

Assignment7

Lokesh Arora, Ankita Shinde

10/29/2020

<https://github.com/ankita1598/Walmart>

```
#Loading Packages  
library(mvtnorm)
```

```
## Warning: package 'mvtnorm' was built under R version 3.6.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 3.6.3
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.6.3
```

```
##  
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':  
##  
##    date, intersect, setdiff, union
```

```
library("plyr")
```

```
## Warning: package 'plyr' was built under R version 3.6.3
```

```
## -----
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)
```

```
## -----
```

```
##  
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':  
##  
##    arrange, count, desc, failwith, id, mutate, rename, summarise,  
##    summarize
```

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
##  
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':  
##  
##    %+%, alpha
```

```
library(RColorBrewer)  
library("dplyr")  
library(carData)
```

```
## Warning: package 'carData' was built under R version 3.6.3
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.6.3
```

```
##  
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':  
##  
##    logit
```

```
## The following object is masked from 'package:dplyr':  
##  
##    recode
```

#Loading Dataset

```
dataset = read.csv("data.csv", header= T)
head(dataset)
```

```
##   Store Dept      Date weeklySales isHoliday Type   Size Temperature
## 1      1      1 2010-02-05    24924.50      False   A 151315      42.31
## 2      1      1 2010-02-12    46039.49       True   A 151315      38.51
## 3      1      1 2010-02-19    41595.55      False   A 151315      39.93
## 4      1      1 2010-02-26    19403.54      False   A 151315      46.63
## 5      1      1 2010-03-05    21827.90      False   A 151315      46.50
## 6      1      1 2010-03-12    21043.39      False   A 151315      57.79
##   Fuel_Price Markdown1 Markdown2 Markdown3 Markdown4 Markdown5      CPI
## 1      2.572         NA         NA         NA         NA         NA 211.0964
## 2      2.548         NA         NA         NA         NA         NA 211.2422
## 3      2.514         NA         NA         NA         NA         NA 211.2891
## 4      2.561         NA         NA         NA         NA         NA 211.3196
## 5      2.625         NA         NA         NA         NA         NA 211.3501
## 6      2.667         NA         NA         NA         NA         NA 211.3806
##   Unemployment
## 1          8.106
## 2          8.106
## 3          8.106
## 4          8.106
## 5          8.106
## 6          8.106
```

#We can see that there are few null values in the data set for column Markdown 1 - 5. We will also split the data column in 3 as Day, Month and Year.

```
dataset$Year <- year(ymd(dataset$Date))
dataset$Month <- month(ymd(dataset$Date))
dataset$Day <- day(ymd(dataset$Date))
dataset$Dept = as.factor(dataset$Dept)
dataset$Store = as.factor(dataset$Store)
dataset$MarkDown1[is.na(dataset$MarkDown1)] = 0
dataset$MarkDown2[is.na(dataset$MarkDown2)] = 0
dataset$MarkDown3[is.na(dataset$MarkDown3)] = 0
dataset$MarkDown4[is.na(dataset$MarkDown4)] = 0
dataset$MarkDown5[is.na(dataset$MarkDown5)] = 0
dataset = fastDummies::dummy_cols(dataset, select_columns = "Type")
dataset$IsHoliday[dataset$isHoliday == "False"] = 0
dataset$IsHoliday[dataset$isHoliday == "True"] = 1
dataset$Dept = as.numeric(as.factor(dataset$Dept))
dataset$Store = as.numeric(as.factor(dataset$Store))
features = c("weeklySales", "Store", "Dept", "IsHoliday", "Type_A", "Type_B", "Type_C", "Size", "Temperature", "Fuel_Price", "MarkDown
1", "MarkDown2", "MarkDown3", "MarkDown4", "MarkDown5", "CPI", "Unemployment", "Year", "Month", "Day")
dataset = select(dataset, features)
```

```
## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(features)` instead of `features` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```

```
head(dataset)
```

```
## weeklySales Store Dept IsHoliday Type_A Type_B Type_C Size Temperature
## 1 24924.50 1 1 0 1 0 0 151315 42.31
## 2 46039.49 1 1 1 1 0 0 151315 38.51
## 3 41595.55 1 1 0 1 0 0 151315 39.93
## 4 19403.54 1 1 0 1 0 0 151315 46.63
## 5 21827.90 1 1 0 1 0 0 151315 46.50
## 6 21043.39 1 1 0 1 0 0 151315 57.79
## Fuel_Price Markdown1 Markdown2 Markdown3 Markdown4 Markdown5 CPI
## 1 2.572 0 0 0 0 0 0 211.0964
## 2 2.548 0 0 0 0 0 0 211.2422
## 3 2.514 0 0 0 0 0 0 211.2891
## 4 2.561 0 0 0 0 0 0 211.3196
## 5 2.625 0 0 0 0 0 0 211.3501
## 6 2.667 0 0 0 0 0 0 211.3806
## Unemployment Year Month Day
## 1 8.106 2010 2 5
## 2 8.106 2010 2 12
## 3 8.106 2010 2 19
## 4 8.106 2010 2 26
## 5 8.106 2010 3 5
## 6 8.106 2010 3 12
```

```
dim(dataset)
```

```
## [1] 421570 20
```

```
names(dataset)
```

```
## [1] "weeklySales" "Store" "Dept" "IsHoliday" "Type_A"
## [6] "Type_B" "Type_C" "Size" "Temperature" "Fuel_Price"
## [11] "Markdown1" "Markdown2" "Markdown3" "Markdown4" "Markdown5"
## [16] "CPI" "Unemployment" "Year" "Month" "Day"
```

```
model = lm(weeklySales~.,data=dataset)
summary(model)
```

```
##
## Call:
## lm(formula = weeklySales ~ ., data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40059 -12789  -6043   5390 673101
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.074e+06  1.855e+05  11.183  < 2e-16 ***
## Store       -1.423e+02  3.098e+00  -45.935  < 2e-16 ***
## Dept        1.038e+02  1.371e+00   75.695  < 2e-16 ***
## IsHoliday    3.395e+02  1.425e+02   2.383  0.01717 *
## Type_A      -5.872e+03  1.846e+02  -31.813  < 2e-16 ***
## Type_B      -6.412e+03  1.476e+02  -43.452  < 2e-16 ***
## Type_C              NA          NA      NA      NA
## Size         9.564e-02  9.766e-04  97.938  < 2e-16 ***
## Temperature  5.913e+00  2.086e+00   2.835  0.00459 **
## Fuel_Price   8.312e+02  1.377e+02   6.038  1.56e-09 ***
## Markdown1    5.015e-02  1.177e-02   4.261  2.03e-05 ***
## Markdown2    5.019e-03  7.048e-03   0.712  0.47634
## Markdown3    1.093e-01  6.374e-03  17.147  < 2e-16 ***
## Markdown4   -3.631e-05  1.627e-02  -0.002  0.99822
## Markdown5    1.033e-01  9.415e-03  10.976  < 2e-16 ***
## CPI          -2.317e+01  9.999e-01  -23.173  < 2e-16 ***
## Unemployment -2.693e+02  2.059e+01  -13.079  < 2e-16 ***
## Year        -1.027e+03  9.241e+01  -11.112  < 2e-16 ***
## Month        1.253e+02  1.164e+01  10.769  < 2e-16 ***
## Day         -1.572e+01  3.911e+00  -4.019  5.84e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21770 on 421551 degrees of freedom
## Multiple R-squared:  0.08124,    Adjusted R-squared:  0.0812
## F-statistic: 2071 on 18 and 421551 DF,  p-value: < 2.2e-16
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.6.3
```

```
##  
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':  
##  
##      select
```

```
stepModelAIC = stepAIC(model, direction = "backward")
```



```

## Start: AIC=8421528
## weeklySales ~ Store + Dept + IsHoliday + Type_A + Type_B + Type_C +
##   Size + Temperature + Fuel_Price + Markdown1 + Markdown2 +
##   Markdown3 + Markdown4 + Markdown5 + CPI + Unemployment +
##   Year + Month + Day
##
##
## Step: AIC=8421528
## weeklySales ~ Store + Dept + IsHoliday + Type_A + Type_B + Size +
##   Temperature + Fuel_Price + Markdown1 + Markdown2 + Markdown3 +
##   Markdown4 + Markdown5 + CPI + Unemployment + Year + Month +
##   Day
##
##
##      Df Sum of Sq      RSS      AIC
## - Markdown4      1 2.3590e+03 1.9978e+14 8421526
## - Markdown2      1 2.4038e+08 1.9978e+14 8421527
## <none>                                1.9978e+14 8421528
## - IsHoliday      1 2.6915e+09 1.9978e+14 8421532
## - Temperature    1 3.8081e+09 1.9978e+14 8421534
## - Day            1 7.6556e+09 1.9979e+14 8421543
## - Markdown1      1 8.6065e+09 1.9979e+14 8421545
## - Fuel_Price     1 1.7279e+10 1.9980e+14 8421563
## - Month          1 5.4956e+10 1.9983e+14 8421642
## - Markdown5      1 5.7094e+10 1.9984e+14 8421647
## - Year           1 5.8515e+10 1.9984e+14 8421650
## - Unemployment   1 8.1070e+10 1.9986e+14 8421697
## - Markdown3      1 1.3933e+11 1.9992e+14 8421820
## - CPI            1 2.5449e+11 2.0003e+14 8422063
## - Type_A         1 4.7964e+11 2.0026e+14 8422537
## - Type_B         1 8.9479e+11 2.0067e+14 8423410
## - Store          1 9.9995e+11 2.0078e+14 8423631
## - Dept           1 2.7154e+12 2.0249e+14 8427218
## - Size           1 4.5457e+12 2.0433e+14 8431011
##
##
## Step: AIC=8421526
## weeklySales ~ Store + Dept + IsHoliday + Type_A + Type_B + Size +
##   Temperature + Fuel_Price + Markdown1 + Markdown2 + Markdown3 +
##   Markdown5 + CPI + Unemployment + Year + Month + Day
##
##
##      Df Sum of Sq      RSS      AIC

