```
import pandas as pd
import numpy as np


data = {'ID': [1, 2, 3, 4, 5],
'Name':['Harsh', 'Shweta', 'Raj', 'Aarya', 'Amit'],
'Gender': ['Male', 'Female', 'Male', 'Female', 'Male'],
'Age':[18, 19, 20, 19, 21],
'Class':['Fresher', 'Fresher','Junior', 'Junior', 'Senior'],
'Major': ['Science', 'Arts', 'Engineering', 'Business', 'Science'],
'GPA':[3.5, 3.5, 3.8, 2.9, 3.6],
'Test Score': [85, 90, 93, 75, 87],
'Attendance':[90, 85, 80, 95, 75]}


df = pd.DataFrame(data)


print(df.isnull().sum())
```

```
ID            0
Name          0
Gender        0
Age           0
Class         0
Major         0
GPA           0
Test Score    0
Attendance    0
dtype: int64
```

```
print(df['Gender'].value_counts())
```

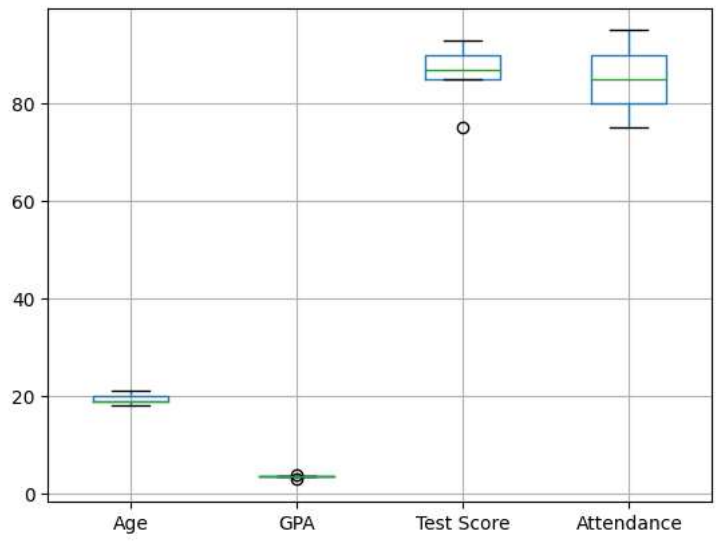
```
Male      3
Female    2
Name: Gender, dtype: int64
```

```
print(df[['Age', 'GPA', 'Test Score', 'Attendance']].describe())
```

```
             Age       GPA  Test Score  Attendance
count   5.000000  5.000000    5.000000    5.000000
mean   19.400000  3.460000   86.000000   85.000000
std     1.140175  0.336155    6.855655    7.905694
min    18.000000  2.900000   75.000000   75.000000
25%    19.000000  3.500000   85.000000   80.000000
50%    19.000000  3.500000   87.000000   85.000000
75%    20.000000  3.600000   90.000000   90.000000
max    21.000000  3.800000   93.000000   95.000000
```

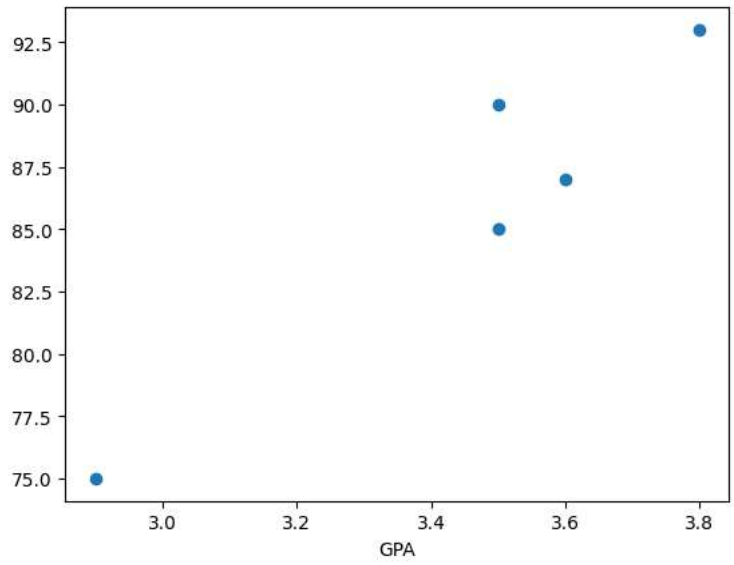
```
import matplotlib.pyplot as plt


df[['Age', 'GPA', 'Test Score', 'Attendance']].boxplot()
plt.show()
```



```
plt.scatter(df['GPA'], df['Test Score'])
plt.xlabel('GPA')
```

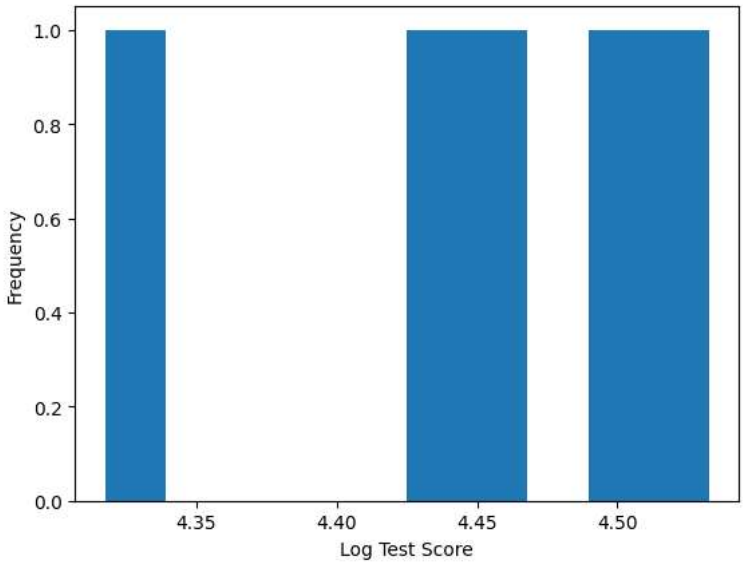
```
Text(0.5, 0, 'GPA')
```



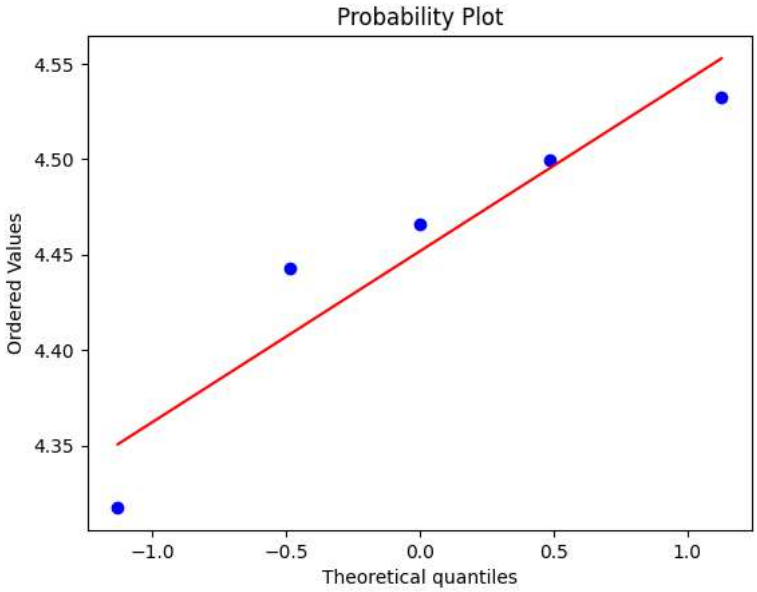

```
df['Log Test Score'] = np.log(df['Test Score'])
```

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

```
plt.hist(df['Log Test Score'])
plt.xlabel('Log Test Score')
plt.ylabel('Frequency')
plt.show()
```




```
stats.probplot(df['Log Test Score'], plot=plt)
plt.show()
```


Probability Plot
Ordered Values
4.55
4.50
4.45
4.40
4.35
−1.0
−0.5
0.0
0.5
1.0
Theoretical quantiles