**Business Report for AIR BNB**

**Submitted by: Veath Prakash B**

**Course: Data Science and Analytics**

**Instructor: Dinesh Kumar**

**Date: 03/01/2025**

---

## 1. Executive Summary

This dataset contains 60,209 Airbnb listings with information about room types, accommodations, pricing, cancellation policies, and review scores. It includes 11 columns, some of which have missing values, particularly in review scores and property attributes like bathrooms, bedrooms, and beds.

---

## 2. Introduction

- **Problem Statement**: Airbnb listings contain various factors influencing their pricing and booking potential.
- **Objective**: This analysis aims to explore key insights into **room types, pricing patterns, review scores, and booking policies**. The study will involve **data preparation, exploratory data analysis (EDA), and interpretation of results**.
- **Dataset Description**:
- Room Information: Includes room_type, accommodates, bathrooms, bedrooms, and beds.
- Pricing: The log_price column represents the log-transformed price.
- Booking & Policies: cancellation_policy, instant_bookable, and cleaning_fee affect listing preferences.
- Reviews: review_scores_rating captures guest feedback.
- Data Issues: Some missing values in review_scores_rating, bathrooms, and beds need handling.

## 3. Data Preparation

Handling Missing Values:

- **Review_scores_rating** has 13,601 missing values (22.6%). These might be imputed or removed based on relevance.
- **bathrooms, bedrooms, and beds** have minor missing values, which can be filled using median values.
- **room_type, cancellation_policy, and cleaning_fee** have very few missing values and can be imputed using the mode.
- **In my Data There is No duplicate value**

---

## 4. Exploratory Data Analysis (EDA)

I have Checked My data Completely by Applying these steps :

* Data import

* Data inspection :- check top 5 rows,last 5 rows

* Data dimenson

* Data information

* Data summarization (statistical)

