Applied Data Science Capstone Project:

Istanbul - The City of Oppurtunities

Raheen Junaid Wani

26-01-21

I have taken my first step towards acquiring skills related to data science by doing the IBM Data Science Professional Certificate course on Coursera. The last module of this course is a capstone project. This project is about using a data science toolset on a real-life problem and demonstrating the use of learned skills.

Introduction:

Istanbul, the capital city of Turkey has a rich cultural history and heritage. It has an immense cultural exchange due to the influence of European and Middle Eastern culture. Istanbul can be considered as the combination of the old and modern city, as well as a mixture of cultures in a captivating atmosphere. There are only few cities in the world that can be so delightful to experience and enjoy as Istanbul. Istanbul is considered to be the bridge that links Asia with Europe from a cultural and geographical standpoint.

Istanbul offers an unforgettable experience for its travelers, with its colorful daily city life and dynamic nightlife. The beautiful silhouette of the city combines historical sites and monuments such as Roman aqueducts, Byzantine churches, Venetian towers, Ottoman palaces from the Byzantine, Ottoman and Turkish periods.

If I start writing about Istanbul, there will be no end to it as it offers so much to see and experience. Therefore, I will directly come to the point of discussion for the problem of this project.

Business Problem:

There won't be any other city where I personally would like to start a cafe or restaurant other than Istanbul, Turkey. If you are in Istanbul, you are one step away from tasting the delicious meals of the Turkish cuisine. Istanbul offers a wide range of Turkish cuisine at the variety of well established, good and luxury **Istanbul Restaurants.**

In this project, I will be considering a problem to recommend places in Istanbul where a cafe or a restaurant can be opened. The interested party or the audience will be the entrepreneurs who want to set up their food chain in the marvelous city of Istanbul. The place can be a crowded one which is already considered a hub of cafes and restaurants or it can be a quiet place where not many entrepreneurs have started cafes or restaurants yet. And this will be a good opportunity in my opinion so as to stand out from the rest.

Data:

The following data will be used for this project:

1. List of the districts of Istanbul.
2. Geo-coordinates of the districts in Istanbul.
3. Top venues of districts.

The details about the data is as follows:

- List of districts will be obtained from Wikipedia. The link used is:

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

- Geo-coordinates of districts will be obtained with the help of the geocoder tool in the notebook which will be imported using the library available in python.
- Data about Top venues of Istanbul will be obtained from Foursquare API.Foursquare is a US tech company from New York focusing on location data. Their technology and data powers apps such as Apple's Maps, Uber, Twitter etc.

Methodology:

First, the data was scrapped from wikepedia which resulted in a pandas dataframe, consisting of information regarding the 39 districts of Istanbul. The data then had to be cleaned which resulted in a dataframe consisting of a single column with all the district names. This is shown below:

Out[6]:

	District
0	Adalar
1	Arnavutköy
2	Ataşehir
3	Avcılar
4	Bağcılar
5	Bahçelievler
6	Bakırköy
7	Başakşehir
8	Bayrampaşa
9	Beşiktaş
10	Beykoz

