Airbnb - New York:

A Python & R Analysis

Final Report

Created by:

Rounak Nischal - 25%

Aakash Bhargava - 25%

Danielle Rowan - 25%

Dhruval Patel - 25%

Due Date:

May 10th 2020

Executive Summary

Our team performed data analysis on the Airbnb New York City dataset that was available on Kaggle. The dataset consists of information about neighborhoods, geographical locations, number of reviews, accommodation types, etc. among several different parameters.

Our task here was to find out the price of an Airbnb booking based on multiple factors. The exploratory analysis that we conducted helped look into the data on a deeper level through extensive data visualization and usage of pandas for data wrangling. Through the techniques mentioned above, we were able to find out relevant insights to evaluate further whether a particular parameter was a critically influencing factor in our dataset or not. After extensive cleaning, visualization & analysis of the data, we were able to study the various parameters in the dataset thoroughly, and we plan on using this analysis, accompanied by multiple modeling techniques to predict further the parameters of a perfect Airbnb that a host can open.

Table of Contents:

Introduction:

Airbnb was founded in 2008 by Brian Chesky, Joe Gebbia, and Nathan Blecharczyk in San Francisco. They eventually became one of the first to peer-to-peer services that specialized in housing accommodations. Airbnb slowly grew into more cities and eventually started to expand worldwide. They now have operations in more than 220 countries and allow customers to be able to rent out a room, an apartment, or even a house for a night or longer.

In New York City, Airbnb has grown exponentially and allowed for regular people to rent out their room or house to visitors. We have taken the data set provided to us by insideairbnb.com for the year of 2019 in order to conduct in-depth analysis on the hosts, geographical availability and other metrics to make predictions and other conclusions. Our main goals are to answer the following four questions:

In order to understand our data set, we must create goals and objectives that we hope to answer the questions to. For this project, we will answer the following questions:

1. How is price impacted by location and physical attributes?
    a. What impact does price have on review scores?
2. Why are there more Airbnb's in certain locations?
    a. If we use Airbnb distance from a landmark/subway station, can we tell where people want to stay?
3. What are the key predictors of price?
    a. Can we get an approximate range for each neighborhood, room type, and other predictors?
4. What kind of model can we create with this information to use as a prediction for future analytics?
    a. What sort of predictions can we make for those hosts?

Methodology:

In order to achieve these managerial insights into our dataset of Airbnb - New York, we need to follow the basic data wrangling and exploration methods. The following is the way we went about answering our objective questions:

1. Business Objectives:
    a. We must identify the problem and how we will achieve our goals.
2. Data Preparation:
    a. We must understand our data through data wrangling and visualization.
3. Exploratory Data Analytics:
    a. We must create a hypothesis and establish insights into our dataset.
4. Feature Selection:
    a. We must determine the variables of utmost importance.
5. Modeling:
    a. We must train our model and finalize a model with proper outcomes.
6. Business Interpretation:
    a. We must be able to properly explain our results and use the information for future benefit.

Data Sources:

The primary data set for this project is from Kaggle (1). In our Preliminary Analysis, our data set contains 16 columns and 48,896 rows with the following information. Essentially, our data set explains the different types of Airbnb's in New York City's boroughs and neighborhoods within each

borough. Across the 5 boroughs, which are called neighborhood groups in the dataset, there are a total of 221 neighborhoods. Each unit in the data set is placed into 3 categories of room type: private room, entire home/apt, and shared room. The dataset goes further in explaining the prices of these hosts, the number of reviews, reviews per month, and unit availability.

The second, supplemental data for our project is sourced from the Inside Airbnb website (2) which hosts Airbnb data for cities across the globe collected from Airbnb posts and organized into datasets. Within the dataset we are provided information about hosts including their name, location, response rate, what percent of offers they accept, and more. The dataset also gives us information about the space itself, which is rentable, what type of building it is, what type of room and bed, how many bathrooms, how many bedrooms and so on. We are also given information about the geographical location where these units are located including neighborhood, neighborhood group, zip code, longitude, and latitude. Other information presented in the data includes review scores from guests, availability of spaces, what amenities are featured in the unit, and the written reviews from guests. We will utilize the extra information provided by this dataset to enhance the primary dataset from Kaggle.

The other two data sources we will use for our analysis comes from the city of New York government website. This information will help us visualize the information provided in the geopandas portion of this analysis. We will use the information in helping us understand where landmarks are in terms of each NYC neighborhood and borough as well as subway stations.

Preliminary Analysis Findings:

Preliminary Analysis:

The analysis of the data frame using Pandas to the right also helps us in explaining the different types of data types provided in this data set such as int64, object, and float64. When describing our data set using the function data.describe() in python, we can now understand information such as the mean, standard deviation, and max/min values. This doesn't apply or make sense for some columns such as ID, host_id, or latitude/longitude. By using this information we are able to understand that the average nightly price for Airbnb's in the NYC area is $152 with the most expensive being $10,000. Similarly, the average number of reviews for a host's place was about 23 with the entire data set ranging from 0 to 629. Another interesting fact is the average number of reviews per month on a host's place was only about 1. Taking this preliminary analysis, we can now

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
id                                48895 non-null int64
name                              48879 non-null object
host_id                           48895 non-null int64
host_name                         48874 non-null object
neighbourhood_group               48895 non-null object
neighbourhood                     48895 non-null object
latitude                          48895 non-null float64
longitude                         48895 non-null float64
room_type                         48895 non-null object
price                             48895 non-null int64
minimum_nights                    48895 non-null int64
number_of_reviews                 48895 non-null int64
last_review                       38843 non-null object
reviews_per_month                 38843 non-null float64
calculated_host_listings_count    48895 non-null int64
availability_365                  48895 non-null int64
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```

go further into our data set and try to understand if there's a difference in these statistics based on the neighborhood, room type, or even host.

	id	host_id	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
count	4.889500e+04	4.889500e+04	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	38843.000000	48895.000000	48895.000000
mean	1.901714e+07	6.762001e+07	40.728949	-73.952170	152.720687	7.029962	23.274466	1.373221	7.143982	112.781327
std	1.098311e+07	7.861097e+07	0.054530	0.046157	240.154170	20.510550	44.550582	1.680442	32.952519	131.622289
min	2.539000e+03	2.438000e+03	40.499790	-74.244420	0.000000	1.000000	0.000000	0.010000	1.000000	0.000000
25%	9.471945e+06	7.822033e+06	40.690100	-73.983070	69.000000	1.000000	1.000000	0.190000	1.000000	0.000000
50%	1.967728e+07	3.079382e+07	40.723070	-73.955680	106.000000	3.000000	5.000000	0.720000	1.000000	45.000000
75%	2.915218e+07	1.074344e+08	40.763115	-73.936275	175.000000	5.000000	24.000000	2.020000	2.000000	227.000000
max	3.648724e+07	2.743213e+08	40.913060	-73.712990	10000.000000	1250.000000	629.000000	58.500000	327.000000	365.000000

Data Descriptive Statistics:

In order to advance with our data analysis of the New York City Airbnb data set, we need to thoroughly analyze each of our data columns. Beginning with host ID, we can see that host ID number 219517861 has about 327 listings in NYC.

	host_id
219517861	327
107434423	232
30283594	121
137358866	103
12243051	96
...	...
1641589	1
4070519	1
208106618	1
235939247	1
1288080	1

37457 rows × 1 columns

Moving on with our data descriptive analysis, we discovered that Williamsburg had the highest number of listings (3920) and Bedford-Stuyvesant closely following with (3714). In addition, another key data column we have is room type which tells us the three main types of rentals in NYC. If one chose to rent an Airbnb in NYC, chances are they would rent an entire home/apt or private room over a shared room according to the value counts with the column to the bottom right.

	room_type
Entire home/apt	25409
Private room	22326
Shared room	1160

With the data table below, we are able to see the difference between the prices of each room type in each neighborhood group. Entire homes and apartments are the most expensive option by far in every neighborhood but especially expensive in Manhattan and cheapest in the Bronx. Private rooms, on the other hand, are also most expensive in Manhattan but cheapest in Staten Island. Shared rooms are also most expensive in Manhattan out of all 5 boroughs, but Brooklyn is the cheapest in this category.

	neighbourhood
Williamsburg	3920
Bedford-Stuyvesant	3714
Harlem	2658
Bushwick	2465
Upper West Side	1971
...	...
New Dorp	1
Richmondtown	1
Woodrow	1
Fort Wadsworth	1
Willowbrook	1

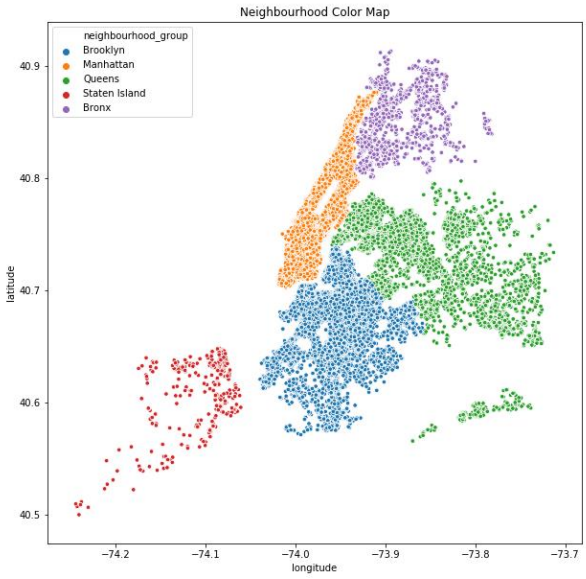
221 rows × 1 columns

	price		
room_type	**Entire home/apt**	**Private room**	**Shared room**
neighbourhood_group			
Bronx	127.506596	66.788344	59.800000
Brooklyn	178.327545	76.500099	50.527845
Manhattan	249.239109	116.776622	88.977083
Queens	147.050573	71.762456	69.020202
Staten Island	173.846591	62.292553	57.444444

Data Visualization of Key Variables:

Similarly, to the previous analysis of our data descriptive statistics, we now will use visualization to help better understand these key variables of our data set. The data set also provides longitude and latitude which could be useful in determining where the most expensive Airbnb's are and why. This could be due to numerous reasons such as walkability

to landmarks, subway stations, or even major tourist destinations. First, we can see by using a scatter plot to visualize the five neighborhoods, each borough has equal amounts of land except for Staten Island but vary in terms of pricing for the types of rooms.



Data Cleansing for physical attributes objective

In this section of our project we are interested in exploring the impact that physical attributes have on listing price per night. Physical attributes refer to the columns where we are given data on the type of room, bed, and property for each listing as well as the number of beds, bedrooms, bathrooms, and people accommodated in the listing, our independent variables. We will explore the data to find trends across the physical areas of New York City set by borders across boroughs, neighborhoods, and zip codes. We will do so by summarizing the average data across all of New York City listings followed by drilling down into each borough to evaluate the borough overall along with the top ten neighborhoods and the top host of the top five neighborhoods of each borough.

We are using the larger dataset from insideairbnb.com which originally was 50,796 rows and 106 columns. After cleaning the dataset, it was reduced to 49,063 rows and 18 columns. This was the result of dropping the columns which did not contribute to the area we are looking to explore as well as columns that had large counts of missing values. We also had to adjust price and zip code to drop missing values and standardize the format of all values in the row. We further adjusted the data to remove outliers in the price column the range originally was from $0 to $10,000 per nights and we limited the range to drop any listing that was $500 per night or more, this dropped 1187 listings from our data. In the figure below we show the original price distribution across NYC and the resulting distribution after the price columns adjustments.

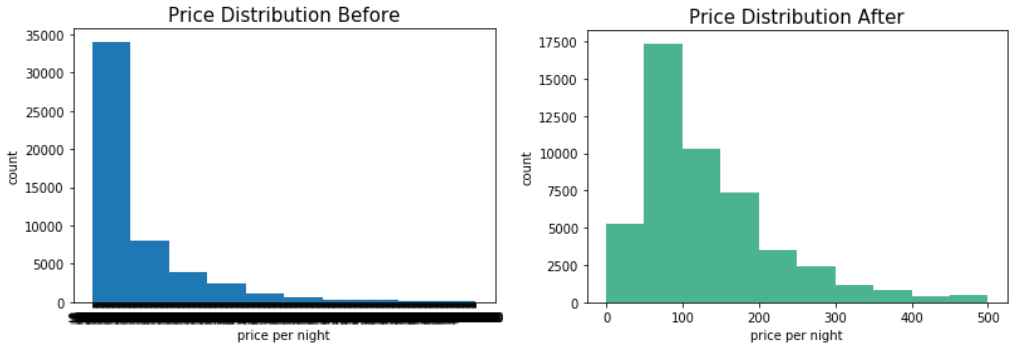


Figure 1

We did the same for the bathrooms column which had a maximum of 6.5 bathrooms per listing and the average listing had approximately 1 bathroom. This column had 54 missing values which we replaced with the column average. We dropped all listings which were equal to or above 5 bathrooms, resulting in a reduction of 19 listings. The average bathroom per listing remained 1 and max dropped to 4. In the figure below we show the original price distribution of bathrooms followed the results after removing this data.

The bedrooms column originally had a maximum of 21 bedrooms per listing and the average was approximately 1 bedroom per listing. This column had 77 missing values which we replaced with the column average. We dropped all listings which were greater than 6 bedrooms, resulting in a reduction of 10 listings. The average bedroom per listing remained 1 and max dropped to 6. In the figure below we show the original price distribution of bathrooms followed the results after removing this data.

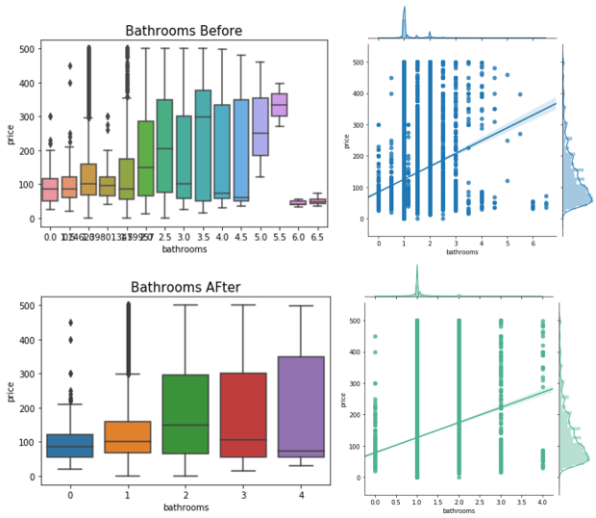


Figure 2

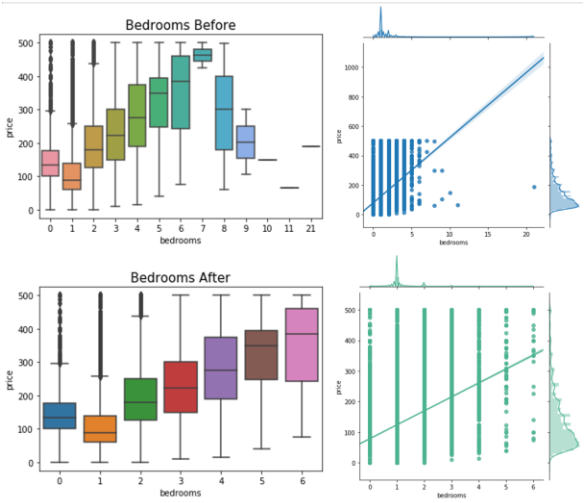


Figure 3

The beds column originally had a maximum of 22 bedrooms per listing and the average was approximately 1.5 bedroom per listing. This column had 482 missing values which we replaced with he column average. We dropped all listings which were greater than 9 beds, resulting in a reduction of 35 listings. The average bedroom per listing remained 1.5 and max dropped to 9. In the figure 4 we show the original price distribution of bathrooms followed the results after removing this data.

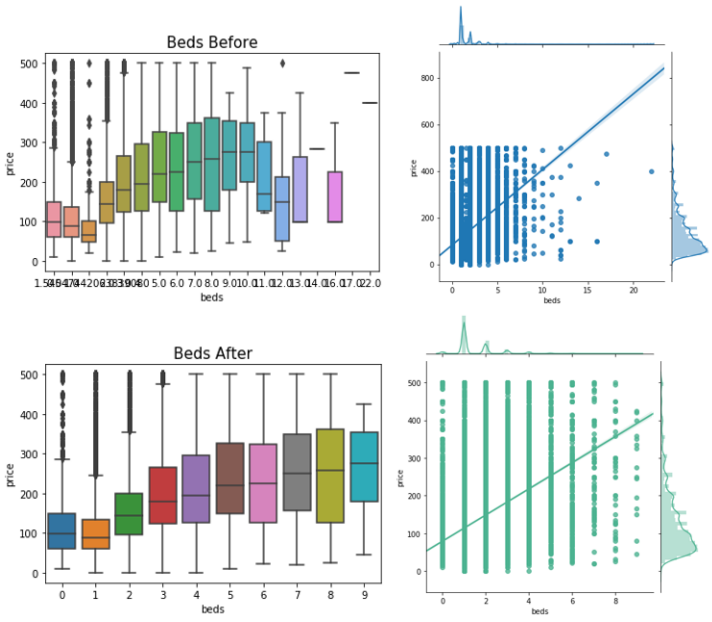


Figure 4

We also evaluated the columns containing information on how many people are accommodated in each listing, how many quests are included in each listed, and the review scores of each listing. We decided not to change these columns. The review score column had 11,431 missing values which we replaced with the column average. The resulting average score is 93.9 with a maximum score of 100 and a minimum of 20. All the review score statistical data before and after replacing the missing values are the same except for a 1.02 reduction in standard deviation. We kept the outliers with much lower scores than average so we can explore if there is a relationship between physical attribute and low scores.

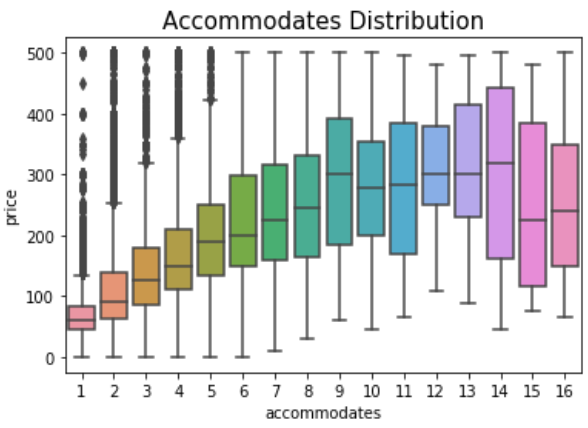


Figure 5

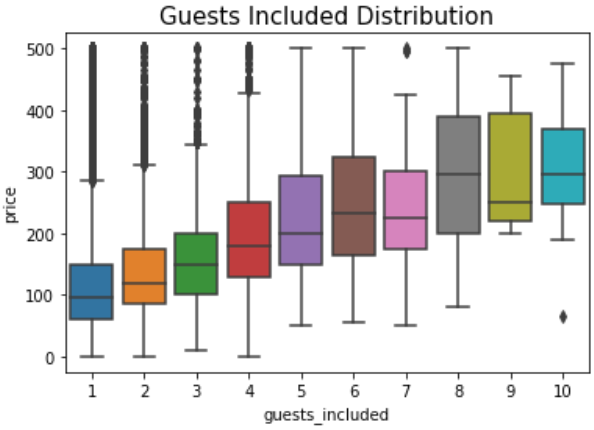


Figure 6

After cleaning the data, we created specific data frames to allow us to explore the variables within specific dimensions which are listed below:

- Each borough
- Top 20 hosts in each borough
- Top 10 neighborhoods in each borough

General Information Across All of New York City Listings

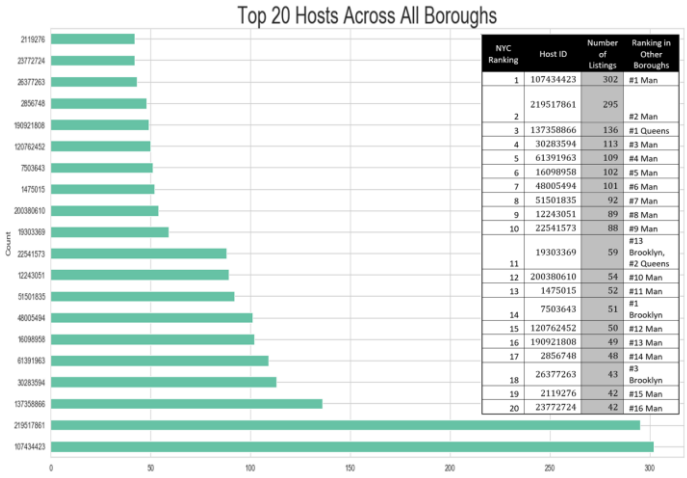
Listings

Within this data we have a total of 49,063 listings. In the figure below we see that Manhattan and Brooklyn have significantly higher amounts of listings, containing 21,354 and 19,973 listings respectively. Queens hosts 6,139 listings, the Bronx has 1,228, and Staten Island contains a significantly small quantity of listings, at 369.

Figure 7

Hosts

Each host can own any number of listings within the dataset, for all of New York City we have a total of 36,877 unique hosts. In the figure below we see the top 20 overall hosts for NYC Airbnb listings as well information about the number of listings these hosts operate as well as these hosts' rankings on the borough level. An interesting discovery from the information collected on hosts that if the host operates numerous listings, it is likely that these listings are across multiple boroughs with a significant amount of their listing within one borough. Host 7503642 is number 14 across NYC and the top host in Brooklyn where this host operates all 51 of their listings. 19303369 has 59 listings with 15 listings in Brooklyn, ranking this host at #13 in Brooklyn, and 37 listings in Queens, ranking this host #2 in Queens. We also see that the majority of the top hosts across NYC are also top hosts in Manhattan.



NYC Ranking	Host ID	Number of Listings	Ranking in Other Boroughs
1	107434423	302	#1 Man
2	219517861	295	#2 Man
3	137358866	136	#1 Queens
4	30283594	113	#3 Man
5	61391963	109	#4 Man
6	16098958	102	#5 Man
7	48005494	101	#6 Man
8	51501835	92	#7 Man
9	12243051	89	#8 Man
10	22541573	88	#9 Man
11	19303369	59	#13 Brooklyn, #2 Queens
12	200380610	54	#10 Man
13	1475015	52	#11 Man
14	7503643	51	#1 Brooklyn
15	120762452	50	#12 Man
16	190921808	49	#13 Man
17	2856748	48	#14 Man
18	26377263	43	#3 Brooklyn
19	2119276	42	#15 Man
20	23772724	42	#16 Man

Figure 8

Price

Across all of NYC Airbnb listings the average price per night is $130.58 with a maximum of $500.00 and listings available for $0.00. In figure 9 below we see that the 75% of the listings have prices between $67.00 and $170.00. We will discuss each borough and top neighborhoods price distribution in more detail later in the paper.

Figure 9

Bathrooms

In the dataset listings can vary from providing zero bathrooms up to 4 bathrooms, the average bathroom per listing is just over one. Figure 10 shows us that in each borough there is a trend that the range of price increased with each bathroom from 0-2 bathrooms provided in the

listing but from 2-4 bathrooms the price range remained around the same. In figure 11 below we see that Manhattan and Brooklyn follow this same trend, but the boroughs with lower volume of listings break away, these three boroughs share the trend of increasing price range from 0-2 bathrooms but vary greatly in the price ranges of 2-4 bathrooms.

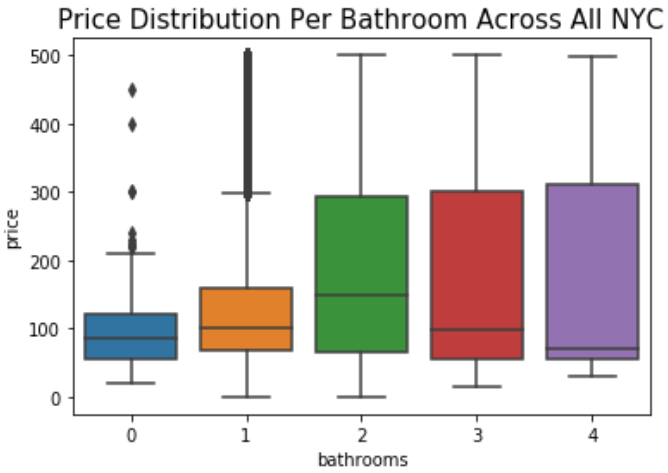


Figure 10

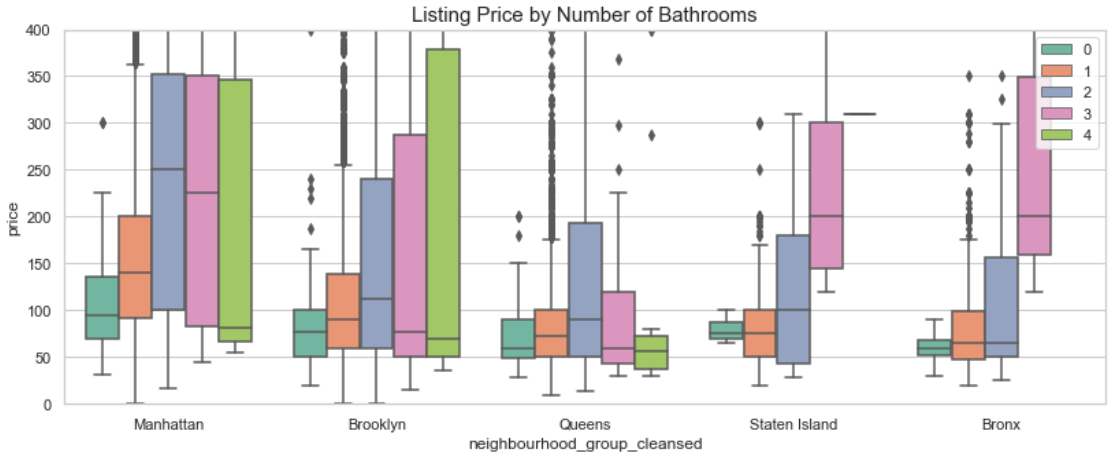


Figure 11

Bedrooms

In the dataset listings can vary from providing zero bedrooms up to 6 bedrooms, the average bedroom per listing is just over one. Figure 12 below shows a positive relationship between price and number of bedrooms per listing from 1-6 bedrooms. As the number of bedrooms increased the

price and the price range increase, except for listings that do not provide a bedroom. Figure 13 below we see that 4 of the 5 boroughs follow this trend, with the exception of the Bronx with shows an increase in minimum price with each additional bedroom, but the price range behaves differently.

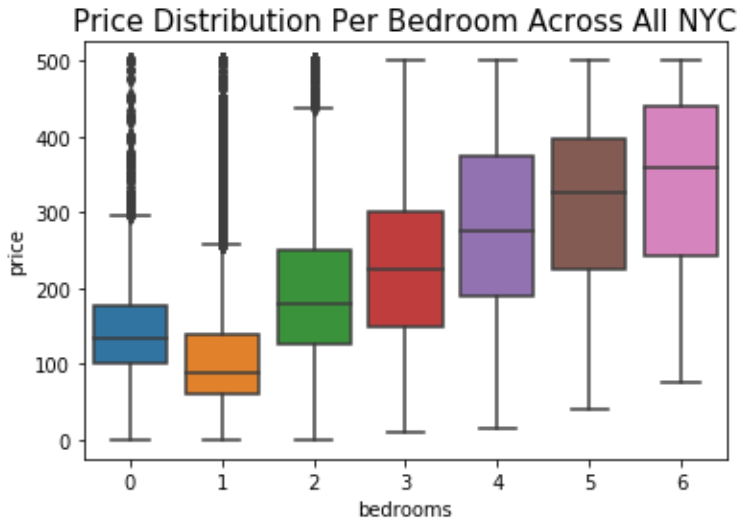


Figure 12

Figure 13

Beds

In the dataset listings can vary from providing zero beds up to 9 beds, the average bed per listing is just over one. In figure 14 below shows a positive relationship between price and number of beds per listing from 1-6 beds. As the number of beds increased the price and the price range increase, except for listings that do not provide a bed. In figure 15 we see that all 5 boroughs follow this trend.

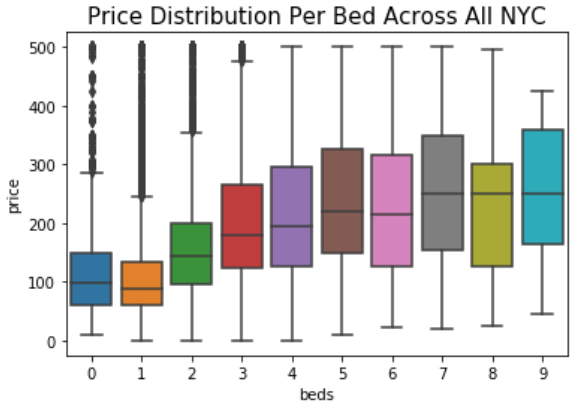


Figure 14

Figure 15

Neighborhoods

There are 220 zip codes used across all NYC Airbnb in this dataset. The figure below illustrates the price distribution of the top 10 neighborhoods of all of NYC, all of which are either in Brooklyn or Manhattan. We will explore further into the top neighborhoods in each borough later in the paper. In figures 16 and 17 we see that Bedford-Stuyvesant has the second largest amount of listings for any neighborhood across all of NYC and the majority of these listings are priced between $50 and $130 per night, amongst the lower prices ranges of the top 10 NYC neighborhoods. We also see that Midtown has the highest price range of the top 10 NYC neighborhoods and has the least listings of these neighborhoods.

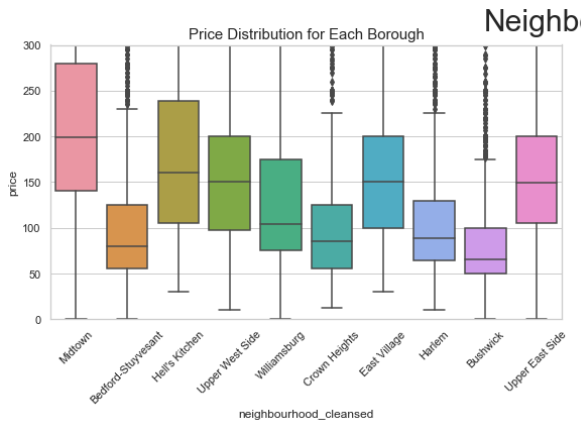


Figure 16

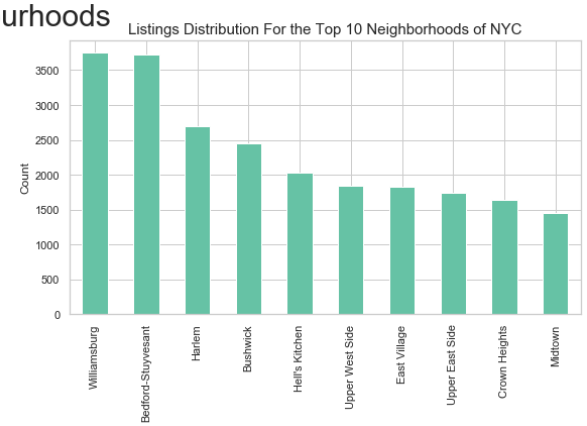


Figure 17

Room Type

There are four room types that each listing can be categorized as: private room, entire home/apartment, shared room, and hotel room. From the figure below we see that the most common room types are entire home/apartments and private rooms. The second below shows the price distribution of each room type across each of the five boroughs. In this figure we see that the largest price range for a room type is a hotel room in Manhattan and that hotel rooms are only available in Manhattan, Brooklyn, and the Bronx. We also see that entire home room types have higher prices than private and shared rooms.

