BUSINESS REPORT

ON

SMDM

By Kshitij Nishant

Table of Contents

Contents

List of Figures

Figures

List of Tables

Tables:

Problem Statement 1:

Cold Storage started its operations in Jan 2016. They are in the business of storing Pasteurized Fresh Whole or Skimmed Milk, Sweet Cream, Flavoured Milk Drinks. To ensure that there is no change of texture, body appearance, separation of fats the optimal temperature to be maintained is between 2º - 4ºC.
In the first year of business, they outsourced the plant maintenance work to a professional company with stiff penalty clauses. It was agreed that if it was statistically proven that the probability of temperature going outside the 2º - 4º C during the one-year contract was above 2.5% and less than 5% then the penalty would be 10% of AMC (annual maintenance case). In case it exceeded 5% then the penalty would be 25% of the AMC fee. The average temperature data at the date level is given in the file : “Cold_Storage_Temp_Data_.csv”

I preformed EDA on the first dataset:

Head of dataset 1:

Season	Month	Date	Temperature
Winter	Jan	1	2.3
Winter	Jan	2	2.2
Winter	Jan	3	2.4
Winter	Jan	4	2.8
Winter	Jan	5	2.5

(Table 1)

Check for null values for 1st dataset:

Season 0
Month 0
Date 0
Temperature 0

(Table 2)

Description of dataset 1:

	count	mean	std	min	25%	50%	75%	max
Date	365.0	15.720548	8.808321	1.0	8.0	16.0	23.0	31.0
Temperature	365.0	3.002466	0.465832	1.7	2.7	3.0	3.3	4.5

(Table 3)

Information on dataset 1:

Data columns (total 4 columns):

#	Column	Non-Null Count	Dtype
0	Season	365 non-null	object
1	Month	365 non-null	object
2	Date	365 non-null	int64
3	Temperature	365 non-null	float64

Data types: float64(1), int64(1), object(2)

(Table 4)

Range of values Temperature: 2.8
Minimum Temperature: 1.7
Maximum Temperature: 4.5
Median value Temperature: 3.0
Mode value Temperature: 0 3.1

Temperature - 1st Quartile (Q1) is: 2.7
Temperature - 3st Quartile (Q3) is: 3.3

Interquartile range (IQR) of Temperature is 0.5999999999999996

1.Histogram-Temperature

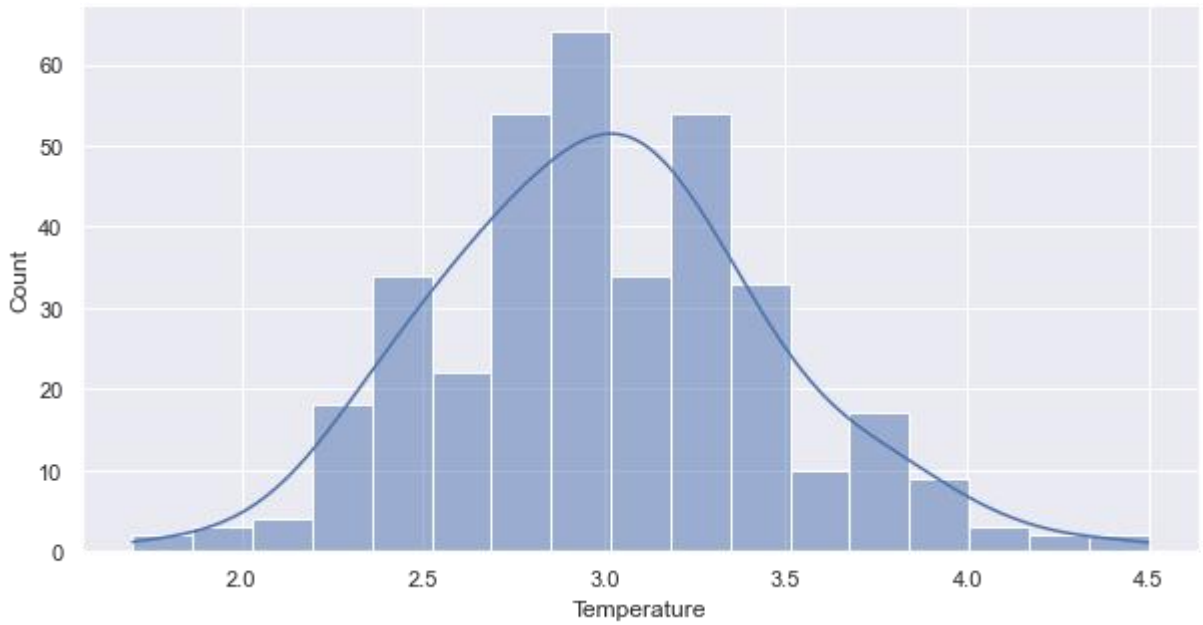

Fig 1

2.BoxPlot- Temperature

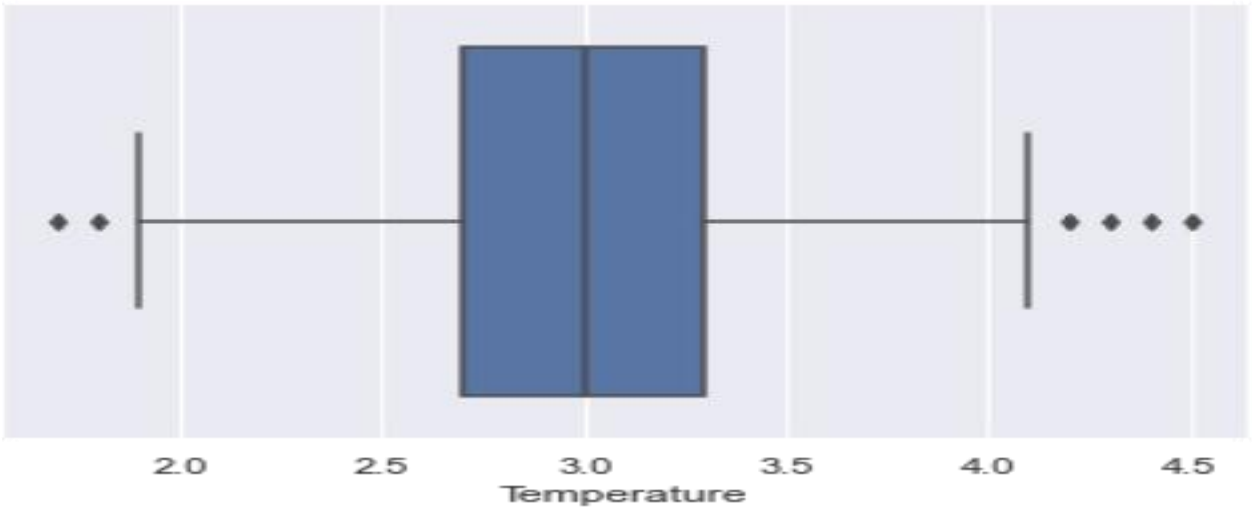

Fig 2

3.BoxPlot - Month v Temperature

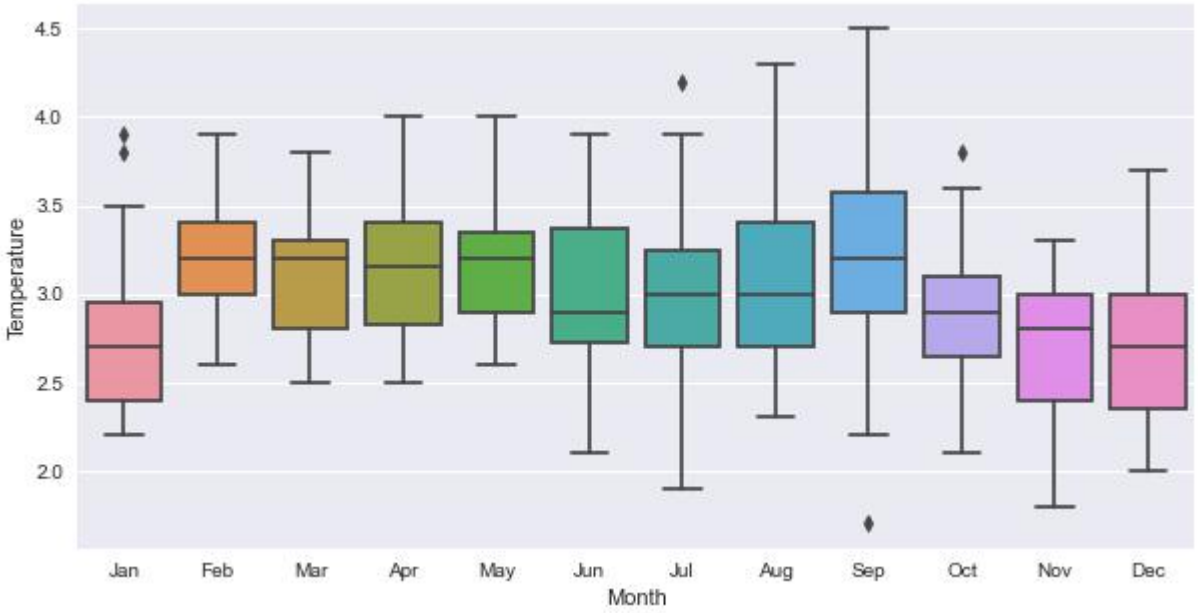

Fig 3

4.BoxPlot - Season v Temperature

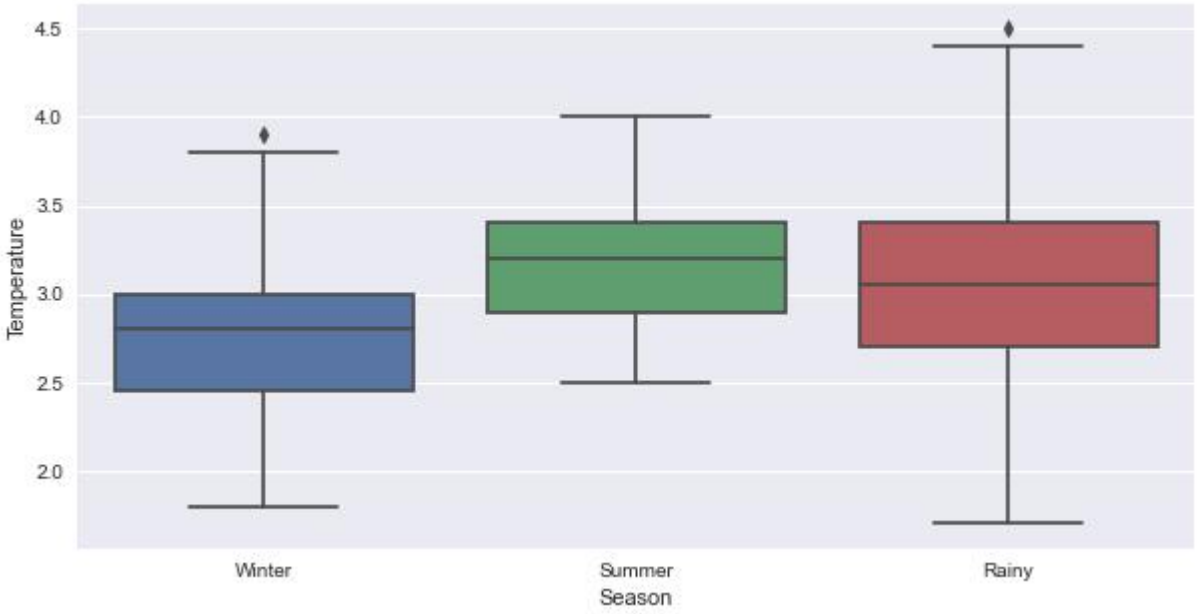