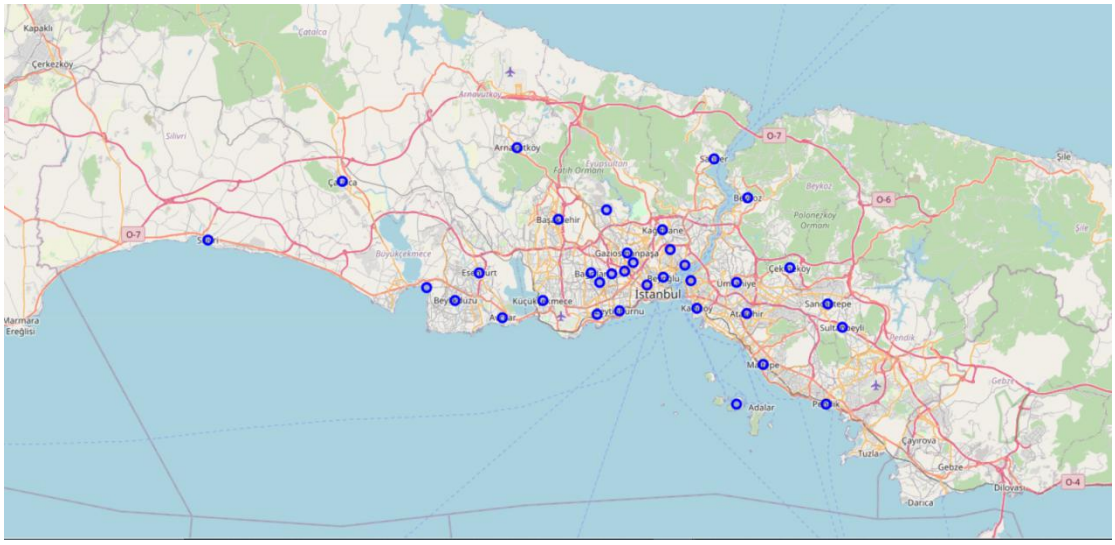
The next step is getting location data for these districts. To do that I used the geopy package with Nominatim. With that package, I get geographical

coordinates of each district in the dataset and add that information as columns Latitude and Longitude.

Out[12]:

	District	Latitude	Longitude
0	Adalar	40.876259	29.091027
1	Arnavutköy	41.184182	28.740729
2	Ataşehir	40.984749	29.106720
3	Avcılar	40.980135	28.717547
4	Bağcılar	41.033899	28.857898
5	Bahçelievler	38.881312	35.627761
6	Bakırköy	40.983541	28.867974
7	Başakşehir	41.097693	28.806163
8	Bayrampaşa	41.035738	28.912260
9	Beşiktaş	41.042847	29.007528

Since, we have information about latitudes and longtitudes, plotting folium map of Istanbul after getting the coordinates of Istanbul itself was done. The map shows the districts marked with a blue circle.

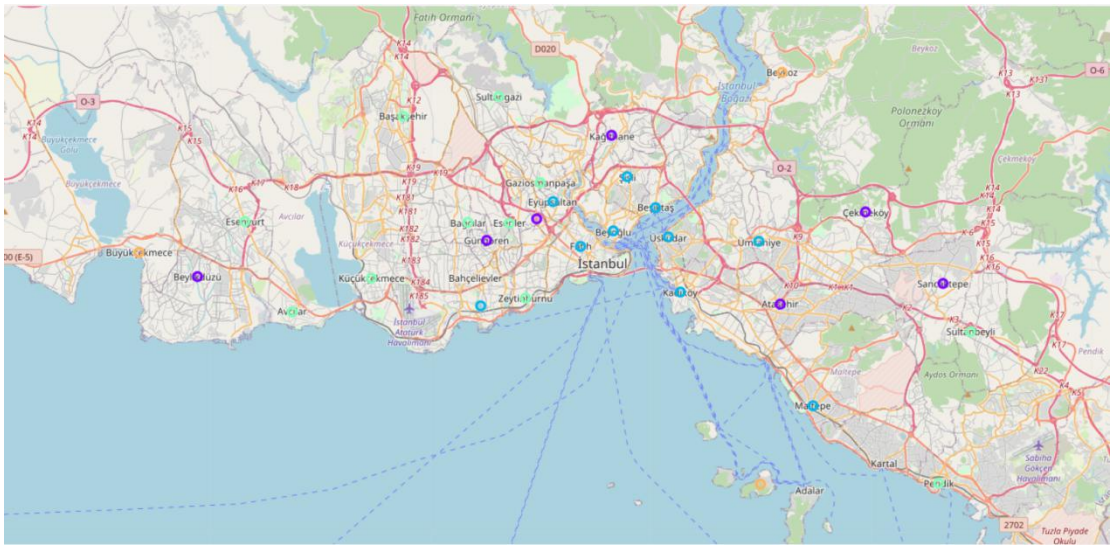


The next step was to use the FourSquare API to get the venues of all districts. A little more processing was done to the data after getting the venues.

