BT1101: Case Study Assignment 2

Leong Maegan

Ow Wen Hui Shirley

1) Improving production using regression analysis
Read the case information on Page 298 of Evans. Use techniques of regression analysis to evaluate the data in the respective worksheets and reach useful conclusions. Summarize your work in a formal report with all appropriate results and analyses.

After analyzing the case information on Page 298 of Evans, we have picked out three main parts to answer this question, in terms of what Elizabeth Burkes would like to know about the company PLE:

1. The company instituted an initiative in August 2011 to work with suppliers to reduce defects, to more closely coordinate deliveries, and to improve materials quality through reengineering supplier production policies. Elizabeth noted that the program appeared to reverse an increasing trend in defects; she would like to predict what might have happened had the supplier initiative not been implemented and how the number of defects might further be reduced in the near future. (Defects After Delivery Sheet)

2. There is a high rate of turnover in its field service staff, and Elizabeth wants to identify the characteristics of individuals that lead to greater retention by studying the effect that years of education, college grade point average and age when hired have on retention. (Employee Retention Sheet)

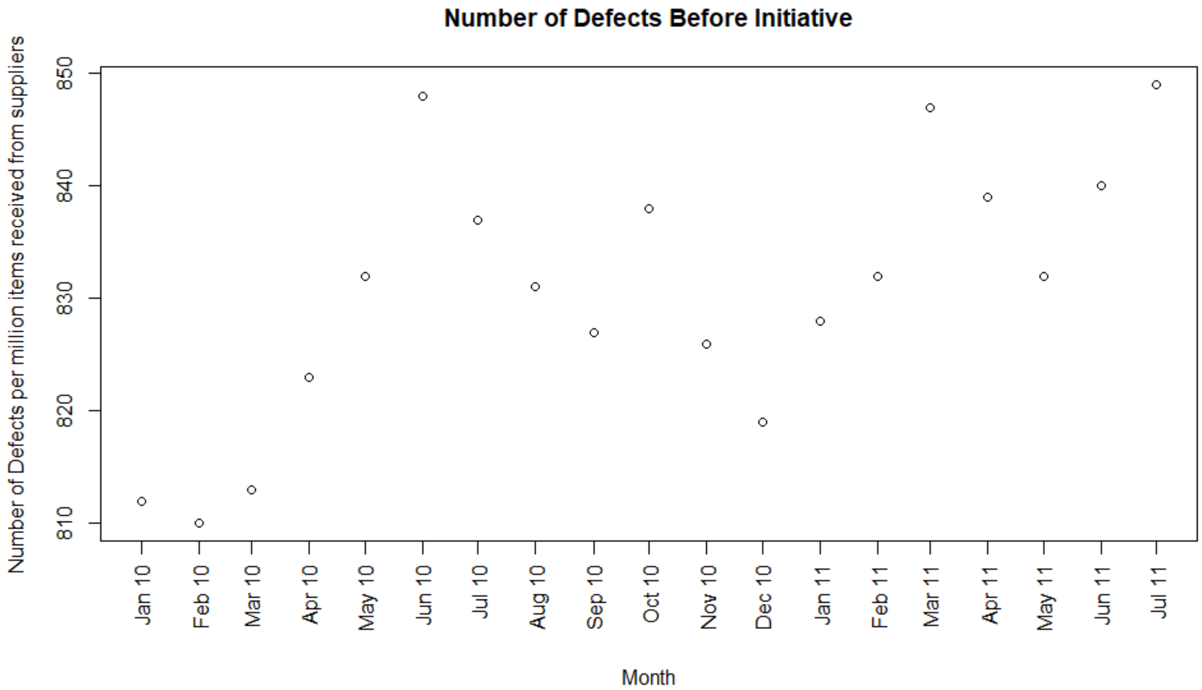
3. Elizabeth would like to understand the rate of learning of the firm in using new production technology. Time required to produce the mowers/tractors using the newer means of technology gradually decreases as more units are made. Rate of improvement declines until the production time levels off. (Engines Sheet)

We will tackle each of the above parts one by one.

Firstly, Elizabeth would like to predict what might have happened had the supplier initiative not been implemented.

Since we would like to examine what might have happened had the initiative not been implemented, the data set is separated into two smaller subsets – one when the supplier initiative has not been implemented (data from Jan 2010 – July 2011) and the other when the supplier initiative has been implemented (data from Aug 2011 – Dec 2014).

Using the first subset of the data when the supplier initiative has not yet been implemented, a scatterplot is plotted to further examine the data.



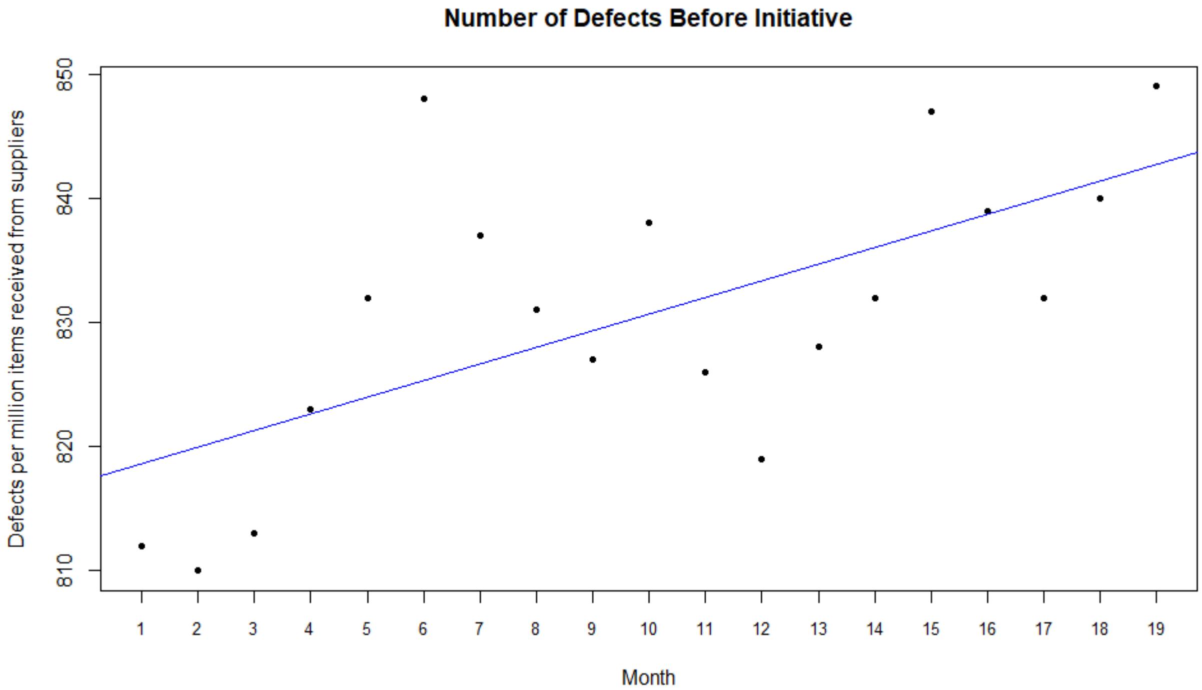
Graph 1: The scatterplot above is plotted with the Number of Defects per million items received from suppliers against the Month the data was collected in, from January 2010 to July 2011.

From Graph 1, we can examine how the number of defects vary with time before the initiative was implemented. The scatterplot generally shows an upward trend for the Number of Defects per million items received from suppliers as time passes.

To predict what might have happened had the initiative not been implemented, we can adopt a simple linear regression model to aid us with the prediction, with the single independent variable being the number of Months from December 2009 the data was collected in, with January 2010 being Month 1, February 2010 being Month 2 etc. till Month 19 in July 2011.

To start off, we will attempt to fit different trendlines to the scatterplot and determine which trendline function will better fit the data.

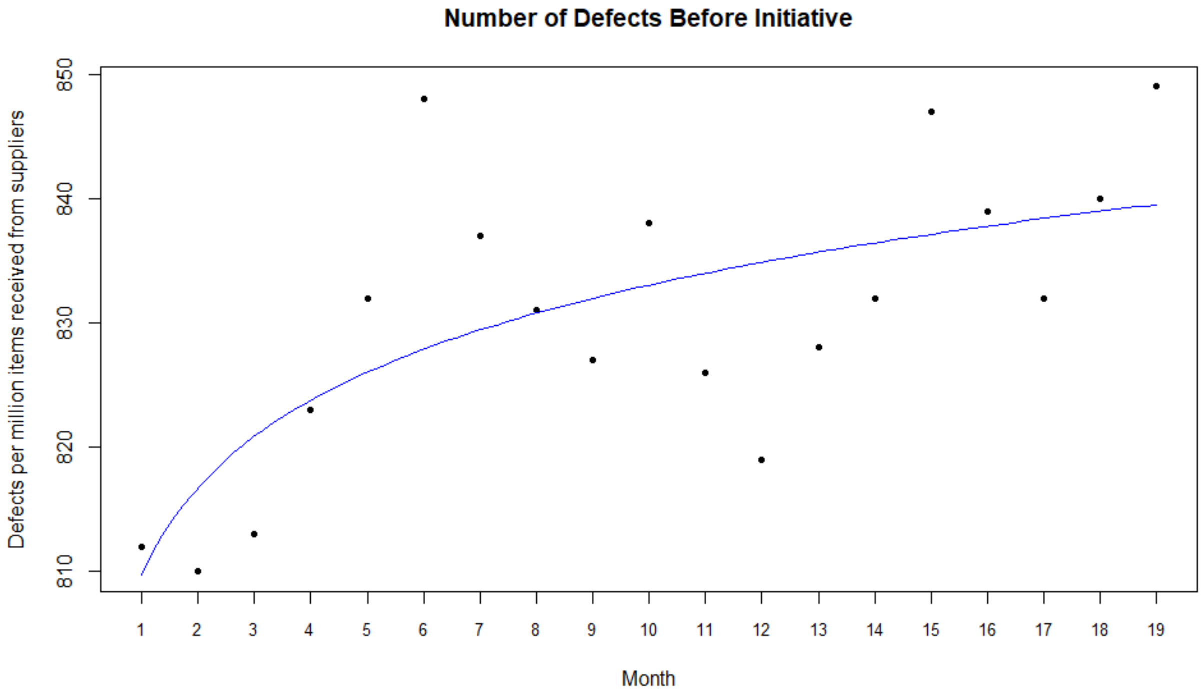
1. Linear



Graph 2: Linear trendline modelled onto the scatterplot.

The multiple R-squared value of this linear model is 0.409 (and the adjusted R-squared value for this linear model is 0.374). The R-squared value gives the proportion of variation in the dependent variable that is explained by the independent variable of the regression model. In this case, the multiple R-squared value is low as it shows that the model only explains 40.9% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean.

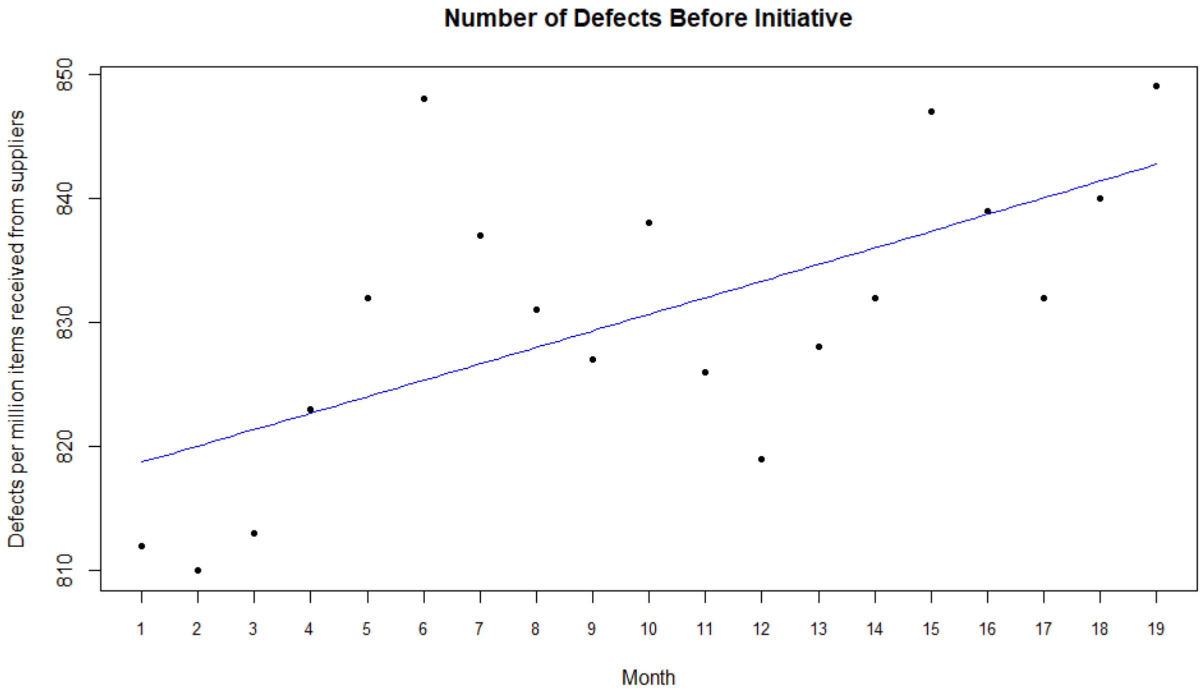
2. Logarithmic



Graph 3: Logarithmic trendline modelled onto the scatterplot.

The multiple R-squared value for the logarithmic model is 0.482. The multiple R-squared value is low as it shows that the model only explains 48.2% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean.

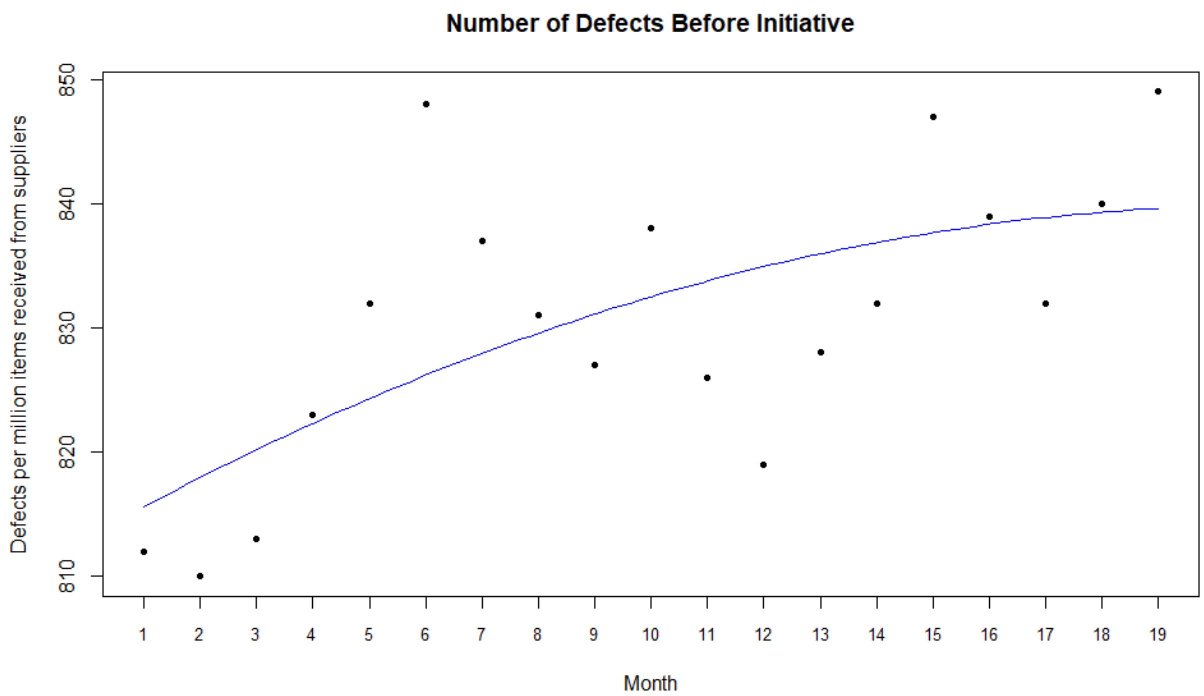
3. Exponential



Graph 4: Exponential trendline modelled onto the scatterplot.

The multiple R-squared value for the logarithmic model is 0.408. The multiple R-squared value is low as it shows that the model only explains 40.8% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean.

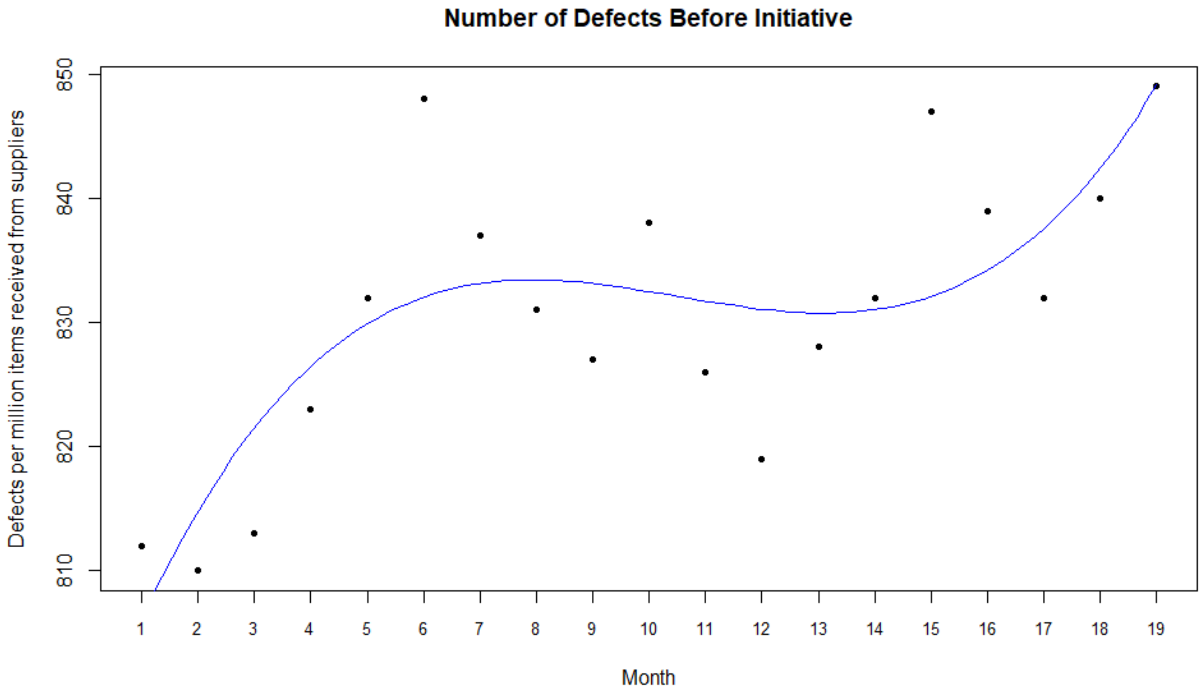
4. Polynomial



Graph 5: Polynomial trendline modelled onto the scatterplot.

The multiple R-squared value for the logarithmic model is 0.429. The multiple R-squared value is low as it shows that the model only explains 42.9% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean.

5. Cubic Polynomial

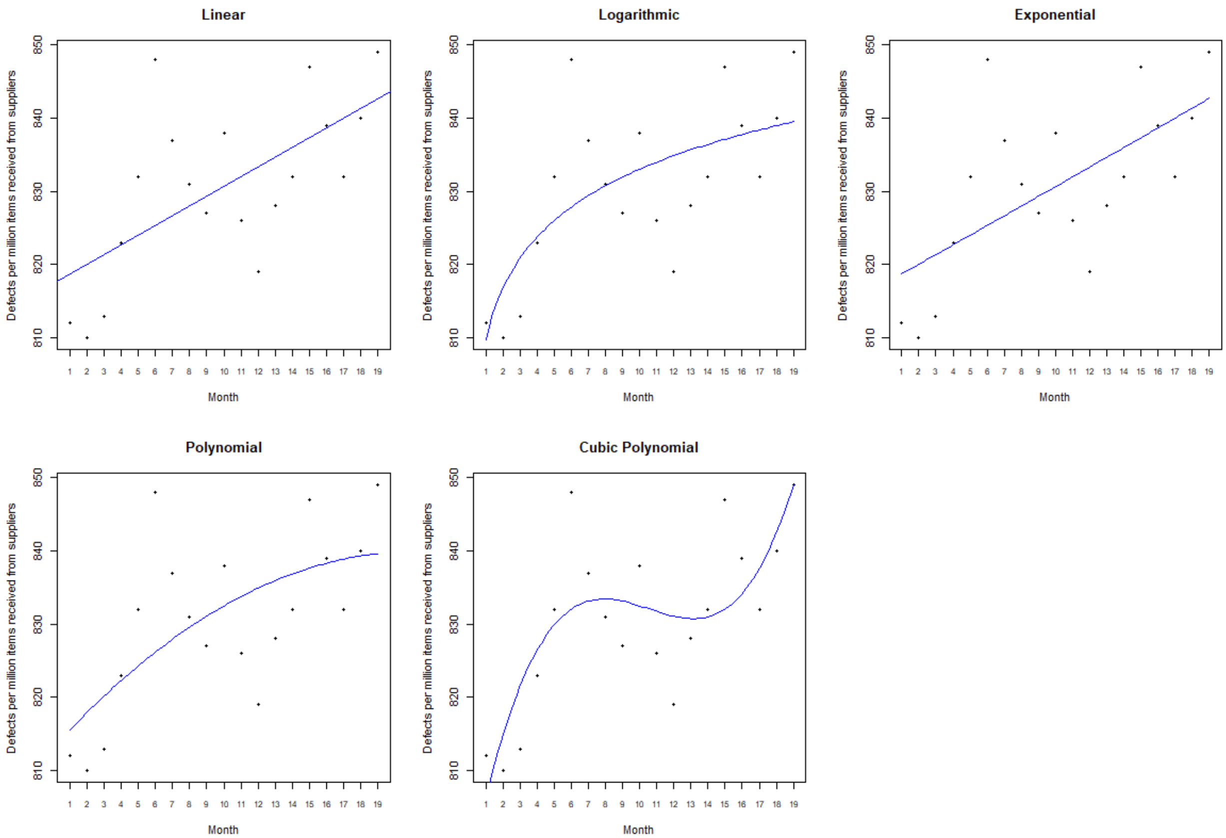


Graph 6: Cubic Polynomial trendline modelled onto the scatterplot.

The multiple R-squared value for the logarithmic model is 0.614. The multiple R-squared value is moderately high as it shows that the model explains 61.4% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean, which is more than 50% of the variability of the data for the dependent variable accounted for.

Model	**R-squared value (3 sig. fig.)**
Linear	0.409
Logarithmic	0.482
Exponential	0.408
Polynomial	0.429
Cubic Polynomial	0.614

Table 1: R-squared values tabulated for the respective models.

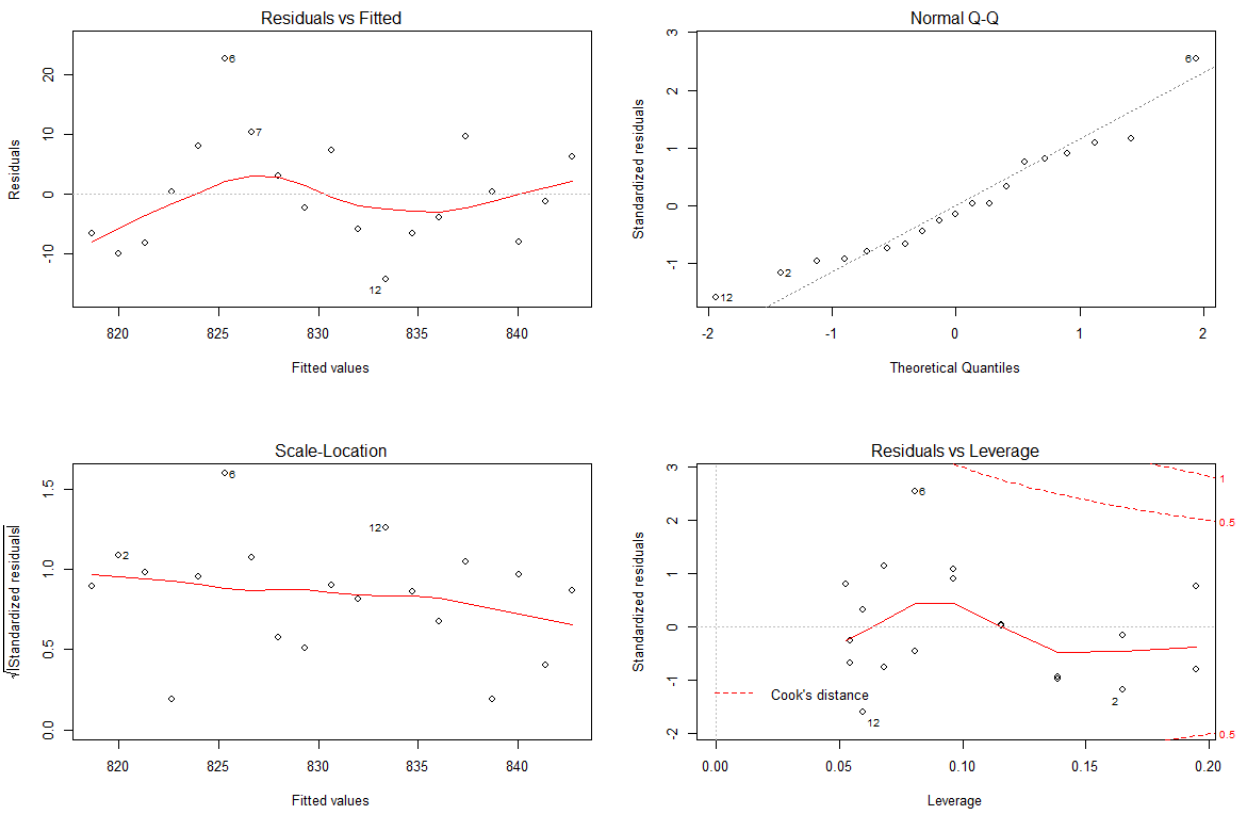


Graph 7: A collation of all the previous graphs in the same window for ease of comparison.

As seen from Table 1 and Graph 7, the Cubic Polynomial model has the highest R-squared value and the best-fitted trendline function as compared to the other models, and hence the cubic polynomial trendline is best-fitted for the data.

Next, we can continue to conduct simple linear regression. We will employ the third-order polynomial trendline to fit our data here.

Before conducting the regression analysis, we must first check for the assumptions associated with regression analysis.



Graph 7: A collation of the Residuals vs Fitted plot, normal Q-Q plot, Scale-Location plot and Residuals vs Leverage plot of the linear model fitted with the cubic polynomial function using our data.

Linearity: The residuals do not exhibit well-defined, non-linear patterns. Points around the line y = 0 seem randomly scattered in the residual plot. Linearity assumption is fulfilled.

Normality of data: Based on the QQ plot, the residuals seem normally distributed, as most of the residuals points on the QQ-plot seem to fall close to a straight line. However, this can be further verified with other tests of normality, such as by conducting the Shapiro-Wilk normality test on the residuals of the model.

```
        Shapiro-Wilk normality test

data:  residuals(defects.old.rate.mod)
W = 0.95262, p-value = 0.4373
```

Based on the p-value from the Shapiro-Wilk normality test, since p-value = 0.4373 > 0.05, we cannot reject the null hypothesis that the residuals in the model are normally distributed. Hence, this model has fulfilled the assumption of normality.

Homoscedasticity: The variation about the regression line appear relatively constant for all values of the independent variable, as the residuals are spread equally along the range of predictors (Scale-Location plot relatively close to a horizontal line with equally, randomly spread points), and hence the model fulfills the assumption of homoscedasticity. We can validate this statement by running a Score Test for Non-Constant Error Variance (ncvTest) using the model.

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.8945435, Df = 1, p = 0.34425
```

In the ncvTest, the null hypothesis is that there is constant error variance and the alternative hypothesis is that the error variance changes with the level of fitted values of the response variable or linear combination of predictors. As p-value is 0.34425 >0.05, we cannot reject the null hypothesis that there is constant error variance. This model thus fulfils the assumption of homoscedasticity.

Independence of errors: Residual plot does not seem to have clusters of residuals with the same sign, and hence autocorrelation does not seem to exist in this case, and errors can be considered independent.

Conclusion: Since the assumptions required for this regression analysis are generally fulfilled, we can proceed on with the construction of our simple linear regression model.

The following shows the summary of the linear model fitted with the cubic polynomial function that we have obtained previously as best-fitted for the data.

```
Call:
lm(formula = defects1$`Defects per million items received from suppliers` ~
    defects1$Month + defects1$Month^2 + defects1$Month^3)

Residuals:
    Min      1Q  Median      3Q     Max
-14.354  -6.679  -1.365   6.808  22.656

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    817.3333     4.4380  184.17  < 2e-16 ***
defects1$Month   1.3351     0.3892    3.43  0.00319 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.293 on 17 degrees of freedom
Multiple R-squared:  0.409,   Adjusted R-squared:  0.3742
F-statistic: 11.76 on 1 and 17 DF,  p-value: 0.003194
```

The multiple R-squared value is still moderately low, at 0.409. However, the predictors are all statistically significant (p-values < 0.05) and the model is on overall significant as well (with p-value = 0.003194 < 0.05). Regardless of the R-squared, the significant coefficients still represent the mean change in the dependent variable for one unit of change in the predictor while holding other predictors in the model constant. Hence, this implies that we can still use the model to predict what might have happened had the supplier initiative not been implemented to a moderate extent of accuracy.

With the coefficients of the cubic polynomial trendline, the following equation is formed:

$$y = 0.0387x^3 - 1.221\,x^2 + 12.067\,x + 795.235$$

where y is the Number of Defects per million items received from suppliers and x is the number of months from December 2009 the data was collected in.

From our trend line equation, we can predict our y value in the near future (where $x \geq 20$). For example, we can use substitute $x = 20$ to find our predicted y-value ($\hat{y}$) in August 2011, had the initiative not been implemented.

$$\hat{y} = 0.0387(20^3) - 1.221(20^2) + 12.067(20) + 795.235 = 857.775$$

Had the initiative not been implemented, the predicted y-value in August 2011 is 857.775, and when rounded down to an integer, $\hat{y} = 857$. This is 8 more than the number of defects per million items received from suppliers in July 2011. This shows that without the initiative implemented, there will be an increase in the number of defects per million items received from suppliers.

The actual number of defects per million items received from suppliers for August 2011 is 857. There is no difference in the predicted y-value and the actual y-value in this case, because the initiative has only just been implemented in August 2011 and the initiative may require some time to take flight. If we were to analyse the difference in the predicted y-value and the actual y-value over a longer period of time, we will be able to see more significant differences.

To see the effects of the implementation of the initiative, we can look at the prediction of $\hat{y}$ in September 2011 if the initiative has not been implemented in August 2011.

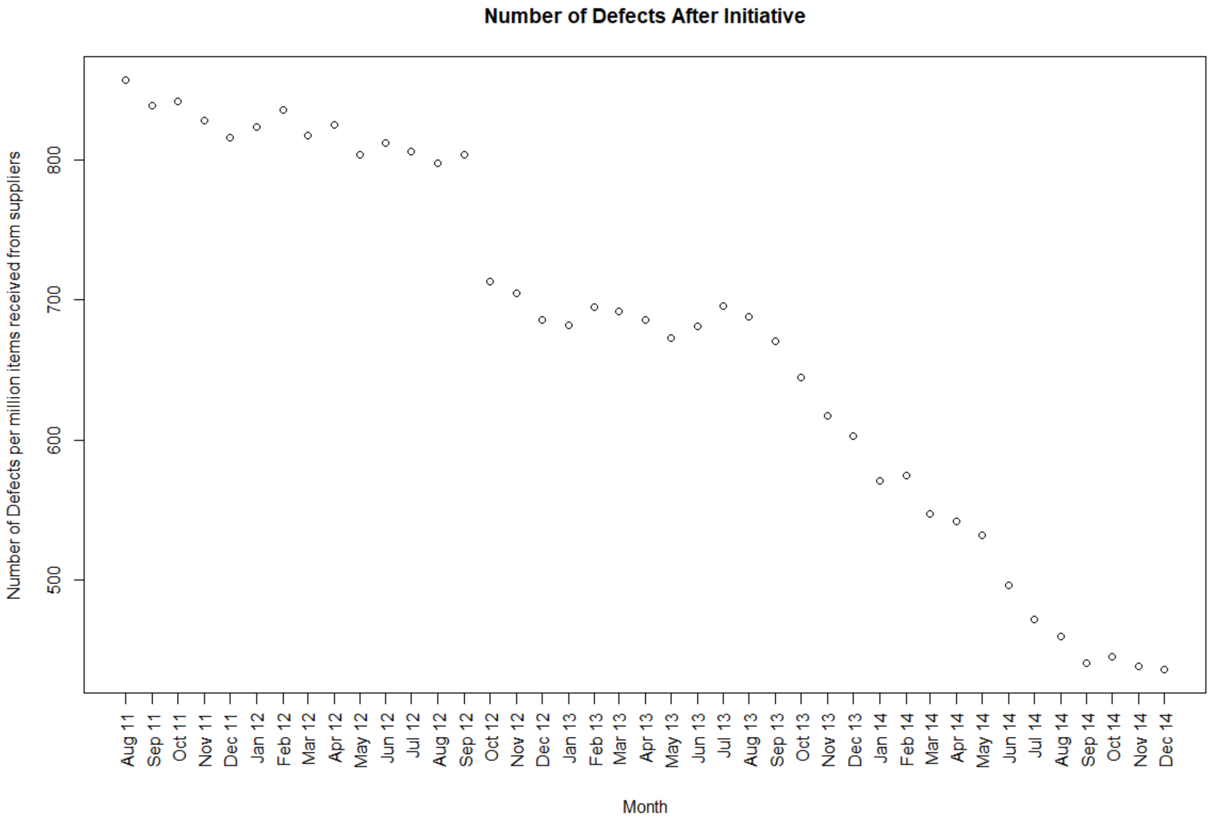
$$\hat{y} = 0.0387(21^3) - 1.221(21^2) + 12.067(21) + 795.235 = 868.582$$

Had the initiative not been implemented, the predicted y-value in September 2011 is 868.582. Meanwhile, the actual y-value (after the initiative has been implemented) was 839. The actual y-value fell from August 2011 to September 2011, while it increased in our prediction model (given no initiative implemented).

Therefore, what might have happened had the supplier initiative not been implemented is the number of defects per million items received from the suppliers will likely to keep increasing as time goes by.

Secondly, Elizabeth would like to know how the number of defects might further be reduced in the near future.

To understand how the number of defects might be further reduced in the near future, we will have to examine the data by plotting a scatterplot of the second subset of the data (Aug 2011 – Dec 2014) after the initiative has been implemented.



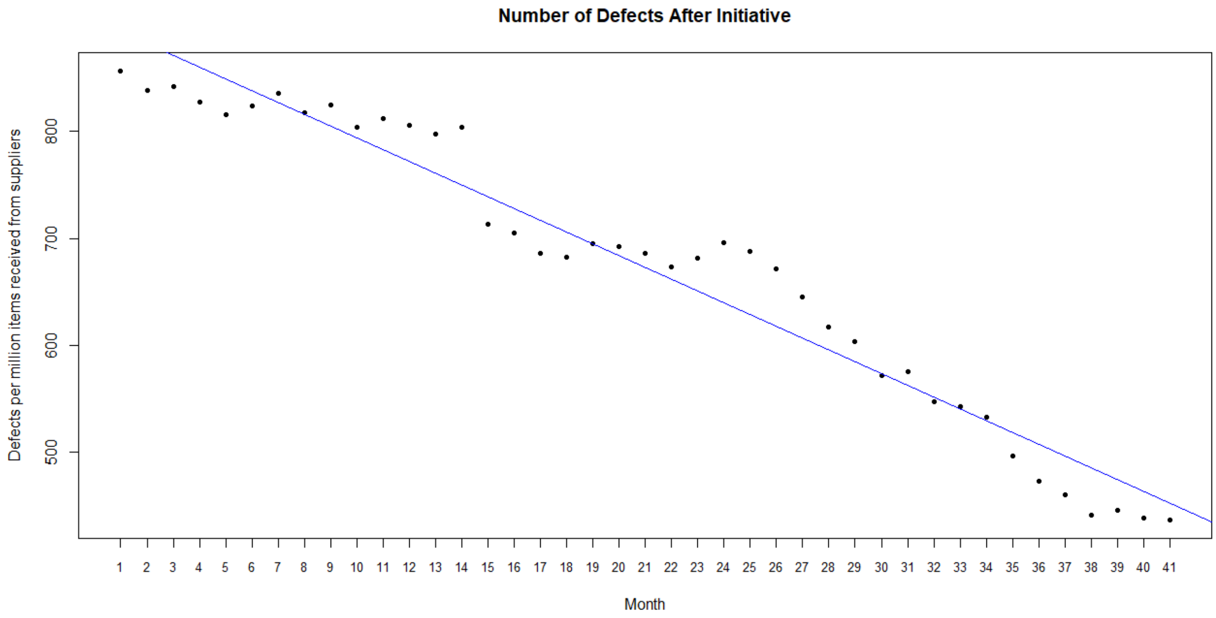
Graph 8: The scatterplot above is plotted with the Number of Defects per million items received from suppliers against the Month the data was collected in, from August 2011 to December 2014.

From Graph 8, we can examine how the number of defects vary with time after the initiative was implemented. The scatterplot generally shows a downward trend for the Number of Defects per million items received from suppliers as time passes.

To predict how the number of defects might be further reduced after the implementation of the initiative, we can adopt a simple linear regression model to aid us with the prediction, with the single independent variable being the number of Months from July 2011 the data was collected in, with August 2011 being Month 1, September 2011 being Month 2 etc. till Month 41 in December 2014.

To start off, we will attempt to fit different trendlines to the scatterplot and determine which trendline function will better fit the data.

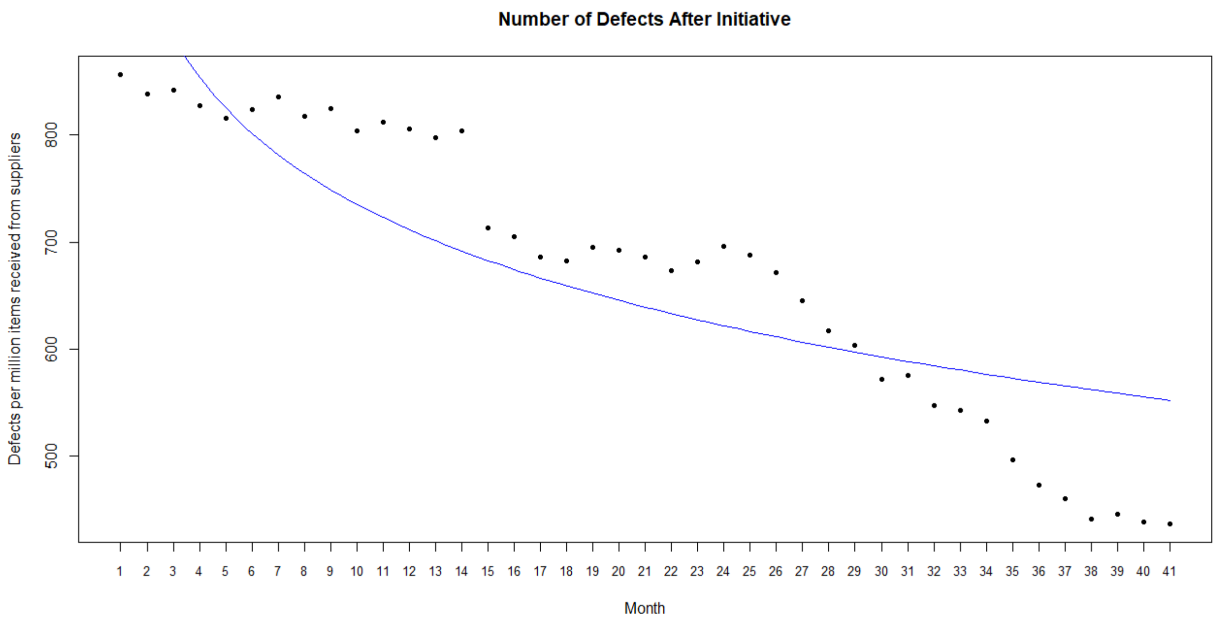
1. Linear



Graph 9: Linear trendline modelled onto the scatterplot.

The multiple R-squared value of this linear model is 0.951 (and the adjusted R-squared value for this linear model is 0.950). The R-squared value gives the proportion of variation in the dependent variable that is explained by the independent variable of the regression model. In this case, the multiple R-squared value is high as it shows that the model explains 95.1% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean.

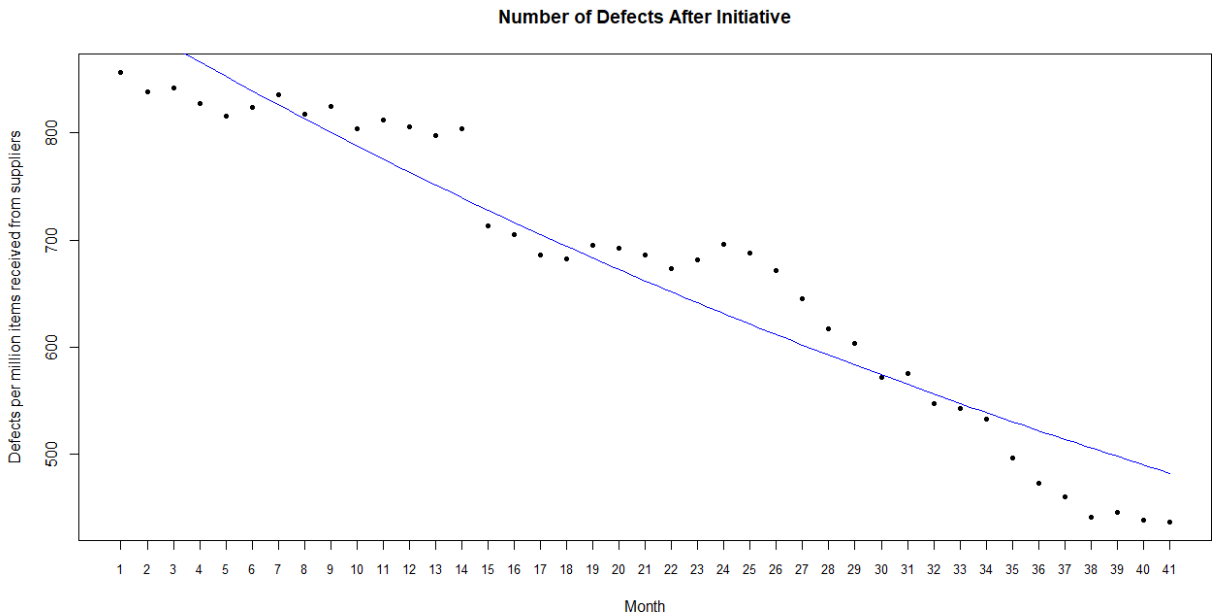
2. Logarithmic



Graph 10: Logarithmic trendline modelled onto scatterplot.

The multiple R-squared value for the logarithmic model is 0.704. The multiple R-squared value is moderately high as it shows that the model explains 70.4% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean.

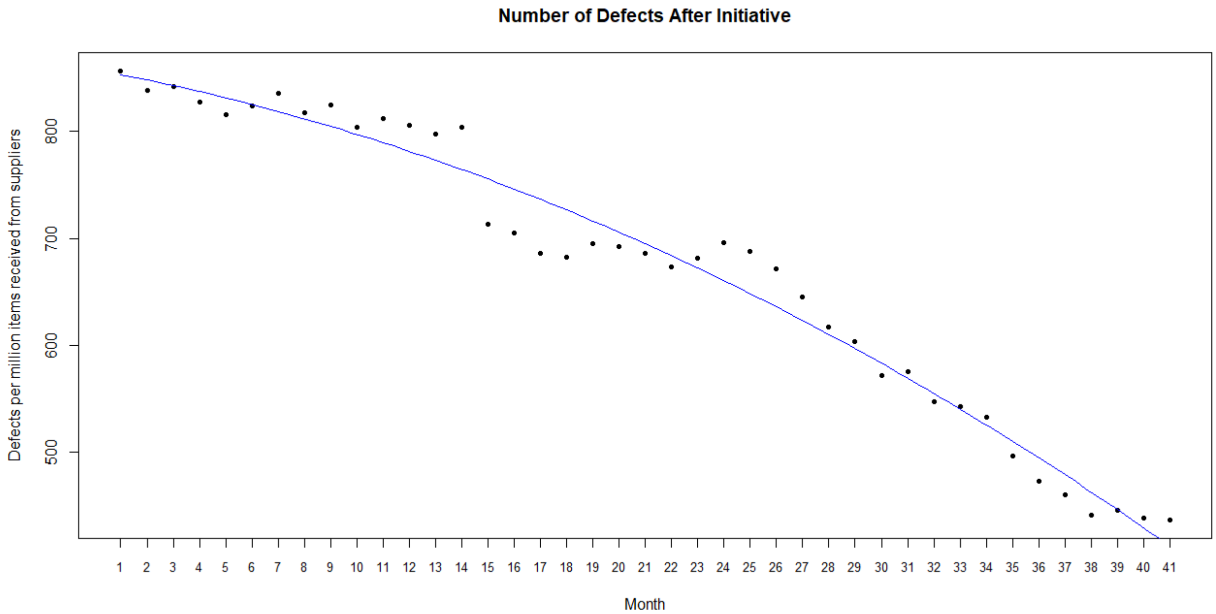
3. Exponential



Graph 11: Exponential trendline modelled onto scatterplot.

The multiple R-squared value for the exponential model is 0.921. The multiple R-squared value is high as it shows that the model explains 92.1% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean.

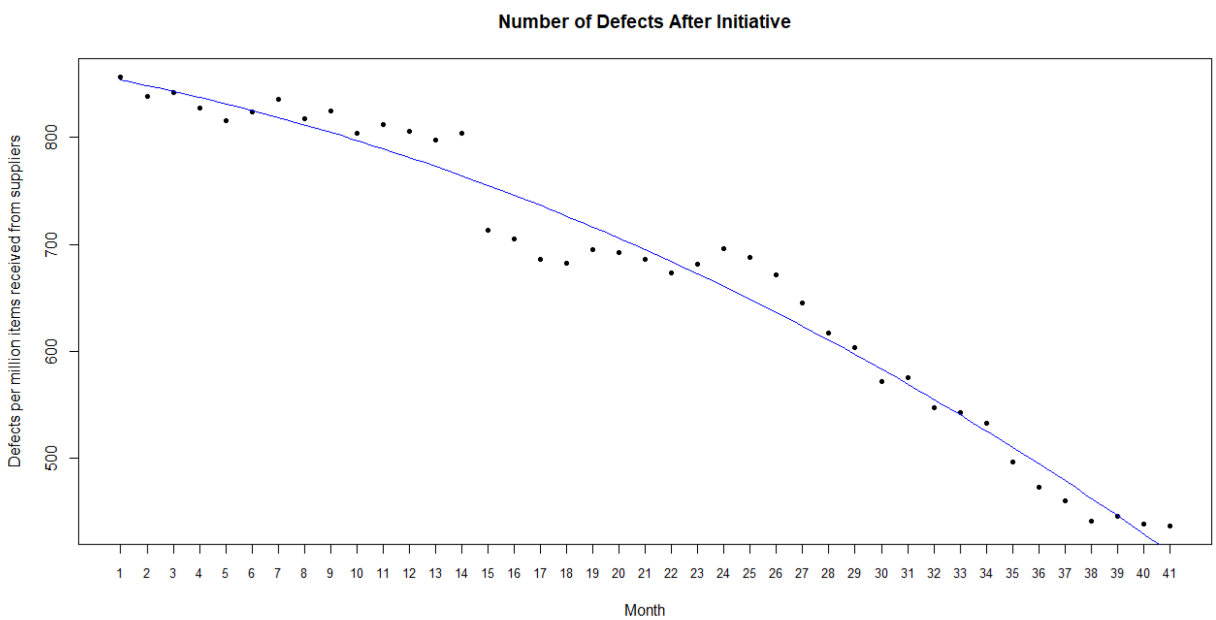
4. Polynomial



Graph 12: Polynomial trendline modelled onto scatterplot.

The multiple R-squared value for the polynomial model is 0.972. The multiple R-squared value is very high as it shows that the model explains 97.2% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean.

5. Cubic Polynomial

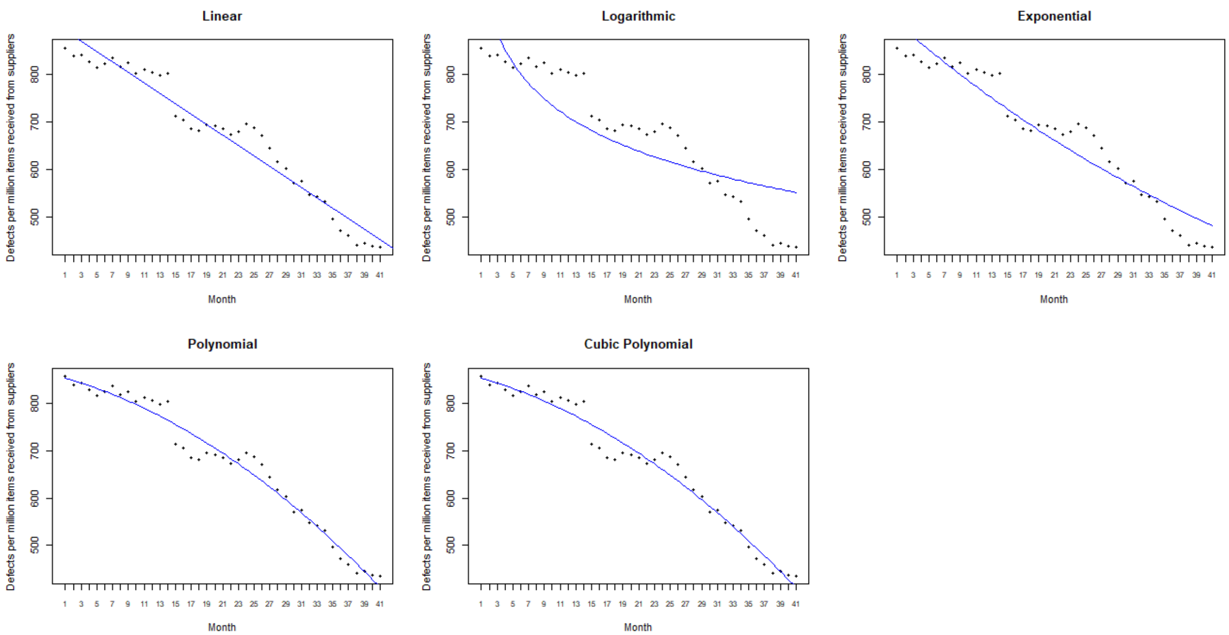


Graph 13: Cubic Polynomial trendline modelled onto scatterplot.

The multiple R-squared value for the polynomial model is 0.972 (this R-squared value has been rounded to 3 significant figures; the actual R-squared value is slightly above the R-squared value of the Polynomial trendline). The multiple R-squared value is very high as it shows that the model explains 97.2% of the variability of the dependent variable (Number of defects per million items received from suppliers) around its mean.

Model	**R-squared value (3 sig. fig.)**
Linear	0.951
Logarithmic	0.704
Exponential	0.921
Polynomial	0.972
Cubic Polynomial	0.972

Table 2: R-squared values tabulated for the respective models.

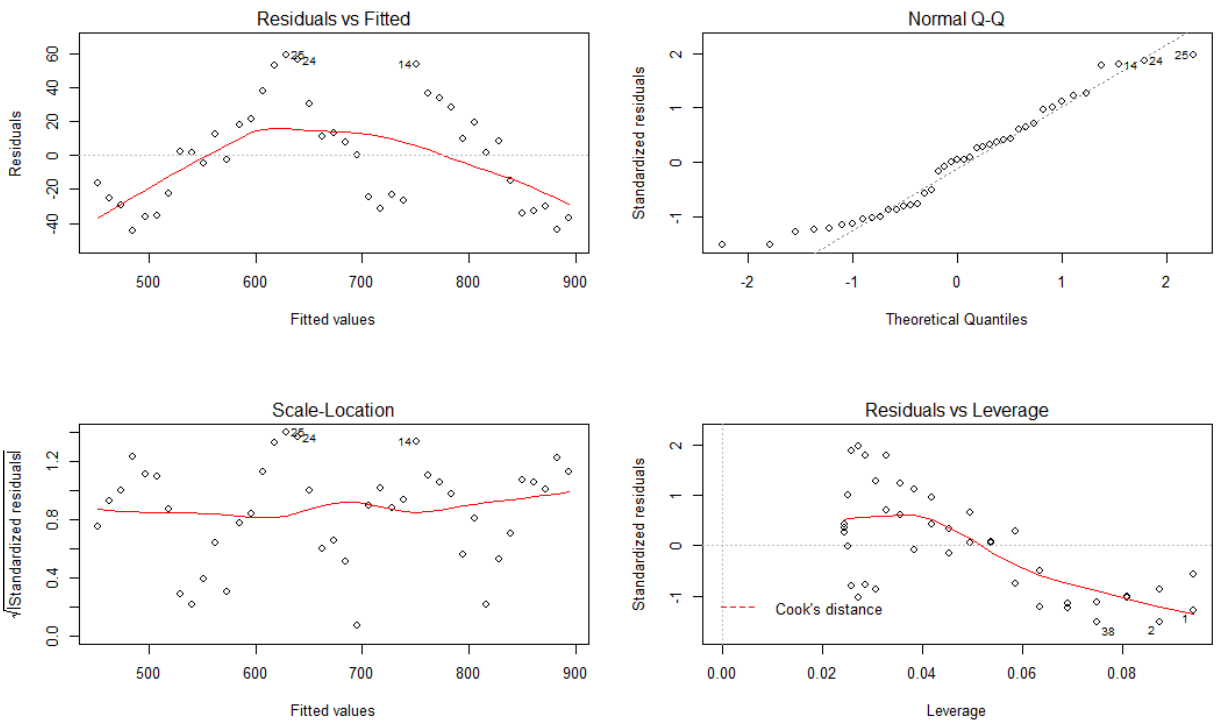


Graph 14: A collation of all the previous graphs in the same window for ease of comparison.

As seen from Table 2 and Graph 15, the Cubic Polynomial model has the highest R-squared value and the best-fitted trendline function as compared to the other models, and hence the cubic polynomial trendline is best-fitted for the data.

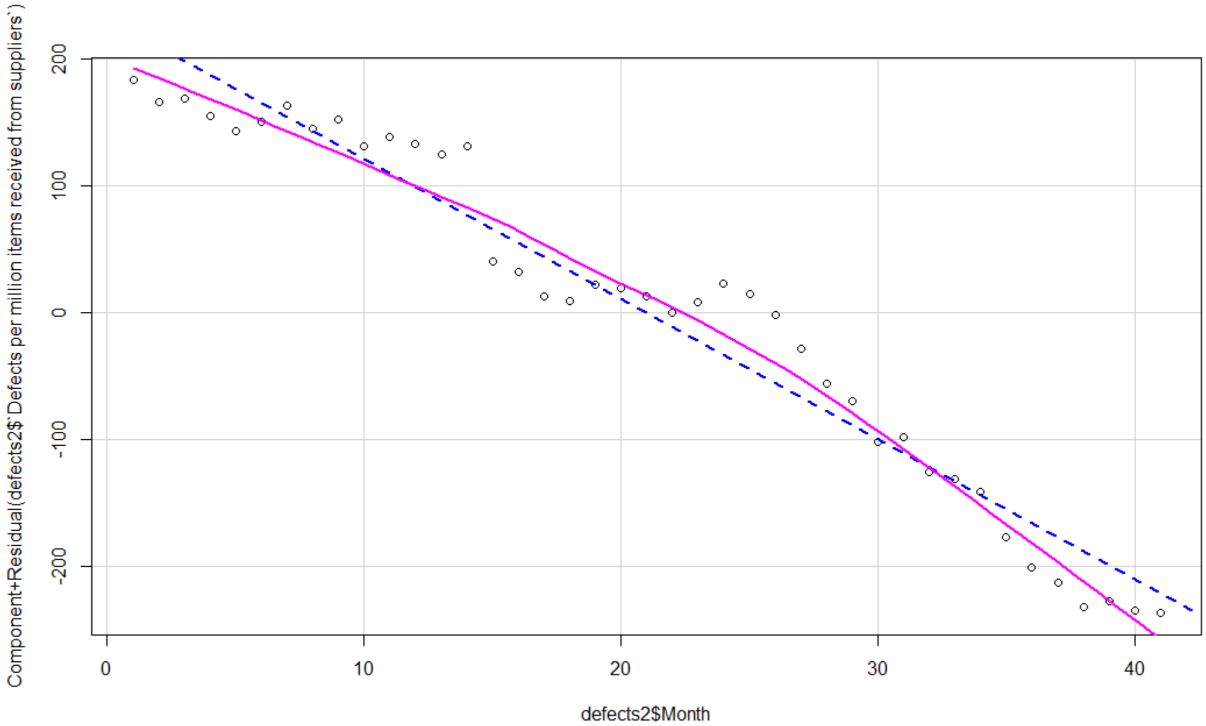
Next, we can continue to conduct simple linear regression. We will employ the third-order polynomial trendline to fit our data here.

Before conducting the regression analysis, we must first check for the assumptions associated with regression analysis.



Graph 15: A collation of the Residuals vs Fitted plot, normal Q-Q plot, Scale-Location plot and Residuals vs Leverage plot of the linear model fitted with the cubic polynomial function using our data.

Linearity: The residuals do not exhibit well-defined, non-linear patterns. Points around the line y = 0 seem randomly scattered in the residual plot. Linearity assumption is fulfilled. Furthermore, the linear relationship between the response variable and the predictor variable can be tested through a Components + Residual Plot (crPlot).



Graph 16: Component + Residual plot of Number of defects per million items received from suppliers (response variable) against the Number of Months from July 2011 the data is collected in (predictor variable).

Since the crPlot shows a graph near linearity, we can accept that the model fulfills the test of linearity.

Normality of data: Based on the QQ plot, the residuals seem normally distributed, as many of the residuals points on the QQ-plot seem to fall close to a straight line. However, this can be further verified with other tests of normality, such as by conducting the Shapiro-Wilk normality test on the residuals of the model.

```
        Shapiro-Wilk normality test

data:  residuals(defects.new.rate.mod)
W = 0.94461, p-value = 0.04534
```

Based on the p-value from the Shapiro-Wilk normality test, since p-value = 0.04534 < 0.05, we reject the null hypothesis that the residuals in the model are normally distributed. But in general, regression analysis is fairly robust against departures from normality, so in most cases this is not a serious issue.

Homoscedasticity: The variation about the regression line appear relatively constant for all values of the independent variable, as the residuals are spread equally along the range of predictors (Scale-Location plot relatively close to a horizontal line with equally, randomly spread points), and hence the model fulfills the assumption of homoscedasticity. We can validate this statement by running a Score Test for Non-Constant Error Variance (ncvTest) using the model.

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.03656066, Df = 1, p = 0.84836
```

In the ncvTest, the null hypothesis is that there is constant error variance and the alternative hypothesis is that the error variance changes with the level of fitted values of the response variable or linear combination of predictors. As p-value is 0.84836 >0.05, we cannot reject the null hypothesis that there is constant error variance. This model thus fulfils the assumption of homoscedasticity.

Independence of errors: Residual plot does not seem to have clusters of residuals with the same sign, and hence autocorrelation does not seem to exist in this case, and errors can be considered independent.

Conclusion: Since the assumptions required for this regression analysis are generally fulfilled, we can proceed on with the construction of our simple linear regression model.

The following shows the summary of the linear model fitted with the cubic polynomial function that we have obtained previously as best-fitted for the data.

```
Call:
lm(formula = defects2$`Defects per million items received from suppliers` ~ 
    defects2$Month + defects2$Month^2 + defects2$Month^3)

Residuals:
    Min      1Q  Median      3Q     Max 
-44.416 -26.338   1.382  19.422  59.063 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    904.9390     9.6729   93.55   <2e-16 ***
defects2$Month -11.0401     0.4013  -27.51   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.4 on 39 degrees of freedom
Multiple R-squared:  0.951,	Adjusted R-squared:  0.9497 
F-statistic: 756.9 on 1 and 39 DF,  p-value: < 2.2e-16
```

The multiple R-squared value of the model is high, at 0.951. This means that the model explains 95.1% of the variation of the dependent variable. With a p-value lower than 0.05 (p-value < 2.2e-16), the predictor variable used here (number of months from July 2011 the data was collected in) is also statistically significant. Similarly, with an overall p-value that is less than 0.05 (p-value < 2.2e-16), the model is statistically significant as well. This implies that the model fits the data well.

With the coefficients of the cubic polynomial trendline, the following equation is formed:

$$y = -1.153\,(10^{-4})\,x^3 - 1.482\,(10^{-1})\,x^2 - 4.633\,x + 858.594$$

where y is the Number of Defects per million items received from suppliers and x is the number of months from July 2011 the data was collected in.

From our trend line equation, we can predict our y value in the near future (where $x \geq 42$). For example, we can use substitute $x = 42$ to find our predicted y-value ($\hat{y}$) in January 2015, had the initiative not been implemented.

$$\hat{y} = -1.153\,(10^{-4})\,(42^3) - 1.482\,(10^{-1})\,(42^2) - 4.633(42) + 858.594 = 394.0409$$

The predicted number of defects per million items received from the supplier is thus 394.0409 from our model, and when rounded down to an integer, $\hat{y} = 394$.

To predict how the number of defects per million items will reduce in the near future, a 6-month prediction using our model is tabulated and shown below.

Jan 2015	394.0409
Feb 2015	376.1860
Mar 2015	358.0051
Apr 2015	339.4973
May 2015	320.6620
June 2015	301.4984

Table 3: Predictions for the number of defects per million items for the next 6 months, from January 2015 to June 2015, given the implementation of the supplier initiative in August 2011.

This shows how the number of defects per million items received from sellers is reduced for the subsequent months.

For the next part of this question, Elizabeth would like to identify the characteristics of individuals that lead to greater retention by studying the effect that years of education, college grade point average and age when hired have on retention.

To answer this question, we will make use of multiple linear regression to form a regression equation with years in PLE (Employee Retention) as the dependent variable while the factors under inspection (years of education, college grade point average and age when hired) as independent variables.

The following shows the summary of the linear model with all the factors under inspection as independent variables and years in PLE as dependent variable.

```
Call:
lm(formula = ER$YearsPLE ~ ER$YrsEducation + ER$`College GPA` +
    ER$Age)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3299 -1.6122 -0.2433  1.8893  4.6312

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.73711    4.50415  -0.608   0.5472
ER$YrsEducation  -0.06705    0.35516  -0.189   0.8513
ER$`College GPA`  0.67998    1.18355   0.575   0.5692
ER$Age            0.29154    0.13504   2.159   0.0376 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.726 on 36 degrees of freedom
Multiple R-squared:  0.1502,  Adjusted R-squared:  0.07939
F-statistic: 2.121 on 3 and 36 DF,  p-value: 0.1146
```

From the summary, we can see that not all predictor variables are statistically significant as some of the variables have a p-value that is greater than 0.05. The adjusted R-squared value of the model is hence low as well, with a value of 0.07939, and the p-value for the overall model is 0.1146, which is greater than 0.05 (implying that the model is statistically insignificant). The independent variable with the highest p-value is Years of Education, with a p-value of 0.8513. Hence, we will remove Years of Education from our linear model in an attempt to improve the model.

The following shows the summary of the linear model with Age when hired and College GPA as independent variables and years in PLE as dependent variable.

```
Call:
lm(formula = ER$YearsPLE ~ ER$`College GPA` + ER$Age)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.3895 -1.5889 -0.1528  1.9828  4.5909 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)  
(Intercept)       -3.2052     3.7112  -0.864   0.3933  
ER$`College GPA`   0.5574     0.9765   0.571   0.5716  
ER$Age             0.2826     0.1248   2.264   0.0295 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.69 on 37 degrees of freedom
Multiple R-squared:  0.1494,	Adjusted R-squared:  0.1034 
F-statistic: 3.248 on 2 and 37 DF,  p-value: 0.05015
```

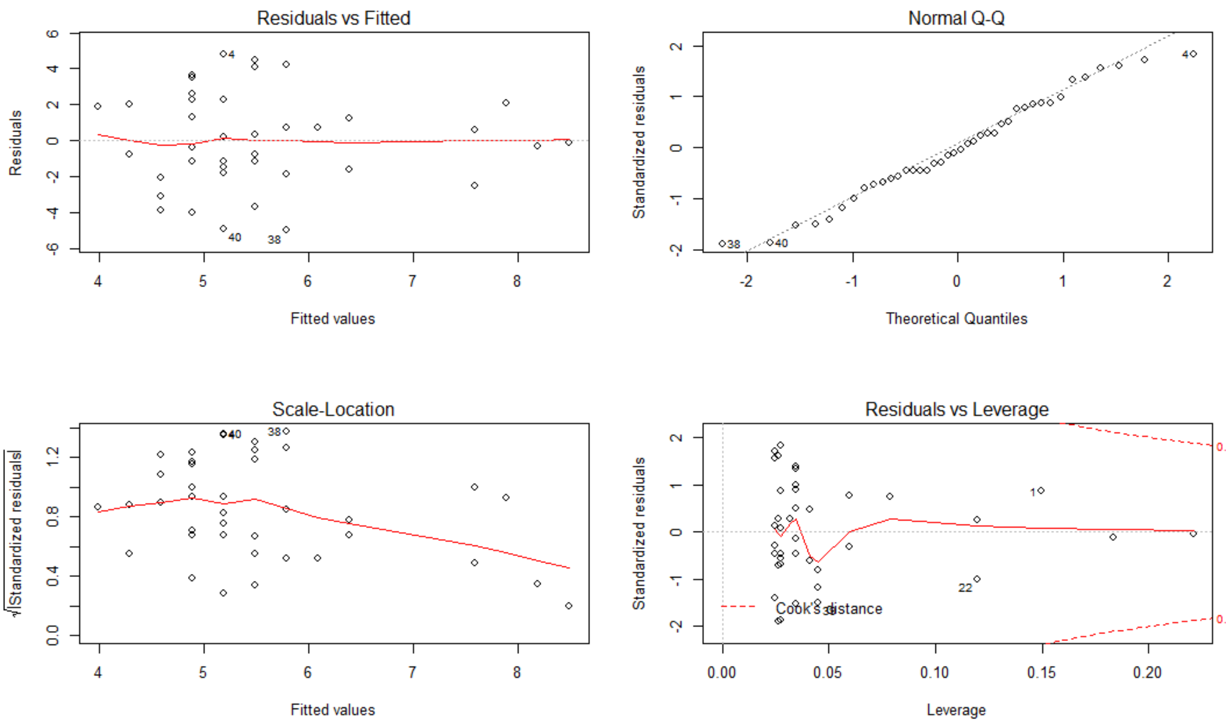
From the summary, we can see that not all predictor variables are statistically significant as College GPA has a p-value that is greater than 0.05, which means that we cannot reject the null hypothesis that the coefficient of College GPA in the regression equation is equals to zero. The adjusted R-squared value of the model is still low but has improved from the previous model – from 0.07939 to 0.1034; the p-value of the overall model has decreased as well from 0.1146 to 0.05015, showing that the model has also improved in statistical significance from the previous model. The independent variable with the highest p-value is College GPA, with a p-value of 0.5716. Hence, we will remove College GPA from our linear model in an attempt to further improve the model.

The following shows the summary of the linear model with Age when hired and College GPA as independent variables and years in PLE as dependent variable.

```
Call:
lm(formula = ER$YearsPLE ~ ER$Age)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.9927 -1.6430 -0.1952  2.0328  4.8078 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  -2.0149     3.0425  -0.662   0.5118  
ER$Age        0.3003     0.1198   2.506   0.0166 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.666 on 38 degrees of freedom
Multiple R-squared:  0.1419,	Adjusted R-squared:  0.1193 
F-statistic: 6.282 on 1 and 38 DF,  p-value: 0.01659
```

From the summary, we can see that now all the independent variables are statistically significant (with a p-value < 0.05). Age is the only independent variable left in the regression equation and it is statistically significant, with a p-value of 0.0166, which is less than 0.05. Adjusted R-squared value remains low, but has increased from 0.1034 to 0.1193, and the overall p-value of the model has dropped from 0.05015 to 0.01659, showing that this model has improved from the previous model. Since the model has assumed statistical significance and all independent variables in the model are statistically significant, this is the best model we can obtain.

Now, before we fit the model into a multiple regression equation, we must first check that the assumptions of regression analysis are met by the model.



Linearity: The residuals do not exhibit well-defined, non-linear patterns. Points around the line y = 0 seem randomly scattered in the residual plot. Graph in residual plot is nearly horizontal at y=0. Linearity assumption is fulfilled.

Normality of data: Based on the QQ plot, the residuals seem normally distributed, as most of the residuals points on the QQ-plot seem to fall close to a straight line. However, this can be further verified with other tests of normality, such as by conducting the Shapiro-Wilk normality test on the residuals of the model.

```
        Shapiro-Wilk normality test

data:  residuals(fit)
W = 0.97632, p-value = 0.5552
```

Based on the p-value from the Shapiro-Wilk normality test, since p-value = 0.5552 > 0.05, we cannot reject the null hypothesis that the residuals in the model are normally distributed. Hence, this model has fulfilled the assumption of normality.

Homoscedasticity: The variation about the regression line appear relatively constant for all values of the independent variable, as the residuals are spread equally along the range of predictors (Scale-Location plot relatively close to a horizontal line with equally, randomly spread points), and hence the model fulfills the assumption of homoscedasticity. We can validate this statement by running a Score Test for Non-Constant Error Variance (ncvTest) using the model.

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.8679938, Df = 1, p = 0.35151
```

In a ncvTest, the null hypothesis is that there is constant error variance and the alternative hypothesis is that the error variance changes with the level of fitted values of the response variable or linear combination of predictors. As p-value is 0.35151 > 0.05, we cannot reject the null hypothesis that there is constant error variance. This model thus fulfils the assumption of homoscedasticity.

Independence of errors: Residual plot does not seem to have clusters of residuals with the same sign, and hence autocorrelation does not seem to exist in this case, and errors can be considered independent.

Conclusion: Since the assumptions required for this regression analysis are generally fulfilled, we can proceed on with the construction of our multiple linear regression model.

Multiple regression line equation fitted:

$$\text{Number of years retained in PLE} = 0.3003 \text{ Age} - 2.0149$$

Hence, the characteristic that will lead to greater employee retention is older age.

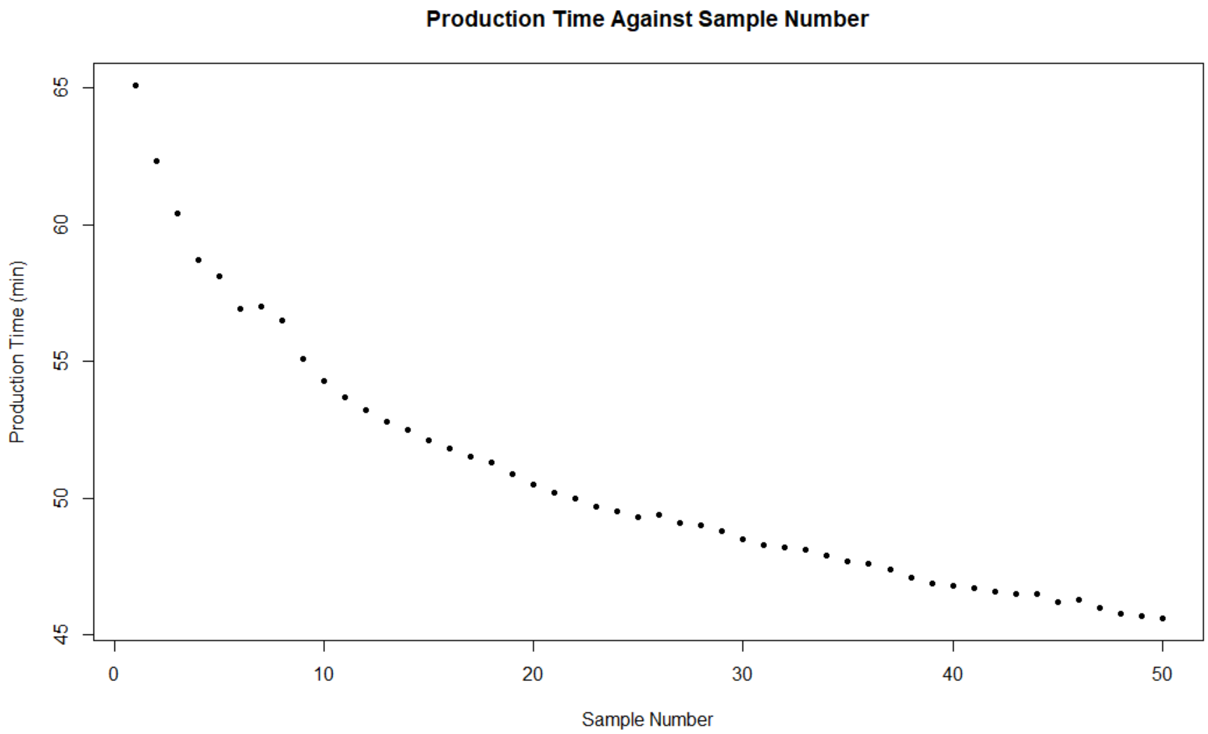
However, this model might not be accurately reflective of the factors affecting the actual employee retention as well, namely because:

- The adjusted R-squared value of the model remains low at 0.1193 even after all the independent variables are statistically significant. The model is hence only able to explain for 11.9% of the variability of the dependent variable (Number of years retained in PLE) around its mean. More than 50% of the variation of the dependent variable cannot be explained by the model.
- We are only examining the statistical significance of 3 factors – Years of Education, College GPA and Age when hired – in determining the number of years an employee is retained in PLE. There are many other factors that can influence the number of years an employee is retained in PLE, and hence the model might not be representative of all the possible predictor variables. Predictions made by the model are hence not necessarily accurate.

As for the third part of the question, Elizabeth would like to understand the rate of learning of the firm in using new production technology.

For this question, we are given the amount of time required to produce each of 50 units of engines that PLE has successively produced on its production line.

We would first plot a scatterplot of the Production Time against the Sample Number (with 1 being the first unit of engine produced on the production line, 2 being the second unit of engine produced etc.) to better examine the data.



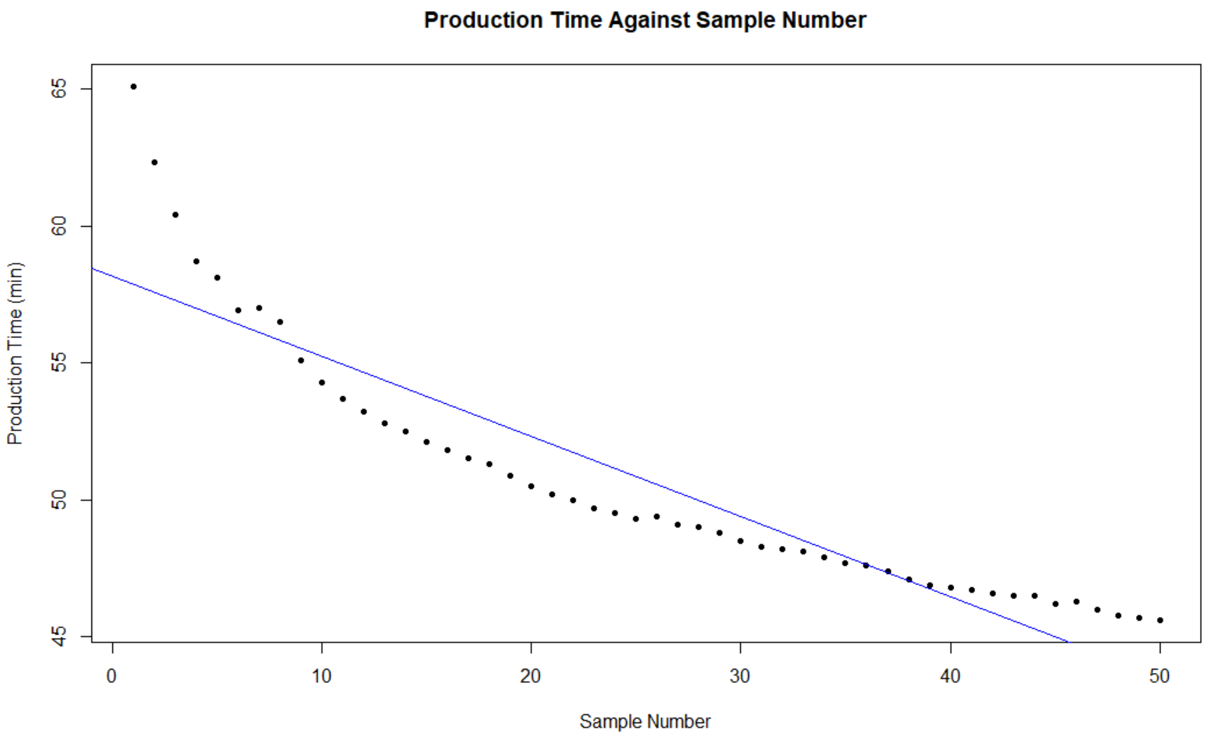
Graph 17: A scatterplot with Production Time (min) plotted against Sample Number, using the data in the Engines sheet.

From Graph 17, we can see a general downward trend in the graph, i.e. as the Sample Number increases, the time required to produce the engine decreases. Rate of improvement, which is the magnitude of the negative gradient of the graph, decreases as Sample Number increases. These correspond to what was mentioned in the extract: “When new production technology is introduced, firms often experience learning, resulting in a gradual decrease in the time required to produce successive units. Generally, the rate of improvement declines until the production time levels off.” (quoted from the extract)

To understand the rate of learning of the firm in using new production technology, simple linear regression can be employed to obtain a regression equation with Production Time (mins) as the dependent variable and the Sample Number as the independent variable. Subsequently, we can analyse the rate of learning of the firm by looking at the amount decreased in Production Time given a unit increase in the Sample Number.

To start off, we will attempt to fit different trendlines to the scatterplot and determine which trendline function will better fit the data.

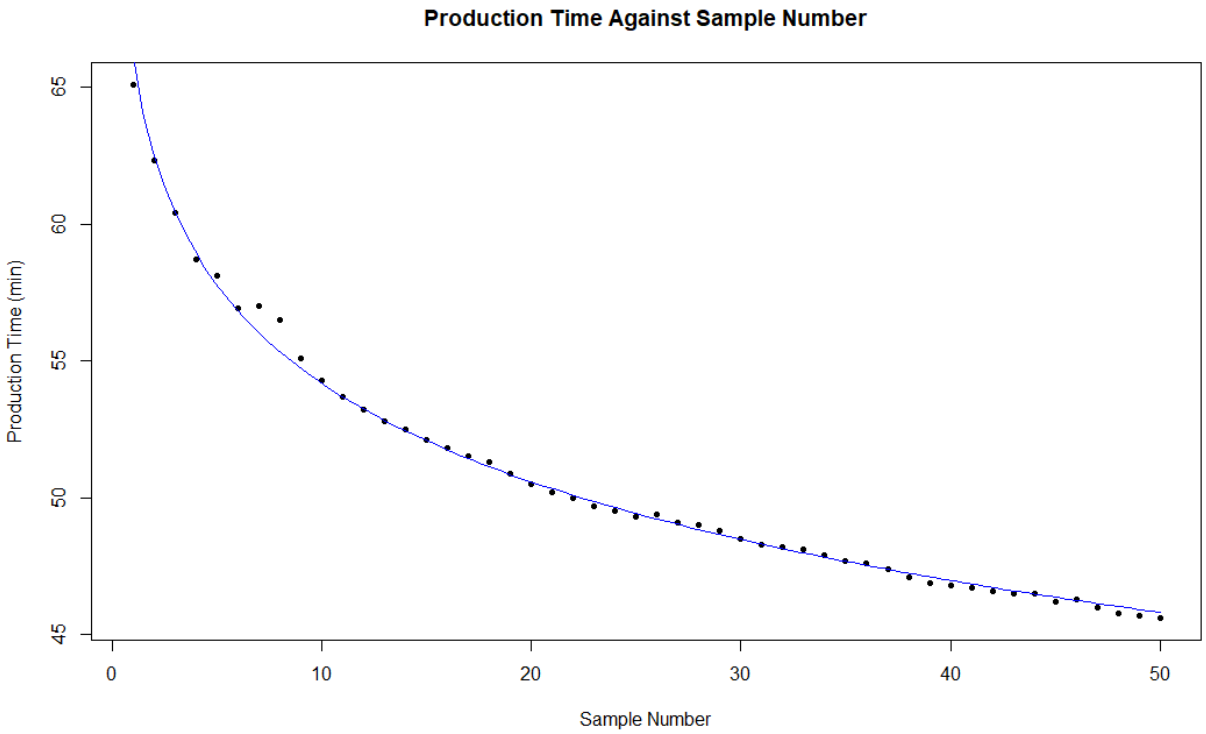
1. Linear



Graph 18: Linear trendline modelled onto the scatterplot.

The multiple R-squared value for the linear trendline model is 0.849, and the adjusted R-squared value for the linear trendline model is 0.846. In this case, the multiple R-squared value is high as it shows that the model explains 84.9% of the variability of the dependent variable (Production Time (mins)) around its mean.

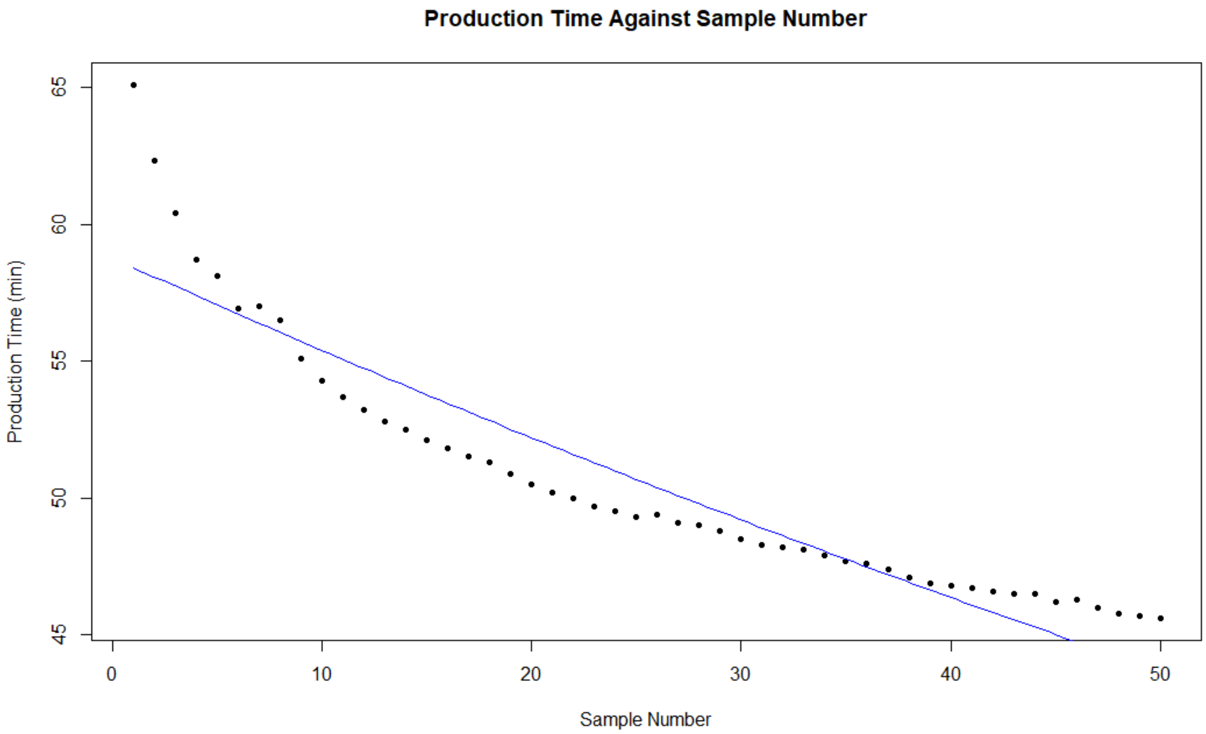
2. Logarithmic



Graph 19: Logarithmic trendline modelled onto the scatterplot.

The multiple R-squared value for the logarithmic trendline model is 0.996. The multiple R-squared value is very high as it shows that the model explains 99.6% of the variability of the dependent variable (Production Time (mins)) around its mean.

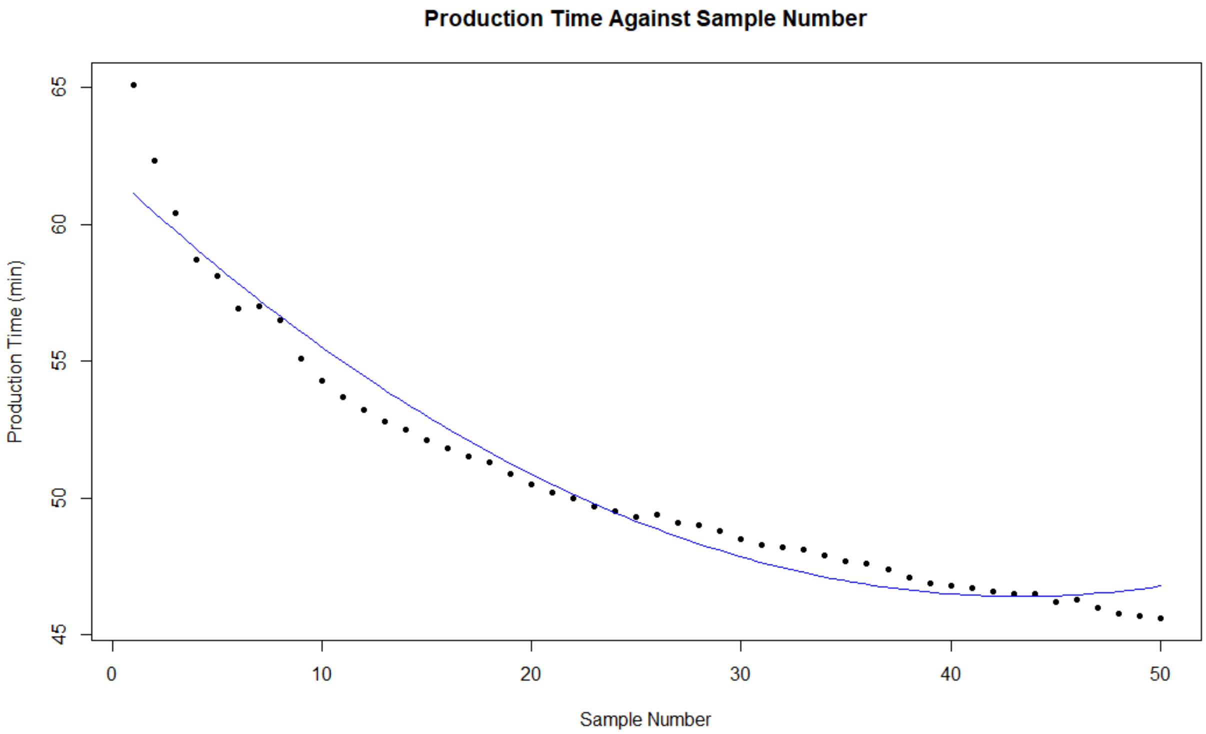
3. Exponential



Graph 20: Exponential trendline modelled onto the scatterplot.

The multiple R-squared value for the exponential trendline model is 0.872. The multiple R-squared value is high as it shows that the model explains 87.2% of the variability of the dependent variable (Production Time (mins)) around its mean.

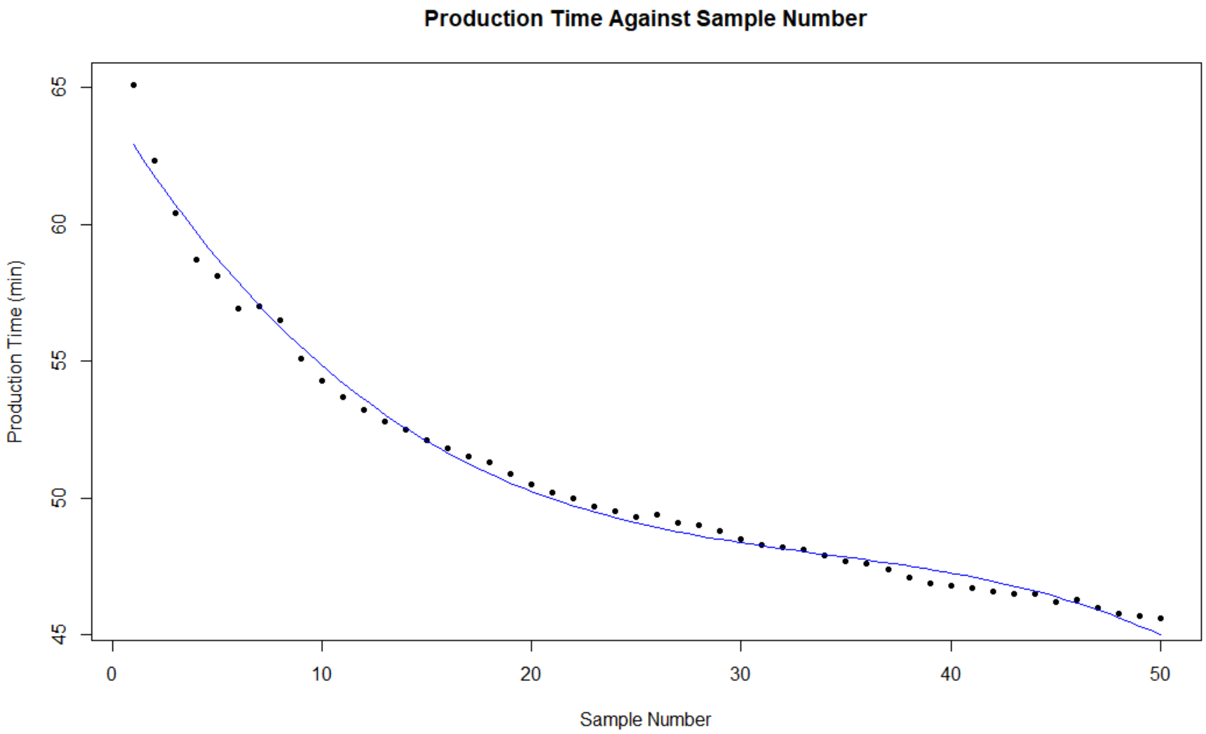
4. Polynomial



Graph 21: Polynomial trendline modelled onto the scatterplot.

The multiple R-squared value for the polynomial trendline model is 0.961. The multiple R-squared value is very high as it shows that the model explains 96.1% of the variability of the dependent variable (Production Time (mins)) around its mean.

5. Cubic Polynomial

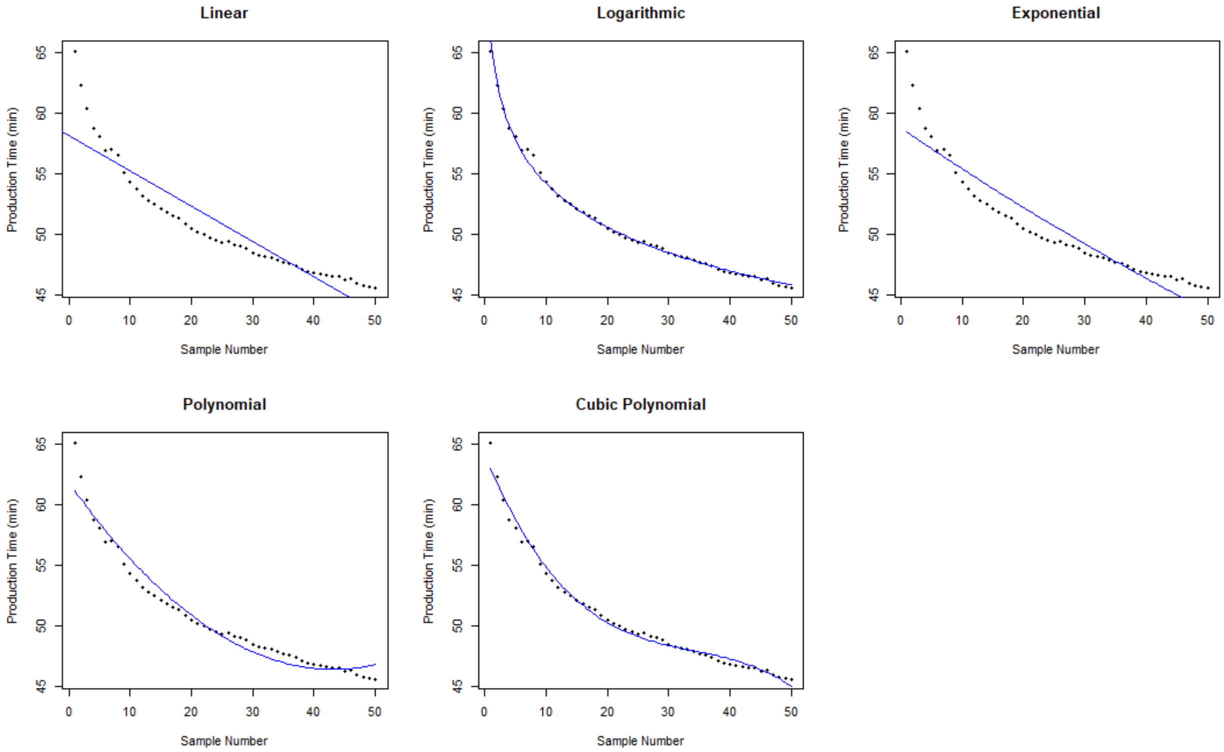


Graph 22: Cubic polynomial trendline modelled onto the scatterplot.

The multiple R-squared value for the cubic polynomial trendline model is 0.989. The multiple R-squared value is very high as it shows that the model explains 98.9% of the variability of the dependent variable (Production Time (mins)) around its mean.

Model	**R-squared value (3 sig. fig.)**
Linear	0.849
Logarithmic	0.996
Exponential	0.872
Polynomial	0.961
Cubic Polynomial	0.989

Table 3: R-squared values tabulated for the respective models.

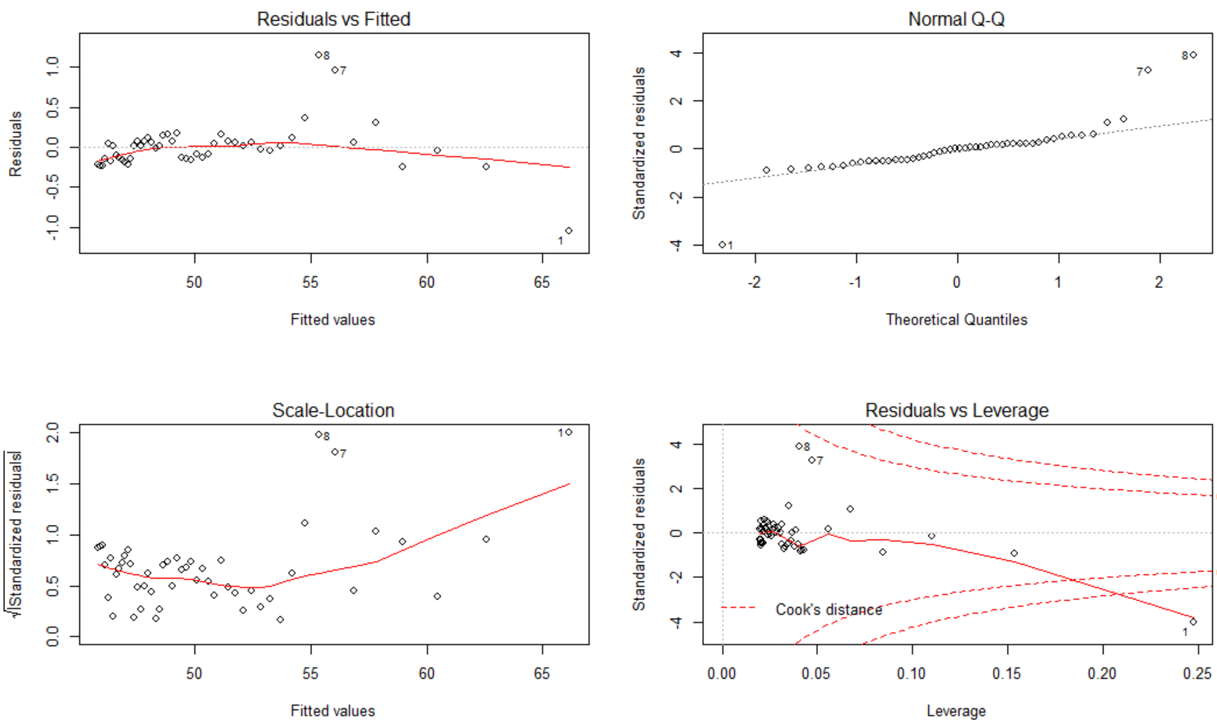


Graph 23: A collation of all the previous graphs in the same window for ease of comparison.

As seen from Table 3 and Graph 23, the Logarithmic model has the highest R-squared value and the best-fitted trendline function as compared to the other models, and hence the logarithmic trendline is best-fitted for the data.

Next, we can continue to conduct simple linear regression. We will employ the logarithmic trendline to fit our data here.

Before conducting the regression analysis, we must first check for the assumptions associated with regression analysis.



Linearity: The residuals do not exhibit well-defined, non-linear patterns. Points around the line y = 0 seem randomly scattered in the residual plot. Linearity assumption is fulfilled.

Normality of data: Based on the QQ plot, residuals appear relatively normally distributed, as most of the points on the QQ-plot seem to fall close to a straight line.

Homoscedasticity: The variation about the regression line does not appear relatively constant for all values of the independent variable, as the residuals are not spread equally along the range of the predictor (with a higher concentration of residuals for x values lower than 55, and a lower concentration of residuals for x values higher than 55, x here being the independent variable). The graph plotted on the Scale-Location plot is thus not close to a horizontal line with equally, randomly spread points. To validate whether the model fulfills the assumption of homoscedasticity, we can run a Score Test for Non-Constant Error Variance (ncvTest) using the model.

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 52.46792, Df = 1, p = 4.3733e-13
```

In the ncvTest, the null hypothesis is that there is constant error variance and the alternative hypothesis is that the error variance changes with the level of fitted values of the response variable or linear combination of predictors. As p-value is 4.3733e-13 < 0.05, we reject the null hypothesis that there is constant error variance. This model thus does not fulfil the assumption of homoscedasticity, but since there are no serious violations of the assumption, regression analysis can still be fairly robust against departures from homoscedasticity.

Independence of errors: Residual plot does not seem to have clusters of residuals with the same sign, and hence autocorrelation does not seem to exist in this case, and errors can be considered independent.

Conclusion: Since the assumptions required for this regression analysis are generally fulfilled, we can proceed on with the construction of our simple linear regression model.

The following shows the summary of the linear model fitted with logarithmic function that we have obtained previously as best-fitted for the data.

```
Call:
lm(formula = Eng$`Production Time (min)` ~ log(Eng$Sample))

Residuals:
     Min       1Q   Median       3Q      Max
-1.05175 -0.14451  0.00909  0.06921  1.15299

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     66.15175    0.14975   441.8   <2e-16 ***
log(Eng$Sample) -5.19598    0.04835  -107.5   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3009 on 48 degrees of freedom
Multiple R-squared:  0.9959,  Adjusted R-squared:  0.9958
F-statistic: 1.155e+04 on 1 and 48 DF,  p-value: < 2.2e-16
```

The multiple R-squared value of the model is very high, at 0.996. This means that the model explains 99.6% of the variation of the dependent variable. With a p-value lower than 0.05 (p-value < 2.2e-16), the predictor variable used here (number of months from July 2011 the data was collected in) is also statistically significant. Similarly, with an overall p-value that is less than 0.05 (p-value < 2.2e-16), the model is statistically significant as well. This implies that the model fits the data very well.

With the coefficients of the logarithmic trendline, the following equation is formed:

$$y = -5.196 \log(x) + 66.152$$

where y is the amount of time required to produce an engine in minutes and x is the sample number, where the first engine produced has a sample number of 1, and the second engine produced a sample number of 2, and etc.

Hence, from our regression equation, you can see that the rate of learning – the rate of decrease in Production Time with respect to the Sample Number – is given by the differentiation of the above equation.

The first derivative of the above mathematical expression is:

$$\frac{d}{dx}(66.152 - 5.196\log(x)) = -\frac{5.196}{x}$$

Note that log(x) here is the natural logarithm of x.

Hence, we can conclude that the rate of learning of the firm (the rate of decrease in Production Time with respect to the Sample Number) is given by $\frac{5.196}{Current\ Sample\ Number}$ minutes per unit increase in Sample Number. But there is a lower-bound asymptote at y=0 (i.e. Production Time > 0), regardless of the value of x.

