

Applied Data Science Capstone Project: Restaurants in Istanbul

By Ali Baran Tasdemir

In the last months of our lives, most of us stuck in our homes due to coronavirus. So I had lots of free time to spend and decided to improve my skills which I can benefit later. Therefore I used my free time to take some courses on Coursera. And I recommend to everyone interested in Data Science this course(specialization) provided by IBM: <https://www.coursera.org/professional-certificates/ibm-data-science>

The final assignment of this course is “Capstone Project”. We are expected to use the tools and methods learned and apply those on a self-chosen challenge or problem in the general idea of “Battle of Neighborhoods”. To pass this assignment, creating a Jupyter notebook with Python and creating a blog-post is required. So I am writing this Medium post. I hope you enjoy it.

INTRODUCTION

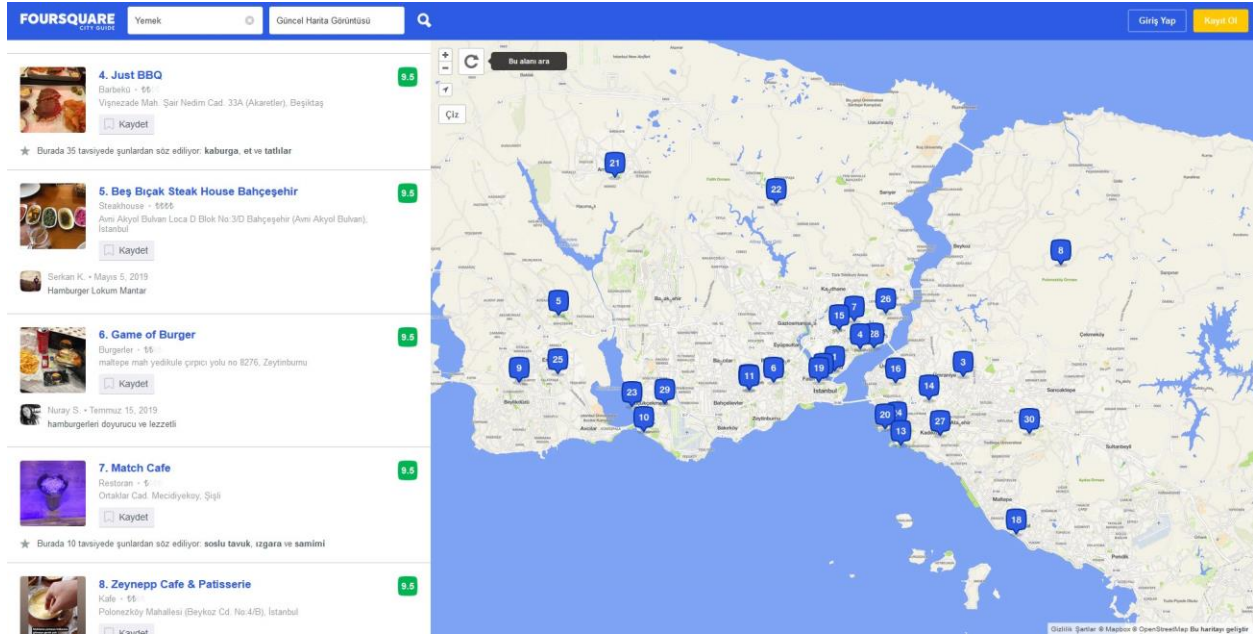
Istanbul is the most crowded city in Turkey and also in Europe with 15.5 million population. Besides that, Istanbul is hosting almost 12 million foreign tourists every year. Istanbul is a crowded and dense city. The city has 39 districts and all of them has different culture and characteristics.

For tourists, finding the right place to eat in Istanbul can be a challenge. Istanbul is a mix of Turkish and European culture. Turkish food includes lots of meat and spice. It may not be the first choice for everyone. So another tourist may want to taste seafood, etc. It is good to have something to check before choosing a place to eat.

