

A decorative graphic on the left side of the slide, consisting of a network of light blue lines and small circles, resembling a circuit board or a data network, extending from the top to the bottom.

# CLUSTERING AND COMPARING VENUES IN THE NEIGHBORHOODS OF NEW YORK CITY AND TORONTO

BY AHMED GAMAL

# CLUSTERING AND COMPARING VENUES IN THE NEIGHBORHOODS OF NEW YORK CITY AND TORONTO





# INTRODUCTION

- ✓ I will study, analyze, cluster, and compare the neighborhoods of two important cities in the world: New York City which is located in United States of America and Toronto which is located in Canada.
- ✓ I will investigate on what kinds of businesses are common in both cities, what kinds of businesses are more common in one of the two cities than the other city, and what kinds of businesses are not common in both cities.

# DATA ACQUISITION AND PREPARATION

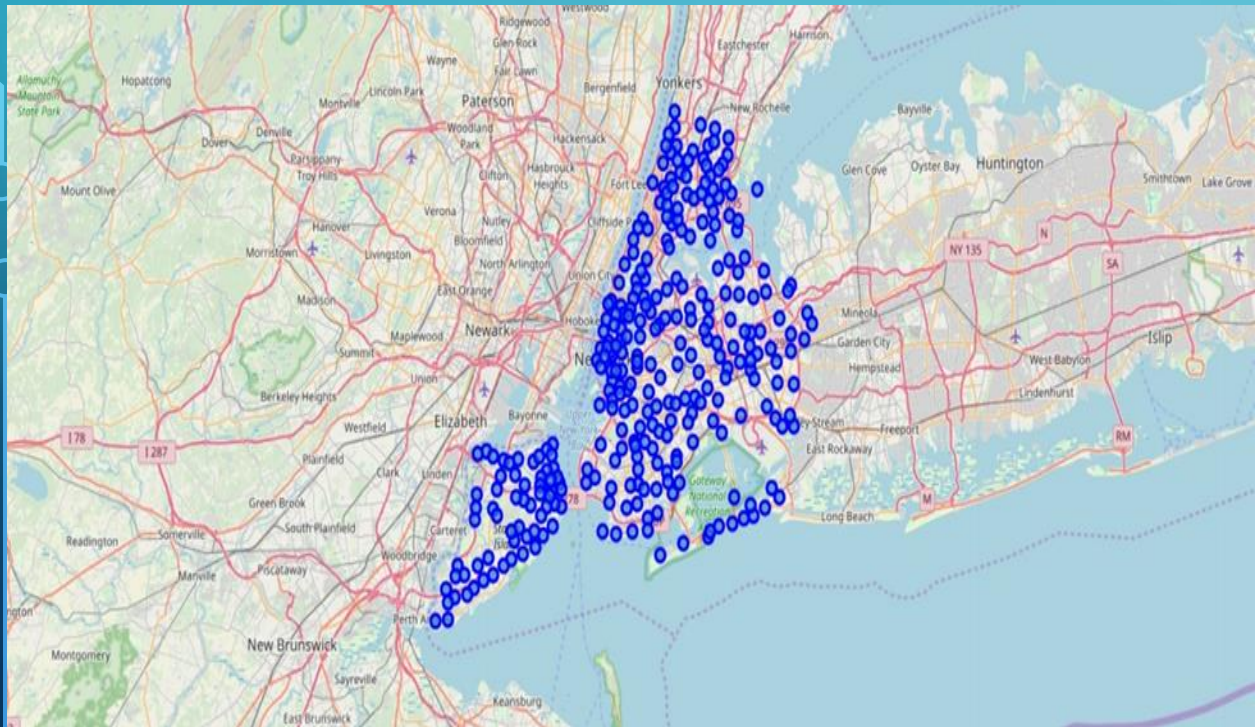
- ✓ **A-Neighborhood Data:** Datasets that lists the names of the neighborhoods of NYC and Toronto and their latitude and longitude coordinates. We have some of this data provided by the coordinators of "IBM Data Science Professional Certificate".
- ✓ **B- Venues data:** Data that describes the top 100 venues (restaurants, cafes, parks, museums, etc.) in each neighborhood of the two cities. The data should list the venues of each neighborhood with their categories.

This data will be retrieved from Foursquare which is one of the world largest sources of location and venue data.

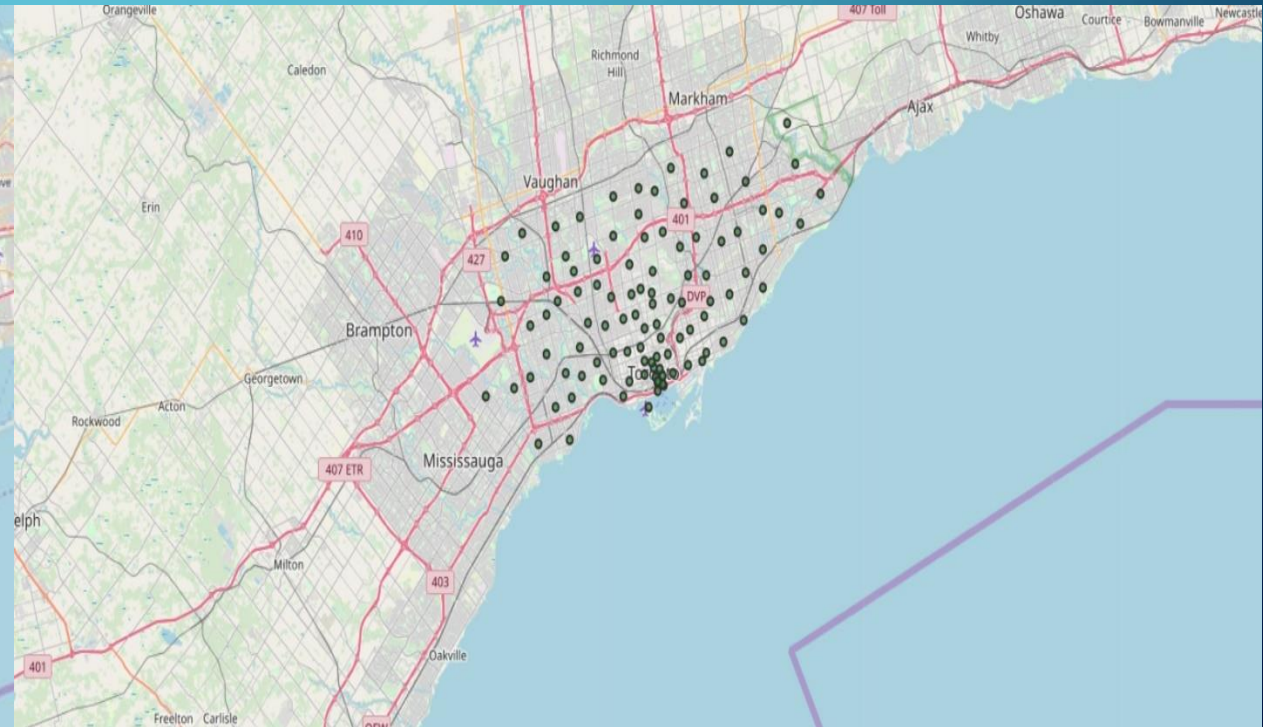
Foursquare API will be utilized to get and download the data.

