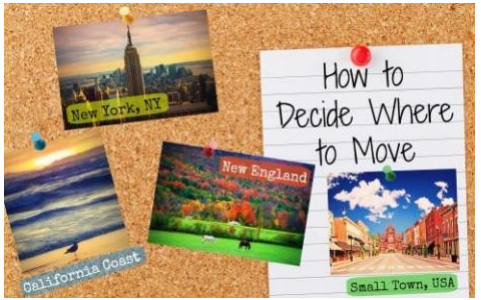
The Battle of Neighborhoods

Introduction

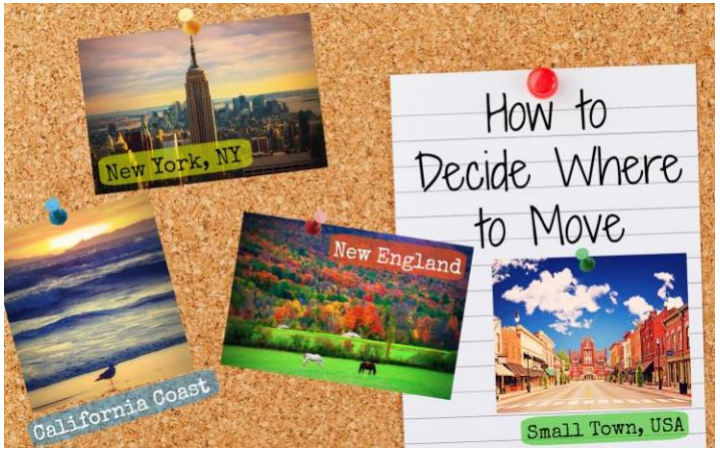
Alex is a software engineer who is based in New York city (US) and he is a user of Foursquare with a lot of posts and shares. Now, he wants to move to Toronto (Canada) and he really wants to find a neighborhood which is similar to his current place.

Is it possible to create a system that helps our users to find out the similarities between the two countries so that they can decide where to settle down?



Objective

Develop a system that can show the similarities in terms of neighborhoods in order to help a Foursquare user decide where to move near the center of Toronto.



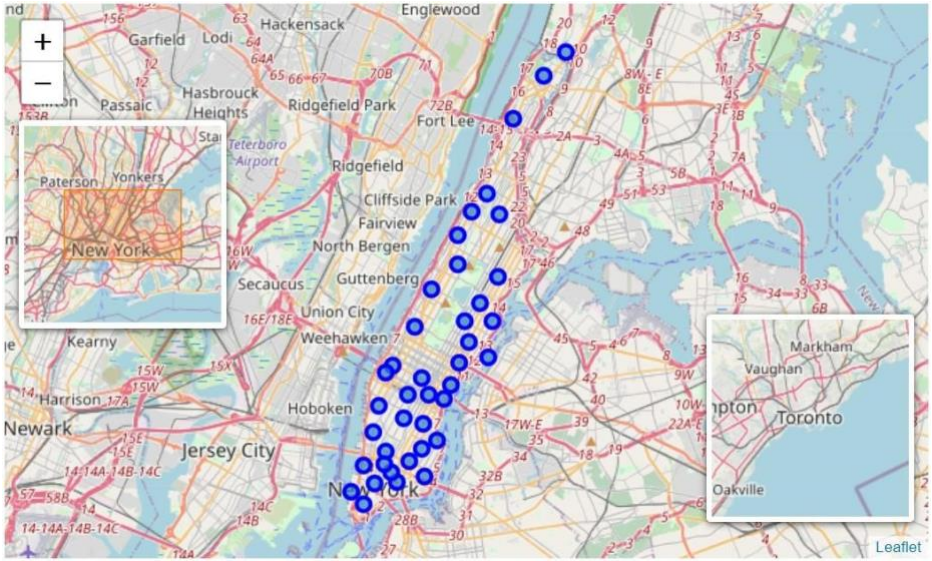
Approach

- 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

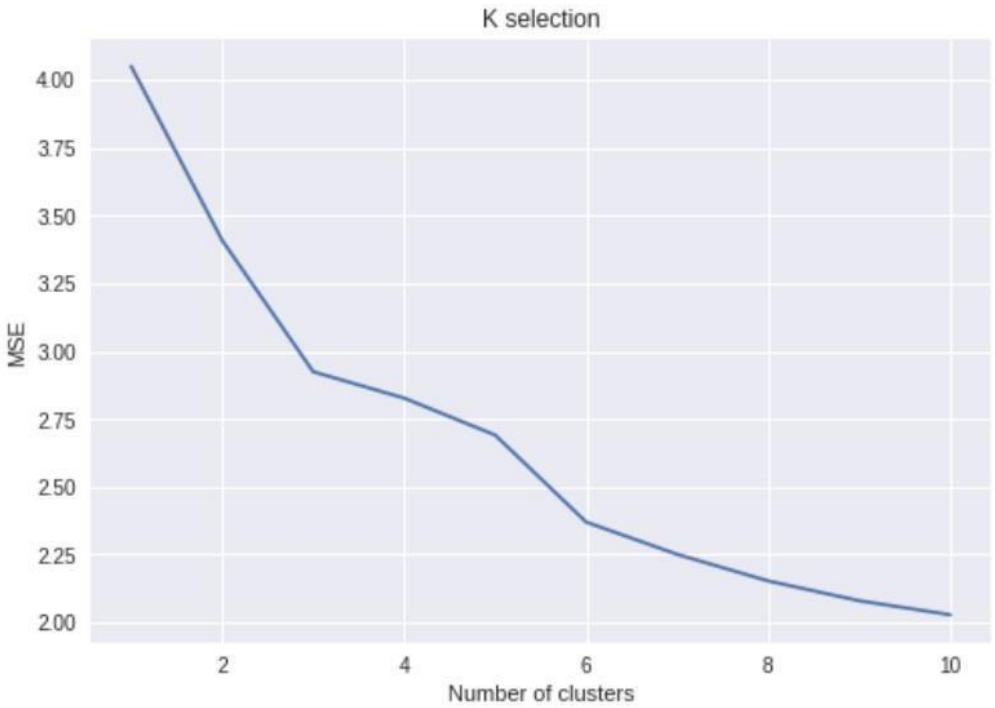
Geographic locations

New York neighborhoods



Toronto neighborhoods

Selection of K

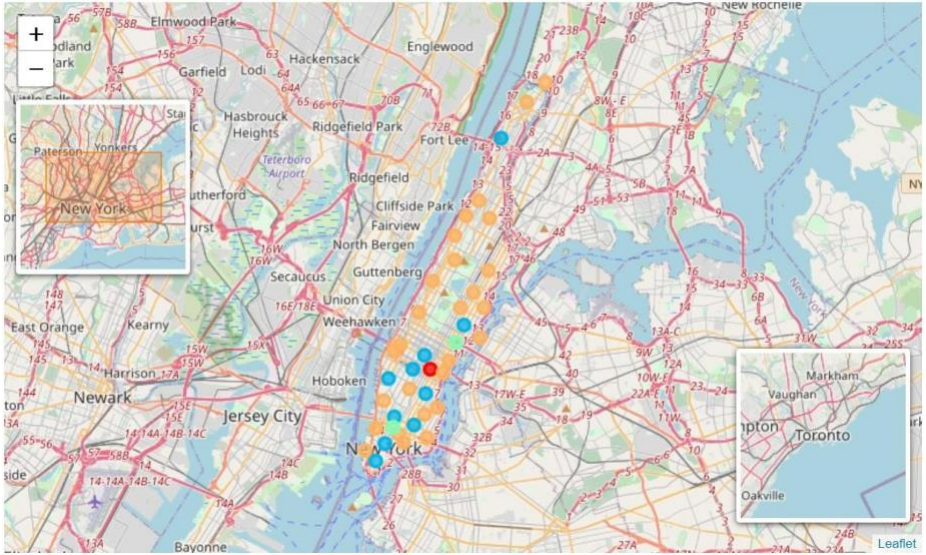


The best number of cluster is 5. That is, where the elbow is located. After that, the mean squared error decrease without big changes.

Geographical Locations (clusters)

