

# The Battle of Neighborhoods

## *Sri Lankan Restaurants*

Danuka Malinda

### INTRODUCTION

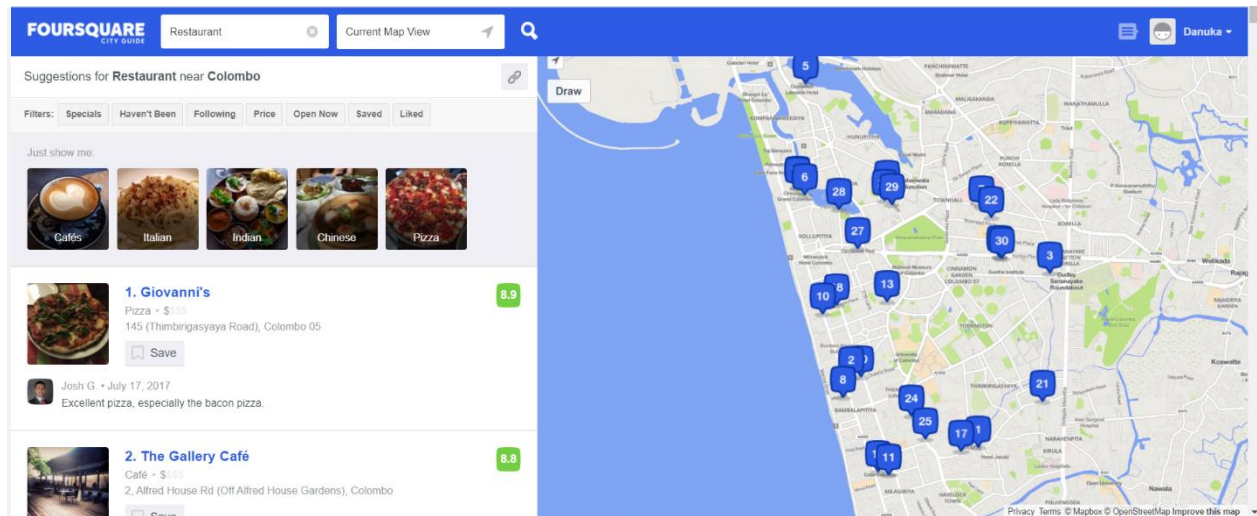
Sri Lanka is one of the most attractive islands in the world. There are 25 districts, and the most populated districts are Colombo and Gampaha. Every district has beautiful places visit and see, therefore the tourism of Sri Lanka is in very high place.

Tourists or local travelers facing challenges to find best place to have meals with good taste and quality. Most of the time, before they visit to those places, they consider about their meals. Restaurants types also consider as a factor.

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

### DATA

I will use Foursquare data about restaurants in Sri Lanka. 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 Sri Lanka

Also, I will use the Wikipedia page of Districts of Sri Lanka to easy-access the table of the districts: [https://en.wikipedia.org/wiki/Districts\\_of\\_Sri\\_Lanka](https://en.wikipedia.org/wiki/Districts_of_Sri_Lanka)

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.

