Toronto Franchise Expansion

Capstone Project

The Battle of Neighborhoods

Introduction

- Toronto is the financial capital and of the most famous cities in Canada.
- In the context of its dynamic business environment, many businesses currently located in Toronto have plans of expanding their geographical presence in New York city.
- This creates the need of creating a tool that can generate data-driven recommendations of which areas in New York are more suitable for each business type and hence, minimize the risk of the investment and increase the potential returns.
- The currently presented tool analyzes the similarities and differences between Toronto and New York in terms of their demographic composition and business environment of their neighborhoods, and it generates recommendations of a successful expansion strategy.
- It aims to serve chain store and franchise owners in Toronto, who are willing to expand their businesses in New York.

Data Acquisition

In this analysis we use the following data types:

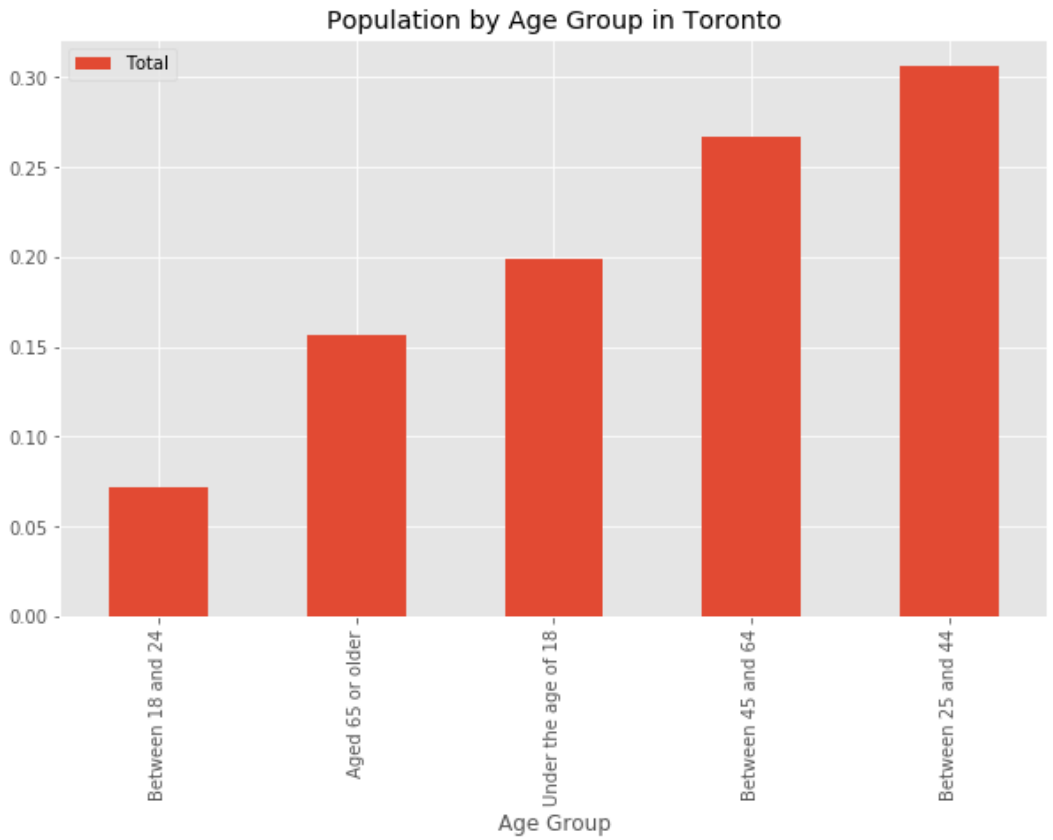
- Population and Age Distribution of the two cities
    - This data will enable us to identify how similar the cities are in terms of their demographic composition, in terms of age distribution of their populations.
- The geographical coordinates of New York and Toronto neighborhoods.
    - This information will enable us to separate the neighborhoods for the analysis but also to eventually visualize the results.
- Foursquare API to retrieve the available venues for each neighborhood.
    - This data is the core of the analysis, as we will retrieve the venues and their types for each neighborhood of the two cities in question.