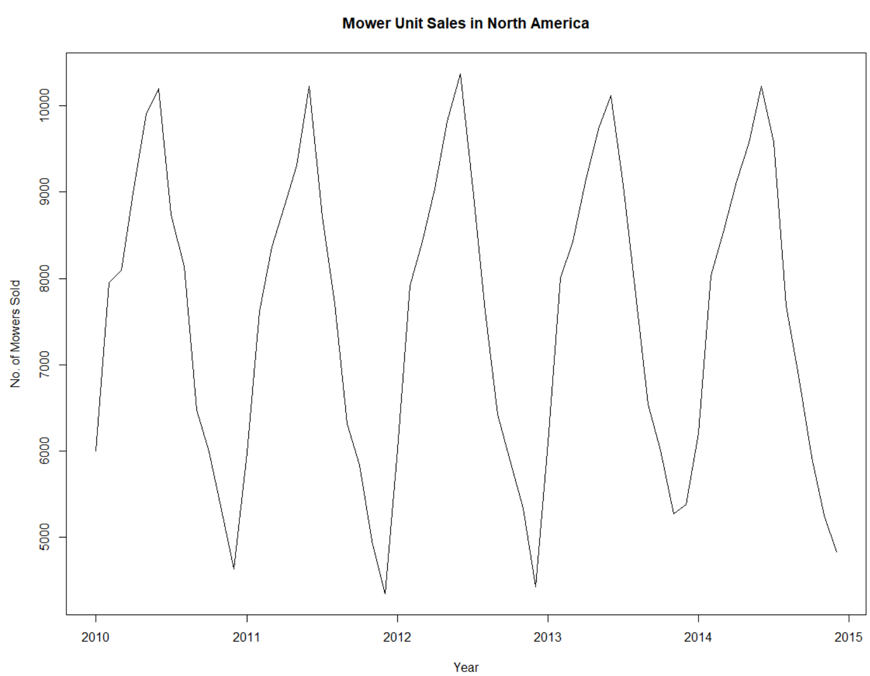
Question 2: Elizabeth Burke is interested in forecasting sales of mowers and tractors in each marketing region as well as industry sales to assess future changes in market share. She also wants to forecast future increases in production costs. Develop forecasting models for these data and prepare a report of your results with appropriate charts and output.

Firstly, we will forecast the sales of mowers in each marketing region using the excel sheet *Mower Unit Sales.*

Forecasting of Unit Sales of Mowers in North America

Time Series:

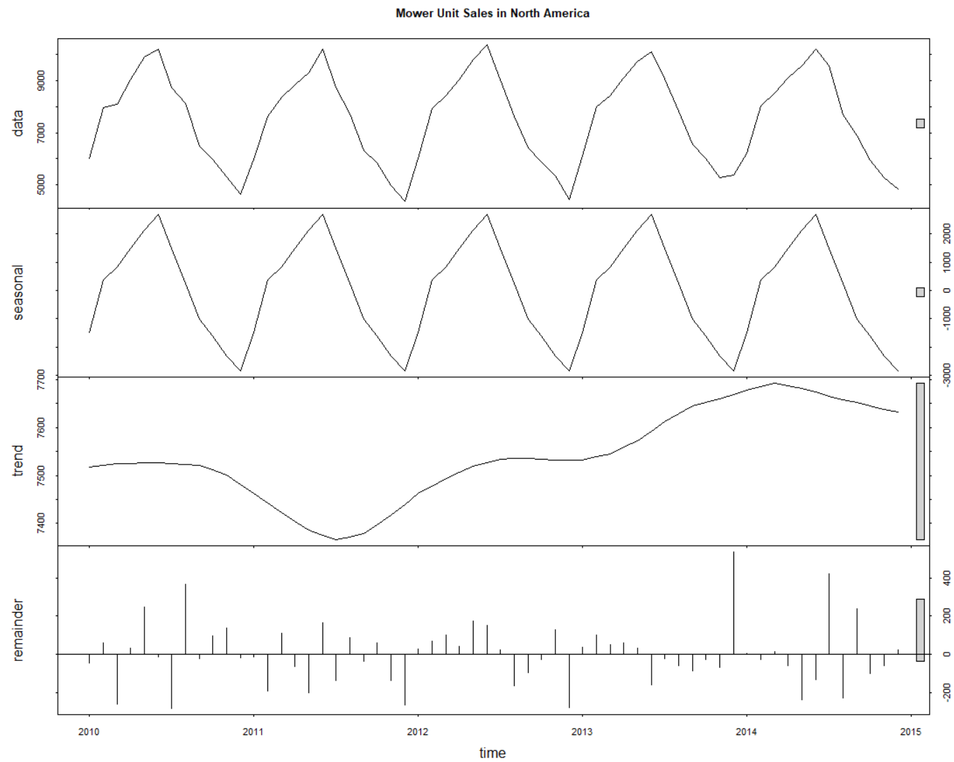
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	6000	7950	8100	9050	9900	10200	8730	8140	6480	5990	5320	4640
2011	5980	7620	8370	8830	9310	10230	8720	7710	6320	5840	4960	4350
2012	6020	7920	8430	9040	9820	10370	9050	7620	6420	5890	5340	4430
2013	6100	8010	8430	9110	9730	10120	9080	7820	6540	6010	5270	5380
2014	6210	8030	8540	9120	9570	10230	9580	7680	6870	5930	5260	4830



Graph 24 : The time series graph above shows the mower unit sales in North America from Janua ry 2010 to December 2014.

From the graph, there is no apparent trend in the data over this period.
However, mower unit sales shows strong seasonality each year. There are periodic fluctuations every year, where the mower unit sales increases till mid-year before decreasing back to around the original volume of mower unit sales that was sold at the start of the year.

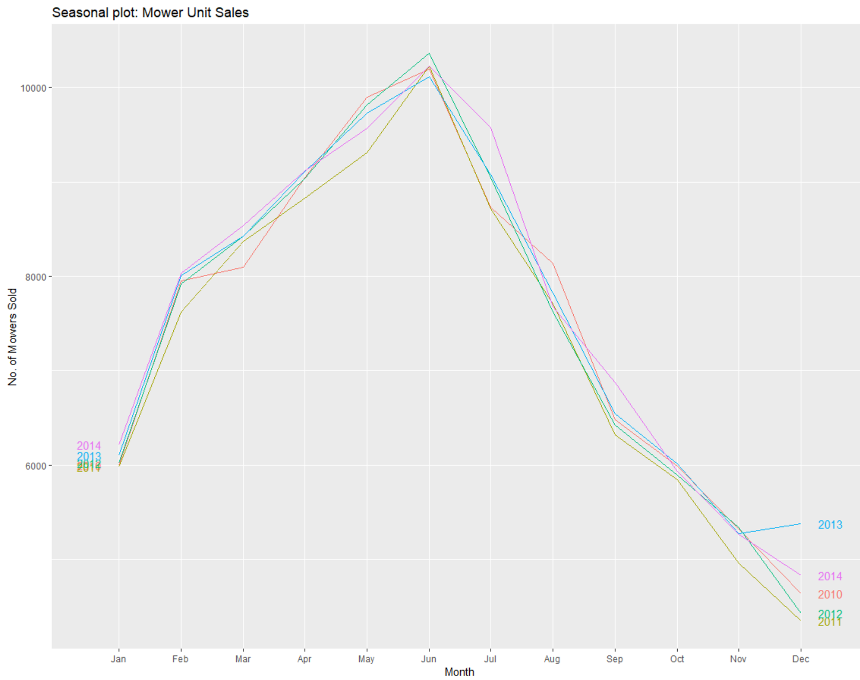
To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 25: Decomposed time series data of mower unit sales in North America from January 2010 to December 2014

Decomposition of the time series further proves the point that there is no apparent trend but there is the presence of a clear seasonality.

A seasonal plot allows the underlying seasonal pattern to be seen more clearly, and is especially useful in identifying years in which the pattern changes. Hence, we can use a seasonal plot to better ascertain the type of seasonality present in the data.

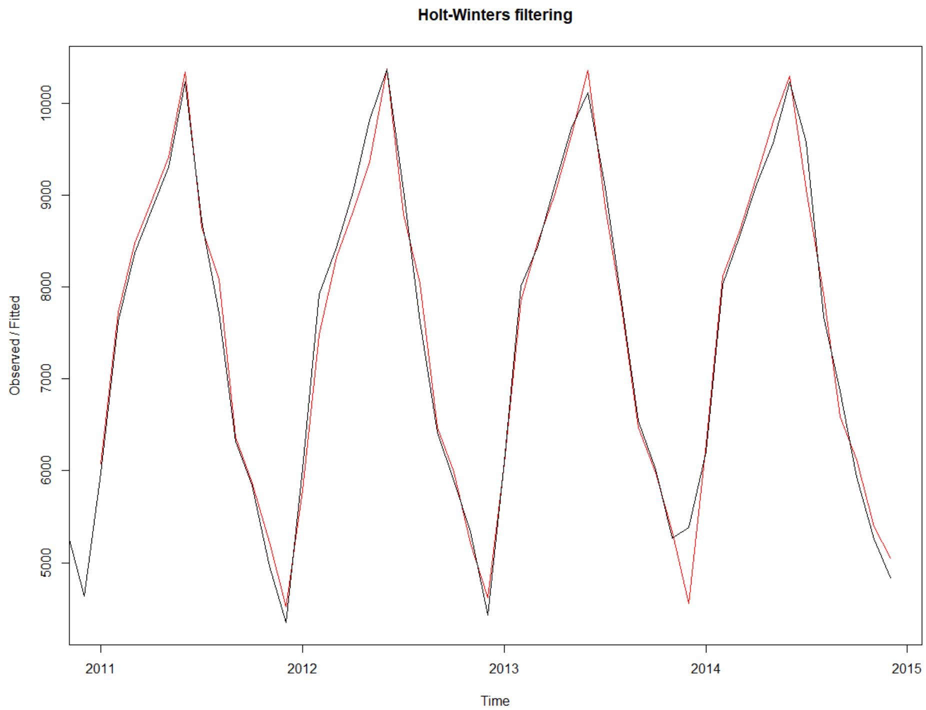


Graph 26: A seasonal plot of mower unit sales in North America from January 2010 to December 2014.

From the seasonal plot, it is clear that there is an increase of mower unit sales till June before it decreases all the way till the end of December. Plus, we can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does not vary with the level of the time series

Hence, we can conclude that the sales of mowers in North America possess no trend but has seasonality.

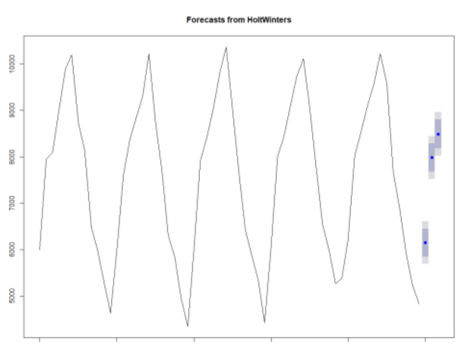
We can use the Holt-Winters no-trend smoothing to forecast future sales.

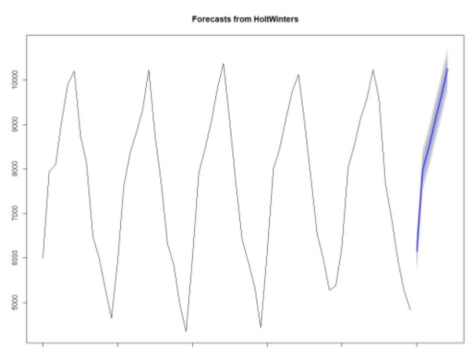


Graph 27: Observed time series data of mower unit sales in North America against the fitted Holt-Winters no-trend smoothing model

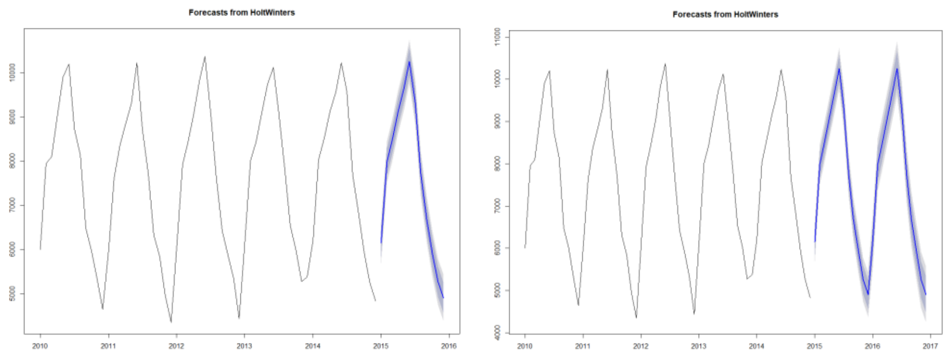
We see from the plot that the Holt-Winters no-trend smoothing model is very successful in predicting the seasonal peaks, which occur every June/July of the year. The model fits well with the observed time series data.

Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months:





Graphs 28 and 29: Forecasted data of unit mower sales in North America over the next 3 and 6 months respectively



Graphs 30 and 31: Forecasted data of unit mower sales in North America over the next 12 and 24 months respectively

The following are forecasted values of mower unit sales in North America:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	6150.553	5850.927	6450.180	5692.314	6608.792
Feb 2015	7985.535	7681.653	8289.417	7520.787	8450.283
Mar 2015	8495.571	8187.492	8803.650	8024.404	8966.737
Apr 2015	9096.602	8784.383	9408.822	8619.104	9574.101
May 2015	9633.307	9317.001	9949.613	9149.558	10117.055
Jun 2015	10252.305	9931.965	10572.646	9762.387	10742.224
Jul 2015	9328.118	9003.794	9652.442	8832.107	9824.129
Aug 2015	7717.801	7389.541	8046.062	7215.771	8219.832
Sep 2015	6689.818	6357.668	7021.967	6181.839	7197.796
Oct 2015	5931.181	5595.188	6267.175	5417.324	6445.039
Nov 2015	5264.650	4924.856	5604.444	4744.980	5784.320
Dec 2015	4900.652	4557.099	5244.204	4375.234	5426.070
Jan 2016	6150.553	5751.479	6549.627	5540.222	6760.884
Feb 2016	7985.535	7583.256	8387.814	7370.302	8600.768
Mar 2016	8495.571	8090.112	8901.030	7875.475	9115.666
Apr 2016	9096.602	8687.989	9505.216	8471.682	9721.523
May 2016	9633.307	9221.562	10045.051	9003.598	10263.016
Jun 2016	10252.305	9837.454	10667.157	9617.844	10886.766
Jul 2016	9328.118	8910.182	9746.054	8688.940	9967.296
Aug 2016	7717.801	7296.804	8138.799	7073.941	8361.661
Sep 2016	6689.818	6265.781	7113.854	6041.309	7338.326
Oct 2016	5931.181	5504.127	6358.236	5278.058	6584.305
Nov 2016	5264.650	4834.599	5694.701	4606.944	5922.357
Dec 2016	4900.652	4467.625	5333.679	4238.394	5562.910

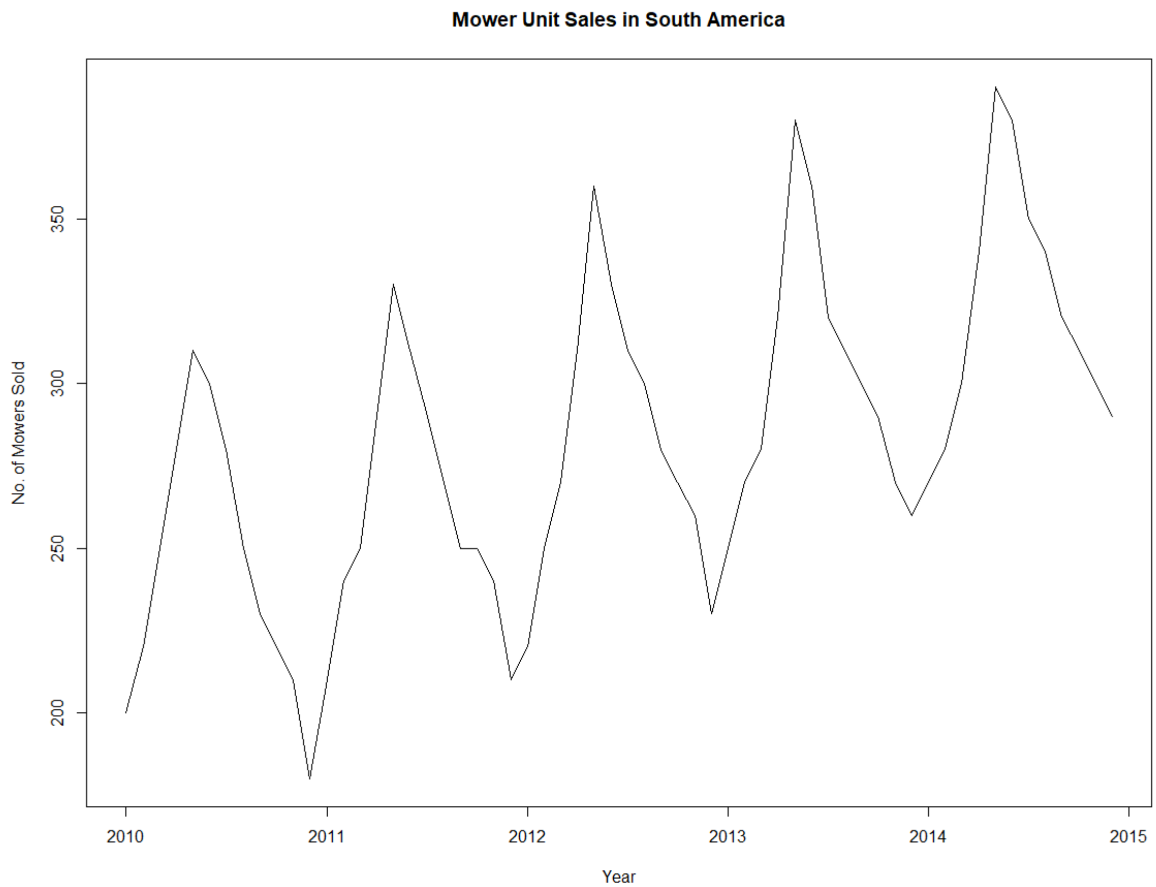
Conclusion:

As seen from the forecasted graphs and values generated, unit mower sales in North America is months predicted to remain the same for the next 24 months to come. Seasonality will remain.

Forecasting of Unit Sales of Mowers in South America

Time Series:

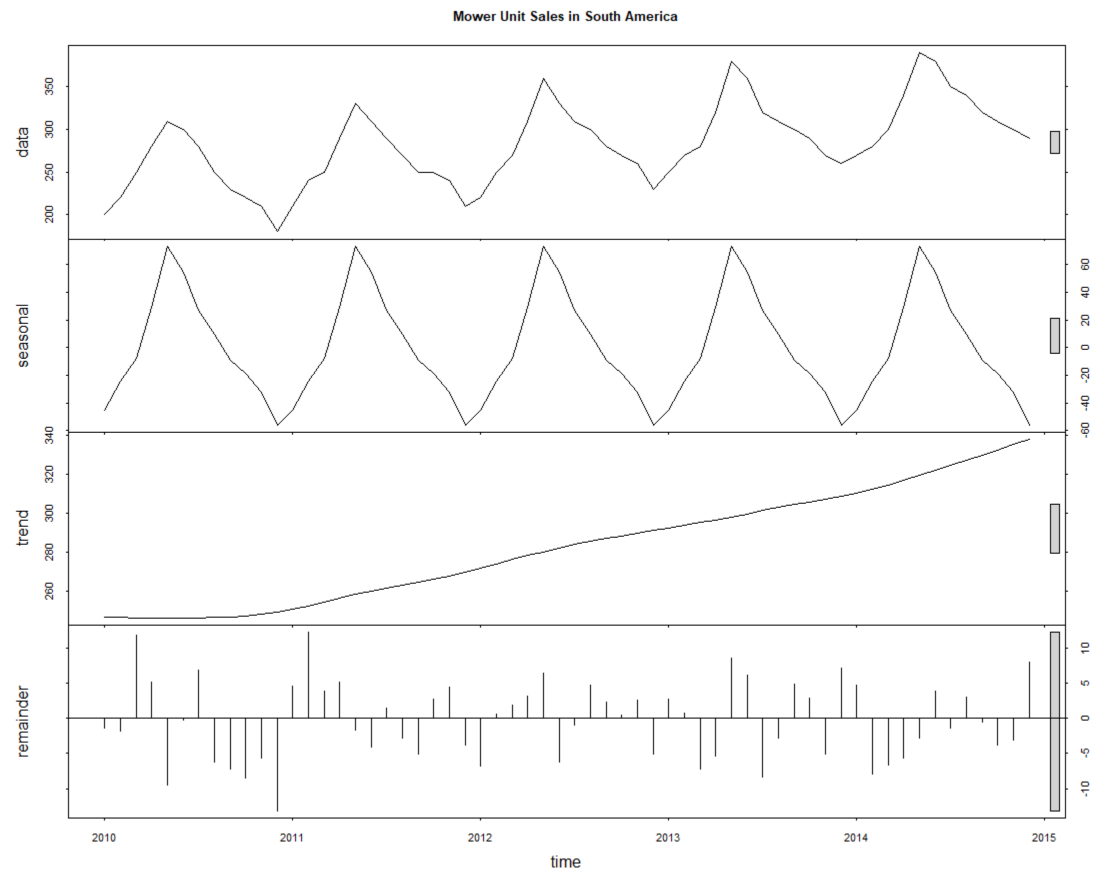
```
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2010 200 220 250 280 310 300 280 250 230 220 210 180
2011 210 240 250 290 330 310 290 270 250 250 240 210
2012 220 250 270 310 360 330 310 300 280 270 260 230
2013 250 270 280 320 380 360 320 310 300 290 270 260
2014 270 280 300 340 390 380 350 340 320 310 300 290
```



Graph 32: The time series graph above shows the mower unit sales in South America from January 2010 to December 2014.

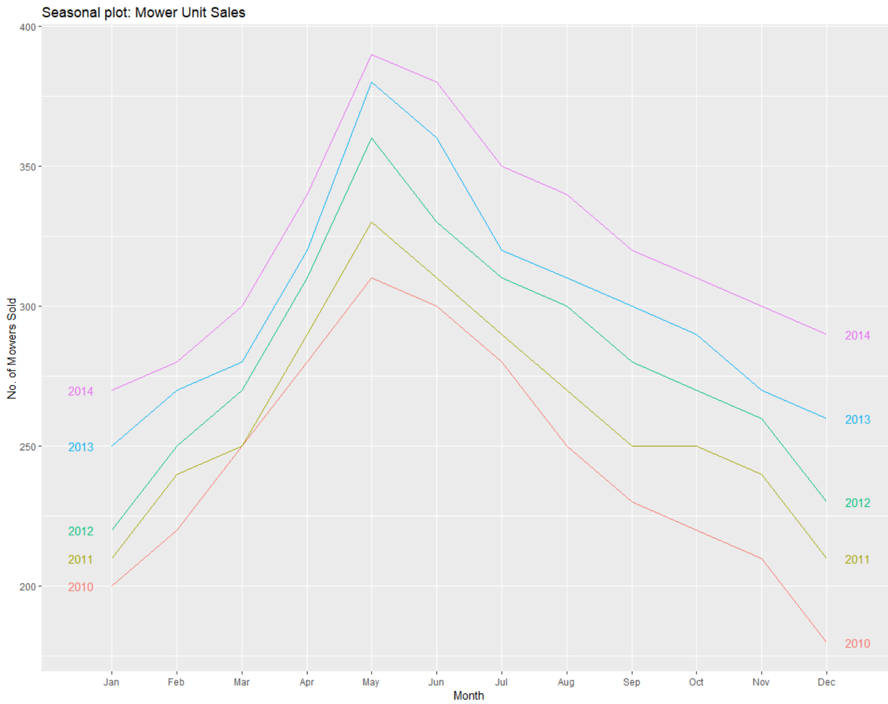
From the graph, there is an upward trend in the data over this period.
Mower unit sales in South America shows strong seasonality each year. There are periodic fluctuations every year, where the mower unit sales increases till mid-year before decreasing back to around the original volume of mower unit sales that was sold at the start of the year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 33: Decomposed time series data of mower unit sales in South America from January 2010 to December 2014

Decomposition of the time series further proves the point that there is a steady upward trend and there is the presence of a clear seasonality which is similar to that of North America's.

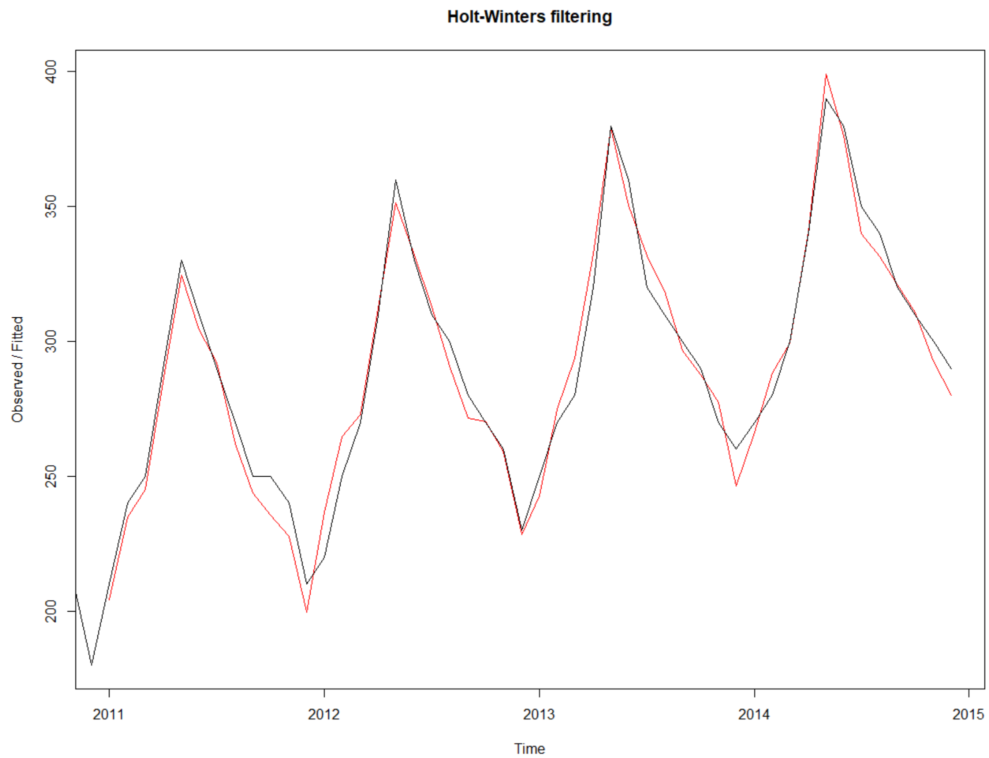


Graph 34: A seasonal plot of mower unit sales in South America from January 2010 to December 2014.

From the seasonal plot, we can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does not vary with the level of the time series.

Hence, we can conclude that the unit sales of mowers in South America possess trend and seasonality.

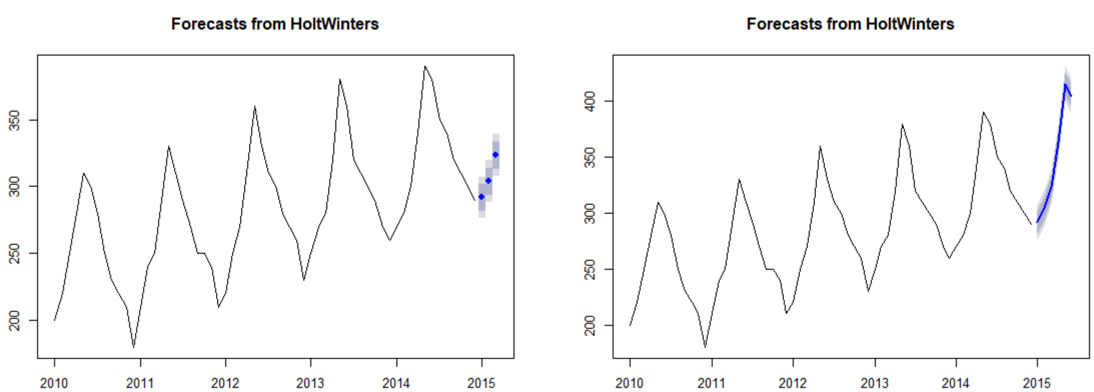
We can use Holt-Winters additive model to forecast future sales.



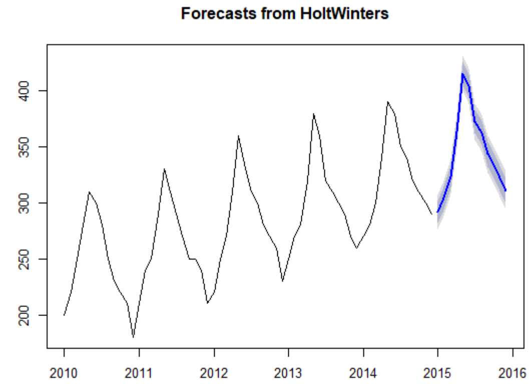
Graph 35: Observed time series data of mower unit sales in South America against the fitted Holt-Winters additive model

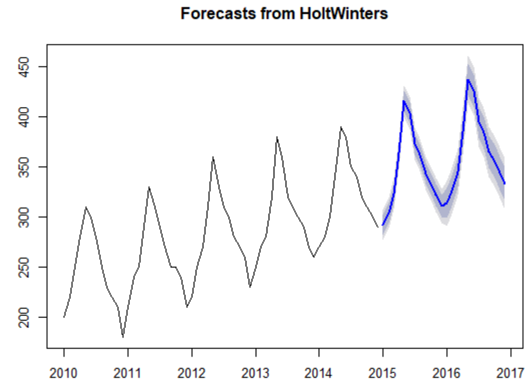
We see from the plot that the Holt-Winters multiplicative model is very successful in predicting the seasonal peaks, which occur every June/July of the year, and the upward trend. The model fits well with the observed time series data.

Hence, we can make use of Holt- Winters additive model to predict data in the next 3,6,12 and 24 months:



Graphs 36 and 37: Forecasted data of unit mower sales in South America over the next 3 and 6 months respectively.





Graphs 38 and 39: Forecasted data of unit mower sales in South America over the next 12 and 24 months respectively.

The following are the forecasted values of mower unit sales in South America:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	291.8960	281.8600	301.9320	276.5472	307.2447
Feb 2015	303.9957	293.8962	314.0951	288.5499	319.4415
Mar 2015	323.3015	313.1314	333.4716	307.7477	338.8553
Apr 2015	363.6535	353.4053	373.9016	347.9802	379.3267
May 2015	415.6133	405.2793	425.9473	399.8088	431.4178
Jun 2015	404.2732	393.8454	414.7010	388.3253	420.2211
Jul 2015	372.8157	362.2859	383.3454	356.7118	388.9195
Aug 2015	362.2294	351.5894	372.8694	345.9569	378.5018
Sep 2015	343.3332	332.5744	354.0919	326.8791	359.7872
Oct 2015	333.5234	322.6373	344.4095	316.8745	350.1723
Nov 2015	322.0859	311.0637	333.1081	305.2289	338.9429
Dec 2015	310.8791	299.7120	322.0461	293.8005	327.9576
Jan 2016	313.6878	298.7641	328.6114	290.8640	336.5115
Feb 2016	325.7875	310.7401	340.8348	302.7745	348.8004
Mar 2016	345.0933	329.9147	360.2719	321.8797	368.3070
Apr 2016	385.4452	370.1278	400.7627	362.0192	408.8713
May 2016	437.4051	421.9410	452.8691	413.7548	461.0553
Jun 2016	426.0650	410.4466	441.6834	402.1787	449.9513
Jul 2016	394.6074	378.8268	410.3881	370.4731	418.7418
Aug 2016	384.0212	368.0705	399.9719	359.6267	408.4156
Sep 2016	365.1250	348.9963	381.2536	340.4583	389.7916
Oct 2016	355.3152	339.0007	371.6297	330.3643	380.2661
Nov 2016	343.8777	327.3694	360.3859	318.6305	369.1249
Dec 2016	332.6709	315.9610	349.3808	307.1153	358.2264

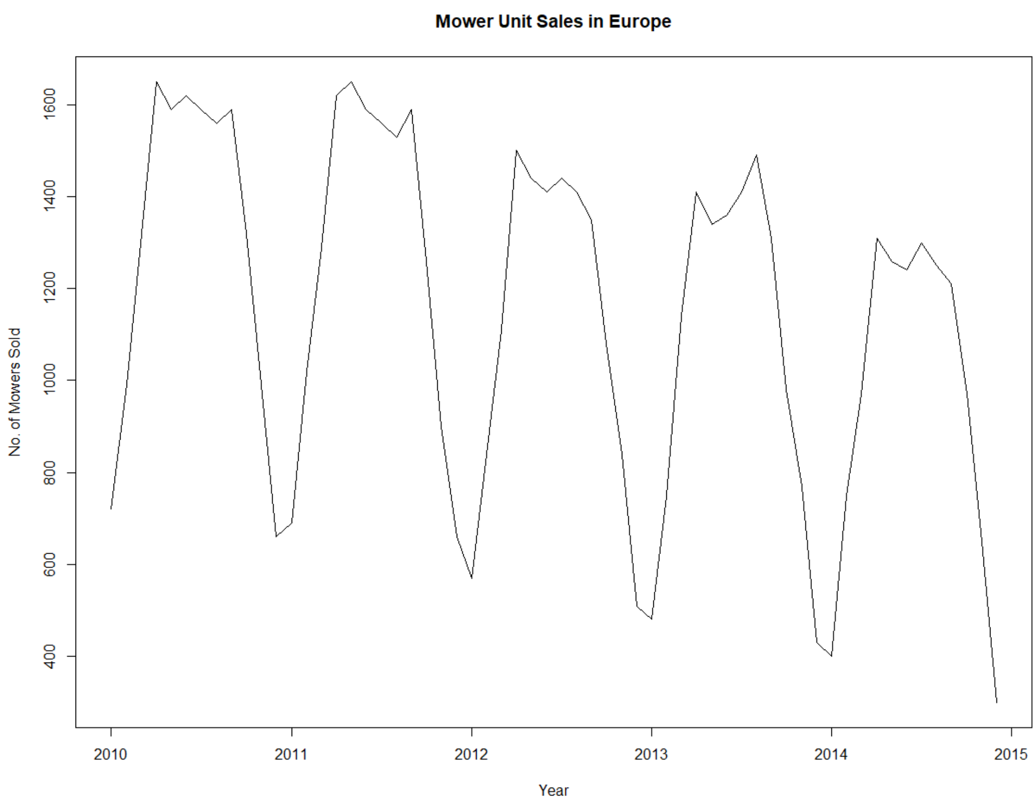
Conclusion:

As seen from the forecasted graphs and values generated, unit mower sales in South America is predicted to increase for the next 24 months to come. Seasonality will remain.

Forecasting of Unit Sales of Mowers in Europe

Time Series:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	720	990	1320	1650	1590	1620	1590	1560	1590	1320	990	660
2011	690	1020	1290	1620	1650	1590	1560	1530	1590	1260	900	660
2012	570	840	1110	1500	1440	1410	1440	1410	1350	1080	840	510
2013	480	750	1140	1410	1340	1360	1410	1490	1310	980	770	430
2014	400	750	970	1310	1260	1240	1300	1250	1210	970	650	300

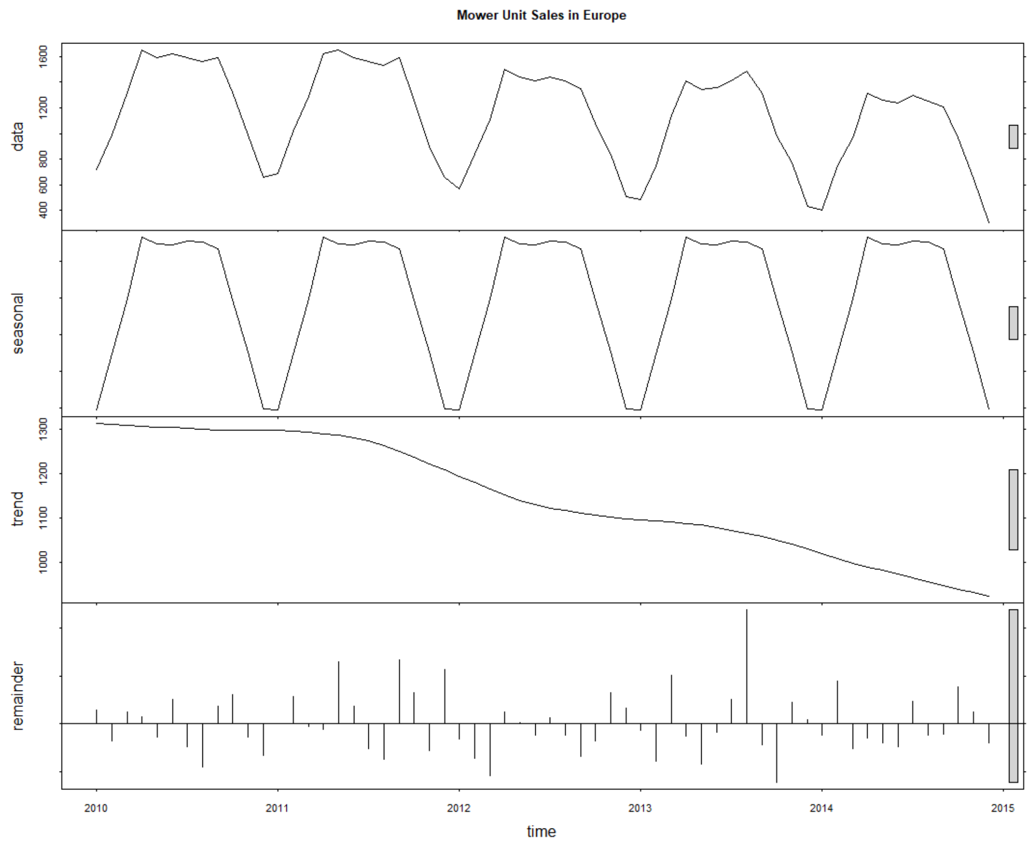


Graph 40: The time series graph above shows the mower unit sales in Europe from January 2010 to December 2014.

From the graph, there is a downward trend in the data over this period.

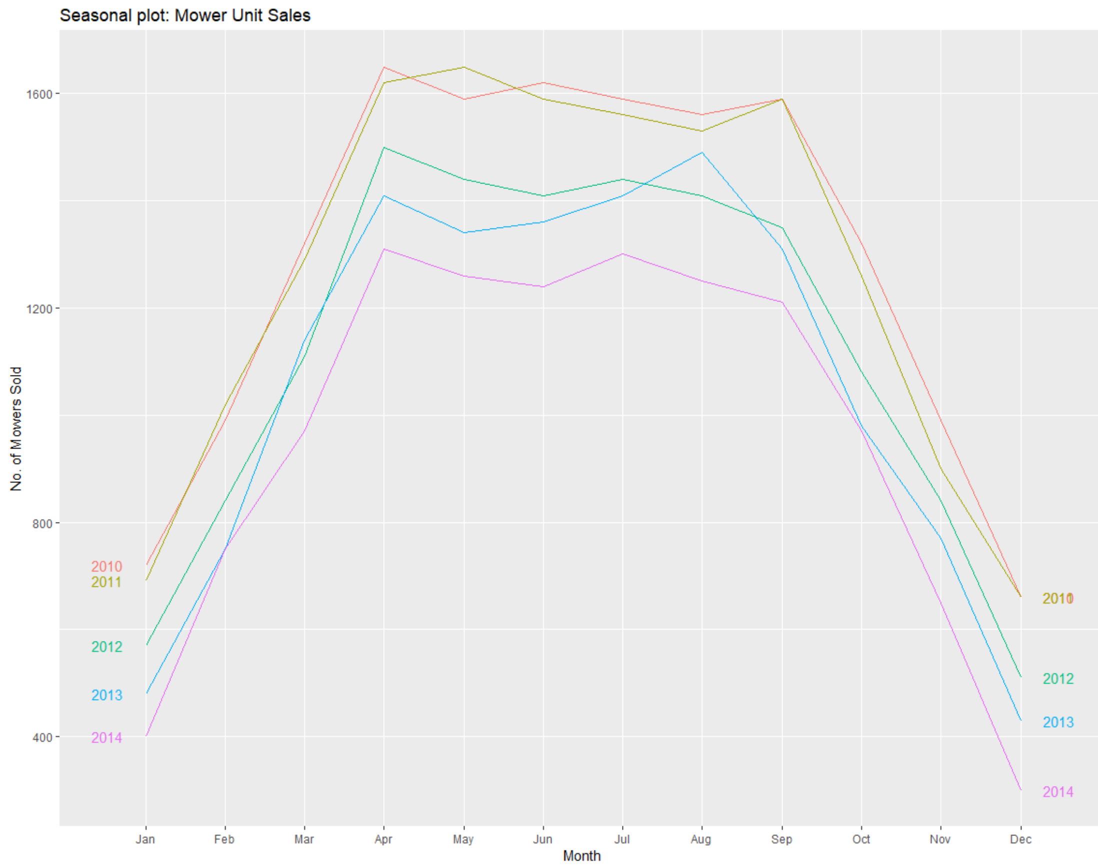
Mower unit sales shows strong seasonality within each year. There are periodic fluctuations every year, where the mower unit sales increases and plateaus till mid-year before decreasing back to around the original volume of mower unit sales that was sold at the start of the year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 41: Decomposed time series data of mower unit sales in Europe from January 2010 to December 2014

Decomposition of the time series further proves the point that there is a steady downward trend and there is the presence of a clear seasonality.

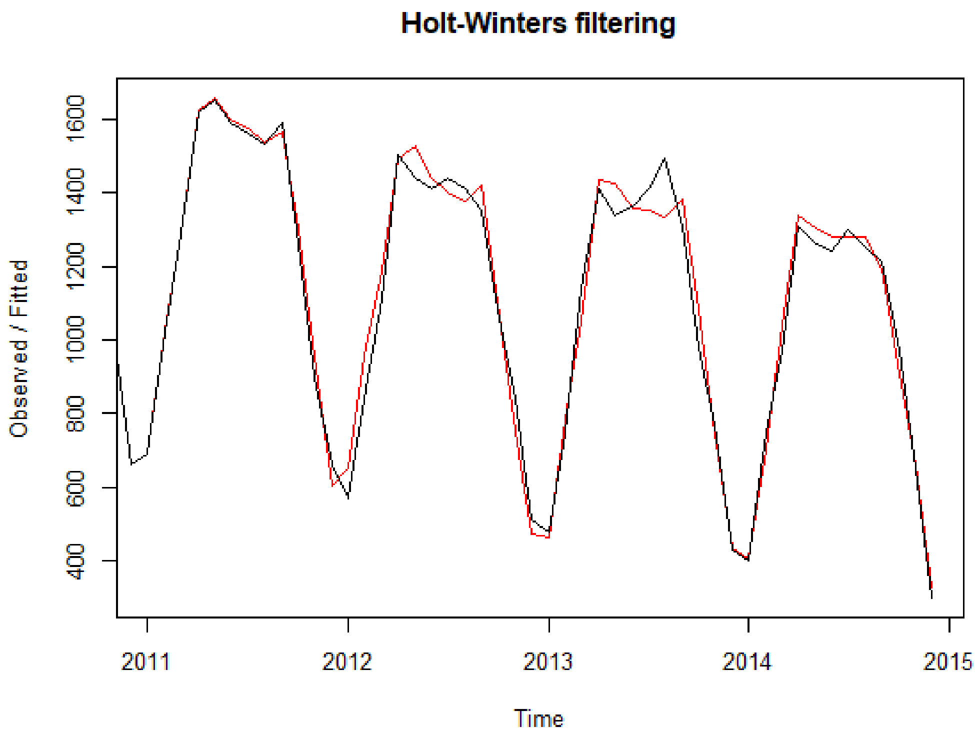


Graph 42: A seasonal plot of mower unit sales in Europe from January 2010 to December 2014.

We can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does not vary with the level of the time series.

Hence, we can conclude that the sales of mowers in Europe possess trend and seasonality.

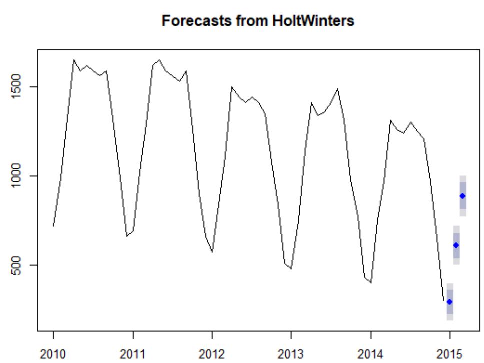
We can use Holt-Winters additive model to forecast future sales.

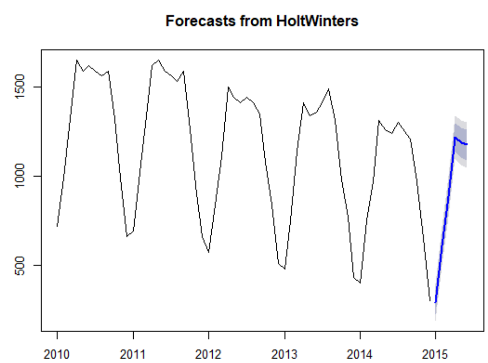


Graph 43: Observed time series data of mower unit sales in Europe against the fitted Holt-Winters additive model

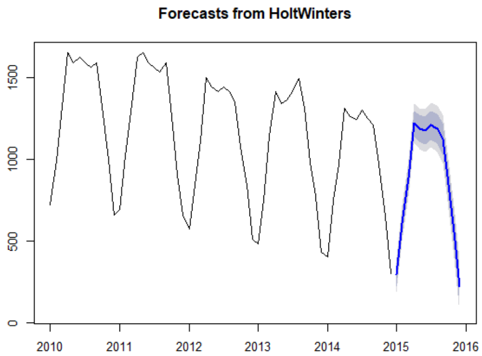
We see from the plot that the Holt-Winters additive model is very successful in predicting the seasonal peaks, which occur from around April to September each year, and the downward trend. The model fits well with the observed time series data.

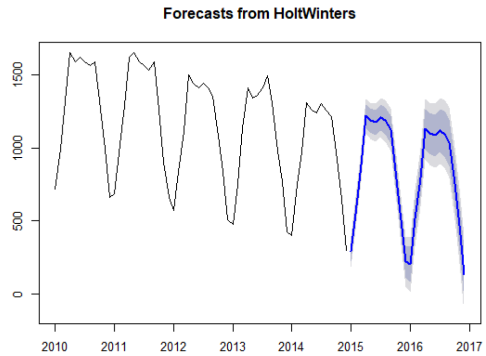
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months





Graphs 44 and 45: Forecasted data of unit mower sales in Europe over the next 3 and 6 months





Graphs 46 and 47: Forecasted data of unit mower sales in Europe over the next 12 and 24 months

The following are the forecasted values of mower unit sales in Europe over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	291.9024	223.25596	360.5488	186.91674	396.8880
Feb 2015	606.7646	535.18605	678.3431	497.29465	716.2345
Mar 2015	886.4436	811.77743	961.1097	772.25156	1000.6356
Apr 2015	1216.8984	1138.99554	1294.8013	1097.75624	1336.0406
May 2015	1186.3354	1105.05283	1267.6180	1062.02444	1310.6464
Jun 2015	1173.9560	1089.15667	1258.7552	1044.26665	1303.6453
Jul 2015	1205.9295	1117.48215	1294.3769	1070.66094	1341.1981
Aug 2015	1184.5869	1092.36543	1276.8084	1043.54633	1325.6275
Sep 2015	1122.9170	1026.80034	1219.0336	975.91928	1269.9146
Oct 2015	846.8561	746.72795	946.9842	693.72333	999.9888
Nov 2015	557.2257	452.97412	661.4774	397.78665	716.6648
Dec 2015	221.8262	113.34313	330.3093	55.91567	387.7367
Jan 2016	202.9574	81.74397	324.1708	17.57746	388.3374
Feb 2016	517.8196	392.46661	643.1726	326.10873	709.5305
Mar 2016	797.4986	667.89479	927.1024	599.28668	995.7105
Apr 2016	1127.9534	993.99112	1261.9158	923.07576	1332.8311
May 2016	1097.3904	958.96525	1235.8156	885.68738	1309.0935
Jun 2016	1085.0110	942.02173	1228.0002	866.32779	1303.6942
Jul 2016	1116.9845	969.33298	1264.6361	891.17098	1342.7981
Aug 2016	1095.6419	943.23267	1248.0512	862.55208	1328.7318
Sep 2016	1033.9720	876.71224	1191.2317	793.46396	1274.4800
Oct 2016	757.9111	595.71059	920.1116	509.84682	1005.9754
Nov 2016	468.2808	301.05155	635.5100	212.52575	724.0358
Dec 2016	132.8812	-39.46243	305.2249	-130.69565	396.4581

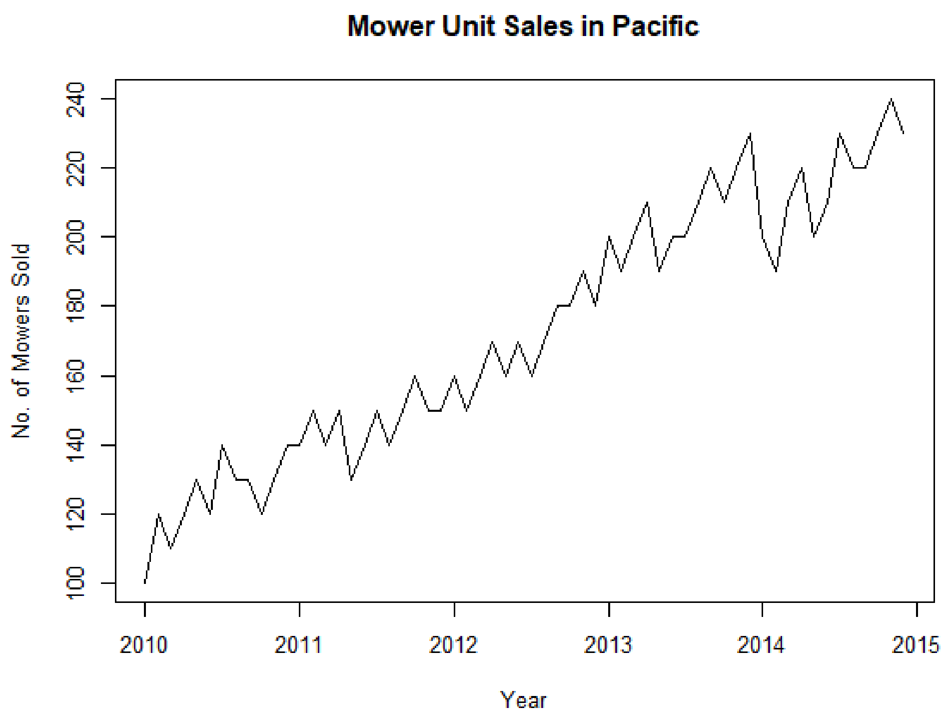
<u>Conclusion:</u>

As seen from the forecasted graphs and values generated, unit mower sales in Europe is predicted to decrease for the next 24 months to come. Seasonality will remain.

Forecasting of Unit Sales of Mowers in Pacific

Time Series:

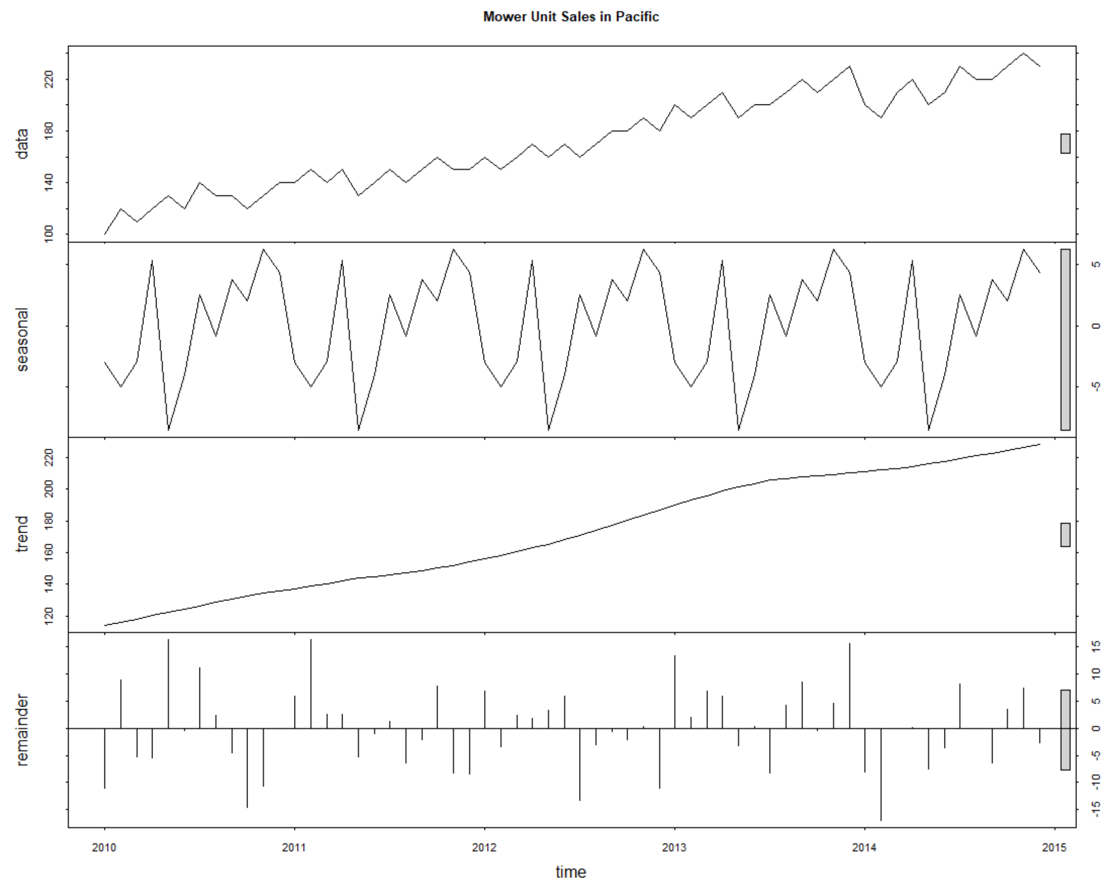
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	100	120	110	120	130	120	140	130	130	120	130	140
2011	140	150	140	150	130	140	150	140	150	160	150	150
2012	160	150	160	170	160	170	160	170	180	180	190	180
2013	200	190	200	210	190	200	200	210	220	210	220	230
2014	200	190	210	220	200	210	230	220	220	230	240	230



Graph 48: The time series graph above shows the mower unit sales in Pacific from January 2010 to December 2014.

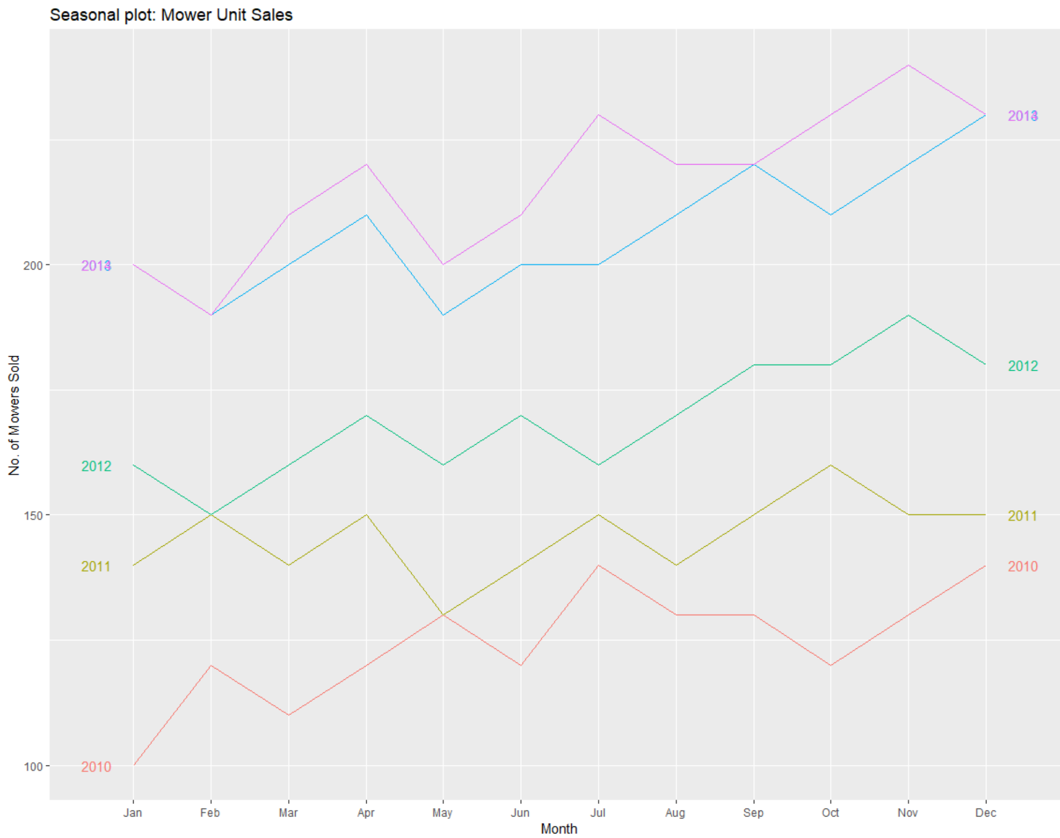
From the graph, there is an upward trend in the data over this period. Mower unit sales shows no seasonality within each year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 49: Decomposed time series data of mower unit sales in Pacific from January 2010 to December 2014

Decomposition of the time series further proves the point that there is a steady upward trend. But we are not certain about the presence of seasonality yet.

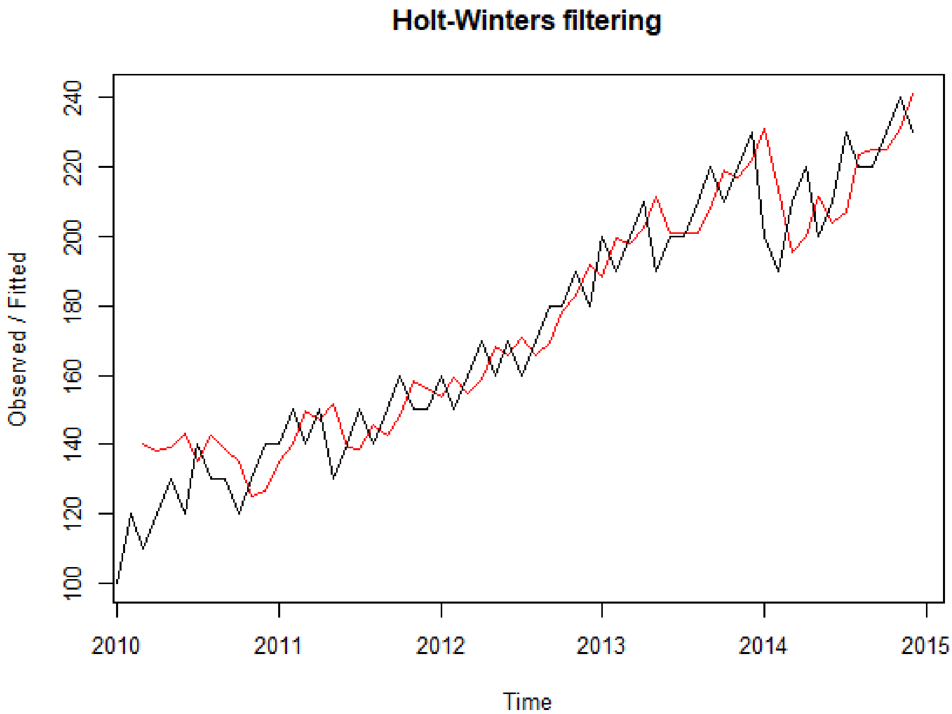


Graph 50: A seasonal plot of mower unit sales in Pacific fom January 2010 to December 2014.

As seen in the seasonal plot, it further proves that there is no seasonality in Pacific due to different peaks in different times of the year.

Hence, we can conclude that the sales of mowers in Pacific possess <u>trend but no seasonality.</u>

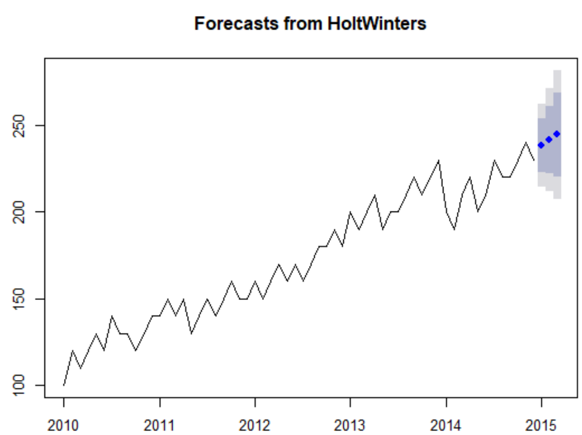
We can use <u>Double Exponential Smoothing</u> to forecast future sales.

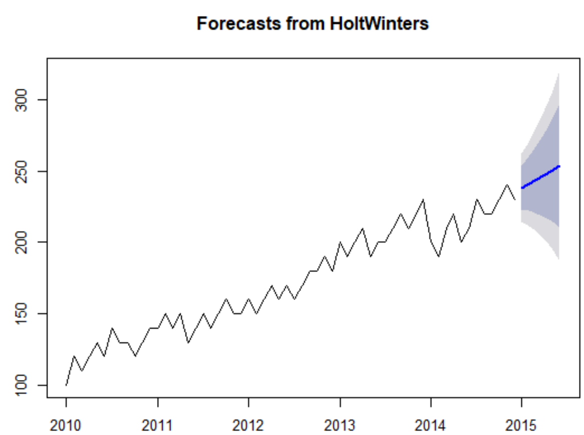


Graph 51: Observed time series data of mower unit sales in Pacific against the fitted double exponential smoothing model

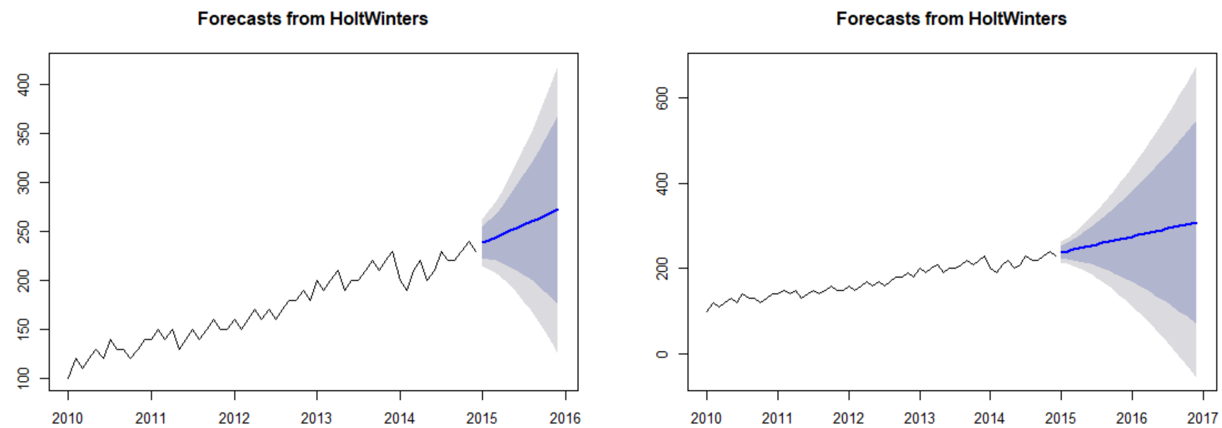
We see from the plot that the model is very successful in predicting the upward trend. The model fits well with the observed time series data.

Hence, we can make use of it to predict data in the next 3,6,12 and 24 months





Graphs 52 and 53: Forecasted data of unit mower sales in Pacific over the next 3 and 6 months respectively.



Graphs 54 and 55: Forecasted data of unit mower sales in Pacific over the next 12 and 24 months respectively.

The following are the forecasted values of mower unit sales in Pacific over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	238.3372	222.59607	254.0784	214.263217	262.4112
Feb 2015	241.4217	221.97231	260.8711	211.676417	271.1670
Mar 2015	244.5062	220.23695	268.7755	207.389583	281.6228
Apr 2015	247.5907	217.60629	277.5751	201.733501	293.4479
May 2015	250.6752	214.23878	287.1116	194.950502	306.3999
Jun 2015	253.7597	210.24224	297.2771	187.205490	320.3139
Jul 2015	256.8442	205.69087	307.9975	178.611936	335.0764
Aug 2015	259.9287	200.63769	319.2197	169.250938	350.6064
Sep 2015	263.0132	195.12227	330.9041	159.182988	366.8434
Oct 2015	266.0977	189.17528	343.0201	148.455029	383.7403
Nov 2015	269.1822	182.82138	355.5430	137.104747	401.2596
Dec 2015	272.2667	176.08092	368.4524	125.163263	419.3701
Jan 2016	275.3512	168.97109	381.7312	112.656876	438.0454
Feb 2016	278.4357	161.50669	395.3646	99.608240	457.2631
Mar 2016	281.5201	153.70070	409.3396	86.037169	477.0031
Apr 2016	284.6046	145.56458	423.6447	71.961220	497.2481
May 2016	287.6891	137.10862	438.2697	57.396115	517.9822
Jun 2016	290.7736	128.34211	453.2052	42.356061	539.1912
Jul 2016	293.8581	119.27351	468.4427	26.853994	560.8623
Aug 2016	296.9426	109.91056	483.9747	10.901768	582.9835
Sep 2016	300.0271	100.26041	499.7938	-5.489693	605.5439
Oct 2016	303.1116	90.32968	515.8935	-22.310271	628.5335
Nov 2016	306.1961	80.12452	532.2677	-39.550554	651.9428
Dec 2016	309.2806	69.65067	548.9105	-57.201755	675.7630

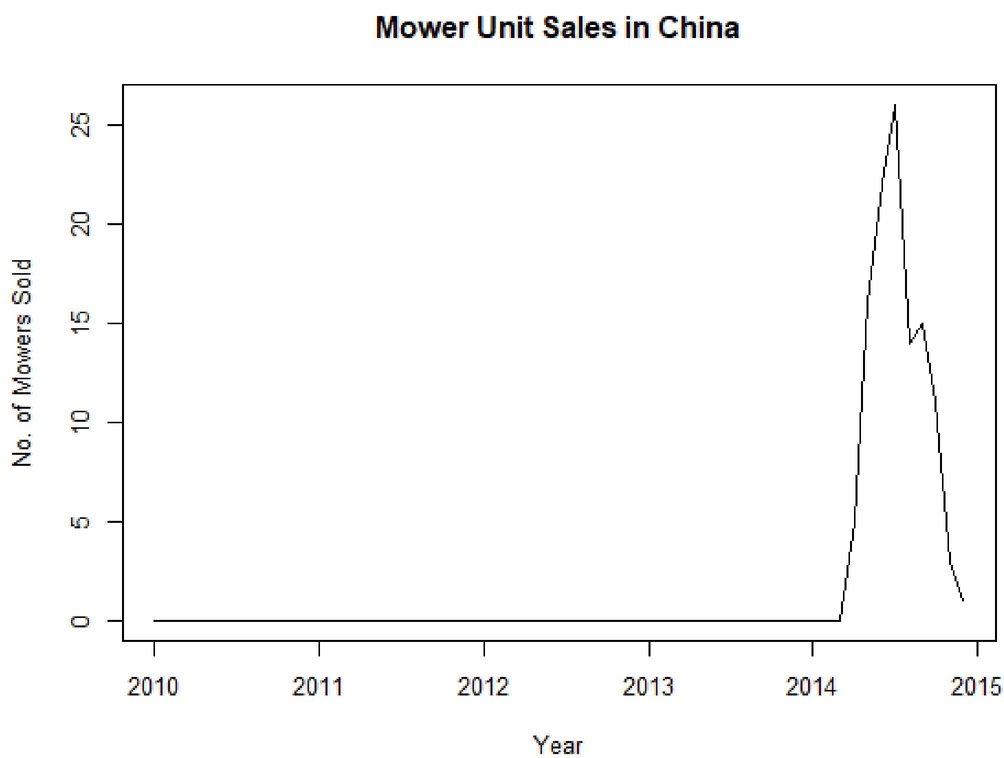
<u>Conclusion:</u>

As seen from the forecasted graphs and values generated, unit mower sales in Pacific is predicted to increase gradually for the next 24 months to come.

Forecasting of Sales of Unit Mowers in China

Time Series:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	0	0	0	0	0	0	0	0	0	0	0	0
2011	0	0	0	0	0	0	0	0	0	0	0	0
2012	0	0	0	0	0	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0	0	0	0	0	0
2014	0	0	0	5	16	22	26	14	15	11	3	1



Graph 56: The time series graph above shows the mower unit sales in China from January 2010 to December 2014.

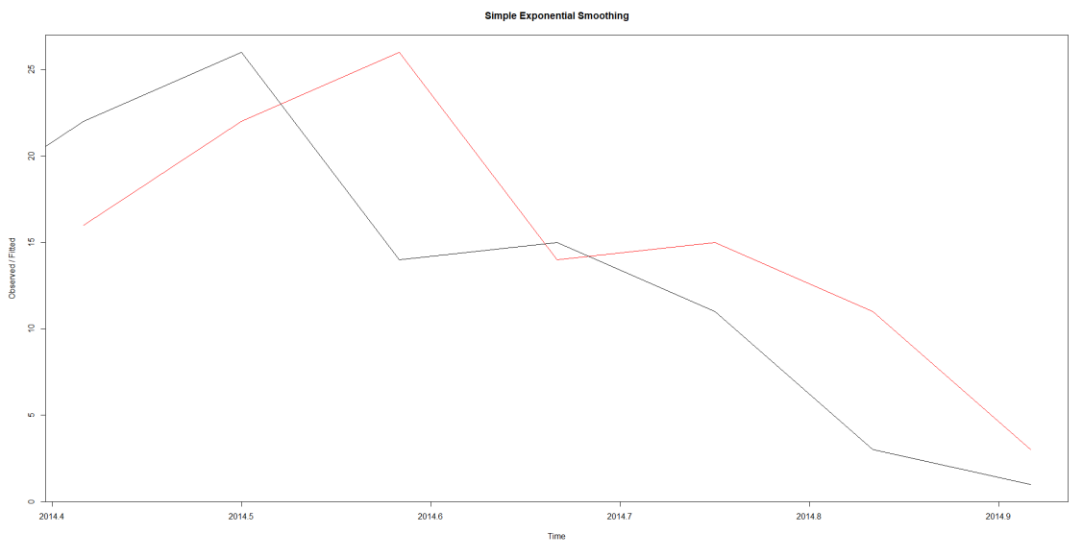
According to the case study given, sales of mowers in China only started much later. Hence, we can subset the data to after May 2014.

Time series:

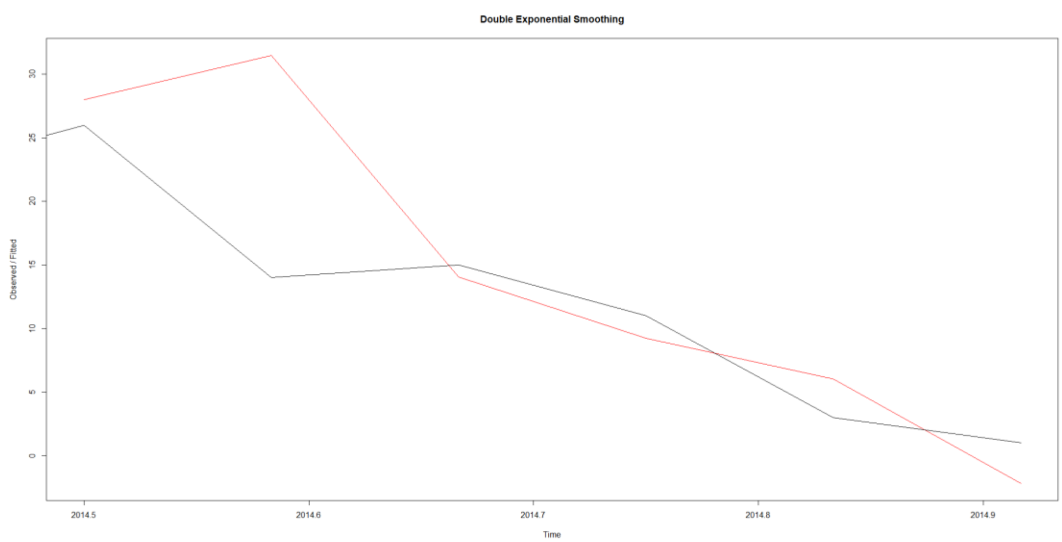
	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2014	16	22	26	14	15	11	3	1

Due to the lack of sufficient data, it is impossible to deduce whether there is a trend and/or seasonality in the data.

Thus, we can must try out different forecast methods to deduce which is the most appropriate.



Graph 57: Observed time series data of mower unit sales in China against the fitted single exponential smoothing model



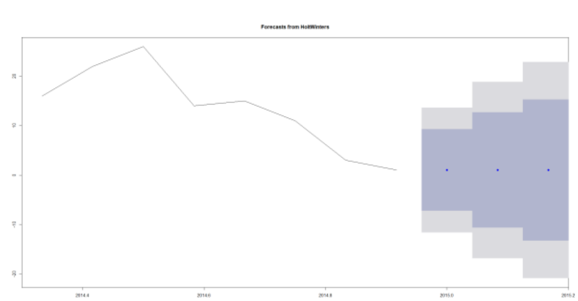
Graph 58: Observed time series data of mower unit sales in China against the fitted double exponential smoothing model

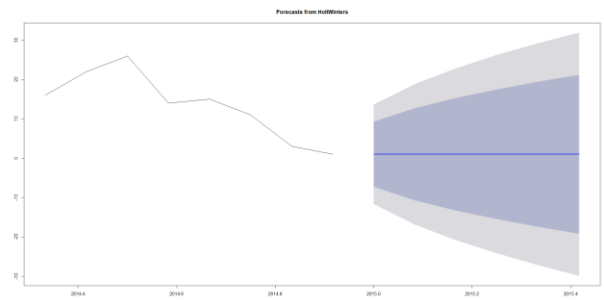
From the graphs plotted, we are unable to tell which graph would be a better fit. Hence, we can calculate the error metrics (using the accuracy function) and see which model gives us an overall lower value of error metrics. The error metrics that we have chosen for comparison are the root-mean-square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

	RMSE	MAE	MAPE
Simple Exponential	6.33	5.28	91.1
Double Exponential	7.44	4.72	95.4

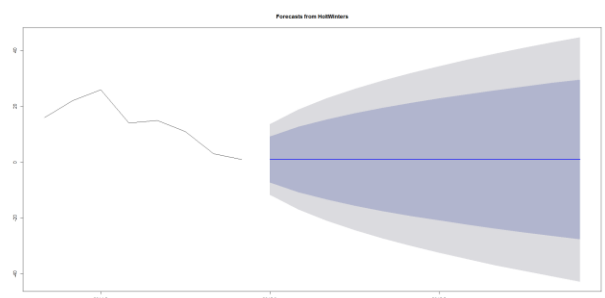
From the table, we can see that using a simple exponential smoothing model will be appropriate because it has lower values of error. Holt-Winters multiplicative/additive model and Holt-Winters no trend smoothing model were not being able to be used due to the lack of 2 full periodic cycles in the time series.

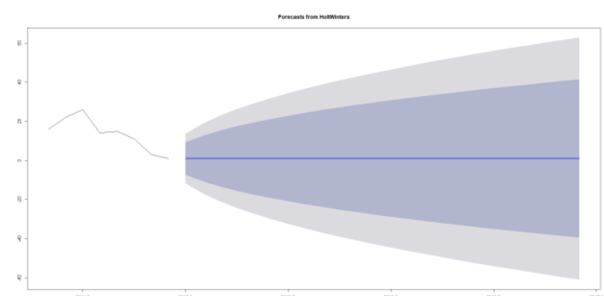
We can make use of Simple Exponential Smoothing to predict data in the next 3,6,12 and 24 months





Graphs 59 and 60: Forecasted data of unit mower sales in China over the next 3 and 6 months





Graphs 61 and 62: Forecasted data of unit mower sales in China over the next 3 and 6 months

The following are the forecasted values of mower unit sales in China over the next 24 months:

```
         Point Forecast      Lo 80     Hi 80     Lo 95    Hi 95
Jan 2015       1.000137  -7.253268  9.253542 -11.62236 13.62263
Feb 2015       1.000137 -10.671540 12.671814 -16.85015 18.85043
Mar 2015       1.000137 -13.294526 15.294800 -20.86166 22.86194
Apr 2015       1.000137 -15.505824 17.506098 -24.24355 26.24383
May 2015       1.000137 -17.454025 19.454299 -27.22307 29.22334
Jun 2015       1.000137 -19.215338 21.215612 -29.91676 31.91704
```

```
Jul 2015       1.000137 -20.835036 22.835310 -32.39388 34.39415
Aug 2015       1.000137 -22.342616 24.342891 -34.69952 36.69980
Sep 2015       1.000137 -23.758568 25.758843 -36.86504 38.86531
Oct 2015       1.000137 -25.097810 27.098084 -38.91323 40.91350
Nov 2015       1.000137 -26.371604 28.371878 -40.86133 42.86160
Dec 2015       1.000137 -27.588699 29.588973 -42.72271 44.72299
Jan 2016       1.000137 -28.756054 30.756328 -44.50803 46.50830
Feb 2016       1.000137 -29.879310 31.879584 -46.22590 48.22618
Mar 2016       1.000137 -30.963117 32.963391 -47.88344 49.88372
Apr 2016       1.000137 -32.011360 34.011634 -49.48659 51.48687
May 2016       1.000137 -33.027327 35.027601 -51.04038 53.04065
Jun 2016       1.000137 -34.013827 36.014101 -52.54910 54.54937
Jul 2016       1.000137 -34.973284 36.973558 -54.01646 56.01674
Aug 2016       1.000137 -35.907807 37.908081 -55.44569 57.44597
Sep 2016       1.000137 -36.819246 38.819520 -56.83962 58.83989
Oct 2016       1.000137 -37.709229 39.709504 -58.20073 60.20100
Nov 2016       1.000137 -38.579206 40.579480 -59.53124 61.53152
Dec 2016       1.000137 -39.430467 41.430741 -60.83314 62.83341
```

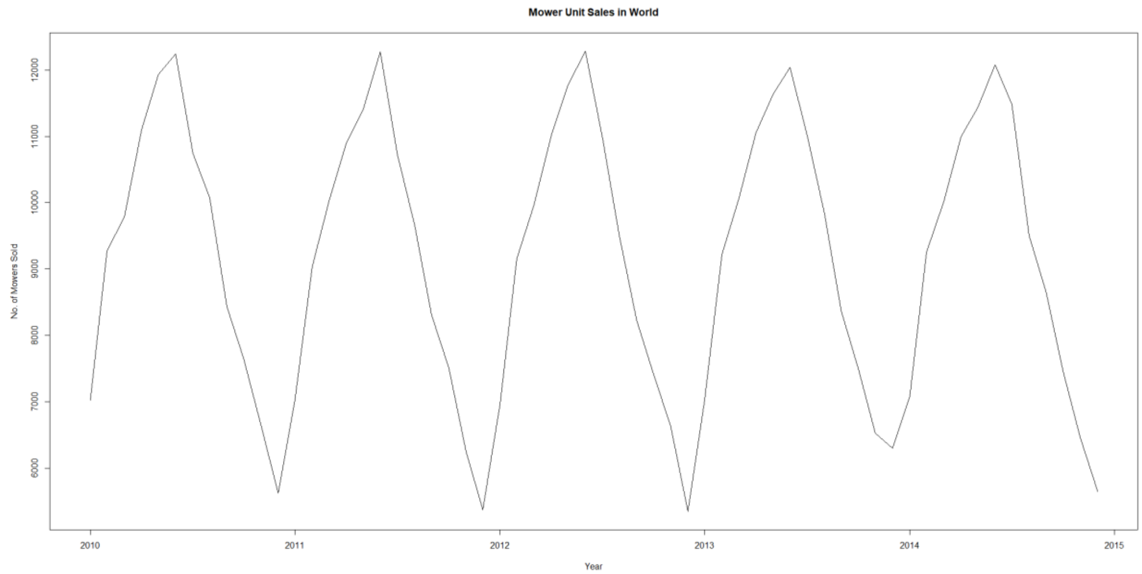
Conclusion:

As seen from the forecasted graphs and values generated, unit mower sales in China is predicted to decrease to zero and remain stagnant at that value for the next 24 months to come.

Forecasting of Unit Sales of Mowers in World

Time Series:

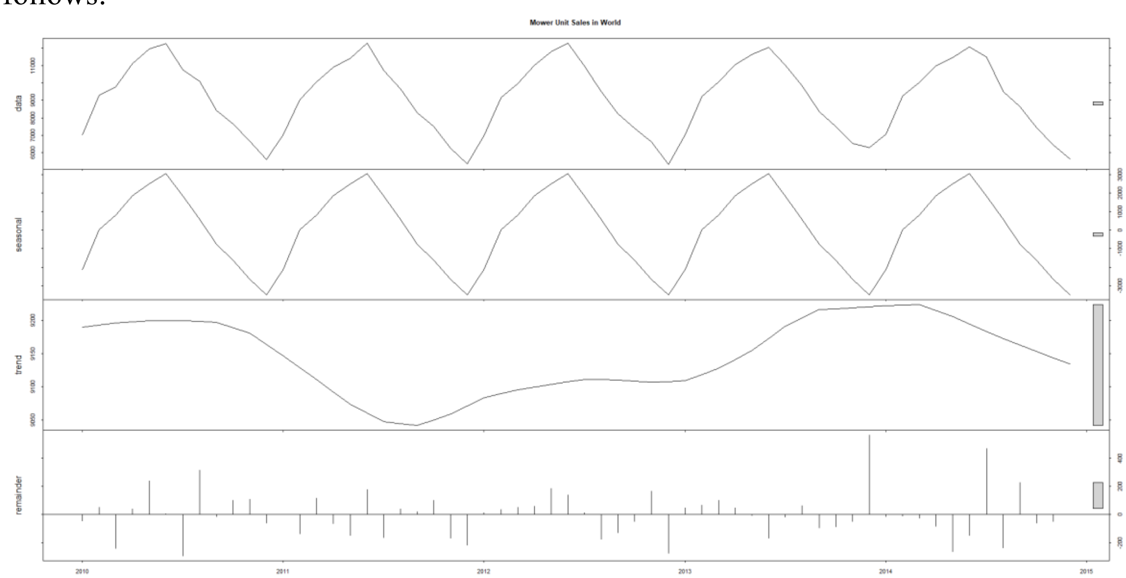
```
       Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
2010  7020  9280  9780 11100 11930 12240 10740 10080  8430  7650  6650  5620
2011  7020  9030 10050 10890 11420 12270 10720  9650  8310  7510  6250  5370
2012  6970  9160  9970 11020 11780 12280 10960  9500  8230  7420  6630  5350
2013  7030  9220 10050 11050 11640 12040 11010  9830  8370  7490  6530  6300
2014  7080  9250 10020 10995 11436 12082 11486  9504  8635  7451  6453  5651
```



Graph 63: The time series graph above shows the mower unit sales in the World from January 2010 to December 2014.

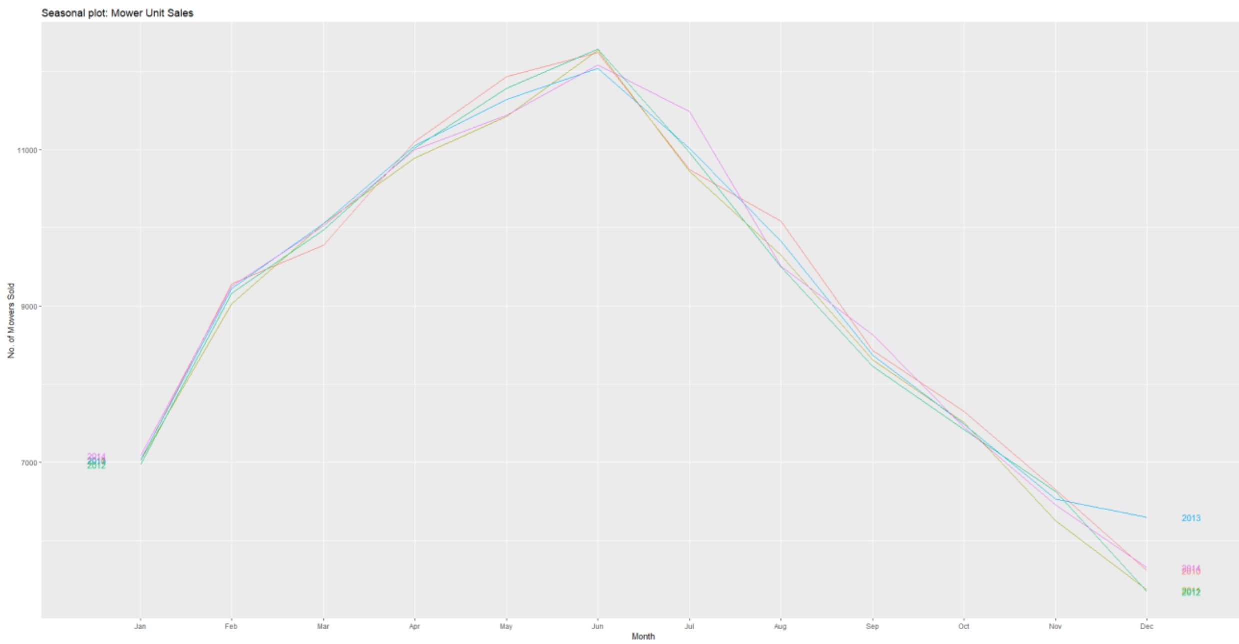
From the graph, there is no apparent trend in the data over this period.
However, mower unit sales shows strong seasonality each year. There are periodic fluctuations every year, where the mower unit sales increases till mid-year before decreasing back to around the original volume of mower unit sales that was sold at the start of the year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 64: Decomposed time series data of mower unit sales in the World from January 2010 to December 2014

Decomposition of the time series further proves the point that there is no apparent trend but there is the presence of a clear seasonality.

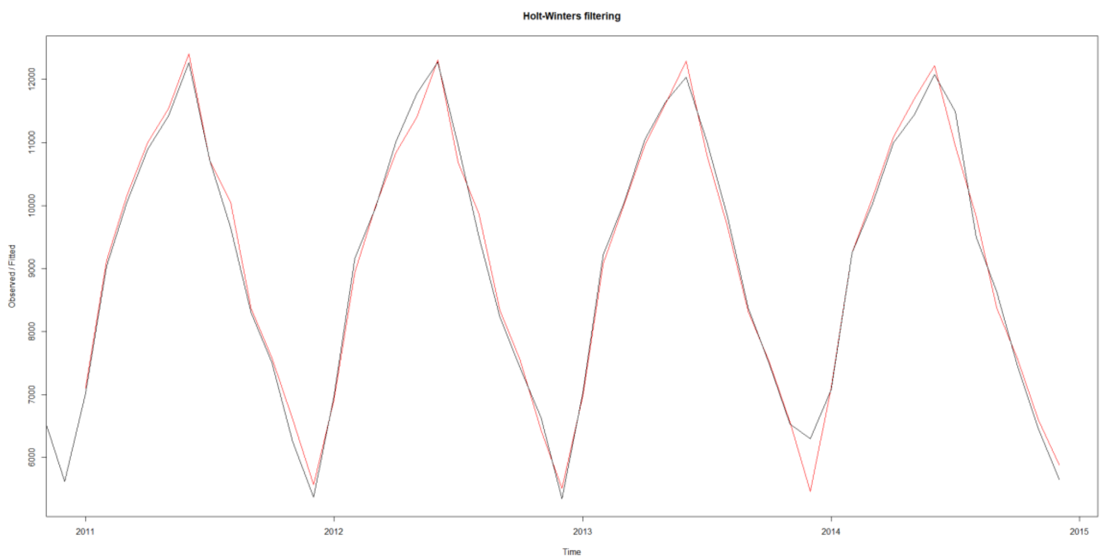


Graph 65: A seasonal plot of mower unit sales in the World from January 2010 to December 2014.

From the seasonal plot, it is clear that there is an increase of mower unit sales till June before it decreases all the way till the end of December. Plus, we can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does not vary with the level of the time series

Hence, we can conclude that the sales of mowers in the World possess no trend but has seasonality.

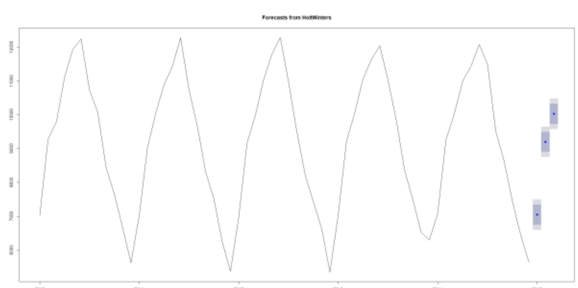
We can use the Holt-Winters no-trend smoothing to forecast future sales.

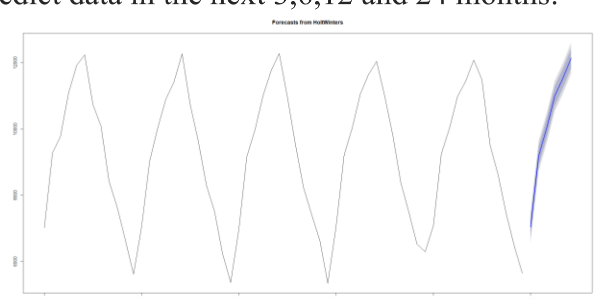


Graph 66: Observed time series data of mower unit sales in the World against the fitted Holt-Winters no-trend smoothing model

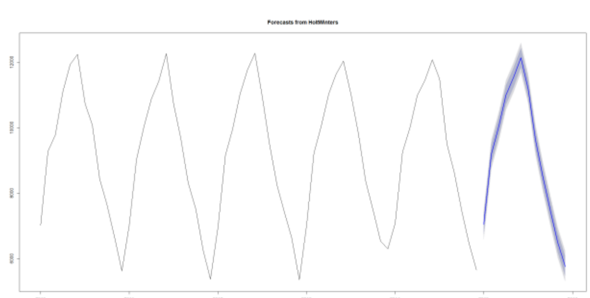
We see from the plot that the Holt-Winters no-trend smoothing model is very successful in predicting the seasonal peaks, which occur every June/July of the year. The model fits well with the observed time series data.

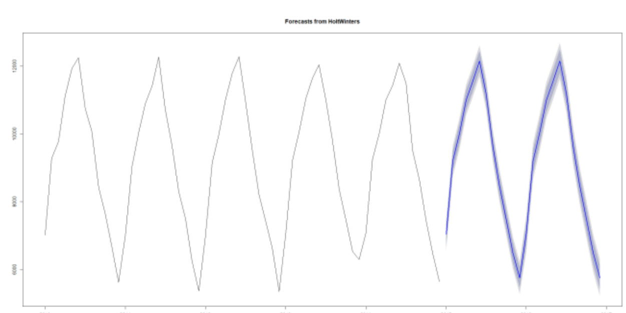
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months:





Graphs 67 and 68: Forecasted data of unit mower sales in the World over the next 3 and 6 months respectively.





Graphs 69 and 70: Forecasted data of unit mower sales in the World over the next 12 and 24 months respectively.

The following are the forecasted values of mower unit sales in the World:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	7043.910	6750.382	7337.437	6594.999	7492.821
Feb 2015	9197.401	8902.741	9492.061	8746.758	9648.045
Mar 2015	10025.170	9729.381	10320.958	9572.800	10477.539
Apr 2015	11000.066	10703.153	11296.978	10545.977	11454.154
May 2015	11537.783	11239.751	11835.816	11081.982	11993.584
Jun 2015	12144.183	11845.035	12443.331	11686.676	12601.691
Jul 2015	11188.010	10887.751	11488.270	10728.803	11647.218
Aug 2015	9630.657	9329.290	9932.024	9169.756	10091.559
Sep 2015	8473.574	8171.103	8776.045	8010.985	8936.163
Oct 2015	7469.531	7165.961	7773.101	7005.261	7933.802
Nov 2015	6499.492	6194.827	6804.158	6033.546	6965.438
Dec 2015	5757.219	5451.462	6062.976	5289.604	6224.834
Jan 2016	7043.910	6699.160	7388.659	6516.661	7571.159
Feb 2016	9197.401	8851.686	9543.116	8668.676	9726.126
Mar 2016	10025.170	9678.493	10371.846	9494.973	10555.366
Apr 2016	11000.066	10652.429	11347.702	10468.401	11531.730
May 2016	11537.783	11189.190	11886.377	11004.655	12070.911
Jun 2016	12144.183	11794.635	12493.731	11609.596	12678.771
Jul 2016	11188.010	10837.511	11538.510	10651.967	11724.053
Aug 2016	9630.657	9279.208	9982.106	9093.162	10168.152
Sep 2016	8473.574	8121.178	8825.969	7934.632	9012.516
Oct 2016	7469.531	7116.192	7822.871	6929.145	8009.918
Nov 2016	6499.492	6145.211	6853.773	5957.666	7041.318
Dec 2016	5757.219	5401.999	6112.439	5213.956	6300.481

<u>Conclusion:</u>

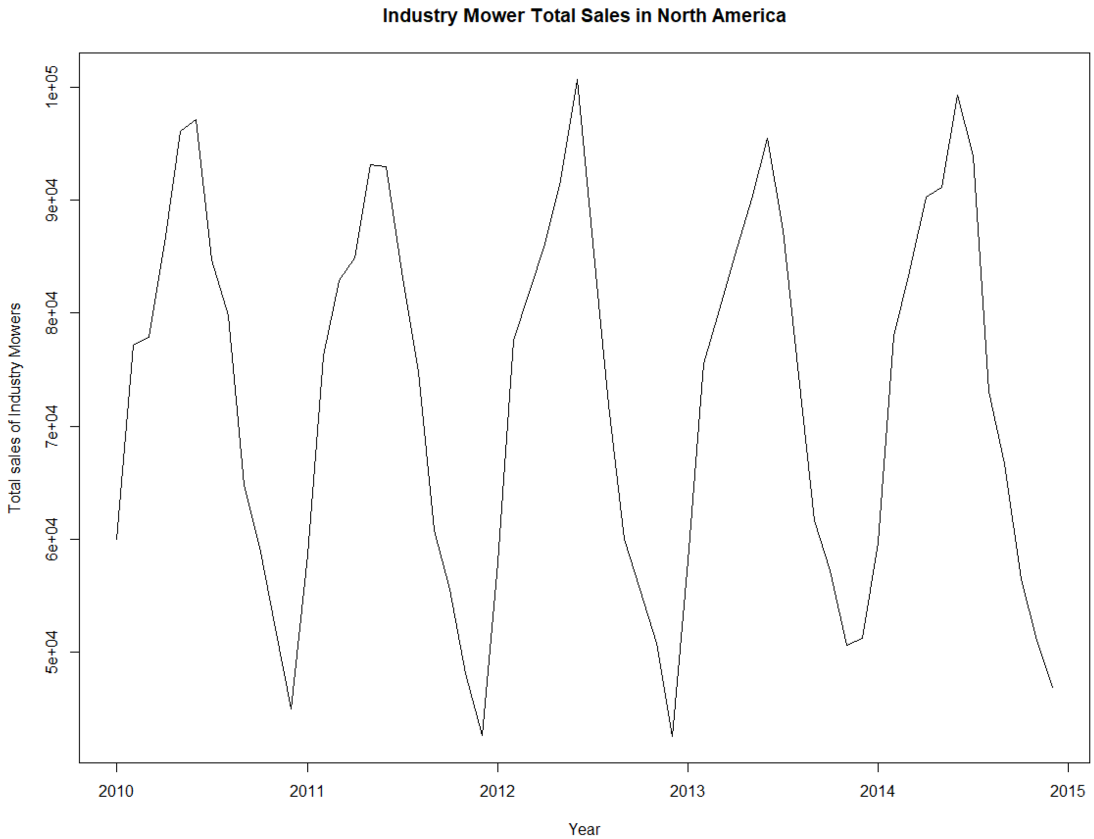
As seen from the forecasted graphs and values generated, unit mower sales in the World is predicted to remain the same for the next 24 months to come. Seasonality will remain.

Next, we will forecast the total industry sales of mowers in each marketing region using the excel sheet Industry Mower Total Sales.

<u>Forecasting of Industry Sales of Mowers in North America</u>

<u>Time Series:</u>

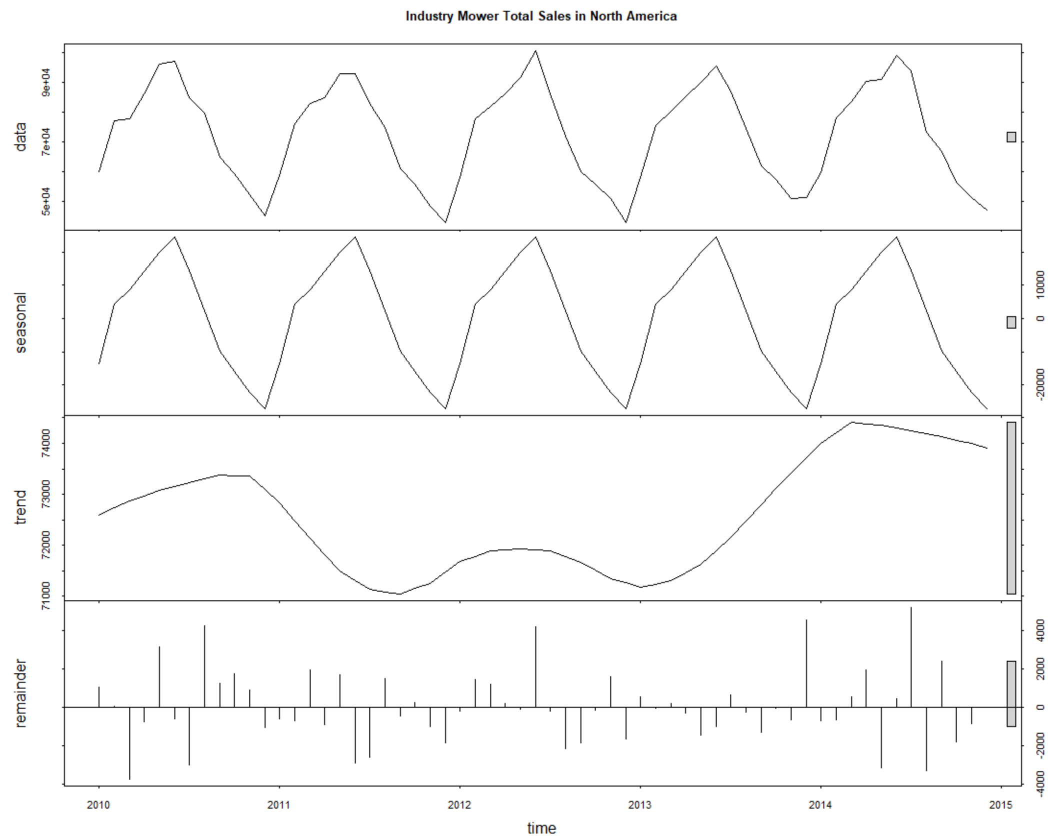
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	60000.00	77184.47	77884.62	86190.48	96116.50	97142.86	84757.28	79803.92	64800.00	59306.93	52156.86	45048.54
2011	58627.45	76200.00	82871.29	84903.85	93100.00	93000.00	83047.62	74854.37	60769.23	55619.05	48155.34	42647.06
2012	57884.62	77647.06	81844.66	86095.24	91775.70	100679.61	86190.48	71886.79	60000.00	55566.04	50857.14	42596.15
2013	58095.24	75566.04	80285.71	85140.19	90092.59	95471.70	87307.69	74476.19	61698.11	57238.10	50673.08	51238.10
2014	59711.54	77961.17	83725.49	90297.03	91142.86	99320.39	93921.57	73142.86	66699.03	56476.19	51067.96	46893.20



Graph 71: The time series graph above shows the mower industry sales in North America from January 2010 to December 2014.

