**import numpy as np import pandas as pd**

**5/7/24, 5:05 PM Assignment 4.ipynb - Colab**

[**https://colab.research.google.com/drive/1PZyz6kV0mriEVRYmm7YTpthE9GJhGX55#scrollTo=UFV2pViVone-&printMode=true**](https://colab.research.google.com/drive/1PZyz6kV0mriEVRYmm7YTpthE9GJhGX55#scrollTo=UFV2pViVone-&printMode=true) **1/3**

**data=pd.read\_csv(Vcontent/HousingData.csv.) df=pd.DataFrame(data)**

**df**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** | **B** | **LSTAT** | **MEDV** |
| **0** | **0.00632** | **18.0** | **2.31** | **0.0** | **0.538** | **6.575** | **65.2** | **4.0900** | **1** | **296** | **15.3** | **396.90** | **4.98** | **24.0** |
| **1** | **0.02731** | **0.0** | **7.07** | **0.0** | **0.469** | **6.421** | **78.9** | **4.9671** | **2** | **242** | **17.8** | **396.90** | **9.14** | **21.6** |
| **2** | **0.02729** | **0.0** | **7.07** | **0.0** | **0.469** | **7.185** | **61.1** | **4.9671** | **2** | **242** | **17.8** | **392.83** | **4.03** | **34.7** |
| **3** | **0.03237** | **0.0** | **2.18** | **0.0** | **0.458** | **6.998** | **45.8** | **6.0622** | **3** | **222** | **18.7** | **394.63** | **2.94** | **33.4** |
| **4** | **0.06905** | **0.0** | **2.18** | **0.0** | **0.458** | **7.147** | **54.2** | **6.0622** | **3** | **222** | **18.7** | **396.90** | **NaN** | **36.2** |
| **501** | **0.06263** | **0.0** | **11.93** | **0.0** | **0.573** | **6.593** | **69.1** | **2.4786** | **1** | **273** | **21.0** | **391.99** | **NaN** | **22.4** |
| **502** | **0.04527** | **0.0** | **11.93** | **0.0** | **0.573** | **6.120** | **76.7** | **2.2875** | **1** | **273** | **21.0** | **396.90** | **9.08** | **20.6** |
| **503** | **0.06076** | **0.0** | **11.93** | **0.0** | **0.573** | **6.976** | **91.0** | **2.1675** | **1** | **273** | **21.0** | **396.90** | **5.64** | **23.9** |
| **504** | **0.10959** | **0.0** | **11.93** | **0.0** | **0.573** | **6.794** | **89.3** | **2.3889** | **1** | **273** | **21.0** | **393.45** | **6.48** | **22.0** |
| **505** | **0.04741** | **0.0** | **11.93** | **0.0** | **0.573** | **6.030** | **NaN** | **2.5050** | **1** | **273** | **21.0** | **396.90** | **7.88** | **11.9** |

**506 rows x 14 columns**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **df.describe()** | **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** |
| **count** | **486.000000** | **486.000000** | **486.000000** | **486.000000** | **506.000000** | **506.000000** | **486.000000** |
| **mean** | **3.611874** | **11.211934** | **11.083992** | **0.069959** | **0.554695** | **6.284634** | **68.518519** |
| **std** | **8.720192** | **23.388876** | **6.835896** | **0.255340** | **0.115878** | **0.702617** | **27.999513** |
| **min** | **0.006320** | **0.000000** | **0.460000** | **0.000000** | **0.385000** | **3.561000** | **2.900000** |
| **25%** | **0.081900** | **0.000000** | **5.190000** | **0.000000** | **0.449000** | **5.885500** | **45.175000** |
| **50%** | **0.253715** | **0.000000** | **9.690000** | **0.000000** | **0.538000** | **6.208500** | **76.800000** |
| **75%** | **3.560263** | **12.500000** | **18.100000** | **0.000000** | **0.624000** | **6.623500** | **93.975000** |
| **max** | **88.976200** | **100.000000** | **27.740000** | **1.000000** | **0.871000** | **8.780000** | **100.000000** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **df.isnull()** | **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** | **B** | **LSD** |
| **0** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **Fah** |
| **1** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **Fah** |
| **2** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **Fah** |
| **3** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **Fah** |
| **4** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **Trt** |
| **501** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **Trt** |
| **502** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **Fah** |
| **503** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **Fah** |
| **504** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **False** | **Fah** |
| **505** | **False** | **False** | **False** | **False** | **False** | **False** | **True** | **False** | **False** | **False** | **False** | **False** | **Fah** |

