Airbnbs in New York City

Team lol: Tamsin Connerly, Hannah Lee, Jasmine Xiang

2025-03-17

Introduction

The rise of short-term rental platforms, particularly Airbnb, has significantly disrupted the traditional hospitality industry and transformed urban housing markets worldwide. In New York City, one of the world's most popular tourist destinations, the impact of Airbnb has been particularly pronounced, raising questions about its effects on local communities, housing affordability, and the broader urban economy.

Previous research has identified several factors that impact Airbnb pricing. One study found that host attributes, site and property attributes, amenities and services, rental rules, and online review ratings all play significant roles in determining listing prices [@wang2017]. Furthermore, recent studies have provided evidence of Airbnb's influence on housing markets. Another study found that a 1% increase in Airbnb listings leads to a 0.018% increase in rents and a 0.026% increase in house prices [@barron2018]. This effect is more pronounced in areas with a lower share of owner-occupiers, suggesting that non-owner-occupiers are more likely to reallocate their properties from long-term to short-term rentals.

Our research question is: "How do various factors, such as bedroom number, room type, review scores, and distance from city center, influence the price of Airbnb listings in New York City?"

Understanding the determinants of Airbnb pricing in New York City is crucial for several reasons. Firstly, it can provide valuable insights for policymakers grappling with the challenges posed by the growth of short-term rentals, including potential impacts on housing affordability and neighborhood character [@toader2021]. Secondly, it can help hosts make more informed pricing decisions, potentially leading to more efficient market outcomes.

Based on existing literature and our understanding of the New York City housing market, we hypothesize that:

- Listings with more bedrooms will command higher prices, reflecting the premium placed on space in urban environments.

- The type of room (entire home/apartment vs. private room) will significantly impact pricing, with entire homes/apartments having a higher price.
- Higher review scores will be associated with higher prices, as positive feedback may justify premium pricing.
- Properties closer to the city center will be priced higher due to their convenient location and proximity to attractions.
- Properties in more affluent neighborhoods like Manhattan will have higher prices compared to less affluent ones like the Bronx because of real estate price differences in each borough.

Exploratory Data Analysis

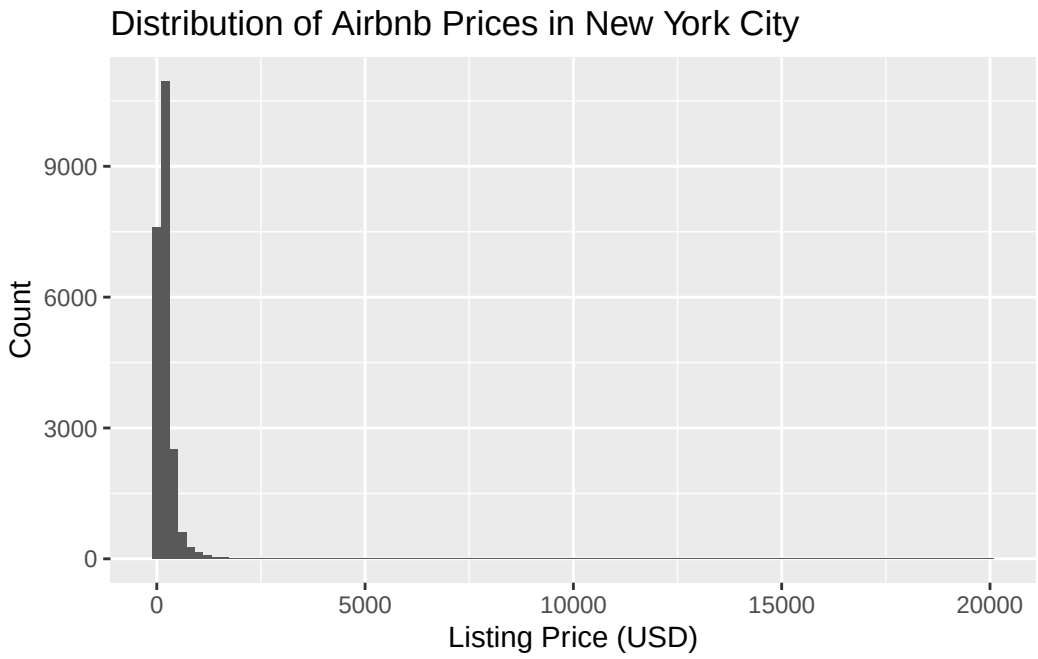
The Airbnb dataset that we are utilizing can be found on Inside Airbnb (https://insideairbnb.com/). Inside Airbnb has collected data on dozens of countries and cities, but we decided to focus on New York City. The data was sourced from publicly available data on the Airbnb website on March 1, 2025.

