Best location for a Indian Restaurant in Stockholm, Sweden

IBM Data Science Capstone Project
Chaudhary Awais Salman

Introduction

While opening a restaurant can be a very lucrative business, a lot of factors cause many restaurants to close within the first year of opening. These factors could be location, competition, and quality of food.

The goal of this project is to use the Foursquare API to determine the optimal location to open a Indian Restaurant. For this problem specifically, location and competition will be determined by where the restaurant will be opened. If there are too many Indian restaurants the profitability of the restaurant will be decreased. Another factor could be starting the restaurant in a location with higher income, this could increase the profitability.

Business Problem. If the client wanted to open a Indian Restaurant in Stockholm, what areas are the best options to open the restaurant?

Data

To answer the business problem, we will use the Stockholm Census data set and the Stockholm neighborhoods data obtain by Wikipedia. The following factors must be extracted from the data sources:

1. Population & Ethnic Distribution of Each Neighborhood (Stockholm Census)
2. Income Distribution of Each Neighborhood (Stockholm Census)
3. Number of Restaurants in Each Neighborhood (Foursquare API)
4. Number of Indian Restaurants in Each Neighborhood (Foursquare API)

Methodology

The first step of the project was to combine the dataset obtained by Wikipedia and the census dataset. The datasets can be seen in the Appendix section.

Using the income distribution for each neighborhood, the spending power of each area was calculated using the median of each category weighted by the number of people in that income category. Thus, the spending power represents the overall capital of each area. Due to the spending power for each area is considerably large, it had to be standardized.

The next step was to visualize the location of the various postal codes within Stockholm to obtain a general understanding the location (Figure 1). As seen from the map, the postal codes are densely clustered near downtown Stockholm and spread out as the distance from downtown increases. This is important because while some postal codes might not have many restaurants, if the area is located near downtown, adjacent regions can heavily impact the profitability of the restaurant.

