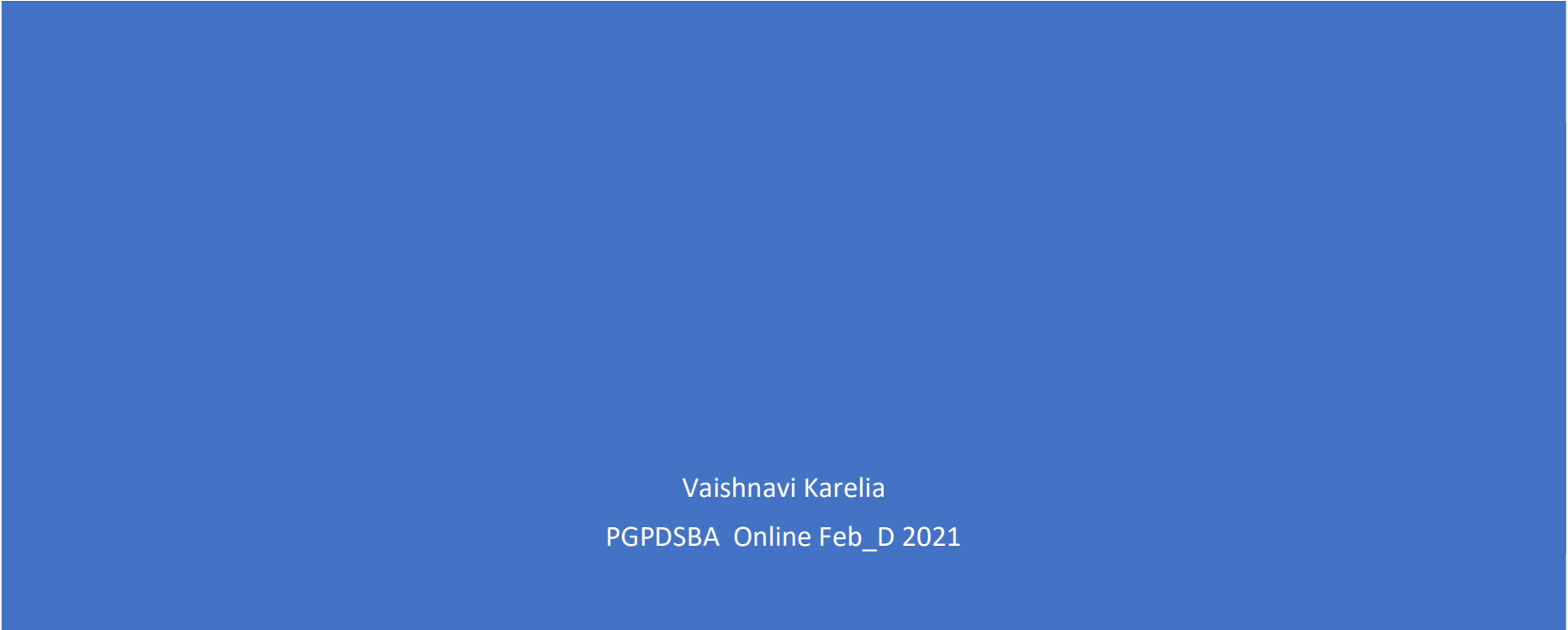# SMDM PROJECT

Vaishnavi Karelia

PGPDSBA Online Feb_D 2021

PROBLEM STATEMENT 1 - WHOLESALE CUSTOMERS ANALYSIS

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?

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 A DESCRIPTIVE MEASURE OF VARIABILITY, WHICH ITEM SHOWS THE MOST INCONSISTENT BEHAVIOUR? WHICH ITEMS SHOW 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 STATEMENT 2 - THE STUDENT NEWS SERVICE

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

2.2. 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?

2.2.2. WHAT IS THE PROBABILITY THAT A RANDOMLY SELECTED CMSU STUDENT WILL BE FEMALE?

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.

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 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?

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 the 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 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.

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 whole Problem 2.

PROBLEM STATEMENT 3 – SHINGLES MANUFACTURERS

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 STATEMENT 1 - WHOLESALE CUSTOMERS ANALYSIS

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).

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?

Region	Lisbon	Oporto	Other
Total	2386813	1555088	10677599

Total spendings as per Region

Channel	Hotel	Retail
Total	7999569	6619931

- Other (10,677,599) region spent most.

- Lisbon (2,386,813) comes second close and Oporto (1,555,088) region spent the least.

- Hotel spent the most (7,999,569) and Retail spent the least (6,619,931)

Total spendings as per Channel
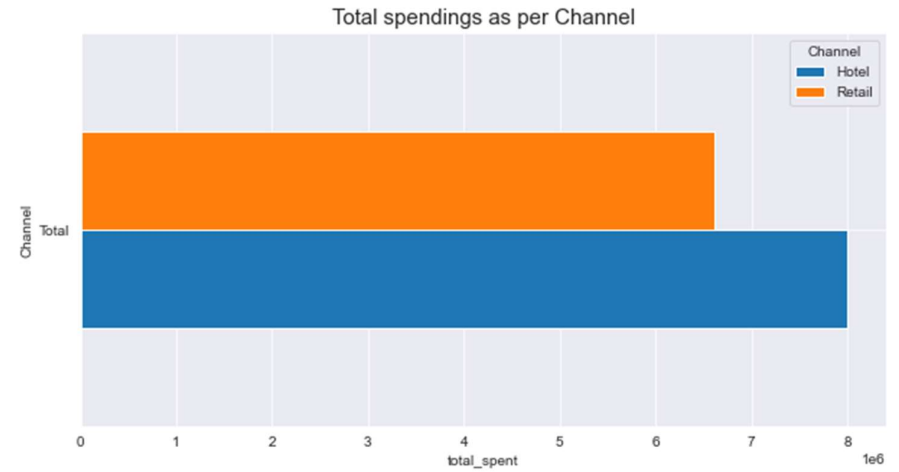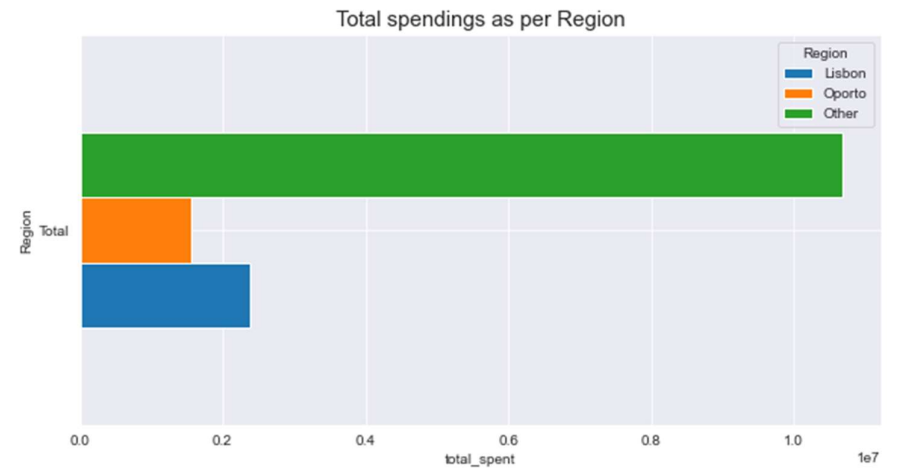
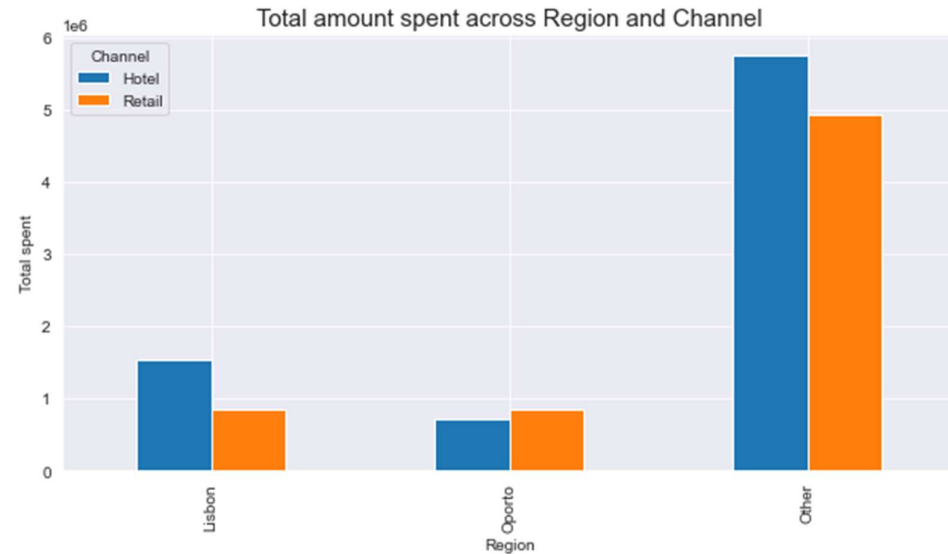| Region	Lisbon	Oporto	Other
Channel			
Hotel	59	28	211
Retail	18	19	105

Figure 1.1.c Total spending across region and channel

With respect to the frequency of the transaction across all the regions and channels, can infer the following:

Channel has two unique values, with Hotel as most frequent with 298 out of 440 transactions i.e., 67.7% of spending and Retail as least frequent with 142 out of 440 transactions i.e., 32.27%

Region has three unique values, with Other as most frequent with 316 out of 440 transactions i.e.,71.8% of spending and Lisbon ranking 2nd with 77 out of 440 transactions i.e., 17.5% of spending. Least frequent is Oporto with 47 out of 440 transactions i.e.,10.68%

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.

- Fresh item has high covariance in Hotel channel for Lisbon and Other regions.

- Grocery has the highest covariance in Oporto region along with Frozen items.

- Delicatessen in both hotel and retail channel has low covariance across all regions.

- Detergents_Paper in hotel channel has low variance for all regions.

Region		Lisbon	Oporto	Other
Channel				
Delicatessen	Hotel	1.488267e+06	1.116781e+06	1.341891e+07
	Retail	2.645459e+06	1.135158e+06	4.490382e+06
Detergents_Paper	Hotel	1.705395e+06	1.808890e+05	1.209935e+06
	Retail	3.042492e+07	6.867020e+07	3.626558e+07
Fresh	Hotel	1.523252e+08	8.044947e+07	2.174614e+08
	Retail	2.932787e+07	4.716852e+07	9.284082e+07
Frozen	Hotel	1.073519e+07	1.312051e+08	2.456779e+07
	Retail	5.879532e+06	6.117047e+06	2.263516e+06
Grocery	Hotel	1.317432e+07	9.292126e+06	1.291329e+07
	Retail	1.084657e+08	1.969940e+08	1.512638e+08
Milk	Hotel	1.847557e+07	8.812756e+06	2.032662e+07
	Retail	4.368181e+07	4.371000e+07	1.118270e+08

	Count	Mean	Standard Deviation	min	Q1 (25%)	Q2 (50%)	Q3 (75%)	max	Variance	IQR	Range
Fresh	440.0	12000.297727	12647.328865	3.0	3127.75	8504.0	16933.75	112151.0	1.599549e+08	13806.00	112148.0
Milk	440.0	5796.265909	7380.377175	55.0	1533.00	3627.0	7190.25	73498.0	5.446997e+07	5657.25	73443.0
Grocery	440.0	7951.277273	9503.162829	3.0	2153.00	4755.5	10655.75	92780.0	9.031010e+07	8502.75	92777.0
Frozen	440.0	3071.931818	4854.673333	25.0	742.25	1526.0	3554.25	60869.0	2.356785e+07	2812.00	60844.0
Detergents_Paper	440.0	2881.493182	4767.854448	3.0	256.75	816.5	3922.00	40827.0	2.273244e+07	3665.25	40824.0
Delicatessen	440.0	1524.870455	2820.105937	3.0	408.25	965.5	1820.25	47943.0	7.952997e+06	1412.00	47940.0
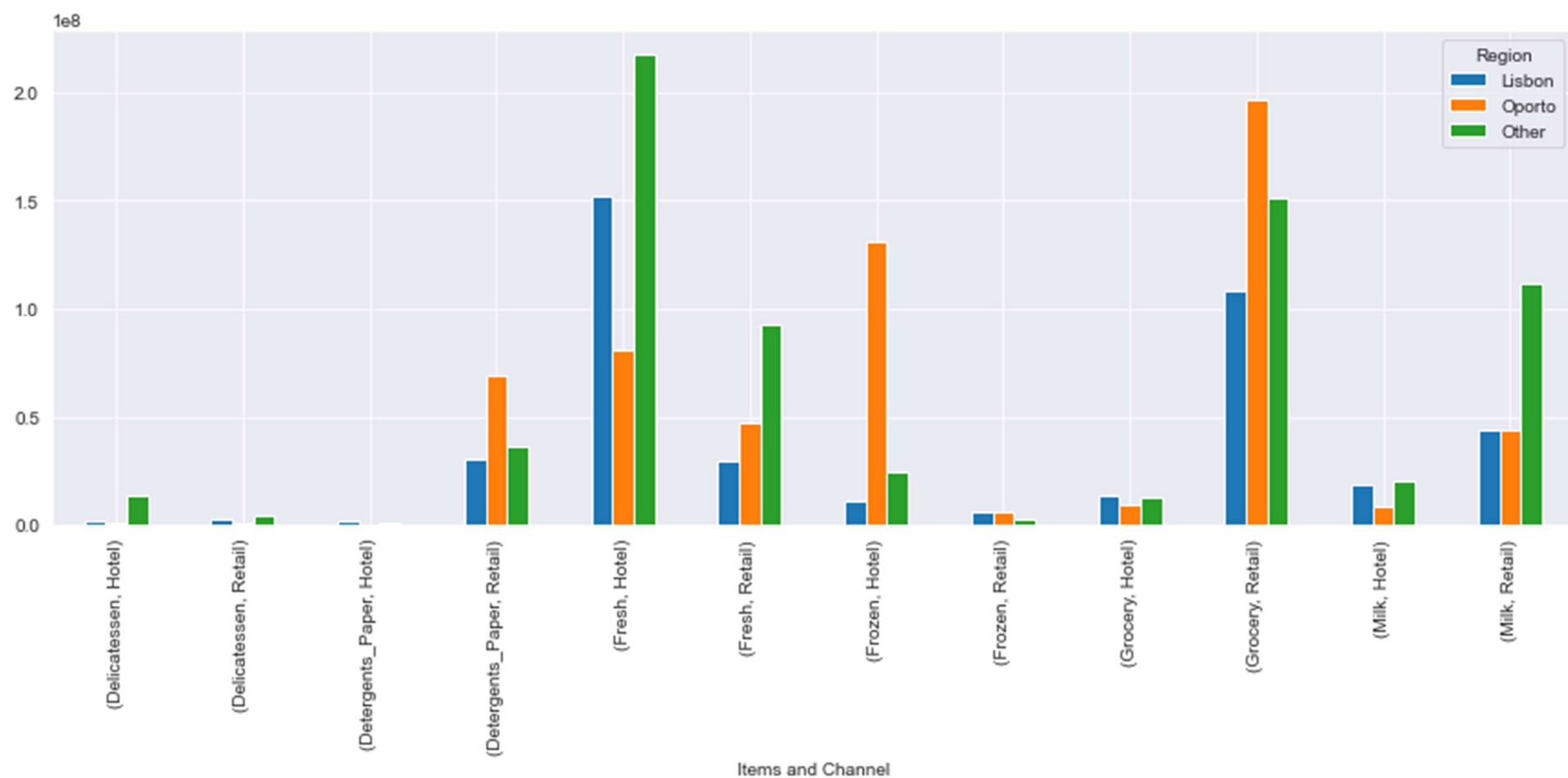
Description of the data

