


The Battle of Neighborhoods

Capstone project

Elisabeth Duijnstee





PROBLEM

- ❑ **Huge plastic waste pile** that grows every day, and primarily consists of **single-use plastic**.
- ❑ Main contributor is **food-sector** due to **grocery packaging**.





Solutions

- ☐ Circular grocery shopping with recyclable packaging in grocery stores
- ☐ Opening venture in New York, with Manhattan, Brooklyn and Queens as urban core.
- ☐ Population seeks high-quality food for home-cooking





Data Analytics

- ☐ Provide analytical view to support decision making for optimal supermarket location.
- ☐ Considering viable options based on specific data and criteria.





Data Section

- ❑ **New York City:** 5 boroughs and 306 neighborhoods.

Link to the dataset is:

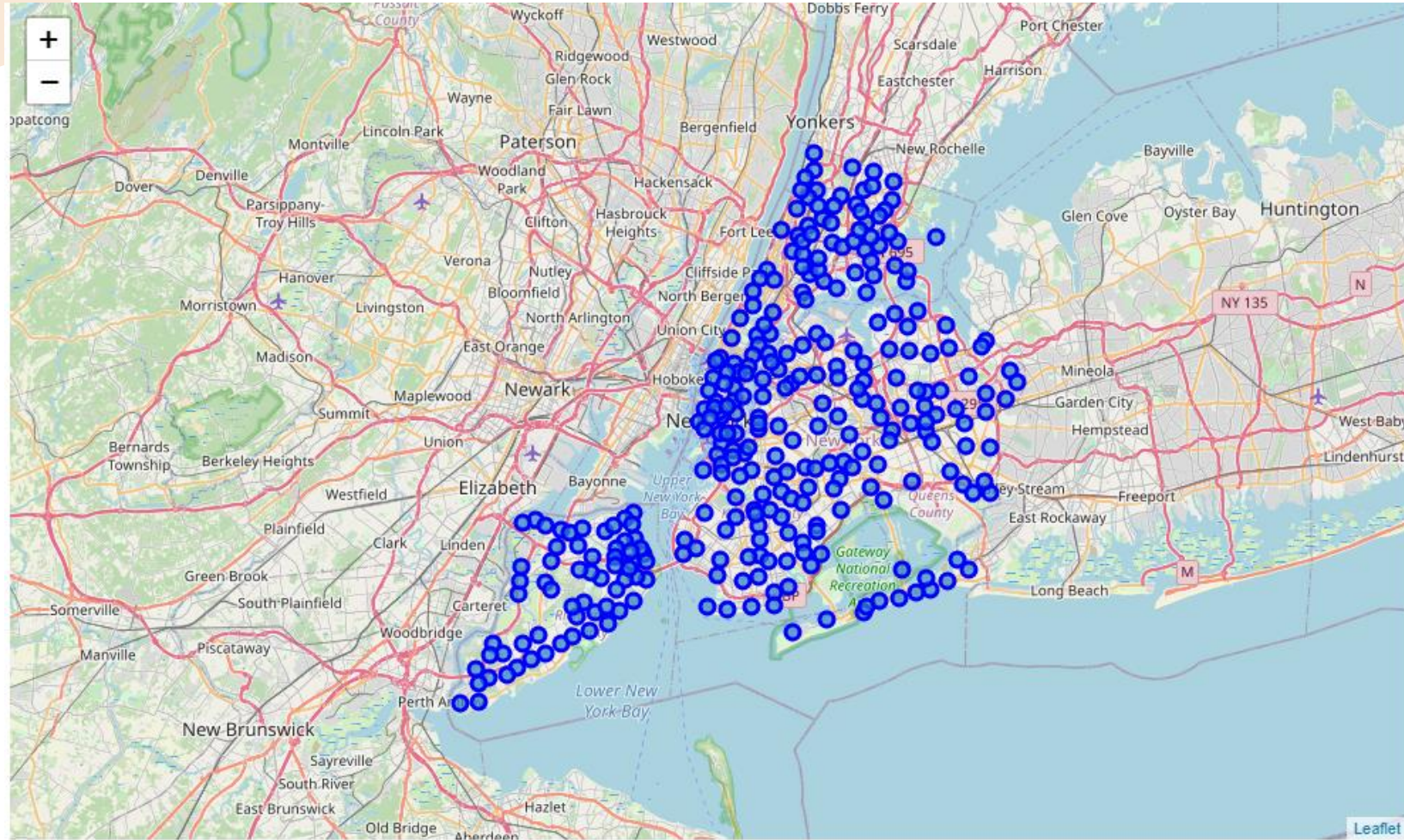
https://geo.nyu.edu/catalog/nyu_2451_34572

