have a general understanding of the location. I used python folium library to visualize geographic details of Istanbul and I created a map of Istanbul.

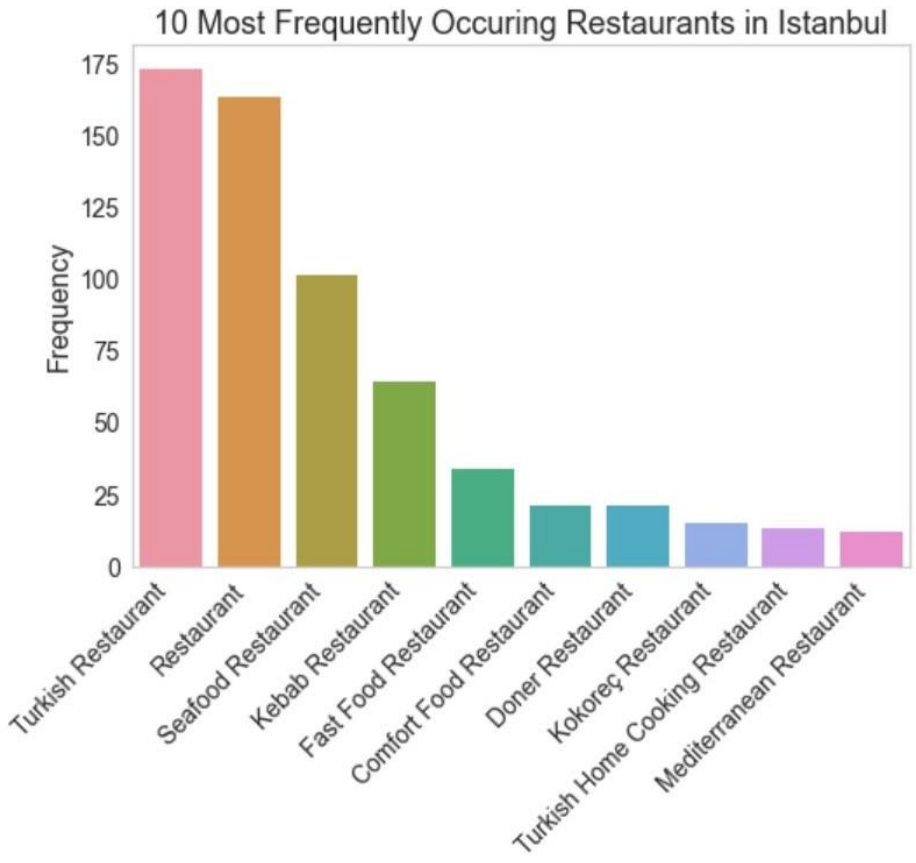


Furthermore, I defined Foursquare API credentials to make API calls in order to explore each neighborhood and returned the top 100 venues within radius of 2000 meters for each district. Now it's time to use the Foursquare API to extract the venues of each district in Istanbul. Here I created a merged table of districts, venues name, category, latitude and longitude information from Foursquare API.

