Practical List Question 3

```
import pandas as pd

from google.colab import drive
drive.mount('/content/gdrive')
df=pd.read_csv('/content/gdrive/MyDrive/Q3_wine.csv')
print(df)
```

```
Mounted at /content/gdrive
     Wine  Alcohol  Malic.acid   Ash  ...  Color.int   Hue    OD  Proline
0       1    14.23        1.71  2.43  ...       5.64  1.04  3.92     1065
1       1    13.20        1.78  2.14  ...       4.38  1.05  3.40     1050
2       1    13.16        2.36  2.67  ...       5.68  1.03  3.17     1185
3       1    14.37        1.95  2.50  ...       7.80  0.86  3.45     1480
4       1    13.24        2.59  2.87  ...       4.32  1.04  2.93      735
..    ...      ...         ...   ...  ...        ...   ...   ...      ...
173     3    13.71        5.65  2.45  ...       7.70  0.64  1.74      740
174     3    13.40        3.91  2.48  ...       7.30  0.70  1.56      750
175     3    13.27        4.28  2.26  ...      10.20  0.59  1.56      835
176     3    13.17        2.59  2.37  ...       9.30  0.60  1.62      840
177     3    14.13        4.10  2.74  ...       9.20  0.61  1.60      560

[178 rows x 14 columns]
```

Check whether all attributes are standardized or not (mean is 0 and standard deviation is 1). If not, standardize the attributes.

```
Mean_result=list()
for i in range(1,df.shape[1]):
  column=df.columns[i]
  Mean_result.append(df[column].mean())
print(Mean_result)
print(len(Mean_result))
```

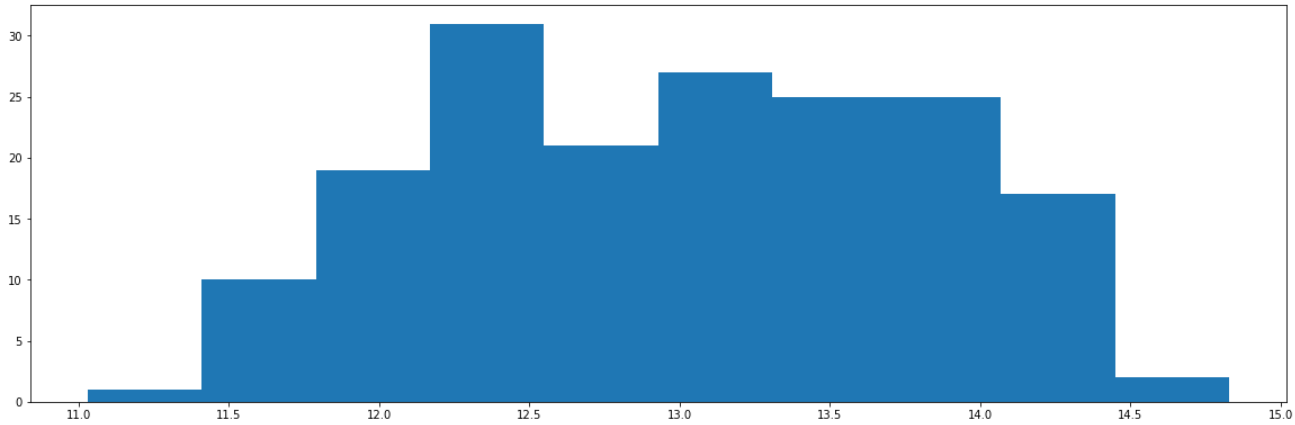
```
[13.000617977528083, 2.336348314606741, 2.3665168539325854, 19.49494382022472, 99.741
13
```

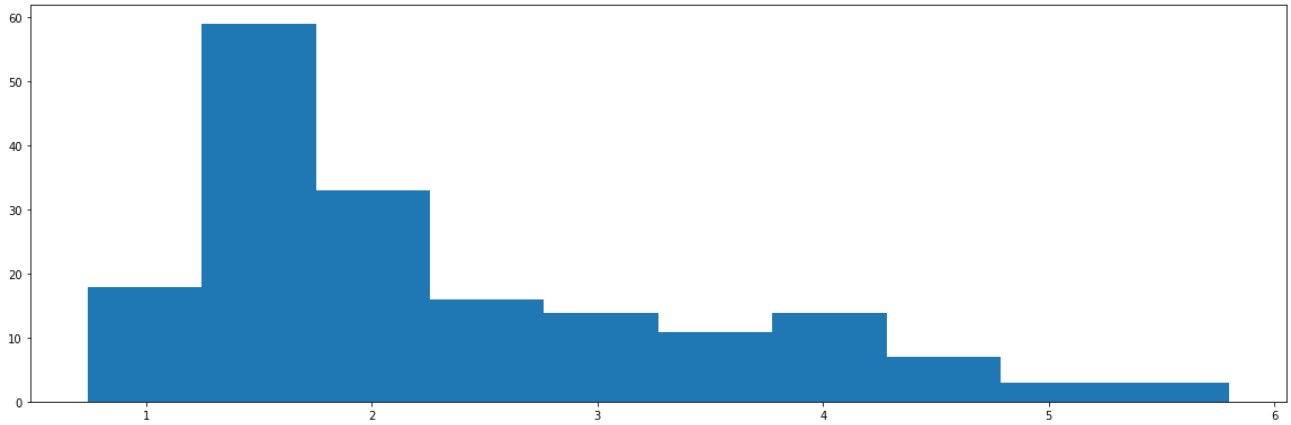
```
SD_result=list()
for i in range(1,df.shape[1]):
  column=df.columns[i]
  SD_result.append(df[column].std())
print(SD_result)
```

