BUSINESS REPORT

SMDM - Project: V.1

Submitted to: Great Learning

Submitted by: Azharuddin Shaikh

Date: 27th March 2022

Contents

Problem 1: Wholesale Customer Data

1.1.1 Use methods of descriptive statistics to summarize data.

    1.1.2 Which Region and which Channel spent the most?

    1.1.3 Which Region and which 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.

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

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

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.

Problem 2: Mountain State University (CMSU)

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 Employmen

    2.1.4. Gender and Computer

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?

2.3. 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

2.4. 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

2.5. 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 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

2.6 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?

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.

2.8.1 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.

2.8.2 Write a note summarizing your conclusions

Problem 3: Manufacturers of ABC asphalt shingles

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

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?

Problem 1

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).

Note: Before reviewing the solutions let's understand the basic Data Structure first.

Database Structure:

Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
Retail	Other	12669	9656	7561	214	2674	1338
Retail	Other	7057	9810	9568	1762	3293	1776
Retail	Other	6353	8808	7684	2405	3516	7844
Hotel	Other	13265	1196	4221	6404	507	1788
Retail	Other	22615	5410	7198	3915	1777	5185
...	...	...	...	...	...	...	...
Hotel	Other	29703	12051	16027	13135	182	2204
Hotel	Other	39228	1431	764	4510	93	2346
Retail	Other	14531	15488	30243	437	14841	1867
Hotel	Other	10290	1981	2232	1038	168	2125
Hotel	Other	2787	1698	2510	65	477	52

Data Information:

RangeIndex: 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)

Buyer/Spender	0
Channel	0
Region	0
Fresh	0
Milk	0
Grocery	0
Frozen	0
Detergents_Paper	0
Delicatessen	0
dtype: int64 : **There is No Null Value**	

1.1.1 Use methods of descriptive statistics to summarize data.

	count	mean	std	min	25%	50%	75%	max
Buyer/Spender	440.0	220.500000	127.161315	1.0	110.75	220.5	330.25	440.0
Fresh	440.0	12000.297727	12647.328865	3.0	3127.75	8504.0	16933.75	112151.0
Milk	440.0	5796.265909	7380.377175	55.0	1533.00	3627.0	7190.25	73498.0
Grocery	440.0	7951.277273	9503.162829	3.0	2153.00	4755.5	10655.75	92780.0
Frozen	440.0	3071.931818	4854.673333	25.0	742.25	1526.0	3554.25	60869.0
Detergents_Paper	440.0	2881.493182	4767.854448	3.0	256.75	816.5	3922.00	40827.0
Delicatessen	440.0	1524.870455	2820.105937	3.0	408.25	965.5	1820.25	47943.0

1.1.2 Which Region and which Channel spent the most?

1.1.3 Which Region and which Channel spent the least?

Let's review the following table and Bar plot to find out the answer

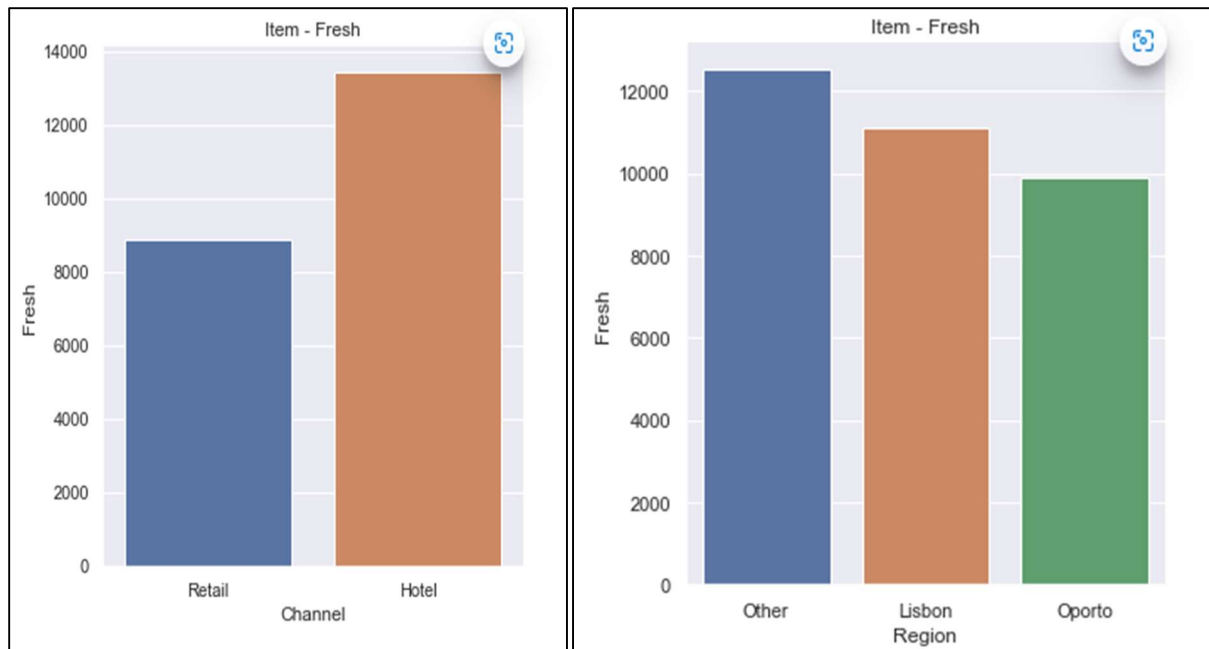| Region				Channel		
Region	Spending	Remark		Channel	Spending	Remark
Other	10677599	**High Spent**		Hotel	7999569	**High Spent**
Lisbon	2386813			Retail	6619931	**Low Spent**
Oporto	1555088	**Low Spent**				
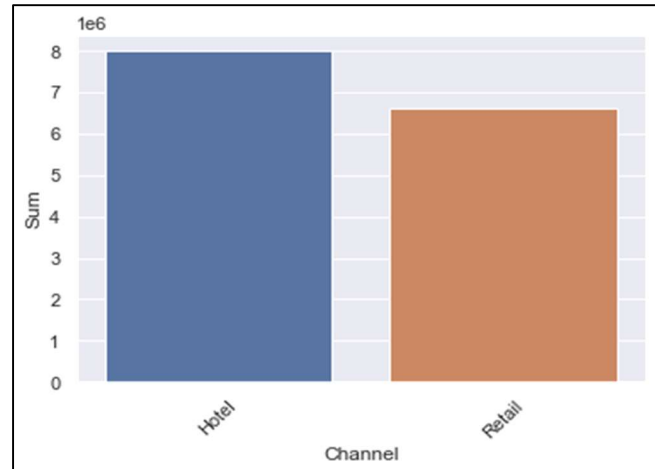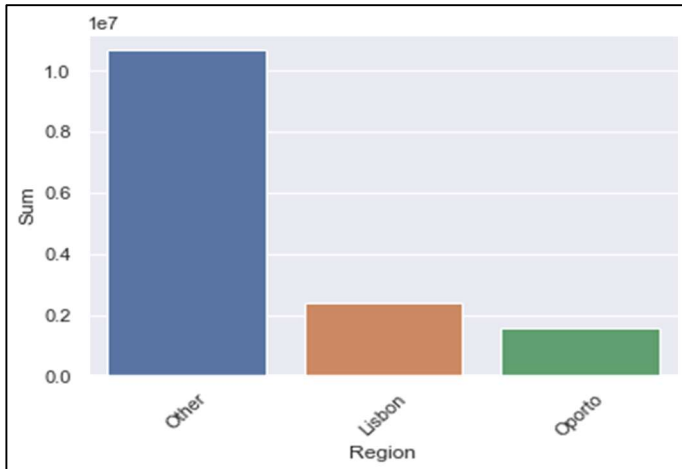
Analytics' Observation:

➢ **1.1.2:** Highest spend in the Region is from **Others** and Highest spend in the Channel is from **Hotel**

➢ **1.1.3:** The lowest spend in the region is from **Oporto** and lowest spend in the Channel is from **Retail**

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.

