Introduction/Business Problem

Canada is a country in the northern part of North America. It is also known as the second largest country by total area. The 2016 Canadian Census shows that between 2011 and 2016 its population grew by 1.7 million people, with immigrants accounting for two-thirds of the increase. Canada is known for its cultural diversity. The three largest metropolitan areas called Toronto, Montreal, and Vancouver accounts for major business developments and economic growth. Hence these places are the primary attractions for immigrants to settle down.

Toronto is an international center for business, finance, multinational organizations, educational institutions and sport activities. Thus, the scope for food industry is undeniably huge. The main objective of this project is to find the best neighborhoods in Toronto area for an entrepreneur to start an Indian restaurant.

Data Acquisition

Source 1: Wikipedia - List of postal codes of Canada

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Source 2: Geospatial Data using Geocoder package

This is the list of geographical coordinates of the postal codes in Toronto. It is a CSV formatted file provided by Congnitive Class https://cocl.us/Geospatial_data

Source 3: Data from Foursquare

Foursquare is a technology company that built a massive dataset of location data. In this project we collect Toronto venue data from Foursquare

Methodology

Data will be collected from Wikipedia and Foursquare. It is then cleansed, modified and upgraded with geographical coordinates using Geocoder.

Python packages for data analysis such as Pandas, Numpy, Beautiful soup, k-means clustering from sklearn, geopy, matplotlib, folium are installed.

The cleaned data is processed into a dataframe and sorted based on rankings. The Indian restaurants near to each main neighborhoods in the City of Toronto is plotted.

The data is converted into clusters using K-means clustering. The data will be analyzed and visualizations will be created to help obtain the insights to the problem and conclusions are made.