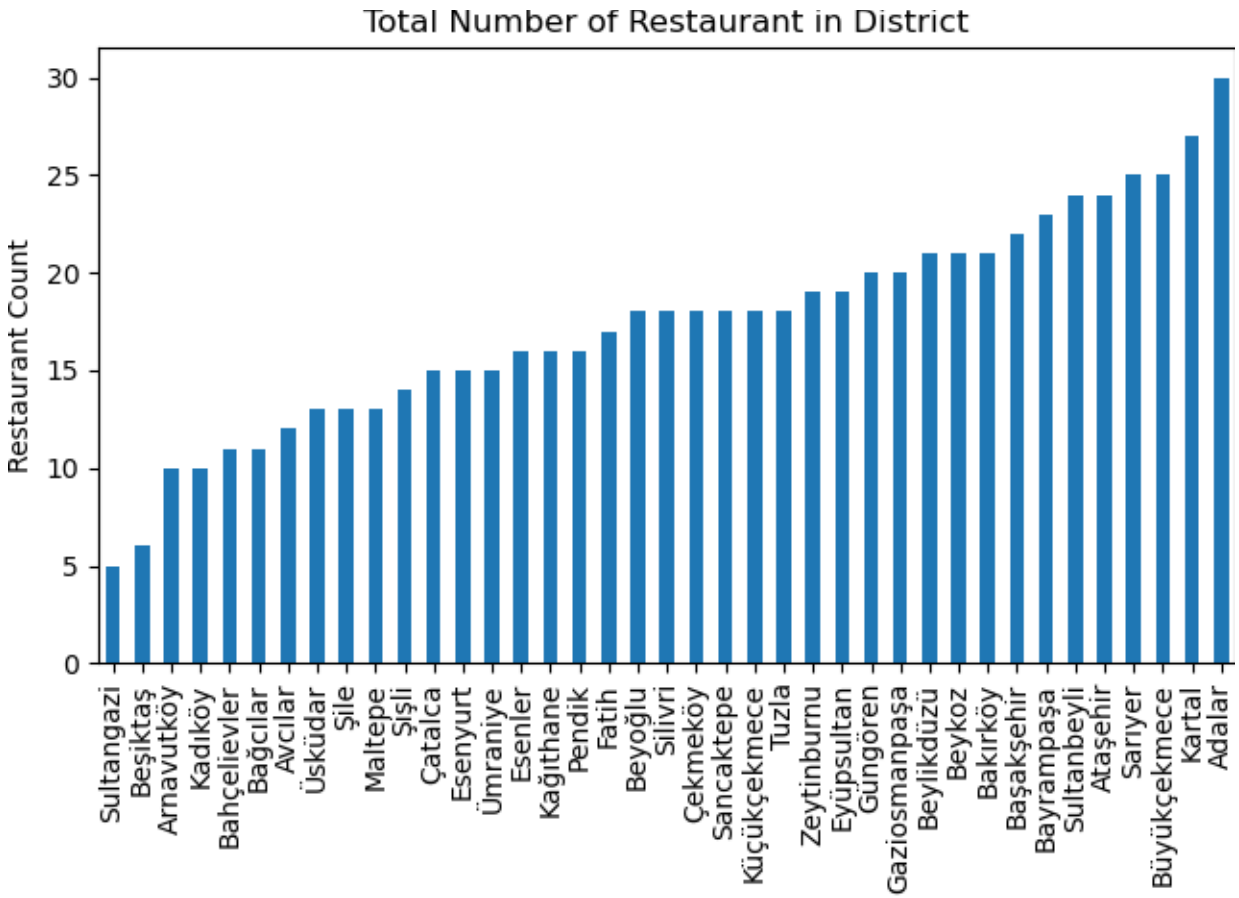
I performed some basic explanatory data analysis and derive some additional information from our raw venues data. We will initially filter out all the restaurant venues from this set. We are getting in total 3796 venues out of which 677 are listed as restaurant on Foursquare.

	District	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Adalar	40.876259	29.091027	İnönü Evi Müzesi	40.878251	29.093647	History Museum
1	Adalar	40.876259	29.091027	Merit Halki Palace Hotel	40.878802	29.090974	Hotel
2	Adalar	40.876259	29.091027	L'isola Guesthouse	40.877038	29.096136	Bed & Breakfast
3	Adalar	40.876259	29.091027	Luz Café	40.877528	29.097877	Café
4	Adalar	40.876259	29.091027	Heybeliada Şafak Askeri Gazino	40.873609	29.099478	Restaurant

According to data returned from Foursquare, restaurants are registered in 34 unique categories. So, I plotted a bar chart with the frequency of the 10 most frequently occurring restaurants in the whole city. The next figure visualizes the top ten restaurant categories in Istanbul:



I worked on the number of restaurants for each neighborhood and created a bar chart and analyze the big picture of it.

