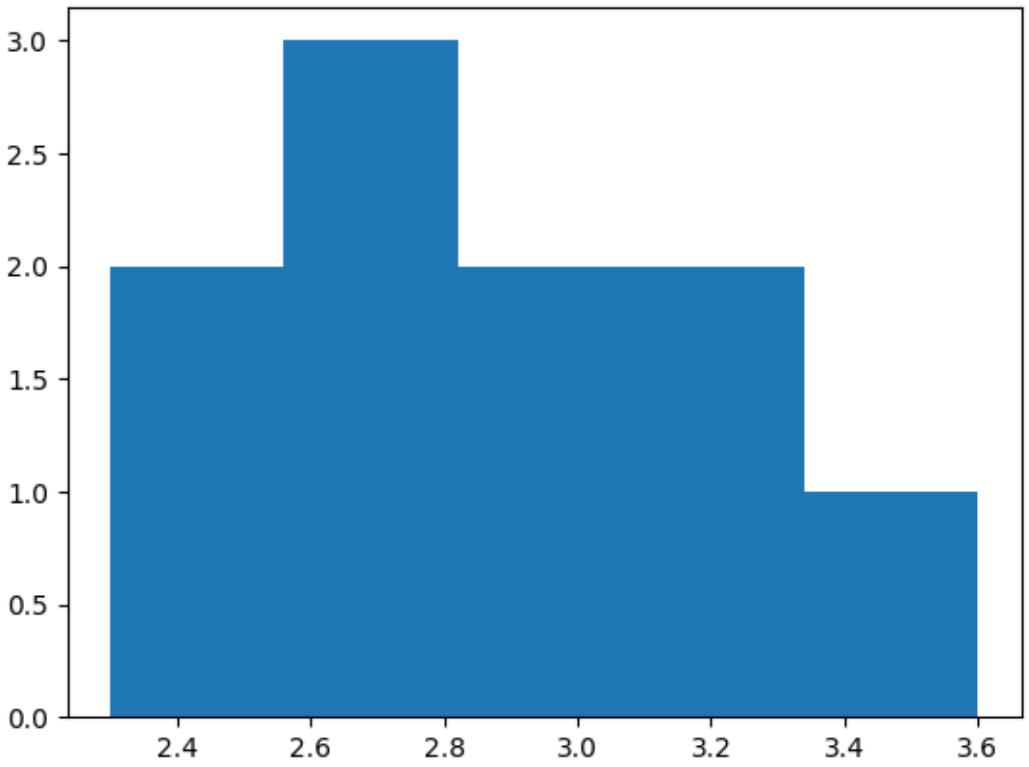
2303a52154-assignment-1

August 9, 2024

[13]:
```
# Question 1
import numpy as np
import matplotlib.pyplot as plt
x=np.array([2.3, 2.5, 3.6, 2.8, 3.1, 2.9, 3.2, 2.7, 2.8, 3.0])
plt.hist(x,bins=5)
plt.show()
```



[ ]:
```
# Find Mean,Median,Mode,Variance,Standard deviation,Range,Interquartile␣
↪Range(IQR), Skewness,Kurtosis of x and y
```

[14]:
```
import numpy as np
from scipy import stats as s
```

```
from scipy.stats import kurtosis
from scipy.stats import skew
import matplotlib.pyplot as plt
x=[4,5,8,2,4,2,5]
y=[5,6,3,8,3,7,8]
```

[15]:
```
print("Mean of x is",np.mean(x))
print("Mean of y is",np.mean(y))
```

```
Mean of x is 4.285714285714286
Mean of y is 5.714285714285714
```

[16]:
```
print("Median of x is",np.median(x))
print("Median of y is",np.median(y))
```

```
Median of x is 4.0
Median of y is 6.0
```

[17]:
```
print("Mode of x is",s.mode(x))
print("Mode of y is",s.mode(y))
```

```
Mode of x is ModeResult(mode=2, count=2)
Mode of y is ModeResult(mode=3, count=2)
```

[18]:
```
print("Variance of x is",np.var(x))
print("Variance of y is",np.var(y))
```

```
Variance of x is 3.6326530612244894
Variance of y is 3.918367346938776
```

[19]:
```
print("Standard deviation of x is",np.std(x))
print("Standard deviation of y is",np.std(y))
```

```
Standard deviation of x is 1.9059520091609048
Standard deviation of y is 1.979486637221574
```

[20]:
```
print("Range of x is",max(x)-min(x))
print("Range of y is",max(y)-min(y))
```

```
Range of x is 6
Range of y is 5
```

[21]:
```
print("Interquartile range of x is",np.percentile(x,75)-np.percentile(x,25))
print("Interquartile range of y is",np.percentile(y,75)-np.percentile(y,25))
```

```
Interquartile range of x is 2.0
Interquartile range of y is 3.5
```

[24]:
```
import numpy as np
x = [4, 5, 8, 2, 4, 2, 5]
y = [5, 6, 3, 8, 3, 7, 8]
n_x = len(x)
mean_x = np.mean(x)
std_dev_x = np.sqrt(np.sum((np.array(x) - mean_x) ** 2) / (n_x - 1))
skewness_x = (n_x / ((n_x - 1) * (n_x - 2))) * np.sum(((np.array(x) - mean_x) /␣
 ↪std_dev_x) ** 3)
n_y = len(y)
mean_y = np.mean(y)
std_dev_y = np.sqrt(np.sum((np.array(y) - mean_y) ** 2) / (n_y - 1))
skewness_y = (n_y / ((n_y - 1) * (n_y - 2))) * np.sum(((np.array(y) - mean_y) /␣
 ↪std_dev_y) ** 3)
print("Skewness of x:", skewness_x)
print("Skewness of y:", skewness_y)
```

```
Skewness of x: 0.749913842326791
Skewness of y: -0.34201087050980616
```

[25]:
```
n_x = len(x)
mean_x = np.mean(x)
std_dev_x = np.sqrt(np.sum((np.array(x) - mean_x) ** 2) / (n_x - 1))
kurtosis_x = ((n_x * (n_x + 1)) / ((n_x - 1) * (n_x - 2) * (n_x - 3))) * np.
 ↪sum(((np.array(x) - mean_x) / std_dev_x) ** 4) - (3 * (n_x - 1) ** 2 / ((n_x␣
 ↪- 2) * (n_x - 3)))
n_y = len(y)
mean_y = np.mean(y)
std_dev_y = np.sqrt(np.sum((np.array(y) - mean_y) ** 2) / (n_y - 1))
kurtosis_y = ((n_y * (n_y + 1)) / ((n_y - 1) * (n_y - 2) * (n_y - 3))) * np.
 ↪sum(((np.array(y) - mean_y) / std_dev_y) ** 4) - (3 * (n_y - 1) ** 2 / ((n_y␣
 ↪- 2) * (n_y - 3)))
print("Kurtosis of x:", kurtosis_x)
print("Kurtosis of y:", kurtosis_y)
```

```
Kurtosis of x: 0.9772250978411812
Kurtosis of y: -1.6898437500000005
```

[26]:
```
import pandas as pd
import numpy as np
from scipy import stats as s
import matplotlib.pyplot as plt
```

[27]:
```
d=pd.read_csv('/content/sample_data/california_housing_train.csv')
d
```

[27]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	\
0	-114.31	34.19	15.0	5612.0	1283.0	

```
1        -114.47     34.40                19.0       7650.0          1901.0
2        -114.56     33.69                17.0        720.0           174.0
3        -114.57     33.64                14.0       1501.0           337.0
4        -114.57     33.57                20.0       1454.0           326.0
…           …        …                    …          …               …
16995    -124.26     40.58                52.0       2217.0           394.0
16996    -124.27     40.69                36.0       2349.0           528.0
16997    -124.30     41.84                17.0       2677.0           531.0
16998    -124.30     41.80                19.0       2672.0           552.0
16999    -124.35     40.54                52.0       1820.0           300.0

       population  households  median_income  median_house_value
0          1015.0       472.0         1.4936             66900.0
1          1129.0       463.0         1.8200             80100.0
2           333.0       117.0         1.6509             85700.0
3           515.0       226.0         3.1917             73400.0
4           624.0       262.0         1.9250             65500.0
…             …          …             …                  …
16995       907.0       369.0         2.3571            111400.0
16996      1194.0       465.0         2.5179             79000.0
16997      1244.0       456.0         3.0313            103600.0
16998      1298.0       478.0         1.9797             85800.0
16999       806.0       270.0         3.0147             94600.0

[17000 rows x 9 columns]
```

[28]:
```
print("1.Mean of Longitude is = ",np.mean(d['longitude']))
print("2.Mean of Latitude is = ",np.mean(d['latitude']))
print("3.Mean of Housing Median Age is = ",np.mean(d['housing_median_age']))
print("4.Mean of Total Rooms is = ",np.mean(d['total_rooms']))
print("5.Mean of Total Bedrooms is = ",np.mean(d['total_bedrooms']))
print("6.Mean of Population is = ",np.mean(d['population']))
print("7.Mean of Households is = ",np.mean(d['households']))
print("8.Mean of Median Income is = ",np.mean(d['median_income']))
print("9.Mean of Median House_value is = ",np.mean(d['median_house_value']))
```

```
1.Mean of Longitude is =  -119.5621082352941
2.Mean of Latitude is =  35.62522470588235
3.Mean of Housing Median Age is =  28.58935294117647
4.Mean of Total Rooms is =  2643.664411764706
5.Mean of Total Bedrooms is =  539.4108235294118
6.Mean of Population is =  1429.5739411764705
7.Mean of Households is =  501.2219411764706
8.Mean of Median Income is =  3.8835781000000007
9.Mean of Median House_value is =  207300.91235294117
```

[29]:
```
print("1.Median of Longitude is = ",np.median(d['longitude']))
print("2.Median of Latitude is = ",np.median(d['latitude']))
print("3.Median of Housing Median Age is = ",np.median(d['housing_median_age']))
print("4.Median of Total Rooms is = ",np.median(d['total_rooms']))
print("5.Median of Total Bedrooms is = ",np.median(d['total_bedrooms']))
print("6.Median of population is = ",np.median(d['population']))
print("7.Median of Households is = ",np.median(d['households']))
print("8.Median of Median Income is = ",np.median(d['median_income']))
print("9.Median of Median House Value is = ",np.median(d['median_house_value']))
```

```
1.Median of Longitude is =  -118.49
2.Median of Latitude is =  34.25
3.Median of Housing Median Age is =  29.0
4.Median of Total Rooms is =  2127.0
5.Median of Total Bedrooms is =  434.0
6.Median of population is =  1167.0
7.Median of Households is =  409.0
8.Median of Median Income is =  3.5446
9.Median of Median House Value is =  180400.0
```

[30]:
```
print("1.Mode of Longitude is = ",s.mode(d['longitude']))
print("2.Mode of Latitude is = ",s.mode(d['latitude']))
print("3.Mode of Housing Median Age is = ",s.mode(d['housing_median_age']))
print("4.Mode of Total Rooms is = ",s.mode(d['total_rooms']))
print("5.Mode of Total Bedrooms is = ",s.mode(d['total_bedrooms']))
print("6.Mode of Population is = ",s.mode(d['population']))
print("7.Mode of Households is = ",s.mode(d['households']))
print("8.Mode of Median Income is = ",s.mode(d['median_income']))
print("9.Mode of Median House Value is = ",s.mode(d['median_house_value']))
```

```
1.Mode of Longitude is =  ModeResult(mode=-118.31, count=136)
2.Mode of Latitude is =  ModeResult(mode=34.06, count=205)
3.Mode of Housing Median Age is =  ModeResult(mode=52.0, count=1052)
4.Mode of Total Rooms is =  ModeResult(mode=1582.0, count=16)
5.Mode of Total Bedrooms is =  ModeResult(mode=280.0, count=48)
6.Mode of Population is =  ModeResult(mode=891.0, count=23)
7.Mode of Households is =  ModeResult(mode=306.0, count=48)
8.Mode of Median Income is =  ModeResult(mode=3.125, count=41)
9.Mode of Median House Value is =  ModeResult(mode=500001.0, count=814)
```

[31]:
```
print("1.Variance of Longitude is = ",np.var(d['longitude']))
print("2.Variance of Latitude is = ",np.var(d['latitude']))
print("3.Variance of Housing Median_age is = ",np.var(d['housing_median_age']))
print("4.Variance of Total Rooms is = ",np.var(d['total_rooms']))
print("5.Variance of Total Bedrooms is = ",np.var(d['total_bedrooms']))
print("6.Variance of Population is = ",np.var(d['population']))
print("7.Variance of Households is = ",np.var(d['households']))
```

```
print("8.Variance of Median Income is = ",np.var(d['median_income']))
print("9.Variance of Median House Value is = ",np.var(d['median_house_value']))
```

```
1.Variance of Longitude is =  4.020455814167474
2.Variance of Latitude is =  4.56795267891903
3.Variance of Housing Median_age is =  158.42166311072666
4.Variance of Total Rooms is =  4751889.69496877
5.Variance of Total Bedrooms is =  177651.3369887336
6.Variance of Population is =  1317488.9119444669
7.Variance of Households is =  147847.57962446712
8.Variance of Median Income is =  3.6408471185562723
9.Variance of Median House Value is =  13451442293.56867
```

[32]:
```
print("1.Standard deviation of Longitude is = ",np.std(d['longitude']))
print("2.Standard deviation of Latitude is = ",np.std(d['latitude']))
print("3.Standard deviation of Housing Median Age is = ",np.
 ↪std(d['housing_median_age']))
print("4.Standard deviation of Total Rooms is = ",np.std(d['total_rooms']))
print("5.Standard deviation of Total Bedrooms is = ",np.
 ↪std(d['total_bedrooms']))
print("6.Standard deviation of Population is = ",np.std(d['population']))
print("7.Standard deviation of Households is = ",np.std(d['households']))
print("8.Standard deviation of Median Income is = ",np.std(d['median_income']))
print("9.Standard deviation of Median House Value is = ",np.
 ↪std(d['median_house_value']))
```

```
1.Standard deviation of Longitude is =  2.005107432076265
2.Standard deviation of Latitude is =  2.1372769307974644
3.Standard deviation of Housing Median Age is =  12.58656677218719
4.Standard deviation of Total Rooms is =  2179.8829544195187
5.Standard deviation of Total Bedrooms is =  421.4870543548563
6.Standard deviation of Population is =  1147.819198281884
7.Standard deviation of Households is =  384.50953125308496
8.Standard deviation of Median Income is =  1.9081003953032116
9.Standard deviation of Median House Value is =  115980.35304985354
```

[33]:
```
print("1.Range of Longitude is = ",max(d['longitude'])-min(d['longitude']))
print("2.Range of Latitude is  = ",max(d['latitude'])-min(d['latitude']))
print("3.Range of Housing Median Age is =␣
 ↪",max(d['housing_median_age'])-min(d['housing_median_age']))
print("4.Range of Total Rooms is =␣
 ↪",max(d['total_rooms'])-min(d['total_rooms']))
print("5.Range of Total Bedrooms is =␣
 ↪",max(d['total_bedrooms'])-min(d['total_bedrooms']))
print("6.Range of Population is = ",max(d['population'])-min(d['population']))
print("7.Range of Households is = ",max(d['households'])-min(d['households']))
```

```
print("8.Range of Median Income is =␣
  ↪",max(d['median_income'])-min(d['median_income']))
print("9.Range of Median House Value is =␣
  ↪",max(d['median_house_value'])-min(d['median_house_value']))
```

```
1.Range of Longitude is =  10.039999999999992
2.Range of Latitude is  =  9.410000000000004
3.Range of Housing Median Age is =  51.0
4.Range of Total Rooms is =  37935.0
5.Range of Total Bedrooms is =  6444.0
6.Range of Population is =  35679.0
7.Range of Households is =  6081.0
8.Range of Median Income is =  14.5002
9.Range of Median House Value is =  485002.0
```