Thus, the goal I want to achieve with this project is to recommend to tourists visiting Istanbul about restaurants. They can find answers to “Where to eat Seafood, Turkish food?”, “In the district, I am visiting which type of restaurants presents?”.

DATA

I will use Foursquare data about restaurants in Istanbul. Foursquare is an application that gives information about places near a location or in a city.



An example search on foursquare for a restaurant in Istanbul

Also, I will use the Wikipedia page of Districts of Istanbul to easy-access the table of the districts: https://en.wikipedia.org/wiki/List_of_districts_of_Istanbul

On this page, there is a table for all districts with information on population, area, density, etc. I will use these districts and the data about restaurants in these districts from foursquare to show the distribution and density of restaurants in them.

Methodology

In this section, I will talk about how I used the data and my initial data analysis.

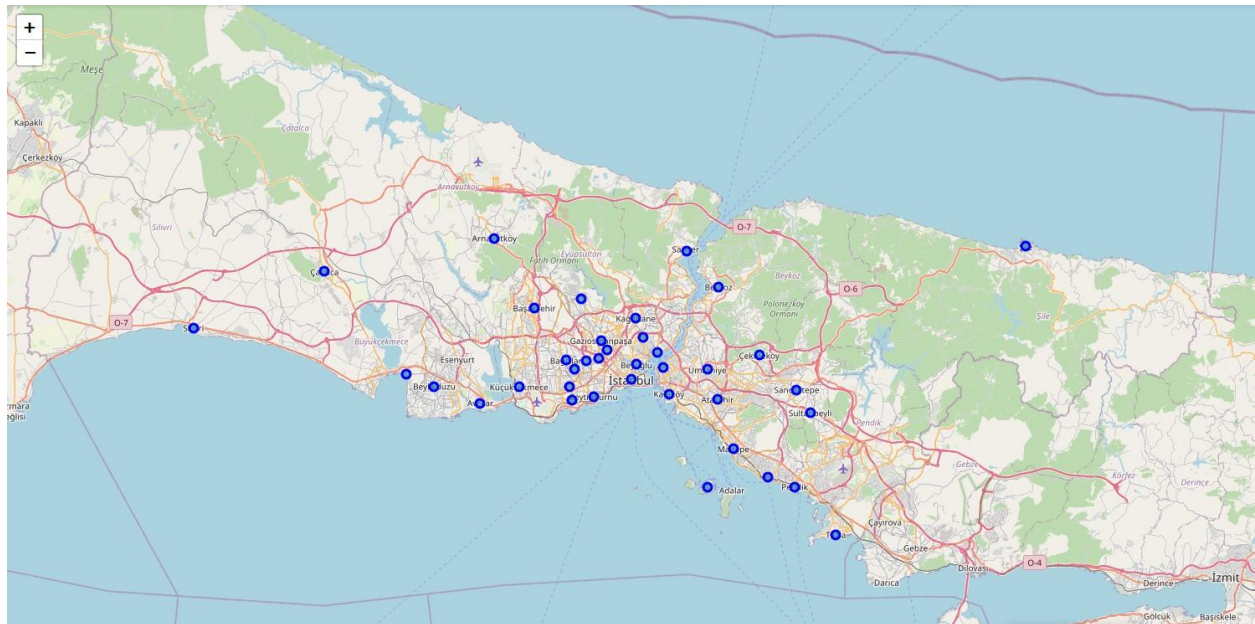
First of all, I started with a little web scraping to get districts of Istanbul. To do that I used pandas HTML read function and after a little cleaning and adjustments, I get this data frame.

