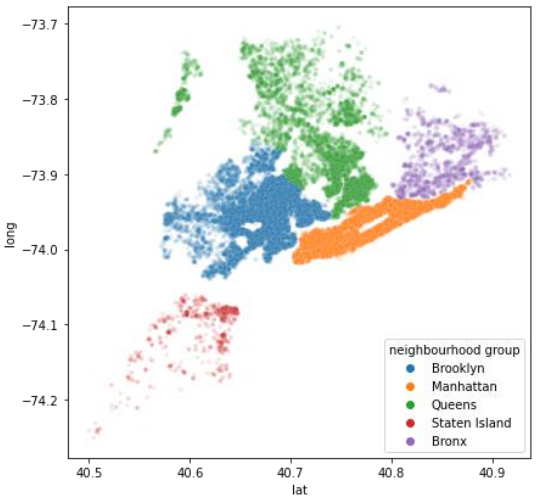
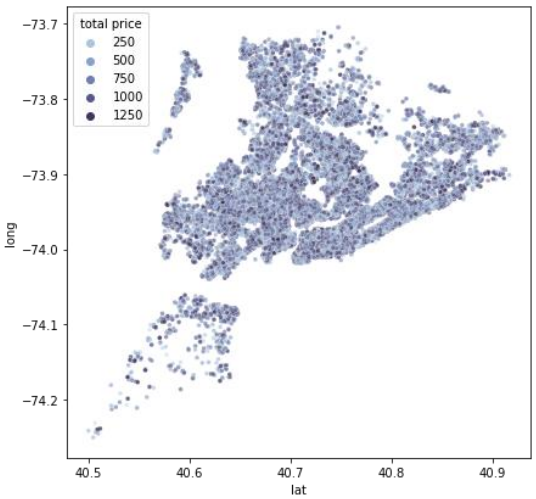
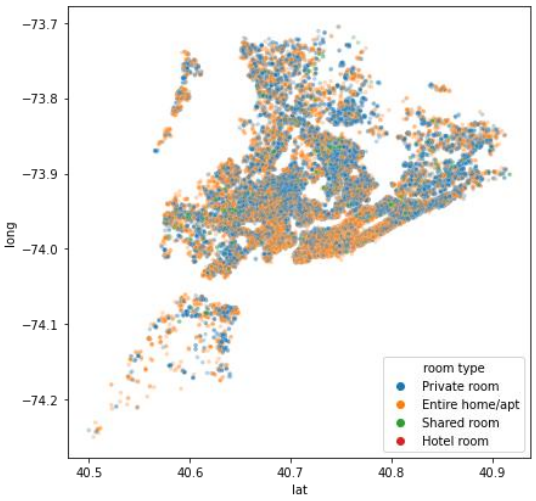
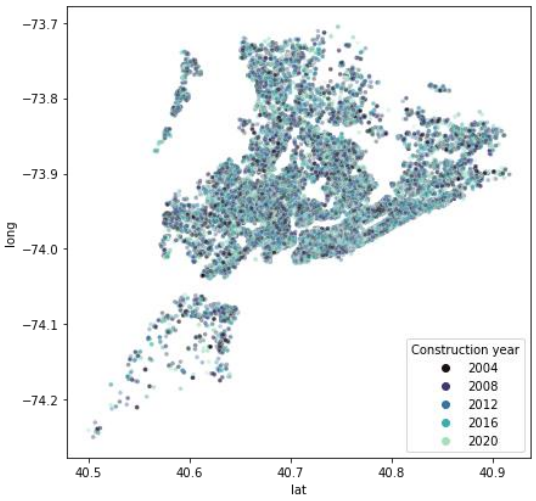
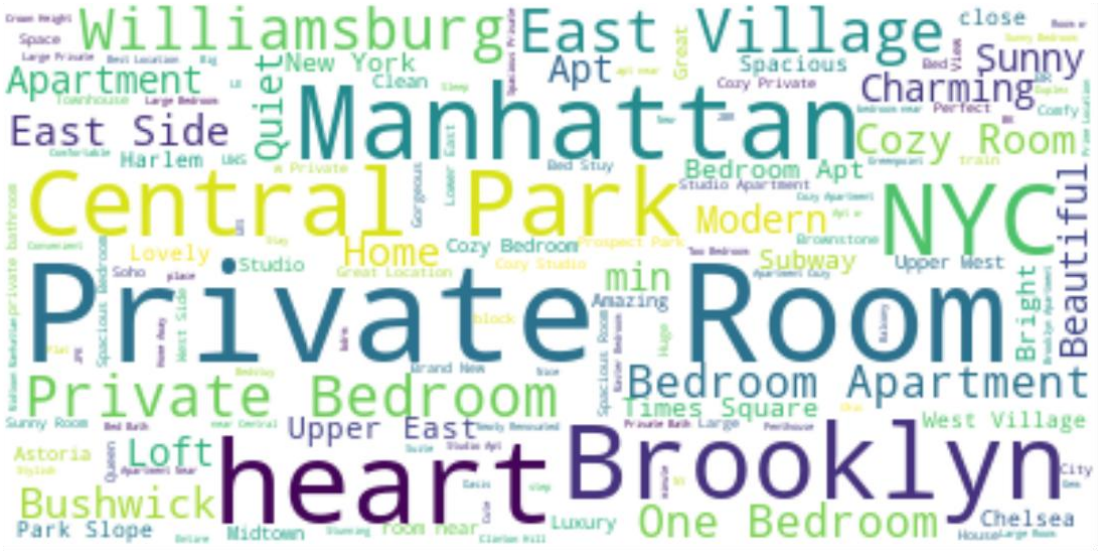
Rental Distribution of Airbnb in NYC











