**Analyzing the Business Viability of a Location for a Restaurant**

Robert Hammond

March 27, 2020

# Introduction

## Background

Restaurant is one of the very viable and profitable business ventures that an investor can invest in, provided certain success-determining factors have been satisfied. One major determining factor is Location. This is also the factor that is tricky to assess. Unless you are very familiar with the particular location under review, it might be difficult to ascertain if the location will be good for the business or not. In order for the restaurant business to thrive, the classification of the location is very important, is it an area marked for Residential or Commercial use. Also the kinds of businesses, organizations or properties in proximity matter (Schools, Hospitals, Homes, Restaurants, Offices, etc.). Having this kind of data or information before setting up a restaurant, gives the business a better chance of survival.

## Problem

Unless you are very familiar with the particular location, it might be difficult to predict if the named location will be good for the running of a restaurant or not. This project seeks to produce the required result in the form of information necessary for the decision-making phase.

## Interest

Definitely, Investment Consultants, Investors, Business Owners would be very interested in the ability to determine the suitability of a location for the setup of a restaurant.

# Data acquisition and cleaning

## Data sources

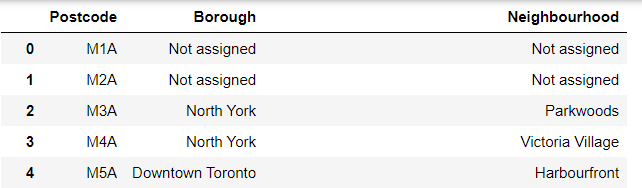
The data for analysis can be found from three datasets.

These three datasets are listed below;

1. The Boroughs and Neighbourhoods data of the city of Toronto was scraped from the following Wikipedia page [here](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).
2. The Geospatial data which contains the coordinates of the various boroughs and neighbourhoods was downloaded from the link [here](http://cocl.us/Geospatial_data).
3. The Foursquare API was utilized as the third dataset, to explore the boroughs and neighborhoods.

## Data cleaning

Two of the data downloaded or scraped from two of the sources were merged into one table. Figure One (1) shows the scraped data from the Wikipedia page. There were a lot of missing values from the “Borough” and “Neighborhood” features denoted by the value “Not assigned”. I decided to only use data with complete records or rows, so I deleted records with missing values.

  
Figure One (1) shows the scraped data from the Wikipedia page.

I checked for duplicate records and removed them from the datasets. The feature on which the two datasets were merged had different names, so they were both renamed to PostalCode.

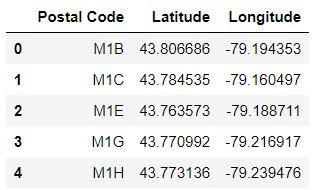


Figure Two (2) shows the downloaded data.

Both datasets were merged on the PostalCode feature. The result is shown below in figure three

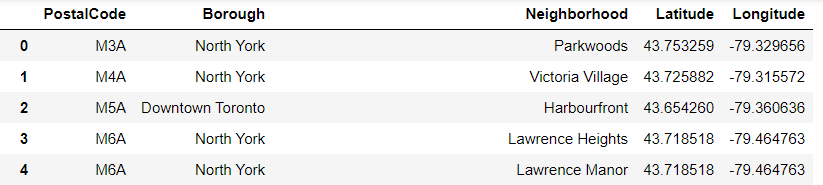


Figure Three (3) shows the merged data.

I utilized the Foursquare API as the third dataset. This API helped to explore the various Boroughs and Neighbourhoods in the City of Toronto. For the purpose of this project, I filtered “Downtown Toronto” as the borough and chose “Harbourfront” as the neighbourhood to explore and analyze.

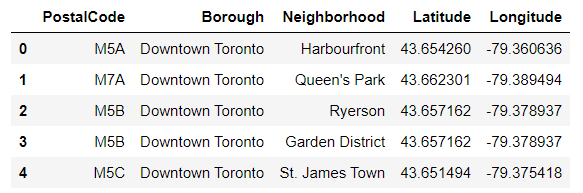


Figure Four (4) shows the borough filtered to “Downtown Toronto”.

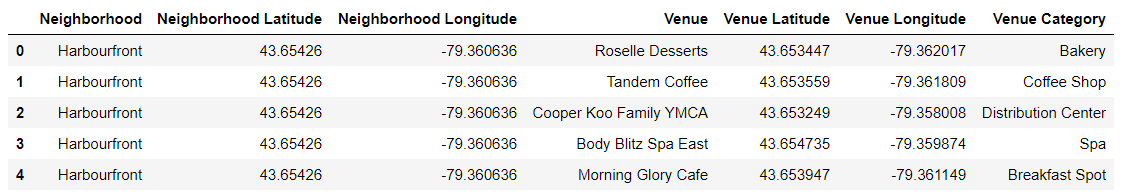


Figure Five (5) shows a section of Venues within the “Harbourfront” neighbourhood.

## Feature selection

After data cleaning for the first two merged datasets, there were 210 samples and 5 features in the data. Upon examining the meaning of each feature, it was clear that all 5 features were important, so all features were selected.