	District	Population (2019)	Area (km ²)	Density (per km ²)	Mensual household income TL(USD)	Annual household income TL(USD)
0	Adalar	15238	11.05	1379	6.652₺ (918\$)	79.821₺ (10,978\$)
1	Arnavutköy	282488	450.35	627	2.030₺ (279\$)	24.360₺ (3,350\$)
2	Ataşehir	425094	25.23	16849	6.577₺ (904\$)	78.924₺ (10,854\$)
3	Avcılar	448882	42.01	10685	3.662₺ (503\$)	43.938₺ (6,064\$)
4	Bağcılar	745125	22.36	33324	3.197₺ (441\$)	38.367₺ (5,295\$)
5	Bahçelievler	611059	16.62	36766	4.674₺ (645\$)	56.088₺ (7,741\$)
6	Bakırköy	229239	29.64	7734	8.845₺ (1,220\$)	106.140₺ (14,650\$)
7	Başakşehir	460259	104.30	4413	4.513₺ (622\$)	54.152₺ (7,474\$)
8	Bayrampaşa	274735	9.61	28588	3.480₺ (480\$)	41.762₺ (5,764\$)
9	Beşiktaş	182649	18.01	10142	10.560₺ (1,457\$)	126.720₺ (17,490\$)

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.

	District	Population (2019)	Area (km ²)	Density (per km ²)	Mensual household income TL(USD)	Annual household income TL(USD)	Latitude	Longitude
0	Istanbul-Adalar	15238	11.05	1379	6.652t (918\$)	79.821t (10.978\$)	40.876259	29.091027
1	Istanbul-Arnautköy	282488	450.35	627	2.030t (279\$)	24.360t (3.350\$)	41.184182	28.740729
2	Istanbul-Ataşehir	425094	25.23	16849	6.577t (904\$)	78.924t (10.854\$)	40.984749	29.106720
3	Istanbul-Avcılar	448882	42.01	10685	3.662t (503\$)	43.938t (6.064\$)	40.980135	28.717547
4	Istanbul-Bağcılar	745125	22.36	33324	3.197t (441\$)	38.367t (5.295\$)	41.033899	28.857898
5	Istanbul-Bahçelievler	611059	16.62	36766	4.674t (645\$)	56.088t (7.741\$)	41.000290	28.863745
6	Istanbul-Bakırköy	229239	29.64	7734	8.845t (1.220\$)	106.140t (14.650\$)	40.983541	28.867974
7	Istanbul-Başakşehir	460259	104.30	4413	4.513t (622\$)	54.152t (7.474\$)	41.097693	28.806163
8	Istanbul-Bayrampaşa	274735	9.61	28588	3.480t (480\$)	41.762t (5.764\$)	41.035738	28.912260
9	Istanbul-Beşiktaş	182649	18.01	10142	10.560t (1.457\$)	126.720t (17.490\$)	41.042847	29.007528

So far I had the districts and the geographical locations of them. That information is enough to create a Folium map to see where these districts on the map.



