A Place to rent in New York

Applied Data Science Capstone Project

Introduction/ Business Problem

- A project of self interest for my own purpose
- Incoming freshman in New York who couldn't secure in-campus dorm, therefore, looking for a good place to rent outside of campus.
- Any people who are looking to move to or around New York might be interested in similar project.

Data acquisition

- New York crime data from NYC OpenData website https://data.cityofnewyork.us/api/views/5uac-w243/rows.csv
- Foursquare API https://api.foursquare.com/v2/venues/
- New York Neighborhood geojson - https://github.com/veltman/snd3/blob/master/data/nyc-neighborhoods.geo.json

Data Cleaning

New York Crime Data

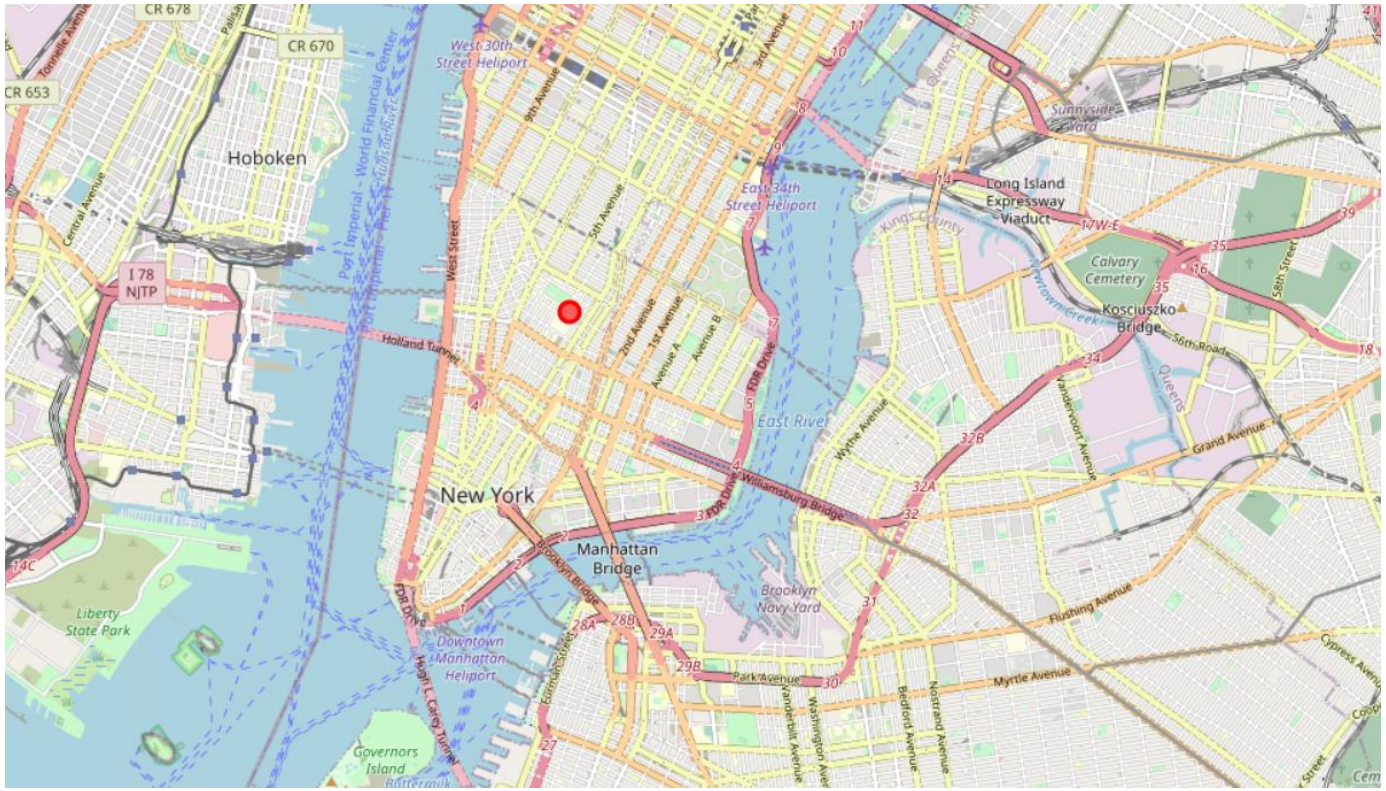
- New York crime data has a total of 306656 rows and 36 features.
- 200 rows of data are selected randomly
- Only 200 rows and 3 attributes remains after cleaning.
- The remaining 3 attributes are Borough name, Latitude and Longitude.

