**Problem Statement:-**

The Student News Service at Marwadi University (MU) has decided to gather data about the undergraduate students that attend MU. MU creates and distributes a survey of 14 questions and receives responses from 62 undergraduates (stored in the Survey data set).

**Importing Necessary Libraries and Dataset :-**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

Dataset = pd.read\_excel("./Tutorial 2 Tk1 A batch\_MU\_Students.xlsx", sheet\_name="Q.2")

**2.1. For this data, construct the following contingency tables (Keep Gender as row variable.**

1. **Gender and Major**

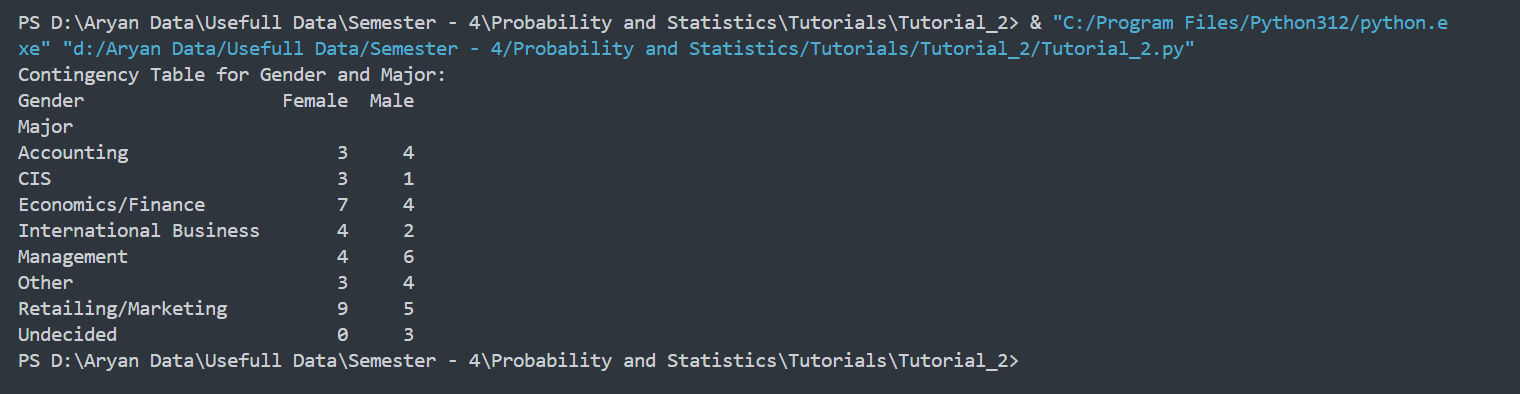
**Code :-**

contingency\_table\_1 = pd.crosstab(Dataset["Major"], Dataset["Gender"])

print("Contingency Table for Gender and Major:")

print(contingency\_table\_1)

**Output :-**

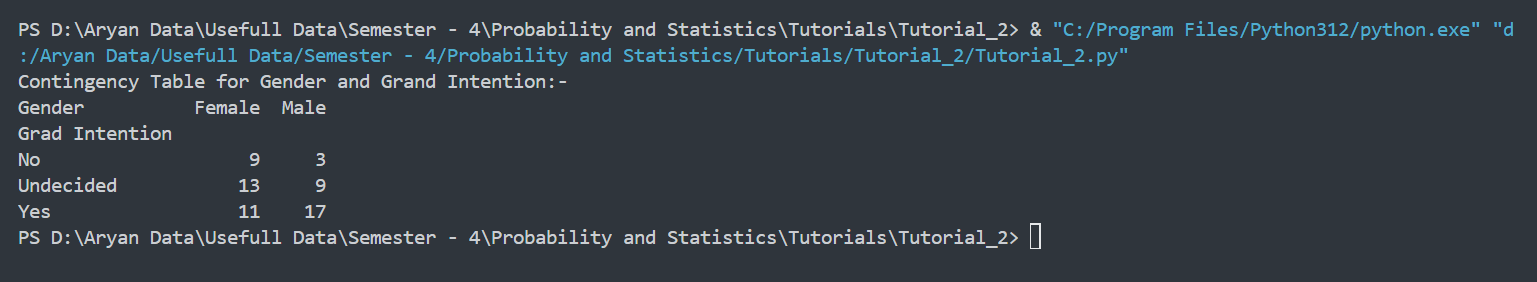
1. **Gender and Grad Intention**

**Code :-**

contingency\_table\_2 = pd.crosstab(Dataset["Grad Intention"], Dataset["Gender"])

print("Contingency Table for Gender and Grand Intention:- ")

print(contingency\_table\_2)