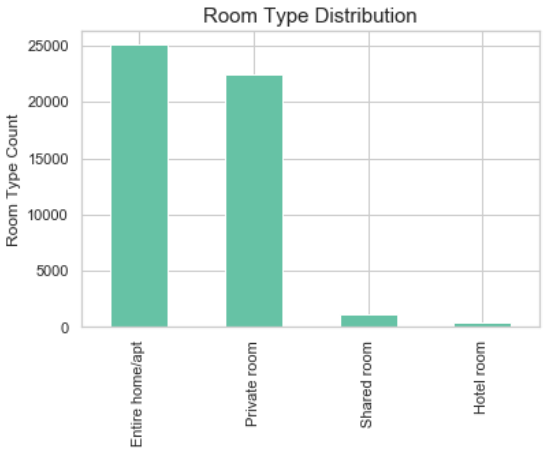


Figure 18

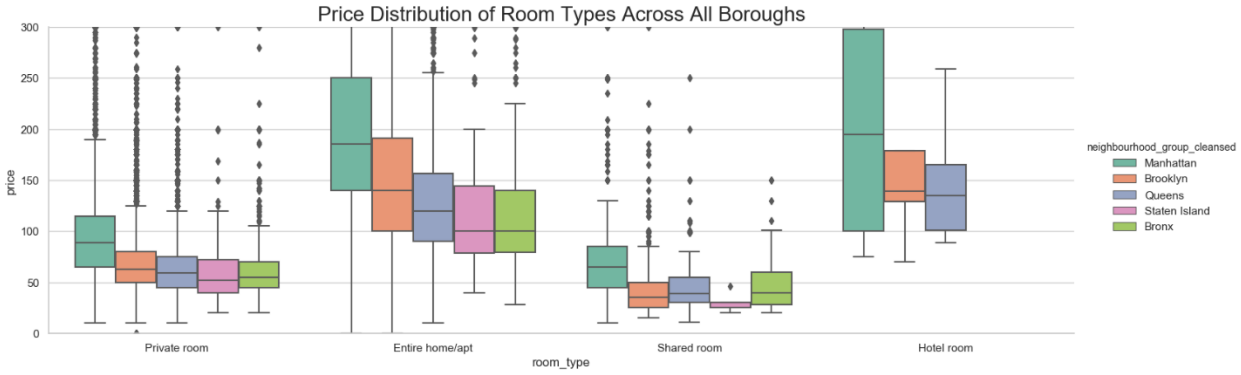


Figure 19

Bed Type

There are five bed types that each listing can be categorized as: real bed, futon, pull-out sofa, airbed, and couch. In figure 20 we see that the most common bed type by a significant amount are real beds, figure 21 is scaled to show the other four types of beds. Figure 22 shows the price distribution of each room type across each of the five boroughs. In this figure we see that the largest price range for a room type is a hotel room in Manhattan and that hotel rooms are only available in Manhattan, Brooklyn, and the Bronx. We also see that entire home room types have higher prices than private and shared rooms.

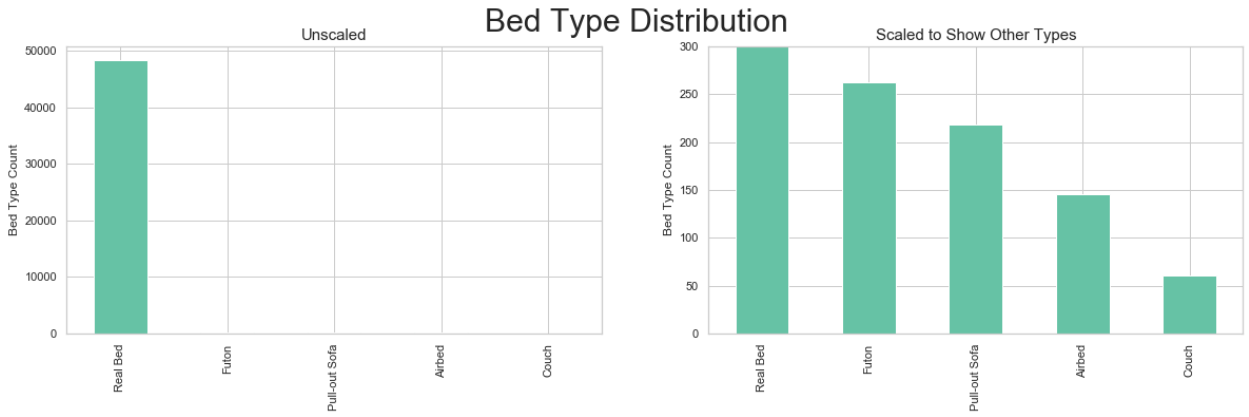


Figure 20

Figure 21

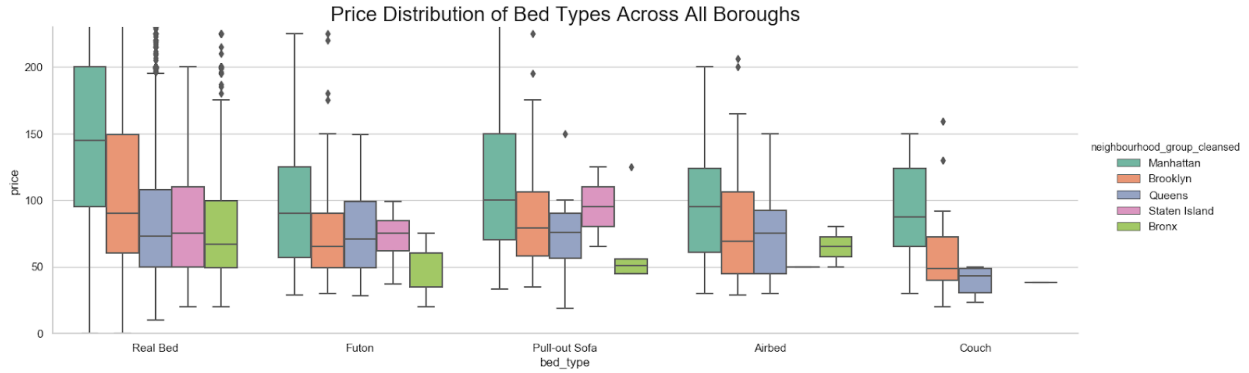


Figure 22

Property Type

There are 25 different property types that the listings are categorized by. For this project we will focus on the top 5 property types: apartment, house, townhouse, condominium, and loft. In figure 23 we see that most listings are categorized as apartments. In figure 24 we see the condominiums and lofts have both the largest price range and the highest prices and that houses have the smallest price range and share the lowest minimum price per night as townhouses. In figure 25 below we see that the across the boroughs apartments are the most common room type except for Staten Island where which houses are the most common room type followed by apartment.

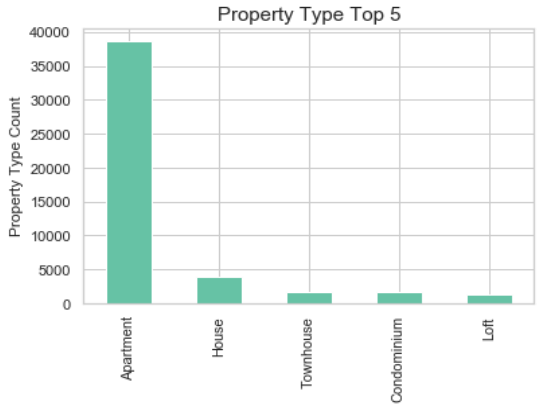


Figure 23

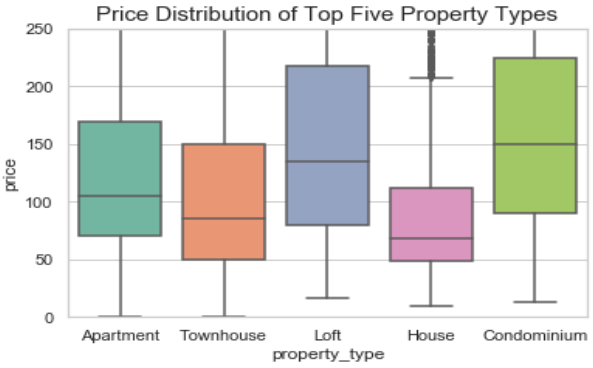


Figure 24

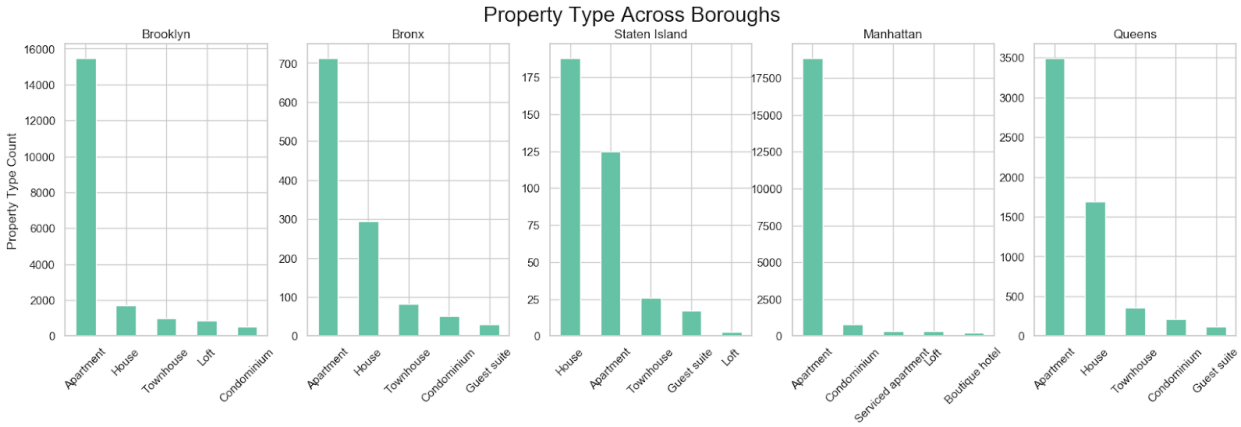


Figure 25

Accommodates

In the dataset listings can vary accommodating 1 up to 16 people, the average number of people accommodatable per listing is just over just approximately 3 people. In figure 27 shows a positive relationship between price and quantity of people accommodatable per listing for 1-13 people. For listings which can accommodate above 13, the price range increased negatively, decreasing the minimum price.

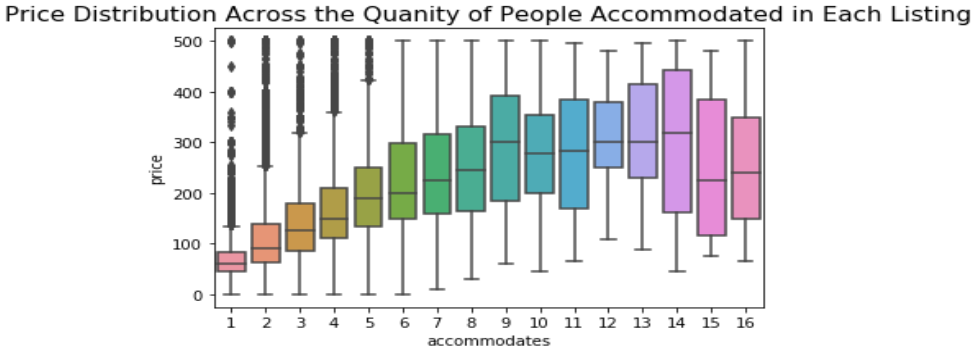


Figure 26

Review Scores

The average review score for all listings across New York City Airbnb listings is 93.92 with a max of 100 and a minimum of 20. This column of data has a significant amount of missing values which were replaced with the average score which changed only the standard deviation of this column, all other statistics remained the same. The review scores have been grouped into 5 bins increasing in score from 1 to 5 as seen in the chart below. In figure 28 we see a positive relationship between price and review score. Bin 1, with the lowest scores, also has the smallest range in range although it's minimum price is approximately equal to that of bins 2 and 3. Bin 5, with the highest scores, has the largest price range and also the highest prices. In figure 29 we see that the boroughs have approximately the same distribution of review scores, later in the paper we will evaluate each borough's review score relationship with price.

Bin	Score Range
1	20-36
2	36-52
3	52-68
4	68-84
5	84-100

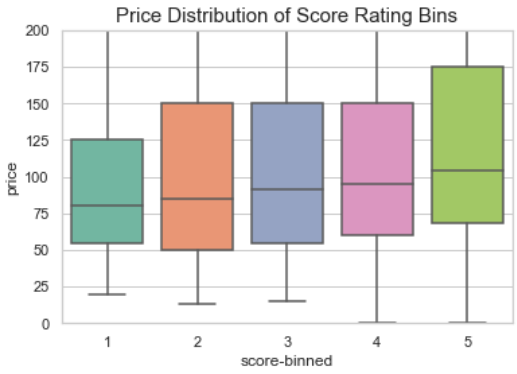


Figure 27

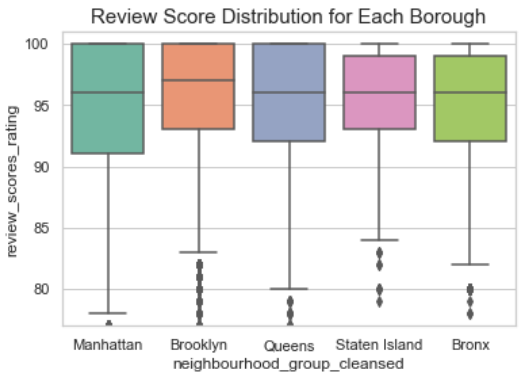


Figure 28

Borough specific profiles

In appendix for objective #1 we have produced visual profiles of each borough and the top ten hosts of the top five neighborhoods within each borough.

- price range
- price distribution
- price distribution for each of the top ten neighborhoods
- 4 plots showing the relationship between beds, bedrooms, bathrooms, and number of people accommodated by each listing and price

- Box plot showing price distribution across the 5 bins of review scores, with the same range as above, 1 is the lowest scores and 5 is the highest scores
- Line plot showing the relationship between price and review score in the borough

Top 5 Neighborhood Host profiles include the following:

- Number of listings in neighborhood and in general
- Price range of host's listings within the neighborhood and in general
- Price relationship with physical attributes
- Price relationship with review score

Exploratory Data Analysis:

In this part of our exploratory data analysis, we aim to understand why there could be more Airbnbs in certain locations. We already understand that Manhattan and Brooklyn have the most Airbnb listings from all of the boroughs. However, what if we use Airbnb distance from a certain landmark or subway station? Can we tell where people want to stay according to the number of Airbnbs within one mile from these places? Answering this kind of question will give us managerial insight into how a major landmark and ease of transportation close has an effect on how many Airbnbs are listed.

Our hypothesis is that more Airbnbs will be listed in areas where the more popular landmarks are regardless of the condition or filter we place on the Airbnb search. Essentially, we assume that the more popular the landmark, the more foot traffic there is and ultimately, more Airbnbs will be available. In order to test our hypothesis, we use seven different conditions to analyze how many Airbnbs are near the landmark in question. First,

we will look at Airbnbs within a one mile radius of the landmark. Second, we will look at those same Airbnbs within a one mile radius and refine the search with only those Airbnbs less than $200. We always use the within one mile radius search as our base for each of the following conditions. Finally, we look at Airbnbs available at least six months out of the year, those with at least ten reviews, those that only offer an entire home or apartment, have a minimum night stay restriction of 3 or less, and are only a quarter of a mile away from the nearest subway station.

In this analysis, we will use geopandas, spatial operations and separate csv files. We are using geopandas to convert our Airbnb dataset into points and polygons using longitude and latitude. This will help us more easily visualize NYC's 5 boroughs and the neighborhoods in each of these boroughs. Geopandas is an extension of the pandas learning library in Python which will help instill the skills we learned in a new way. We are also using spatial operations because this will allow us to make the 1 mile radius away from a landmark. Because the dataset does not already include boundaries, it will help us contain the area we want and further emphasize the distance between two objects. We are combining our Airbnb data set which already gives us certain longitude and latitudes for each Airbnb listing with existing neighborhood boundary information and subway station information from the NYC Government Website.

In our preliminary analysis, we did a seaborn scatter plot of each airbnb listing grouped by neighborhood groups with longitude and latitude as our axis. Now we can import a png file included with the initial Airbnb data set and enclose it with geopandas by creating enclosed polygons. This map (Figure A) in our appendix shows a heat map of the five boroughs demonstrating which borough has the most listings and Manhattan and Brooklyn

lead the rest. Now we have a NYC neighborhood CSV file that we import from an external source mentioned above and through python we can plot it using longitude and latitude to demonstrate each approximate neighborhood boundary. Finally, we take that same dataset and join it with our external NYC boundary information file to enclose each neighborhood and make boundaries. After plotting this new dataset, we can see a heat map (Figure B) which shows the most densely populated neighborhood in terms of number of Airbnb listings in NYC. As you can see, all of Manhattan and portions of Queens and Brooklyn are the most populated. To be specific, Williamsburg in Brooklyn and Bedford-Stuyvesant are the most popular overall and in Brooklyn. However, NYC has the most popular boroughs such as Harlem, Upper West Side, Hell's Kitchen and Midtown. Our question arises again as we are curious to see if landmarks in these areas are what might attract Airbnb customers.

After gathering information from different internet sources on what are the most popular landmarks in New York City, we determined that (Figure C) eighteen landmarks are important to consider for this analysis. Please refer to the python analysis to see the list of all landmarks and each analysis closely. For this report, we will focus on the World Trade Center (WTC). Many of the top landmarks in NYC happen to be in Manhattan, specifically Midtown just as our preliminary analysis indicates. If you are not familiar with where WTC is, it is located in Manhattan and in the Financial District (Figure D). Because this area is primarily business offices, we do not expect to see too many Airbnbs in this area.

We also have the external subway lines with subway stations information in order to determine which Airbnbs are a quarter of a mile away from the landmark we are analyzing. Out of the many subway lines in NYC (Figure E) we can use Google Maps transit feature to see which subway line has the closest stop to WTC. We determine that the R-Line

is the most popular stop near the WTC and someone looking for ease of transportation to other places in NYC would choose somewhere near this stop. This will help us analyze this condition later in the analysis.

Our next step is to utilize spatial operations and create a distance buffer of a one mile radius from the center of the World Trade Center. Finally, we are able to see how many Airbnbs are available within 1 mile from WTC and under different conditions. There are 1247 Airbnbs within one mile away from WTC, and only 591 of those are under $200. Only 509/1247 Airbnbs are available at least 6 months out of the year while the others are taken off the market by their hosts. About 283/1247 (a very low number) have at least 10 reviews. This is certainly disappointing and must mean that these are generally newer Airbnbs. About 924/1247 offer an entire home or apartment and 729/1247 have a minimum night stay of 3 or less.

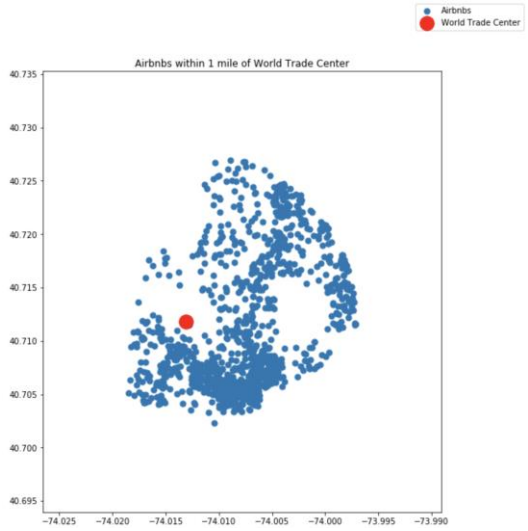
```
Starting number of airbnbs: 1247
Number of airbnbs after cutting price to less than $200: 591
Number of airbnbs after selecting those that are available at least 6 months out of the year: 509
Number of airbnbs after selecting those with at least 10 reviews: 283
Number of airbnbs after selecting those that offer the entire home/apt: 924
Number of airbnbs left after selecting airbnbs that have a minimum night stay of 3 or less: 729
```

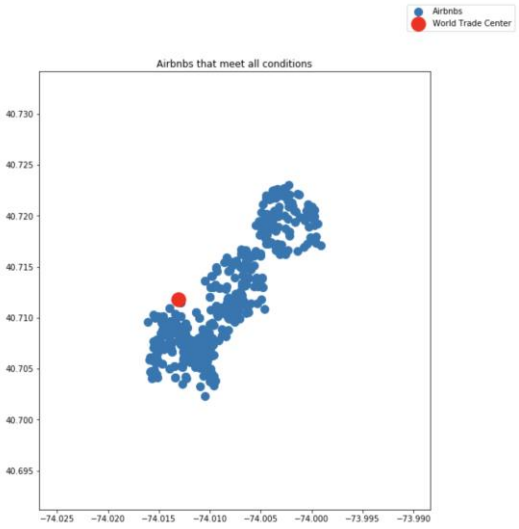
If we decide to filter like a search engine or like Airbnb's website filters, every time we filter or place a new condition, we would get a smaller and smaller number. Every step means they must follow each and every condition. In this scenario, we find that only 5 Airbnbs satisfy all the conditions.

```
Number of airbnbs after cutting price to less than $200: 591
Number of airbnbs after selecting those that are available at least 6 months out of the year: 182
Number of airbnbs after selecting those with at least 10 reviews: 36
Number of airbnbs after selecting those that offer the entire home/apt: 12
Number of airbnbs left after selecting airbnbs that have a minimum night stay of 3 or less: 5
```

Let's only focus on the Airbnbs within one mile away from World Trade Center and

see how many of those are also a quarter of a mile away from the nearest subway station on the R-Line. Clearly the Airbnbs are now more concentrated to the southeast of WTC now.





In order to really compare all of our landmarks, I decided to convert it all into an excel file and move it to R to analyze (Figure F). Each of our analyses are organized by color essentially telling us which borough each landmark is in. Blue plots are from Manhattan, green from Brooklyn, yellow from Bronx, pink from Queens and red from Staten Island. Let's take a look at the first quadrant in Figure F, where you can see the total number of Airbnbs for each landmark. Clearly Manhattan leads in Airbnbs within 1 mile away from the landmarks. Empire State Building and Times Square are the most popular in terms of most Airbnbs near these locations, while High Rock Park and Rockaway Beach are the least popular. In the next couple of graphs, we use percentage as our scale and benchmark to assess what percentage of the total number of Airbnbs within one mile away from each landmark satisfies each condition.

By doing this and not looking at the total number, we are giving each landmark an even base to accurately assess our objective. Looking at our next quadrant, we see that 100% of Staten Island's total number is less than $200 and Manhattan has the most expensive Airbnbs near the landmarks with those near World Trade Center being the most expensive. In the next quadrant, Staten Island leads in available for more than half of the year with many in Prospect Park in Brooklyn not available often along with the majority of Airbnbs. This could be because most of the Airbnbs might serve as listings during summer months while the owner is on vacation.

Next, we see that very few percent of Manhattan landmarks have at least ten reviews with Staten Island and Queens having a large percentage. For Airbnbs that offer an entire home or apartment, surprisingly, Manhattan offers a large percentage of complete apartments compared to the majority of locations outside of Manhattan that might offer shared or other apartments. Last, we see that the percentage of Airbnb locations that offer a subway station close to the listing vary between neighborhoods with Queen's Rockaway Beach ranking first and Bronx's Yankee Stadium ranking last. Now that we've compared each landmark, let's see where the most popular Airbnb locations are in terms of landmarks.

Overall, Manhattan and Brooklyn are number one and number two respectively. In Manhattan, the Empire State Building, Times Square and Madison Square Garden have the most airbnbs that fulfill our conditions within one mile away. For Brooklyn, its Brooklyn Bridge, Prospect Park and Coney Island. For the Bronx, its Yankee Stadium, New York Botanical Garden and Bronx Zoo. For Queens, it's Citi Field and Rockaway Beach. For Staten Island, it's Staten Island Ferry and High Rock Park. Please take a look at Figure G to see the complete list of rankings. The first six rankings are in Manhattan with the first five in Midtown and the sixth ranking in Chelsea, as expected.

However, number seven and number nine end up in Brooklyn, while number ten (WTC) is in Manhattan's Financial District.

Ultimately, we can take what we learned here and compare it to our initial hypothesis. Were more airbnbs available in areas where the greatest known landmarks were regardless of the condition we require? Yes, however, we did see quite a percentage decrease for many of the conditions. It is clear to summarize that Manhattan and Brooklyn are the most popular tourist attractions because of their landmarks ultimately exhibiting more airbnb listings.

Feature Selection:

Predictive Analytics:

The entire idea behind doing predictive analytics for this dataset is to understand the consumer mindset. We're going to perform predictive analytics on multiple parameters listed in the dataset that will in turn help us plan from a host's point of view. We want to help future hosts understand that if they plan on purchasing a new property in New York City with the idea of renting it out, what is going to be the best location i.e. w which borough of manhattan should they purchase the property, what the best price per night would be in that location, minimum number of nights someone needs to rent the property, what would be the best neighborhood in that particular borough of New York City, example Greenwich Village in Manhattan & room type, which could be anything from an entire apartment to a shared room in a large apartment.

Methodology

- Data Preprocessing
- Using domain knowledge and existing EDA ton determine best columns for prediction
- Executing different regressions models to predict the price.

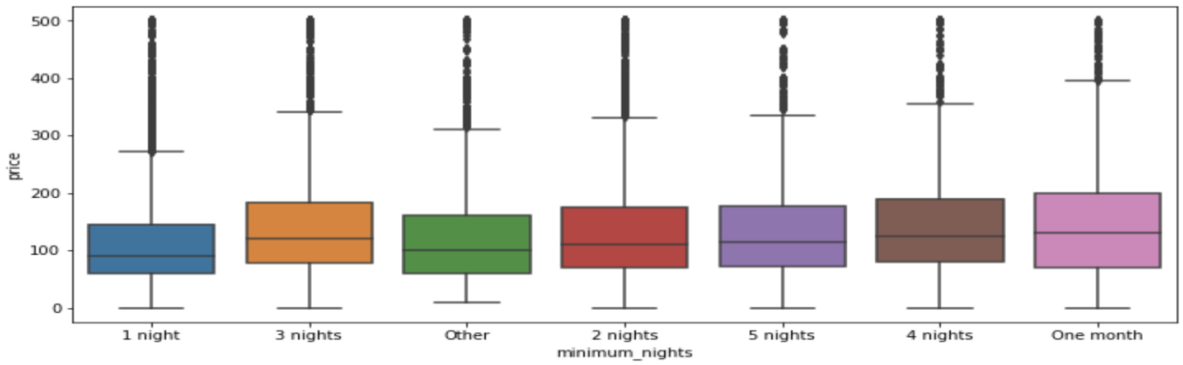
Neighborhood Groups:

In our dataset, there are 5 different neighborhood groups, which represented the five boroughs of New York City. They are Manhattan, Brooklyn, Staten Island, Queens and the Bronx. While calculating the mean prices of these various neighborhoods, we could see a considerable difference between all of them, including Brooklyn & Manhattan. And in turn, Brooklyn & Manhattan were leaps and bounds ahead of the other neighborhood groups in terms of pricing. Out of the one thousand odd listings in the Bronx, the mean came at around $81. For Brooklyn, which had around twenty thousand listings, the mean was around $112. Manhattan was the most expensive of the lot with a mean price of $163 from its more than twenty thousand listings. Staten Island being isolated, had the least number of listings at three hundred and sixty seven, with a mean of $91. And lastly, we had Queens with over five thousand five hundred listings with a mean price of around $92. Therefore an AirBnB in Brooklyn or Manhattan should be priced higher than an AirBnB in any of the other neighborhood groups owing to the customer preference.

Minimum Nights:

As the heading says, minimum nights was a requirement set up by hosts with respect to the

least number of nights a guest has to stay at the property to rent it out. Minimum nights was a numerical variable. Since we had a lot of values that were repeating themselves such as three nights for small listings and one month for large properties, we decided to convert some of the data into a categorical value and treat the others as a category. In the box plot that we plotted for the minimum nights -



Listings that had a minimum night requirement for one month and three nights seem to be on the higher portion of the price spectrum while those that had a minimum requirement of one night seemed to be on the lower end.

Number of Reviews:

During our analysis, we noticed that a lot of Airbnb properties did not have any reviews. These were properties that weren't frequently sought by customers either. This makes sense since from a customer's point of view, they would want to stay at a property where both the listing and the owner had some credibility. So when we think of suggesting to a new investor on how they should have their property, getting in reviews from customers is going to be a very important aspect here. Also, when we plotted the graphs, we saw a negative correlation between the price and the number of reviews. More the number of reviews, lower the price

seemed to be. This also makes sense since most customers are economical and would want to spend as little as possible while trying to spend most during their experiences exploring the city.

Availability:

In this variable, we saw that the majority of the dataset had listings that were available throughout the year. Some of them, such as the waterfront properties in Staten Island, would not be available during the winter months when it's snowing. We could also see here that Airbnbs that did not have year round availability, tended to be on the lower price range to attract more crowds during the times that they were open. They had to do this since they would have a lesser operational duration as compared to other listings and if they decided to keep pricing high, they wouldn't be able to attract a lot of customers.

Neighborhoods:

The last variable which we look at predicting as the best case scenario is the neighborhood in which the listing should be. The areas of Midtown Manhattan, West Village and the downtown financial district have a higher price range compared to all other areas in all other neighborhoods. All three of these areas are in the Manhattan borough of the city. Sunnyside, Bushwick and Ridgewood are areas that have the lowest prices amongst all areas in all neighborhoods. The area of Sunnyside is in the neighborhood of Queens. The area of Bushwick is in the neighborhood of Brooklyn and Ridgewood is in Queens.

Modeling:

Building up on the previous dataset engineering, we decided to predict the prices for the airbnb using regression models such as linear regression, random forest and svm and to figure out which one would be the best model.

Since we decided to use regression, there were a couple of steps that we had to do before we could start the analysis.

The first step that we had to take was to use dummy variables for categorical variables.

Dummy variables for categorical variables

```
data=pd.get_dummies(data,drop_first=True,columns=['neighbourhood_group', 'neighbourhood', 'room_type',
       'minimum_nights', 'number_of_reviews','availability_365'])
```

First, we need to understand what categorical variables are. Categorical variables (also known as factor or qualitative variables) are variables that classify observations into groups. Regression analysis requires numerical variables. So, when a researcher wishes to include a categorical variable in a regression model, additional steps are required to make the results interpretable. Therefore, we use dummy variables. Technically, dummy variables are quantitative variables. Their range of values is small; they can take on only two quantitative values. As a practical matter, regression results are easiest to interpret when dummy variables are limited to two specific values, 1 or 0.Typically, 1 represents the presence of a qualitative attribute, and 0 represents the absence.

Once a categorical variable has been recorded as a dummy variable, the dummy variable can be used in regression analysis just like any other quantitative variable.

The next step that we did was to separate X & Y for training. So, when we're working on modeling and we want to train it, we obviously need a dataset. However, after training, we need to test the model on a dataset. For this, we need a dataset which is different from the one used before in training. But since it's not always possible to get a different dataset, we use another way around this. What we do is, we split the dataset into two sets. One for training & one for testing. This needs to be done before we start training and when it comes to data splitting, there are a few parameters that we need to know.

Seperate X and Y for training

```
X=data.drop(["id","price"],axis=1)
y=data["price"]
```

```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.33,random_state=100)
```

test_size — This parameter decides the size of the data that has to be split as the test dataset. This is given as a fraction. For example, if we pass 0.5 as the value, the dataset will be split 50% as the test dataset. If we specify this parameter, we can ignore the next parameter.

train_size — We have to specify this parameter only if we're not specifying the test_size.

random_state — Here we pass an integer, which will act as the seed for the random number generator during the split. Or, we can also pass an instance of the RandomState class, which will become the number generator. If we don't pass anything, the RandomState instance used by np.random will be used instead.

Next, we conducted linear regression. a simple linear regression is a linear regression model with a single explanatory variable. Based on this, we were able to find out top 5 positive and negative influencers who were affecting prices of the listings. The top 5 negative

were, 1) room_type_shared room, 2) room_type_private room, 3) neighborhood_washington heights, 4)neighborhood_harlem & 5)neighborhood_morningside heights. The top 5 positive are 1)number_of_reviews_unreviewed, 2)neighborhood_group_brooklyn, 3)neighborhood_midtown, 4)availability_365_available_whole_year & 5)neighborhood_group_manhattan.

