



Clustering NYC: Mid-Term Project

By: Annie Stanley & Josh Ho

Agenda

1. Objective
2. Project Workflow
3. Exploratory Data Analysis
4. Modeling Process
5. Results



Objective

Objective

1

Provide relevant real estate information and insights to various stakeholders

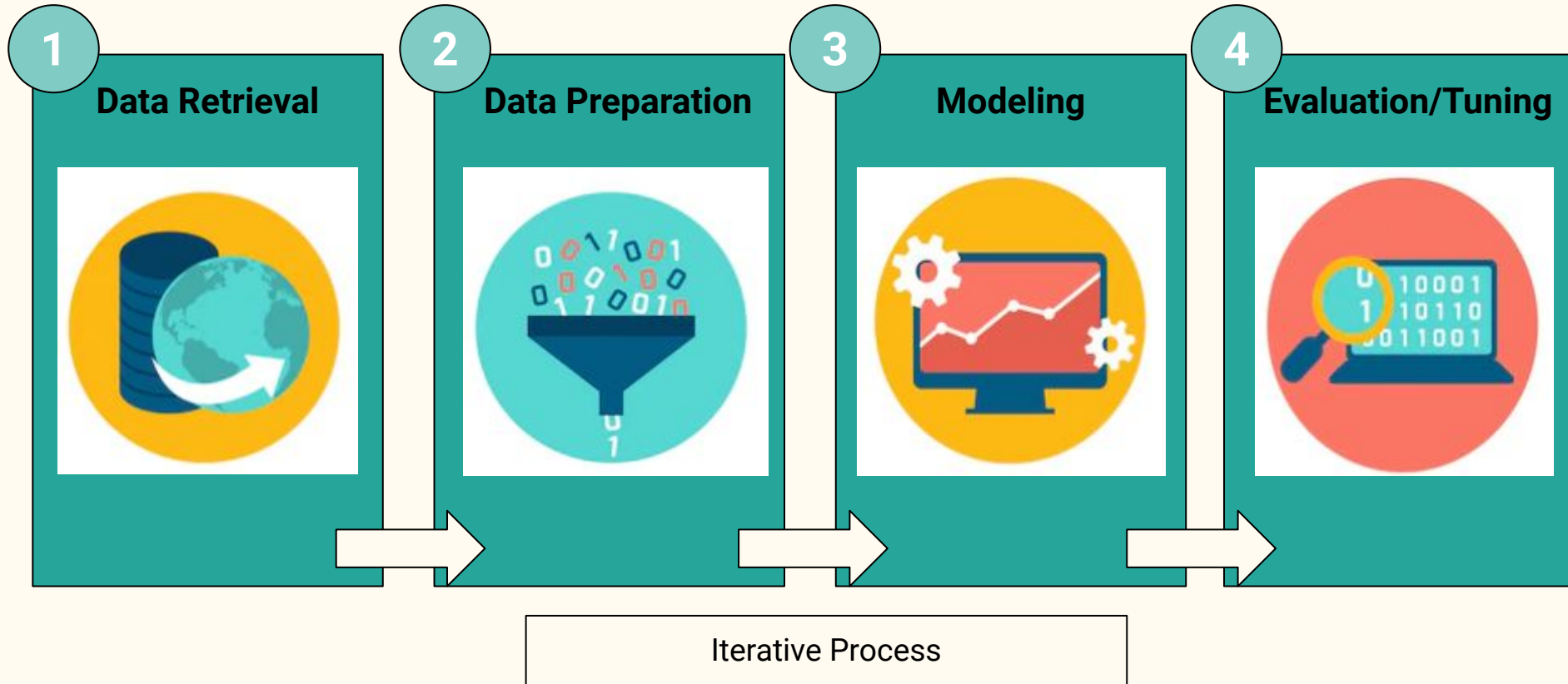
2

Ultimately, provide a big picture aspect to help stakeholders make informed decisions when moving to New York City

