



THE BATTLE OF NEIGHBORHOODS

Project Report

Sankar Ramamurthy
Aug-2020

Table of Contents

Table of Contents	1
Introduction	2
Business Problem	2
Target Audience	2
Data Acquisition	3
Methodology.....	3
Data pre-processing and cleansing	4
Visualizing the Neighborhood data.....	4
Obtaining the Venue Data	5
Analysis	5
Analyze the Venues in the Neighborhood	5
Analyze the Neighborhoods.....	6
Cluster the Neighborhoods.....	6
Analysis of each Cluster	9
Cluster 1 (Red).....	10
Cluster 2 (Purple)	10
Cluster 3 (Turquoise).....	11
Cluster 4 (Dark Khakhi)	11
Results and Discussions	11
Conclusion.....	12

Introduction



Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571 in 2016 it is the most populous city in Canada and the fourth most populous city in North America. Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Toronto encompasses a geographical area formerly administered by many separate municipalities. These municipalities have each developed a distinct history and identity over the years, and their names remain in common use among Torontonians. Throughout the city there exist hundreds of small neighborhoods and some larger neighborhoods covering a few square kilometers.

Having such a diverse mix of people and culture, there exists tough competition the restaurant business in Toronto. Therefore, it becomes a challenging task for hoteliers to decide the optimal location to start a new restaurant in a specific neighborhood.

Business Problem

This project uses data science methodologies to find an optimal location for starting a new restaurant in Toronto, Canada. It tries to identify the most suitable neighborhood, by taking into consideration the competitors, cuisine provided by other restaurants, proximity to the city center etc.

In summary, this project aims at analyzing the restaurants and their cuisines in various neighborhoods of Toronto and figure out the best possible neighborhood to start a new restaurant.

Target Audience

This project is beneficial to anyone who is planning to start a new restaurant business in Toronto. The project can also benefit already running restaurants to shift to some other cuisine for better prospects.

Data Acquisition

The data requirements for this project are as follows:

1. The Neighborhoods details of Toronto shall be obtained from Wikipedia:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Data scraping was done from the above web page and stored into a Pandas dataframe:

	PostalCode	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

2. The geographical location data using Geocoder Package: https://cocl.us/Geospatial_data

This data is in CSV format. The CSV file was read and stored into a Pandas dataframe:

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

3. All the existing venues in the neighborhood along with their category: obtained using the FourSquare API

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

Methodology

The following Data Science and Machine Learnings are applied:

- Data pre-processing & Cleansing
- Data Visualization in geographical maps using Folium

- Transform the Categorical data into Numerical Data for performing ML algorithms by using One-hot-encoding and K-Means Clustering
- Determine the optimum number of clusters using Elbow point

Data pre-processing and cleansing

The following data cleansing activities were performed:

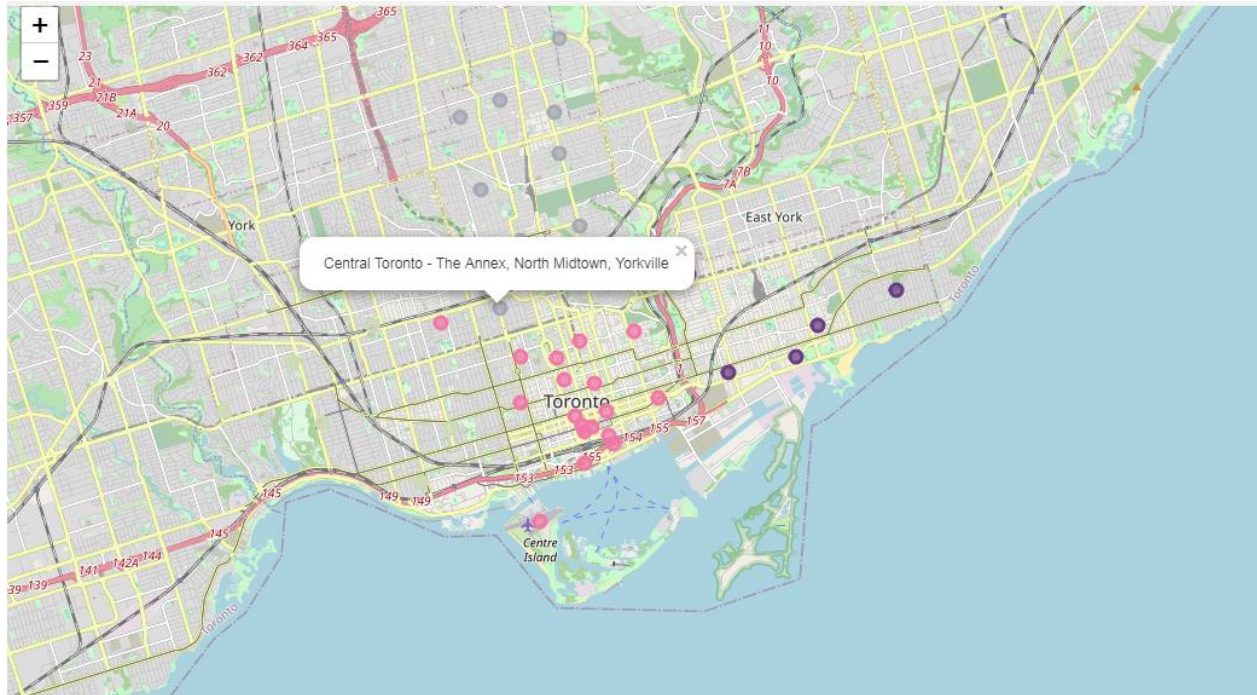
1. Data that had a Borough "Not Assigned" were removed
2. Postal code areas that had more than one neighborhood were combined into a single row with the Neighborhood names separated by a comma
3. If the Neighborhood name of a Borough is "Not Assigned", then it will be given the same name as the Borough

After the data was cleaned as per the above steps, it was merged with the co-ordinate dataframe that was populated using the Geocoder package. The resultant dataframe was then filtered to contain only the rows that had Toronto in their Borough.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
19	M4E	East Toronto	The Beaches	43.676357	-79.293031

Visualizing the Neighborhood data

The Neighborhood data was then visualized using a Folium map.



Obtaining the Venue Data

The Venue data was pulled using the FourSquare API. The top 100 venues around a radius of 500 meters for each Neighborhood, were obtained.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

Analysis

The data analysis was done on the Venues in the Neighborhoods. The Machine learning algorithms were then applied on the data and observations were derived from it.

Analyze the Venues in the Neighborhood

The venues were first grouped so as to obtain the count of venues in each Neighborhood.

There were a total of 234 unique venue categories. E.g. Coffee shop, Hotel, Theatre, Indian/Italian/<Other cuisine> Restaurants, Banks etc.

Analyze the Neighborhoods

The venues were categorized using **One-hot-encoding** method

The resulting data was then grouped by Neighborhood, by taking the mean of the frequency of occurrence of each venue category

	Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theme Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Wine Bar	V
0	Berczy Park	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.017241	0.0	0.0	0.000000	
1	Brockton, Parkdale Village, Exhibition Place	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	
2	Business reply mail Processing Centre, South C...	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	
3	CN Tower, King and Spadina, Railway Lands, Har...	0.0	0.0625	0.0625	0.0625	0.125	0.125	0.125	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	
4	Central Bay Street	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.015385	0.0	0.0	0.015385	

