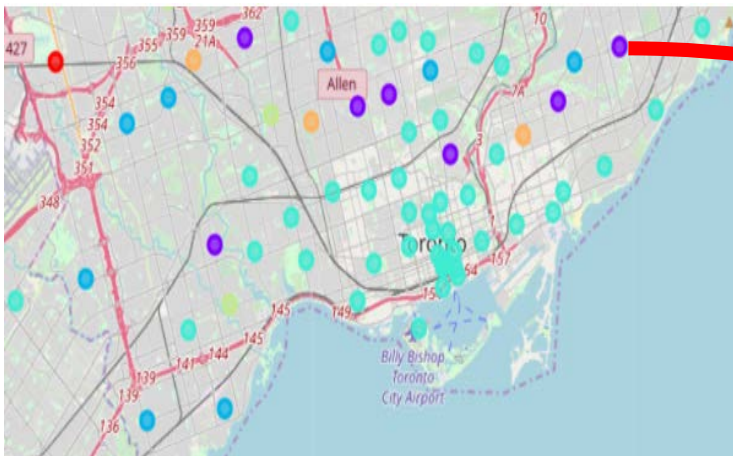# Where Should I Move?

A moving guidance app based on foursquare and clustering in python

Case study: Toronto to Queens/Staten Island

- This report uses a hypothetical move
- The hypothetical person is moving from Manhattan, having previously lived in Toronto
- The hypothetical person is considering moving to Staten Island or Queens
- The goal is to give the user a "from-to" table where the "from" is neighborhoods in Toronto or Manhattan, and the "to" is neighborhoods that are similar in Queens or Staten Island.

Data Acquisition and Cleaning

- Two parts: acquiring neighborhood/postal code data and associating venues
- Neighborhood/postal code data was scraped from Wikipedia articles. Then, each neighborhood was associated with an address using geocoder algorithms
- Once address were obtained, Foursquare was used to find 100 venues for each address
- Tables for each location (Toronto, Manhattan, Staten Island and Queens) were generated with the top 10 venue for each neighborhood
- Finally, a macro table was create with all location neighborhoods considered simultaneously (as if they were one location)

|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Bar	Steakhouse	Thai Restaurant	Restaurant	Burger Joint	Hotel	Sushi Restaurant	Asian Restaurant
1	Agincourt	Latin American Restaurant	Lounge	Skating Rink	Breakfast Spot	Women's Store	Dumpling Restaurant	Dog Run	Doner Restaurant	Donut Shop	Drugstore
2	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Bakery	Playground	Drugstore	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant

Describing each location: Toronto and Manhattan

- KMeans with 5 clusters
- Toronto cluster radiate from city center
- Manhattan dominated by quadrants
- 5 clusters seems sufficient
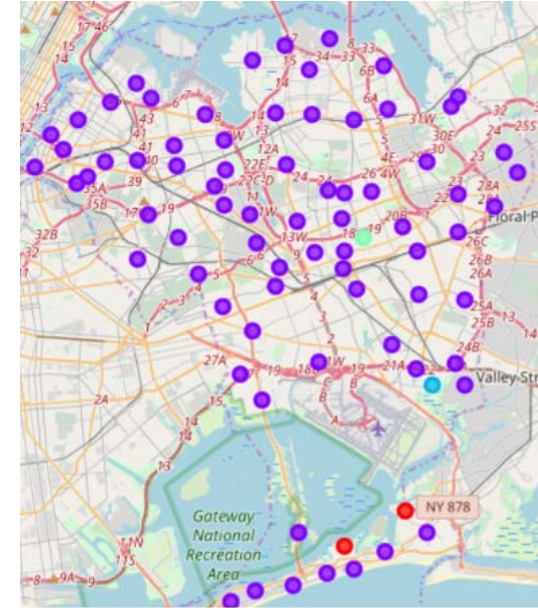

