Applied Data Science Capstone

Toronto and Manhattan neighborhoods comparison

Eduardo Palomero López, 2020

Index

- Introduction
- Data description
- Methodology
- Results
- Discussion
- Conclusion

Introduction

- During this Capstone course, we had the opportunity of analyse neighbourhoods of Manhattan (NYC) and Toronto
- We used k-means with k=5 to get clusters of neighbourhoods in these two cities
- For NYC we have observed that there were variety of neighbourhoods
  - There are 2 cluster with higher number of nodes, but, mostly, they are balanced
- But for Toronto, most neighbourhoods fall into one cluster while all other barely have 1 or 2 neighbourhoods
  - Does this mean that Toronto neighbourhoods are quite similar between them?
- If we were able to run this analysis with both cities, which NYC area will be more similar to Toronto?
- Under this assumption, would we still have most Toronto neighbourhoods in same cluster? Or they will be divided and falling into other clusters?

- This exercise has been thought as theoretical work for algorithm evaluation

Data description

- Data used is the same as it was used during this Capstone.

- NYC (Manhattan)
  - 5 boroughs and 306 neighbourhoods
  - In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the coordinates of each
  - Dataset can be found: https://geo.nyu.edu/catalog/nyu_2451_34572

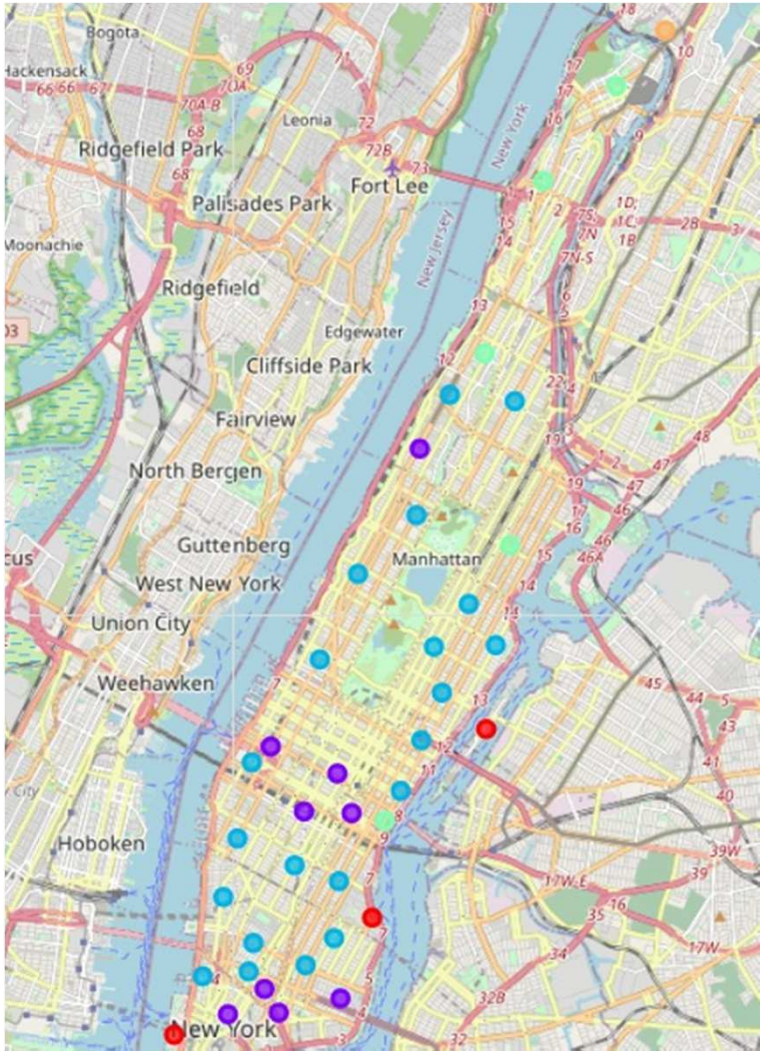
- Toronto
  - Data will be obtained from table in Wikipedia
  - We will use BeautifulSoup to parse the data.
  - The table is a list of the postal codes and neighbourhoods in Toronto

Methodology

- Data
  - Data sources have different formats
  - So they are imported and pre-process differently
  - At the end, both are imported to a Pandas dataframe to handle them
- Analysis
  - K-means algorithm will be used for clustering the neighbourhoods
  - Initially k=5 for all analysis
  - Analysis for Manhattan and Toronto
  - Final analysis for comparison

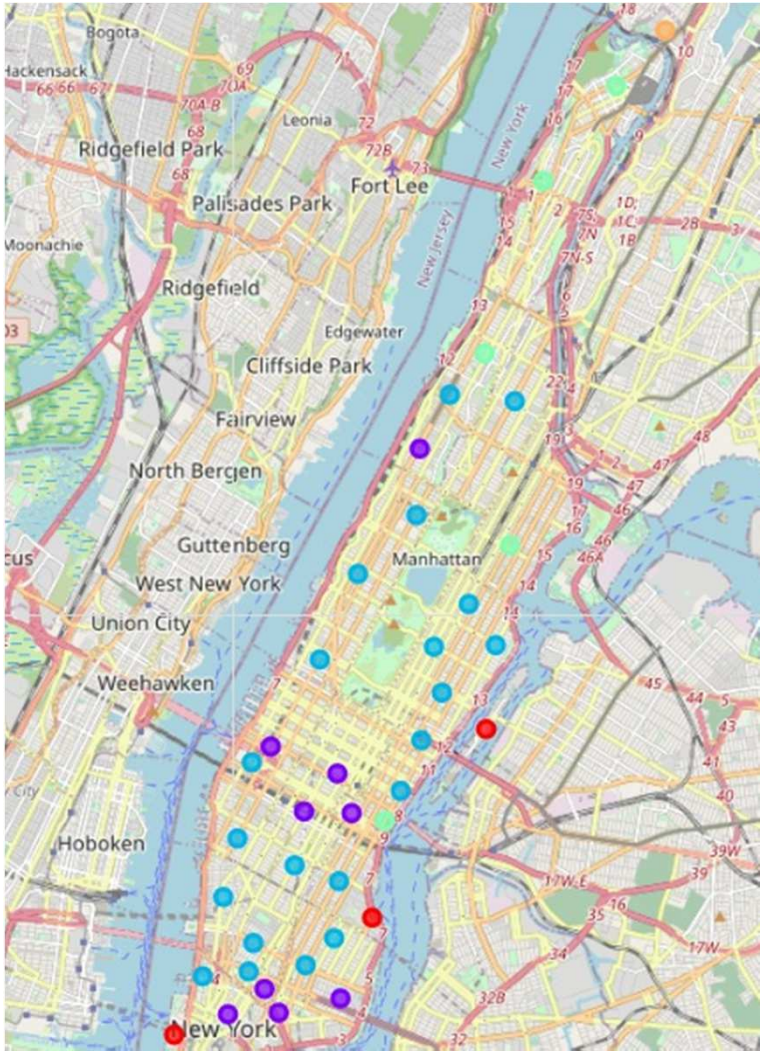
Results Manhattan (1)

- Cluster 0 (Red):
  - 3 neighbourhood in the cluster, spotted in the East and South, all 3 river side
- Cluster 1 (Purple):
  - Second most common, with 9 neighbourhoods, mostly at centre and South-East
- Cluster 2 (Blue):
  - Most of the neighbourhoods of Manhattan are grouped in this cluster