Fresh item (440 records):

With the mean of 12000.297727, standard deviation of 12647.328865, range is (max-min) 112148.0 (112151.0 - 3.0).

The Q1(25%) of the data is up to 3127.75, Q3(75%) is 16933.8, with Q2(50%) 8504

IQR = Q3-Q1 = 16933.8-3127.75 = 13,806.05

Milk item (440 records):

Mean of 5796.27, standard deviation of 7380.38, range is (max-min) 73443.0.

The Q1(25%) of the data is up to 1533, Q3(75%) is 7190.25, with Q2(50%) 3627

IQR = Q3-Q1 = 7190.25-1533 = 5657.25

Grocery item (440 records)

Mean of 7951.28, standard deviation of 9503.16, range = max-min =92780-3=92777

The Q1(25%) is 2153, Q3(75%) is 10655.8, with Q2(50%) 4755.5

IQR = Q3-Q1 = 10655.8-2153 = 8502.8

Frozen (440 records):

Mean of 3071.93, standard deviation of 4854.67, range = max-min =60869-25=60844

Q1(25%) is 742.25, Q3(75%) is 3554.25, with Q2(50%) 1526

IQR = Q3-Q1 = 3554.25-742.25 = 2812

Detergents_Paper (440 records):

Mean of 2881.49, standard deviation of 4767.85, range = max-min =40827-3=40824

Q1(25%) is 256.75, Q3(75%) is 3922, with Q2(50%) 816.5

IQR = Q3-Q1 = 3922-256.75 = 3665.25

Delicatessen (440 records):

Mean of 1524.87, standard deviation of 2820.11, range = max-min =47943-3=47940

Q1(25%) is 408.25, Q3(75%) is 1820.25, with Q2(50%) 965.5

IQR = Q3-Q1 = 1820.25-408.25 = 1412

1.3 ON THE BASIS OF A DESCRIPTIVE MEASURE OF VARIABILITY, WHICH ITEM SHOWS THE MOST INCONSISTENT BEHAVIOUR? WHICH ITEMS SHOW THE LEAST INCONSISTENT BEHAVIOUR?

```
Item's behaviour based on standard deviation:

Fresh 12647.3289
Milk 7380.3772
Grocery 9503.1628
Frozen 4854.6733
Detergents_Paper 4767.8544
Delicatessen 2820.1059

Item's behaviour based on coefficient of variation:

Fresh 1.0527
Milk 1.2719
```

Grocery 1.1938
Frozen 1.5785
Detergents_Paper 1.6528
Delicatessen 1.8473

Based on Standard Deviation: `Delicatessen` has the lowest standard deviation of 2820.1059 and thus is the consistent.

`Frozen` has the highest standard deviation of 12647.3289 and thus is the most inconsistent.

Based on coefficient of variation:

`Fresh` has lowest variance of 1.0527 | It has the least inconsistent behaviour means less varying.

`Delicatessen` has the highest variance of 1.8473 | It has the most inconsistent behaviour which means most varying.

