COMPARING NEIGHBORHOODS

THE PROBLEM

- It’s hard nowadays to find good and organized information on the venues around an specific place;
- When buying a house or starting a business, it’s important to be able to find neighborhoods similar to those in other places, or even to restrict the group of possible neighborhoods when deciding where to go.

Data acquisition and cleaning

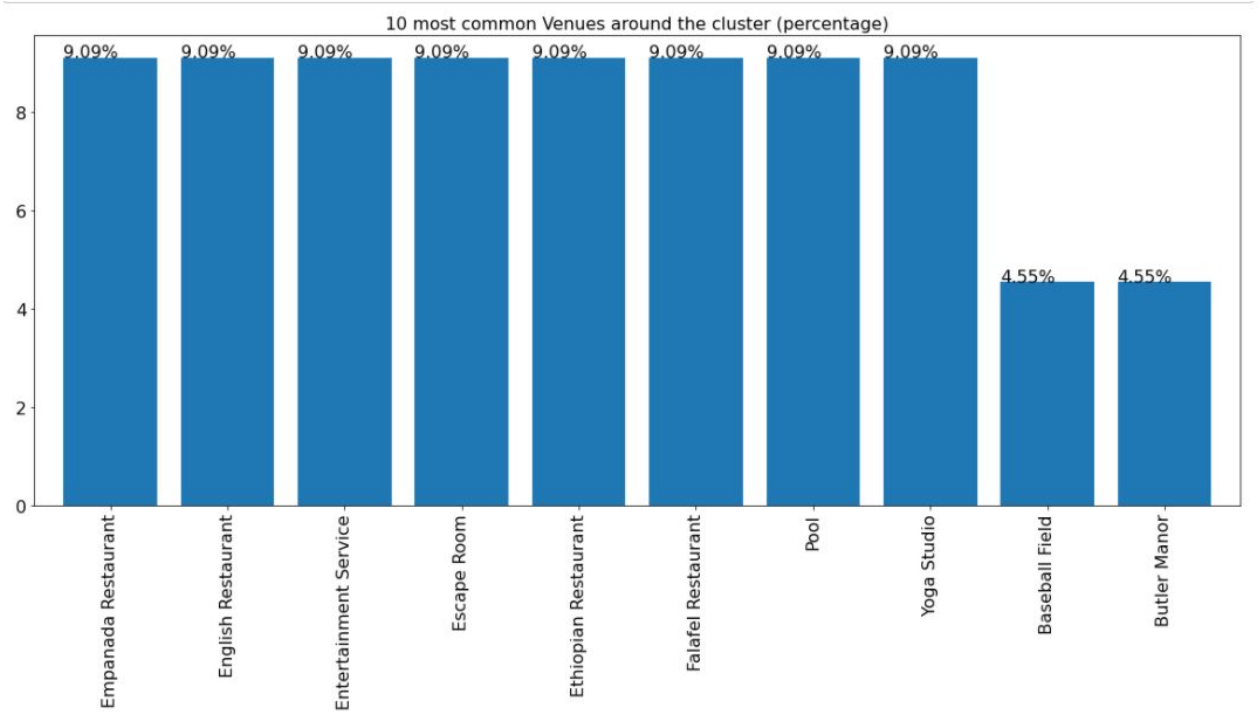
- Data from two different cities (Toronto and New York City), obtained from wikipedia, cocl.us and documents provided by the course;
- Data on the venues obtained from FourSquare API;
- Organization of data into tables and preparation for clustering (by putting the datasets together and generating features based on the presence of venue categories)

Methods

- Organization of data and preparation of functions to access FourSquare API and to take the needed features;
- Definition of methods (Elbow method) to find ideal number of clusters;
- Clusterization and description of the clusters using bar charts;
- Plotting of choropleth map showing the neighborhoods and the clusters each one belongs to.