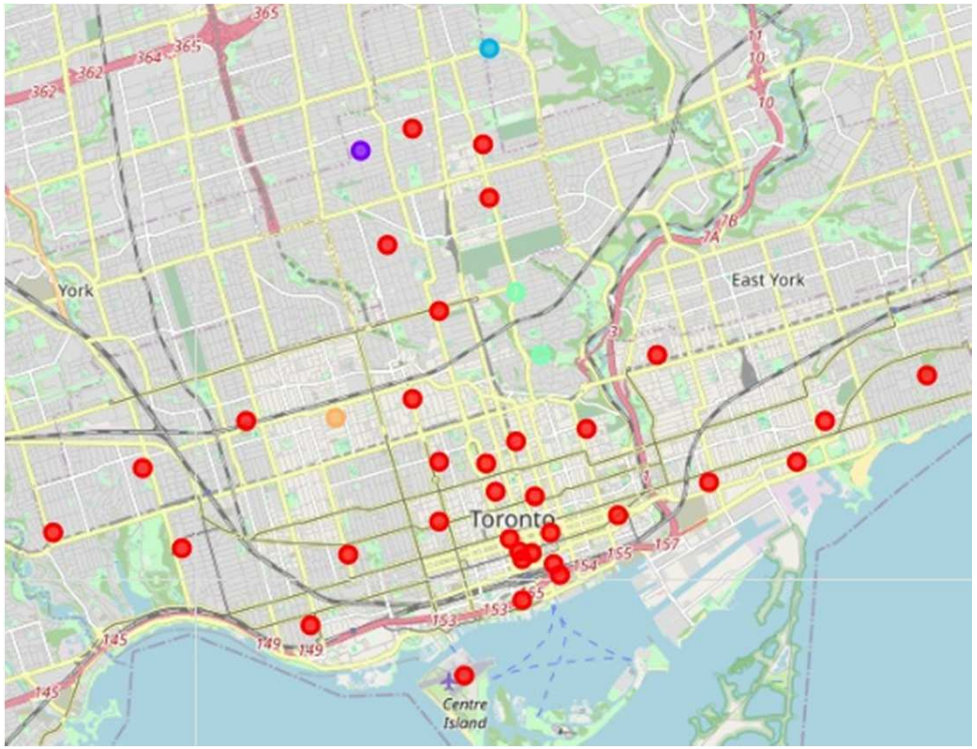
Results Manhattan (2)

- Cluster 3 (Green):
  - 5 neighbourhood in the cluster, 4 of them at the North
- Cluster 4 (Orange):
  - Only one neighbourhood in the cluster, Marble Hill, placed at the North of Manhattan



Results Toronto

- Cluster 0 (Red):
  - Almost all neighbourhoods in Toronto fall into this cluster
- Only 1 neighbourhood in the following clusters:
  - Cluster 1 (Purple)
  - Cluster 2 (Blue)
  - Cluster 4 (Orange)
- Cluster 3 (Green): 2 adjacent neighbourhoods



Results Manhattan and Toronto individually - **highlights**

- Manhattan
  - Variety of neighbourhoods
  - Cluster 2 and 4 are obviously unbalanced, but others are

