**IBM APPLIED DATA SCIENCE CAPSTONE PROJECT**

**The Battle of Neighborhood -- Week 2**

Finding the Best Neighborhoods to Open a Few African Restaurants in Toronto

Contents

[1 Introduction and Business Problem 2](#_Toc42335860)

[1.1 Introduction 2](#_Toc42335861)

[1.2 Business Problem 4](#_Toc42335862)

[2 Data Sources 4](#_Toc42335863)

[2.1 Toronto Data 4](#_Toc42335864)

[2.2 Geocoder 4](#_Toc42335865)

[2.3 Foursquare API 2.3 4](#_Toc42335866)

[3 Approach 5](#_Toc42335867)

[4 Methodology 5](#_Toc42335868)

[4.1 Toronto Data frame Creation and visualization 6](#_Toc42335869)

[4.2 Explore the Neighborhoods in Toronto Based on Top Venues 7](#_Toc42335870)

[4.3 Cluster the Neighborhoods in Toronto Based on Top Venues in Borough 8](#_Toc42335871)

[4.3.1 Cluster 1 8](#_Toc42335872)

[4.3.2 Cluster 2 8](#_Toc42335873)

[4.3.3 Cluster 3 9](#_Toc42335874)

[4.3.4 Cluster 4 9](#_Toc42335875)

[4.3.5 Cluster 5 10](#_Toc42335876)

[5 Results 10](#_Toc42335877)

[Therefore, the suggested clusters for opening the a few restaurants in Toronto are cluster 1,and 2, and specifica 10](#_Toc42335878)

[6 Discussions 10](#_Toc42335879)

[7 Conclusion 11](#_Toc42335880)

# Introduction and Business Problem

## Introduction

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,7M as of 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5.9 million, making it Canada's most populous CMA and the fourth most populous city in North America. Experts predict that by 2035, the city’s population will have gained another 1 million residents.



Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. Nearly half the population of Toronto is foreign-born, with 50% of the population identifies as white, while the rest identify as East Indian, South Asian, Black, Southeast Asian, Latin American, West Asian, and Arab. While the majority of Torontonians speak English as their primary language, over 160 languages are spoken in the city.

Reflecting this diverse population, Toronto has many ethnic neighborhoods where immigrant populations find comfort and familiarity, this includes neighborhoods like Little India, Greektown, Little Italy, Chinatown, Little Portugal, and Little Jamaica. Toronto is a prominent center for music, theatre, motion picture production, and television production, and is home to the headquarters of Canada's major national broadcast networks and media outlets and varied cultural institutions, which include numerous museums and galleries, festivals and public events, entertainment districts, national historic sites, and sports activities that attract over 43 million tourists each year.

## Business Problem

Being the fastest growing city in North America, Toronto is home to more than 140 languages. According to Statistics Canada, while English is the predominant language in Toronto, other languages such as Cantonese, Mandarin, Tagalog, Italian, Spanish, Farsi, Russian, Korean, Tamil, Urdu, Polish, Somali, Arabic, Panjabi,

Vietnamese and more each have tens of thousands of speakers. With this diverse culture, naturally comes diverse food and tastes. Therefore, a consortium of African restaurants having the goal to build up an African restaurant chain around the world is exploring opportunities to open a few restaurants in Toronto, and need some suggestion

# Data Sources

The following data are used for the project:

## Toronto Data

I have used beautifulSoup to scrape a data frame of Toronto from the csv file on Wikipedia page below. This allowed to obtain the information about the boroughs and neighborhoods along with the postal codes.

Data source: h ttps://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M (https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M)

## Geocoder

I have used to map the geographical coordinates of the neighborhoods through the postal code. Link of the csv file: h ttp://cocl.us/Geospatial\_data (<http://cocl.us/Geospatial_data>)

## Foursquare API 2.3

Foursquare API was used to retrieve data from the Foursquare database, and explore venues and types of restaurant in Toronto Neighborhoods.

Data source: h ttps://developer.foursquare.com/docs/places-api/ (https://developer.foursquare.com/docs/places- a pi/

# Approach

There are many types or international restaurants in Toronto, such as Chinese, Indian, and French. The approach is first to find out whether there are African restaurants in Toronto. If yes, determine the corresponding neighborhoods and the stories behind them, which might be used to open similar restaurant in other neighborhoods. Secondly, determine the top neighborhoods with the most international restaurants or multicultural communities where people may try diverse types of cuisines, which then associated with the top communities having the highest percentage of black population to ascertain the rights clusters for opening the few restaurants in Toronto.

