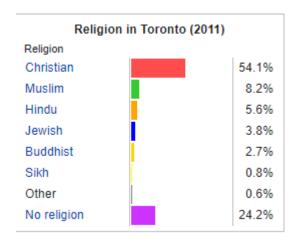# Coursera Capstone

Opening New Indian Restaurant in Toronto City, Canada

By

Santosh Sawant

06.08.2020

Introduction:

Toronto is the capital city of the Canadian province of Ontario. It is the most populous city in Canada and fourth most populous city in North America. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. If we look at the overall religion status, around 15% population belongs to asian community (Fig. pertaing to 2011 as per Wikipedia). The city is home to the Toronto Stock Exchange, the headquarters of Canada's five largest banksand the headquarters of many large Canadian and multinational corporations. Due to this population density of immigramts is more in Toronto than other Canadian cities. Also statistics indicates that, immigrants to Canda from India are highest after China. So There is high probability of successful running of Indian restaurants in Toronto.

Business Problem:

The objective of this capstone project is to analyse and select the best locations in the city of Toronto, Canada to open a new Indian Restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Toronto, Canada, if a popular restaurant chain owner is looking to open a new Indian restaurant, where would you recommend that they open it?

Target Audience of this project :

This project is particularly useful to Restaurant owners, investors looking to open or invest in new Indian restaurant in the Toronto city in Canada.

Data:

To solve the problem, we will need the following data:

- List of neighborhoods in Toronto. This defines the scope of this project which is confined to the city of Toronto in Canada.

- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to Indian restaurants. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) contains a list of neighborhoods in Toronto, with a total of 103 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods. After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare API will provide many categories of the venue data, we are particularly interested in the Indian restaurant category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used