Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/131/2024.20711

What Factors Influence the Price of Airbnb Listings in New York City?

Haoran Wang[1,*], Zejia Su[2], Yang Yang[3]

[1]Department of Math, Sun Yat-sen University, Guangzhou, China
[2]School of Arts and Sciences, Rutgers University-New Brunswick, New Brunswick, United States of America
[3]Department of Statistics, Ohio State University, Columbus, United States of America

[1]wanghr63@mail2.sysu.edu.cn
[2]kyrie11su@gmail.com
[3]Yang.7048@osu.edu
*corresponding author

Abstract. This paper explores what factors influence the price of Airbnb listings in New York City. As a key participant in the short-term rental business, Airbnb contributes significantly to the local economy by providing a platform for connecting property owners and tourists. The purpose of this study is to improve our understanding of the ways in which these factors impact Airbnb listings' profitability and to offer suggestions for the most effective pricing tactics. This study uses a publicly available dataset from Kaggle that contains details on Airbnb listings in New York City. This dataset includes information such as room type, location, neighborhood, cleanliness, availability of bedrooms and bathrooms, accommodation capacity, available days of the year, and the number of beds, along with feedback from guests about their stays. By conducting a thorough correlation analysis, the research examines how these different factors affect nightly prices. In addition, the research studies the relation between occupancy and price using Times Square as a center point, which is calculated by Haversine formula. The findings indicate a strong relationship between these factors and Airbnb's economic performance. According to the results, properties in prominent locations, with higher cleanliness ratings, more bedrooms and baths, bigger accommodation capacity, and more available days, tend to command higher costs and higher occupancy rates. This highlights the importance of these traits in making Airbnb rentals more profitable. Additionally, the study provides helpful advice to property owners on how to improve their listings. Hosts may considerably enhance room occupancy and total revenues by modifying parameters like as pricing, location, room type, and amenities in response to the findings. This study contributes to a better knowledge of short-term rental market dynamics and provides useful advice for optimizing economic returns in the Airbnb marketplace. correlation analysis; Haversine formula.

Keywords: Airbnb; economic benefits; hospitality sector; correlation analysis; Haversine formula.

1. Introduction

Airbnb is an online marketplace for short-and-long-term homestays and experiences in various countries and regions[1], allowing property owners to rent out their spaces to guests. Founded in 2008, Airbnb operates through an online marketplace accessible via its website and mobile app. The concept for Airbnb arose from the necessity to address the scarcity of cheap hotel choices during popular events. The company was founded by Brian Chesky, Joe Gebbia, and Nathan Blecharczyk, who saw an opportunity to create a platform that could connect travelers with hosts offering spare rooms, entire homes, or unique accommodations[2]. Recognizing the broader potential of this concept, they expanded it to include various types of properties, such as homes, apartments, private rooms, and unique spaces like boats and treehouses.

Several factors contributed to Airbnb's rapid growth. First, economic conditions had a big impact. During the 2008 financial crisis, renting out vacant areas became an appealing opportunity for homeowners to generate additional money. This economic incentive was crucial in encouraging many property owners to list their spaces on Airbnb. Additionally, technical developments contributed to the platform's growth. The widespread availability of the internet and the rise of smartphones made it easier for people to access and use Airbnb. The platform's user-friendly design and strong mobile app increased its use and convenience.

Furthermore, there was a significant shift in travel preferences. Many travelers began seeking unique, local experiences rather than staying in traditional hotels. Airbnb capitalized on this trend by offering a wide range of accommodation options that catered to diverse preferences and budgets. These qualities produced a sense of trust and dependability, which was critical to the platform's growth and acceptability among users. Airbnb quickly expanded beyond the United States, entering markets across Europe, Asia, and Latin America. By 2011, the platform had facilitated over one million nights booked[3].

This study focused on Airbnb listing in New York City for numerous convincing reasons. Firstly, New York is one of the most visited cities in the world, attracting millions of tourists every year. The strong demand for tourism produces a great requirement for a variety of hotel choices. New York's unique neighborhoods offer a wide range of experiences, making it an ideal city to study the impact of location on Airbnb pricing and guest preferences. The city's dynamic economy, which combines business, culture, and entertainment, makes it an important market for determining the economic advantages of Airbnb. Moreover, the presence of numerous hotels and other lodging options makes New York City a highly competitive market. This competitiveness provides a rich environment for examining Airbnb's strategies and performance. Therefore a more convincing understanding of Airbnb's strategic responses and modifications can be gained by examining the company's activities in a highly regulated and competitive market.

Airbnb's rise from a simple notion to a worldwide hospitality behemoth demonstrates the potential of new solutions to satisfy changing customer demands. Airbnb has transformed how people travel and experience the globe by harnessing technology, building community trust, and adjusting to market realities. New York City serves as an exemplary case study due to its tourism appeal, economic importance, and regulatory landscape, making it an ideal focus for this research. Through this study, we can gain a deeper understanding of Airbnb's impact on the hospitality industry and the various factors that contribute to its success in one of the most vibrant and complex markets in the world. This research aims to provide a comprehensive analysis of Airbnb's operations in New York City, examining the interplay of economic, regulatory, and competitive factors that shape the platform's performance and influence its strategic decisions.

2. Literature review

Airbnb is an online platform that allows individuals to rent out their spaces as tourist accommodations. The prices of Airbnb listings are influenced by a variety of factors, with the most critical determinants frequently being location, accommodation attributes, host characteristics, and market dynamics, as identified in numerous studies.

