

PROBLEM 1: Wholesale Customers Analysis

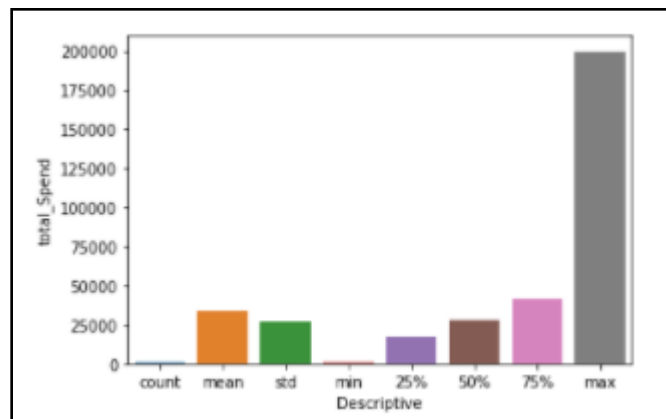
Problem Statement:

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).

1.1.1 Use methods of descriptive statistics to summarize data.

Summarizing descriptive statistics in a table showing full information of the data.

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	total_Spend
count	440.000000	440.000000	440.000000	440.000000	440.000000	440.000000	440.000000	440.000000
mean	220.500000	12000.297727	5796.265909	7951.277273	3071.931818	2881.493182	1524.870455	33226.136364
std	127.161315	12647.328865	7380.377175	9503.162829	4854.673333	4767.854448	2820.105937	26356.301730
min	1.000000	3.000000	55.000000	3.000000	25.000000	3.000000	3.000000	904.000000
25%	110.750000	3127.750000	1533.000000	2153.000000	742.250000	256.750000	408.250000	17448.750000
50%	220.500000	8504.000000	3627.000000	4755.500000	1526.000000	816.500000	965.500000	27492.000000
75%	330.250000	16933.750000	7190.250000	10655.750000	3554.250000	3922.000000	1820.250000	41307.500000
max	440.000000	112151.000000	73498.000000	92780.000000	60869.000000	40827.000000	47943.000000	199891.000000



1.1.2 Which Region and which Channel spent the most?

Max total expenditure is 199891

Buyer/Spender	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	total_Spend	
85	86	Retail	Other	16117	46197	92780	1026	40827	2944	199891

From 440 total numbers of large retailers, Buyer/ Spender number 88 has the most annual spending on 6 different varieties of products that is in region (Other) and sales channel (Retail).

1.1.3 Which Region and which Channel spent the least?

Buyer/Spender	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	total_Spend	
154	155	Hotel	Other	622	55	137	75	7	8	904

From 440 total numbers of large retailers, Buyer/ Spender number 155 has the least annual spending on 6 different varieties of products that is in region (Other) and sales channel (Hotel).

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.

The variability result for all the varieties is : -

Coefficient of Variation of Fresh is 1.0527196084948243

Coefficient of Variation of Milk is 1.27185083074245

Coefficient of Variation of Grocery is 1.1938154477492668

Coefficient of Variation of Frozen is 1.578535529860776

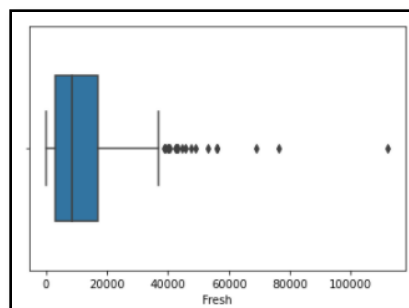
Coefficient of Variation of Detergents_Paper is 1.6527657881041735

Coefficient of Variation of Delicatessen is 1.8473041039189302

The behavior of all the varieties is overall same. The variability is high in all the varieties as all the value of CV exceeded 1 or 100%, Simply means that the standard deviation exceeds the mean value. Into the varieties the least variability is in Fresh = 1.05(105%), followed by Grocery = 1.19(119%), Milk = 1.27(127%), Frozen = 1.57(157%), Detergents paper = 1.65(165%) and the most variability is in Delicatessen = 1.84(184%).

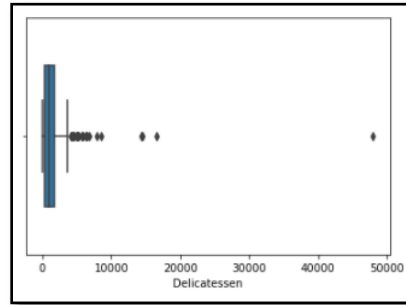
1.3 On the basis of a descriptive measure of variability,

1.3.1 which item shows the most inconsistent behavior?