Clustered Toronto Neighborhoods
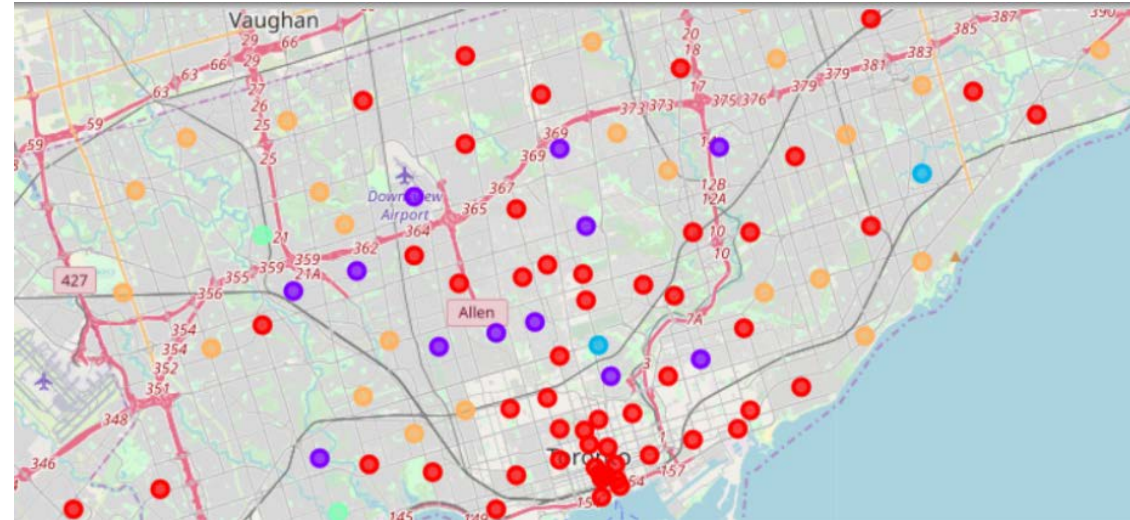
Clustered Manhattan Neighborhoods
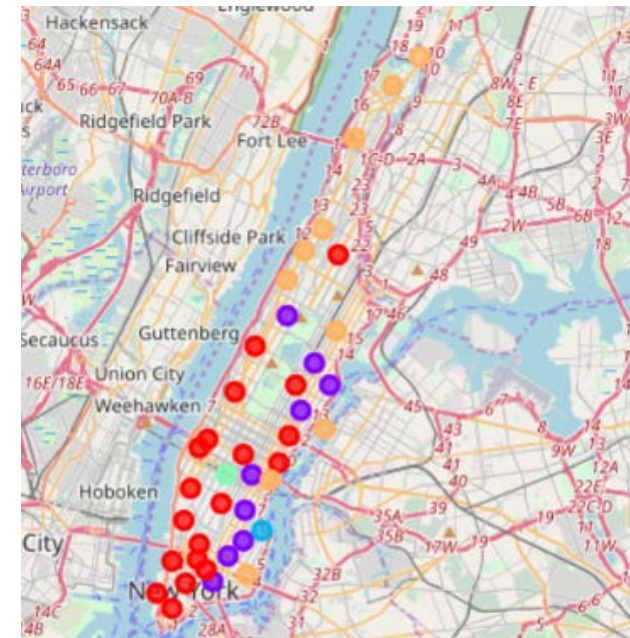
Describing each location: Queens and Staten Island

- KMeans with 5 clusters

- Queens is dominated with one cluster, with a few unique neighborhoods

- Staten Island has a large coastal cluster, a large internal cluster, and a few outlier neighborhoods

