# Venue Similarity between Neighborhoods in New York, Toronto, and Detroit

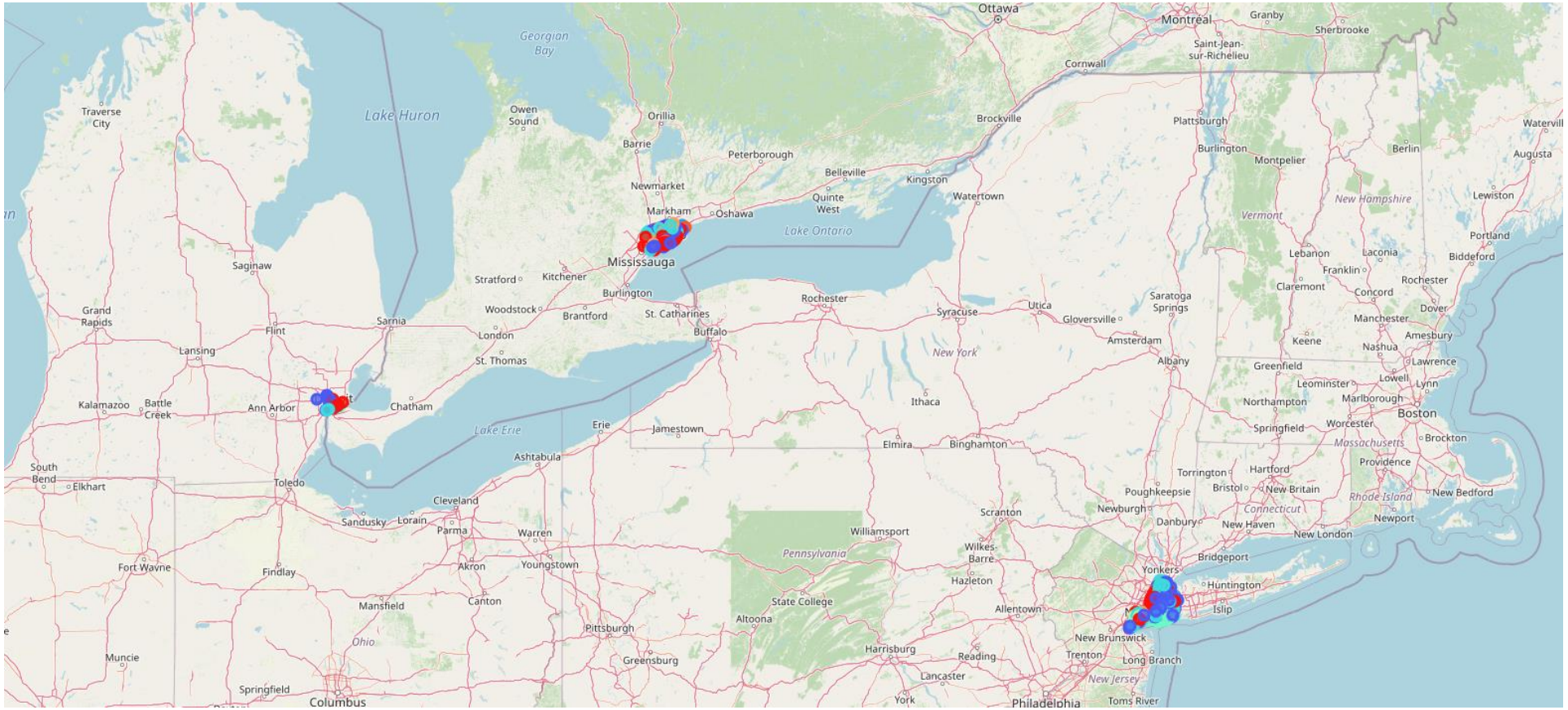
# Recommending cities to move to

- Many residents of large cities such as New York, Toronto or Detroit love the surrounding night life, cuisine, retail outlets and more
- Each city has its own personality that makes it unique
- Moving can be a stressful task
  - Unfamiliar surrounding venues can add to this stress
  - Familiar surroundings can be comforting
- Recommendations of cities to move to based on similarity of venue category may push movers in a direction that makes their move feel less stressful and more familiar

