

# **Business Report for AIR BNB**

**Submitted by: Vaitheeswaran .M**

**Course: Data Science and Analytics**

**Instructor: Dinesh**

**Date: 10/02/2025**

---

## **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.

---

## **2. 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.

## **3. Data Preparation**

- Review Scores Rating Missing Values
  - 13,601 missing values (22.6%)
  - Possible imputation 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.

---

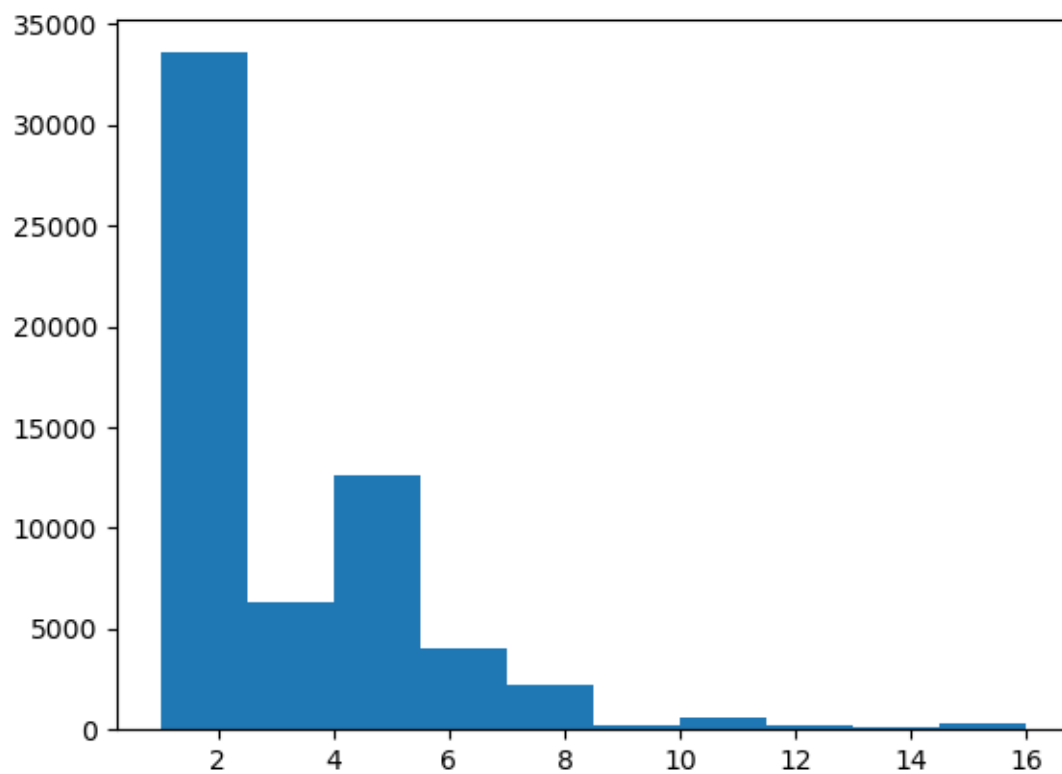
## 4. Exploratory Data Analysis (EDA)

Data Checking Process:

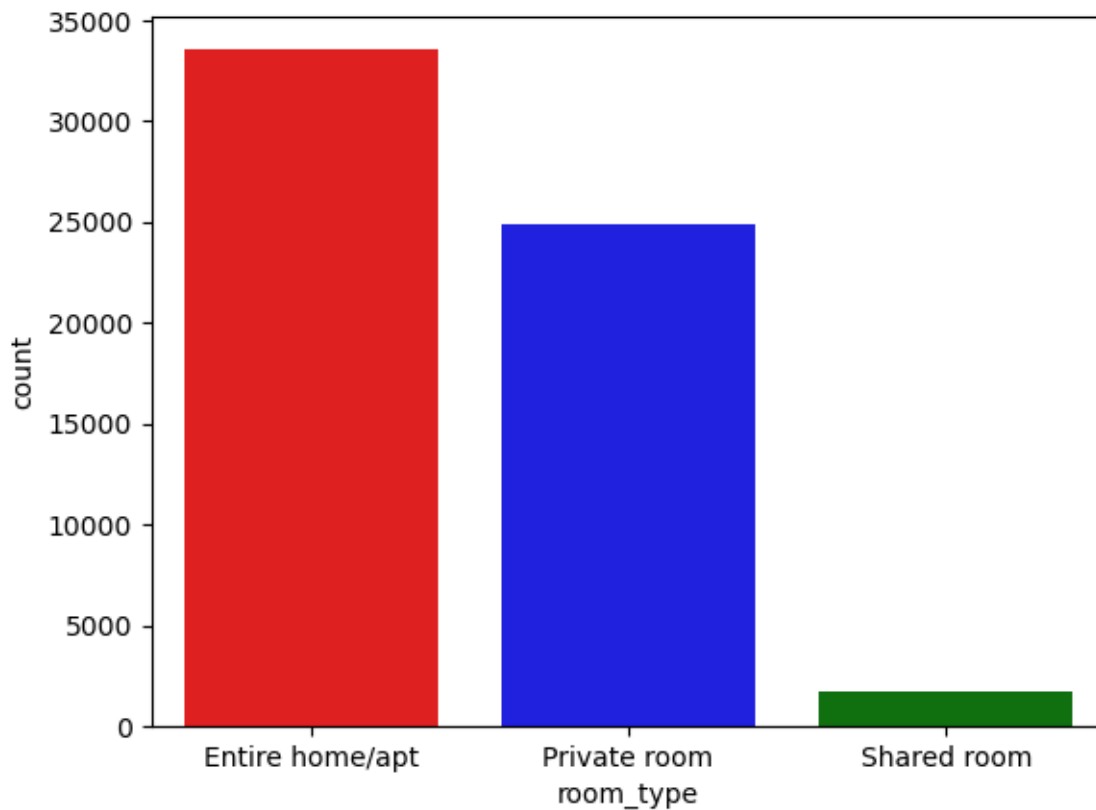
- Import data.
- Inspect top and last rows.
- Determine data dimension.
- Gather data information.
- Summarize data for statistical analysis.
- Understand data types.
- Check data columns.
- Check missing values and duplicates.

## Univariate Analysis:

- The Histogram is used to analyze the distribution of accommodates.
- The data shows a skew distribution, with a dominant preference for lower values, a long tail showing a minority preference, and a sharp drop-off indicating a potential imbalance.



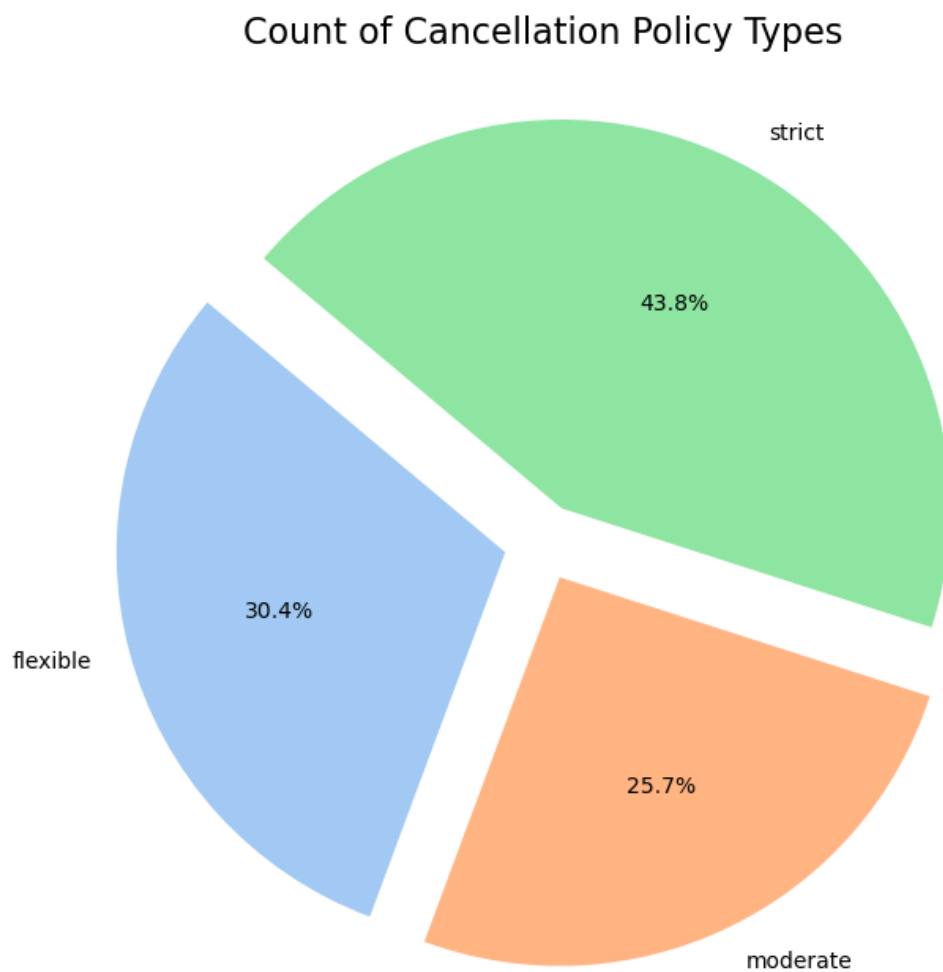
- The most preferred room type is "Entire home/apt," followed by "Private room" as the second most popular option. The least popular is "Shared room," with a significant gap between its counts and other room types, indicating limited appeal.



