Coursera Capstone Project on the battle of Neighborhoods

About the city of Toronto

Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario’s northwestern shore. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower. Toronto also has many green spaces, from the orderly oval of Queen’s Park to 400-acre High Park and its trails, sports facilities and zoo.

Now we want to concentrate on the business problem that we want to solve and that is to find the most optimum location or neighborhood to set up a new Indian Restaurant in Toronto. Is it possible to predict the success of a new restaurant even before establishing it? Who would benefit from this Project?

* In this Project people who are looking to invest into Restaurants in Toronto region will be benefited.

Let's see how?

The Steps taken for this project:

1. Business Understanding
2. Analytical Approach
3. Data Requirement
4. Data Collection
5. Data Understanding
6. Data Preparation
7. Modeling
8. Conclusion or Results

1. Business Understanding

Before starting the project we will first define our goal and objective for the project.

Goal

* To find the optimum location in Toronto to set up a new Indian Restaurant.

Objective

* To find the all the locations in Toronto
* Find the venues in the neighborhoods
* Select one of the locations from the list having most venues and popualation density.

2. Analytical Approach

* The analytical approach would be clustering the locations as by clustering we will get the maximum number of density of venues and population

3. Data Requirements

* We will get our required data of toronto from internet and then from a a website we can pull the data in the form of a table

4. Data Collection

* We need to collect two types of datas one for the neighborhood and second for the coordinates of those neighborhoods

5. Data Understanding

* In this step we will focus on which values should be selected and which values are to be ignored.

6. Data Preparation

* Here we focus on getting the most useful data by adding,removing or replacing the data as required in the table.
* Now we go to the next step of using Foursquare to get all the venues around the Downtown Toronto region.

7. Modeling and Clustering the Neighborhoods of Toronto

* Now we will be fitting all our data into the K-means clustering model to get the clusters of our neighborhood.

8. Conclusion or Results

* We can say that cluster 2 and cluster 5 are the ones with the highest number of venues but cluster 5 has more coffee shops than restaurants so we will set up an Indian restaurant in cluster 5 as cluster 5 has more population density and the frequency of restaurants is lower than coffee shops.
* In cluster 2 we can see that even though coffee shop and café are most common but there are Restaurants as well compared to the cluster 5 where frequency of Restaurants was not so prominent.