**Output :-**

1. **Gender and Employment**

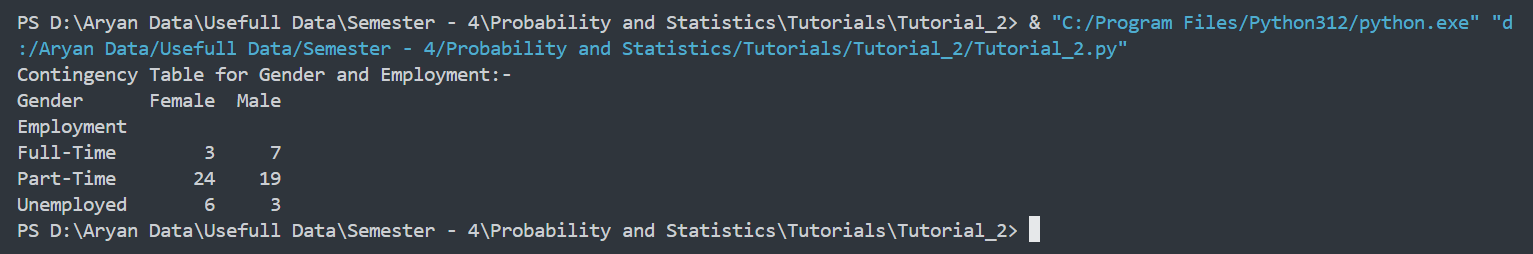
**Code :-**

contingency\_table\_3 = pd.crosstab(Dataset["Employment"], Dataset["Gender"])

print("Contingency Table for Gender and Employment:- ")

print(contingency\_table\_3)

**Output :-**

****

1. **Gender and Computer**

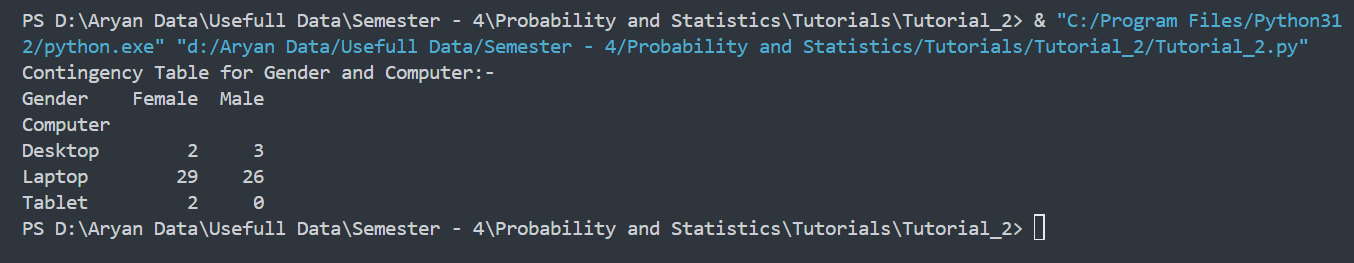
**Code :-**

contingency\_table\_4 = pd.crosstab(Dataset["Computer"], Dataset["Gender"])

print("Contingency Table for Gender and Computer:- ")

print(contingency\_table\_4)

**Output :-**

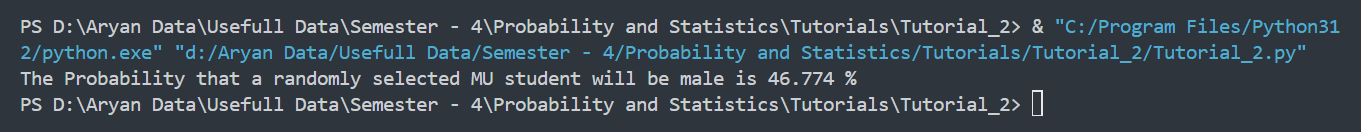
****

**2.2. Assume that the sample is representative of the population of MU. Based on the data, answer the following question:**