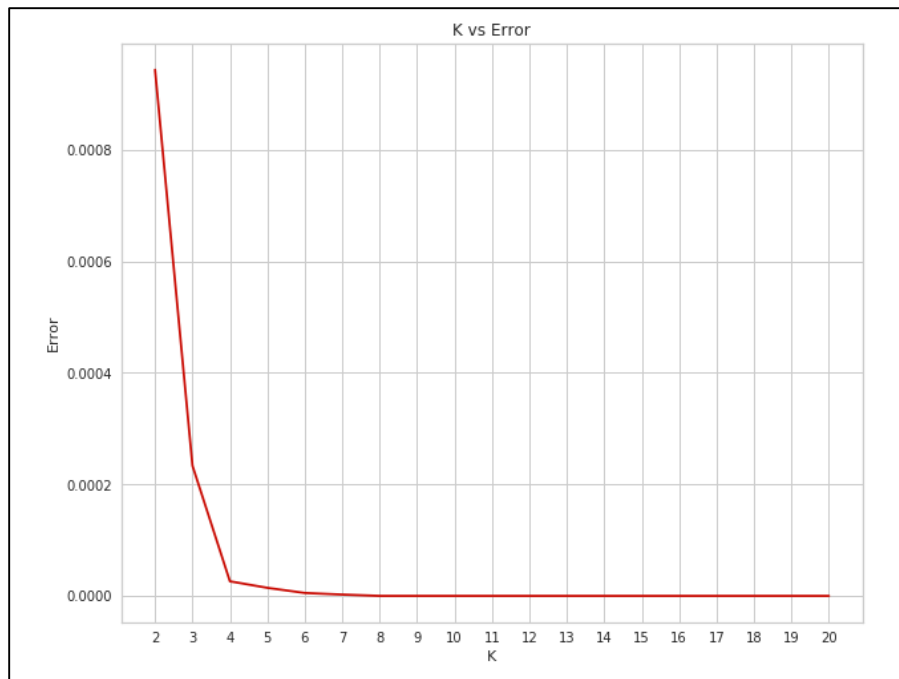
5 rows x 235 columns

The data pertaining to Indian restaurants were then pulled out. This allowed the data to be summarized based on each individual Neighborhood and made the data much simpler to analyze.

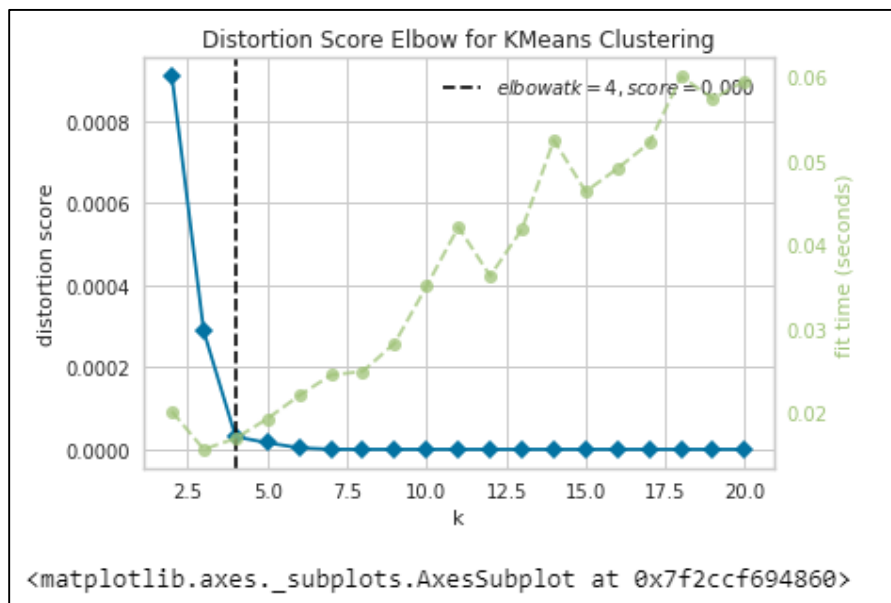
	Neighborhood	Indian Restaurant
0	Berczy Park	0.000000
1	Brockton, Parkdale Village, Exhibition Place	0.000000
2	Business reply mail Processing Centre, South C...	0.000000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000
4	Central Bay Street	0.015385

Cluster the Neighborhoods

The next step is to cluster the neighborhoods that has similar averages of Indian restaurants in that neighborhood. The **K-Means clustering** methodology. The **Elbow point** technique was used to obtain the optimum K value that was neither overfitting or under fitting the model. This test was ran with different number of K values and the best value was chosen at the point in which the line has the sharpest turn. In this case, the Elbow point was observed at K = 4. This means there will be a total of 4 clusters.



The KElbowVisualizer from the YellowBrick package was fitted over the model in order to calculate the distortion score.



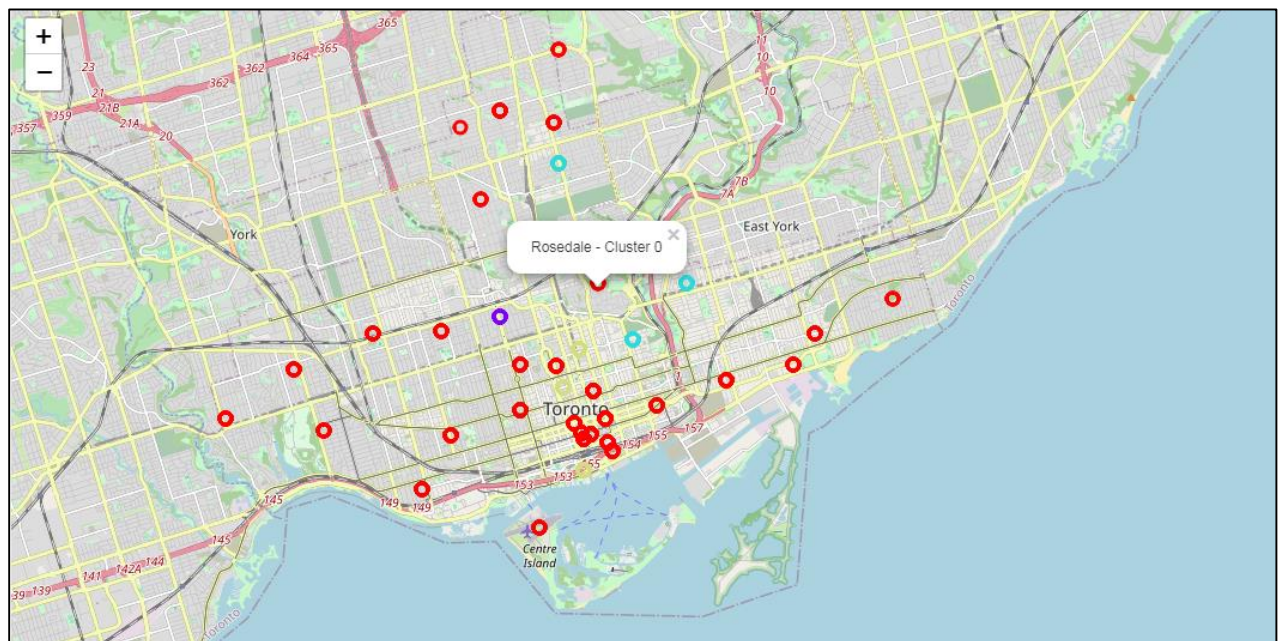
The Neighborhoods that had a similar mean frequency of Indian Restaurants were divided into 4 clusters. Each of these clusters was labelled from 0 to 3 as the indexing of labels begins with 0 instead of 1.

	Neighborhood	Indian Restaurant	Cluster Labels
0	Berczy Park	0.000000	0
1	Brockton, Parkdale Village, Exhibition Place	0.000000	0
2	Business reply mail Processing Centre, South C...	0.000000	0
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0
4	Central Bay Street	0.015385	3

The venue data is now merged with the above table to create a new table that would be the basis for analyzing the opportunities for opening a new Indian restaurant in Toronto.

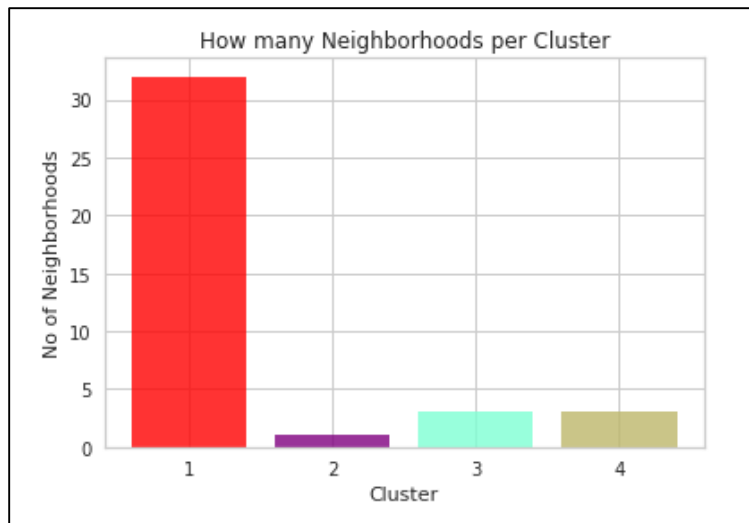
	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berczy Park	0.0	0	43.644771	-79.373306	The Keg Steakhouse + Bar - Esplanade	43.646712	-79.374768	Restaurant
0	Berczy Park	0.0	0	43.644771	-79.373306	LCBO	43.642944	-79.372440	Liquor Store
0	Berczy Park	0.0	0	43.644771	-79.373306	Fresh On Front	43.647815	-79.374453	Vegetarian / Vegan Restaurant
0	Berczy Park	0.0	0	43.644771	-79.373306	Meridian Hall	43.646292	-79.376022	Concert Hall
0	Berczy Park	0.0	0	43.644771	-79.373306	Goose Island Brewhouse	43.647329	-79.373541	Beer Bar

