**MDSC-102-Final Lab Test Report-ESE**

Name : Gade Sai Shravan.

Regd No : 23904.

Subject Code : Inferential Statistics(P)-MDSC-102.

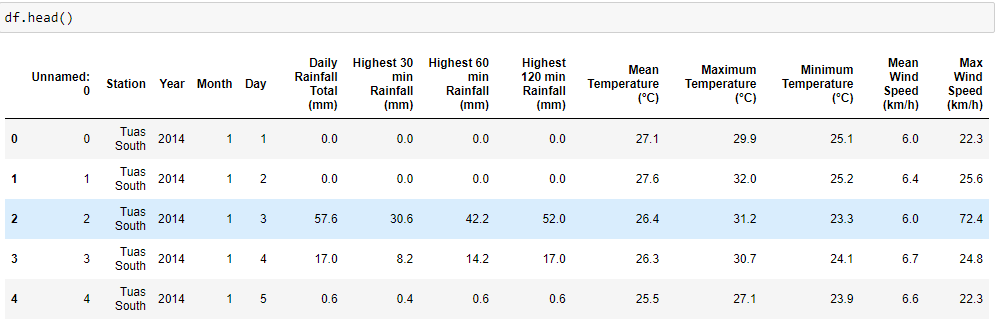
**Problem Statement:**

**Choose a dataset where it has missing values and then pre-process the data. Now Visualize the data from the dataset using matplotlib by plotting the suitable graphs for the data. After visualization, make the inferences from the plots. Next, check the skewness and transform the data to the normal and make the testing hypothesis from it.**

Data Set I have chosen is about the Singapore weather, where it has the 14 features in the dataset. The dataset is about the daily rain, minimum and maximum temperature, mean and max of wind speed in a city of Singapore.

**Dataset:** [**Dataset Link**](https://github.com/saishravan-ss/MDSC-102--Inferential-Statistics--Assignments/blob/main/tuassouth.csv)

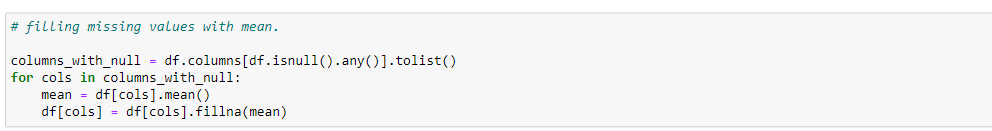
The first five attributes of the dataset are:



There are also missing values present in the dataset about 2% in all the features except Year, Month and Day.

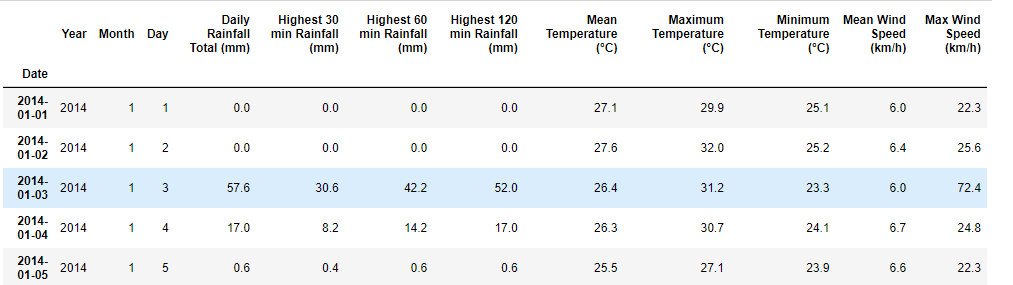
-> **Data Pre-processing:**

When doing data pre-processing, first filled all the missing values with mean of their respective columns. Filling the missing values with mean, median, mode and with the average of their neighbour values gives the good result when working with the dataset.



It is easy to work with the dataset where the most of the features in the dataset are the numeric. Therefore, all the features in the dataset are float except Year, Month, and Day.