**SUR rnws x 14 columns**

■

**df.isnull().sum()**

**5/7/24, 5:05 PM Assignment 4.ipynb - Colab**

[**https://colab.research.google.com/drive/1PZyz6kV0mriEVRYmm7YTpthE9GJhGX55#scrollTo=UFV2pViVone-&printMode=true**](https://colab.research.google.com/drive/1PZyz6kV0mriEVRYmm7YTpthE9GJhGX55#scrollTo=UFV2pViVone-&printMode=true) **2/3**

CRIM 20

ZN 20

INDUS 20

CHAS 20

NOX 0

RM 0

AGE 20

DIS 0

RAD 0

TAX 0

PTRATIO 0

B 0

LSTAT 20

MEDV 0

dtype: int64

df['AGE']=dfrAGE'Lfillna(np.mean(dfrAGE'D) df.isnull().sum()

CRIM 20

ZN 20

INDUS 20

CHAS 20

NOX 0

RM 0

AGE 0

DIS 0

RAD 0

TAX 0

PTRATIO 0

B 0

LSTAT 20

MEDV 0

dtype: int64

dff'CRIM']=dfUCRIM1.fillna(dfUCRIM'Imode()[0]) df['ZN']=df['ZN'].fillna(df['ZN'].mode()[0]) df['INDUS']=df['INDUS'].fillna(df['INDUS'].mode()[0]) df['CHAS']=dfUCHAS1.fillna(dfUCHAS1.mode()[0]) df['LSTAT']=df['LSTAT'].fillna(df['LSTAT'].mode()[0]) df.isnull().sum()

CRIM 0

ZN 0

INDUS 0

CHAS 0

NOX 0

RM 0

AGE 0

DIS 0

RAD 0

TAX 0

PTRATIO 0

B 0

LSTAT 0

MEDV 0

dtype: int64

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| df.head() | **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** | **B** | **LSTAT** |
| **0** | **0.00632** | **18.0** | **2.31** | **0.0** | **0.538** | **6.575** | **65.2** | **4.0900** | **1** | **296** | **15.3** | **396.90** | **4.98** |
| **1** | **0.02731** | **0.0** | **7.07** | **0.0** | **0.469** | **6.421** | **78.9** | **4.9671** | **2** | **242** | **17.8** | **396.90** | **9.14** |
| **2** | **0.02729** | **0.0** | **7.07** | **0.0** | **0.469** | **7.185** | **61.1** | **4.9671** | **2** | **242** | **17.8** | **392.83** | **4.03** |
| **3** | **0.03237** | **0.0** | **2.18** | **0.0** | **0.458** | **6.998** | **45.8** | **6.0622** | **3** | **222** | **18.7** | **394.63** | **2.94** |
| **4** | **0.06905** | **0.0** | **2.18** | **0.0** | **0.458** | **7.147** | **54.2** | **6.0622** | **3** | **222** | **18.7** | **396.90** | **6.36** |
|  |  |  |  |  |  |  |  |  |  |  |  |  | ■ |

df.tail()

**CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO**

**5/7/24, 5:05 PM Assignment 4.ipynb - Colab**

[**https://colab.research.google.com/drive/1PZyz6kV0mriEVRYmm7YTpthE9GJhGX55#scrollTo=UFV2pViVone-&printMode=true**](https://colab.research.google.com/drive/1PZyz6kV0mriEVRYmm7YTpthE9GJhGX55#scrollTo=UFV2pViVone-&printMode=true) **3/3**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **501** | 0.06263 | 0.0 | 11.93 | 0.0 | 0.573 | 6.593 | 69.100000 | 2.4786 | 1 | 273 | 21.0 | 391.99 |
| **502** | 0.04527 | 0.0 | 11.93 | 0.0 | 0.573 | 6.120 | 76.700000 | 2.2875 | 1 | 273 | 21.0 | 396.90 |

'13','LSTAT']]

y=df[['MEDV']]

from sklearn.model\_selection import train\_test\_split from sklearn.linear\_model import LinearRegression

X\_train,X\_test,y\_train,y\_test=train\_test\_split(X,y,test\_size=0.4)

lm=LinearRegression() lm.fit(X\_train,y\_train)

LinearRegression linearRegression01

LinearRegression(copy\_X=True,fit\_intercept=True,n\_jobs=None)

LinearRegression:

LinearRegressionpl

predictions=lm.predict(X\_test)

from sklearn import metrics

print('MAE:',metrics.mean\_absolute\_error(y\_test,predictions)) print('MSE:',metrics.mean\_squared\_error(y\_test,predictions))

MAE: 3.1206149796393485   
MSE: 20.155998556981956

print('RMSE:',np.sqrt(metrics.mean\_absolute\_error(y\_test,predictions))) RMSE: 1.7665262465186722

Start coding or generate with AI.