```
#REG NO:1076_DS-01
#NAME:KUNAPURI UMESHCHANDRA
#DATE:13-10-2022
#IDENTEFY VARIABLES AND THEIR TYPES (QUANTITATIVE OR QUALITATIVE)
import numpy as np
import pandas as pd
```

```
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
```

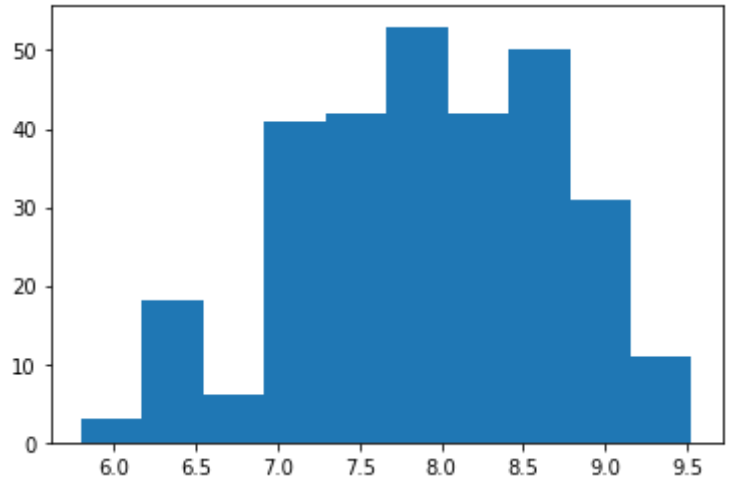
```
#Size of Data (No. of Columns and Rows)
rows=len(data)
cols=len(data.axes[1])
```

```
print("rows:",str(rows))
print("cols:",str(cols))
```

```
rows: 297
cols: 5
```

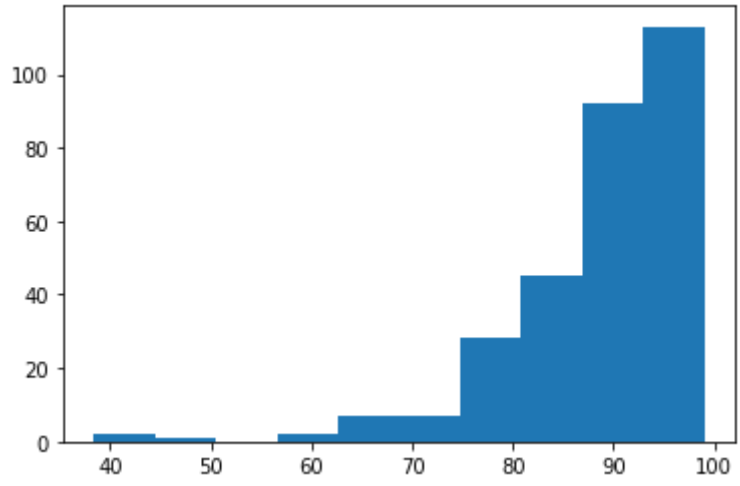
```
import matplotlib.pyplot as plt
import statistics as stat


# Prepare Histogram for Degree, Inter and 10th Class
plt.hist(data['DEGREE'])
plt.show
```

```
<function matplotlib.pyplot.show(*args, **kw)>
```



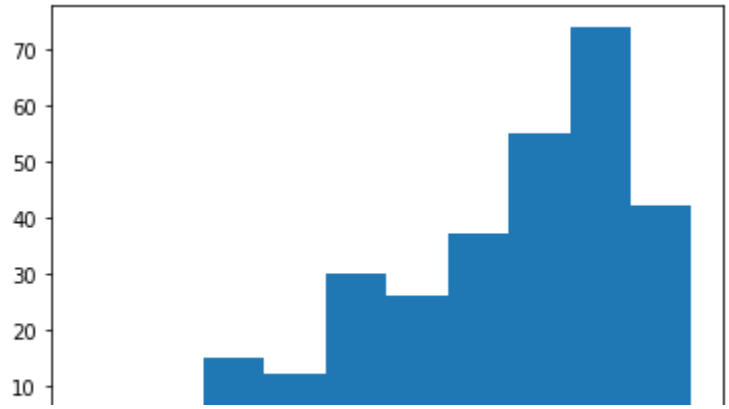

```
plt.hist(data['SSC'])
plt.show
```

```
<function matplotlib.pyplot.show(*args, **kw)>
```

