[11]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis
```

[13]:
```
df = pd.read_csv("Boston.csv")
```

[14]:
```
# Display my some top5 sample data
print("Shape of dataset:", df.shape)   # in my data present 506 row and  column
print("\nFirst 5 rows:\n", df.head())
```

```
Shape of dataset: (506, 15)

First 5 rows:
    Unnamed: 0     crim    zn  indus  chas    nox     rm   age     dis  rad  \
0           1  0.00632  18.0   2.31     0  0.538  6.575  65.2  4.0900    1
1           2  0.02731   0.0   7.07     0  0.469  6.421  78.9  4.9671    2
2           3  0.02729   0.0   7.07     0  0.469  7.185  61.1  4.9671    2
3           4  0.03237   0.0   2.18     0  0.458  6.998  45.8  6.0622    3
4           5  0.06905   0.0   2.18     0  0.458  7.147  54.2  6.0622    3
```

```
   tax  ptratio   black  lstat  medv
0  296     15.3  396.90   4.98  24.0
1  242     17.8  396.90   9.14  21.6
2  242     17.8  392.83   4.03  34.7
3  222     18.7  394.63   2.94  33.4
4  222     18.7  396.90   5.33  36.2
```

5]:
```
# Select only numeric columns
num_cols = df.select_dtypes(include=['float64', 'int64']).columns
print("\nNumeric Columns:\n", num_cols)
```

```
Numeric Columns:
 Index(['Unnamed: 0', 'crim', 'zn', 'indus', 'chas', 'nox', 'rm', 'age', 'dis',
       'rad', 'tax', 'ptratio', 'black', 'lstat', 'medv'],
      dtype='object')
```

8]:
```
 # Univariate Analysis Histograms + Skewness + Kurtosis ---
print("\nSkewness and Kurtosis of Numeric Columns:")
stats = pd.DataFrame({
    'Skewness': df[num_cols].skew(),
    'Kurtosis': df[num_cols].kurt()})
print(stats)
```

```
Skewness and Kurtosis of Numeric Columns:
```

```
Skewness and Kurtosis of Numeric Columns:
            Skewness   Kurtosis
Unnamed: 0  0.000000  -1.200000
crim        5.223149  37.130509
zn          2.225666   4.031510
indus       0.295022  -1.233540
chas        3.405904   9.638264
nox         0.729308  -0.064667
rm          0.403612   1.891500
age        -0.598963  -0.967716
dis         1.011781   0.487941
rad         1.004815  -0.867232
tax         0.669956  -1.142408
ptratio    -0.802325  -0.285091
black      -2.890374   7.226818
lstat       0.906460   0.493240
medv        1.108098   1.495197
```

#for col in num_cols: #print(f"{col:10s} Skewness = {df[col].skew():.2f} | Kurtosis = {df[col].kurt():.2f}")

```
1->Skewness
   skew = 0 approximately symmetric
   kew > 0 right-skewed (long tail on the right)
   skew < 0 left-skewed (long tail on the left)
```

```
2->Kurtosis
   kurtosis ~ 3 → normal distribution
   kurtosis > 3 → leptokurtic (peaked)
   kurtosis < 3 → platykurtic (flat)
```

```
#Univariate Analysis Histograms + Skewness + Kurtosis ---
plt.figure(figsize=(15, len(num_cols)*3))
```

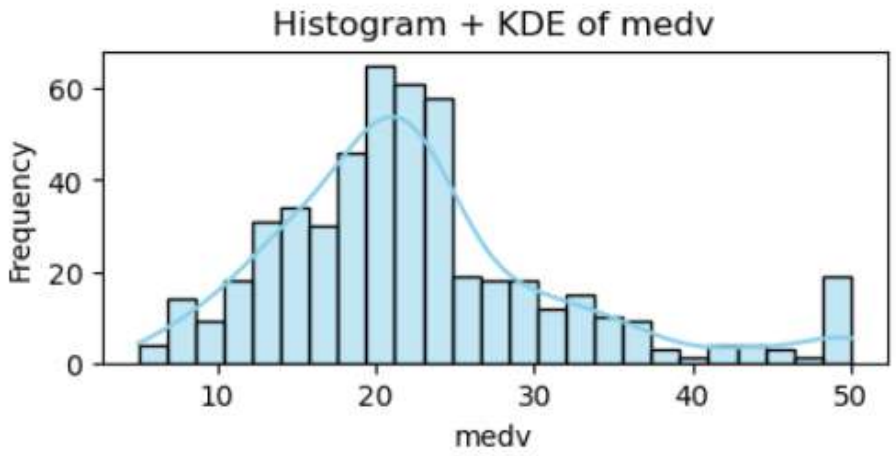
```
<Figure size 1500x4500 with 0 Axes>
```

<Figure size 1500x4500 with 0 Axes>

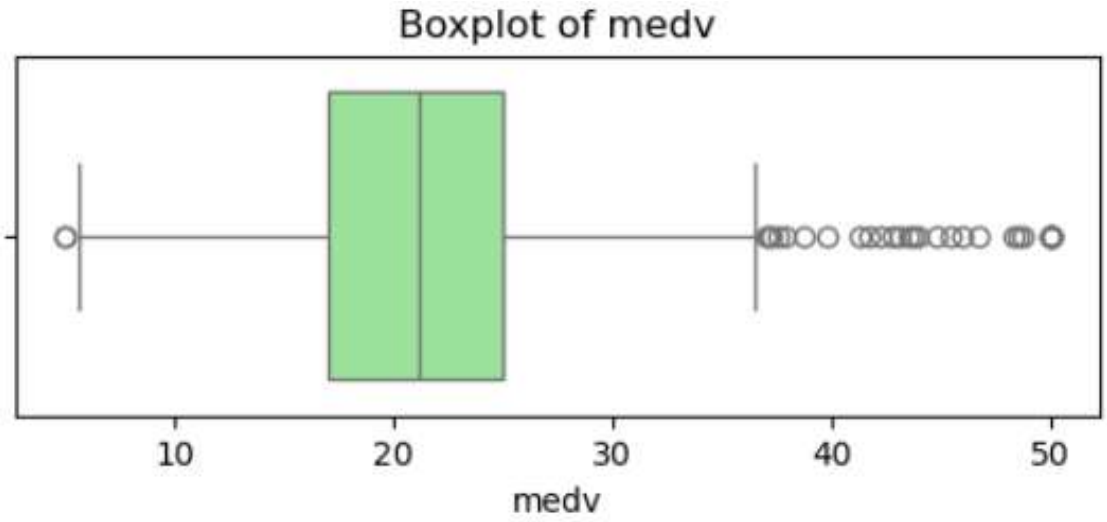
```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(5,2))  # Figure size first

sns.histplot(df[col], kde=True, bins=25, color='skyblue', edgecolor='black')
plt.title(f"Histogram + KDE of {col}")
plt.xlabel(col)
plt.ylabel("Frequency")
plt.show()
```



[19]:
```
#Boxplot
plt.figure(figsize=(6,2))  # smaller figure for boxplot
sns.boxplot(x=df[col], color='lightgreen')
plt.title(f"Boxplot of {col}")
plt.show()
```




```
#handle outlier with boxplot
```

9]:
```
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5*IQR
    upper_bound = Q3 + 1.5*IQR
```

5]:
```
outlier= df[(df[col] >= lower_bound) & (df[col] <= upper_bound)]
```

[37]:
```
plt.figure(figsize=(6,2))  # smaller figure for boxplot
sns.boxplot(x=df[col], color='lightgreen')
plt.title(f"Boxplot of {col}")
plt.show()
```