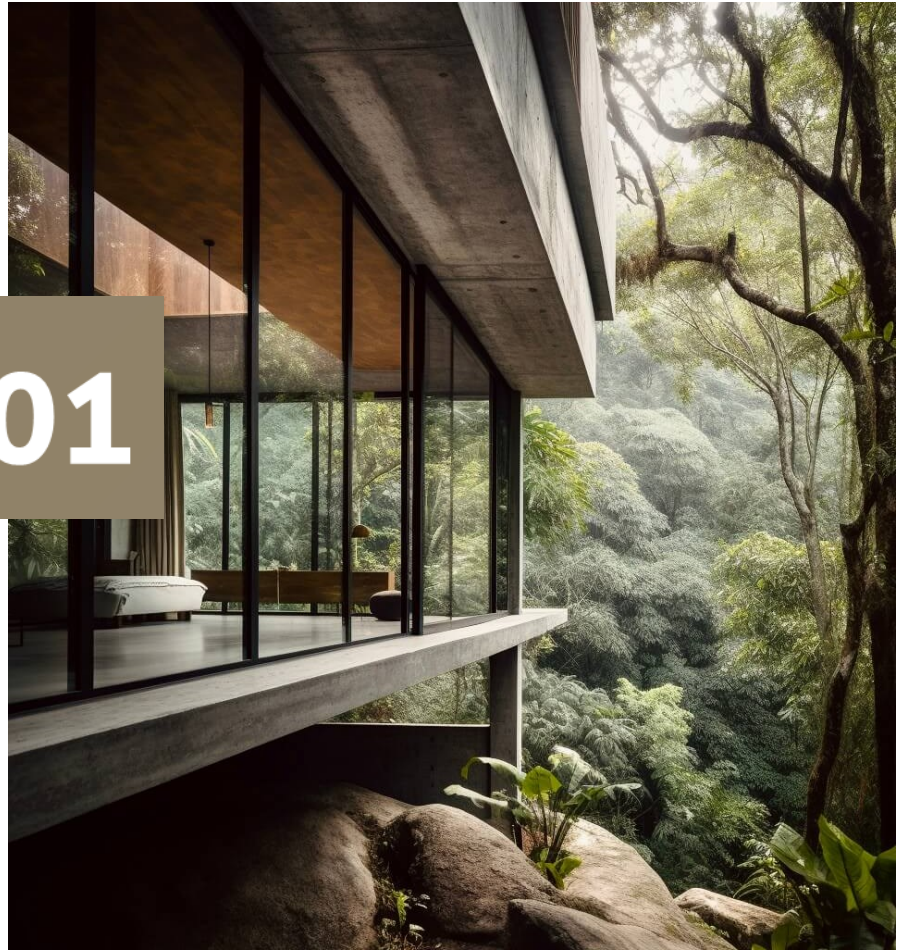
HOUSE RENT PREDICTION

Team Members:

- Allieah Ferrao (2315)
- Shannon Sequeira (2340)

