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.

Pre-Plan:

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.

Methodology :

Firstly, we need to get list of neigbourhoods in Istanbul. Fortunately, the list is avaliable in the wikipedia page [1]. We will do web scraping using python requests and beautifulsoup packages to extract the list of neigbourhoods data. However, this is just a list of names. We need to get geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API.To do so, We will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the datas into the pandas dataframe and visualize the neigbourhood in a map using Folium package. This allows us to make sure that geographical data returned by geocoder are correctly plotted in Istanbul city map.

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare developer account in order to obtain the foursquare ID and Foursquare secret key. Then, we make API calls, passing geographical coordinates of the neigbourhoods in loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues are returned for each neigbourhood. Then, we will analyse each neigbourhood by grouping the rows by neigbourhood and taking the mean of the frequency of occurence of each venue category. By doing so,we are also preparing the data for use in clustring.

Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. According to inertia analyse, we will cluster the neighbourhoods into 3 clusters based on density,population,avarage house prices and popular venues. The results will allow us to identify which neighbourhoods will be the best option for customer who wants be apartment owner.

Result:

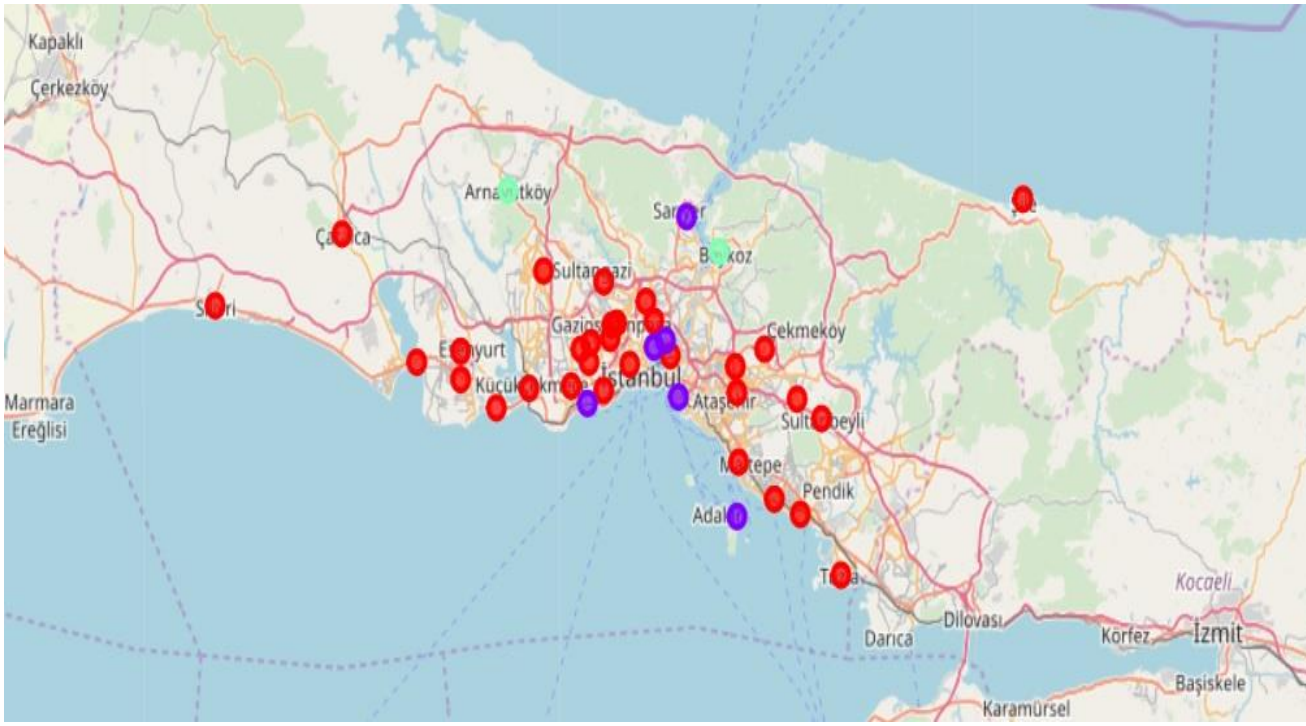
The result of the k-means clustering show that we can categorize the neigbourhood into 3 clusters based on density,population,avarage house prices and popular venues.

Cluster 1 = High Density, House Prices below avarage - kinda cheap-, Generally cafe and Turkish restaurants

Cluster 2 = Avagare density, High House prices, Have social areas like art gallery,theater,bar and lounges

Cluster 3 = Very Low density, avarage House prices, generally seafood restaurants,cafe and social activity areas.

The cluster result of the clustering are visualized in the map with cluster 0 in red, cluster 1 in purple, cluster 2 in green.



Discussion:

As observation noted from the map in the result section, most of the cluster 2 which is avarage density, high house prices and good social areas, placed on seaside. Cluster 3 placed on upstate where density is very low, avarage house prices and peaceful places. Cluster 1 placed inside of the city which is not seaside. Low house prices but high density and social places are generally cafe.

So If you want to escape from the city noise, Cluster 3 is the best option for you. Few people, classy places, avarage house prices are plus. But on the other hand, It is too far from the city center and you would not be actually live in Istanbul.

If you want to enjoy Istanbul, having fun, nightlife and city that never sleeps, cluster 2 is the best option for you. Bars, lounges, art galleries. Totally culturel and nightlife. On the cons side, House prices is high. Life in this cluster could be expensive.

In cluster 1, life is cheap and social areas are generally cafe or restaurant. For this cluster cons are, density and population are too high there. It means you feel the noise of city and crowded city so much. It could be exhaustive.

Conclusion:

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting, cleaning and preparing the data, visualization, performing machine learning by clustering the data into 3 clusters based on their density, avarage house prices and popular venues. Lastly, proveding recommendations to the relevant stakeholders to pick the best locations to live. I aimed that even if you have no idea Istanbul, we will give options to select the best place for you and this project will be guide of you.

References:

[1] - Disctrict list of Istanbul, Received from Wikipedia,

http://www.wikizero.biz/index.php?q=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnL3dpa2kvTGlzdF9vZl9kaXN0cmljdHNfb2ZfSXN0YW5idWw

[2] - Foursquare Developers Documentation, Received from Foursquare,

https://developer.foursquare.com/docs

[3] - Lastest House Prices, Received from Hürriyet Emlak

https://www.hurriyetemlak.com/Emlak-Endeksi/Detayli-Analiz/Istanbul