The Battle of Neighborhoods

Project Presentation

By

Sankar Ramamurthy – Aug2020

Introduction

- Toronto, the 4th most populous city in North America
- It is a multi-cultural and cosmopolitan city
- Comprises of hundreds of large and small Neighborhoods
- Tough competition in restaurant business
- 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 etc.

Data Acquisition

- The Neighborhoods details of Toronto shall be obtained from Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- The geographical location data using Geocoder Package: https://cocl.us/Geospatial_data
- All the existing venues in the neighborhood along with their category: obtained using the FourSquare API

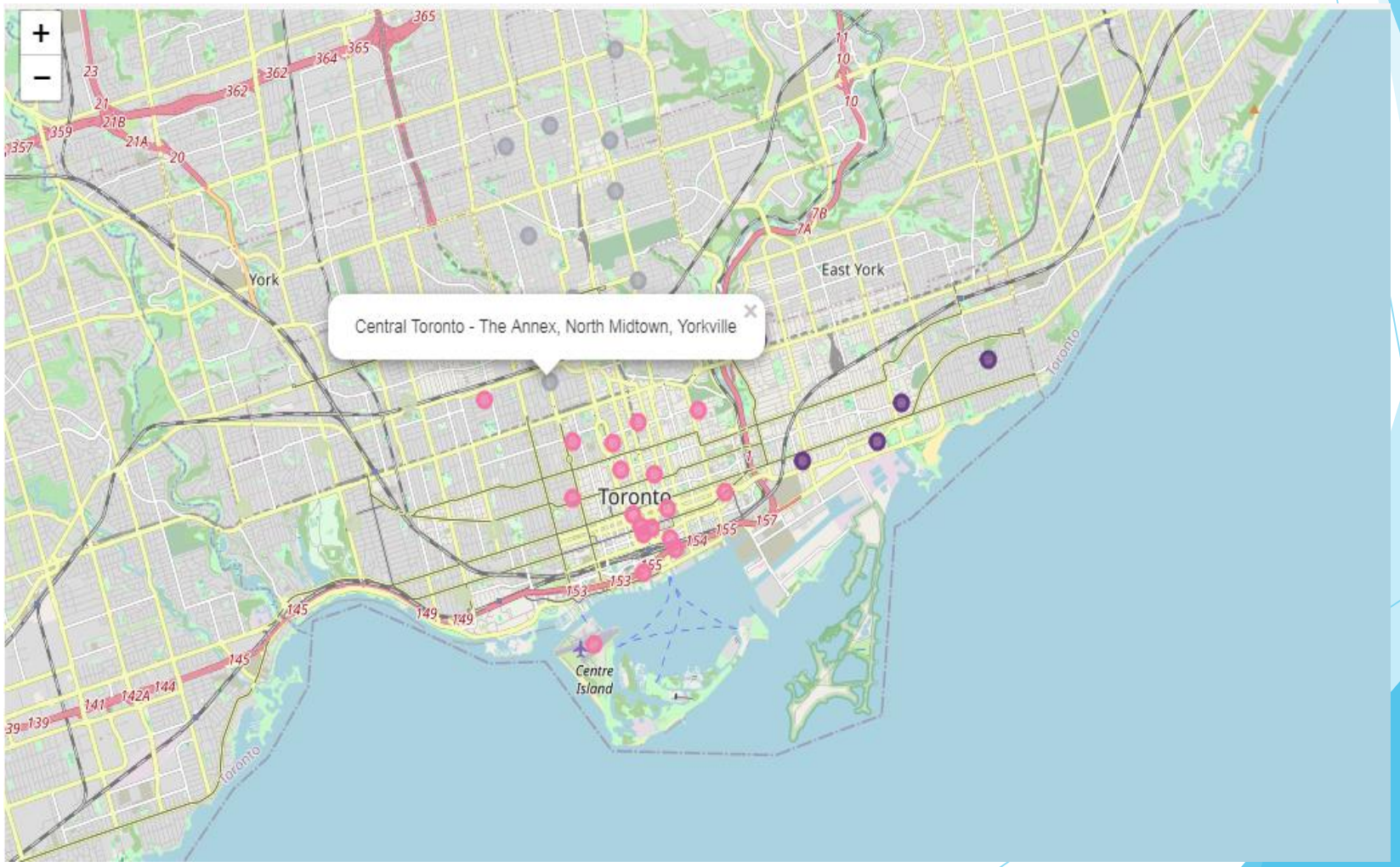
Methodology

- 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 & Cleansing

	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



Venue Data – Foursquare API

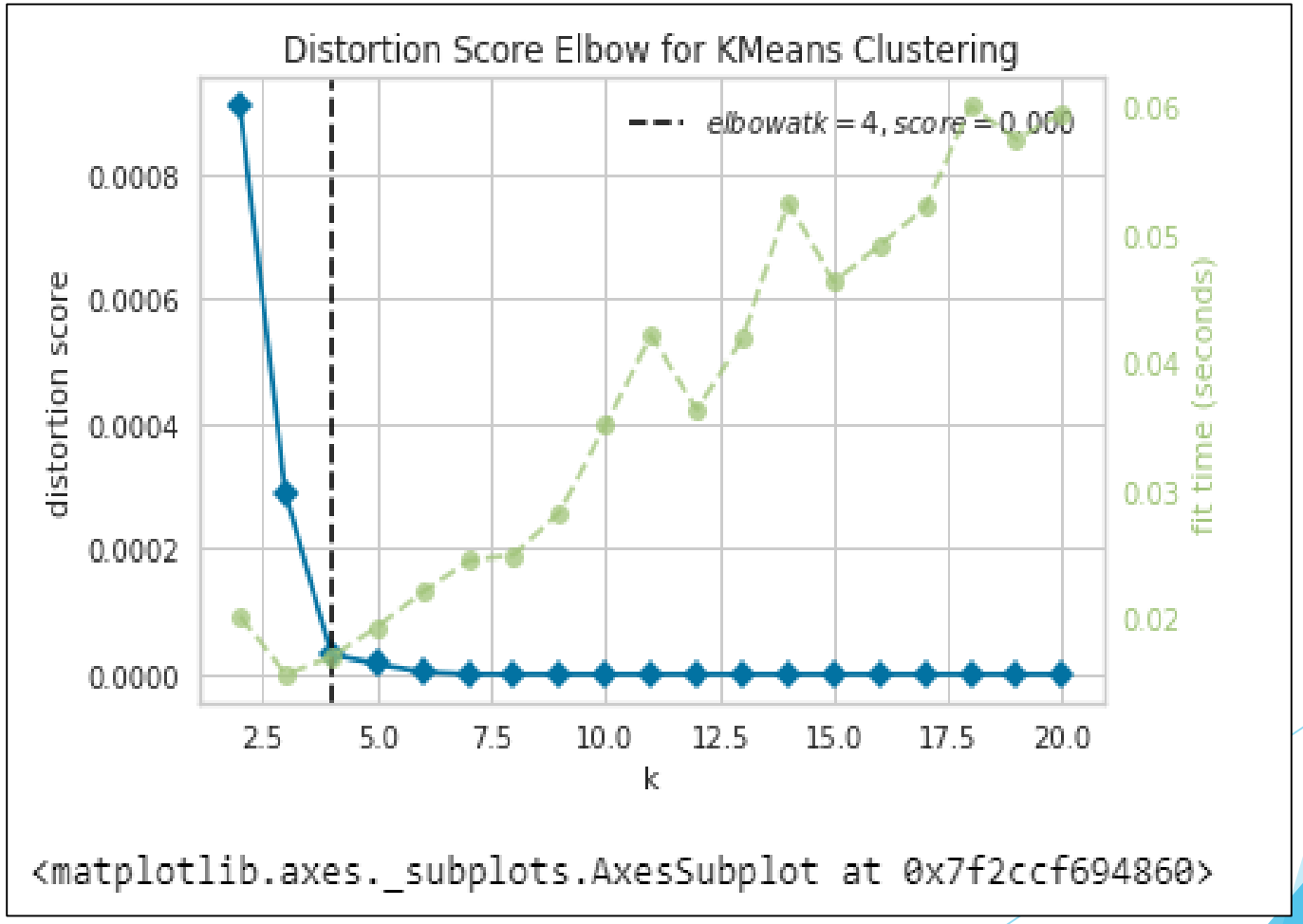
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent 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