```

```

## - Markdown2      1 2.4119e+08 1.9978e+14 8421525
## <none>           1.9978e+14 8421526
## - IsHoliday      1 2.6940e+09 1.9978e+14 8421530
## - Temperature    1 3.8091e+09 1.9978e+14 8421532
## - Day            1 7.6919e+09 1.9979e+14 8421541
## - Fuel_Price     1 1.7283e+10 1.9980e+14 8421561
## - Markdown1      1 2.4390e+10 1.9980e+14 8421576
## - Month          1 5.5281e+10 1.9983e+14 8421641
## - Markdown5      1 5.7151e+10 1.9984e+14 8421645
## - Year           1 5.8992e+10 1.9984e+14 8421649
## - Unemployment   1 8.1138e+10 1.9986e+14 8421696
## - Markdown3      1 1.3935e+11 1.9992e+14 8421818
## - CPI            1 2.5455e+11 2.0003e+14 8422061
## - Type_A         1 4.7992e+11 2.0026e+14 8422536
## - Type_B         1 8.9610e+11 2.0068e+14 8423411
## - Store          1 9.9996e+11 2.0078e+14 8423629
## - Dept           1 2.7154e+12 2.0249e+14 8427216
## - Size           1 4.5471e+12 2.0433e+14 8431012
##
## Step: AIC=8421525
## weeklySales ~ Store + Dept + IsHoliday + Type_A + Type_B + Size +
##   Temperature + Fuel_Price + Markdown1 + Markdown3 + Markdown5 +
##   CPI + Unemployment + Year + Month + Day
##
##           Df  Sum of Sq      RSS      AIC
## <none>           1.9978e+14 8421525
## - IsHoliday      1 3.1333e+09 1.9978e+14 8421530
## - Temperature    1 3.6095e+09 1.9978e+14 8421531
## - Day            1 7.5358e+09 1.9979e+14 8421539
## - Fuel_Price     1 1.7075e+10 1.9980e+14 8421559
## - Markdown1      1 2.4934e+10 1.9980e+14 8421576
## - Month          1 5.5691e+10 1.9984e+14 8421640
## - Markdown5      1 5.7483e+10 1.9984e+14 8421644
## - Year           1 5.8844e+10 1.9984e+14 8421647
## - Unemployment   1 8.0935e+10 1.9986e+14 8421694
## - Markdown3      1 1.3931e+11 1.9992e+14 8421817
## - CPI            1 2.5475e+11 2.0003e+14 8422060
## - Type_A         1 4.8053e+11 2.0026e+14 8422536
## - Type_B         1 8.9693e+11 2.0068e+14 8423411
## - Store          1 1.0021e+12 2.0078e+14 8423632

```

```
## - Dept      1 2.7155e+12 2.0250e+14 8427214
## - Size      1 4.5577e+12 2.0434e+14 8431032
```

```
summary(stepModelAIC)
```

```
##
## Call:
## lm(formula = weeklySales ~ Store + Dept + IsHoliday + Type_A +
##      Type_B + Size + Temperature + Fuel_Price + Markdown1 + Markdown3 +
##      Markdown5 + CPI + Unemployment + Year + Month + Day, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40049 -12790  -6043   5390 673078
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.061e+06  1.837e+05  11.215 < 2e-16 ***
## Store        -1.424e+02  3.096e+00  -45.984 < 2e-16 ***
## Dept         1.038e+02  1.371e+00   75.696 < 2e-16 ***
## IsHoliday    3.592e+02  1.397e+02   2.571  0.01013 *
## Type_A       -5.875e+03  1.845e+02 -31.843 < 2e-16 ***
## Type_B       -6.414e+03  1.474e+02 -43.504 < 2e-16 ***
## Size          9.567e-02  9.756e-04  98.067 < 2e-16 ***
## Temperature  5.692e+00  2.063e+00   2.760  0.00578 **
## Fuel_Price    8.239e+02  1.373e+02   6.002  1.95e-09 ***
## Markdown1     5.053e-02  6.966e-03   7.253  4.07e-13 ***
## Markdown3     1.089e-01  6.353e-03  17.145 < 2e-16 ***
## Markdown5     1.036e-01  9.404e-03  11.013 < 2e-16 ***
## CPI          -2.318e+01  9.997e-01 -23.185 < 2e-16 ***
## Unemployment -2.688e+02  2.057e+01 -13.068 < 2e-16 ***
## Year         -1.020e+03  9.154e+01 -11.143 < 2e-16 ***
## Month         1.257e+02  1.160e+01  10.840 < 2e-16 ***
## Day          -1.552e+01  3.891e+00  -3.988  6.67e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21770 on 421553 degrees of freedom
## Multiple R-squared:  0.08124,    Adjusted R-squared:  0.0812
## F-statistic: 2330 on 16 and 421553 DF,  p-value: < 2.2e-16
```

```
stepModelAIC2 = stepAIC(model, direction = "backward", k=log(nrow(dataset)))
```

```

## Start: AIC=8421736
## weeklySales ~ Store + Dept + IsHoliday + Type_A + Type_B + Type_C +
##   Size + Temperature + Fuel_Price + Markdown1 + Markdown2 +
##   Markdown3 + Markdown4 + Markdown5 + CPI + Unemployment +
##   Year + Month + Day
##
##
## Step: AIC=8421736
## weeklySales ~ Store + Dept + IsHoliday + Type_A + Type_B + Size +
##   Temperature + Fuel_Price + Markdown1 + Markdown2 + Markdown3 +
##   Markdown4 + Markdown5 + CPI + Unemployment + Year + Month +
##   Day
##
##
##      Df Sum of Sq      RSS      AIC
## - Markdown4    1 2.3590e+03 1.9978e+14 8421724
## - Markdown2    1 2.4038e+08 1.9978e+14 8421724
## - IsHoliday    1 2.6915e+09 1.9978e+14 8421729
## - Temperature  1 3.8081e+09 1.9978e+14 8421732
## <none>                                1.9978e+14 8421736
## - Day          1 7.6556e+09 1.9979e+14 8421740
## - Markdown1    1 8.6065e+09 1.9979e+14 8421742
## - Fuel_Price   1 1.7279e+10 1.9980e+14 8421760
## - Month        1 5.4956e+10 1.9983e+14 8421839
## - Markdown5    1 5.7094e+10 1.9984e+14 8421844
## - Year         1 5.8515e+10 1.9984e+14 8421847
## - Unemployment 1 8.1070e+10 1.9986e+14 8421895
## - Markdown3    1 1.3933e+11 1.9992e+14 8422017
## - CPI          1 2.5449e+11 2.0003e+14 8422260
## - Type_A       1 4.7964e+11 2.0026e+14 8422734
## - Type_B       1 8.9479e+11 2.0067e+14 8423607
## - Store        1 9.9995e+11 2.0078e+14 8423828
## - Dept         1 2.7154e+12 2.0249e+14 8427415
## - Size         1 4.5457e+12 2.0433e+14 8431208
##
## Step: AIC=8421724
## weeklySales ~ Store + Dept + IsHoliday + Type_A + Type_B + Size +
##   Temperature + Fuel_Price + Markdown1 + Markdown2 + Markdown3 +
##   Markdown5 + CPI + Unemployment + Year + Month + Day
##
##
##      Df Sum of Sq      RSS      AIC

```

```

## - Markdown2      1 2.4119e+08 1.9978e+14 8421711
## - IsHoliday      1 2.6940e+09 1.9978e+14 8421716
## - Temperature    1 3.8091e+09 1.9978e+14 8421719
## <none>           1.9978e+14 8421724
## - Day            1 7.6919e+09 1.9979e+14 8421727
## - Fuel_Price     1 1.7283e+10 1.9980e+14 8421747
## - Markdown1      1 2.4390e+10 1.9980e+14 8421762
## - Month          1 5.5281e+10 1.9983e+14 8421827
## - Markdown5      1 5.7151e+10 1.9984e+14 8421831
## - Year           1 5.8992e+10 1.9984e+14 8421835
## - Unemployment   1 8.1138e+10 1.9986e+14 8421882
## - Markdown3      1 1.3935e+11 1.9992e+14 8422005
## - CPI            1 2.5455e+11 2.0003e+14 8422247
## - Type_A         1 4.7992e+11 2.0026e+14 8422722
## - Type_B         1 8.9610e+11 2.0068e+14 8423597
## - Store          1 9.9996e+11 2.0078e+14 8423815
## - Dept           1 2.7154e+12 2.0249e+14 8427402
## - Size           1 4.5471e+12 2.0433e+14 8431198
##
## Step: AIC=8421711
## weeklySales ~ Store + Dept + IsHoliday + Type_A + Type_B + Size +
##   Temperature + Fuel_Price + Markdown1 + Markdown3 + Markdown5 +
##   CPI + Unemployment + Year + Month + Day
##
##              Df  Sum of Sq      RSS      AIC
## - IsHoliday    1 3.1333e+09 1.9978e+14 8421705
## - Temperature  1 3.6095e+09 1.9978e+14 8421706
## <none>          1.9978e+14 8421711
## - Day          1 7.5358e+09 1.9979e+14 8421714
## - Fuel_Price   1 1.7075e+10 1.9980e+14 8421734
## - Markdown1    1 2.4934e+10 1.9980e+14 8421751
## - Month        1 5.5691e+10 1.9984e+14 8421816
## - Markdown5    1 5.7483e+10 1.9984e+14 8421819
## - Year         1 5.8844e+10 1.9984e+14 8421822
## - Unemployment 1 8.0935e+10 1.9986e+14 8421869
## - Markdown3    1 1.3931e+11 1.9992e+14 8421992
## - CPI          1 2.5475e+11 2.0003e+14 8422235
## - Type_A       1 4.8053e+11 2.0026e+14 8422711
## - Type_B       1 8.9693e+11 2.0068e+14 8423587
## - Store        1 1.0021e+12 2.0078e+14 8423807
## - Dept         1 2.7155e+12 2.0250e+14 8427390

```

```

## - Size          1 4.5577e+12 2.0434e+14 8431208
##
## Step: AIC=8421705
## weeklySales ~ Store + Dept + Type_A + Type_B + Size + Temperature +
##      Fuel_Price + Markdown1 + Markdown3 + Markdown5 + CPI + Unemployment +
##      Year + Month + Day
##
##              Df Sum of Sq      RSS      AIC
## - Temperature  1 2.6061e+09 1.9979e+14 8421697
## <none>                1.9978e+14 8421705
## - Day          1 7.1959e+09 1.9979e+14 8421707
## - Fuel_Price   1 1.6282e+10 1.9980e+14 8421726
## - Markdown1    1 2.5494e+10 1.9981e+14 8421746
## - Markdown5    1 5.6354e+10 1.9984e+14 8421811
## - Year         1 5.7883e+10 1.9984e+14 8421814
## - Month        1 6.1045e+10 1.9984e+14 8421821
## - Unemployment 1 7.9521e+10 1.9986e+14 8421860
## - Markdown3    1 1.5935e+11 1.9994e+14 8422028
## - CPI          1 2.5369e+11 2.0004e+14 8422227
## - Type_A       1 4.8099e+11 2.0026e+14 8422706
## - Type_B       1 8.9975e+11 2.0068e+14 8423586
## - Store        1 1.0044e+12 2.0079e+14 8423806
## - Dept         1 2.7155e+12 2.0250e+14 8427383
## - Size         1 4.5551e+12 2.0434e+14 8431196
##
## Step: AIC=8421697
## weeklySales ~ Store + Dept + Type_A + Type_B + Size + Fuel_Price +
##      Markdown1 + Markdown3 + Markdown5 + CPI + Unemployment +
##      Year + Month + Day
##
##              Df Sum of Sq      RSS      AIC
## <none>                1.9979e+14 8421697
## - Day          1 7.0267e+09 1.9979e+14 8421699
## - Fuel_Price   1 1.8865e+10 1.9980e+14 8421724
## - Markdown1    1 2.5395e+10 1.9981e+14 8421738
## - Markdown5    1 5.5416e+10 1.9984e+14 8421801
## - Year         1 5.8515e+10 1.9984e+14 8421808
## - Month        1 7.2113e+10 1.9986e+14 8421836
## - Unemployment 1 7.6932e+10 1.9986e+14 8421847
## - Markdown3    1 1.5698e+11 1.9994e+14 8422015
## - CPI          1 2.5542e+11 2.0004e+14 8422223

