**Business Report for AIR BNB**

**Submitted by: Vaitheeswaran .M**

**Course: Data Science and Analytics**

**Instructor: Dinesh**

**Date: 03/01/2025**

1. **Executive Summary**

The dataset contains 60,209 Airbnb listings, including room types, accommodations, pricing, cancellation policies, and review scores, with 11 missing columns, especially in review scores and property attributes.

1. **Introduction**

**. Problem Statement**: The problem statement highlights the various factors that influence the pricing and booking potential of Airbnb listings.

**.Objective**: The study will investigate room types, pricing patterns, review scores, and booking policies through data preparation, exploratory data analysis (EDA), and interpretation of results.

**.Dataset Description**:

\*Room Information Overview:

Includes room\_type, accommodates, bathrooms, bedrooms, beds

\* Log\_price column represents log-transformed price.

\* Booking & Policies:

cancellation\_policy, instant\_bookable, cleaning\_fee affect listing preference.

\* Review scores:

captures guest feedback.

\*Data Issues:

missing values in review\_scores\_rating, bathrooms, and beds.

1. **Data Preparation**

* Review Scores Rating Missing Values  
   • 13,601 missing values (22.6%)  
   • Possible imputement or removal based on relevance.
* "Minor Missing Values in Bathrooms, Bedrooms, Beds"  
  • Filling with median values.
* "Room Type, Cancellation Policy, Cleaning Fee Imputed"  
  • Few missing values.  
  • Mode can impute.
* Data Analysis:  
  • No duplicate values.

1. **Exploratory Data Analysis (EDA)**

Data Checking Process:  
 • Import data.

• Inspect top and last rows.

• Determine data dimension.

• Gather data information.

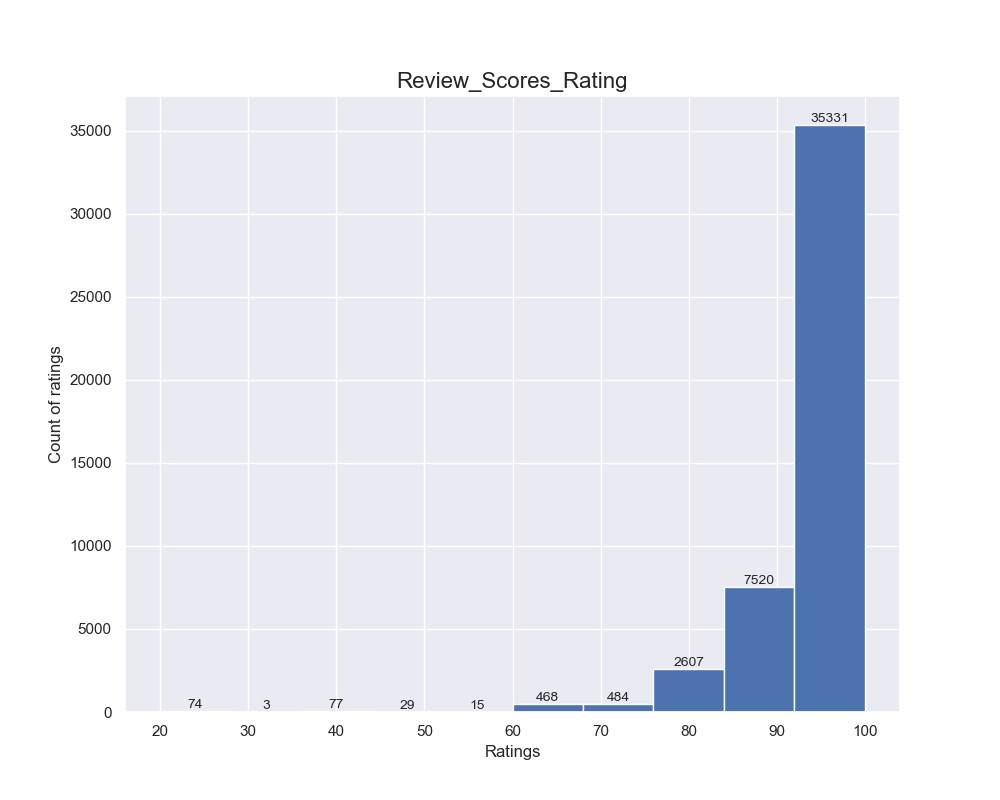
• Summarize data for statistical analysis.