Neighborhood Analysis

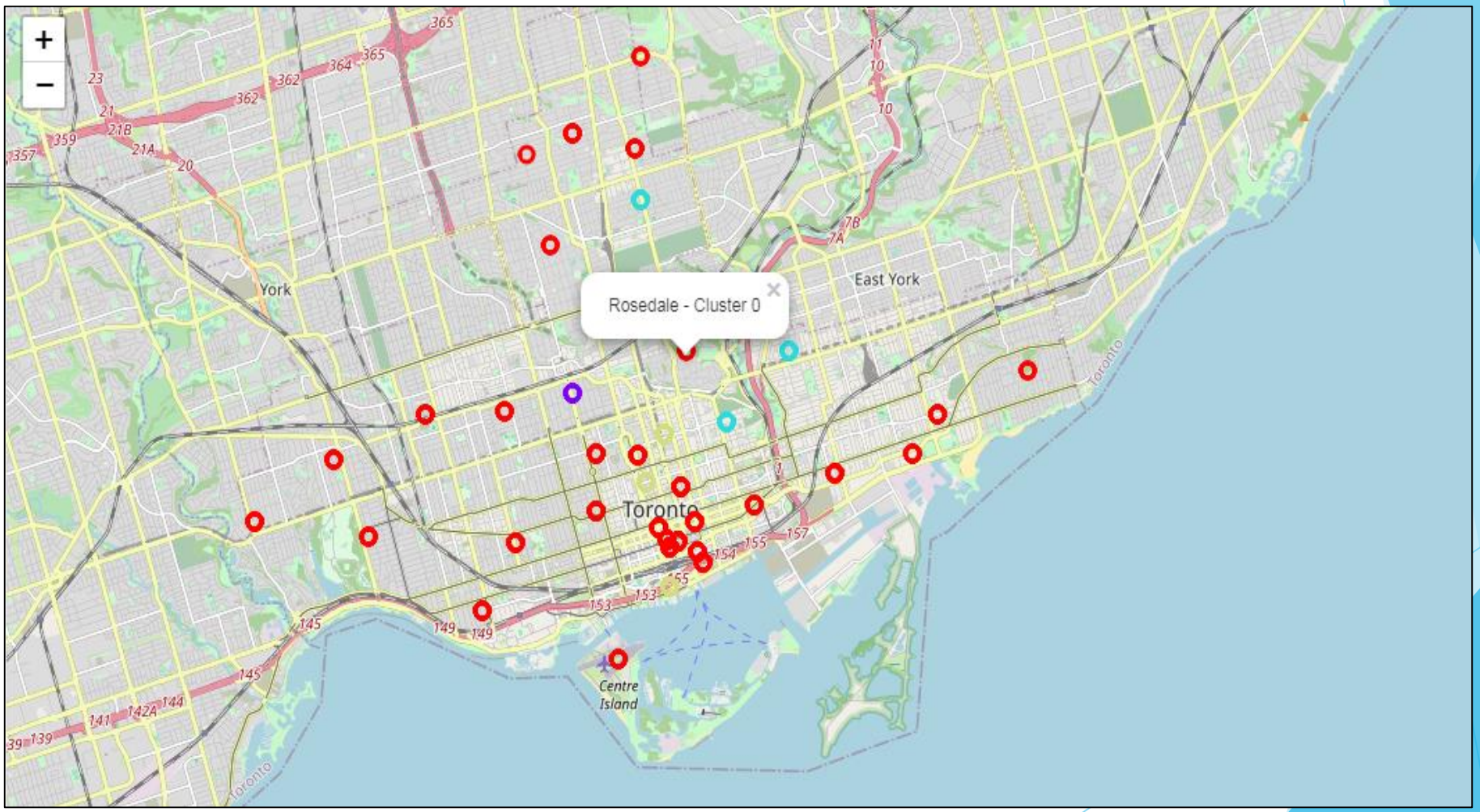
	Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theme Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Wine Bar	V
0	Berczy Park	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 × 235 columns