One significant study focused on four Spanish Mediterranean Arc cities, identifying key factors influencing the daily price of Airbnb listings, including accommodation attributes, advertisements, host features, tourism-related environmental characteristics, and the listing's location[4]. The study highlighted the critical role that location plays in setting prices, particularly in places where there is a high demand for tourism.

In addition to location, neighborhood characteristics also play a crucial role in determining the average nightly price and monthly earnings of Airbnb listings. Transportation accessibility, in particular, has the most significant impact on both average overnight prices and average monthly earnings. Furthermore, professionally managed Airbnb listings generally yield higher monthly revenue, indicating the importance of management practices in maximizing profitability[5].

The price is also heavily impacted by the host's cumulative experience and market demand on particular booking dates. This finding validates the idea that seasoned hosts are more adept at adjusting their prices responsively to maximize the performance of their listings[6]. In a more comprehensive analysis of 180,533 lodging rental listings across 33 cities on Airbnb.com, the factors influencing prices were classified into five categories: host attributes, site and property attributes, amenities and services, rental rules, and online review ratings[7]. This comprehensive approach demonstrates the varied range of elements that influence Airbnb pricing in different contexts.

Interestingly, the influence of these elements can differ significantly depending on the scale of operations. For example, owners who manage multiple listings are more influenced by the unique characteristics of each listing when setting prices than those who manage a single property. Moreover, the reputation of a property can negatively impact the price, supporting the view that price determinants are highly context-dependent and may vary by region[8].

Given the importance of location in determining Airbnb prices, particularly in competitive urban markets, it is crucial to examine how these dynamics play out in a city as diverse and economically significant as New York. As Airbnb's third-largest market worldwide, New York City generated over $650 million in host revenue in a single year, underscoring the platform's significant economic impact[9]. The distribution of active listings across the city reveals hotspots in Midtown Manhattan, the Lower East Side, Williamsburg, and Bushwick in Brooklyn, emphasizing the importance of specific neighborhoods in influencing pricing strategies. This suggests that in high-demand markets like New York City, location plays a crucial role, alongside other factors such as accommodation attributes and host management practices.

Overall, existing hedonic research on Airbnb has identified several important factors that influence prices and listing performance, but the results are frequently inconsistent[10]. Some scholars have attempted to develop global models based on the price determinants of housing rentals in the sharing economy. However, these models often fail to consider the differences among cities, as the interactions between city-specific variables and other factors have not been thoroughly explored.

To determine the factors that influence Airbnb accommodation prices, particularly in New York City, this paper builds on previous research by applying correlation analysis and multiple linear regression models. This research will use data that includes information on room types, locations, cleanliness, and guest feedback, as these factors are likely to influence Airbnb prices within the unique context of New York City. Considering the dynamic socio-economic environment, the elements influencing prices are expected to evolve over time. Therefore, future research should continue improving and revising these models to accurately represent present trends.

3. Methodology

In this study data collection and preparation stage we utilized the dataset from Kaggle that focuses on room types and locations of Airbnb in New York City, along with cleanliness ratings and guest feedback for analysis purposes. We cleaned the data by removing duplicates and handling missing values through imputing average value. Data visualization were used in the research, such as geographical distribution map showing listing distributions across different neighborhoods and bar charts for room type and neighborhood group distribution. Correlation analysis was also used in this research to identify key factors influencing prices. These visualizations and analyses provided a solid foundation for the subsequent application of a multiple linear regression model to optimize pricing strategies. In this research the main tools and techniques of handling and examining data utilized is Python programming with the assistance of libraries like Pandas and Scikit Learn. The dataset can be categorized into two parts, one contains information of listings and the other consists of feedback from guests. The listing data encompasses both aspects such as neighborhood and room type as well as numerical attributes like latitude, longitude and price. The guest feedback data comprises reviews from visitors along with numerical ratings, for various aspects of their stay. There is a lot of information in these databases that can be used to analyze the factors that influence the prices of Airbnb listings, in New York City.

3.1. Data Collection

Since 2008, Airbnb has enabled guests and hosts to enhance travel opportunities, offering a more unique and personalized way to experience the world. The two datasets for this study were obtained from Kaggle, a well-known data repository, that provides comprehensive information on Airbnb listings in New York City. The first dataset describes 47,906 listings activity and metrics in New York City for 2019. The dataset contains numeric features such as latitude, longitude, and price and categorical features such as neighborhood, and room type. The other dataset includes 48,864 guest feedback and numerical ratings on these listings.

3.2. Data Cleaning and Preprocessing

3.2.1. Merging the two datasets The two datasets were merged based on the listing ID to create a comprehensive dataset with 30 columns. This merging process was accomplished by using the Pandas library in Python. This combined dataset allows for a more detailed analysis by incorporating both listing attributes and guest feedback.

3.2.2. Handling Missing Values 19,444 Missing values in the new dataset were identified and imputed using the mean values of the respective columns. This was accomplished using the Pandas library in Python with the 'fillna()' function, ensuring the dataset remained complete and reducing the risk of bias from incomplete data.

3.2.3. Removing Duplicates 37 Duplicate records were removed using the 'drop_duplicates()' function from Pandas. This step was essential to maintain the uniqueness and accuracy of the data, ensuring each listing was represented only once.