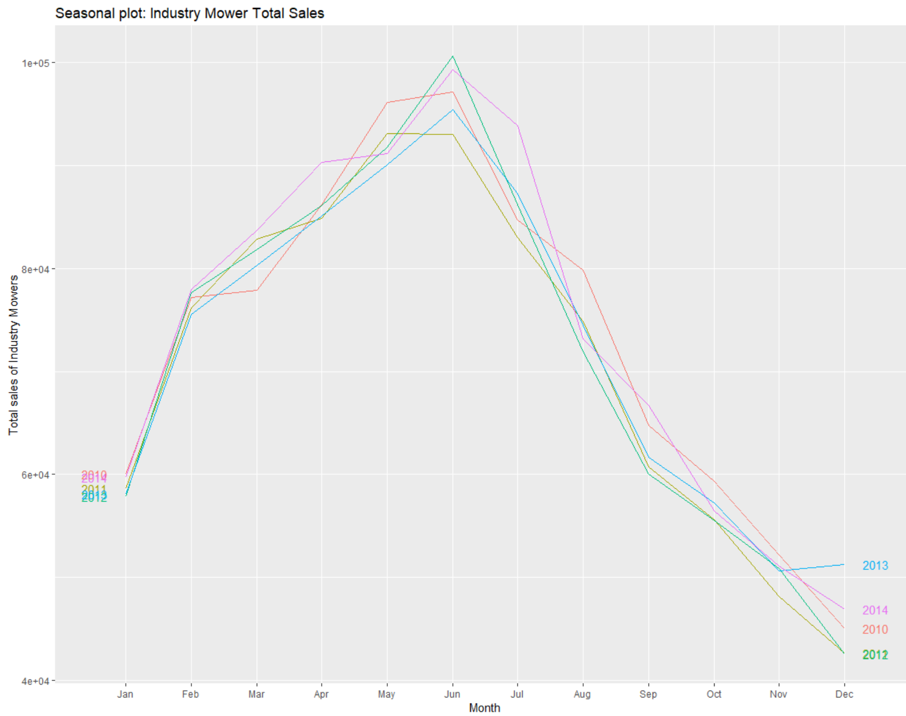
From the graph, there is no apparent trend in the data over this period.
However, mower industry sales shows strong seasonality each year. There are periodic fluctuations every year, where the mower industry sales increases till mid-year before decreasing back to around the original volume of mower industry sales that was sold at the start of the year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 72: Decomposed time series data of mower industry sales in North America from January 2010 to December 2014

Decomposition of the time series further proves the point that there is no apparent trend but there is the presence of a clear seasonality.

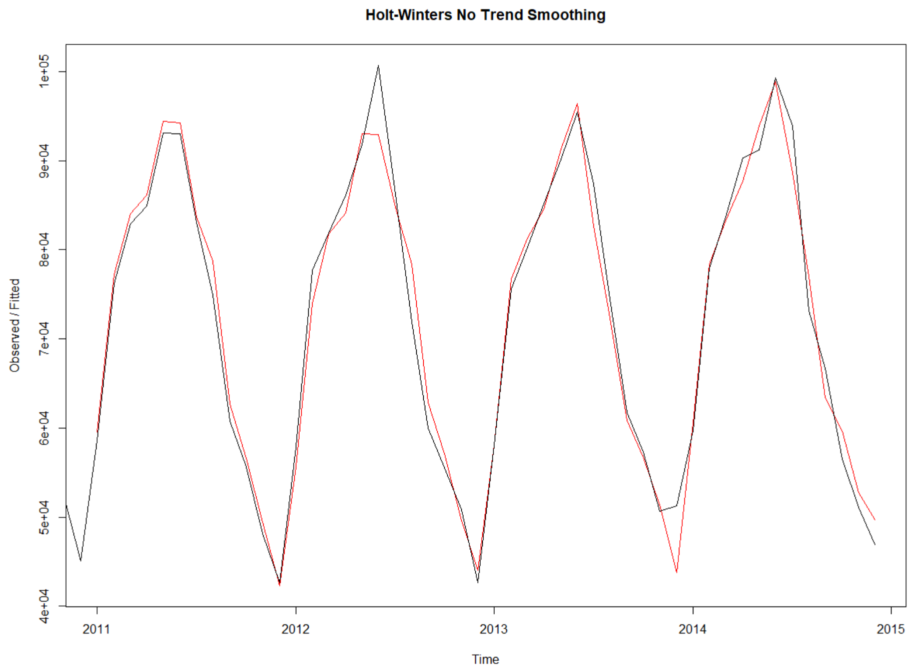


Graph 73: A seasonal plot of mower industry sales in North America from January 2010 to December 2014.

From the seasonal plot, it is clear that there is an increase of mower industry sales till June before it decreases all the way till the end of December. Plus, we can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does not vary with the level of the time series

Hence, we can conclude that the industry sales of mowers in North America possess no trend but has seasonality.

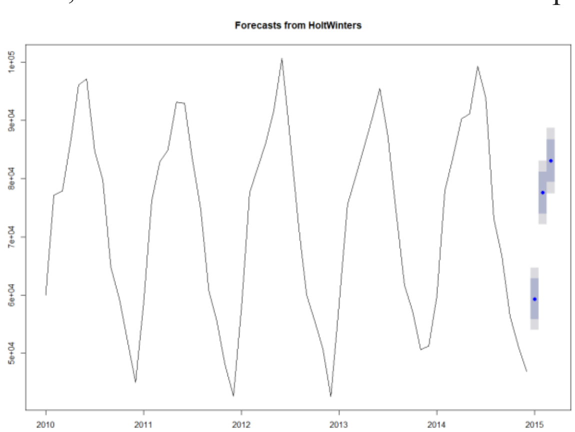
We can use the Holt-Winters no-trend smoothing to forecast future sales.

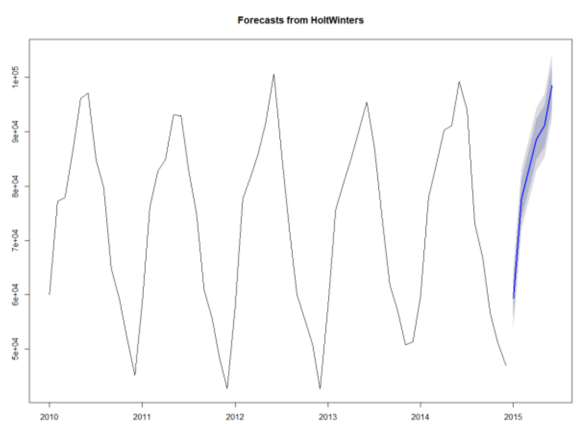


Graph 74: Observed time series data of mower industry sales in North America against the fitted Holt-Winters no-trend smoothing model

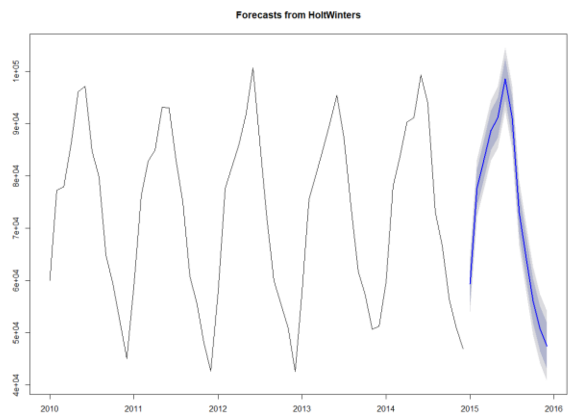
We see from the plot that the Holt-Winters no-trend smoothing model is very successful in predicting the seasonal peaks, which occur every June/July of the year. The model fits well with the observed time series data.

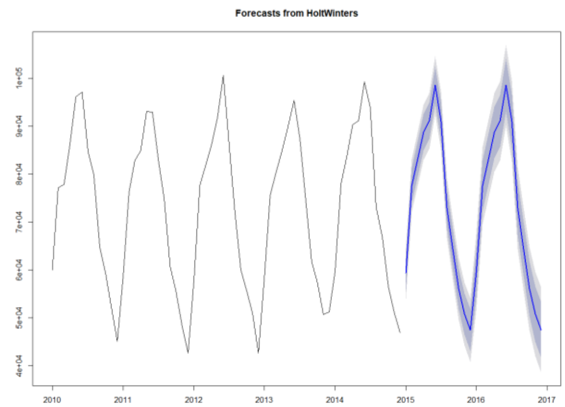
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months:





Graphs 75 and 76: Forecasted data of industry mower sales in North America over the next 3 and 6 months respectively.





Graphs 77 and 78: Forecasted data of industry mower sales in North America over the next 12 and 24 months respectively.

The following are the forecasted values of mower industry sales in North America:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	59348.27	55850.47	62846.08	53998.84	64697.70
Feb 2015	77598.03	74013.64	81182.42	72116.18	83079.88
Mar 2015	83071.87	79402.95	86740.80	77460.73	88683.01
Apr 2015	88674.75	84923.19	92426.31	82937.23	94412.27
May 2015	91144.30	87311.89	94976.72	85283.13	97005.48
Jun 2015	98586.66	94675.06	102498.26	92604.39	104568.93
Jul 2015	91080.79	87091.58	95070.00	84979.82	97181.76
Aug 2015	72917.83	68852.49	76983.16	66700.43	79135.22
Sep 2015	64368.72	60228.66	68508.79	58037.04	70700.41
Oct 2015	56088.57	51875.10	60302.04	49644.62	62532.52
Nov 2015	50763.74	46478.12	55049.36	44209.45	57318.03
Dec 2015	47427.45	43070.87	51784.02	40764.64	54090.25
Jan 2016	59348.27	54155.56	64540.98	51406.71	67289.83
Feb 2016	77598.03	72346.61	82849.45	69566.68	85629.39
Mar 2016	83071.87	77762.39	88381.36	74951.72	91192.03
Apr 2016	88674.75	83307.83	94041.67	80466.76	96882.74
May 2016	91144.30	85720.56	96568.05	82849.40	99439.20
Jun 2016	98586.66	93106.68	104066.64	90205.75	106967.56
Jul 2016	91080.79	85545.15	96616.44	82614.75	99546.83
Aug 2016	72917.83	67327.07	78508.58	64367.50	81468.15
Sep 2016	64368.72	58723.39	70014.06	55734.94	73002.51
Oct 2016	56088.57	50389.19	61787.95	47372.12	64805.02
Nov 2016	50763.74	45010.82	56516.67	41965.40	59562.08
Dec 2016	47427.45	41621.47	53233.42	38547.98	56306.92

Conclusion:

As seen from the forecasted graphs and values generated, industry mower sales in North America is predicted to remain the same for the next 24 months to come. Seasonality will remain.

Future Predictions of Market Share in North America:

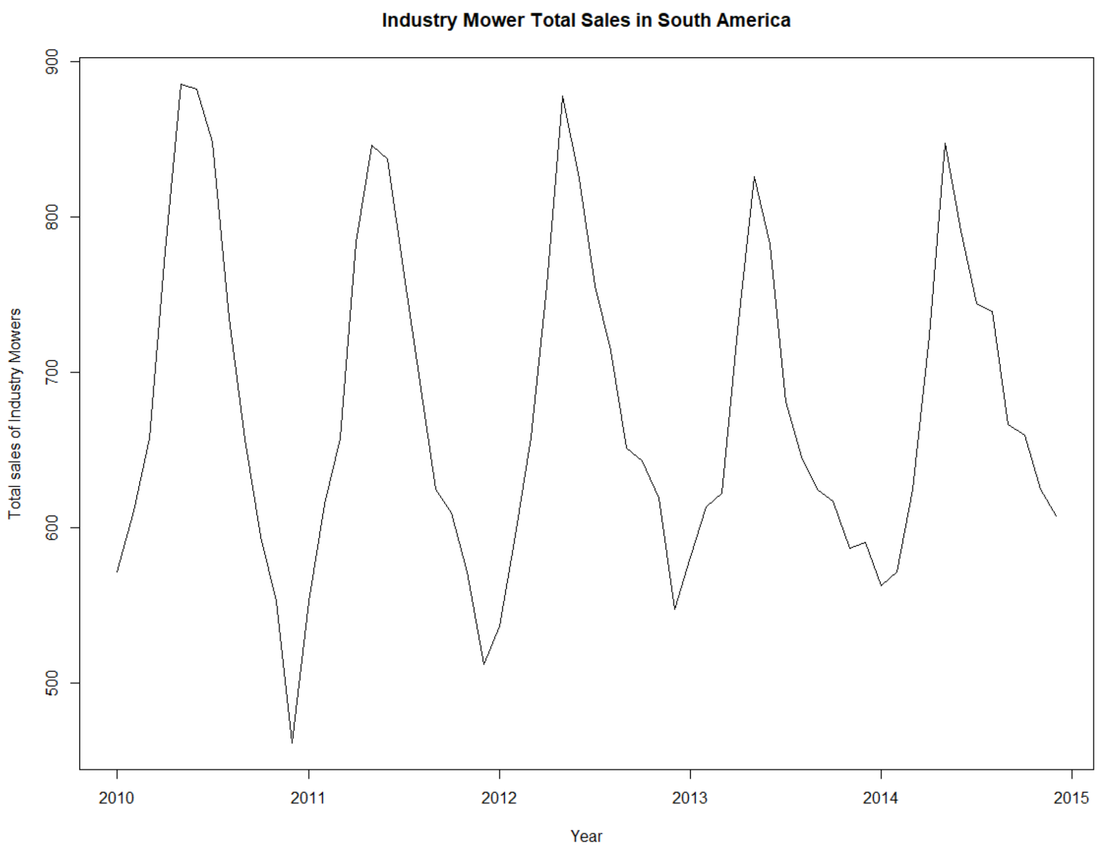
Both forecasted mower sales (Industry and Unit) in North America had similar values. Both did not see any increase/decrease in sales in the next 24 months. Meanwhile, seasonality remained where where the mower sales increases till mid-year before decreasing back to around the original volume of mower sales that was sold at the start of the year.

Since *market share = unit sales/industry sales*, there will be <u>no predicted change</u> in market shares in North America for the sales of mowers.

Forecasting of Industry Sales of Mowers in South America

<u>Time Series:</u>

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	571.4286	611.1111	657.8947	777.7778	885.7143	882.3529	848.4848	735.2941	657.1429	594.5946	552.6316	461.5385
2011	552.6316	615.3846	657.8947	783.7838	846.1538	837.8378	763.1579	694.0000	625.0000	609.7561	571.4286	512.1951
2012	536.5854	595.2381	658.5366	756.0976	878.0488	825.0000	756.0976	714.2857	651.1628	642.8571	619.0476	547.6190
2013	581.3953	613.6364	622.2222	727.2727	826.0870	782.6087	680.8511	645.8333	625.0000	617.0213	586.9565	590.9091
2014	562.5000	571.4286	625.0000	723.4043	847.8261	791.6667	744.6809	739.1304	666.6667	659.5745	625.0000	608.0000

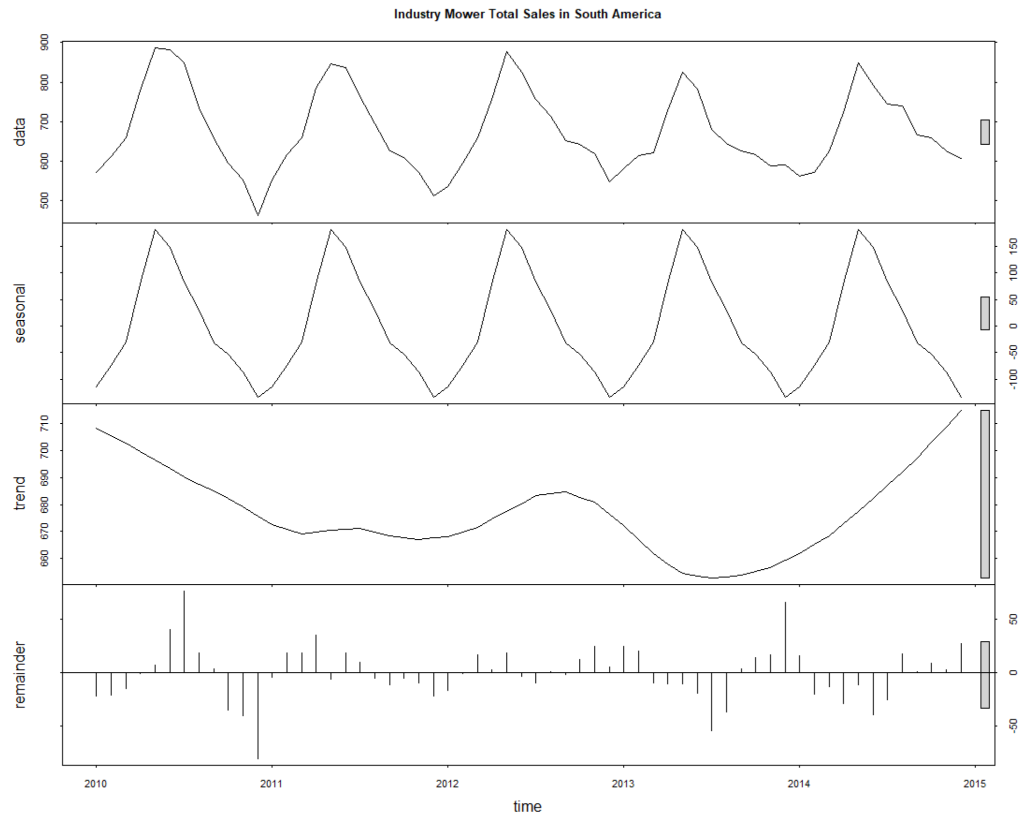


Graph 80: The time series graph above shows the mower industy sales in South America from January 2010 to December 2014.

From the graph, there is no apparent trend in the data over this period.
However, mower industry sales shows strong seasonality each year. There are periodic fluctuations every year, where the mower industry sales increases till aounrd May before

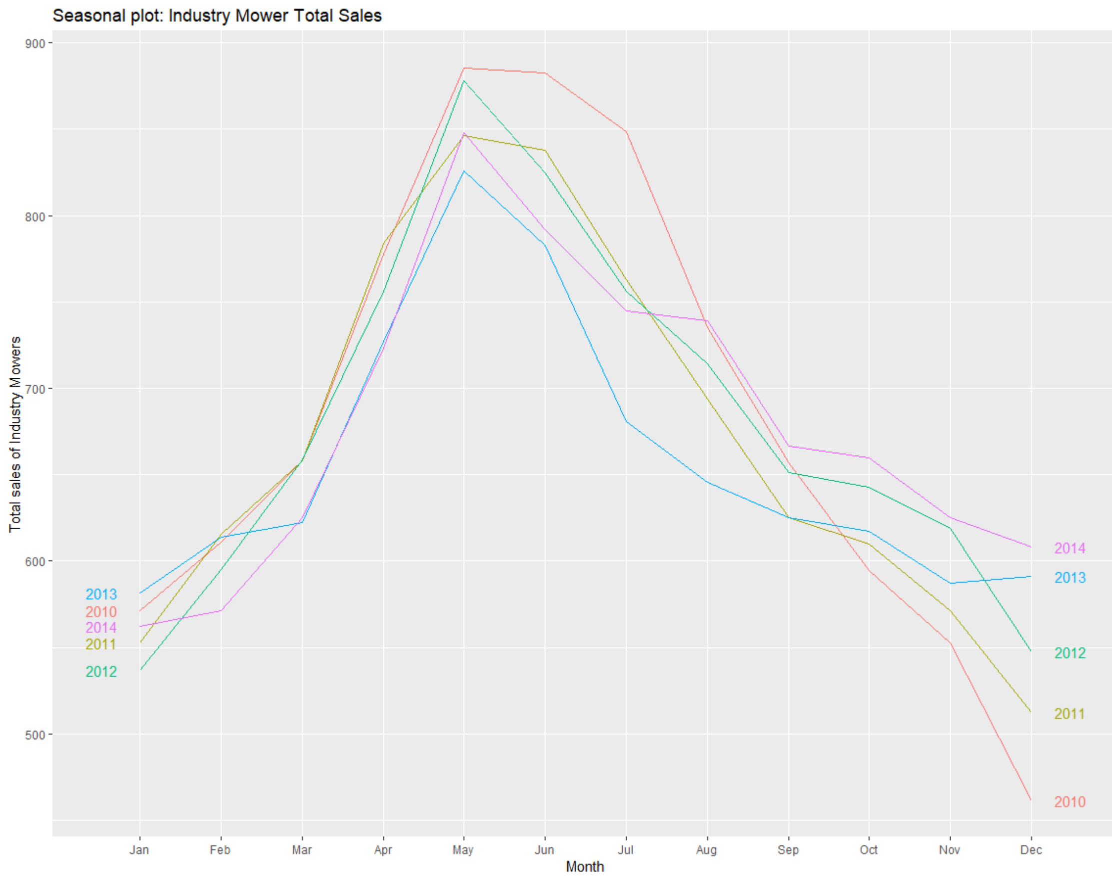
decreasing back to around the original volume of mower industry sales that was sold at the start of the year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 81: The time series graph above shows the mower industry sales in South America from January 2010 to December 2014.

Decomposition of the time series further proves the point that there is no apparent trend but there is the presence of a clear seasonality.

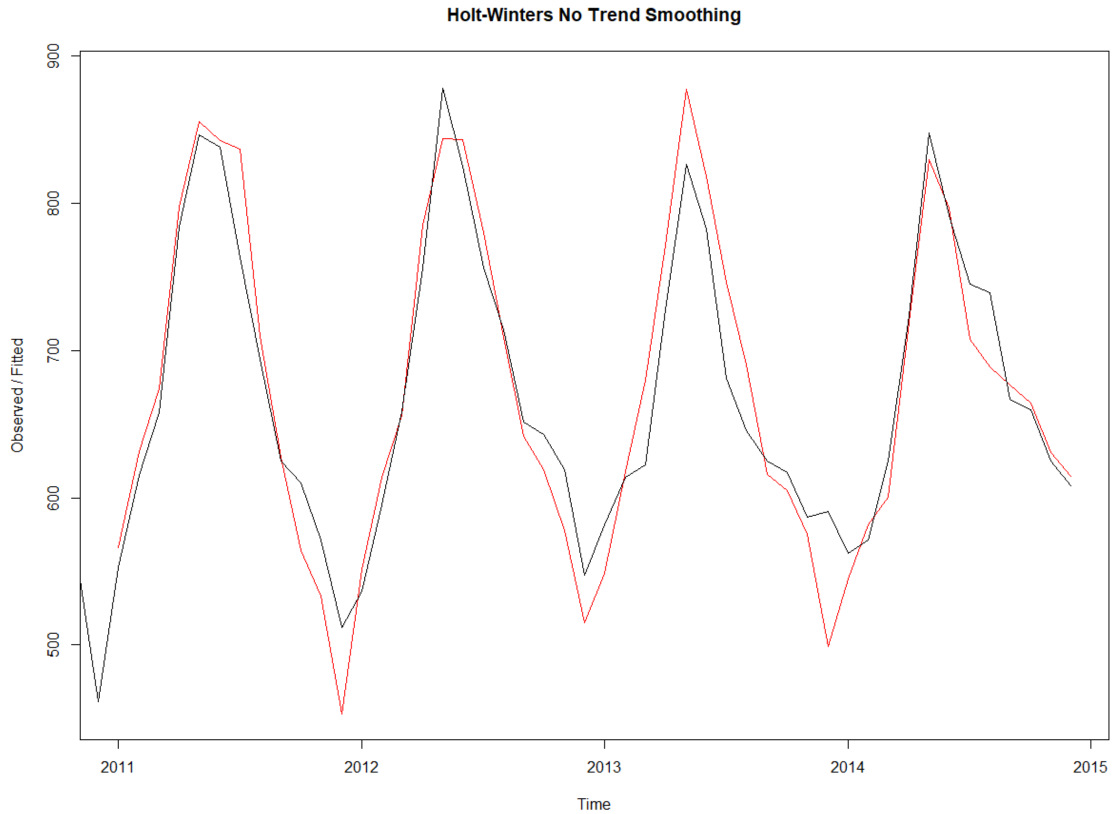


Graph 82: A seasonal plot of mower industry sales in South America from January 2010 to December 2014.

From the seasonal plot, we can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does not vary with the level of the time series.

Hence, we can conclude that the industry sales of mowers in South America possess no <u>trend but has seasonality.</u>

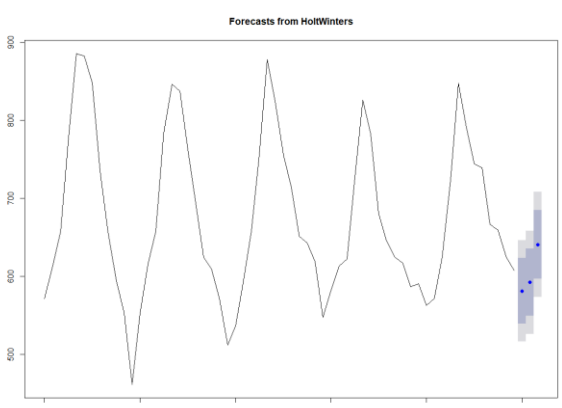
We can use <u>Holt-Winters No Trend Smoothing Model</u>.

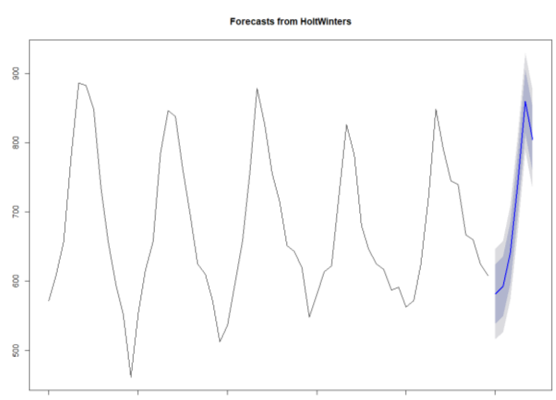


Graph 83: Observed time series data of mower industry sales in South America against the fitted Holt-Winters no-trend smoothing model

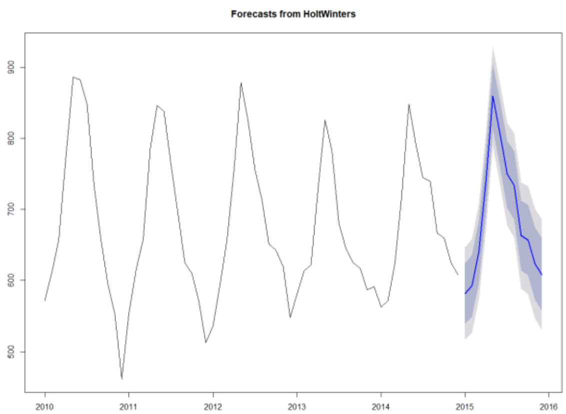
We see from the plot that the Holt-Winters no-trend smoothing model is very successful in predicting the seasonal peaks, which occur every May/June of the year. The model fits well with the observed time series data.

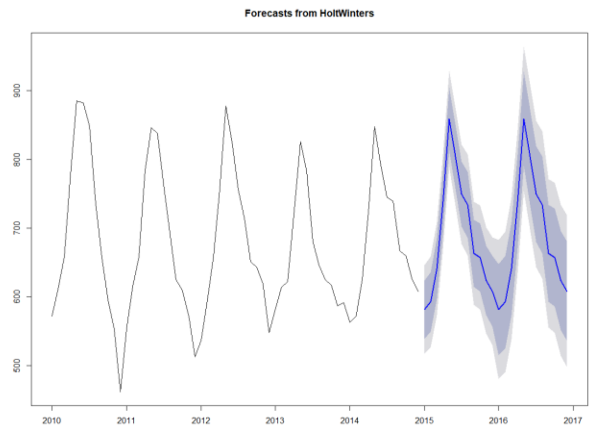
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months:





Graphs 84 and 85: Forecasted data of industry mower sales in South America over the next 3 and 6 months respectively





Graphs 86 and 87: Forecasted data of industry mower sales in South America over the next 12 and 24 months respectively.

The following are the forecasted values of mower industry sales in South America over the next 24 months

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	581.4114	539.0372	623.7856	516.6056	646.2172
Feb 2015	592.4573	549.2109	635.7037	526.3176	658.5969
Mar 2015	640.9407	596.8393	685.0420	573.4935	708.3878
Apr 2015	738.4957	693.5557	783.4357	669.7659	807.2255
May 2015	859.1575	813.3942	904.9208	789.1686	929.1464
Jun 2015	804.0485	757.4764	850.6205	732.8226	875.2743
Jul 2015	749.4560	702.0890	796.8230	677.0144	821.8976
Aug 2015	733.6830	685.5342	781.8318	660.0457	807.3203
Sep 2015	663.2364	614.3182	712.1546	588.4225	738.0503
Oct 2015	657.0552	607.3796	706.7308	581.0830	733.0275
Nov 2015	623.6259	573.2043	674.0475	546.5127	700.7391
Dec 2015	608.0000	556.8432	659.1568	529.7624	686.2376
Jan 2016	581.4114	514.9841	647.8387	479.8196	683.0032
Feb 2016	592.4573	525.4702	659.4443	490.0094	694.9052
Mar 2016	640.9407	573.3985	708.4828	537.6438	744.2375
Apr 2016	738.4957	670.4030	806.5885	634.3569	842.6346
May 2016	859.1575	790.5186	927.7964	754.1834	964.1316
Jun 2016	804.0485	734.8677	873.2292	698.2456	909.8513
Jul 2016	749.4560	679.7376	819.1744	642.8310	856.0811
Aug 2016	733.6830	663.4311	803.9349	626.2420	841.1240
Sep 2016	663.2364	592.4550	734.0178	554.9856	771.4872
Oct 2016	657.0552	585.7483	728.3622	548.0006	766.1098
Nov 2016	623.6259	551.7972	695.4546	513.7734	733.4784
Dec 2016	608.0000	535.6534	680.3466	497.3554	718.6446

Conclusion:

As seen from the forecasted graphs and values generated, industry mower sales in South America is predicted to remain the same for the next 24 months to come. Seasonality will remain.

Future Predictions of Market Share in South America:

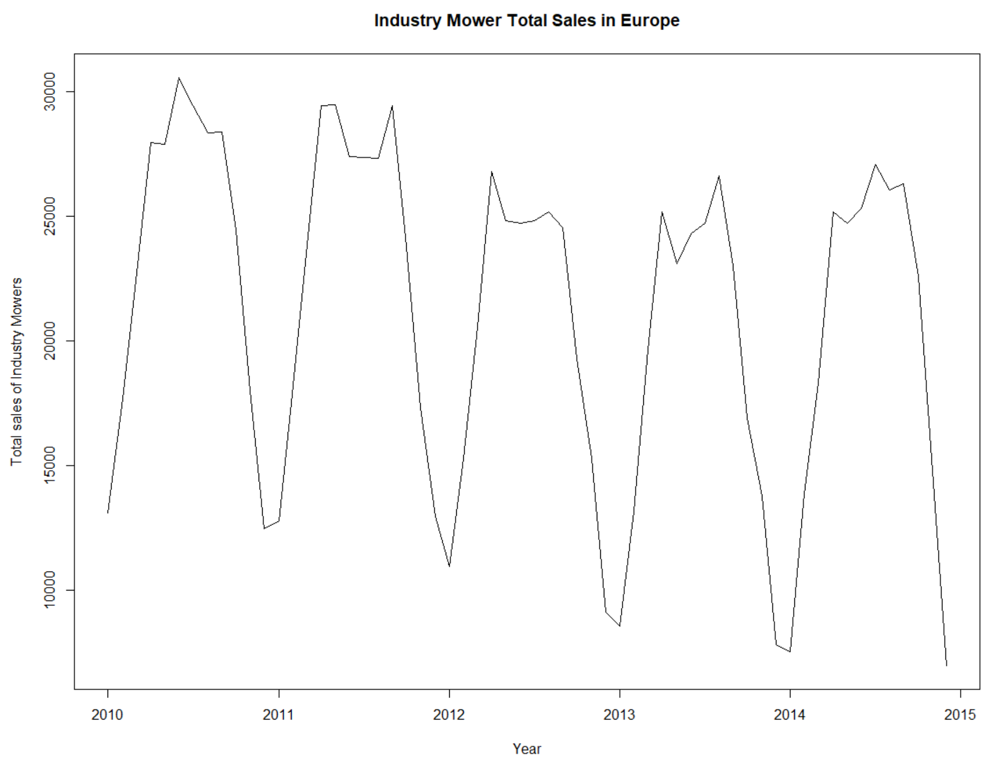
Forecasted unit mower sales in South America showed an upward trend while forecasted industry mower sales in South America showed similar sales output as compared to previous years. Both sales showed signs of similar seasonality patterns as well.

Since *market share = unit sales/industry sales*, there will be a predicted increase in market shares in South America for the sales of mowers.

Forecasting of Industry Sales of Mowers in Europe

Time Series:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	13090.909	17678.571	22758.621	27966.102	27894.737	30566.038	29444.444	28363.636	28392.857	24444.444	18000.000	12452.830
2011	12777.778	18214.286	23888.889	29454.545	29464.286	27413.793	27368.421	27321.429	29444.444	23773.585	17307.692	12941.176
2012	10961.538	15272.727	20555.556	26785.714	24827.586	24736.842	24827.586	25178.571	24545.455	19285.714	15272.727	9107.143
2013	8571.429	13157.895	19655.172	25178.571	23103.448	24285.714	24736.842	26607.143	22982.456	16896.552	13750.000	7818.182
2014	7547.170	13888.889	18301.887	25192.308	24705.882	25306.122	27083.333	26041.667	26304.348	22558.140	14772.727	6976.744

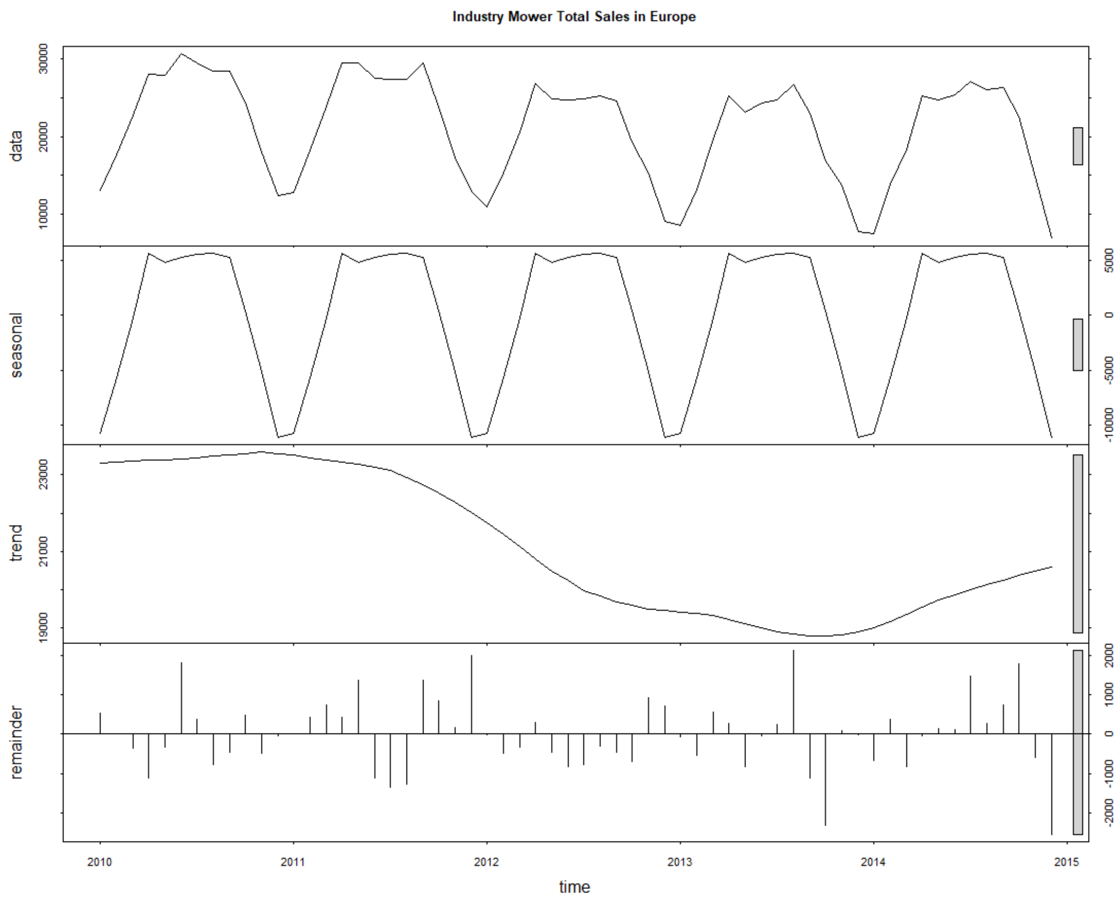


Graph 88: The time series graph above shows the mower industry sales in Europe from January 2010 to December 2014.

From the graph, there is a downward trend in the data over this period.

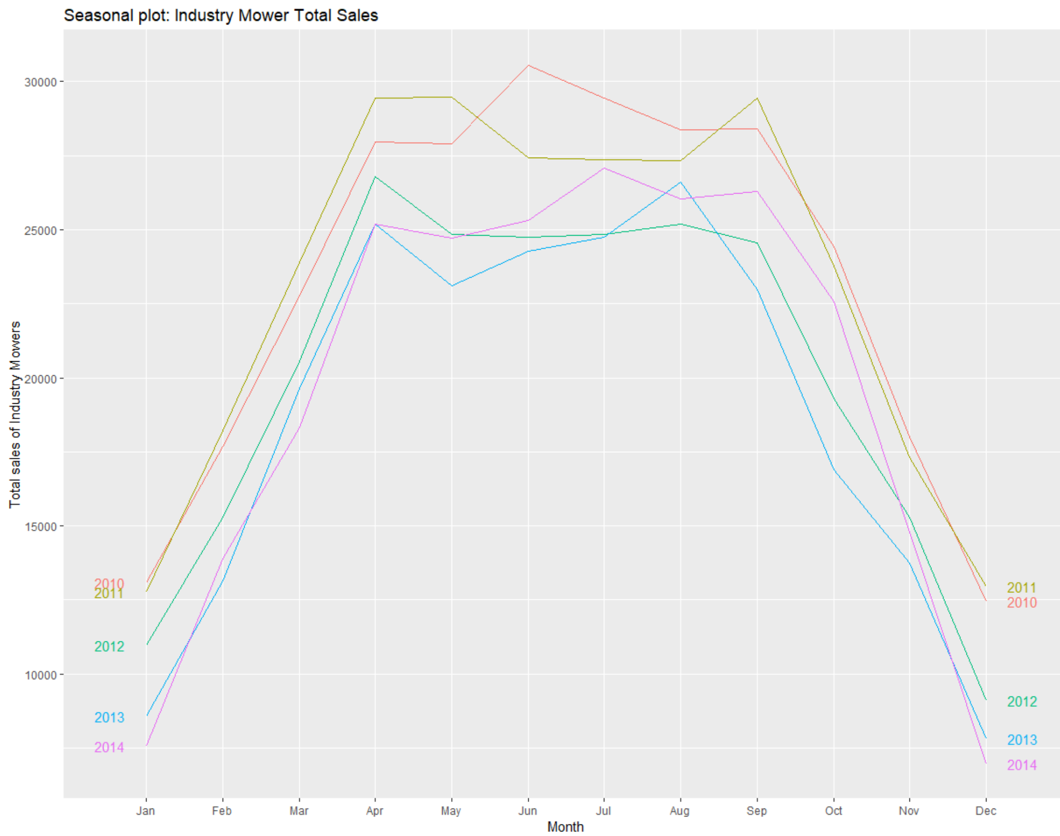
Mower unit sales shows strong seasonality within each year. There are periodic fluctuations every year, where the mower industry sales increases and plateaus till mid-year before decreasing back to around the original volume of mower industry sales that was sold at the start of the year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 89: Decomposed time series data of mower industry sales in Europe from January 2010 to December 2014

Decomposition of the time series further proves the point that there is a steady downward trend and there is the presence of a clear seasonality.

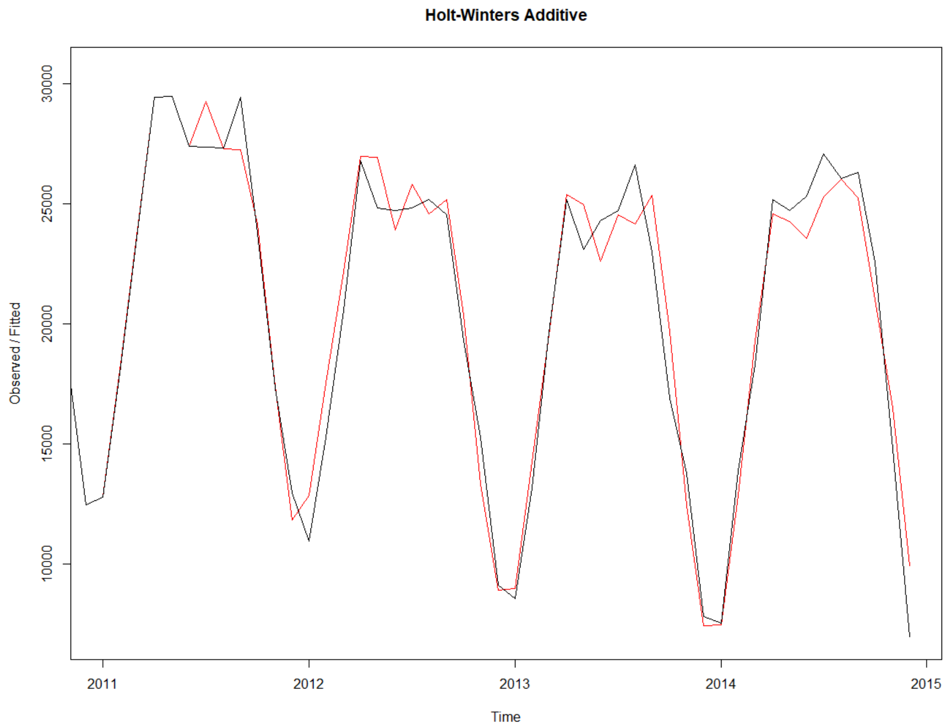


Graph 90: A seasonal plot of mower industry sales in Europe from January 2010 to December 2014.

We can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does not vary with the level of the time series.

Hence, we can conclude that the industry sales of mowers in Europe possess trend and seasonality.

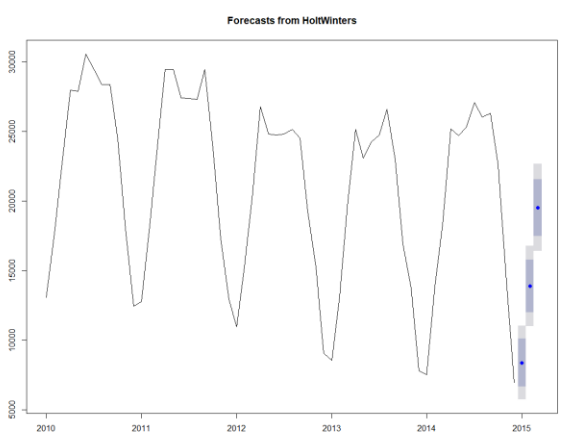
We can use Holt-Winters additive model to forecast future sales.

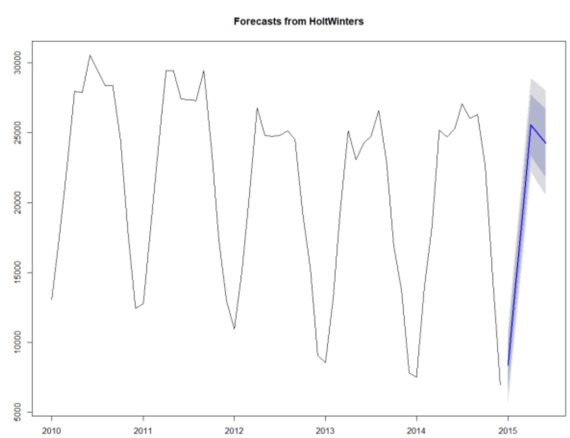


Graph 91: Observed time series data of mower industry sales in Europe against the fitted Holt-Winters additive model

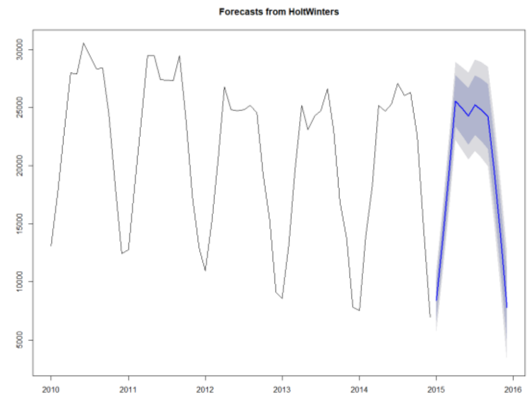
We see from the plot that the Holt-Winters additive model is very successful in predicting the seasonal peaks, which occur from around April to September each year, and the downward trend. The model fits well with the observed time series data.

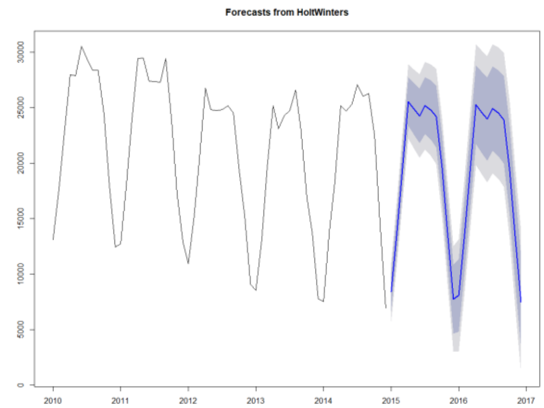
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months





Graphs 92 and 93: Forecasted data of industry mower sales in Europe over the next 3 and 6 months respectively.





Graphs 94 and 95: Forecasted data of industry mower sales in Europe over the next 12 and 24 months respectively.

The following are the forecasted values of mower industry sales in Europe over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	8404.004	6678.688	10129.32	5765.362	11042.65
Feb 2015	13913.799	12021.323	15806.27	11019.507	16808.09
Mar 2015	19539.518	17493.494	21585.54	16410.394	22668.64
Apr 2015	25564.909	23376.081	27753.74	22217.385	28912.43
May 2015	24927.252	22604.384	27250.12	21374.731	28479.77
Jun 2015	24268.003	21818.417	26717.59	20521.684	28014.32
Jul 2015	25201.366	22631.303	27771.43	21270.794	29131.94
Aug 2015	24803.847	22118.708	27488.99	20697.281	28910.41
Sep 2015	24220.900	21425.417	27016.38	19945.578	28496.22
Oct 2015	19612.814	16711.181	22514.45	15175.149	24050.48
Nov 2015	13774.434	10770.399	16778.47	9180.159	18368.71
Dec 2015	7771.622	4668.563	10874.68	3025.902	12517.34
Jan 2016	8111.437	4822.791	11400.08	3081.888	13140.99
Feb 2016	13621.232	10241.892	17000.57	8452.977	18789.49
Mar 2016	19246.951	15779.287	22714.62	13943.617	24550.29
Apr 2016	25272.342	21718.549	28826.14	19837.285	30707.40
May 2016	24634.686	20996.802	28272.57	19071.023	30198.35
Jun 2016	23975.436	20255.362	27695.51	18286.074	29664.80
Jul 2016	24908.799	21108.312	28709.29	19096.456	30721.14
Aug 2016	24511.281	20632.047	28390.51	18578.505	30444.06
Sep 2016	23928.333	19971.920	27884.75	17877.522	29979.14
Oct 2016	19320.247	15288.132	23352.36	13153.659	25486.84
Nov 2016	13481.868	9375.445	17588.29	7201.636	19762.10
Dec 2016	7479.056	3299.647	11658.46	1087.202	13870.91

Conclusion:

As seen from the forecasted graphs and values generated, industry mower sales in Europe is predicted to decrease for the next 24 months to come. Seasonality will remain.

Future Predictions of Market Share in Europe:

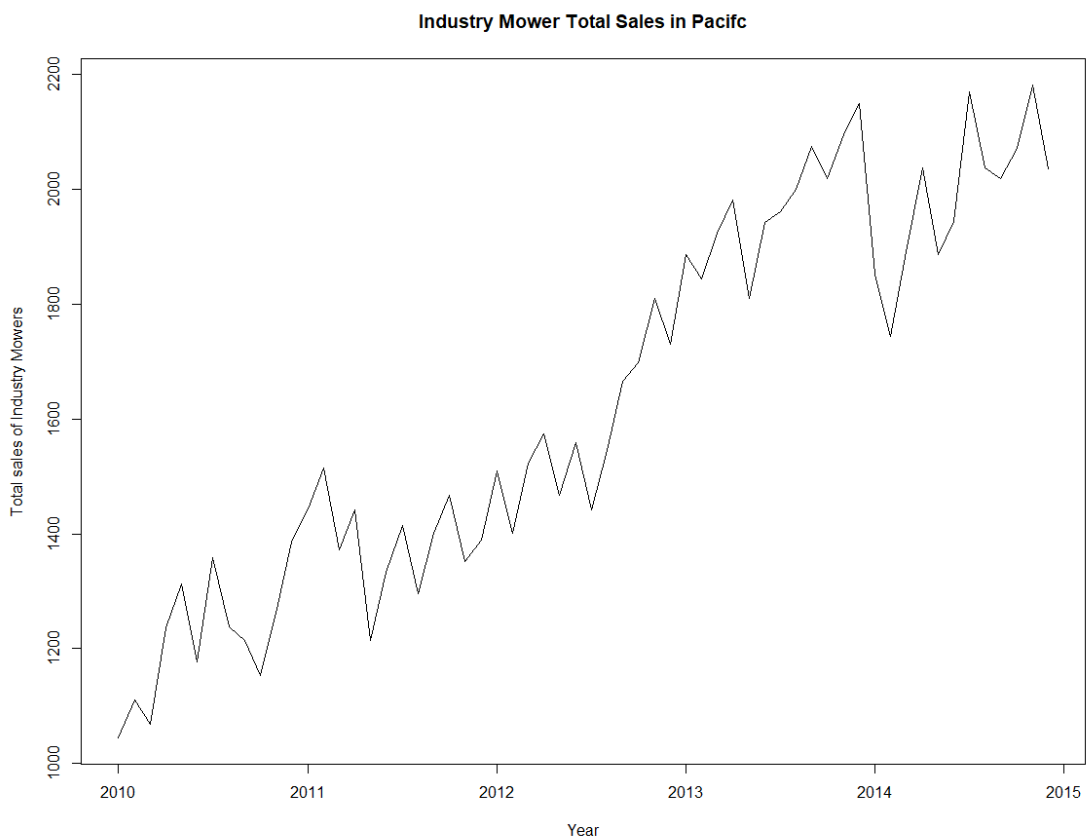
Both forecasted mower sales (Industry and Unit) in Europe had similar values. Both saw a similar rate and magnitude of decrease in sales in the next 24 months. Meanwhile, seasonality remained where where the mower sales increases till April to September before decreasing back to around the original volume of mower sales that was sold at the start of the year.

Since *market share = unit sales/industry sales*, there will be no predicted change in market shares in Europe for the sales of mowers.

Forecasting of Industry Sales of Mowers in Pacific

Time Series:

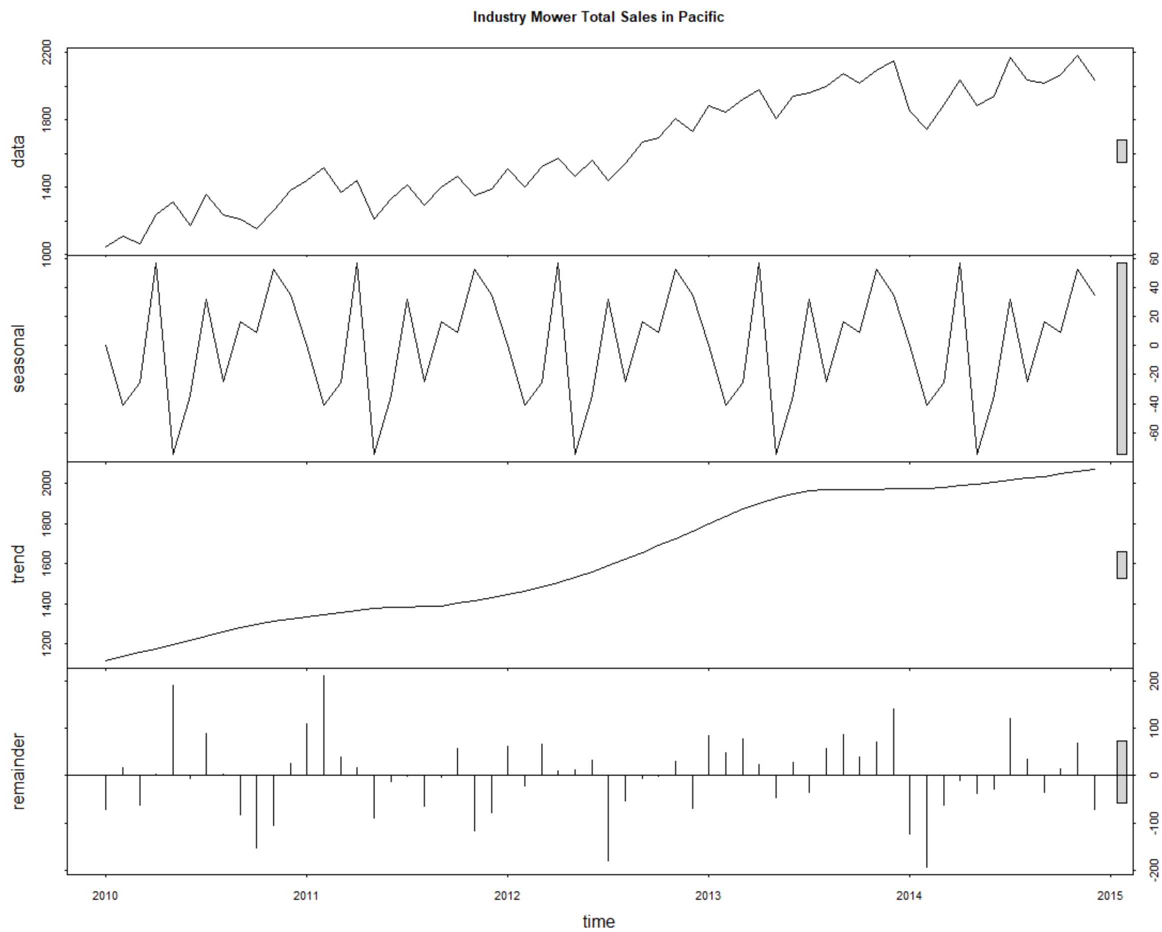
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	1045.000	1111.111	1067.961	1237.113	1313.131	1176.471	1359.223	1238.095	1214.953	1153.846	1262.136	1386.139
2011	1443.299	1515.152	1372.549	1442.308	1214.953	1333.333	1415.094	1296.296	1401.869	1467.890	1351.351	1388.889
2012	1509.434	1401.869	1523.810	1574.074	1467.890	1559.633	1441.441	1545.455	1666.667	1698.113	1809.524	1730.769
2013	1886.792	1844.660	1923.077	1981.132	1809.524	1941.748	1960.784	2000.000	2075.472	2019.231	2095.238	2149.533
2014	1851.852	1743.119	1891.892	2037.037	1886.792	1944.444	2169.811	2037.037	2018.349	2072.072	2181.818	2035.398



Graph 96: The time series graph above shows the mower industry sales in Pacific from January 2010 to December 2014

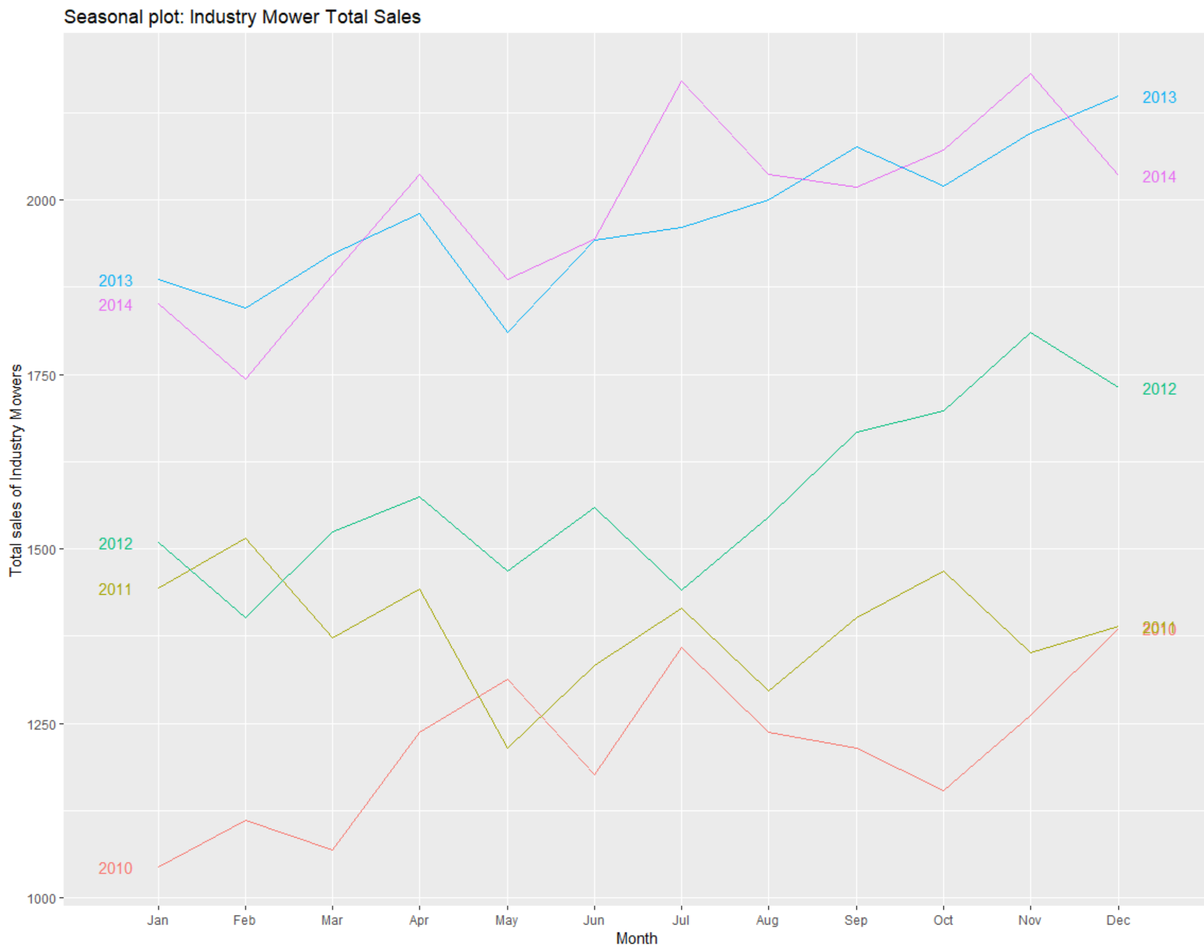
From the graph, there is an upward trend in the data over this period.
Mower Industry Sales shows no seasonality within each year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 97: Decomposed time series data of mower industry sales in Pacific from January 2010 to December 2014

Decomposition of the time series further proves the point that there is a steady upward trend. But we are not certain about the presence of seasonality yet.

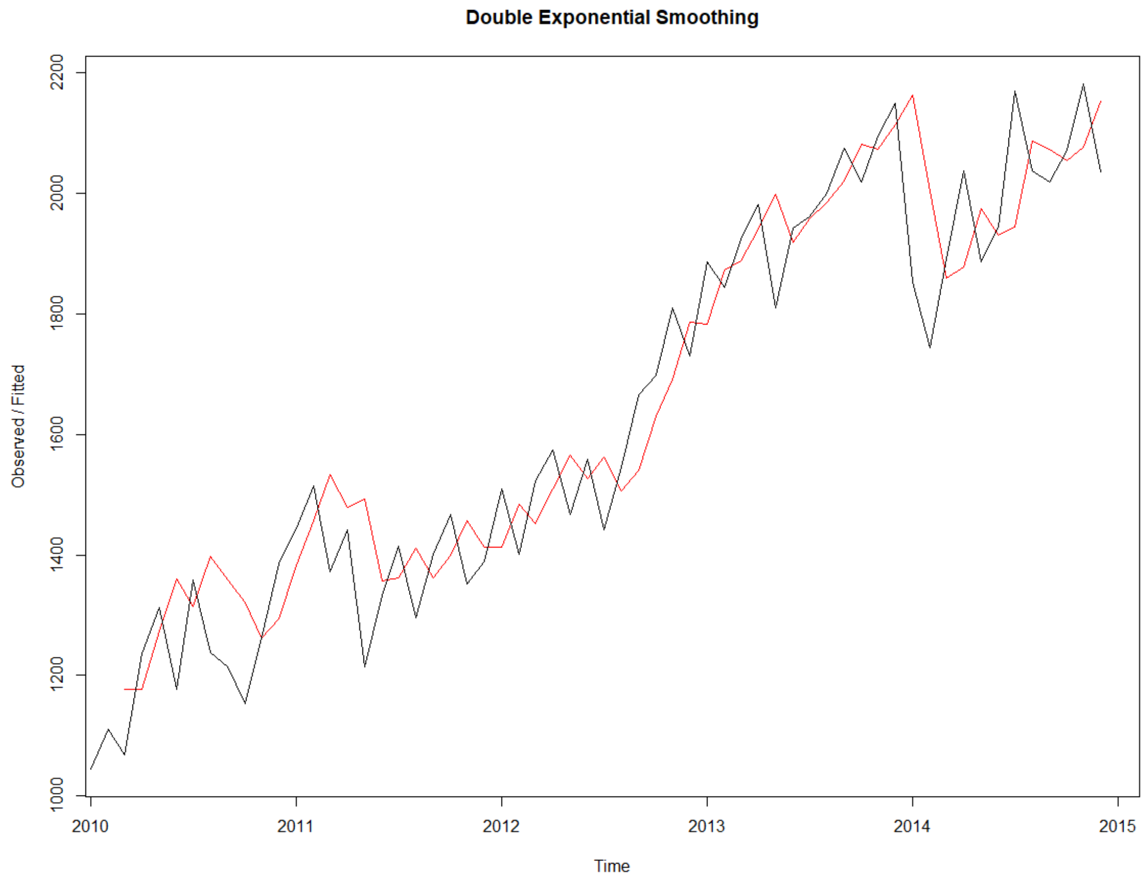


Graph 98: A seasonal plot of mower industry sales in Pacific fom January 2010 to December 2014.

As seen in the seasonal plot, it further proves that there is no seasonality in Pacific due to different peaks in different times of the year.

Hence, we can conclude that the industry sales of mowers in Pacific possess trend but no seasonality.

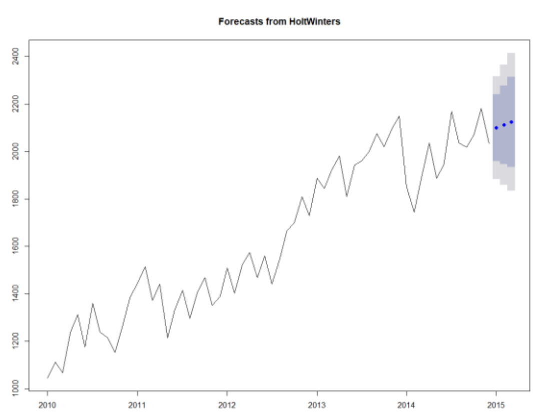
We can use Double Exponential Smoothing to forecast future sales.

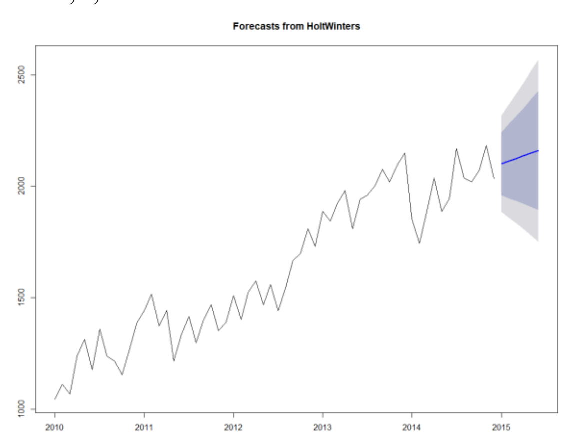


Graph 99: Observed time series data of mower industry sales in Pacific against the fitted double exponential smoothing model

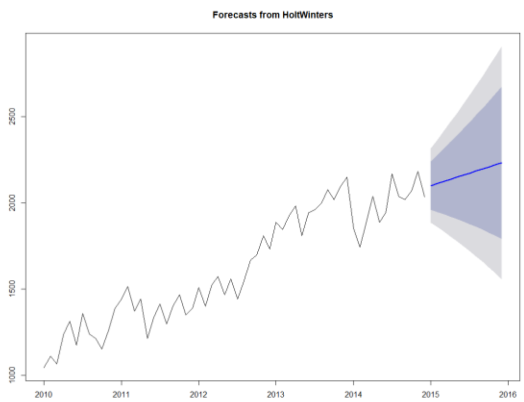
We see from the plot that the model is very successful in predicting the upward trend. The model fits well with the observed time series data.

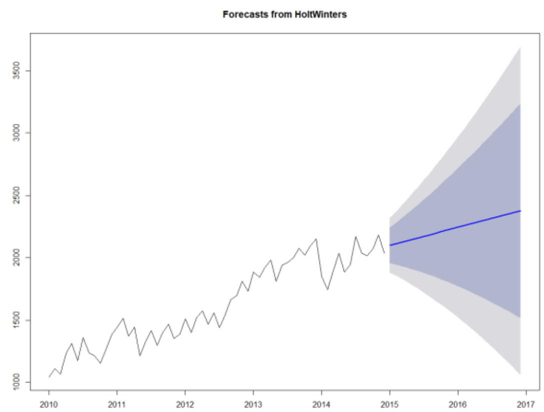
Hence, we can make use of it to predict data in the next 3,6,12 and 24 months





Graphs 100 and 101: Forecasted data of industry mower sales in Pacific over the next 3 and 6 months respectively.





Graphs 102 and 103: Forecasted data of industry mower sales in Pacific over the next 12 and 24 months respectively.

The following are the forecasted values of mower industry sales in Pacific over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	2100.203	1958.728	2241.679	1883.836	2316.571
Feb 2015	2112.248	1946.751	2277.745	1859.142	2365.354
Mar 2015	2124.293	1934.218	2314.368	1833.598	2414.988
Apr 2015	2136.338	1921.051	2351.625	1807.085	2465.591
May 2015	2148.383	1907.208	2389.557	1779.538	2517.227
Jun 2015	2160.427	1892.669	2428.185	1750.927	2569.928
Jul 2015	2172.472	1877.427	2467.518	1721.239	2623.705
Aug 2015	2184.517	1861.479	2507.555	1690.473	2678.561
Sep 2015	2196.562	1844.830	2548.294	1658.634	2734.490
Oct 2015	2208.607	1827.486	2589.728	1625.733	2791.481
Nov 2015	2220.652	1809.455	2631.848	1591.781	2849.522
Dec 2015	2232.696	1790.747	2674.646	1556.793	2908.600
Jan 2016	2244.741	1771.370	2718.112	1520.783	2968.699
Feb 2016	2256.786	1751.336	2762.236	1483.767	3029.805
Mar 2016	2268.831	1730.654	2807.008	1445.760	3091.901
Apr 2016	2280.876	1709.334	2852.417	1406.778	3154.973
May 2016	2292.920	1687.386	2898.455	1366.835	3219.006
Jun 2016	2304.965	1664.819	2945.112	1325.946	3283.985
Jul 2016	2317.010	1641.643	2992.377	1284.125	3349.895
Aug 2016	2329.055	1617.867	3040.243	1241.386	3416.724
Sep 2016	2341.100	1593.499	3088.701	1197.743	3484.456
Oct 2016	2353.145	1568.549	3137.741	1153.209	3553.081
Nov 2016	2365.189	1543.024	3187.355	1107.795	3622.584
Dec 2016	2377.234	1516.932	3237.537	1061.515	3692.953

Conclusion:

As seen from the forecasted graphs and values generated, industry mower sales in Pacific is predicted to increase gradually for the next 24 months to come.

Future Predictions of Market Share in Pacific:

Both forecasted mower sales (Industry and Unit) in Pacific had similar values. Both saw a similar rate and magnitude of a gradual increase in sales in the next 24 months.

Since *market share = unit sales/industry sales*, there will be no predicted change in market shares in Pacific for the sales of mowers.

Forecasting of Industry Sales of Mowers in China

There was not data given for the industry sales of mowers. Hence, we cannot ascertain the future changes in market share of mowers in China.

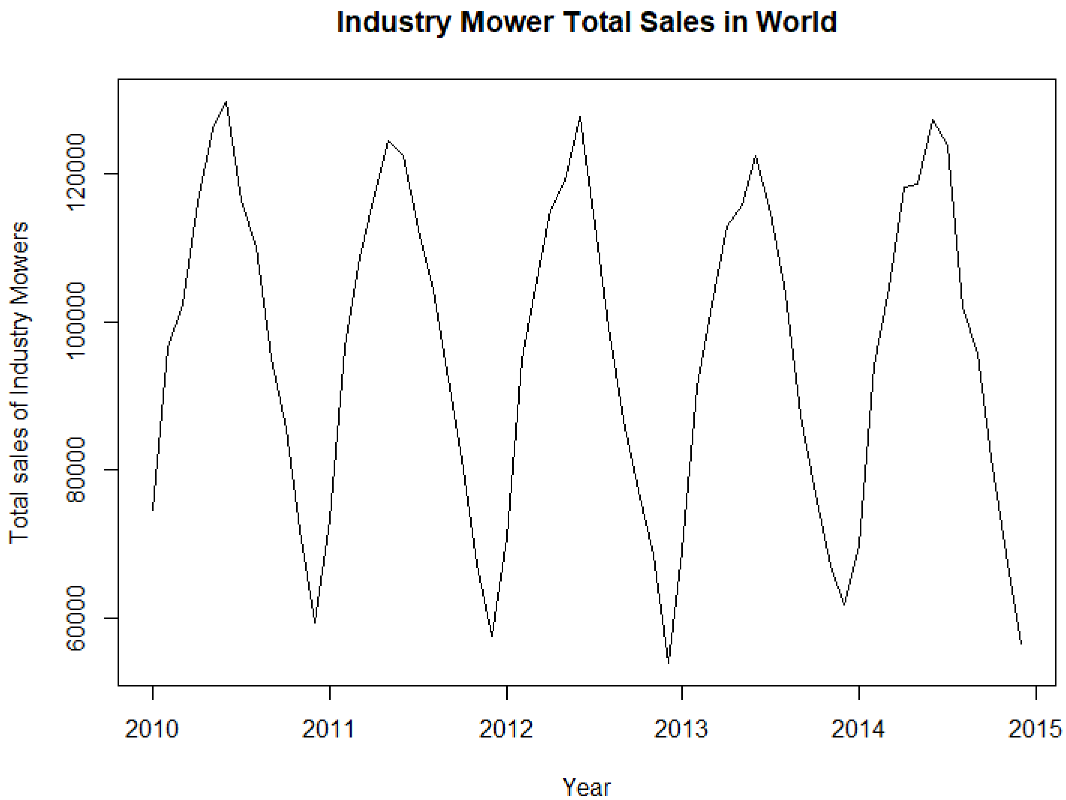
However, by looking simply at the forecasted values of unit mower sales in China earlier on, we can conclude that sales are looking bleak. Predicted sales in the next 24 months dipped all the way near zero.

Hence, assuming that industry sales of mowers stayed the same in China, we can conclude that market share of mowers in China will decrease.

Forecasting of Industry Sales of Mowers in the World

Time Series:

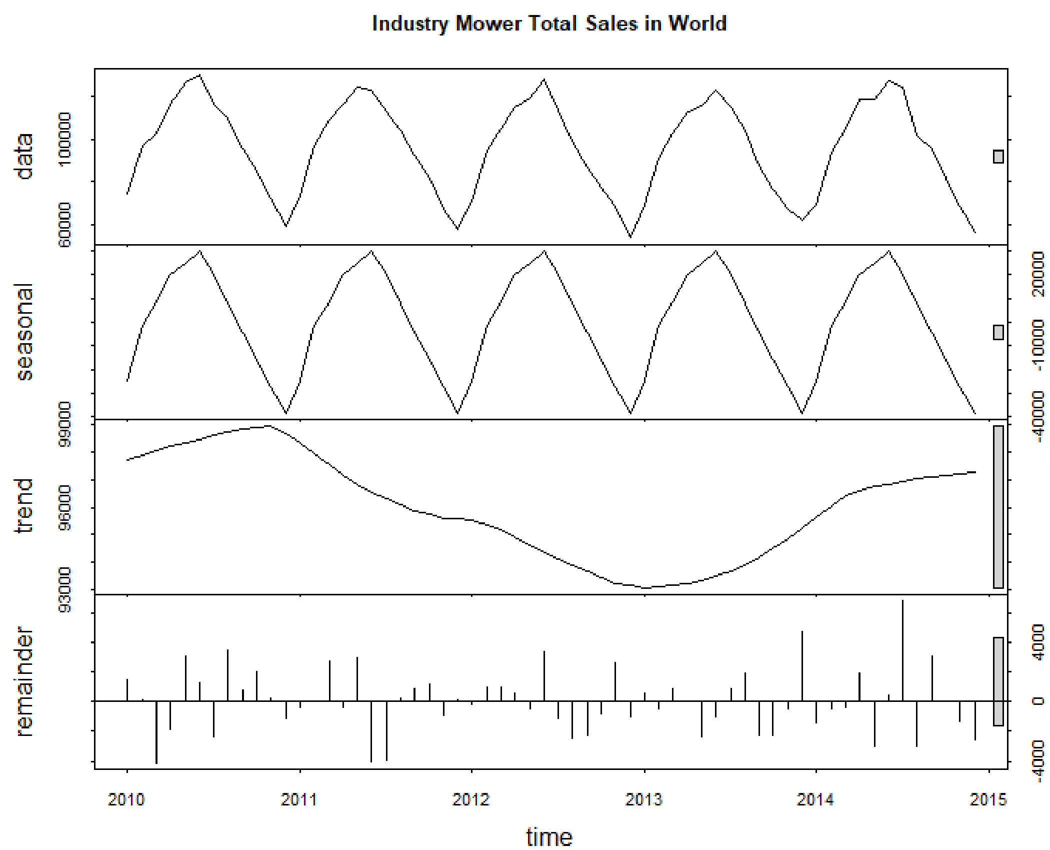
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	74662.34	96585.26	102369.09	116171.47	126210.09	129767.72	116409.43	110140.95	95064.95	85499.82	71971.63	59349.05
2011	73401.16	96544.82	108790.62	116584.48	124625.39	122584.96	112594.29	104164.40	92240.54	81470.28	67385.81	57489.32
2012	70892.17	94916.89	104582.56	115211.12	118949.23	127801.09	113215.60	99325.10	86863.28	77192.72	68558.44	53981.68
2013	69134.85	91182.23	102486.19	113027.16	115831.65	122481.77	114686.17	103729.17	87381.04	76770.90	67105.27	61796.72
2014	69673.06	94164.60	104544.27	118249.78	118583.36	127362.62	123919.39	101960.69	95688.39	81765.98	68647.51	56509.51



Graph 104: The time series graph above shows the mower industry sales in the World from January 2010 to December 2014.

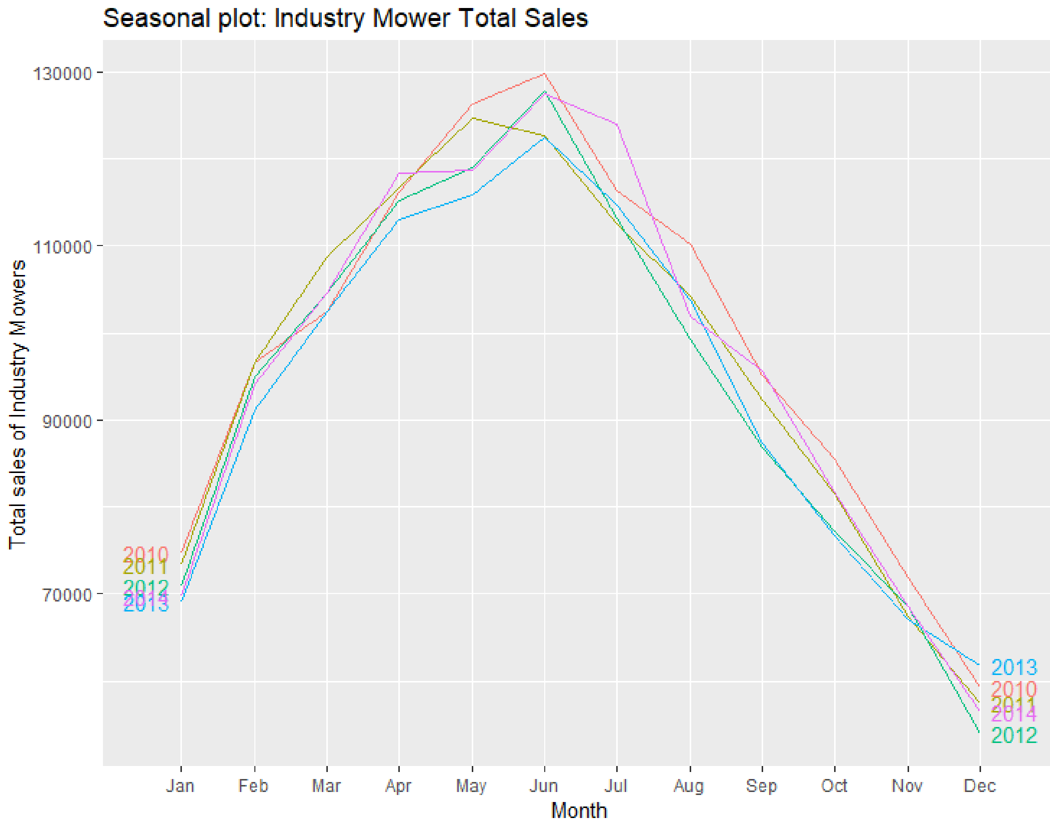
From the graph, there is no apparent trend in the data over this period.
However, mower unit sales shows strong seasonality each year. There are periodic fluctuations every year, where the mower industry sales increases till mid-year before decreasing back to around the original volume of mower industry sales that was sold at the start of the year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 105: Decomposed time series data of mower industry sales in the World from January 2010 to December 2014

Decomposition of the time series further proves the point that there is no apparent trend but there is the presence of a clear seasonality.

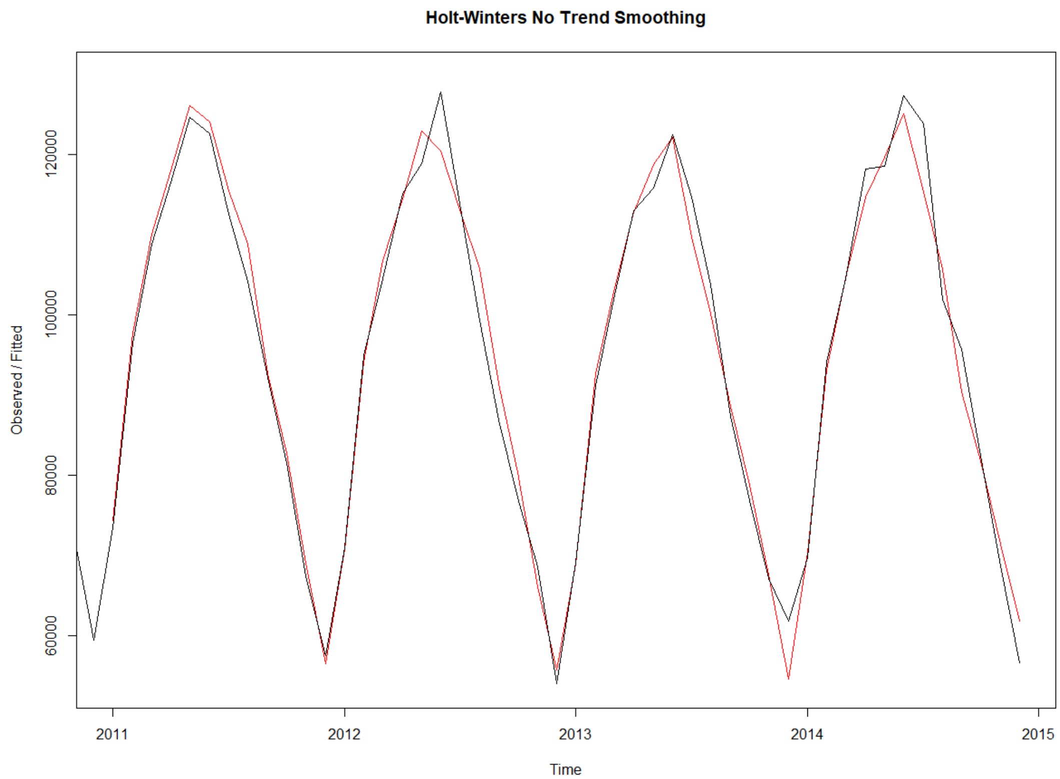


Graph 106: A seasonal plot of mower industry sales in the World from January 2010 to December 2014.

From the seasonal plot, it is clear that there is an increase of mower industry sales till June before it decreases all the way till the end of December. Plus, we can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does not vary with the level of the time series

Hence, we can conclude that the industry sales of mowers in the World possess no trend but has seasonality.

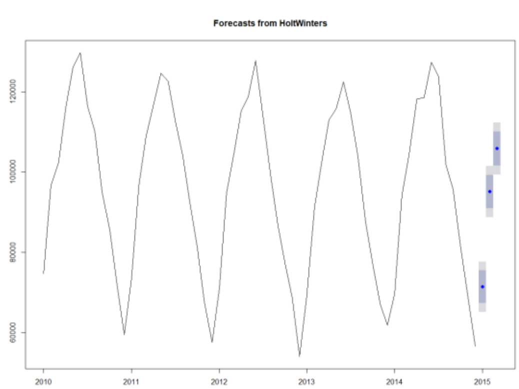
We can use the Holt-Winters no-trend smoothing to forecast future sales.

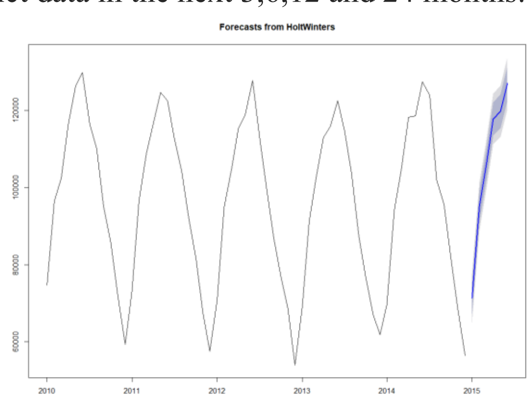


Graph 107: Observed time series data of mower industry sales in the World against the fitted Holt-Winters no-trend smoothing model

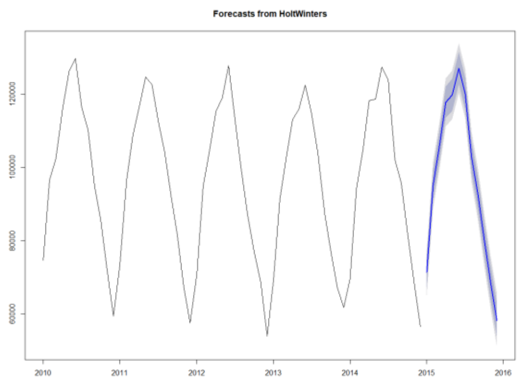
We see from the plot that the Holt-Winters no-trend smoothing model is very successful in predicting the seasonal peaks, which occur every June/July of the year. The model fits well with the observed time series data.

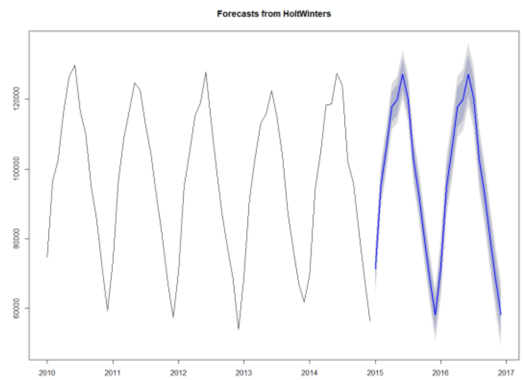
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months:





Graphs 108 and 109: Forecasted data of industry mower sales in the World over the next 3 and 6 months respectively.





Graphs 110 and 111: Forecasted data of industry mower sales in the World over the next 12 and 24 months respectively.

The following are the forecasted values of mower industry sales in the World over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	71364.52	67268.09	75460.94	65099.58	77629.46
Feb 2015	95108.96	90945.23	99272.69	88741.09	101476.83
Mar 2015	105847.35	101617.40	110077.31	99378.19	112316.51
Apr 2015	117790.76	113495.59	122085.93	111221.87	124359.65
May 2015	119790.14	115430.74	124149.54	113123.01	126457.27
Jun 2015	127090.73	122668.02	131513.43	120326.79	133854.67
Jul 2015	120117.90	115632.78	124603.01	113258.51	126977.29
Aug 2015	102706.20	98159.54	107252.87	95752.68	109659.73
Sep 2015	92633.20	88025.80	97240.59	85586.80	99679.60
Oct 2015	80180.87	75513.53	84848.20	73042.79	87318.94
Nov 2015	68541.63	63815.11	73268.14	61313.05	75770.21
Dec 2015	58174.89	53389.93	62959.86	50856.93	65492.86
Jan 2016	71364.52	65814.73	76914.30	62876.85	79852.18
Feb 2016	95108.96	89509.31	100708.61	86545.04	103672.88
Mar 2016	105847.35	100198.28	111496.42	97207.85	114486.86
Apr 2016	117790.76	112092.70	123488.82	109076.33	126505.19
May 2016	119790.14	114043.50	125536.77	111001.42	128578.86
Jun 2016	127090.73	121295.92	132885.53	118228.34	135953.12
Jul 2016	120117.90	114275.32	125960.47	111182.45	129053.34
Aug 2016	102706.20	96816.24	108596.16	93698.29	111714.12
Sep 2016	92633.20	86696.24	98570.16	83553.40	101713.00
Oct 2016	80180.87	74197.27	86164.47	71029.74	89331.99
Nov 2016	68541.63	62511.76	74571.50	59319.73	77763.53
Dec 2016	58174.89	52099.10	64250.69	48882.77	67467.02

Conclusion:
As seen from the forecasted graphs and values generated, industry mower sales in the World is predicted to remain the same for the next 24 months to come. Seasonality will remain.

Future Predictions of Market Share in the World:
Both forecasted mower sales (Industry and Unit) in the World had similar values. Both did not see any increase/decrease in sales in the next 24 months. Meanwhile, seasonality remained where where the mower sales increases till mid-year before decreasing back to around the original volume of mower sales that was sold at the start of the year.

Since *market share = unit sales/industry sales,* there will be no predicted change in market shares in the World for the sales of mowers.

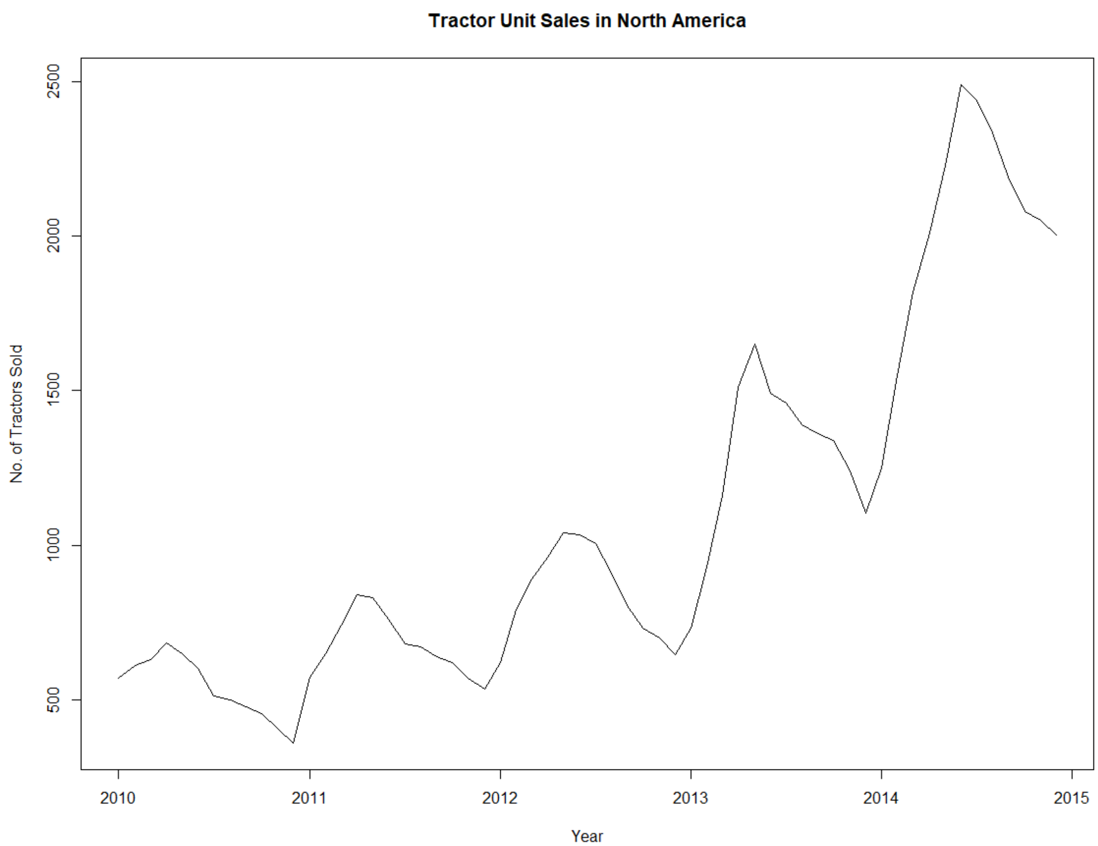
After we have predicted the future changes in market shares in the sales of mowers, we can now move on to predicting the future changes in market shares in the sales of tractors.

Firstly, we will forecast the sales of tractors in each marketing region using the excel sheet Tractor Unit Sales.

Forecasting of Unit Sales of Tractors in North America

Time Series:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	570	611	630	684	650	600	512	500	478	455	407	360
2011	571	650	740	840	830	760	681	670	640	620	570	533
2012	620	792	890	960	1040	1032	1006	910	803	730	699	647
2013	730	930	1160	1510	1650	1490	1460	1390	1360	1340	1240	1103
2014	1250	1550	1820	2010	2230	2490	2440	2334	2190	2080	2050	2004

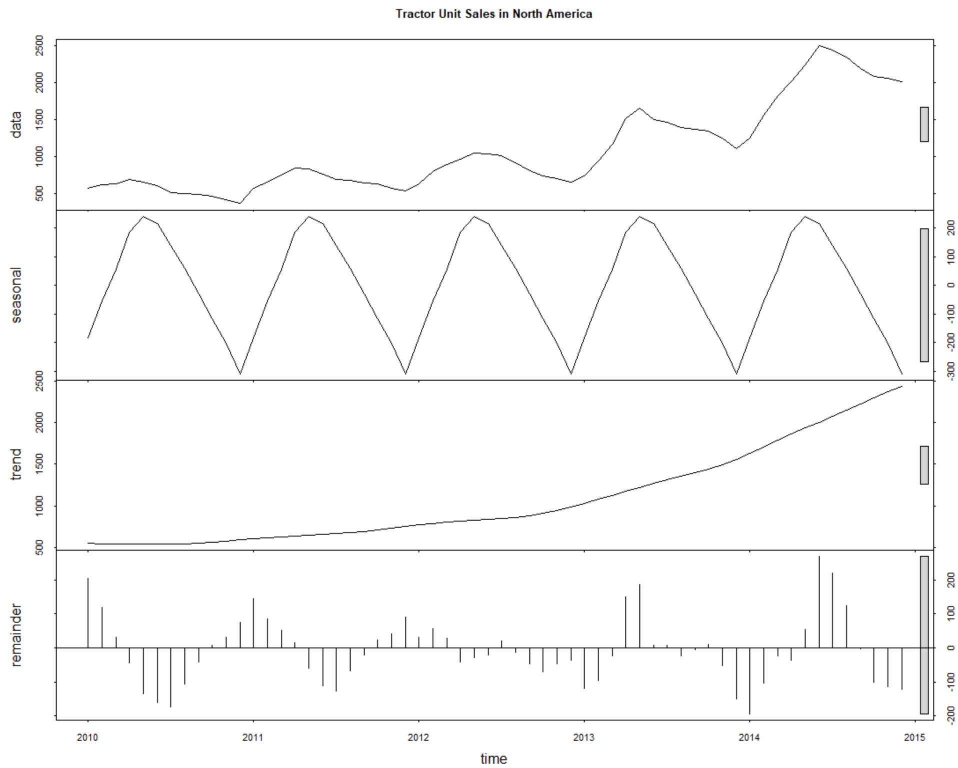


Graph 112: The time series graph above shows the tractor unit sales in North America from January 2010 to December 2014.

From the graph, there is an upward trend in the data over this period.

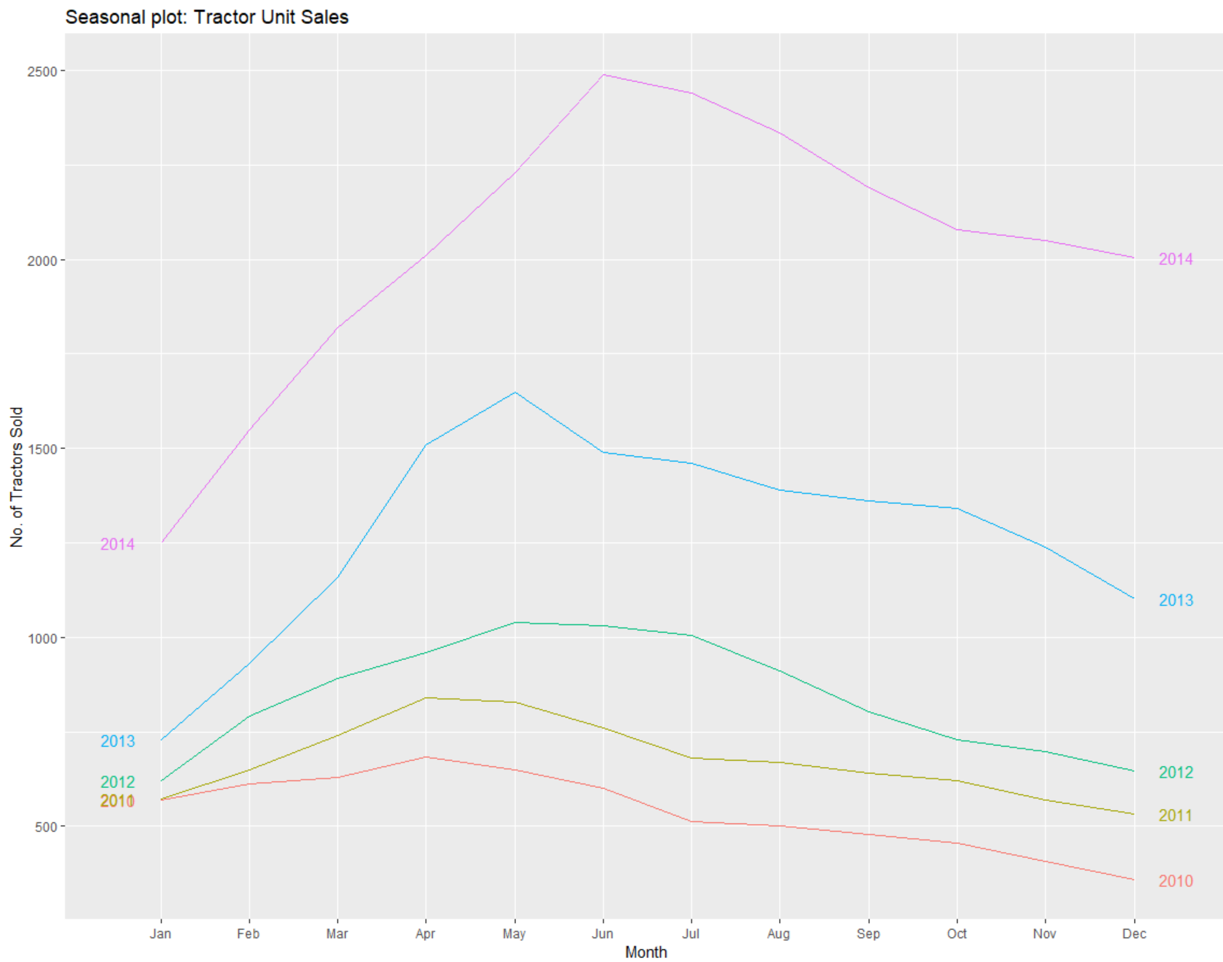
Tractor unit sales in North America shows strong seasonality each year. There are periodic fluctuations every year, where the mower unit sales increases till mid-year before decreasing back to around the original volume of mower unit sales that was sold at the start of the year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 113: The time series graph above shows the tractor unit sales in North America from January 2010 to December 2014.

Decomposition of the time series further proves the point that there is a steady upward trend and there is the presence of a clear seasonality.

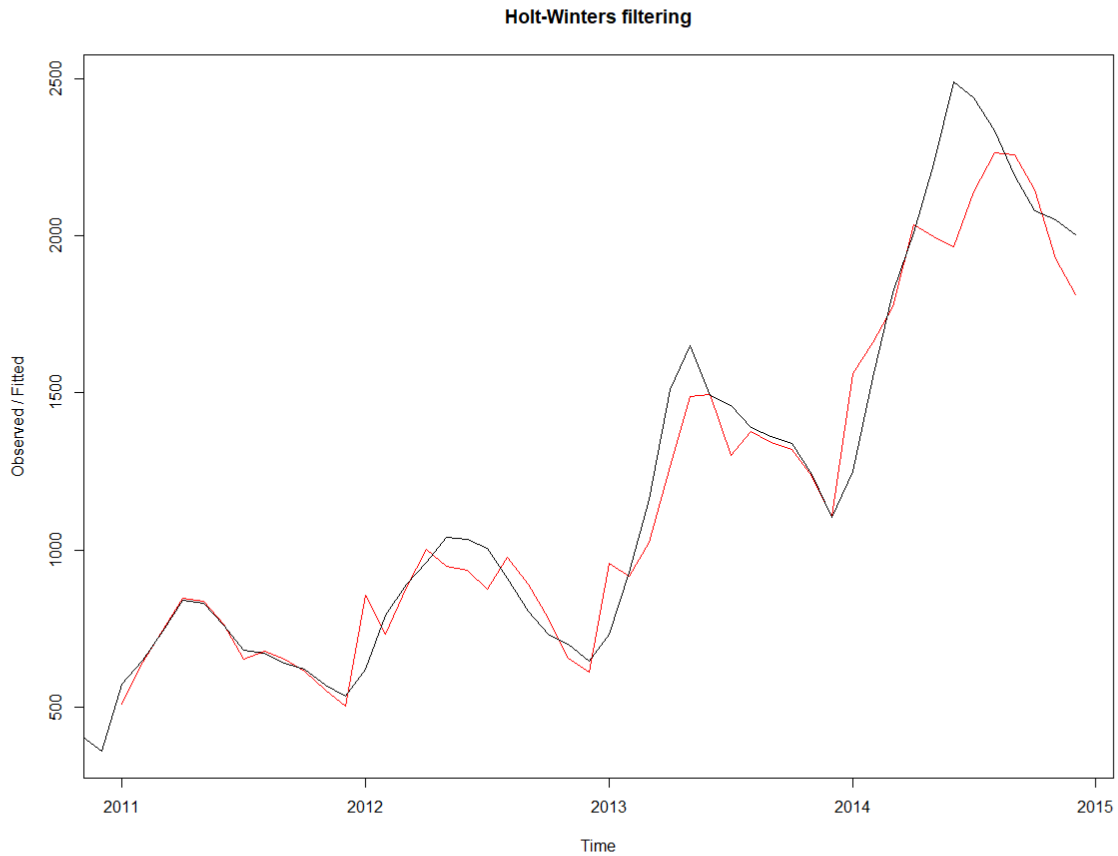


Graph 114: A seasonal plot of tractor unit sales in North America from January 2010 to December 2014.

From the seasonal plot, we can deduce that multiplicative decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does increase with the level of the time series.

Hence, we can conclude that the unit sales of tractors in North America possess <u>trend and seasonality.</u>

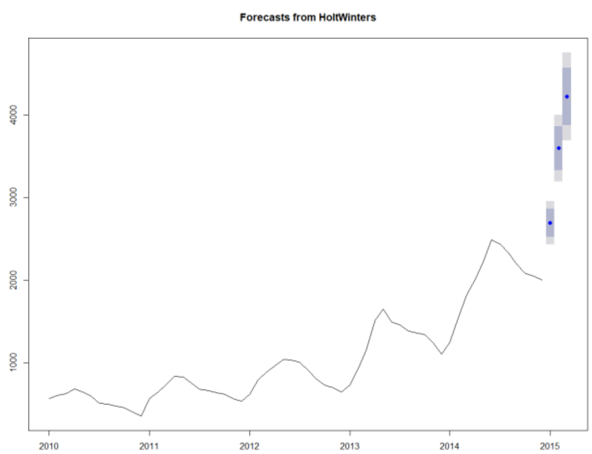
We can use <u>Holt-Winters multiplicative smoothing</u>.

