Capstone Project:
The Battle of Neighbourhoods

Naiara Marcos Gonzalez

May 28, 2020

1. Introduction

1. Background

Finding similarities or dissimilarities between cities can help in expanding business, copying infrastructure or so on.

2. Problem

Comparing three cities: London, New York and Toronto. We explore the cities through the 10 most common venues that contain neighbourhoods from the cities. We find the patterns using k-means algorithm.

2. Data

We use different Website for each cities:

- 1. New York
  `https://geo.nyu.edu/catalog/nyu_2451_34572`
- 2. Toronto
  `https://en.wikipedia.org/wiki/List_of_city-designated_neighbourhoods_in_Toronto`
- 3. London
  `https://www.doogal.co.uk/london_postcodes.php` Note:

  We only select the neighbourhoods which belong to 1- 3 zones

We use The Foursquare API to get information about the venues of the neighbourhoods.

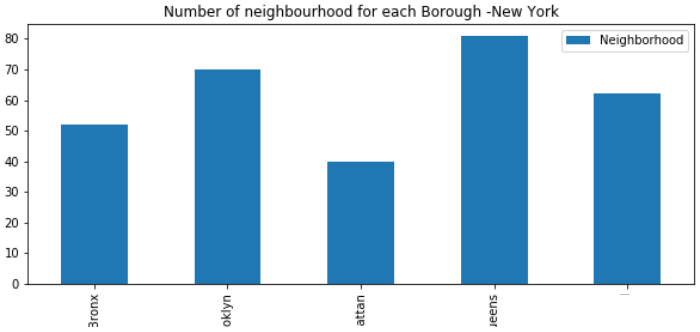
3. Methodology

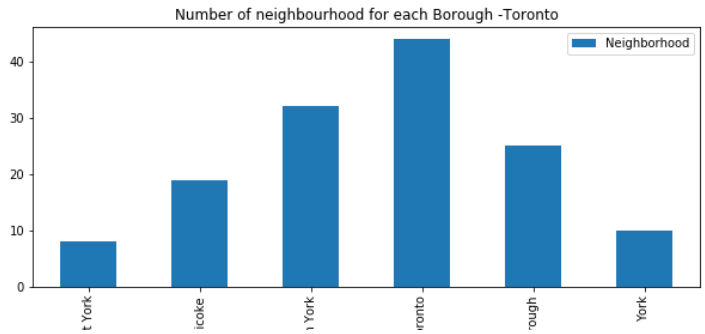
We follow the next steps:

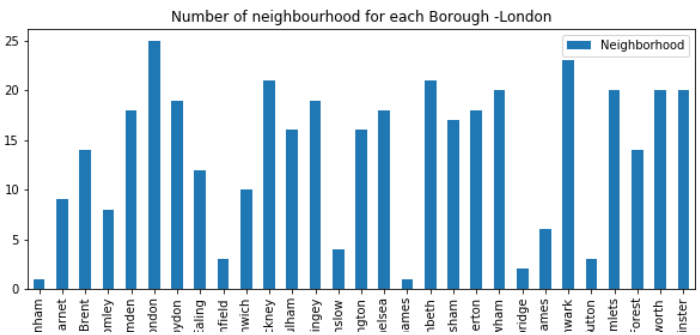
- 1. We analyse the neighbourhoods' venues, data from the Foursquare API, using different plots.
- 2. Selecting the 10 most common places for each neighbourhood.
- 3. We apply k-means algorithm from scikit-learn library to an unsupervised data.

4. Results

1. The neighbourhoods distribution around the boroughs from three cities.

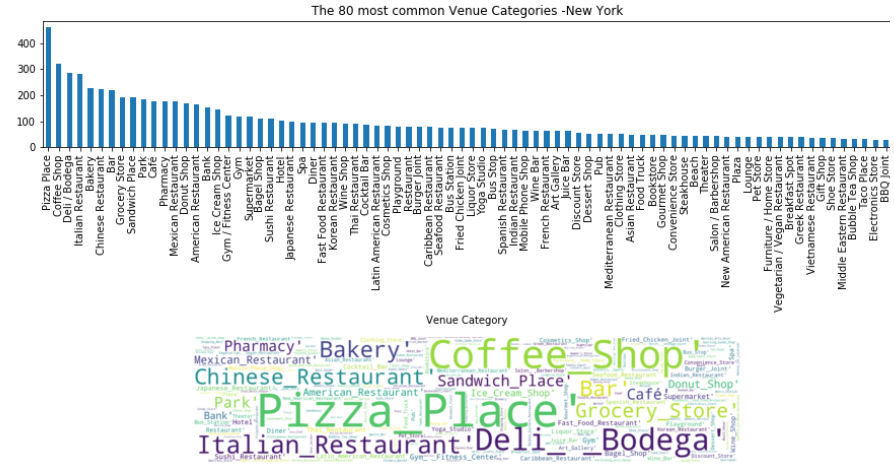


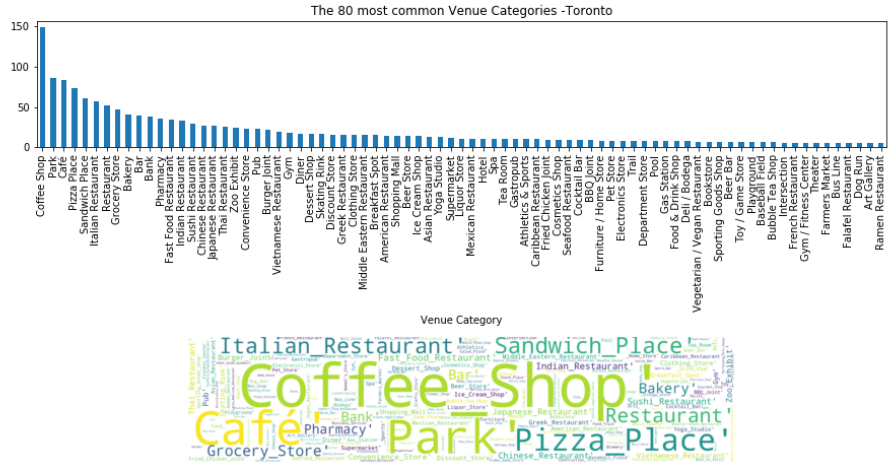


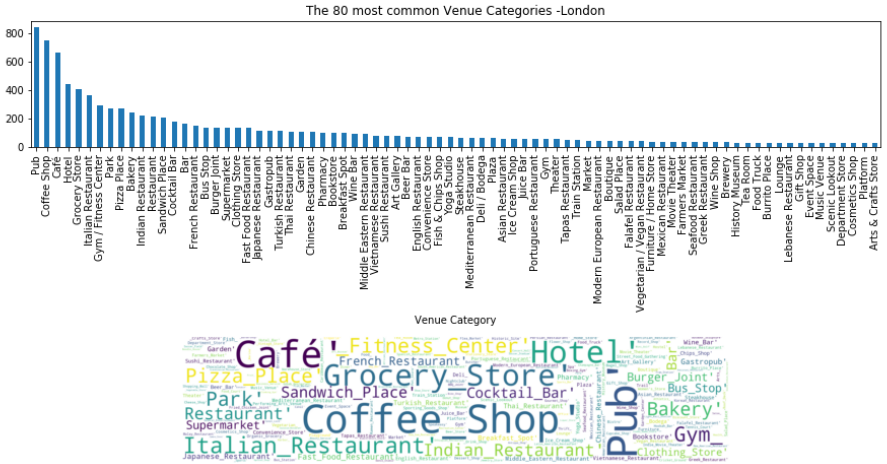


4. Results

2. Venues' category.



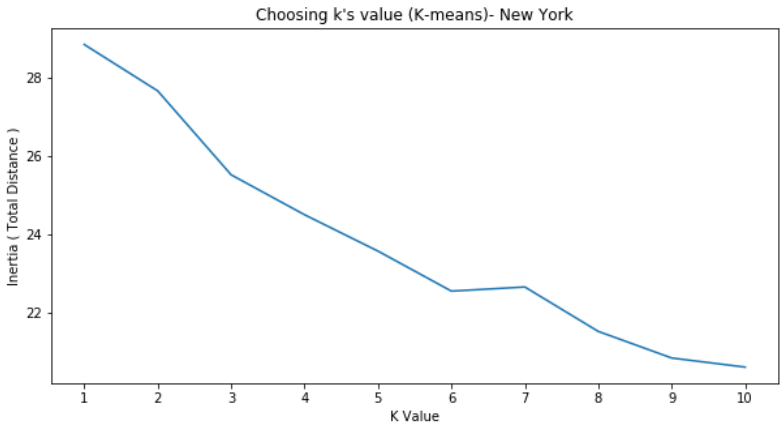


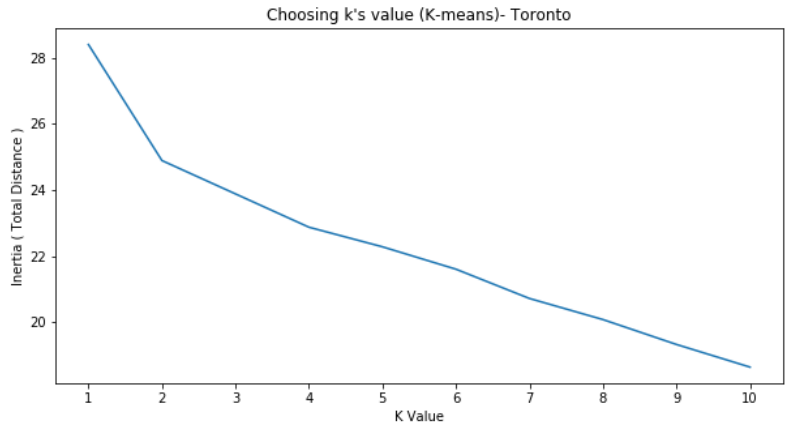


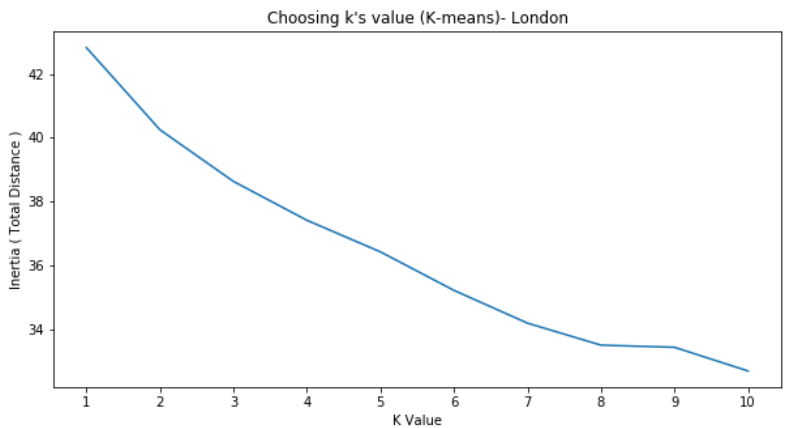