The cluster and venue data is then plotted on a Folium map, with each neighborhood colored based on the cluster label.

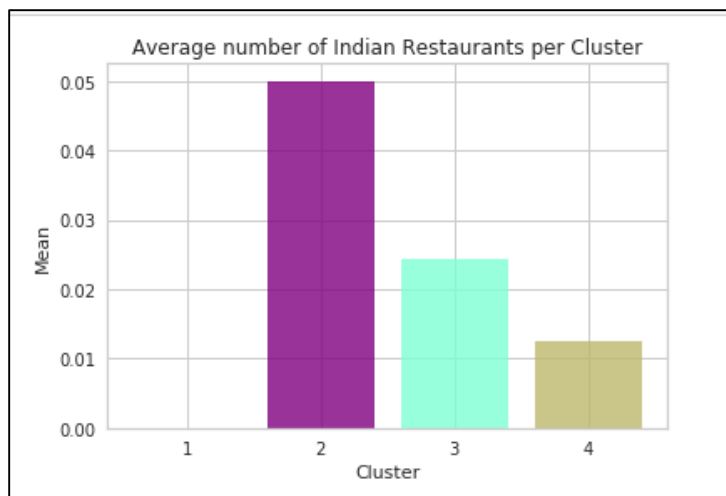


The map above shows the different clusters that had a similar mean frequency of Indian restaurants.

We have a total of 4 clusters (0,1,2,3). Before we analyze them one by one let's check the total amount of neighborhoods in each cluster. A bar graph is plotted using Matplotlib.



Now let's compare it with the average number of Indian restaurants per cluster.



As we can see from the above charts, even though there is only 1 neighborhood in Cluster 2, it has the highest average of Indian Restaurants (0.0500). While Cluster 1 has the most neighborhoods but has the least average of Indian Restaurants (0.0009).

[Analysis of each Cluster](#)

Now let's analyze the Clusters individually (Note: these are just snippets of the data).

Cluster 1 (Red)

	Borough	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
5	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Corktown Common	43.655618	-79.356211	Park
6	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	The Distillery Historic District	43.650244	-79.359323	Historic Site

- There are 32 Neighborhoods in Cluster 1
- There are 798 venues in Cluster 1
- There are 0 Indian Restaurants in Cluster 1

Cluster 2 (Purple)

	Borough	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	Ezra's Pound	43.675153	-79.405858	Café
1	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	Roti Cuisine of India	43.674618	-79.408249	Indian Restaurant
18	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	Subway	43.675650	-79.410255	Sandwich Place
17	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	Tim Hortons	43.675800	-79.403532	Coffee Shop
16	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	Martino's Pizza	43.675560	-79.403558	Pizza Place
15	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	LCBO	43.675344	-79.405327	Liquor Store
14	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	Subway	43.675071	-79.406877	Sandwich Place
13	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	Haute Coffee	43.675818	-79.402793	Café
12	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	Pour House	43.675641	-79.403821	Pub
11	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	The Annex Hodgepodge	43.674975	-79.406543	Sandwich Place
10	Central Toronto	The Annex, North Midtown, Yorkville	0.05	1	43.67271	-79.405678	Toronto Archives	43.676447	-79.407509	History Museum

- There are 2 Neighborhoods in Cluster 2
- There are 19 venues in Cluster 2
- There are 1 Indian Restaurants in Cluster 2

Cluster 3 (Turquoise)

	Borough	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	East Toronto	The Danforth West, Riverdale	0.024390	2	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop
1	East Toronto	The Danforth West, Riverdale	0.024390	2	43.679557	-79.352188	Dolce Gelato	43.677773	-79.351187	Ice Cream Shop
2	East Toronto	The Danforth West, Riverdale	0.024390	2	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant
3	East Toronto	The Danforth West, Riverdale	0.024390	2	43.679557	-79.352188	Cafe Fiorentina	43.677743	-79.350115	Italian Restaurant
4	East Toronto	The Danforth West, Riverdale	0.024390	2	43.679557	-79.352188	Louis Cifer Brew Works	43.677663	-79.351313	Brewery
5	East Toronto	The Danforth West, Riverdale	0.024390	2	43.679557	-79.352188	La Diperie	43.677702	-79.352265	Ice Cream Shop
6	East Toronto	The Danforth West, Riverdale	0.024390	2	43.679557	-79.352188	The Big Carrot Organic Juice Bar	43.677438	-79.352683	Juice Bar

- There are 3 Neighborhoods in Cluster 3
- There are 106 venues in Cluster 3
- There are 3 Indian Restaurants in Cluster 3

Cluster 4 (Dark Khakhi)

	Borough	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Downtown Toronto	Central Bay Street	0.015385	3	43.657952	-79.387383	Jimmy's Coffee	43.658421	-79.385613	Coffee Shop
1	Downtown Toronto	Central Bay Street	0.015385	3	43.657952	-79.387383	Tim Hortons	43.658570	-79.385123	Coffee Shop
2	Downtown Toronto	Central Bay Street	0.015385	3	43.657952	-79.387383	Somethin' 2 Talk About	43.658395	-79.385338	Middle Eastern Restaurant
3	Downtown Toronto	Central Bay Street	0.015385	3	43.657952	-79.387383	Hailed Coffee	43.658833	-79.383684	Coffee Shop
4	Downtown Toronto	Central Bay Street	0.015385	3	43.657952	-79.387383	Neo Coffee Bar	43.660140	-79.385870	Coffee Shop
5	Downtown Toronto	Central Bay Street	0.015385	3	43.657952	-79.387383	The Queen and Beaver Public House	43.657472	-79.383524	Gastropub
6	Downtown Toronto	Central Bay Street	0.015385	3	43.657952	-79.387383	The Elm Tree Restaurant	43.657397	-79.383761	Modern European Restaurant

- There are 3 Neighborhoods in Cluster 4
- There are 219 venues in Cluster 4
- There are 3 Indian Restaurants in Cluster 4

Results and Discussions

- Most of the Indian Restaurants are in cluster 2 represented by the purple clusters
- The Neighborhoods located in the Central Toronto area that have the highest average of Indian Restaurants are: The Annex, North Midtown and Yorkville
- Even though there is a huge amount of Neighborhoods in cluster 1, there is little to no Indian Restaurant
- In the East Toronto area (cluster 3) has the second last average of Indian Restaurants
- Looking at the high availability of nearby venues, the optimum place to put a new Indian Restaurant is in Downtown Toronto as there are many Neighborhoods in the area but little to no Indian Restaurants therefore, eliminating any competition

Conclusion

This concludes the optimal findings for this project and recommends the entrepreneur to open an authentic Indian restaurant in the locations mentioned above. In conclusion, to end off this project, we had an opportunity on a business problem, and it was tackled in a way that it was similar to how a genuine data scientist would do. We utilized numerous Python libraries to fetch the information, control the content and break down and visualize those datasets. We have utilized Foursquare API to investigate the settings in neighborhoods of Toronto, get a great measure of data from Wikipedia which we scraped with the BeautifulSoup Web scraping Library. We also visualized utilizing different plots present in seaborn and Matplotlib libraries. Similarly, we applied AI strategy to anticipate the error given the information and utilized Folium to picture it on a map.

The analysis and the accuracy of the findings can be definitely enhanced by bringing in other parameters such as proximity to the city center, the economy of the locality etc. Nonetheless, if the food is authentic, affordable and good quality, I am confident that it will have great following everywhere.