1.4 ARE THERE ANY OUTLIERS IN THE DATA? BACK UP YOUR ANSWER WITH A SUITABLE PLOT/TECHNIQUE WITH THE HELP OF DETAILED COMMENTS.
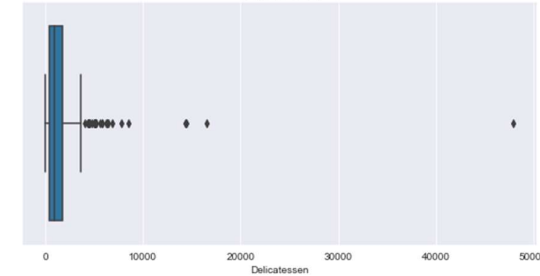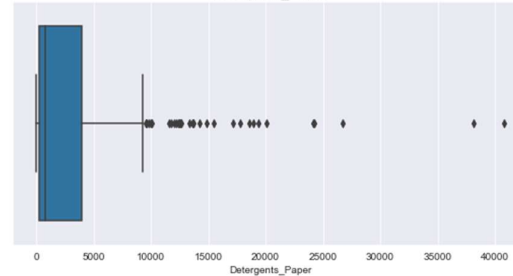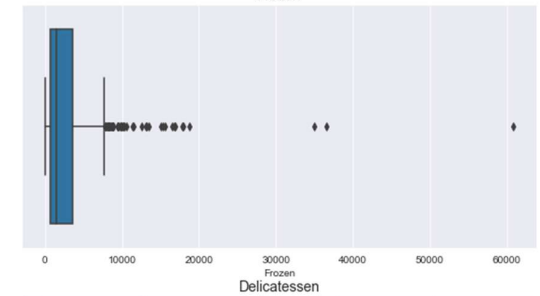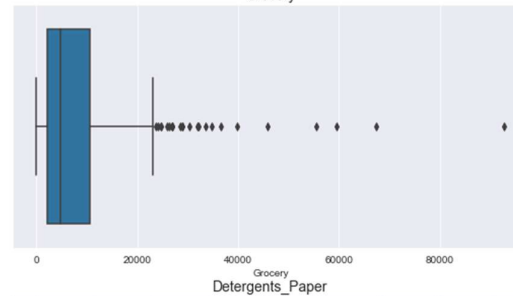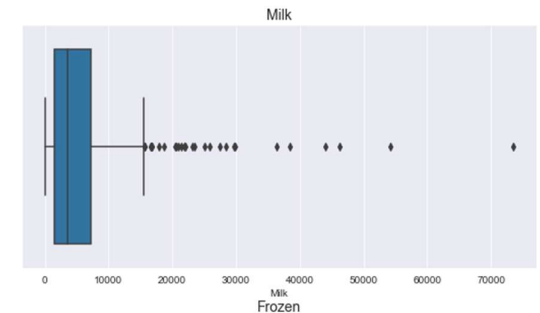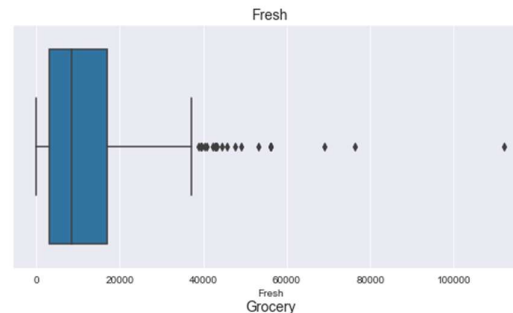
The following columns has outliers:

```
'Fresh', 'Milk', 'Grocery', 'Frozen', 'Detergents_Paper', 'Delicatessen'
```

IQR of the items

Item	Value
Fresh	13806.00
Milk	5657.25
Grocery	8502.75
Frozen	2812.00
Detergents_Paper	3665.25
Delicatessen	1412.00

dtype: float64

	Columns	Outliers
2	Delicatessen	True
3	Detergents_Paper	True
4	Fresh	True
5	Frozen	True
6	Grocery	True
7	Milk	True

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.

Channel has two unique values, with "Hotel" as most frequent with 298 out of 440 transactions i.e., 67.7% of spending and Retail as least frequent with 142 out of 440 transactions i.e., 32.27%.

Region has three unique values, with "Other" as most frequent with 316 out of 440 transactions i.e.,71.8% of spending and Lisbon ranking 2nd with 77 out of 440 transactions i.e., 17.5% of spending. Least frequent is Oporto with 47 out of 440 transactions i.e.,10.68%

The distributer should focus more on Hotel channel in other regions provided most of its transactions are carried out from there.

For retail channel the focus should be more on Oporto region. Items such as fresh, grocery and frozen have the highest demand.

Since Fresh items has the least inconsistent behaviour the distributor should continue to supply across all regions and channels.

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).

2.1. FOR THIS DATA, CONSTRUCT THE FOLLOWING CONTINGENCY TABLES (KEEP GENDER AS ROW VARIABLE)
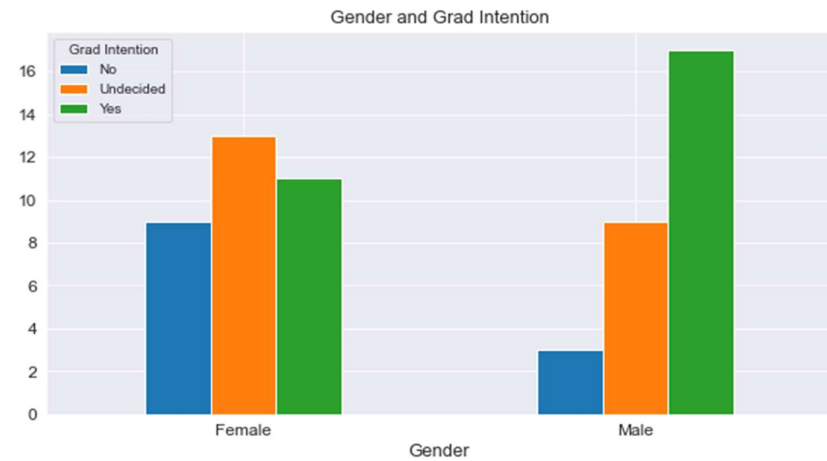
2.1.1. GENDER AND MAJOR

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

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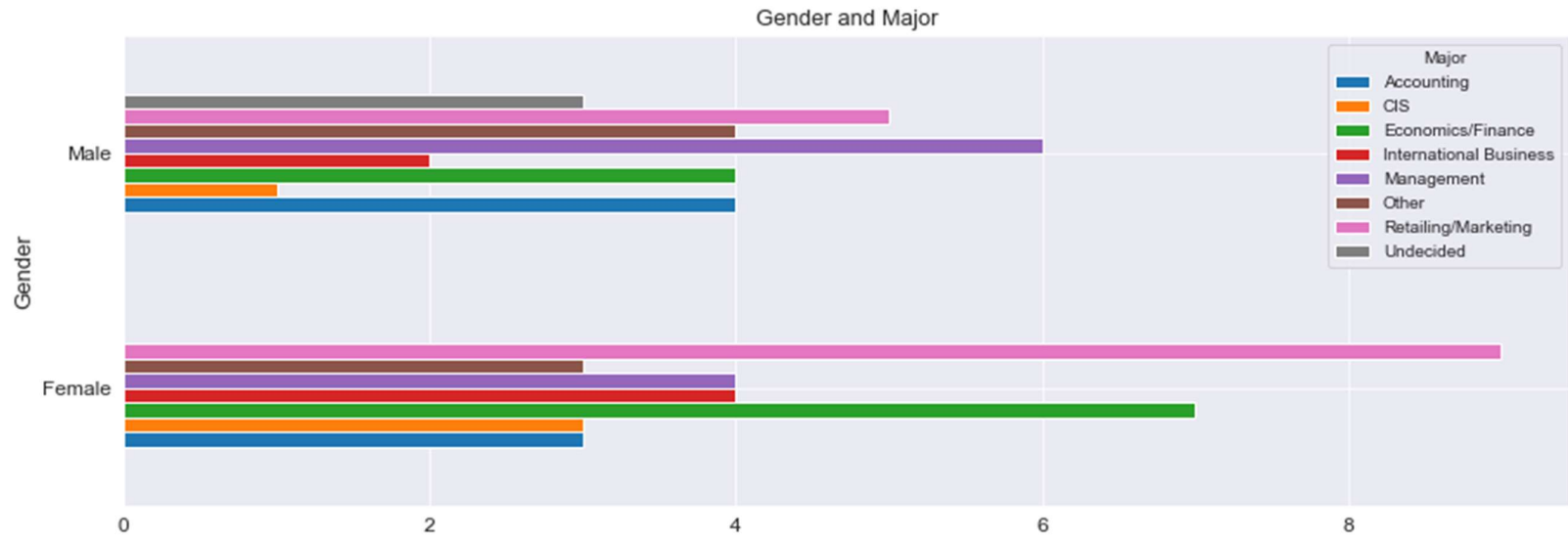
Grad Intention Gender	No	Undecided	Yes
Female	9	13	11
Male	3	9	17

Gender and Grad Intention

Employment	Full-Time	Part-Time	Unemployed
Gender			
Female	3	24	6
Male	7	19	3
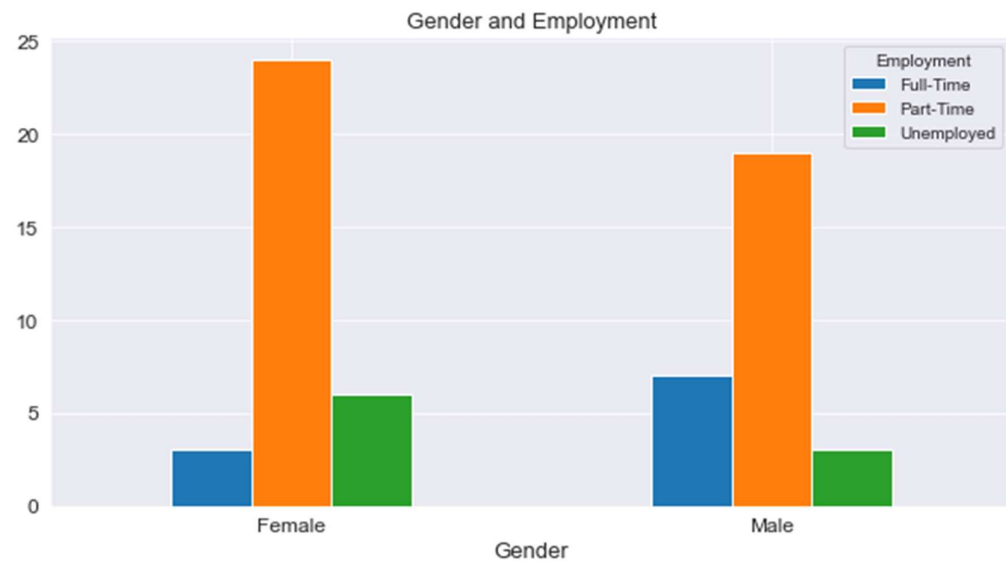
2.1.4. GENDER AND COMPUTER

Computer	Desktop	Laptop	Tablet
Gender			
Female	2	29	2
Male	3	26	0
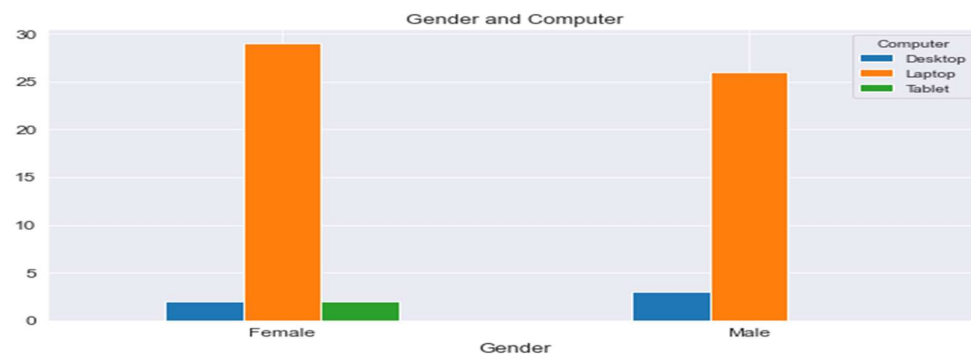
2.2. ASSUME THAT THE SAMPLE IS REPRESENTATIVE OF THE POPULATION OF CMSU. BASED ON THE DATA, ANSWER THE FOLLOWING QUESTION:

2.2.1. WHAT IS THE PROBABILITY THAT A RANDOMLY SELECTED CMSU STUDENT WILL BE MALE?

Total number of Males in CMSU = 29
Total number of students in CMSU = 62
Probability that a randomly selected CMSU student will be male is: 0.4677 or 46.77 %

2.2.2. WHAT IS THE PROBABILITY THAT A RANDOMLY SELECTED CMSU STUDENT WILL BE FEMALE?

Total number of Females in CMSU = 33
Total number of students in CMSU = 62
Probability that a randomly selected CMSU student will be female is: 0.5323 or 53.23 %

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.
Total number of Males in CMSU = 29

Total number of males in Accounting major: 4
 Probability of Accounting majors among the male students of CMSU: 0.1379 or 13.79 %

Total number of males in CIS major: 1
 Probability of CIS majors among the male students of CMSU: 0.0345 or 3.45 %

Total number of males in Economics/Finance major: 4
 Probability of Economics/Finance majors among the male students of CMSU: 0.1379 or 13.79 %

Total number of males in International Business major: 2
 Probability of International Business majors among the male students of CMSU: 0.069 or 6.9 %

Total number of males in Management major: 6
 Probability of Management majors among the male students of CMSU: 0.2069 or 20.69 %

Total number of males in Other major: 4
 Probability of Other majors among the male students of CMSU: 0.1379 or 13.79 %

Total number of males in Retailing/Marketing major: 5
 Probability of Retailing/Marketing majors among the male students of CMSU: 0.1724 or 17.24 %

Total number of males in Undecided major: 3
 Probability of Undecided majors among the male students of CMSU: 0.1034 or 10.34 %

2.3.2 FIND THE CONDITIONAL PROBABILITY OF DIFFERENT MAJORS AMONG THE FEMALE STUDENTS OF CMSU.

Total number of Females in CMSU = 33

Total number of females in Accounting major: 3
 Probability of Accounting majors among the female students of CMSU: 0.0909 or 9.09 %

Total number of females in CIS major: 3
 Probability of CIS majors among the female students of CMSU: 0.0909 or 9.09 %

Total number of females in Economics/Finance major: 7
 Probability of Economics/Finance majors among the female students of CMSU: 0.2121 or 21.21 %

Total number of females in International Business major: 4
 Probability of International Business majors among the female students of CMSU: 0.1212 or 12.12 %

Total number of females in Management major: 4
 Probability of Management majors among the female students of CMSU: 0.1212 or 12.12 %

Total number of females in Other major: 3
 Probability of Other majors among the female students of CMSU: 0.0909 or 9.09 %