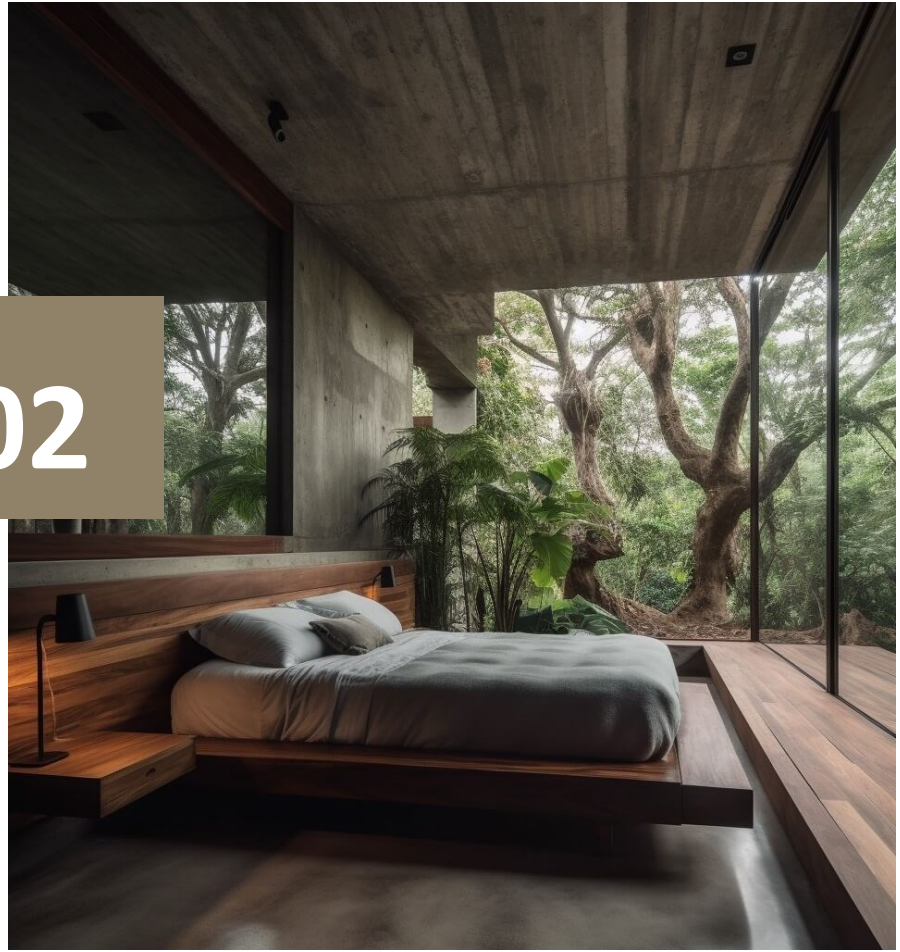
TABLE OF CONTENTS



01

UNDERSTANDING THE WORK

- **Problem Statement:**
  Predicting House Rents based on various features.
- **Objective:**
  To develop a predictive model to estimate house rents based on various property attributes.
- **Significance:**
  Facilitates better decision-making for tenants and landlords.



02

UNDERSTANDING THE DATASET

- The dataset comprises information on over **4700** houses, apartments, and flats available for rent.
- **Key parameters** include BHK (bedrooms, hall, kitchen), rent, size, number of floors, area type, area locality, city, furnishing status, tenant preferences, number of bathrooms, and point of contact.
- Exploratory data analysis reveals the diversity and complexity of the housing market, reflecting the wide range of property types and attributes.

03

OUR APPROACH TO ACHIEVE WORK OBJECTIVE

- Machine learning approach to predict rent prices based on the available dataset.
- **Data preprocessing** involves handling outliers, encoding categorical variables, and scaling numerical features.
- **Feature engineering** may include extracting additional information from existing features or creating new features to enhance model performance.
- Experimented with various regression algorithms such **as linear regression, random forest,** to identify the most effective model for rent prediction.

STEP-BY-STEP APPROACH

1. **Data Exploration and Cleaning:**
   - Load and inspect the dataset, handle missing values, duplicates, and outliers.
   - Extract relevant features and drop unnecessary columns.
2. **Feature Engineering and Encoding:**
   - Engineer new features if needed and encode categorical features into numerical values.
3. **Data Splitting:**
   - Split the dataset into features and target variables.
   - Further split the data into training and testing sets.
4. **Model Building:**
   - Train and evaluate different regression models such as Linear Regression, Random Forest Regressor, and Support Vector Machine (SVM).
5. **Model Evaluation:**
   - Evaluate each model using metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2) Score.
6. **Error Analysis and Conclusion:**
   - Analyze model errors to identify strengths and weaknesses.
   - Select the best-performing model based on evaluation results and draw conclusions.

04

WHERE OUR APPROACH/MODEL IS FAILING: ERROR ANALYSIS

Error Analysis:

- Model evaluation reveals discrepancies between predicted and actual rent prices.
- Common errors include underestimation or overestimation of rent values, particularly for properties with unique attributes or in outlier scenarios.
- Challenges arise from data quality issues, such as inaccurate or incomplete property information, as well as the inherent variability of the housing market.

