MODULE – 2
SMDM PROJECT

Name: Sudheendra

PGP-DSBA.O.APR22C

DATE: 26th June, 2022

Table of Contents

Contents

Wholesale Customers Analysis

Introduction

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 and Others) and across different sales channel (Hotel and Retail).

Let's do few exploratory data analysis and explore the dataset using central tendency and other parameters. This assignment should help us in exploring different attributes, the summarized statistics, contingency tables, conditional probabilities and hypothesis testing.

Macro View of Data Frame

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

Table 1. Dataset Sample

From the above dataset we can observe that there are nine variables with two categorical variables and remaining all seven continuous variables.

Exploratory Data Analysis

```
<class 'pandas.core.frame.DataFrame'>
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)
memory usage: 31.1+ KB
```

From the above information, we see that

- Dataset contains 440 & 9 Rows and Columns respectively
- 440 counts in all the variables
- 2 Object data types
- 7 integer data types
- From the above results we can also see that there is no missing value present in the dataset.

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

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Buyer/Spender	440	NaN	NaN	NaN	220.5	127.1613	1	110.75	220.5	330.25	440
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	NaN	NaN	NaN	12000.3	12647.33	3	3127.75	8504	16933.75	112151
Milk	440	NaN	NaN	NaN	5796.266	7380.377	55	1533	3627	7190.25	73498
Grocery	440	NaN	NaN	NaN	7951.277	9503.163	3	2153	4755.5	10655.75	92780
Frozen	440	NaN	NaN	NaN	3071.932	4854.673	25	742.25	1526	3554.25	60869
Detergents_Paper	440	NaN	NaN	NaN	2881.493	4767.854	3	256.75	816.5	3922	40827
Delicatessen	440	NaN	NaN	NaN	1524.87	2820.106	3	408.25	965.5	1820.25	47943

Table 2. Summary of the Data

From the descriptive statistics, we can see there are:

- Two unique values in Channel Variable
- Three unique values in Region Variable
- The minimum value seems to be 3.0 for Fresh, Grocery, Detergents Paper and Delicatessen
- From the IQR values we understand the range of data lies in the 25%, 50% and 75%
- The maximum value seems to be 112151 holding by Fresh
- We can see that there are 3 Region Variables:
    - Other 316 counts
    - Lisbon 77 counts
    - Oporto 47 counts
- Also, we can see there are 2 types of Channels:
    - Hotel 298 counts
    - Retail 142 counts

Note: NaN shows that the values cannot be calculated for those particular variables.

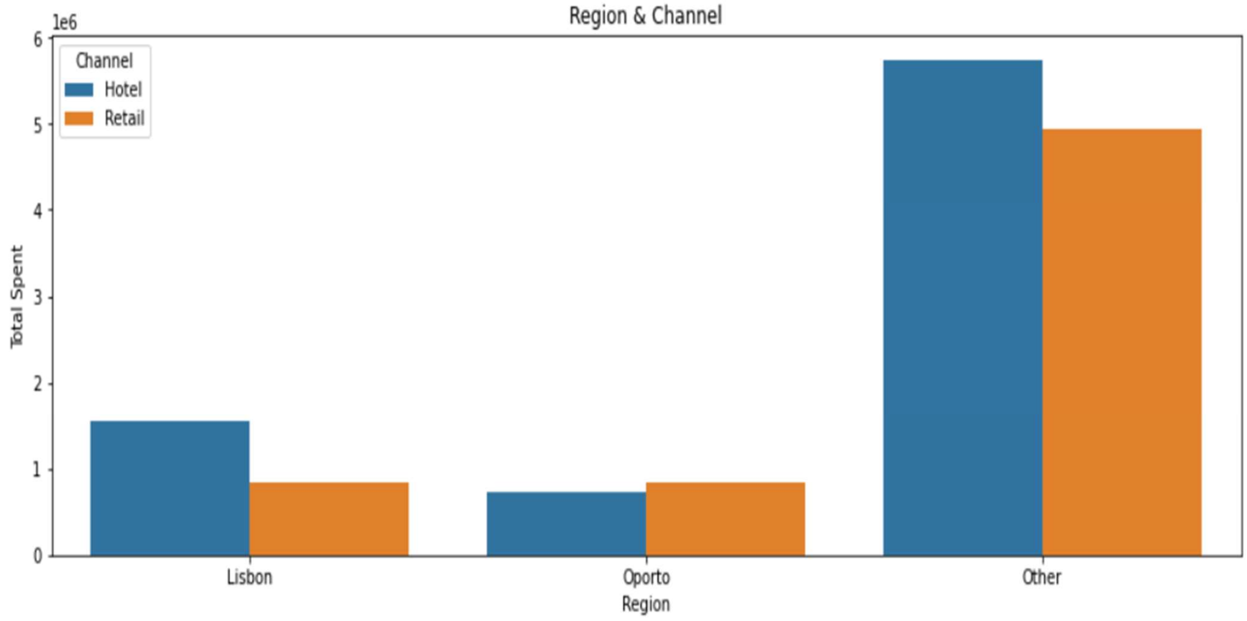


Fig.1

From the above graph we can conclude the following:

- Buyers from **Region-Others** spends more
- Buyers from **Channel-Hotel** spends more
- Buyers from **Region-Oporto** spends less
- Buyers from **Channel-Hotel** spends less

And to conclude with we can see that

- Buyer/Spenders from Other-Region & through Hotel Channel spends more.
- Buyer/Spenders from Oporto-Region & through Hotel Channel spends 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.

	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

Table.3 Summary as per Region

Looking at six different items across all the given Region, from the below Fig.2

- Consumers spending is maximum on all varieties in **Other-Region**
- The maximum is spent on **Fresh** products and minimum is spent on **Delicatessen**
- Consumers spending is less on all varieties in **Oporto-Region** compared to the rest of two other regions

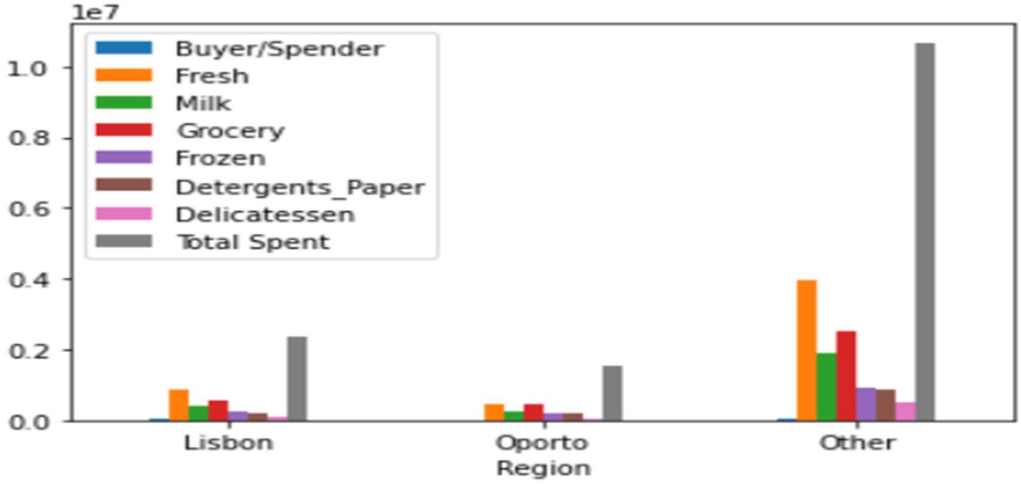


Fig.2

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
Channel						
Hotel	4015717	1028614	1180717	1116979	235587	421955
Retail	1264414	1521743	2317845	234671	1032270	248988

Table.3 Summary as per Channel

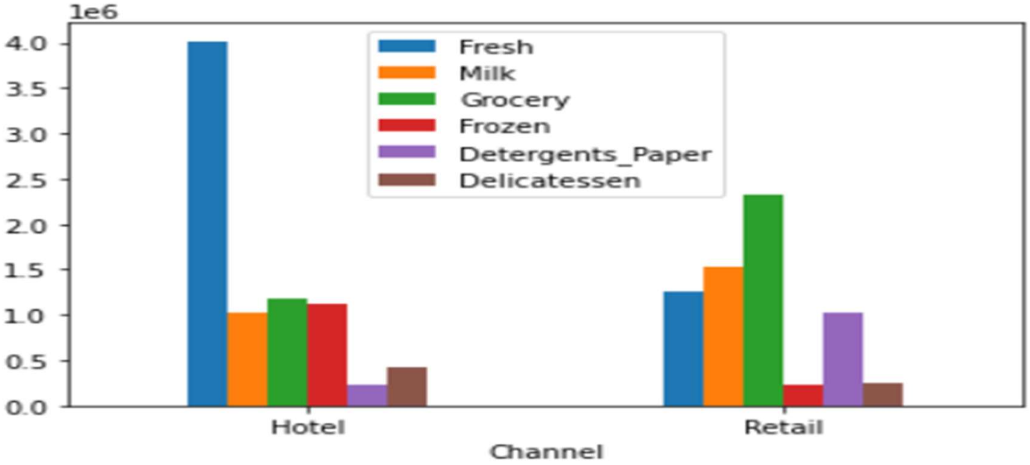


Fig.3

Now from this above plot(Fig) we can see all six varieties across Channel,

- Hotel users are spending more on **Fresh** products
- Retail users are spending normally on all varieties but we can notice maximum spent is on **Grocery**

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?

