AIRBNB PRICE PREDICTION.

OVERVIEW

Subject Area : **What triggers the Airbnb rental price?**

- Number of beds?
- Review score?
- Number of guests allowed?
- Cancellation policy?

Goal : Analyse the factors that influence the rental price and develop a model that captures the complexity of the pricing system.

Opportunity : Provides interesting insights that can benefit a host looking to maximize their profit.

PROPOSED VISION FOR TACKLING THE PROBLEM.

- **Feature selection** : Identify the key features that are likely to influence.
- **Scaling** : If necessary, scale numerical data to ensure they are on a similar scale.
- **Model training** : Implement linear regression on the preprocessed dataset.
- **Model evaluation** : Analyze the coefficients of the model to understand the impact of each features on rental price.

IMPACT OF THE PROPOSED SOLUTION.

Potential impact of a successful solution includes:

- Hosts can optimize **pricing strategies**.
- Guests can make **better decisions** that fits their interests and financial limitations.
- Transparent Airbnb marketplace.

DATASET AND PREPROCESSING.

Dataset:

- Comprises information on Airbnb listings in USA with **74111 rows** and **29 columns**

Preprocessing:

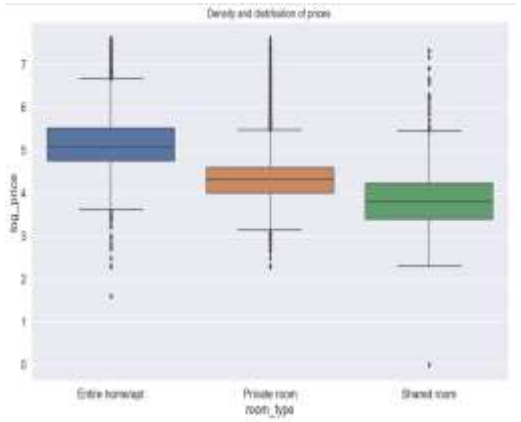
- **Data cleaning:** Almost clean !
- **Scaling:** Scale numerical features to similar range.
- **Encoding:** Label encoding.
- **Feature selection** : Identify the key features that are likely to influence.

IMPORTANT FINDINGS IN EDA.

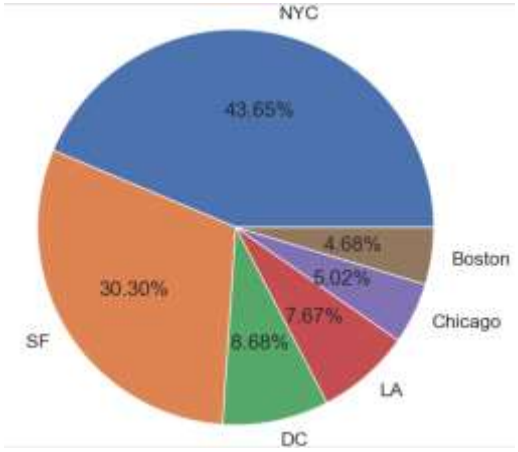
- **Data Distribution**: Normal distribution.
- **Data Quality:** Missing values and outliers were handled appropriately.
- **Patterns and trends:** To understand underlying relationships and phenomena of data through visualization.
- **Correlation**: Analyze multicollinearity with heatmaps
- **Feature importance**: Determining which features are most relevant for the target variable through visual inspection.

