Import Library

```
In [2]:  import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt
```

Panggil dataset ke var df

```
In [8]:  df = pd.read_csv('HousingData.csv')
```

Show head untuk df

```
In [10]:  df.head()
```

Out[10]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90

Hitung sum kolom yang null

```
In [12]:  df.isnull().sum()
```

```
Out[12]:  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
```

Mofidy kolom dengan nama baru

```
In [14]:  new_columns_dist = {
              'CRIM' : 'rate_of_crime',
              'ZN' : 'residential_zone_pct',
              'INDUS' : 'business_zone_pct',
```

```
        'CHAS' : 'borders_river',
        'NOX' : 'oxide_concentration',
        'RM' : 'average_rooms',
        'AGE' : 'owner_occup_pct',
        'DIS' : 'dist_to_work',
        'RAD' : 'access_to_highway',
        'TAX' : 'property_tax',
        'PTRATIO' : 'student_teacher_ratio',
        'LSTAT' : 'pct_underclass',
        'MEDV' : 'home_median_value'
    }
df.rename(columns=new_columns_dist, inplace=True)
df.head()
```

Out[14]:

	rate_of_crime	residential_zone_pct	business_zone_pct	borders_river	oxide_concentration
0	0.00632	18.0	2.31	0.0	0.538
1	0.02731	0.0	7.07	0.0	0.469
2	0.02729	0.0	7.07	0.0	0.469
3	0.03237	0.0	2.18	0.0	0.458
4	0.06905	0.0	2.18	0.0	0.458

Describe statistic dengan df

In [16]: `df.describe()`

Out[16]:

|	rate_of_crime	residential_zone_pct	business_zone_pct	borders_river	oxide_concentr
count	486.000000	486.000000	486.000000	486.000000	506.00
mean	3.611874	11.211934	11.083992	0.069959	0.55
std	8.720192	23.388876	6.835896	0.255340	0.11
min	0.006320	0.000000	0.460000	0.000000	0.38
25%	0.081900	0.000000	5.190000	0.000000	0.44
50%	0.253715	0.000000	9.690000	0.000000	0.53
75%	3.560263	12.500000	18.100000	0.000000	0.62
max	88.976200	100.000000	27.740000	1.000000	0.87