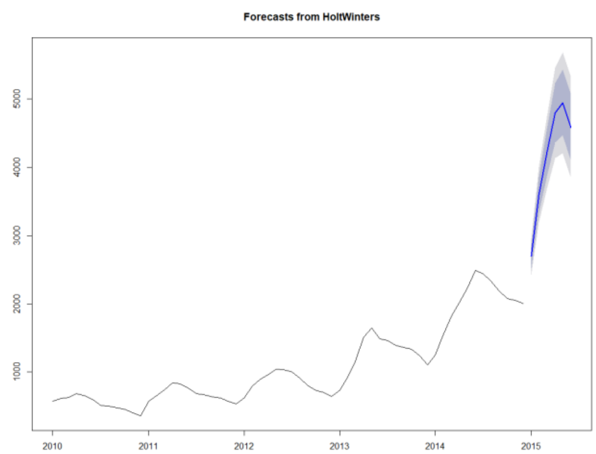


Graph 115: Observed time series data of tractor unit sales in North America against the fitted Holt-Winters multiplicative model

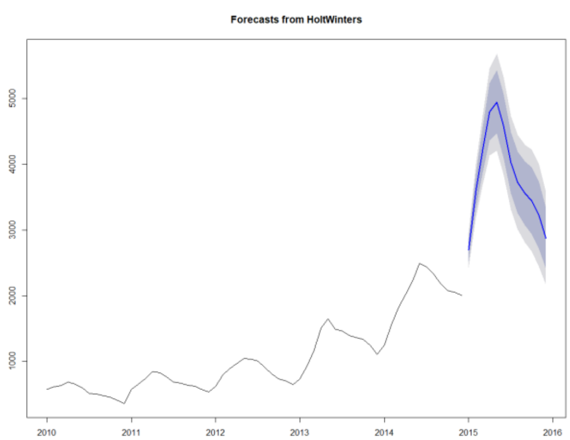
We see from the plot that the Holt-Winters multiplicative model is very successful in predicting the seasonal peaks, which occur every June/July of the year, and the upward trend. The model fits well with the observed time series data.

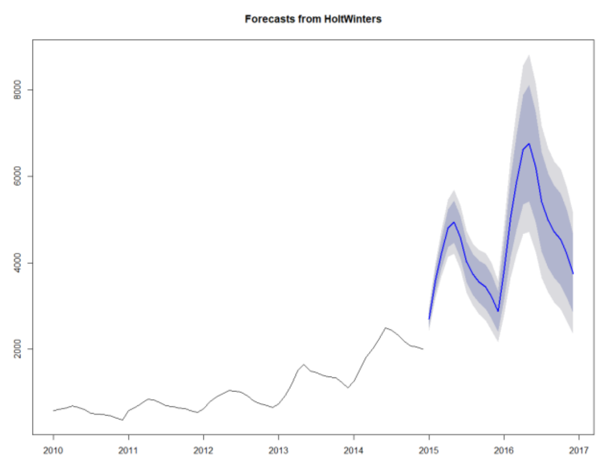
Hence, we can make use of Holt- Winters multiplicative model to predict data in the next 3,6,12 and 24 months:





Graphs 116 and 117: Forecasted data of unit tractor sales in North America over the next 3 and 6 months respectively.





Graphs 118 and 119: Forecasted data of unit tractor sales in North America over the next 12 and 24 months respectively.

The following are the forecasted values of tractor unit sales in North America over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	2692.911	2520.560	2865.262	2429.323	2956.499
Feb 2015	3595.098	3331.593	3858.603	3192.103	3998.093
Mar 2015	4220.906	3874.107	4567.706	3690.522	4751.291
Apr 2015	4799.059	4368.053	5230.065	4139.892	5458.226
May 2015	4948.677	4466.099	5431.255	4210.638	5686.717
Jun 2015	4581.751	4095.609	5067.892	3838.261	5325.240
Jul 2015	4021.650	3555.109	4488.191	3308.138	4735.162
Aug 2015	3724.185	3253.767	4194.602	3004.743	4443.626
Sep 2015	3554.243	3067.977	4040.510	2810.562	4297.924
Oct 2015	3445.537	2937.570	3953.503	2668.669	4222.404
Nov 2015	3214.646	2703.725	3725.568	2433.259	3996.033

Dec 2015	2872.838	2409.521	3336.156	2164.255	3581.421
Jan 2016	3819.717	3147.694	4491.740	2791.947	4847.487
Feb 2016	5048.723	4135.482	5961.963	3652.042	6445.404
Mar 2016	5871.937	4775.190	6968.683	4194.608	7549.266
Apr 2016	6616.979	5342.654	7891.305	4668.067	8565.892
May 2016	6765.909	5422.119	8109.699	4710.760	8821.059
Jun 2016	6214.284	4939.896	7488.671	4265.276	8163.291
Jul 2016	5413.291	4265.146	6561.436	3657.355	7169.227
Aug 2016	4976.771	3885.885	6067.658	3308.404	6645.139
Sep 2016	4717.080	3649.396	5784.764	3084.198	6349.962
Oct 2016	4542.890	3481.852	5603.927	2920.173	6165.606
Nov 2016	4211.994	3195.385	5228.602	2657.225	5766.762
Dec 2016	3741.676	2831.655	4651.697	2349.919	5133.433

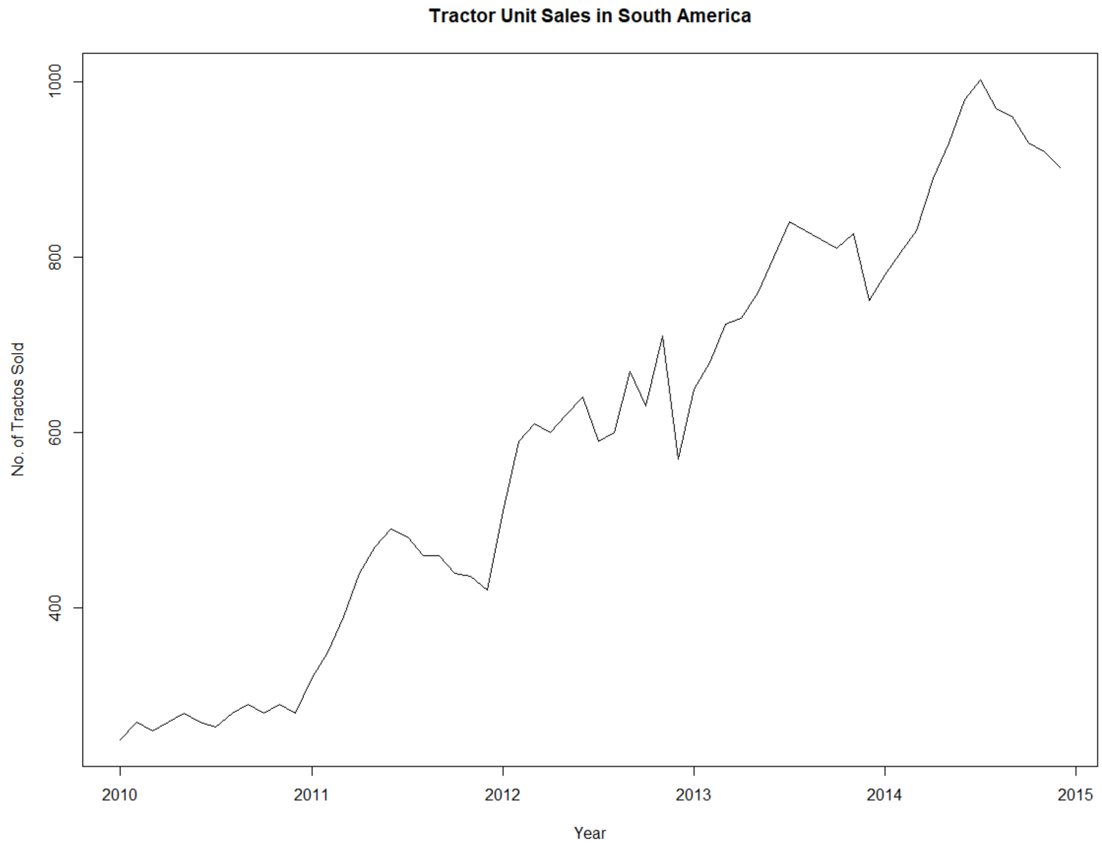
Conclusion:

As seen from the forecasted graphs and values generated, unit tractor sales in North America is predicted to increase for the next 24 months to come. Rate of increase will become bigger as magnitude increases. Seasonality will remain.

Forecasting of Unit Sales of Tractors in South America

Time Series:

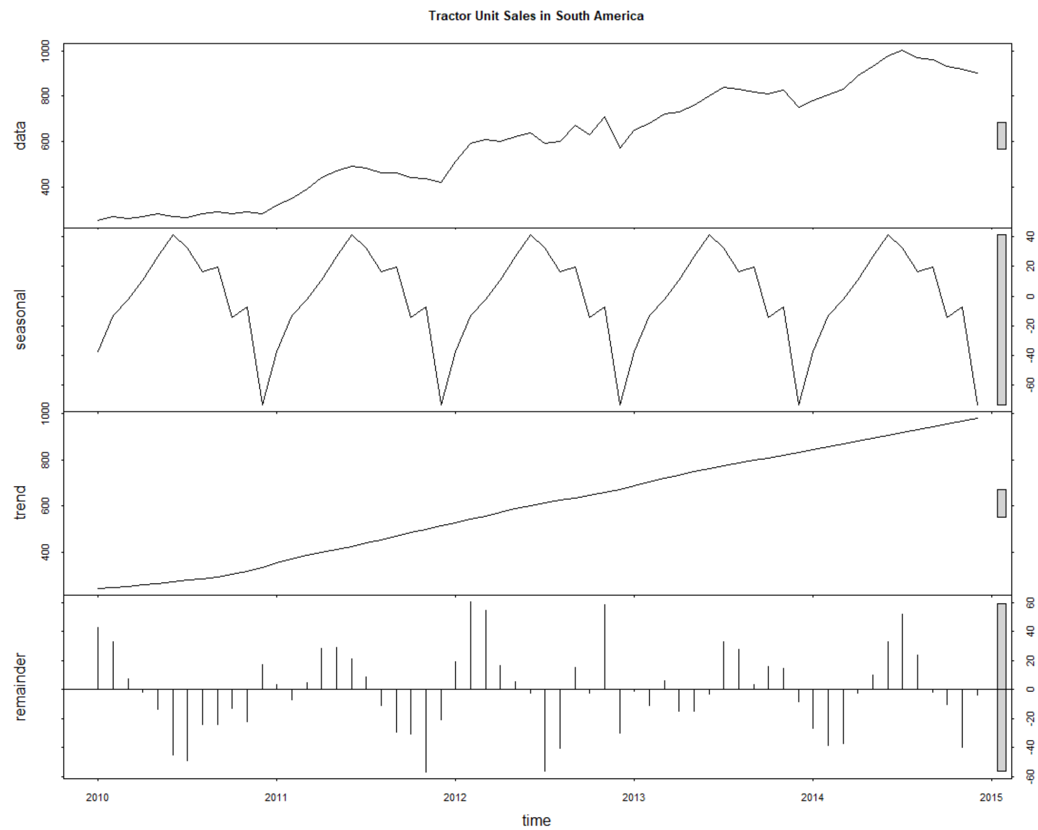
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	250	270	260	270	280	270	264	280	290	280	290	280
2011	320	350	390	440	470	490	481	460	460	440	436	420
2012	510	590	610	600	620	640	590	600	670	630	710	570
2013	650	680	724	730	760	800	840	830	820	810	827	750
2014	780	805	830	890	930	980	1002	970	960	930	920	902



Graph 120: The time series graph above shows the tractor unit sales in South America from January 2010 to December 2014.

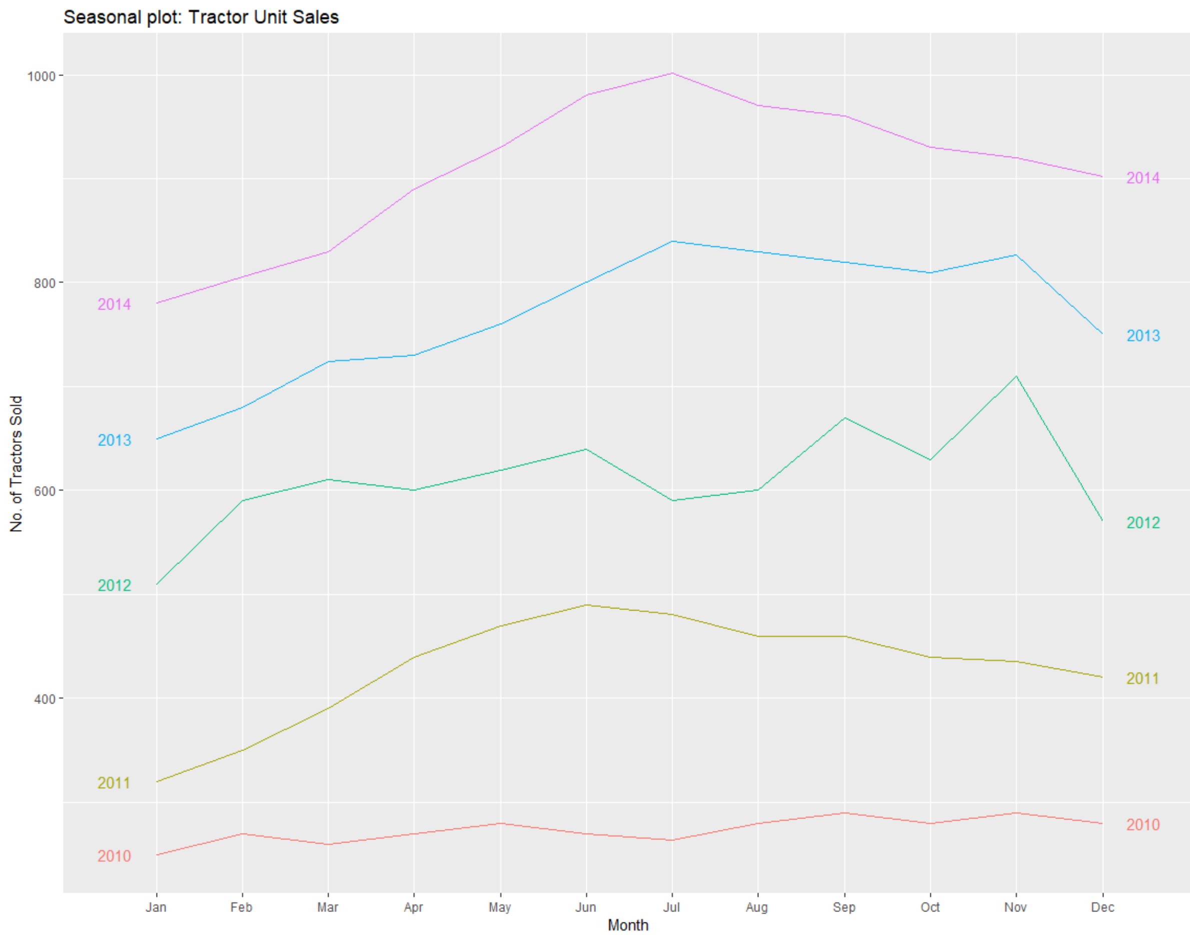
From the graph, there is an upward trend in the data over this period.
However, it is not clear whether there is the presence of seasonality due the differing fluctuation in sales every year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 121: The time series graph above shows the tractor unit sales in South America from January 2010 to December 2014.

Decomposition of the time series further proves the point that there is a steady upward trend. But we are not certain about the presence of seasonality yet.

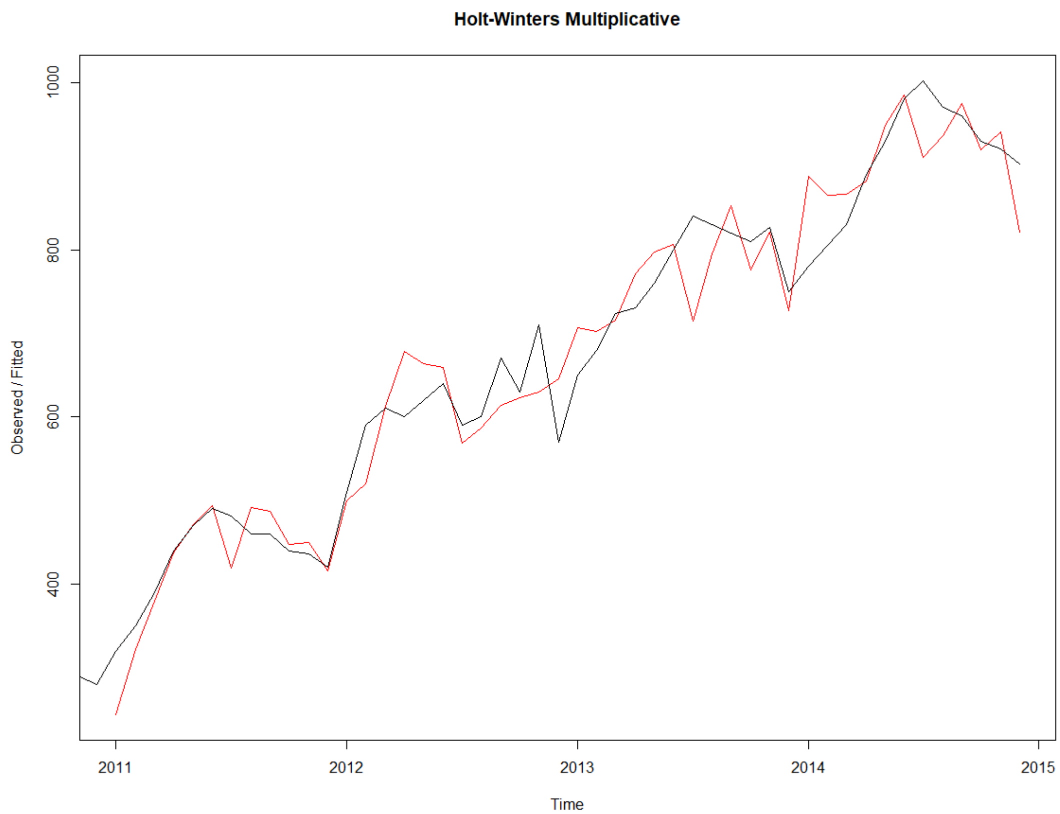


Graph 122: A seasonal plot of tractor unit sales in South America from January 2010 to December 2014.

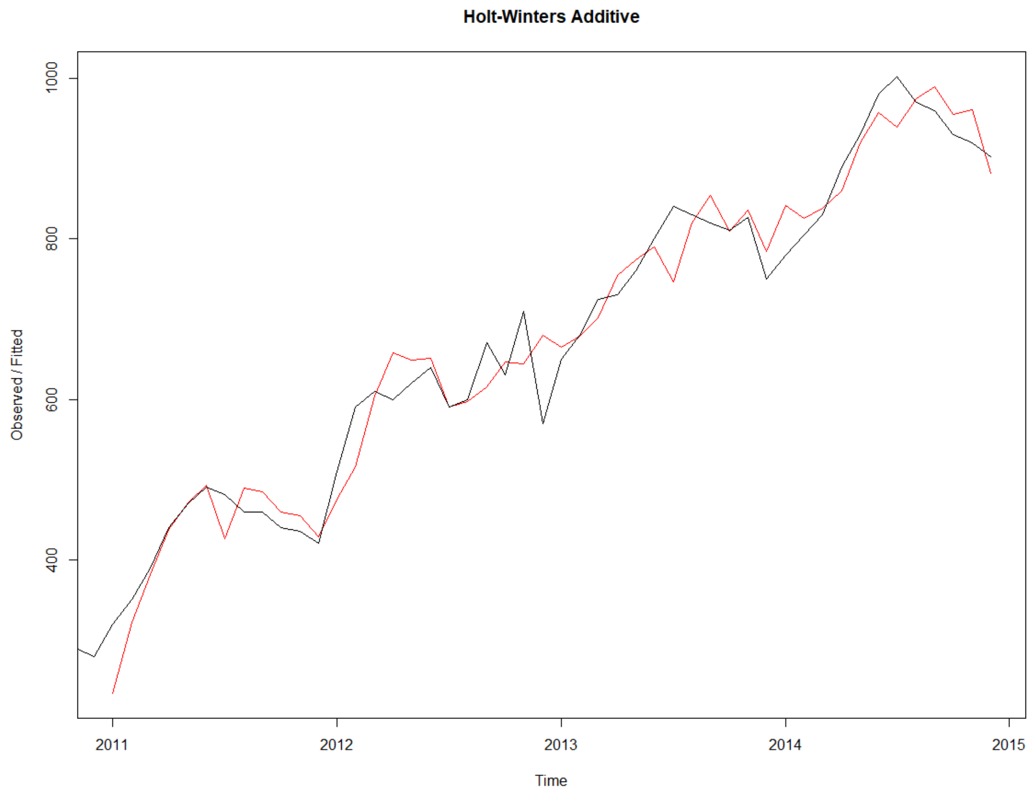
From the seasonal plot, it is not clear of the presence of seasonality. Only the last two years, 2013 and 2014, shows some form of seasonality.

Hence, we can conclude that the unit sales of tractors in South America possess trend but no seasonality or both trend and seasonality.

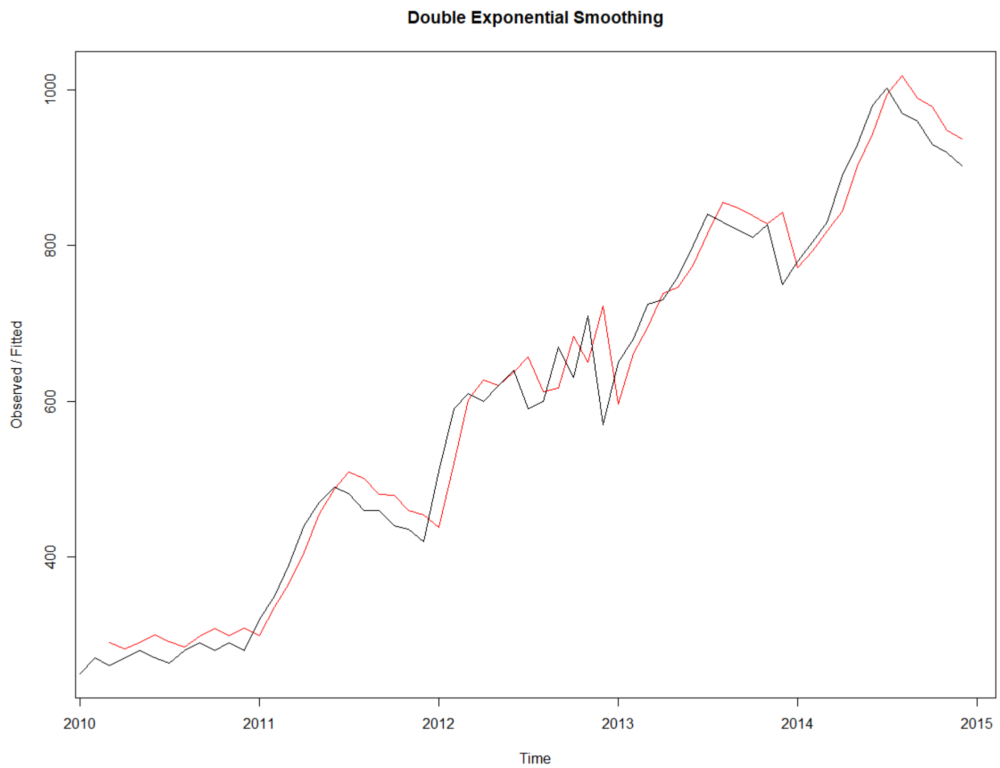
Thus, we can try using Double Exponential smoothing or Holt-Winters additive/multiplicative model.



Graph 123: Observed time series data of mower tractor sales in South America against the fitted Holt-Winters multiplicative model.



Graph 124: Observed time series data of mower tractor sales in South America against the fitted Holt-Winters additive model.



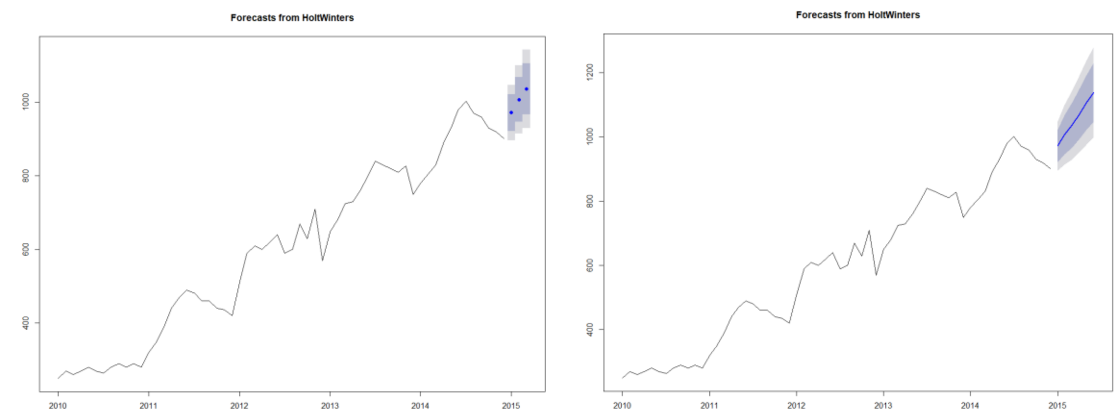
Graph 125: Observed time series data of tractor unit sales in South America against the fitted double exponential smoothing model

From the graphs plotted, the double exponential smoothing model (Graph 125) seems like a better fit. However, we are unable to tell which model would be a better fit solely based on the graph. Hence, we can calculate the error metrics (using the accuracy function) and see which model gives us an overall lower value of error metrics.

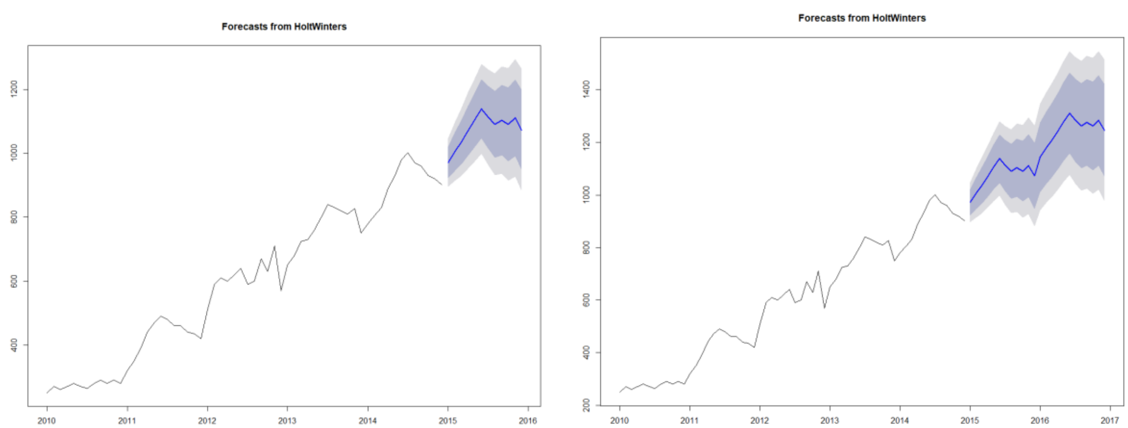
	RMSE	MAE	MAPE
Holt-Winters multiplicative	46.1	34.6	5.33
Double Exponential	39.1	30.1	5.49
Holt-Winters additive	38.2	28.0	0.17

From the table, we can see that using a Holt-Winters additive model will be appropriate because it has lower values of error.

We can make use of Holt-Winters additive model to predict data in the next 3,6,12 and 24 months



Graphs 126 and 127: Forecasted data of unit tractor sales in South America over the next 3 and 6 months respectively.



Graphs 128 and 129: Forecasted data of unit tractor sales in South America over the next 12 and 24 months respectively.

The following are the forecasted values of tractor unit sales in South America over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	971.3986	921.9361	1020.861	895.7521	1047.045
Feb 2015	1006.8554	946.4327	1067.278	914.4468	1099.264
Mar 2015	1035.9686	966.2890	1105.648	929.4028	1142.535
Apr 2015	1070.9949	993.1514	1148.838	951.9436	1190.046
May 2015	1105.8764	1020.6475	1191.105	975.5301	1236.223
Jun 2015	1138.8351	1046.8116	1230.859	998.0973	1279.573
Jul 2015	1112.7421	1014.3924	1211.092	962.3291	1263.155
Aug 2015	1090.6403	986.3473	1194.933	931.1379	1250.143
Sep 2015	1103.8175	993.9021	1213.733	935.7165	1271.919
Oct 2015	1090.5689	975.3051	1205.833	914.2882	1266.850
Nov 2015	1111.1561	990.7813	1231.531	927.0587	1295.253
Dec 2015	1072.7482	947.4707	1198.026	881.1528	1264.344
Jan 2016	1144.1661	1011.1987	1277.134	940.8099	1347.522

Feb 2016	1179.6229	1042.2012	1317.045	969.4546	1389.791
Mar 2016	1208.7361	1067.0002	1350.472	991.9698	1425.502
Apr 2016	1243.7624	1097.8398	1389.685	1020.5930	1466.932
May 2016	1278.6439	1128.6514	1428.636	1049.2501	1508.038
Jun 2016	1311.6026	1157.6476	1465.558	1076.1488	1547.056
Jul 2016	1285.5096	1127.6917	1443.327	1044.1480	1526.871
Aug 2016	1263.4078	1101.8193	1424.996	1016.2796	1510.536
Sep 2016	1276.5850	1111.3119	1441.858	1023.8216	1529.348
Oct 2016	1263.3364	1094.4591	1432.214	1005.0608	1521.612
Nov 2016	1283.9236	1111.5174	1456.330	1020.2510	1547.596
Dec 2016	1245.5157	1069.6514	1421.380	976.5545	1514.477

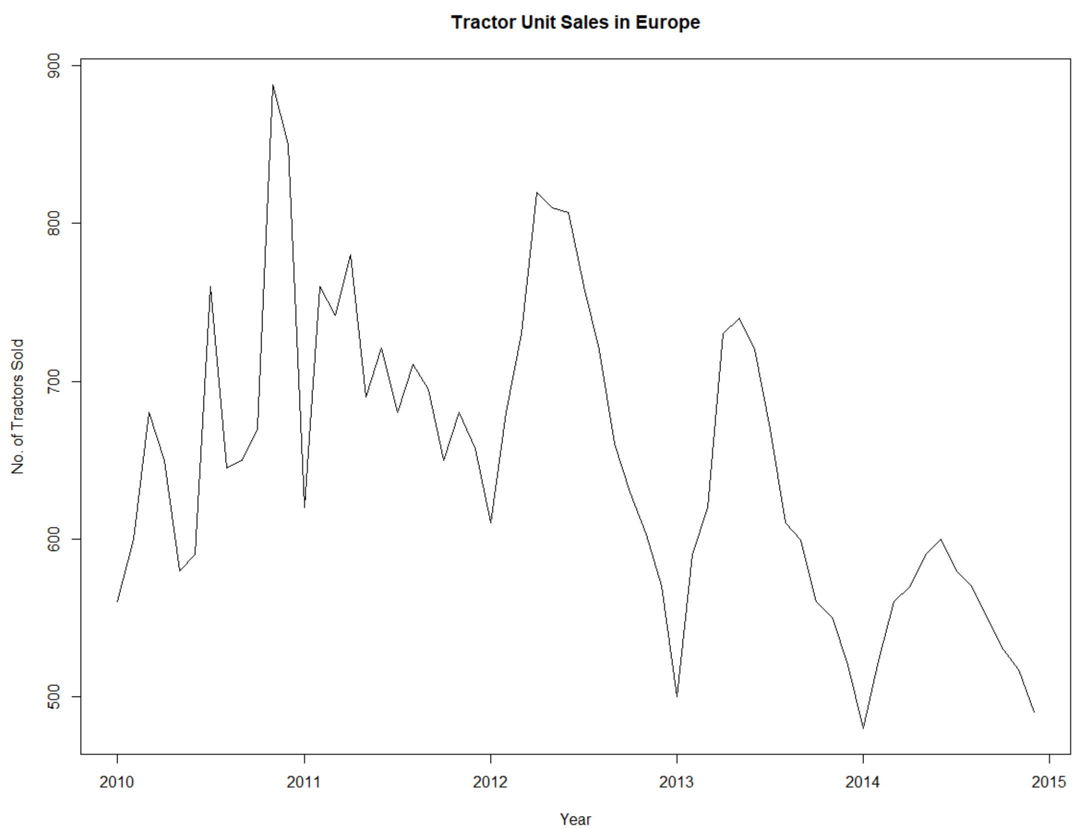
Conclusion:

As seen from the forecasted graphs and values generated, unit tractor sales in South America is predicted to increase gradually for the next 24 months to come.

<u>Forecasting of Unit Sales of Tractors in Europe</u>

<u>Time Series:</u>

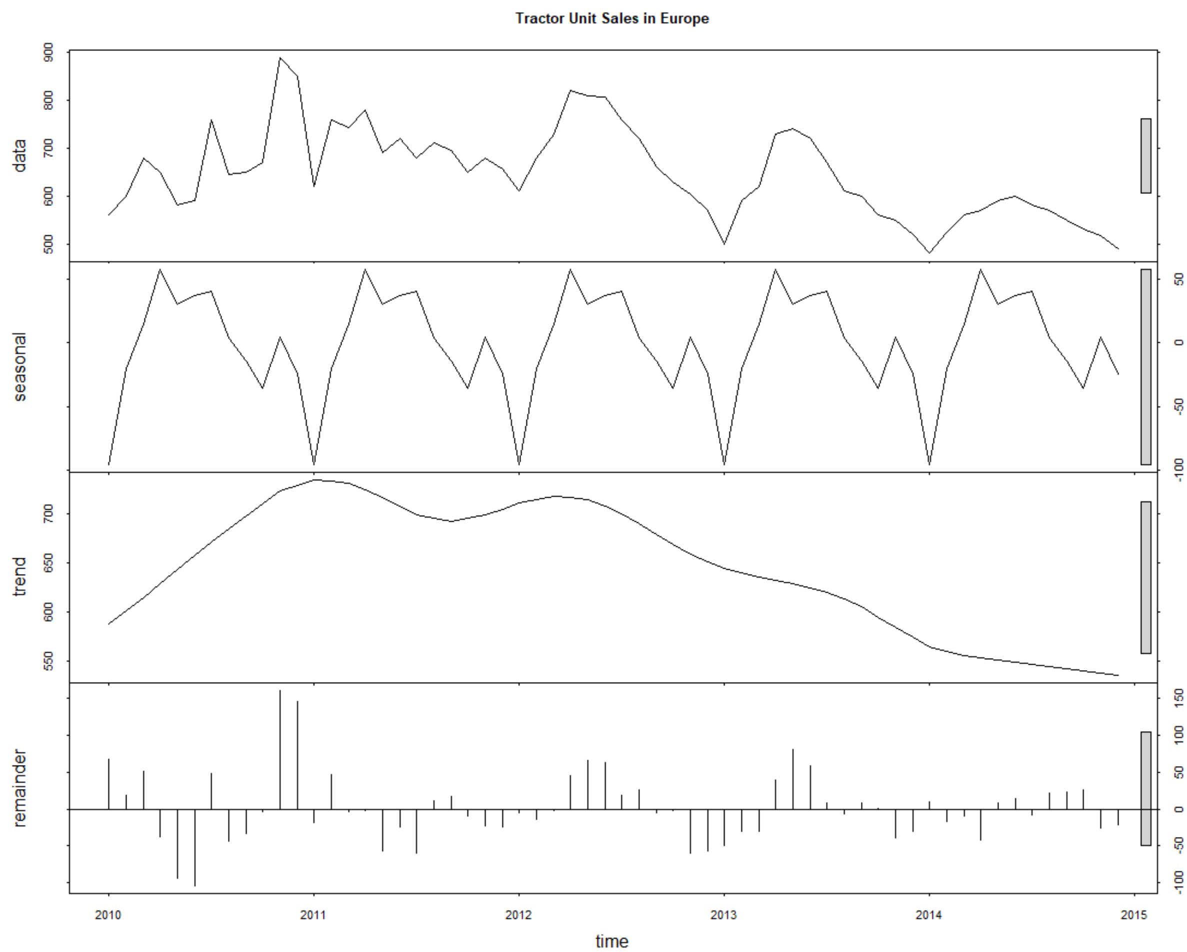
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	560	600	680	650	580	590	760	645	650	670	888	850
2011	620	760	742	780	690	721	680	711	695	650	680	657
2012	610	680	730	820	810	807	760	720	660	630	603	570
2013	500	590	620	730	740	720	670	610	599	560	550	520
2014	480	523	560	570	590	600	580	570	550	530	517	490



Graph 130 : The time series graph above shows the tractor unit sales in Europe from January 2010 to December 2014.

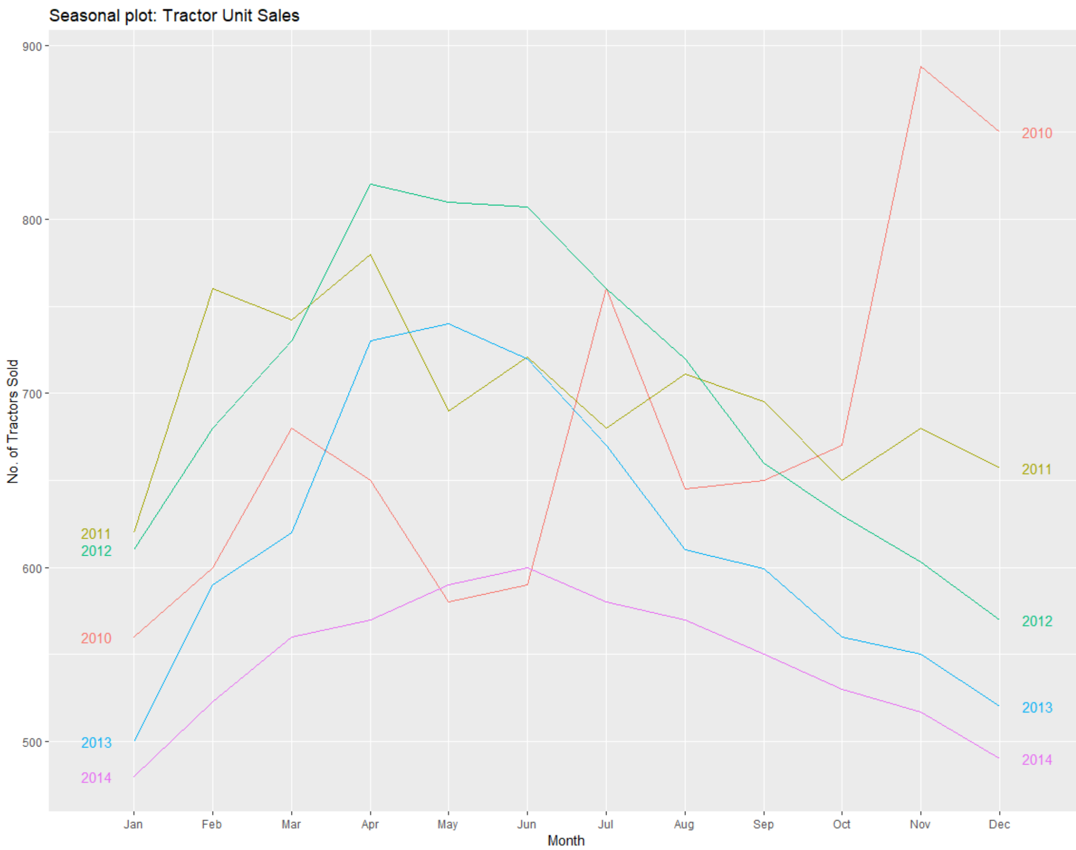
From the graph, there is no clear trend and seasonality in the data over this period.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 131: Decomposed time series data of tractor unit sales in Europe from January 2010 to December 2014

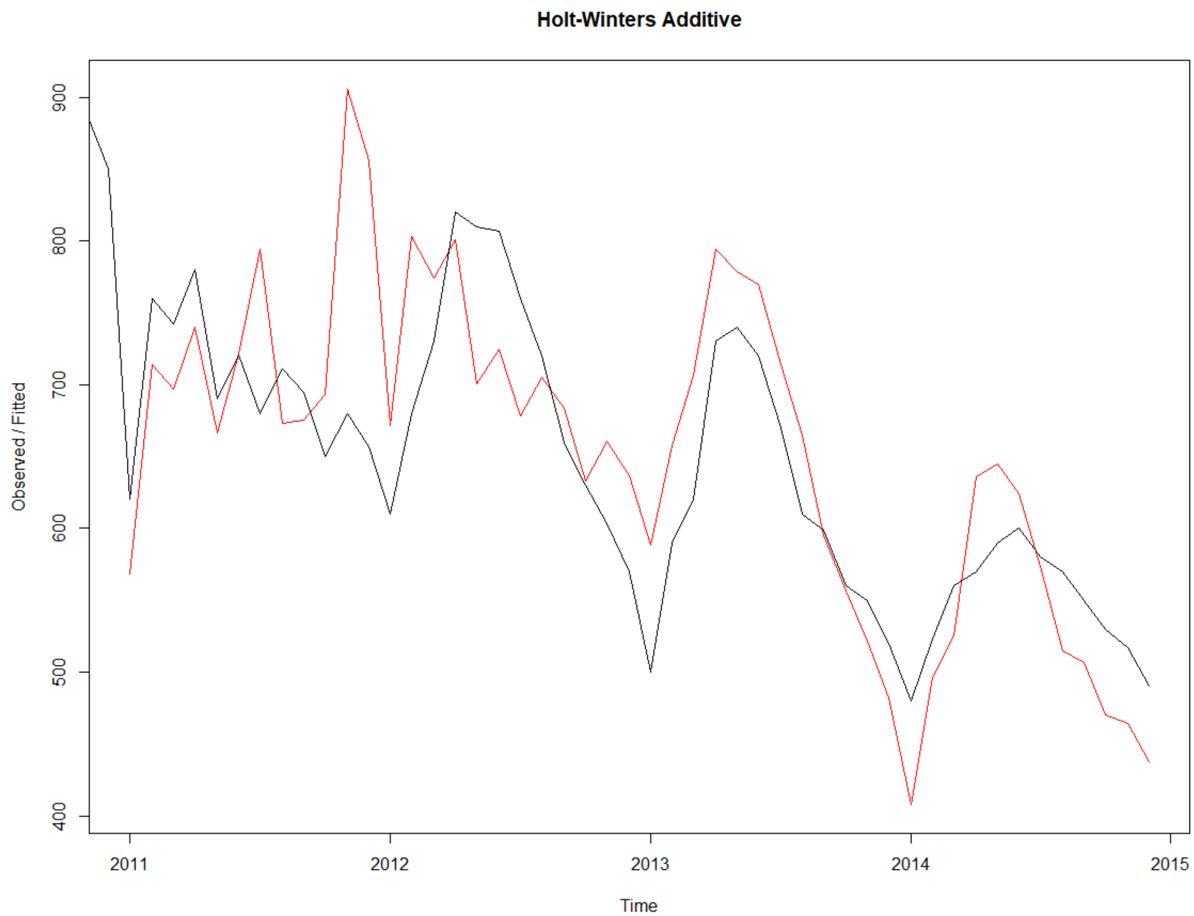
Decomposition of the time series shows how there is no clear overall trend in the data.



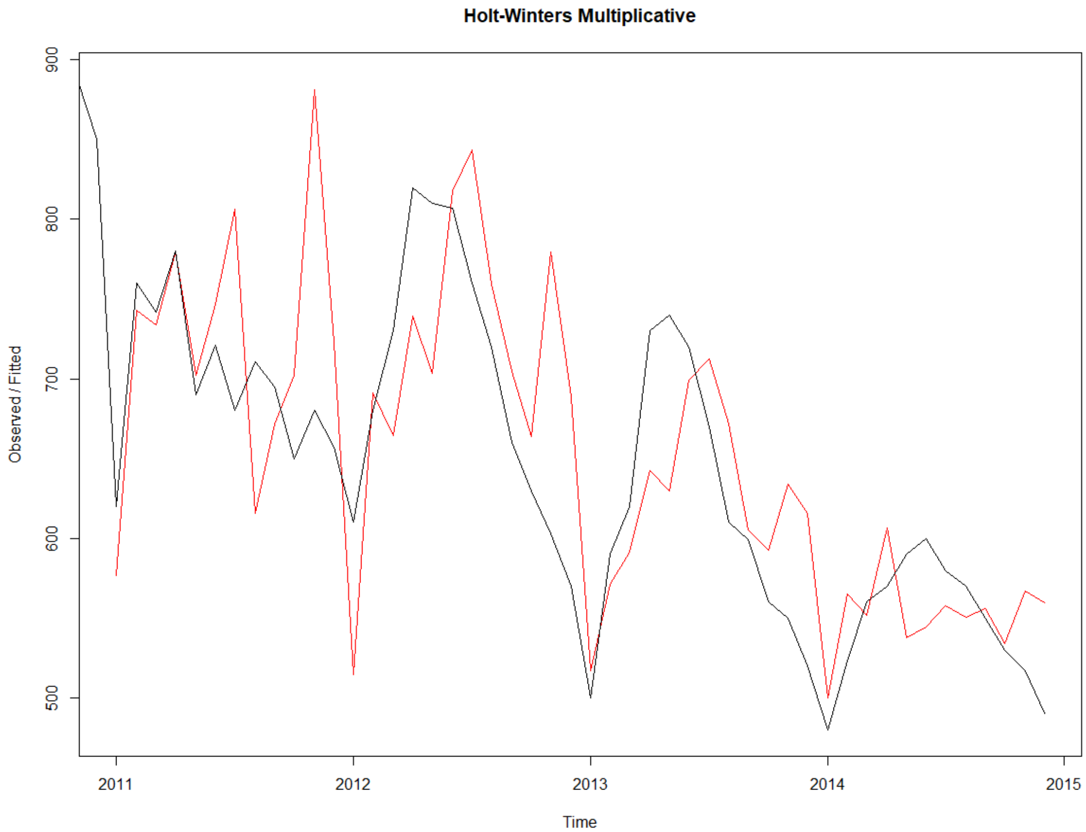
Graph 132: A seasonal plot of tractor unit sales in Europe from January 2010 to December 2014.

From the seasonal plot, it is not clear of the presence of seasonality. However, there is the presence of a possible seasonality in the last three years where sales increased till mid-year before dropping

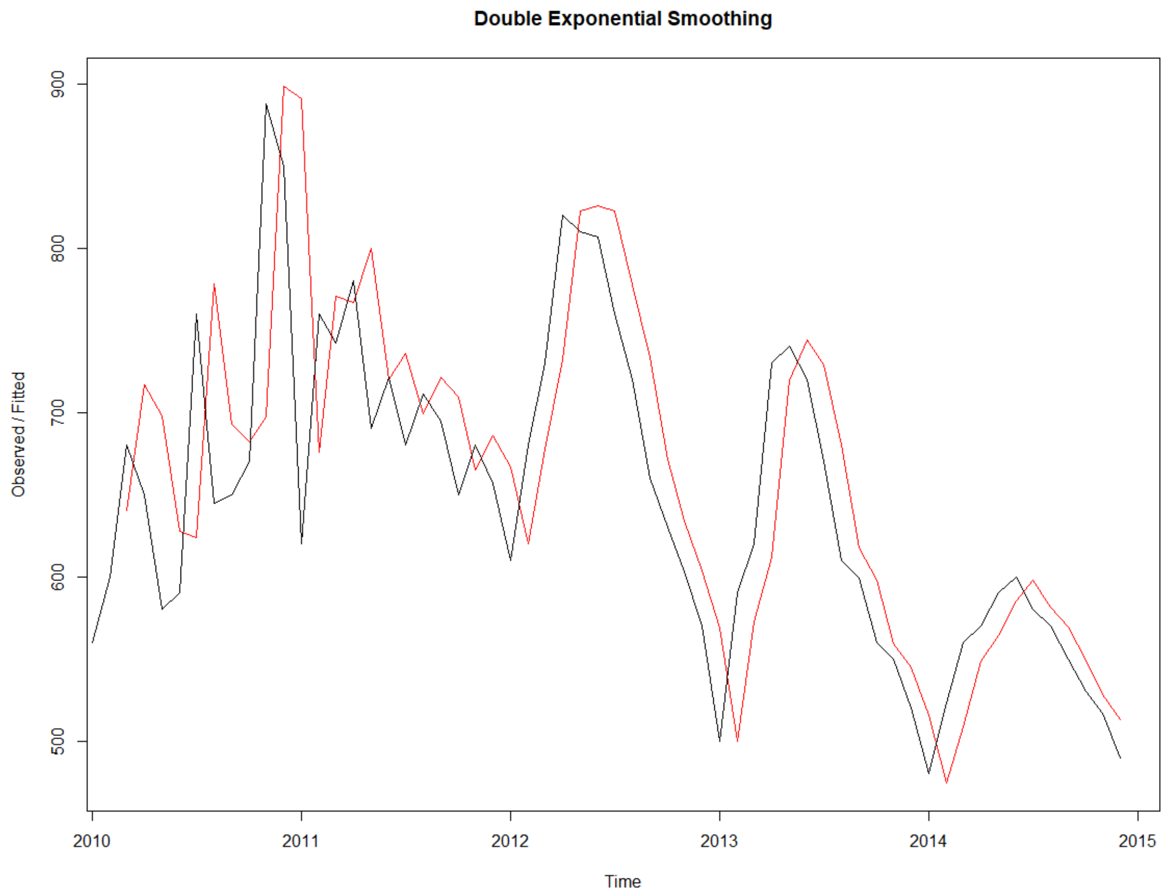
Thus, we can experiment use Simple Exponential smoothing, Double Exponential Smoothing, Holt-Winters no-trend smoothing, additive/multiplicative models.



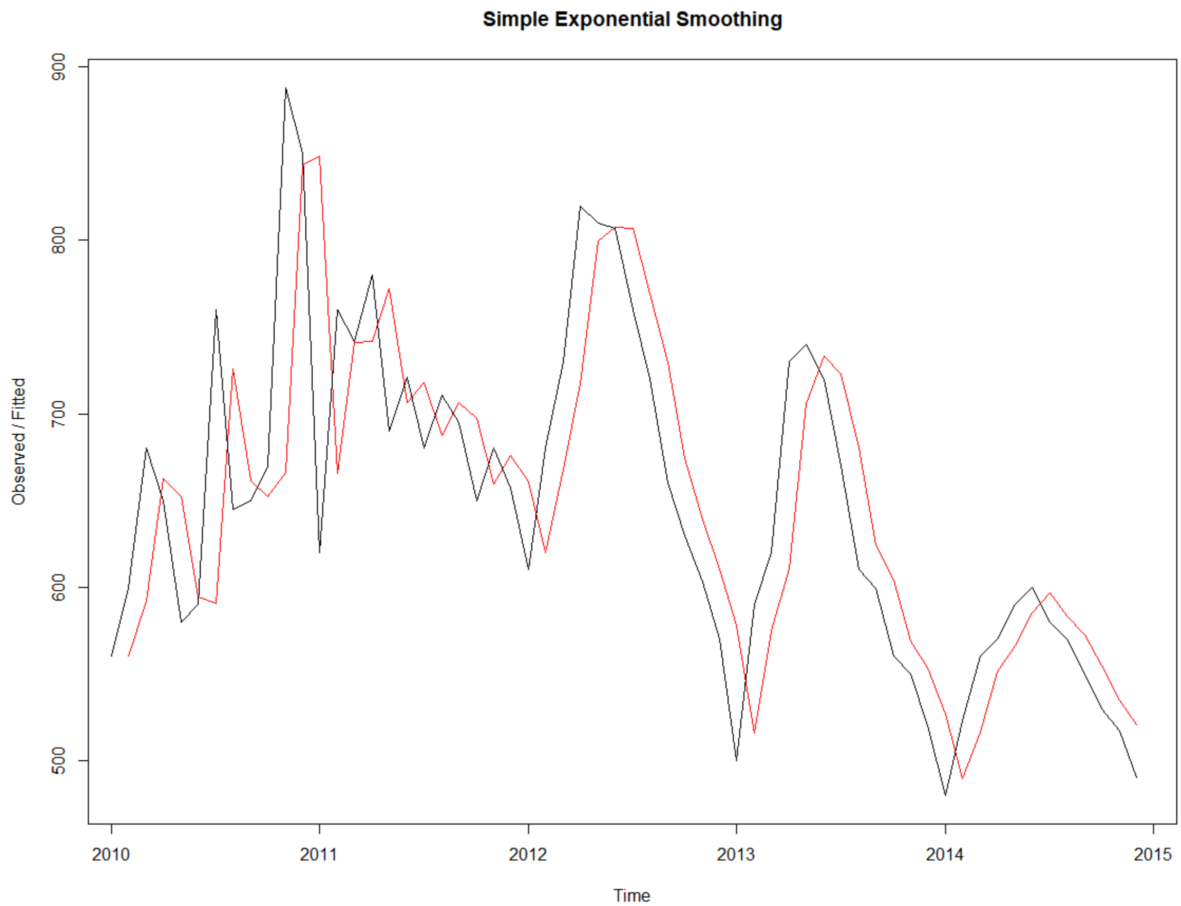
Graph 133: Observed time series data of tractor unit sales in Europe against the fitted Holt-Winters Additive model



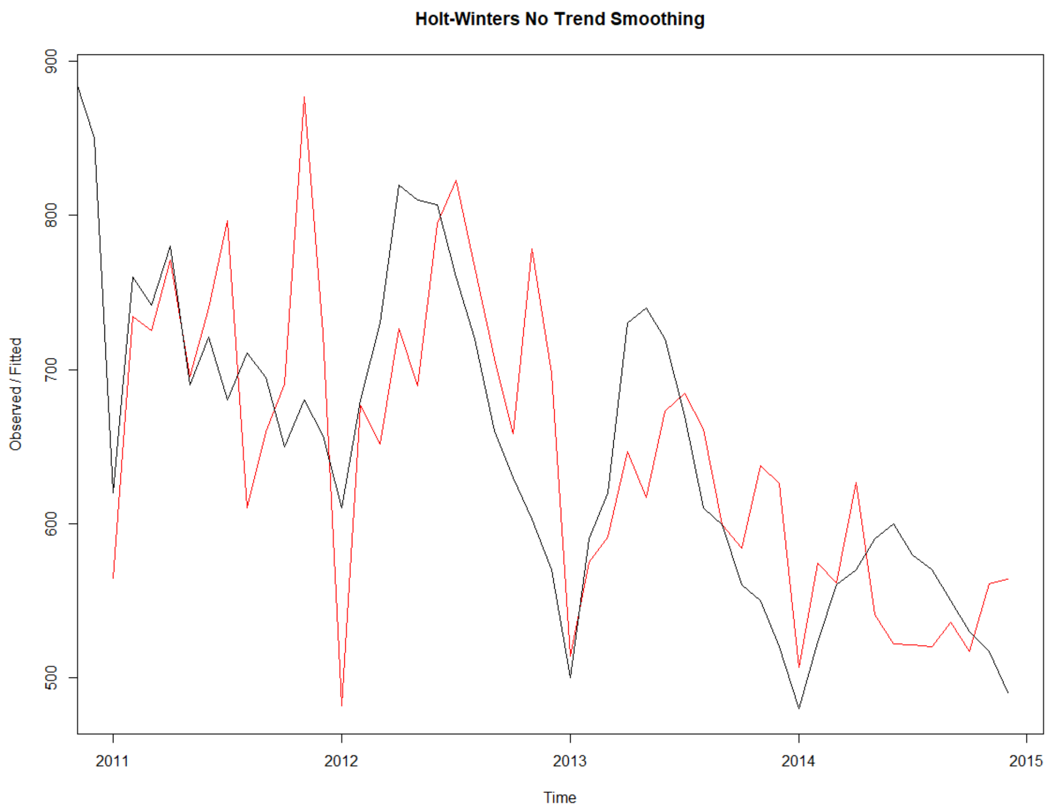
Graph 134: Observed time series data of tractor unit sales in Europe against the fitted Holt-Winters Multiplicative model



Graph 135: Observed time series data of tractor unit sales in Europe against the fitted Double Exponential Smoothing model.



Graph 136: Observed time series data of tractor unit sales in Europe against the fitted Simple Exponential smoothing model



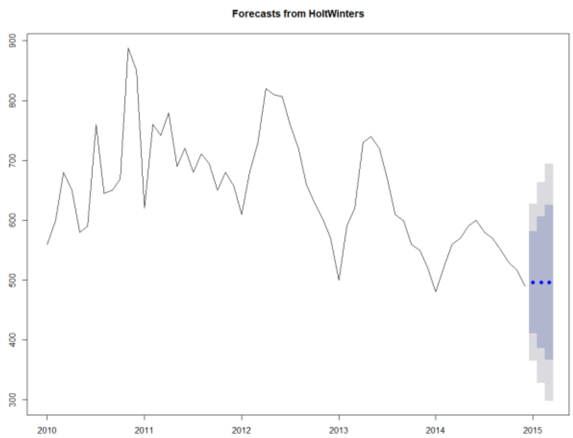
Graph 137: Observed time series data of tractor unit sales in Europe against the fitted Holt-Winters no-trend smoothing model

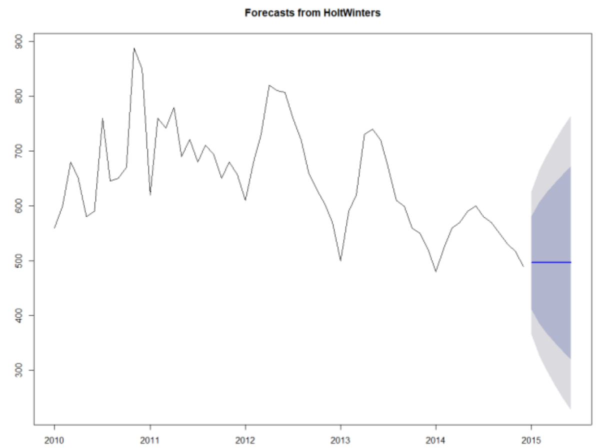
From the graphs plotted, the double and simple exponential smoothing models seem like a better fit. However, we are unable to tell which model would be a better fit solely based on the graphs. Hence, we can calculate the error metrics (using the accuracy function) and see which model gives us an overall lower value of error metrics.

	RMSE	MAE	MAPE
Holt-Winters multiplicative	68.7	52.6	8.27
Double Exponential	70.1	51.8	7.9
Holt-Winters additive	70.1	55.2	8.70
Simple Exponential	66.3	47.9	7.31
Holt-Winters no-trend	72.5	56.7	8.93

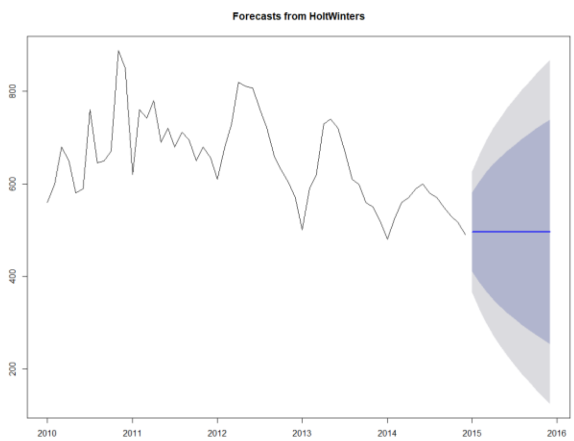
From the table, we can see that using a Simple Exponential model will be appropriate because it has lower values of error.

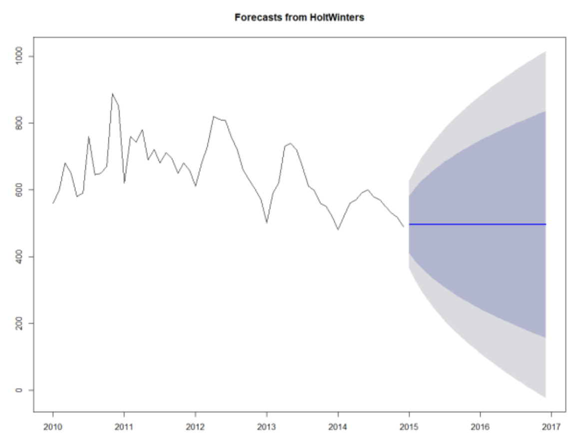
We can make use of Simple Exponential Smoothing to predict data in the next 3,6,12 and 24 months





Graphs 138 and 139: Forecasted data of unit tractor sales in Europe over the next 3 and 6 months respectively.





Graphs 140 and 141: Forecasted data of unit tractor sales in Europe over the next 12 and 24 months respectively.

The following are the forecasted values of tractor unit sales in Europe over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	496.1355	410.4448	581.8262	365.082881	627.1881
Feb 2015	496.1355	386.4269	605.8441	328.350698	663.9203
Mar 2015	496.1355	366.7947	625.4763	298.325767	693.9452
Apr 2015	496.1355	349.7725	642.4985	272.292607	719.9784
May 2015	496.1355	334.5335	657.7375	248.986574	743.2844
Jun 2015	496.1355	320.6126	671.6584	227.696416	764.5746
Jul 2015	496.1355	307.7175	684.5535	207.974968	784.2960
Aug 2015	496.1355	295.6500	696.6210	189.519383	802.7516
Sep 2015	496.1355	284.2688	708.0022	172.113289	820.1577
Oct 2015	496.1355	273.4685	718.8025	155.595717	836.6753
Nov 2015	496.1355	263.1684	729.1026	139.843072	852.4279
Dec 2015	496.1355	253.3048	738.9662	124.758004	867.5130
Jan 2016	496.1355	243.8265	748.4445	110.262211	882.0088

Feb 2016	496.1355	234.6916	757.5794	96.291599	895.9794
Mar 2016	496.1355	225.8653	766.4057	82.792911	909.4781
Apr 2016	496.1355	217.3183	774.9527	69.721328	922.5497
May 2016	496.1355	209.0255	783.2455	57.038704	935.2323
Jun 2016	496.1355	200.9657	791.3053	44.712254	947.5587
Jul 2016	496.1355	193.1202	799.1508	32.713557	959.5574
Aug 2016	496.1355	185.4727	806.7983	21.017780	971.2532
Sep 2016	496.1355	178.0090	814.2619	9.603077	982.6679
Oct 2016	496.1355	170.7165	821.5545	-1.549891	993.8209
Nov 2016	496.1355	163.5839	828.6871	-12.458345	1004.7293
Dec 2016	496.1355	156.6010	835.6700	-23.137693	1015.4087

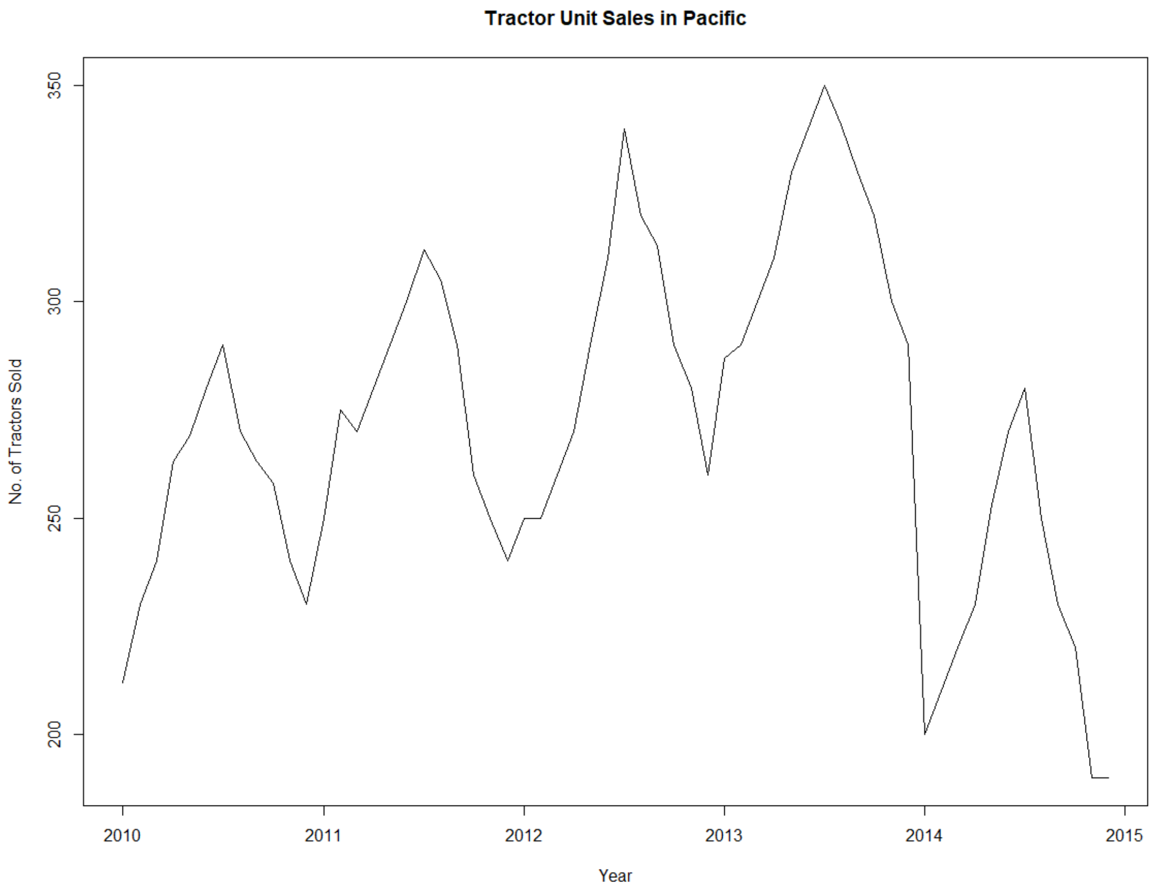
Conclusion:

As seen from the forecasted graphs and values generated, unit tractor sales in Europe is predicted to remain stagnant at 496 for the next 24 months to come.

Forecasting of Unit Sales of Tractors in Pacific

Time Series:

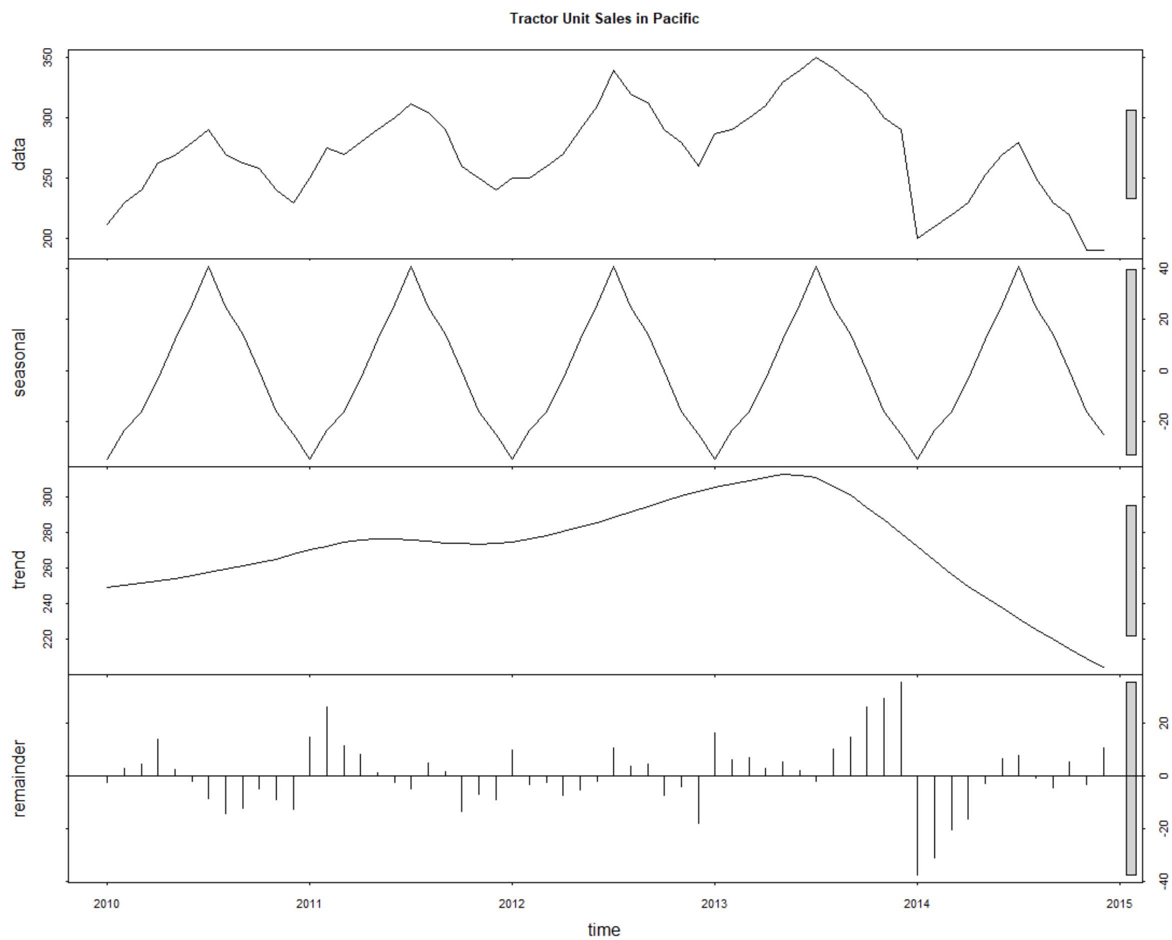
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	212	230	240	263	269	280	290	270	263	258	240	230
2011	250	275	270	280	290	300	312	305	290	260	250	240
2012	250	250	260	270	290	310	340	320	313	290	280	260
2013	287	290	300	310	330	340	350	341	330	320	300	290
2014	200	210	220	230	253	270	280	250	230	220	190	190



Graph 143: The time series graph above shows the tractor unit sales in Pacific from January 2010 to December 2014.

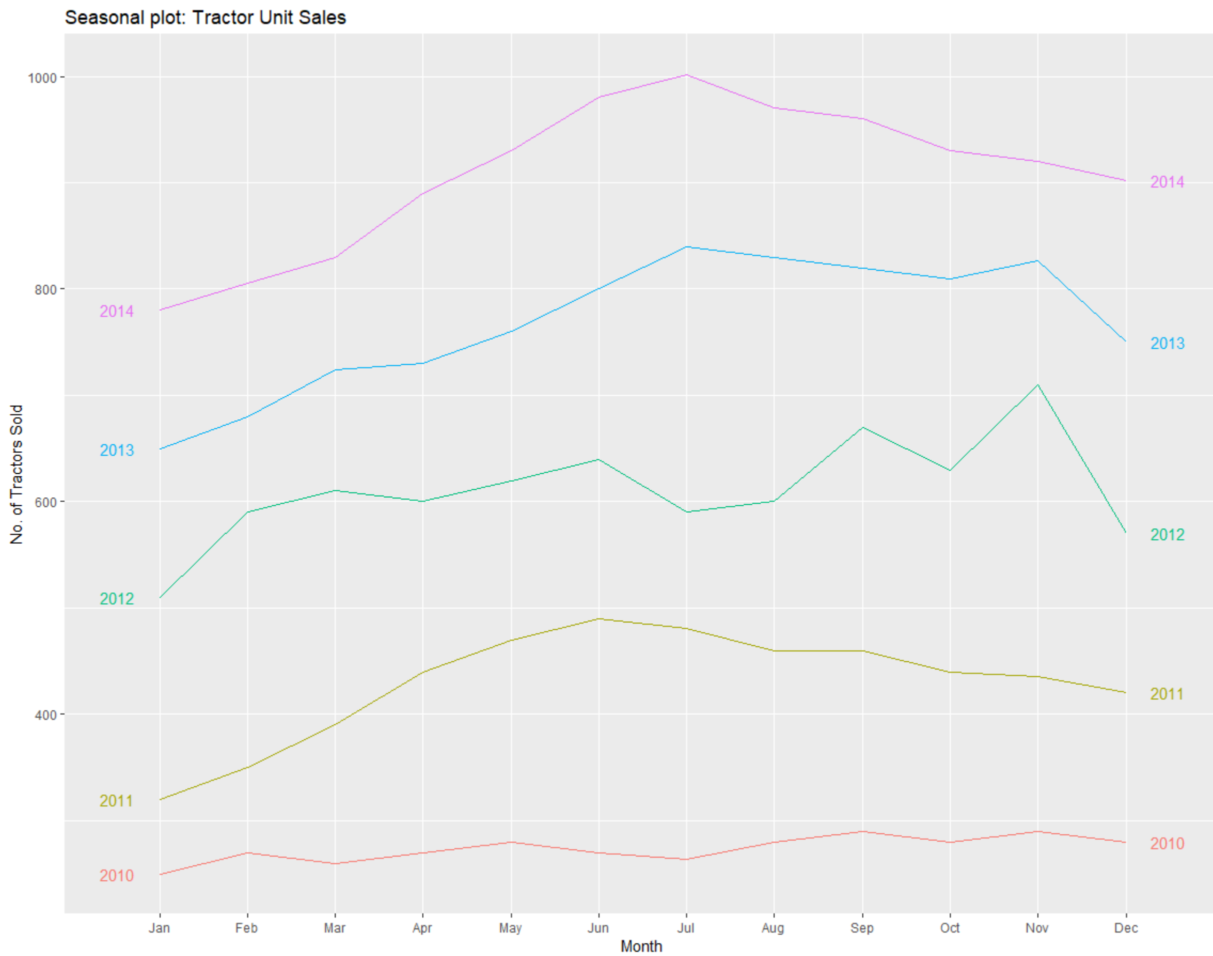
From the graph, there is no clear trend in the data over this period.
However, there is a clear presence of seasonality due the constant fluctuation in sales every middle of the year.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 144: The time series graph above shows the tractor unit sales in Pacific from January 2010 to December 2014.

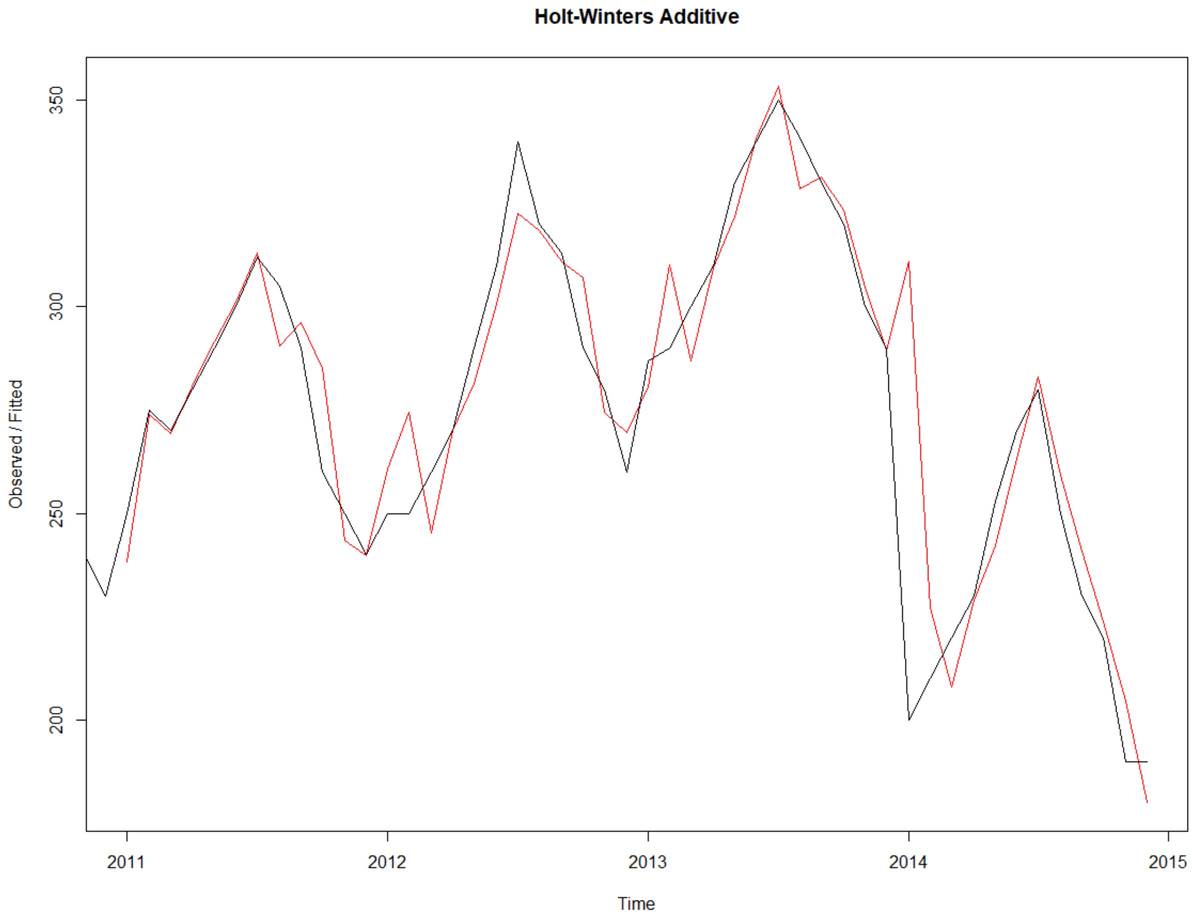
Decomposition of the time series shows segmented trends: an upward trend till around 2013 before going into a decreasing trend. Hence, we cannot be certain if the data overall has an increasing trend with an anomaly in the last two years or there is no trend at all



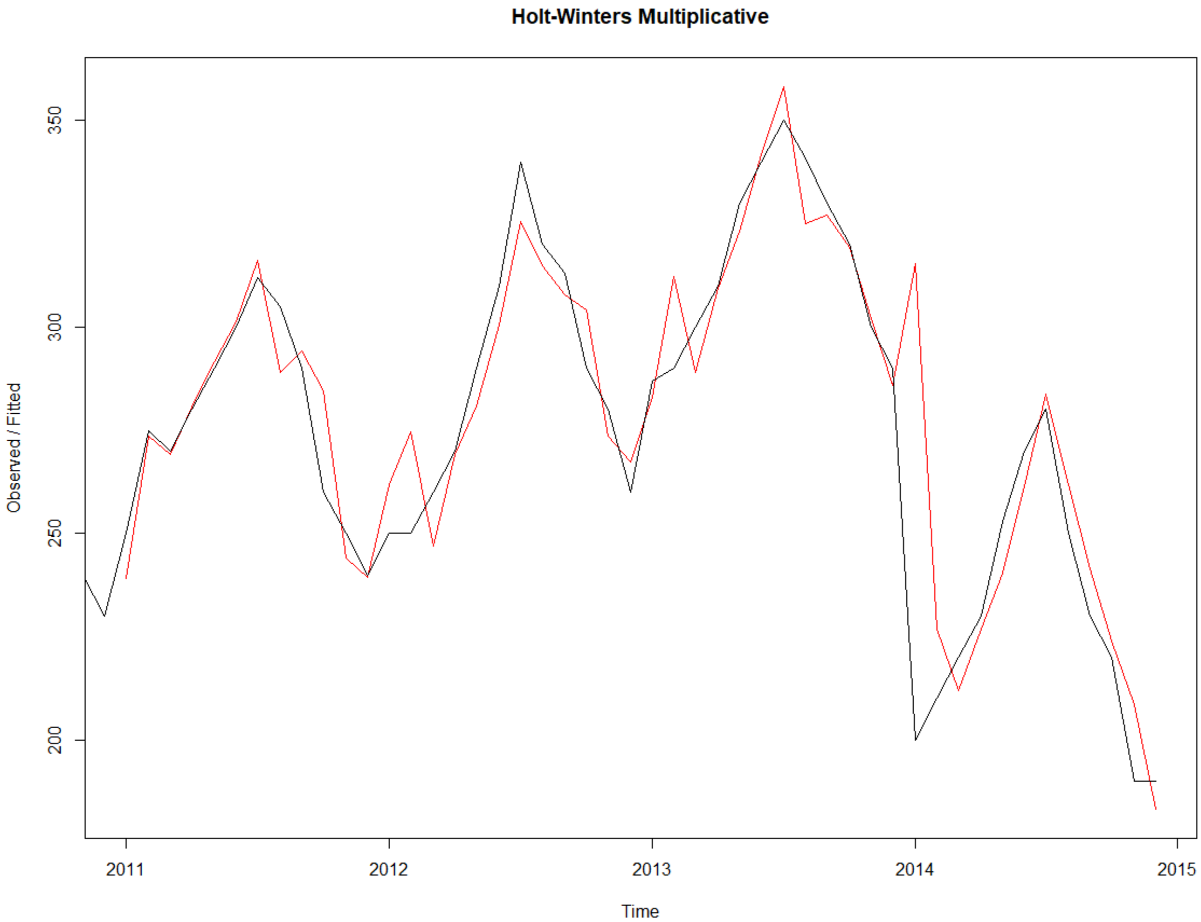
Graph 145: A seasonal plot of tractor unit sales in Pacific from January 2010 to December 2014.

From the seasonal plot, it is not clear of the presence of seasonality.

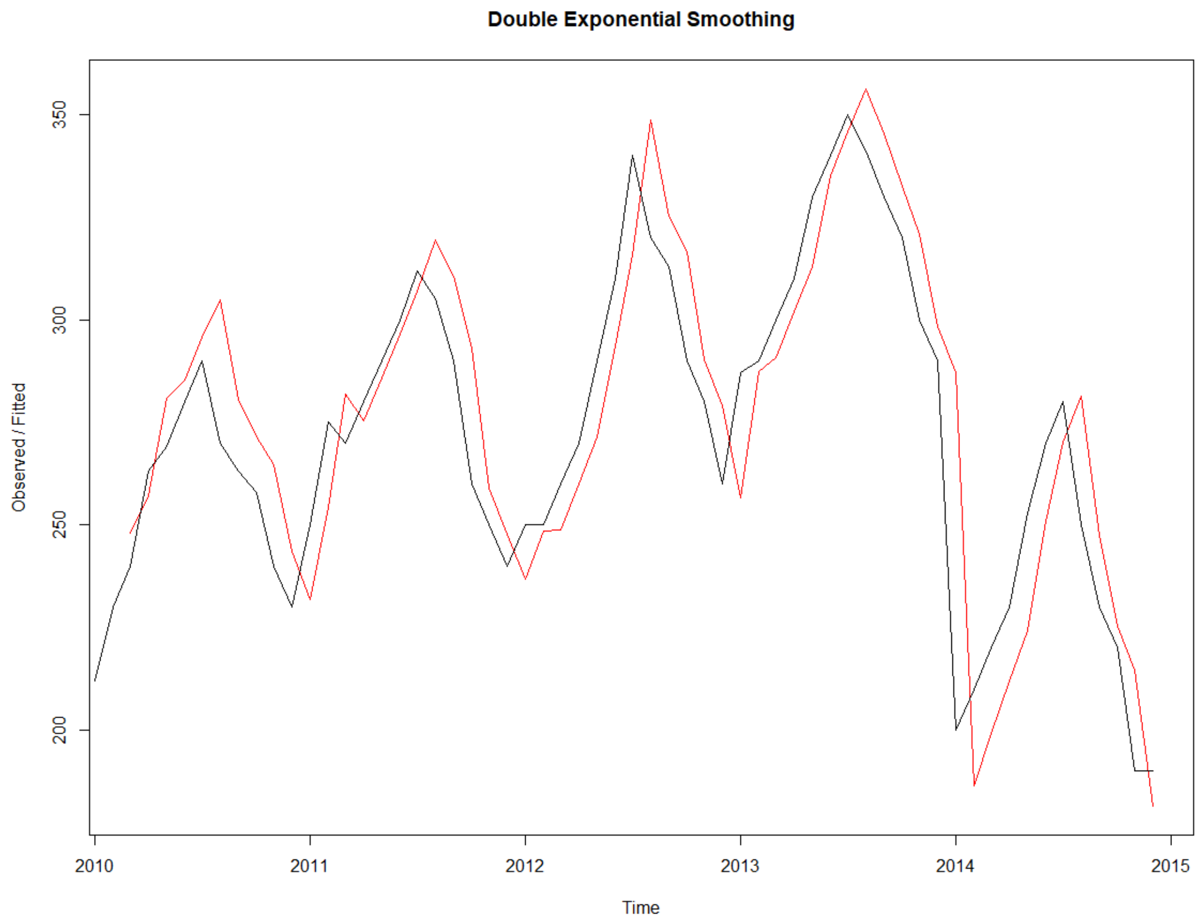
Thus, we can experiment use Simple Exponential smoothing, Double Exponential Smoothing, Holt-Winters no-trend smoothing, additive/multiplicative models.



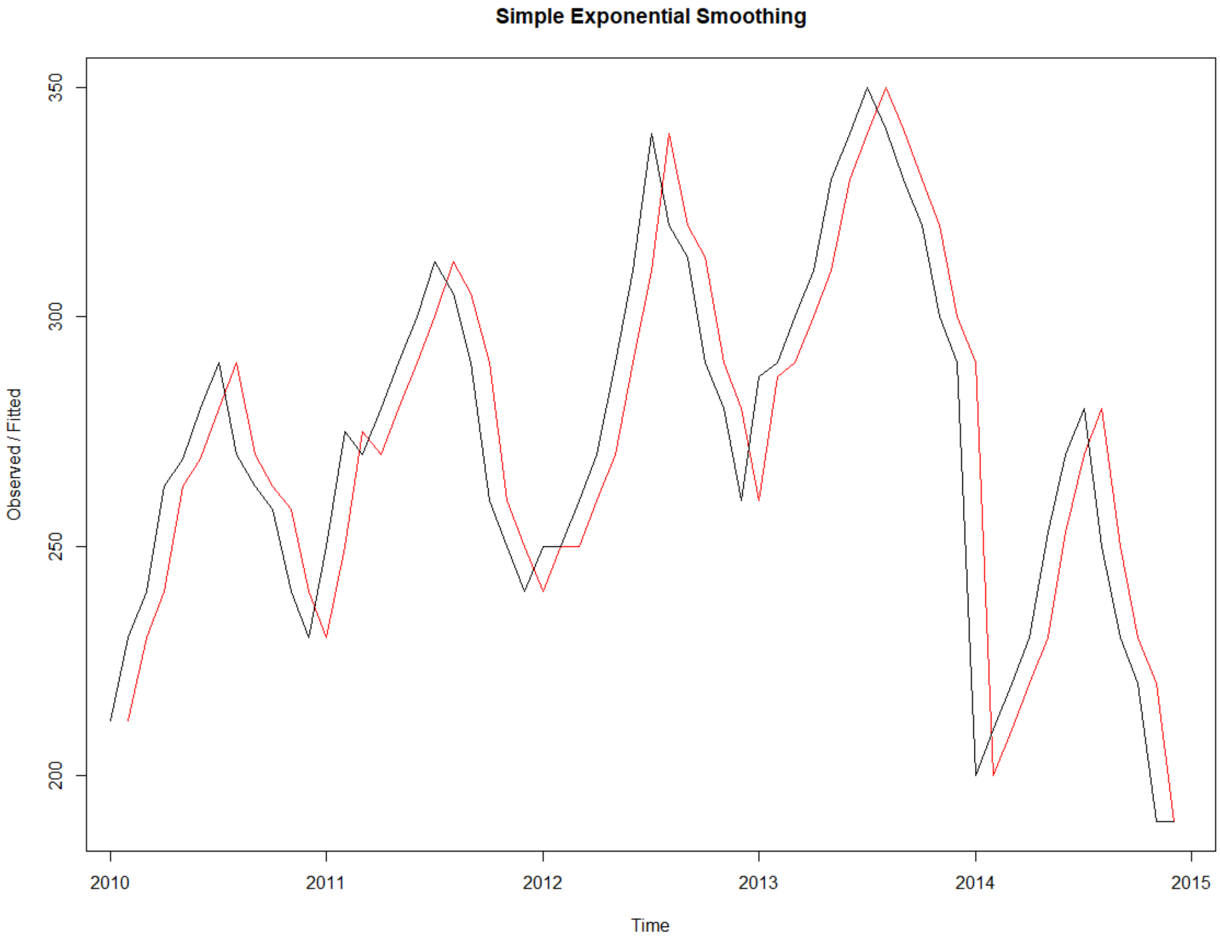
Graph 146: Observed time series data of tractor unit sales in Pacific against the fitted Holt-Winters Additive model



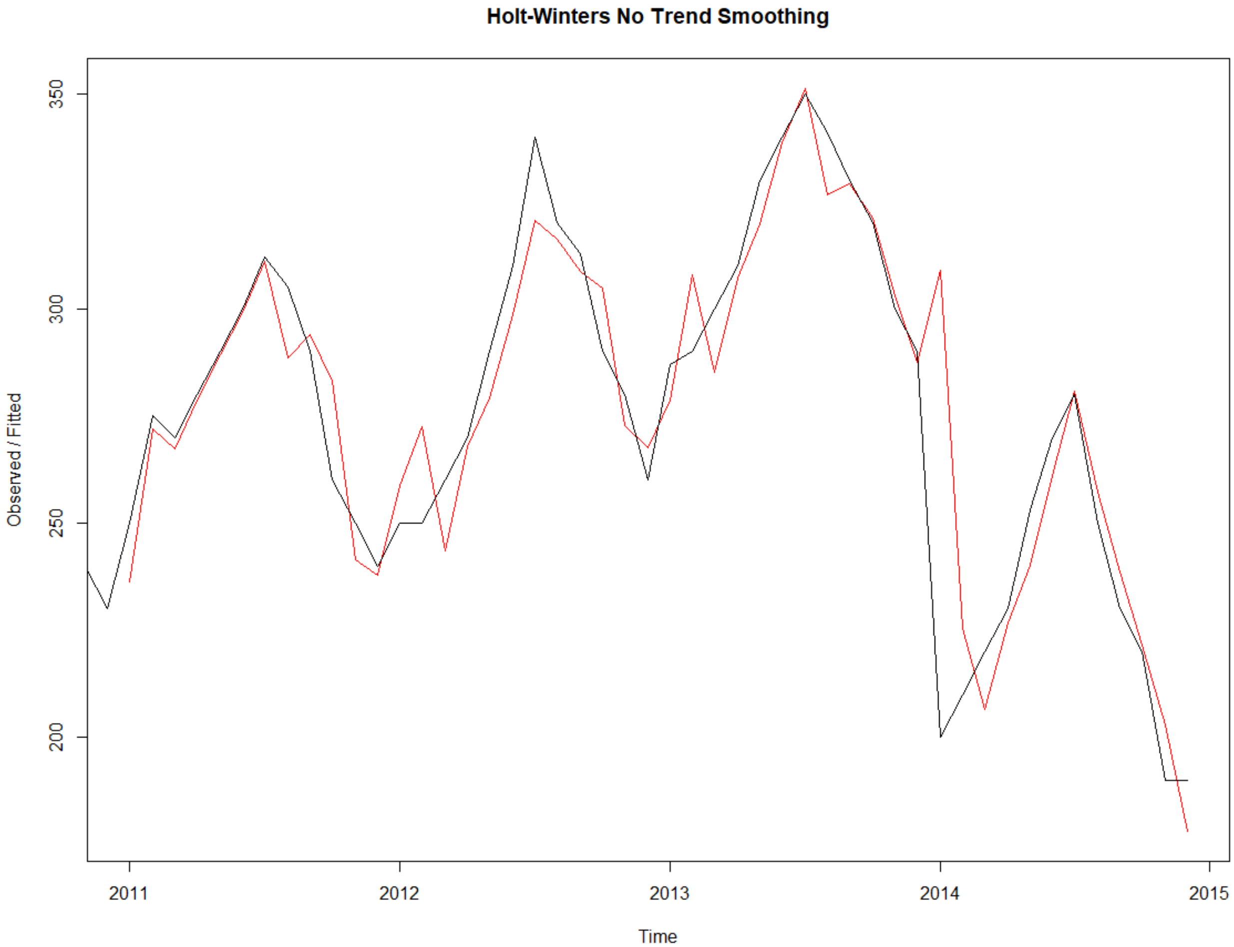
Graph 147: Observed time series data of tractor unit sales in Pacific against the fitted Holt-Winters Multiplicative model



Graph 148: Observed time series data of tractor unit sales in Pacific against the fitted Double Exponential Smoothing model.



Graph 149: Observed time series data of tractor unit sales in Pacific against the fitted Simple Exponential smoothing model



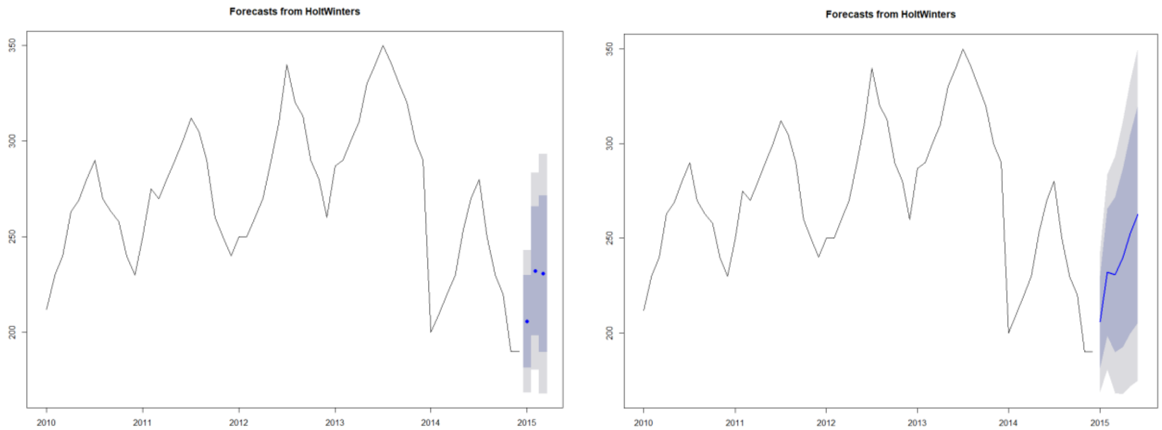
Graph 150: Observed time series data of tractor unit sales in Pacific against the fitted Holt-Winters No Trend Smoothing model

From the graphs plotted, we are unable to tell which model would be a better fit solely based on the graphs. Hence, we can calculate the error metrics (using the accuracy function) and see which model gives us an overall lower value of error metrics.

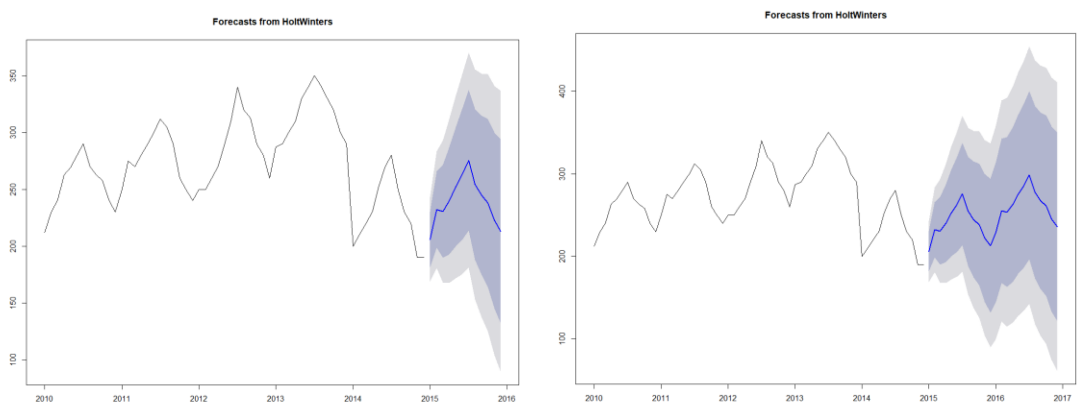
	RMSE	MAE	MAPE
Holt-Winters multiplicative	19.5	10.3	4.19
Double Exponential	20.4	16.0	6.19
Holt-Winters additive	9.92	4.04	0.25
Simple Exponential	19.4	15.1	5.80
Holt-Winters no-trend	18.7	10.2	4.14

From the table, we can see that using a Holt Winters Additive model will be appropriate because it has lower values of error.

We can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months.



Graphs 151 and 152: Forecasted data of unit tractor sales in Pacific over the next 3 and 6 months respectively.



Graphs 153 and 154: Forecasted data of unit tractor sales in Pacific over the next 12 and 24 months respectively.

The following are the forecasted values of tractor unit sales in Pacific over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	205.7989	181.5087	230.0891	168.65026	242.9475
Feb 2015	232.1058	198.4650	265.7467	180.65662	283.5550
Mar 2015	230.6644	189.7572	271.5716	168.10222	293.2265
Apr 2015	239.7000	192.6353	286.7648	167.72072	311.6794
May 2015	252.1753	199.6703	304.6804	171.87576	332.4749
Jun 2015	262.7289	205.2966	320.1613	174.89370	350.5641
Jul 2015	275.5465	213.5774	337.5156	180.77292	370.3200
Aug 2015	254.3803	188.1846	320.5759	153.14280	355.6178
Sep 2015	244.6733	174.5053	314.8414	137.36055	351.9861
Oct 2015	238.2300	164.3027	312.1573	125.16792	351.2920
Nov 2015	222.3152	144.8107	299.8196	103.78233	340.8480
Dec 2015	212.8444	131.9208	293.7680	89.08241	336.6064
Jan 2016	228.6433	144.1528	313.1338	99.42622	357.8604
Feb 2016	254.9502	167.3127	342.5878	120.92020	388.9803
Mar 2016	253.5088	162.8334	344.1842	114.83270	392.1849
Apr 2016	262.5444	168.9297	356.1592	119.37303	405.7159
May 2016	275.0198	178.5552	371.4843	127.48990	422.5496
Jun 2016	285.5733	186.3407	384.8059	133.81017	437.3365

```
Jul 2016       298.3909 196.4654 400.3164 142.50933 454.2724
Aug 2016       277.2247 172.6757 381.7737 117.33079 437.1186
Sep 2016       267.5177 160.4094 374.6260 103.70972 431.3257
Oct 2016       261.0744 151.4665 370.6822  93.44366 428.7051
Nov 2016       245.1596 133.1079 357.2112  73.79136 416.5278
Dec 2016       235.6888 121.2455 350.1321  60.66292 410.7147
```

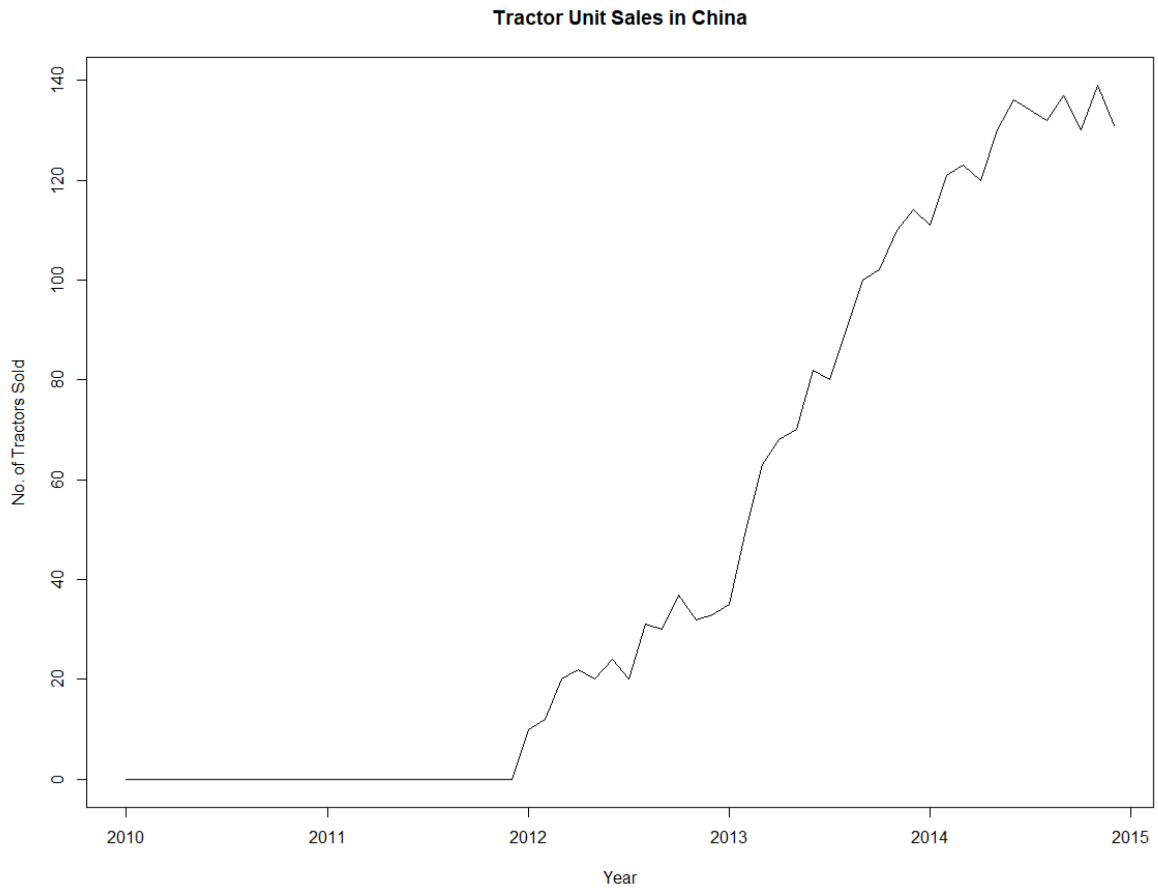
Conclusion:

As seen from the forecasted graphs and values generated, unit tractor sales in Pacific is predicted to remain around the same values for the next 24 months to come. Seasonality will remain.

