**ZUHAL DURAN**

**REPORT OF THE FINAL PROJECT OF THE**

**“APPLIED DATA SCIENCE CAPSTONE” COURSE’S**

**PEER GRADED ASSIGNMENT**

**SUBJECT:**

**“WHAT” to OPEN “WHERE” in ISTANBUL**

**JUNE 10, 2020**

**Introductıon**

Istanbul, Turkey Situated in the city and one of the country's 81 provinces. It is the most populous, economical, historical and socio-culturally important city in the country. The city ranks first in Europe and sixth in the world in terms of economic size, taking into account the municipal boundaries in terms of population. Within its borders, there are 40 municipalities together with the metropolitan municipality. If you want to invest in such a large and developed city, how to proceed will be examined.

Demands and preferences of people change according to the districts of Istanbul. Because the lifestyle of people living in the district, income level, education level, etc. there are differences. So you can ask which county to open and what? These investments can be places such as cafe, hotel, gym. Therefore, the most preferred places and sectors will be grouped according to the districts of Istanbul.

**Data**

In this section, we will examine what data is used for clustering and where it is obtained from. There are differences between their districts, there will be different preferences regarding investment. The places preferred by people will be determined from foursquare according to the districts. District coordinates of Istanbul province were obtained from <http://dunya.arztalep.com/tr-TR/istanbul/istanbul-haritasi.aspx>. I used the geolocator library to find the latitudes and longitude of Istanbul. The data are given in table 1 below. Istanbul province has a total of 38 borough. The latitude and longitude values of these 38 borough and borough are given below. The georapical coordinate of İstanbul are 41.0766019, 29.052495.



Table 1 Latitude and Longitude in Istanbul

I used the folium library to indicate the distribution of the districts of Istanbul on the map. I used the folium library to indicate the distribution of the districts of Istanbul on the map. The map is located below.

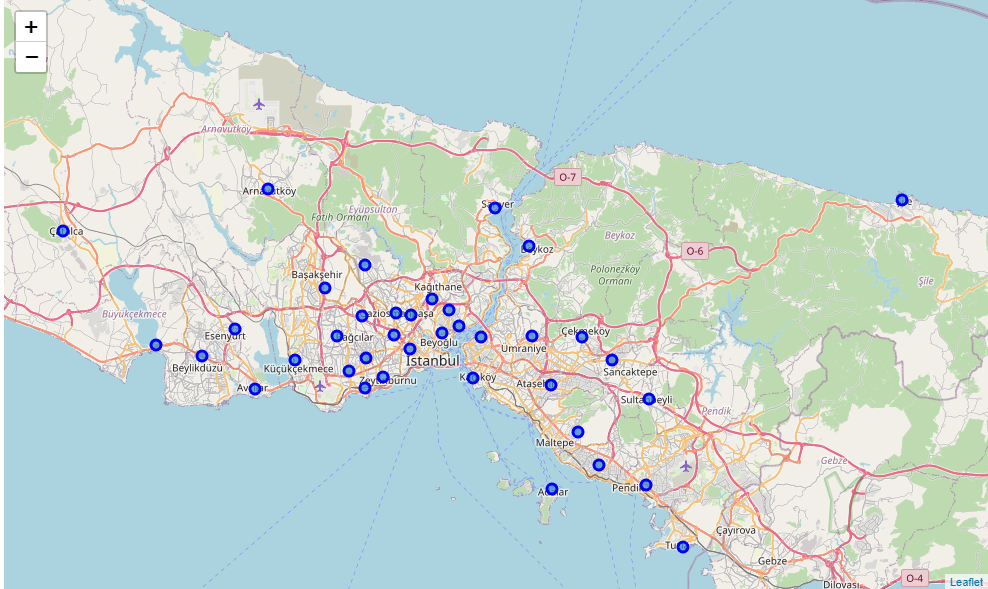
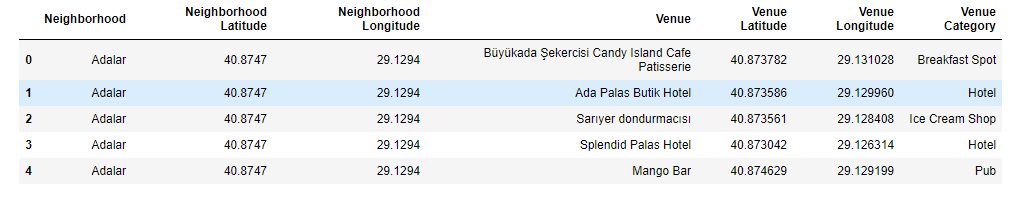


Figure 1 Borough of Istanbul on the map

After drawing the map, we will examine the preferred places for each district. The data was obtained using the Foursquare API. A total of 3891 venues were obtained. The data at the beginning and end of the created list are given in the figures below.



