IBM Data Science Capstone Project

PRIME LOCATION TO OPEN RESTAURANT

Introduction/ Business Problem

Despite the substantial risks and the tough hours, owning a restaurant business is one of the most rewarding experiences of life. Considering the massive opportunity that the F&B industry holds, opening a restaurant can be a very lucrative business. There are many different factors that can account for a restaurant's success such as location, competition and quality of the food. This is an important question that every business owner must face when choosing whether to open a restaurant or not, as well as location of the business. Also, a lack of demand causes many restaurants to close within the first year of opening.

This aim of the project is to help facilitate the process of choosing a location and opening a restaurant for a business owner. For this project we'll try to find out the best location/ areas to open an Indian restaurant in the city of Toronto. The success of the restaurant will be determined by the location and the area where it is opened. The financial gain of the restaurant will be decreased if the there are too many Indian restaurants in the local vicinity. Additionally, starting a restaurant in a location with higher income would increase the profitability of the business over starting in a poorer area.

Data

To find the solution for the proposed business problem, we'll have to collect data from various sources:

- Population & Ethnic Distribution of Each Neighborhood (Toronto Census)
- Income Distribution of Each Neighborhood (Toronto Census)
- Number of Restaurants in Each Neighborhood (Foursquare API)
- Number of Indian Restaurants in Each Neighborhood (Foursquare API)

Toronto Census Data will help us to determine the population density, ethnic distribution and income distribution of each neighborhood. Since the problem statement is to find the prime location to open Indian restaurant, we'll focus on the data corresponding to Indian ethnicity.

The Toronto Census data is extracted from https://www.toronto.ca/city-government/dataresearch-maps/open-data/open-data-catalogue/#8c732154-5012-9afe-d0cd-ba3ffc813d5a.

Methodology

The first step of the project was to combine the Toronto dataset, containing the postal code, borough, neighborhood name, latitude and longitude for each postal code in Toronto, and the census dataset.

Toronto Dataset:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront, Regent Park	43.654260	-79.360636
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494

Toronto Census Dataset:

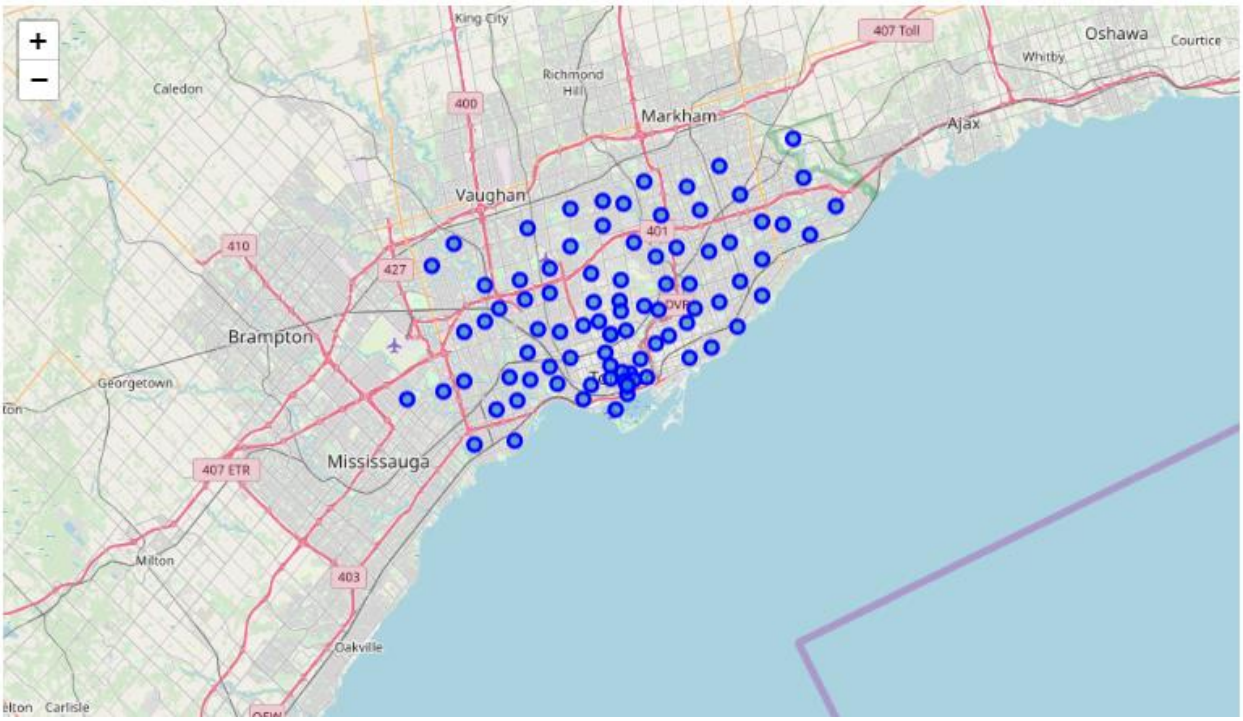
	PostalCode	Borough	Neighbourhood Number	Population, 2016	Population density per square kilometre	Land area in square kilometres	Total - Household total income groups in 2015	Under $5,000	5, 000to 9,999	10, 000to 14,999	15, 000to 19,999	20, 000to 24,999	25, 000to 29,999	30, 000to 34,999
0	M1B	Scarborough	263.0	90290.0	6208.0	45.74	26825.0	290.0	240.0	420.0	720.0	730.0	925.0	955.0
1	M1C	Scarborough	134.0	12494.0	2403.0	5.20	3700.0	60.0	25.0	45.0	60.0	70.0	80.0	90.0
2	M1E	Scarborough	411.0	54764.0	8570.0	19.04	19855.0	315.0	540.0	815.0	970.0	880.0	890.0	905.0
3	M1G	Scarborough	137.0	53485.0	4345.0	12.31	18445.0	435.0	455.0	685.0	1170.0	825.0	960.0	910.0
4	M1H	Scarborough	127.0	29960.0	4011.0	7.47	10765.0	615.0	220.0	255.0	450.0	370.0	475.0	465.0
5	M1J	Scarborough	139.0	16724.0	5395.0	3.10	5920.0	105.0	180.0	305.0	330.0	325.0	345.0	370.0

Merged Dataset:

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Population	Density	Area	< 5k	5k - 10k	10k - 15k	15k - 20k	20k - 25k	25k - 30k	30k - 35k
0	M3A	North York	Parkwoods	43.753259	-79.329656	34805.0	4691.0	7.42	345.0	185.0	315.0	520.0	590.0	600.0	615.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	17510.0	3710.0	4.72	180.0	205.0	280.0	540.0	420.0	385.0	400.0
2	M5A	Downtown Toronto	Harbourfront, Regent Park	43.654260	-79.360636	76716.0	25823.0	8.01	1975.0	1180.0	1535.0	1695.0	1520.0	1440.0	1370.0
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763	6577.0	3614.0	1.82	50.0	60.0	160.0	190.0	185.0	170.0	190.0
4	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	90290.0	6208.0	45.74	290.0	240.0	420.0	720.0	730.0	925.0	955.0

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 (i.e. total income of the inhabitants). Since the spending power for each area is considerably large and the relative strength is difficult to visualize, the spending power for each area was standardized.

The next step was to visualize the location of the various postal codes within Toronto to obtain a general understanding the location.