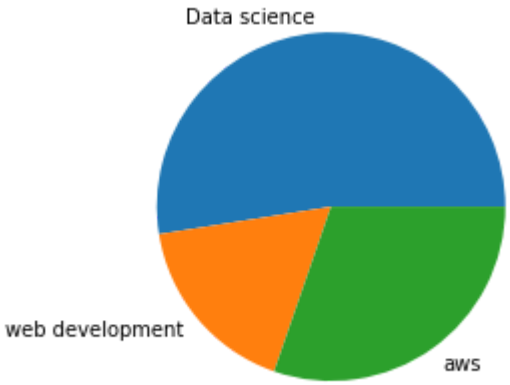



```
plt.hist(data['INTERMEDIATE'])
plt.show
```

```
<function matplotlib.pyplot.show(*args, **kw)>
```




```
#Create Pie-Chart to represent the Enrollments for each Internship Program
interncourses=['Data science','web development','aws']
enrollments=[156,51,90]
plt.pie(enrollments,labels=interncourses)
plt.show()
```




```
#Find No. of Enrollments for each Internship Program
data['INTERNSHIP'].value_counts()
```

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

```
#Find Measure of Central Tendency: MEAN, MEDIAN, MODE for Degree,Inter and 10th
print("DEGREE")
print("mean=",np.mean(data['DEGREE']))
print("median=",np.median(data['DEGREE']))
print("mode=",stat.mode(data['DEGREE']))
print("INTERMEDIATE")
print("mean=",np.mean(data['INTERMEDIATE']))
print("median=",np.median(data['INTERMEDIATE']))
print("mode=",stat.mode(data['INTERMEDIATE']))
print("SSC")
print("mean=",np.mean(data['SSC']))
print("median=",np.median(data['SSC']))
```

```
print("mode=",stat.mode(data['SSC']))
```

```
DEGREE
mean= 7.928080808080809
median= 8.0
mode= 7.0
INTERMEDIATE
mean= 88.66262626262626
median= 90.8
mode= 95.0
SSC
mean= 88.10673400673402
median= 90.0
mode= 95.0
```

```
#Find Measure of Variance: Minimum, Maximum, Range, Mean Deviation,
#Standard Deviation, Co-efficient of Variation for Degree, Inter and 10th
cv= lambda x: np.std(x, ddof=1)/np.mean(x)*100
print("DEGREE")
print("Range=",max(data['DEGREE'])-min(data['DEGREE']))
print("Co-efficient of variation=",cv(data['DEGREE']))
data['DEGREE'].describe()
```

```
DEGREE
Range= 3.7299999999999995
Co-efficient of variation= 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
```

```
print("INTERMEDIATE")
print("Range=",max(data['DEGREE'])-min(data['INTERMEDIATE']))
print("Co-efficient of variation=",cv(data['INTERMEDIATE']))
data['INTERMEDIATE'].describe()
```

```
INTERMEDIATE
Range= -55.47
Co-efficient of variation= 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
```

```
print("SSC")
print("Range=",max(data['DEGREE'])-min(data['SSC']))
print("Co-efficient of variation=",cv(data['SSC']))
data['SSC'].describe()
```

```
SSC
Range= -28.869999999999997
Co-efficient 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
```

```
#Measures of Position: Standard Scores for Degree, Inter and 10th
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 Ssc")
```

```
print(stats.zscore(data['SSC']))
```

```
     Standard scores of Ssc
     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
```

```
#Identify Outliers for Degree, Inter and 10th
def outlier(a):
  q1 = np.quantile(a,0.25)
  q2 = np.quantile(a,0.75)
  m = np.median(a)
  iqr = q2-q1
  u_bound = q2+(1.5*iqr)
  l_bound = q1-(1.5*iqr)
  print(iqr,u_bound,l_bound)
  print("Inter Quartile Range:",iqr)
  outliers = a[(a<= l_bound)|(a>= u_bound)]
  print("outliers in boxplot:\n{}".format(outliers))
outlier(data['DEGREE'])
```

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

```
outlier(data['INTERMEDIATE'])
```

```
     11.599999999999994 111.99999999999999 65.60000000000001
     Inter Quartile Range: 11.599999999999994
     outliers in boxplot:
     271    65.0
     Name: INTERMEDIATE, dtype: float64
```

