Introduction/ Business Problem

The project is thought out for my own purpose with less thought about other audience and business side of things. Assuming that I am an incoming international student at New York University and it’s unfortunate that I didn’t get to secure myself an in-campus student dorm so I have to rent a place and live outside of campus. When trying to look for a place to rent, there are several things that I’d consider. The first thing to think of is the crime rate of the area. Secondly, I am an Asian and a foodie so I’d love if there are good Japanese restaurants and Chinese restaurants nearby that’d be convenient for me to go or get fast delivery. At last, I love to play arcade games and go there regularly so I’d want to have an arcade nearby. Not just arcades and restaurants, any kind of venues can be searched for from Foursquare API, such as café, theme parks and many more. There might be university students around New York just like me having similar thoughts about nearby venues when trying to look for a place to rent that might be interested in this kind of data analysis. Not just freshman but any people or family looking to move or rent a place in New York might also be interested in the project, not exactly being interested in the restaurants and arcades, but things like transport, groceries stores, shopping mall, metro which data could all be obtained from Foursquare API and last but not least, the price index of the neighborhoods.

Data

There are two data sources that will be used, first is the crime rate data of New York from NYC OpenData website,” https://data.cityofnewyork.us/api/views/5uac-w243/rows.csv”. The Neighborhood boundaries data layer is a combination of zoning neighborhood boundaries, zip code boundaries and 2010 Census tract boundaries. There are a total of 36 attributes in the dataset such as date of occurrence, type of offence, level of offence and a lot more but only three of them will be used in the project which are borough of crime occurrence, latitude and longitude of occurrence. And secondly, using Foursquare API, we can look for data of nearby venues such as restaurants and arcades, along with their data such as their names, ratings, website, exact location, tips given by other users and a lot more. The data will be obtained as json format and for convenience we can change it into a pandas dataframe.

Lastly, it is worth mentioning that geojson for boundaries of New York neighborhoods was also found from <https://github.com/veltman/snd3/blob/master/data/nyc-neighborhoods.geo.json> but due to unforseen circumstances like not being able to add the boundary layer to the folium, it isn’t used in the project.

Methodology

All the analysis is done through visualization of data points of the map using folium library. First and foremost, location of NYU is added onto the map of New York in red circle. Next, 1000 data are selected randomly from a collection of 306656 data in total since it’s too much data points to visualize on the map. Then, these 1000 data are clustered together to have a better visualization and a general idea on crime rate of neighborhoods in New York can be seen. The crime clusters can be seen in the screenshot below with bigger green and yellow circles with numbers on them, showing number of crime data points in that area. Then, depending on the crime rate visualization made, a few places are chosen depending on whether they have lower level of crime rate happening, to look for further analysis, the said locations are pinpointed using purple circles onto the map. Then, interested venues are looked for with a radius of 1km around the purple circle point locations, just a comfortable distance of walk. The venues that are searched were Chinese restaurants with yellow circle, Japanese restaurant with blue circle, arcades with orange circle added onto the folium map. More locations and more venues can be looked for if desired. Then, with the visualization made this way, we can decided which particular area we will want to look for a place to rent.

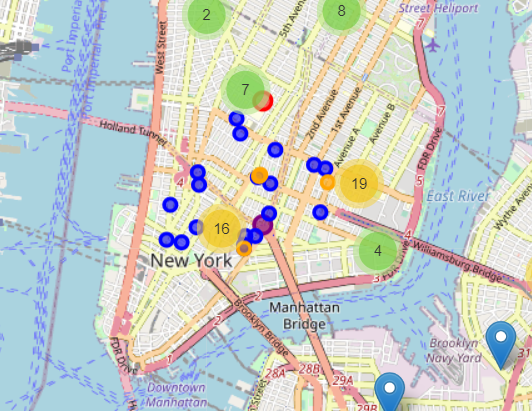


Figure 1- Data Visualization for the project

Discussion

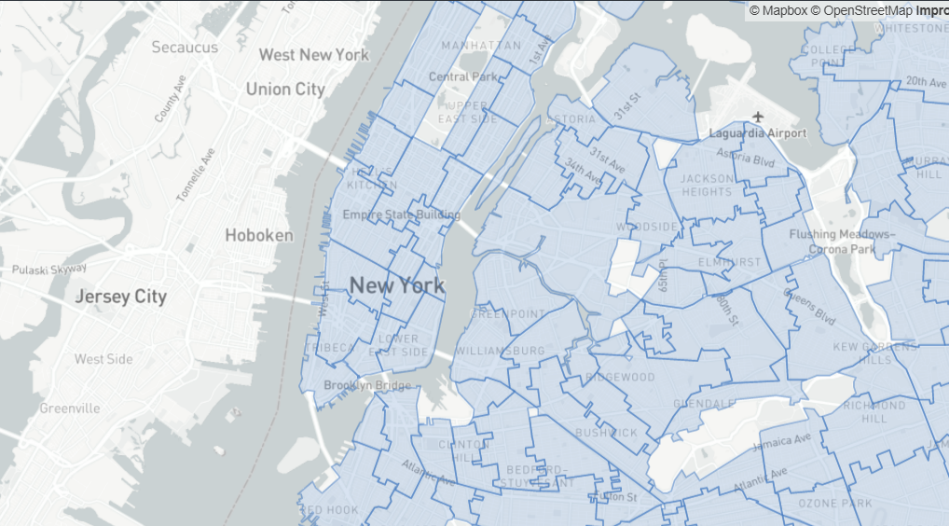
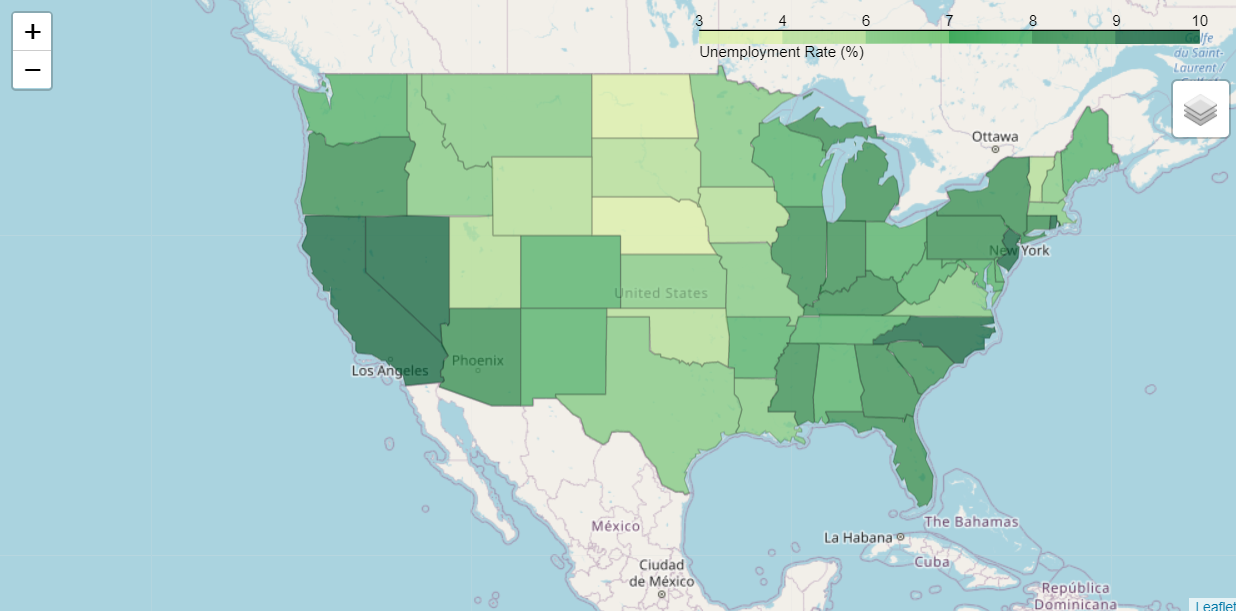
It’s unfortunate that I couldn’t add the boundary layer of New York Neighborhood even though I have found the geojson of it but with this, a choropleth folium map can be made which would be a better and accurate way of visualizing crime rate data of neighborhood utilizing all of the 306656 data instead of 1000 random data points. Because of this, every time the notebook is restarted and the code run again, a new series of 1000 data points are chosen again and the clusters of crime data changes in some way, making the previous analysis made inaccurate. Therefore, the chosen purple circle locations might not be the most optimized choice.