```

```
## - Type_A      1 4.8500e+11 2.0027e+14 8422706
## - Type_B      1 9.3108e+11 2.0072e+14 8423644
## - Store       1 1.0360e+12 2.0082e+14 8423865
## - Dept        1 2.7155e+12 2.0250e+14 8427376
## - Size        1 4.5596e+12 2.0434e+14 8431198
```

```
model2 = lm(weeklySales~Store,data=dataset)
summary(model2)
```

```
##
## Call:
## lm(formula = weeklySales ~ Store, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20136 -13343  -8334   4339 675272
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19341.006     69.835   276.95  <2e-16 ***
## Store       -151.336      2.726  -55.52  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22630 on 421568 degrees of freedom
## Multiple R-squared:  0.007258,    Adjusted R-squared:  0.007256
## F-statistic: 3082 on 1 and 421568 DF,  p-value: < 2.2e-16
```

```
stepModelAIC_forward = stepAIC(model2,direction = "forward",scope=list(lower=model2, upper=~Store+Dept+IsHoliday+Type_A+Type_B+Type_C+Size+Temperature+Fuel_Price+MarkDown1+MarkDown2+MarkDown3+MarkDown4+MarkDown5+CPI+Unemployment+Year+Month+Day))
```



```
## Start: AIC=8454143
## weeklySales ~ Store
##
##           Df  Sum of Sq      RSS      AIC
## + Size      1 1.1720e+13 2.0415e+14 8430611
## + Type_A     1 7.2328e+12 2.0863e+14 8439778
## + Type_B     1 5.2513e+12 2.1061e+14 8443763
## + Dept       1 2.8879e+12 2.1298e+14 8448467
## + Type_C     1 9.0571e+11 2.1496e+14 8452372
## + Markdown5  1 5.3082e+11 2.1534e+14 8453107
## + Markdown1  1 3.8631e+11 2.1548e+14 8453390
## + CPI        1 3.4447e+11 2.1552e+14 8453472
## + Markdown3  1 2.9507e+11 2.1557e+14 8453568
## + Markdown4  1 2.4927e+11 2.1562e+14 8453658
## + Month      1 1.7656e+11 2.1569e+14 8453800
## + Markdown2  1 6.9241e+10 2.1580e+14 8454010
## + IsHoliday  1 3.5224e+10 2.1583e+14 8454076
## + Year       1 2.1123e+10 2.1585e+14 8454104
## + Unemployment 1 1.4901e+10 2.1585e+14 8454116
## + Temperature 1 9.4396e+09 2.1586e+14 8454126
## + Day        1 8.3272e+09 2.1586e+14 8454129
## + Fuel_Price  1 6.4675e+09 2.1586e+14 8454132
## <none>                2.1587e+14 8454143
##
## Step: AIC=8430611
## weeklySales ~ Store + Size
##
##           Df  Sum of Sq      RSS      AIC
## + Dept      1 2.7922e+12 2.0135e+14 8424808
## + Type_C     1 8.3414e+11 2.0331e+14 8428887
## + Type_B     1 3.5287e+11 2.0379e+14 8429884
## + Markdown3  1 1.9289e+11 2.0395e+14 8430215
## + CPI        1 1.9184e+11 2.0395e+14 8430217
## + Month      1 1.7958e+11 2.0397e+14 8430242
## + Markdown5  1 4.2472e+10 2.0410e+14 8430526
## + IsHoliday  1 3.4583e+10 2.0411e+14 8430542
## + Temperature 1 1.9162e+10 2.0413e+14 8430574
## + Year       1 1.6880e+10 2.0413e+14 8430579
## + Day        1 8.0823e+09 2.0414e+14 8430597
## + Type_A     1 7.2342e+09 2.0414e+14 8430599
```

```

## + Markdown4      1 6.9420e+09 2.0414e+14 8430599
## + Markdown1      1 4.6676e+09 2.0414e+14 8430604
## <none>           2.0415e+14 8430611
## + Fuel_Price     1 7.3046e+08 2.0415e+14 8430612
## + Unemployment   1 2.2242e+08 2.0415e+14 8430613
## + Markdown2      1 1.3595e+08 2.0415e+14 8430613
##
## Step:  AIC=8424808
## weeklySales ~ Store + Size + Dept
##
##              Df  Sum of Sq      RSS      AIC
## + Type_C      1 7.7838e+11 2.0058e+14 8423177
## + Type_B      1 3.1062e+11 2.0104e+14 8424159
## + Markdown3   1 1.9018e+11 2.0116e+14 8424411
## + CPI         1 1.8629e+11 2.0117e+14 8424419
## + Month       1 1.7853e+11 2.0118e+14 8424436
## + Markdown5   1 4.0749e+10 2.0131e+14 8424724
## + IsHoliday   1 3.4075e+10 2.0132e+14 8424738
## + Year        1 1.8250e+10 2.0134e+14 8424771
## + Temperature 1 1.7500e+10 2.0134e+14 8424773
## + Day         1 7.8943e+09 2.0135e+14 8424793
## + Markdown4   1 6.1959e+09 2.0135e+14 8424797
## + Markdown1   1 4.0625e+09 2.0135e+14 8424801
## + Type_A      1 3.7441e+09 2.0135e+14 8424802
## <none>         2.0135e+14 8424808
## + Fuel_Price   1 5.6712e+08 2.0135e+14 8424808
## + Unemployment 1 3.4197e+08 2.0135e+14 8424809
## + Markdown2   1 8.9540e+07 2.0135e+14 8424809
##
## Step:  AIC=8423177
## weeklySales ~ Store + Size + Dept + Type_C
##
##              Df  Sum of Sq      RSS      AIC
## + CPI         1 2.6156e+11 2.0031e+14 8422629
## + Markdown3   1 1.9548e+11 2.0038e+14 8422768
## + Month       1 1.7951e+11 2.0040e+14 8422801
## + Markdown5   1 4.8975e+10 2.0053e+14 8423076
## + IsHoliday   1 3.4311e+10 2.0054e+14 8423107
## + Year        1 1.8817e+10 2.0056e+14 8423139
## + Markdown1   1 1.1092e+10 2.0056e+14 8423155
## + Unemployment 1 1.0044e+10 2.0057e+14 8423158

```

```

## + Markdown4      1 1.0020e+10 2.0057e+14 8423158
## + Day            1 7.8411e+09 2.0057e+14 8423162
## + Fuel_Price     1 2.0671e+09 2.0057e+14 8423174
## <none>           2.0058e+14 8423177
## + Type_A         1 3.1620e+08 2.0058e+14 8423178
## + Type_B         1 3.1620e+08 2.0058e+14 8423178
## + Markdown2      1 2.7323e+08 2.0058e+14 8423178
## + Temperature    1 1.5281e+08 2.0058e+14 8423178
##
## Step: AIC=8422629
## weeklySales ~ Store + Size + Dept + Type_C + CPI
##
##              Df Sum of Sq      RSS      AIC
## + Markdown3    1 1.9168e+11 2.0012e+14 8422227
## + Month         1 1.8201e+11 2.0013e+14 8422247
## + Markdown5     1 6.8274e+10 2.0025e+14 8422487
## + Unemployment  1 6.4252e+10 2.0025e+14 8422495
## + IsHoliday     1 3.3928e+10 2.0028e+14 8422559
## + Markdown1     1 1.2149e+10 2.0030e+14 8422605
## + Type_A        1 1.1418e+10 2.0030e+14 8422607
## + Type_B        1 1.1418e+10 2.0030e+14 8422607
## + Year          1 9.6319e+09 2.0030e+14 8422610
## + Markdown4     1 9.6198e+09 2.0030e+14 8422610
## + Day           1 7.5871e+09 2.0031e+14 8422615
## + Temperature   1 5.0773e+09 2.0031e+14 8422620
## + Fuel_Price    1 1.0722e+09 2.0031e+14 8422628
## <none>          2.0031e+14 8422629
## + Markdown2     1 1.6654e+08 2.0031e+14 8422630
##
## Step: AIC=8422227
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3
##
##              Df Sum of Sq      RSS      AIC
## + Month         1 1.4312e+11 1.9998e+14 8421927
## + Unemployment  1 6.0843e+10 2.0006e+14 8422101
## + Markdown5     1 5.9807e+10 2.0006e+14 8422103
## + Markdown1     1 1.4281e+10 2.0011e+14 8422199
## + Day           1 1.4172e+10 2.0011e+14 8422199
## + Type_A        1 1.2639e+10 2.0011e+14 8422202
## + Type_B        1 1.2639e+10 2.0011e+14 8422202
## + Markdown4     1 1.1134e+10 2.0011e+14 8422206

```

```

## + Year      1 1.0304e+10 2.0011e+14 8422207
## + Temperature 1 9.0458e+09 2.0011e+14 8422210
## + IsHoliday  1 4.9000e+09 2.0012e+14 8422219
## + Fuel_Price 1 1.6611e+09 2.0012e+14 8422226
## <none>      2.0012e+14 8422227
## + Markdown2  1 2.8611e+08 2.0012e+14 8422228
##
## Step: AIC=8421927
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
##      Month
##
##           Df Sum of Sq      RSS      AIC
## + Unemployment 1 5.8991e+10 1.9992e+14 8421805
## + Markdown5     1 5.0672e+10 1.9993e+14 8421823
## + Markdown1     1 2.3615e+10 1.9996e+14 8421880
## + Markdown4     1 2.1340e+10 1.9996e+14 8421884
## + Day           1 1.4905e+10 1.9996e+14 8421898
## + Type_A        1 1.2474e+10 1.9997e+14 8421903
## + Type_B        1 1.2474e+10 1.9997e+14 8421903
## + IsHoliday     1 1.1315e+09 1.9998e+14 8421927
## <none>          1.9998e+14 8421927
## + Year          1 7.4565e+08 1.9998e+14 8421928
## + Fuel_Price    1 5.9911e+08 1.9998e+14 8421928
## + Markdown2     1 5.6925e+08 1.9998e+14 8421928
## + Temperature   1 3.8023e+04 1.9998e+14 8421929
##
## Step: AIC=8421805
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
##      Month + Unemployment
##
##           Df Sum of Sq      RSS      AIC
## + Markdown5     1 4.0965e+10 1.9988e+14 8421721
## + Markdown1     1 1.7534e+10 1.9990e+14 8421770
## + Markdown4     1 1.6774e+10 1.9990e+14 8421772
## + Day           1 1.5024e+10 1.9990e+14 8421775
## + Type_A        1 9.5159e+09 1.9991e+14 8421787
## + Type_B        1 9.5159e+09 1.9991e+14 8421787
## + Year          1 7.7631e+09 1.9991e+14 8421791
## + Fuel_Price    1 2.2526e+09 1.9992e+14 8421802
## + Temperature   1 1.4538e+09 1.9992e+14 8421804
## + IsHoliday     1 1.4192e+09 1.9992e+14 8421804

