The Battle of Neighborhoods

Author: Haroldo Vélez Lora

Motivation

Suppose a person wants to move from New York to Toronto for a job. This person does not know anything about Toronto and he would like to move into a place similar to the place where he lives now.

Is it possible to create a system that can help our user showing to him the similarities between this two countries?

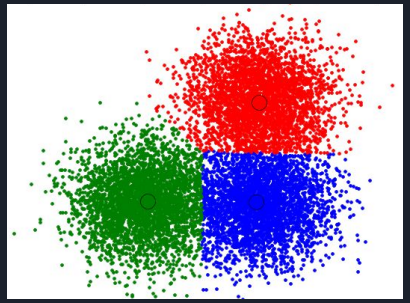
Objectives

Develop a system able to show similarities in terms of neighborhoods in order to help a user decide whether to move near the center of Toronto or not.

PROPOSAL

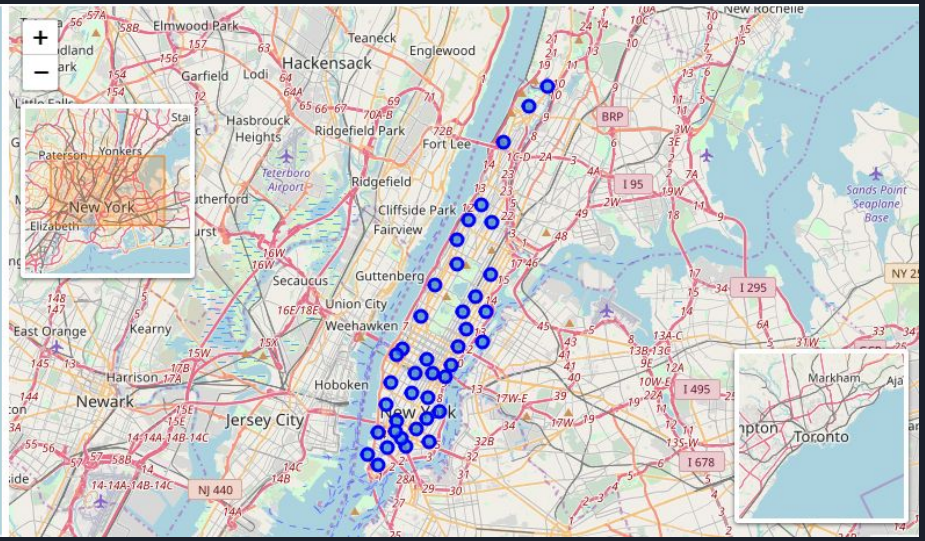
Approach

- Neighborhoods are downloaded
- Venues are requested using Foursquare API
- The categories of venues are encoded using One Hot
- K-means algorithm is used for finding similarities
- The elbow method is used for select K