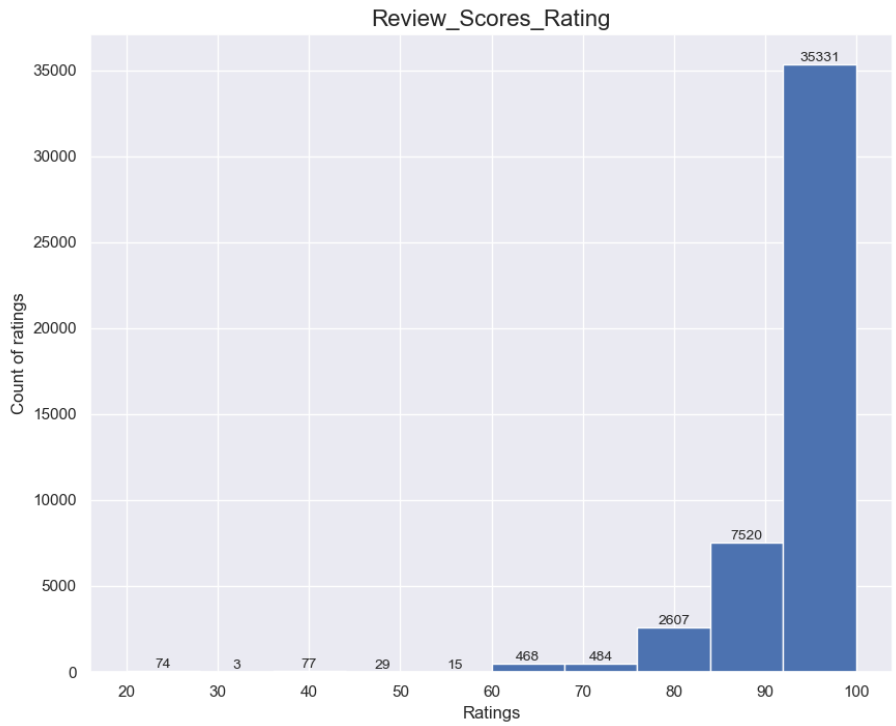
* Data types

* Data columns

* Check missing values

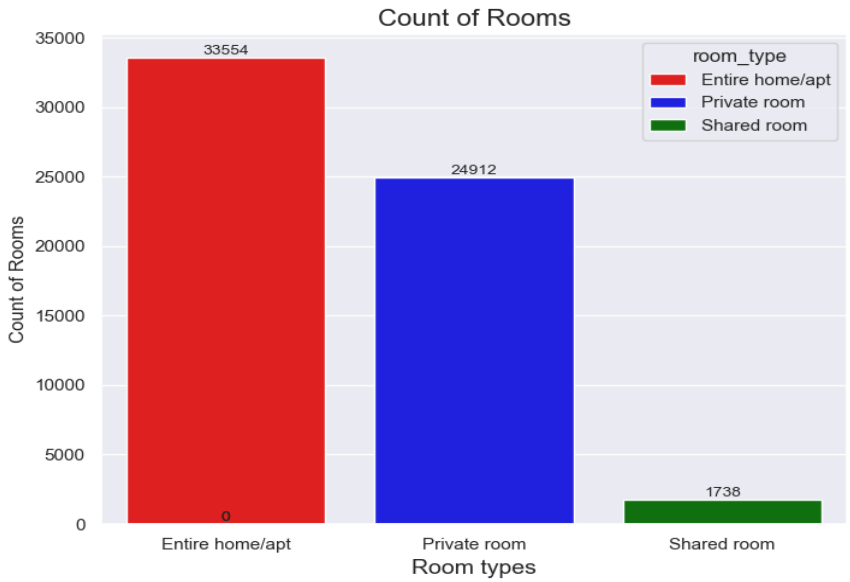
* Check duplicates

- **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.