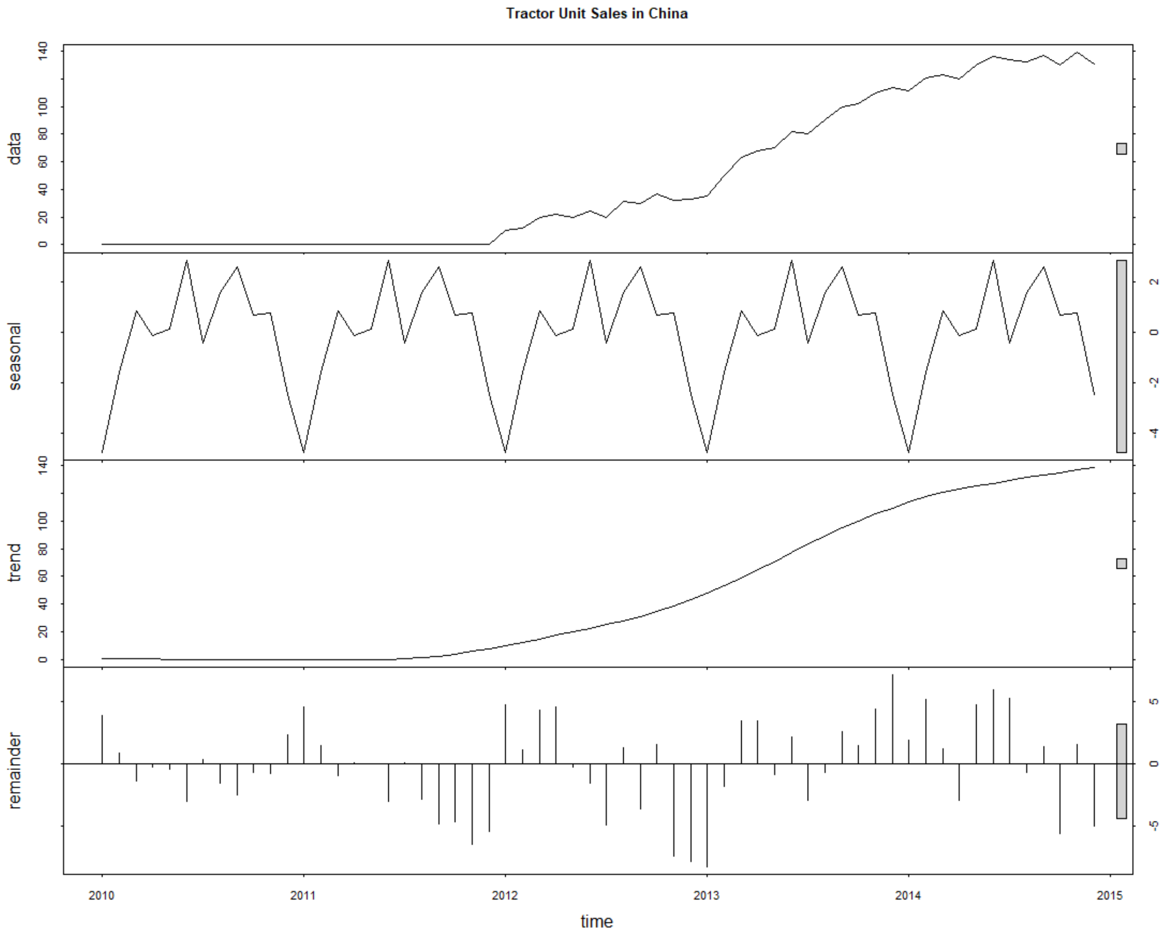
Forecasting of Unit Sales of Tractors in China:

Time Series:

```
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2010   0   0   0   0   0   0   0   0   0   0   0   0
2011   0   0   0   0   0   0   0   0   0   0   0   0
2012  10  12  20  22  20  24  20  31  30  37  32  33
2013  35  50  63  68  70  82  80  90 100 102 110 114
2014 111 121 123 120 130 136 134 132 137 130 139 131
```



Graph 155: The time series graph above shows the tractor unit sales in China from January 2010 to December 2014.



Graph 156: The time series graph above shows the tractor unit sales in China from January 2010 to December 2014.

Due to the lack of sufficient data, it is impossible to deduce whether there is a trend and/or seasonality in the data.

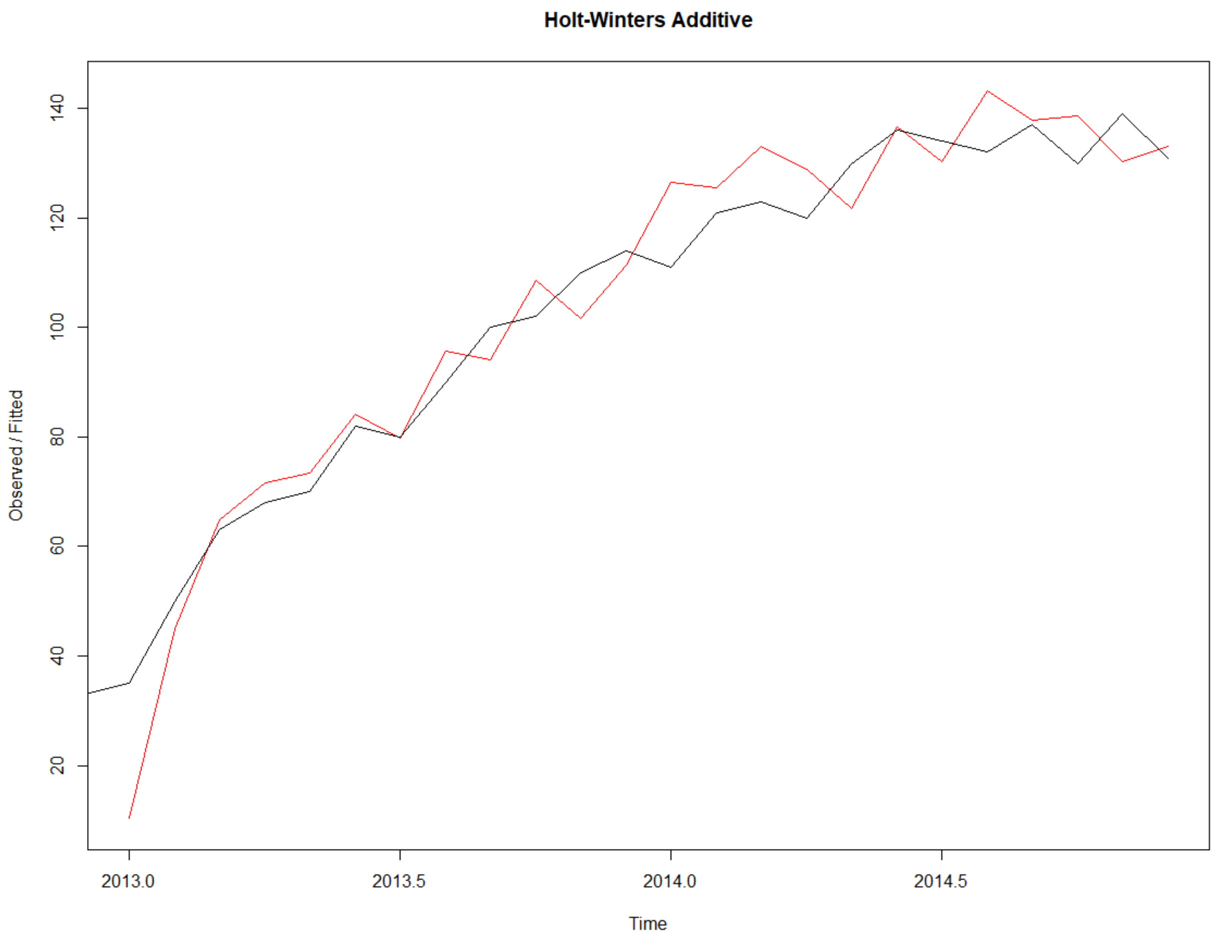
Thus, we can must try out different forecast methods to deduce which is the most appropriate.

We subset the data to after January 2012.

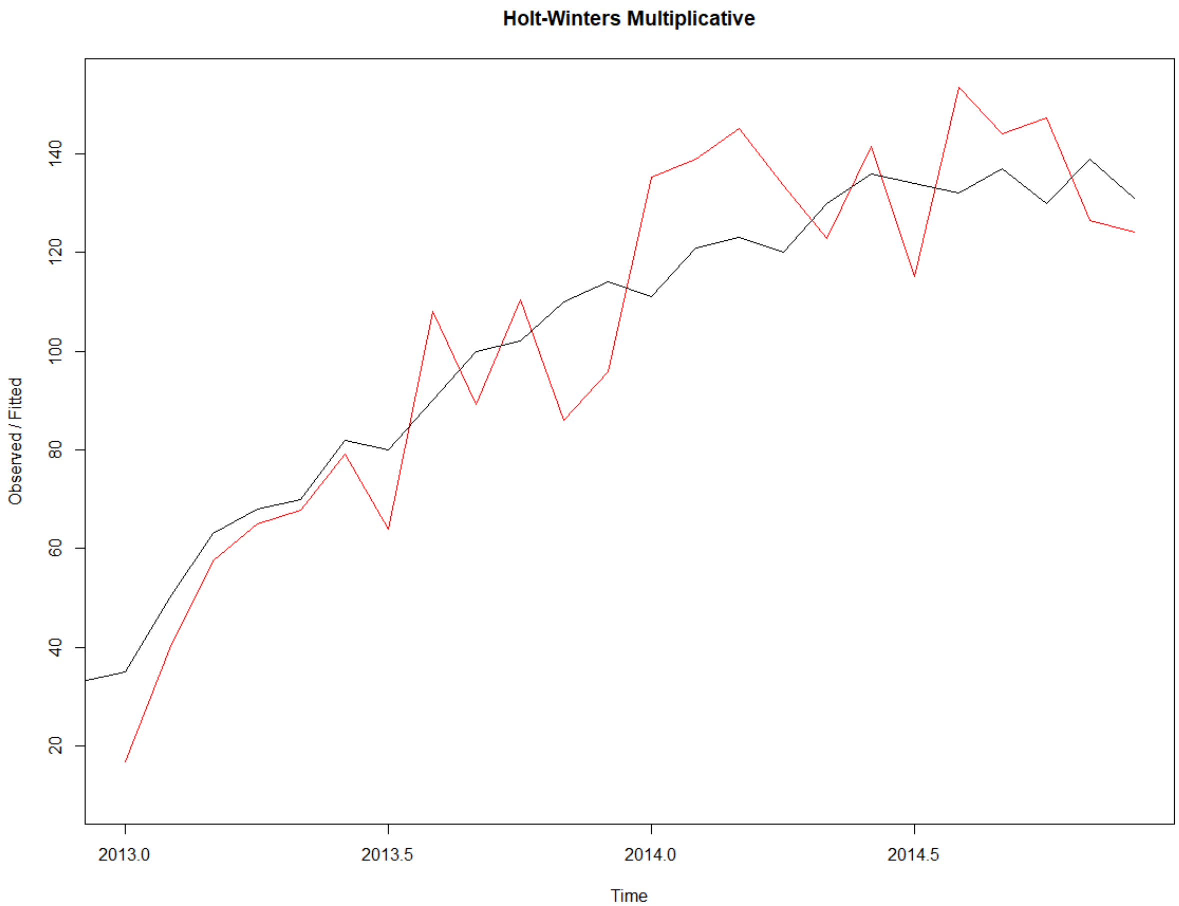
Time series:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	10	12	20	22	20	24	20	31	30	37	32	33
2013	35	50	63	68	70	82	80	90	100	102	110	114
2014	111	121	123	120	130	136	134	132	137	130	139	131

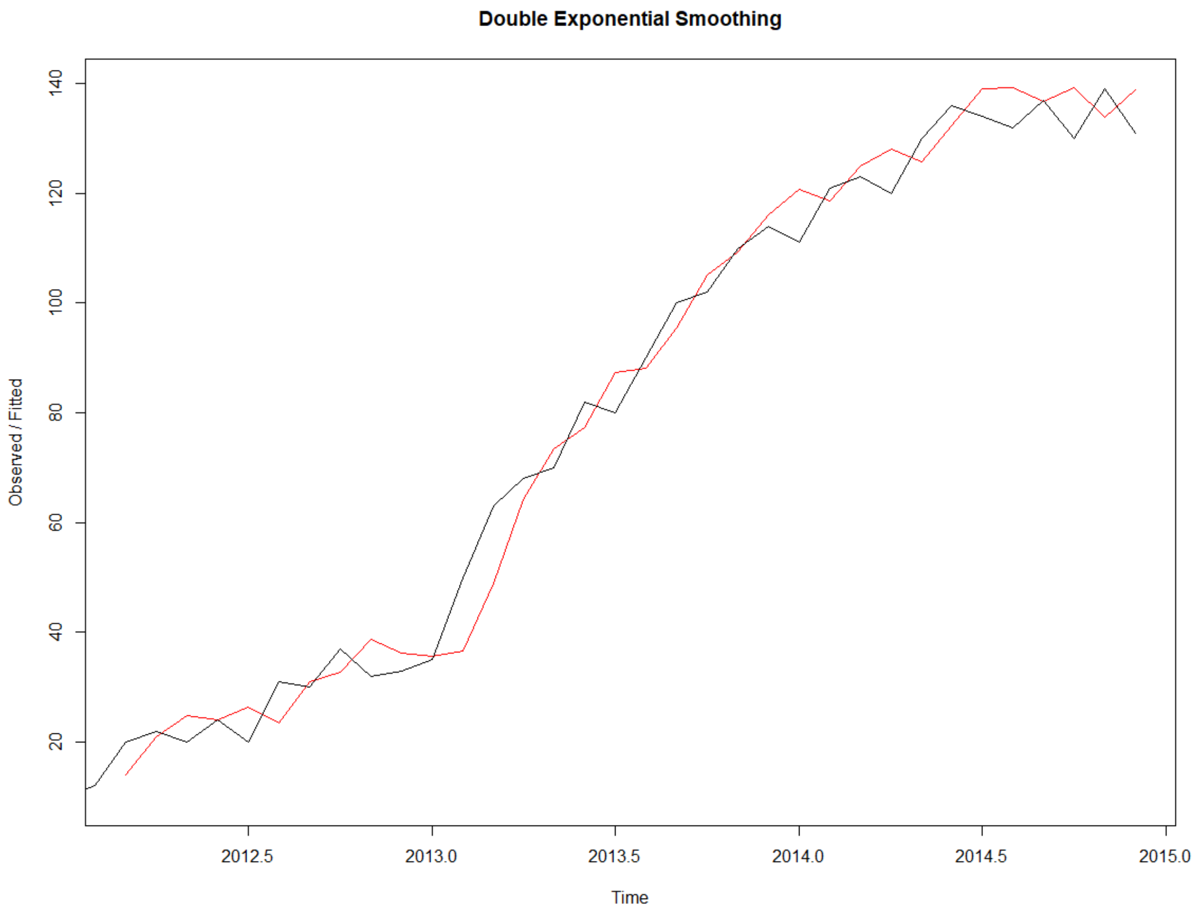
Thus, we can must try out different forecast methods to deduce which is the most appropriate. We use the time series data after January 2012 for the forecasting



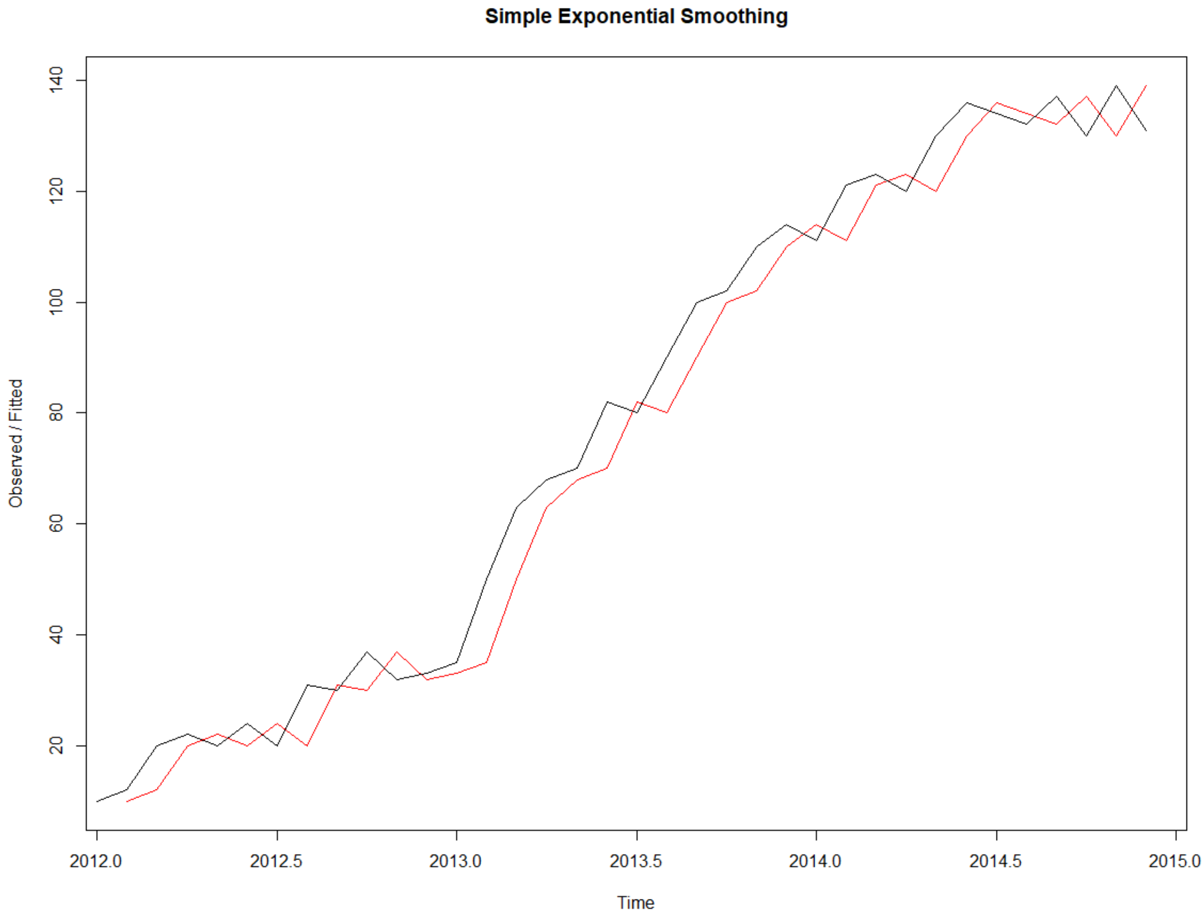
Graph 157: Observed time series data of tractor unit sales in China against the fitted Holt-Winters Additive model



Graph 158: Observed time series data of tractor unit sales in China against the fitted Holt-Winters Multiplicative model



Graph 159: Observed time series data of tractor unit sales in China against the fitted double exponential smoothing model



Graph 160: Observed time series data of tractor unit sales in China against the fitted single exponential smoothing model

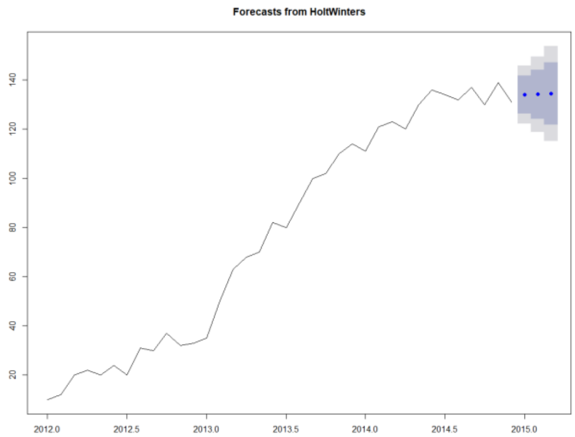
Graph 161: Observed time series data of tractor unit sales in China against the fitted Holt-Winters no-trend smoothing model

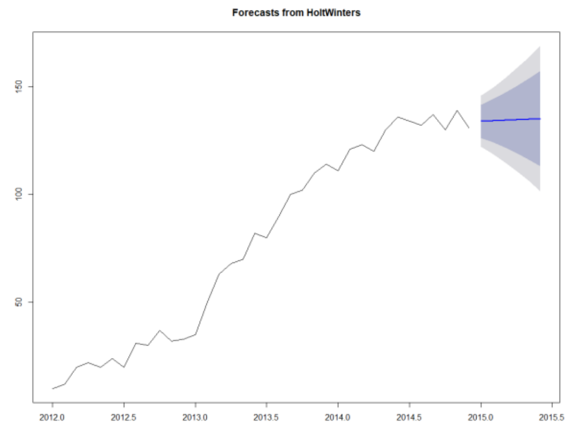
From the graphs plotted, we are unable to tell which model would be a better fit solely based on the graphs. Hence, we can calculate the error metrics (using the accuracy function) and see which model gives us an overall lower value of error metrics.

	RMSE	MAE	MAPE
Holt-Winters multiplicative	14.7	12.9	13.6
Double Exponential	5.94	4.86	8.83
Holt-Winters additive	8.31	6.49	7.98
Simple Exponential	6.86	5.68	10.2
Holt-Winters no-trend	10.2	8.08	10.3

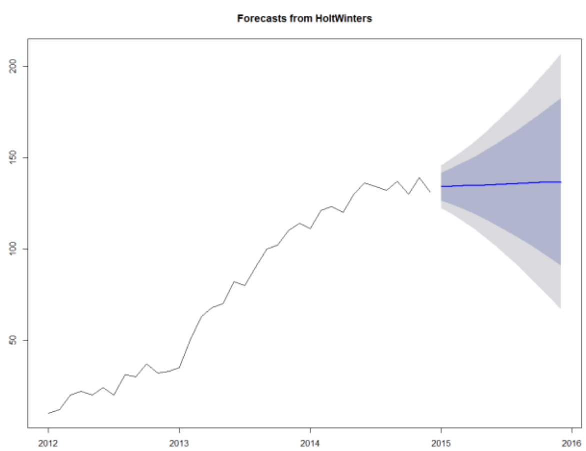
From the table, we can see that using a Double Exponential Smoothing model will be appropriate because it has lower values of error.

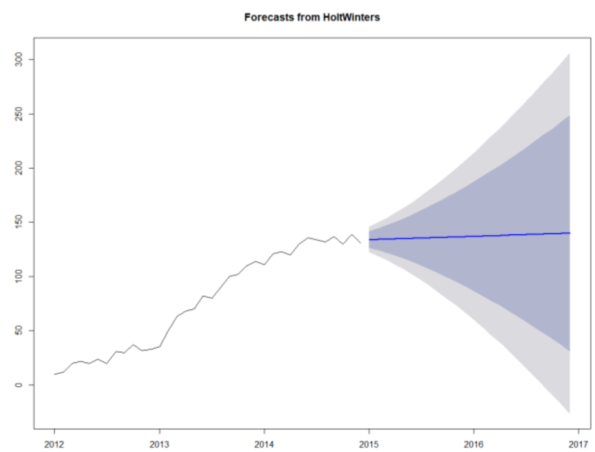
Hence, we can make use of the model to predict data in the next 3,6,12 and 24 months.





Graphs 162 and 163: Forecasted data of unit tractor sales in China over the next 3 and 6 months respectively.





Graphs 164 and 165: Forecasted data of unit tractor sales in China over the next 3 and 6 months respectively.

The following are the forecasted values of tractor unit sales in China over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	133.9610	126.23531	141.6866	122.1455888	145.7764
Feb 2015	134.2154	124.19741	144.2333	118.8942208	149.5365
Mar 2015	134.4698	121.83550	147.1041	115.1473161	153.7923
Apr 2015	134.7242	119.20094	150.2474	110.9834277	158.4650
May 2015	134.9786	116.32686	153.6303	106.4532257	163.5040
Jun 2015	135.2330	113.23614	157.2299	101.5916967	168.8743
Jul 2015	135.4874	109.94558	161.0292	96.4245533	174.5503
Aug 2015	135.7418	106.46820	165.0154	90.9716774	180.5120
Sep 2015	135.9962	102.81444	169.1780	85.2490648	186.7434
Oct 2015	136.2506	98.99299	173.5083	79.2699863	193.2313
Nov 2015	136.5050	95.01122	177.9988	73.0457178	199.9643
Dec 2015	136.7594	90.87551	182.6434	66.5860220	206.9329
Jan 2016	137.0138	86.59147	187.4362	59.8994784	214.1282
Feb 2016	137.2682	82.16410	192.3724	52.9937192	221.5428
Mar 2016	137.5227	77.59787	197.4474	45.8756014	229.1697

```
Apr 2016       137.7771  72.89685 202.6573  38.5513376 237.0028
May 2016       138.0315  68.06474 207.9982  31.0265959 245.0363
Jun 2016       138.2859  63.10495 213.4668  23.3065793 253.2652
Jul 2016       138.5403  58.02062 219.0599  15.3960886 261.6845
Aug 2016       138.7947  52.81465 224.7747   7.2995743 270.2898
Sep 2016       139.0491  47.48976 230.6084  -0.9788218 279.0770
Oct 2016       139.3035  42.04847 236.5585  -9.4352305 288.0422
Nov 2016       139.5579  36.49315 242.6226 -18.0660261 297.1818
Dec 2016       139.8123  30.82604 248.7986 -26.8678010 306.4924
```

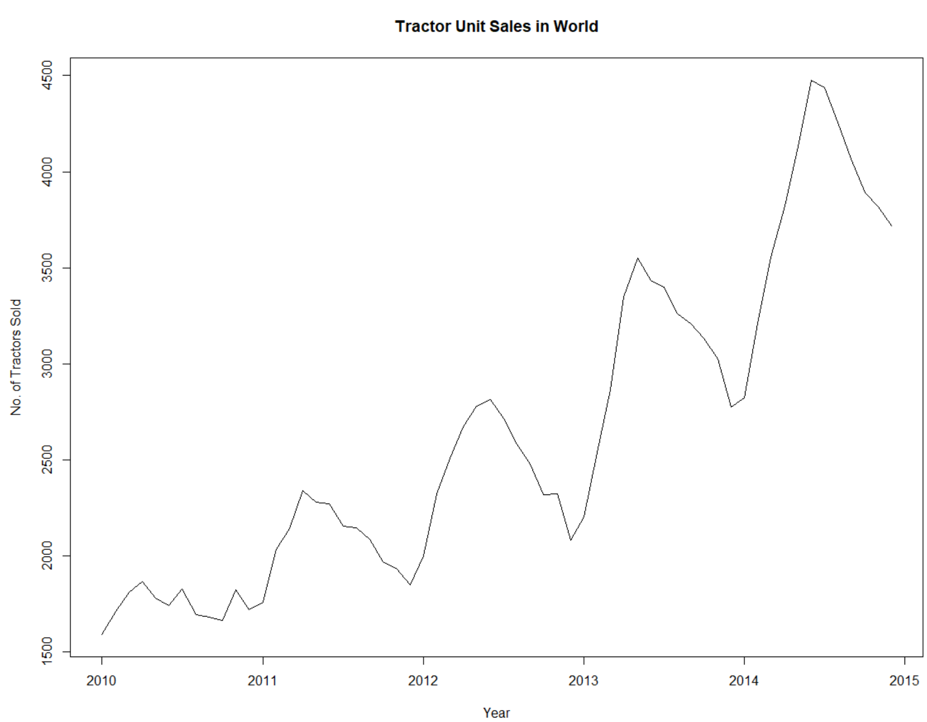
Conclusion:

As seen from the forecasted graphs and values generated, unit tractor sales in China is predicted to reach a plateau at a value of around 137-130 with very slow increase over the next 24 months.

Forecasting of Unit Sales of Tractor in World

Time Series:

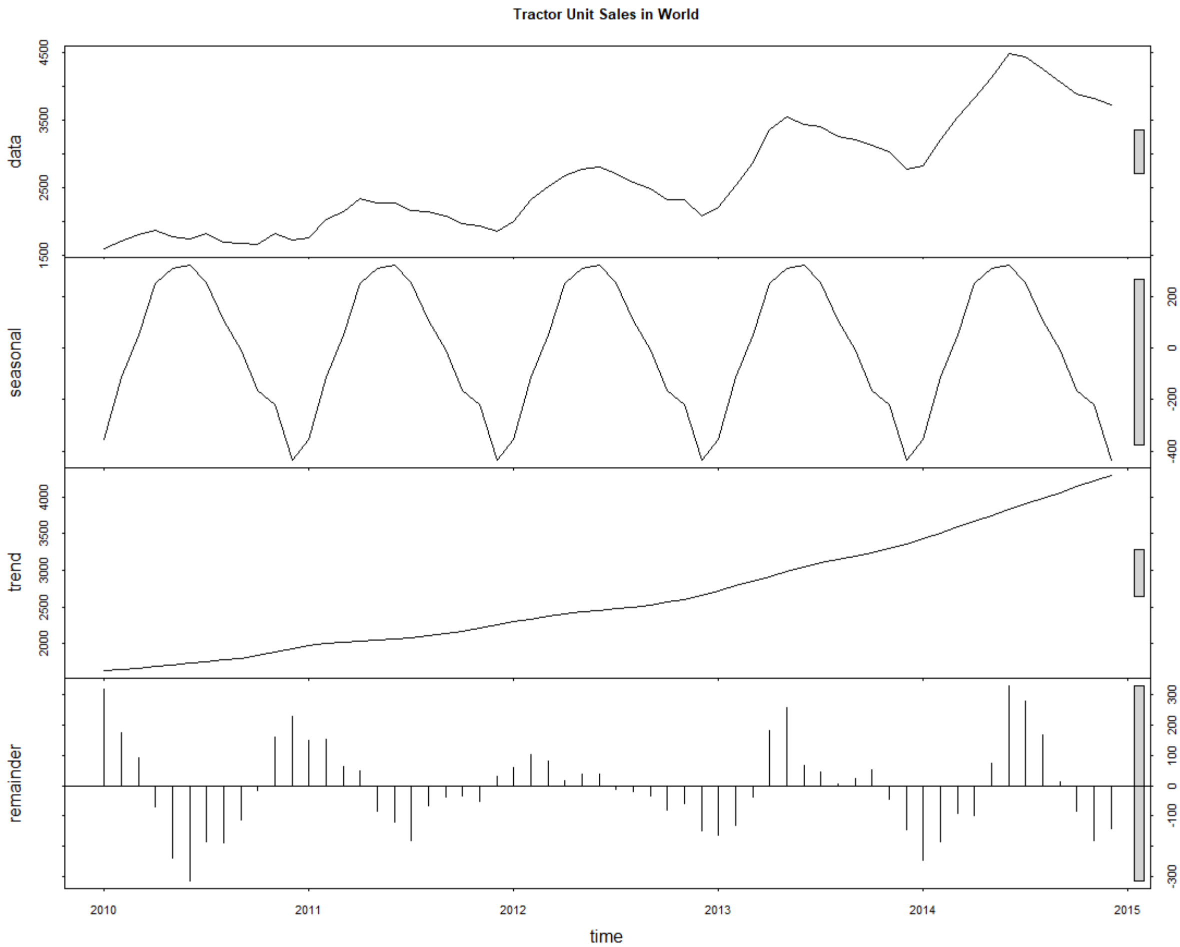
```
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2010 1592 1711 1810 1867 1779 1740 1826 1695 1681 1663 1825 1720
2011 1761 2035 2142 2340 2280 2271 2154 2146 2085 1970 1936 1850
2012 2000 2324 2510 2672 2780 2813 2716 2581 2476 2317 2324 2080
2013 2202 2540 2867 3348 3550 3432 3400 3261 3209 3132 3027 2777
2014 2821 3209 3553 3820 4133 4476 4436 4256 4067 3890 3816 3717
```



Graph 166: The time series graph above shows the tractor unit sales in the World from January 2 010 to December 2014.

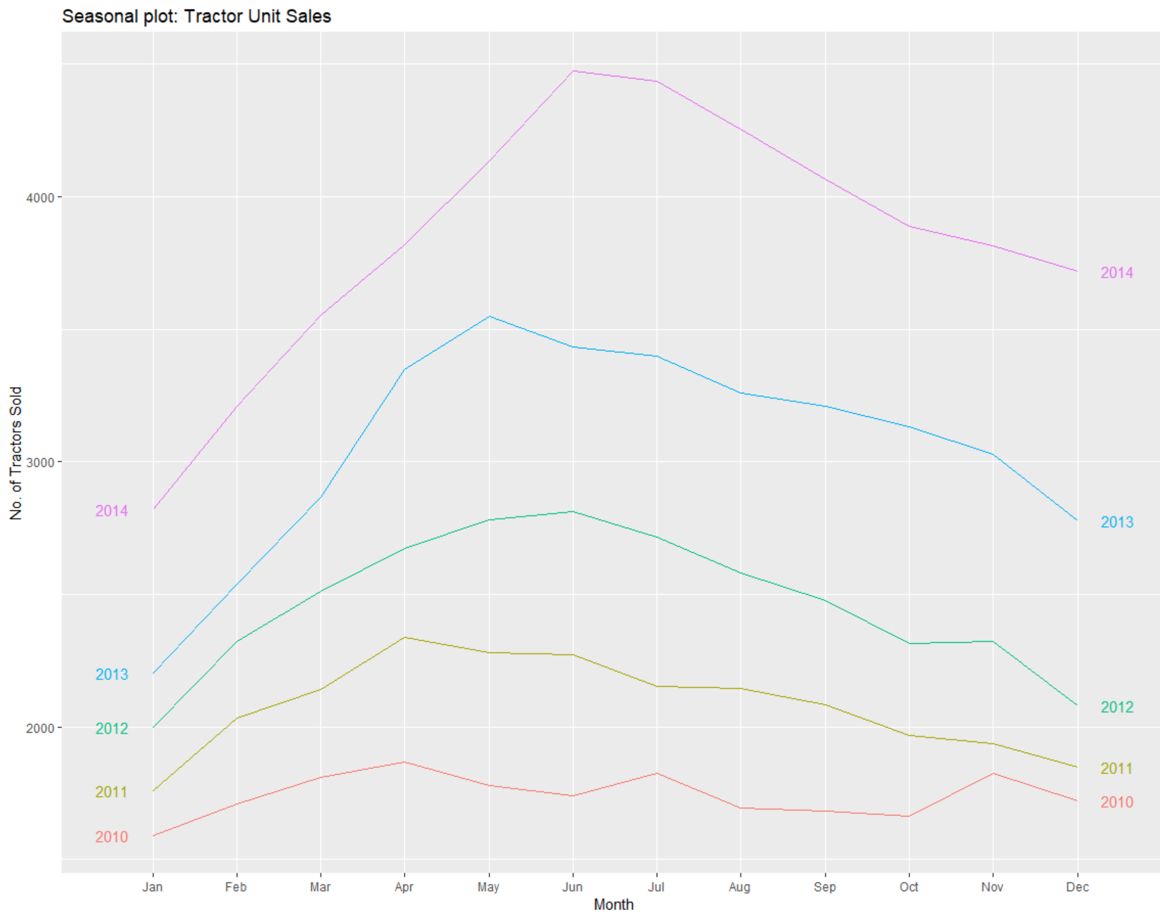
From the graph, there is an upward trend in the data over this period.
There is also a slight presence of seasonality.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 167: Decomposed time series data of tractor unit sales in the World from January 2010 to December 2014

Decomposition of the time series further proves the point that there is an upward trend.

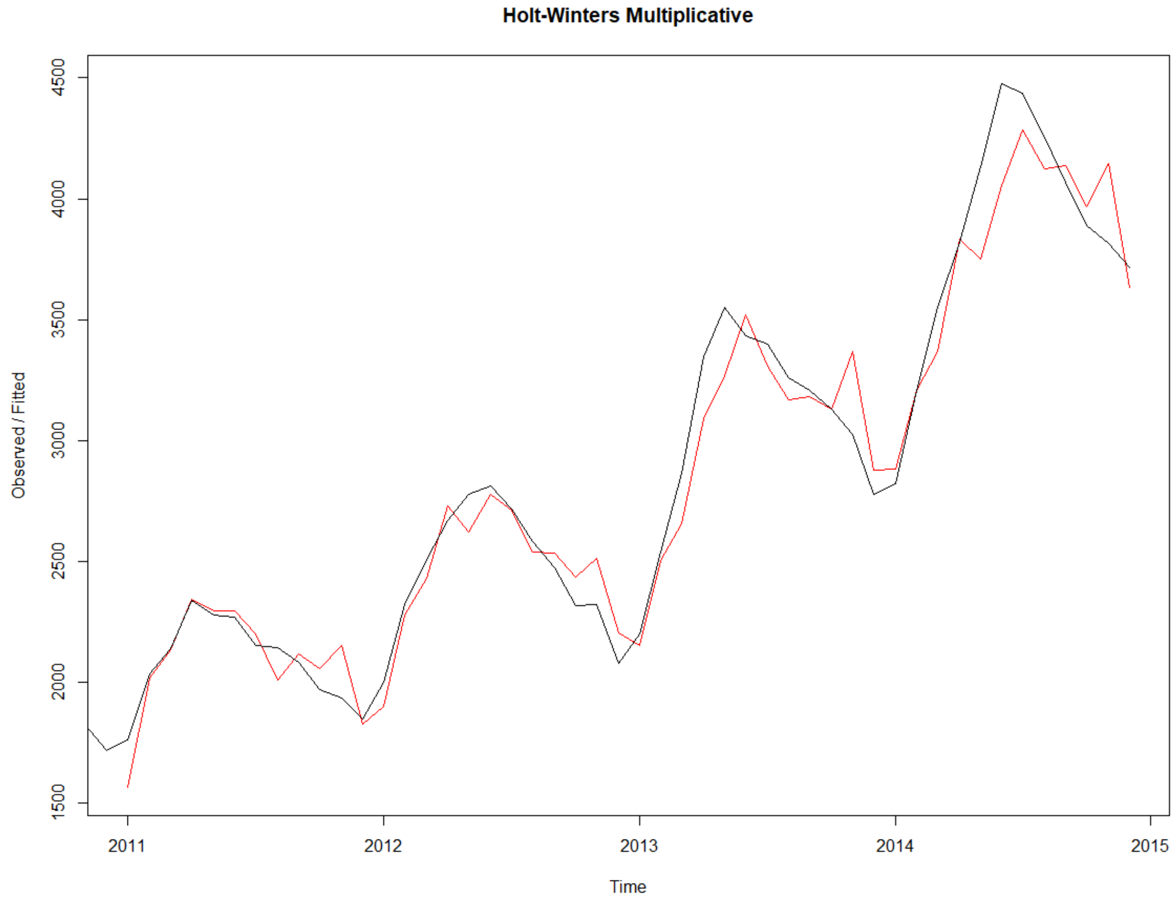


Graph 168: A seasonal plot of tractor unit sales in the World from January 2010 to December 2014.

From the seasonal plot, there seems to be the presence of seasonality due to constant fluctuations at a specific time of every year. We can deduce that multiplicative decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations varies with the level of the time series.

Hence, we can conclude that the sales of tractors in World possess <u>trend and seasonality.</u>

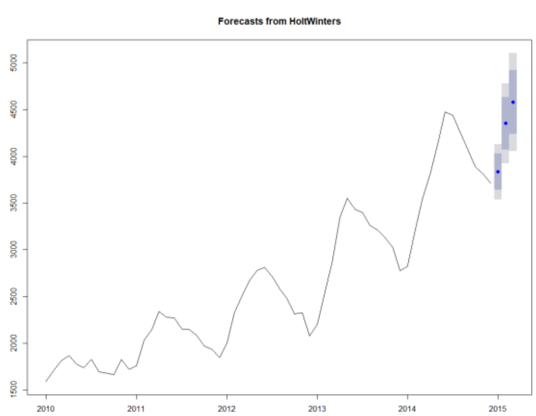
We can use <u>Holt-Winters Multiplicative Model</u> to forecast future sales

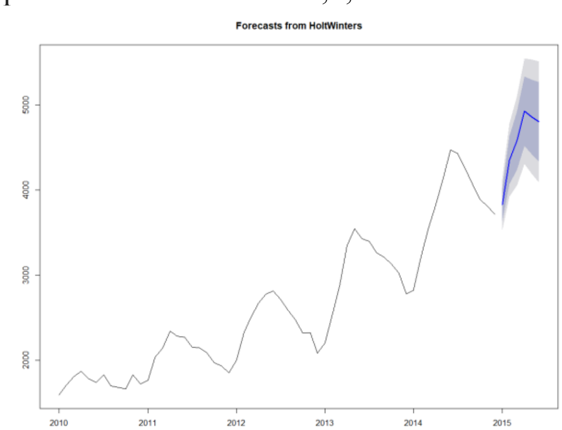


Graph 169: Observed time series data of tractor unit sales in the World against the fitted Holt-Winters multiplicative model

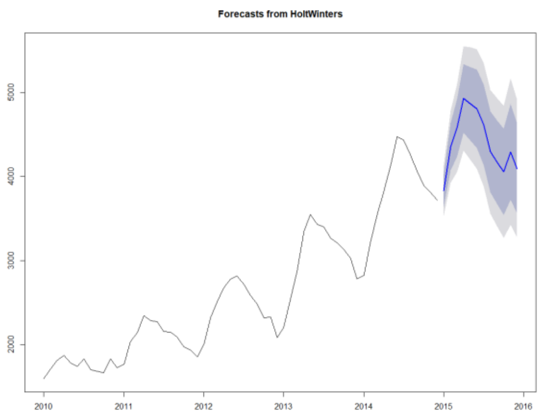
The model fits well with the observed time series data.

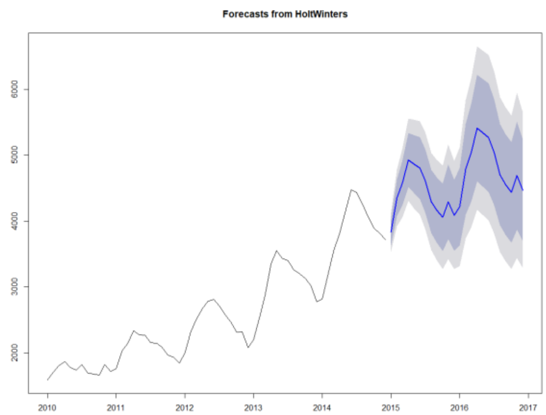
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months.





Graphs 170 and 171: Forecasted data of unit tractor sales in the World over the next 3 and 6 months respectively.





Graphs 172 and 173: Forecasted data of unit tractor sales in the World over the next 12 and 24 months respectively.

The following are the forecasted values of tractor unit sales in the World over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	3833.994	3640.108	4027.880	3537.471	4130.517
Feb 2015	4353.170	4073.477	4632.864	3925.416	4780.925
Mar 2015	4580.715	4238.616	4922.814	4057.520	5103.910
Apr 2015	4929.902	4523.469	5336.335	4308.317	5551.487
May 2015	4867.860	4430.265	5305.455	4198.617	5537.104
Jun 2015	4805.801	4340.144	5271.459	4093.639	5517.963
Jul 2015	4609.810	4130.315	5089.305	3876.486	5343.134
Aug 2015	4296.271	3816.249	4776.292	3562.141	5030.400
Sep 2015	4167.997	3670.976	4665.018	3407.869	4928.125
Oct 2015	4057.579	3543.671	4571.486	3271.625	4843.533
Nov 2015	4293.727	3724.347	4863.107	3422.936	5164.519
Dec 2015	4092.670	3557.249	4628.091	3273.814	4911.526
Jan 2016	4218.252	3630.648	4805.856	3319.589	5116.915
Feb 2016	4785.848	4100.967	5470.730	3738.412	5833.284
Mar 2016	5032.270	4294.228	5770.312	3903.532	6161.008
Apr 2016	5411.919	4602.866	6220.972	4174.579	6649.259
May 2016	5339.964	4525.124	6154.804	4093.774	6586.155
Jun 2016	5268.150	4447.845	6088.455	4013.601	6522.698
Jul 2016	5049.776	4246.111	5853.441	3820.677	6278.875
Aug 2016	4703.076	3935.623	5470.529	3529.358	5876.795
Sep 2016	4559.567	3796.935	5322.199	3393.223	5725.912
Oct 2016	4435.814	3675.232	5196.396	3272.604	5599.024
Nov 2016	4690.891	3871.000	5510.781	3436.977	5944.804
Dec 2016	4468.340	3697.278	5239.402	3289.103	5647.577

Conclusion:

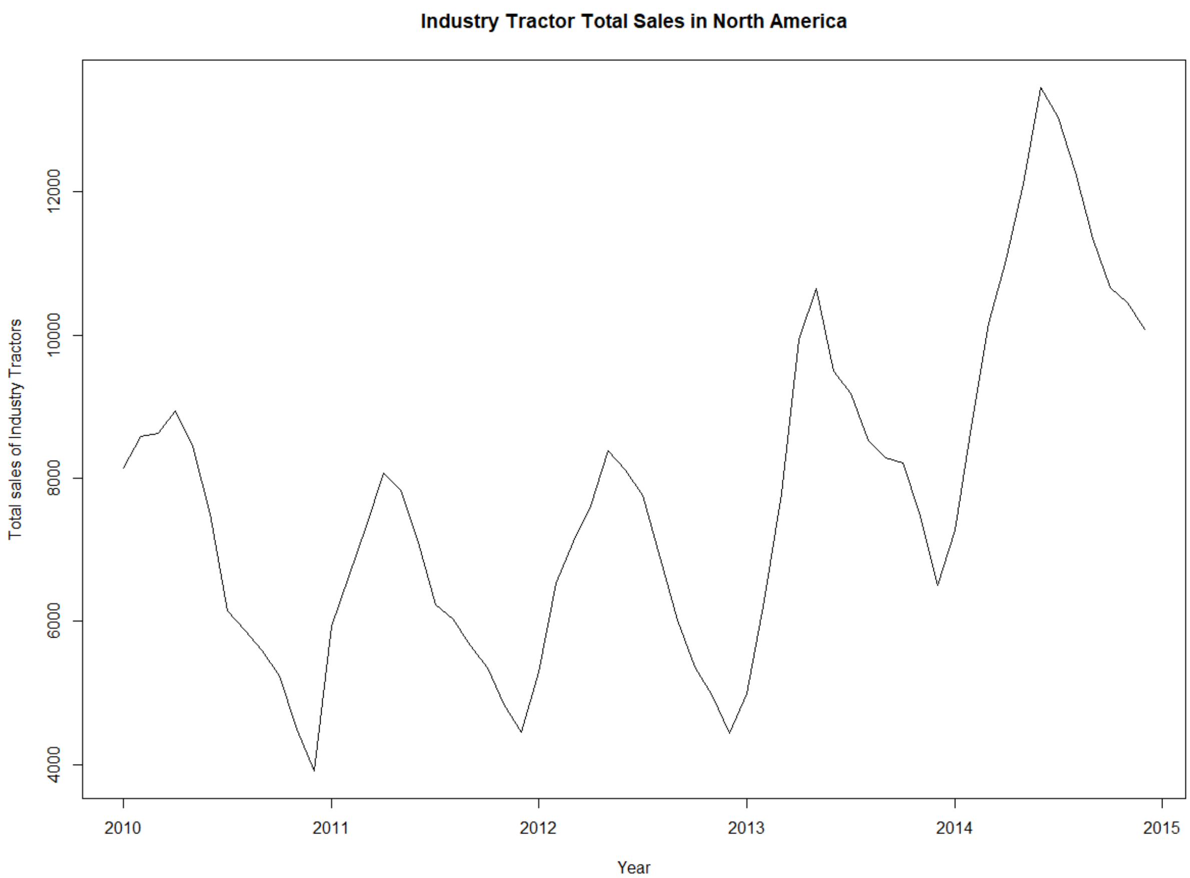
As seen from the forecasted graphs and values generated, unit tractor sales in the World is predicted to increase steadily for the next 24 months to come. Seasonality will remain.

Next, we will forecast the total industry sales of tractors in each marketing region using the excel sheet *Industry Tractor Total Sales*

<u>Forecasting of Industry Sales of Tractors in North America</u>

<u>Time Series:</u>

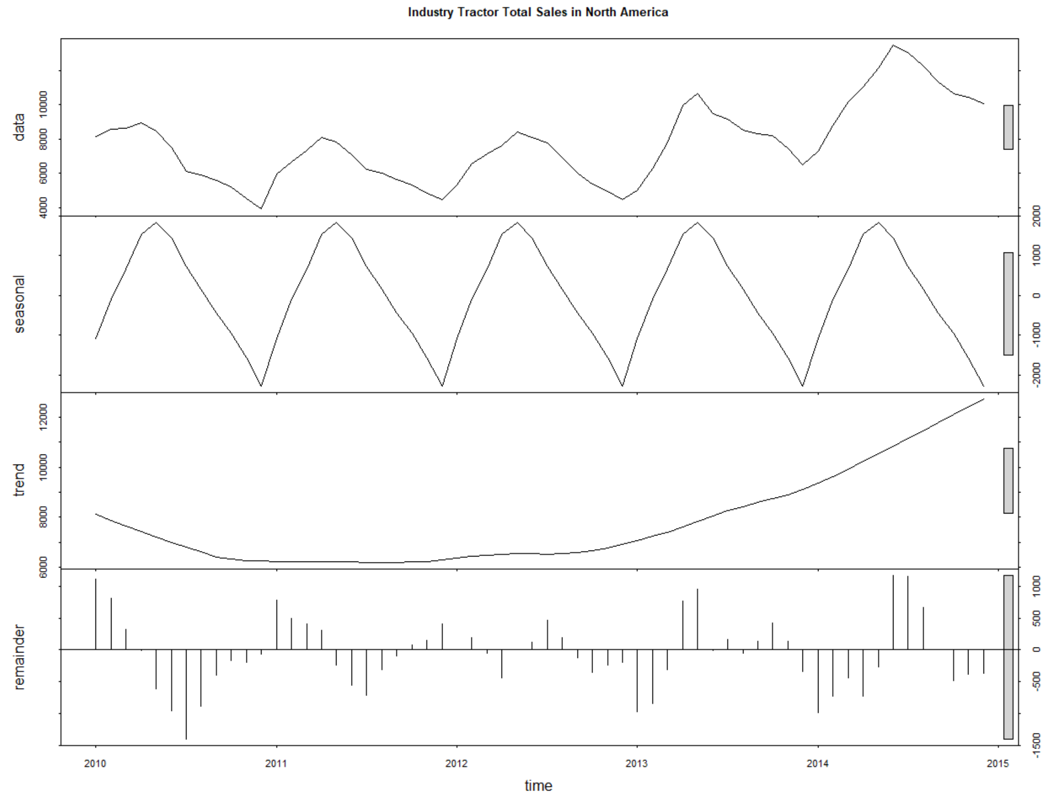
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	8142.857	8591.549	8630.137	8947.368	8441.558	7500.000	6144.578	5882.353	5595.238	5232.558	4494.382	3913.043
2011	5937.500	6632.653	7326.733	8076.923	7830.189	7102.804	6238.532	6036.036	5663.717	5344.828	4830.508	4453.782
2012	5299.145	6528.926	7120.000	7619.048	8387.097	8110.236	7751.938	6893.939	6015.038	5367.647	4964.029	4444.444
2013	5000.000	6283.784	7785.235	9934.211	10645.161	9491.000	9182.390	8527.607	8292.683	8220.859	7469.880	6508.876
2014	7267.442	8806.818	10167.598	11043.956	12119.565	13459.459	13048.128	12275.132	11347.150	10666.667	10459.184	10082.000



Graph 174: The time series graph above shows the tractor industry sales in North America from January 2010 to December 2014.

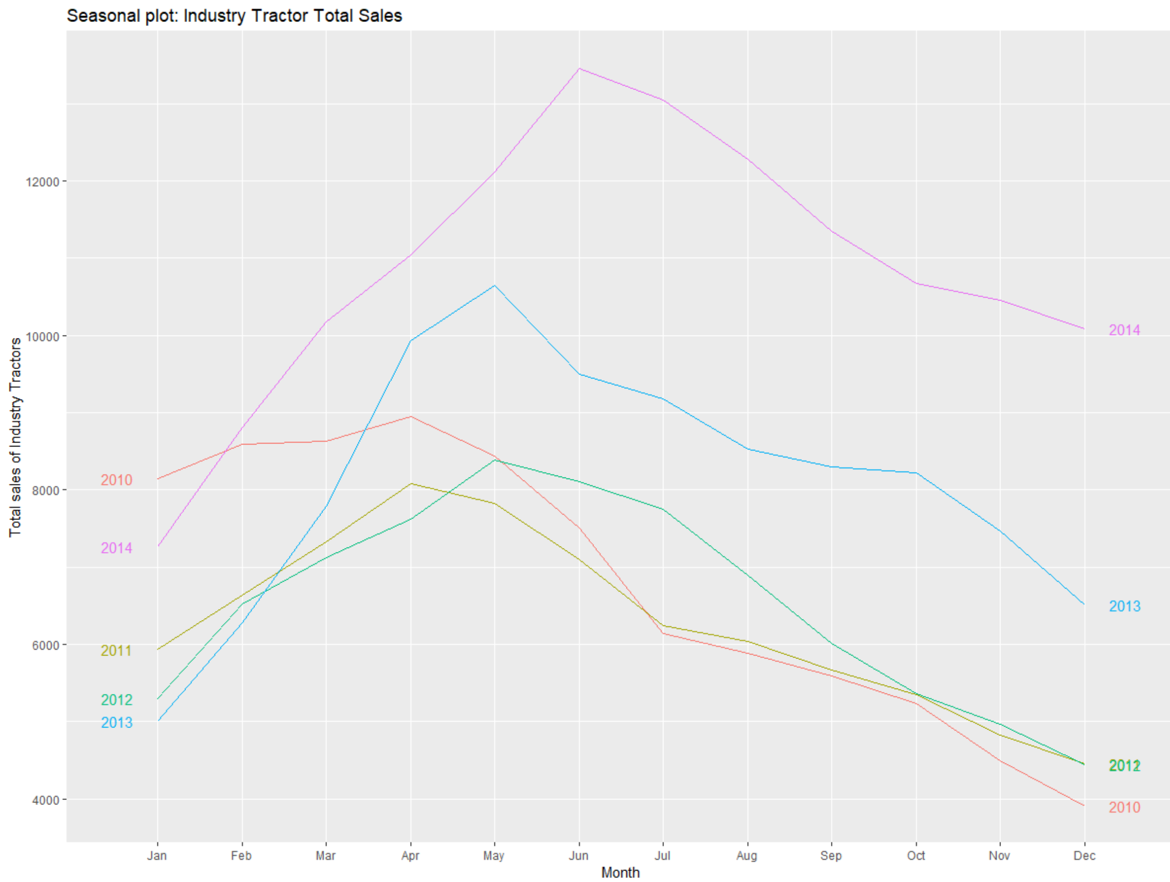
From the graph, there is an upward trend in the data over this period.
There is also a slight presence of seasonality.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 175: Decomposed time series data of tractor industry sales in North America from January 2010 to December 2014

Decomposition of the time series further proves the point that there is an upward trend

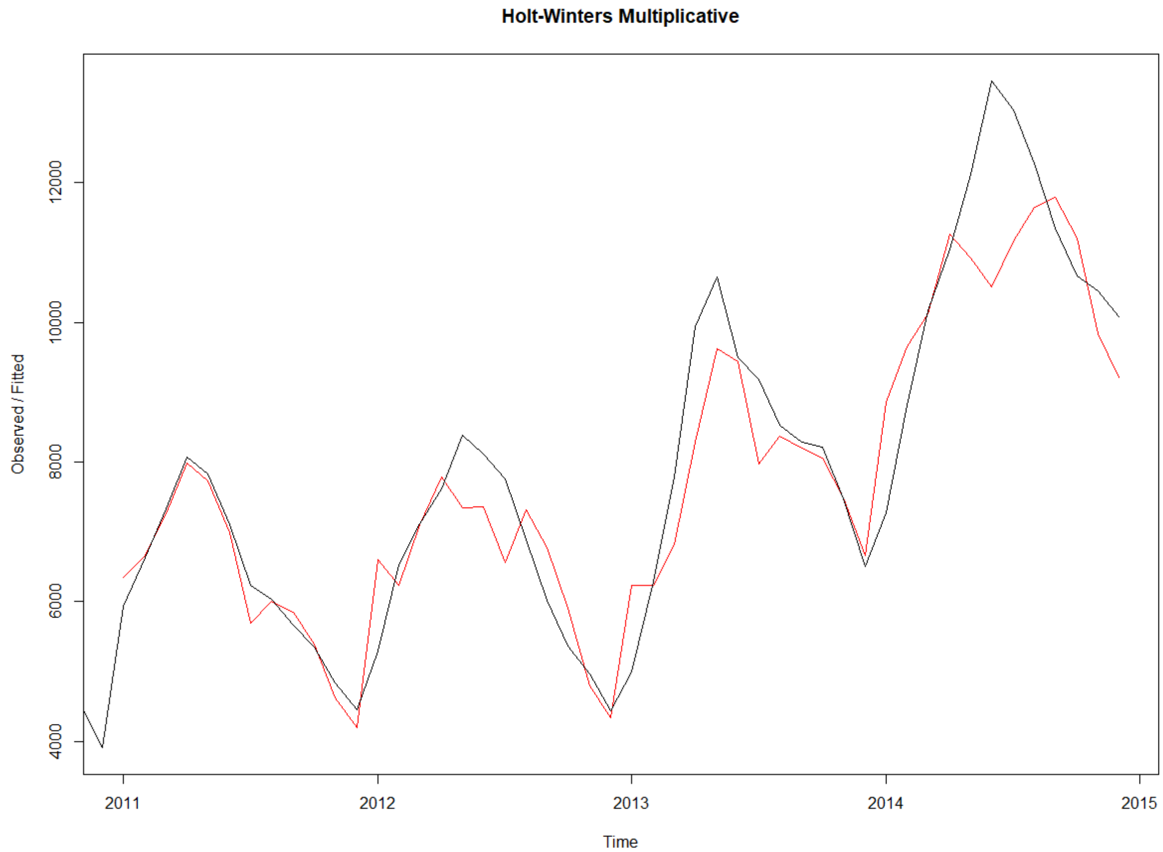


Graph 176: A seasonal plot of tractor industry sales in North America from January 2010 to December 2014.

From the seasonal plot, there seems to be the presence of seasonality (besides the anomaly of the first year) due to constant fluctuations at a specific time of every year. We can deduce that multiplicative decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations varies with the level of the time series.

Hence, we can conclude that the industry sales of tractors in North America possess trend and seasonality.

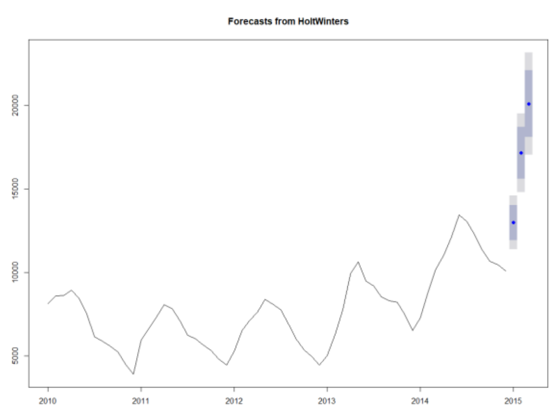
We can use Holt-Winters Multiplicative Model to forecast future sales.

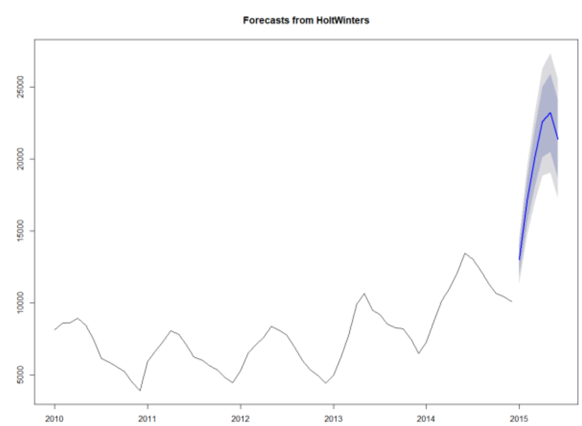


Graph 177: Observed time series data of tractor industry sales in North America against the fitted Holt-Winters Multiplicative Model

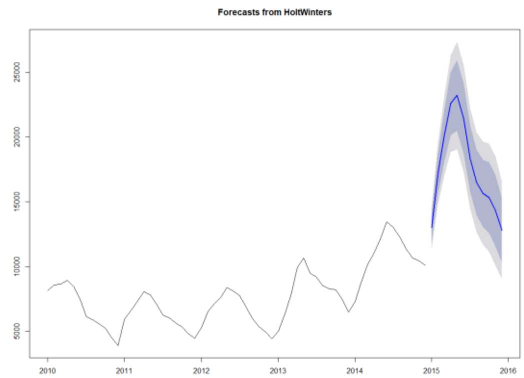
The model fits well with the observed time series data.

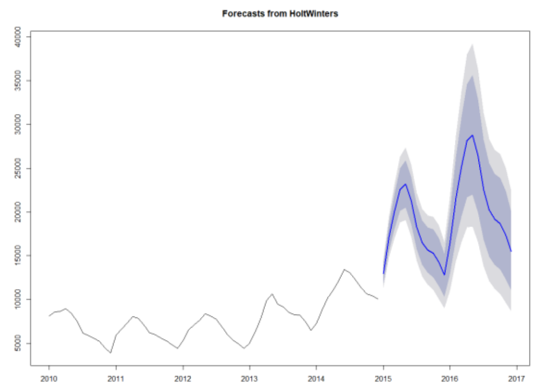
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months





Graphs 178 and 179: Forecasted data of industry tractor sales in North America over the next 3 and 6 months respectively.





Graphs 179 and 180: Forecasted data of industry tractor sales in North America over the next 12 and 24 months respectively.

The following are the forecasted values of tractor industry sales in North America over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	12984.63	11935.57	14033.69	11380.237	14589.03
Feb 2015	17157.69	15609.49	18705.88	14789.924	19525.45
Mar 2015	20099.24	18094.77	22103.72	17033.659	23164.83
Apr 2015	22571.54	20131.72	25011.36	18840.158	26302.93
May 2015	23199.12	20492.46	25905.78	19059.647	27338.60
Jun 2015	21349.48	18645.34	24053.61	17213.860	25485.09
Jul 2015	18311.80	15767.03	20856.58	14419.911	22203.70
Aug 2015	16496.30	13985.12	19007.48	12655.780	20336.82
Sep 2015	15668.67	13076.07	18261.27	11703.634	19633.71
Oct 2015	15293.23	12563.70	18022.77	11118.772	19467.70
Nov 2015	14284.57	11530.92	17038.21	10073.230	18495.90
Dec 2015	12784.23	10322.58	15245.87	9019.470	16548.98
Jan 2016	16388.80	12891.78	19885.82	11040.571	21737.03
Feb 2016	21559.73	16863.99	26255.47	14378.213	28741.25
Mar 2016	25148.04	19517.78	30778.30	16537.301	33758.78
Apr 2016	28125.11	21654.24	34595.98	18228.774	38021.45
May 2016	28792.42	21976.81	35608.03	18368.850	39215.99
Jun 2016	26395.44	19948.87	32842.01	16536.267	36254.62
Jul 2016	22556.22	16849.28	28263.16	13828.206	31284.23
Aug 2016	20247.45	14938.88	25556.02	12128.699	28366.20
Sep 2016	19165.37	13966.57	24364.17	11214.486	27116.25
Oct 2016	18643.83	13419.14	23868.52	10653.357	26634.31
Nov 2016	17358.06	12319.84	22396.28	9652.769	25063.35
Dec 2016	15486.45	11002.73	19970.17	8629.194	22343.71

<u>Conclusion:</u>

As seen from the forecasted graphs and values generated, industry tractor sales in North America is predicted to increase drastically for the next 24 months to come. Seasonality will remain.

Future Predictions of Market Share in North America:

Both forecasted tractor sales (Industry and Unit) in North America had similar trends. Both will see an increase in sales in the next 24 months. Meanwhile, seasonality remained where where the tractor sales increases till mid-year before decreasing back to around the original volume of tractor sales that was sold at the start of the year.

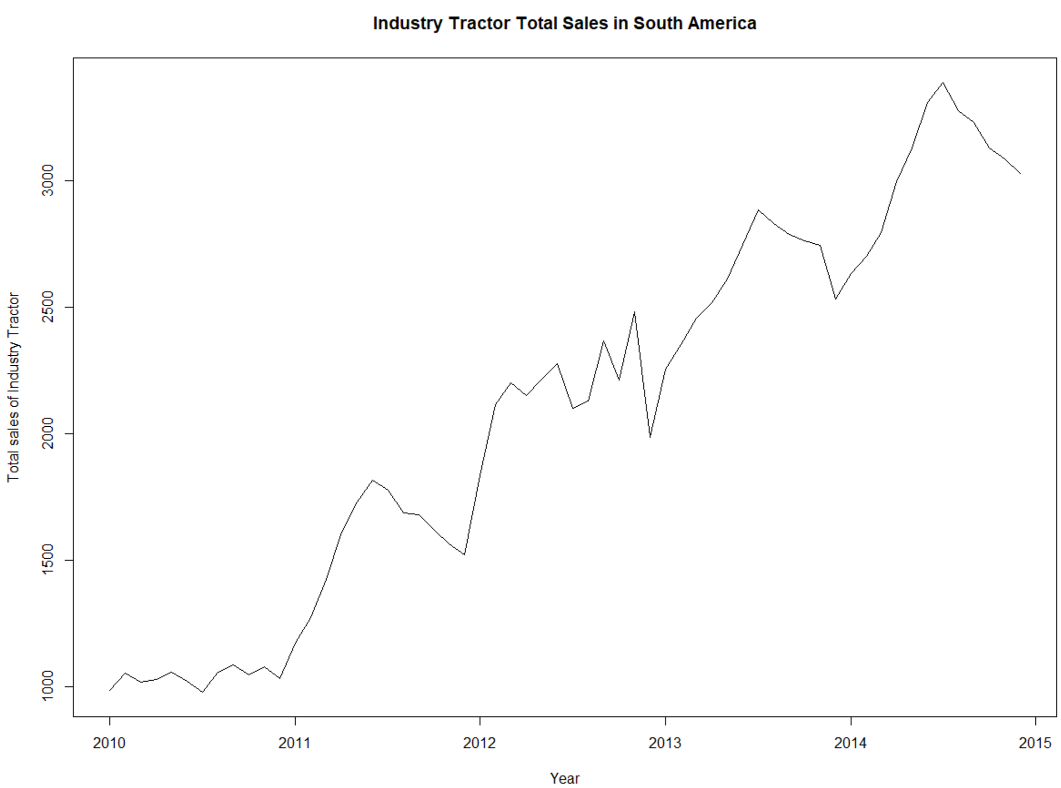
Since *market share = unit sales/industry sales*, there will be no predicted change in market shares in North America for the sales of tractor since both are predicted to increase at around the same rate.

Forecasting of Industry Sales of Tractors in South America

Time Series:

u

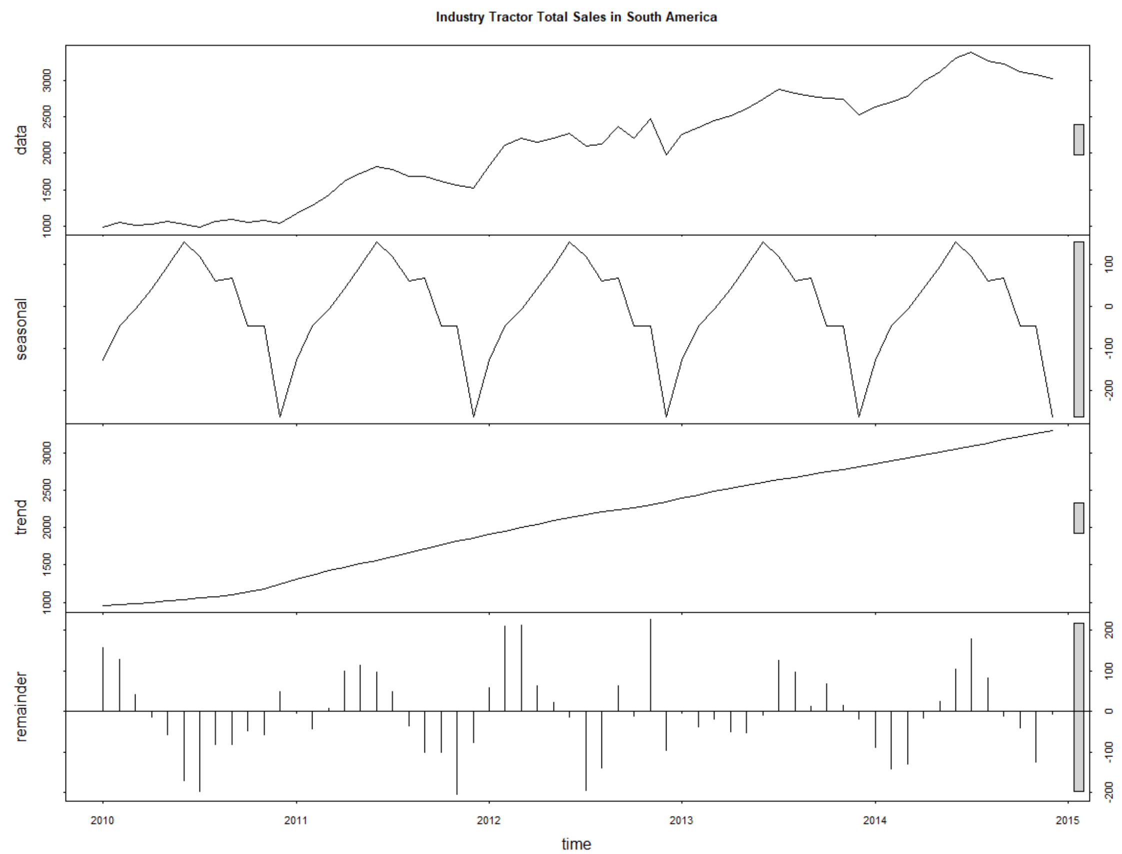
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	984.0000	1050.5837	1015.6250	1026.6160	1056.6038	1018.8679	977.4436	1056.6038	1086.1423	1044.7761	1078.0669	1029.4118
2011	1172.1612	1272.7273	1423.3577	1611.7216	1727.9412	1814.8148	1776.0000	1684.9817	1678.8321	1617.6471	1563.6364	1521.7391
2012	1834.5324	2114.6953	2202.1661	2150.5376	2214.2857	2277.5801	2099.6441	2127.6596	2367.4912	2210.5263	2482.5175	1986.0627
2013	2256.9444	2352.9412	2456.7474	2517.2414	2611.6838	2749.1409	2886.5979	2832.7645	2789.1156	2764.5051	2745.7627	2533.7838
2014	2635.1351	2702.7027	2794.6128	2996.6330	3131.3131	3310.8108	3389.8305	3277.0270	3232.3232	3131.3131	3087.2483	3030.3030



Graph 181: The time series graph above shows the tractor industy sales in South America from January 2010 to December 2014.

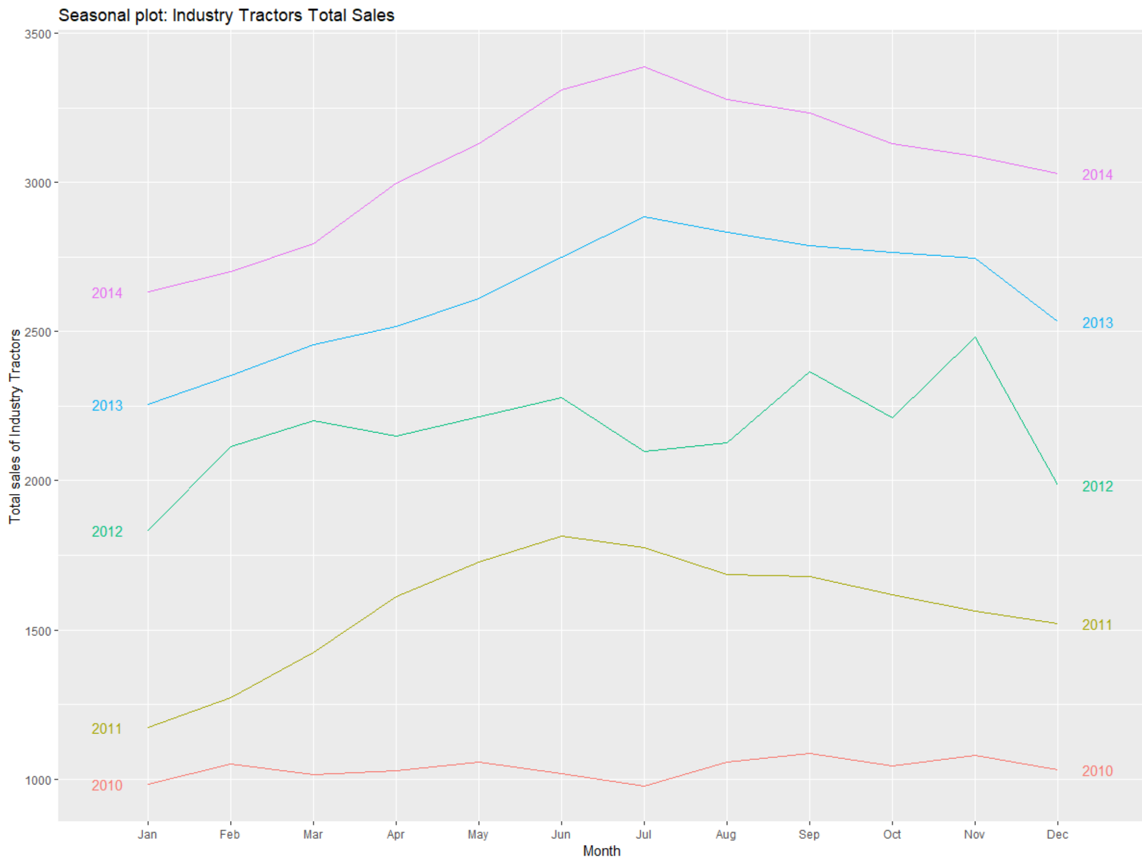
From the graph, there is an upward trend in the data over this period.
There is also a slight presence of seasonality.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 182: The time series graph above shows the tractor industry sales in South America from January 2010 to December 2014.

Decomposition of the time series further proves the point that there is an upward trend.

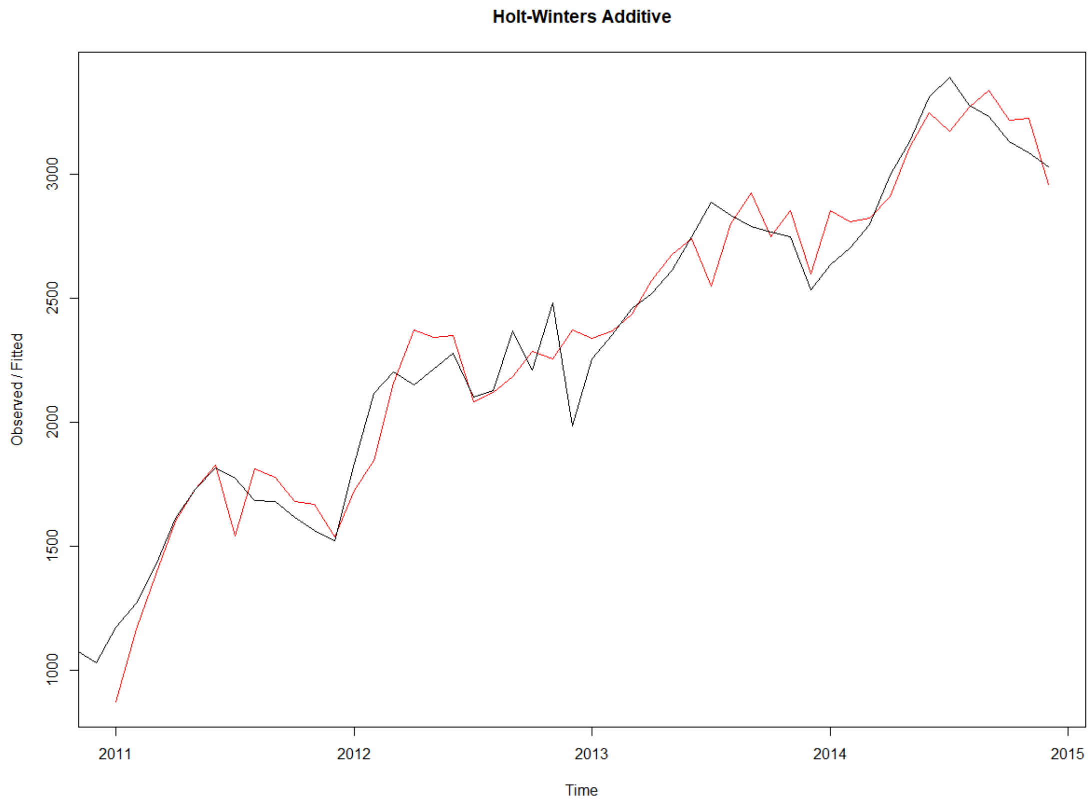


Graph 183: A seasonal plot of tractor industry sales in South America from January 2010 to December 2014.

From the seasonal plot, there seems to be the presence of seasonality due to constant fluctuations at a specific time of every year. We can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations varies with the level of the time series.

Hence, we can conclude that the industry sales of tractors in South America possess trend and seasonality.

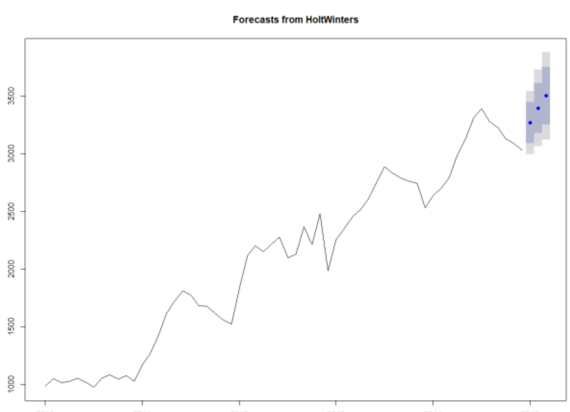
Thus, we can use Holt-Winters Additive Model.

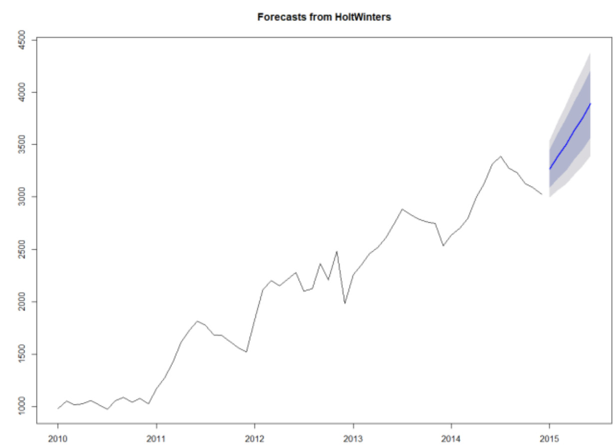


Graph 184: Observed time series data of mower industry sales in South America against the fitted Holt-Winters additive model

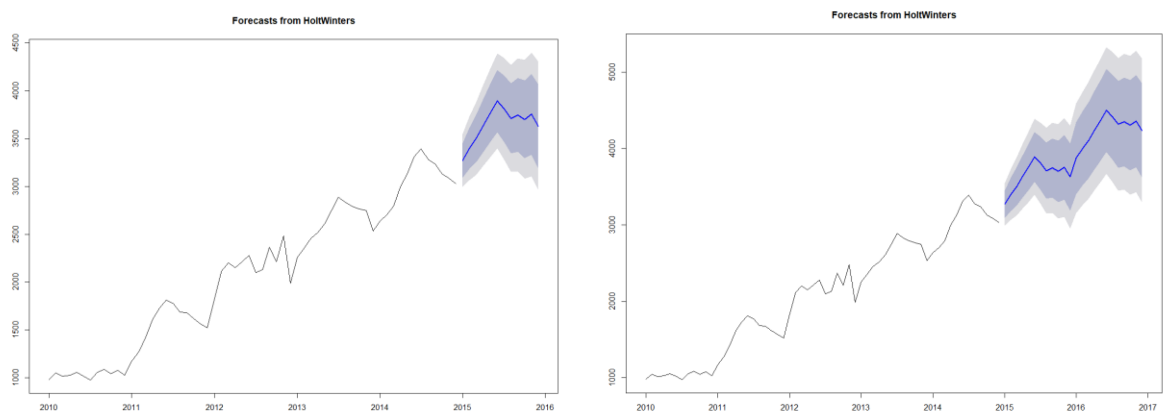
The model fits well with the observed time series data.

Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months.





Graphs 185 and 186: Forecasted data of industry tractor sales in South America over the next 3 and 6 months respectively.



Graphs 187 and 188: Forecasted data of industry tractor sales in South America over the next 12 and 24 months respectively.

The following are the forecasted values of tractor industry sales in South America over the next 24 months

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	3269.268	3089.904	3448.633	2994.954	3543.582
Feb 2015	3396.882	3180.649	3613.114	3066.183	3727.580
Mar 2015	3501.880	3254.208	3749.552	3123.099	3880.661
Apr 2015	3633.364	3357.817	3908.911	3211.951	4054.777
May 2015	3757.027	3456.177	4057.878	3296.916	4217.139
Jun 2015	3891.254	3567.069	4215.439	3395.456	4387.052
Jul 2015	3809.346	3463.397	4155.295	3280.262	4338.429
Aug 2015	3709.648	3343.226	4076.071	3149.253	4270.044
Sep 2015	3744.407	3358.595	4130.218	3154.359	4334.455
Oct 2015	3699.941	3295.669	4104.212	3081.661	4318.221
Nov 2015	3751.640	3329.716	4173.565	3106.362	4396.918
Dec 2015	3627.594	3188.726	4066.462	2956.403	4298.785
Jan 2016	3872.928	3404.557	4341.299	3156.616	4589.240
Feb 2016	4000.541	3516.851	4484.232	3260.800	4740.282
Mar 2016	4105.540	3607.001	4604.079	3343.090	4867.990
Apr 2016	4237.024	3724.065	4749.982	3452.521	5021.526
May 2016	4360.687	3833.704	4887.670	3554.736	5166.638
Jun 2016	4494.914	3954.270	5035.558	3668.070	5321.758
Jul 2016	4413.006	3859.037	4966.974	3565.784	5260.227
Aug 2016	4313.308	3746.329	4880.288	3446.188	5180.429
Sep 2016	4348.067	3768.368	4927.766	3461.494	5234.640
Oct 2016	4303.601	3711.456	4895.746	3397.993	5209.208
Nov 2016	4355.300	3750.965	4959.635	3431.050	5279.550
Dec 2016	4231.254	3614.971	4847.538	3288.730	5173.778

<u>Conclusion:</u>

As seen from the forecasted graphs and values generated, industry tractor sales in South America is predicted to increase steadily for the next 24 months to come. Seasonality will remain.

Future Predictions of Market Share in South America:

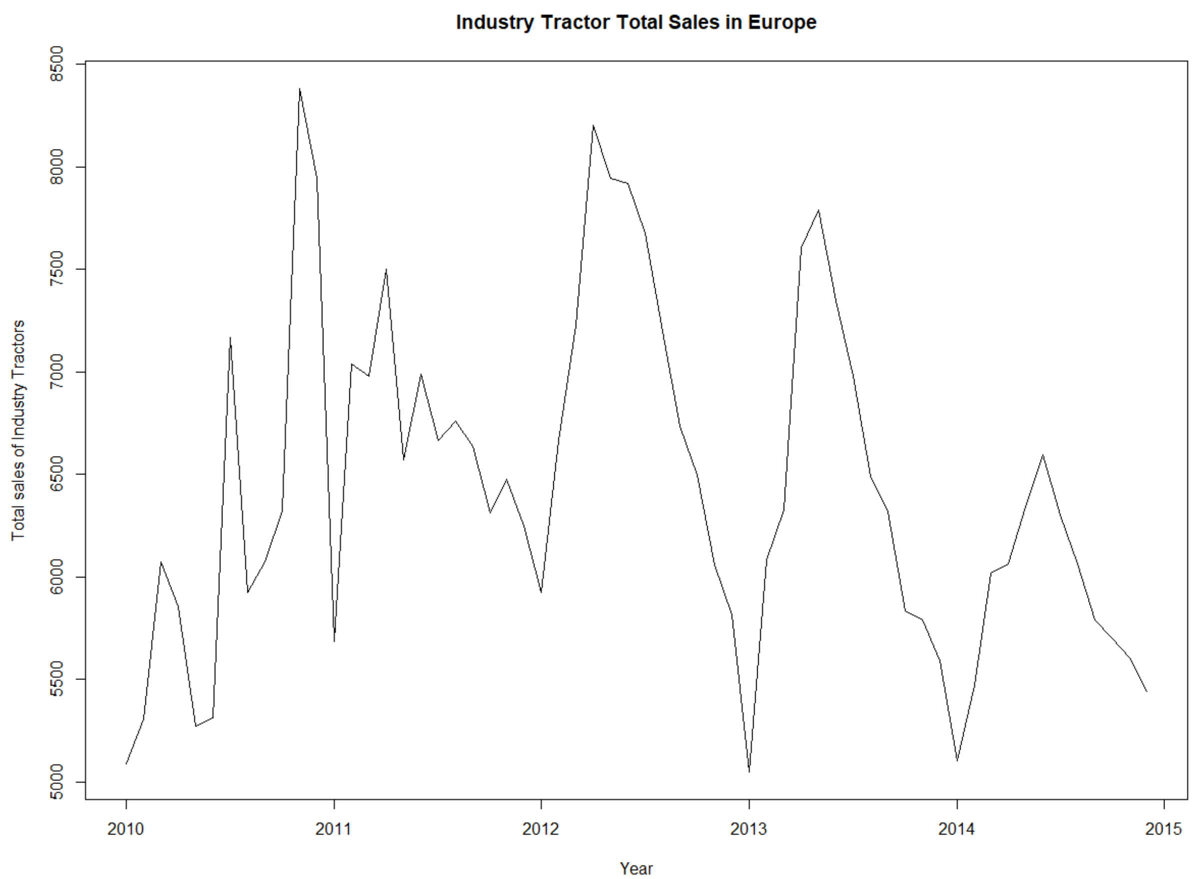
Forecasted unit tractor sales in South America showed an upward trend while forecasted industry tractor sales in South America showed similar upward trend in sales output. Both sales showed signs of similar seasonality patterns as well.

Since *market share = unit sales/industry sales*, there will be no predicted change in market shares in South America for the sales of tractors. It is predicted that both types of sales will increase at a similar rate for the next 24 months.

Forecasting of Industry Sales of Tractors in Europe

Time Series:

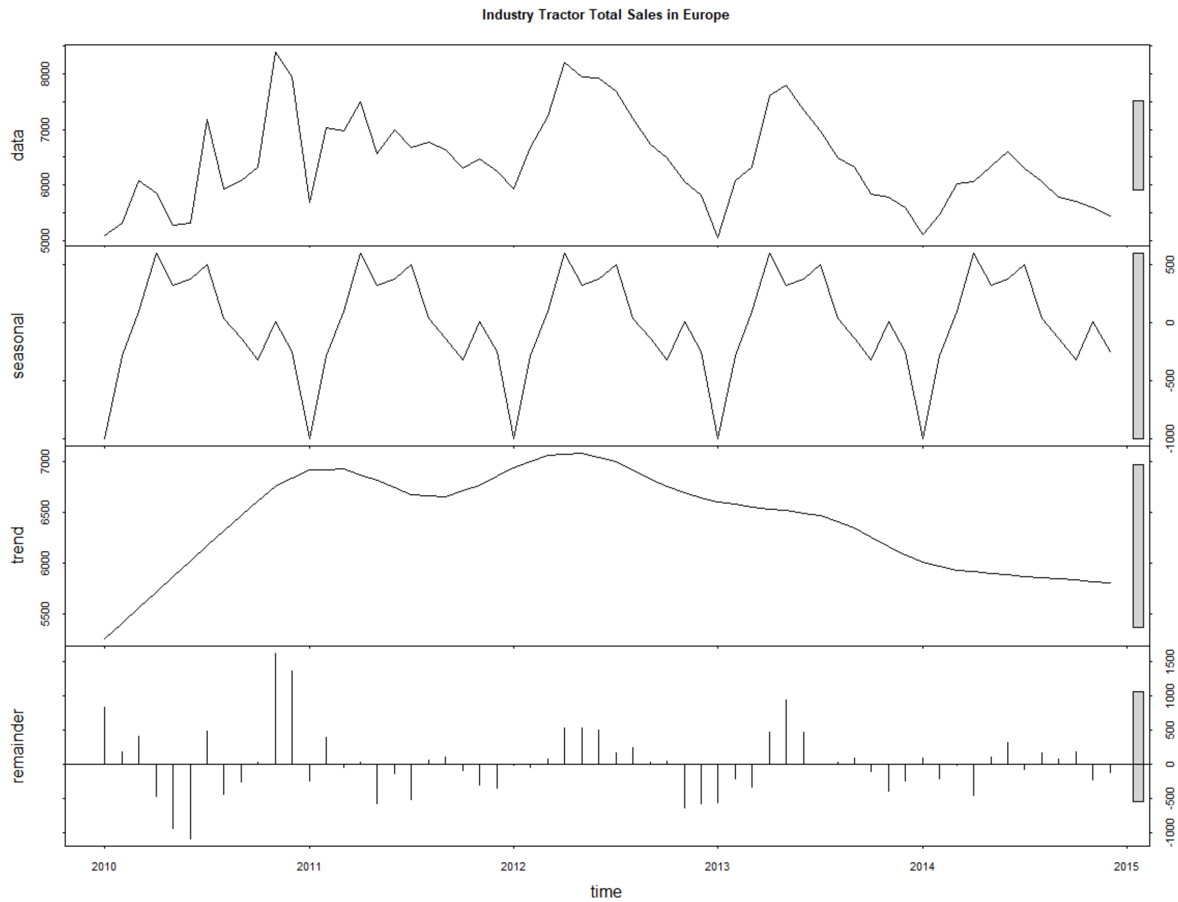
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	5090.909	5309.735	6071.429	5855.856	5272.727	5315.315	7169.811	5925.926	6074.766	6320.755	8380.952	7943.925
2011	5688.073	7037.037	6981.132	7500.000	6571.429	6990.291	6666.667	6761.905	6634.615	6310.680	6476.190	6250.000
2012	5922.330	6666.667	7227.723	8200.000	7941.176	7920.792	7676.768	7200.000	6734.694	6494.845	6060.606	5816.327
2013	5050.505	6082.474	6326.531	7604.167	7789.474	7346.939	6979.167	6489.362	6315.789	5833.333	5789.474	5591.398
2014	5106.383	5473.684	6021.505	6063.830	6344.086	6593.407	6304.348	6063.830	5789.474	5698.925	5604.396	5444.444



Graph 189: The time series graph above shows the tractor industry sales in Europe from January 2010 to December 2014.

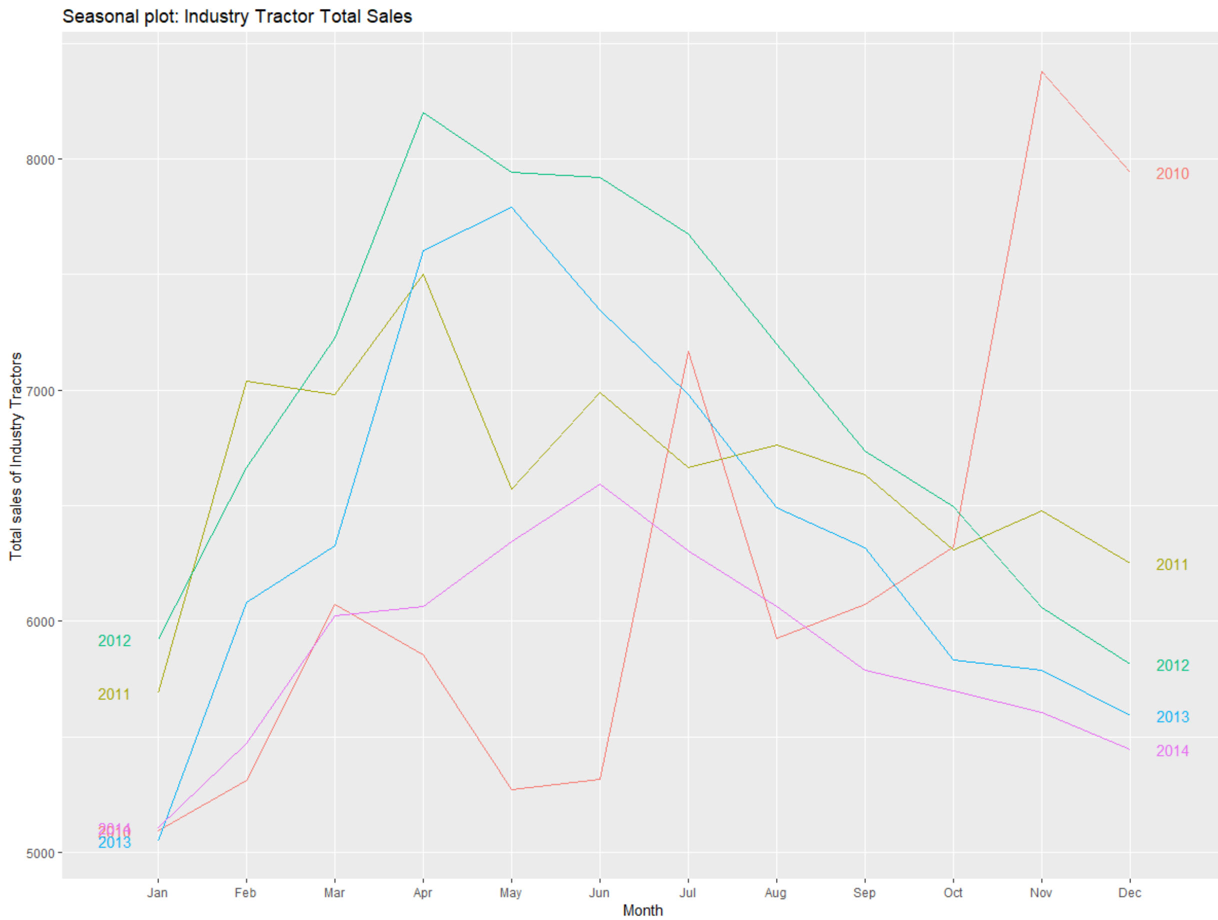
From the graph, there is no trend in the data over this period. However, there is also a slight presence of seasonality, especially in the years 2012-2014. Industry sales increase to a peak in mid year before decreasing again.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 190: Decomposed time series data of tractor industry sales in Europe from January 2010 to December 2014

Decomposition of the time series further proves the point that there is no trend

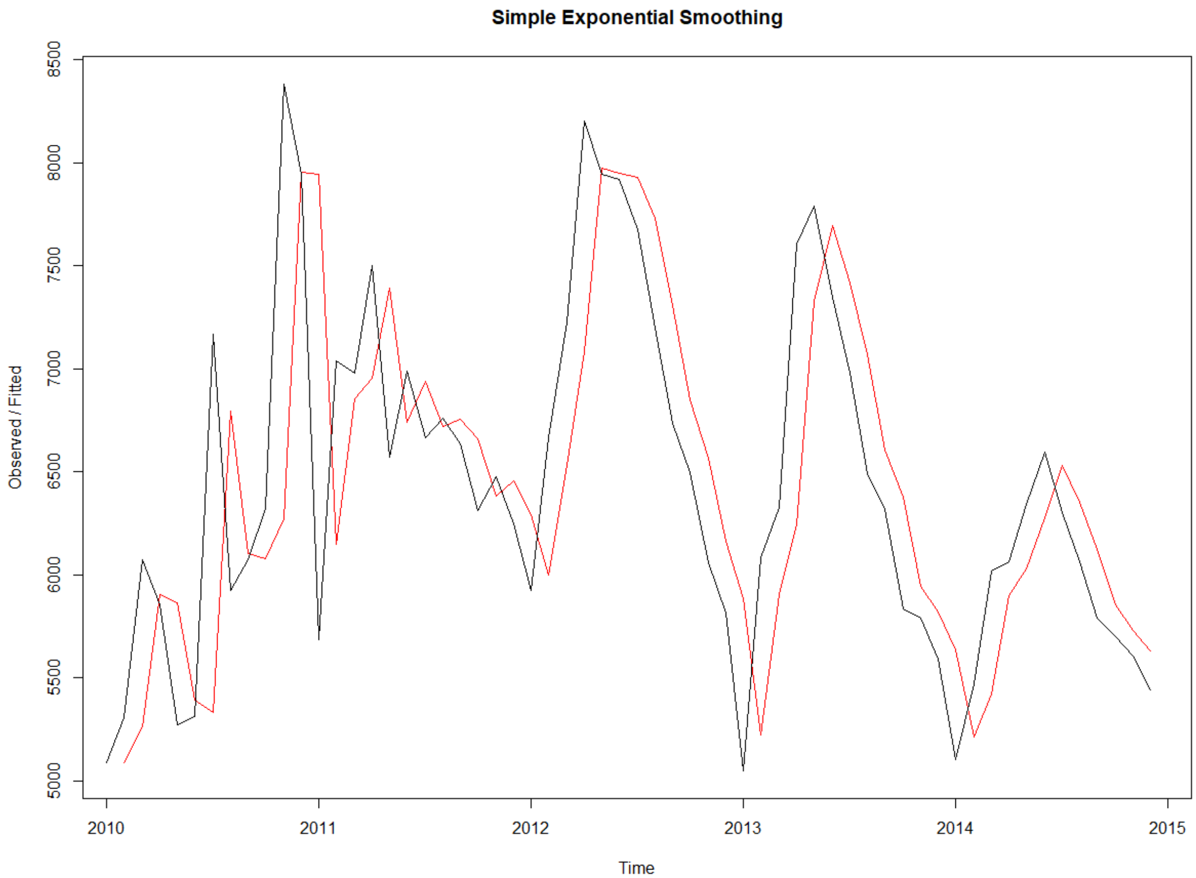


Graph 191: A seasonal plot of tractor industry sales in Europe from January 2010 to December 2014.

From the seasonal plot, there seems to be no presence of seasonality

Hence, we can conclude that the industry sales of tractors in Europe possess no trend and seasonality.

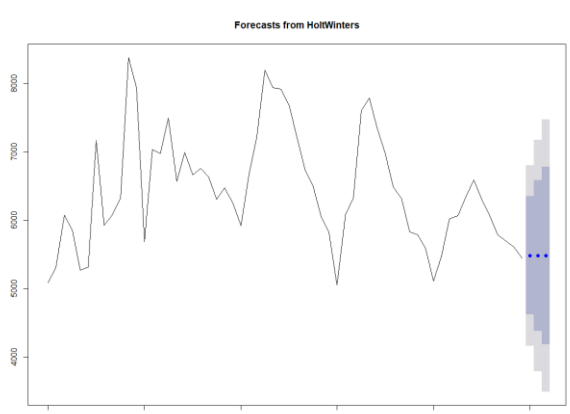
Thus, we can use Simple Exponential Smoothing to forecast future sales.

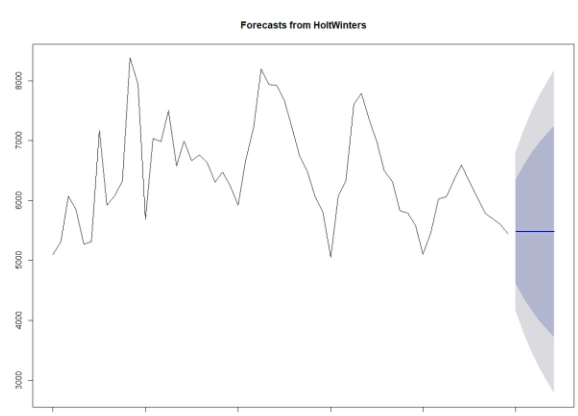


Graph 192: Observed time series data of tractor industry sales in Europe against the fitted Simple Exponential Smoothing model

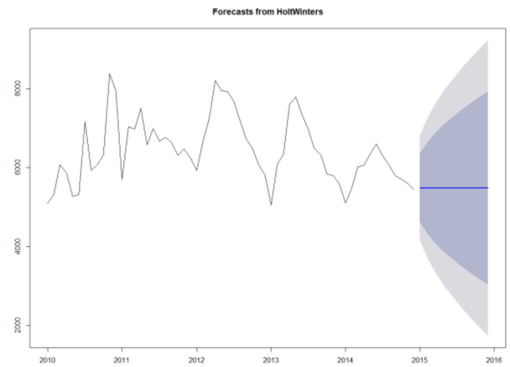
The model fits well with the observed time series data.

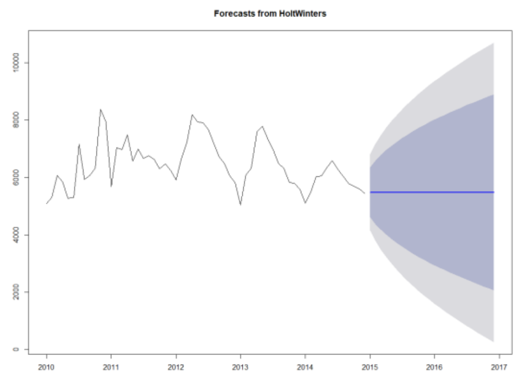
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months.





Graphs 193 and 194: Forecasted data of industry tractor sales in Europe over the next 3 and 6 months respectively.





Graphs 195 and 196: Forecasted data of industry tractor sales in Europe over the next 12 and 24 months respectively.

