# model.R

*atchirc*

*Fri May 05 17:42:28 2017*

```r
# *************************************************************************
#                   Load Library ----
# *************************************************************************
library(MASS)




# *************************************************************************
#                   LOad Data ----
# *************************************************************************
camera_accessory_data_nrm <- read.csv('./intrim/cameraAccessory.csv')
home_audio_data_nrm <-       read.csv('./intrim/homeAudio.csv')
gaming_accessory_data_nrm <- read.csv('./intrim/gamingAccessory')




# *************************************************************************
#                     LINEAR MODEL : Camera_accessory ----
# *************************************************************************

# do we really need to predict something,
# I feel, we are trying to explain effects of indenpendant variables
# set.seed(100)
indices=sample(1:nrow(camera_accessory_data_nrm),1.0*nrow(camera_accessory_data_nrm))
train=camera_accessory_data_nrm[indices,]
test=camera_accessory_data_nrm[-indices,]

# Initial Model
model_cam1 <- lm(gmv~ .,data=train)

# summary(model_cam1)
step_cam <- stepAIC(model_cam1, direction = "both",trace=FALSE,k=2)


summary(step_cam)
```

```
##
## Call:
## lm(formula = gmv ~ sla + procurement_sla + TV + Digital + Sponsorship +
##     ContentMarketing + Affiliates + SEM + Radio, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0018 -0.3900 -0.0508  0.2848  2.3954
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.517e-16  9.149e-02   0.000 1.000000
## sla               1.375e-01  9.847e-02   1.397 0.169874
```

```
## procurement_sla     1.545e-01   9.560e-02    1.616 0.113491
## TV                  -1.072e+00   2.966e-01   -3.614 0.000801 ***
## Digital              3.252e+00   1.128e+00    2.882 0.006194 **
## Sponsorship          1.700e+00   3.242e-01    5.243 4.82e-06 ***
## ContentMarketing    -1.983e+00   8.301e-01   -2.389 0.021480 *
## Affiliates           1.501e+00   3.511e-01    4.276 0.000107 ***
## SEM                 -2.368e+00   1.078e+00   -2.196 0.033664 *
## Radio                6.730e-01   2.446e-01    2.751 0.008725 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6598 on 42 degrees of freedom
## Multiple R-squared:  0.6415, Adjusted R-squared:  0.5647
## F-statistic: 8.351 on 9 and 42 DF,  p-value: 5.293e-07
```

```
# *****************************************************************************
#                       LINEAR MODEL : gaming_accessory ----
# *****************************************************************************

indices=sample(1:nrow(home_audio_data_nrm),1.0*nrow(home_audio_data_nrm))
train=home_audio_data_nrm[indices,]
test=home_audio_data_nrm[-indices,]

# Initial Model
model_ga1 <- lm(gmv~ .,data=train)

# summary(model_ga1)
step_ga <- stepAIC(model_ga1, direction = "both",trace=FALSE)

summary(step_ga)
```

```
##
## Call:
## lm(formula = gmv ~ discount + sla + procurement_sla + TV + Digital +
##       Sponsorship + ContentMarketing + OnlineMarketing + SEM +
##       Radio + Other + NPS, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.80632 -0.27195 -0.00747  0.25673  2.57773
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.905e-15  8.188e-02    0.000 1.000000
## discount          1.069e+00  1.498e-01    7.137 1.86e-08 ***
## sla               4.204e-01  1.249e-01    3.366 0.001789 **
## procurement_sla  -3.585e-01  9.868e-02   -3.633 0.000845 ***
## TV                2.218e+00  1.149e+00    1.931 0.061166 .
## Digital           7.220e+00  2.288e+00    3.156 0.003177 **
## Sponsorship       8.694e-01  3.254e-01    2.672 0.011156 *
## ContentMarketing  2.319e+00  1.513e+00    1.533 0.133696
## OnlineMarketing  -1.497e+00  8.405e-01   -1.781 0.083099 .
## SEM              -1.031e+01  3.505e+00   -2.943 0.005585 **
## Radio             7.623e+00  2.919e+00    2.611 0.012957 *
## Other            -7.965e+00  3.208e+00   -2.483 0.017696 *
```

```
## NPS                -4.934e-01  3.250e-01  -1.518 0.137539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.579 on 37 degrees of freedom
## Multiple R-squared:  0.7469, Adjusted R-squared:  0.6648
## F-statistic: 9.097 on 12 and 37 DF,  p-value: 8.82e-08
```

```
# ****************************************************************************
#                       LINEAR MODEL : home_audio ----
# ****************************************************************************

indices=sample(1:nrow(gaming_accessory_data_nrm),1.0*nrow(gaming_accessory_data_nrm))
train=gaming_accessory_data_nrm[indices,]
test=gaming_accessory_data_nrm[-indices,]

# Initial Model
model_ha1 <- lm(gmv~ .,data=train)

# summary(model_ha1)
step_ha <- stepAIC(model_ha1, direction = "both",trace=FALSE)

summary(step_ha)
```

```
##
## Call:
## lm(formula = gmv ~ discount + procurement_sla + TV + Digital +
##     Sponsorship + ContentMarketing + OnlineMarketing + Affiliates +
##     SEM + Radio + NPS, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82764 -0.27903 -0.04518  0.15915  1.36044
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.251e-15  7.102e-02   0.000 1.000000
## discount          3.128e-01  1.112e-01   2.812 0.007513 **
## procurement_sla   1.440e-01  9.087e-02   1.584 0.120817
## TV               -1.811e+00  3.120e-01  -5.805 8.19e-07 ***
## Digital           4.053e+00  1.005e+00   4.034 0.000233 ***
## Sponsorship       2.600e+00  3.652e-01   7.119 1.12e-08 ***
## ContentMarketing -3.962e+00  7.773e-01  -5.096 8.22e-06 ***
## OnlineMarketing  -2.882e+00  1.742e+00  -1.655 0.105553
## Affiliates        5.204e+00  1.718e+00   3.030 0.004222 **
## SEM              -1.560e+00  9.141e-01  -1.706 0.095511 .
## Radio             1.325e+00  2.235e-01   5.928 5.48e-07 ***
## NPS               5.061e-01  2.305e-01   2.195 0.033849 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5171 on 41 degrees of freedom
## Multiple R-squared:  0.7892, Adjusted R-squared:  0.7327
## F-statistic: 13.96 on 11 and 41 DF,  p-value: 1.343e-10
```

"

"