**FINAL REPORT**

**Capstone Project - The Battle of Neighbourhoods**

**Introduction:**

In this project of mine, I aim to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Indian restaurant in Toronto, Canada.

Since Toronto is a big city, it obviously is crowded with a lot of restaurants. So, in this project, I will try to detect locations that are not already crowded with restaurants and also areas with no Indian restaurants in vicinity.

I will use my Data Science and Machine Learning skills that I've learnt so far in this IBM-DATA SCIENCE specialization to generate a few most promising neighborhoods that match the criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

**Problem:**

To find the answers to the following questions:

Q1) List and visualize all major neighborhoods of Toronto, Canada.

Q2) What is best location in New York City for Indian Cuisine?

Q3) Which areas have potential Indian Restaurant Market?

Q4) Which all areas lack Indian Restaurants?

Q5) Which is the best place to stay if you prefer Indian Cuisine?

**Data Description**

## The lists of things or data that I need to know for my project are as follows:

1.list of neighborhoods in Toronto

2.the co-ordinates(latitudes and longitudes) of these neighborhoods

3.the venue data related to the Indian restaurants already present in Toronto

I have decided to use regularly spaced grid of locations, centered around city center, to define the neighborhoods.

## Following data sources will be needed to extract/generate the required information:

1.scrapping of all neighborhoods in Toronto via Wikipedia

2.getting latitudes and longitudes of data of these neighborhoods via Geocoder package

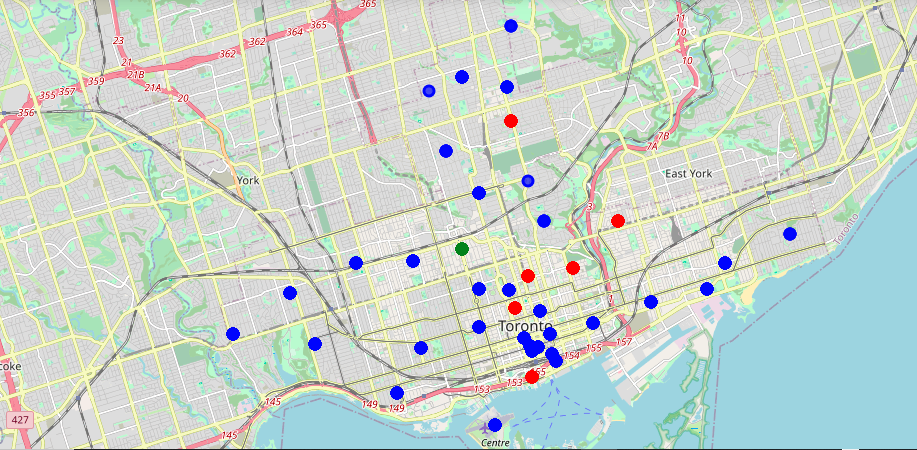
3.using Foursquare API to generate venue data, related to these neighborhoods

**Methodology:**

* Firstly, I’ve looked for all the neighborhoods in Toronto, using the list of neighborhoods from Wikipedia: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* Then comes the web scraping. I did it by utilizing pandas HTML table scraping method as it is easier and more convenient to pull tabular data directly from a web page into the data frame.
* However, it is only a list of neighborhood names and postal codes. And I need to get their coordinates, i.e. the latitudes and longitudes to utilize Foursquare to pull the list of venues near these neighborhoods. And therefore, I used the CSV file provided by this specialization, to match the coordinates of the Toronto neighborhoods.
* After generating these coordinates, I visualized the map of Toronto using the Folium package, to verify whether these are correct coordinates or not.
* Next, I used the Foursquare API to pull the list of top 100 venues within 500 meters radius.
* And from my Foursquare developer account, I was able to obtain my Client ID and Secret, to pull the data.
* From Foursquare, I was able to pull the names, categories, latitudes and longitudes of the venues.
* Then, I analyzed each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.
* Lastly, I used the clustering method: k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well.
* I have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for “Indian food”. Based on the results (the concentration of clusters), I will be able to recommend the ideal/optimal location to open the restaurant.

**Result:**

The results from k-means clustering shows that we can categorize Toronto neighborhoods into 3 clusters based on how many Indian restaurants are in each neighborhood:



# 

# Analysis and Inference

From the tables above, we can conclude that most of the Indian restaurants are in cluster 1, which is around Toronto Islands, Riverdale, Davisville etc.

And, the least number of Indian restaurants are in clusters 2 and 3, i.e. The Annex, North Midtown, Yorkville.

Thus, we can safely conclude that there's a very good opportunity to open an Indian restaurant in or near The Annex, North Midtown, Yorkville.