The following are the forecasted values of tractor industry sales in Europe over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	5482.307	4617.245	6347.369	4159.3094	6805.305
Feb 2015	5482.307	4376.533	6588.081	3791.1722	7173.442
Mar 2015	5482.307	4179.564	6785.050	3489.9338	7474.681
Apr 2015	5482.307	4008.692	6955.923	3228.6067	7736.008
May 2015	5482.307	3855.671	7108.944	2994.5812	7970.033
Jun 2015	5482.307	3715.856	7248.759	2780.7531	8183.861
Jul 2015	5482.307	3586.324	7378.291	2582.6505	8381.964
Aug 2015	5482.307	3465.092	7499.523	2397.2426	8567.372
Sep 2015	5482.307	3350.744	7613.870	2222.3628	8742.252
Oct 2015	5482.307	3242.226	7722.389	2056.3983	8908.216
Nov 2015	5482.307	3138.727	7825.888	1898.1104	9066.504
Dec 2015	5482.307	3039.609	7925.005	1746.5234	9218.091
Jan 2016	5482.307	2944.360	8020.254	1600.8519	9363.763
Feb 2016	5482.307	2852.558	8112.056	1460.4532	9504.161
Mar 2016	5482.307	2763.855	8200.760	1324.7931	9639.821
Apr 2016	5482.307	2677.956	8286.659	1193.4218	9771.193
May 2016	5482.307	2594.611	8370.004	1065.9566	9898.658
Jun 2016	5482.307	2513.605	8451.010	942.0686	10022.546
Jul 2016	5482.307	2434.751	8529.863	821.4724	10143.142
Aug 2016	5482.307	2357.887	8606.728	703.9188	10260.696
Sep 2016	5482.307	2282.869	8681.746	589.1886	10375.426
Oct 2016	5482.307	2209.570	8755.045	477.0875	10487.527
Nov 2016	5482.307	2137.877	8826.737	367.4427	10597.172
Dec 2016	5482.307	2067.689	8896.925	260.0995	10704.515

Conclusion:

As seen from the forecasted graphs and values generated, industry tractor sales in Europe is predicted to remain stagnant at 5482 for the next 24 months to come.

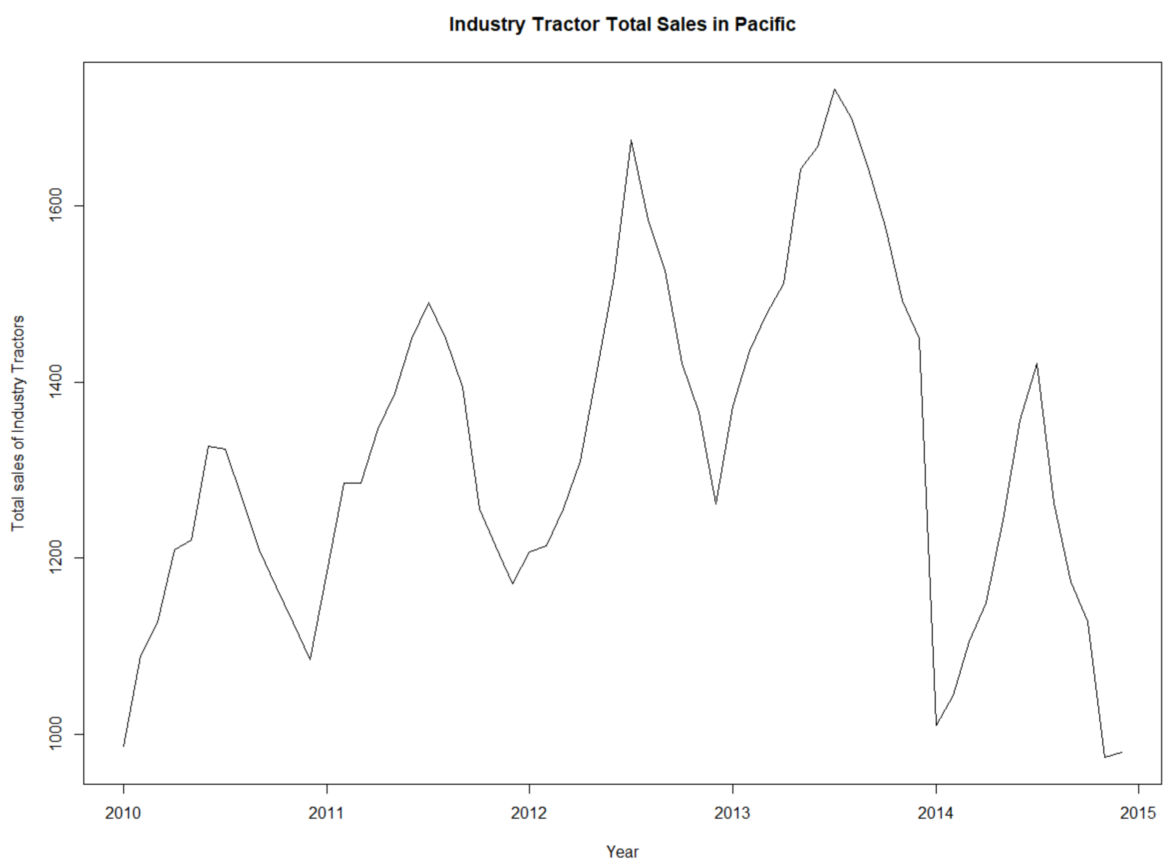
Future Predictions of Market Share in Europe:
Both forecasted tractor sales (Industry and Unit) in Europe had similar trend. Both predicted a stagnant sales value for the next 24 months with slight/little change/

Since *market share = unit sales/industry sales*, there will be no predicted change in market shares in Europe for the sales of tractor.

Forecasting of Industry Sales of Tractors in Pacific

Time Series:

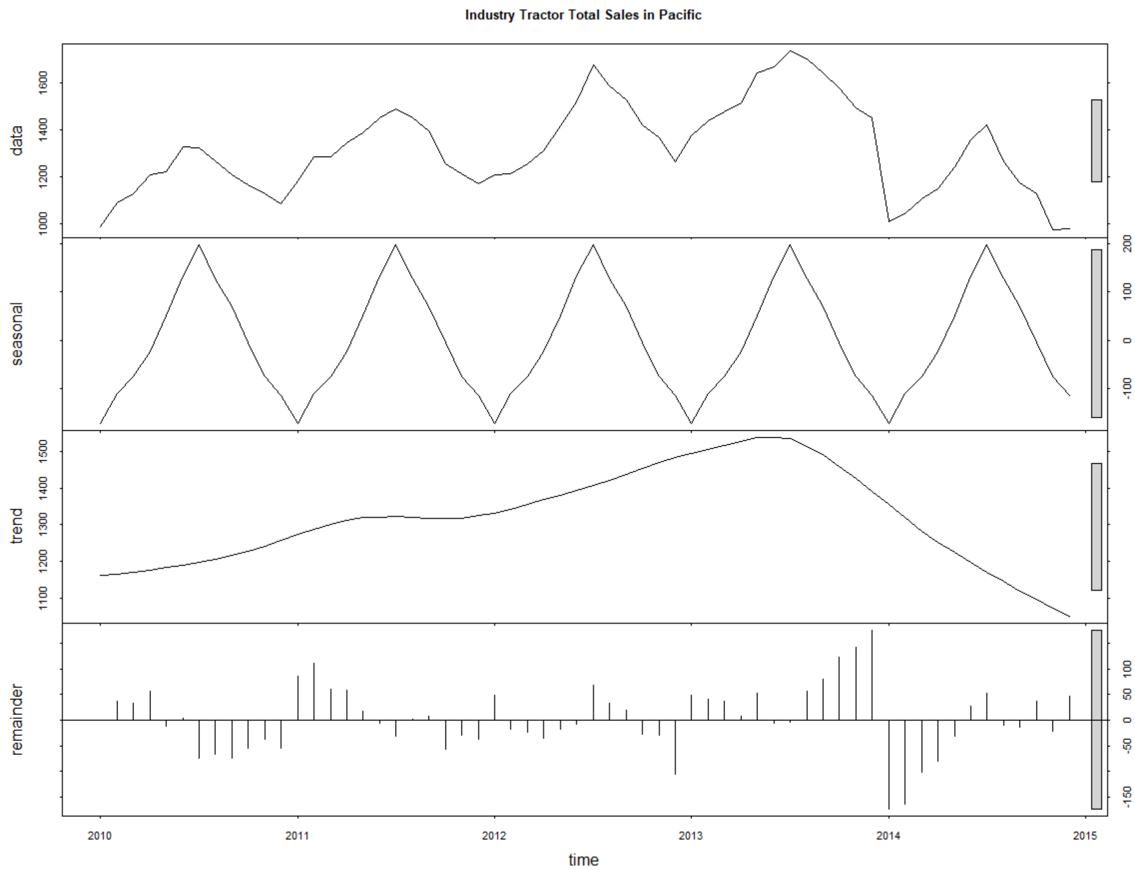
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	987.0000	1090.0474	1126.7606	1209.3023	1220.6573	1327.0142	1324.2009	1267.6056	1209.3023	1168.2243	1126.7606	1084.9057
2011	1184.8341	1285.7143	1285.7143	1346.1538	1387.5598	1449.2754	1490.3846	1449.2754	1394.2308	1256.0386	1213.5922	1170.7317
2012	1207.7295	1213.5922	1256.0386	1310.6796	1414.6341	1519.6078	1674.8768	1584.1584	1527.0936	1421.5686	1365.8537	1262.1359
2013	1372.5490	1435.6436	1477.8325	1512.1951	1641.7910	1666.6667	1732.6733	1700.0000	1641.7910	1576.3547	1492.5373	1450.0000
2014	1010.1010	1044.7761	1105.5276	1150.0000	1243.7811	1356.7839	1421.3198	1262.6263	1173.4694	1128.2051	974.3590	979.3814



Graph 197: The time series graph above shows the tractor industry sales in Pacific from January 2010 to December 2014

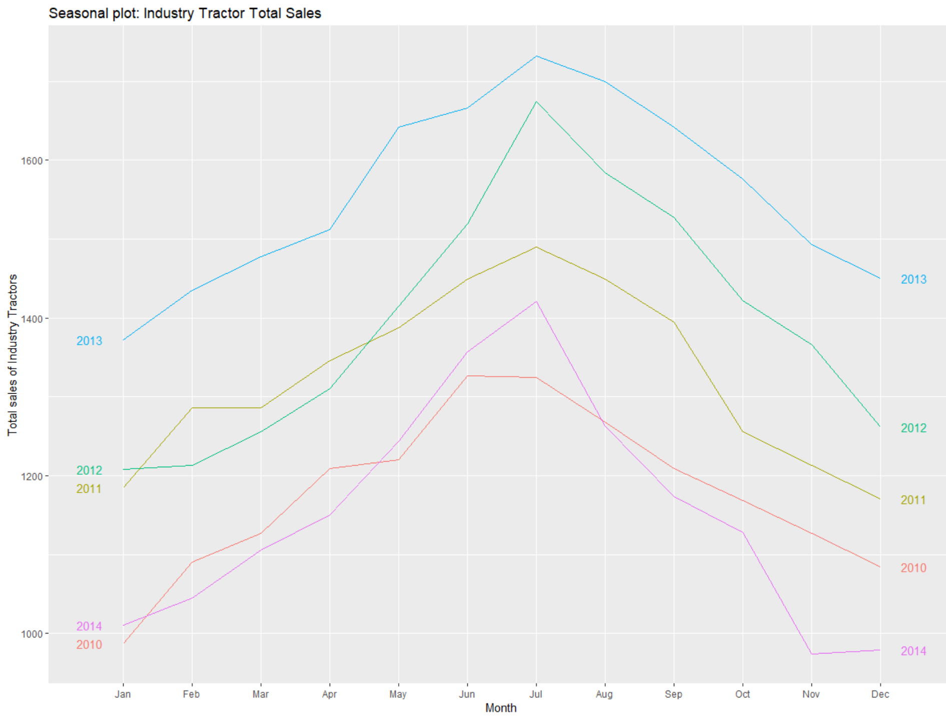
From the graph, there is no trend in the data over this period. However, there is a presence of seasonality where sales will peak in the middle of the year before dipping again.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 198: Decomposed time series data of tractor industry sales in Pacific from January 2010 to December 2014

Decomposition of the time series further proves the point that there is an upward trend overall, except for anomaly in the year 2014.

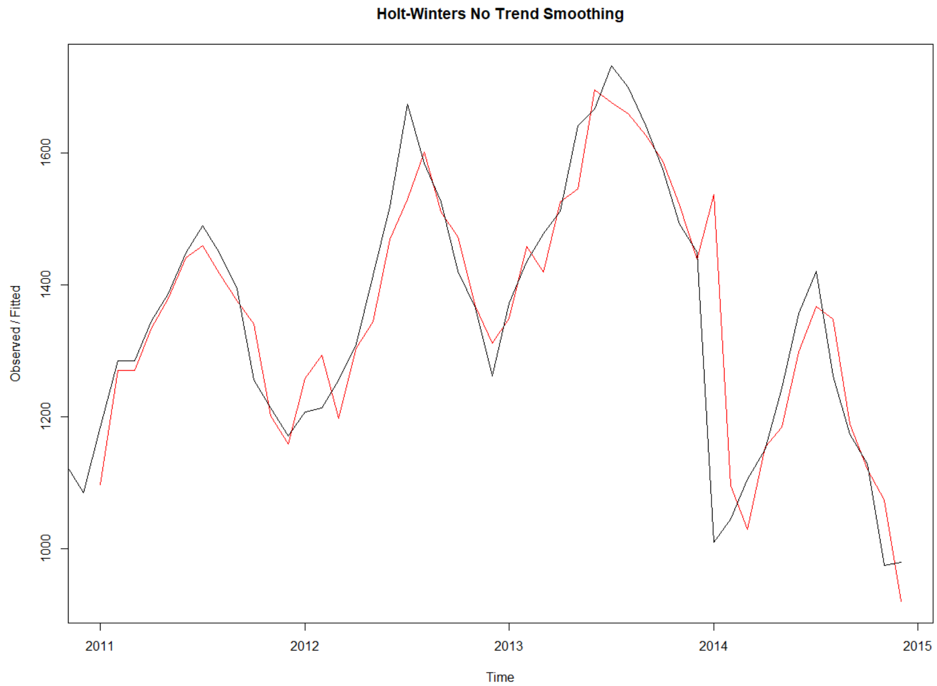


Graph 199: A seasonal plot of tractor industry sales in Pacific from January 2010 to December 2014.

From the seasonal plot, there seems to be the presence of seasonality due to constant fluctuations at a specific time of every year. We can deduce that multiplicative decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations varies with the level of the time series.

Hence, we can conclude that the industry sales of tractors in Pacific possess no trend but has seasonality.

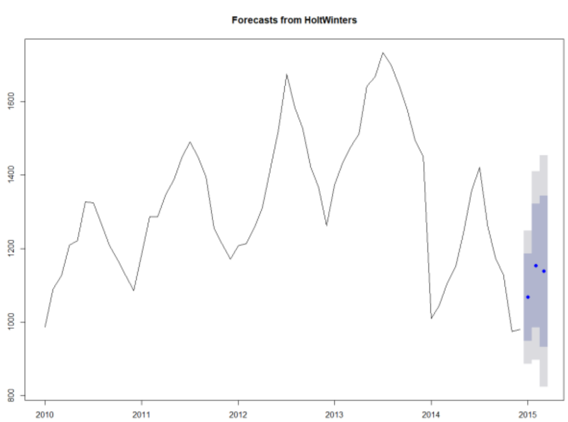
Thus, we can use Holt-Winters No-Trend Smoothing Model to forecast future sales.

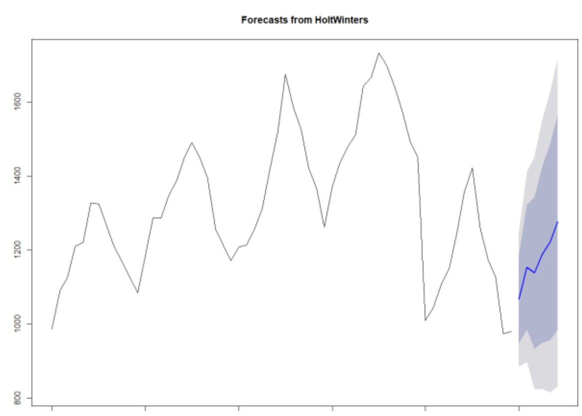


Graph 200: Observed time series data of tractor industry sales in Pacific against the fitted Holt-Winters No-Trend Smoothing Model

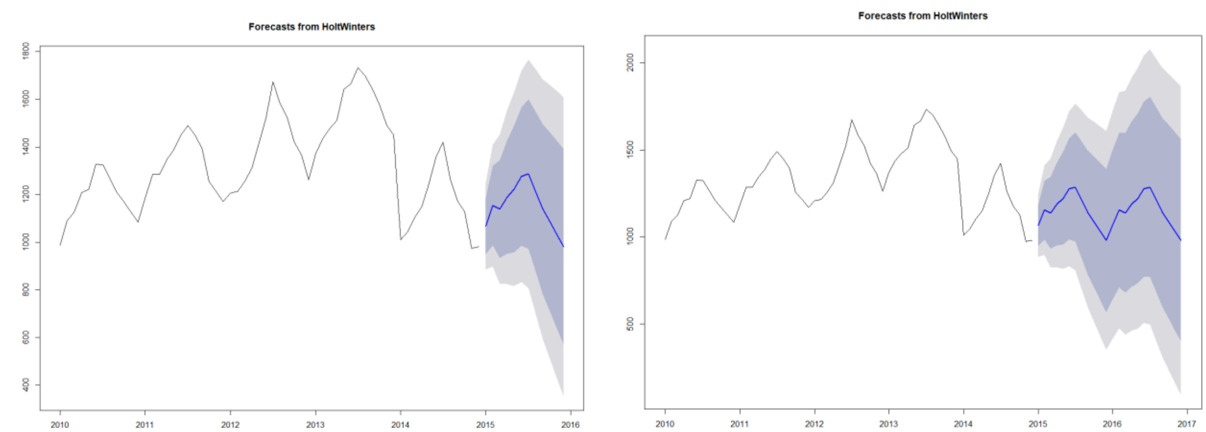
The model fits well with the observed time series data.

Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months.





Graphs 201 and 202: Forecasted data of industry tractor sales in Pacific over the next 3 and 6 months respectively.



Graphs 203 and 204: Forecasted data of industry tractor sales in Pacific over the next 12 and 24 months respectively.

The following are the forecasted values of tractor industry sales in Pacific over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	1067.2914	948.5949	1185.988	885.76079	1248.822
Feb 2015	1153.6776	985.8155	1321.540	896.95463	1410.401
Mar 2015	1138.4027	932.8144	1343.991	823.98252	1452.823
Apr 2015	1187.4780	950.0851	1424.871	824.41683	1550.539
May 2015	1221.6070	956.1937	1487.020	815.69231	1627.522
Jun 2015	1276.1285	985.3827	1566.874	831.47122	1720.786
Jul 2015	1286.8786	972.8372	1600.920	806.59378	1767.163
Aug 2015	1213.8874	878.1631	1549.612	700.44137	1727.333
Sep 2015	1140.8082	784.7188	1496.898	596.21648	1685.400
Oct 2015	1087.4050	712.0538	1462.756	513.35488	1661.455
Nov 2015	1033.2848	639.6132	1426.956	431.21600	1635.354
Dec 2015	979.3814	568.2048	1390.558	350.54106	1608.222
Jan 2016	1067.2914	639.3252	1495.258	412.77354	1721.809
Feb 2016	1153.6776	709.5561	1597.799	474.45239	1832.903
Mar 2016	1138.4027	678.6933	1598.112	435.33779	1841.468
Apr 2016	1187.4780	712.6921	1662.264	461.35566	1913.600
May 2016	1221.6070	732.2090	1711.005	473.13727	1970.077
Jun 2016	1276.1285	772.5421	1779.715	505.95948	2046.298
Jul 2016	1286.8786	769.4926	1804.264	495.60509	2078.152
Aug 2016	1213.8874	683.0607	1744.714	402.05795	2025.717
Sep 2016	1140.8082	596.8727	1684.744	308.93059	1972.686
Oct 2016	1087.4050	530.6692	1644.141	235.95108	1938.859
Nov 2016	1033.2848	464.0366	1602.533	162.69474	1903.875
Dec 2016	979.3814	397.8899	1560.873	90.06684	1868.696

Conclusion:

As seen from the forecasted graphs and values generated, industry tractor sales in Pacific is predicted to remain around the same values of sales as in 2014 for the next 24 months to come. Seasonality remains.

Future Predictions of Market Share in Pacific:

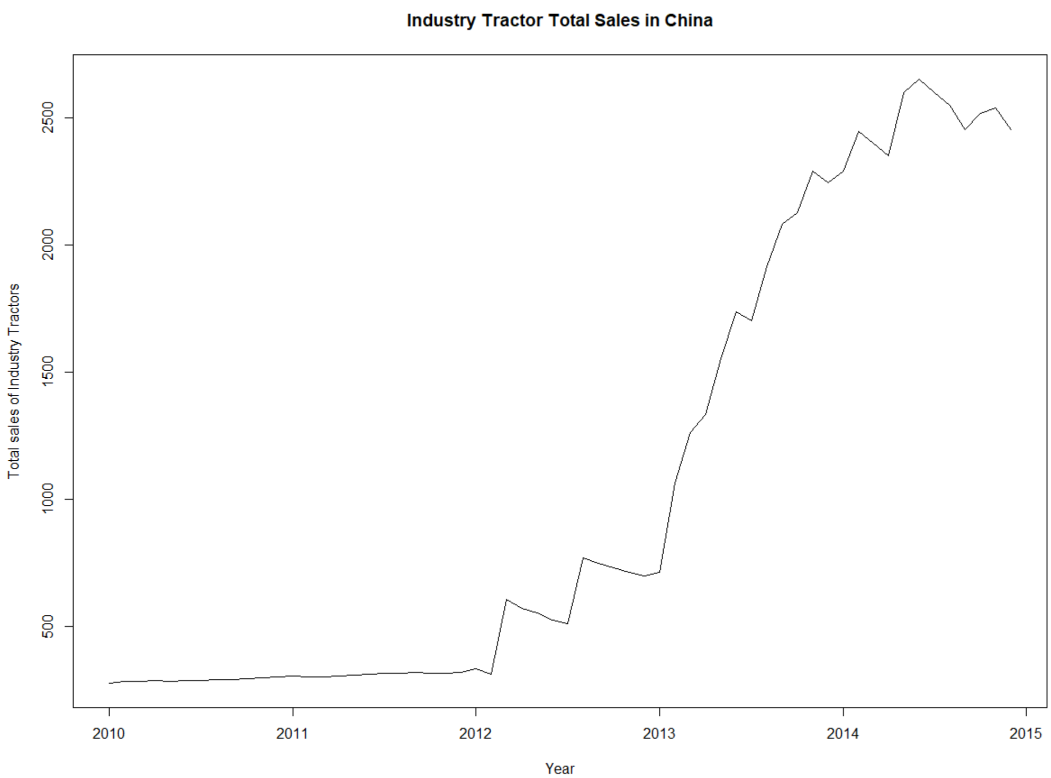
Both forecasted tractor sales (Industry and Unit) in Pacific had similar trends. Both saw forecasted sales stabilizing and hovering around similar values as of 2014.

Since *market share = unit sales/industry sales*, there will be no predicted change in market shares in Pacific for the sales of tractors.

Forecasting of Industry Sales of Tractors in China

Time Series:

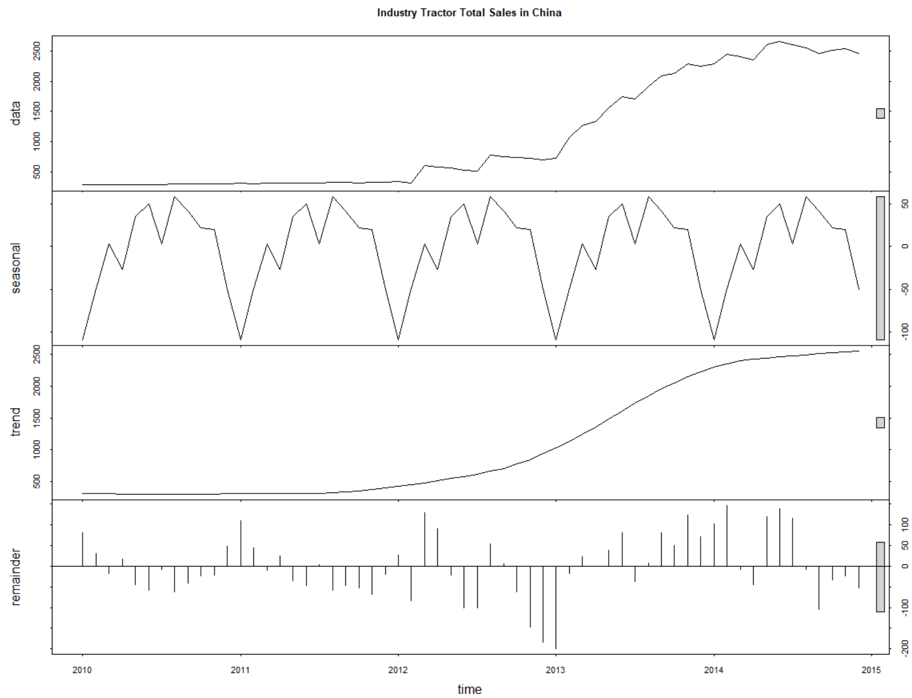
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	278.0000	283.0000	285.0000	288.0000	286.0000	287.0000	289.0000	290.0000	293.0000	295.0000	298.0000	301.0000
2011	306.0000	302.0000	303.0000	307.0000	309.0000	312.0000	315.0000	318.0000	321.0000	315.0000	318.0000	320.0000
2012	333.3333	312.5000	606.0606	571.4286	555.5556	526.3158	512.8205	769.2308	750.0000	731.7073	714.2857	697.6744
2013	714.2857	1063.0000	1264.0000	1333.3333	1555.5556	1739.1304	1702.1277	1914.8936	2083.3333	2127.6596	2291.6667	2244.8980
2014	2291.6667	2448.9796	2400.0000	2352.9412	2600.0000	2653.0612	2600.0000	2549.0196	2452.8302	2517.0000	2541.0000	2452.8302



Graph 205: The time series graph above shows the tractor industry sales in China from January 2010 to December 2014

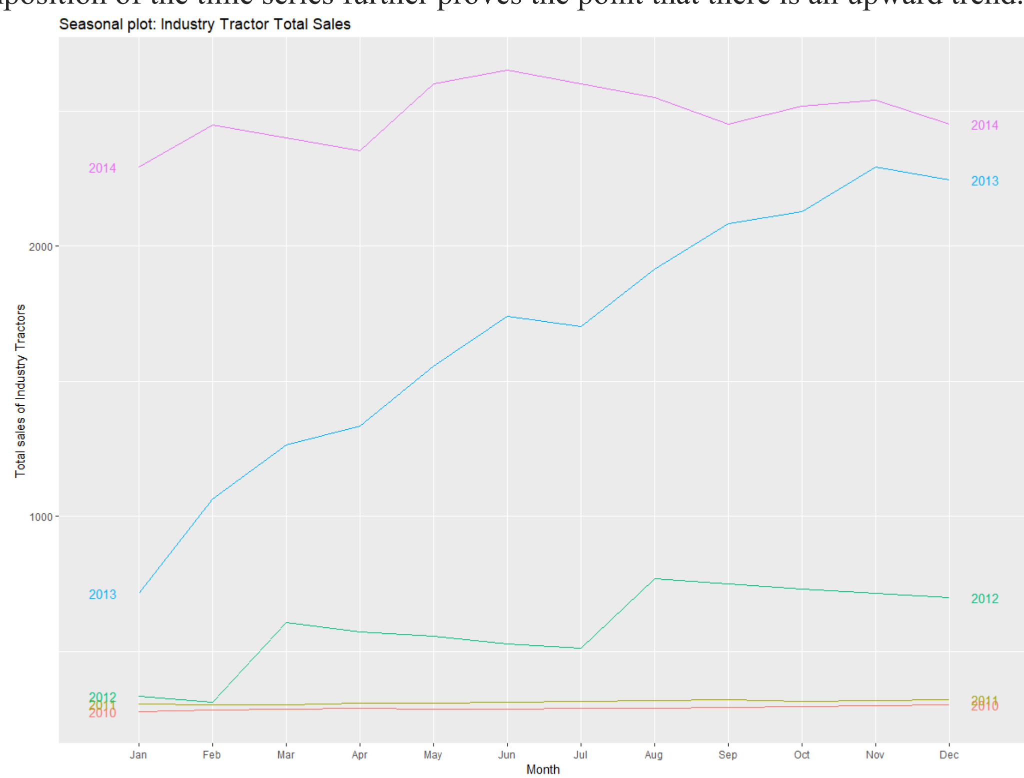
From the graph, there is an upward trend in the data from 2012 to 2014. However, there is no presence of seasonality.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 206: Decomposed time series data of tractor industry sales in China from January 2010 to December 2014

Decomposition of the time series further proves the point that there is an upward trend.

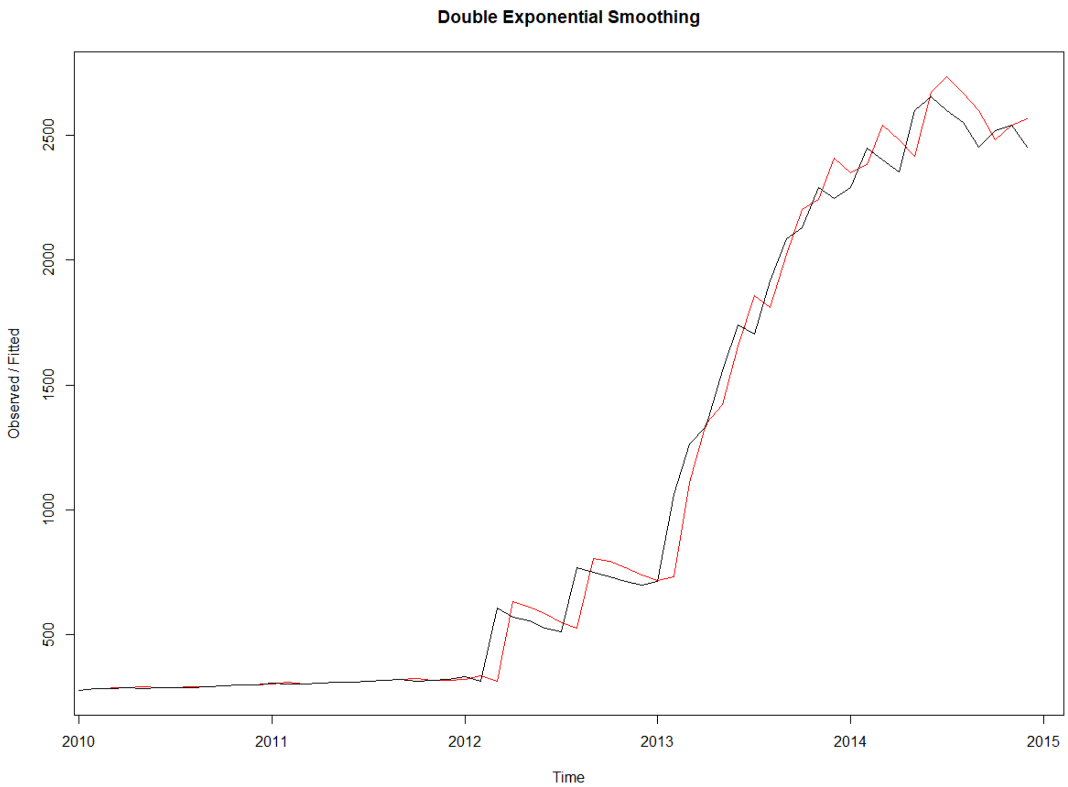


Graph 207: A seasonal plot of tractor industry sales in China from January 2010 to December 2014.

From the seasonal plot, there seems to be no presence of seasonality

Hence, we can conclude that the industry sales of tractors in China possess trend but no seasonality.

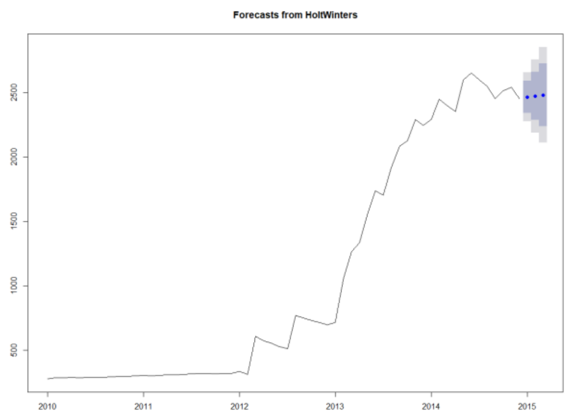
Thus, we can use Double Exponential Smoothing Model to forecast future sales.

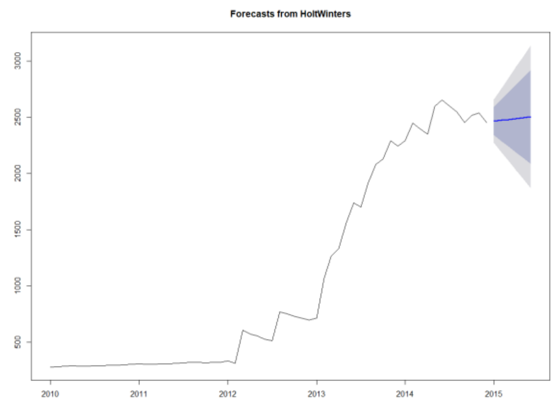


Graph 208: Observed time series data of mower industry sales in China against the fitted Double Exponential Smoothing model.

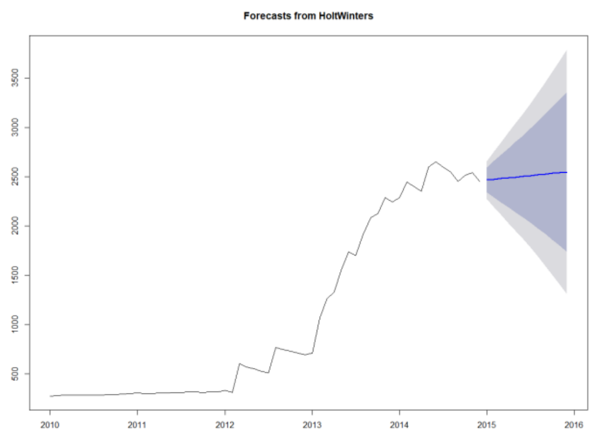
The model fits well with the observed time series data.

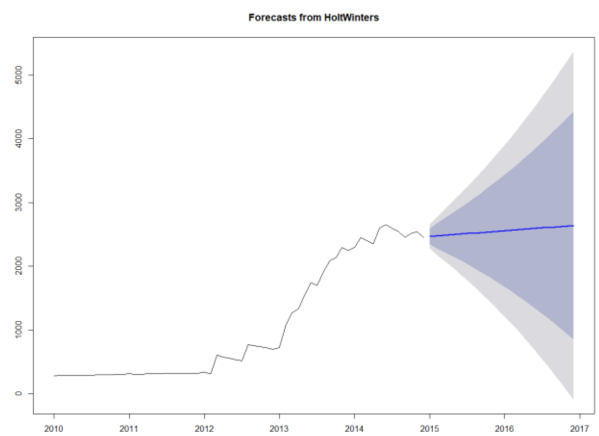
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months.





Graphs 209 and 210: Forecasted data of industry tractor sales in China over the next 3 and 6 months respectively.





Graphs 211 and 212: Forecasted data of industry mower sales in China over the next 12 and 24 months respectively.

The following are the forecasted values of tractor industry sales in China over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	2466.901	2341.5199	2592.282	2275.14718	2658.655
Feb 2015	2474.390	2288.4120	2660.368	2189.96133	2758.818
Mar 2015	2481.879	2238.7229	2725.034	2110.00409	2853.753
Apr 2015	2489.367	2189.2424	2789.492	2030.36580	2948.369
May 2015	2496.856	2138.8303	2854.882	1949.30284	3044.409
Jun 2015	2504.345	2086.9830	2921.707	1866.04497	3142.645
Jul 2015	2511.834	2033.4571	2990.210	1780.21992	3243.447
Aug 2015	2519.322	1978.1332	3060.511	1691.64501	3347.000
Sep 2015	2526.811	1920.9567	3132.665	1600.23678	3453.385
Oct 2015	2534.300	1861.9090	3206.691	1505.96682	3562.633
Nov 2015	2541.789	1800.9923	3282.585	1408.83848	3674.739
Dec 2015	2549.277	1738.2210	3360.334	1308.87374	3789.681
Jan 2016	2556.766	1673.6166	3439.915	1206.10556	3907.426
Feb 2016	2564.255	1607.2048	3521.305	1100.57320	4027.936
Mar 2016	2571.743	1539.0135	3604.473	992.31934	4151.168

Apr 2016	2579.232	1469.0717	3689.393	881.38824	4277.076
May 2016	2586.721	1397.4086	3776.033	767.82459	4405.617
Jun 2016	2594.210	1324.0531	3864.366	651.67277	4536.747
Jul 2016	2601.698	1249.0338	3954.363	532.97635	4670.421
Aug 2016	2609.187	1172.3785	4045.996	411.77788	4806.597
Sep 2016	2616.676	1094.1141	4139.238	288.11862	4945.233
Oct 2016	2624.165	1014.2669	4234.062	162.03853	5086.291
Nov 2016	2631.653	932.8620	4330.445	33.57617	5229.731
Dec 2016	2639.142	849.9238	4428.361	-97.23126	5375.516

Conclusion:
As seen from the forecasted graphs and values generated, industry mower sales in China is predicted to increase linearly and steadily for the next 24 months to come.

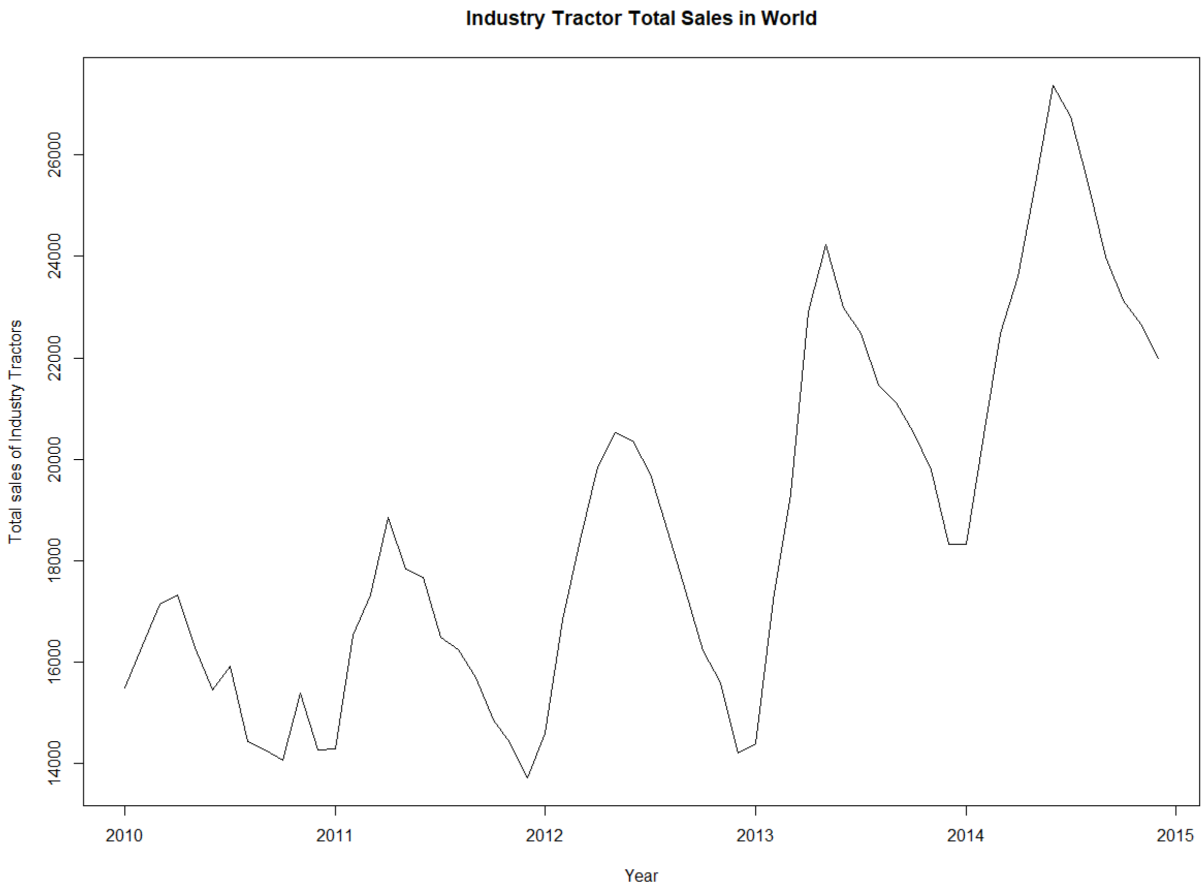
Future Predictions of Market Share in China:
Forecasted unit tractor sales in China saw a smaller rate of increase than industry sales. Forecasted values of industry tractor sales increased by the hundreds while forecasted values of the unit tractor sales increased by the ones. Both displayed linearity in their sales increase over the 24 months.

Since *market share = unit sales/industry sales*, there will be a predicted decrease in market shares in China for the sales of tractors.

Forecasting of Industry Sales of Tractors in the World

Time Series:

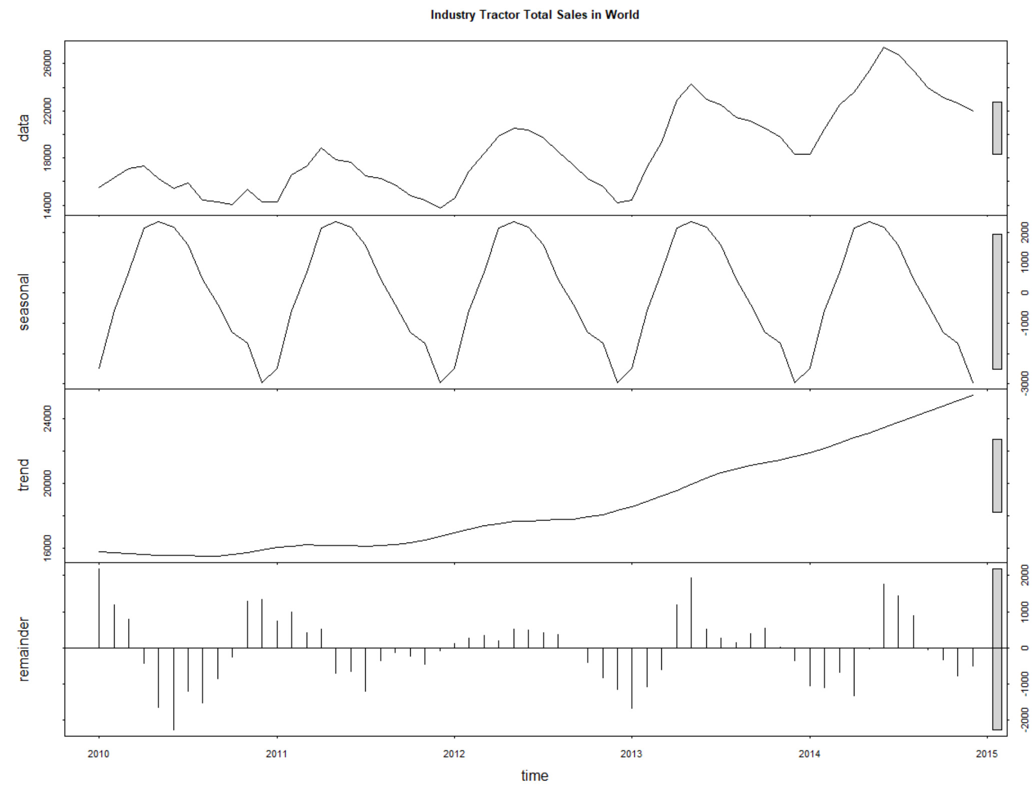
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	15482.77	16324.91	17128.95	17327.14	16277.55	15448.20	15905.03	14422.49	14258.45	14061.31	15378.16	14272.29
2011	14288.57	16530.13	17319.94	18841.80	17826.12	17669.19	16486.58	16250.20	15692.40	14844.19	14401.93	13716.25
2012	14597.07	16836.38	18411.99	19851.69	20512.75	20354.53	19716.05	18574.99	17394.32	16226.29	15587.29	14206.64
2013	14394.28	17217.84	19310.35	22901.15	24243.67	22992.88	22482.96	21464.63	21122.71	20522.71	19789.32	18328.96
2014	18310.73	20476.96	22489.24	23607.36	25438.75	27373.52	26763.63	25427.63	23995.25	23142.11	22666.19	21988.96



Graph 213: The time series graph above shows the tractor industry sales in the World from Janua ry 2010 to December 2014.

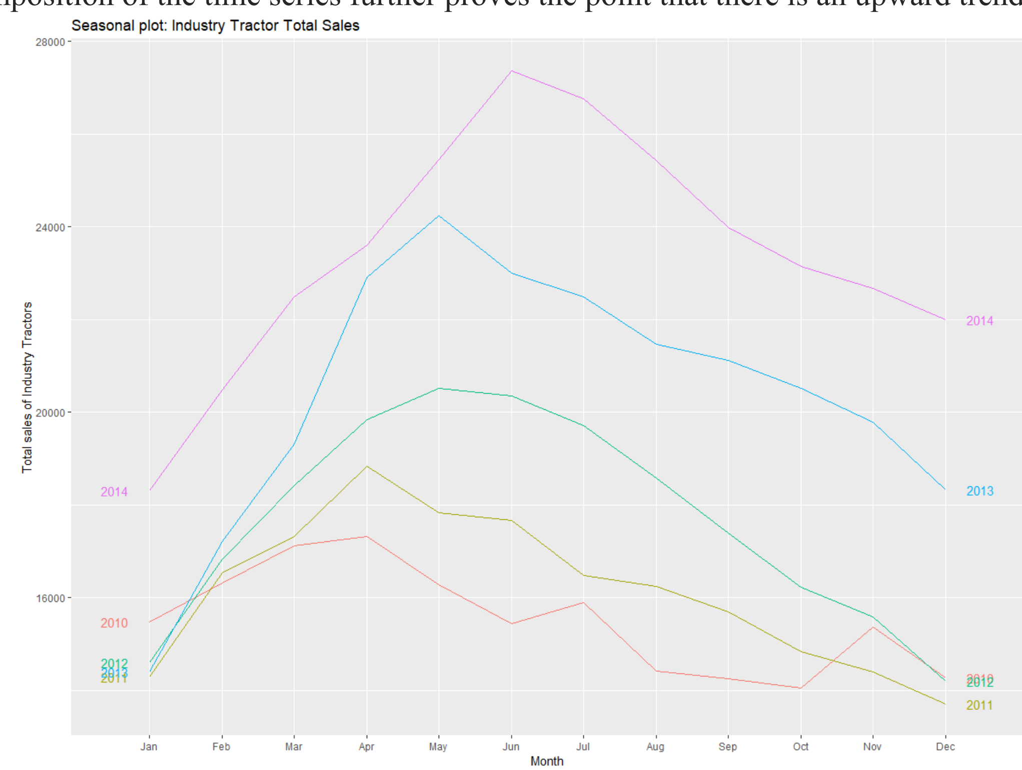
From the graph, there is an upwawrd trend in the data over this period.
There is also a slight presence of seasonality.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 214: Decomposed time series data of tractor industry sales in the World from January 2010 to December 2014

Decomposition of the time series further proves the point that there is an upward trend.

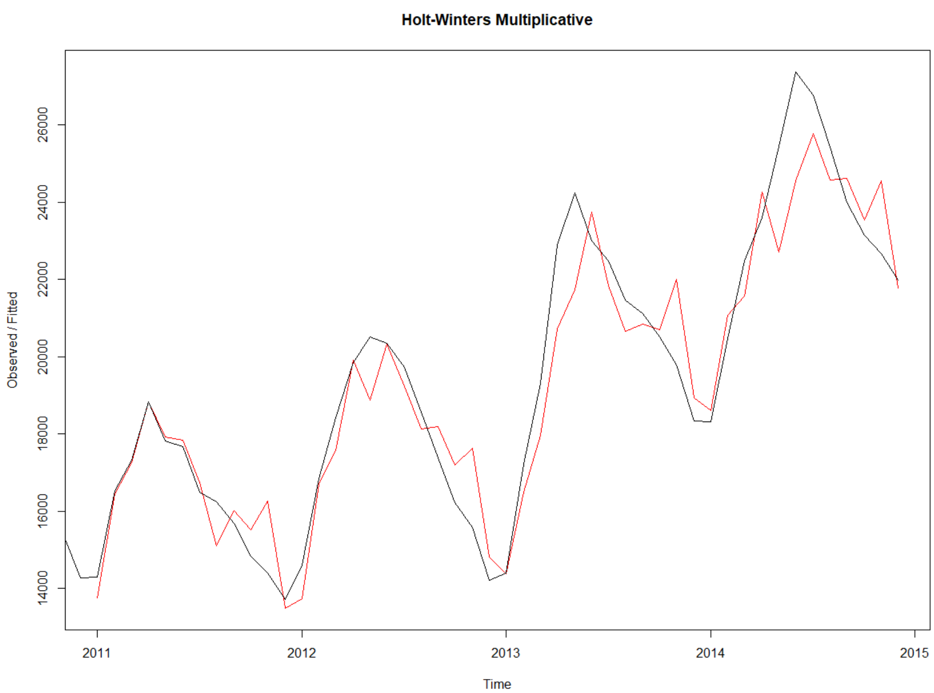


Graph 215: A seasonal plot of tractor industry sales in the World from January 2010 to December 2014.

From the seasonal plot, there seems to be the presence of seasonality due to constant fluctuations at a specific time of every year. We can deduce that multiplicative decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations varies with the level of the time series.

Hence, we can conclude that the industry sales of tractors in the World possess <u>trend and seasonality.</u>

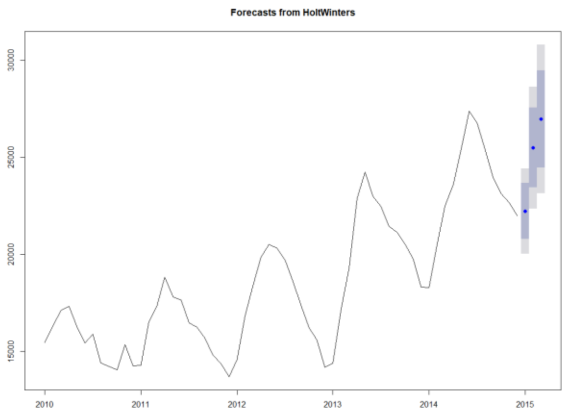
Thus, we can use <u>Holt-Winters Multiplicative Model</u>.

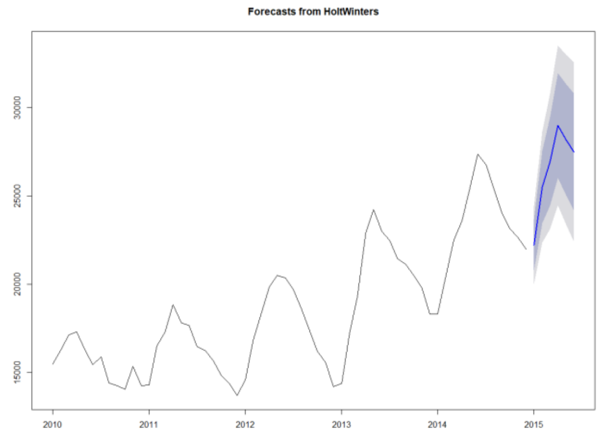


Graph 216: Observed time series data of tractor industry sales in the World against the fitted Holt-Winters multiplicative model

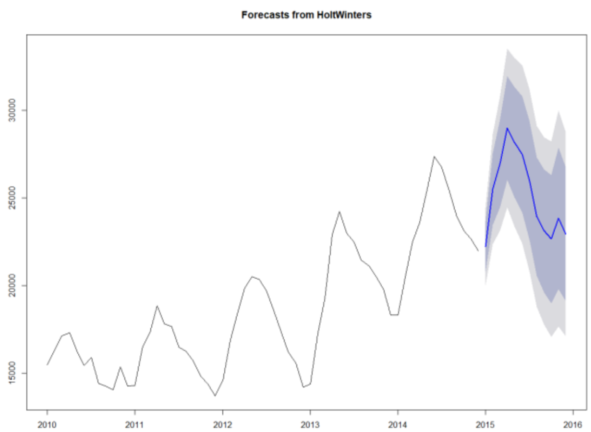
We see from the plot that the Holt-Winters multiplicative model is very successful in predicting the seasonal peaks, which occur every June/July of the year. The model fits well with the observed time series data.

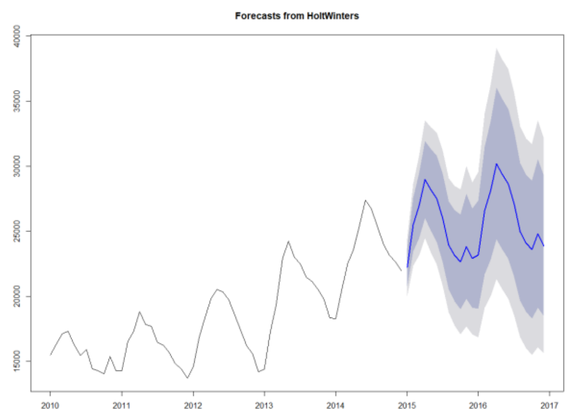
Hence, we can make use of Holt-Winter multiplicative model to predict data in the next 3,6,12 and 24 months:





Graphs 217 and 218: Forecasted data of industry tractor sales in the World over the next 3 and 6 months respectively.





Graphs 219 and 220: Forecasted data of industry tractor sales in the World over the next 12 and 24 months respectively.

The following are the forecasted values of tractor industry sales in the World over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	22237.44	20803.97	23670.91	20045.13	24429.75
Feb 2015	25503.72	23453.35	27554.09	22367.95	28639.49
Mar 2015	26969.51	24465.84	29473.18	23140.48	30798.55
Apr 2015	28988.04	26025.57	31950.50	24457.34	33518.74
May 2015	28182.16	25034.52	31329.80	23368.26	32996.05
Jun 2015	27487.53	24166.75	30808.32	22408.83	32566.24
Jul 2015	25993.72	22604.86	29382.58	20810.90	31176.53
Aug 2015	23949.49	20574.39	27324.59	18787.72	29111.25
Sep 2015	23127.37	19630.94	26623.80	17780.04	28474.71
Oct 2015	22657.70	19008.39	26307.01	17076.56	28238.84
Nov 2015	23838.71	19808.11	27869.30	17674.44	30002.97
Dec 2015	22931.81	19116.79	26746.84	17097.24	28766.39
Jan 2016	23187.55	19035.03	27340.08	16836.82	29538.29
Feb 2016	26589.53	21699.19	31479.87	19110.40	34068.65
Mar 2016	28113.66	22811.78	33415.54	20005.14	36222.18
Apr 2016	30213.49	24402.13	36024.84	21325.79	39101.19
May 2016	29369.36	23590.75	35147.97	20531.73	38206.99

```
Jun 2016       28641.42 22876.39 34406.45 19824.57 37458.27
Jul 2016       27081.09 21490.77 32671.42 18531.43 35630.75
Aug 2016       24947.87 19645.08 30250.66 16837.95 33057.79
Sep 2016       24088.14 18820.60 29355.69 16032.12 32144.16
Oct 2016       23595.71 18291.75 28899.67 15484.01 31707.42
Nov 2016       24822.22 19122.08 30522.36 16104.61 33539.83
Dec 2016       23874.67 18479.41 29269.93 15623.34 32126.00
```

Conclusion:
As seen from the forecasted graphs and values generated, industry tractor sales in the World is predicted to increase steadily for the next 24 months to come. Seasonality will remain.

Future Predictions of Market Share in the World:

Both forecasted tractor sales (Industry and Unit) in the World had similar trends. Both saw an increase in sales in the next 24 months. Meanwhile, seasonality remained where where the mower sales increases till mid-year before decreasing back to around the original volume of mower sales that was sold at the start of the year. However, the magnitude of increase in unit tractor sales was more than of industry's

Since *market share = unit sales/industry sales*, there will be a predicted increase in market shares in the World for the sales of tractors.

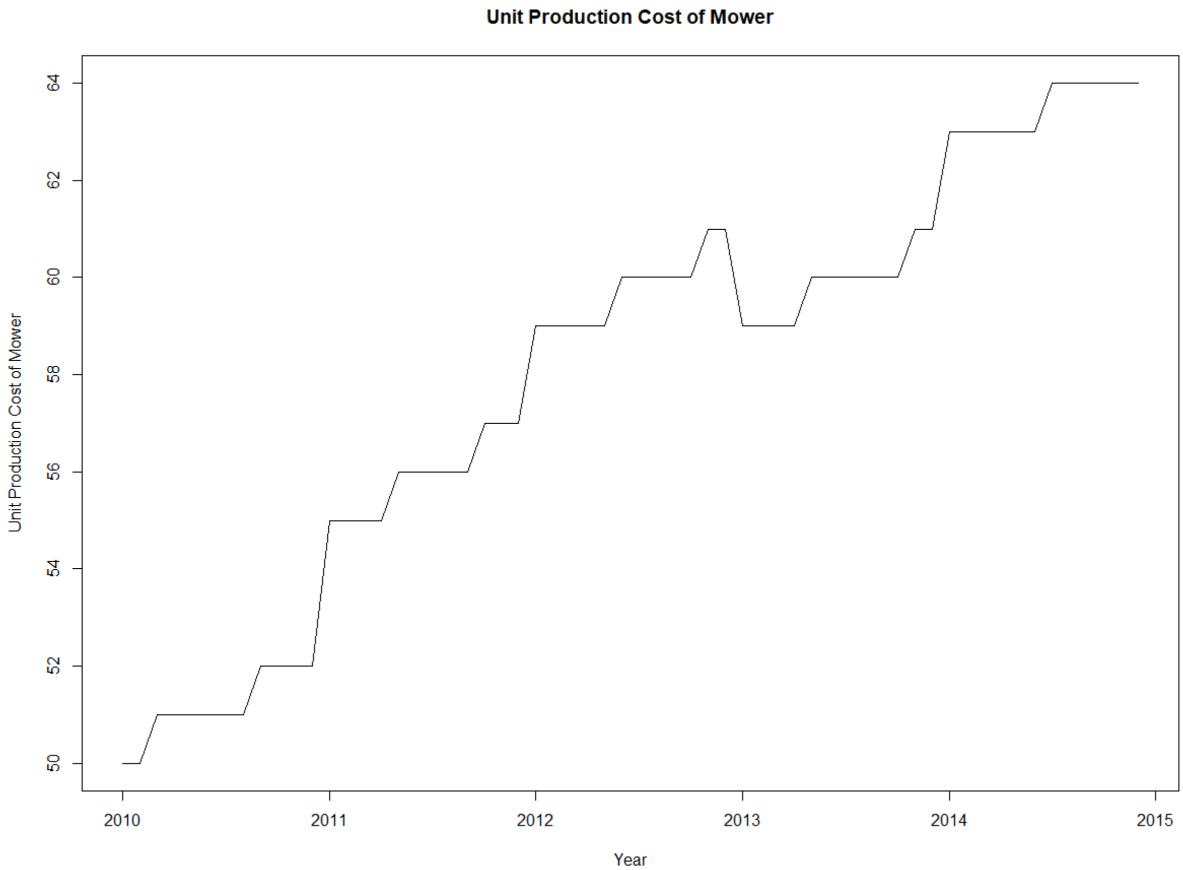
Lastly, we will proceed onto forecast future increases in **unit production costs of mowers** and tractors using the excel sheet *Unit Production Costs*.

Firstly, we will forecast the unit production costs of mowers.

Forecasting of Unit Production Costs of Mowers

Time Series:

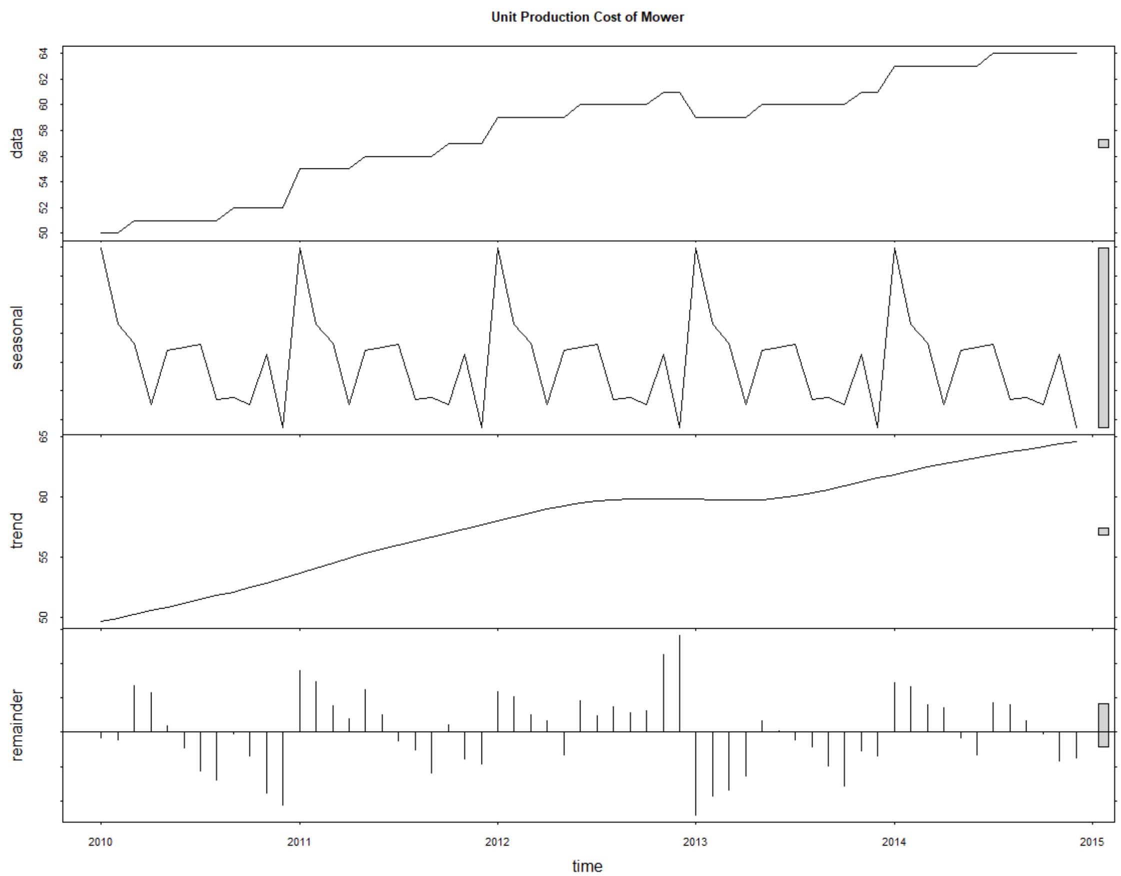
```
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2010  50  50  51  51  51  51  51  51  52  52  52  52
2011  55  55  55  55  56  56  56  56  56  57  57  57
2012  59  59  59  59  59  60  60  60  60  60  61  61
2013  59  59  59  59  60  60  60  60  60  60  61  61
2014  63  63  63  63  63  63  64  64  64  64  64  64
```



Graph 221: The time series graph above shows the unit production costs of mowers from January 2010 to December 2014.

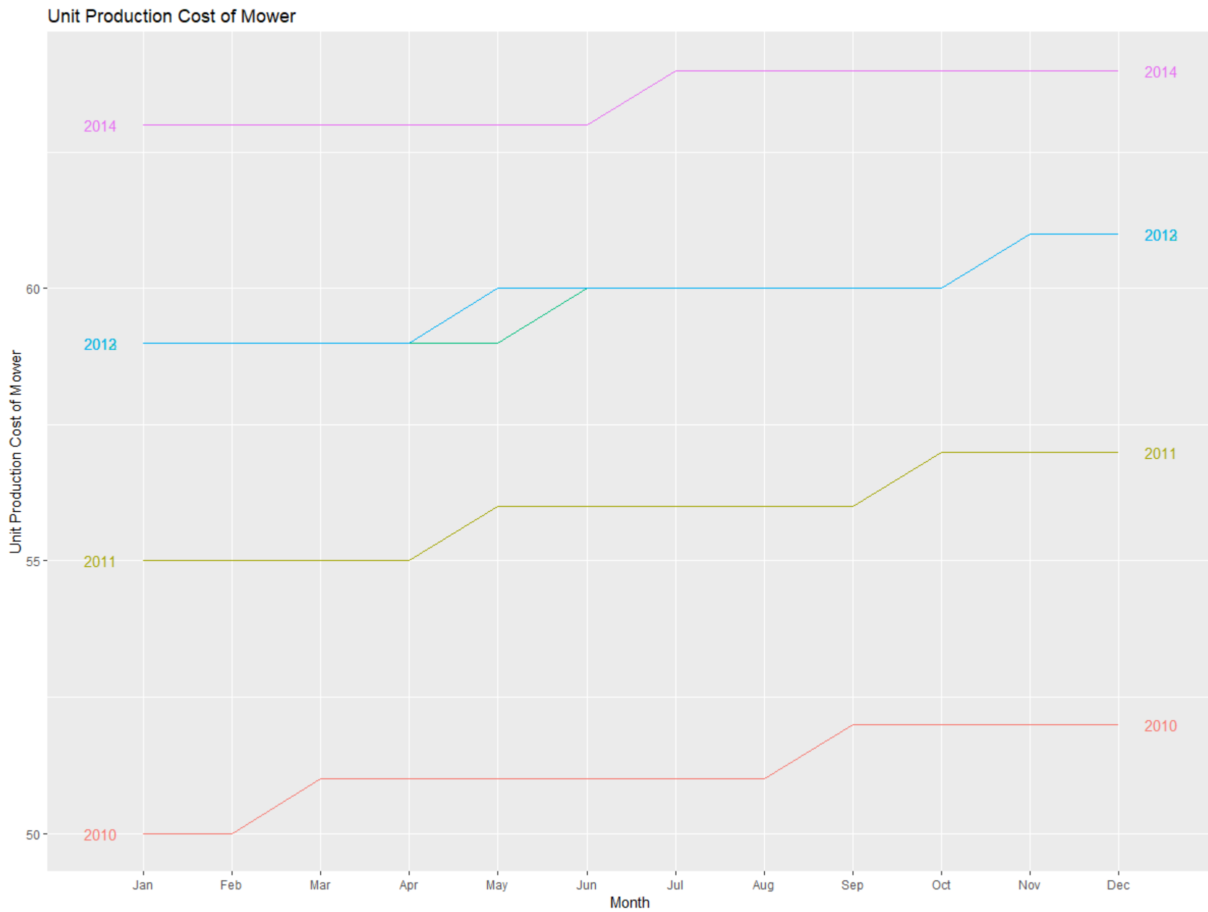
From the graph, there is an upward trend in the data over this period.
There is also a slight presence of seasonality.

To further access the seasonality and trend of the time series data, we can decompose the data as follows



Graph 222: The time series graph above shows the unit production costs of mowers from January 2010 to December 2014.

Decomposition of the time series further proves the point that there is an upward trend.

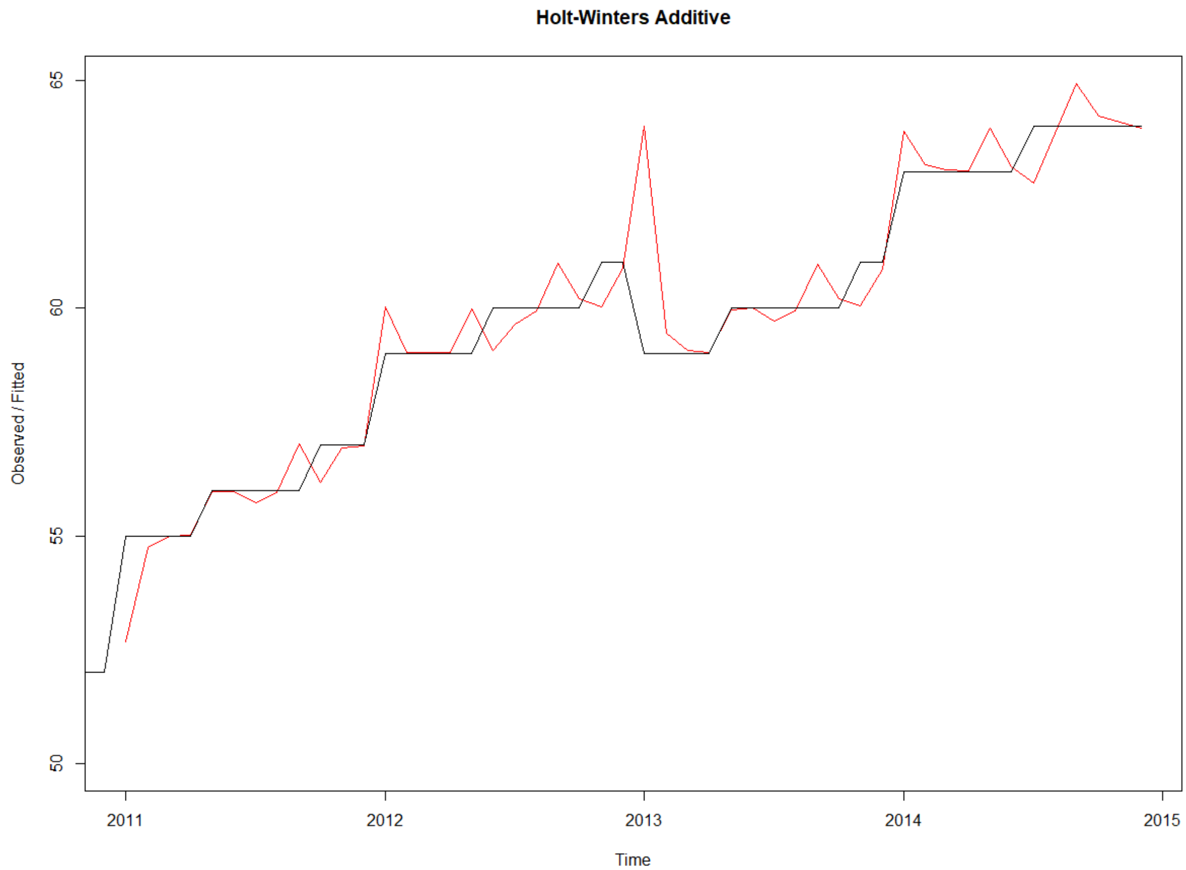


Graph 223: A seasonal plot of unit production costs of mowers from January 2010 to December 2014.

From the graph, there seems to be the presence of seasonality due to constant fluctuations at a specific time of every year. We can deduce that additive decomposition is the most appropriate as seen in the seasonal plot where the magnitude of the seasonal fluctuations does not vary with the level of the time series.

Hence, we can conclude that the unit production cost of mowers possesses <u>trend and seasonality</u>.

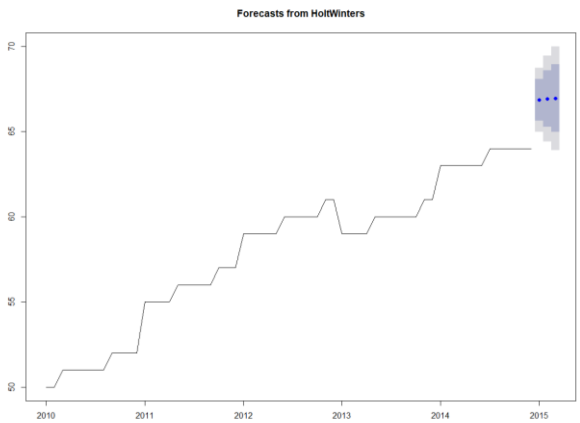
Thus, we can use <u>Holt-Winters Additive Model</u> to forecast future production costs.

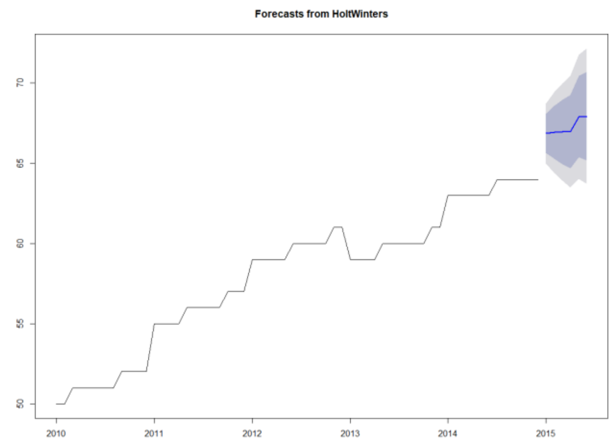


Graph 224: Observed time series data of unit production costs of mowers against the fitted Holt-Winters additive model

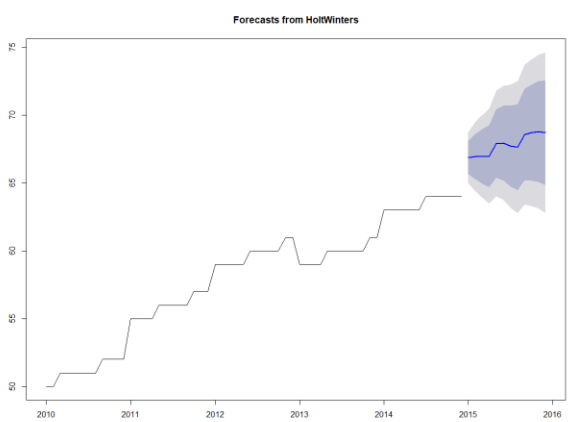
The model fits well with the observed time series data.

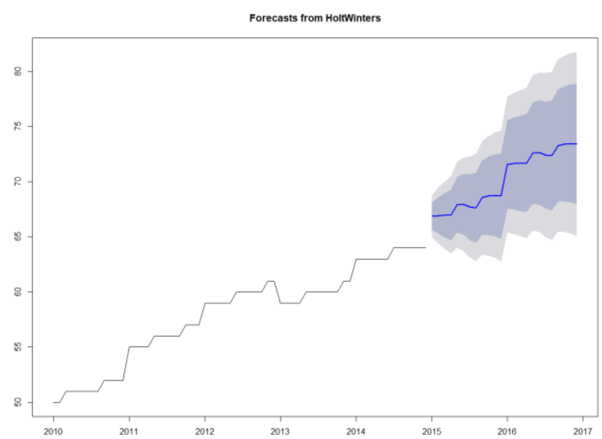
Hence, we can make use of Holt-Winters to predict data in the next 3,6,12 and 24 months.





Graphs 225 and 226: Forecasted data of unit production costs of mowers over the next 3 and 6 months respectively.





Graphs 227 and 228: Forecasted data of unit production costs of mowers over the next 12 and 24 months respectively.

The following are the forecasted values of unit production costs of mowers over the next 24 months:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	66.86335	65.63511	68.09160	64.98491	68.74180
Feb 2015	66.93196	65.27949	68.58442	64.40473	69.45919
Mar 2015	66.95547	64.96733	68.94360	63.91487	69.99606
Apr 2015	66.97355	64.69874	69.24835	63.49453	70.45256
May 2015	67.90989	65.38070	70.43908	64.04183	71.77795
Jun 2015	67.94546	65.18524	70.70569	63.72406	72.16687
Jul 2015	67.70135	64.72799	70.67472	63.15398	72.24872
Aug 2015	67.63681	64.46460	70.80903	62.78533	72.48830
Sep 2015	68.56072	65.20141	71.92004	63.42309	73.69835
Oct 2015	68.71047	65.17394	72.24700	63.30182	74.11913
Nov 2015	68.76617	65.06089	72.47145	63.09943	74.43291
Dec 2015	68.70732	64.84065	72.57399	62.79375	74.62088
Jan 2016	71.57489	67.54576	75.60403	65.41286	77.73692
Feb 2016	71.64349	67.46546	75.82153	65.25374	78.03325
Mar 2016	71.66700	67.34520	75.98881	65.05737	78.27664

Apr 2016	71.68509	67.22414	76.14604	64.86265	78.50752
May 2016	72.62143	68.02554	77.21731	65.59263	79.65022
Jun 2016	72.65700	67.93004	77.38396	65.42774	79.88627
Jul 2016	72.41289	67.55838	77.26740	64.98857	79.83721
Aug 2016	72.34835	67.36957	77.32713	64.73396	79.96274
Sep 2016	73.27226	68.17223	78.37229	65.47244	81.07209
Oct 2016	73.42201	68.20355	78.64048	65.44106	81.40296
Nov 2016	73.47771	68.14344	78.81198	65.31965	81.63577
Dec 2016	73.41886	67.97124	78.86647	65.08745	81.75026

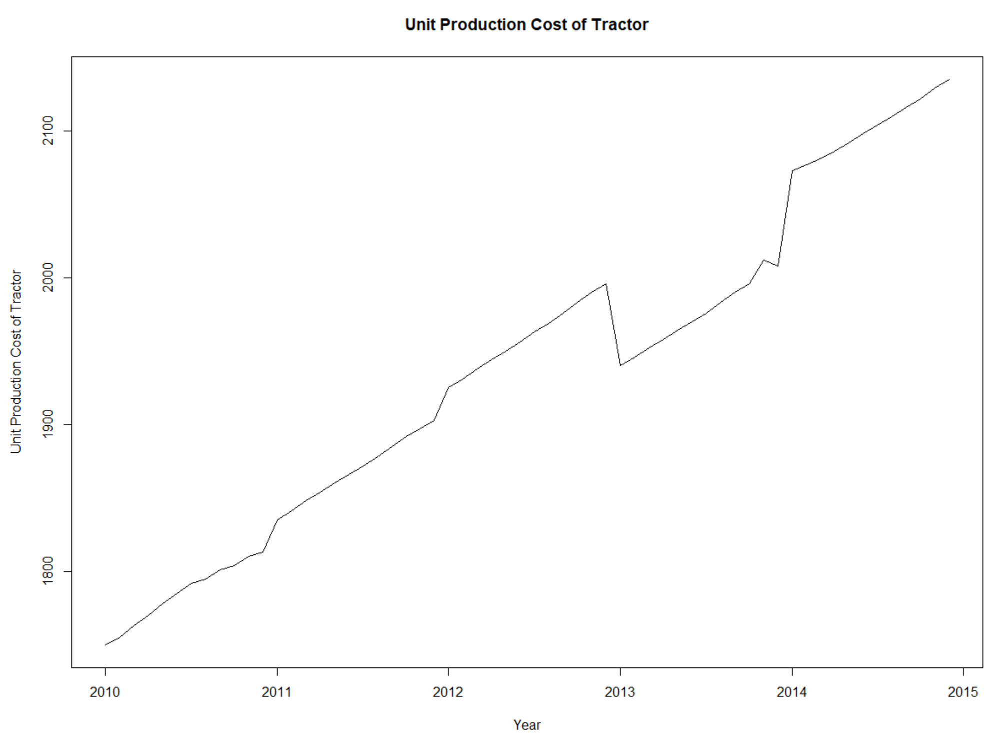
Conclusion:
As seen from the forecasted graphs and values generated, unit production costs of mowers is predicted to increase gradually in the next 24 months.

Next, we will forecast the *unit production costs of tractors*.

Forecasting of Unit Production Costs of Tractors

Time Series:

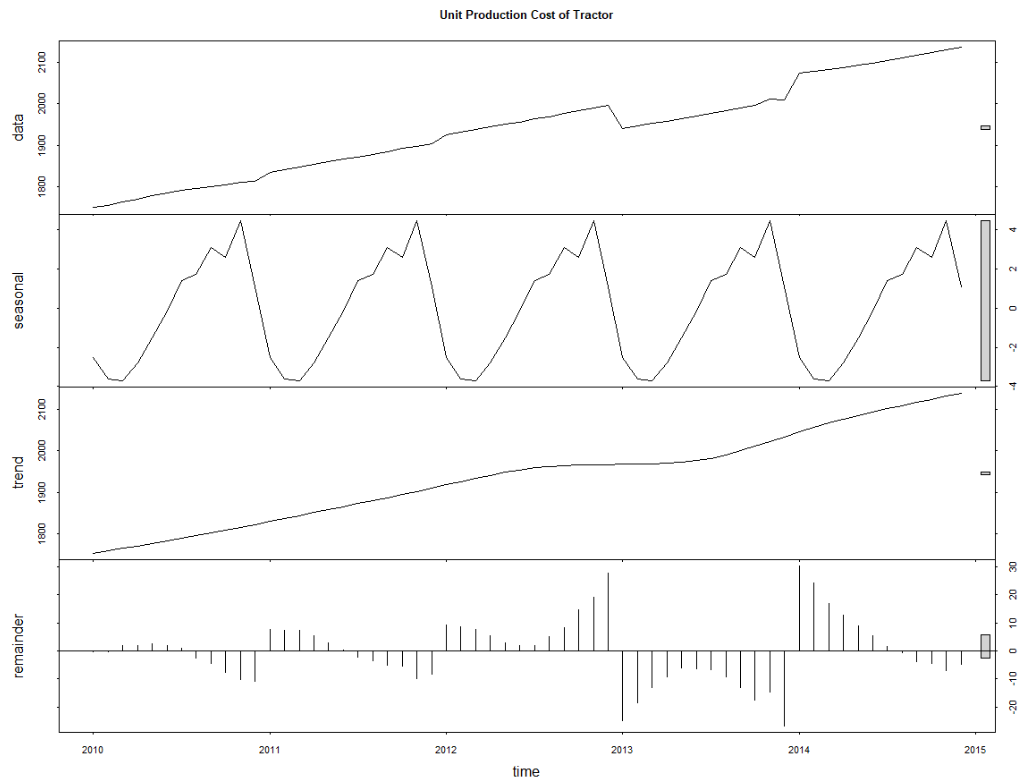
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	1750	1755	1763	1770	1778	1785	1792	1795	1801	1804	1810	1813
2011	1835	1841	1848	1854	1860	1866	1872	1878	1885	1892	1897	1903
2012	1925	1931	1938	1944	1950	1956	1963	1969	1976	1983	1990	1996
2013	1940	1946	1952	1958	1964	1970	1976	1983	1990	1996	2012	2008
2014	2073	2077	2081	2086	2092	2098	2104	2110	2116	2122	2129	2135



Graph 229: The time series graph above shows the unit production costs of tractors from January 2010 to December 2014.

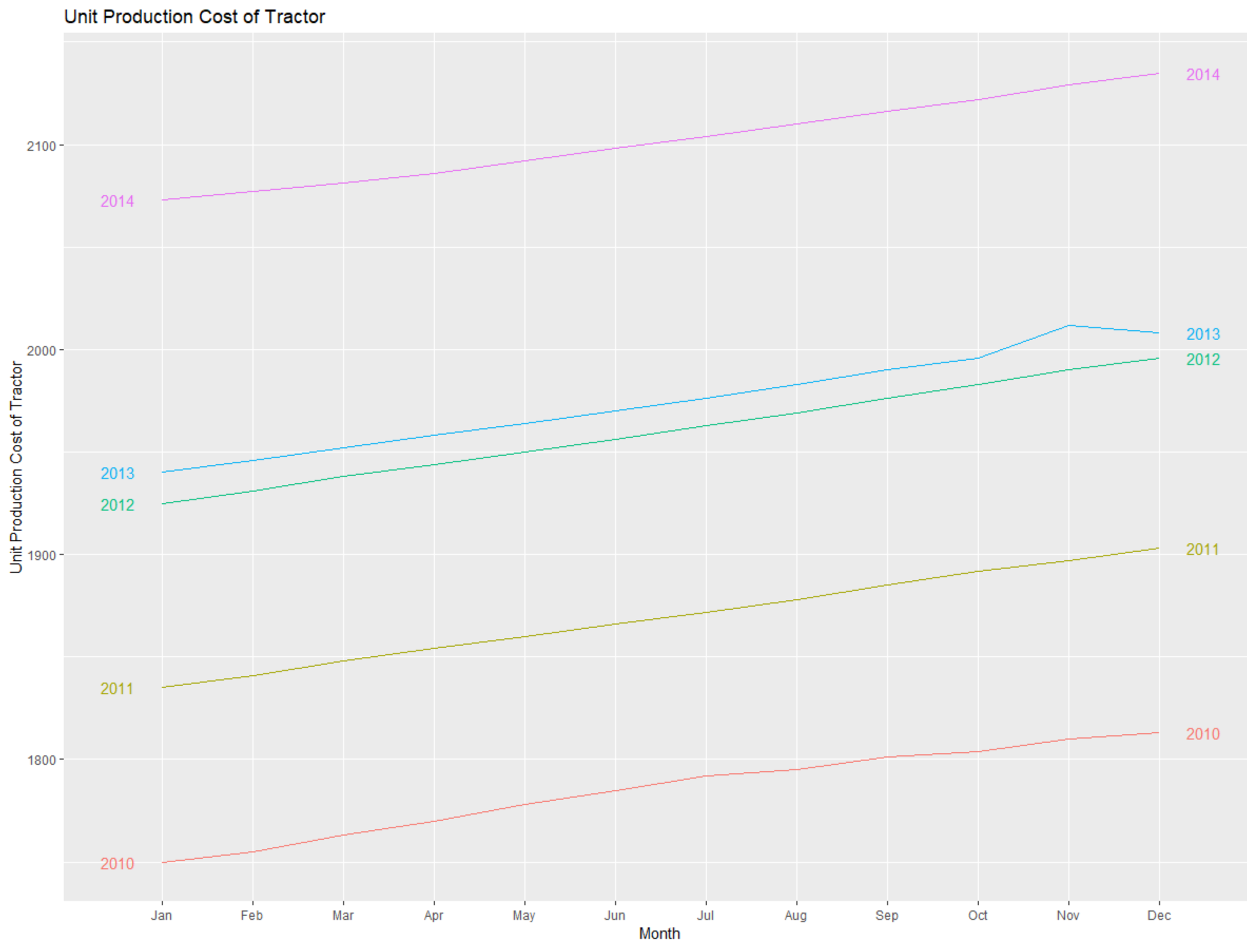
From the graph, there is an upward (and somewhat linear) trend in the data over this period. There is also a slight presence of seasonality.

To further access the seasonality and trend of the time series data, we can decompose the data as follows:



Graph 230: The time series graph above shows the unit production costs of tractors from January 2010 to December 2014

Decomposition of the time series further proves the point that there is an upward trend.

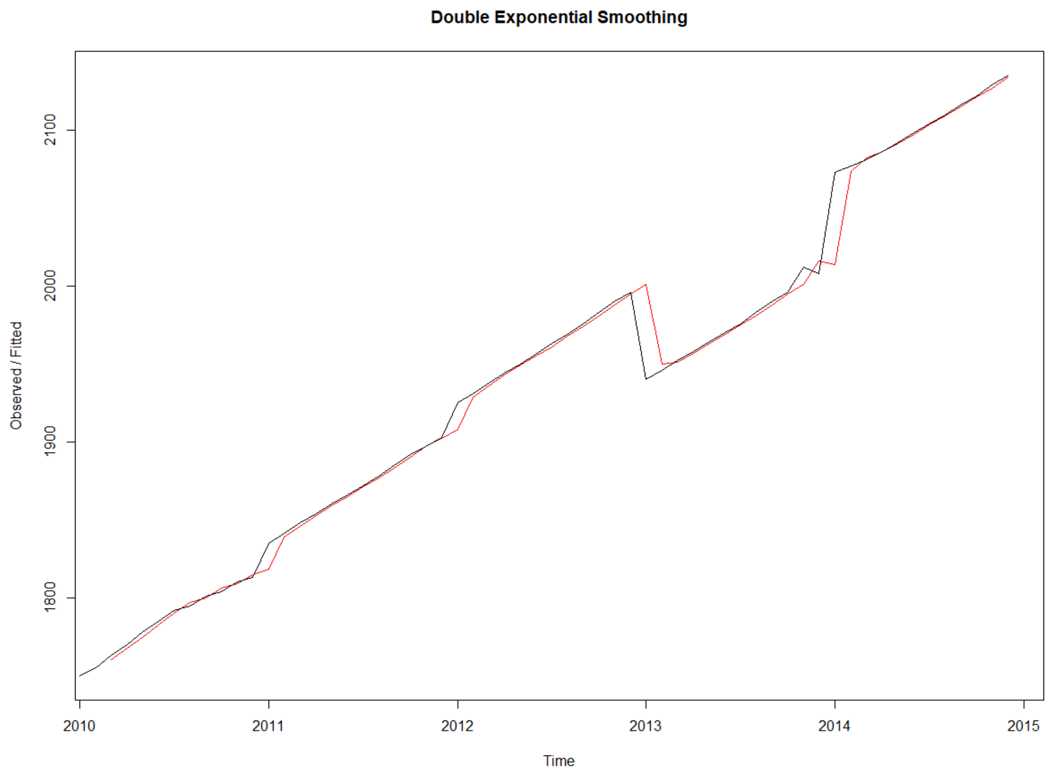


Graph 231: The time series graph above shows the unit production costs of tractors from January 2010 to December 2014.

From the seasonal plot, we can conclude that there is no seasonality in the data.

Hence, we can conclude that the unit production costs of tractors possess <u>trend but no seasonality.</u>

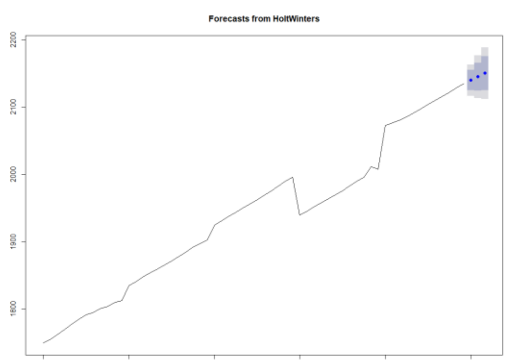
We can use <u>Double Exponential Smoothing Model</u> to forecast future unit production costs of tractors.

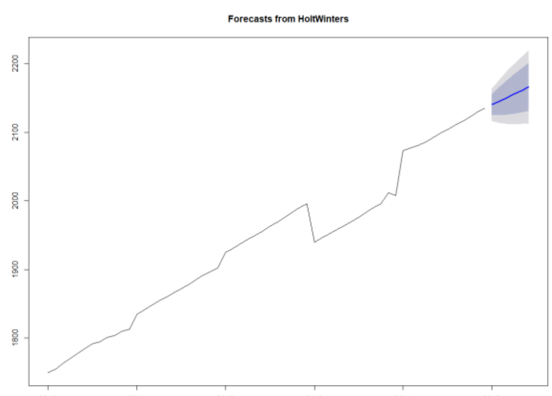


Graph 232: Observed time series data of unit production costs of tractors against the fitted Double Exponential Smoothing model.

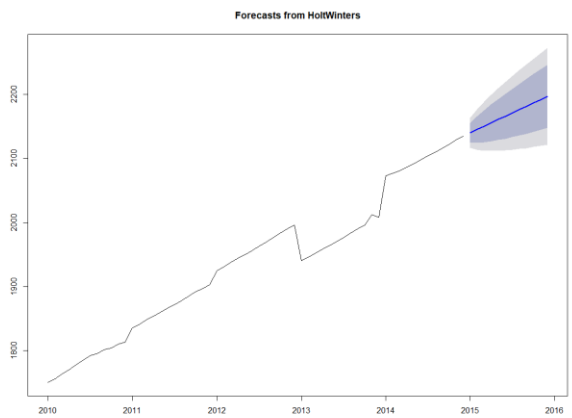
The model fits well with the observed time series data.

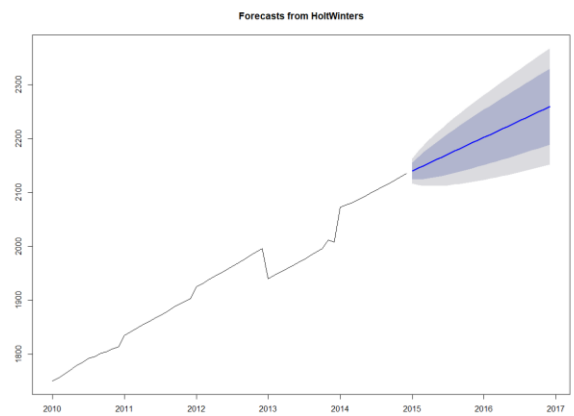
Hence, we can make use of Holt-Winters double exponential smoothing model to predict data in the next 3,6,12 and 24 months.





Graphs 233 and 234: Forecasted data of unit production costs of tractors over the next 3 and 6 months respectively.





Graphs 235 and 236: Forecasted data of unit production costs of tractors over the next 12 and 24 months respectively.

The following are the forecasted unit production costs of tractors over the next 24 months:

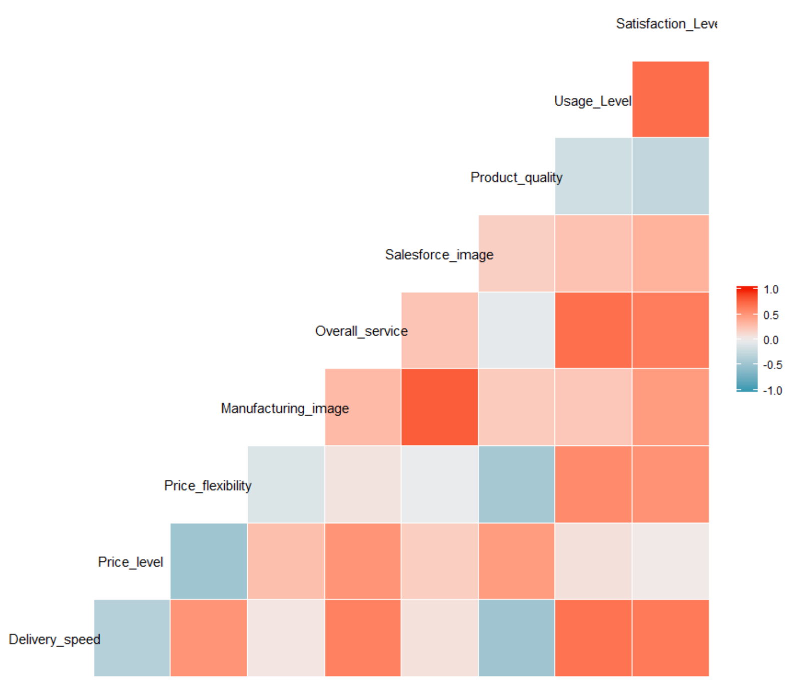
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	2140.123	2124.927	2155.320	2116.883	2163.364
Feb 2015	2145.322	2124.648	2165.997	2113.703	2176.942
Mar 2015	2150.522	2125.524	2175.520	2112.290	2188.753
Apr 2015	2155.721	2127.028	2184.414	2111.839	2199.603
May 2015	2160.920	2128.942	2192.898	2112.014	2209.826
Jun 2015	2166.119	2131.150	2201.088	2112.639	2219.599
Jul 2015	2171.318	2133.583	2209.054	2113.606	2229.030
Aug 2015	2176.517	2136.193	2216.842	2114.846	2238.189
Sep 2015	2181.717	2138.948	2224.485	2116.308	2247.125
Oct 2015	2186.916	2141.826	2232.006	2117.957	2255.875
Nov 2015	2192.115	2144.807	2239.423	2119.764	2264.466
Dec 2015	2197.314	2147.878	2246.750	2121.709	2272.919
Jan 2016	2202.513	2151.028	2253.998	2123.774	2281.252
Feb 2016	2207.712	2154.248	2261.177	2125.946	2289.479
Mar 2016	2212.912	2157.530	2268.293	2128.213	2297.610
Apr 2016	2218.111	2160.868	2275.354	2130.565	2305.657
May 2016	2223.310	2164.256	2282.364	2132.994	2313.625
Jun 2016	2228.509	2167.690	2289.328	2135.494	2321.524
Jul 2016	2233.708	2171.166	2296.250	2138.059	2329.357
Aug 2016	2238.907	2174.682	2303.133	2140.683	2337.132
Sep 2016	2244.106	2178.232	2309.980	2143.361	2344.852
Oct 2016	2249.306	2181.816	2316.795	2146.090	2352.521
Nov 2016	2254.505	2185.431	2323.578	2148.866	2360.144
Dec 2016	2259.704	2189.075	2330.333	2151.686	2367.722

Conclusion:
As seen from the forecasted graphs and values generated, unit production costs of tractors are predicted to increase steadily and almost linearly for the next 24 months to come.

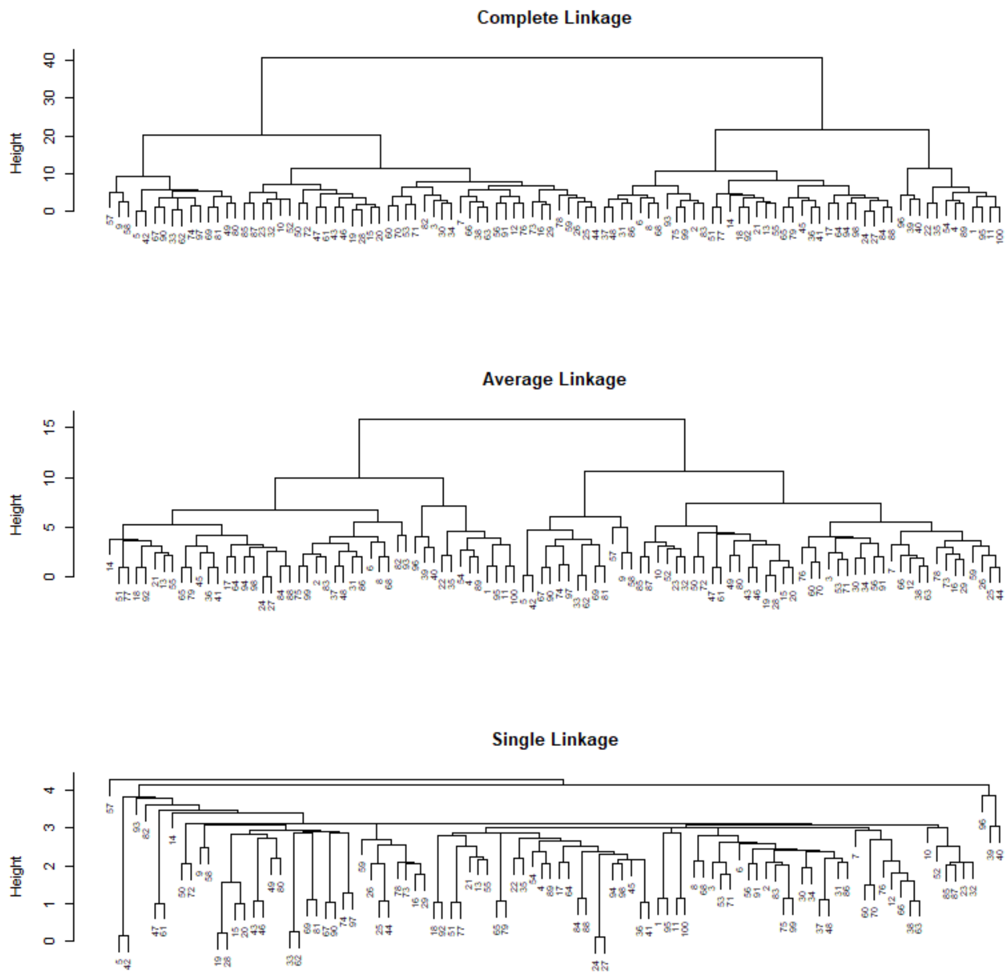
3) Generating customer insights using data-mining techniques
Analyze the respective data provided by apply appropriate data-mining techniques, to generate customer insights, such as segmenting PLE customers with similar perceptions about the company into groups, or insights about the drivers of satisfaction and usage level, etc. Summarize your results in a report.

For this question, we are given the purchasing survey results of 100 PLE customers (each customer being a purchasing manager representing a firm), where each customer rates each of the seven PLE attributes, namely Delivery Speed, Price Level, Price Flexibility, Manufacturing Image, Overall Service, Sales Force Image and Product Quality; using a scale from 0.0 to 10.0, rounded to one decimal place. Other data from the survey included the rating of the customer's Usage Level of PLE products (ranging from 0% to 100%), how satisfied the customer is with past purchases from PLE (ranging from 1 to 7), and information about the firm the customer is making purchases for, namely the size of firm, purchasing structure, industry and buying type.

To generate customer insights, we will first segment the PLE customers with similar perceptions about the company into groups (clustering). We would group the PLE customers into different clusters by using the customers' responses for the 9 numerical variables in the data, namely Delivery Speed, Price Level, Price Flexibility, Manufacturing Image, Overall Service, Sales Force Image, Product Quality, Usage Level and Satisfaction Level. Below is the correlation matrix of these 9 variables in the data set, to better understand the relationships between these variables.



