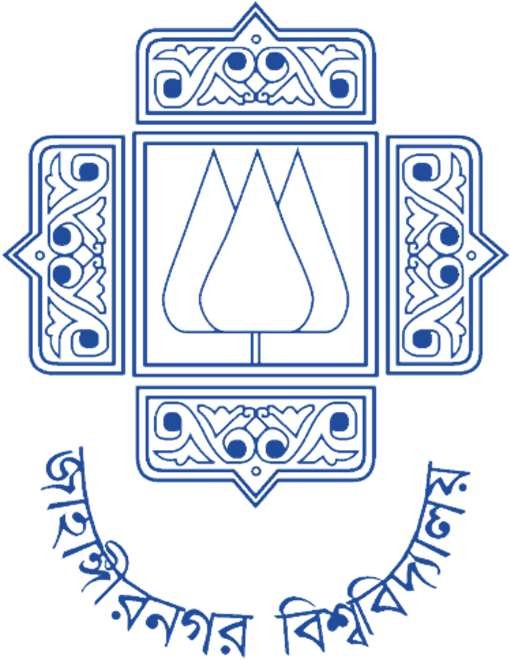
Institute of Information Technology (IIT)

Jahangirnagar University



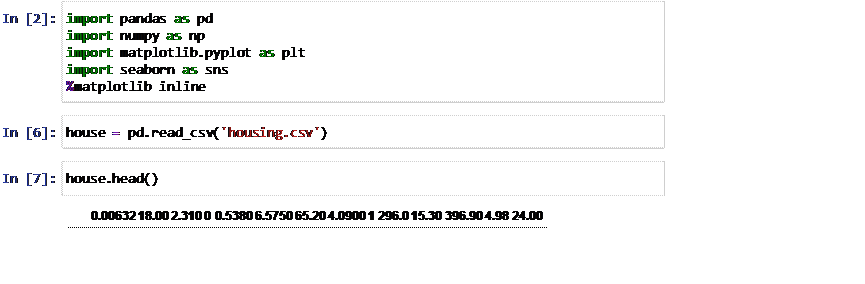
**Lab Report: 04**

Submitted by:

Name: Md. Shakil Hossain

Roll No: 2023

Lab Date: 11-July-2023 Submission Date: 11-Aug-2023



Out[7]:

1. 0.02731 0.00 7.070 0 0.4690 6.4210 78...
2. 0.02729 0.00 7.070 0 0.4690 7.1850 61...
3. 0.03237 0.00 2.180 0 0.4580 6.9980 45...
4. 0.06905 0.00 2.180 0 0.4580 7.1470 54...
5. 0.02985 0.00 2.180 0 0.4580 6.4300 58...

In

[8]:

house

.

columns

Out[8]: Index([' 0.00632 18.00 2.310 0 0.5380 6.5750 65.20 4.0900 1 296.0

15.30 396.90 4.98 24.00'], dtype='object')

Out[20]:

In

[16]:

In

[19]:

In

[20]:

**CRIM**

**ZN**

**INDUS**

**CHAS**

**NOX**

**RM**

**AGE**

**DIS**

**RAD**

**TAX**

**PTRATIO**

**B**

**LSTA**

**0**

0.00632

18.0

2.31

0

0.538

6.575

65.2

4.0900

1

296.0

15.3

396.90

4.9

col

**=**

[

'CRIM'

,

'ZN'

,

'INDUS'

,

'CHAS'

,

'NOX'

,

'RM'

,

'AGE'

,

'DIS'

,

'RAD'

,

'TAX'

,

house\_data

**=**

pd

.

read\_csv

(

'housing.csv'

,

header

**=**

**None**

,

delimiter

**=**

r"\s+"

,

names

**=**

house\_data

.

head

()

