



An insight on two neighborhood types in New York City based on their most prevalent venue categories

Introduction





Setting up shop in NYC is no trivial matter

- New York City (NYC) is the top global metropolis
- Lease space in global cities is limited and expensive
- Venue opening decisions need to be informed by quality, actionable data
- This report provides guidance by offering a machine learning-driven segmentation of the neighborhoods in NYC based on their venue composition

Data and methods





Data sources

- Source 1: a dataset available from IBM that offers a breakdown of boroughs and neighborhoods of NYC, with coordinates information
- Source 2: information on venues obtained through Foursquare using the Python requests library (see attached Jupyter notebook)



Data treatment

- Neighborhood-level data were obtained through source 1 in json format. Variables:
 - Borough
 - Neighborhood
 - Coordinates (latitude & longitude)
- Data were scraped into a pandas dataframe, and visualized on a map through folium
- Venues for each neighborhood fetched through source 2 (limit of 100 venues in a 500 m radius for each neighborhood)



Data treatment contd

- For each neighborhood, fetched venues were grouped by category
- Categories were used to train a K-means clustering algorithm using $k = 5$
- Two main clusters emerged, plus three minor clusters of no more than two neighborhoods each. Only the main clusters, coded here as A and B, were used in further analysis
- For the two main clusters, the aggregate neighborhood modes of the top 10 categories were computed
- The modes were grouped into three broad venue types: food locations, shopping locations, and transportation

