BATTLE OF THE NEIGHBORHOODS

USING FOURSQUARE’s DATA IN TIMES OF COVID 19

Roberto Arriaga Omacell

August, 2020

GETTING TO THE PROBLEM

EXECUTIVE SUMMARY:

DURING 2020, COVID 19 EPIDEMIC HAS BECOME A GLOBAL CONCERN.

MANY COUNTRIES HAVE IMPLEMENTED DIFFERENT MEASURES TO TRY TO STOP CONTAGION, MAINLY QUARANTINE AND SOCIAL DISTANCING.

ACCESS TO HEALTH CARE VENUES, SUCH AS HOSPITALS AND PHARMACIES, IS CRITICAL FOR ANY POPULATION.

DURING THE FIRST STAGES OF THE PANDEMIC, IN THE US, THE CITY OF NEW YORK WAS ONE OF THE MOST AFFECTED, AND HAVE BEEN WORKING TO GET THINGS UNDER CONTROL.

SO THE QUESTION IS…?

- Can we use data from FOURSUARE, a platform mostly focused on categories related to entertainment, leisure and shopping, to evaluate if a neighborhood has low or high access to healthcare venues?.

DATA

And scope delimitation.

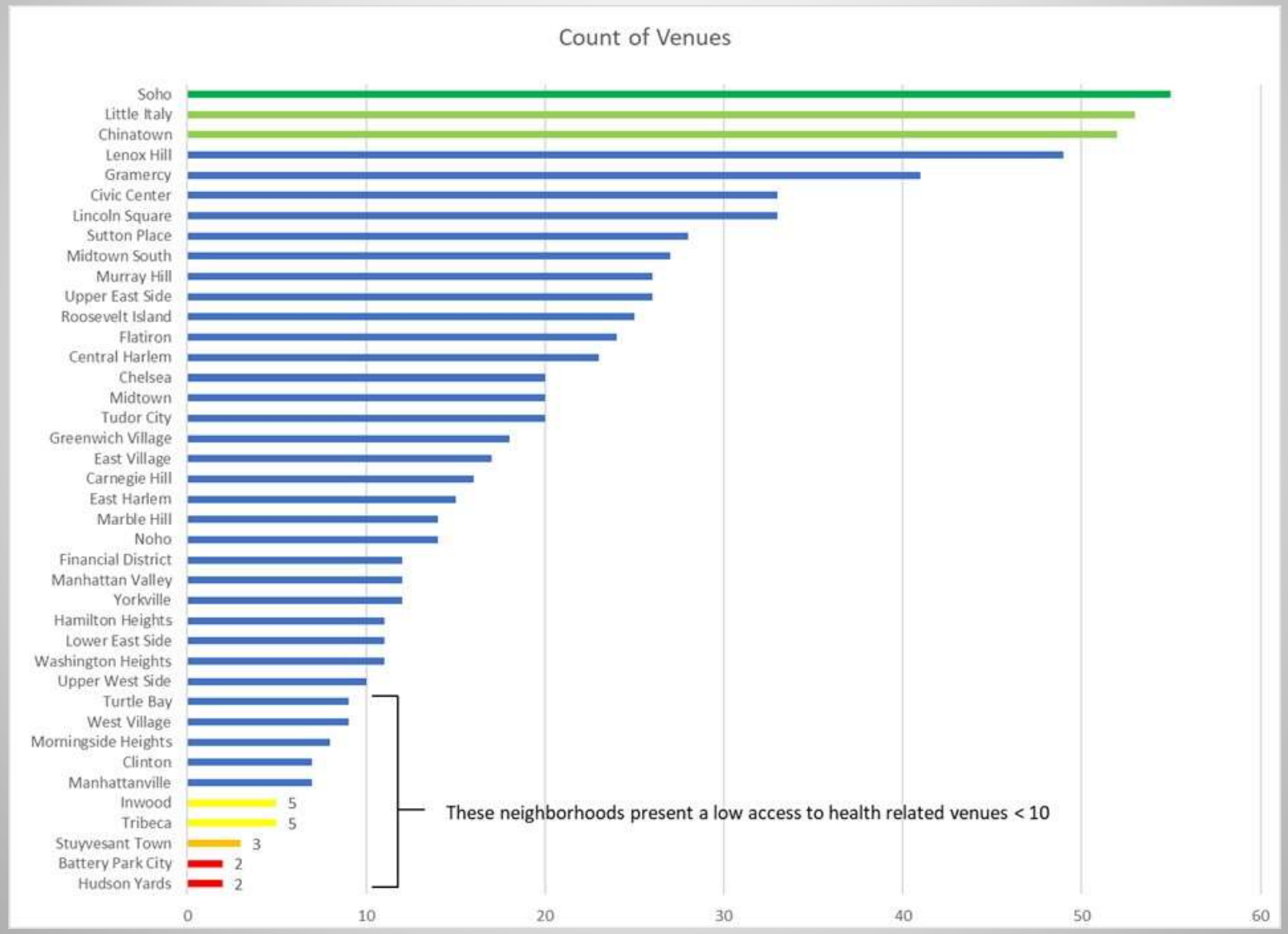
- This project will use information provided by **FOURSQUARE** regarding venues with tags like **'hospital'** (both, public and private) and **‘pharmacy’**.(≡)
- The focus will be New York City, USA; and using **Manhattan** for the first sample of neighborhoods (40).
- Location information for the neighborhoods will be obtained from the NYU Spatial Data Repository (2014)