Linear Regression

```
reg=LinearRegression()
reg.fit(X_train,y_train)
y_pred=reg.predict(X_test)
```

```
reg_coefs=pd.Series(dict(zip(X.columns.tolist(),reg.coef_.tolist()))).sort_values()
print("\nTop 5 strong negative influencers")
print(reg_coefs[:5].to_frame())
print("\nTop 5 strong positive influencers")
print(reg_coefs[-5:].to_frame())
```

For the simple linear regression, we calculated the RMSE and MAE.

```
print("RMSE : "+str(sqrt(mean_squared_error(y_test,y_pred))))
print("MAE : "+str(mean_absolute_error(y_test,y_pred)))
```

```
RMSE : 65.38906167000958
MAE : 43.875837104036826
```

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit. MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It's the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.

The next type of regression that we conducted was a random forest regression. Random forest is a type of a learning algorithm which uses combined learning methods for classification and regression. Random forest is more a bagging technique and not essentially a boosting technique. The trees in random forests are run in parallel.

```
from sklearn.ensemble import RandomForestRegressor
reg1=RandomForestRegressor()
reg1.fit(X_train,y_train)
y_pred1=reg1.predict(X_test)
```

```
print("RMSE : "+str(sqrt(mean_squared_error(y_test,y_pred1))))
print("MAE : "+str(mean_absolute_error(y_test,y_pred1)))
```

```
RMSE : 66.05799438835079
MAE : 43.06709592277255
```

And the last regression method that we used was SVM. SVM or support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new text.

```
reg2=SVR()
reg2.fit(X_train,y_train)
y_pred2=reg2.predict(X_test)
```

Now, based on the three regression methods, simple linear regression was the best option. Low MAE implies low Bias in the model. This is desirable. Low RMSE means high precision which is also desirable. Therefore, we chose the model with lowest bias and highest precision, which was simple linear regression.

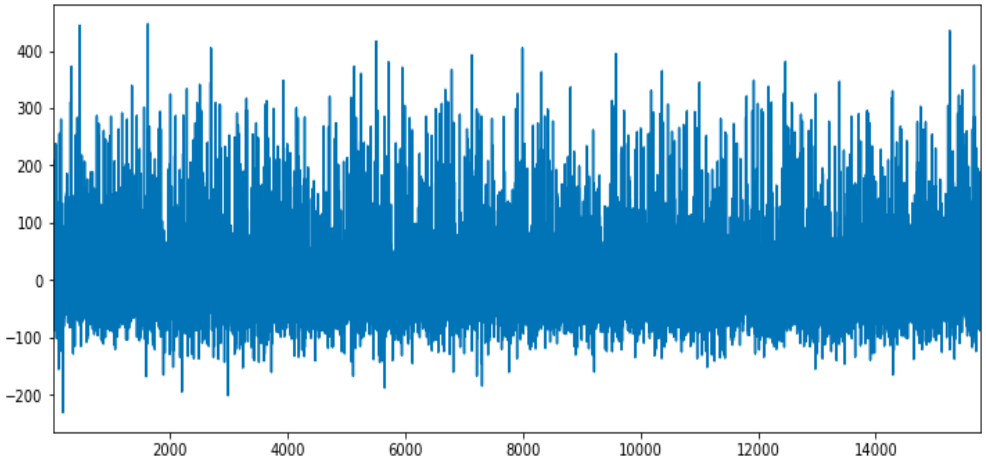
Residual Analysis:

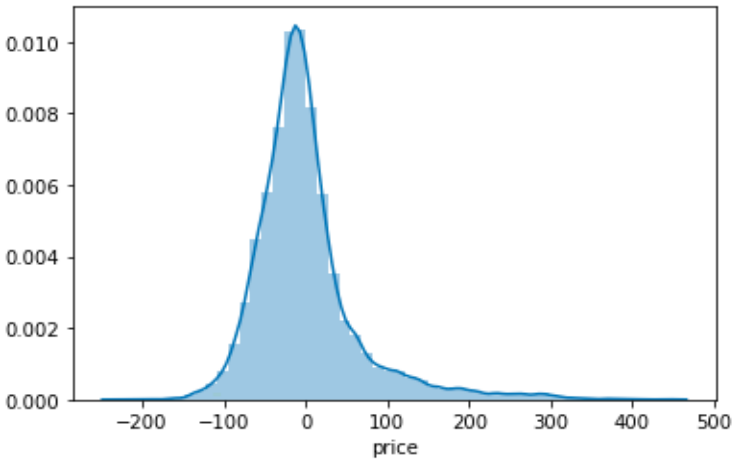
Just to make sure that simple linear regression was the best choice, we run the residual analysis through a LINE test.

1. **Linearity**: The relationship between X and the mean of Y is linear in nature.
2. **Independence**: The individual residuals are independent of each other.

3. **Normality**: For any fixed value of X, Y is normally distributed.

4. **Equal Variance**: The variance of residual is the same for any value of X.





Conclusion & Business Interpretation:

Through the analysis that our group conducted, the first thing that we ended up realizing that New York City in itself is like a small country. All the five boroughs of the city had such differences in the values of the categorical variables of the dataset. Manhattan, without a doubt was the top choice in terms of both, the number of listings on AirBnB as well as the number of customers coming into the city. Most of them want to stay in Manhattan for obvious reasons. Manhattan was followed by Brooklyn, then Queens, then Bronx and then Staten Island, which is the furthest of all the boroughs. The pricing of the listings in general differed a lot amongst the five boroughs of the city.

Based on the predictive analytics we did, linear regression seemed to be the best model for future analysis. So based on our findings, if a new person wants to purchase a property to list on AirBnB in New York City, the ideal variables for this would be the borough being Manhattan, the property being available through the year, the property booking to be for a minimum of three nights and lastly, they need to focus on getting a good number of reviews for their listings. Customers even while visiting New York City are opting for cheap housing options in the two most popular boroughs, even though the cheapest options were in the Bronx, most people did not opt for it.

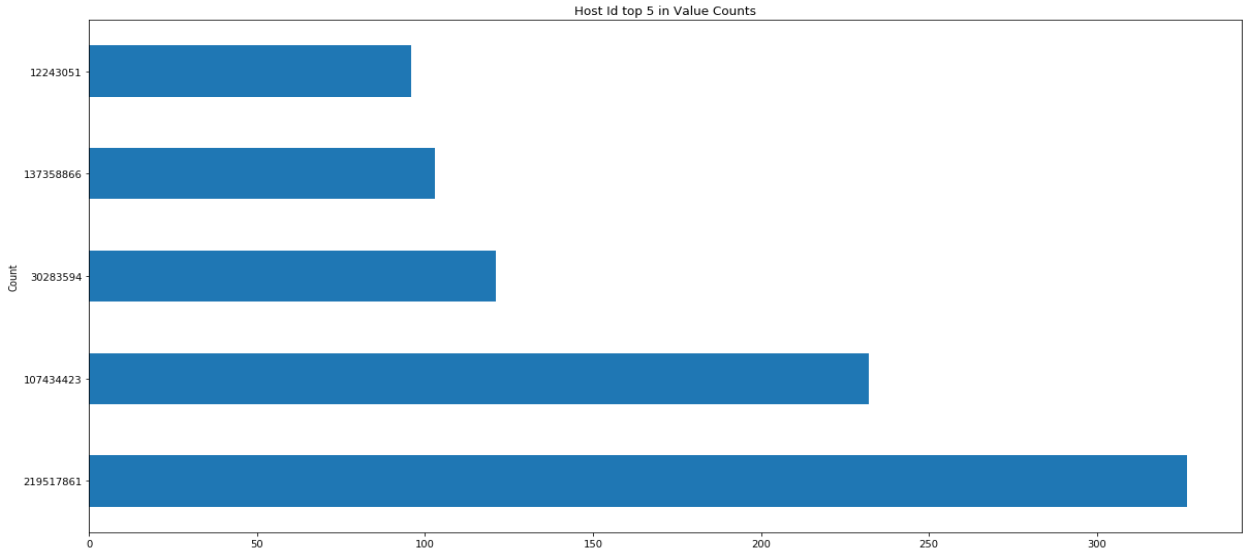
References

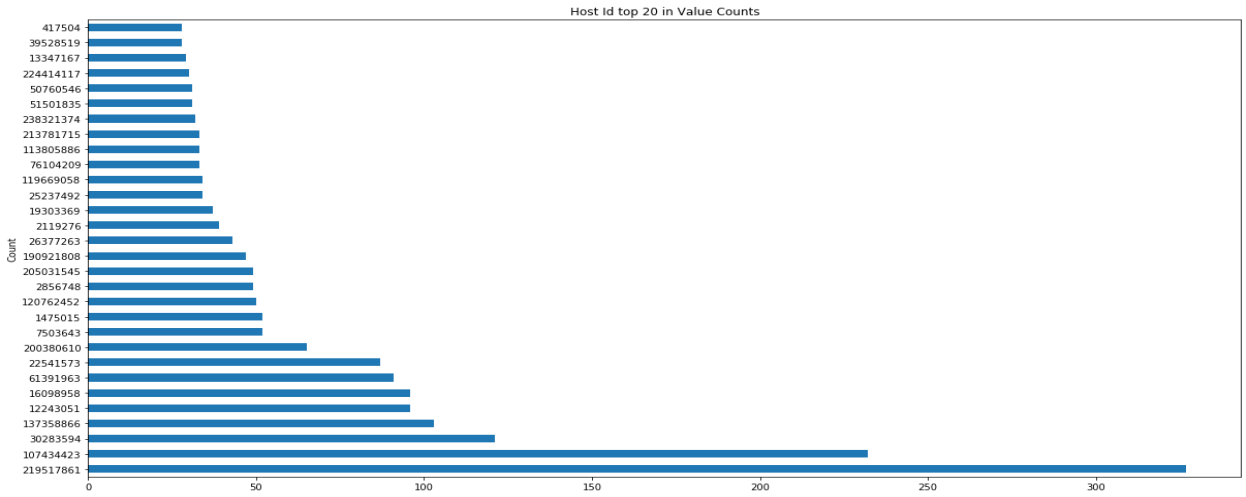
1. https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data
2. **https://data.cityofnewyork.us/City-Government/Neighborhood-Tabulation-Areas/cpf4-rkhq**
3. **https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49**
4. **http://insideairbnb.com/get-the-data.html**

Appendices:

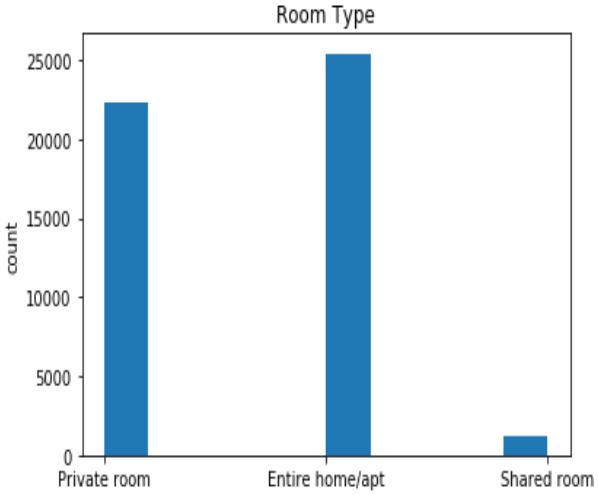
Appendix - Part 1 Visualizations

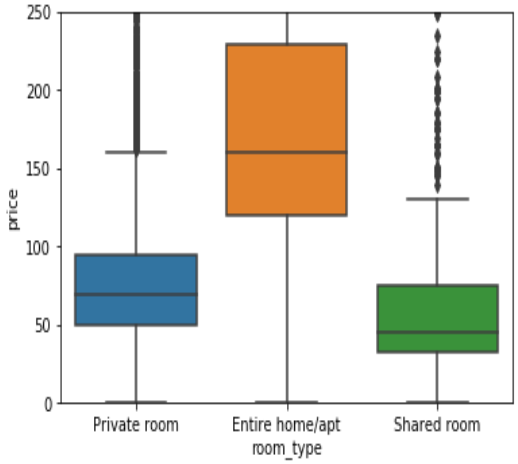
Host ID



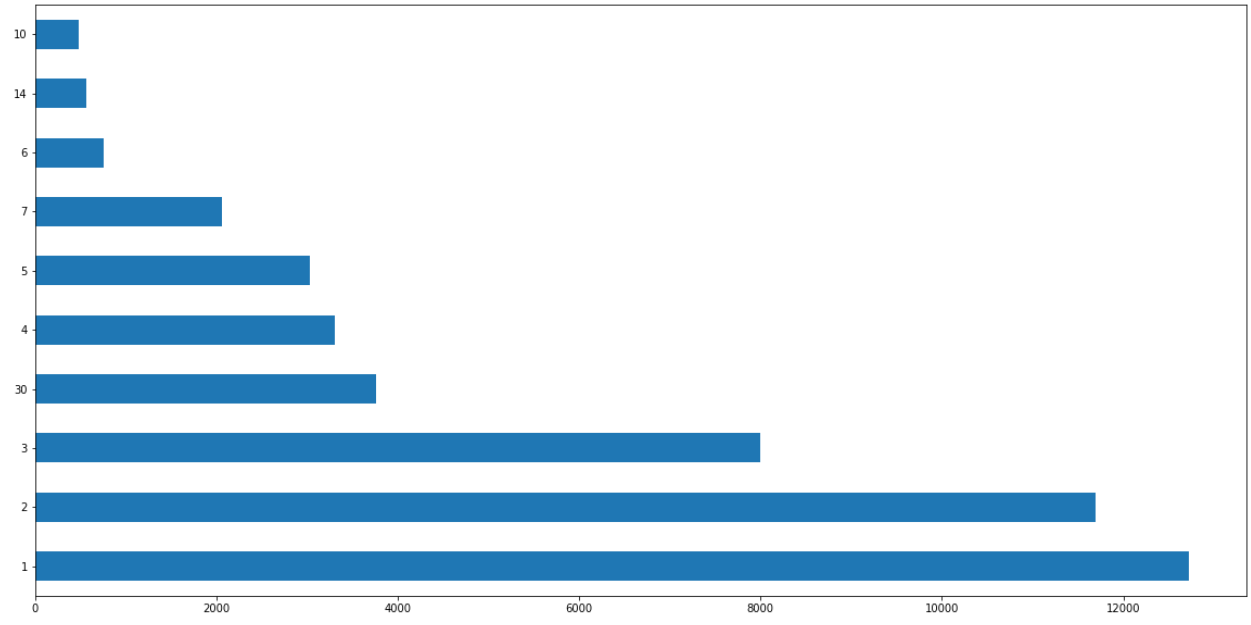


Room Type

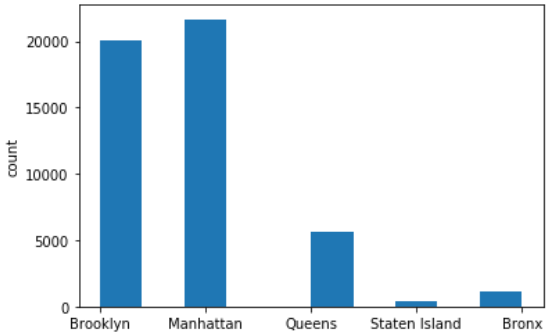


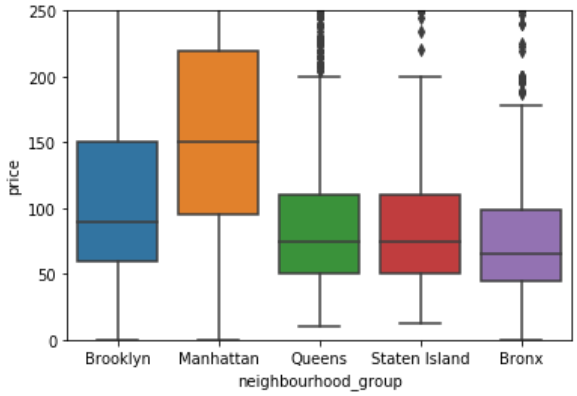


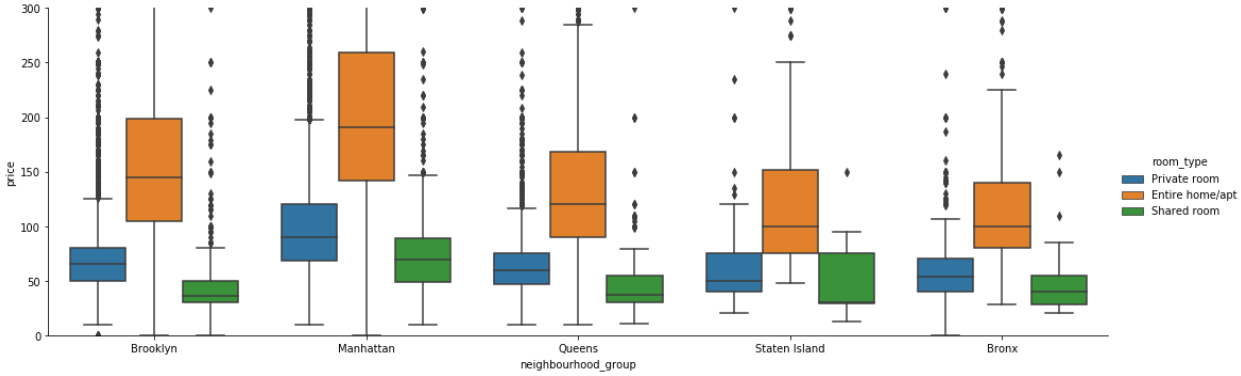
Minimum Nights

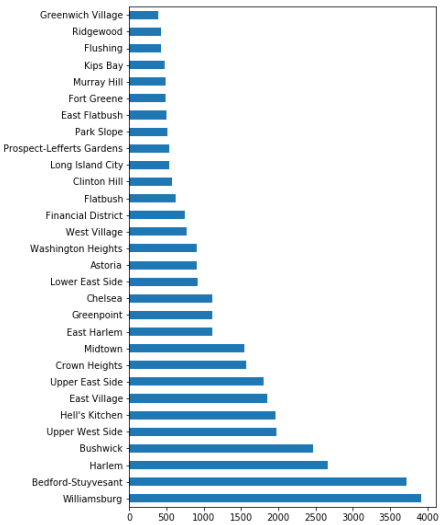


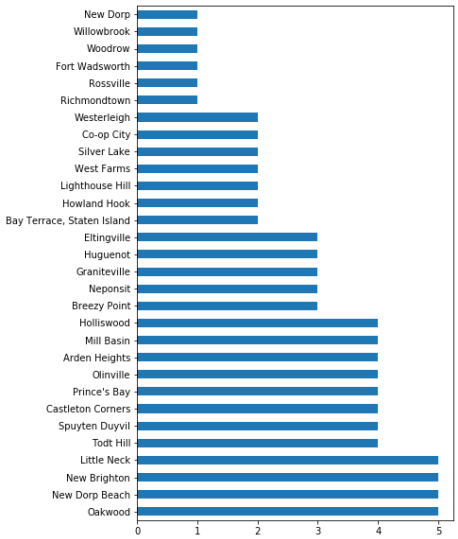
Neighborhood Groups





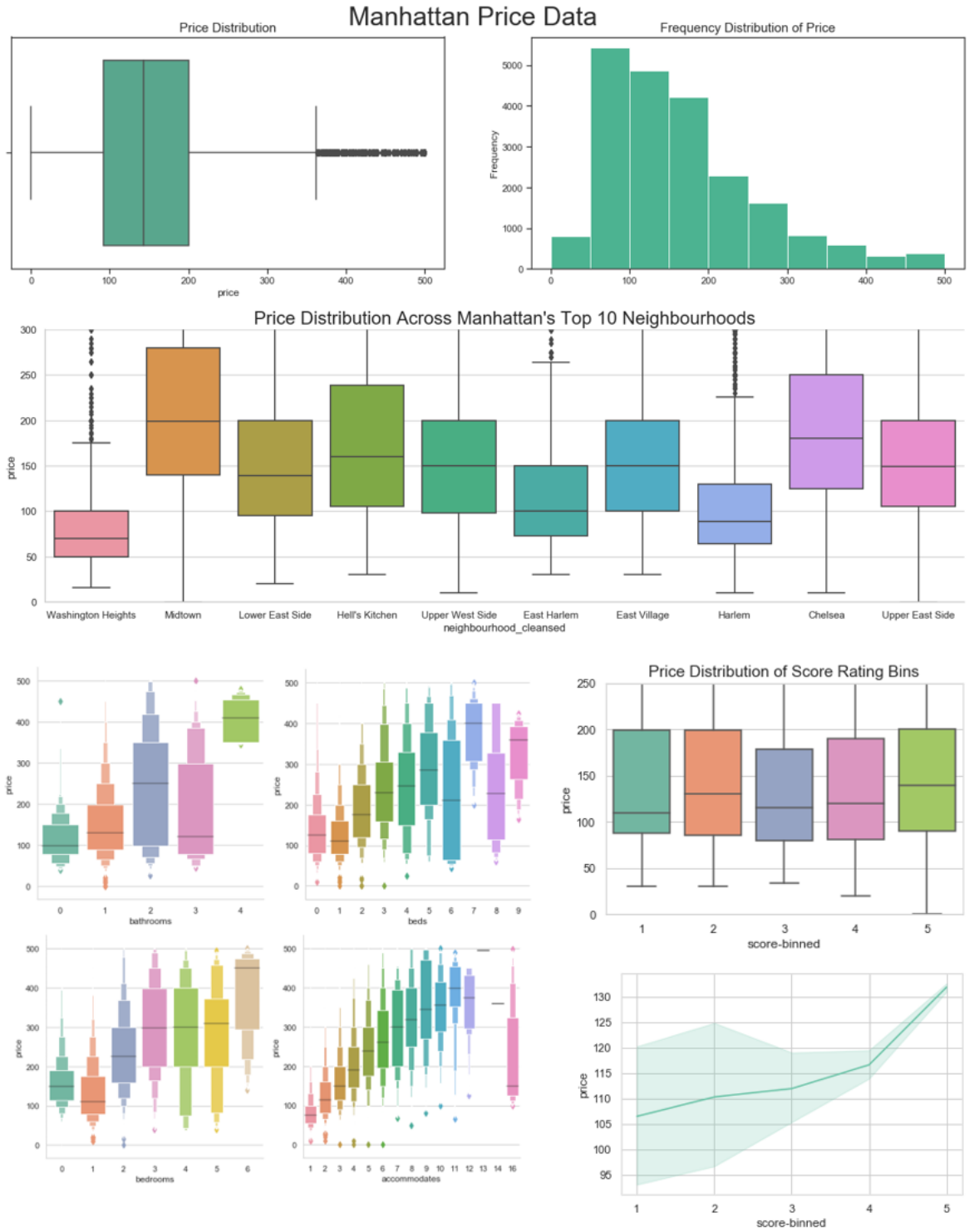






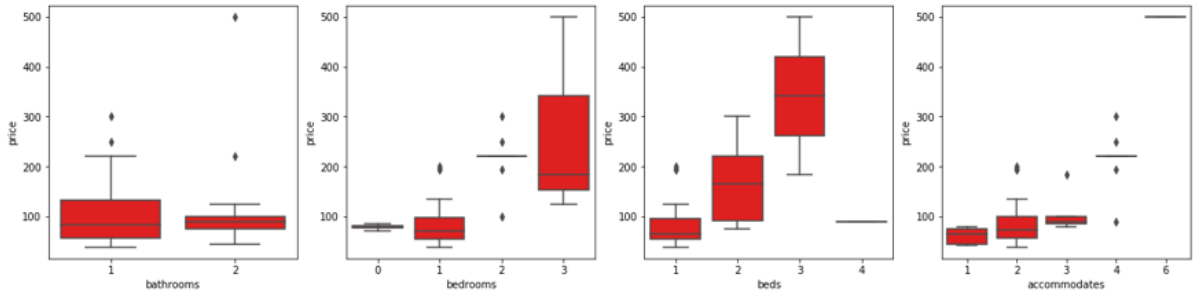
Appendix - Part 2 Objective #1

Manhattan

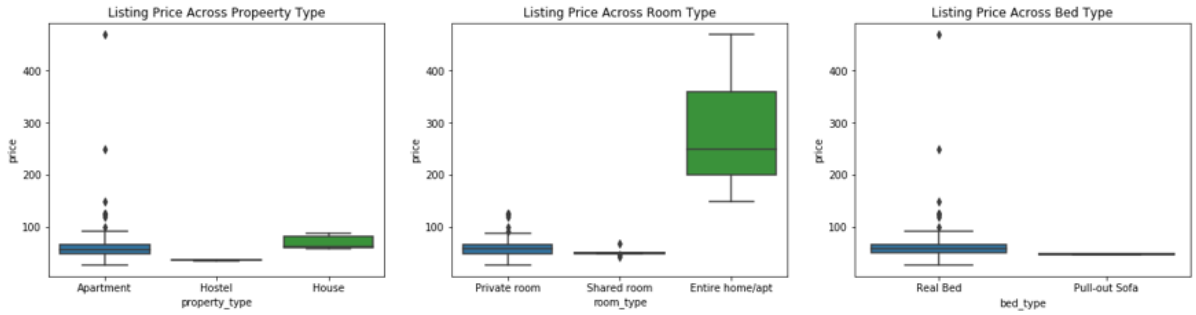


Manhattan: Harlem Profile of Top Ten Hosts

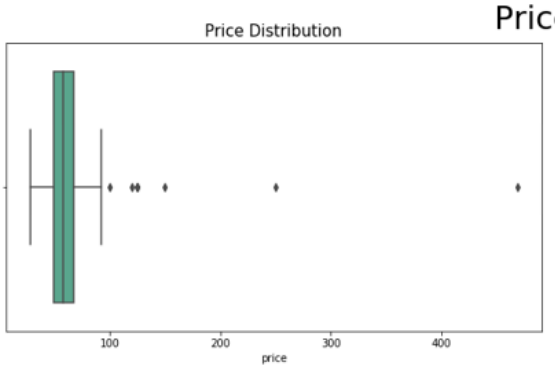
Listings' Physical Attributes Relationship to Price

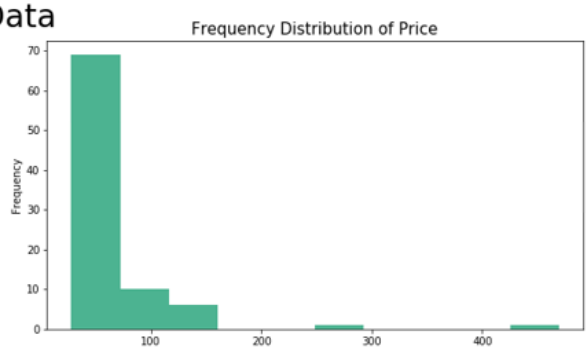


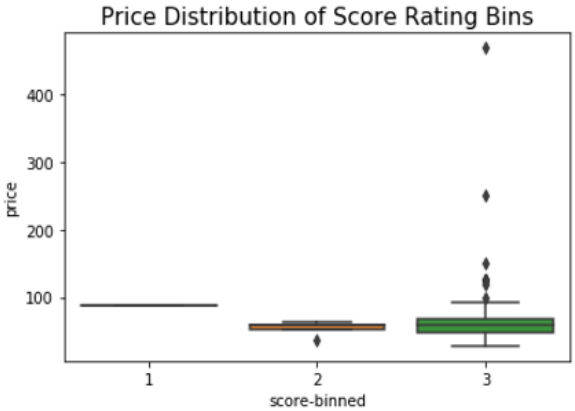
Listings' Physical Attributes Relationship to Price



Price Data



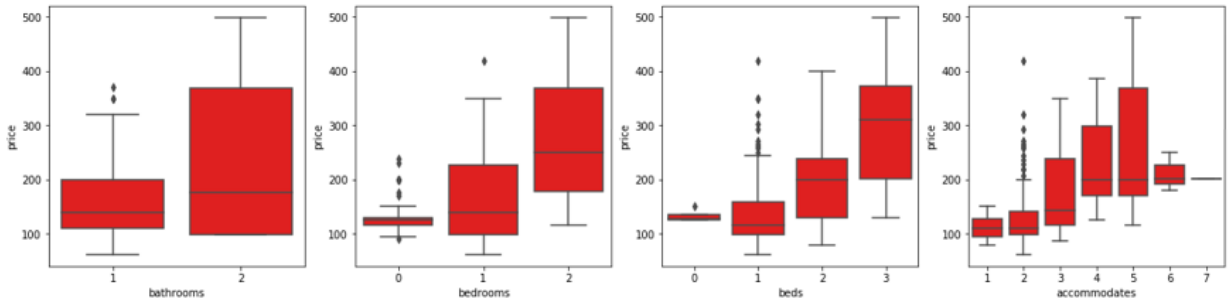




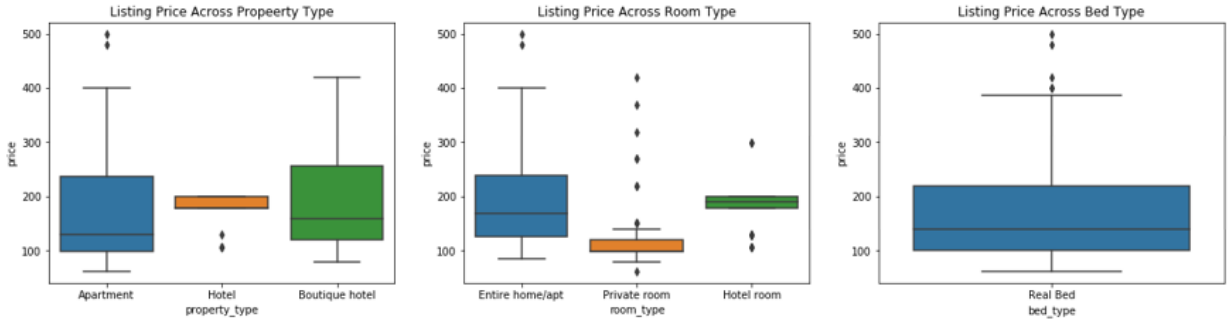
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	87.000000	87.000000	87.000000	87.000000	87.000000	87.000000	87.000000
mean	2.310345	1.218391	1.091954	1.494253	68.448276	1.195402	89.741167
std	1.564953	0.468180	0.497253	1.021553	53.475436	0.833026	11.450738
min	1.000000	0.000000	1.000000	0.000000	28.000000	1.000000	20.000000
25%	1.500000	1.000000	1.000000	1.000000	49.000000	1.000000	88.000000
50%	2.000000	1.000000	1.000000	1.000000	58.000000	1.000000	93.903137
75%	2.000000	1.000000	1.000000	2.000000	67.000000	1.000000	93.903137
max	10.000000	2.000000	4.000000	6.000000	469.000000	6.000000	100.000000

Manhattan: Hell's Kitchen Profile of Top Ten Hosts

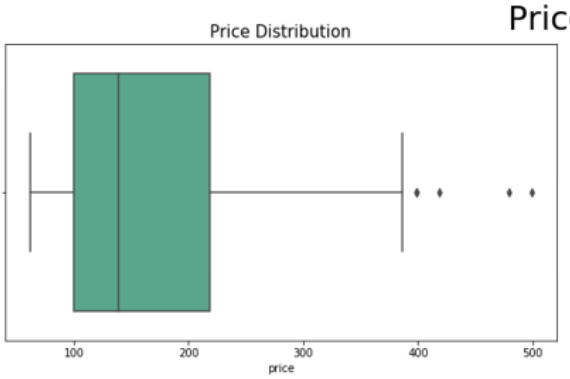
Listings' Physical Attributes Relationship to Price

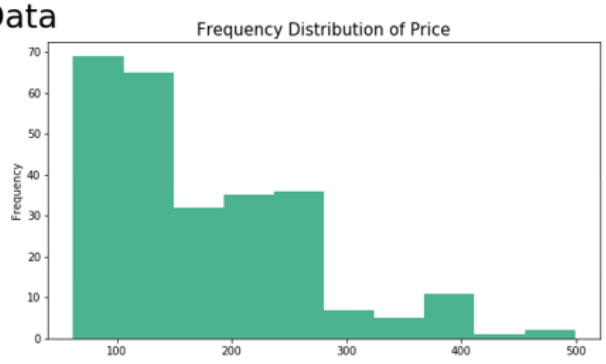


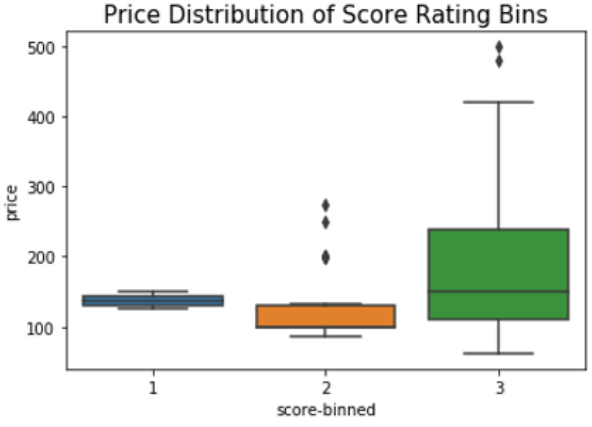
Listings' Physical Attributes Relationship to Price