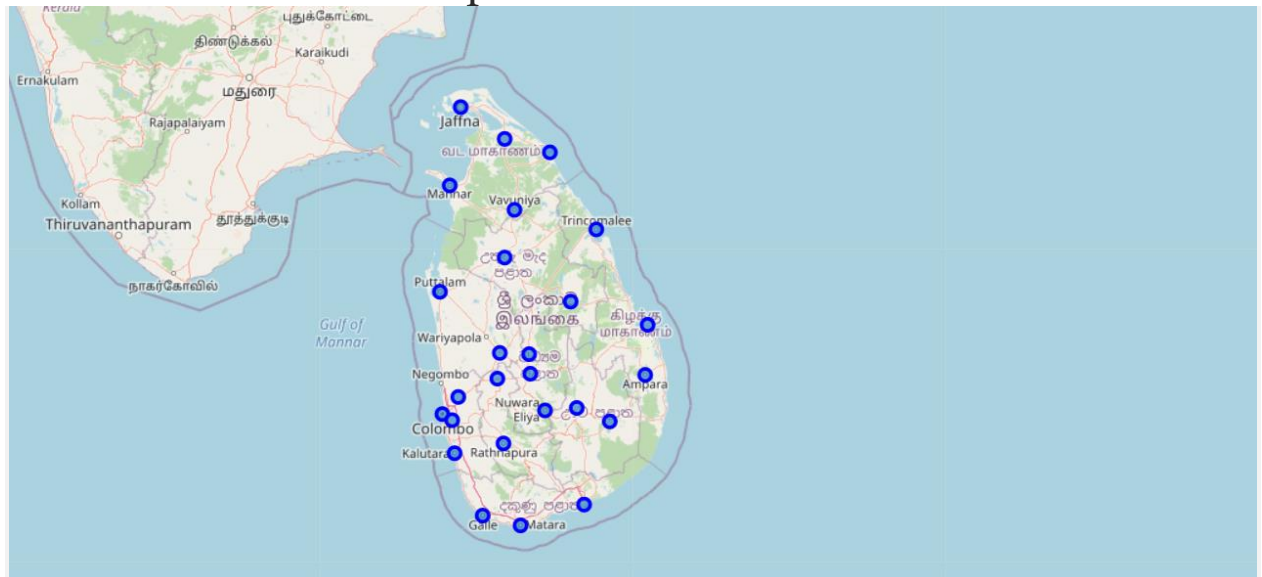
First of all, I started with a little web scraping to get districts of Sri Lanka. To do that I used pandas HTML read function and after a little cleaning and adjustments, I get this data frame.

	District	Area map	Province	Districtcapital	Landareain km2 (mi2)[24]	Inlandwaterareain km2 (mi2)[24]	Totalareain km2 (mi2)[24]	Population(2012)[25]	Populationdensityper km2(per mi2)[a]
0	Ampara	NaN	Eastern	Ampara	4,222 (1,630)	193 (75)	4,415 (1,705)	649402	154 (400)
1	Anuradhapura	NaN	North Central	Anuradhapura	6,664 (2,573)	515 (199)	7,179 (2,772)	860575	129 (330)
2	Badulla	NaN	Uva	Badulla	2,827 (1,092)	34 (13)	2,861 (1,105)	815405	288 (750)
3	Batticaloa	NaN	Eastern	Batticaloa	2,610 (1,010)	244 (94)	2,854 (1,102)	526567	202 (520)
4	Colombo	NaN	Western	Colombo	676 (261)	23 (8.9)	699 (270)	2324349	3,438 (8,900)

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.

	Neighborhood	Area map	Province	Districtcapital	Landareain km2 (mi2) [24]	Inlandwaterareain km2 (mi2) [24]	Totalareain km2 (mi2) [24]	Population(2012) [25]	Populationdensityper km2(per mi2) [a]	Latitude	Longitude
0	SriLanka-Ampara	NaN	Eastern	Ampara	4,222 (1,630)	193 (75)	4,415 (1,705)	649402	154 (400)	7.291123	81.672395
1	SriLanka-Anuradhapura	NaN	North Central	Anuradhapura	6,664 (2,573)	515 (199)	7,179 (2,772)	860575	129 (330)	8.334985	80.410610
2	SriLanka-Badulla	NaN	Uva	Badulla	2,827 (1,092)	34 (13)	2,861 (1,105)	815405	288 (750)	6.989820	81.056943
3	SriLanka-Batticaloa	NaN	Eastern	Batticaloa	2,610 (1,010)	244 (94)	2,854 (1,102)	526567	202 (520)	7.735603	81.694196
4	SriLanka-Colombo	NaN	Western	Colombo	676 (261)	23 (8.9)	699 (270)	2324349	3,438 (8,900)	6.934997	79.853846
5	SriLanka-Galle	NaN	Southern	Galle	1,617 (624)	35 (14)	1,652 (638)	1063334	658 (1,700)	6.032814	80.214955
6	SriLanka-Gampaha	NaN	Western	Gampaha	1,341 (518)	46 (18)	1,387 (536)	2304833	1,719 (4,450)	7.092560	79.995140
7	SriLanka-Hambantota	NaN	Southern	Hambantota	2,496 (964)	113 (44)	2,609 (1,007)	599903	240 (620)	6.124913	81.124256
8	SriLanka-Jaffna	NaN	Northern	Jaffna	929 (359)	96 (37)	1,025 (396)	583882	629 (1,630)	9.665093	80.009303
9	SriLanka-Kalutara	NaN	Western	Kalutara	1,576 (608)	22 (8.5)	1,598 (617)	1221948	775 (2,010)	6.583522	79.961251

