Forecasting Tractor Sales

Scenario:

PowerHorse, a tractor and farm equipment manufacturing company, was established a few years after World War II. 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 it's inventory and production cost down because of variability in sales and tractor demand. The management at PowerHorse is under enormous pressure from the shareholders and board to reduce the production cost.

Objective of the project:

The objective of this project is to forecast the tractor sales in the next 36 months using Time Series Analysis- ARIMA method to bring effectiveness in production planning to maintain healthy business margins and effective inventory management.

Approach/Activities:

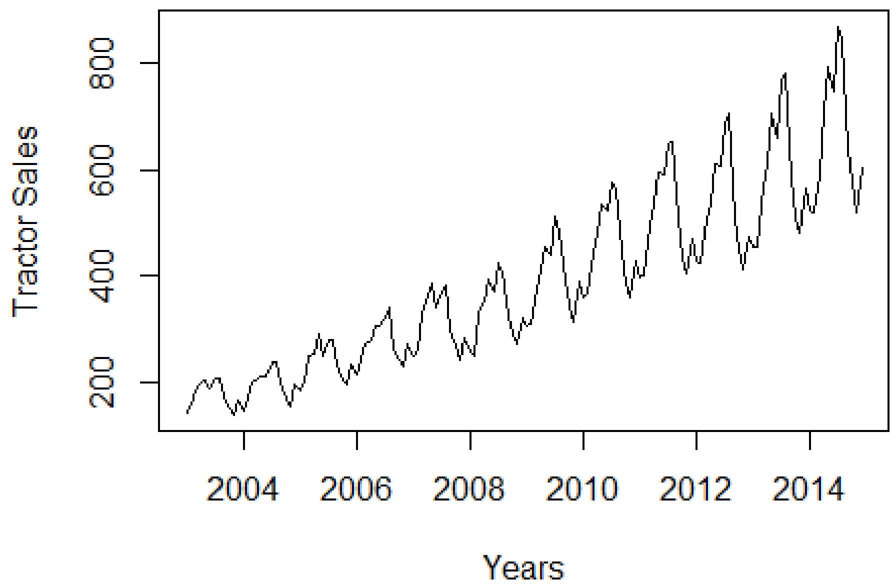
Time series decomposition of data into 4 components: Trend, Seasonality, Cycle and Irregular remainder, ARIMA modelling including estimating p,q and d levels and plotting ACF and PACF plots

Data Set Information:

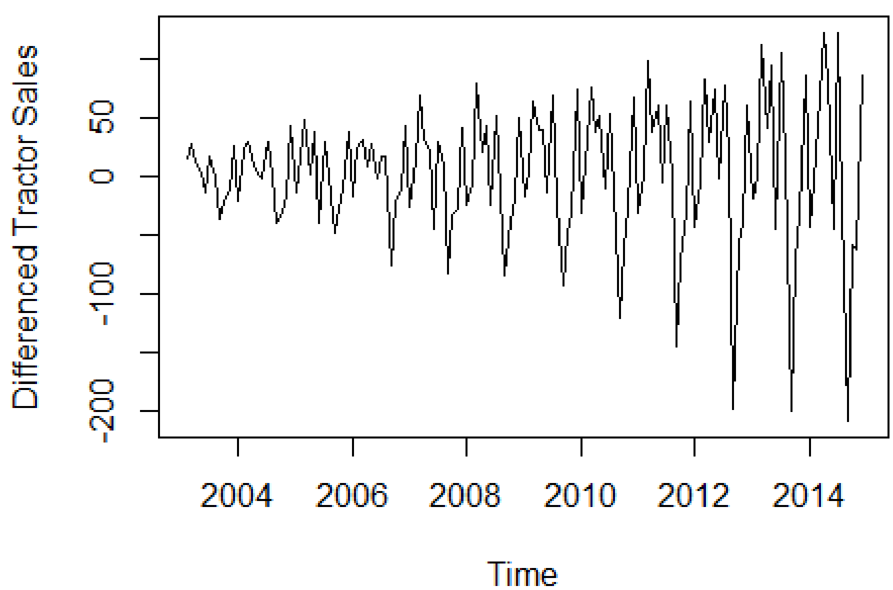
The MIS team of PowerHorse shared the month on month (MoM) sales figures (number of tractors sold). The dataset consists of 144 observations having the total month wise sales data of Tractors for a period of past 12 years

