**PROJECT REPORT**

BY : PRANJAL SHRIVASTAVA

The report is part of the IBM Applied Data Science Specialization Capstone Project. The main objectives of this project were to define a business problem, look for data on the web and use Foursquare location data to compare to different neighborhoods of Toronto to figure out which neighborhood is suitable for starting a new restaurant business.

INTRODUCTION –

**Toronto** **city**, capital of the province of Ontario, southeastern Canada. It is the most populous city in Canada and the country’s financial and commercial center. The demographics of Toronto make Toronto one of the most multicultural and multiracial cities in the world. More than half of the entire Indian-Canadian population resides in Toronto, people from India love food and I love to eat, thus Toronto is one of the best places to start an INDIAN RESTAURANT.

BUSINESS PROBLEM –

In this capstone project, we will analyze the neighborhoods in Toronto to identify the most profitable neighborhood for opening an Indian Restaurant, by using Web Scraping, Data Pre-processing, Machine learning algorithms like K-Means clustering algorithm, and Foursquare API Service.

TARGET AUDIENCE –

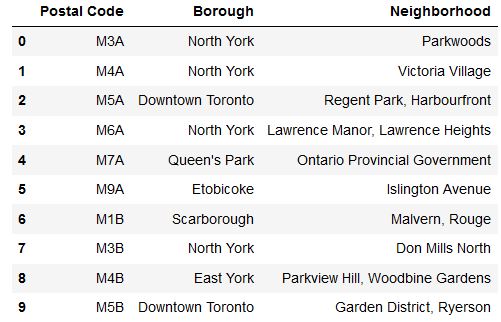
* The business owner who wants to invest or open a start-up company or restaurant.
* The freelancer who loves to have their own small company or restaurant as a side business.
* Indian crowd who wants to find neighborhoods with lots of options for Indian restaurants.
* Tourists who want to eat Indian food.

Data Sources –

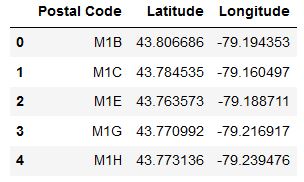
* Toronto City Neighbourhoods Data –<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* Geographical Coordinates of the Neighbourhoods – <https://cocl.us/Geospatial_data>
* Location Data of Neighbourhood – Foursquare API Services

Methodology –

* First, web scraping for data using the ‘BeautifulSoup’ package is done.
* Then a dataframe is created which contains Postal Code, Borough, Neighbourhood.
* Then further processing of the dataframe is done, like removing unassigned values, merging different neighborhoods with the same borough.



* The next step is to add geographical coordinates, for that the Geospatial\_data.csv is used.



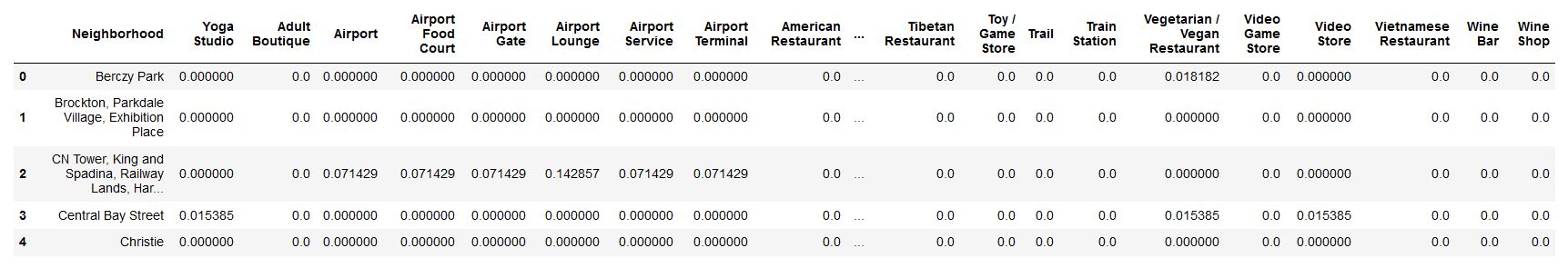
* Then merging is done based on Neighbourhood.



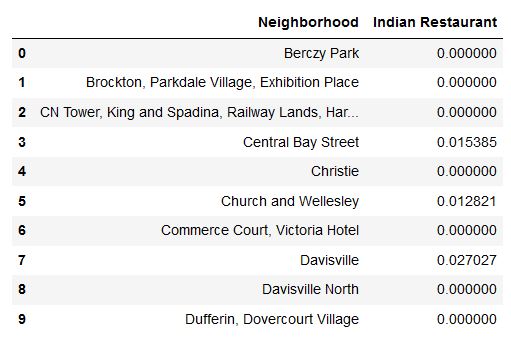
* Finally, the dataframe is modified, in which Borough contains ‘Toronto’.



