IBM Data Science Professional Certificate Capstone Project

The Battle of Neighborhoods

Introduction: Business Problem

With a population of approximately 15 million people in the metropolitan area, Istanbul is one of the world's fourteenth largest city and one of the most populous cities in the world.

Despite such a crowded city, vegetarian or vegan individuals may find it difficult to find vegetarian or vegan restaurants in Istanbul. Turkish food culture is predominantly meat and meat products. For this reason, vegetarianism and veganism have not yet played an important role in Turkish culture.

In this project, 38 districts of Istanbul and vegetarian and vegan restaurants were found within 1km (thought to be walking distance) from the center. At the same time, a clustering analysis was performed with k-means by using the populations of the districts of Istanbul and the information of the restaurants in these districts.

The questions we want to answer at the end of the project are:

- Which districts in Istanbul have vegetarian or vegan restaurants?
- What are the number of restaurants by district?
- What are the popular districts for someone who wants to open a new restaurant, taking into account district populations and restaurant numbers?

Data

Based on definition of our problem, factors that will influence our decisions are:

- number of existing restaurants in the borough (vegetarian or vegan)
- population of borough
- which boroughs have restaurants

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

The dataset which can be reached at https://geo.nyu.edu/catalog/stanford-nj696zj1674, contains all the cities and borough lists of Turkey and information on the location of these cities and towns.

Total population, female and male population percentages data of all these boroughs to the boroughs data. We can access the population data for 2018 boroughs at https://www.nufusu.com/ilceleri/istanbul-ilceleri-nufusu.

- Using **Geopy** web service, location information of Istanbul was obtained
- **Folium** and districts of Istanbul, restaurants in districts and clusters obtained as a result of k-means are shown on the map
- Population information was obtained from the website with **BeautifulSoup**
- Restaurant information around coordinates determined with **GoogleMaps Place API**

Methodology and Analysis

First, to find vegetarian and vegan restaurants in Istanbul and Turkey's "Second-level Administrative Divisions, Turkey, 2015 "dataset containing the geographical coordinates of all provinces and districts obtained from the address https://geo.nyu.edu/catalog/stanford-nj696zj1674 used. With this data set, latitude and longitude data of Istanbul and 38 districts were obtained.

```
ist_data.shape
print('Istanbul has 38 boroughs, so shape of ist_data is {}.'.format(ist_data.shape))
```

```
Istanbul has 38 boroughs, so shape of ist_data is (38, 3).
```

```
ist_data.head()
```

	Borough	Latitude	Longitude
0	Çatalca	41.302761	28.384248
1	Çekmeköy	41.079035	29.260164
2	Adalar	40.866249	29.058751
3	Ümraniye	41.039030	29.129102
4	Üsküdar	41.157230	28.571349

Coordinates of Istanbul were obtained by using Geopy. The districts of Istanbul were shown on the map using Folium.

A data set for 2018 containing the population in Istanbul and its districts, the population of men and women, and the percentages of men and women was obtained from https://www.nufusu.com/ilceleri/istanbul-ilceleri-nufusu. This dataset was taken from this address with BeautifulSoup and cleaned up to obtain population data.

```
population_data.head()
```

	Borough	Population	Male	Female
0	Esenyurt	891.120	%51,23	%48,77
1	Küçükçekmece	770.317	%50,00	%50,00
2	Bağcılar	734.369	%50,70	%49,30
3	Pendik	693.599	%50,50	%49,50
4	Ümraniye	690.193	%50,06	%49,94

Population data was combined with data obtained from districts and coordinates previously obtained.

```
new_data.shape
```

