Hospital Neighborhoods in New York City

IBM Data Science Coursera Capstone Project

6 Sep 2020

Katherine Terlesky

Presentation Overview

- This presentation reviews the data science project performed for the IBM Data Science Capstone Project on Coursera.

- Presentation Outline
  - Business Case – Problem Statement
  - Solution Statement
  - Data Sources
  - Methodology
  - Results
  - Conclusion

Business Case- Problem Statement

- Traveling nurses are healthcare professionals who choose to work in cities with short term nurse staffing needs.
- The benefits of being a traveling nurse are higher pay and the ability to travel around the country exploring cities.
- Challenges with this traveling profession are identification of cities that meet your interests and professional needs.

Solution Statement

- Develop a tool that incorporates ratings of hospitals with neighborhood venue data from Foursquare to help traveling nurses select the best Hospital Neighborhood.

- This tool will help prospective traveling nurses identify hospital neighborhoods with appealing venues within a radius of the hospital.

- This data science project will explore integration of hospital rating data available from the Center for Medicare and Medicaid services with Foursquare venue data.

Data Sources

Presentation of the Data Sources used in the Analysis

How to Evaluate Hospitals?

- In the United States, the Medicare program is a government sponsored health insurance program for adults age 65 and older and some people with disabilities.

- Medicaid is a government-sponsored health program for low-income people.

- The U.S. Gov has developed metrics and requirements hospitals must follow to continue to receive Federal aid.

- The Overall Hospital Rating and 'Safety of Care' Ratings will be used in this analysis.

Hospital Measures Data Source

- The complete data set of all Hospital Measures is available at https://data.medicare.gov/data/hospital-compare.

- Hospitals in the United States (excluding Veterans Affairs and Dept of Defense Hospitals) report data to the Centers for Medicare & Medicaid Services.

- Hospitals in the United States are rated based on 7 areas of Quality including Mortality, **Safety of Care**, Readmission, Patient Experience, Effectiveness of Care, Timeliness of Care, and Efficient Use of Medical Imaging.

- The ratings in these seven areas are combined into a star rating for each hospital (range of 1-5 stars).

Hospital Measures Data Set

Metadata fields used in this analysis

7 Categories of Measurement for Hospital Quality. Only 'Safety of Care' used in this analysis

Hospital Ratings Dataset Fields (from Center for Medicare and Medicaid Services)	
Data Field	**Description and Type**
Facility ID	6 characters
Facility Name	72 characters
Address	51 characters -Street address of hospital
City	20 characters City of hospital
State	2 characters
ZIP Code	8 digit number
County Name	25 character county name
Phone Number	14 character phone number
Hospital Type	34 character type of hospital
Hospital Ownership	43 character type of hospital ownership
Emergency Services	3 character (Yes/No) for Emergency Services
Meets criteria for promoting interoperability of EHRs	1 character (Y/N) for meets interoperability
Hospital overall rating	13 character 1-5 or not available
Hospital overall rating footnote	8 digit number
Mortality national comparison	28 character rating Below, At or Above National average or Not Available
Mortality national comparison footnote	8 digit number
Safety of care national comparison	28 character rating Below, At or Above National average or Not Available
Safety of care national comparison footnote	8 digit number
Readmission national comparison	28 character rating Below, At or Above National average or Not Available
Readmission national comparison footnote	8 digit number
Patient experience national comparison	28 character rating Below, At or Above National average or Not Available
Patient experience national comparison footnote	8 digit number
Effectiveness of care national comparison	28 character rating Below, At or Above National average or Not Available
Effectiveness of care national comparison footnote	8 digit number
Timeliness of care national comparison	28 character rating Below, At or Above National average or Not Available
Timeliness of care national comparison footnote	8 digit number
Efficient use of medical imaging national comparison	28 character rating Below, At or Above National average or Not Available
Efficient use of medical imaging national comparison footnote	8 digit number

1
2
3
4
5
6
7

Foursquare Venue Data

- The Foursquare API will be used to retrieve venue information around the hospitals in New York City. https://foursquare.com/

- The Foursquare data tended to return mostly restaurant data.

