Boston Housing Prices Project

In [ ]:

```
# Importing Libraries
```

In [1]:

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

In [2]:

```
df = pd.read_csv("housing.csv")
df
```

Out[2]:

	CRIM	ZN	INDUS	CHAS	NOX	RN	AGE	DIS	RED	TAX	PTRATIO	BK	LSTA
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.9
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.1
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.0
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.9
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.3
...	...	...	...	...	...	...	...	...	...	...	...	...	.
501	0.06263	0.0	11.93	0	0.573	6.593	69.1	2.4786	1	273	21.0	391.99	9.6
502	0.04527	0.0	11.93	0	0.573	6.120	76.7	2.2875	1	273	21.0	396.90	9.0
503	0.06076	0.0	11.93	0	0.573	6.976	91.0	2.1675	1	273	21.0	396.90	5.6
504	0.10959	0.0	11.93	0	0.573	6.794	89.3	2.3889	1	273	21.0	393.45	6.4
505	0.04741	0.0	11.93	0	0.573	6.030	80.8	2.5050	1	273	21.0	396.90	7.8

506 rows × 14 columns

In [7]:

```
df.head(15)
```

Out[7]:

	CRIM	ZN	INDUS	CHAS	NOX	RN	AGE	DIS	RED	TAX	PTRATIO	BK	LSTA
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.9
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.1
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.0
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.9
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.3
5	0.02985	0.0	2.18	0	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.2
6	0.08829	12.5	7.87	0	0.524	6.012	66.6	5.5605	5	311	15.2	395.60	12.4
7	0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505	5	311	15.2	396.90	19.1
8	0.21124	12.5	7.87	0	0.524	5.631	100.0	6.0821	5	311	15.2	386.63	29.9
9	0.17004	12.5	7.87	0	0.524	6.004	85.9	6.5921	5	311	15.2	386.71	17.1
10	0.22489	12.5	7.87	0	0.524	6.377	94.3	6.3467	5	311	15.2	392.52	20.4
11	0.11747	12.5	7.87	0	0.524	6.009	82.9	6.2267	5	311	15.2	396.90	13.2
12	0.09378	12.5	7.87	0	0.524	5.889	39.0	5.4509	5	311	15.2	390.50	15.7
13	0.62976	0.0	8.14	0	0.538	5.949	61.8	4.7075	4	307	21.0	396.90	8.2
14	0.63796	0.0	8.14	0	0.538	6.096	84.5	4.4619	4	307	21.0	380.02	10.2

In [4]:

```
df.isnull().sum()
```

Out[4]:

```
CRIM       0
ZN         0
INDUS      0
CHAS       0
NOX        0
RN         0
AGE        0
DIS        0
RED        0
TAX        0
PTRATIO    0
BK         0
LSTAT      0
MEDV       0
dtype: int64
```

In [5]:

```
df.dtypes
```

Out[5]:

```
CRIM       float64
ZN         float64
INDUS      float64
CHAS         int64
NOX        float64
RN         float64
AGE        float64
DIS        float64
RED          int64
TAX          int64
PTRATIO    float64
BK         float64
LSTAT      float64
MEDV       float64
dtype: object
```