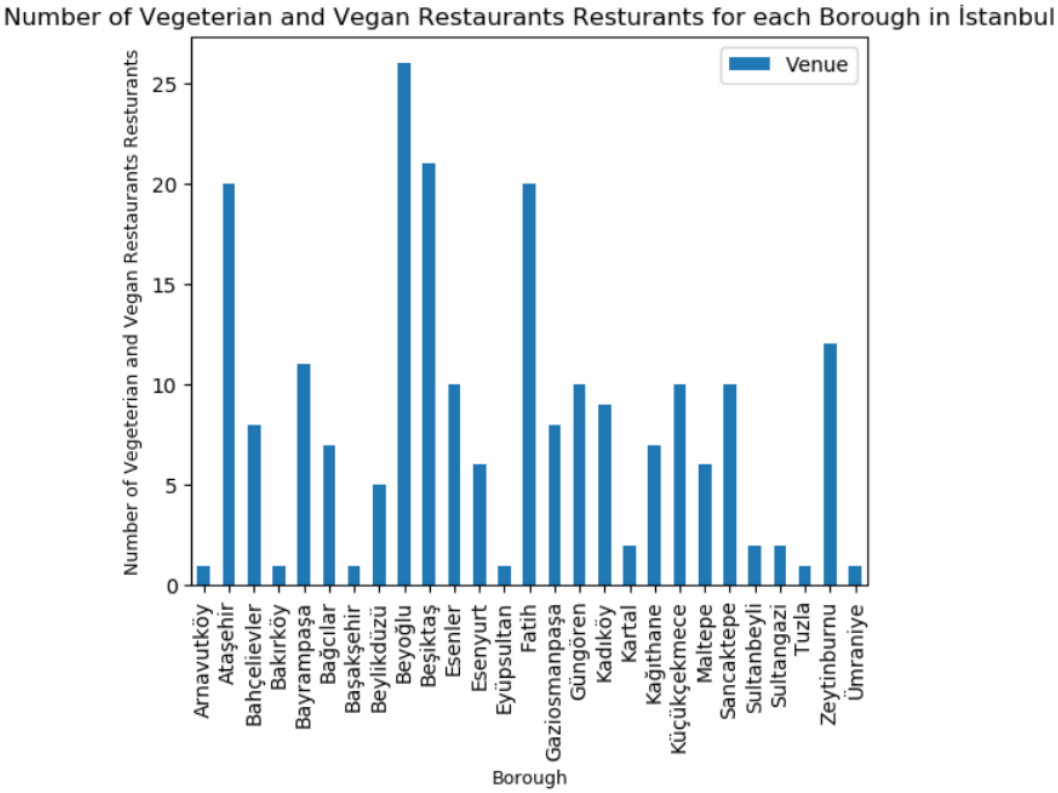
```
(38, 6)
```

```
new_data
```

	Borough	Latitude	Longitude	Population	Male	Female
0	Çatalca	41.302761	28.384248	72.966	%51,63	%48,37
1	Çekmeköy	41.079035	29.260164	251.937	%50,15	%49,85
2	Adalar	40.866249	29.058751	16.119	%53,37	%46,63
3	Ümraniye	41.039030	29.129102	690.193	%50,06	%49,94
4	Üsküdar	41.157230	28.571349	529.145	%48,73	%51,27
5	Arnavutköy	41.237003	28.691689	270.549	%51,99	%48,01
6	Ataşehir	41.000820	29.128503	416.318	%48,93	%51,07
7	Avcılar	41.032429	28.727993	435.625	%49,96	%50,04
8	Şişli	41.095581	28.988744	274.289	%48,67	%51,33
9	Şile	41.110001	29.606633	36.516	%51,50	%48,50
10	Büyükçekmece	41.050030	28.537505	247.736	%49,55	%50,45
11	Bağcılar	41.047117	28.840071	734.369	%50,70	%49,30
12	Bahçelievler	41.013857	28.837940	594.053	%50,19	%49,81
13	Bakırköy	40.982445	28.822385	222.668	%46,56	%53,44
14	Başakşehir	41.099737	28.760093	427.835	%50,41	%49,59

Using GoogleMaps Place API; for each district 1km (walking distance) from the district coordinates of the vegetarian (total 197) and vegan restaurants (total 34) located in areas separately obtained by combining information (total 218). They were marked on the map with Folium.

