Production Planning and Control

Mini Project on Forecasting Technique

Weather Forecasting

Team Members:

- 16IM10033 - Shubham Mawa
- 16IM10034 - Bhargav D
- 16IM10035 - Prerit Jain
- 16IM30022 - Aman Aggarwal

Abstract

The aim of the study is to use forecasting methods such as ARIMA for time series forecasting learned in the course Production Planning and Control and apply them to a suitable forecasting problem. We chose the problem of short-term Weather Prediction in Dry Regions of India, specifically, Jaipur. We use seasonal ARIMA models to predict different parameters of the Weather. We can use these to predict if there is going to be extreme weather conditions such as high Rain Fall or an extreme case of drought. This prediction can help Government combat droughts and floods and prepare well in advance by organizing alternative sources of water in case of droughts or ensure better relief measures in case of floods.

Introduction

Every year we constantly hear stories of extreme weather conditions wreaking havoc in different parts of India. These tend to be extreme where extreme weather conditions are persistent. Droughts are one of the most difficult problems faced by the Indian population and they have very severe effects. They cause severe acute water shortage, loss of agricultural plants and crops and many more adverse effects. The other extreme end of the spectrum to droughts, i.e., Floods have similar adverse effects.

We plan to address the problem by building a reliable short-term weather forecasting system which can predict the different weather parameters such as Temperature, Pressure, Dew Percentage, Humidity and, Precipitation using Forecasting Techniques learned as part of the course Production Planning and Control such as ARMA, ARIMA, etc. Since data in

this problem tends to be seasonal in nature because it is, in fact, data of the parameters of the Weather Data are inherently seasonal in nature, we prefer to use variants of the traditional models such as Seasonal ARIMA to better capture the patterns in the data.

We are using the Jaipur Weather Forecasting dataset for the purpose of weather prediction for in this case for dry regions. This dataset provides the weather conditions in the city of Jaipur, India from 1st May 2016 to 11 March 2018. The data fields include mean temperature, mean dew percentage, mean pressure, maximum humidity, minimum humidity, maximum pressure, minimum pressure, maximum temperature, minimum temperature, maximum dew percentage, minimum dew percentage, and precipitation levels.

Using these predictions, organizations and responsible authorities can be better prepared to preempt and face these adverse weather conditions.

Forecasting Methods

Introduction

ARIMA Model

ARIMA (Autoregressive Integrated Moving Average) consists of three major components namely, **AR(p), I(d), MA(q)**. A simple ARIMA model is denoted by **ARIMA(p,d,q).**

The AR(p) part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The auto-regressive parameter *p* specifies the number of lags used in the **autoregressive** component

The MA(q) part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The order q determines the number of error terms include in the **moving average** component.

The I(d) (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values. The d represents the degree of differencing in the **integrated (**$I(d)$**)** component. The expression for a simple ARIMA(p,d,q) is as follows:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Seasonal ARIMA Model

ARIMA models are also capable of a wide range of seasonal data. A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far. It is denoted by **ARIMA (p,d,q) $(P,D,Q)_m$.**

Here, **(p,d,q)** denotes the order of **non-seasonal part** of the model and $(P,D,Q)_m$ denotes the **seasonal part** of the model. **'m'** denotes the period of seasonality.

The seasonal part of the model consists of terms that are similar to the non-seasonal components of the model, but involve backshifts of the seasonal period.
A general expression for ARIMA (p,d,q) $(P,D,Q)_m$ can be represented in backshift notation as:

$$(1-\phi_1 B)\,(1-\Phi_1 B_m)(1-B)(1-B_m)y_t = (1+\theta_1 B)\,(1+\Theta_1 B_m)\,\varepsilon_t.$$

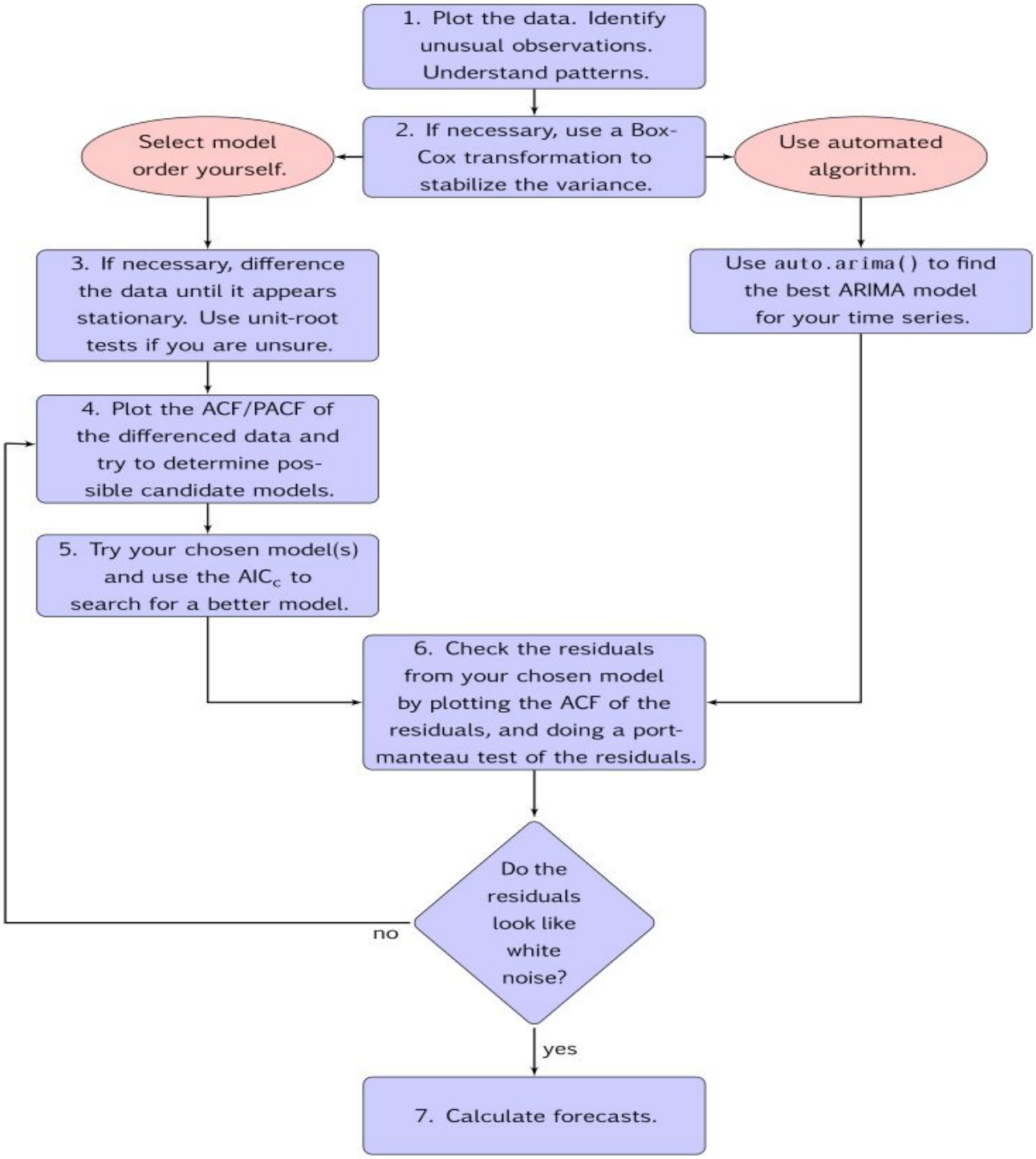
Implementation:

The process of fitting an ARIMA model is sometimes referred to as the **Box-Jenkins method**. The different stages of ARIMA modeling are identification, estimation, diagnostic checking and finally forecasting.

The statistical tools applied during the **identification stage** are transformation, differencing, order determination, analyses of autocorrelations, partial autocorrelations, and inverse autocorrelations, and stationarity tests. In the **estimation** part, parameters of the model are estimated by following a suitable estimation algorithm and the significance of parameters are statistically tested. In the **diagnostic checking stage,** the adequacy of the fitted model is tested by evaluating the residuals.

We have used **sarima()** a statistical function available in R documentation for application of the seasonal ARIMA model. For estimation of orders, the plots of ACF and PACF are analyzed and the best model is determined on the basis of RMSE values.

Flow Chart

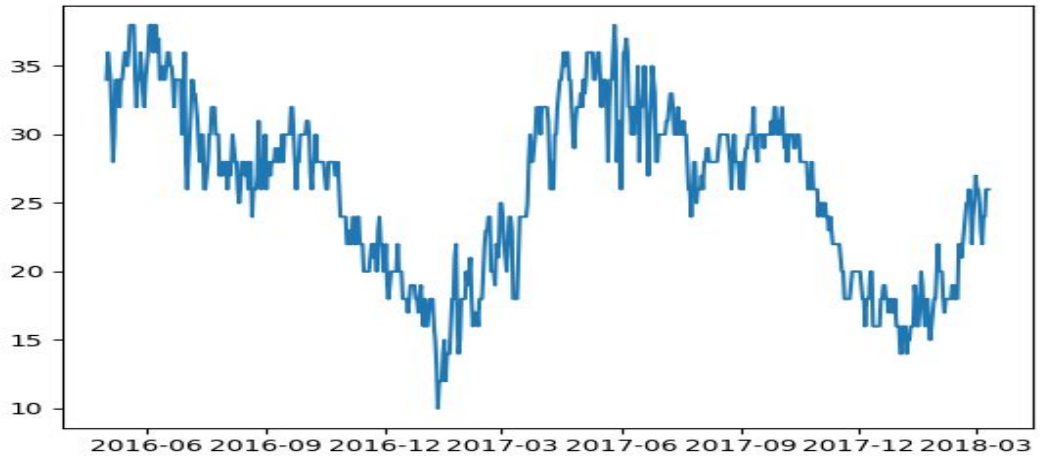

1. Plot the data. Identify unusual observations. Understand patterns.
2. If necessary, use a Box-Cox transformation to stabilize the variance.
Select model order yourself.
Use automated algorithm.
3. If necessary, difference the data until it appears stationary. Use unit-root tests if you are unsure.
Use auto.arima() to find the best ARIMA model for your time series.
4. Plot the ACF/PACF of the differenced data and try to determine possible candidate models.
5. Try your chosen model(s) and use the AIC_c to search for a better model.
6. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals.
Do the residuals look like white noise?
no
yes
7. Calculate forecasts.

Results

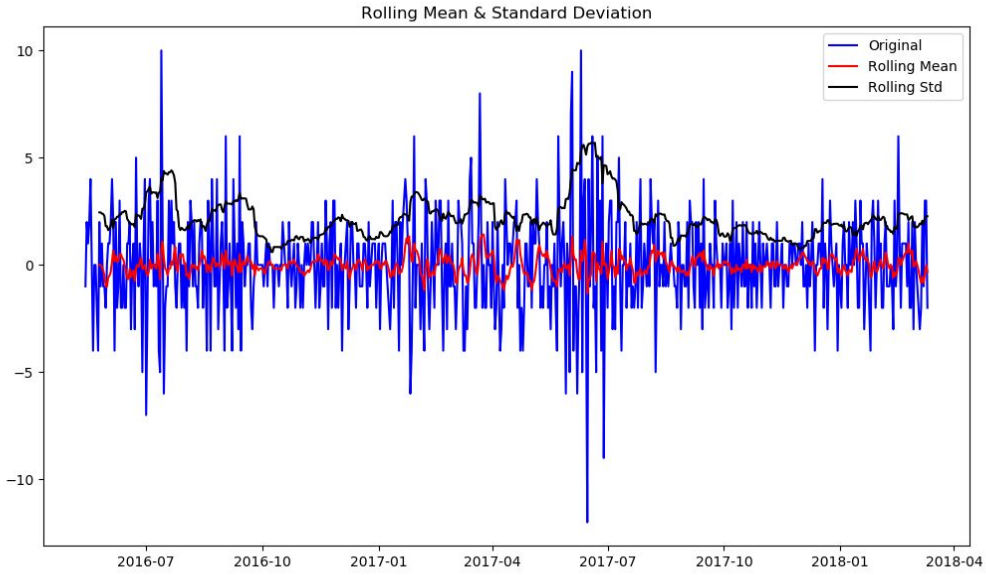
The dataset contains 12 weather parameters. We run 12 separate seasonal ARIMA Models on all these 12 weather parameter values and report the results here.

Mean Temperature

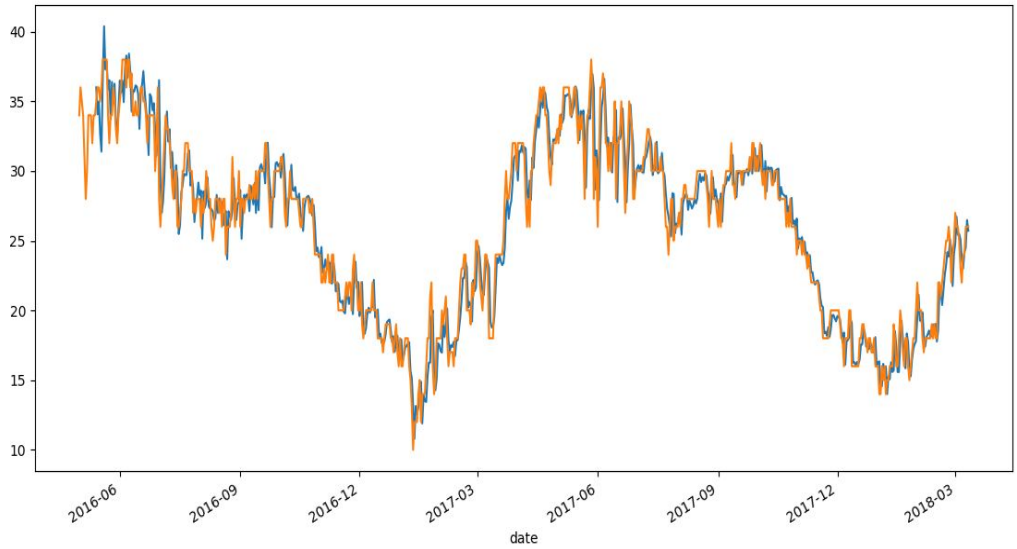
Original Data



Converting the Data into a Stationary Series



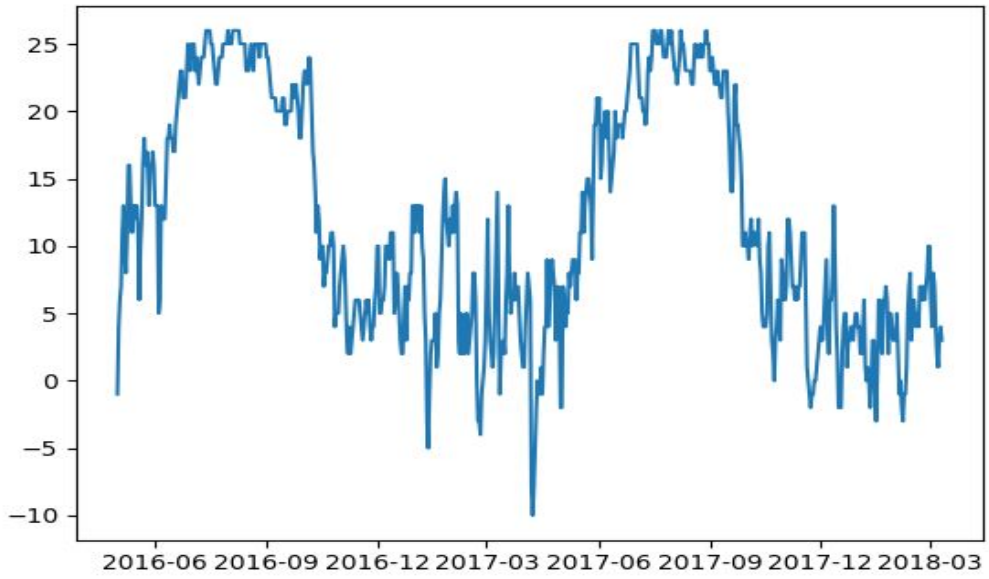
Predictions with the Original Data



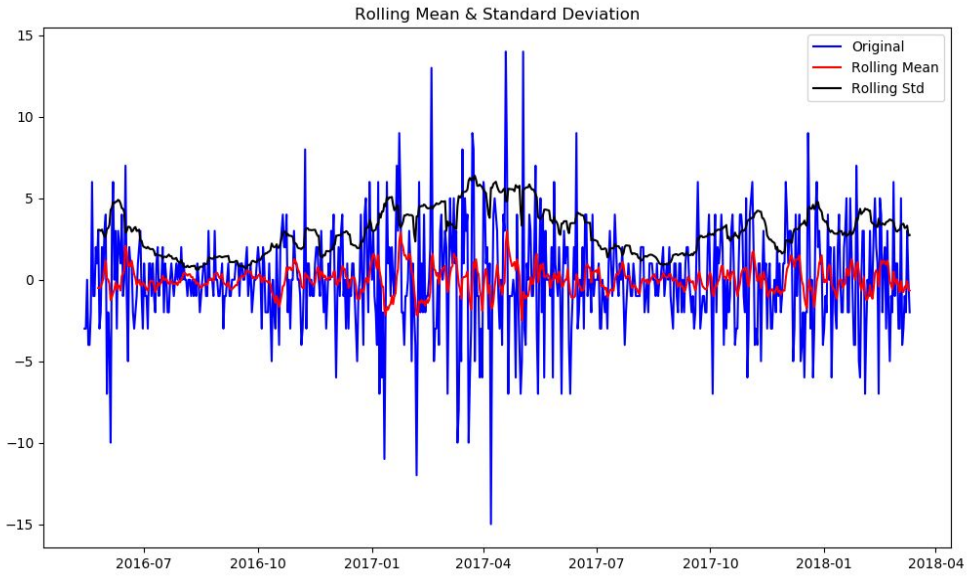
- **RMSE Value (Prediction Accuracy) : 1.5868184**

Mean Dew

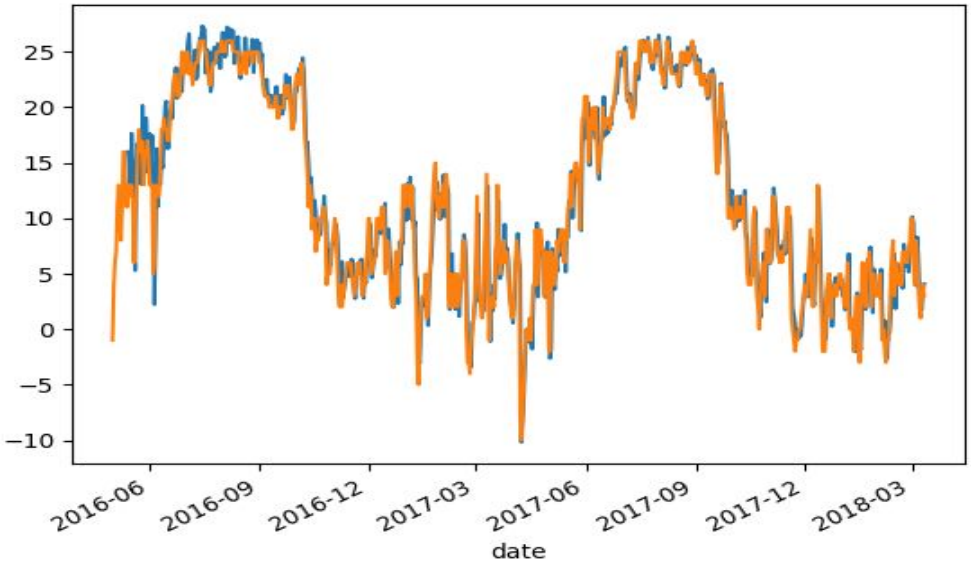
Original Data



Converting the Data into a Stationary Series



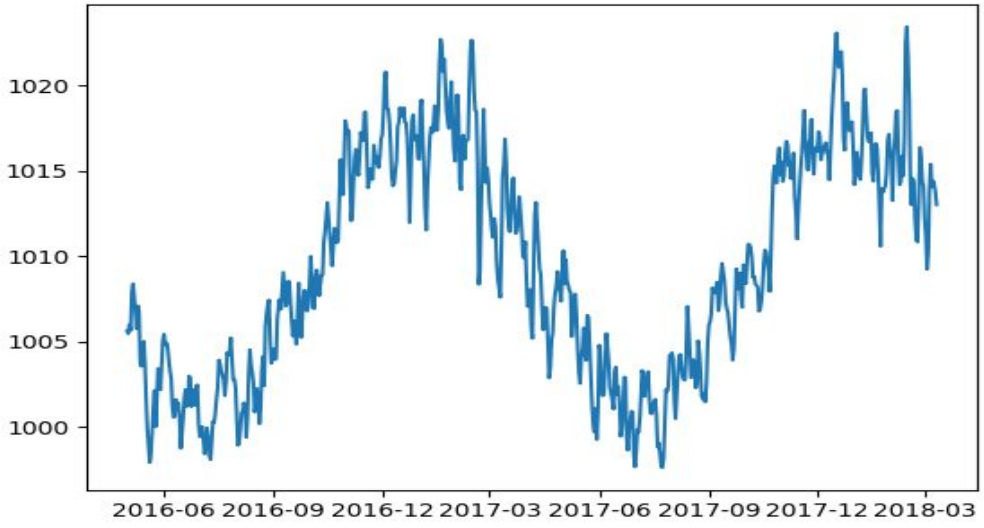
Predictions with the Original Data



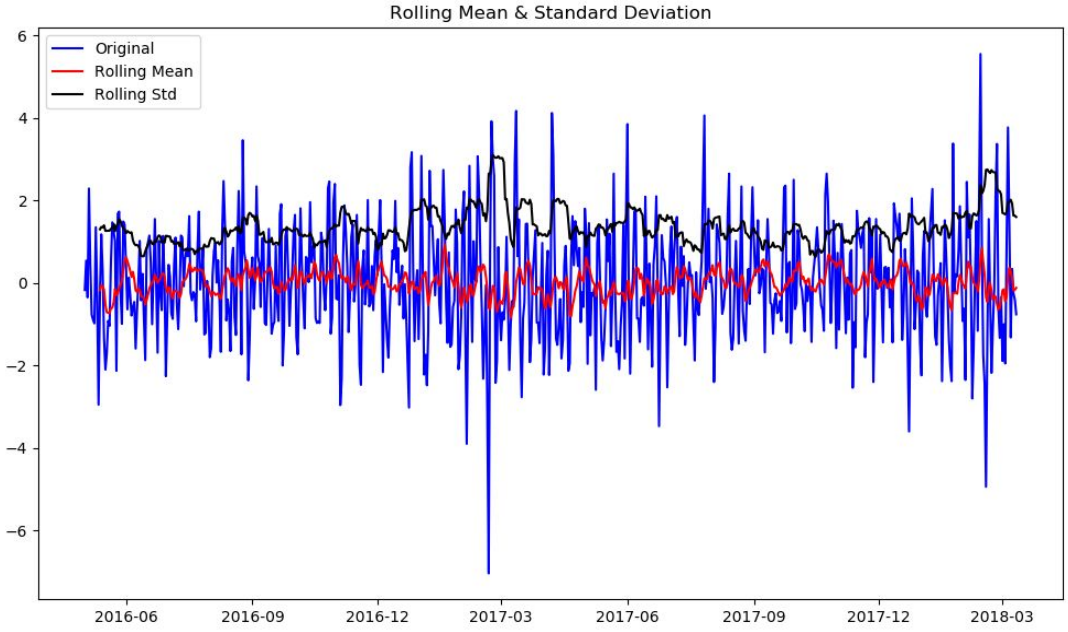
- **RMSE Value (Prediction Accuracy) : 2.198283652**

Mean Pressure

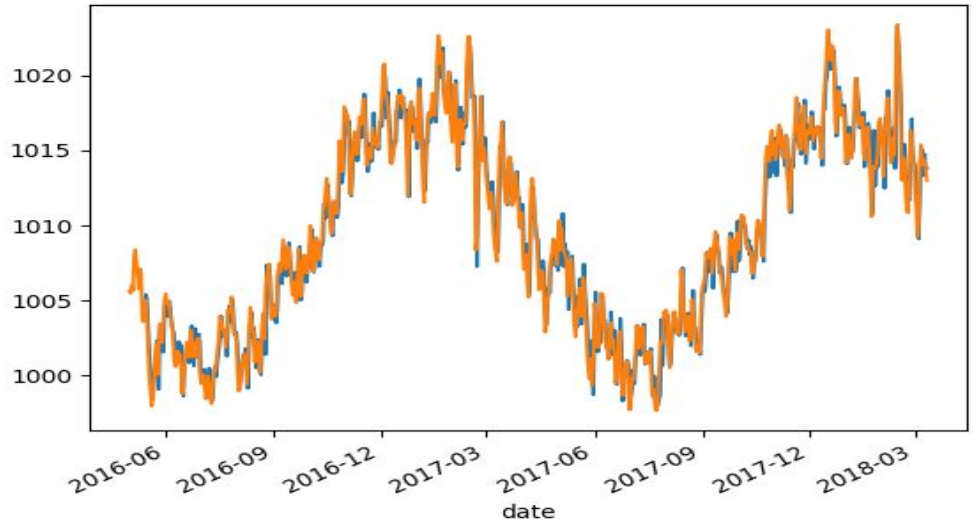
Original Data



Converting the Data into a Stationary Series



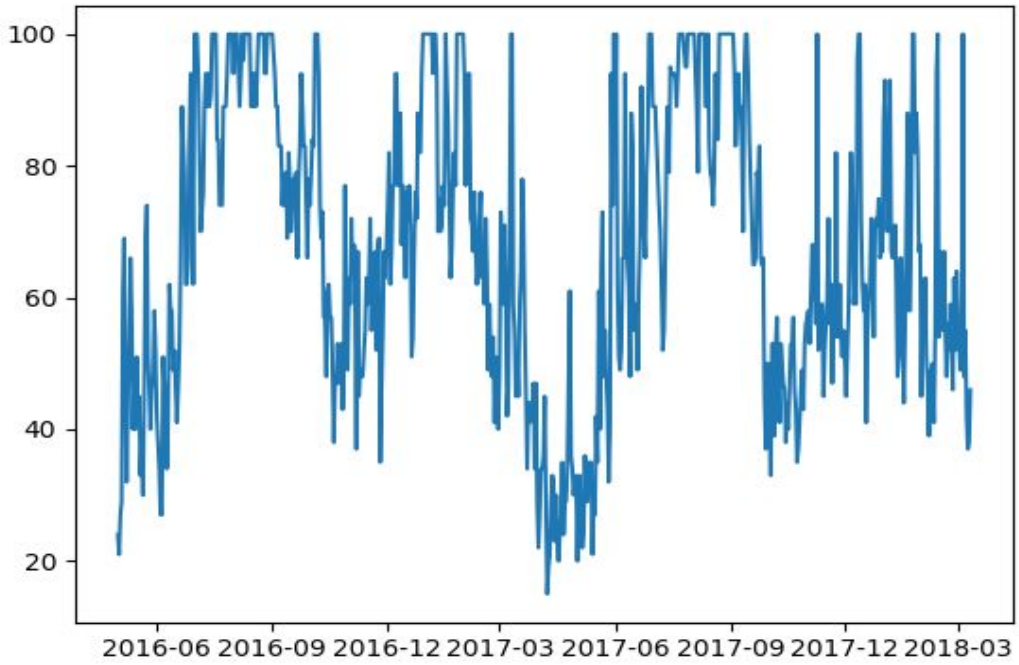
Predictions with the Original Data



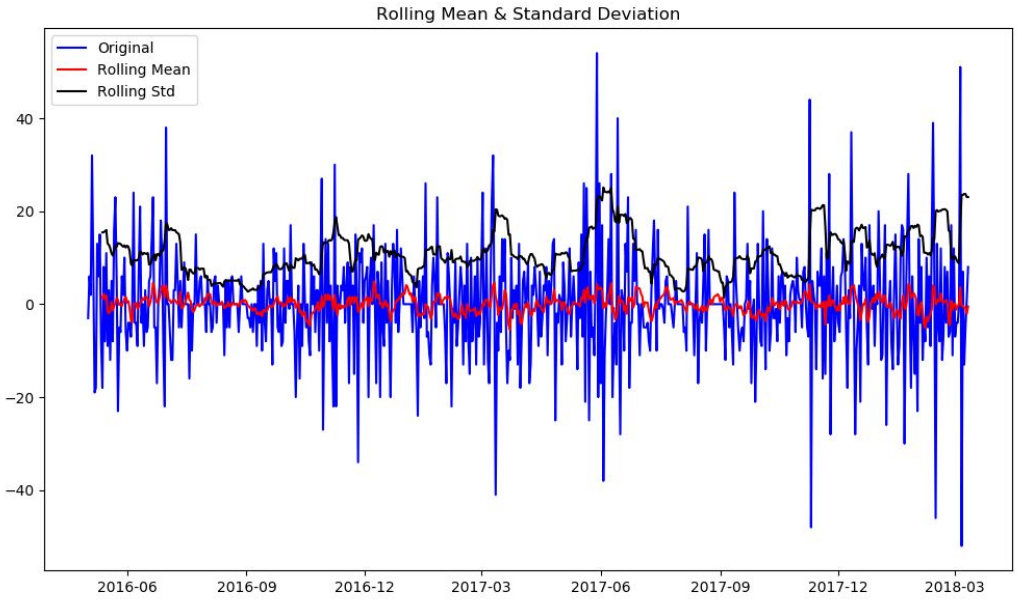
- **RMSE Value (Prediction Accuracy) : 1.21906752**

Maximum Humidity

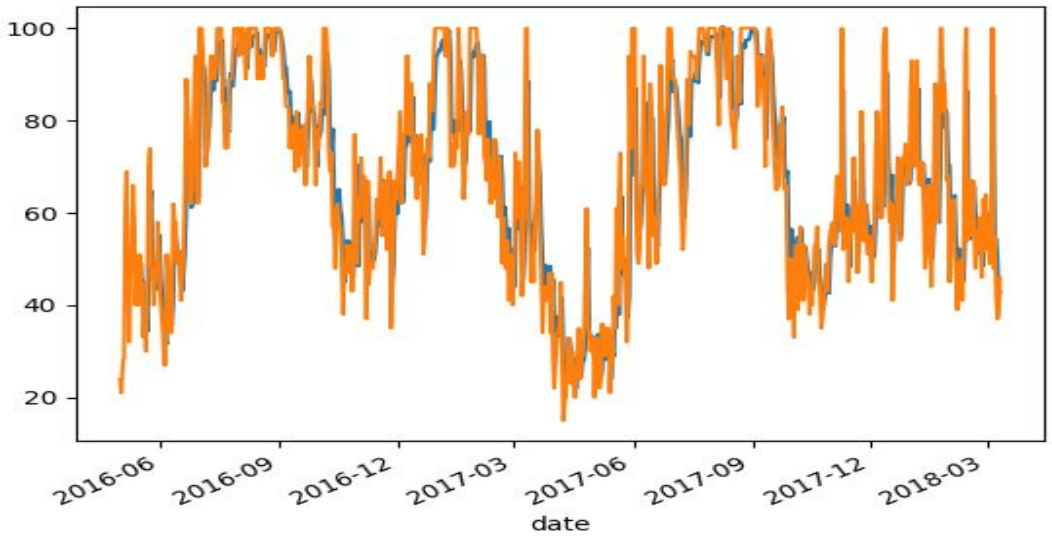
Original Data



Converting the Data into a Stationary Series



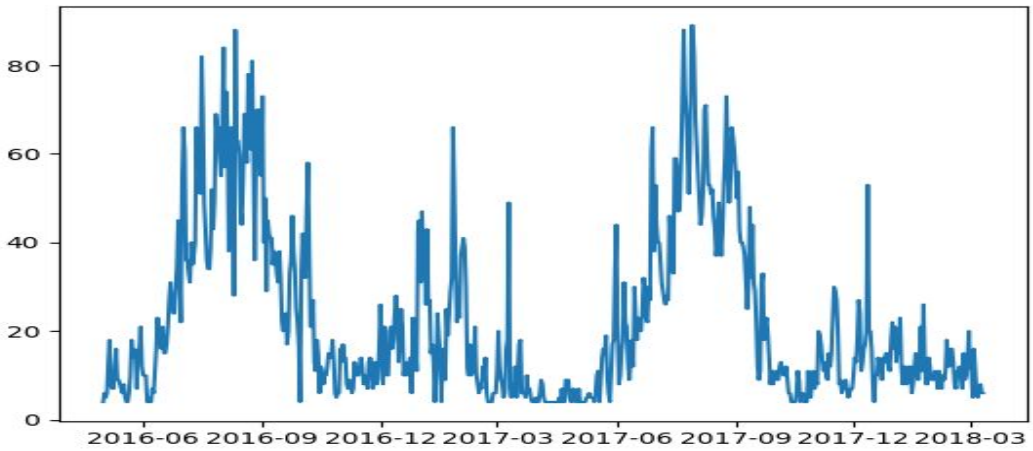
Predictions with the Original Data



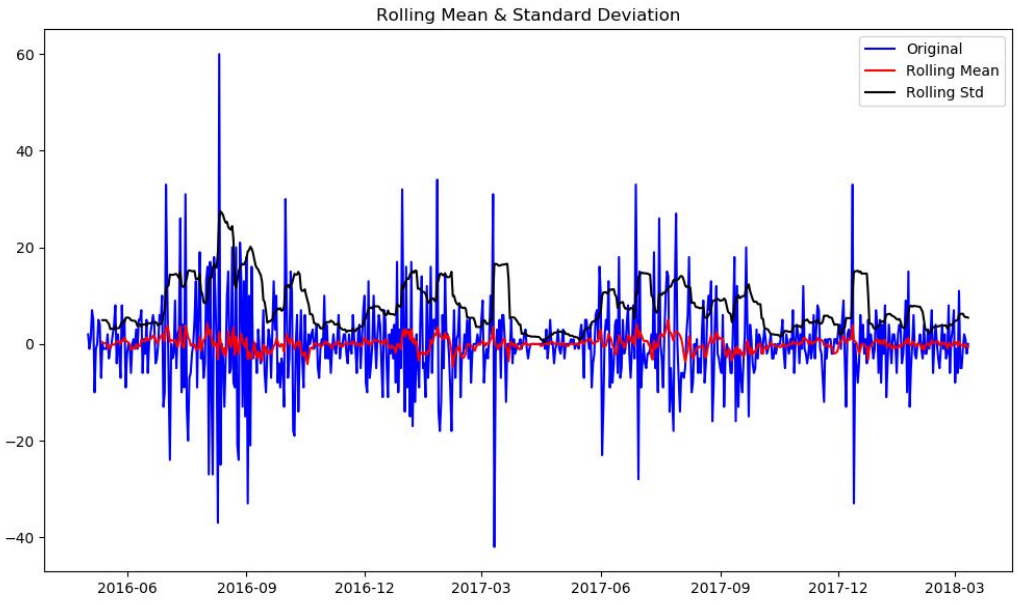
- **RMSE Value (Prediction Accuracy) : 10.77388657**

Minimum Humidity

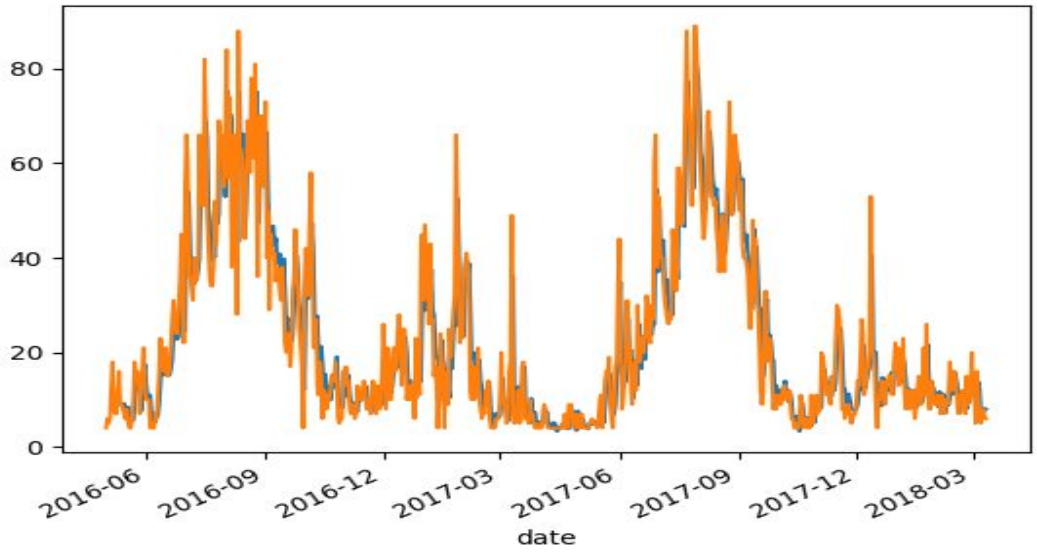
Original Data



Converting the Data into a Stationary Series



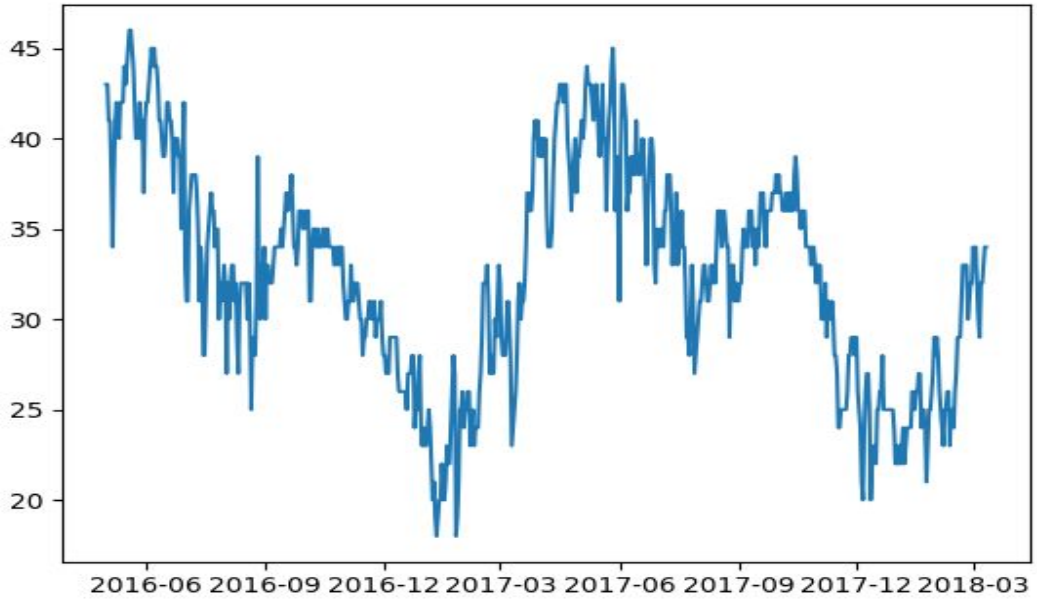
Predictions with the Original Data



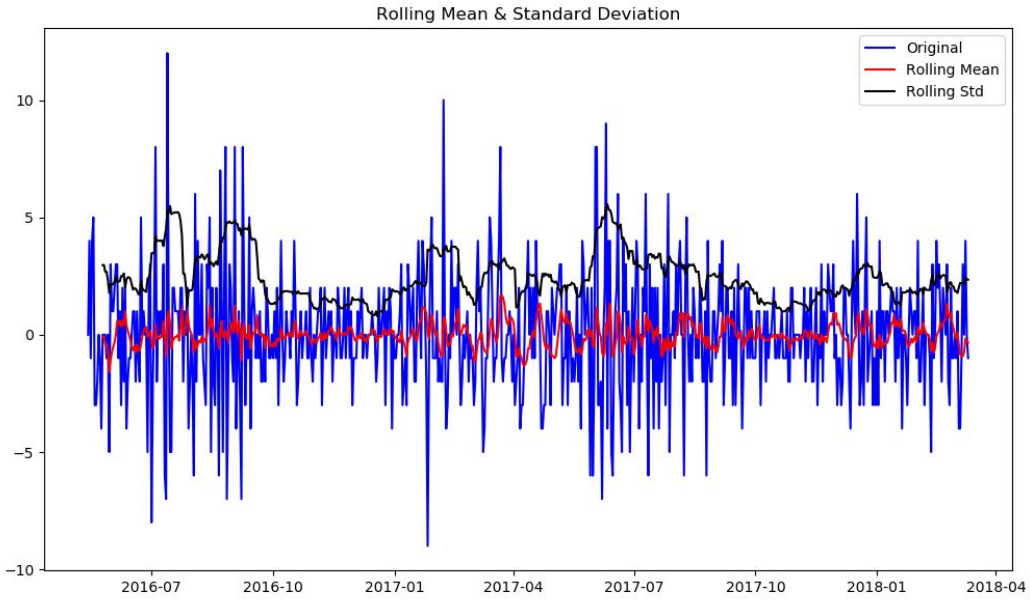
- **RMSE Value (Prediction Accuracy) : 8.26056058**

Maximum Temperature

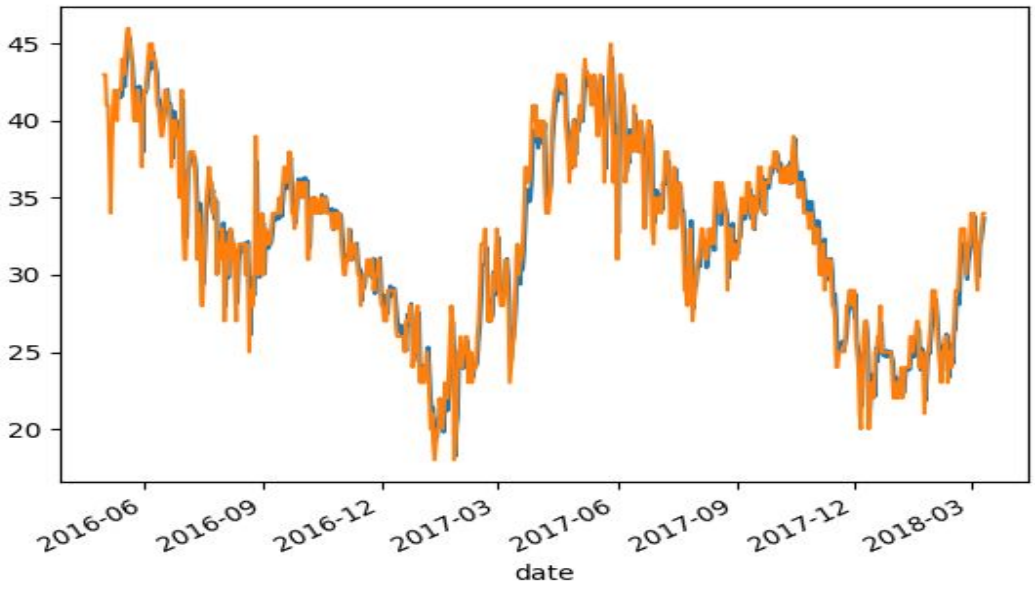
Original Data



Converting the Data into a Stationary Series



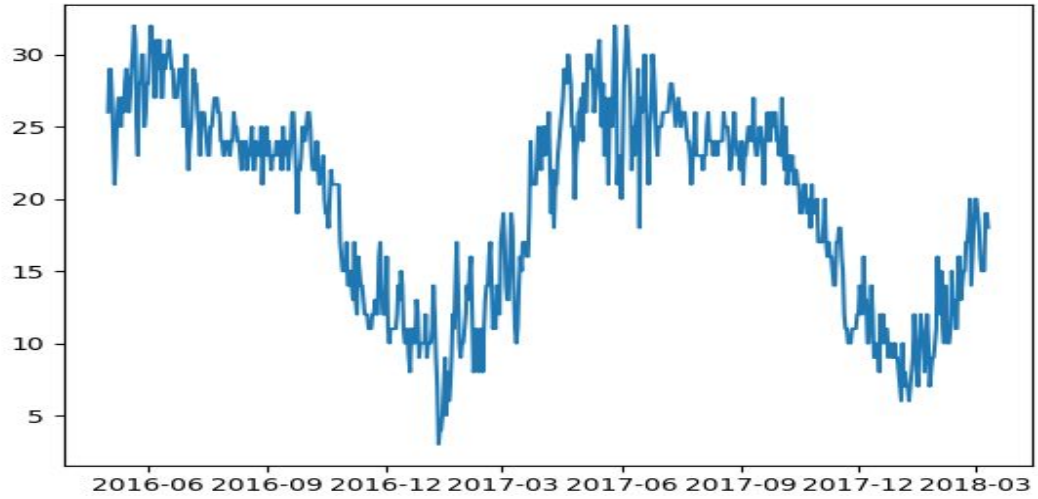
Predictions with the Original Data



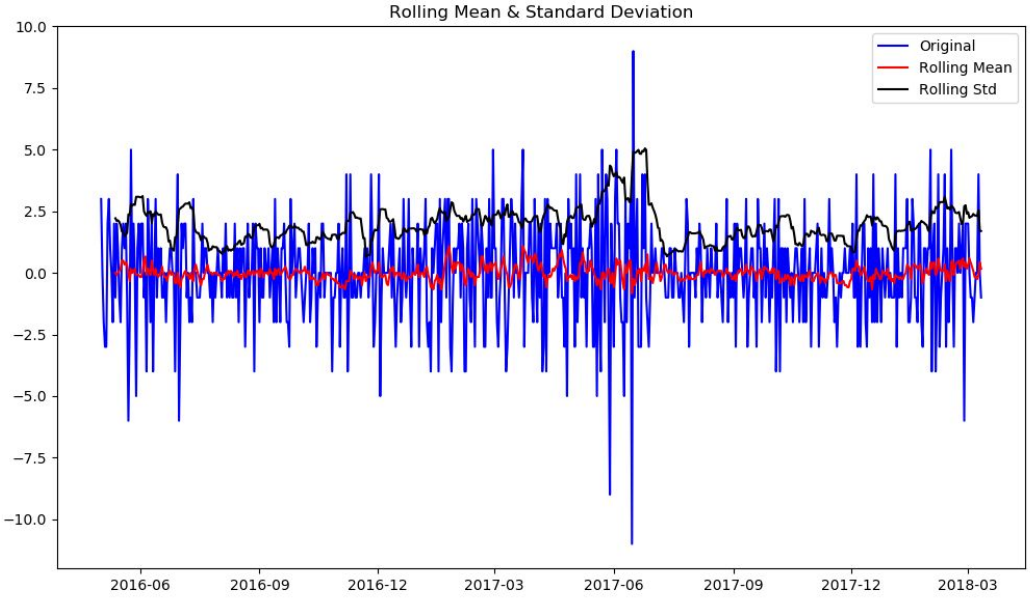
- **RMSE Value (Prediction Accuracy) : 1.727910026**

Minimum Temperature

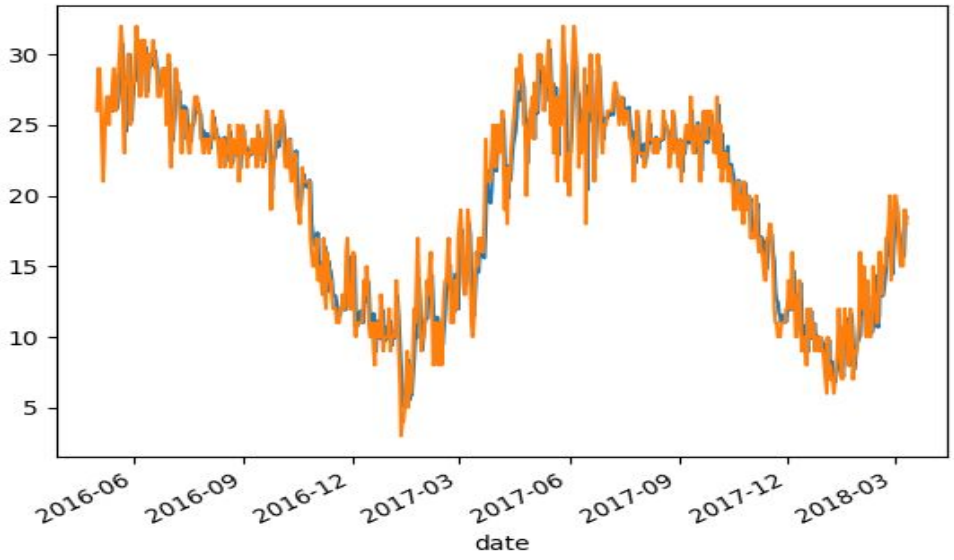
Original Data



Converting the Data into a Stationary Series



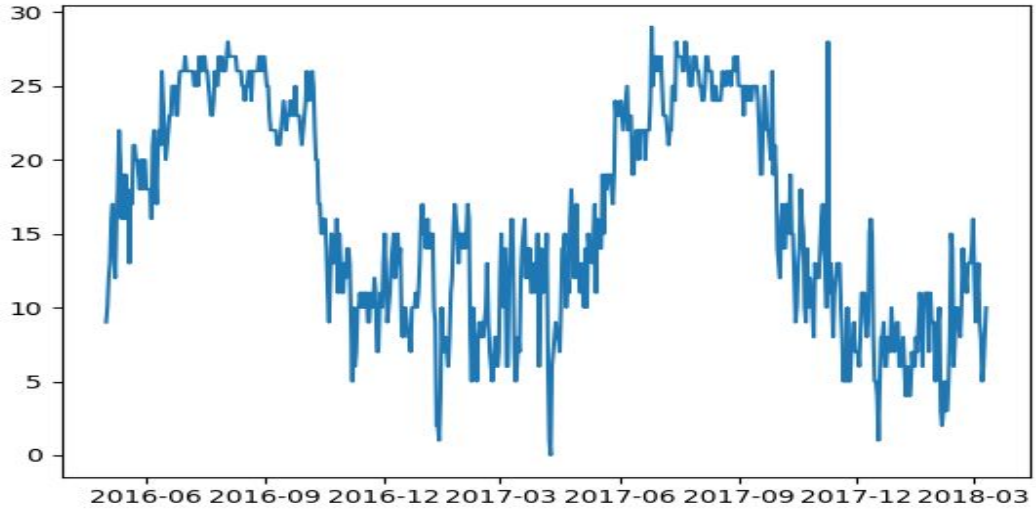
Predictions with the Original Data



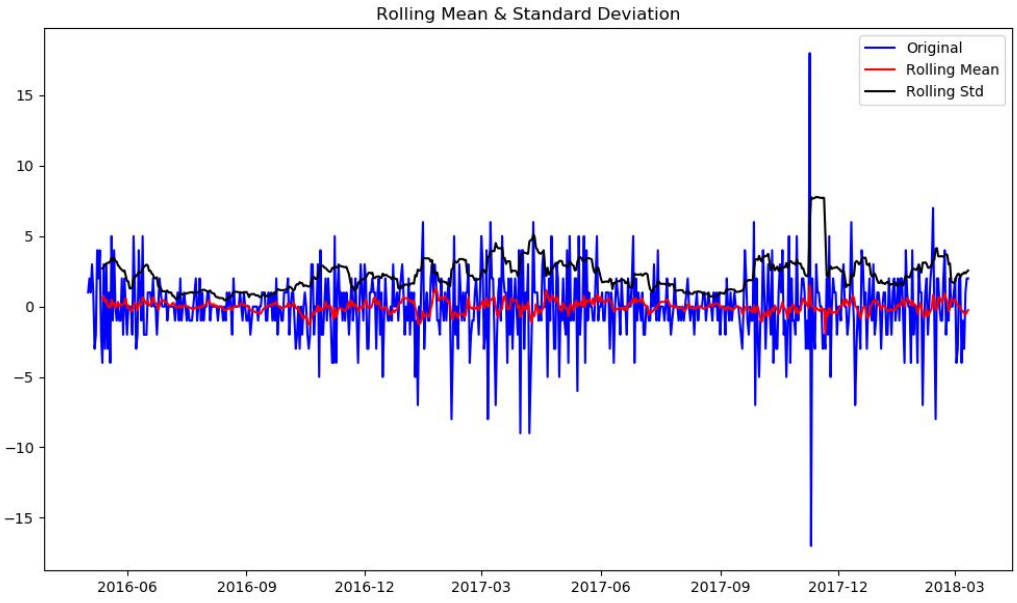
- **RMSE Value (Prediction Accuracy) : 1.883211292**

Maximum Dew

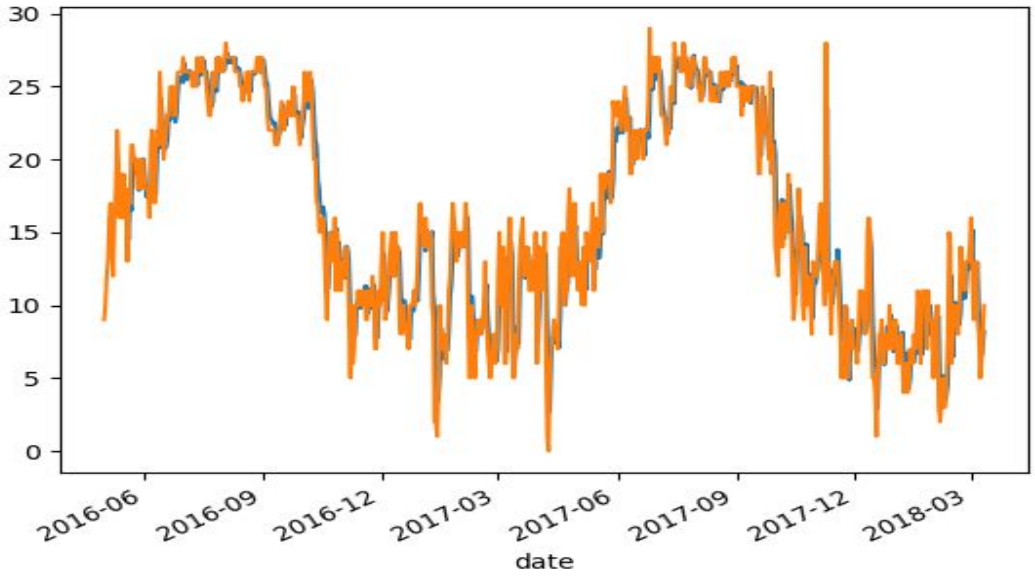
Original Data



Converting the Data into a Stationary Series



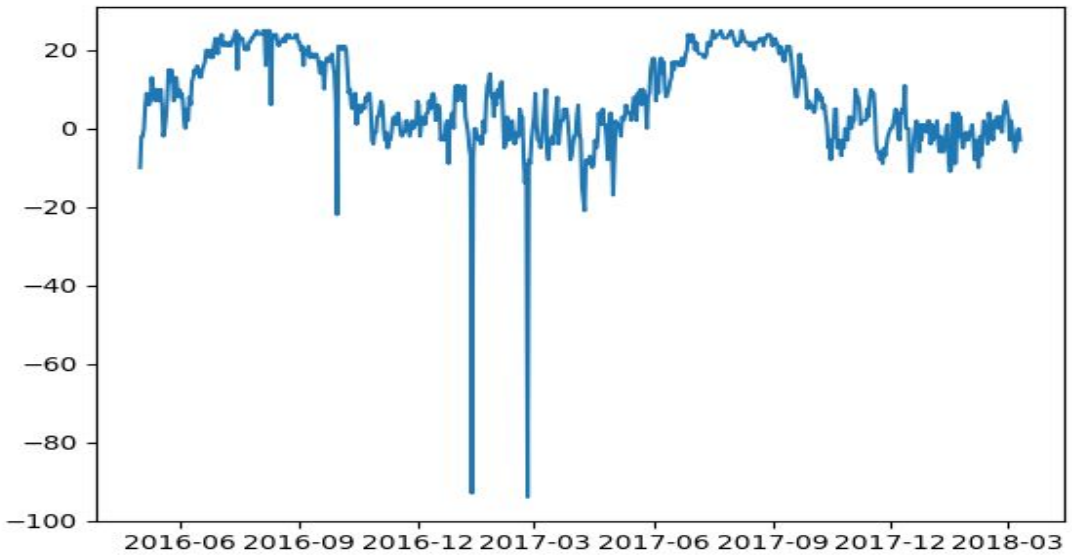
Predictions with the Original Data



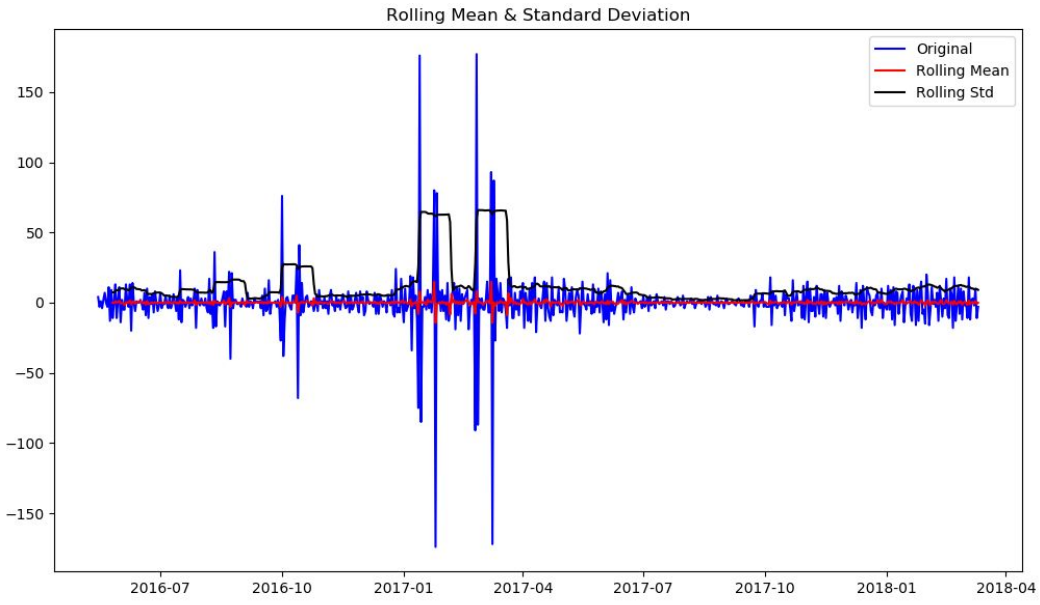
- **RMSE Value (Prediction Accuracy) : 2.290671862**

Minimum Dew

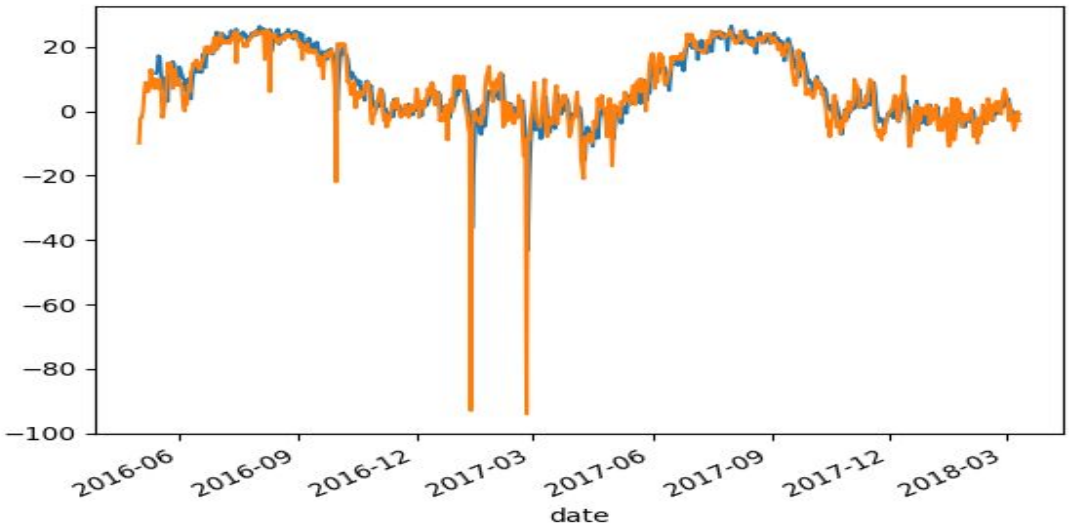
Original Data



Converting the Data into a Stationary Series



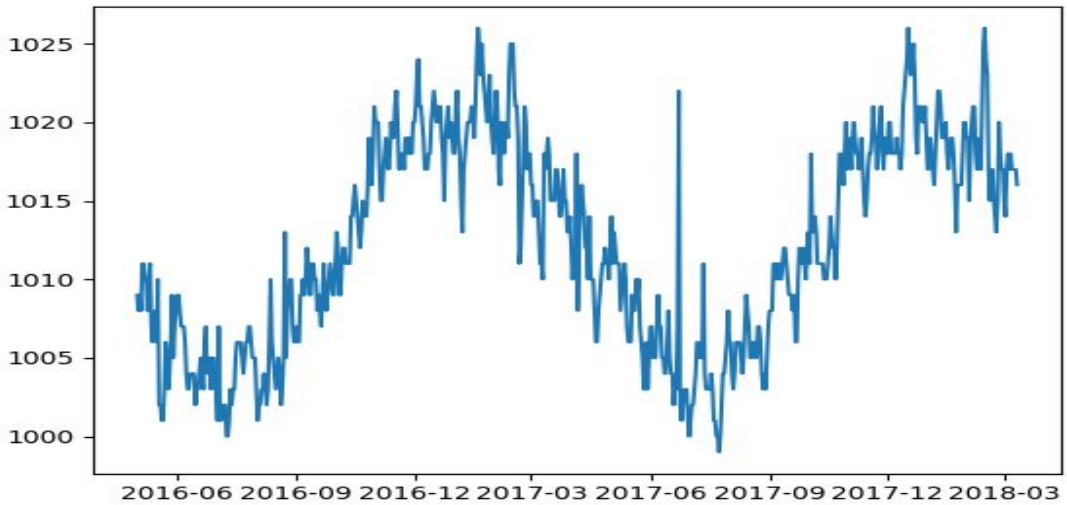
Predictions with the Original Data



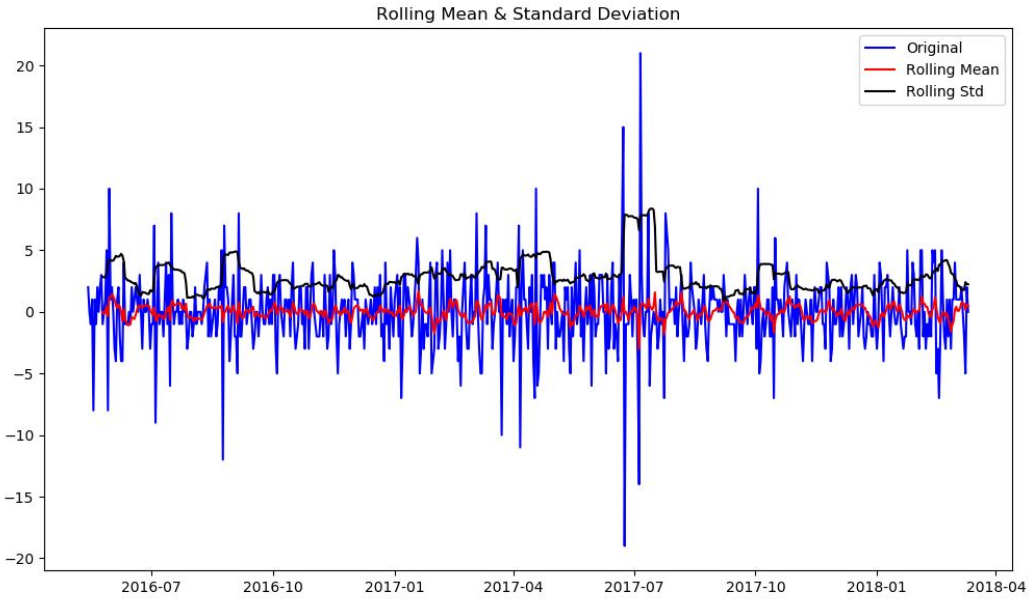
- **RMSE Value (Prediction Accuracy) : 6.827054357**

Maximum Pressure

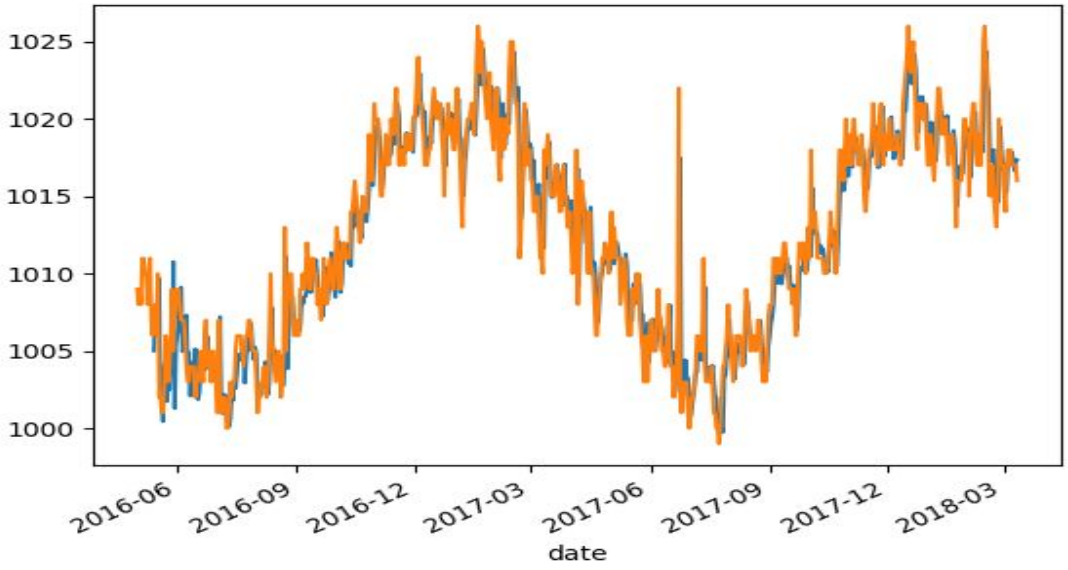
Original Data



Converting the Data into a Stationary Series



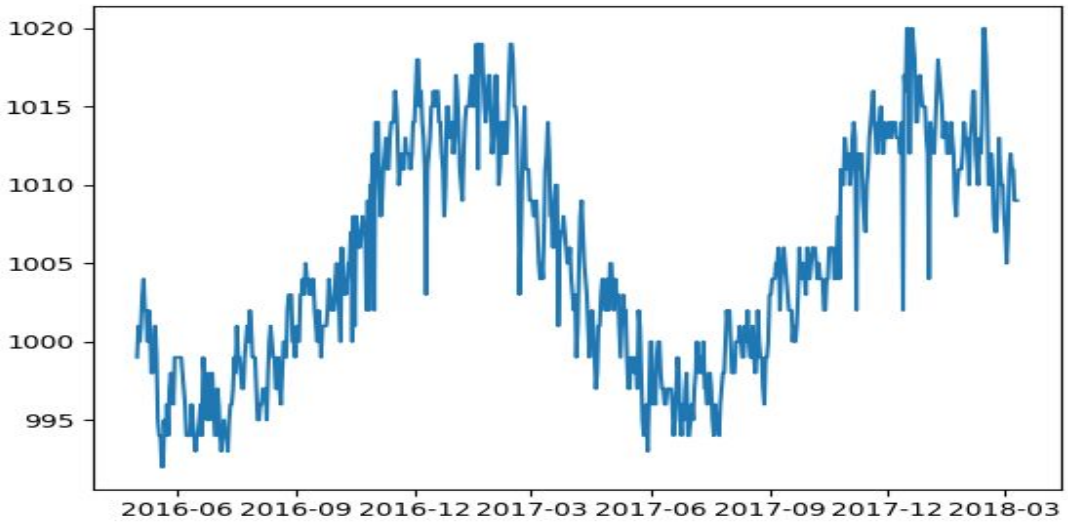
Predictions with the Original Data



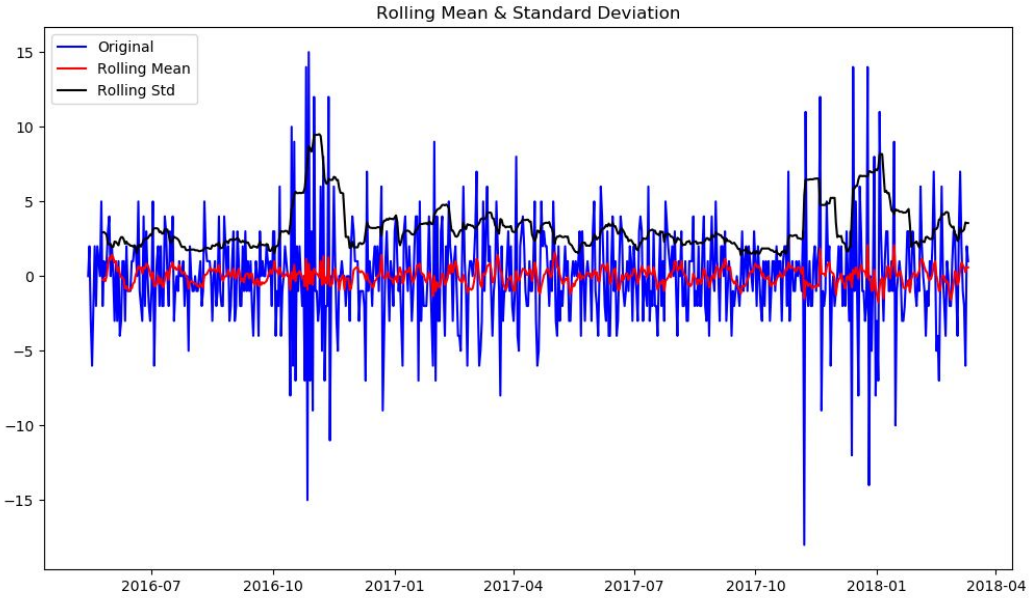
- **RMSE Value (Prediction Accuracy) : 2.066841852**

Minimum Pressure

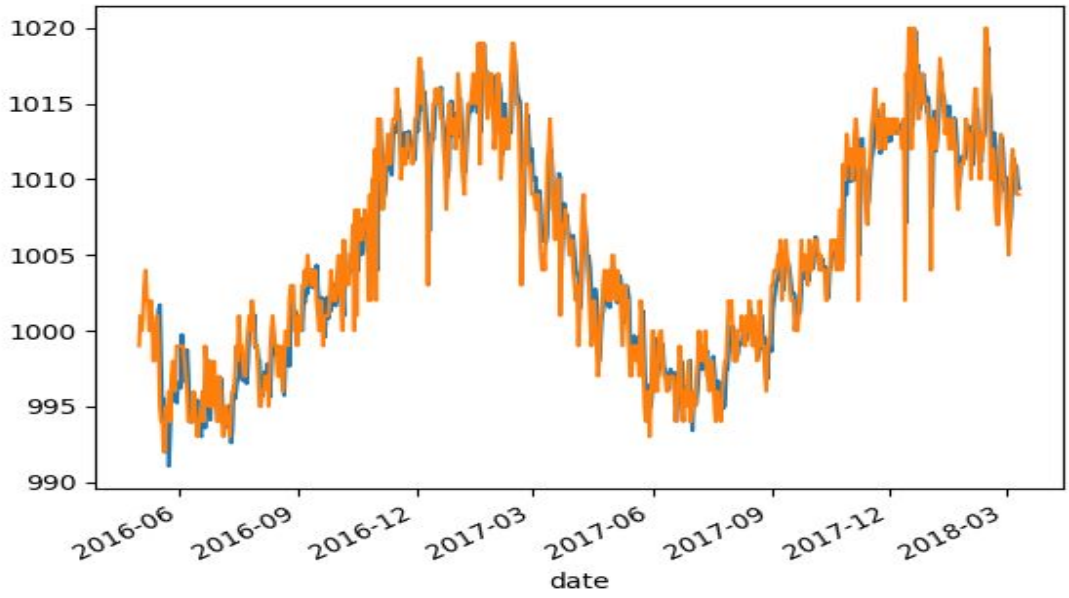
Original Data



Converting the Data into a Stationary Series



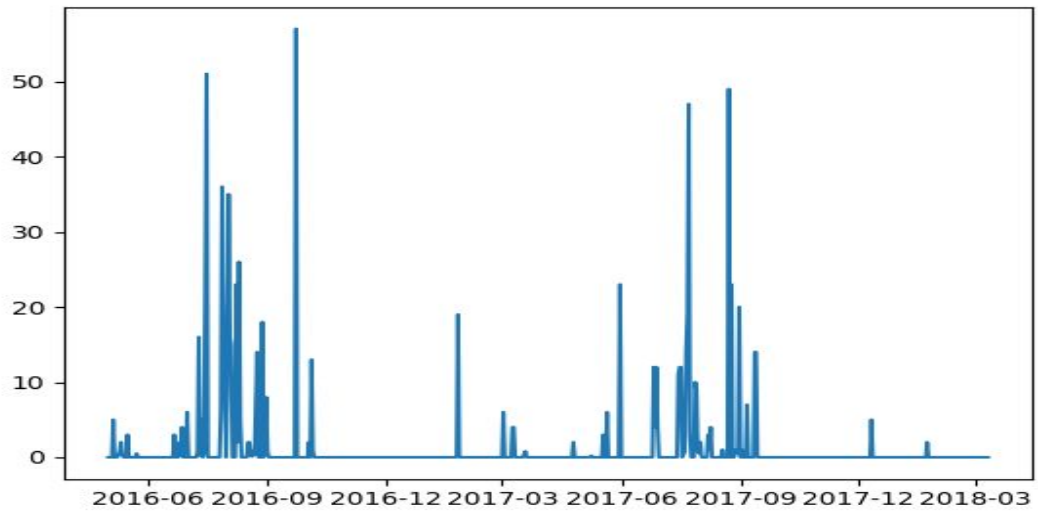
Predictions with the Original Data



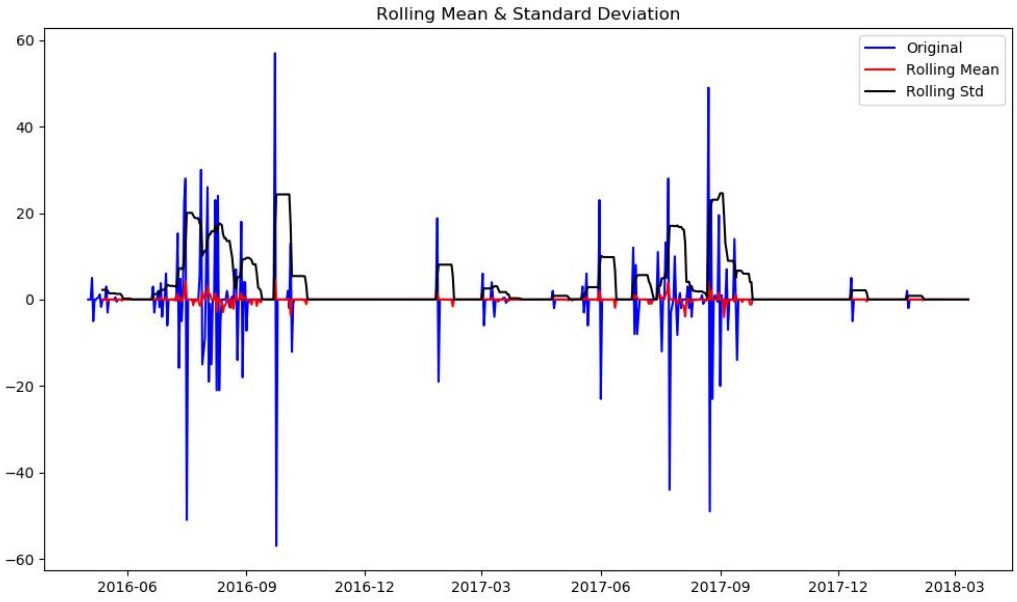
- **RMSE Value (Prediction Accuracy) : 2.269115739**

Precipitation

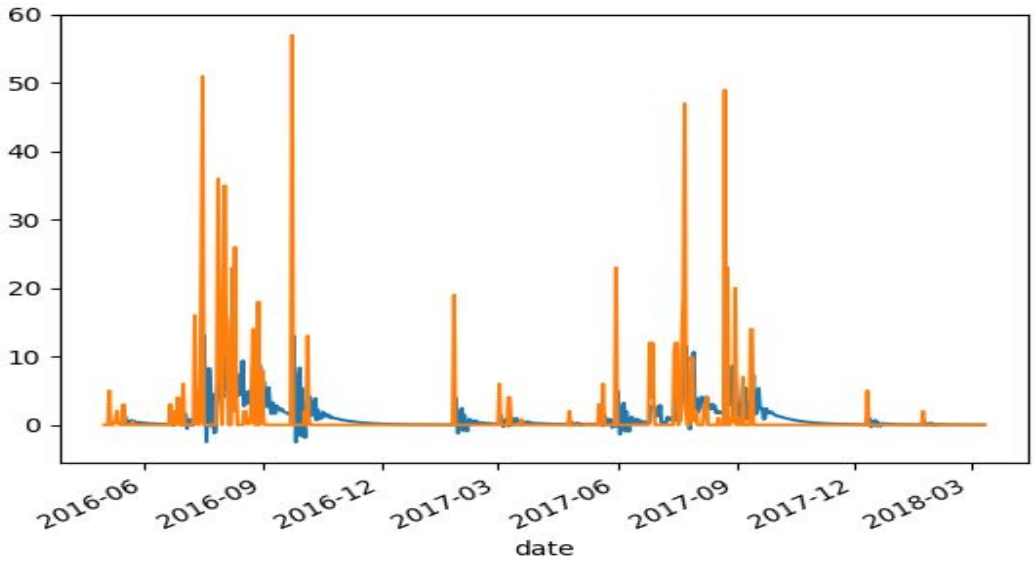
Original Data



Converting the Data into a Stationary Series



Predictions with the Original Data



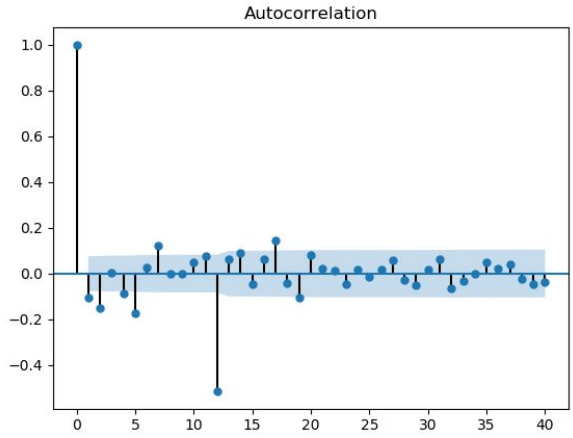
- **RMSE Value (Prediction Accuracy) : 5.133352687**

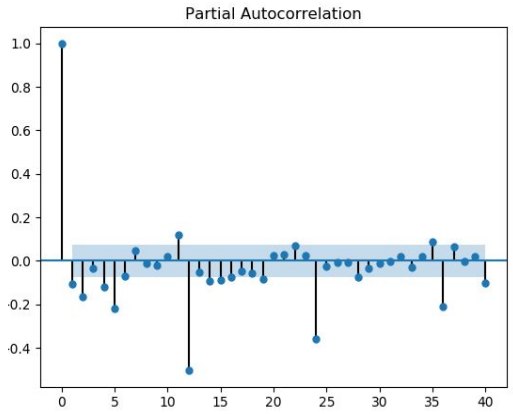
Findings

- So, our model gives us very high-quality predictions. Using these, we can do a lot of good. Such as predicting the rainfall of a dry region, which helps farmer ultimately for growing crops. Thus our model can help a farmer by saving him from huge losses during cultivation time by knowing about the rainfall from days before and taking respective measures.

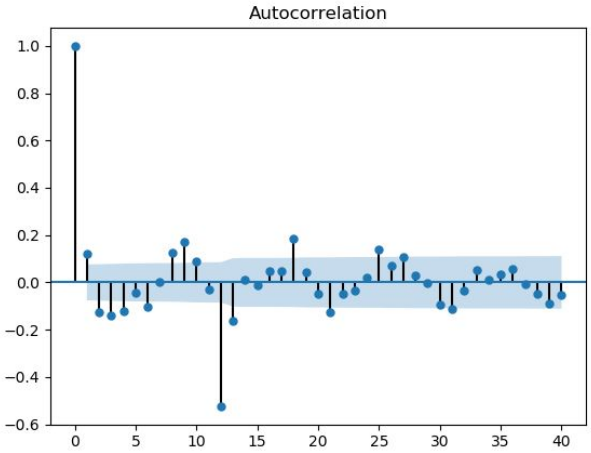
- Plotting the data of all the 12 parameters we could clearly see that the data pattern was repeating with frequency 2 in the given time interval, this makes sense because the data is collected is over the period of 2 years and we expect the weather parameter values to repeat on a yearly basis based on different seasons over the course of a year. So, the data is seasonal. Thus a normal ARIMA model would fail in this case and that's why need its variant, Seasonal ARIMA Model.
- We get an idea of how the Model works by converting the Data into a stationary series by removing things like trend, seasonality, etc. and plotting the stationary series and the corresponding rolling mean and rolling standard deviation.
- So, as expected the Seasonal ARIMA Model fits the original data very well, so it gives very less RMSE values, an accuracy measure. [when scaled, very low RMSE corresponds to very low error percentages which in turn corresponds to High Accuracy]
- Both the methods of finding a model, using an automated algorithm or choosing the model yourself seem to be in agreement in the case of all of the 12 variables.
- You choose the model parameters yourself by looking at the number of outliers in the ACF & PACF. Below are the plots of them :

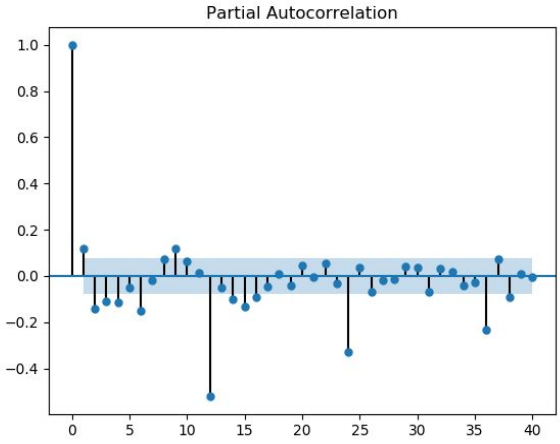
Mean Temperature



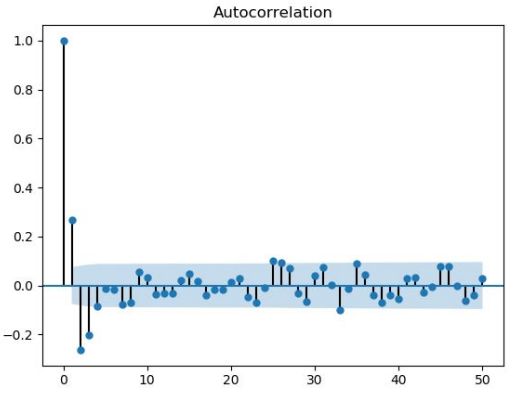


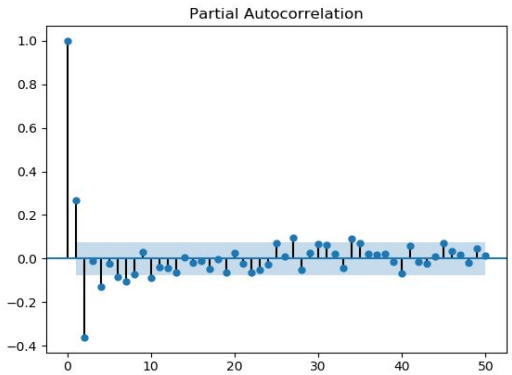
Mean Dew



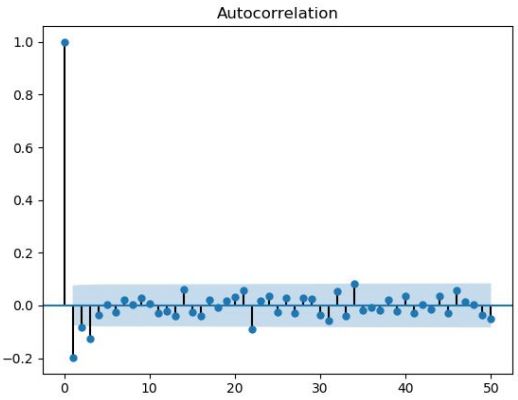


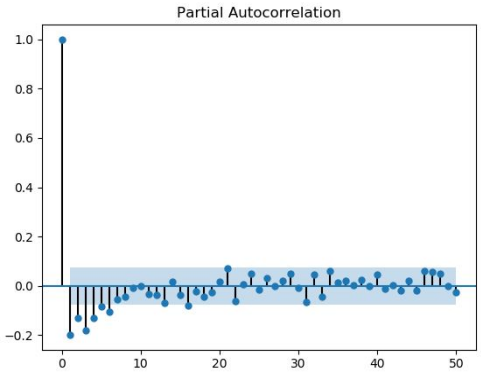
Mean Pressure



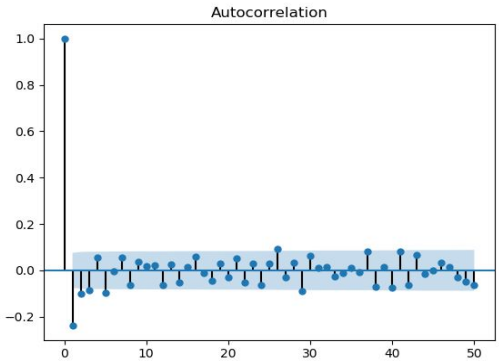


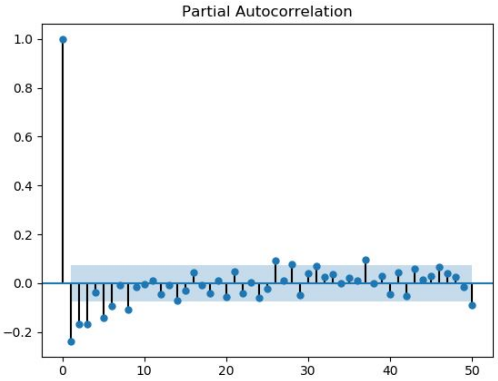
Max Humidity



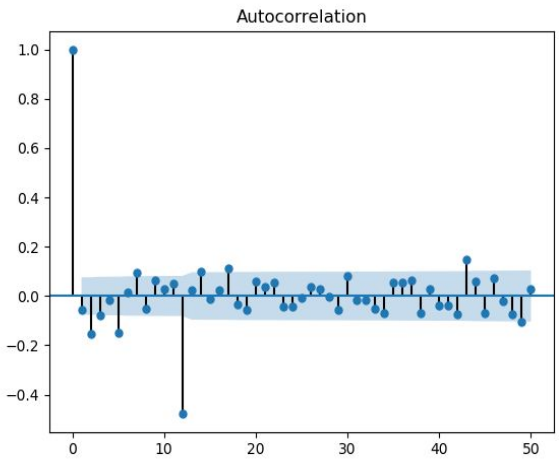


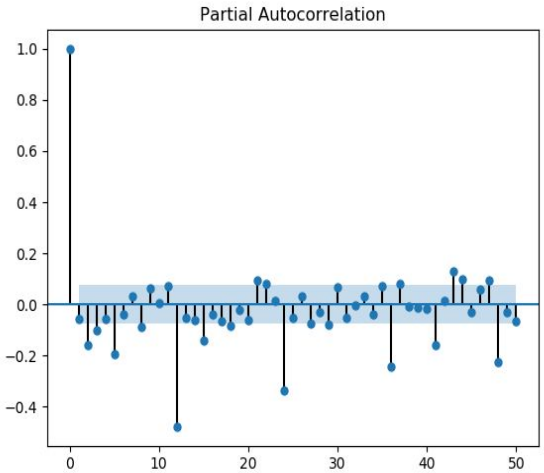
Minimum Humidity



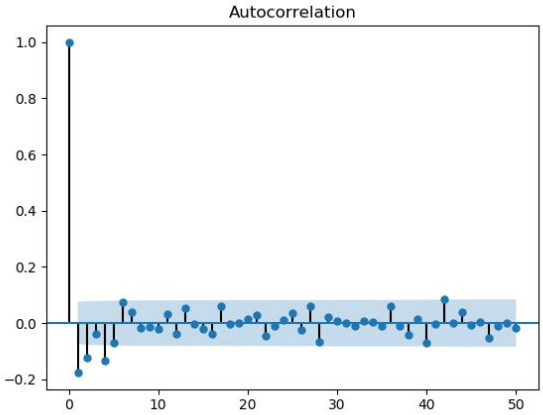


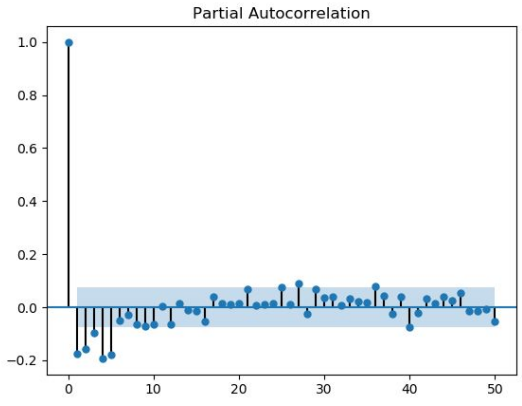
Maximum Temperature



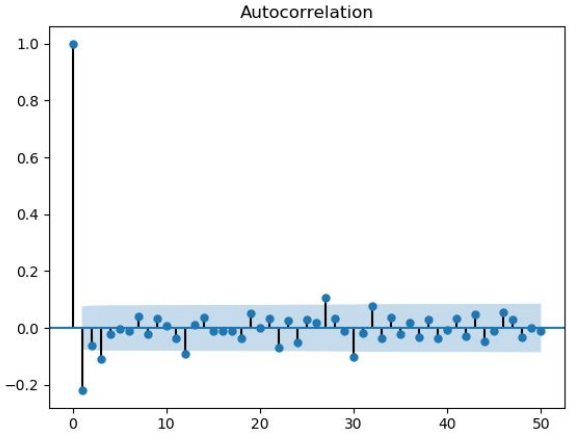


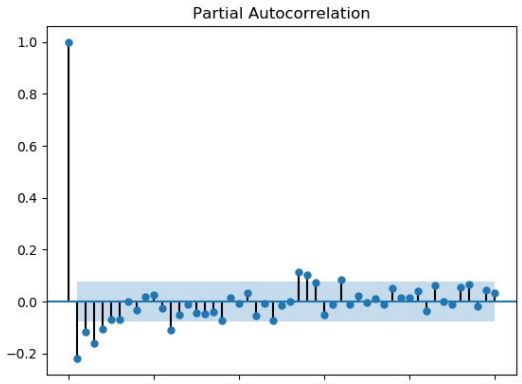
Minimum Temperature



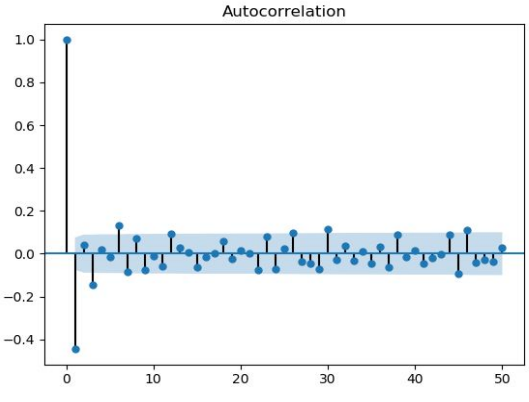


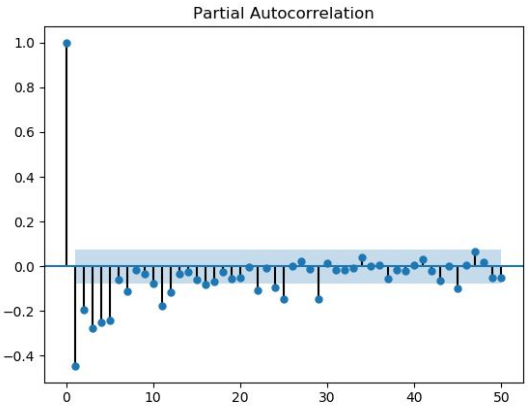
Maximum Dew



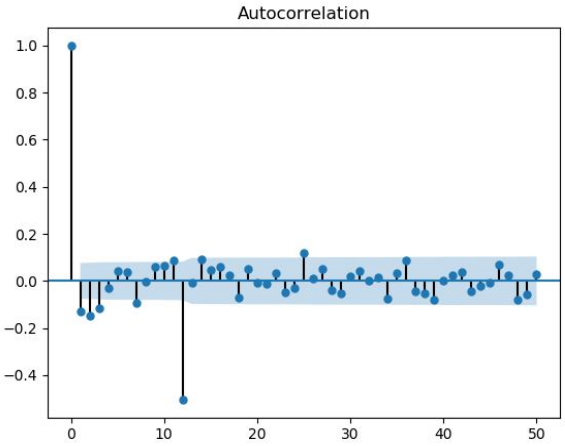


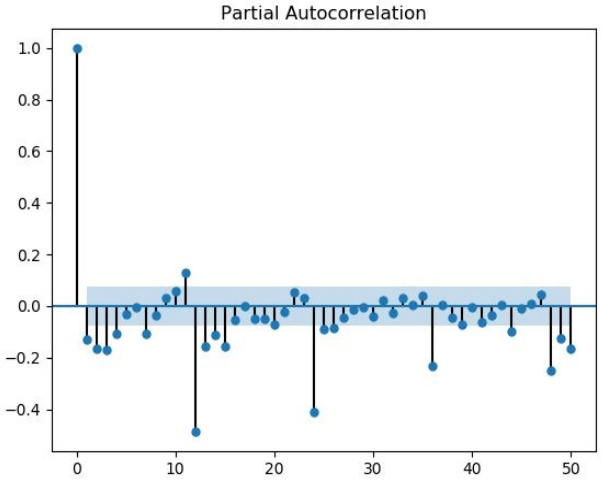
Precipitation



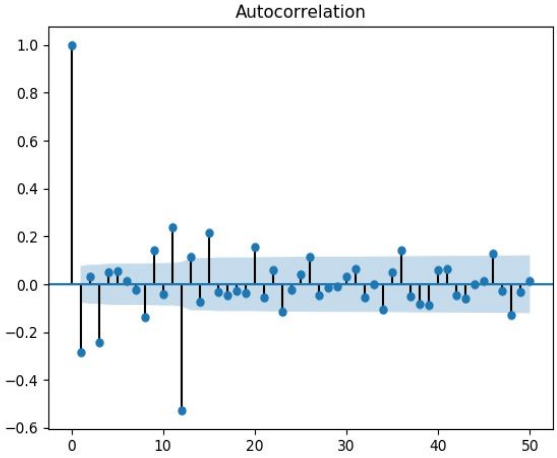


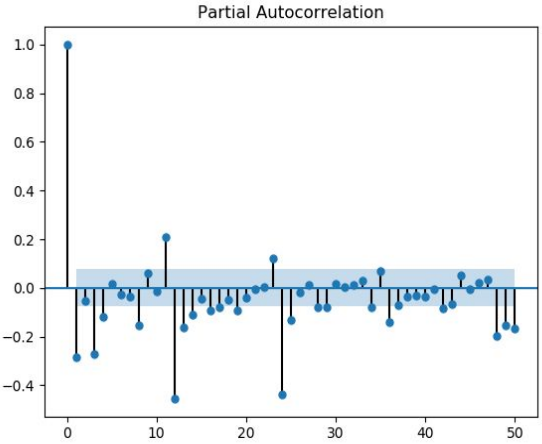
Maximum Pressure





Minimum Pressure





References

- **Paper:**
  Raman, R.K., Sathianandan, T.V., Sharma, A.P. and Mohanty, B.P., 2017. Modeling and Forecasting Marine Fish Production in Odisha Using Seasonal ARIMA Model. *National Academy Science Letters,* 40(6), pp.393-397.
- **Data-Set:**
  https://www.kaggle.com/rajatdey/jaipur-weather-forecasting