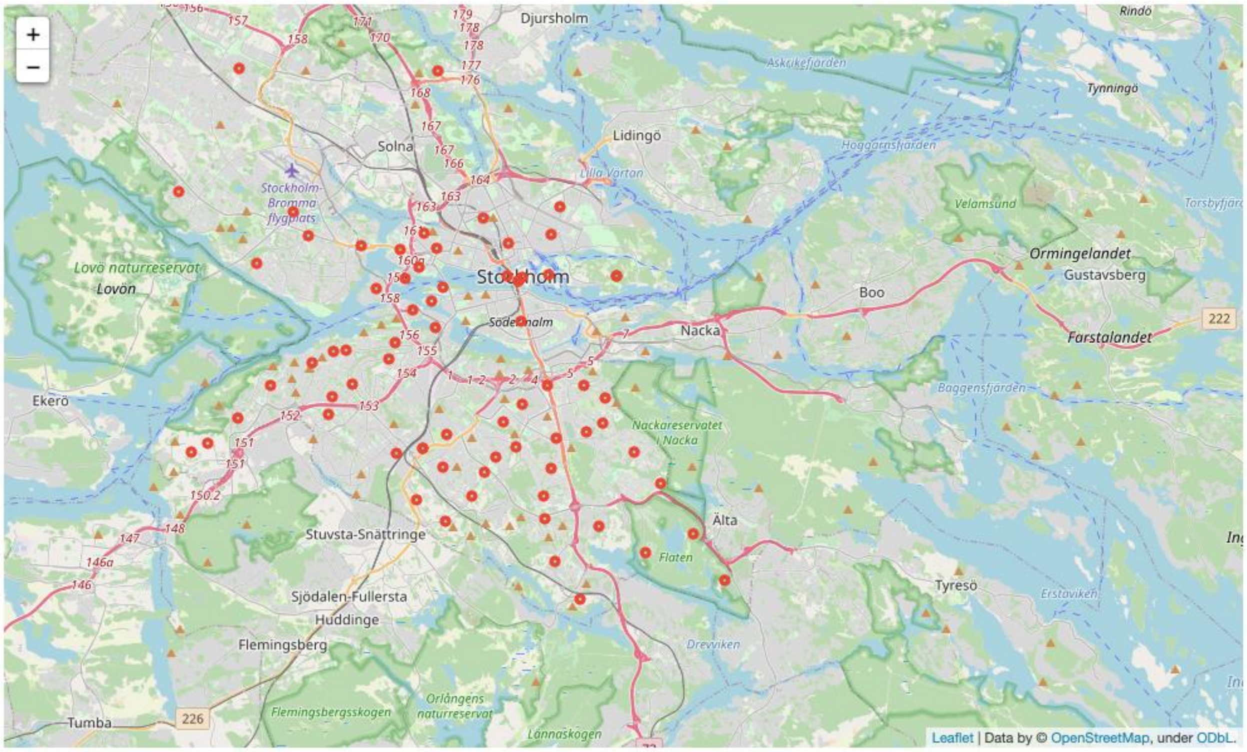


Figure 1 Location of each postal code within Stockholm

Now that the region has been clearly visualized, the Foursquare API was used to explore each neighborhood and return the top 200 venues within 1,000 meters of the longitude and latitude for each postal code. The extracted venue categories were encoded using one-hot encoding and the total restaurants and Indian restaurants in each region were calculated (Figure 2).

With the resulting data, the Postal Code, Borough name, Latitude, Longitude and Density columns of each region were dropped from the dataframe. Then, the population, area, spending power, total number of restaurants and the number of Indian restaurants were used to train a k-Means clustering algorithm with 5 clusters (Figure 3).

[28]:

	Neighborhood	Total Restaurants	Mexican Restaurants
0	Adelaide,King,Richmond	33	0
1	Agincourt	26	0
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	11	0
3	Albion Gardens,Beaumond Heights,Humbergate,Jam...	6	0
4	Alderwood,Long Branch	4	0