Let's review the flowing catplot against each varieties of items along with observation
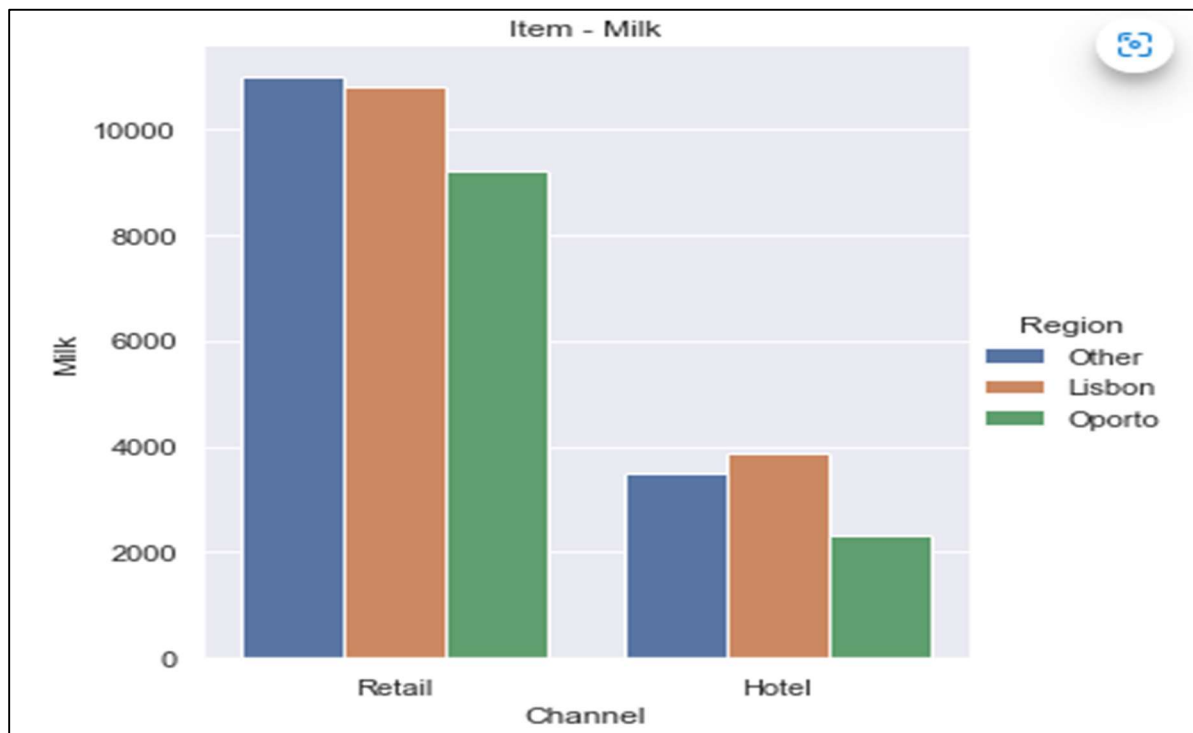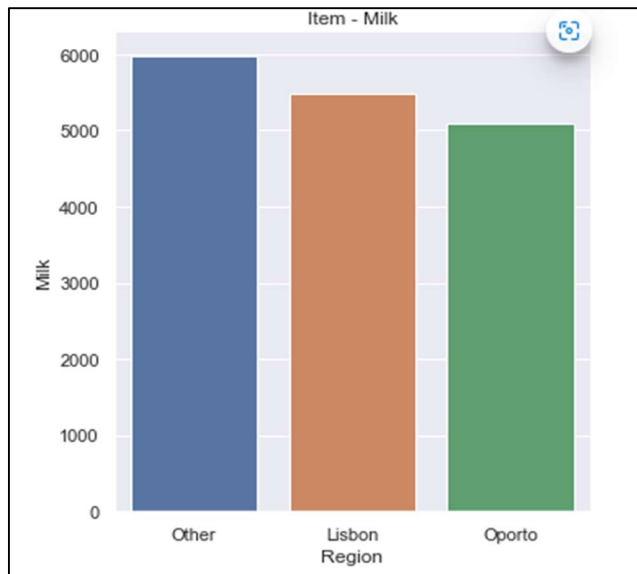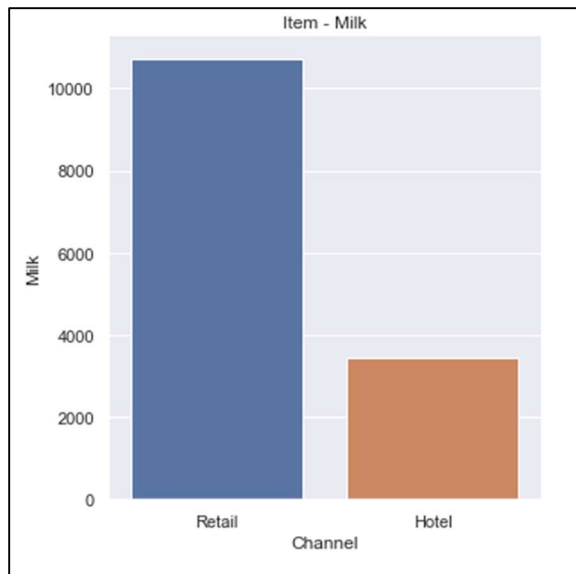
Item 1 – Fresh

Analytics' Observation:

➢ Based on the plot, Fresh item is sold more in hotel channel, on a regional level it has low numbers for Oporto as compared to Lisbon and other region.
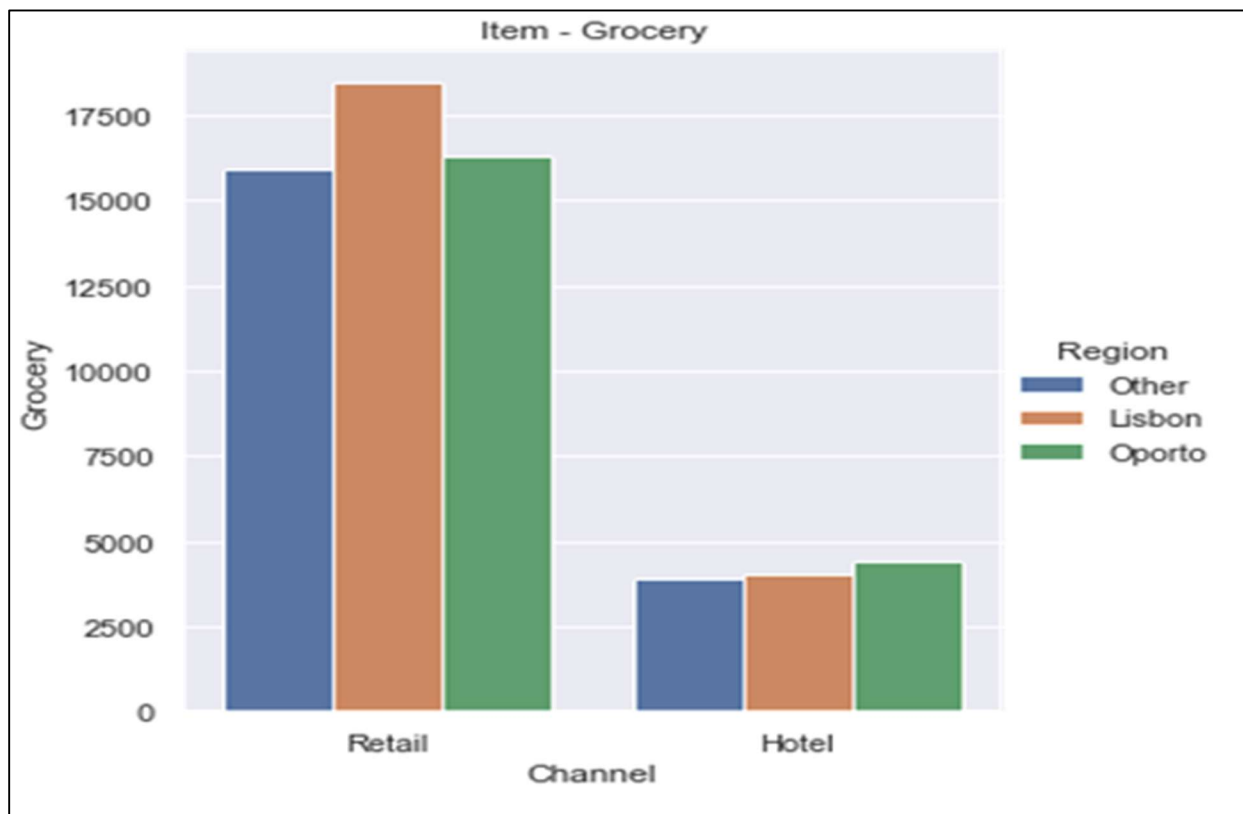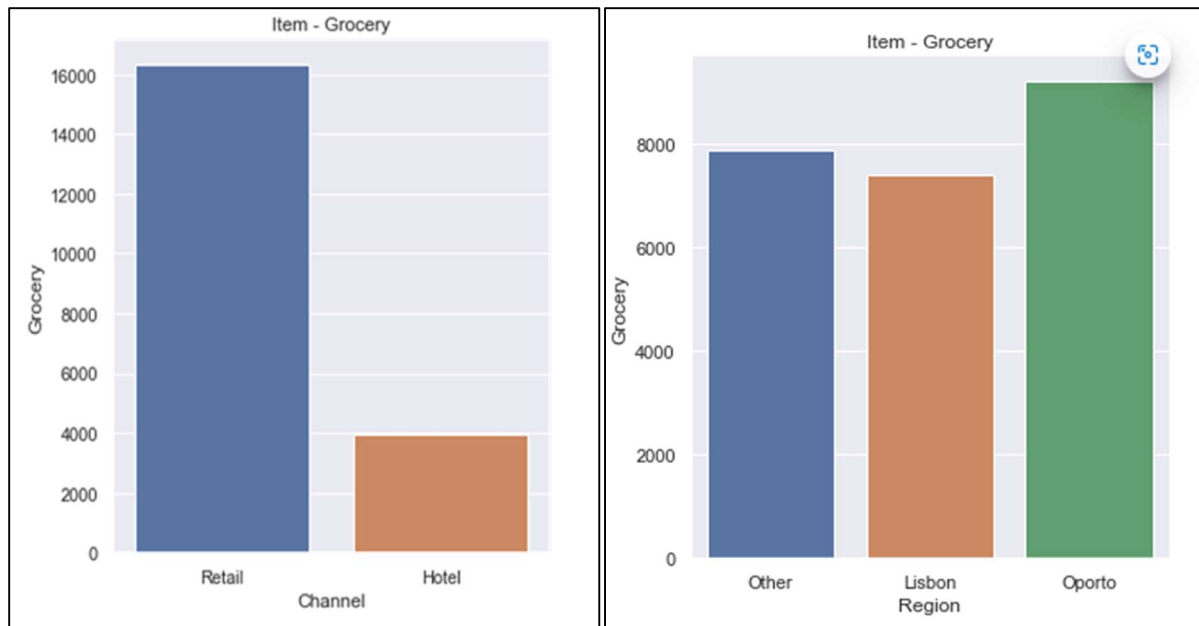
Item 2 – Milk

Analytics' Observation:

➢ Based on the plot, <u>Milk</u> item is sold more in <u>Retail channel</u>, on a regional level it has low numbers for <u>Oporto</u> as compared to Lisbon and other region.

Item 3 – Grocery

Analytics' Observation:

- ➢ Based on the plot, Grocery item is sold more in Retail channel, on regional level it has low numbers for Lisbon as compared to Oporto and Other region.

