Segmenting and Clustering Neighborhoods in Istanbul

Ayaz Ul Haq

1. Introduction

1.1 Background

Moonlight Tours is a travel agency for people to travel abroad for holidays with packages. They provide different packages i.e. Classic, Premium and Luxury and their Luxury package includes Hotel stay, food at top restaurants and sightseeing of top venues around the city. They make customer satisfaction their top priority and are famous for these services. They are branching out to Turkey as their new travel destination. To maintain their quality they are to make list of nearest venues with restaurants, hotels, cafes, shopping stores for their customers.

1.2 Problem

Analysis might contribute to determining the location with venues best suited for the travelers. Factors that might influence in selecting the location for stay and visit are different venues and their location towards one another i.e. hotels, restaurants, cafes, gym, festivals, and different places for the entertainment and visit.

1.3 Interest

Moonlight Tours would be very interested in the determination of different venues which might help them in attaining and attracting new customers as well. This will increase in the business value of the agency and will continue to be so in the future.

2. Data acquisition and cleaning

2.1 Data sources

The data for the neighborhoods is collected from Wikipedia. To complement these datasets, I scraped https://en.wikipedia.org/wiki/List_of_districts_of_Istanbul and I used geopy geocoding web services to get the coordinates of these neighborhoods.

2.2 Data cleaning

Data downloaded or scraped from Wikipedia cleaned and transformed into a panda's data frame. There were a lot of missing values from, because of unarranged record keeping.

Dataset returned population Area and density for these areas which were not useful for my purpose. Only neighborhood column was kept. Data scrapped from the source is shown in below figure:

	District	Population (2018)	Area (km²)	Density (per km²)
1	Adalar	16,119	11.05	1,458
2	Arnavutköy	270,549	450.35	600
3	Ataşehir	416,318	25.20	16,520
4	Avcılar	435,625	42.01	10,369
5	Bağcılar	734,369	22.36	32,842

. https://en.wikipedia.org/wiki/List_of_districts_of_Istanbul

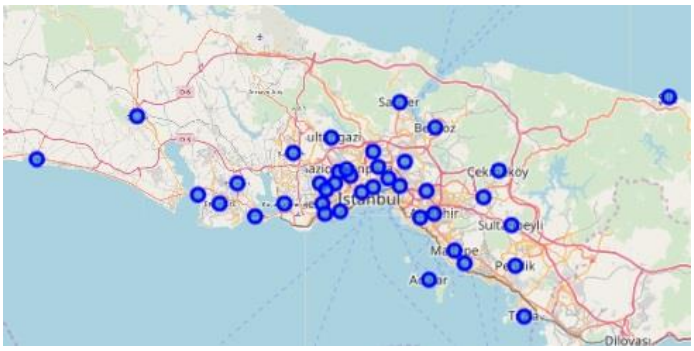
2.3 Feature selection

After data cleaning, I used geopy geo location web service to get the latitude and longitude for these areas. After I merged the neighborhood and their respective coordinates into one panda's data frame as shown in below figure.

	Neighborhood	Latitude	Longitude
0	Adalar,TR	40.875931	29.094742
1	Arnavutköy,TR	41.068394	29.041154
2	Ataşehir,TR	40.984749	29.106720
3	Avcılar,TR	40.980135	28.717547
4	Bağcılar,TR	41.033899	28.857898

3. Visualization:

After getting the coordinates of all the neighborhoods I get the coordinates of Istanbul and used folium library. **Folium** is a great visualization library. Feel free to zoom into the above map, and click on each circle mark to reveal the name of the neighborhood. I used marker feature of the folium library and iterate over all the neighborhoods to visualize the data we have gathered. Figure shown below:



4. Finding Venues

After getting the coordinates of all the neighborhoods I used Foursquare API to fetch the venues of the areas. I get top 100 venues within the 500 meter radius of first neighborhood. After fetching the data from the service I stored the data into the panda's data frame. Figure of the data frame is shown below:

	venue.name	venue.categories	venue.location.lat	venue.location.lng
0	L'isola Guesthouse	Bed & Breakfast	40.877038	29.096136
1	İnönü Evi Müzesi	History Museum	40.878251	29.093647
2	Luz Café	Café	40.877528	29.097877
3	Heybeliada Şafak Askeri Gazino	Restaurant	40.873609	29.099478
4	Merit Halki Palace Hotel	Hotel	40.878802	29.090974

5. Finding Venues

I repeated the same for all the neighborhoods in Istanbul and loop through each one of them to fine the top 10 most common venues. Figure of the data frame is shown below:

	Neighborhood	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	10th Most Common Venue
0	0	Café	Turkish Restaurant	Restaurant	Dessert Shop	Seafood Restaurant	Coffee Shop	Steakhouse	Bakery	Gym	Gym / Fitness Center
1	1	Wings Joint	Diner	Fast Food Restaurant	Farmers Market	Farm	Factory	Fabric Shop	Entertainment Service	Electronics Store	Eastern European Restaurant