```
1  import numpy as np
2  import pandas as pd
3  from pandas import Series,DataFrame
4  #import matplotlib.pyplot as plt
5  import seaborn as sns
6  from sklearn.datasets import load_boston
7
8  sns.set_style('whitegrid')
9  #%matplotlib inline
```

```
1  boston = load_boston()
```

```
1  print(boston.DESCR)
```

```
.. _boston_dataset:

Boston house prices dataset
---------------------------

**Data Set Characteristics:**

    :Number of Instances: 506

    :Number of Attributes: 13 numeric/categorical predictive. Median Value (attribute 14)

    :Attribute Information (in order):
        - CRIM     per capita crime rate by town
        - ZN       proportion of residential land zoned for lots over 25,000 sq.ft.
        - INDUS    proportion of non-retail business acres per town
        - CHAS     Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
        - NOX      nitric oxides concentration (parts per 10 million)
        - RM       average number of rooms per dwelling
        - AGE      proportion of owner-occupied units built prior to 1940
        - DIS      weighted distances to five Boston employment centres
        - RAD      index of accessibility to radial highways
        - TAX      full-value property-tax rate per $10,000
        - PTRATIO  pupil-teacher ratio by town
        - B        1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
        - LSTAT    % lower status of the population
        - MEDV     Median value of owner-occupied homes in $1000's

    :Missing Attribute Values: None

    :Creator: Harrison, D. and Rubinfeld, D.L.

This is a copy of UCI ML housing dataset.
https://archive.ics.uci.edu/ml/machine-learning-databases/housing/


This dataset was taken from the StatLib library which is maintained at Carnegie Mellon Ur

The Boston house-price data of Harrison, D. and Rubinfeld, D.L. 'Hedonic
prices and the demand for clean air', J. Environ. Economics & Management,
```

```
    vol.5, 81-102, 1978.   Used in Belsley, Kuh & Welsch, 'Regression diagnostics
    ...', Wiley, 1980.   N.B. Various transformations are used in the table on
    pages 244-261 of the latter.

    The Boston house-price data has been used in many machine learning papers that address re
    problems.

    .. topic:: References

       - Belsley, Kuh & Welsch, 'Regression diagnostics: Identifying Influential Data and Sou
       - Quinlan,R. (1993). Combining Instance-Based and Model-Based Learning. In Proceedings
```

```
1   df = pd.DataFrame(boston.data)
2   df.columns = boston.feature_names
3   df['PRICE'] = boston.target
```

```
1   df.head()
2   df.PRICE.describe()
```

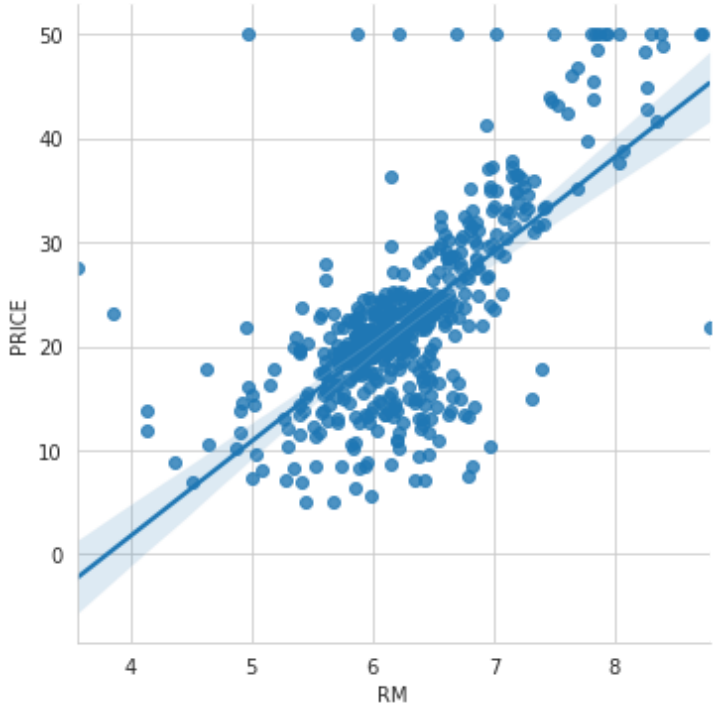
```
    count    506.000000
    mean      22.532806
    std        9.197104
    min        5.000000
    25%       17.025000
    50%       21.200000
    75%       25.000000
    max       50.000000
    Name: PRICE, dtype: float64
```

```
1   #sns.distplot(df.PRICE, bins=50, kde=False)
2   sns.distplot(df.RM, bins=10, kde=False)
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/distributions.py:2557: FutureWarning: `dis
```

```
1  sns.lmplot('RM', 'PRICE', data=df)
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the
  FutureWarning
<seaborn.axisgrid.FacetGrid at 0x7f33d7efa748>
```



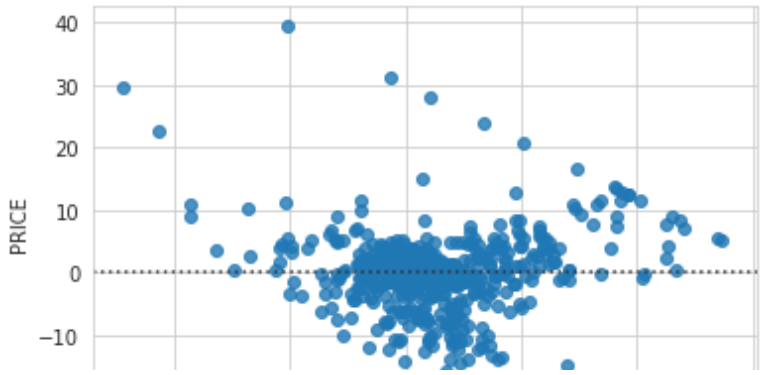

```
1  #X = np.vstack(df.RM)
2  x = df.RM
3  X = np.vstack([x, np.ones(len(x))]).T # T attribute in numpy is a transpose?
4  X
```

```
array([[6.575, 1.   ],
       [6.421, 1.   ],
       [7.185, 1.   ],
       ...,
       [6.976, 1.   ],
       [6.794, 1.   ],
       [6.03 , 1.   ]])
```

```
1  y = df.PRICE
```

```
1  sns.residplot(x, y)
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f33d7efad68>
```




```
1  resid = np.linalg.lstsq(X, y)[1]
2  r2 = 1 - resid / (y.size * y.var())
3  r2[0]
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: FutureWarning: `rcond` pa
To use the future default and silence this warning we advise to pass `rcond=None`, to kee
  """Entry point for launching an IPython kernel.
0.4845461566711933
```

```
1  m, b = np.linalg.lstsq(X, y)[0]
2  m, b
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: FutureWarning: `rcond` pa
To use the future default and silence this warning we advise to pass `rcond=None`, to kee
  """Entry point for launching an IPython kernel.
(9.102108981180313, -34.67062077643857)
```

```
1  sns.scatterplot(x, y)
2  sns.lineplot(x, m*x + b, color='r')
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the
  FutureWarning
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f33cf368f98>
```