New York

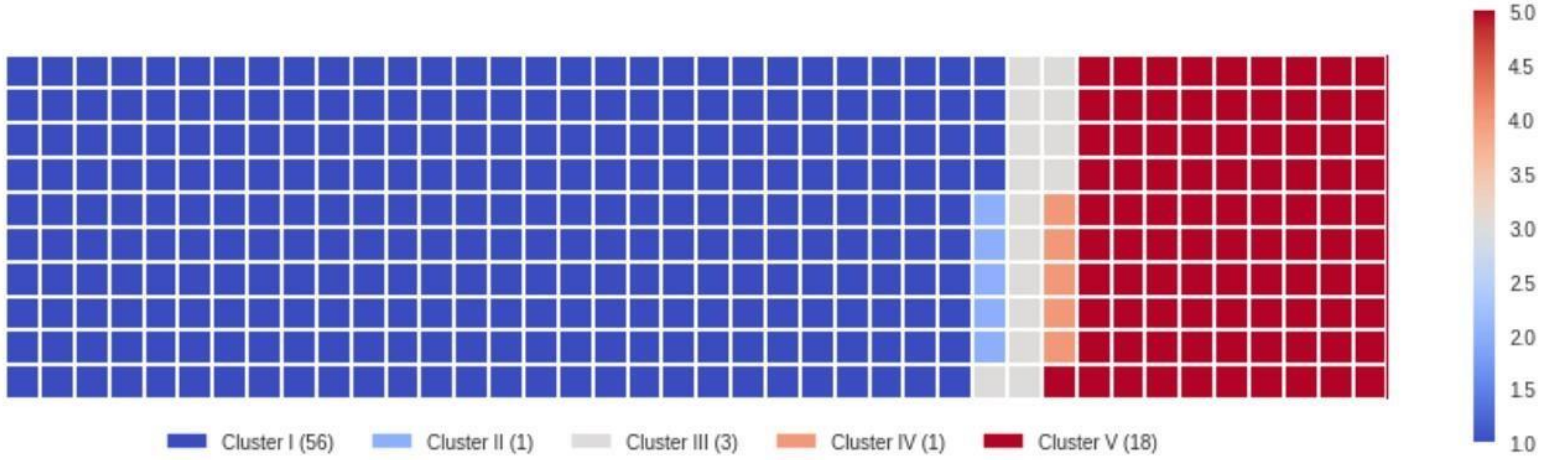
Toronto


Elmwood Park
Englewood
Garfield
Lodi
Hackensack
Hasbrouck Heights
Ridgefield Park
Fort Lee
Teterboro Airport
Ridgefield
Cliffside Park
Fairview
North Bergen
Guttenberg
Secaucus
Union City
Weehawken
East Orange
Kearny
Harrison
Hoboken
Newark
Jersey City
New York
Bayonne
Paterson
Yonkers
New York
Markham
Vaughan
Toronto
Oakville
Leaflet