Fig 4

5.PointPlot - Month v Temperature

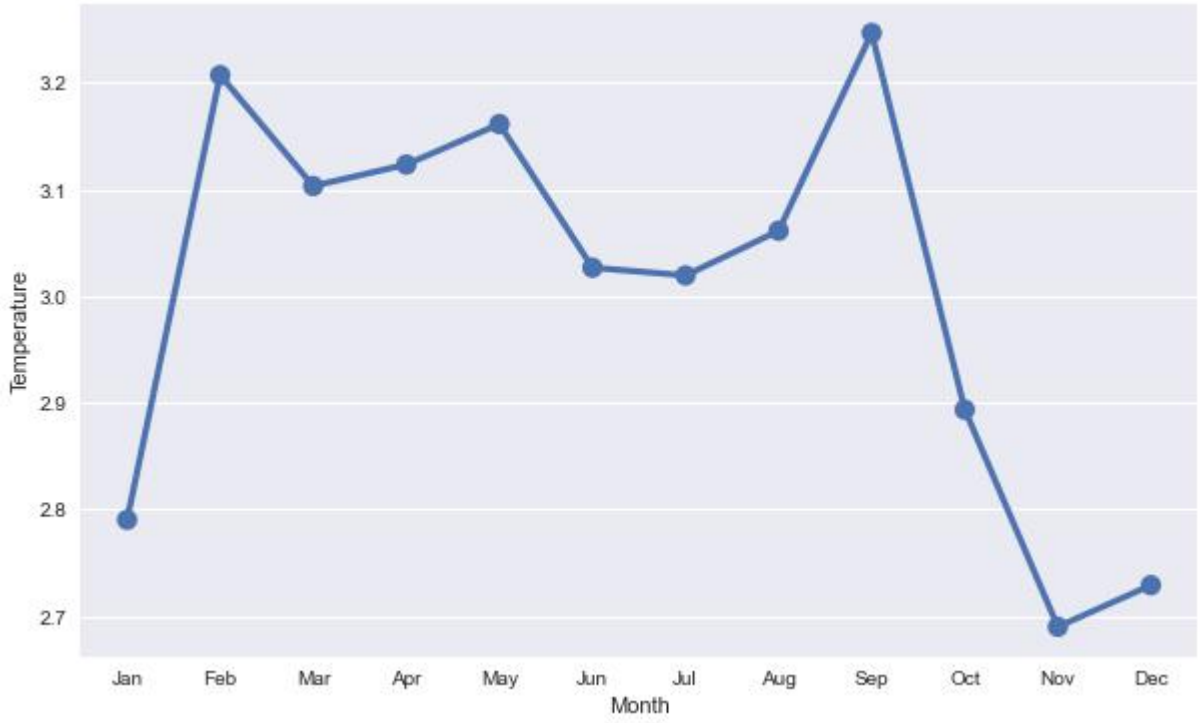

Fig 5

6.CountPlot - Season(hue: Temperature)

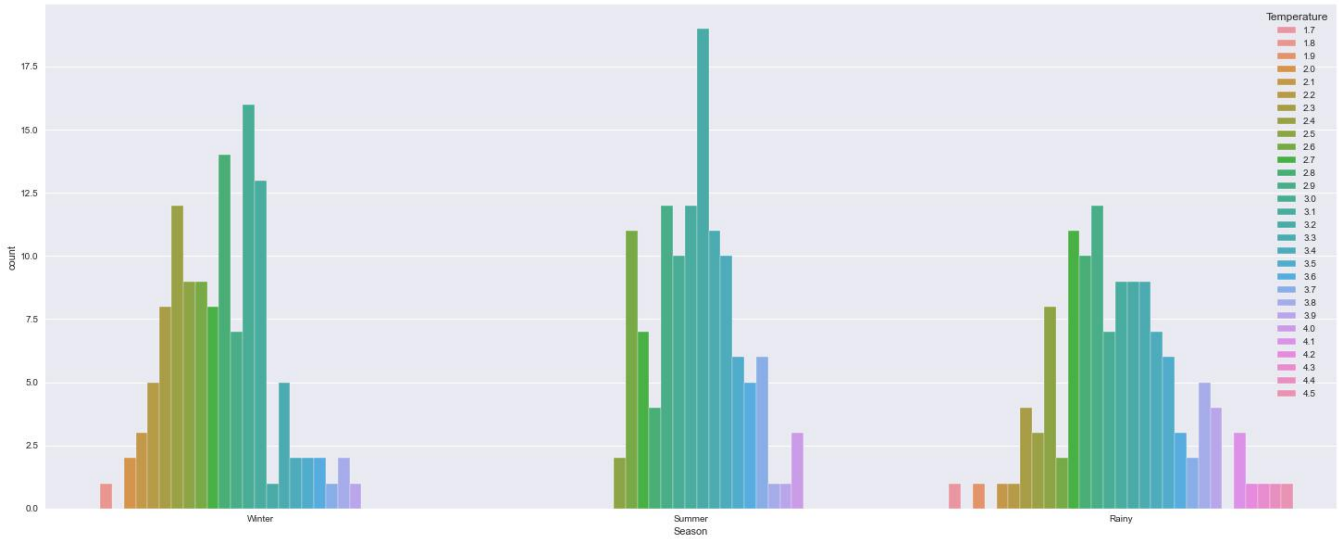

Fig 6

7.CatPlot - Month v Temperature(hue-Season)

