**Recommendation on Indian Restaurant in Toronto**

**Introduction:**

Toronto is a well-developed capital city of Canada, with lots of business opportunities and business friendly environment, it has no issue in attracting many different players into the market. However, that also means the market is highly competitive and as a well-developed city, the cost of doing business is also one of the highest in the country. And thus, any new business venture or expansion in the country needs to be reviewed carefully and strategically targeted so that the return on investment will be sustainably reasonable and more importantly the investment can be considerably less risker.

**Problem Description:**

This is clearly a problem that Indian Restaurant (i.e. ABC Restaurant) needs to review and resolve as part of their new business venture in the country. they have sufficient funds, but they need to choose their first starting location in the country carefully for the points highlighted above and more importantly, if this is successful, the location should allow them to replicate the same success fairly quickly; so, first mover advantage is critical for this business and thereby the choice of location (i.e. neighbourhood) is also important to them.

**Target Audience:**

To solve this problem, data scientist team led by myself has been engaged by ABC Restaurant. The objective is to locate and recommend to the management which neighbourhood will be the best choice to start off their first Indian Restaurant.

**Success Criteria:**

The success criteria of this project will be a good recommendation of the neighbourhoods choice to the management of ABC Restaurant based on Indian population and higher number of residences presented (higher demand) and it should allow easy replication of the business model (similarities among the neighbourhoods).

**Data Description:**

As we need to explore, segment, and cluster the neighbourhoods in the city of Toronto, the Toronto neighbourhoods’ data is key for this project. Unfortunately, the data is for the Toronto neighbourhood data is not widely available on the Internet in the structured format, hence we need to scrap it through an existing Wikipedia page exists that has all the information we need to explore and cluster the neighbourhoods in Toronto. The data should contain the coordinates for each of the neighbourhood in Toronto that will help us to further obtaining more information critical for this project. We will also like to obtain the key information like below; such as number of residences information for each neighbourhood which is one of the key factors for the neighbourhood of choice in the final report. The data needs to be clean up and eventually in a structured format like the example below.

1. Neighbourhood Name

2. Neighbourhood Latitude

3. Neighbourhood Longitude

4. Number of residences in each neighbourhood.

**Data Features:**

We will be leveraging on features in a reliable location information provider such as the Foursquare.com to explore the various types of venues and its categories available in each neighbourhood. We will also need to understand the type of these venues nearby (i.e. within 1500M) in each of the respective neighbourhood. The information obtained per neighbourhood will be as such like below and must be in a structured format so to allow for further computation:

1. Neighbourhood

2. Neighbourhood Latitude

3. Neighbourhood Longitude

4. Venue Name

5. Venue Category

6. Venue Latitude

7. Venue Longitude

**Methodology:**

****

Data scrapping from the Wikipedia page that contains the up-to-date population statistics of Toronto neighbourhoods has been used. This is critical to understand the population of each Toronto neighbourhood which is one of the key elements in the neighbourhood of choice in this project.

Furthermore, we need to know the coordinates and locations of this neighbourhoods, and therefore the geolocator API has been used for achieving this objective. This is important so that we can input this information into the location information provider such as Foursquare.com to obtain venue information in these neighbourhoods, and this is precisely what we have done for it in this project.

We will also use machine learnings techniques such as the K-Means to segment and cluster these neighbourhoods so that we can group them together to understand their similarities. This is critical as we need to recommend to the management the regions of the neighbourhoods of the choice in our recommendation so that ABC Restaurant can easily replicate their business model across multiple neighbourhoods of similarities easily and quickly as part of their business growth plan.

Finally, with all these methodologies, we will then be able to come up with a best recommendation to the management of ABC Restaurant to their problem which is where is the best regions of neighbourhoods for them to first start off to offer their services based on neighbourhood’s similarities,

high population and low competition (i.e. fewer Indian Restaurants.)

**Result:**

With K-Means clustering technique, the top 5 clusters of similar neighbourhoods have been apparent in the result, see below. These clusters are group together based on the similar nearby venues in each of the neighbourhoods. This information is critical so that we can target on the cluster that offer the largest business expansion and growth opportunity as the management of ABC Restaurant is interested to replicate their business model fairly quickly upon success in their first service offering in the selected neighbourhood.

