Untitled6

October 13, 2022

Reg.no. 2145 Name : CH.VARUN REDDY Date : 13/10/2022

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

[ ]:
```
data=open("/content/Enrollments_28092022.csv")
```

[ ]:
```
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 x 5 columns]

[ ]:
```
#for quantitative it gives numerical data type
#for qualitative it gives descriptive datatype
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
```

[ ]:
```
#no.of columns and rows
data.shape
```

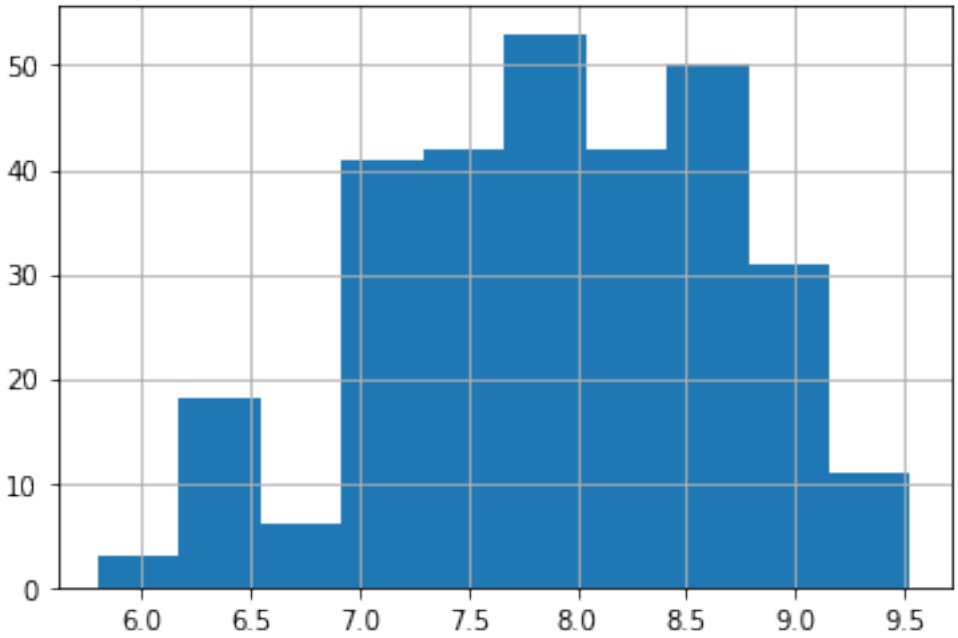
[ ]:
```
(297, 5)
```

[ ]:
```
#type of data
type(data)
```

[ ]:
```
pandas.core.frame.DataFrame
```

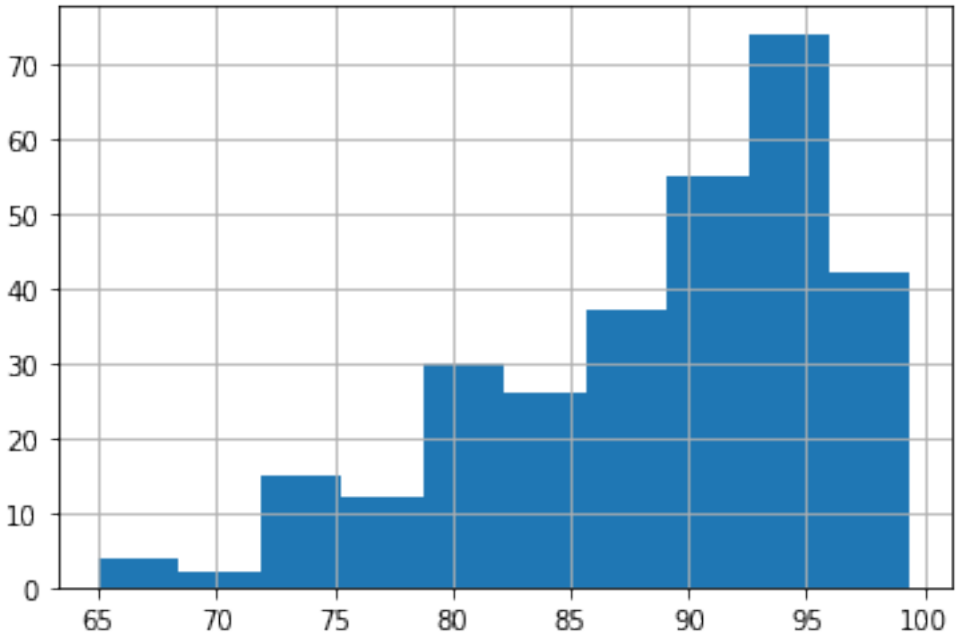
[ ]:
```
#Histogram of Degree column in data
data.DEGREE.hist()
```

[ ]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9562d3cb10>
```



