Capstone Project Report:
Urban Planning of Financial Capitals; A Cluster Analysis of Neighbourhoods

I.   Introduction

The purpose of this report is to compare the dissimilarities in urban planning of City of Toronto, and City of New York. Particularly focusing on their commercial centres, Inner Toronto, and Manhattan Island.

Financial centres are known for their particular urban planning schemes. Toronto, financial centre of Canada, may wish to align its urban planning with that of New York, considered the financial capital of the world. However, New York is not known for systematically addressing the needs of its citizens. Instead, the city is a result of centuries of iterative changes. Thus, it can be hypothesized, that the city is an example of a "natural" financial centre.

Therefore, this report carries out a clustering analysis of both Toronto, and Manhattan neighbourhoods, to find dissimilarities in structure to help Toronto's future plans to compete with New York as a financial centre.

II.   Data

There are three matching datasets for each city. First, a list of neighbourhoods, boroughs, and postcodes obtained from respective postal service listings. Second, geographical coordinates based on aforementioned postcodes, obtained from the open-source database Nominatim. Third, a list of venues found around each geographical coordinate obtained from Foursquare's database.

Combined, these datasets enable comparison between types of venues in Toronto, and Manhattan neighbourhoods. Specifically, it allows for similar neighbourhoods to be grouped together based on frequency of particular venues. For example, two neighbourhoods containing similar occurrence of restaurants focused on a single cuisine may be defined as similar, and to be within a single group.

III.   Methodology

a. Pre-processing and Features

In order to analyse neighbourhood venues with machine learning methodologies, categorical variables were hot coded into binary form, with a code of 1 if a venue type was found in a neighbourhood. This allowed for summation of frequencies, and their averages. The final processed dataset contained neighbourhood, borough, postcode, latitude, longitude, and the average amount of venues by type.

b. Models, and Techniques

K-means clustering was used to group neighbourhoods into clusters based on venue occurrence. The main drawback of using K-means was that there were no clearly defined clusters in the dataset. This was found through the elbow method for computing the optimal amounts of clusters. Results generated lacked a clearly defined point where marginal decrease in mean squared error was severely reduced.

Figure 1: Optimal K for both cities
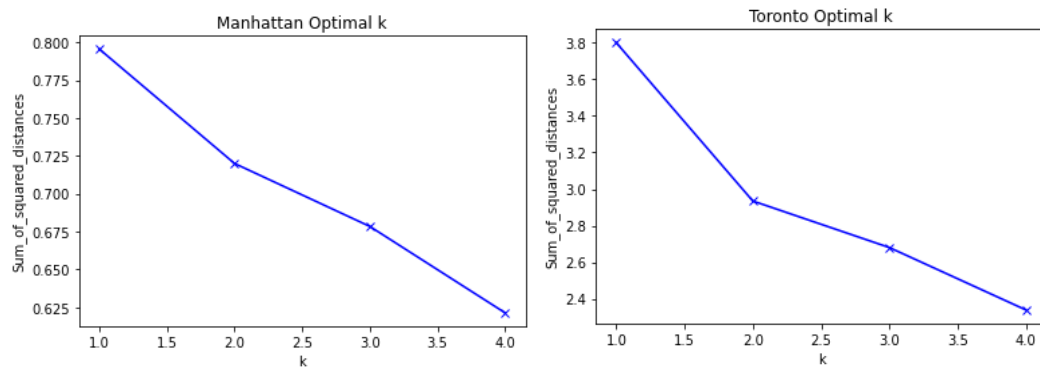
In order to account for this agglomerative clustering was additionally used. This allows for neighbourhoods to be clustered from the bottom-up, allowing a measurement of similarity.

IV. Findings

K-Means clustering resulted in 3 clusters for each city based on type of venues in the area. It seems that Toronto neighbourhoods have little in common with its New York counterparts. Toronto citizens seems to prefer coffee shops opposed to their New York counterpart's preference towards venues that serve alcoholic beverages.

One similarity between the cities is that both cities have a distinct neighbourhood dedicated to parks, and greenery of the city. It seems urban planners of both cities have adopted the same approach. Additionally, it seems Manhattan has a more varied offerings of venues by type of cuisine than Toronto.

