Business Report Project SMDM

Shubham Gupta

# Synopsis

Problem 1
Wholesale Customers Analysis

 Problem Statement:

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).

Q-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?

Ans.

|	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
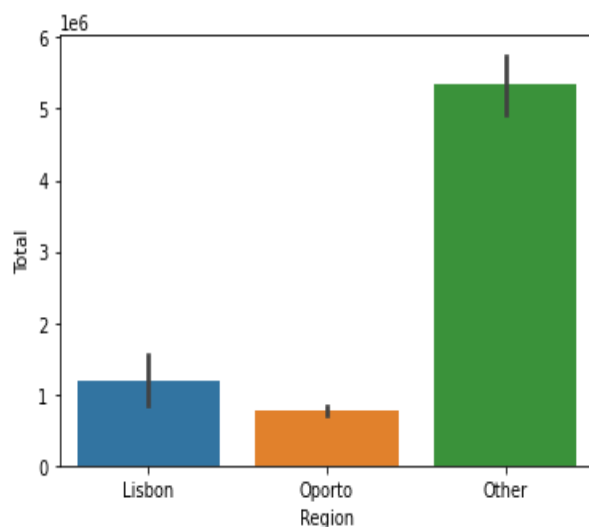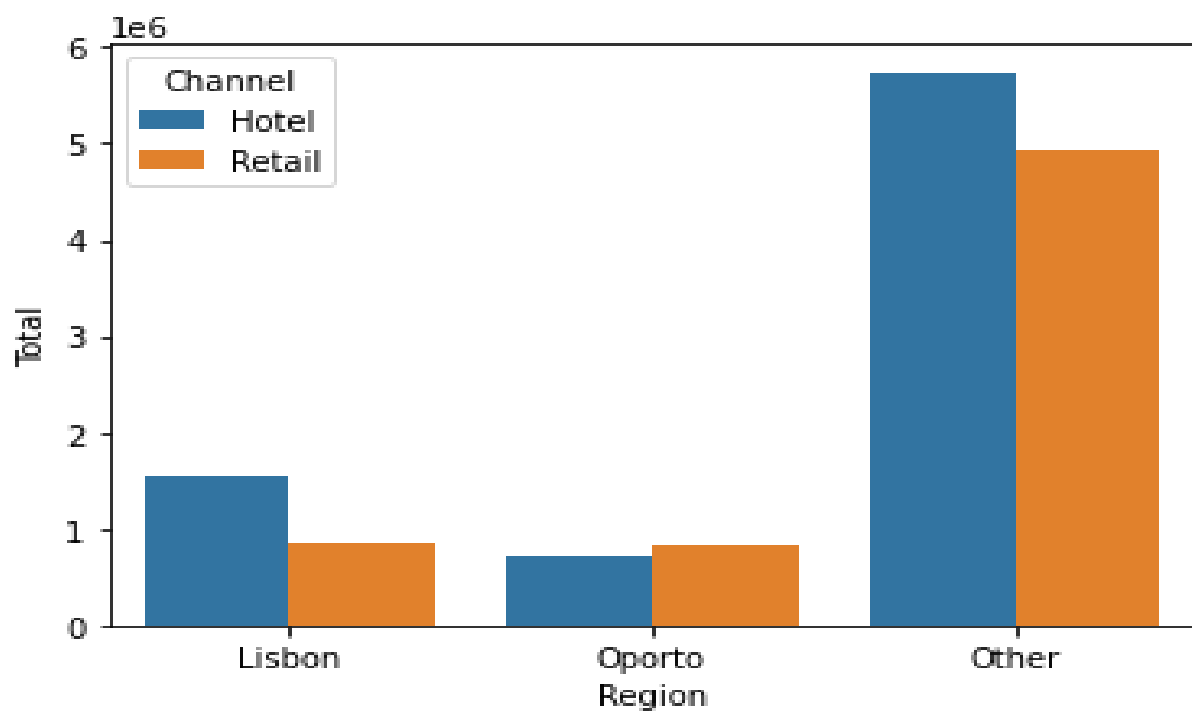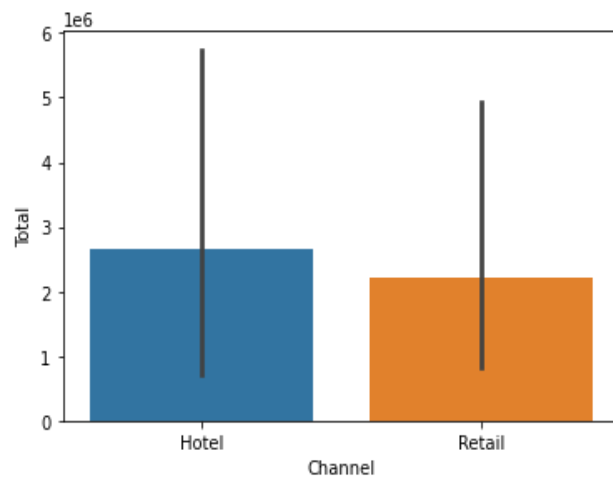
From the above graph we have conclude that
Region – Other spends more
Channel- Hotel spends more
Region – Oporto spends less
Channel – Retail spends less