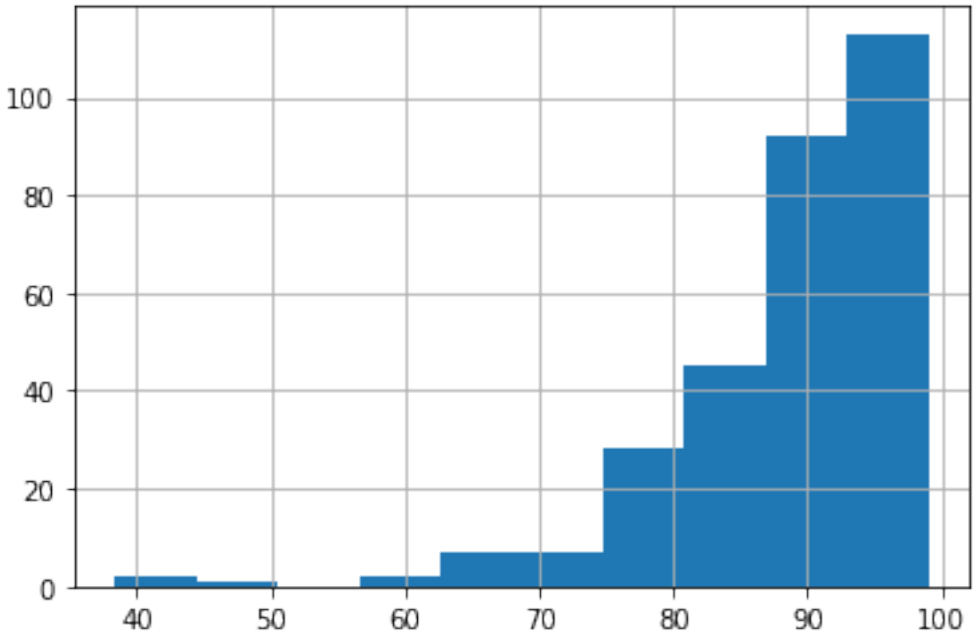
[ ]:
```
# Histogram of Intermediate coulum in data
data.INTERMEDIATE.hist()
```

[ ]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9562c3bc90>
```



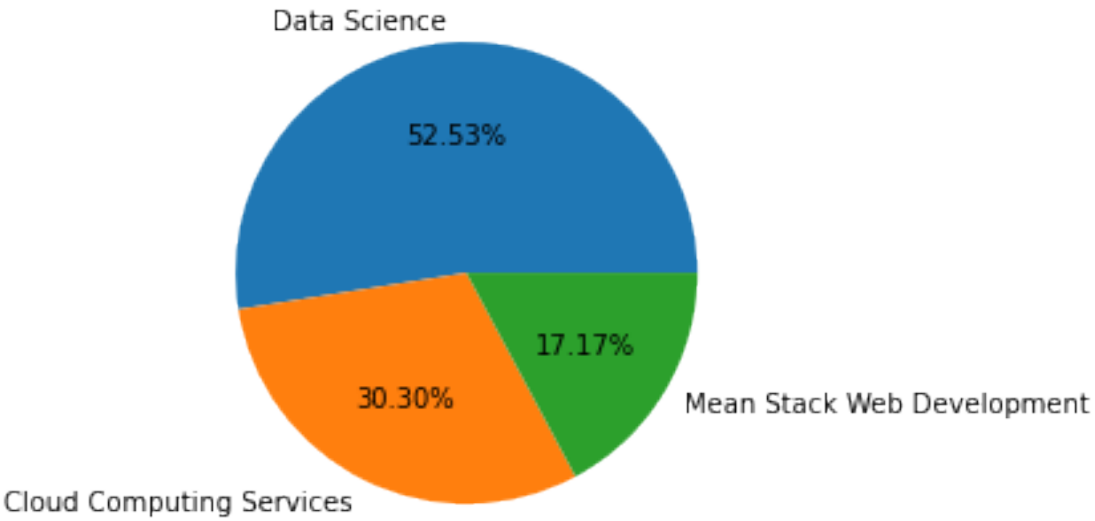
[ ]:
```
#Histogram of SSC column in data
data.SSC.hist()
```

[ ]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f95627a01d0>
```