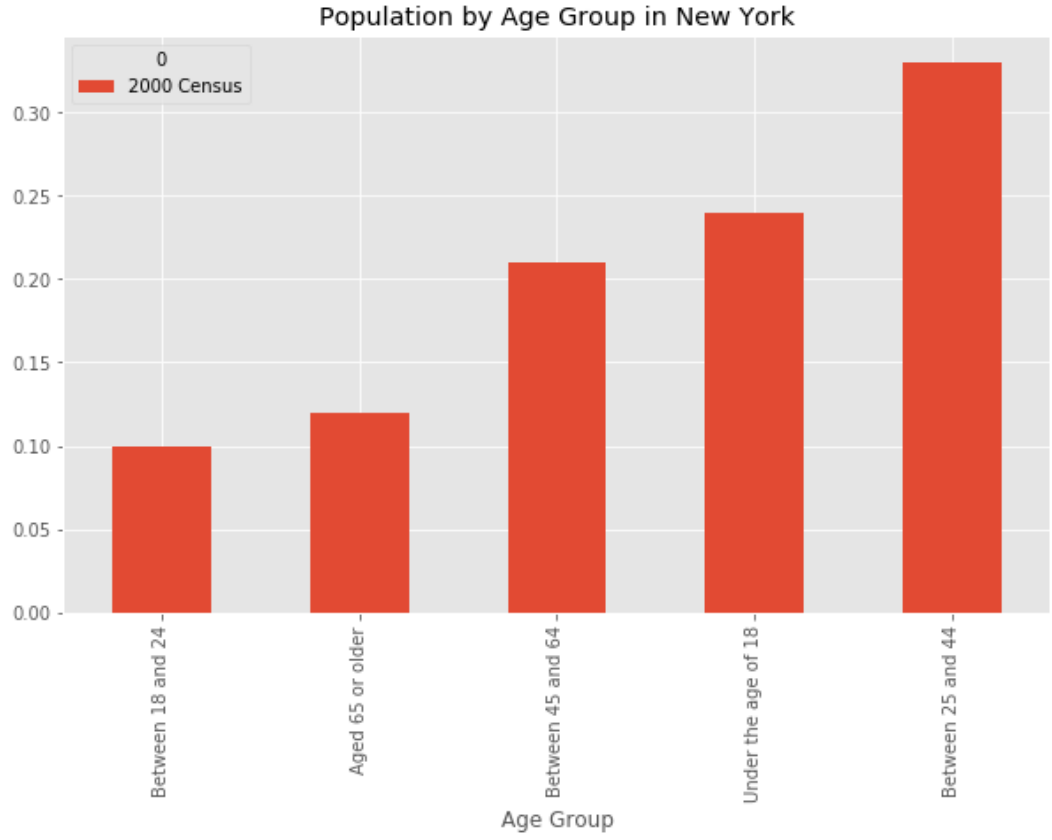
Demographic composition of Toronto and New York

Toronto Age Distribution	
Age Group	**Split**
Under the age of 18	20%
Between 18 and 24	7%
Between 25 and 44	31%
Between 45 and 64	27%
Aged 65 or older	16%

New York Age Distribution	
Age Group	**Split**
Under the age of 18	24%
Between 18 and 24	10%
Between 25 and 44	33%
Between 45 and 64	21%
Aged 65 or older	12%