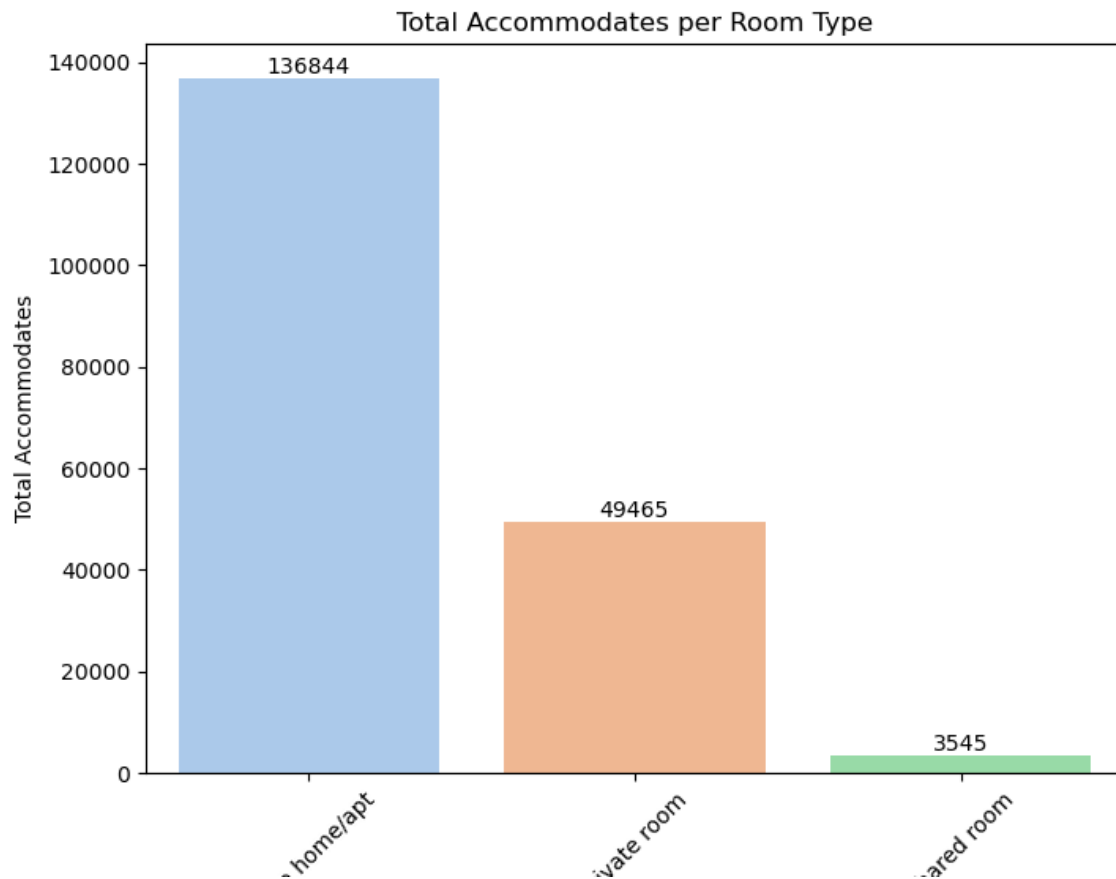
### Policy Distribution Analysis

- **Strict Policies:** Majority of cancellations fall under strict policies, indicating preference for definitive terms.
- **Flexible Policies:** Majority of cancellations are handled under flexible policies, indicating value for flexibility.
- **Moderate Policies:** Small yet significant portion of cancellations, indicating balanced approach.
- **Strategic Insights:** Analysis can optimize cancellation policies to meet user demands and maintain operational efficiency.

