**Executive Summary**

The analysis identifies the key variables influencing the nightly rental price for Airbnb listings in New York City. The most significant factors include location, room type, listing characteristics, and host activity. Listings in Manhattan command the highest prices, while Staten Island properties are priced the lowest. Room type also plays a critical role; entire homes/apartments charge significantly higher prices than private or shared rooms. Additionally, listings available year-round are priced higher, reflecting their appeal to a broader audience. Although the number of reviews and reviews per month negatively correlate with price, this may reflect increased competition among popular listings. Based on these insights, Airbnb management should optimize pricing strategies by focusing on properties in high-demand locations, promoting entire home offerings, and ensuring consistent availability. These recommendations aim to enhance revenue and competitiveness.

Findings and Report

The analysis reveals critical insights from descriptive statistics, correlations, and regression modeling. The mean rental price is $152.74, but the distribution is heavily skewed, with outliers driving the average above the median of $106. Correlations show weak relationships between price and variables such as availability and review metrics. Regression results indicate that location and room type are the most significant predictors, with listings in Manhattan and entire homes commanding the highest prices. Availability positively impacts price, while metrics like the number of reviews and minimum nights have smaller, negative effects. Visualizations such as pairplots and heatmaps further illustrate these relationships, emphasizing the importance of room type and location in determining price. The regression model explains 49% of the variance in price, underscoring the utility of these variables in pricing strategies.

Future Prediction Discussion

These findings provide a robust foundation for predicting nightly rental prices even before a property is listed. By incorporating variables like location, room type, and availability into predictive models, Airbnb can estimate competitive prices aligned with market dynamics. This enables hosts to benchmark their offerings and optimize for maximum revenue. Predictive tools leveraging these insights can enhance Airbnb’s competitive edge, ensuring responsive and accurate pricing strategies in a dynamic market. The integration of such data-driven approaches will enable Airbnb to cater to diverse customer demands while maintaining profitability.

**Deliverable 1**

Dependent and Independent Variables

The goal of this analysis is to explain or predict the price charged for an Airbnb rental in New York City. The dependent variable is the price of one night’s rental. The following are the independent variables identified as the most likely to explain the price:

Dependent Variable:

- Price: The price charged for one night’s rental.

Independent Variables:

- Neighbourhood Group: Broad location of the rental, e.g., Manhattan, Brooklyn.

- Neighbourhood: More specific location within the neighbourhood group.

- Room Type: Type of room being rented, e.g., entire home, private room, shared room.

- Minimum Nights: Minimum number of nights required for a booking.

- Reviews Per Month: Number of reviews received per month, reflecting the popularity of the rental.

- Availability 365: Number of days the rental is available per year, indicating flexibility and demand.

**Deliverable 2**



*Fig 1.1: Descriptive Statistics the Airbnb data set.*

Based on the descriptive statistics, several insights and unusual patterns emerge. The average price per night for an Airbnb rental is $152.74, though the potential presence of outliers suggests that some listings might be significantly more expensive, potentially skewing the mean. Manhattan dominates as the most frequent neighbourhood group, accounting for 21,643 listings, while Williamsburg is the most common specific neighbourhood with 3,917 listings, highlighting it as a hotspot for Airbnb activity. The most prevalent room type is "Entire home/apt," with 25,393 listings, indicating that most hosts offer private spaces for travelers.

The average minimum nights required is 7 nights, which is relatively high, potentially influenced by outliers enforcing long-term rental requirements. Similarly, reviews per month average 1.09, indicating moderate activity overall, though some properties may have zero reviews, possibly due to being newer or less popular. The mean annual availability for listings is 112.8 days, suggesting many hosts offer their spaces part-time or seasonally rather than year-round.

Unusual patterns include potential price outliers, listings with excessively high minimum night requirements, and properties with no reviews, which might reflect lower demand or recent additions. Additionally, the limited annual availability for many listings could significantly impact pricing trends, as properties available year-round typically face greater competition. Further exploration of these outliers and unusual patterns could provide deeper insights into pricing dynamics in the NYC Airbnb market.

A group of graphs with numbers

Description automatically generated with medium confidence

*Fig 1.2: Box plots of the selected numeric independent variables.*

**Deliverable 3**

A group of blue and white data

Description automatically generated

*Fig 2.1: Pairplot showing relationships between independent variables*

The pairplot visualization provides a comprehensive overview of the relationships between numerical variables and their individual distributions. The diagonal histograms reveal the skewness and spread of each variable. For instance, the price distribution is likely skewed, with most rentals priced lower and a few high-priced outliers. Similarly, variables like minimum\_nights and reviews\_per\_month may display right-skewed distributions, as most listings have short minimum night requirements or fewer reviews. The scatter plots on the off-diagonal highlight pairwise relationships. A positive relationship is expected between price and number\_of\_reviews or reviews\_per\_month, as more popular listings tend to command higher prices. However, variables like minimum\_nights may show weak or inconsistent relationships with price. The relationship between availability\_365 and price might indicate that highly available listings cater to broader demand and potentially higher pricing. This visualization serves as a foundational step in predicting how independent variables influence the price, highlighting key patterns and outliers that could inform more advanced modeling and predictions.

**Deliverable 4**

A screenshot of a graph

Description automatically generated

*Fig 3.1: Pivot chart showing the relationship between room type and average price.*

The first pivot table examines the relationship between room type and the average price. It reveals that entire homes or apartments command the highest average price at **$211.81**, reflecting their desirability for groups or families seeking privacy and autonomy during their stay. Private rooms are priced at an average of **$89.79**, offering a mid-range option for travelers on a budget. Shared rooms are the most economical, with an average price of **$70.08**, catering primarily to cost-conscious travelers. This analysis highlights that the type of room plays a significant role in determining its price, with privacy and exclusivity being premium factors.



*Fig 3.2: Relationship between number or reviews and price.*

Fig 3.2 table explores how neighborhood groups influence the average price and the number of reviews. Manhattan stands out with the highest average price of **$196.90**, underscoring its status as a prime location for tourists and business travelers. Brooklyn follows at **$124.41**, with a slightly lower but still substantial demand reflected in its high review count of **20,089 reviews**. The Bronx, with the lowest average price of **$87.47**, appeals to budget travelers and has the least reviews (**1,089 reviews**) among the three areas. This analysis indicates that neighborhood desirability significantly impacts both pricing and demand, with Manhattan leading due to its premium status and central location.

A screenshot of a computer

Description automatically generated

*Fig 3.3: Relationship between number of nights spent in the Airbnb and the price.*

The third pivot table highlights the relationship between neighborhood groups, average minimum nights, and average price. Manhattan once again emerges as a premium destination with the longest average minimum stay requirement of **8.54 nights**, aligning with its high average price of **$196.90**. Brooklyn, with a slightly shorter average stay of **6.06 nights** and an average price of **$124.41**, provides a more moderate option for visitors. The Bronx has the shortest average minimum stay requirement of **4.56 nights**, making it an accessible choice for budget-conscious or short-term travelers, reflected in its lowest average price of **$87.47**. This analysis suggests that areas with higher prices often correlate with longer minimum stay requirements, emphasizing their value as premium destinations.

A screenshot of a computer

Description automatically generated

*Fig 3.4: Relationship between number of reviews and price.*

The fourth pivot table focuses on room types, their average prices, and their respective review counts. Entire homes/apartments emerge as the most expensive option, averaging **$211.81**, and also garner the highest number of reviews (**25,393**), reflecting their popularity among travelers seeking exclusive accommodations. Private rooms, with an average price of **$89.79** and **22,306 reviews**, are also in high demand, offering a balance between cost and comfort. Shared rooms, while the cheapest at **$70.08**, have significantly fewer reviews (**1,159 reviews**), indicating limited demand. This analysis underscores the strong preference for entire homes/apartments, which combine privacy with higher pricing, while shared rooms, despite their affordability, cater to a smaller market segment.

A screenshot of a computer screen

Description automatically generated

*Fig 3.5: Correlation Matrix*

The correlation matrix provides insights into the relationships between numerical variables in the dataset. The price of a one-night rental (price) is expected to have a weak to moderate positive correlation with variables like number\_of\_reviews and reviews\_per\_month, as popular listings may charge higher prices. However, the relationship may not be strong due to variability in customer preferences and property characteristics. The correlation between price and minimum\_nights is likely weak or nonexistent, as the minimum stay requirement reflects host policies rather than directly influencing pricing. A moderate positive correlation might exist between price and availability\_365, as listings available year-round may cater to more customers and command higher prices. Additionally, a strong positive correlation is expected between number\_of\_reviews and reviews\_per\_month, as listings with more total reviews tend to have higher review frequencies. This analysis highlights which variables are more likely to influence pricing and sets the stage for more advanced modeling and prediction efforts.

A screenshot of a computer screen

Description automatically generated

*Fig 3.6: Regression model*

The OLS model explains 49% of the variance in log-transformed rental prices, indicating moderately strong explanatory power. Significant predictors include location (latitude, longitude, and neighborhood group), room type, availability, and minimum nights. Compared to the Bronx, listings in Manhattan are 26.34% more expensive, while Staten Island is 82.33% cheaper. Private and shared rooms are priced 74.86% and 116.97% lower, respectively, compared to entire homes/apartments. Continuous predictors like latitude and longitude show modest negative effects, reflecting the importance of precise location on pricing. While nearly all predictors are statistically significant, residual diagnostics highlight potential issues like non-normality and multicollinearity, suggesting opportunities for refinement. Despite these limitations, the model provides valuable insights into the pricing dynamics of short-term rentals, emphasizing the critical roles of location, room type, and availability in shaping prices.