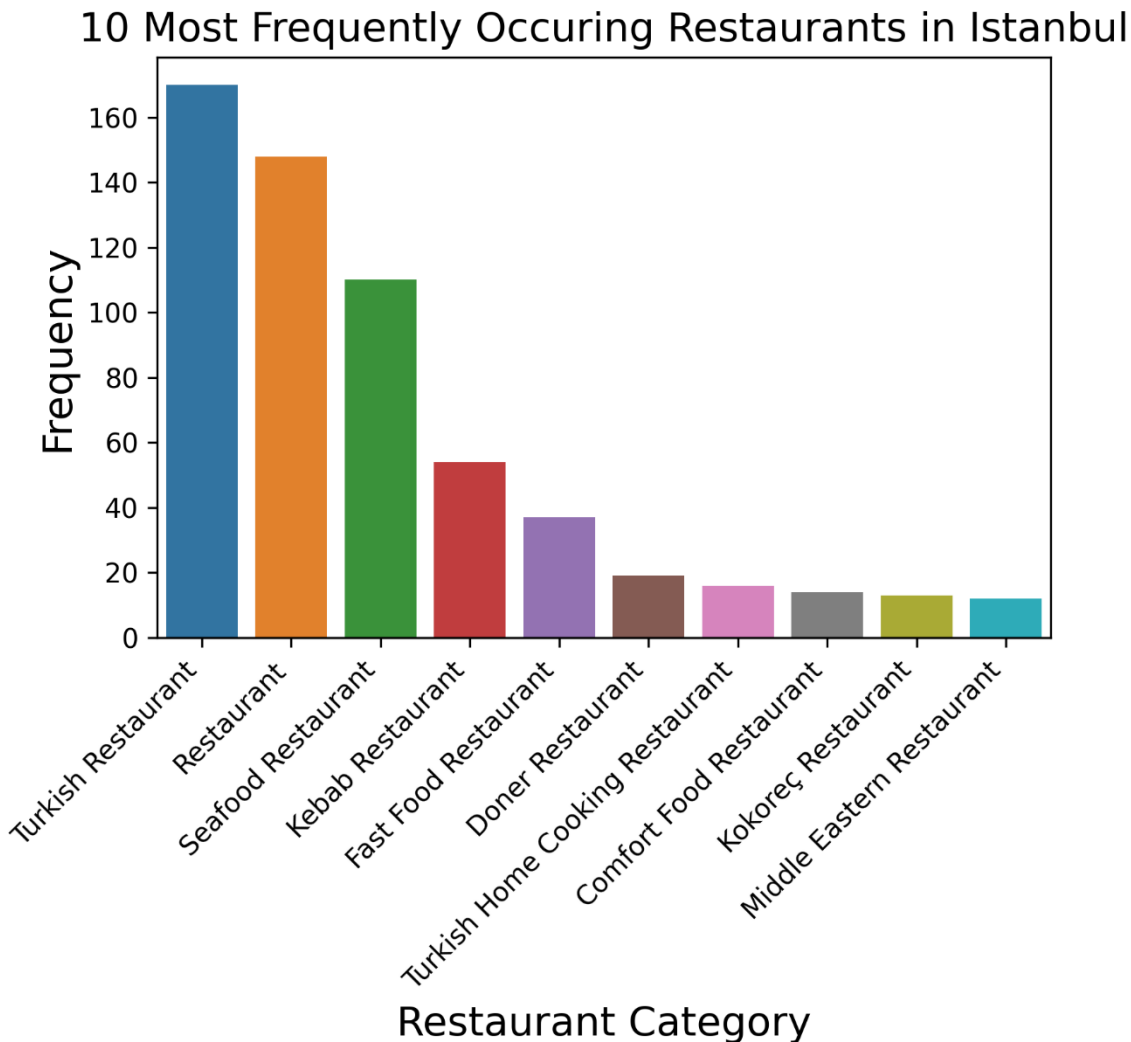
Istanbul map with district centers pointed

Until that point, I never used foursquare data. Now its time to use foursquare API to get venues for each district.

I search for a maximum of 1000 venues in a 2km radius for each district center. That resulted in 3741 venues from 38 districts. In 3741 venues there are 652 restaurants. And these 652 restaurants serve 33 different types of foods such as Turkish, Kebab, and Seafood.

To see the distribution, I plotted a bar plot for 10 most frequently observed restaurants in the city. We can see that the most occurring restaurant in Istanbul is Turkish restaurants as expected. The second one is the restaurants labeled as “Restaurant”. I assume that these restaurants are the

generic type or labeled as just “Restaurant” by owners (not specified the type). And the third one is the Seafood restaurants. It is reasonable too because Istanbul is a city surrounded by sea from everywhere. That’s why fresh fishes reachable all the time and you can eat your seafood with an amazing view of Bosphorus.



To make ready the data for k-means analysis, I applied one-hot encoding to restaurants that I get with Foursquare API. In the end, I get 33 columns(unique types of restaurants in total) of 0/1’s for 652 restaurants.