# USING VENUES DISTRIBUTION AS THE MAIN MEASURE

New York City

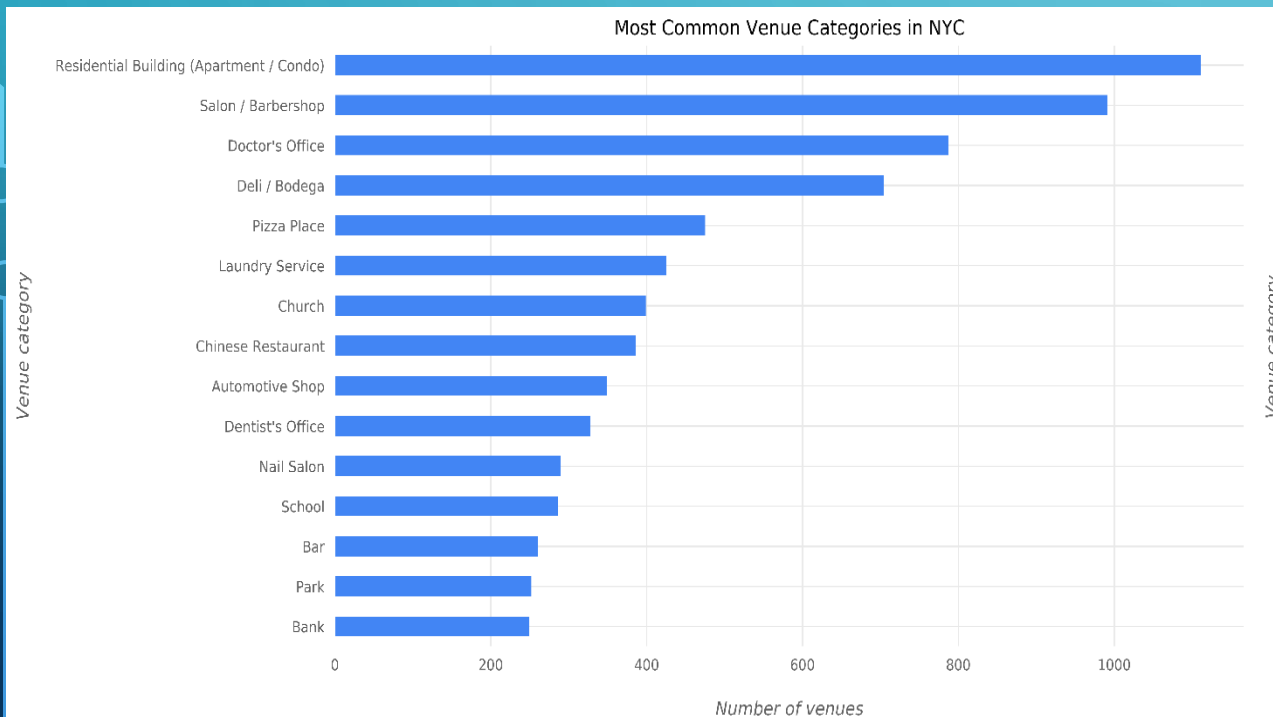


Toronto City

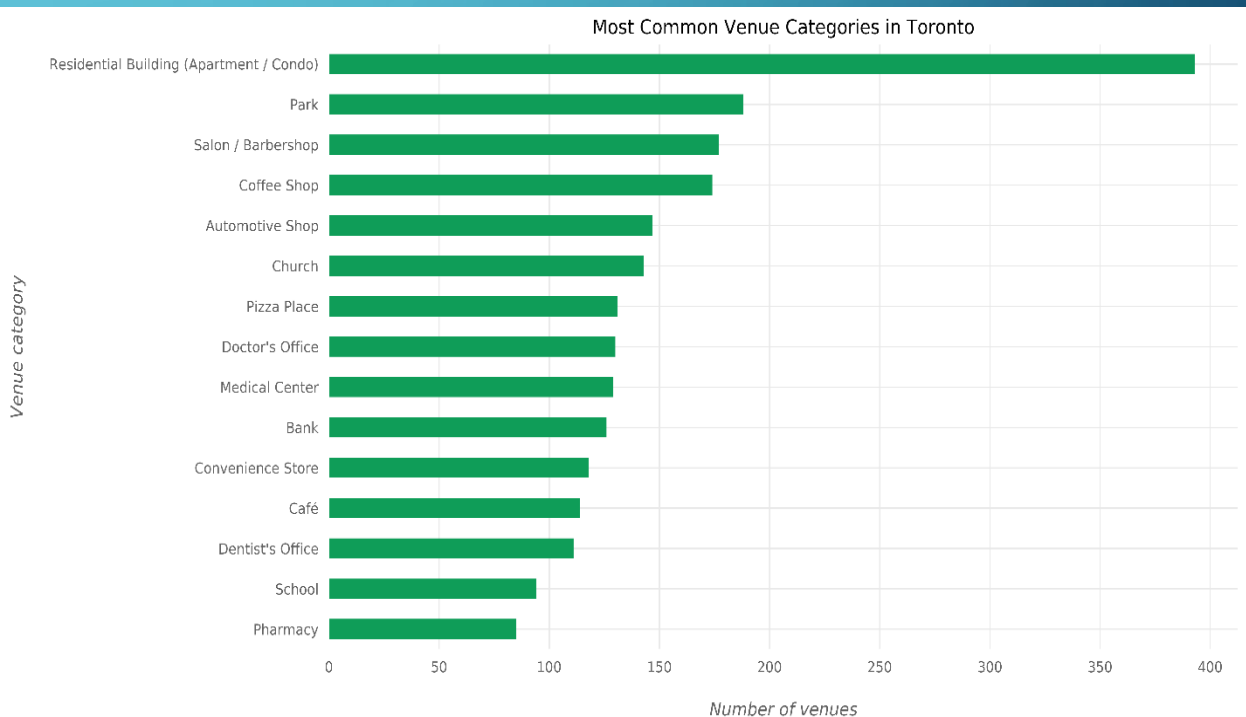


# MOST COMMON VENUES IN NYC AND TORONTO CITIES:

## New York City

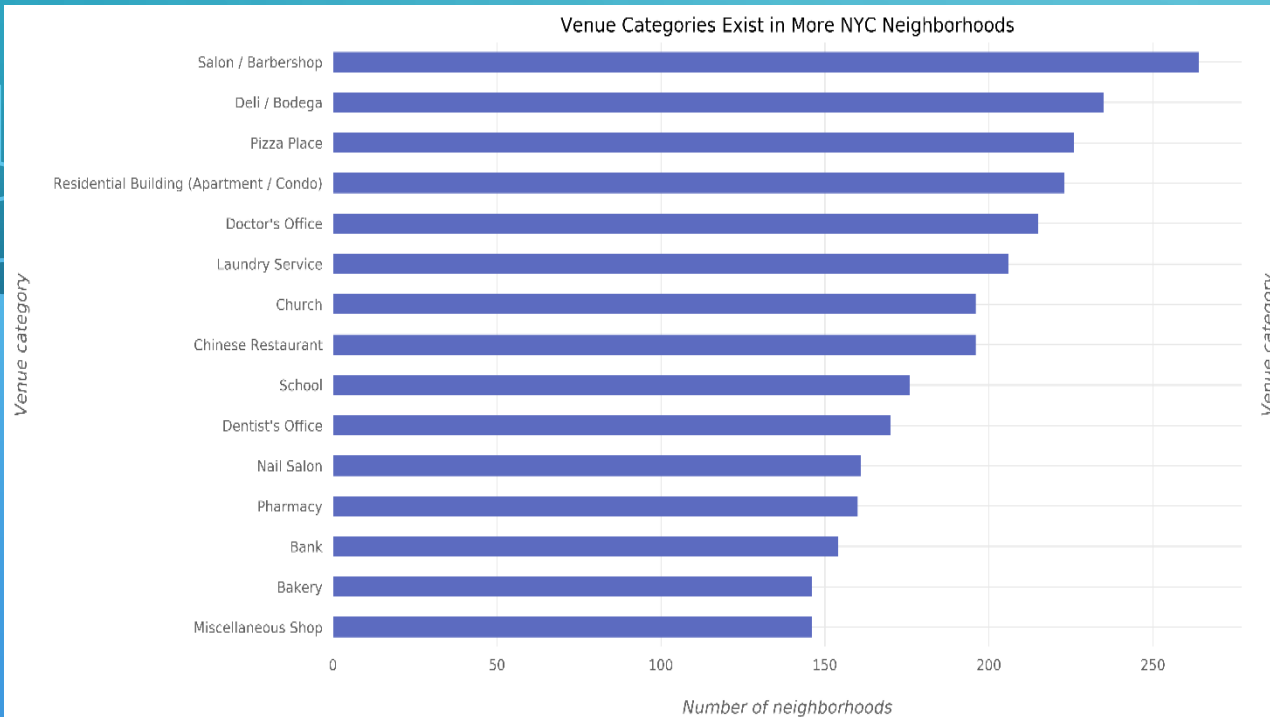


## Toronto City

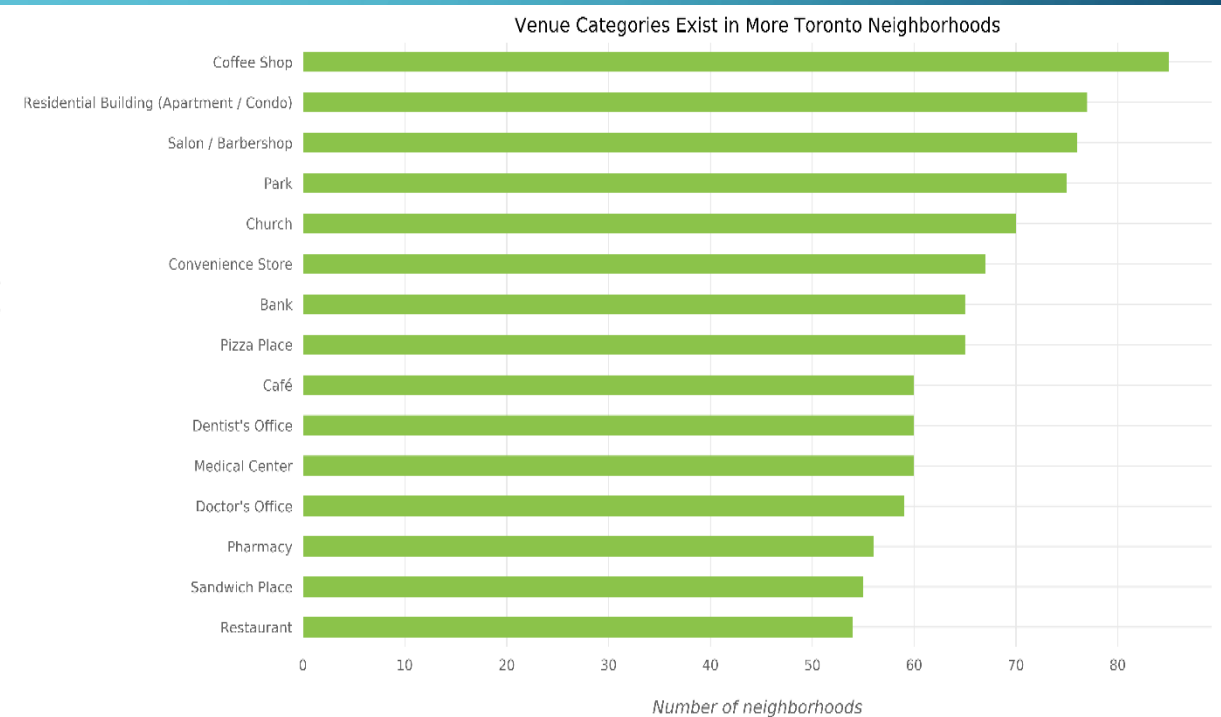


# MOST WIDESPREAD VENUES IN NYC AND TORONTO CITIES:

## New York City



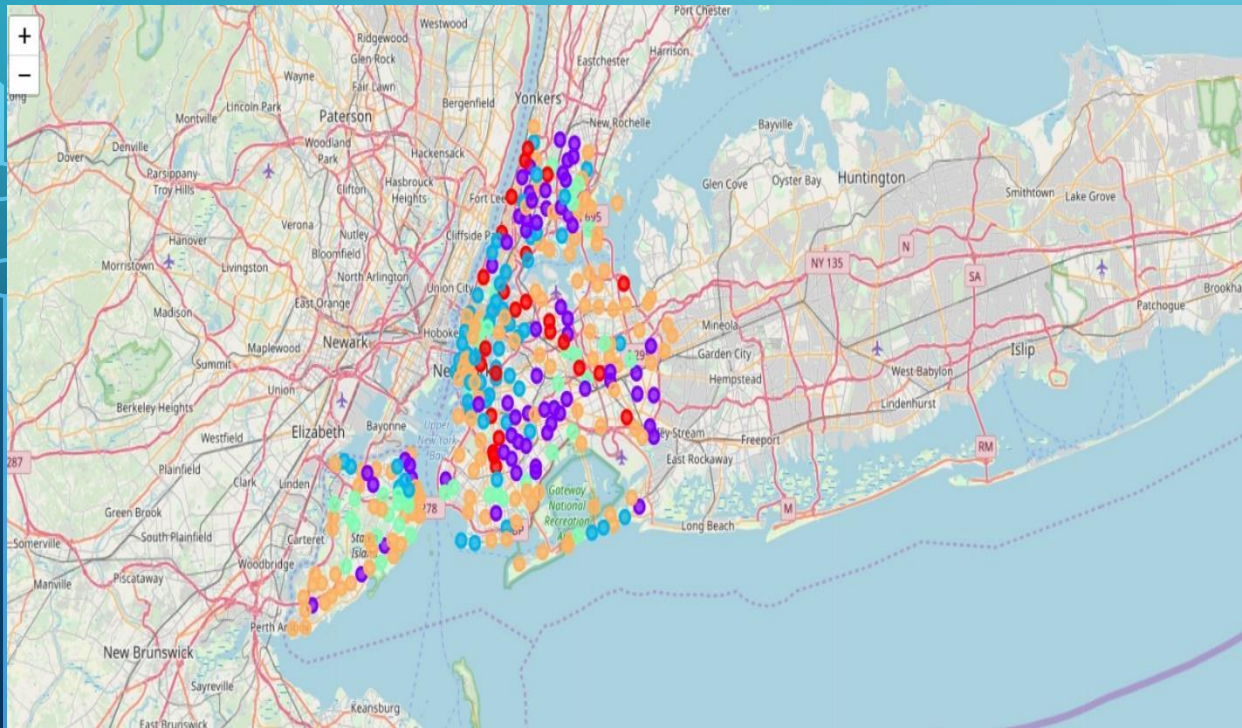
