Coursera Capstone: Analysing Toronto Neighbourhoods to Determine the Locations to Open New Shops

Introduction

This project is to help the companies and entrepreneurs that are planning to expand their business, to open new shops or franchise in Toronto, Canada in an optimal manner. However, the majority of the companies and entrepreneurs are facing a problem, what is the best type of shop to open in certain locations of Toronto, Canada. Moreover, companies and entrepreneurs also very concern about the number of competitors around the vicinity of their business. In order to solve this problem, the data about the shops of each neighbourhood will be collected to show the current shops at each locations, how many similar shops there are, demand of the shops, and what type of shops. The target audience of this report will be the companies, entrepreneurs or people that interested in expanding their business, open new shops or franchise in Toronto, Canada.

Data

In order to solve this problem, the data that consisted of postal codes, neighbourhood and borough of Canada will be collected. The number of shops in each neighbourhood will be collected through Foursquare API by combining with the data that consisted of postal codes, neighbourhood and borough of Canada. There are 103 observations in the data. After combining the data, the data will be show what shops are in each neighbourhood and what type of shops are in high demand. For example, location A having cafe A, pizza, gym, cafe B, cafe C, Japanese restaurant, gym and more. As we can see from the list, cafe is the most popular in the location A, which is high demand but also many competitors. The end result will be showed in the list or table format, so that it is easy for people to understand and see it.

Data that consisted of postal codes, neighbourhood and borough of Canada : <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

Methodology

To help get to the results exploratory data analysis and machine learning is used to predict the rank of the most frequently visited types of cuisine in each neighbourhood.

Data Cleaning

Both the borough and neighbourhood columns contained missing values that needed to be filled in. For the missing values in the borough column, these rows were ignored and only rows containing a borough were considered. However for missing values in the neighbourhood column the value listed under borough will be assigned in place of the missing value. For example, if there is no neighbourhood listed for a postcode in Scarborough, then Scarborough is assigned to the neighbourhood as well.

Data Transformation

The neighbourhoods that having the same postal code area will be combined together with a comma and put in the same postal code area. For example, the postcode M1B contains both the neighbourhoods of Rouge and Malvern so these are combined into one row as ‘Rouge, Malvern’. In addition, the Geocoder API was used to acquired to the longitude and latitude coordinates for each postcode in order to later utilize the Foursqure API. The number of venues associated with each postcode were then acquired using the Fourquare API, and were subsequently ranked and added to the table.

Machine Learning

The machine learning algorithm that was used was K means clustering. After manually adding the number of desired clusters, in this case 6 clusters were chosen, the algorithm identifies a centroid and predicts the location of each new data point based on its similarity to the existing data points. This keeps iterating until the position of the centroids is stabilized and 6 clusters are created.

Result and Discussion

The final result is in the table format which listed out all the top 10 most common shops in that location by the postal code area, borough, neighbourhood, longitude and latitude. The user able to understand the table easily by looking at the location they want. The final result may not be accurate in the future as there are many new shops, shops closed down, changing of demand and more factors will affect the locations. Hence, some update needed to do every year in order to help the companies and entrepreneurs for their business.

Conclusion

In conclusion, this report able to help the companies and entrepreneurs that planning to expand their business, open new shops or franchise in Toronto, Canada. This will increase the job opportunity and attract many customers to visit.