Stakeholders

- Buyers
- Renters
- Investors
- Real Estate Brokers

Project Workflow



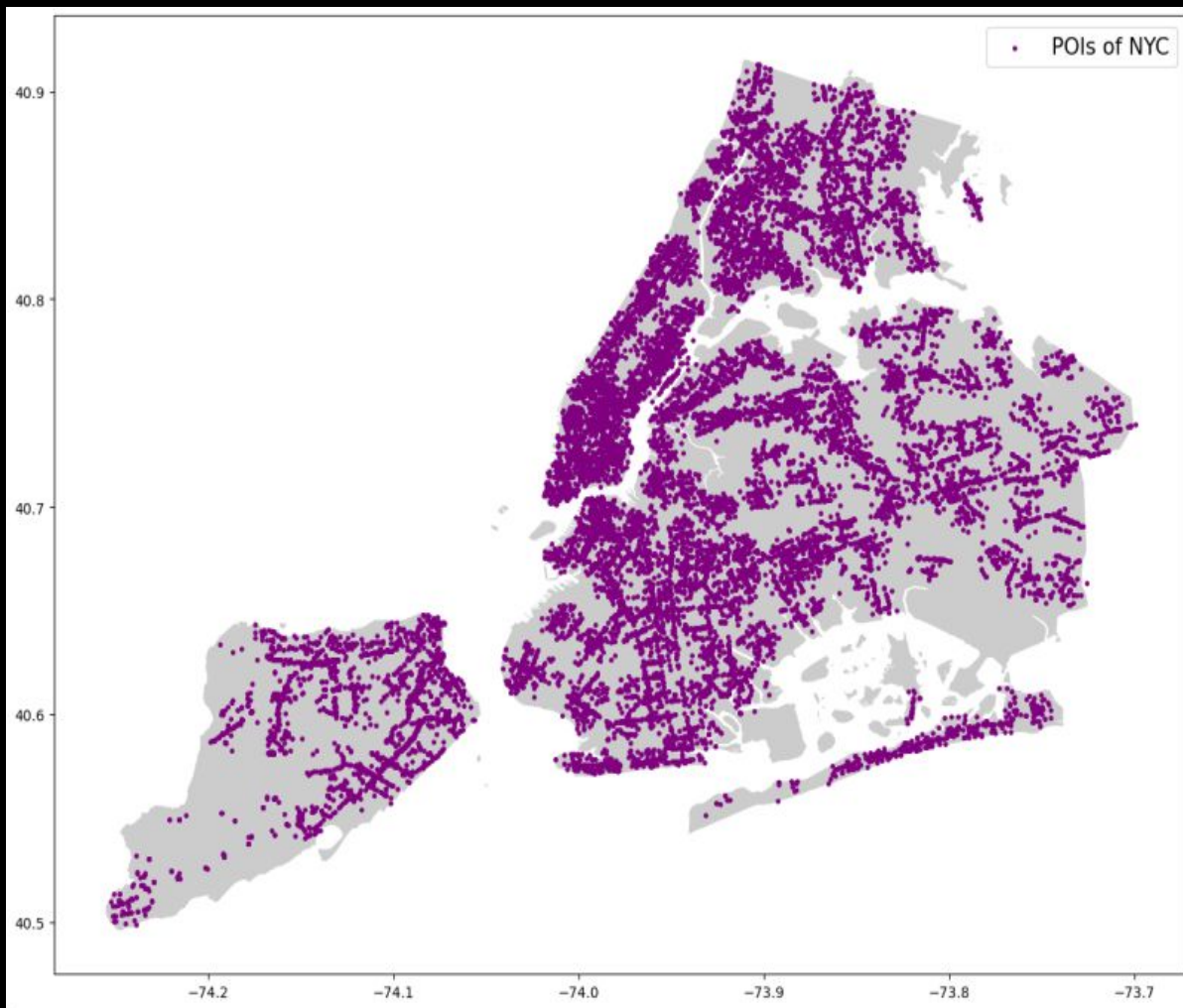
Sources of Data

Source	Content
NYC JSON, GeoSpatial	<ul style="list-style-type: none">○ Boroughs○ Neighbourhoods○ Zip Codes○ Geometry (Polygons)
Housing Prices	<ul style="list-style-type: none">○ Median Housing Prices 2021○ Average Housing Prices 2020
API's	<ul style="list-style-type: none">○ Yelp POI○ Foursquare POI○ Google Places POI
NYC OpenData (Public Repo Operated by City of New York)	<ul style="list-style-type: none">○ Uber Rides○ Rat Reports○ Subway Lines○ Community Districts

Exploratory Data Analysis (EDA)

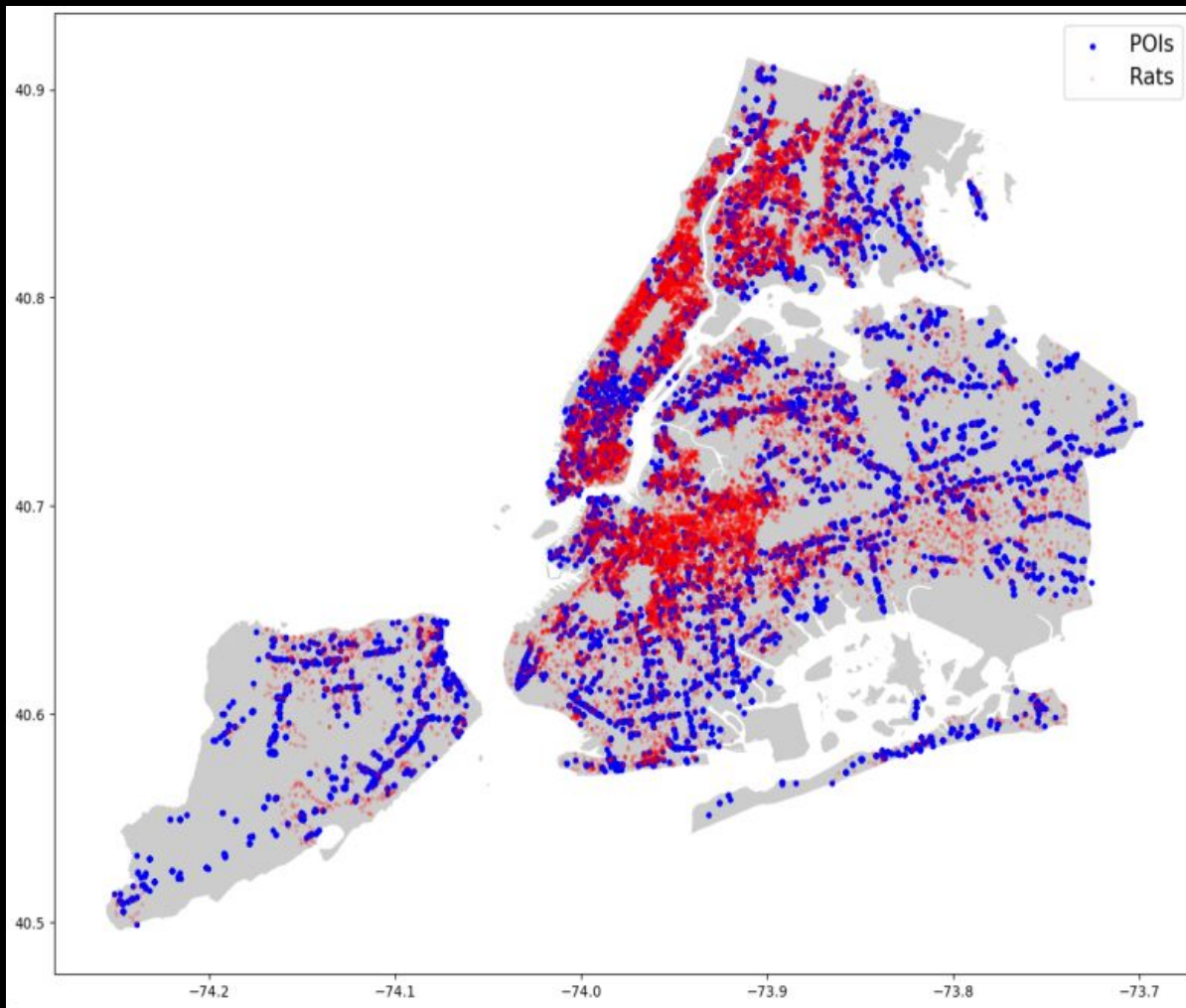
EDA

1. Places of Interest in NYC



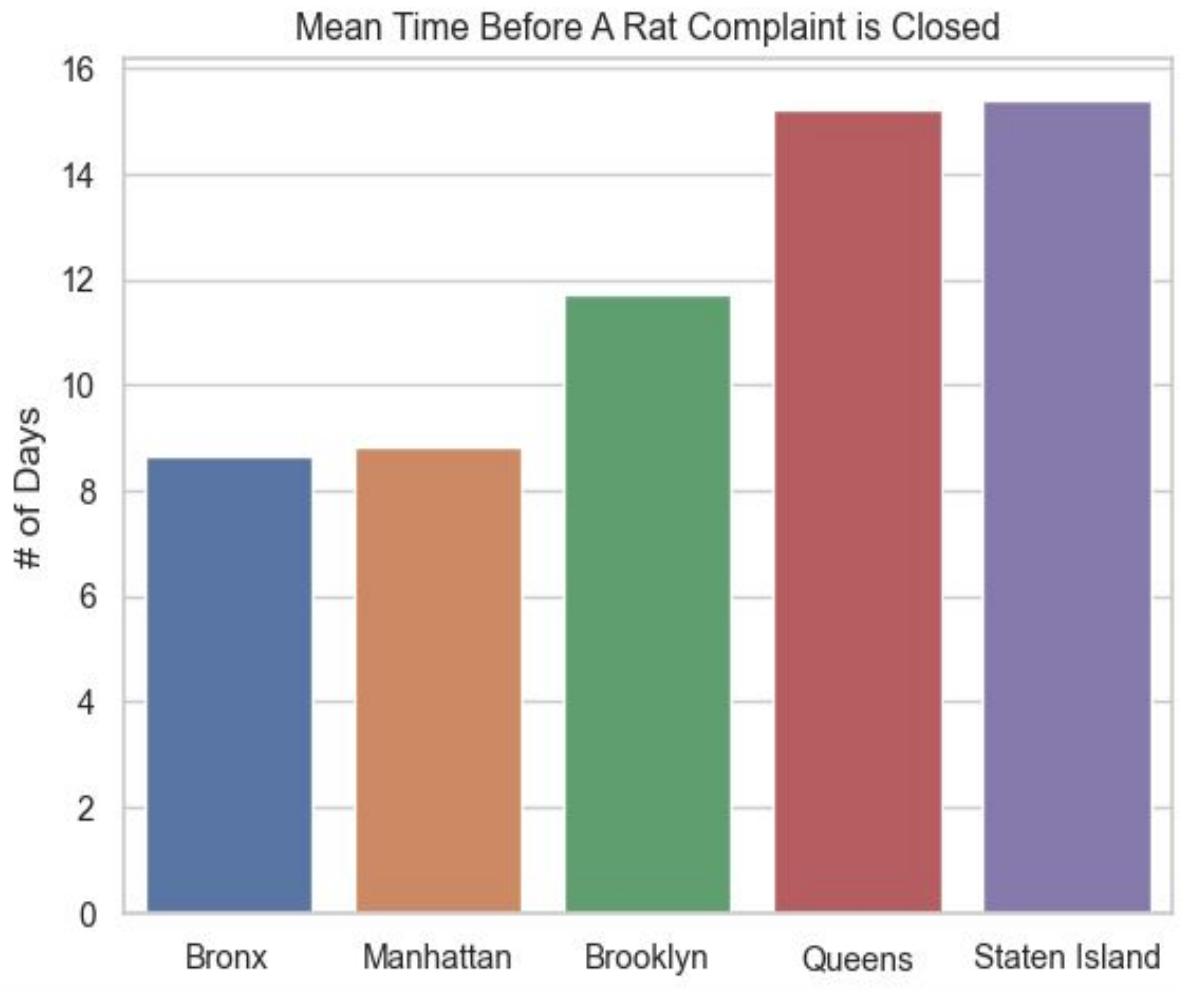
EDA

2. Rat Sightings vs Points of Interest in NYC



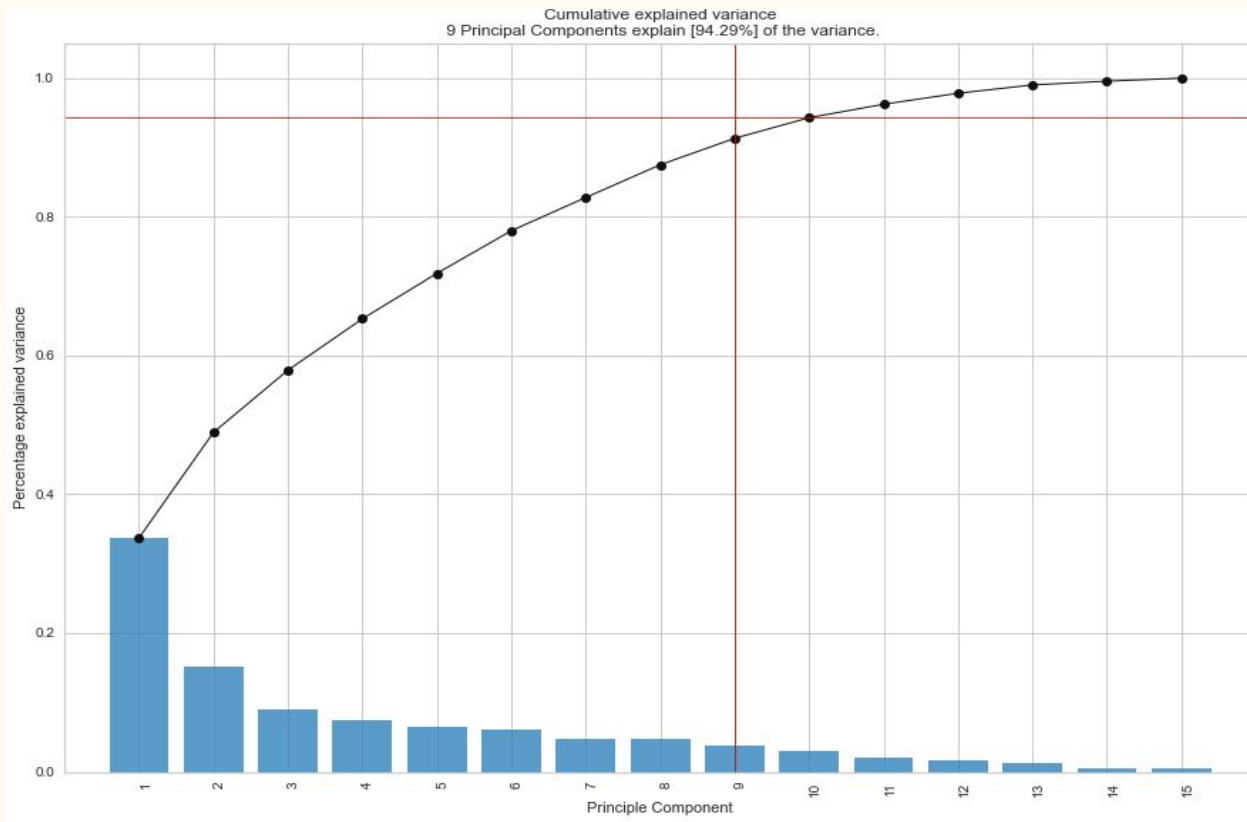
