Reg No:3010

Name:Tarun Mallik

Date:19/10/2022

```
In [33]: #Importing Libraries
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
```

Loading Dataset

```
In [34]: data =pd.read_csv("/content/Enrollments_28092022.csv")
         data
```

Out[34]:

	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

Q1. Identify Variables and their Types (Quantitative or Qualitative)

In [5]: `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
```

In [35]: `data['StudentNo'] = data['StudentNo'].apply(str)`

In [36]: `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    object
 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), object(2)
memory usage: 11.7+ KB
```

In [38]: `data.describe()`

Out[38]:

	DEGREE	INTERMEDIATE	SSC
count	297.000000	297.000000	297.000000
mean	7.928081	88.662626	88.106734
std	0.785579	7.355733	9.027984
min	5.800000	65.000000	38.400000
25%	7.400000	83.000000	85.000000
50%	8.000000	90.800000	90.000000
75%	8.560000	94.600000	95.000000
max	9.530000	99.400000	99.000000

Q1. Answer

Qualitative Data : StudentNo and Internship

Quantitative Data : Degree, Intermediate and SSC

Q2. Size of Data (No. of Rows and Columns)
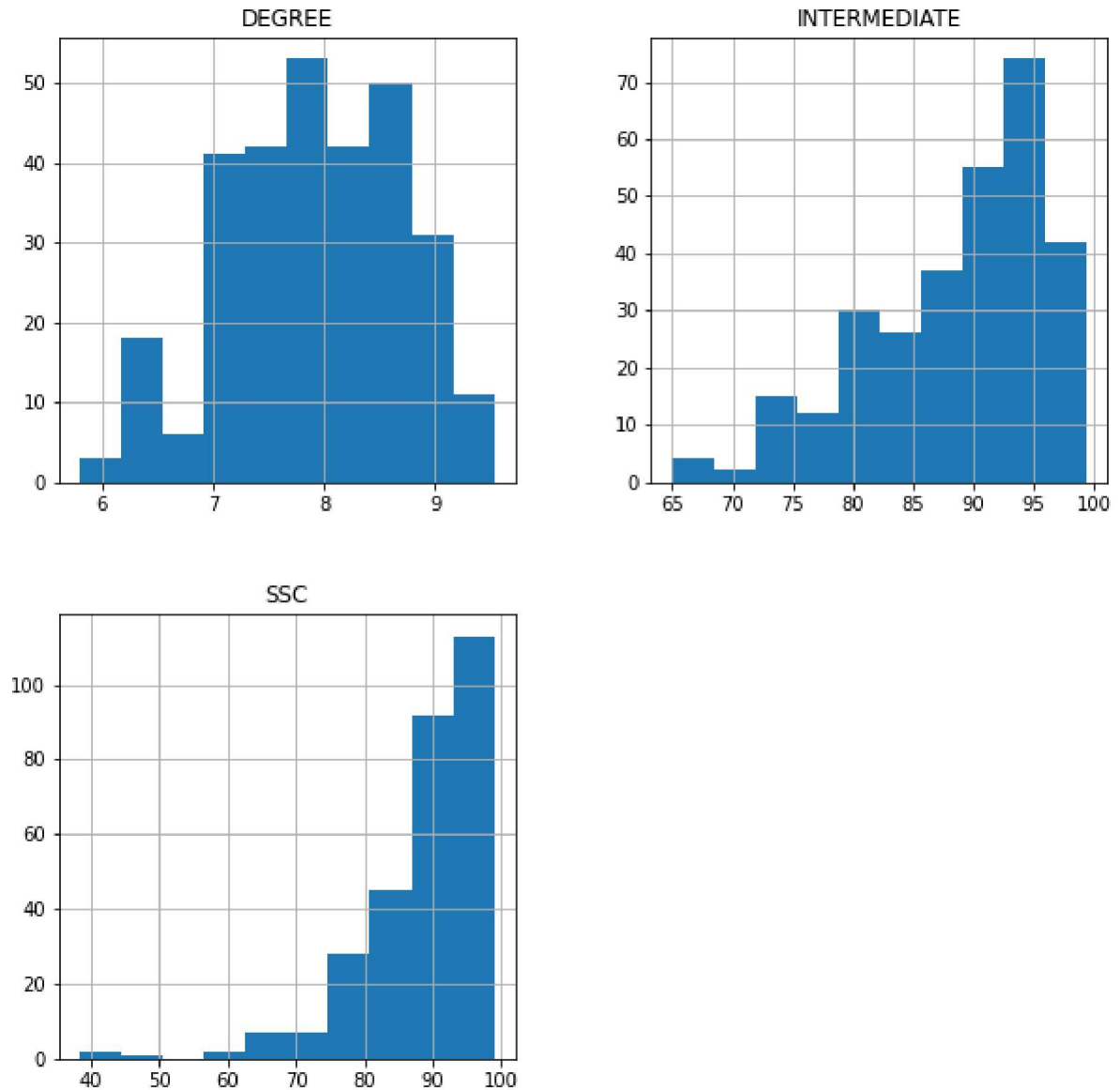
```
In [40]: data.shape
Out[40]: (297, 5)
```

