

Comparing Neighbourhoods in Toronto and Manhattan Using K-Means Clustering

Dr Mehmood Syed

30th May 2020



Introduction

- John Smith is a 37 year old IT Professional living in the neighbourhood of St. James Town in Toronto, Canada.
- John has a particular affinity to his current neighbourhood because of the eclectic mix of restaurants, bars, cafes and other venues in his immediate vicinity.
- For personal reasons however, John is required to relocate to Manhattan, New York. He is therefore seeking a shortlist of suitable neighbourhoods in Manhattan to which to relocate which are similar to St. James Town in terms of the local mix of venues.



Data

- To provide a suitable recommendation to John Smith on the neighbourhoods in Manhattan which are similar to St. James Town, all neighbourhoods in both Toronto and Manhattan will be clustered using k-means clustering.
- The clustering will be based on the composition of the top 10 venues
- within a 500m radius of the geographic center of each neighbourhood. Those neighbourhoods in Manhattan which are also in the same cluster as St. James Town will then comprise the shortlist of neighbourhoods that John should consider relocating to.

Data

- The FourSquare.com database has been queried previously and the top 10 venues within 500m of each neighbourhood in Manhattan have been identified. The latitude and longitude of each venue and each neighbourhood have been identified (see Coursera Capstone Project Week 3). Similarly, the top 10 venues within 500m of each neighbourhood of Toronto have been identified in the Capstone Project Week 3 Student Assignment.
- The link to the Jupyter Notebook explaining that analysis can be found here:
- https://github.com/msyed187/Coursera_Capstone/blob/master/Neighbourhood%20Clustering%20MSyed.ipynb

Method

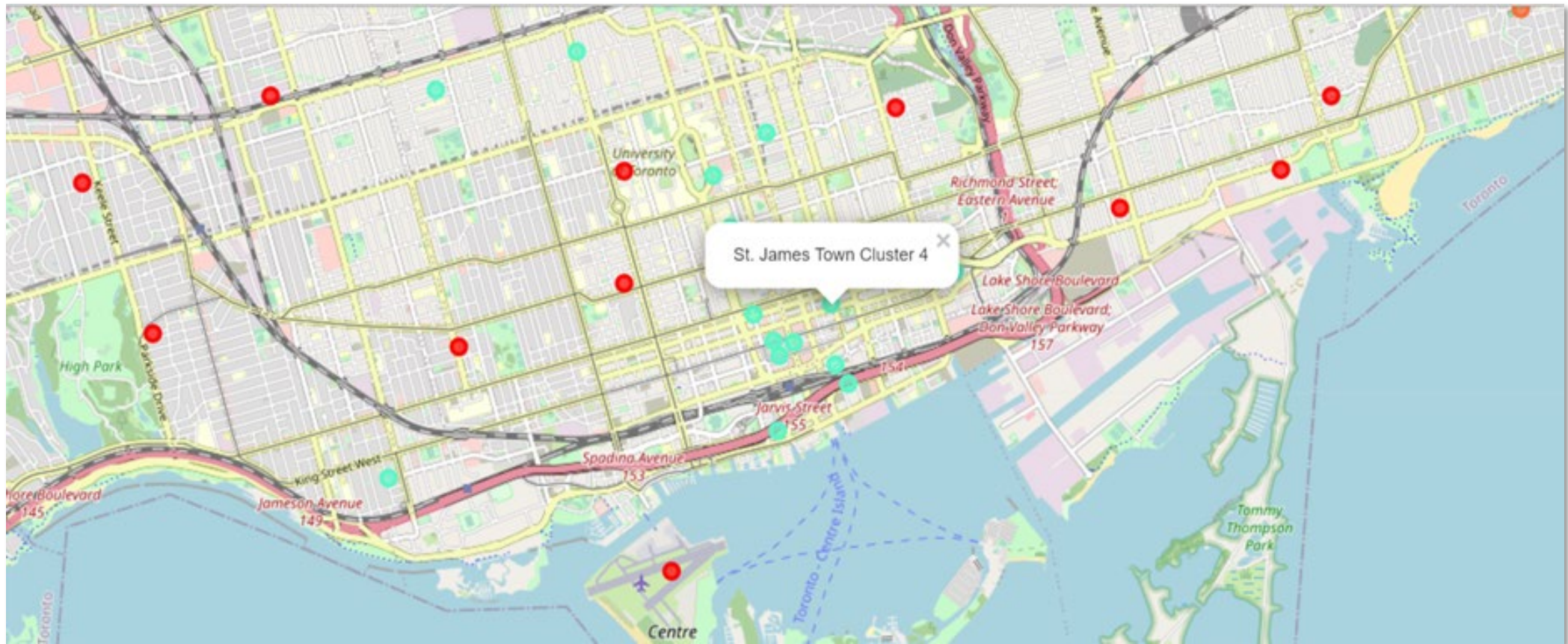
- The following approach has been used to derive the results. Details are provided in the Final Report and Jupyter Notebook associated with this assignment:
 - Reload data showing neighbourhoods and top 10 venues within 500m of each neighbourhood in both Manhattan and Toronto
 - Data are combined into a single dataframe
 - K-Means Clustering is conducted on the combined dataframe
 - The Cluster in which John Smiths' current hometown of St. James Town is identified.
 - The neighbourhoods in Manhattan in the same cluster are identified, thus identifying the shortlist of possible neighbourhoods for John to consider in his search for a new home in Manhattan



Results

- 4 possible neighbourhoods in Manhattan have been identified as being in the same cluster as
- St. James Town.
- These are:
 - Marble Hill
 - Morningside Heights
 - Battery Park City
 - Financial District
 - Carnegie Hill

Map of Clusters in Toronto



Map of Clusters in Manhattan



Discussion & Conclusion

- The above clustering of neighbourhoods in both Toronto and Manhattan successfully segments the neighbourhoods according to the mix of top 10 venues within a 500m radius of each neighbourhood. It is notable however, that several of the clusters contain multiple neighbourhoods and a lower value of k would result in too few clusters to make a meaningful recommendation.
- K-means clustering is however an appropriate method of segmenting geographical neighbourhoods according to some distinct feature of each neighbourhood. In this example we have been able to make a shortlist of 4 neighbourhoods in Manhattan which are similar to John Smith's current location of St. James Town, Toronto. Thus, John is able to confidently limit his search for a new home to the four areas listed in the Results Section.