Foursquare API Venue data

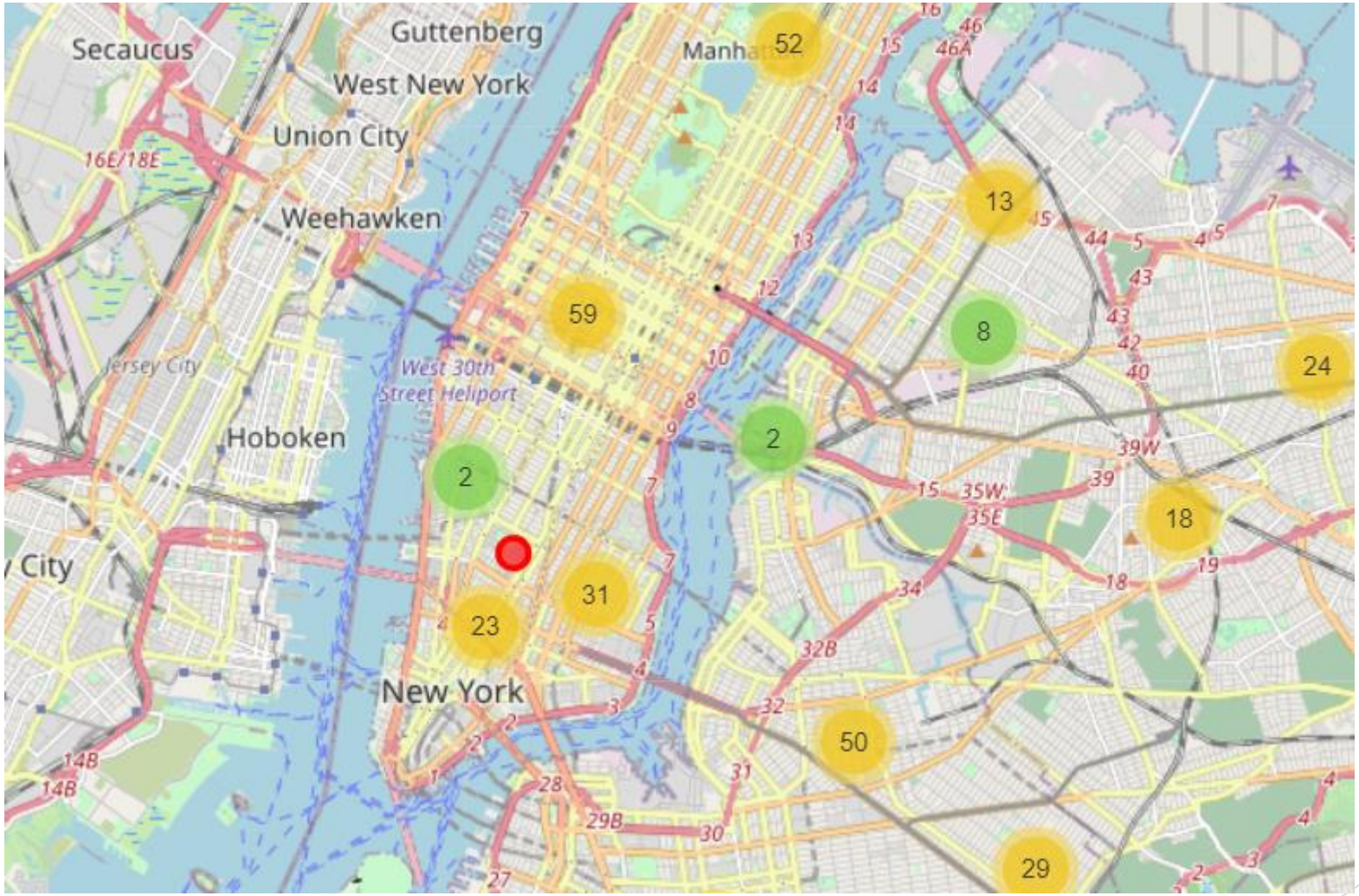
- By making a get request for venue search from Foursquare API we can get a json containing information data about the desired searched venues.
- Json can be made into pandas dataframe by using json_normalize() function.
- The dataframe contains 25 attributes and after cleaning only 6 will remain.
- The remaining 6 are name, categories, latitude, longitude, distance from search point and Address.

