CAPSTONE FINAL PROJECT

IBM DATA SCIENCE

ON

COURSERA

USING LOCATION DATA TO EXPLORE ENTREPRENEURIAL OPPORTUNITIES IN TORONTO

SUBMITTED BY:

ARYAM BAHUKHANDI

INTRODUCTION:

This project is aimed at someone who wants to open a new business in Toronto and wants ideas regarding the same. In this project I have used location data of Toronto and Foursquare API to generate venues in the city. Further, I have used data analysis to figure out which are the most common venues in the city. The data reveals that cafés and coffee shops are quite popular and spread all over Toronto. This suggests high demand for beverages in Toronto. Consequently, I have considered a hypothetical case if someone wants to open a Tea Shop, how he can use the data related to coffee shops to figure out the best location for him to open the tea shop.

BUSINESS PROBLEM:

The aim of the project is to suggest ideas to an Entrepreneur who may be interested in exploring new ventures in Toronto. Alternatively, the aim of the project is to leverage location data to figure out trends in Toronto, which can help an Entrepreneur identify some business opportunities.

TARGET AUDIENCE:

Any Entrepreneur who is seeking some new business ventures in Toronto

DATA:

The data used in the project includes:

1) List of neighborhoods in Toronto (Wikipedia)
2) Latitudes and Longitudes corresponding to the neighborhoods
3) Venues related to the neighborhoods

DATA EXTRACTION:

1) Web scrapping for extracting data from Wikipedia
2) Use of Geocoder package to extract the latitudinal and longitudinal data
3) Use of Foursquare API to get venues related to the neighborhoods

METHODOLOGY:

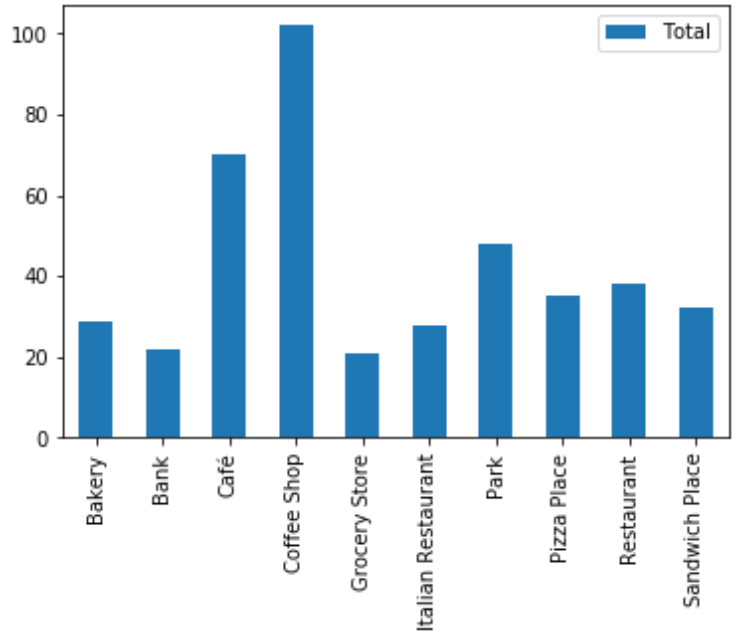
The first step in the process was to scrap data related to the neighborhoods in Toronto from Wikipedia in the form of a table. For this, I used BeautifulSoup. After scrapping the data, the second step involved adding latitudes and longitudes of the neighborhoods to the table. Once the neighborhoods data was updated with the geographical coordinates, I used foursquare API to get a list of venues for the neighborhoods.

Further, I sorted the data to find out unique venue categories and venue categories which were very frequent. The data revealed that cafés and coffee shops were very popular and spread across Toronto. This suggested to me that beverages were in high demand in Toronto. Consequently, I considered a hypothetical case where if someone would want to open a Tea Shop, how can they use the location data to figure out the best location for them to open their tea shop. Subsequently, I visualized Tea shops and Coffee Shops in Toronto on maps by using the folium library.