4. Results

3. Cluster neighbourhoods.

3.1. Choosing k-means

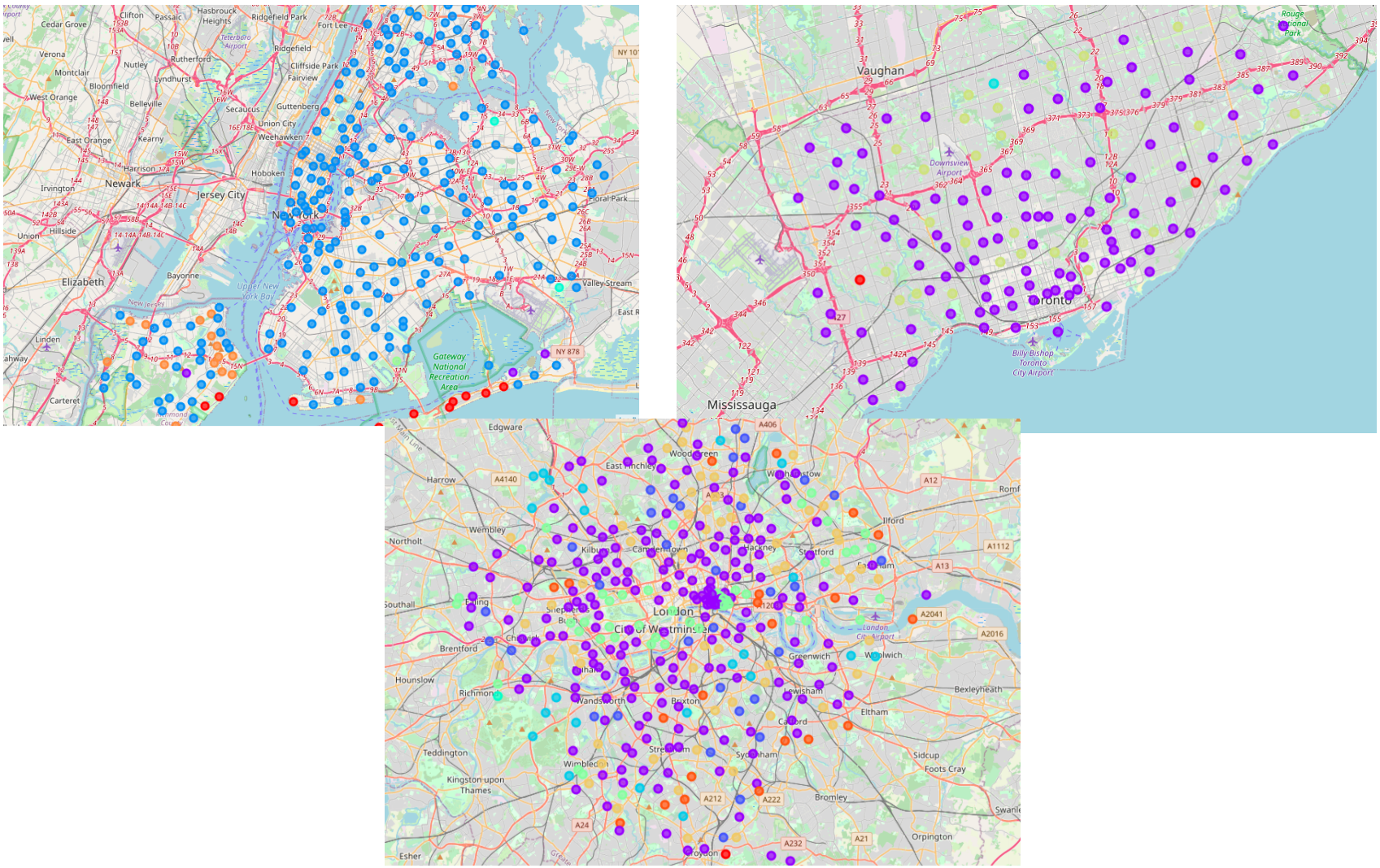






4. Results

3.2. Clusters' representation



5. Conclusion

Summing up:

- Comparing the three cities, we only use the 10 most common venues' category for each neighbourhoods.
- The places' category are very similar between the cities. Then, the clusters are very alike between them.
- It is very difficult to say whether London is more similar to New York or Toronto.

Interesting things to improve the analysis:

- To get a larger dataset, with more venues' category. Furthermore, to add new features such as: related to people who live in the neighbourhoods, the incomes by family, housing prices, etc.