Coursera Capstone Project

THE BATTLE OF NEIGHBORHOODS

JUNE 2020

Table of Contents

1 - Introduction & Business Problem

A common problem facing many people is choosing a place to live. Even after you have decided on a city, there can be an overwhelming number of options and tradeoffs to consider. This is also a space with a large amount of data that could help you to make better informed choices if you can integrate it in the right way.

The focus of this project will be on New York City's rental real estate market. A user will be able to specify certain criteria for their preferences and using this data science tool, provide back a set of potential housing options. Criteria include:

- Max rent per month (i.e. $5,000 per month)
- Number of bedrooms (i.e. 3 bedrooms)
- Proximity to a metro station in the Manhattan area (i.e. within 1 mi radius)
- Type of amenities in the neighborhood (i.e. Asian restaurants, coffee shops, gym, etc)

The goal of this project is to find a suitable apartment for rent in Manhattan NY that meets the requirements of price, and venue. The following section will detail the data sources required for the analysis. This problem is common to many large cities in the US and abroad. While the data used in this project is specific to Manhattan, the approach is common and can be applied more broadly.

2 - Data

Below are the list of datasets required for this project:

- Apartments for rent including price, address, # of beds
- Manhattan neighborhoods including latitude and longitude
- Venues for each neighborhood
- Subway stations in Manhattan including address

Addresses from rental realestate locations will be converted to lat/lon coordinates. Using Foursquare API we will map the top venues for all manhattan neighborhoods and cluster into groups. We will also convert the address of the subway to lat/long to determine the proximity to realestate listings. Data will be sourced from open data sources where available and using Foursquare api for venue information.

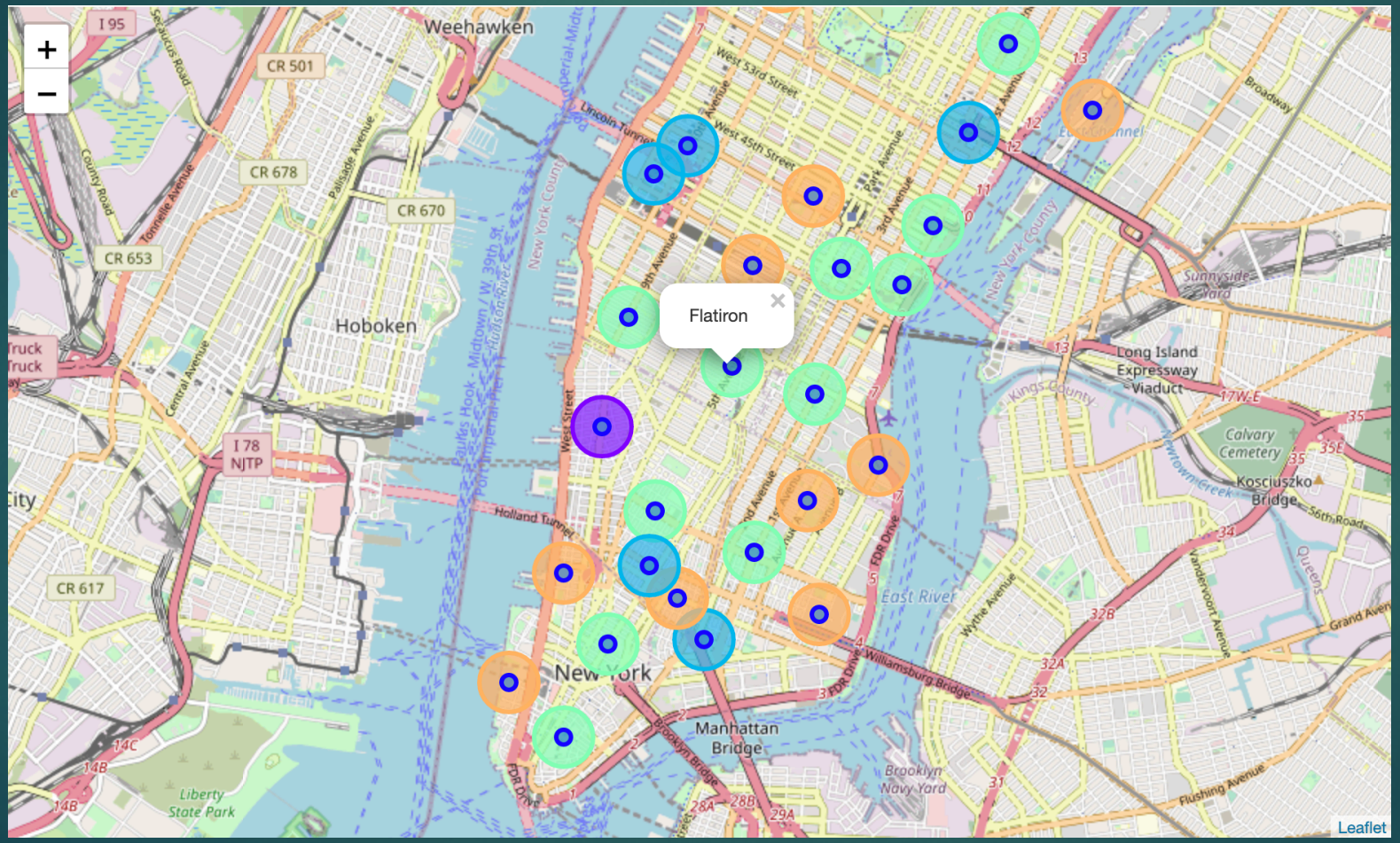
In addition to solving the primary problem of identifying the optimum place to live, the dataset will allow a user to look at:

- What is the average rental price in each Manhattan neighborhood?
- Is there a correlation between the price of real estate and proximity to subway stations?
- What types of venues make up each Manhattan neighborhood?
- Is there a correlation between the price and number of bedrooms or other realestate metrics?

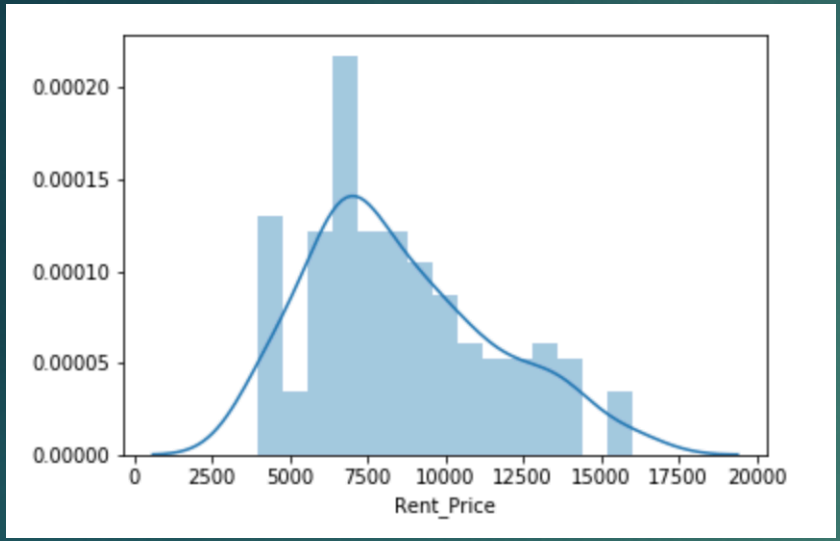
3 - Methodology

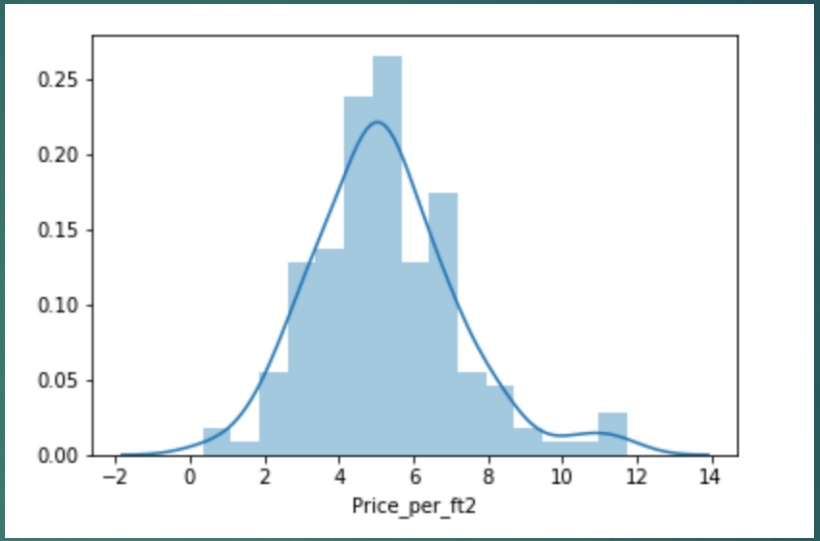
The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the criteria defined: location near a subway, rental price and desired venues. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection intuitive. The tools described are used here and the Notebook cells indicates the execution of steps.

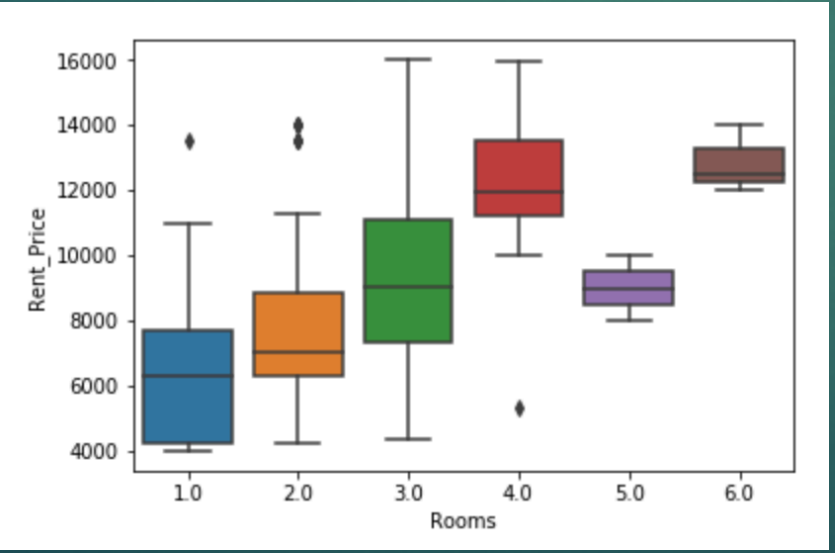
Map of Manhattan neighborhoods with top 10 clustered venues



