Introduction & Business Problem

In this project, we will explore the city of Toronto and determine which location is optimal for a new restaurant.

The city of Toronto, as the most populous city in Canada, is very multicultural and provides many opportunities for business. However, it can be very competitive at the same time to start a new business. It is very important to analyze the market and the environment of Toronto before any business decision. There are many types of restaurants in every neighbourhood in Toronto, such as French, Italian, Japanese, etc. This project will provide insights to stakeholders about which neighbourhood is the best location to open a new restaurant and to gain the highest return in Toronto. This report is targeting a group of stakeholders who are interested in opening a new restaurant in Toronto, Canada.

Data

Various factors need to be investigated in order to determine the best location, including but not limited to:

- Toronto population;
- Demographics;
- Segmentation of the borough and all existing venues in every neighbourhood like restaurant, supermarkets, schools, parks where people flow rate is high;
- The saturation of the market in specific neighbourhood

Following data sources will be used to extract the required information:

For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

For the geographical coordinates of each postal code: http://cocl.us/Geospatial_data

For the number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API: https://foursquare.com/explore?mode=url&ne=44.418088%2C-78.362732&q=Restaurant&sw=42.742978%2C-80.554504

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

- Name of the venue
- Venue Latitude
- Venue Longitude
- Venue Category

Methodology & Exploratory Data Analysis

Our main goal of this project to find the optimal location for new restaurant business in Toronto. Different data set mentioned above are used. K-means Clustering algorithm is used. K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the K-means clustering algorithm are: The centroids of the K clusters, which can be used to label new data Labels for the training data (each data point is assigned to a single cluster). K-mean is used to identify clusters of neighbourhoods in the city of Toronto and to help explore the venues within them.

Postal code data of Toronto city will be scraped from Wikipedia and is transformed into a dataframe. The data file that contains the geographical coordinates of Toronto neighbourhood will be also transformed into a pandas dataframe. These 2 data frames will be merged together with details of post code, borough, neighbourhood, latitude and longitude. These data will be used to get venues data from Foursquare in order to explore neighbourhood in city of Toronto. We will also use Geopy and Folium libraries to create a map of Toronto with neighbourhood shown on top.

K-Cluster algorithm is used to come up with 5 different clusters in Toronto with similar set of Venues, with an assumption of 5 clusters. We will explore each cluster and determine the discriminating venue categories that distinguish each cluster. We will also identify the clusters and neighbourhoods with highest number restaurants and their types.

Result and Discussion

There are 4 boroughs and 74 neighborhoods inside the city of Toronto. Although there are so many restaurants in Toronto, there are areas near the city center but with lower amount of restaurant. In neighbourhoods such as Adelaide, Yorkville, Harbourfront, and TD centers which have significantly higher restaurant density, there is probably a higher demand for a new restaurant, regardless of the competition between similar types of restaurants. Based on the clustering, Cluster 0 and Cluster 4 have higher number of restaurants than rest of the clusters. In neighbourhoods such as University of Toronto, the restaurant density is low, which means lower market saturation. It is also a very potential location for a new restaurant because there are so many students who probably don’t have time to cook and need to eat out. The location for a new restaurant will also largely depends on the type of the restaurant itself.

Conclusion

This project identifies the areas in Toronto with different number of venues (restaurant specifically), which can help stakeholders in narrowing down the options for finding optimal location for a new restaurant. The restaurant density (numbers of venues) from Foursquare

data was calculated to identify which neighbourhoods have the higher number of restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.
However, this project is performed on limited data sources, which may not provide enough insights. In order to decide an optimal location for a new restaurant, more data sources can be explored such as the demographics, the availability of schools and gyms and office buildings and parks, and the popular cuisines in each borough or neighbourhood.