

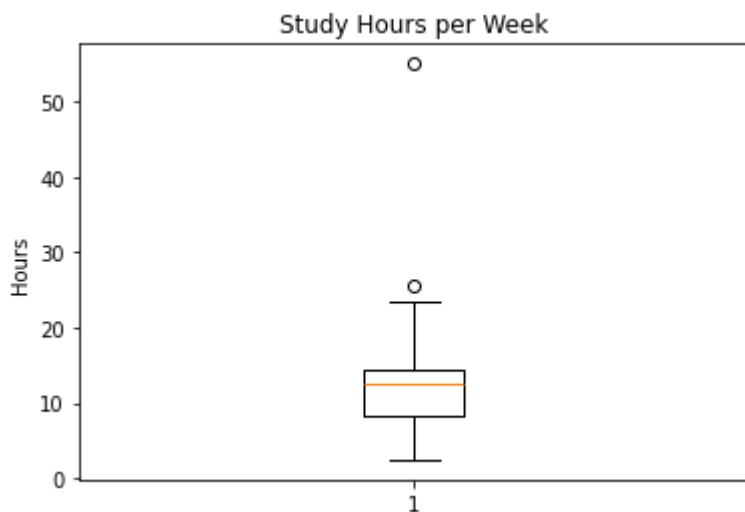
```
In [5]: df.isnull().sum()
```

```
Out[5]: Student_ID      0  
Age          0  
Gender       0  
Attendance_Percentage 1  
Study_Hours_per_Week   1  
Math_Score     1  
Science_Score   0  
English_Score    0  
dtype: int64
```

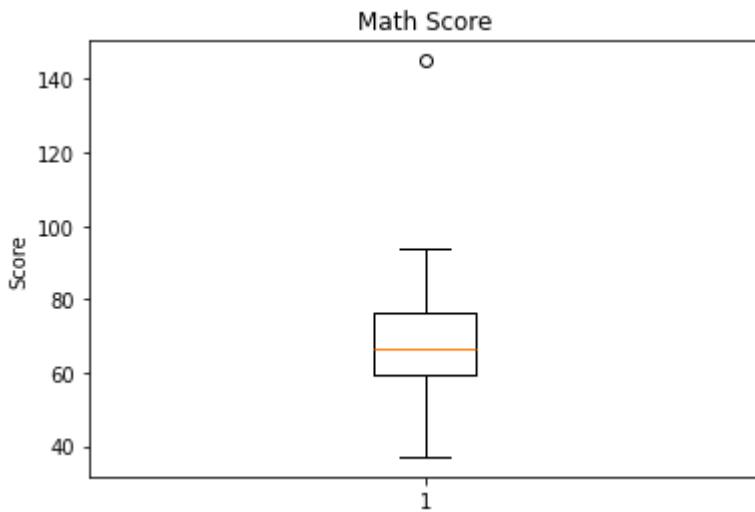
```
In [8]: df["Gender"] = df["Gender"].str.lower()  
df["Gender"] = df["Gender"].replace({"male": "Male", "female": "Female"})
```

```
In [9]: df.loc[df["Age"] < 0, "Age"] = df["Age"].median()
```

```
In [10]: plt.figure()  
plt.boxplot(df["Study_Hours_per_Week"])  
plt.title("Study Hours per Week")  
plt.ylabel("Hours")  
plt.show()
```



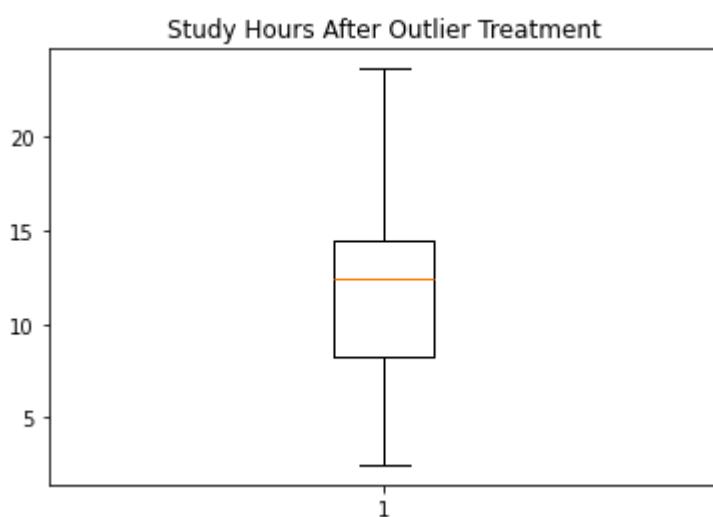
```
In [11]: plt.figure()  
plt.boxplot(df["Math_Score"])  
plt.title("Math Score")  
plt.ylabel("Score")  
plt.show()
```



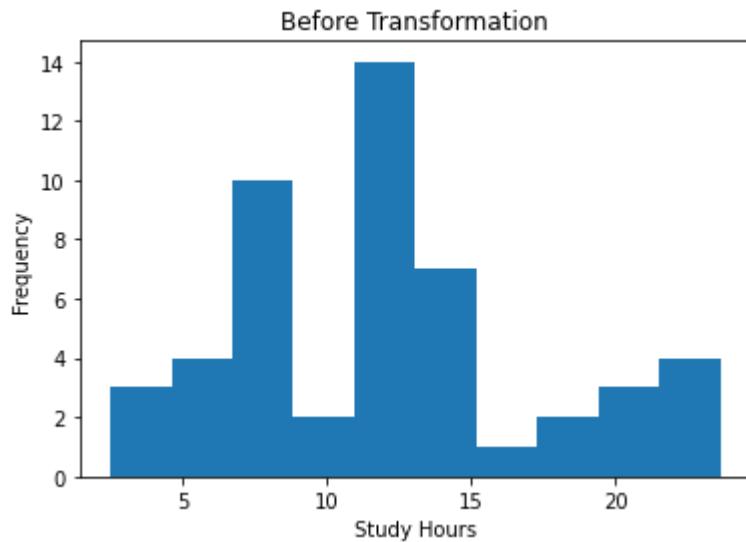
```
In [12]: def treat_outliers(column):
    Q1 = column.quantile(0.25)
    Q3 = column.quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR
    return column.clip(lower, upper)
```

```
In [13]: df["Study_Hours_per_Week"] = treat_outliers(df["Study_Hours_per_Week"])
df["Math_Score"] = treat_outliers(df["Math_Score"])
```

```
In [14]: plt.figure()
plt.boxplot(df["Study_Hours_per_Week"])
plt.title("Study Hours After Outlier Treatment")
plt.show()
```

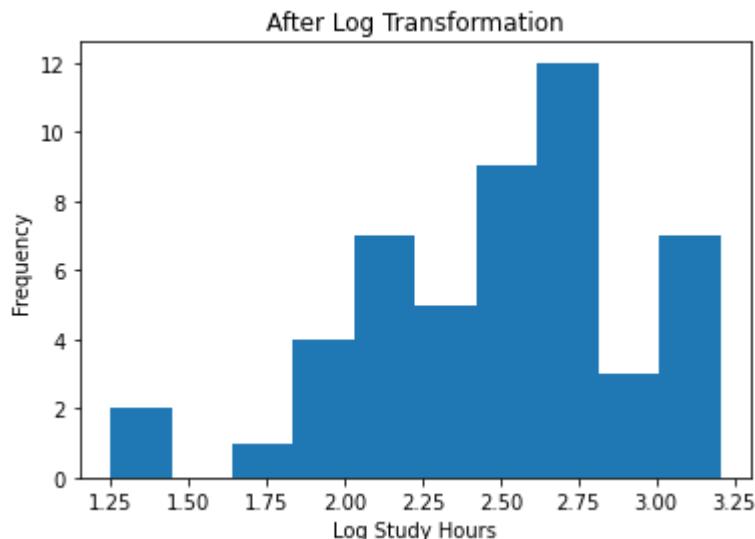


```
In [15]: plt.figure()
plt.hist(df["Study_Hours_per_Week"], bins=10)
plt.title("Before Transformation")
plt.xlabel("Study Hours")
plt.ylabel("Frequency")
plt.show()
```



```
In [16]: df["Log_Study_Hours"] = np.log1p(df["Study_Hours_per_Week"])
```

```
In [17]: plt.figure()
plt.hist(df["Log_Study_Hours"], bins=10)
plt.title("After Log Transformation")
plt.xlabel("Log Study Hours")
plt.ylabel("Frequency")
plt.show()
```



```
In [19]: df.to_csv("academic_performance_cleaned.csv", index=False)
```