Survey Analysis

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 receive responses from 62 undergraduates (stored in the ***Survey*** data set).

Exploratory Data Analysis:

ID	Gender	Age	Class	Major	Grad Intention	GPA	Employment	Salary	Social Networking	Satisfaction	Spending	Computer	Text Messages
1	Female	20	Junior	Other	Yes	2.9	Full-Time	50.0	1	3	350	Laptop	200
2	Male	23	Senior	Management	Yes	3.6	Part-Time	25.0	1	4	360	Laptop	50
3	Male	21	Junior	Other	Yes	2.5	Part-Time	45.0	2	4	600	Laptop	200
4	Male	21	Junior	CIS	Yes	2.5	Full-Time	40.0	4	6	600	Laptop	250
5	Male	23	Senior	Other	Undecided	2.8	Unemployed	40.0	2	4	500	Laptop	100

Dataset has 14 variables, which has the different values for the particular response. ID is the variable which has the unique row number for each response.

Let us check the types of variables in the data frame.

```
ID                   int64
Gender              object
Age                  int64
Class               object
Major               object
Grad Intention      object
GPA                float64
Employment          object
Salary             float64
Social Networking    int64
Satisfaction         int64
Spending             int64
Computer            object
Text Messages        int64
dtype: object
```

Check for missing values in the dataset:

```
RangeIndex: 62 entries, 0 to 61
Data columns (total 14 columns):
ID                  62 non-null int64
Gender              62 non-null object
Age                 62 non-null int64
Class               62 non-null object
Major               62 non-null object
Grad Intention      62 non-null object
GPA                 62 non-null float64
Employment          62 non-null object
Salary              62 non-null float64
Social Networking   62 non-null int64
Satisfaction        62 non-null int64
Spending            62 non-null int64
Computer            62 non-null object
Text Messages       62 non-null int64
dtypes: float64(2), int64(6), object(6)
memory usage: 6.9+ KB
```

From the above description we see that there is no missing value present in the dataset.

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

Contingency Table: A cross-classification table showing the distribution of one row variable and a column variable. Contingency tables are useful to understand bivariate relationship between the constituent variables. Contingency tables may be constructed with more than 2 categorical variables.

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 a 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?

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

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

Prob (Male)= (Total number of male students)/ (Total number of students at the university).

Prob (Male)= 29/62 = 0.468

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

Prob (Female)= (Total number of female students)/ (Total number of students at the university)

Prob (Female)= 33/62 = 0.532 = 1 – Prob (Male)

2.3 Assume that the sample is a 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.

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

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

Count of Males = 29

P (Accounting| Male) = count of males selecting account/male count = 4/29 = 0.138
P (CIS| Male) = count of males selecting CIS /male count = 1/29 = 0.034
P (Economics| Male) = count of males selecting Economics /male count = 4/29 = 0.138
P (International| Male) = count of males selecting International /male count = 2/29 = 0.069
P (Mgmt.| Male) = count of males selecting Mgmt. /male count = 6/29 = 0.20
P (Other| Male) = count of males selecting other /male count = 4/29 = 0.138
P (Retail| Male) = count of males selecting Retail /male count = 5/29 = 0.172
P (Undecided| Male) = count of males are Undecided /male count = 3/29 = 0.103

2.3.2 Find the conditional probability of different majors among the female students of CMSU.
Note that sum of the above conditional probabilities is 1

Count of Female = 33

P (Accounting| Female) = count of Female selecting account/ Female count = 3/33 = 0.091
P (CIS| Female) = count of Female selecting CIS / Female count = 3/33 = 0.091
P (Economics| Female) = count of Female selecting Economics / Female count = 7/33 = 0.21
P (International| Female) = count of Female selecting Intl / Female count = 4/33 = 0.12
P (Mgmt.| Female) = count of Female selecting Mgmt. / Female count = 4/33 = 0.121
P (Other| Female) = count of Female selecting other / Female count = 3/33 = 0.091
P (Retail| Female) = count of Female selecting Retail / Female count = 9/33 = 0.28
P (Undecided| Female) = count of Female are Undecided / Female count = 0/33 = 0

Note that sum of the above conditional probabilities is 1

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

2.4.1 Let the event that a randomly chosen students is Male be denoted by M
The event that a randomly chosen student Intends to graduate be denoted by G
Prob (Male AND Intends to graduate) = P(M ∩ G)

From the contingency table Gender and Grad Intention, there are 17 male students who intend to graduate

Hence
P(M ∩ G) = 17 / 62 = 0.274

2.4.2 Let the event that a randomly chosen students is Female be denoted by F

The event that a randomly chosen student has a laptop be denoted by L
Hence the event that a randomly chosen student does not have a laptop be denoted by L^c

Prob(Female AND Does not have a laptop) = $P(F \cap L^c)$

From the contingency table gender and computer the number of female students not having a laptop is 2 + 2 = 4. (having desktops and tablets)

Hence
$P(F \cap L^c) = 4 / 62 = 0.06$

2.5 Assume that the sample is a 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 a full-time employment?
2.5.2 Find the conditional probability that given a female student is randomly chosen, she is majoring in international business or management.

2.5.1 Let the event that a randomly chosen students is Male be denoted by M
Let the event that a randomly chosen students has full-time employment be denoted by E
Prob(Male OR full-time employment) = P(M U E) = P(M) + P(E) - P(M ∩ E)
Where (M ∩ E) denotes the event that a randomly chosen student is a male AND has full-time employment.

P(M) = 29/62 = 0.468
P(E) = 10/62 = 0.16
P(M ∩ E) = 7/62 = 0.11

Hence P(M U E) = P(M) + P(E) - P(M ∩ E) = 0.468 + 0.16 – 0.11 = 0.518

2.5.2 **When dealing with conditional probability that the students chosen is a female, only the row where gender = Female in the table Gender and Major is of concern.**
Prob(International Business OR Management) = (4 + 4) / 33 = 0.242

	No	Yes	Total
Female	9	11	20
Male	3	17	20
Grand Total	12	28	40

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 graduate intention and being female are independent events?

Refer to the table above.

P(F) = 20/40 = 0.5
P(Yes) = 28/40 = 0.7

If being female and graduate intention are independent, the P(F ∩ Yes) = P(F)P(Yes)

P(F ∩ Yes) = 11 / 40 = 0.275
P(F)P(Yes) = 0.5(0.7) = 0.35 ≠ P(F ∩ Yes)

The two events are not independent

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

2.7.1 If a student is chosen randomly, what is the probability that his/her GPA is less than 3?
2.7.2 Find conditional probability that a randomly selected male earns 50 or more. Find conditional probability that a randomly selected female earns 50 or more.

2.7.1 Prob(GPA < 3) = 17 / 62 = 0.274

2.7.2 Prob(Salary ≥ 50 | Male) = 14/29 = 0.48
Prob(Salary ≥ 50 | Female) = 18/33 = 0.545

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.

For this we will test empirical rule: The empirical rule states that for a normal distribution, nearly all of the data will fall within three standard deviations of the mean. The empirical rule can be broken down into three parts:

- 68% of data falls within the first standard deviation from the mean.
- 95% fall within two standard deviations from the mean
- 99.7% fall within three standard deviations from the mean

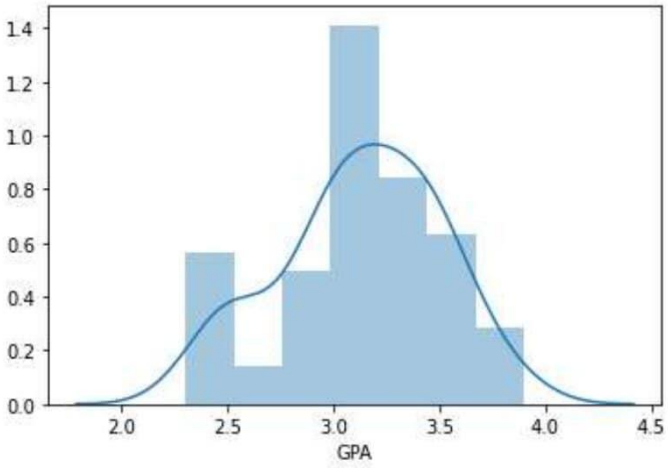
The rule is also called the 68-95-99.7 Rule or the Three Sigma Rule.

First we will calculate the mean and median and standard deviation for the variables.

GPA Variable:

GPA Mean: 3.13
GPA Median: 3.15
GPA Standard Deviation: 0.38

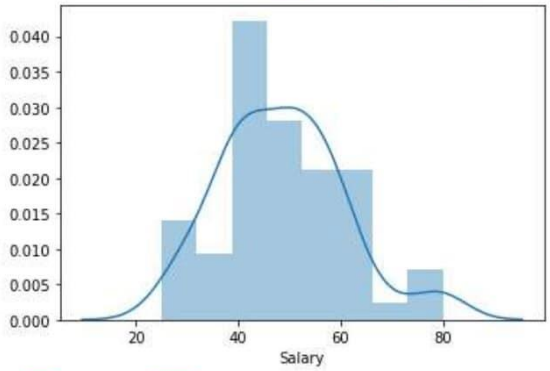
GPA Histogram:



Salary Variable:

Salary Mean: 48.55
Salary Median: 50.0
Salary Standard Deviation: 12.08

Histogram for Salary variable:

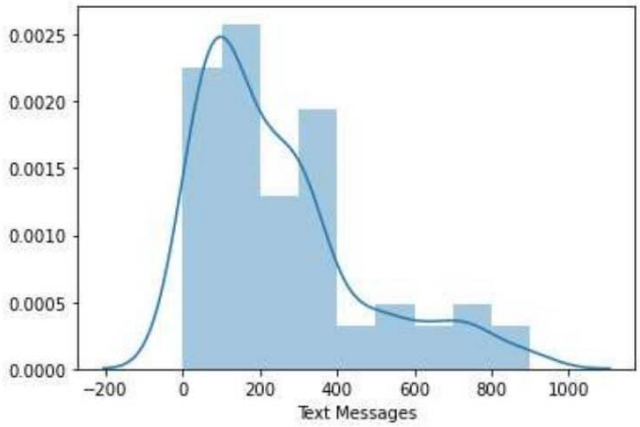


Text Messages Variable:

Text Messages Mean: 246.21
Text Messages Median: 200.0
Text Messages Standard Deviation: 214.47

Since mean and median of the Text Messages column has huge difference. It results that that data is highly skewed.

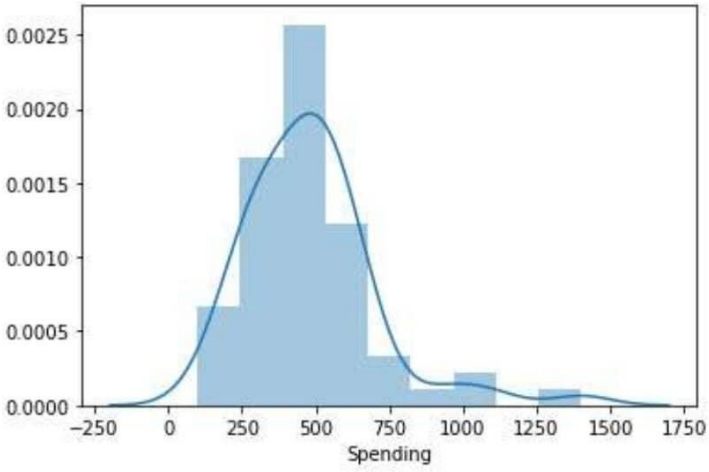
Histogram of Text messages



Spending Variable:

Spending Mean: 482.02
Spending Median: 500.0
Spending Standard Deviation: 221.95

Histogram of Spending



From the above analysis, we came to the result that variable (Salary, Text messages and Spending) are not normally distributed. Since the data is skewed (mean =! median) and the empirical rule also failed to propose that the data is normally distributed.

Appendix Code

Survey Analysis

Basic python packages load

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from scipy.stats import iqr #To calculate the IQR - Interquartile Range
import statsmodels.api as sm # to plot qqplot
from numpy.random import seed # To generate random seed
```

```
data = pd.read_csv("Survey.csv")
```

Exploratory Data Analysis

```
data.head()
```

	ID	Gender	Age	Class	Major	Grad Intention	GPA	Employment	Salary	Social Networking	Satisfaction	Spending	Computer	Tex Message
0	1	Female	20	Junior	Other	Yes	2.9	Full-Time	50.0	1	3	350	Laptop	20
1	2	Male	23	Senior	Management	Yes	3.6	Part-Time	25.0	1	4	360	Laptop	5
2	3	Male	21	Junior	Other	Yes	2.5	Part-Time	45.0	2	4	600	Laptop	20
3	4	Male	21	Junior	CIS	Yes	2.5	Full-Time	40.0	4	6	600	Laptop	25
4	5	Male	23	Senior	Other	Undecided	2.8	Unemployed	40.0	2	4	500	Laptop	10

```
data.dtypes
```

```
ID                      int64
Gender                 object
Age                     int64
Class                  object
Major                  object
Grad Intention         object
GPA                   float64
Employment             object
Salary                float64
Social Networking       int64
Satisfaction            int64
Spending                int64
Computer               object
Text Messages           int64
dtype: object
```

```
row, col = data.shape
print("There are total {}".format(row), "rows and {}".format(col), "columns in the dataset")
```

```
There are total 62 rows and 14 columns in the dataset
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 62 entries, 0 to 61
Data columns (total 14 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   ID                  62 non-null     int64  
 1   Gender              62 non-null     object 
 2   Age                 62 non-null     int64  
 3   Class               62 non-null     object 
 4   Major               62 non-null     object 
 5   Grad Intention      62 non-null     object 
 6   GPA                 62 non-null     float64
 7   Employment          62 non-null     object 
 8   Salary              62 non-null     float64
 9   Social Networking   62 non-null     int64  
 10  Satisfaction        62 non-null     int64  
 11  Spending            62 non-null     int64  
 12  Computer            62 non-null     object 
 13  Text Messages       62 non-null     int64  
dtypes: float64(2), int64(6), object(6)
memory usage: 6.9+ KB
```

```
data.describe(include = 'all')
```

	ID	Gender	Age	Class	Major	Grad Intention	GPA	Employment	Salary	Social Networking	Satisfaction
count	62.000000	62	62.000000	62	62	62	62.000000	62	62.000000	62.000000	62.000000
unique	NaN	2	NaN	3	8	3	NaN	3	NaN	NaN	NaN
top	NaN	Female	NaN	Senior	Retailing/Marketing	Yes	NaN	Part-Time	NaN	NaN	NaN
freq	NaN	33	NaN	31	14	28	NaN	43	NaN	NaN	NaN
mean	31.500000	NaN	21.129032	NaN	NaN	NaN	3.129032	NaN	48.548387	1.516129	3.741935
std	18.041619	NaN	1.431311	NaN	NaN	NaN	0.377388	NaN	12.080912	0.844305	1.213793
min	1.000000	NaN	18.000000	NaN	NaN	NaN	2.300000	NaN	25.000000	0.000000	1.000000
25%	16.250000	NaN	20.000000	NaN	NaN	NaN	2.900000	NaN	40.000000	1.000000	3.000000
50%	31.500000	NaN	21.000000	NaN	NaN	NaN	3.150000	NaN	50.000000	1.000000	4.000000
75%	46.750000	NaN	22.000000	NaN	NaN	NaN	3.400000	NaN	55.000000	2.000000	4.000000
max	62.000000	NaN	26.000000	NaN	NaN	NaN	3.900000	NaN	80.000000	4.000000	6.000000

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 Contingency Table showing relation between Gender and Major
d1 = pd.crosstab(data['Gender'], data['Major'], margins = True)
d1
```

]:

Major	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided	All
Gender									
Female	3	3	7	4	4	3	9	0	33
Male	4	1	4	2	6	4	5	3	29
All	7	4	11	6	10	7	14	3	62

]:
```
# 2.1.2 Contingency Table showing relation between Gender and Grad Intention
d2 = pd.crosstab(data['Gender'], data['Grad Intention'], margins = True)
d2
```

]:

Grad Intention	No	Undecided	Yes	All
Gender				
Female	9	13	11	33
Male	3	9	17	29
All	12	22	28	62

]:
```
# 2.1.3 Contingency Table showing relation between Gender and Employment
d3 = pd.crosstab(data['Gender'], data['Employment'], margins = True)
d3
```

]:

Employment	Full-Time	Part-Time	Unemployed	All
Gender				
Female	3	24	6	33
Male	7	19	3	29
All	10	43	9	62

]:
```
# 2.1.4 Contingency Table showing relation between Gender and Computer
d4 = pd.crosstab(data['Gender'], data['Computer'], margins = True)
d4
```

]:

Computer	Desktop	Laptop	Tablet	All
Gender				
Female	2	29	2	33
Male	3	26	0	29
All	5	55	2	62

2.2 Assume that the sample is a 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?

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

]:
```
#2.2.1 : probability that a randomly selected CMSU student will be male
Prob_Male = 29/62
print("probability that a randomly selected CMSU student will be male is: {}".format(Prob_Male))

# 2.2.2 probability that a randomly selected CMSU student will be female
Prob_FeMale = 33/62
print("probability that a randomly selected CMSU student will be female is: {}".format(Prob_FeMale))
```

```
probability that a randomly selected CMSU student will be male is: 0.46774193548387094
probability that a randomly selected CMSU student will be female is: 0.532258064516129
```

2.3 Assume that the sample is a 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.

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

```
d1
```

Major	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided	All
Gender									
Female	3	3	7	4	4	3	9	0	33
Male	4	1	4	2	6	4	5	3	29
All	7	4	11	6	10	7	14	3	62

```
#Male
round(d1.loc["Male"]/d1.loc['Male']['All'],2) #ignore  "ALL" in output
```

```
Major
Accounting                0.14
CIS                       0.03
Economics/Finance         0.14
International Business    0.07
Management                0.21
Other                     0.14
Retailing/Marketing       0.17
Undecided                 0.10
All                       1.00
Name: Male, dtype: float64
```

```
#Female
round(d1.loc["Female"]/d1.loc['Female']['All'],2)
```

```
Major
Accounting                0.09
CIS                       0.09
Economics/Finance         0.21
International Business    0.12
Management                0.12
Other                     0.09
Retailing/Marketing       0.27
Undecided                 0.00
All                       1.00
Name: Female, dtype: float64
```

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.

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

```
d2
```

Grad Intention	No	Undecided	Yes	All
Gender				
Female	9	13	11	33
Male	3	9	17	29
All	12	22	28	62

```
## 2.4.1 Find the probability That a randomly chosen student is a male and intends to graduate.
round(d2.loc['Male']['Yes']/d2.loc['All']['All'],3)
```

```
0.274
```

```
d4
```

Computer	Desktop	Laptop	Tablet	All
Gender				
Female	2	29	2	33
Male	3	26	0	29
All	5	55	2	62

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

round((d4.loc['Female']['Desktop']+d4.loc['Female']['Tablet'])/d4.loc['All']['All'],2)
```

```
0.06
```

2.5 Assume that the sample is a 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 a full-time employment?

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

1]:
```
d3
```

1]:

Employment	Full-Time	Part-Time	Unemployed	All
Gender				
Female	3	24	6	33
Male	7	19	3	29
All	10	43	9	62

P(M U E) = P(M) + P(E) - P(M ∩ E)

8]:
```
round((d3.loc['Male']['All'] + d3.loc['All']['Full-Time'] - d3.loc['Male']['Full-Time']) / d3.loc['All'
]['All'],3)
```

8]:
```
0.516
```

0]:
```
## 2.5.2 Find the conditional probability that given a female student is randomly chosen,
## she is majoring in international business or management.
d1
```

0]:

Major	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided	All
Gender									
Female	3	3	7	4	4	3	9	0	33
Male	4	1	4	2	6	4	5	3	29
All	7	4	11	6	10	7	14	3	62

3]:
```
round((d1.loc['Female']['International Business'] + d1.loc['Female']['Management']) / d1.loc['Female']['
'All'],3)
```

3]:
```
0.242
```

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 graduate intention and being female are independent events?

5]:
```
d = pd.crosstab(data['Gender'], data['Grad Intention'])
d.drop('Undecided', axis = 1, inplace=True) # Dropping Undecided column so that We have Intent to gradu
ate as Yes and No
d['Total'] = d.sum(axis=1) # Adding column totals
d
```

5]:

Grad Intention	No	Yes	Total
Gender			
Female	9	11	20
Male	3	17	20

P(F ∩ Yes)

9]:
```
d.loc['Female']['Yes']/d['Total'].sum()
```

9]:
```
0.275
```

P(F)P(Yes)

)3]:
```
(d.loc['Female']['Total']/d['Total'].sum()) * (d['Yes'].sum()/d['Total'].sum())
```

)3]:
```
0.35
```

)5]:
```
## The two events are not independent since P(F ∩ Yes) != P(F)P(Yes)
```

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

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

2.7.2 Find conditional probability that a randomly selected male earns 50 or more. Find conditional probability that a randomly selected female earns 50 or more.

7]:
```
data.head()
```

7]:

	ID	Gender	Age	Class	Major	Grad Intention	GPA	Employment	Salary	Social Networking	Satisfaction	Spending	Computer	Tex Message
0	1	Female	20	Junior	Other	Yes	2.9	Full-Time	50.0	1	3	350	Laptop	20
1	2	Male	23	Senior	Management	Yes	3.6	Part-Time	25.0	1	4	360	Laptop	5
2	3	Male	21	Junior	Other	Yes	2.5	Part-Time	45.0	2	4	600	Laptop	20
3	4	Male	21	Junior	CIS	Yes	2.5	Full-Time	40.0	4	6	600	Laptop	25
4	5	Male	23	Senior	Other	Undecided	2.8	Unemployed	40.0	2	4	500	Laptop	10

Prob(GPA < 3)

6]:
```
round(data[data['GPA']<3].groupby(['Gender']).count()['GPA'].sum()/data.Gender.count(),3)
```

6]:
```
0.274
```

2.7.2

Prob(Salary ≥ 50 | Female)

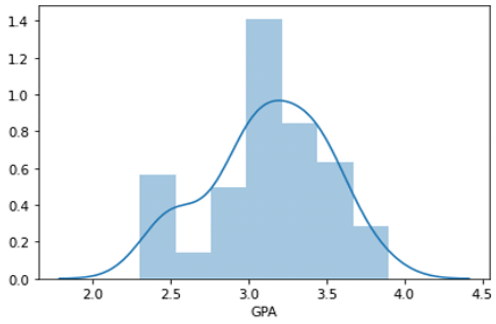
Prob(Salary ≥ 50 | Male)

7]:
```
round((data[data['Salary']>=50].groupby(['Gender']).count()['Salary'])/data.groupby('Gender').size(),3
)
```

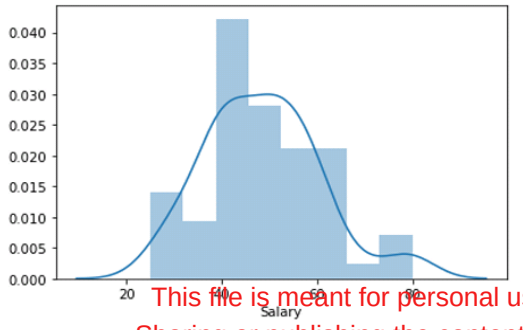
7]:
```
Gender
Female    0.545
Male      0.483
dtype: float64
```

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.

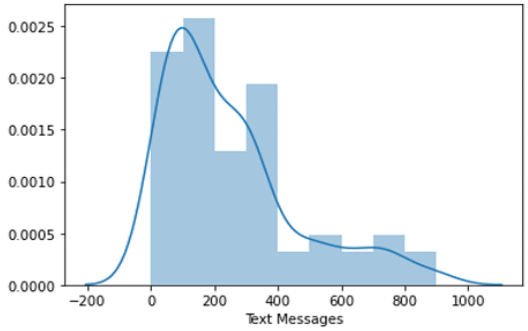
3]:
```
## Distrbution of GPA
sns.distplot(data['GPA'])
plt.show()
```



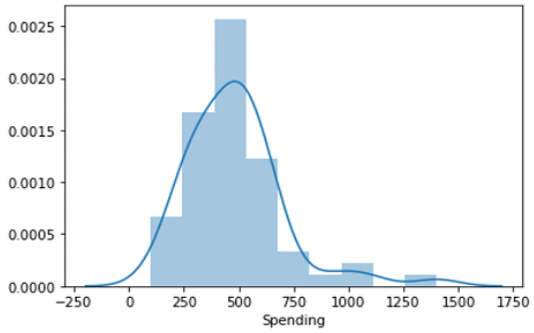
3]:
```
## Distribution of Salary
sns.distplot(data['Salary'])
plt.show()
```



```
## Distribution of Text Messages
sns.distplot(data['Text Messages'])
plt.show()
```




```
## Distribution of Spending
sns.distplot(data['Spending'])
plt.show()
```




```
df=data[['Salary','Spending','Text Messages']]
```

```
df.mean()
```

```
Salary            48.548387
Spending         482.016129
Text Messages    246.209677
dtype: float64
```

```
df.std()
```

```
Salary            12.080912
Spending         221.953805
Text Messages    214.465950
dtype: float64
```

```
def empirical(x):
    '''
      This custom function calculates the 68-95-99.7 Rule or the Three Sigma Rule.

    '''

    sd=x.std()
    mean=x.mean()

    #this will print the Lower Interval (Mu Minus One Sigma) and Upper Interval (Mu Minus One Sigma)
    print ('68% data should lie between {} and {}'.format(mean-sd,mean+sd))
    #this will print the % of values Lower & Upper Interval
    print('{}% data lies between LL and UL for 68%'.format(pd.Series((x> mean-sd) &
                                                                     (x< mean+sd)).value_counts(normali
ze=True).values[0]*100))
    print('\n')

    #this will print the Lower Interval (Mu Minus two Sigma) and Upper Interval (Mu Minus two Sigma)
    print ('95% data should lie between {} and {}'.format(mean-(2*sd),mean+(2*sd)))
    #this will print the % of values Lower & Upper Interval
    print('{}% data lies between LL and UL for 95%'.format(pd.Series((x> mean-(2*sd)) &
                                                                     (x< mean+(2*sd))).value_counts(nor
malize=True).values[0]*100))
    print('\n')

    #this will print the Lower Interval (Mu Minus three Sigma) and Upper Interval (Mu Minus three Sigm
a)
    print ('99% data should lie between {} and {}'.format(mean-(3*sd),mean+(3*sd)))
    #this will print the % of values Lower & Upper Interval
    print('{}% data lies between LL and UL for 99%'.format(pd.Series((x> mean+(3*sd)) &
                                                                     (x< mean+(3*sd))).value_counts(nor
malize=True).values[0]*100))
```

```
]: empirical(df['Salary'])
```

```
68% data should lie between 36.46747488043692 and 60.62929931311147
79.03225806451613% data lies between LL and UL for 68%


95% data should lie between 24.38656266409964 and 72.71021152944874
95.16129032258065% data lies between LL and UL for 95%


99% data should lie between 12.30565044776236 and 84.79112374578602
100.0% data lies between LL and UL for 99%
```

```
]: empirical(df['Spending'])
```

```
68% data should lie between 260.062324066296 and 703.9699339982201
80.64516129032258% data lies between LL and UL for 68%


95% data should lie between 38.10851910033398 and 925.9237389641821
95.16129032258065% data lies between LL and UL for 95%


99% data should lie between -183.84528586562806 and 1147.8775439301442
100.0% data lies between LL and UL for 99%
```

```
']: empirical(df['Text Messages'])
```

```
68% data should lie between 31.74372711665876 and 460.67562772205093
79.03225806451613% data lies between LL and UL for 68%


95% data should lie between -182.72222318603733 and 675.141578024747
91.93548387096774% data lies between LL and UL for 95%


99% data should lie between -397.18817348873336 and 889.6075283274431
100.0% data lies between LL and UL for 99%
```