



Clustering Similar Neighborhoods in New York City and Toronto



Importance of finding similar neighborhoods

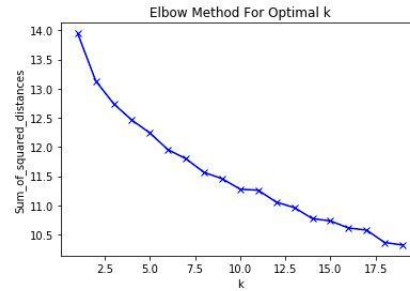
- Businesses want to determine similar locations to existing stores to build at in different cities
- Certain types of locations can lead to higher profits for a business
- Someone may want to move somewhere similar to where they live now



Data Acquisition and Cleaning

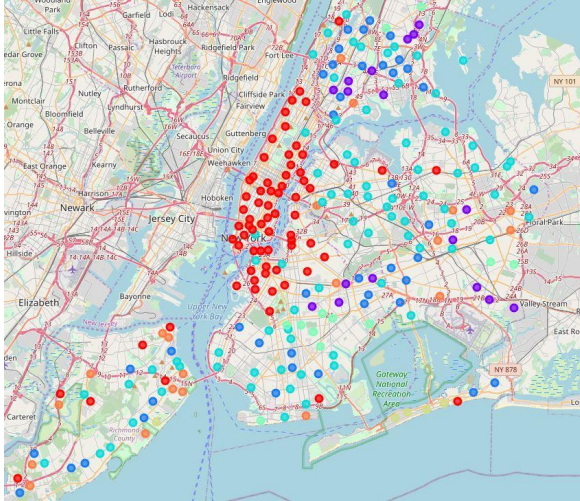
- Local venue data from Foursquare API
- Neighborhood names and burrows scraped from wikipedia, locations from geopy
- Neighborhoods with low amounts of local venues (< 10) dropped
- 250 neighborhoods in NYC and 100 in Toronto in cleaned data

Using elbow method to find optimal k



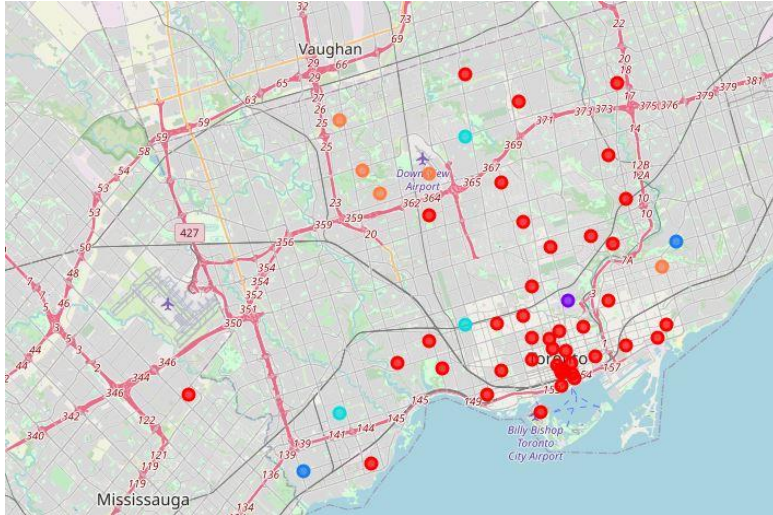
- No clear elbow
- More data or better data cleaning needed to cluster more accurately

NYC clustering



- Cluster 3 neighborhoods are almost all close together downtown
- Cluster 0 and 1 make up most neighborhoods that are not downtown
- Clusters aside from cluster 3 are spread out

Toronto clustering



- Most neighborhoods belong to cluster 3
- Like NYC most cluster 3 neighborhoods are close together downtown
- Much less diverse than NYC



Conclusion and Future directions

- Downtown areas of both cities are similar
- New York Cities is much more varied in its types of neighborhoods than Toronto
- Include income of neighborhoods, population growth, demographics, and other factors to improve clustering in the future