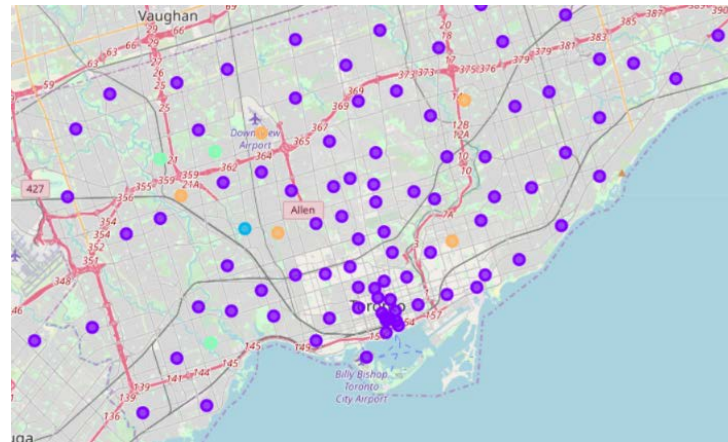- 5 clusters seems sufficient

Clustered Queens Neighborhoods

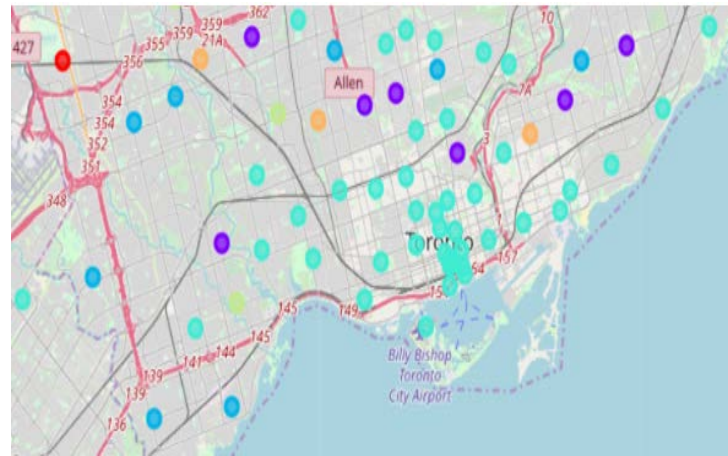
Clustered Staten Island Neighborhoods

Macro Dataset: parameter refinement

- Clustering the enter data set with 5 clusters is very sparse

- Clustering with 9 clusters reveals distinct regions that are consistent with the individually clustered on the previous slides.
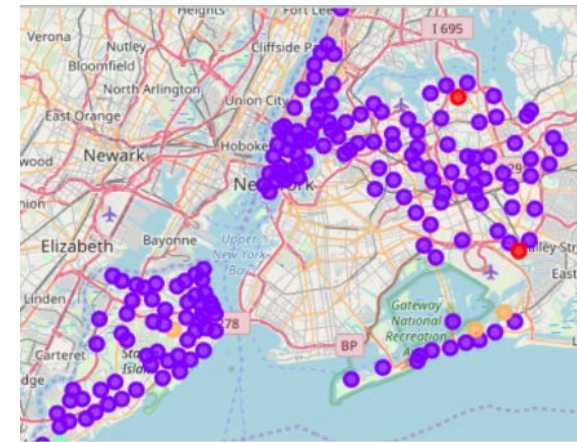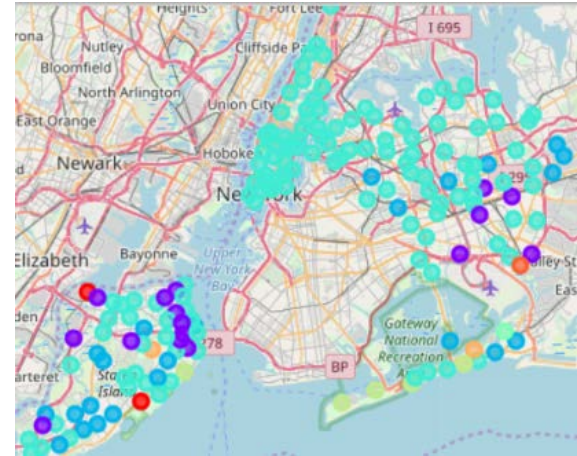
Toronto region, 5 clusters

New York region, 5 clusters

Toronto region, 9 clusters

New York region, 9 clusters

Final recommendation

- 3 clusters are most useful (represented in "from" as well as "to" locations

- The table below summarizes the movement guidance for this hypothetical example

Preferred cluster	Preferred Cluster Description	Recommended Move location
Cluster 4	Downtown Toronto and/or Manhattan	Northeast Queens or Northeast Staten Island
Cluster 3	Mid-town Toronto	Queens, either Laurelton Cluster or Bayswater neighborhoods
Cluster 1	Western Toronto	Southern Staten Island

Conclusion and future directions

- Hypothetical case successful: clear guidance for moving location was generated using the neighborhood/foursquare/clustering approach

- Case study is scalable, as data acquisition, cleaning and clustering can be automated for arbitrary "to" "from" locations

- Future work may include a human readable description of clusters. I had to manual describe the clusters in this report, but using compos directions and map hints may automate this process.

Thanks!