Supervised Learning — Concept Document

Name: Sambathini Paul Jonathan
Day: 1
Topic: Supervised Learning – Regression and Classification

Steps:

1. **Imported necessary libraries**
(numpy, pandas, matplotlib, sklearn) for data handling, visualization, and modeling.

2. **Loaded dataset**
Used the **California Housing dataset** from scikit-learn as the labeled data for regression.

3. **Selected features and target variable**
Independent features: MedInc, HouseAge, AveRooms, AveOccup
Target variable: MedHouseVal (Median house value).

4. **Split the data**
Divided data into **80% training** and **20% testing** sets using train_test_split().

5. **Trained the Linear Regression model**
Initialized LinearRegression() and fitted it on the training data (X_train, y_train).

6. **Made predictions**
Predicted house values using the test data and stored results in y_pred.

7. **Evaluated model performance**
Calculated **Mean Squared Error (MSE)** and **R^2 score** to measure model accuracy.

8. **Visualized results**
Created a scatter plot of **Actual vs Predicted** house prices to observe model performance visually.

9. **Interpreted coefficients**
Displayed feature coefficients to understand how each feature influences house price.

Reflected on learning
Understood how Linear Regression finds relationships between variables and how feature scaling and data quality impact accuracy.