

# Using Analytical and Machine Learning techniques to find best location to open an Indian Restaurant

## 1. Introduction

### 1.1 Background

For this Capstone project, I am creating a hypothetical scenario for opening an Indian Restaurant in city of Toronto. As we all know out of all population, there lives huge amount of Indians in Toronto. So it will open the gates of a nice restaurant for every Indian foodie living there. And also it might present a great opportunity for the entrepreneur who is based in Canada. Entrepreneur will think of opening the restaurant in the place where there is less Indian restaurant in the neighbourhood. With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

### 1.2 Business Problem

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a brand new Indian restaurant in Toronto, Canada. By using data science analytical methods and machine learning algorithms such as clustering, this project aims to provide solutions to answer the business question: In Toronto, if an entrepreneur wants to open an Indian restaurant, where should they consider opening it?

### 1.3 Target Audience

The entrepreneur who needs a best location to open a brand new Indian restaurant.

## 2. Data

To solve this problem I will be needing the following data:

1. List of Neighbourhoods of Toronto
2. Latitudes and Longitudes of all the neighbourhoods in Toronto
3. Venue data of all the corresponding neighbourhoods including their name, category, longitudes and latitudes.

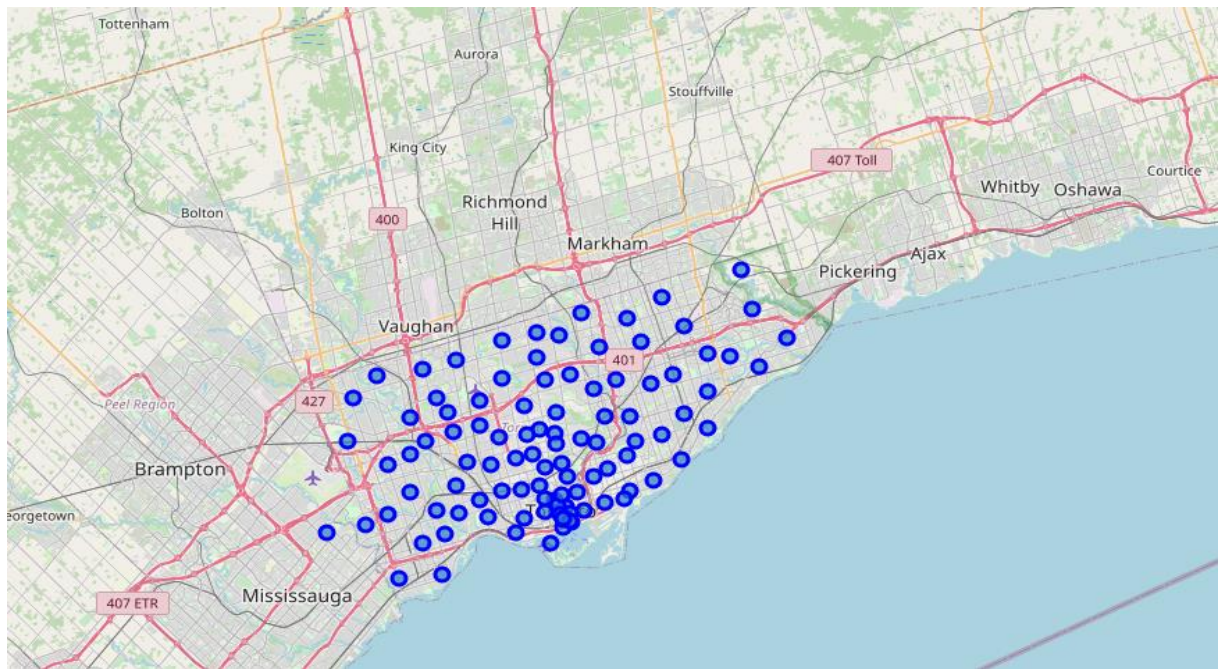
### 3. Extracting Data

- Neighbourhood's data of Toronto will be scraped from Wikipedia.
- Longitude and Latitude of Toronto neighbourhood will be fetched from geocoder package
- Foursquare API will be used to get venue data such as venue name, venue category, venue latitudes and venue longitudes.