# Methodology

The overall methodology can be summarized by the following steps

1. Scraped data and build a data frame of Toronto that contains its Boroughs and Neighborhoods’ list along with their latitude and longitude.
2. Use the Geocoder to map the geographical coordinates of the neighborhoods through the postal code
3. Use Foursquare API to explore the neighborhoods in Toronto by venues while focusing only on restaurants
4. Check whether there are African restaurant in Toronto. If yes check out the corresponding neighborhood and discard them from the list if not among the top neighborhoods
5. List and plot the 10 top international restaurants by frequency in Toronto
6. List the top international restaurants for the selected each neighborhoods
7. Cluster, visualize, and explore the top neighborhoods based on the frequency of international restaurants by using k-means
8. Discuss the results and select the ideal location to open the few African restaurants
9. Conclusion

## Toronto Data frame Creation and visualization

I have created Toronto data frame by scrapping and wrangling data from Wikipedia using BeautifulSoup. In this process, only the cells that have an assigned borough are processed, and combined the results more than one neighborhood were found for one postal code area, then separate the neighborhoods with a comma.

Then the geocoder is used to get the longitude and latitude of Toronto and merge it with the above scrapped data frame to obtain a final data frame that contains the postal code, borough, neighborhood, neighborhood latitude, and neighborhood longitude and mapped with Folium.

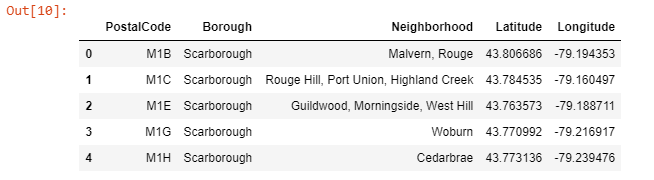


Figure 1: Toronto data frame

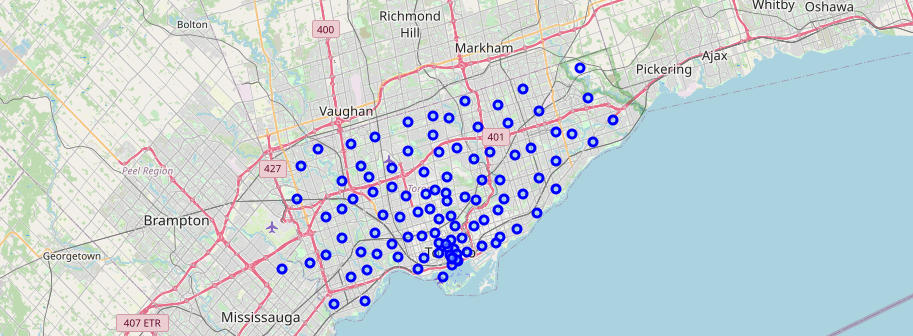


Figure 2: Toronto neighborhood map

## Exploring the Neighborhoods in Toronto Based on Top Venues

In this section, foursquare API is used to explore the neighborhoods in Toronto and create a final data frame with venues categories focusing on international restaurants.

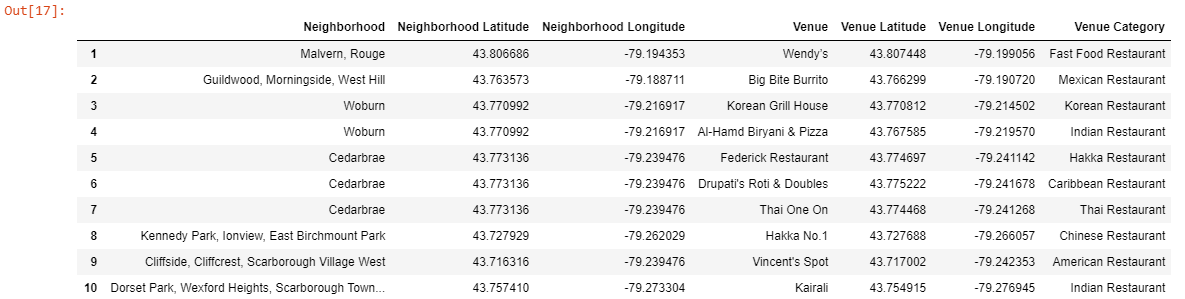


Figure 3: Toronto neighborhood complete data frame with venue categories

The result shows that there are 40 unique types or restaurants in Toronto including one Ethiopian restaurant, called the Ethiopian house at "Church and Wellesley", which is the top 20th neighborhoods with the most number of international restaurants.