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.



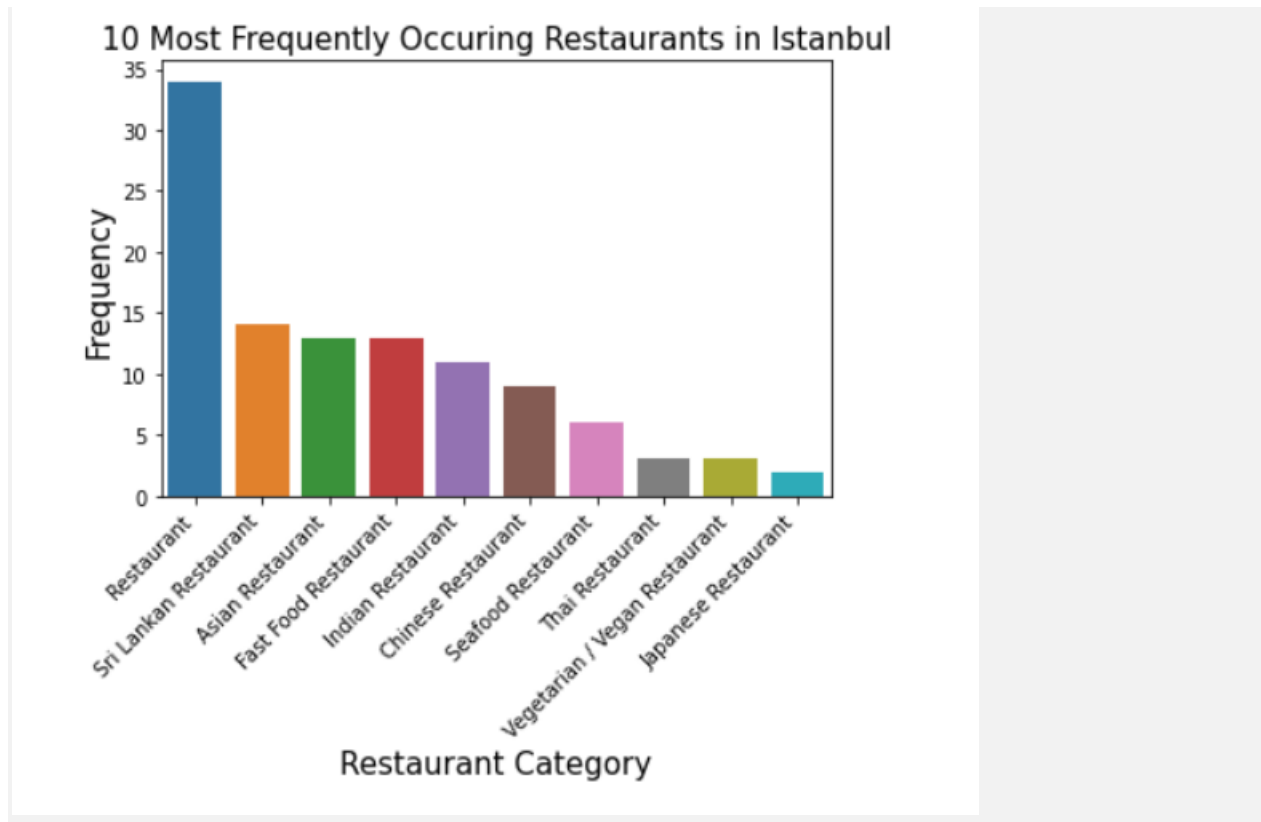
Sri Lanka map with district centers pointed