In [6]:

```
df.shape
```

Out[6]:

```
(506, 14)
```

In [8]:

```
df.describe()
```

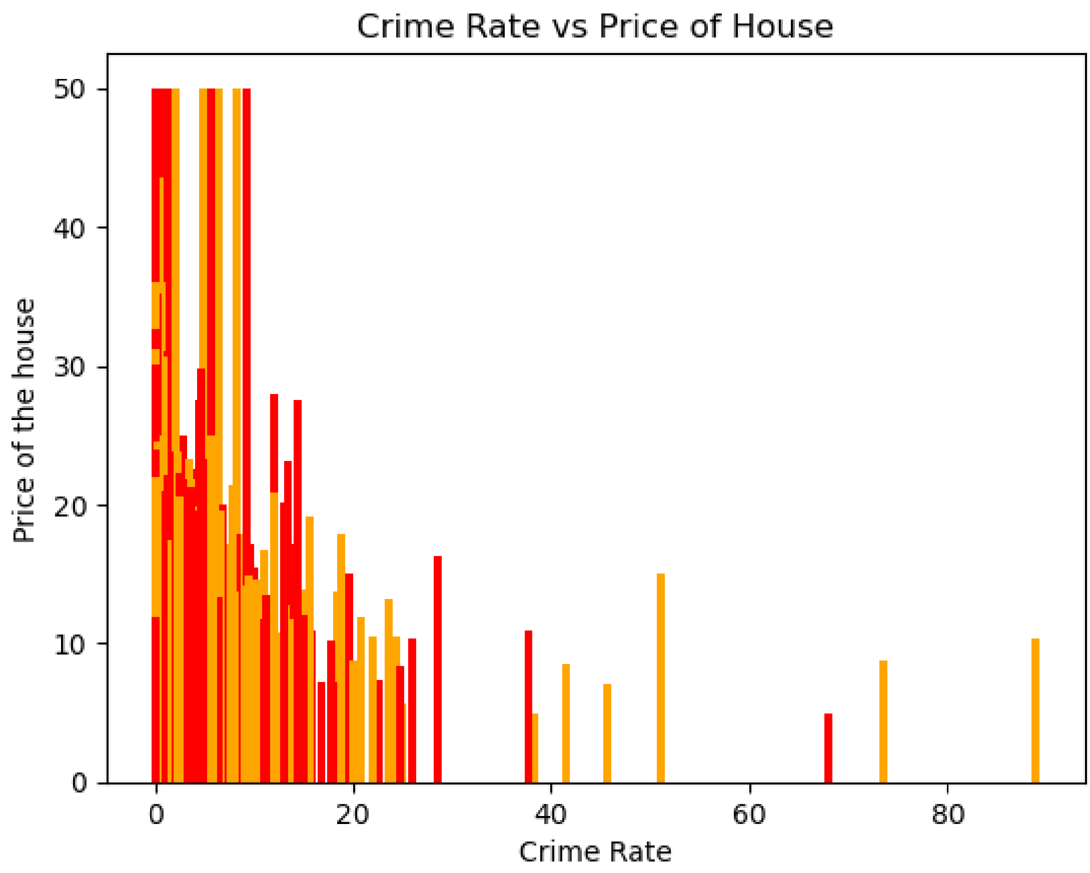
Out[8]:

M	ZN	INDUS	CHAS	NOX	RN	AGE	DIS	RED
00	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000
24	11.363636	11.136779	0.069170	0.554695	6.284634	68.574901	3.795043	9.549407
45	23.322453	6.860353	0.253994	0.115878	0.702617	28.148861	2.105710	8.707259
20	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000	1.129600	1.000000
45	0.000000	5.190000	0.000000	0.449000	5.885500	45.025000	2.100175	4.000000
10	0.000000	9.690000	0.000000	0.538000	6.208500	77.500000	3.207450	5.000000
33	12.500000	18.100000	0.000000	0.624000	6.623500	94.075000	5.188425	24.000000
00	100.000000	27.740000	1.000000	0.871000	8.780000	100.000000	12.126500	24.000000

when crime rate increase the Price of Houses is Down continuosly

In [9]:

```
plt.bar(df.CRIM, df.MEDV, color=['orange', 'red'])
plt.xlabel('Crime Rate')
plt.ylabel('Price of the house')
plt.title('Crime Rate vs Price of House')
plt.show()
```



observation

When the crime rate increases in a particular area, it can lead to a decrease in the desirability of that area, and as a result, the demand for housing in that area can decrease. When demand for housing decreases, sellers may have to lower their prices in order to attract buyers.