Visualizations:

- Actual vs Predicted House Prices
- Scatter plot of Actual vs Predicted Prices

04

WHERE OUR APPROACH/MODEL IS FAILING: ERROR ANALYSIS

Mean Squared Error (MSE):
MSE measures the average of the squares of the errors or deviations.
It provides an overall idea of how close the predictions are to the actual values.
High MSE indicates large errors between predicted and actual values.

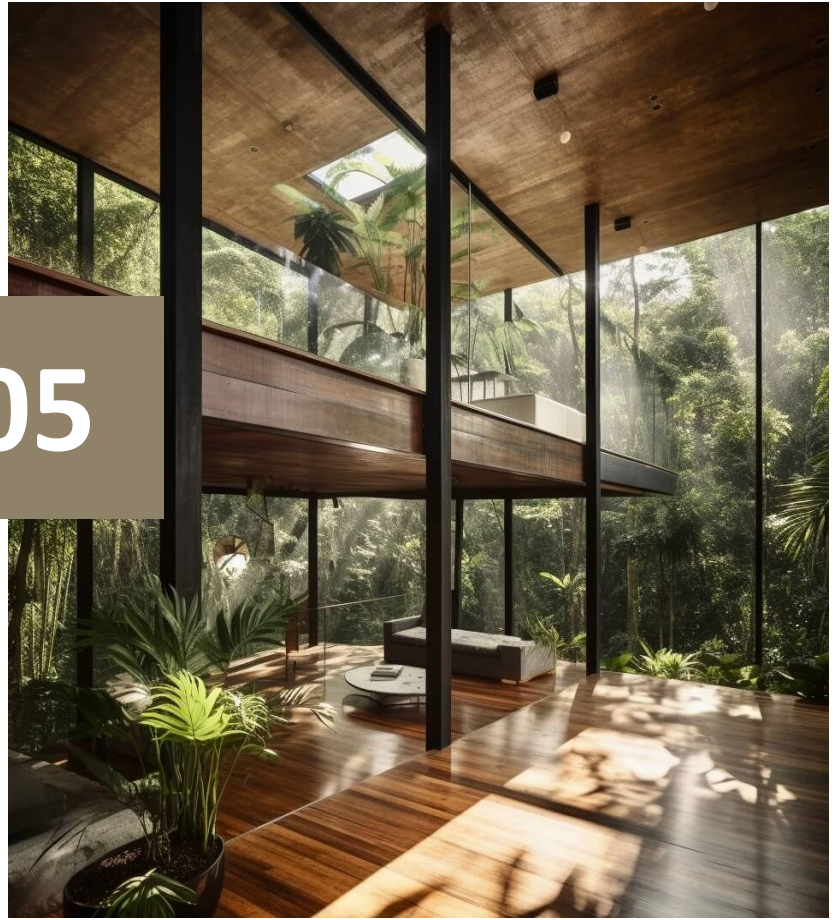
Mean Absolute Error (MAE):
MAE measures the average of the absolute errors between predicted and actual values.
It is more robust to outliers compared to MSE.
MAE gives a clear indication of the average magnitude of errors.

04

WHERE OUR APPROACH/MODEL IS FAILING: ERROR ANALYSIS

Root Mean Squared Error (RMSE):
RMSE is the square root of the MSE.
It provides a measure of the spread of errors in the predicted values.
RMSE is interpreted in the same units as the target variable.

R-squared (R2) Score:
R2 score measures the proportion of the variance in the dependent variable that is predictable from the independent variables.
It ranges from 0 to 1, where 1 indicates a perfect fit.
A higher R2 score indicates a better fit of the model to the data.



05

COMPARING WITH EXISTING APPROACH

Traditional methods rely on simplistic models or heuristic pricing strategies, lacking scalability and struggling with complex datasets.

Our machine learning approach offers:

- **Scalability**: Efficient handling of large, diverse datasets.
- **Automation**: Automated prediction process, reducing manual effort.
- *Adaptability*: Handling various property attributes and market dynamics.

In comparative analysis, our approach outperforms traditional methods in predictive accuracy, efficiency, and scalability, providing valuable insights for informed decision-making in real estate.