Results

1. The Indian restaurants can be seen as blue circle markers On the city of Toronto map.
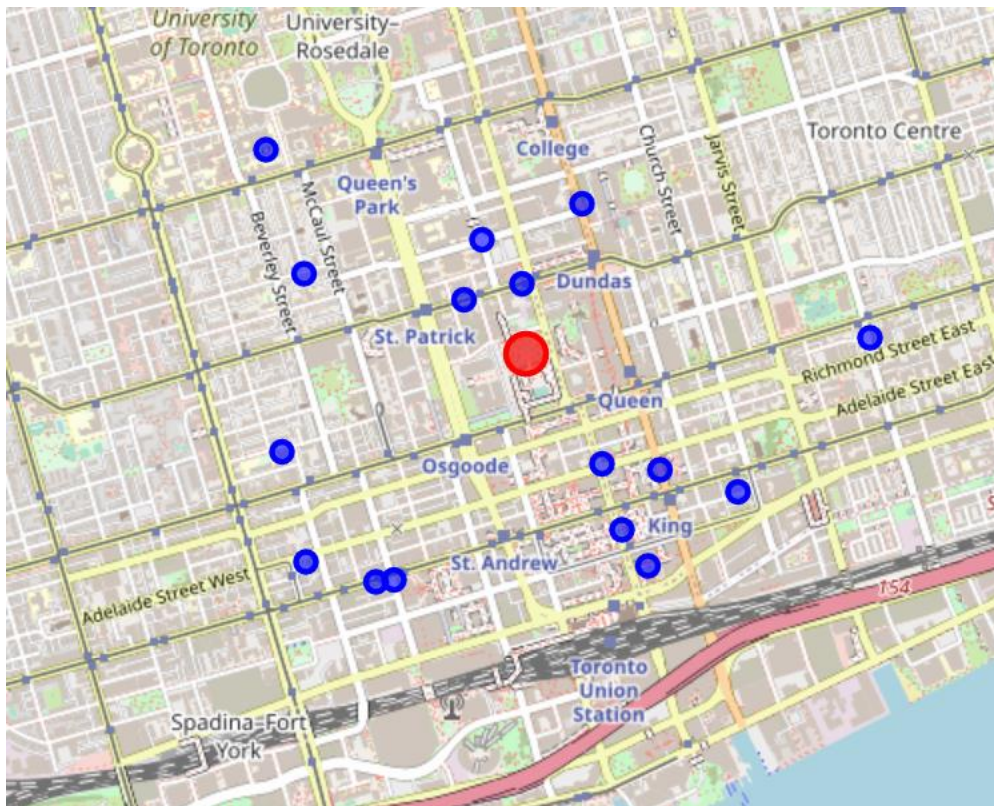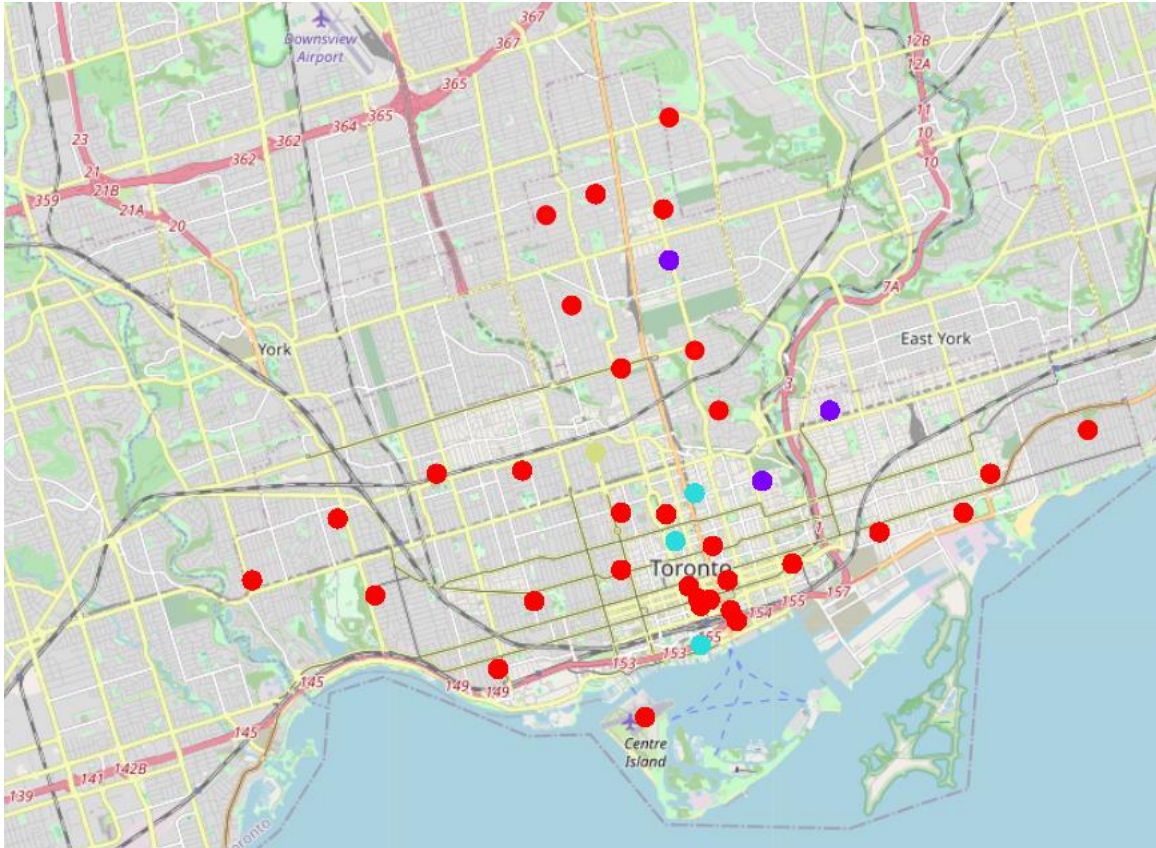
2. The value of K is obtained as 4 from the Elbow point figure below. Thus four clusters are made for the analysis
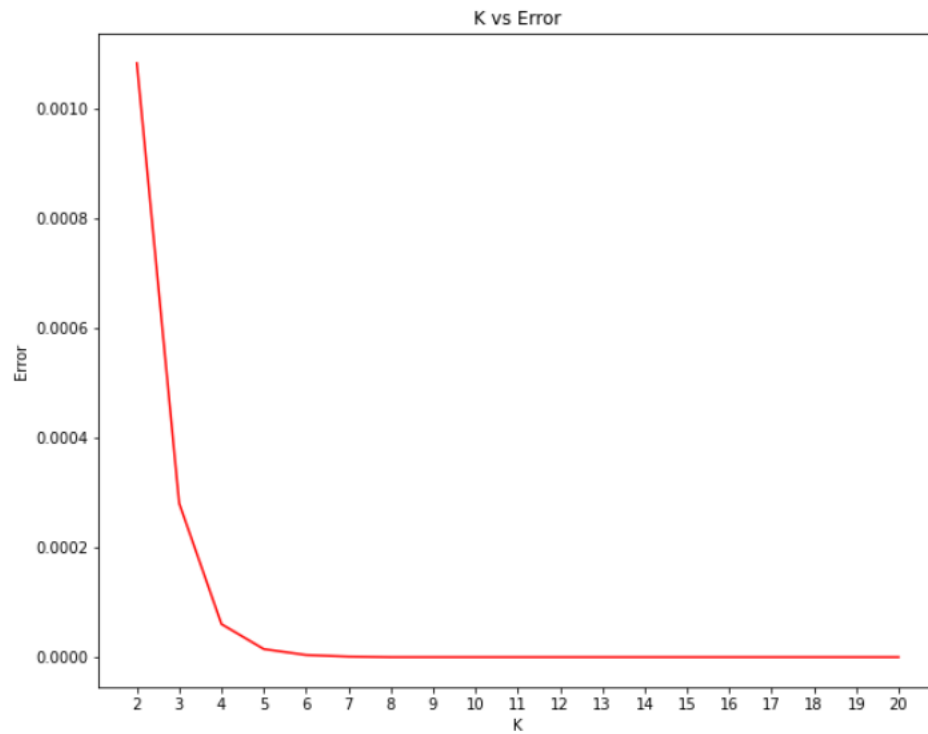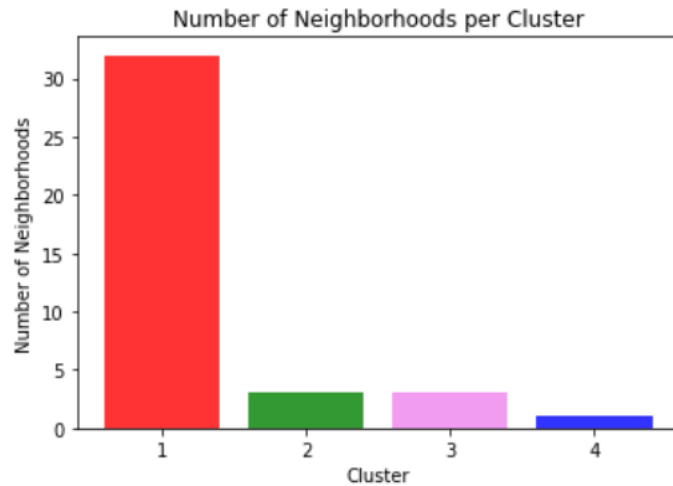

