

```
pip install pandas
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/r
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (1.3
Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dis
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (fr
```

```
import pandas as pd
import numpy as np
import seaborn as sns
```

```
from google.colab import files
uploaded=files.upload()
```

Enrollments...092022.csv

- **Enrollments\_28092022.csv**(text/csv) - 11329 bytes, last modified: 10/6/2022 - 100% done  
Saving Enrollments\_28092022.csv to Enrollments\_28092022.csv

```
dataset = pd.read_csv('Enrollments_28092022.csv')
```

```
dataset
```

	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

## 1. Identify Variables and their types(Qualitative and Quantitative analysis)

```
dataset.columns
```

```
Index(['StudentNo', 'DEGREE', 'INTERMEDIATE', 'SSC', 'INTERNSHIP'], dtype='object')
```

Double-click (or enter) to edit

```
dataset.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

```
dataset.shape
```

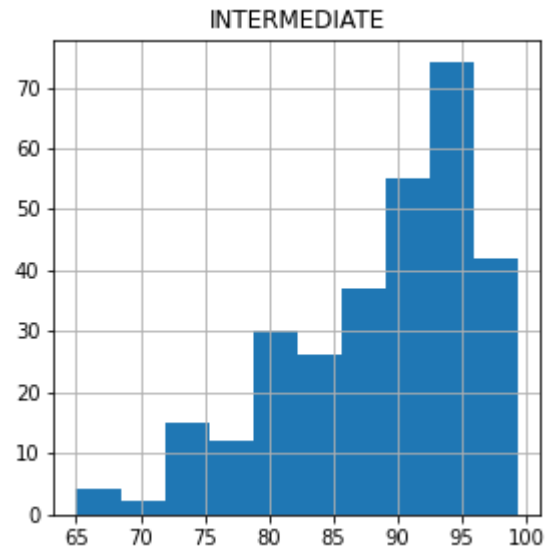
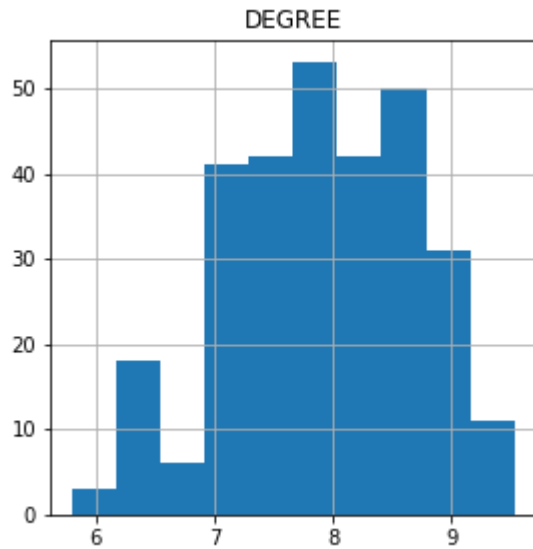
```
(297, 5)
```

## 3. Histogram for Degree, Inter and 10th class

```
import matplotlib.pyplot as plt
```

```
df=dataset.drop('StudentNo',axis=1)
df.hist(figsize=(10,10),bins=10)
```

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7fcc7076f5d0>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x7fcc70730b90>],
      [<matplotlib.axes._subplots.AxesSubplot object at 0x7fcc706f41d0>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x7fcc706af7d0>]],
      dtype=object)
```



SSC

#### 4. Create Pie-Chart to represent the Enrollments for each Internship program

```
| | | | | | | | | |
```

```
dataset['INTERNSHIP'].value_counts()['Data Science']
```

156

```
| | | | | | | | | |
```

```
dataset['INTERNSHIP'].value_counts()['MEAN Stack Web Development']
```

51

```
| | | | | | | | | |
```

```
dataset['INTERNSHIP'].value_counts()['Cloud Computing Services (AWS)']
```

90

```
INTERNSHIPS = ['Data Science', 'MEAN Stack Web Development', 'Cloud Computing Services (AWS)']
```

```
data = [156, 51, 90]
```

```
plt.pie(data, labels = INTERNSHIPS)
```

```
# show plot
```

```
plt.show()
```

Data Science



### 5. Find No. of enrollments in each program



```
dataset['INTERNSHIP'].value_counts()['Data Science'] #no. of enrollments for data science
```

156

```
dataset['INTERNSHIP'].value_counts()['MEAN Stack Web Development'] #no. of enrollments for
```

51

```
dataset['INTERNSHIP'].value_counts()['Cloud Computing Services (AWS)'] #no. of enrollments
```

90

### 6. Find Measure of Central Tendency: Mean, Mode, Median for Degree, 10th and Inter

```
dataset.describe() #for mean
```

	StudentNo	DEGREE	INTERMEDIATE	SSC
<b>count</b>	297.000000	297.000000	297.000000	297.000000
<b>mean</b>	1727.585859	7.928081	88.662626	88.106734
<b>std</b>	502.019415	0.785579	7.355733	9.027984
<b>min</b>	1001.000000	5.800000	65.000000	38.400000
<b>25%</b>	1075.000000	7.400000	83.000000	85.000000
<b>50%</b>	2044.000000	8.000000	90.800000	90.000000
<b>75%</b>	2118.000000	8.560000	94.600000	95.000000
<b>max</b>	2192.000000	9.530000	99.400000	99.000000



```
median_degree = dataset['DEGREE'].median() #median for degree
print(median_degree)
```

8.0

```
median_inter = dataset['INTERMEDIATE'].median() #median for inter
print(median_inter)
```