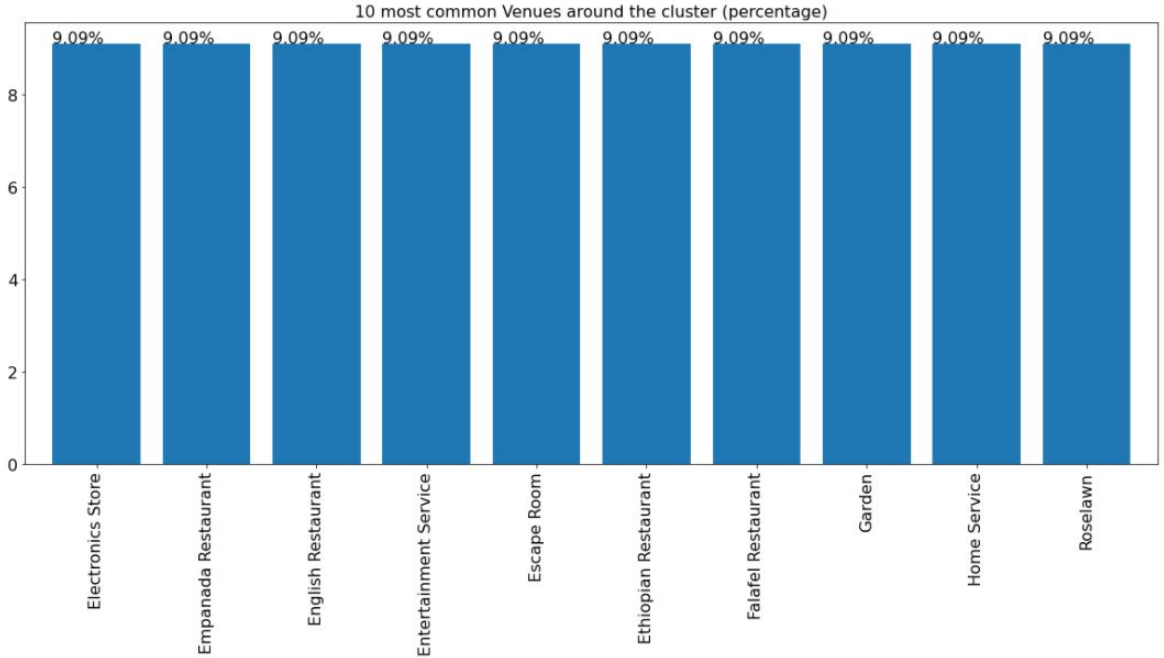
Bar charts, choropleth map and interpretation

- Cluster 1



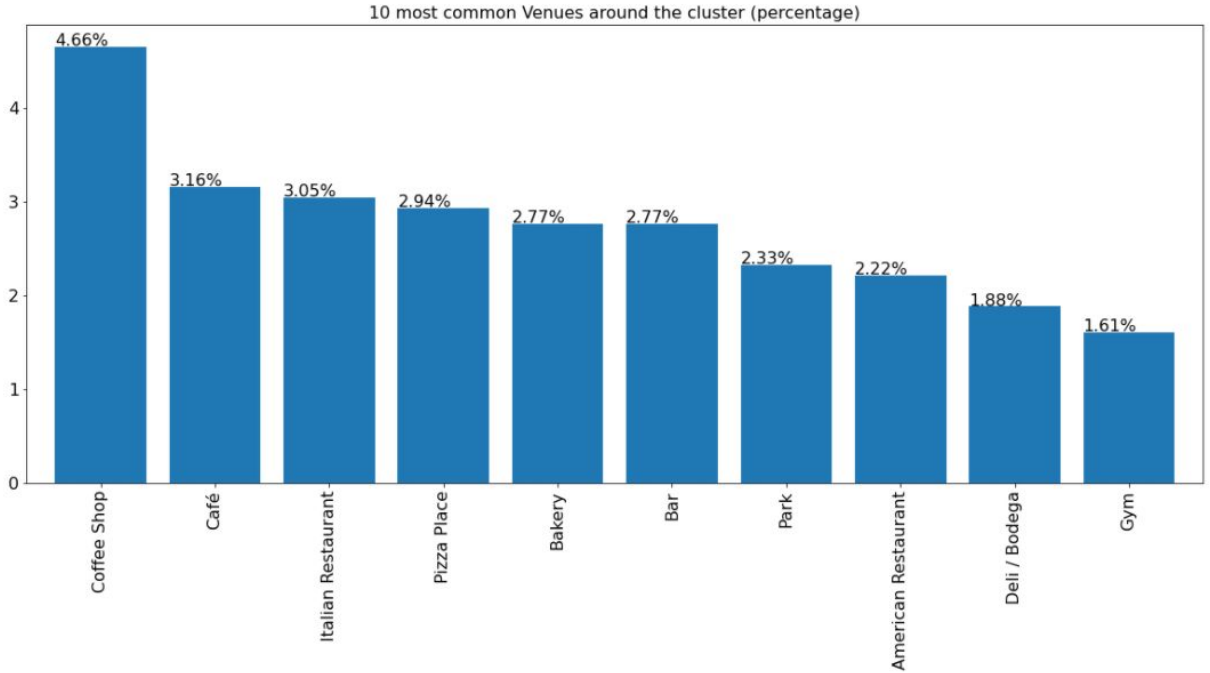
Bar charts, choropleth map and interpretation