(≡) After a initial review of the site, tags like 'clinic' although related to health, were mostly associated with cosmetic and even veterinary, so will be out of this scope

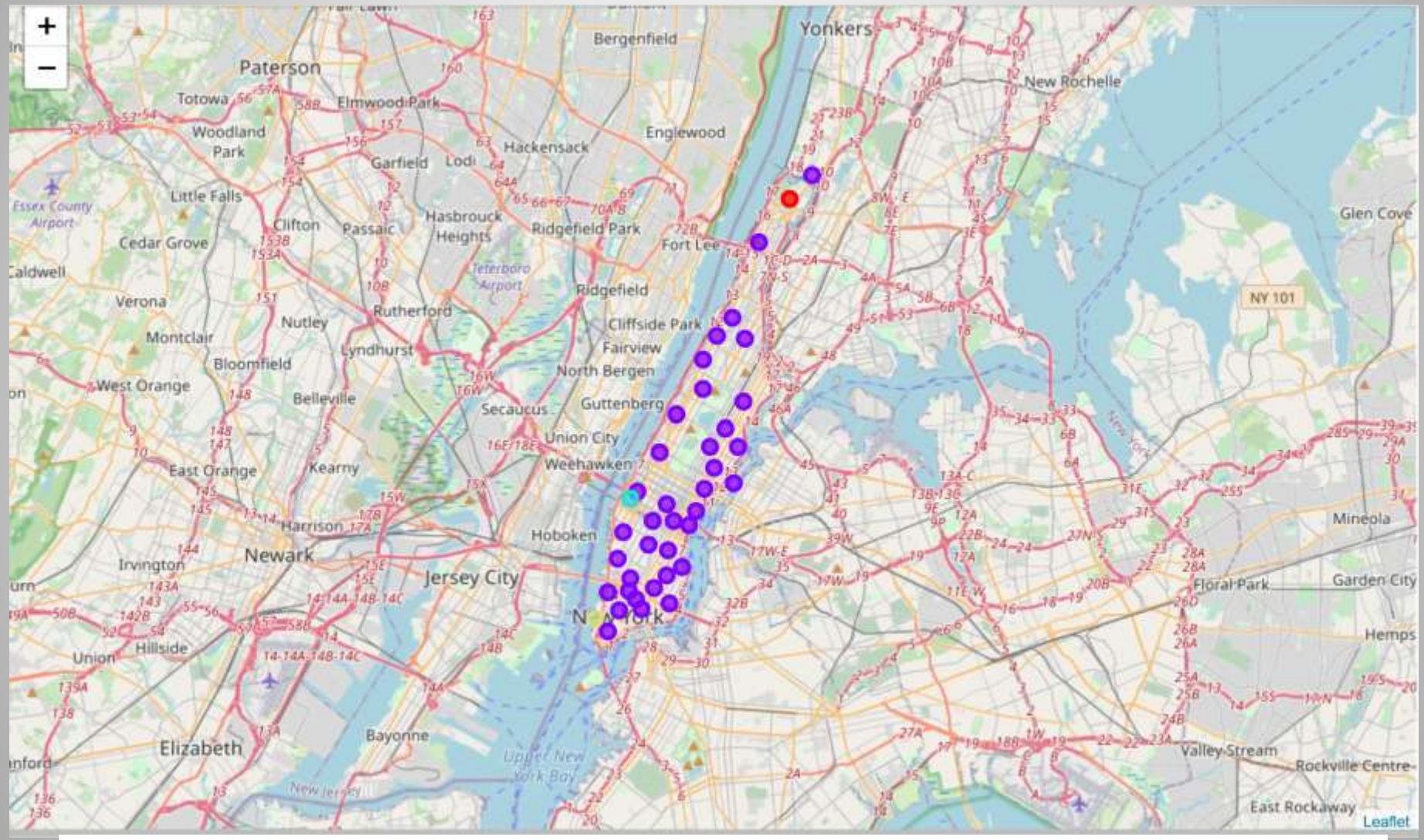
METHODOLOGY

- Initial Data
  - Data for NY City containing Borough's Name, Neighborhood and location (latitude and longitude) were loaded from json file.
  - After transforming to a panda, a slice for Borough=Manhattan was created.
  - With the location information of Manhattan's neighborhoods, FOURSQUARE was queried for venues using categories 'hospital' and 'pharmacy', radius = 500, limit = 100.
- Exploration
  - The venue data was assigned to their respective neighborhoods, counted and estimated frequency.
- Clustering
  - K-means technique was used, with k = 4.
  - The resulting clusters were mapped and each analyzed.