Figure 2 Number of restaurants in each region

Results

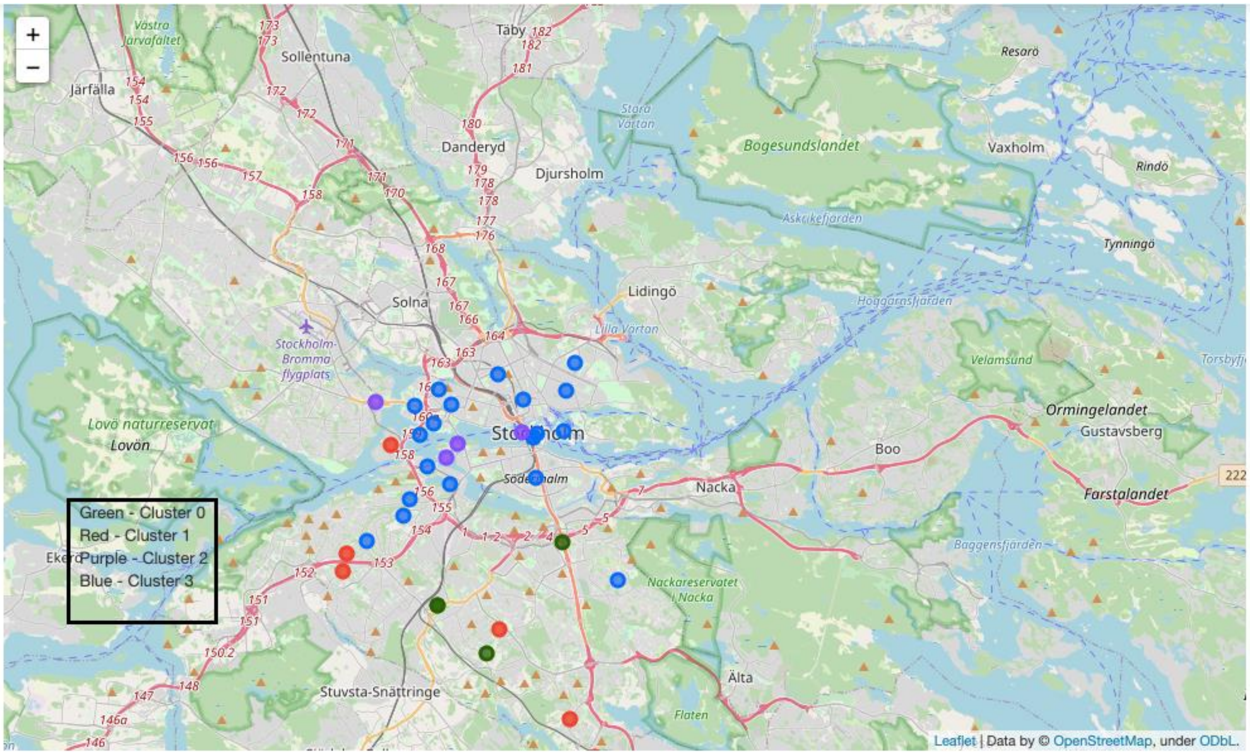
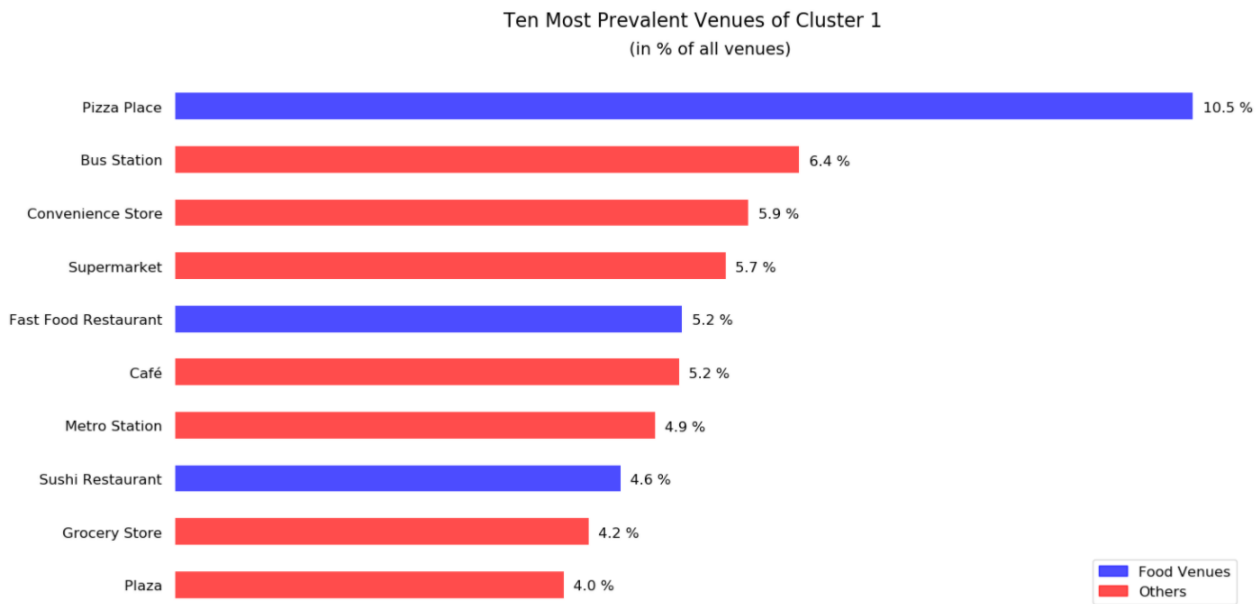
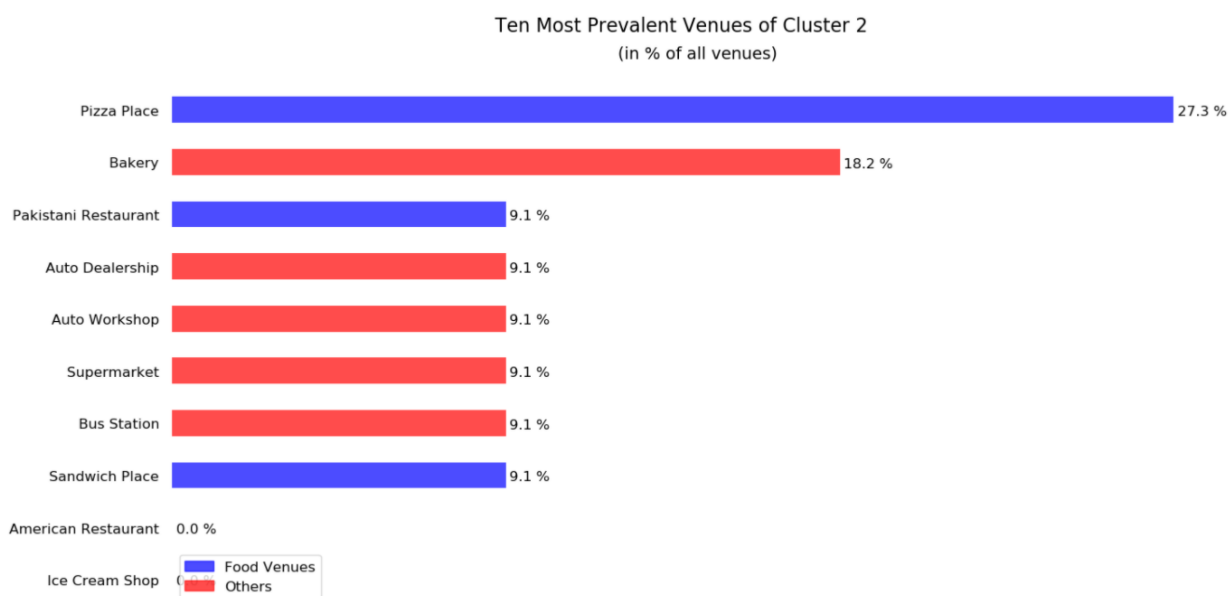