Step 1: Read and Plot tractor sales data as time series

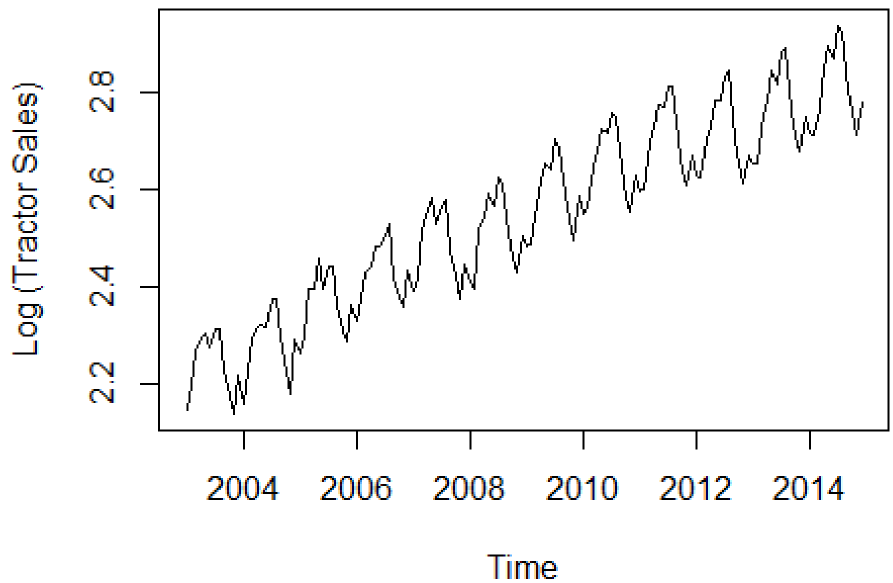
```r
data = read.csv("TractorSale.csv")
#data is converted into time series.
data = ts(data[,2],start = c(2003,1),frequency = 12)
plot(data, xlab='Years', ylab = 'Tractor Sales')
```



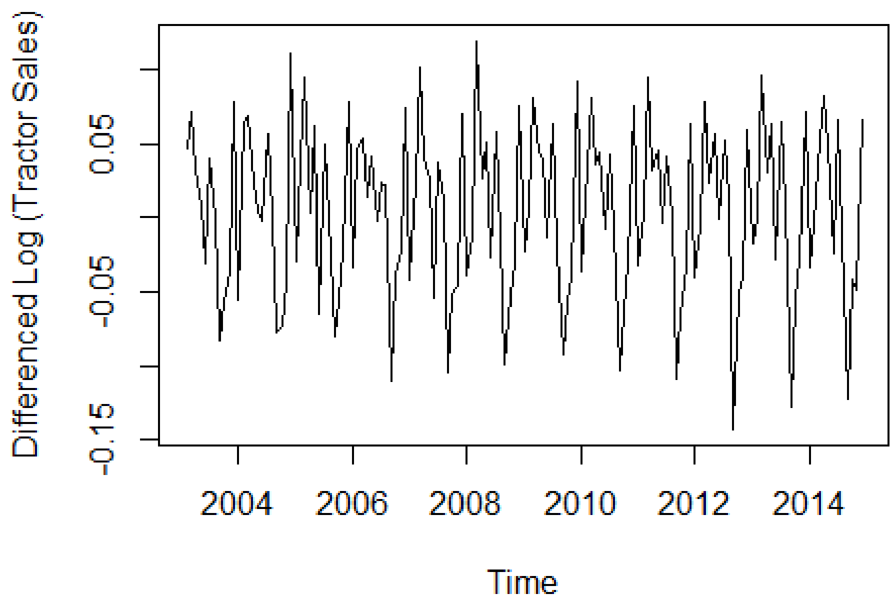

```r
#Step 2:Difference data to make data stationary on mean (remove trend)
plot(diff(data),ylab='Differenced Tractor Sales')
```



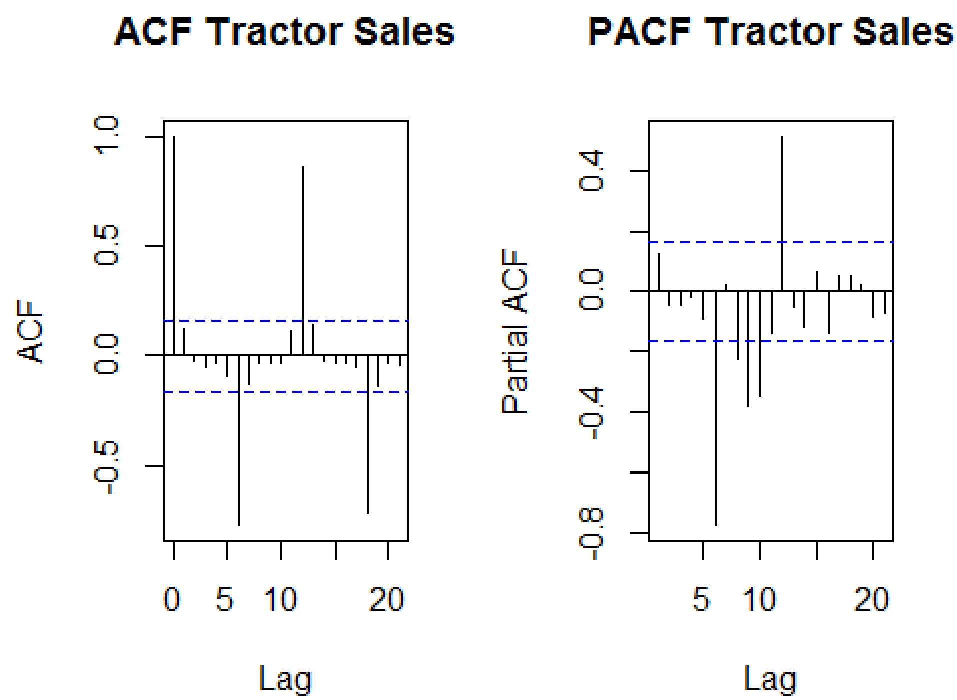
```r
#Step 3:Log transform data to make data stationary on variance
plot(log10(data),ylab='Log (Tractor Sales)')
```




```r
#Step 4:Difference log transform data to make data stationary on both mean and variance

plot(diff(log10(data)),ylab='Differenced Log (Tractor Sales)')
```




```r
#Step5:Plot ACF and PACF to identify potential AR and MA model

par(mfrow = c(1,2))
acf(ts(diff(log10(data))),main='ACF Tractor Sales')
pacf(ts(diff(log10(data))),main='PACF Tractor Sales')
```

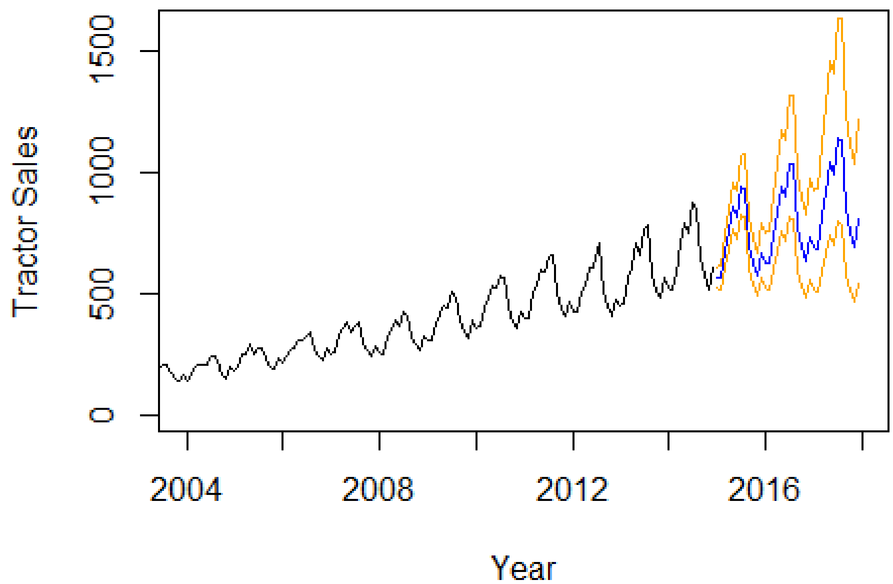



```r
#Step 6: Identification of best fit ARIMA model
library("forecast")

## Warning: package 'forecast' was built under R version 3.5.1

ARIMAfit = auto.arima(log10(data), approximation=FALSE,trace=FALSE)
summary(ARIMAfit)

## Series: log10(data)
## ARIMA(0,1,1)(0,1,1)[12]
##
## Coefficients:
##           ma1     sma1
##       -0.4047  -0.5529
## s.e.   0.0885   0.0734
##
## sigma^2 estimated as 0.0002571:  log likelihood=354.4
## AIC=-702.79   AICc=-702.6   BIC=-694.17
##
## Training set error measures:
##                        ME       RMSE        MAE         MPE      MAPE
## Training set 0.0002410698 0.01517695 0.01135312 0.008335713 0.4462212
##                   MASE       ACF1
## Training set 0.2158968 0.01062604

#Step 6: Forecast sales using the best fit ARIMA model
```

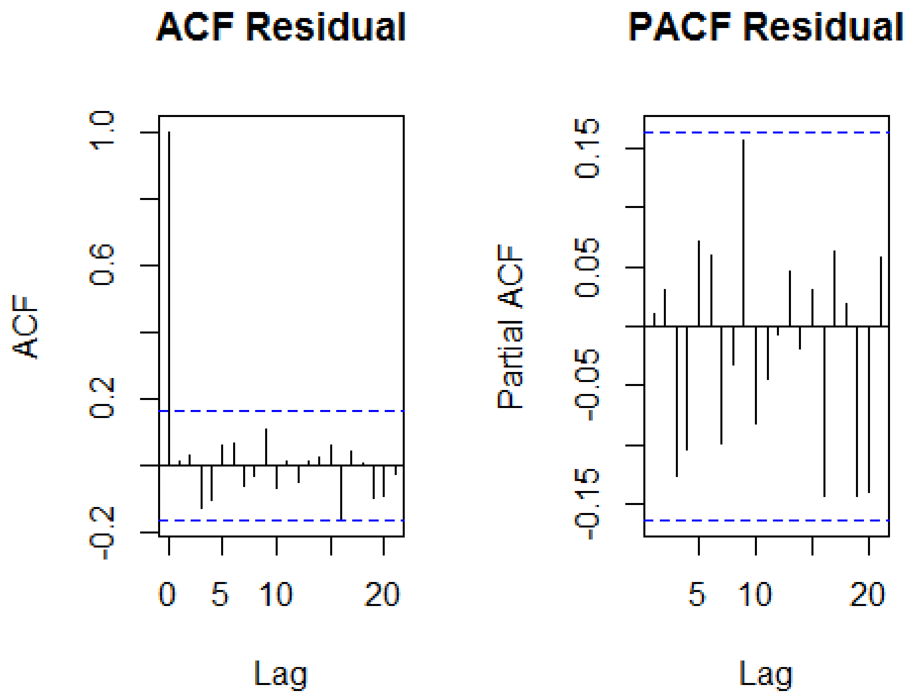
```r
par(mfrow = c(1,1))
pred = predict(ARIMAfit, n.ahead = 36)
pred
```

```
## $pred
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 2015 2.754168 2.753182 2.826608 2.880192 2.932447 2.912372 2.972538
## 2016 2.796051 2.795065 2.868491 2.922075 2.974330 2.954255 3.014421
## 2017 2.837934 2.836948 2.910374 2.963958 3.016213 2.996138 3.056304
##           Aug      Sep      Oct      Nov      Dec
## 2015 2.970585 2.847264 2.797259 2.757395 2.825125
## 2016 3.012468 2.889147 2.839142 2.799278 2.867008
## 2017 3.054351 2.931030 2.881025 2.841161 2.908891
##
## $se
##             Jan        Feb        Mar        Apr        May        Jun
## 2015 0.01603508 0.01866159 0.02096153 0.02303295 0.02493287 0.02669792
## 2016 0.03923008 0.04159145 0.04382576 0.04595157 0.04798329 0.04993241
## 2017 0.06386474 0.06637555 0.06879478 0.07113179 0.07339441 0.07558934
##             Jul        Aug        Sep        Oct        Nov        Dec
## 2015 0.02835330 0.02991723 0.03140337 0.03282229 0.03418236 0.03549035
## 2016 0.05180825 0.05361850 0.05536960 0.05706700 0.05871534 0.06031866
## 2017 0.07772231 0.07979828 0.08182160 0.08379608 0.08572510 0.08761165
```

```r
plot(data,type='l',xlim=c(2004,2018),ylim=c(1,1600),xlab = 'Year',ylab =
'Tractor Sales')
lines(10^(pred$pred),col='blue')
lines(10^(pred$pred+2*pred$se),col='orange')
lines(10^(pred$pred-2*pred$se),col='orange')
```




```
#Step 7: Plot ACF and PACF for residuals of ARIMA model to ensure no more
information is left for extraction
par(mfrow=c(1,2))
acf(ts(ARIMAfit$residuals),main='ACF Residual')
pacf(ts(ARIMAfit$residuals),main='PACF Residual')
```


ACF Residual
ACF
1.0
0.6
0.2
-0.2
0
5
10
20
Lag
PACF Residual
Partial ACF
0.15
0.05
-0.05
-0.15
5
10
20
Lag