Figure 3- Example of chropleth map with US's unemployment rate

Figure 2- Boundary Layer for New York Neighborhood from Github

Result

Even though it might not be optimized, but the choices were made according to the analysis made in shown screenshots below. The analysis are made around Upper West Side and Chinatown shown as purple circles. NYU is in red circle. Chinese restaurants are not searched around Chinatown cuz, I mean…. It’s Chinatown……

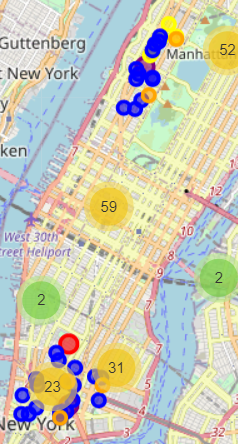


Figure 4- Overall Data visualization made for the project

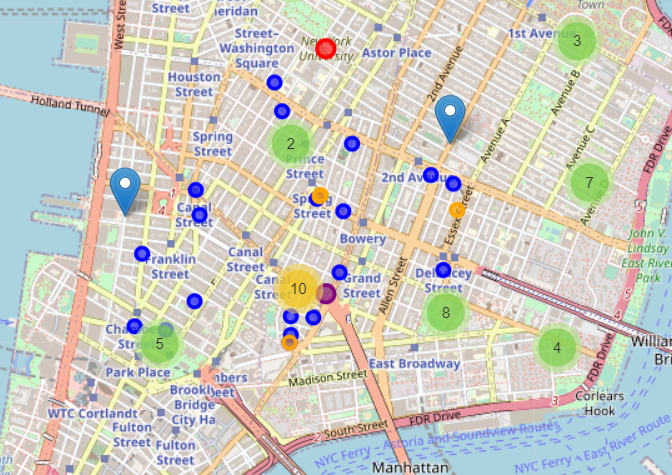
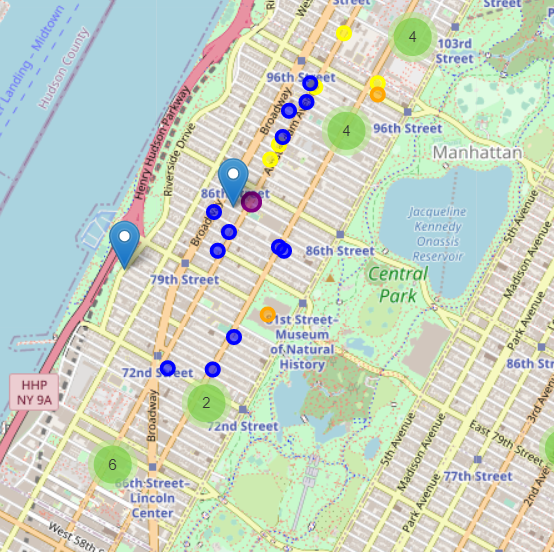


Figure 6- Data centered around Upper West Side

Figure 5- Data centered around Chinatown

Conclusion

Personally, I like the area around Chinatown because it is closer to the University and have more variation and choices of restaurant and also closer to the arcades. But compared to Upper West Side, Chinatown has more crime rate. So I guess, for the sake of being at the safer side, I guess I’ll look for places to rent in Upper West Side.