From the above figure the value for K = 4. Thus we run K-Means to cluster the Neighborhood into 4 clusters

3. The main venues in the city of Toronto are differentiated into 4 clusters.
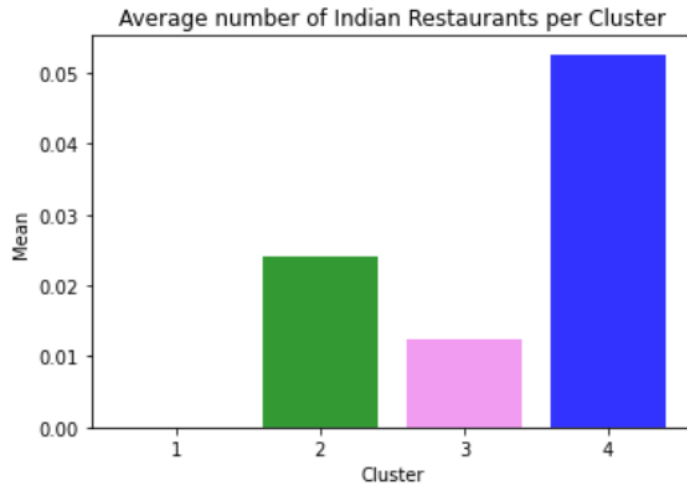
4. The number of neighbourhood associated with each cluster is visualized

Number of Neighborhoods per Cluster

```
ind['Cluster Labels'].value_counts()
```

```
5]: 0    32
    2     3
    1     3
    3     1
    Name: Cluster Labels, dtype: int64
```

5. Average number of Indian restaurants per cluster is visualized

Observations

We have analyzed 4 clusters through K-means clustering.

The Cluster 1 has the maximum number of neighbourhoods (32) but least (Zero) number of Indian restaurants.

The Cluster 2 and Cluster 3 have three(3) neighborhoods each with average Indian restaurants of (0.023) and (0.014) as well

The Cluster 4 has the least number of neighborhoods (1) but highest number of Indian restaurants (0.052) on average

Conclusion

To conclude, it is no doubt that the scope of food business in a city like Toronto is limitless. The multicultural city is highly populated and diversified.

Our analysis shows that, it is better recommended to start a restaurant in Cluster 1 neighbourhoods since they lack Indian cuisines eliminating potential competition. Eventhough, there are variety of choices and multiple food outlets we found that the number of Indian restaurants are comparatively less. Since India contribute a plethora of immigrants and students to the population of Toronto, the success rate of establishing an Indian restaurant is very significant.