- Cluster 2



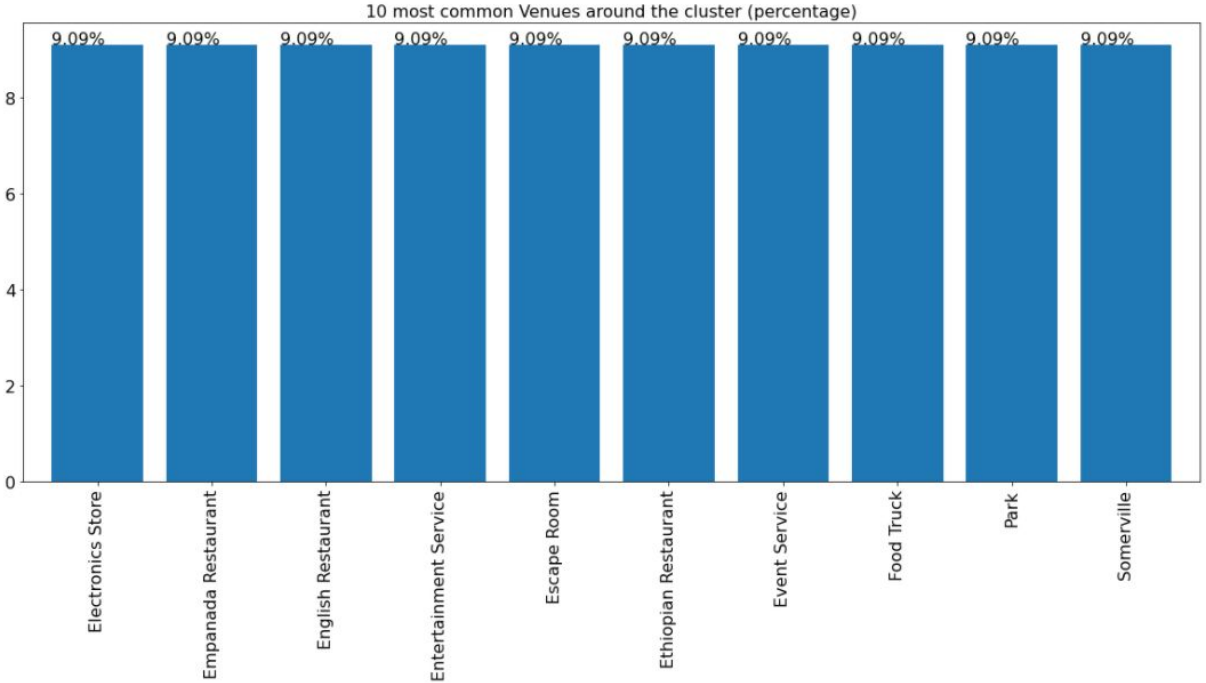
Bar charts, choropleth map and interpretation