[ ]:
```
dataFrame = pd.DataFrame({
   "INTERNSHIPS": ['Data Science','Cloud Computing Services','Mean Stack Web␣
 ↪Development'],"REPRTITIONS": [156,90,51]
})
```

[ ]:
```
#Piechart for given data
plt.pie(dataFrame["REPRTITIONS"], labels = dataFrame["INTERNSHIPS"],autopct="%1.
 ↪2f%%")
plt.show()
```



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

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

[ ]:
```
#MEAN, MEDIAN, MODE for Degree
data.DEGREE.mean()
```

[ ]:
```
7.928080808080809
```

[ ]:
```
data.DEGREE.mode()
```

[ ]:
```
0    7.0
dtype: float64
```

[ ]:
```
data.DEGREE.median()
```

[ ]:
```
8.0
```

[ ]:
```
#MEAN, MEDIAN, MODE for inter
data.INTERMEDIATE.mean()
```

[ ]:
```
88.66262626262626
```

[ ]:
```
data.INTERMEDIATE.mode()
```

[ ]:
```
0    95.0
dtype: float64
```

[ ]:
```
data.INTERMEDIATE.median()
```

[ ]:
```
90.8
```

[ ]:
```
#MEAN, MEDIAN, MODE for 10th
data.SSC.mean()
```

[ ]:
```
88.10673400673402
```

[ ]:
```
data.SSC.mode()
```

[ ]:
```
0    95.0
dtype: float64
```

[ ]:
```
data.SSC.median()
```

[ ]:
```
90.0
```

[ ]:
```
#Measure of Variance
data.describe()
```

[ ]:

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

[ ]:
```
import scipy.stats as stats
```

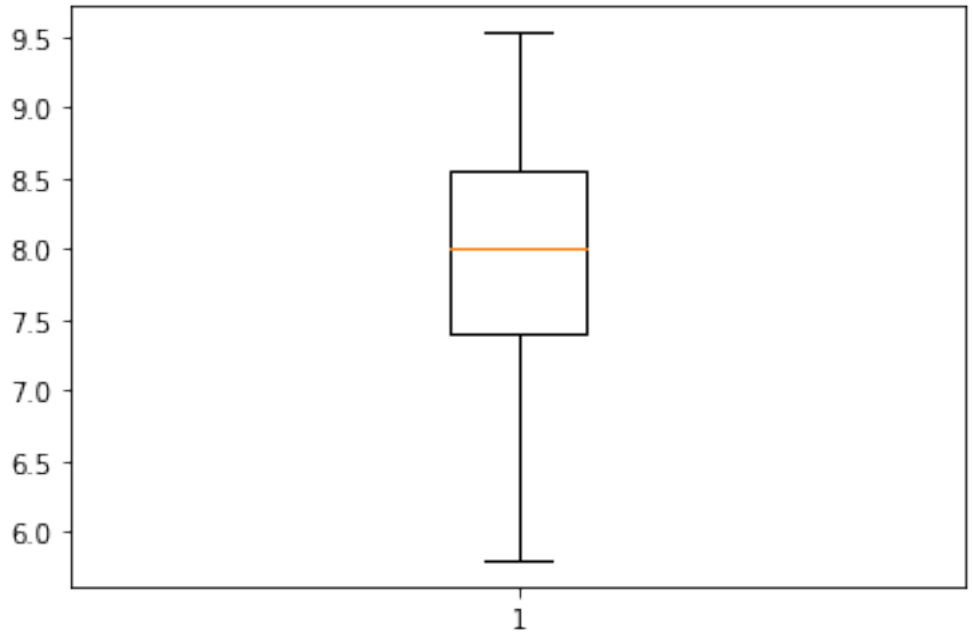
[ ]:
```
print("Standard scores of DEGREE:")
print(stats.zscore(data['DEGREE']))
print("Standard scores of INTERMEDIATE:")
print(stats.zscore(data['INTERMEDIATE']))
print("Standard scores of SSC:")
print(stats.zscore(data['SSC']))
```

```
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
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
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
```