Total number of females in Retailing/Marketing major: 9
 Probability of Retailing/Marketing majors among the female students of CMSU: 0.2727 or 27.27 %

Total number of females in Undecided major: 0
 Probability of Undecided majors among the female students of CMSU: 0.0 or 0.0 %

2.4. ASSUME THAT THE SAMPLE IS A REPRESENTATIVE OF THE POPULATION OF CMSU. BASED ON THE DATA, ANSWER THE FOLLOWING QUESTION:

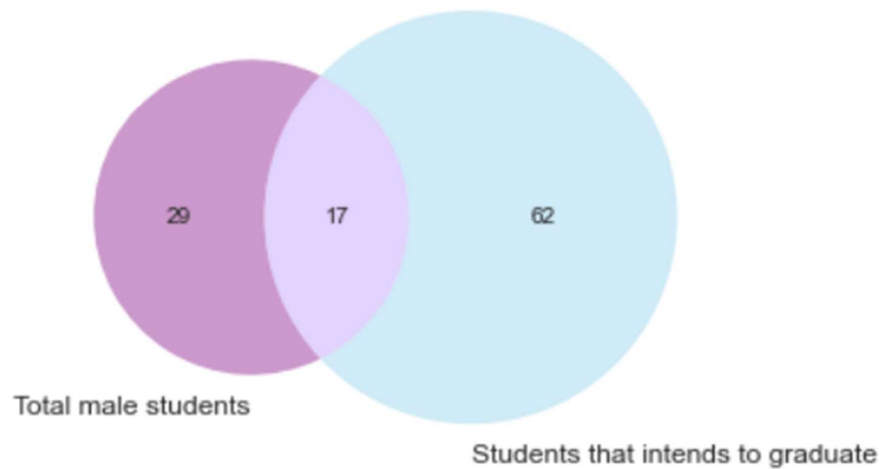## 2.4.1. FIND THE PROBABILITY THAT A RANDOMLY CHOSEN STUDENT IS A MALE AND INTENDS TO GRADUATE.

```
Total number of students that are male and intends to graduate = 17
Total number of students that intends to graduate in CMSU = 62
Probability that a randomly chosen student is a male and intends to graduate:  0.2742 or 27.42 %
```

Total number of students that are male and intends to graduate

Total male students

Students that intends to graduate

2.4.2 FIND THE PROBABILITY THAT A RANDOMLY SELECTED STUDENT IS A FEMALE AND DOES NOT HAVE A LAPTOP.
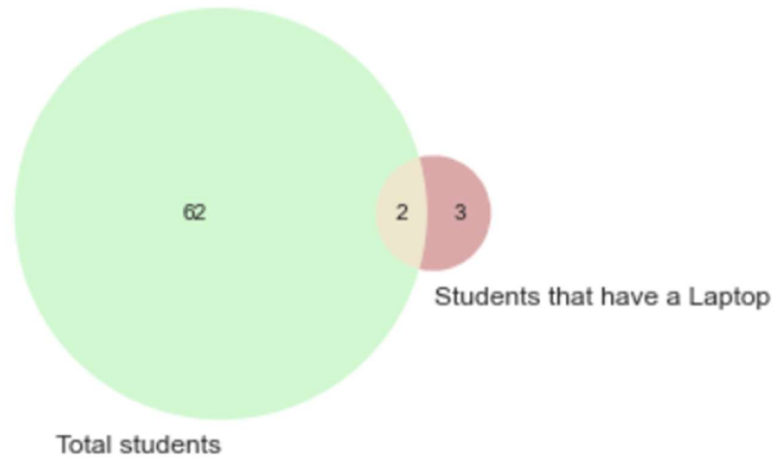
Total number of students that are Female and have a desktop = 2
Total number of students that are Female and have a tablet = 2
Probability that a randomly selected student is a female and <u>does NOT have a laptop:</u> 0.0645 or 6.45 %

Total number of students that are female and have a Laptop

62   2   3

Students that have a Laptop

Total students

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?
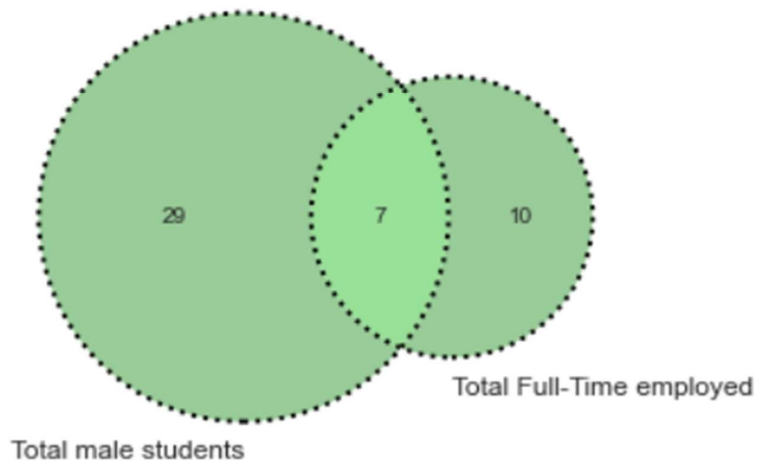
Probability that a randomly chosen student is a male: 0.4677 or 46.77 %
Probability that a randomly chosen student has full-time employment: 0.1613 or 16.13 %
Probability that a randomly chosen student is a male and has full-time employment: 0.1129 or 11.29 %
Probability that a randomly chosen student is either a male or has full-time employment: 0.5161 or 51.61 %

Student is either a male or has full-time employment

29   7   10

Total Full-Time employed

Total male students

2.5.2. FIND THE CONDITIONAL PROBABILITY THAT GIVEN A FEMALE STUDENT IS RANDOMLY CHOSEN, SHE IS MAJORING IN INTERNATIONAL BUSINESS OR MANAGEMENT.

Total number of Females in CMSU = 33

Total number of Females majoring in International Business = 4 / 33
Probability that given a female student is randomly chosen and she is majoring in International Business: 0.1212 or 12.12 %

Total number of Females majoring in Management = 4 / 33
Probability that given a female student is randomly chosen and she is majoring in Management: 0.1212 or 12.12 %

Probability that given a female student is randomly chosen, she is majoring in international business or management: 0.2424 or 24.24 %

Female student is either a international business or management major
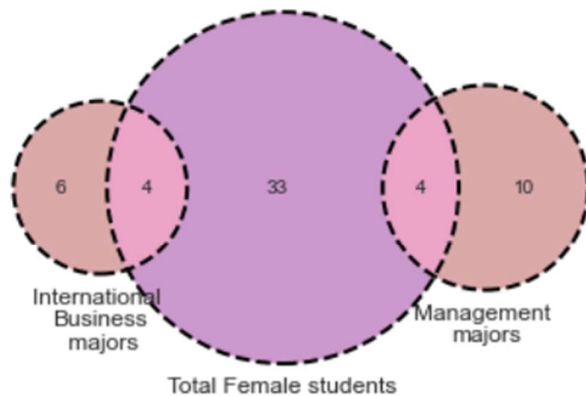
International Business majors

Management majors

Total Female students

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
Gender		
Female	9	11
Male	3	17

Probability that a randomly selected CMSU student will be female is: 33 / 62 = 0.5323 or 53.23 %

Probability that a student intends to graduate from CMSU = 28 / 62 = 0.4516 or 45.16 %

Product of the probabilities that a student will be female and that a student intends to graduate from CMSU is= 0.5323 * 0.4516 = 0.2404 or 24.0387 %

Probability that a student is female and intends to graduate from CMSU = 11 / 62 = 0.1774 or 17.74 %

Events are independent if the probability of an event occurring together is the product of their individual probabilities.