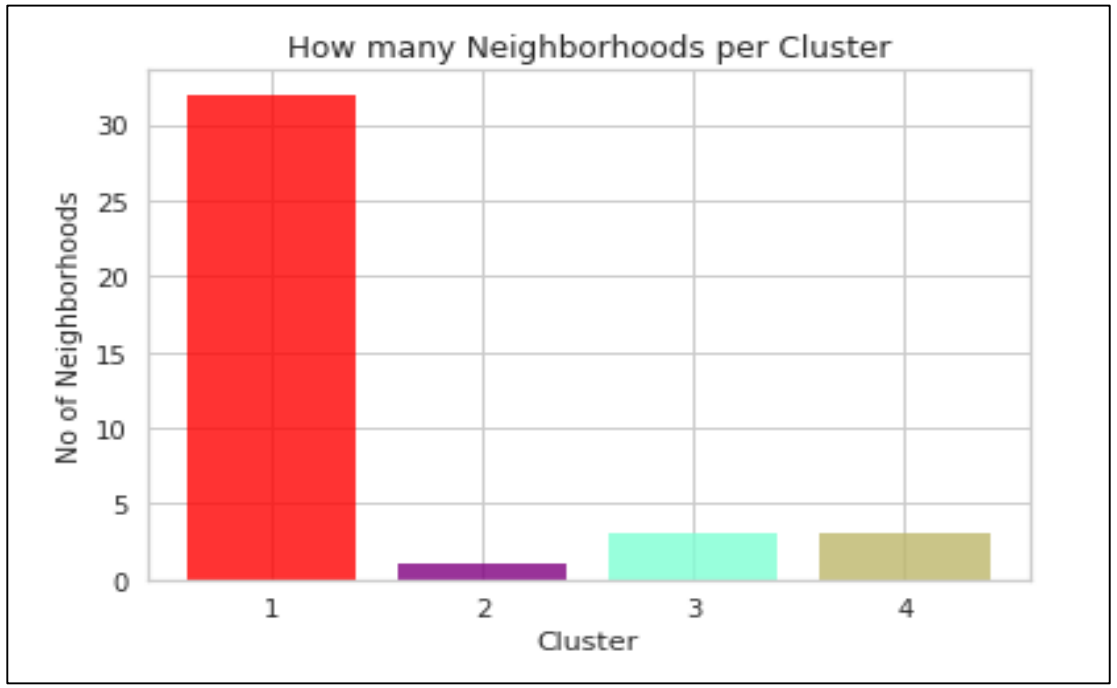
Optimum K-value: Elbow Point

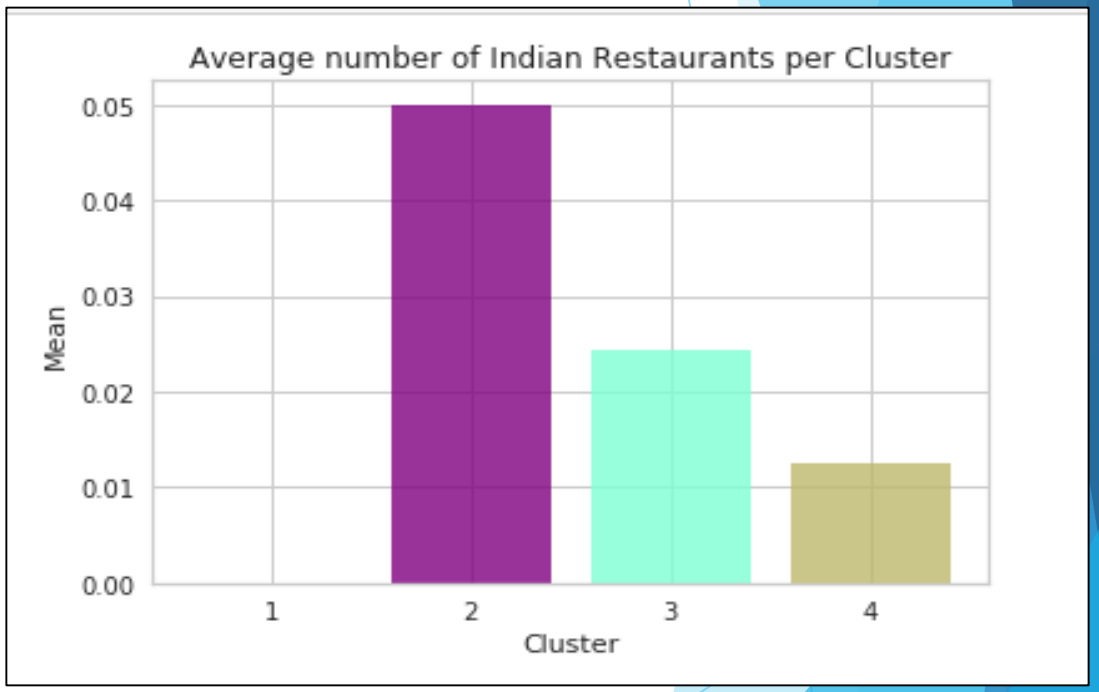


Clustering the Neighborhoods



Cluster Comparison





Results

- 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

- Utilized numerous Python libraries to fetch the information, control the content and break down and visualize those datasets
- Utilized Foursquare API to investigate the settings in neighborhoods of Toronto
- Data obtained from Wikipedia was scraped with the Beautifulsoup Web scraping Library
- Visualized the data utilizing different plots present in seaborn and Matplotlib libraries
- 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.