- **Bivariate/Multivariate Analysis:**



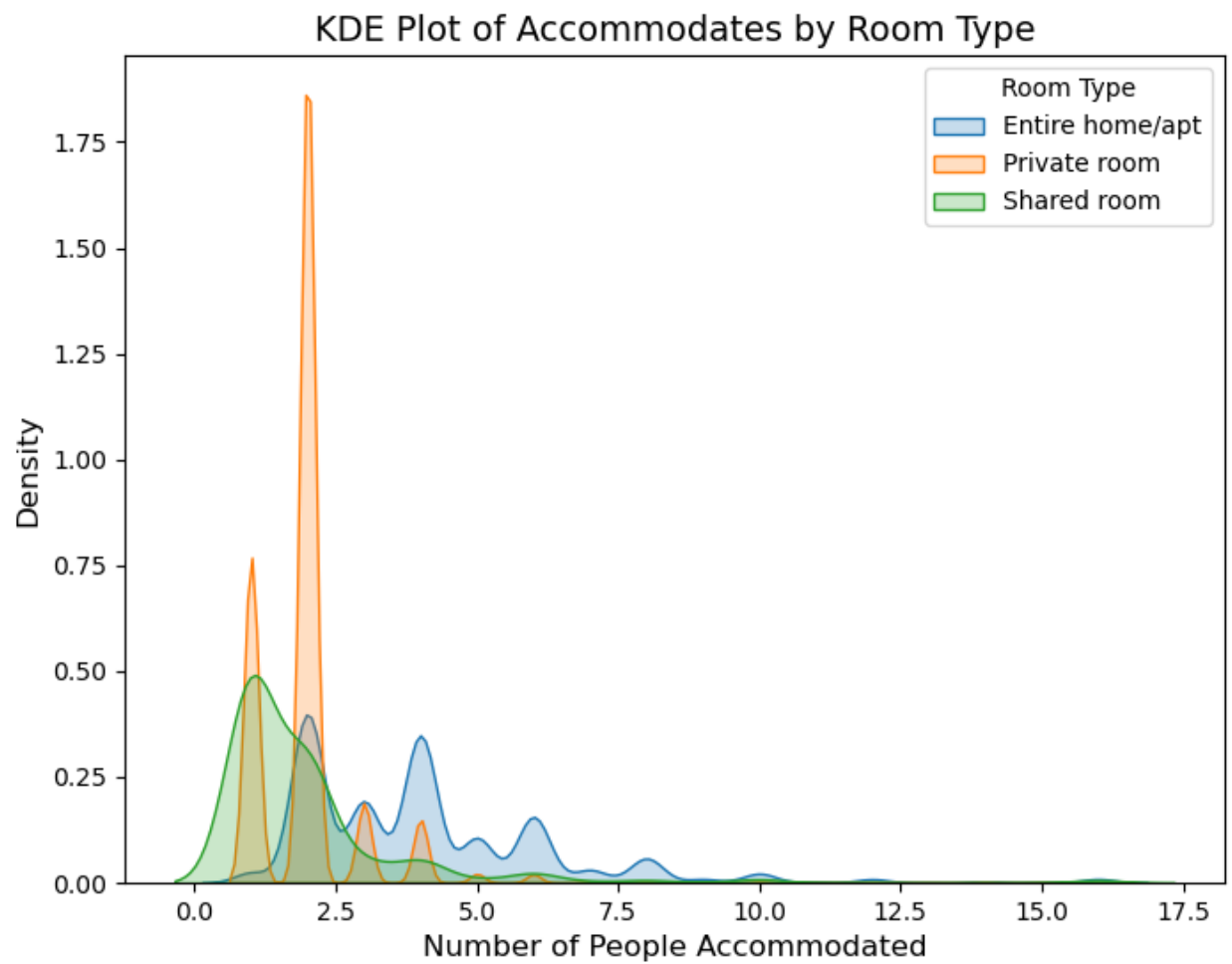
- The "Entire home/apt" category has the highest number of accommodations, with 136,844 units. Private rooms are less popular than entire homes or apartments, with 49,465.
- Shared rooms have the least popularity, with 3,545 accommodations. The data indicates a preference hierarchy, with entire homes/apts yielding the most opportunities for hosts or renters.