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | 0.02731 | 0.0 | 7.07 | 0 | 0.469 | 6.421 | 78.9 | 4.9671 | 2 | 242.0 | 17.8 | 396.90 | 9. |
| **2** | 0.02729 | 0.0 | 7.07 | 0 | 0.469 | 7.185 | 61.1 | 4.9671 | 2 | 242.0 | 17.8 | 392.83 | 4.0 |
| **3** | 0.03237 | 0.0 | 2.18 | 0 | 0.458 | 6.998 | 45.8 | 6.0622 | 3 | 222.0 | 18.7 | 394.63 | 2.9 |
| **4** | 0.06905 | 0.0 | 2.18 | 0 | 0.458 | 7.147 | 54.2 | 6.0622 | 3 | 222.0 | 18.7 | 396.90 | 5.3 |

[21]:

house\_data

.

info

()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 506 entries, 0 to 505 Data columns (total 14 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

1. CRIM 506 non-null float64
2. ZN 506 non-null float64
3. INDUS 506 non-null float64
4. CHAS 506 non-null int64
5. NOX 506 non-null float64
6. RM 506 non-null float64
7. AGE 506 non-null float64
8. DIS 506 non-null float64
9. RAD 506 non-null int64
10. TAX 506 non-null float64
11. PTRATIO 506 non-null float64
12. B 506 non-null float64
13. LSTAT 506 non-null float64
14. MEDV 506 non-null float64 dtypes: float64(12), int64(2) memory usage: 55.5 KB

Out[22]:

In

[22]:

**CRIM**

**ZN**

**INDUS**

**CHAS**

**NOX**

**RM**

**AGE**

house\_data

.

describe

()

**count** 506.000000 506.000000 506.000000 506.000000 506.000000 506.000000 506.000000 506

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **mean** | 3.613524 | 11.363636 | 11.136779 | 0.069170 | 0.554695 | 6.284634 | 68.574901 | 3 |
| **std** | 8.601545 | 23.322453 | 6.860353 | 0.253994 | 0.115878 | 0.702617 | 28.148861 | 2 |
| **min** | 0.006320 | 0.000000 | 0.460000 | 0.000000 | 0.385000 | 3.561000 | 2.900000 | 1 |
| **25%** | 0.082045 | 0.000000 | 5.190000 | 0.000000 | 0.449000 | 5.885500 | 45.025000 | 2 |
| **50%** | 0.256510 | 0.000000 | 9.690000 | 0.000000 | 0.538000 | 6.208500 | 77.500000 | 3 |
| **75%** | 3.677083 | 12.500000 | 18.100000 | 0.000000 | 0.624000 | 6.623500 | 94.075000 | 5 |
| **max** | 88.976200 | 100.000000 | 27.740000 | 1.000000 | 0.871000 | 8.780000 | 100.000000 | 12 |
|  |  |  |  |  |  |  |  |  |

In

[24]:

house\_data

.

columns

Out[24]: Index(['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TA

X',

'PTRATIO', 'B', 'LSTAT', 'MEDV'], dtype='object')

In

[25]:

house\_data

.

index

Out[25]: RangeIndex(start=0, stop=506, step=1)

[27]: house\_data.isnull().sum()

Out[27]: 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

In

[29]:

corr

**=**

house\_data

.

corr

()

print

(

house\_data

)

# CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX \

1. 0.00632 18.0 2.31 0 0.538 6.575 65.2 4.0900 1 296.0
2. 0.02731 0.0 7.07 0 0.469 6.421 78.9 4.9671 2 242.0
3. 0.02729 0.0 7.07 0 0.469 7.185 61.1 4.9671 2 242.0
4. 0.03237 0.0 2.18 0 0.458 6.998 45.8 6.0622 3 222.0
5. 0.06905 0.0 2.18 0 0.458 7.147 54.2 6.0622 3 222.0 .. ... ... ... ... ... ... ... ... ... ...
6. 0.06263 0.0 11.93 0 0.573 6.593 69.1 2.4786 1 273.0
7. 0.04527 0.0 11.93 0 0.573 6.120 76.7 2.2875 1 273.0
8. 0.06076 0.0 11.93 0 0.573 6.976 91.0 2.1675 1 273.0
9. 0.10959 0.0 11.93 0 0.573 6.794 89.3 2.3889 1 273.0
10. 0.04741 0.0 11.93 0 0.573 6.030 80.8 2.5050 1 273.0

