Coursera Capstone

IBM Applied Data Science Capstone

Opening a New North Indian restaurant in Toronto, Canada

By: Akshya Magotra

January , 2020

Introduction

For this venture, I am expecting a speculative situation of an idea Indian business visionary who needs to investigate opening a solid , North Indian eatery in Toronto, Canada.

The thought behind this venture is that there may not be sufficient Indian restaurant in Toronto and it may show an incredible open door for this business person who is situated in Toronto. The business visionary needs to open the eatery close to other Indian eateries since individuals who like Indian nourishment may will in general like North Indian nourishment.

Finding the area to open the eatery is the most significant and critical choice for the business person and this task is tied in with helping him locate the ideal area.

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 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 property developer is looking to open a new Restaurants , where should he consider opening it?

Target Audience of this project

The entrepreneur who wants to find the location to open a reliable North Indian Restaurant

This project is particularly useful to property developers and investors looking to open or invest in new north Indian restaurant's in the Toronto city of Canada. This project is timely as the city is currently suffering from oversupply of restaurants .

Data

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

•   List of neighbourhoods in Toronto . This defines the scope of this project which is confined to the city of Toronto, the city of the country of Canada .

•   Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

•   Venue data, particularly data related to restaurant particular of Indian restaurant . We will use this data to perform clustering on the neighbourhoods .

Sources of data and methods to extract them

- Scrapping of Toronto Neighborhood data via [Wikipedia](#)

- Getting the latitude and Longitude of this neighborhood using the Geocoder package.

- Using Foursquare API to get the venues related to this neighborhood.

- Scrapping of Toronto Neighborhood data via [Wikipedia](#)

- Getting the latitude and Longitude of this neighborhood using the Geocoder package.

- Using Foursquare API to get the venues related to this neighborhood.

Methodology

First, I need to get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipediapage(" https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M ").

I did the web scratching by using the pandas HTML table scratching technique as it is simpler and increasingly advantageous to pull unthinkable information straightforwardly from a website page into an information outline. In any case, it is just a rundown of neighborhood names and postal codes. I should get their directions to use Foursquare to pull the rundown of scenes close to these areas.

To get the directions, it is possible that I can utilize the Geocoder or essentially utilize the CSV document gave by Coursera itself. In the wake of social affair every one of these directions, I imagined the guide of Toronto utilizing the Folium bundle to check whether these are right organizes. Next, I utilize Foursquare API to pull the rundown of top 100 settings inside 500 meters span. I have made a Foursquare designer account prior so as to acquire a record ID and API key to pull the information.

From Foursquare, I am ready to pull the names, classifications, scope, and longitude of the settings. With this information, I can likewise check what number of exceptional classifications that I can get from these scenes.

At that point, I dissect every area by gathering the lines by neighborhood and taking the mean on the recurrence of event of every scene class. This is to get ready bunching to be done later.

Here, I made a support to explicitly search for "Indian Restaurants" on the grounds that the quantity of results for "North Indian Restaurants" was less.

Ultimately, I played out the bunching strategy by utilizing k-implies grouping. K-implies bunching calculation distinguishes k number of centroids, and afterward dispenses each datum point to the closest group while keeping the centroids as little as could be allowed. It is one of the least complex and famous solo AI calculations and it is profoundly appropriate for this undertaking too. I have bunched the areas in Toronto into 4 groups dependent on their recurrence of event for "Indian nourishment".

In light of the outcomes (the grouping of bunches), I can ready to prescribe the perfect area to open the eatery.