```
Reg no-1081
Name-Rishitha Bandi
Date-12/10/2022
```

```
import pandas as pd
import numpy as np
```

```
data =pd.read_csv("/content/Enrollments_28092022.csv")
data
```

	StudentNo	DEGREE	INTERMEDIATE	SSC	INTERNSHIP
0	1001	8.10	76.0	92.0	Data Science
1	1002	8.10	76.0	92.0	MEAN Stack Web Development
2	1003	7.80	94.6	92.0	MEAN Stack Web Development
3	1004	9.03	89.5	89.0	Data Science
4	1005	8.38	87.0	90.0	MEAN Stack Web Development
...	...	...	...	...	...
292	2188	8.70	94.1	93.0	Data Science
293	2189	8.45	90.0	93.0	Data Science
294	2190	8.40	94.9	98.0	Data Science
295	2191	7.06	90.6	88.0	Cloud Computing Services (AWS)
296	2192	7.50	95.5	95.0	Cloud Computing Services (AWS)

297 rows × 5 columns

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 297 entries, 0 to 296
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   StudentNo     297 non-null    int64  
 1   DEGREE        297 non-null    float64
 2   INTERMEDIATE  297 non-null    float64
 3   SSC           297 non-null    float64
 4   INTERNSHIP    297 non-null    object 
dtypes: float64(3), int64(1), object(1)
memory usage: 11.7+ KB
```

```
rows=len(data)
```

```
print("no.of rows:",rows)
```

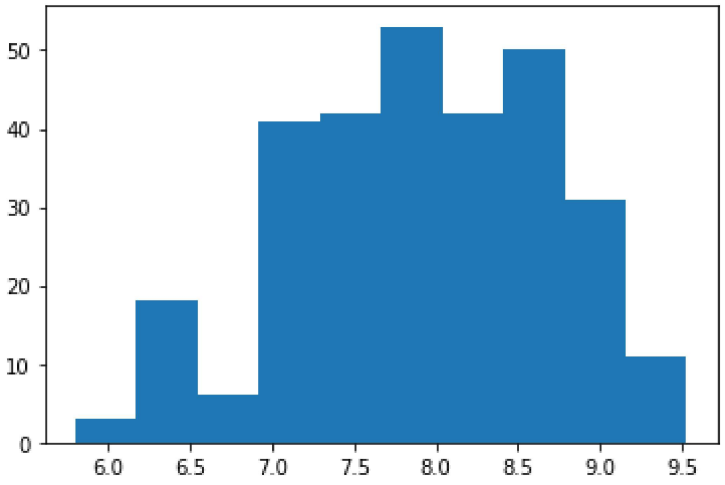
```
no.of rows: 297
```

```
cols=len(data.axes[1])
print("no.of columns:",str(cols))
```

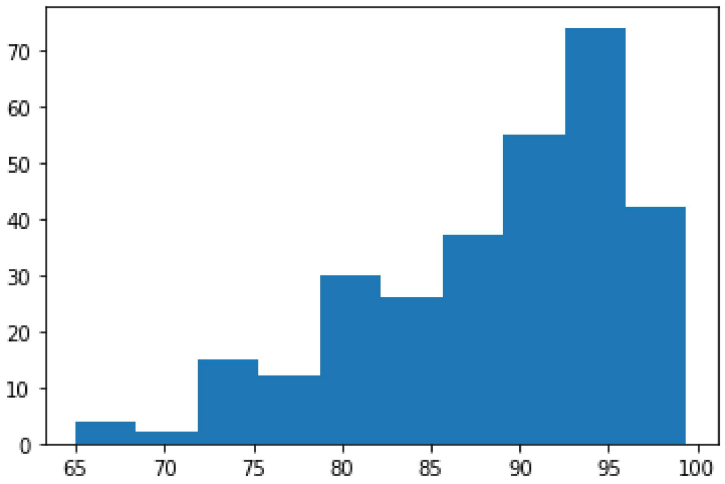
```
no.of columns: 5
```