PTRATIO B LSTAT MEDV 0 15.3 396.90 4.98 24.0

1. 17.8 396.90 9.14 21.6
2. 17.8 392.83 4.03 34.7
3. 18.7 394.63 2.94 33.4
4. 18.7 396.90 5.33 36.2 .. ... ... ... ...
5. 21.0 391.99 9.67 22.4
6. 21.0 396.90 9.08 20.6
7. 21.0 396.90 5.64 23.9
8. 21.0 393.45 6.48 22.0
9. 21.0 396.90 7.88 11.9

[506 rows x 14 columns]

[34]:

plt

.

figure

(

figsize

**=**

(

20

,

15

))

sns

.

heatmap

(

house\_data

.

corr

().

abs

()

,

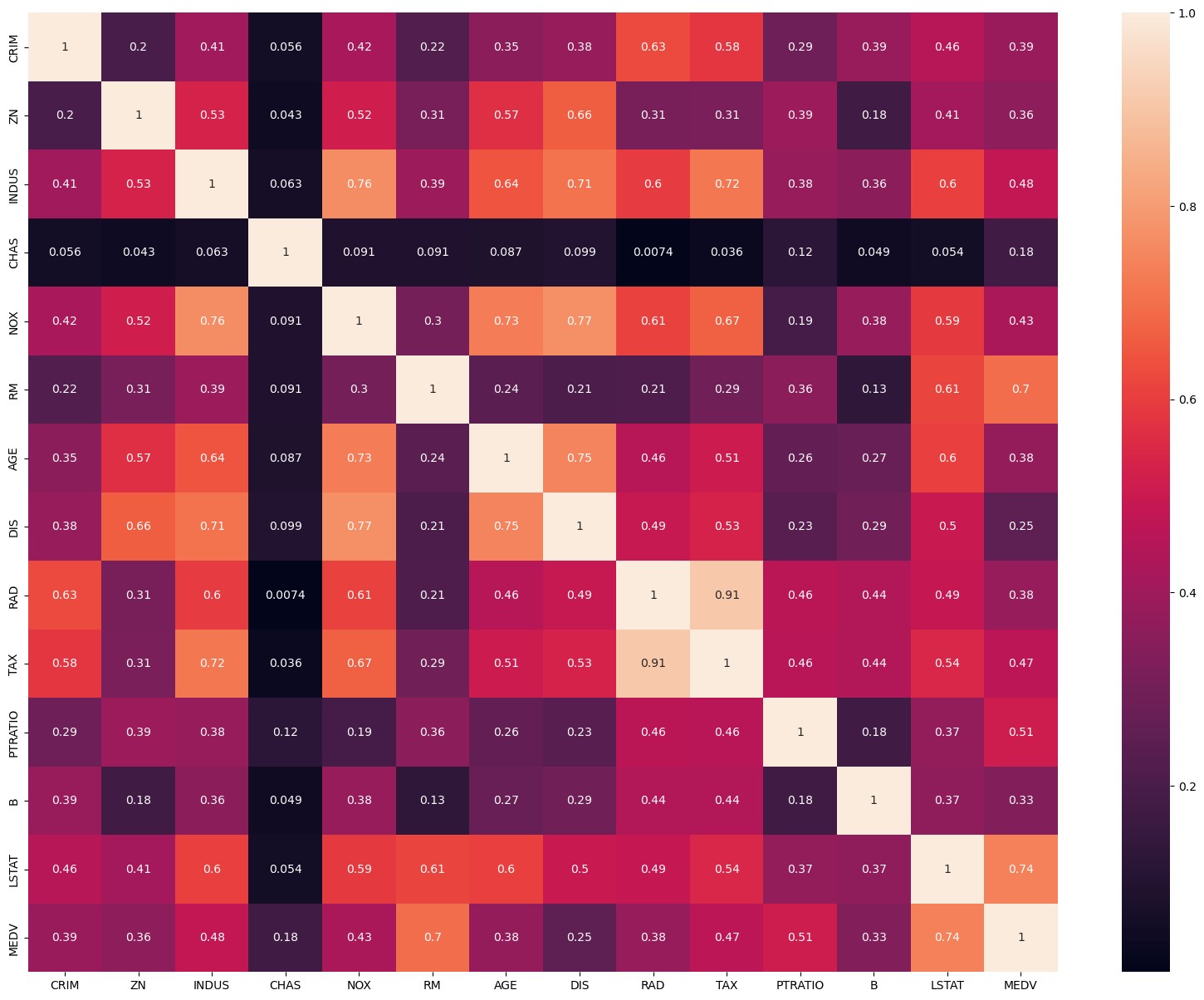
annot

**=**

**True**

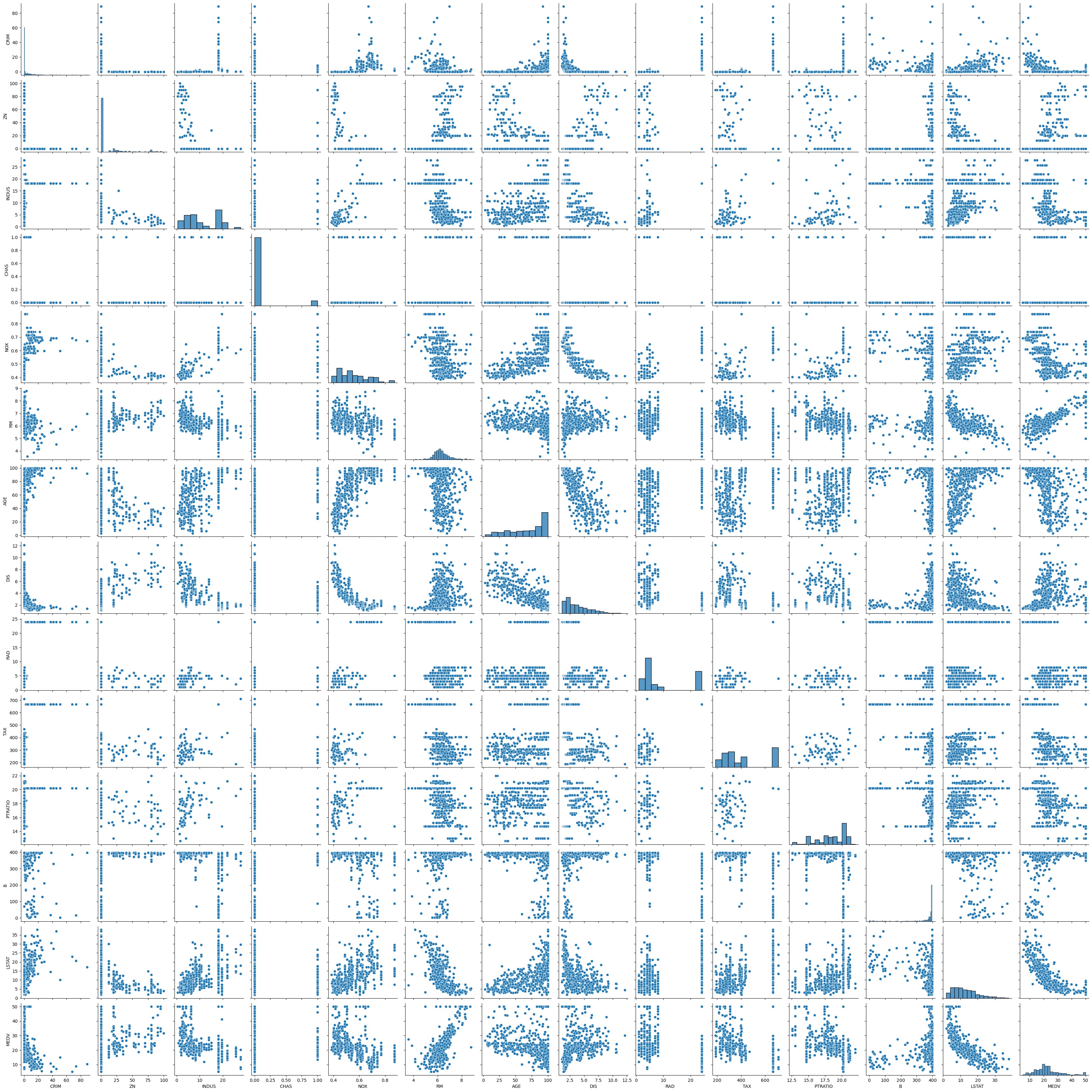
)

