**THE BATTE OF NEIGHBORHOODS: COMPARING SERVICE BUSINESSES IN THE CITIES OF NEW DELHI AND TORONTO**

by

Afshan Anam Saeed

**INTRODUCTION**

New Delhi and Toronto are the capital cities of India and Canada, respectively. They are also the major financial cities of their countries. New Delhi, being the capital of a very densely populated country boasts a population of about 1.9 crore and is highly densely populated. Canada on the other hand has a total population of 3.76 crore, with Toronto holding 29.3 lakh individuals amongst them. Different cultures in different cities tend to affect the mannerism of the functioning of the economies of these cities. The objective of this project is to analyze the service economies of these two cities on different corners of the world and see how they compare with one another.

With the difference in the population of these two cities, the top services that are up and running in the neighborhoods and contributing majorly to the economy of that neighborhood may differ. The type of services that are preferred around a certain neighborhood speaks volumes about the income and lifestyle of the inhabitants residing over there. Thus, information about neighborhoods of the form are crucial in making decisions about setting up major businesses in the area.

This project can act as a key source of information to parties considering investing in a certain service business in these two major cities. It can widely be helpful for an investor living in one of these cities and willing to invest in another.

**DATA ACQUISTION AND CLEANING**

**DATA SOURCES**

The data of the Borough, Neighborhoods, geospatial co-ordinates, and businesses running in the city neighborhoods are required for the task at hand. The data of neighborhoods of the New Delhi have been obtained from a Kaggle set ([link](https://www.kaggle.com/shaswatd673/delhi-neighborhood-data?select=delhi_dataSet.csv)) and that for Toronto have been obtained using a Wikipedia page ([link](file:///C:\Users\Aqdus\Desktop\IBM%20Data%20Science\(https:\en.wikipedia.org\wiki\List_of_postal_codes_of_Canada:_M)).

The latitude and longitude data for New Delhi and Toronto have been obtained from a python package called geopy.

To locate the venues around different Boroughs and neighborhoods present in the city of New Delhi and Toronto, I have used Foursquare. Foursquare is an open source site that can bring in information of different locations to explore around a certain neighborhood. This may include tourist attraction sites, restaurants, shops, etc.

**DATA CLEANING**

The information that was obtained on the neighborhoods of the two cities was a clean data that listed the Boroughs and Neighborhoods of the two cities. The required data is the Borough Name, Neighborhood name, latitude, and the longitude of the neighborhood. This data will be stored in data frames of the two different cities.

On obtaining the data for the neighborhood businesses using Foursquare, the results were returned in the form of a json file with several credentials listing business referral ID, name, reviews, etc. The only data that was specifically required for the analysis that is being done was the data on the name, category of business, latitude, and longitude of the certain venue. This data was scraped and converted into a neat and readable pandas Data Frame.

The total neighborhoods returned from the New Delhi database were 185, of which only 163 of them had complete information about them listed. Thus, the neighborhoods with incomplete data were annulled. As for Toronto, a total number of 103 Boroughs were returned with full geospatial location data. Some of the neighborhoods that did not have a Borough name mentioned were annulled from the data frame.

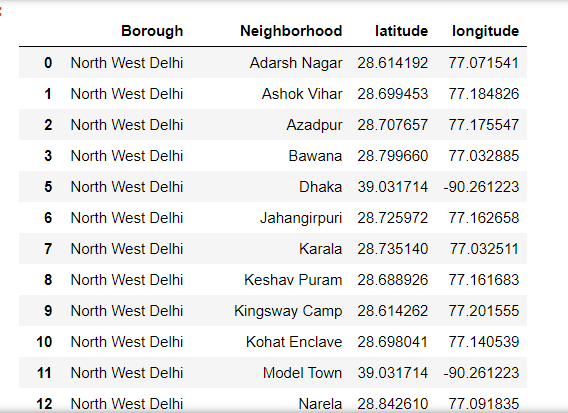
**METHODOLOGY**

The analysis in the project have been done in a Python environment. I have used Jupyter notebook to implement my codes and have used several Python Libraries such as Pandas, NumPy, Beautiful Soup, geopy, Matplotlib, Seaborn, etc. The first step in the project is to gather data from all the sources that were listed. The following steps will be undertaken as a part of the analysis that will be done in the project:

* Neighborhood Data Gathering
* Neighborhood Data cleaning
* Venues Data gathering using Foursquare
* Venue Data Cleaning
* Listing top businesses in a particular neighborhood
* Using K-Means to Cluster the neighborhoods based on their top businesses
* Analyzing the Clusters obtained

