Introduction

This report is apart of the IBM Applied Data Science Capstone project. For this project, we are required to come up with a hypothetical business problem where we can utilise our data science knowledge and skills to solve a business problem. Within this report, I talk about the business problem, the data getting evaluated to solve the business problem, the method being used, and results of the data getting analysed.

Business problem

The Gupta's have planned to migrate to Toronto, Canada. They have been running a family hospitality business for many years in India. Since they are moving to Toronto, they are planning on opening their hospitality business in the city of Toronto, where they can grow their business whilst profiting. This is a new place for them, so they are not sure what the best location is for opening a Indian restaurant. Therefore, the objective of our project is to help the Gupta's find a great location for opening a business in Toronto using the skills and knowledge of data science.

Target Audience

Individuals or group of individuals who are wanting to open continental restaurants in the city of Toronto.

Data

The data getting used for analysis are as follows:

List of neighbourhoods in Toronto. We were given a Wikipedia link with the list of neighbourhoods in Toronto, where I was able to extract the data via we scrapping.

Geospatial coordinates of neighbourhoods in Toronto. I used the Geocoder Package within Python to get the latitude and longitude coordinates.

Data of venues related to Indian restaurants. With the use of Foursquare API to get venue data.

Methodology

First and foremost, we will need to use the Foursquare location data (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and postal code list data (Geospatial Coordinates data) in conjunction with each other. The Wikipedia website contains the Canadian postal code starting with the letter M and the corresponding Borough and Neighborhood. This list gets consolidated with the Geospatial Coordinates data which includes the longitude and latitude of the Neighborhoods.

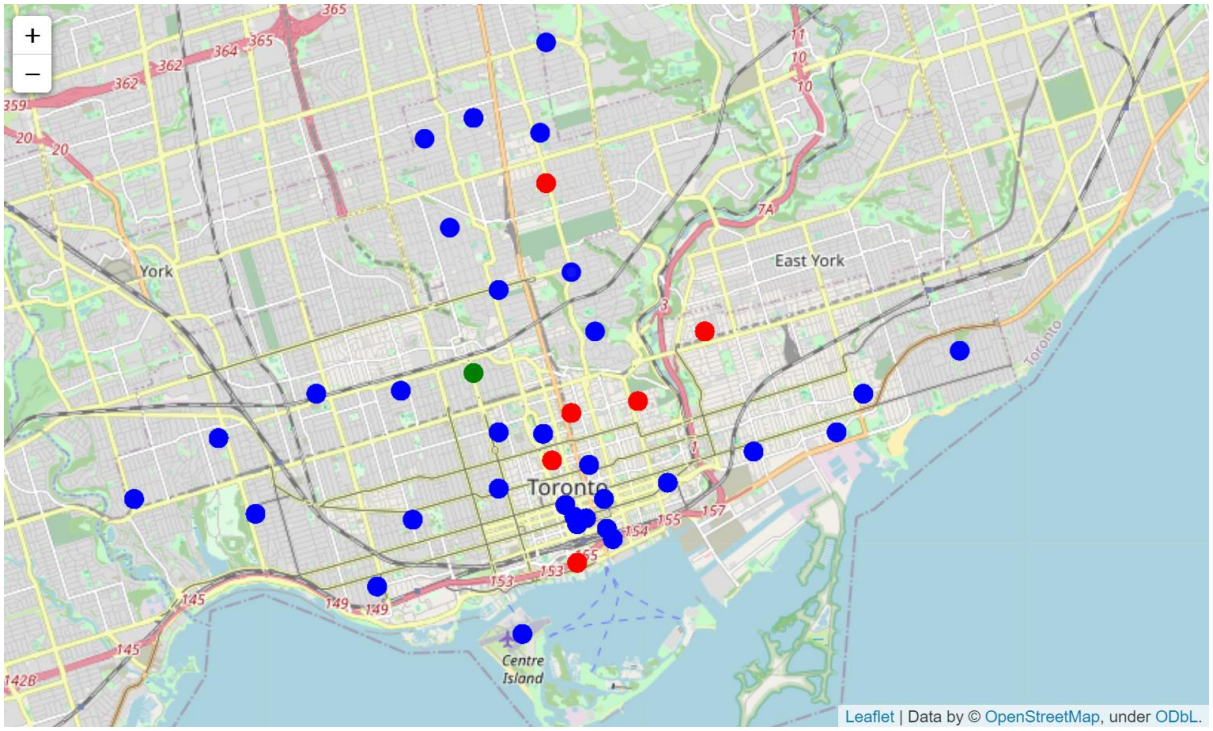
We then explore the neighborhoods in Toronto using the Foursquare location data. We will ask the stakeholders for a preference of location, and depending on that, we cluster the neighborhoods by the different suburbs. This is essential as by clustering, we can identify and group similar data points in larger datasets without concern for the specific outcome. After analysing the data frame, we can suggest different locations to the Gupta's.

Results

Below is a summary of the clustering of Indian restaurants in the city of Toronto, Canada.

I used the K-Means to build 3 different clusters on how many Indian restaurants are in each neighbourhood.

- Cluster 0 had the highest number of Indian restaurants.
- Cluster 1 had no Indian restaurants.
- Cluster 2 had low number of restaurants.



Conclusion

There is a large number of Indian restaurants already in cluster 0 which includes places like Riverdale, Davisville and Cabbagetown. So, opening a new Indian restaurant in that area would create conflicts with other already established Indian restaurants, so I would not recommend opening a new shop in this neighbourhood. Cluster 1 has no effective Indian restaurants, therefore, I would highly recommend opening an Indian restaurant in this area, as there will be a new option of cuisine for the residents in this neighbourhood and since it is the only Indian restaurant in that area, there will high sale and demand for Indian food.