RESULTS

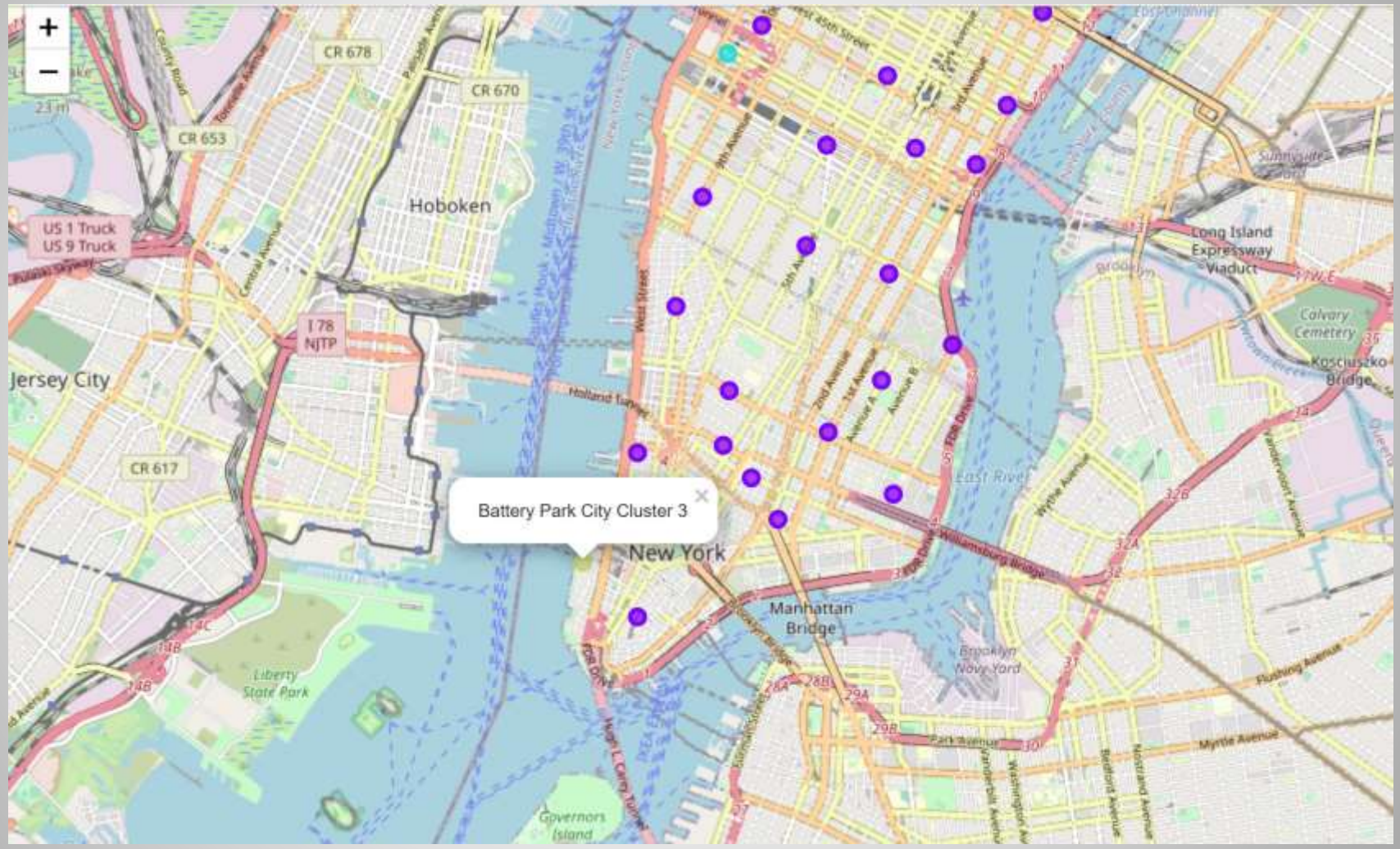

Count of Venues
Soho
Little Italy
Chinatown
Lenox Hill
Gramercy
Civic Center
Lincoln Square
Sutton Place
Midtown South
Murray Hill
Upper East Side
Roosevelt Island
Flatiron
Central Harlem
Chelsea
Midtown
Tudor City
Greenwich Village
East Village
Carnegie Hill
East Harlem
Marble Hill
Noho
Financial District
Manhattan Valley
Yorkville
Hamilton Heights
Lower East Side
Washington Heights
Upper West Side
Turtle Bay
West Village
Morningside Heights
Clinton
Manhattanville
Inwood 5
Tribeca 5
Stuyvesant Town 3
Battery Park City 2
Hudson Yards 2
These neighborhoods present a low access to health related venues < 10
0
10
20
30
40
50
60