The numbers of restaurants found in each district were obtained. It was reached that 27 of 38 districts had vegetarian or vegan restaurants.



K-means clustering was applied to cluster regions according to their populations. To find the best number of k clusters; Silhouette Score was used before. The best result was obtained for cluster = 4.

Then k-means clustering was applied with k = 4 in the dataset. 4 clusters were obtained with the following characteristics;

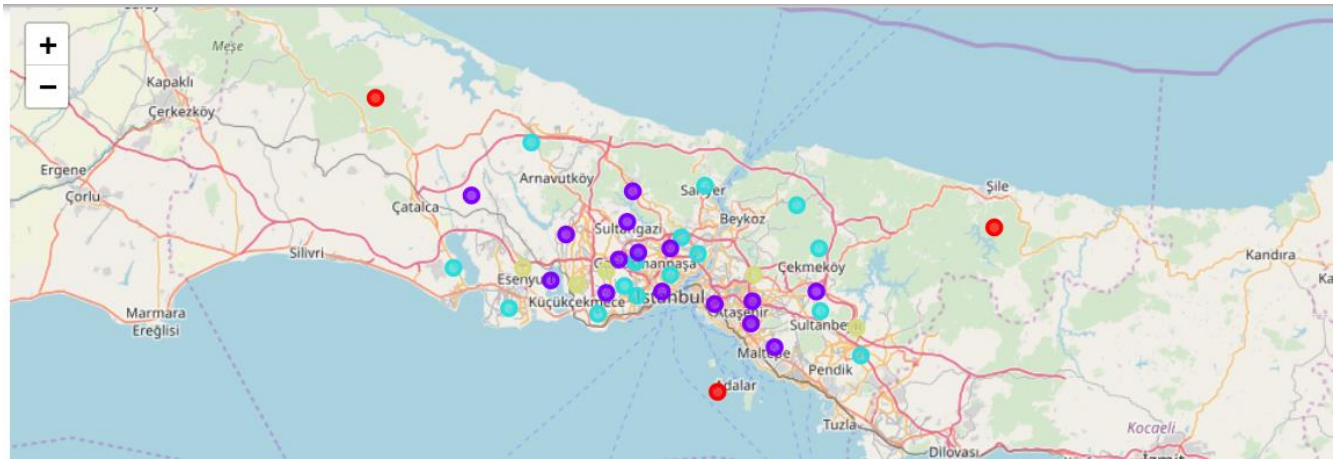
Cluster0: population between 16K and 72K, 30 venues

Cluster1: population between 181K and 342K, 48 venues

Cluster2: population between 383K and 594K, 121 venues

Cluster3: population between 690K and 891K, 18 venues

Created clusters were marked on the map using Folium. This map can be seen below.



Results and Discussion

By using the location information of the districts for Istanbul, we obtained vegetarian and vegan restaurants and locations of these restaurants within 1 km from each district center. Then, we created clusters with k-means by using the population information of Istanbul districts and the number of vegetarian and vegan restaurants in those regions. As a result of the experiments, we observed that 4 clusters were formed:

Cluster0: population between 16K and 72K, 30 venues
Cluster1: population between 181K and 342K, 48 venues
Cluster2: population between 383K and 594K, 121 venues
Cluster3: population between 690K and 891K, 18 venue

Starting from here;

Cluster3 includes the 5 most populated regions in Istanbul. However, there are only 18 vegetarian or vegan restaurants.

Cluster0; Unlike Cluster3, it contains 3 regions with the least population. The total population of the regions in this cluster is not as large as a region in Cluster3. However, there are 30 restaurants in this area.

From here; we can say that the number of vegetarian or vegan restaurants in the regions where population density is low is due to the socioeconomic and sociocultural structures of the people living in that region.

Conclusion

This project answers the following questions;

- What is the best location in Istanbul for vegan or vegetarian Cuisine?
- Which areas have potential vegan or vegetarian restaurant?
- Which areas in İstanbul lack vegan or vegetarian?