- Cluster 3



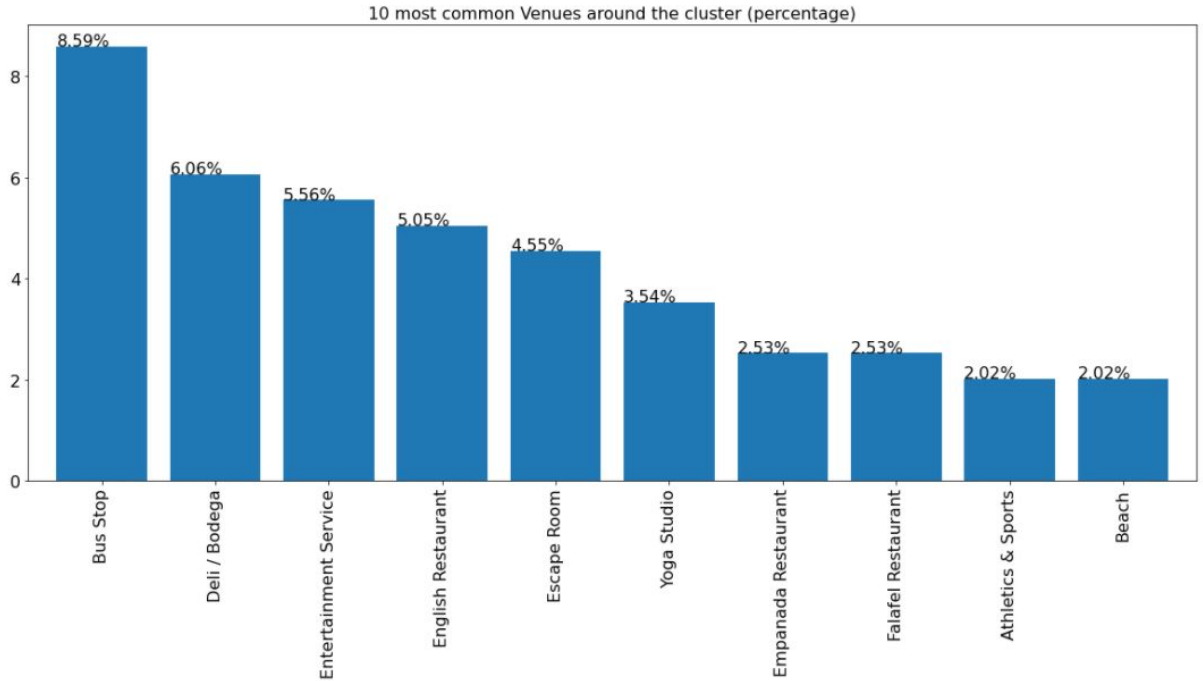
Bar charts, choropleth map and interpretation

- Cluster 4



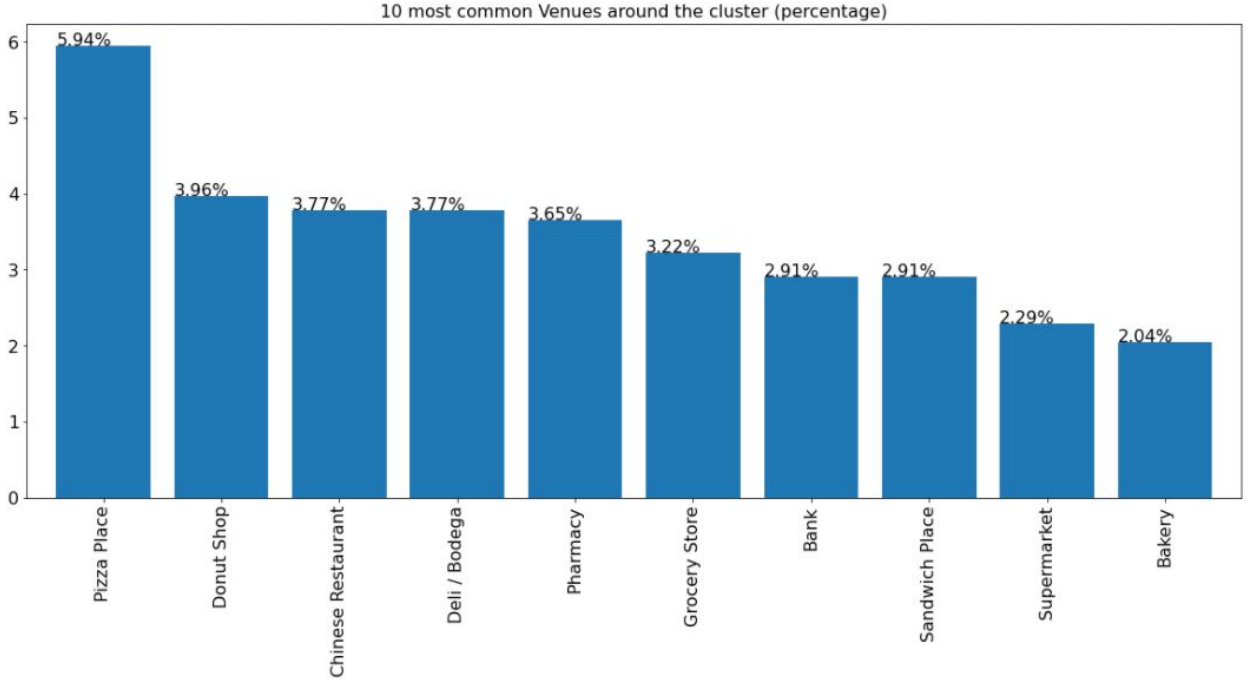
Bar charts, choropleth map and interpretation