RESULTS



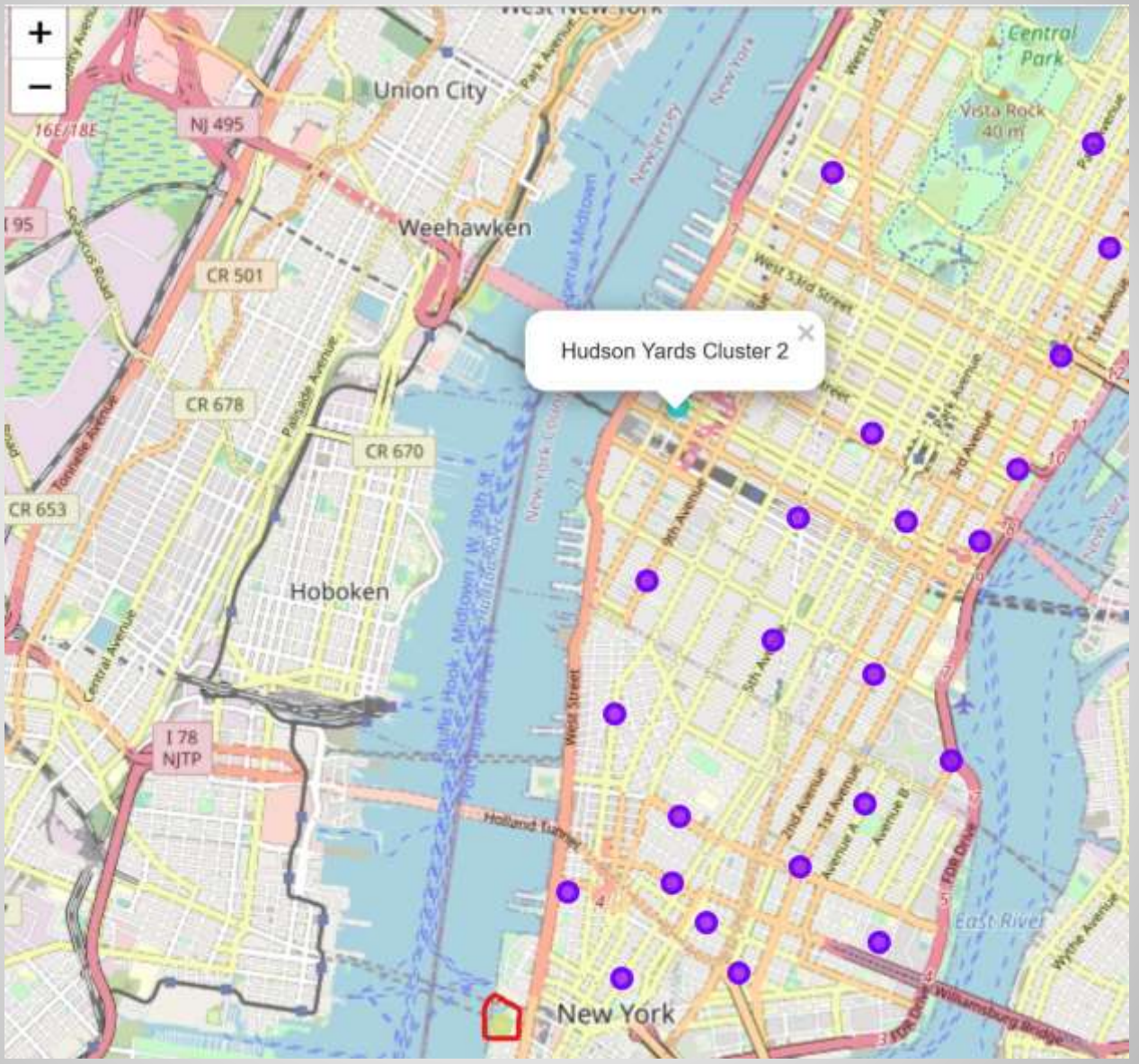
General Map of Manhattan showing clustered neighborhoods using K-Means

RESULTS



Detail of cluster 3: Battery Park City

RESULTS



Detail of cluster 2 Hudson Yards,
with Battery Park highlighted in the south

DISCUSSION

It can be seen that some of the venues are related to veterinaries (animal) and there are some 'hospitalities', this may be product of stating initial query limit of 100 venues [100 venues*40 neighborhoods*2 categories = 8000 possible, and near 519 were found, so the depth of the query was very high]. Instead of eliminating them, each case was analyzed in context.

Soho, **Little Italy** and **Chinatown** are neighborhoods with ***high access*** to hospitals and pharmacies (>50 venues).

Battery Park City and **Hudson Yards** are the neighborhoods with the ***least access*** to healthcare facilities (each have 2 venues, but actually just a pharmacy each, as the other venue is an animal hospital).

DISCUSSION

K-means included **Hudson Yards** as cluster 2; and **Battery Park City** as cluster 3, and definitely these two present interesting characteristics as being the ones with less access, both are coastal and really just 1 pharmacy. It is possible that an algorithm such as DBSCAN would have clustered these two together.

Inwood, or cluster 0, has also a relatively low access (5 venues), but they are all in the main category of *hospitals* or *pharmacies*

By increasing k, eventually a set of clusters with neighborhoods from the lowest to the highest access, could be obtained.

CONCLUSION

Or the answer to the question

FOURSQUARE (FS) data can be used to estimate the ‘accessibility’ of neighborhoods to healthcare venues, at least, with some considerations:

- Some of the data in **FS** is tagged ‘ambiguously’ or wrong, and deep searches could bring these values, producing undesired noise.
- There are other platforms more focused on the Health field with more current data.
- K-means worked well identifying 2 interesting data points.

THANK YOU!

And Keep Safe