- To increase the impact of venues in Arts and & Entertainment and Outdoors, the 'section' constraint was used in the Foursquare call

Methodology

Methodology used in the Analysis

Methodology Phases

- Phase 1 -Exploration and Refinement of the Hospital data set to establish Hospital Neighborhoods
  - Milestone: Definition of the final hospitals with geolocation for the analysis
- Phase 2- Integration of Foursquare venue data into the Hospital Neighborhoods focusing on Arts and Outdoors venues
  - Milestone: Collection of the Foursquare top venues data into a DataFrame
- Phase 3 -Use K-means clustering algorithm to group hospital neighborhoods with like characteristics
  - Milestone: Final recommendation of Hospital Neighborhood for the Use Case

Methodology Details- Data Preprocessing

- Pre-processing was performed on the hospital data set to remove unnecessary columns, hospital types, and hospital data outside of the target area of interest (New York City)
- The data set was evaluated statistically at different pre-processing steps to understand the impact of pre-processing.
- The data set was sequentially narrowed down as follows:
- 5319 hospitals in starting data set
  - 3262 Acute- care hospitals
    - 143 hospitals in the state of New York
      - 36 hospitals in the 5 boroughs of New York City
        - 26 hospitals with a Safety of Care rating of "1".

Methodology Details- Geolocating

- The geolocation service, Nominatum, was used through geopy.
- The addresses of the hospitals in New York City selected for the final analysis were converted to latitude and longitude readings.
- The latitude and longitude of each hospital were used as the center for the Hospital neighborhood for the Foursquare venue calls.

Methodology- Foursquare

- To limit the number of restaurants retrieved in the Foursquare data, the following two variables were defined and used.

- def getNearbyOutdoorsVenues (names, latitudes, longitudes, radius=500, section = 'outdoors', LIMIT = 100):
  - This variable defined the 'outdoors' section as a constraint in the venue retrieval from foursquare.

- def getNearbyArtsVenues (names, latitudes, longitudes, radius=500, section = 'arts', LIMIT = 100):
  - This variable defined the 'arts' section as a constraint in the venue retrieval from foursquare.

Methodology Details- Selection of Machine Learning Method

- Options on Machine Learning involve Supervised or Unsupervised learning.
- Unsupervised learning was selected because there was no training set available.
- Supervised learning results in clusters. Multiple algorithms are available for clustering including Hierarchical, K-Means, and neural networks.
  - K-means was determined to be the most suitable machine-learning algorithm because a defined number of clusters was desired to support the decision making.
  - The data set was narrowed using sequential feature removal. Additional feature removal necessary for hierarchical clustering was not easily possible.
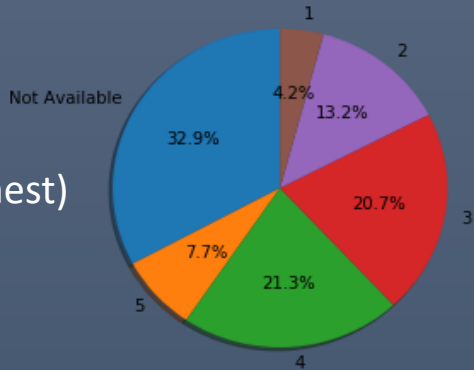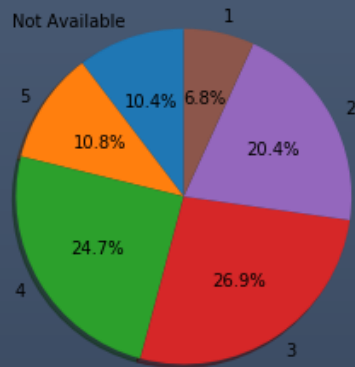
Results

Findings of each phase performed in the Methodology

Distribution of Safety Ratings in Overall Rating

- The Safety of Care Rating was related to the Overall Hospital Rating.
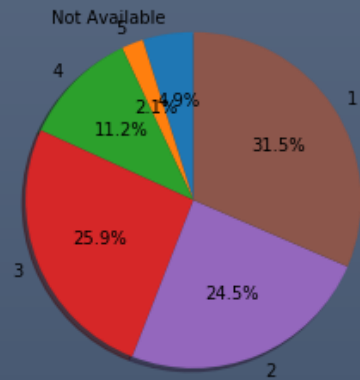- The Safety of Care Ratings "below the national average" were most predominant in hospitals with an overall rating of 1 or 2.