Since the probability that a student is female and intends to graduate from CMSU does <u>NOT equal</u> to the product of the probabilities that a student will be female and that a student intends to graduate from CMSU.

Therefore, it concludes that the graduate intention and being female 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.

2.7.1. IF A STUDENT IS CHOSEN RANDOMLY, WHAT IS THE PROBABILITY THAT HIS/HER GPA IS LESS THAN 3?
Total number of students with GPA less than 3: 17 / 62
Probability that a student is chosen randomly and that his/her GPA is less than 3: 0.2742 or 27.4194 %

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.
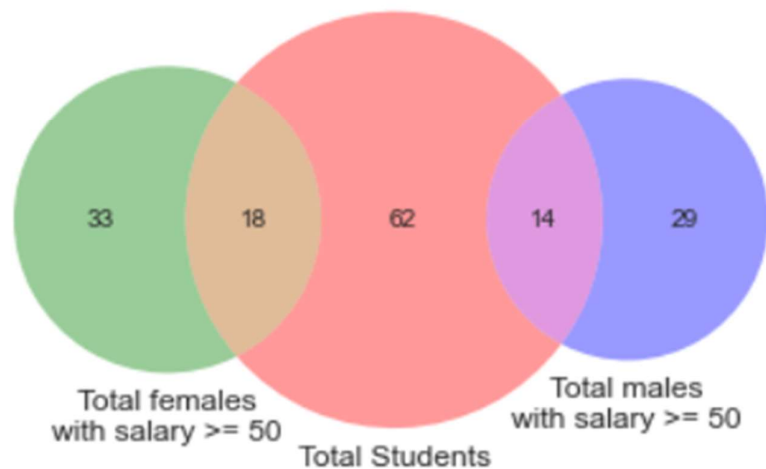
Salary	False	True
Gender		
Female	15	18
Male	15	14

Probability that a randomly selected male earns 50 or more: 14 / 29 = 0.4828 or 48.2759 %

Probability that a randomly selected female earns 50 or more: 18 / 33 = 0.5455 or 54.5455 %

Randomly selected male/female earns 50 or more

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 WHOLE PROBLEM 2.

GPA: skewed = -0.3146
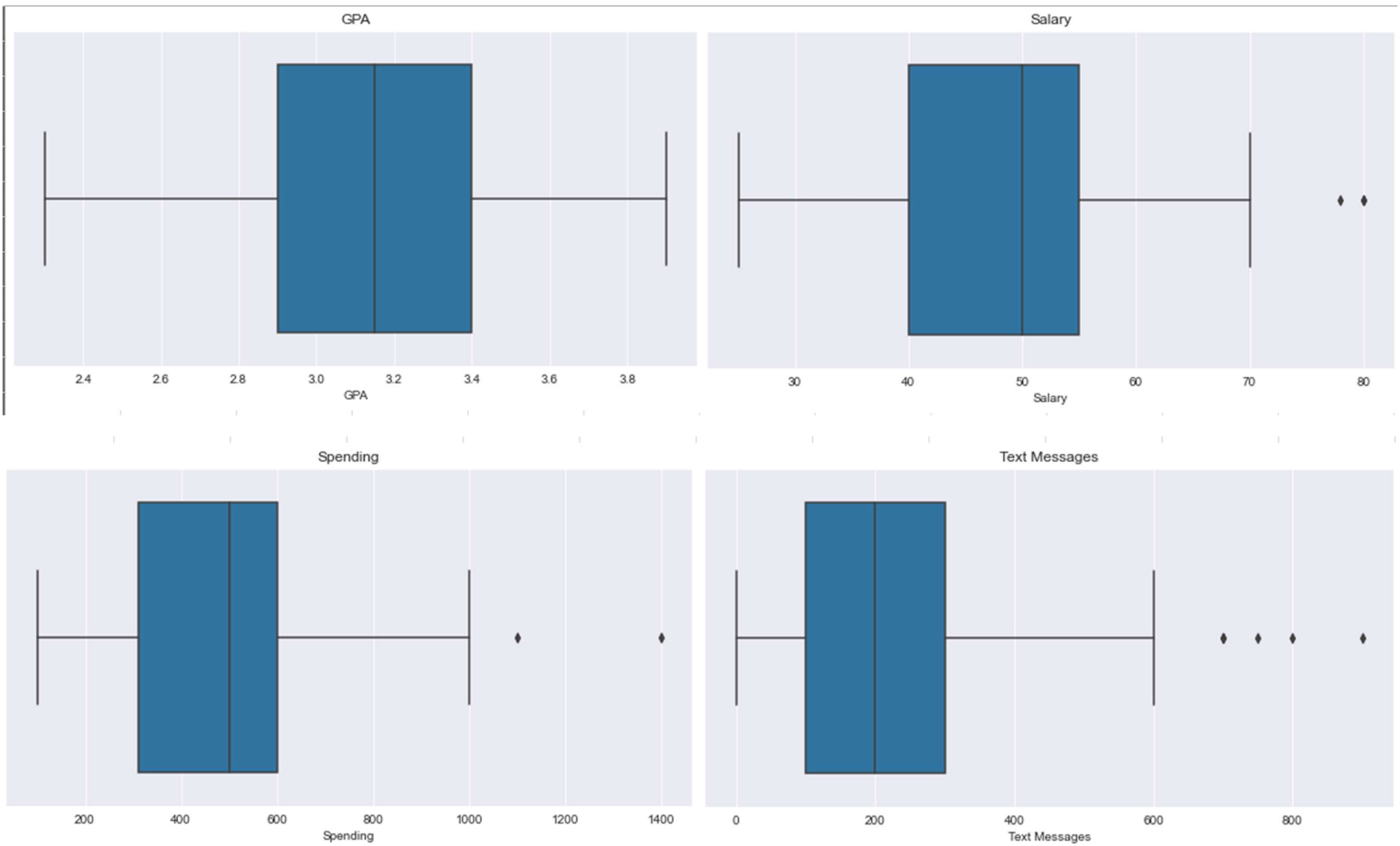GPA is negatively skewed or left skewed