1. **Neighborhood Data Gathering and Cleaning**

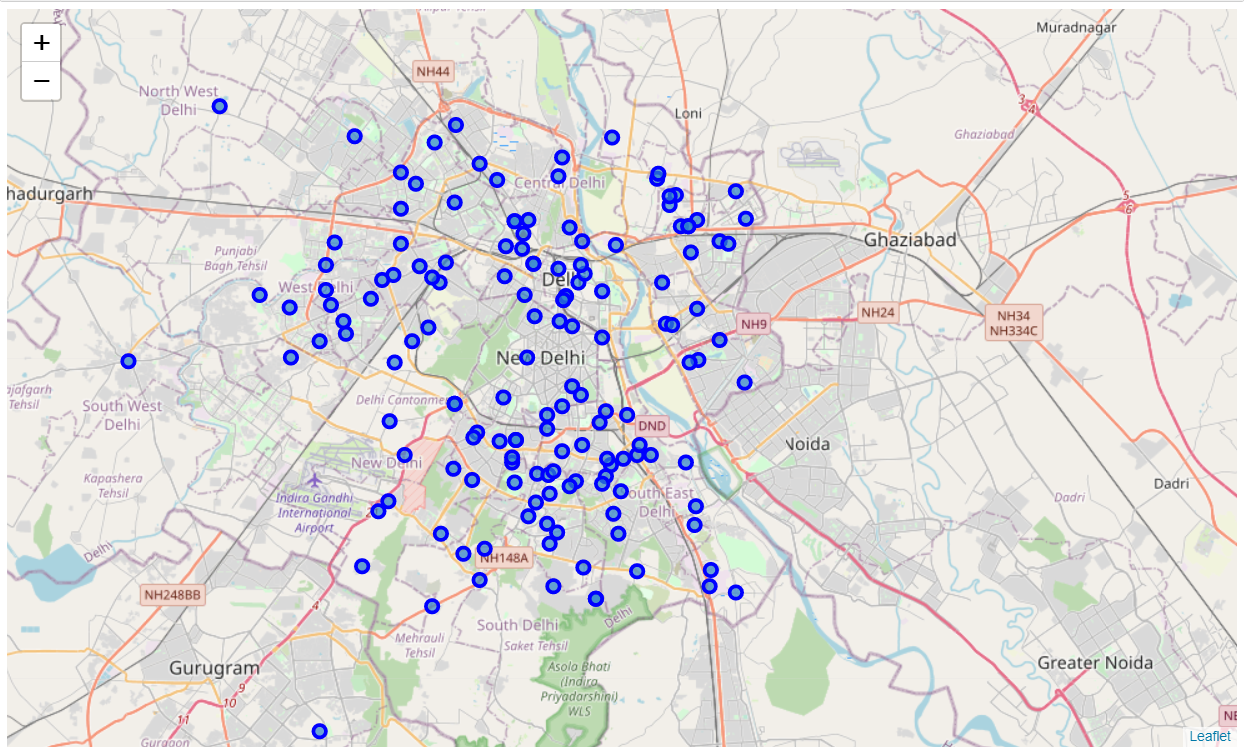
As mentioned in the Data Acquisition section, the data was gathered and converted into their respective data frames. Upon converting the data into data frames, this data was cleaned to include only the necessary information and was made into a readable format. For both the datasets of New Delhi and Toronto, the Borough name, Neighborhoods and the latitude and longitude were retained. The dataset of New Delhi and Toronto looked as follows



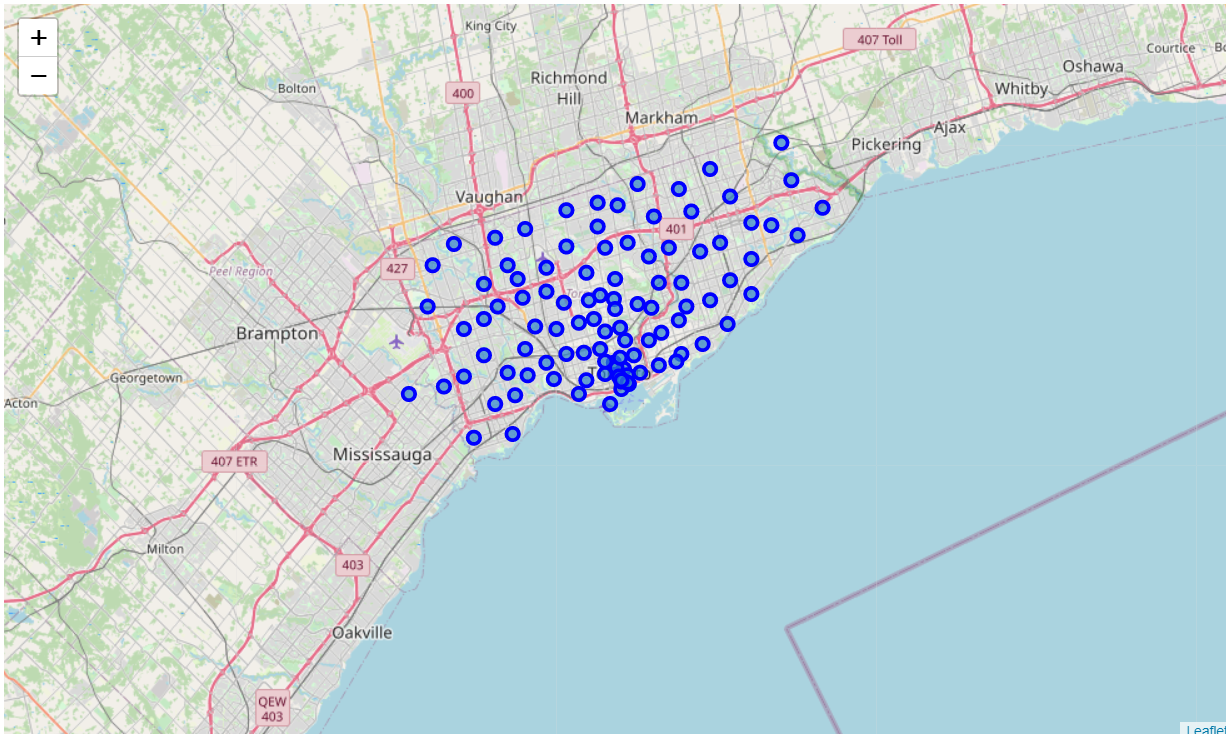
**Pandas Data Frame of New Delhi**

**Pandas Data Frame of Toronto**

The neighborhoods of both these cities are visualized with the help of Folium. Folium helps to visualize our data on an interactive map and helps place markers to label each data point. The maps for New Delhi and Toronto are given below. Each circle represents a neighborhood of their respective cities.