To make further analysis, I grouped this data for districts and show frequencies of the types of restaurants and listed the most common 5 types of restaurants for each district.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Istanbul-Adalar	Seafood Restaurant	Restaurant	Turkish Restaurant	Fast Food Restaurant	Mediterranean Restaurant
1	Istanbul-Arnavutköy	Turkish Restaurant	Restaurant	Turkish Home Cooking Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant
2	Istanbul-Ataşehir	Restaurant	Seafood Restaurant	Kebab Restaurant	Italian Restaurant	Turkish Home Cooking Restaurant
3	Istanbul-Avcılar	Restaurant	Seafood Restaurant	Fast Food Restaurant	Theme Restaurant	Kebab Restaurant
4	Istanbul-Bahçelievler	Seafood Restaurant	Italian Restaurant	Turkish Restaurant	Turkish Home Cooking Restaurant	Caribbean Restaurant
5	Istanbul-Bakırköy	Turkish Restaurant	Restaurant	Fast Food Restaurant	Kebab Restaurant	Italian Restaurant
6	Istanbul-Bayrampaşa	Turkish Restaurant	Restaurant	Comfort Food Restaurant	Kebab Restaurant	Seafood Restaurant
7	Istanbul-Bağcılar	Kebab Restaurant	Turkish Restaurant	Mediterranean Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant
8	Istanbul-Başakşehir	Kebab Restaurant	Fast Food Restaurant	Restaurant	Middle Eastern Restaurant	Turkish Restaurant
9	Istanbul-Beykoz	Seafood Restaurant	Restaurant	Comfort Food Restaurant	Turkish Restaurant	Fast Food Restaurant

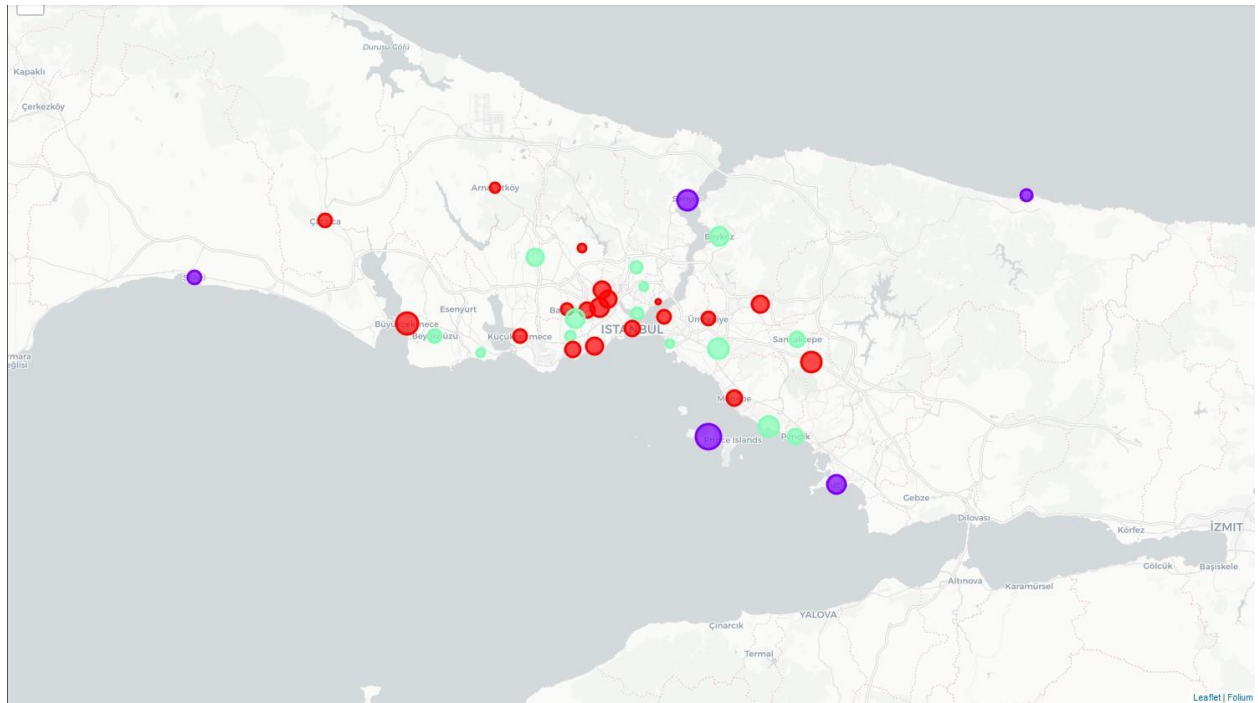
With all this data, finally, I can run a clustering algorithm to reach the goal of this project. I choose the k-means clustering algorithm from the scikit-learn package. To start the k-means algorithm I need to choose a k(num. of clusters) value. I can pick this number more systematically but with my initial observations, I decided to choose 3.