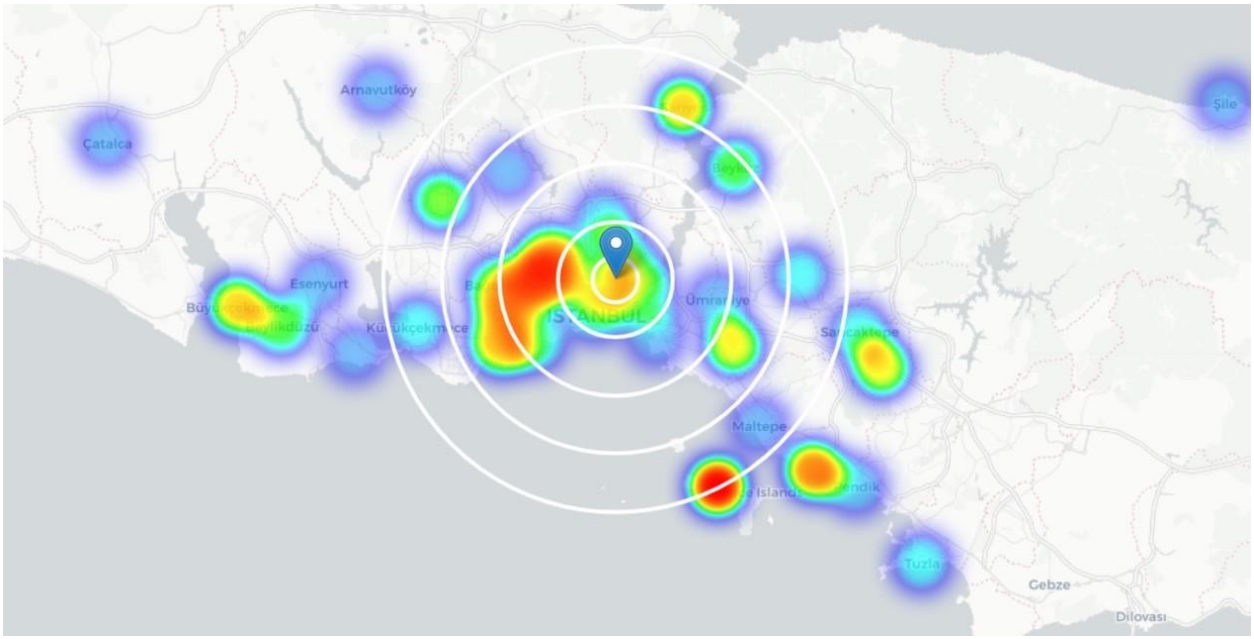


We can see that Adalar, Kartal, Büyükçekmece, Sarıyer, Ataşehir, Sultanbeyli, Bayrampaşa, Başakşehir, Bakırköy, Beykoz and Beylikdüzü boroughs are above 20 venues in our given coordinates.

The extracted venue categories will be encoded using one-hot encoding to handle any categorical values in the dataset. I applied one hot encoding on the obtained dataset and use it to find the 10 most common venue category in each district.