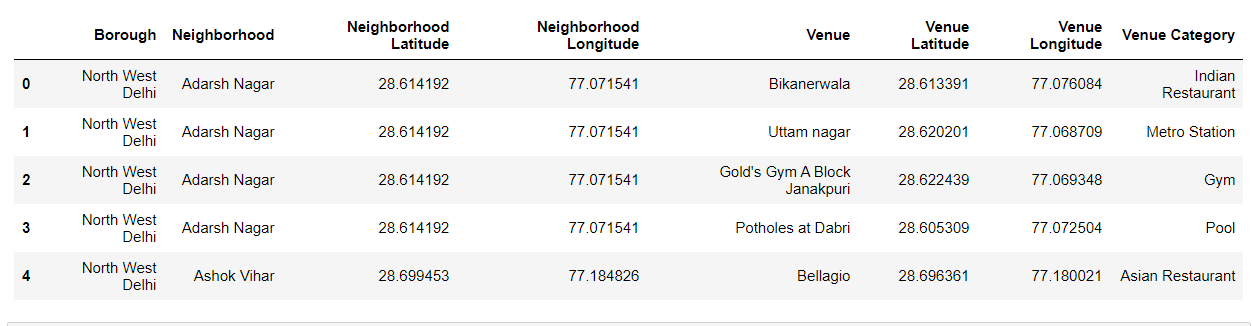
**Neighborhoods of New Delhi visualized with the help of Folium**



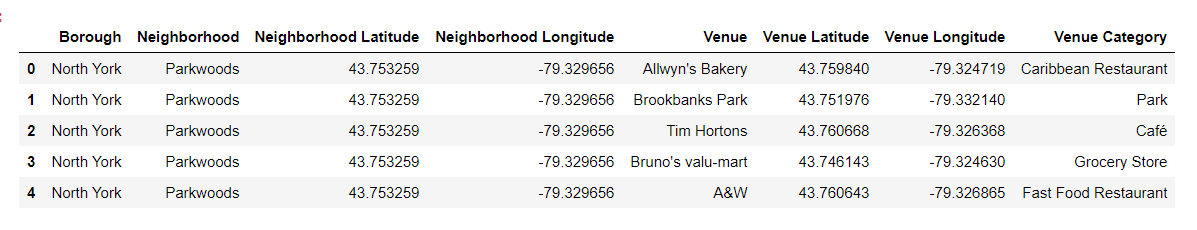
**Neighborhoods of Toronto visualized with the help of Folium**

1. **Extracting Venues with the help of Foursquare for different neighborhoods**

Foursquare is a platform that is useful in obtaining location-based data. Its API can be used to obtain data on neighborhoods of a certain city, such as places to explore, their exact locations, their reviews, their services, etc. The first step in using Foursquare is to set up a developer account on Foursquare to use their API and obtain your credentials. On working in a Jupyter notebook, a URL was set up to use the Foursquare API for locating venues around the neighborhoods of the cities of New Delhi and Toronto and all the credentials were included in the same. The result was returned in the form of a json file, that included the venue categories, name, location, ID, etc. Only the useful information was extracted and converted into a pandas data frame. The data frame heads have been shown in the figure below. This includes the Borough name, Neighborhood, Latitude and Longitude of the neighborhood, the Venue name, the venue location and the venue category.



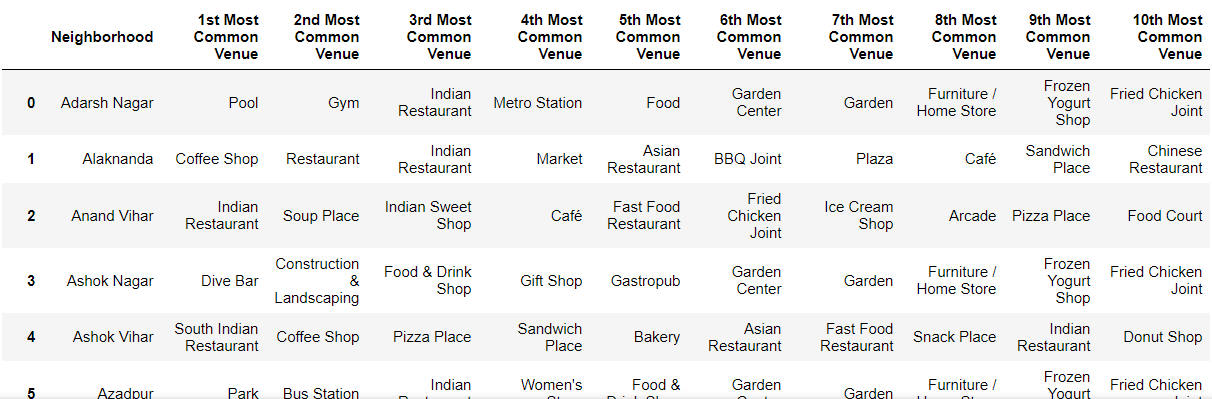
**Venue data of New Delhi neighborhoods obtained with the help of Foursquare**