Item 4 - Frozen
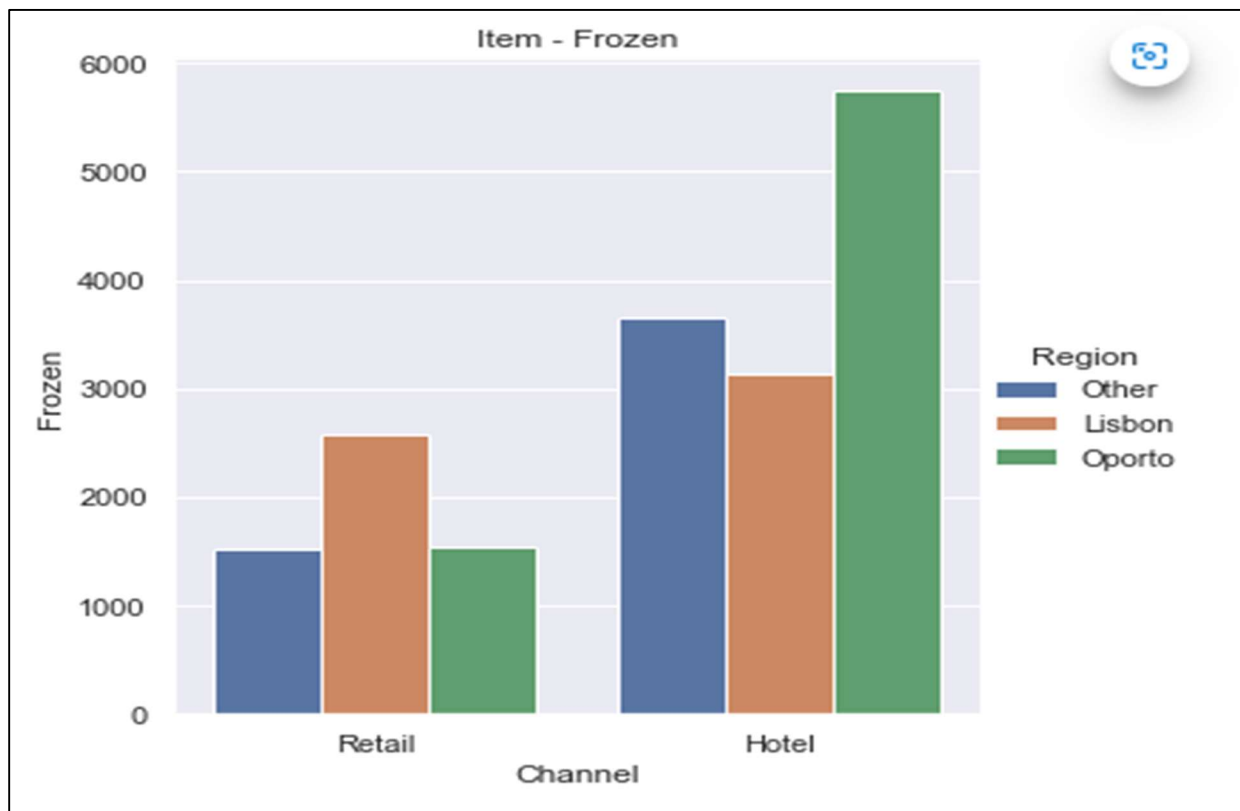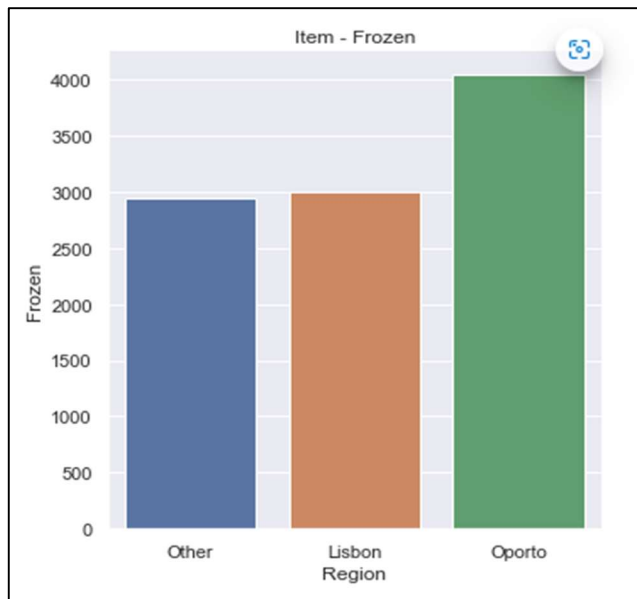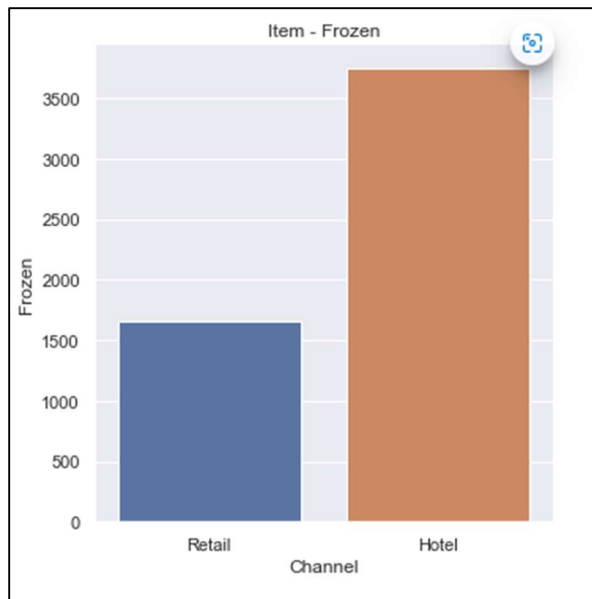
Analytics' Observation:

➢ Based on the plot, Frozen item is sold more in hotel channel, on regional level it has low numbers for Other region as compared to Oporto and Lisbon region.

Item 5 - Detergents_Paper

Analytics' Observation:

- ➢ Based on the plot, Detergents_Paper item is sold more in Retail channel, on regional level it has low numbers for Lisbon as compared to Oporto and other region.

Item 6 – Delicatessen
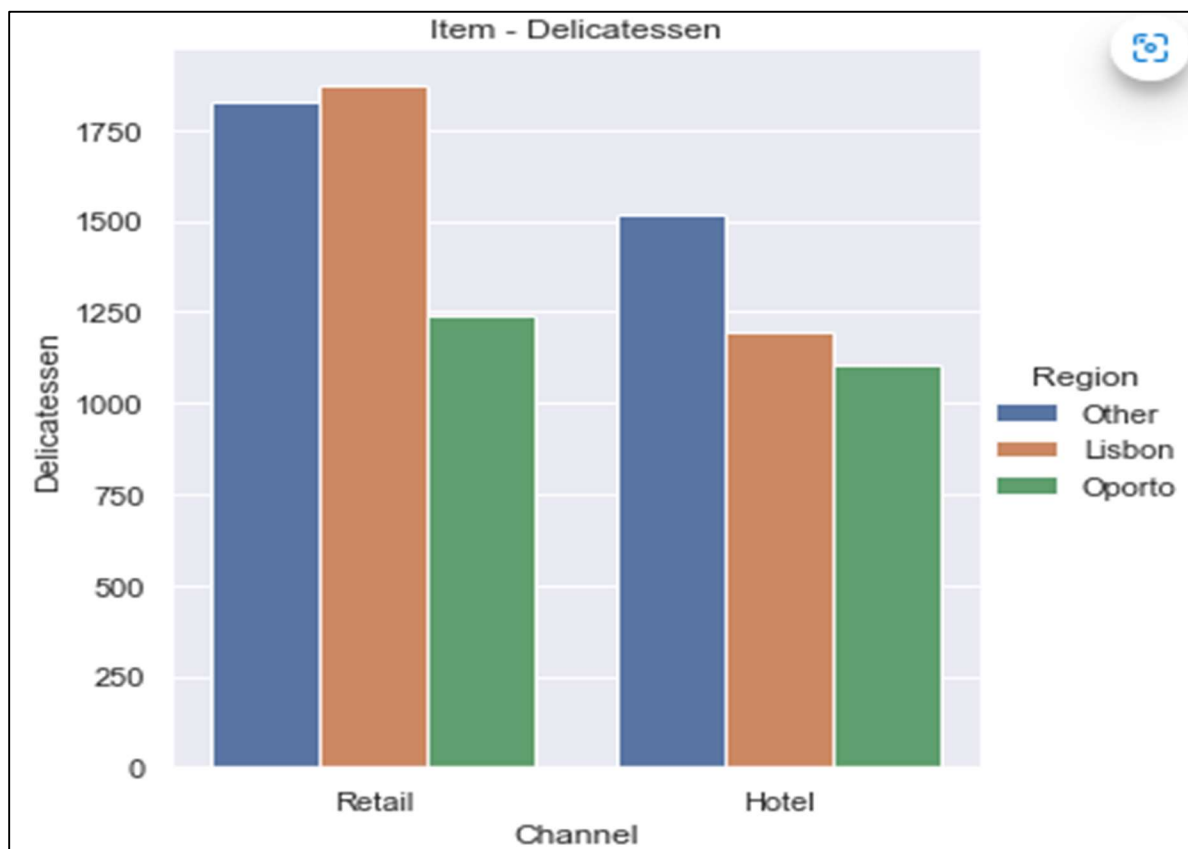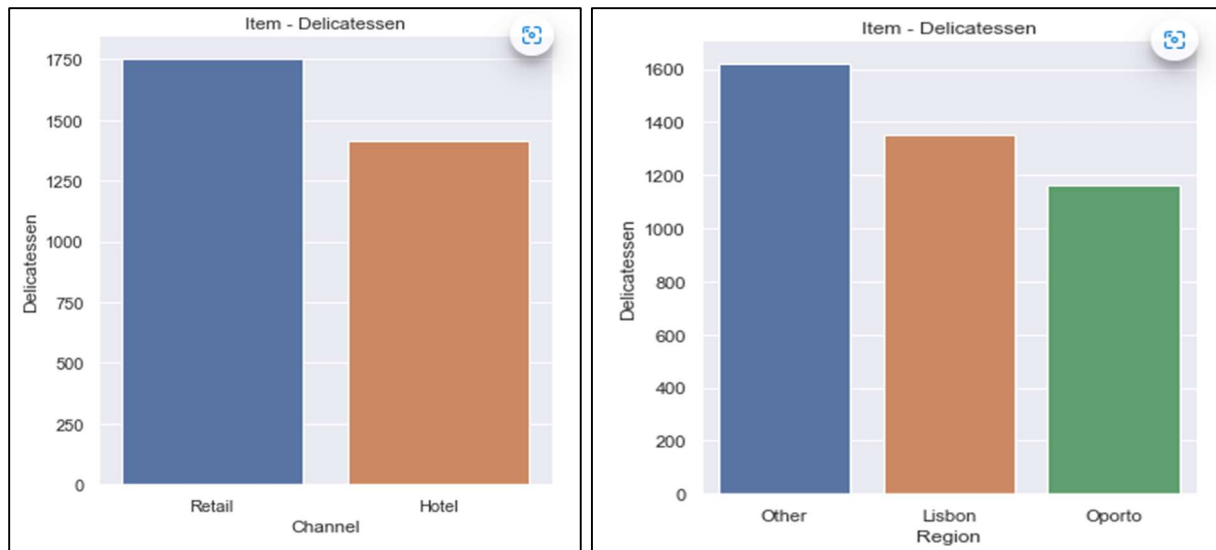
Analytics' Observation:

➤ Based on the plot, <u>Delicatessen</u> item is sold more in <u>Retail</u> channel, on a regional level it has low numbers for <u>Oporto</u> as compared to Lisbon and other region.

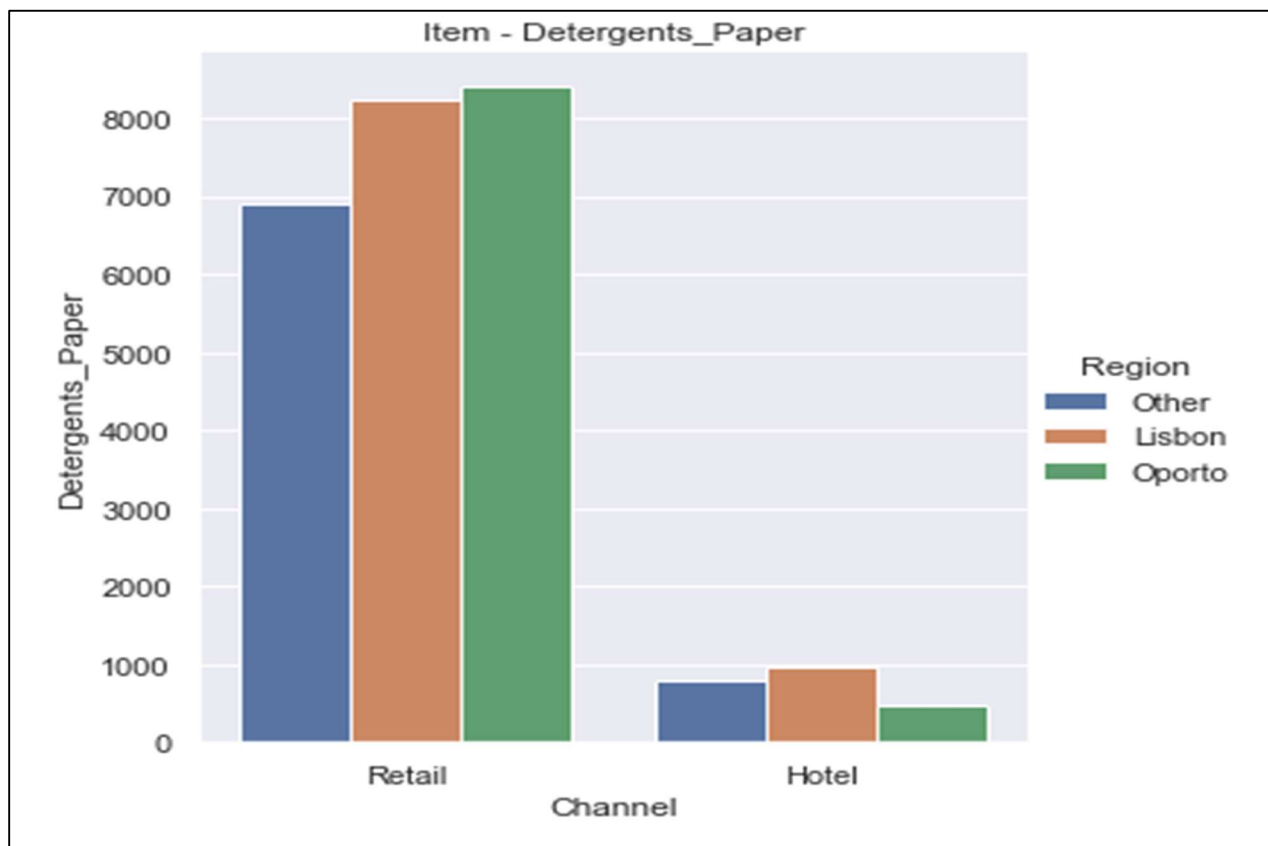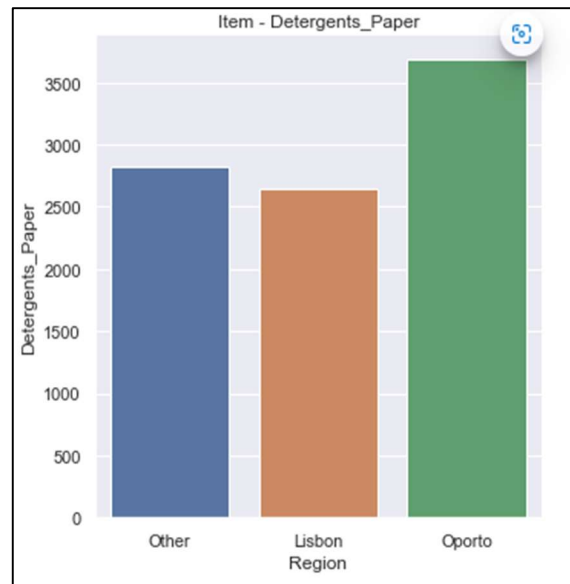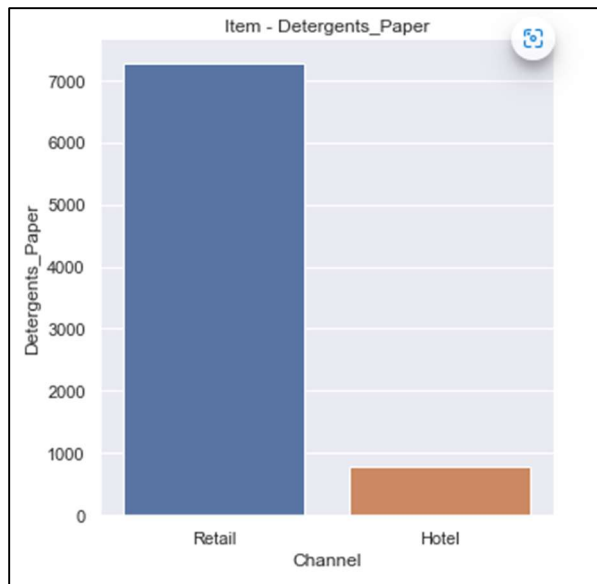➤ Based on above analysis, it can be concluded that considering all the 6 variety of items, all varieties do not show similar behavior across Region and Channel also based on Coefficient of variable there is inconsistency observed.