Each row in the dataset represents a unique Airbnb listing in New York City. Each of these correspond to individual properties available for rental on the platform and have many (53) variables such as name of the listing, latitude and longitude, room type, price, minimum number of nights required for booking, total number of reviews the listing has reviewed, and more. We are particularly interested in the following explanatory variables that we believe could impact the price of an Airbnb listing:

- Number of bedrooms
  - These could give insights into the size and comfort level of the Airbnb, likely affecting the price.
- Room type
  - Type of room (whether the listing is an entire home/apartment or a private room) can impact pricing
- Review scores (overall rating)
  - Listings with better or higher number of reviews could be priced higher because of higher perceived value and trustworthiness
- Distance from city center
  - Proximity to central locations can impact pricing (since it is more convenient and desirable)

- Neighborhood
  - Which New York City Neighborhood the Airbnb is located in. More affluent neighborhoods like Manhattan may have higher prices than neighborhoods like the Bronx.

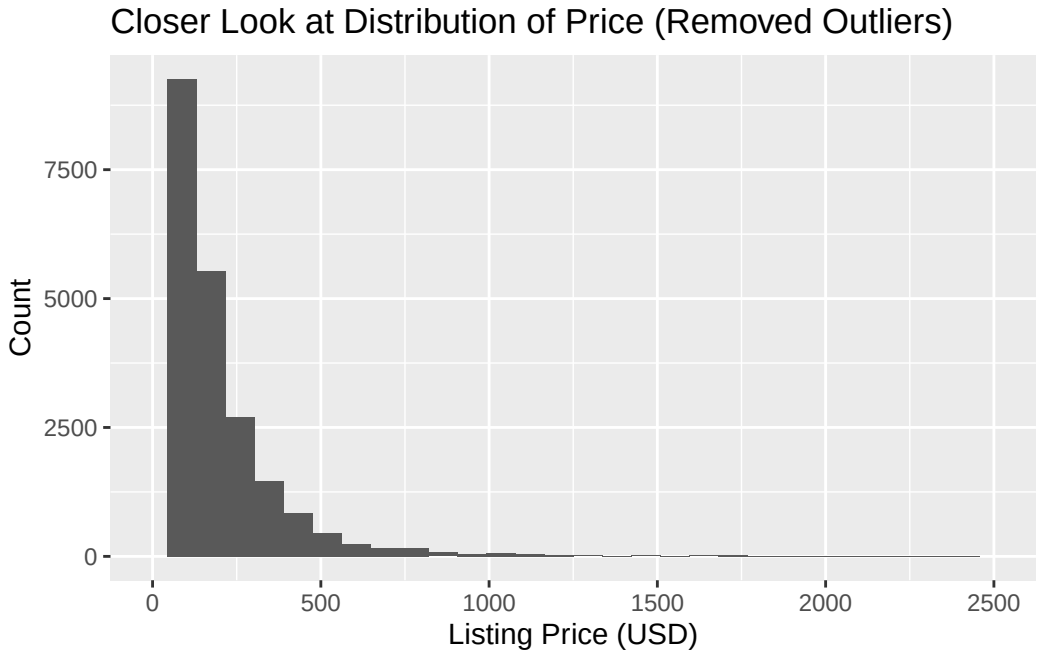
Response Variable - Price




```
# A tibble: 7 x 2
  Statistic Value
  <chr>     <table>
1 Min.          7.0000
2 1st Qu.      85.0000
3 Median      140.0000
4 Mean        213.8352
5 3rd Qu.     240.0000
6 Max.      20000.0000
7 NA's      15126.0000
```

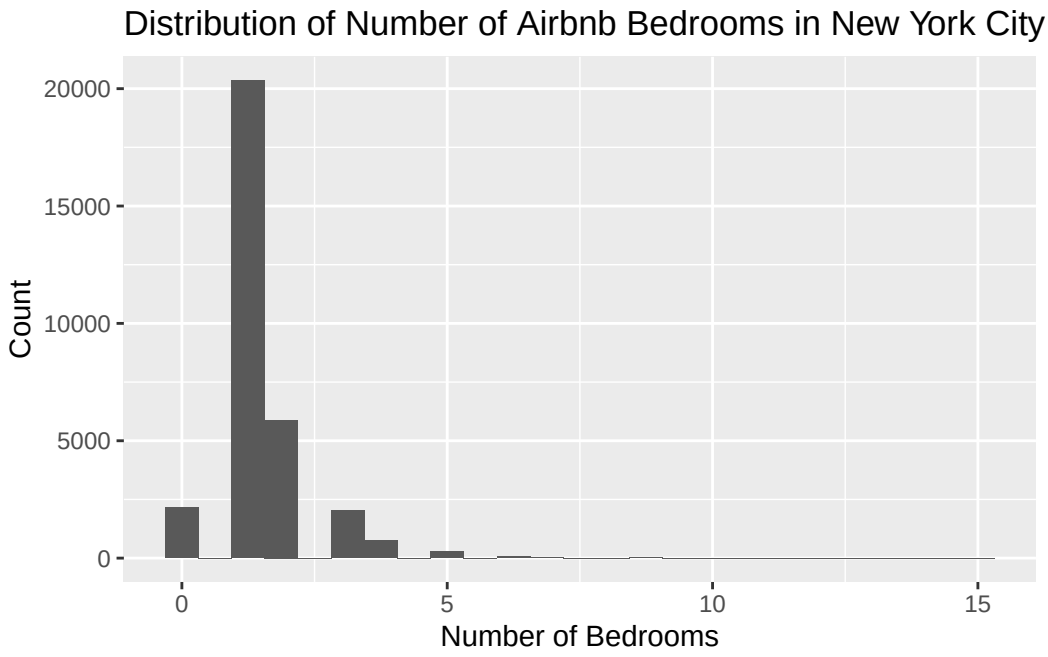
The distribution is pretty heavily right skewed, as can be seen from both of the histograms. It is difficult to analyze the distribution in the first histogram because there is an outlier at $20,000 and makes the bins and binwidth very narrow and zoomed out (since the range of the