- Cluster 5



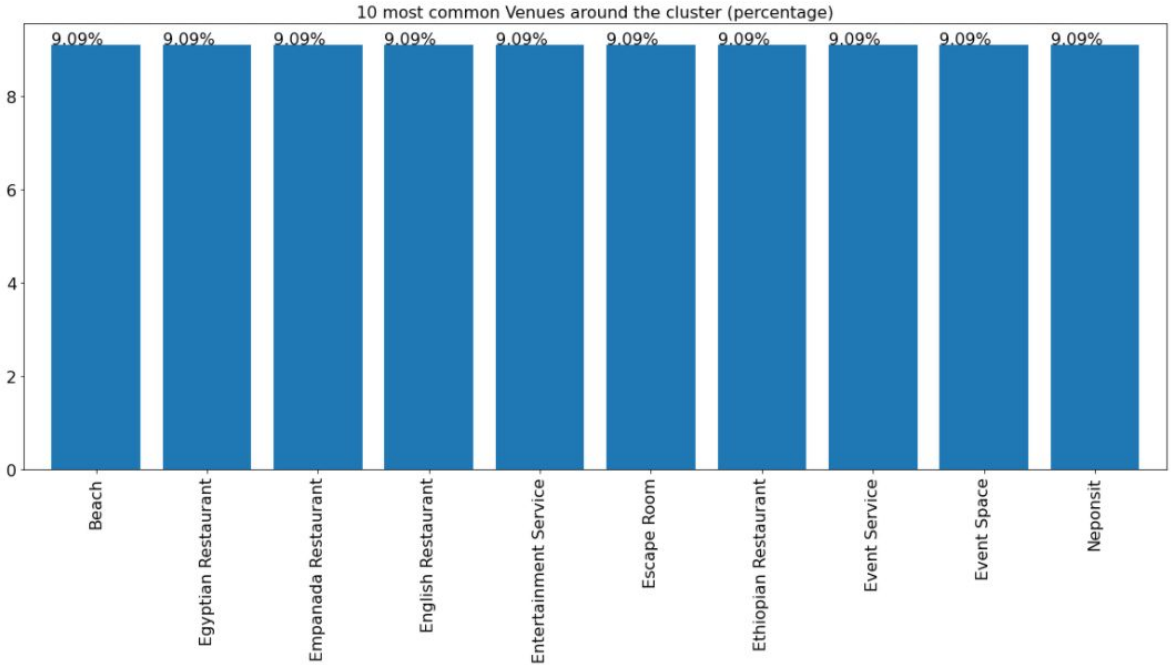
Bar charts, choropleth map and interpretation