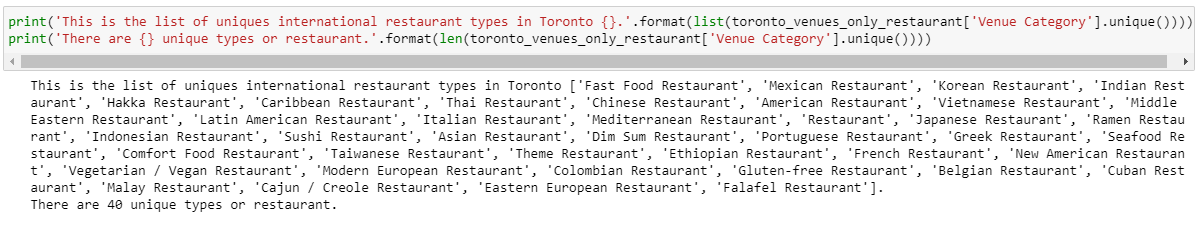


Figure 4: list of unique international restaurants in Toronto

The below figures show the top 10 international restaurants by frequency in Toronto (with the Italian restaurants leading the lists), and the top 10 most common international restaurants in each neighborhood.

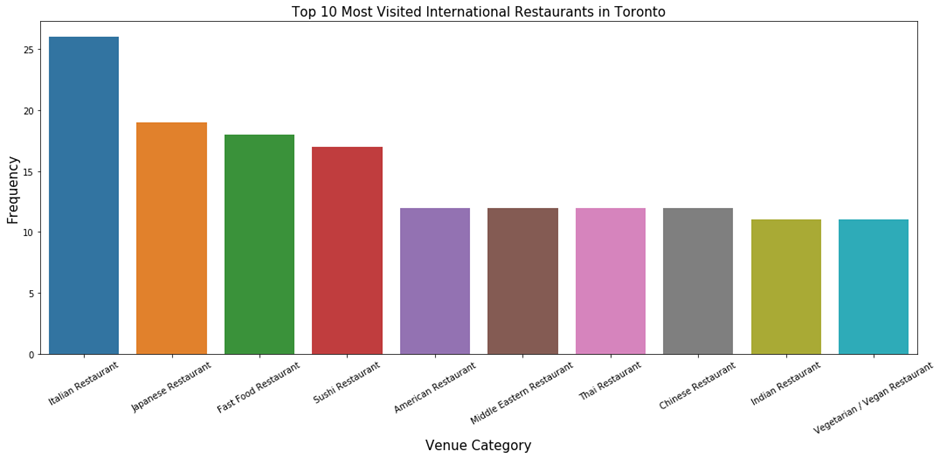


Figure 5: top international restaurants in Toronto



Figure 6: list of the top 10 international restaurants in each neighborhoods

## Cluster the Neighborhoods in Toronto Based on Top Venues in Borough

K-means clustering method is used to cluster the neighborhood into 5 clusters from the new data of the top 10 common international restaurants in each. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms.