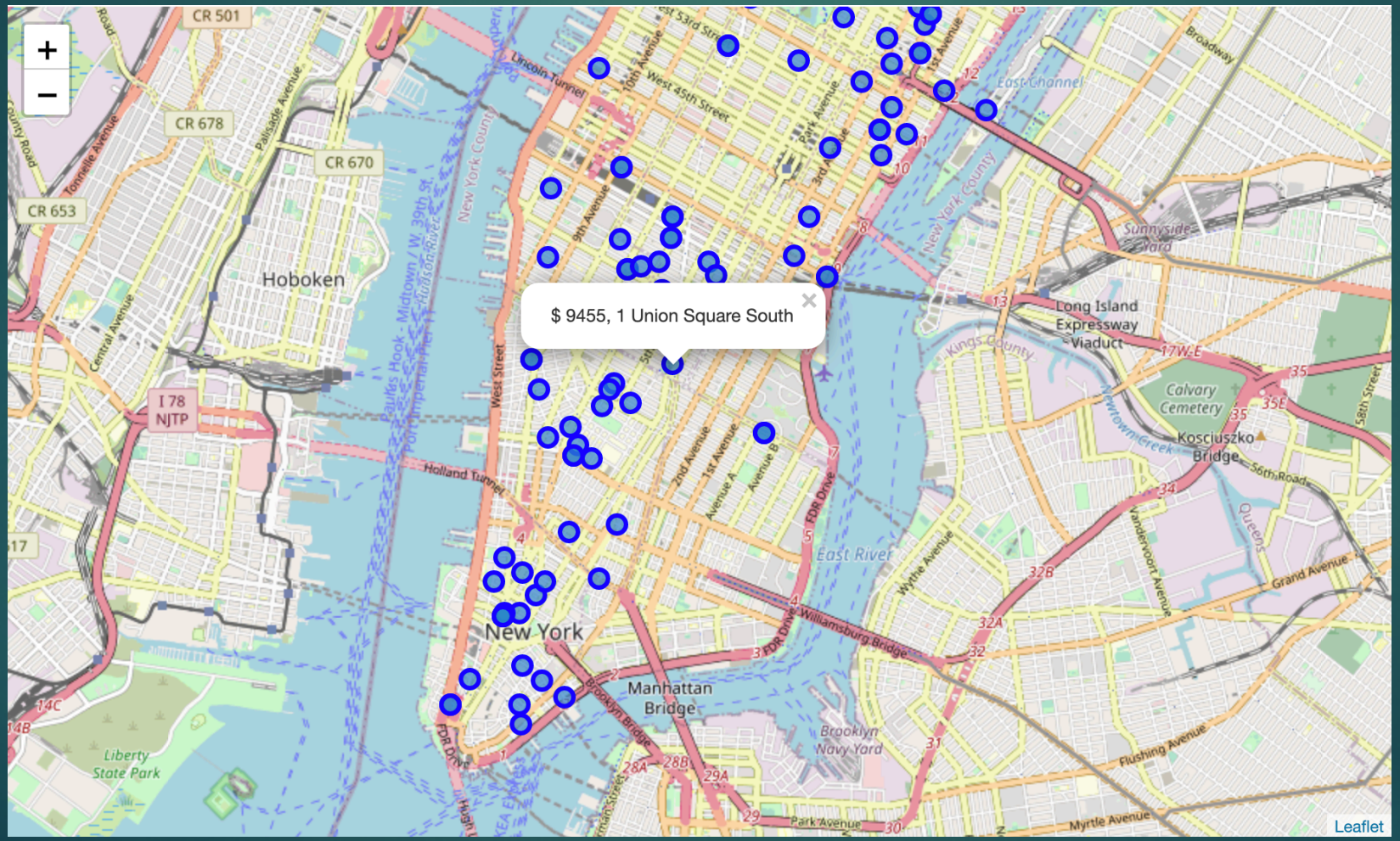
Manhattan apartment rent price statistics