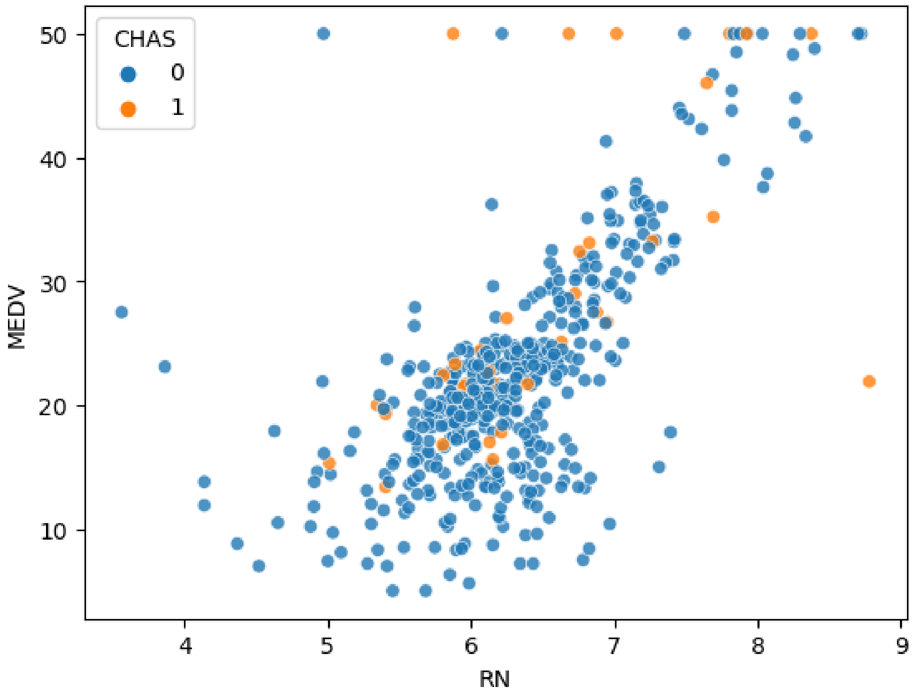
When lands are closer to Charles River the Price of the land will be increse and land are further away from the river the cost will be down

In [10]:

```
sns.scatterplot(y='MEDV', x='RN', data=df, alpha=0.8, hue='CHAS')
plt.suptitle('plots showing MEDV and Room Relatioship based on Charles River Dummy variable'

plt.show()
```

plots showing MEDV and Room Relatioship based on Charles River Dummy variable



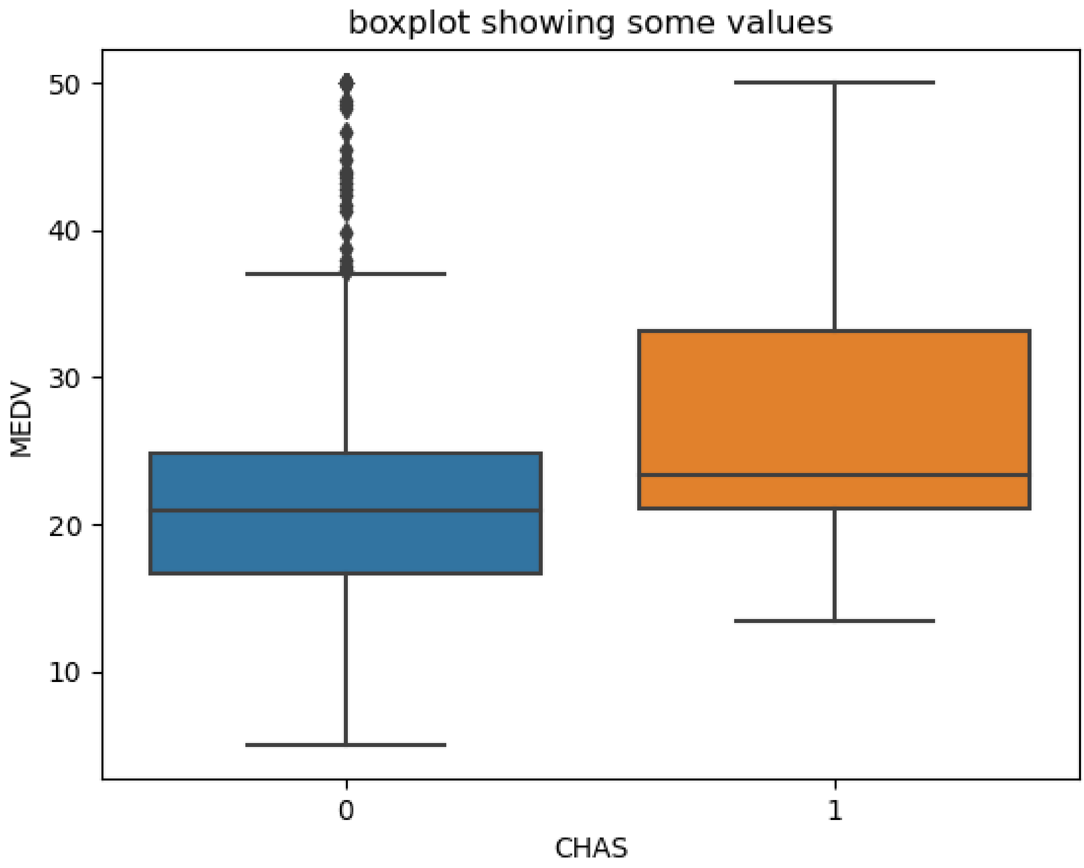
observation

In the Boston area. The land near the river is considered prime Location because of the view of the river and proximity to water-based activities in this land there are sufficent water for daily life activity. Therefore, properties close to the Charles River tend to be more more expensive as campared to land which is further away from Charles River

When lands are closer to Charles River the Price of the land will be increse and land are further away from the river the cost will be down using boxplot

In [6]:

```
sns.boxplot(x= 'CHAS',y='MEDV',data= df )
plt.title('boxplot showing some values' )
plt.show()
```



observation

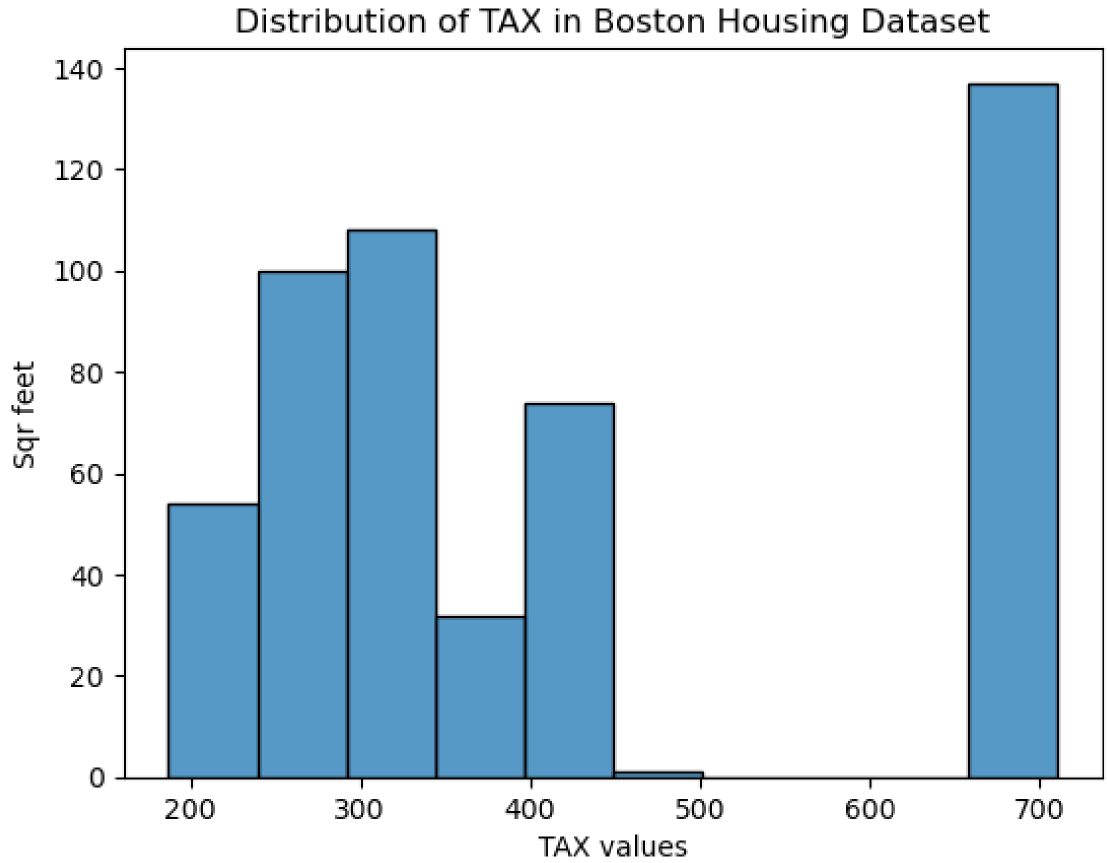
Price of houses closer to Charles River are found to be higher ingeneral with higher value of 1st Quantile Median and 3rd Quantile Even the min or max values are higher if the house is located in closer of Charles River

Number of Outlier are found to be unusely high in case of when the houses are further away from Charles River this could be affet the predictablity of pricing which is further away from the Charles River

Location wise Tax Distribution

In [4]:

```
sns.histplot(data=df, x='TAX')
plt.xlabel('TAX values')
plt.ylabel('Sqr feet')
plt.title('Distribution of TAX in Boston Housing Dataset')
plt.show()
```



observation

1 Location-wise tax distribution refers to the way in which property taxes are distributed across different geographic locations or regions. Property taxes are typically assessed by local governments

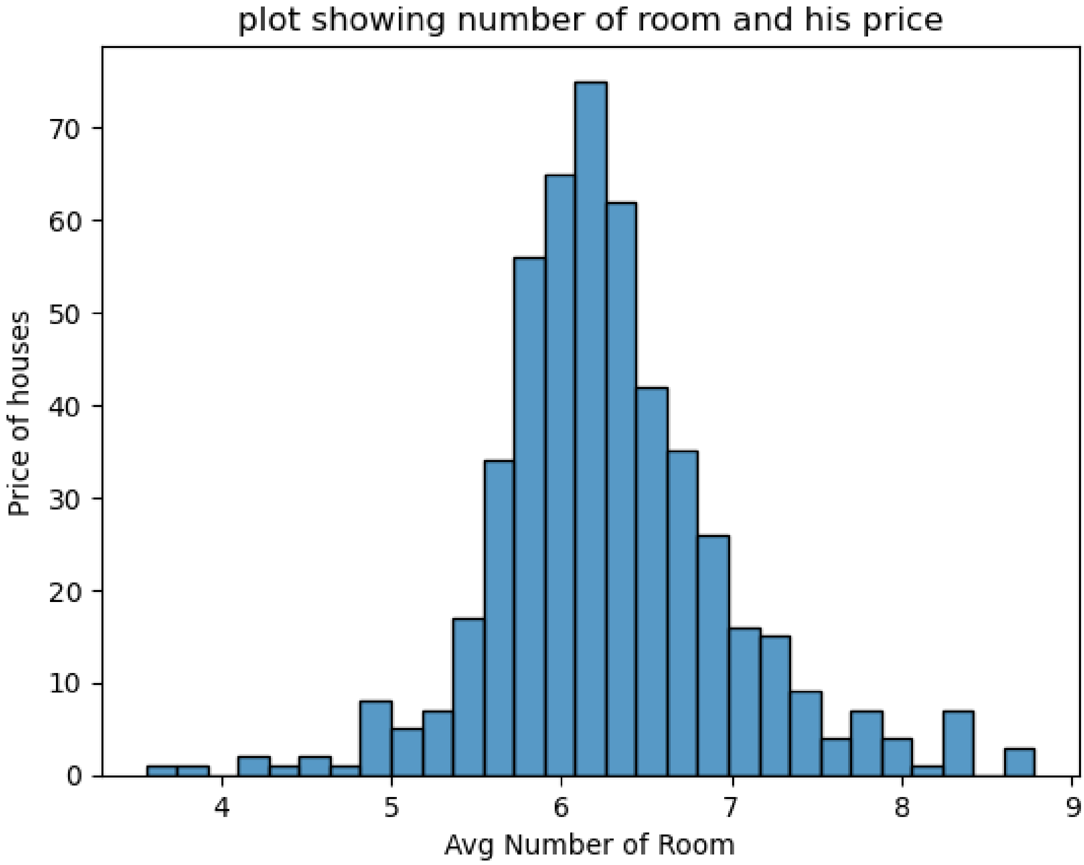
2 For example, areas with higher property values may have higher tax rates in order to generate more revenue for local governments.

3 Therefore Taxes are Totally depended on the Lcation

The Avg Number of Rooms are Incressing the price of the house is also incresing

In [13]:

```
sns.histplot(data=df, x='RN')
plt.xlabel('Avg Number of Room')
plt.ylabel('Price of houses')
plt.title('plot showing number of room and his price')
plt.show()
```



observation

1 there is a positive correlation between the number of rooms in a house and its price.

2 the fact that larger houses with more rooms generally have more square footage, which can increase the overall value of the property.

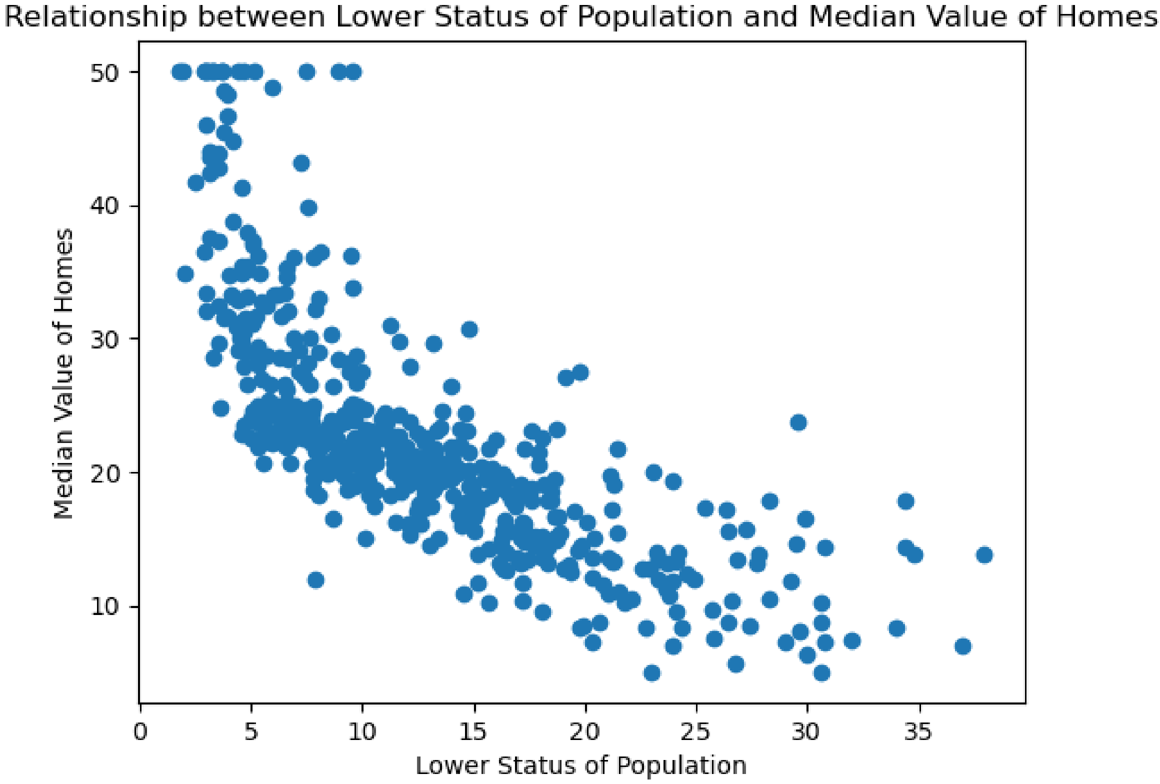
The median value of homes tends to decrease as the lower status of the population increases

In [13]:

```
plt.scatter(df[ 'LSTAT'], df['MEDV'])

plt.xlabel ('Lower Status of Population')
plt.ylabel('Median Value of Homes')
plt.title('Relationship between Lower Status of Population and Median Value of Homes')

plt.show()
```



observation

1 The column LSTAT represents the percentage of lower status of the population in a given area, while MEDV represents the median value of owner-occupied homes in thousands of dollars.

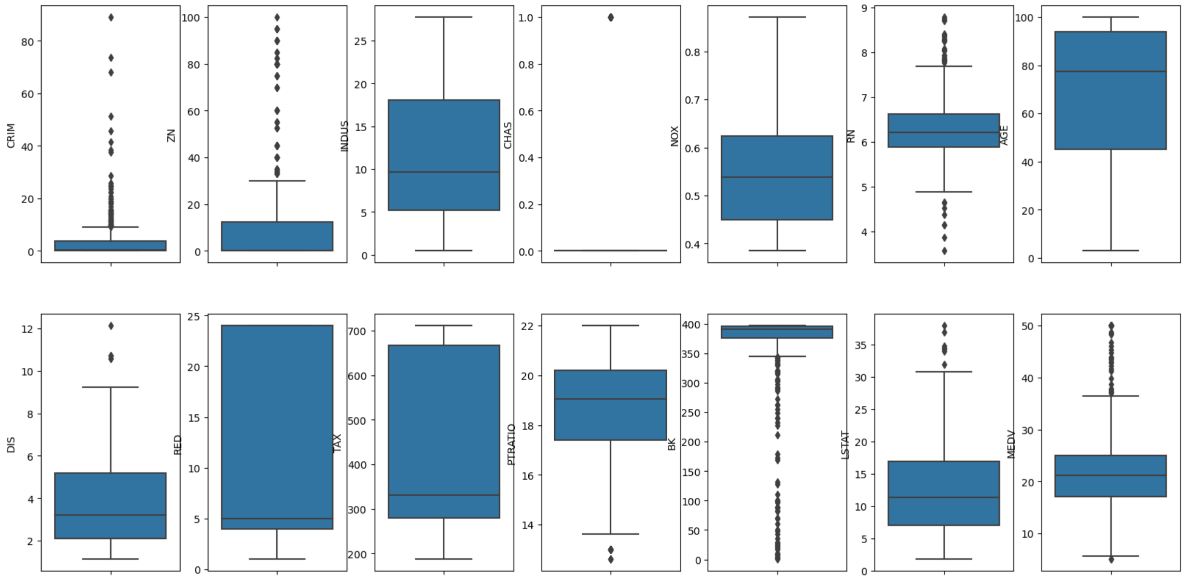
2 There is a negative correlation between LSTAT and MEDV.

3 This negative correlation between LSTAT and MEDV can be explained by a variety of factors. Areas with a higher percentage of lower status population may have lower income levels, which can make it more difficult for residents to afford higher- priced homes. Additionally, areas with higher poverty rates may also have higher crime rates, worse school systems, and less desirable living conditions, all of which can negatively impact home values.

Checking Outliers

In [14]:

```
fig, ax = plt.subplots(ncols=7, nrows=2, figsize=(20,10))
index = 0
ax = ax.flatten()
for col in df.items():
    sns.boxplot(y=col[0], data=df, ax=ax[index])
    index += 1
```



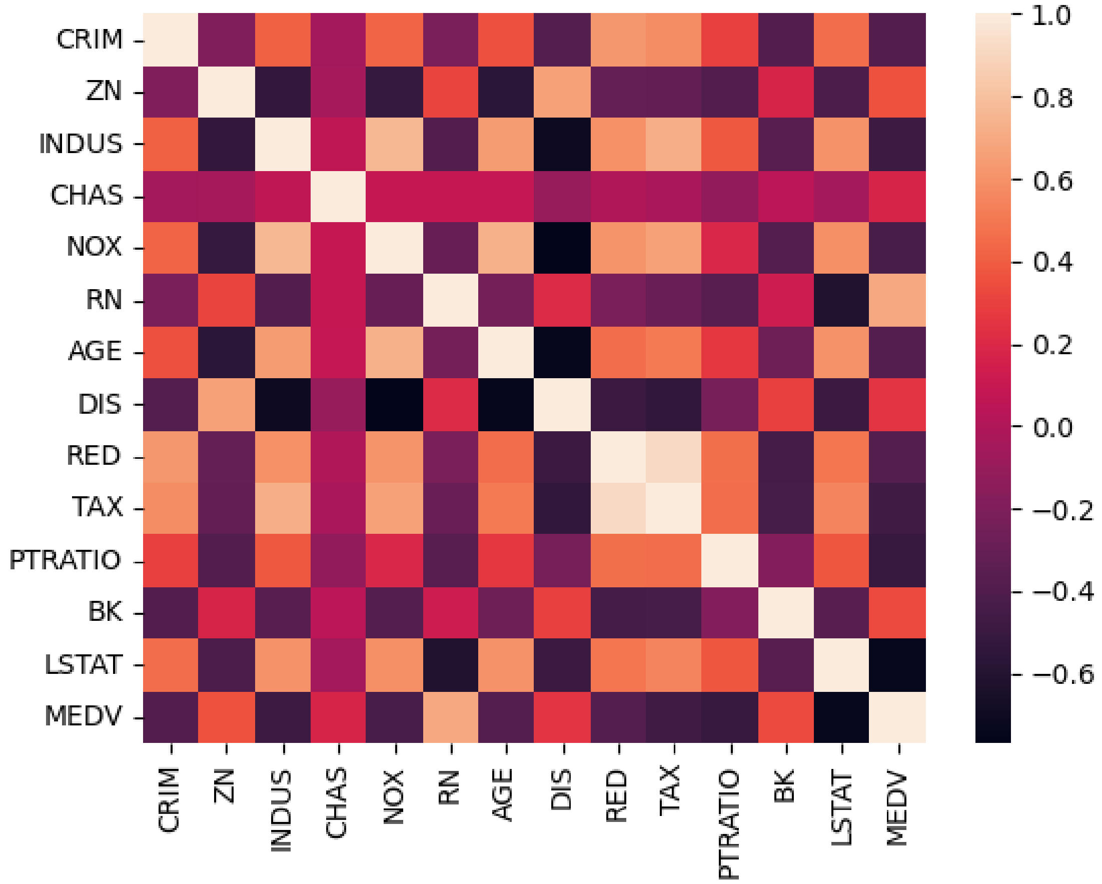
Checking Correation

In [15]:

```
sns.heatmap(df.corr())
```

Out[15]:

```
<Axes: >
```



Conclusion

1 There is a negative correlation between the crime rate and the median value of homes. This means that as the crime rate increases, the median value of homes tends to decrease.

2 Homes that are located closer to the Charles River tend to have a higher median value than those located further away.

3 There is a negative correlation between the median value of homes and the percentage of the lower status of the population. This means that as the percentage of lower status population increases, the median value of homes tends to decrease.

4 The Avg Number of Rooms are Incressing the price of the house is also incresing Becuse ther are sufficent space for the buyer buyer give lots of money to the land onwer and wants to good fecility in the house

In [ ]:

In [ ]:

```

```