1. **What is the probability that a randomly selected MU student will be male?**

**Code :-**

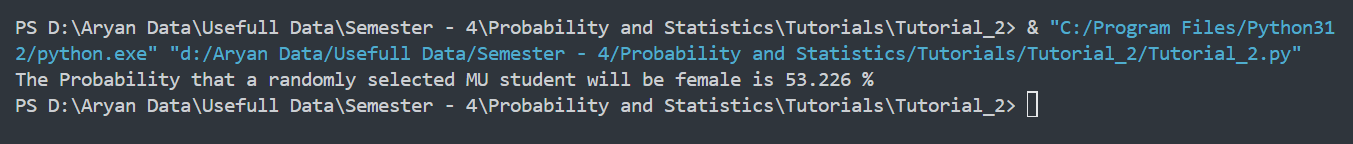
print(f"The Probability that a randomly selected MU student will be male is {"{:.3f}".format((Dataset[(Dataset['Gender'] == "Male")].shape[0] / Dataset.shape[0]) \* 100)} %")

**Output :-**

1. **What is the probability that a randomly selected MU student will be female?**

**Code :-**

print(f"The Probability that a randomly selected MU student will be female is {"{:.3f}".format((Dataset[(Dataset['Gender'] == "Female")].shape[0] / Dataset.shape[0]) \* 100)} %")

**Output :-**

**2.3. Assume that the sample is representative of the population of MU. Based on the data, answer the following question:**

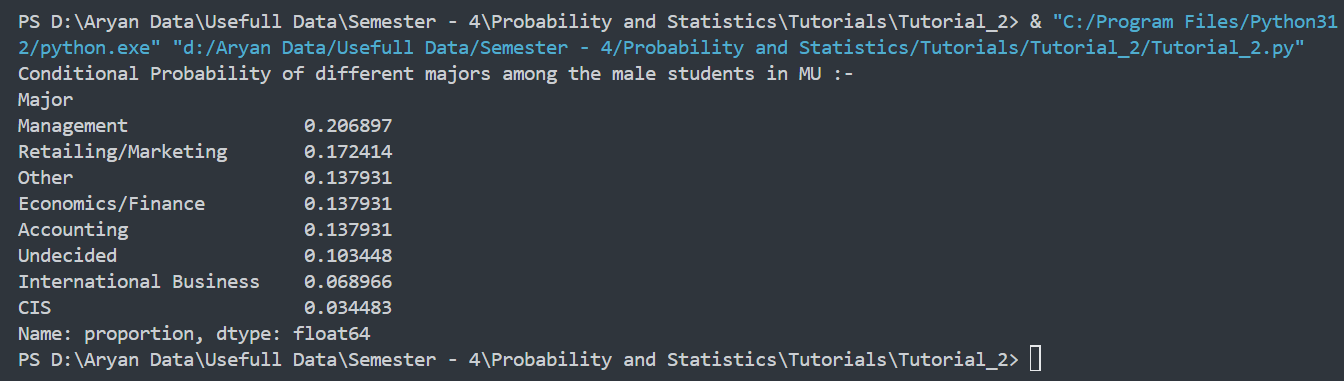
1. **Find the conditional probability of different majors among the male students in MU.**

**Code :-**

Males\_Data = Dataset[(Dataset['Gender'] == "Male")]

print("Conditional Probability of different majors among the male students in MU :-")

print(Males\_Data["Major"].value\_counts(normalize=True))

