```python
import pandas as pd  import numpy as np
import matplotlib.pyplot as plt import seaborn
as sns
from sklearn.preprocessing import
MinMaxScaler, LabelEncoder
from sklearn.model_selection import
train_test_split from sklearn.linear_model
import LinearRegression from sklearn.metrics
import mean_squared_error, r2_score from
google.colab import files  files.upload()
df = pd.read_csv('HousePricePrediction.xlsx -
Sheet1.csv') print(df.head()) print(df.describe())
# Check for missing values
print(df.isnull().sum())
# Correlation matrix to understand feature
relationships correlation_matrix =
df.select_dtypes(include='number').corr()
sns.heatmap(correlation_matrix, annot=True,
cmap='coolwarm') plt.title("Correlation Matrix")
plt.show()
# Preprocessing: Selecting features and target
variable
X = df[['bedrooms', 'bathrooms', 'sqft_living',
'sqft_lot', 'floors',
'waterfront', 'view', 'condition']] y = df['price']
# Splitting the dataset into training and testing
sets
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,
random_state=42)
```