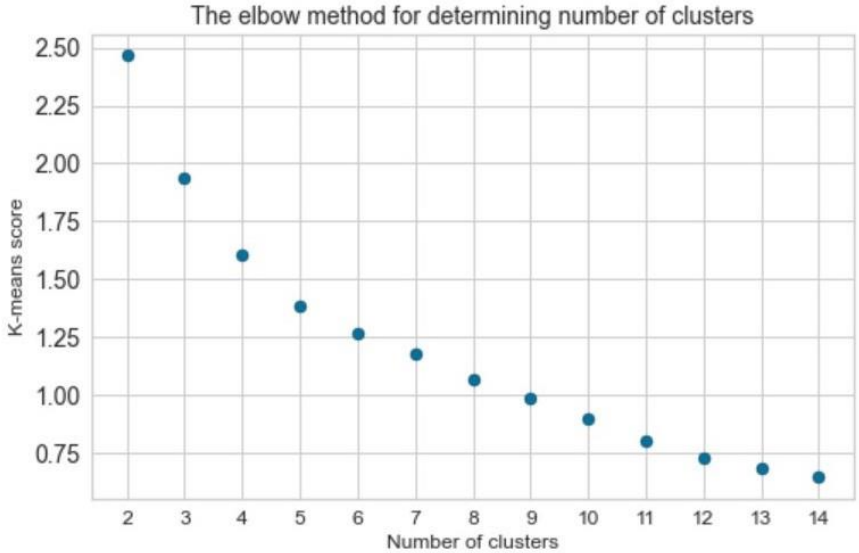
	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adalar	Seafood Restaurant	Restaurant	Turkish Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Argentinian Restaurant	Asian Restaurant	Bosnian Restaurant	Caucasian Restaurant	Arepa Restaurant
1	Arnavutköy	Turkish Restaurant	Restaurant	Turkish Home Cooking Restaurant	Fast Food Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Bosnian Restaurant	Caucasian Restaurant
2	Ataşehir	Restaurant	Kebab Restaurant	Turkish Restaurant	Seafood Restaurant	Comfort Food Restaurant	Turkish Home Cooking Restaurant	Italian Restaurant	Fast Food Restaurant	Doner Restaurant	Vegetarian / Vegan Restaurant
3	Avcılar	Restaurant	Comfort Food Restaurant	Turkish Restaurant	Theme Restaurant	Kebab Restaurant	Doner Restaurant	Seafood Restaurant	Fast Food Restaurant	Chinese Restaurant	Eastern European Restaurant
4	Bahçelievler	Turkish Restaurant	Italian Restaurant	Seafood Restaurant	Comfort Food Restaurant	Turkish Home Cooking Restaurant	Restaurant	Fast Food Restaurant	Kokoreç Restaurant	Vegetarian / Vegan Restaurant	Eastern European Restaurant

I created a heatmap showing the density of restaurants and try to extract some meaningful info from that. Also, we can add a few circles indicating distance of 2km, 5km, 10km, 15km and 20km from city center.