Q2. Answer

Rows: 297

Attributes: 5

Q3. Create Histogram

In [42]: `data.hist(figsize=(10,10))`

Out[42]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7fac892f8e10>,
                 <matplotlib.axes._subplots.AxesSubplot object at 0x7fac892bb290>],
                [<matplotlib.axes._subplots.AxesSubplot object at 0x7fac892ef890>,
                 <matplotlib.axes._subplots.AxesSubplot object at 0x7fac892a5e90>]],
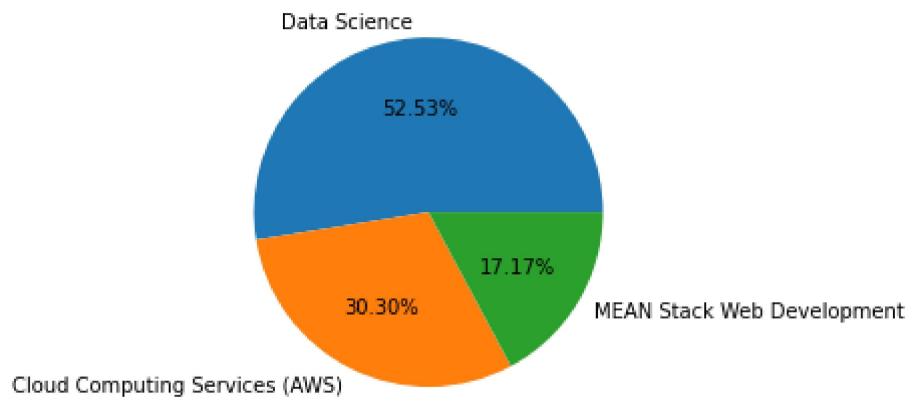   dtype=object)

DEGREE

INTERMEDIATE

SSC

Q4. Create Pie-Chart to represent the Enrollments for each Internship Program

In [48]: `data['INTERNSHIP'].value_counts()`

Out[48]:
```
Data Science                     156
Cloud Computing Services (AWS)    90
MEAN Stack Web Development         51
Name: INTERNSHIP, dtype: int64
```

```
In [46]: courses=['Data Science','Cloud Computing Services (AWS)','MEAN Stack Web Devel
         opment']
         students=[156,90,51]
         plt.pie(students, labels = courses,autopct='%1.2f%%')
         plt.show
```

Out[46]: `<function matplotlib.pyplot.show(*args, **kw)>`

Q5. Find No. of Enrollments for each Internship Program

```
In [49]: data['INTERNSHIP'].value_counts()
```

```
Out[49]: Data Science                    156
         Cloud Computing Services (AWS)   90
         MEAN Stack Web Development        51
         Name: INTERNSHIP, dtype: int64
```

Q6. Find Measure of Central Tendency: MEAN, MEDIAN, MODE

```
In [50]: # MEAN
         print(data.mean(numeric_only= True))
```

```
DEGREE           7.928081
INTERMEDIATE    88.662626
SSC             88.106734
dtype: float64
```

```
In [52]: # MEDIAN
         print(data.median(numeric_only= True))
```

```
DEGREE           8.0
INTERMEDIATE    90.8
SSC             90.0
dtype: float64
```

```
In [53]:  # MODE
          print(data.mode(numeric_only= True))
```

```
       DEGREE   INTERMEDIATE    SSC
0      7.0             95.0   95.0
```

Q7. Find Measure of Variance: Minimum, Maximum, Range, Mean Deviation, Standard Deviation, Co- efficient of Variation

```
In [55]:  # Minimum
          print(data.min(numeric_only= True))
```

```
DEGREE            5.8
INTERMEDIATE     65.0
SSC              38.4
dtype: float64
```

```
In [56]:  # Maximum
          print(data.max(numeric_only= True))
```

```
DEGREE            9.53
INTERMEDIATE     99.40
SSC              99.00
dtype: float64
```

```
In [58]:  # Range
          print(data.max(numeric_only= True)-data.min(numeric_only= True))
```

```
DEGREE            3.73
INTERMEDIATE     34.40
SSC              60.60
dtype: float64
```

```
In [59]:  # Standard  Deviation
          print(data.std(numeric_only= True))
```

```
DEGREE          0.785579
INTERMEDIATE    7.355733
SSC             9.027984
dtype: float64
```

```
In [60]:  # Co - effienct of Variation
          print(data.std(numeric_only= True)/data.mean(numeric_only= True))
```

```
DEGREE          0.099088
INTERMEDIATE    0.082963
SSC             0.102466
dtype: float64
```