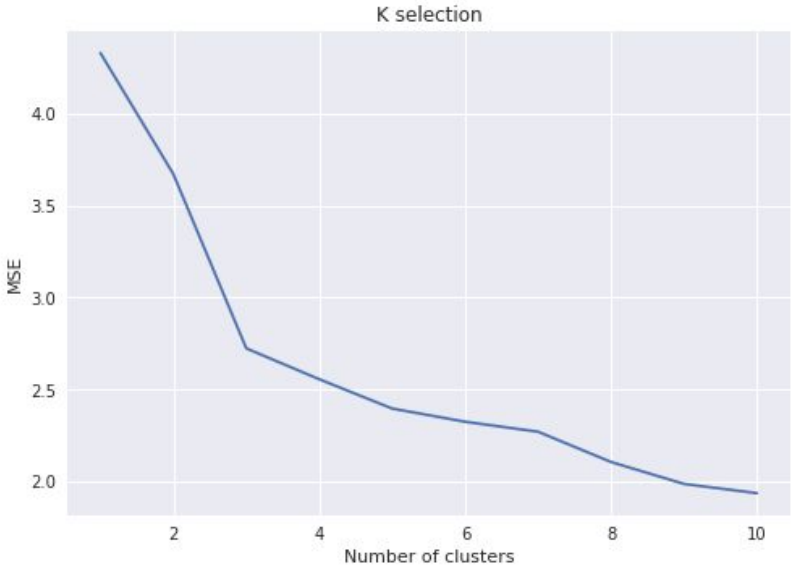


RESULTS

Geographical Location

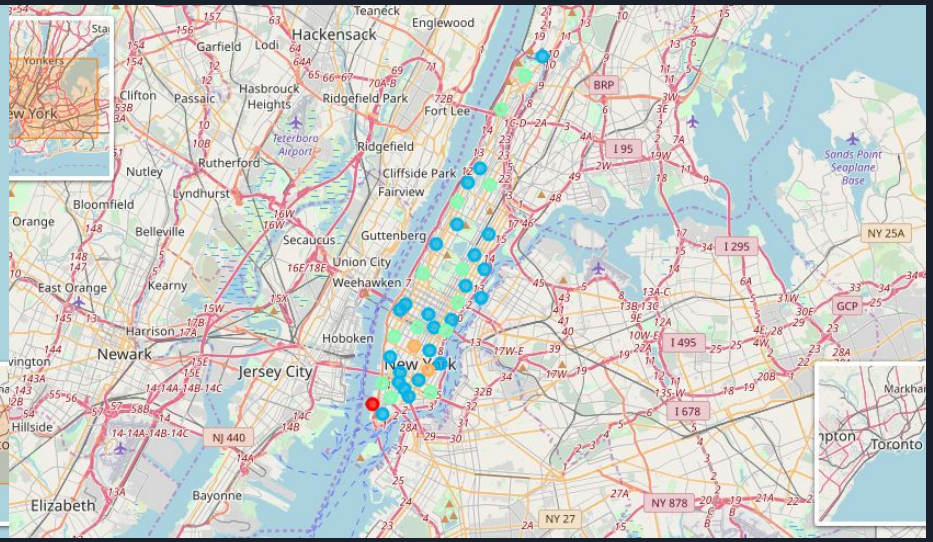


Selection of K

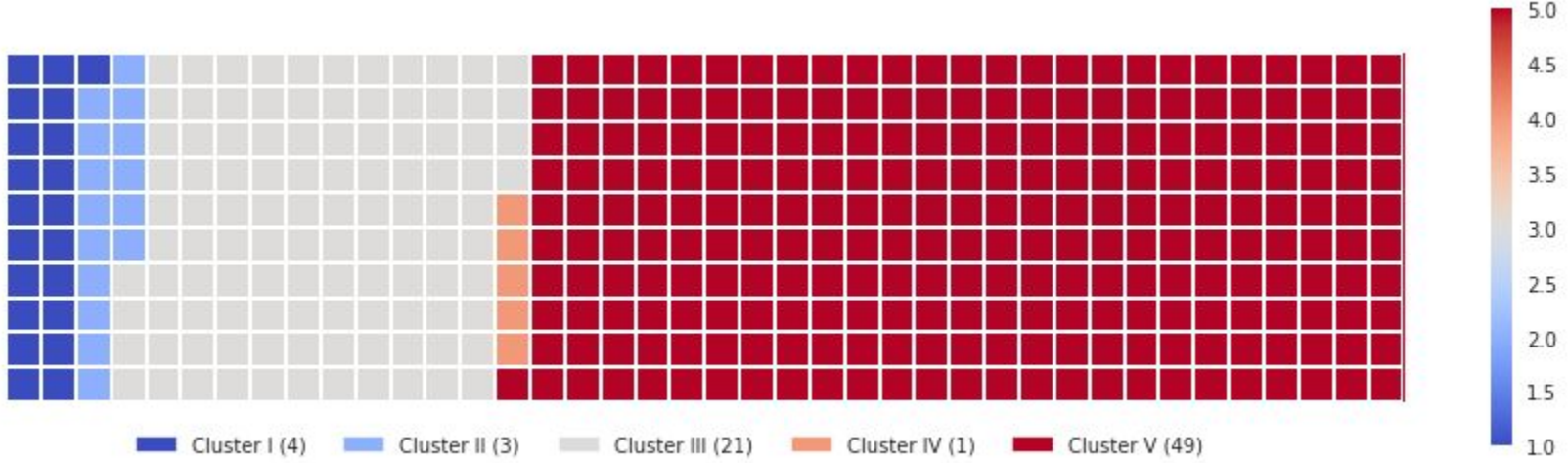


The best number of cluster is 5. That is, where the elbow is located. After that, the mean squared error decrease without big changes.