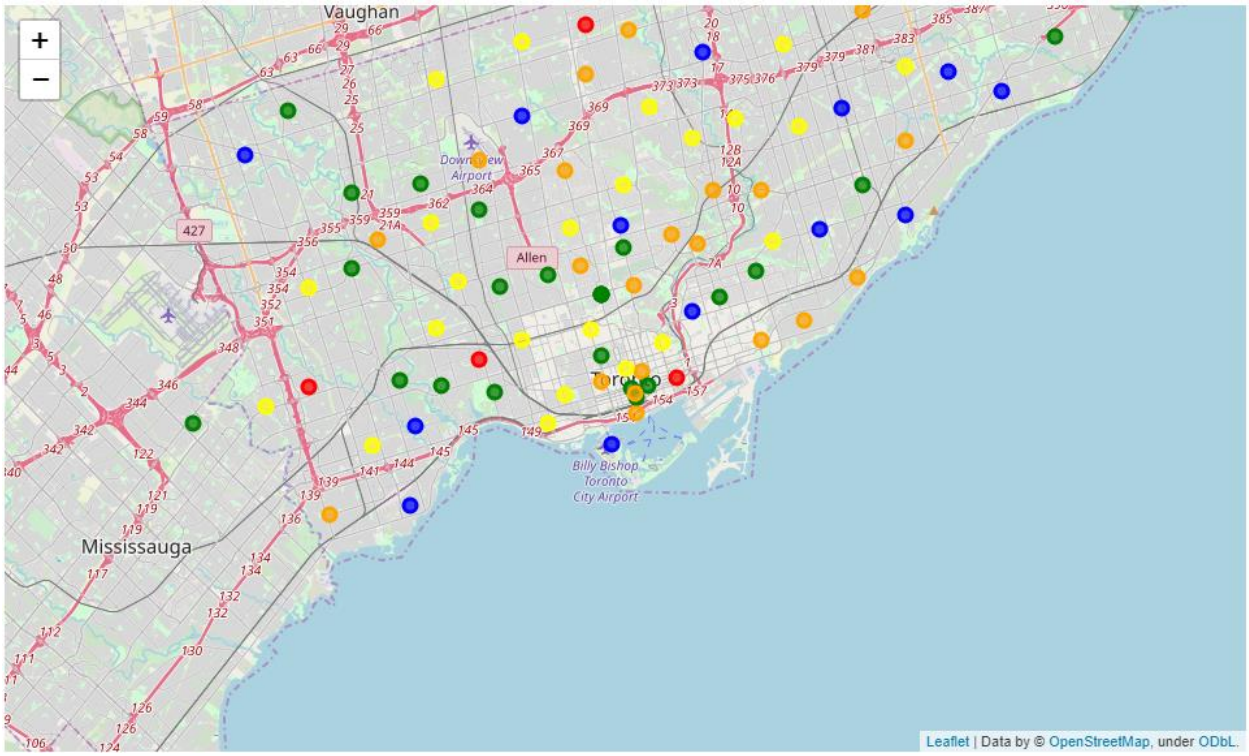
As seen from the map, the postal codes are densely clustered near downtown Toronto and spread out as the distance from downtown increases. This is important because while some postal codes might not have many restaurants, if the area is located near downtown, adjacent regions can heavily impact the profitability of the restaurant.

Now that the region has been clearly visualized, the Foursquare API was used to explore each neighborhood and return the top 200 venues within 2,000 meters (1.2 miles) 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.

	Neighborhood	Total Restaurants	Indian Restaurants
0	Adelaide, King, Richmond	16	0
1	Agincourt	32	1
2	Agincourt North, L'Amoreaux East, Milliken, St...	27	2
3	Albion Gardens, Beaumond Heights, Humbergate, ...	17	6
4	Alderwood, Long Branch	18	0
5	Bathurst Manor, Downsview North, Wilson Heights	22	0
6	Bayview Village	24	1
7	Bedford Park, Lawrence Manor East	17	1
8	Birch Cliff, Cliffside West	18	0
9	Bloordale Gardens, Eringate, Markland Wood, Ol...	13	0
10	Brockton, Exhibition Place, Parkdale Village	11	0

With the resulting data, the Postal Code, Borough name, Latitude, Longitude and Density columns of each region were dropped from the data frame. 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.

Results after clustering:



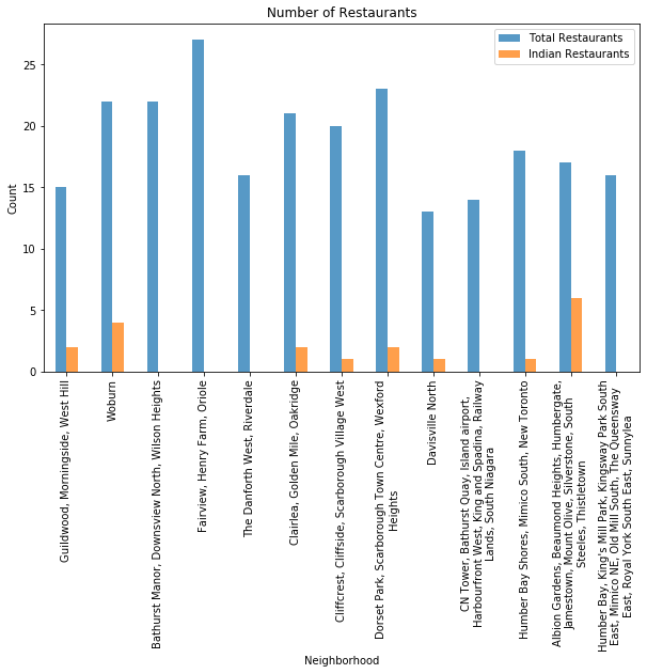
CLUSTER	SPENDING POWER RANGE		CHARACTERISTICS
Cluster 1	0.221065	-0.829808	Positive Spending Power
Cluster 2	3.85978	1.7253	High Spending Power
Cluster 3	1.8576	0.250744	Positive Spending Power
Cluster 4	-0.412131	-1.11835	Negative Spending Power
Cluster 5	0.63426	-0.483407	Near Zero Spending Power

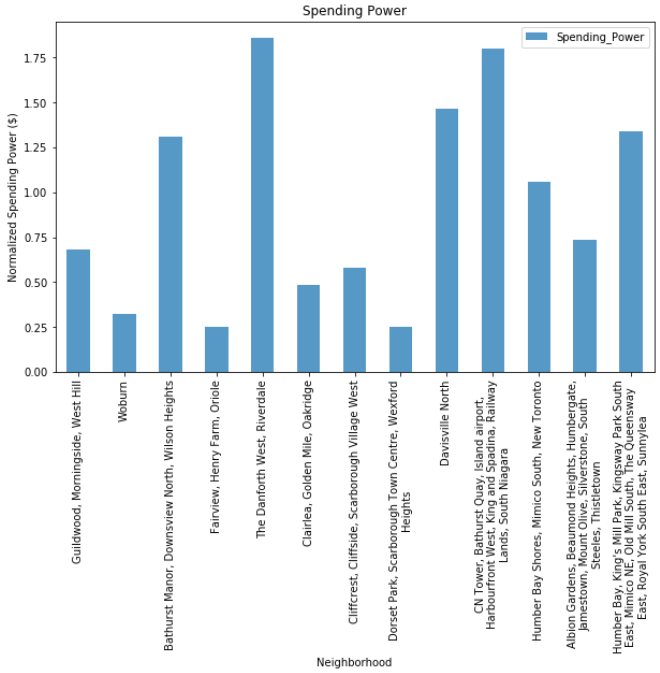
Cluster 3 is the favorable choice here, because of the Positive Spending Power and the number of Indian Restaurants in the vicinity.