Statistical Consistency of data can be viewed in many ways and now let's evaluate the consistency of the data by standard error of the mean (i.e., the standard deviation of the sampled population divided by the square root of the sample size)

By performing Measures of Variance, we can know the reliability of all products and we get the below output.

- The coefficient of Variance for Fresh is: 1.05
- The coefficient of Variance for Milk is: 1.27
- The coefficient of Variance for Grocery is: 1.2
- The coefficient of Variance for Frozen is: 1.58
- The coefficient of Variance for Detergents_Paper is: 1.65
- The coefficient of Variance for Delicatessen is: 1.85

Therefore, from the above information we can infer that

- **Fresh** varieties seem to have **the least inconsistent/more consistent behaviour** in terms of spending by the buyers
- **Delicatessen** products possess to have **most inconsistent behaviour** in terms of spending by the buyers

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

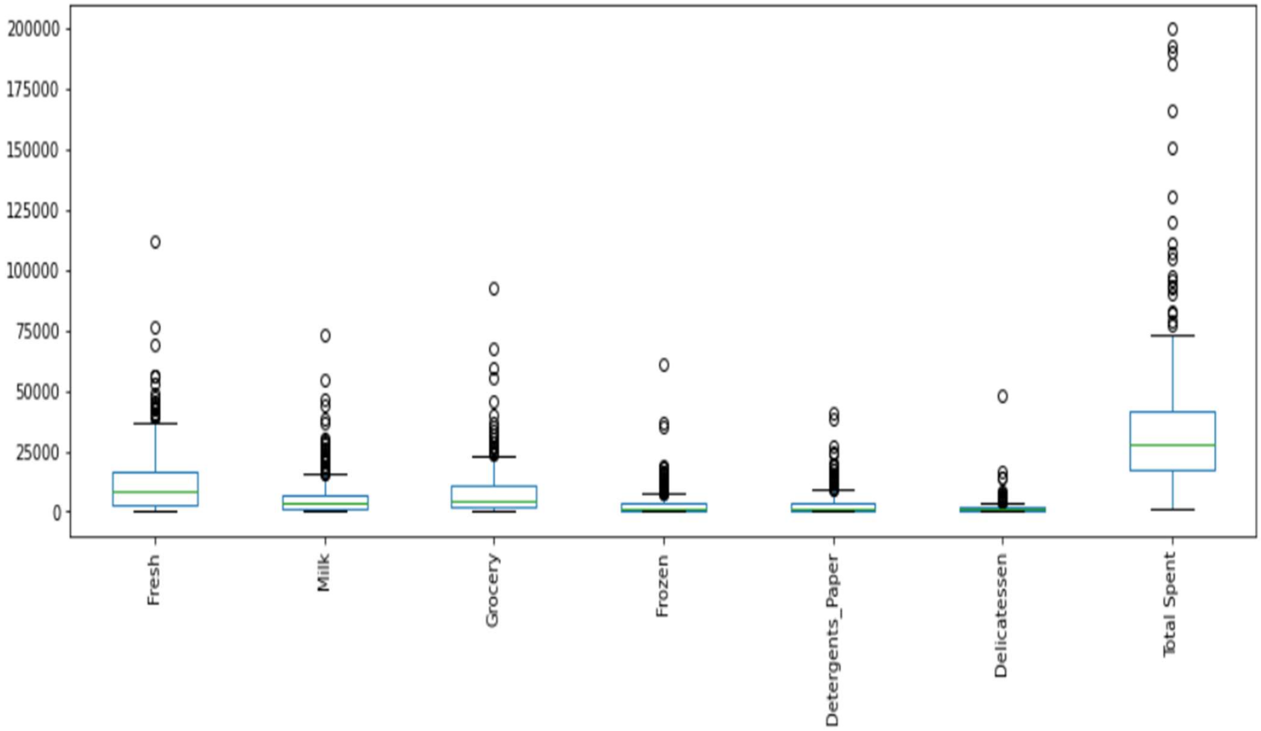


Fig.4

There are multiple ways to figure out outliers in a data like IQR Method, Bar Plotting Method and many more. However, by plotting bar graph we could represent outliers for the given data. From the above plot Fig.4, we are getting to know that outliers are present in all products

(Please ignore total spent in the plot)

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

So far from all the insights, it has been observed from the data there are only two modes of sale it is either from **Retail** or from **Hotel** and it is indeed there are large number of consumers in Other-Region from Hotel Channel, comparatively it is less in retail channel from Lisbon and Oporto as well. So, it is advisable to focus on retail base in Lisbon and Oporto

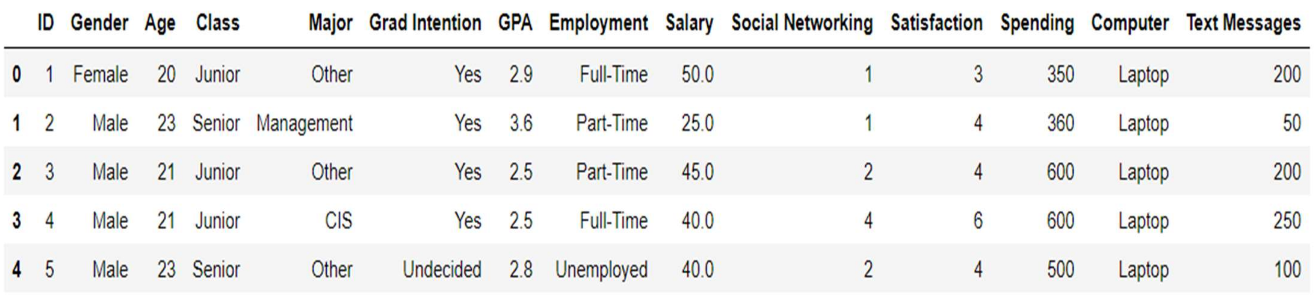
Here, we can also find that **Fresh** products are considerably used more by many retailers among all the regions whereas products like **Frozen, Detergents Paper and Delicatessen** are not strongly admired among the retailer. Therefore, it is recommended that these kinds of unengaging items can may be maximised where there is more demand with new multiple selling methods could be implemented to expand the customer base in all region and reach out to them.

Student Survey

Introduction

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.

Exploratory Data Analysis



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

Table.4 Data Sample

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 62 entries, 0 to 61
Data columns (total 14 columns):
#      Column               Non-Null Count      Dtype
---    -------              -----------         -----
0      ID                   62 non-null         Int64
1      Gender               62 non-null         Object
2      Age                  62 non-null         Int64
3      Class                62 non-null         Object
4      Major                62 non-null         Object
5      Grad Intention       62 non-null         Object
6      GPA                  62 non-null         float64
7      Employment           62 non-null         Object
8      Salary               62 non-null         float64
9      Social Networking    62 non-null         Int64
10     Satisfaction         62 non-null         Int64
11     Spending             62 non-null         Int64
12     Computer             62 non-null         Object
13     Text Messages        62 non-null         Int64
dtypes: float64(2), int64(6), object(6)
memory usage: 6.9+ KB
```

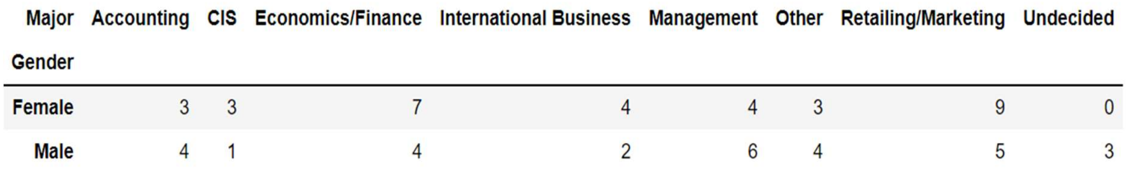
Basically, from the EDA we could understand that

- There are 62 & 14 rows and columns respectively
- There are 3 datatypes namely 2 float integers, 6 integers and 6 object types
- As we can see there are no null values

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

Contingency Table: A table showing the distribution of one variable in rows and another in columns, used to study the correlation between the two variables.

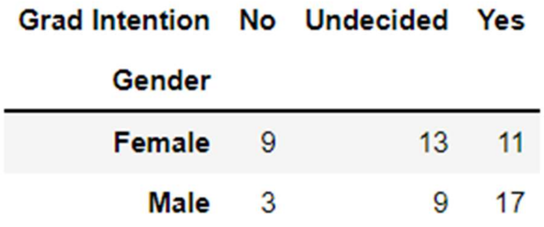
2.1.1. Gender and Major



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

Table.5 Contingency Table: Gender Vs Major

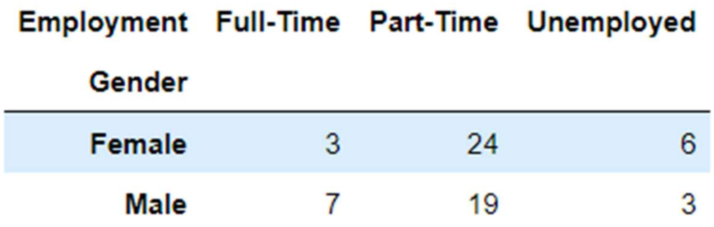
2.1.2. Gender and Grad Intention



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

Table 6. Contingency Table: Gender Vs Intention

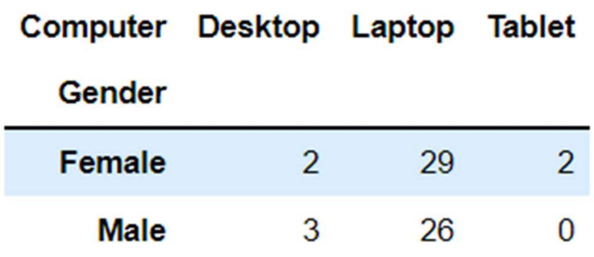
2.1.3. Gender and Employment



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

Table 7. Contingency Table: Gender Vs Employment

2.1.4. Gender and Computer



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

Table 8. Contingency Table: 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?

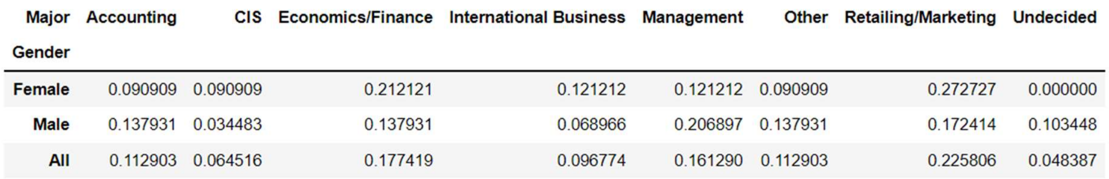
Total number of males(M): 29 P (selected is male) =(M/T)*100=> (29/62)*100 = 46.77 %
Total number of males (F): 33
Total value of Gender (T): 62
The probability that a randomly selected CMSU student will be male is approximately **46.7742 %**

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

Total number of males (M): 29 P (selected is female)=(F/T)*100=> (33/62)*100 = 53.23 %
Total number of males (F) : 33
Total value of Gender (T) : 62
The probability that a randomly selected CMSU student will be male is approximately **53.2258 %**

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.



Major	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided
Gender								
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

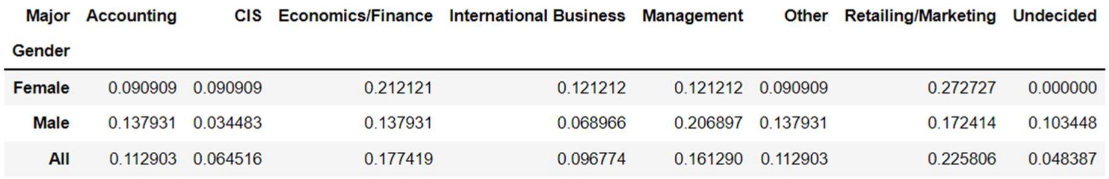
Table 9 Conditional Probability: Gender Vs Major

Conditional Probability of different majors among the male students P(Different Majors/Male) in CMSU can be seen from the above table

The snippet shows the probability of male choosing different majors

- **20.68%** of male students selected **Management** as their major as stands at the top of the selection list
- Least selection by male students is **International Business** of **6.89 %**

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



Major	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided
Gender								
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

Table 9 Conditional Probability: Gender Vs Major

Conditional Probability of different majors among the female students P(Different Majors/Female) in CMSU can be seen from the above table

The snippet shows the probability of male choosing different majors

- **21.21 %** of female students find the highest interest in **Economic/Finance** major
- Female students find least interest in **Accounting, CIS and in Other** majors

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

Table 10. Gender Vs Grad Intention

P(Intends to Graduate/Male) => 17/29 = **0.5862**

The probability that a randomly chosen student is a male and intends to graduate is **58.62%**

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

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

Table 11. Gender Vs Laptop

P(Have laptop/Female) = 29/33 = **0.8787**

But we need to know the probability of female not having a laptop, therefore it shall be

1 minus P(Have laptop/Female)=> 1-0.8787 = **0.1212**

Hence, the probability that a randomly selected student is a female and does not have a laptop is

12.12%

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?

Probability of randomly selected student is male P (M) = **46.77 %**

Probability of randomly selected student has a fulltime employment P (E) = **16.13 %**

Probability of male having a fulltime employment P (M & E) = **11.29 %**

P (M or E) = P(M) + P(E) - P(M&E) => 51.61290 %

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

51.61290 %

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 majoring in international business (IB) = **4**

Total number of females majoring in management (Man) = **4**

Total number of females majoring(M) = **33**

P (IB + Man/Female) = [(IB + Man)/M] * 100 => [(4+4)/33] * 100 = **24.24 %**

Therefore, probability that given a female student is randomly chosen, she is majoring in international business or management is **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	All
Gender			
Male	3	17	20
Female	9	11	20
Total	12	28	40

Grad Intention	No	Yes
Gender		
Male	0.15	0.85
Female	0.45	0.55
Total	0.3	0.7

Table 12.

The probability that a randomly selected Student is Female **50.0 %**

[(20/40) *100]

The probability that a randomly selected student is female and intends to graduate is **55 %**

[(11/20) * 100]

They are not independent events because the probability changes depending on female having graduate intention hence, they are dependent 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?

No. of students that his/her GPA is less than 3 = **17**

Total value = **62**

P (S<3) = (17/62) * 100 => **23.42 %**

Hence, if a student is chosen randomly, the probability that his/her GPA is less than 3 is **27.42 %**

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.

Gender	More than 50	Total	CP
Male	14	29	0.482759
Female	18	33	0.545455
Total	32	62	

Table 13

From the data we know that,

- Total No. of males (M) who earns more than 50 = **14**
- Total No. of males = **29**
- Total No. of females(F) who earns more than 50 = **18**
- Total No. of females = **33**

P (>= 50/M) => 14/29 = **0.4827**

Probability that a randomly selected male earns 50 or more is **48.27%**

P (>=50/F) => 18/33 = **0.5454**

Probability that a randomly selected female earns 50 or more is **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.

The probability plot is a graphical technique for assessing whether the dataset follows any given distribution, the data are plotted against a theoretical distribution examines in such a way that points should form a straight line.

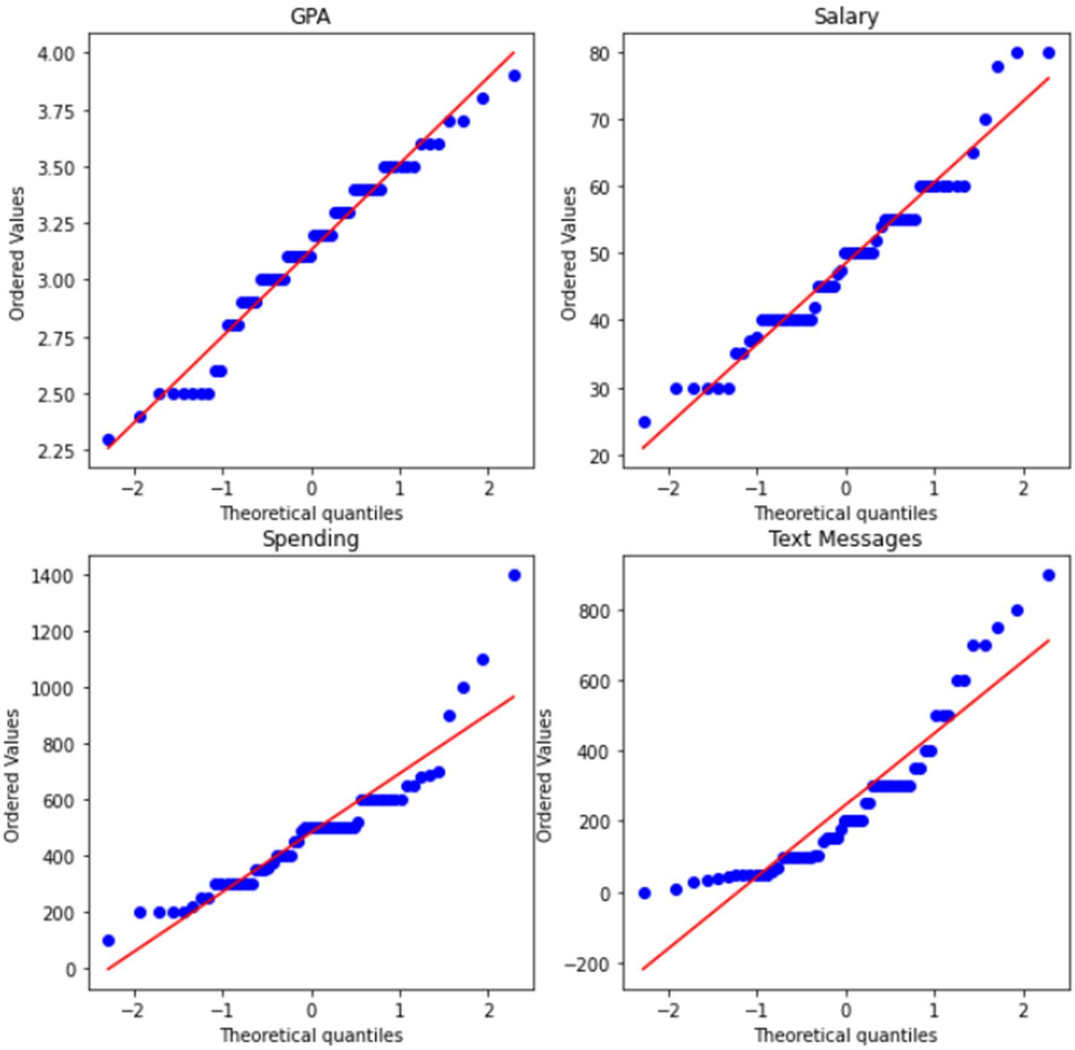


Fig 5

From the above snippet if we examine and we can evaluate that all the four points follow a straight line and we can say that all the GPA, salary, spending and text messages follow a normal distribution.

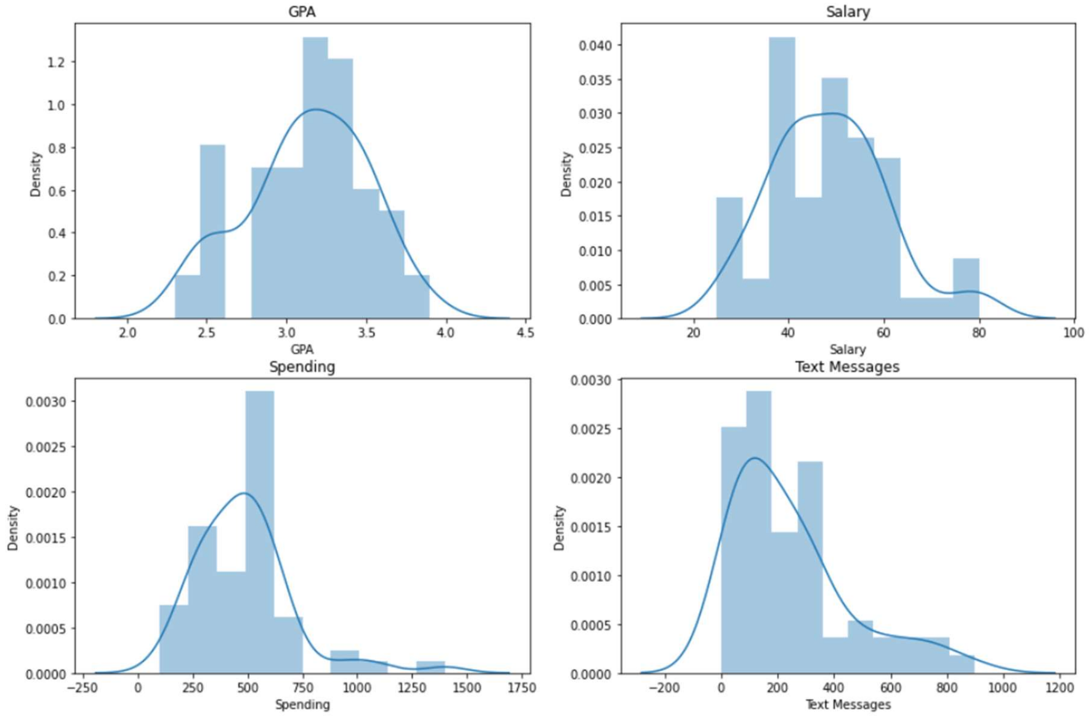


Fig 6.

Skewness is used along with kurtosis to better judge the likelihood of events falling in the tails of a probability distribution and it can be used to obtain approximate probabilities and quantiles of distributions

For the above snippet, the skewness values are as follows

- skew value of GPA is -**0.3146**
- skew value of Salary is **0.5347**
- skew value of Spending is **1.5859**
- skew value of Text Message is **1.2958**

Conclusion

From the survey that has been made about under graduate students who attend CMSU out of the maximum students we have got 62 respondents. Among the 62 respondents where male and female respondents were 29 and 33 respectively.

Out of all the majors the most preferred by the students is Retailing/Marketing, and students were opting part time employment more compared to full time employment and the dataset follows the normal distribution with average GPA with **3.1**.

Shingles

Introduction

An important qualified 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 are calculated. The company would like to show that the mean moisture content is less than 0.35 pounds per 100 square feet.

Exploratory Data Analysis

We have two variable A and B

