**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

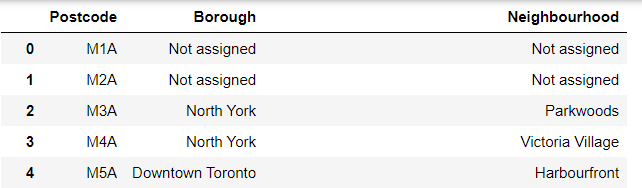
The data for analysis can be found from three datasets.

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.

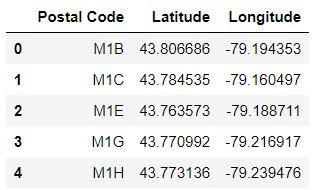


Figure Two (2) shows the downloaded data.

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

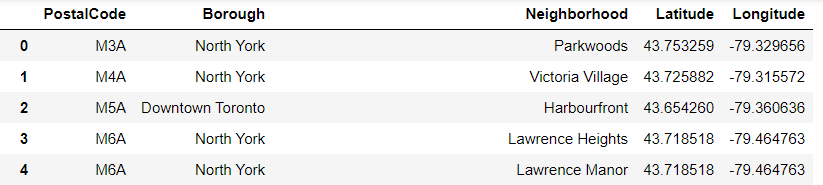


Figure Three (3) shows the merged data.

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.

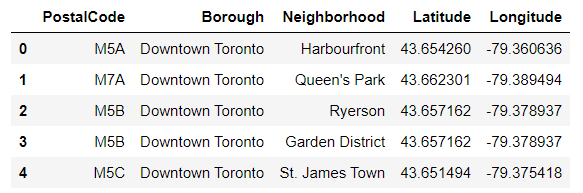


Figure Four (4) shows the borough filtered to “Downtown Toronto”.

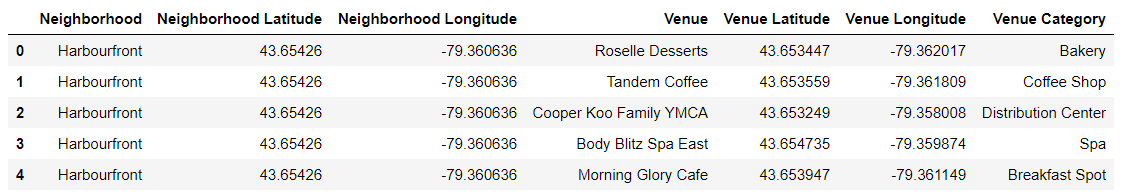


Figure Five (5) shows a section of Venues within the “Harbourfront” neighbourhood.

## Feature selection

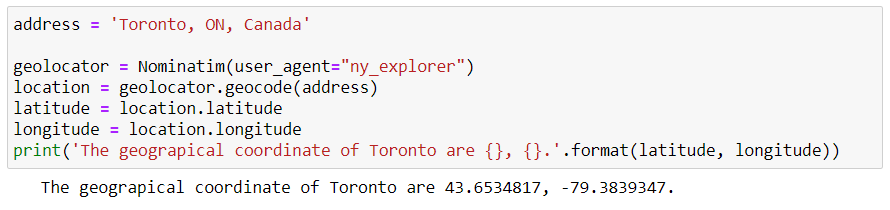
After data cleaning for the first two merged datasets, there were 210 samples and 5 features in the data. Upon examining the meaning of each feature, it was clear that all 5 features were important, so all features were selected.

# Methodology

After identifying and analyzing the problem, the next step is to detail the methodology/processes needed to arrive at the expected solution. I have already described how I acquired and cleaned the data, and from which sources. Please refer to Chapter two (2).

Now for the next part, I used Geocoder from Geopy library to get the latitude and longitude values of the City of Toronto.

Then a map of Toronto is created with its neighborhoods superimposed on it, using the Folium Library.

  
Figure Six (6) shows the code to extract the longitude and latitude of the City of Toronto.