## Toronto City



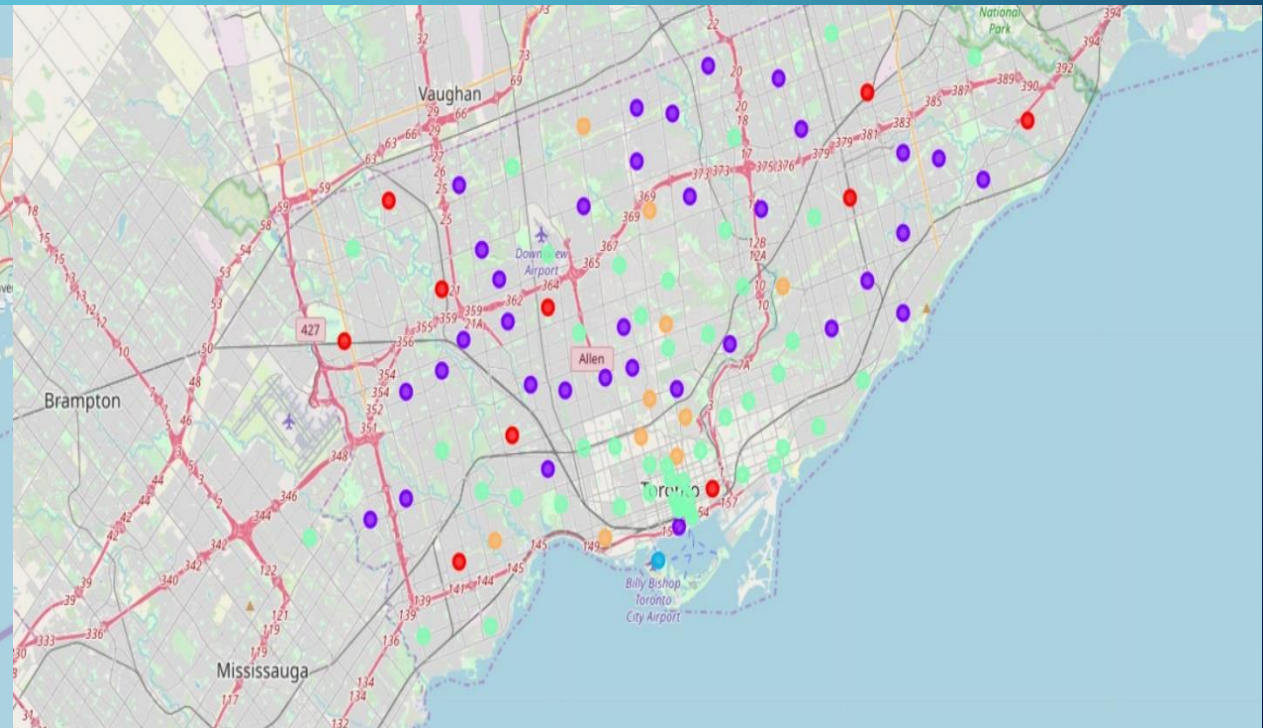


# CLUSTER VENUES DISTRIBUTION IN NYC AND TORONTO CITIES:

New York City



Toronto City





# CONCLUSION AND FUTURE DIRECTIONS

- ✓ In this project, the neighborhoods of New York City and Toronto were clustered into multiple groups based on the categories (types) of the venues in these neighborhoods.
- ✓ The results showed that there are venue categories that are more common in some cluster than the others; the most common venue categories differ from one cluster to the other.
- ✓ The results also showed the most common venues categories in each city along with the widespread of each category so by using the clustering algorithm the project can detect the distribution of similar venues or business category on map.
- ✓ By these results you have the best and widespread business category and also their location in each city and this can be applied on the major cities around the world.
- ✓ I think a deeper analysis can be done according to our client's business category and also using more data sources can give the project more intuition and prediction of each business category future.