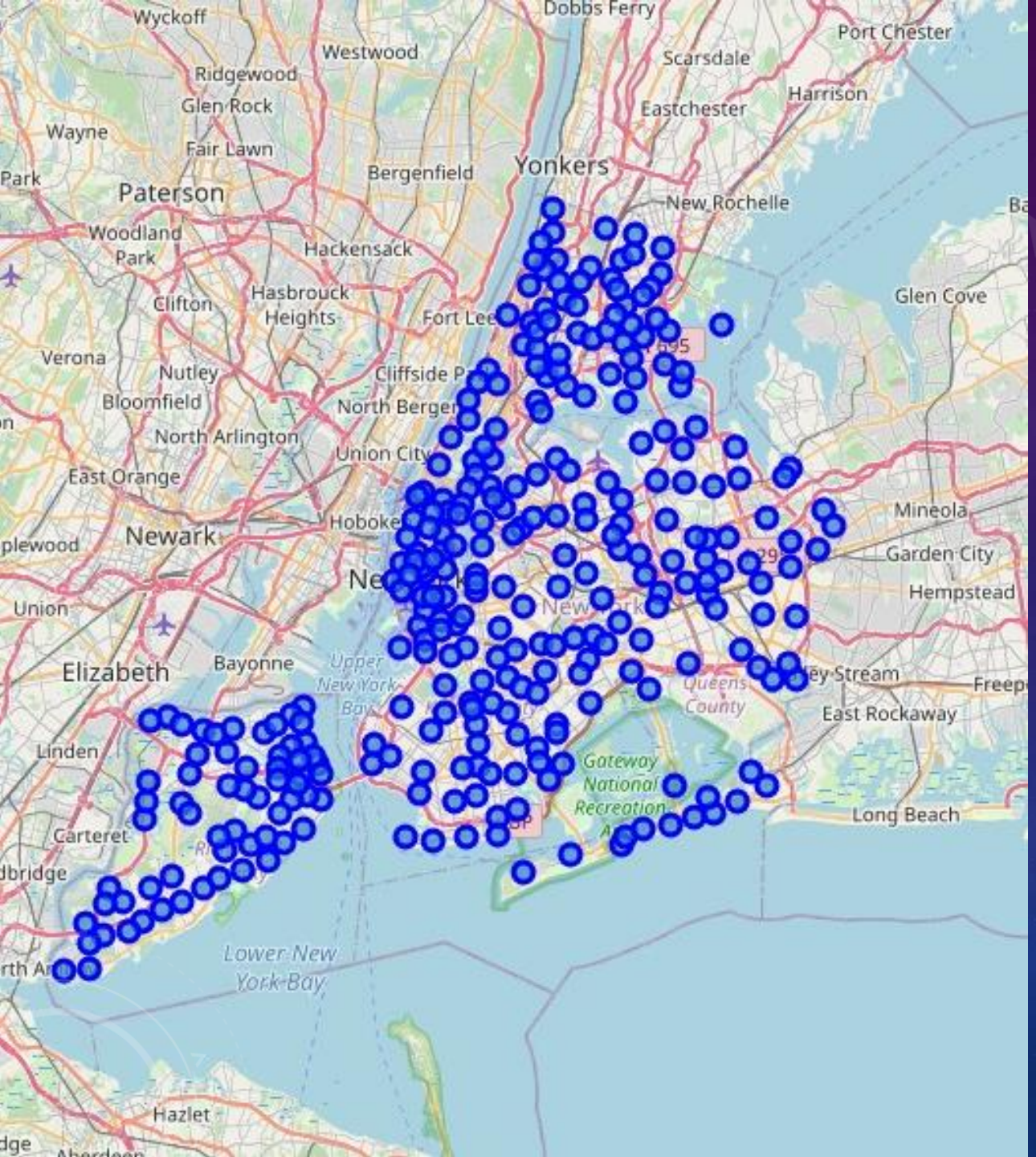
SIMILARITIES BETWEEN NEIGHBORHOODS IN NEW YORK AND TORONTO

DETERMINING NEIGHBORHOOD SIMILARITY

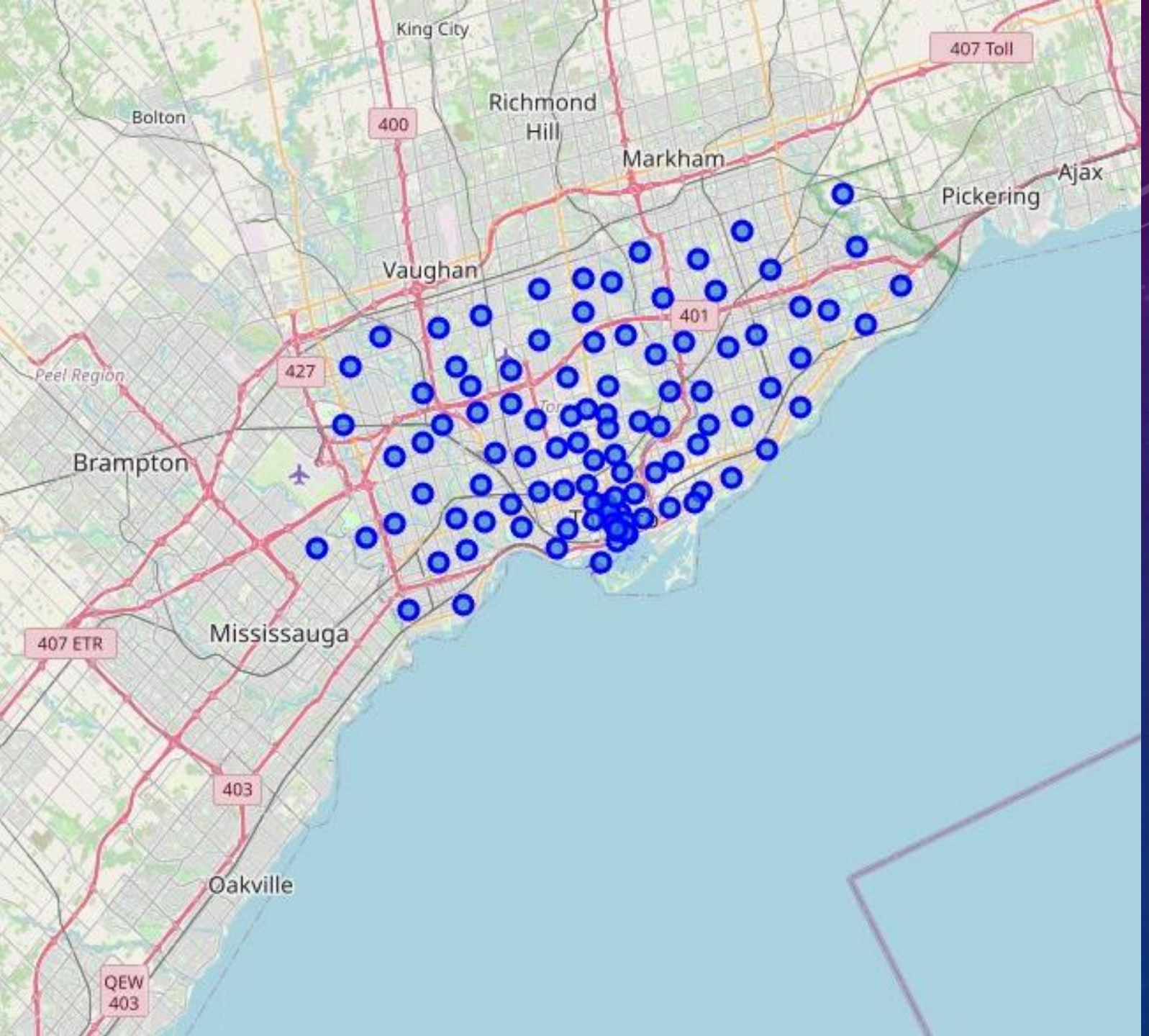
- A neighborhood is made up of various features such as:
  - Restaurants
  - Parks
  - Museums
  - Businesses
  - Other attractions
- A business that performs well in one neighborhood may not succeed in another
- Neighborhood's similarity can determine where to successfully expand

