The Battle of Neighborhoods

Sebastián Toledo

Manhattan Vs. New York

▶ Through Data Analysis and some Machine Learning algorithms we will look for patterns that help us understand the similarities and differences that the neighborhoods of both cities have. Information that is very valuable for tourists, business people or interested in living there.

Data acquisition and cleaning

- For this project we will use information provided by two sources that I will cite below.

- Dataset containing relevant information from the city of New York: https://geo.nyu.edu/catalog/nyu_2451_34572

- Website containing relevant information about the city of Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

- This information is essential to respond to our analysis as it contains longitude, latitude and corresponding neighborhood of each city, we also have the Foursquare API that provides us with additional information as more relevant places that make up each neighborhood and their respective information provided by visitors.

Manhattan Vs. New York
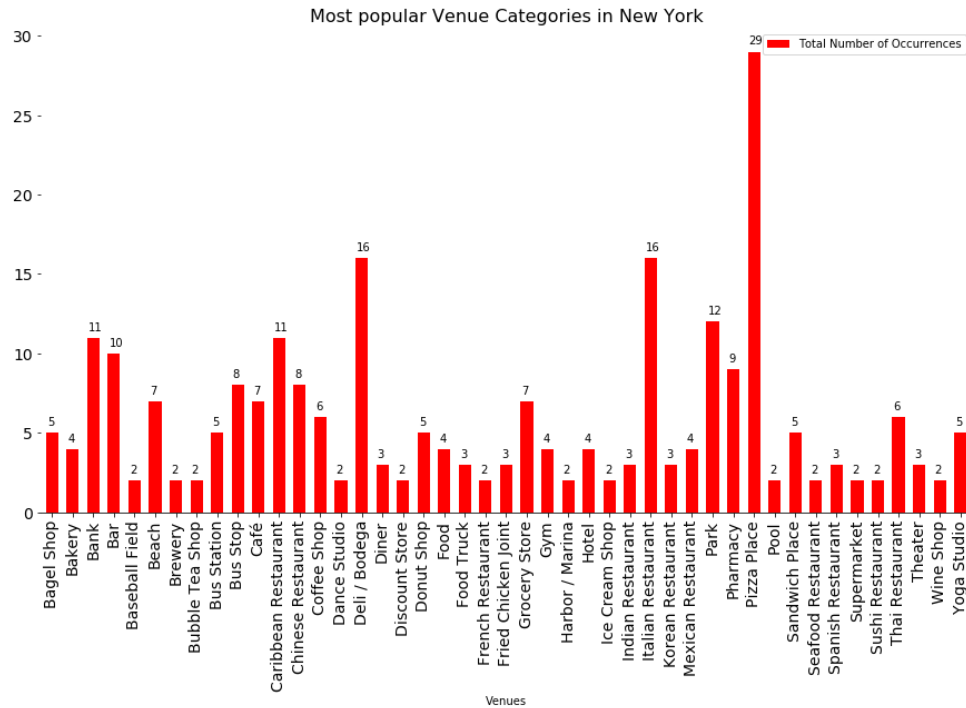
Toronto with neighborhoods superimposed on top

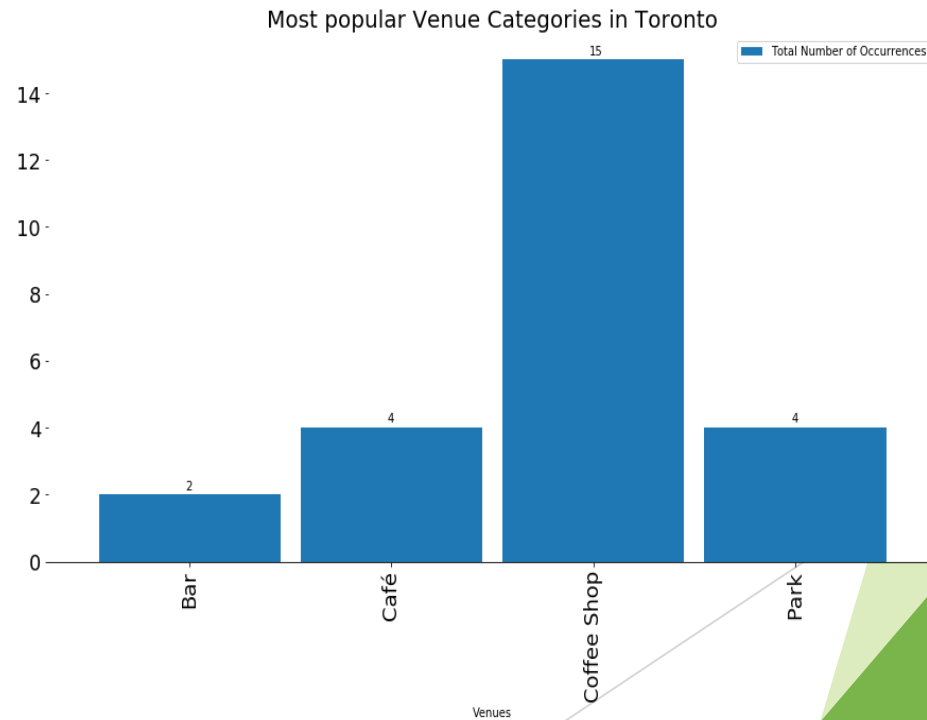**New York map with neighborhoods superimposed on top.**
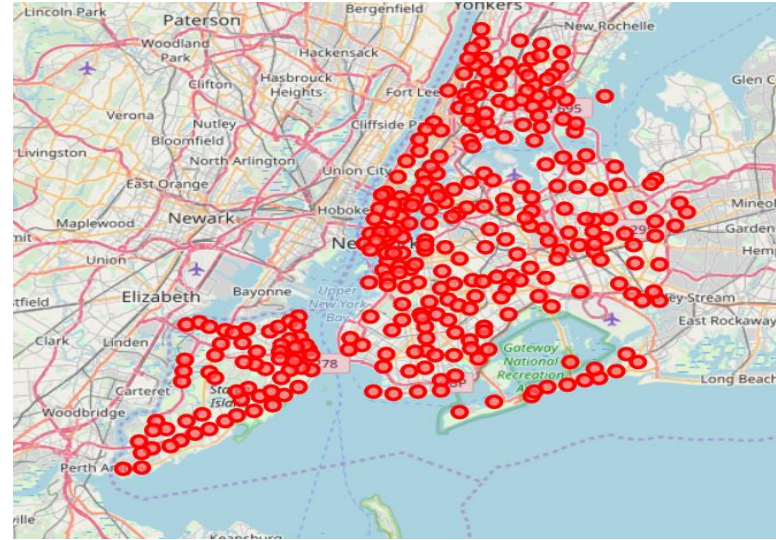
Manhattan Vs. New York

Clustering Model

K-Means Algorithm

▶ As a predictive model we will use a Machine Learning algorithm belonging to the unsupervised learning called K-Means, which mainly consists of the following steps.

1. Randomly placing k centroids, one for each cluster.

2. Calculate the distance of each point from each centroid.

3. Assign each data point (object) to its closest centroid, creating a cluster.

4. Recalculate the position of the k centroids.

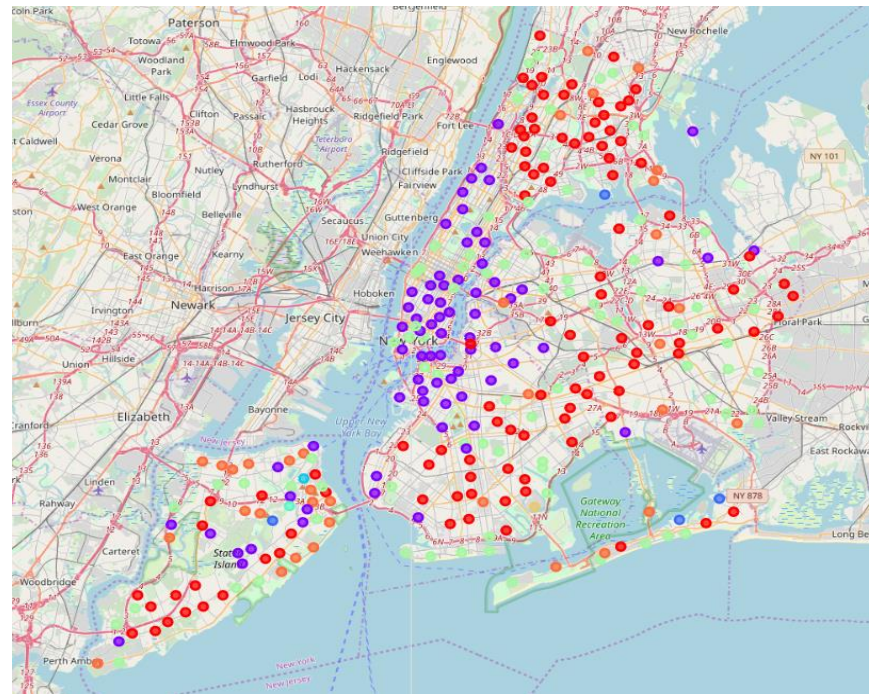5. Repeat the steps 2-4, until the centroids no longer move.
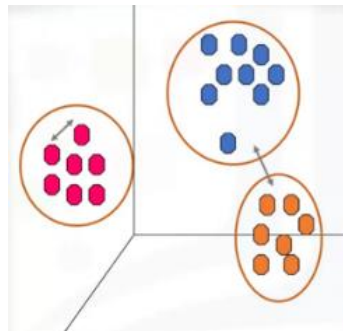
Clustering Results

- **Toronto clustering results**
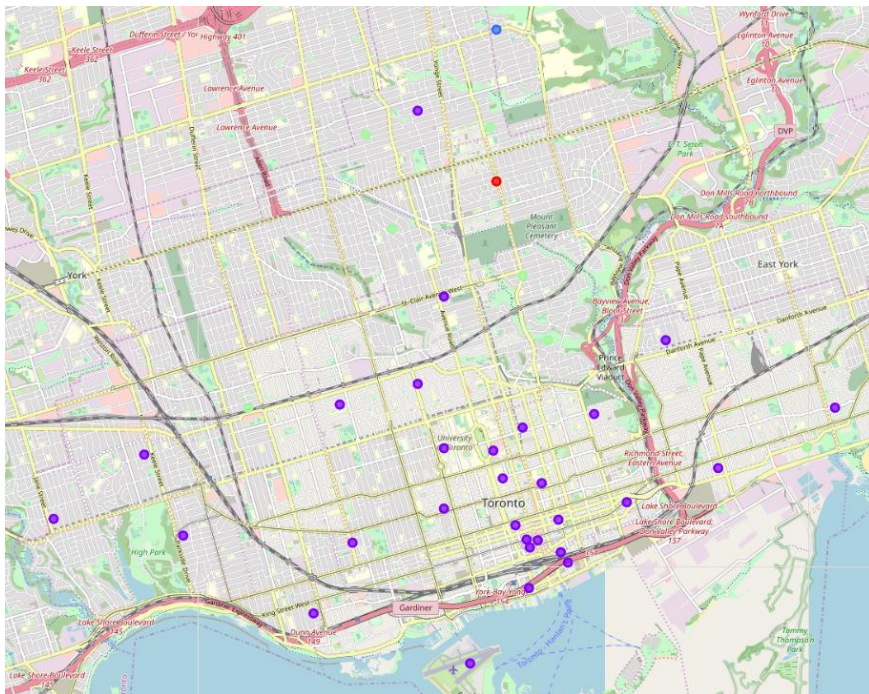
- **New York clustering results**

Conclusion

▶ Purpose of this project was to identify the similarities and differences between the city of New York and the city of Toronto, a task that was not simple due to the large number of factors that can influence the outcome in addition to the vision and purpose have when analyzing the data. As a result of this project we can conclude that despite the geographic differences between the two cities together with the population difference, we were able to successfully segment both cities into eight clusters which show us remarkable characteristics that the neighborhoods between both cities share, mainly for the most common places that thanks at foursquare we have been able to recover.

▶ I hope this notebook is useful to someone who aspires to visit or live in either of the two New York or Toronto cities, or even to someone who has a business interest.