- ❑ **Food store related venues:** Foursquare API, allows querying places and users. New York City geographical coordinates data will be used as input to leverage information on venues for each neighborhood.

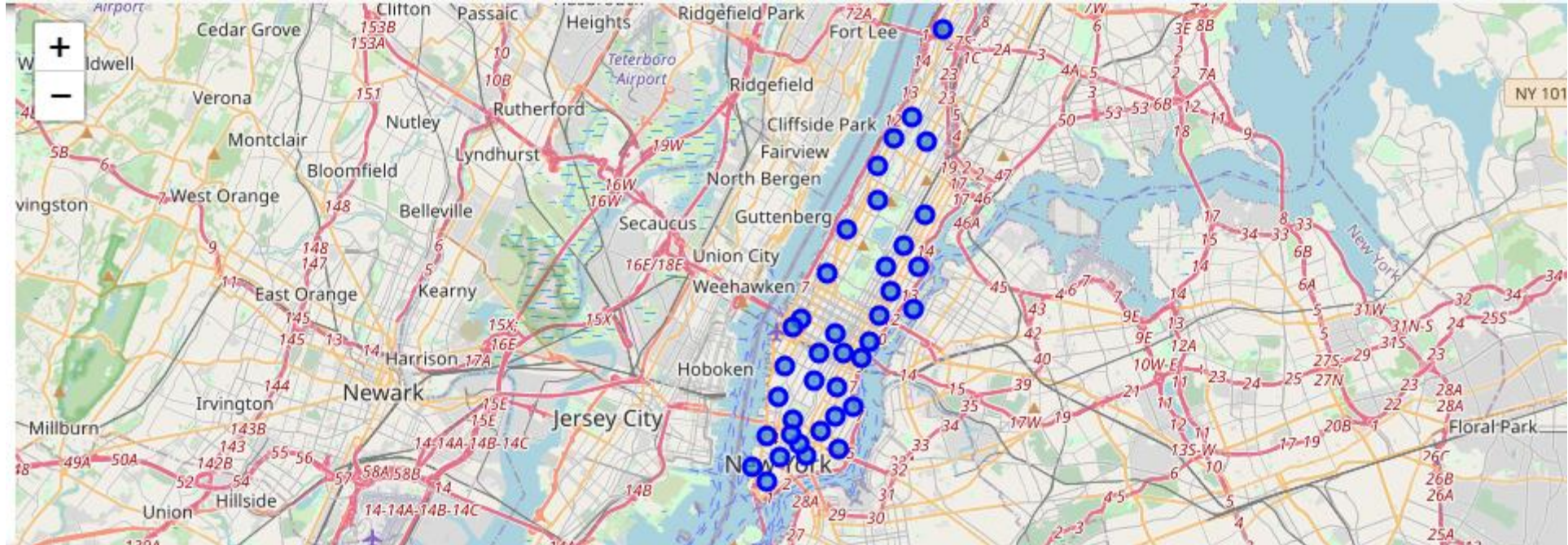
- ❑ **Data science techniques:** One-hot encoding, k-means clustering, Folium visualization library