RESULTS

	Neighborhood	Population (2019)	Area (km²)	Density (per km²)	Mensual household income TL(USD)	Annual household income TL(USD)	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Istanbul-Adalar	15238	11.05	1379	6.652t (918\$)	79.821t (10.978\$)	40.876259	29.091027	1.0	Seafood Restaurant	Restaurant	Turkish Restaurant	Fast Food Restaurant	Mediterranean Restaurant
1	Istanbul-Arnavutköy	282488	450.35	627	2.030t (279\$)	24.360t (3.350\$)	41.184182	28.740729	0.0	Turkish Restaurant	Restaurant	Turkish Home Cooking Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant
2	Istanbul-Ataşehir	425094	25.23	16849	6.577t (904\$)	78.924t (10.854\$)	40.984749	29.106720	2.0	Restaurant	Seafood Restaurant	Kebab Restaurant	Italian Restaurant	Turkish Home Cooking Restaurant
3	Istanbul-Avcılar	448882	42.01	10685	3.662t (503\$)	43.938t (6.064\$)	40.980135	28.717547	2.0	Restaurant	Seafood Restaurant	Fast Food Restaurant	Theme Restaurant	Kebab Restaurant
4	Istanbul-Bağcılar	745125	22.36	33324	3.197t (441\$)	38.367t (5.295\$)	41.033899	28.857898	0.0	Kebab Restaurant	Turkish Restaurant	Mediterranean Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant

Cluster labels for districts with most common types of restaurants

In this table, we see that cluster labels assigned by the k-means clustering algorithm. We can see the 5 most common types of restaurants before. Now we can see the categories(clusters) of each district. We can use a map to see clusters easily.



On this map, you will see 3 different colors of points on district centers. Each color represents a different cluster. Now we will inspect these clusters in more detail and try to give a name for each one.

Cluster 1 (Red)

