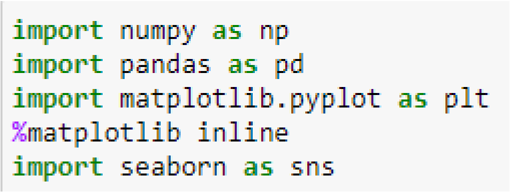
PROBLEM 2 – SURVEY

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

IMPORTING NESSCEARY LIBRARIES

For performing basic EDA we need to import pandas, numpy, and matplotlib and seaborn modules



```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

EDA:

In basic EDA we could understand more about the data such as

SHAPE-(62, 14)

HEAD-We can use this to understand the first five rows and columns in the dataset

INFO-we have 62 entries and 13 column and we don't have any Null Values

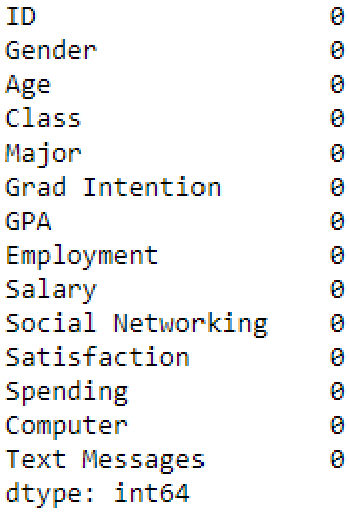
We have,

GPA and SALARY as float values

ID, AGE, SOCIAL NETWORKING, SATISFACTION, SPENDING, TEXT MESSAGES as integer values

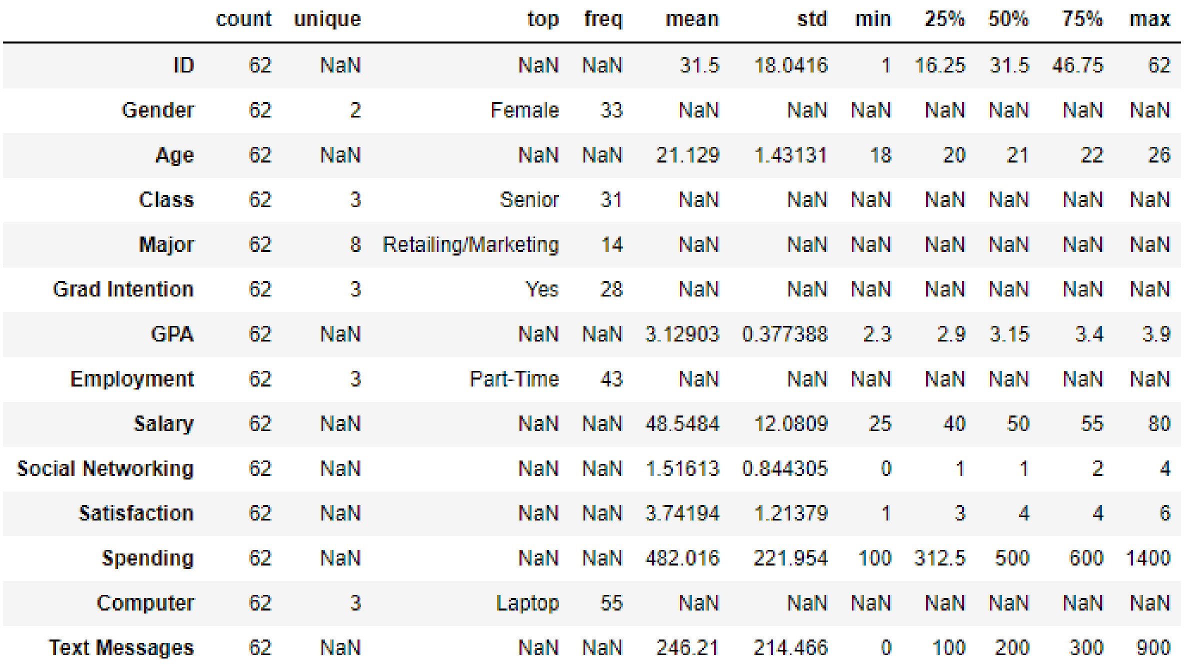
GENDER, CLASS, MAJOR, GRAD INTENTION, EMPLOYMENT AND COMPUTER as object values

NULL VALUES = 0



```
ID                   0
Gender               0
Age                  0
Class                0
Major                0
Grad Intention       0
GPA                  0
Employment           0
Salary               0
Social Networking    0
Satisfaction         0
Spending             0
Computer             0
Text Messages        0
dtype: int64
```

DESCRIPTIVE STATICS OF THE DATASET



	count	unique	top	freq	mean	std	min	25%	50%	75%	max
ID	62	NaN	NaN	NaN	31.5	18.0416	1	16.25	31.5	46.75	62
Gender	62	2	Female	33	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Age	62	NaN	NaN	NaN	21.129	1.43131	18	20	21	22	26
Class	62	3	Senior	31	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Major	62	8	Retailing/Marketing	14	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Grad Intention	62	3	Yes	28	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GPA	62	NaN	NaN	NaN	3.12903	0.377388	2.3	2.9	3.15	3.4	3.9
Employment	62	3	Part-Time	43	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Salary	62	NaN	NaN	NaN	48.5484	12.0809	25	40	50	55	80
Social Networking	62	NaN	NaN	NaN	1.51613	0.844305	0	1	1	2	4
Satisfaction	62	NaN	NaN	NaN	3.74194	1.21379	1	3	4	4	6
Spending	62	NaN	NaN	NaN	482.016	221.954	100	312.5	500	600	1400
Computer	62	3	Laptop	55	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Text Messages	62	NaN	NaN	NaN	246.21	214.466	0	100	200	300	900

We have unique values in Gender, class, Major, Grad Intention, Employment and Computer

From descriptive statistics or five point summary,

- No of female is 33
- No of male is 29
- The max age of students is 26 and median age is 21
- Class we have 3 unique values senior, junior , sophomore
- In majors we have 8 majors
- Students have scored 3.9 GPA but median remains around to be 3.15
- Retailing/marketing is the most preferred Major by students
- 28/62 have grad intent
- Part-time seems to be more when compared to fulltime job
- 55 have laptop for the education

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

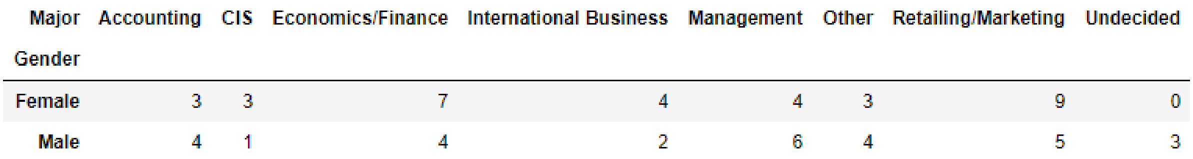
2.1.1. Gender and Major

2.1.2. Gender and Grad Intention

2.1.3. Gender and Employment

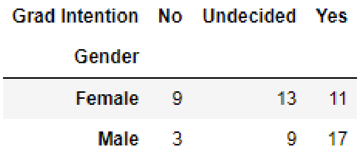
2.1.4. Gender and Computer

2.1.1. Gender and Major



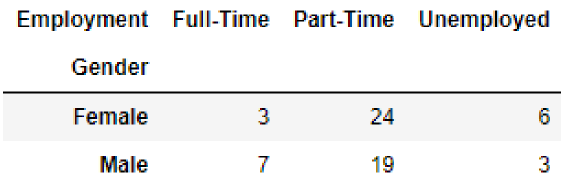
Major	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided
Gender								
Female	3	3	7	4	4	3	9	0
Male	4	1	4	2	6	4	5	3

2.1.2. Gender and Grad Intention



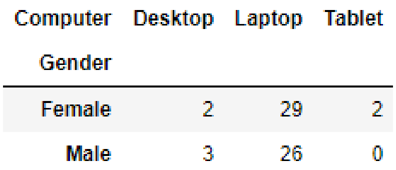
Grad Intention	No	Undecided	Yes
Gender			
Female	9	13	11
Male	3	9	17

2.1.3. Gender and Employment



Employment	Full-Time	Part-Time	Unemployed
Gender			
Female	3	24	6
Male	7	19	3

2.1.4. Gender and Computer



Computer	Desktop	Laptop	Tablet
Gender			
Female	2	29	2
Male	3	26	0

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

2.2.1. What is the probability that a randomly selected CMSU student will be male?

Number of male (A) = 29

Total Number of students (B) = 62

P (A/B) =29/62

The probability that a randomly selected CMSU student will be male is **46.77419354 8387096 %**

2.2.2. What is the probability that a randomly selected CMSU student will be female?

Number of female (A) = 33

Total Number of students (B) = 62

P (A/B) =33/62

The probability that a randomly selected CMSU student will be female is **53.2258064516129 %**

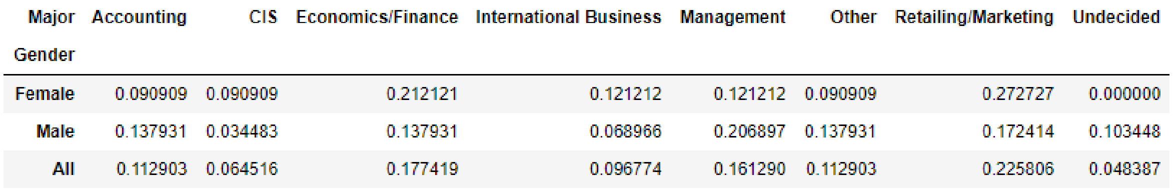
2.3. Assume that the sample is representative of the population of CMSU. Based on the data, answer the following question:

2.3.1. Find the conditional probability of different majors among the male students in CMSU.

Conditional probability of different Majors

P (Different Majors/ Male)

The snippet shows the probability of male choosing different majors

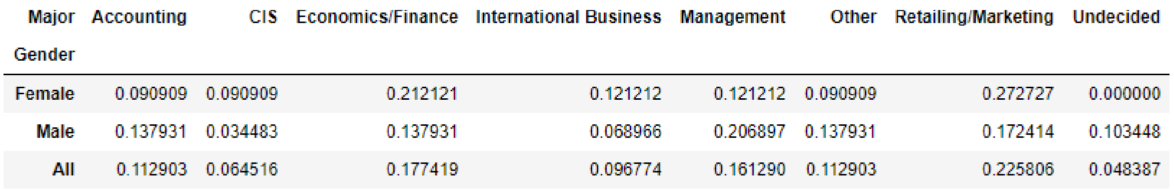


Major	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided
Gender								
Female	0.090909	0.090909	0.212121	0.121212	0.121212	0.090909	0.272727	0.000000
Male	0.137931	0.034483	0.137931	0.068966	0.206897	0.137931	0.172414	0.103448
All	0.112903	0.064516	0.177419	0.096774	0.161290	0.112903	0.225806	0.048387

2.3.2 Find the conditional probability of different majors among the female students of CMSU.

P (Conditional Majors/ Female)

The snippet shows the probability of female choosing different majors

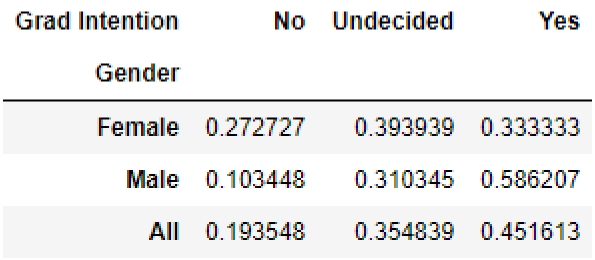


Major	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided
Gender								
Female	0.090909	0.090909	0.212121	0.121212	0.121212	0.090909	0.272727	0.000000
Male	0.137931	0.034483	0.137931	0.068966	0.206897	0.137931	0.172414	0.103448
All	0.112903	0.064516	0.177419	0.096774	0.161290	0.112903	0.225806	0.048387

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

2.4.1. Find the probability that a randomly chosen student is a male and intends to graduate.

P (Grad Intent Yes/ Male) = 17/29



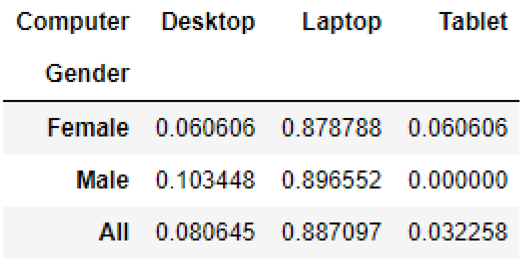
Grad Intention	No	Undecided	Yes
Gender			
Female	0.272727	0.393939	0.333333
Male	0.103448	0.310345	0.586207
All	0.193548	0.354839	0.451613

The probability that a randomly chosen student is a male and intends to graduate is **58.62%**

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

P (Have a laptop/ female) = 29/33

P (does not have a laptop/ female) = 1- P (Have a laptop/ female) = 1-0.88=12%



Computer	Desktop	Laptop	Tablet
Gender			
Female	0.060606	0.878788	0.060606
Male	0.103448	0.896552	0.000000
All	0.080645	0.887097	0.032258

The probability that a randomly selected student is a female and does not have laptop is 1-0.88

The probability that a randomly selected student is a female and does NOT have a laptop is **12%**

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

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

Probability of randomly selected student is male P (A) = 46.77%

Probability of randomly selected student has a fulltime job P (B) = 16.13%

Probability of male having a fulltime job P (A and B) = 11.29%

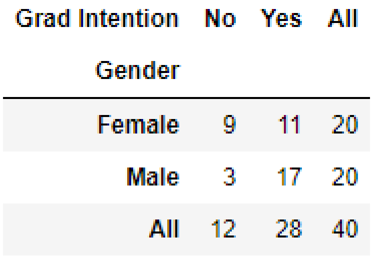
P = p_of_male_stu+p_of_fulltime_emp-p_of_male_fulltime_emp = 51.61%

The probability that a randomly chosen student is either a male or has full-time empl oyment **51.61290322580645 %**

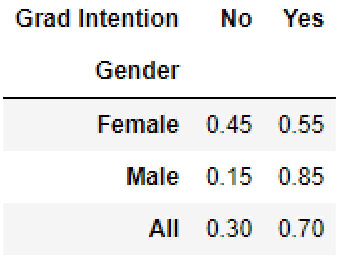
2.5.2. Find the conditional probability that given a female student is randomly chosen, she is majoring in international business or management.

Probability that given a female student is randomly chosen, she is majoring in international business or management **24.24 %**

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?



Grad Intention	No	Yes	All
Gender			
Female	9	11	20
Male	3	17	20
All	12	28	40



Grad Intention	No	Yes
Gender		
Female	0.45	0.55
Male	0.15	0.85
All	0.30	0.70

CONCLUSION:

The probability that a randomly selected Student is Female 50.0
The probability that a randomly selected student is female and intends to graduate 55.0 %
They are not independent events

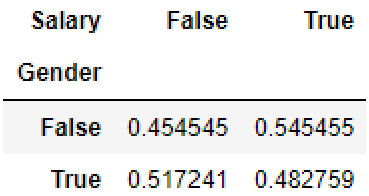
2.7. Note that there are four numerical (continuous) variables in the data set, GPA, Salary, Spending, and Text Messages.

Answer the following questions based on the data

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

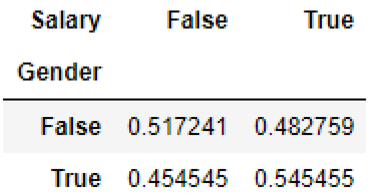
The probability that his/her GPA is less than 3 is **27.419354838709676 %**

2.6.2. 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.



Salary	False	True
Gender		
False	0.454545	0.545455
True	0.517241	0.482759

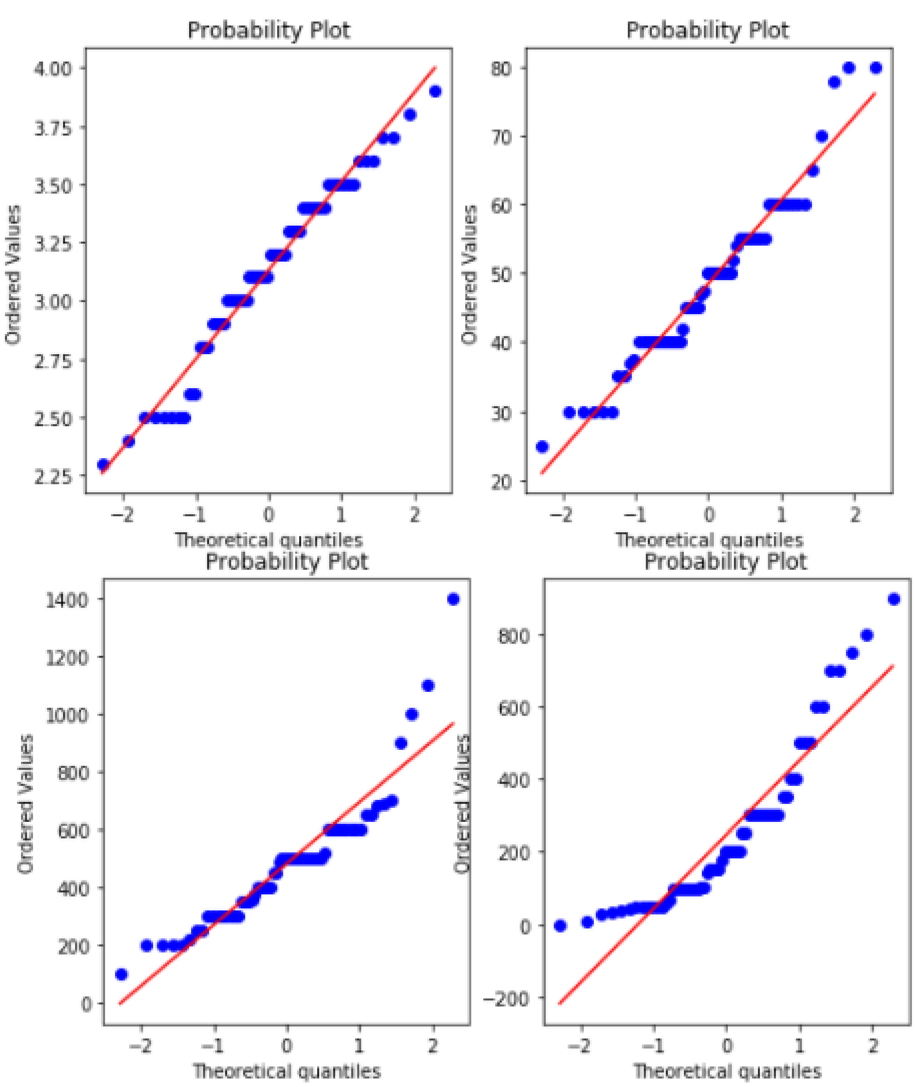
Probability that a randomly selected male earns 50 or more is 48%

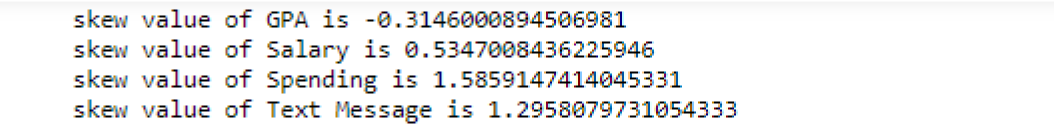


Salary	False	True
Gender		
False	0.517241	0.482759
True	0.454545	0.545455

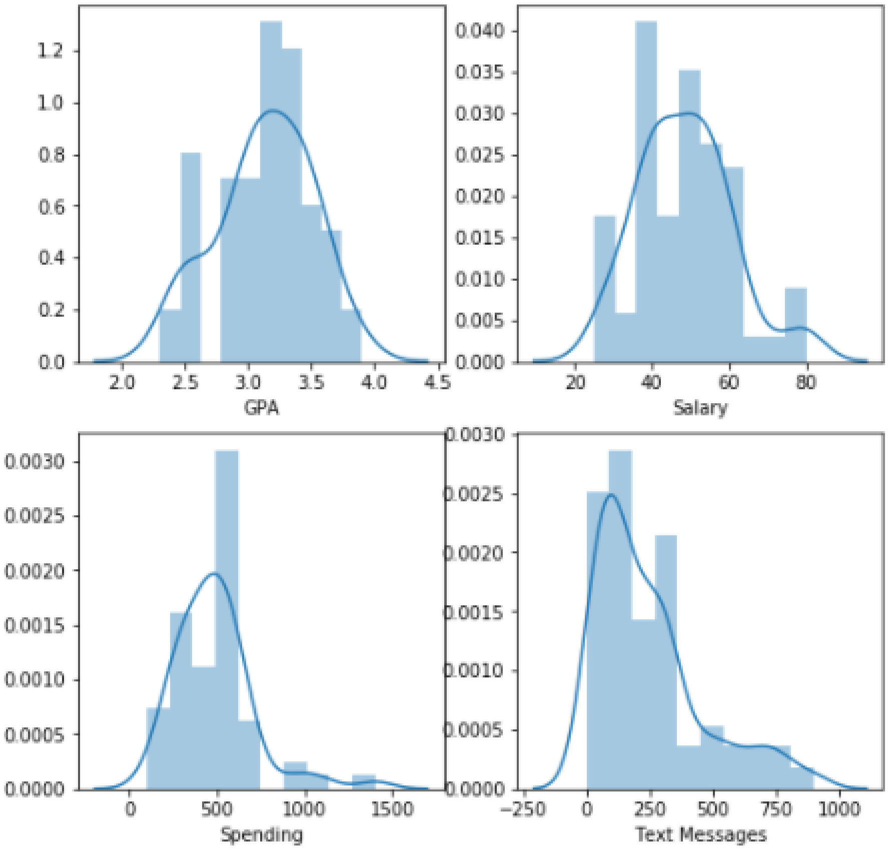
Probability that a randomly selected female earns 50 or more is 54%

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.





```
skew value of GPA is -0.3146000894506981
skew value of Salary is 0.5347008436225946
skew value of Spending is 1.5859147414045331
skew value of Text Message is 1.2958079731054333
```



The probability plot can be used to find the dataset follows a normal distribution or not, in our dataset we can find the points follow a straight line and we can say that all the GPA, salary, spending and text messages follow a normal distribution.

Looking at the skew value if the value is zero it is symmetric data, if we have a negative value for the skew that indicates that the data are skewed left and positive value of skew indicates the data are skewed towards right.

CONCLUSION

We have dataset of students answering to the survey and we have 62 responses from the students both male and female. We have almost equal number of male and female students. Many students have intention of graduating the retailing and marketing seem to have chosen by quite number of students. 2/3 of the students are looking for a part time job. The mean salary means to be around 50.