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

Let's review the following Coefficient of Variation result to find out the answer

Item	Mean	Std	Coefficient of Variation
Fresh	12000.29773	12647.32887	1.053917924
Milk	5796.265909	7380.377175	1.273298584
Grocery	7951.277273	9503.162829	1.195174373
Frozen	3071.931818	4854.673333	1.580332384
Detergents_Paper	2881.493182	4767.854448	1.654647138
Delicatessen	1524.870455	2820.105937	1.849406897

Analytics' Observation:

➤ Fresh item, have lowest coefficient of Variation therefore it's consistent. And Delicatessen" item have highest coefficient of Variation, so therefore it has inconsistent behavior the most.

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

Let's review the following Boxplot to deep drive the outlier's query
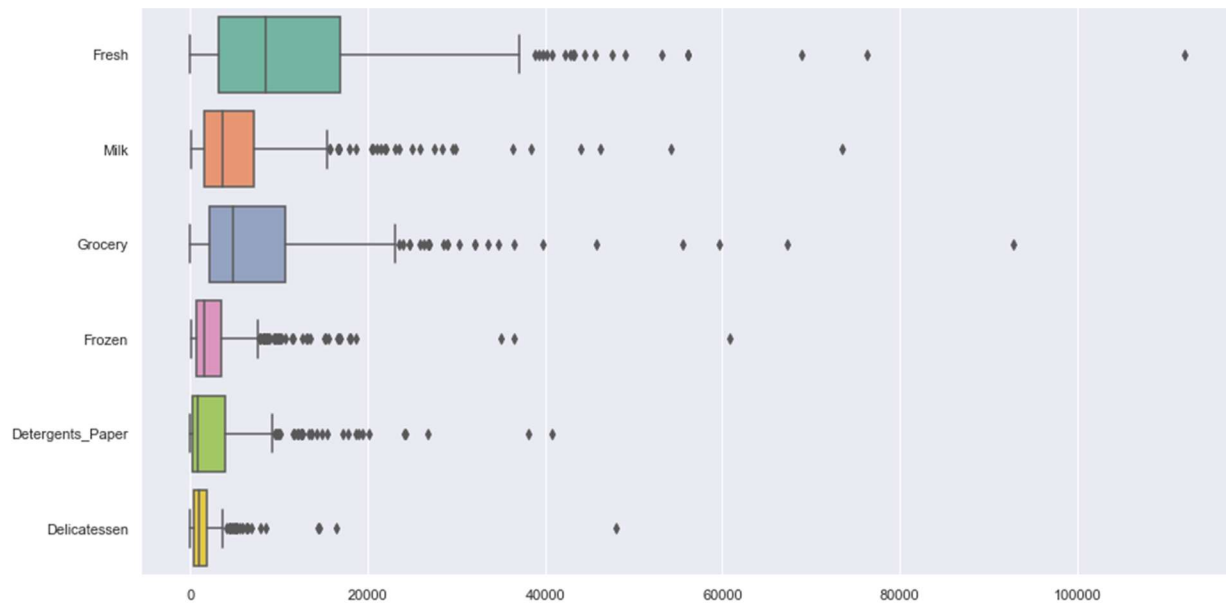
Analytics' Observation:

➢ Reference to above boxplot visuals, there is an Outliers in all the Variables across the product range. Outliers are detected but not necessarily removed, it depends on the situation

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.

Based on the analysis the following recommendations can be made

Analytics' Observation:

➢ Based on the analysis, the region **Other** and the channel **Hotel** have Higher spending than other Channel and Regions. Hence From the Business perspective if a new business is to be opened it Should be opened in the **other** region with Channel **Hotel** as the Other region is absorbing maximum numbers of sales and this can boast up the Revenue compared to opening a new business in Lisbon or Oporto and with the Channel Retail.

➢ In all the regions the Food Items Fresh has the highest spending followed by Grocery and Milk. Hence these food products are strongly recommended to be available simultaneously at all the businesses with priority of availability being Fresh food products.

➢ Also, the food item Delicatessen shows inconsistent behavior across all regions and channels which should be minimized. So, Delicatessen is also recommended to be always available in all the Businesses.

Problem 2

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).

Database Structure:

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	1	3	350	Laptop	200
2	Male	23	Senior	Management	Yes	3.6	Part-Time	25	1	4	360	Laptop	50
3	Male	21	Junior	Other	Yes	2.5	Part-Time	45	2	4	600	Laptop	200
4	Male	21	Junior	CIS	Yes	2.5	Full-Time	40	4	6	600	Laptop	250
5	Male	23	Senior	Other	Undecided	2.8	Unemployed	40	2	4	500	Laptop	100
...	...	...	...	...	...	...	...	...	...	...	...	...	...
58	Female	21	Senior	International Business	No	2.4	Part-Time	40	1	3	1000	Laptop	10
59	Female	20	Junior	CIS	No	2.9	Part-Time	40	2	4	350	Laptop	250
60	Female	20	Sophomore	CIS	No	2.5	Part-Time	55	1	4	500	Laptop	500
61	Female	23	Senior	Accounting	Yes	3.5	Part-Time	30	2	3	490	Laptop	50
62	Female	23	Senior	Economics/Finance	No	3.2	Part-Time	70	2	3	250	Laptop	0

Data Information:

#	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
int64 dtypes: float64(2), int64(6), object(6) memory usage: 6.9+ KB			

There is No Null value in data

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

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	Total
Gender									
Female	3	3	7	4	4	3	9	0	33
Male	4	1	4	2	6	4	5	3	29
Total	7	4	11	6	10	7	14	3	62

Analytics' Observation:

➢ The Majority of Female students is higher than a male student.

2.1.2. Gender and Grad Intention

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

Analytics' Observation:

➢ There are 28 students across all the genders has the intention for Grad. Which comes around 45.16 %

2.1.3. Gender and Employment

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

Analytics' Observation:

➢ There are 9 students who at present comes under the Unemployed category. Which comes around 14.51 %.

2.1.4. Gender and Computer

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

Analytics' Observation:

➢ Most of the students used Laptop device. Which comes around 88.70 %.

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?

Gender	Student's Count	Probability
Female	33	0.532258064516129
Male	29	0.46774193548387094
Total	62	100

Total Number of Students are 62, Total numbers of Male out of 62 is 29

X=62, Y=29

The probability that a randomly selected CMSU student will be male Is, Y/X = 0.46774193548387094

Analytics' Observation:

➢ The probability that a randomly selected CMSU student will be male is **0.46774193548387094**, in % it's 46.77%

2.2.2 <u>What is the probability that a randomly selected CMSU student will be female?</u>

Gender	Student's Count	Probability
Female	33	0.532258064516129
Male	29	0.46774193548387094
Total	62	100

Total Number of Students are 62, Total numbers of Female out of 62 is 33

X=62, Z=33

The probability that a randomly selected CMSU student will be Female Is, Z/X = 0.532258064516129

Analytics' Observation:

➢ The probability that a randomly selected CMSU student will be Female is **0.532258064516129** , in % it's 53.22%

2.3. <u>Assume that the sample is a representative of the population of CMSU. Based on the data, answer the following question:</u>

2.3.1 <u>Find the conditional probability of different majors among the male students in CMSU.</u>

Major	Accounting	CIS	Economics/ Finance	International Business	Management	Other	Retailing/Marketi ng	Undecided	Total
Gender									
Male	4	1	4	2	6	4	5	3	29
Probability	0.137931034	0.034482759	0.13793103	0.068965517	0.206896552	0.13793103	0.172413793	0.103448276	

Analytics' Observation:

Hence from the calculations done in Python we conclude that:

➢ Probability of Accounting among the male student **0.13793103448275862**
➢ Probability of CIS among the male students **0.034482758620689655**
➢ Probability of Economics/Finance among the male students **0.13793103448275862**
➢ Probability of International Business among the male students **0.06896551724137931**
➢ Probability of Management among the male students Management **0.20689655172413793**
➢ Probability of Other among the male students Other **0.13793103448275862**
➢ Probability of Retailing/Marketing among the male students **0.1724137931034483**
➢ Probability of Undecided among the male students **0.10344827586206896**

2.3.2 <u>Find the conditional probability of different majors among the female students of CMSU</u>

Major	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided	Total
Gender									
Female	3	3	7	4	4	3	9	0	33
Probability	0.090909091	0.090909091	0.212121212	0.121212121	0.121212121	0.09090909	0.272727273	0	

Analytics' Observation:

Hence from the calculations done in Python we conclude that:

➢ Probability of Accounting among the female students **0.09090909090909091**
➢ Probability of CIS among the female students **0.09090909090909091**
➢ Probability of Economics/Finance among the female students **0.21212121212121213**
➢ Probability of International Business among the female students **0.12121212121212122**
➢ Probability of Management among the female students Management **0.12121212121212122**
➢ Probability of Other among the female students Other **0.09090909090909091**
➢ Probability of Retailing/Marketing among the female students **0.2727272727272727**
➢ Probability of Undecided among the female students **0.0**

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.

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

Probability that a randomly chosen student is a Male = 29/62

Probability of Male that intends to Graduate = 17/29

X=17, Y=62

X/Y= **0.27419354838709675**

Analytics' Observation:

➢ The probability That a randomly chosen student is a male and intends to graduate is **0.27419354838709675**

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

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

Probability that a randomly chosen student is a Female = 33/62

Probability of Female with No Laptop = 4/62

x1=62, y1=4

y1/x1= 0.06451612903225806

Analytics' Observation:

> The probability that a randomly selected student is a female and does NOT have a laptop is **0.06 451612903225806**

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 a male or has a full-time employment

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

Probability of a Student being Male = 29/62

Probability of a student having Fulltime Employment = 10/62

Probability of a Male having Fulltime Employment = 7/62

a1 =29/62

b1=10/62

c1=7/62

a1+b1-c1 = 0.5161290322580645

Analytics' Observation:

> The probability that a randomly chosen student is either a male or has a full-time employment **0.5161290 322580645**

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

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

Probability of international business given Female = 4/33

Probability of management given Female = 4/33

d1 = 4/33

e1 = 4/33

d1/e1=0.24242424242424243

Analytics' Observation:

➢ The conditional probability that given a female student is randomly chosen, she is majoring in international business or management is **0.2424242424242424**

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?

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

P(Grad Intention Yes) = 28/40

P(Grad Intention Yes female) = 11/40

Probability that a randomly selected candidate is female, 28/40 = 0.7
Probability that a female candidate has Grad intention, 11/40 = 0.275

Analytics' Observation:

➢ From the calculations done in Python we conclude that: These probabilities are not equal. This suggests that the **Two events are 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?

GPA	<3	>3	Total
	17	45	62

➢ Total number students have GPA is 62, and out of 62 there are 17 Students which has GPA less than a 3.

The probability of his/her GPA is less than 3 is', 17/62 = **0.27419354838709675**

Analytics' Observation:

➢ The probability of his/her GPA is less than 3 is **0.27419354838709675,** % wise it will be 27.41%.

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.

Please refer following strip plot to visualize the Salary earning gender wise.
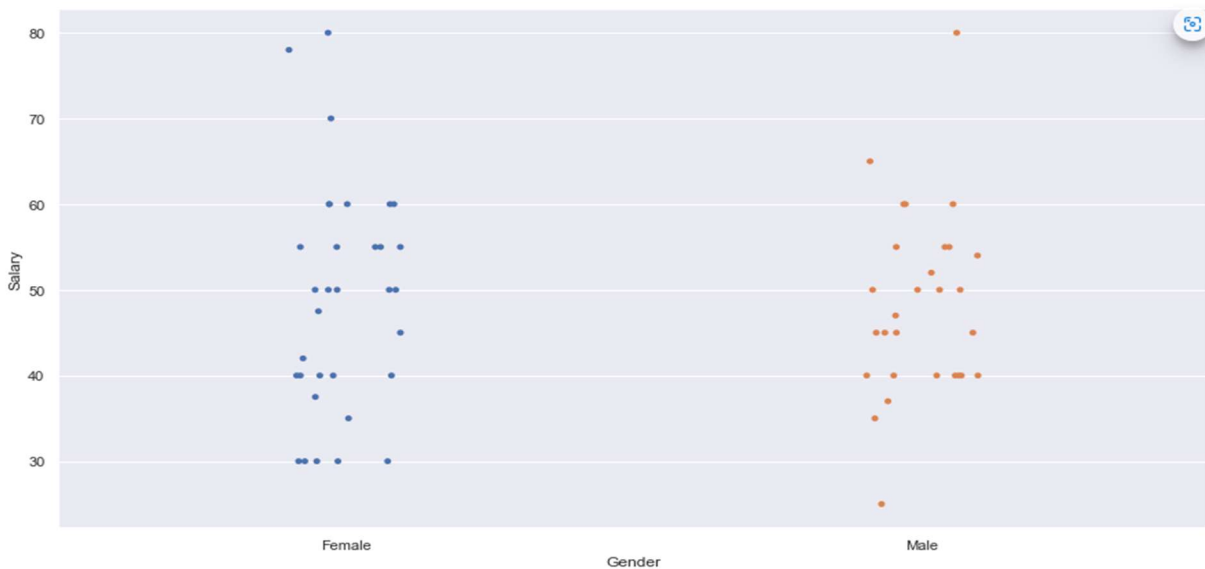
| Gender	Salary			Probability of 50 or more
	<50	>50	Total	
Female	15	18	33	0.545454545
Male	15	14	29	0.482758621
Total	30	32	62	

➢ Total number Male is 29, and out of 29 there are 14 males who earns 50 or more

   Probability that a randomly selected male earns 50 or more is',14/29 = **0.4827586206896552**

➢ Total number Female is 33, and out of 33 there are 18 females who earns 50 or more

   Probability that a randomly selected Female earns 50 or more is',18/33 = **0.5454545454545454**

Analytics' Observation:

➢ The Conditional probability that a randomly selected **Male** earns 50 or more is
   0.4827586206896552
➢ The Conditional probability that a randomly selected **Female** earns 50 or more is
   0.5454545454545454

2.8.1 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.

Let's review the following Box plot to conclude the Normal Distribution
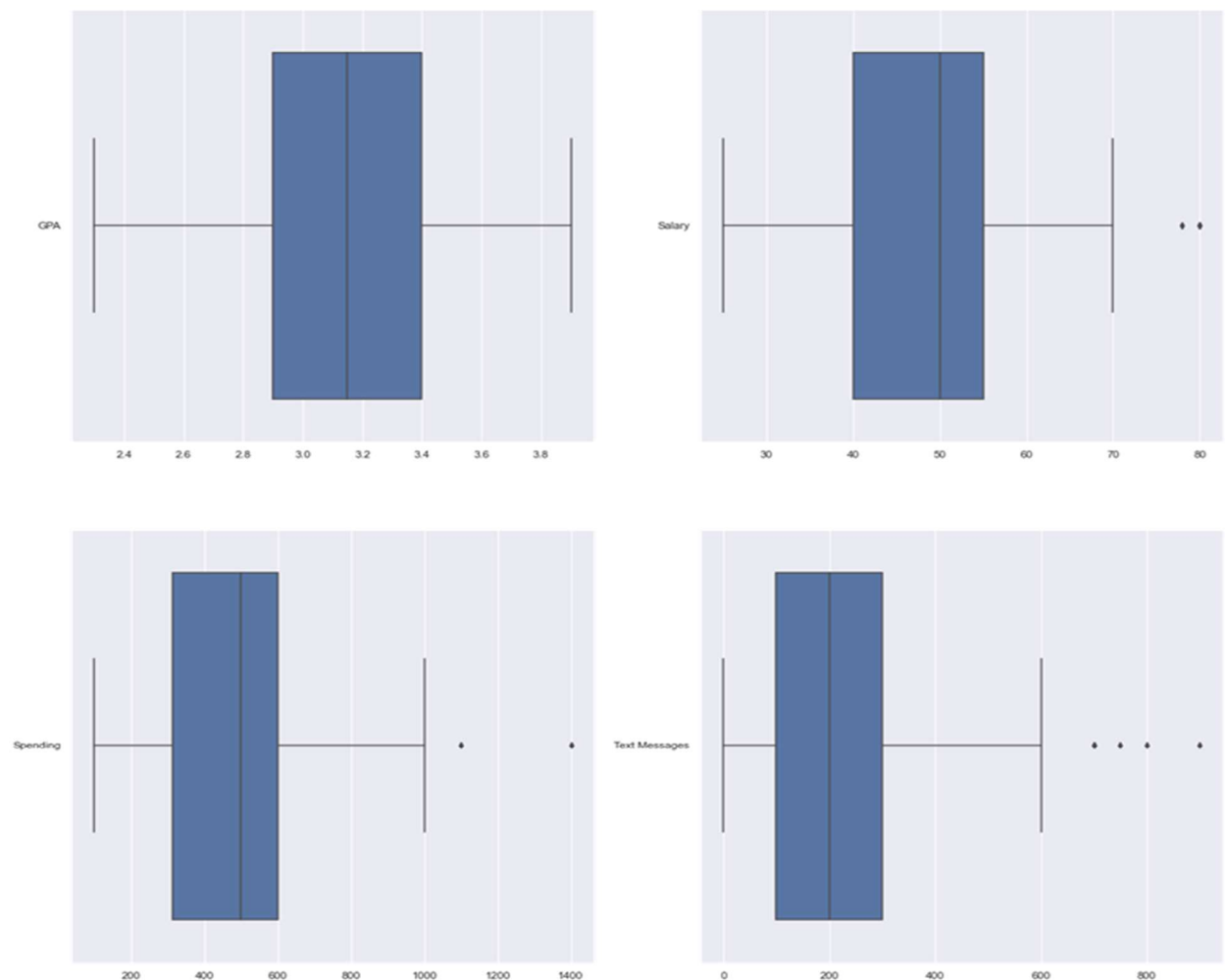
Analytics' Observation:

From the above boxplot for the continuous variables GPA, Salary, Spending and Text Messages we can see that:

➢ GPA is almost Normally Distributed with a slight skewness toward the left.
➢ Salary is also Normally Distributed with a slight skewness towards the right.
➢ Spending is not Normally distributed and highly Right Skewed
➢ Text message is not Normally distributed and highly Right Skewed
➢ By these details we confirm that out of the given four data sets 'GPA' and 'Salary' are following normal distribution whereas other two 'Spending' and 'Text Messages' are not following the normal distribution

2.8.2 Write a note summarizing your conclusions

Analytics' Observation:

➢ The Majority of Female students are higher than a male student, there is little improvement is needed in terms of diversity
➢ The graduate intention is 45.16 %, which is very low, we need to motivate the Undecided students which are approx. 22 to become intent to graduate. This will improve the % of graduate intention drastically.
➢ Need to focus on Unemployed students, which comes around 14.51 %. to facilitate employment based on skills.
➢ Spending & Text message is not Normally distributed, therefore requires more focus to improve the numbers.

Problem 3

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.

Note: Before reviewing the data, Query and solutions let's understand the basic Data Structure first.

Database Structure:

	A	B
0	0.44	0.14
1	0.61	0.15
2	0.47	0.31
3	0.30	0.16
4	0.15	0.37
5	0.24	0.18
6	0.16	0.42
7	0.20	0.58
8	0.20	0.25
9	0.20	0.41
10	0.26	0.17
11	0.14	0.13
12	0.33	0.23
13	0.13	0.11
14	0.72	0.10
15	0.51	0.19
16	0.28	0.22
17	0.39	0.44
18	0.39	0.11

Data information:

```
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   A           36 non-null     float64
 1   B           31 non-null     float64
dtypes: float64(2)
```

	count	mean	std	min	25%	50%	75%	max
A	36.0	0.316667	0.135731	0.13	0.2075	0.29	0.3925	0.72
B	31.0	0.273548	0.137296	0.10	0.1600	0.23	0.4000	0.58

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

A Shingle conditions

HO=> mu<=0.35

Ha=> mu>=0.35

Step 1: Since a value is not given, therefor we will be considering 0.05 as a value considering global standard

Step 2: Based on condition decided to have a one sample t-test.

Step 3: Post t-test with the help of python script, we got the following result.

  t statistic: 1.4735046253382782 , and p value : 0.07477633144907513

Test summary: The Level of significance is 0.05 and the **p value is 0.074776** and it's in higher side of level of significance. So, the based on the statistics the test is failing to reject the null hypothesis at 5% level of significance.

Analytics' Observation:

➢ Hence, at 95% confidence level, and based on sufficient evidence we can conclude that the content in A Shingles is more than 0.35 pound per 100 square feet.

B Shingle conditions

HO=> mu<=0.35

Ha=> mu>=0.35

Step 1: Since a value is not given, therefor we will be considering 0.05 as a value considering global standard and the Sample size (n) for this case would be 31,

Step 2: Based on condition decided to have a one sample t-test.

Step 3: Post t-test with the help of python script, we got the following result.

t statistic: -3.1003313069986995 p value: 0.0020904774003191826

Test summary: The Level of significance is 0.05 and the **p value: 0.00209** and it's less than 5% of Level of significance. So, the based on the statistics the test is rejected the null hypothesis at 5% level of significance.

Analytics' Observation:

➤ Hence, at 95% confidence level, based on sufficient evidence we can conclude that the content in B Shingles is less than 0.35 pound per 100 square feet.

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?

H0 : $\mu(A) = \mu(B)$

Ha : $\mu(A) != \mu(B)$

$\alpha = 0.05$

Step 1: since a value is not given, therefor we will be considering 0.05 as a value considering global standard

Step 2: Based on condition decided to have an in-depended Sample ttest

Step 3: Post t-test with the help of python script, we got the following p Value.

p value: 0.2017496571835328

Test Summary: With the help of in-depended Sample ttest, we got p **value: 0.2017496571835328** and it's in higher side of level of significance. So, the based on the statistics the test is failing to reject the null hypothesis at 5% level of significance.

Analytics' Observation:

➤ Hence, at 95% confidence level, and based on sufficient evidence we can conclude that the content in A Shingles is equal to mean moisture content in B

Thank you 😊