Q8. Measures of Position: Standard Scores, Inter-quartile Range for Degree, Inter and 10th

In [61]:
```python
# 1st Quartile
data.quantile(q=0.25, numeric_only=True)
```

Out[61]:
```
DEGREE            7.4
INTERMEDIATE     83.0
SSC              85.0
Name: 0.25, dtype: float64
```

In [63]:
```python
# 2nd Quartile
data.quantile(q=0.5, numeric_only=True)
```

Out[63]:
```
DEGREE            8.0
INTERMEDIATE     90.8
SSC              90.0
Name: 0.5, dtype: float64
```

In [65]:
```python
# 3rd Quartile
data.quantile(q=0.75, numeric_only=True)
```

Out[65]:
```
DEGREE            8.56
INTERMEDIATE     94.60
SSC              95.00
Name: 0.75, dtype: float64
```
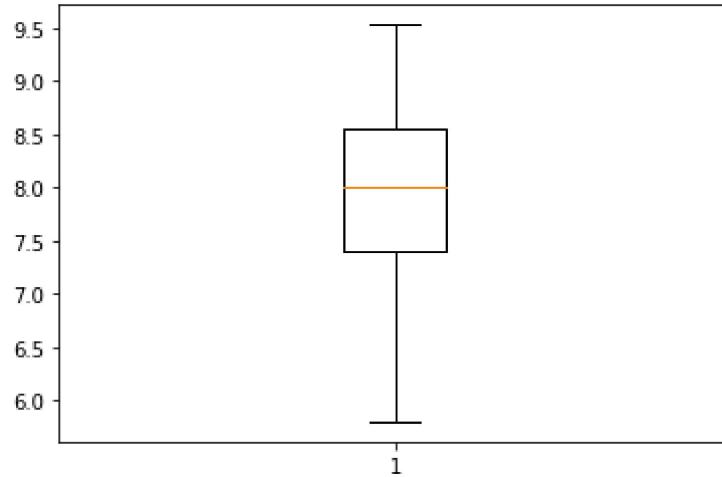
In [66]:
```python
# Inter-Quartile = Q3 - Q1
data.quantile(q=0.75, numeric_only=True) - data.quantile(q=0.25, numeric_only=True)
```

Out[66]:
```
DEGREE            1.16
INTERMEDIATE     11.60
SSC              10.00
dtype: float64
```

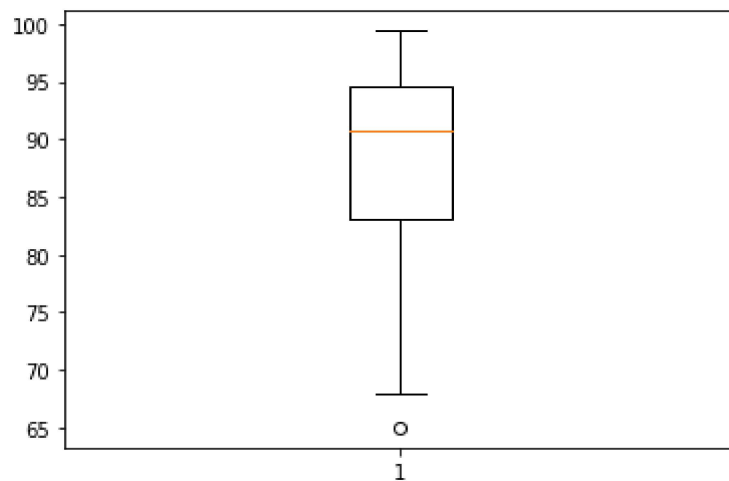Q9. Create Box Plot and Identify Outliers

