THE BATTLE OF NEIGHBORHOODS: ISTANBUL CASE

GANI YAMAN

TABLE OF CONTENTS

1.Introduction: Business Problem

Istanbul is the most populated city in Turkey with an increasing migration constantly. In order to meet the needs of the growing population, new amnesties established continuously, thus, arises new opportunities for entrepreneurs. The aim of this project is the get insight for business purposes to find the best location to start a business such as restaurants, cafe, arcade centers, and to get the best possible profitability out of the investment.

The tool that will be utilized in this project is the Foursquare API. The geographical data acquired from Foursquare API will be used to access the venues information and so to compare the neighborhoods by applying clustering algorithms and finally to choose best option among boroughs.

2. Description of the Data

Istanbul has 39 districts in total. The coordinates of the boroughs of Istanbul attained from the Second-level Administrative Divisions of the Turkey from Spatial Data Repository of NYU. Also, the venues information of the boroughs is attained from Foursquare API. Furthermore, the center coordinates of each Borough are determined by using Google Map, 'Search Nearby' option.

I will be using K-Means clustering model to analyze different clusters of boroughs and evaluate the venue information to establish a model which will determine the best location to start a business in Istanbul.

Latitude and longtitude values attained by using geopy library and by using Latitude and Longtitude values, the map of Istanbul is created by Folium library.
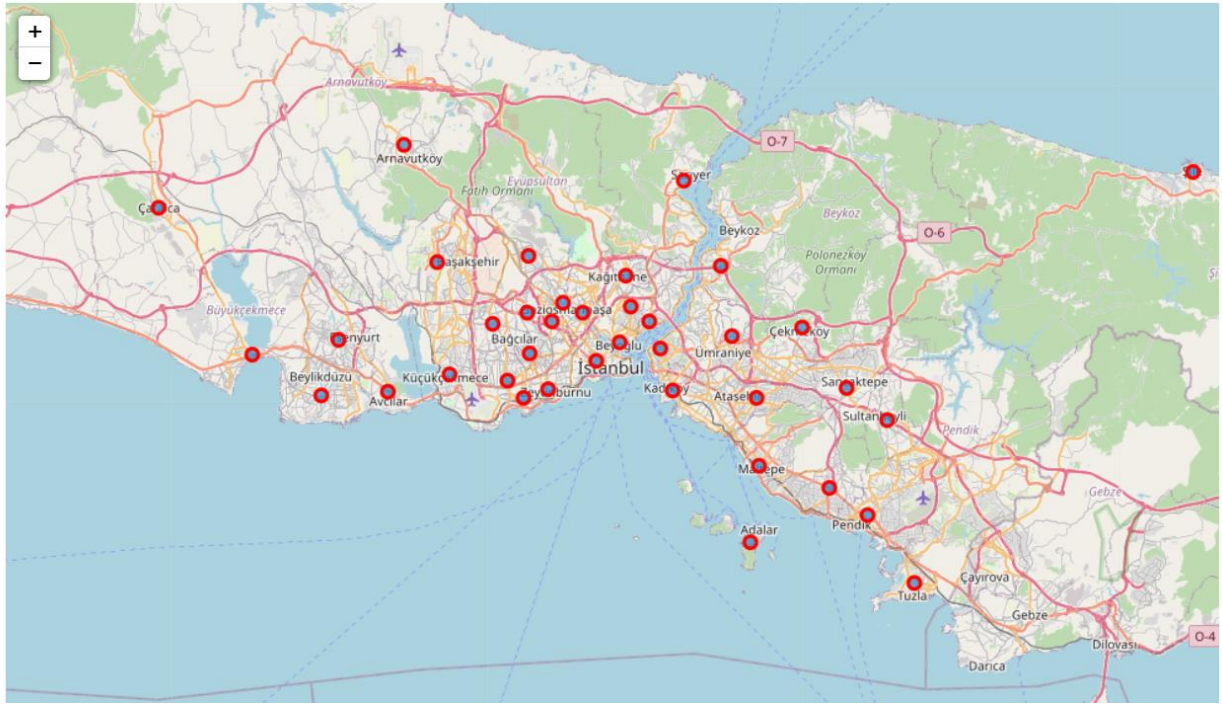
We can explore the boroughs of Istanbul by using Foursquare API