3

Q-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.

Ans.
We have 6 varieties so, if we see across region, we can find insights of 6 varieties.

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
Region						
Lisbon	854833	422454	570037	231026	204136	104327
Oporto	464721	239144	433274	190132	173311	54506
Other	3960577	1888759	2495251	930492	890410	512110
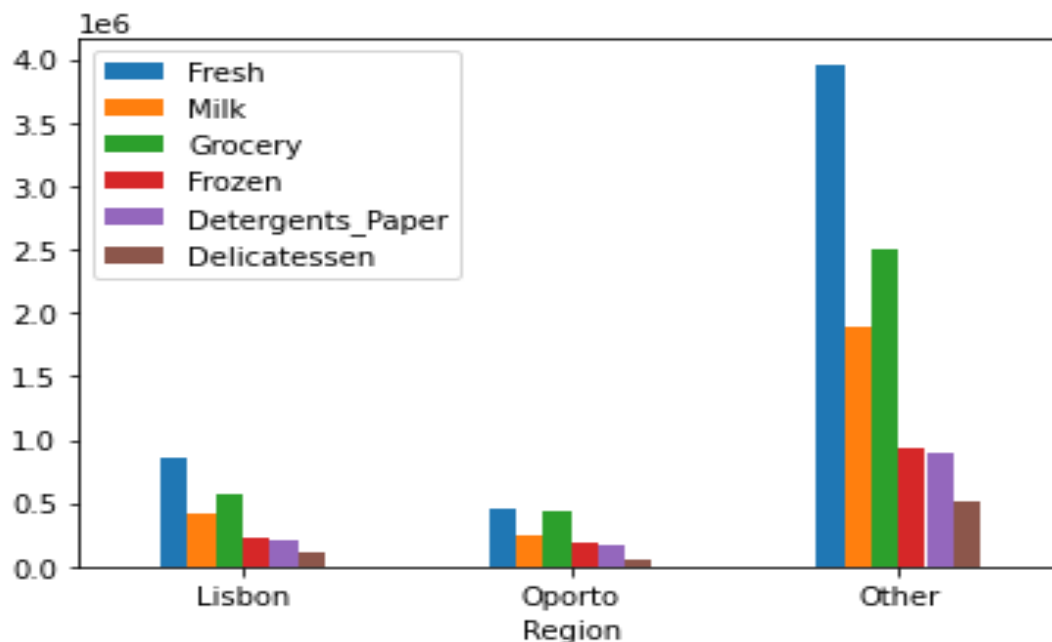
In Other Region we are spending maximum & Oporto Region we are spending less.

We have 6 varieties so, if we see across Channel, we can find insights of 6 varieties.