	Population (2019)	Annual household income TL (USD)	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	282488	24.360₺ (3,350\$)	41.184182	28.740729	0.0	Turkish Restaurant	Restaurant	Turkish Home Cooking Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant
4	745125	38.367₺ (5,295\$)	41.033899	28.857898	0.0	Kebab Restaurant	Turkish Restaurant	Mediterranean Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant
6	229239	106.140₺ (14,650\$)	40.983541	28.867974	0.0	Turkish Restaurant	Restaurant	Fast Food Restaurant	Kebab Restaurant	Italian Restaurant
8	274735	41.762₺ (5,764\$)	41.035738	28.912260	0.0	Turkish Restaurant	Restaurant	Comfort Food Restaurant	Kebab Restaurant	Seafood Restaurant
9	182649	126.720₺ (17,490\$)	41.042847	29.007528	0.0	Restaurant	Turkish Restaurant	Falafel Restaurant	Arepa Restaurant	Bosnian Restaurant
13	254103	44.049₺ (6,079\$)	41.015691	28.595524	0.0	Turkish Restaurant	Seafood Restaurant	Restaurant	Italian Restaurant	Turkish Home Cooking Restaurant
14	73718	25.536₺ (3,524\$)	41.143563	28.461969	0.0	Turkish Restaurant	Restaurant	Turkish Home Cooking Restaurant	Fast Food Restaurant	Kebab Restaurant
15	264508	42.033₺ (5,801\$)	41.040210	29.175059	0.0	Turkish Restaurant	Restaurant	Doner Restaurant	Middle Eastern Restaurant	Fast Food Restaurant
16	450344	34.164₺ (4,715\$)	41.033254	28.890953	0.0	Restaurant	Turkish Restaurant	Kebab Restaurant	Turkish Home Cooking Restaurant	Tantuni Restaurant
18	400513	56.044₺ (7,735\$)	41.046044	28.925324	0.0	Turkish Restaurant	Restaurant	Kebab Restaurant	Vegetarian / Vegan Restaurant	Fast Food Restaurant
19	443090	63.378₺ (8,747\$)	41.009633	28.965165	0.0	Turkish Restaurant	Restaurant	Kebab Restaurant	Italian Restaurant	Turkish Home Cooking Restaurant
20	491962	36.228₺ (5,000\$)	41.057526	28.915650	0.0	Turkish Restaurant	Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant
25	792821	42.804₺ (5,908\$)	41.000214	28.780889	0.0	Turkish Restaurant	Seafood Restaurant	Restaurant	Greek Restaurant	Doner Restaurant
26	513316	69.259₺ (9,559\$)	40.923542	29.132836	0.0	Turkish Restaurant	Seafood Restaurant	Restaurant	Gluten-free Restaurant	Fast Food Restaurant
31	336021	26.064₺ (3,597\$)	40.968087	29.258789	0.0	Turkish Restaurant	Restaurant	Kebab Restaurant	Theme Restaurant	Seafood Restaurant
32	534565	26.244₺ (3,622\$)	41.109240	28.882614	0.0	Turkish Restaurant	Middle Eastern Restaurant	Tibetan Restaurant	Fast Food Restaurant	Restaurant
36	710280	43.641₺ (6,023\$)	41.022269	29.090073	0.0	Turkish Restaurant	Restaurant	Arepa Restaurant	Kebab Restaurant	Kokoreç Restaurant
37	531825	83.839₺ (11,572\$)	41.024905	29.017528	0.0	Turkish Restaurant	Restaurant	Turkish Home Cooking Restaurant	Fast Food Restaurant	Kokoreç Restaurant
38	293574	6.036₺ (6,036\$)	40.988118	28.903635	0.0	Restaurant	Turkish Restaurant	Doner Restaurant	Turkish Home Cooking Restaurant	Fast Food Restaurant

Most of the districts labeled as cluster 1(0 in code and table) have Turkish Restaurants as the most common restaurant type. So we can assume that this cluster represents **Turkish Restaurants**.

Cluster 2 (Purple)

	Population (2019)	Annual household income TL(USD)	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	15238	79.821k (10,978\$)	40.876259	29.091027	1.0	Seafood Restaurant	Restaurant	Turkish Restaurant	Fast Food Restaurant	Mediterranean Restaurant
29	347214	87.696k (12,104\$)	41.168592	29.056649	1.0	Seafood Restaurant	Turkish Restaurant	Kebab Restaurant	American Restaurant	Kokoreç Restaurant
30	193680	28.464k (3,928\$)	41.073204	28.246922	1.0	Seafood Restaurant	Turkish Restaurant	Kokoreç Restaurant	Doner Restaurant	Turkish Home Cooking Restaurant
33	37692	29.789k (4,115\$)	41.174407	29.612522	1.0	Seafood Restaurant	Restaurant	Turkish Home Cooking Restaurant	Fast Food Restaurant	Comfort Food Restaurant
35	267400	40884k (5,643\$)	40.816667	29.300833	1.0	Seafood Restaurant	Restaurant	Kebab Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant

All of the districts labeled as cluster 2(1 in code and table) have Seafood Restaurants as the most common restaurant type. So we can assume that this cluster represents **Seafood Restaurants**.