	Borough	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adalar	40.8619	29.1208	Büyükada Tepesi	40.861107	29.117418	Mountain
1	Adalar	40.8619	29.1208	Eski Rum Yetimhanesi	40.861705	29.123323	Historic Site
2	Adalar	40.8619	29.1208	Aşıklar butik by şükrü	40.862570	29.118003	Hotel
3	Adalar	40.8619	29.1208	Büyükada Bisiklet Parkuru	40.865000	29.116861	Bike Trail
4	Adalar	40.8619	29.1208	Nizam Butik Otel & Bistro	40.863322	29.116257	Bed & Breakfast

1812 venues were found, and we can count venues per borough

|	Borough	Count
0	Arnavutkoy	9
1	BuyukCekmece	12
2	Pendik	13
3	Adalar	14
4	Tuzla	15

3. Methodology

Now, we have the neighborhoods data of Istanbul (i.e total 39 neighborhoods). We also have the most popular venues in each neighborhood obtained using Foursquare API. A total of 1812 venues have been obtained in the whole city and 254 UNIQUE CATEGORIES. The least venue number is 9 which is in Arnavutkoy.

We can perform one hot encoding on the obtained data set and use it find the 9 most common venue category in each neighborhood. Then clustering can be performed on the dataset. K - Nearest Neighbor clustering method have been used and to find the optimal number of clusters silhouette score metric technique is used. The clusters obtained can be analyzed to find the major type of venue categories in each cluster. The outcome then can be used to suggest business people, potential locations based on the category.