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 570 venues from 25 districts. In 570 venues there are 118 restaurants. And these 118 restaurants serve 19 different types of foods such as Sri Lanka, Chinese, Indian 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 Sri Lanka labeled as “Restaurant”. I assume that these restaurants are the generic type or labeled as just “Restaurant” by owners (not specified the type). The second one is the restaurants labeled as “SriLankan restaurents”. And the third one is the Asian restaurants. It is reasonable too because Sri Lankan people going to Asian restaurants to have their meals with variety of tastes.

	Venue_Category	Frequency
0	Restaurant	34
1	Sri Lankan Restaurant	14
2	Asian Restaurant	13
3	Fast Food Restaurant	13
4	Indian Restaurant	11
5	Chinese Restaurant	9
6	Seafood Restaurant	6
7	Thai Restaurant	3
8	Vegetarian / Vegan Restaurant	3
9	Japanese Restaurant	2



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 19 columns(unique types of restaurants in total) of 0/1's for 118 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	SriLanka-Ampara	Asian Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Chinese Restaurant	Dumpling Restaurant
1	SriLanka-Anuradhapura	Sri Lankan Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Chinese Restaurant
2	SriLanka-Badulla	Restaurant	Asian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant
3	SriLanka-Batticaloa	Indian Restaurant	Restaurant	Halal Restaurant	Asian Restaurant	Chinese Restaurant
4	SriLanka-Colombo	Seafood Restaurant	Indian Restaurant	Restaurant	Chinese Restaurant	Fast Food Restaurant
5	SriLanka-Galle	Restaurant	Sri Lankan Restaurant	Indian Restaurant	Asian Restaurant	Chinese Restaurant
6	SriLanka-Jaffna	Vegetarian / Vegan Restaurant	Asian Restaurant	Indian Restaurant	Sri Lankan Restaurant	American Restaurant
7	SriLanka-Kalutara	Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	German Restaurant	Asian Restaurant
8	SriLanka-Kandy	Restaurant	Fast Food Restaurant	Indian Restaurant	Chinese Restaurant	Sri Lankan Restaurant
9	SriLanka-Kegalle	Sri Lankan Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Asian Restaurant	Chinese Restaurant

