Classifying New York City Subway Stations

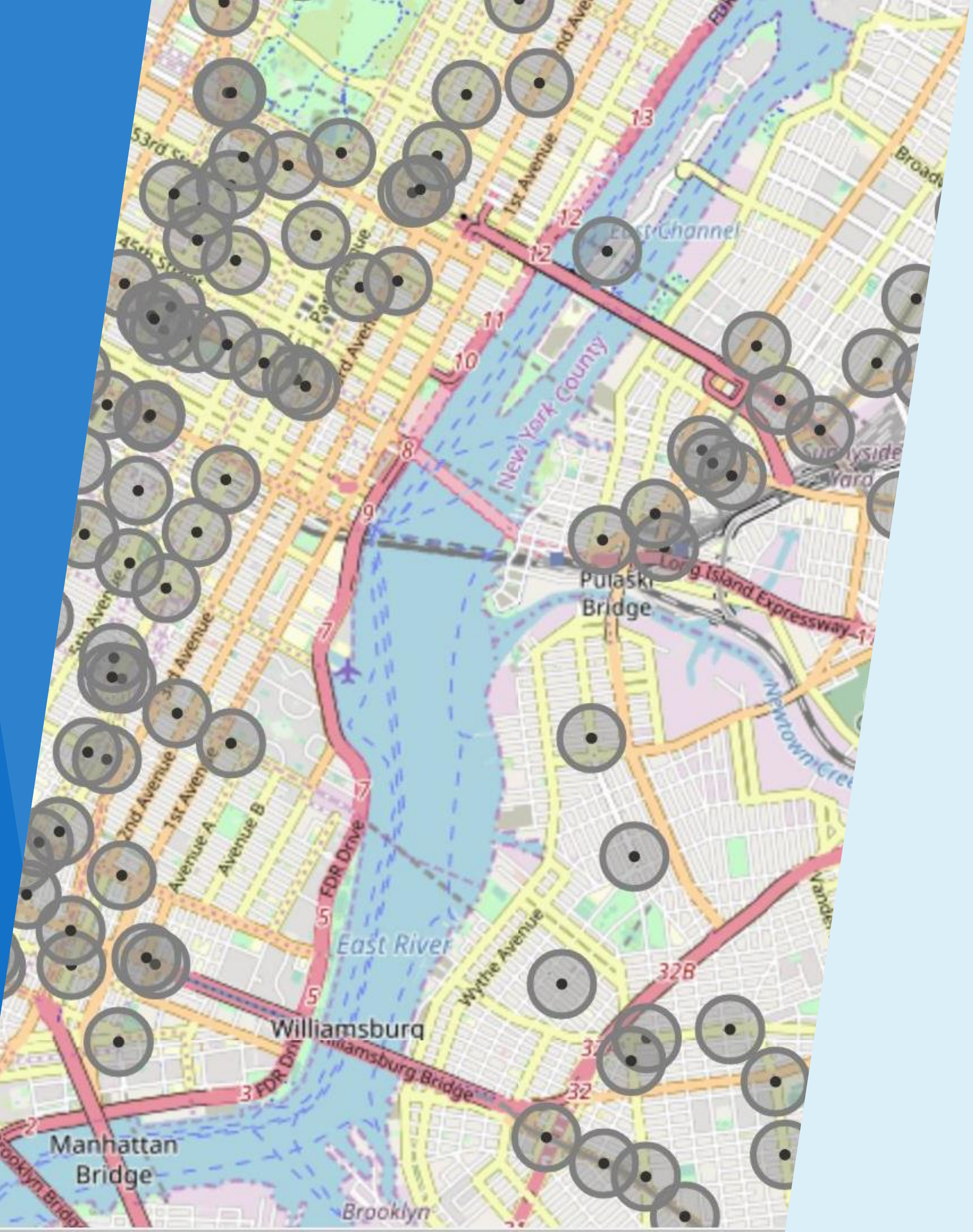
Ryan Albertson

What’s the Purpose of this Project?

- *473* subway stations in NYC will be analyzed along with venues that surround each station.
- **The goal is to**
- Provide city planners with useful data for estimating human traffic.
  - The amount of venues surrounding each subway station.
  - The categories of each of these venues
  - The most common venue categories for each station
- Discover clusters of subway stations that may hold significance.

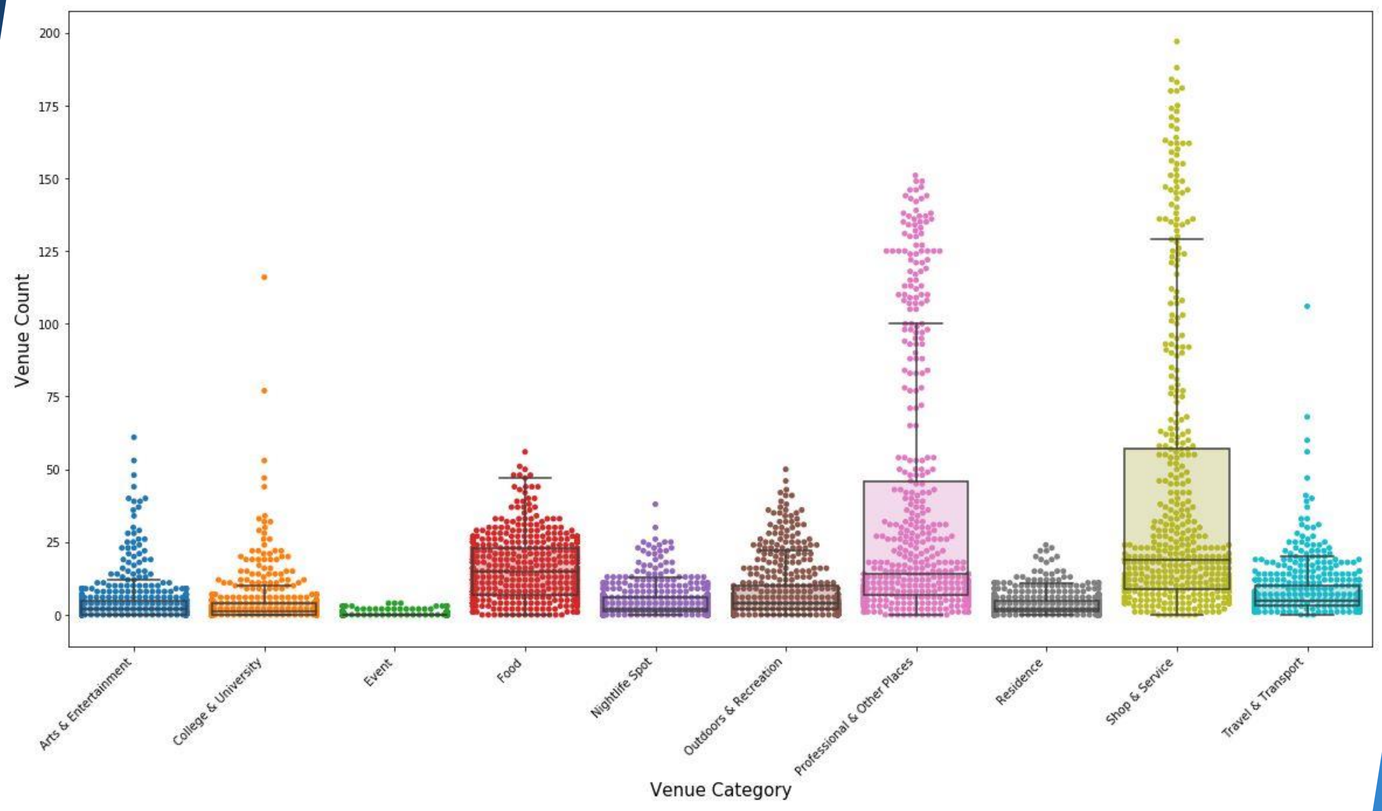
Data

- Subway station names and coordinates were taken from NYC Open Data
- There are 473 total subway stations in the dataset
- Venue data was taken from *Foursquare API*
- The *Event* venue category was dropped because of insufficient data



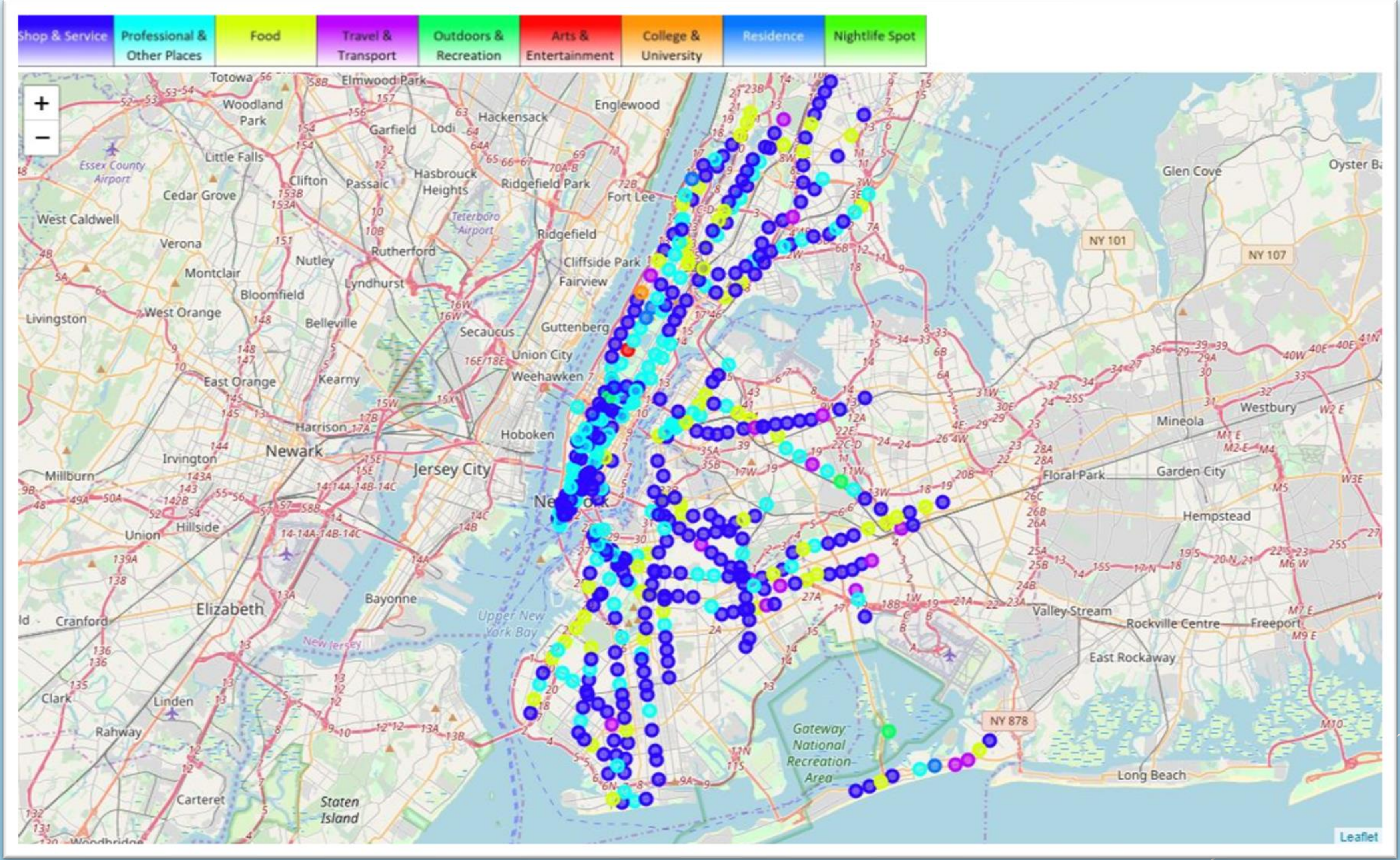
Map of Some Subway Stations and the Venue Search Radius around each

Visualizing the Count of Venues in Each Category for each Station



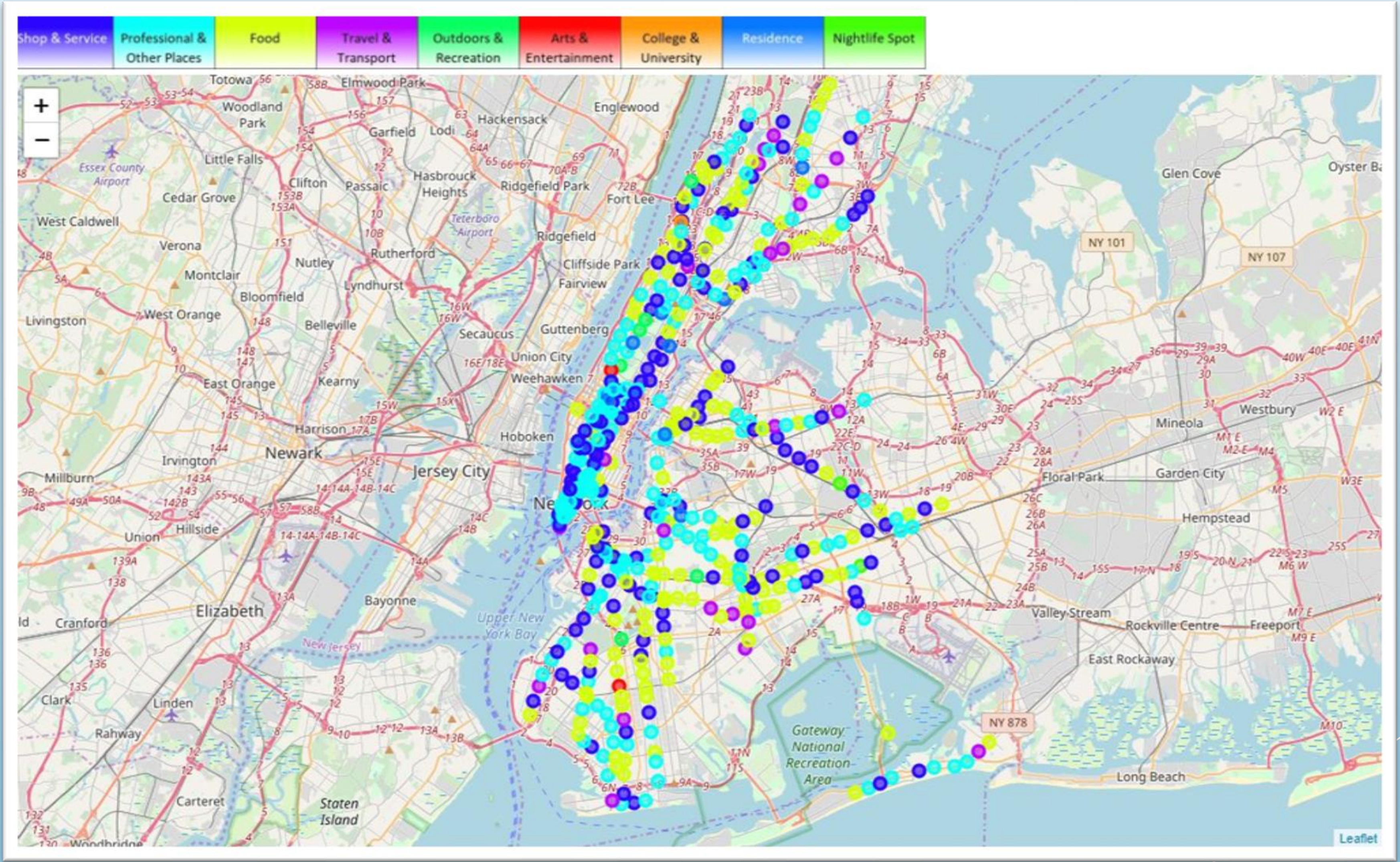
1st Most Common Venue Categories

for each Subway Station



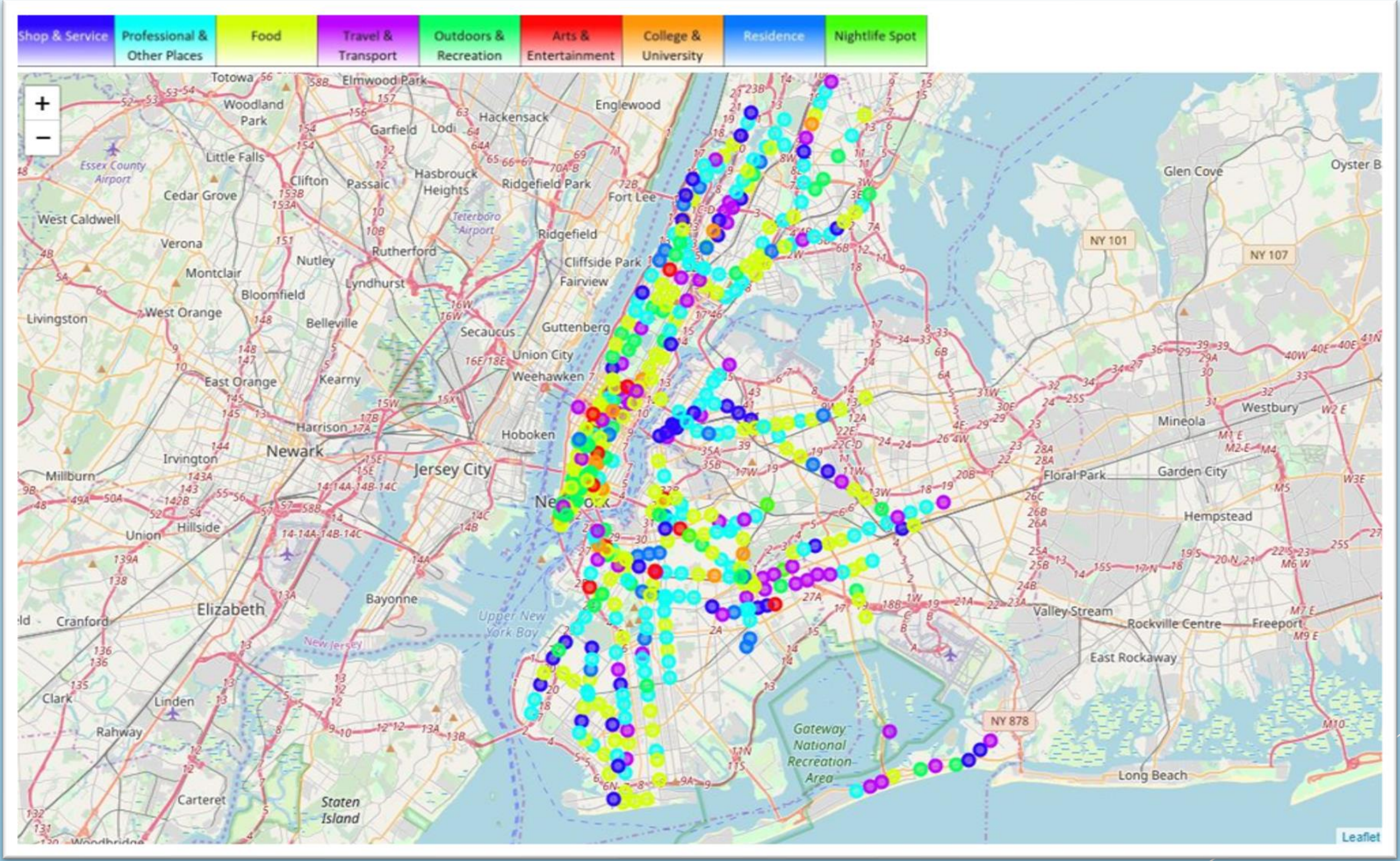
2nd Most Common Venue Categories

for each Subway Station



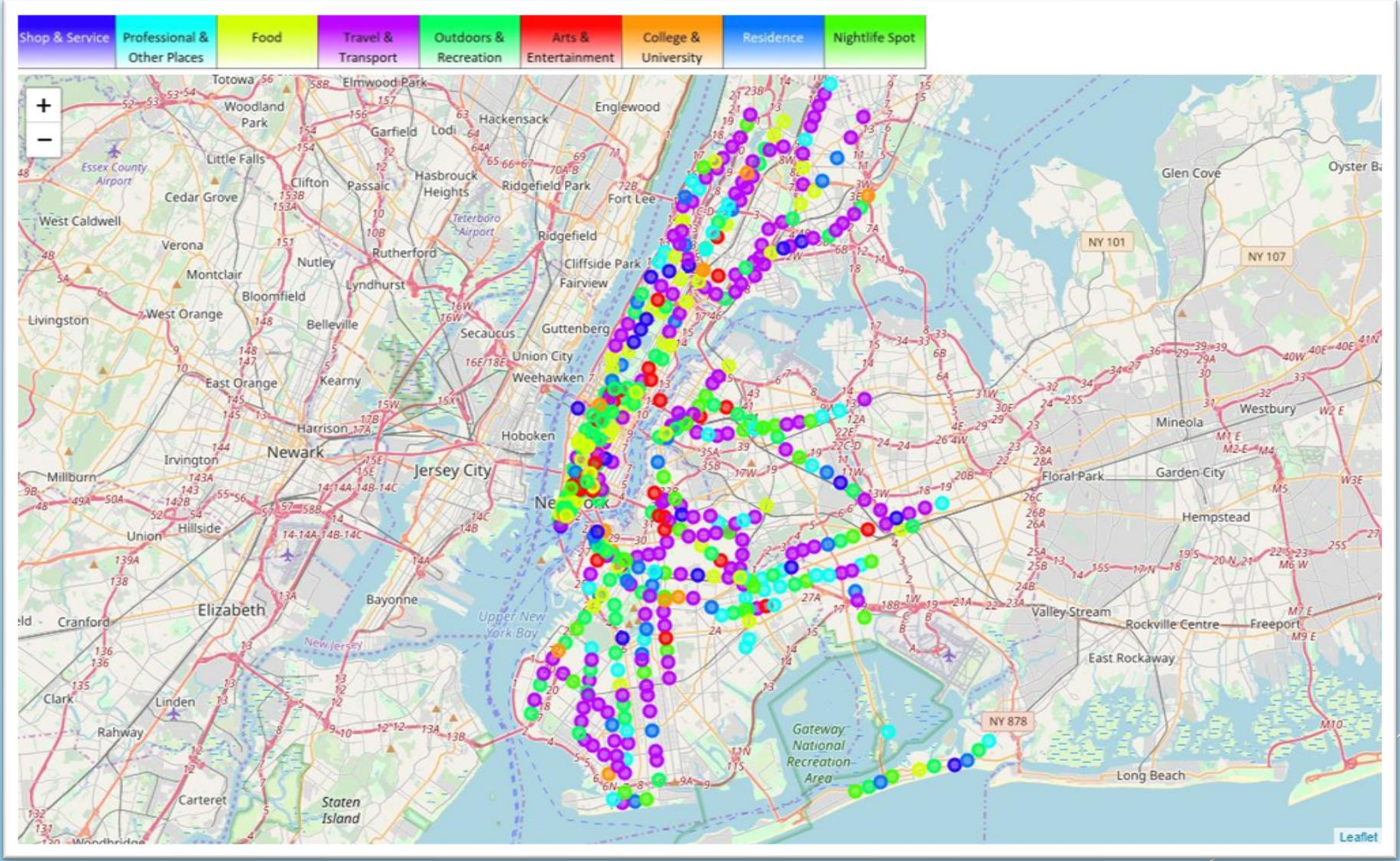
3rd Most Common Venue Categories

for each Subway Station



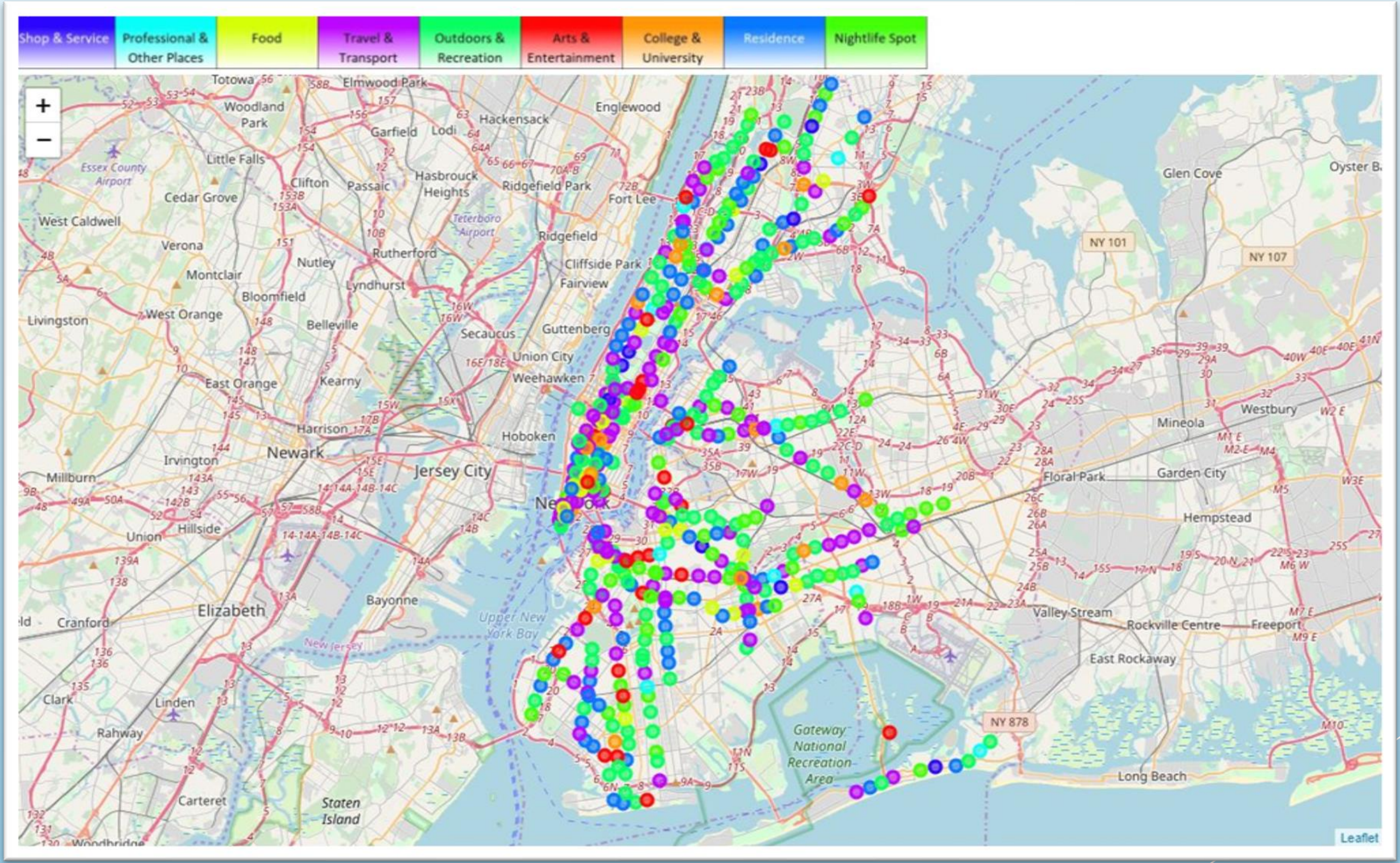
4th Most Common Venue Categories

for each Subway Station



5th Most Common Venue Categories

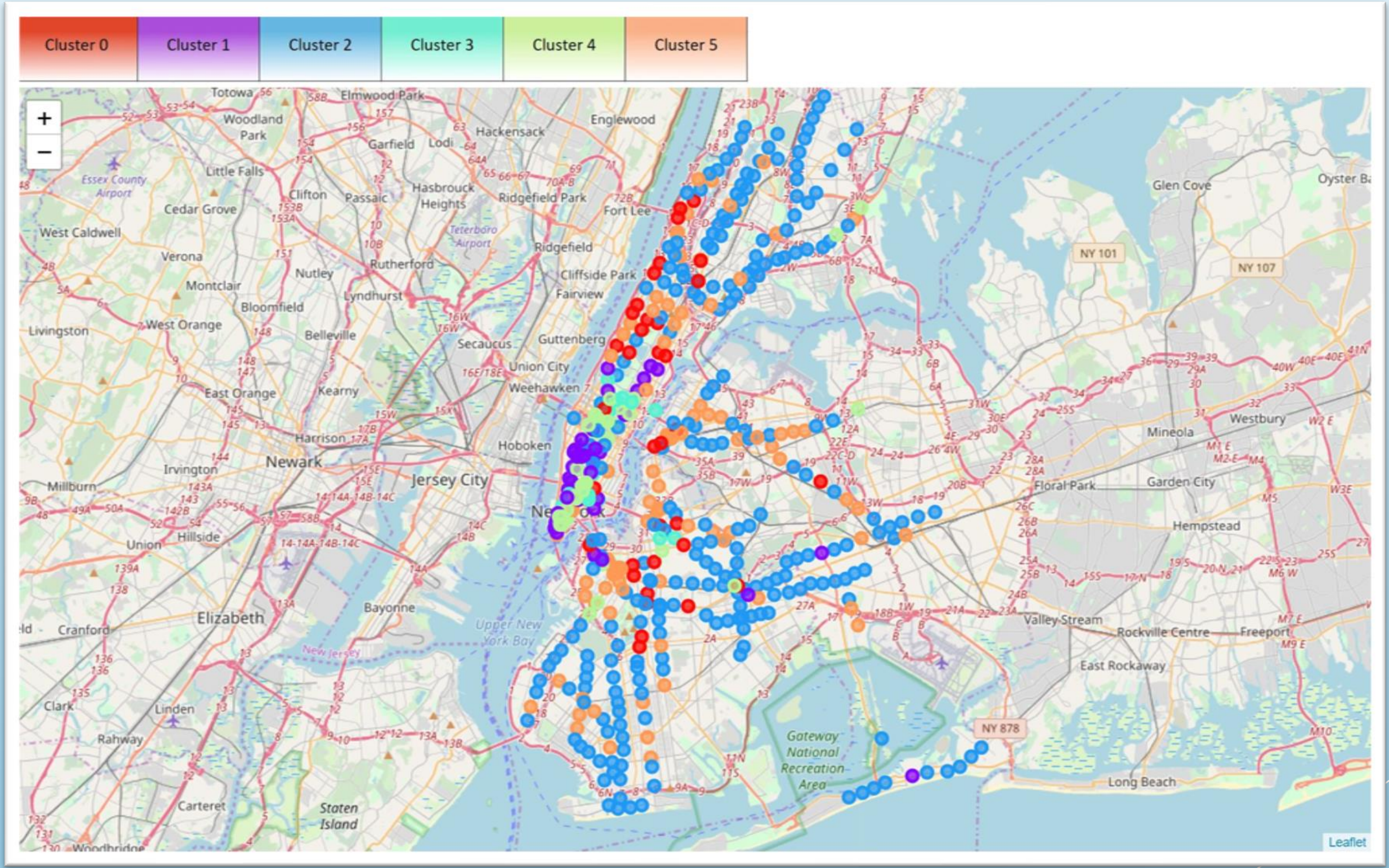
for each Subway Station



K-Means Clustering

- This algorithm will group subway stations into clusters based upon the Euclidean distance between them.
- The silhouette method suggested a k-value of 6.
- Clusters 1, 3, and 4 have similarities with their venue categories, as well as being similar in location.

Map of the Clusters



Discussion

- Venue categories can be associated with times of congestion for stations.
    - This allows city planners to estimate when certain areas will be busy.
- Sub-categories could be explored to provide more insight, at the cost of less datapoints.
- Clusters 1, 3, and 4 all have similar compositions of venue categories.
    - *Shop & Service* and *Professional & Other* are the two prominent categories.
    - In addition, these clusters are grouped well geographically.
    - It’s possible that there’s more insight to be found in these clusters.