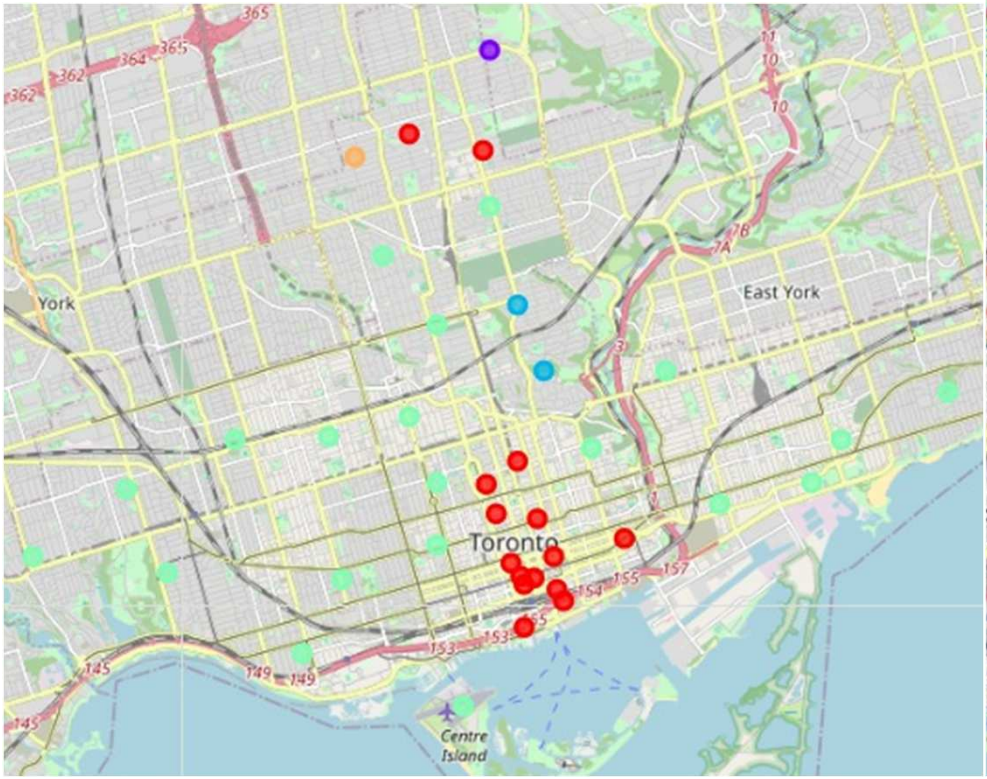
Cluster	# of Neighbourhoods
Cluster 0	3
Cluster 1	9
Cluster 2	22
Cluster 3	5
Cluster 4	1

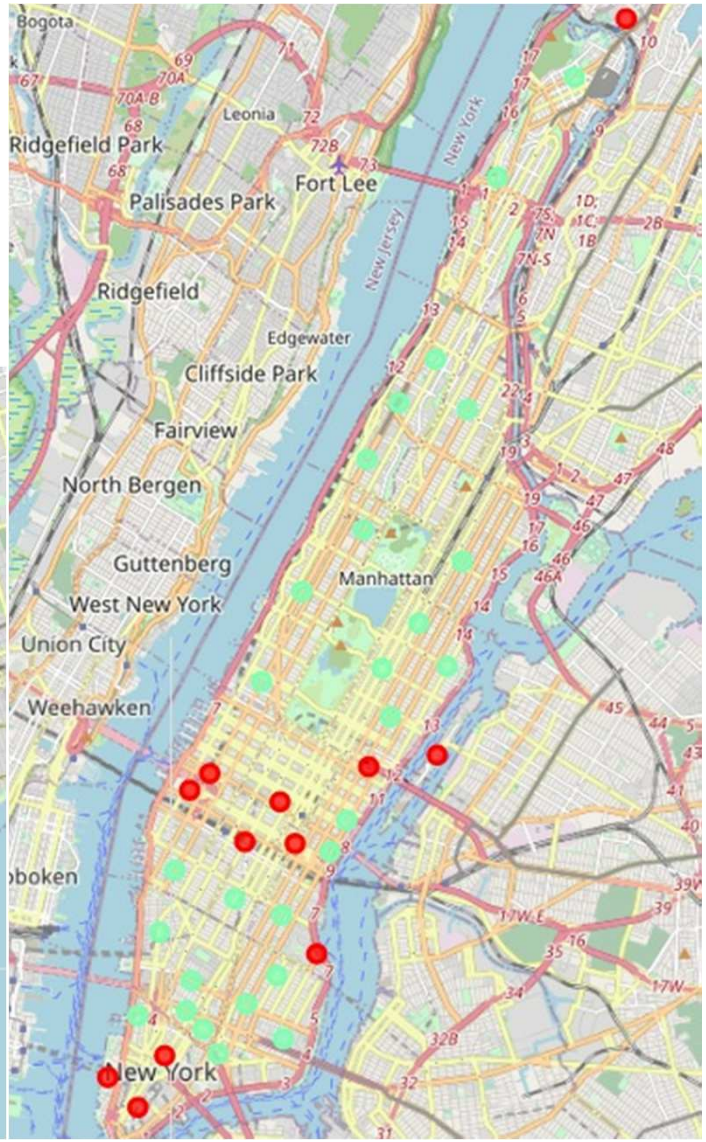
- Toronto
  - Totally unbalanced
  - 1 cluster with most of the neighbourhoods
  - All others having 1 or 2