**Output :-**

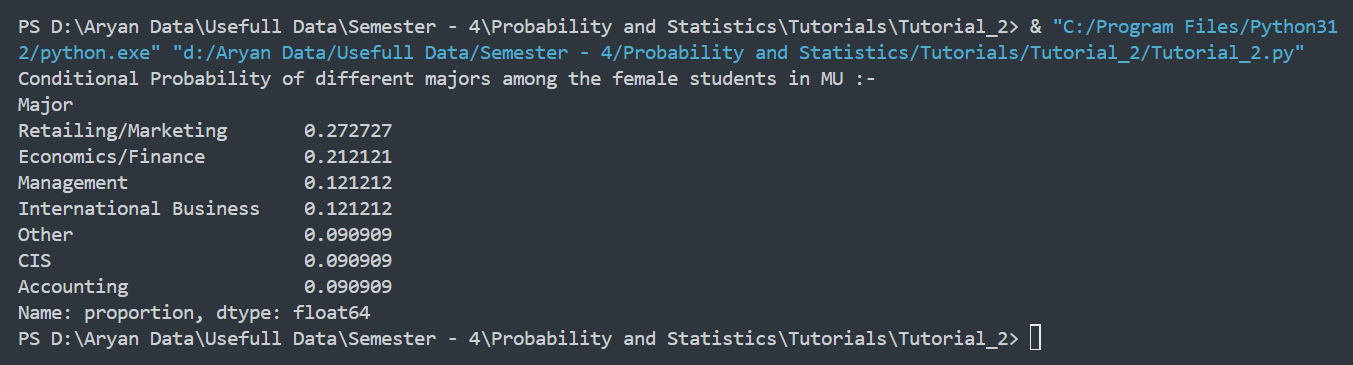
1. **Find the conditional probability of different majors among the female students of MU.**

**Code :-**

Females\_Data = Dataset[(Dataset["Gender"] == "Female")]

print("Conditional Probability of different majors among the female students in MU :-")

print(Females\_Data["Major"].value\_counts(normalize=True))

**Output :-**

**2.4 Assume that the sample is a representative of the population of MU. Based on the data, answer the following question:**

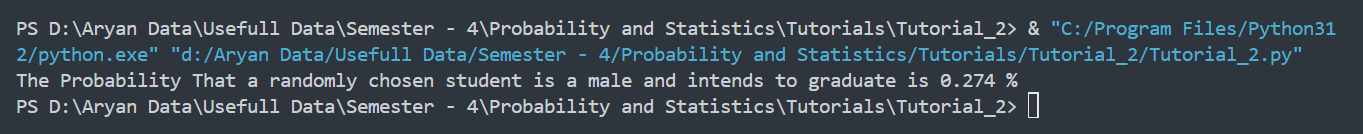
1. **Find the probability That a randomly chosen student is a male and intends to graduate.**

**Code :-**

Graduate\_Male = Dataset[(Dataset["Gender"] == "Male") & (Dataset["Grad Intention"] == "Yes")]

probability = '{:.3f}' .format(Graduate\_Male.shape[0] / Dataset.shape[0])

print(f"The Probability That a randomly chosen student is a male and intends to graduate is {probability} %")

**Output :-**

1. **Find the probability that a randomly selected student is a female and does NOT have a laptop.**

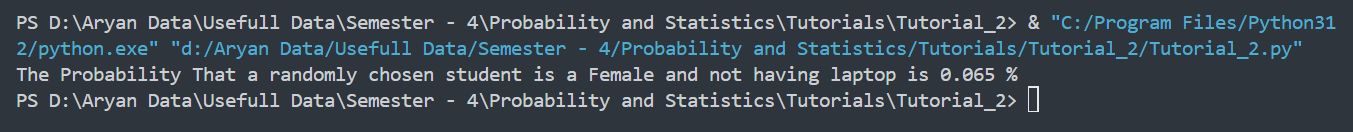
**Code :-**

Not\_Laptop\_Feale = Dataset[(Dataset["Gender"] == "Female") & (Dataset["Computer"] != "Laptop")]

probability = "{:.3f}".format(Not\_Laptop\_Feale.shape[0] / Dataset.shape[0])

print(f"The Probability That a randomly chosen student is a Female and not having laptop is {probability} %")

**Output :-**



**2.5 Assume that the sample is representative of the population of MU. Based on the data, answer the following question:**

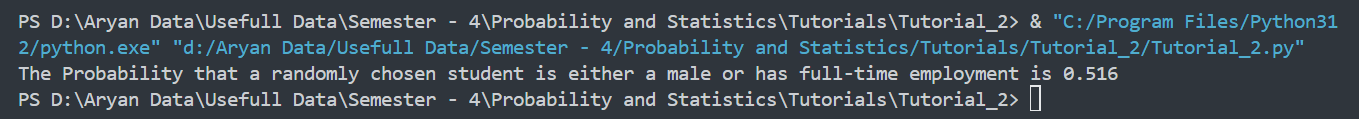
1. **Find the probability that a randomly chosen student is either a male or has full-time employment?**