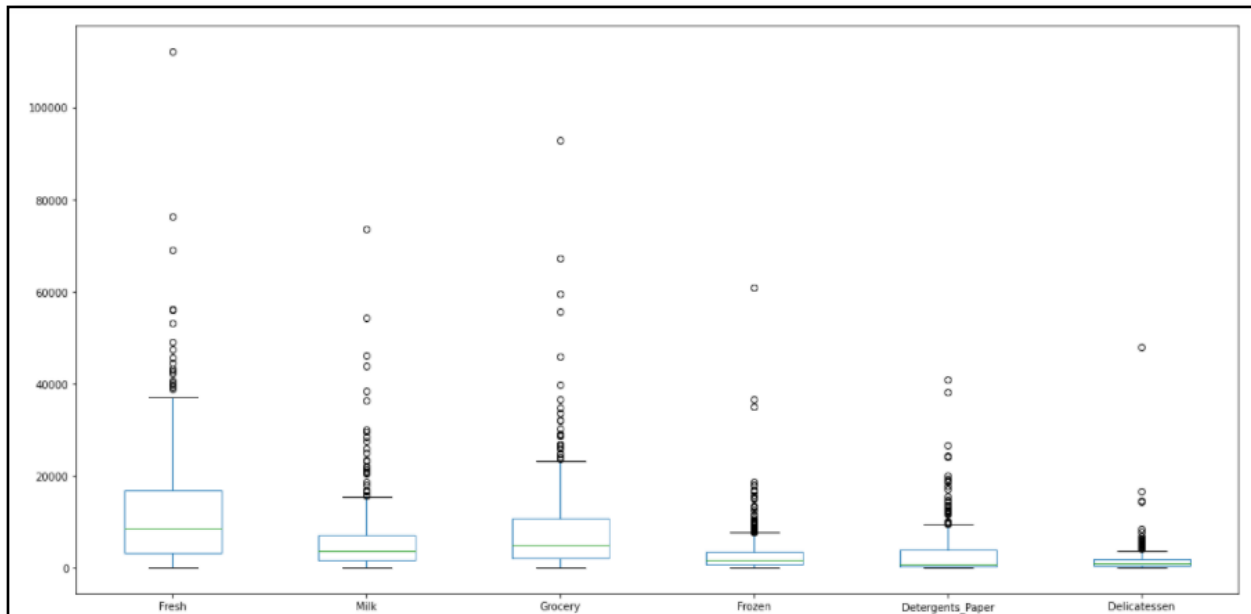
The smaller the IQR or smaller the box size, the more consistent the data values are closer to the median of the data. So, In this case we have the most inconsistent behavior under **Fresh**, as the IQR and outliers are widely spread.

1.3.2 Which items show the least inconsistent behavior?



The least inconsistent behavior can be seen under **Delicatessen**, as it has the smallest box among all other boxes and also less outlier value compare to the other's outlier.

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



****Comment.**

All the variety of products has outlier, graphically presented in the box plot. There are lot of outliers above the maximum range i.e., ($Q3 + 1.5 * IQR$) in all the variety of products.

1.4 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.

A high level of outliers can lead to the chances of getting more events and new opportunities for the business to grow, so in such cases more the outlier if to be seen into a variety of product buyers/spenders should focus on the growth of that product. Here is this case we get the insights as the Fresh named variety has the higher number of outliers as compared to any of the other products so investments should be done more on Fresh product variety by the customers.

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.

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
Gender								
Female	3	3	7	4	4	3	9	0
Male	4	1	4	2	6	4	5	3

2.1.2. Gender and Grad Intention

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

2.1.3. Gender and Employment

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

2.1.4. Gender and Computer

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

2.2. Assume that the sample is 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?

Class	Junior	Senior	Sophomore	All
Gender				
Female	14	17	2	33
Male	11	14	4	29
All	25	31	6	62

Here, Total No. of CMSU student is 62.

Total number of male is 29.

So, this is total number of the male divided by the total students.

Or, $29/62 = 0.46774193548387094 * 100 = 46.77 \%$

Hence, 46.77 % will be male if randomly selected CMSU students.

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

Here, Total No. of CMSU student is 62. Total number of female is 33.

So, this is total number of the female divided by the total students.

Or, $33/62 = 0.532258064516129 * 100 = 53.22 \%$

Hence, 53.22 % will be female if randomly selected CMSU students.

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.

Probability of male candidate in the Accounting Major is 13. %

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

Probability of male candidate in the CIS Major is 3.44 %
 Probability of male candidate in the Economics/Finance Major is 13.79 %
 Probability of male candidate in the International Business Major is 6.89 %
 Probability of male candidate in the Management Major is 20.68 %
 Probability of male candidate in the Other Major is 13.79 %
 Probability of male candidate in the Retailing/Marketing Major is 17.24 %
 Probability of male candidate in the Undecided Major is 10.34 %

** Comment = Conditional probability of different majors among male = $P(\text{All the different Majors} \mid \text{Total number of male})$

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

Probability of female candidate in the Accounting Major is 9.09 %
 Probability of female candidate in the CIS Major is 9.09 %
 Probability of female candidate in the Economics/Finance Major is 21.21 %
 Probability of female candidate in the International Business Major is 12.12 %
 Probability of female candidate in the Management Major is 12.12 %
 Probability of female candidate in the Other Major is 9.09 %
 Probability of female candidate in the Retailing/Marketing Major is 27.27 %
 Probability of female candidate in the Undecided Major is 0.0 %

** Comment = Conditional probability of different majors among the female = $P(\text{All the different Majors} \mid \text{Total number of female})$

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	All
Gender				
Female	9	13	11	33
Male	3	9	17	29
All	12	22	28	62

Probability that a randomly chosen student is a male and intends to graduate is 58.62 %.

** Comment = $P(\text{Grad intention} \mid \text{Total Male}) * 100 = P(17/29) * 100$

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

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

Probability that a randomly selected student is a female and does NOT have a laptop 12.12 %

** Comment = $P(\text{Female with desktop} + \text{Female with Tablet} \mid \text{Total Female}) * 100 = ((2+2)/33) * 100$

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?

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

Probability that a randomly chosen student is a male 46.77 %

Probability that a randomly chosen student has full-time employment 16.12 %

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	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

Conditional probability that given a female student is randomly chosen, she is majoring in international business or management 24.24 %

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	Column_Total
Gender			
Female	9	11	20
Male	3	17	20
Row_Total	12	28	40

Probability of female get selected randomly = $20/40 \times 100 = 50\%$

Probability of female intend to graduate = $11/20 \times 100 = 55\%$

These even are not depending on each other, acted as 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?

Probability of getting randomly select a student as his/her GPA can be less than 3:- 27.41 %

** Comment = Calculated by the total value of the male and female students.

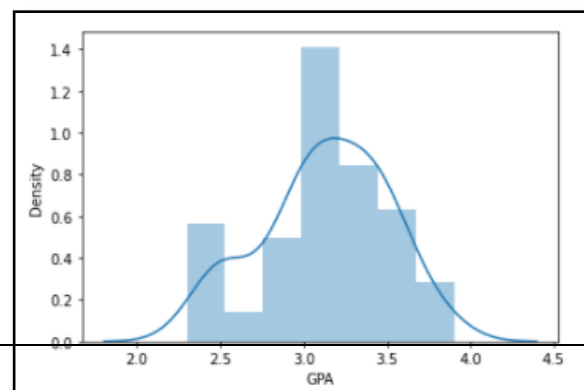
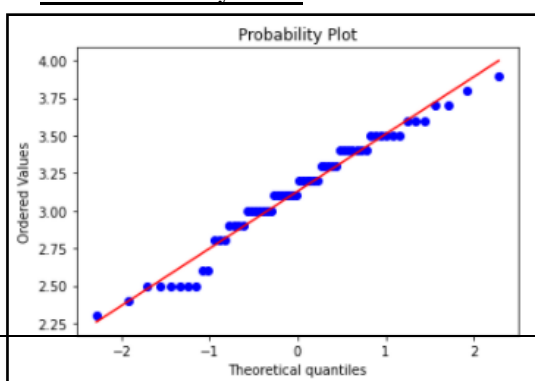
2.7.2. Find the conditional probability that a randomly selected male earns 50 or more. Find the conditional probability that a randomly selected female earns 50 or more.

Conditional probability that a randomly selected male earns 50 or more 48.27 %

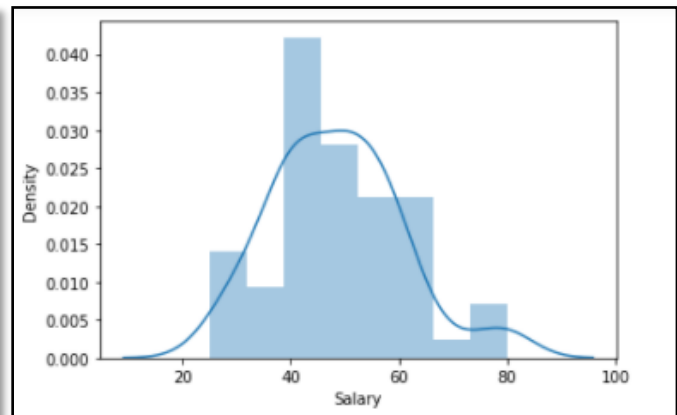
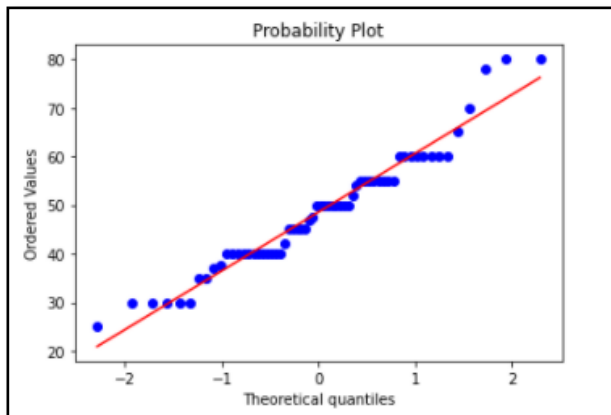
Conditional probability that a randomly selected female earns 50 or more 54.54 %

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.

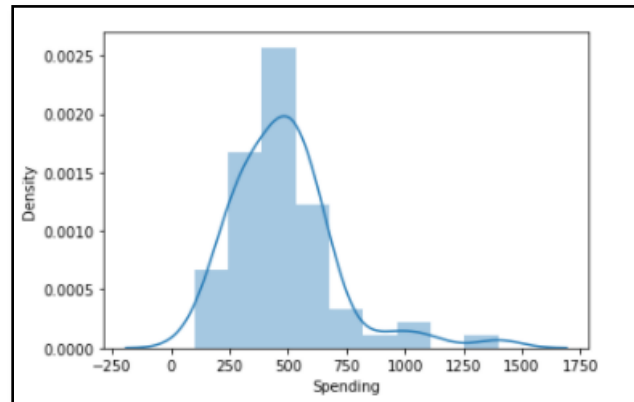
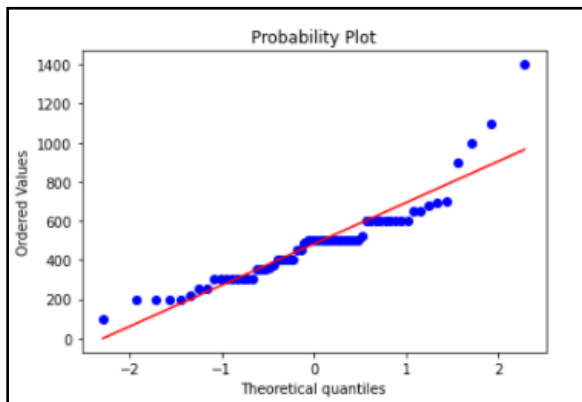
GPA Probability Plot:-



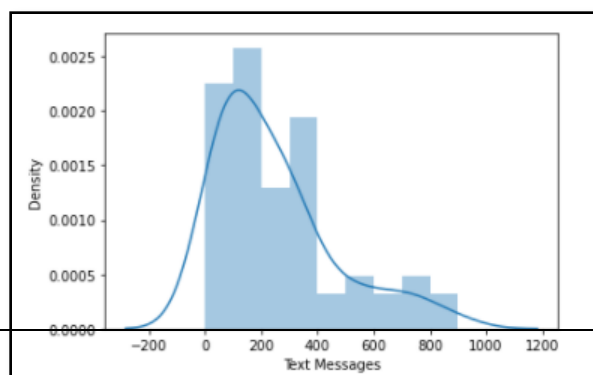
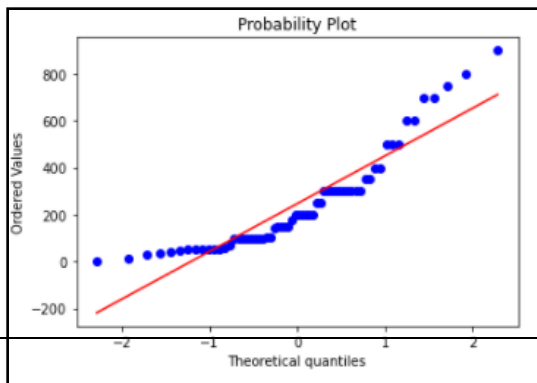
Salary Probability Plot:- (Positively skewed)



Spending Probability Plot:-



Text Messages Probability Plot:-



We have skewed data towards right so that the majority data stays with positive values, indicates positive skewness.

CONCLUSION

With 62 responses both from male and female students, Conditional probability shows that a random selection for male or female both has earning probability near to 50 % or more. Many students wants to graduate from retailing/marketing and economic finance.

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.

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?

Please reflect on all that you have learnt while working on this project. This step is critical in cementing all your concepts and closing the loop.