4. Analysis of the Data
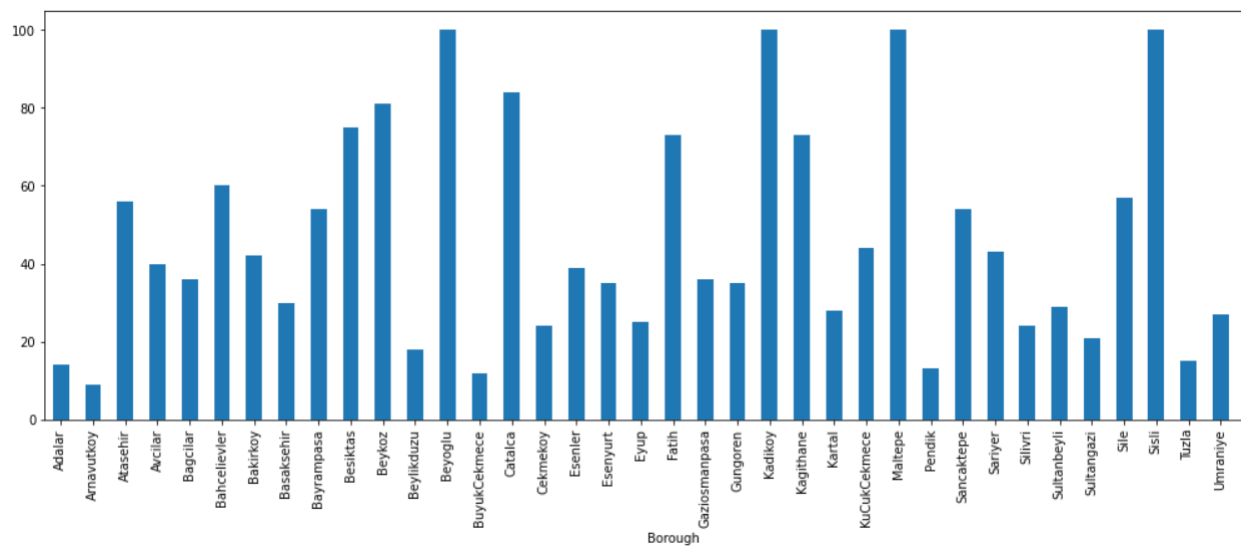
Some boroughs have not sufficient venues while others such as Beyoglu, Kadıkoy, Sisli have significantly higher number of venues. So, we have to cluster neighborhoods here by using k-Nearest Neighborhoods clustering technique. I will use the silhouette score to obtain the best value for the number of clusters.
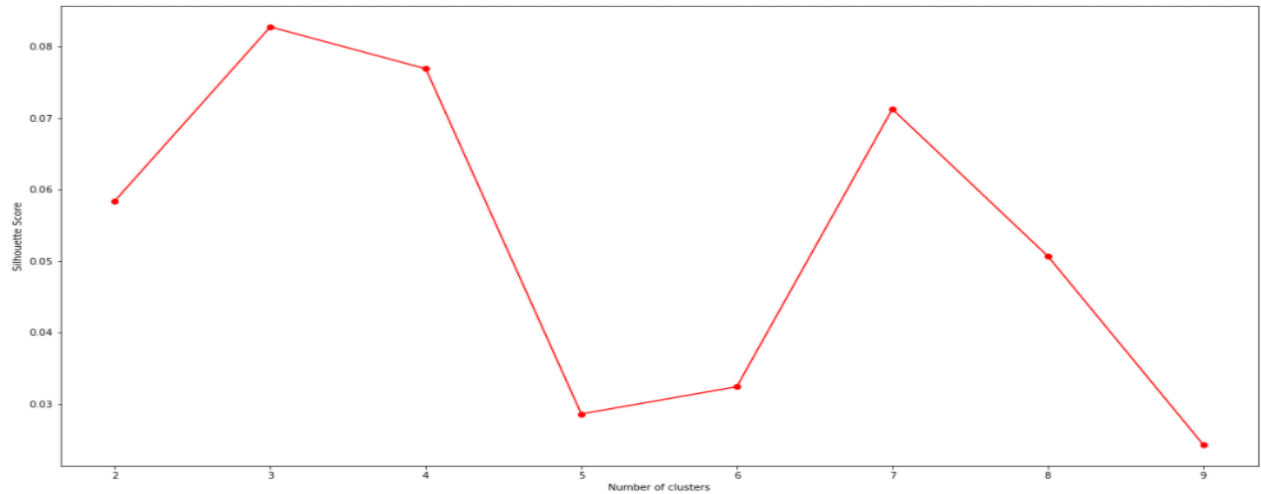
As seen from the above line plot, the best number of clusters having the highest silhouette score is 3.
So, we could consider the number of clusters as 3.

	Borough	Latitude	Longitude	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
0	Adalar	40.8619	29.1208	1	Forest	Café	Garden Center	Nature Preserve	Historic Site	Hotel	Mountain	Tea Room	Bike Trail
1	Arnavutkoy	41.1956	28.7352	0	Shopping Mall	Big Box Store	Diner	Construction & Landscaping	Arcade	Convenience Store	Restaurant	Kofte Place	Leather Goods Store
2	Atasehir	40.9831	29.1279	2	Café	Coffee Shop	Pool	Spa	Soccer Stadium	Soccer Field	Breakfast Spot	Gym / Fitness Center	Restaurant
3	Avcilar	40.9880	28.7170	1	Café	Hotel	Turkish Restaurant	Cosmetics Shop	Pizza Place	Perfume Shop	Fast Food Restaurant	Moving Target	Modern European Restaurant
4	Bagcilar	41.0450	28.8338	2	Café	Turkish Restaurant	Kebab Restaurant	Gym	Steakhouse	Supermarket	Dessert Shop	Food Court	Department Store
5	Bahcelievler	40.9976	28.8500	2	Turkish Restaurant	Café	Kebab Restaurant	Bakery	Turkish Home Cooking Restaurant	Steakhouse	Gym	Men's Store	Motorcycle Shop
6	Bakirkoy	40.9835	28.8679	2	Gym	Turkish Restaurant	Park	Cosmetics Shop	Bookstore	Spa	Athletics & Sports	Comfort Food Restaurant	Food Court
7	Basaksehir	41.0969	28.7718	1	Café	Shopping Mall	Comfort Food Restaurant	Restaurant	Dessert Shop	Bakery	Halal Restaurant	Pool	Beer Garden
8	Bayrampasa	41.0477	28.8992	2	Clothing Store	Sporting Goods Shop	Restaurant	Café	Shopping Mall	Toy / Game Store	Cosmetics Shop	Department Store	Gym
9	Besiktas	41.0469	29.0078	2	Coffee Shop	Café	Pub	Hookah Bar	Chocolate Shop	Park	Turkish Restaurant	Performing Arts Venue	Music Venue

We can also visualize the density of neighborhood by cluster label in a chart below.
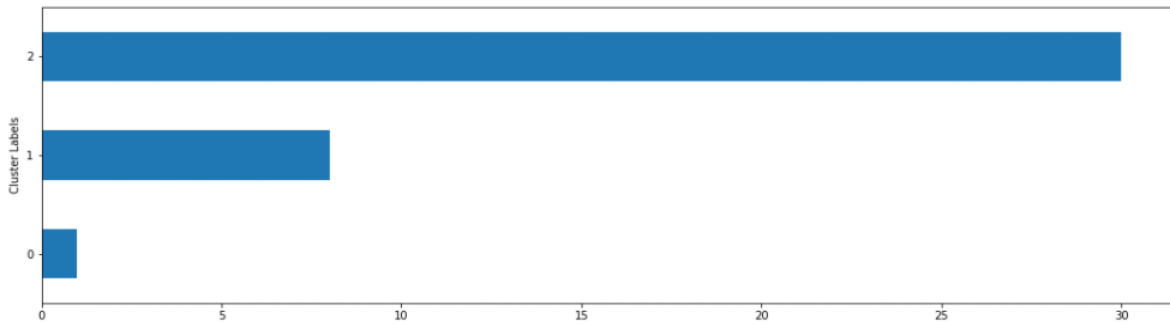
As we can see in above data frame Cluster2 contains the highest cluster density. We need to find the geographic centroid for this cluster. This is the most suitable location for a new business set-up.

With the help of reverse lookup, we can find the best suitable Neighborhood to locate a new business is in Fatih while we know the geographical coordinate of Fatih are (41.0252221, 28.9491824) which is located around Balat.

5. Results

We need to dig more to find out complimentary business set-ups in each cluster in order to suggest a most suitable cluster for new business set-up. So, let's examine the clusters and find the discriminating venue categories that distinguish each cluster. For this purpose, we should also print the nine most common venue category in each cluster.
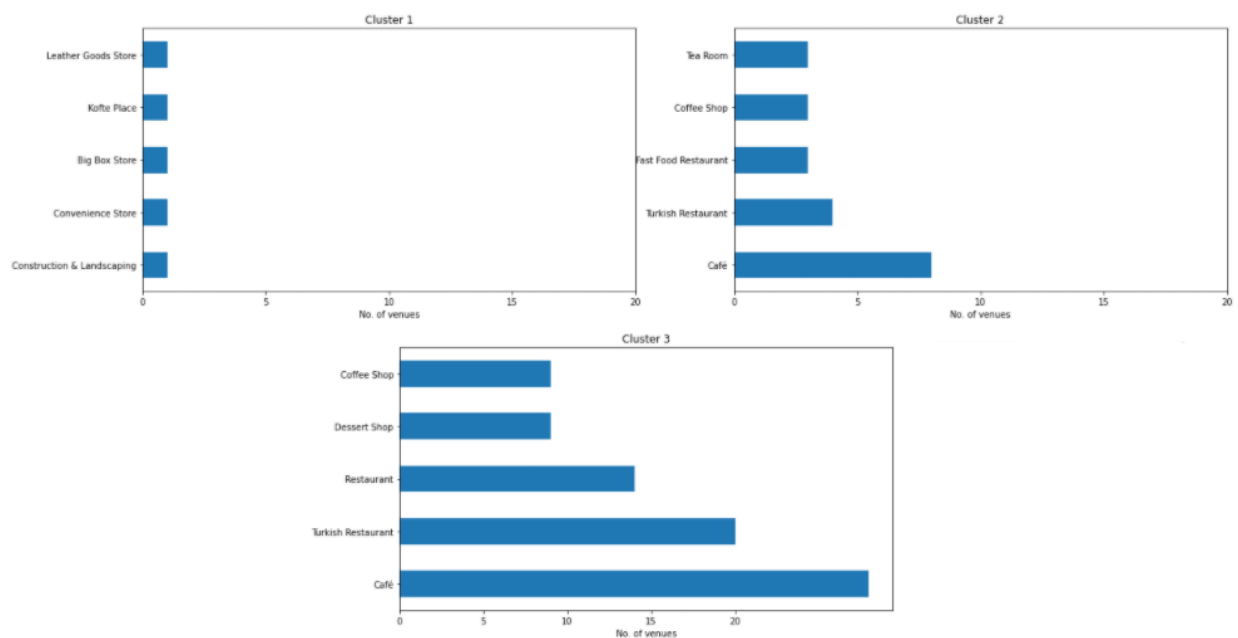
These above plots can be used to suggest valuable information to Business persons. Let's discuss a few examples considering they would like to start the following category of business.

6. Discussion

Let us discuss about appropriate business opportunity by Venue Category for certain targeted business set-ups

A. Cafe

Cluster 3 has the highest number of cafes. To start a successful business we should find a place which has least number of the amnesty. Thus, Cluster 1 which is Arnavutkoy could be the best option for opening a cafe. Also, Cluster 2 could be the other option.

B. Fast Food Restaurant

Cluster 3 seems to be best attraction for people as most of amnesties established in that region. This could be a good sign of locating a place to start-up a business. Also, our best place Fatih that we found before is also located in that cluster. However, there is not much Fast Food Restaurant in that region. So, it could be a good option to run a business in cluster 3.

C. Tea Room

Another popular amnesty is the room that is on the list in Cluster 2. However, we could not see much tea room in cluster 3. So, it could be good option for cluster 3.

D. Dessert Shop

We can see the high similarity between cluster 2 and cluster 3 in terms of amnesties. Therefore, we could suggest the Dessert Shop for cluster 2

Map of Istanbul with the clusters superimposed on top

This map can be used to find a suitable location to start a new business based on the Venue category. I have added a most suitable location on Istanbul map based on above analytics for a new business set-up. This map can also be used to select a vast suggestion area for a particular type of business based on the category.

7. Conclusion

The objective of this project was to analyze the boroughs of Istanbul and create a clustering model to suggest potential places to start a new business based on the category. The neighborhoods data was obtained from the Second-level Administrative Divisions of the Turkey from Spatial Data Repository of NYU and the Foursquare API was used to find the major venues in each neighborhood. Locations were used to create a clustering model. The best number of clusters i.e. 3 was obtained using the silhouette score. Each cluster was examined to find the most venue categories present, that defines the characteristics for that particular cluster.

A few examples for the applications that the clusters can be used for have also been discussed. A map showing the clusters have been provided. Both these can be used by stakeholders to decide the location for the particular type of business. A major drawback of this project was that the Foursquare API returned only few venues in each neighborhood. As a future improvement, better data sources can be used to obtain more venues in each neighborhood. This way the neighborhoods that were filtered out can be included in the clustering analysis to create a better decision model