Price Data



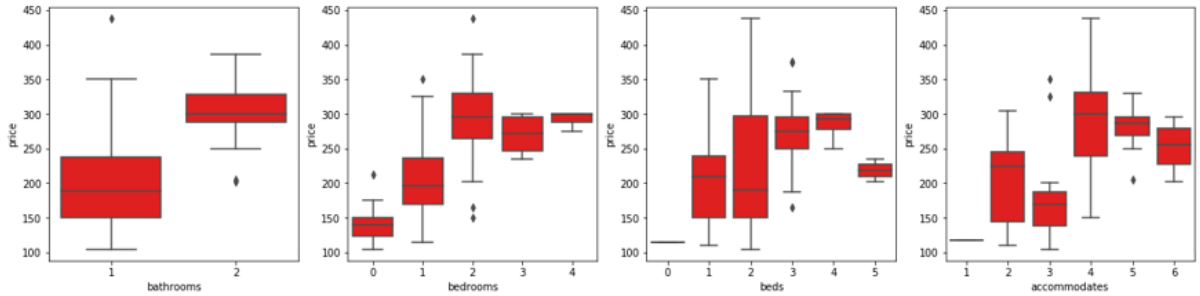




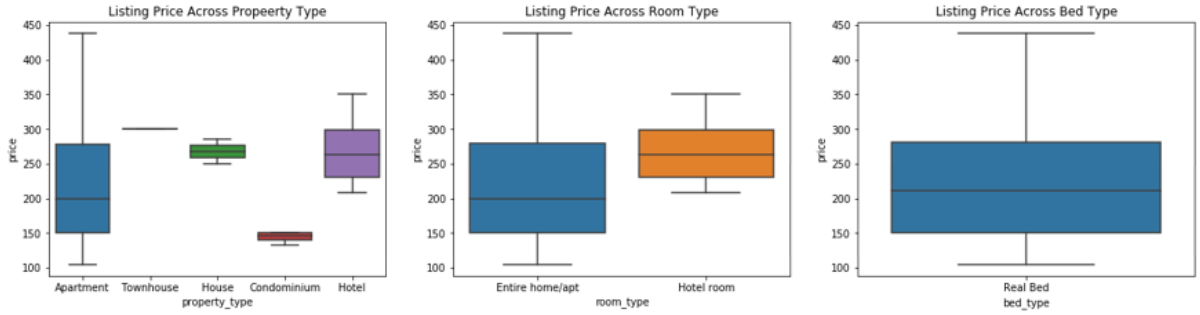
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	263.000000	263.000000	263.000000	263.000000	263.000000	263.000000	263.000000
mean	2.920152	1.167300	0.935361	1.475285	172.566540	1.368821	88.945642
std	1.079516	0.373955	0.629737	0.634833	83.218066	0.702298	11.789967
min	1.000000	1.000000	0.000000	0.000000	62.000000	1.000000	20.000000
25%	2.000000	1.000000	1.000000	1.000000	100.000000	1.000000	83.000000
50%	3.000000	1.000000	1.000000	1.000000	139.000000	1.000000	93.903137
75%	3.000000	1.000000	1.000000	2.000000	219.000000	2.000000	96.000000
max	7.000000	2.000000	2.000000	3.000000	499.000000	5.000000	100.000000

Manhattan: Upper West Side Profile of Top Ten Hosts

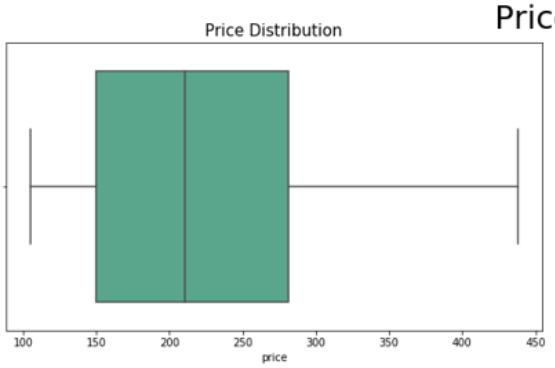
Listings' Physical Attributes Relationship to Price

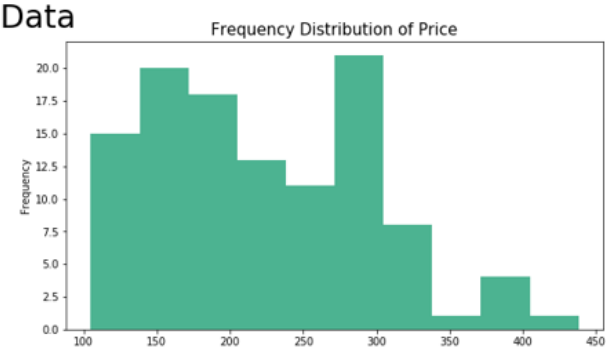


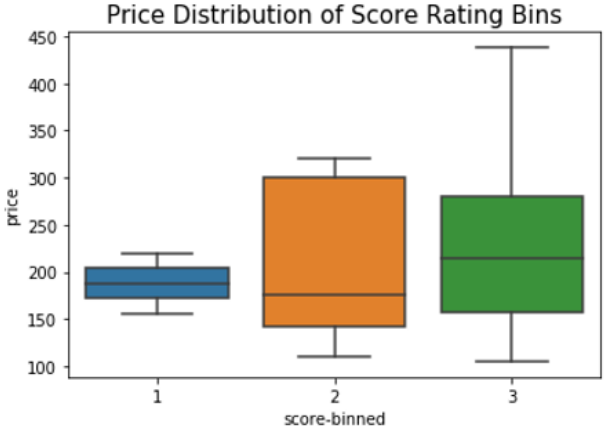
Listings' Physical Attributes Relationship to Price



Price Data



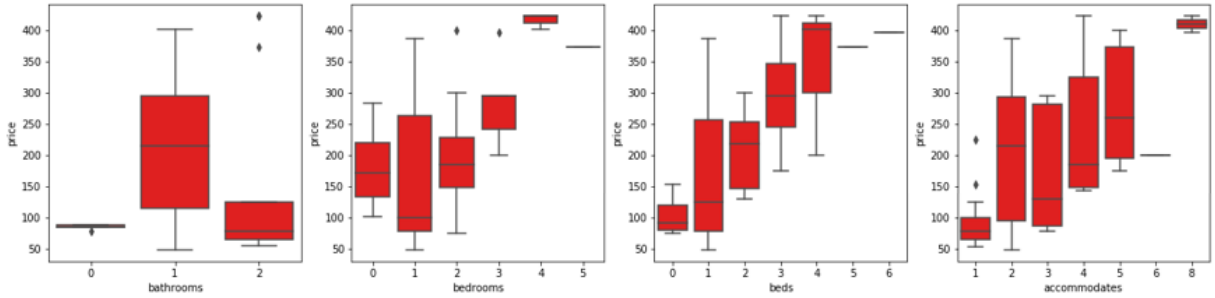




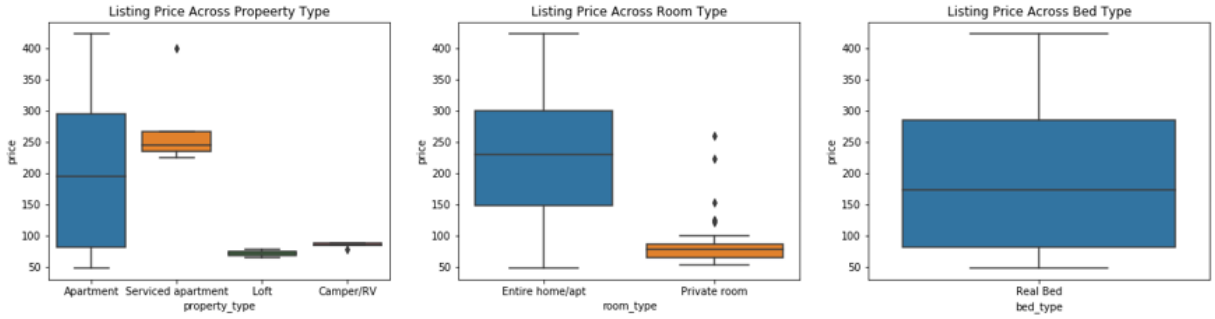
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	112.000000	112.000000	112.000000	112.000000	112.000000	112.000000	112.000000
mean	3.258929	1.232143	1.250000	1.875000	222.250000	1.660714	92.766383
std	1.121016	0.424097	0.864724	0.969118	74.649029	1.095063	6.794248
min	1.000000	1.000000	0.000000	0.000000	105.000000	1.000000	60.000000
25%	2.000000	1.000000	1.000000	1.000000	150.000000	1.000000	92.750000
50%	3.000000	1.000000	1.000000	2.000000	210.500000	1.000000	93.903137
75%	4.000000	1.000000	2.000000	2.000000	281.250000	2.000000	96.000000
max	6.000000	2.000000	4.000000	5.000000	438.000000	5.000000	100.000000

Manhattan: East Village Profile of Top Ten Hosts

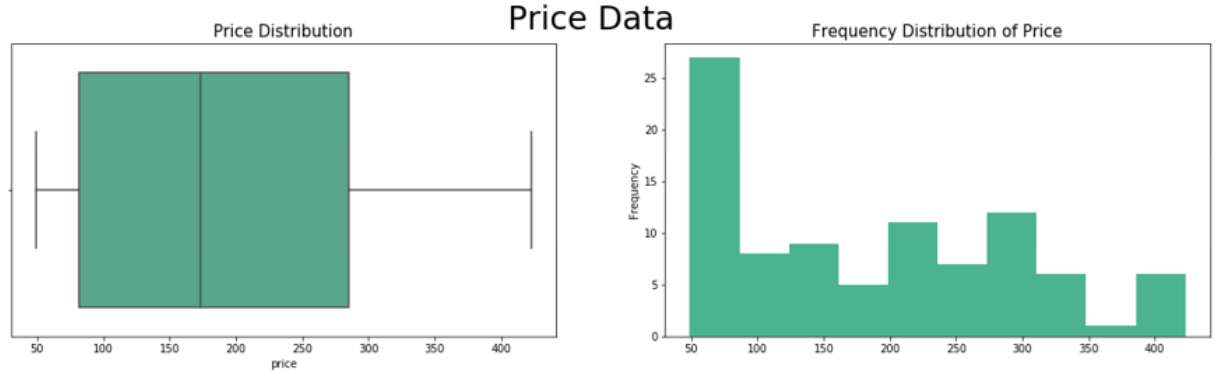
Listings' Physical Attributes Relationship to Price

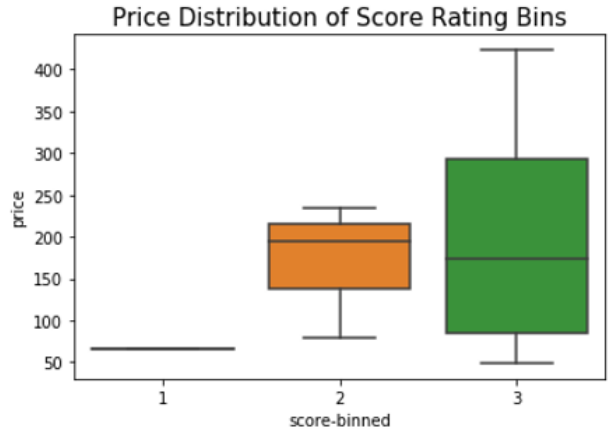


Listings' Physical Attributes Relationship to Price



Price Data

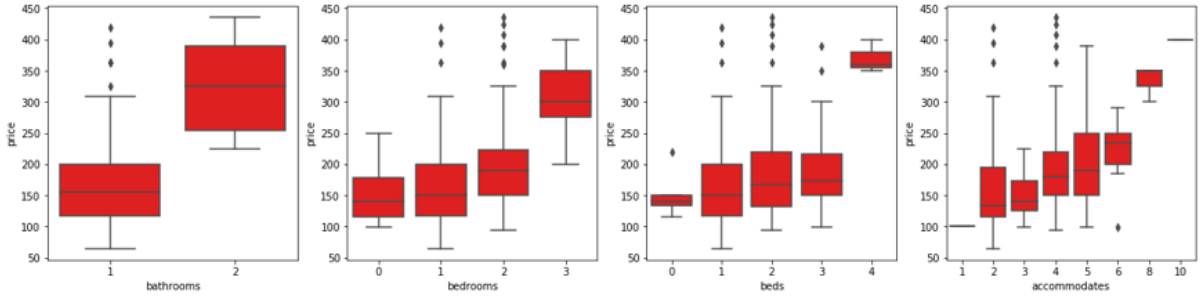




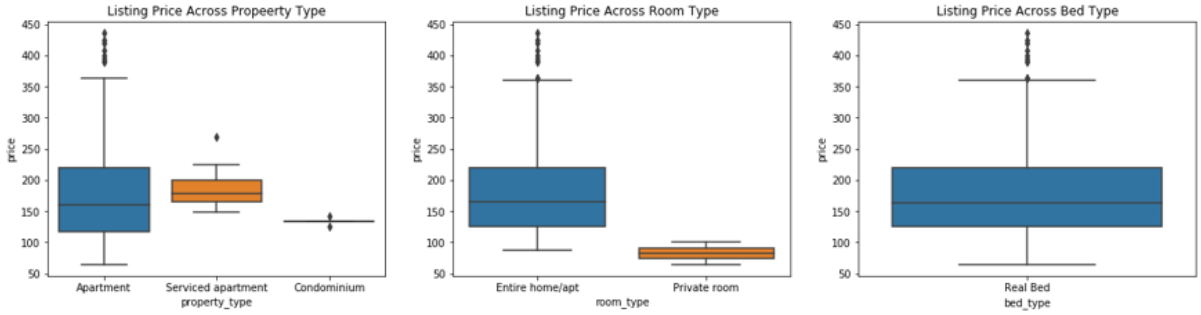
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	92.000000	92.000000	92.000000	92.000000	92.000000	92.000000	92.000000
mean	2.445652	1.086957	1.282609	1.402174	188.641304	1.173913	93.449804
std	1.385673	0.460582	0.953030	1.038382	108.156539	0.459543	4.725073
min	1.000000	0.000000	0.000000	0.000000	49.000000	1.000000	60.000000
25%	2.000000	1.000000	1.000000	1.000000	81.750000	1.000000	93.903137
50%	2.000000	1.000000	1.000000	1.000000	173.000000	1.000000	93.903137
75%	3.000000	1.000000	1.000000	1.250000	285.250000	1.000000	93.903137
max	8.000000	2.000000	5.000000	6.000000	423.000000	3.000000	100.000000

Manhattan: Upper East Side Profile of Top Ten Hosts

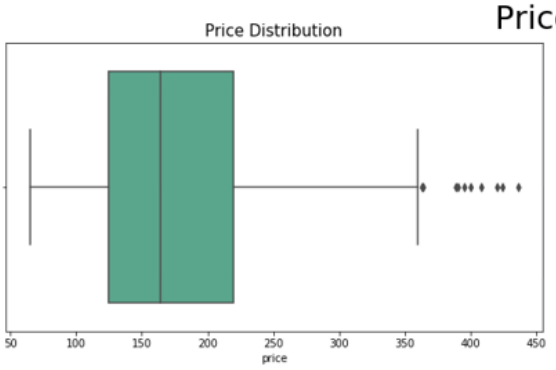
Listings' Physical Attributes Relationship to Price

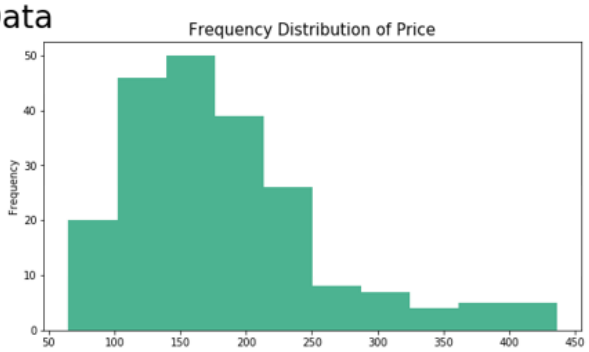


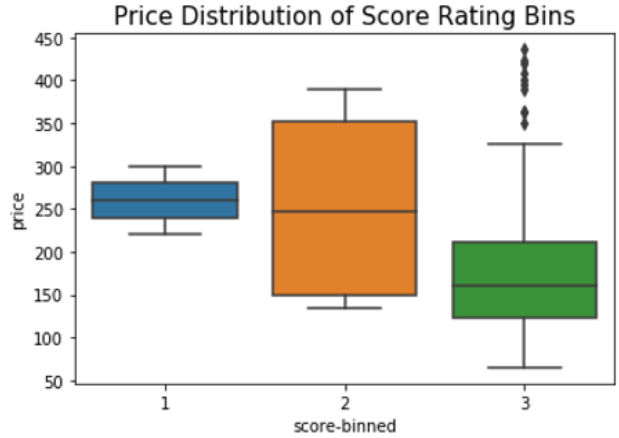
Listings' Physical Attributes Relationship to Price



Price Data







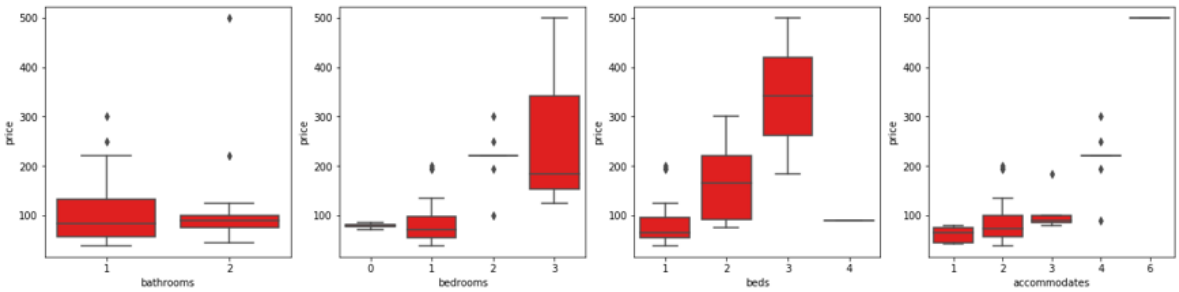
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	210.000000	210.000000	210.000000	210.000000	210.000000	210.000000	210.000000
mean	3.423810	1.085714	1.271429	1.742857	182.352381	1.523810	91.632995
std	1.402801	0.280611	0.781143	0.818782	75.406139	1.031515	10.341100
min	1.000000	1.000000	0.000000	0.000000	65.000000	1.000000	20.000000
25%	2.000000	1.000000	1.000000	1.000000	125.000000	1.000000	90.000000
50%	3.000000	1.000000	1.000000	2.000000	164.000000	1.000000	93.903137
75%	4.000000	1.000000	2.000000	2.000000	220.000000	2.000000	97.000000
max	10.000000	2.000000	3.000000	4.000000	436.000000	5.000000	100.000000

Brooklyn

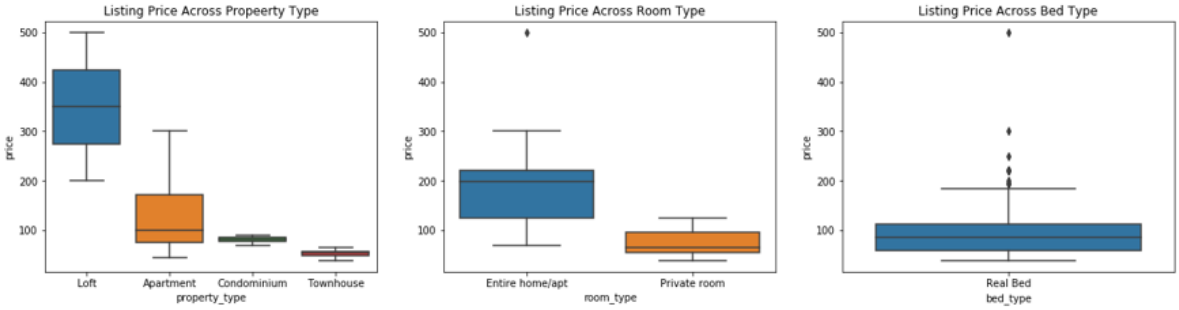

Booklyn Price Data
Price Distribution
price
Frequency Distribution of Price
Frequency
Price Distribution Across Brooklyn's Top 10 Neighbourhoods
price
Clinton Hill
Bedford-Stuyvesant
Williamsburg
Crown Heights
Park Slope
Bushwick
Flatbush
Prospect-Lefferts Gardens
Greenpoint
East Flatbush
neighbourhood_cleansed
bathrooms
accommodates
bedrooms
beds
Price Distribution of Score Rating Bins
score-binned

Brooklyn: Williamsburg Profile of Top Ten Hosts

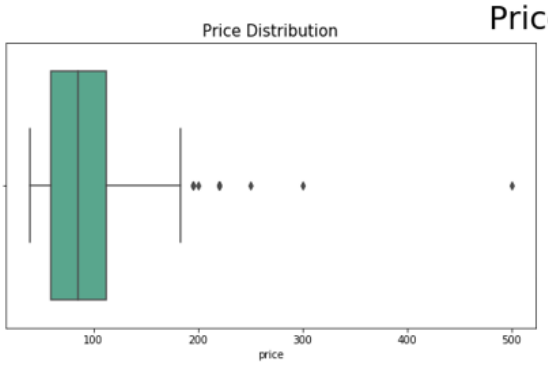
Listings' Physical Attributes Relationship to Price

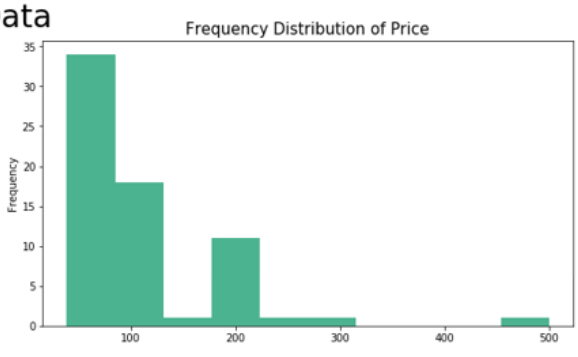


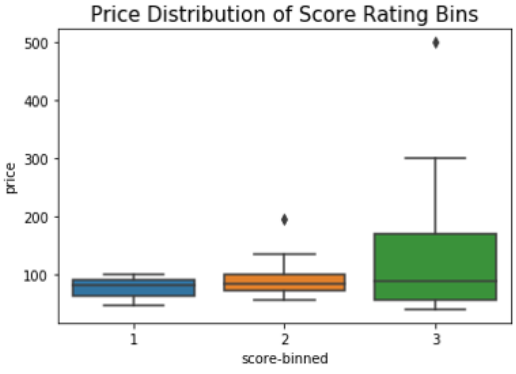
Listings' Physical Attributes Relationship to Price



Price Data



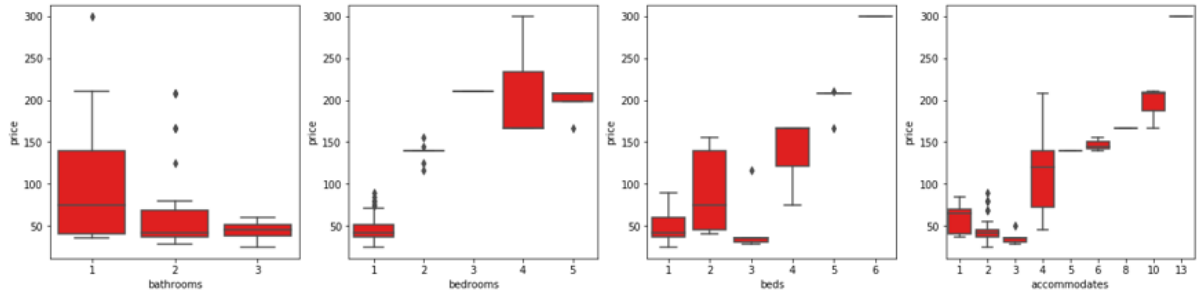




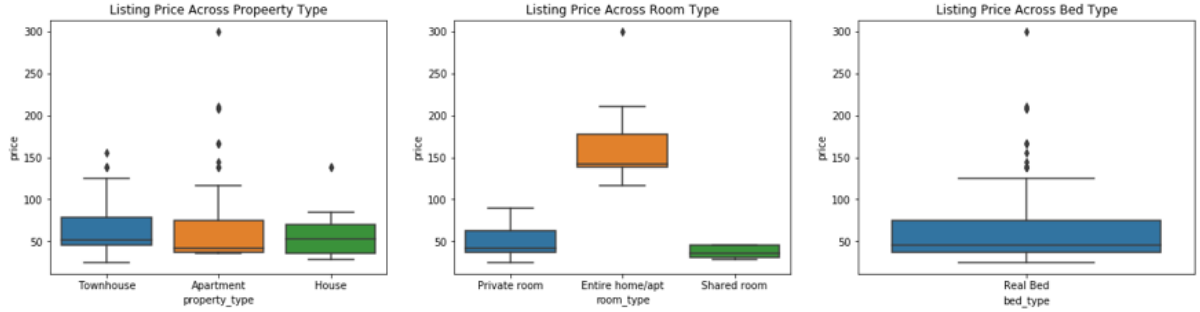
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000
mean	2.761062	1.858407	1.345133	1.681416	68.159292	1.327434	89.962284
std	2.071380	0.624889	0.913880	1.128095	50.150945	1.113573	13.860322
min	1.000000	1.000000	1.000000	1.000000	25.000000	1.000000	20.000000
25%	2.000000	1.000000	1.000000	1.000000	37.000000	1.000000	89.000000
50%	2.000000	2.000000	1.000000	1.000000	45.000000	1.000000	93.903137
75%	3.000000	2.000000	1.000000	2.000000	75.000000	1.000000	100.000000
max	13.000000	3.000000	5.000000	6.000000	300.000000	10.000000	100.000000

Brooklyn: Bedford- Stuyvesant Profile of Top Ten Hosts

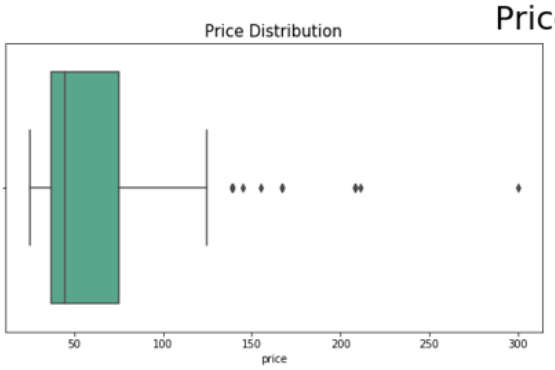
Listings' Physical Attributes Relationship to Price

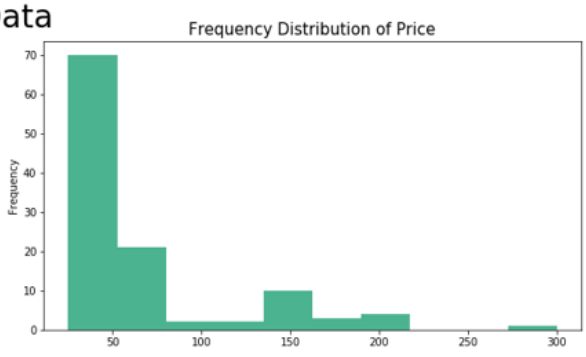


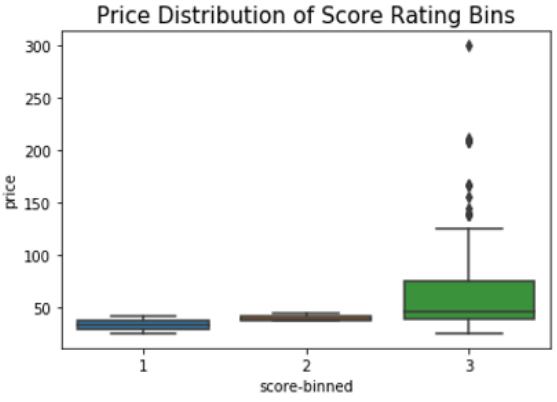
Listings' Physical Attributes Relationship to Price



Price Data



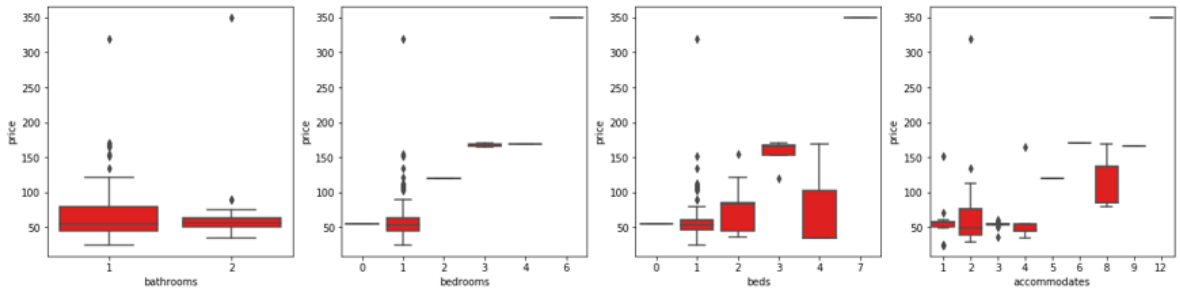




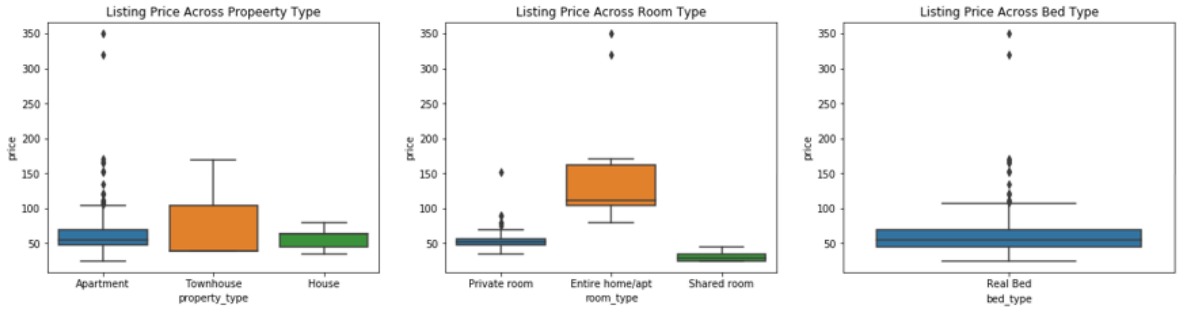
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000
mean	2.761062	1.858407	1.345133	1.681416	68.159292	1.327434	89.962284
std	2.071380	0.624889	0.913880	1.128095	50.150945	1.113573	13.860322
min	1.000000	1.000000	1.000000	1.000000	25.000000	1.000000	20.000000
25%	2.000000	1.000000	1.000000	1.000000	37.000000	1.000000	89.000000
50%	2.000000	2.000000	1.000000	1.000000	45.000000	1.000000	93.903137
75%	3.000000	2.000000	1.000000	2.000000	75.000000	1.000000	100.000000
max	13.000000	3.000000	5.000000	6.000000	300.000000	10.000000	100.000000

