Task 01

Explanation of code:

House Price Prediction:

This project uses different features such as lot size, number of rooms, year built, and neighborhood to predict house prices.

Step-by-Step Breakdown:

Loading the Data:

- Data is loaded from the CSV file using pandas.read_csv(), and simple exploration (head(), tail(), describe(), info()) is performed.

Handling Missing Values:

- Numerical missing values (such as GarageYrBlt, MasVnrArea, LotFrontage) are filled with the column mean.
- Categorical missing values (such as Electrical, BsmtQual, GarageType) are filled with the mode.

Dropping Unnecessary Columns:

- Certain columns (Alley, PoolQC, Fence, MiscFeature, FireplaceQu) are dropped because of too many missing values.

Encoding Categorical Data:

- Categorical columns are encoded into numbers with LabelEncoder().

Splitting Data for Training:

- The data is separated into features (X) and target (y, which is SalePrice).
- The data is split into training set and test set using train_test_split().

Building the Model:

- A RandomForestRegressor (a machine learning model for predicting numerical data) is fitted on the data.
- The model is stored with pickle for later use.

Model Evaluation:

- Predictions are done on the test set.
- Root Mean Squared Error (RMSE) is used to evaluate prediction accuracy.

Predictions on New Data:

- Test dataset is preprocessed similarly.

- Predictions are done, and a submission file (submission.csv) is prepared.

Root Mean Square Error: