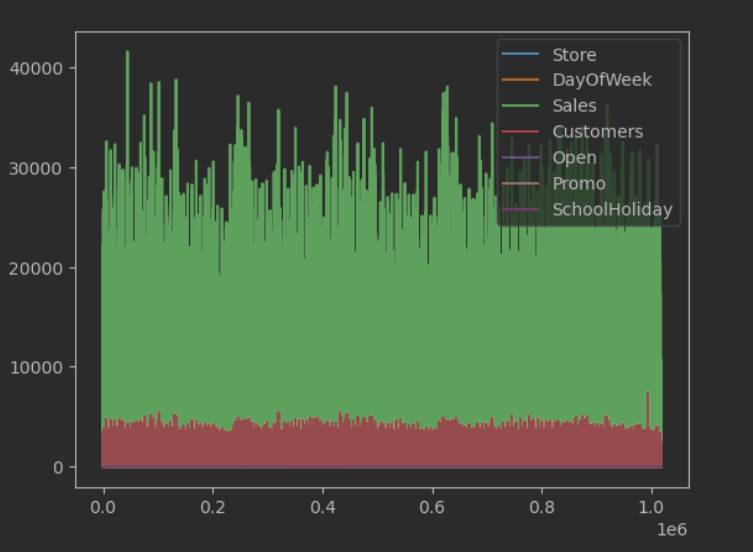
In this analysis, I am going to discuss the ARIMA and SARIMA models, and going to use Rossmann Store Sales from Kaggle.

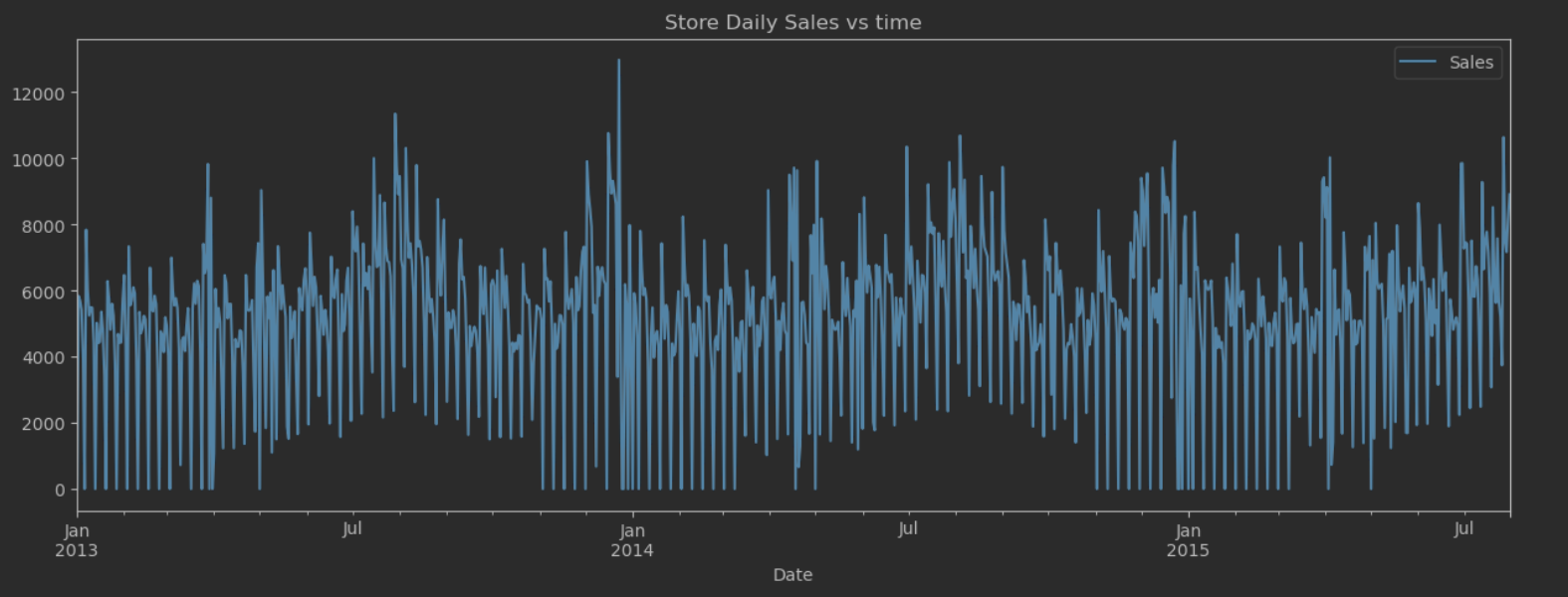
* Data Description

Rossmann operates over 3,000 drug stores in 7 European countries. Currently, Rossmann store managers are tasked with predicting their daily sales for up to six weeks in advance.



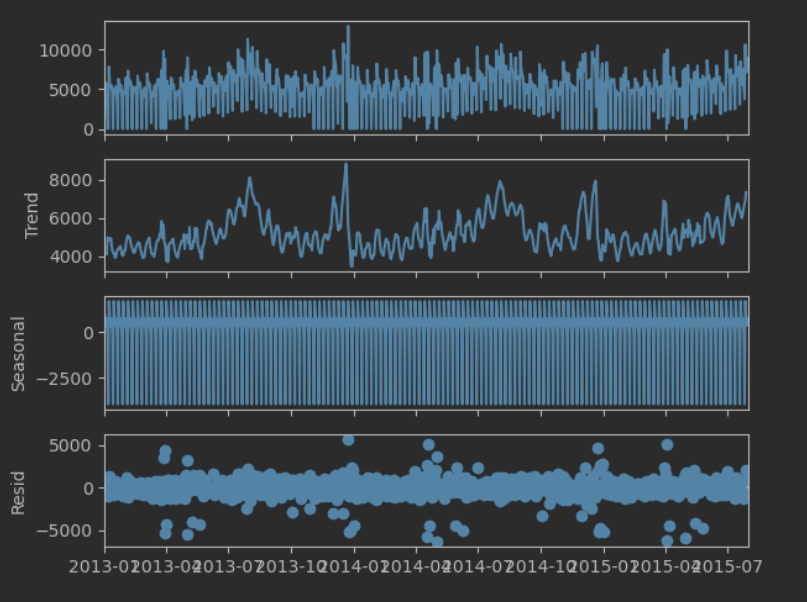
* Build and Analysis of the Time Series Data

We are going to create a sales time series from the store-sales data-frame

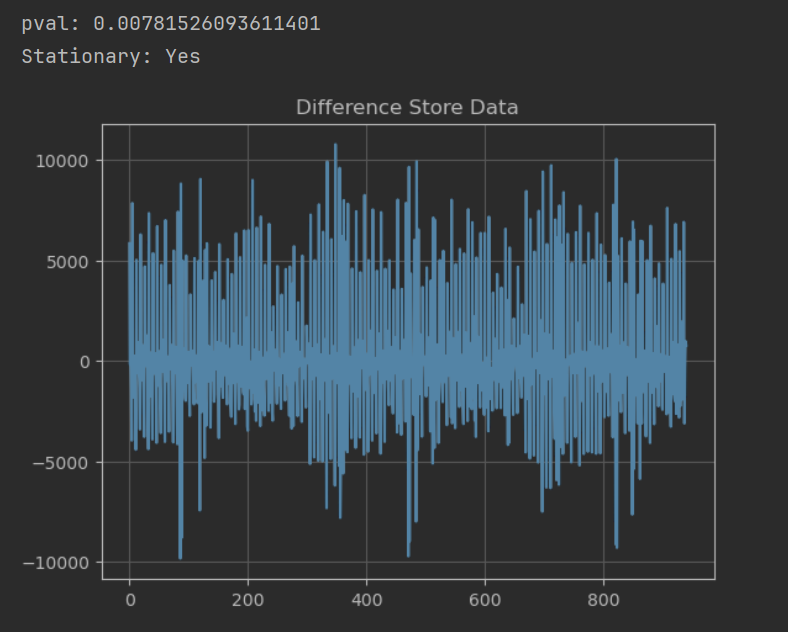


From the plot we can see there is not any increasing or decreasing trend, but there some zero sales cycles at the begging and the end of the year, because of that we need to check if the time series stationary, which will allow us to chose p and q autocorrelation metrics.

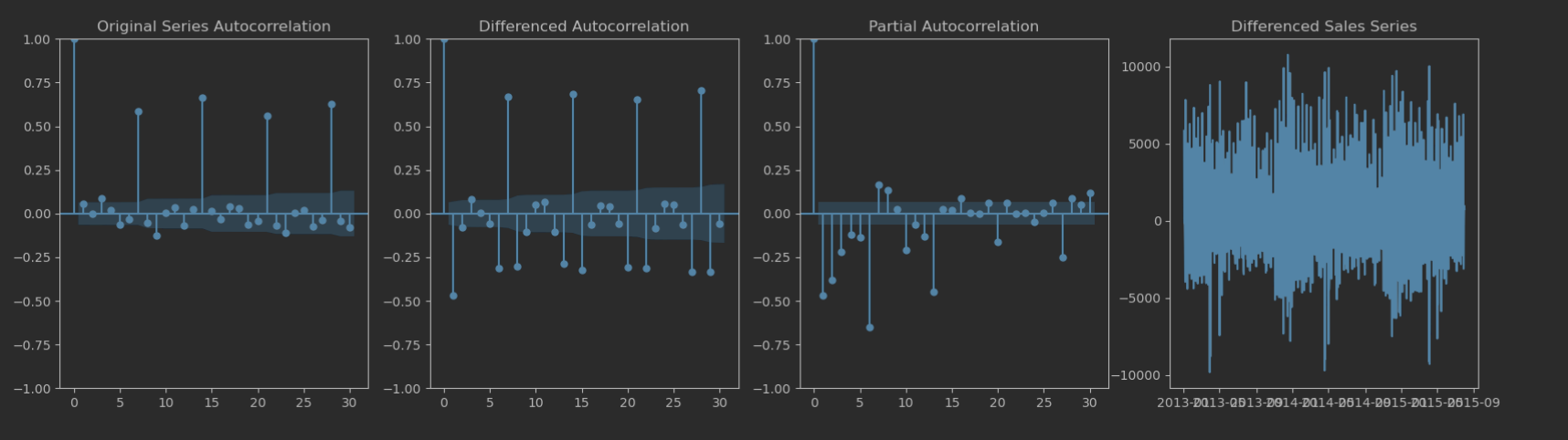
* + Decomposition



We can see there is not any seasonality, the trend is neither increasing nor decreasing, it looks like is stationary, but we need to check for that next.



* + Autocorrelation and Partial Autocorrelation



From the original timeseries autocorrelation we can identify 2 as a value for p, and from partial autocorrelation we can identify 5 as value for q, as the timeseries is stationary we will use zero for d.

Those autocorrelation parameters are going to be used in ARIMA and SARIMA models.

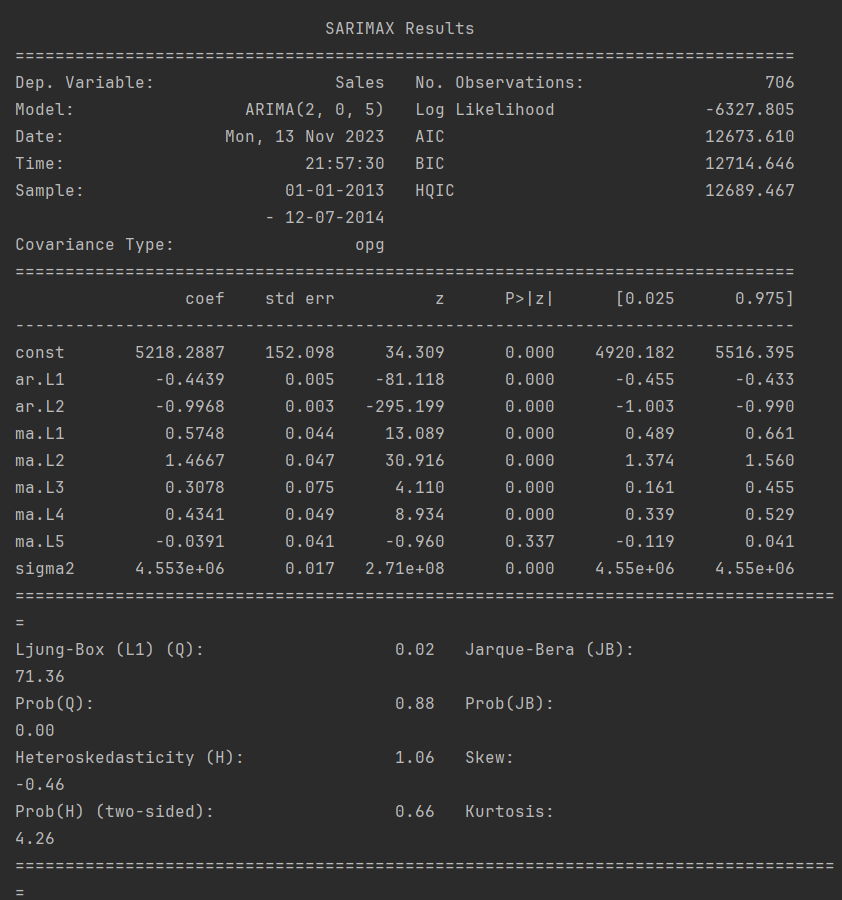
p = 2

q = 5

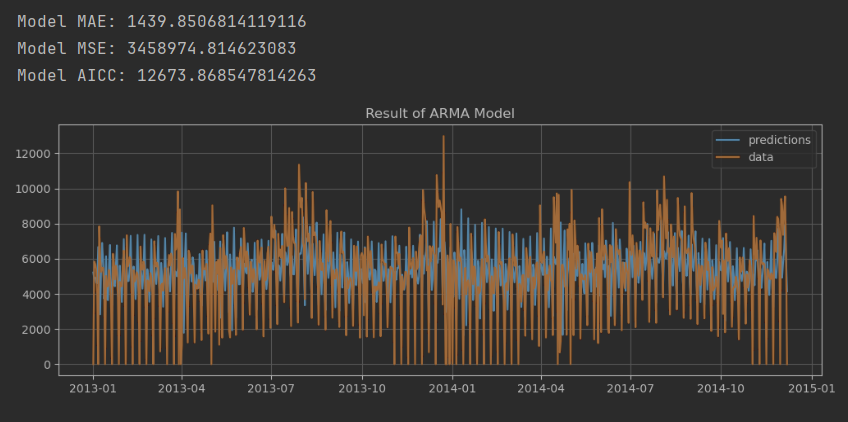
d = 0

* Forecasting with ARIMA model

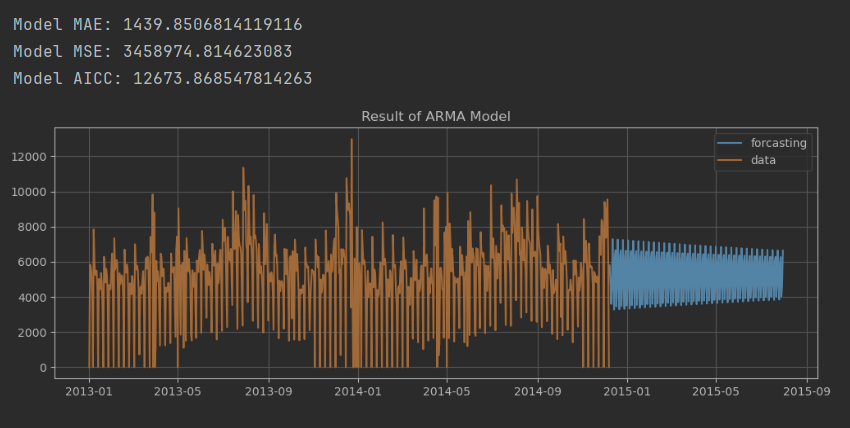
We are going to use all periods to the forecast & predict of the sales timeseries.



* + Prediction

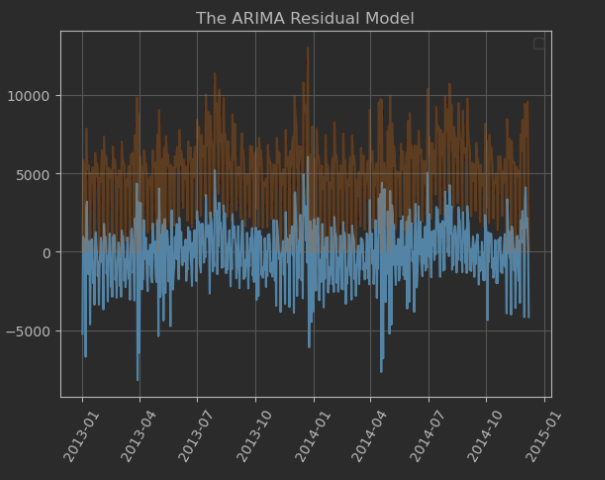


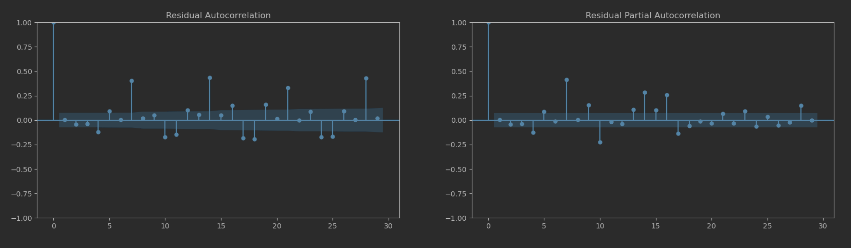
* + Forecasting



From the above plots we can see the forecasting sales are around average 5000 - 6000 per month

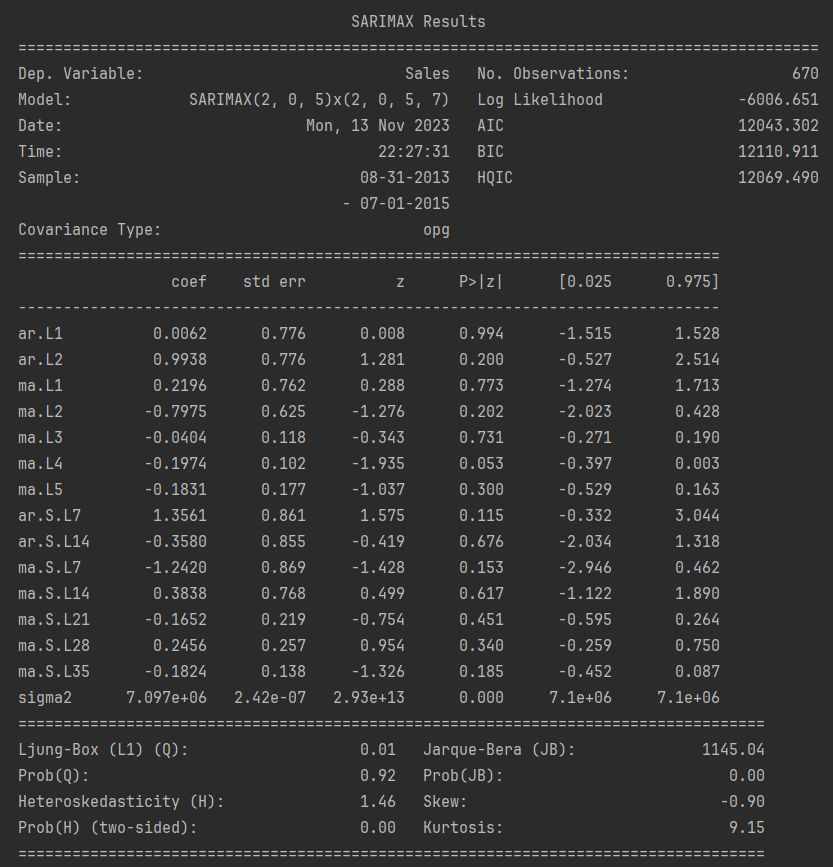
* + Residuals and Residual Autocorrelation



