Coursera Capstone Project

IBM Applied Data Science Capstone Project

A apartment for everyone in Istanbul, Turkey

Alihan Karadağ

April 2019

Introduction:

Istanbul is one of the largest metropolises in the world where over 15 million people live and it has a population density of 2.813 people per square kilometer. As a Turkish citizen I decided to use Istanbul which is the most crowded city in Turkey in my project. The city is divided into 39 districts in total. However, the fact that the districts are squeezed into an area of approximately 72 square kilometers causes the city to have a very intertwined and mixed structure.

As you can see Istanbul is a city with high population and population density and as a foreigner it is hard to pick somewhere to live. As a apartment owner, everybody has own motivation to choose apartment. They may want to choose the district according to the social places density, may want high or low population, density and ofcourse house prices.

Objective of this project is analyse and select the best location according to customer's desires in Istanbul,Turkey to live. Using data science methodology and machine learning approach like clustring, this project aims to provide solutions to answer the business questions : if a person totaly foreing to Istanbul looking for apartment according to him/her desires, where would you recommend ?

Target Audience of This Project :

This project is useful for everyone who wants to buy a apartment in Istanbul,Turkey. You could be foreign or not, have knowledge about Istanbul or not, this project could help you to choose the best place to live according to your desires like house prices, density, social places.

Data:

To consider the problem we can list the datas as below:

- I used wikipedia's tablue of district of Istanbul to get Istanbul's districts.
- Geocoder for getting coordinates.
- I used Forsquare API to get the most common venues of given Borough of Istanbul.
- There are not too many public datas related to demographic and social parameters for the city of Istanbul. Therefor you must set-up your own data tables in most cases. In this case, I collected latest per square meter Housing Sales Price (HSP) Averages for each Borough of Istanbul from housing retail web page.

Methodology:

Wikipedia page of Istanbul contains a list of districts in Istanbul with a total of 39 districts. We will use web scraping techniques to extract the data from the Wikipedia page with python request and beautifulsoup packages. Then We will get geographical coordinates of the districts by using python Geocoder package which gives latitude and longitude coordinates of the districts.

After that, we will use Foursquare API to get the venue data for those districts. Foursquare is the one of the largest database of 105+ million places and is by over 125.000 developers. Foursquare API will provide many categories of the venue data.

We will use python folium library to visualize geographic details of Istanbul and its boroughs and we will create a map of Istanbul with boroughs.

After all data cleaning and data processing, we will use K-means machine learning approach.