90.8

```

median_ssc = dataset['SSC'].median() #median for 10th
print(median_ssc)

90.0

mode_degree = dataset['DEGREE'].mode() #median for degree
print(mode_degree)

0    7.0
dtype: float64

mode_inter = dataset['INTERMEDIATE'].mode() #median for inter
print(mode_inter)

0    95.0
dtype: float64

mode_ssc = dataset['SSC'].mode()
print(mode_ssc)

0    95.0
dtype: float64

```

## 7. Find Measure of Variance: Minimum, Maximum, Range, Mean Deviation, Standard Deviation, Co-efficient of Variation for Degree, Inter And 10th class

```
dataset.describe() #for min,max values,mean,std deviation
```

	StudentNo	DEGREE	INTERMEDIATE	SSC
<b>count</b>	297.000000	297.000000	297.000000	297.000000
<b>mean</b>	1727.585859	7.928081	88.662626	88.106734
<b>std</b>	502.019415	0.785579	7.355733	9.027984
<b>min</b>	1001.000000	5.800000	65.000000	38.400000
<b>25%</b>	1075.000000	7.400000	83.000000	85.000000
<b>50%</b>	2044.000000	8.000000	90.800000	90.000000
<b>75%</b>	2118.000000	8.560000	94.600000	95.000000
<b>max</b>	2192.000000	9.530000	99.400000	99.000000



```
range_degree=dataset.DEGREE.max()-dataset.DEGREE.min() #range for degree
print(range_degree)
```

```
3.7299999999999995
```

```
range_inter=dataset.INTERMEDIATE.max()-dataset.INTERMEDIATE.min() #range for inter
print(range_inter)
```

```
34.400000000000006
```

```
range_ssc=dataset.SSC.max()-dataset.SSC.min() #range for 10th  
print(range_ssc)
```

```
60.6
```

```
#Co-efficient of Variation for Degree  
cov_degree = 0.784255/7.928080808080809*100 #Formula for co-efficient of variation is stan  
print(cov_degree)
```

```
9.89211662929365
```

```
#Co-efficient of Variation for Intermediate  
cov_inter = 7.343339/88.66262626262626*100  
print(cov_inter)
```

```
8.282338691669706
```


```
#Co-efficient of Variation for SSC  
cov_ssc = 9.027984/88.10673400673402*100  
print(cov_ssc)
```

```
10.246644710845814
```

## 8. Measures of Position

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

```
data_score = pd.DataFrame(np.random.randint(0, 10, size=(298,3 )), columns=['DEGREE', 'INT  
data_score  
data_score.apply(stats.zscore)
```

	DEGREE	INTERMEDIATE	SSC	
0	-0.282674	1.586798	-1.173672	
1	-1.675019	-0.511719	0.496597	

## Inter Quartile Range

```
q1 = np.percentile(dataset.DEGREE, 25)
q3 = np.percentile(dataset.DEGREE, 75)
iqr = q3 - q1
print('Q1 :' + str(q1))
print('Q3 :' + str(q3))
print("IQR equals: " + str(iqr))
```

```
Q1 :7.4
Q3 :8.56
IQR equals: 1.1600000000000001
```

## 9.Create Box Plot and Identify Outliers for Degree, Inter and 10th calss

```
df3=dataset.drop('StudentNo',axis=1)
df3.boxplot(figsize=(10,10))
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fcc70096490>



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



```
np.percentile(dataset.SSC, .90)
```

97.0



```
np.percentile(dataset.DEGREE, 90)
```

8.9



```
np.percentile(dataset.INTERMEDIATE, 90)
```

96.5



```
dataset[dataset.DEGREE>8.9]
```





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

```
dataset[dataset.INTERMEDIATE>96.5]
```



	StudentNo	DEGREE	INTERMEDIATE	SSC	INTERNSHIP
<b>16</b>	1017	9.53	98.2	93.1	MEAN Stack Web Development
<b>19</b>	1020	9.16	98.0	87.0	Data Science
<b>20</b>	1021	9.00	96.8	99.0	Data Science
<b>24</b>	1025	9.08	97.2	97.0	Data Science
<b>40</b>	1041	9.34	97.7	97.0	Data Science
<b>60</b>	1061	8.80	97.3	92.0	Data Science
<b>96</b>	1097	9.20	97.8	90.0	Data Science
<b>127</b>	2023	9.20	97.2	92.0	MEAN Stack Web Development
<b>138</b>	2034	9.38	98.2	87.0	Data Science

```
dataset[dataset.SSC>97.0]
```

StudentNo	DEGREE	INTERMEDIATE	SSC	INTERNSHIP
-----------	--------	--------------	-----	------------



14	1015	9.07	95.00	98.0	Cloud Computing Services (AWS)
20	1021	9.00	96.80	99.0	Data Science
21	1022	8.00	95.20	98.0	Data Science
76	1077	8.70	93.00	98.0	Data Science
95	1096	8.83	94.70	98.0	Data Science
122	2018	7.00	92.00	98.0	Cloud Computing Services (AWS)
126	2022	7.03	72.00	98.0	Cloud Computing Services (AWS)
145	2041	8.76	94.50	98.0	Data Science
147	2043	8.60	94.00	98.0	MEAN Stack Web Development
153	2049	8.38	87.00	98.0	MEAN Stack Web Development
162	2058	9.10	94.00	98.0	Data Science
174	2070	6.33	84.80	98.0	Cloud Computing Services (AWS)
188	2084	7.76	97.40	98.0	Cloud Computing Services (AWS)
191	2087	7.30	98.80	98.0	Cloud Computing Services (AWS)
196	2092	6.50	77.62	98.0	MEAN Stack Web Development



0s

completed at 7:56 PM