Brooklyn: Bushwick Profile of Top Ten Hosts

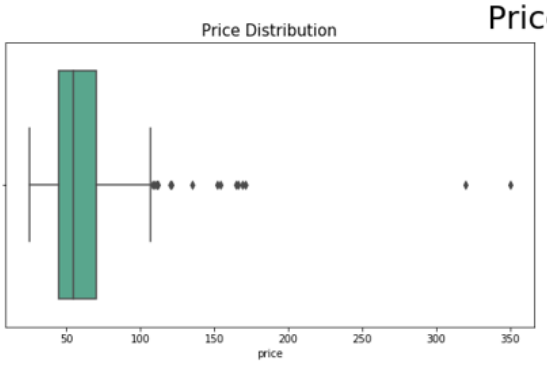
Listings' Physical Attributes Relationship to Price

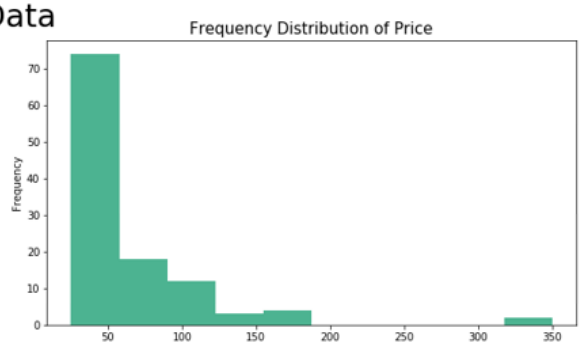


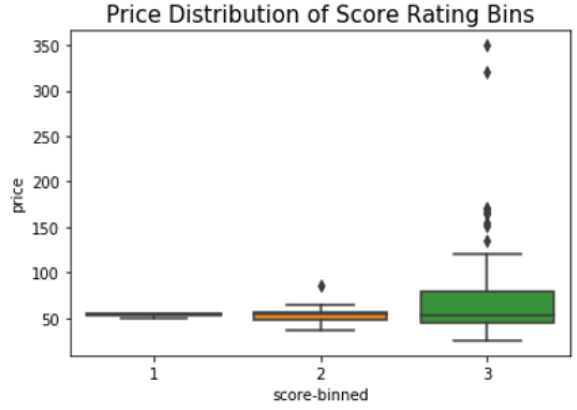
Listings' Physical Attributes Relationship to Price



Price Data



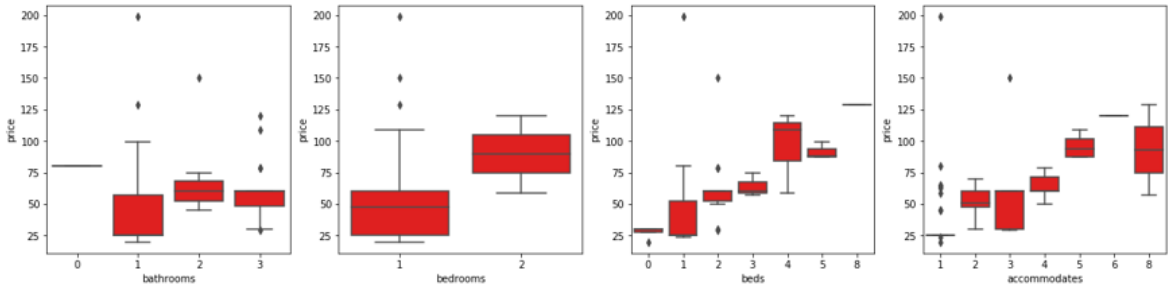




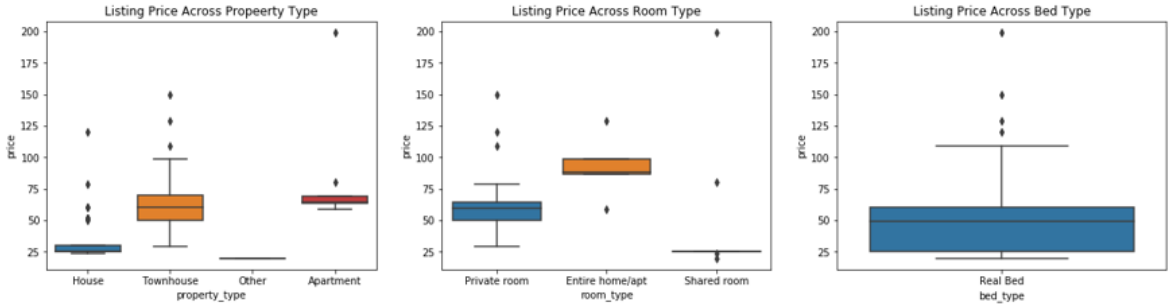
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000
mean	2.663717	1.212389	1.115044	1.265487	68.300885	1.362832	89.941735
std	2.002724	0.410821	0.651265	0.855838	48.594511	0.802473	12.843249
min	1.000000	1.000000	0.000000	0.000000	25.000000	1.000000	30.000000
25%	2.000000	1.000000	1.000000	1.000000	45.000000	1.000000	91.000000
50%	2.000000	1.000000	1.000000	1.000000	55.000000	1.000000	93.903137
75%	3.000000	1.000000	1.000000	1.000000	70.000000	2.000000	96.000000
max	12.000000	2.000000	6.000000	7.000000	350.000000	8.000000	100.000000

Brooklyn: Crown Heights Profile of Top Ten Hosts

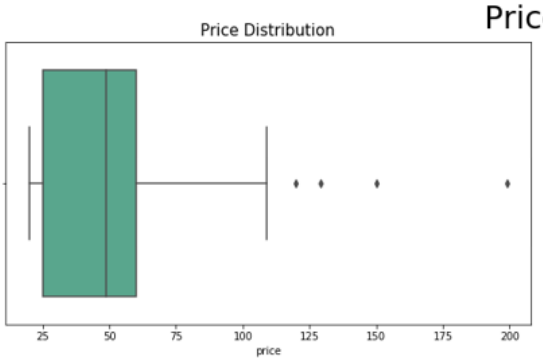
Listings' Physical Attributes Relationship to Price

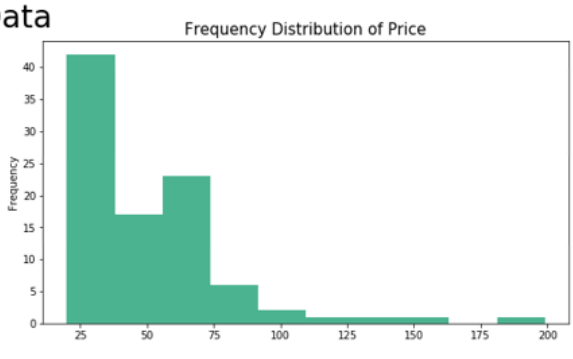


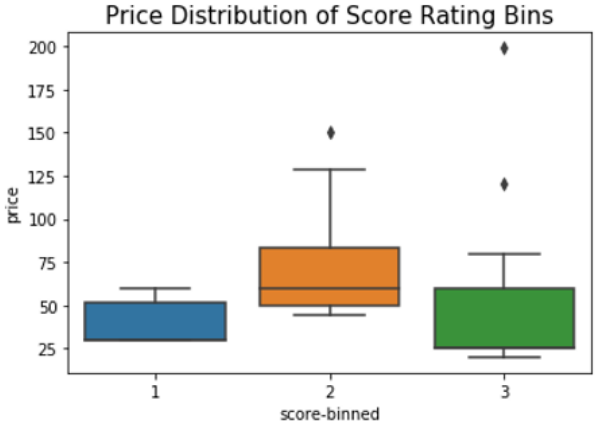
Listings' Physical Attributes Relationship to Price



Price Data



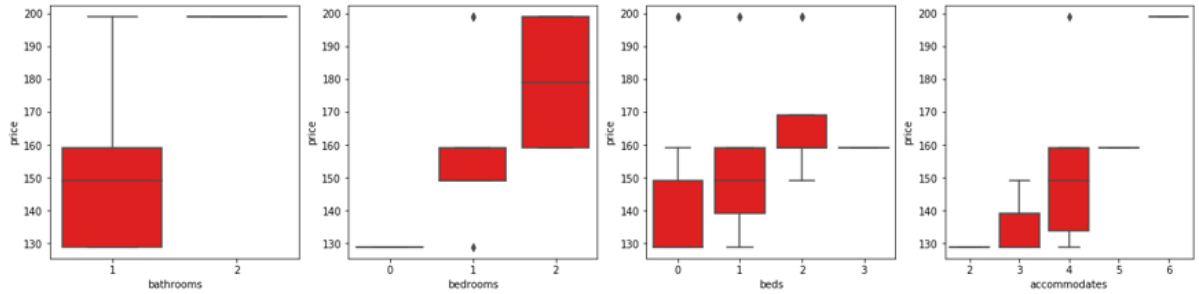




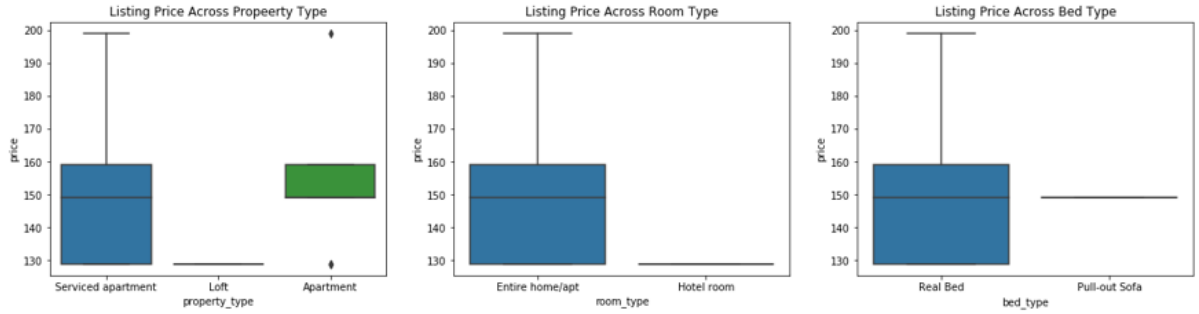
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000
mean	2.095745	1.500000	1.021277	1.468085	49.436170	1.159574	92.608748
std	1.510317	0.799866	0.145079	1.197687	30.042822	0.514378	4.361884
min	1.000000	0.000000	1.000000	0.000000	20.000000	1.000000	79.000000
25%	1.000000	1.000000	1.000000	1.000000	25.000000	1.000000	90.000000
50%	2.000000	1.000000	1.000000	1.000000	49.000000	1.000000	93.903137
75%	2.000000	2.000000	1.000000	1.750000	60.000000	1.000000	93.903137
max	8.000000	3.000000	2.000000	8.000000	199.000000	5.000000	100.000000

Brooklyn: Greenpoint Profile of Top Ten Hosts

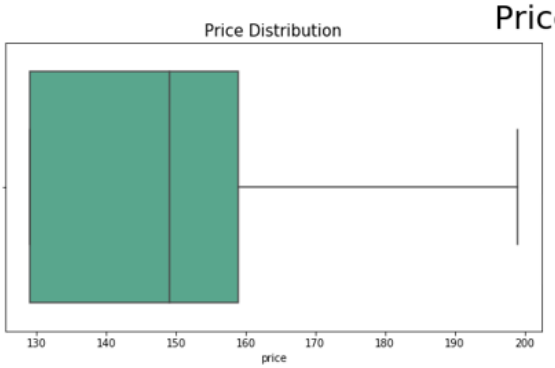
Listings' Physical Attributes Relationship to Price

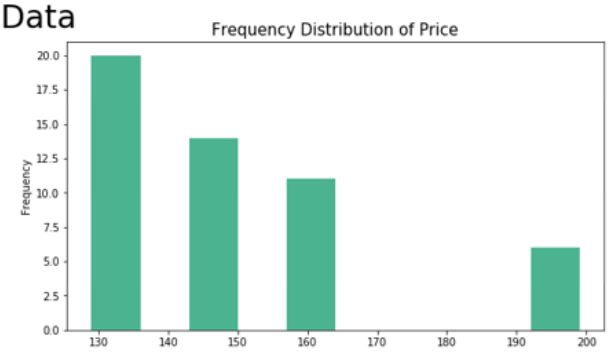


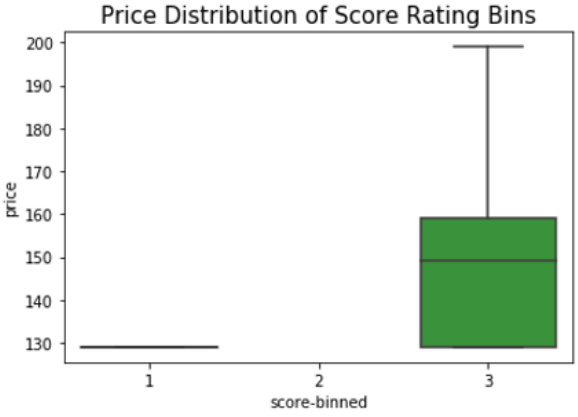
Listings' Physical Attributes Relationship to Price



Price Data

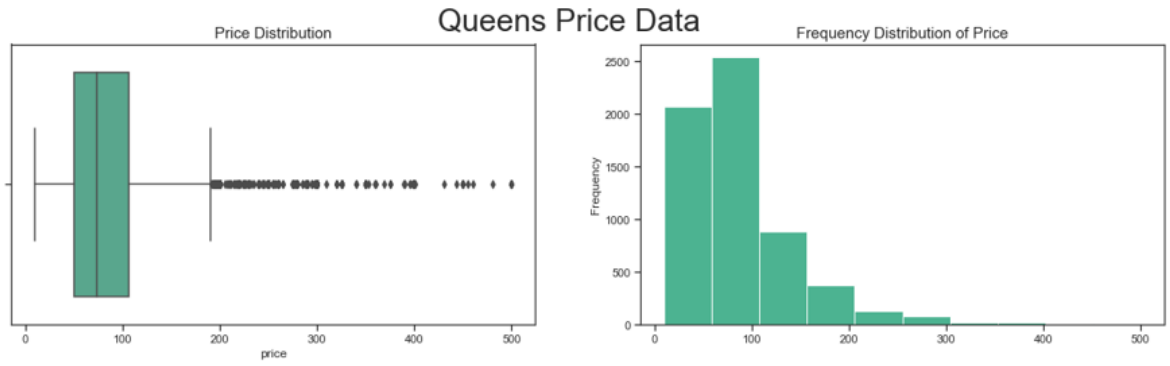


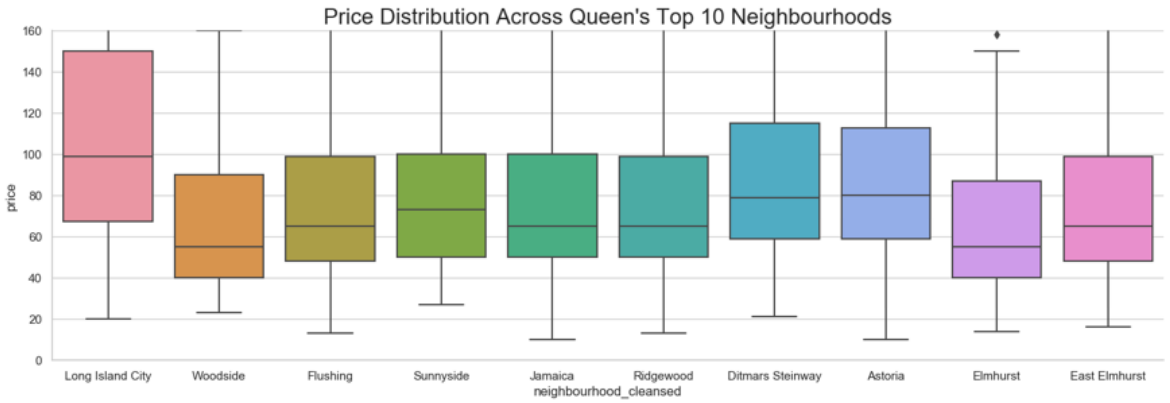


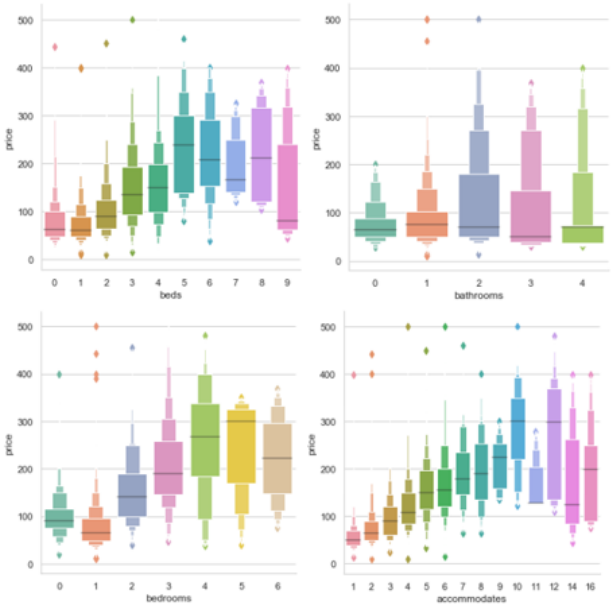


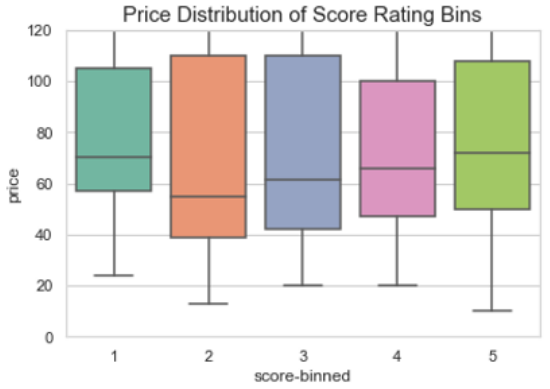
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	51.000000	51.000000	51.000000	51.000000	51.000000	51.0	51.000000
mean	3.803922	1.019608	0.784314	0.588235	149.196078	1.0	94.860846
std	1.131717	0.140028	0.701818	0.828890	21.953605	0.0	8.829740
min	2.000000	1.000000	0.000000	0.000000	129.000000	1.0	40.000000
25%	3.500000	1.000000	0.000000	0.000000	129.000000	1.0	92.500000
50%	4.000000	1.000000	1.000000	0.000000	149.000000	1.0	97.000000
75%	4.000000	1.000000	1.000000	1.000000	159.000000	1.0	100.000000
max	6.000000	2.000000	2.000000	3.000000	199.000000	1.0	100.000000

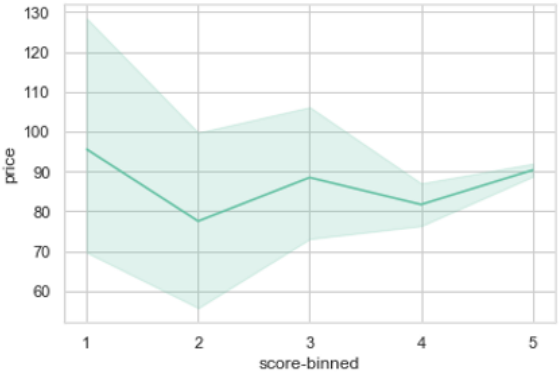
Queens


Queens Price Data
Price Distribution
price
Frequency Distribution of Price
Frequency


Price Distribution Across Queen's Top 10 Neighbourhoods
price
Long Island City
Woodside
Flushing
Sunnyside
Jamaica
Ridgewood
Ditmars Steinway
Astoria
Elmhurst
East Elmhurst
neighbourhood_cleansed

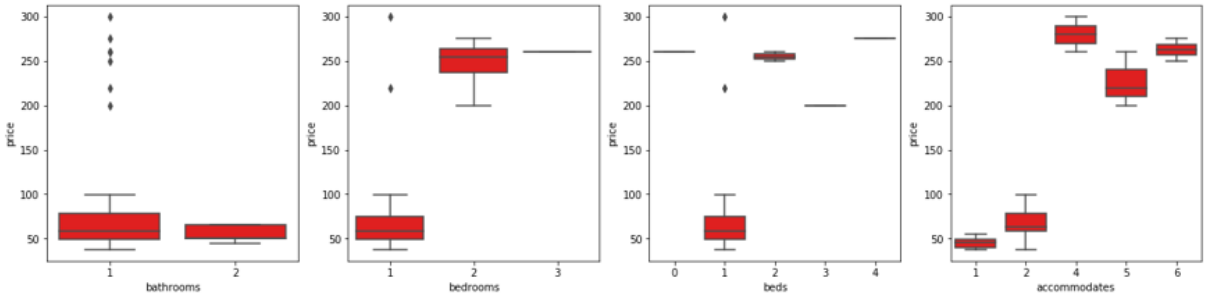

price
beds
bathrooms
bedrooms
accommodates


Price Distribution of Score Rating Bins
price
score-binned

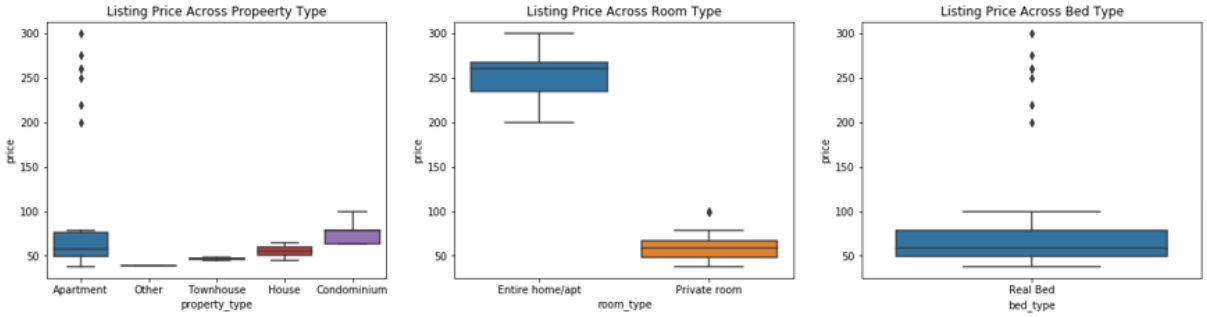

price
score-binned

Queens: Astoria Profile of Top Ten Hosts

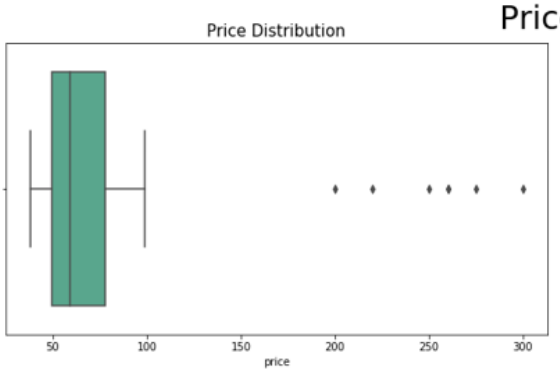
Listings' Physical Attributes Relationship to Price

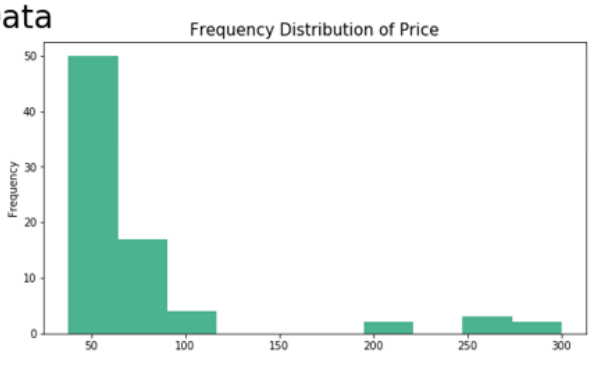


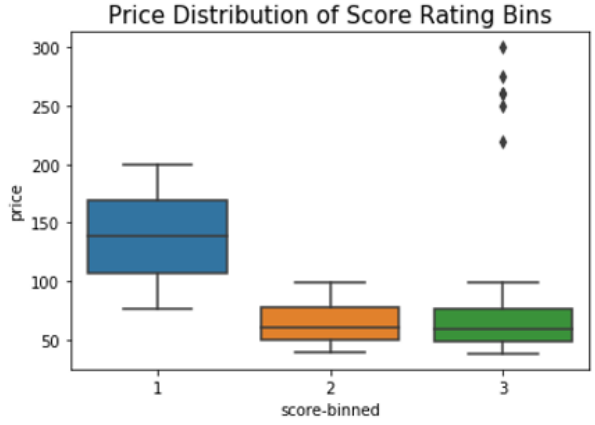
Listings' Physical Attributes Relationship to Price



Price Data



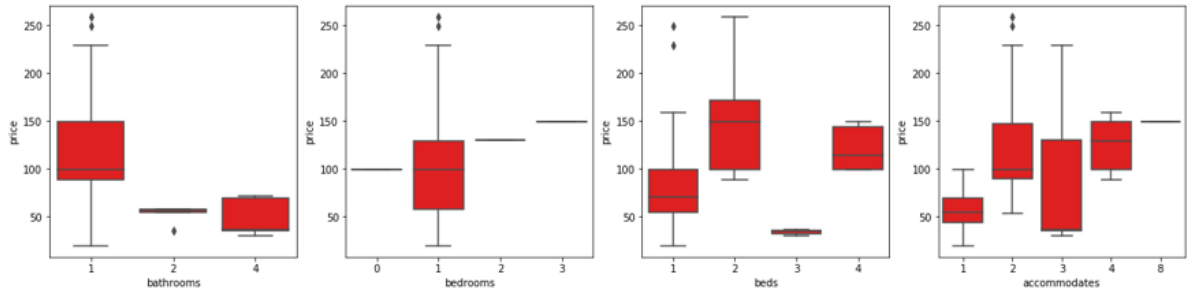




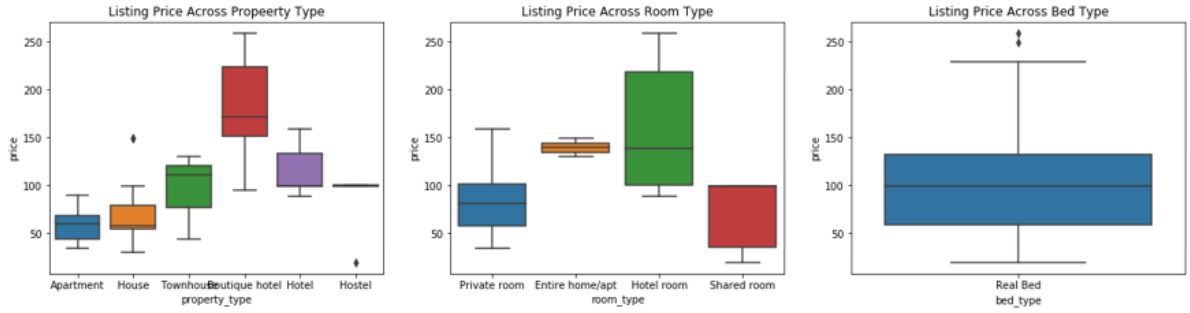
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000	113.000000
mean	2.761062	1.858407	1.345133	1.681416	68.159292	1.327434	89.962284
std	2.071380	0.624889	0.913880	1.128095	50.150945	1.113573	13.860322
min	1.000000	1.000000	1.000000	1.000000	25.000000	1.000000	20.000000
25%	2.000000	1.000000	1.000000	1.000000	37.000000	1.000000	89.000000
50%	2.000000	2.000000	1.000000	1.000000	45.000000	1.000000	93.903137
75%	3.000000	2.000000	1.000000	2.000000	75.000000	1.000000	100.000000
max	13.000000	3.000000	5.000000	6.000000	300.000000	10.000000	100.000000

Queens: Long Island City Profile of Top Ten Hosts

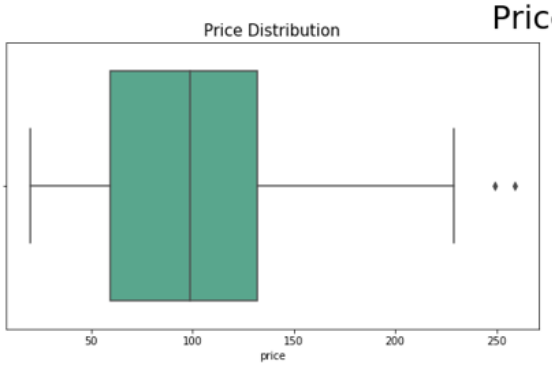
Listings' Physical Attributes Relationship to Price

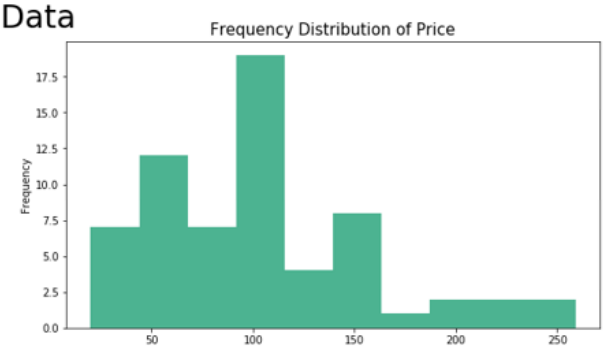


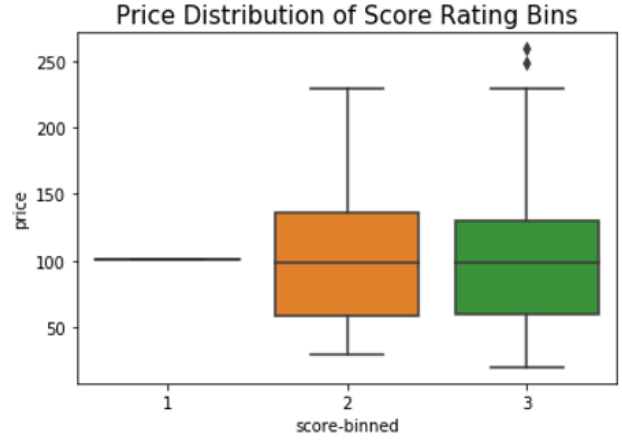
Listings' Physical Attributes Relationship to Price