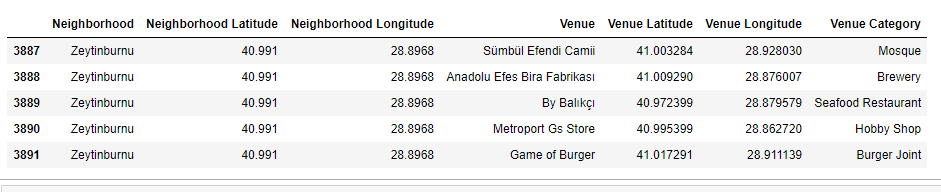


Table 2 First and Last Rows of the Data Frame

We will group according to the number of venues in the borough.

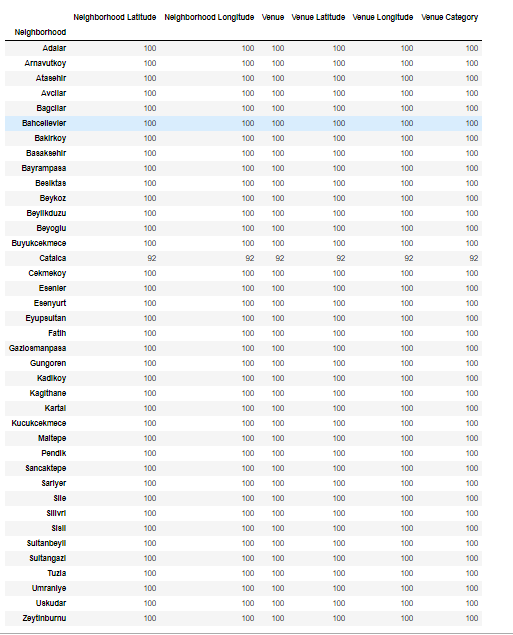


Table 3 Number of Venues at Boroughs

We need to categorize venues. So we will create a new database using the clustering algorithm.

**Methodology**

In this section, we will examine the venues we group by boroughs. We will create a new data frame and examine the categories of venue by boroughs. A total of 283 different venues have been created. The first 5 lines of these places are below.

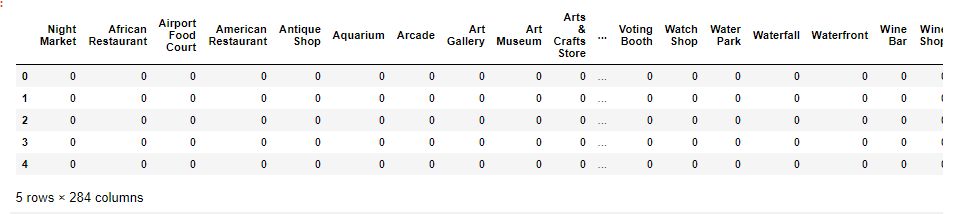


Table 4 Head of the istanbul\_onehot\_encode Data Frame

Its categories were grouped according to districts. Mean values were examined. The table below also shows the calculated average values.