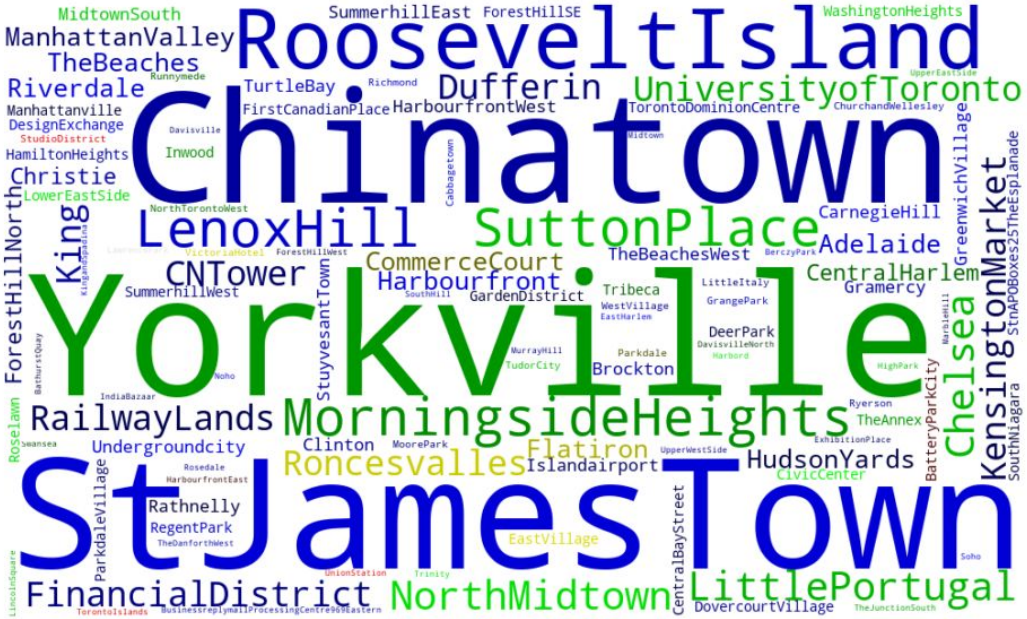
Geographical Location (Clustered)