[34]:
```
print("1.Interquartile range of Longitude is = ",np.
  ↪percentile(d['longitude'],75)-np.percentile(d['longitude'],25))
print("2.Interquartile range of Latitude is = ",np.
  ↪percentile(d['latitude'],75)-np.percentile(d['latitude'],25))
print("3.Interquartile range of Housing Median Age is = ",np.
  ↪percentile(d['housing_median_age'],75)-np.
  ↪percentile(d['housing_median_age'],25))
print("4.Interquartile range of Total Rooms is = ",np.
  ↪percentile(d['total_rooms'],75)-np.percentile(d['total_rooms'],25))
print("5.Interquartile range of Total Bedrooms is = ",np.
  ↪percentile(d['total_bedrooms'],75)-np.percentile(d['total_bedrooms'],25))
print("6.Interquartile range of Population is = ",np.
  ↪percentile(d['population'],75)-np.percentile(d['population'],25))
print("7.Interquartile range of Households is = ",np.
  ↪percentile(d['households'],75)-np.percentile(d['households'],25))
print("8.Interquartile range of Median Income is = ",np.
  ↪percentile(d['median_income'],75)-np.percentile(d['median_income'],25))
print("9.Interquartile range of Median House_value is = ",np.
  ↪percentile(d['median_house_value'],75)-np.
  ↪percentile(d['median_house_value'],25))
```

```
1.Interquartile range of Longitude is =  3.7900000000000063
2.Interquartile range of Latitude is =  3.789999999999999
3.Interquartile range of Housing Median Age is =  19.0
4.Interquartile range of Total Rooms is =  1689.25
5.Interquartile range of Total Bedrooms is =  351.25
6.Interquartile range of Population is =  931.0
7.Interquartile range of Households is =  323.25
8.Interquartile range of Median Income is =  2.2006250000000005
9.Interquartile range of Median House_value is =  145600.0
```

[35]:
```
A=['longitude','latitude','housing_median_age','total_rooms','total_bedrooms','population','ho
for i in A:
    values=d[i]
    mean=np.mean(values)
    std=np.std(values)
    skewness=np.mean(((values-mean)/std)**3)
    print("Skewness of the given ",i,"is ",skewness)
```

```
Skewness of the given  longitude is  -0.3039761523070952
Skewness of the given  latitude is  0.47175948982753174
Skewness of the given  housing_median_age is  0.06488830685009046
Skewness of the given  total_rooms is  4.00237680794326
Skewness of the given  total_bedrooms is  3.322343534496144
Skewness of the given  population is  5.1867541718637655
Skewness of the given  households is  3.3423734139781573
Skewness of the given  median_income is  1.6265495626993864
Skewness of the given  median_house_value is  0.972950775194606
```

[12]:
```
A=['longitude','latitude','housing_median_age','total_rooms','total_bedrooms','population','ho
for i in A:
    values=d[i]
    mean=np.mean(values)
    std=np.std(values)
    kurtosis=np.mean(((values-mean)/std)**4)-3
    print("Kurtosis of the given ",i,"is",kurtosis)
```

```
Kurtosis of the given  longitude is -1.3222937000181345
Kurtosis of the given  latitude is -1.1122523149717822
Kurtosis of the given  housing_median_age is -0.8009436273526562
Kurtosis of the given  total_rooms is 29.50685142754096
Kurtosis of the given  total_bedrooms is 19.6866056565209
Kurtosis of the given  population is 80.83786311937729
Kurtosis of the given  households is 20.686206054602916
Kurtosis of the given  median_income is 4.76239090620088
Kurtosis of the given  median_house_value is 0.30355527450221276
```