**Peer-graded Assignment: Restaurant Capstone (Week 1)**

# INTRODUCTION

## Description of the Project

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571, it is the most populous city in Canada and the fourth most populous city in North America. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario, while the Greater Toronto Area (GTA) proper had a 2016 population of 6,417,516. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

According to the United Nations Development Programme, Toronto has the second-highest percentage of constant foreign-born population among world cities, after Miami, Florida. While Miami's foreign-born population has traditionally consisted primarily of Cubans and other Latin Americans, no single nationality or culture dominates Toronto's immigrant population, placing it among the most diverse cities in the world. In 2010, it was estimated over 100,000 immigrants arrive in the Greater Toronto Area each year.

In this project we will try to analyse all the **Indian Restaurants** currently present in Toronto's different neighbourhood and find top Indian Restaurant’s based on popularity from its Likes, rating, tips etc and will able to provide an final result/analysis at the end of this project.

# BUSINESS PROBLEM

The objective is to find a suitable location(s) to open an Indian Restaurant in Toronto City, Ontario, Canada. This project makes use of various Data Science and Machine Learning methodologies (K-means Clustering) to provide a Solution to the client.

The project aims to provide a Solution to the Question: ‘**Where should you consider opening an Indian Restaurant in Toronto City?’**

# TARGET AUDIENCE

1. Business personnel who wants to invest or open an Indian restaurant.
2. Budding Data Scientists, who want to refer the analysis.

# DATA SECTION

For this project we need the following data:

a) Toronto Neighbourhood data that contains Borough, Neighbourhoods along with their latitudes and longitudes

1. Data Source: Geospatial Co-ordinates CSV File
2. Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of Toronto.

b) Toronto Postal Code Data that contains Postal codes for neighbourhood present withing Toronto City.

1. Data Source: [*https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M*](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
2. Description: By using this API we will get all the Postal Code information for all the neighbourhood within Toronto City.

c) Indian restaurants Toronto city.

1. Data Source: Foursquare API
2. Description: By using this API we will get all the venues in the Toronto. We can filter these venues to get only Indian restaurants.

# METHODOLOGY

1. We will first use, the data **Beautiful Soup Package** to scrape the Postal Code information of Toronto neighbourhood from the Wikipedia page, and store the same information within **Pandas** dataframe.
2. Then data cleansing will be done to remove unassigned values in Borough and Neighbourhood.
3. We will then import the contents of Geospatial Co-ordinates csv file into the dataframe. We will then use the two data frames created, to create one single Dataframe.
4. Foursquare API will help to get the venues within all the neighbourhoods of Toronto city.
5. Filter out the Venues of all Indian Restaurant present in Toronto city.
6. We will then collect the likes, rating , tips etc for each of the Indian restaurant and then sort them according to their values, again through Four square API .
7. The comparison data and the ranking will be plotted on Bar Chart graphs using matplotlib library, Seaborn library.
8. We will Visualize the data of top Neighbourhood based on the venue categories in a Toronto City map using *Folium Package*.
9. Finally, we will cluster these top Neighbourhood based on the venue categories and use K-Means clustering Machine learning technique to build a model using Elbow point method.

# RESULTS

a) ***Downtown Toronto*** Borough has max no of Indian Restaurant in Toronto City

b) ***Downtown Toronto*** Borough has ***10*** Indian Restaurant, where as, ***York*** Borough has least Indian Restaurant present its count is only **1** in Toronto City respectively.

c) ***India Bazaar, The Beaches West*** Neighbourhood has maximum no of Indian Restaurant with a count of ***6***.

d) We can observe ***Roti Cuisine of India*** of Indian Restaurant got the Maximum Rating. It belongs to ***The Annex, North Midtown, Yorkville*** Neighbourhood and of ***Central Toronto*** Borough.

e) ***Cluster 2*** has most of the ***Indian Restaurant*** followed by ***Cluster 5*** and ***Cluster 3.***

## DISCUSSION

In this project, I tried to use all methods I learned through courses like Data cleansing, Web scraping, Data handling, Exploratory analysis, and obtaining results with Machine learning Algorithms

Analysing that is almost free, you only need the data and a computer. In the end, I wish I had more clear and larger data (like demographic information (Ethnicity, salary information, Cuisine preferences), Financial information (property rents)) about Toronto. We can do a more thorough analysis using this information.

## CONCLUSION

In the end, I reached the goal that I declared in the first section. I think with the map on the Results section a tourist can see a simple guide about restaurants in different districts.

But there is scope in future to fetch more demographic and Financial Data to get more concrete analysis. In total, I hope you enjoyed these results of the Capstone project.