ASSIGNMENT-2

Course :Statistical machine learning

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import skew, kurtosis


dp=pd.read_csv('/content/housing.csv')
dp
```

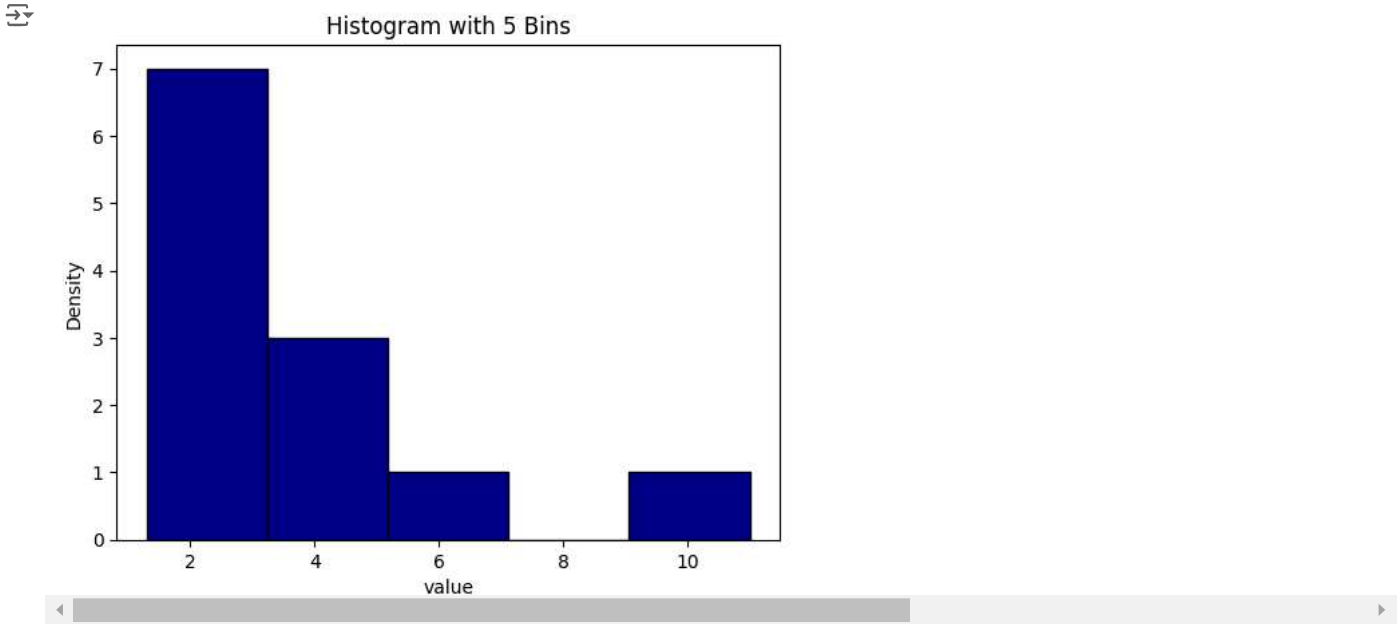
	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_val
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	45260
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	35850
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	35210
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	34130
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	34220
...	...	...	...	...	...	...	...	...	
20635	-121.09	39.48	25.0	1665.0	374.0	845.0	330.0	1.5603	7810
20636	-121.21	39.49	18.0	697.0	150.0	356.0	114.0	2.5568	7710
20637	-121.22	39.43	17.0	2254.0	485.0	1007.0	433.0	1.7000	9230
20638	-121.32	39.43	18.0	1860.0	409.0	741.0	349.0	1.8672	8470
20639	-121.24	39.37	16.0	2785.0	616.0	1387.0	530.0	2.3886	8940

20640 rows × 10 columns

Next steps: Generate code with dp | View recommended plots | New interactive sheet

1Q

```
X={1.3,1.5,2.6,2.6,3.2,3.9,4.2,3.7,3.10,3.0,11,6.7,1.9}
plt.hist(X,bins=5,color="darkblue",edgecolor="black" )
plt.xlabel('value')
plt.ylabel('Density')
plt.title('Histogram with 5 Bins')
plt.show()
```



2Q

```
X=[12,13,56,23,41,28,51]
Y=[51,63,37,88,33,72.81]

def mean(data):
  return np.mean(data)

def median(data):
  return np.median(data)

def mode(data):
  return stats.mode(data)

def variance(data):
  return np.var(data)

def range(data):
  return max(data) - min(data)

def iqr(data):
  q1, q3 = np.percentile(data, [25,75])

def skewness(data):
  return stats.skew(data)

def kurtosis(data):
  return stats.kurtosis(data)

def stddev(data):
  return np.std(data)

print("Statistics for X:")
print("Mean:", mean(X))
print("Median:", median(X))
print("Mode:", mode(X))
print("Variance:", variance(X))
print("Standard deviation:", stddev(X))
print("Range:", range(X))
print("Interquartile range (IQR):", iqr(X))
print("Skewness:", skewness(X))
print("Kurtosis:", kurtosis(X))
```

```
print("\nStatistics for Y:")
print("Mean:", mean(Y))
print("Median:", median(Y))
print("Mode:", mode(Y))
print("Variance:", variance(Y))
print("Standard deviation:", stddev(Y))
print("Range:", range(Y))
print("Interquartile range (IQR):", iqr(Y))
print("Skewness:", skewness(Y))
print("Kurtosis:", kurtosis(Y))
```

```
Statistics for X:
Mean: 32.0
Median: 28.0
Mode: ModeResult(mode=12, count=1)
Variance: 268.0
Standard deviation: 16.3707055437449
Range: 44
Interquartile range (IQR): None
Skewness: 0.18755223867066048
Kurtosis: -1.47684021258314

Statistics for Y:
Mean: 57.468333333333334
Median: 57.0
Mode: ModeResult(mode=33.0, count=1)
Variance: 376.2733472222222
Standard deviation: 19.397766552420983
Range: 55
Interquartile range (IQR): None
Skewness: 0.19971617107804526
Kurtosis: -1.2801310517130808
```

3Q

```
Y = [51,63,37,88,33,72.81]
X_dependent = [12,13,56,23,41,28,51]
X_dependent = X_dependent[:len(Y)]
XY_df = pd.DataFrame({'Y': Y, 'X': X_dependent})
XY_stats = XY_df.describe().T
XY_stats['IQR'] = XY_stats['75%'] - XY_stats['25%']
XY_stats['Skewness'] = XY_df.apply(lambda x: skew(x), axis=0)
XY_stats['Kurtosis'] = XY_df.apply(lambda x: kurtosis(x), axis=0)
print("\nStatistical measures for the features in the given table:")
print(XY_stats)
```

```
Statistical measures for the features in the given table:
   count       mean        std   min   25%   50%      75%   max      IQR  \
Y    6.0  57.468333  21.249189  33.0  40.5  57.0  70.3575  88.0  29.8575
X    6.0  28.833333  17.057745  12.0  15.5  25.5  37.7500  56.0  22.2500

   Skewness  Kurtosis
Y  0.199716 -1.280131
X  0.569986 -0.984739
```

Done By KODAM SHISHIR BHAGATH [2303A52164]