Demographic composition of Toronto and New York

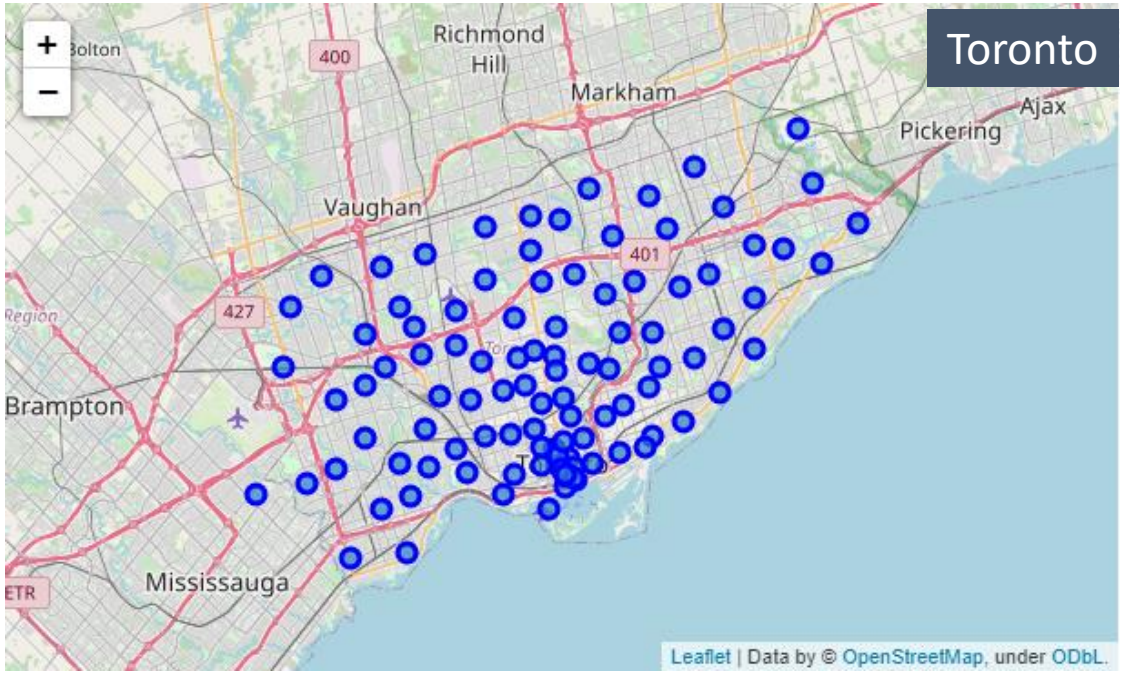


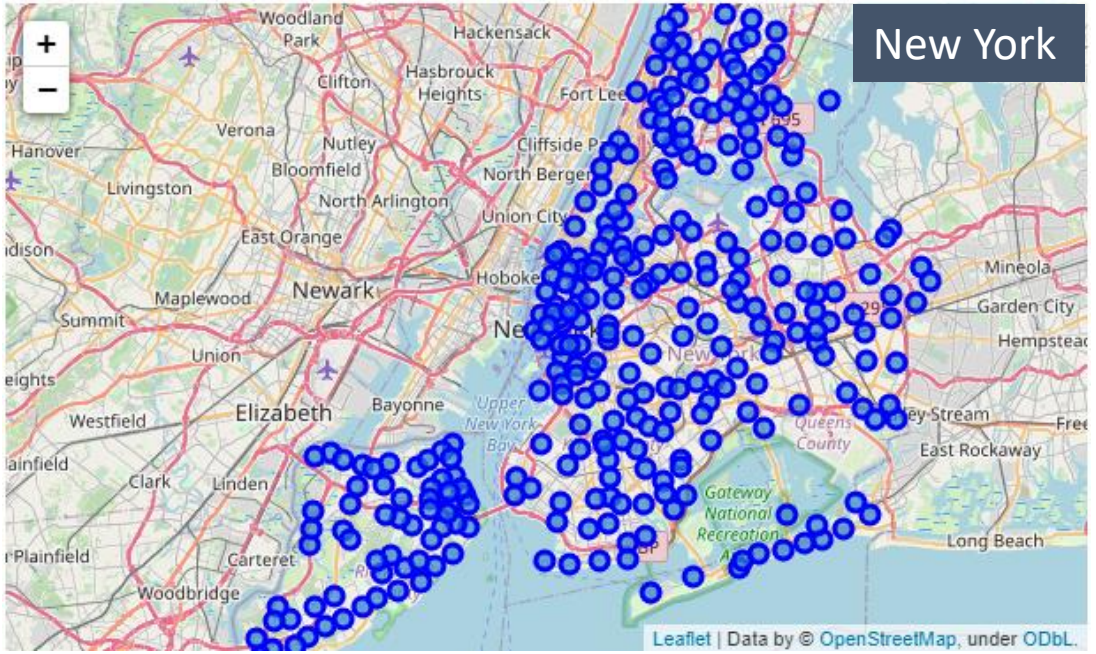


Toronto has more people with an age of 45 years old and older than New York. On the contrary, New York's population appears to be younger, especially for the people less than 24 years old.

Neighborhoods' Insights

Having concluded the first insights of our analysis, we proceed by having a look at how the neighborhoods of the cities look like.





Toronto has clearly less neighborhoods that New York. Moreover, the neighborhoods of Toronto appear to be more uniformly distributed across the city, while in New York the pattern does not follow any visible rule. We could infer that New York might have neighborhoods of different sizes and types, namely a bigger variety and culture in the city. Toronto, on the other side, is expected to be more consistent in these terms.

Neighborhoods' Insights

To verify these assumptions, we can retrieve the venue types for each city by using Foursquare's API.

We can the calculate how many different venue types we have in each city and validate our previous assumption.

- Toronto has 265 unique venue categories
- New York has 440 unique venue categories

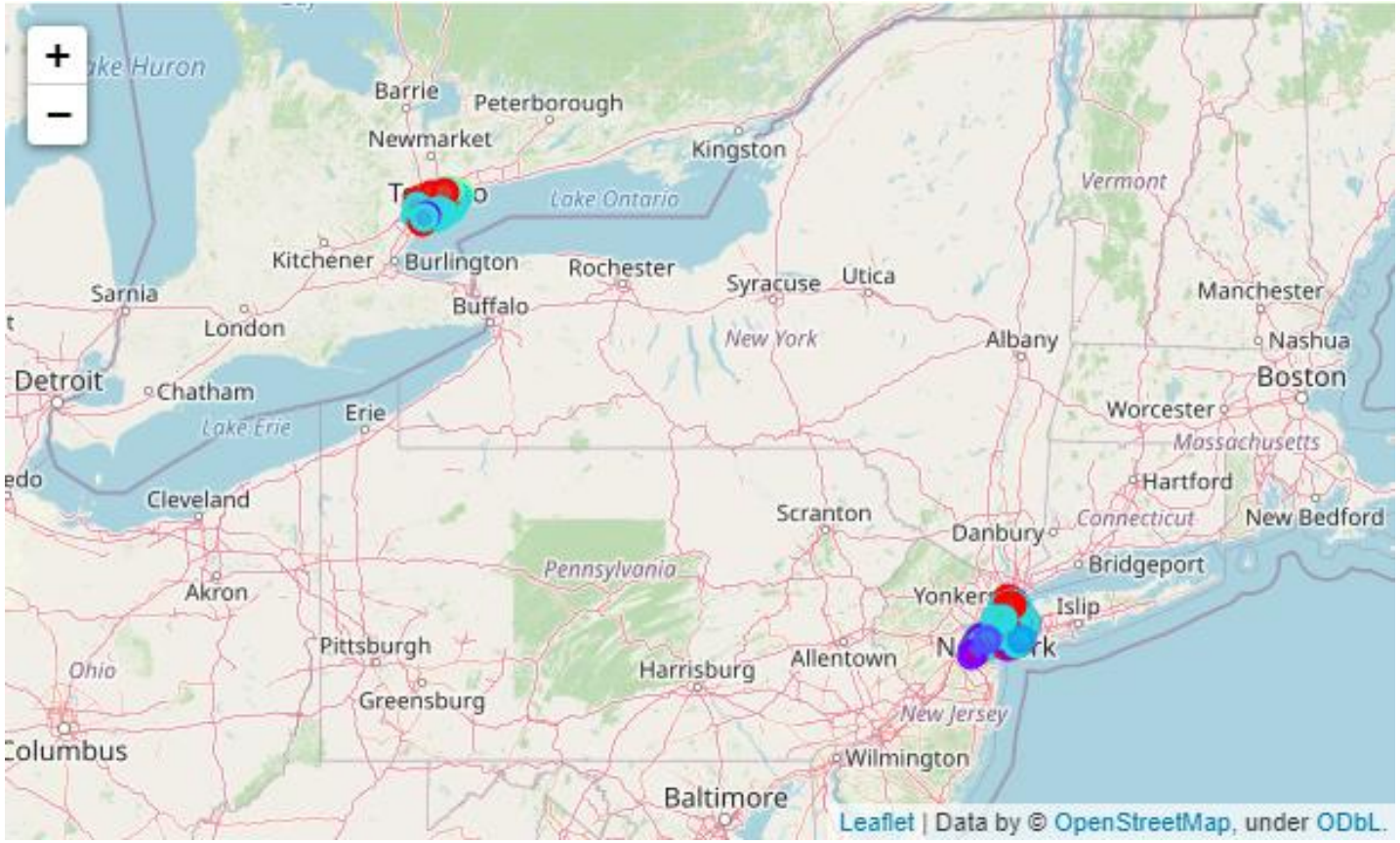
✓ We can then validate our hypothesis that New York offers a more versatile business environment than Toronto.

