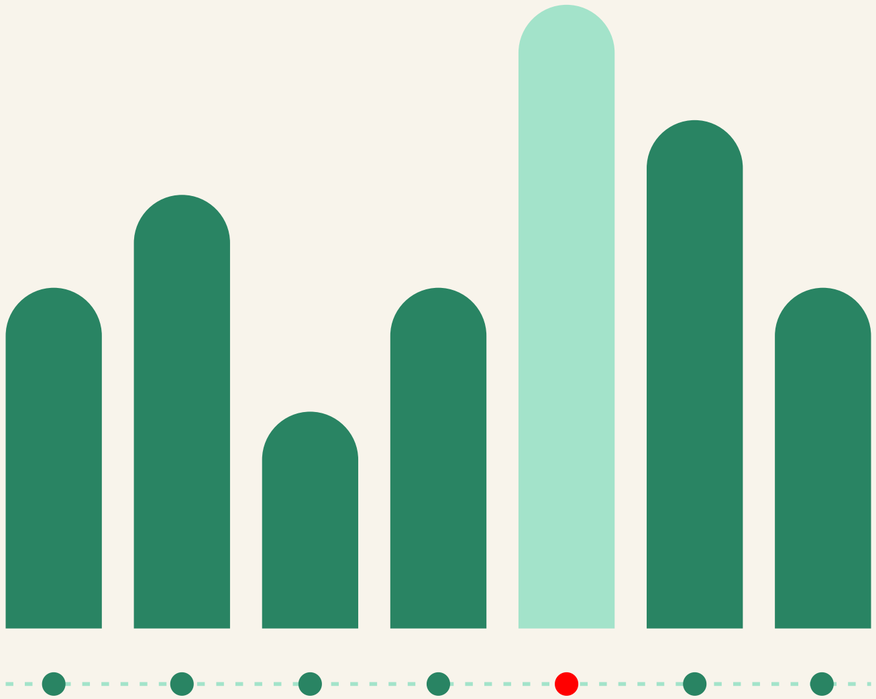
SUPERSTORE DATA

TIMES SERIES SALES FORECASTING

- AMAN VERMA

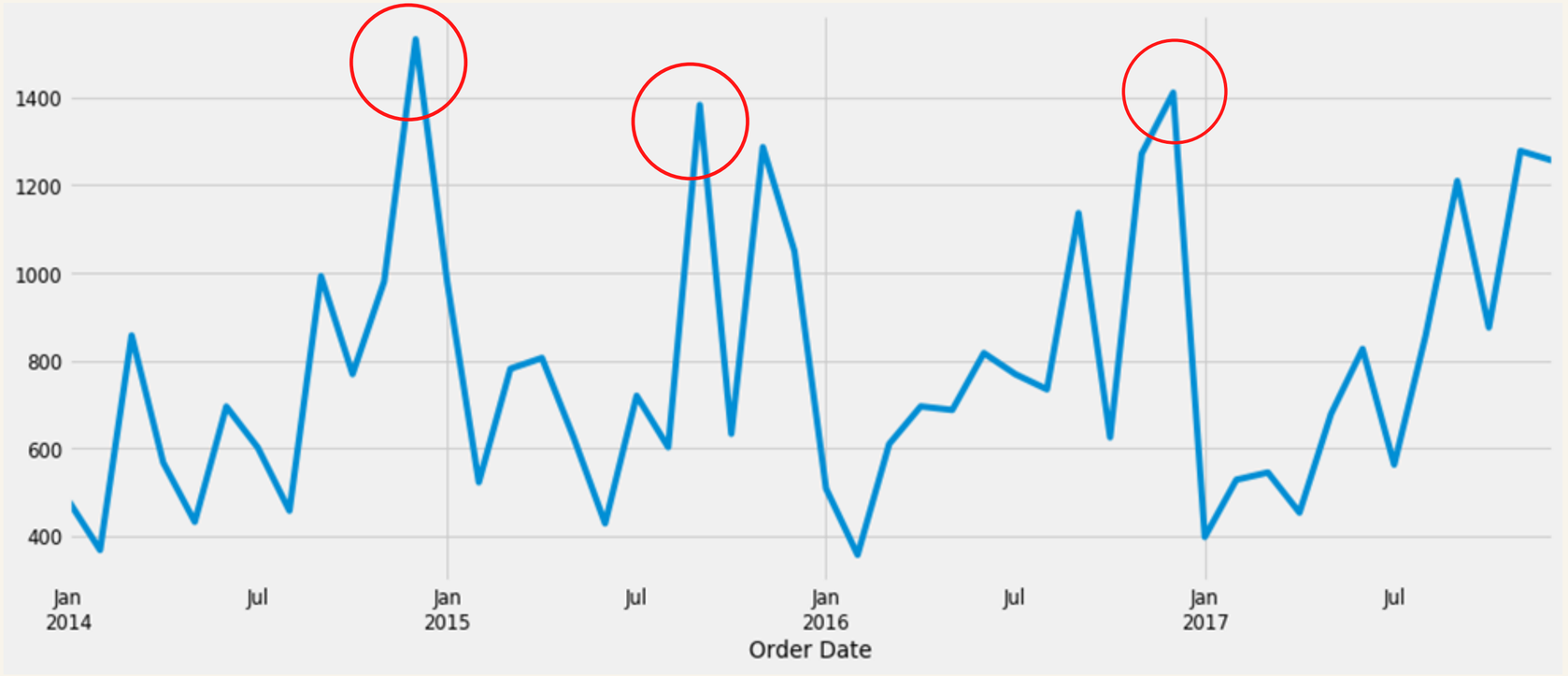
GitHub Link



WHAT IS TIME SERIES FORCATSING

Time series forecasting occurs when you make scientific predictions based on historical time stamped data.

"*Time series analysis involves developing models that best capture or describe an observed time series in order to understand the underlying causes. This field of study seeks the “why” behind a time series dataset*".



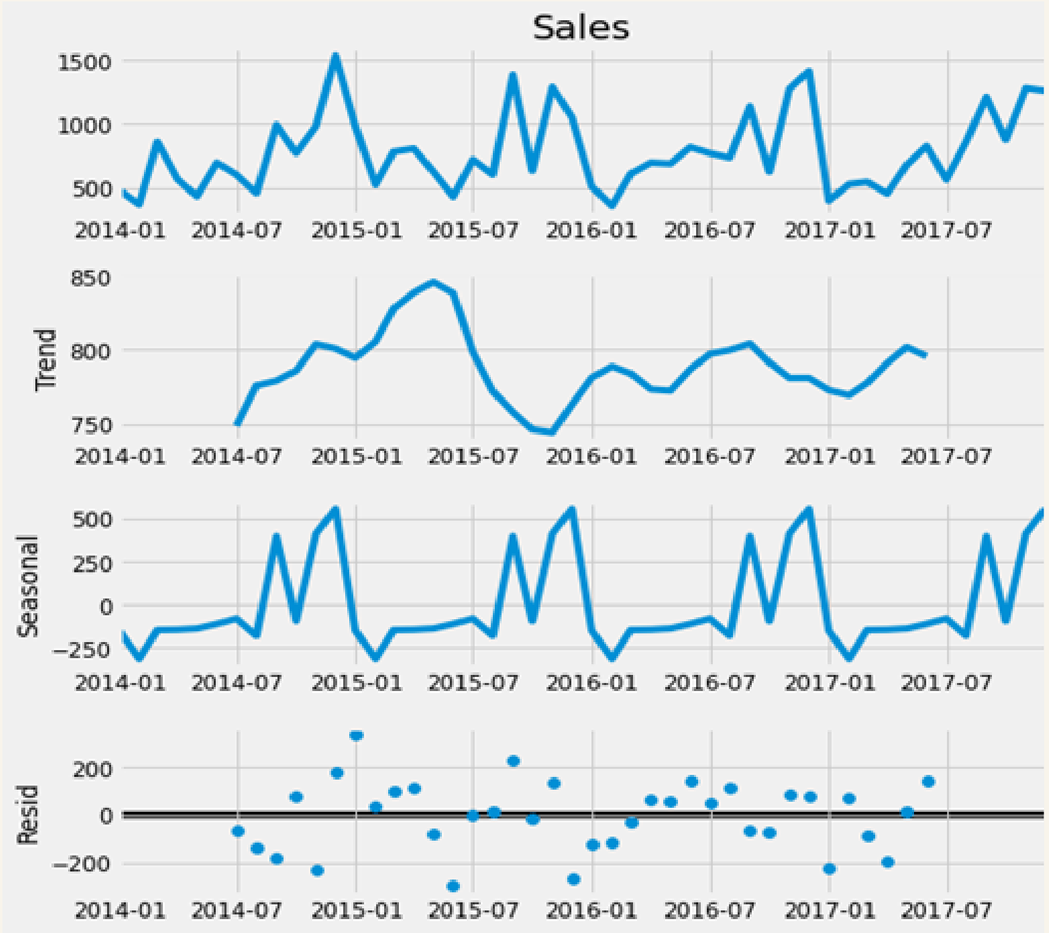
- Non Seasonal Order - Jan 2018 = f(Dec 2018)
- Seasonal Order - Jan 2018 = f(Jan 2017) -> Red Circles Are Seasonal Order

Components of Time Series

Time series analysis provides a body of techniques to better understand a dataset.
Perhaps the most useful of these is the decomposition of a time series into 4 constituent parts:

1. TREND = LONG TERM SMOOTH MOVEMENT OF A DATA - UPWARD OR DOWNWARD DIRECTION.
2. SEASONAL = PERIODIC FLUCTATIONS IN THE DATA, WHICH LESS THAN 1 YEAR
3. CYCLICAL = PERIODIC FLUCTATIONS IN THE DATA, WHICH MORE THAN 1 YEAR
4. IRREGULAR = RANDOM MOVEMENT OF A TIME SERIES DATA

Decomposition of Series in Time Components



Residual Series = Actual Series - Trend - Seasonal

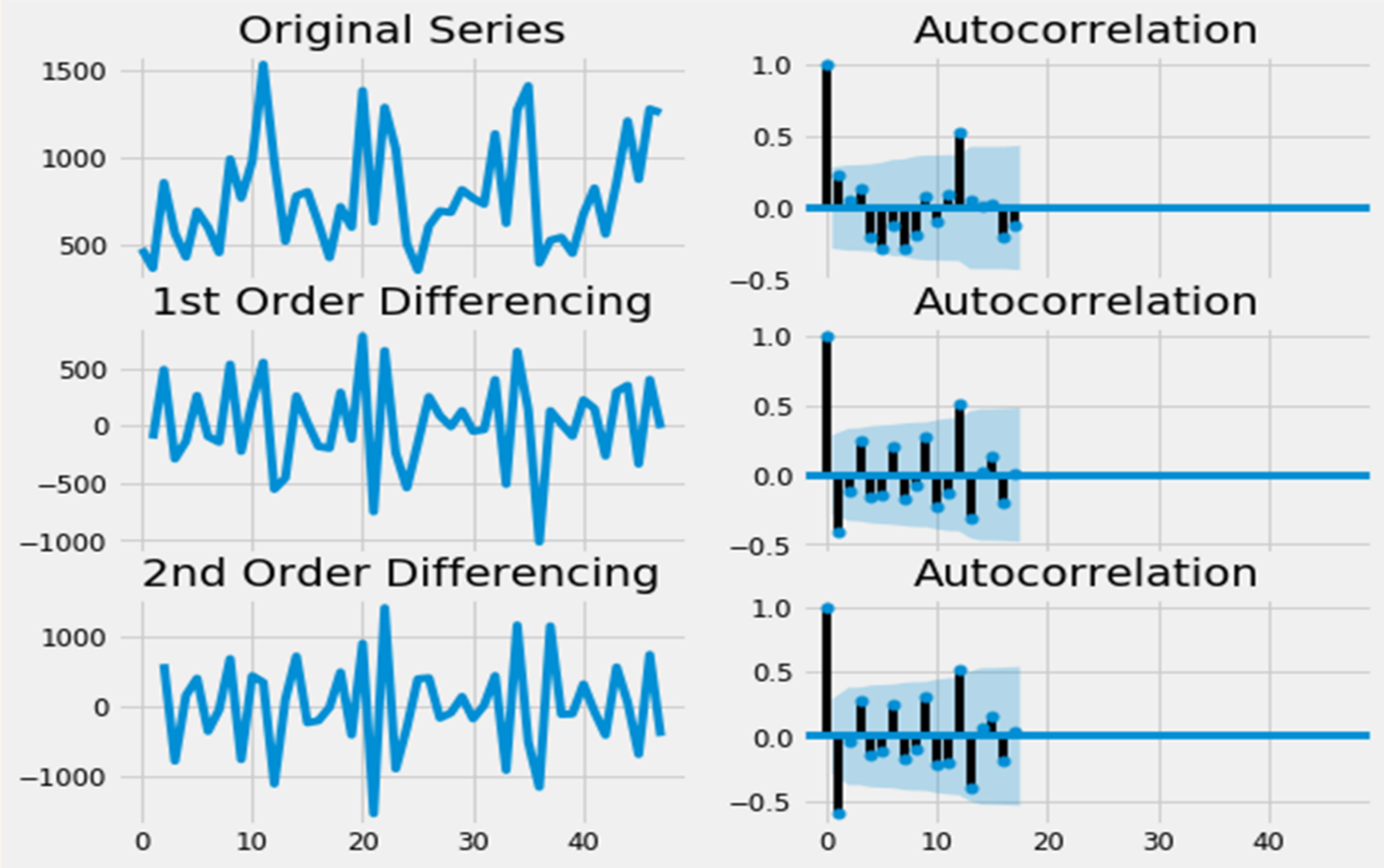
Actulal Series = Trend + Seasonal + Residual

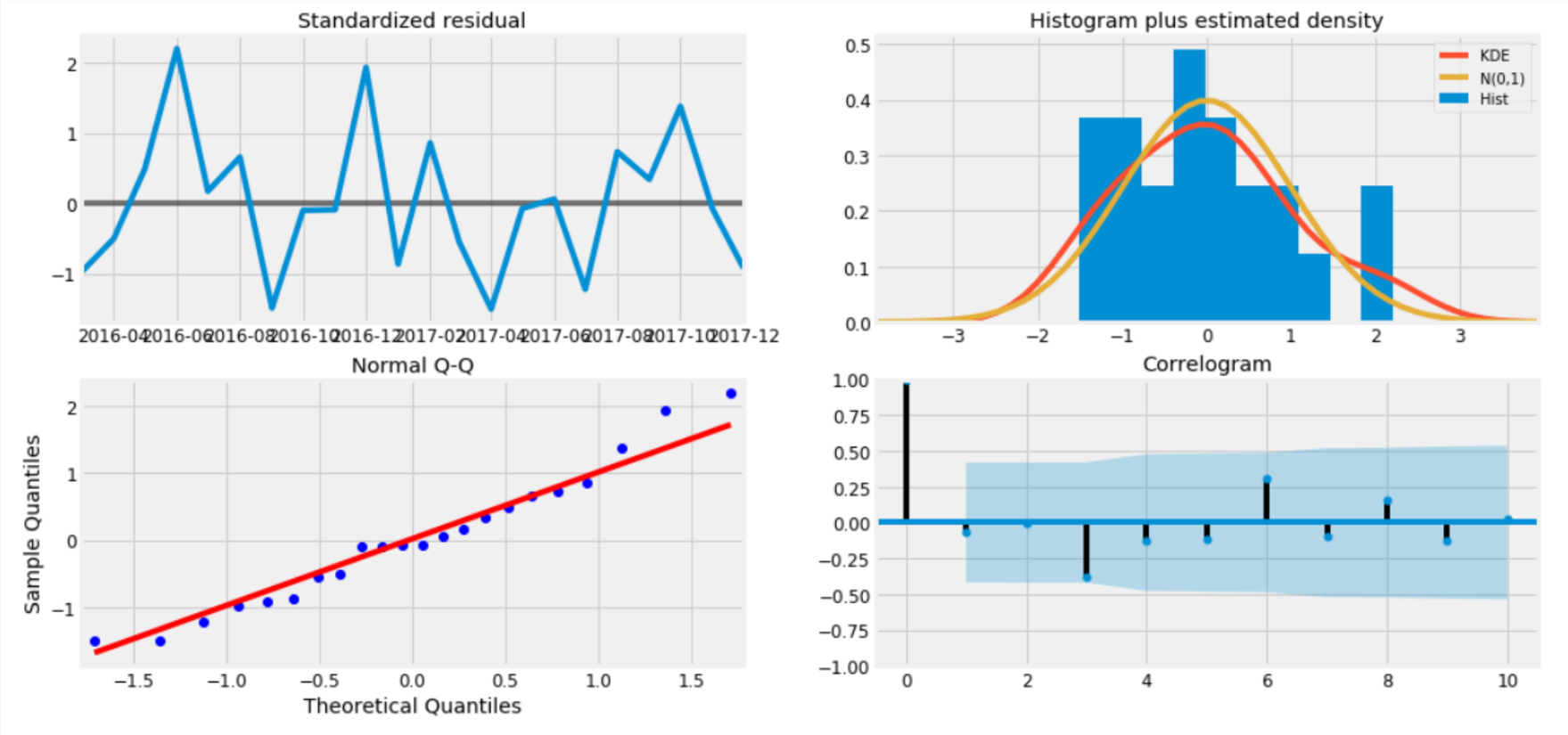
Time series forecasting with ARIMA

ARIMA stands for Auto Regressive Integrated Moving Average. There are seasonal and Non-seasonal ARIMA models that can be used for forecasting.

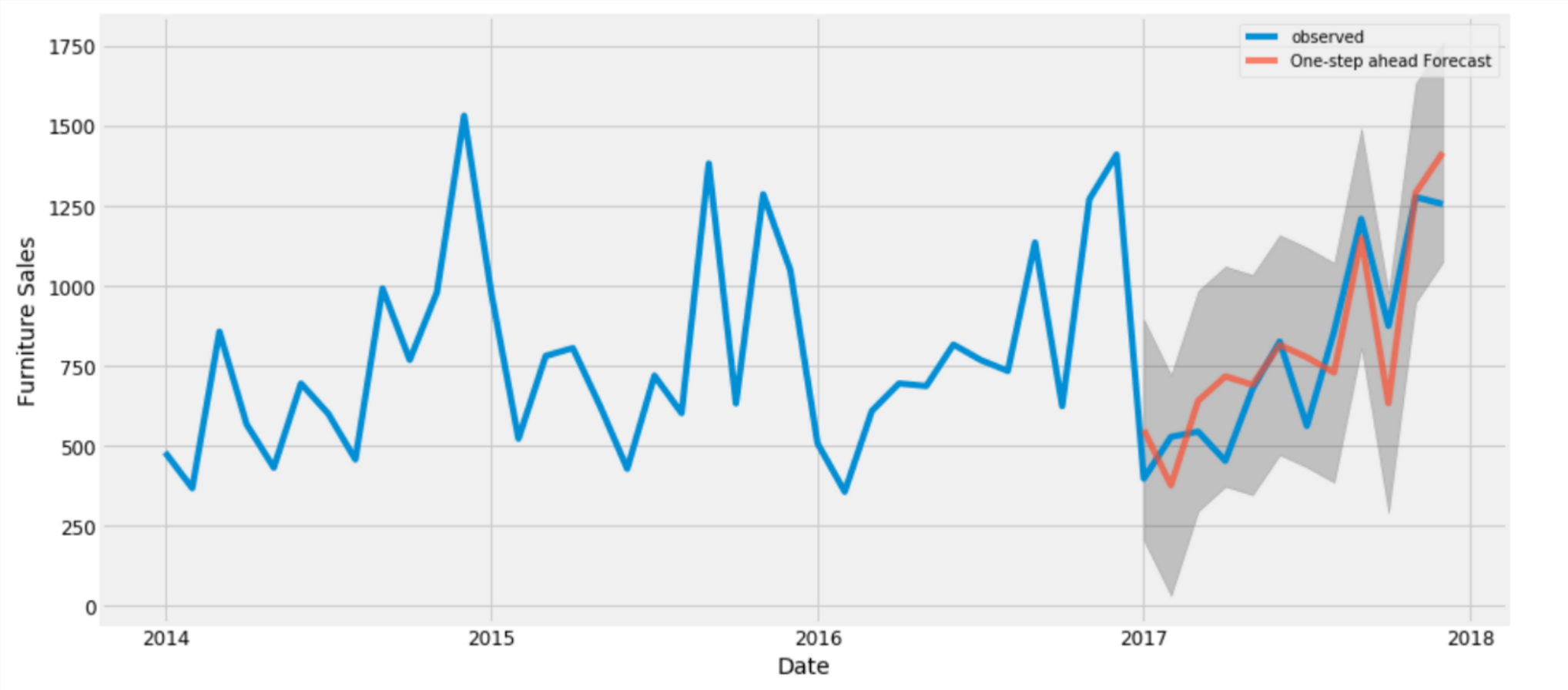
- Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
- Integrated (I): represents the differencing of raw observations to allow for the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
- Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

To sum up, these three parameters ARIMA(p, d, q)account respectively for seasonality, trend, and noise in data.

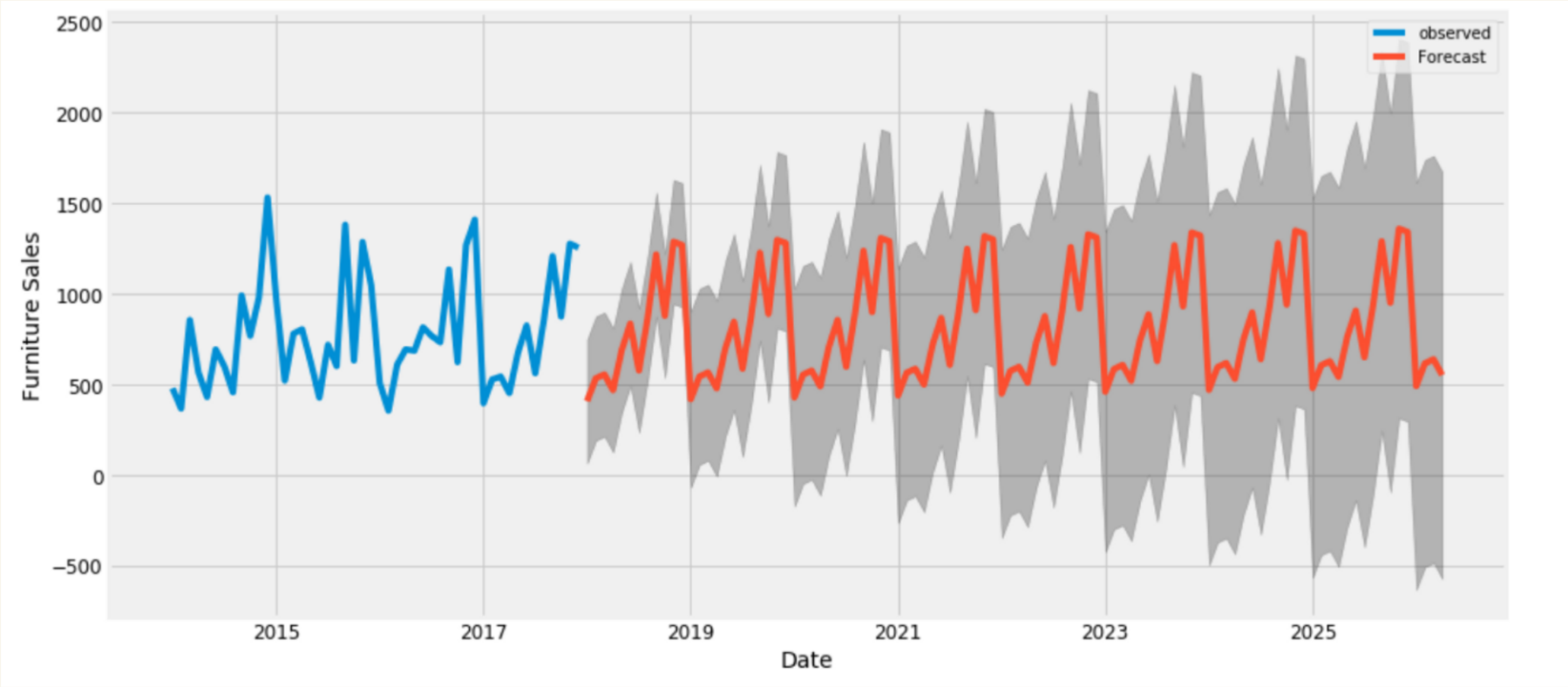

Original Series
1500
1000
500
1st Order Differencing
500
0
−500
−1000
2nd Order Differencing
1000
0
−1000
0
10
20
30
40
Autocorrelation
1.0
0.5
0.0
−0.5
Autocorrelation
1.0
0.5
0.0
−0.5
Autocorrelation
1.0
0.5
0.0
−0.5
0
10
20
30
40



- We should always run model diagnostics to investigate any unusual behavior.
- It is not perfect, however, our model diagnostics suggests that the model residuals are near normally distributed.



The line plot is showing the observed values compared to the rolling forecast predictions. Overall, our forecasts align with the true values very well, showing an upward trend starts from the beginning of the year and captured the seasonality toward the end of the year.



Our model clearly captured furniture sales seasonality. As we forecast further out into the future, it is natural for us to become less confident in our values. This is reflected by the confidence intervals generated by our model, which grow larger as we move further out into the future.