SOME VISUALS

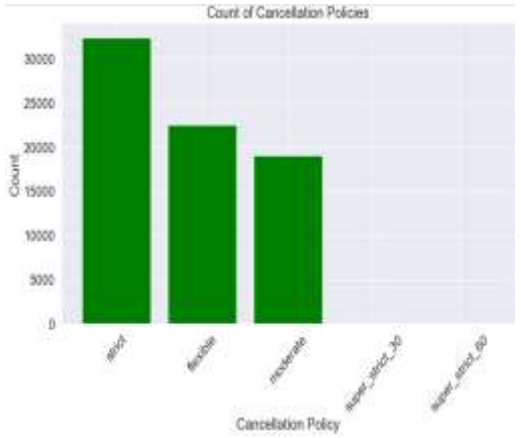
BOXPLOTS



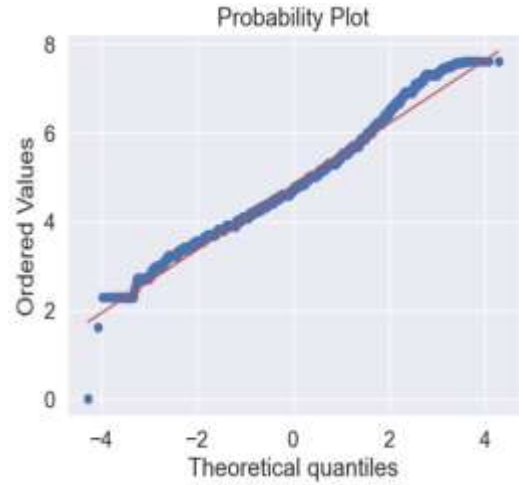
PIE CHARTS



BAR GRAPHS



DISTRIBUTION



MODELS USED

- Linear Regression
- Decision Tree
- Random Forest Regressor
- Neural Networks

BASELINE MODEL AND EVALUATION METRICS.