Next, created the column “Date” by combining the Year, Month and Day, and, replaced the index of the dataset with the Date feature.



Now, Visualization.

-> **Visualization:**

Since, the dataset deals with the continuous values and it’s more about the weather details, we can choose the Time Series plot for the given dataset, and plotted the Time Series plot for the seven features and made inferences.

The features are:

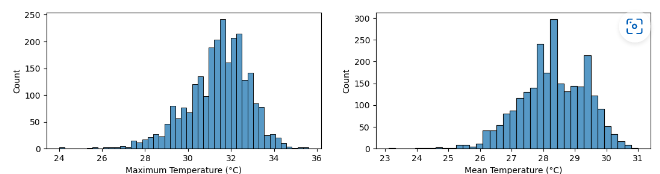
1) Daily Rainfall. 2) Mean Temperature.

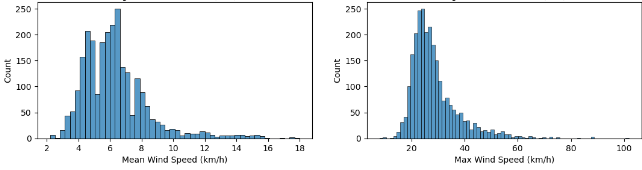
3) Mean Wind Speed. 4) Highest 30 min Rainfall.

5) Highest 60 min Rainfall. 6) Maximum Temperature

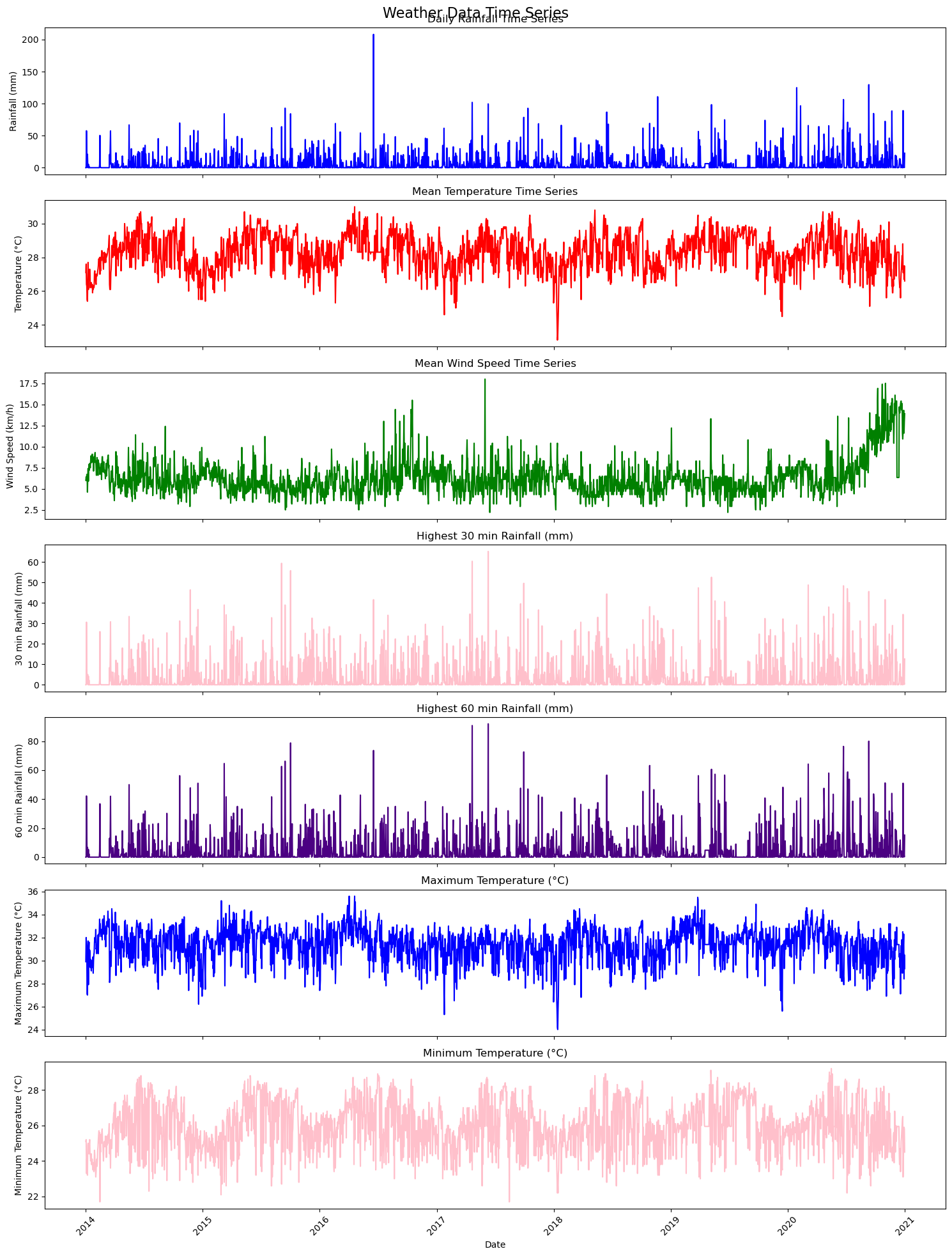
7) Minimum Temperature.

Plots before Normalising:





We can clearly see that the maximum and minimum temperature are left-skewed. Whereas, features Mean Wind and Max wind are left-skewed.



The Inference for few feature are as follows:

1) **Daily Rainfall:** We can see that the rainfall across the years 2014,2015 the highest rainfall was around 100 mm and on the other days it was around 25(mm) to 75(mm) of rainfall. In 2016 we can also see that like in June or July month the highest rainfall was there and that is 200(mm) of rainfall, and in other days of 2016 year it was normal like around 60. Following years from 2018 to 2021 the rainfall in (mm) got increased. It was around 110(mm) to 140(mm) of rainfall.

2) **Mean Temperature:** In the initial years of the 2014 the average temperature was from 25 to 27.8 degrees Celsius. But later on the average temperature was 31 degrees Celsius. The same followed for the years 2015,2016, and 2017. But the starting year of the 2018 the temperature was around less than 24 degrees Celsius. Later, the temperature increased to 31 degrees. And we can see that the average temperature in the city of Tuas South is around 31-32 degrees Celsius.

3) **Maximum Temperature:** The maximum temperature in the 2014 was around 34 degrees Celsius, and in the later months of that year the temperature was around 25 to 32 degree Celsius. In the year 2015 the highest temperature was around 35 degrees Celsius. The maximum temperature in the year 2016 was 35 degrees Celsius. Later we can observe that average temperature in the year 2017 and 2018 was around the 32 to 34 degree Celsius. The highest temperature in the year 2019 was 35 degrees Celsius. And in the year 2020 the temperature was around 33 degrees Celsius. By this we can observe that the minimum temperature in the city Tuas South is 24 degrees and highest is 35 degrees.

4) **Minimum Temperature:** We can clearly observe that the minimum temperature across the years in the city Tuas South is 22 degrees and highest among all of them it can go is around 29 degrees.

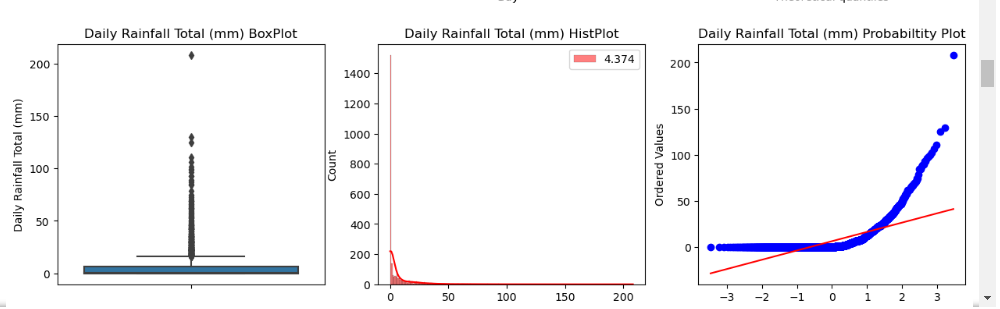
**Transforming Data to Normal:**

-> First, calculate the skewness of each of the features using skew () function.