```
import matplotlib.pyplot as plt
import statistics as stat
```

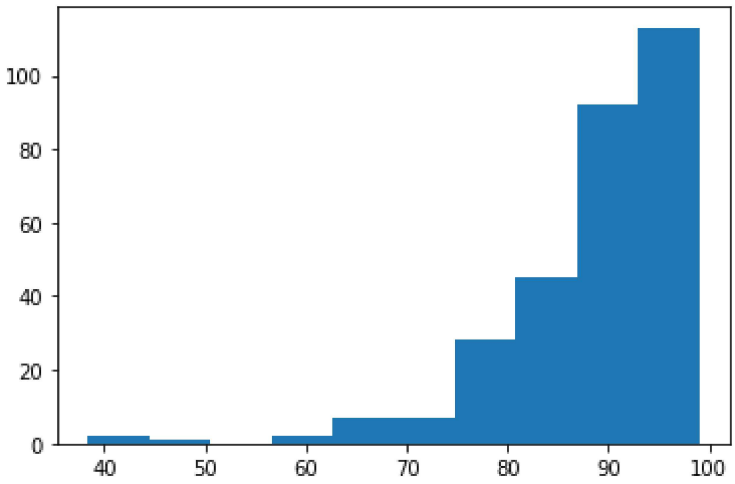
```
plt.hist(data["DEGREE"])
plt.show()
```




```
plt.hist(data["INTERMEDIATE"])
plt.show()
```



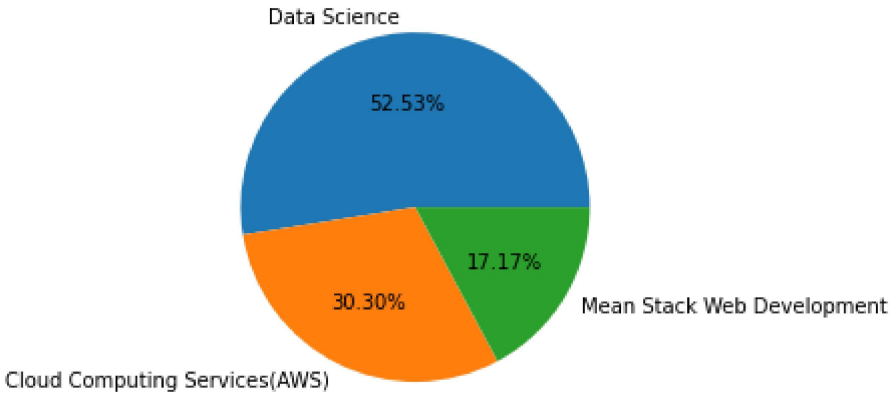

```
plt.hist(data["SSC"])
plt.show()
```




```
data["INTERNSHIP"].value_counts()
```

```
Data Science                       156
Cloud Computing Services (AWS)      90
MEAN Stack Web Development          51
Name: INTERNSHIP, dtype: int64
```

```
courses=["Data Science","Cloud Computing Services(AWS)","Mean Stack Web Development"]
students=[156,90,51]
plt.pie(students,labels=courses,autopct="%1.2f%%")
plt.show()
```




```
#Degree
print("Mean=",np.mean(data["DEGREE"]))
print("Median=",np.median(data["DEGREE"]))
print("Mode=",stat.mode(data["DEGREE"]))
```

```
Mean= 7.928080808080809
Median= 8.0
Mode= 7.0
```

```
#Intermediate
```

```
print("Mean=",np.mean(data["INTERMEDIATE"]))
print("Median=",np.median(data["INTERMEDIATE"]))
print("Mode=",stat.mode(data["INTERMEDIATE"]))
```

```
Mean= 88.66262626262626
Median= 90.8
Mode= 95.0
```

```
#10th class
print("Mean=",np.mean(data["SSC"]))
print("Median=",np.median(data["SSC"]))
print("Mode=",stat.mode(data["SSC"]))
```