Figure 3 Map of the resulting clusters. Cluster 0= Red Cluster 1= Purple Cluster 2= Blue Cluster 3= Turquoise Cluster 4= Orange

Cluster	Characteristics
Cluster 0	Negative Spending Power (-1.1- -0.4)
Cluster 1	Positive Spending Power (0.2 - 1.8)
Cluster 2	High Positive Spending Power (1.8 - 3.8)
Cluster 3	Near Zero Spending Power (-0.4 - 0.5)
Cluster 4	Near Zero Spending Power (-0.8 – 0.2)

Table 1 Spending Power of the clusters resulting from K-Means clustering algorithm


Ten Most Prevalent Venues of Cluster 1
(in % of all venues)
Pizza Place
Bus Station
Convenience Store
Supermarket
Fast Food Restaurant
Café
Metro Station
Sushi Restaurant
Grocery Store
Plaza
10.5 %
6.4 %
5.9 %
5.7 %
5.2 %
5.2 %
4.9 %
4.6 %
4.2 %
4.0 %
Food Venues
Others


Ten Most Prevalent Venues of Cluster 2
(in % of all venues)
Pizza Place
Bakery
Pakistani Restaurant
Auto Dealership
Auto Workshop
Supermarket
Bus Station
Sandwich Place
American Restaurant
Ice Cream Shop
27.3 %
18.2 %
9.1 %
9.1 %
9.1 %
9.1 %
9.1 %
9.1 %
0.0 %
Food Venues
Others

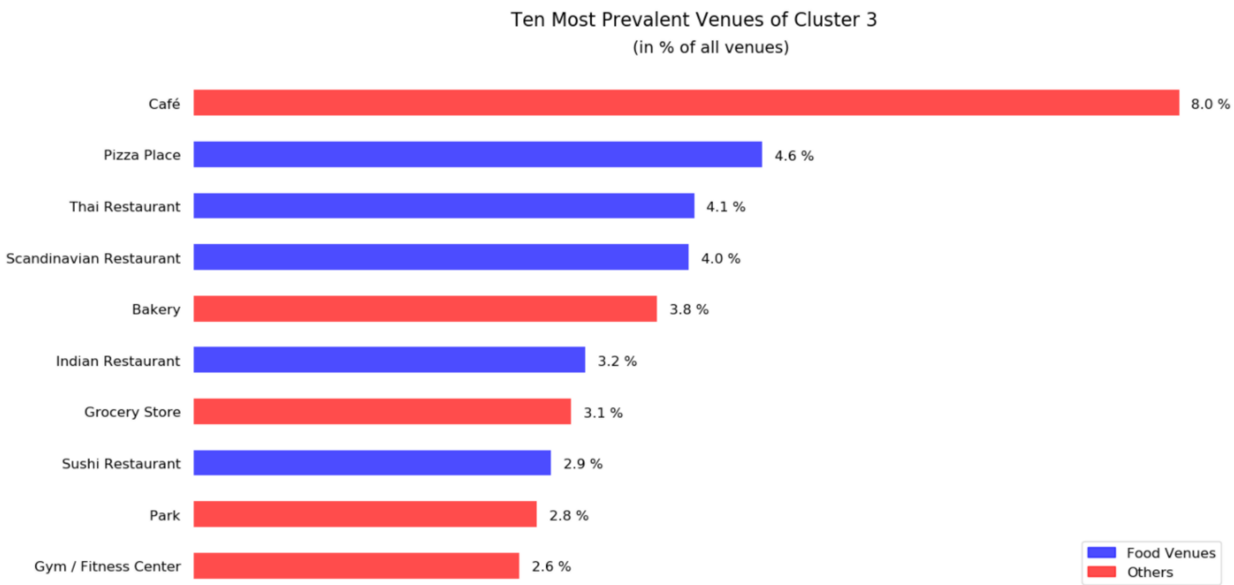


Figure 4 Characteristics of neighborhoods belonging to clusters

Discussion

From the results of the clustering algorithm, it was determined that neighborhoods corresponding to cluster 2 were the best choice for opening an Indian restaurant based on the normalized spending power and population. This narrowed down possible locations to six different areas. Using the results in Figure 5, the High Park, The Junction South region; the Cloverdale, Islington, Martin Grove, Princess region and the Harbourfront region were eliminated due to the large number of restaurants in the area.