Cluster	# of Neighbourhoods
Cluster 0	34
Cluster 1	1
Cluster 2	1
Cluster 3	2
Cluster 4	1

- Does this mean that most Toronto neighbourhoods are very similar?

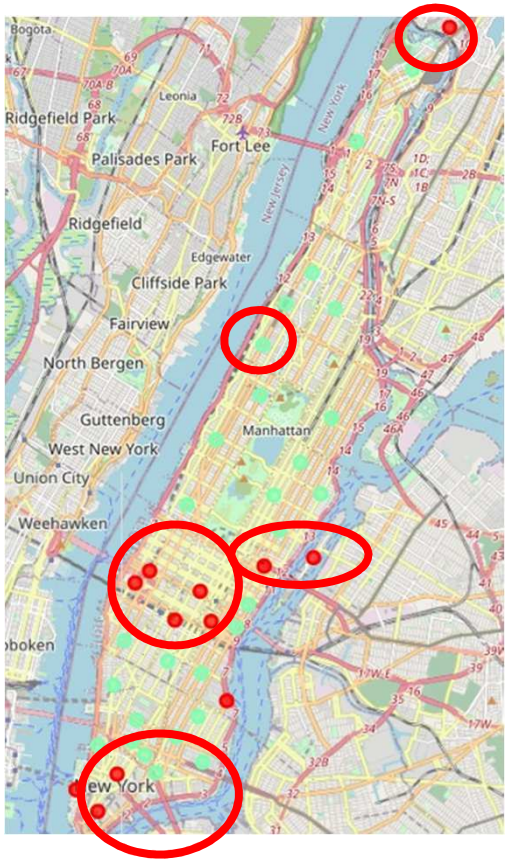
Results – Toronto and Manhattan clustering (1)





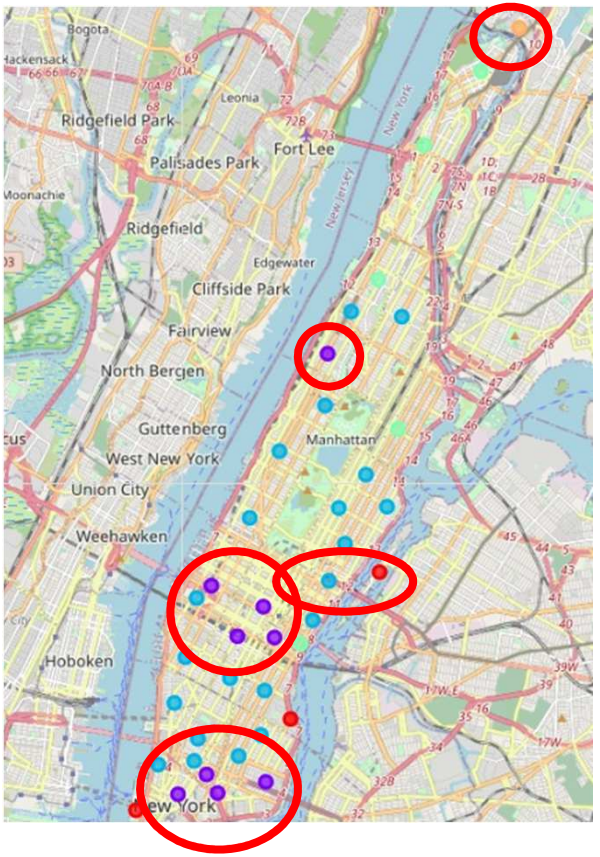
Results – Toronto and Manhattan clustering (2)