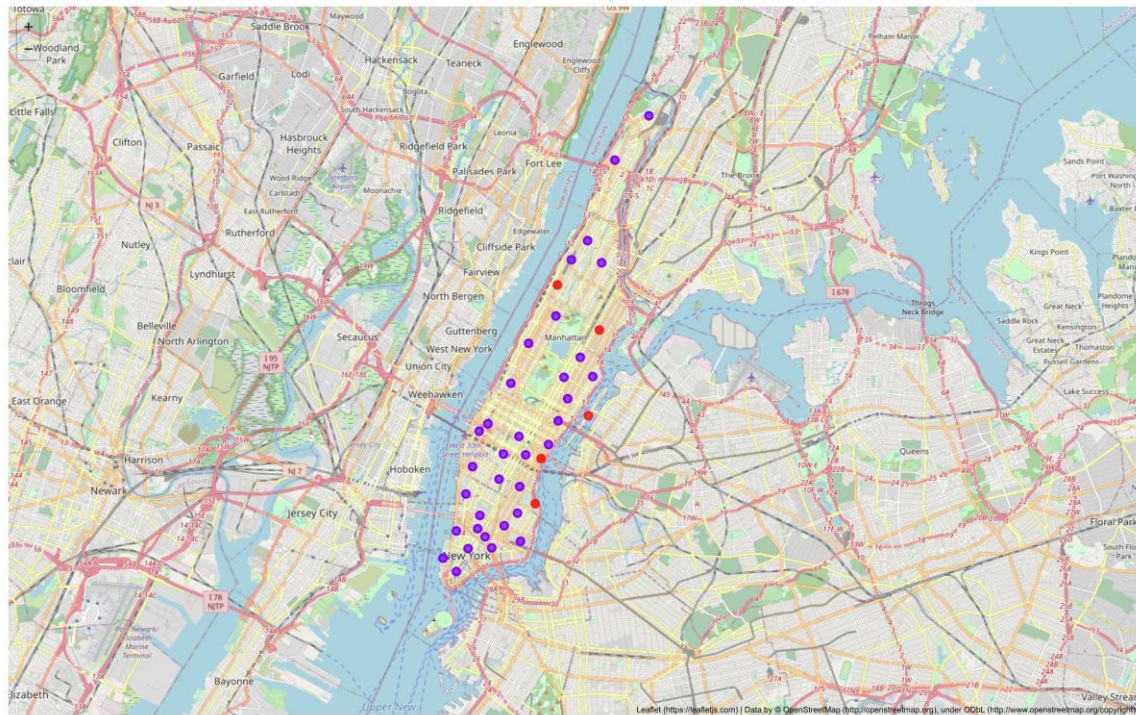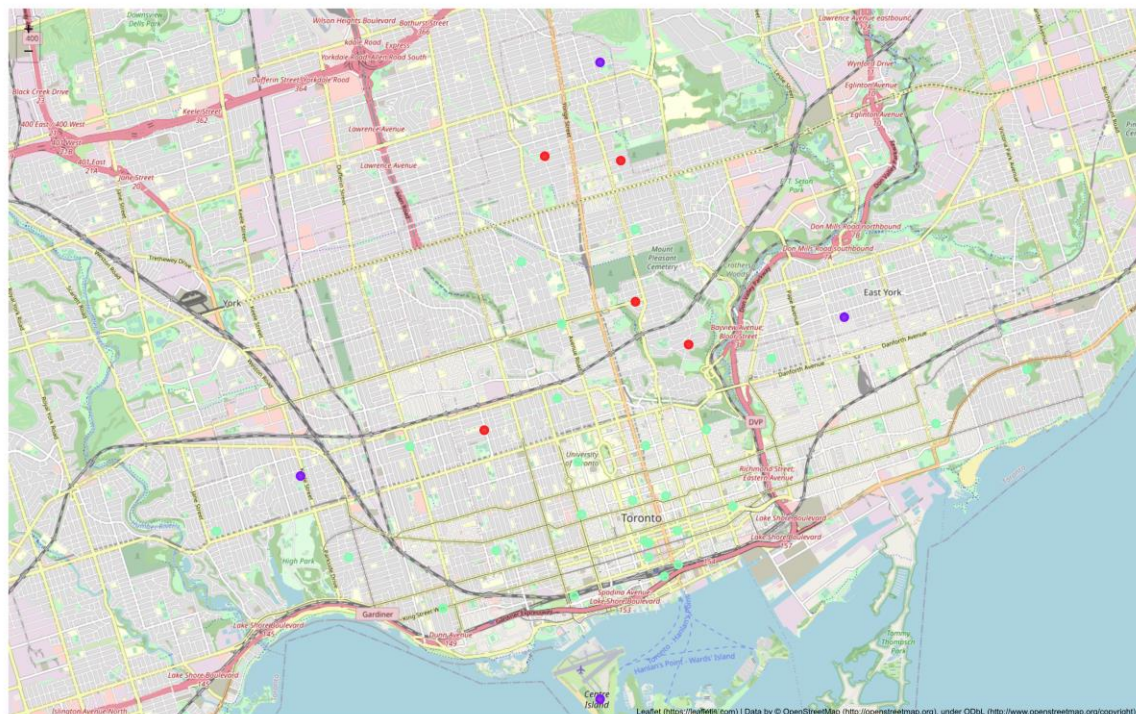*Figure 2: Clustering results for Manhattan*

Figure 3: Clustering results for Toronto

Agglomerative clustering reveals the real dissimilarities within the two cities. Manhattan seems to be more homogenous compared to Toronto. While the highest-level clusters in Manhattan have a distance of 0.4, the highest cluster in Toronto has a distance of 1.2. Additionally, with Manhattan, the highest-level cluster composes the neighbourhood of Tribeca, and everything else.

Figure 4: Dendrograms for the two cities
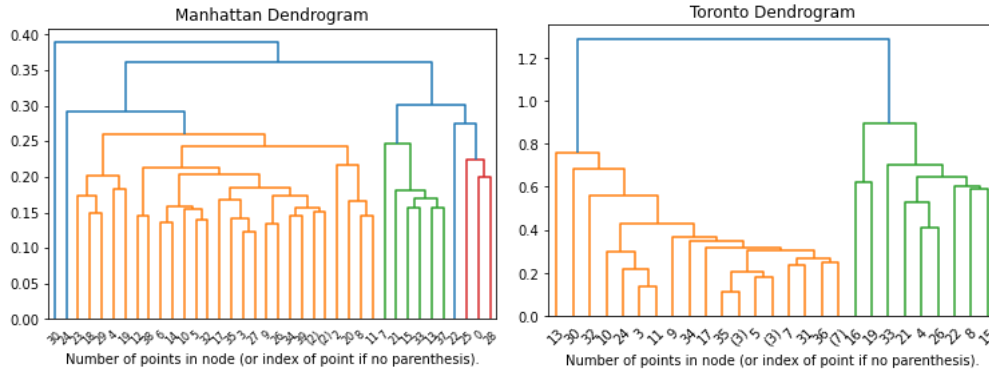
V.   Discussions

The findings seem to confirm the notion that Manhattan naturally developed into a financial centre rather than planned as one. Being more homogeneous indicates that there was no master plan to partition areas into distinct neighbourhoods. This is contrasted by Toronto, being three times more heterogenous than Manhattan.

This may be an indication that the secret to a performing financial centre is iterated responses to problems of the citizenry as they arise, rather than adhering to a master plan. This is also supported by the levels at which homogeneous clusters start to form in each city. In Toronto, there is a step-wise progression in clusters, while in Manhattan, clusters seem to appear at around the same distance.

Based on these findings, in order for Toronto to better emulate Manhattan, it should not rigidly follow a strict urban planning principle. Instead, it should opt for a more natural progression by responding to citizen needs in an iterative manner.

VI.   Conclusion

This report was commissioned to study the similarities, and dissimilarities of Toronto, and Manhattan as financial centres. It was found that clustering neighbourhoods by venues was useful in comparing the two cities from each other. Additionally, agglomerative clustering was used to measure homogeneity within the two cities. It was found that Manhattan likely developed naturally, resulting in more homogeneous neighbourhoods, while Toronto likely followed a rigid urban plan, resulting in heterogeneous neighbourhoods. This finding supports the recommendation that cities seeking to emulated Manhattan's success as financial centre should form a more iterative approach to urban planning.