3.3. Data Visualization

3.3.1. Geographical Distribution Maps Geographical distribution maps provide a visual representation of space, allowing us to understand the layout and relationships between different geographic features. By studying maps, distances, and relative positions of various locations can be reveal clearly, facilitating easier navigation and spatial orientation. Using Matplotlib and Seaborn libraries of Python, a geographical distribution map was created to visualize the distribution of Airbnb listings across different neighborhoods in New York City (fig.1), which can help understanding the spatial distribution of Airbnb listings.

DOI: 10.54254/2755-2721/131/2024.20711

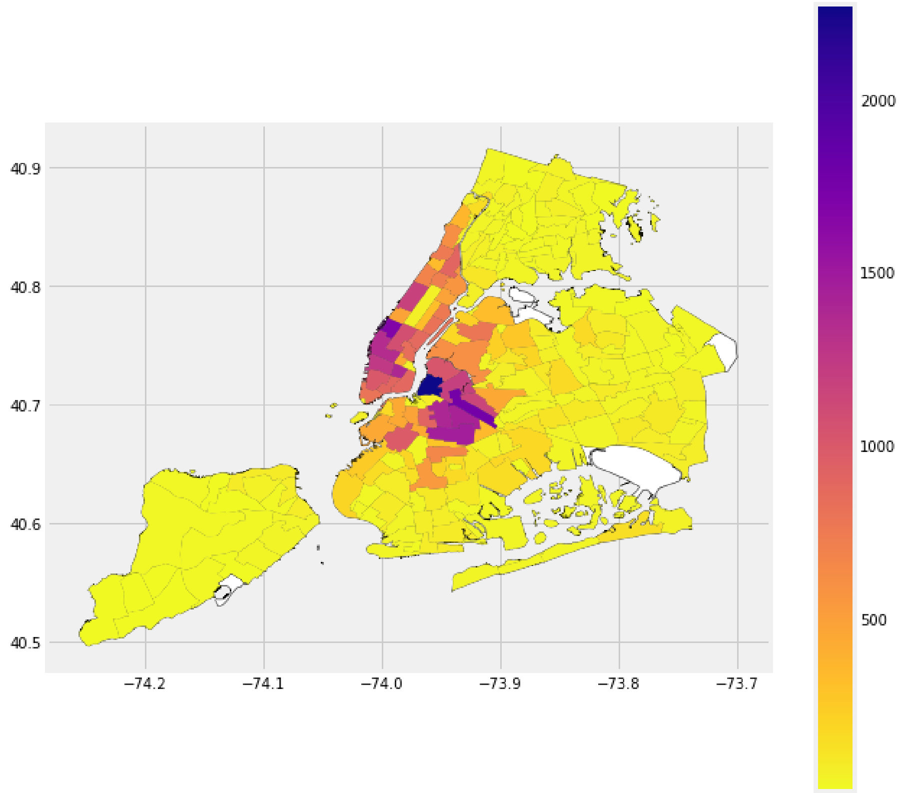


Figure 1: Number of Airbnb listings by neighborhood

3.3.2. Bar Charts Bar charts are a powerful tool for data visualization, providing a clear, effective, and versatile way to present and interpret data. They can clearly show the differences between categories or groups. The length of each bar is proportional to the value it represents, making it easy to compare data briefly.

Bar charts (fig.2) were generated to show the distribution of different room types (entire homes/apartments, private rooms, shared rooms) within each neighborhood, highlighting trends and preferences in different neighborhoods. The map illustrates that Manhattan has the most rooms. The majority of room kinds in Manhattan are Entire Home, which is a unique feature. Because in other communities, the number of private rooms exceeds that of the complete home.

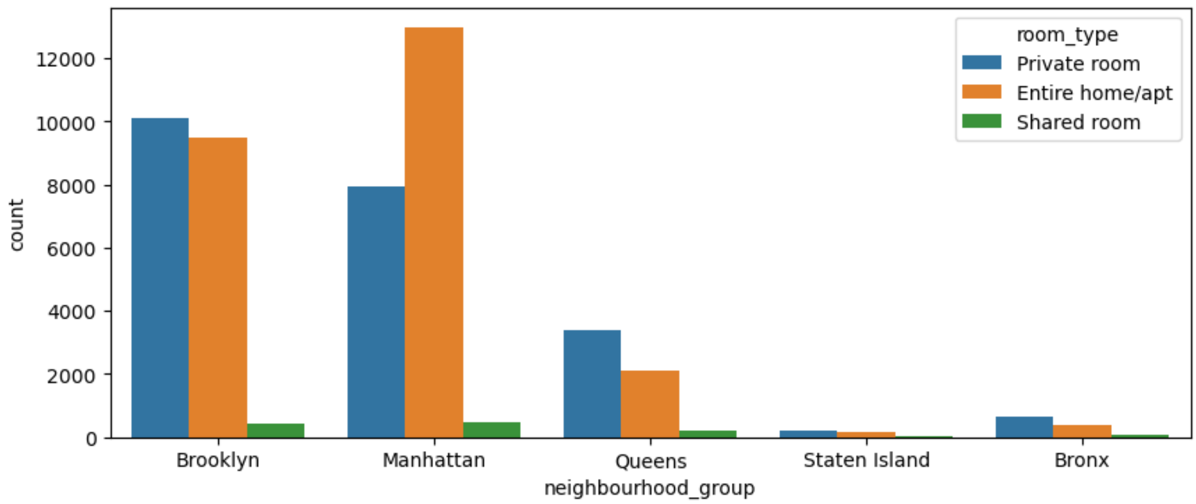


Figure 2: Distribution of room types in each neighborhood group

3.3.3. Box Plots Box plots are particularly useful for identifying outliers in a dataset. Outliers are shown as individual points outside the whiskers of the box, making them easy to spot. Meanwhile, Box plots are effective for comparing the distribution of data across multiple groups or categories. By placing multiple box plots side by side, medians, variability, and the presence of outliers can be easily compared between different datasets

Box plots (fig.3 and fig.4) were used to display the price distribution across different neighborhoods and room types. This visualization helped identify median prices and outliers, providing a clear picture of pricing trends across the city.