• Understand data types.

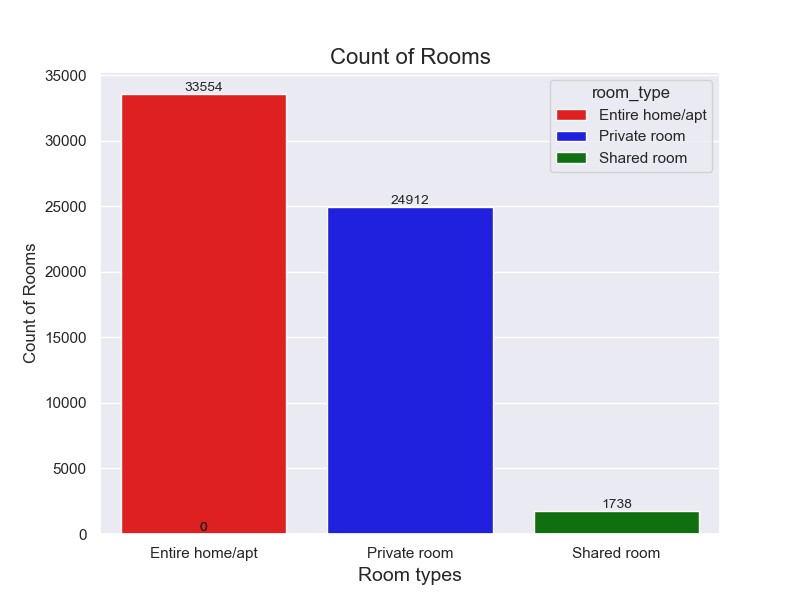
• Check data columns.

• Check missing values and duplicates.

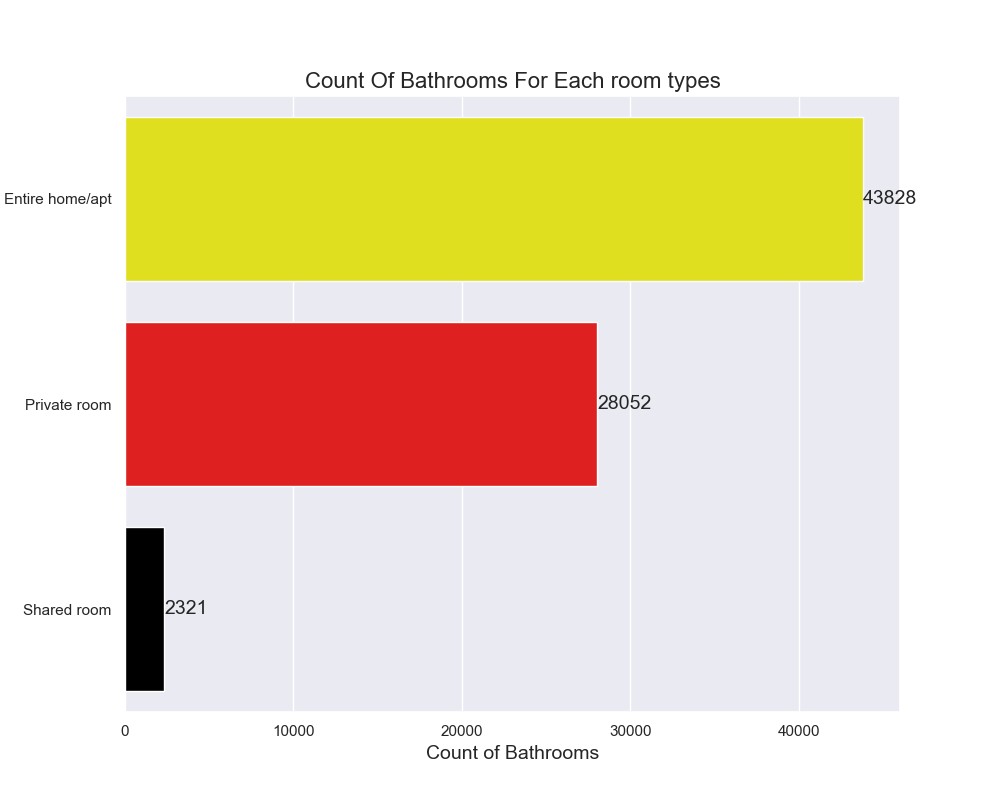
* **Univariate Analysis**: The distribution of Review\_scores\_Rating is analysed using Histogram
* Summary statistics such as mean, median, and standard deviation for Review\_scores\_rating were calculated, showing a mean of with a 94.08,standard deviation of 7.81.



