The Battle of Neighborhoods:

Analysis of Cuisine Diversity in Istanbul

Applied Data Science Capstone Project Report

Fuat Akal

February, 2020

Problem

- Istanbul is the most populous and diverse city in Turkey.
- It attracts millions of people as residents from several different ethnicities.
- Food Diversity is an important aspect of mega cities.
- The objective of this project is to segment districts of Istanbul into major clusters based on their their cuisines
  - probably to investigate food habits and taste of these clusters.

Data Set

- In this project, two main data sources are used.
  - The first one is a wiki page containing list of districts of Istanbul.
  - The second one is the Foursquare API to get venues in districts of Istanbul.

Methodology

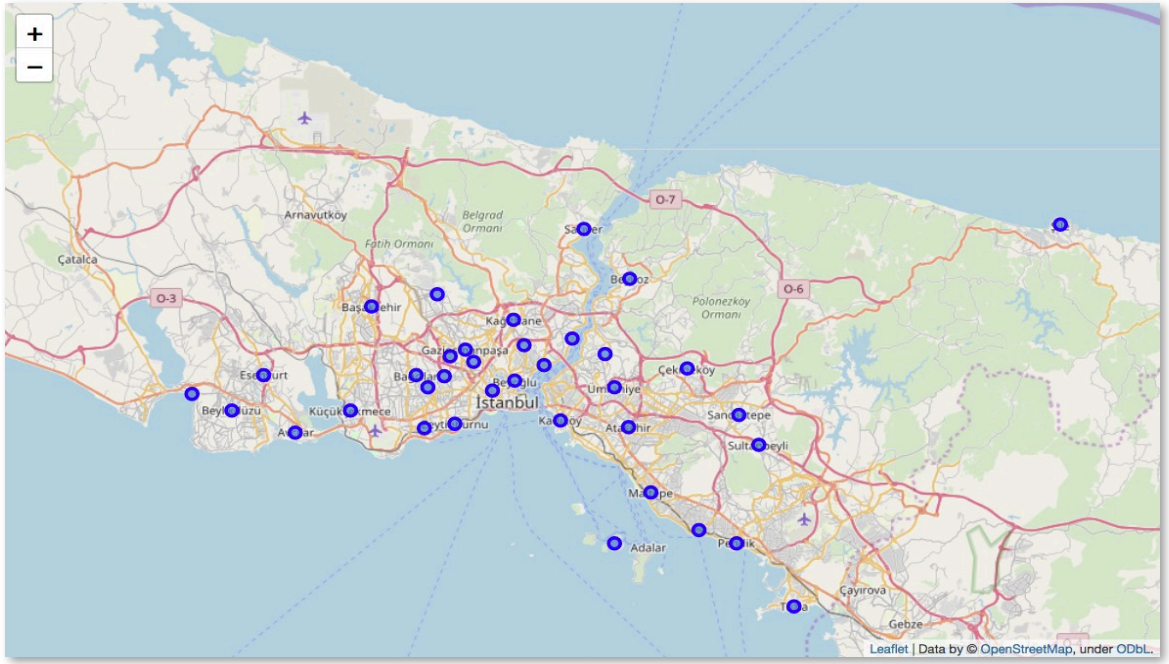
- I partly obtained the data from the Web, which is the list of districts in Istanbul.
- I was not able to go in to neighborhood detail due to lack of public data. then, i construct a data frame of districts along with their coordinates.

	district	latitude	longitude
0	Adalar	40.876259	29.091027
1	Arnavutköy	41.068394	29.041154
2	Ataşehir	40.984749	29.106720
3	Avcılar	40.980135	28.717547
4	Bağcılar	41.033899	28.857898

Methodology (cont'd)

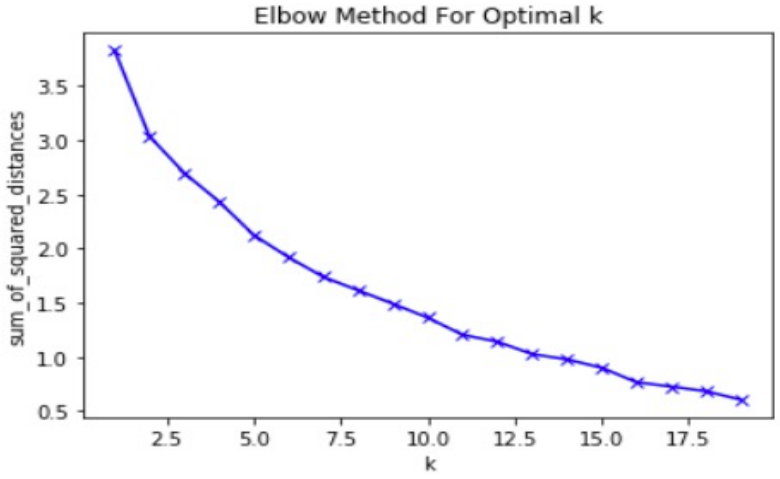
- Then, i used the Foursquare API to get the list of venues in those districts. The venues I am interested in are about food. Furthermore, I removed general food places like cafes.