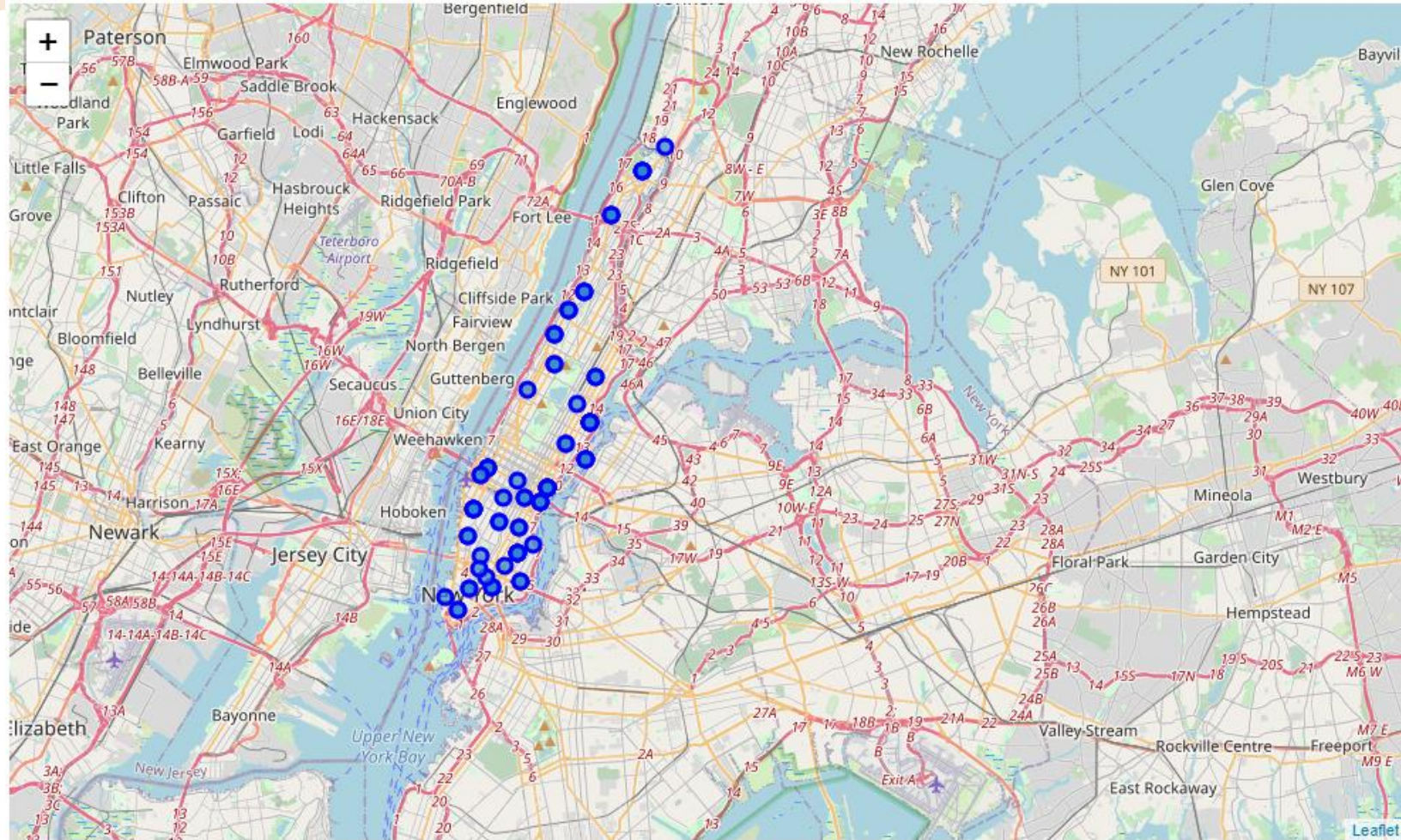
Map of NYC using latitude and longitude values



