# Project Cold Storage

# Table of Contents

1 **Project Objective**

1. The objective of the report is to explore the Cold Storage data set ("Cold_Storage_Mar2018") and

("Cold_Storage_Temp_Data") importing the dataset in R and generate insights about the data set. This exploration report will consist of the following:

- Importing the dataset in R
- Understanding the structure of dataset
- Graphical exploration
- Descriptive statistics
- Insights from the dataset

**2.Assumptions:**

1. "Cold_Storage_Temp_Data" -> This dataset declares the summary of all seasons from 2016,365days temperature.
2. History of temperature is distributed equally.
3. We are expecting the null value 3.9-degree c should not be exceed by cold storage as we have maintained the most favourable level at the plant.

## 3.Exploratory Data Analysis – Step by Step Approach:

A Typical Data exploration activity consist of the following steps.

1. Environment Set up and Data Import
2. Variable Identification
3. Univariate Analysis
4. Bi-Variate Analysis
5. Variable Transformation / Feature Creation
6. Feature Exploration

### 3.1 Environment Set up and Data Import

#### 3.1.1 Install necessary Packages and Invoke Libraries

Use this section to install necessary packages and invoke associated libraries. Having all the packages at the same places increases code readability.

#### 3.1.2 Set up working Directory

Setting a working directory on starting of the R session makes importing and exporting data files and code files easier. Basically, working directory is the location/ folder on the PC where you have the data, codes etc. related to the project.
Please refer Appendix A for Source Code.

#### 3.1.3 Read and Import the Dataset

The given dataset is in .csv format. Hence, the command 'read.csv' is used for importing the file.
Please refer Appendix A for Source Code.

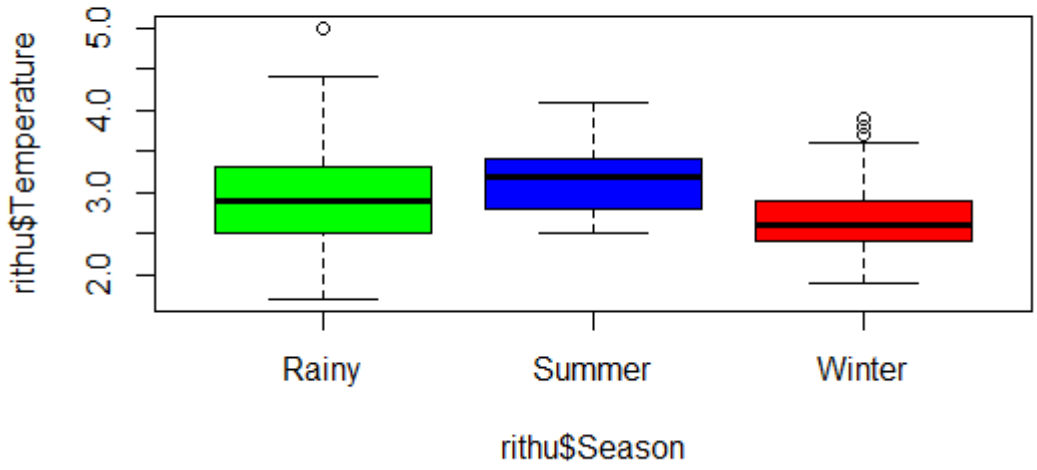
### 1.1 Variable Identification

Mean() : mean of the dataset
Summary () : Complete summary of the dataset
Hist() : To represent the frequency of the values into ranges, and also it represent the total number of values present in that dataset
Read.csv() : To understand the csv file
By() : To identify the summary of seasons.
Pnorm(): To Calculate the probability of going below temperature 2deg C
1-pnorm() : To raise the temperature of above 4deg c
Sd() : Standard Deviation
Str() : It is used to check the structure of dataset
t.test() : One Sample T Test observation
Boxplot() : comparing the distribution of data across dataset by drawing boxplot for each of them.

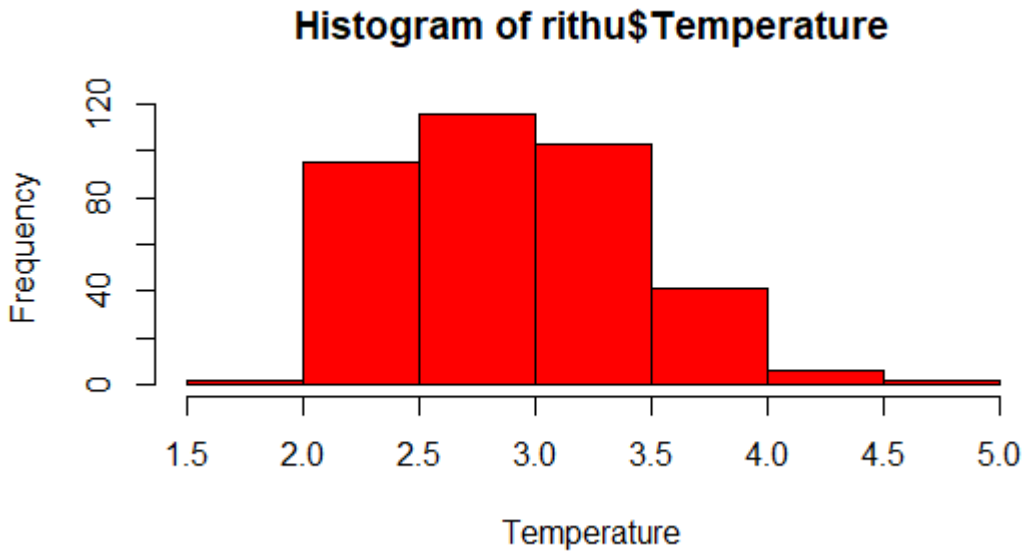
Pvalue(): To check the hypothesis testing for rejecting or accepting the null value.
Zstat(): Have declared this function along with formula to do z test
Nested “if-else” : To check how much penalty should they charge for AMC Company

## 1.2 Summary of observation about each Variable

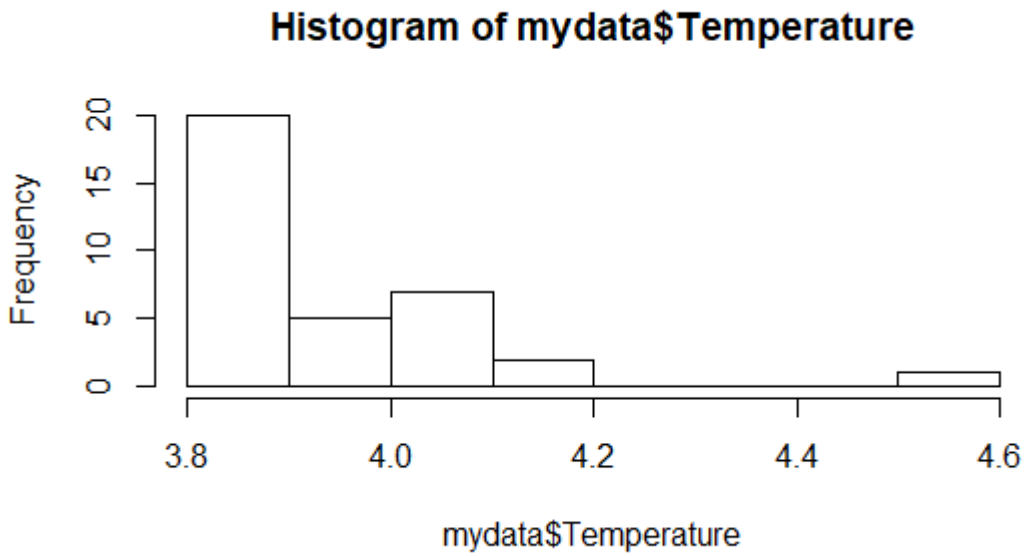
Box Plot of “Cold_Storage_Temp_Data”



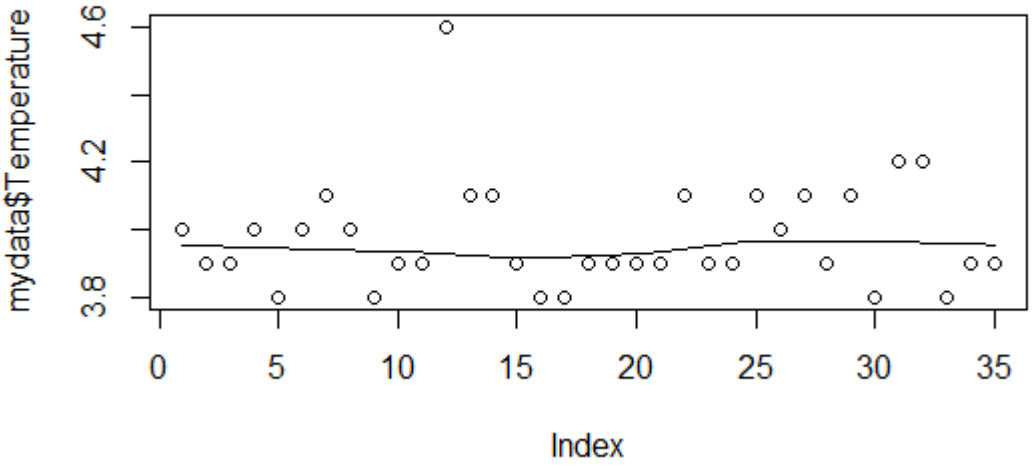

Histogram of “Cold_Storage_Temp_Data”




Histogram of “Cold_Storage_Mar2018”



Scatterplot of “Cold_Storage_Mar2018”



## 4. Diagnosis of Problem 1&2

Problem 1 :

As per the problem if temperature will go above 2.5 % to =<5% then their should be penalty charges 10%.

As per the problem if temperature goes above 5% then their should be penalty charges 25% .

So from the calculation done in R Studio there is chances of going above temperature 4 degree 2.07% (0.02072629) which is within 0.05 ie 5%

And from calculation again we got the data temperature will go less than 2degree (0.02914744) 2.91% which is between 2.5% and 5% .

Hence above solution we can come to an conclusion that there must be a chance of giving 10% penalty charges.

Problem 2:

We can say Null Hypothesis = 3.9

Alternate Hypothesis is greater than 3.9

From the z test and T Test we have got P Value is 0.004711 and from problem alpha is 0.1

So as P Value is less than Alpha we can conclude we can reject the null hypothesis and accept the alternate hypothesis with a confidence level 90%.

**Hence from both the test we can come an conclusion that the temperature will go above 3.9 Degree C.**

## 5. Appendix A – Source Code

**Solution 1:**

```
###mean for all year - 2nd ques
setwd("G:/Projects")
getwd()
rithu=read.csv("Cold_Storage_Temp_Data.csv")
attach(rithu)
rithu
summary(rithu)

###sd for full year - 3rd ques

sd(Temperature)

###mean for all season - 1st ques

by(rithu,INDICES = Season,FUN = summary)
###Below 2 degree c - 4th ques

p1= pnorm(2,mean = 2.963,sd=0.508589)
p1
###Above 4 degree c - 5th ques

p2=1-pnorm(4,mean = 2.963,sd=0.508589)
p2
pnorm(2,mean = 2.963,sd=0.508589,lower.tail = TRUE)

#### Question 6
q=p1
q
if(q>0.025 && q<=0.05){
  (penalty = '10%')
}else if    (q>0.05){
```

```
  (penalty = '25%')
} else  {
  penalty='0%'
}
print(penalty)

boxplot(rithu$Temperature~rithu$Season,col=c("green","blue","red"
))
hist(rithu$Temperature,xlab = "Temperature",col="red")
```

**Solution 2:**

```
setwd("G:/Projects")
getwd()
mydata=read.csv("Cold_Storage_Mar2018.csv")
str(mydata)
summary(mydata)
sd=sd(mydata$Temperature)
sd
hist(mydata$Temperature)
boxplot(mydata$Temperature,horizontal = T)
scatter.smooth(mydata$Temperature)

### T Test

t.test(mydata$Temperature,alternative = "greater",paired =FALSE)
sample_mean=mean(mydata$Temperature)
sample_mean
sqrt=sqrt(35)

### Z Test
xbar=3.9
Mu=sample_mean
n=35
sd
```

```
zstat=(xbar-Mu)/(sd/(n^0.5))
zstat

### P Valve
pvalve=pt(zstat,34)
pvalve

hist(mydata$Temperature)
boxplot(mydata$Temperature)
scatter.smooth(mydata$Temperature)
```