### 4. Methodology

First I will scrape the list of neighbourhoods of Toronto, Canada from the Wikipedia page. Scarping will be done using BeautifulSoup and Selenium library. Using these libraries data from the url: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) if fetched and then converted into a DataFrame using Pandas library.

Now my DataFrame contains Postal Codes, Borough and Neighbourhoods data of Toronto. I need to get coordinates corresponding to every Postal code in order to solve my problem. First I will get the coordinates of Toronto using Geopy library. Now I will be needing coordinates of each Postal code, I tried it first using Geopy library but it doesn't work for me so I will be using a csv file containing coordinates data corresponding to each Postal Code. Now merging both the DataFrames based on Postal Code and I get my coordinates for every location in Toronto.



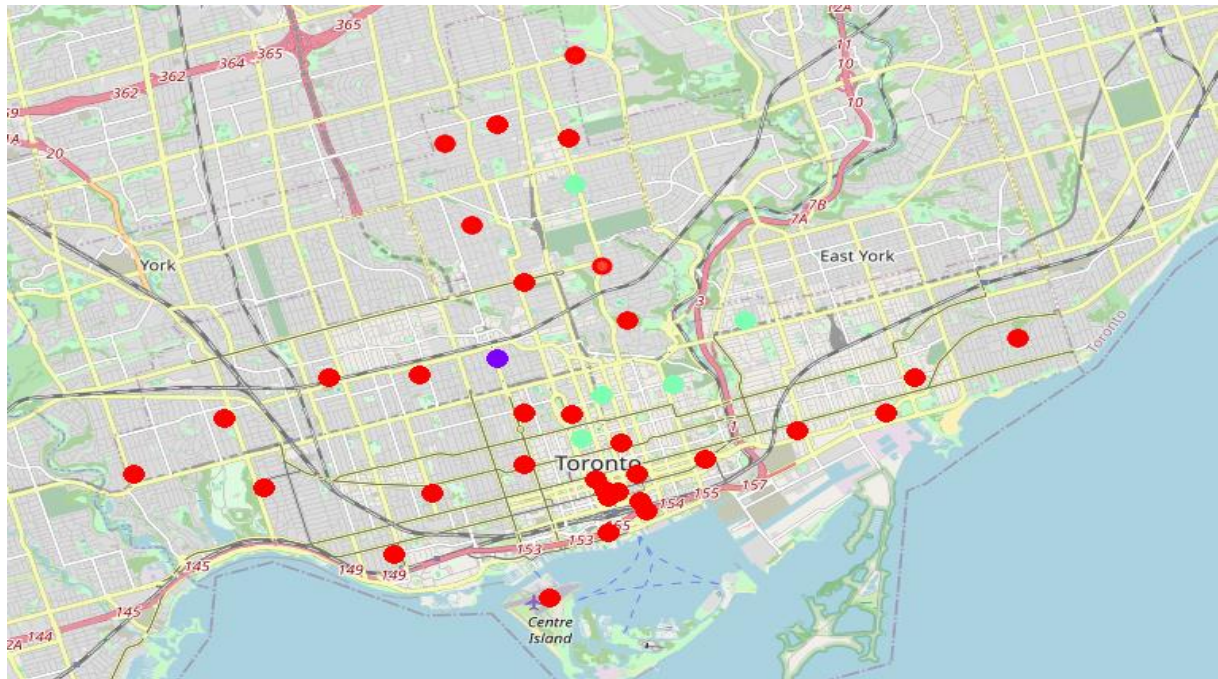
**Location of Neighbourhoods in Toronto, Canada**

After this I will be using Foursquare API to pull the list of venues near all the locations. Using Foursquare API I need to pull 100 venues within 500 m of radius. After REST call to the API I get venue name, venue category, venue latitudes, venue longitudes, etc. , from it. Now I analyse each neighbourhood by grouping the rows by neighbourhood and mean of count of each venue category. This needs to be done for applying clustering algorithm.