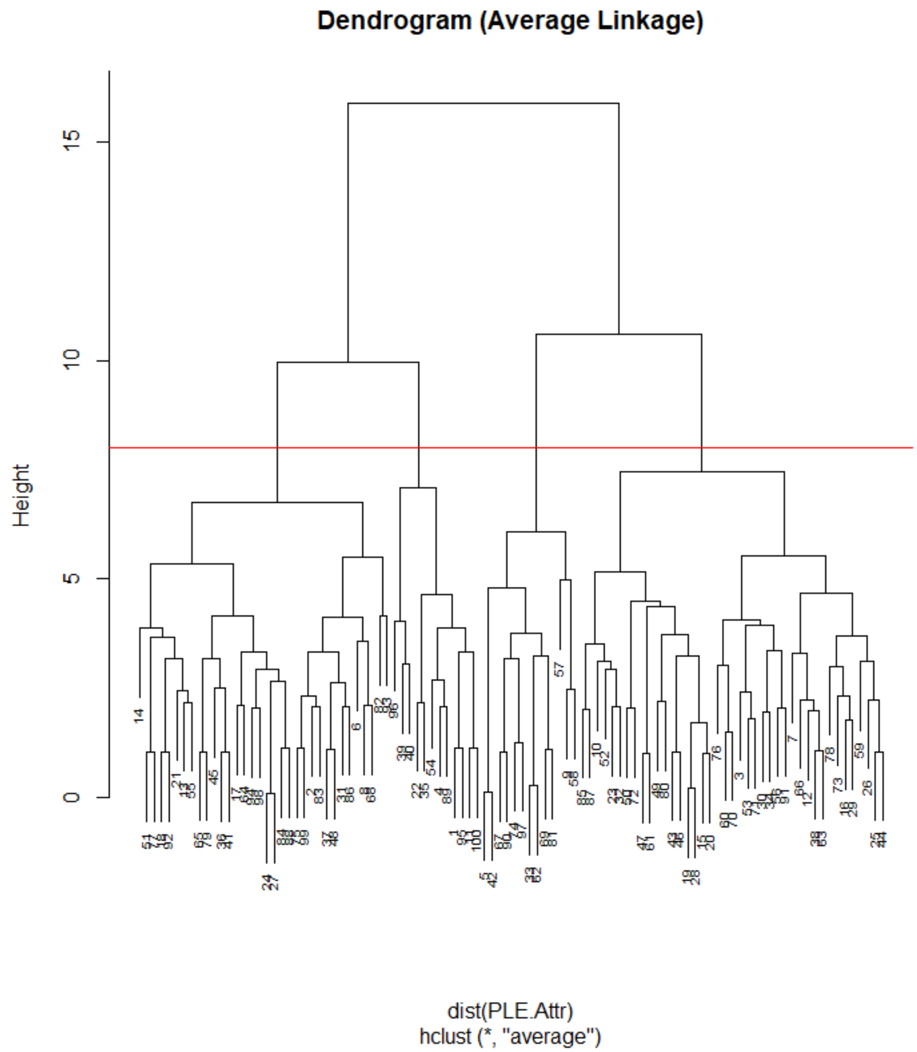
Using hierarchical clustering of the data, the following dendrograms can be plotted, differentiated by the different methods of clustering, namely Complete Linkage, Average Linkage and Single Linkage.



Graph 237: Collation of dendrograms formed by different methods of hierarchical clustering, namely Single Linkage, Average Linkage and Complete Linkage in one window.

Generally, we would use Average Linkage method to perform Hierarchical Clustering, as this method typically has fewer disadvantages as compared to the other methods.

Using the Average Linkage method, we can cut the dendrogram such that we would obtain 4 clusters of customers. In this case, a horizontal line at y=8 will cut the dendrogram into 4 clusters as there are 4 intersections between the dendrogram and the horizontal red line (y=8) as seen below. The dendrogram is divided into 4 clusters here because as shown in the following dendrogram, it seems that 4 groups can separate our data into clusters relatively nicely.



Graph 238: Dendrogram formed using average linkage clustering cut into 4 clusters by a horizontal line at y = 8.

After cutting the dendrogram to form 4 clusters, each of the 100 PLE customers were classified under one of the 4 clusters. The results are as shown below.

```
 [1] 1 2 3 1 4 2 3 2 4 3 1 3 2 2 3 3 2 2 3 3 2 1 3 2 3 3 2 3 3 3 2 3 4 3 1 2 2 3 1 1 2 4 3 3 2 3 3 2 3 3 2 3 3 2 3 3 1 2 3 4 4 3 3 3 4 3 2
[65] 2 3 4 2 4 3 3 3 3 4 2 3 2 3 2 3 4 2 2 2 3 2 3 2 1 4 3 2 2 2 1 1 4 2 2 1
```

The following tables are constructed to show the cluster each PLE customer is classified under.

hc.clusters	1	10	100	11	12	13	14	15	16	17	18	19	2	20	21	22	23	24	25	26	27	28	29	3	30	31	32	33	34	35	36	37	38	39	4	40	41	42	43	44	45
1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	0	0
2	0	0	0	0	0	1	1	0	0	1	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1
3	0	1	0	0	1	0	0	1	1	0	0	1	0	1	0	0	1	0	1	1	0	1	1	1	1	0	1	0	1	0	0	0	1	0	0	0	0	0	1	1	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0

hc.clusters	46	47	48	49	5	50	51	52	53	54	55	56	57	58	59	6	60	61	62	63	64	65	66	67	68	69	7	70	71	72	73	74	75	76	77	78	79	8	80	81	82
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	1
3	1	1	0	1	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0	1	0	0	1	0	0	0	1	1	1	1	1	0	0	1	0	1	0	0	1	0	0
4	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0

```
hc.clusters 83 84 85 86 87 88 89 9 90 91 92 93 94 95 96 97 98 99
          1  0  0  0  0  0  0  1 0  0  0  0  0  0  1  1  0  0  0
          2  1  1  0  1  0  1  0 0  0  0  1  1  1  0  0  0  1  1
          3  0  0  1  0  1  0  0 0  0  1  0  0  0  0  0  0  0  0
          4  0  0  0  0  0  0  0 1  1  0  0  0  0  0  0  1  0  0
```

To see which customers (with which attributes) are clustered into the same clusters, we appended a column to the original data to state the group each customer has been classified into. Below shows a screenshot of the modified data.

	Delivery speed	Price level	Price flexibility	Manufacturing image	Overall service	Salesforce image	Product quality	Usage Level	Satisfaction Level	Size of firm	Purchasing Structure	Industry	Buying Type	Classification
1	4.1	0.6	6.9	4.7	2.4	2.3	5.2	32	4.2	0	0	1	1	1
2	1.8	3.0	6.3	6.6	2.5	4.0	8.4	43	4.3	1	1	0	1	2
3	3.4	5.2	5.7	6.0	4.3	2.7	8.2	48	5.2	1	1	1	2	3
4	2.7	1.0	7.1	5.9	1.8	2.3	7.8	32	3.9	1	1	1	1	1
5	6.0	0.9	9.6	7.8	3.4	4.6	4.5	58	6.8	0	0	1	3	4
6	1.9	3.3	7.9	4.8	2.6	1.9	9.7	45	4.4	1	1	1	2	2
7	4.6	2.4	9.5	6.6	3.5	4.5	7.6	46	5.8	0	0	1	1	3
8	1.3	4.2	6.2	5.1	2.8	2.2	6.9	44	4.3	1	1	0	2	2
9	5.5	1.6	9.4	4.7	3.5	3.0	7.6	63	5.4	0	0	1	3	4
10	4.0	3.5	6.5	6.0	3.7	3.2	8.7	54	5.4	1	1	0	2	3
11	2.4	1.6	8.8	4.8	2.0	2.8	5.8	32	4.3	0	0	0	1	1
12	3.9	2.2	9.1	4.6	3.0	2.5	8.3	47	5.0	0	0	1	2	3
13	2.8	1.4	8.1	3.8	2.1	1.4	6.6	39	4.4	1	1	0	1	2
14	3.7	1.5	8.6	5.7	2.7	3.7	6.7	38	5.0	0	0	1	1	2
15	4.7	1.3	9.9	6.7	3.0	2.6	6.8	54	5.9	0	0	0	3	3
16	3.4	2.0	9.7	4.7	2.7	1.7	4.8	49	4.7	0	0	0	3	3
17	3.2	4.1	5.7	5.1	3.6	2.9	6.2	38	4.4	0	1	1	2	2
18	4.9	1.8	7.7	4.3	3.4	1.5	5.9	40	5.6	0	0	0	2	2
19	5.3	1.4	9.7	6.1	3.3	3.9	6.8	54	5.9	0	0	1	3	3
20	4.7	1.3	9.9	6.7	3.0	2.6	6.8	55	6.0	0	0	0	3	3
21	3.3	0.9	8.6	4.0	2.1	1.8	6.3	41	4.5	0	0	0	2	2
22	3.4	0.4	8.3	2.5	1.2	1.7	5.2	35	3.3	0	0	0	1	1
23	3.0	4.0	9.1	7.1	3.5	3.4	8.4	55	5.2	0	1	0	3	3
24	2.4	1.5	6.7	4.8	1.9	2.5	7.2	36	3.7	1	1	0	1	2
25	5.1	1.4	8.7	4.8	3.3	2.6	3.8	49	4.9	0	0	0	2	3
26	4.6	2.1	7.9	5.8	3.4	2.8	4.7	49	5.9	0	0	1	3	3

Next, we will then be able to generate customer insights by looking at what are some of the attributes that the customers grouped under the same cluster have in common.

Firstly, for group one, below is a screenshot of all the rows of customers in the data set who have been classified under Cluster 1.

	Delivery speed	Price level	Price flexibility	Manufacturing image	Overall service	Salesforce image	Product quality	Usage Level	Satisfaction Level	Size of firm	Purchasing Structure	Industry	Buying Type	Classification
1	4.1	0.6	6.9	4.7	2.4	2.3	5.2	32	4.2	0	0	1	1	1
2	2.7	1.0	7.1	5.9	1.8	2.3	7.8	32	3.9	1	1	1	1	1
3	2.4	1.6	8.8	4.8	2.0	2.8	5.8	32	4.3	0	0	0	1	1
4	3.4	0.4	8.3	2.5	1.2	1.7	5.2	35	3.3	0	0	0	1	1
5	2.4	1.0	7.7	3.4	1.7	1.1	6.2	35	4.1	1	1	0	1	1
6	0.0	2.1	6.9	5.4	1.1	2.6	8.9	29	3.9	1	1	1	1	1
7	2.4	2.0	6.4	4.5	2.1	2.2	8.8	28	3.3	1	1	1	1	1
8	2.8	2.4	6.7	4.9	2.5	2.6	9.2	32	3.7	1	1	1	1	1
9	2.9	1.2	7.3	6.1	2.0	2.5	8.0	34	4.0	1	1	1	1	1
10	4.0	0.5	6.7	4.5	2.2	2.1	5.0	31	4.0	0	0	1	1	1
11	0.6	1.6	6.4	5.0	0.7	2.1	8.4	25	3.4	1	1	1	1	1
12	2.5	1.8	9.0	5.0	2.2	3.0	6.0	33	4.4	0	0	0	1	1

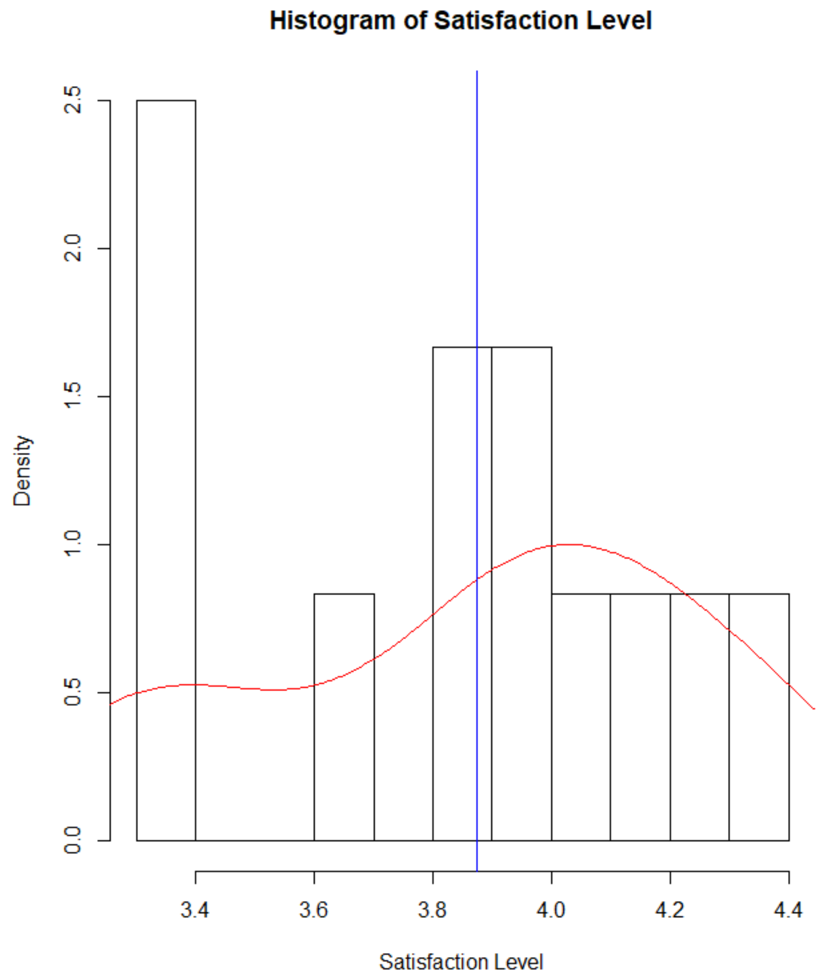
Below are the descriptive statistics of the customers in Cluster 1:

```
                     vars  n  mean   sd median trimmed  mad  min  max range  skew kurtosis   se
Delivery speed          1 12  2.52 1.20   2.60    2.61 0.37  0.0  4.1   4.1 -0.73    -0.36 0.35
Price level             2 12  1.35 0.67   1.40    1.34 0.74  0.4  2.4   2.0  0.00    -1.54 0.19
Price flexibility       3 12  7.35 0.90   7.00    7.28 0.67  6.4  9.0   2.6  0.68    -1.15 0.26
Manufacturing image     4 12  4.72 0.99   4.85    4.81 0.52  2.5  6.1   3.6 -0.76    -0.08 0.29
Overall service         5 12  1.82 0.56   2.00    1.87 0.37  0.7  2.5   1.8 -0.69    -0.91 0.16
Salesforce image        6 12  2.27 0.51   2.30    2.32 0.37  1.1  3.0   1.9 -0.77    -0.01 0.15
Product quality         7 12  7.04 1.62   7.00    7.03 2.37  5.0  9.2   4.2  0.02    -1.88 0.47
Usage Level             8 12 31.50 2.94  32.00   31.80 2.22 25.0 35.0  10.0 -0.75    -0.41 0.85
Satisfaction Level      9 12  3.88 0.38   3.95    3.88 0.37  3.3  4.4   1.1 -0.34    -1.38 0.11
Size of firm           10 12  0.58 0.51   1.00    0.60 0.00  0.0  1.0   1.0 -0.30    -2.06 0.15
Purchasing Structure   11 12  0.58 0.51   1.00    0.60 0.00  0.0  1.0   1.0 -0.30    -2.06 0.15
Industry               12 12  0.67 0.49   1.00    0.70 0.00  0.0  1.0   1.0 -0.62    -1.74 0.14
Buying Type            13 12  1.00 0.00   1.00    1.00 0.00  1.0  1.0   0.0   NaN      NaN 0.00
Classification         14 12  1.00 0.00   1.00    1.00 0.00  1.0  1.0   0.0   NaN      NaN 0.00
```

Analysis of each of the variables:

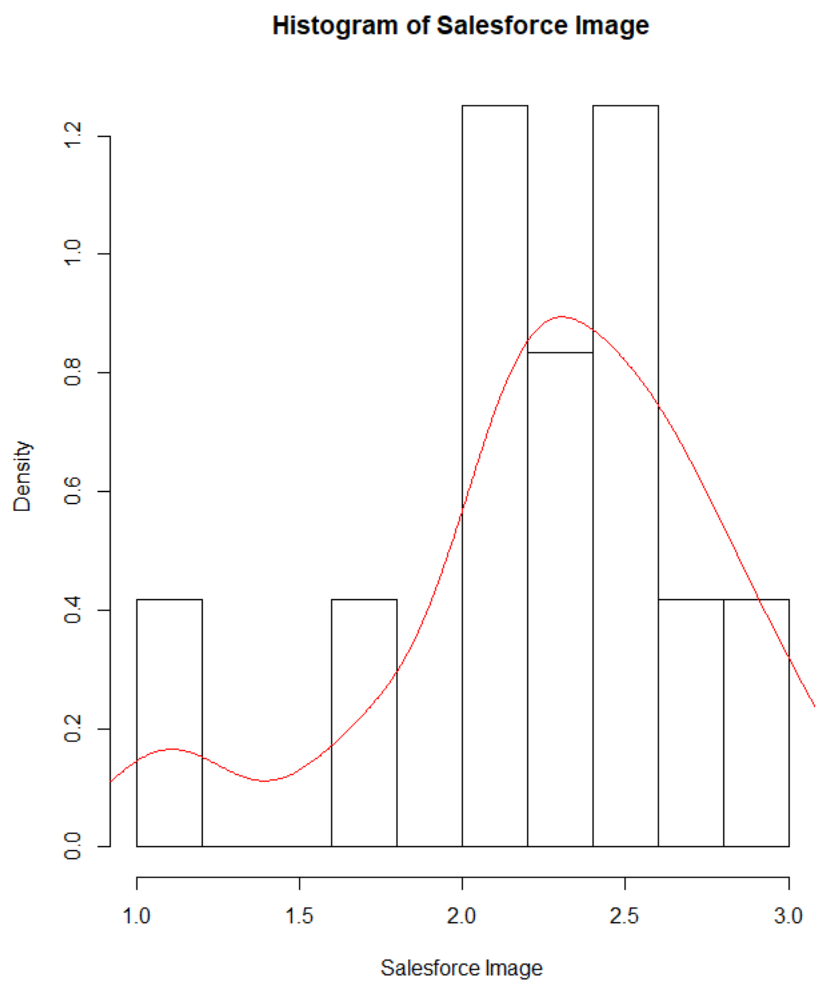
- **Buying Type:** As seen from the standard deviation of the variables in Cluster 1 in the above table, all customers belong to Buying Type group 1, which represents 'New Purchase' (i.e. standard deviation of Buying Type = 0.00). Thus, if a customer is classified under Cluster 1, he/she is likely to be classified under Buying Type group 1 as well.

- **Satisfaction Level:** Satisfaction Level in cluster 1 is the variable with the second smallest standard deviation among all the other variables (standard deviation of Satisfaction Level as seen from the above table is 0.38). Satisfaction level varies from 3.3 to 4.4 in cluster 1 and having a small standard deviation for the variable implies that the data points form a narrow spread about the mean value of 3.88. Most customers grouped under cluster 1 thus have satisfaction levels close to 3.88. Magnitude of the coefficient of skewness is less than 0.5, which means that the variable displays relative symmetry in its distribution. Coefficient of kurtosis is less than 3 (-1.38), showing that the variable's distribution is flat with a wide degree of dispersion. To better understand the distribution of the Satisfaction Level in cluster 1 customers, we plotted the following density histogram and density plot:



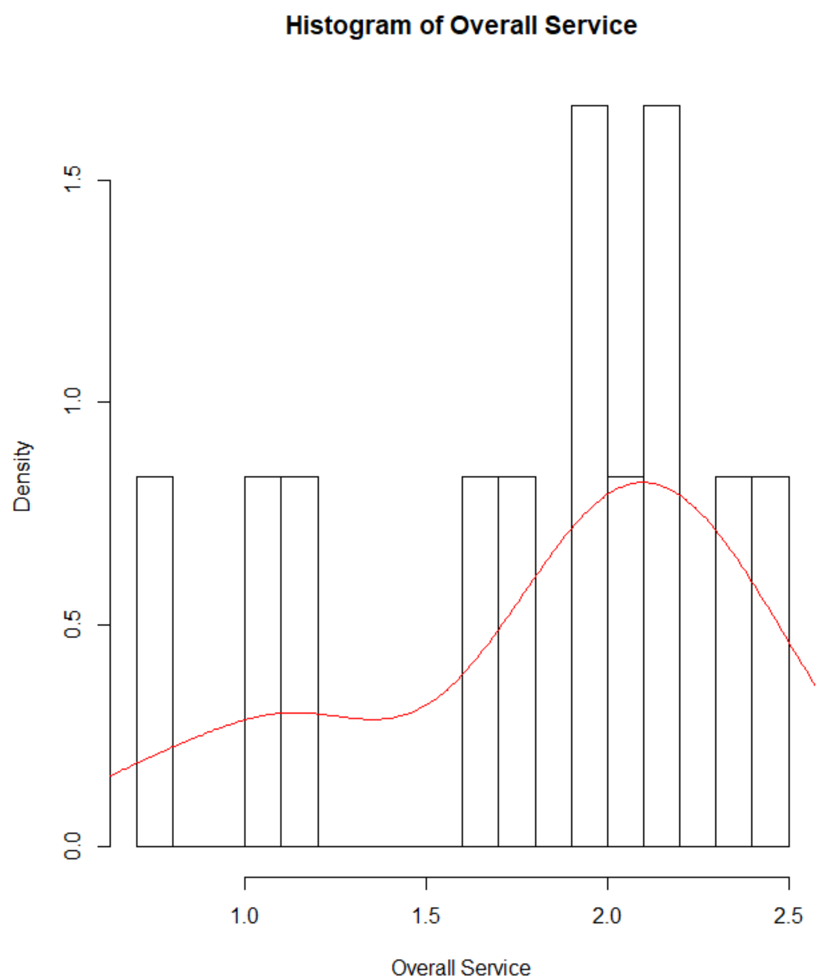
Graph 239: Above shows a histogram of the Satisfaction Level among cluster 1 customers, with the red line showing the density plot of the distribution and the blue line showing the mean Satisfaction Level of the distribution.

- **Salesforce Image:** With a standard deviation of 0.51, the distribution has a moderately narrow spread about its mean. Salesforce image varies from 1.1 to 3.0 and has a mean value of 2.27. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 1 tend to have rated Salesforce image low (ratings range from 1.1 to 3.0 out of 10.0, with an average value of 2.27). Below is the histogram and density plot of Salesforce Image in cluster 1.



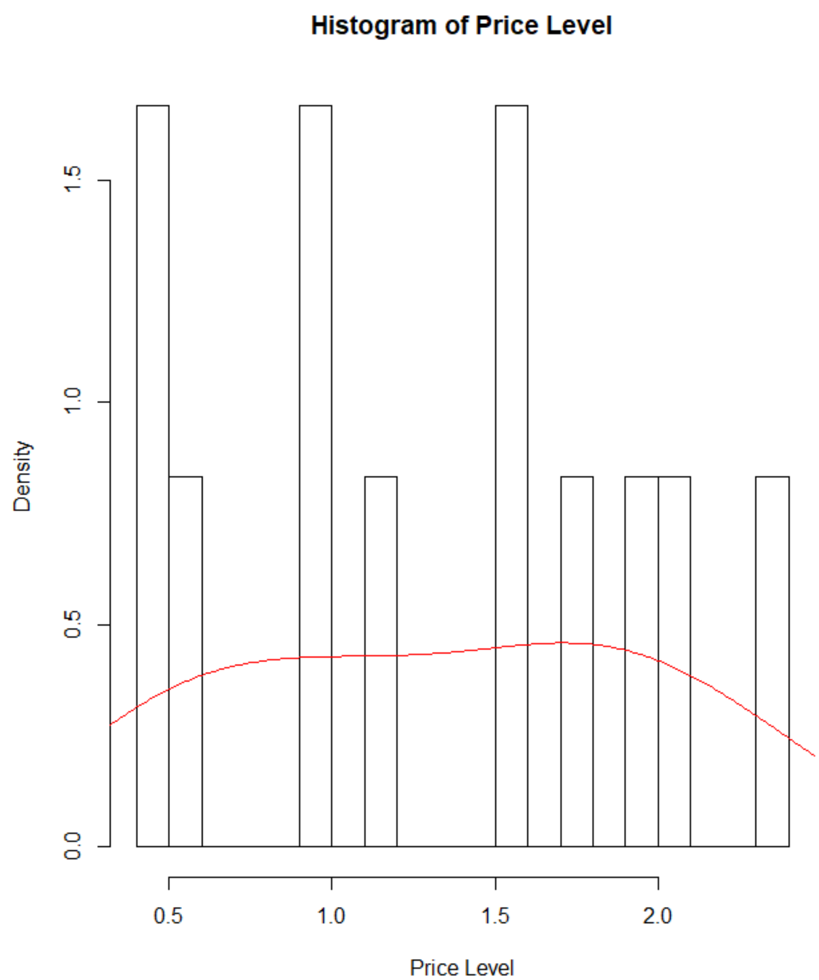
Graph 240: Histogram of Salesforce Image ratings in Cluster 1, with density plot shown as the red line.

- **Overall Service:** With a standard deviation of 0.56, the distribution has a moderately narrow spread about its mean. Overall Service varies from 0.7 to 2.5 and has a mean value of 1.82. Rating was done on a 10cm scale, rounded to 1 decimal place. Similar to Salesforce Image, customers in cluster 1 also tend to have rated Overall Service low (ratings range from 0.7 to 2.5 out of 10.0, with a mean value of 1.82). Below is the histogram and density plot of Overall Service in cluster 1.



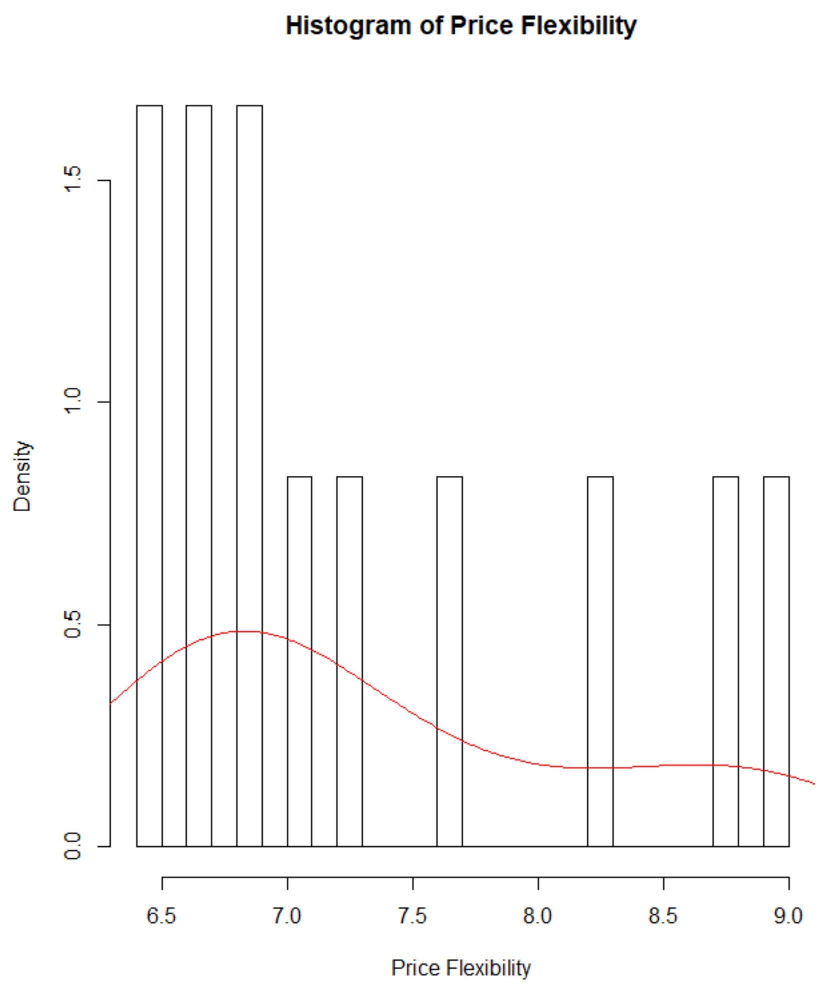
Graph 241: Histogram of Overall Service ratings in Cluster 1, with density plot shown as the red line.

- **Price Level:** With a standard deviation of 0.67, the distribution has a moderately narrow spread about its mean. Price Level varies from 0.4 to 2.4 and has a mean value of 1.35. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 1 also tend to have rated Price Level low (ratings range from 0.4 to 2.4 out of 10.0, with a mean value of 1.35). Price level depicts the perceived level of price charged by PLE. Coefficient of skewness is 0, which means that the distribution is symmetric. Below is the histogram and density plot of Price Level in cluster 1.



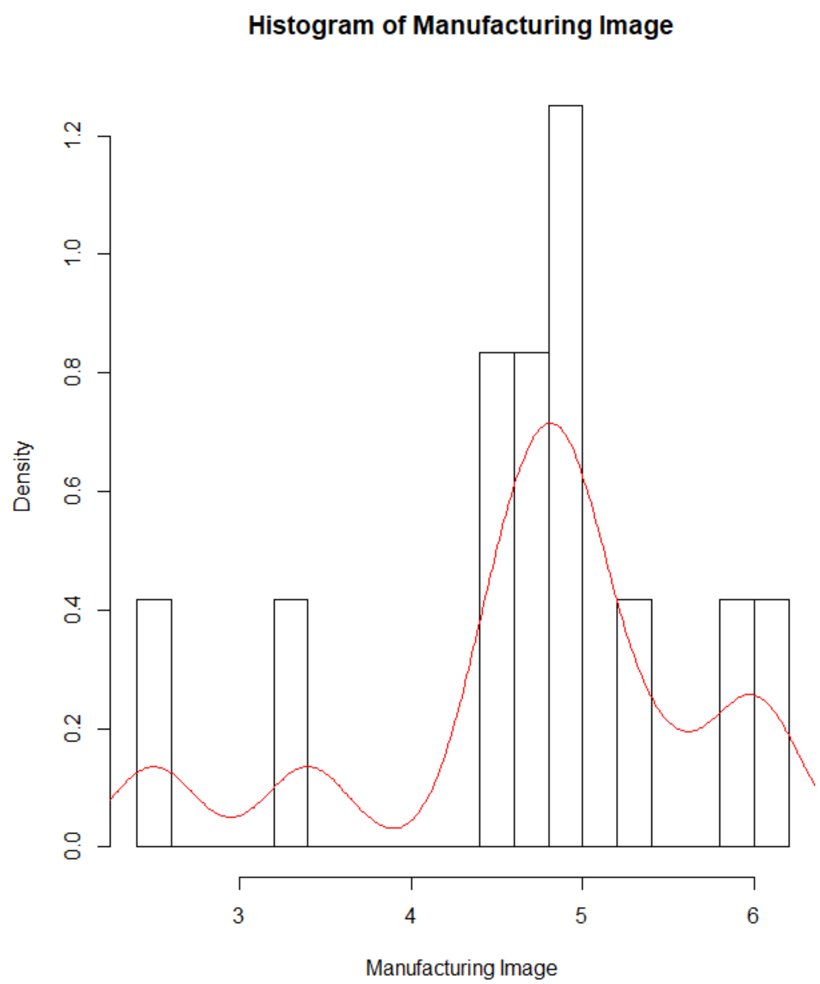
Graph 242: Histogram of Price Level ratings in Cluster 1, with density plot shown as the red line.

- **Price Flexibility:** With a standard deviation of 0.90, Price Flexibility varies from 6.4 to 9.0 and has a mean value of 7.35. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 1 tend to have rated Price Flexibility moderately high (ratings ranging from 6.4 to 9.0 out of 10.0, with a mean value of 7.35). Price flexibility depicts the perceived willingness of PLE representatives to negotiate prices on all types of purchases. Below is the histogram and density plot of Price Flexibility in cluster 1.



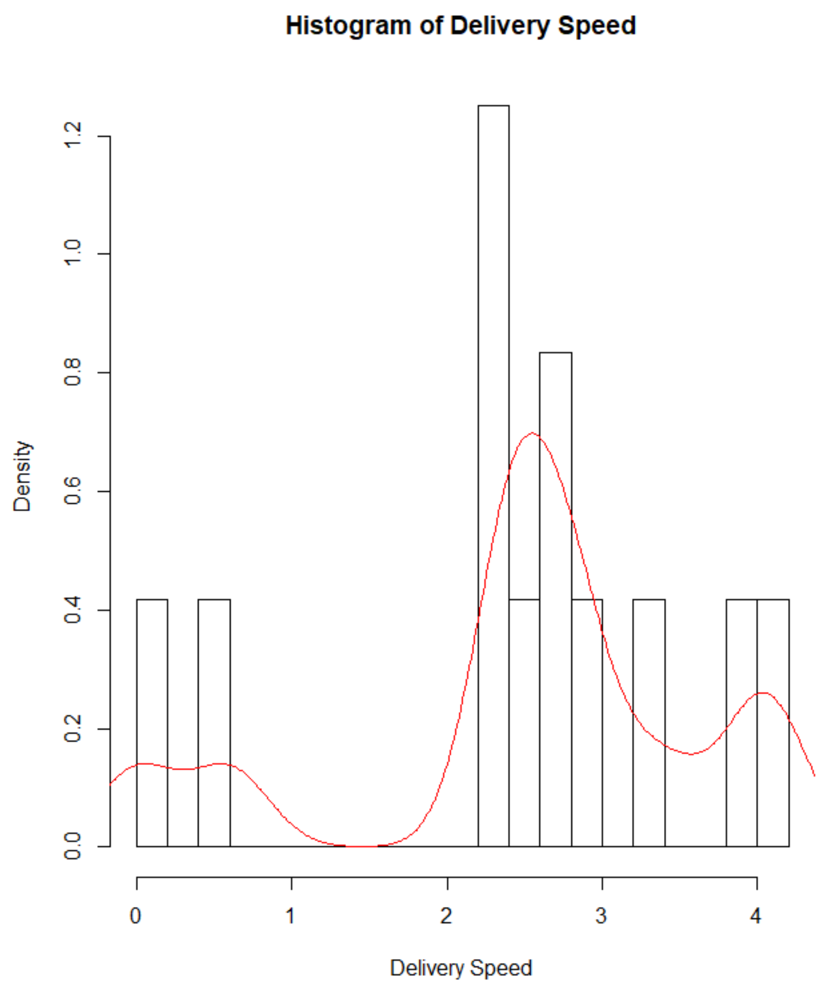
Graph 243: Histogram of Price Flexibility ratings in Cluster 1, with density plot shown as the red line.

- **Manufacturing Image:** With a standard deviation of 0.99, Manufacturing Image varies from 2.5 to 6.1 and has a mean value of 4.72. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 1 tend to have rated Manufacturing Image moderately low (ratings ranging from 2.5 to 6.1 out of 10.0, with a mean value of 4.72). Manufacturing Image depicts the overall image of the manufacturer. Below is the histogram and density plot of Manufacturing Image in cluster 1.



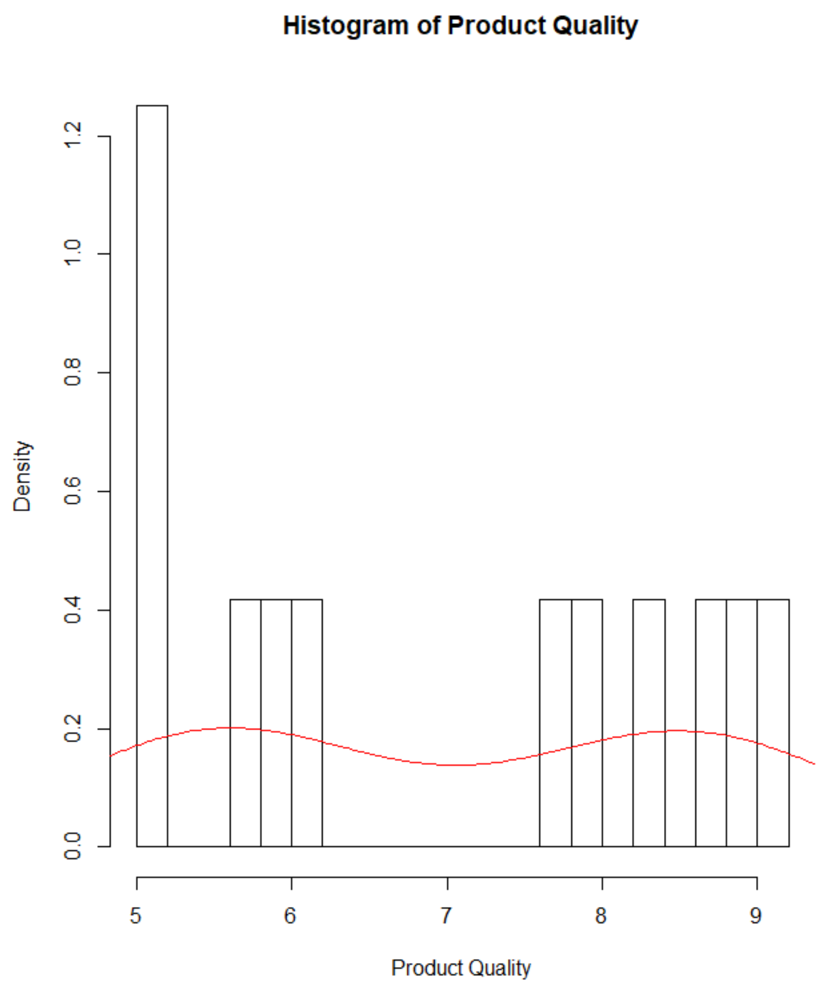
Graph 244: Histogram of Manufacturing Image ratings in Cluster 1, with density plot shown as the red line.

- **Delivery Speed:** With a standard deviation of 1.20, Delivery Speed varies from 0.0 to 4.1 and has a mean value of 2.52. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 1 tend to have rated Delivery Speed low (ratings ranging from 0.0 to 4.1 out of 10.0, with a mean value of 2.52). Delivery Speed depicts the amount of time it takes to deliver product once an order is confirmed. Below is the histogram and density plot of Delivery Speed in cluster 1.



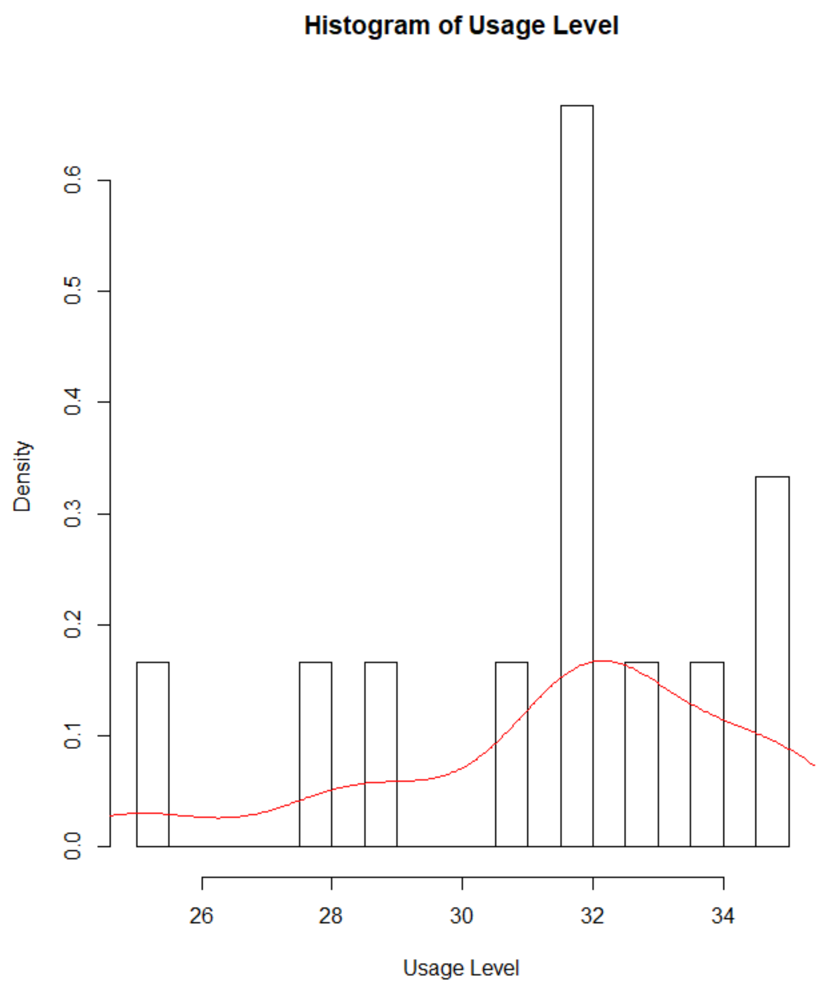
Graph 245: Histogram of Delivery Speed ratings in Cluster 1, with density plot shown as the red line.

- **Product Quality:** With a standard deviation of 1.62, Product Quality varies from 5.0 to 9.2 and has a mean value of 7.04. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 1 tend to have rated Product Quality moderately high (ratings ranging from 5.0 to 9.2 out of 10.0, with a mean value of 7.04). Product Quality depicts the perceived level of quality of PLE products. Below is the histogram and density plot of Product Quality in cluster 1.



Graph 246: Histogram of Product Quality ratings in Cluster 1, with density plot shown as the red line.

- **Usage Level:** With a standard deviation of 2.94, Usage Level varies from 25.0 to 35.0 and has a mean value of 31.50. Usage Level depicts how much of the firm's total product is purchased from PLE, measured on a 100-point scale, ranging from 0% to 100%. Customers in cluster 1 tend to rate the usage level moderately low as well (ratings ranging from 25.0% to 35.0% out of 100.0%, with a mean value of 31.5%). Below is the histogram and density plot of Usage Level in cluster 1.



Graph 247: Histogram of Usage Level ratings in Cluster 1, with density plot shown as the red line.

Secondly, for group two, below is a screenshot of some of the rows of customers in the data set who have been classified under Cluster 2.

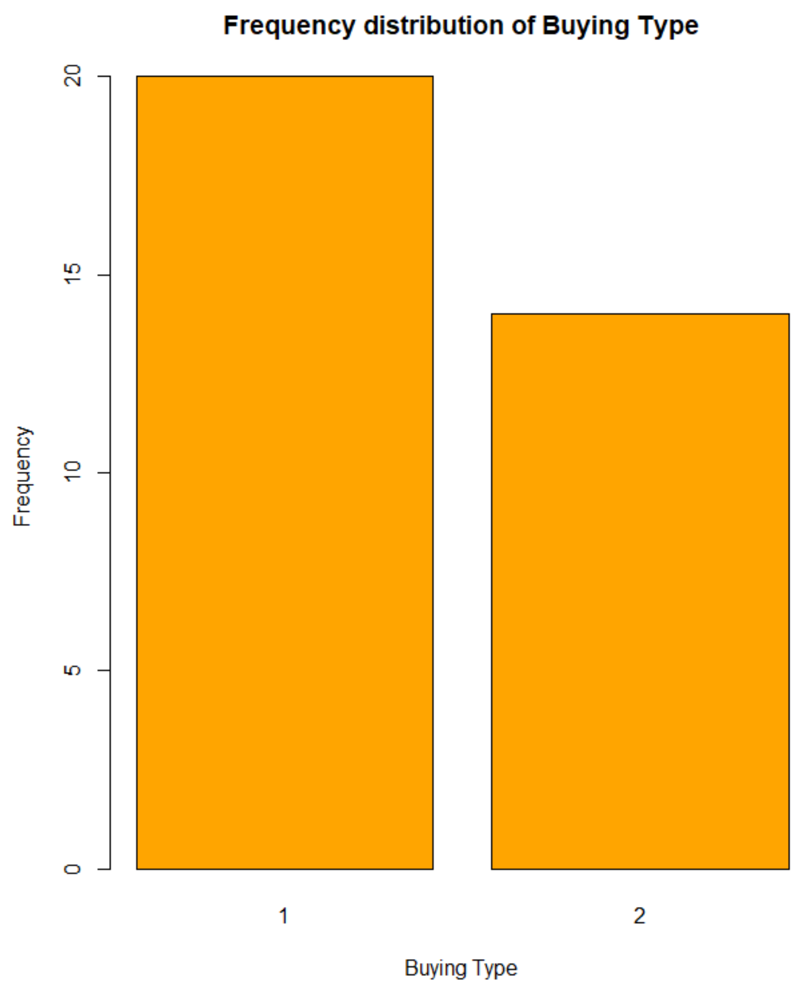
	Delivery speed	Price level	Price flexibility	Manufacturing image	Overall service	Salesforce image	Product quality	Usage Level	Satisfaction Level	Size of firm	Purchasing Structure	Industry	Buying Type	Classification
2	1.8	3.0	6.3	6.6	2.5	4.0	8.4	43	4.3	1	1	0	1	2
6	1.9	3.3	7.9	4.8	2.6	1.9	9.7	45	4.4	1	1	1	2	2
8	1.3	4.2	6.2	5.1	2.8	2.2	6.9	44	4.3	1	1	0	2	2
13	2.8	1.4	8.1	3.8	2.1	1.4	6.6	39	4.4	1	1	0	1	2
14	3.7	1.5	8.6	5.7	2.7	3.7	6.7	38	5.0	0	0	1	1	2
17	3.2	4.1	5.7	5.1	3.6	2.9	6.2	38	4.4	0	1	1	2	2
18	4.9	1.8	7.7	4.3	3.4	1.5	5.9	40	5.6	0	0	0	2	2
21	3.3	0.9	8.6	4.0	2.1	1.8	6.3	41	4.5	0	0	0	2	2
24	2.4	1.5	6.7	4.8	1.9	2.5	7.2	36	3.7	1	1	0	1	2
27	2.4	1.5	6.6	4.8	1.9	2.5	7.2	36	3.7	1	1	0	1	2
31	3.0	3.2	6.0	5.3	3.1	3.0	8.0	43	3.3	1	1	0	1	2
36	1.8	3.3	7.5	4.5	2.5	2.4	7.6	39	3.6	1	1	1	1	2
37	3.6	4.0	5.8	5.8	3.7	2.5	9.3	44	4.8	1	1	1	2	2
41	1.9	3.4	7.6	4.6	2.6	2.5	7.7	40	3.7	1	1	1	1	2
45	2.0	2.6	6.5	3.7	2.4	1.7	8.5	38	3.2	1	1	1	1	2
48	3.4	3.9	5.6	5.6	3.6	2.3	9.1	43	4.7	1	1	1	2	2
51	3.7	0.7	8.2	6.0	2.1	2.5	5.2	41	5.0	0	0	0	2	2
55	3.8	0.8	8.7	2.9	1.6	2.1	5.6	39	3.7	0	0	0	1	2
64	3.0	3.8	5.5	4.9	3.4	2.6	6.0	36	4.2	0	1	1	2	2
65	1.1	2.0	7.2	4.7	1.6	3.2	10.0	40	3.4	1	1	1	1	2
68	1.6	4.5	6.4	5.3	3.0	2.5	7.1	46	4.5	1	1	0	2	2
75	3.0	2.0	6.6	6.6	2.4	2.7	8.2	41	4.1	1	1	0	1	2
77	3.8	0.8	8.3	6.1	2.2	2.6	5.3	42	5.1	0	0	0	2	2
79	1.0	1.9	7.1	4.5	1.5	3.1	9.9	39	3.3	1	1	1	1	2
82	3.4	4.6	5.5	8.2	4.0	4.4	6.3	47	5.6	0	1	1	2	2
83	1.6	2.8	6.1	6.4	2.3	3.8	8.2	41	4.1	1	1	0	1	2

Below are the descriptive statistics of the customers in Cluster 2:

```
                     vars  n  mean   sd median trimmed  mad  min  max range  skew kurtosis   se
Delivery speed          1 34  2.68 0.98   2.65    2.65 1.11  1.0  4.9   3.9  0.30    -0.63 0.17
Price level             2 34  2.69 1.17   2.80    2.71 1.48  0.7  4.6   3.9 -0.12    -1.28 0.20
Price flexibility       3 34  6.84 1.08   6.70    6.84 1.33  5.0  8.7   3.7  0.04    -1.19 0.19
Manufacturing image     4 34  5.20 1.13   4.95    5.13 0.96  2.9  8.2   5.3  0.68     0.36 0.19
Overall service         5 34  2.66 0.65   2.60    2.65 0.67  1.5  4.0   2.5  0.19    -0.88 0.11
Salesforce image        6 34  2.61 0.73   2.50    2.58 0.52  1.4  4.4   3.0  0.56    -0.06 0.13
Product quality         7 34  7.45 1.36   7.30    7.42 1.63  5.2 10.0   4.8  0.15    -1.08 0.23
Usage Level             8 34 40.38 3.04  40.00   40.25 2.97 36.0 47.0  11.0  0.24    -0.89 0.52
Satisfaction Level      9 34  4.27 0.67   4.30    4.24 0.89  3.2  5.6   2.4  0.33    -0.85 0.12
Size of firm           10 34  0.62 0.49   1.00    0.64 0.00  0.0  1.0   1.0 -0.46    -1.84 0.08
Purchasing Structure   11 34  0.79 0.41   1.00    0.86 0.00  0.0  1.0   1.0 -1.39    -0.06 0.07
Industry               12 34  0.41 0.50   0.00    0.39 0.00  0.0  1.0   1.0  0.34    -1.94 0.09
Buying Type            13 34  1.41 0.50   1.00    1.39 0.00  1.0  2.0   1.0  0.34    -1.94 0.09
Classification         14 34  2.00 0.00   2.00    2.00 0.00  2.0  2.0   0.0   NaN      NaN 0.00
```

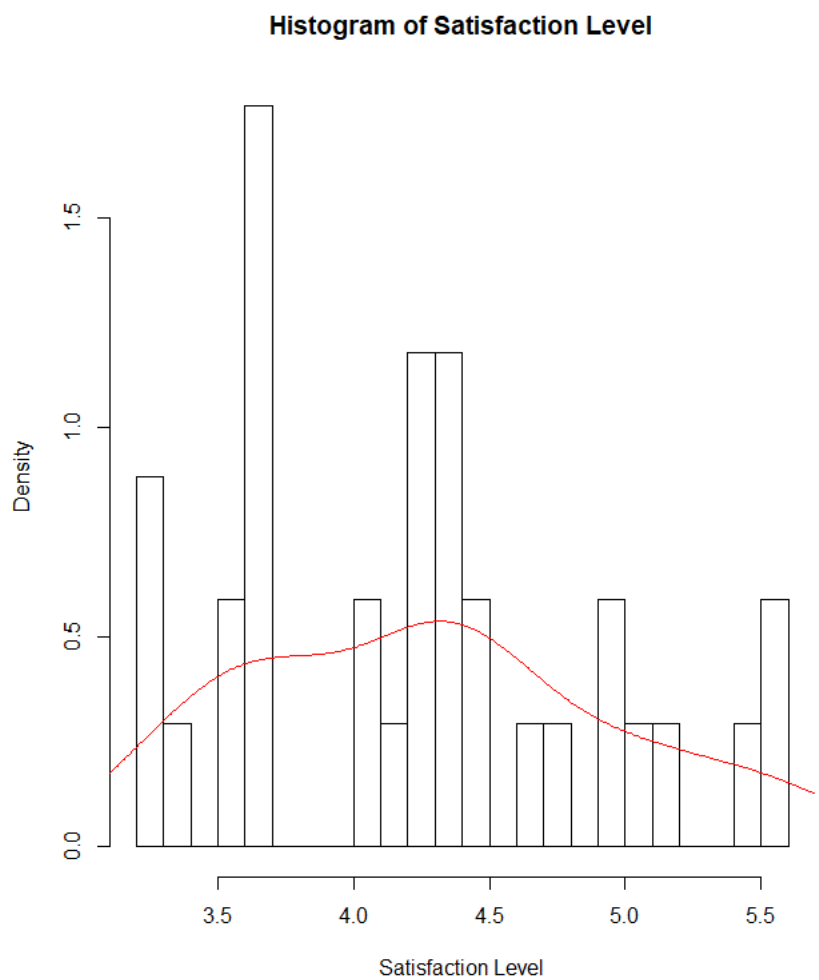
Analysis of each of the variables:

- **Buying Type:** Standard deviation of the Buying Type variable is 0.50. Customers in Cluster 2 are either in Buying Type group 1 (‘New Purchase’) or group 2 (‘Modified Rebuy’). Below is the frequency distribution of Buying Type.



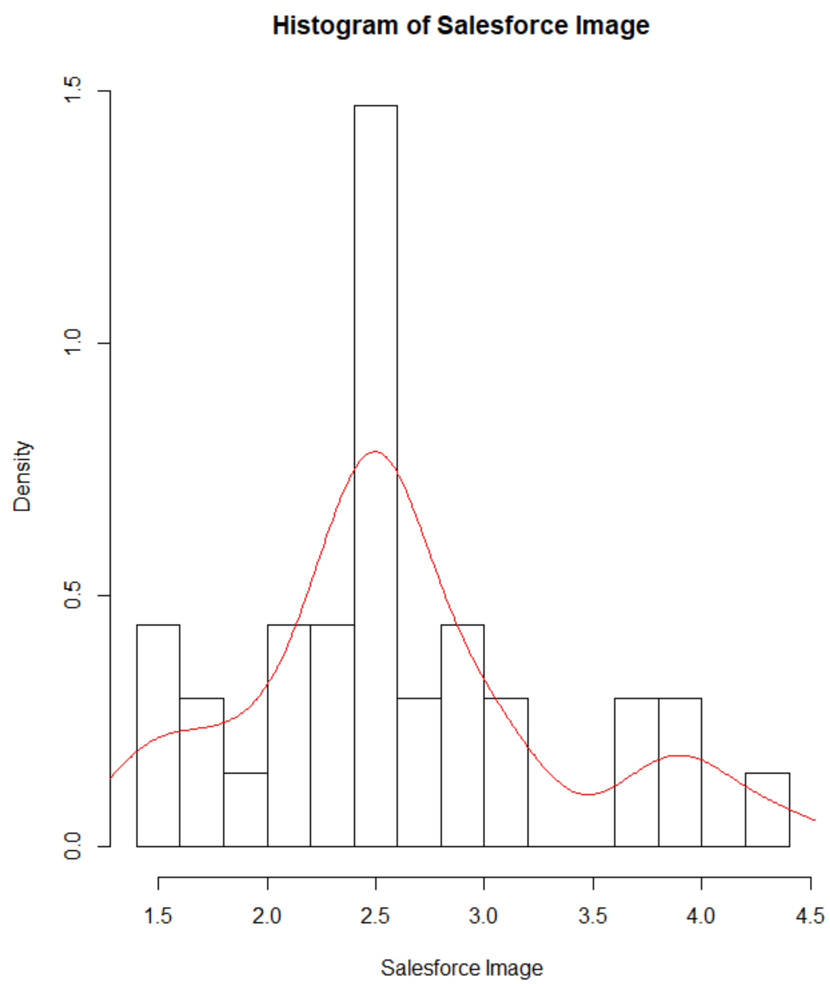
Graph 248: Frequency distribution of number of customers in each of the 2 Buying Type groups.

- **Satisfaction Level:** With a standard deviation of 0.67, Satisfaction Level varies from 3.2 to 5.6 and has a mean value of 4.27. Satisfaction level depicts how satisfied the purchaser is with past purchases from PLE, rated from 1 to 7. Customers in cluster 2 tend to rate their satisfaction levels moderately high (with ratings ranging from 3.2 to 5.6, and a mean value of 4.27). Below is the histogram and density plot of Satisfaction Level in cluster 2.



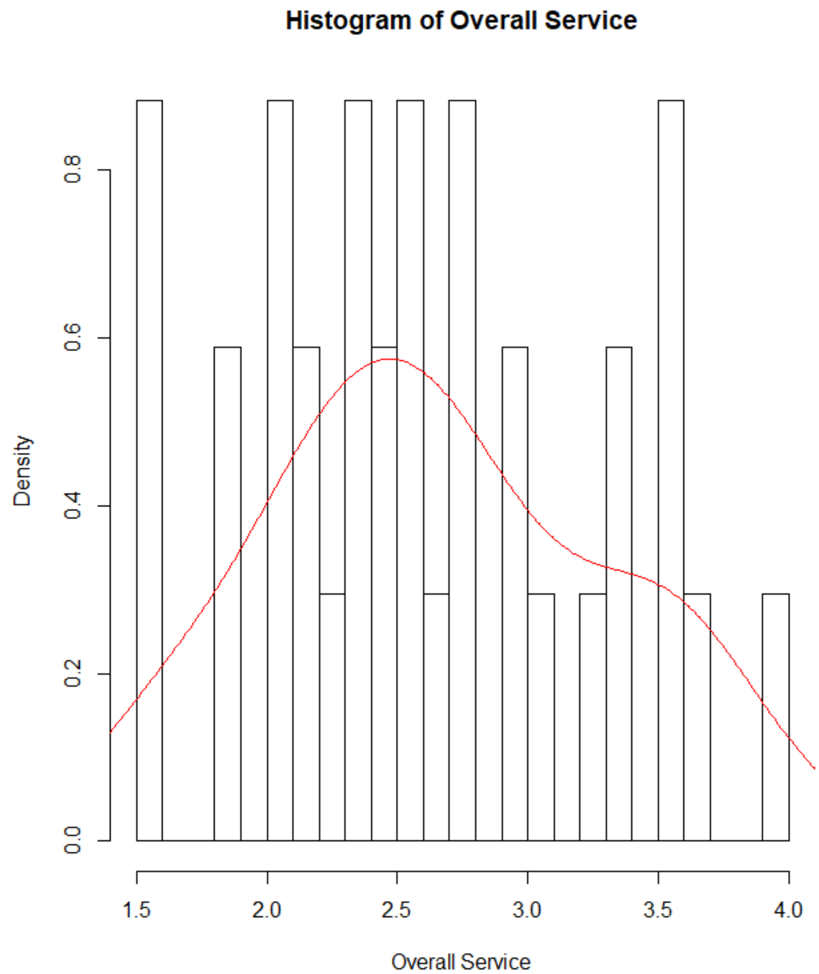
Graph 249: Histogram of Satisfaction Level ratings in Cluster 2, with density plot shown as the red line.

- **Salesforce Image:** With a standard deviation of 0.73, the distribution has a moderately narrow spread about its mean. Salesforce image varies from 1.4 to 4.4 and has a mean value of 2.61. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 2 tend to have rated Salesforce image low (ratings ranging from 1.1 to 4.4 out of 10.0, with a mean value of 2.61). Below is the histogram and density plot of Salesforce Image in cluster 2.



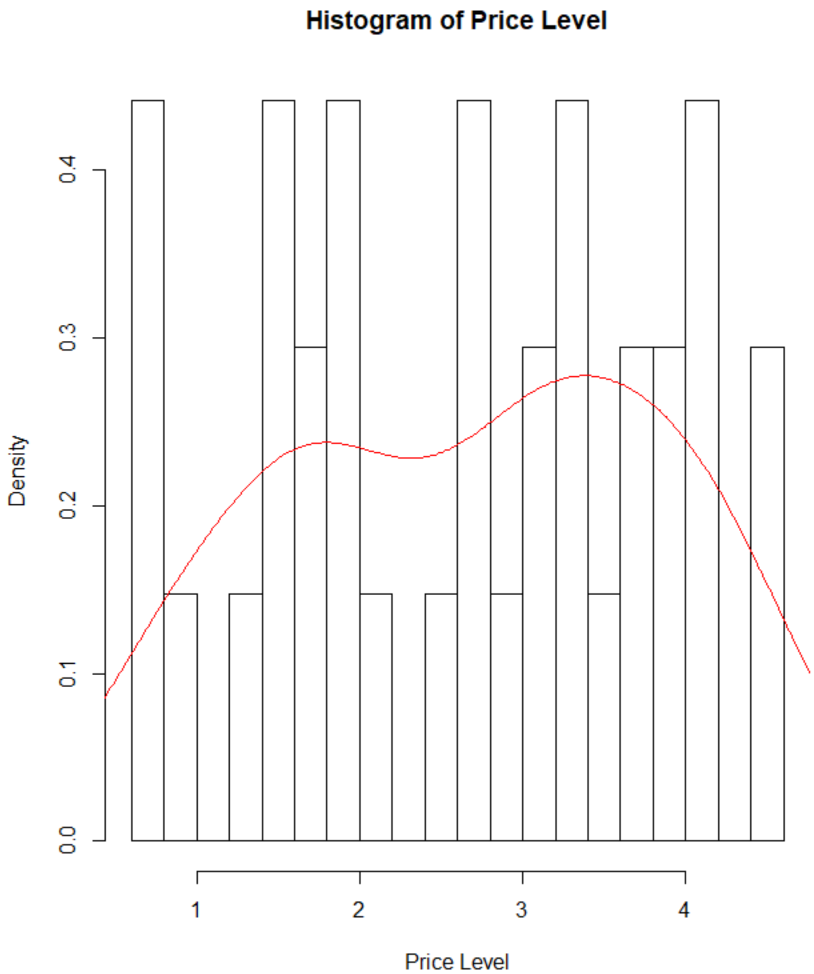
Graph 250: Histogram of Salesforce Image ratings in Cluster 2, with density plot shown as the red line.

- **Overall Service:** With a standard deviation of 0.65, the distribution has a moderately narrow spread about its mean. Overall Service varies from 1.5 to 4.0 and has a mean value of 2.66. Rating was done on a 10cm scale, rounded to 1 decimal place. Similar to Salesforce Image, customers in cluster 2 also tend to have rated Overall Service low (ratings ranging from 1.5 to 4.0 out of 10.0, with a mean value of 2.66). Below is the histogram and density plot of Overall Service in cluster 2.



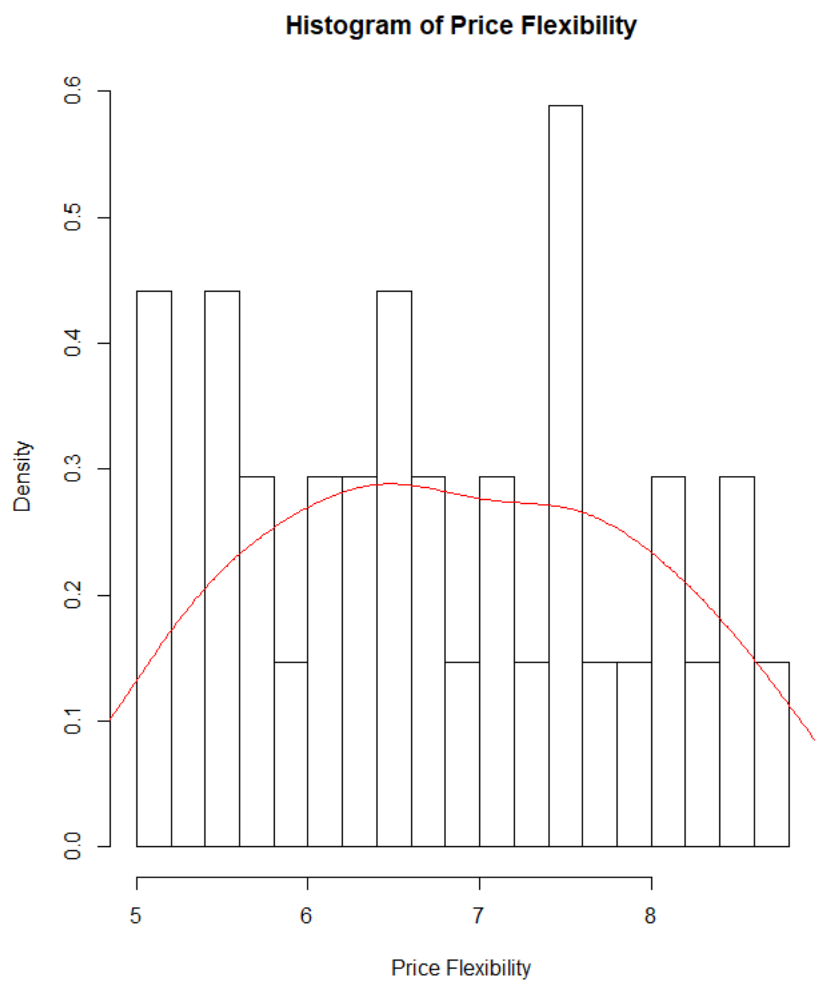
Graph 251: Histogram of Overall Service ratings in Cluster 2, with density plot shown as the red line.

- **Price Level:** With a standard deviation of 1.17, the distribution has a moderately narrow spread about its mean. Price Level varies from 0.7 to 4.6 and has a mean value of 2.69. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 2 tend to have rated Price Level low (ratings ranging from 0.7 to 4.6 out of 10.0, with a mean value of 2.69). Price level depicts the perceived level of price charged by PLE. Below is the histogram and density plot of Price Level in cluster 2.



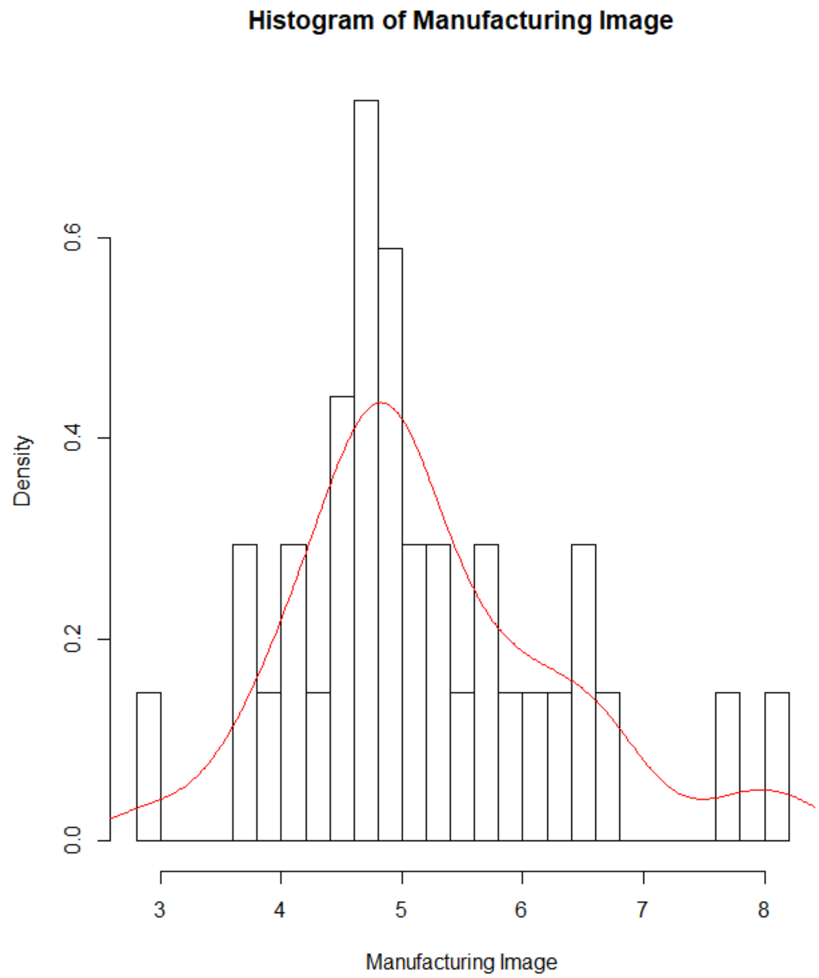
Graph 252: Histogram of Price Level ratings in Cluster 2, with density plot shown as the red line.

- **Price Flexibility:** With a standard deviation of 1.08, Price Flexibility varies from 5.0 to 8.7 and has a mean value of 6.84. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 2 tend to have rated Price Flexibility moderately high (ratings ranging from 5.0 to 8.7 out of 10.0, with a mean value of 6.84). Price flexibility depicts the perceived willingness of PLE representatives to negotiate prices on all types of purchases. Below is the histogram and density plot of Price Flexibility in cluster 2.



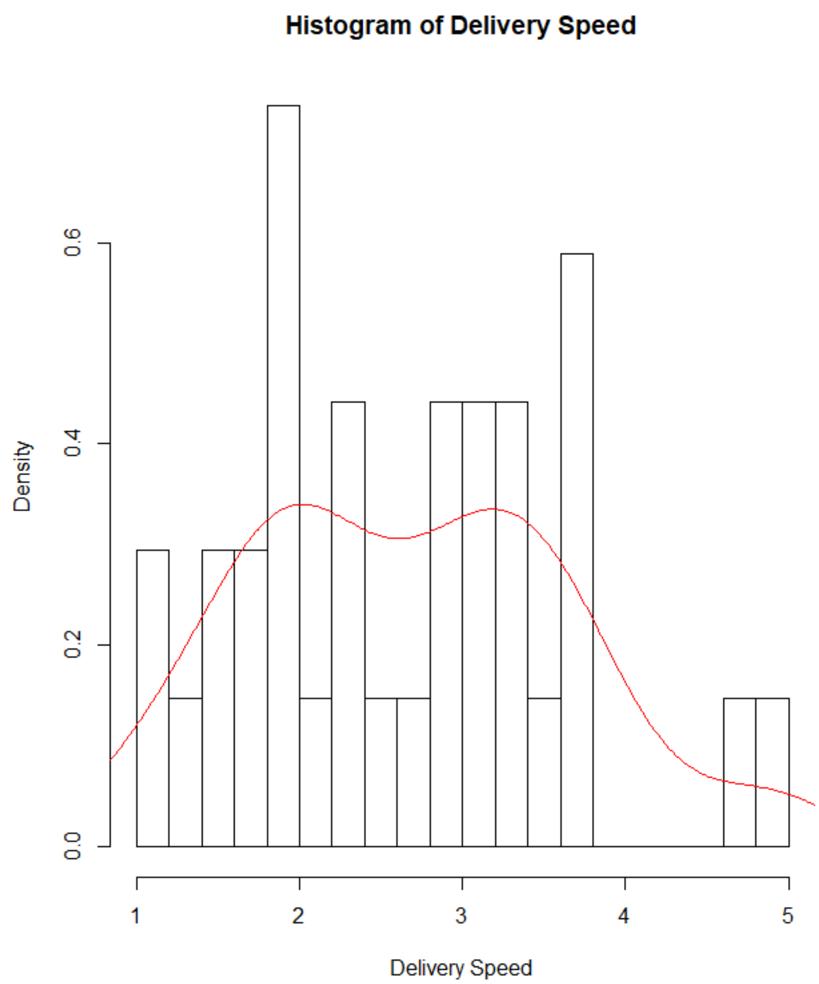
Graph 253: Histogram of Price Flexibility ratings in Cluster 2, with density plot shown as the red line.

- **Manufacturing Image:** With a standard deviation of 1.13, Manufacturing Image varies from 2.9 to 8.2 and has a mean value of 5.20. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 2 tend to have rated Manufacturing Image moderately low (ratings ranging from 2.9 to 8.2 out of 10.0, with a mean value of 5.20). Manufacturing Image depicts the overall image of the manufacturer. Below is the histogram and density plot of Manufacturing Image in cluster 2.



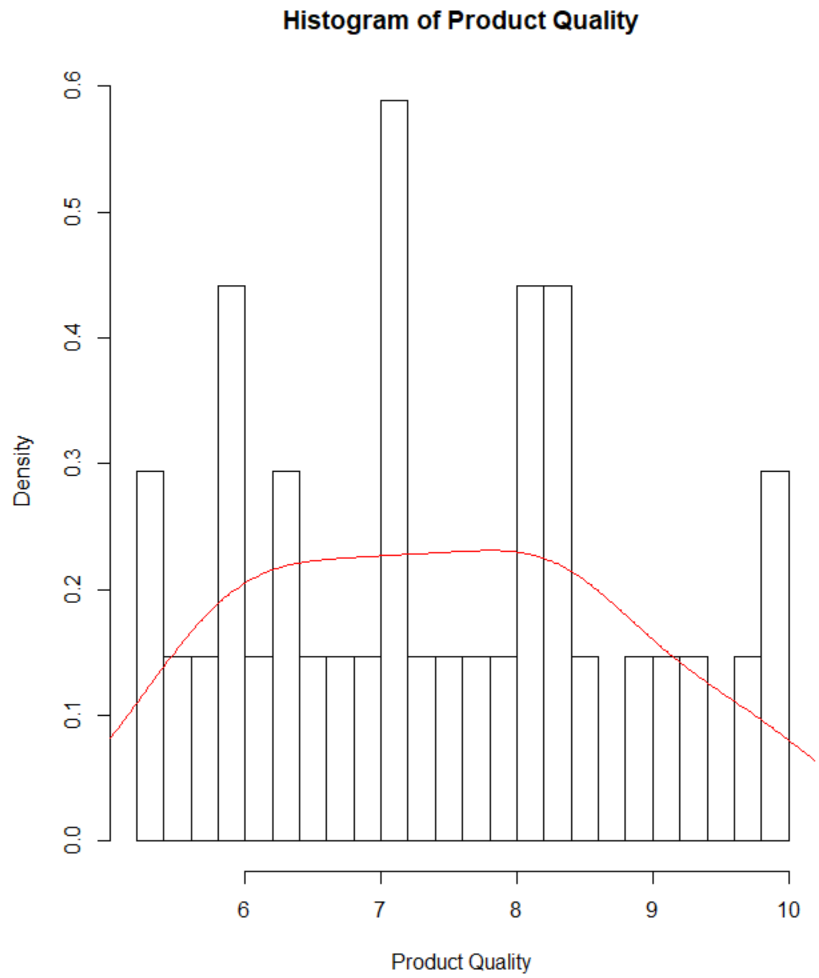
Graph 254: Histogram of Manufacturing Image ratings in Cluster 2, with density plot shown as the red line.

- **Delivery Speed:** With a standard deviation of 0.98, Delivery Speed varies from 1.0 to 4.9 and has a mean value of 2.68. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 2 tend to have rated Delivery Speed low (ratings ranging from 1.0 to 4.9 out of 10.0, with a mean value of 2.68). Delivery Speed depicts the amount of time it takes to deliver product once an order is confirmed. Below is the histogram and density plot of Delivery Speed in cluster 2.



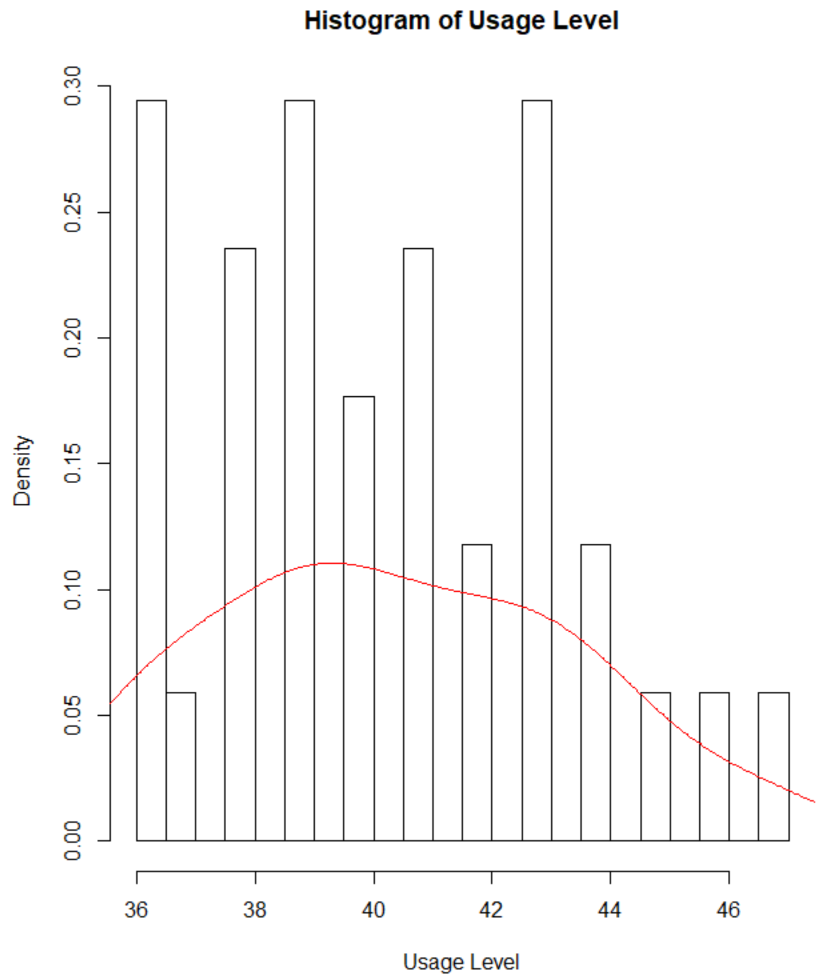
Graph 255: Histogram of Delivery Speed ratings in Cluster 2, with density plot shown as the red line.

- **Product Quality:** With a standard deviation of 1.36, Product Quality varies from 5.2 to 10.0 and has a mean value of 7.45. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 2 tend to have rated Product Quality moderately high (ratings ranging from 5.2 to 10.0 out of 10.0, with a mean value of 7.45). Product Quality depicts the perceived level of quality of PLE products. Below is the histogram and density plot of Product Quality in cluster 2.



Graph 256: Histogram of Product Quality ratings in Cluster 2, with density plot shown as the red line.

- **Usage Level:** With a standard deviation of 3.04, Usage Level varies from 36.0 to 47.0 and has a mean value of 40.38. Usage Level depicts how much of the firm's total product is purchased from PLE, measured on a 100-point scale, ranging from 0% to 100%. Customers in cluster 1 tend to rate the usage level moderately low (ratings ranging from 36.0% to 47.0% out of 100.0%, with a mean value of 40.38%). Below is the histogram and density plot of Usage Level in cluster 2.



Graph 257: Histogram of Usage Level ratings in Cluster 2, with density plot shown as the red line.

Thirdly, for group three, below is a screenshot of some of the rows of customers in the data set who have been classified under Cluster 3.

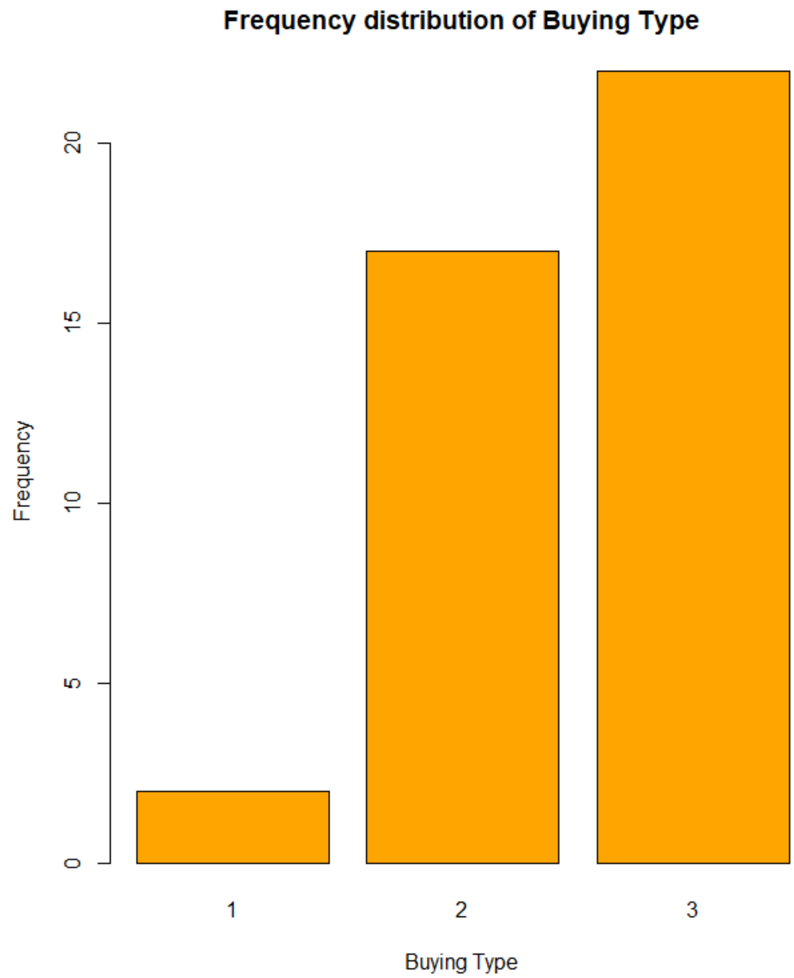
	Delivery speed	Price level	Price flexibility	Manufacturing image	Overall service	Salesforce image	Product quality	Usage Level	Satisfaction Level	Size of firm	Purchasing Structure	Industry	Buying Type	Classification
3	3.4	5.2	5.7	6.0	4.3	2.7	8.2	48	5.2	1	1	1	2	3
7	4.6	2.4	9.5	6.6	3.5	4.5	7.6	46	5.8	0	0	1	1	3
10	4.0	3.5	6.5	6.0	3.7	3.2	8.7	54	5.4	1	1	0	2	3
12	3.9	2.2	9.1	4.6	3.0	2.5	8.3	47	5.0	0	0	1	2	3
15	4.7	1.3	9.9	6.7	3.0	2.6	6.8	54	5.9	0	0	0	3	3
16	3.4	2.0	9.7	4.7	2.7	1.7	4.8	49	4.7	0	0	0	3	3
19	5.3	1.4	9.7	6.1	3.3	3.9	6.8	54	5.9	0	0	1	3	3
20	4.7	1.3	9.9	6.7	3.0	2.6	6.8	55	6.0	0	0	0	3	3
23	3.0	4.0	9.1	7.1	3.5	3.4	8.4	55	5.2	0	1	0	3	3
25	5.1	1.4	8.7	4.8	3.3	2.6	3.8	49	4.9	0	0	0	2	3
26	4.6	2.1	7.9	5.8	3.4	2.8	4.7	49	5.9	0	0	1	3	3
28	5.2	1.3	9.7	6.1	3.2	3.9	6.7	54	5.8	0	0	1	3	3
29	3.5	2.8	9.9	3.5	3.1	1.7	5.4	49	5.4	0	0	1	3	3
30	4.1	3.7	5.9	5.5	3.9	3.0	8.4	46	5.1	1	1	0	2	3
32	2.8	3.8	8.9	6.9	3.3	3.2	8.2	53	5.0	0	1	0	3	3
34	3.4	3.7	6.4	5.7	3.5	3.4	8.4	47	3.8	1	1	0	1	3
38	4.0	0.9	9.1	5.4	2.4	2.6	7.3	46	5.1	0	0	1	3	3
43	4.9	2.3	9.3	4.5	3.6	1.3	6.2	53	5.9	0	0	0	3	3
44	5.0	1.3	8.6	4.7	3.1	2.5	3.7	48	4.8	0	0	0	2	3
46	5.0	2.5	9.4	4.6	3.7	1.4	6.3	54	6.0	0	0	0	3	3
47	3.1	1.9	10.0	4.5	2.6	3.2	3.8	55	4.9	0	0	1	3	3
49	5.8	0.2	8.8	4.5	3.0	2.4	6.7	57	4.9	0	0	1	3	3
50	5.4	2.1	8.0	3.0	3.8	1.4	5.2	53	3.8	0	0	1	3	3
52	2.6	4.8	8.2	5.0	3.6	2.5	9.0	53	5.2	1	1	1	2	3
53	4.5	4.1	6.3	5.9	4.3	3.4	8.8	50	5.5	1	1	0	2	3
56	2.9	2.6	7.7	7.0	2.8	3.6	7.7	47	4.2	0	1	1	2	3

Below are the descriptive statistics of the customers in Cluster 3:

```
                     vars  n  mean   sd median trimmed  mad  min  max range  skew kurtosis   se
Delivery speed          1 41  3.94 0.97    4.0    3.93 1.04  2.3  5.8   3.5  0.05    -1.12 0.15
Price level             2 41  2.54 1.22    2.3    2.45 1.33  0.2  5.4   5.2  0.55    -0.42 0.19
Price flexibility       3 41  8.55 1.23    8.8    8.70 1.19  5.7 10.0   4.3 -0.88    -0.25 0.19
Manufacturing image     4 41  5.31 1.07    5.4    5.35 1.04  3.0  7.1   4.1 -0.21    -0.61 0.17
Overall service         5 41  3.24 0.50    3.1    3.20 0.44  2.4  4.5   2.1  0.61    -0.22 0.08
Salesforce image        6 41  2.66 0.75    2.6    2.65 0.74  1.3  4.5   3.2  0.13    -0.40 0.12
Product quality         7 41  6.77 1.64    6.8    6.87 2.22  3.7  9.1   5.4 -0.41    -1.10 0.26
Usage Level             8 41 50.66 3.39   50.0   50.64 4.45 45.0 57.0  12.0  0.08    -1.39 0.53
Satisfaction Level      9 41  5.14 0.57    5.1    5.19 0.44  3.8  6.0   2.2 -0.57    -0.03 0.09
Size of firm           10 41  0.27 0.45    0.0    0.21 0.00  0.0  1.0   1.0  1.01    -1.01 0.07
Purchasing Structure   11 41  0.37 0.49    0.0    0.33 0.00  0.0  1.0   1.0  0.54    -1.75 0.08
Industry               12 41  0.56 0.50    1.0    0.58 0.00  0.0  1.0   1.0 -0.24    -1.99 0.08
Buying Type            13 41  2.49 0.60    3.0    2.55 0.00  1.0  3.0   2.0 -0.64    -0.64 0.09
Classification         14 41  3.00 0.00    3.0    3.00 0.00  3.0  3.0   0.0   NaN      NaN 0.00
```

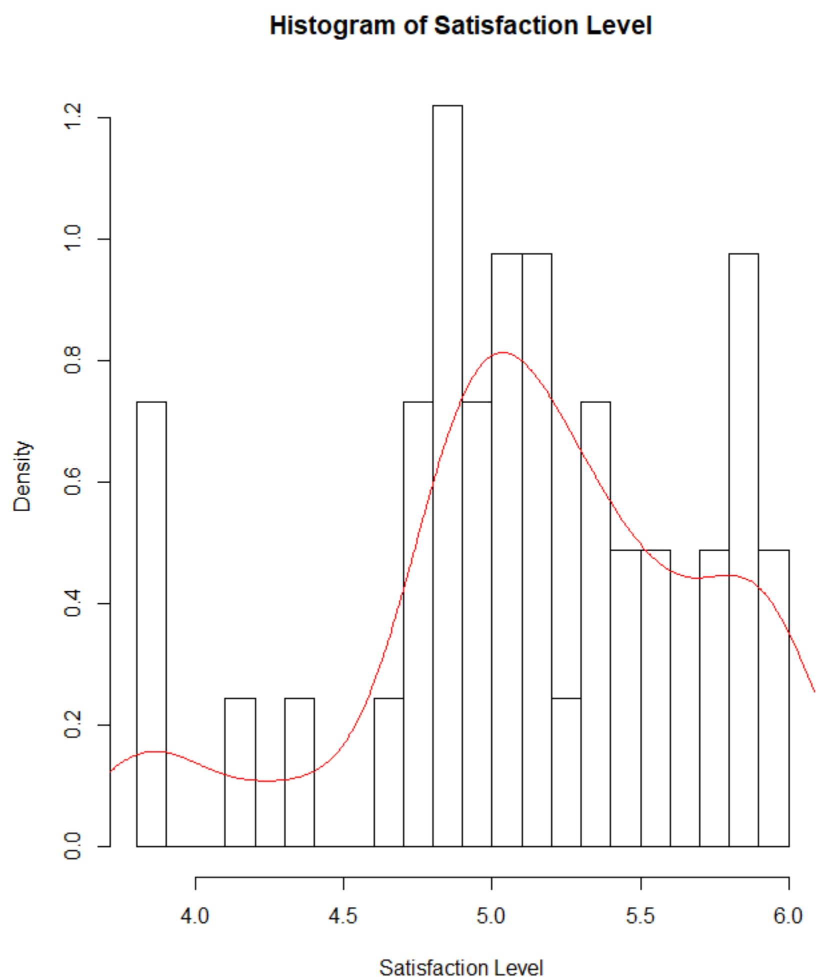
Analysis of each of the variables:

- **Buying Type:** Standard deviation of the Buying Type variable is 0.60. Customers in Cluster 3 are either in Buying Type group 1 (‘New Purchase’) or group 2 (‘Modified Rebuy’) or in Buying Type group 3 (‘straight rebuy’). Below is the frequency distribution of the number of customers under each Buying Type.



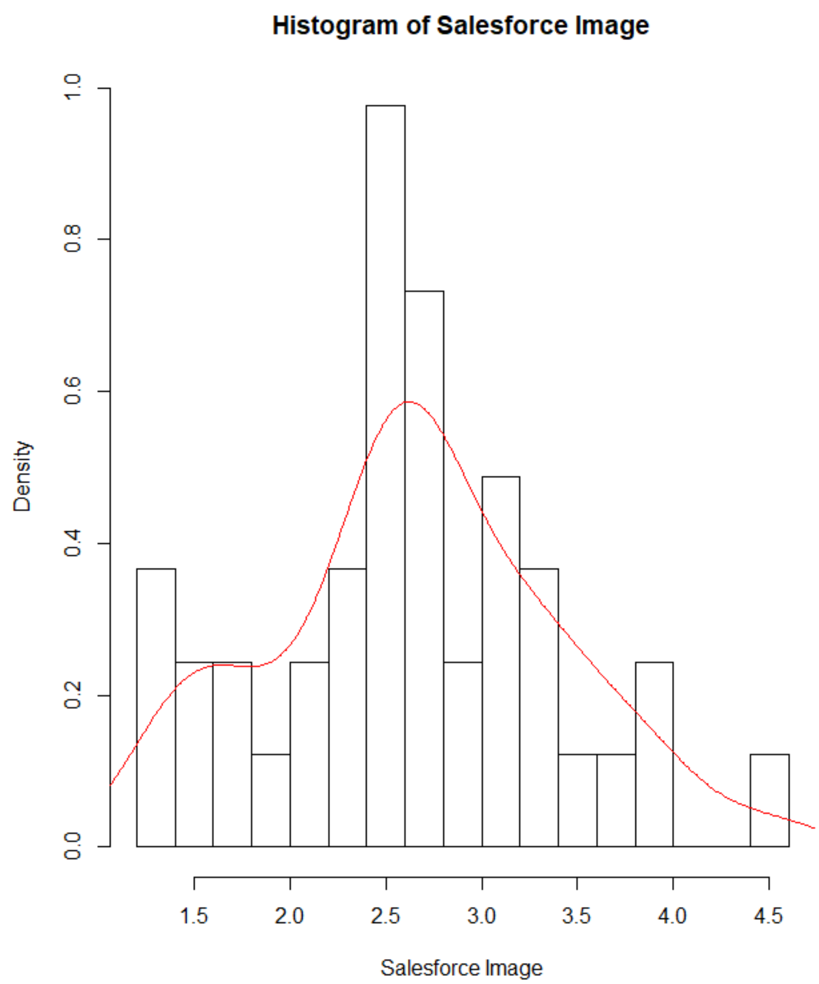
Graph 258: Frequency distribution of number of customers in each of the 3 Buying Type groups in cluster 3.

- **Satisfaction Level:** With a standard deviation of 0.57, Satisfaction Level varies from 3.8 to 6.0 and has a mean value of 5.14. Satisfaction level depicts how satisfied the purchaser is with past purchases from PLE, rated from 1 to 7. Customers in cluster 3 tend to give a moderately high rating to satisfaction level (ratings ranging from 3.8 to 6.0, with a mean of 5.14). Below is the histogram and density plot of Satisfaction Level in cluster 3.



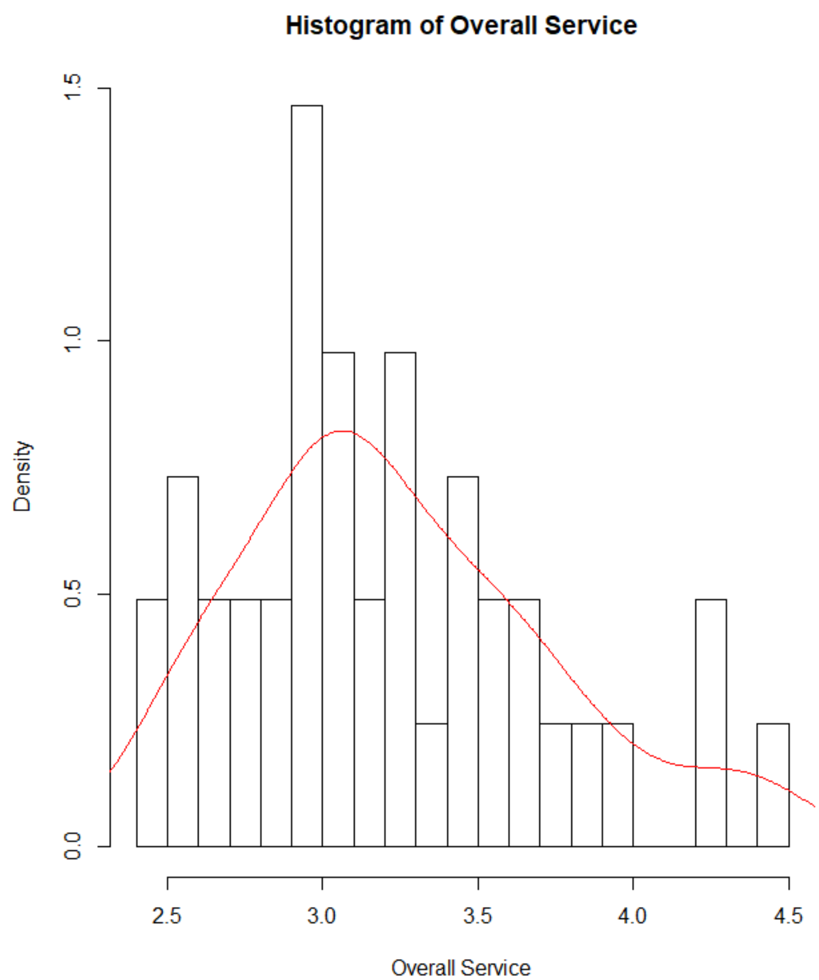
Graph 259: Histogram of Satisfaction Level ratings in Cluster 3, with density plot shown as the red line.

- **Salesforce Image:** With a standard deviation of 0.75, the distribution has a moderately narrow spread about its mean. Salesforce image varies from 1.3 to 4.5 and has a mean value of 2.66. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 3 tend to have rated Salesforce image low (ratings ranging from 1.3 to 4.5 out of 10.0, with a mean value of 2.66). Below is the histogram and density plot of Salesforce Image in cluster 3.



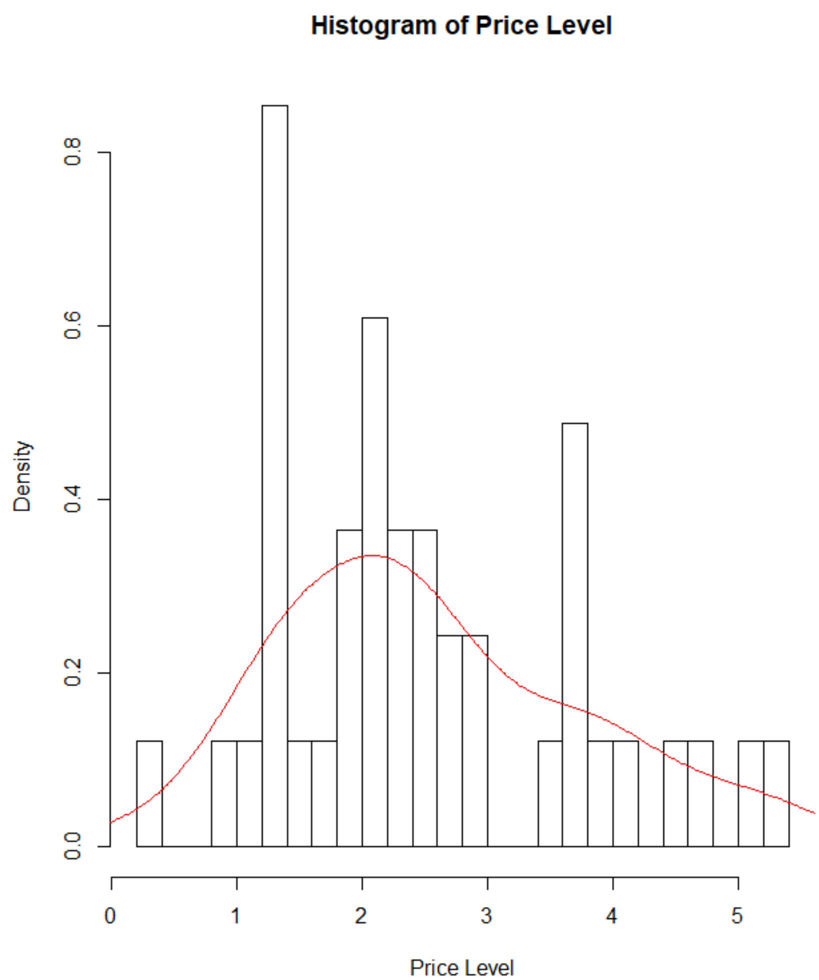
Graph 260: Histogram of Salesforce Image ratings in Cluster 3, with density plot shown as the red line.

- **Overall Service:** With a standard deviation of 0.50, the distribution has a moderately narrow spread about its mean. Overall Service varies from 2.4 to 4.5 and has a mean value of 3.24. Rating was done on a 10cm scale, rounded to 1 decimal place. Similar to Salesforce Image, customers in cluster 3 tend to have rated Overall Service low (ratings ranging from 2.4 to 4.5 out of 10.0, with a mean value of 3.24). Below is the histogram and density plot of Overall Service in cluster 3.



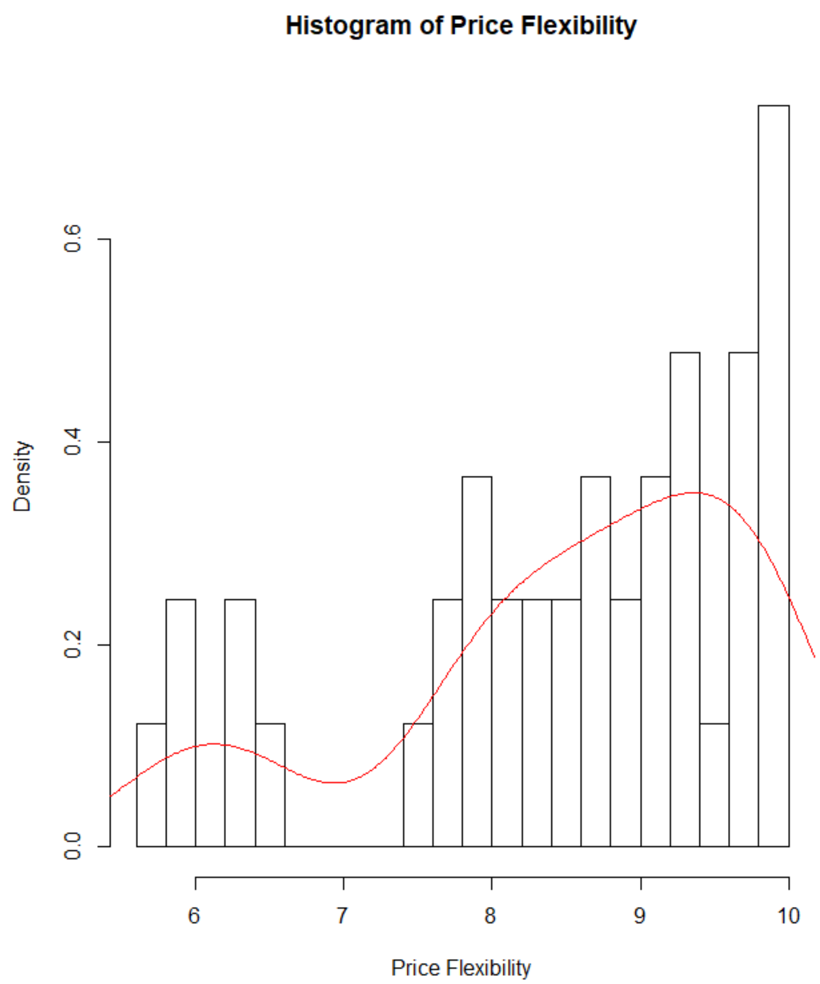
Graph 261: Histogram of Overall Service ratings in Cluster 3, with density plot shown as the red line.

- **Price Level:** With a standard deviation of 1.22, Price Level varies from 0.2 to 5.4 and has a mean value of 2.54. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 3 tend to have rated Price Level low (ratings ranging from 0.2 to 5.4 out of 10.0, with a mean value of 2.54), with quite a broad range of 5.2. Price level depicts the perceived level of price charged by PLE. Below is the histogram and density plot of Price Level in cluster 3.



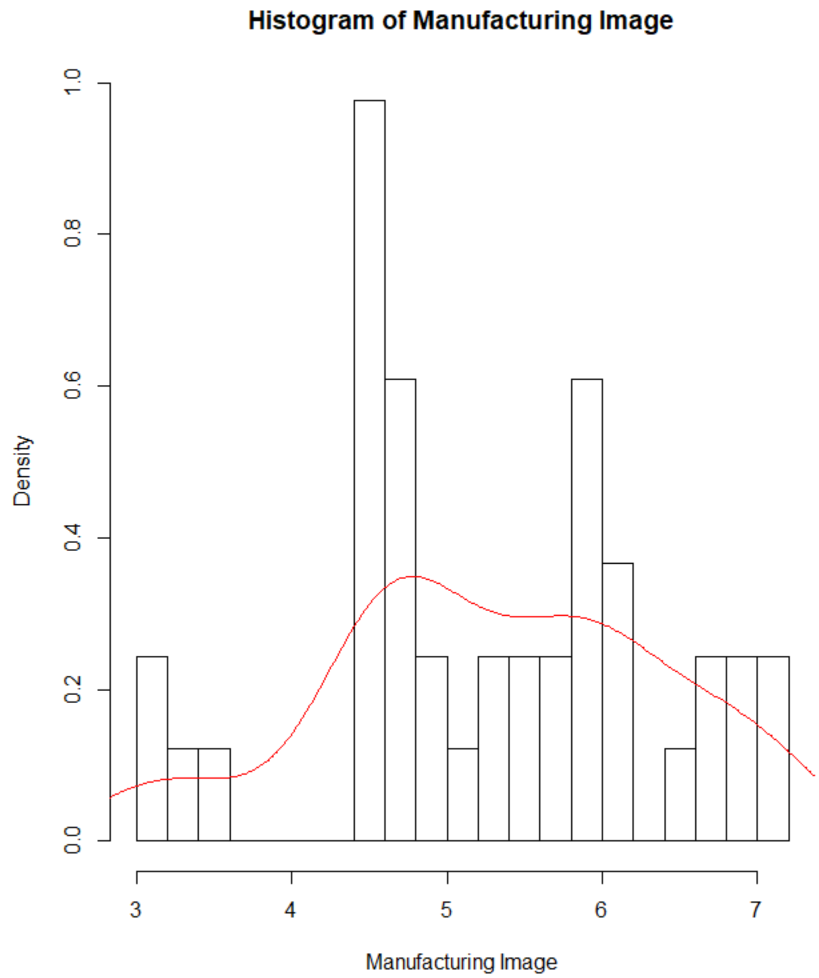
Graph 262: Histogram of Price Level ratings in Cluster 3, with density plot shown as the red line.