Out[34]: <Axes: >



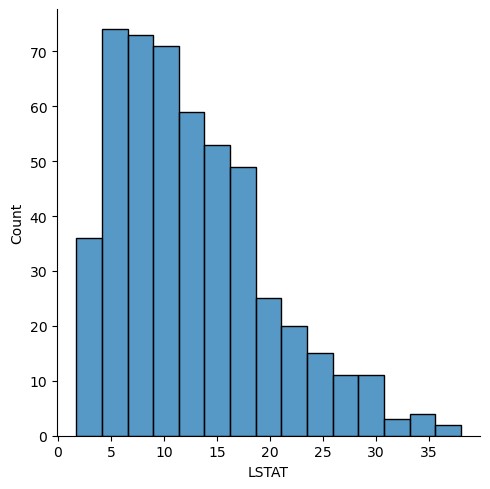
[35]: sns.pairplot(house\_data)

Out[35]: <seaborn.axisgrid.PairGrid at 0x289de2236d0>



[36]: sns.displot(house\_data['LSTAT'])

Out[36]: <seaborn.axisgrid.FacetGrid at 0x289ea07a230>



[37]: sns.distplot(house\_data['LSTAT'])

C:\Users\User\AppData\Local\Temp\ipykernel\_1800\3880656448.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see [https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (https://gis](https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

[t.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)](https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)  sns.distplot(house\_data['LSTAT'])

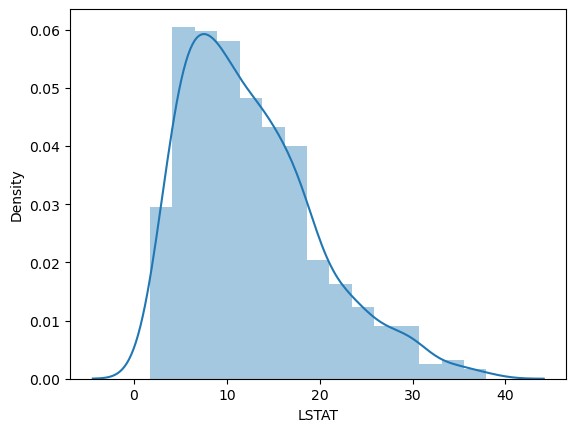
Out[37]: <Axes: xlabel='LSTAT', ylabel='Density'>

In

[39]:

In

[41]:



X

**=**

house\_data

.

drop

(

'MEDV'

,

axis

**=**

1

)

y

**=**

house\_data

[

'MEDV'

]

