The Most Suitable Place to Open a Cafe in Istanbul

Kubilay Çıtak

1. Introduction and Business Problem

Istanbul is the largest and the most crowded city in Turkey. Therefore, you can find some many cafes, museums, operas, historical attractions in the city. Speaking of cafes, opening a cafe in Istanbul might be difficult but deciding to do so is even more difficult. It is obvious even before you decide to do so, that you will have many competitors. Not only that, you also must have great strategies and come up with great ideas to successfully open your café and operate it. For example, you must consider every detail of the place that you will open your café, such as borough, neighborhood, its local people, your target audience, the list goes on and on. Besides, you must come up with good ideas that will make you better than your competitors. The more crowded your café's place is, the more competition you will encounter. The less crowded your café's place is, the less your chance will be to attract customers. Therefore, your café's place is one of the most important criteria you must think of. This project aims to reach out to people involved in this business and give them a better idea of the potential location of their future café.

2. Data Acquisition

To make recommendations about Istanbul, first we need to get information about it. Boroughs and neighborhoods, and their populations and areas can be found in "atlasbig.com". Also, to be able to create a map, we need geographic coordination information of locations as well. To get latitude and longitude information of every borough, Python's geopy library was more than enough. And to get venues and their categories, a Foursquare Developer Application was created, and its client information was used to use Foursquare API and get the required values. Thereby, all the required data was successfully acquired.

3. Methodology

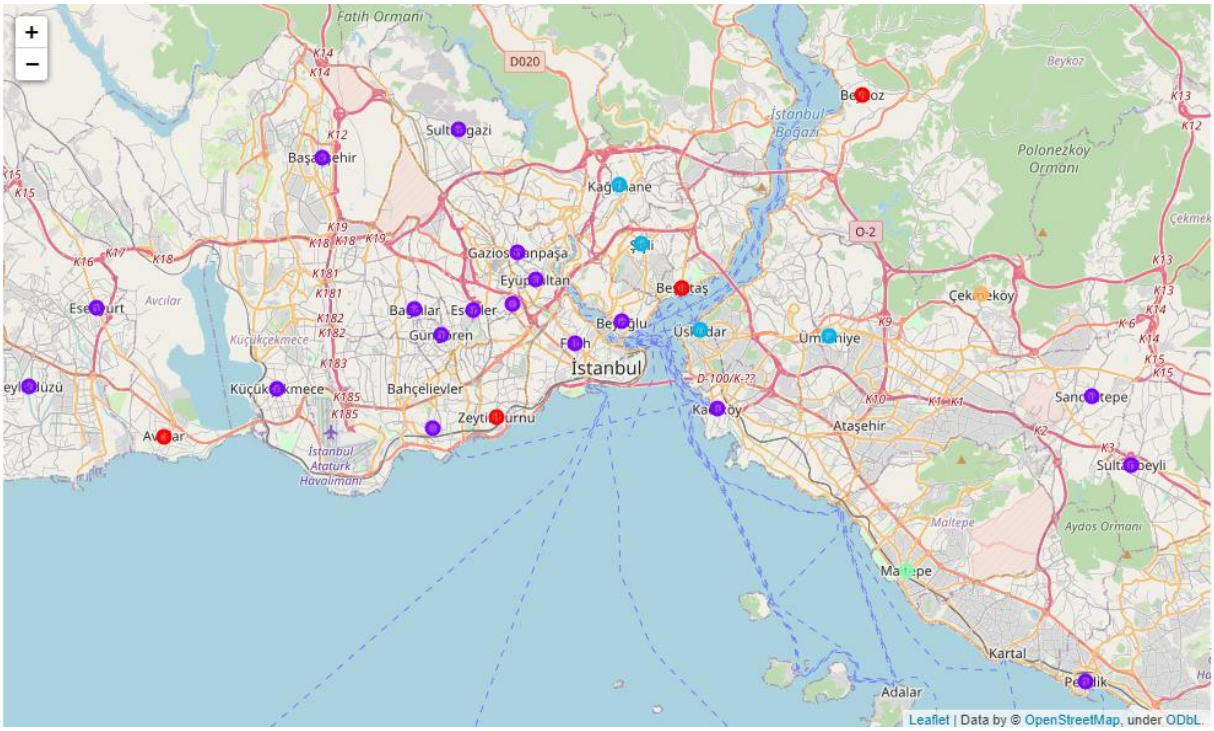
First, I used web scraping for the table on the source web site, which contains information about Istanbul's boroughs. After putting the table in a new DataFrame, I used geopy to get geographic information about the boroughs and added that information to the DataFrame as well. Especially in the data preparation part, I heavily used Python's Pandas and Numpy libraries. After having the data in a DataFrame, it gets quite simple to manipulate the data in any way you wish.

The next goal was to get boroughs' the most popular venue type into the DataFrame, and to do so I created a Foursquare application and used Foursquare API with that application's information. I added this data to the DataFrame too, and after having the DataFrame with all the necessary data, I applied one-hot encoding to DataFrame to turn its values into numeric values because the standard k-means clustering algorithm cannot be directly applicable to categorical data.

The next step was clustering. I used K-Means clustering model to fit my DataFrame and cluster venues and added each borough's cluster label to the main DataFrame. And by using this cluster labels, I created a map with folium library which shows the result of the project.

4. Results

The final map contains venues of boroughs which were located and clustered previously by the tools and the model. With the help of the map many results could be acquired, such as how close some venues and boroughs are to others, how left out some venues are, how clustering worked and separated venues etc.



5. Discussion

By looking at the final map, we can conclude that opening a cafe in a borough which is not too close or too away from others could be the best option. This way the cafe would not be into a great competition, also would be able to get customers as well. According to the map and the algorithm, new cafes in the light blue cluster could be successful.

6. Conclusion

In this project, Istanbul's boroughs and their many kind of information was analyzed by using several tools. To be able to recommend the best place to open a café, many tools were used such as Pandas, Numpy, folium, matplotlib, geopy, K-Means clustering model etc. As a future directions, information like population and areas of the boroughs, number of venues and neighborhoods could be used to get a better result.