There are one main word cloud plot showing popular word in rental names and four side scatter plots showing the geographic distribution of Airbnb rentals classified by neighborhood, price, room type and construction year in New York City. In scatter plots, the x-axis represents the latitude and the y-axis represents the longitude, so each scatter in the plot represent the a rental.

Findings

- In word cloud, most popular words in Airbnb are “private room”, “Manhattan”, “heart”, “Central Park” etc.

1. In the first plot, rentals located in different neighborhoods have different colors.
    - Most rentals located in Manhattan and Brooklyn with very dense distribution.
2. In the second plot, rentals with different price range are labeled with different colors.
    - Rentals with different price are random distributed in all neighborhoods.
3. In the third plot, rentals of different room type are labeled with different colors.
    - Most rentals are private room and entire home/apt.
    - Only a few of rentals are shared room and barely no hotel room (because most hotel room are not posted on Airbnb).
4. In the forth plot, rentals built in different years are labeled with different colors.
    - Most rentals were built in 10 years.

Data and method

The dataset was download from Airbnb Open Data. There are 26 columns and here are some important columns: 'id' - unique identifier for a rental, 'NAME', 'host id', 'host_identity_verified', 'host name', 'neighbourhood', 'lat' - Latitude, 'long' - Longitude, 'instant_bookable', 'cancellation_policy', 'room type', 'Construction year', 'price', 'service fee', 'number of reviews' and so on. They describe the basic information for each rental in NYC.

I started with looking in the dataset and found that there are some dirty data. Therefore, I cleaned the duplicated rows and converted some columns into a uniform formatting. For example, some hosts set a comparatively lower rental price to attract more people, but the discounted price will be added to the service fee. So I use the sum of rental price and service fee as total price. Then I picked four columns, which are the four most important references when I picked a rental in Airbnb. They are neighbourhood, price, room type and construction year.

Significance statement

As the most popular tourist city and one of the most expensive places in the world, the hotel price is extremely high in New York City, so many tourists tend to find a comparatively cheaper rental. In order to find a ideal rental, tourists need to take price, neighborhood and many other attributes into consideration. Therefore, I visualized the information of rentals in NYC and hope it could be helpful for my next trip to NYC.

Github page: https://github.com/player654012/2415IV

Data is compressed due to oversize