data is too large). It is also clear that this outlier impacts the mean, since the median of $140 is quite a bit less than the mean of around $213.84, and the mean is roughly equal to the 3rd quartile which is also around $240. We plan to to remove this outlier as a result when doing our analysis, and we will go into more depth later and check for additional outliers.



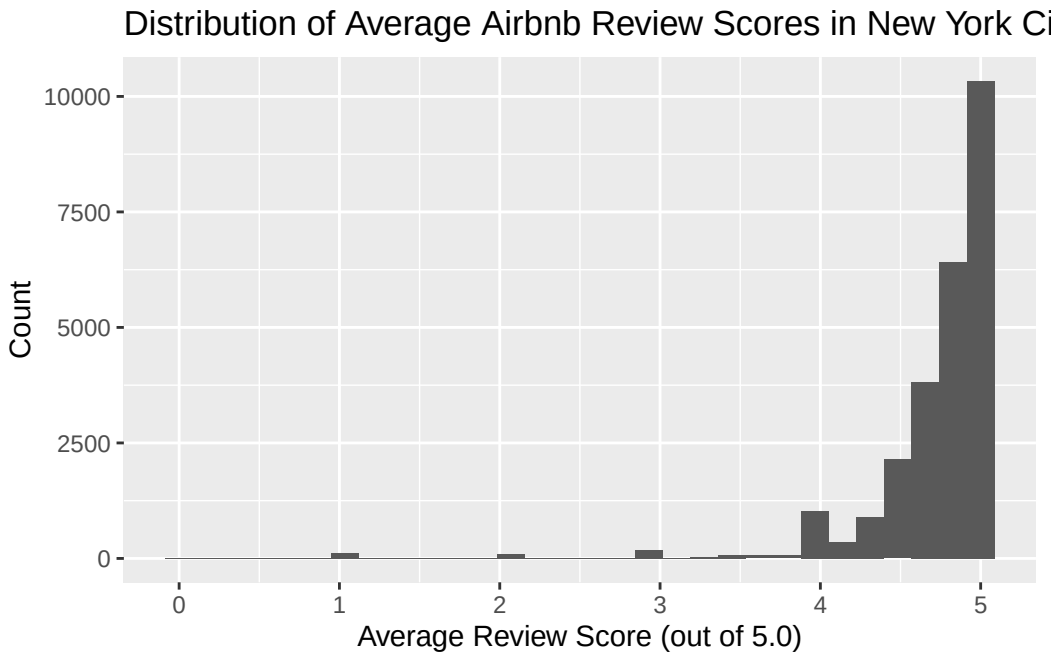
From the initial histogram, since the majority of the listings seem to fall between 0 and 2500 dollars, we visualized the distribution of the listings between this price range to get a better view of it. We can see that the distribution is still right skewed, and the vast majority of the listings seem to cost between $50-$200. There are a few more points visible that were not in the initial histogram that appear to be quite far from where the majority of the listings are clustered, so we will go back and check to see if they are greater than 1.5 x IQR and can be classified as outliers. We also plan to apply log transformation to this variable to address the skew of the response variable.

Predictor Variable - Bedrooms




```
# A tibble: 7 x 2
  Statistic Value
  <chr>     <table>
1 Min.         0.00000
2 1st Qu.      1.00000
3 Median       1.00000
4 Mean         1.37414
5 3rd Qu.      2.00000
6 Max.        15.00000
7 NA's      5911.00000
```

Predictor Variable - Review Scores




```
# A tibble: 7 x 2
  Statistic Value
  <chr>     <table>
1 Min.          0.000000
2 1st Qu.       4.650000
3 Median        4.860000
4 Mean          4.724751
5 3rd Qu.       5.000000
6 Max.          5.000000
7 NA's      11787.000000
```

Predictor Variable - Room Type

Predictor Variable - Neighborhood