**CONTENTS**

1. Problem Statement
2. Data
3. Data Understanding
4. Data summary
5. Dimension of data
6. Preparation of data
7. Visualization of data
8. Checking stationarity
9. Rolling Statistics test
10. Dickey-fuller test
11. Autocorrelation function and partial autocorrelation function
12. Prediction of data
13. Training using 1 month data
14. Training using 2 month data

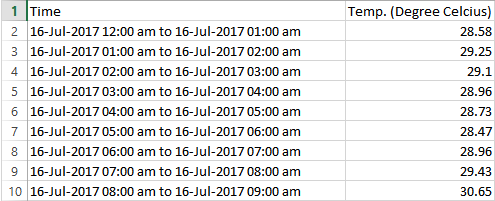
# Problem Statement

**Temperature prediction model for a location**

To forecast the temperature of a particular location based on the time series analysis. The accuracy of the model must be greater than 95%.

# Data

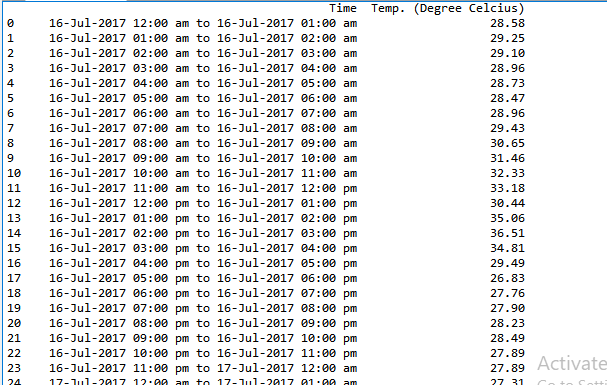
Now, we need to load the data from csv file using read\_csv().



# 

# Data Understanding

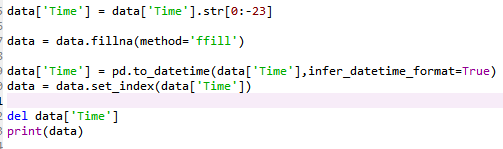
### Data Summary

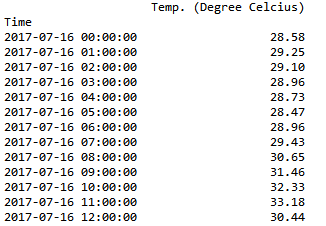


### Dimension of data

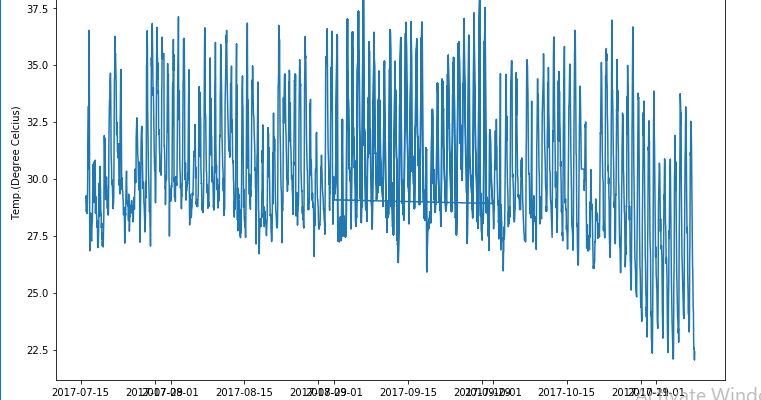
3487 rows and 2 columns

### Preparation of data





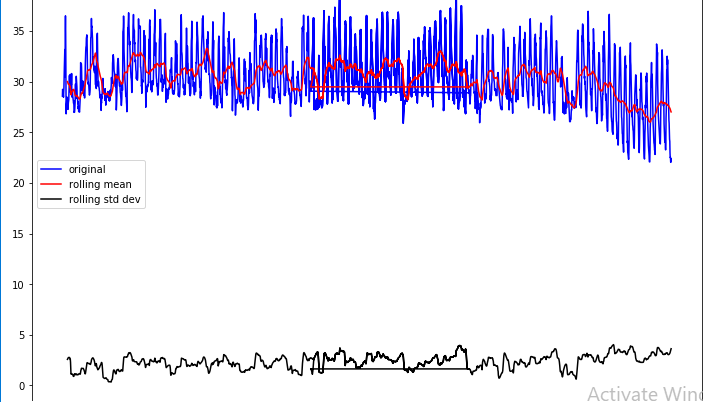
1. Visualization of data



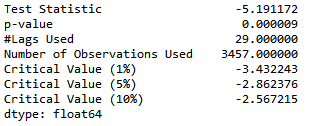
### Checking stationarity on original data