```
Mean= 88.10673400673402
Median= 90.0
Mode= 95.0
```

```
df = lambda x:np.std(x, ddof=1)/np.mean(x)*100
```

```
#Degree
print("Range=",max(data["DEGREE"])-min(data["DEGREE"]))
print("co-effecient of variations =",df(data["DEGREE"]))
data["DEGREE"].describe()
```

```
Range= 3.7299999999999995
co-effecient of variations = 9.90881225818308
count    297.000000
mean       7.928081
std        0.785579
min        5.800000
25%        7.400000
50%        8.000000
75%        8.560000
max        9.530000
Name: DEGREE, dtype: float64
```

```
#Intermediate
print("Range=",max(data["INTERMEDIATE"])-min(data["INTERMEDIATE"]))
print("co-effecient of variations=",df(data["INTERMEDIATE"]))
data["INTERMEDIATE"].describe()
```

```
Range= 34.400000000000006
co-effecient of variations= 8.29631726338337
count    297.000000
mean      88.662626
std        7.355733
min       65.000000
25%       83.000000
50%       90.800000
75%       94.600000
```

```
max       99.400000
Name: INTERMEDIATE, dtype: float64
```

```
#10th class
print("Range=",max(data["SSC"])-min(data["SSC"]))
print("co-effecient of variation=",df(data["SSC"]))
data["SSC"].describe()
```

```
Range= 60.6
co-effecient of variation= 10.24664491920062
count    297.000000
mean      88.106734
std        9.027984
min       38.400000
25%       85.000000
50%       90.000000
75%       95.000000
max       99.000000
Name: SSC, dtype: float64
```

```
def outlier(a):
  q1 = np.quantile(a,0.25)
  q3 = np.quantile(a,0.75)
  med = np.median(a)
  iqr = q3-q1
  upper_bound = q3+(1.5*iqr)
  lower_bound = q1-(1.5*iqr)
  print(iqr,upper_bound,lower_bound)
  print("Inter-Quartile Range:",iqr)
  outliers = a[(a<=lower_bound) | (a>=upper_bound)]
  print("The following are the outliers in the boxplot:\n{}".format(outliers))
```

```
#Degree
outlier(data["DEGREE"])
```

```
1.1600000000000001 10.3 5.66
Inter-Quartile Range: 1.1600000000000001
The following are the outliers in the boxplot:
Series([], Name: DEGREE, dtype: float64)
```

```
#Intermediate
outlier(data["INTERMEDIATE"])
```

```
11.599999999999994 111.99999999999999 65.60000000000001
Inter-Quartile Range: 11.599999999999994
The following are the outliers in the boxplot:
271    65.0
Name: INTERMEDIATE, dtype: float64
```

```
#10th class
outlier(data["SSC"])
```

```
10.0 110.0 70.0
Inter-Quartile Range: 10.0
The following are the outliers in the boxplot:
5      64.0
7      70.0
31     60.0
51     68.0
69     60.0
82     65.6
86     50.0
107    64.0
236    38.4
237    67.0
243    40.2
270    65.0
288    65.0
Name: SSC, dtype: float64
```

```
import scipy.stats as stats
```

```
print("Standard Scores of Degree:")
print(stats.zscore(data["DEGREE"]))
```

```
Standard Scores of Degree:
0      0.219213
1      0.219213
2     -0.163315
3      1.405052
4      0.576240
          ...
292    0.984271
293    0.665497
294    0.601742
295   -1.106886
296   -0.545844
Name: DEGREE, Length: 297, dtype: float64
```

