**Experiment-12**

AIM: Exploratory Data Analysis for Classification using Pandas or Matplotlib

import pandas as pd

import matplotlib.pyplot as plt

from scipy.stats import f\_oneway

data = {

'education': ['High School', 'College', 'High School', 'College', 'College'],

'vote': ['Yes', 'No', 'No', 'Yes', 'Yes'],

'population': [1000, 1500, 800, 1200, 2000]

}

DF = pd.DataFrame(data)

print(DF.describe())

y = list(DF['population'])

plt.boxplot(y)

plt.show()

print(DF["education"].value\_counts())

print(DF.groupby(['education', 'vote']).mean())

group1 = [5, 7, 3, 4, 8]

group2 = [9, 12, 11, 13, 10]

group3 = [14, 16, 19, 17, 15]

f\_statistic, p\_value = f\_oneway(group1, group2, group3)

print("F-statistic:", f\_statistic)

print("p-value:", p\_value)

Output:

population

count 5.000000

mean 1300.000000

std 469.041576

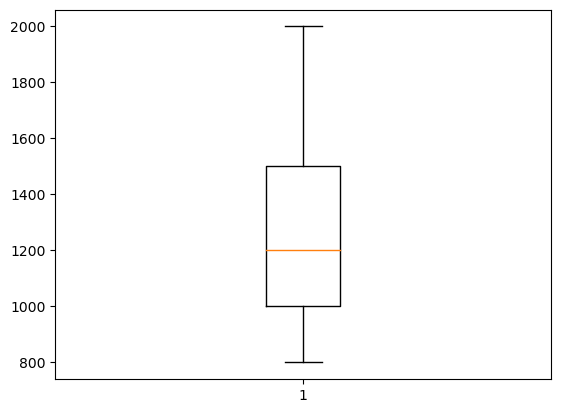
min 800.000000

25% 1000.000000

50% 1200.000000

75% 1500.000000

max 2000.000000



College 3

High School 2

Name: education, dtype: int64

population

education vote

College No 1500.0

Yes 1600.0

High School No 800.0

Yes 1000.0

F-statistic: 41.67619047619048

p-value: 3.972813930868759e-06