Methodology

- Using folium library, data are visualized on the map of New York.
- The latitude and longitude of a specific location is found with the help of geopy library. Like the red circle below which shows the location of NYU.



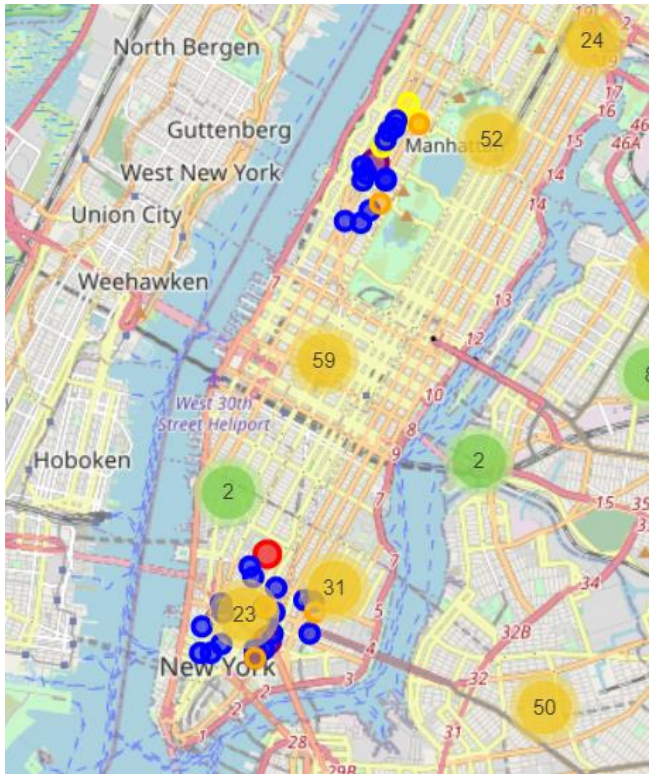
- Then, the clustered objects of crime data points are imposed onto the map.