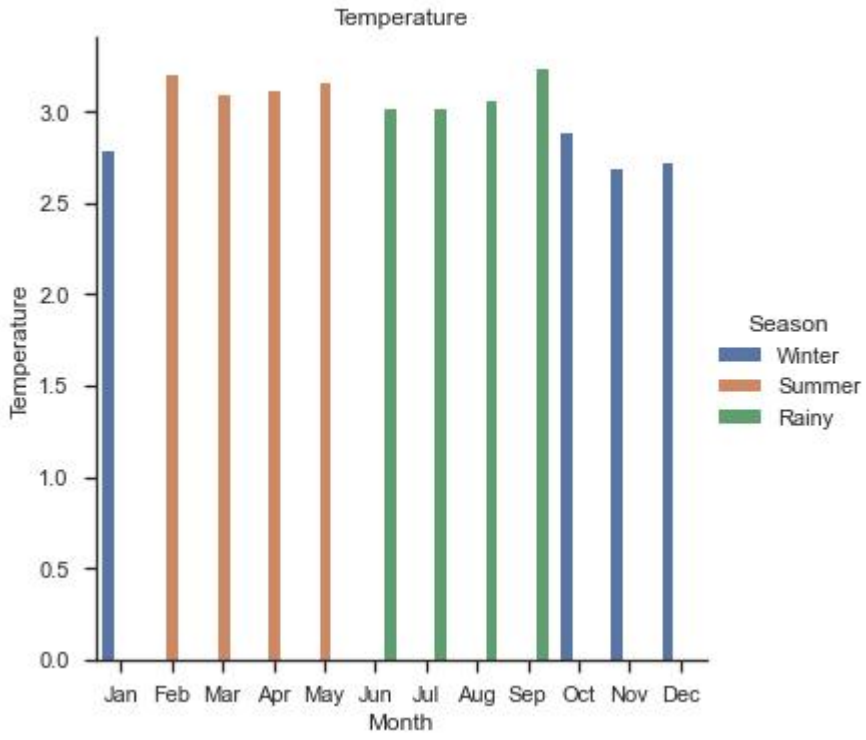

Fig 7

8.Histogram - By Seasons for Temperature

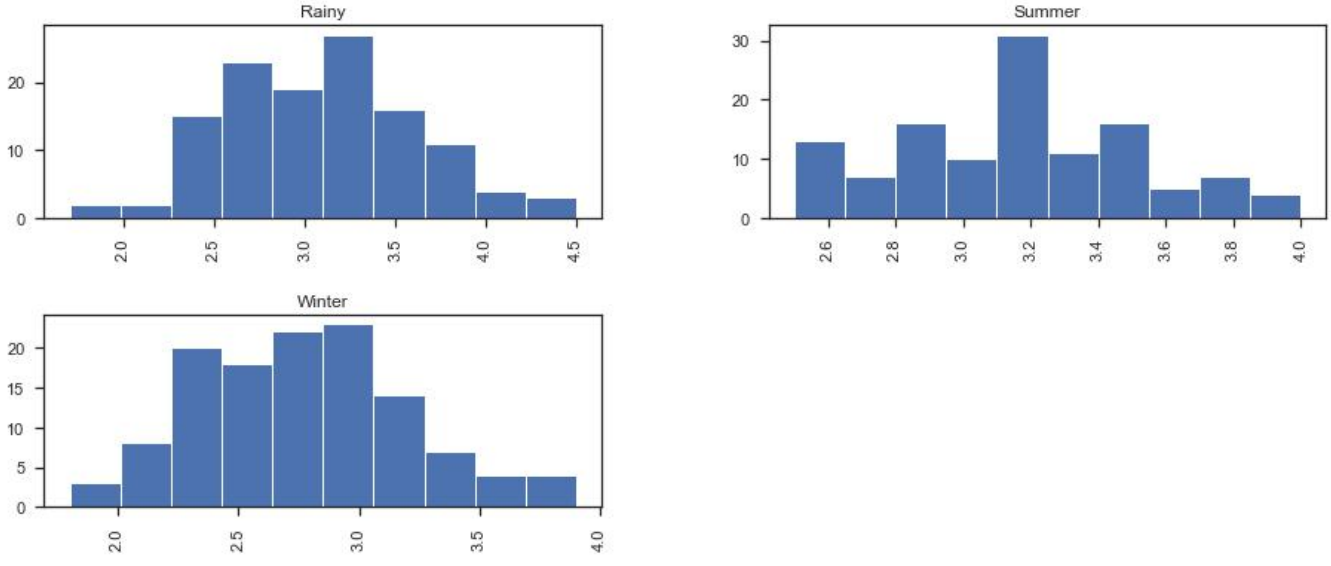

Fig 8

9.Histogram - By Months for Temperature

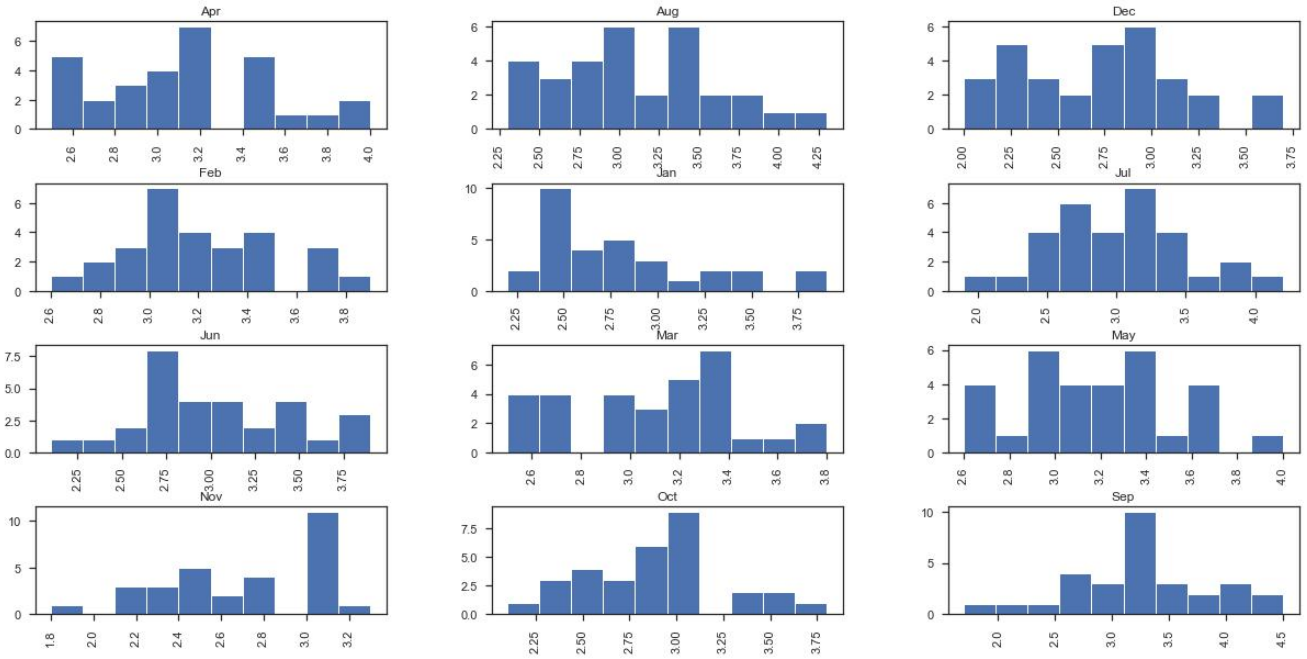

Fig 9

10.ScatterPlot - Month v Temperature

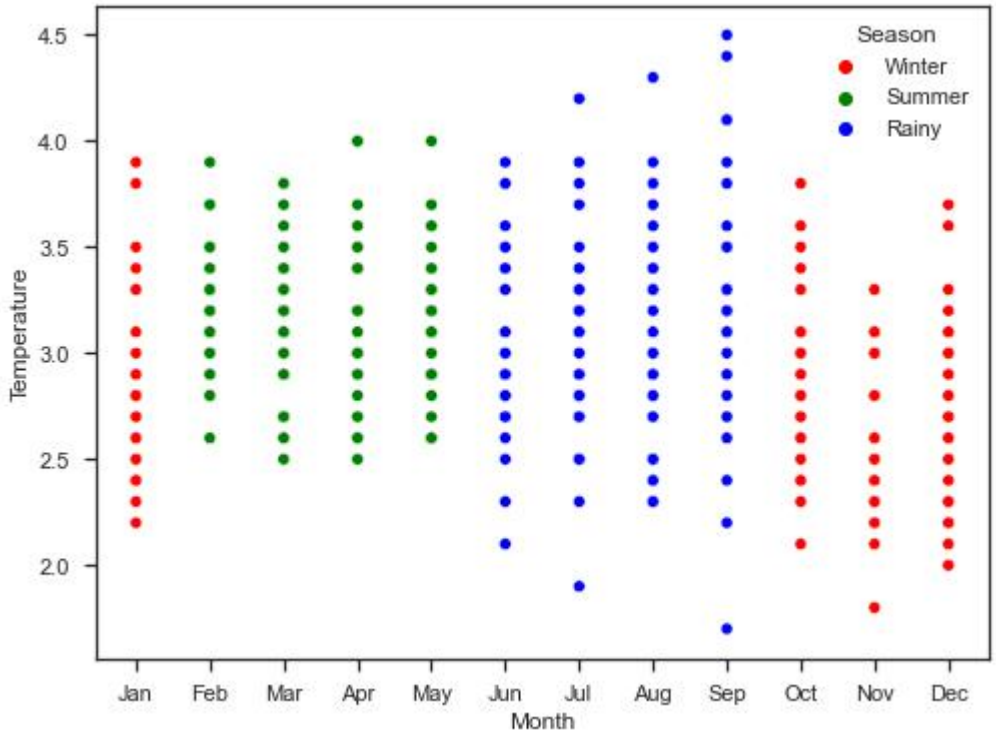

Fig 10

11.Bell Curve - Temperature

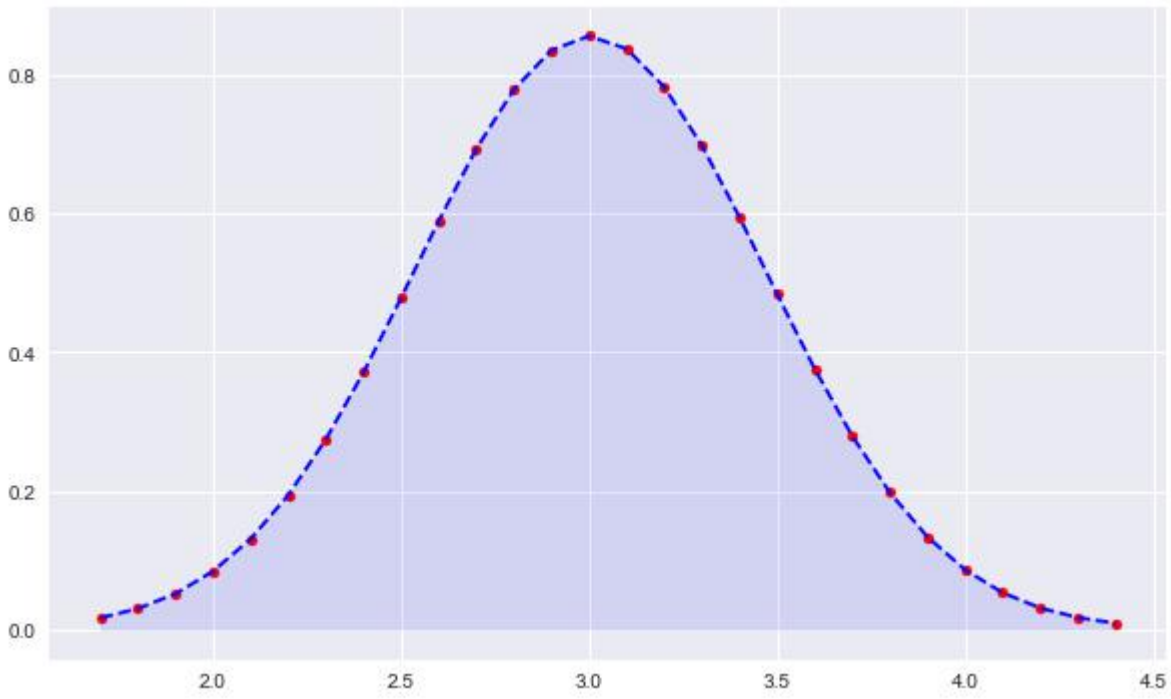

Fig 11

1.1 Find mean cold storage temperature for Summer, Winter, and Rainy Season.

Answer:

Mean temperatures by Seasons are as follows:

Season	Mean Temperature
Rainy	3.087705
Summer	3.147500
Winter	2.776423

(Table 5)

1.2 Find the overall mean for the full year.

Answer:

Overall mean for full year: 3.0024657534246546

1.3 Find Standard Deviation for the full year.

Answer:

Standard Deviation for the full year: 0.4658319416510761 ~ 0.5

1.4 Assume Normal distribution, what is the probability of temperature having fallen below 2º C?

Answer:

Probability of temperature having fallen below 2º C: 2.248518493937247 %