Analysis With Toronto



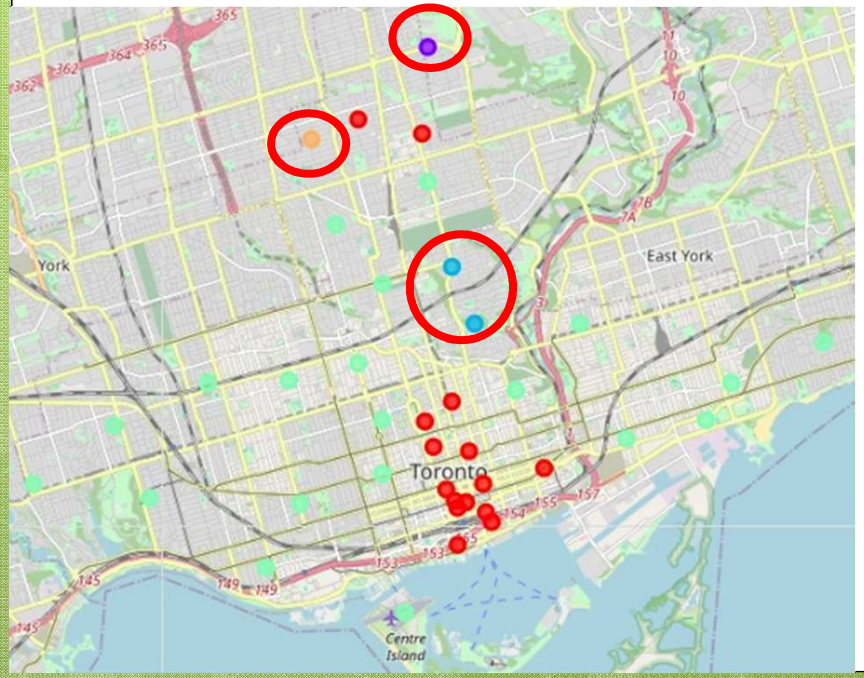
- Now we have ONLY 2 clusters present in Manhattan
- All other 3 clusters have been merged

Analysis ONLY Manhattan

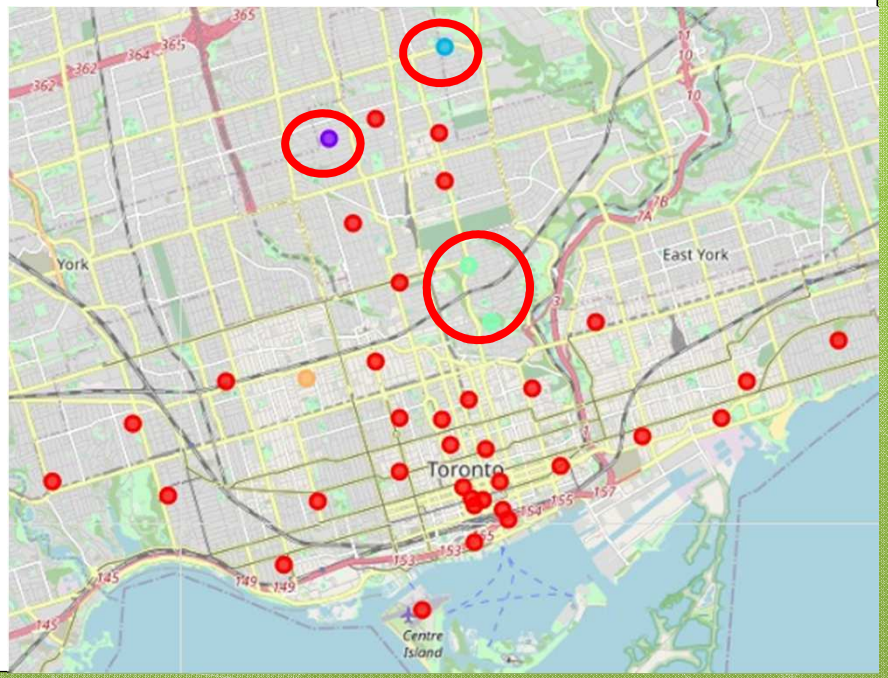


Results – Toronto and Manhattan clustering (3)