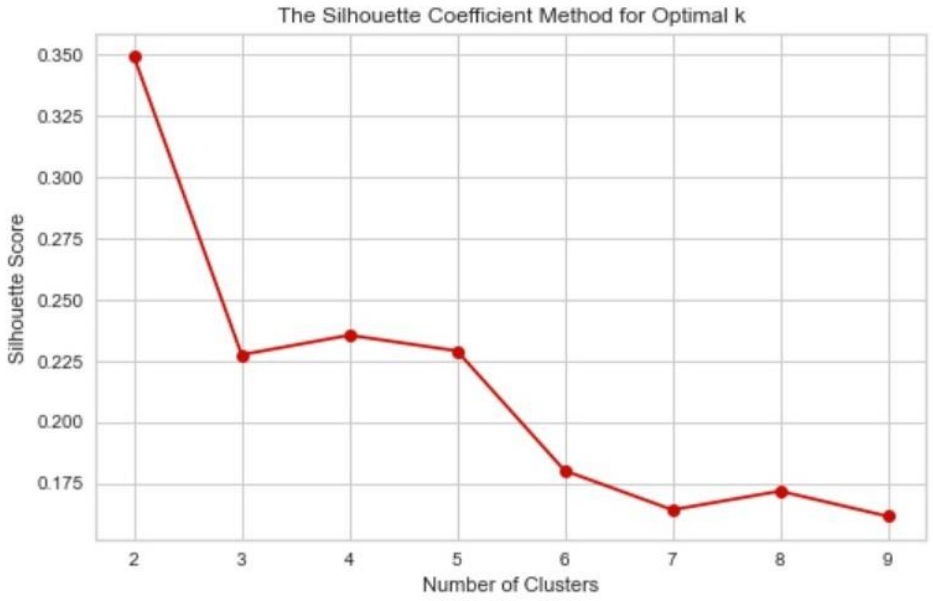


K-Means clustering

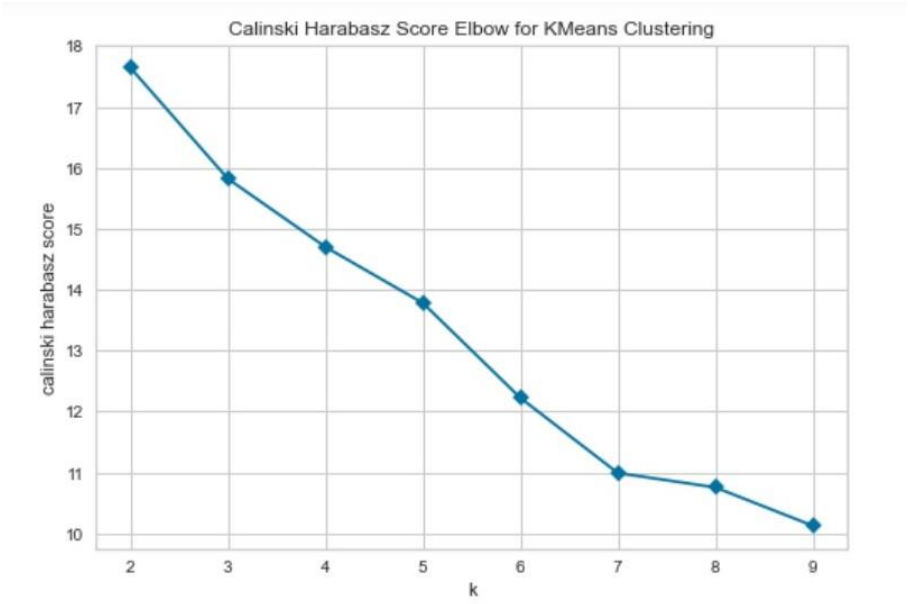
Since we have common venues in districts, I could run an unsupervised machine learning algorithm, known as k-means clustering algorithm, which groups the unlabeled dataset into different clusters. To determine the optimal number of clusters, first I used the "Elbow" method.



The results are not good enough to decide between k=3 and k=5. So, we can use Silhouette coefficients or another else methods for better vision. Silhouette score is used to evaluate the quality of clusters created using clustering algorithms such as K-Means in terms of how well samples are clustered with other samples.



KElbowVisualizer visualizes the clusters according to a scoring function, looking for an "elbow" in the curve. In the following code, I used the *calinski_harabasz score*.



By contrast, Silhouette and Calinski-Harabasz Index method provide a better insight under this context and it works better. We can inspect that **the best k** value for this task is **3**. Hence, we will have 3 cluster neighborhoods at the end. 39 districts of Istanbul were classified into 3 clusters.

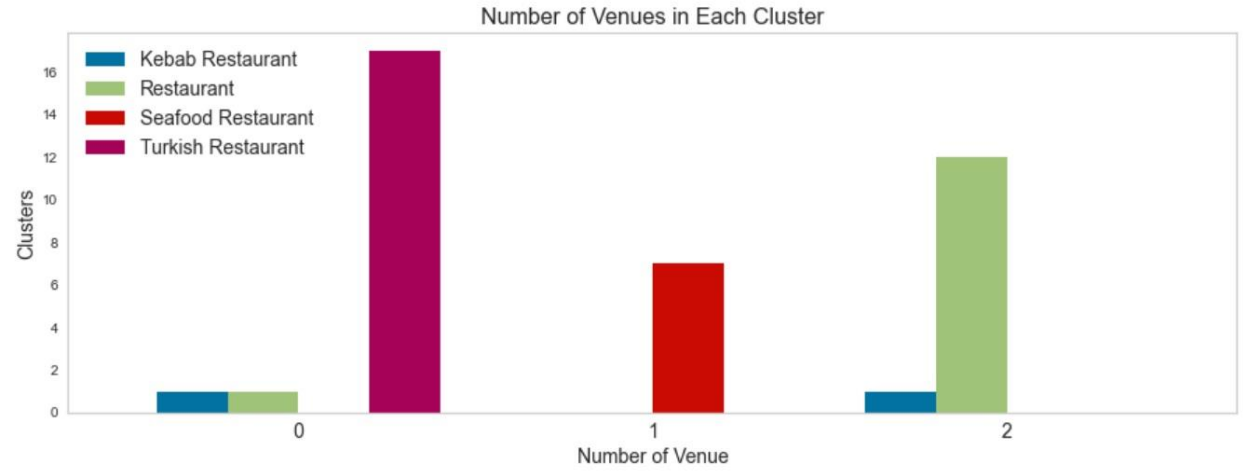
I created a new merged data table that includes the cluster as well as the top 10 venues for each district.