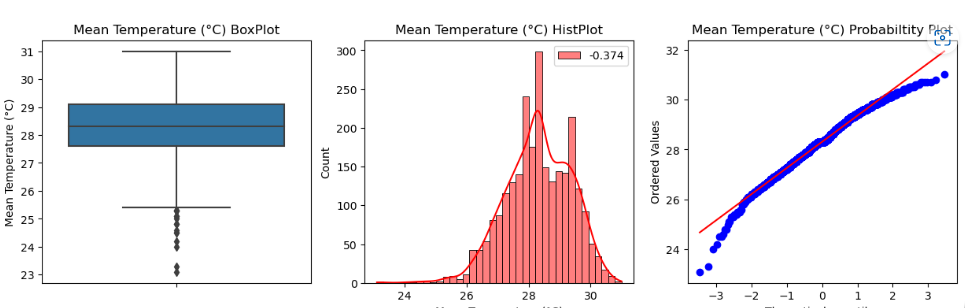
-> Then next plot the Box Plot, Hist Plot and Prob Plot (graphical technique for assessing whether or not a data set follows a given distribution such as normal). We can see that clearly the data is not normal and has many outliers and can identify the outliers.

For example,

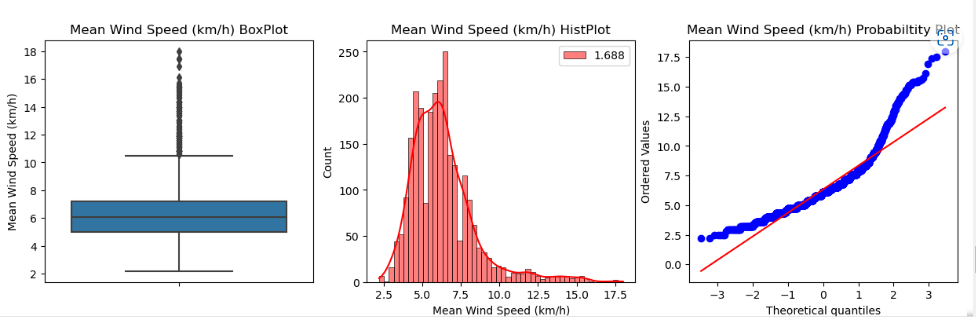
**Daily Rainfall:**



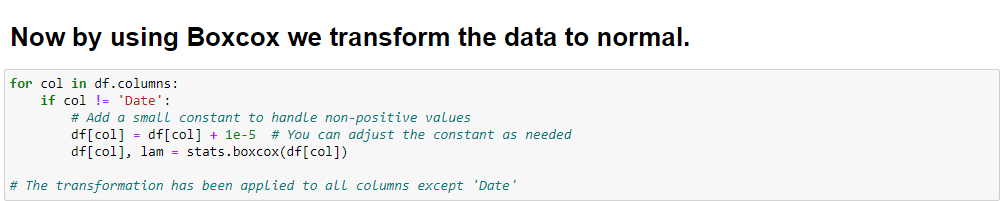
**Mean Temperature:**

****

**Mean wind Speed:**

****

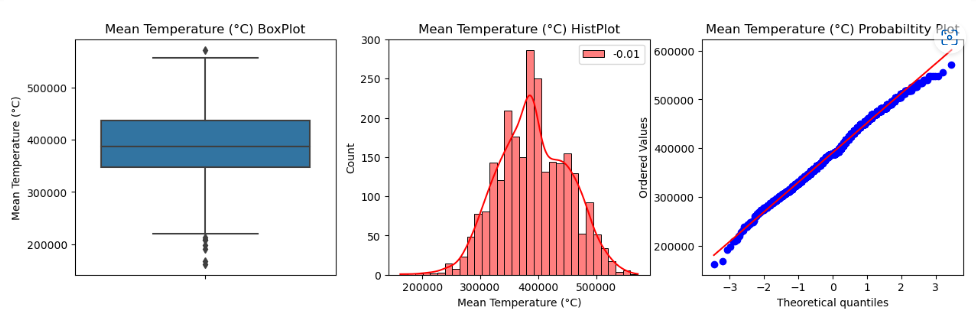
Now, we can see that the data is not at all normal, then we transform the data to the normal using boxcox.



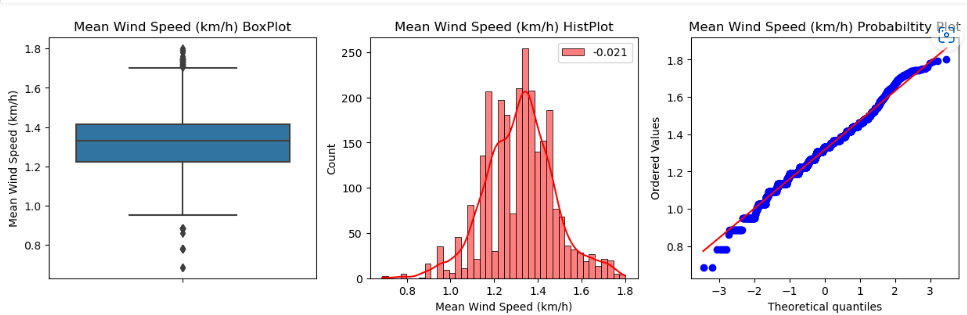
Now again, plot the same graphs, where we can observe the difference of the plots previously posted above.