- 3 out of 5 clusters in Toronto are fully maintained same as ONLY Toronto analysis
- Another one has been integrated
- The big one has been split into two, matching the two clusters in Manhattan



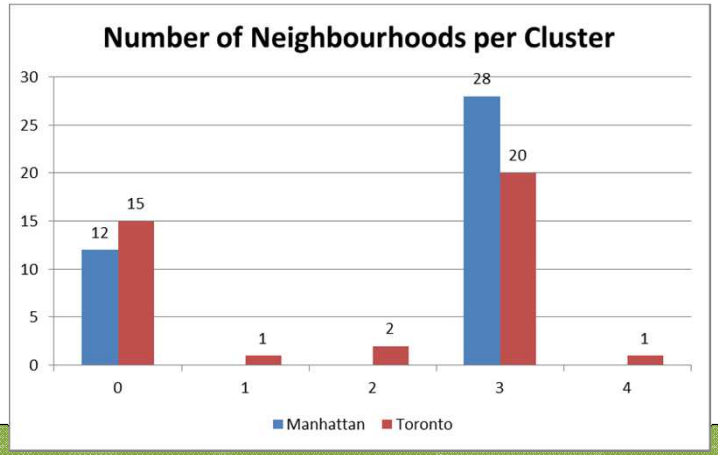
Analysis With Manhattan



Analysis ONLY Toronto

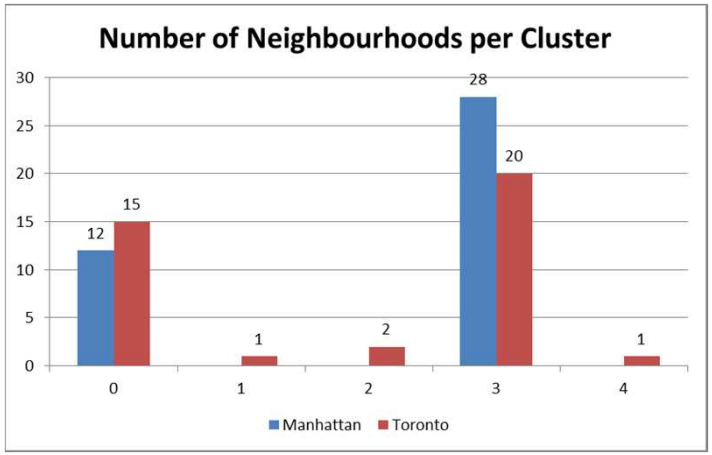
Discussion

- Manhattan had several defined areas different between them
- When compared with Toronto, Manhattan has not so many differences and all neighbourhoods fall into 2 clusters



Discussion

- Toronto was very similar, with 34 neighbourhoods in the same cluster
- When compared with Manhattan, it is observed 2 different areas (clusters 0 and 3)
- Clusters 1, 2 and 3 from ONLY Toronto analysis are kept into new clusters 1, 2 and 4



Conclusion

- We can conclude that Toronto and Manhattan are quite similars in terms of venues, which is what was analysed here
- The following Toronto Neighbourhoods seem to be quite different to the rest of the city as when compared with Manhattan they have no other neighbourhood similar to them
  - Lawrence Park (cluster 1)
  - Moore Park, Summerhill East & Rosedale (cluster 2)
  - Roselawn (cluster 4)