**Location of Neighbourhoods which contains Toronto in it**

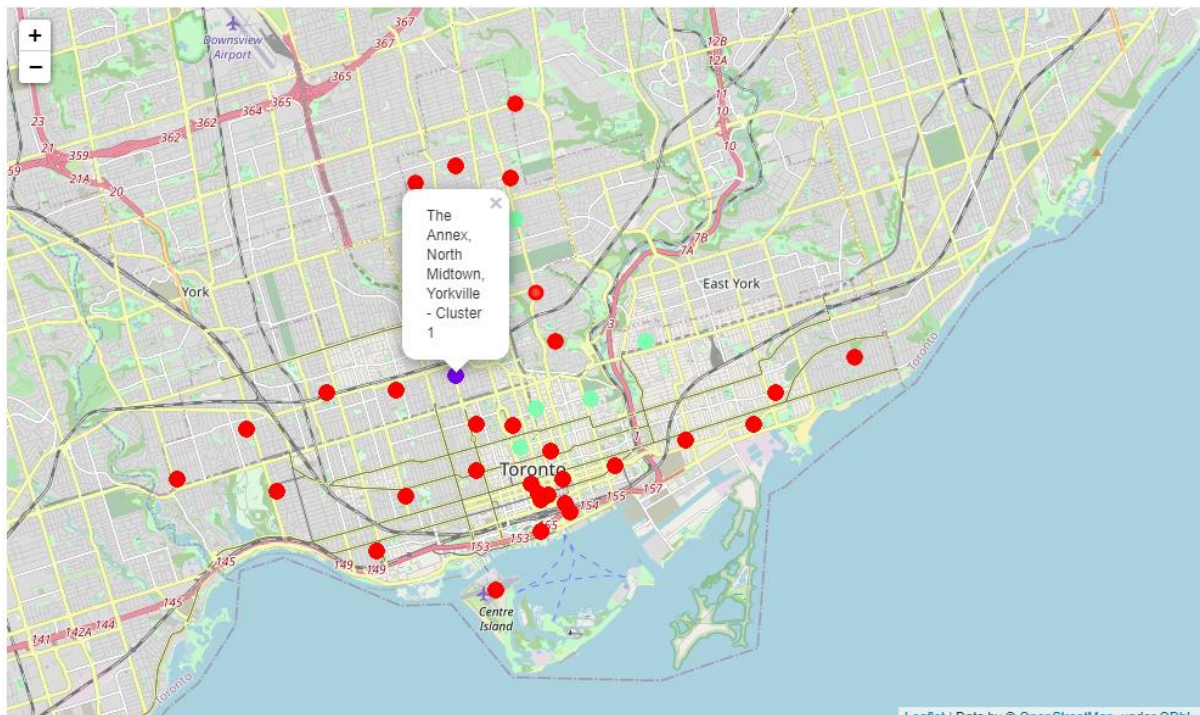
Now I need to fetch only Indian Restaurants from the Venue category column. Now created my new DataFrame which only contained venue category as Indian Restaurants. Now it's time to apply my Machine Learning Algorithm K-Means clustering to it. K-means cluster creates cluster labels to each neighbourhoods based on their location and thus created different clusters. It is one the simplest and popular unsupervised machine learning algorithm. I have clustered the neighbourhood into three clusters based on the frequency of Indian food.



**Three different clusters**



## 5. Results



Result from K-Means clustering shows that we can categorize Toronto neighbourhoods into 3 categories based on count of Indian restaurants in respective neighbourhoods:

Cluster 0: Neighbourhoods with maximum no. of Indian restaurants.

Cluster 1: Neighbourhoods with least Indian restaurants.

Cluster 2: Neighbourhoods with little no. of Indian restaurants

## 6. Recommendations

Thus from the above map it is recommended to open Indian restaurant around The Annex, North Midtown, Yorkville places and if the food quality and ambience is good then the business will do good.

## 7. Conclusions

In this project, we have gone through the process of identifying the business problem, specifying data which required, scraping and processing data, performing machine learning technique and providing recommendation to the entrepreneur.