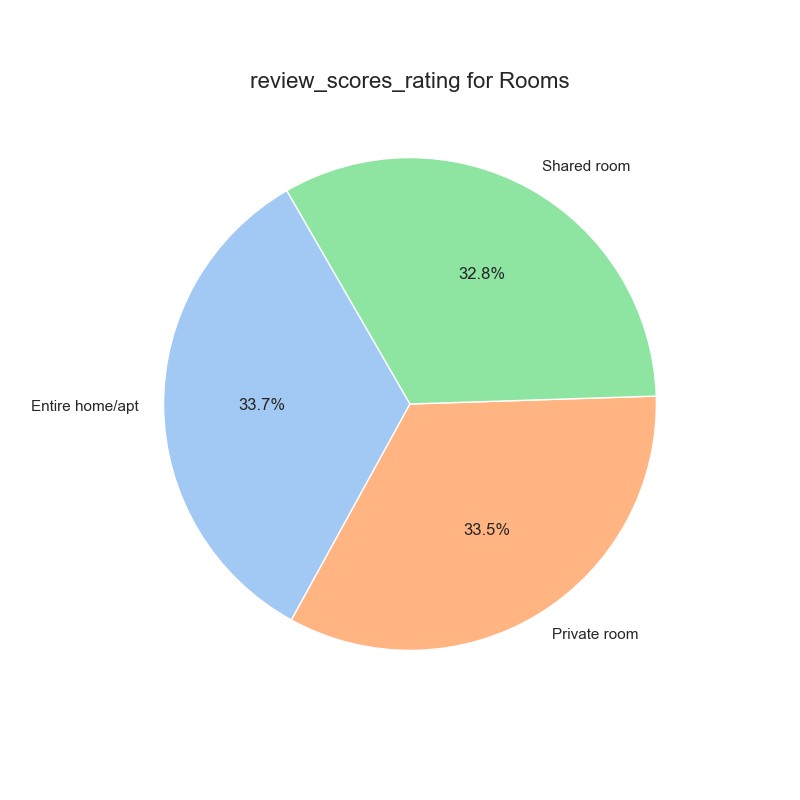
* The Univariant analysis of room\_types generates a Count plot to visualize the frequency distribution of Room\_types.
* Using This Analysis We can conclude That There are more no.of.Entire home/apt count is 33554 and less count of room is Shared room 1738.



