Title : Create a Linear Regression Model using Python/R to predict home prices using Boston Housing Dataset (https://www.kaggle.com/c/boston-housing). The Boston Housing dataset contains information about various houses in Boston through different parameters. There are 506 samples and 14 feature variables in this dataset. The objective is to predict the value of prices of the house using the given features

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

ad=pd.read\_csv('/home/ubuntu/DSBDA/Boston.csv')

print(ad)

RM

LSTAT

PTRATIO

MEDV

0

6.575

4.98

15.3

504000.0

1

6.421

9.14

17.8

453600.0

2

7.185

4.03

17.8

728700.0

3

6.998

2.94

18.7

701400.0

4

7.147

5.33

18.7

760200.0

..

…

…

…

…

484

6.593

9.67

21.0

470400.0

485

6.120

9.08

21.0

432600.0

486

6.976

5.64

21.0

501900.0

487

6.794

6.48

21.0

462000.0

488

6.030

7.88

21.0

249900.0

[489 rows x 4 columns]

print(ad.info())

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

RangeIndex: 489 entries, 0 to 488

Data columns (total 4 columns):

#

Column

Non-Null Count

Dtype

---

------

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

-----

0

RM

489 non-null

float64

1

LSTAT

489 non-null

float64

2

PTRATIO

489 non-null

float64

3

MEDV

489 non-null

float64

dtypes: float64(4)

memory usage: 15.4 KB

None

print(ad.keys())

Index(['RM', 'LSTAT', 'PTRATIO', 'MEDV'], dtype='object')

print(ad.head())

RM

LSTAT

PTRATIO

MEDV

0

6.575

4.98

15.3

504000.0

1

6.421

9.14

17.8

453600.0

2

7.185

4.03

17.8

728700.0

3

6.998

2.94

18.7

701400.0

4

7.147

5.33

18.7

760200.0

.

print(ad.tail())

RM

LSTAT

PTRATIO

MEDV

484

6.593

9.67

21.0

470400.0

485

6.120

9.08

21.0

432600.0

486

6.976

5.64

21.0

501900.0

487

6.794

6.48

21.0

462000.0

488

6.030

7.88

21.0

249900.0

print(ad.describe())