Price Data

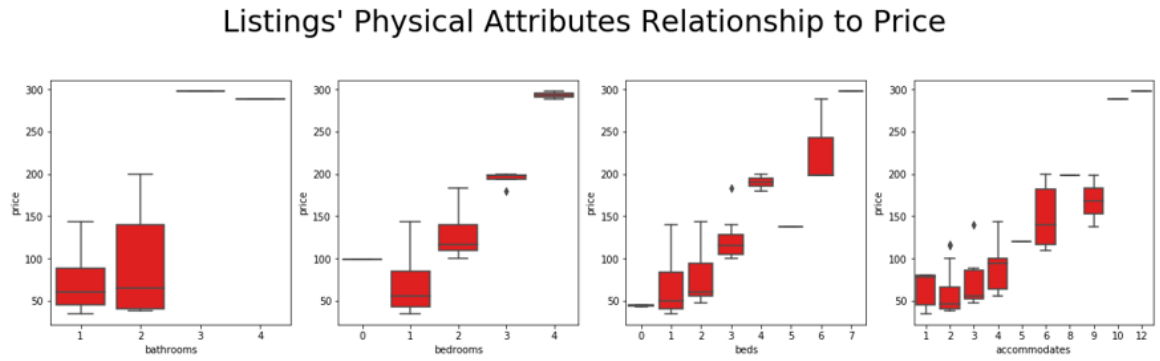


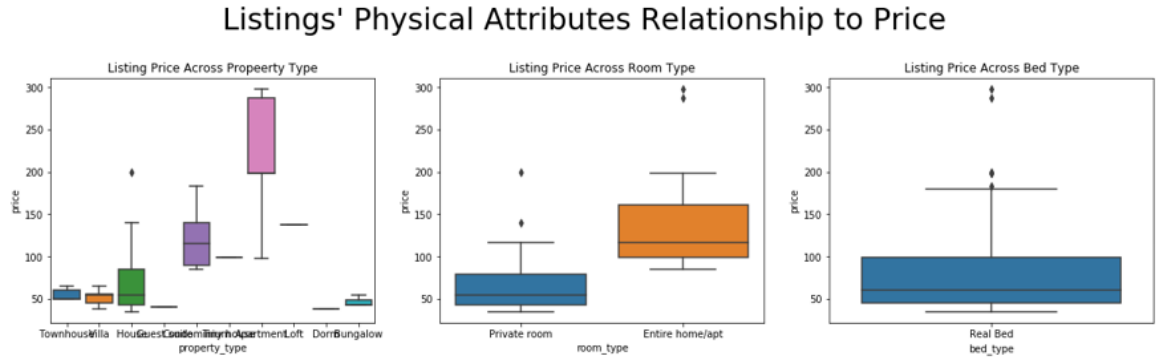


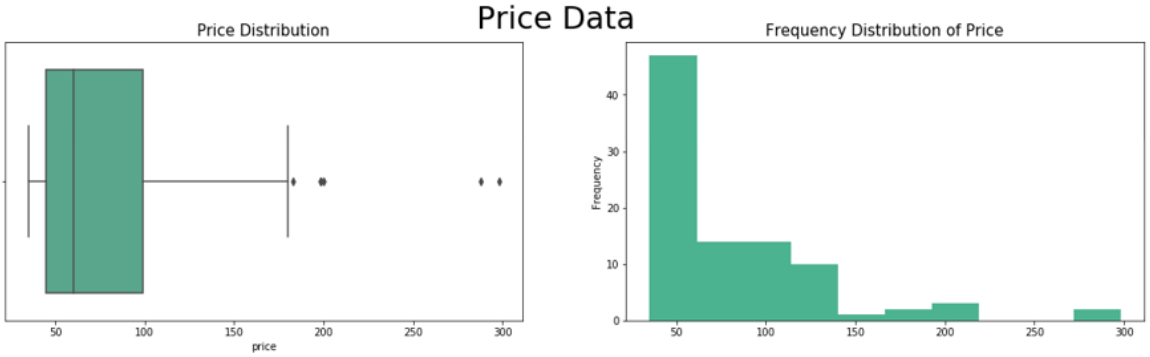


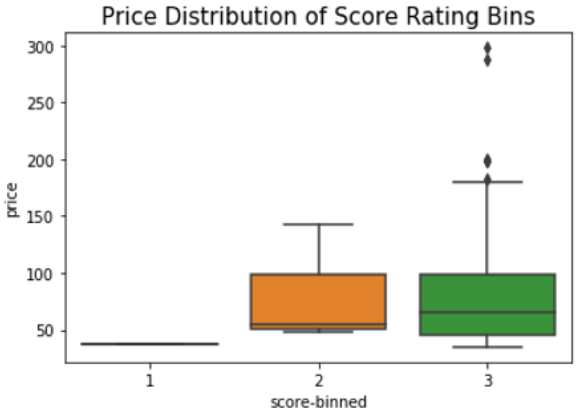
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	64.000000	64.000000	64.000000	64.000000	64.000000	64.000000	64.000000
mean	2.375000	1.328125	1.031250	1.671875	103.968750	1.171875	90.125490
std	1.315355	0.836749	0.306995	0.943729	55.154896	0.380254	6.910237
min	1.000000	1.000000	0.000000	1.000000	20.000000	1.000000	67.000000
25%	1.000000	1.000000	1.000000	1.000000	59.500000	1.000000	86.000000
50%	2.000000	1.000000	1.000000	1.000000	99.000000	1.000000	91.000000
75%	3.250000	1.000000	1.000000	2.000000	132.250000	1.000000	93.903137
max	8.000000	4.000000	3.000000	4.000000	259.000000	2.000000	100.000000

Queens: Flushing Profile of Top Ten Hosts





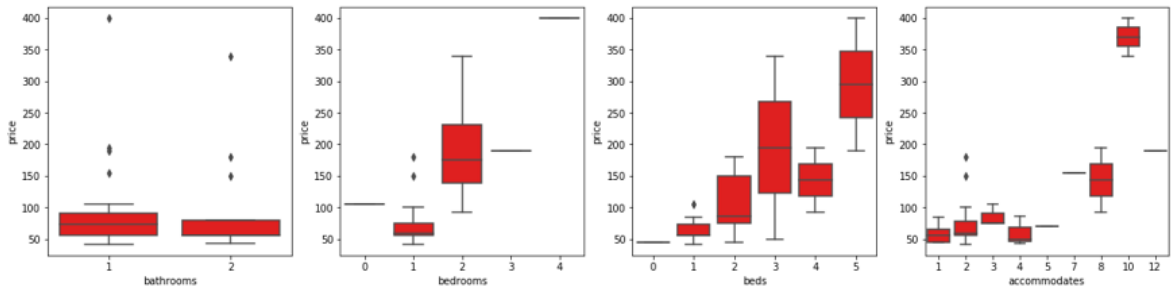




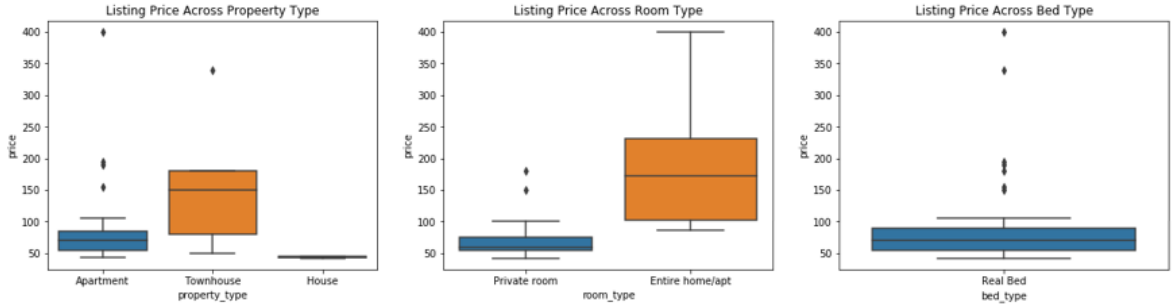
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	39.000000	39.000000	39.000000	39.000000	39.000000	39.000000	39.000000
mean	3.230769	1.333333	1.205128	1.666667	94.384615	1.538462	94.374681
std	2.795311	0.477567	0.656124	1.154701	76.825371	0.995943	6.426840
min	1.000000	1.000000	0.000000	0.000000	42.000000	1.000000	69.000000
25%	2.000000	1.000000	1.000000	1.000000	55.000000	1.000000	92.000000
50%	2.000000	1.000000	1.000000	1.000000	70.000000	1.000000	93.903137
75%	3.500000	2.000000	1.000000	2.000000	89.000000	2.000000	100.000000
max	12.000000	2.000000	4.000000	5.000000	400.000000	6.000000	100.000000

Queens: Ridgewood Profile of Top Ten Hosts

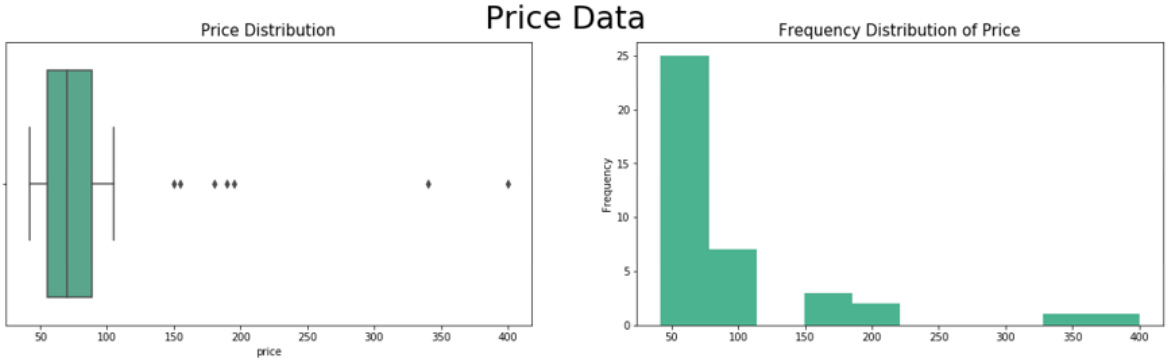
Listings' Physical Attributes Relationship to Price

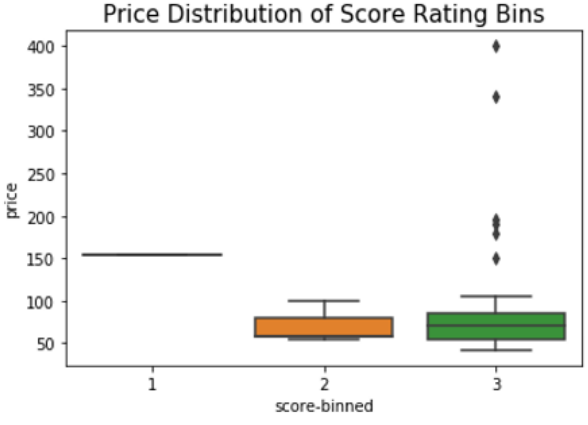


Listings' Physical Attributes Relationship to Price



Price Data

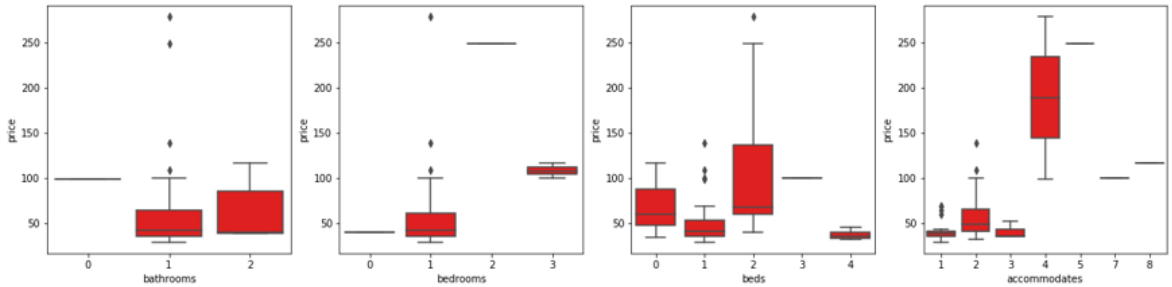




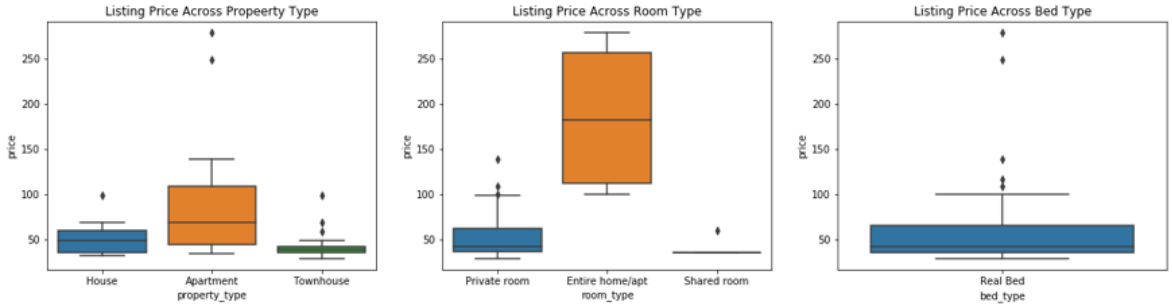
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	64.000000	64.000000	64.000000	64.000000	64.000000	64.000000	64.000000
mean	1.937500	1.078125	1.062500	1.250000	58.625000	1.359375	91.434963
std	1.295597	0.323899	0.393398	0.776643	44.034799	0.932094	8.615260
min	1.000000	0.000000	0.000000	0.000000	29.000000	1.000000	50.000000
25%	1.000000	1.000000	1.000000	1.000000	35.750000	1.000000	87.750000
50%	2.000000	1.000000	1.000000	1.000000	42.000000	1.000000	93.451569
75%	2.000000	1.000000	1.000000	1.000000	65.000000	1.000000	100.000000
max	8.000000	2.000000	3.000000	4.000000	279.000000	6.000000	100.000000

Queens: Sunnyside Profile of Top Ten Hosts

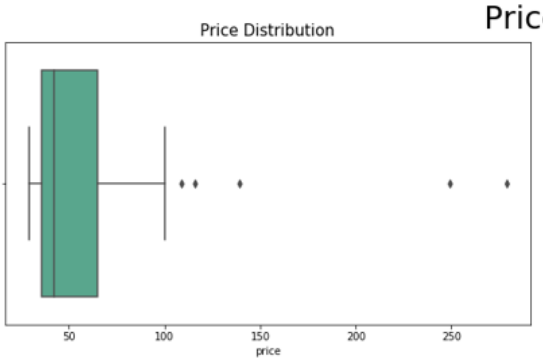
Listings' Physical Attributes Relationship to Price

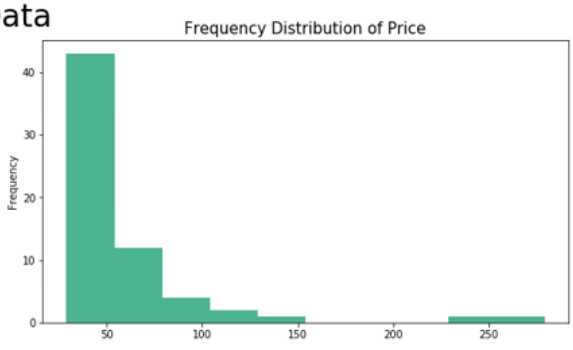


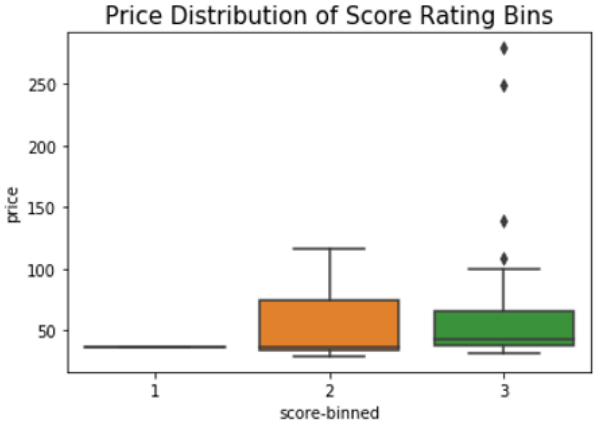
Listings' Physical Attributes Relationship to Price



Price Data



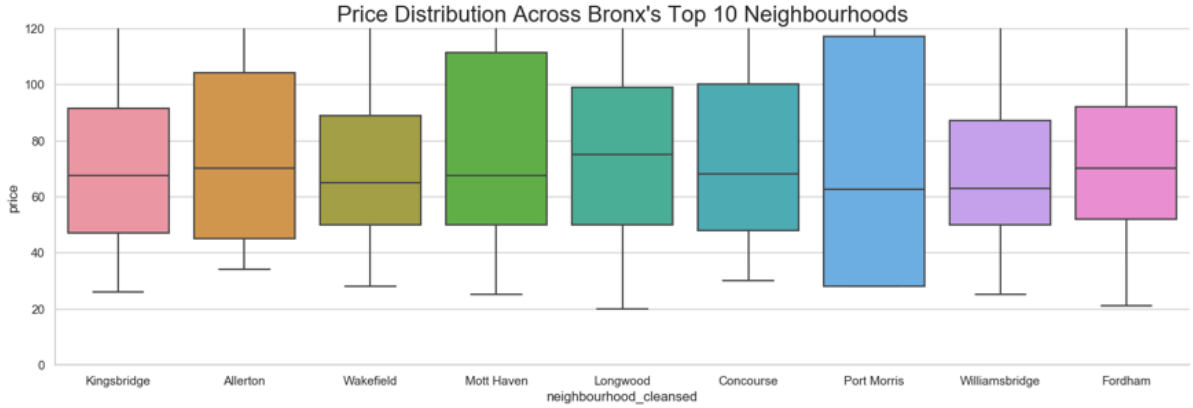


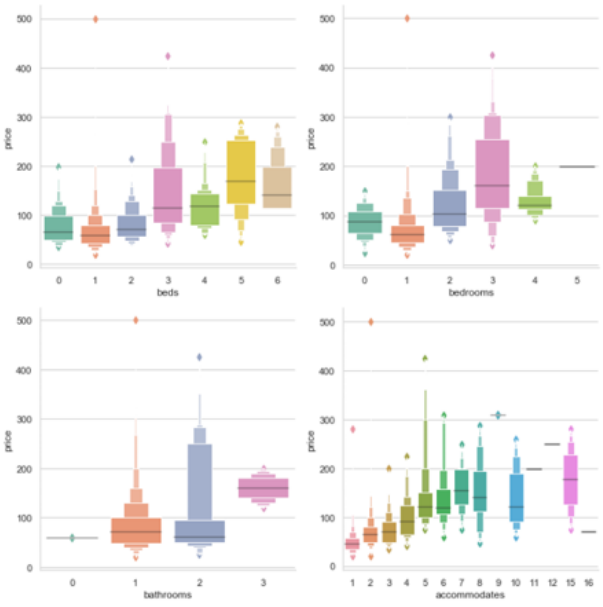


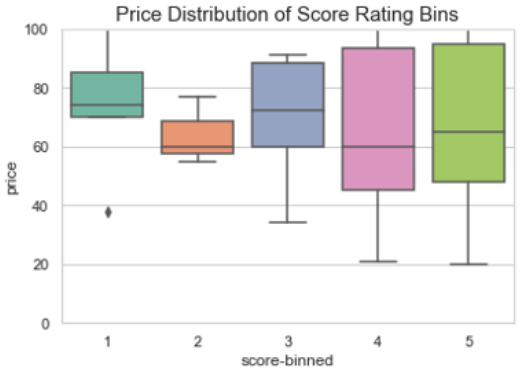
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000	93.000000
mean	3.172043	1.279570	1.236559	1.709677	81.462366	2.064516	94.286494
std	2.019598	0.539033	0.649378	1.281818	50.875927	1.240673	5.451863
min	1.000000	1.000000	0.000000	0.000000	35.000000	1.000000	60.000000
25%	2.000000	1.000000	1.000000	1.000000	45.000000	1.000000	93.000000
50%	2.000000	1.000000	1.000000	1.000000	60.000000	2.000000	94.000000
75%	4.000000	1.000000	1.000000	2.000000	99.000000	2.000000	98.000000
max	12.000000	4.000000	4.000000	7.000000	298.000000	6.000000	100.000000

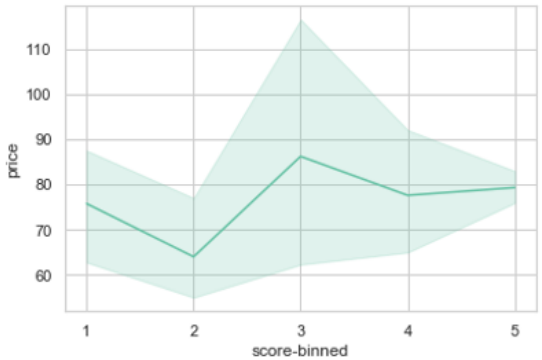
Bronx


Bronx Price Data
Price Distribution
price
Frequency Distribution of Price
Frequency


Price Distribution Across Bronx's Top 10 Neighbourhoods
price
Kingsbridge
Allerton
Wakefield
Mott Haven
Longwood
Concourse
Port Morris
Williamsbridge
Fordham
neighbourhood_cleansed

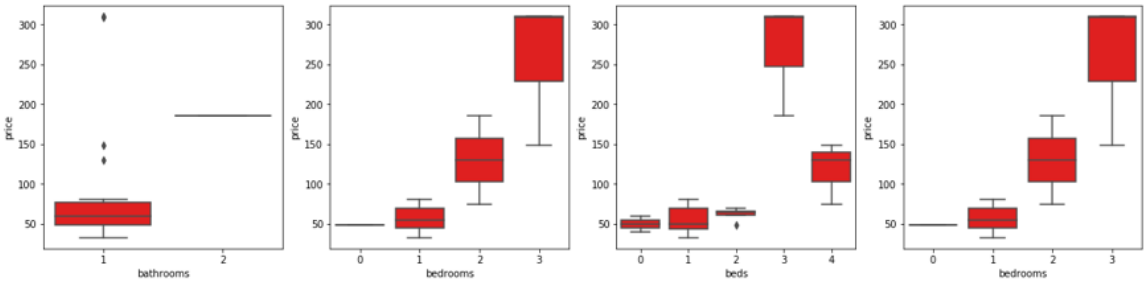

beds
bedrooms
bathrooms
accommodates
price


Price Distribution of Score Rating Bins
price
score-binned

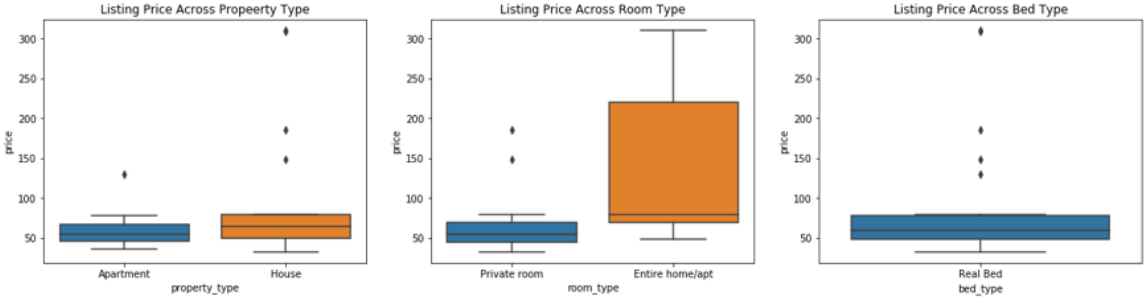

price
score-binned

Bronx: Wakefield Profile of Top Ten Hosts

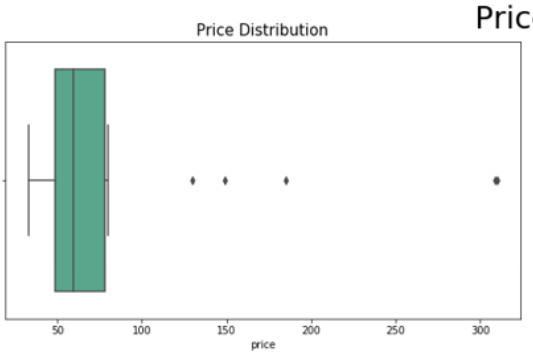
Listings' Physical Attributes Relationship to Price

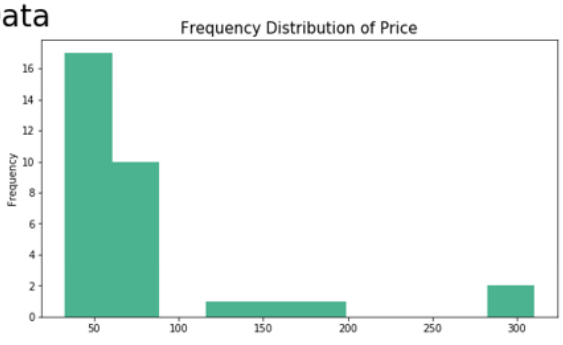


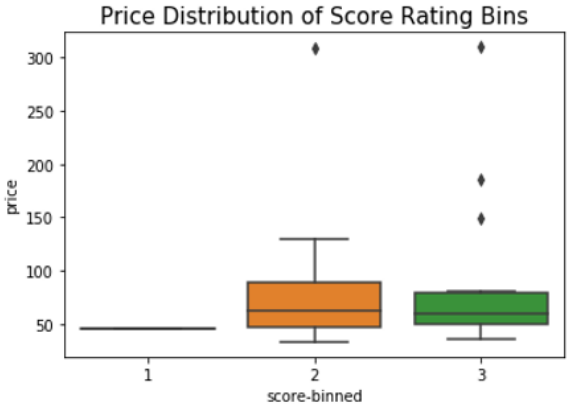
Listings' Physical Attributes Relationship to Price



Price Data



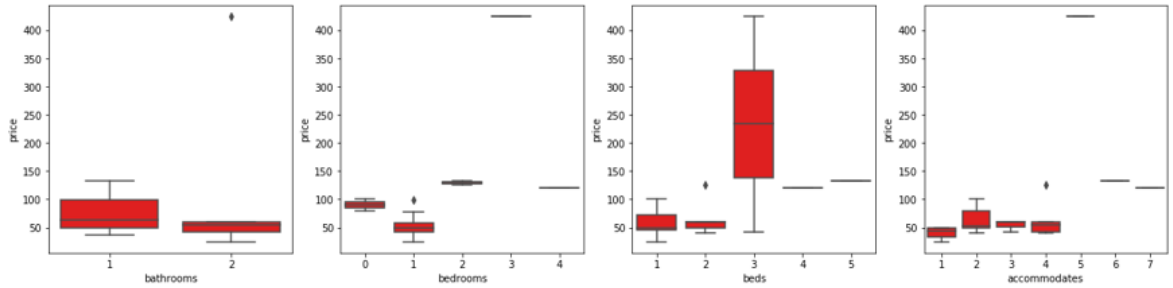




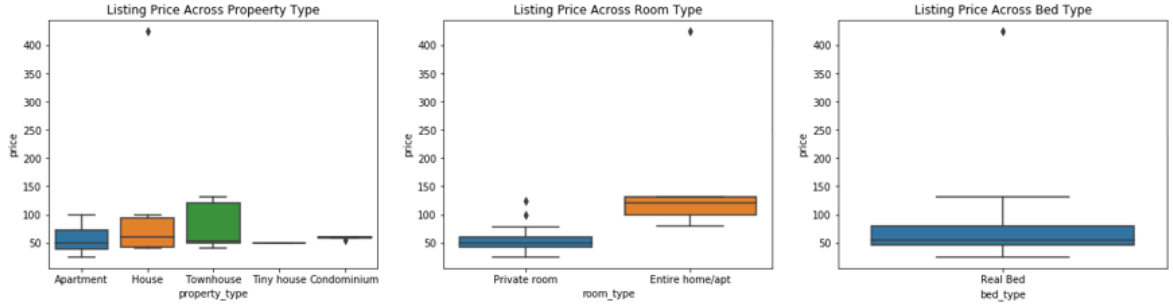
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000
mean	3.250000	1.031250	1.250000	1.531250	81.375000	1.406250	95.328615
std	2.676059	0.176777	0.672022	1.077164	68.275887	0.756024	4.312622
min	1.000000	1.000000	0.000000	0.000000	33.000000	1.000000	80.000000
25%	2.000000	1.000000	1.000000	1.000000	48.750000	1.000000	93.000000
50%	2.000000	1.000000	1.000000	1.000000	59.500000	1.000000	95.000000
75%	4.000000	1.000000	1.000000	2.000000	78.250000	2.000000	100.000000
max	10.000000	2.000000	3.000000	4.000000	310.000000	4.000000	100.000000

Bronx: Mott Haven Profile of Top Ten Hosts

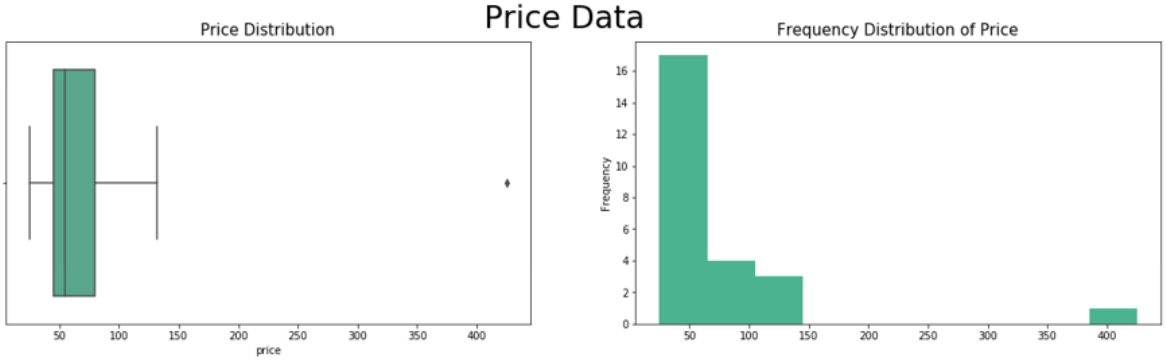
Listings' Physical Attributes Relationship to Price

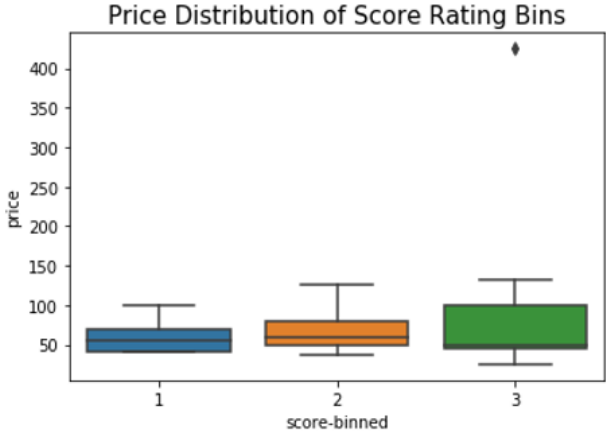


Listings' Physical Attributes Relationship to Price



Price Data

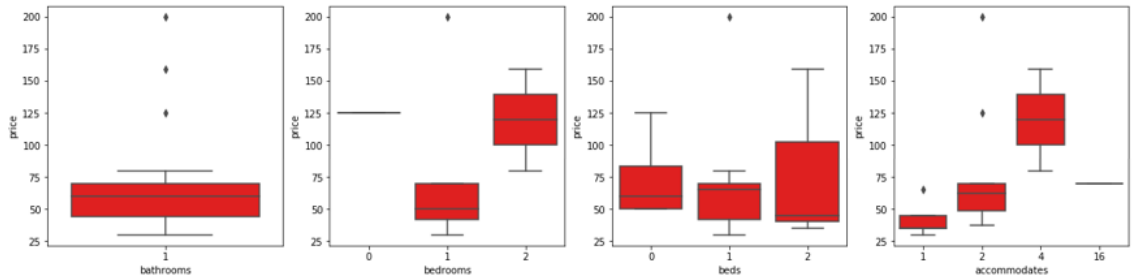




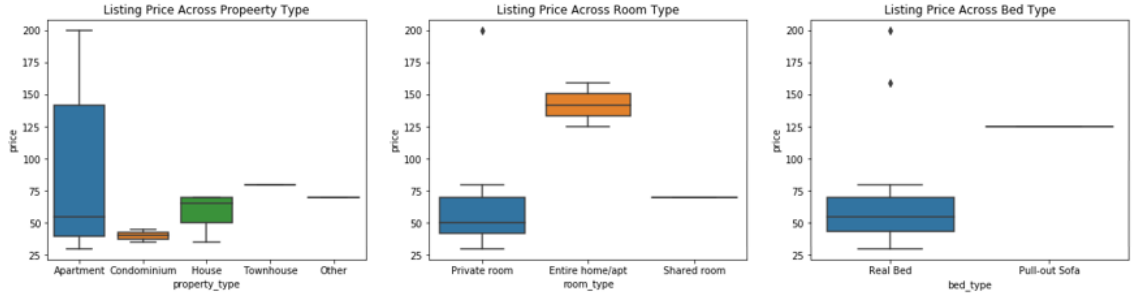
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000
mean	2.840000	1.440000	1.200000	1.720000	78.800000	1.880000	95.152251
std	1.572683	0.506623	0.816497	1.061446	77.789245	1.563117	3.957202
min	1.000000	1.000000	0.000000	1.000000	25.000000	1.000000	88.000000
25%	2.000000	1.000000	1.000000	1.000000	45.000000	1.000000	92.000000
50%	2.000000	1.000000	1.000000	1.000000	55.000000	1.000000	95.000000
75%	4.000000	2.000000	1.000000	2.000000	80.000000	2.000000	99.000000
max	7.000000	2.000000	4.000000	5.000000	425.000000	7.000000	100.000000

Bronx: Kingsbridge Profile of Top Ten Hosts

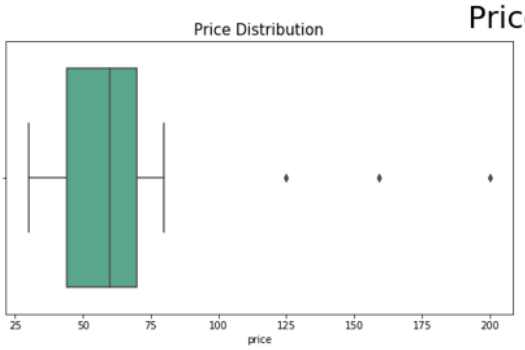
Listings' Physical Attributes Relationship to Price