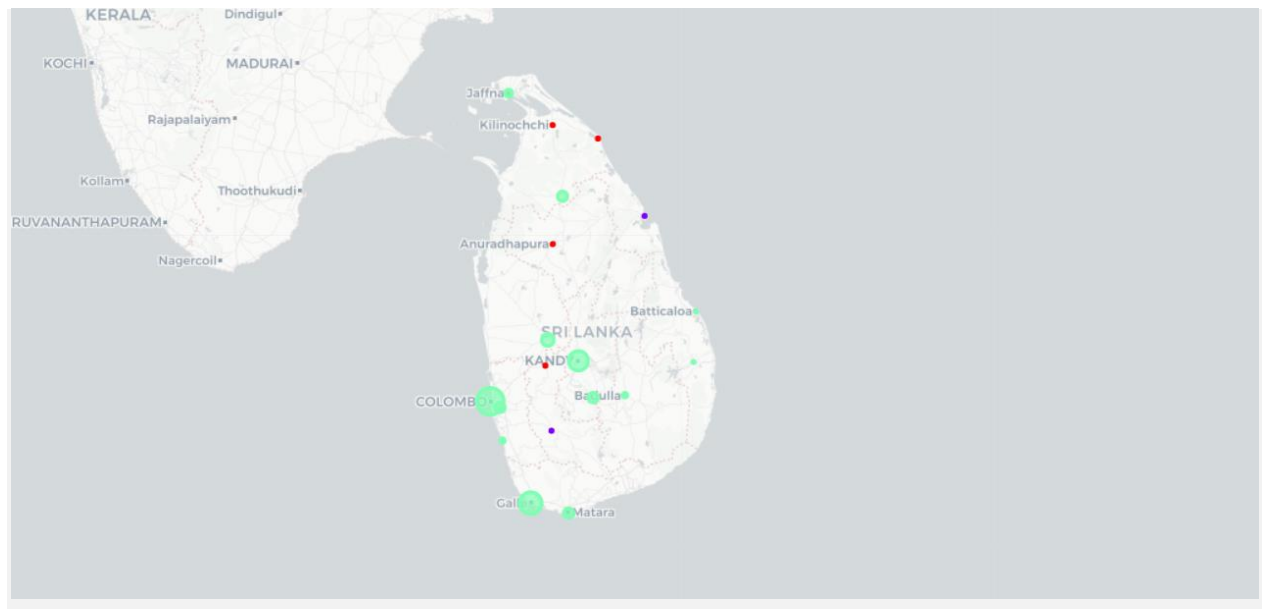
With all this data, finally, I can run a clustering algorithm to reach the goal of this project. I choose the k-means clustering algorithm from the scikit-learn package. To start the k-means algorithm I need to choose a k(num. of clusters) value. I can pick this number more systematically but with my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

## RESULTS

	Neighborhood	Area map	Province	Districtcapital	Landareain km2 (mi2) [24]	Inlandwaterareain km2 (mi2)[24]	Totalareain km2 (mi2) [24]	Population(2012) [26]	Populationdensityper km2(per mi2)[a]	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	SriLanka-Ampara	NaN	Eastern	Ampara	4,222 (1,630)	193 (75)	4,415 (1,705)	649402	154 (400)	7.291123	81.672395	2.0	Asian Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Chinese Restaurant	Dump Restau
1	SriLanka-Anuradhapura	NaN	North Central	Anuradhapura	6,664 (2,573)	515 (199)	7,179 (2,772)	860575	129 (330)	8.334985	80.410610	0.0	Sri Lankan Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Chin Restau
2	SriLanka-Badulla	NaN	Uva	Badulla	2,827 (1,092)	34 (13)	2,861 (1,105)	815405	288 (750)	6.989820	81.056943	2.0	Restaurant	Asian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	H Restau
3	SriLanka-Batticaloa	NaN	Eastern	Batticaloa	2,610 (1,010)	244 (94)	2,854 (1,102)	526567	202 (520)	7.735603	81.694196	2.0	Indian Restaurant	Restaurant	Halal Restaurant	Asian Restaurant	Chin Restau
4	SriLanka-Colombo	NaN	Western	Colombo	676 (261)	23 (8.9)	699 (270)	2324349	3,438 (8,900)	6.934997	79.853846	2.0	Seafood Restaurant	Indian Restaurant	Restaurant	Chinese Restaurant	Fast F Restau

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)

	Area map	Inlandwaterareain km2 (mi2)[24]	Totalareain km2 (mi2)[24]	Population(2012) [25]	Populationdensityper km2(per mi2)[a]	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	NaN	515 (199)	7,179 (2,772)	860575	129 (330)	8.334985	80.410610	0.0	Sri Lankan Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Chinese Restaurant
11	NaN	8 (3.1)	1,693 (654)	840648	499 (1,290)	7.253201	80.345413	0.0	Sri Lankan Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Asian Restaurant	Chinese Restaurant
12	NaN	74 (29)	1,279 (494)	113510	94 (240)	9.384007	80.408722	0.0	Vegetarian / Vegan Restaurant	Sri Lankan Restaurant	Halal Restaurant	Asian Restaurant	Chinese Restaurant
18	NaN	202 (78)	2,617 (1,010)	92238	38 (98)	9.269853	80.814535	0.0	Sri Lankan Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Asian Restaurant	Chinese Restaurant

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

## Cluster 2 (Purple)

	Area map	Inland water area in km2 (mi2)[24]	Total area in km2 (mi2)[24]	Population (2012) [28]	Population density per km2 (per mi2)[a]	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
22	NaN	39 (15)	3,275 (1,264)	1088007	336 (870)	6.680369	80.402298	1.0	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Asian Restaurant	Chinese Restaurant
23	NaN	198 (76)	2,727 (1,053)	379541	150 (390)	8.576425	81.234495	1.0	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Asian Restaurant	Chinese Restaurant

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

## Cluster 3 (Green)

	Area map	Inland water area in km2 (mi2)[24]	Total area in km2 (mi2)[24]	Population (2012) [28]	Population density per km2 (per mi2)[a]	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	NaN	193 (75)	4,415 (1,705)	649402	154 (400)	7.291123	81.672395	2.0	Asian Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant	Chinese Restaurant	Dumpling Restaurant
2	NaN	34 (13)	2,861 (1,105)	815405	288 (750)	6.989820	81.056943	2.0	Restaurant	Asian Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant
3	NaN	244 (94)	2,854 (1,102)	526567	202 (520)	7.735603	81.694196	2.0	Indian Restaurant	Restaurant	Halal Restaurant	Asian Restaurant	Chinese Restaurant
4	NaN	23 (8.9)	699 (270)	2324349	3,438 (8,900)	6.934997	79.853846	2.0	Seafood Restaurant	Indian Restaurant	Restaurant	Chinese Restaurant	Fast Food Restaurant
5	NaN	35 (14)	1,652 (638)	1063334	658 (1,700)	6.032814	80.214955	2.0	Restaurant	Sri Lankan Restaurant	Indian Restaurant	Asian Restaurant	Chinese Restaurant
8	NaN	96 (37)	1,025 (396)	583882	629 (1,630)	9.665093	80.009303	2.0	Vegetarian / Vegan Restaurant	Asian Restaurant	Indian Restaurant	Sri Lankan Restaurant	American Restaurant
9	NaN	22 (8.5)	1,598 (617)	1221948	775 (2,010)	6.583522	79.961251	2.0	Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	German Restaurant	Asian Restaurant
10	NaN	23 (8.9)	1,940 (750)	1375382	716 (1,850)	7.293092	80.635077	2.0	Restaurant	Fast Food Restaurant	Indian Restaurant	Chinese Restaurant	Sri Lankan Restaurant
13	NaN	192 (74)	4,816 (1,859)	1618465	350 (910)	7.487046	80.364908	2.0	Restaurant	Sri Lankan Restaurant	Asian Restaurant	Chinese Restaurant	Fast Food Restaurant
16	NaN	13 (5.0)	1,283 (495)	814048	641 (1,660)	5.947822	80.548292	2.0	Restaurant	Fast Food Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Halal Restaurant
19	NaN	35 (14)	1,741 (672)	711644	417 (1,080)	6.973886	80.767127	2.0	Sri Lankan Restaurant	Restaurant	Indian Restaurant	Italian Restaurant	Thai Restaurant
24	NaN	106 (41)	1,967 (759)	172115	92 (240)	8.759352	80.500078	2.0	Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	Seafood Restaurant
25	NaN	2,905 (1,122)	65,610 (25,330)	20359439	325 (840)	6.877829	79.934959	2.0	Chinese Restaurant	Asian Restaurant	Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant

Most of the restaurants in this category under first common venue is labeled as “Restaurants”. With this fact, we can consider these areas have common type of restaurants rather than special types.

## 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.