****

With bar chart visualization technique, we can easily tell what the top population are (i.e. higher number of residences) in the neighbourhood cluster. This is also critical as we will like to recommend to the management of ABC Restaurant of the neighbourhood with the higher number of population so that there will be a higher demand for their service offering. The top 10 neighbourhoods with highest number of populations are as follows.

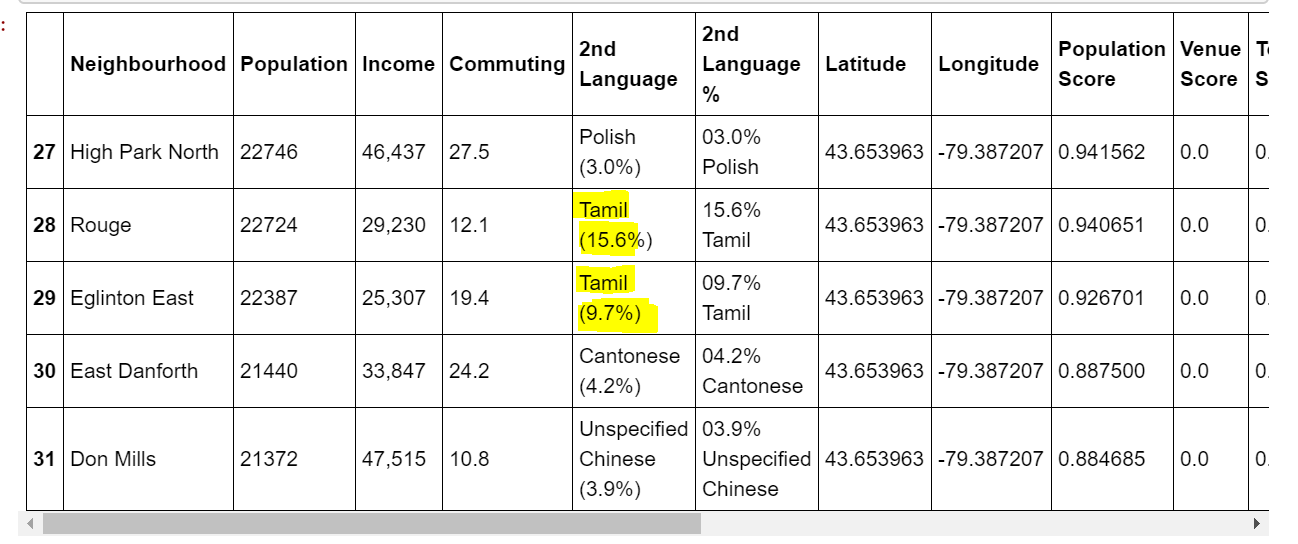
****

With Foursquare.com API, we are also able to leverage on the data to find out the top common nearby venues and their categories in each of these neighbourhoods. This is critical as we want to recommend a neighbourhood whereby the supply is low (lower competition). As shown below, these neighbourhoods have fewer grocery choices available giving ABC Restaurant a higher advantage and chance to succeed upon entry.

**Discussion:**

Based on the result above, the first cluster looks to offer a higher number of similar neighbourhoods that has more Indian population (more specifically south Indians\_Tamil)

Within the first cluster, we will like to recommend a neighbourhood with higher demand and lower supply to give ABC Restaurant a higher advantage and chance to succeed upon their first service offering. Hence, with this in mind, it is apparent that neighbourhood **'Rouge'** looks to be the choice as it is the highest Tamil populated and very few Indian restaurants in the neighbourhood (i.e. close to none for the first few most common venues in this neighbourhood).



It is also apparent that there is a high number of Tamil people in that neighbourhood and hence, we will also like to encourage the management of ABC restaurant to offer Tamil food or related Groceries shop

**Conclusion:**

With that, we have concluded that the best recommendation for ABC restaurant to first offer their services in Toronto will be neighbourhood **'Rouge'** with the key factors to consider such as higher demand, lower competition, easy replication for business expansion. See the recommendation summary below.

Region: First Cluster.

Neighbourhood: **'Rouge'**

Additional Offering: Tamil food or related supplies.