- **Price Flexibility:** With a standard deviation of 1.23, Price Flexibility varies from 5.7 to 10.0 and has a mean value of 8.55. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 3 tend to have rated Price Flexibility relatively high (ratings ranging from 5.7 to 10.0 out of 10.0, with a high mean of 8.55). Price flexibility depicts the perceived willingness of PLE representatives to negotiate prices on all types of purchases. Below is the histogram and density plot of Price Flexibility in cluster 3.



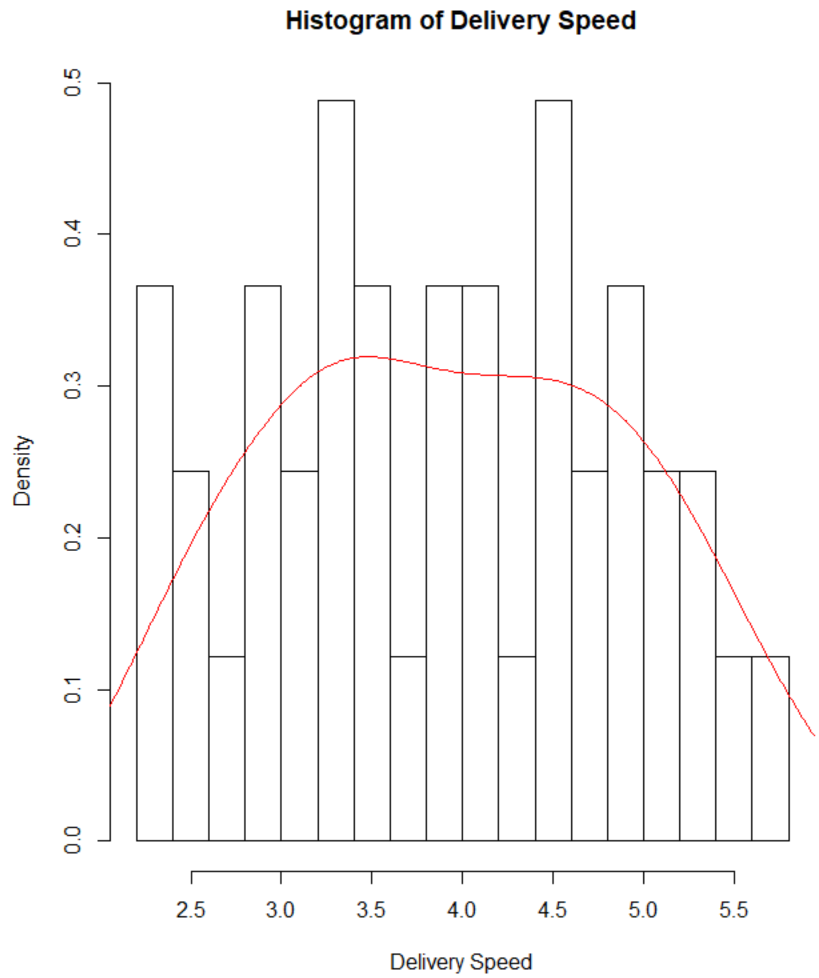
Graph 263: Histogram of Price Flexibility ratings in Cluster 3, with density plot shown as the red line.

- **Manufacturing Image:** With a standard deviation of 1.07, Manufacturing Image varies from 3.0 to 7.1 and has a mean value of 5.31. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 3 tend to have rated Manufacturing Image moderately low (ratings ranging from 3.0 to 7.1 out of 10.0, with a mean of 5.31). Manufacturing Image depicts the overall image of the manufacturer. Below is the histogram and density plot of Manufacturing Image in cluster 3.



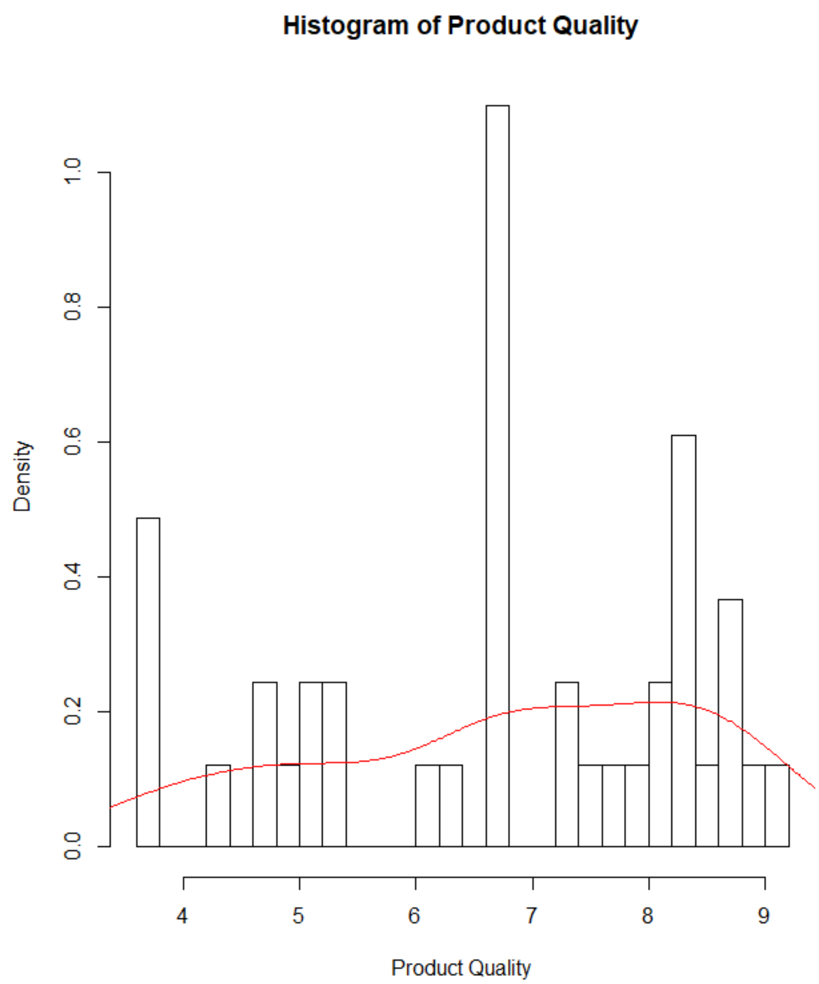
Graph 264: Histogram of Manufacturing Image ratings in Cluster 3, with density plot shown as the red line.

- **Delivery Speed:** With a standard deviation of 0.97, Delivery Speed varies from 2.3 to 5.8 and has a mean value of 3.94. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 3 tend to have rated Delivery Speed low (ratings ranging from 2.3 to 5.8 out of 10.0, with a mean value of 3.94). Delivery Speed depicts the amount of time it takes to deliver product once an order is confirmed. Below is the histogram and density plot of Delivery Speed in cluster 3.



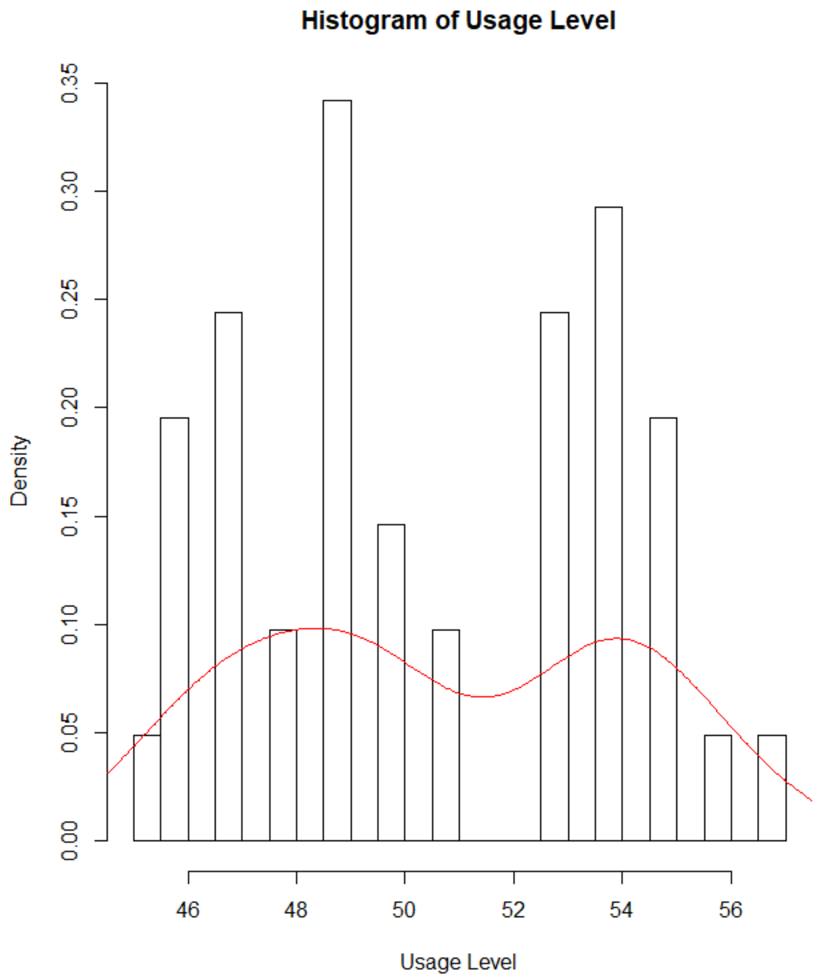
Graph 265: Histogram of Delivery Speed ratings in Cluster 3, with density plot shown as the red line.

- **Product Quality:** With a standard deviation of 1.64, Product Quality varies from 3.7 to 9.1 and has a mean value of 6.77. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 3 tend to have rated Product Quality moderately high (ratings ranging from 3.7 to 9.1 out of 10.0, with a mean of 6.77). Product Quality depicts the perceived level of quality of PLE products. Below is the histogram and density plot of Product Quality in cluster 3.



Graph 266: Histogram of Product Quality ratings in Cluster 3, with density plot shown as the red line.

- **Usage Level:** With a standard deviation of 3.39, Usage Level varies from 45.0 to 57.0 and has a mean value of 50.66. Usage Level depicts how much of the firm's total product is purchased from PLE, measured on a 100-point scale, ranging from 0% to 100%. Customers in cluster 3 tend to rate the usage level moderately high - as compared to the customers in cluster 1 and 2 - (ratings ranging from 45.0% to 57.0% out of 100.0%, with a mean value of 50.66). Below is the histogram and density plot of Usage Level in cluster 3.



Graph 267: Histogram of Usage Level ratings in Cluster 3, with density plot shown as the red line.

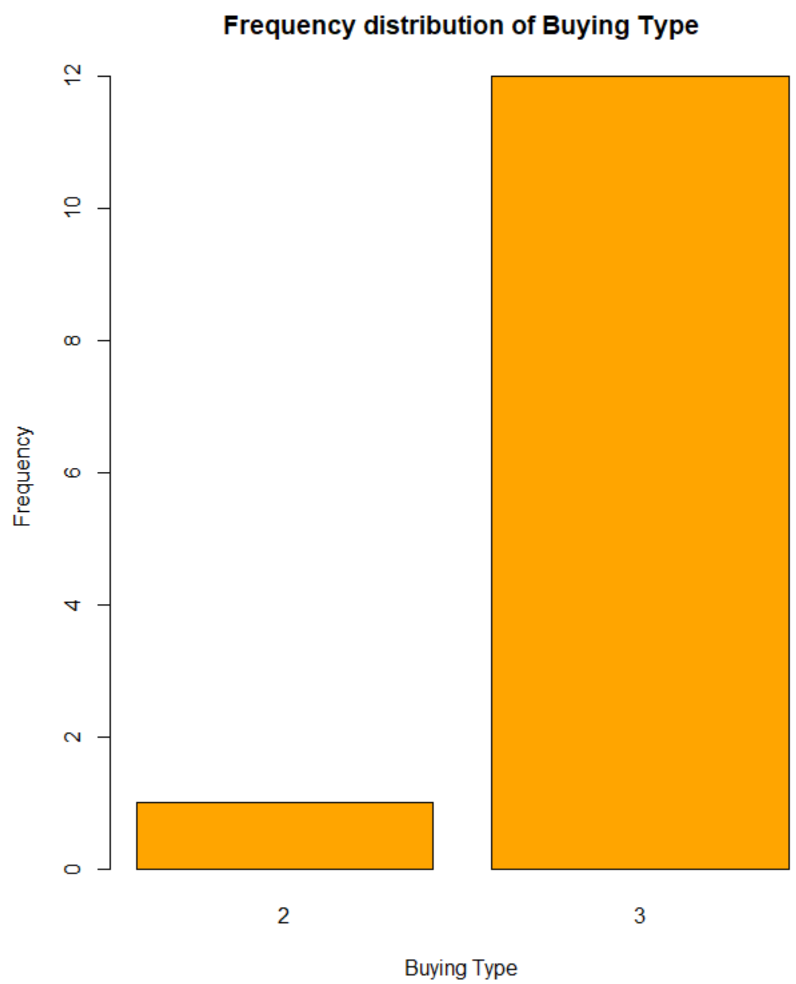
Fourthly, for group four, below is a screenshot of some of the rows of customers in the data set who have been classified under Cluster 4.

	Delivery speed	Price level	Price flexibility	Manufacturing image	Overall service	Salesforce image	Product quality	Usage Level	Satisfaction Level	Size of firm	Purchasing Structure	Industry	Buying Type	Classification
1	6.0	0.9	9.6	7.8	3.4	4.6	4.5	58	6.8	0	0	1	3	4
2	5.5	1.6	9.4	4.7	3.5	3.0	7.6	63	5.4	0	0	1	3	4
3	5.2	2.0	9.3	5.9	3.7	2.4	4.6	60	6.1	0	0	0	3	4
4	5.9	0.9	9.6	7.8	3.4	4.6	4.5	58	6.7	0	0	1	3	4
5	4.9	4.4	7.4	6.9	4.6	4.0	9.6	62	6.2	1	1	0	2	4
6	5.4	2.5	9.6	5.5	4.0	3.0	7.7	65	6.0	0	0	0	3	4
7	5.1	1.9	9.2	5.8	3.6	2.3	4.5	60	6.1	0	0	0	3	4
8	4.2	2.5	9.2	6.2	3.3	3.9	7.3	59	6.0	0	0	0	3	4
9	5.3	1.7	8.5	3.7	3.5	1.9	4.8	58	4.3	0	0	0	3	4
10	5.2	1.3	9.1	4.5	3.3	2.7	7.3	60	5.1	0	0	1	3	4
11	5.5	1.8	8.7	3.8	3.6	2.1	4.9	59	4.5	0	0	0	3	4
12	4.3	2.5	9.3	6.3	3.4	4.0	7.4	60	6.1	0	0	0	3	4
13	6.1	0.5	9.2	4.8	3.3	2.8	7.1	60	5.2	0	0	1	3	4

Below are the descriptive statistics of the customers in Cluster 4:

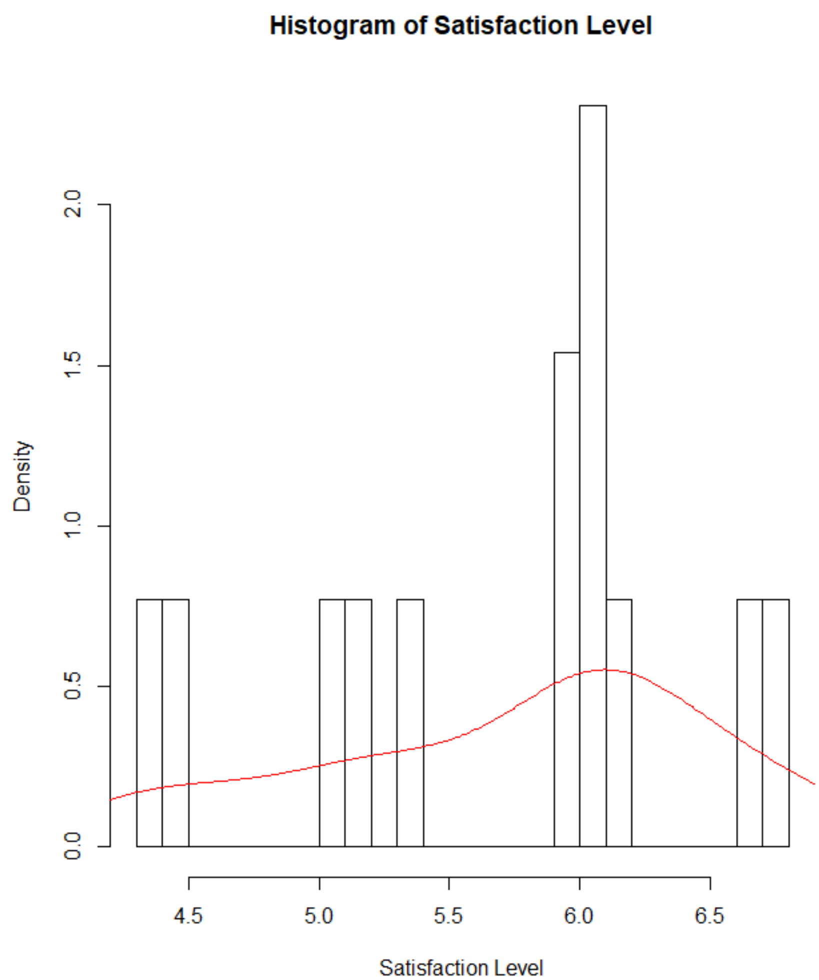
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Delivery speed	1	13	5.28	0.58	5.3	5.30	0.30	4.2	6.1	1.9	-0.42	-0.81	0.16
Price level	2	13	1.88	0.99	1.8	1.78	1.04	0.5	4.4	3.9	0.93	0.67	0.27
Price flexibility	3	13	9.08	0.60	9.2	9.19	0.30	7.4	9.6	2.2	-1.61	1.91	0.17
Manufacturing image	4	13	5.67	1.35	5.8	5.65	1.63	3.7	7.8	4.1	0.13	-1.27	0.37
Overall service	5	13	3.58	0.36	3.5	3.52	0.15	3.3	4.6	1.3	1.66	1.97	0.10
Salesforce image	6	13	3.18	0.94	3.0	3.16	1.33	1.9	4.6	2.7	0.24	-1.56	0.26
Product quality	7	13	6.29	1.71	7.1	6.15	3.26	4.5	9.6	5.1	0.26	-1.41	0.48
Usage Level	8	13	60.15	2.08	60.0	59.91	1.48	58.0	65.0	7.0	0.95	-0.11	0.58
Satisfaction Level	9	13	5.73	0.78	6.0	5.76	0.89	4.3	6.8	2.5	-0.46	-1.07	0.21
Size of firm	10	13	0.08	0.28	0.0	0.00	0.00	0.0	1.0	1.0	2.82	6.44	0.08
Purchasing Structure	11	13	0.08	0.28	0.0	0.00	0.00	0.0	1.0	1.0	2.82	6.44	0.08
Industry	12	13	0.38	0.51	0.0	0.36	0.00	0.0	1.0	1.0	0.42	-1.96	0.14
Buying Type	13	13	2.92	0.28	3.0	3.00	0.00	2.0	3.0	1.0	-2.82	6.44	0.08
Classification	14	13	4.00	0.00	4.0	4.00	0.00	4.0	4.0	0.0	NaN	NaN	0.00

- **Buying Type:** Standard deviation of the Buying Type variable is 0.28. Customers in Cluster 4 are mostly from Buying Type group 3 (‘straight rebuy’) except for 1 customer who is in Buying Type group 2 (‘modified rebuy’). Below is the frequency distribution of Buying Type.



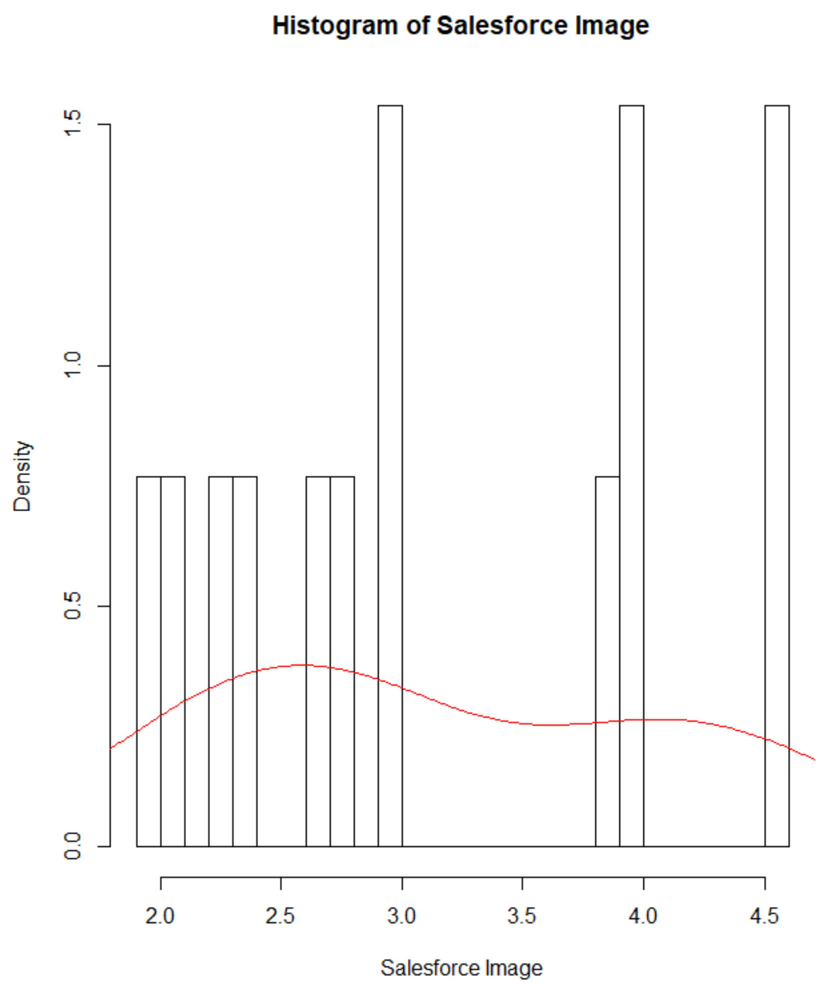
Graph 268: Frequency distribution of number of customers in each of the 2 Buying Type groups in cluster 4.

- **Satisfaction Level:** With a standard deviation of 0.78, Satisfaction Level varies from 4.3 to 6.8 and has a mean value of 5.73. Satisfaction level depicts how satisfied the purchaser is with past purchases from PLE, rated from 1 to 7. Customers in cluster 4 tend to give a moderately high rating to satisfaction level (ratings ranging from 4.3 to 6.8, with a mean value of 5.73). Below is the histogram and density plot of Satisfaction Level in cluster 4.



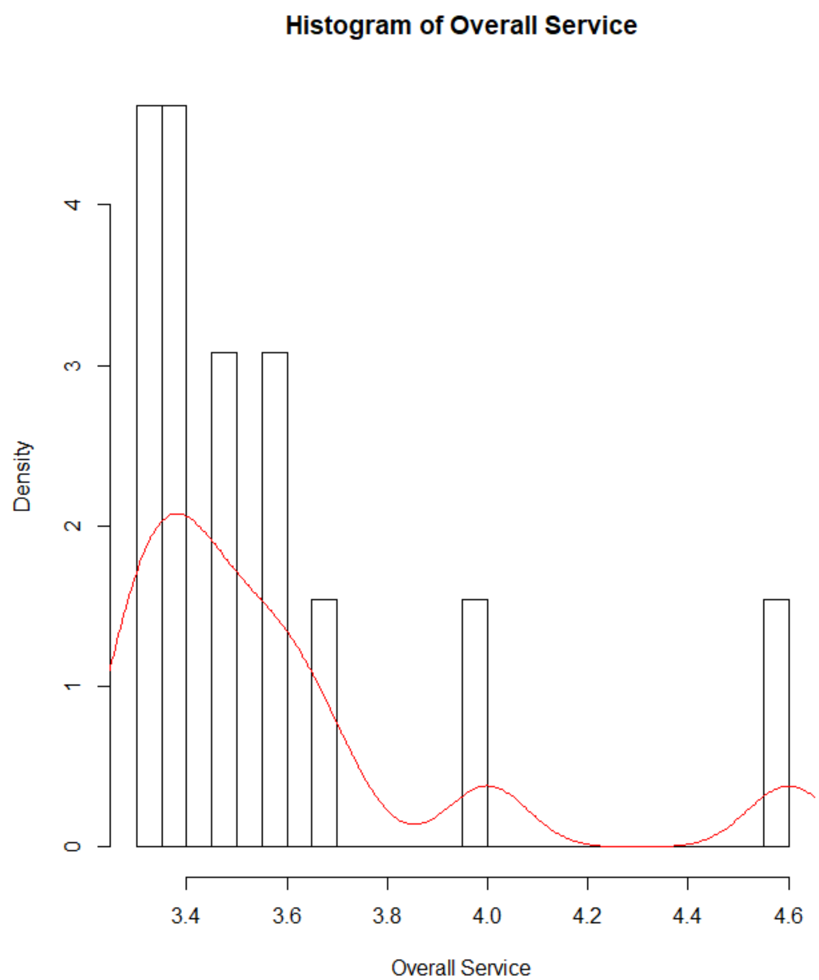
Graph 269: Histogram of Satisfaction Level ratings in Cluster 4, with density plot shown as the red line.

- **Salesforce Image:** With a standard deviation of 0.94, Salesforce image varies from 1.9 to 4.6 and has a mean value of 3.18. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 4 tend to have rated Salesforce image low (ratings ranging from 1.9 to 4.6 out of 10.0, with a mean value of 4.6). Below is the histogram and density plot of Salesforce Image in cluster 4.



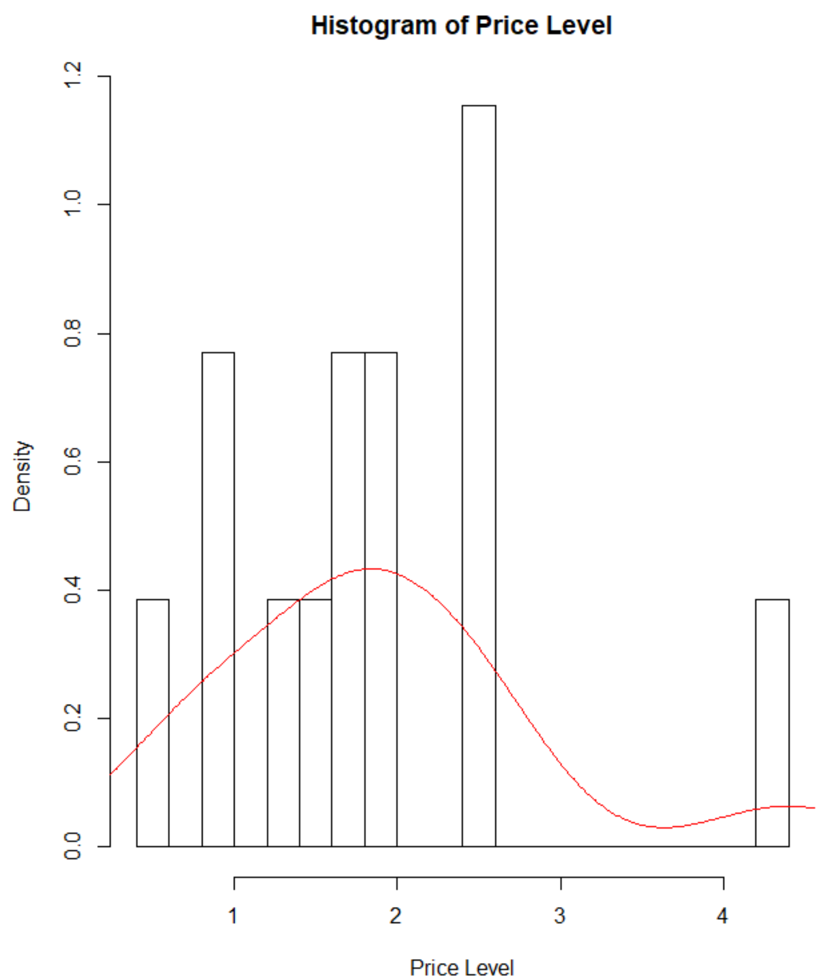
Graph 270: Histogram of Salesforce Image ratings in Cluster 4, with density plot shown as the red line.

- **Overall Service:** With a standard deviation of 0.36, the distribution has a moderately narrow spread about its mean. Overall Service varies from 3.3 to 4.6 and has a mean value of 3.58. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 4 tend to have rated Overall Service moderately low (ratings ranging from 3.3 to 4.6 out of 10.0, with a mean value of 3.58). Below is the histogram and density plot of Overall Service in cluster 4.



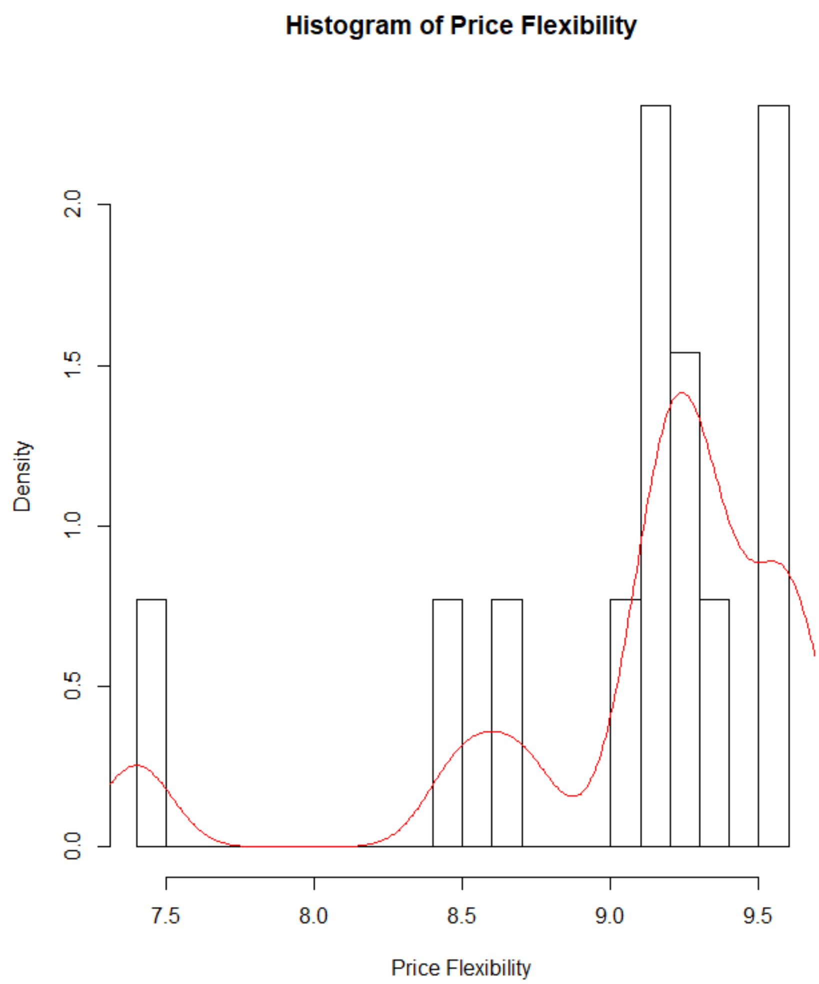
Graph 271: Histogram of Overall Service ratings in Cluster 4, with density plot shown as the red line.

- **Price Level:** With a standard deviation of 0.99, Price Level varies from 0.5 to 4.4 and has a mean value of 1.88. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 4 tend to have rated Price Level low (ratings ranging from 0.5 to 4.4 out of 10.0, with a mean of 1.88 and a standard deviation of 0.99). Price level depicts the perceived level of price charged by PLE. Below is the histogram and density plot of Price Level in cluster 4.



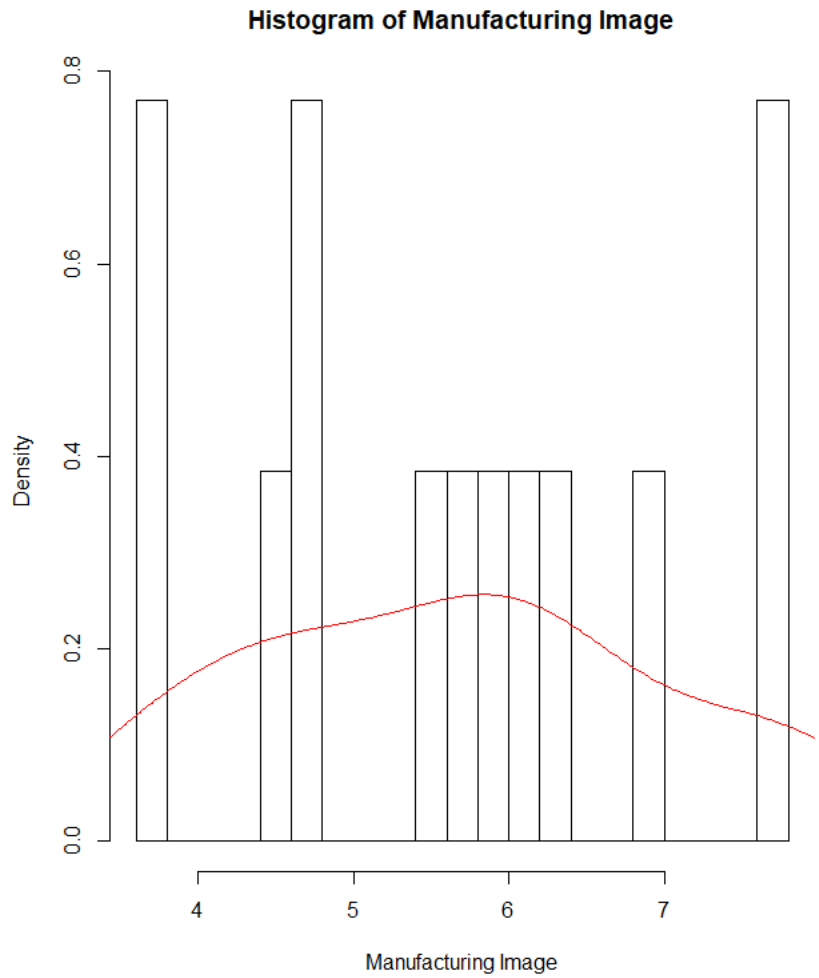
Graph 272: Histogram of Price Level ratings in Cluster 4, with density plot shown as the red line.

- **Price Flexibility:** With a standard deviation of 0.60, Price Flexibility varies from 7.4 to 9.6 and has a mean value of 9.08. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 4 tend to have rated Price Flexibility high (ratings ranging from 7.4 to 9.6 out of 10.0, with a mean of 9.08). Price flexibility depicts the perceived willingness of PLE representatives to negotiate prices on all types of purchases. Below is the histogram and density plot of Price Flexibility in cluster 4.



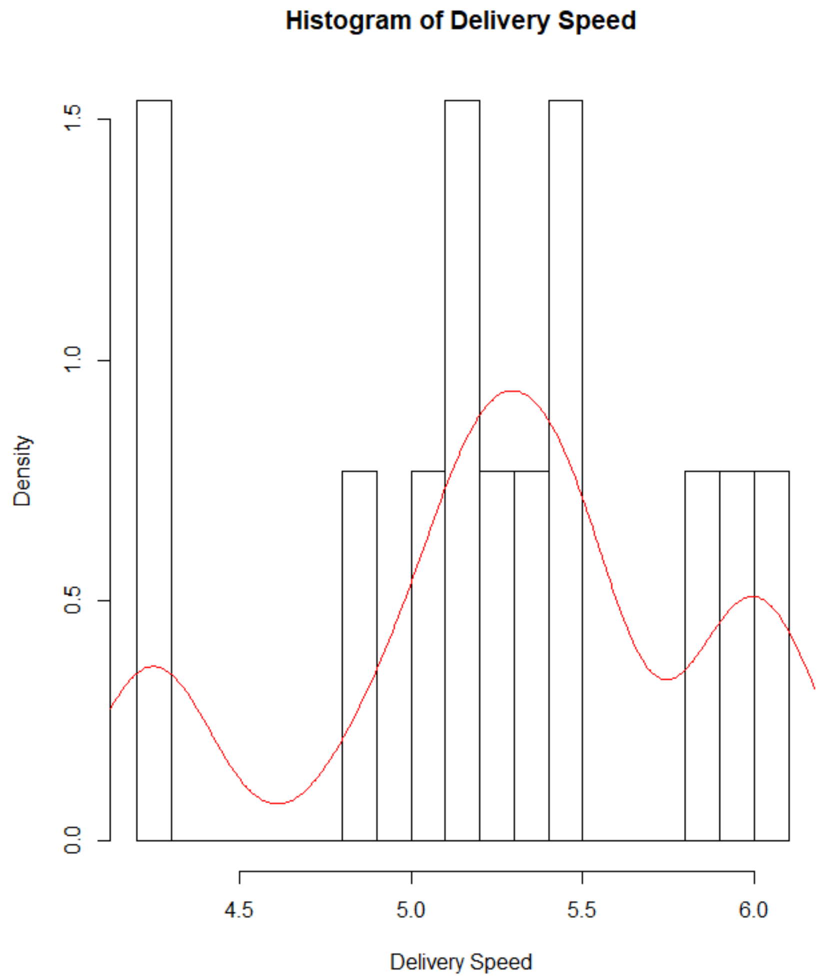
Graph 273: Histogram of Price Flexibility ratings in Cluster 4, with density plot shown as the red line.

- **Manufacturing Image:** With a standard deviation of 1.35, Manufacturing Image varies from 3.7 to 7.8 and has a mean value of 5.67. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 4 tend to have rated Manufacturing Image moderately low (ratings ranging from 3.7 to 7.8 out of 10.0, with a mean of 5.67). Manufacturing Image depicts the overall image of the manufacturer. Below is the histogram and density plot of Manufacturing Image in cluster 4.



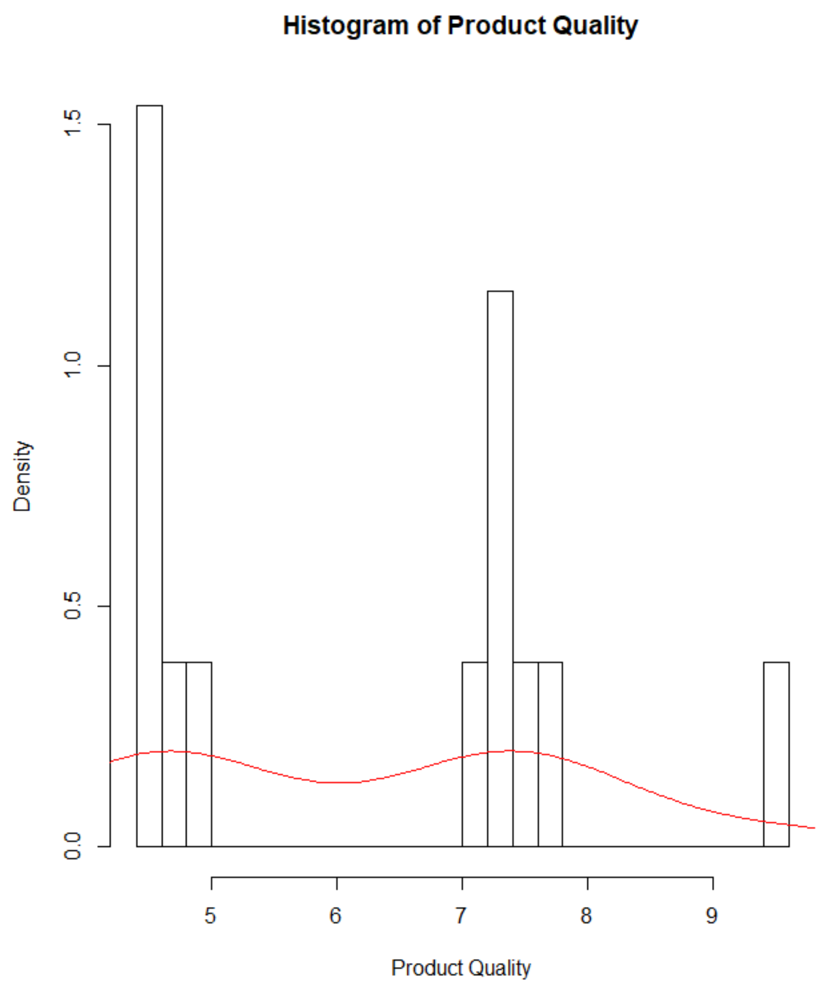
Graph 274: Histogram of Manufacturing Image ratings in Cluster 4, with density plot shown as the red line.

- **Delivery Speed:** With a standard deviation of 0.58, Delivery Speed varies from 4.2 to 6.1 and has a mean value of 5.28. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 4 tend to have rated Delivery Speed moderately high (ratings ranging from 4.2 to 6.1 out of 10.0, with a mean of 5.28). Delivery Speed depicts the amount of time it takes to deliver product once an order is confirmed. Below is the histogram and density plot of Delivery Speed in cluster 4.



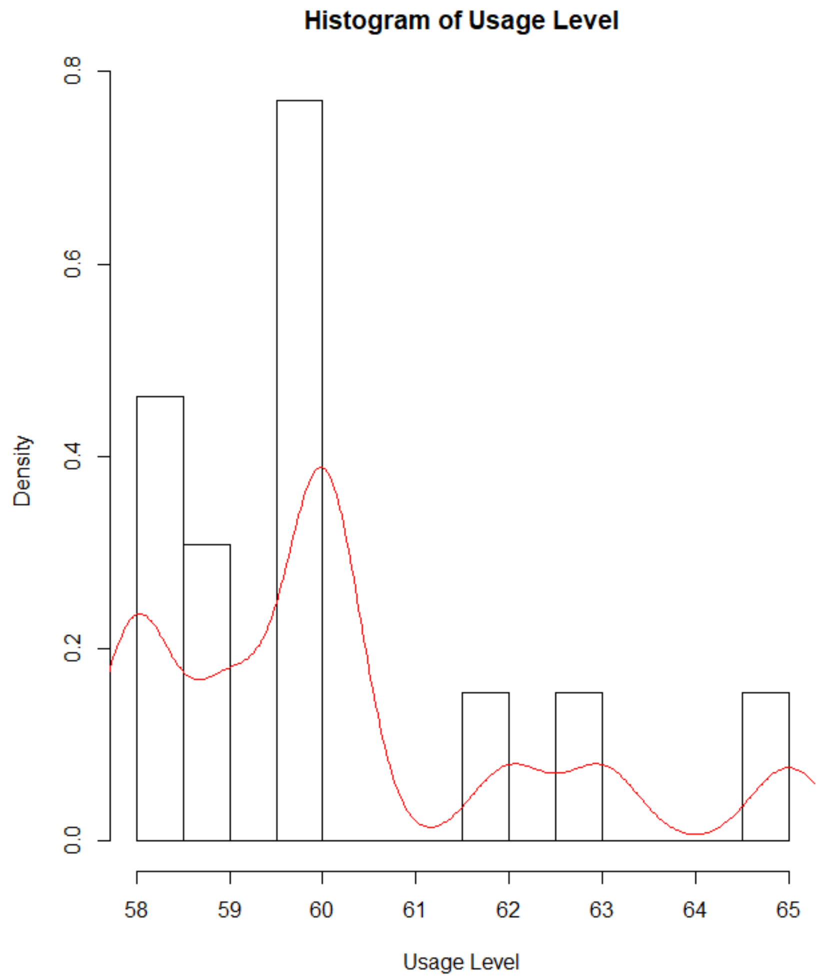
Graph 275: Histogram of Delivery Speed ratings in Cluster 4, with density plot shown as the red line.

- **Product Quality:** With a standard deviation of 1.71, Product Quality varies from 4.5 to 9.6 and has a mean value of 6.29. Rating was done on a 10cm scale, rounded to 1 decimal place. Customers in cluster 4 tend to have rated Product Quality moderately high (ratings ranging from 4.5 to 9.6 out of 10.0, with a mean of 6.29). Product Quality depicts the perceived level of quality of PLE products. Below is the histogram and density plot of Product Quality in cluster 4.



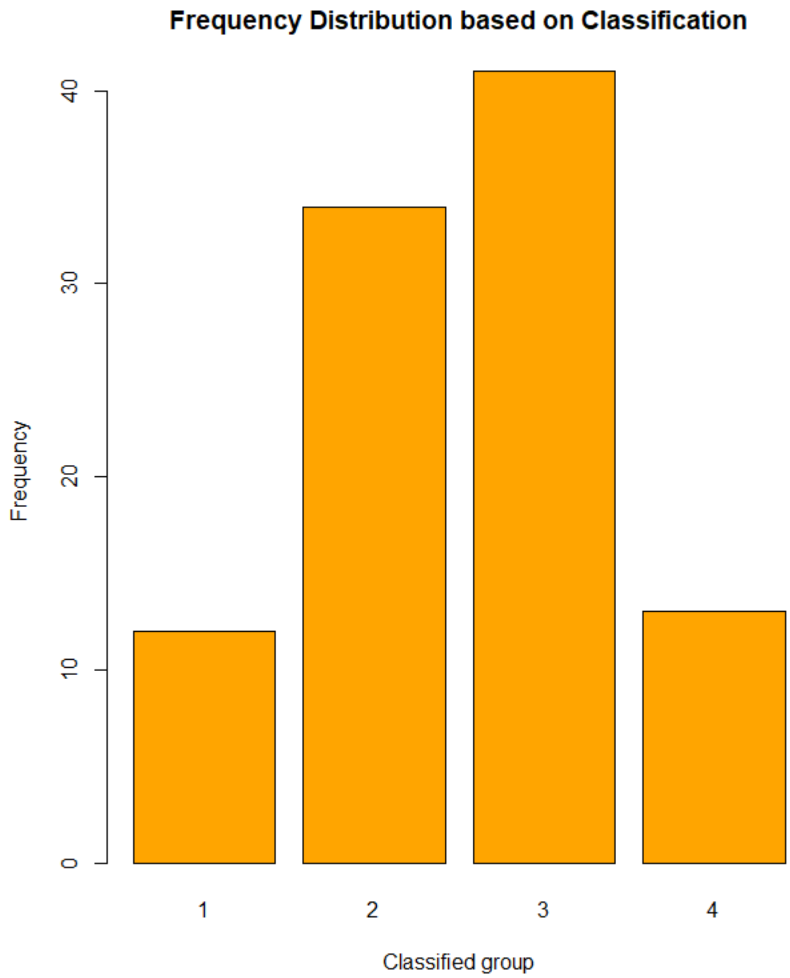
Graph 276: Histogram of Product Quality ratings in Cluster 4, with density plot shown as the red line.

- **Usage Level:** With a standard deviation of 2.08, Usage Level varies from 58.0 to 65.0 and has a mean value of 60.15. Usage Level depicts how much of the firm's total product is purchased from PLE, measured on a 100-point scale, ranging from 0% to 100%. Customers in cluster 4 tend to rate the usage level moderately high (ratings ranging from 58.0% to 65.0% out of 100.0%, with a mean of 60.15%). Below is the histogram and density plot of Usage Level in cluster 4.



Graph 277: Histogram of Usage Level ratings in Cluster 4, with density plot shown as the red line.

It is the combined ratings of these variables that have resulted in customers being classified into the same groups. Below shows the frequency distribution of the 100 customers being classified into the above 4 groups (using the data).



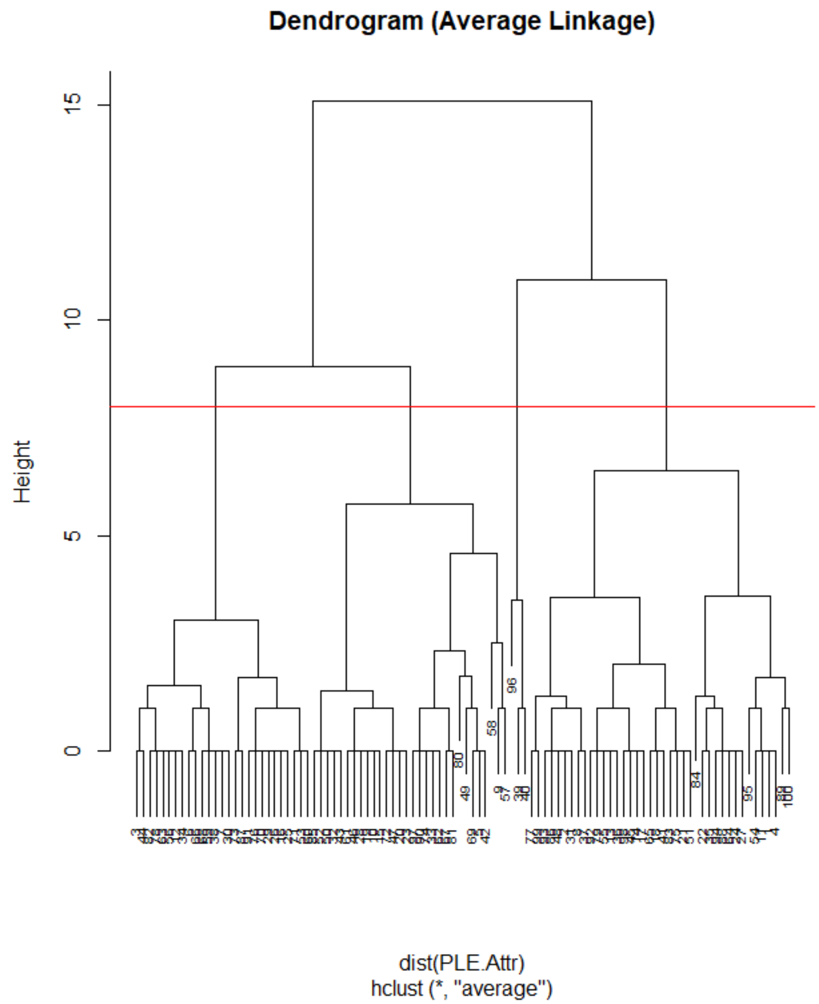
Graph 278: Frequency distribution of number of customers in each of the 4 classified groups.

By classifying customers with similar attributes together, the company can identify customers with similar profiles, and in this case, work on the attributes of the company that were poorly rated by the majority of the customers. Most customers are grouped under clusters 2 and 3 as seen from the frequency distribution above in plot 41. Both groups of customers from clusters 2 and 3 have relatively low ratings for Salesforce Image, Overall Service, Price Level, Manufacturing Image and Delivery Service of PLE. Hence, since these two groups of customers form the bulk of PLE's customer base, PLE should then focus on improving these attributes of the company to increase total revenue.

To further generate insights of the drivers of Usage levels and Satisfaction levels among the customers, we can make use of hierarchical clustering to cluster customers with similar usage levels/similar satisfaction levels into the same group.

We perform a similar method of hierarchical clustering using Average Linkage, but this time only clustering based on Usage Levels.

Using the Average Linkage method, we can cut the dendrogram such that we would obtain 4 clusters of customers. In this case, a horizontal line at y=8 will cut the dendrogram into 4 clusters as there are 4 intersections between the dendrogram and the horizontal red line (y=8) as seen below. The dendrogram is divided into 4 clusters here because as shown in the following dendrogram, it seems that 4 groups can separate our data into clusters relatively nicely.



Graph 279: Dendrogram formed using average linkage clustering cut into 4 clusters by a horizontal line at y = 8.

Similarly, we can examine the descriptive statistics of the customers under each cluster, clustered based on similar usage levels.

Below are the descriptive statistics of clusters 1, 2, 3 and 4 respectively.

Cluster 1

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Delivery speed	1	40	2.79	0.93	2.80	2.77	0.96	1.0	4.9	3.9	0.19	-0.45	0.15
Price level	2	40	2.24	1.18	2.00	2.21	1.48	0.4	4.2	3.8	0.16	-1.31	0.19
Price flexibility	3	40	7.03	1.09	7.05	7.04	1.04	5.0	9.0	4.0	-0.04	-0.95	0.17
Manufacturing image	4	40	5.01	1.07	4.90	5.00	0.96	2.5	7.8	5.3	0.15	0.21	0.17
Overall service	5	40	2.47	0.64	2.40	2.45	0.59	1.2	3.7	2.5	0.31	-0.71	0.10
Salesforce image	6	40	2.51	0.67	2.50	2.48	0.44	1.1	4.0	2.9	0.27	0.09	0.11
Product quality	7	40	7.22	1.42	7.20	7.18	1.78	5.0	10.0	5.0	0.15	-1.15	0.22
Usage Level	8	40	38.27	3.76	39.00	38.44	4.45	31.0	44.0	13.0	-0.29	-1.03	0.59
Satisfaction Level	9	40	4.17	0.61	4.15	4.13	0.67	3.2	5.6	2.4	0.50	-0.42	0.10
Size of firm	10	40	0.57	0.50	1.00	0.59	0.00	0.0	1.0	1.0	-0.29	-1.96	0.08
Purchasing Structure	11	40	0.70	0.46	1.00	0.75	0.00	0.0	1.0	1.0	-0.84	-1.33	0.07
Industry	12	40	0.42	0.50	0.00	0.41	0.00	0.0	1.0	1.0	0.29	-1.96	0.08
Buying Type	13	40	1.27	0.45	1.00	1.22	0.00	1.0	2.0	1.0	0.97	-1.08	0.07
Classification	14	40	1.77	0.42	2.00	1.84	0.00	1.0	2.0	1.0	-1.27	-0.40	0.07
Classification based on Usage Level	15	40	1.00	0.00	1.00	1.00	0.00	1.0	1.0	0.0	NaN	NaN	0.00

Cluster 2

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Delivery speed	1	27	3.56	0.91	3.6	3.59	0.89	1.6	5.1	3.5	-0.36	-0.64	0.17
Price level	2	27	2.87	1.27	2.6	2.83	1.63	0.9	5.4	4.5	0.35	-0.98	0.24
Price flexibility	3	27	8.06	1.41	8.3	8.11	1.48	5.5	9.9	4.4	-0.43	-1.21	0.27
Manufacturing image	4	27	5.41	1.05	5.4	5.39	0.89	3.3	8.2	4.9	0.44	0.42	0.20
Overall service	5	27	3.21	0.57	3.1	3.17	0.44	2.4	4.5	2.1	0.79	-0.34	0.11
Salesforce image	6	27	2.71	0.75	2.6	2.66	0.44	1.5	4.5	3.0	0.68	0.00	0.14
Product quality	7	27	6.97	1.75	7.4	7.04	1.63	3.7	9.7	6.0	-0.45	-1.18	0.34
Usage Level	8	27	47.89	1.76	48.0	47.87	1.48	45.0	51.0	6.0	0.08	-1.21	0.34
Satisfaction Level	9	27	5.04	0.50	5.1	5.06	0.44	3.8	5.9	2.1	-0.44	-0.26	0.10
Size of firm	10	27	0.37	0.49	0.0	0.35	0.00	0.0	1.0	1.0	0.51	-1.81	0.09
Purchasing Structure	11	27	0.48	0.51	0.0	0.48	0.00	0.0	1.0	1.0	0.07	-2.07	0.10
Industry	12	27	0.63	0.49	1.0	0.65	0.00	0.0	1.0	1.0	-0.51	-1.81	0.09
Buying Type	13	27	2.22	0.58	2.0	2.26	0.00	1.0	3.0	2.0	-0.01	-0.52	0.11
Classification	14	27	2.89	0.32	3.0	2.96	0.00	2.0	3.0	1.0	-2.34	3.61	0.06
Classification based on Usage Level	15	27	2.00	0.00	2.0	2.00	0.00	2.0	2.0	0.0	NaN	NaN	0.00

Cluster 3

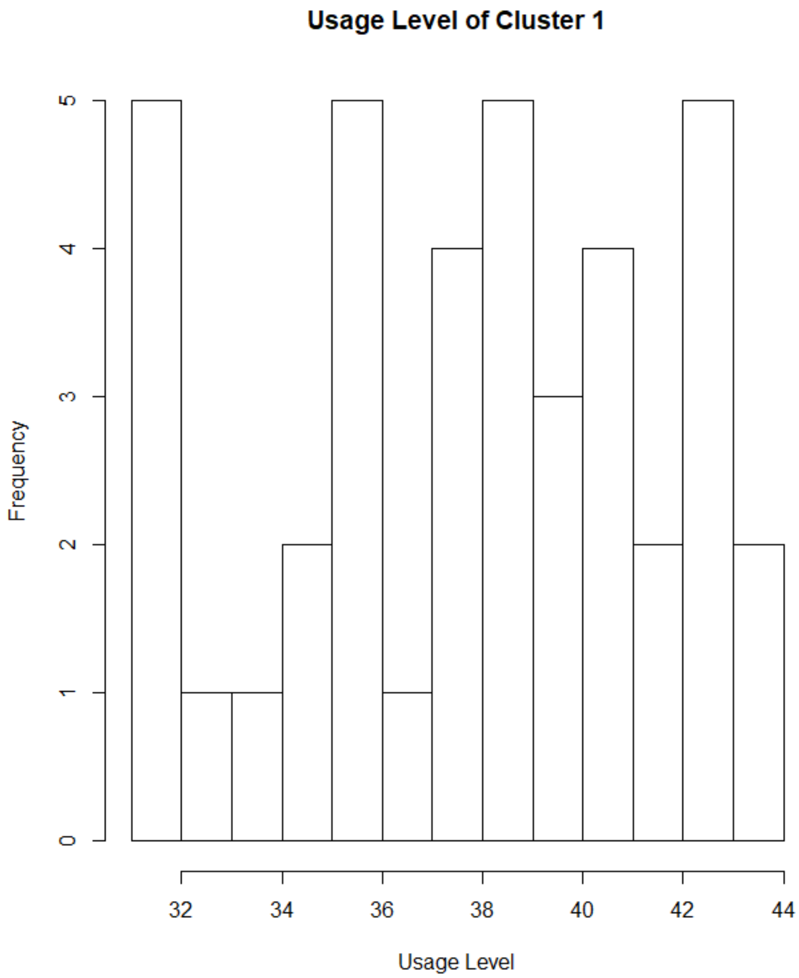
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Delivery speed	1	30	4.70	1.06	5.05	4.79	0.74	2.6	6.1	3.5	-0.77	-0.72	0.19
Price level	2	30	2.12	1.11	1.90	2.03	0.89	0.2	4.8	4.6	0.69	-0.10	0.20
Price flexibility	3	30	9.03	0.79	9.20	9.13	0.67	6.5	10.0	3.5	-1.28	1.61	0.14
Manufacturing image	4	30	5.45	1.29	5.65	5.46	1.56	3.0	7.8	4.8	-0.03	-0.95	0.24
Overall service	5	30	3.41	0.42	3.40	3.41	0.30	2.6	4.6	2.0	0.30	0.68	0.08
Salesforce image	6	30	2.86	0.91	2.80	2.86	0.82	1.3	4.6	3.3	0.12	-0.84	0.17
Product quality	7	30	6.47	1.57	6.80	6.45	1.70	3.8	9.6	5.8	-0.03	-0.99	0.29
Usage Level	8	30	56.80	3.38	55.50	56.46	3.71	53.0	65.0	12.0	0.60	-0.76	0.62
Satisfaction Level	9	30	5.46	0.75	5.50	5.50	0.74	3.8	6.8	3.0	-0.42	-0.49	0.14
Size of firm	10	30	0.13	0.35	0.00	0.04	0.00	0.0	1.0	1.0	2.05	2.28	0.06
Purchasing Structure	11	30	0.20	0.41	0.00	0.12	0.00	0.0	1.0	1.0	1.43	0.04	0.07
Industry	12	30	0.43	0.50	0.00	0.42	0.00	0.0	1.0	1.0	0.26	-2.00	0.09
Buying Type	13	30	2.87	0.35	3.00	2.96	0.00	2.0	3.0	1.0	-2.05	2.28	0.06
Classification	14	30	3.43	0.50	3.00	3.42	0.00	3.0	4.0	1.0	0.26	-2.00	0.09
Classification based on Usage Level	15	30	3.00	0.00	3.00	3.00	0.00	3.0	3.0	0.0	NaN	NaN	0.00

<u>Cluster 4</u>

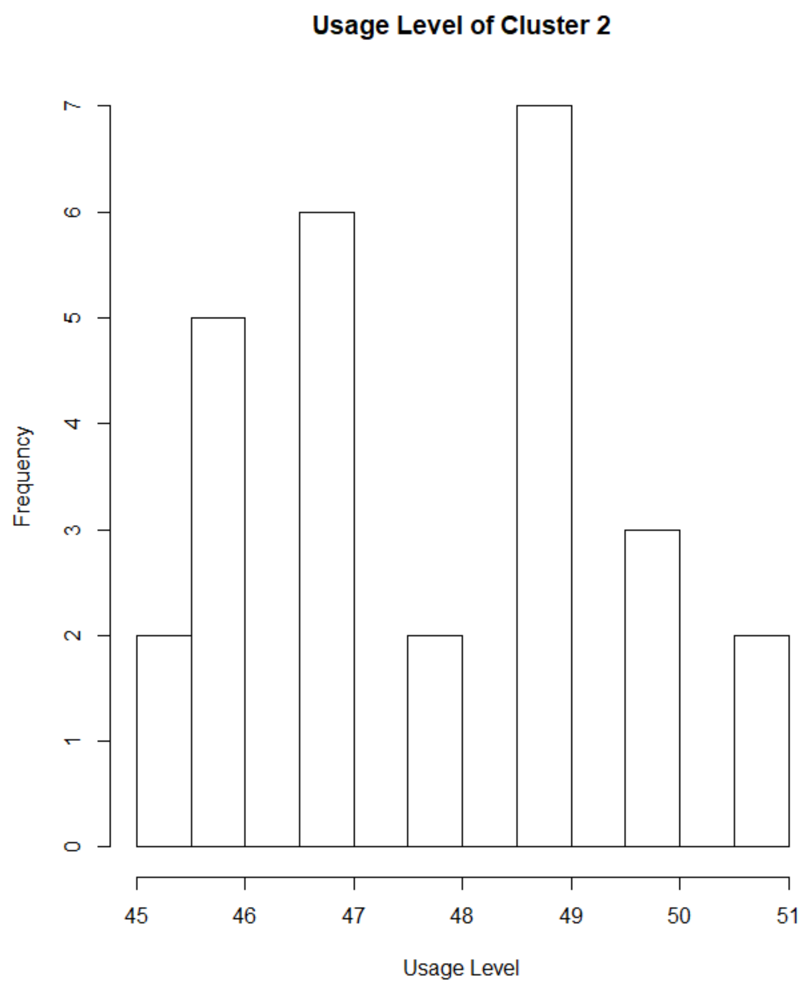
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Delivery speed	1	3	1.00	1.25	0.6	1.00	0.89	0.0	2.4	2.4	0.29	-2.33	0.72
Price level	2	3	1.90	0.26	2.0	1.90	0.15	1.6	2.1	0.5	-0.32	-2.33	0.15
Price flexibility	3	3	6.57	0.29	6.4	6.57	0.00	6.4	6.9	0.5	0.38	-2.33	0.17
Manufacturing image	4	3	4.97	0.45	5.0	4.97	0.59	4.5	5.4	0.9	-0.07	-2.33	0.26
Overall service	5	3	1.30	0.72	1.1	1.30	0.59	0.7	2.1	1.4	0.26	-2.33	0.42
Salesforce image	6	3	2.30	0.26	2.2	2.30	0.15	2.1	2.6	0.5	0.32	-2.33	0.15
Product quality	7	3	8.70	0.26	8.8	8.70	0.15	8.4	8.9	0.5	-0.32	-2.33	0.15
Usage Level	8	3	27.33	2.08	28.0	27.33	1.48	25.0	29.0	4.0	-0.29	-2.33	1.20
Satisfaction Level	9	3	3.53	0.32	3.4	3.53	0.15	3.3	3.9	0.6	0.34	-2.33	0.19
Size of firm	10	3	1.00	0.00	1.0	1.00	0.00	1.0	1.0	0.0	NaN	NaN	0.00
Purchasing Structure	11	3	1.00	0.00	1.0	1.00	0.00	1.0	1.0	0.0	NaN	NaN	0.00
Industry	12	3	1.00	0.00	1.0	1.00	0.00	1.0	1.0	0.0	NaN	NaN	0.00
Buying Type	13	3	1.00	0.00	1.0	1.00	0.00	1.0	1.0	0.0	NaN	NaN	0.00
Classification	14	3	1.00	0.00	1.0	1.00	0.00	1.0	1.0	0.0	NaN	NaN	0.00
Classification based on Usage Level	15	3	4.00	0.00	4.0	4.00	0.00	4.0	4.0	0.0	NaN	NaN	0.00

The descriptive statistics shown above can help us analyse what are some of the similar ratings the customers in each cluster have given to the various attributes of PLE, and hence resulting in differing usage levels.

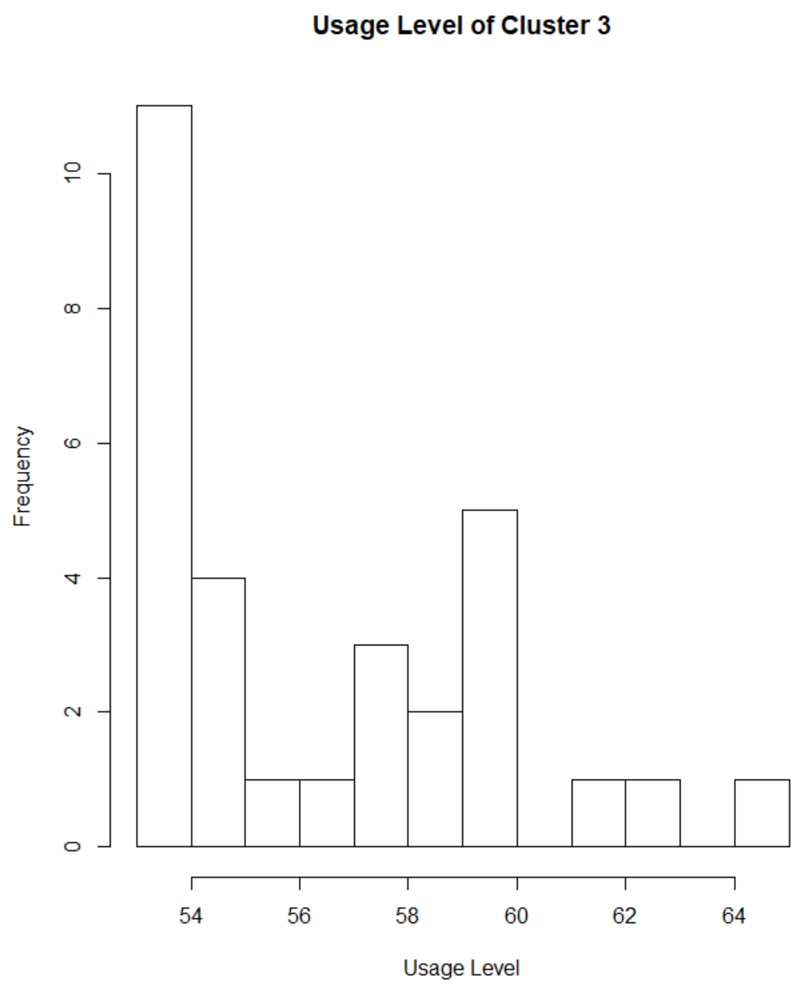
Below show histogram plots of the usage levels of cluster 1,2,3 and 4 respectively.



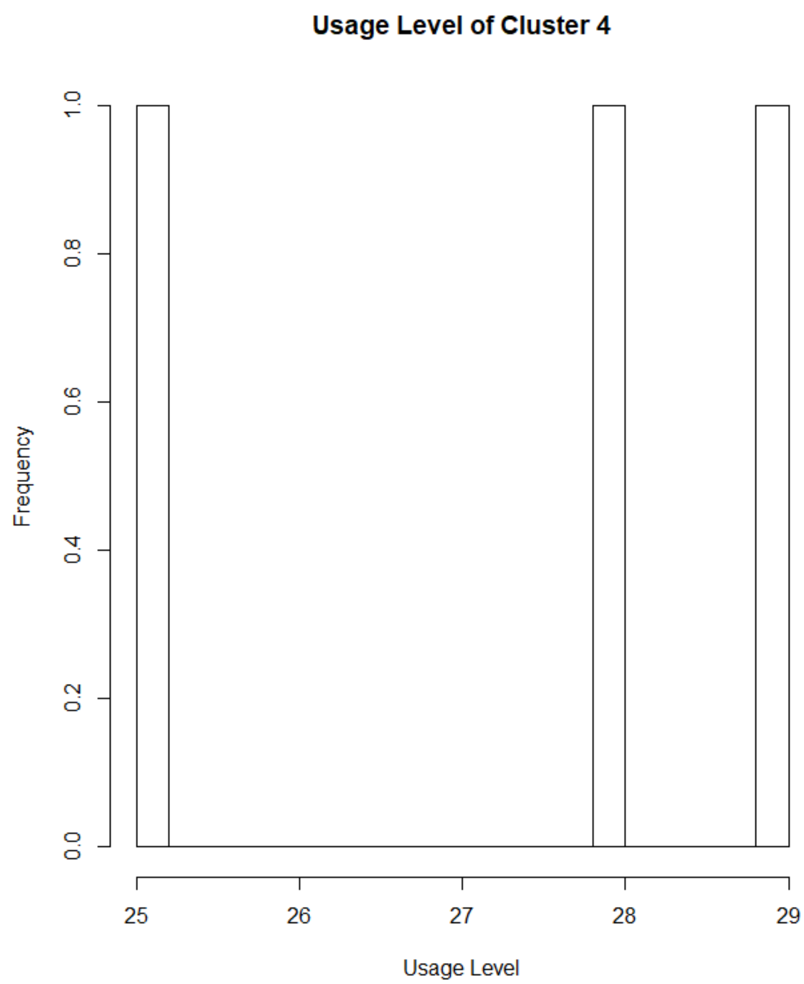
Graph 280: Histogram of Usage Level of Cluster 1



Graph 281: Histogram of Usage Level of Cluster 2



Graph 282: Histogram of Usage Level of Cluster 3



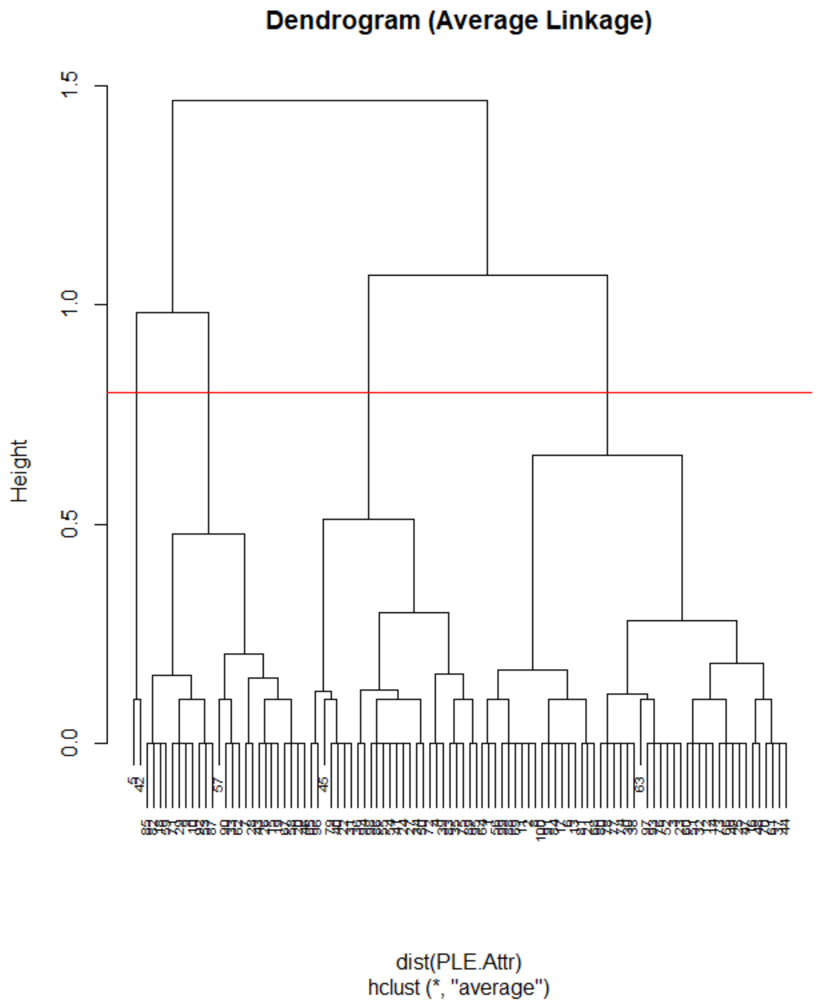
Graph 283: Histogram of Usage Level of Cluster 4

Since Cluster 3 has the highest usage levels among the other clusters, the company can look at what could have driven the higher usage levels of customers in cluster 3, by looking at the descriptive statistics of customers in cluster 3 (as shown above) and analysing the combination of ratings of the different PLE attributes by customers in cluster 3 that could possibly drive higher usage levels.

Lastly, we will use hierarchical clustering again to look at which combination of ratings of the different PLE attributes would drive higher Satisfaction levels.

We perform a similar method of hierarchical clustering using Average Linkage, but this time only clustering based on Satisfaction Levels.

Using the Average Linkage method, we can cut the dendrogram such that we would obtain 4 clusters of customers. In this case, a horizontal line at y=0.8 will cut the dendrogram into 4 clusters as there are 4 intersections between the dendrogram and the horizontal red line (y=0.8) as seen below. The dendrogram is divided into 4 clusters here because as shown in the following dendrogram, it seems that 4 groups can separate our data into clusters relatively nicely.



Graph 284: Dendrogram formed using average linkage clustering cut into 4 clusters by a horizontal line at y = 0.8.

Similarly, we can examine the descriptive statistics of the customers under each cluster, clustered based on similar satisfaction levels.

Below are the descriptive statistics of clusters 1, 2, 3 and 4 respectively.

Cluster 1

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Delivery speed	1	47	3.45	1.11	3.3	3.40	1.04	1.3	6.1	4.8	0.44	-0.26	0.16
Price level	2	47	2.49	1.34	2.2	2.47	1.63	0.2	5.2	5.0	0.21	-1.25	0.20
Price flexibility	3	47	8.04	1.37	8.5	8.11	1.04	5.1	10.0	4.9	-0.63	-0.83	0.20
Manufacturing image	4	47	5.20	0.98	4.9	5.15	0.59	3.3	7.8	4.5	0.69	0.03	0.14
Overall service	5	47	2.96	0.53	3.0	2.96	0.59	2.0	4.3	2.3	0.13	-0.60	0.08
Salesforce image	6	47	2.61	0.60	2.5	2.58	0.44	1.4	4.0	2.6	0.43	-0.05	0.09
Product quality	7	47	6.81	1.63	6.9	6.87	2.08	3.7	9.7	6.0	-0.24	-0.96	0.24
Usage Level	8	47	46.43	7.34	47.0	46.46	7.41	32.0	60.0	28.0	-0.04	-0.67	1.07
Satisfaction Level	9	47	4.76	0.35	4.8	4.76	0.44	4.2	5.3	1.1	-0.18	-1.52	0.05
Size of firm	10	47	0.28	0.45	0.0	0.23	0.00	0.0	1.0	1.0	0.97	-1.09	0.07
Purchasing Structure	11	47	0.47	0.50	0.0	0.46	0.00	0.0	1.0	1.0	0.12	-2.03	0.07
Industry	12	47	0.51	0.51	1.0	0.51	0.00	0.0	1.0	1.0	-0.04	-2.04	0.07
Buying Type	13	47	2.13	0.71	2.0	2.15	1.48	1.0	3.0	2.0	-0.18	-1.06	0.10
Classification	14	47	2.60	0.74	3.0	2.59	1.48	1.0	4.0	3.0	-0.17	-0.35	0.11
Classification based on Usage Level	15	47	1.85	0.78	2.0	1.82	1.48	1.0	3.0	2.0	0.25	-1.35	0.11
Classification based on Satisfaction Level	16	47	1.00	0.00	1.0	1.00	0.00	1.0	1.0	0.0	NaN	NaN	0.00

Cluster 2

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Delivery speed	1	26	2.54	1.28	2.40	2.47	0.89	0.0	5.6	5.6	0.49	0.30	0.25
Price level	2	26	2.07	0.92	2.05	2.08	1.04	0.4	3.7	3.3	-0.09	-1.04	0.18
Price flexibility	3	26	6.92	0.87	6.80	6.94	0.59	5.0	8.7	3.7	-0.08	-0.21	0.17
Manufacturing image	4	26	4.65	1.07	4.75	4.66	0.82	2.5	6.6	4.1	-0.21	-0.66	0.21
Overall service	5	26	2.22	0.79	2.20	2.19	0.59	0.7	4.0	3.3	0.43	-0.14	0.15
Salesforce image	6	26	2.40	0.61	2.50	2.40	0.52	1.1	3.8	2.7	0.05	-0.19	0.12
Product quality	7	26	7.70	1.45	8.10	7.73	1.11	5.0	10.0	5.0	-0.55	-0.80	0.28
Usage Level	8	26	37.88	6.90	38.00	37.45	4.45	25.0	55.0	30.0	0.57	0.30	1.35
Satisfaction Level	9	26	3.69	0.27	3.70	3.69	0.30	3.2	4.1	0.9	-0.18	-1.15	0.05
Size of firm	10	26	0.81	0.40	1.00	0.86	0.00	0.0	1.0	1.0	-1.47	0.18	0.08
Purchasing Structure	11	26	0.81	0.40	1.00	0.86	0.00	0.0	1.0	1.0	-1.47	0.18	0.08
Industry	12	26	0.58	0.50	1.00	0.59	0.00	0.0	1.0	1.0	-0.29	-1.99	0.10
Buying Type	13	26	1.15	0.54	1.00	1.00	0.00	1.0	3.0	2.0	2.99	7.25	0.11
Classification	14	26	1.77	0.65	2.00	1.73	0.00	1.0	3.0	2.0	0.23	-0.85	0.13
Classification based on Usage Level	15	26	1.54	1.07	1.00	1.36	0.00	1.0	4.0	3.0	1.52	0.59	0.21
Classification based on Satisfaction Level	16	26	2.00	0.00	2.00	2.00	0.00	2.0	2.0	0.0	NaN	NaN	0.00

Cluster 3

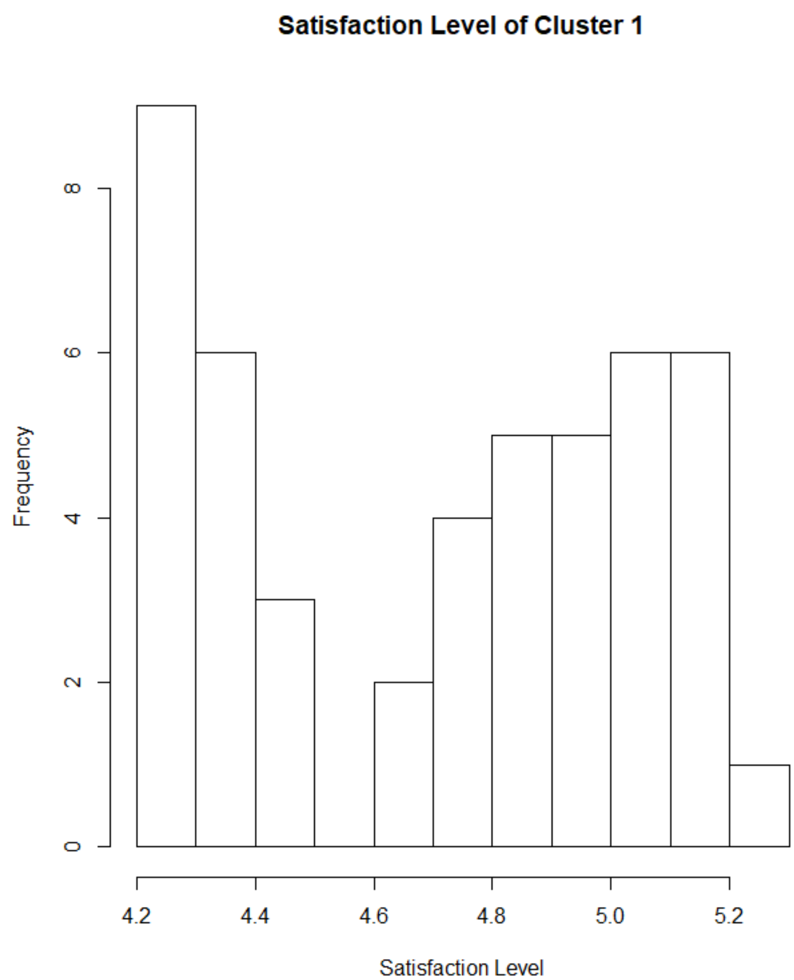
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Delivery speed	1	2	5.95	0.07	5.95	5.95	0.07	5.9	6.0	0.1	0	-2.75	0.05
Price level	2	2	0.90	0.00	0.90	0.90	0.00	0.9	0.9	0.0	NaN	NaN	0.00
Price flexibility	3	2	9.60	0.00	9.60	9.60	0.00	9.6	9.6	0.0	NaN	NaN	0.00
Manufacturing image	4	2	7.80	0.00	7.80	7.80	0.00	7.8	7.8	0.0	NaN	NaN	0.00
Overall service	5	2	3.40	0.00	3.40	3.40	0.00	3.4	3.4	0.0	NaN	NaN	0.00
Salesforce image	6	2	4.60	0.00	4.60	4.60	0.00	4.6	4.6	0.0	NaN	NaN	0.00
Product quality	7	2	4.50	0.00	4.50	4.50	0.00	4.5	4.5	0.0	NaN	NaN	0.00
Usage Level	8	2	58.00	0.00	58.00	58.00	0.00	58.0	58.0	0.0	NaN	NaN	0.00
Satisfaction Level	9	2	6.75	0.07	6.75	6.75	0.07	6.7	6.8	0.1	0	-2.75	0.05
Size of firm	10	2	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	NaN	NaN	0.00
Purchasing Structure	11	2	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	NaN	NaN	0.00
Industry	12	2	1.00	0.00	1.00	1.00	0.00	1.0	1.0	0.0	NaN	NaN	0.00
Buying Type	13	2	3.00	0.00	3.00	3.00	0.00	3.0	3.0	0.0	NaN	NaN	0.00
Classification	14	2	4.00	0.00	4.00	4.00	0.00	4.0	4.0	0.0	NaN	NaN	0.00
Classification based on Usage Level	15	2	3.00	0.00	3.00	3.00	0.00	3.0	3.0	0.0	NaN	NaN	0.00
Classification based on Satisfaction Level	16	2	3.00	0.00	3.00	3.00	0.00	3.0	3.0	0.0	NaN	NaN	0.00

Cluster 4

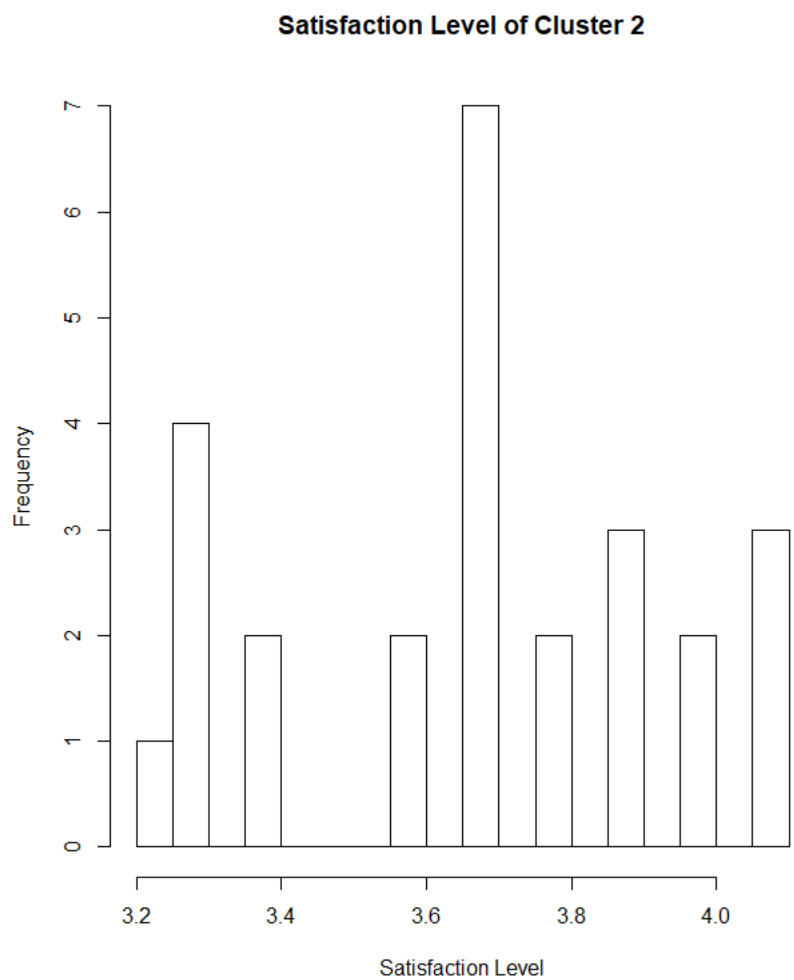
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Delivery speed	1	25	4.46	0.82	4.7	4.56	0.59	2.4	5.5	3.1	-1.02	0.22	0.16
Price level	2	25	2.54	1.10	2.4	2.43	0.89	1.3	5.4	4.1	0.99	0.08	0.22
Price flexibility	3	25	8.50	1.35	9.2	8.64	1.04	5.5	9.9	4.4	-0.80	-0.72	0.27
Manufacturing image	4	25	5.76	1.02	5.9	5.77	0.59	3.5	8.2	4.7	-0.17	0.07	0.20
Overall service	5	25	3.51	0.49	3.4	3.48	0.44	2.7	4.6	1.9	0.59	-0.29	0.10
Salesforce image	6	25	2.88	0.96	2.8	2.88	0.89	1.3	4.5	3.2	-0.07	-1.09	0.19
Product quality	7	25	6.71	1.38	6.8	6.69	1.19	4.4	9.6	5.2	0.06	-0.68	0.28
Usage Level	8	25	53.08	6.68	54.0	53.33	7.41	39.0	65.0	26.0	-0.21	-0.58	1.34
Satisfaction Level	9	25	5.77	0.26	5.8	5.77	0.30	5.4	6.2	0.8	-0.06	-1.55	0.05
Size of firm	10	25	0.24	0.44	0.0	0.19	0.00	0.0	1.0	1.0	1.15	-0.71	0.09
Purchasing Structure	11	25	0.28	0.46	0.0	0.24	0.00	0.0	1.0	1.0	0.92	-1.19	0.09
Industry	12	25	0.36	0.49	0.0	0.33	0.00	0.0	1.0	1.0	0.55	-1.76	0.10
Buying Type	13	25	2.56	0.58	3.0	2.62	0.00	1.0	3.0	2.0	-0.83	-0.45	0.12
Classification	14	25	3.16	0.62	3.0	3.19	0.00	2.0	4.0	2.0	-0.10	-0.65	0.12
Classification based on Usage Level	15	25	2.52	0.65	3.0	2.62	0.00	1.0	3.0	2.0	-0.93	-0.35	0.13
Classification based on Satisfaction Level	16	25	4.00	0.00	4.0	4.00	0.00	4.0	4.0	0.0	NaN	NaN	0.00

The descriptive statistics shown above can help us analyse what are some of the similar ratings the customers in each cluster have given to the various attributes of PLE, and hence resulting in differing satisfaction levels.

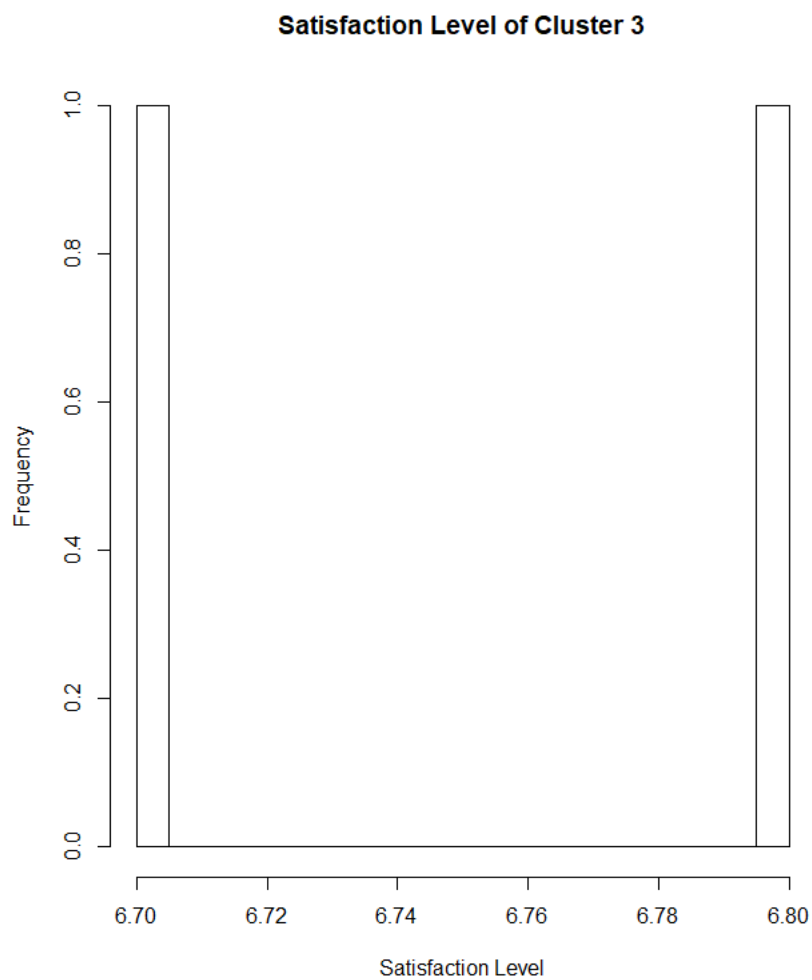
Below show histogram plots of the satisfaction levels of cluster 1,2,3 and 4 respectively.



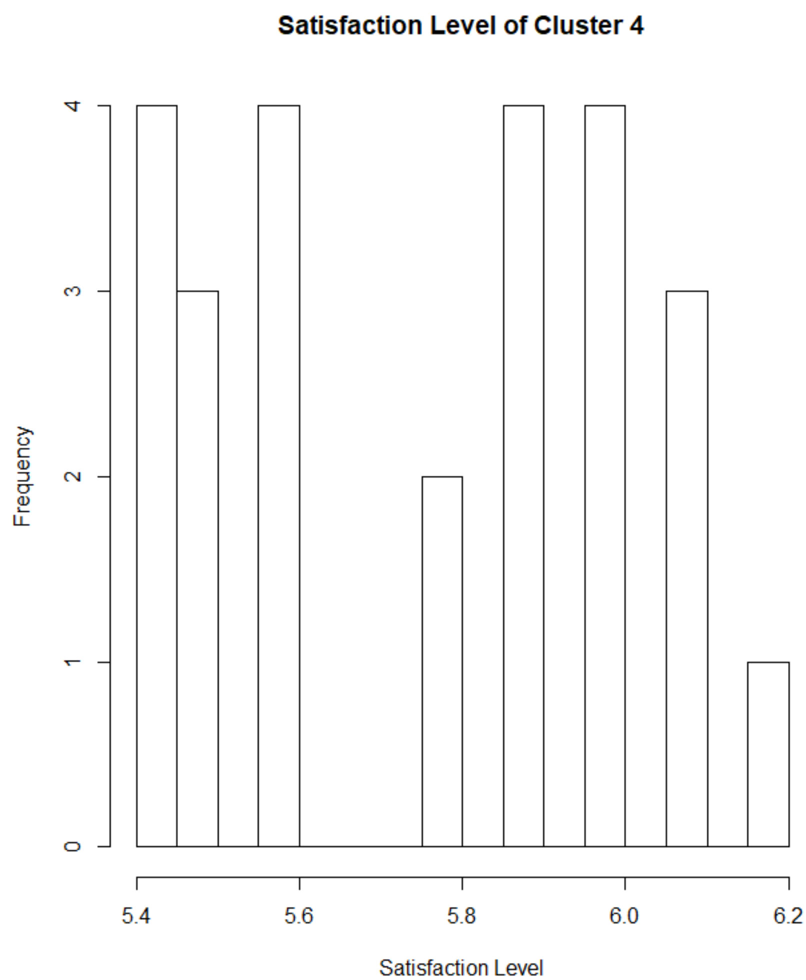
Graph 285: Histogram of Satisfaction Level of Cluster 1



Graph 286: Histogram of Satisfaction Level of Cluster 2



Graph 287: Histogram of Satisfaction Level of Cluster 3



Graph 288: Histogram of Satisfaction Level of Cluster 4

Since Clusters 3 and 4 have the highest satisfaction levels among the other clusters, the company can look at what could have driven the higher satisfaction levels of customers in clusters 3 and 4 by looking at the descriptive statistics of customers in these clusters (as shown above) and analysing the combination of ratings of the different PLE attributes by customers in clusters 3 and 4 that could possibly have driven higher satisfaction levels.

Question 4: Develop a model to simulate 260 working days (1 year), and count the number of additional shifts that are required. Assume that the initial additional inventory is 100 units. Find the distribution of the number of shifts that the company would expect over the next year. Explain and summarize your findings in a report to the plant manager and make a recommendation as to how many shifts to plan in next year's budget.

From the case study, we are given the following information:

<u>Develop a model to simulate 260 working days (1 year)</u>:

- Planned production capacity for one component is 100 units per shift, and the plant operates one shift per day
- Demand fluctuates and is historically between 80 and 130 units per day
- Considering a policy to run a second shift the next day if inventory falls to 50 or below at the end of a day (after the daily demand is known)
- Formula: ending inventory = beginning inventory+ production – demand

We can conduct a monte carlo simulation using for loops to develop numbers to make up a model to predict the number of additional shifts required.

<u>Monte Carlo Simulation</u>

Firstly, we set prepare the data needed for the simulation.

```
set.seed(51669841)
thld = 50
#Set inventory threshold - if the end-of-day inventory level is less than 50
units, then run an extra shift tomorrow.

for(i_sim in 1:1000) {
  for(day in 1:260) {           #for each 260 working days in a year
    if(day == 1) {              #if on day 1:
           beg_invty <- 100          #initial additional inventory
           num_shifts <- 1           #no. of shifts =1
           production <- num_shifts * 100   #production output in a day

           demand <- floor(runif(1, min = 80, max = 131))
#generate 1 random number representing demand


           end_invty <- beg_invty + production - demand #end level of
inventory

           insuff_invty_ind = FALSE  # still got inventory left
           if (end_invty < 0) {
             end_invty <- 0
             insuff_invty_ind = TRUE
           }
      curr_day_entry <- data.frame(day,
                                   beg_invty,
                                   num_shifts,
                                   production,
                                   demand,
                                   end_invty,
                                   insuff_invty_ind)
```

```
      #save current day's inventory log entry record
      invty_log <- curr_day_entry

    } else {                #if on days other than day 1:
      beg_invty <- end_invty
      num_shifts <- ifelse(beg_invty <= thld, 2, 1)

#if BOD inventory <= threshold, add 2nd shift

      production <- num_shifts * 100

      demand <- floor(runif(1, min = 80, max = 131))


      end_invty <- beg_invty + production - demand

      insuff_invty_ind = FALSE
      if (end_invty < 0) {
        end_invty <- 0
        insuff_invty_ind = TRUE
      }
      curr_day_entry <- data.frame(day,
                             beg_invty,
                             num_shifts,
                             production,
                             demand,
                             end_invty,
                             insuff_invty_ind)

      invty_log <- rbind(invty_log, curr_day_entry)

#append today's entry to inventory log

    }
  }

annu_num_shifts <- sum(invty_log$num_shifts)

#at end of the year, calculate annual total no. of shifts required

annu_days_insuff_invty <- sum(invty_log$insuff_invty_ind)

#calculate annual total no. of days with insufficient inventory

if(i_sim == 1) {

#for every simulation, record the above two metrics in a log

    sim_log <- data.frame(i_sim, annu_num_shifts, annu_days_insuff_invty)
  } else {
    sim_log <- rbind(sim_log,
                     data.frame(i_sim,
annu_num_shifts,annu_days_insuff_invty))
       }
}
```

Next, we can then compute the average number of shifts annually.

```
avg_annu_num_shifts <- mean(sim_log$annu_num_shifts)
```

Average annual number of shifts = 273.207 = 274 (round up to nearest whole number)

To find the number of additional shifts needed, we simply take the calculated average annual number of shifts and deduct 260 (the number of shifts based on the normal operation of the factory which operates one shift per day)

```
additional_shifts <- avg_annu_num_shifts - 260
```

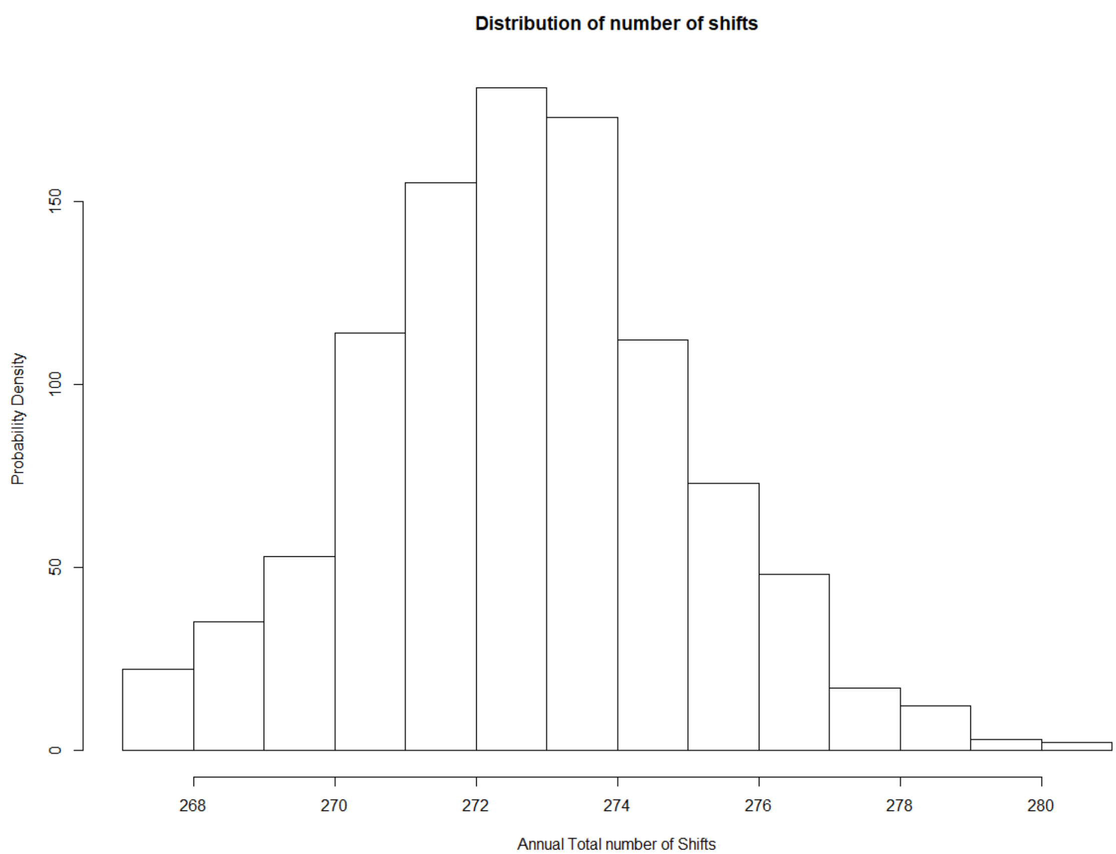
Additional number of shifts needed = 14 (round up to the nearest whole number)

Conclusion

The number of additional shifts required is 14.

Next, we can find the distribution of the number of shifts that the company would expect over the next year.

Using the data generated from the monte carlo simulation above, we can plot a histogram to visualize the distribution of the number of shifts



Graph 289: Histogram showing the distribution of annual total number of shifts in the next year.

From the histogram, we can deduce that the distribution of the number of shifts is approximately normal about the mean.

Furthermore, from the histogram, we can see that the annual total number of shifts peaks at about 273.

Conclusion

Hence, it is recommended for the plant manager to target **273 shifts** the next year to fully optimize resources and manpower.