**SCRIPT FOR PROJECT EVALUATION VAISHNAV AND BHUSHAN**

**Bhushan:**

**Slide 1: Title Slide**  
**Speech:**  
“Good morning everyone,  
We are here to present our academic project titled *‘House Price Prediction.’* My name is Bhushan, and this is Vaishnav. Together, we have been working on this year-long project as part of our academic curriculum.

Our goal is to predict house prices using data-driven techniques, focusing on uncovering patterns in data and leveraging machine learning models to provide accurate predictions.

Today, we’ll be sharing our progress so far, which includes:

1. An introduction to the project and its objectives.
2. A detailed description of the dataset we’re using.
3. Key insights derived from exploratory data analysis (EDA).
4. Challenges we faced during the analysis.
5. Future steps we plan to take.

Let’s start with an introduction to the project.”

**Vaishnav: Introduction**

**Slide 2: Introduction**  
**Speech:**  
“Thank you, Bhushan.  
Our project aims to address a common challenge in the real estate market—predicting house prices accurately.

**Why is this important?**  
House prices depend on various factors like size, location, and amenities. These prices can often be unpredictable, making it challenging for buyers, sellers, and investors to make informed decisions. With the help of data analysis and machine learning, we aim to create a predictive model that provides better price estimations.

**Scope:**  
Currently, we are in the initial stages of our project. We have focused on data collection, cleaning, and exploratory data analysis to uncover patterns and relationships. Moving forward, these insights will help us build a robust predictive model.”

**Bhushan: Dataset Overview**

**Slide 3: Dataset Description**  
**Speech:**  
“To build our model, we are using a dataset that we sourced from [source, e.g., Kaggle or government real estate data].

This dataset contains **[X rows] and [Y columns],** representing thousands of properties. Key features include:

1. **Price:** Our target variable, representing the price of a house.
2. **Size (in square feet):** The area of the house.
3. **Location:** The geographical area where the property is located.
4. **Number of Bedrooms and Bathrooms:** Key factors influencing the price.
5. **Other Features:** Such as the type of property (e.g., apartment, villa), year built, and amenities like parking or garden access.

Here is a quick snapshot of the dataset: [Refer to a screenshot in your slide].

This dataset provides a rich set of variables to analyze and build our model. However, as with most real-world data, we encountered challenges that required cleaning and preprocessing, which Vaishnav will explain next.”

**Vaishnav: Data Cleaning**

**Slide 4: Data Cleaning and Preprocessing**  
**Speech:**  
“Thank you, Bhushan.  
Before conducting analysis, we noticed several issues in the dataset, such as:

1. **Missing Values:** Some important features, like Price and Size, had missing data.
2. **Outliers:** Extreme values in features like Price that could distort our results.
3. **Duplicate Entries:** A few rows in the dataset were exact duplicates.

**Steps Taken:**

1. For missing values, we applied techniques like replacing them with the median for numerical data or the most frequent value for categorical data.
2. Outliers were identified using boxplots and treated by capping them within a reasonable range.
3. Duplicate entries were removed to maintain data integrity.

This preprocessing step ensured that the dataset was clean and ready for analysis.”

**Bhushan: Exploratory Data Analysis (EDA)**

**Slide 5: Exploratory Data Analysis**  
**Speech:**  
“With the cleaned dataset, we moved on to exploratory data analysis, or EDA, to identify patterns and relationships.

Here’s what we found:

1. **Distribution of House Prices:**
   * A histogram revealed that most properties fall within a mid-price range, with a few high-end properties as outliers. [Point to histogram on slide.]
2. **Correlation Between Features:**
   * A heatmap showed a strong positive correlation between Size and Price, indicating that larger houses tend to cost more.
   * Other features like Location also showed significant impact on prices. [Point to heatmap.]
3. **Scatter Plot Analysis:**
   * A scatter plot of Price vs. Size highlighted this relationship. However, we also noticed diminishing returns—beyond a certain size, price increases at a slower rate. [Point to scatter plot.]

These visuals helped us understand the data better and paved the way for deriving insights.”

**Vaishnav: Key Insights**

**Slide 6: Key Insights from EDA**  
**Speech:**  
“From our analysis, we derived several insights:

1. **Location Matters:** Urban properties are significantly more expensive than rural ones, likely due to higher demand and better infrastructure.
2. **Size and Price Relationship:** Larger homes generally have higher prices, but the relationship isn’t linear.
3. **Strong Predictors:** Features like Location, Size, and Number of Bedrooms show a strong correlation with price, making them important for our model.

These findings will guide us as we move to the modeling phase.”

**Bhushan: Challenges Faced**

**Slide 7: Challenges**  
**Speech:**  
“Of course, we faced several challenges along the way:

1. **Missing Data:** Missing values in critical features required careful imputation to avoid bias.
2. **Outliers:** Price outliers posed a significant challenge, as they could distort the results if not handled properly.
3. **Imbalanced Data:** Certain categories, like luxury homes, were underrepresented in the dataset, which may affect predictions.

We addressed these challenges during preprocessing, but some, like class imbalance, may require further adjustments during model building.”

**Vaishnav: Future Scope**

**Slide 8: Future Scope**  
**Speech:**  
“Moving forward, our focus will shift to:

1. **Feature Engineering:**
   * Creating derived features like price per square foot or clustering locations into urban, suburban, and rural areas.
2. **Model Building:**
   * We plan to use regression-based models, starting with Linear Regression, and later explore advanced techniques like Decision Trees or Random Forests.
3. **Model Evaluation:**
   * Evaluating the models using metrics such as RMSE (Root Mean Squared Error) and R² to ensure accuracy and reliability.

These steps will bring us closer to building a predictive model that meets the project’s objectives.”

**Bhushan: Conclusion**

**Slide 9: Conclusion**  
**Speech:**  
“To conclude, we’ve made significant progress in understanding the data through exploratory analysis. The patterns we uncovered will guide the next phases of feature engineering and model building.

Our ultimate goal remains to deliver a robust prediction model that can assist buyers, sellers, and investors in making informed decisions.

Thank you for your attention. We now invite any questions or feedback you might have.”