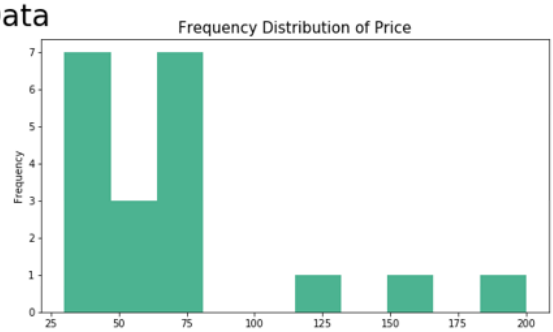


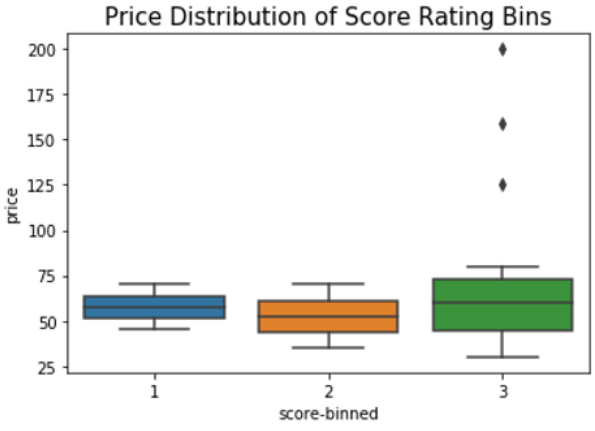
Listings' Physical Attributes Relationship to Price



Price Data



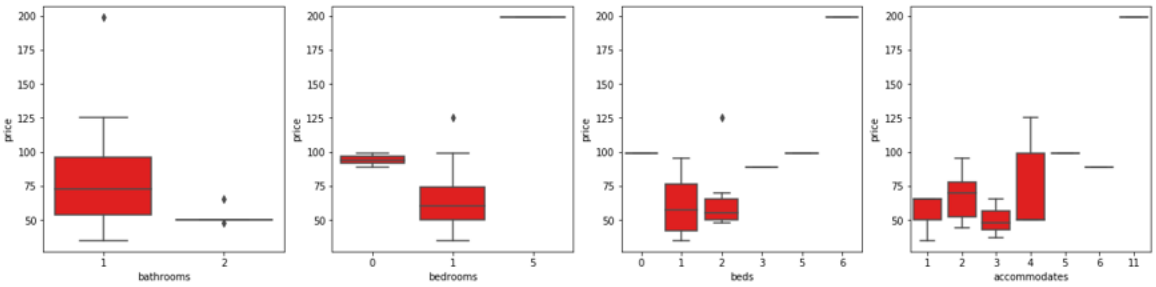




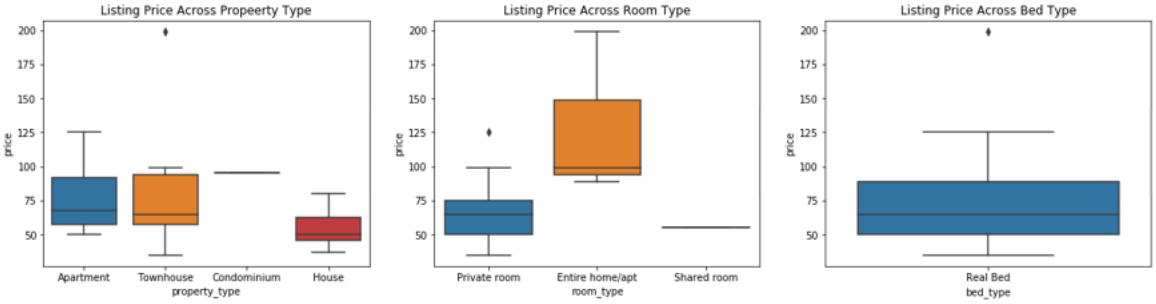
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	20.000000	20.0	20.000000	20.000000	20.000000	20.000000	20.000000
mean	2.650000	1.0	1.050000	0.950000	70.150000	1.300000	94.240314
std	3.248886	0.0	0.394034	0.604805	43.630355	0.801315	7.745504
min	1.000000	1.0	0.000000	0.000000	30.000000	1.000000	70.000000
25%	1.750000	1.0	1.000000	1.000000	44.250000	1.000000	93.427353
50%	2.000000	1.0	1.000000	1.000000	60.000000	1.000000	95.000000
75%	2.000000	1.0	1.000000	1.000000	70.000000	1.000000	100.000000
max	16.000000	1.0	2.000000	2.000000	200.000000	4.000000	100.000000

Bronx: Longwood Profile of Top Ten Hosts

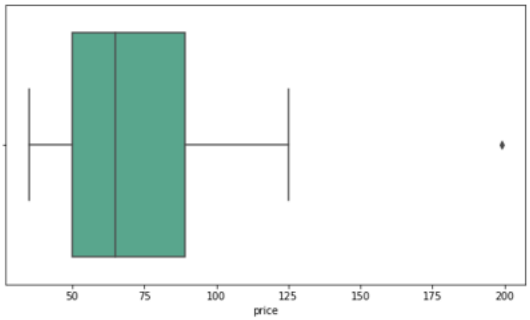
Listings' Physical Attributes Relationship to Price

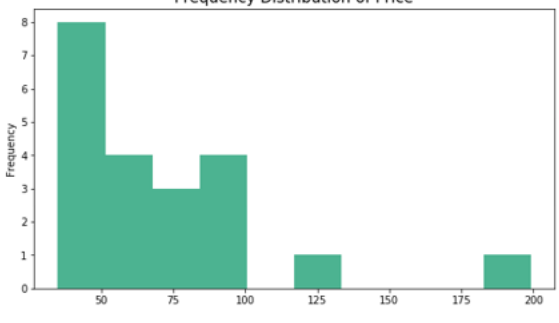


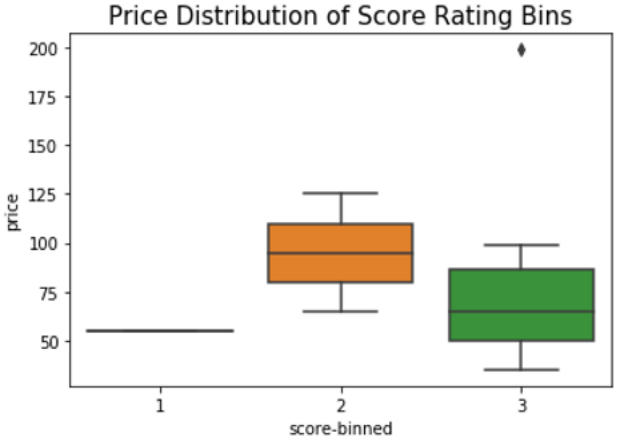
Listings' Physical Attributes Relationship to Price



Price Data



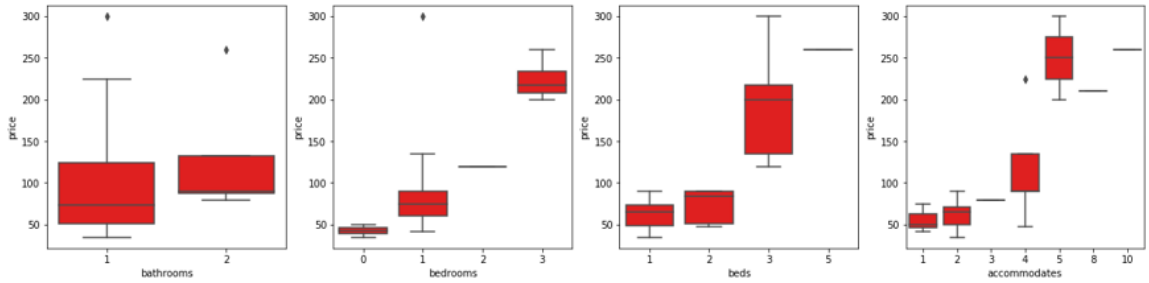




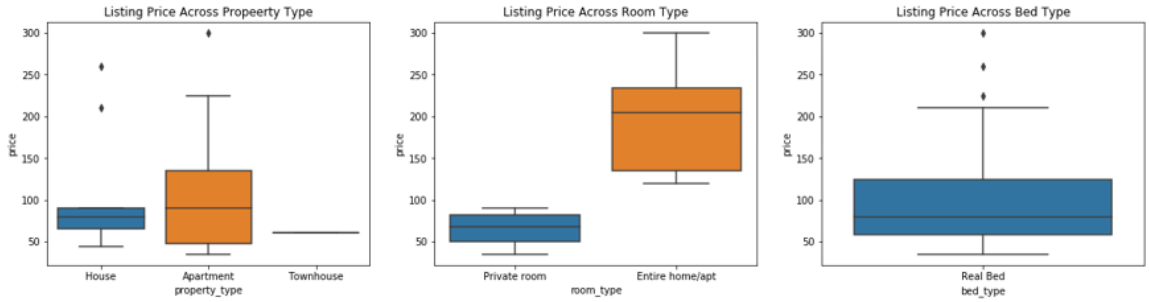
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	21.000000	21.000000	21.000000	21.000000	21.000000	21.000000	21.000000
mean	3.238095	1.238095	1.095238	1.904762	73.571429	1.285714	92.791074
std	2.233937	0.436436	0.943650	1.374946	37.124886	0.462910	9.905418
min	1.000000	1.000000	0.000000	0.000000	35.000000	1.000000	57.000000
25%	2.000000	1.000000	1.000000	1.000000	50.000000	1.000000	91.000000
50%	3.000000	1.000000	1.000000	2.000000	65.000000	1.000000	94.000000
75%	4.000000	1.000000	1.000000	2.000000	89.000000	2.000000	99.000000
max	11.000000	2.000000	5.000000	6.000000	199.000000	2.000000	100.000000

Bronx: Fordham Profile of Top Ten Hosts

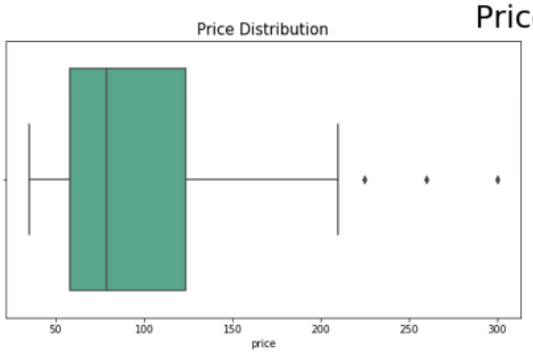
Listings' Physical Attributes Relationship to Price

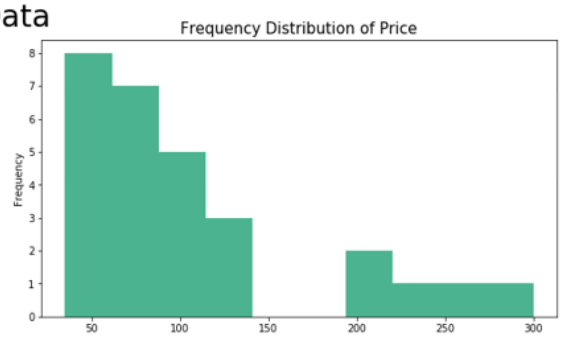


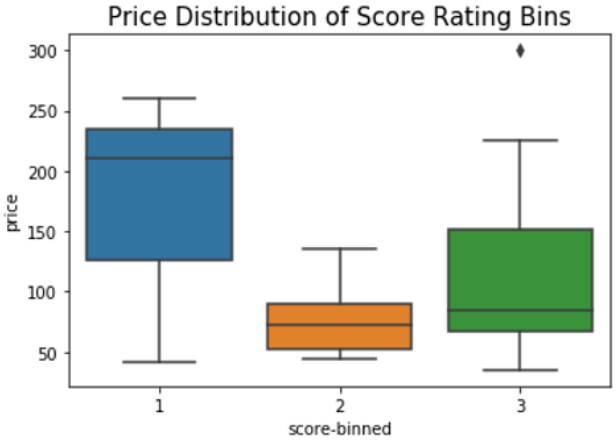
Listings' Physical Attributes Relationship to Price



Price Data

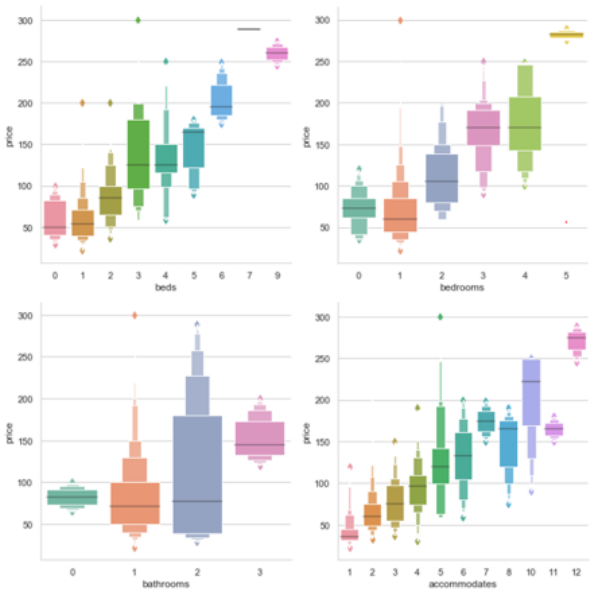


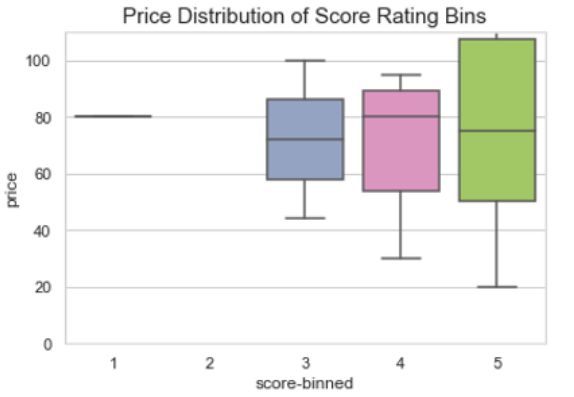


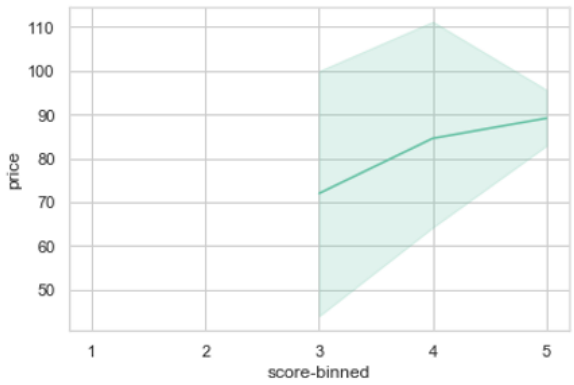


	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	28.000000	28.000000	28.000000	28.000000	28.000000	28.000000	28.000000
mean	3.285714	1.142857	1.250000	1.928571	104.250000	2.107143	92.496541
std	2.034101	0.356348	0.799305	1.015749	70.552646	1.640670	5.391071
min	1.000000	1.000000	0.000000	1.000000	35.000000	1.000000	80.000000
25%	2.000000	1.000000	1.000000	1.000000	58.000000	1.000000	89.000000
50%	2.500000	1.000000	1.000000	2.000000	79.000000	2.000000	92.500000
75%	4.000000	1.000000	1.000000	3.000000	123.750000	2.250000	96.250000
max	10.000000	2.000000	3.000000	5.000000	300.000000	8.000000	100.000000

Staten Island

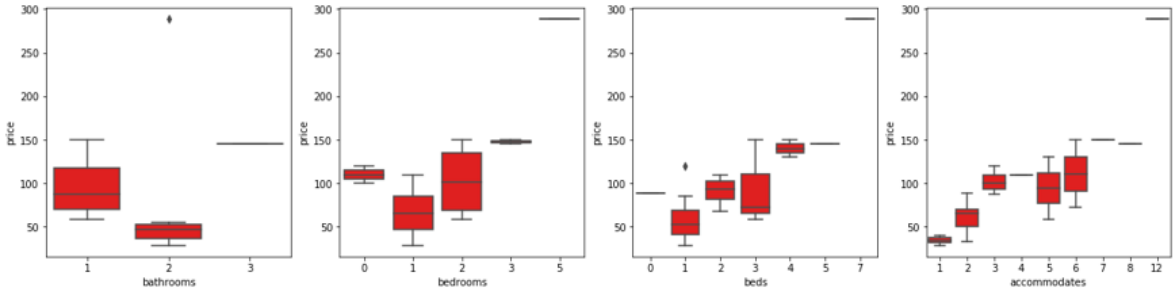




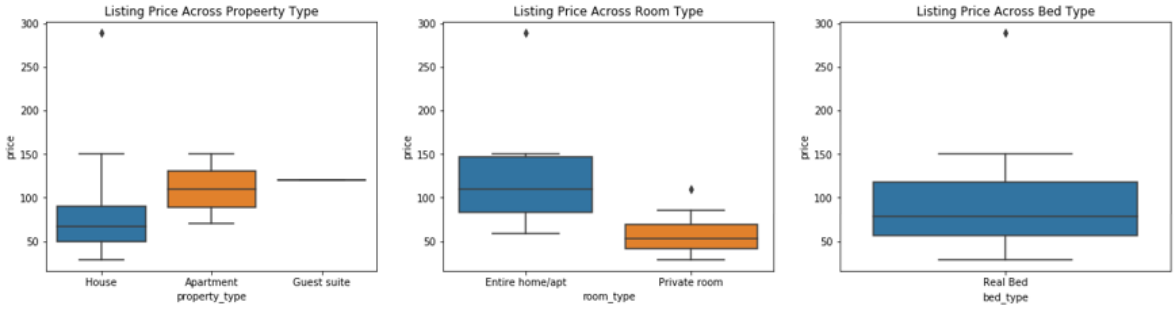


Staten Island: St. George Profile of Top Ten Hosts

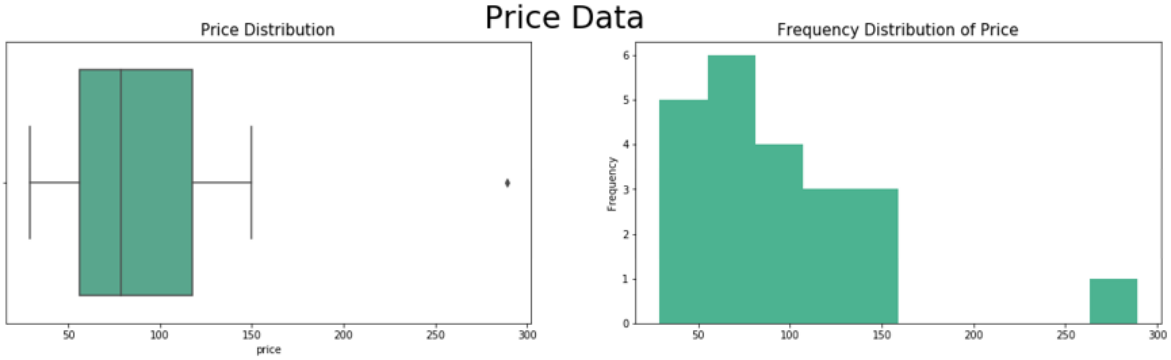
Listings' Physical Attributes Relationship to Price

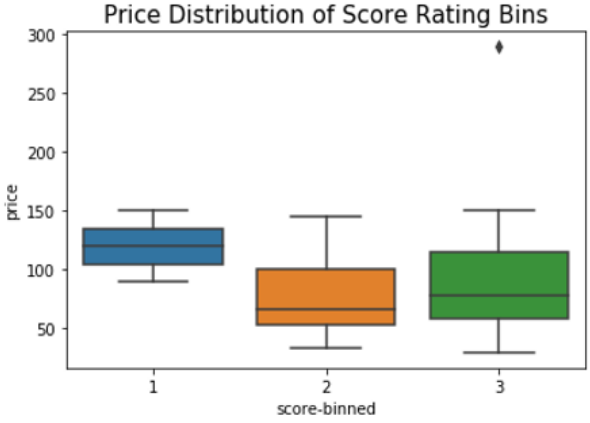


Listings' Physical Attributes Relationship to Price



Price Data

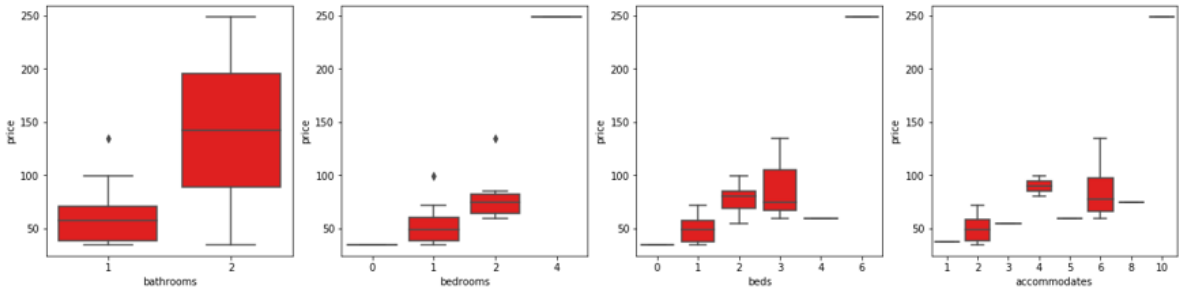




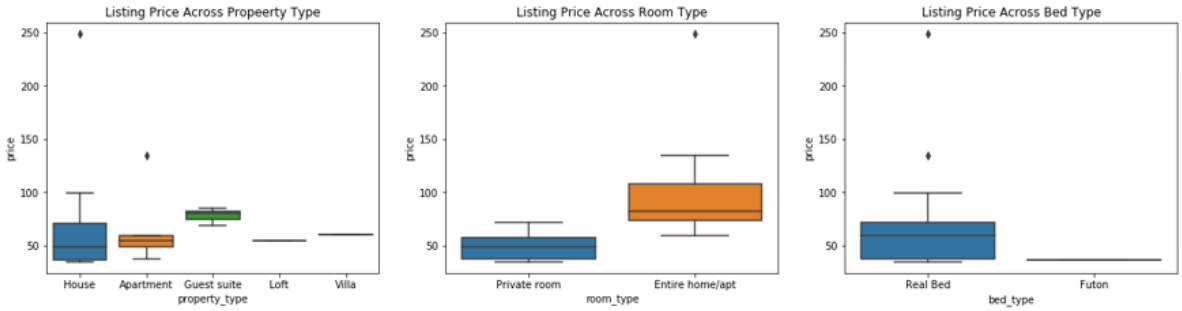
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	22.000000	22.000000	22.000000	22.000000	22.000000	22.000000	22.000000
mean	3.727273	1.409091	1.454545	2.136364	92.863636	1.909091	98.173012
std	2.711208	0.590326	1.100964	1.670342	57.577522	1.477098	1.913735
min	1.000000	1.000000	0.000000	0.000000	29.000000	1.000000	93.903137
25%	2.000000	1.000000	1.000000	1.000000	56.000000	1.000000	97.000000
50%	2.500000	1.000000	1.000000	1.500000	78.500000	1.000000	98.500000
75%	5.000000	2.000000	2.000000	3.000000	117.500000	2.000000	100.000000
max	12.000000	3.000000	5.000000	7.000000	289.000000	6.000000	100.000000

Staten Island: Tompkinsville Profile of Top Ten Hosts

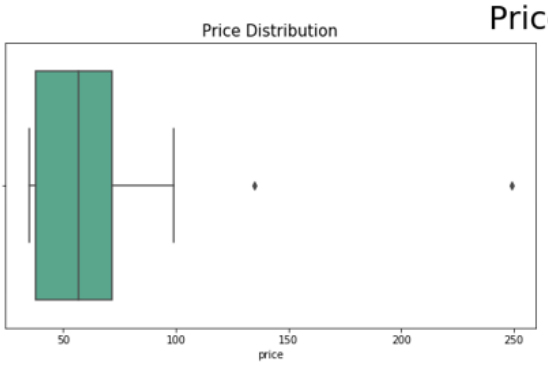
Listings' Physical Attributes Relationship to Price

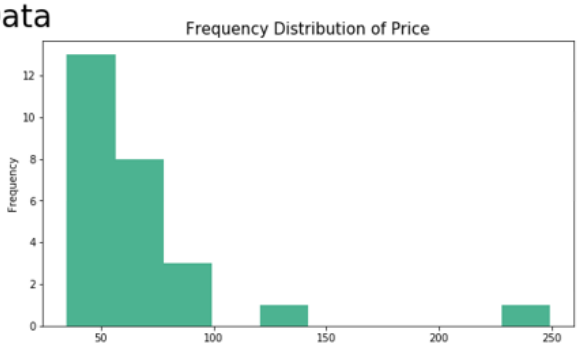


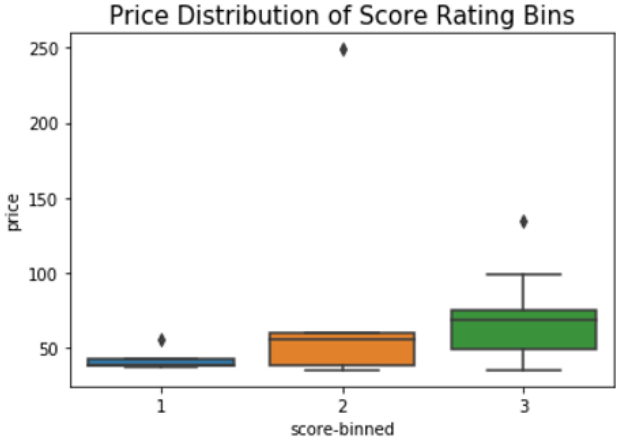
Listings' Physical Attributes Relationship to Price



Price Data



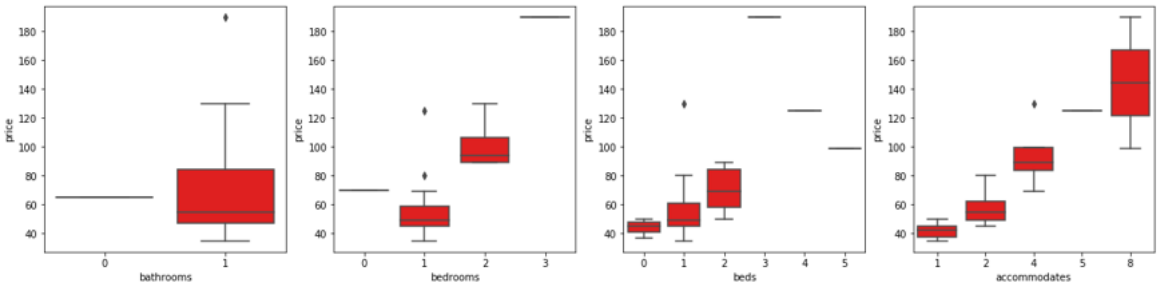




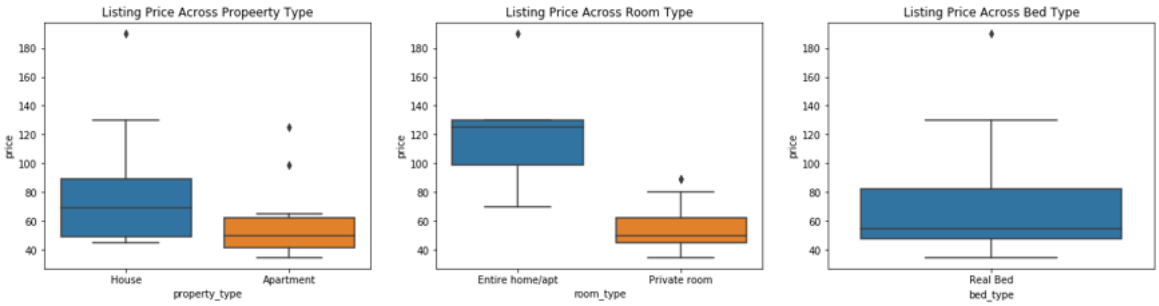
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000
mean	3.384615	1.076923	1.346154	1.692308	66.692308	2.076923	94.611659
std	2.316496	0.271746	0.745241	1.257592	43.855918	1.741794	4.253385
min	1.000000	1.000000	0.000000	0.000000	35.000000	1.000000	84.000000
25%	2.000000	1.000000	1.000000	1.000000	38.000000	1.000000	92.000000
50%	2.000000	1.000000	1.000000	1.000000	57.000000	1.000000	95.500000
75%	4.750000	1.000000	2.000000	2.000000	71.750000	3.500000	97.750000
max	10.000000	2.000000	4.000000	6.000000	249.000000	8.000000	100.000000

Staten Island: West Brighton Profile of Top Ten Hosts

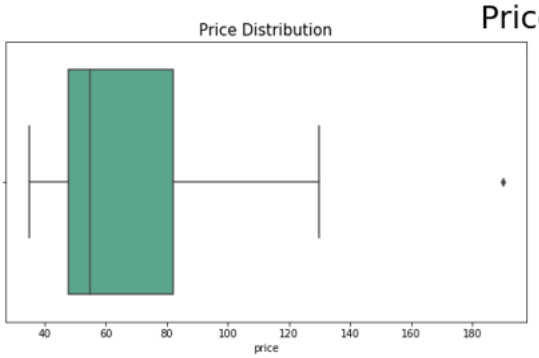
Listings' Physical Attributes Relationship to Price

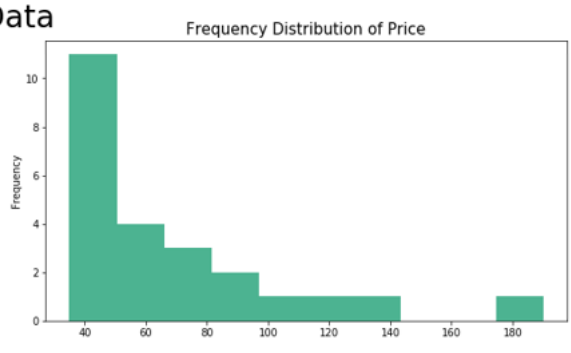


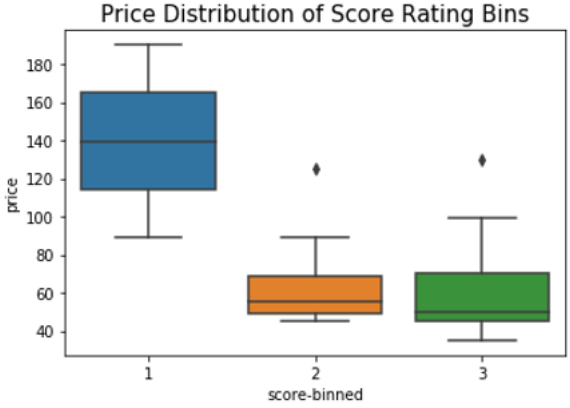
Listings' Physical Attributes Relationship to Price



