IBM Data Science Coursera Capstone Project Report

Report By:

Anurag Agrawal

Introduction:

For this Capstone Project I am creating model for a restauranteur to open an Indian Restaurant in Toronto, Canada. Indian food is very popular in Canada because a large number of immigrants. As a restaurateur I must be convinced that my restaurant will make profit in the competition. With this purpose, I will use machine learning techniques to find an ideal location to start my restaurant.

Business Problem:

This purpose of this project is to find an ideal location to start an Indian restaurant in Canada. In this project classification methods will be used to address the problem of finding an ideal location.

Target Audience:

Restaurateur who wants to start an Indian restaurant in Toronto.

Data:

To solve this problem we will need:

List of neighborhoods in Toronto.

Latitude and Longitude of the neighborhood

Venue data of Indian restaurant.

Extracting Data:

Scrapping Toronto Neighborhood via Wikipedia

Getting Latitude and Longitude via geocoder package

Venue Data from the four square API

Methodology:

Firstly, it is very important to collect and tidy the data. Hence, I have first collected the data of the neighborhoods in Toronto using Wikipedia. These neighborhoods contains the postal codes and neighborhood data. However, collecting the data for the coordinates was very crucial. This data was

collected using the Foursquare Developer API. Next, I used the foursquare API to pull the venues in the radius of 500 meters. (FourSquare developer account is required for getting the data). Next thing, I was only concerned with Indian restaurant and pulled the data using the API. Then, I reorganized the data of Indian restaurant a have a look at it. Lastly, I performed the clustering algorithm (k-means clustering algorithm). This algorithm will generate a centroid and cluster the data based on these centroid into their respective cluster (into 3 cluster in this project). Based on this data it will be easy to recommend the place to open a new Indian restaurant.

Results:

The results show different number of Indian restaurant in different neighborhoods. These neighborhoods are classified into 3 clusters; each cluster represents the number of indian restaurants in different neighborhood.

Recommendation:

As we can see that the lowest number of Indian restaurants are in cluster 1. Hence, the competition is low in cluster 1 to open an Indian restaurant. Hence, it would be profitable to start an Indian restaurant at cluster 1 from a data stand point. The restaurant should be open at The Danforth West, Riverdale because of less clustering.

Limitation:

This project is limited to only one factor. The restaurant preference can be depended upon various other factors such as the taste and preference of the consumer itself, the income of people in different location. Hence, there lies a prospect of future research to make this model more robust by including other factors into consideration as well.

Conclusion:

In this Capstone project, I have introduced the business problem, extracted the data from various sources and created a k-means algorithm and gave my suggestion on the problem using the data and clustering methodology.

References:

List of neighborhood in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

ForeSquare Developer API: https://foursquare.com/developers