The average price in Manhattan surpasses that of other areas, which is not surprising. It is logical that entire homes are priced higher than other types of accommodations, and similarly, private rooms tend to be more expensive than shared rooms. One contributing factor to Manhattan's higher prices could be the prevalence of entire home listings. Since entire homes generally command higher prices, and Manhattan has the largest number of these listings, this likely drives up the average price in the area.

DOI: 10.54254/2755-2721/131/2024.20711

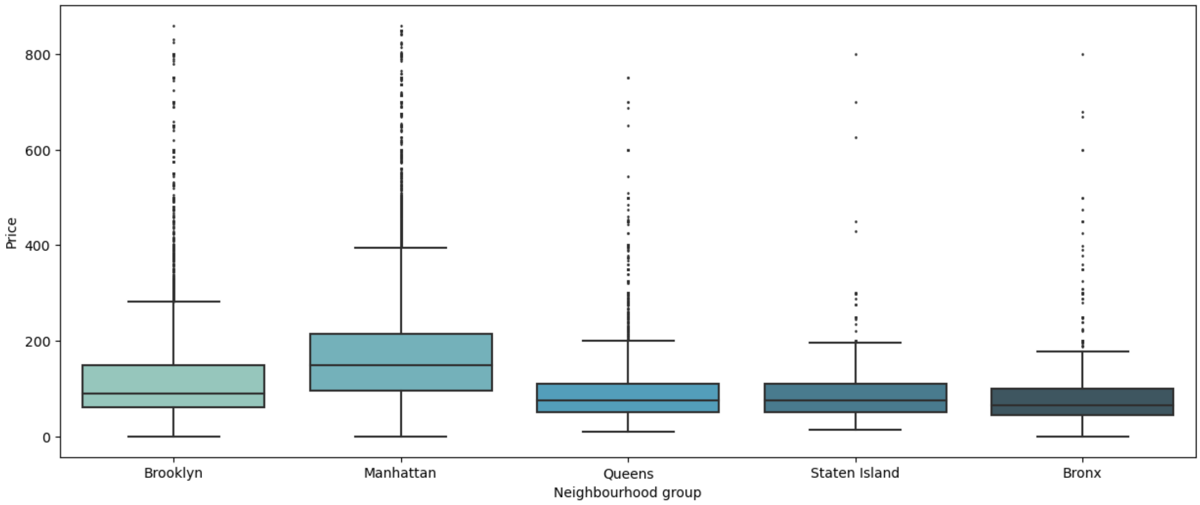


Figure 3: Density and distribution for each neighborhood group

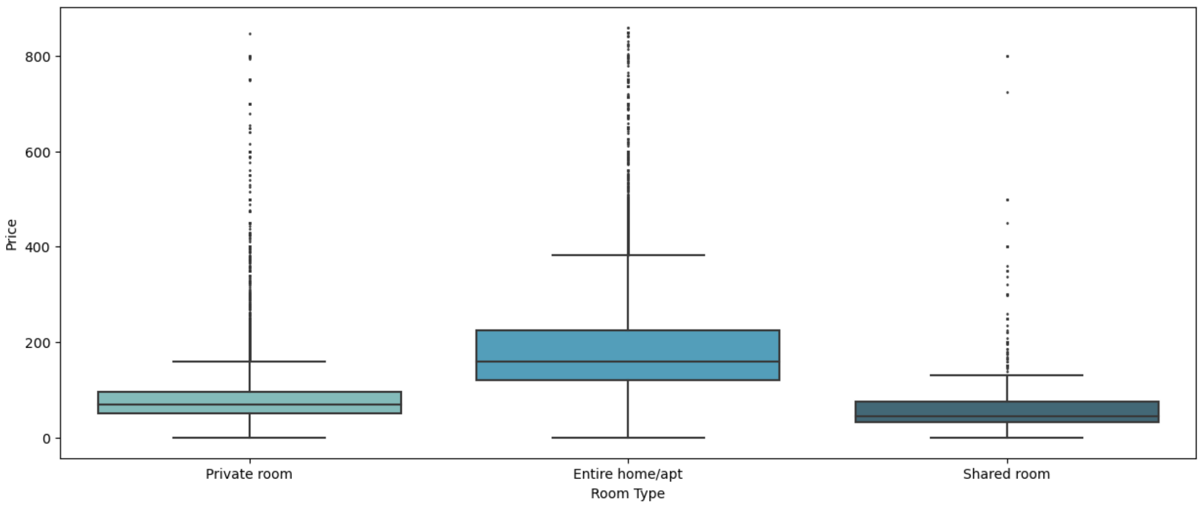


Figure 4: Density and distribution for each room type

4. Data Analysis And Results

4.1. The correlation analysis

4.1.1. Calculated Distance The distances between listings and Times Square were calculated by the Haversine formula. The latitude and longitude of Times Square is (40.71643, -74.00666). The coordinates were converted from degrees to radians and then applies the Haversine formula to compute the great-circle distance between two points on the Earth's surface, with the Earth's radius taken as 6371 kilometers. This was applied to each row in the dataset which contains latitude and longitude data. The results were added to a new column that displays the calculated distance from each point to the fixed point.

DOI: 10.54254/2755-2721/131/2024.20711

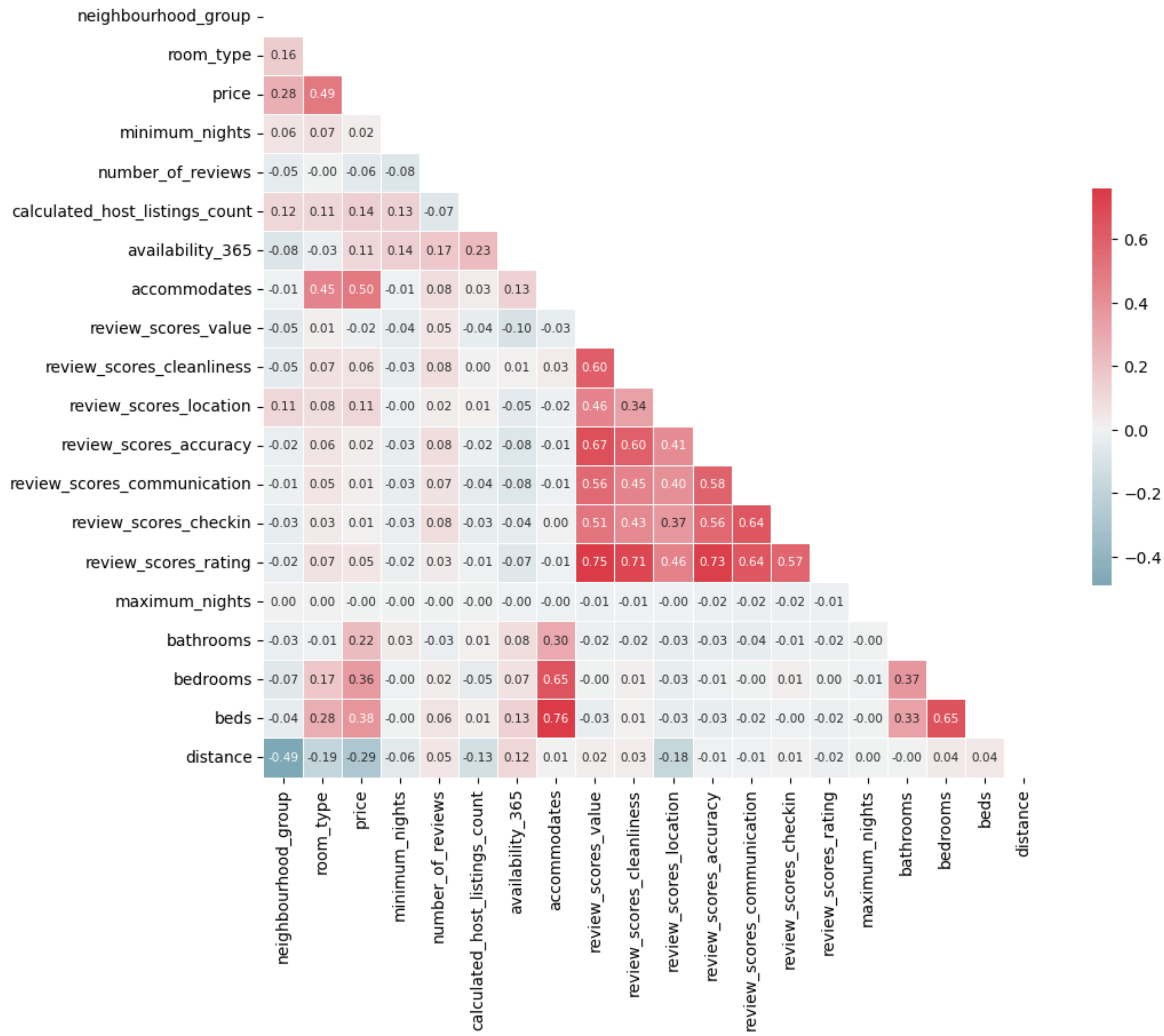


Figure 5: Heatmap of all the Features

4.1.2. Results The heatmap shows significant relationships between the several factors influencing Airbnb listing prices. It's interesting to note that the number of beds, baths, and bedrooms has a significant positive link with pricing; this means that larger listings usually command higher costs. Additionally, there is a substantial negative link between price and proximity to Times Square, indicating that listings near Times Square usually cost more. Furthermore, there is a significant association between greater review scores in terms of value, cleanliness, location, and correctness and higher pricing, suggesting that listings with better reviews can fetch higher prices. Along with showing a positive correlation between reviews per month and the total number of reviews, the analysis also shows substantial positive correlations between various review scores, underscoring the tendency for frequently reviewed listings to have more reviews overall. This thorough investigation emphasizes how important it is for both qualitative and physical characteristics to have an impact on Airbnb prices.