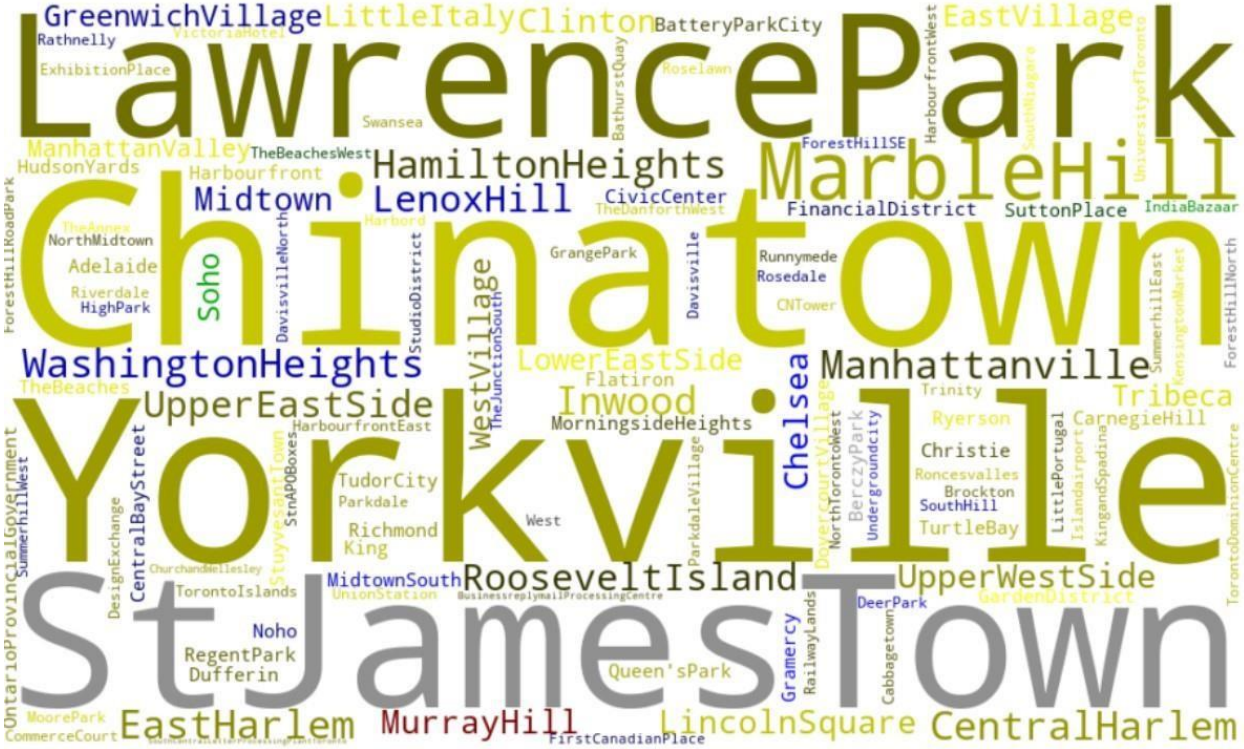

Downsview Airport
Allen
Toronto
Billy Bishop Toronto City Airport
Paterson
Yonkers
New York
Markham
Vaughan
Toronto
Oakville

Proportion of data segments

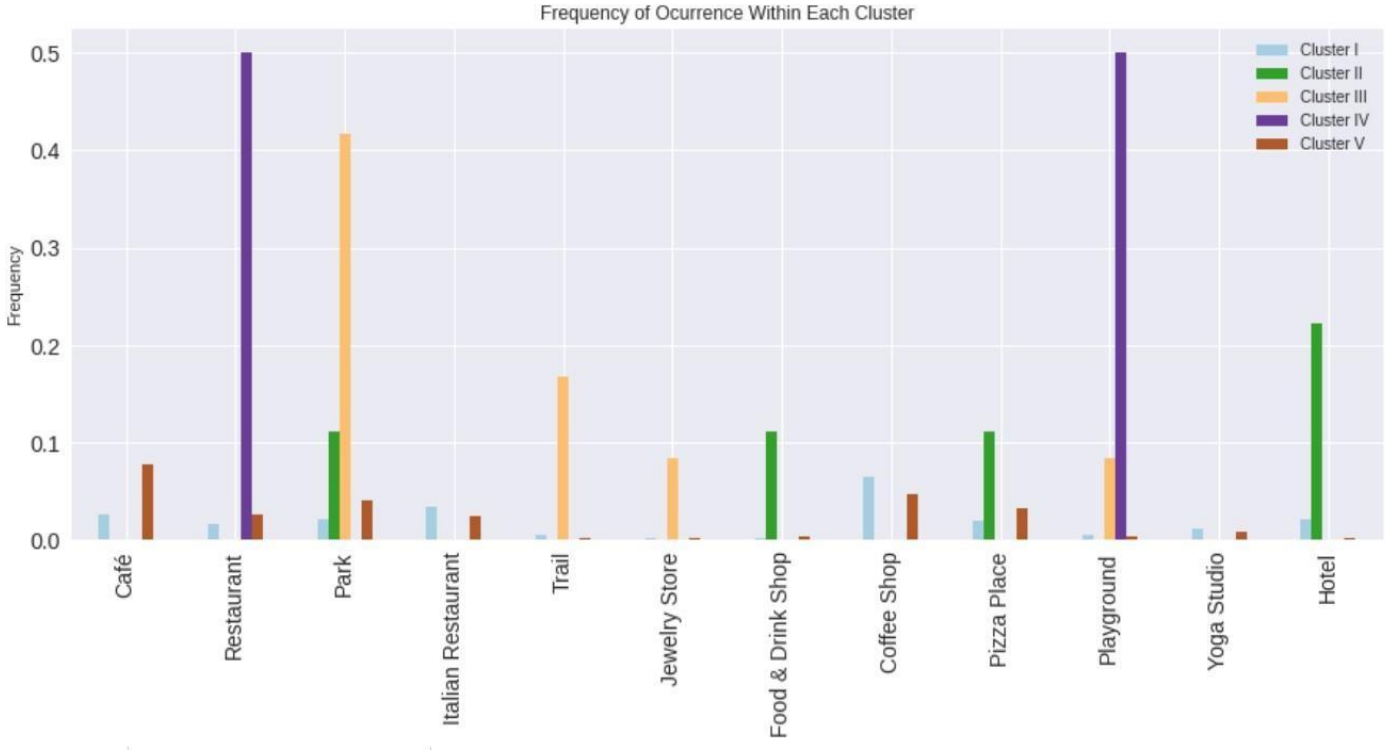
```
Total number of tiles is 400
Cluster I: 284
Cluster II: 5
Cluster III: 15
Cluster IV: 5
Cluster V: 91
<Figure size 576x396 with 0 Axes>
```