**Venue data of Toronto neighborhoods obtained with the help of Foursquare**

1. **Listing top businesses operating in a neighborhood**

Following the collection of relevant data, dummies were created for each type of venue categories and the mean of all these venue categories were taken to sort each type of category in the ascending order within each neighborhood. This brings out the most common businesses running within each neighborhood. In the project, a data frame of the top 10 businesses running in each neighborhood were listed in both New Delhi and Toronto.



**A data frame of the top 10 businesses operating in the different neighborhoods of New Delhi**

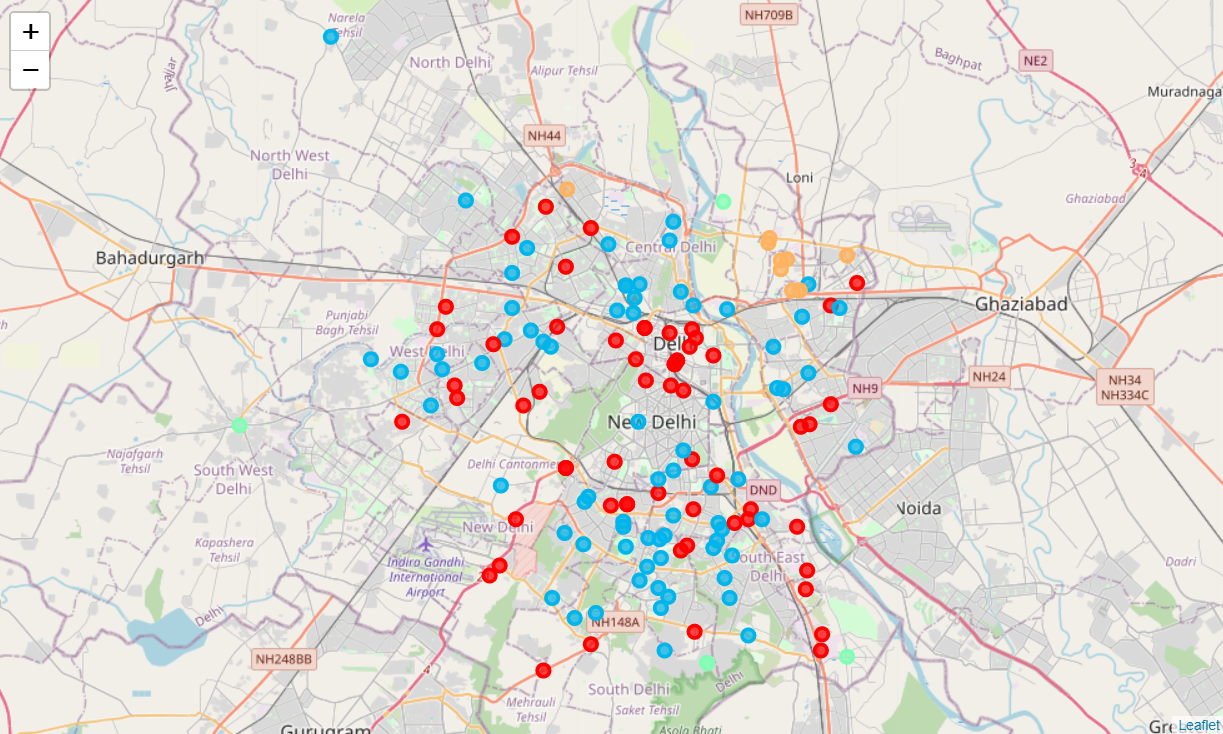
A picture containing diagram

Description automatically generated

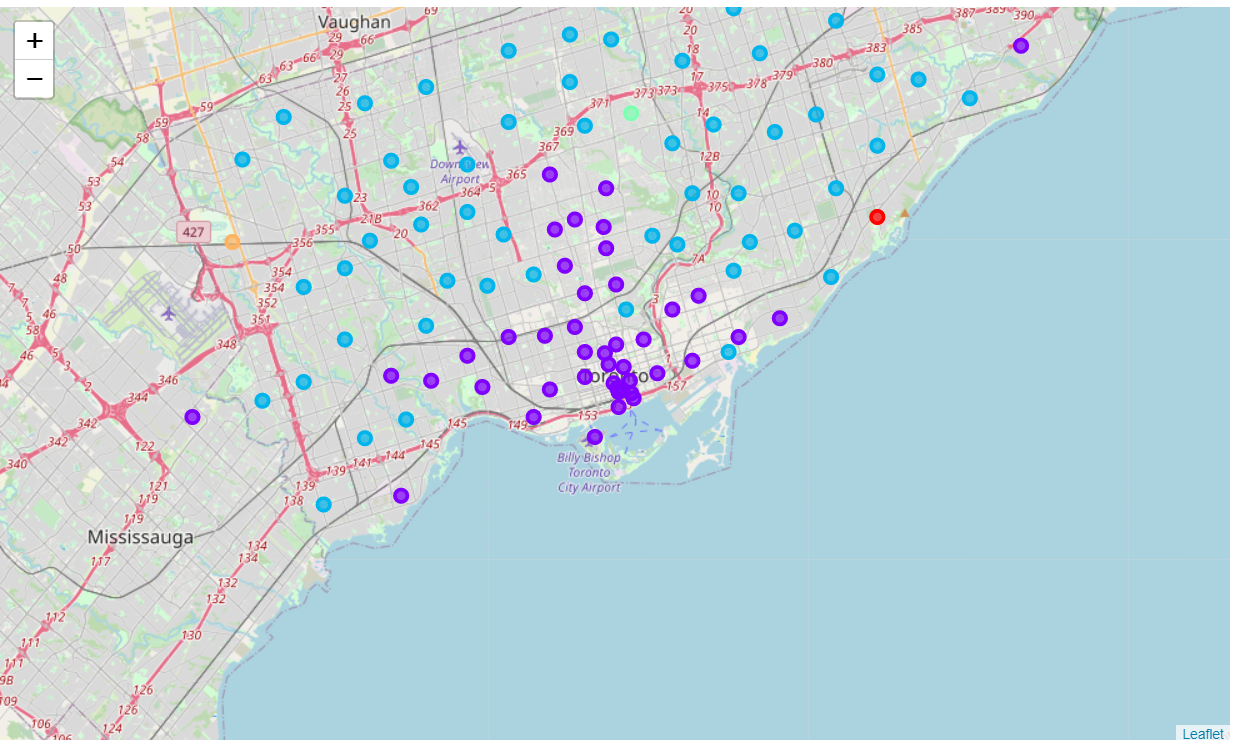
**A data frame of the top 10 businesses operating in the different neighborhoods of Toronto**

1. **Clustering of neighborhoods of New Delhi and Toronto**