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adalar	1	Seafood Restaurant	Restaurant	Turkish Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Argentinian Restaurant	Asian Restaurant	Bosnian Restaurant	Caucasian Restaurant	Arepa Restaurant
1	Arnavutköy	0	Turkish Restaurant	Restaurant	Turkish Home Cooking Restaurant	Fast Food Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Bosnian Restaurant	Caucasian Restaurant
2	Ataşehir	2	Restaurant	Kebab Restaurant	Turkish Restaurant	Seafood Restaurant	Comfort Food Restaurant	Turkish Home Cooking Restaurant	Italian Restaurant	Fast Food Restaurant	Doner Restaurant	Vegetarian / Vegan Restaurant
3	Avcılar	2	Restaurant	Comfort Food Restaurant	Turkish Restaurant	Theme Restaurant	Kebab Restaurant	Doner Restaurant	Seafood Restaurant	Fast Food Restaurant	Chinese Restaurant	Eastern European Restaurant
4	Bağcılar	0	Kebab Restaurant	Turkish Restaurant	Fast Food Restaurant	Mediterranean Restaurant	French Restaurant	Falafel Restaurant	Eastern European Restaurant	Dumpling Restaurant	Doner Restaurant	Vegetarian / Vegan Restaurant

We can also anticipate the number of "1st Most Common Venue" in each cluster.

1st Most Common Venue	Kebab Restaurant	Restaurant	Seafood Restaurant	Turkish Restaurant
0	1	1	0	17
1	0	0	7	0
2	1	12	0	0

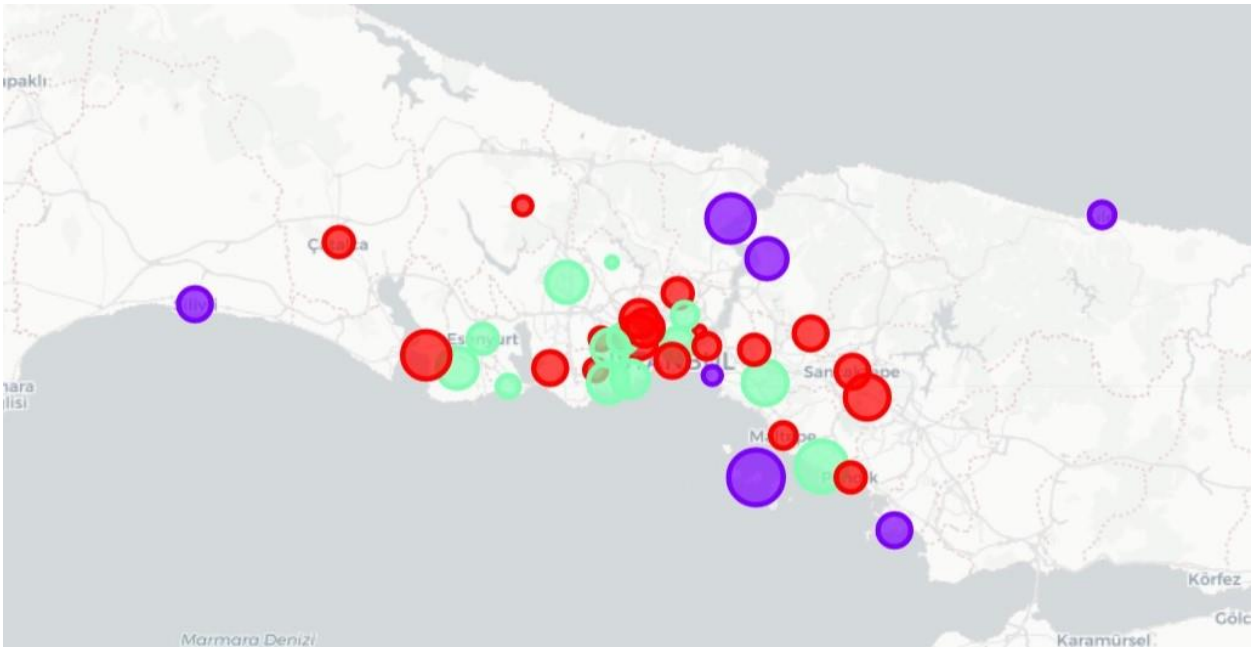
In this way, we can create a bar chart which may be helpful to us to find proper labels for each cluster.



As you can see from above bar chart, we can label each cluster as follows:

- Cluster 0 - Turkish Restaurant cluster
- Cluster 1 - Seafood Restaurant cluster
- Cluster 2 - Mixed or Generic Restaurant cluster

We can now use the cluster labels to show the districts marked with a cluster-specific color on a map by using folium. You will see bubbles for the city districts, with three different colors for the three different clusters. Finally, let's visualize the resulting clusters:



Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, we can then assign a name to each cluster. Let’s analyze these clusters in detail:

Cluster #1 - RED:

Most of the districts labeled as cluster 1 (0 in code and table) have Turkish Restaurants as the most common restaurant type. So we can assume that this cluster represents Turkish Restaurants.

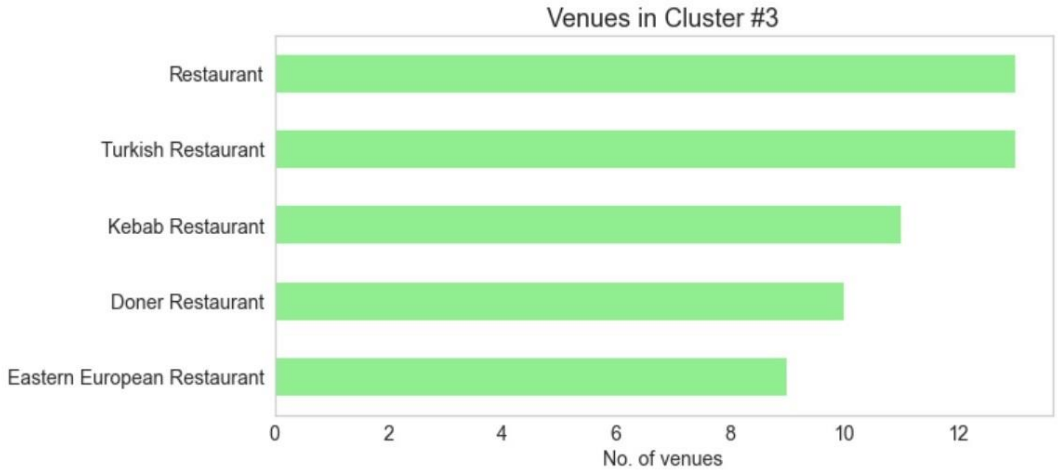
Cluster #2 - PURPLE:

All of the districts labeled as cluster 2 (1 in code and table) have Seafood Restaurants as the most common restaurant type. So we can assume that this cluster represents Seafood Restaurants.

Cluster #3 - GREEN:

At first glance, in 1st Most Common Venue, Restaurants seem to be ahead. On the other hand, when considering all the most common venues (from 1 to 10) there are Restaurants, Turkish Restaurants, Kebab, and Doner restaurants. Although “Restaurant” here is the most common venue, this cluster seems to be a *generic* one. It represents a *mixed* or *other* cluster.

