**Location finding for a food related Business in Toronto**

Adebayo Ogunmoriyele August 5, 2019

# Introduction

## Problem Definition

A business location could simply be defined as the place where it is situated. There are several factors that need to be considered in choosing a location for a business. One of the earliest decisions any entrepreneur must make is where to locate his or her business to break even easily. In order to do this, he or she must make a careful assessment of the area. The ideal location would be one where the targets are. The entrepreneur would need to look at the benefits which each area had to offer as well as other factors like security and Labor which might be available. Many business owners have gone bankrupt because of poor location and choice of business. Having a good business placed in a wrong location is as bad as having a bad business put in a good Location.

## Problem

Toronto is a big City. Getting a good location for a good business manually could be cumbersome, tiering, discouraging, time consuming and could yield bad result at the end of the stress and money invested This project is based on finding different areas in the city of Toronto with their existing businesses in other to give my client who wants a food related business a professional advice on what to do and where to place it

# Data acquisition and cleaning

## Data sources

## Data was extracted from [https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M%20) . This link which comprises of a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area. Also, a geospatial\_cordinates which is allocated along each postal code in Toronto to get the coordinates of different areas. forsquare location data was also implored to get different area name and the type of business available . The link to the coordinate is shown bellow <https://github.com/Boyombo1/Cousera_Capstone/blob/master/Geospatial_Coordinates.csv>

## Data cleaning

Data were downloaded and scraped from the two links earlier mentioned and were combined into one After the data scraped from the first hyperlink provided in **2.1** was turned to Dataframe using Pandas, it was noticed that some Postal code have “Not Assigned” value under Borough and Neighborhood colum of the dataframe may be due to insufficient information. These rows with “Not Assigned” information was removed. Thus, only the rows and columns with complete information were utilized. After the Cleaning was completed, I ended up having 103 rows and 3 columns which means 103 areas was reviewed for proper location.

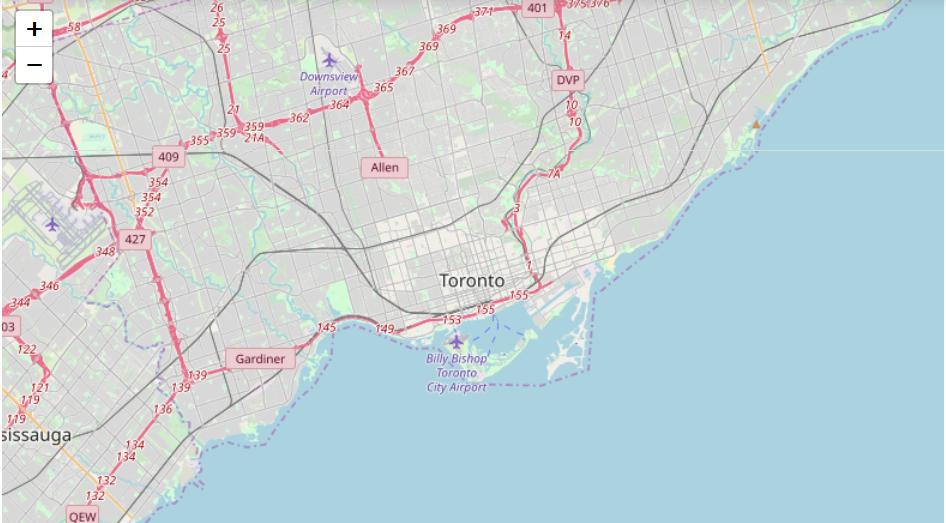
Job not done on data. I went ahead to assign the coordinates extracted from the dataset downloaded from the second link shown in 2.1 by using the join command. After joining was completed, I was able to achieve 103 rows and 5 columns to work with