EDA

3. Average Time It Takes
For a Rat Complaint To Be
Addressed



Modeling Process

Modeling Process | PCA

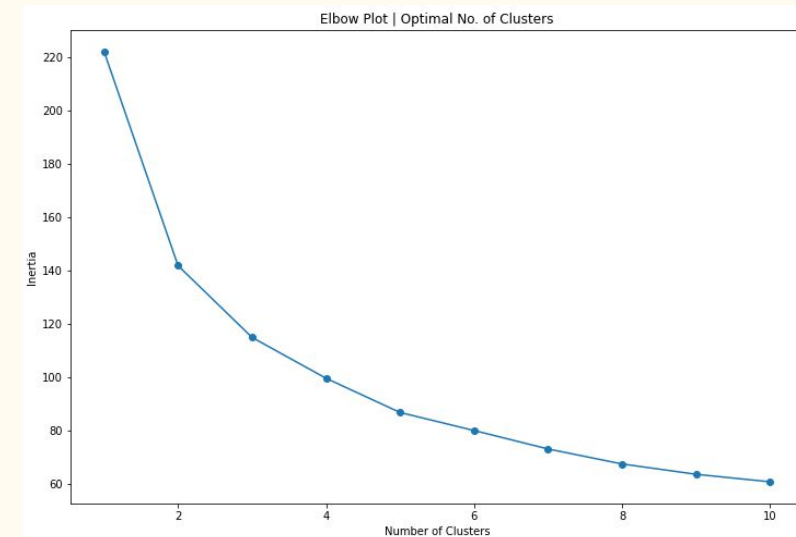
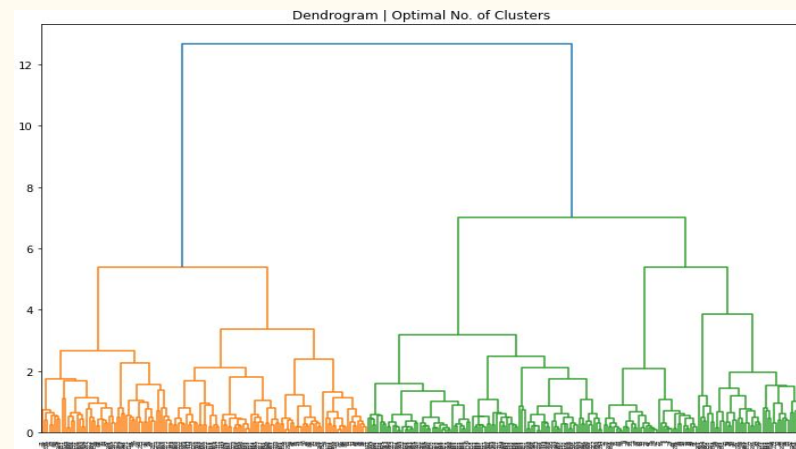
