

# PRODIGY\_ML\_01

## TASK\_01

### IMPORT LIBRARIES

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

### LOADING DATASET

```
#load the training dataset
train_data=pd.read_csv('train.csv')
#displaying the first few rows of the datasets
train_data.head()
```

### EXPLORING THE DATASET

```
#checking the missing value
train_data.isnull().sum()
#basis statistics of the dataset
train_data.describe()
```

### FEATURE SELECTION

```
# Select features and target variable
features = ['GrLivArea', 'BedroomAbvGr', 'FullBath']
X = train_data[features]
y = train_data['SalePrice']
```

## **HANDLING MISSING VALUES**

```
#Filling the missing values with the median(if any)
X=X.fillna(X.median())
```

## **SPLIT THE DATASET**

```
#splitting it in training and testing set
X_train, X_test, y_train, y_test =
train_test_split(X,y,test_size=0.2,random_state=42)
```

## **TRAIN THE LINEAR REGRESSION MODEL**

```
#Initialize the linear regression model
model=LinearRegression()
#Train the model
model.fit(X_train,y_train)
```

## **MAKE PREDICTION**

```
y_pred=model.predict(X_test)
```

## **EVALUATE THE MODEL**

```
# Calculate Mean Squared Error (MSE)
mse=mean_squared_error(y_test,y_pred)
# Calculate R_squared
r2=r2_score(y_test,y_pred)
print(f'Mean Squared Error: {mse}')
print(f'R-squared: {r2}')
```

## **VISUALIZE THE RESULTS**

```
plt.scatter(y_test, y_pred)
plt.xlabel('Actual Prices')
plt.ylabel('Predicted Prices')
plt.title('Actual vs Predicted Prices')
plt.show()
```