```

```

## <none> 1.9992e+14 8421805
## + Markdown2 1 2.2782e+08 1.9992e+14 8421807
##
## Step: AIC=8421721
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
## Month + Unemployment + Markdown5
##
##      Df Sum of Sq      RSS      AIC
## + Year      1 3.5798e+10 1.9984e+14 8421647
## + Day        1 1.2200e+10 1.9987e+14 8421697
## + Type_A      1 1.0269e+10 1.9987e+14 8421701
## + Type_B      1 1.0269e+10 1.9987e+14 8421701
## + Fuel_Price  1 9.1878e+09 1.9987e+14 8421703
## + Markdown4   1 5.3477e+09 1.9987e+14 8421711
## + Markdown1   1 3.0877e+09 1.9988e+14 8421716
## + IsHoliday   1 1.9237e+09 1.9988e+14 8421719
## + Temperature 1 1.6292e+09 1.9988e+14 8421719
## <none> 1.9988e+14 8421721
## + Markdown2   1 9.1165e+07 1.9988e+14 8421723
##
## Step: AIC=8421647
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
## Month + Unemployment + Markdown5 + Year
##
##      Df Sum of Sq      RSS      AIC
## + Markdown1   1 2.0551e+10 1.9982e+14 8421606
## + Markdown4   1 1.4316e+10 1.9983e+14 8421619
## + Day          1 1.0793e+10 1.9983e+14 8421626
## + Fuel_Price   1 1.0387e+10 1.9983e+14 8421627
## + Type_A       1 9.4997e+09 1.9983e+14 8421629
## + Type_B       1 9.4997e+09 1.9983e+14 8421629
## + Temperature  1 5.5424e+09 1.9984e+14 8421638
## + IsHoliday    1 1.4553e+09 1.9984e+14 8421646
## <none> 1.9984e+14 8421647
## + Markdown2   1 4.5224e+07 1.9984e+14 8421649
##
## Step: AIC=8421606
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
## Month + Unemployment + Markdown5 + Year + Markdown1
##
##      Df Sum of Sq      RSS      AIC

```

```

## + Fuel_Price    1 1.6682e+10 1.9981e+14 8421573
## + Type_A        1 1.1472e+10 1.9981e+14 8421584
## + Type_B        1 1.1472e+10 1.9981e+14 8421584
## + Day           1 7.2010e+09 1.9982e+14 8421593
## + Temperature   1 6.5431e+09 1.9982e+14 8421594
## + IsHoliday     1 1.0299e+09 1.9982e+14 8421606
## <none>          1.9982e+14 8421606
## + Markdown2     1 5.1139e+07 1.9982e+14 8421608
## + Markdown4     1 2.6196e+07 1.9982e+14 8421608
##
## Step: AIC=8421573
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
##      Month + Unemployment + Markdown5 + Year + Markdown1 + Fuel_Price
##
##              Df Sum of Sq      RSS      AIC
## + Type_A      1 1.3659e+10 1.9979e+14 8421546
## + Type_B      1 1.3659e+10 1.9979e+14 8421546
## + Day         1 7.2132e+09 1.9980e+14 8421559
## + Temperature 1 3.8906e+09 1.9980e+14 8421566
## + IsHoliday   1 1.8858e+09 1.9980e+14 8421571
## <none>        1.9981e+14 8421573
## + Markdown2   1 5.2024e+07 1.9981e+14 8421575
## + Markdown4   1 5.0100e+07 1.9981e+14 8421575
##
## Step: AIC=8421546
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
##      Month + Unemployment + Markdown5 + Year + Markdown1 + Fuel_Price +
##      Type_A
##
##              Df Sum of Sq      RSS      AIC
## + Day         1 7026704801 1.9979e+14 8421533
## + Temperature 1 2436958306 1.9979e+14 8421543
## + IsHoliday   1 1884973689 1.9979e+14 8421544
## <none>        1.9979e+14 8421546
## + Markdown2   1 105162361 1.9979e+14 8421548
## + Markdown4   1 33509426 1.9979e+14 8421548
##
## Step: AIC=8421533
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
##      Month + Unemployment + Markdown5 + Year + Markdown1 + Fuel_Price +
##      Type_A + Day

```

```
##
##           Df Sum of Sq      RSS      AIC
## + Temperature  1 2606113653 1.9978e+14 8421530
## + IsHoliday    1 2129887598 1.9978e+14 8421531
## <none>                1.9979e+14 8421533
## + Markdown2    1  269591938 1.9979e+14 8421534
## + Markdown4    1    96234 1.9979e+14 8421535
##
## Step:  AIC=8421530
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
##      Month + Unemployment + Markdown5 + Year + Markdown1 + Fuel_Price +
##      Type_A + Day + Temperature
##
##           Df Sum of Sq      RSS      AIC
## + IsHoliday  1 3133319107 1.9978e+14 8421525
## <none>                1.9978e+14 8421530
## + Markdown2  1  680496496 1.9978e+14 8421530
## + Markdown4  1    88383 1.9978e+14 8421532
##
## Step:  AIC=8421525
## weeklySales ~ Store + Size + Dept + Type_C + CPI + Markdown3 +
##      Month + Unemployment + Markdown5 + Year + Markdown1 + Fuel_Price +
##      Type_A + Day + Temperature + IsHoliday
##
##           Df Sum of Sq      RSS      AIC
## <none>                1.9978e+14 8421525
## + Markdown2  1 241185510 1.9978e+14 8421526
## + Markdown4  1   805326 1.9978e+14 8421527
```

```
finalModel = lm(formula = weeklySales ~ Store + Dept + IsHoliday + Type_A +
  Type_B + Size + Temperature + Fuel_Price + Markdown1 + Markdown3 +
  Markdown5 + CPI + Unemployment + Year + Month + Day, data = dataset)
summary(finalModel)
```

```
##
## Call:
## lm(formula = weeklySales ~ Store + Dept + IsHoliday + Type_A +
##      Type_B + Size + Temperature + Fuel_Price + Markdown1 + Markdown3 +
##      Markdown5 + CPI + Unemployment + Year + Month + Day, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40049 -12790  -6043   5390 673078
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.061e+06  1.837e+05  11.215 < 2e-16 ***
## Store        -1.424e+02  3.096e+00  -45.984 < 2e-16 ***
## Dept         1.038e+02  1.371e+00   75.696 < 2e-16 ***
## IsHoliday    3.592e+02  1.397e+02   2.571  0.01013 *
## Type_A       -5.875e+03  1.845e+02 -31.843 < 2e-16 ***
## Type_B       -6.414e+03  1.474e+02 -43.504 < 2e-16 ***
## Size          9.567e-02  9.756e-04  98.067 < 2e-16 ***
## Temperature  5.692e+00  2.063e+00   2.760  0.00578 **
## Fuel_Price    8.239e+02  1.373e+02   6.002  1.95e-09 ***
## Markdown1     5.053e-02  6.966e-03   7.253  4.07e-13 ***
## Markdown3     1.089e-01  6.353e-03  17.145 < 2e-16 ***
## Markdown5     1.036e-01  9.404e-03  11.013 < 2e-16 ***
## CPI          -2.318e+01  9.997e-01 -23.185 < 2e-16 ***
## Unemployment -2.688e+02  2.057e+01 -13.068 < 2e-16 ***
## Year         -1.020e+03  9.154e+01 -11.143 < 2e-16 ***
## Month         1.257e+02  1.160e+01  10.840 < 2e-16 ***
## Day          -1.552e+01  3.891e+00  -3.988  6.67e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21770 on 421553 degrees of freedom
## Multiple R-squared:  0.08124,    Adjusted R-squared:  0.0812
## F-statistic: 2330 on 16 and 421553 DF,  p-value: < 2.2e-16
```



```
finalModel2 = lm(formula = weeklySales ~ Store + Dept + Type_A + Type_B + Size + Fuel_Price +
  Markdown1 + Markdown3 + Markdown5 + CPI + Unemployment +
  Year + Month + Day, data = dataset)
summary(finalModel2)
```

```
##
## Call:
## lm(formula = weeklySales ~ Store + Dept + Type_A + Type_B + Size +
##     Fuel_Price + Markdown1 + Markdown3 + Markdown5 + CPI + Unemployment +
##     Year + Month + Day, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40149 -12794  -6045   5389  673360
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.053e+06  1.836e+05  11.184 < 2e-16 ***
## Store       -1.435e+02  3.069e+00  -46.755 < 2e-16 ***
## Dept         1.038e+02  1.371e+00   75.696 < 2e-16 ***
## Type_A      -5.897e+03  1.843e+02  -31.990 < 2e-16 ***
## Type_B      -6.470e+03  1.460e+02  -44.324 < 2e-16 ***
## Size         9.549e-02  9.735e-04   98.087 < 2e-16 ***
## Fuel_Price   8.538e+02  1.353e+02   6.309 2.81e-10 ***
## Markdown1    5.097e-02  6.963e-03   7.320 2.48e-13 ***
## Markdown3    1.116e-01  6.134e-03  18.200 < 2e-16 ***
## Markdown5    1.015e-01  9.385e-03  10.813 < 2e-16 ***
## CPI         -2.259e+01  9.731e-01  -23.215 < 2e-16 ***
## Unemployment -2.581e+02  2.026e+01  -12.741 < 2e-16 ***
## Year        -1.016e+03  9.144e+01  -11.112 < 2e-16 ***
## Month        1.369e+02  1.110e+01  12.335 < 2e-16 ***
## Day         -1.497e+01  3.888e+00   -3.851 0.000118 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21770 on 421555 degrees of freedom
## Multiple R-squared:  0.08121,    Adjusted R-squared:  0.08118
## F-statistic: 2662 on 14 and 421555 DF,  p-value: < 2.2e-16
```

```
anova(finalModel, finalModel2)
```

```
## Analysis of Variance Table
##
## Model 1: weeklySales ~ Store + Dept + IsHoliday + Type_A + Type_B + Size +
##   Temperature + Fuel_Price + Markdown1 + Markdown3 + Markdown5 +
##   CPI + Unemployment + Year + Month + Day
## Model 2: weeklySales ~ Store + Dept + Type_A + Type_B + Size + Fuel_Price +
##   Markdown1 + Markdown3 + Markdown5 + CPI + Unemployment +
##   Year + Month + Day
##   Res.Df      RSS Df Sum of Sq    F Pr(>F)
## 1 421553 1.9978e+14
## 2 421555 1.9979e+14 -2 -5739432760 6.0554 0.002345 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

FinalModel is based upon the results of setpAIC without log whereas the FinalModel2 is based on logs. When we compare F-Statistic value we see FinalModel2 is better and has less variables so we go ahead with finalModel2.

```
confint(finalModel2, level=0.95)
```

```
##           2.5 %      97.5 %
## (Intercept)  1.693171e+06  2.412691e+06
## Store       -1.494945e+02 -1.374651e+02
## Dept        1.010714e+02  1.064445e+02
## Type_A      -6.257810e+03 -5.535270e+03
## Type_B      -6.756434e+03 -6.184209e+03
## Size        9.358250e-02  9.739868e-02
## Fuel_Price   5.885366e+02  1.118974e+03
## Markdown1    3.732249e-02  6.461654e-02
## Markdown3    9.960888e-02  1.236525e-01
## Markdown5    8.309265e-02  1.198828e-01
## CPI         -2.449684e+01 -2.068250e+01
## Unemployment -2.977840e+02 -2.183804e+02
## Year        -1.195309e+03 -8.368585e+02
## Month        1.151441e+02  1.586469e+02
## Day         -2.258912e+01 -7.349830e+00
```

```
anova(finalModel2)
```

```
## Analysis of Variance Table
##
## Response: weeklySales
##
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
## Store	1	1.5783e+12	1.5783e+12	3330.1743	< 2.2e-16	***
## Dept	1	2.8879e+12	2.8879e+12	6093.5075	< 2.2e-16	***
## Type_A	1	7.0934e+12	7.0934e+12	14967.3693	< 2.2e-16	***
## Type_B	1	8.4366e+06	8.4366e+06	0.0178	0.8938592	
## Size	1	5.3098e+12	5.3098e+12	11203.8944	< 2.2e-16	***
## Fuel_Price	1	2.1998e+09	2.1998e+09	4.6416	0.0312064	*
## Markdown1	1	9.4115e+09	9.4115e+09	19.8586	8.341e-06	***
## Markdown3	1	1.9770e+11	1.9770e+11	417.1556	< 2.2e-16	***
## Markdown5	1	2.9944e+10	2.9944e+10	63.1835	1.888e-15	***
## CPI	1	2.9882e+11	2.9882e+11	630.5167	< 2.2e-16	***
## Unemployment	1	4.9724e+10	4.9724e+10	104.9189	< 2.2e-16	***
## Year	1	1.2349e+11	1.2349e+11	260.5617	< 2.2e-16	***
## Month	1	7.1403e+10	7.1403e+10	150.6624	< 2.2e-16	***
## Day	1	7.0267e+09	7.0267e+09	14.8266	0.0001179	***
## Residuals	421555	1.9979e+14	4.7392e+08			
## ---						

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
vcov(finalModel2)
```

```

##          (Intercept)          Store          Dept          Type_A
## (Intercept)  3.369204e+10 -4.417228e+03  2.612390e+02 -6.514329e+05
## Store       -4.417228e+03  9.417318e+00 -2.325496e-02  1.176382e+02
## Dept        2.612390e+02 -2.325496e-02  1.878874e+00  1.529770e+00
## Type_A      -6.514329e+05  1.176382e+02  1.529770e+00  3.397550e+04
## Type_B      -1.925183e+06  1.739275e+02  3.696822e+00  2.185204e+04
## Size        -2.044463e+01  2.565914e-04 -1.090940e-07 -1.270913e-01
## Fuel_Price   2.013814e+07 -1.295611e+01  1.268995e-01  8.861375e+01
## Markdown1    4.848389e+02  2.819733e-04 -2.462896e-05  5.703316e-05
## Markdown3    2.123614e+01  2.136255e-04 -1.679348e-05  5.229726e-03
## Markdown5    5.074525e+02 -1.288895e-03 -1.443512e-05 -2.897863e-02
## CPI          4.337495e+04  6.009657e-01  1.011550e-02  3.687422e-01
## Unemployment -1.034956e+06 -5.727717e+00 -1.913184e-02  4.961846e+02
## Year         -1.678464e+07  2.006077e+00 -1.682586e-01  3.162371e+02
## Month        -6.580002e+05  7.512584e-03 -1.846031e-02  1.662935e+01
## Day          3.666027e+04 -1.495636e-02  1.361997e-03 -1.126594e+00
##          Type_B          Size          Fuel_Price          Markdown1
## (Intercept) -1.925183e+06 -2.044463e+01  2.013814e+07  4.848389e+02
## Store        1.739275e+02  2.565914e-04 -1.295611e+01  2.819733e-04
## Dept         3.696822e+00 -1.090940e-07  1.268995e-01 -2.462896e-05
## Type_A       2.185204e+04 -1.270913e-01  8.861375e+01  5.703316e-05
## Type_B       2.130960e+04 -5.013086e-02 -9.768694e+02 -5.852475e-02
## Size        -5.013086e-02  9.477651e-07 -1.188718e-02 -9.217885e-07
## Fuel_Price   -9.768694e+02 -1.188718e-02  1.831089e+04  1.686138e-01
## Markdown1    -5.852475e-02 -9.217885e-07  1.686138e-01  4.848163e-05
## Markdown3    -4.602696e-03 -1.576063e-07  4.608419e-03  1.215414e-06
## Markdown5    -3.383026e-02 -7.271202e-07  2.165132e-01 -1.324183e-05
## CPI          1.274779e+01  6.494699e-05  3.929365e+01  5.258729e-04
## Unemployment  3.960356e+02 -7.568735e-04 -5.190184e+02 -9.380040e-03
## Year          9.481101e+02  1.015991e-02 -1.004414e+04 -2.413893e-01
## Month         3.343502e+01  3.621898e-04 -3.686108e+02 -2.046053e-03
## Day          -5.095799e+00 -8.348029e-05  7.295294e-02  3.696690e-03
##          Markdown3          Markdown5          CPI          Unemployment
## (Intercept)  2.123614e+01  5.074525e+02  4.337495e+04 -1.034956e+06
## Store        2.136255e-04 -1.288895e-03  6.009657e-01 -5.727717e+00
## Dept        -1.679348e-05 -1.443512e-05  1.011550e-02 -1.913184e-02
## Type_A       5.229726e-03 -2.897863e-02  3.687422e-01  4.961846e+02
## Type_B      -4.602696e-03 -3.383026e-02  1.274779e+01  3.960356e+02
## Size        -1.576063e-07 -7.271202e-07  6.494699e-05 -7.568735e-04
## Fuel_Price   4.608419e-03  2.165132e-01  3.929365e+01 -5.190184e+02

```

```
## Markdown1      1.215414e-06 -1.324183e-05  5.258729e-04 -9.380040e-03
## Markdown3      3.762177e-05 -1.757272e-06  8.481479e-05  7.249009e-04
## Markdown5     -1.757272e-06  8.808558e-05  1.181797e-05 -5.420574e-03
## CPI            8.481479e-05  1.181797e-05  9.468498e-01  3.599661e+00
## Unemployment   7.249009e-04 -5.420574e-03  3.599661e+00  4.103184e+02
## Year          -1.054189e-02 -2.526094e-01 -2.174167e+01  5.134507e+02
## Month         -7.522236e-03 -1.914220e-02 -7.762165e-01  2.292037e+01
## Day           -1.611988e-03  1.217049e-03 -8.019315e-03 -3.599031e-01
##              Year           Month           Day
## (Intercept) -1.678464e+07 -6.580002e+05  3.666027e+04
## Store        2.006077e+00  7.512584e-03 -1.495636e-02
## Dept        -1.682586e-01 -1.846031e-02  1.361997e-03
## Type_A       3.162371e+02  1.662935e+01 -1.126594e+00
## Type_B       9.481101e+02  3.343502e+01 -5.095799e+00
## Size         1.015991e-02  3.621898e-04 -8.348029e-05
## Fuel_Price  -1.004414e+04 -3.686108e+02  7.295294e-02
## Markdown1    -2.413893e-01 -2.046053e-03  3.696690e-03
## Markdown3    -1.054189e-02 -7.522236e-03 -1.611988e-03
## Markdown5    -2.526094e-01 -1.914220e-02  1.217049e-03
## CPI          -2.174167e+01 -7.762165e-01 -8.019315e-03
## Unemployment  5.134507e+02  2.292037e+01 -3.599031e-01
## Year         8.361795e+03  3.273927e+02 -1.834203e+01
## Month        3.273927e+02  1.231616e+02 -7.003441e-01
## Day         -1.834203e+01 -7.003441e-01  1.511370e+01
```

```
cov2cor(vcov(finalModel2))
```

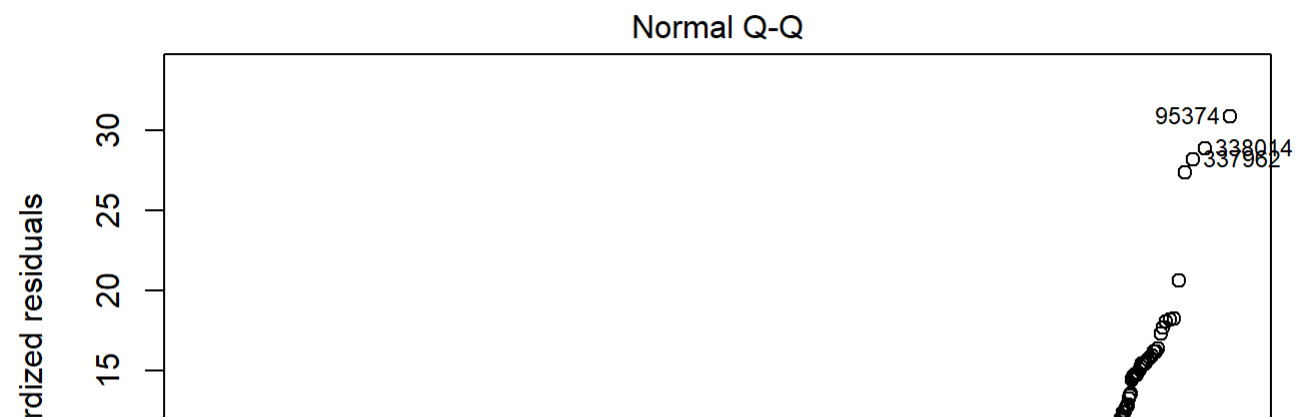
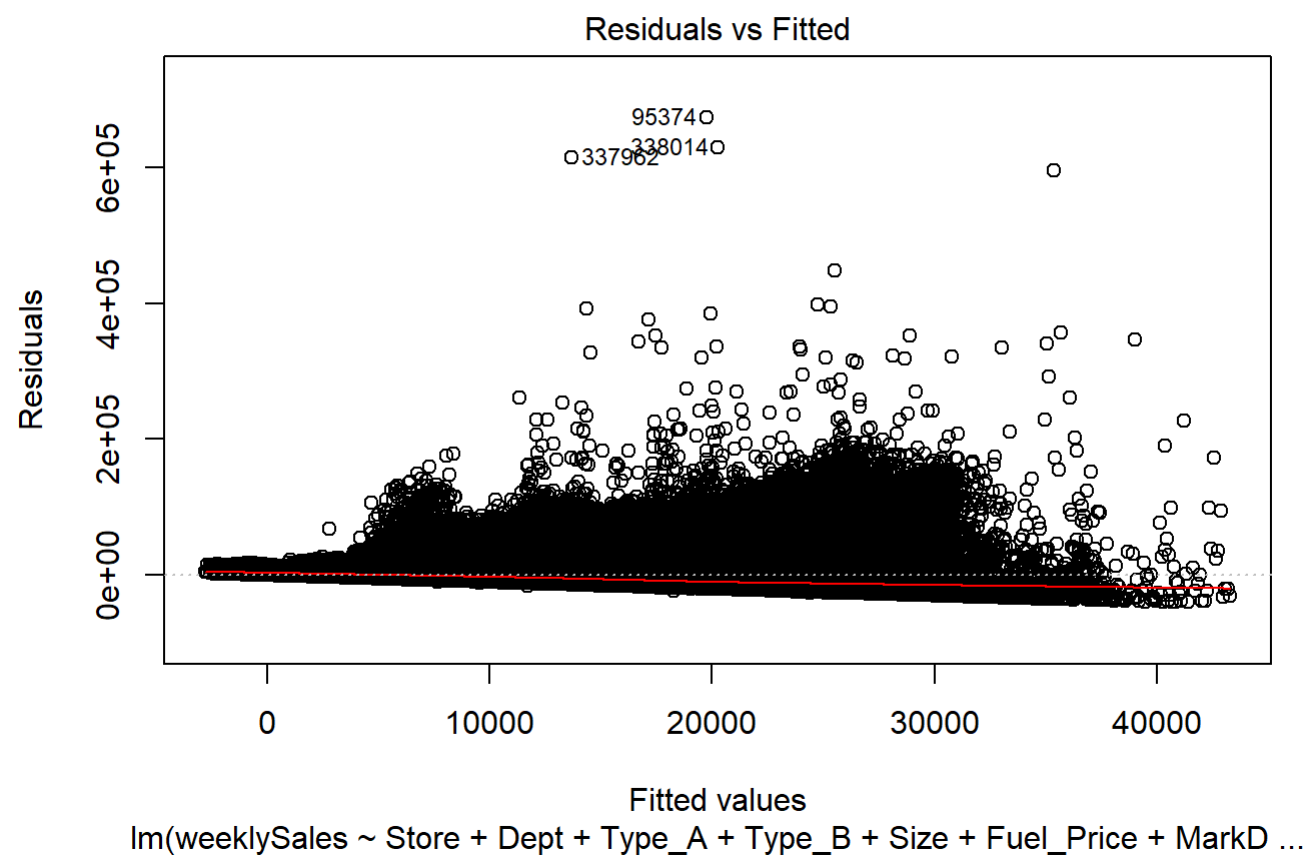
```

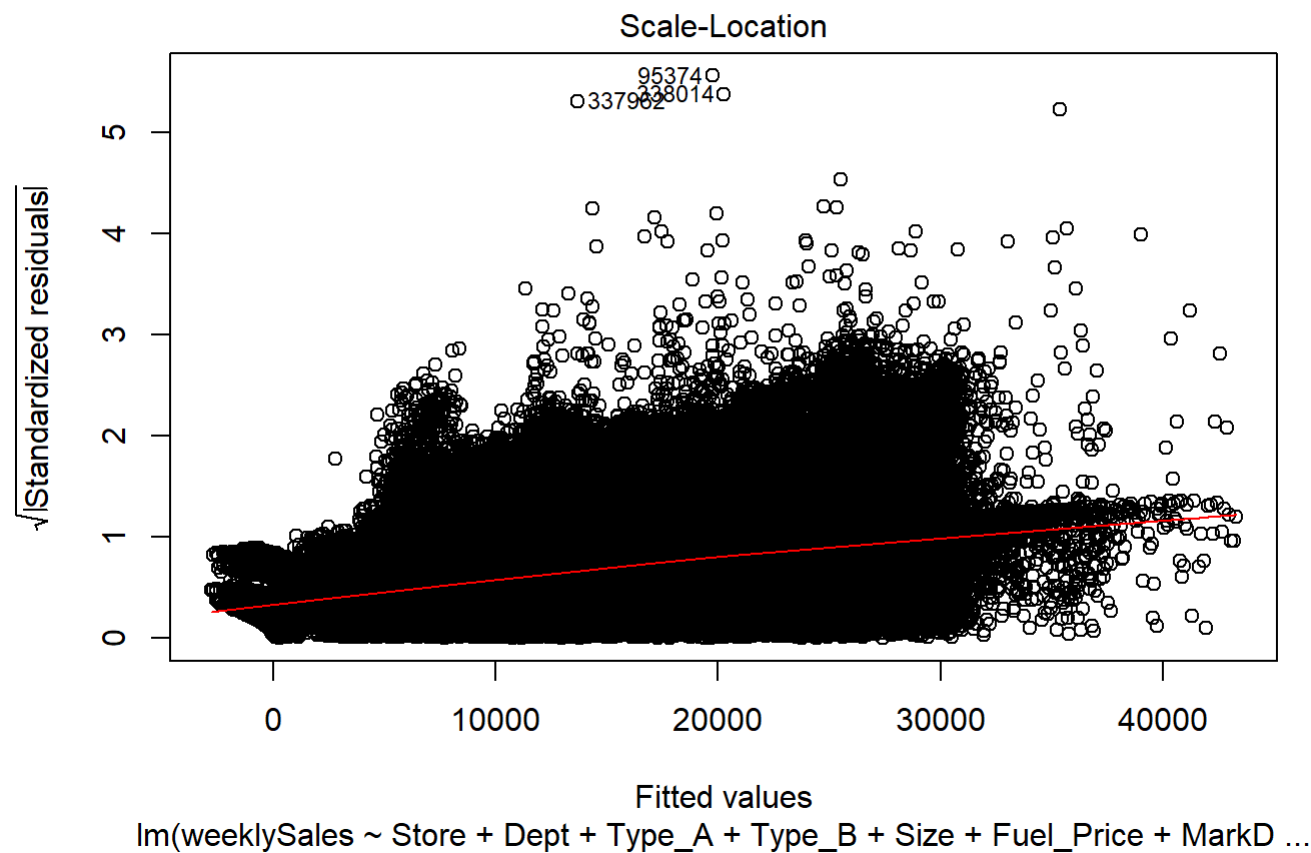
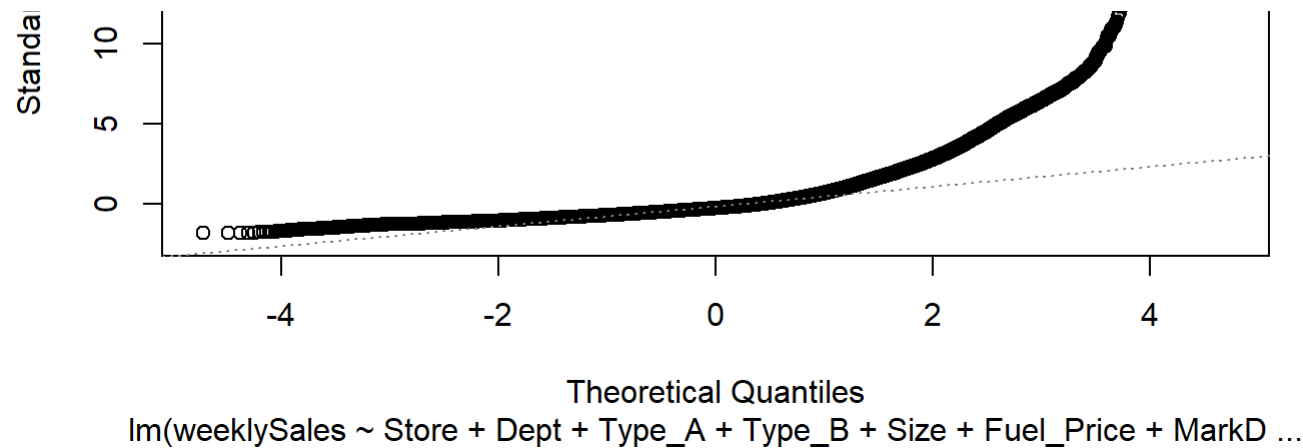
##          (Intercept)          Store          Dept          Type_A
## (Intercept)  1.000000000 -0.0078419224  1.038306e-03 -1.925409e-02
## Store      -0.007841922  1.00000000000 -5.528446e-03  2.079706e-01
## Dept       0.001038306  -0.0055284456  1.000000e+00  6.054722e-03
## Type_A     -0.019254094  0.2079705821  6.054722e-03  1.000000e+00
## Type_B     -0.071848974  0.3882548166  1.847532e-02  8.121223e-01
## Size       -0.114410326  0.0858871526  -8.175262e-05 -7.082436e-01
## Fuel_Price  0.810776028  -0.0312001232  6.841581e-04  3.552740e-03
## Markdown1   0.379354408  0.0131964094  -2.580528e-03  4.443815e-05
## Markdown3   0.018862197  0.0113493178  -1.997433e-03  4.625688e-03
## Markdown5   0.294563762  -0.0447508839  -1.122068e-03  -1.675107e-02
## CPI         0.242848139  0.2012544693  7.583995e-03  2.055888e-03
## Unemployment -0.278354278  -0.0921419534  -6.890447e-04  1.328923e-01
## Year        -0.999997001  0.0071488164  -1.342390e-03  1.876204e-02
## Month       -0.323016499  0.0002205912  -1.213535e-03  8.129327e-03
## Day         0.051374381  -0.0012536517  2.555889e-04  -1.572169e-03
##          Type_B          Size          Fuel_Price          Markdown1
## (Intercept) -0.071848974 -1.144103e-01  0.8107760275  3.793544e-01
## Store        0.388254817  8.588715e-02 -0.0312001232  1.319641e-02
## Dept         0.018475321  -8.175262e-05  0.0006841581  -2.580528e-03
## Type_A       0.812122329  -7.082436e-01  0.0035527404  4.443815e-05
## Type_B       1.000000000  -3.527501e-01 -0.0494531659  -5.757889e-02
## Size        -0.352750066  1.000000e+00 -0.0902347703  -1.359854e-01
## Fuel_Price  -0.049453166  -9.023477e-02  1.0000000000  1.789574e-01
## Markdown1   -0.057578888  -1.359854e-01  0.1789574277  1.000000e+00
## Markdown3   -0.005140495  -2.639391e-02  0.0055523605  2.845876e-02
## Markdown5   -0.024692507  -7.957990e-02  0.1704814873  -2.026316e-01
## CPI         0.089744248  6.855958e-02  0.2984194748  7.761606e-02
## Unemployment 0.133932533  -3.838066e-02 -0.1893510023  -6.650520e-02
## Year        0.071026647  1.141273e-01 -0.8117240852  -3.791226e-01
## Month       0.020638410  3.352341e-02 -0.2454571672  -2.647833e-02
## Day        -0.008979231  -2.205709e-02  0.0001386764  1.365649e-01
##          Markdown3          Markdown5          CPI          Unemployment          Year
## (Intercept) 0.018862197  0.294563762  0.242848139 -0.2783542783 -0.999997001
## Store       0.011349318  -0.044750884  0.201254469 -0.0921419534  0.007148816
## Dept        -0.001997433  -0.001122068  0.007583995 -0.0006890447 -0.001342390
## Type_A      0.004625688  -0.016751071  0.002055888  0.1328923064  0.018762041
## Type_B      -0.005140495  -0.024692507  0.089744248  0.1339325327  0.071026647
## Size        -0.026393910  -0.079579905  0.068559577 -0.0383806648  0.114127332
## Fuel_Price  0.005552361  0.170481487  0.298419475 -0.1893510023 -0.811724085

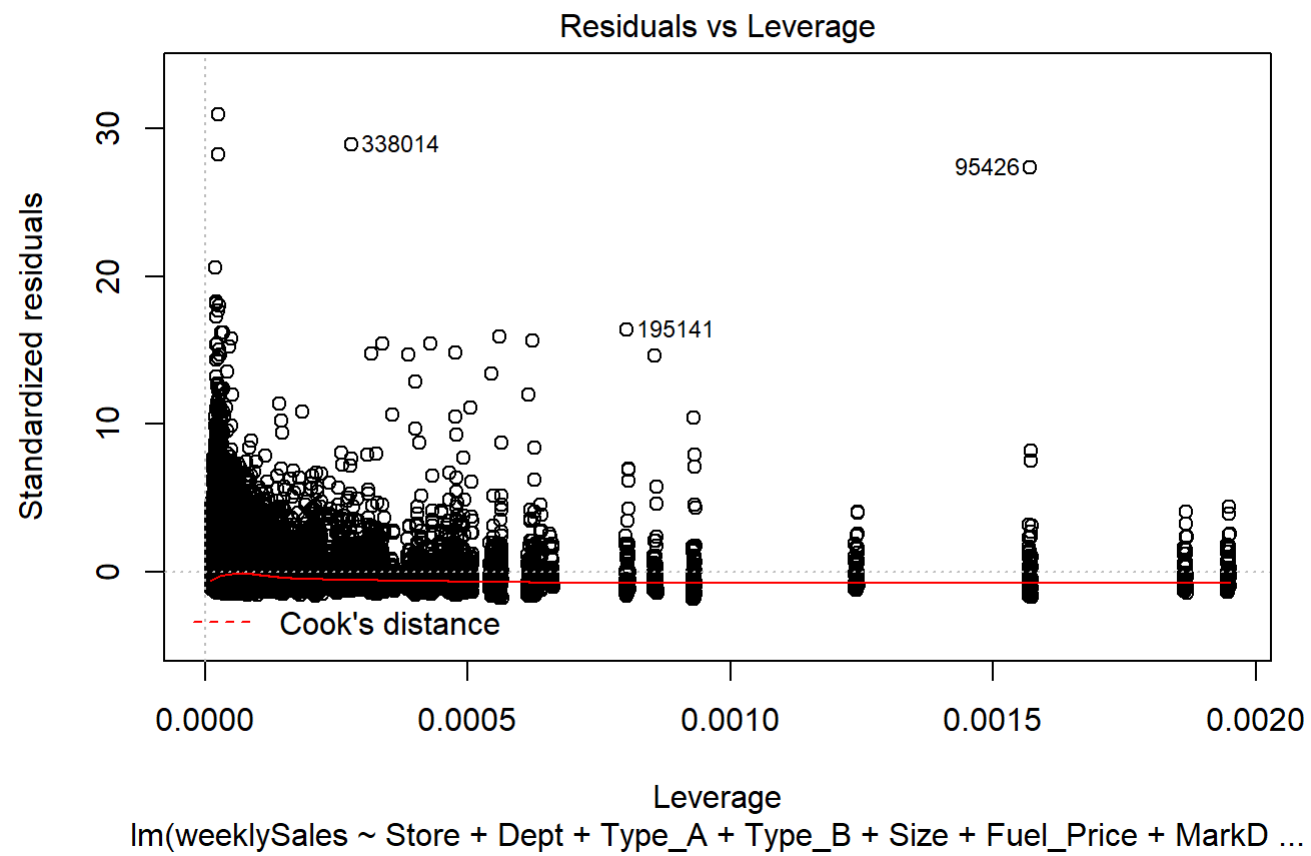
```

```
## Markdown1      0.028458761 -0.202631563  0.077616065 -0.0665051998 -0.379122647
## Markdown3      1.000000000 -0.030525786  0.014210568  0.0058344312 -0.018795284
## Markdown5     -0.030525786  1.000000000  0.001294047 -0.0285123094 -0.294338652
## CPI            0.014210568  0.001294047  1.000000000  0.1826251589 -0.244344513
## Unemployment   0.005834431 -0.028512309  0.182625159  1.0000000000  0.277197005
## Year          -0.018795284 -0.294338652 -0.244344513  0.2771970048  1.000000000
## Month         -0.110506968 -0.183781465 -0.071879442  0.1019585202  0.322612781
## Day          -0.067601581  0.033355703 -0.002119880 -0.0045702470 -0.051595541
##              Month      Day
## (Intercept) -0.3230164985  0.0513743811
## Store        0.0002205912 -0.0012536517
## Dept        -0.0012135353  0.0002555889
## Type_A       0.0081293269 -0.0015721689
## Type_B       0.0206384096 -0.0089792314
## Size         0.0335234142 -0.0220570884
## Fuel_Price  -0.2454571672  0.0001386764
## Markdown1    -0.0264783295  0.1365648681
## Markdown3    -0.1105069677 -0.0676015808
## Markdown5    -0.1837814649  0.0333557028
## CPI          -0.0718794416 -0.0021198800
## Unemployment  0.1019585202 -0.0045702470
## Year         0.3226127805 -0.0515955405
## Month        1.0000000000 -0.0162326125
## Day         -0.0162326125  1.0000000000
```

```
plot(finalModel2)
```



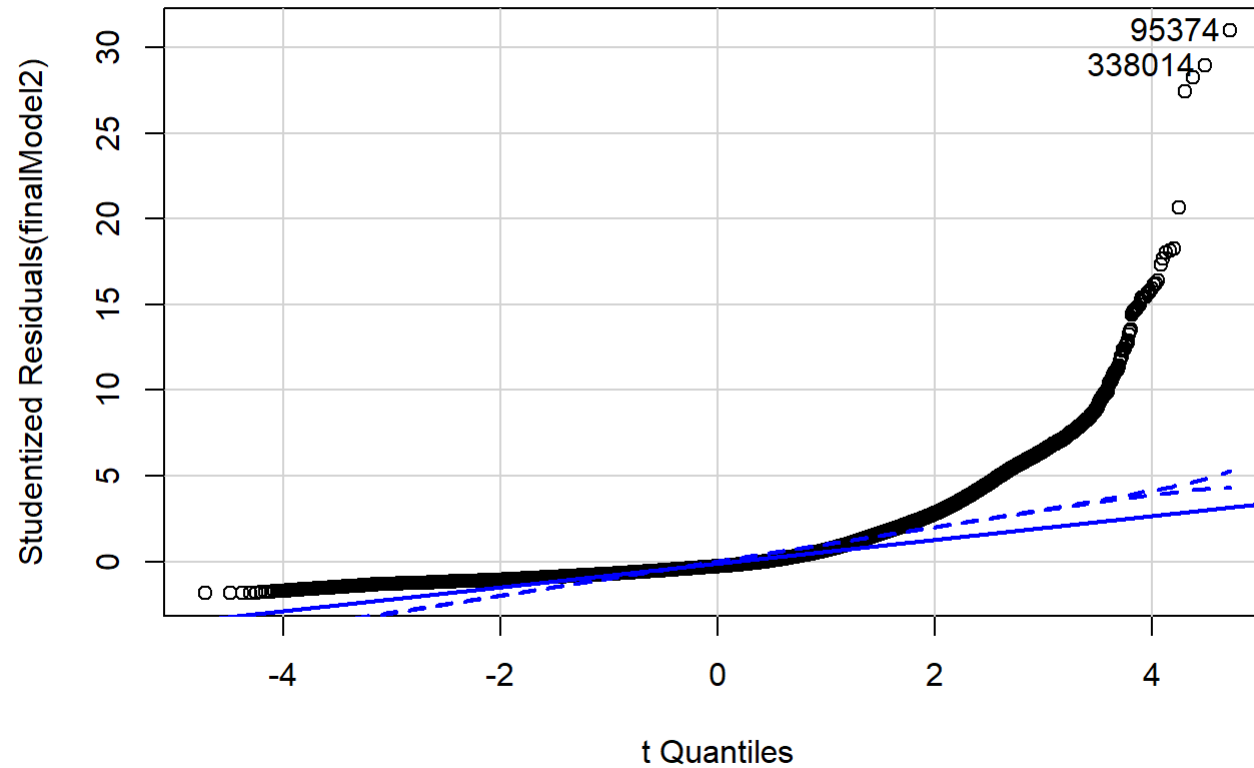


```
outlierTest(finalModel2)
```

```
##          rstudent unadjusted p-value Bonferroni p
## 95374  30.96642      2.6358e-210  1.1112e-204
## 338014 28.94929      4.3465e-184  1.8323e-178
## 337962 28.24315      2.5206e-175  1.0626e-169
## 95426  27.40525      3.3199e-165  1.3996e-159
## 135666 20.62576      1.7952e-94   7.5680e-89
## 195089 18.26824      1.5832e-74   6.6745e-69
## 264391 18.16307      1.0798e-73   4.5522e-68
## 88429  18.04270      9.5915e-73   4.0435e-67
## 95378  17.65884      9.2204e-70   3.8870e-64
## 214433 17.30299      4.7112e-67   1.9861e-61
```

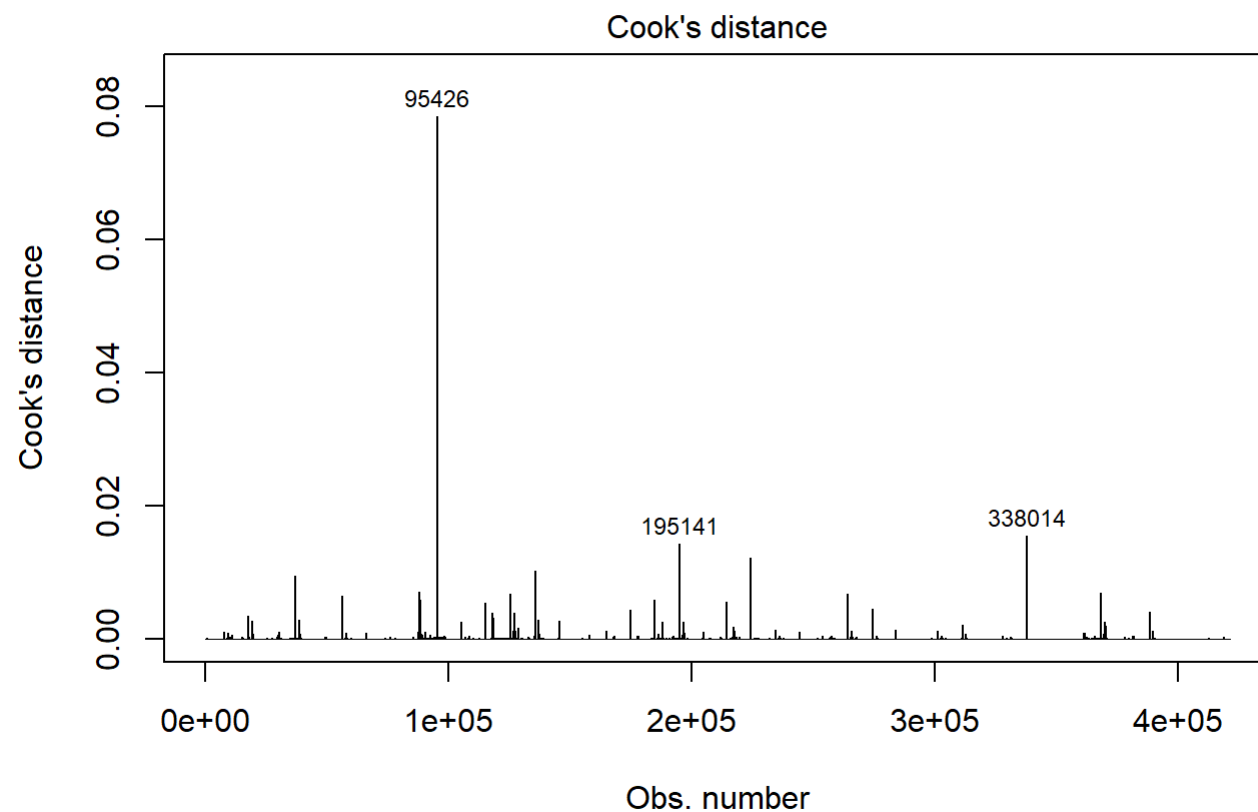
```
qqPlot(finalModel2, main="QQ Plot")
```

QQ Plot



```
## [1] 95374 338014
```

```
# Cook's Distance plot
cutoff <- 4/((nrow(dataset)-length(finalModel2$coefficients)-2))
plot(finalModel2, which=4, cook.levels=cutoff)
```



lm(weeklySales ~ Store + Dept + Type_A + Type_B + Size + Fuel_Price + MarkD ...

```
influencePlot(finalModel2, id.method="identify", main="Influence Plot", sub="Circle s
ize is proportional to Cook's Distance" )
```

```
## Warning in plot.window(...): "id.method" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "id.method" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "id.method" is not  
## a graphical parameter
```

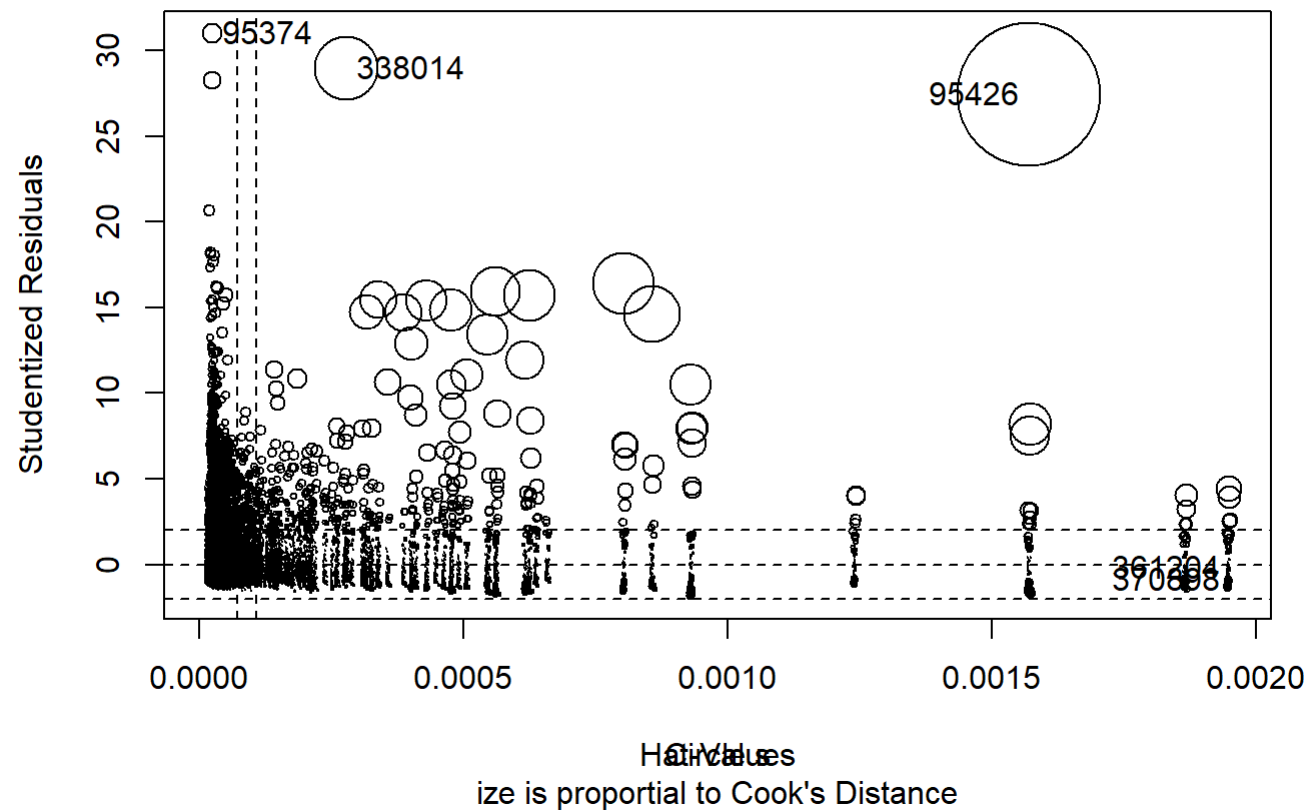
```
## Warning in axis(side = side, at = at, labels = labels, ...): "id.method" is not  
## a graphical parameter
```

```
## Warning in box(...): "id.method" is not a graphical parameter
```

```
## Warning in title(...): "id.method" is not a graphical parameter
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "id.method" is not a  
## graphical parameter
```

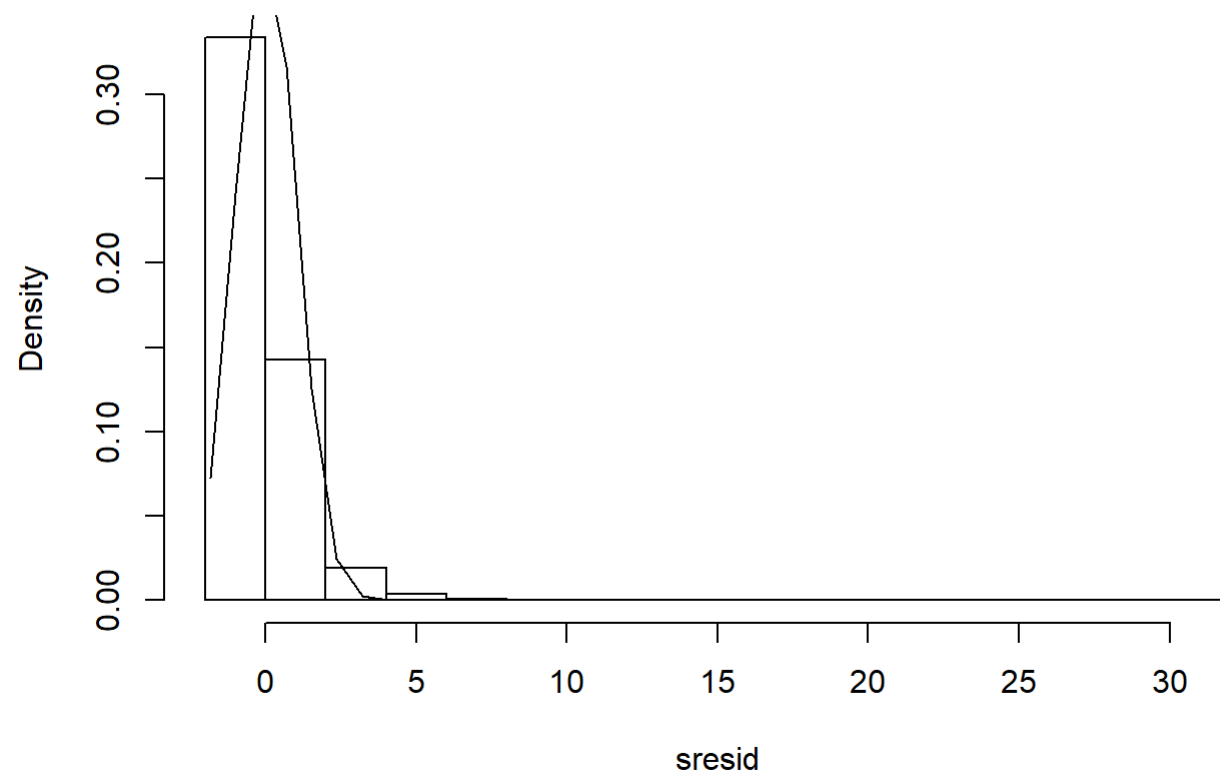
Influence Plot



##	StudRes	Hat	CookD
## 95374	30.9664199	2.416956e-05	1.541645e-03
## 95426	27.4052465	1.568981e-03	7.854232e-02
## 338014	28.9492911	2.784490e-04	1.553065e-02
## 361204	-0.2316797	1.950300e-03	6.992539e-06
## 370898	-0.9449781	1.950280e-03	1.163315e-04


```
sresid <- studres(finalModel2)
hist(sresid, freq=FALSE,
     main="Distribution of Studentized Residuals")
xfit<-seq(min(sresid),max(sresid),length=40)
yfit<-dnorm(xfit)
lines(xfit, yfit)
```

Distribution of Studentized Residuals



```
ncvTest(finalModel2)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 109821.1, Df = 1, p = < 2.22e-16
```

```
vif(finalModel2)
```

```
##      Store      Dept      Type_A      Type_B      Size      Fuel_Price
##  1.369330  1.001011  7.551806  4.500355  3.135040  3.424324
##  Markdown1  Markdown3  Markdown5      CPI  Unemployment      Year
##  1.579755  1.022991  1.387200  1.291545  1.267197  4.723243
##      Month      Day
##  1.152356  1.030145
```

```
sqrt(vif(finalModel2)) > 2
```

```
##      Store      Dept      Type_A      Type_B      Size      Fuel_Price
##      FALSE      FALSE      TRUE      TRUE      FALSE      FALSE
##  Markdown1  Markdown3  Markdown5      CPI  Unemployment      Year
##      FALSE      FALSE      FALSE      FALSE      FALSE      TRUE
##      Month      Day
##      FALSE      FALSE
```

Type_A, Type_B and Year have multi-collinearity