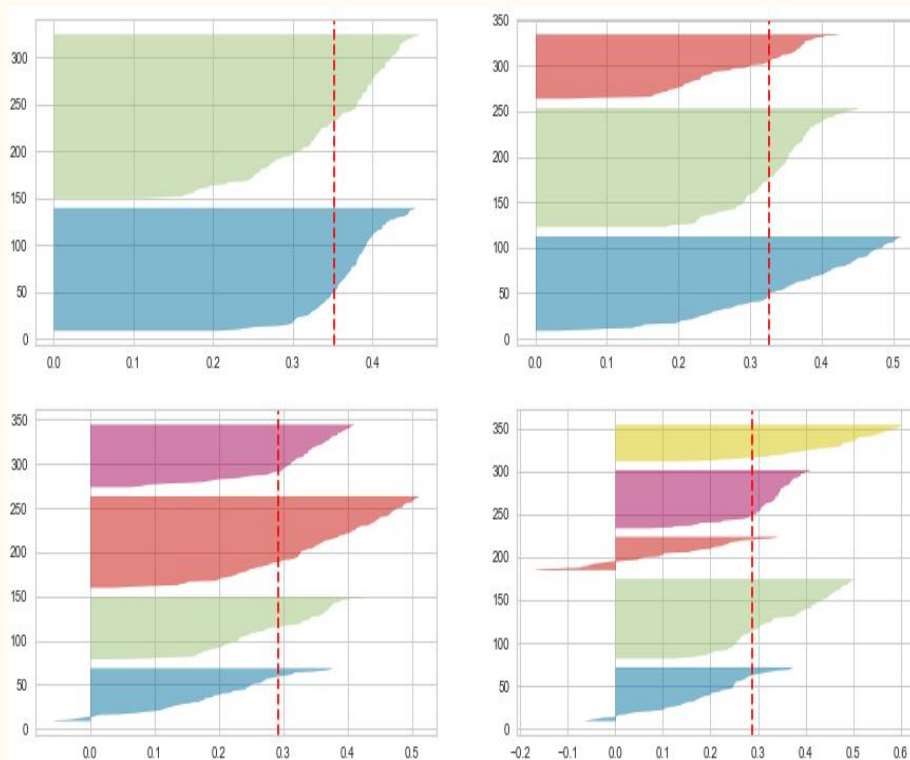


9 Principal Components
Explain 94.29% of the Variance

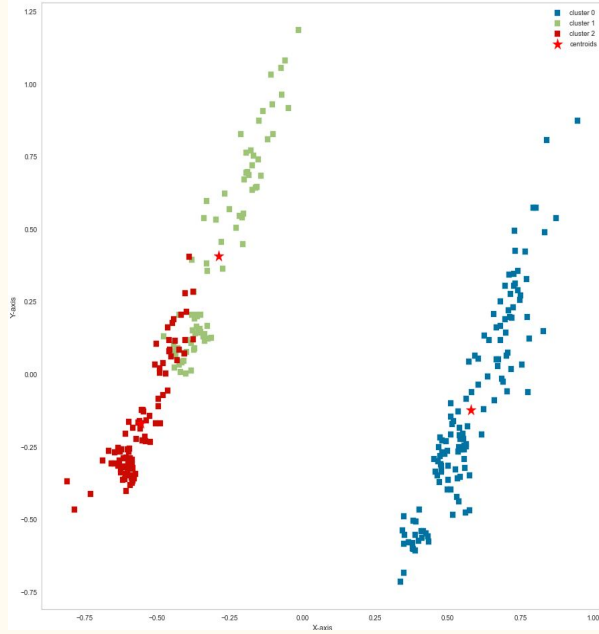
Version Displayed:
Based on a Clustering of a
Combination of All Features