Let's visualize the top 3 most common venue categories in each of the cluster 3.



4. Results and Discussion

We have got a glimpse of the Restaurants in Istanbul. Now in this case we are ready to find out some interesting insights which might be useful to stakeholders, investors or to people with business interests.

We used Foursquare API and we searched for all restaurants locations for a maximum of 100 venues in a radius of 2000 meters from each neighborhood. Venues dataset is limited to those available in Foursquare. We obtain the following results:

- We found 677 different types of restaurants registered in 34 unique categories on Foursquare.
- In Foursquare, 173 Turkish Restaurants and 163 Restaurants were recorded in separate categories.
- Turkish restaurants are the most common venues in all districts
- Seafood restaurants are among the top three.

A significant number of restaurants on Foursquare have been labeled as "Restaurant" rather than a specific restaurant category. These may be similar or different in themselves. Most of the foreigners who come to Istanbul think that restaurants serve only Kebap or Doner. Contrary to popular belief, the number of other types of restaurants is significantly high.

We have used the k-means algorithm to cluster the districts in Istanbul, resulting in 3 clusters with the information of most common of restaurant venues in each district. As a result, we can examine venues listed inside each cluster and define the discriminating venue categories that distinguish them. The clusters are:

Cluster 1: Cluster 1 has Turkish Restaurants as the most common restaurant type. So we can assume that this cluster represents Turkish Restaurants. The neighborhoods in cluster 1 has the greatest number of Turkish Restaurants, hence opening one here is not the best choice.

Cluster 2: All of the districts labeled as cluster 2 have Seafood Restaurants as the most common restaurant type. So, we can assume that this cluster represents Seafood Restaurants. As seen from the clustering calculations and chart given above, the neighborhoods 2 has notable number of seafood restaurants. Therefore, opening a new one here is not the best option.

Cluster 3: It is not an easy task to interpret the 3rd cluster. At first glance, in 1st Most Common Venue, Restaurants seem to be ahead. On the other hand, when considering all the most common venues (from 1 to 10) there are Restaurants, Turkish Restaurants, Kebab, and Doner restaurants. It is seen that their numbers are approximately close to each other. You can observe that there are restaurants representing the main cuisines in this part of Istanbul. So with a simple approach, we can say that this cluster seems to be *generic* one that represents "mixed or others". Clearly, this cluster deserves a more detailed analysis.

The results at the end of the analysis did not surprise us. As seen on the map the best seafood restaurants are located in three districts of Istanbul as Adalar, Sariyer and Beykoz. The clustering is completely based on the most common venues obtained from Foursquare API data. The real condition could be different as there might be place that have not yet included in Foursquare. Moreover, these results also could potentially vary if we use some other techniques. The results will allow us to identify which neighborhoods have higher concentration of restaurants while which have fewer.

5. Conclusion

Based on the results described above, we can conclude that: Purpose of this project was to identify districts which fits best to the diverse requirements of the customer. For finding the perfect location, now we must go deeper and analyze the top 10 to 15 districts more detailed. We could compare specific neighborhoods and add more detailed data like... Neighboring businesses can affect the profitability both positively and negatively.

Finding the best location to start a business can be challenging and quite frustrating due to many uncertainties. However, we can quickly gain meaningful insights into the city and its neighborhoods with data available today. This helps everyone, including entrepreneurs, business owners, and stakeholders, to make solid decisions based on facts.

I believe that we have gained new insights about questions mentioned in the introduction and business problem section. We also have found some answers to some of them. We observed that the restaurants in Istanbul can be gathered in 3 separate groups: Turkish Restaurants, Seafood and Others or Mixtures. In addition districts near to the sea, have the most seafood restaurants as expected.

We have found that the numbers of some restaurants are very low, such as French, Vegan, or Fusion. This could be a potential opportunity for investors after doing a detailed analysis.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like demographic structure, crime index, proximity to hotels or parks, proximity to main street, traffic, real estate prices or rents, social and economic dynamics of every neighborhood.