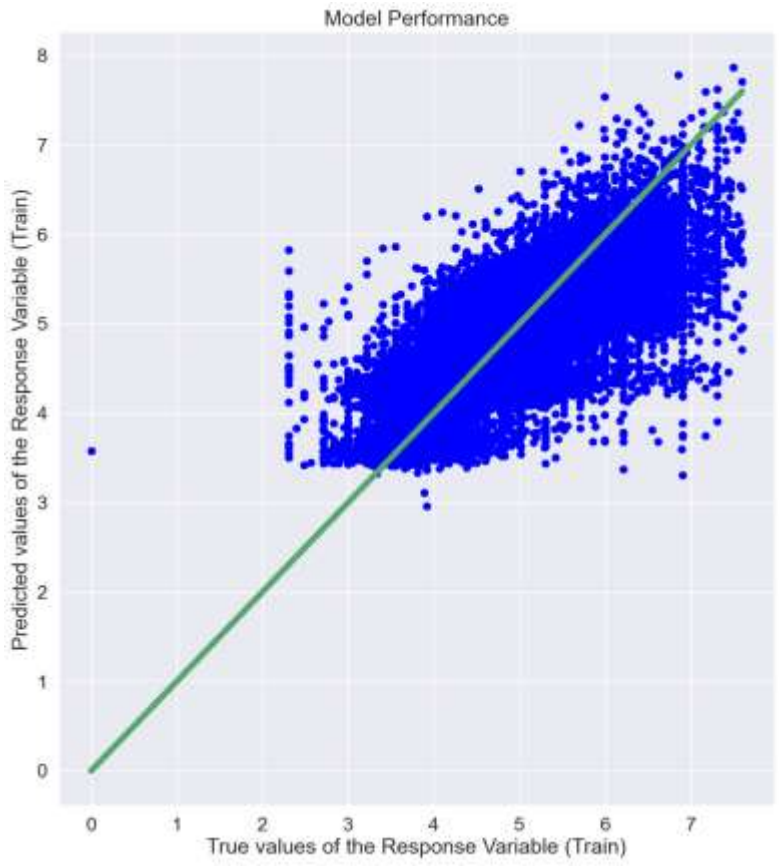
Baseline Model : Linear Regression

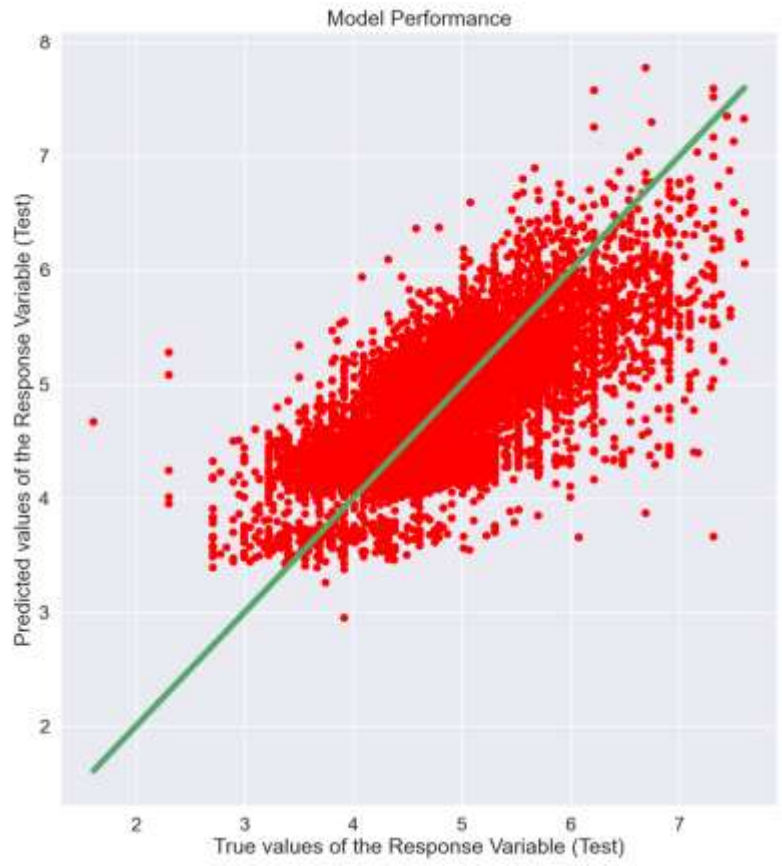
Mean Absolute Error (MAE): 0.36

Mean Squared Error (MSE): 0.23

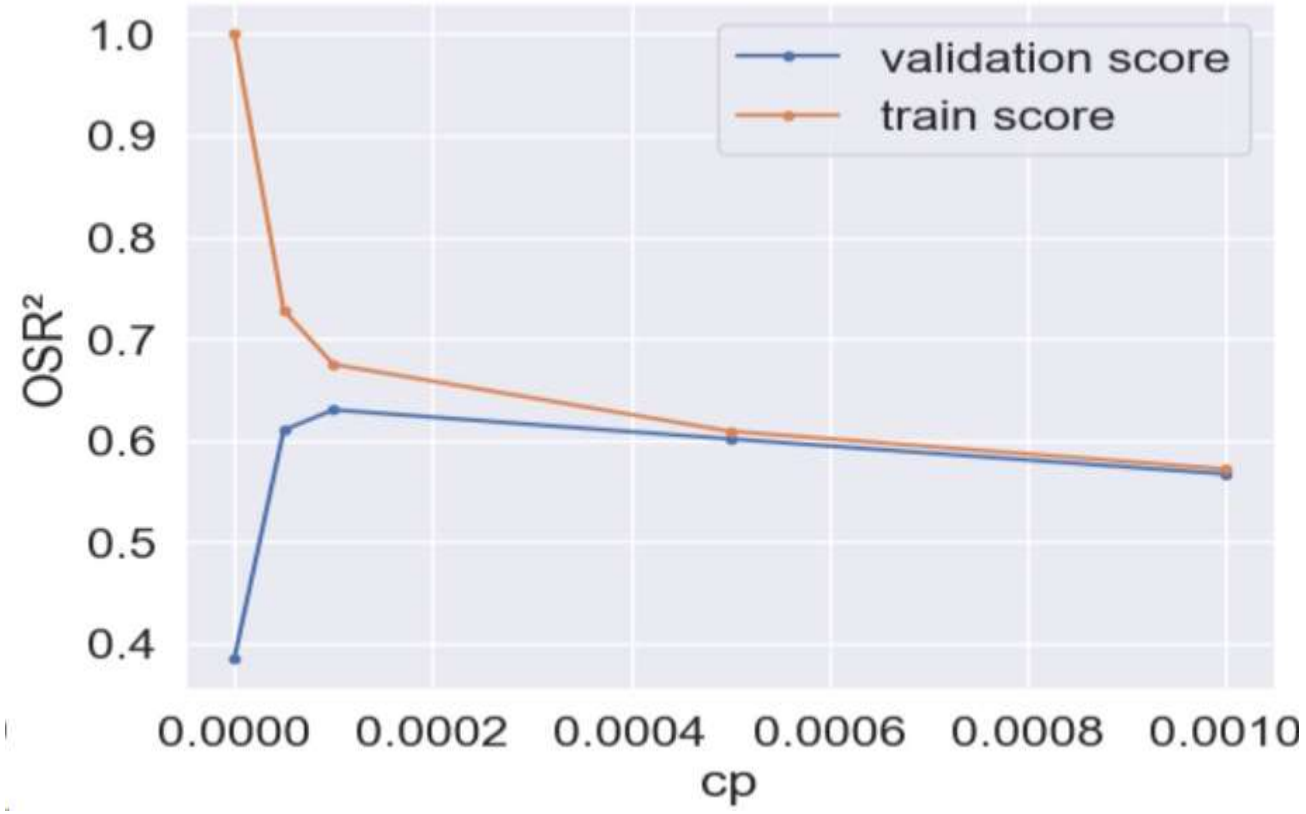
Root Mean Squared Error (RMSE): 0.48