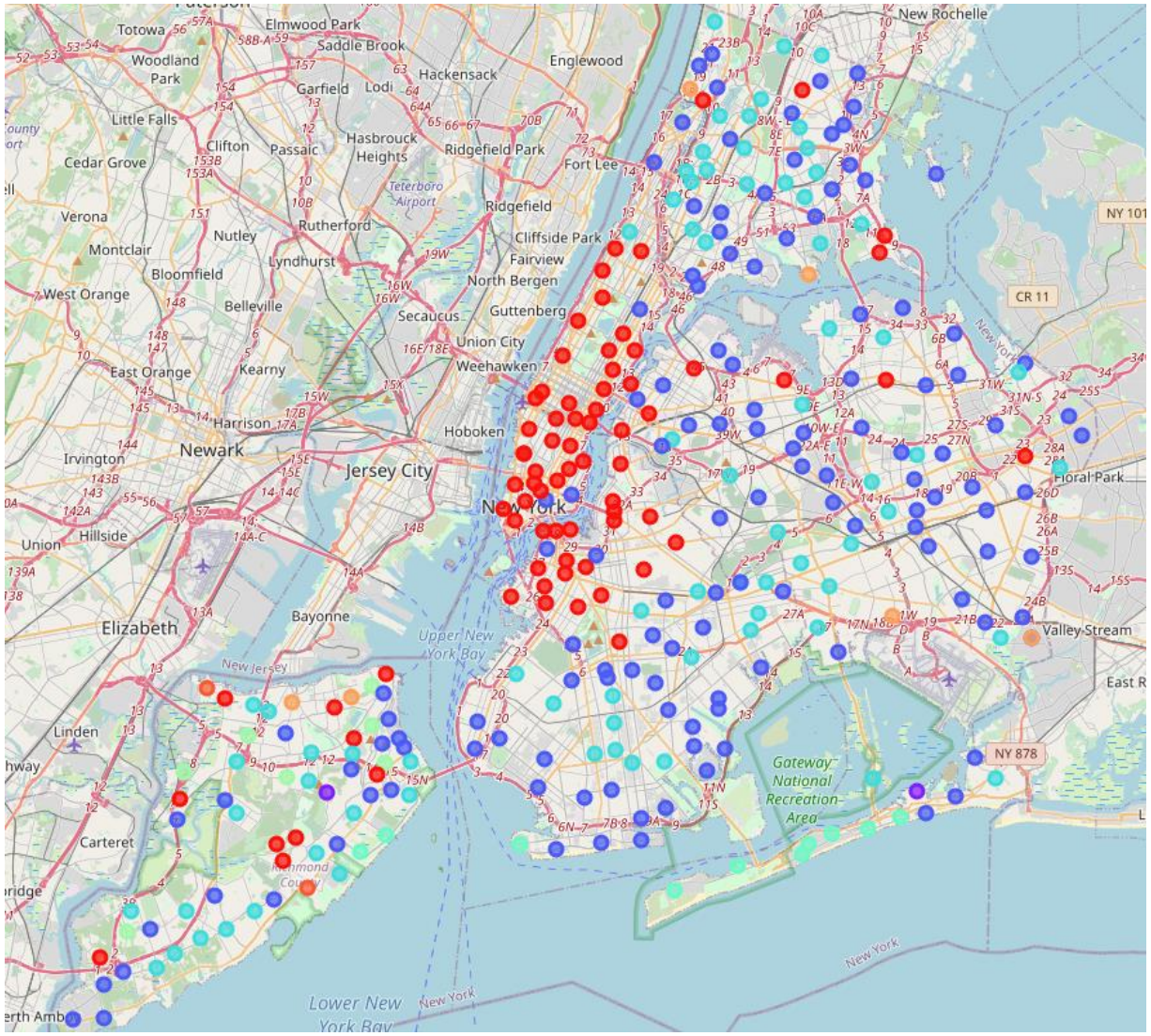
# Data acquisition, cleaning and combination

- City, borough, neighborhood, latitude and longitude data was gathered for each city through various means
  - New York data was gathered from a publicly available JSON file: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json
  - Toronto data was gathered by scraping the following Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. This data was combined with latitude and longitude data in a publicly available CSV file: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json
  - Detroit data was gathered by scraping the following Wikipedia page: https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Detroit
- This data was used to call Foursquare's Venues Explore API to gather data on venues in each neighborhood
- This data was then modified to contain averages of each venue category. Columns and rows containing no data were dropped. All cities were then combined into one dataframe to be used with kmeans clustering.

# Using kmeans to cluster neighborhoods

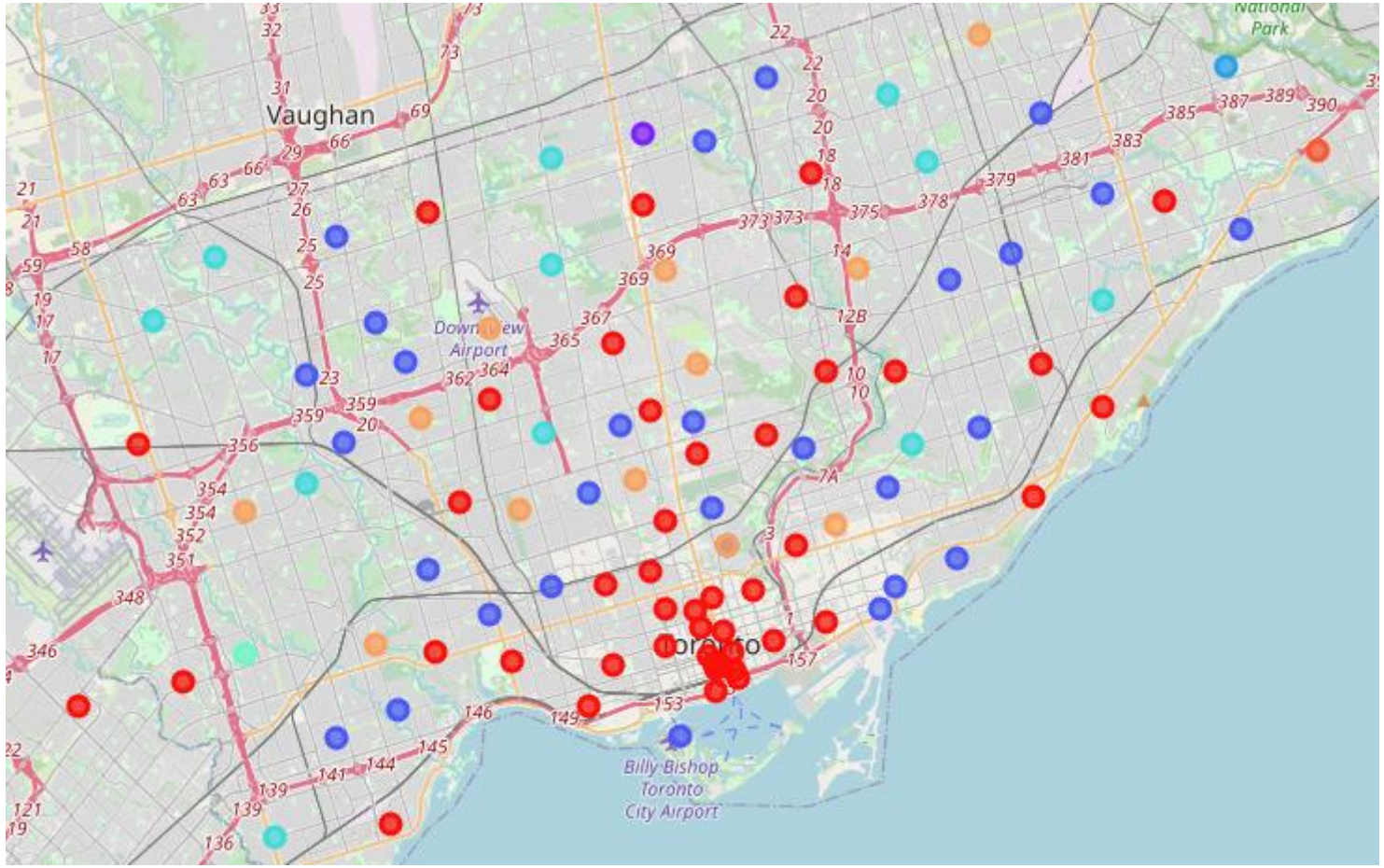


# New York clusters



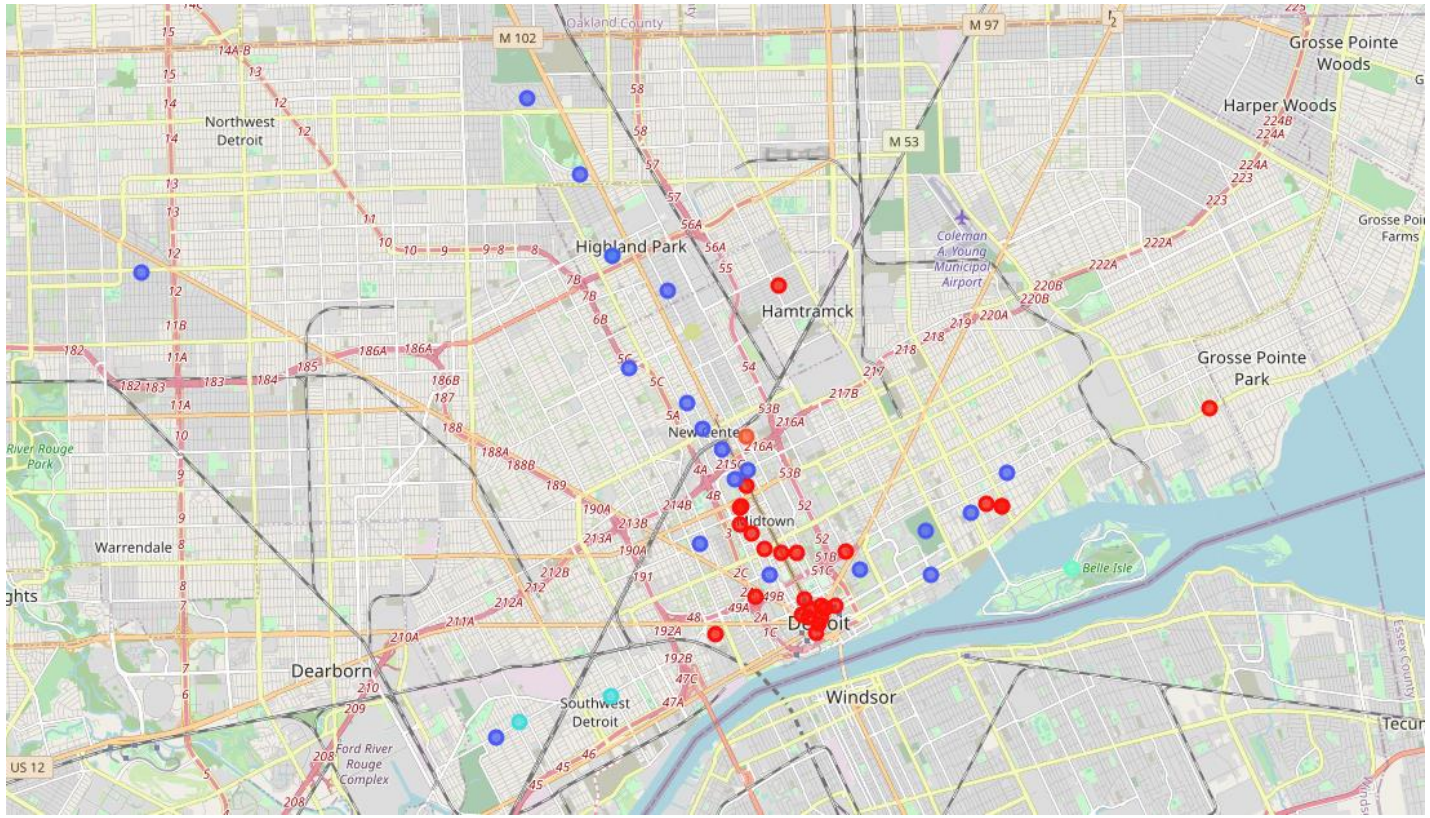

- Large groupings
  - Red - Cluster 0 – common venues include bars and coffee shops
  - Blue - Cluster 2 – common venues include parks/playgrounds, restaurants and transportation
  - Cyan – Cluster 4 – common venues include pharmacies and pizza places

# Toronto clusters




- Large groupings
  - Red - Cluster 0 – common venues include bars and coffee shops
  - Blue - Cluster 2 – common venues include parks/playgrounds, restaurants and transportation
  - Cyan – Cluster 4 – common venues include pharmacies and pizza places
  - Orange – Cluster 8 – common venues include parks, transportation and retail

# Detroit clusters




- Large groupings
  - Red - Cluster 0 – common venues include bars and coffee shops
  - Blue - Cluster 2 – common venues include parks/playgrounds, restaurants and transportation

# Conclusion

- New York and Toronto were found to have the most similar venue categories
- New York and Detroit were found to have the second most similar venue categories
- Toronto and Detroit were found to have the least similar venue categories
- Room for improvement:
  - Normalize number of neighborhoods used in each calculation
  - Include additional cities to expand options