**Code :-**

Male\_or\_full\_time = Dataset[(Dataset["Gender"] == "Male") | (Dataset["Employment"] == "Full-Time")]

probability = "{:.3f}".format(Male\_or\_full\_time.shape[0] / Dataset.shape[0])

print(f"The Probability that a randomly chosen student is either a male or has full-time employment is {probability}")

**Output :-**

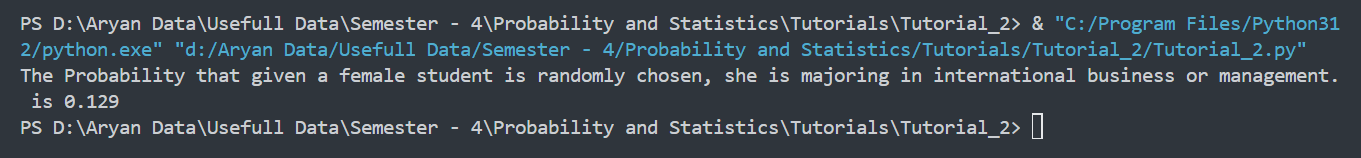
1. **Find the conditional probability that given a female student is randomly chosen, she is majoring in international business or management.**

**Code :-**

Female\_or\_full\_time = Dataset[(Dataset["Gender"] == "Female")& ((Dataset["Major"] == "Management")| (Dataset["Major"] == "International Business"))]

probability = "{:.3f}".format(Female\_or\_full\_time.shape[0] / Dataset.shape[0])

print(f"The Probability that given a female student is randomly chosen, she is majoring in international business or management. is {probability}")

**Output :-**

**2.6 Construct a contingency table of Gender and Intent to Graduate at 2 levels (Yes/No). The Undecided students are not considered now and the table is a 2x2 table. Do you think the graduate intention and being female are independent events?**

**Code :-**

subset = Dataset[Dataset["Grad Intention"].isin(["Yes", "No"])]

cont\_table = pd.crosstab(subset["Gender"], subset["Grad Intention"])

print(cont\_table)

sum = 0

for i in range(0,2) :

for j in range(0,2) :

sum = sum + cont\_table.iloc[i, j]

for i in range(0, 2):

for j in range(0, 2):

cont\_table.iloc[i, j] = cont\_table.iloc[i, j] / sum

print(cont\_table)

p\_female = 0

graduate = 0

for i in range(0, 2):

p\_female = p\_female + cont\_table.iloc[0, i]

for i in range(0, 2):

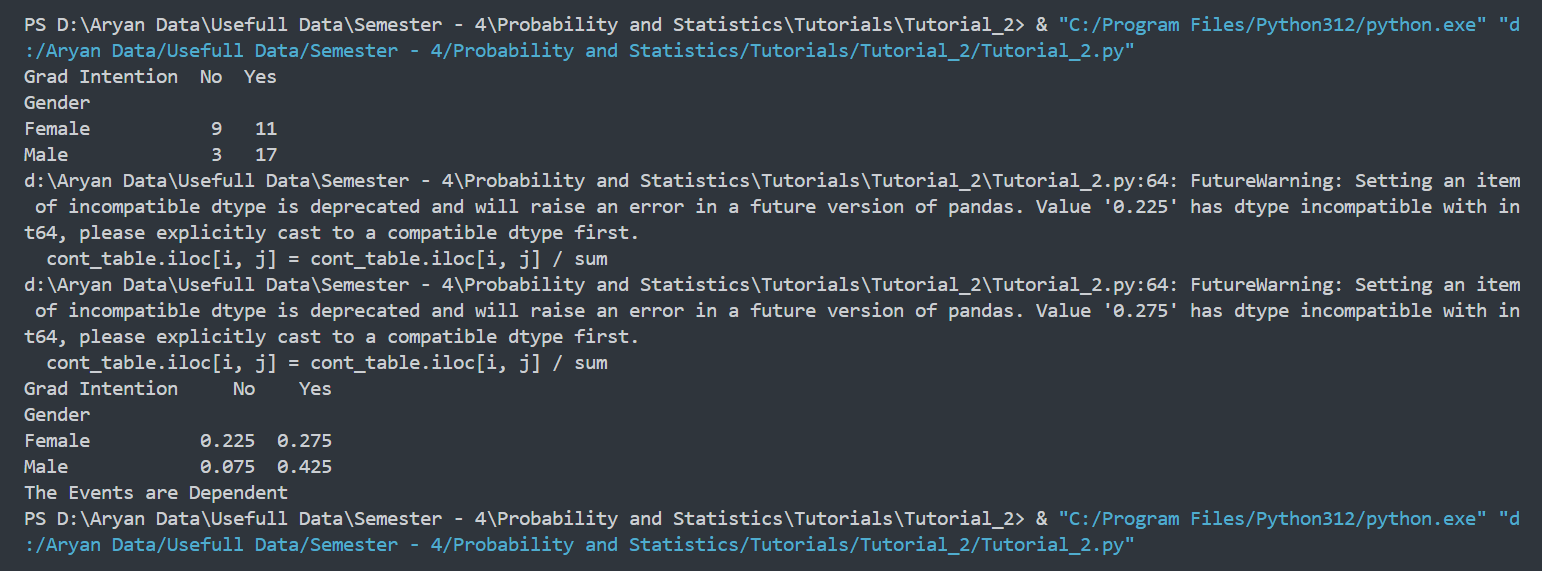
graduate = graduate + cont\_table.iloc[i, 1]