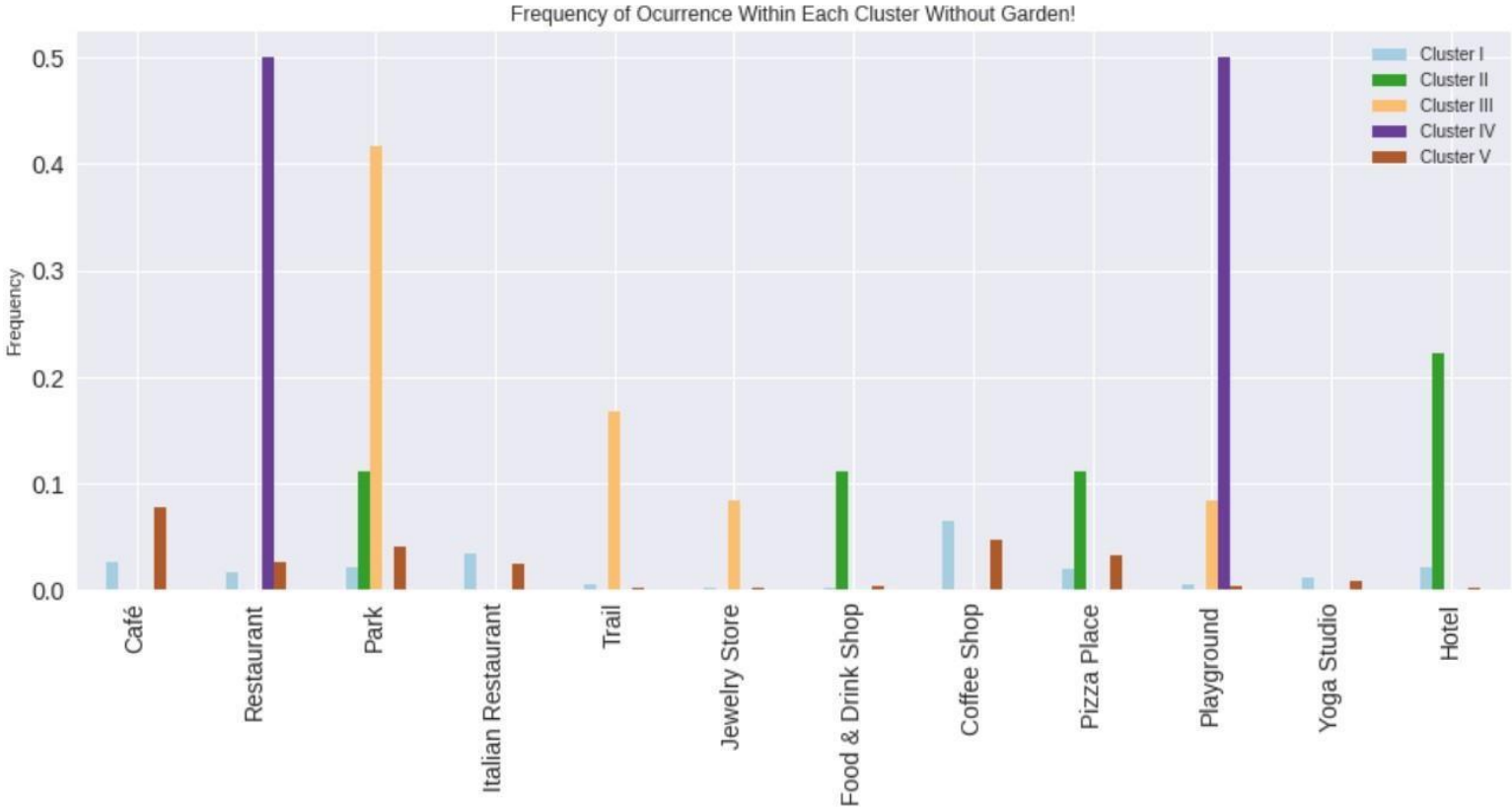
Neighborhoods segment (words cloud)



Bar chart (frequently visited venues)



Bar chart (without garden)



Conclusion

- Cluster I: Neighborhoods that have around coffee shops, Itatian restaurants and parks.
- Cluster II: Neighborhoods that have around hotels, food & drink shops and parks.
- Cluster III: Neighborhoods that have around coffee parks, trails and jewelry stores.
- Cluster IV: Neighborhood that have around restaurants and playgrounds.
- Cluster V: Neighborhoods that have around café and parks.