Membuat boxplot dengan data rate of crime

In [18]: `sns.boxplot(df['rate_of_crime'])`

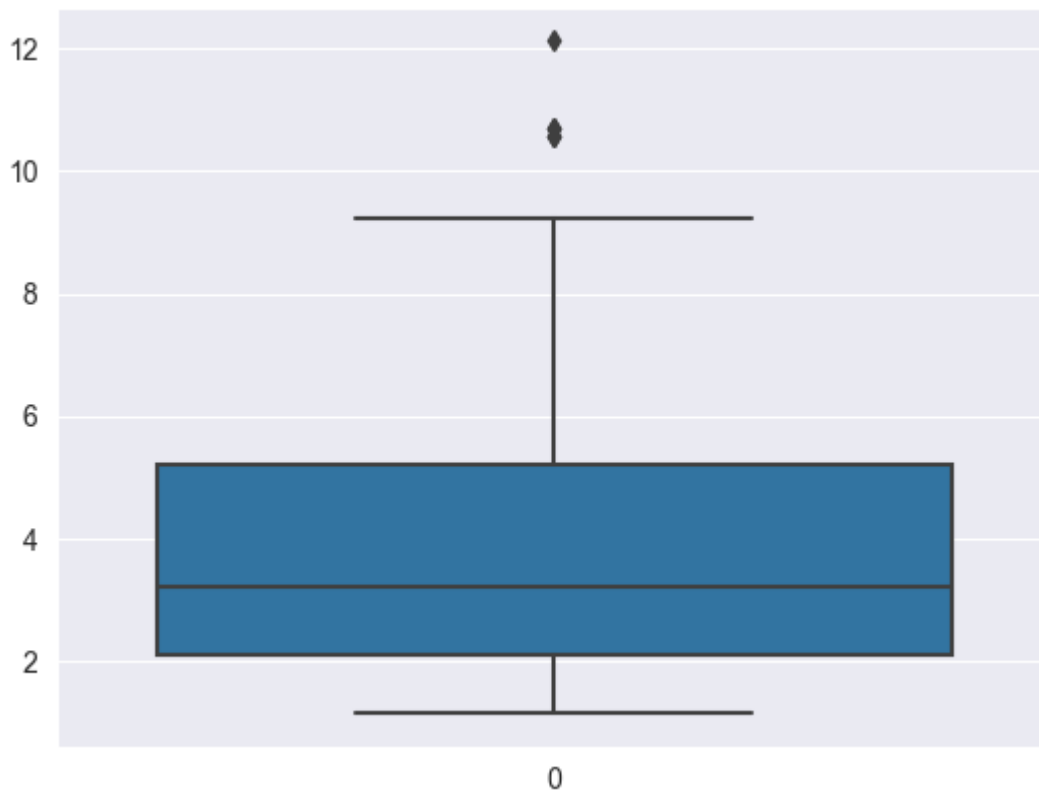Out[18]: `<AxesSubplot:>`
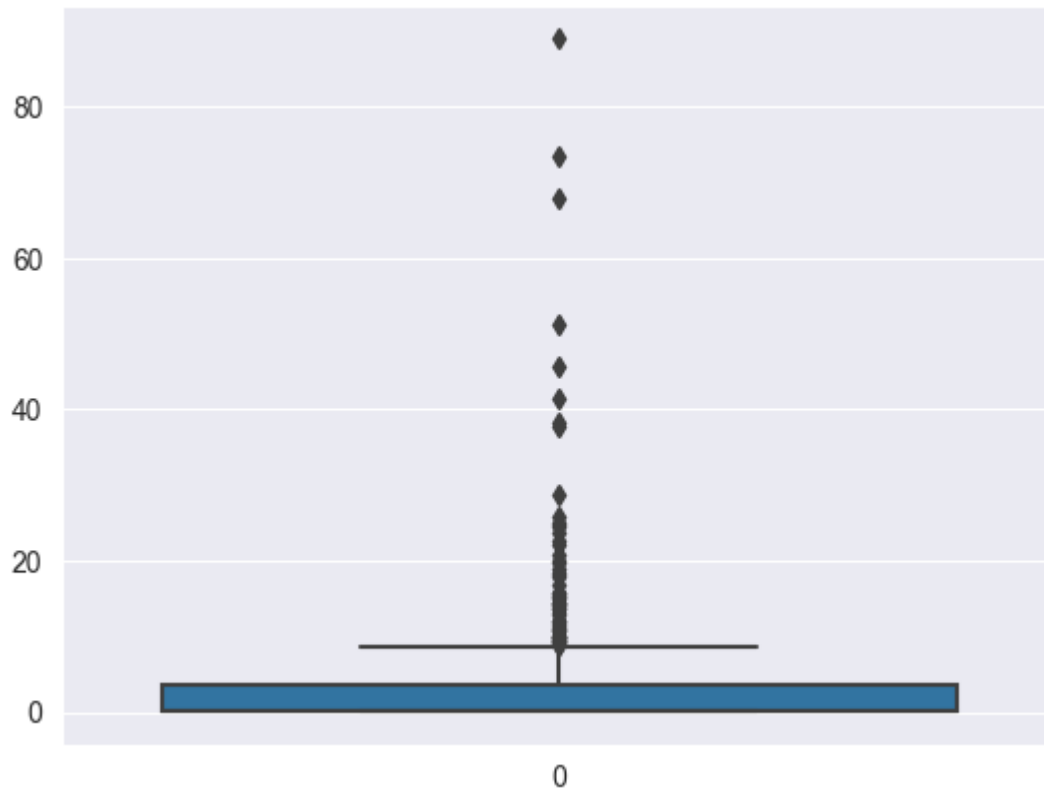
Membuat boxplot dengan data dist to work

```
In [20]: sns.boxplot(df['dist_to_work'])
```
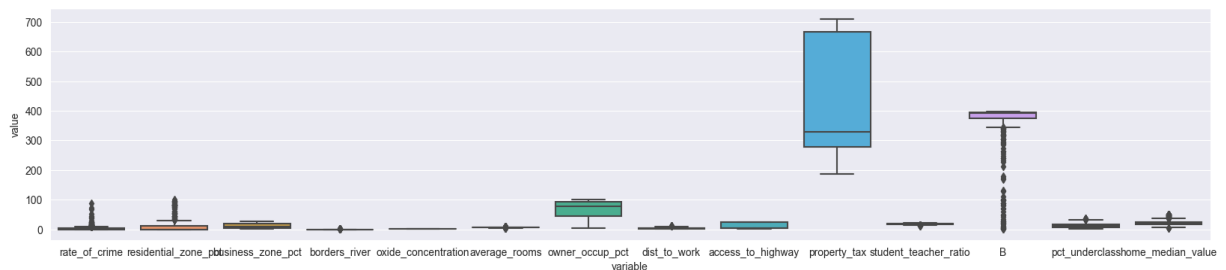
Out[20]: <AxesSubplot:>