Modeling Process | Evaluation

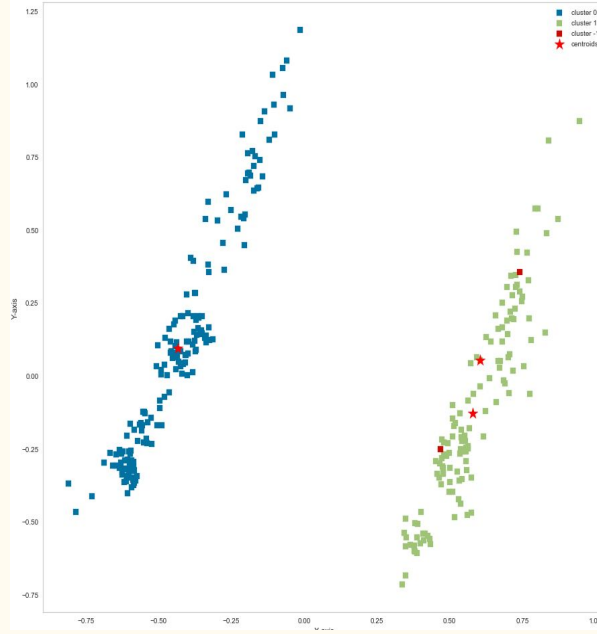
Optimal Clusters for 9 Components by Metric



