Use Case AGUC002

Time Series Analysis

Manufacturing Use Case

Agilytics Technologies Pvt Ltd.

Table of Contents

Introduction

In this use case, we will carry out time series analysis for a manufacturing operation. Time series analysis and modeling have many business and social applications. It is extensively used to forecast company sales, product demand, stock market trends, agricultural production etc.

Time Series Analysis

The basic idea of time series analysis is to decompose the original time series of sales, stock market trends etc. into several independent components. A business time series is divided into the following four components:

1. **Trend** – overall direction of the series i.e. upwards, downwards etc.
2. **Seasonality** – monthly or quarterly patterns.
3. **Cycle** – long-term business cycles.
4. **Irregular remainder** – random noise left after extraction of all the components.

The final series is produced by the interference of above four components.

Manufacturing Use Case Example

PowerBull is a tractor and farm equipment manufacturing. The company has shown a consistent growth in its revenue from tractor sales since its inception. However, over the years the company has struggled to keep the inventory and production cost down because of variability in sales and tractor demand. The management at PowerBull is under enormous pressure from the shareholders and board to reduce the production cost. Additionally, they are also interested in understanding the impact of their marketing and farmer connect efforts towards overall sales. In the same effort, they have hired you as a data science and predictive analytics consultant.

Eventually, you will develop an ARIMA model to forecast the demand for next year. Additionally, you will also investigate the impact of marketing program on sales by using an exogenous variable ARIMA model.

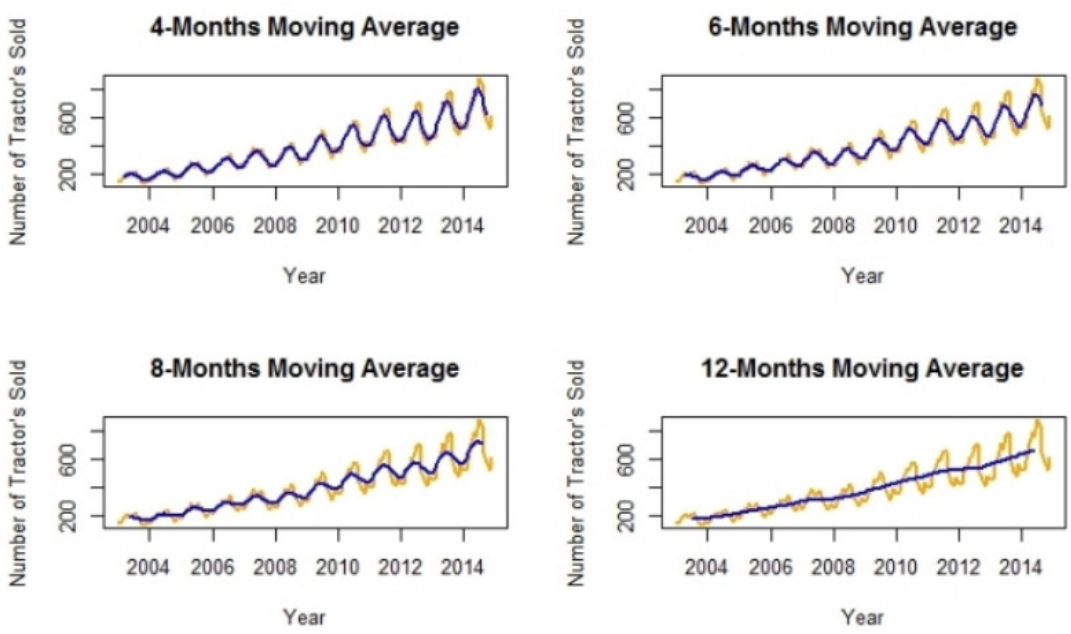
Deciphering the Trend

One of the usual procedures to decipher trends embedded in the tractor sales time series is by using Moving Average. The idea with moving average is to remove all the zigzag motion from the time series to produce a steady trend through averaging adjacent values of a time period. Hence, the formula for moving average is:

$$Moving\ Average = \frac{\sum_{i=-m}^{m} Y_{t+i}}{2m}$$

Now, if we consider 4, 6, 8 and 12 months of moving average of Tractor sales figures, we get following plots: -

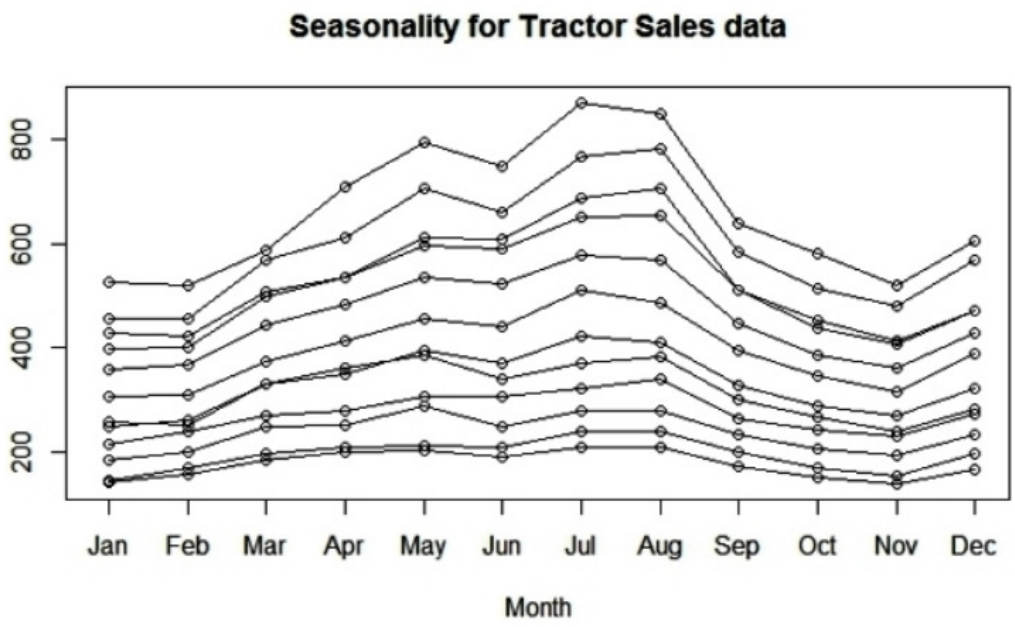
Blue Line: Moving Average; **Orange Line**: Actual Sales Figure



As evident from the plots, the 12 months moving average produced the most smooth curve.

Seasonality by Time Series Decomposition

The first thing to do is to see how number of tractors sold vary on a month on month basis. We will plot a stacked annual plot to observe seasonality in our data. As you can observe from the plot that there is a fairly consistent month on month variation with July and August as the peak months for tractor sales.



To decipher underlying patterns in tractor sales, you build a multiplicative time series decomposition model with the following equation

Y(t) = Trend(t) X **Seasonality (t)** X **Remainder (t)**

Instead of multiplicative model you could have chosen additive model as well. However, it would have made very little difference in terms of conclusion you will draw from this time series decomposition exercise. Additionally, you are also aware that plain vanilla decomposition models like these are rarely used for forecasting. Their primary purpose is to understand underlying patterns in temporal data to use in more sophisticated analysis like Holt-Winters seasonal method or ARIMA.

ARIMA Modeling

The process of making sugar cane juice has similarities with ARIMA modeling. It is prepared by crushing a long piece of sugar cane through the juicer with two large cylindrical rollers. However, it is difficult to extract all the juice from a tough sugar cane in one go hence the process is repeated multiple times.

Consider your time series data as a sugar cane and ARIMA models as sugar cane juicers. The idea with ARIMA models is that the final residual should look like white noise (juice-less-residual) otherwise, there is juice or information available in the data to extract.

ARIMA is a combination of three parts i.e. AR (AutoRegressive), I (Integrated), and MA (Moving Average). A convenient notation for ARIMA model is ARIMA(p,d,q). Here p,d, and q are the levels for each of the AR, I, and MA parts. Each of these three parts is an effort to make the final residuals display a white noise pattern (or no pattern at all). In each step of ARIMA modeling, time series data is passed through these three parts like a sugar cane through a sugar cane juicer to produce juice-less residual. The sequence of three passes for ARIMA analysis is as follows:

First Pass of ARIMA to Extract Information

Integrated (I) – subtract time series with its lagged series to extract trends from the data.

In this pass of ARIMA, we extract trend(s) from the original time series data. Differencing is one of the most commonly used mechanisms for extraction of trends. Here, the original series is subtracted from it's lagged series e.g. November's sales values are subtracted from October's values to produce trend-less residual series. The formulae for different orders of differencing are as follows:

No Differencing (d=0)	$Y_t' = Y_t$
1st Differencing (d=1)	$Y_t' = Y_t - Y_{t-1}$
2nd Differencing (d=2)	$Y_t' = Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2}) = Y_t - 2 \times Y_{t-1} + Y_{t}$

Second Pass of ARIMA to Extract Information

***AutoRegressive* (AR)** – extract the influence of the previous periods' values on the current period. After the time series data is made stationary through the *integrated* (I) pass, the AR part of the ARIMA gets activated. As the name auto-regression suggests, here we try to extract the influence of the values of previous periods on the current period e.g. the influence of the September and October's sales value on the November's sales. This is done through developing a regression model with the time-lagged period values as independent or predictor variables.

The general form of the equation for this regression model is shown below.

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

Third Pass of ARIMA to Extract Information

Moving Average* (MA)** – extract the influence of the previous period's error terms on the current period's error. Finally, the last component of ARIMA i.e. MA involves finding relationships between the previous periods' error terms on the current period's error term. ***Moving Average (MA) part of ARIMA is developed with the following simple multiple linear regression values with the lagged error values as independent or predictor variables.

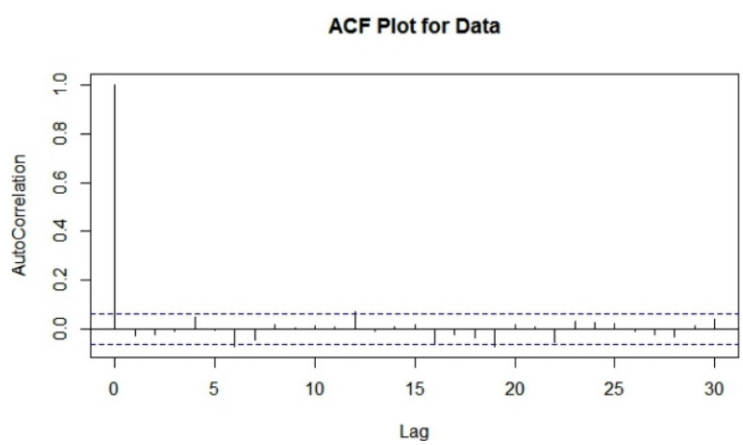
$$Y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

White Noise and ARIMA

White noise is a funny thing, if you look at it for long you will start seeing some false patterns. This is because the human brain is wired to find patterns, and at times confuses noises with signals. The biggest proof of this is how people lose money every day on the stock market.

This is precisely the reason why we need a mathematical or logical process to distinguish between a white noise and a signal (information). For example, consider the following simulated white noise:

A good way to distinguish between signal and noise is ACF (AutoCorrelation Function). This is developed by finding the correlation between a series of its lagged values. In the following ACF plot, you could see that for lag = 0 the ACF plot has the perfect correlation i.e. ρ=1. This makes sense because any data with itself will always have the perfect correlation. However as expected, our white noise doesn't have a significant correlation with its historic values (lag≥1). The dotted horizontal lines in the plot show the threshold for the insignificant region i.e. for a significant correlation the vertical bars should fall outside the horizontal dotted lines.



Conclusion

Time series modeling and ARIMA forecasting are scientific ways to predict the future. The ARIMA models and time series analysis can be used to forecast tractor sales. The best fit ARIMA model is selected based on Akaike Information Criterion (AIC) , and Bayesian Information Criterion (BIC) values. The idea is to choose a model with minimum AIC and BIC values.

Time series:	$\log_{10}$(Tractor Sales)	
Best fit Model:	**ARIMA(0,1,1)(0,1,1)[12]**	
	ma1	**sma1**
Coefficients:	-0.4047	-0.5529
s.e.	0.0885	0.0734
log likelihood=354.4		
AIC=-702.79	AICc=-702.6	BIC=-694.17