In the next few steps, the method of K-Means was used to cluster the neighborhoods of New Delhi and Toronto. Several methods of clustering are available such as Hierarchical Clustering, Density Based Clustering, etc. K-Means was used because of its simplicity and the guarantee of convergence in large data sets. In this form of clustering method, the number of clusters are to be pre-defined. In this case, I opted to segment the neighborhoods into 5 clusters each, although New Delhi should be preferably divided into greater number of clusters because of the larger number of neighborhoods present and the vast population that it holds. But to match numbers, 5 clusters have been created for both New Delhi and Toronto. These clusters are represented on the folium map shown below.



**A folium map showing the clustering of the various neighborhoods of New Delhi**



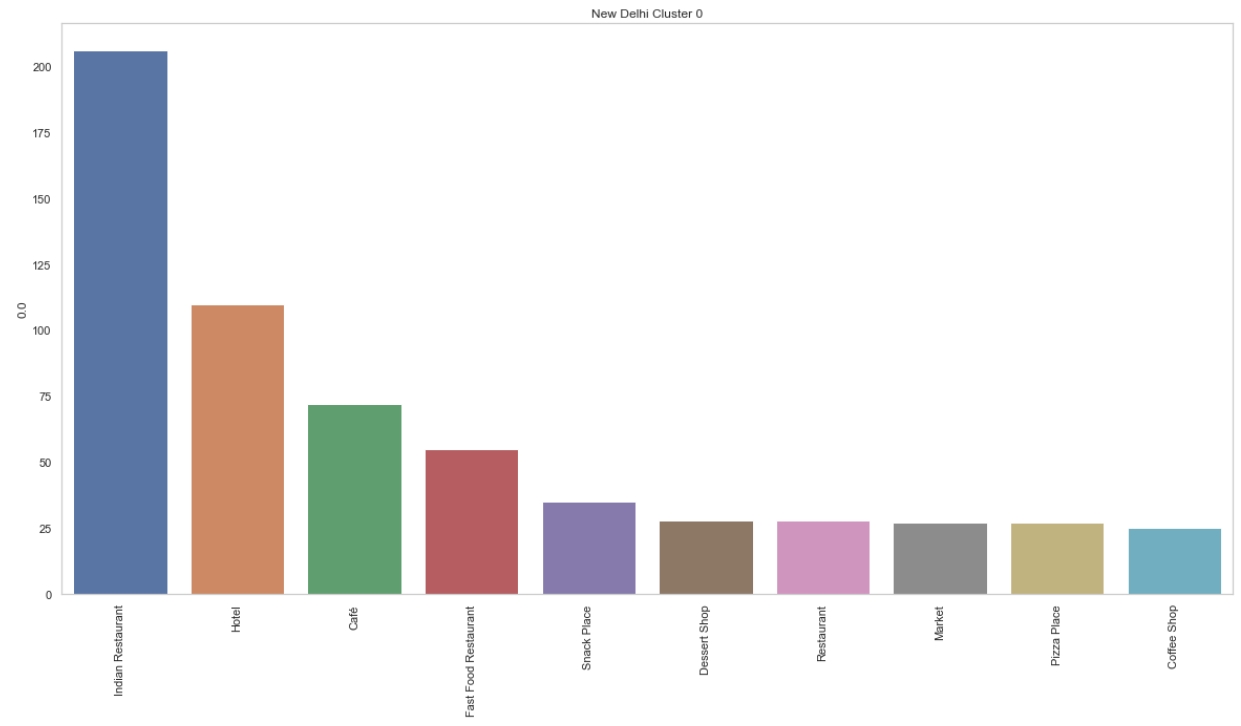
**A folium map showing the clustering of the neighborhoods of Toronto**

**RESULT**

**Exploring the clusters of New Delhi**

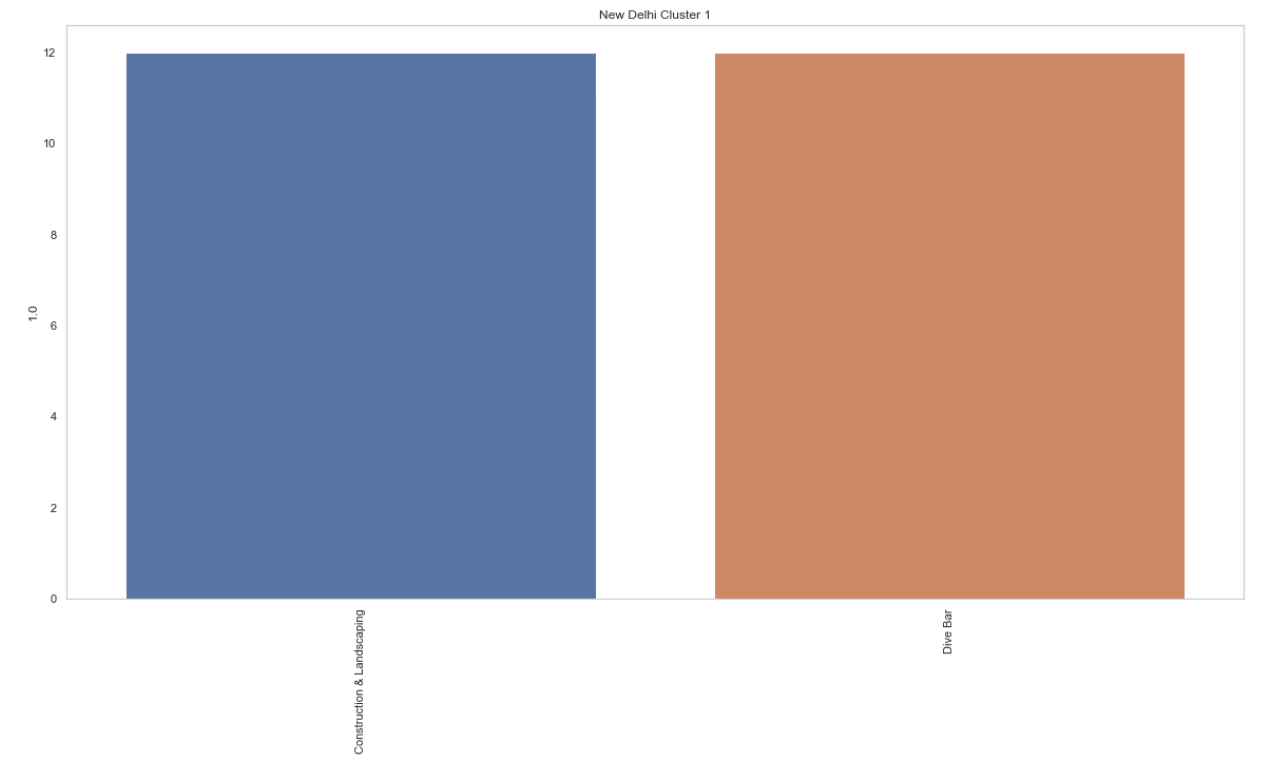
Post clustering of the neighborhoods in both the cities, we can explore each cluster one by one and list down the top businesses functioning in each cluster.

Cluster 0:



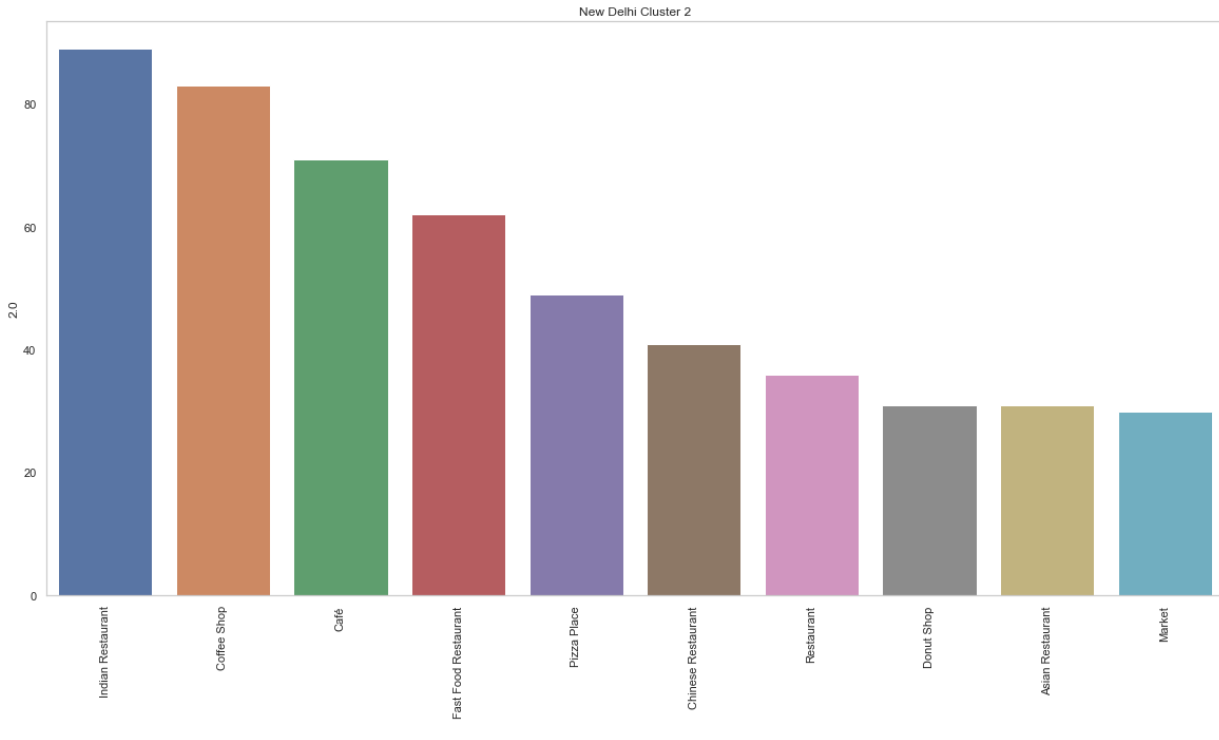
This cluster lists neighborhoods with their top functioning business as Indian Restaurants, followed by hotels and other eateries. This is the biggest cluster of New Delhi with a total of 60 neighborhoods listed in this cluster. The number of businesses in this cluster exceeds that of any other cluster due to the greater number of neighborhoods listed here. This cluster includes the major attractions of New Delhi due to the greater number of hotels that are present and contributing to the economy of this cluster.

Cluster 1:



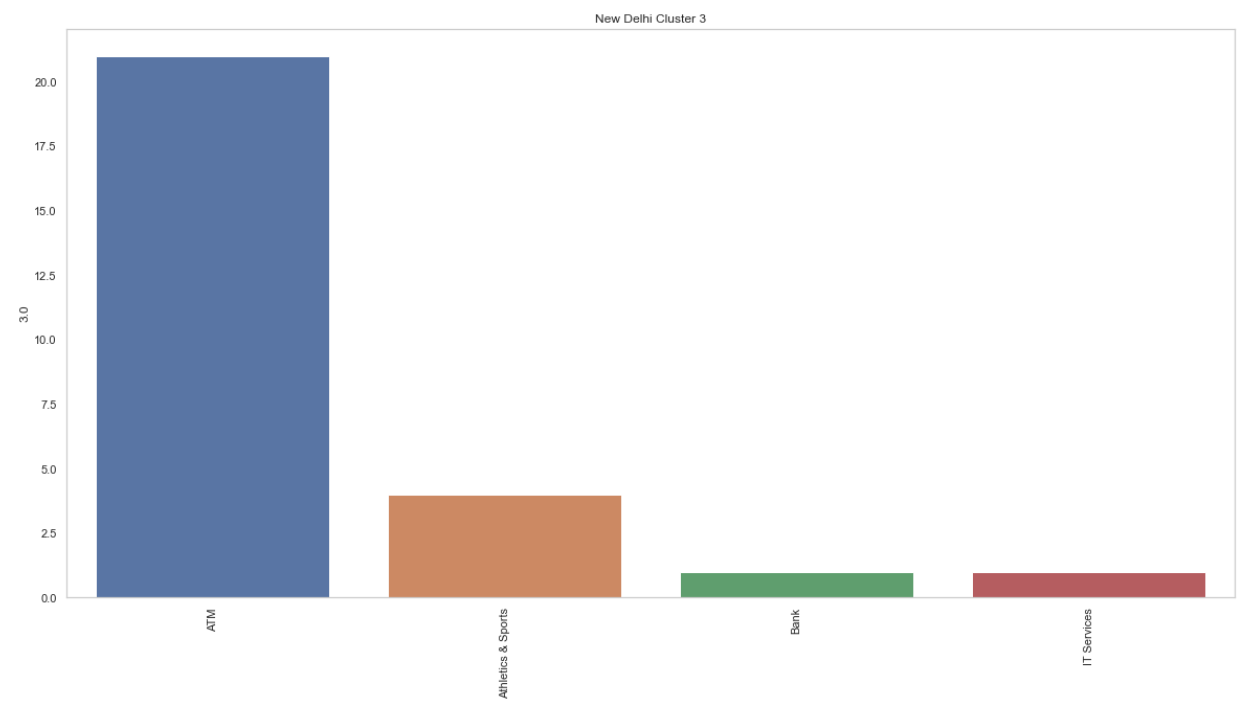
This cluster includes the developing portions of the city with major construction businesses present in this cluster.