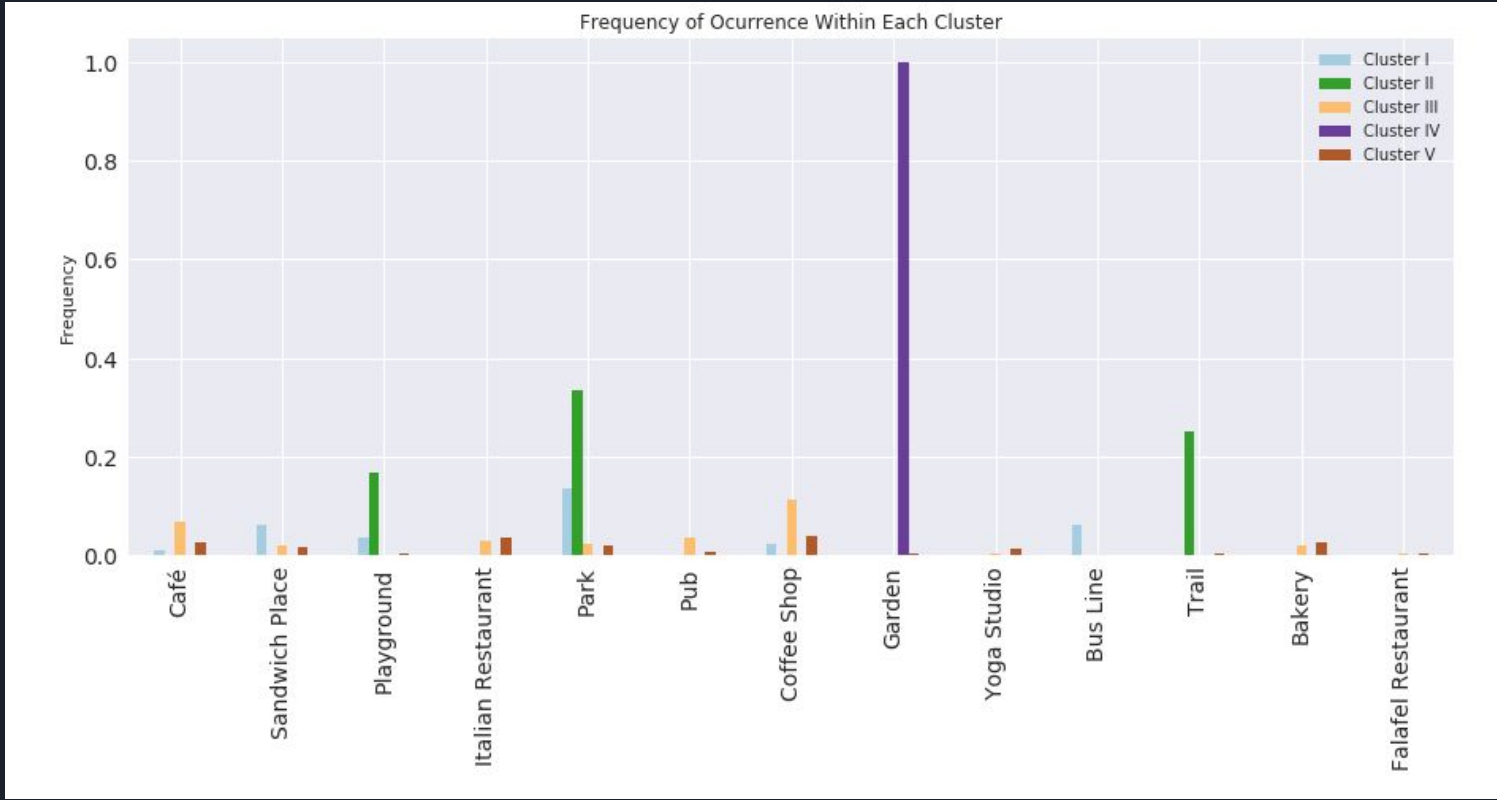
Proportion of Data Segmented



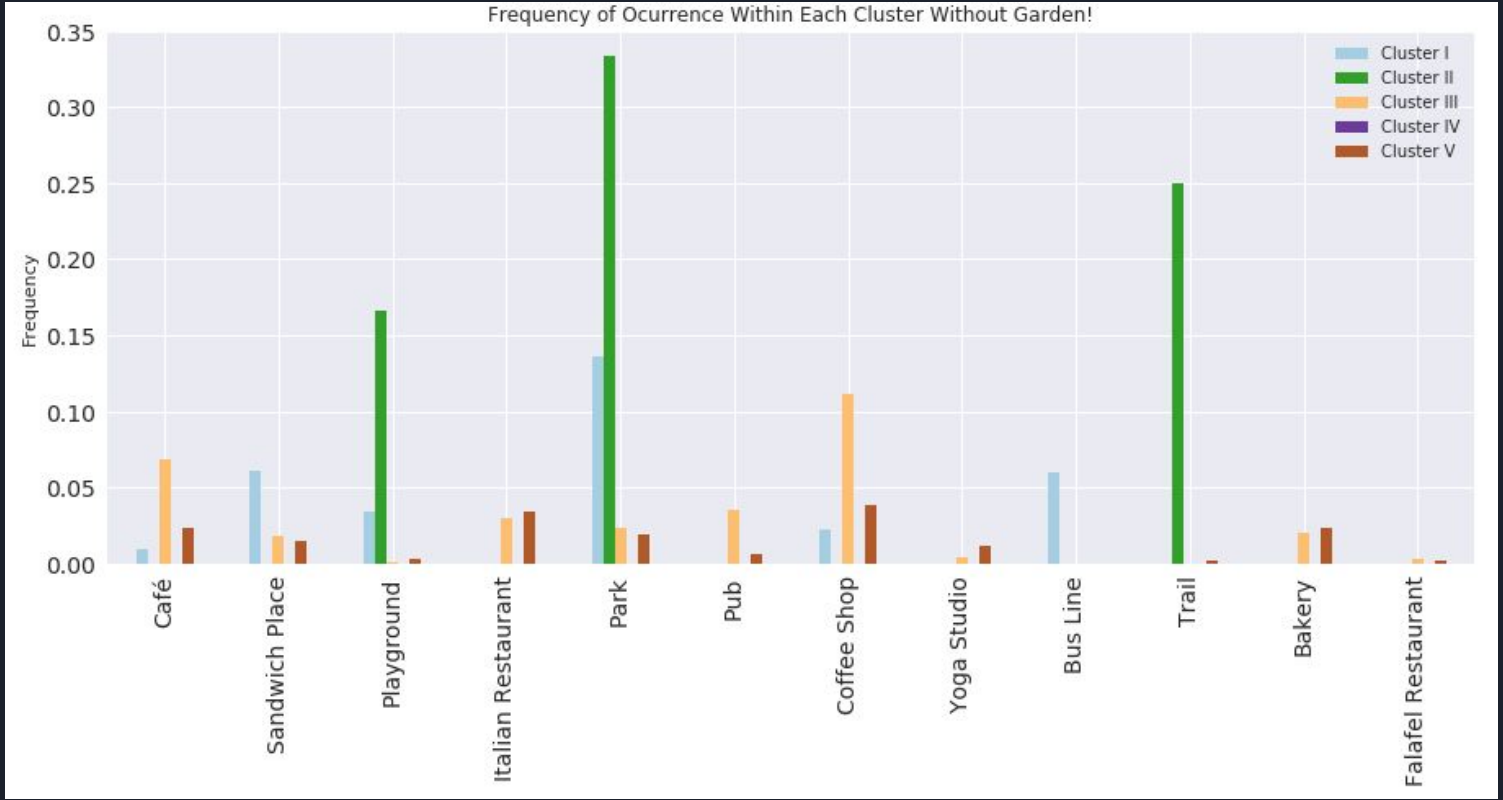
Neighborhoods Segmented by Colors



Bar Chart (Frequent Venues)



Bar Chart (Without Garden)



Conclusion

- I: Neighborhoods that have around parks, bus lines and sandwich places.
- II: Neighborhoods that have around parks, playgrounds and trails.
- III: Neighborhoods that have around coffee shops, pubs and italian restaurants.
- IV: Neighborhood that have around gardens.
- V: Neighborhoods that have around coffee shops, parks and bakeries.