We need to check the stationarity of data as this is the assumption of time series model. There are two ways to check the stationarity:

1. Using Rolling Statistics test



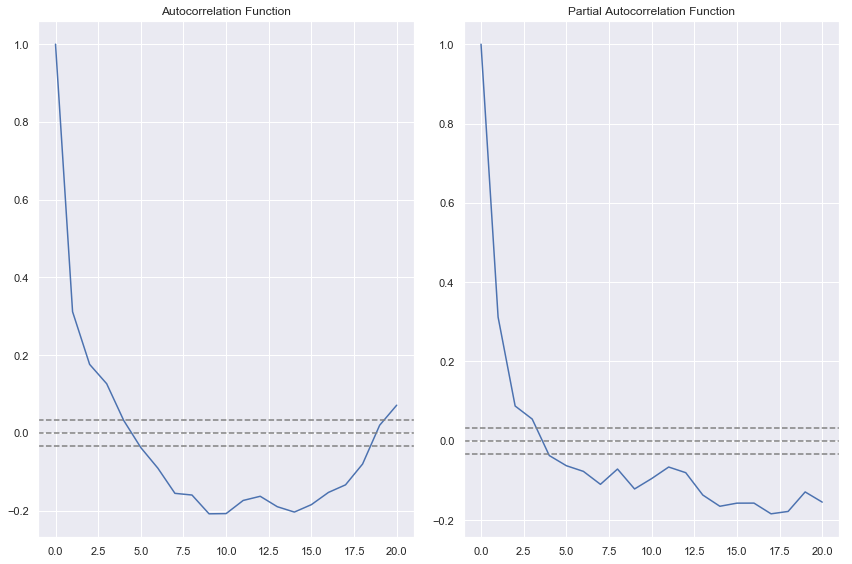
1. Using Dickey-fuller test



It is clearly visible that the data is almost stationary, as in the result of rolling statistics there is very less variations in mean and standard deviation. And from dickey-fuller test we can see that p-value is very less, test-statistics is less than the critical value. But we can improve the results by removing trends from the original data. And we will get pure stationary data to be used for fitting the model.

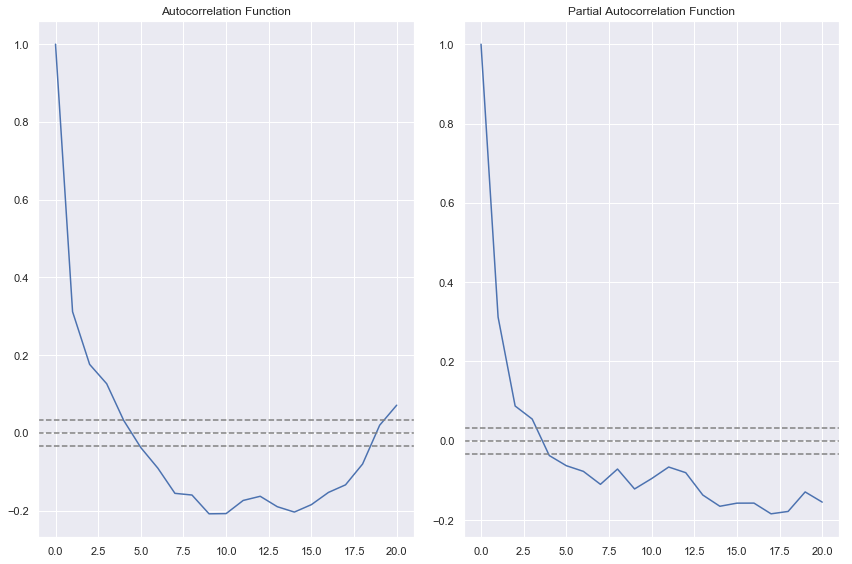
For original data

1. Autocorrelation Function 2. Partial Autocorrelation Function



For transformed data

1. Autocorrelation Function 2. Partial Autocorrelation Function



Training model

We will train the model using 3 month data and then test the model on rest 1 month data. Here, we will use AR, MA and combined model for training. And we will compare the accuracy to check the best fit model.