- Cluster 6



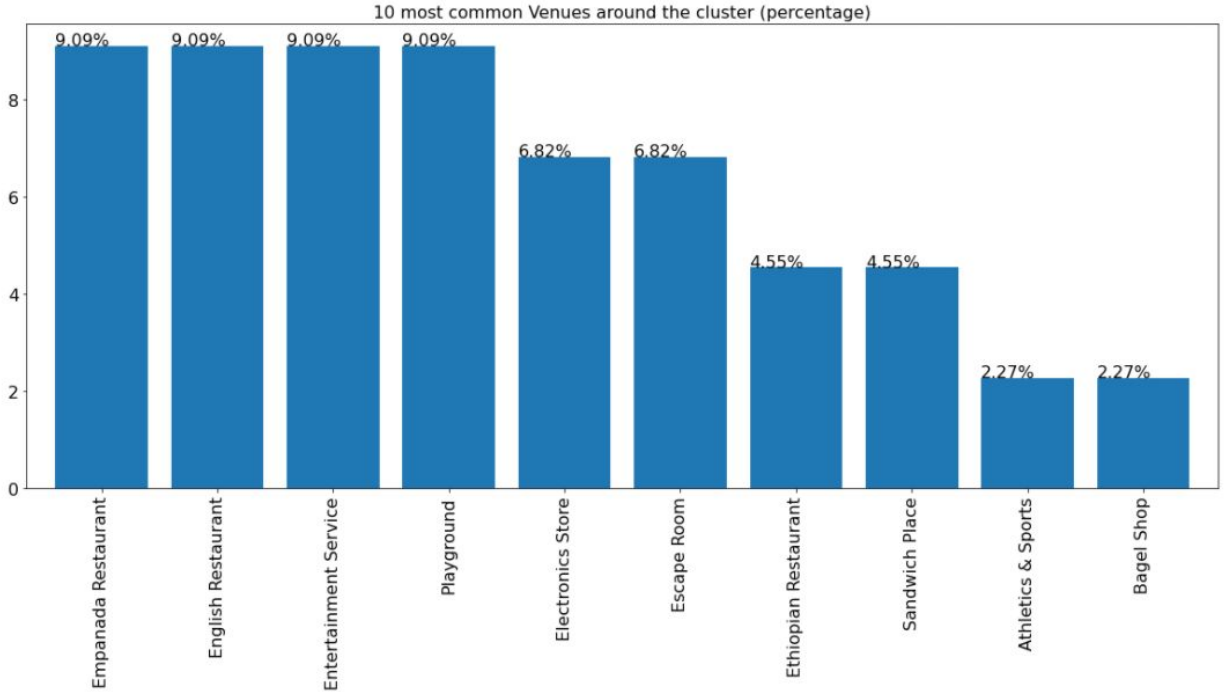
Bar charts, choropleth map and interpretation

- Cluster 7



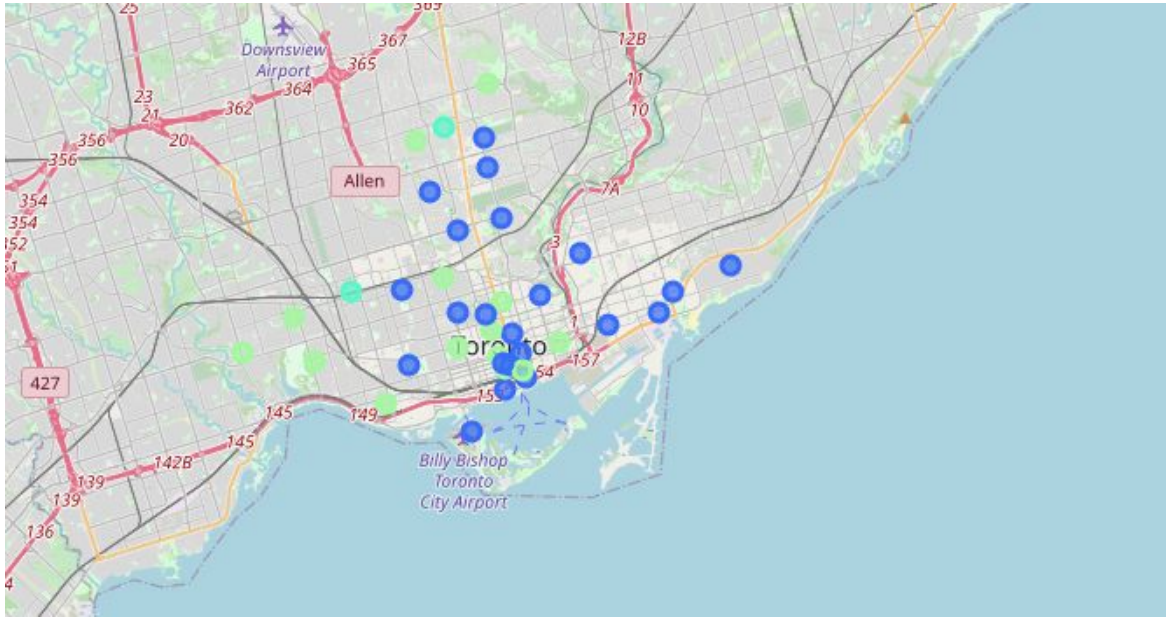
Bar charts, choropleth map and interpretation

