SMDM PROJECT REPORT

Contents

Problem 1.....4

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? 4
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.....7
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?.....13
1.4. Are there any outliers in the data? Back up your answer with a suitable plot/technique with the help of detailed comments.....13
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.....14

Problem 2.....3
2.1. For this data, construct the following contingency tables(Keep Gender as row variable).....15
2.1.1. Gender and Major.....15
2.1.2. Gender and Grad Intention.....15
2.1.3. Gender and Employment.....15
2.1.4. Gender and Computer 16
2.2. Assume that the sample is a representative of the population of CMSU. Based on the data, answer the following questions:.....16
2.2.1. What is the probability that a randomly selected CMSU student will be male?.....16
2.2.2. What is the probability that a randomly selected CMSU student will be female? 16
2.3. Assume that the sample is representative of the population of CMSU. Based on the data, answer the following question:.....16
2.3.1. Find the conditional probability of different majors among the male students in CMSU...17
2.3.2 Find the conditional probability of different majors among the female students of CMSU…….....…..17
2.4. Assume that the sample is a representative of the population of CMSU. Based on the data, answer the following question:.....18
2.4.1. Find the probability That a randomly chosen student is a male and intends to graduate..18
2.4.2 Find the probability that a randomly selected student is a female and does NOT have a laptop…….....…19
2.5. Assume that the sample is representative of the population of CMSU. Based on the data, answer the following question:.....19
2.5.1. Find the probability that a randomly chosen student is a male or has full-time employment?.....19
2.5.2. Find the conditional probability that given a female student is randomly chosen, she is majoring in international business or management.……....19

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?.....20
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……..….20
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…….....21
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……..…….21

Problem 3.....23
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……..……..23
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?.....24

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).

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?

	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
count	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
std	127.161315	12647.328865	7380.377175	9503.162829	4854.673333	4767.854448	2820.105937
min	1.000000	3.000000	55.000000	3.000000	25.000000	3.000000	3.000000
25%	110.750000	3127.750000	1533.000000	2153.000000	742.250000	256.750000	408.250000
50%	220.500000	8504.000000	3627.000000	4755.500000	1526.000000	816.500000	965.500000
75%	330.250000	16933.750000	7190.250000	10655.750000	3554.250000	3922.000000	1820.250000
max	440.000000	112151.000000	73498.000000	92780.000000	60869.000000	40827.000000	47943.000000

- The above small dataframe gives us the descriptive statistics of all the numerical columns in the wholesale customer dataframe.
- From the above table, we can get the count, mean, standard deviation, minimum value, maximum value, percentile values of the given data.
- When we compare the several items given, we can see the below information:
  Mean value is highest for the item Fresh
  Highest Maximum spending is for the item Fresh

```
<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 data, it is clear that there are 440 rows(0 to 439) and 9 columns.
- All the columns has 440 as the no. of rows and hence we can say that there are no missing values in the data imported.
- The data types in the DataFrame are int and object.

After calculating the total for the given items, below is the head of the DataFrame:

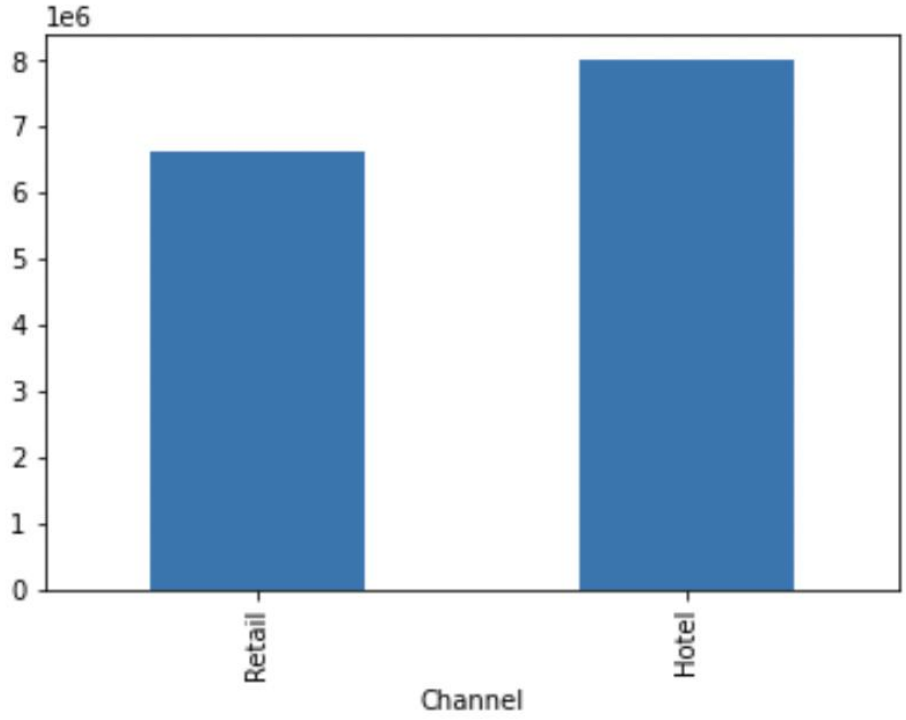
	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen	Total
0	Retail	Other	12669	9656	7561	214	2674	1338	34112
1	Retail	Other	7057	9810	9568	1762	3293	1776	33266
2	Retail	Other	6353	8808	7684	2405	3516	7844	36610
3	Hotel	Other	13265	1196	4221	6404	507	1788	27381
4	Retail	Other	22615	5410	7198	3915	1777	5185	46100

Which Region and which Channel spent the most?
Which Region and which Channel spent the least?

Below is the total amount spent channel wise:

```
Channel
Retail    6619931
Hotel     7999569
Name: Total, dtype: int64
```

Plotting bar plot to visualize the total amount spent channel wise:

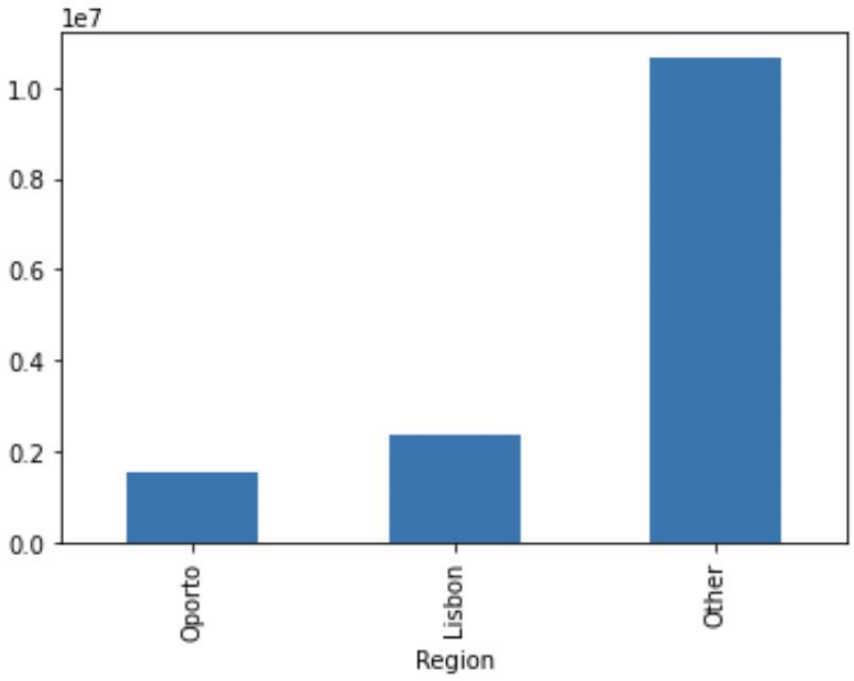


From the above bar plot, it is clear that Hotel channel spent the most and Retail channel spent the least

Below is the total amount spent channel wise:

```
Region
Oporto      1555088
Lisbon      2386813
Other      10677599
Name: Total, dtype: int64
```

Plotting bar plot to visualize the total amount spent region wise:



From the above bar plot, it is clear that Oporto region spent the least and Other region spent the most.

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.

Descriptive statistics of all the different varieties of items across Region and Channel are shown in detail in the below table.(next page)

	Channel	Hotel			Retail		
	Region	Lisbon	Oporto	Other	Lisbon	Oporto	Other
Fresh	count	59.000000	28.000000	211.000000	18.000000	19.000000	105.000000
	mean	12902.254237	11650.535714	13878.052133	5200.000000	7289.789474	9831.504762
	std	12342.008901	8969.362752	14746.572913	5415.521495	6867.934548	9635.394129
	min	514.000000	3.000000	3.000000	18.000000	161.000000	23.000000
	25%	4437.500000	4938.250000	3702.500000	2378.250000	2368.000000	2343.000000
	50%	8656.000000	9787.000000	9612.000000	2926.000000	6468.000000	7362.000000
	75%	18135.000000	17031.500000	18821.000000	5988.000000	9162.000000	15076.000000
	max	56083.000000	32717.000000	112151.000000	20782.000000	27082.000000	44466.000000
Milk	count	59.000000	28.000000	211.000000	18.000000	19.000000	105.000000
	mean	3870.203390	2304.250000	3486.981043	10784.000000	9190.789474	10981.009524
	std	4298.321195	2968.628697	4508.505269	6609.221463	6611.354136	10574.827178
	min	258.000000	333.000000	55.000000	2527.000000	928.000000	1124.000000
	25%	1071.000000	1146.000000	1188.500000	6253.250000	4148.500000	6128.000000
	50%	2280.000000	1560.500000	2247.000000	8866.000000	6817.000000	7845.000000
	75%	4995.500000	2344.750000	4205.000000	13112.250000	13127.500000	11114.000000
	max	23527.000000	16784.000000	43950.000000	28326.000000	25071.000000	73498.000000
Grocery	count	59.000000	28.000000	211.000000	18.000000	19.000000	105.000000
	mean	4026.135593	4395.500000	3886.734597	18471.944444	16326.315789	15953.809524
	std	3629.644143	3048.298815	3593.506056	10414.687844	14035.453775	12298.935356
	min	489.000000	1330.000000	3.000000	5265.000000	2743.000000	4523.000000
	25%	1620.000000	2373.750000	1666.000000	10634.250000	9318.500000	9170.000000
	50%	2576.000000	3352.000000	2642.000000	16106.000000	12469.000000	12121.000000
	75%	5172.500000	5527.500000	4927.500000	23478.750000	19785.500000	19805.000000
	max	16966.000000	13626.000000	21042.000000	39694.000000	67298.000000	92780.000000
Frozen	count	59.000000	28.000000	211.000000	18.000000	19.000000	105.000000
	mean	3127.322034	5745.035714	3656.900474	2584.111111	1540.578947	1513.200000
	std	3276.460124	11454.478518	4956.590848	2424.774577	2473.266471	1504.498737
	min	91.000000	264.000000	25.000000	61.000000	131.000000	33.000000
	25%	966.000000	962.250000	779.000000	923.500000	639.500000	437.000000
	50%	1859.000000	2696.500000	1960.000000	1522.000000	934.000000	1059.000000
	75%	4479.000000	4617.000000	4542.500000	3843.000000	1410.000000	2194.000000
	max	18711.000000	60869.000000	36534.000000	8321.000000	11559.000000	8132.000000
Detergents_Paper	count	59.000000	28.000000	211.000000	18.000000	19.000000	105.000000
	mean	950.525424	482.714286	786.682464	8225.277778	8410.263158	6899.238095
	std	1305.907616	425.310506	1099.970640	5515.878798	8286.748255	6022.091110
	min	5.000000	15.000000	3.000000	788.000000	332.000000	523.000000
	25%	237.000000	182.750000	176.500000	4818.250000	3900.000000	3537.000000

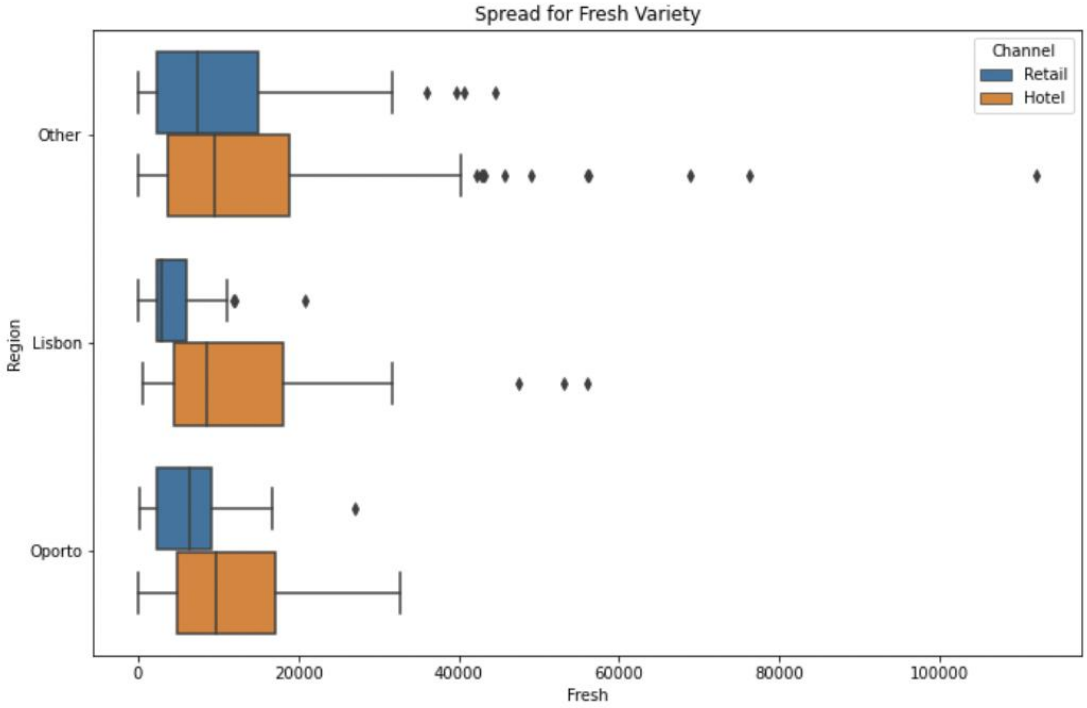
	50%	412.000000	325.000000	375.000000	6177.000000	6236.000000	5121.000000
	75%	874.000000	707.000000	948.500000	11804.750000	9837.500000	7677.000000
	max	5828.000000	1679.000000	6907.000000	19410.000000	38102.000000	40827.000000
Delicatessen	count	59.000000	28.000000	211.000000	18.000000	19.000000	105.000000
	mean	1197.152542	1105.892857	1518.284360	1871.944444	1239.000000	1826.209524
	std	1219.945304	1056.778800	3663.183304	1626.486667	1065.438042	2119.052222
	min	7.000000	51.000000	3.000000	120.000000	59.000000	3.000000
	25%	374.000000	567.250000	378.500000	746.000000	392.500000	545.000000
	50%	749.000000	883.000000	823.000000	1414.000000	1037.000000	1386.000000
	75%	1621.500000	1146.000000	1582.000000	2456.500000	1815.000000	2158.000000
	max	6854.000000	5609.000000	47943.000000	6372.000000	3508.000000	16523.000000

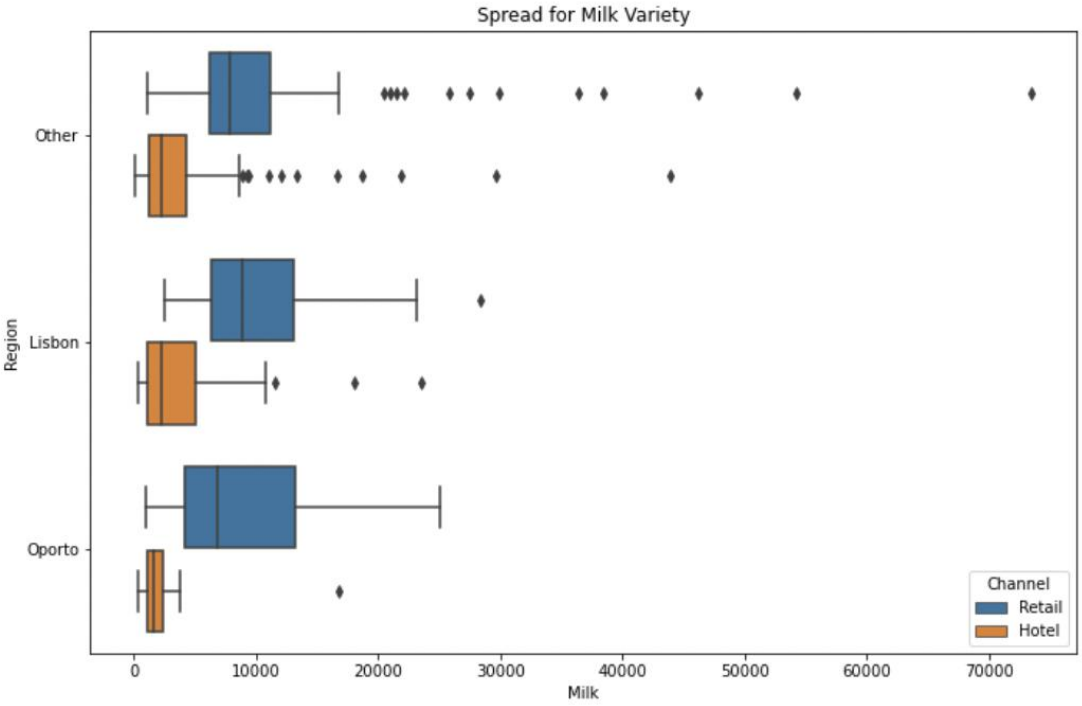
The above table shows us the description of all the varieties across region and channel.

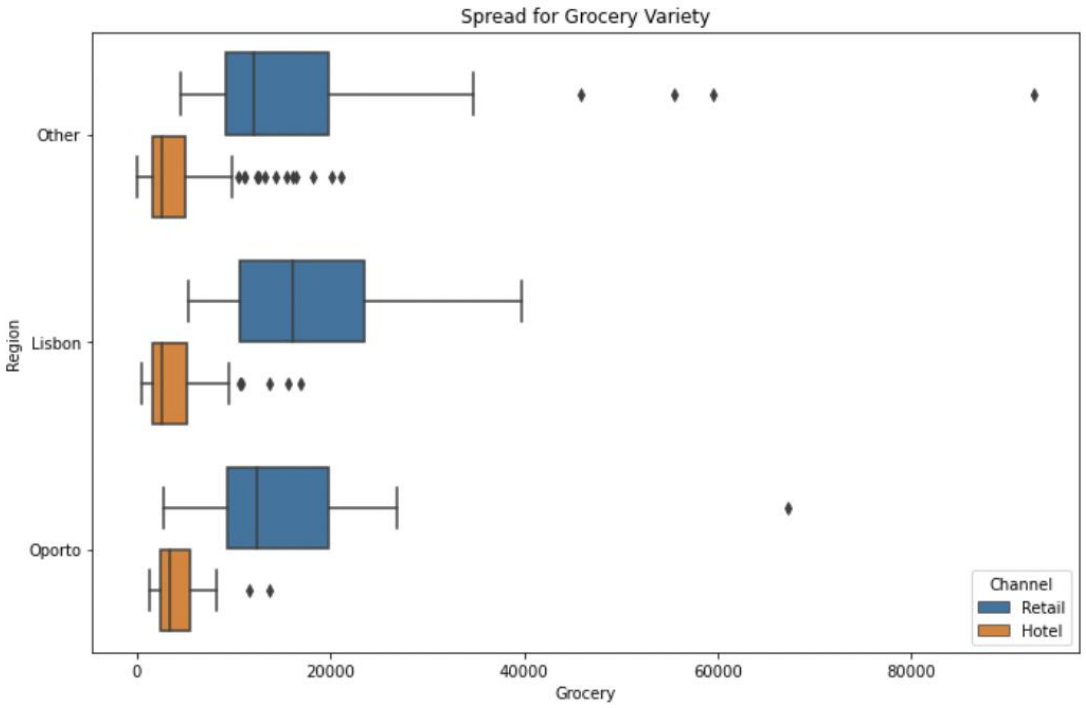
		Delicatessen	Detergents_Paper	Fresh	Frozen	Grocery	Milk
Channel	Region						
Hotel	Lisbon	1197.152542	950.525424	12902.254237	3127.322034	4026.135593	3870.203390
	Oporto	1105.892857	482.714286	11650.535714	5745.035714	4395.500000	2304.250000
	Other	1518.284360	786.682464	13878.052133	3656.900474	3886.734597	3486.981043
Retail	Lisbon	1871.944444	8225.277778	5200.000000	2584.111111	18471.944444	10784.000000
	Oporto	1239.000000	8410.263158	7289.789474	1540.578947	16326.315789	9190.789474
	Other	1826.209524	6899.238095	9831.504762	1513.200000	15953.809524	10981.009524

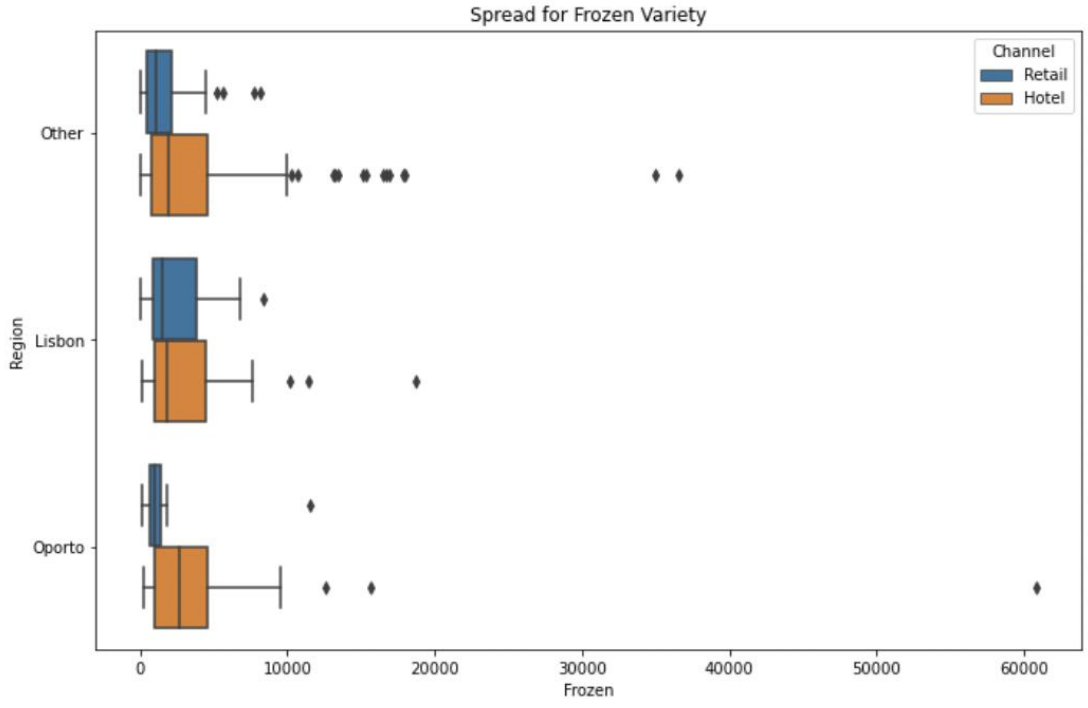
Looking at the above table, we see that some categories like Milk, Grocery & Detergents_Paper have higher spend in the Retail channel when compared with Hotel channel, across all the three regions.
Whereas, Fresh and Frozen have higher spend in the Hotel channel than Retail channel, across all the three regions.

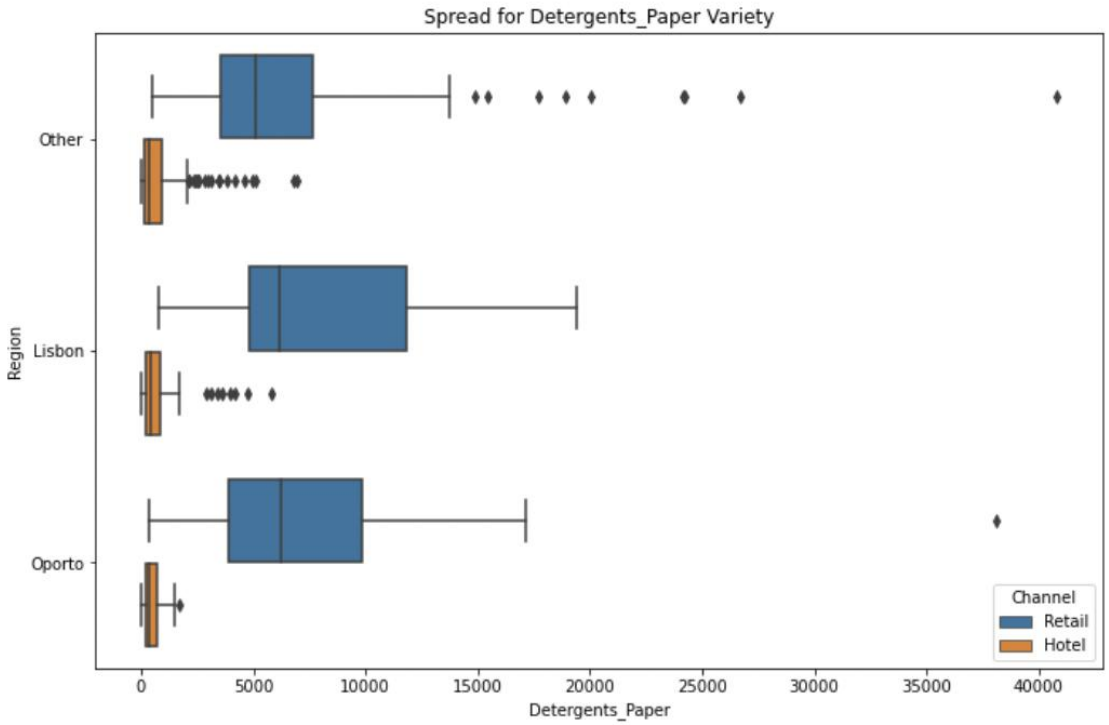
Box Plots for the spread of all the varieties across region and channel:

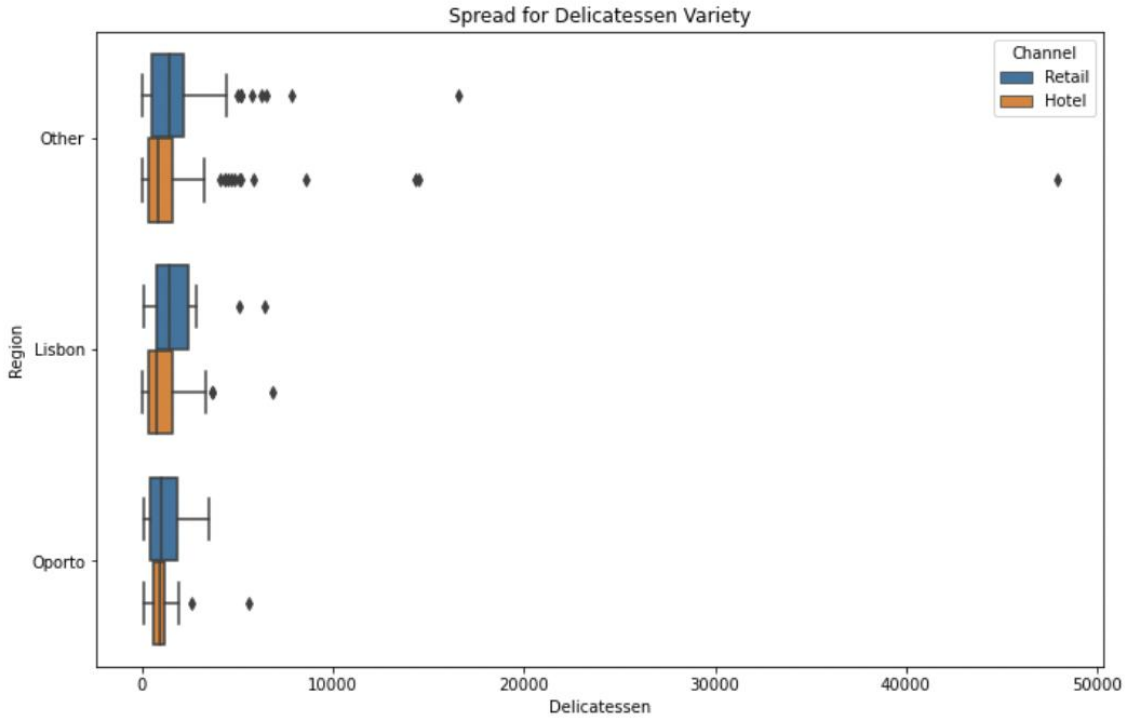





Spread for Grocery Variety
Other
Lisbon
Oporto
Region
Channel
Retail
Hotel
0
20000
40000
60000
80000
Grocery


Spread for Frozen Variety
Other
Lisbon
Oporto
Region
Channel
Retail
Hotel
0
10000
20000
30000
40000
50000
60000
Frozen





From the above box plot we can summarize that the spend for Fresh and groceries is the maximum across region and channel while for Delicatessen it is the least across region and channel.

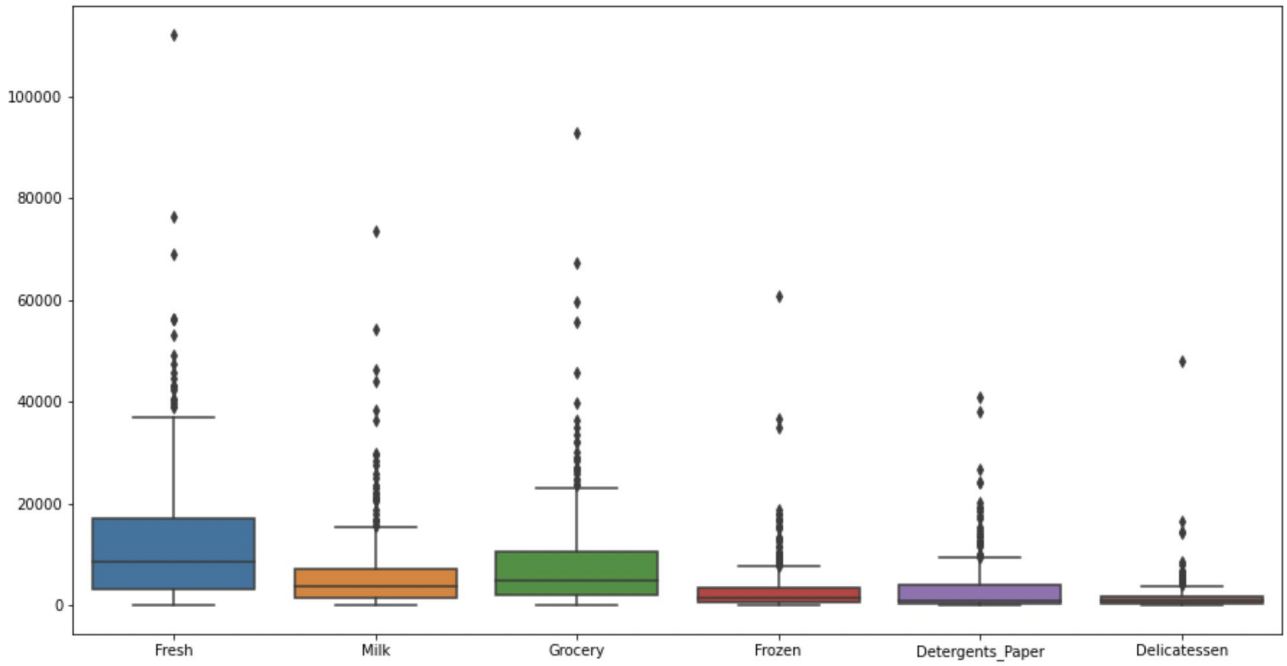
1.3. On the basis of the descriptive measure of variability, which item shows the most inconsistent behaviour?
Which items shows the least inconsistent behaviour?

Coefficient of Variation = standard deviation/mean. Below are the calculated CV for all the varieties

```
Coefficient of Variation for Fresh is : 1.05392
Coefficient of Variation for Milk is : 1.2733
Coefficient of Variation for Grocery is : 1.19517
Coefficient of Variation for Frozen is : 1.58033
Coefficient of Variation for Detergents_Paper is : 1.65465
Coefficient of Variation for Delicatessen is : 1.84941
```

- The higher the coefficient of variation, greater the level of dispersion around the mean.
- The lower the value of the coefficient of variation, the more precise the estimate.
- By calculating Coefficient of Variation, we know that least value is for Category “Fresh”(1.05) and highest value is for Category “Delicatessen”(1.84)
- From the above data, it is clear that most inconsistent behavior is shown by item 'Delicatessen' and least inconsistent behavior is shown by item 'Fresh'

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



As per the above box plot, we can clearly see that there are outliers in the data.
Outliers are present for all the items in the given dataset.

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.

- As per the above analysis, there are inconsistencies in the amount spent for different items (concluded this by calculating Coefficient of Variation). Inconsistency can be reduced.
- There is a huge difference in spending for Hotel and Retail channels for few items such as Fresh, Milk, Grocery, Detergents_paper. The difference in spending can be reduced.
- For the items 'Fresh' and 'Grocery', there is a large spread for the amount spent.
- Hence, business can also focus on other items in order to increase their spends.
- 'Other' region has more spending compared to 'Oporto' and 'Lisbon'. The business can focus on increasing the spends for retailers in the regions 'Oporto' and 'Lisbon'.

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.csv file).

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

2.1.2. Gender and Grad Intention:

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

2.1.3. Gender and Employment:

Employment Gender	Full-Time	Part-Time	Unemployed
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 a representative of the population of CMSU. Based on the data, answer the following questions:

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

no. of female students: 33
no. of male students: 29
Total no. of students: 62

The probability that a randomly selected CMSU student will be male is = 29/62 = 0.4677

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

no. of female students: 33
no. of male students: 29
Total no. of students: 62

The probability that a randomly selected CMSU student will be female is = 33/62 = 0.5323

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

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

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

no. of male students: 29

The probability of students choosing Accounting among the male students in CMSU = 4/29 = 0.1379

The probability of students choosing CIS among the male students in CMSU = 1/29 = 0.0345

The probability of students choosing Economics/Finance among the male students in CMSU = 4/29 = 0.1379

The probability of students choosing International Business among the male students in CMSU = 2/29 = 0.069

The probability of students choosing Management among the male students in CMSU = 6/29 = 0.2069

The probability of students choosing Other among the male students in CMSU = 4/29 = 0.1379

The probability of students choosing Retailing/Marketing among the male students in CMSU = 5/29 = 0.1724

The probability of students choosing Undecided among the male students in CMSU = 3/29 = 0.1034

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

no. of female students: 33

The probability of students choosing Accounting among the female students in CMSU = 3/33 = 0.0909

The probability of students choosing CIS among the female students in CMSU = 3/33 = 0.0909

The probability of students choosing Economics/Finance among the female students in CMSU = 7/33 = 0.2121

The probability of students choosing International Business among the female students in CMSU = 4/33 = 0.1212

The probability of students choosing Management among the female students in CMSU = 4/33 = 0.1212

The probability of students choosing Other among the female students in CMSU = 3/33 = 0.0909

The probability of students choosing Retailing/Marketing among the female students in CMSU = 9/33 = 0.2727

The probability of students choosing Undecided among the female students in CMSU = 0/33 = 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
Gender			
Female	9	13	11
Male	3	9	17

Total number of students : 62
Probability that a randomly chosen student is male and intends to graduate = 17/62 = 0.2742

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

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

Total number of students : 62
Probability that a randomly selected student is female and does not have laptop = 4/62 = 0.0645

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
Gender			
Female	3	24	6
Male	7	19	3

Probability that a randomly chosen student is a male or has full-time employment = ((7+19+3)/62)+((3+7)/62)-(7/62) = 0.5161

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

number of female students : 33
Probability that given a female student is randomly chosen, she is majoring in international business or management = 8/33 = 0.2424

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

For graduate intention and being female to be independent events, it should satisfy the condition, Prob(Female and Yes) = P(Female) * P(Yes)

- probability of being female and grad intention to be yes [Prob(Female and Yes)] = 11/40 = 0.275
- probability of grad intention to be yes = 28/40 = 0.7
- probability of being a female = 20/40 = 0.5
- P_yes * P_female : 0.35

1. 0.275 is not equal to 0.35
2. Prob(Female and Yes) is not equal to P(Female) * P(Yes)

Hence, 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?

number of students with GPA less than 3 is = 17
Total number of students = 62
Probability that his/her GPA is less than 3 when a student chosen randomly = 17/62 = 0.2742

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.

number of students whose salary is 50 or more:

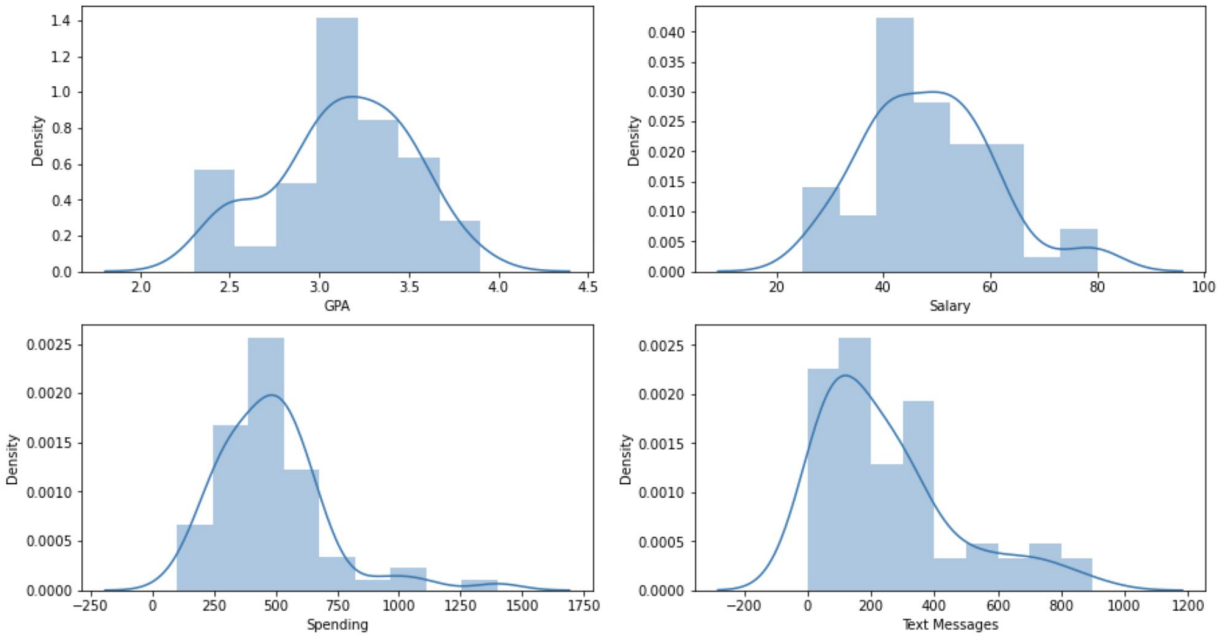
Female 18
Male 14

total number of students:

Female 33
Male 29

probability that a randomly selected male earns 50 or more = 14/29 0.4828
probability that a randomly selected female earns 50 or more = 18/33 = 0.5455

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 above kde and distribution plot shows us the distribution of all the 4 variables in the dataset. To calculate if the variables are normally distributed or not, we are checking the empirical rule below for all the variables individually:

For GPA:

mean of gpa : 3.129032258064516
standard deviation of gpa : 0.3773883926969118

Fraction of values within one standard deviation for GPA : 0.683
Fraction of values within two standard deviation for GPA : 0.954
Fraction of values within three standard deviation for GPA : 0.997

For Salary:

mean of salary : 48.54838709677419
standard deviation of salary : 12.080912216337277

Fraction of values within one standard deviation for Salary : 0.683
Fraction of values within two standard deviation for Salary : 0.954
Fraction of values within three standard deviation for Salary : 0.997

For Spending:

mean of spending : 482.01612903225805
standard deviation of spending : 221.95380496596204

Fraction of values within one standard deviation for Spending : 0.683
Fraction of values within two standard deviation for Spending : 0.954
Fraction of values within three standard deviation for Spending : 0.997

For Text Messages:

mean of text messages : 246.20967741935485
standard deviation of text messages : 214.4659503026961

Fraction of values within one standard deviation for text messages : 0.683
Fraction of values within two standard deviation for text messages : 0.954
Fraction of values within three standard deviation for text messages : 0.997

Conclusion:
For all the numerical variables, below is observed:

- 68% of the observed values lie between 1 standard deviation around the mean
- 95% of the observed values lie between 2 standard deviations around the mean
- 99.7% of the observed values lie between 3 standard deviation around the mean

Hence, we can say that GPA, Salary, Spending and Text Messages follow normal distribution.

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.

H0: mu <= 0.35
H1: mu > 0.35

A type Shingle:

t statistic value is: -1.4735046253382782
p value divided by 2 is: 0.07477633144907513

pval/2 > 0.05

pval is greater than alpha(assuming 5% significance level)
Failed to reject null hypothesis(H0)
Hence, there is no evidence that the mean moisture contents in type A shingles is greater than 0.35

B type Shingle:

t statistic value is: -3.1003313069986995
p value divided by 2 is: 0.0020904774003191813

pval < 0.05

Since pval is less than alpha(0.05), Reject null hypothesis(H0).
Hence, there is enough evidence to conclude that mean moisture content for Sample B shingles is greater than 0.35.

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
H1: mu_A != mu_B

alpha = 0.05

t statistic value is: 1.289628271966112
p value is: 0.2017496571835328
p_value>alpha(0.05)

p value is greater than alpha(0.05)
Hence, failed to reject H0.
There is enough evidence that the population mean for shingles A and B are equal.