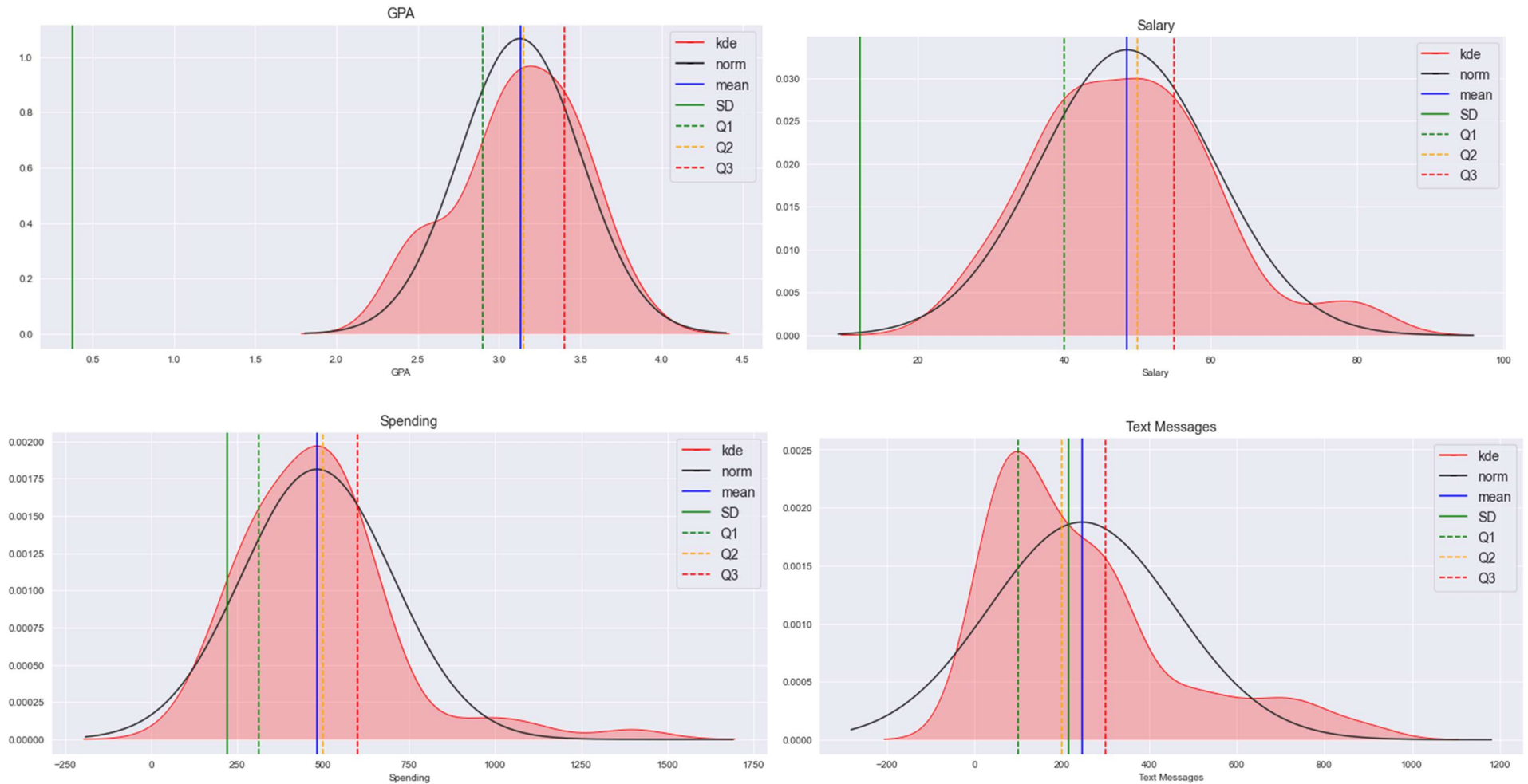Salary: skewed = 0.5347
Salary is positively skewed or right skewed

Spending: skewed = 1.5859
Spending is positively skewed or right skewed

Text Messages: skewed = 1.2958
Text Messages is positively skewed or right skewed

None of the continuous variables follow normal distribution

Summary

Clear Mountain State University (CMSU) has total 62 students out of which 29 of them are male and 33 of them are female. The Management program is mostly dominated by males - probability that a male will major in management is 20.69% and the probability of retail/marketing major by a female is 27.27%; a female opting for economics/finance major is 21.21%. Even so, most of the female's intention to graduate is still undecided with 17.74% have intentions to graduate, whereas the probability of a male and intentions to graduate are 60.71%.

Most of the students have a part-time job, 16.13% of a student has a full-time employment. 27.4194% of the students probably has the GPS less than 3.

PROBLEM STATEMENT 3 – SHINGLES MANUFACTURERS

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 colouring 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 is calculated. The company would like to show that the mean moisture content is less than 0.35 pound per 100 square feet.

The file (A & B shingles.csv) includes 36 measurements (in pounds per 100 square feet) for A shingles and 31 for B 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.

Ho = The mean moisture content is less than equal to 0.35 pound per 100 square feet.

H_A = The mean moisture content is more than 0.35 pound per 100 square feet.

OR

$$Ho = \mu \leqq 0.35$$

$$H_A = \mu > 0.35$$

Level of significance or alpha 0.05

A shingles

Sample Mean of A shingles: 0.3167
Sample Std Deviation of A shingles: 0.3167
Total number of observations for A shingles: 36

One sample t test for A shingles
t statistic: -1.4735 p value: 0.1496

We have no evidence to reject the null hypothesis since p value > Level of significance
Our one-sample t-test p-value= 0.1496

which is greater than 5% level of significance

At 95% confidence level, there is sufficient evidence to prove that mean moisture content is less than equal to 0.35 pound per 100 square feet.

B shingles

Sample Mean of B shingles: 0.2735
Sample Std Deviation of B shingles: 0.2735
Total number of observations for B shingles: 31

One sample t test for B shingles
t statistic: -3.1003 p value: 0.0042

We have evidence to reject the null hypothesis since p value < Level of significance
Our one-sample t-test p-value= 0.0042
which is less than 5% level of significance

Hence, at 5% level of significance, there is not enough evidence that the mean moisture content is less than equal to 0.35 pound per 100 square feet.

$$Ho = \mu_o = \mu_A$$

$$H_A = \mu_o \neq \mu_A$$

Mean for Shingles A is 0.3167
Mean for Shingles B is 0.2735
t_statistic= 1.2896 and pvalue= 0.2017

Level of significance or alpha 0.05

We have no evidence to reject the null hypothesis since p value > Level of significance
Our one-sample t-test p-value= 0.2017
which is greater than 5% level of significance

At 95% confidence level, there is sufficient evidence to prove that population mean for shingles A and B are equal.