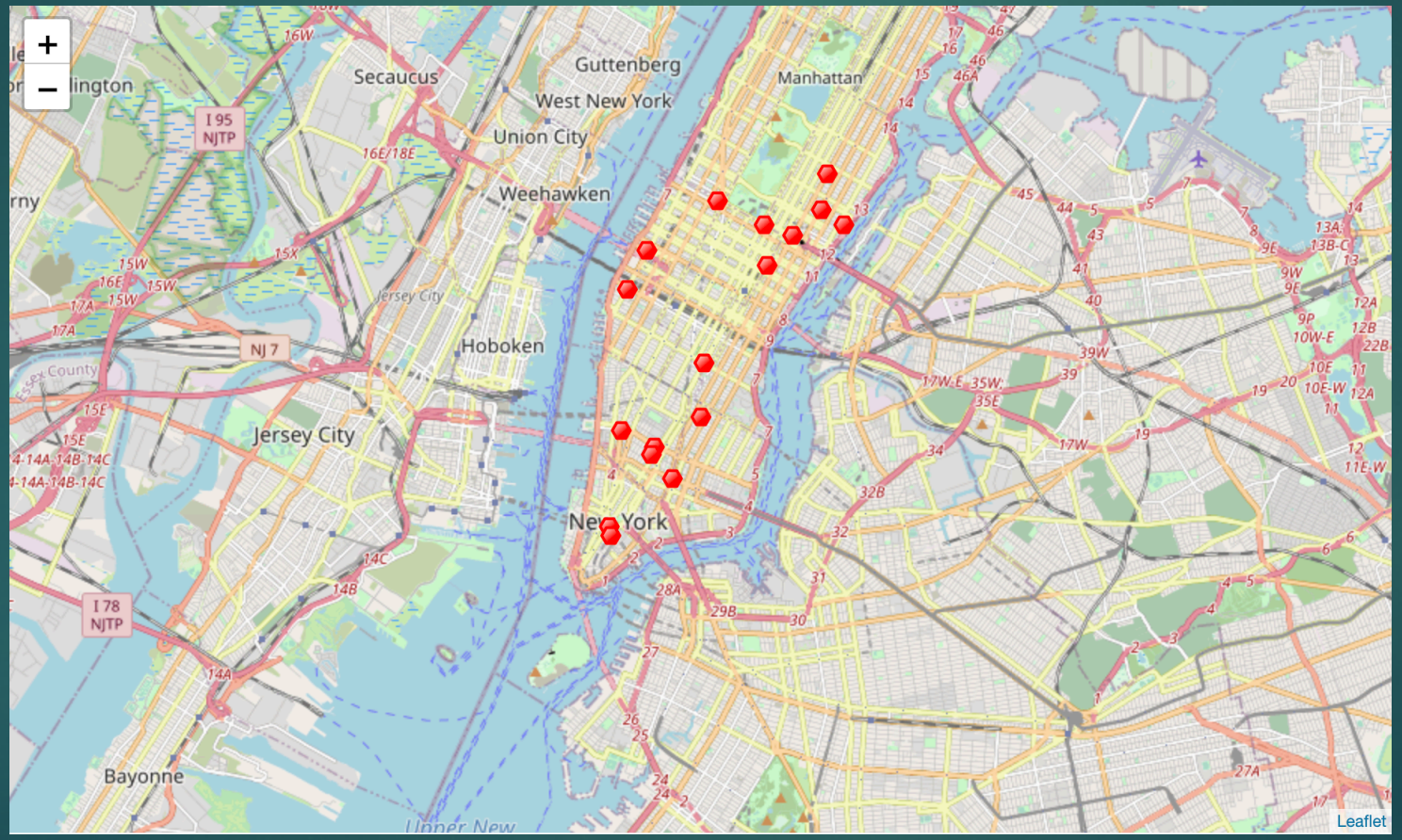


Map of Manhattan apartments for rent

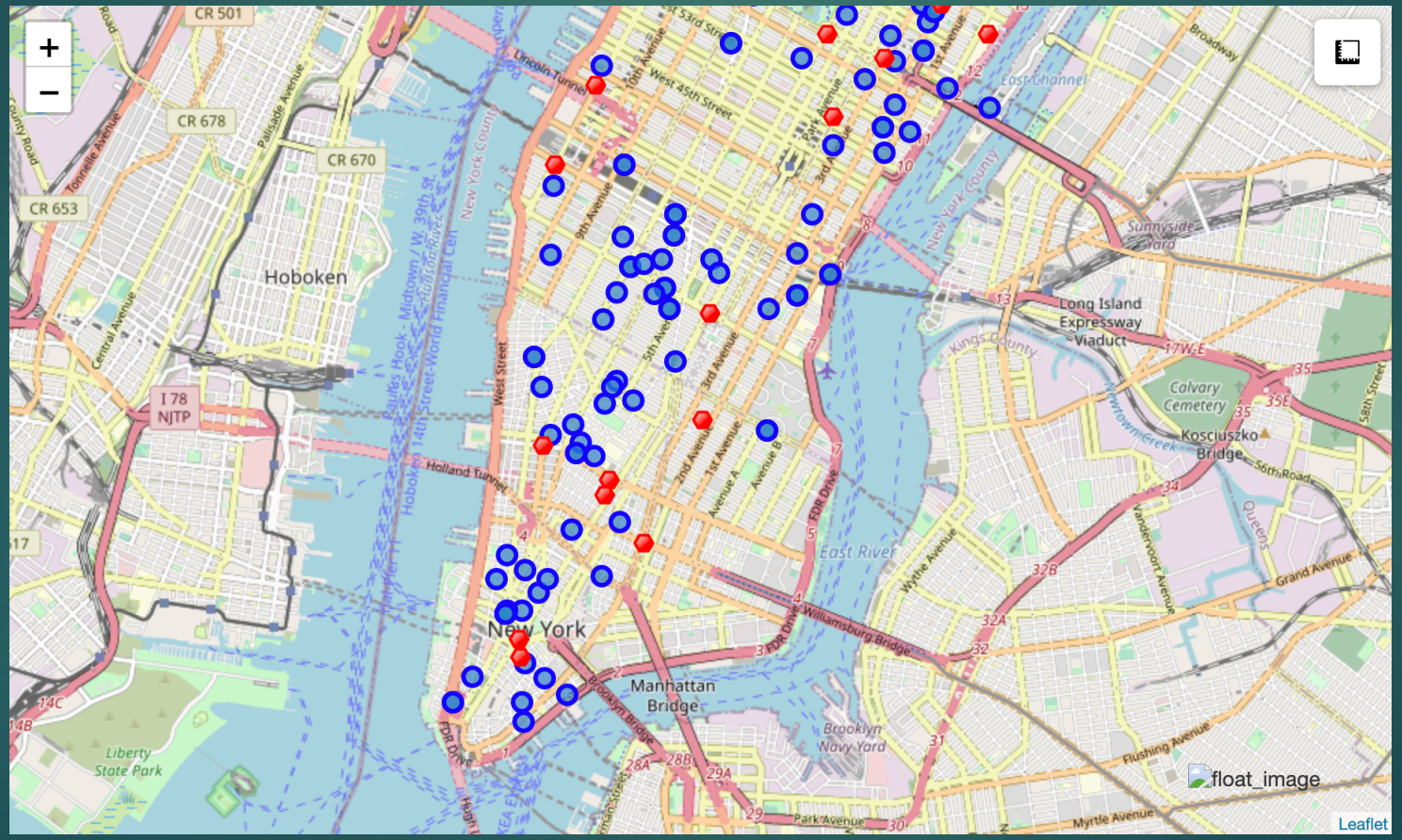


Map of Manhattan showing the places for rent and the cluster of venues

Map of Manhattan Subway Stations



Map of Manhattan showing places for rent and the subway locations nearby



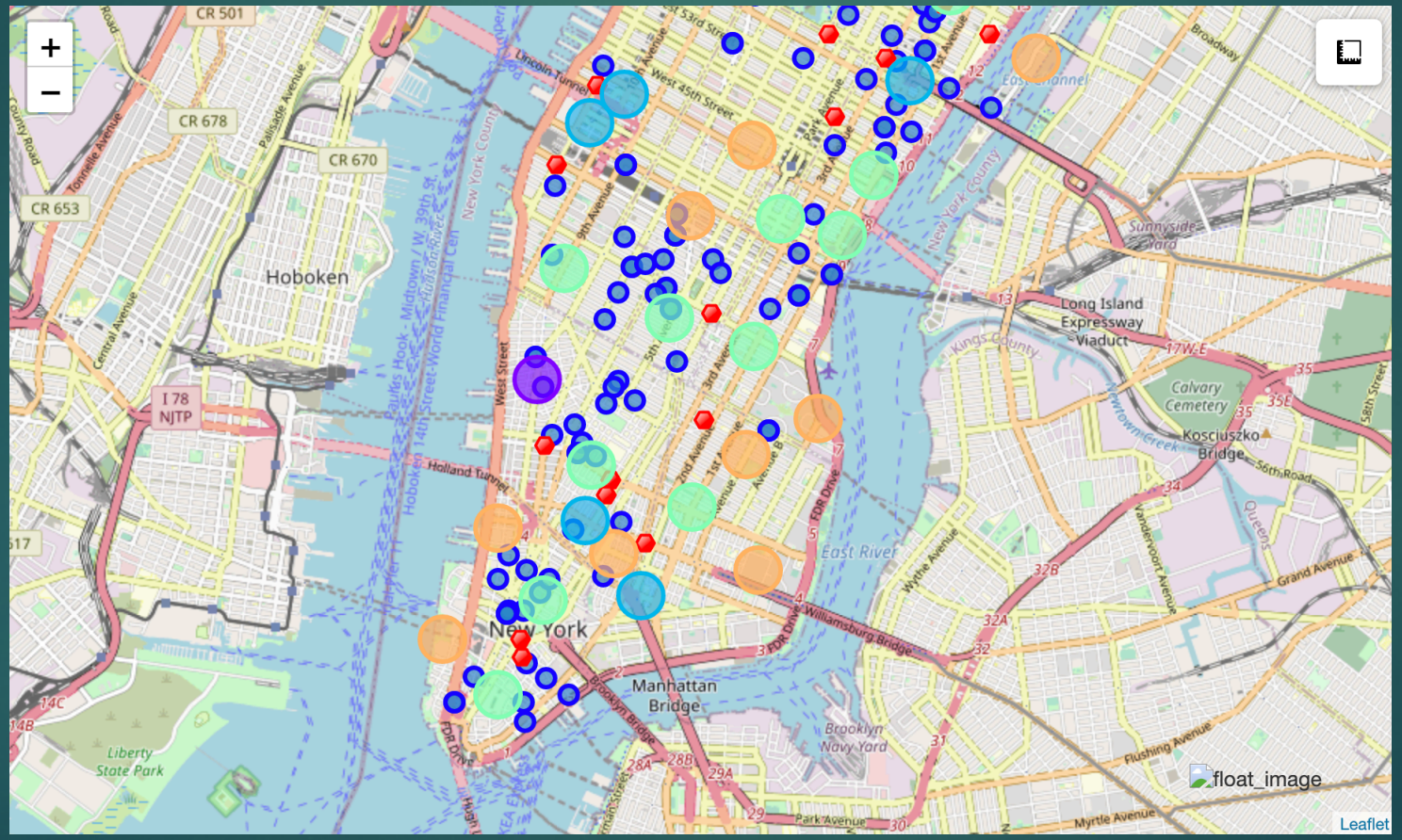
4 - Results

Visualize the desirable rental places and their nearest subway station. Popups display rental address and monthly rental price and the subway station name.

Use one consolidated map to serve as a basis for evaluating rental places by observing their proximity to venues and other amenities.

Once a set of possible apartments are determined, further explore the venues in that cluster to learn more about the neighborhood.

Map of Manhattan with rental places, subway locations and cluster of venues



- Blue dots = Rental apartments
- Red dots = Subway Stations
- Larger circles = cluster of venues

5 - Discussion

Section where you discuss any observations you noted and any recommendations you can make based on the results.

This course provides a good foundational set of data science skills. In this project we leverage some of those to explore a common problem of visualizing and comparing geographic data. Our foundational data skills enabled us to source public data and input it into our notebook. Following some basic data cleanup we were able to map this to lat/long coordinates which ultimately allow us to map them. The mapping features of Folium provide a quick and powerful way to visualize information and for this project link together multiple different datasets.

6 - Conclusion

This project has provided a practical example of some of the skills learned in the Coursera Data Science class. Data sourcing, manipulation in pandas, and geo data mapping are all very important skills to understand for real-life use cases.

Areas to expand this work would include:

- Further analysis of trends in the data (i.e. regression of price vs number of bedrooms)
- Ability to automatically select apartments based on a set of input criteria
- Ability to identify market value opportunities (i.e. rent price lower than predicted compared to other comparable apartments)