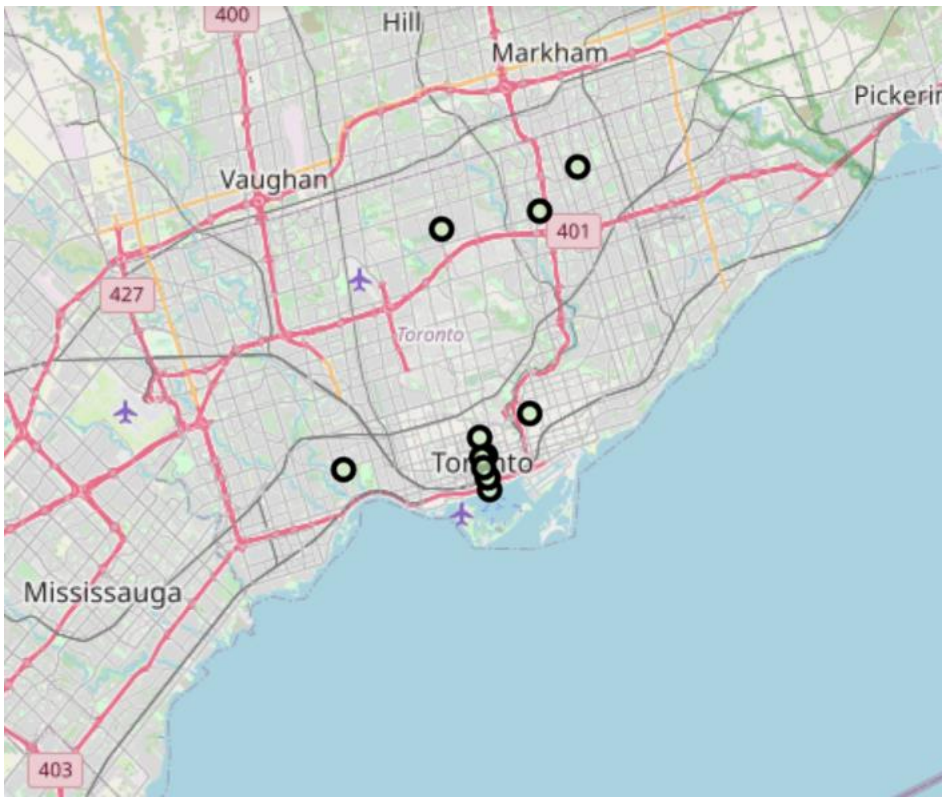
Later, I used k-Means clustering to form clusters for coffee shops. I consider coffee shops as indirect competition to tea shops, as a person who wants to drink a beverage can substitute tea for coffee. The clusters resulting from k-Means gave an insight into the various groupings of coffee shops, which can be transposed to tea shops to figure out a suitable location to open a new one.

RESULTS:

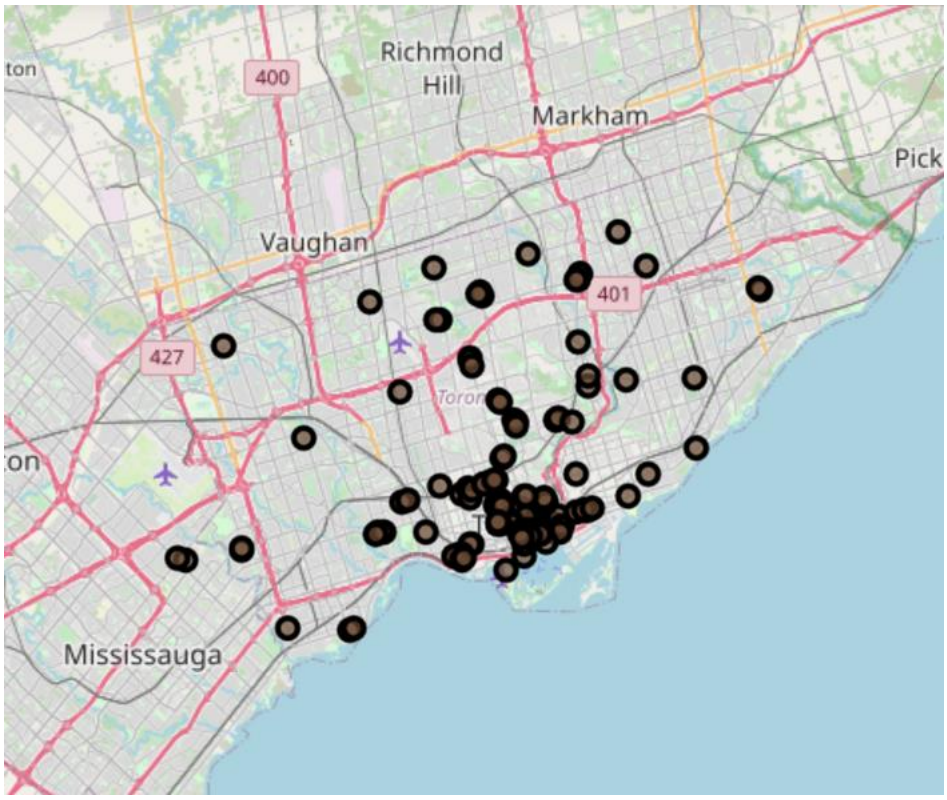
High frequency occurrences of Venue Categories:



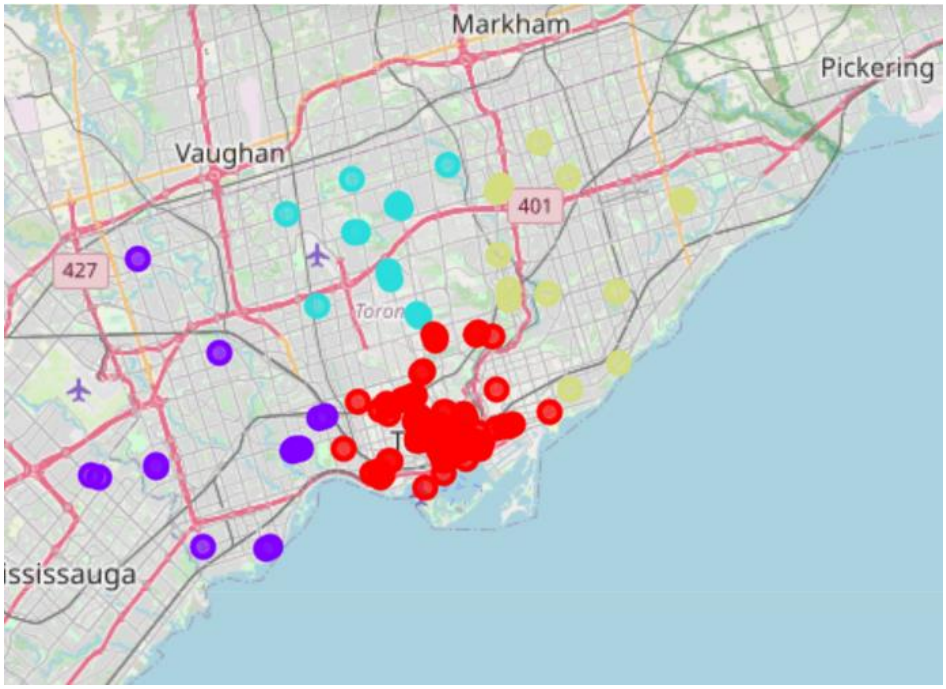
Visualizing Tea Shops in Toronto:



Visualizing Coffee Shops in Toronto:



Visualizing Clusters:



	Venue	Observations
Cluster Labels		
0 (RED)	124	Densely Packed - may be chosen by someone with a high risk appetite.
1 (PURPLE)	16	Scattered – well suited to set up a new Tea Shop (Personal commendation)
2 (BLUE)	16	Evenly spaced – difficult to find an empty space
3 (YELLOW)	16	Less Scattered – this area can be a second choice

In my opinion, Cluster 1 should be chosen to set up a new Tea Shop, because it has very less coffee shops, and a single tea shop (from map_tea). Moreover, the coffee shops are scattered and there is enough area with no coffee or tea shop in the vicinity. Therefore, this area has a benefit of low competition compared to other areas in the map.