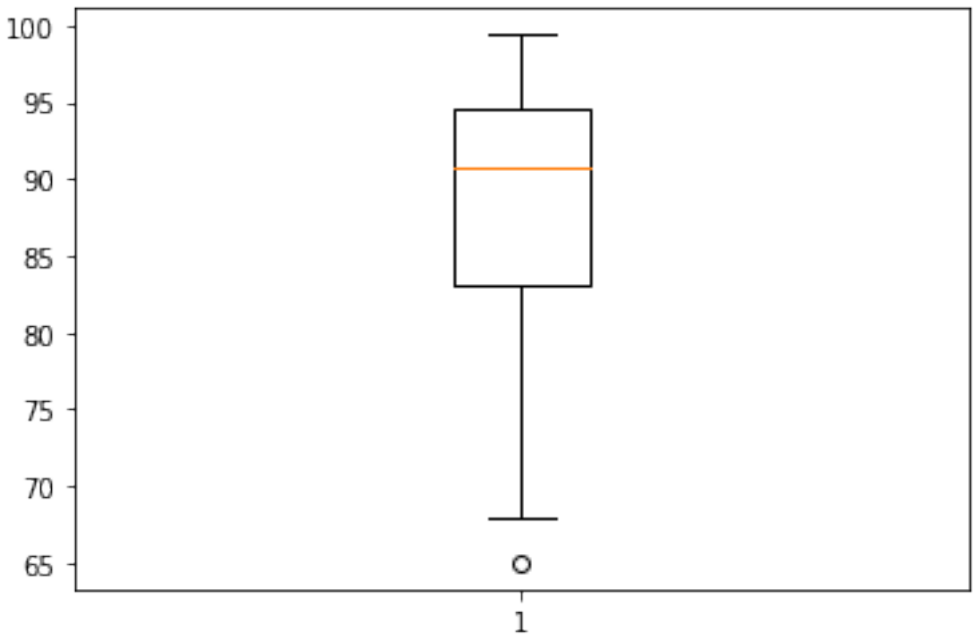
[ ]:
```
plt.boxplot(data['DEGREE'])
plt.show
```

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



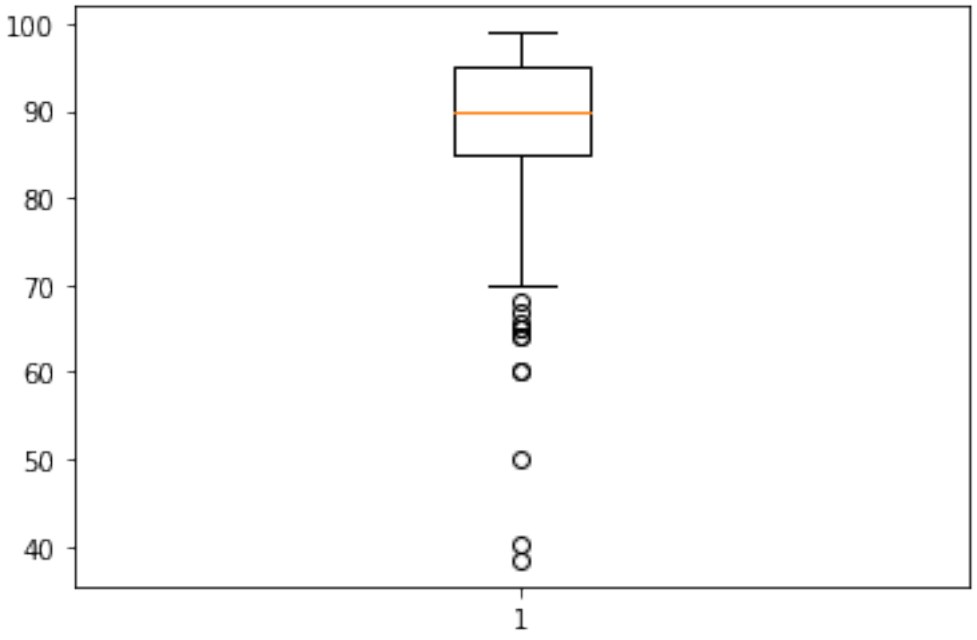
[ ]:
```
plt.boxplot(data['INTERMEDIATE'])
plt.show
```

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



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

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



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

[ ]:
```
outlier(data['DEGREE'])
```

```
Inter-Quartile Range: 1.1600000000000001
the following are outliers in box plot:
Series([], Name: DEGREE, dtype: float64)
```

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

```
Inter-Quartile Range: 11.599999999999994
the following are outliers in box plot:
271    65.0
Name: INTERMEDIATE, dtype: float64
```

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

```
Inter-Quartile Range: 10.0
the following are outliers in box plot:
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
```

[49]:
```
#90% percentile in cgpa is 9.4 so it is taken as greater than(>) 9.4
data.DEGREE[data.DEGREE>9.4].count()
data.INTERMEDIATE[data.INTERMEDIATE>90].count()
data.SSC[data.SSC>90].count()
```

[49]: 141

[ ]: