

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('https://github.com/ybifoundation/Dataset/raw/main/Boston.csv')
```

```
df.head()
```

	CRIM	ZN	INDUS	CHAS	NX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
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	

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 506 entries, 0 to 505
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0   CRIM        506 non-null    float64
1   ZN          506 non-null    float64
2   INDUS       506 non-null    float64
3   CHAS        506 non-null    int64
4   NX          506 non-null    float64
5   RM          506 non-null    float64
6   AGE         506 non-null    float64
7   DIS         506 non-null    float64
8   RAD         506 non-null    int64
9   TAX         506 non-null    float64
10  PTRATIO     506 non-null    float64
11  B           506 non-null    float64
12  LSTAT       506 non-null    float64
13  MEDV       506 non-null    float64
dtypes: float64(12), int64(2)
memory usage: 55.5 KB
```

```
df.describe()
```

	CRIM	ZN	INDUS	CHAS	NX	RM	AGE
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000
mean	3.613524	11.363636	11.136779	0.069170	0.554695	6.284634	68.574900
std	8.601545	23.322453	6.860353	0.253994	0.115878	0.702617	28.148800
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000
25%	0.082045	0.000000	5.190000	0.000000	0.449000	5.885500	45.025000
50%	0.256510	0.000000	9.690000	0.000000	0.538000	6.208500	77.500000
75%	0.677000	10.500000	10.400000	0.000000	0.604000	6.600500	84.075000

```
df.isna().sum()
```

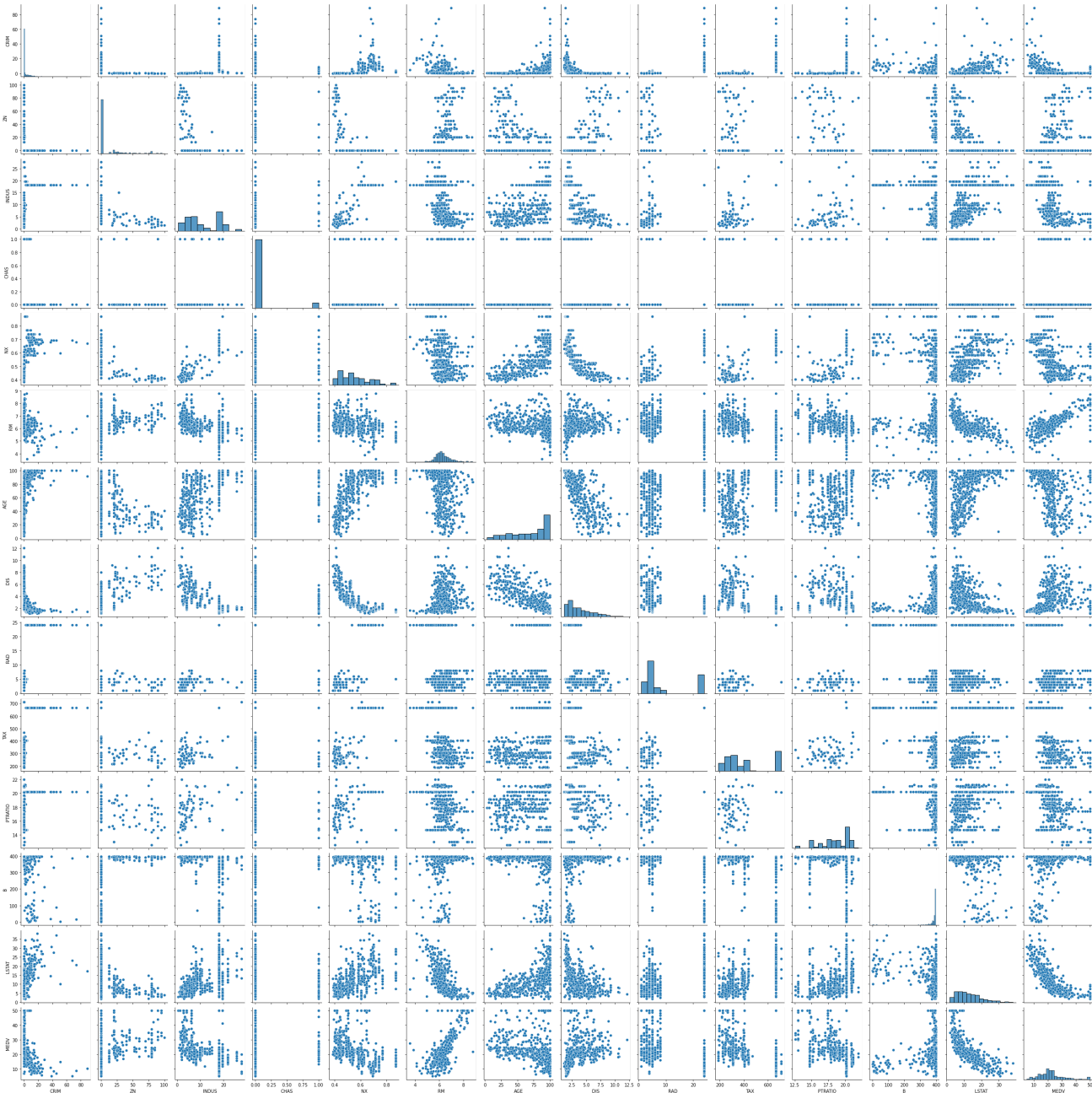
```
CRIM      0
ZN        0
INDUS     0
CHAS      0
NX        0
RM        0
AGE       0
DIS       0
RAD       0
TAX       0
PTRATIO   0
B         0
LSTAT     0
MEDV     0
dtype: int64
```

```
df.nunique()
```

```
CRIM      504
ZN        26
INDUS     76
CHAS       2
NX        81
RM       446
AGE       356
DIS       412
RAD        9
TAX       66
PTRATIO   46
B        357
LSTAT    455
MEDV     229
dtype: int64
```

```
sns.pairplot(df)
```

<seaborn.axisgrid.PairGrid at 0x7f5861013390>



```
df.columns
```

```
Index(['CRIM', 'ZN', 'INDUS', 'CHAS', 'NX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX',  
      'PTRATIO', 'B', 'LSTAT', 'MEDV'],  
      dtype='object')
```

```
y = df['MEDV']
```

```
X = df[['CRIM', 'ZN', 'INDUS', 'CHAS', 'NX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX',  
      'PTRATIO', 'B', 'LSTAT']]
```

```
from sklearn.model_selection import train_test_split  
X_train, X_test, y_train, y_test = train_test_split(  
    X, y, test_size=.20)
```

```
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

```
((404, 13), (102, 13), (404,), (102,))
```

```
from sklearn.linear_model import LinearRegression  
model = LinearRegression()
```

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

```
LinearRegression()
```

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

```
from sklearn.metrics import mean_absolute_error, mean_absolute_percentage_error, mean_squa
```

```
mean_absolute_error(y_test, y_pred)
```

```
3.2635674962986227
```

```
mean_absolute_percentage_error(y_test, y_pred)
```

```
0.15620169046127025
```

```
mean_squared_error(y_test, y_pred)
```

```
23.607008294613323
```

```
sample = df.sample()  
sample
```

CRTM

ZN

INDUS

CHAS

NX

RM

AGE

DIS

RAD

TAX

PTRATIO

B

X\_new = sample.loc[:,X.columns]

X\_new

	CRIM	ZN	INDUS	CHAS	NX	RM	AGE	DIS	RAD	TAX	PTRATIO	B
345	0.03113	0.0	4.39	0	0.442	6.014	48.5	8.0136	3	352.0	18.8	385.64

```
model.predict(X_new)

array([16.07004304])
```

[Colab paid products](#) - [Cancel contracts here](#)