**Clustering the Neighborhood Areas of Toronto**

**Satyam Kumar**

**3rd Feb,2021**

**1. Introduction:**

**1.1 Background:** Toronto is the capital city of Canadian province of Ontario. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. It is one of the International Centre of Business, Finance, Art, Hospitality, Tourism etc. and recognized as most Multicultural and Cosmopolitan cities in the World.

**1.2** **Business Problem:** As mentioned above about the diverse population and business expansion, one can find it hard to determine which sort of business will be suitable for him to begin with in respect to their neighborhood and frequency of the venues around him. The Business in which, one can get more exposure and diversified people with minimum competition to meet with high demands with less supplies of that area.

**1.3** **Approach:** We will be exploring Toronto Neighborhood Venues and Clustering them into Segmentation based on different Radius and Neighbor Clusters to find out how they are relatable to each other and to find out which type of Venue Category will be suitable to begin Business with in different Locations or Neighborhoods.

**2.** **Data Collecting and Cleaning:**

###### First, We will be scrapping Data of Postal Codes with Boroughs and Neighborhood from Wikipedia <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

1. Then, we will be cleaning the data to remove any not assigned boroughs from the data and group it distinct Postal Codes having multiple Neighborhood
2. Then, we will fetch the Latitude and Longitude of the Postal Codes and merge the data with the existing above data to in respect to the Postal Codes.

**3.** **Exploratory Data Analysis (EDA):**

1. First, we plotted the map of all the Postal Codes with the help of Folium Library, by providing the Latitudes and Longitudes to see the dispersion of the location or Neighborhoods
2. Then, we find the data having most frequent venue category in respect to Each Neighborhood and try to analyze it with Bar Chart
3. Then, we went on ahead finding the mean of each venue category in respect of each neighborhood and derived the Top 10 Most Common Venue for Each Neighborhood
4. Then, we used K-Neighbor algorithm for Clustering the Neighborhood on the basis of their relatability to know which place has diversity more look-alike in order to determine or choose the neighborhood to set up the business and category of venue which is popular among there.

**4. Result:**

On the basis of the above modeling and EDA, we find the clustering among the neighborhoods which will give the basic idea of anyone to look for the location area and the type of category of business, one want to proceed and expand with it.

**5. Discussion:**

In this note, the observation we shared is that there are few neighborhoods which is same type of venue categories filled with it and overpowering the whole area like Indian Restaurant being in the single neighborhood and owning the whole business of that area while on the contrary there are neighborhoods with Café or Coffee Shops are available in abundance in some neighborhoods while the rest of the Venue Categories are just can be counted on fingers.

**6. Conclusion:**

On the end note, we will just like to admit, that this was the somewhat the basic idea to get some sort of insights in regard to this business problem but there can also be some more modified ways or techniques to analyze it and come out it more information which can be helpful for the society to access it and nurture on it.