Results



Folium visualization allows identifying NYC neighborhoods

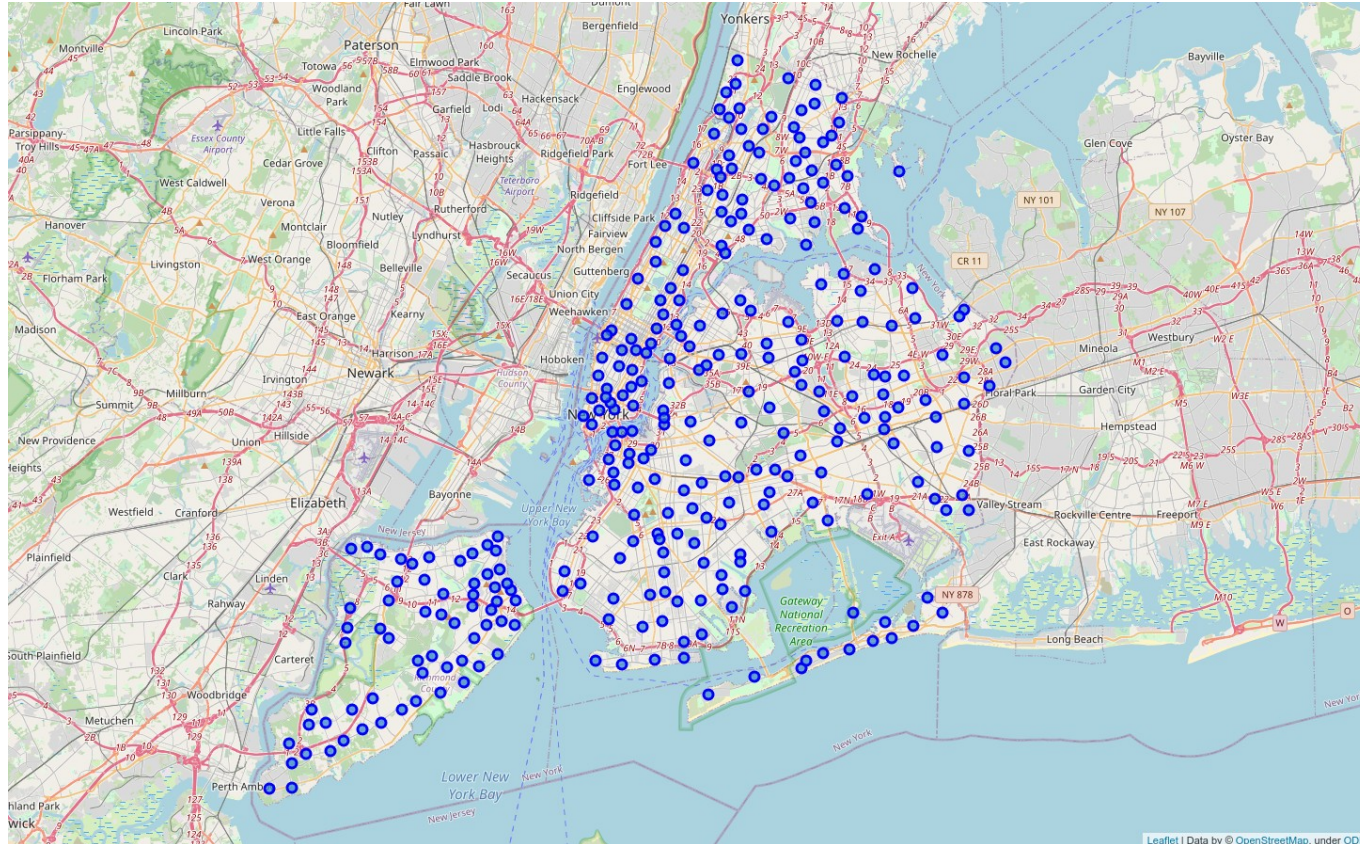
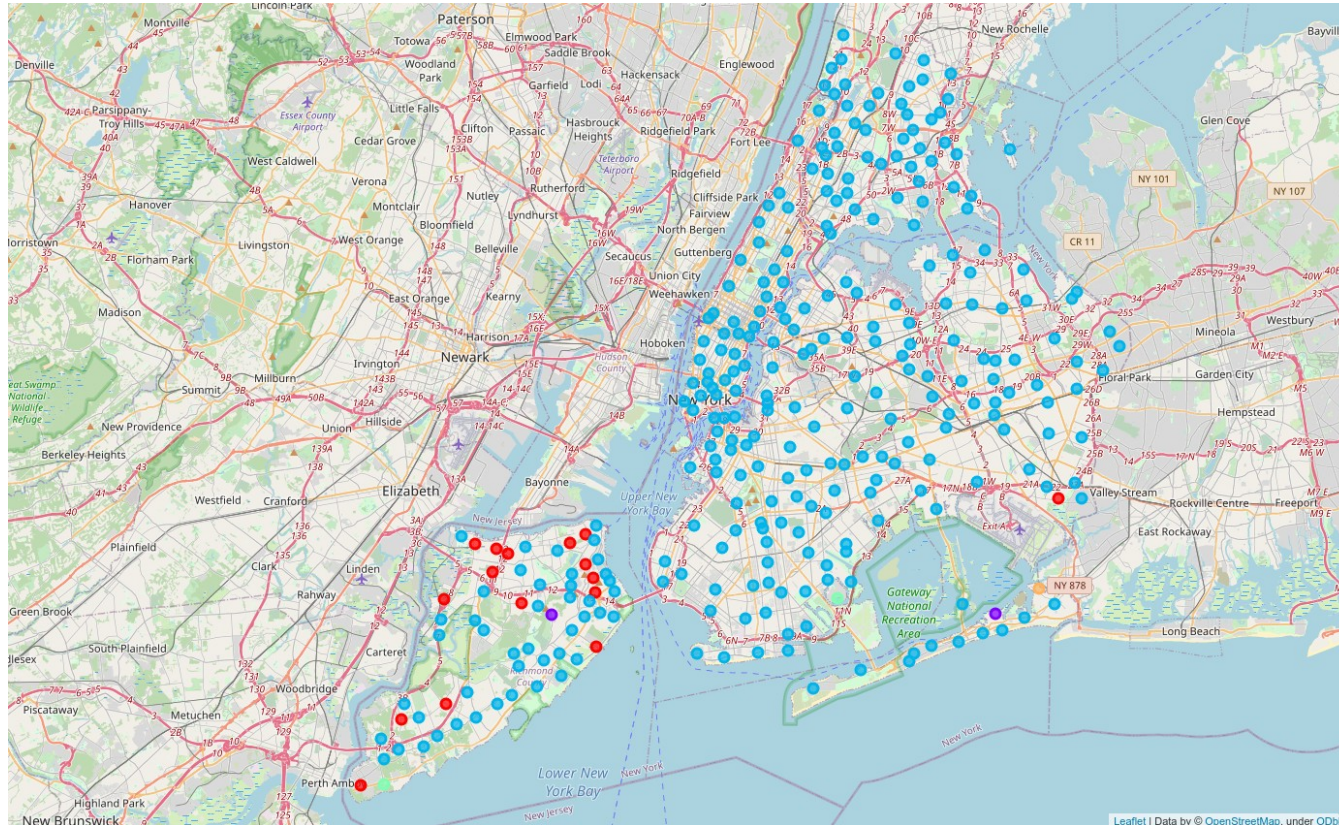