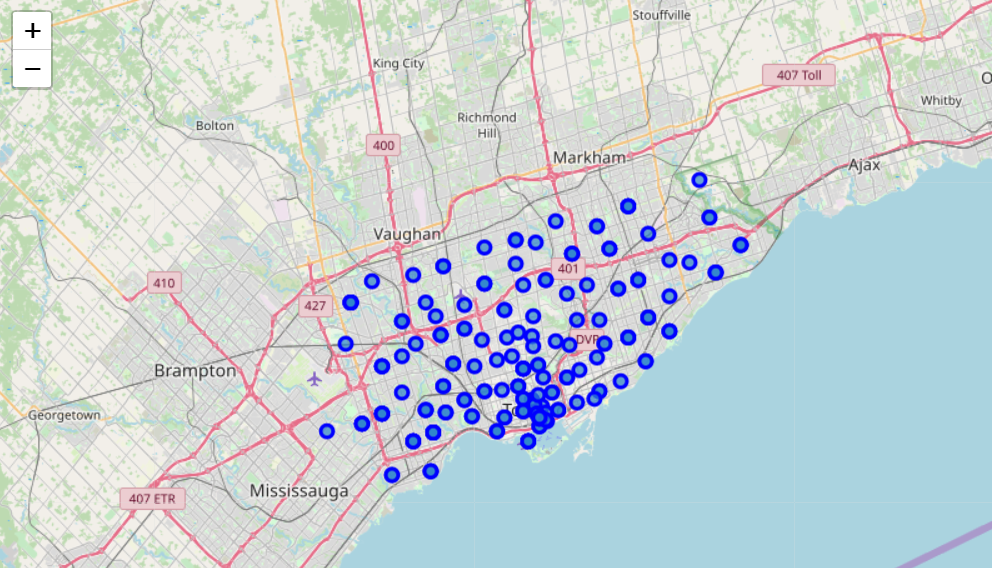


Figure Seven (7) shows the Map of The City of Toronto with the various Neighbourhoods plotted on it.

The next step is to define the Foursquare API credentials and version.

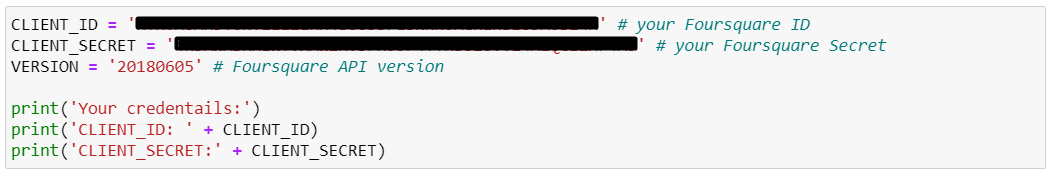


Figure Eight (8) shows the code for defining the API credentials for Foursquare.

Let’s pick the first Neighbourhood and explore it.

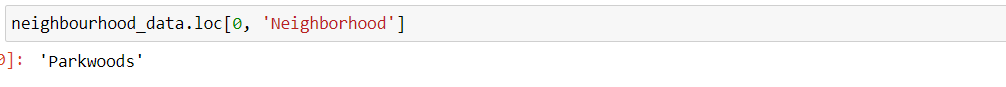
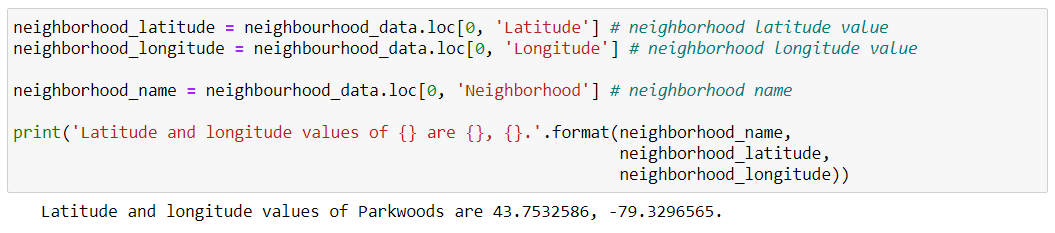


Figure Nine (9) Get the name of the first Neighbourhood in the dataframe  
  
  
Figure Ten (10) Get the Neighbourhood’s Latitude and Longitude.

#### Now, let's get the top 100 venues that are in Parkwoods within a radius of 500 meters.

We first, make a url request to the Foursquare API and receive a response in JSON format.