- After that, I marked the food venues on the Istanbul map.



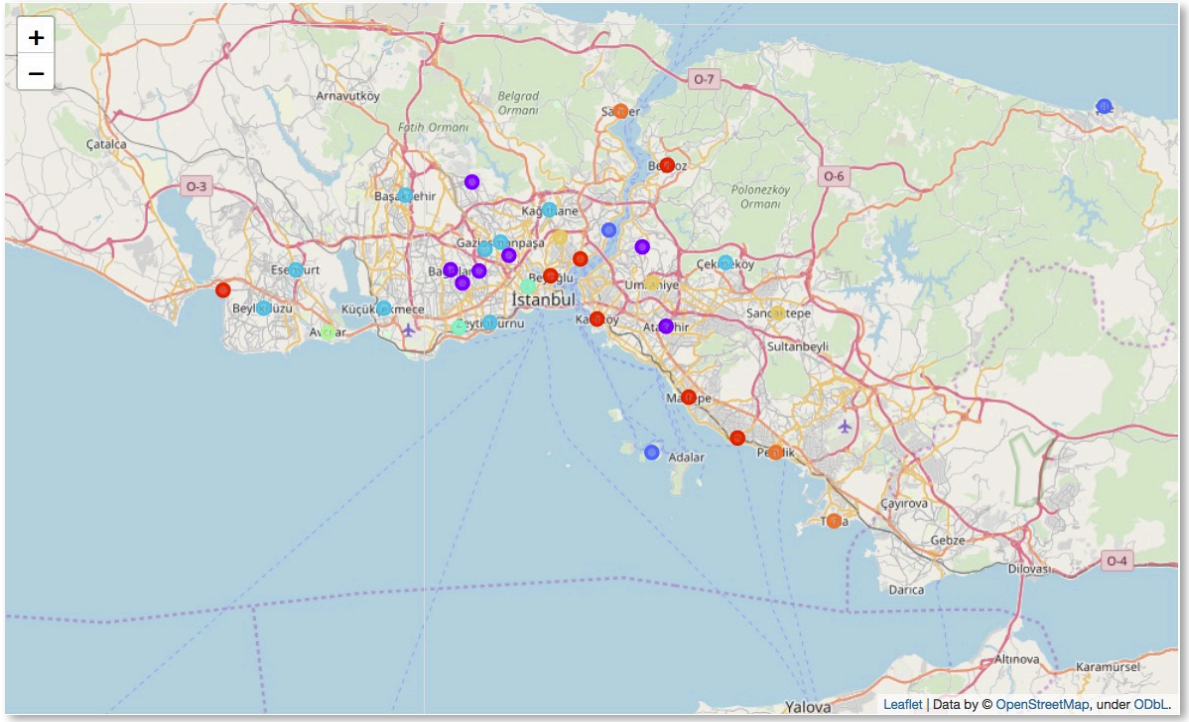
Methodology (cont'd)

- Once the places are determined, I used the k-means clustering algorithm.
- Before that, I tried the elbow method to find the optimum number of clusters, which was 7 in my experiments.



Methodology (cont'd)

- Finally, I found clusters and marked them on the Istanbul map.



Analysis Results

- In the analysis section, I investigated the clusters separately. After this analysis, it is clearly seen that which districts are enjoying which food.
- Results have two aspects.
  - First, we can easily pick a district if we have a specific food preference by using the result of analysis.
  - Second, the results might give an idea about opportunities to open a food place.

Conclusions

- This project really helped me improve especially my data handling skills. Getting and shaping the data was the most challenging part of the project.

- In addition, I feel happy as I produced some results which may help visitors and/or residents and investors of a mega city.

- However, I was not able to find detailed information publicly available for Istanbul. I stayed at the district level. That is why I see room in my study to further explore the clusters at the neigborhood level.