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [42]: Out[42]: | X | **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** | **B** | **LS** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 0.00632 | 18.0 | 2.31 | 0 | 0.538 | 6.575 | 65.2 | 4.0900 | 1 | 296.0 | 15.3 | 396.90 |
| **1** | 0.02731 | 0.0 | 7.07 | 0 | 0.469 | 6.421 | 78.9 | 4.9671 | 2 | 242.0 | 17.8 | 396.90 |
| **2** | 0.02729 | 0.0 | 7.07 | 0 | 0.469 | 7.185 | 61.1 | 4.9671 | 2 | 242.0 | 17.8 | 392.83 |
| **3** | 0.03237 | 0.0 | 2.18 | 0 | 0.458 | 6.998 | 45.8 | 6.0622 | 3 | 222.0 | 18.7 | 394.63 |
| **4** | 0.06905 | 0.0 | 2.18 | 0 | 0.458 | 7.147 | 54.2 | 6.0622 | 3 | 222.0 | 18.7 | 396.90 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **501** | 0.06263 | 0.0 | 11.93 | 0 | 0.573 | 6.593 | 69.1 | 2.4786 | 1 | 273.0 | 21.0 | 391.99 |
| **502** | 0.04527 | 0.0 | 11.93 | 0 | 0.573 | 6.120 | 76.7 | 2.2875 | 1 | 273.0 | 21.0 | 396.90 |
| **503** | 0.06076 | 0.0 | 11.93 | 0 | 0.573 | 6.976 | 91.0 | 2.1675 | 1 | 273.0 | 21.0 | 396.90 |
| **504** | 0.10959 | 0.0 | 11.93 | 0 | 0.573 | 6.794 | 89.3 | 2.3889 | 1 | 273.0 | 21.0 | 393.45 |
| **505** | 0.04741 | 0.0 | 11.93 | 0 | 0.573 | 6.030 | 80.8 | 2.5050 | 1 | 273.0 | 21.0 | 396.90 |

In [43]: y

Out[43]: 0 24.0 1 21.6

1. 34.7
2. 33.4
3. 36.2 ... 501 22.4
4. 20.6
5. 23.9
6. 22.0
7. 11.9

In

[44]:

In

[45]:

In

[46]:

Name: MEDV, Length: 506, dtype: float64

**from**

sklearn

.

model\_selection

**import**

train\_test\_split

X\_train

,

X\_test

,

y\_train

,

y\_test

**=**

train\_test\_split

(

X

,

y

,

test\_size

**=**

0.3

,

random\_stat

**from**

sklearn

.

linear\_model

**import**

LinearRegression

model

**=**

LinearRegression

()

[47]: model.fit(X\_train,y\_train)

Out[47]: LinearRegression()

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.

On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In

[48]:

predictions

**=**

model

.

predict

(

X\_test

)

# **Model Evaluation**

In

[49]:

print

(

model

.

intercept\_

)

31.63108403569186

In

[50]:

print

(

model

.

coef\_

)

[-1.33470103e-01 3.58089136e-02 4.95226452e-02 3.11983512e+00 -1.54170609e+01 4.05719923e+00 -1.08208352e-02 -1.38599824e+00 2.42727340e-01 -8.70223437e-03 -9.10685208e-01 1.17941159e-02 -5.47113313e-01]

In [52]: coeff

In

[51]:

Out[52]:

**Coefficient**

**CRIM**

-0.133470

coeff

**=**

pd

.

DataFrame

(

model

.

coef\_

,

X

.

columns

,

columns

**=**

[

"Coefficient"

])

**ZN** 0.035809

**INDUS** 0.049523

**CHAS** 3.119835

**NOX** -15.417061

**RM** 4.057199

**AGE** -0.010821

**DIS** -1.385998

**RAD** 0.242727

**TAX** -0.008702

**PTRATIO** -0.910685

**B** 0.011794

**LSTAT** -0.547113

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [53]: house\_data.head()  Out[53]: **CRIM ZN INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** | **B** | **LSTA** |

**0** 0.00632 18.0 2.31 0 0.538 6.575 65.2 4.0900 1 296.0 15.3 396.90 4.9 **1** 0.02731 0.0 7.07 0 0.469 6.421 78.9 4.9671 2 242.0 17.8 396.90 9.