Membuat figure 20 x 4 Membuat boxplot dengan x dan y itu

```
In [22]:  plt.figure(figsize=(20,4))
          sns.boxplot(x='variable', y='value', data=pd.melt(df))
```
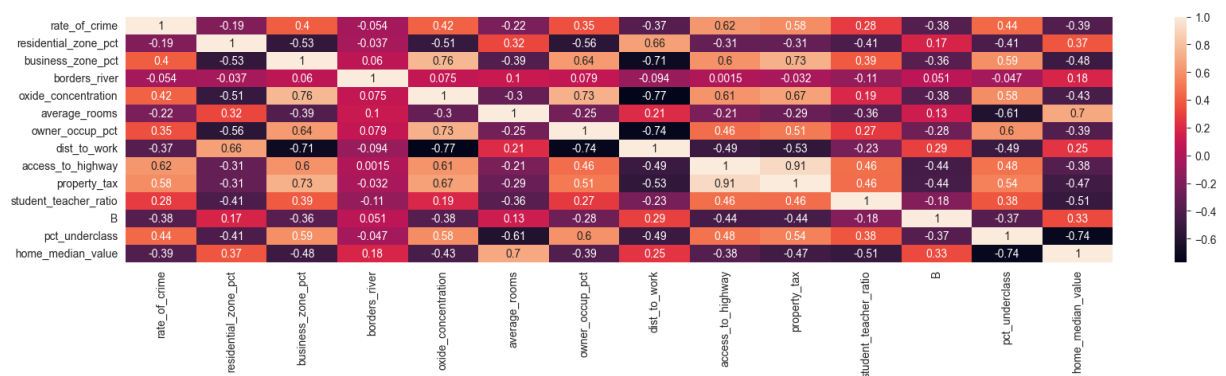
Out[22]: <AxesSubplot:xlabel='variable', ylabel='value'>

Membuat heatmap, dengan data hasil korelasi

```
In [25]:  plt.subplots(figsize=(20,4))
          sns.heatmap(df.corr(), annot=True)
```

Out[25]: <AxesSubplot:>

Menampilkan corelasi dari home_median_value secara asc

```
In [27]:  df.corr()['home_median_value'].sort_values()
```

Out[27]:
```
pct_underclass         -0.735822
student_teacher_ratio  -0.507787
business_zone_pct      -0.481772
property_tax           -0.468536
oxide_concentration    -0.427321
owner_occup_pct        -0.394656
rate_of_crime          -0.391363
access_to_highway      -0.381626
borders_river           0.181391
dist_to_work            0.249929
B                       0.333461
residential_zone_pct    0.373136
average_rooms           0.695360
home_median_value       1.000000
Name: home_median_value, dtype: float64
```

Membuat pairplot dari data df

In [28]: `sns.pairplot(df)`

Out[28]: `<seaborn.axisgrid.PairGrid at 0x29d7b47f130>`