The plots are:

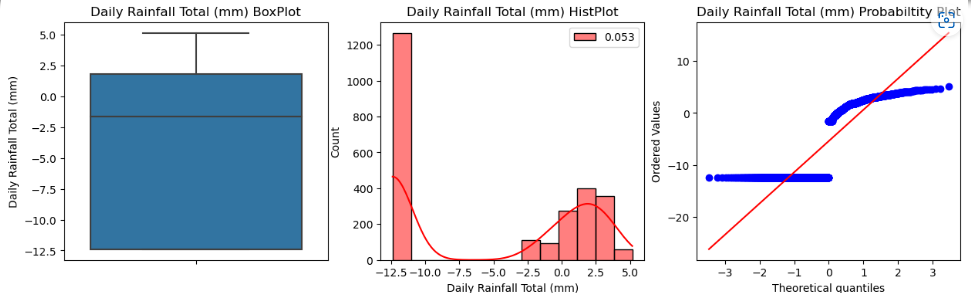
**Mean Temperature:**



**Mean Wind Speed:**

****

**Daily Rainfall:**

****

**Testing Hypothesis:**

Let us choose 2 features Daily Rainfall, and Maximum Temperature from testing hypothesis.

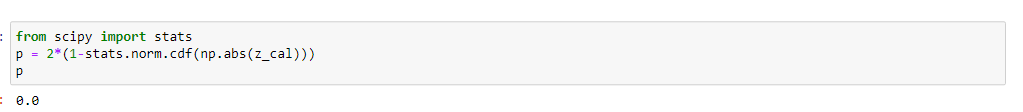
First calculated the mean of the both the features and performed the z-test and t-test by making the hypothesis.

-> Mean of the Daily rainfall is 6.33105

And hypothesis is:



Then performed the z-test using its formula, the z-calculated values is 9.56953, and the p-values is:



If p-values is less than or equal to the alpha value, we reject the H0 else we Do not the H0. 

-------------------------------------- Thank You-----------------------------------