## Methodology

## 3.1 Clustering and Segmentation

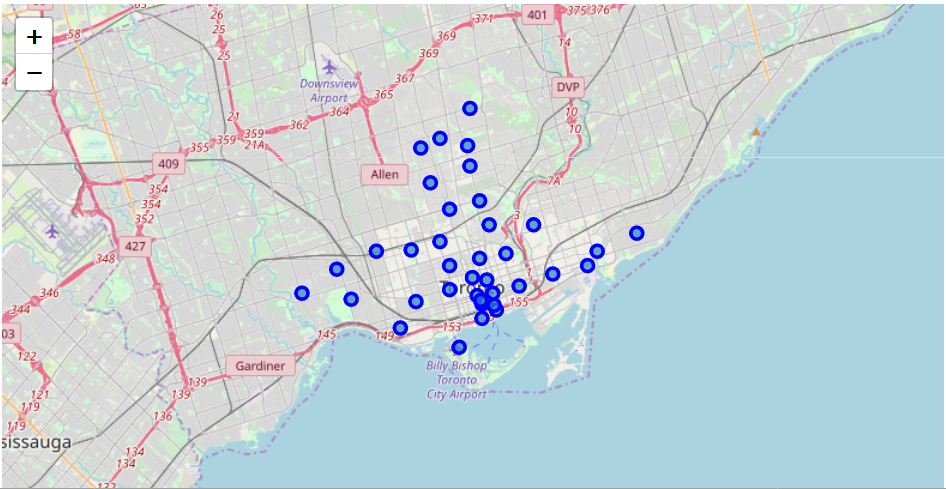
After data Cleaning, there was a need to explore and cluster the Neighborhood in “Toronto, Ontario Canada. Since the address is known, I was able to get the coordinates of the address using geolocator. After the coordinate was obtained, folium was used to plot the area map which is shown bellow

Fig1.1 **Toronto, Ontario Canada Before Clustering**



The Dataframe with different neighborhood and their coordinates was iterated to get the cluster done .

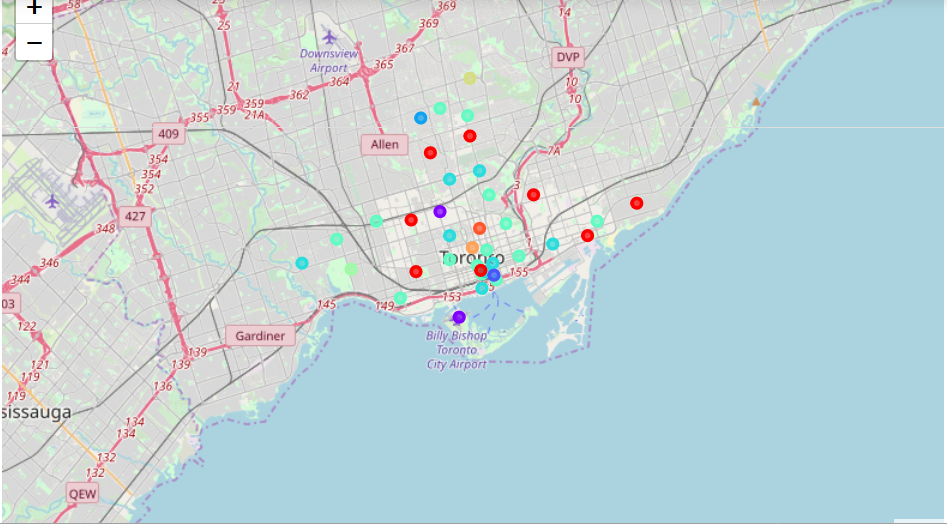
Fig1.2 **Toronto, Ontario Canada After Clustering**



After Clustering, foursquare API was implored for neighborhood segmentation and nearby venues were fetched. The nearby venues were converted to dataframe and were grouped by neighborhood and the clustering was completed by implementing the **Kmeans Machine learning Algorithm**. This machine learning algorithm was adopted because

* It Guarantees convergence
* Can warm-start the positions of centroids.
* With a large number of variables, K-Means may be computationally faster than hierarchical clustering (if K is small).

**Fig 1.3 Clustering and Segmentation Using Kmeans**



1. **Result**

From the clustering and Segmentation, a total, 191 unique categories were achieved. From the data frame, the type of each business available in each area was gotten and recommendation was done

* 1. **Discussion Section**

From the analysis it was observed that different food businesses are existing in many of the neighborhood. If we are to go by factors affecting businesses in general which could be target customers, security, labor and competition. Starting a Bakery in **Harbord,University of Toronto** would be a very good idea because there is no bakery in the whole university, thus no initial competition. Security is guaranteed labor isn’t on the high side and the population is there.

5.0 **Conclusion**

From the analysis and report, we have noticed and discovered that a data driven business in survey and decision making is the best.

Special Thanks to Wikipedia, forsquare data collection and also to cousera