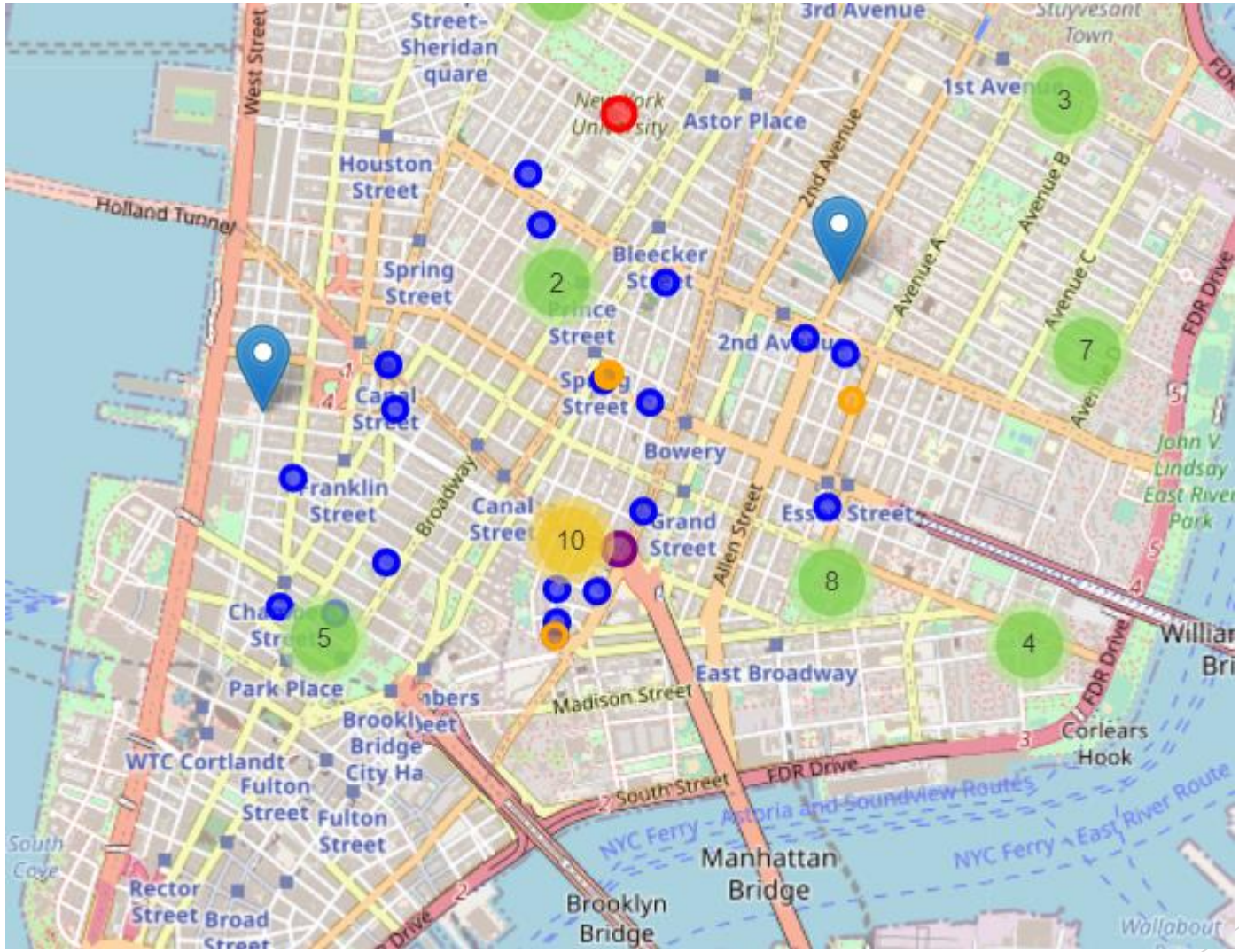
- By zooming in and out, the clustered object will dispersed as following, by looking at these, a general idea of crime rate in specific area can be seen.
- Area of lower level crime rate is chosen.