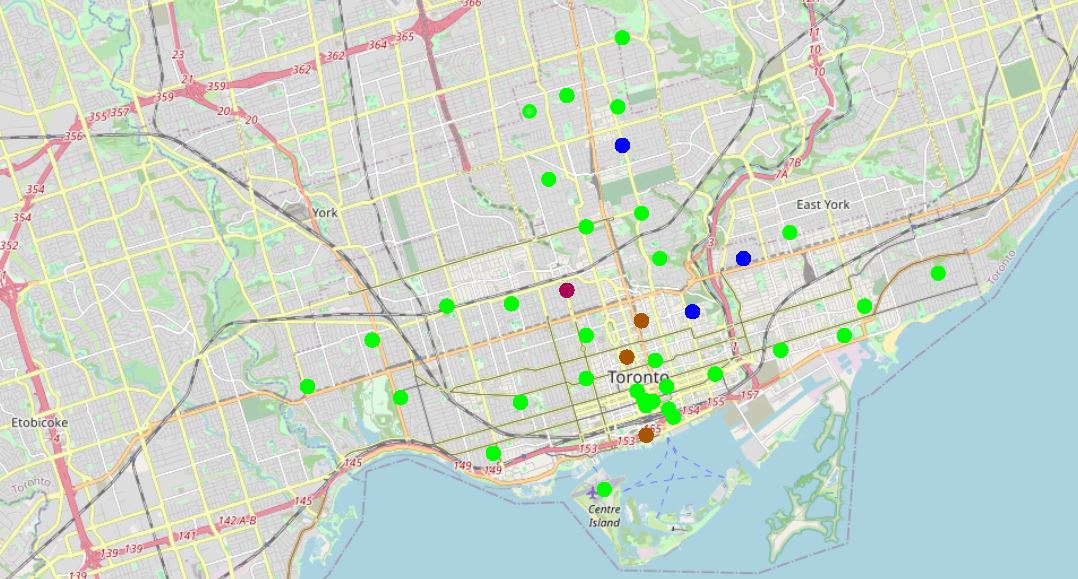
* Then using Foursquare API developer services, 100 venues were explored under the radius of 500m. A Foursquare developer account to obtain a Client ID and Client Secret key to pull the data. From Foursquare, the names, categories, latitude, and longitude of the venues were pulled.
* Then one-hot encoding was performed, for each of the neighborhood’s venues were turned into the frequency at how many of those Venues were located in each neighborhood. Then, the values in each venue category were grouped by the average of the frequency of each venue category.



* Then, a new dataframe was created, which stored Neighbourhood and Indian Restaurant column. It was created to make further operations simpler.



* At last, clustering of Indian Restaurant was done, it was based on K-Means Clustering Algorithm. Here k-value was taken as 4, then the model was fitted. And finally, the clustered map was displayed.

Results & Discussion–

* Cluster 1 – Green Colour
* Cluster 2 – Blue Colour
* Cluster 3 – Purple Colour
* Cluster 4 – Brown Colour

Cluster 1 - Depicts the least frequency of Indian Restaurants among the neighborhoods.

Cluster 3(The Annex, North Midtown, Yorkville) - Depicts the maximum frequency of Indian Restaurants among the neighborhoods.

After analyzing, it is found that The Annex, North Midtown, Yorkville has the highest frequency amongst all other neighbourhoods, followed by Davisville. Approximately 80 percentage of the neighborhood has no authentic Indian Restaurant, thus it gives a good opportunity for business owner and freelancer to open a new Restaurant. The green cluster(Cluster 1) can be a good option to open an Indian Restaurant, for example The Beaches, St.James Town, India Bazaar, Forest Hill and Parkdale Village are some good options. This concludes the findings for the location and recommends the business owner and freelancer to open an authentic Indian restaurant in these locations.

Conclusion –

Finally, to conclude this project, I have got a glimpse of how data-science project look-like. I have used various libraries like, folium, pandas, sklearn, requests. I have also used BeautifulSoup package for web scraping. Here, I have also used Foursquare API services to explore the neighborhoods. And finally, I have used machine learning algorithm, K-Means Clustering Algorithm, to predict the most profitable neighborhood for opening an Indian Restaurant.