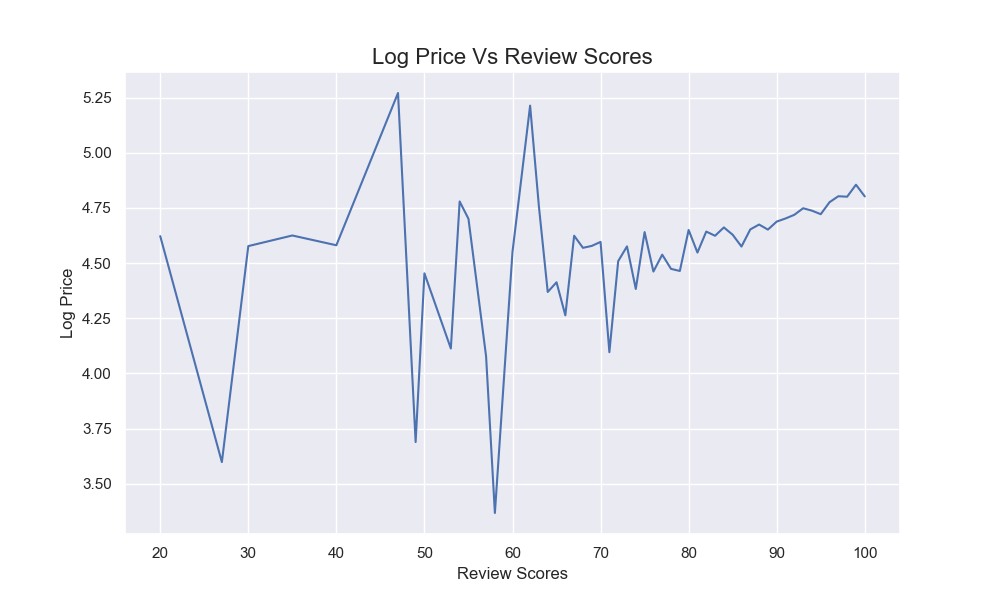
* The Univariant analysis of Bathrooms for Room\_types generates a bar plot to visualize the frequency distribution of Bathrooms for different Room\_types.
* Using This Analysis We can conclude That There are more no.of.bathrooms in Entire home/apt count is 43828 and less count of room is Shared room 2321.

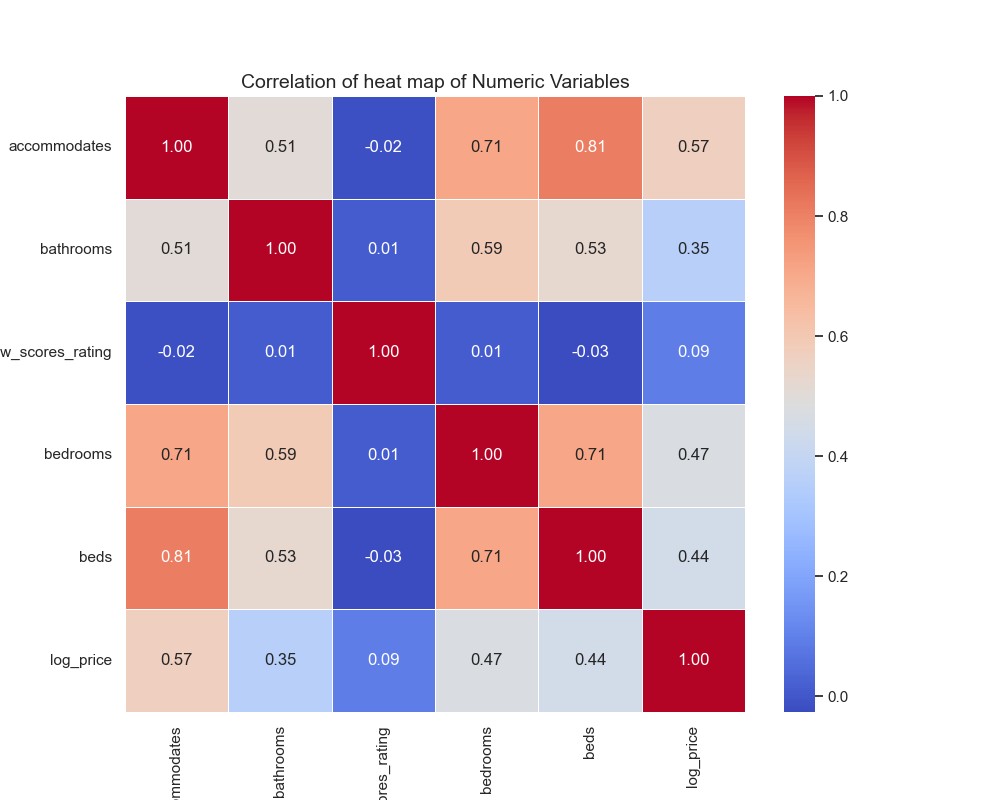


* **Bivariate/Multivariate Analysis**:
* This type of Analysis is used to analyse between a categorical and numerical value. So, I have analysed review\_score for different room types using matplot (Pie chart).
* The Reviews are similar to all the Rooms may be it differs in more than 0.1% and less than 0.9%.



