Capstone Project Report - The Battle of the Neighborhoods

I. Introduction

Matilda needs to relocate from Yorkville, Toronto to New York City due family reason. She really likes the neighborhood of Yorkville in Toronto. The goal of this project is to help Matilda to find similar neighborhood in New York City so that she can better enjoy the life after relocation.

II. Data

Based on description of the problem, we need to find out the location data for all the neighborhoods in New York City and Yorkville, Toronto. They we need to find the most popular venues in both the Yorkville Toronto neighborhood and the New York City neighborhood. The venue data will be obtained using FourSquare.

III. Methodology

1. We need to get the geographic location data for both New York City and Yorkville, Toronto.
2. Data needs to be cleaned and transformed to be able to use.
3. Use FourSquare to get the top 100 venues in all the neighborhood of New York City and Yorkville, Toronto.
4. Find the mean of the frequency of venues in each category in each neighborhood.
5. Use k-means clustering algorism to cluster the neighborhood and find the most similar neighborhood in New York City to Yorkville in Toronto.

IV. Results

The clustering results showed that Gramercy is the most similar neighborhood in Manhattan New York to Yorkville in Toronto. By looking at the venues in the two neighborhoods, we do find many similarities.

V. Discussion

Due to the limitation of FourSquare, only Manhattan location data is used in this project. To better support our conclusion, or to find a more similar neighborhood in New York City, all the neighborhoods in New York City should be included.

The same methodology could be applied to a broader scope. We can find similarities of one neighborhood to another neighborhood in any cities in the world.

VI. Conclusion

In this project, we used machine learning techniques to find the most similar neighborhood in New York City compared with Yorkville in Toronto. Though the problem discussed here is quire specific, the same methodology could be applied to a broader scope. We can find similarities of one neighborhood to another neighborhood in any cities in the world.