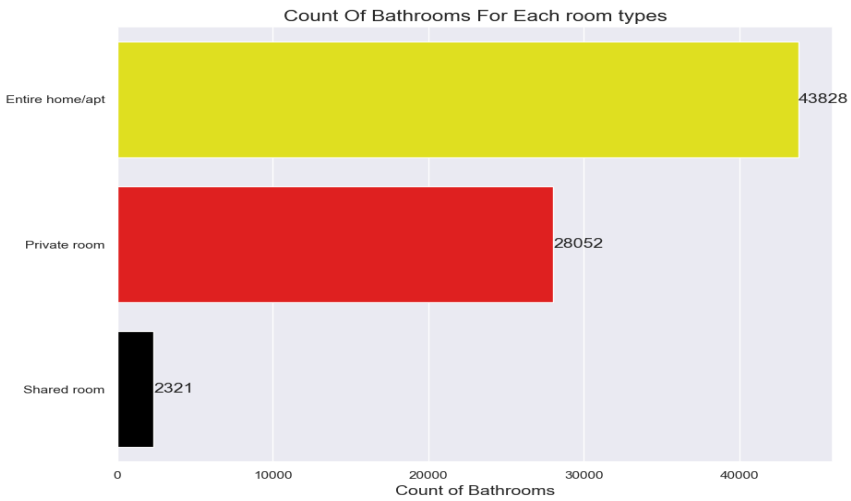



- The Review Score Starts From 20 itself, But Maximum Ratings is between 90- 100

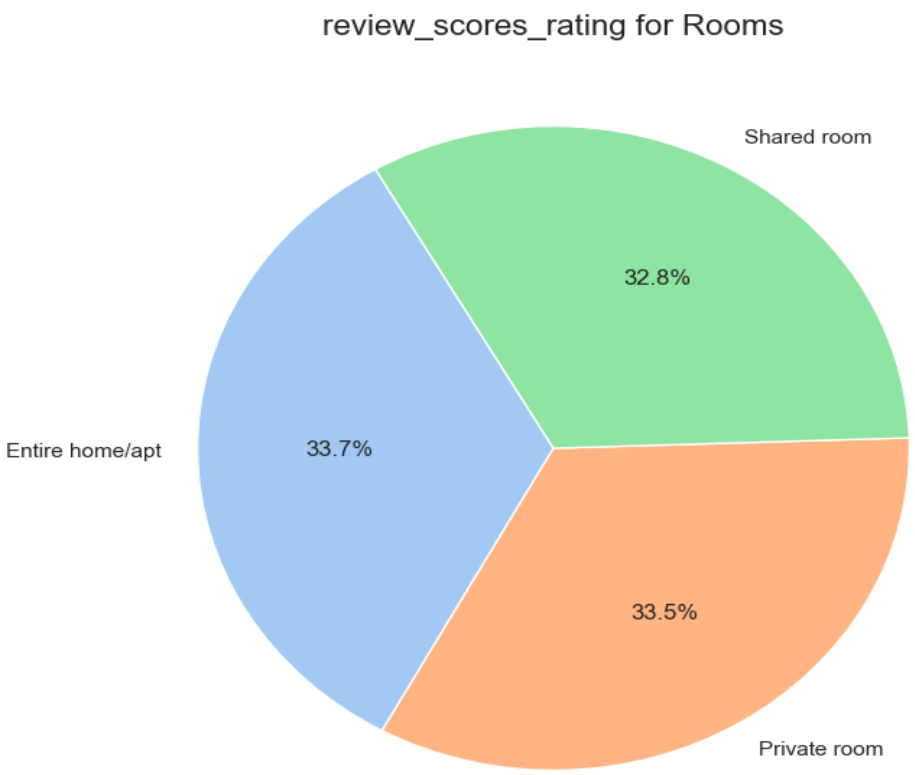
- The Univariant analysis of room_types generates a Count plot to visualize the frequency distribution of Room_types.
- Using This Analysis We can conclude That There are more no.of.Entire home/apt count is 33554 and less count of room is Shared room 1738.



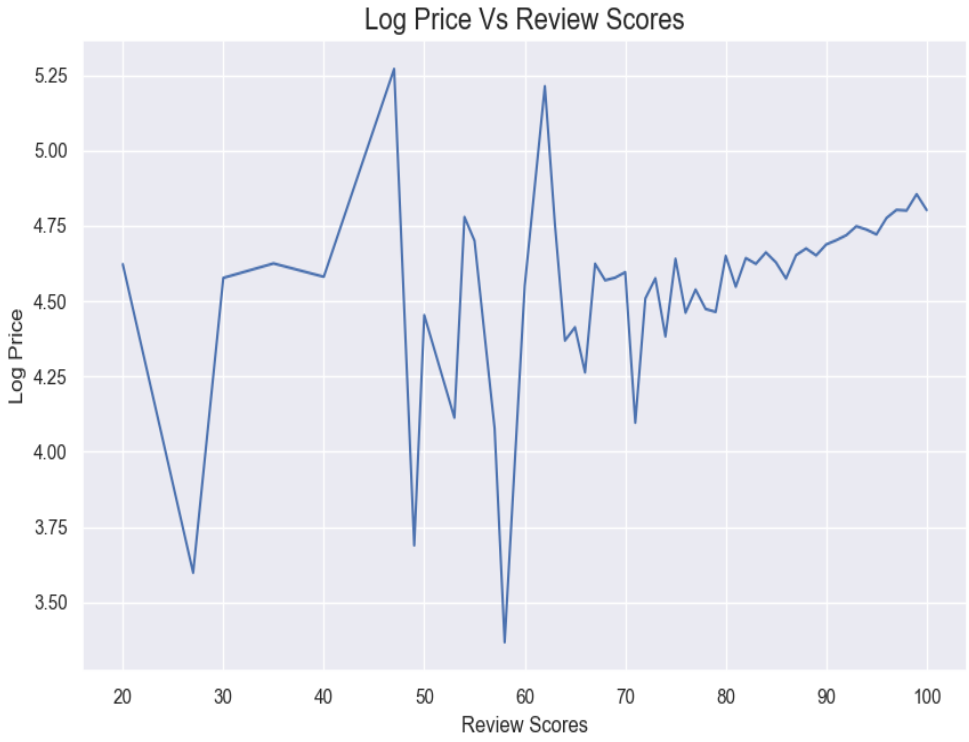

- The Univariant analysis of Bathrooms for Room_types generates a bar plot to visualize the frequency distribution of Bathrooms for different Room_types.
- Using This Analysis We can conclude That There are more no.of.bathrooms in Entire home/apt count is 43828 and less count of room is Shared room 2321.

