

Capstone Week 5 Assignment

Comparing Neighbourhoods in Toronto, Canada and Manhattan, New York, using K-Means Clustering

By Dr Mehmood Syed

1 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.

2 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 centre 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.

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

3 Results

Four 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

The clusters are identified on the two maps below:

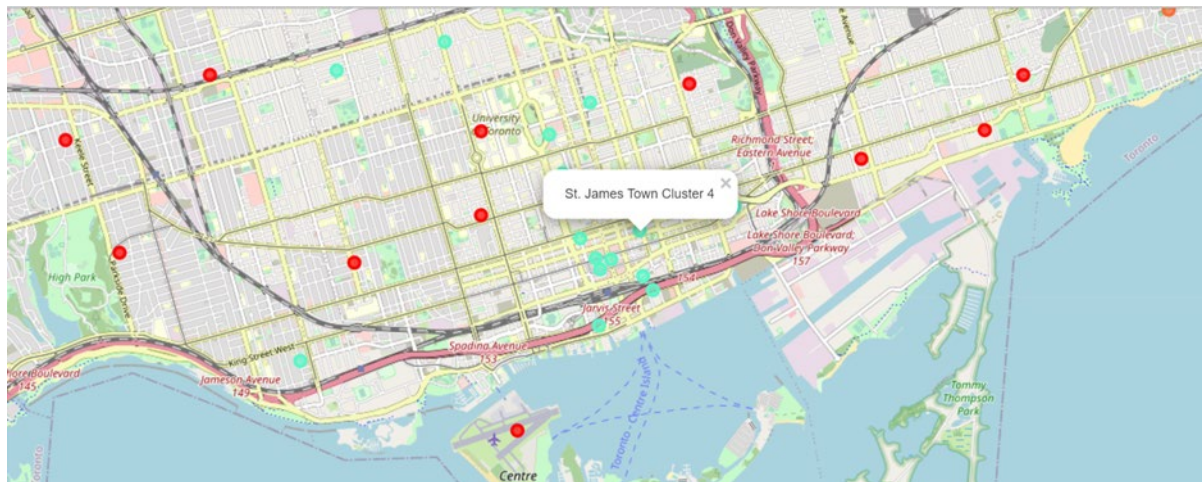


Fig.3.1: Map of Neighbourhood Clusters in Toronto

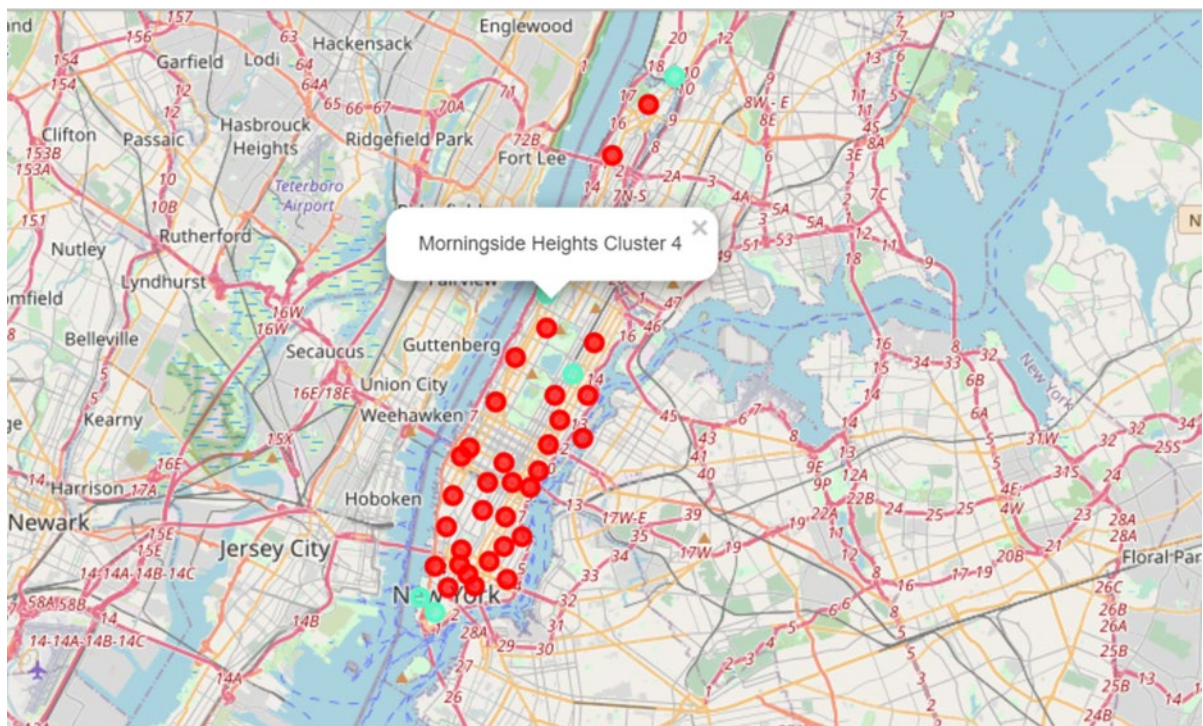


Fig 3.2: Map of Neighbourhood Clusters in Manhattan

Discussion

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.

Conclusion

K-means clustering is 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.