Cluster 3 (Green)

	Population (2019)	Annual household income TL(USD)	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	425094	78.924k (10,854\$)	40.984749	29.106720	2.0	Restaurant	Seafood Restaurant	Kebab Restaurant	Italian Restaurant	Turkish Home Cooking Restaurant
3	448882	43.938k (6,064\$)	40.980135	28.717547	2.0	Restaurant	Seafood Restaurant	Fast Food Restaurant	Theme Restaurant	Kebab Restaurant
5	611059	56.088k (7,741\$)	41.000290	28.863745	2.0	Seafood Restaurant	Italian Restaurant	Turkish Restaurant	Turkish Home Cooking Restaurant	Caribbean Restaurant
7	460259	54.152k (7,474\$)	41.097693	28.806163	2.0	Kebab Restaurant	Fast Food Restaurant	Restaurant	Middle Eastern Restaurant	Turkish Restaurant
10	248260	44.316k (6,116\$)	41.123936	29.108315	2.0	Seafood Restaurant	Restaurant	Comfort Food Restaurant	Turkish Restaurant	Fast Food Restaurant
11	352412	51.924k (7,166\$)	41.001026	28.641984	2.0	Restaurant	Turkish Restaurant	Doner Restaurant	Fast Food Restaurant	Kokoreç Restaurant
12	233323	57.275k (7,905\$)	41.028423	28.973681	2.0	Restaurant	Vegetarian / Vegan Restaurant	Caucasian Restaurant	Greek Restaurant	Turkish Restaurant
21	289441	40.656k (5,611\$)	41.022871	28.872629	2.0	Turkish Restaurant	Restaurant	Kebab Restaurant	Seafood Restaurant	Comfort Food Restaurant
22	482713	108.300k (14,948\$)	40.991572	29.027017	2.0	Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	Doner Restaurant	Turkish Home Cooking Restaurant
23	448025	50.260k (6,937\$)	41.085980	28.972621	2.0	Turkish Restaurant	Kebab Restaurant	Italian Restaurant	Seafood Restaurant	Doner Restaurant
24	470676	49.443k (6,824\$)	40.888492	29.189602	2.0	Turkish Restaurant	Seafood Restaurant	Restaurant	Doner Restaurant	Fast Food Restaurant
27	711894	36.664k (5,060\$)	40.876589	29.233342	2.0	Turkish Restaurant	Fast Food Restaurant	Seafood Restaurant	Restaurant	Eastern European Restaurant
28	436733	31.602k (4,361\$)	40.996887	29.236601	2.0	Turkish Restaurant	Comfort Food Restaurant	Restaurant	Kebab Restaurant	Seafood Restaurant
34	279817	93.864k (12,955\$)	41.061273	28.985020	2.0	Italian Restaurant	Restaurant	Turkish Restaurant	Mediterranean Restaurant	Doner Restaurant

I am having a hard time commenting on this cluster. Because there are some Turkish, Kebab, Seafood, and one Italian Restaurant at 1st most common column. It looks like a mix of all kinds of restaurants. So with a lazy approach, we can say that this cluster represents “others”.

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 new to me and I tried to learn more with documentation.

Analyzing that is almost free, you only need the data and a computer. In the end, I wish I had more clear and larger data about the other cities in Turkey. I hope we can see more data in the coming years.

CONCLUSION

In the end, I reached the goal that I declared in the first section. I think with the map on the Results section a tourist can see a simple guide about restaurants in different districts. But the number of clusters can be determined with a more systematic way to improve results. And can make further analysis of the restaurants labeled just as “Restaurant” and make manual labels for them can increase the efficiency. In total, I hope you enjoyed these results of the capstone project.