Business report
on
Survey
dataset

Neha Mishra
PGP-DSBA Online
February 22
Date:27/02/2022

Table of Contents

Contents

Wholesale Customers Analysis

Clear Mountain State University

Quality Shingles

List of Figures

List of Tables

Executive Summary

The wholesaler wants to find out the relation between product along the different region. The different channel in the dataset are Retail and Hotel. Anddifferent regions are Other, Lisbon, Oporto.

Introdction

A wholesale distributor operating in different regions of Portugal has information on annual spending of several items in their stores across different regions and channels. The data consists of 440 large retailers' annual spending on 6 different varieties of products in 3 different regions (Lisbon, Oporto, Other) and across different sales channel (Hotel, Retail).

Data Description

1. Buyer/Spender : continuous from 1 to 440.
2. Channel : Retail and Hotel.
3. Region : Other, Lisbon and Oporto.
4. Fresh : continuous from 3.00 to 112151.00.
5. Milk : continuous from 3.00 to 112151.00.
6. Grocery : continuous from 3.00 to 112151.00
7. Frozen : continuous from 3.00 to 112151.00
8. Detergents Paper: continuous from 3.00 to 112151.00.
9. Delicatessen : continuous from 3.00 to 112151.00.

Sample of the dataset

	Buyer/Spender	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
0	1	Retail	Other	12669	9656	7561	214	2674	1338
1	2	Retail	Other	7057	9810	9568	1762	3293	1776
2	3	Retail	Other	6353	8808	7684	2405	3516	7844
3	4	Hotel	Other	13265	1196	4221	6404	507	1788
4	5	Retail	Other	22615	5410	7198	3915	1777	5185

Table no 1: Sample Dataset

Exploratory Data Analysis

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

Buyer/Spender : int64
Channel : object
Region : object
Fresh : int64
Milk : int64
Grocery : int64
Frozen : int64
Detergents Paper : int64
Delicatessen : int64

Check for missing values in the dataset:

Range Index: 440 entries, 0 to 439
Data columns (total 9 columns):

#	Column	Non-Null	Count	Dtype
---	---	---	- -------	--------
0	Buyer/Spender	440	non-null	int64
1	Channel	440	non-null	object
2	Region	440	non-null	object
3	Fresh	440	non-null	int64
4	Milk	440	non-null	int64
5	Grocery	440	non-null	int64
6	Frozen	440	non-null	int64
7	Detergents_Paper	440	non-null	int64
8	Delicatessen	440	non-null	int64

dtypes: int64(7), object (2)

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

1.1 Use methods of descriptive statistics to summarize data. Which Region and which Channel spent the most? Which Region and which Channel spent the least?

	count	unique	Top	freq	mean	std	min	25%	50%	75%	max
Buyer/Spender	440.0	NaN	NaN	NaN	220.5	127.161315	1.0	110.75	220.5	330.25	440.0
Channel	440	2	Hotel	298	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Region	440	3	Other	316	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Fresh	440.0	NaN	NaN	NaN	12000.297727	12647.328865	3.0	3127.75	8504.0	16933.75	112151.0
Milk	440.0	NaN	NaN	NaN	5796.265909	7380.377175	55.0	1533.0	3627.0	7190.25	73498.0
Grocery	440.0	NaN	NaN	NaN	7951.277273	9503.162829	3.0	2153.0	4755.5	10655.75	92780.0
Frozen	440.0	NaN	NaN	NaN	3071.931818	4854.673333	25.0	742.25	1526.0	3554.25	60869.0
Detergents_Paper	440.0	NaN	NaN	NaN	2881.493182	4767.854448	3.0	256.75	816.5	3922.0	40827.0
Delicatessen	440.0	NaN	NaN	NaN	1524.870455	2820.105937	3.0	408.25	965.5	1820.25	47943.0

Table no.2: Summarize table

So, before going to region and channel which product spent least and most, we will get some result from above descriptive table.

25%, 50%, 75%, min, max, standard deviation and mean for the variable fresh, milk, grocery, frozen, detergentspaper, delicatessen is clear.

	Region	Channel	Fresh	Milk	Grocery	Frozen	Detergents_ Paper	Delicatessen	Total
0	Lisbon	Hotel	761233	228342	237542	184512	56081	70632	1538342
1	Lisbon	Retail	93600	194112	332495	46514	148055	33695	848471
2	Oporto	Hotel	326215	64519	123074	160861	13516	30965	719150
3	Oporto	Retail	138506	174625	310200	29271	159795	23541	835938
4	Other	Hotel	2928269	735753	820101	771606	165990	320358	5742077
5	Other	Retail	1032308	1153006	1675150	158886	724420	191752	4935522

Table no.3: most least summarize table

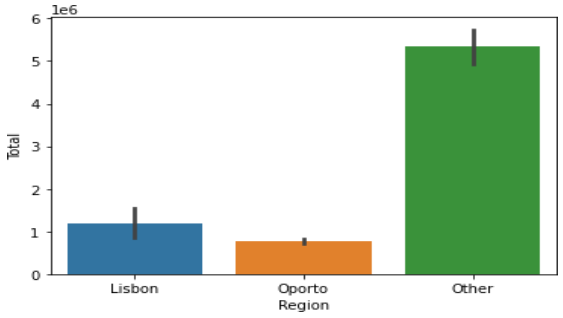

Fig1: total vs region

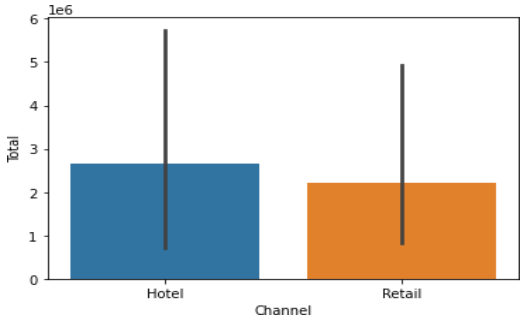

Fig2: total vs channel

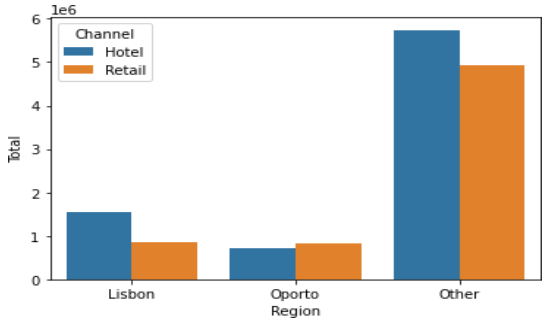

Fig3: total vs region vs channel

From above table and graph, it is clear that

1)other region spent the most,

2)oporto region spent the least,

3)hotel channel spent the most,

4)retail channel spent the least.

1.2 There are 6 different varieties of items that are considered. Describe and comment/explain all the varieties across Region and Channel? Provide a detailed justification for your answer.

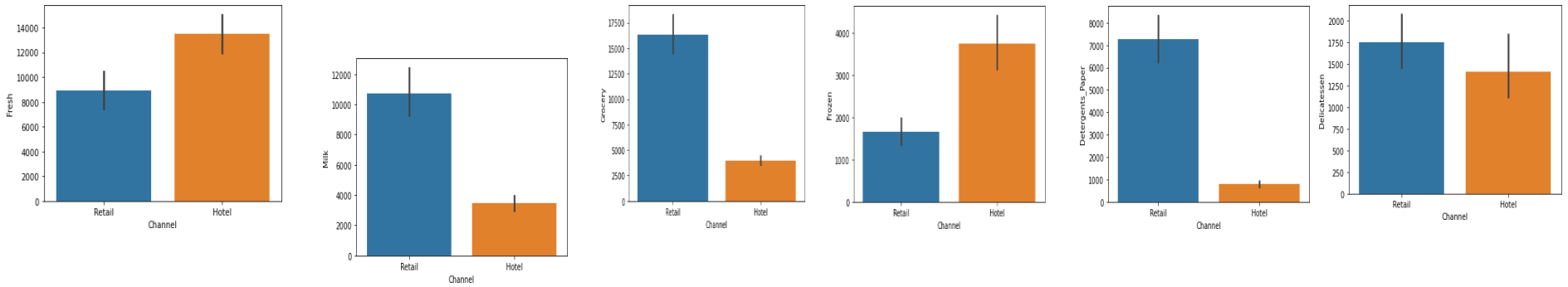

Fig4: channel vs individual varieties

From the above graph,

fresh and frozen spent max in hotel region and
milk, grocery, detergent paper and delicatessen spent max in retail region individually

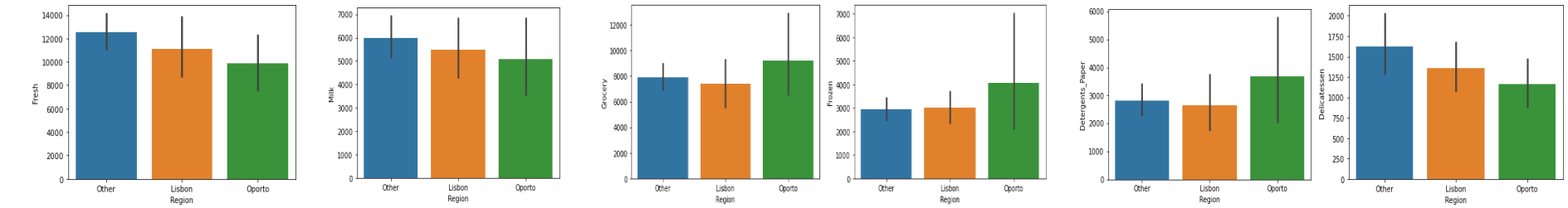

fig5: region vs individual varieties

From the above graph,

Fresh, milk and delicatessen spent max in other region and
grocery, frozen, detergent paper and delicatessen spent max in oporto region
All varieties spent moderate in Lisbon region individually

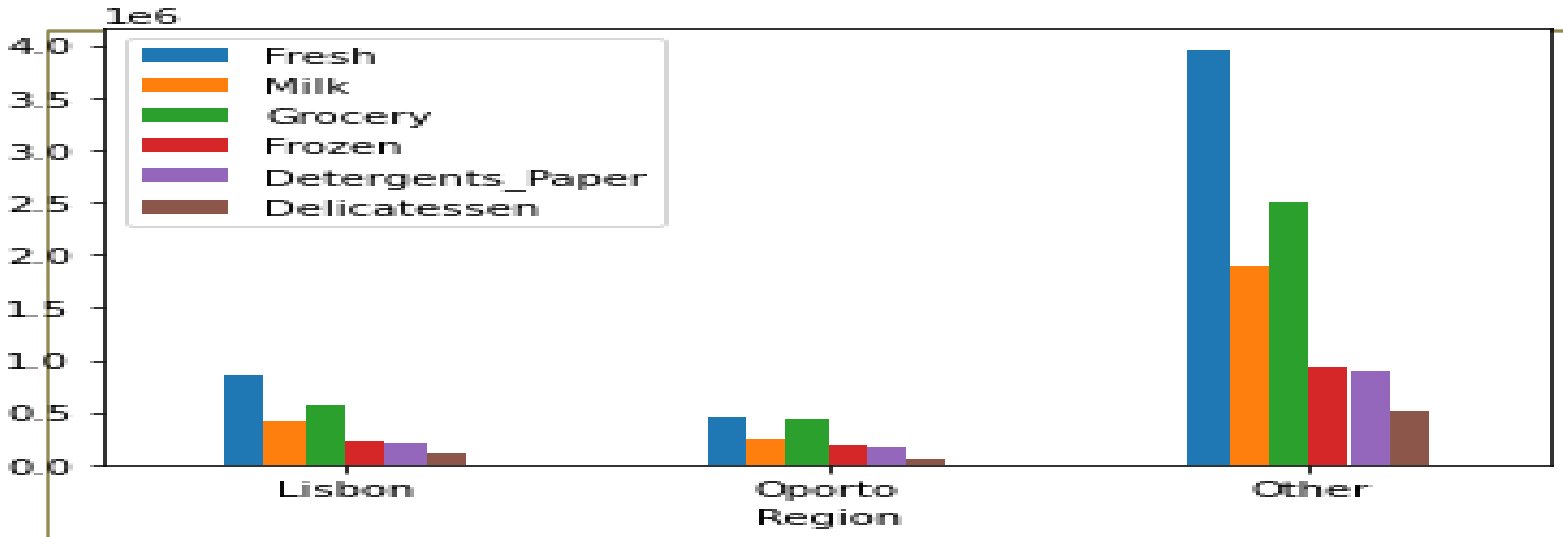

Fig6 : region vs varieties

So, the graph says other region spent max in all varieties and oporto region spent less in all varieties.

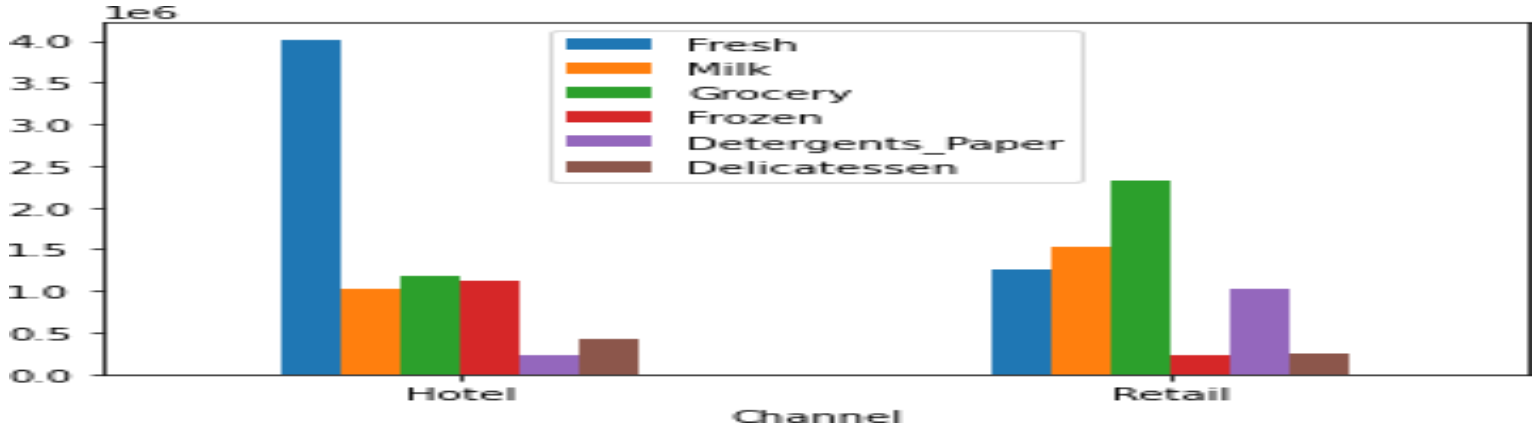

Fig7: channel vs varieties

From the above graph fresh and grocery spent max in hotel and retail channel and delicatessen spent least in both

1.3 On the basis of a descriptive measure of variability, which item shows the most inconsistent behavior? Which items show the least inconsistent behavior?

Table no 4: IQR Table	
Fresh	13806.00
Milk	5657.25
Grocery	8502.75
Frozen	2812.00
Detergents Paper	3665.25
Delicatessen	1412.0
Total	23858.75

Table no 5: Standard Deviation	
Fresh	12647.328865
Milk	7380.377175
Grocery	9503.162829
Frozen	4854.673333
Detergents Paper	4767.854448
Delicatessen	2820.105937
Total	26356.301730

In the IQR Table and Standard deviation,
Fresh varieties having max IQR and std so we can say that it is most inconsistentAnd
Delicatessen have least IQR and std so we can say that it is least inconsistent.

1.4 Are there any outliers in the data? Back up your answer with a suitable plot/technique with the help of detailed comments.

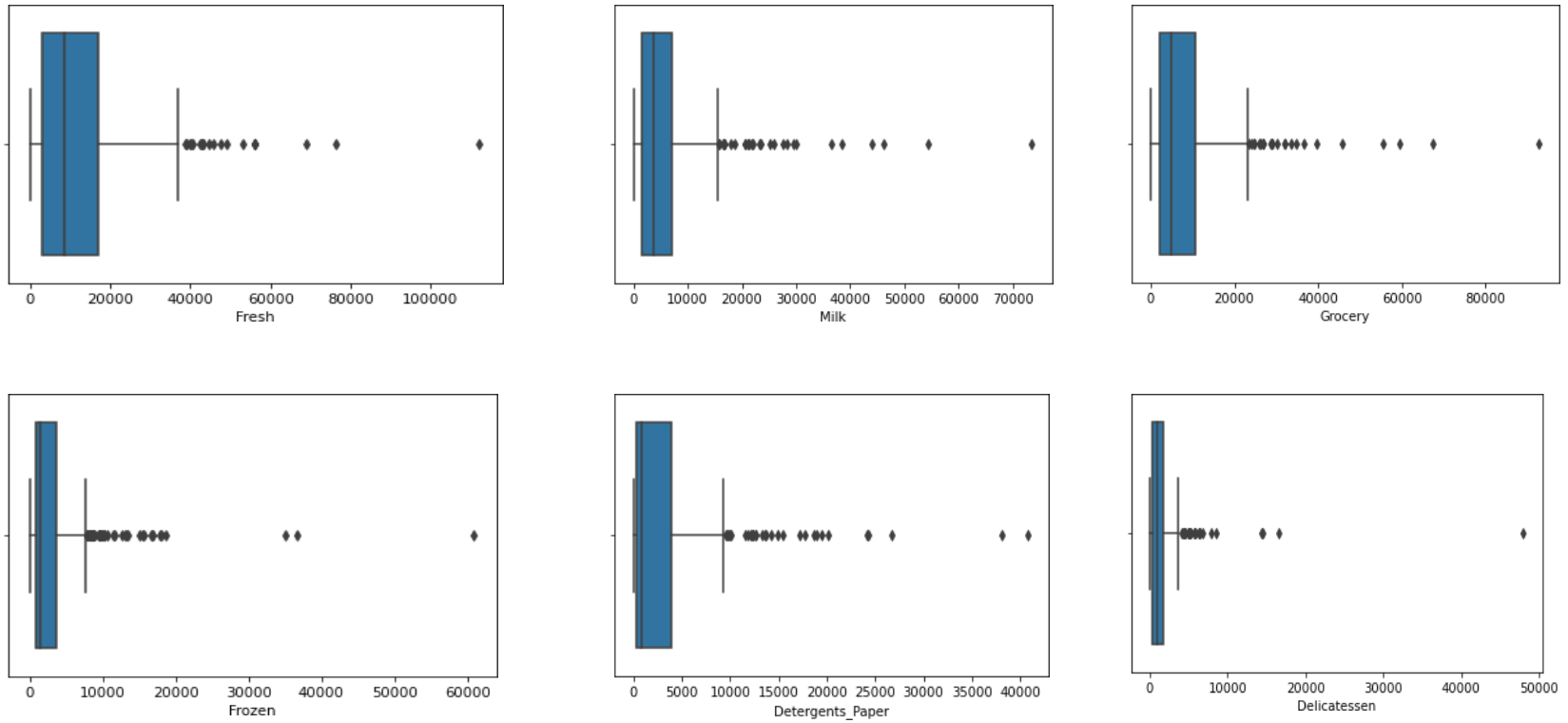

Fig8 : box plot

From the boxplot for all six varieties each have outliers.

1.5 On the basis of your analysis, what are your recommendations for the business? How can your analysis help the business to solve its problem? Answer from the business perspective

On the basis of analysis, plots and table between different variable along the region and channel, region wise other region showing good result but oporto region needs to make better strategies to get more spender. As per the varieties fresh varieties having good spender over different channel and region. Frozen, Detergents Paper and Delicatessen are less popular over spender.

Problem-2
Executive Summary

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.

Data Description

1. Age : continuous from 18 to 26.
2. Gender : Male, Female
3. ID : Retail and Hotel.
4. Class : Junior, Senior, Sophomore.
5. Major : Other, Management, CIS, Economics/Finance, Undecided, International Business, Retailing/Marketing, Accounting.
6. Grad Intention : Yes, No, Undecided.
7. GPA :continuous from 2.3 to 3.9.
8. Employment : Full time, Part time, Unemployed.
9. Salary : continuous from 25 to 80.
10. Social networking : continuous from 0 to 4.
11. Satisfaction : continuous from 1 to 6.
12. Spending : continuous from 100 to 1400
13. Computer : Laptop, Tablet, Desktop.
14. Text messages : continuous from 0 to 900.

Sample of the dataset

	ID	Gender	Age	Class	Major	Grad Intention	GPA	Employment	Salary	Social Networking	Satisfaction	Spending	Computer	Text Messages
0	1	Female	20	Junior	Other	Yes	2.9	Full-Time	50.0	1	3	350	Laptop	200
1	2	Male	23	Senior	Management	Yes	3.6	Part-Time	25.0	1	4	360	Laptop	50
2	3	Male	21	Junior	Other	Yes	2.5		45.0	2	4	600	Laptop	200
3	4	Male	21	Junior	CIS	Yes	2.5	Full-Time	40.0	4	6	600	Laptop	250
4	5	Male	23	Senior	Other	Undecided	2.8	Unemployed	40.0	2	4	500	Laptop	100

Table no 6: Sample Dataset

Exploratory Data Analysis

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:

Range Index: 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)

From the above results we can 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

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

Table no 7 : Gender vs Major

From above contingency table, it is clear that male and female participating in different major are 29 and 33 respectively.

2.1.2. Gender and Grad Intention

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

Table no 8: Gender vs Grad Intention

2.1.3 Gender and Employment

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

Table no 9: Gender vs Employment

2.1.4. Gender and Computer

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

Table no 10: Gender vs Computer

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?

from the contingency table 2, the distribution of male and female is clear. So, the probability of selection of male student we will denote by p_male

total_no_student = 62
total_no_male_student = 29
p_male=(total_no_male_student/total_no_student)*100
the probability of selection of male candidate is: 46.774193548387096

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

The probability of selection of female student we will denote by p_female

total_no_student = 62
total_no_female_student = 33
p_female=total_no_female_student/total_no_student*100

the probability of selection of female candidate 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

From the contingency table between gender and major we will get the right information. As the total no. of male candidate are 29
The participation male candidate in accounting is 4. we will denote it by m_acc.
The participation male candidate in CIS is 1. we will denote it by m_cis.
The participation male candidate in Economics/Finance is 4. we will denote it by m_ef.
The participation male candidate in International Business is 2. we will denote it by m_ib.
The participation male candidate in Management is 6. we will denote it by m_mg.
The participation male candidate in Other is 4. we will denote it by m_ot.
The participation male candidate in Retailing/Marketing is 5. we will denote it by m_rm.
The participation male candidate in Undecided is 3. we will denote it by m_un.

p_acc=((m_acc/total_no_male_student)/p_male)*100
p_cis=((m_cis/total_no_male_student)/p_male)*100
p_ef=((m_ef/total_no_male_student)/p_male)*100
p_ib=((m_ib/total_no_male_student)/p_male)*100
p_mg=((m_mg/total_no_male_student)/p_male)*100
p_ot=((m_ot/total_no_male_student)/p_male)*100
p_rm=((m_rm/total_no_male_student)/p_male)*100
p_un=((m_un/total_no_male_student)/p_male)*100

The probability of selection accounting as a male is : 29.488703923900122
The probability of selection CIS as a male is : 7.3721759809750305
The probability of selection Economics/Finance as a male is : 29.488703923900122
The probability of selection International Business as a male is : 14.744351961950061
The probability of selection Management as a male is : 44.23305588585018
The probability of selection Other as a male is : 29.488703923900122
The probability of selection Retailing/Marketing as a male is : 36.860879904875155
The probability of selection undecided as a male is : 22.11652794292509

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

From the contingency table 2 between gender and major we will get the right information. As the total no. of female candidate are 33

The participation female candidate in accounting is 4. we will denote it by f_acc.
The participation female candidate in CIS is 1. we will denote it by f_cis.
The participation female candidate in Economics/Finance is 4. we will denote it by f_ef.
The participation female candidate in International Business is 2. we will denote it by f_ib.
The participation female candidate in Management is 6. we will denote it by f_mg.
The participation female candidate in Other is 4. we will denote it by f_ot.
The participation female candidate in Retailing/Marketing is 5. we will denote it by f_rm.

pf_acc=((f_acc/total_no_female_student)/p_female)*100
pf_cis=((f_cis/total_no_female_student)/p_female)*100
pf_ef=((f_ef/total_no_female_student)/p_female)*100
pf_ib=((f_ib/total_no_female_student)/p_female)*100
pf_mg=((f_mg/total_no_female_student)/p_female)*100
pf_ot=((f_ot/total_no_female_student)/p_female)*100
pf_rm=((f_rm/total_no_female_student)/p_female)*100

The probability of selection accounting as a female is: 17.079889807162534
The probability of selection CIS as a female is: 17.079889807162534
The probability of selection Economics/Finance as a female is: 39.85307621671259
The probability of selection International Business as a female is: 22.773186409550046
The probability of selection Management as a female is: 22.773186409550046
The probability of selection Other as a female is: 17.079889807162534
The probability of selection Retailing/Marketing as a female is: 51.2396694214876
As the undecided female are zero the probability will be zero

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.

total_no_male_student = 29
total_no_student_male_intends_to_graduate = 17
pm_grad=(total_no_student_male_intends_to_graduate/total_no_male_student)*100

The probability of male student want to graduate is : 58.620689655172406

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

total_no_female_student = 33
total_no_student_female_intends_to_graduate = 11
pf_grad=(total_no_student_female_intends_to_graduate/total_no_female_student)*100

The probability of female student want to graduate is: 33.33333333333333

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 a male or has full-time employment?

total_no_male_student = 29
total_no_ft_emp = 10
male_no_ft_emp = 7
total_ss = 62

pm_select =((total_no_male_student/total_ss)+(total_no_ft_emp/total_ss)-(male_no_ft_emp/total_ss))*100

Hence the probability that a randomly chosen student is a male or has full-time employment : 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.

total_no_female_student = 33
f_ib = 4

pib_select = ((f_ib/total_no_female_student)/p_female) *100

Hence the conditional probability that given a female student is randomly chosen, she is majoring in international business or management: 22.773186409550046

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	33
Male	3	17	29
All	12	28	62

Table no 11: Gender vs Grad Intention

the_no_student_want_to_graduate = 28
the_no_female_want_to_graduate = 11
total_ss = 62
prob_stu_gradute=(the_no_student_want_to_graduate/total_ss) *100
prob_fe_stu_gradute=(the_no_female_want_to_graduate/total_ss) *100

Probability of student graduate: 45.16129032258064
Probability of female student graduate: 17.741935483870968

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?

As we have calculated from code, the no of GPA less than 3 (df_3) is 17

prob_gpa_less3 = (df_3/total_ss) *100

the probability that his/her GPA is less than 3: 27.419354838709676

2.7.2. 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.

	ID	Gender	Age	Class	Major	Grad Intention	GPA	Employment	Salary	Social Networking	Satisfaction	Spending	Computer	Text Messages
0	1	Female	20	Junior	Other	Yes	2.9	Full-Time	50.0	1	3	350	Laptop	200
5	6	Female	22	Senior	Economics/Finance	Undecided	2.3	Unemployed	78.0	3	2	700	Laptop	30
6	7	Female	21	Junior	Other	Undecided	3.0	Part-Time	50.0	1	3	500	Laptop	50
7	8	Female	22	Senior	Other	Undecided	3.1	Full-Time	80.0	1	2	200	Tablet	300
10	11	Female	23	Senior	Economics/Finance	Yes	2.8	Full-Time	50.0	2	5	400	Laptop	200

Table no 12: Gender earns 50 or more

Salary	50.0	52.0	54.0	55.0	60.0	65.0	70.0	78.0	80.0	All
Gender										
Female	5	0	0	5	5	0	1	1	1	18
Male	4	1	1	3	3	1	0	0	1	14
All	9	1	1	8	8	1	1	1	2	32

Table no : Gender vs salary earn 50 or more

from the above contingency table, the no. male and female candidate earn 50 or more are 14 and 18 respectively.

total_male_50 = 14
total_female_50 = 18

prob_male_50=((total_male_50/total_ss)/p_male)*100

prob_female_50=((total_female_50/total_ss)/p_female)*100

the conditional probability that a randomly selected male earns 50 or more: 48.275862068965516

the conditional probability that a randomly selected female earns 50 or more: 54.545454545454554

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.

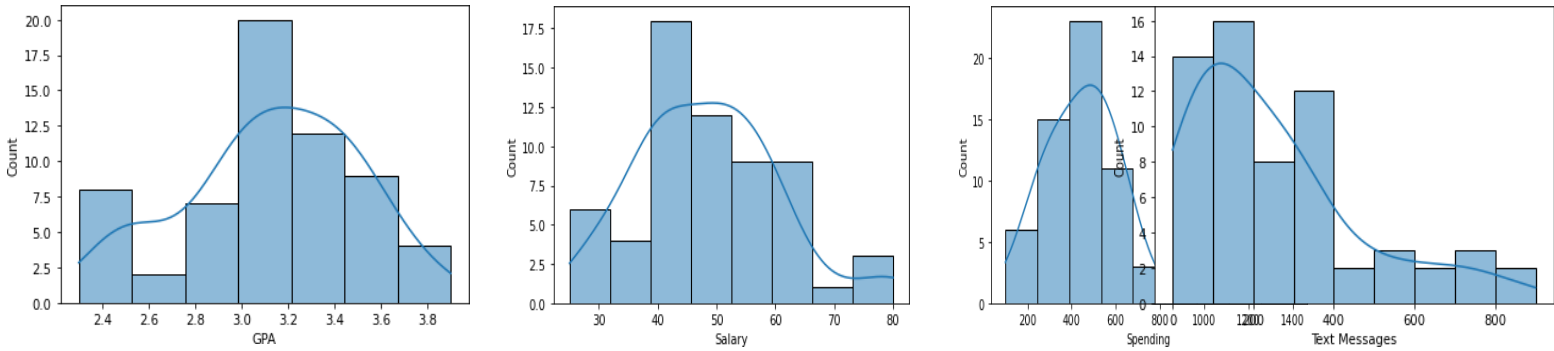

Fig 9 : histogram for variable

```
The skewness

of GPA is:-

0.3146000894

506981 The

skewness of

Salary is

:0.534700843

6225946

The skewness of Spending is :1.5859147414045331

The skewness of Text Messages is :1.2958079731054333
```

All four continuous variables GPA, Salary, Spending, Text Messages showing continuous line on Histogram. So, we can say than they follow a normal distribution, from the graph and skewness GPA is left skewed and Salary, Spending and Text messages are right skewed.

As per the analysis, the participation of female candidate is more than the male candidate but male intended to graduate more. As we have calculated male preferring full time employment. If we talk about the earning part both the gender person earn quite good.

Problem-3
Executive Summary

An important quality characteristic used by the manufacturers of ABC asphalt shingles is the amount of moisture the shingles contain when they are packaged. Customers may feel that they have purchased a product lacking in quality if they find moisture and wet shingles inside the packaging. In some cases, excessive moisture can cause the granules attached to the shingles for texture and coloring purposes to fall off the shingles resulting in appearance problems. To monitor the amount of moisture present, the company conducts moisture tests. A shingle is weighed and then dried. The shingle is then reweighed, and based on the amount of moisture taken out of the product, the pounds of moisture per 100 square feet are calculated. The company would like to show that the mean moisture content is less than 0.35 pounds per 100 square feet. Exploratory.

A	B
0.44	0.14
0.61	0.15
0.47	0.31
0.3	0.16
0.15	0.37

Dataset has 2 variables A & B and both are float data type.

Descriptive Statistics for the dataset:

	A	B
count	36.000000	31.000000
mean	0.316667	0.273548
std	0.135731	0.137296
min	0.130000	0.100000
25%	0.207500	0.160000
50%	0.290000	0.230000
75%	0.392500	0.400000
max	0.720000	0.580000

Check for Null values:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   A       36 non-null     float64
 1   B       31 non-null     float64
dtypes: float64(2)
memory usage: 640.0 bytes
```

From the above results, it is evident that there is no null values present in the dataset.

A have 36 record and B have 31 data record.

3.1 Do you think there is evidence that mean moisture contents in both types of shingles are within the permissible limits? State your conclusions clearly showing all steps.

❖ Calculate population mean moisture content

for shingles A: Step 1: Define null and alternative

hypotheses

Since the population standard deviation (Sigma) is unknown, we have to use a Tstat test in this hypothesis testing.

In testing the population mean moisture content for shingles A:

- Null hypothesis: H0 = $\mu \geq 0.35$.
- Alternative hypothesis: Ha = $\mu < 0.35$

Step 2: Decide the significance level

Here we select $\alpha = 0.05$.

The sample size for this problem is 36

Step 3: Identify the test statistic

We do not know the population standard deviation and n-1 = 36-1=35. So, we use the t distribution and the tSTAT test statistic.

Step 4: Calculate the p - value and test statistic

One sample t test
t-statistic: -1.4735046253382782 p-value: 0.14955266289815025

Step 5: Decide to reject or accept null hypothesis

p value is 0.14955266289815025 and it is greater than 5% level of significance
So, the statistical decision is failing to reject the null hypothesis at 5% level of significance

Hence, there is not enough evidence to support the claim that the population mean moisture content of shingles A is less than 0.35 pound per 100 square feet, at the 0.05 significance level.

❖ **Calculate population mean moisture content for shingles B:**

Step 1: Define null and alternative hypotheses

Since the population standard deviation (Sigma) is unknown, we have to use a Tstat test in this hypothesis testing.

In testing the population mean moisture content for shingles B:

- Null hypothesis: H0 = $\mu \geq 0.35$.
- Alternative hypothesis: Ha = $\mu < 0.35$

Step 2: Decide the significance level

Here we select $\alpha = 0.05$.

The sample size for this problem is 31

Step 3: Identify the test statistic

We do not know the population standard deviation and n-1 = 31-1=30. So, we use the t distribution and the tSTAT test statistic.

Step 4: Calculate the p - value and test statistic

One sample t test

t-statistic: -3.1003313069986995 p-value: 0.004180954800638363

Step 5: Decide to reject or accept null hypothesis

Level of significance: 0.05

We have evidence to reject the null hypothesis since p value < Level of significance

Our one-sample t-test p-value= 0.004180954800638363

Hence, there is enough evidence to support the claim that the population mean moisture content of shingles B is less than 0.35 pound per 100 square feet, at the 0.05 significance level.

3.2 Do you think that the population mean for shingles A and B are equal? Form the hypothesis and conduct the test of the hypothesis. What assumption do you need to check before the test for equality of means is performed?

Step 1: Define null and alternative hypotheses

In testing whether the population mean for shingles A and B are equal, the null hypothesis states that the population mean of shingles A and shingles B are the same, equals. The alternative hypothesis states that the population mean of shingles A and shingles B are different, is not equals.

- H_0: $\mu A = \mu B$
- H_A: $\mu A \neq \mu B$

Step 2: Decide the significance level

Here we select = 0.05 and the population standard deviation is not known.

Step 3: Identify the test statistic

- We have two samples and we do not know the population standard deviation.
- Sample sizes for both samples are not same.
- Degree of freedom of A (n) = 36 and Degree of freedom of B (n) = 31
- This is a two - tailed test.

Step 4: Calculate the p - value and test statistic

Two sample t test
t-statistic: 1.289628271966112 p-value: 0.2017496571835328

Step 5: Decide to reject or accept null hypothesis

Level of significance: 0.05
We have no evidence to reject the null hypothesis since p value > Level of significance
Our two-sample t-test p-value= 0.2017496571835328

The results given indicate that there is no significant difference between the population averages,

Whereas the Null Hypothesis of equality of means is accepted.

Hence, there is enough evidence to support the claim that the population means for shingles A and B are equal

Assumptions made:

*** The data are normally distributed**

*** For two sample means, the population variances of shingles A and B are not equal**

Conclusion:

Because we did not have the population standard deviation given, we have to go with T-test in this case, which states that the population means for shingles A and B are equal.