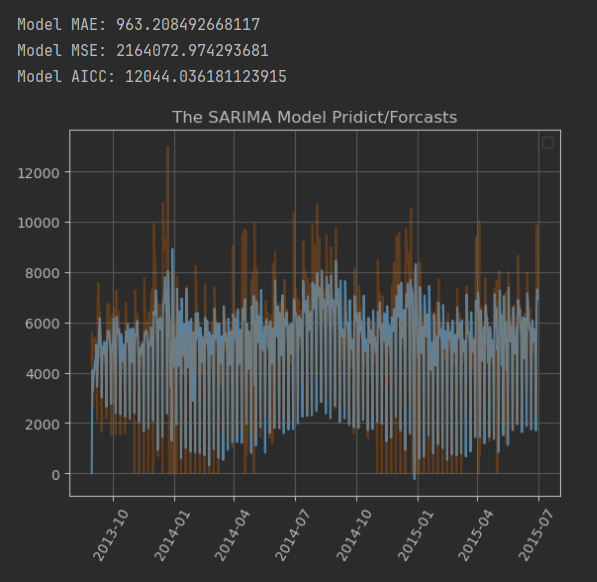


* Forecasting with SARIMA model

For this model we will try to forecast a 30-day period and going to use 700 days history.

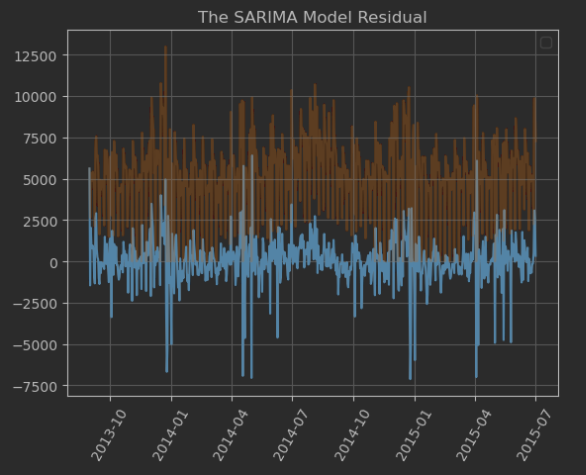


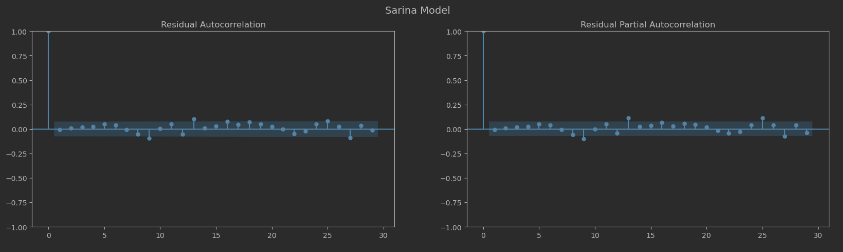
* + Forecasting



From above plot we can see the forecast for the following moth is around 4000 sales, which is much lower than the long run forecast by ARIMA model.

* + Residuals and Residual Autocorrelation





The “residuals” in a time series model are what is left over after fitting a model. For many (but not all) time series models, the residuals are equal to the difference between the observations and the corresponding fitted values

Having a negative residual means that the predicted value is too high, similarly if you have a positive residual, it means that the predicted value was too low.

The Conclusion is the Sarina model looks better according to the model’s metrics (MAE, residual, residual autocorrelation, etc.)

Regarding the forecast of the sales in the short term is 4000 per month which is much lower to the long-term period with average sales around 5000 per month.

Regarding the residuals in our case the residuals are zero and very close to zero, but ARIMA model has higher variance. So our model’s forecast looks good.