Figure 7: neighborhood data frame with the cluster labels

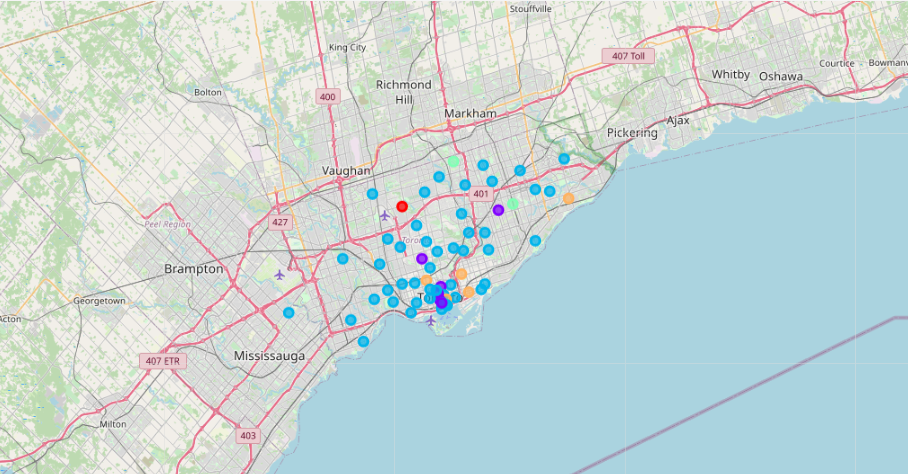


Figure 8: Toronto neighborhoods clusters map

### Cluster 1

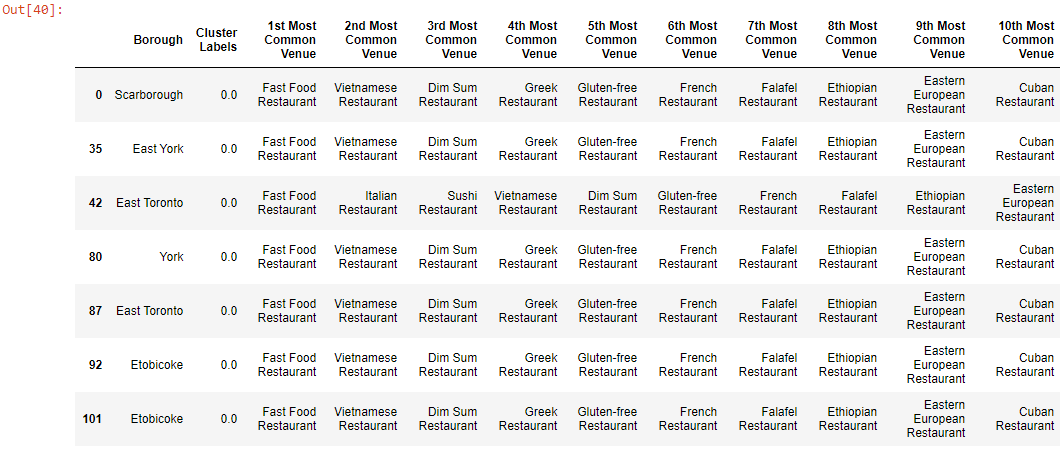


Figure 9: neighborhood cluster 1

### Cluster 2

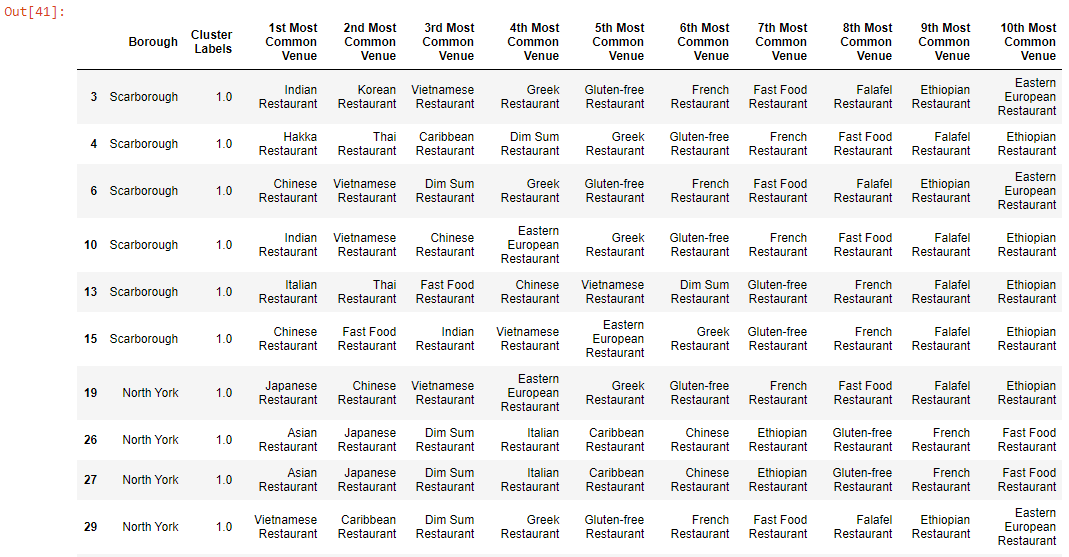


Figure 10: neighborhood cluster 2

### Cluster 3

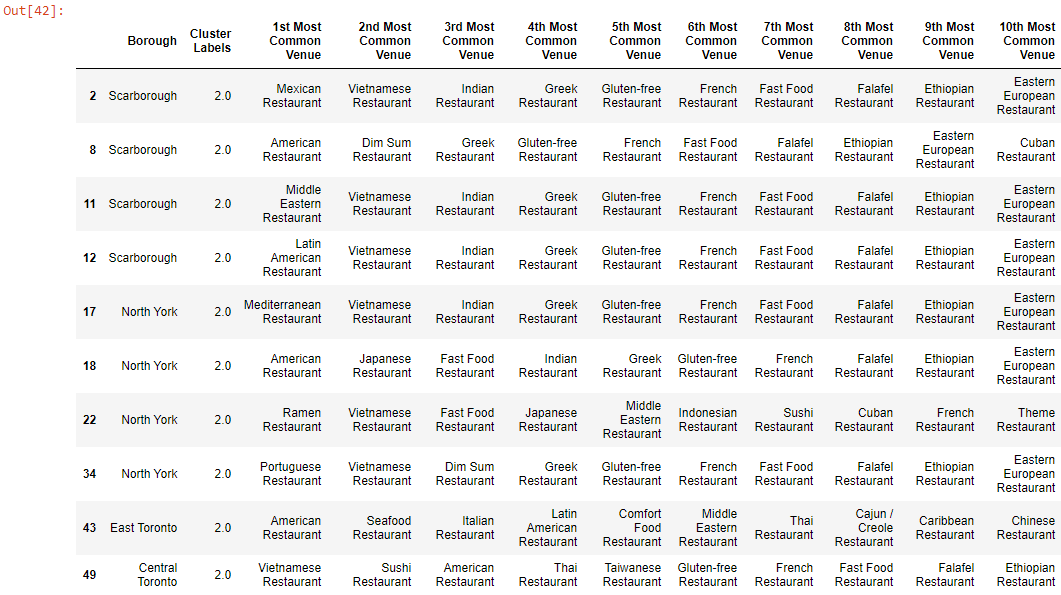


Figure 11: neighborhood cluster 3

### Cluster 4

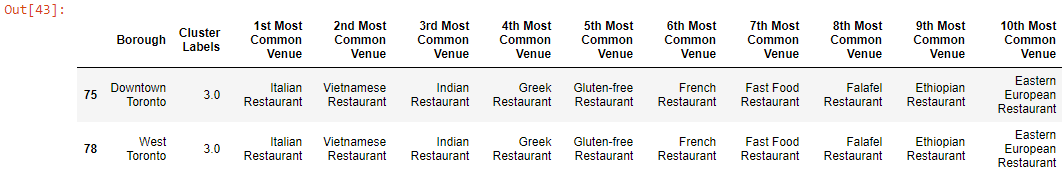


Figure 12: neighborhood cluster 4

### Cluster 5



Figure 13: neighborhood cluster 5

# Results

K-means clustering show the results into 5 clusters based on the most venues in each borough. As per Toronto census 2016, the top visible-minority groups per Community Council are Toronto & East York, North York,Scarborough, and Etobicoke York, with the latter two having the highest percentage of black population 10.8% and 15.7% respectively. The percentage of black population in Toronto and East York is 5.6%, and 5.2% in Etobicoke. Consequently the results of the clustering can be classified as follows:

* + Cluster 1: 4/8 of the boroughs are communities with the highest black percentage
  + Cluster 2: 13/22 of the boroughs are communities with the highest black percentage
  + Cluster 3: 11/29 of the boroughs are communities with the highest black percentage
  + Cluster 4: no neighborhoods with the highest black percentage, but the only Ethiopian Restaurant in Toronto is opened in Toronto Downtown, at Church and Wellesle, which is little weird
  + Cluster 5: 1/4 of boroughs are communities with the highest black percentage

Therefore, the suggested clusters for opening the a few restaurants in Toronto are cluster 1 and 2, and specifications.

# Discussions

The selection of cluster 1 and 2 were based on the number of borough having the highest number of black population. The Cluster 3 could also be considered, since it's the cluster with the highest number of boroughs, but just lag behind in terms of number of communities having the highest percentage of black population. Moreover, 9/29 of the boroughs are Toronto Downtown, where the Ethiopian Restaurants is actually opened. The overall, results could be better if the population and nationalities data per neighborhood were available so that we can spot and target the right neighborhoods with African communities.

# Conclusion

There are many ways to conduct this analysis, but I chose the methods I chose the method I selected as it was a straightforward way to narrow down the options, not complicating what is actually simple in many ways such as focusing of international restaurants only. Form this project, it can be concluded that data science and machine learning are great ways predictions or decisions making on such risky projects. Without leveraging data, the project could end up opening the restaurants in the clusters 4, and 5, which are the boroughs with no or east communities with black population. I also tried to incorporate the data of Toronto's population categorized per immigrants, immigrants’ percentage, and neighborhoods in order to improve the research quality, but could not find the right data. However, this is not an end point of this research, but rather a starting point that will guide the next part of the process to spot the right locations for such a project. The next part will involve domain knowledge of the restaurant industry.