Once the venues were ready, I started the process of k-means clustering. To make data ready for clustering one hot encoding was done. Tha data was then divided into 5 clusters and then plotted again using folium. Different colors were used to mark the clusters.

	Neighbourhood	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	Adalar	Seafood Restaurant	Café	Restaurant	Campground	Pool	History Museum	Trail	Steakhouse	Other Great Outdoors	Bed & Breakfast
1	Arnavutköy	Café	Turkish Restaurant	Restaurant	Fish & Chips Shop	Dessert Shop	Kofte Place	Bakery	Electronics Store	Fast Food Restaurant	Gym / Fitness Center
2	Ataşehir	Restaurant	Kebab Restaurant	Basketball Stadium	Clothing Store	Coffee Shop	Shopping Mall	Spa	Motorcycle Shop	Steakhouse	Tennis Court
3	Avcılar	Café	Dessert Shop	Mobile Phone Shop	Men's Store	Bar	Restaurant	Burger Joint	Souvenir Shop	Sporting Goods Shop	Steakhouse
4	Bahçelievler	Stables	Bay	Garden	Farm	Scenic Lookout	Lake	Food	Flea Market	Fishing Spot	Fish Market

In this table, we see that cluster labels assigned by the k-means clustering algorithm.
The map after clustering was done is also shown below:



Results:

The clusters were then analysed. It was found that cluster 3 and 4 were most appropriate for opening up a cafe.

Cluster 3:

	Neighbourhood	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
6	Bakırköy	Café	Coffee Shop	Hotel	Bookstore	Pizza Place	Bakery	Pool	Food Court	Spa	Salsa Club
9	Beşiktaş	Hotel	Coffee Shop	Café	Beer Garden	Roof Deck	Lounge	Spa	Falafel Restaurant	Chinese Restaurant	Chocolate Shop
12	Beyoğlu	Hotel	Café	Coffee Shop	Bar	Historic Site	Plaza	Concert Hall	History Museum	Kebab Restaurant	Music Store
18	Eyüpsultan	Café	Historic Site	Pide Place	Dessert Shop	Turkish Restaurant	Kebab Restaurant	Jewelry Store	Plaza	Pharmacy	Pedestrian Plaza
19	Fatih	Mosque	Café	Dessert Shop	Bakery	Hookah Bar	Hotel	Turkish Restaurant	Bridal Shop	Fish & Chips Shop	Steakhouse
22	Kadıköy	Coffee Shop	Café	Pizza Place	Candy Store	Ice Cream Shop	Theater	Mosque	Escape Room	Opera House	Doner Restaurant
26	Maltepe	Coffee Shop	Café	Jewelry Store	Tea Room	Gym / Fitness Center	Food Court	Middle Eastern Restaurant	Diner	Dessert Shop	Cosmetics Shop
34	Şişli	Coffee Shop	Hotel	Steakhouse	Restaurant	Historic Site	Chinese Restaurant	Lounge	French Restaurant	Café	Snack Place
36	Ümraniye	Coffee Shop	Café	Turkish Restaurant	Jewelry Store	Bistro	Accessories Store	Bath House	Dessert Shop	Restaurant	Electronics Store
37	Üsküdar	Coffee Shop	Café	Historic Site	Mosque	Turkish Restaurant	Miscellaneous Shop	Sporting Goods Shop	Smoke Shop	Motorcycle Shop	Candy Store

Cluster 4:

	Neighbourhood	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
1	Arnavutköy	Café	Turkish Restaurant	Restaurant	Fish & Chips Shop	Dessert Shop	Kofte Place	Bakery	Electronics Store	Fast Food Restaurant	Gym / Fitness Center
3	Avcılar	Café	Dessert Shop	Mobile Phone Shop	Men's Store	Bar	Restaurant	Burger Joint	Souvenir Shop	Sporting Goods Shop	Steakhouse
4	Bağcılar	Café	Gym	Dessert Shop	Coffee Shop	Convenience Store	Steakhouse	Jewelry Store	Garden	Furniture / Home Store	Mediterranean Restaurant
7	Başakşehir	Café	Dessert Shop	Steakhouse	Çöp Şiş Place	Jewelry Store	Sandwich Place	Park	Nail Salon	Music Venue	Kuruyemişçi
16	Esenler	Café	Gym	Restaurant	Ice Cream Shop	College Quad	Spa	Burger Joint	Snack Place	Seafood Restaurant	Clothing Store
17	Esenyurt	Café	Hotel	Restaurant	Turkish Restaurant	Men's Store	Bookstore	Burger Joint	Kebab Restaurant	Smoke Shop	Public Art
20	Gaziosmanpaşa	Café	Turkish Restaurant	Restaurant	Arcade	Food & Drink Shop	Coffee Shop	Buffet	Gym	Beer Garden	Steakhouse
25	Küçükçekmece	Café	Turkish Restaurant	Gym	Seafood Restaurant	Gym / Fitness Center	Soup Place	Ice Cream Shop	Theater	Coffee Shop	Doner Restaurant
27	Pendik	Café	Dessert Shop	Gym / Fitness Center	Salon / Barbershop	Art Gallery	Coffee Shop	Mobile Phone Shop	Meyhane	Steakhouse	Food Court
29	Sarıyer	Café	Bakery	Gym	Steakhouse	Kebab Restaurant	Seafood Restaurant	Turkish Restaurant	Soccer Stadium	Fish & Chips Shop	Burger Joint
31	Sultanbeyli	Café	Restaurant	Coffee Shop	Turkish Restaurant	Kebab Restaurant	Electronics Store	Hookah Bar	Cosmetics Shop	Tea Room	Fish & Chips Shop

Clusters 2 and 5 were most appropriate for opening restaurants.

Cluster 5:

	Neighbourhood	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	Adalar	Seafood Restaurant	Café	Restaurant	Campground	Pool	History Museum	Trail	Steakhouse	Other Great Outdoors	Bed & Breakfast
10	Beykoz	Restaurant	Seafood Restaurant	Fast Food Restaurant	Waterfront	Scenic Lookout	Café	Sculpture Garden	Food Court	Cheese Shop	Other Great Outdoors
13	Büyükçekmece	Seafood Restaurant	Turkish Restaurant	Café	Dessert Shop	Salon / Barbershop	Scenic Lookout	Gym / Fitness Center	Coffee Shop	Kokoreç Restaurant	Rest Area
30	Silivri	Seafood Restaurant	Steakhouse	Turkish Restaurant	Café	Gym / Fitness Center	Kokoreç Restaurant	Tea Room	Dessert Shop	Bar	Salon / Barbershop

Discussion:

In this project, I tried to use all methods I learned through courses like data cleaning, scraping, handling, analysis, and getting results with machine learning algorithms. Using location data and Foursquare API was exciting and new to me and I tried to learn more with documentation.

Conclusion:

In the end, I reached the goal that I declared in the first section. With the details given in the result section, an entrepreneur coming into the city of Istanbul can see which place will be the most approprite for opening a cafe or restaurant, depending on different conditions and wishes. But the number of clusters can be determined with a more systematic way to improve results. And to make further analysis.

References:

The link to the jupyter notebook of this project can be found on github:
https://github.com/raheen14/Coursera_Capstone/blob/main/The%20Battle%20Of%20Neighborhoods%20Continued%20(code).ipynb

The End!