* An analysis is done between Log\_price Vs Review Scores.There is Rise in log\_price when the review is after 70, Before that the Log\_price is up and down,but after 70 the log\_price is rising.





**Purpose of the Chart**

* This is a **correlation heatmap** that visualizes the relationships between different numerical variables in a dataset.

**Color Interpretation**

* **Red shades (closer to 1)** → Strong positive correlation (as one variable increases, the other also increases).
* **Blue shades (closer to -1)** → Strong negative correlation (as one variable increases, the other decreases).
* **Neutral colors (closer to 0)** → Weak or no correlation

**Key Observations from the Chart**

* **Beds & Accommodates (0.81):** Strong positive correlation, meaning more beds generally mean higher accommodations.
* **Bedrooms & Accommodates (0.71):** A good positive correlation, showing that more bedrooms lead to higher accommodates.
* **Bathrooms & Bedrooms (0.59):** Moderate positive correlation, implying that listings with more bedrooms often have more bathrooms.
* **Log Price & Accommodates (0.57):** Indicates that properties accommodating more people tend to have higher prices.
* **w\_scores\_rating & Other Variables (~0):** Very weak correlation, meaning review scores do not significantly affect the other numerical variables.
* **Key Insights:**

Higher accommodation capacity (accommodates, beds, bedrooms) is positively correlated with log\_price, indicating that listings with more space tend to be priced higher.

A strong correlation (**0.81 between beds and accommodates** & **0.71 between beds and bedrooms**) suggests that properties with more beds and bedrooms accommodate more guests.

**Bathrooms and log\_price correlation (0.35)** is weaker compared to other factors, indicating that the number of bathrooms has a **lesser**

**5. Results and Interpretation**

* **Key Findings**: **For higher revenue,** hosts should focus on increasing accommodation capacity (beds and bedrooms) rather than just adding more bathrooms.

**Customer ratings are not directly linked to property attributes**, so other factors like service quality, location, or cleanliness might be influencing reviews.

* **Visualizations**:

**Uni-varient Analysis**

Histogram – Review scores,

Countplot – Count of Rooms,

Barplot – Count of Bathroom

**Bi-Varient Analysis**

**Pie Chart – Review Scores for Room,**

**Line plot – Log price vs Review score,**

**Heat map – co-relation Between accommodates,bathrooms,reviewscoresrating,bedrooms,beds, log\_price.**

* **Recommendations**:

**More Beds & Bedrooms = Higher Price**

Listings with more beds and bedrooms have a strong positive correlation with price.

Increase guest capacity with extra beds or optimized room layouts.

**Bathrooms Have Less Impact on Price**

Bathrooms show weaker correlation with pricing compared to bedrooms.Focus on luxury amenities and cleanliness instead of adding more bathrooms.

**Ratings Aren’t Affected by Property Size**

Review scores have little correlation with attributes like beds or price. Improve guest experience through better service, cleanliness, and communication.

**6. Conclusion**

* **Summary of Analysis**:
* **Pricing is strongly influenced** by the number of **beds, bedrooms, and accommodates**, meaning larger properties tend to be priced higher.
* **Bathrooms have a weaker impact on pricing**, suggesting that additional bathrooms do not significantly drive up listing value.
* **Review scores do not correlate well** with numerical attributes, indicating that factors like service quality and cleanliness play a bigger role in guest satisfaction.
* **Limitations**:

The analysis only considers numerical variables; factors like location, amenities, and property type are not included.

External market trends (seasonality, demand, competitor pricing) are not factored into the correlations.

Causation cannot be established—correlation only shows relationships, not direct influence.

* **Next Steps**:

**Expand analysis** by incorporating **location, property type, and customer reviews** for a more holistic view.

**Perform sentiment analysis** on guest reviews to understand nonnumerical factors affecting ratings.