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
Channel						
Hotel	4015717	1028614	1180717	1116979	235587	421955
Retail	1264414	1521743	2317845	234671	1032270	248988
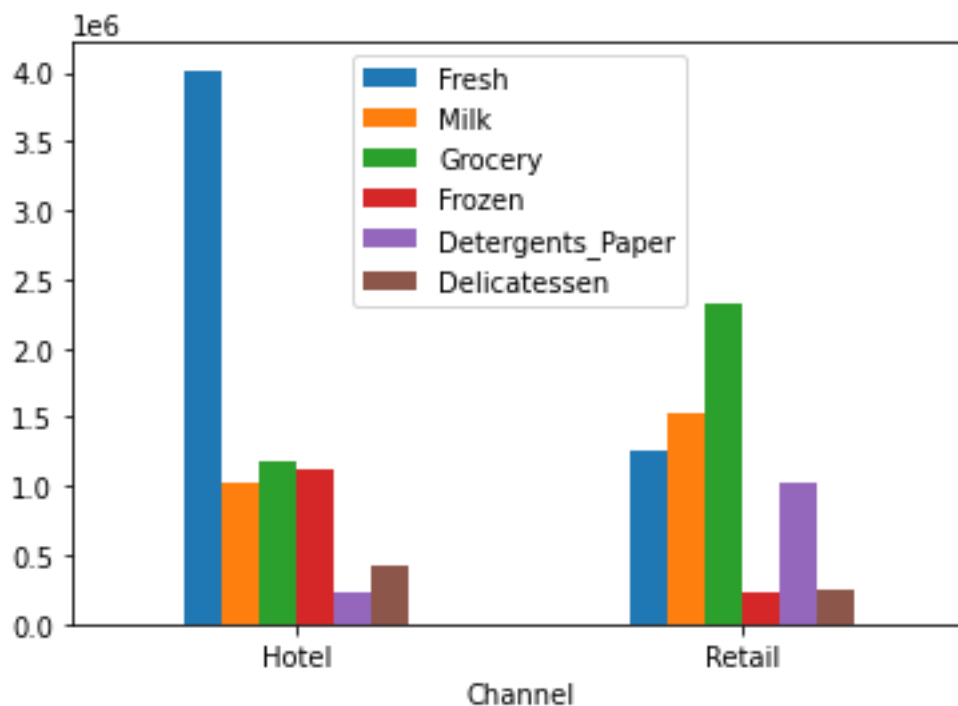
We are spending less in Retail Channel as compare to Hotel Channel.

Q- 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?

Ans.
We can use IQR method or STD to find the item of most & least in inconsistent behaviour.

IQR for all 6 items

|	
Buyer/Spender	219.50
Fresh	13806.00
Milk	5657.25
Grocery	8502.75
Frozen	2812.00
Detergents Paper	3665.25
Delicatessen	1412.00

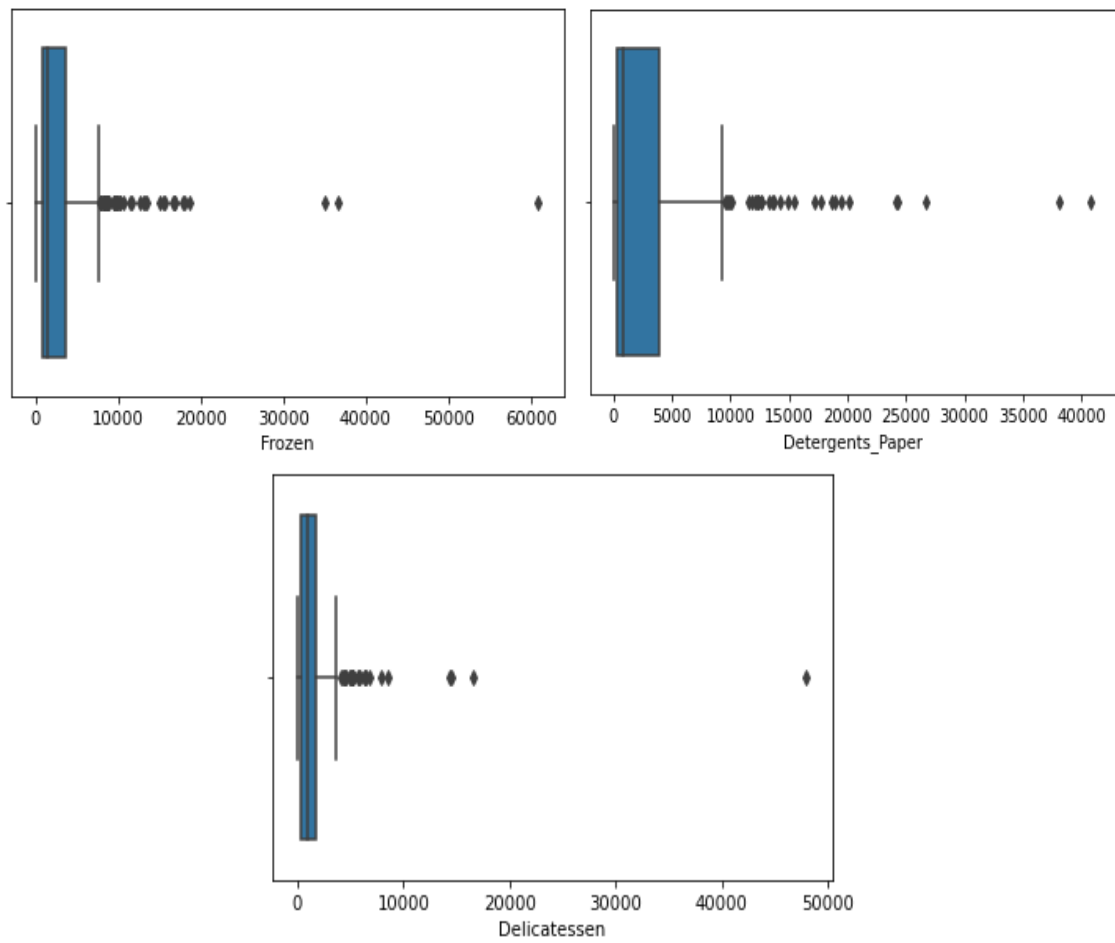Fresh item seems to be the most inconsistent behaviour.
Delicatessen item seems to be the least inconsistent behaviour.

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

Ans.

Yes, all 6 varieties of outliers in dataset.
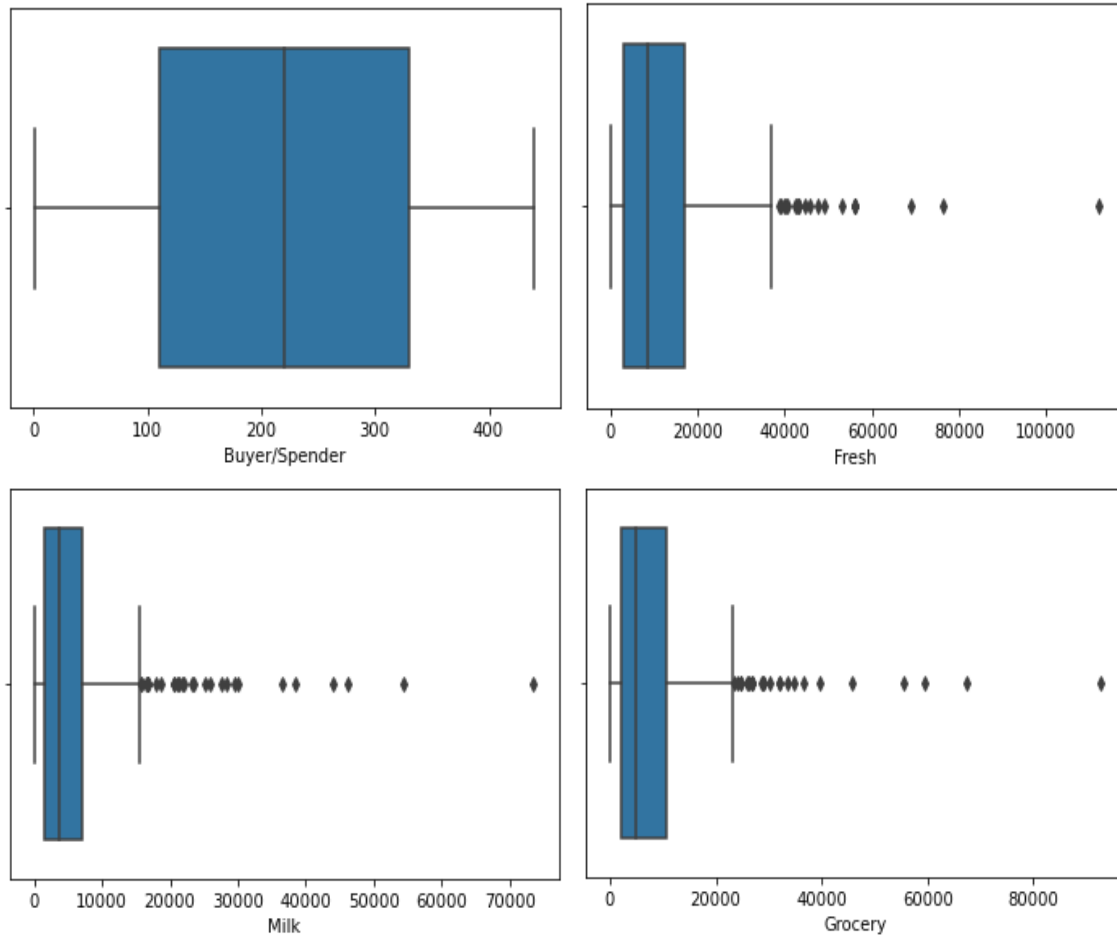
Q-1.5 On the basis of your analysis, what are your recommendations for the business? How can your analysis help the business to solve its problem? Answer from the business perspective.

Ans.

On the basis of the analysis the following recommendations can be made:

• On the basis of the analysis, it can be seen that the region Other and the channel Retail have higher spending than the Channel and Region. Hence from the Business perspective if a new business is to setup it should be opened in the Other region with Channel Retail as the Other reason absorbing maximum amount of sell and this can be boast up the Revenue compared to opening new business in Lisbon or Oporto and with the Channel Hotel.

• In all the region the Food Items Fresh has the highest spending followed by Grocery and Milk. Hence, these food products are strongly recommended to be available.

• Also, the food Item Delicatessen shows least inconsistent behaviour across all the regions and the channels. So, Delicatessen is also recommended to be available all the times in the business.

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).

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

Q-2.1.1. Gender and Major

Ans.

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

Q-2.1.2. Gender and Grad Intention

Ans.

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

Q-2.1.3. Gender and Employment

Ans.

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

Q-2.1.4. Gender and Computer

Ans.

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

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

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

Ans. The probability that a randomly selected CMSU student will be male 46.77.

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

Ans. The probability that a randomly selected CMSU student will be female 53.22.

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

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

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

Ans.

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

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

Q-2.4.1. Find the probability That a randomly chosen student is a male and intends to graduate.

Ans.

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

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

Ans.

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

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

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

Ans. The probability that a randomly chosen student is either a male or has full-time employment 51.6%.

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

Ans.

Probability that given a female student is randomly chosen, she is majoring in international business or management 24

Q-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?

Ans.

The probability that a randomly selected Student is Female 50.0
The probability that a randomly selected student is female and intends to graduate 55.00000000000001 %
They are not independent events

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

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

Ans.

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

Q-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.

Ans.

Salary	False	True
Gender		
False	0.517241	0.482759
True	0.454545	0.545455

Q-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.
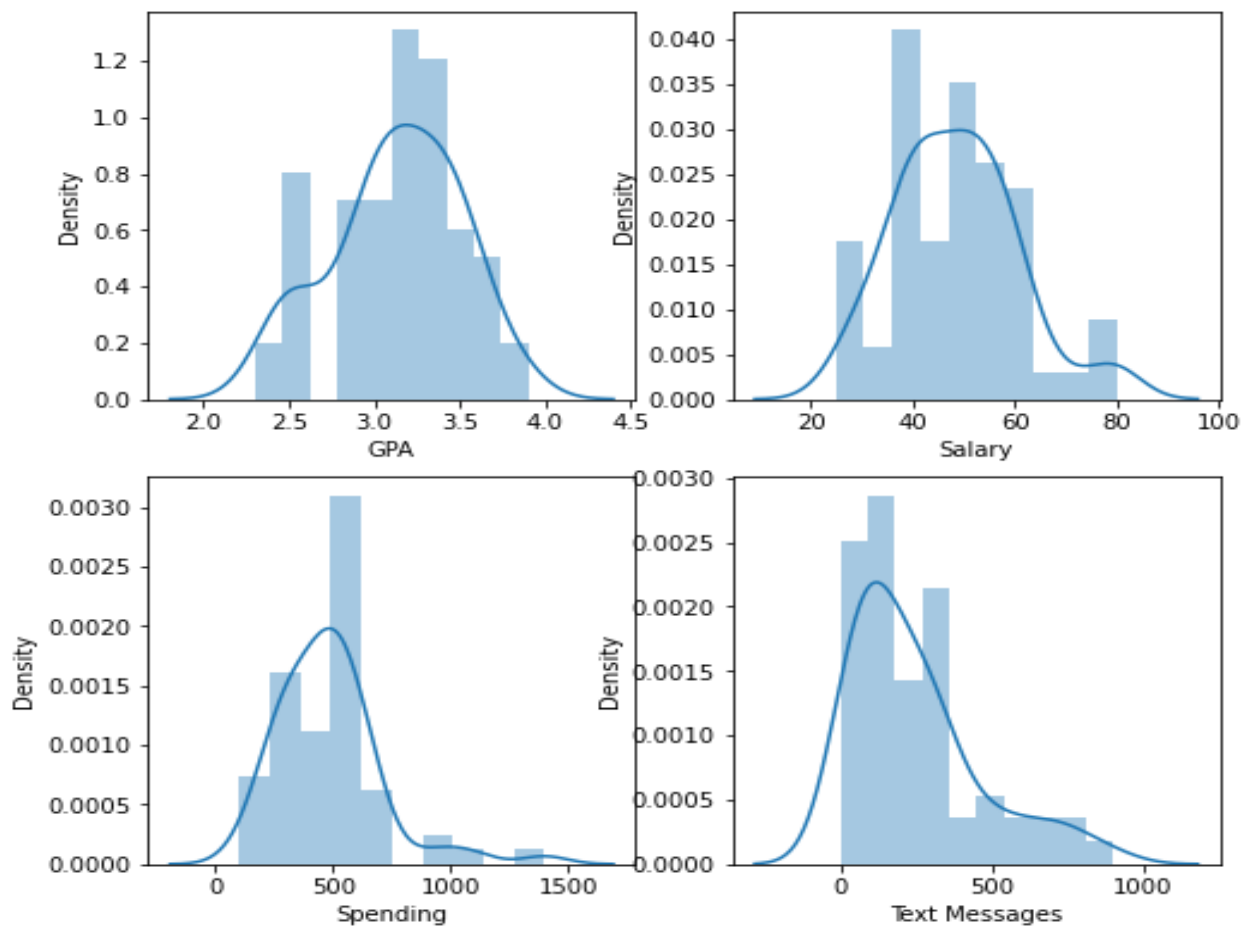
Ans.

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

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.

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

Q-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

Ans.

<div align="center">1 sample t test for the A shingles</div>

Step 1	H0	mean moisture content <= 0.35
	HA	mean moisture content > 0.35

Step 2   0.05   level of significance

$$t = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

Step 3   1 sample t test

Test statistic	-1.4735	1.473505

xbar	0.316667
s	0.135731
n	36
mu	0.35

Step 4	0.07477633	p_value	0.074776

Step 5   p_value > 0.05   We fail to reject the H0

on the basis of the hypothesis test performed for the given sample of 36 observations at 95% confidence

we fail to reject the null hypothesis i.e. the mean moisture content is less than or equal to 0.35

Step 1   H0   mean moisture content <= 0.35

                  HB   mean moisture content > 0.35

Step 2   0.05   level of significance

Step 3   1 sample t test

$$t = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

Test statistic   -3.10033   3.100331

                  xbar   0.273548

                  s   0.137296

                  n   31

                  mu   0.35

Step 4   0.00209   p_value   0.00209

Step 5   p_value > 0.05   We reject the H0

on the basis of the hypothesis test performed for the given sample of 31 observations at 95% confidence

we reject the null hypothesis i.e. the mean moisture content is less than or equal to 0.35

P value (0.0021) is less than level of significance (0.05)

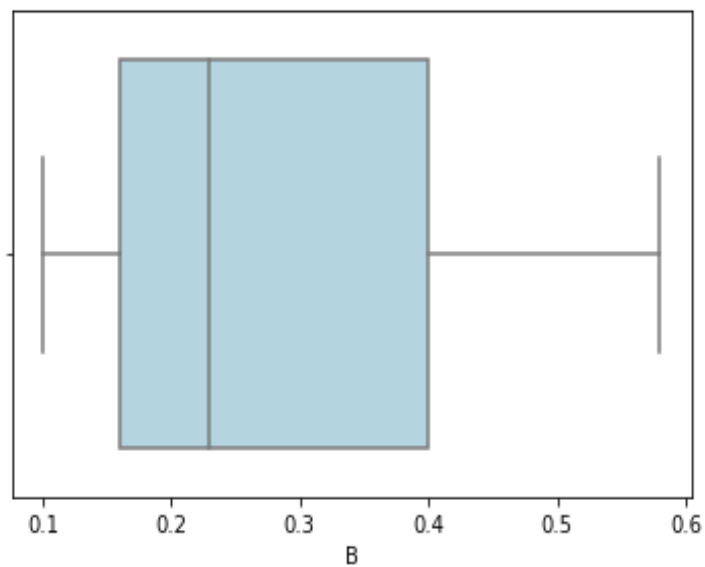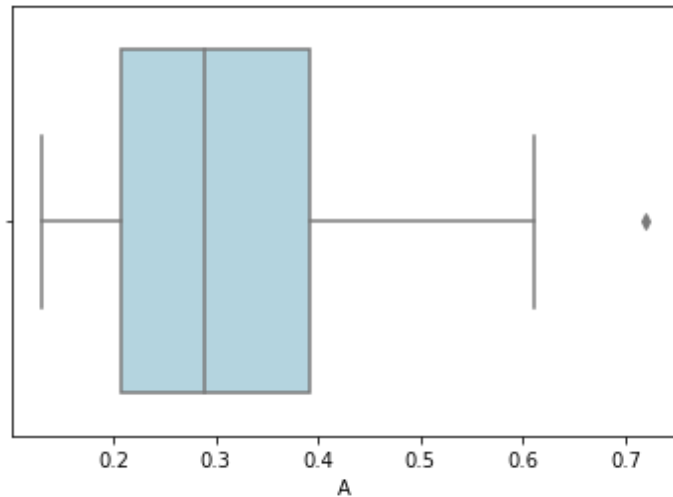Hence, we reject H0 mean moisture content is <=0.35

Q- 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?
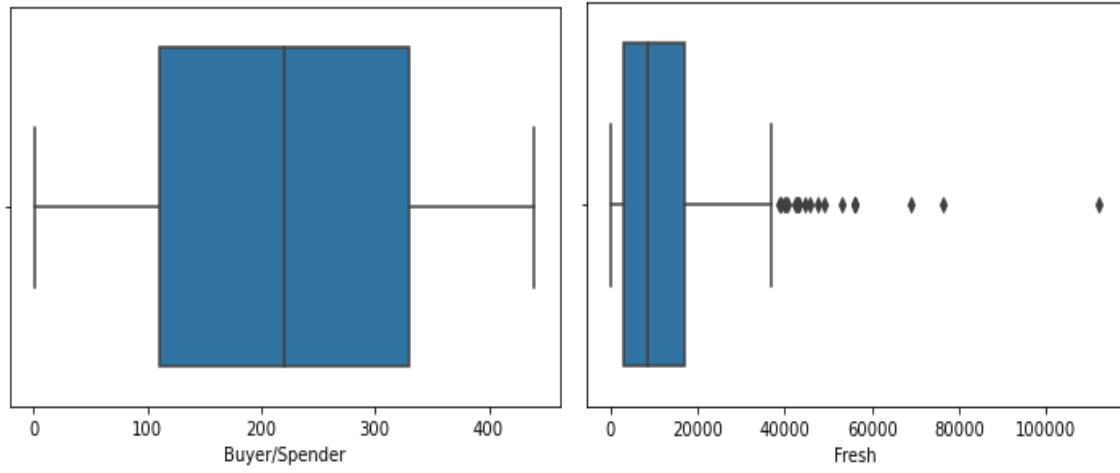
Ans.

t stat 1.2896282719661123
p-value 0.2017496571835306
we failed to reject null hypothesis

THANK YOU