```
print("Standard Scores of Intermediate:")
print(stats.zscore(data["INTERMEDIATE"]))
```

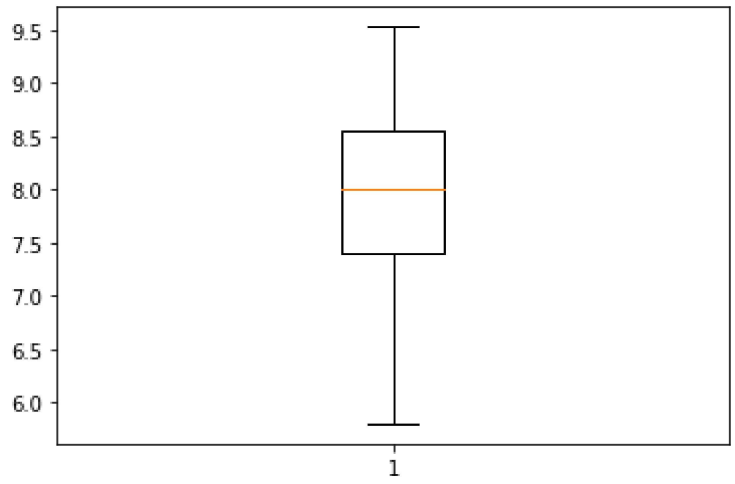
```
Standard Scores of Intermediate:
0     -1.724369
1     -1.724369
2      0.808539
3      0.114032
4     -0.226413
          ...
292    0.740450
293    0.182121
```

```
294    0.849392
295    0.263827
296    0.931099
Name: INTERMEDIATE, Length: 297, dtype: float64
```

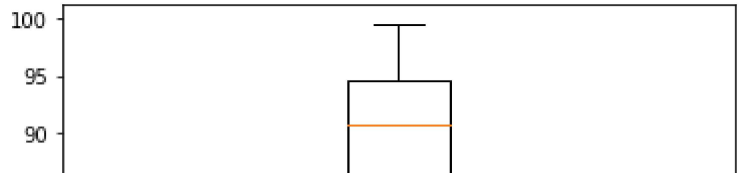
```
print("Standard Scores of 10th class:")
print(stats.zscore(data["SSC"]))
```

```
Standard Scores of 10th class:
0      0.431972
1      0.431972
2      0.431972
3      0.099111
4      0.210065
         ...
292    0.542926
293    0.542926
294    1.097694
295   -0.011843
296    0.764833
Name: SSC, Length: 297, dtype: float64
```

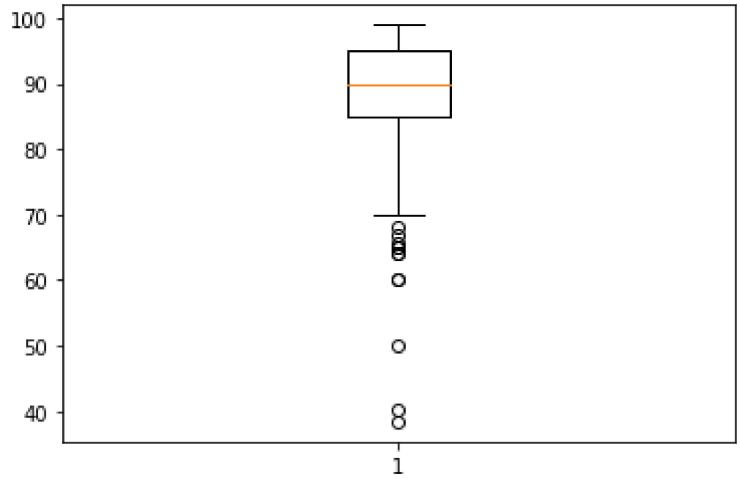
```
plt.boxplot(data["DEGREE"])
plt.show()
```




```
plt.boxplot(data["INTERMEDIATE"])
plt.show()
```




```
plt.boxplot(data["SSC"])
plt.show()
```




```
def func(b):
  q9 = np.quantile(b,0.9)
  li=b[b==q9]
  print("No.of students with 90% percentile:",li.count())


#Degree
func(data['DEGREE'])
```

```
     No.of students with 90% percentile: 3
```

```
#Intermediate
func(data['INTERMEDIATE'])
```

```
     No.of students with 90% percentile: 3
```

```
#10th Class
func(data['SSC'])
```

```
     No.of students with 90% percentile: 19
```

Colab paid products - Cancel contracts here

✓ 0s completed at 7:27 PM