Price Data



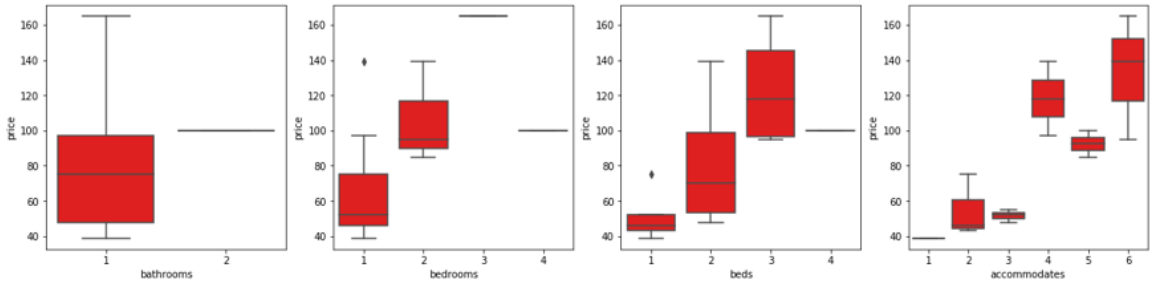




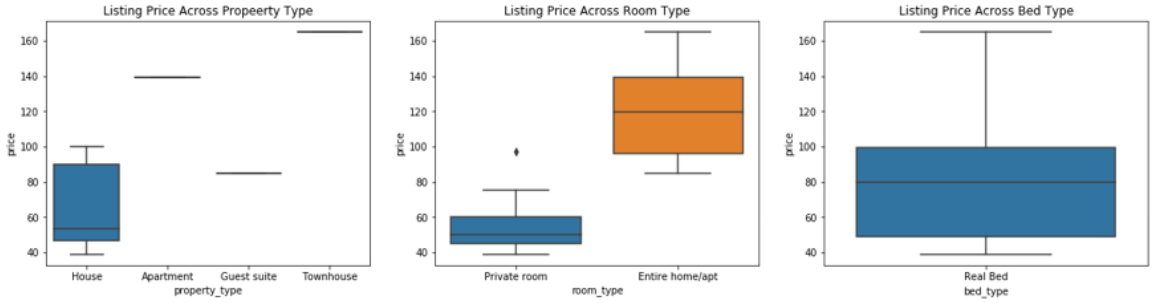
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000
mean	2.708333	0.958333	1.208333	1.500000	69.541667	1.333333	93.154559
std	1.988645	0.204124	0.588230	1.179536	36.434550	0.761387	5.959551
min	1.000000	0.000000	0.000000	0.000000	35.000000	1.000000	76.000000
25%	1.750000	1.000000	1.000000	1.000000	48.000000	1.000000	91.000000
50%	2.000000	1.000000	1.000000	1.000000	55.000000	1.000000	93.451569
75%	4.000000	1.000000	1.000000	2.000000	82.250000	1.000000	98.250000
max	8.000000	1.000000	3.000000	5.000000	190.000000	4.000000	100.000000

Staten Island: Stapleton Profile of Top Ten Hosts

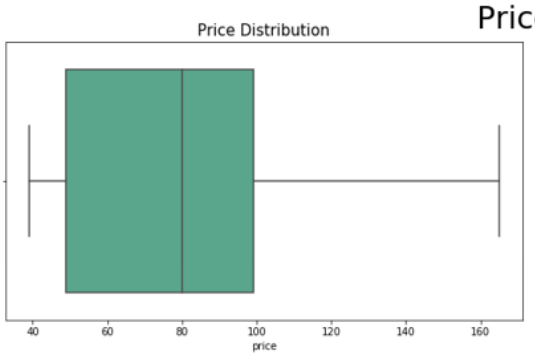
Listings' Physical Attributes Relationship to Price

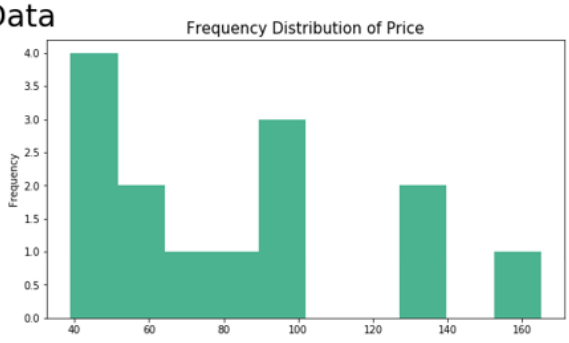


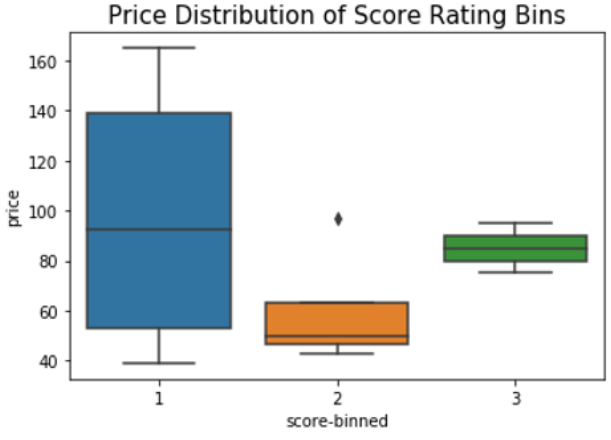
Listings' Physical Attributes Relationship to Price



Price Data



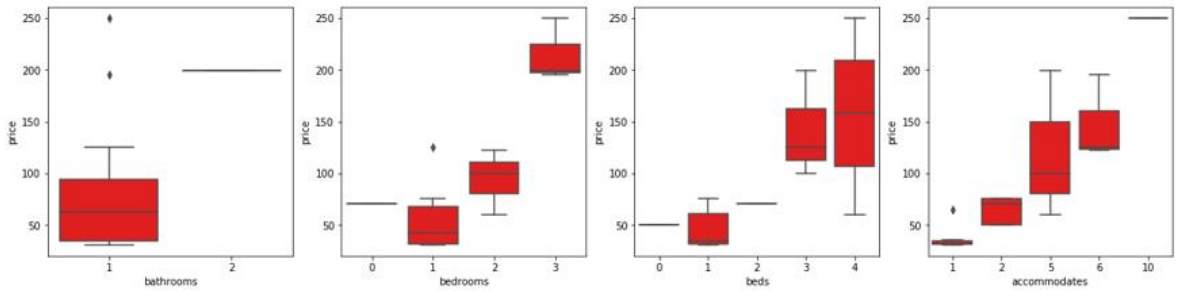




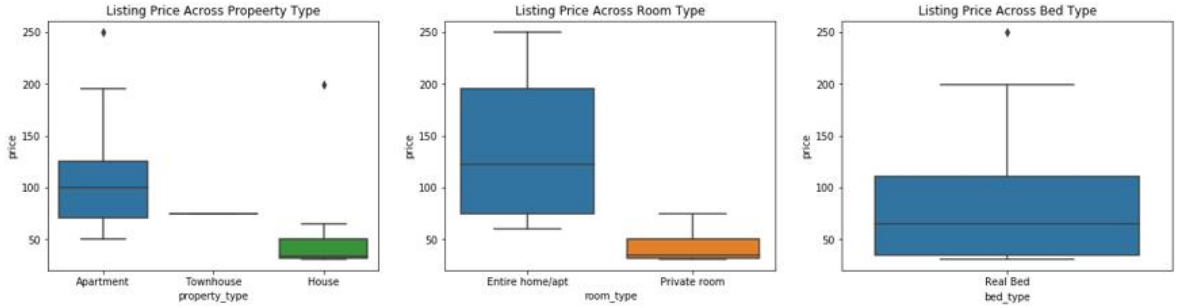
	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	14.000000	14.000000	14.000000	14.000000	14.000000	14.000000	14.000000
mean	3.714286	1.071429	1.571429	2.071429	84.142857	2.214286	95.486162
std	1.683795	0.267261	0.937614	0.997249	40.714498	1.251373	1.883067
min	1.000000	1.000000	1.000000	1.000000	39.000000	1.000000	93.000000
25%	2.250000	1.000000	1.000000	1.000000	49.000000	1.000000	94.000000
50%	3.500000	1.000000	1.000000	2.000000	80.000000	2.000000	95.000000
75%	5.000000	1.000000	2.000000	3.000000	99.250000	3.500000	96.000000
max	6.000000	2.000000	4.000000	4.000000	165.000000	4.000000	100.000000

Staten Island: Arrochar Profile of Top Ten Hosts

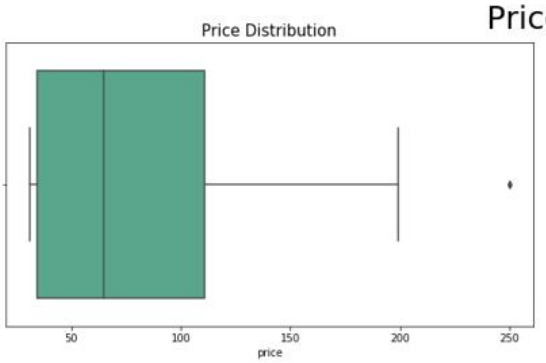
Listings' Physical Attributes Relationship to Price

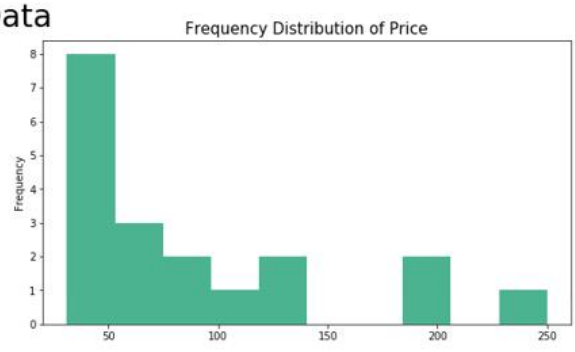


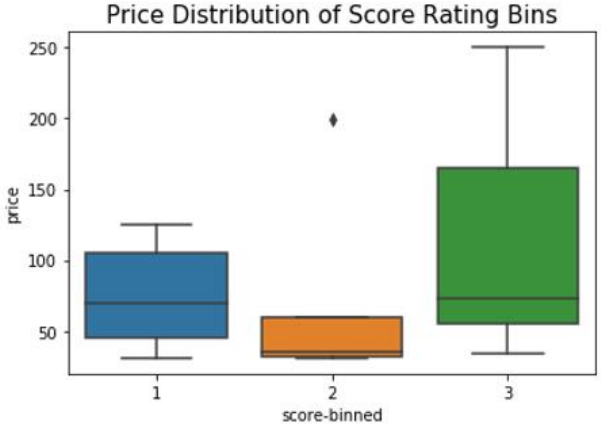
Listings' Physical Attributes Relationship to Price



Price Data







	accommodates	bathrooms	bedrooms	beds	price	guests_included	review_scores_rating
count	19.000000	19.000000	19.000000	19.000000	19.000000	19.000000	19.000000
mean	3.157895	1.052632	1.421053	1.947368	85.894737	2.052632	95.884541
std	2.587645	0.229416	0.837708	1.352926	64.945186	1.715086	2.872672
min	1.000000	1.000000	0.000000	0.000000	31.000000	1.000000	91.000000
25%	1.000000	1.000000	1.000000	1.000000	34.500000	1.000000	93.903137
50%	2.000000	1.000000	1.000000	1.000000	65.000000	1.000000	96.000000
75%	5.000000	1.000000	2.000000	3.000000	111.000000	3.000000	98.500000
max	10.000000	2.000000	3.000000	4.000000	250.000000	5.000000	100.000000

Appendix - Part 3 Objective #2

Figure A

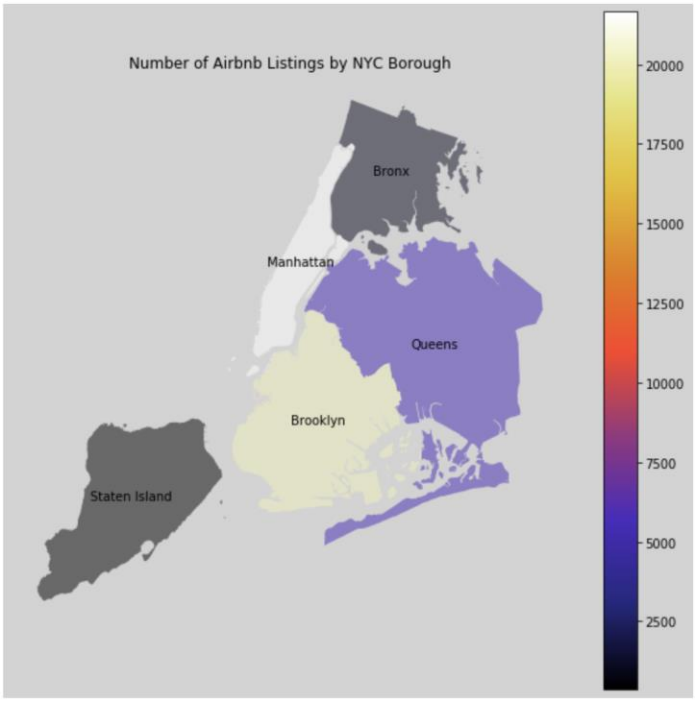


Figure B

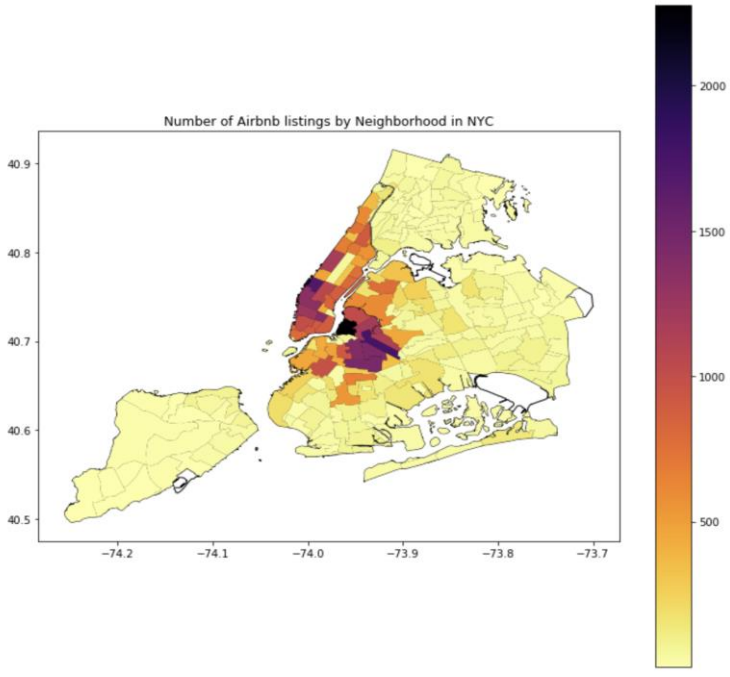


Figure C

	Name	Long	Lat	Borough	geometry
0	Empire State Building	-73.985700	40.748400	Manhattan	POINT (-73.98570 40.74840)
1	High Line	-74.004800	40.748000	Manhattan	POINT (-74.00480 40.74800)
2	World Trade Center	-74.013100	40.711800	Manhattan	POINT (-74.01310 40.71180)
3	Grand Central Terminal	-73.977295	40.752655	Manhattan	POINT (-73.97729 40.75265)
4	Madison Square Garden	-73.993400	40.750500	Manhattan	POINT (-73.99340 40.75050)
5	Times Square	-73.985500	40.758000	Manhattan	POINT (-73.98550 40.75800)
6	Rockefeller Center	-73.978700	40.758700	Manhattan	POINT (-73.97870 40.75870)
7	Central Park	-73.965400	40.782900	Manhattan	POINT (-73.96540 40.78290)
8	New York Botanical Garden	-73.877200	40.862400	Bronx	POINT (-73.87720 40.86240)
9	Yankee Stadium	-73.926186	40.829659	Bronx	POINT (-73.92619 40.82966)
10	Bronx Zoo	-73.877000	40.850600	Bronx	POINT (-73.87700 40.85060)
11	Brooklyn Bridge	-73.996900	40.706100	Brooklyn	POINT (-73.99690 40.70610)
12	Prospect Park	-73.969000	40.660200	Brooklyn	POINT (-73.96900 40.66020)
13	Coney Island	-73.970700	40.575500	Brooklyn	POINT (-73.97070 40.57550)
14	Citi Field	-73.845800	40.757100	Queens	POINT (-73.84580 40.75710)
15	Rockaway Beach	-73.816600	40.586000	Queens	POINT (-73.81660 40.58600)
16	High Rock Park	-74.123100	40.582500	Staten Island	POINT (-74.12310 40.58250)
17	Staten Island Ferry	-74.072300	40.644300	Staten Island	POINT (-74.07230 40.64430)

Figure D

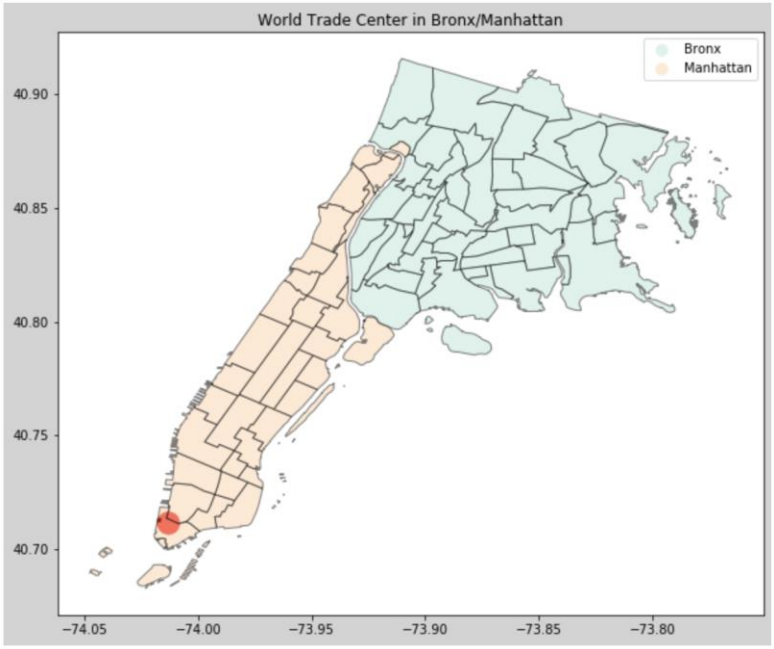


Figure E

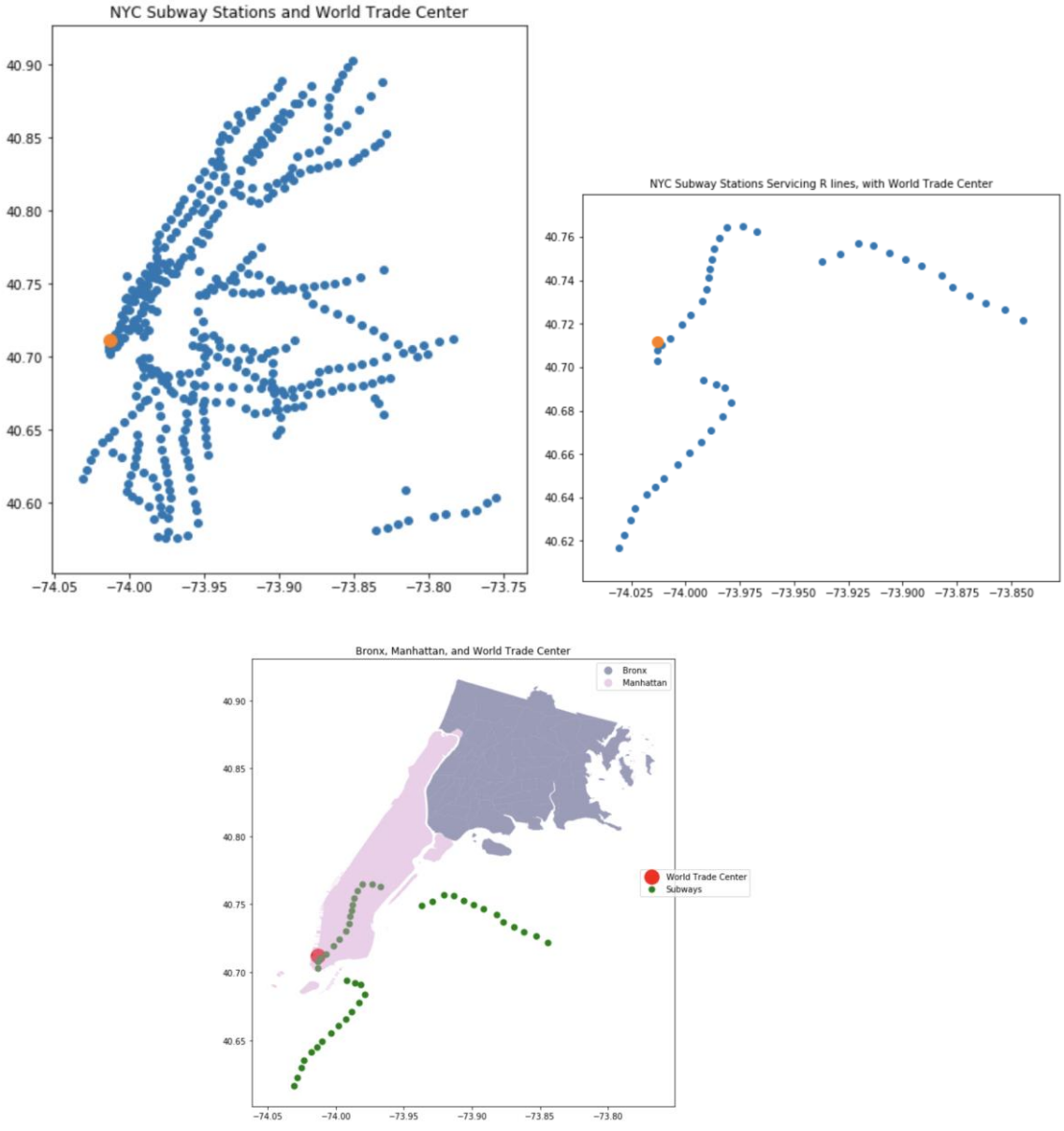
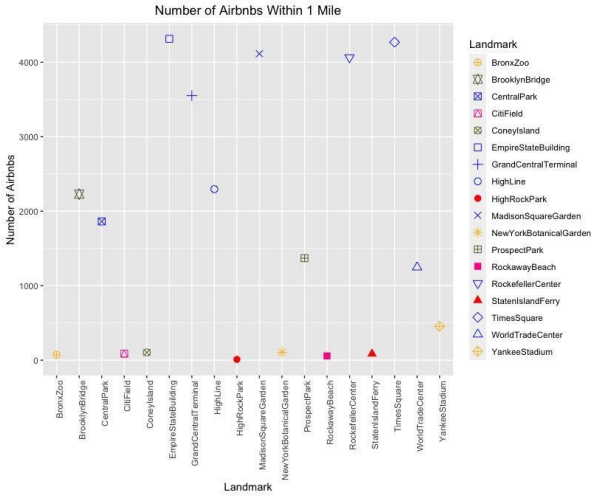
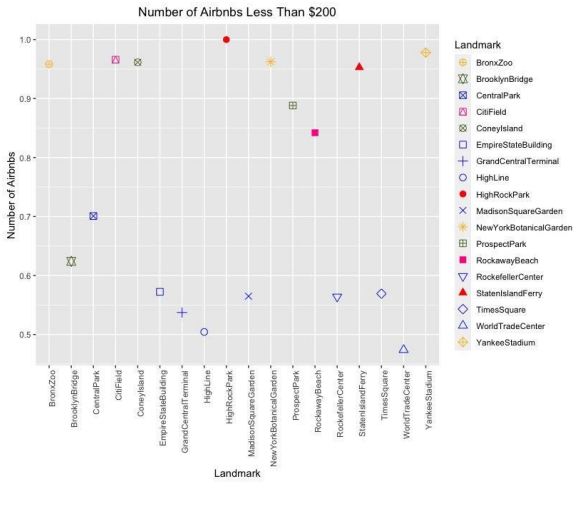
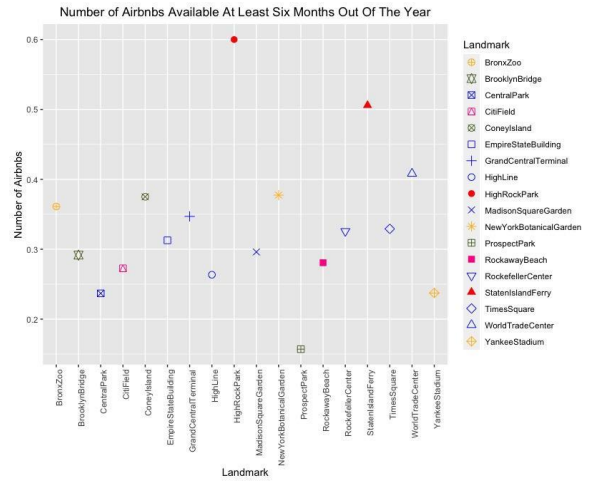
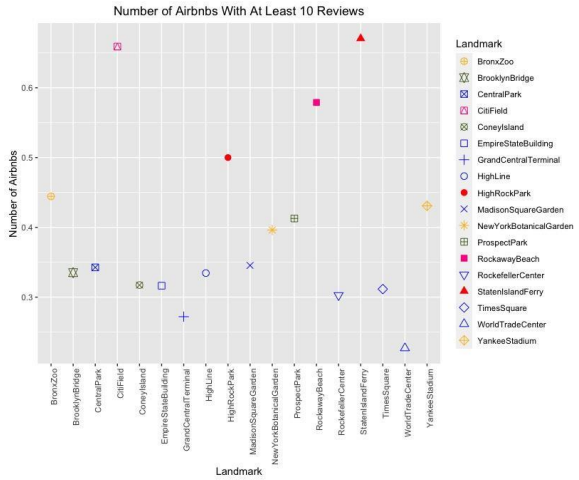

NYC Subway Stations and World Trade Center
NYC Subway Stations Servicing R lines, with World Trade Center
Bronx, Manhattan, and World Trade Center
Bronx
Manhattan
World Trade Center
Subways

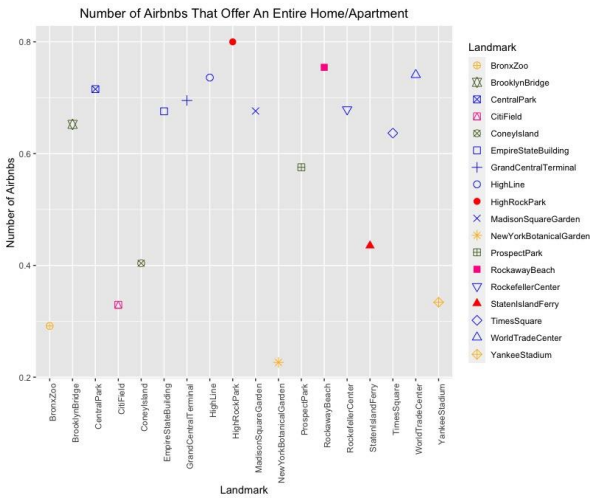
Figure F

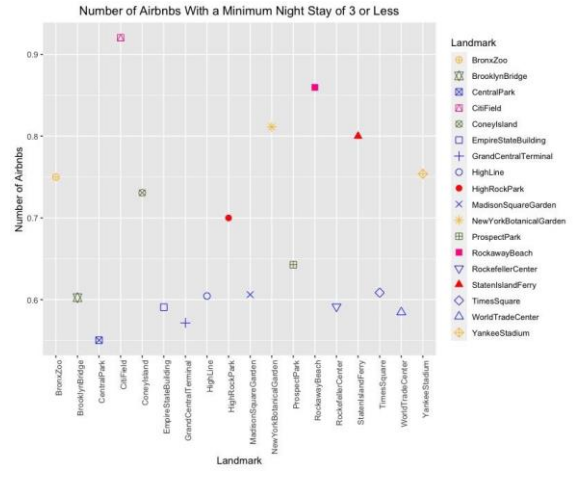

Number of Airbnbs Within 1 Mile
Number of Airbnbs
Landmark
BronxZoo
BrooklynBridge
CentralPark
CitiField
ConeyIsland
EmpireStateBuilding
GrandCentralTerminal
HighLine
HighRockPark
MadisonSquareGarden
NewYorkBotanicalGarden
ProspectPark
RockawayBeach
RockefellerCenter
StatenIslandFerry
TimesSquare
WorldTradeCenter
YankeeStadium


Number of Airbnbs Less Than $200
Number of Airbnbs
Landmark
BronxZoo
BrooklynBridge
CentralPark
CitiField
ConeyIsland
EmpireStateBuilding
GrandCentralTerminal
HighLine
HighRockPark
MadisonSquareGarden
NewYorkBotanicalGarden
ProspectPark
RockawayBeach
RockefellerCenter
StatenIslandFerry
TimesSquare
WorldTradeCenter
YankeeStadium


Number of Airbnbs Available At Least Six Months Out Of The Year
Number of Airbnbs
Landmark
BronxZoo
BrooklynBridge
CentralPark
CitiField
ConeyIsland
EmpireStateBuilding
GrandCentralTerminal
HighLine
HighRockPark
MadisonSquareGarden
NewYorkBotanicalGarden
ProspectPark
RockawayBeach
RockefellerCenter
StatenIslandFerry
TimesSquare
WorldTradeCenter
YankeeStadium


Number of Airbnbs With At Least 10 Reviews
Number of Airbnbs
Landmark
BronxZoo
BrooklynBridge
CentralPark
CitiField
ConeyIsland
EmpireStateBuilding
GrandCentralTerminal
HighLine
HighRockPark
MadisonSquareGarden
NewYorkBotanicalGarden
ProspectPark
RockawayBeach
RockefellerCenter
StatenIslandFerry
TimesSquare
WorldTradeCenter
YankeeStadium


Number of Airbnbs That Offer An Entire Home/Apartment
Number of Airbnbs
Landmark
BronxZoo
BrooklynBridge
CentralPark
CitiField
ConeyIsland
EmpireStateBuilding
GrandCentralTerminal
HighLine
HighRockPark
MadisonSquareGarden
NewYorkBotanicalGarden
ProspectPark
RockawayBeach
RockefellerCenter
StatenIslandFerry
TimesSquare
WorldTradeCenter
YankeeStadium


Number of Airbnbs With a Minimum Night Stay of 3 or Less
Number of Airbnbs
Landmark
BronxZoo
BrooklynBridge
CentralPark
CitiField
ConeyIsland
EmpireStateBuilding
GrandCentralTerminal
HighLine
HighRockPark
MadisonSquareGarden
NewYorkBotanicalGarden
ProspectPark
RockawayBeach
RockefellerCenter
StatenIslandFerry
TimesSquare
WorldTradeCenter
YankeeStadium

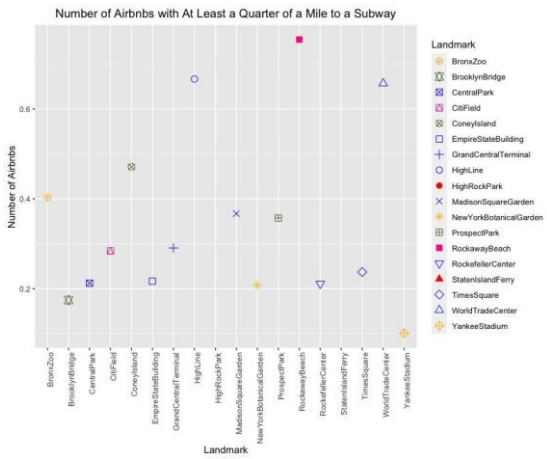


Figure G

Landmark	Borough	Neighborhood	Within 1 Mile	Price Less Than Two Hundred	Available At Least Six Months Out Of The Year	Atleast Ten Reviews	Offer An Entire Home Apartment	Minimum Night Stay Of Three Or Less	Atleast A Quarter Of A Mile To A Subway	Overall Ranking
Empire State Building	Manhattan	Midtown	1	1	2	2	1	2	5	1
High Line	Manhattan	Chelsea	6	9	7	6	6	6	1	6
World Trade Center	Manhattan	Financial District	10	10	8	10	9	10	7	10
Grand Central Terminal	Manhattan	Midtown	5	5	4	5	5	5	3	5
MadisonSquareGarden	Manhattan	Midtown	3	3	5	1	2	3	2	3
Times Square	Manhattan	Midtown	2	2	1	3	4	1	4	2
Rockefeller Center	Manhattan	Midtown	4	4	3	4	3	4	6	4
Central Park	Manhattan	Upper West Side	8	7	9	8	8	8	9	8
New York Botanical Garden	Bronx	Allerton	12	12	13	14	16	12	16	12
Yankee Stadium	Bronx	Concourse	11	11	11	11	11	11	12	11
Bronx Zoo	Bronx	Little Yemen	16	16	15	17	17	16	14	16
Brooklyn Bridge	Brooklyn	Brooklyn Heights	7	6	6	7	7	7	10	7
Prospect Park	Brooklyn	Prospect Lefferts Gardens	9	8	10	9	10	9	8	9
Coney Island	Brooklyn	Brighton Beach	13	13	14	15	13	14	11	13
Citi Field	Queens	Queens	14	14	16	12	15	13	15	14
Rockaway Beach	Queens	Queens	17	17	17	15	12	17	13	17
High Rock Park	Staten Island	Staten Island	18	18	18	18	18	18	N/A	18
Staten Island Ferry	Staten Island	Staten Island	15	15	12	13	14	15	N/A	15