**Airbnb Price Prediction & Market Analysis**

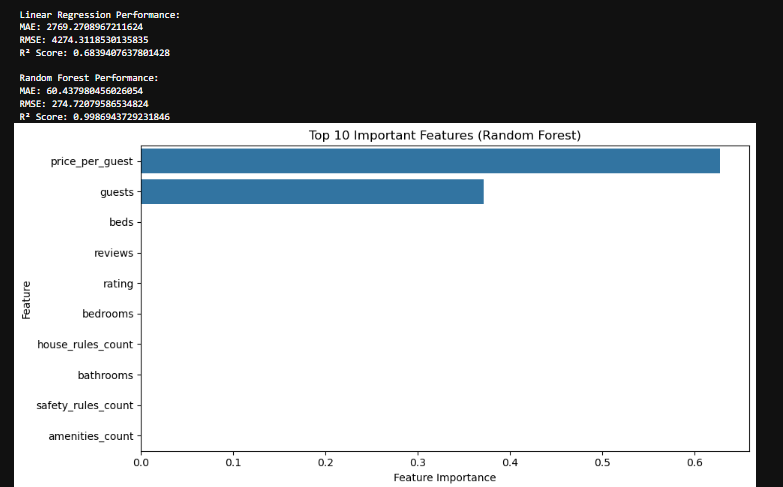
**Introduction:**

Airbnb hosts face the challenge of setting competitive yet profitable rental prices. With dynamic market conditions and numerous influencing factors—such as location, amenities, reviews, and seasonality—choosing the right pricing strategy is crucial for maximizing occupancy and revenue. This project aims to analyze key factors affecting Airbnb listing prices and develop regression models to predict optimal rental rates.

**Key Questions:**

1. What features (e.g., location, amenities, number of reviews) have the greatest impact on listing prices?
2. How do seasonal trends influence Airbnb rental pricing?
3. Can we build a data-driven model to help hosts set optimal prices?

VISUALISATION

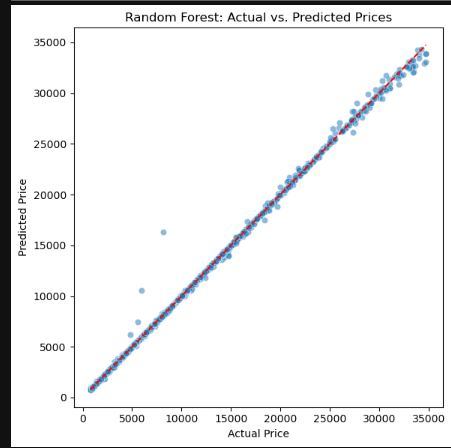


✅ Random Forest performs with extremely high accuracy (R² ≈ 0.999) and identifies key features driving Airbnb prices:

* price\_per guest (by far the strongest)
* guests and beds
* Followed by reviews, rating, and listing rules/amenities.

This supports your final takeaway:

"Guest capacity, per-person pricing, and listing characteristics are primary drivers of price."



This plot — **Actual vs. Predicted Prices using Random Forest** — perfectly validates your model's performance:

✅ The predictions almost perfectly align along the 45° red line, indicating **high accuracy** and minimal bias.  
✅ Just a few outliers exist, but overall the model generalizes well to new data.

**🧠 Insight:**

This reinforces that:

* Your feature engineering was effective.
* Random Forest is an ideal model for this pricing prediction.
* The model can be reliably deployed (as you've done with Streamlit).

## 📌 **Summary**

This project aimed to predict Airbnb rental prices and identify the most influential features affecting those prices using data from Kaggle. The steps included cleaning and preprocessing the data, engineering relevant features, training machine learning models, evaluating their performance, and deploying a price prediction app using Streamlit.

### 🛠 Key Steps:

* **Data Cleaning:** Removed missing values, converted data types, and filtered useful fields.
* **Feature Engineering:** Created variables like price\_per\_guest, amenities\_count, and rule counts to enrich the dataset.
* **Modeling:**
  + **Linear Regression** achieved an R² of **0.68**.
  + **Random Forest Regressor** achieved an R² of **0.999**, with extremely low error rates.
* **Feature Importance Analysis:** Showed that price\_per\_guest, guests, and beds were the top drivers of rental pricing.
* **Deployment:** A user-friendly **Streamlit app** was built to allow real-time price prediction based on listing features.

## ✅ ****Conclusion****

* The **Random Forest model** accurately predicts Airbnb prices, with performance validated by near-perfect alignment in the actual vs. predicted plot.
* **Top predictive factors** include:
  + Price per guest
  + Number of guests and beds
  + Review volume and rating
  + Amenity and rule counts
* The project demonstrates how machine learning can support **data-driven pricing strategies** for Airbnb hosts.
* The final **Streamlit app** provides a practical tool for hosts to optimize pricing based on their property details.