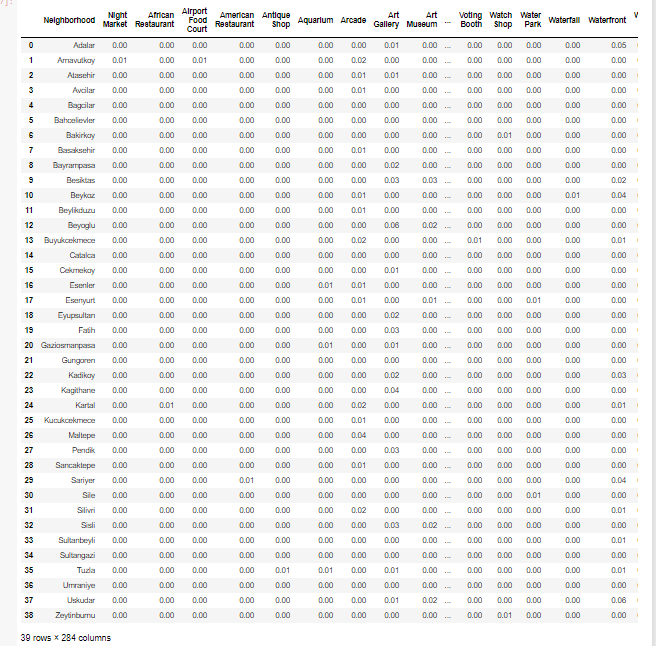


Table 5 Mean Venue Type of Each Borough

283 venue do not make much sense to us. So we will analyze the top 10 most common places. Below are the most common venue 10 values in the table.

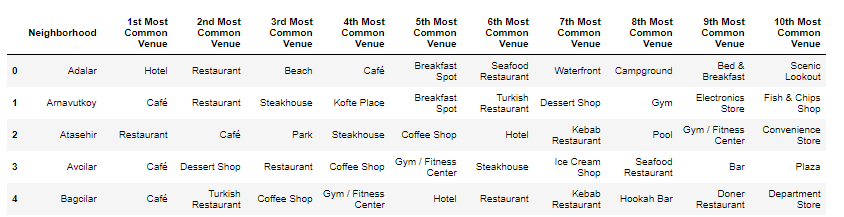


Table 6 Most Visited Venues at Boroughs

And now our data is ready for clustering algorithm. I have set 5 clusters. I use Kmeans algorithm from sklearn library.



Table 7 Head of The Clustered Data Frame

**Results and Discussion**

At this stage, we will examine the clusters we mentioned in the methodology. We will give advice based on the results. There are a total of 5 clusters and we will examine them individually.

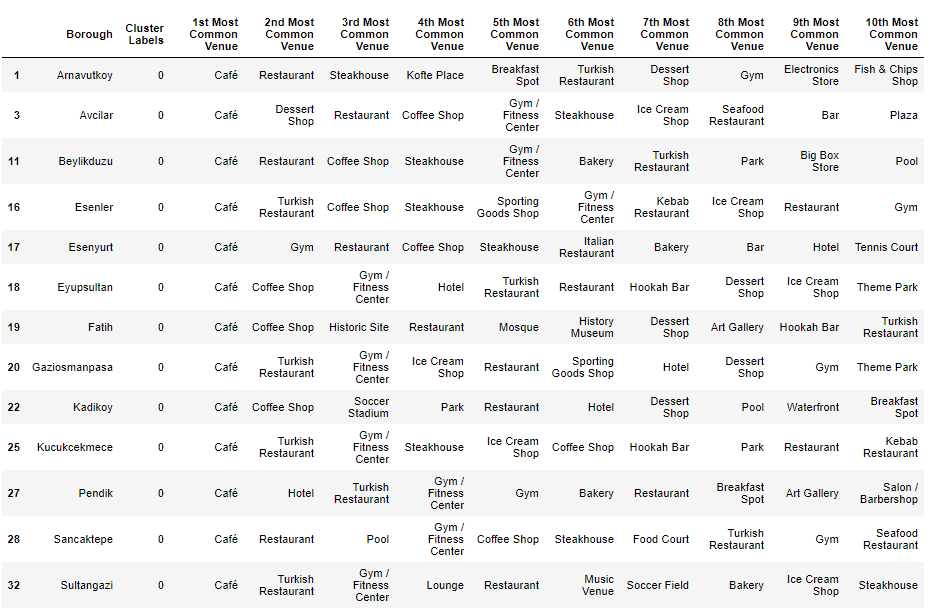


Table 8 Boroughs and Venues in the First Cluster

The first cluster is one of the most crowded clusters. There are 13 boroughs.

* If you want to invest in one of the boroughs in the first cluster, we see that the most preferred venue is cafe. There are different places in the second most common venue.
* If you want to open a restaurant, you can choose these boroughs such as:Arnavutkoy, Beylikduzu,Esenler,Gaziosmanpasa,Kucukcekmece,Sancaktepe and Sultangazi.
* If you want to open a gym, you can choose Esenyurt borough. The 2nd most common venue was obtained as a gym.
* In Pendik, the most preferred place after the cafe was obtained as a hotel.

Now, lets’s discuss about second cluster.



Table 9 Boroughs and Venues in the Second Cluster

In this cluster, the first choices differ as cafe and seafood restaurant. There are 6 districts in total. The seafood restaurant is preferred in Buyukcekmece and Sarıyer, and the cafe with Beykoz, Kartal, Silivri and Tuzla is preferred.



Table 10 Boroughs and Venues in the Third Cluster

There are only 4 boroughs in this cluster. While Beşiktaş and Beyoğlu are the first common venues hotels, Şişli and Üsküdar are also the second common venue hotels.

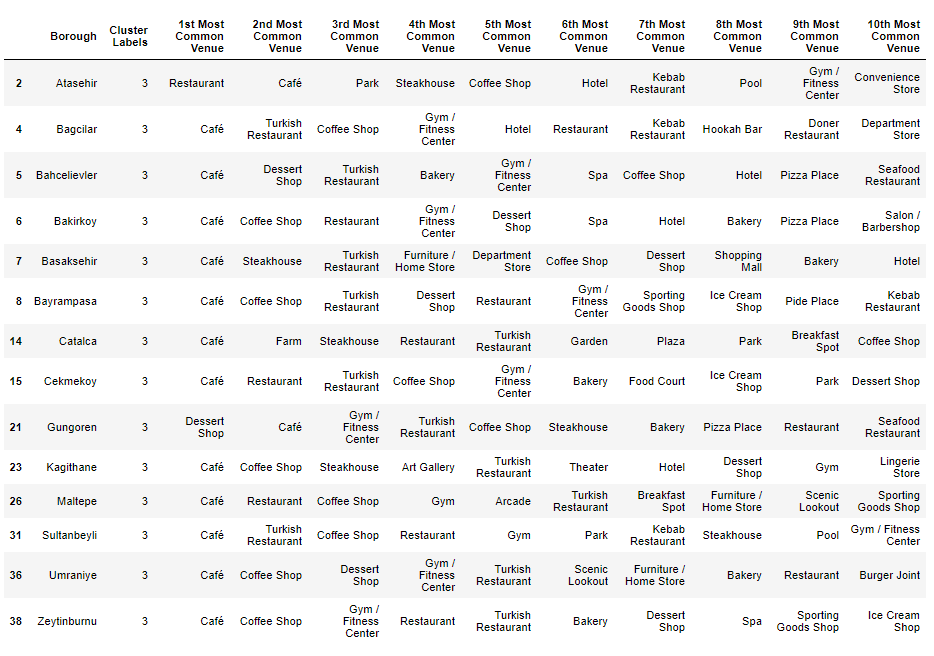


Table 11 Boroughs and Venues in the Fourth Cluster

The most crowded cluster among the clusters is the fourth cluster. The first choice was again the café in all boroughs except Gungoren and Atasehir. The first place to be preferred in Güngören was the dessert shop.

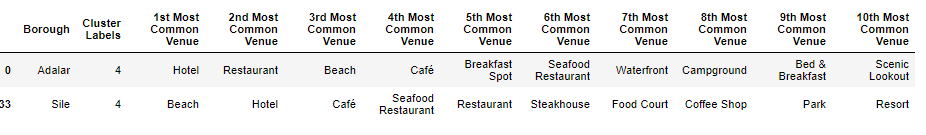


Table 12 Boroughs and Venues in the Fifth Cluster

There are only 2 boroughs in the last cluster. While the first choice in Adalar borough was a hotel, it was a beach in Şile.

**Conclusion**

At this Project we have analyzed the borougs of Istanbul according t oto the number and type of the venues in each borough. We started this project with findind the latitude and longitude values of each boroughs. Then we have used Foursquare API to find the venues at each borough. And we hav divided our data to the 5 clusters. We can say the conclusion below.

* When we look at the whole of Istanbul, we see that the most preferred venue is the cafe. We see that people in Istanbul mostly visit food and hotel venues.If you want to open a cafe, you can choose many borough of Istanbul. The first choice of 29 borough was the cafe.
* If you want to open a seafood restaurant, you can open it in one of the borough in the 2nd cluster.
* If you want to open a hotel, you can choose one of the borough in the 3rd cluster.
* Investments can be made in the dessert shop in Güngören borough.
* The 2 most preferred venues in Çatalca were farms. People who come to visit here prefer to visit nature.
* Unlike other borough, the park is the most visited venue in Üsküdar.

That is all fort his Project.Thank you for your read and for your feedbacks.