```
 <class 'pandas.core.frame.DataFrame'>
 RangeIndex: 36 entries, 0 to 35
 Data columns (total 2 columns):
 #   Column       Non Null Count           Dtype
 0   A            36 non-null              float64
 1   B            31 non-null              float64
 dtypes: float64(2)
 memory usage: 704.0 bytes
```

From the above information it can be seen that null values are present in column B.

	Count	Mean	Std	Min	25%	50%	75%	Max
A	36	0.316667	0.135731	0.13	0.2075	0.29	0.3925	0.72
B	31	0.273548	0.137296	0.1	0.16	0.23	0.4	0.58

Table No 14. Statistical Summary

From the table we can infer

- Mean of A and B is 0.3166 and 0.2735 respectively
- Standard Deviation of A and B is 0.13 and 0.1 respectively

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.

Observing the problem,

- We can know that there are two **independent** samples of shingles A & B
- Population **Standard Deviation** is **unknown**
- We cannot perform Z-Test
- Let's proceed with one tailed left side **T-Test**

Subsequently, we have to figure out mean moisture contents is within the permissible limit for both samples, we have to perform T-Test for both the samples step wise;

FOR SAMPLE A

STEP 1: - DEFINING NULL AND ALTERNATE HYPOTHESIS

The null hypothesis(H0) states that the moisture content of sample A is greater or than equal to the permissible limit, $\mu \geq$ **0.35**

The alternative hypothesis(H1) states that the moisture content of sample A is less than permissible limit, $\mu <$ **0.35**

- H_0: $\mu \geq$ **0.35**
- H_1: $\mu <$ **0.35**

STEP 2: - DECIDING THE SIGNIFICANCE LIMIT

- Since alpha value is not given in the question, generally we assume alpha(α) = **0.05**

STEP 3: - IDENTIFY THE TEST STATISTIC

- We have sample A and we do not know the population standard deviation.
- Sample size **n=36.**
- We use the t distribution and the tSTAT test statistic for one sample t-test.

STEP 4: - CALCULATE THE P-VALUE AND TEST STATISTIC

- **t = x̄ - μ / (s/ √n)** x̄ = 0.316667; Std = 0.135731; N=36; μ = 0.35

By applying we get,

- Tstat = -**1.4735**
- P-Value = **0.0747** **α=0.05**

STEP 5: - DECIDE TO REJECT OR ACCEPT NULL HYPOTHESIS

- Since p_value is greater than alpha, we fail to reject the null hypothesis and the results are insignificant
- We conclude that the moisture content is greater than permissible limit in sample A.

Concluding that the Sample A's the mean moisture content is greater than 0.35 pounds per 100 square feet and it is evident that Sample A of Shingles has more mean moisture content than the permissible limit

FOR SAMPLE B

STEP 1: - DEFINE NULL AND ALTERNATE HYPOTHESIS

The null hypothesis states that the moisture content of sample B is greater or than equal to the permissible limit, $\mu \geq 0.35$
The alternative hypothesis states that the moisture content of sample B is less than permissible limit, $\mu < 0.35$
It shall be one tail left side T - test

- H_0: $\mu \geq$ **0.35**
- H_1: $\mu <$ **0.35**

STEP 2: - DECIDE THE SIGNIFICANCE LIMIT

- Since alpha value is not given in the question, we assume it has alpha = **0.05**

STEP 3: - IDENTIFY THE TEST STATISTIC

- We have sample A and we do not know the population standard deviation. Sample size **n=31.**
- We use the t distribution and the $tSTAT$ test statistic for one sample t-test.

STEP 4: - CALCULATE THE P - VALUE AND TEST STATISTIC

- $t = \bar{x} - \mu / (s/ \sqrt{n})$ $\bar{x} = 0.2735$; Std = 0.1372; N=31; $\mu = 0.35$

By applying we get,

- Tstat = -**3.1003**
- P-Value = **0.0020** **α=0.05**

STEP 5: - DECIDE TO REJECT OR ACCEPT NULL HYPOTHESIS

- Since p_value is lesser than alpha, we reject the null hypothesis and the results are significant
- We conclude that the moisture content is less than permissible limit in sample B

Concluding that the Sample B's mean moisture content is lesser than 0.35 pounds per 100 square feet it is evident that Sample B of Shingles has mean moisture with in the permissible limit

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?

Before getting into solving let us look into few points,

- We know that there are two samples
- More significantly population deviation is not known
- Sample size is more than 30, but sample size for both samples are not same
- Henceforth we can take up T-Distribution and the tSTAT test statistic for two sample
- It shall be a Two Tailed Test
- Also, let's assume significance limit alpha = **0.05(Step 2)**

STEP 1: - DEFINING NULL AND ALTERNATIVE HYPOTHESIS

- **$H_0 = \mu A \neq \mu B$**
- **$H_1 = \mu A = \mu B$**

STEP 2: - DECIDE THE SIGNIFICANCE LIMIT

- Since alpha value is not given in the question, we assume it has alpha = **0.05**

STEP 3: - CALCULATION OF P-VALUE AND TEST STATISTIC

- **$t = (\bar{x}\ 1 - \bar{x}\ 2)/ \surd(s12n1+s22n2)$** Degrees of freedom (DF)=n_1+n_2-2
- $\bar{x}_1 = 0.32$ $\bar{x}_2=0.027$
- $s_1^2 = 0.02$ $s_2^2=0.02$
- $n_1 = 36$ $n_2= 31$
- $DF_1= 35$ $DF_2 = 30$

By applying we get,

- Tstat = **1.2896282719661123**
- P Value = **0.2017496571835306** **α=0.05**

STEP 4: - DECIDE TO REJECT OR ACCEPT THE NULL HYPOTHESIS

- Since p_value is greater than alpha, we fail to reject the hypothesis we conclude that the result is insignificant
- We conclude that mean for shingles A and singles B are not the same.

Concluding that the Mean sample of A is not equal to Mean sample of B