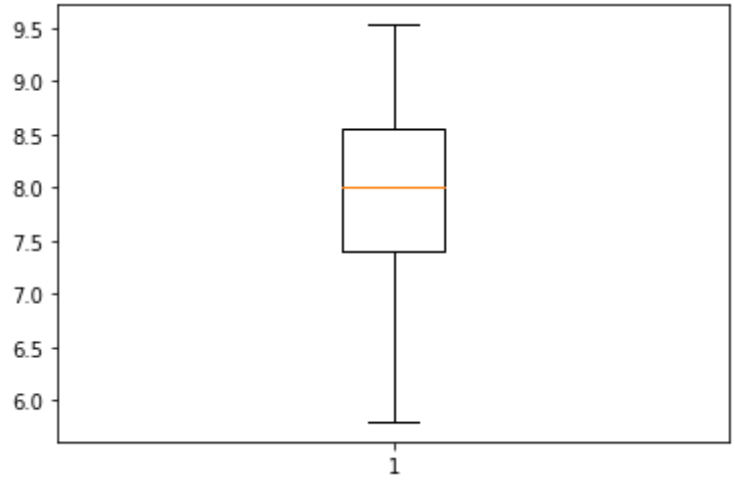
```
outlier(data['SSC'])
```

```
     10.0 110.0 70.0
     Inter Quartile Range: 10.0
     outliers in 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
```

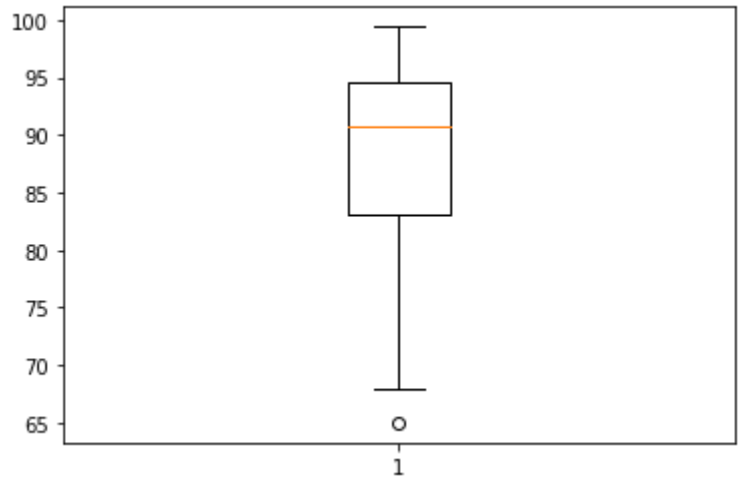
```
#Create Box Plot for Degree, Inter and 10th
plt.boxplot(data['DEGREE'])
plt.show
```

```
<function matplotlib.pyplot.show(*args, **kw)>
```



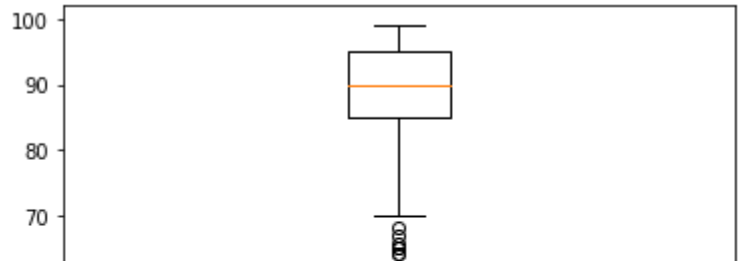

```
plt.boxplot(data['INTERMEDIATE'])
plt.show
```

```
<function matplotlib.pyplot.show(*args, **kw)>
```




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

```
<function matplotlib.pyplot.show(*args, **kw)>
```




```
#Identify No. of Students with 90% percentile for Degree, Inter and 10th Class
def func(c):
  quantile = np.quantile(c, 0.9)
  Data=c[c==quantile]
  print("Students with 90% percentile:",Data.count())
func(data['DEGREE'])
```

```
Students with 90% percentile: 3
```

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

```
Students with 90% percentile: 3
```

```
func(data['SSC'])
```

```
Students with 90% percentile: 19
```

Colab paid products - Cancel contracts here