Consider only the below columns and prepare a prediction model for predicting Price.

Corolla<-Corolla[c("Price","Age\_08\_04","KM","HP","cc","Doors","Gears","Quarterly\_Tax","Weight")]

**Ans :**

**R Code :**

## Multi Linear Regression

########## Toyota Corolla Data Set #########

Cars1 <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Multi Linear Regression\\ToyotaCorolla.csv')

Cars <- Cars1[c("Price","Age\_08\_04","KM","HP","cc","Doors","Gears","Quarterly\_Tax","Weight")]

#Model Building

#Regression Model and Summary

model.car<-lm(Price~.,data = Cars)

summary(model.car)

car::vif(model.car)

#Diagnostic Plots:

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

plot(model.car)

#Residuals vs Regressors

library(car)

residualPlots(model.car)

#Added Variable Plots

avPlots(model.car)

#QQ plots of studentized residuals

qqPlot(model.car)

influenceIndexPlot(model.car) # Index Plots of the influence measures

####Iteration 1

#Remove 77th observation

Cars1<-Cars[-81,]

model1<-lm(Price~.,data = Cars1)

summary(model1)

car::vif(model1)

plot(model1)

residualPlots(model1)

qqPlot(model1)

influenceIndexPlot(model1)

#iteration2

Cars2<-Cars[-c(81,222,961),]

model2<-lm(Price~.,data = Cars2)

summary(model2)

car::vif(model2)

plot(model2)

residualPlots(model2)

qqPlot(model2)

influenceIndexPlot(model2)

#iteration3

Cars3<-Cars[-c(81,222,961,602,148),]

model3<-lm(Price~.,data = Cars3)

summary(model3)

car::vif(model3)

plot(model3)

residualPlots(model3)

qqPlot(model3)

influenceIndexPlot(model3)

#iteration4

Cars4<-Cars[-c(81,222,961,602,148,655,524),]

model4<-lm(Price~.,data = Cars4)

summary(model4)

car::vif(model4)

plot(model4)

residualPlots(model4)

qqPlot(model4)

influenceIndexPlot(model4)

#iteration5

Cars5<-Cars[-c(81,222,961,602,148,655,524,193,192,992),]

model5<-lm(Price~.,data = Cars5)

summary(model5)

car::vif(model5)

plot(model5)

residualPlots(model5)

qqPlot(model5)

influenceIndexPlot(model5)

**Results :**

> model.car<-lm(Price~.,data = Cars)

> summary(model.car)

Call:

lm(formula = Price ~ ., data = Cars)

Residuals:

Min 1Q Median 3Q Max

-9366.4 -793.3 -21.3 799.7 6444.0

Coefficients:

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

(Intercept) -5.573e+03 1.411e+03 -3.949 8.24e-05 \*\*\*

Age\_08\_04 -1.217e+02 2.616e+00 -46.512 < 2e-16 \*\*\*

KM -2.082e-02 1.252e-03 -16.622 < 2e-16 \*\*\*

HP 3.168e+01 2.818e+00 11.241 < 2e-16 \*\*\*

cc -1.211e-01 9.009e-02 -1.344 0.17909

Doors -1.617e+00 4.001e+01 -0.040 0.96777

Gears 5.943e+02 1.971e+02 3.016 0.00261 \*\*

Quarterly\_Tax 3.949e+00 1.310e+00 3.015 0.00262 \*\*

Weight 1.696e+01 1.068e+00 15.880 < 2e-16 \*\*\*

---

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

Residual standard error: 1342 on 1427 degrees of freedom

Multiple R-squared: 0.8638, Adjusted R-squared: 0.863

F-statistic: 1131 on 8 and 1427 DF, p-value: < 2.2e-16

> car::vif(model.car)

Age\_08\_04 KM HP cc Doors Gears Quarterly\_Tax

1.884620 1.756905 1.419422 1.163894 1.156575 1.098723 2.311431

Weight

2.516420

> residualPlots(model.car)

Test stat Pr(>|Test stat|)

Age\_08\_04 16.4550 < 2.2e-16 \*\*\*

KM 1.7226 0.085186 .

HP 4.2000 2.835e-05 \*\*\*

cc 8.3001 2.385e-16 \*\*\*

Doors -1.2269 0.220068

Gears -1.0422 0.297474

Quarterly\_Tax 2.7799 0.005508 \*\*

Weight 0.0188 0.985010

Tukey test 21.2798 < 2.2e-16 \*\*\*

---

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

> qqPlot(model.car)

[1] 81 222

> ####Iteration 1

> #Remove 81th observation

> Cars1<-Cars[-81,]

> model1<-lm(Price~.,data = Cars1)

> summary(model1)

Call:

lm(formula = Price ~ ., data = Cars1)

Residuals:

Min 1Q Median 3Q Max

-11455.7 -761.7 -32.7 739.3 6739.7

Coefficients:

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

(Intercept) -6.285e+03 1.383e+03 -4.545 5.95e-06 \*\*\*

Age\_08\_04 -1.205e+02 2.562e+00 -47.021 < 2e-16 \*\*\*

KM -1.785e-02 1.277e-03 -13.973 < 2e-16 \*\*\*

HP 3.935e+01 2.911e+00 13.516 < 2e-16 \*\*\*

cc -2.524e+00 3.072e-01 -8.216 4.67e-16 \*\*\*

Doors -2.723e+01 3.924e+01 -0.694 0.48788

Gears 5.239e+02 1.929e+02 2.717 0.00667 \*\*

Quarterly\_Tax 9.044e+00 1.425e+00 6.348 2.93e-10 \*\*\*

Weight 2.017e+01 1.116e+00 18.076 < 2e-16 \*\*\*

---

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

Residual standard error: 1313 on 1426 degrees of freedom

Multiple R-squared: 0.8694, Adjusted R-squared: 0.8686

F-statistic: 1186 on 8 and 1426 DF, p-value: < 2.2e-16

> car::vif(model1)

Age\_08\_04 KM HP cc Doors Gears Quarterly\_Tax

1.887229 1.909570 1.583835 2.754405 1.163178 1.100907 2.859861

Weight

2.864117

> residualPlots(model1)

Test stat Pr(>|Test stat|)

Age\_08\_04 15.6083 < 2.2e-16 \*\*\*

KM 2.3562 0.0185997 \*

HP 4.1629 3.331e-05 \*\*\*

cc 7.6206 4.584e-14 \*\*\*

Doors -2.4444 0.0146288 \*

Gears -1.2190 0.2230339

Quarterly\_Tax 3.5641 0.0003772 \*\*\*

Weight -4.3022 1.806e-05 \*\*\*

Tukey test 18.1533 < 2.2e-16 \*\*\*

---

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

> qqPlot(model1)

222 961

221 960

> #iteration2

> Cars2<-Cars[-c(81,222,961),]

> model2<-lm(Price~.,data = Cars2)

> summary(model2)

Call:

lm(formula = Price ~ ., data = Cars2)

Residuals:

Min 1Q Median 3Q Max

-8756.8 -761.3 -31.7 720.6 6306.6

Coefficients:

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

(Intercept) -1.474e+04 1.433e+03 -10.289 < 2e-16 \*\*\*

Age\_08\_04 -1.120e+02 2.479e+00 -45.185 < 2e-16 \*\*\*

KM -1.699e-02 1.200e-03 -14.160 < 2e-16 \*\*\*

HP 3.661e+01 2.745e+00 13.334 < 2e-16 \*\*\*

cc -3.795e+00 3.021e-01 -12.562 < 2e-16 \*\*\*

Doors -1.225e+02 3.748e+01 -3.270 0.00110 \*\*

Gears 4.650e+02 1.810e+02 2.569 0.01029 \*

Quarterly\_Tax 5.213e+00 1.371e+00 3.802 0.00015 \*\*\*

Weight 3.064e+01 1.290e+00 23.748 < 2e-16 \*\*\*

---

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

Residual standard error: 1231 on 1424 degrees of freedom

Multiple R-squared: 0.8852, Adjusted R-squared: 0.8845

F-statistic: 1372 on 8 and 1424 DF, p-value: < 2.2e-16

> car::vif(model2)

Age\_08\_04 KM HP cc Doors Gears Quarterly\_Tax

2.007438 1.914560 1.599075 3.022852 1.203959 1.101509 2.997686

Weight

3.846677

> #iteration3

> Cars3<-Cars[-c(81,222,961,602,148),]

> model3<-lm(Price~.,data = Cars3)

> summary(model3)

Call:

lm(formula = Price ~ ., data = Cars3)

Residuals:

Min 1Q Median 3Q Max

-6085.7 -760.6 -42.4 714.1 6000.5

Coefficients:

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

(Intercept) -1.617e+04 1.406e+03 -11.497 < 2e-16 \*\*\*

Age\_08\_04 -1.098e+02 2.423e+00 -45.316 < 2e-16 \*\*\*

KM -1.770e-02 1.171e-03 -15.118 < 2e-16 \*\*\*

HP 3.437e+01 2.687e+00 12.788 < 2e-16 \*\*\*

cc -3.693e+00 2.942e-01 -12.554 < 2e-16 \*\*\*

Doors -1.470e+02 3.656e+01 -4.021 6.11e-05 \*\*\*

Gears 4.690e+02 1.760e+02 2.665 0.00779 \*\*

Quarterly\_Tax 3.796e+00 1.349e+00 2.813 0.00497 \*\*

Weight 3.215e+01 1.272e+00 25.285 < 2e-16 \*\*\*

---

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

Residual standard error: 1197 on 1422 degrees of freedom

Multiple R-squared: 0.8904, Adjusted R-squared: 0.8898

F-statistic: 1444 on 8 and 1422 DF, p-value: < 2.2e-16

> car::vif(model3)

Age\_08\_04 KM HP cc Doors Gears Quarterly\_Tax

2.020616 1.920131 1.619237 3.018721 1.209814 1.101486 3.068381

Weight

3.907992

> #iteration4

> Cars4<-Cars[-c(81,222,961,602,148,655,524),]

> model4<-lm(Price~.,data = Cars4)

> summary(model4)

Call:

lm(formula = Price ~ ., data = Cars4)

Residuals:

Min 1Q Median 3Q Max

-5254.5 -756.7 -44.7 706.3 4863.1

Coefficients:

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

(Intercept) -1.697e+04 1.399e+03 -12.126 < 2e-16 \*\*\*

Age\_08\_04 -1.090e+02 2.387e+00 -45.661 < 2e-16 \*\*\*

KM -1.761e-02 1.150e-03 -15.315 < 2e-16 \*\*\*

HP 3.349e+01 2.652e+00 12.628 < 2e-16 \*\*\*

cc -3.698e+00 2.890e-01 -12.792 < 2e-16 \*\*\*

Doors -1.550e+02 3.611e+01 -4.294 1.88e-05 \*\*\*

Gears 4.711e+02 1.729e+02 2.725 0.00651 \*\*

Quarterly\_Tax 3.206e+00 1.350e+00 2.376 0.01764 \*

Weight 3.301e+01 1.275e+00 25.889 < 2e-16 \*\*\*

---

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

Residual standard error: 1176 on 1420 degrees of freedom

Multiple R-squared: 0.8939, Adjusted R-squared: 0.8933

F-statistic: 1496 on 8 and 1420 DF, p-value: < 2.2e-16

> residualPlots(model4)

Test stat Pr(>|Test stat|)

Age\_08\_04 11.3776 < 2.2e-16 \*\*\*

KM 1.7707 0.0768257 .

HP 3.5014 0.0004772 \*\*\*

cc 5.1519 2.941e-07 \*\*\*

Doors -5.2278 1.972e-07 \*\*\*

Gears -2.1636 0.0306649 \*

Quarterly\_Tax -1.1889 0.2346969

Weight 4.5060 7.149e-06 \*\*\*

Tukey test 13.3866 < 2.2e-16 \*\*\*

---

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

> qqPlot(model4)

192 193

190 191

> #iteration5

> Cars5<-Cars[-c(81,222,961,602,148,655,524,193,192,992),]

> model5<-lm(Price~.,data = Cars5)

> summary(model5)

Call:

lm(formula = Price ~ ., data = Cars5)

Residuals:

Min 1Q Median 3Q Max

-4466.7 -759.2 -51.5 699.8 4822.1

Coefficients:

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

(Intercept) -1.670e+04 1.390e+03 -12.018 < 2e-16 \*\*\*

Age\_08\_04 -1.098e+02 2.353e+00 -46.651 < 2e-16 \*\*\*

KM -1.703e-02 1.134e-03 -15.016 < 2e-16 \*\*\*

HP 3.185e+01 2.634e+00 12.091 < 2e-16 \*\*\*

cc -3.571e+00 2.851e-01 -12.525 < 2e-16 \*\*\*

Doors -1.571e+02 3.562e+01 -4.410 1.11e-05 \*\*\*

Gears 3.238e+02 1.783e+02 1.816 0.0696 .

Quarterly\_Tax 1.849e+00 1.342e+00 1.378 0.1684

Weight 3.355e+01 1.256e+00 26.705 < 2e-16 \*\*\*

---

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

Residual standard error: 1156 on 1417 degrees of freedom

Multiple R-squared: 0.8973, Adjusted R-squared: 0.8967

F-statistic: 1547 on 8 and 1417 DF, p-value: < 2.2e-16

> car::vif(model5)

Age\_08\_04 KM HP cc Doors Gears Quarterly\_Tax

2.041727 1.918129 1.656363 3.022426 1.227613 1.114435 3.244722

Weight

4.042234

> residualPlots(model5)

Test stat Pr(>|Test stat|)

Age\_08\_04 11.0101 < 2.2e-16 \*\*\*

KM 1.6508 0.09901 .

HP 4.1836 3.045e-05 \*\*\*

cc 4.6839 3.084e-06 \*\*\*

Doors -5.5116 4.219e-08 \*\*\*

Gears -0.0614 0.95103

Quarterly\_Tax -1.7409 0.08191 .

Weight 4.5350 6.247e-06 \*\*\*

Tukey test 13.3747 < 2.2e-16 \*\*\*

---

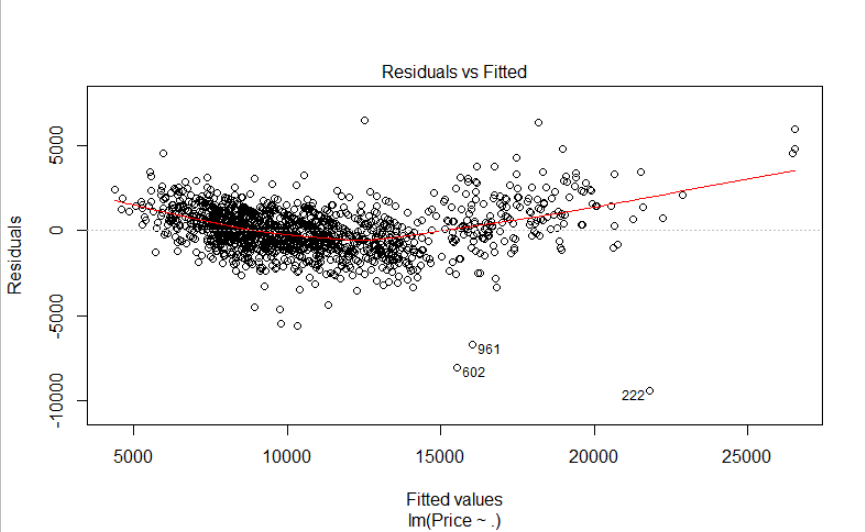
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

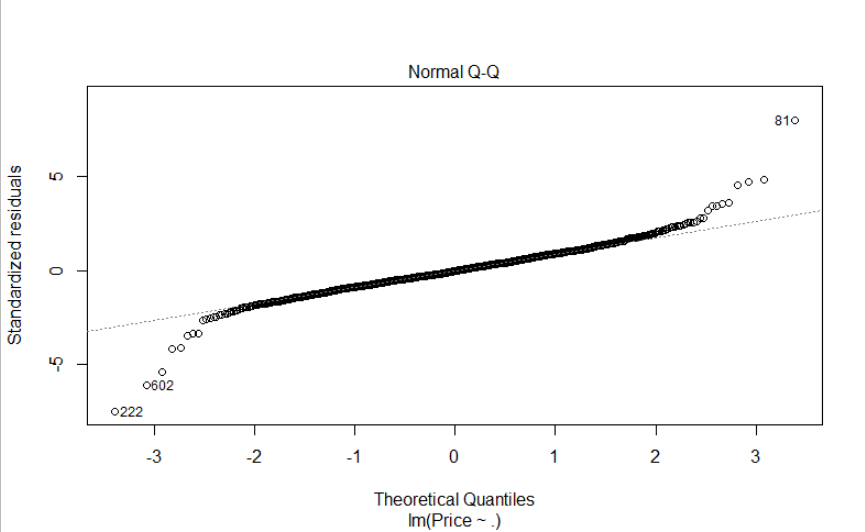
> qqPlot(model5)

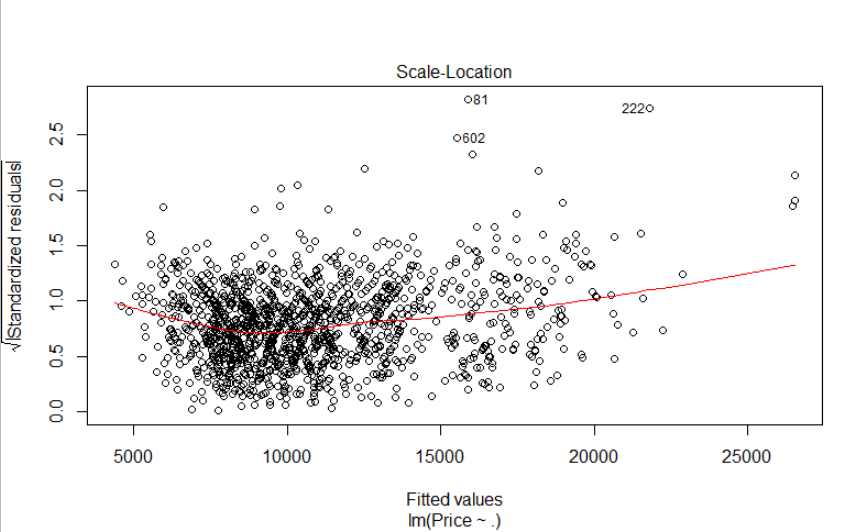
172 394

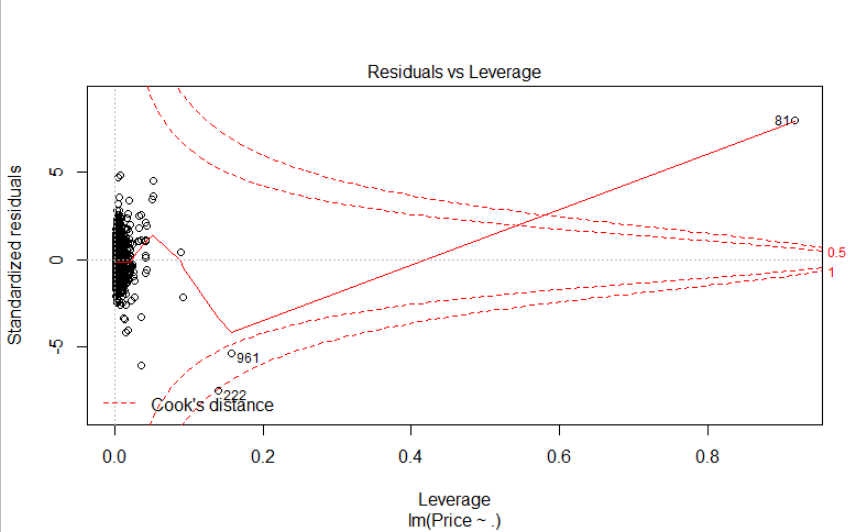
170 389

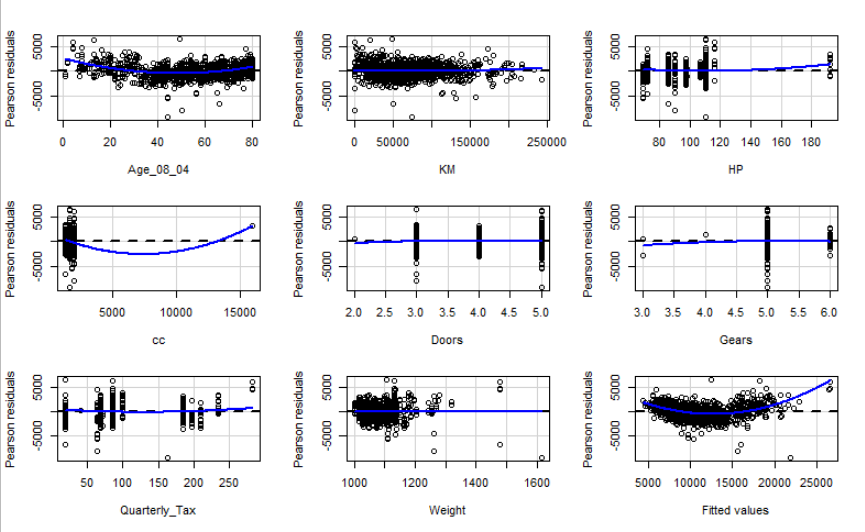
**Plots :**

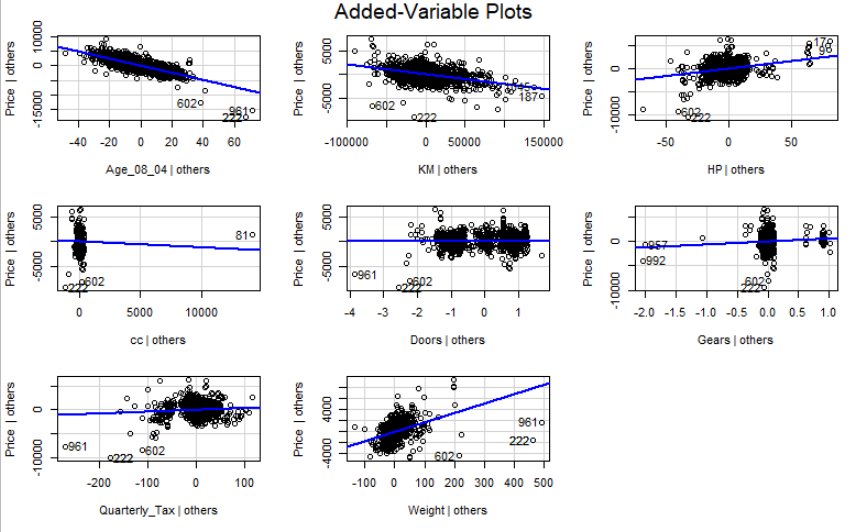


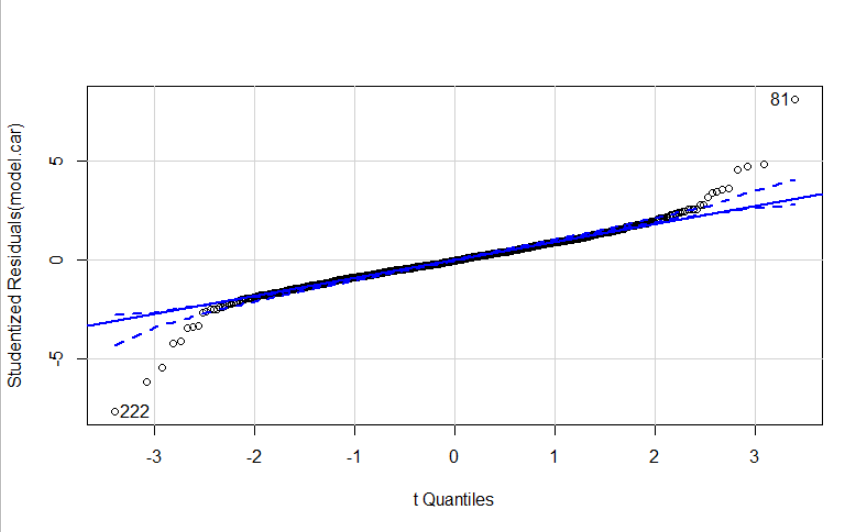


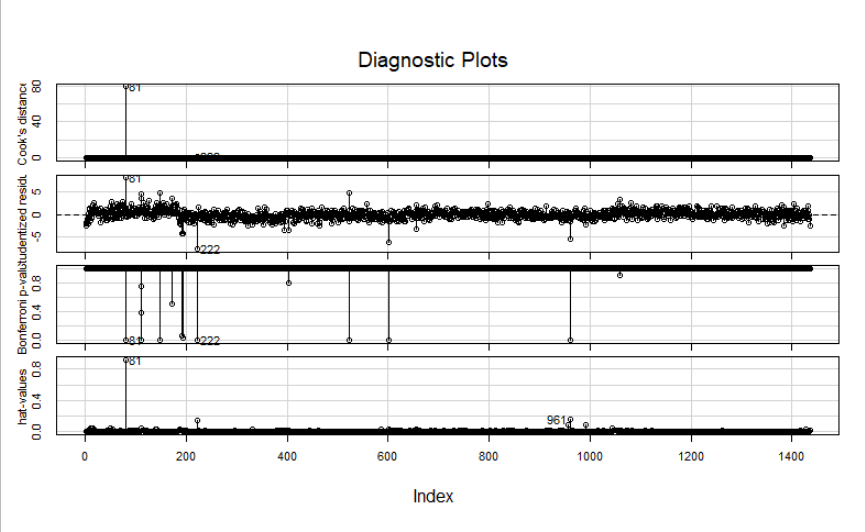












**Inference :**

Getting good Accuracy in Model.

|  |  |  |  |
| --- | --- | --- | --- |
| Sr No. | Model No. | Multiple R-squared | Adjusted R-squared |
| 1 | Model.car | 0.8638 | 0.863 |
| 2 | Model1 | 0.8694 | 0.8686 |
| 3 | Model2 | 0.8852 | 0.8845 |
| 4 | Model3 | 0.8904 | 0.8898 |
| 5 | Model4 | 0.8939 | 0.8933 |
| 6 | Model5 | 0.8973 | 0.8967 |

Weight, Quarterly Tax, CC, Age were some of the important features found.