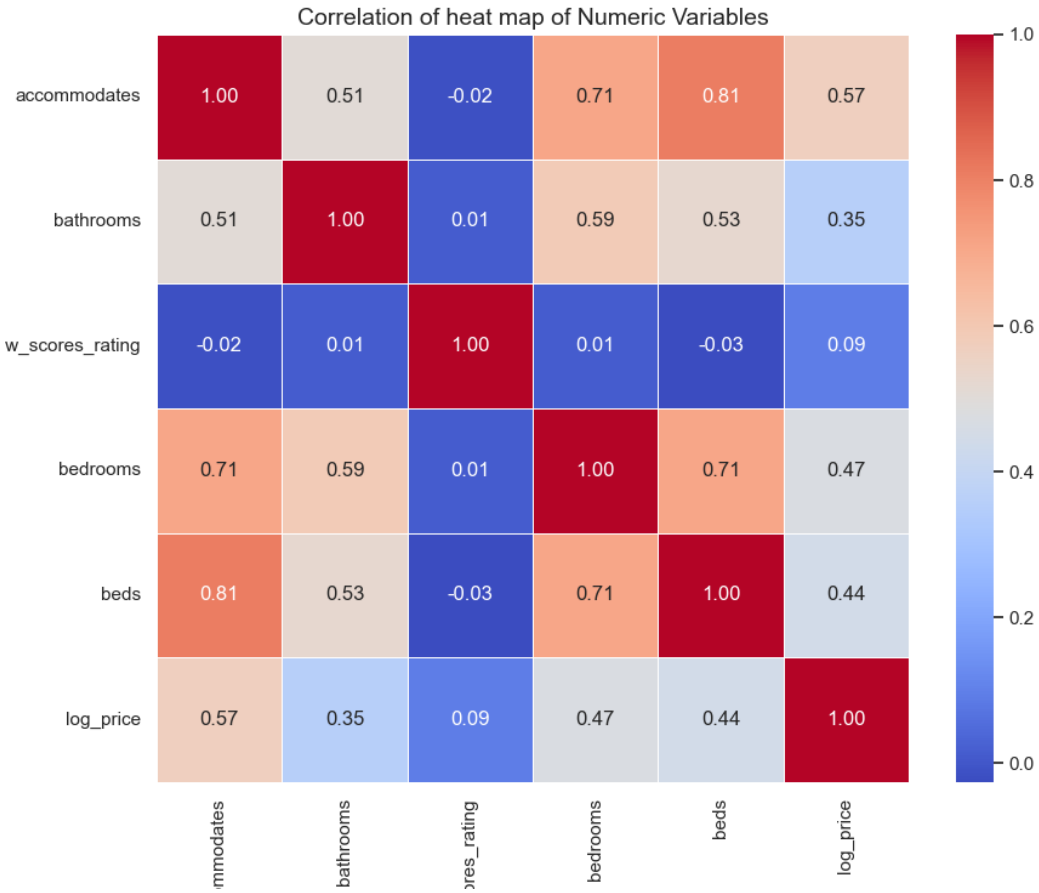


- **Bivariate/Multivariate Analysis**:
- This type of Analysis is used to analyse between a categorical and numerical value. So, I have analysed review_score for different room types using matplot (Pie chart).
- The Reviews are similar to all the Rooms may be it differs in  more than 0.1% and less than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

---