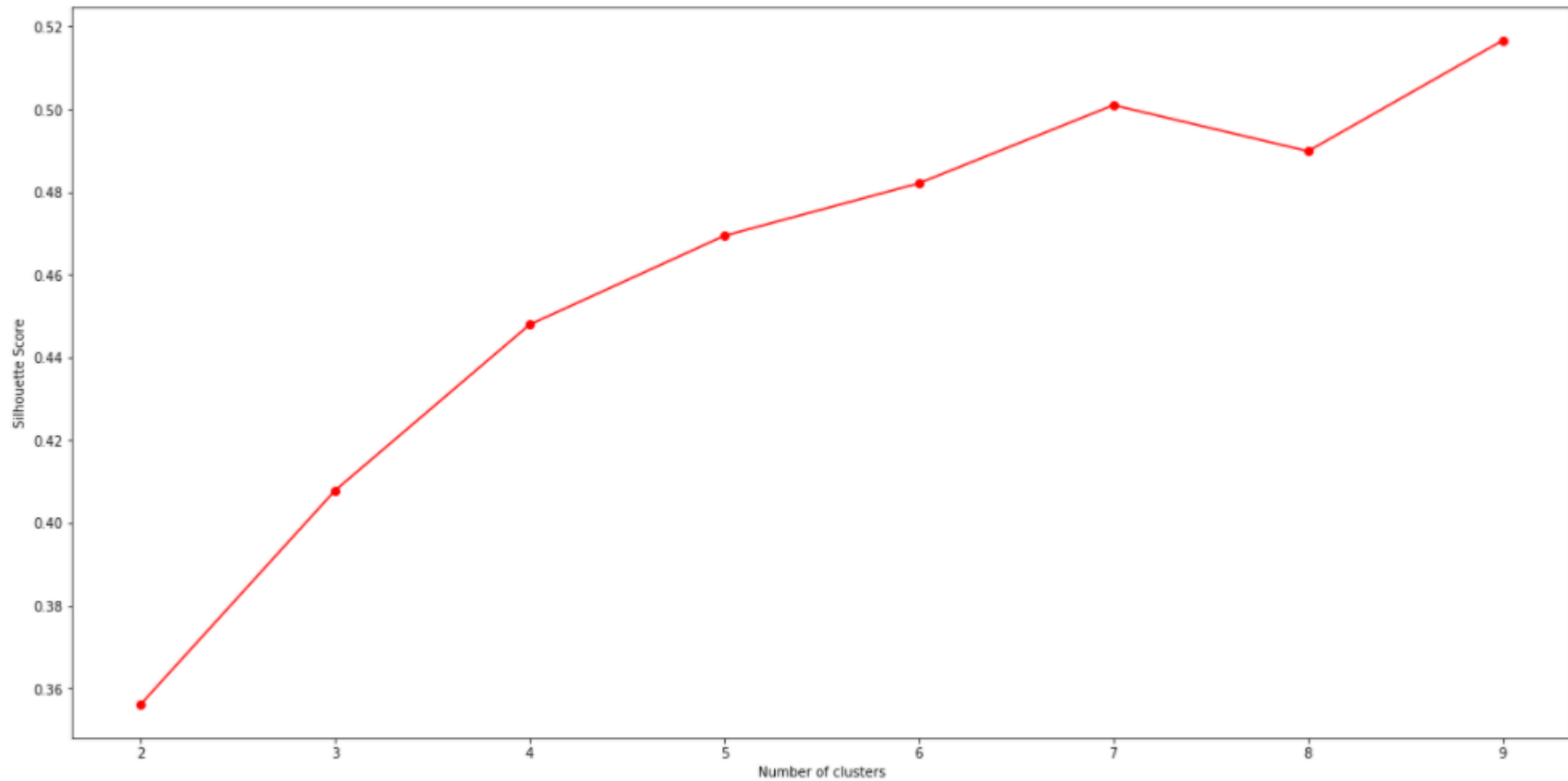
Manhattan neighborhoods



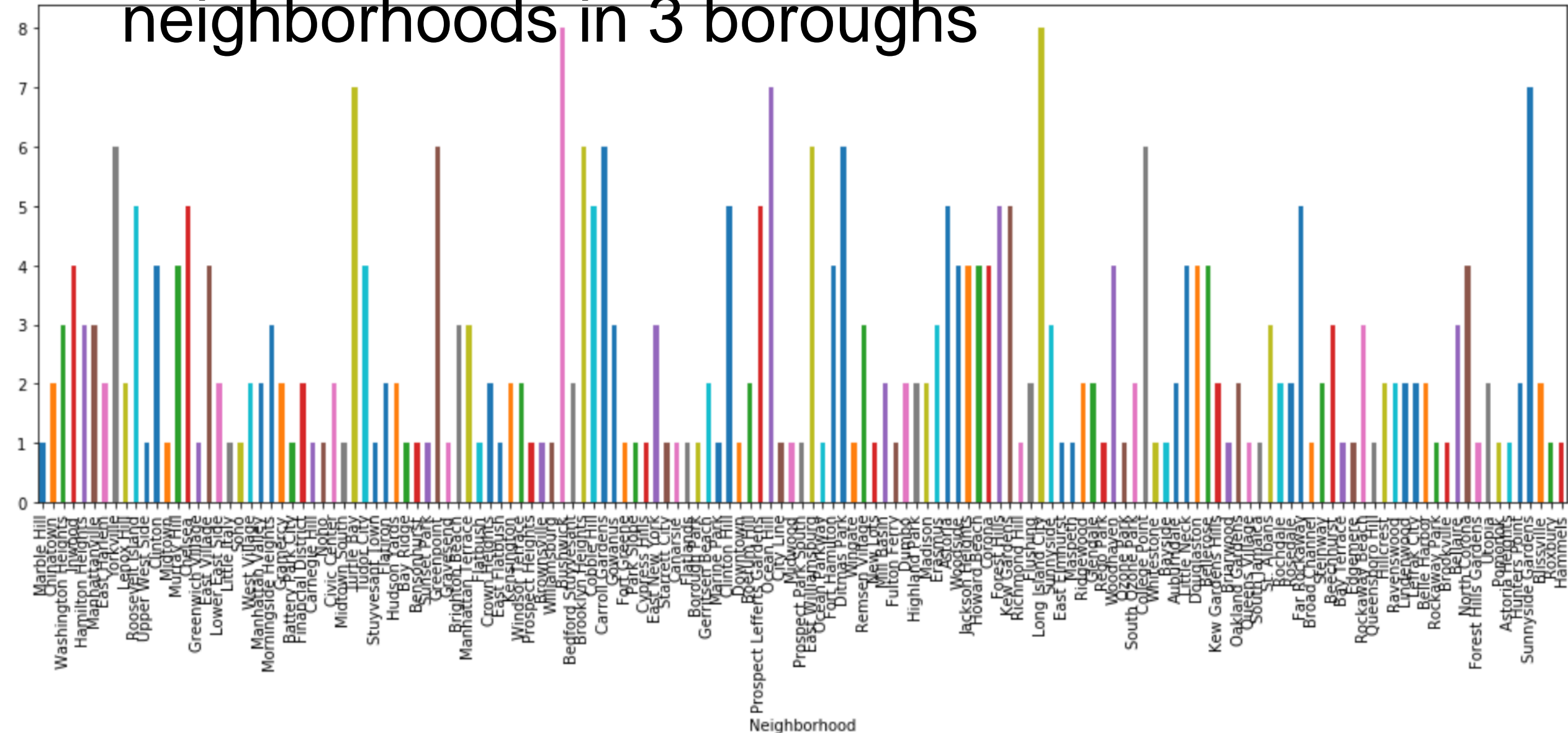