For this we need to give three parameters p, d and q.

In this plot, the two dotted lines on either sides of 0 are the confidence intervals. These can be used to determine the ‘p’ and ‘q’ values as:

1. **p** – The lag value where the **PACF** chart crosses the upper confidence interval for the first time. If you notice closely, in this case p=1.
2. **q** – The lag value where the **ACF** chart crosses the upper confidence interval for the first time. If you notice closely, in this case q=1.

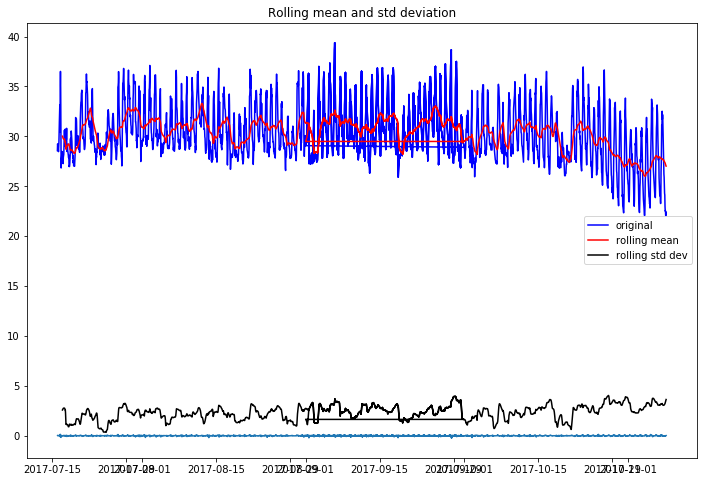
e.g. - model = model\_name(history, order=(1,1,0))

model\_fit = model.fit(disp=0)

output = model\_fit.forecast()

After differentiation

1. Using Rolling Statistics test



1. Using Dickey-fuller test

Test Statistic -5.191172

p-value 0.000009

#Lags Used 29.000000

Number of Observations Used 3457.000000

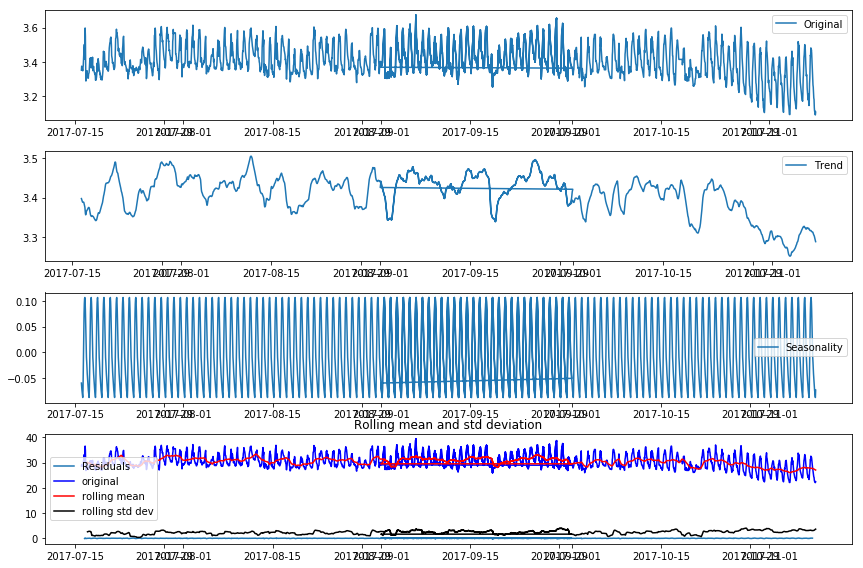
Critical Value (1%) -3.432243

Critical Value (5%) -2.862376

Critical Value (10%) -2.567215

After decomposition

1. Using Rolling Statistics test



1. Using Dickey-fuller test

Test Statistic -5.191172

p-value 0.000009

#Lags Used 29.000000

Number of Observations Used 3457.000000

Critical Value (1%) -3.432243

Critical Value (5%) -2.862376

Critical Value (10%) -2.567215

dtype: float64

Prediction of data

The model which has given the best result.

AR model

Mean Squared Error: 1.047

Root Mean Squared Error: 1.023

Mean Absolute Error: 0.7 degrees.

Accuracy: 97.72 %.

