Coursera IBM Data Science Professional Certificate

Capstone Project

The Battle of Neighborhoods

Toronto

Potential areas to open a New Indian Restaurant

Introduction

Toronto is a highly multicultural city and has residents from various nationalities. One of the most important time out for the people is eating out with friends and family. People from different nationalities like to visit the restaurant of their ethnicity as it makes them feel closer to home and their culture, apart from satisfying their taste buds. Hence, restaurant owners and want-to-be owners are always on constant lookout for places where they can set up the restaurants. Toronto has a high Indian diaspora and the people are scattered through out the city. Although, some areas cater well to the diaspora in terms of restaurant offerings, others lack them. This report tries to highlight potential areas where such restaurants can be set up.

Objectives

Location is the biggest investment for a restaurant and the choice can define whether a restaurant will be successful or not. Any investment requires careful evaluation of all variables and decide the optimum scenario. The purpose of this report is to highlight potential areas based on available data to aid in such a decision making. It tries to find under-served areas which can be potential markets.

Limitations

A lot of considerations go into the selection of venue and this report focuses on one of the aspects only. This report also highlights the potential areas and further evaluation requires careful evaluation with other variables. Also, only areas are suggested and not their potential.

Data requirement

Following data is required to analyse the business problem:

- List of neighbourhoods in Toronto. The scope of this report is throughout the Greater Toronto Area, which combines neighbour cities and have great inter-connectivity.
- Latitude and longitudes of the neighborhoods. These co-ordinates will be required to plat the data on the map and also analyse them.
- Venue data, which provides different available amenities. Indian restaurants will be selected among them and then will be clustered to analyse the results.

Sources of Data

- Postal Codes, Borough and Neighborhood Data is available on Wikipedia.
- Latitude and Longitude information can be found out using Geoencoder or downloaded from here. In this report, downloaded data is used as Geoencoder is not very reliable.
- Venue information can be gathered using Foursquare API. A personal Foursquare developer account is used. Although, it gives limited access, it is enough to satisfy the requirements of this report.

Methodology

Wikipedia page is scraped for data on Postal codes, Borough and neighborhood data using Pandas. Data is cleaned by removing rows with null values and grouping the neighbourhoods as per postal codes. This data is then combined with co-ordinates information to tabulate a dataframe with Postal Codes, Borough, neighbourhood and corresponding co-ordinates. Data from Foursquare is downloaded for Toronto and then filtered for Indian restaurants only. These restaurants are then clustered together as per neighbourhoods into 5 different clusters. For clustering K-means clustering methodology is used. Each cluster is then analysed for the number of restaurants currently operating there. Areas with low densities offer a potential investment opportunity to open a new restaurant. Each cluster is visualised on the map of Toronto and will be provide in the report to aid in decision making.