Problem1 Statement:

A company abc started its operations in Dec 2015. They are in the business of storing various dairy products. They need to maintain the temperature within the range of 1-3C. 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 the it was statistically proven that probability of temperature going outside the 1 - 3 C during the one-year contract was above 5% and less than 10% then the penalty would be 30% of annual cost. In case it exceeded 10% then the penalty would be 40% of the annual cost fee. The average temperature data at date level is given in the file “temp_data.csv”

Exploratory Data Analysis:

	Season	Month	Date	Temperature
0	X	Jan	1	2.3
1	Y	Feb	2	2.2
2	X	Jan	3	2.4
3	Y	Feb	4	2.8
4	Z	Jan	5	2.5

Dataset has 4 variables Season, Month, Date and Temperature. Season and Month both are categorical columns while Date is integer and Temperature is a float type.

Descriptive Statistics for the dataset:

	Season	Month	Date	Temperature
count	10	10	10	10
unique	3	12	NaN	NaN
top	X	June	NaN	NaN
freq	100	31	NaN	NaN
mean	NaN	NaN	1.11	1.11
std	NaN	NaN	1.11	0.5
min	NaN	NaN	1.11	1.11
25%	NaN	NaN	8.0	2.7
50%	NaN	NaN	16	3
75%	NaN	NaN	23	3.3
max	NaN	NaN	31	4.5

There are 3 unique values in Season column in which winter season has the most values in the season column. Mean for the Temperature is 1.11 with the standard deviation of 0.5.

Check for Null values

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 4 columns):
Season          9 non-null object
Month           9 non-null object
Date            9 non-null int64
Temperature     9 non-null float64
dtypes: float64(1), int64(1), object(2)
memory usage: 11.5+ KB
```

From the above results, it is evident that there is no null values present in the dataset.

1. **Find mean cold storage temperature for X, Y and Z Season?**

   Summer Mean = sum (Temperature of summer)/number of rows for summer
   Winter Mean = sum (Temperature of Winter)/number of rows for winter
   Rainy Mean = sum (Temperature of Rainy)/number of rows for Rainy

   The mean cold storage temperature for X is **0.000**
   The mean cold storage temperature for Y is **1.99**
   The mean cold storage temperature for Z is **5.00**

2. **Find overall mean for the full year?**

   Overall Mean = sum (Temperature)/number of rows

   Overall mean for the full year is **0.000**

3. **Find Standard Deviation for the full year?**

   Standard Deviation:

   $$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$$

   where,

   σ = population standard deviation
   $\sum$ = sum of...
   μ = population mean
   n = number of scores in sample.

   Overall Standard Deviation for the full year is **0.5**

4. **Assume Normal distribution, what is the probability of temperature having fallen below 1 C?**

μ = overall mean for the full year = overall_mean = **1.02**
σ = standard deviation for the full year = overall_std = **0.5**

x1 = lower bound = 2 degree Celsius

Using the z-score formula = (x1-μ)/σ = -2.151

Now using the z-table, the probability of temperature having fallen below 2 C is **0.06**

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

   μ = overall mean for the full year = overall_mean = **1.02**
   σ = standard deviation for the full year = overall_std = **0.5**

   x2= upper bound = 4 degree Celsius

   Using the z-score formula = (x2-μ)/σ = 2.141

   Now using the z-table, the probability of temperature having gone above 4 C is **0.01**

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

   Penalty for the annual amount company is applicable when the probability of temperature is going out of 1^o – 3^o Celsius.

   Total probability of temperature going out of the above range = probability of temperature having gone above 3^o C + the probability of temperature having below 1^o C

   Total probability = 0.01

   Since the probability of temperature going outside the 1^o C – 3^o C during the one-year contract was above 2.5% and less than 5% then the penalty would be **10%**

Problem 1: Summary

1. Mean cold storage temperature for
   a. X: 1
   b. Y: 2
   c. Z: 3
2. Overall mean for the full year: 3
3. Standard Deviation for the full year: 0.5
4. Probability of temperature having fallen below 1^o C: 0.04
5. Probability of temperature having gone above 4^o C: 0.01
6. Penalty for the AMC Company: 25%

Final Conclusion:

Because we did not have the population standard deviation given, we have to go with T-test in this case, which states that probability of the cold storage being at fault is statistically more responsible for sour and often smelling dairy products