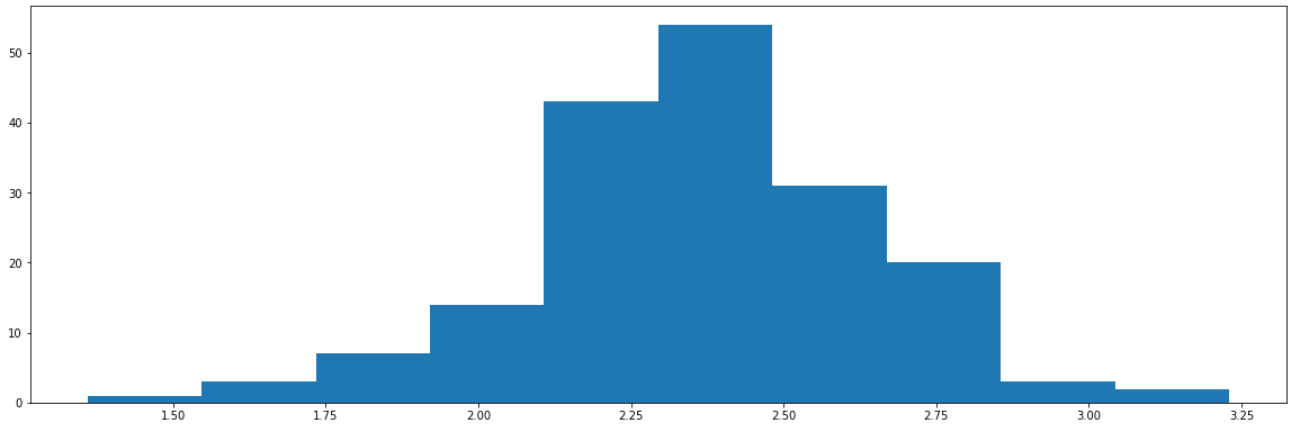
```
[0.8118265380058577, 1.1171460976144627, 0.2743440090608148, 3.3395637671735052, 14.2
```

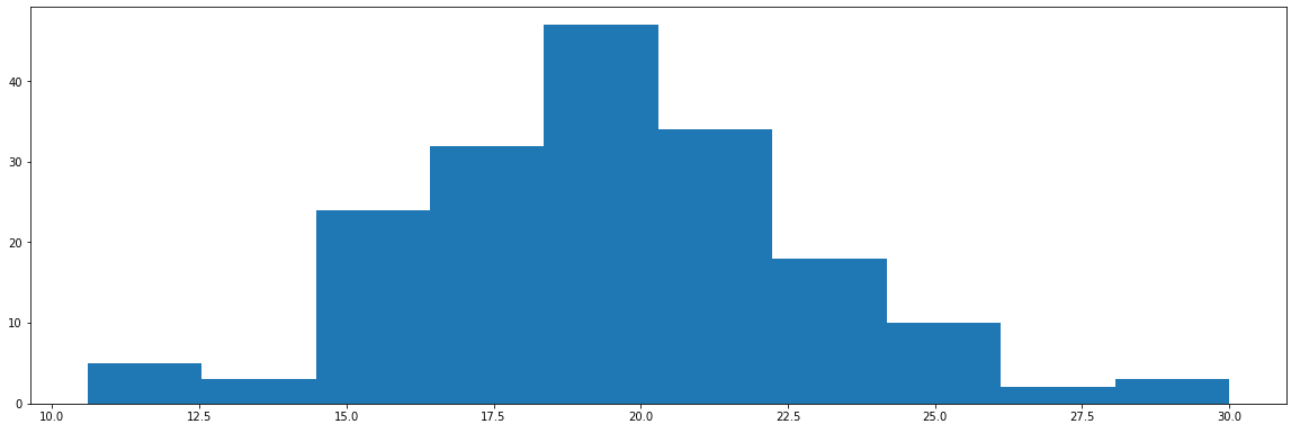
```
from matplotlib import pyplot as plt
plt.figure(figsize = [20, 100])
for i in range(1,df.shape[1]):
```

```
    column=df.columns[i]
    plt.subplot(13, 1, i)
    plt.hist(df[column])
```

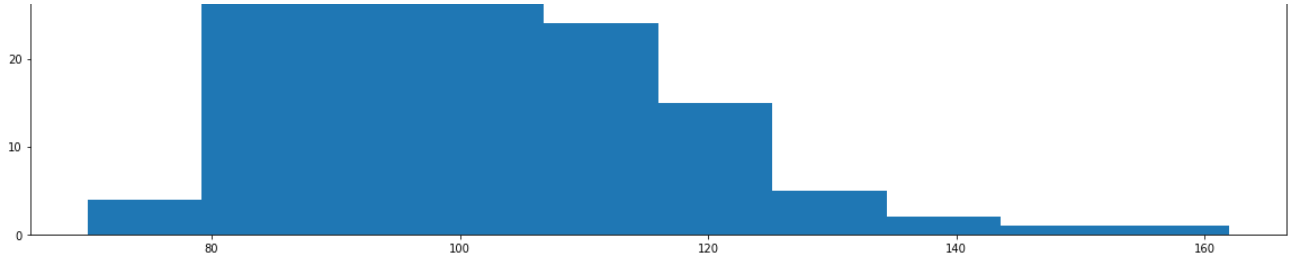

30
25
20
15
10
5
0
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0

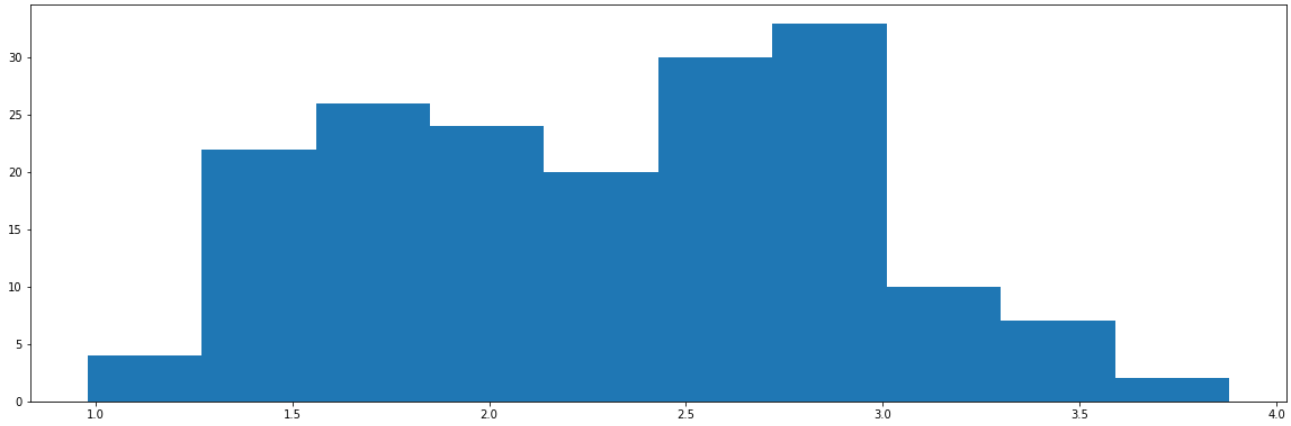

60
50
40
30
20
10
0
1
2
3
4
5
6

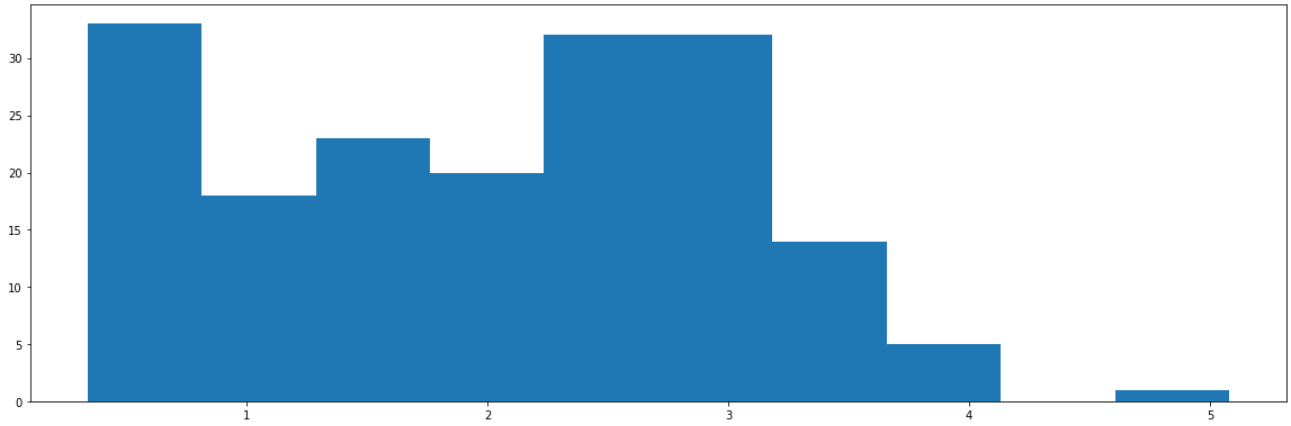

50
40
30
20
10
0
1.50
1.75
2.00
2.25
2.50
2.75
3.00
3.25

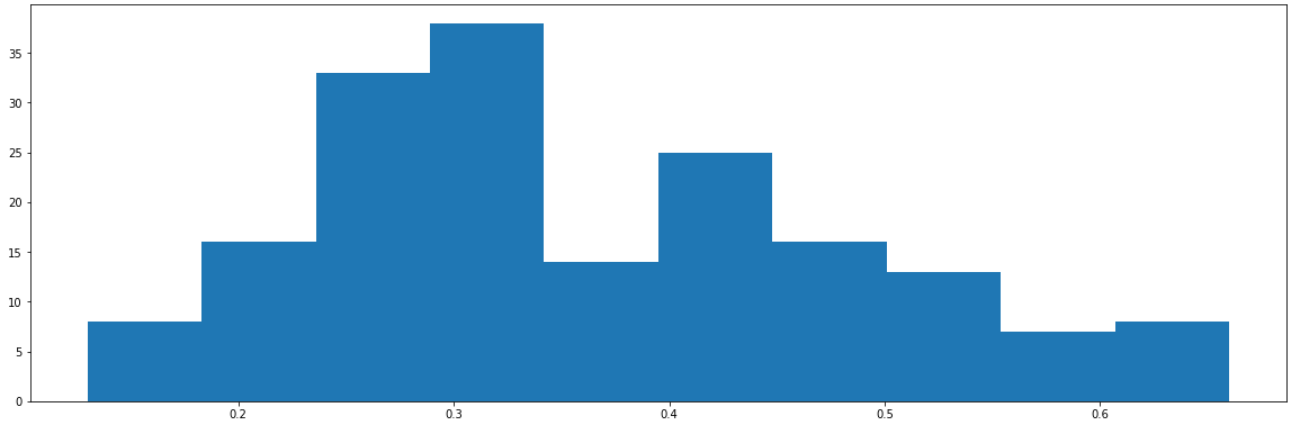

40
30
20
10
0
10.0
12.5
15.0
17.5
20.0
22.5
25.0
27.5
30.0


40
30

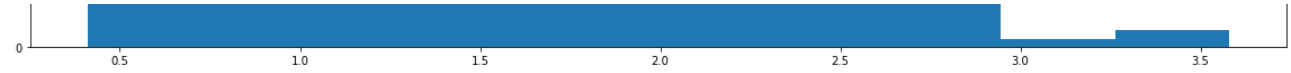

20
10
0
80
100
120
140
160

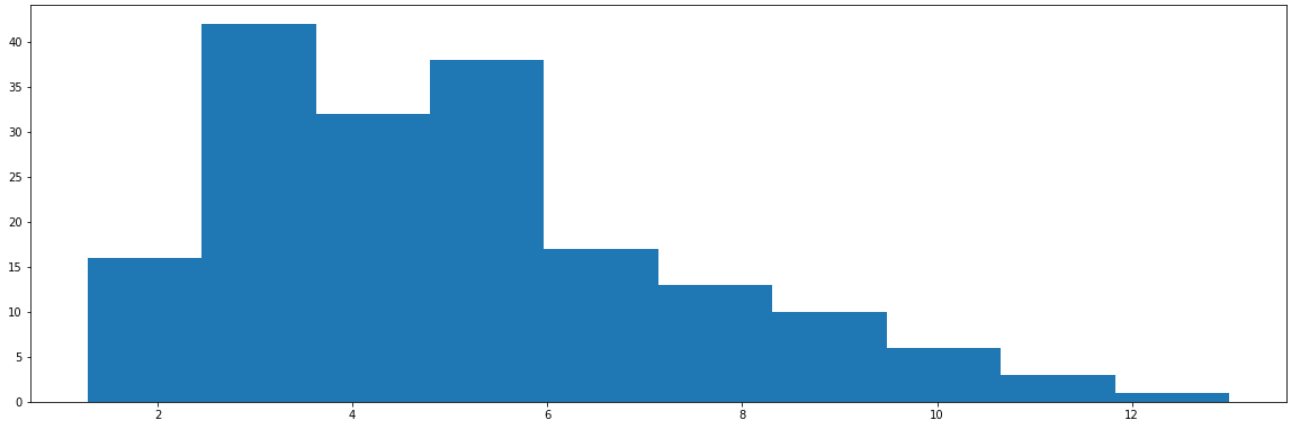

30
25
20
15
10
5
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0

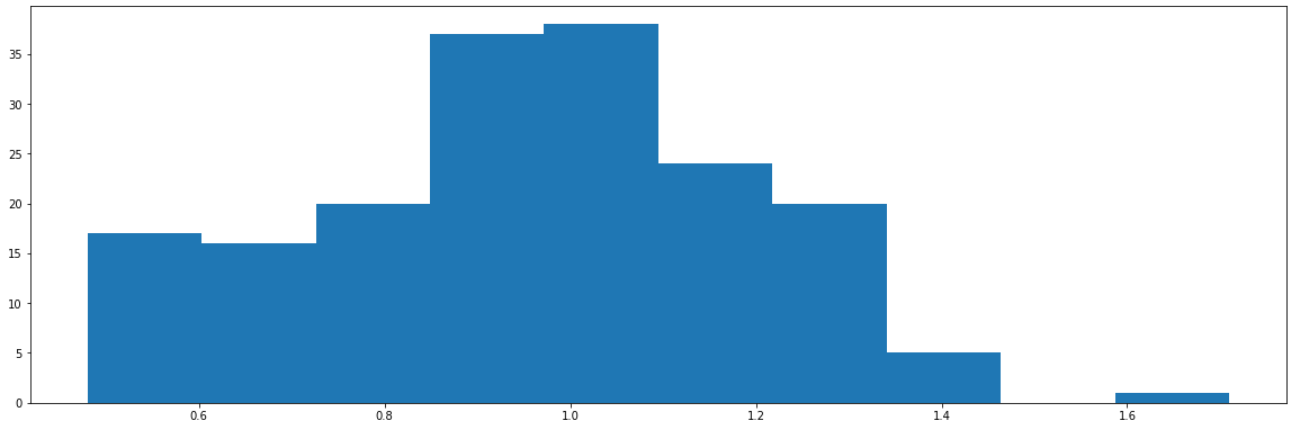

30
25
20
15
10
5
0
1
2
3
4
5

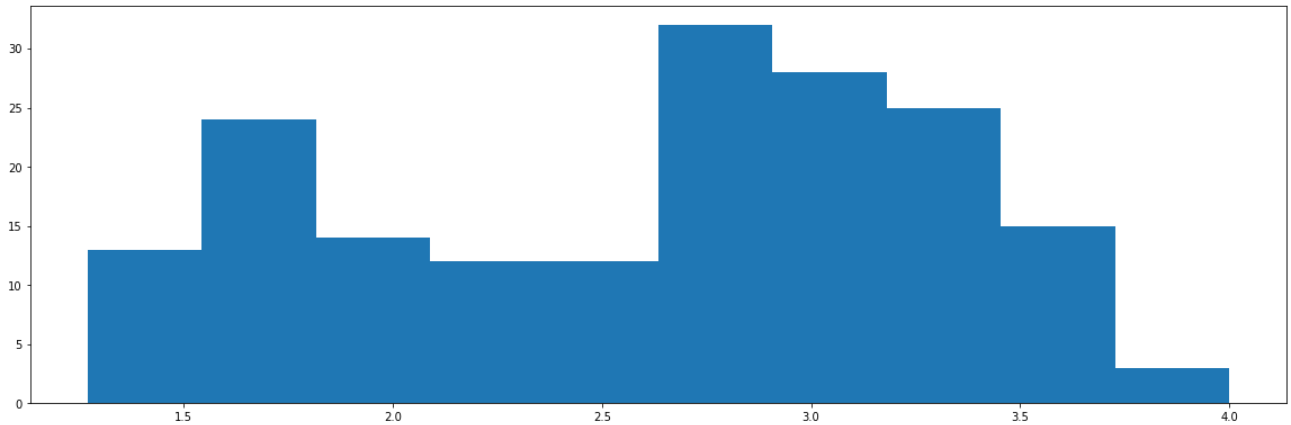

35
30
25
20
15
10
5
0
0.2
0.3
0.4
0.5
0.6

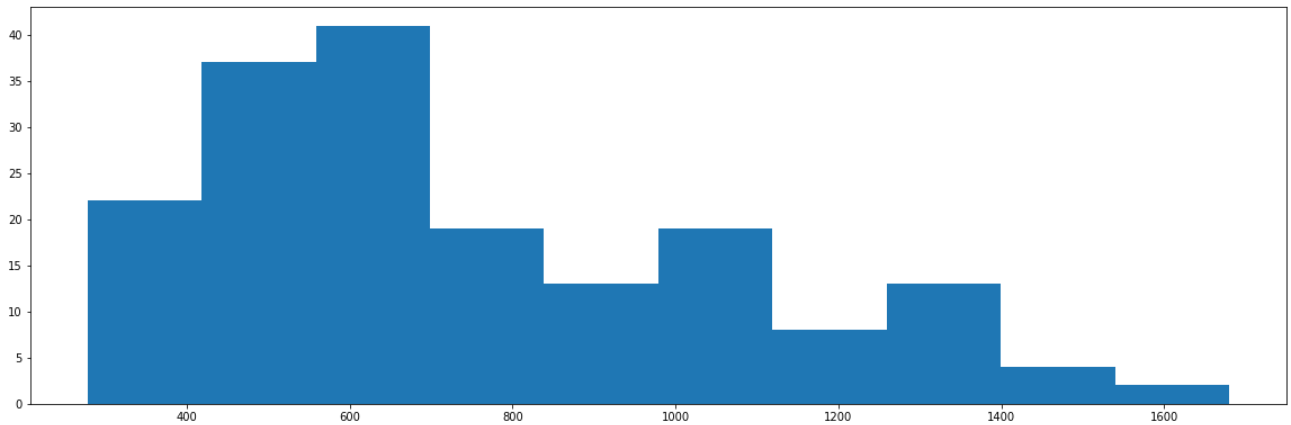

40
30
20
10


0
0.5
1.0
1.5
2.0
2.5
3.0
3.5


0
5
10
15
20
25
30
35
40
2
4
6
8
10
12


0
5
10
15
20
25
30
35
0.6
0.8
1.0
1.2
1.4
1.6


0
5
10
15
20
25
30
1.5
2.0
2.5
3.0
3.5
4.0


0
5
10
15
20
25
30
35
40
400
600
800
1000
1200
1400
1600

```
secwine_data=df
from sklearn import preprocessing
for i in range(1,secwine_data.shape[1]):
  column=df.columns[i]
  secwine_data[column]=preprocessing.scale(df[column])
secwine_data
```

	Wine	Alcohol	Malic.acid	Ash	Acl	Mg	Phenols	Flavanoids
0	1	1.518613	-0.562250	0.232053	-1.169593	1.913905	0.808997	1.034819
1	1	0.246290	-0.499413	-0.827996	-2.490847	0.018145	0.568648	0.733629
2	1	0.196879	0.021231	1.109334	-0.268738	0.088358	0.808997	1.215533
3	1	1.691550	-0.346811	0.487926	-0.809251	0.930918	2.491446	1.466525
4	1	0.295700	0.227694	1.840403	0.451946	1.281985	0.808997	0.663351
...	...	...	...	...	...	...	...	...
173	3	0.876275	2.974543	0.305159	0.301803	-0.332922	-0.985614	-1.424900
174	3	0.493343	1.412609	0.414820	1.052516	0.158572	-0.793334	-1.284344
175	3	0.332758	1.744744	-0.389355	0.151661	1.422412	-1.129824	-1.344582
176	3	0.209232	0.227694	0.012732	0.151661	1.422412	-1.033684	-1.354622
177	3	1.395086	1.583165	1.365208	1.502943	-0.262708	-0.392751	-1.274305

178 rows × 14 columns

```
SMean_result=list()
from sklearn import preprocessing
for i in range(1,secwine_data.shape[1]):
  column=secwine_data.columns[i]
  SMean_result.append(round(secwine_data[column].mean()))
print(SMean_result)
```

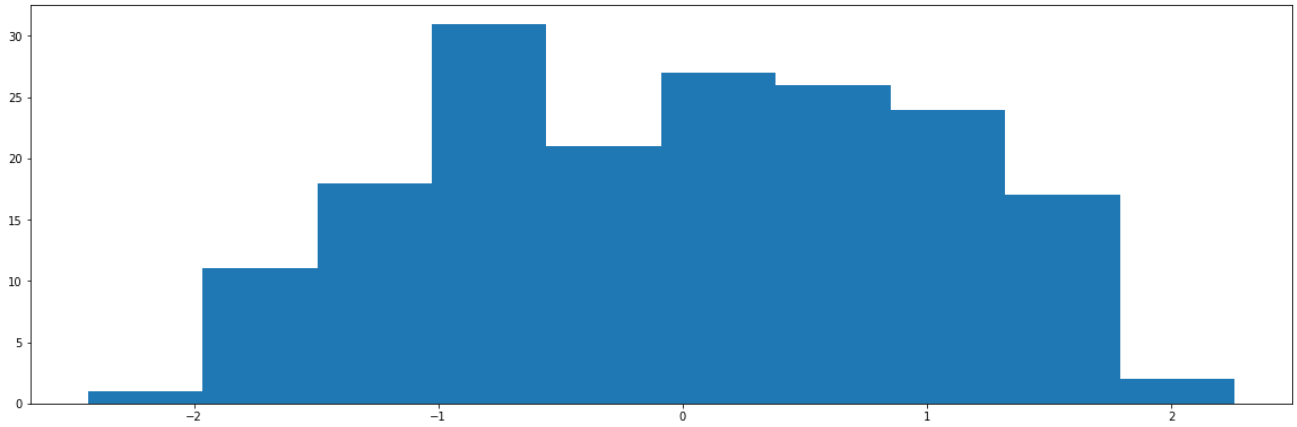
```
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

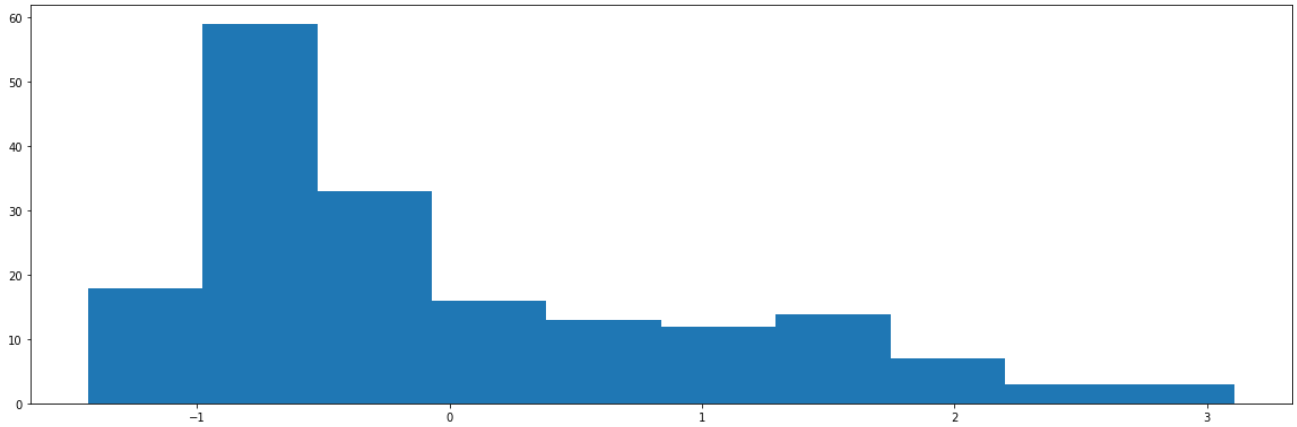
```
SSD_result=list()
from sklearn import preprocessing
for i in range(1,secwine_data.shape[1]):
  column=secwine_data.columns[i]
  SSD_result.append(round(secwine_data[column].std()))
print(SSD_result)
```

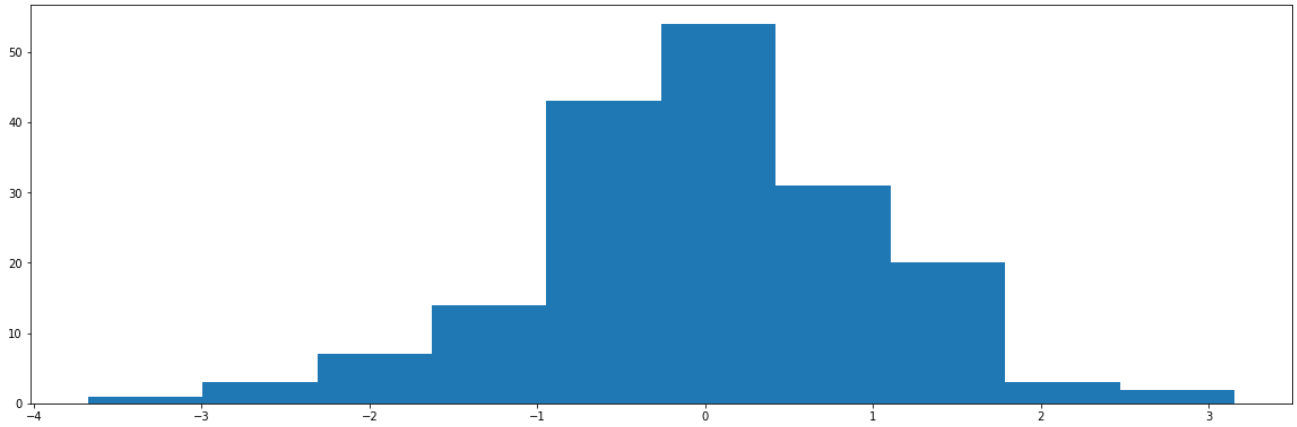
```
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
```

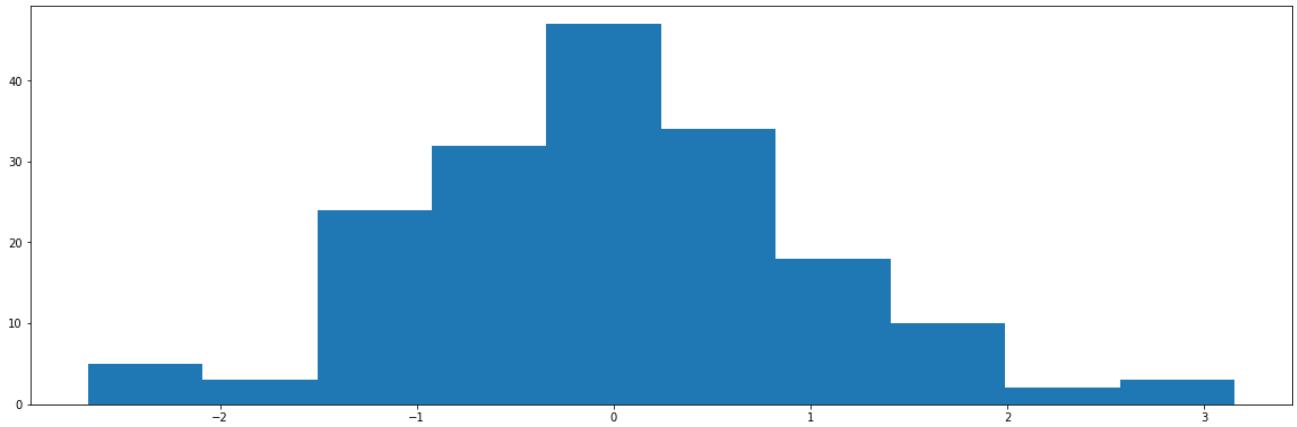
```
plt.figure(figsize = [20, 100])
for i in range(1,secwine_data.shape[1]):
  column=secwine_data.columns[i]
```

```
    column = secwine_data.columns[i]
    plt.subplot(13, 1, i)
    plt.hist(secwine_data[column])
```

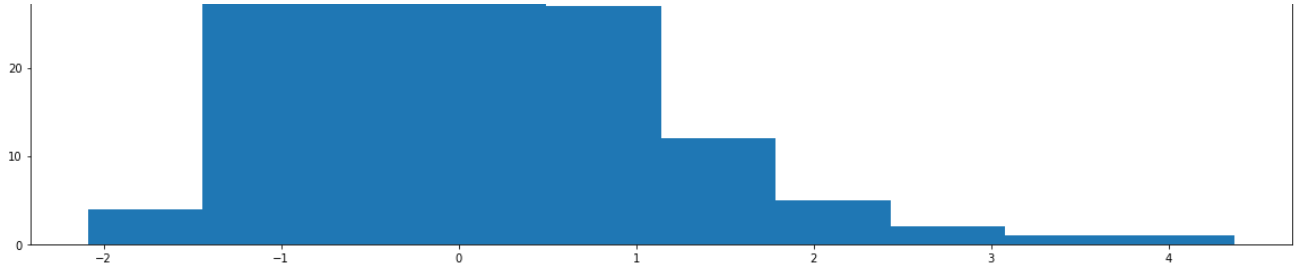

30
25
20
15
10
5
0
−2
−1
0
1
2

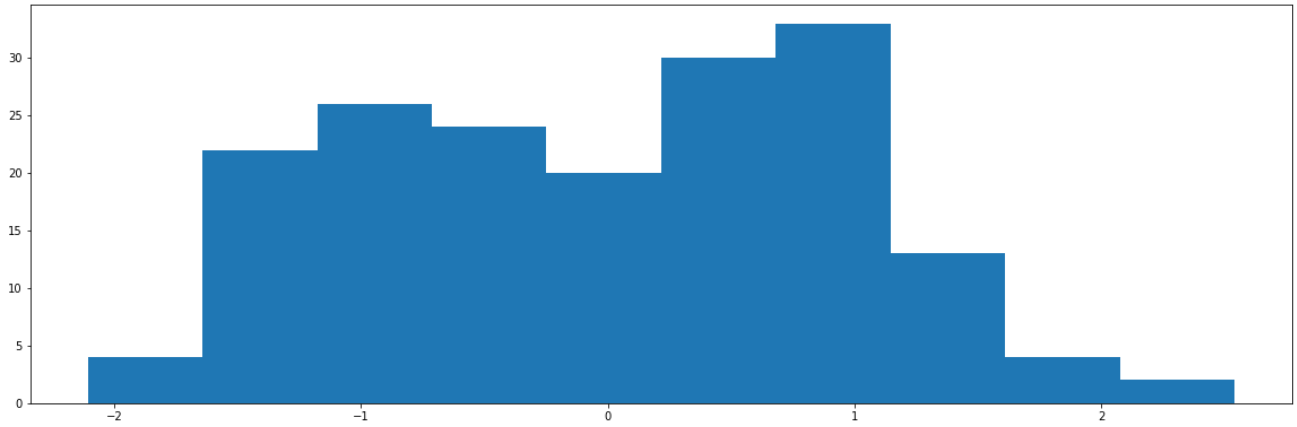

60
50
40
30
20
10
0
−1
0
1
2
3

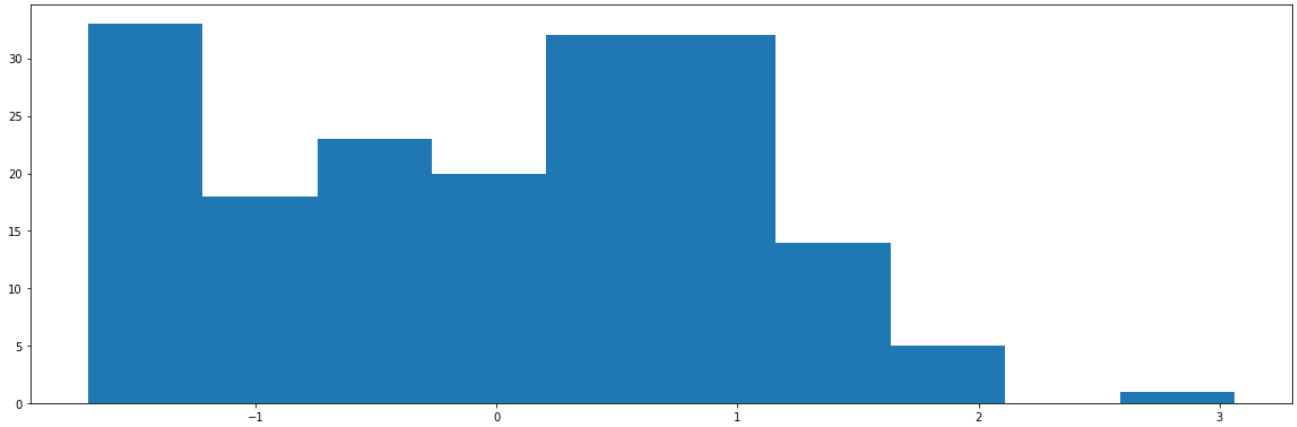

50
40
30
20
10
0
−4
−3
−2
−1
0
1
2
3

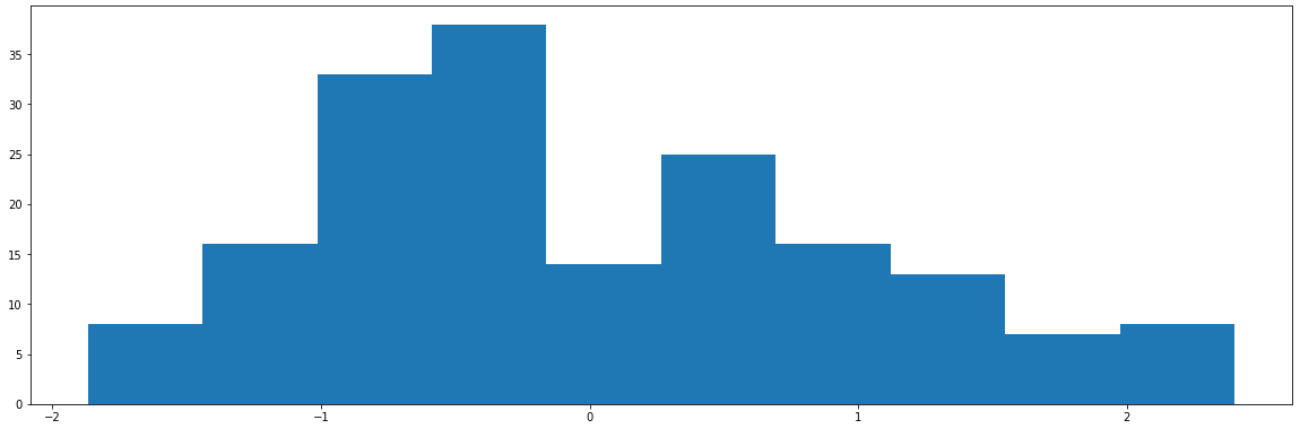

40
30
20
10
0
−2
−1
0
1
2
3


40
30


20
10
0
−2
−1
0
1
2
3
4

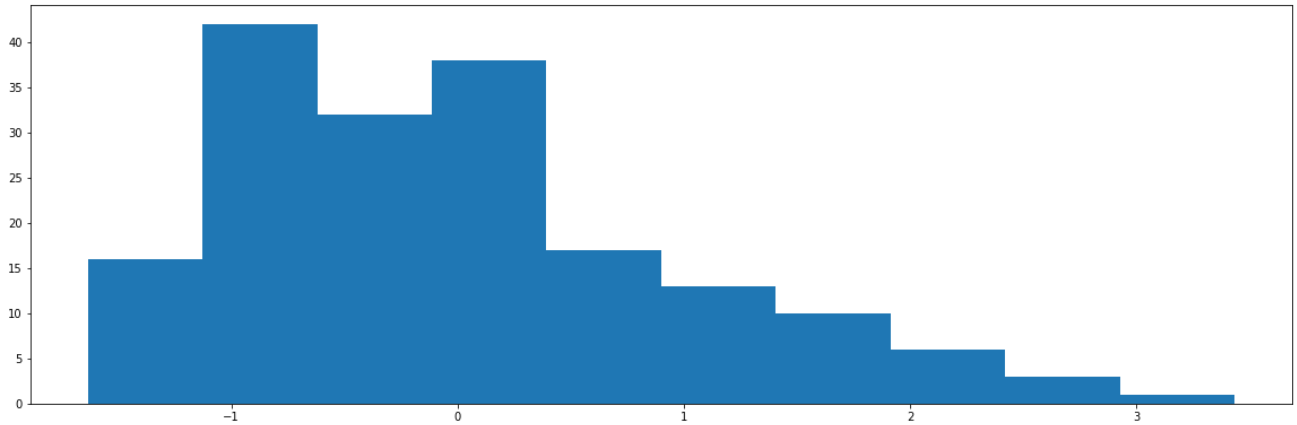

30
25
20
15
10
5
0
−2
−1
0
1
2

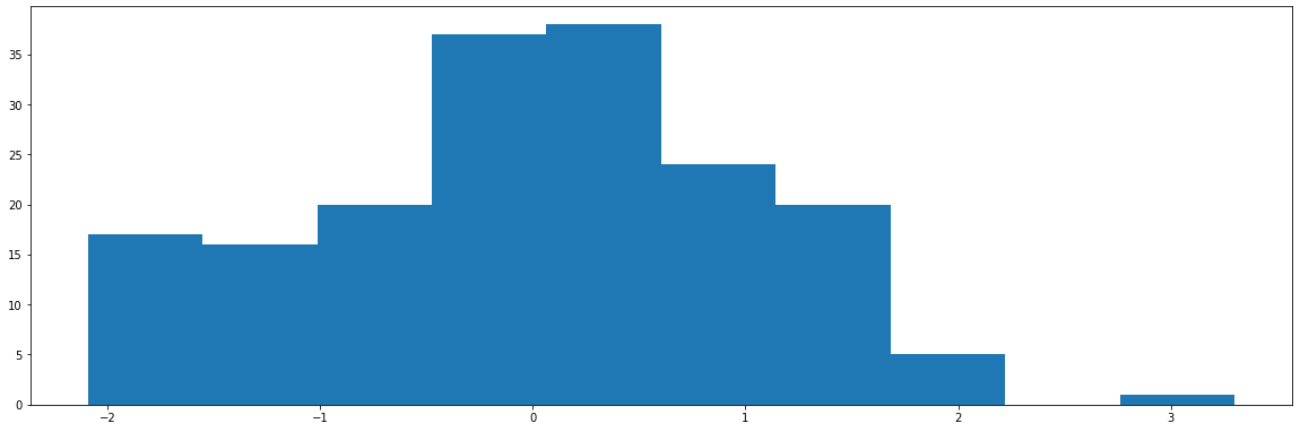

30
25
20
15
10
5
0
−1
0
1
2
3

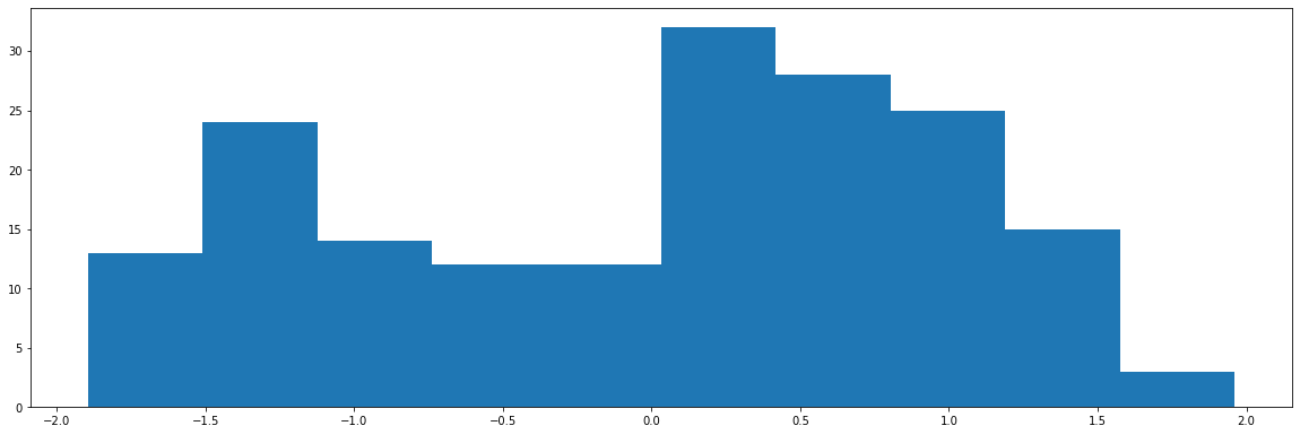

35
30
25
20
15
10
5
0
−2
−1
0
1
2

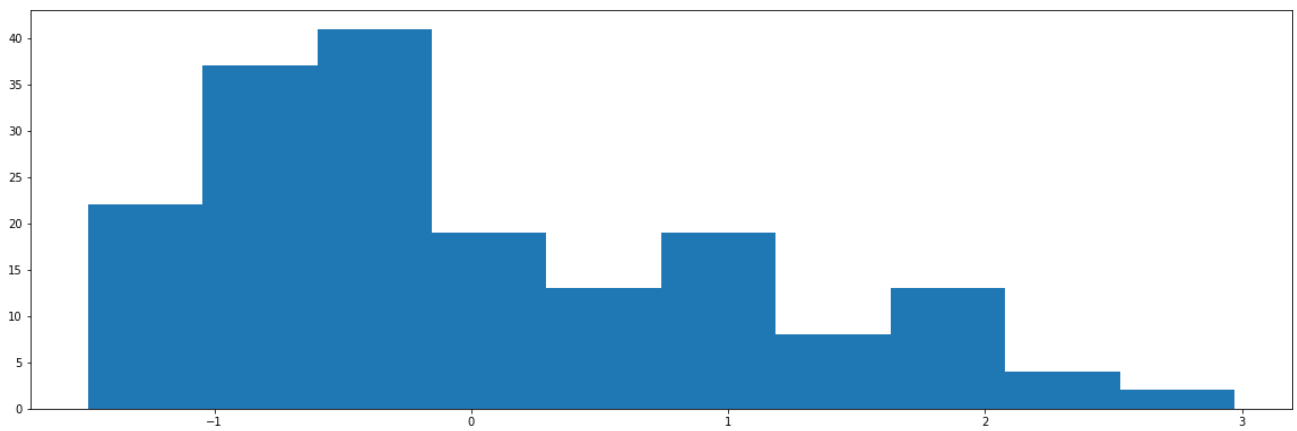

40
30
20
10


0
−2
−1
0
1
2
3


0
5
10
15
20
25
30
35
40
−1
0
1
2
3


0
5
10
15
20
25
30
35
−2
−1
0
1
2
3


0
5
10
15
20
25
30
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0


0
5
10
15
20
25
30
35
40
−1
0
1
2
3