DATA

- A table of postal codes and neighborhoods in Toronto (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- A csv of the latitude and longitude of postal codes in Toronto (Geospatial_Coordinates.csv)
- A json file of New York neighborhood latitude and longitude data (newyork_data.json)
- To obtain the number of venues in each neighborhood (Foursquare API)
- Number of neighborhoods:
  - New York 300
  - Toronto 95
- Number of venues types
  - Both 461



NEW YORK NEIGHBORHOODS

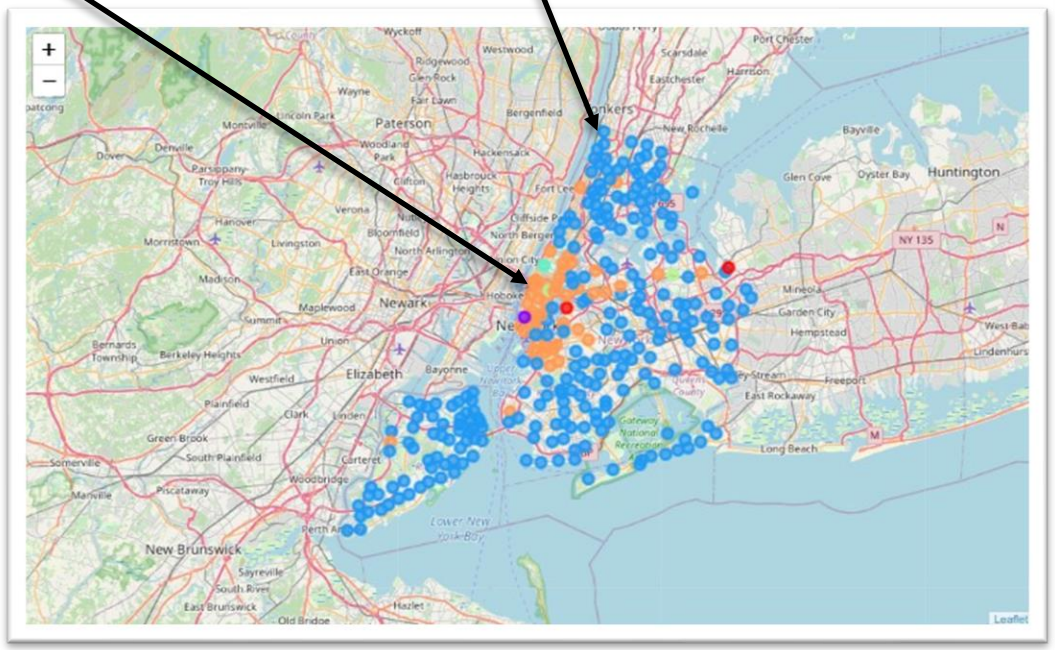


TORONTO NEIGHBORHOODS

Manhattan cluster 5

Outlying areas cluster 2

Outlying areas cluster 0



KMEANS CLUSTERING OF CITIES (K=6)

CONCLUSIONS

- Largest cluster in Toronto also contains Greenpoint and Little Neck from New York
- Largest New York cluster contains Commerce Court, Victoria Hotel, First Canadian Place, Underground city, Richmond, Adelaide, King, Toronto Dominion Centre, and Design Exchange from Toronto.
- Outlying areas in Toronto(red) and New York(blue) belong to the largest clusters and are most similar.
- New York is interesting as Manhattan appears to have its own cluster as well (shown in orange).
- Shows analysis can be used situationally for a business looking to move from New York to Toronto and can allow them to rule out most neighborhoods for expansion.

FUTURE WORK

- Additional variables may prove helpful for analysis:
  - crime rates
  - household income
  - property value
  - average education level
  - access to public transportation
- Other clustering methods may prove to be more useful:
  - hierarchical clustering
  - density based clustering
- Additionally cities should be considered in the analysis
  - interesting to see if other cities are more similar and follow the trend of outlying areas being predominantly of one cluster.