From the three remaining regions, I would recommend that the client open his/her restaurant in either the Rouge, Malvern region or Newtoonbrook, Willowdale region. Both regions have very few restaurants and are farther away from the downtown area. Also, both regions have a good percentage of Latin American people.

	Neighbourhood	Latitude	Longitude	Distance from stockholm center (in km)
0	Stora Essingen	59.321747	17.990692	4.496574
1	Bandhagen	59.270305	18.049588	6.637193
2	Västertorp	59.291315	17.966692	7.155108
3	Fruängen	59.286468	17.964876	7.565392
4	Farsta	59.245347	18.088366	9.391358
5	Farsta strand	59.235049	18.102066	10.640258

Figure 5 Regions cluster 2

Conclusion

Opening a restaurant is a complex task that can lead to a large monetary loss if not done properly. Thus, the selection of the area would greatly increase the likelihood of the restaurant succeeding. From the project above, I demonstrated the workflow necessary for a client to determine what area the restaurant should open. For specifically, I determined that the optimal location to open a Indian restaurant in Stockholm should be either the Stora Essingen, Bandhage, Västerstorp, Farsta or Farsta strand.

Appendix

	PostCode	Borough	Neighborhood	Latitude	Longitude	Population	Density	Area	< 5k	5k - 10k	10k - 15k	15k - 20k
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	90290.0	6208.0	45.74	290.0	240.0	420.0	720.0
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	12494.0	2403.0	5.20	60.0	25.0	45.0	60.0
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	54764.0	8570.0	19.04	315.0	540.0	815.0	970.0
3	M1G	Scarborough	Woburn	43.770992	-79.216917	53485.0	4345.0	12.31	435.0	455.0	685.0	1170.0
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	29960.0	4011.0	7.47	615.0	220.0	255.0	450.0

20k - 25k	25k - 30k	30k - 35k	35k - 40k	40k - 45k	45k - 50k	50k - 60k	60k - 70k	70k - 80k	80k - 90k	90k - 100k	100k - 125k	125k - 150k	150k - 200k	> 200k	South Asian
730.0	925.0	955.0	1090.0	1055.0	1110.0	2330.0	2150.0	1930.0	1845.0	1640.0	3355.0	2315.0	2390.0	1300.0	41.64
70.0	80.0	90.0	120.0	80.0	115.0	230.0	230.0	200.0	195.0	210.0	490.0	410.0	550.0	440.0	36.14
880.0	890.0	905.0	885.0	905.0	815.0	1565.0	1360.0	1255.0	1140.0	1050.0	1970.0	1320.0	1390.0	915.0	18.74
825.0	960.0	910.0	950.0	955.0	815.0	1725.0	1405.0	1240.0	1070.0	865.0	1660.0	1030.0	855.0	430.0	40.28
370.0	475.0	465.0	520.0	495.0	530.0	935.0	845.0	765.0	615.0	575.0	1015.0	700.0	635.0	275.0	27.72

Chinese	Black	Filipino	Latin American	Arab	Southeast Asian	West Asian	Korean	Japanese	White	Spending Power
6.00	16.49	9.92	1.41	0.84	0.55	1.32	0.16	0.15	14.64	2.331712e+09
7.64	12.41	6.44	1.64	0.68	0.68	0.80	1.04	0.28	25.49	3.970375e+08
3.44	15.05	8.04	1.74	0.50	0.90	1.29	0.37	0.53	43.03	1.511462e+09
6.95	10.91	7.65	1.39	1.14	0.59	2.47	0.39	0.19	23.36	1.240412e+09
14.69	6.38	9.63	1.77	1.12	1.03	2.72	0.68	0.52	26.77	7.651875e+08

Figure 7 Dataframe used in the project. This dataframe is a combination of the two databases used during the project