DOI: 10.54254/2755-2721/131/2024.20711

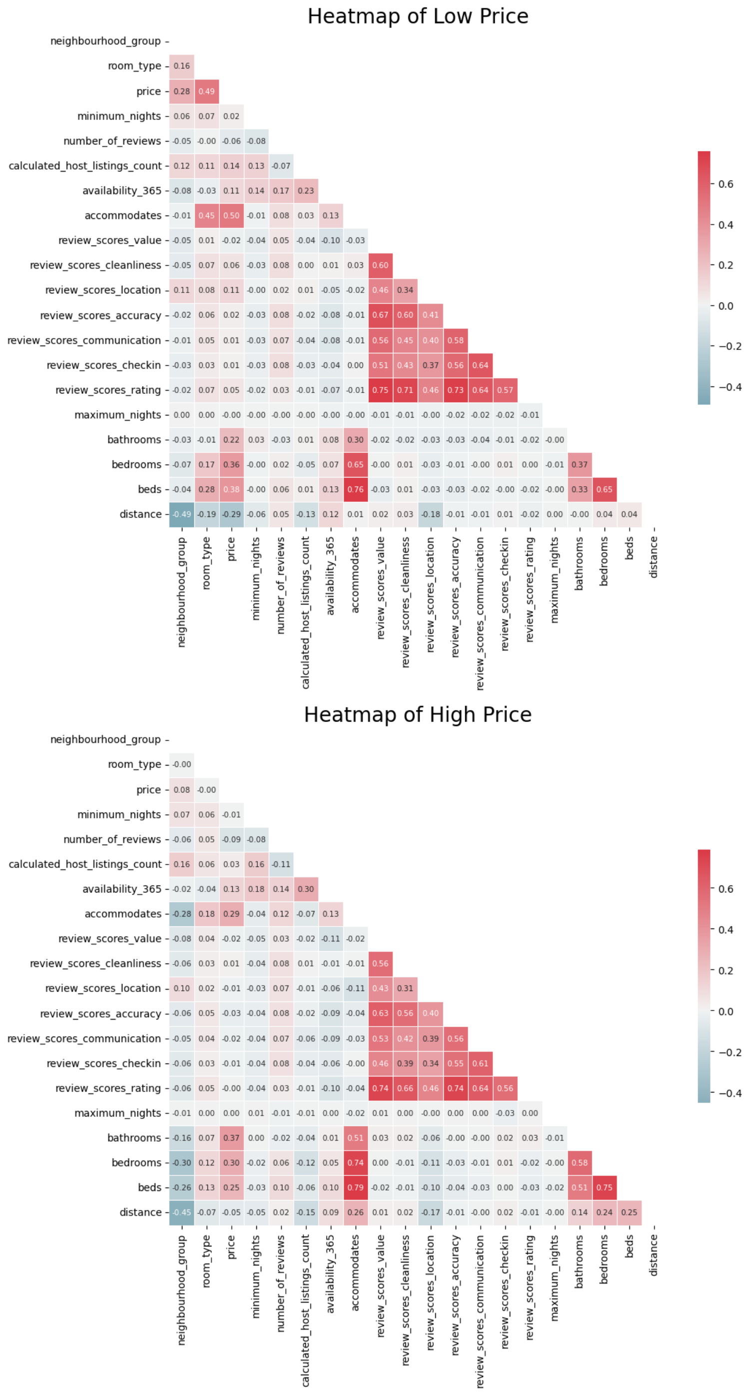


Figure 6: Heatmap of all the Features After Dividing

4.1.3. Advanced Analysis The two heat maps, which split the listings into two categories according to price points, show the relationships between several characteristics of Airbnb listings and their costs. The importance of lodging capacity—which includes the number of beds, visitors, bathrooms, and bedrooms—in influencing Airbnb listing rates is demonstrated by the two heatmaps. This indicates that listings with more sleeping arrangements tend to be priced higher. It is noteworthy that the distance from key locations is inversely correlated with the price of low-priced listings. Conversely, the room type and review scores regarding locations demonstrate a moderate positive correlation with low-priced

DOI: 10.54254/2755-2721/131/2024.20711

listings. However, these features have a negligible impact on the price of high-priced listings, suggesting the existence of disparate pricing dynamics across market segments. These findings indicate that the majority of individuals who select low-priced listings are likely to be tourists seeking a residence in close proximity to renowned attractions at a reduced cost.

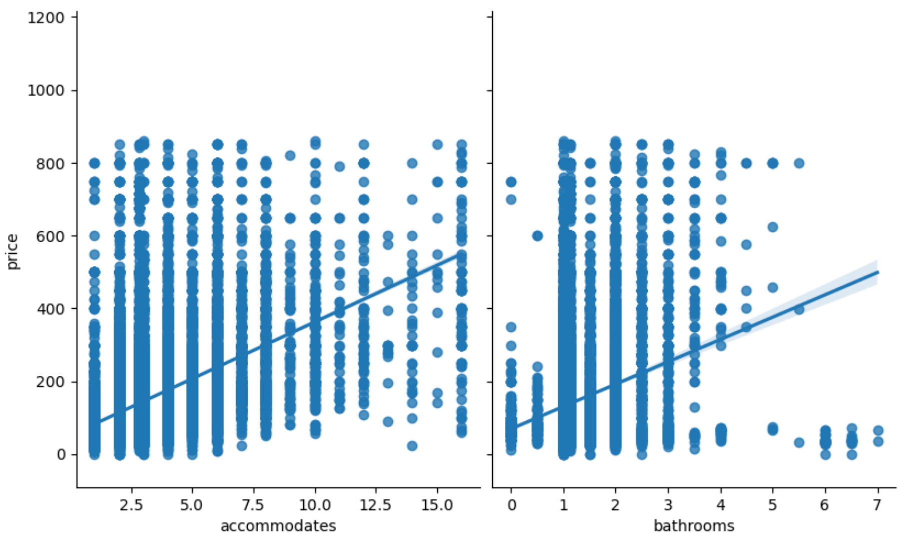
4.2. Multiple Linear Regression Model

4.2.1. Model Selection The multiple linear regression model was chosen due to its ability to quantify the relationship between multiple independent variables and the dependent variable (price). The independent variables included the physical attributes of the listings such as the number of accommodations, bedrooms, bathrooms, beds, and distance.