1.5 Assume Normal distribution, what is the probability of temperature having gone above 4º C?

Answer:

Probability of temperature having gone above 4º C: 2.301770505698253 %

1.6 What will be the penalty for the AMC Company?

Answer:

1. As particular Temperature can never attain a value “lower than 2” and “higher than 4” at the same time, therefore these 2 are mutually exclusive events, thus P (A U B) = P(A) + P(B).

2. So, P = P(Temp<2) + P(Temp>4)

3. Therefore, penalty = 10% of AMC, since the probability of Temps going outside of the range of 2 – 4 C falls between the 2.5% and 5% boundary as mentioned in the problem statement.

4. probability of Temps going outside of the range of 2º C – 4º C falls between the 2.5% and 5% boundary: 4.5502889996355 %

End of 1st Problem Statement

Problem 2

In Mar 2018, Cold Storage started getting complaints from their clients that they have been getting complaints from end consumers of the dairy products going sour and often smelling. On getting these complaints, the supervisor pulls out data of the last 35 days' temperatures. As a safety measure, the Supervisor has been vigilant to maintain the mean temperature 3.9º C or below. Assume 3.9º C as the upper acceptable mean temperature and at alpha = 0.1 do you feel that there is a need for some corrective action in the Cold Storage Plant or is it that the problem is from the procurement side from where Cold Storage is getting the Dairy Products. The data of the last 35 days is in "Cold_Storage_Mar2018_.csv"

Head of dataset 2

Season	Month	Date	Temperature
Summer	Feb	11	4.0
Summer	Feb	12	3.9
Summer	Feb	13	3.9
Summer	Feb	14	4.0
Summer	Feb	15	3.8

(Table 6)

Check for Null Values for 2nd dataset

Season	0
Month	0
Date	0
Temperature	0

(Table 7)

Description of 2nd dataset

	Temperature
count	35.000000
mean	3.974286
std	0.159674
min	3.800000
25%	3.900000
50%	3.900000
75%	4.100000
max	4.600000

(Table 8)

2.1 Which Hypothesis test shall be performed to check if corrective action is needed at the cold storage plant? Justify your answer.

Answer:

A. *Observations*:

1) The dataset 1 contains the temperatures throughout the year of 2016

2) The dataset 2 contains the temperatures from a sample of 35 days in 2018 – 02/11 to 03/17 that was procured on back of customer complaints

3) From this, we can infer, the dataset 2 is not a sample of the dataset 1.

4) Therefore, we cannot get the population mean/standard deviation.

5) As the dataset 2 is independent, and we have not been provided with the population data for the same, cannot deduce the Population standard deviation.

B. *Assumptions*:

1)We are assuming the population mean and standard deviation to be same as Sample mean Mu and Sample Standard deviation Sigma, thus assuming our Sample estimation will be reflective of the reality/population sampling

C. *Approach*:
1) Since the population standard deviation is unknown, the best Statistic test to perform would be the T-statistic Test

2) Since we are talking about potential corrective actions, we intend to be more exhaustive and detail oriented.

2.2 State the Hypothesis and do the necessary calculations to accept or reject the corresponding null hypothesis.

Answer:

• Null Hypothesis, H0 : Mean Temp<=3.9
• Alternate Hypothesis, H1 : Mean Temp >3.9

the p-value of this test to be less than the α value.(0.1)

Assumed Population mean = 3.9
Sample mean:3.974285714285715
Sample Standard Deviation :0.15967403771223335
Sample Size is 35
Degree of freedom: 34
Sampling Error: 0.07428571428571518
Standard Error: 0.026989838468435126

One sample t test:
t statistic: 2.752358609800241
p value: 0.0047111977021322155

Check for t-test using alternative = 'greater' param:

One sample t test:
t statistic: 2.752358609800241
p value: 0.0047111977021322155

Level of significance: 0.10
We have evidence to reject the null hypothesis since p value < Level of significance
Our one-sample t-test p-value= 0.0047111977021322155

Since Pvalue < alpha, the Null Hypothesis is rejected, and Alternative Hypothesis is accepted, thus statistically concluding (via T Test) that the Temperature in the Cold Storage is greater than 3.9^{o}

C with 90% confidence [1 – 0.1], thus causing the products go sour and often smelling.

I found the actual confidence by subtracting the Pvalue from 1.
Actual Confidence = (1 - Pvalue) * 100 = 99.52888%

2.3 Give your inference.

Answer:

We have seen from the dataset 1 that holds the values from the year 2016, the average temperature throughout the year is 3.002 C. However, as months went by, the working quality of the Cold Storage seems to have degraded. And from the samples taken in 2018, without even putting it through any Statistical analyses, we can see a mean temperature to be 3.97 C, which is 0.97 degree higher, and going by the working principle of Cold storages, that does not look good, which is why the complaints of products going sour and smelling kept pouring in. However, we hold our judgement before doing an analysis through Statistical analysis and concluding the result.
To do a root cause analysis via statistical hypothesis procedures, we performed T-statistic, which gave us the following results:

1. Having performed T tests, we have confirmed rejecting the Null Hypothesis of Mean Temp < 3.9º C.

2. With 90% confidence, we can conclude the temperature indeed crossed the permissible limit of 3.9º C.

3. With 99.53% actual confidence, we can conclude the above statement.

4. With only 0.47% confidence, we can conclude that the temperature is equal to or lesser than 3.9º C .

5. Thus, statistically we can conclude that there needs to be corrective measures taken to keep the Cold Storage function properly, and there is no apparent problem (statistically speaking)

from procurement side from where Cold Storage is getting the Dairy Products.

6. We can submit the results to the owner of the Cold Storage and they need to figure out the resolution path, is it the lackadaisical approach in work by the Supervisor or some inherent problems with the machines being used. This, we cannot conclude statistically due to lack of necessary data.

Also, as we have seen earlier, there is ‘almost’ a 5% probability of the temperatures to be outside of the permissible range of 2º - 4º C, thus attracting a hefty fine of 10% AMC, and until immediate necessary measures are taken, it could cross the 5º C mark and attract even a heftier fine of 25% AMC.

End Of Business Report