if (cont\_table.iloc[0, 1] == p\_female \* graduate) :

print("The Events are Independent")

else :

print("The Events are Dependent")

**Output :-**

**2.7. Note that there are four numerical (continuous) variables in the data set, GPA, Salary, Spending, and Text Messages. Answer the following questions based on the data**

1. **If a student is chosen randomly, what is the probability that his/her GPA is less than 3?**

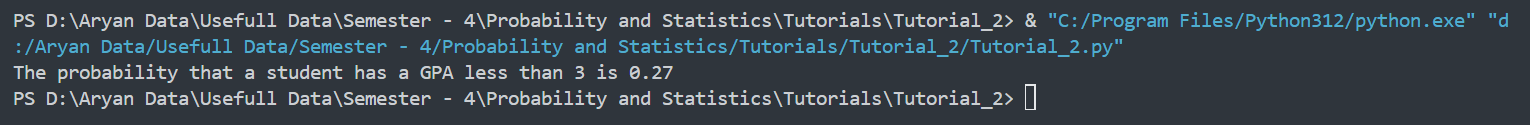
**Code :-**

num\_low\_gpa = Dataset[Dataset["GPA"] < 3]["GPA"].count()

total\_students = Dataset["GPA"].count()

prob\_low\_gpa = num\_low\_gpa / total\_students

print(f"The probability that a student has a GPA less than 3 is {prob\_low\_gpa:.2f}")

**Output :-**

1. **Find the conditional probability that a randomly selected male earns 50 or more. Find the conditional probability that a randomly selected female earns 50 or more.**

**Code :-**

table = pd.crosstab(index=Dataset["Gender"],columns=pd.cut(Dataset["Wealth"], bins=[-1, 49, 100], labels=["<50", ">=50"]),)

total\_males = table.loc["Male"].sum()

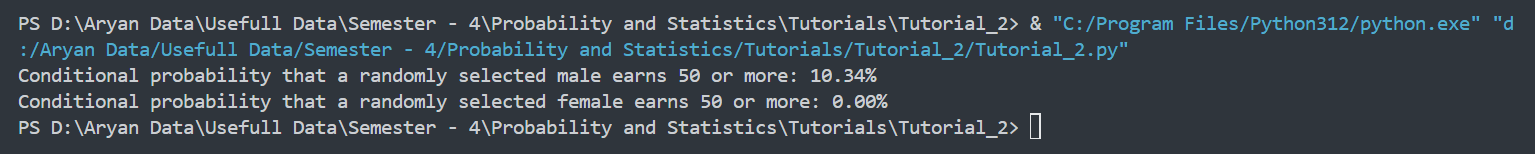
total\_females = table.loc["Female"].sum()

prob\_male\_50\_or\_more = table.loc["Male", ">=50"] / total\_males

prob\_female\_50\_or\_more = table.loc["Female", ">=50"] / total\_females

print("Conditional probability that a randomly selected male earns 50 or more: {:.2%}".format(prob\_male\_50\_or\_more))

print("Conditional probability that a randomly selected female earns 50 or more: {:.2%}".format(prob\_female\_50\_or\_more))

