Vishal.S

G3 PGPDSBA JULY

SMDM PROJECT DSBA

COLD STORAGE PROJECT

|  |  |  |
| --- | --- | --- |
| SNO | TITLE | PAGE NUMBER |
| A | LIST OF FIGURES | 2 |
| 1 | PROJECT OBJECTIVE | 3 |
| 2 | ASSUMPTIONS | 3 |
| 2.1 | PROBLEM 1 ASSUMPTIONS | 3 |
| 2.2 | PROBLEM 2 ASSUMPTIONS | 3 |
| 3 | DATA ANALYSIS AND HYPOTHESIS TESTING | 4 |
| 3.1 | ENVIRONMENT SETUP AND DATA IMPORT | 4 |
| 3.1.1 | INSTALL NECESSARY PACKAGES AND IMPORT LIBRARIES | 4 |
| 3.1.2 | SETUP WORKING DIRECTORY | 4 |
| 3.1.2.1 | PROBLEM 1 | 4 |
| 3.1.2.2 | PROBLEM 2 | 6 |
| 4 | SOLUTION | 9 |
| 4.1 | PROBLEM 1 | 9 |
| 4.2 | PROBLEM 2 | 10 |
| 5 | CONCLUSION | 12 | |
| 5.1 | PROBLEM 1 | 12 | |
| 5.2 | PROBLEM 2 | 12 | |

TABLE OF CONTENTS

LIST OF FIGURES AND SNIPPETS

|  |  |  |  |
| --- | --- | --- | --- |
| SNO | FIGURE NUMBER | TITLE | PAGE NUMBER |
| 1 | 3.1.2.1.1 | head () | 4 |
| 2 | 3.1.2.1.2 | Shape | 5 |
| 3 | 3.1.2.1.3 | info () | 5 |
| 4 | 3.1.2.1.4 | describe () | 5 |
| 5 | 3.1.2.1.5 | isnull () | 6 |
| 6 | 3.1.2.1.6 | Checking for outliers using Boxplot | 6 |
| 7 | 3.1.2.2.1 | head () | 6 |
| 8 | 3.1.2.2.2 | Shape | 7 |
| 9 | 3.1.2.2.3 | info () | 7 |
| 10 | 3.1.2.2.4 | describe () | 7 |
| 11 | 3.1.2.2.5 | isnull () | 8 |
| 12 | 3.1.2.2.6 | Checking for outliers using Boxplot | 8 |
| 13 | 4.1.4 | Code Snippet for finding Probability of temperature below 2º C | 9 |
| 14 | 4.1.5 | Code Snippet for finding Probability of temperature above 4º C | 10 |
| 15 | 4.1.6.1 | Code Snippet for finding Probability of temperature below 2º C | 10 |
| 16 | 4.1.6.2 | Code Snippet for finding Probability of temperature above 4º C | 10 |
| 17 | 4.1.6.3 | Code Snippet for finding Probability of temperature between 2º C and 4º C | 10 |
| 18 | 4.2.2.1 | Code Snippet for finding the P value | 11 |
| 19 | 5.1.1 | The Mean Temperature for various seasons | 12 |

1.PROJECT OBJECTIVE

The Objective of the project is to explore and analyse the cold storage data sets (“Cold\_Storage\_Temp\_Data\_.csv” and “Cold\_Storage\_Mar2018\_.csv”) in python and perform Descriptive Analysis in Data Set 1 and Hypothesis testing in Data Set 2 to gain insights about the plant maintenance

This Project report consists of the following

* Loading and reading of the Datasets in Python
* Understanding the Shape of the Datasets
* Graphical Exploration of the Datasets
* Descriptive Statistics in Dataset 1
* Hypothesis Testing in Dataset 2
* Inference from the Analysis

2.ASSUMPTIONS

The Following Assumptions were made for this project on the individual datasets

2.1 PROBLEM 1: DATASET 1: Cold\_Storage\_Temp\_Data

* Temperature data for the whole year is normally distributed
* There is no mistake in recording the temperature data during the whole year period
* The extreme value of temperature that may be present is recorded as it is

2.2 PROBLEM 2: DATASET 2: Cold\_Storage\_Mar2018

* The Population Standard Deviation of Temperature Data is not known
* The Sample size of 35 is large enough to assume the central limit theorem applies here
* There is no mistake in recording the temperature data during the whole year period
* The extreme value of temperature that may be present is recorded as it is

3.DATA ANALYSIS AND HYPOTHESIS TETSTING

3.1 Environment Setup and Data Import

3.1.1 Install Necessary packages and import libraries

The below packages were used in the code for the Descriptive Analysis of the Datasets

* Pandas
* Matplotlib
* Seaborn
* Scipy.Stats

3.1.2 Setup Working Directory

Setting a working directory on starting of the Python session makes importing and exporting data files and code files easier. The below lines of codes set the working directory to the PC location where the datasets are present

3.1.2.1 Problem 1

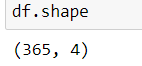
3.1.2.1.1 head ()

Table

Description automatically generated

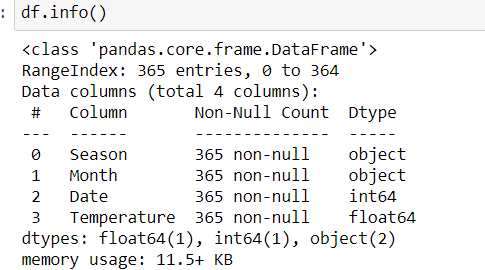
Inference: All variables are present in a proper format

3.1.2.1.2 shape



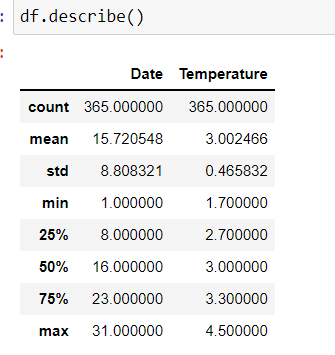
Inference: The Dataset has 365 observations and 4 variables

3.1.2.1.3 info ()



Inference: The various Datatypes present in the datasets are shown

3.1.2.1.4 describe ()



Inference: The 5-point summary is shown

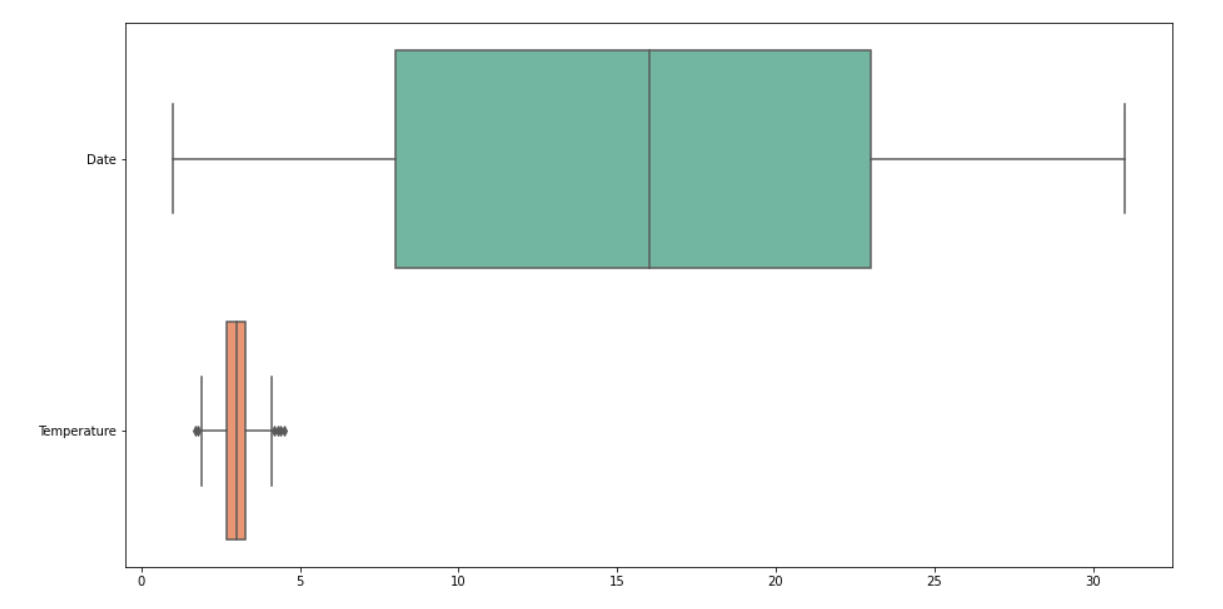
3.1.2.1.5 isnull ()

Text

Description automatically generated with medium confidence

Inference: There are no null values present in the dataset

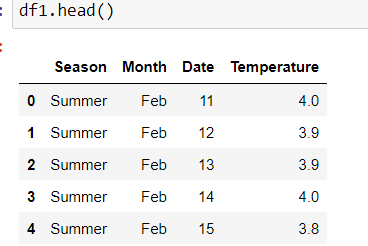
3.1.2.1.6 Checking Outliers using Boxplot



Inference: It has been observed that there is an outlier value in the variable temperature

3.1.2.2 Problem 2

3.1.2.2.1 head ()



Inference: All variables are present in a proper format

3.1.2.2.2 shape

Graphical user interface, application

Description automatically generated

Inference: The Dataset has 35 observations and 4 variables

3.1.2.2.3 info ()

Text

Description automatically generated

Inference: The various Datatypes present in the datasets are shown

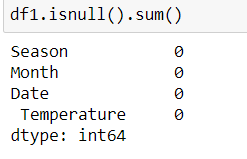
3.1.2.2.4 describe ()

Table

Description automatically generated

Inference: The 5-point Summary is shown

3.1.2.2.5 isnull ()



Inference: There are no null values present in the dataset

3.1.2.2.6 Checking Outliers using Boxplot

Chart, box and whisker chart

Description automatically generated

Inference: It has been observed that there is an outlier value in the variable temperature

4.SOLUTION

**4.1 Problem 1:** Cold\_Storage\_Temp\_Data\_.csv

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 - 4º C. 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”

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

A screenshot of a computer

Description automatically generated with low confidence

2. Find the overall mean for the full year.



3. Find Standard Deviation for the full year.



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

By using the norm.cdf () function in python we find the probability of temperature having fallen below 2º C

A picture containing timeline

Description automatically generatedThe probability of temperature having fallen below 2º C is 1.569%

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

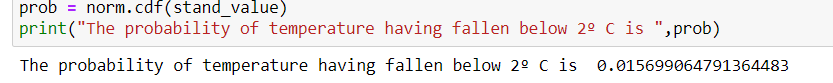
By using the norm.cdf () function in python we find the probability of temperature having gone above 4º C

Graphical user interface, text

Description automatically generated

The probability of temperature having gone above 4º C is 1.612%

6. What will be the penalty for the AMC Company?



Graphical user interface, text

Description automatically generated

Graphical user interface, text

Description automatically generated

Since the Probability is around 3.18% which is above 2.5% and less than 5% then the penalty would be 10% of AMC (annual maintenance case)

**4.2 Problem 2**: Cold\_Storage\_Mar2018\_.csv

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”

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

The Hypothesis test that will be applied to test this dataset is one tailed T (Right Tailed 1 sample T test) because of the below reasons

* The distribution of the dataset shows it is skewed to the right
* The sample size is large (n=35) so that the central limit theorem applies to the dataset
* The Population standard deviation is not known in this case
* Also, the test here is a one-sided test since we need to check whether the temperature is less than the upper acceptable value of 3.9 and greater

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

NULL and ALTERNATE HYPOTHESIS

NULL HYPOTHESIS H0: The Temperature is below 3.9 degrees and there is no problem with the cold storage

ALTERNATE HYPOTHESIS H1: The Temperature is above 3.9 degrees and there is problem with the cold storage

Graphical user interface, text, application

Description automatically generated

Since the p value 0.00295 is less than the alpha = 0.1, we can reject the Null Hypothesis

3.Give your inference

The One Sample T test is performed in the dataset to find out whether any corrective action is needed in the plant or the problem is from procurement side. The result of one sample hypothesis test states that P value 0.0047 which is less than the alpha value 0.1.So the null Hypothesis is rejected in this case the P value is less than the alpha value, we can 99.7% say that the null hypothesis is rejected and there is problem with the cold storage plant, and it needs corrective action

5.CONCLUSION

5.1 Problem 1

5.1.1. The Mean Overall Temperature in Various Seasons

|  |  |
| --- | --- |
| **Season** | **Mean Temperature** |
| Summer | 3.1475ºC |
| Winter | 2.776ºC |
| Rainy | 3.08872ºC |

2.The Overall Mean for the full year is 3.002465

3.The Overall Standard Deviation for the full year is 0.465831

4.The Probability of Temperature having fallen below 2º C is 1.56

5. The Probability of Temperature having gone above 4º C is 1.612%

6. The Probability is around 3.18% which is above 2.5% and less than 5% then the penalty would be 10% of AMC (annual maintenance case)

5.2 Problem 2

The null Hypothesis is rejected in this case the P value is less than the alpha value, we can 99.7% say that the null hypothesis is rejected and there is problem with the cold storage plant, and it needs corrective action