Predict Price of the computer

**Ans :**

**R Code :**

## Multi Linear Regression

########## Computer Data Set #########

######################

computers <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Multi Linear Regression\\Computer\_Data\_r.csv')

computer <- computers[,-c(1)]

#Model Building

#Regression Model and Summary

model<-lm(price~.,data = computer)

summary(model)

car::vif(model)

#Diagnostic Plots:

#Residual Plots, QQ-Plos, Std. Residuals vs Fitted

plot(model)

#Residuals vs Regressors

library(car)

residualPlots(model)

#Added Variable Plots

avPlots(model)

#QQ plots of studentized residuals

qqPlot(model)

influenceIndexPlot(model) # Index Plots of the influence measures

####Iteration 1

computer1<-computer[-c(1441,1701),]

model1<-lm(price~.,data = computer1)

summary(model1)

car::vif(model1)

plot(model1)

residualPlots(model1)

qqPlot(model1)

influenceIndexPlot(model1)

####Iteration 2

computer2<-computer[-c(1441,1701,994,20,25,3784,4478),]

model2<-lm(price~.,data = computer2)

summary(model2)

car::vif(model2)

plot(model2)

residualPlots(model2)

qqPlot(model2)

influenceIndexPlot(model2)

####Iteration 3

computer3<-computer[-c(1441,1701,994,20,25,3784,4478,982,5961,28,310),]

model3<-lm(price~.,data = computer3)

summary(model3)

car::vif(model3)

plot(model3)

residualPlots(model3)

qqPlot(model3)

influenceIndexPlot(model3)

####Iteration 4

computer4<-computer[-c(1441,1701,994,20,25,3784,4478,982,5961,28,310,1043,208,314,1689,1785),]

model4<-lm(price~.,data = computer4)

summary(model4)

car::vif(model4)

plot(model4)

residualPlots(model4)

qqPlot(model4)

influenceIndexPlot(model4)

####Iteration 5

computer5<-computer[-c(1441,1701,994,20,25,3784,4478,982,5961,28,310,1043,208,314,1689,1785,1123,175,224,721,1049),]

model5<-lm(price~.,data = computer5)

summary(model5)

car::vif(model5)

plot(model5)

residualPlots(model5)

qqPlot(model5)

influenceIndexPlot(model5)

**Results :**

> model<-lm(price~.,data = computer)

> summary(model)

Call:

lm(formula = price ~ ., data = computer)

Residuals:

Min 1Q Median 3Q Max

-1093.77 -174.24 -11.49 146.49 2001.05

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 307.98798 60.35341 5.103 3.44e-07 \*\*\*

speed 9.32028 0.18506 50.364 < 2e-16 \*\*\*

hd 0.78178 0.02761 28.311 < 2e-16 \*\*\*

ram 48.25596 1.06608 45.265 < 2e-16 \*\*\*

screen 123.08904 3.99950 30.776 < 2e-16 \*\*\*

cd 60.91671 9.51559 6.402 1.65e-10 \*\*\*

multi 104.32382 11.41268 9.141 < 2e-16 \*\*\*

premium -509.22473 12.34225 -41.259 < 2e-16 \*\*\*

ads 0.65729 0.05132 12.809 < 2e-16 \*\*\*

trend -51.84958 0.62871 -82.470 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 275.3 on 6249 degrees of freedom

Multiple R-squared: 0.7756, Adjusted R-squared: 0.7752

F-statistic: 2399 on 9 and 6249 DF, p-value: < 2.2e-16

> car::vif(model)

speed hd ram screen cd multi premium ads trend

1.265364 4.207395 2.974628 1.081644 1.859370 1.290568 1.109388 1.217218 2.022790

> residualPlots(model)

Hit <Return> to see next plot:

Hit <Return> to see next plot:

Test stat Pr(>|Test stat|)

speed -15.5971 < 2.2e-16 \*\*\*

hd -20.5563 < 2.2e-16 \*\*\*

ram -14.2167 < 2.2e-16 \*\*\*

screen 11.1308 < 2.2e-16 \*\*\*

cd 1.9685 0.04906 \*

multi -1.4962 0.13466

premium 0.5336 0.59364

ads -7.8401 5.259e-15 \*\*\*

trend -12.1804 < 2.2e-16 \*\*\*

Tukey test 17.6144 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #Added Variable Plots

> avPlots(model)

> #QQ plots of studentized residuals

> qqPlot(model)

[1] 1441 1701

> ####Iteration 1

> computer1<-computer[-c(1441,1701),]

> model1<-lm(price~.,data = computer1)

> summary(model1)

Call:

lm(formula = price ~ ., data = computer1)

Residuals:

Min 1Q Median 3Q Max

-1095.65 -172.78 -10.84 146.42 1510.65

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 337.1635 59.9242 5.627 1.92e-08 \*\*\*

speed 9.2992 0.1835 50.664 < 2e-16 \*\*\*

hd 0.7749 0.0274 28.286 < 2e-16 \*\*\*

ram 48.5222 1.0576 45.878 < 2e-16 \*\*\*

screen 121.0926 3.9714 30.492 < 2e-16 \*\*\*

cd 60.4964 9.4400 6.409 1.58e-10 \*\*\*

multi 104.7703 11.3195 9.256 < 2e-16 \*\*\*

premium -509.8352 12.2409 -41.650 < 2e-16 \*\*\*

ads 0.6510 0.0509 12.791 < 2e-16 \*\*\*

trend -51.6496 0.6238 -82.793 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 273.1 on 6247 degrees of freedom

Multiple R-squared: 0.7777, Adjusted R-squared: 0.7774

F-statistic: 2428 on 9 and 6247 DF, p-value: < 2.2e-16

> qqPlot(model1)

[1] 20 25

> ####Iteration 2

> computer2<-computer[-c(1441,1701,994,20,25,3784,4478),]

> model2<-lm(price~.,data = computer2)

> summary(model2)

Call:

lm(formula = price ~ ., data = computer2)

Residuals:

Min 1Q Median 3Q Max

-1090.71 -172.42 -10.65 146.67 1423.78

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 324.12233 59.57581 5.441 5.52e-08 \*\*\*

speed 9.28668 0.18248 50.891 < 2e-16 \*\*\*

hd 0.77100 0.02784 27.694 < 2e-16 \*\*\*

ram 48.57719 1.06184 45.748 < 2e-16 \*\*\*

screen 121.57335 3.94753 30.797 < 2e-16 \*\*\*

cd 60.22212 9.38348 6.418 1.48e-10 \*\*\*

multi 104.72510 11.25066 9.308 < 2e-16 \*\*\*

premium -509.59558 12.16690 -41.884 < 2e-16 \*\*\*

ads 0.66475 0.05068 13.116 < 2e-16 \*\*\*

trend -51.38983 0.62425 -82.323 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 271.4 on 6242 degrees of freedom

Multiple R-squared: 0.7787, Adjusted R-squared: 0.7784

F-statistic: 2440 on 9 and 6242 DF, p-value: < 2.2e-16

> car::vif(model2)

speed hd ram screen cd multi premium ads trend

1.265720 4.336312 3.034391 1.081753 1.858726 1.290558 1.109421 1.220360 2.049149

> qqPlot(model2)

28 310

26 308

> ####Iteration 3

> computer3<-computer[-c(1441,1701,994,20,25,3784,4478,982,5961,28,310),]

> model3<-lm(price~.,data = computer3)

> summary(model3)

Call:

lm(formula = price ~ ., data = computer3)

Residuals:

Min 1Q Median 3Q Max

-1086.67 -172.19 -10.54 146.47 1328.04

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 313.32373 59.26310 5.287 1.29e-07 \*\*\*

speed 9.31265 0.18151 51.307 < 2e-16 \*\*\*

hd 0.76624 0.02799 27.376 < 2e-16 \*\*\*

ram 48.61951 1.06061 45.841 < 2e-16 \*\*\*

screen 121.93421 3.92618 31.057 < 2e-16 \*\*\*

cd 60.93335 9.33869 6.525 7.34e-11 \*\*\*

multi 104.97699 11.19961 9.373 < 2e-16 \*\*\*

premium -509.90120 12.10072 -42.138 < 2e-16 \*\*\*

ads 0.67443 0.05045 13.367 < 2e-16 \*\*\*

trend -51.19896 0.62278 -82.210 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 269.9 on 6238 degrees of freedom

Multiple R-squared: 0.7801, Adjusted R-squared: 0.7798

F-statistic: 2459 on 9 and 6238 DF, p-value: < 2.2e-16

> car::vif(model3)

speed hd ram screen cd multi premium ads trend

1.264623 4.396218 3.059151 1.081702 1.860354 1.291703 1.109496 1.221329 2.059027

> residualPlots(model3)

Hit <Return> to see next plot:

Hit <Return> to see next plot:

Test stat Pr(>|Test stat|)

speed -15.8169 < 2e-16 \*\*\*

hd -21.7130 < 2e-16 \*\*\*

ram -14.1876 < 2e-16 \*\*\*

screen 10.7956 < 2e-16 \*\*\*

cd -2.0661 0.03886 \*

multi -1.5135 0.13019

premium 0.5255 0.59923

ads -8.2882 < 2e-16 \*\*\*

trend -12.8034 < 2e-16 \*\*\*

Tukey test 17.8225 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> qqPlot(model3)

208 314

205 310

> ####Iteration 4

> computer4<-computer[-c(1441,1701,994,20,25,3784,4478,982,5961,28,310,1043,208,314,1689,1785),]

> model4<-lm(price~.,data = computer4)

> summary(model4)

Call:

lm(formula = price ~ ., data = computer4)

Residuals:

Min 1Q Median 3Q Max

-1084.0 -171.5 -10.4 146.5 1286.6

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 302.07917 58.99017 5.121 3.13e-07 \*\*\*

speed 9.31905 0.18066 51.584 < 2e-16 \*\*\*

hd 0.75479 0.02833 26.646 < 2e-16 \*\*\*

ram 48.86959 1.06562 45.860 < 2e-16 \*\*\*

screen 122.38607 3.90738 31.322 < 2e-16 \*\*\*

cd 60.96660 9.29814 6.557 5.94e-11 \*\*\*

multi 104.90919 11.14591 9.412 < 2e-16 \*\*\*

premium -509.66803 12.04116 -42.327 < 2e-16 \*\*\*

ads 0.68199 0.05029 13.562 < 2e-16 \*\*\*

trend -50.91655 0.62318 -81.704 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 268.6 on 6233 degrees of freedom

Multiple R-squared: 0.7812, Adjusted R-squared: 0.7808

F-statistic: 2472 on 9 and 6233 DF, p-value: < 2.2e-16

> car::vif(model4)

speed hd ram screen cd multi premium ads trend

1.264802 4.522071 3.117268 1.081715 1.861178 1.291888 1.109419 1.223861 2.079760

> qqPlot(model4)

175 224

172 220

> ####Iteration 5

> computer5<-computer[-c(1441,1701,994,20,25,3784,4478,982,5961,28,310,1043,208,314,1689,1785,1123,175,224,721,1049),]

> model5<-lm(price~.,data = computer5)

> summary(model5)

Call:

lm(formula = price ~ ., data = computer5)

Residuals:

Min 1Q Median 3Q Max

-1081.27 -171.40 -10.61 147.13 1277.84

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 290.95065 58.73688 4.953 7.48e-07 \*\*\*

speed 9.31626 0.17988 51.791 < 2e-16 \*\*\*

hd 0.74319 0.02864 25.951 < 2e-16 \*\*\*

ram 49.12902 1.06864 45.973 < 2e-16 \*\*\*

screen 122.83598 3.89002 31.577 < 2e-16 \*\*\*

cd 60.97361 9.26076 6.584 4.95e-11 \*\*\*

multi 104.80640 11.09583 9.446 < 2e-16 \*\*\*

premium -509.35215 11.98533 -42.498 < 2e-16 \*\*\*

ads 0.68982 0.05013 13.761 < 2e-16 \*\*\*

trend -50.61746 0.62389 -81.131 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 267.3 on 6228 degrees of freedom

Multiple R-squared: 0.782, Adjusted R-squared: 0.7817

F-statistic: 2482 on 9 and 6228 DF, p-value: < 2.2e-16

> car::vif(model5)

speed hd ram screen cd multi premium ads trend

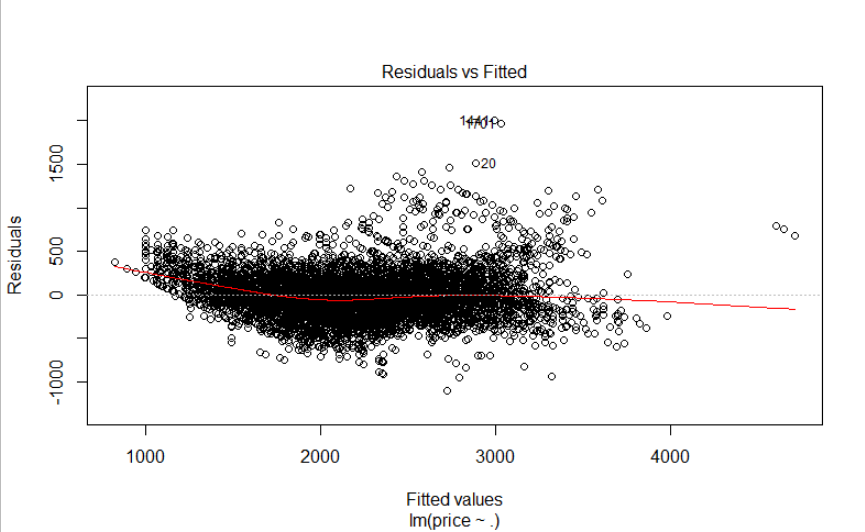
1.265119 4.639050 3.163363 1.081767 1.862145 1.292121 1.109347 1.226108 2.100929

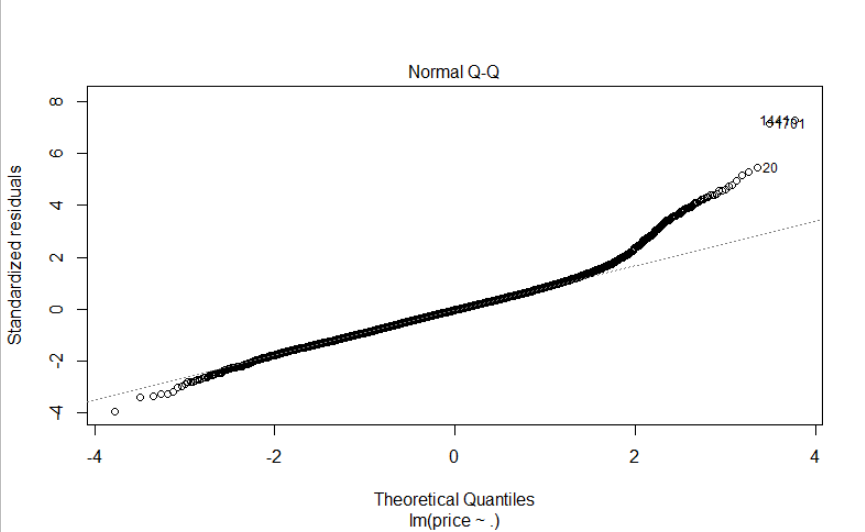
> qqPlot(model5)

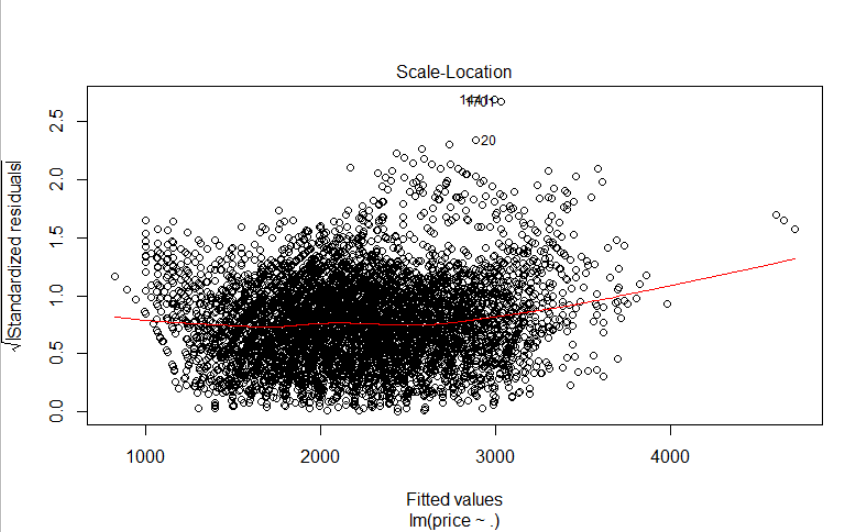
291 1806

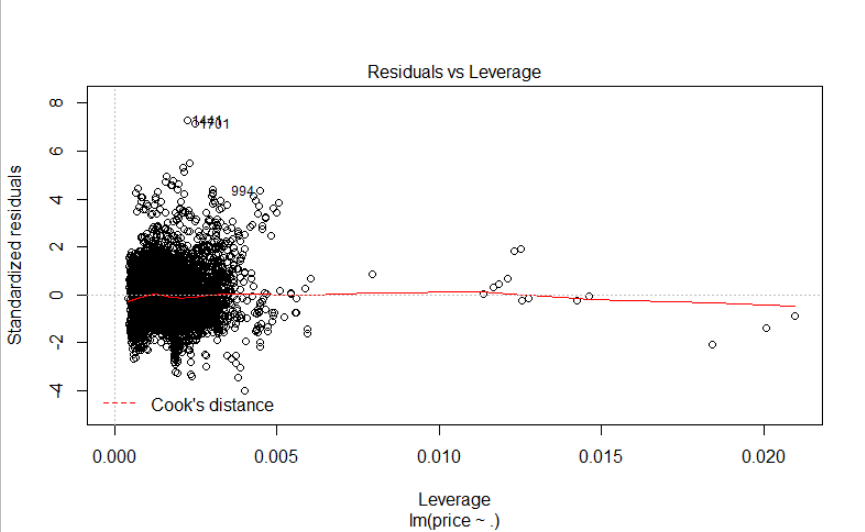
285 1788

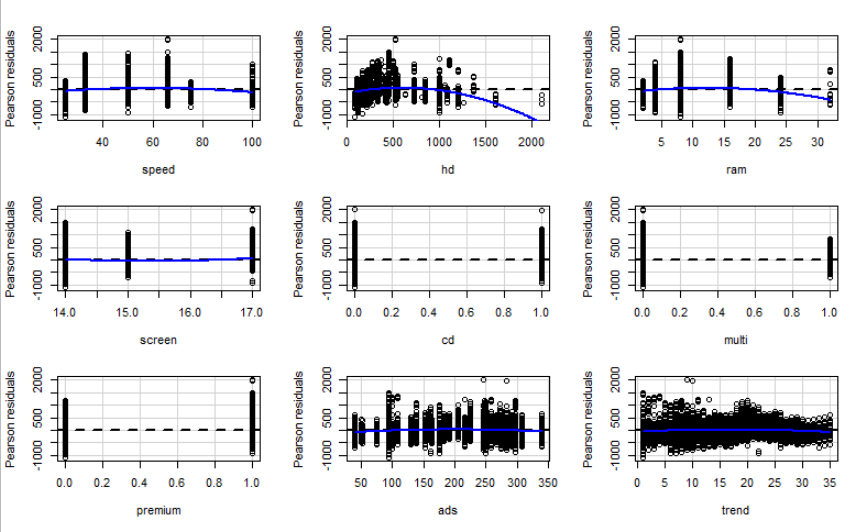
**Plots :**

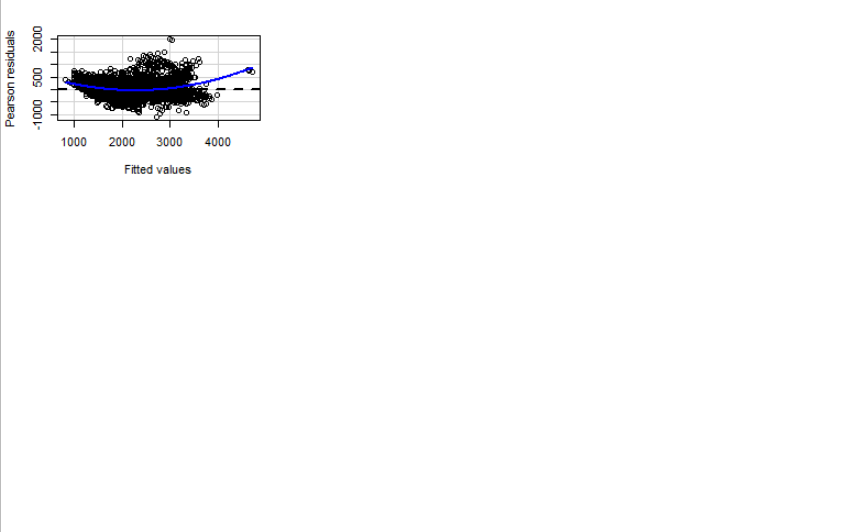


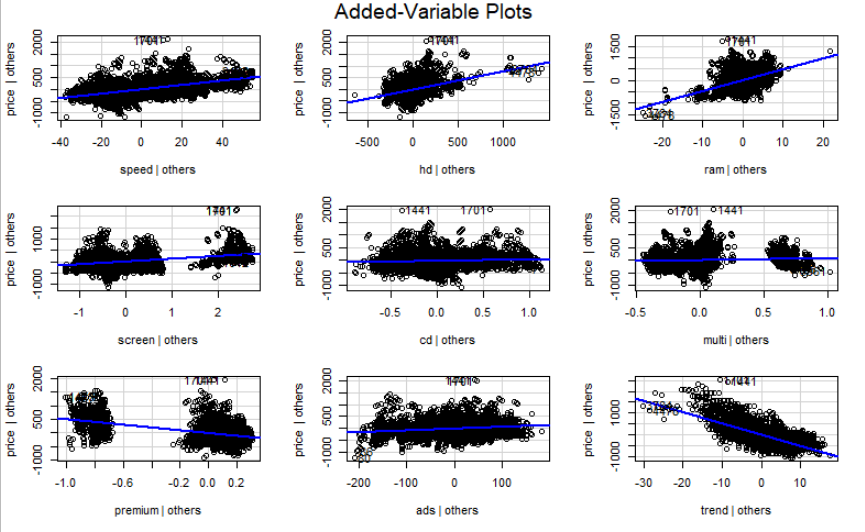


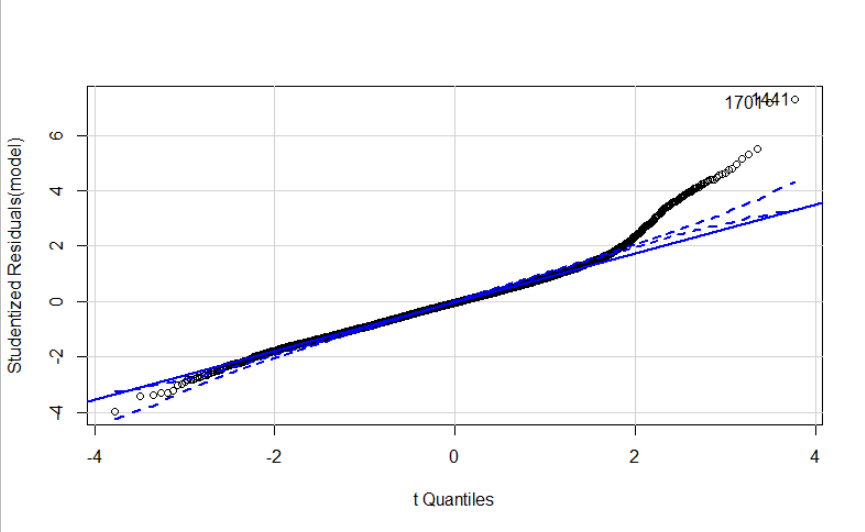


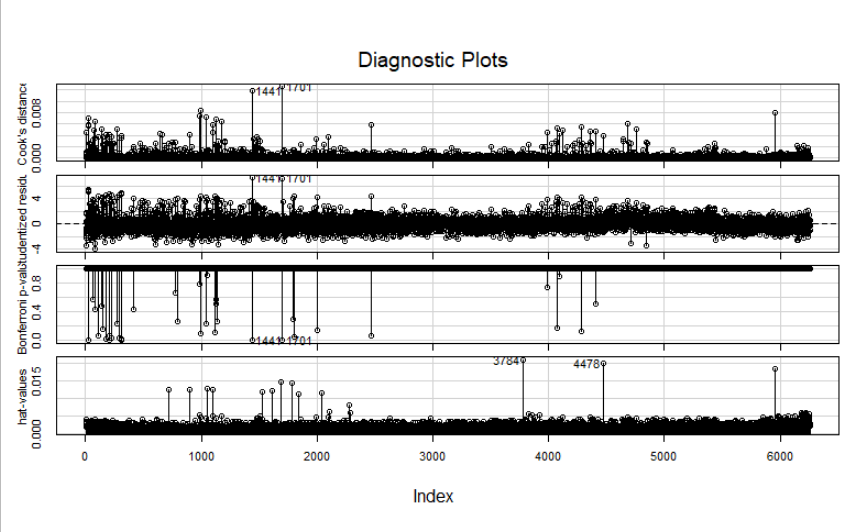












**Inference :**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr No. | Model No. | Multiple R-squared | Adjusted R-squared |
| 1 | Model | 0.7756 | 0.7752 |
| 2 | Model1 | 0.7777 | 0.7774 |
| 3 | Model2 | 0.7787 | 0.7784 |
| 4 | Model3 | 0.7801 | 0.7798 |
| 5 | Model4 | 0.7812 | 0.7808 |
| 6 | Model5 | 0.782 | 0.7817 |

Hd, ram, trend, cd were some of the important features found.