Cluster 2:



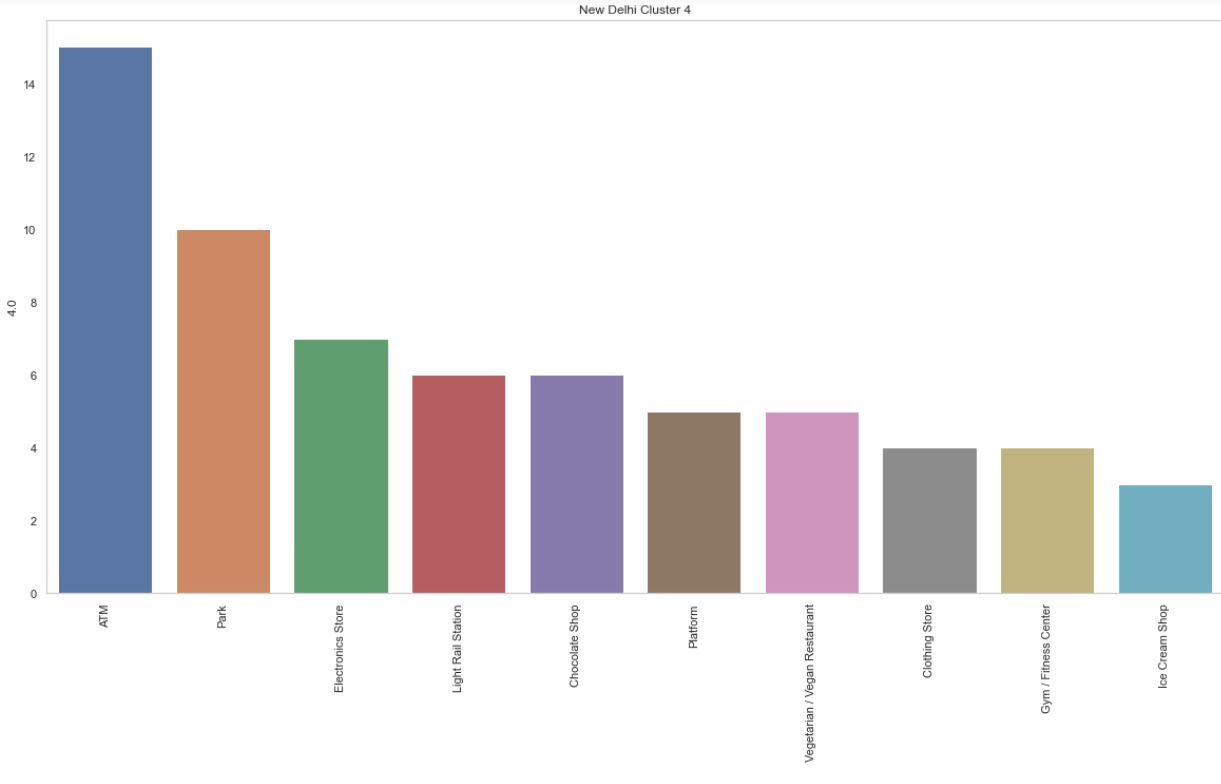
This cluster includes 74 no. of neighborhoods which make up 45% of the city neighborhoods. This cluster runs on eateries and is the major market area of New Delhi. The preference in eateries that the residents of New Delhi prefer is visible in this cluster.

Cluster 3:



This cluster profits on recreational activities and provides bank and IT services.

Cluster 4:



This cluster is a small section of Delhi which includes transport stations, parks and all sort of businesses complementing the transport business.

**Exploring the clusters of Toronto**

Cluster 0:

Chart, bar chart

Description automatically generated

This cluster includes recreational and leisure activities such as beaches and other businesses that compliment it.

Cluster 1:

Chart, bar chart

Description automatically generated

This is one of the major clusters of Toronto. It includes 44 Boroughs in total and the major businesses operating in this cluster include eateries of one form or another.

Cluster 2:

Chart, bar chart

Description automatically generated

This cluster includes 56 Boroughs which make up 54 % of the total boroughs of Toronto. This cluster pertains to businesses found around residential areas such as Grocery stores, Pharmacies, Gas stations, etc.

Cluster 3:

Chart, bar chart

Description automatically generated

This cluster, just like cluster number 0 is a recreational sector that includes Parks and Pool.

Cluster 4:

Chart, bar chart, treemap chart

Description automatically generated

The neighborhood included here includes a simple lounging area and a coffee shop.

**CONCLUSION**

The neighborhoods of New Delhi and Toronto have been segmented into 5 clusters. Each of these clusters represents different parts of the city that help us compare the types of businesses that function and majorly contribute to the economy of the two cities.

The clusters in New Delhi were clearly differentiated based upon the level of activities that pertain to each cluster. The tourism-rich, residential and the developing sections of the city were clearly separated. The bulk of neighborhoods of Toronto, on the other hand, were listed in only two clusters.

The businesses operating in New Delhi are numerically greater than those operating in Toronto, but the population to business ratio of Toronto far exceeds that of New Delhi. New Delhi displays a more tourism-based economy than Toronto and the variety of businesses is greater in New Delhi than in Toronto. This shows that there is a good scope of setting up businesses in New Delhi to serve its vast population.

The most dominant form of Business in both New Delhi and Toronto is that of eateries. While Indian Restaurants and Cafes make up the bulk of them in New Delhi, Coffee shops are dominating in the Toronto food space. The diversity in restaurant types is greater in Toronto than it is in New Delhi, preferably due to the greater ethnic diversity in the population of Toronto.

The marketplace of New Delhi far exceeds that of Toronto. But Toronto is better planned than New Delhi due to the proper ratio of the essential goods store that are in the residential areas of Toronto. This shows that New Delhi can benefit from the setting up of a greater number of essential goods and services stores around its residential areas.