Manhattan food venues



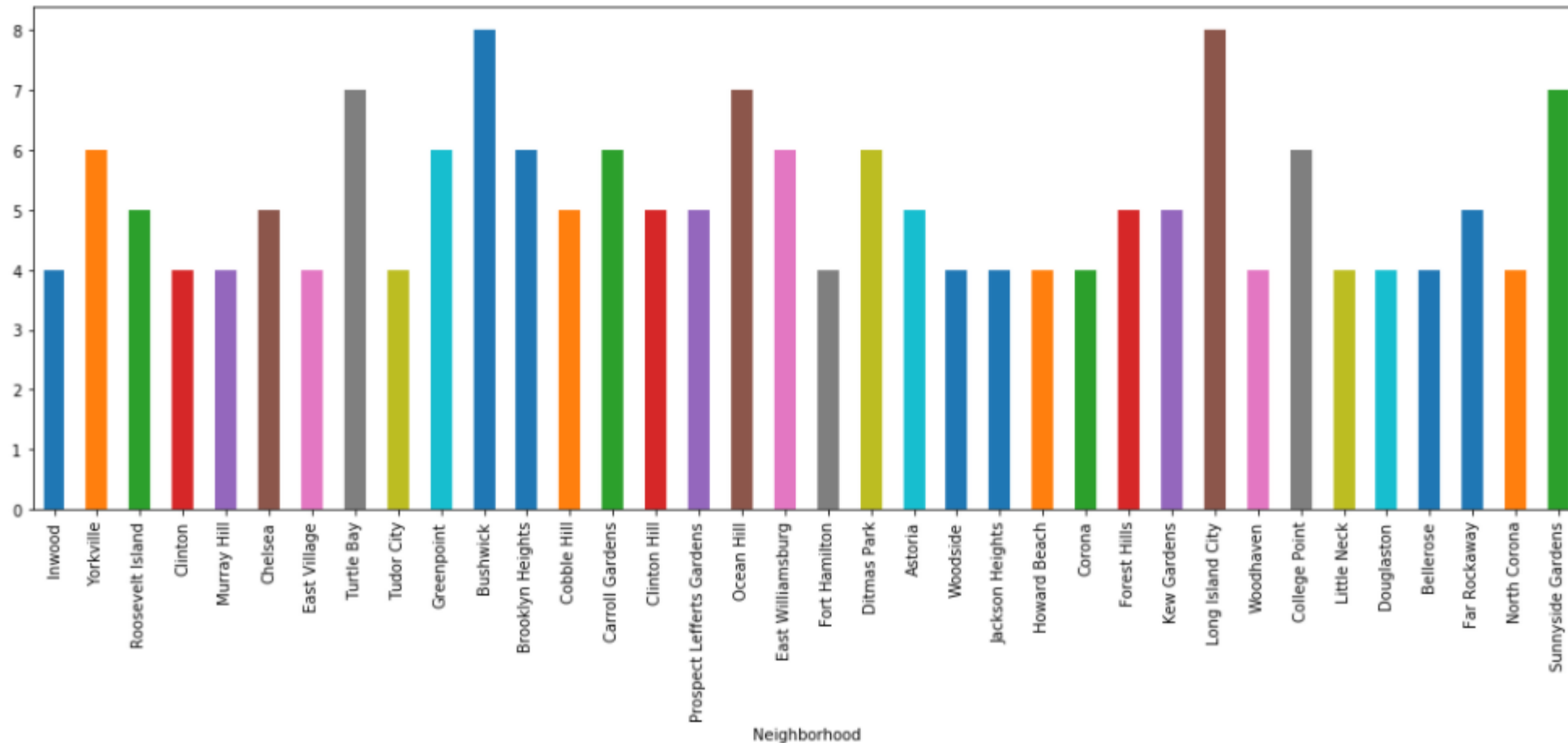
Sillhouette score



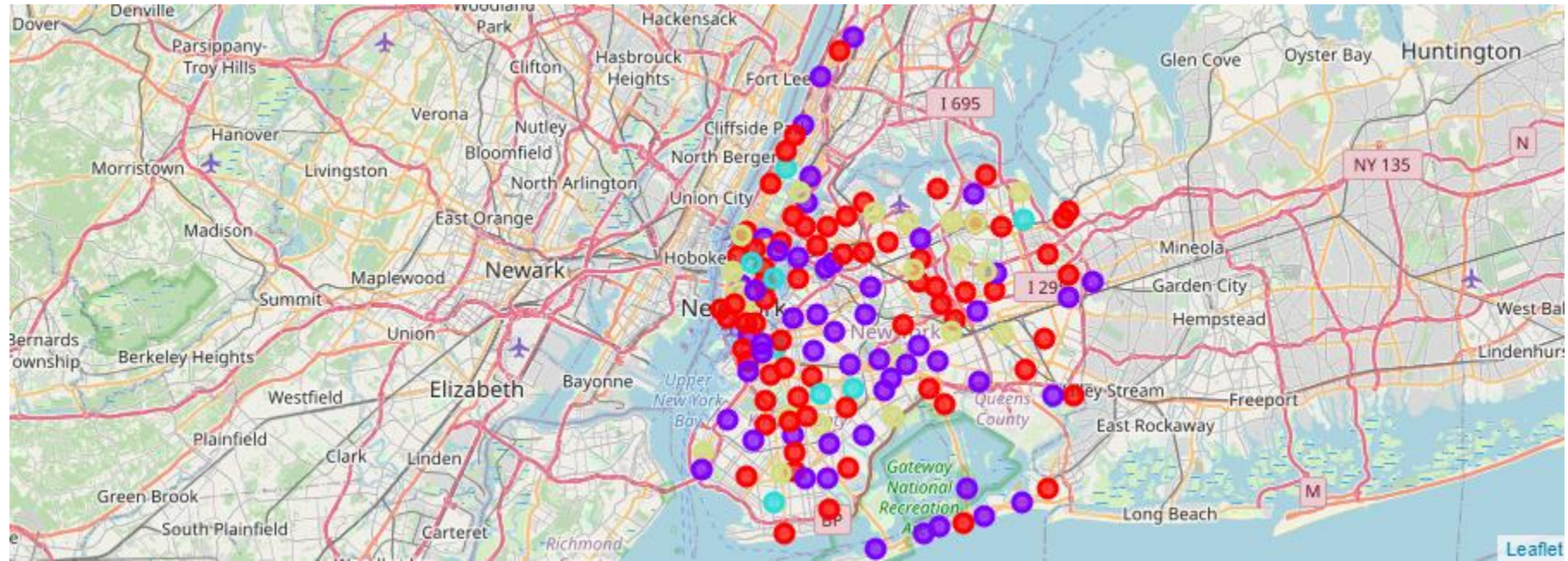
number of venues obtained in all the neighborhoods in 3 boroughs



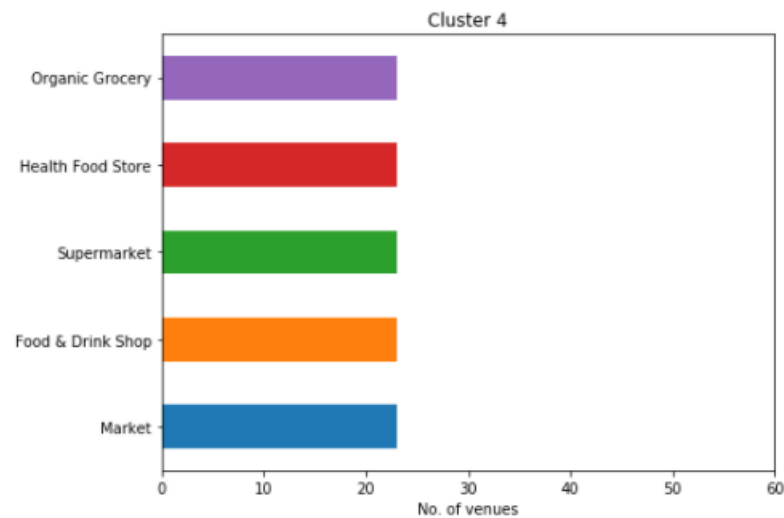
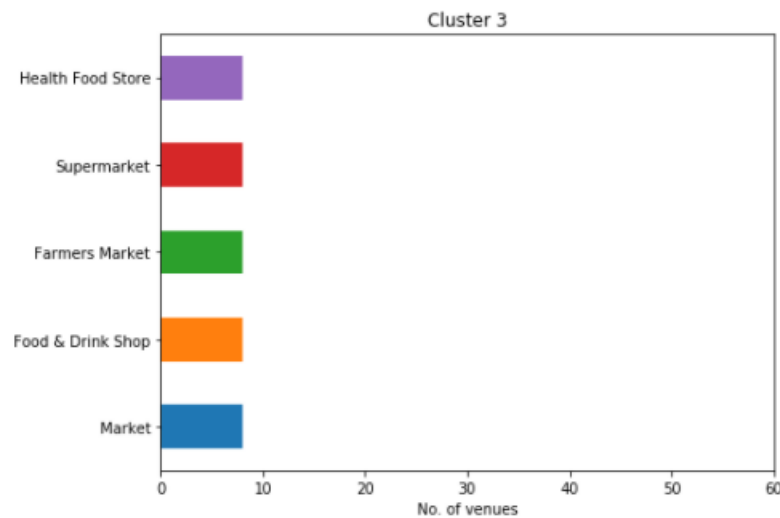
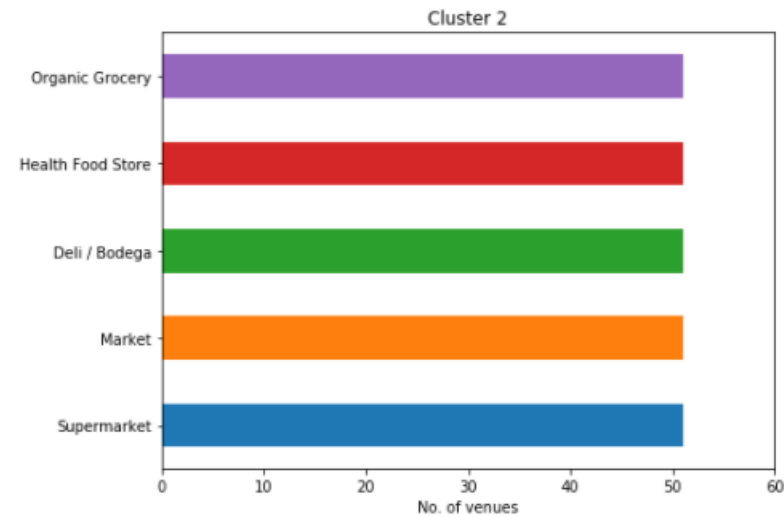
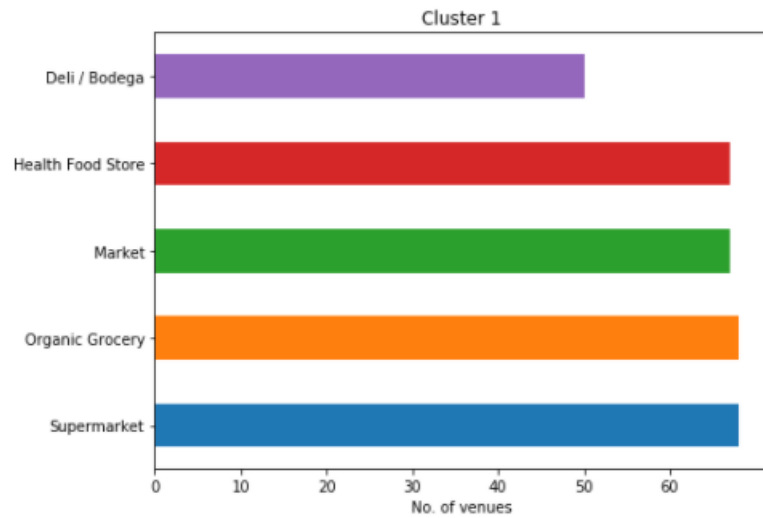
Neighborhoods with more than 3 venues



K-clustering



Top 5 most common food venue categories in each of the clusters



Conclusion

The purpose of this project is to analyze the neighborhoods of NYC and create a clustering model to suggest the best location for a packaging free circular grocery shop. The neighborhoods data was obtained from an online source and the Foursquare API was used to find the major venues in each neighborhood. We merged the three most popular boroughs, Manhattan, Brooklyn and Queens, and build a data science model. We selected our targets related to food markets and organic grocery shopping in all neighborhoods in the respective boroughs and used this as input to create a clustering model. The best number of clusters was obtained using the silhouette score. Each cluster was examined to find the most venue categories present, that defines the characteristics for that particular cluster.

A few examples for the applications that the clusters can be used for have also been discussed. A map showing the clusters have been provided. Both these can be used by stakeholders to decide the location for grocery store. Our suggestion would be cluster 4, as it has (1) the least competition in terms of total venues, but (2) interest in organic markets and healthy foods (given the current venues). A major drawback of this project was that the Foursquare API returned only few venues in each neighborhood. As a future improvement, better data sources can be used to obtain more venues in each neighborhood. This way the neighborhoods that were filtered out can be included in the clustering analysis to create a better decision model.