Clustering

- Our target is to find similar neighborhoods across the two cities in terms of the currently available venue types and be able to recommend which neighborhoods in New York will be more suitable for an existing business in Toronto willing to expand their presence in the former city.
- To achieve this result, we first merge the neighborhoods of the two cities in a common dataframe, along with their geographical coordinates and venues.
- This dataset will be used for unsupervised clustering using the k-Means algorithm. We will use 10 clusters to get an adequate granularity on the different neighborhood types.

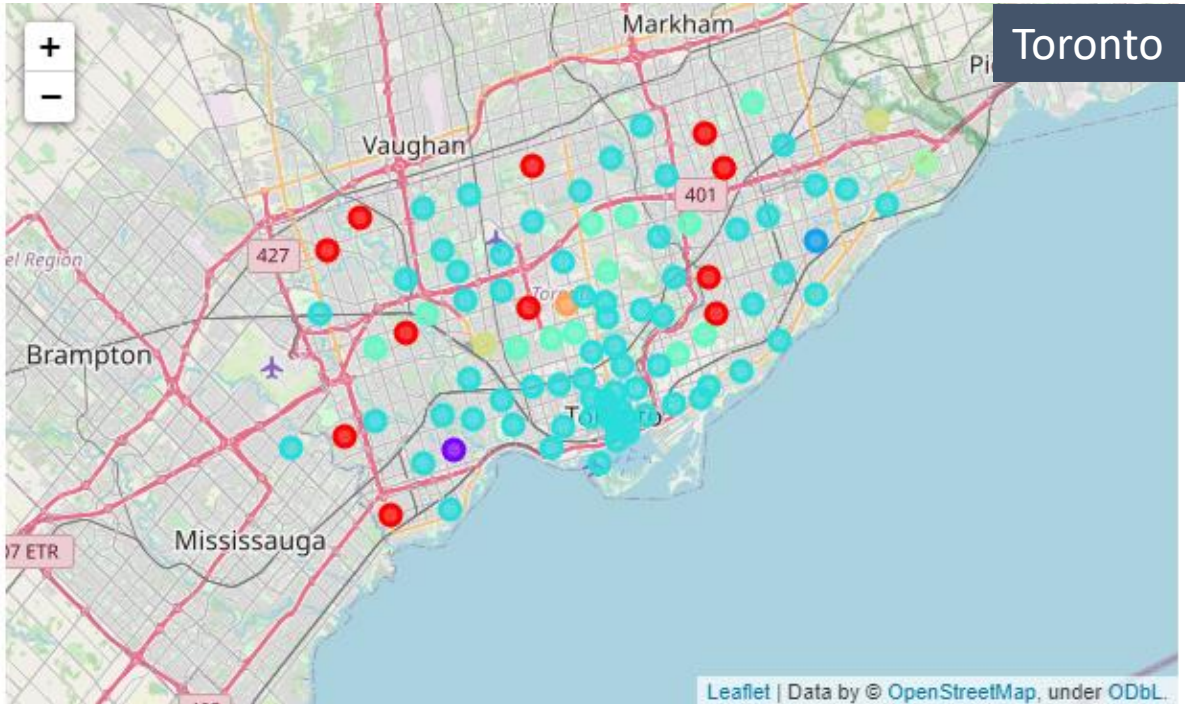
Clustering

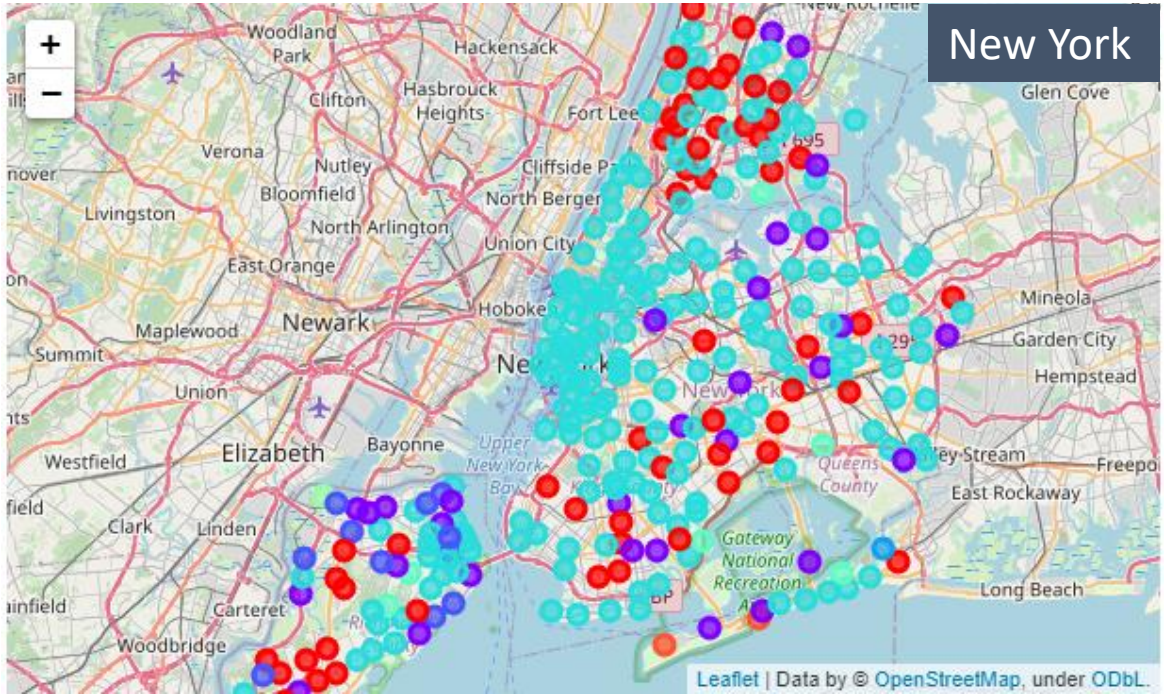
- As you may see in the below figure, the neighborhoods of the two cities have been clustered based on the same rules.



Clustering

- As you may see in the below figure, the neighborhoods of the two cities have been clustered based on the same rules.





Results

- We can see that in both cities, we have three main clusters that describe most of the neighborhoods, based on the currently available venue types.
- We also observe that most of the neighborhoods in Toronto's city center belong to Cluster 4, highlighted with a light blue-green color. The same cluster is also present in New York, mainly in Mid/Lower Manhattan.
- Exploring in depth this cluster, we can identify that the most frequent venue types in these neighborhoods are Coffee Shops, Clothing Stores, Restaurants and Gyms. In addition to that, we can also observe that this type of neighborhoods can be found in both cities in more de-centralized locations but less frequently.

✓ **If we assume that our client has a Restaurant business in Toronto's city center, we are now able to recommend which locations in New York are more suitable for expanding their business in.**

Results

- Additionally, we can enhance our recommendation by exploring how the competition looks like in the targeted neighborhoods.
  - Namely, how many restaurants exist, what is ratio of the restaurants per capita for each location, and how many restaurants of a particular type exist in each neighborhood.
  - For example, if our client has a Mexican restaurant, we can identify the suitable areas and then sort them based on how many Mexican restaurants they have.
  - This can provide us with adequate information of choosing a promising area with the lowest possible level of competition.