

Task 3

Name: Jaavanika L

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

data = pd.read_csv("C:/Users/JAAVANIKA L/Fall semester 22-23/Downloads/Housing.csv")

data.head()
```

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
0	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
1	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
2	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
3	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
4	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished

```
data.describe()
```

	price	area	bedrooms	bathrooms	stories	parking
count	5.450000e+02	545.000000	545.000000	545.000000	545.000000	545.000000
mean	4.766729e+06	5150.541284	2.965138	1.286239	1.805505	0.693578
std	1.870440e+06	2170.141023	0.738064	0.502470	0.867492	0.861586
min	1.750000e+06	1650.000000	1.000000	1.000000	1.000000	0.000000
25%	3.430000e+06	3600.000000	2.000000	1.000000	1.000000	0.000000
50%	4.340000e+06	4600.000000	3.000000	1.000000	2.000000	0.000000
75%	5.740000e+06	6360.000000	3.000000	2.000000	2.000000	1.000000
max	1.330000e+07	16200.000000	6.000000	4.000000	4.000000	3.000000

```
le = LabelEncoder()
for col in data.columns:
    if data[col].dtype == 'object':
        data[col] = le.fit_transform(data[col])

X = data.drop(columns=['furnishingstatus'])
y = data['furnishingstatus']

data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 545 entries, 0 to 544
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   price                 545 non-null    int64
 1   area                  545 non-null    int64
 2   bedrooms              545 non-null    int64
 3   bathrooms             545 non-null    int64
 4   stories               545 non-null    int64
 5   mainroad              545 non-null    int32
 6   guestroom             545 non-null    int32
 7   basement              545 non-null    int32
 8   hotwaterheating       545 non-null    int32
 9   airconditioning       545 non-null    int32
10   parking               545 non-null    int64
11   prefarea              545 non-null    int32
12   furnishingstatus      545 non-null    int32
dtypes: int32(7), int64(6)
memory usage: 40.6 KB
```

```
data.head()
```

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
0	13300000	7420	4	2	3	1	0	0	0	1	2	1	0
1	12250000	8960	4	4	4	1	0	0	0	1	3	0	0
2	12250000	9960	3	2	2	1	0	1	0	0	2	1	1
3	12215000	7500	4	2	2	1	0	1	0	1	3	1	0
4	11410000	7420	4	1	2	1	1	1	0	1	2	0	0

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
model = LinearRegression()  
model.fit(X_train, y_train)
```

```
▼ LinearRegression ⓘ ⓘ  
LinearRegression()
```

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

```
mae = mean_absolute_error(y_test, y_pred)  
mse = mean_squared_error(y_test, y_pred)  
r2 = r2_score(y_test, y_pred)  
print(f"Mean Absolute Error: {mae}")  
print(f"Mean Squared Error: {mse}")  
print(f"R2 Score: {r2}")
```

```
Mean Absolute Error: 0.6528965180210892  
Mean Squared Error: 0.6121348260126974  
R2 Score: 0.02848331981607577
```

```
plt.scatter(y_test, y_pred)  
plt.xlabel("Actual")  
plt.ylabel("Predicted")  
plt.title("Actual vs Predicted Furnishing Status")  
plt.show()
```

```
plt.scatter(y_test, y_pred)  
plt.xlabel("Actual")  
plt.ylabel("Predicted")  
plt.title("Actual vs Predicted Furnishing Status")  
plt.show()
```