4.2.2. Variable Identification $Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 + b$

Y : "$prices$"($unit$: $dollar$)
X_1 : "$accommodates$"
X_2 : "$bathrooms$"
X_4 : "$bedrooms$"
X_5 : "$beds$"
X_6 : "$distance$"($unit$: $kilometer$)

4.2.3. Model Training and Testing The dataset was split into training sets and testing sets to train and validate the model. The training sets account for a quarter of the total. The 'train_test_split' function from Scikit-learn was used for this purpose, ensuring that the model's performance could be assessed on unseen data. The 'LinearRegression' model from Scikit-learn was used to train the model on the training set using the 'fit()' method.



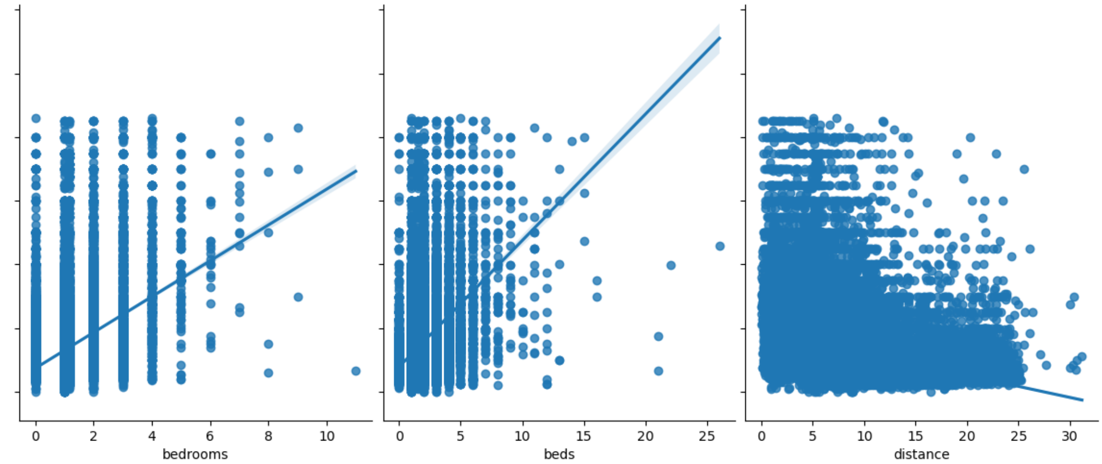


Figure 7: Regression Images Between Price and Other Variables

4.2.4. Results and Coefficient Analysis Regression coefficient:

$'accommodates'$: 27.76540599779547
$'bathrooms'$: 19.45066147141844
$'bedrooms'$: 10.604834224277209
$'beds'$: -1.140172143042945
$'distance'$: -7.003006919889155

The regression coefficients were analyzed to understand the impact of each independent variable on the dependent variable (price). These coefficients represent the marginal effect of each independent variable on the price.

According to the regression coefficients, the number of accommodations, bathrooms, bedrooms, and distance can greatly affect the price while the number of beds does not, indicating that larger listings and easy access to Time Square tend to command higher prices.

5. Discussions And Limitations

Our analysis highlights that properties located in prime areas, boasting high cleanliness ratings, multiple bedrooms and bathrooms, larger accommodation capacities, and greater availability tend to command higher prices and enjoy better occupancy rates. These findings underscore the importance of these characteristics in enhancing the profitability of Airbnb rentals.

The insights from this study are valuable for both travelers and homeowners. For travelers, the findings offer a range of accommodation options tailored to various needs and budgets, thereby enhancing their living experience in New York City. The more they know about the details of the listing, the lower their risk of information asymmetry, and this facilitates easier identification of suitable places to stay. Homeowners can utilize our model and findings to optimize their profits. By improving the quality of their listings and adjusting prices based on the identified factors. They can also promote their listings by emphasizing the factors mentioned in the model to attract the attention of travelers to increase occupancy. As the market of Airbnb continues to expand and develop, as its main competitor, the hotel industry, is bound to cause a potential impact. Some cutting-edge research has mentioned the effect of Airbnb on hotel performance (Dogru et al., 2021).

Indeed, this study has limitations. The first significant limitation is that the dataset used is only updated to 2019 and does not reflect real-time data. Databases are constantly being updated, and the lack of real-time data may affect the relevance and applicability of our findings to the current market. The dynamic nature of the short rental market means that trends influencing prices and occupancy rates can change rapidly. However, since most of these data are missing randomly, this has a slight effect on our overall results, although it leads to larger standard errors. While, the method of dealing with missing values by estimating the mean, while practical, can be biased. While this approach ensures the completeness of the data set, it may not accurately reflect the variability of the data. Future studies should incorporate more recent data to ensure the continued relevance of their conclusions.

Additionally, the dataset includes some default five-star reviews that may not accurately reflect guest experiences. Many users avoid leaving critical reviews for various reasons, making it difficult to distinguish between default and genuine five-star reviews. While removing default reviews could improve accuracy, it risks losing valuable information. Future research should address these challenges to enhance data accuracy.

Moreover, our study focuses solely on the New York market. While this provides detailed insights into a specific area, it limits the generalizability of the findings to other locations. Different cities and regions may have unique factors influencing Airbnb rental prices and occupancy rates. the significant presence of Airbnb users in the U.S., despite Airbnb's global reach. As the platform expands and the user base becomes more international. Therefore, future research should examine multiple markets to provide a more comprehensive understanding of the dynamics at play.

Still, another limitation is the possible presence of unobserved variables that affect the results. While we included many factors in our analysis, there may be other variables affecting Airbnb prices and occupancy rates that were not accounted for. For instance, local events, seasonal trends, and changes in tourism patterns can all impact rental dynamics. More factors will be included in future studies.

6. Conclusion

Our analysis investigates what factors influence the price of Airbnb listings in New York City and hopes to see some causal relationship. Based on a review of relevant literature, we conducted a more comprehensive statistics and investigation. We used two datasets from Kaggle, one detailing 47,906 Airbnb listings in New York City, including room type, cleanliness, availability, and number of bedrooms and toilets, and the other containing 48,864 pieces of feedback from residents, and to ensure that the data was accurate in some ways including data collection, data cleaning and interpolating missing values using the Pandas library in Python to calculate the averages to ensure the accuracy. Various visualization techniques were used to help us understand the spatial and categorical distribution of Airbnb listings in New York City, including bar charts and box plots. Descriptive statistics, including means, standard deviations, minimums, and maximums, were calculated to summarize the data set. Correlation analysis identified key factors that influence price, revealing that larger listings with more bedrooms and bathrooms, as well as those closer to Times Square, tend to be priced higher. Additionally, higher cleanliness, location, accuracy, and value scores were correlated with higher prices. In addition, the heatmap analyzes and shows how each different factor influences Airbnb pricing.

Multiple linear regression models were used to quantify the relationship between several independent variables and listing prices. The independent variables included the number of accommodations, bathrooms, bedrooms, beds, and distance from Times Square. The model was trained and tested using Scikit-learn's "train_test_split" function, with the training set comprising a quarter of the total dataset. This evidence shows that pricing accommodations in the NYC Airbnb is a multi-faceted and intricate task. Eventually, we reflect on some limitations in the dataset and results and then suggest further studies to enrich the conclusion. We find that the number of bathrooms and bedrooms has a significant positive effect on price and occupancy, while the number of beds has no effect. And, there is also an indispensable negative phase correlation between price and reviews and distance to Times Square, This may be an artifact of residents' preferences for cheaper sharing accommodations, resulting in a higher volume of reviews for an estate at the lower endpoint of the price spectrum. which indicates that properties that are larger and have easier access to prime locations are more expensive.

References

[1] ”Company Overview of Airbnb Inc”. Bloomberg L.P. Archived from the original on January 8, 2018. [Link](https://www.sciencedirect.com/science/article/abs/pii/S0278431916305618).
[2] ”Nathan Blecharczyk”. Forbes. Archived from the original on April 13, 2019. Retrieved April 11, 2019.
[3] Reddy, S. N. (2023, October 18). How Airbnb became a global phenomenon: The growth story you need to know. *Medium*. [Link](https://medium.com/@satya_/how-airbnb-became-a-global-phenomenon-the-growth-story-you-need-to-know-f5436cfd5ed8).
[4] Perez-Sanchez, V. R., Serrano-Estrada, L., Marti, P., & Mora-Garcia, R.-T. (2018, December 5). The what, where, and why of Airbnb price determinants. *MDPI*. [Link](https://www.mdpi.com/2071-1050/10/12/4596).
[5] Deboosere, R. (n.d.). Full article: Location, location, and professionalization: A multilevel hedonic analysis of Airbnb listing prices and revenue. [Link](https://www.tandfonline.com/doi/full/10.1080/21681376.2019.1592699).
[6] Magno, F., Cassia, F., & Ugolini, M. M. (2018). Accommodation prices on Airbnb: Effects of host experience and market demand. *The TQM Journal*, 30(5), 608-620.
[7] Balaguer, J., Becerra, M., Belk, R., Ert, E., Fang, B., Gkiotsalitis, K., Heo, C. Y., Hung, W. T., Israeli, A. A., Kaplan, A. M., Karlsson, L., Lee, C. G., Masiero, L., Ren, L., Saló, A., ... Gansky, L. (2017, January 18). Price determinants of sharing economy-based accommodation rental: A study of listings from 33 cities on airbnb.com. *International Journal of Hospitality Management*.
[8] Toader, V., Negrușa, A. L., Bode, O. R., & Rus, R. V. (2022). Analysis of price determinants in the case of Airbnb listings. *Economic Research-Ekonomska Istraživanja*, 35(1), 2493-2509.
[9] Wachsmuth, D. (2018, June 20). Airbnb and the Rent Gap: Gentrification through the Sharing Economy: Upgo-mcgill. *UPGo*. [Link](https://upgo.lab.mcgill.ca/publication/airbnb-and-the-rent-gap/).
[10] Dogru, T., Mody, M., Line, N., Hanks, L., Suess, C., & Bonn, M. (2021). The effect of Airbnb on hotel performance: Comparing single- and multi-unit host listings in the United States. *Cornell Hospitality Quarterly*, 63(3), 297–312. [Link](https://doi.org/10.1177/1938965521993083).