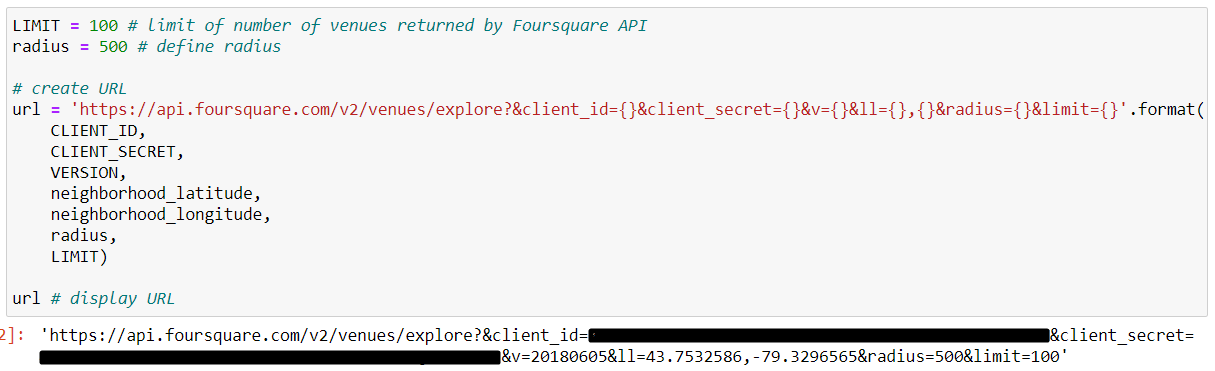


Figure Eleven (11)

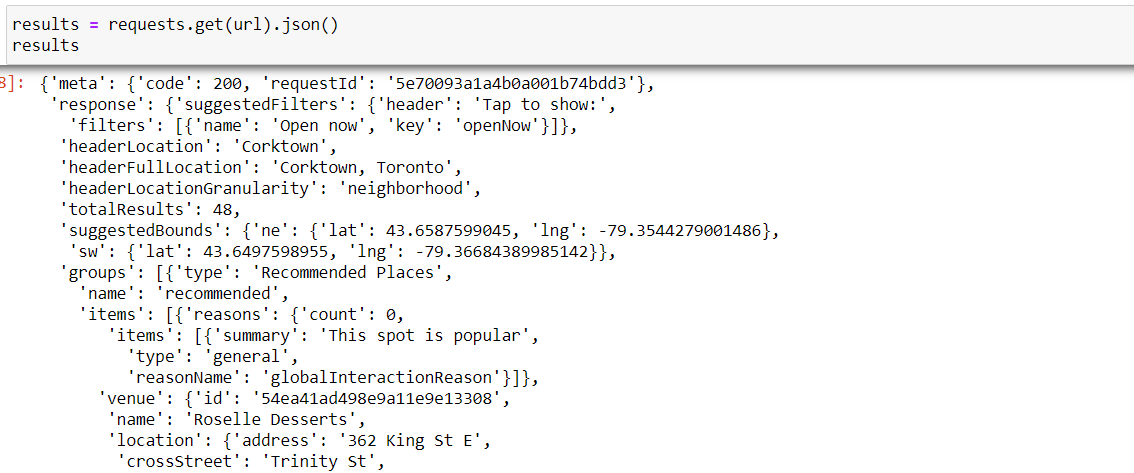


Figure Twelve (12)

Next, we extract the categories of the returned venues. We then clean the returned json and structure it into a pandas dataframe.

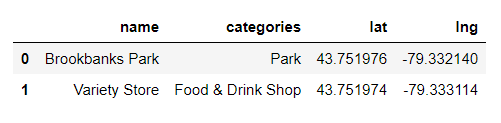


Figure Thirteen (13)

We next run the same process for all the Neighbourhoods in Toronto City, and below is a section of the dataframe results.  


Figure Fourteen (14)

In total, 4253 Venues were returned for the City of Toronto. These venues are grouped by their Neighbourhoods in the Figure below.