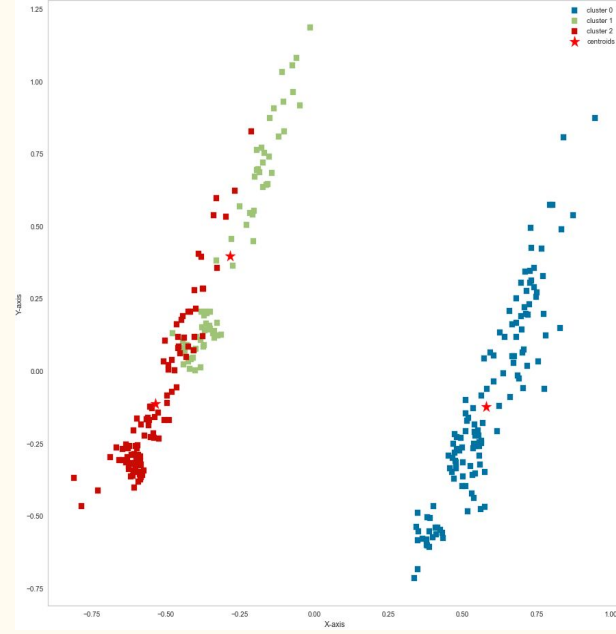
Modeling Process | Clustering



K-Means



DBSCAN

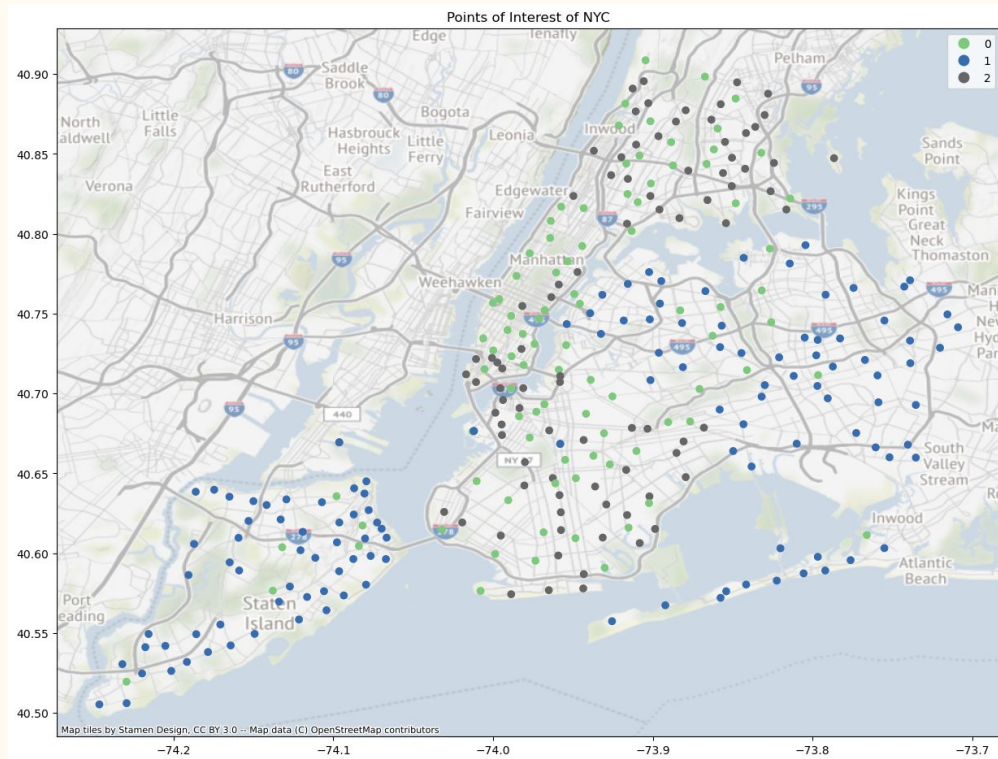


Hierarchical

Results

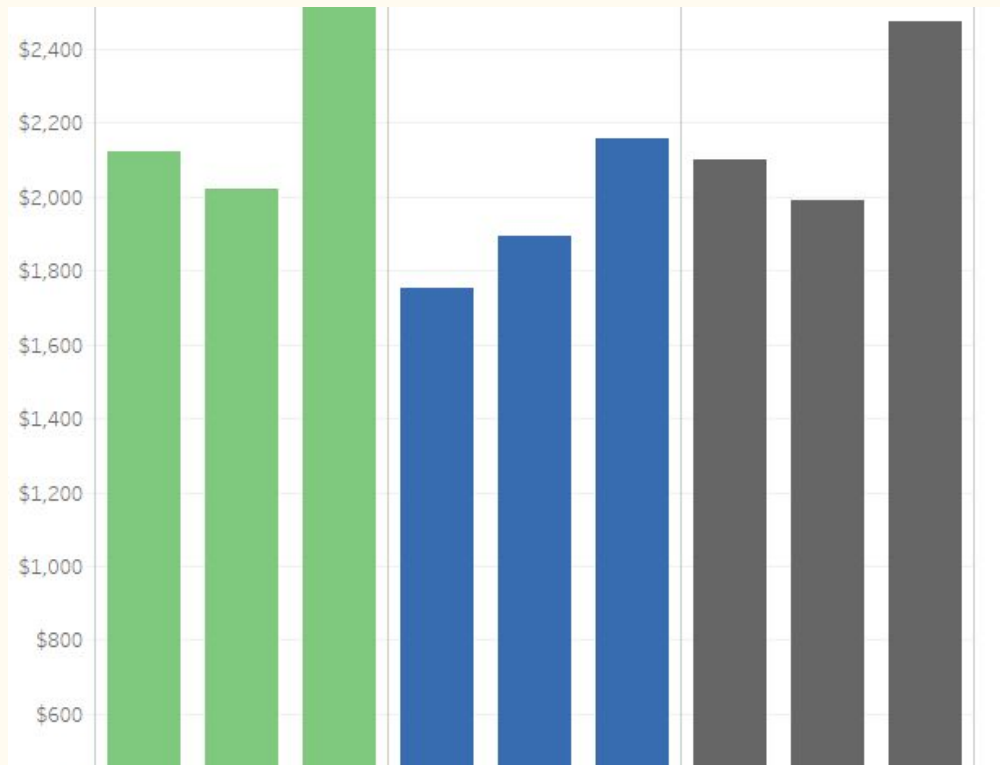
Results (Clustered on Points of Interest)

- **C0:**
 - Studio : \$2124
 - 1 Bed: \$2021
 - 2 Bed: \$2831
 - Restaurants: 75
 - Schools: 26
 - Parks: 18
 - Transit Stations: 32
- **C1:**
 - Studio : \$1751
 - 1 Bed: \$1893
 - 2 Bed: \$2158
 - Restaurants: 31
 - Schools: 5
 - Parks: 1
 - Transit Stations: 7
- **C2:**
 - Studio : \$2099
 - 1 Bed: \$1988
 - 2 Bed: \$2476
 - Restaurants: 26
 - Schools: 2
 - Parks: 1
 - Transit Stations: 1



Results (Clustered on Points of Interest)

- **C0:**
 - Studio : \$2124
 - 1 Bed: \$2021
 - 2 Bed: \$2831
 - Restaurants: 75
 - Schools: 26
 - Parks: 18
 - Transit Stations: 32
- **C1:**
 - Studio : \$1751
 - 1 Bed: \$1893
 - 2 Bed: \$2158
 - Restaurants: 31
 - Schools: 5
 - Parks: 1
 - Transit Stations: 7
- **C2:**
 - Studio : \$2099
 - 1 Bed: \$1988
 - 2 Bed: \$2476
 - Restaurants: 26
 - Schools: 2
 - Parks: 1
 - Transit Stations: 1



Results (Clustered on Rat Sightings & Restaurants)

C0:

- Studio : \$1711
- 1 Bed: \$1957
- 2 Bed: \$2065
- Restaurants: 28
- Restaurant Distance (m): 561m
- Rat Complaint Time (Days): 23

• C1:

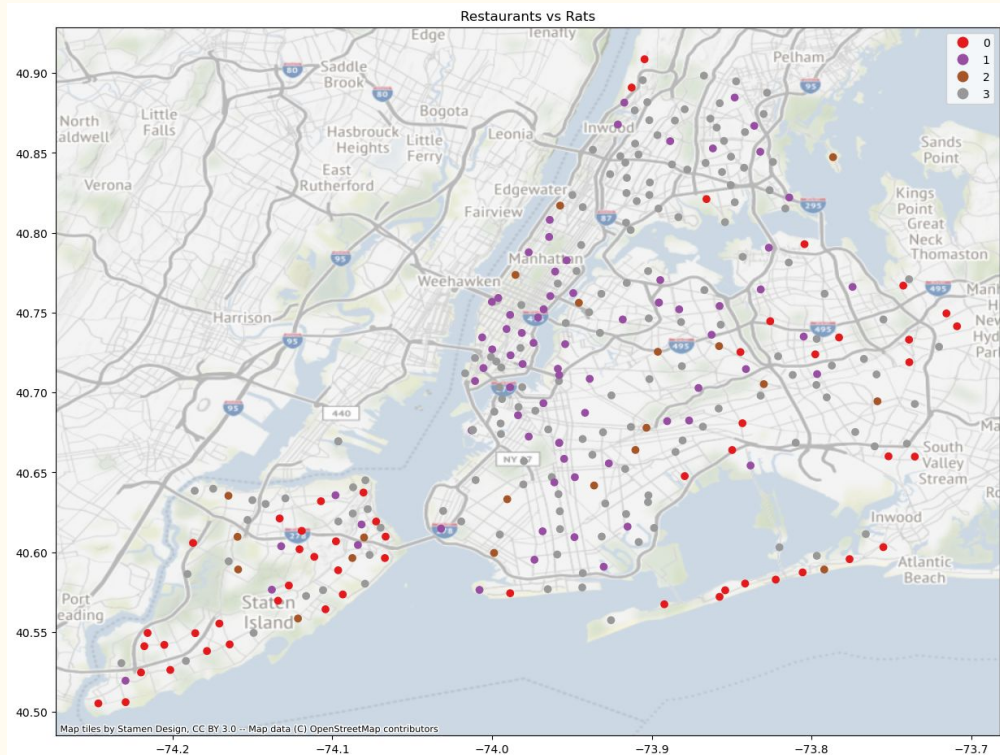
- Studio : \$2096
- 1 Bed: \$2106
- 2 Bed: \$3161
- Restaurants: 76
- Restaurant Distance (m): 658m
- Rat Complaint Time (Days): 11

C2:

- Studio : \$1860
- 1 Bed: \$1779
- 2 Bed: \$1977
- Restaurants: 40
- Restaurant Distance (m): 593m
- Rat Complaint Time (Days): 12

• C3:

- Studio : \$1990
- 1 Bed: \$1932
- 2 Bed: \$2337
- Restaurants: 31
- Restaurant Distance (m): 31m
- Rat Complaint Time (Days): 9



Results (Clustered on Rat Sightings & Restaurants)

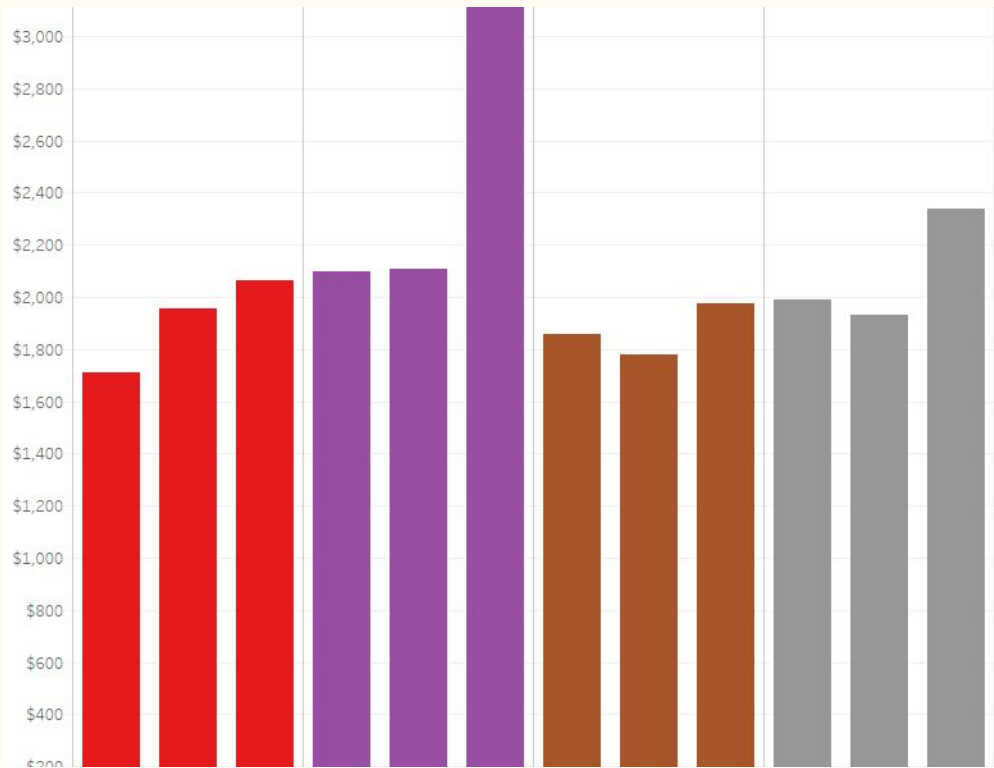
- C0:**

 - Studio : \$1711
 - 1 Bed: \$1957
 - 2 Bed: \$2065
 - Restaurants: 28
 - Restaurant Distance (m): 561m
 - Rat Complaint Time (Days): 23
- **C1:**

 - Studio : \$2096
 - 1 Bed: \$2106
 - 2 Bed: \$3161
 - Restaurants: 76
 - Restaurant Distance (m): 658m
 - Rat Complaint Time (Days): 11
- C2:**

 - Studio : \$1860
 - 1 Bed: \$1779
 - 2 Bed: \$1977
 - Restaurants: 40
 - Restaurant Distance (m): 593m
 - Rat Complaint Time (Days): 12
- **C3:**

 - Studio : \$1990
 - 1 Bed: \$1932
 - 2 Bed: \$2337
 - Restaurants: 31
 - Restaurant Distance (m): 31m
 - Rat Complaint Time (Days): 9



Thank You!

