Finding the best location in Manhattan, NY for an Italian Restaurant Supply business

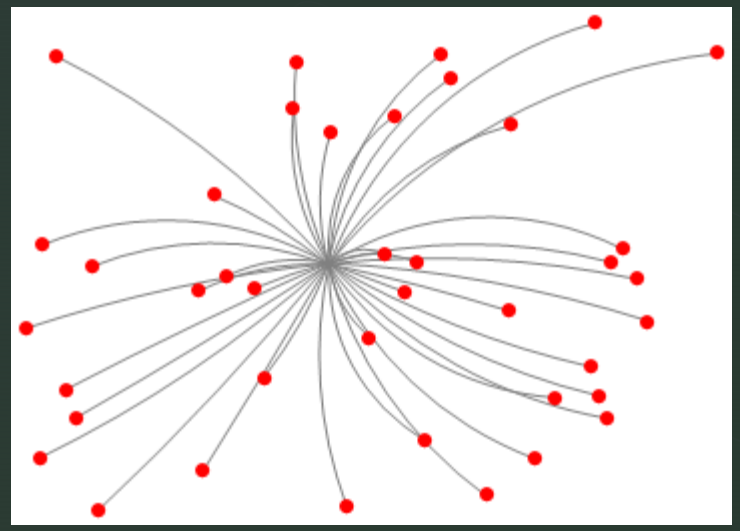
A data science analysis

Background

This Photo CC BY-SA

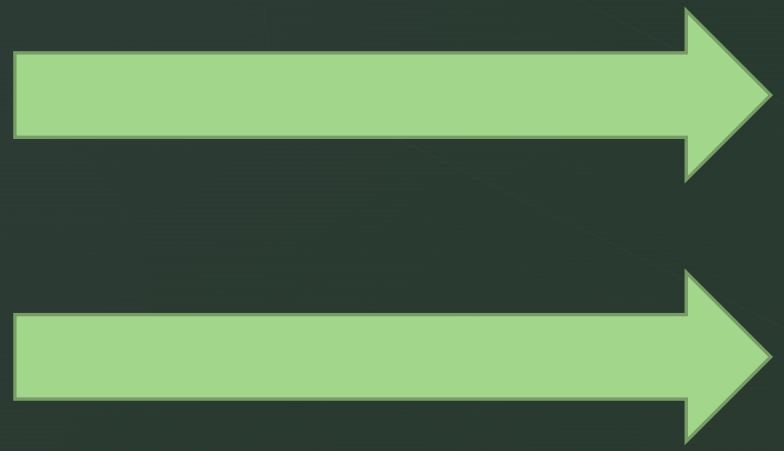
- Manhattan, NY has the best Italian and Pizza restaurants
- Their supplies and ingredients come from suppliers
- Suppliers want to be close to customers to minimize delivery time and expense

The Problem



- Find a cluster of potential customers and then find the central location from which to serve them.

Approach and Methodology



- Use online resources to find the neighborhood with the most potential customers
- Use data science to establish customer clusters and identify the optimized center of those clusters

Data Sources

- **Venue type and location data** – Foursquare, an online venue location and reviewer tool which has both free and upgraded accounts available.
- **Neighborhood data** – An online file with names and geo coordinates for Manhattan neighborhoods namely https://geo.nyu.edu/catalog/nyu_2451_34572

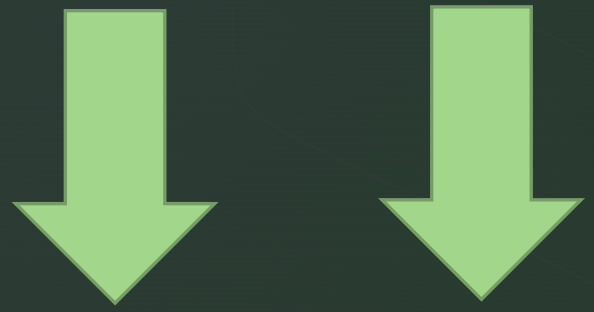
Data Tools

- **Mapping and data visualization** – Geopy libraries from Nominatim for translating address and geo locations and Folium an open source tool for mapping.
- **Source Code** – Python
- **Source Code Storage** – Github, an online source code repository that also provides branching and merging for code.
- **Data manipulation** – Pandas and Numpy, both open source libraries for Python.
- **K means analysis** – Sklearn, another open source library for Python.

Results from the neighborhood survey

Neighborhood data

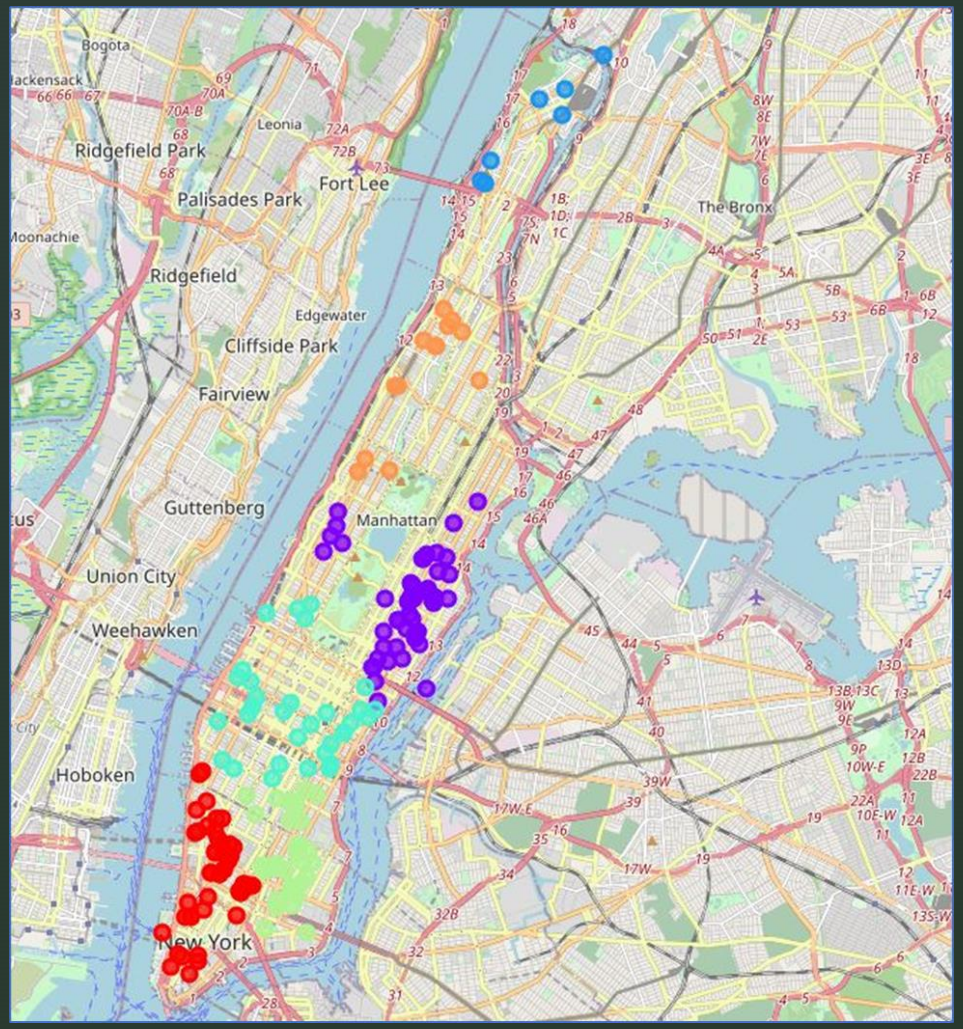
Foursquare data



Neighborhood	Customer Count
NoHo	14
Greenwich Village	12
Lenox Hill	12
Upper East Side	11
Yorkville	10
West Village	9