1. 0.02729 0.0 7.07 0 0.469 7.185 61.1 4.9671 2 242.0 17.8 392.83 4.0
2. 0.03237 0.0 2.18 0 0.458 6.998 45.8 6.0622 3 222.0 18.7 394.63 2.9

In

[55]:

**4**

0.06905

0.0

2.18

0

0.458

7.147

54.2

6.0622

3

222.0

18.7

396.90

5.3

model

.

predict

([[

0.00632

,

18.0

,

2.31

,

0

,

0.538

,

6.575

,

65.2

,

4.0900

,

1

,

296.0

,

15.3

,

396.9

C:\Users\User\anaconda3\lib\site-packages\sklearn\base.py:420: UserWarning: X does not have valid feature names, but LinearRegression was fitted with featu re names

warnings.warn(

Out[55]: array([30.08649576])

In [56]: model.predict([[0.06905,0.0,2.18,0,0.458,7.147,54.2,6.0622,3,222.0,18.7,396.90

C:\Users\User\anaconda3\lib\site-packages\sklearn\base.py:420: UserWarning: X does not have valid feature names, but LinearRegression was fitted with featu re names

warnings.warn(

Out[56]: array([28.20837173])

In [57]: model.predict([[0.06905,0.0,2.18,0,0.458,7.147,54.2,6.0622,3,222.0,18.7,396.90

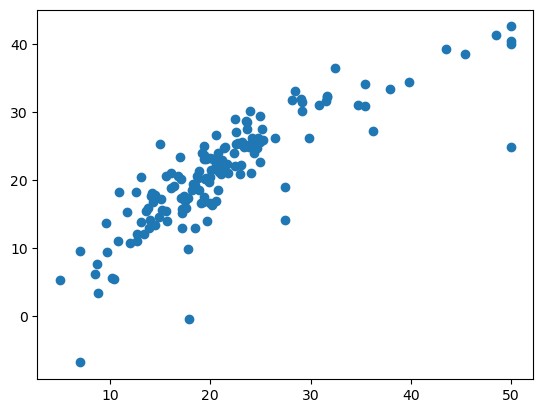
C:\Users\User\anaconda3\lib\site-packages\sklearn\base.py:420: UserWarning: X does not have valid feature names, but LinearRegression was fitted with featu re names

warnings.warn(

Out[57]: array([28.20837173])

[59]: plt.scatter(y\_test,predictions)

Out[59]: <matplotlib.collections.PathCollection at 0x289f48bb0d0>



[73]: sns.distplot((y\_test**-**predictions),kde **=** **True**,bins **=** 20)

C:\Users\User\AppData\Local\Temp\ipykernel\_1800\248592448.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see [https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751 (https://gis](https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

[t.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)](https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)  sns.distplot((y\_test-predictions),kde = True,bins = 20)

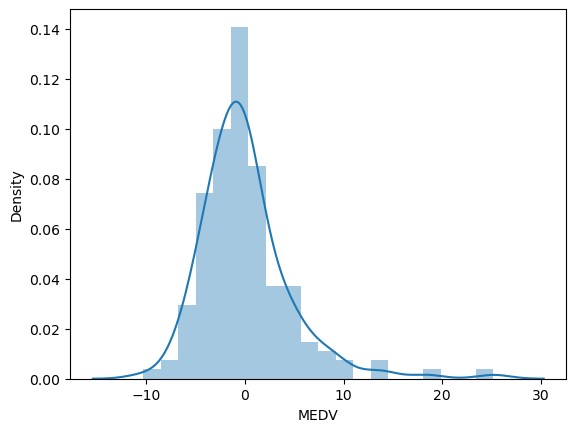
Out[73]: <Axes: xlabel='MEDV', ylabel='Density'>

In

[60]:

In

[72]:



MAE : 3.1627098714574253

MSE : 21.517444231177432

**from**

sklearn

**import**

metrics

print

(

'MAE :'

,

metrics

.

mean\_absolute\_error

(

y\_test

,

predictions

))

print

(

'MSE :'

,

metrics

.

mean\_squared\_error

(

y\_test

,

predictions

))

print

(

'RMSE: '

,

np

.

sqrt

(

metrics

.

mean\_squared\_error

(

y\_test

,

predictions

)))

RMSE: 4.63868