Result

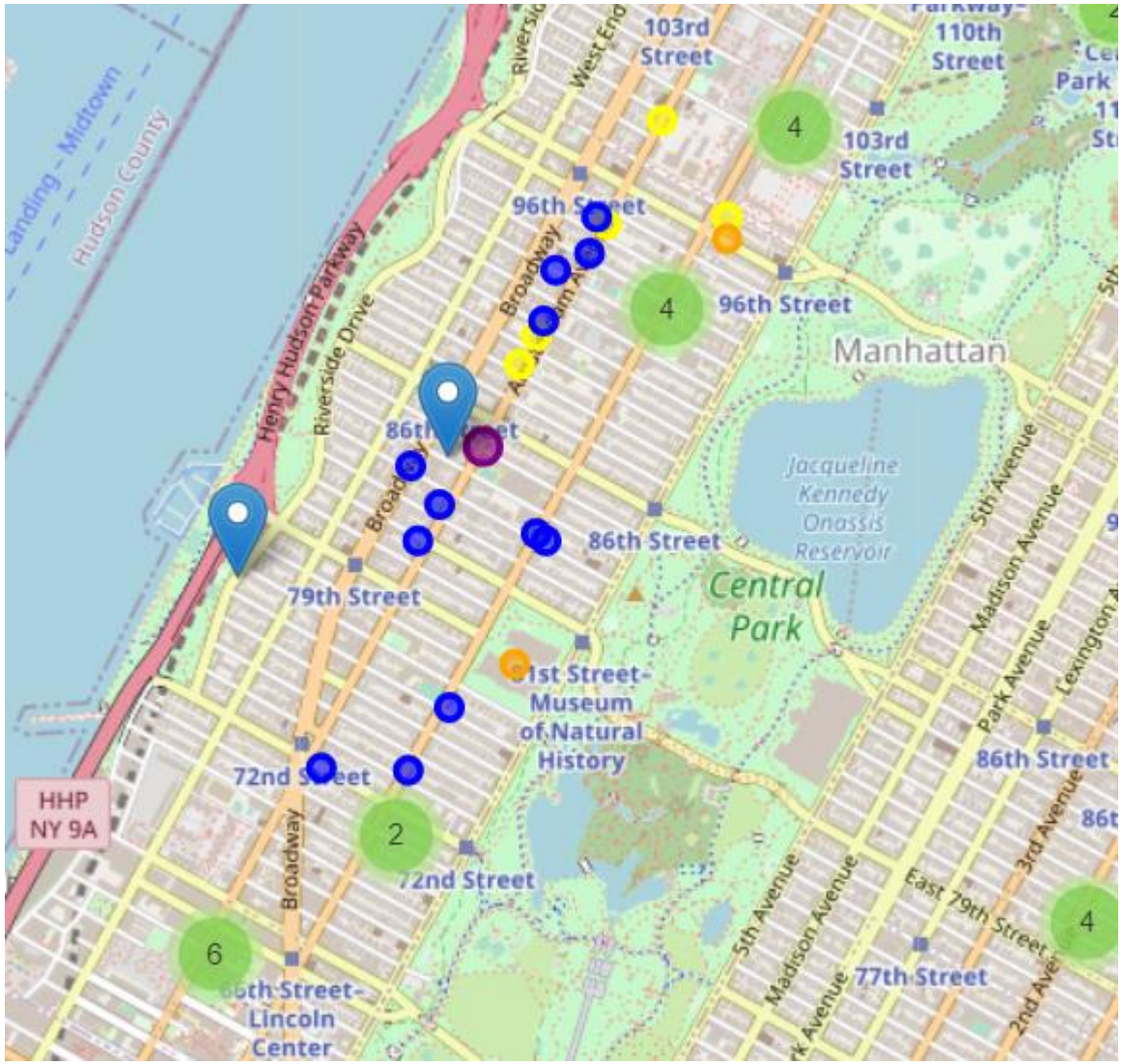
- After investigating, Chinatown and Upper West Side were chosen as choices for area to look for places to rent. They are shown as purple circles.
- Further investigation is done by looking for nearby venues such as restaurants and arcades in both area. Yellow circles for Chinese restaurants, blue circles for Japanese restaurants and orange circles for arcades.



- Data around Chinatown

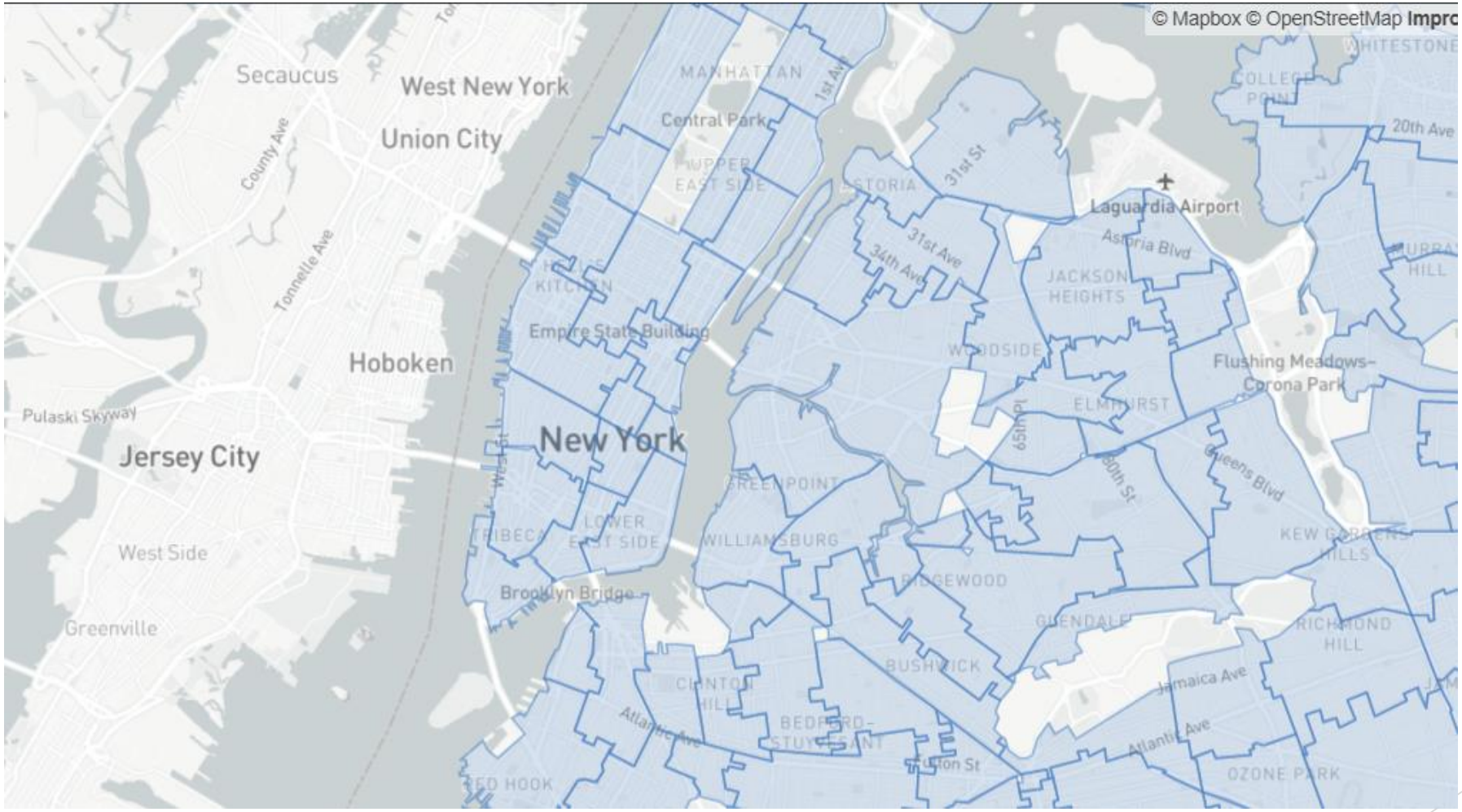


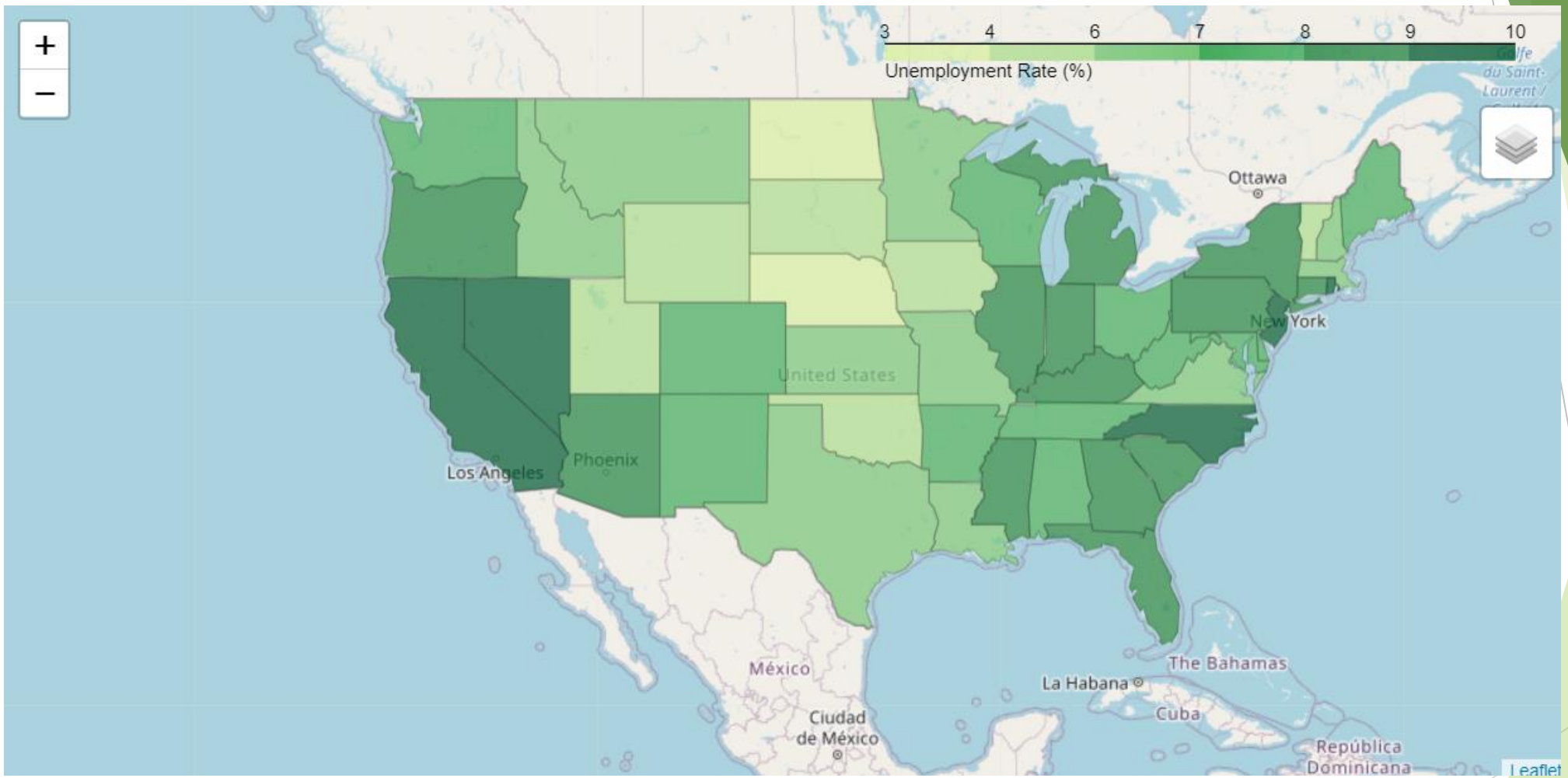
- Data around Upper West Side



Discussion

- Crime rate can be shown in choropleth map which can utilize all of 306656 rows of data, which is better for visualization and way more accurate than choosing 200 rows of data randomly.
- But, I wasn’t able to execute this due to some errors.
- Example of geojson boundary layer of New York and choropleth map of unemployed rate will be shown in next slides.


© Mapbox © OpenStreetMap
Secaucus
West New York
Union City
Hoboken
Jersey City
New York
Central Park
MANHATTAN
Empire State Building
Laguardia Airport
Flushing Meadows–
Corona Park
Brooklyn Bridge
West Side
Greenville
Pulaski Skyway
County Ave
Tonnelle Ave
1st Ave
31st St
31st Ave
34th Ave
Astoria Blvd
20th Ave
Queens Blvd
Jamaica Ave
Atlantic Ave
JACKSON HEIGHTS
ELMHURST
WOODSIDE
ASTORIA
LONG ISLAND CITY
WILLIAMSBURG
GREENPOINT
BUSHWICK
OZONE PARK
RICHMOND HILL
KEW GARDENS HILLS


3
4
6
7
8
9
10
Unemployment Rate (%)
Ottawa
New York
United States
Phoenix
Los Angeles
México
Ciudad
de México
La Habana
The Bahamas
Cuba
República
Dominicana
Leaflet

Conclusion

- Chinatown is closer to university and has more variety of venues nearby.
- Upper West side is safer compared to Chinatown.