**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

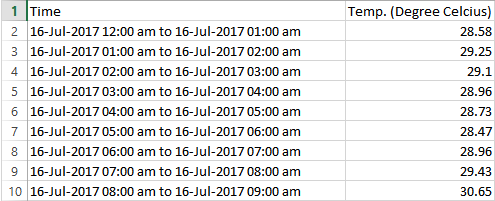
# Problem Statement

**Temperature prediction model for a location**

Model must be trained from the data of 1 month, compare predicted values against actual values of temperature for the rest 3 months (The R-square value to be compared).

# Data

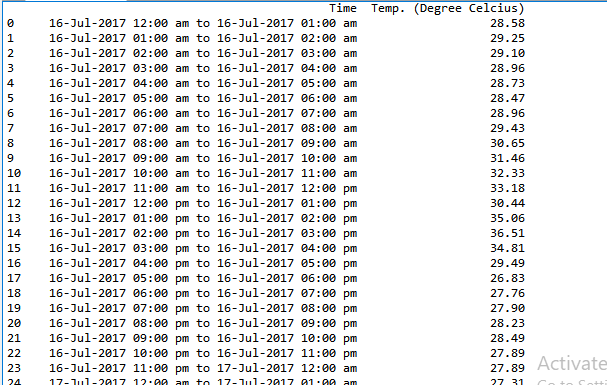
Now, we need to load the data from csv file using read\_csv().



# 

# Data Understanding

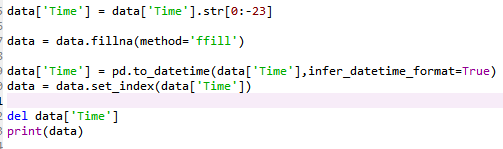
### Data Summary

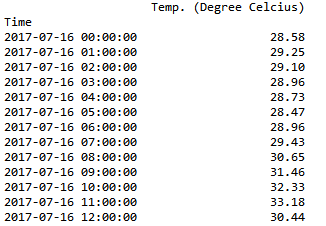


### Dimension of data

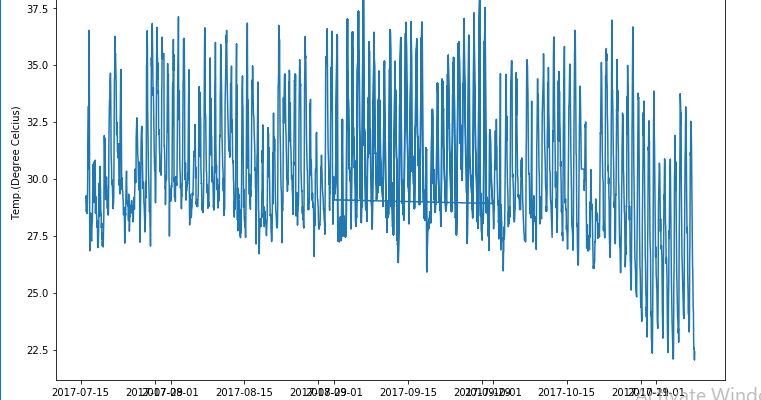
3487 rows and 2 columns

### Preparation of data





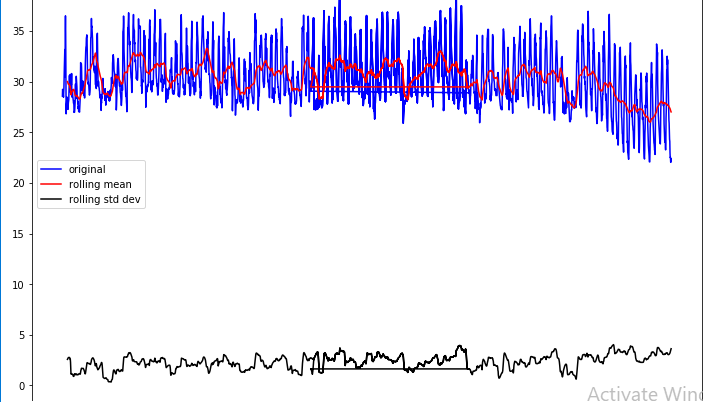
1. Visualization of data



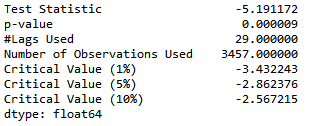
### Checking stationarity

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



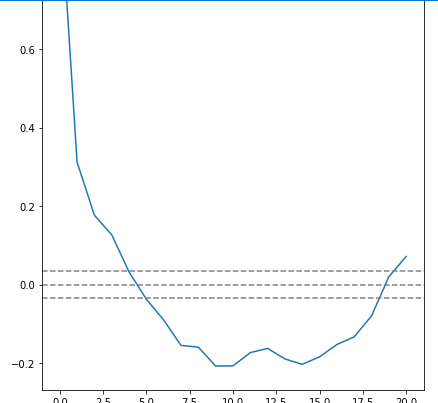
1. Using Dickey-fuller test



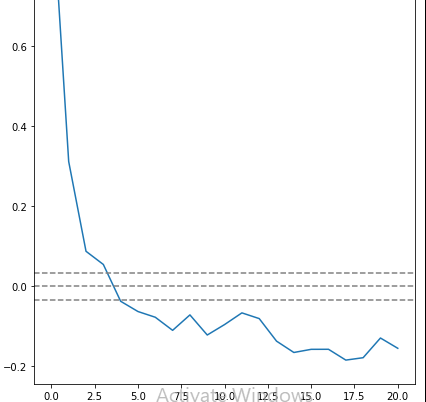
It is clearly visible that the data is stationary, as in the result of rolling statistics there is very less variations in mean and standard deviation. And from dickey-fuller test we can see that p-value is very less, test-statistics is less than the critical value.

Autocorrelation Function and Partial Autocorrelation Function

1. Autocorrelation Function



1. Partial Autocorrelation Function



Training model

We will train the model using 1 month data and then test the data of rest 3 months. Here, we will choose ARIMA model for training. For thius 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 interevals. 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.

model = ARIMA(history, order=(1,1,0))

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

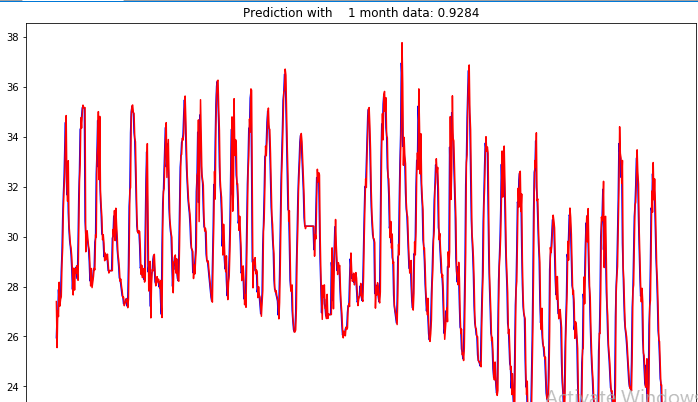
output = model\_fit.forecast()

Prediction of data

1. Training with 1 month data

Mean Squared Error: 0.862

Root Mean Squared Error: 0.928



1. Training with 2 month data

Mean Squared Error: 1.039

Root Mean Squared Error: 1.020