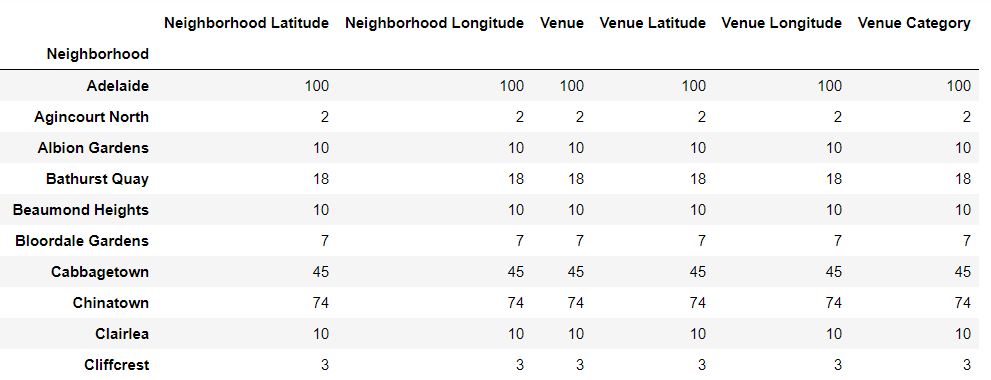


Figure Fifteen (15)

270 Unique Venue categories were returned from the City of Toronto.

Next, let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category.

We then print each neighborhood along with the top 5 most common venues

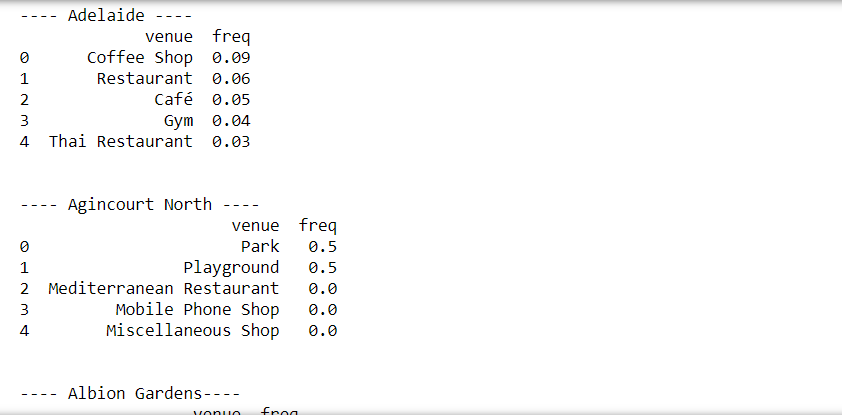


Figure Sixteen (16)

Now let's create the new dataframe and display the top 10 venues for each neighborhood.

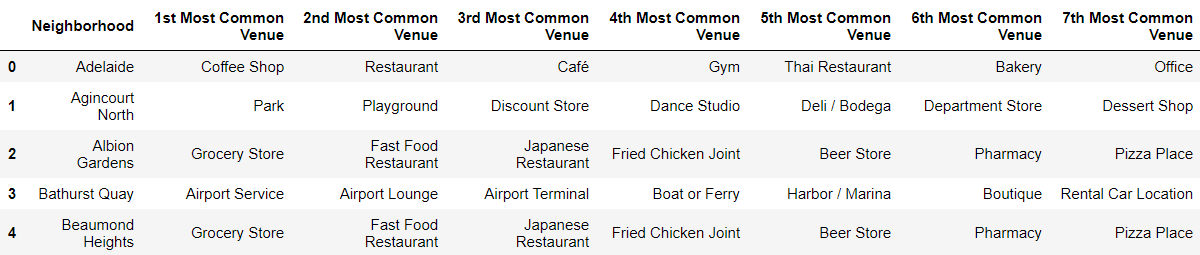


Figure Seventeen (17)

We will now run k-means to cluster the neighborhood into 5 clusters.

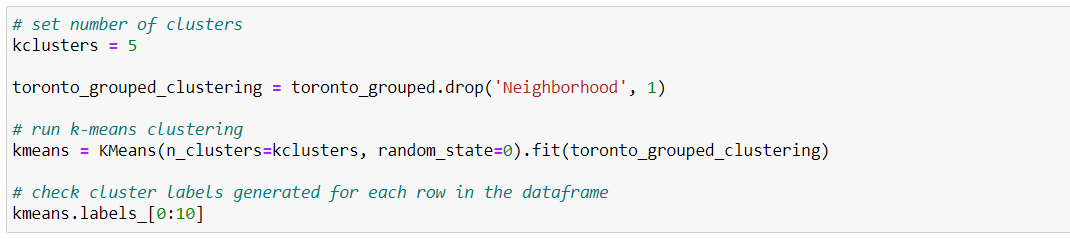


Figure Eighteen (18)

We will create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

Figure Nineteen (19)

# Results

Finally, let's visualize the resulting clusters

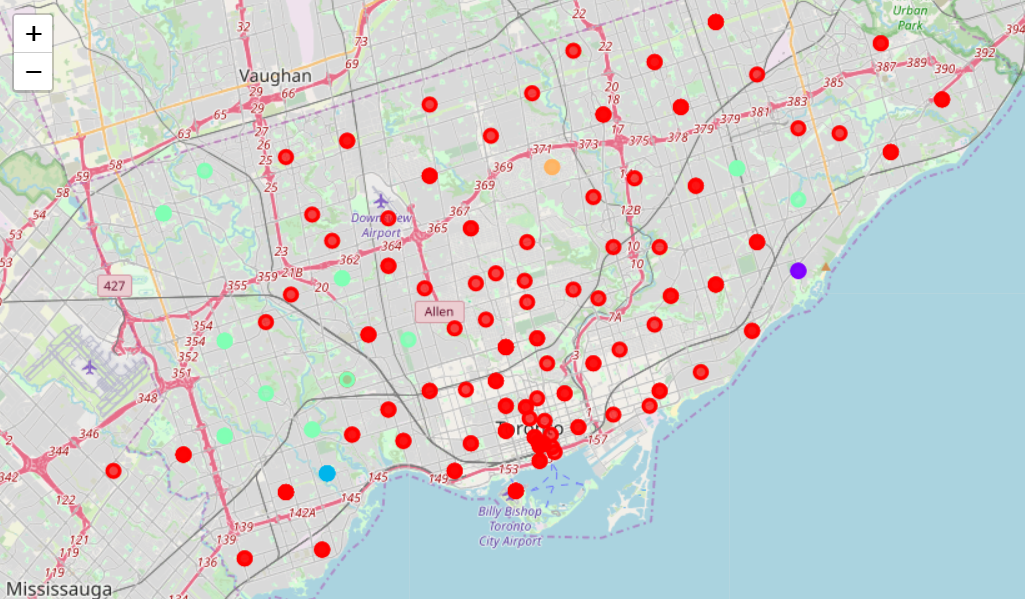


Figure Twenty (20) shows a visualization of the various clusters on the Map of the City of Toronto.

Below are the five clusters that were created



Figure Twenty-One (21) shows result of Cluster 1.

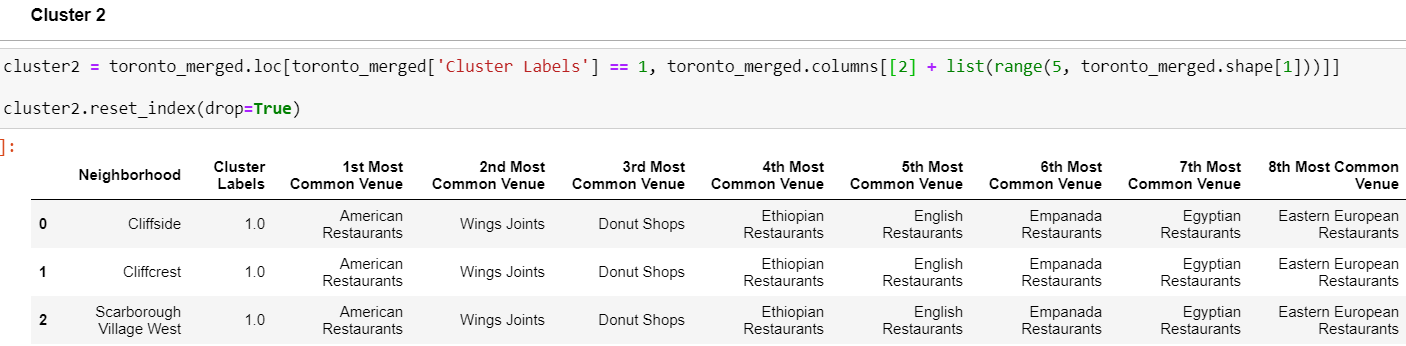


Figure Twenty-Two (22) shows result of Cluster 2.