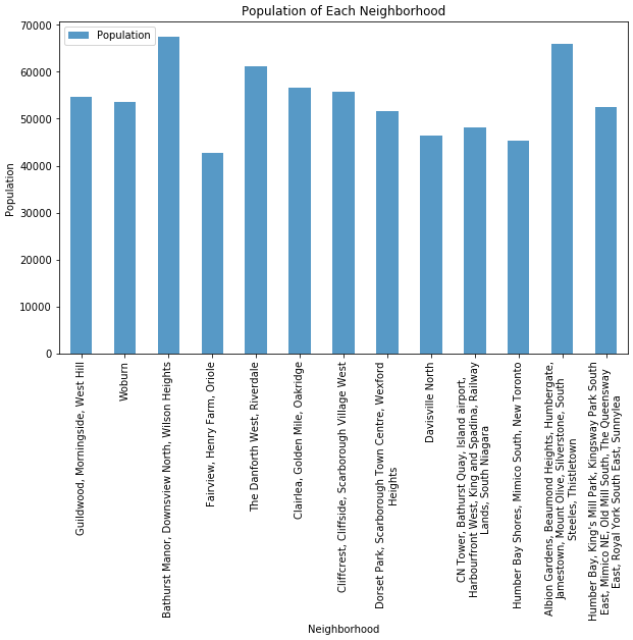
Cluster 3 Data

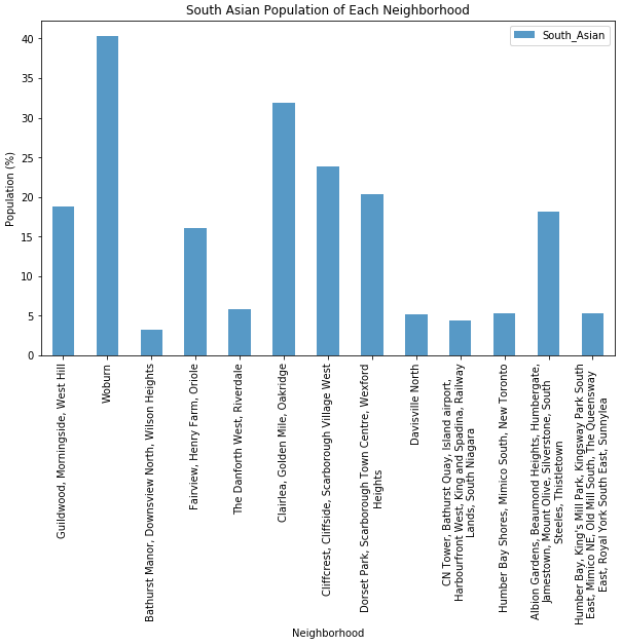
	Borough	Cluster Labels	Population	Area	South_Asian	Spending_Power	Total Restaurants	Indian Restaurants
15	Scarborough	2	54764.0	19.04	18.74	0.680704	15	2
18	Scarborough	2	53485.0	12.31	40.28	0.320774	22	4
23	North York	2	67397.0	23.84	3.28	1.311792	22	0
28	North York	2	42774.0	6.80	16.04	0.250744	27	0
36	East Toronto	2	61173.0	13.31	5.88	1.857595	16	0
39	Scarborough	2	56512.0	10.30	31.84	0.482497	21	2
44	Scarborough	2	55834.0	13.83	23.88	0.578090	20	1
56	Scarborough	2	51575.0	15.42	20.29	0.252819	23	2
57	Central Toronto	2	46433.0	4.45	5.22	1.462559	13	1
77	Downtown Toronto	2	48217.0	6.28	4.46	1.797922	14	0
78	Etobicoke	2	45427.0	10.34	5.32	1.060635	18	1
79	Etobicoke	2	65924.0	14.75	18.08	0.736360	17	6
86	Etobicoke	2	52434.0	15.10	5.34	1.336923	16	0

Cluster 3 characteristics:









Discussion

From the results of the clustering algorithm, it was determined that neighborhoods corresponding to cluster 3 were the best choice for opening an Indian restaurant based on the normalized spending power and population.

This narrowed down possible locations to 5 different areas.

Neighborhoods	Indian Population	Spending Power	Indian Restaurants
Clairlea, Golden Mile, Oakridge	31.84	0.482497	2
Cliffcrest, Cliffside, Scarborough Village West	23.88	0.57809	1
Dorset Park, Scarborough Town Centre, Wexford ...	20.29	0.252819	2
Guildwood, Morningside, West Hill	18.74	0.680704	2
Fairview, Henry Farm, Oriole	16.04	0.250744	0

From the analysis the mentioned Neighborhoods in the above table will be the best locations to open a restaurant based on Indian population, Spending Power and Indian Restaurants.

Conclusion

Opening a restaurant is a complex task that can lead to a large monetary loss if not done properly. Thus, extensive research about the area would greatly increase the likelihood of the restaurant succeeding.