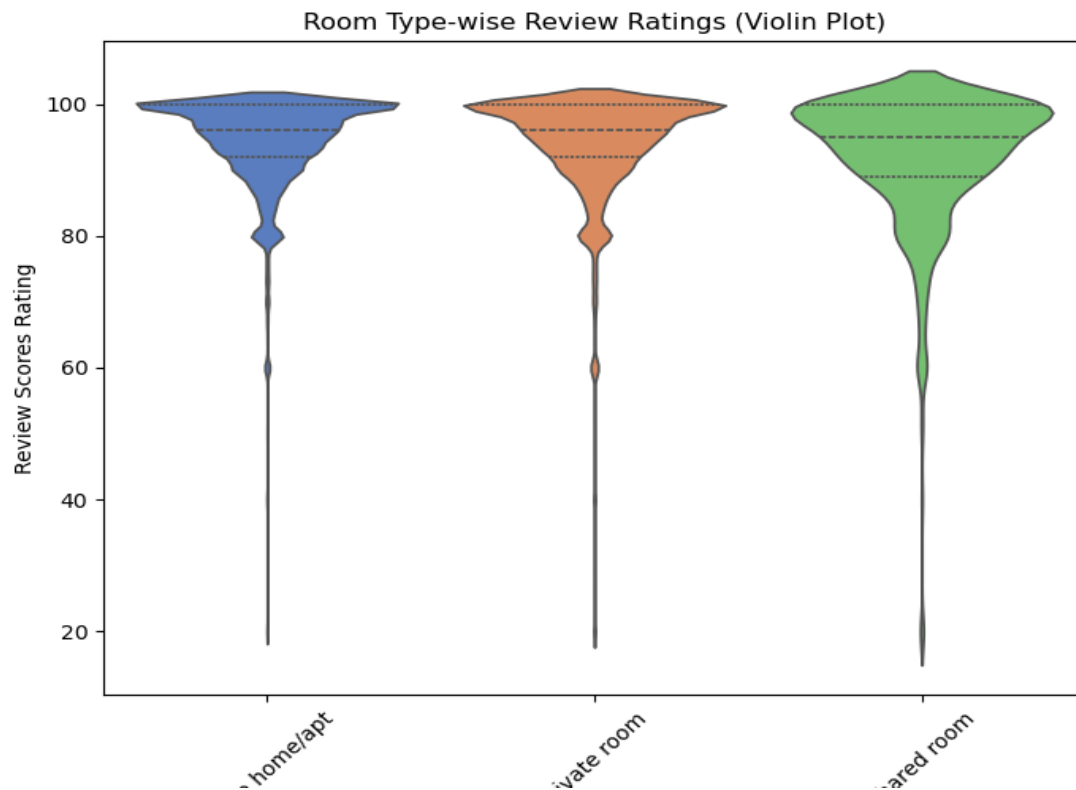


### Peak Density and Distribution of Rooms

- Private rooms have the highest density peak around 2 people, indicating they are designed for small groups or couples.
- Entire homes/apts have a broader distribution, accommodating a wider range of group sizes.
- Shared rooms have a sharp peak for accommodating 1-2 people, primarily targeting individuals or small groups.

- Entire homes/apts are more versatile in terms of group size.
- For larger groups, focus on entire homes/apts. For individual travelers or couples, private and shared rooms are better options.





### Purpose of the Chart

This violin plot visualizes the distribution of review score ratings for different room types, providing insights into the spread, density, and central tendency of ratings.

### Key Observations

- **High Median Ratings Across All Room Types:**

The median ratings (dashed lines) for all room types are close to 100, indicating high satisfaction levels across the board.

- **Entire Home/Apt (Blue):**

Consistent ratings with a narrow spread, suggesting high and stable satisfaction among guests. Few outliers exist at lower ratings.



- **Private Room (Orange):**

Slightly wider distribution compared to entire homes, showing more variability in ratings. However, the majority are clustered near the top, indicating generally positive feedback.

- **Shared Room (Green):**

The widest spread among the three categories, with more occurrences of lower ratings. This suggests that shared rooms may have more variability in guest experiences.

- **Key Insights:**

- Focus on Entire Home/Apt: High guest satisfaction, stable ratings suggest consistent positive outcomes.
- Private Rooms: High ratings, room for improvement through better amenities or service.
- Shared Rooms: High variability, lower ratings, need standardization and guest concerns addressed.

---

## 5. Results and Interpretation

### Key Findings:

- Hosts should increase accommodation capacity for higher revenue.
- Customer ratings aren't directly linked to property attributes.
- Service quality, location, cleanliness may influence reviews.

- **Visualizations:**

### Uni-varient Analysis

Histogram – accommodates,

Countplot – Count of Room types

### **Bi-Varient Analysis**

Pie-chart- Count of cancellation policies

Bar plot- Total Accommodates per Room Type,

KDE plot – Accommodates by Room Type,

violinplot - Room Type-wise Review Ratings

### **Property Pricing Recommendations**

- More Beds & Bedrooms: Strong positive correlation with higher prices.
- Less Impact of Bathrooms: Focus on luxury amenities and cleanliness.
- Ratings: Little correlation with property size.
- Improvements: Enhance guest experience through better service, cleanliness, and communication.

---

## **6. Conclusion**

- ❖ Summary of Analysis:
- ❖ Data primarily concentrated around 2 values.
- ❖ High frequency of low values and low frequency of high values.
- ❖ Relevant in economics, where income distribution often shows high-income individuals.
- ❖ "Entire Home/Apt" category is most prevalent, followed by "Private Room," and "Shared Room."
- ❖ Most prevalent cancellation policy type is "strict," accounting for nearly half of the total (43.8%).

- ❖ "Flexible" policy type represents about a third (30.4%), and "moderate" policy type is the least common (25.7%).
  - ❖ "Entire Home/Apt" category has the highest number of accommodations (136,844), "Private Room" with 49,465 accommodations, and "Shared Room" with 3,545.
  - ❖ Private rooms are most commonly designed to accommodate 2 people, while entire homes or apartments offer more flexibility.
  - ❖ All three room types receive high review scores, with slight variations reflecting differences in guest satisfaction levels.
-