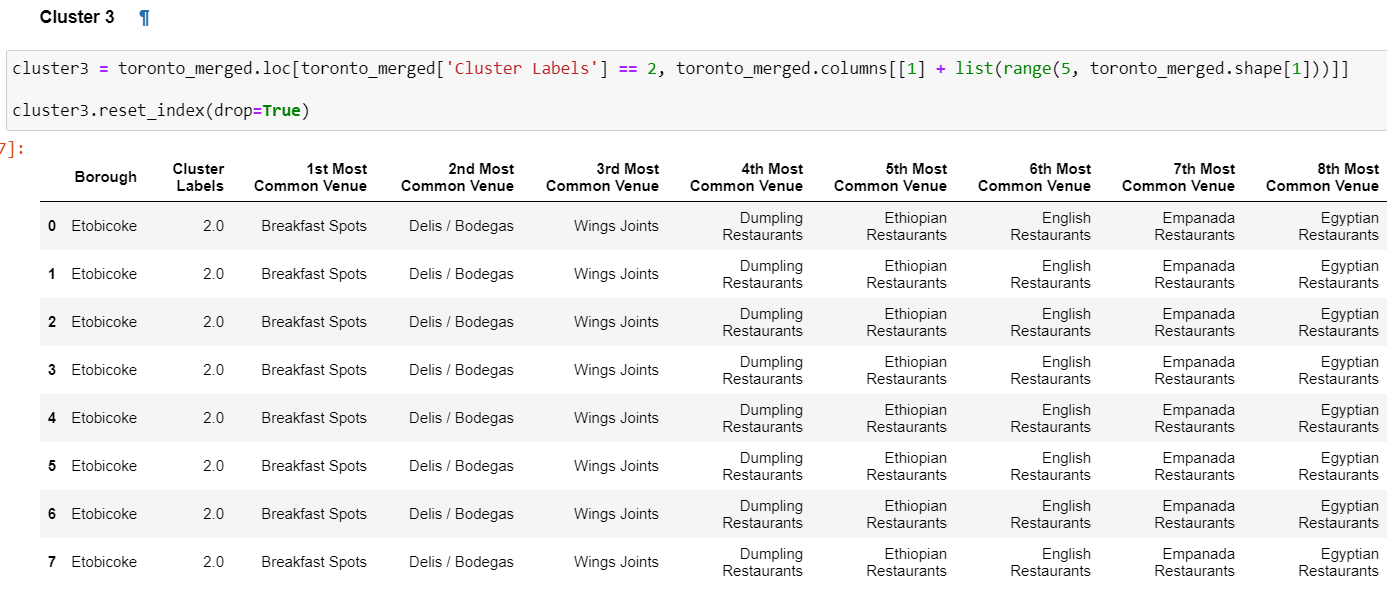


Figure Twenty-Three (23) shows result of Cluster 3.