In [67]:
```
plt.boxplot(data['DEGREE'])
plt.show
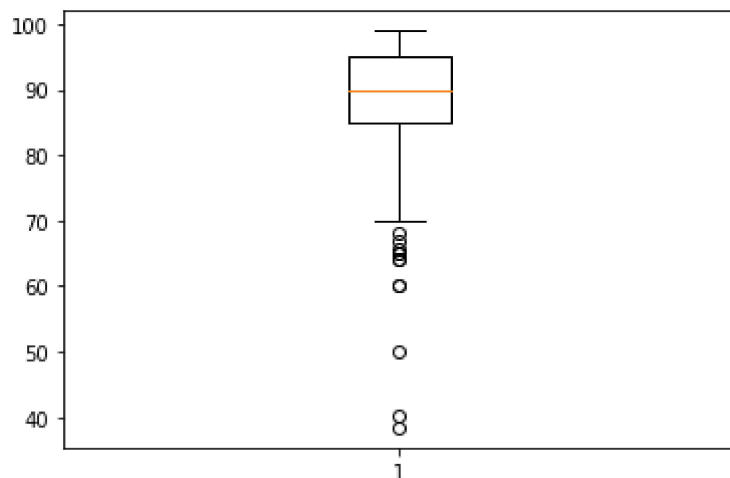```

Out[67]: `<function matplotlib.pyplot.show(*args, **kw)>`



In [68]: 
```
plt.boxplot(data['INTERMEDIATE'])
plt.show
```

Out[68]: `<function matplotlib.pyplot.show(*args, **kw)>`

In [69]:
```
plt.boxplot(data['SSC'])
plt.show
```

Out[69]: <function matplotlib.pyplot.show(*args, **kw)>



Q10. Identify No. of Students with 90% percentile for Degree, Inter and 10th Class

In [70]:
```
# 90th Percentile or Quantile
data.quantile(q=0.9, numeric_only=True)
```

Out[70]:
```
DEGREE   8.9
INTERMEDIATE   96.5
SSC   97.0
Name: 0.9, dtype: float64
```

In [75]:
```
df1 = data[data.DEGREE>8.9]
df1.head(10)
```

Out[75]:

|    | StudentNo | DEGREE | INTERMEDIATE | SSC | INTERNSHIP |
|----|-----------|--------|--------------|------|------------|
| 3  | 1004 | 9.03 | 89.5 | 89.0 | Data Science |
| 14 | 1015 | 9.07 | 95.0 | 98.0 | Cloud Computing Services (AWS) |
| 16 | 1017 | 9.53 | 98.2 | 93.1 | MEAN Stack Web Development |
| 17 | 1018 | 9.08 | 95.2 | 83.0 | MEAN Stack Web Development |
| 19 | 1020 | 9.16 | 98.0 | 87.0 | Data Science |
| 20 | 1021 | 9.00 | 96.8 | 99.0 | Data Science |
| 24 | 1025 | 9.08 | 97.2 | 97.0 | Data Science |
| 28 | 1029 | 8.92 | 89.7 | 83.0 | Cloud Computing Services (AWS) |
| 39 | 1040 | 9.03 | 96.5 | 97.0 | Data Science |
| 40 | 1041 | 9.34 | 97.7 | 97.0 | Data Science |

In [85]: 
```
df1.shape
```

Out[85]: (29, 5)

In [83]: 
```
df2 = data[data.INTERMEDIATE>96.5]
df2.head(10)
```

Out[83]:

|     | StudentNo | DEGREE | INTERMEDIATE | SSC  | INTERNSHIP |
| --- | --- | --- | --- | --- | --- |
| 16  | 1017 | 9.53 | 98.2 | 93.1 | MEAN Stack Web Development |
| 19  | 1020 | 9.16 | 98.0 | 87.0 | Data Science |
| 20  | 1021 | 9.00 | 96.8 | 99.0 | Data Science |
| 24  | 1025 | 9.08 | 97.2 | 97.0 | Data Science |
| 40  | 1041 | 9.34 | 97.7 | 97.0 | Data Science |
| 60  | 1061 | 8.80 | 97.3 | 92.0 | Data Science |
| 96  | 1097 | 9.20 | 97.8 | 90.0 | Data Science |
| 127 | 2023 | 9.20 | 97.2 | 92.0 | MEAN Stack Web Development |
| 138 | 2034 | 9.38 | 98.2 | 87.0 | Data Science |
| 154 | 2050 | 9.20 | 97.2 | 90.0 | Cloud Computing Services (AWS) |

In [84]: 
```
df2.shape
```

Out[84]: (29, 5)

In [80]: 
```
df3 = data[data.SSC   > 97.0]
df3.head(10)
```

Out[80]:

|     | StudentNo | DEGREE | INTERMEDIATE | SSC  | INTERNSHIP |
| --- | --- | --- | --- | --- | --- |
| 9   | 1010 | 8.30 | 87.0 | 98.0 | MEAN Stack Web Development |
| 14  | 1015 | 9.07 | 95.0 | 98.0 | Cloud Computing Services (AWS) |
| 20  | 1021 | 9.00 | 96.8 | 99.0 | Data Science |
| 21  | 1022 | 8.00 | 95.2 | 98.0 | Data Science |
| 76  | 1077 | 8.70 | 93.0 | 98.0 | Data Science |
| 95  | 1096 | 8.83 | 94.7 | 98.0 | Data Science |
| 122 | 2018 | 7.00 | 92.0 | 98.0 | Cloud Computing Services (AWS) |
| 126 | 2022 | 7.03 | 72.0 | 98.0 | Cloud Computing Services (AWS) |
| 145 | 2041 | 8.76 | 94.5 | 98.0 | Data Science |
| 147 | 2043 | 8.60 | 94.0 | 98.0 | MEAN Stack Web Development |

In [82]: 
```
df3.shape
```

Out[82]: (21, 5)