Figure 1: Map of New York City neighborhoods. Each blue dot represents a neighborhood

K-means clustering reveals two main clusters, A and B, and three outlier clusters



Red: cluster A
Cyan: cluster B

Figure 2: Map of New York City neighborhoods segmented into clusters based on most common venue types

Clusters A and B are dominated by distinct types of venues

- For each neighborhood in both clusters, the mode for each of the top 10 most common venue categories was computed
- Each of the 10 modes was assigned to one of three broad venue types: food, shops, and transportation

Types of venue	Cluster A share of top 10 modes	Cluster B share of top 10 modes
Food	2	8
Shops	5	2
Transportation	3	0

Clusters A and B are dominated by distinct types of venues

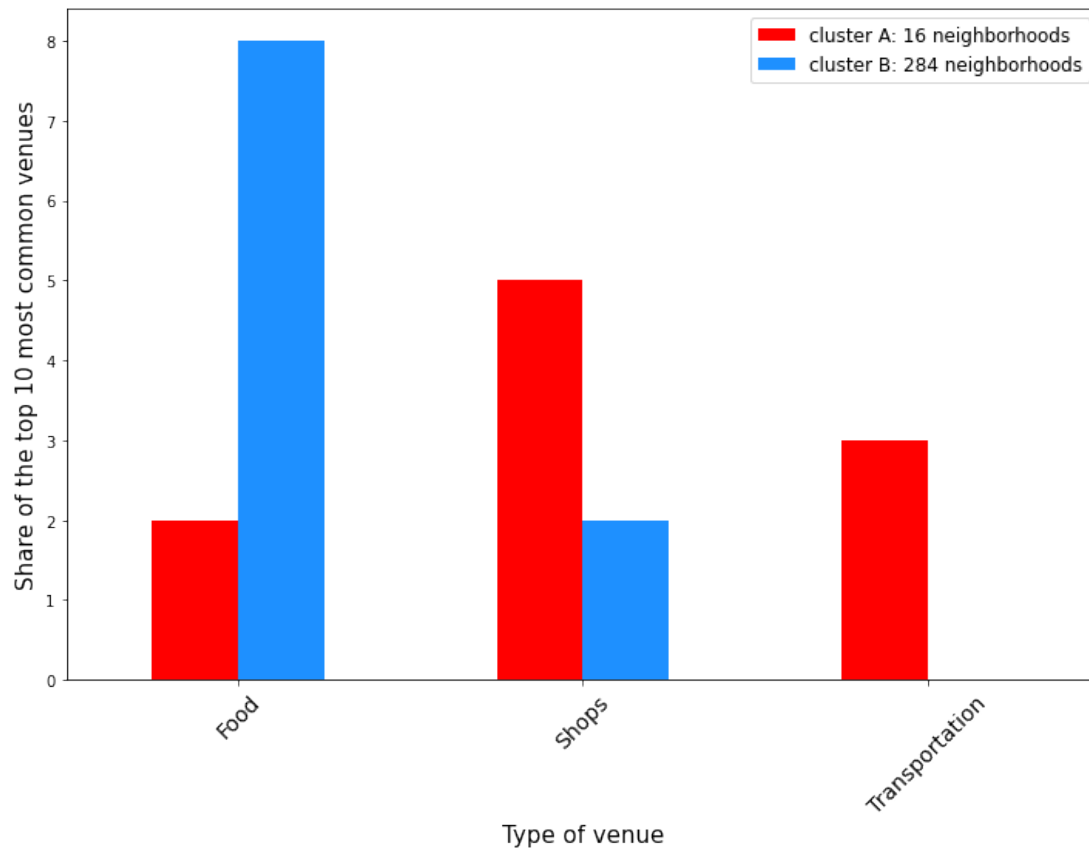


Figure 3: Venue type composition of the top 10 neighborhood-level category modes

Conclusion





Conclusion

- Two main types of neighborhoods can be defined:
 - Cluster A: neighborhoods that are heavily dominated by shops and frequent transportation,
 - Cluster B: neighborhoods that chiefly feature eating places, such as cafeterias, restaurants, and bars
- Cluster A is dominant throughout New York, but cluster B seems to be chiefly located within the Staten Island borough, possibly due to the unique transportation features of this area
- Prospective business owners should leverage this along other data sources to inform their venue opening decisions