- Cluster 8



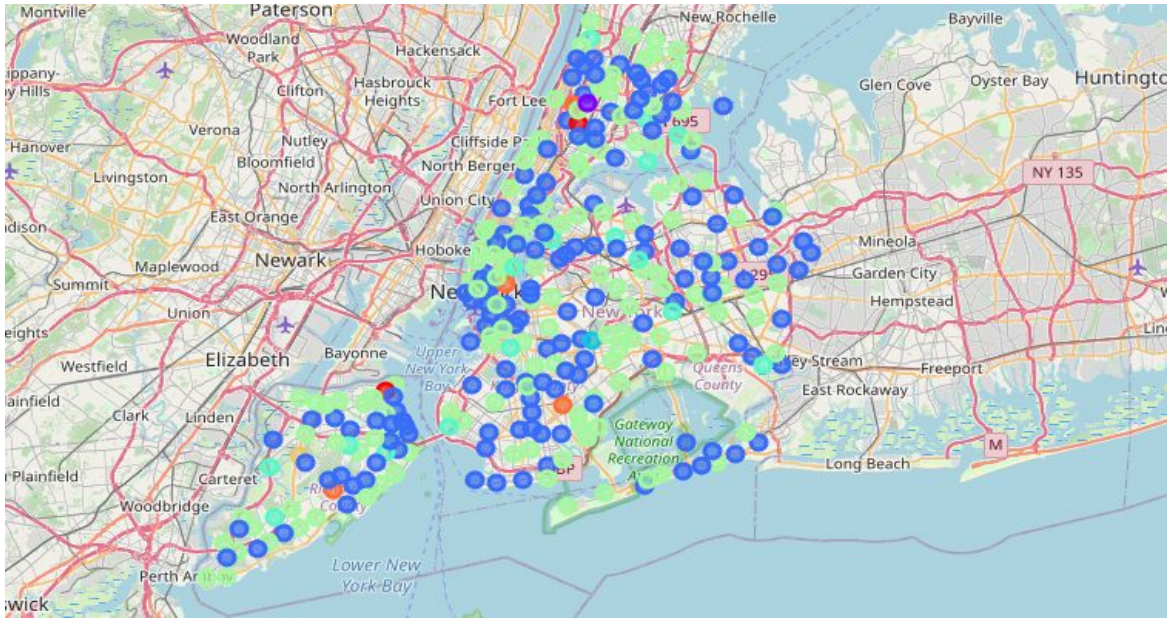
Bar charts, choropleth map and interpretation

- Toronto choropleth map



Bar charts, choropleth map and interpretation

- New York City choropleth map



Conclusion and next steps

- Project achieves its main goals, since it makes really easier for the client to choose a place or define their priorities better;
- Charts bring a lot of information about the neighborhoods in each cluster and makes it possible to restrict the search for places to a single cluster;
- Information on cities similarities/differences is also available through deeper analysis of the bar charts;
- New functions/improvements may be implemented.