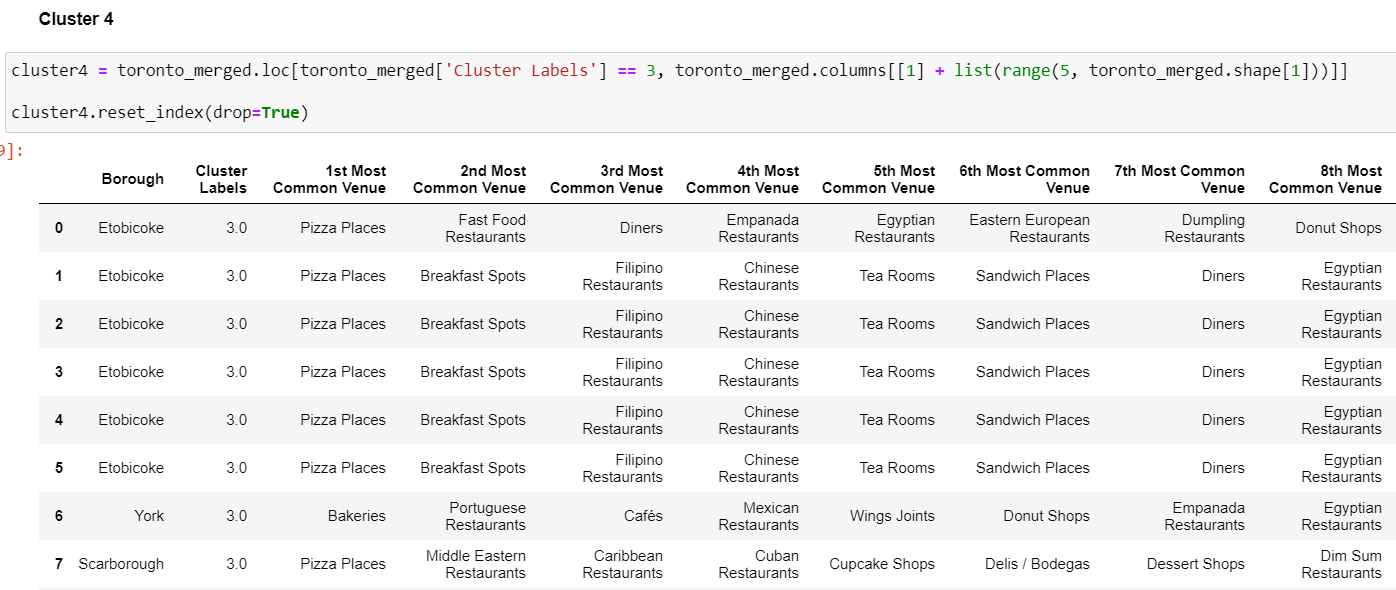


Figure Twenty-Four shows result of Cluster 4.

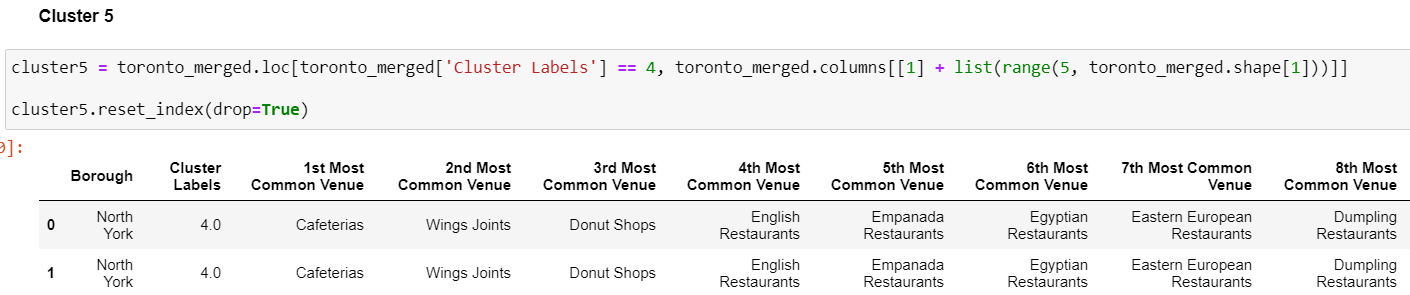


Figure Twenty-Five (25) shows result of Cluster 5.

#### In Cluster 2, there are no African or Caribbean Restaurants, this is not an ideal option.

#### Cluster 3 is same as cluster 2, and therefore is not an ideal option either.

#### Cluster 5 is same as cluster 2 and 3, therefore is not an ideal option either.

#### Cluster 4 does not look like the ideal option. Though this cluster has a high percentage of Caribbean Restaurants, there is no African Restaurant found.

#### Cluster 1 emerged as the most ideal option to site the African Restaurant. As expected, just a few people in cluster 1 are inclined towards African and Carribean Dishes. Now let’s find out the neighborhoods where there are no African Restaurants in this cluster, which would be the list of neighborhoods an entrepreneur can invest and start an African Restaurant.