Distribution of Overall & Safety Ratings

Overall Hospital Rating

3262 hospitals

↓

Removed all hospitals not in New York State
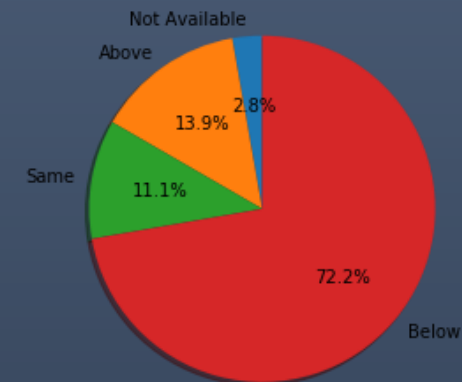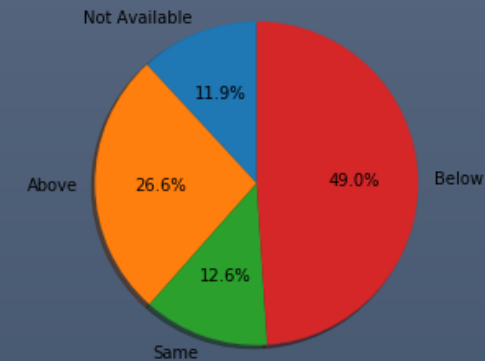
↓

143 NY state hospitals

↓
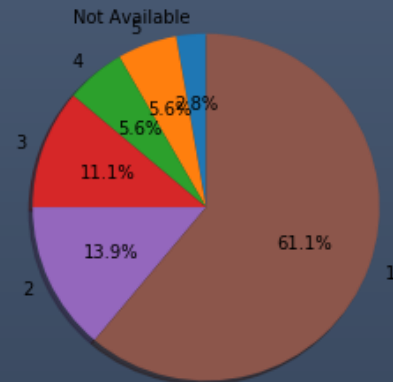
Removed all hospitals not in New York City

↓

36 NY City hospitals

Overall Safety Rating

72% of New York City hospitals have a 'Safety of Care' rating of Below the National Average.
This results in a final hospital set of 26 hospitals for this traveling nurse Use Case.

Final 26 Hospitals in New York City

	Facility	City	Overall Rating	Safety Rating
0	BRONX-LEBANON HOSPITAL CENTER - CONCOURSE DIVI...	BRONX	1	1
1	MONTEFIORE MEDICAL CENTER	BRONX	1	1
2	LINCOLN MEDICAL & MENTAL HEALTH CENTER	BRONX	1	1
3	JACOBI MEDICAL CENTER	BRONX	1	1
4	ST BARNABAS HOSPITAL	BRONX	1	1
5	NEW YORK-PRESBYTERIAN HOSPITAL	NEW YORK	3	1
6	LENOX HILL HOSPITAL	NEW YORK	3	1
7	MOUNT SINAI BETH ISRAEL	NEW YORK	2	1
8	BELLEVUE HOSPITAL CENTER	NEW YORK	1	1
9	NEW YORK UNIVERSITY LANGONE MEDICAL CENTER	NEW YORK	5	1
10	JAMAICA HOSPITAL MEDICAL CENTER	QUEENS	1	1
11	ELMHURST HOSPITAL CENTER	QUEENS	1	1
12	FLUSHING HOSPITAL MEDICAL CENTER	QUEENS	1	1
13	ST JOHN'S EPISCOPAL HOSPITAL AT SOUTH SHORE	LONG ISLAND	1	1
14	BROOKLYN HOSPITAL CENTER - DOWNTOWN CAMPUS	BROOKLYN	1	1
15	MAIMONIDES MEDICAL CENTER	BROOKLYN	1	1
16	NYC HEALTH + HOSPITALS/CONEY ISLAND	BROOKLYN	1	1
17	KINGSBROOK JEWISH MEDICAL CENTER	BROOKLYN	1	1
18	KINGS COUNTY HOSPITAL CENTER	BROOKLYN	1	1
19	WYCKOFF HEIGHTS MEDICAL CENTER	BROOKLYN	1	1
20	BROOKDALE HOSPITAL MEDICAL CENTER	BROOKLYN	1	1
21	NEW YORK-PRESBYTERIAN/BROOKLYN METHODIST HOSPITAL	BROOKLYN	3	1
22	SUNY/DOWNSTATE UNIVERSITY HOSPITAL OF BROOKLYN	BROOKLYN	1	1
23	INTERFAITH MEDICAL CENTER	BROOKLYN	1	1
24	RICHMOND UNIVERSITY MEDICAL CENTER	STATEN ISLAND	1	1
25	STATEN ISLAND UNIVERSITY HOSPITAL	STATEN ISLAND	1	1

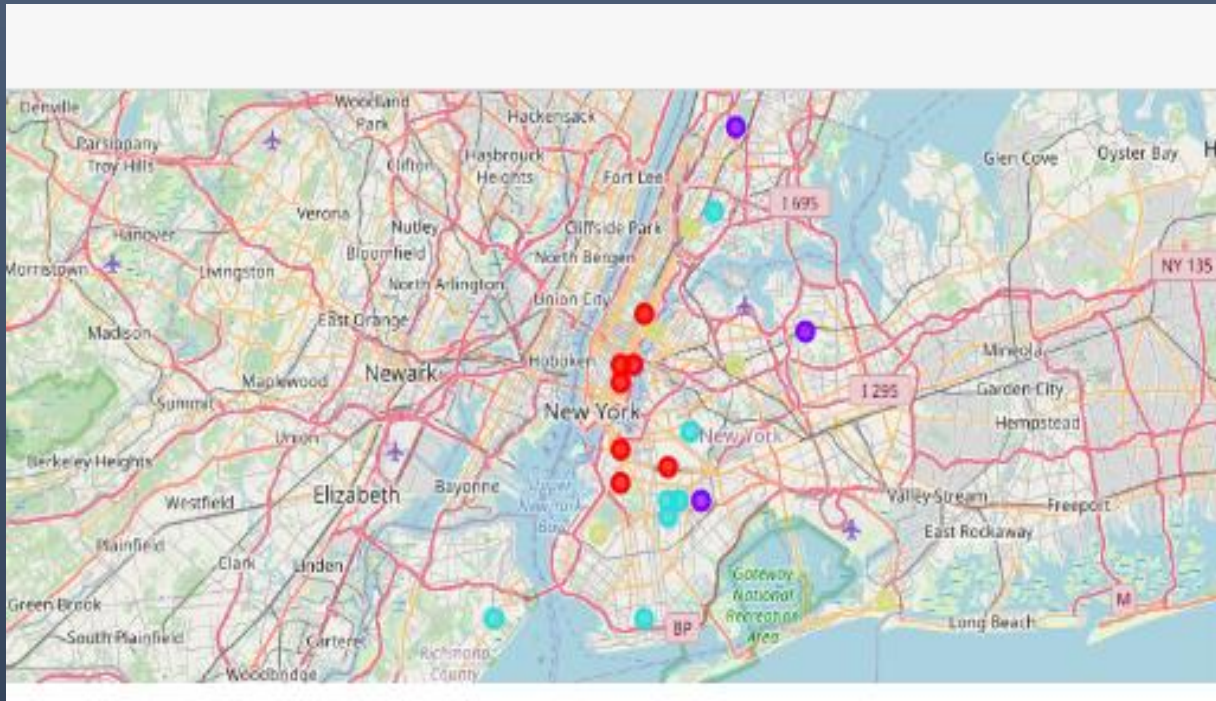# Mapping of the 26 Hospital Lat/Long using Folium
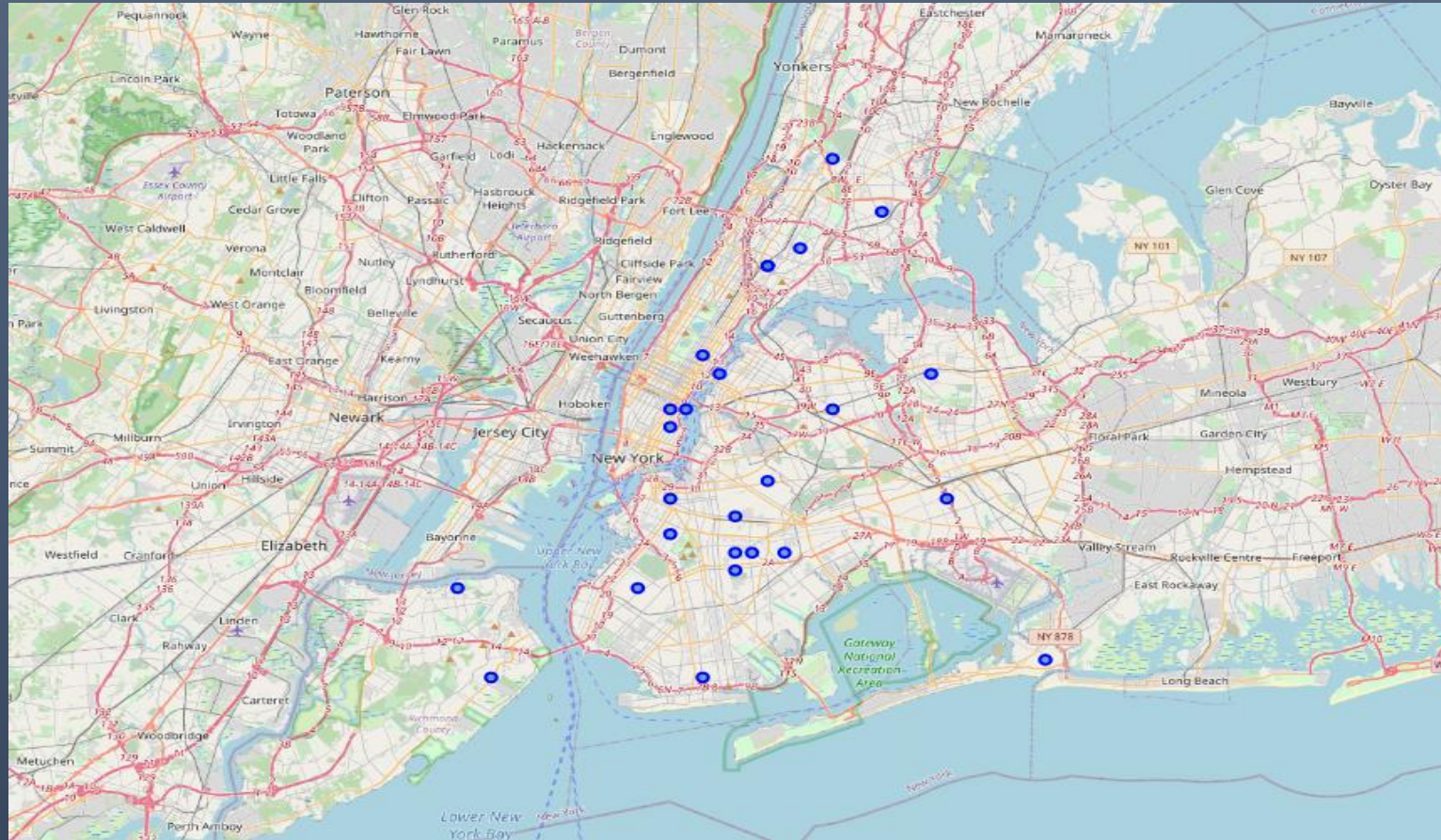
- Hospital locations are shown with a blue circle on the map.

K-means Clustering

- 4 clusters selected
- Clusters mapped using Folium

- Cluster 0 = 7 hospitals, Red circles
- Cluster 1 = 3 hospitals, Purple circles
- Cluster 2 = 7 hospitals, Blue circles
- Cluster 3 = 6 hospitals, Yellow circles

Characteristics of Cluster 0

Cluster 0 (Red) is the Downtown living Cluster with high density of Arts & Entertainment.

Cluster 0 = Downtown living Cluster with high density of Arts & Entertainment					
Number	Hospital Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
6	LENOX HILL HOSPITAL	Art Gallery	Gym / Fitness Center	Theater	Gym	Art Museum
7	MOUNT SINAI BETH ISRAEL	Art Gallery	Theater	Playground	Indie Theater	Gym / Fitness Center
8	BELLEVUE HOSPITAL CENTER	Theater	Art Gallery	Gym / Fitness Center	Gym	Comedy Club
9	NEW YORK UNIVERSITY LANGONE MEDICAL CENTER	Arts & Entertainment	Art Gallery	Outdoor Sculpture	Scenic Lookout	Garden
14	BROOKLYN HOSPITAL CENTER - DOWNTOWN CAMPUS	Gym / Fitness Center	Theater	Performing Arts Venue	Dance Studio	Opera House
21	NEW YORK-PRESBYTERIAN/BROOKLYN METHODIST HOSPITAL	Art Gallery	Gym	Theater	Yoga Studio	Gym / Fitness Center
23	INTERFAITH MEDICAL CENTER	Art Gallery	Museum	Theater	Playground	Park

Characteristics of Cluster 1

Cluster 1 (Purple) is Cluster 1 (purple) is High Park density over Arts & Entertainment

Cluster 1 = High Park density over Arts & Entertainment						
Number	Hospital Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	MONTEFIORE MEDICAL CENTER	Park	Dance Studio	History Museum	Performing Arts Venue	Museum
12	FLUSHING HOSPITAL MEDICAL CENTER	History Museum	Plaza	Park	Public Art	Yoga Studio
20	BROOKDALE HOSPITAL MEDICAL CENTER	Art Gallery	Performing Arts Venue	Park	Pool	Music Venue

Characteristics of Cluster 2

Cluster 2 (light blue) is the Music Cluster with balanced access to Outdoors venues

Number	Hospital Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
	Cluster 2 = Music Cluster with balanced access to Outdoors venues					
0	BRONX-LEBANON HOSPITAL CENTER - CONCOURSE DIVI...	Park	Movie Theater	Music Venue	Playground	Gym
16	NYC HEALTH + HOSPITALS/CONEY ISLAND	Concert Hall	Performing Arts Venue	Music Venue	Playground	Gym / Fitness Center
17	KINGSBROOK JEWISH MEDICAL CENTER	Music Venue	Yoga Studio	Gym / Fitness Center	Dance Studio	Beach
18	KINGS COUNTY HOSPITAL CENTER	Music Venue	Dance Studio	Gym	Playground	Park
19	WYCKOFF HEIGHTS MEDICAL CENTER	Gym	Music Venue	Art Gallery	Park	Circus
22	SUNY/DOWNSTATE UNIVERSITY HOSPITAL OF BROOKLYN	Music Venue	Museum	Gym	Dog Run	Lake
25	STATEN ISLAND UNIVERSITY HOSPITAL	Dance Studio	Yoga Studio	Park	Gym	Scenic Lookout

Characteristics of Cluster 3

Cluster 3 (light yellow) is the Art Gallery Cluster with balanced access to outdoors recreation

Cluster 3 = Art Gallery Cluster with balanced access to outdoors recreation						
Number	Hospital Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	LINCOLN MEDICAL & MENTAL HEALTH CENTER	Art Gallery	Movie Theater	Gym	Recreation Center	Baseball Field
4	ST BARNABAS HOSPITAL	Art Gallery	Movie Theater	Gym	Recreation Center	Baseball Field
5	NEW YORK-PRESBYTERIAN HOSPITAL	Art Gallery	Indie Theater	Gym	Baseball Field	Park
11	ELMHURST HOSPITAL CENTER	Art Gallery	Music Venue	Martial Arts School	Park	Playground
13	ST JOHN'S EPISCOPAL HOSPITAL AT SOUTH SHORE	Art Gallery	Martial Arts School	Park	Cycle Studio	Beach
15	MAIMONIDES MEDICAL CENTER	Art Gallery	Dance Studio	Martial Arts School	Park	Skate Park

Conclusions

- The 4 clusters had representation across more than 1 NYC borough.
- Using 4 clusters resulted in clusters with unique characteristics that supported the final decision process.
- The traveling nurse for this use case selected to investigate open travelling nurse positions for hospitals in Cluster 3, with Cluster 0 as second choice.
- This analysis and visualization allowed this nurse to narrow down their decision.