R-squared (R2) Score: 0.54





DECISION TREE

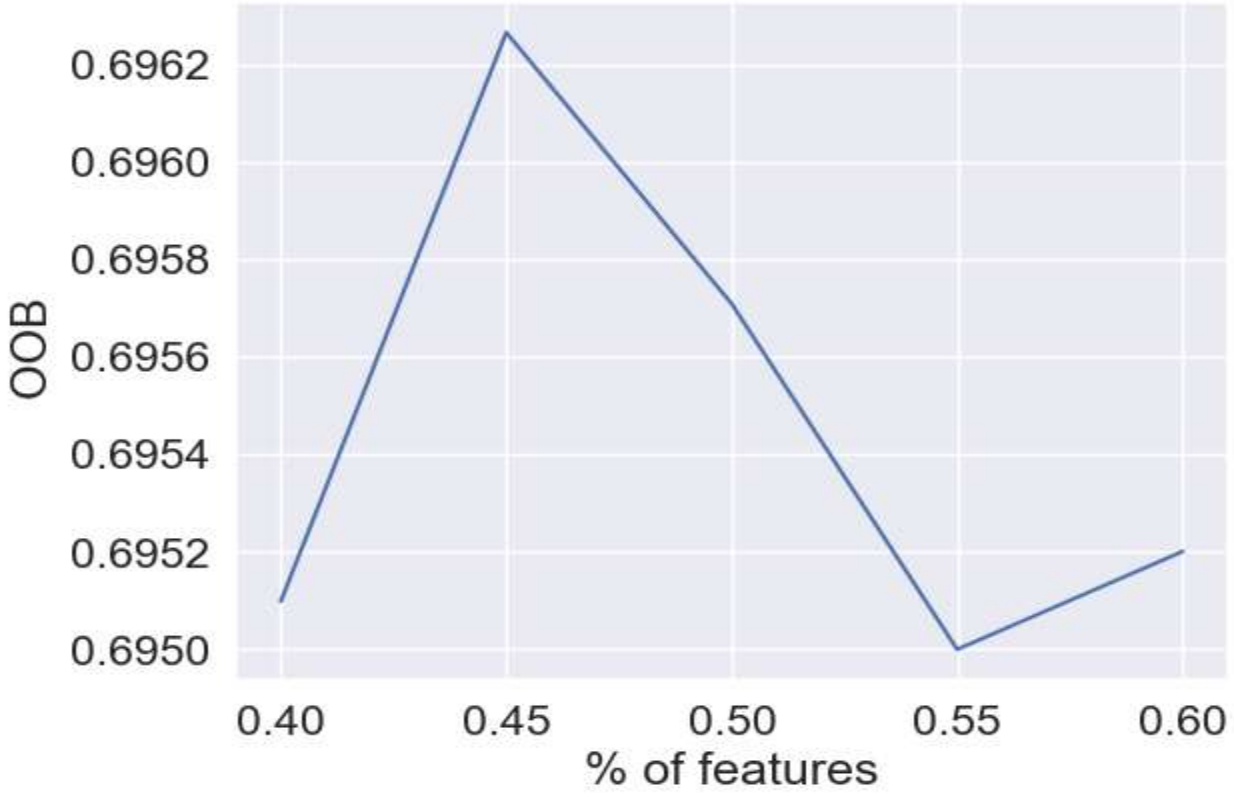


	feature	importance
1	room_type	0.279940
14	longitude	0.124422
13	latitude	0.106926
17	bedrooms	0.100541
3	accommodates	0.084841
4	bathrooms	0.063458

OSR Score: 64%

Importance

RANDOM FOREST REGRESSOR



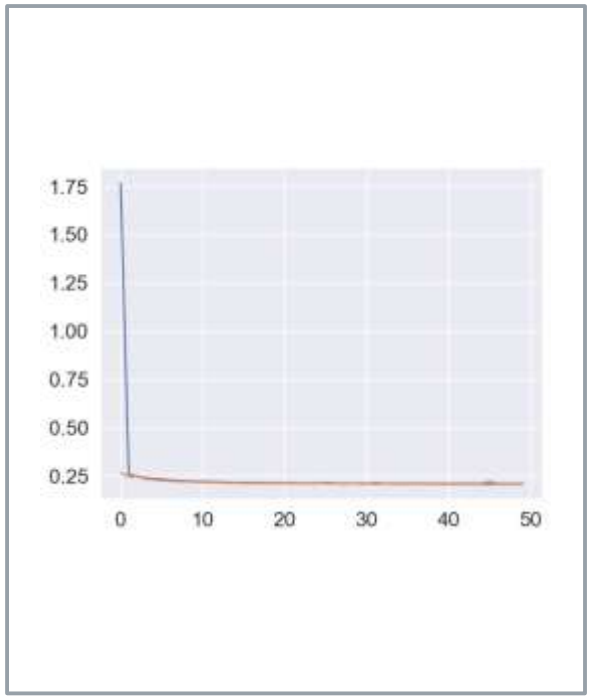
	feature	importance
1	room_type	0.279940
14	longitude	0.124422
13	latitude	0.106926
17	bedrooms	0.100541
3	accommodates	0.084841
4	bathrooms	0.063458

Importance

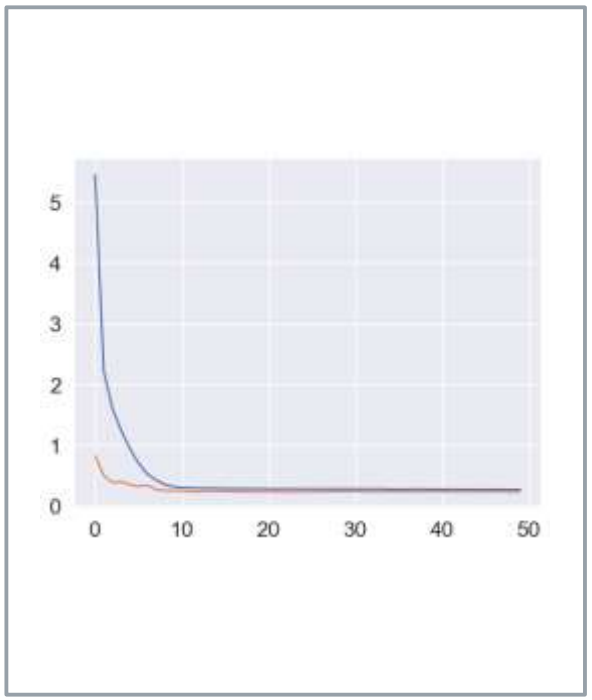
OOB Score: 70%

NEURAL NETWORK- MULTI LAYERED PERCEPTRON

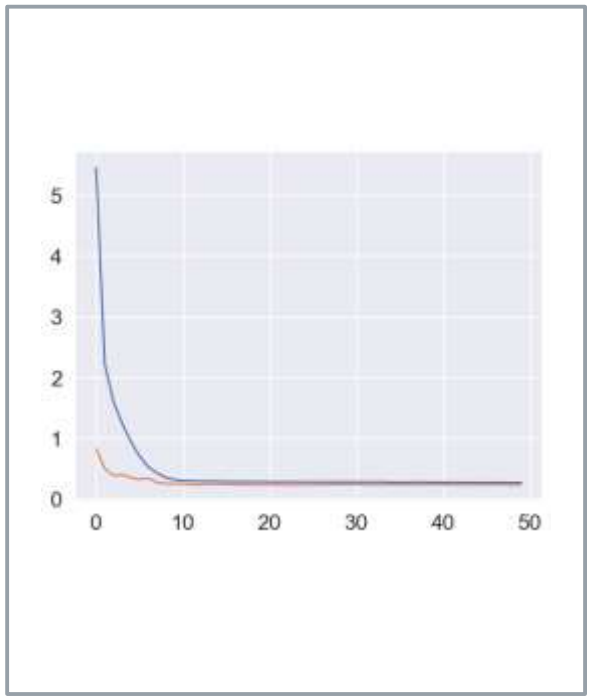
No feature selection
MAE: 0.36
Accuracy 56%



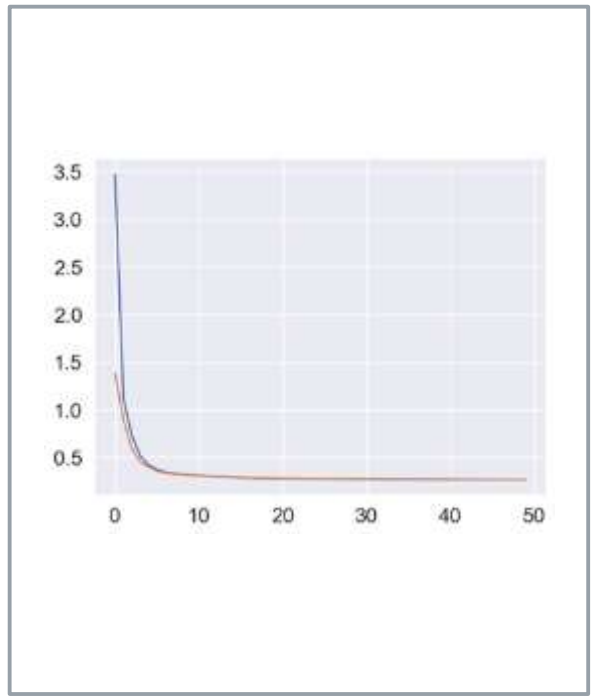
With More Layers
MAE: 0.37
Accuracy 55%



With Drop-out
MAE: 0.38
Accuracy 53%



With Regularisation
MAE: 0.37
Accuracy 56%



X-axis = Number of epochs , Y-axis = Loss

COMPARISON OF MODELS

Linear Regression: 54%
Decision Tree: 64%
Random Forest: 70%
Neural Network
- Benchmark: 56%
- With more layers: 55%
- With Drop-out: 53%
- With L1/L2 Regularization: 56%

Random Forest Regressor → Best Performing Model with 70% Accuracy!
Triggering Features in order → Room Type, Location, Accommodates, Bedrooms, Bathrooms

THANK YOU!

ANY QUESTIONS?