['Parkwoods', 'Victoria Village', 'Regent Park ', ' Harbourfront', 'Lawrence Manor ', ' Lawrence Heights', "Queen's Park ", ' Ontario Provincial Government', 'Malvern ', ' Rouge', 'Don Mills', 'Parkview Hill ', ' Woodbine Gardens', 'Garden District, Ryerson', 'Glencairn', 'Don Mills', 'Woodbine Heights', 'St. James Town', 'Humewood-Cedarvale', 'Eringate ', ' Bloordale Gardens ', ' Old Burnhamthorpe ', ' Markland Wood', 'Guildwood ', ' Morningside ', ' West Hill', 'The Beaches', 'Berczy Park', 'Woburn', 'Leaside', 'Central Bay Street', 'Christie', 'Cedarbrae', 'Hillcrest Village', 'Bathurst Manor ', ' Wilson Heights ', ' Downsview North', 'Thorncliffe Park', 'Richmond ', ' Adelaide ', ' King', 'Dufferin ', ' Dovercourt Village', 'Fairview ', ' Henry Farm ', ' Oriole', 'Northwood Park ', ' York University', 'East Toronto', 'Harbourfront East ', ' Union Station ', ' Toronto Islands', 'Little Portugal ', ' Trinity', 'Kennedy Park ', ' Ionview ', ' East Birchmount Park', 'Bayview Village', 'Downsview', 'The Danforth West ', ' Riverdale', 'Toronto Dominion Centre ', ' Design Exchange', 'Brockton ', ' Parkdale Village ', ' Exhibition Place', 'Golden Mile ', ' Clairlea ', ' Oakridge', 'Downsview', 'India Bazaar ', ' The Beaches West', 'Commerce Court ', ' Victoria Hotel', 'Downsview', 'Studio District', 'Bedford Park ', ' Lawrence Manor East', 'Del Ray ', ' Mount Dennis ', ' Keelsdale and Silverthorn', 'Birch Cliff ', ' Cliffside West', 'Willowdale', 'Downsview', 'Lawrence Park', 'Roselawn', 'Runnymede ', 'Weston', 'Davisville North', 'Forest Hill North & West', 'High Park ', ' The Junction South', 'Westmount', 'Wexford ', ' Maryvale', 'Willowdale', 'North Toronto West', 'The Annex ', ' North Midtown ', ' Yorkville', 'Parkdale ', ' Roncesvalles', 'Canada Post Gateway Processing Centre', 'Agincourt', 'Davisville', 'University of Toronto ', ' Harbord', 'Runnymede ', ' Swansea', 'Clarks Corners ', " Tam O'Shanter ", ' Sullivan', 'Moore Park ', ' Summerhill East', 'Kensington Market ', ' Chinatown ', ' Grange Park', 'Milliken ', ' Agincourt North ', ' Steeles East ', " L'Amoreaux East", 'Summerhill West ', ' Rathnelly ', ' South Hill ', ' Forest Hill SE ', ' Deer Park', 'CN Tower ', ' King and Spadina ', ' Railway Lands ', ' Harbourfront West ', ' Bathurst Quay ', ' South Niagara ', ' Island airport', 'New Toronto ', ' Mimico South ', ' Humber Bay Shores', 'Steeles West ', " L'Amoreaux West", 'Rosedale', 'Stn A PO Boxes', 'Alderwood ', ' Long Branch', 'St. James Town ', ' Cabbagetown', 'First Canadian Place ', ' Underground city', 'Church and Wellesley', 'Business reply mail Processing CentrE', 'Mimico NW ', ' The Queensway West ', ' South of Bloor ', ' Kingsway Park South West ', ' Royal York South West']

# Discussions

I observed that all the three (3) African Restaurants are scattered in three (3) different Neighbourhoods, but belong to the same Borough. My recommendation will therefore be that, the neighbourhoods from Cluster 1 that belong to the borough Scarborough, but do not have the three African Restaurants should be considered.

# Conclusion

In this project, I analyzed location-based data to ascertain and draw conclusions on what locations will be viable for the running of an African Restaurant within the City of Toronto. The results were satisfactory.