**Output :-‘**

**2.8. Note that there are four numerical (continuous) variables in the data set, GPA, Salary, Spending, and Text Messages. For each of them comment whether they follow a normal distribution. Write a note summarizing your conclusions.**

**Code :-**

plt.figure(figsize=(12, 8))

plt.subplot(2, 2, 1)

plt.hist(Dataset["GPA"], bins=20)

plt.title("GPA Histogram")

plt.subplot(2, 2, 2)

plt.hist(Dataset["Salary"], bins=20)

plt.title("Salary Histogram")

plt.subplot(2, 2, 3)

plt.hist(Dataset["Spending"], bins=20)

plt.title("Spending Histogram")

plt.subplot(2, 2, 4)

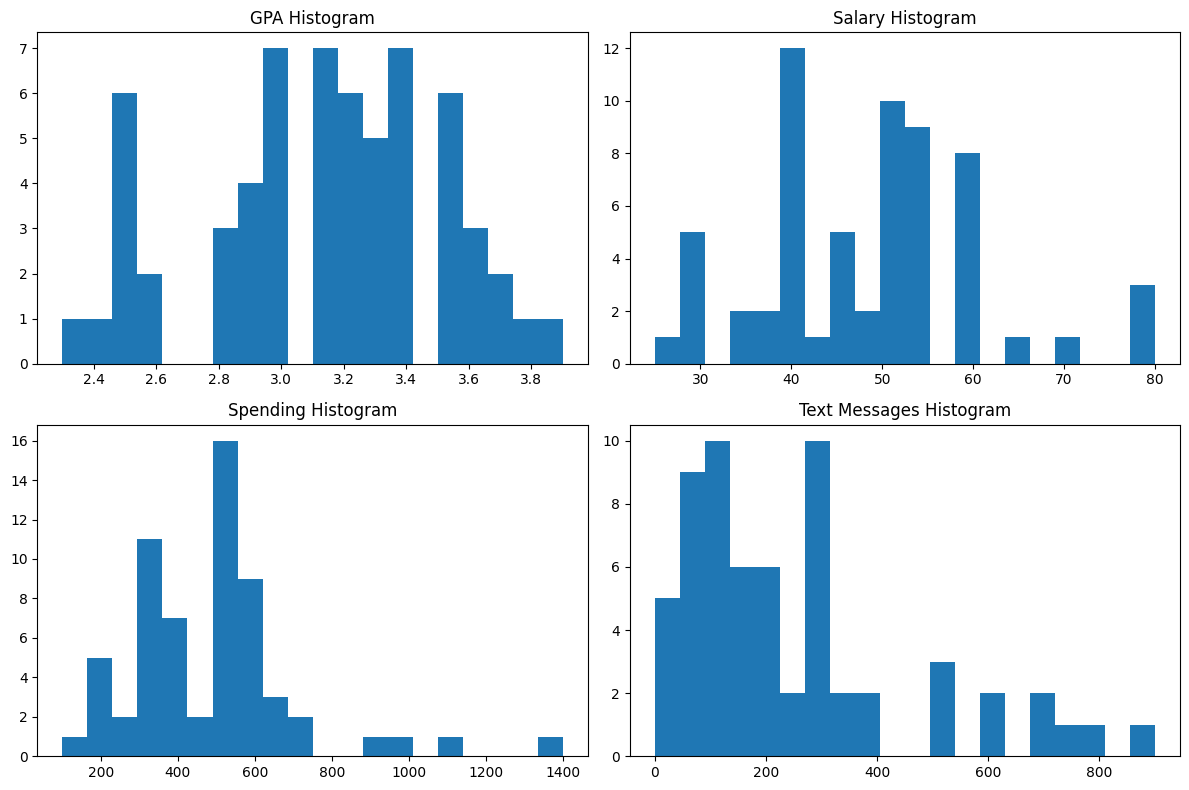
plt.hist(Dataset["Text Messages"], bins=20)

plt.title("Text Messages Histogram")

plt.tight\_layout()

plt.show()

**Output :-**



In conclusion, GPA and Salary appears to be approximately normally distributed, while Spending, and Text Messages are not normally distributed.