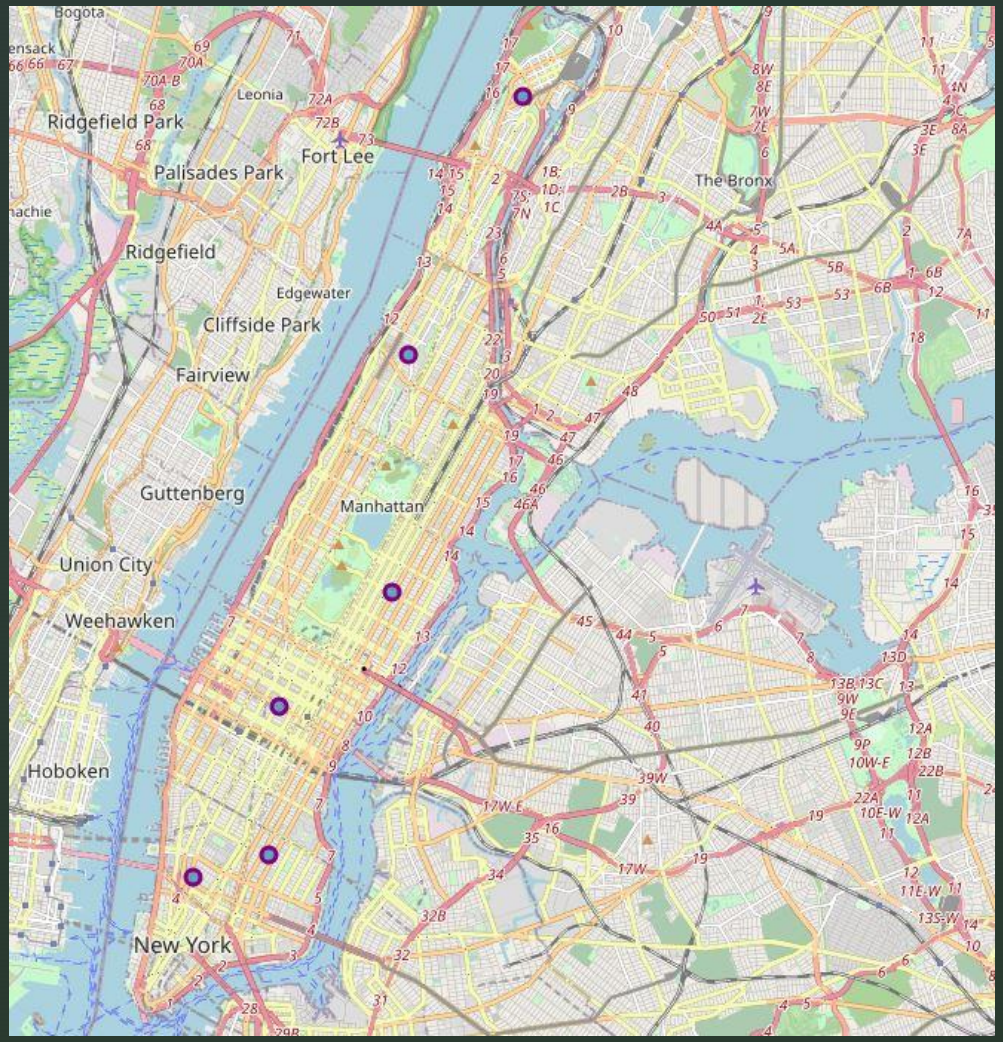
Combined online neighborhood geo data with Foursquare venue information to find the neighborhoods with the highest frequency of Italian or pizza restaurants

Results from the K means analysis



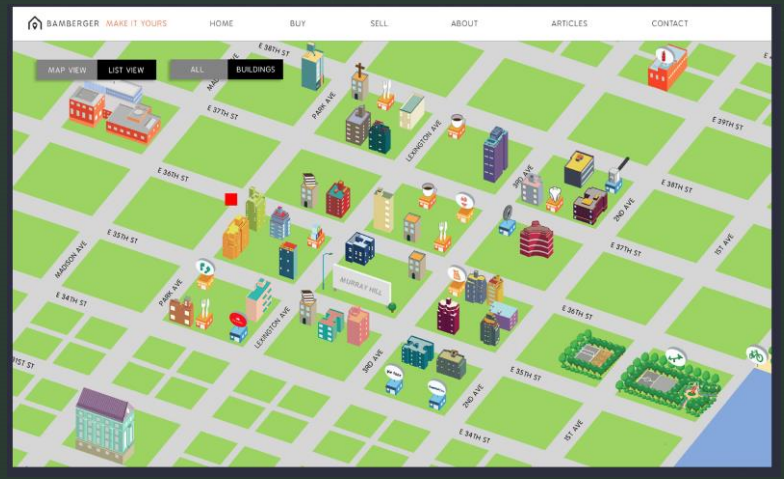
- Six clusters identified from 195 venue locations

Resulting Centroids



- Six mathematically optimized centers for those potential customer clusters

Discussion



- Other site selection factors
  - Cost / benefit
  - Supply chain
  - Actual transportation networks
  - Real estate availability/ suitability

Conclusion

- Not the final answer but:
  - Helps visualize customer locations
  - Helps identify starting points for site location search