Trying to find a good location to open a Indian restaurant using Data Science and ML Techniques

Introduction

According to a study of cuisine trade, Indian food is the fourth most popular cuisine the world. Its simplicity resulted in its popularity and it is a very diverse cuisine. Most of these restaurants abroad have been opened by the native Indians who have settled abroad and do not want to loose touch with their food and culture, but surprisingly these restaurants are visited more by the international customers than the native Indians. However there are not many Indian restaurants out there in certain places and these places tend to miss an opportunity to explore a beautiful, colourful and rich cuisine. In this project our goal is to find a great location to open an Indian restaurant where it would gain popularity amongst the locals but at the same time not run on loss as the cuisine is not really popular or likeable there.

Business Problem

The objective of this project is to find the perfect place to open an Indian restaurant in Toronto, Canada. Though Toronto has a lot of Indians, there are not many Indian restaurants, I am put in a situation where I want to find suitable locations to open my Indian restaurant. My intention is to find the answer to my business problem “Where should I open my indian restaurant?”. In this capstone project, with the help to data science techniques and machine learning methods (like clustering), we intend to find a solution to our business problem.

Target Audience

The audience that we are targeting are all the Indian households or entrepreneurs who want to open or expand their Indian restaurant in great places as it is a great opportunity for not just entrepreneurs but stay-at-home parents or partners who are not working but are looking for means to earn.

Data Description

To solve this problem, I would be requiring the following data:

1. List of neighborhoods in Toronto. This defines the scope of this project which is confined to the city of Toronto, Canada.
2. Coordinates of these neighborhoods. This is required in order to plot the map and also to get the venue data.
3. Venue data related to our business. We will use this data to perform clustering on the neighbourhoods.

Sources of Data

This Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) contains a list of neighbourhoods in Toronto Canada with a total of 103 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page. Then we will get the geographical coordinates of the neighbourhoods to get the latitudes ad longitudes of the neighbourhood.

This project will also use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

This project 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).