Airbnb Price Predictions: Linear Regression Machine Learning Model

Executive Summary: A supervised machine-learning model to predict log-transformed Airbnb prices. Here's a detailed summary of the contents, including Exploratory Data Analysis (EDA), and the machine learning model used:

1. Introduction

- The notebook's main goal is to develop a model to predict Airbnb prices by training a supervised machine learning model. The price data is transformed using logarithms (log_price) to normalize the distribution of prices.

2. Exploratory Data Analysis (EDA)

- **Correlation Heatmap**: A heatmap is plotted to visualize the correlation between different features. Features highly correlated with the target variable (log_price) are explored in more detail.
- **Feature Distributions**: Various feature distributions such as Price, Error, and Predicted_Price are visualized to understand their spread and characteristics.
- **Outlier Detection**: Scatter plots are used to detect outliers and errors in the predicted vs. actual price.
- **Missing Data**: Missing data handling is performed, through specific imputation techniques based on central tendency.
- **Categorical Encoding**: Categorical variables are handled, through the exact method (e.g., one-hot encoding, label encoding).

3. Feature Engineering

- The notebook seems to involve feature transformations and engineering, particularly with log_price transformation. The extracted sections also indicate the importance of feature scaling and normalization.

4. Machine Learning Model

- **Train-Test Split**: The data is split into training and test sets.
- **Model Used**: Based on the imported libraries and code cells, the model seems to be a **Linear Regression** model or similar, given the focus on regression tasks.
- **Evaluation Metrics**:
    - Error vs. Predicted price scatter plot is used to evaluate the model's performance.
    - Evaluation metrics like **RMSE** (Root Mean Squared Error) and **MAE** (Mean Absolute Error) may be used, though not explicitly visible in the extracted section.

5. Visualization

- Several visualizations are used throughout the notebook:
    - Correlation heatmaps for feature relationships.
    - Scatter plots to visualize predictions, error distributions, and comparisons between predicted and actual prices.

6. Markdown Sections

- Some markdown sections likely contain theoretical explanations or background, but they were not fully extracted from the file.

Conclusion

The notebook demonstrates a systematic approach to predicting Airbnb prices using a supervised machine learning model. It employs basic EDA techniques, feature engineering, and regression models (likely Linear Regression). The notebook includes several visualizations to understand the data, evaluate model performance, and explore prediction errors.