```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('HousingData.csv')
df

df.shape

df.isnull().sum()

df['CRIM'] = df['CRIM'].fillna(df['CRIM'].mean())
df['ZN'] = df['ZN'].fillna(df['ZN'].mean())
df['INDUS'] = df['INDUS'].fillna(df['INDUS'].mean())
df['CHAS'] = df['CHAS'].fillna(df['CHAS'].mean())
df['AGE'] = df['AGE'].fillna(df['NOX'].mean())
df['LSTAT'] = df['LSTAT'].fillna(df['LSTAT'].mean())
df

df.isnull().sum()

x = df.drop('MEDV', axis=1)
x

y = df['MEDV']
y

from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=0)

x_train
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(x_train.values, y_train.values)

y_train_pred = model.predict(x_train.values)
y_test_pred = model.predict(x_test.values)
```

```
y_train_pred

model.predict([[0.00632, 18.0, 2.31, 0.0, 0.538, 6.575, 65.2, 4.0900, 1.0, 296.0, 15.3, 396.90, 4.98]])

model.predict([[0.00632, 18.0, 2.31, 0.0, 0.538, 6.575, 65.2, 4.0900, 1.0, 296.0, 15.3, 396.90, 4.98]])

y_train

y_test

df1 = pd.DataFrame({'Actual': y_train, 'Predicted': y_train_pred})
df2 = pd.DataFrame({'Actual': y_test, 'Predicted': y_test_pred})

df1
df2
from sklearn.metrics import mean_squared_error, r2_score
mse = mean_squared_error(y_test, y_test_pred)
mse

mse = mean_squared_error(y_train, y_train_pred)
mse

r2 = r2_score(y_test, y_test_pred)
r2
plt.scatter(y_test, y_test_pred, c='red', marker='o', label='Test Data')
plt.scatter(y_train, y_train_pred, c='blue', marker='s', label='Train Data')
plt.xlabel('Actual Prices')
plt.ylabel('Predicted Prices')
plt.title('Actual Prices vs Predicted Prices')
plt.legend(loc='upper left')
plt.plot()
plt.show()
```