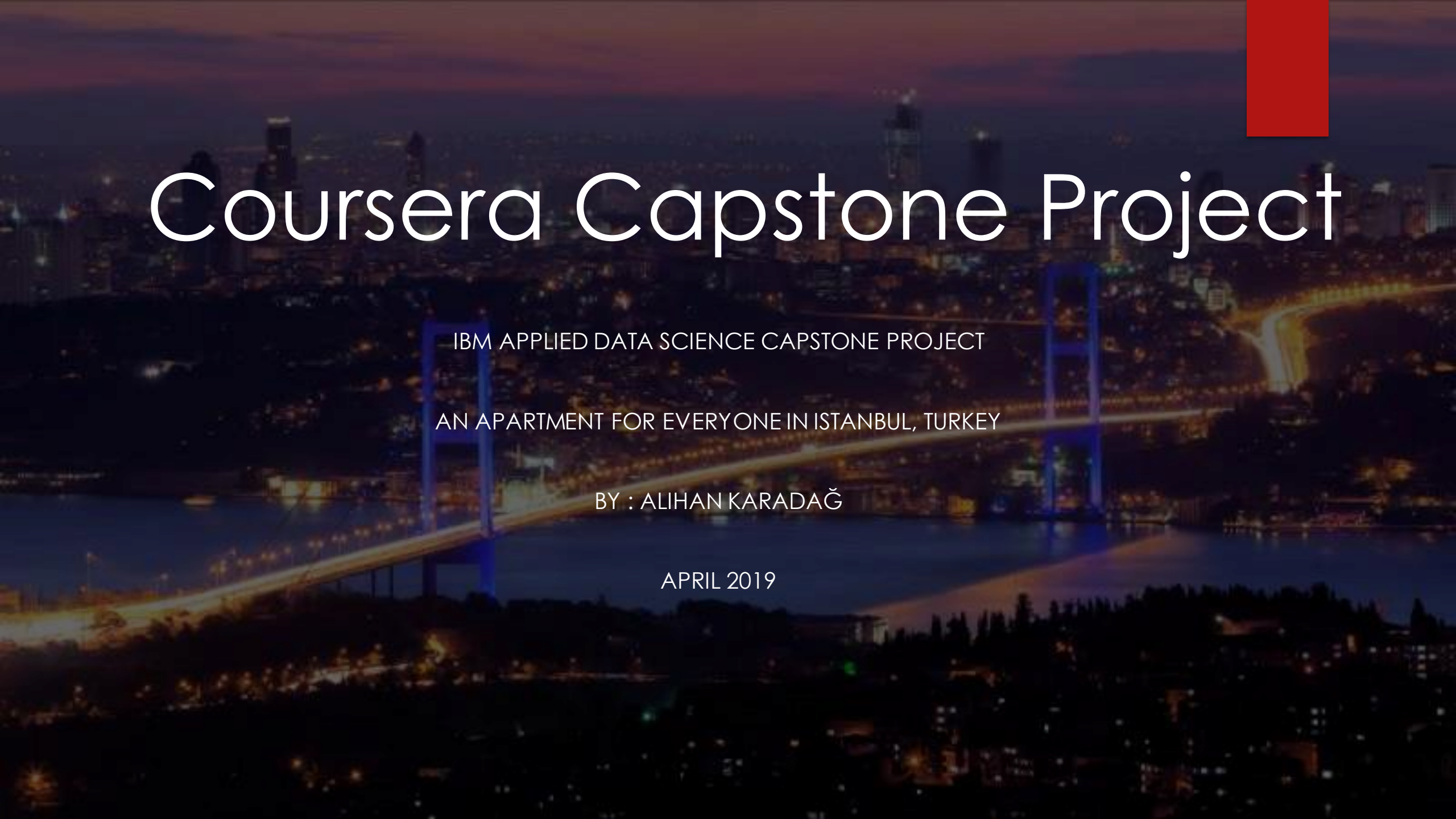


Coursera Capstone Project

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

AN APARTMENT FOR EVERYONE IN ISTANBUL, TURKEY

BY : 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 an 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 foreign to Istanbul looking for apartment according to him/her desires, where would you recommend ?

Target Audience of This Project :

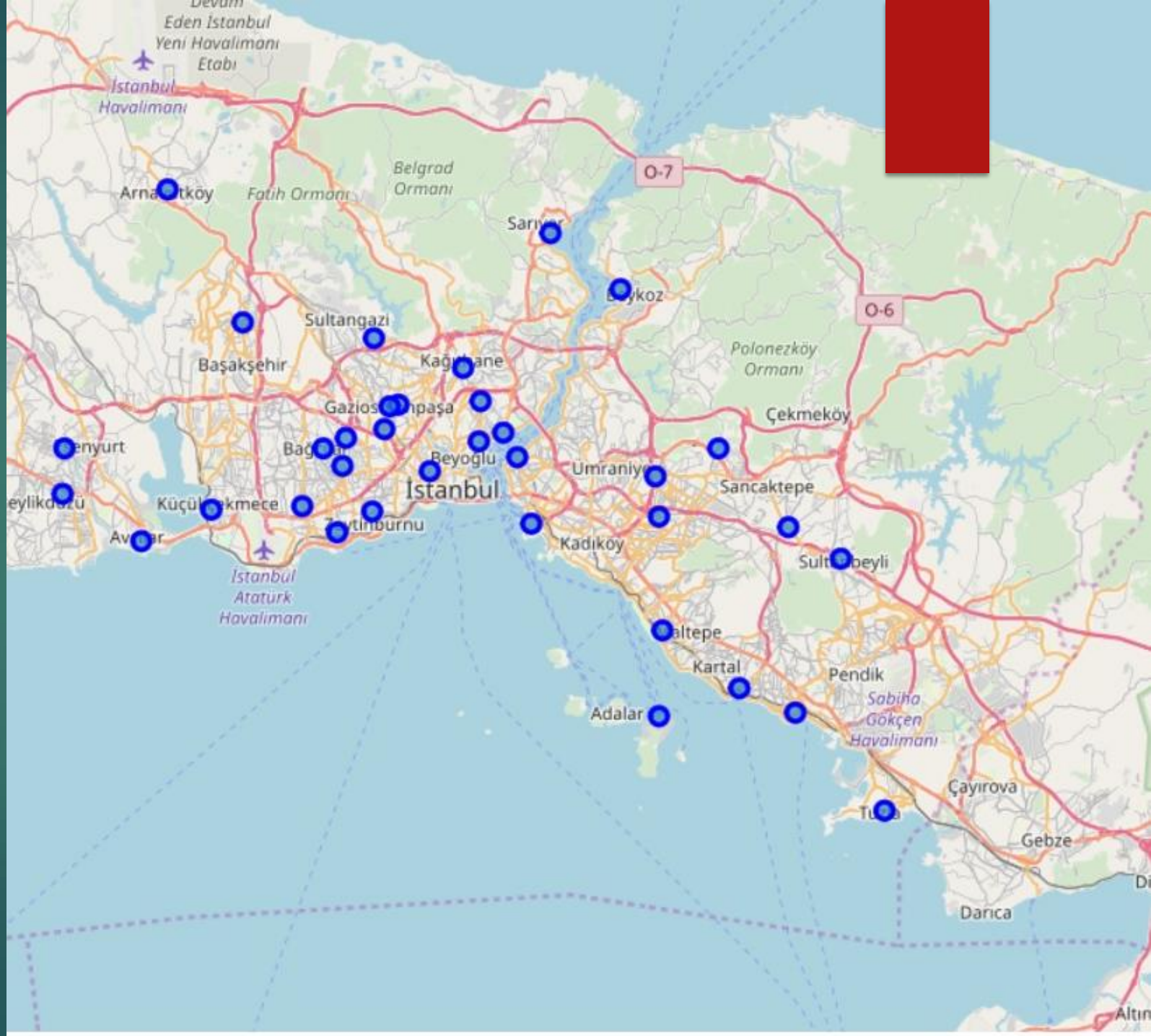
This project is useful for everyone who wants to buy an apartment in Istanbul,Turkey. You could be foreign or not, have knowledge about Istanbul or not, this project could help you to choose the best place to live according to your desires like house prices, density, social places.

Methodology :

- Firstly, we need to get list of neigbourhoods in Istanbul. Fortunately, the list is a v aliable in the wikipedia page. We will do web scraping using python request and beautifulsoup packages to extract the list of neigbourhoods data. Howe v er, 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 Geocoder package that will allow us to conv ert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the datas into the pandas datafra me and v isualize the neigbourhood in a map using Folium package. This allo ws us to make sure that geographical data returned by geocoder are correc tly plotted in Istanbul city map.

	District	Population	Density	Latitude	Longitude	Avg House Price(m2)
0	Adalar	16.119	1.458	40.87373	29.12797	6.728
1	Arnavutköy	270.549	600.000	41.18558	28.74147	2.124
2	Ataşehir	416.318	16.520	40.99248	29.12777	4.679
3	Avcılar	435.625	10.369	40.97813	28.72101	2.239
4	Bağcılar	734.369	32.842	41.03323	28.86351	3.918

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

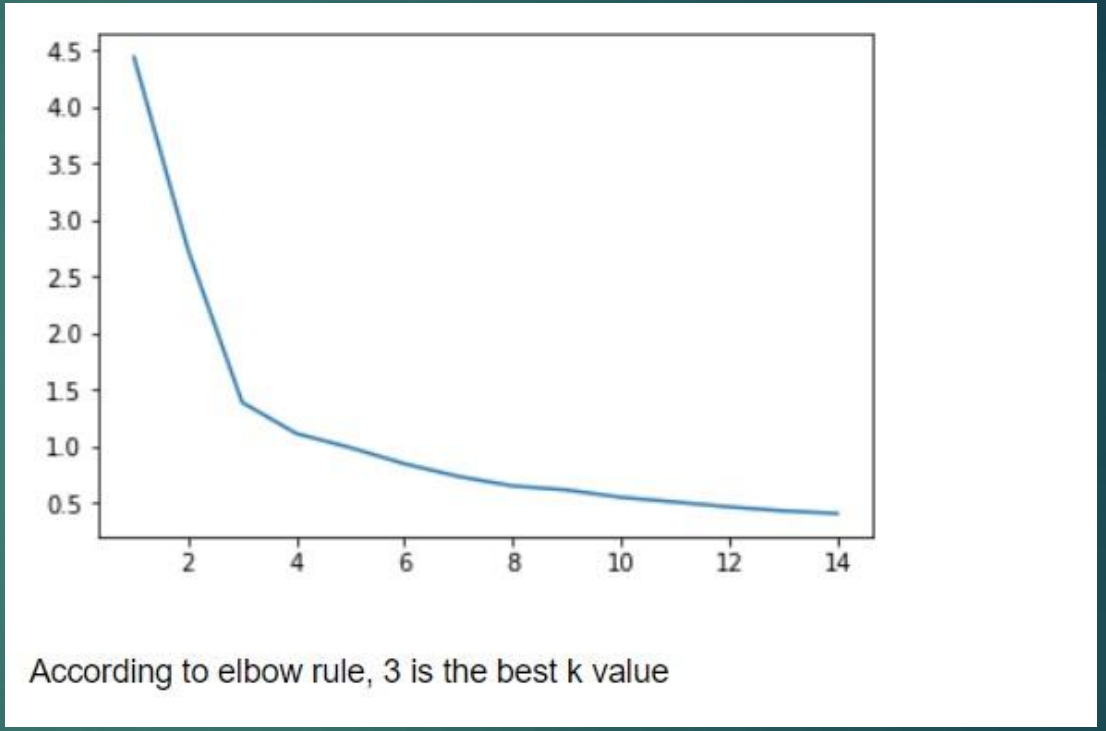


Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare developer account in order to obtain the foursquare ID and Foursquare secret key. Then, we make API calls, passing geographical coordinates of the neigbourhoods in loop. 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.

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
Hotel	Café	Beach	Boat or Ferry	Breakfast Spot	Bed & Breakfast	Harbor / Marina	Dessert Shop	Seafood Restaurant	Ice Cream Shop
Café	Turkish Restaurant	Restaurant	Gym	Fish & Chips Shop	Arcade	Dessert Shop	Diner	Kofte Place	Fast Food Restaurant
Café	Gym	Park	Pool	Seafood Restaurant	Kebab Restaurant	Coffee Shop	Dance Studio	Breakfast Spot	Pharmacy
Café	Dessert Shop	Gym / Fitness Center	Steakhouse	Coffee Shop	Breakfast Spot	Bar	Gym	Comfort Food Restaurant	Restaurant
Café	Turkish Restaurant	Gym	Bakery	Restaurant	Dessert Shop	Seafood Restaurant	Steakhouse	Middle Eastern Restaurant	Burger Joint
Café	Restaurant	Coffee Shop	Bakery	Gym	Turkish Restaurant	Pizza Place	Fast Food Restaurant	Yoga Studio	Clothing Store

We have some common venue categories in boroughs. In this reason I used unsupervised learning **K-means algorithm** to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning.

First, I will run K-Means to cluster the boroughs into **3** clusters because when I analyze the K-Means with elbow method it ensured me the 3 degree for optimum k of the K-Means.



According to elbow rule, 3 is the best k value

Here is my merged table with cluster labels for each borough.

	District	Population	Density	Latitude	Longitude	Avg House Price(m2)	Cluster Labels	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	8t Cc
0	Adalar	16.119	1.458	40.87373	29.12797	6.728	1	Hotel	Café	Beach	Boat or Ferry	Breakfast Spot	Bed & Breakfast	Harbor / Marina	[
1	Arnavutköy	270.549	600.000	41.18558	28.74147	2.124	2	Café	Turkish Restaurant	Restaurant	Gym	Fish & Chips Shop	Arcade	Dessert Shop	
2	Ataşehir	416.318	16.520	40.99248	29.12777	4.679	0	Café	Gym	Park	Pool	Seafood Restaurant	Kebab Restaurant	Coffee Shop	
3	Avcılar	435.625	10.369	40.97813	28.72101	2.239	0	Café	Dessert Shop	Gym / Fitness Center	Steakhouse	Coffee Shop	Breakfast Spot	Bar	
4	Bağcılar	734.369	32.842	41.03323	28.86351	3.918	0	Café	Gym	Turkish Restaurant	Kebab Restaurant	Dessert Shop	Steakhouse	Coffee Shop	

Result:

Cluster - 1

- High Density,
- House Prices below av arage, consider cheap
- Generally cafe and Turkish restaurants

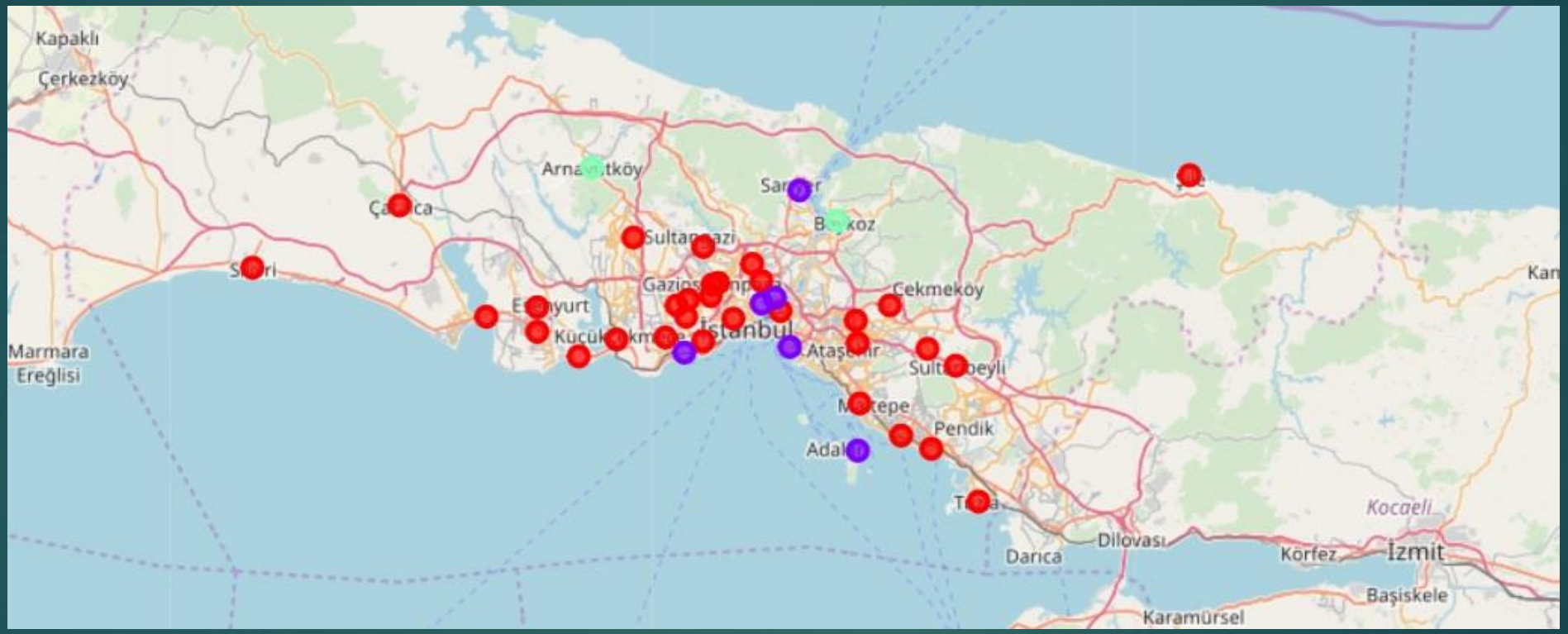
Cluster – 2

- Av agare density,
- High House prices,
- Hav e social areas like art gallery,theater,bar and lounges

Cluster - 3

- V ery Low density,
- Av arage House prices,
- Generally seafood restaurants, cafe and social activ ity areas.

Istanbul map with to clusters



The cluster result of the clustering are visualized in the map with cluster 1 in red, cluster 2 in purple, cluster 3 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 2 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 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. 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.

- Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

Referances:

- [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