3. Methodology

3.1 Data Preparation

3.2 Scraping Istanbul Wards Table from Wikipedia

I first make use of Special Wards of Istanbul page from Wiki to scrap the table to create a data-frame. For this, I’ve used pandas to transform the data in the table on the Wikipedia page into a dataframe .

3.3 Getting Coordinates of Major Districts : Geopy Client

Next objective is to get the coordinates of these major districts using geocoder class of Geopy client.

I used python folium library to visualize geographic details of Istanbul and its major districts and I created a map of Istanbul with boroughs superimposed on top. I used latitude and longitude values to get the visual .

3.4 Using Foursquare Location Data

3.5 cluster the neighborhood into 5 clusters.

4. Results & discussion

We got a glimpse of the Restaurants in Istanbul and were able to find out some interesting insights which might be useful to travelers as well as people with business interests. Let’s summarize our findings:

· Turkish Restaurant top the charts of most common venues in the districts.

· Avcılar and Bakırkö has maximum number of restaurants.

· Since the clustering was based only on the category of restaurants on each district, Istanbul central 5 wards (C5W) all fall in the same cluster, which indicate that each of those districts presents a similar experience to the traveler in terms of category of food.

· Tuzla has the least number of restaurants.

The clustering is completely based on the most common venues obtained from Foursquare data.

However, in our analysis, we have ignored other factors like distance of the venues from closest stations, range of prices of restaurants, since we don’t have such data and it would be difficult to farm it for a small exploratory study like ours. Hence, our analysis only helps travelers to get an overview of Restaurants distribution by categories in the major districts of Istanbul .

Furthermore, this results also could potentially vary if we use some other clustering techniques like DBSCAN.

5. Conclusion

In a fast-moving world, there are many real-life problems or scenarios where data can be used to find solutions to those problems. Like seen in the example above, data was used to cluster neighborhoods in Istanbul based on the most common food venues (Restaurants) in its major districts. The results can help a traveler to decide about the district that fit the most his needs.

I have made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the major districts of Istanbul and saw the results of segmentation of districts

Similarly, data can also be used to solve other problems, which most people face in metropolitan cities. Potential for this kind of analysis in a real-life problem is discussed in great detail. Also, some of the drawbacks and chance for improvements to represent even more realistic pictures are mentioned.