

Modeling and Evaluation

Sunday BENJAMIN

June 22, 2021

```
library(caTools)
library(nnet)
library(plyr)
library(dplyr)
library(caret)
```

Row and Column Count of Data Set respectively

```
nrow(Grand_Total_for_modeling);ncol(Grand_Total_for_modeling)
```

```
## [1] 16096
```

```
## [1] 63
```

Using caTools to split data into 75% for training and 25% for testing

Splitting

```
set.seed(109)
```

```
# Use caTools package for splitting with SplitRatio of 70%:30%
```

```
dff= sample.split(Grand_Total_for_modeling$id,SplitRatio = 0.75)
```

Subsetting into inTrain data

```
inTrain = subset(Grand_Total_for_modeling,dff==TRUE)
```

```
nrow(inTrain);ncol(inTrain)
```

```
## [1] 12072
```

```
## [1] 63
```

```
as.data.frame(head(inTrain,1L))
```

```
## id cons_12m cons_gas_12m cons_last_month
```

```
## 1 0002203ffbb812588b632b9e628cc38d 4.343113 0 3.489255
```

```
## forecast_cons_12m forecast_discount_energy forecast_meter_rent_12m
```

```
## 1 2.863359 0 2.145973
```

```
## forecast_price_energy_p1 forecast_price_energy_p2 forecast_price_pow_p1
```

```
## 1 0.1169 0.100015 40.6067
```

```
## has_gas imp_cons margin_gross_pow_ele margin_net_pow_ele nb_prod_act
```

```
## 1 0 1.620968 43.08 43.08 1
```

```
## net_margin pow_max churn days_since months_activ months_to_end
```

```

## 1 81.42 17.25 0 2224 72 2
## months_modif_prod months_renewal id2 activity_apd activity_ckf
activity_clu
## 1 72 10 1 0 0 0
## activity_cwo activity_fmw activity_kkk activity_kwu activity_sfi
activity_wxe
## 1 0 0 0 0 0 0
## channel_epu channel_ewp channel_fix channel_foo channel_lmk channel_sdd
## 1 0 0 0 1 0 0
## channel_usi origin_ewx origin_kam origin_ldk origin_lxi origin_usa
## 1 0 0 1 0 0 0
## mean_year_price_p1_var mean_year_price_p2_var mean_year_price_p3_var
## 1 0.1243384 0.1037938 0.07316033
## mean_year_price_p1_fix mean_year_price_p2_fix mean_year_price_p3_fix
## 1 40.70173 24.42104 16.28069
## mean_year_price_p1 mean_year_price_p2 mean_year_price_p3 G.Total_1_yr
## 1 40.82607 24.52483 16.35385 1235018
## G.Total_6_months G.Total_3_months discount_1 discount_2 discount_3
active_1Y
## 1 741893.5 494715.7 988014 593514.8 395772.5 1
## active_6m active_3m
## 1 1 1

```

`tail(inTrain,1L)`

```

## id cons_12m cons_gas_12m cons_last_month
## 16096 ffff7fa066f1fb305ae285bb03bf325a 4.705924 0 3.739731
## forecast_cons_12m forecast_discount_energy forecast_meter_rent_12m
## 16096 3.016908 0 2.12064
## forecast_price_energy_p1 forecast_price_energy_p2 forecast_price_pow_p1
## 16096 0.11691 0.100572 40.6067
## has_gas imp_cons margin_gross_pow_ele margin_net_pow_ele nb_prod_act
## 16096 0 2.017117 23.72 23.72 1
## net_margin pow_max churn days_since months_activ months_to_end
## 16096 132.2 19 0 1461 43 6
## months_modif_prod months_renewal id2 activity_apd activity_ckf
## 16096 26 6 16096 0 0
## activity_clu activity_cwo activity_fmw activity_kkk activity_kwu
## 16096 0 0 0 0 0
## activity_sfi activity_wxe channel_epu channel_ewp channel_fix channel_foo
## 16096 0 0 0 0 0 1
## channel_lmk channel_sdd channel_usi origin_ewx origin_kam origin_ldk
## 16096 0 0 0 0 0 0
## origin_lxi origin_usa mean_year_price_p1_var mean_year_price_p2_var
## 16096 1 0 0.1253601 0.1048949
## mean_year_price_p3_var mean_year_price_p1_fix mean_year_price_p2_fix
## 16096 0.075635 40.64743 24.38846
## mean_year_price_p3_fix mean_year_price_p1 mean_year_price_p2
## 16096 16.25897 40.77279 24.49335
## mean_year_price_p3 G.Total_1_yr G.Total_6_months G.Total_3_months

```

```
## 16096 16.33461 2277470 1368140 912411.8
## discount_1 discount_2 discount_3 active_1Y active_6m active_3m
## 16096 1821976 1094512 729929.4 1 1 1
```

Subsetting into Test data

```
test = subset(Grand_Total_for_modeling,dff==FALSE)
```

```
nrow(test);ncol(test)
```

```
## [1] 4024
```

```
## [1] 63
```

```
head(test,1L)
```

```
## id cons_12m cons_gas_12m cons_last_month
## 3 0010bcc39e42b3c2131ed2ce55246e3c 3.871631 0 3.026533
## forecast_cons_12m forecast_discount_energy forecast_meter_rent_12m
## 3 3.117987 30 1.28713
## forecast_price_energy_p1 forecast_price_energy_p2 forecast_price_pow_p1
## 3 0.19923 0 45.80688
## has_gas imp_cons margin_gross_pow_ele margin_net_pow_ele nb_prod_act
## 3 0 2.331953 38.58 38.58 2
## net_margin pow_max churn days_since months_activ months_to_end
## 3 81.61 13.856 0 1165 35 4
## months_modif_prod months_renewal id2 activity_apd activity_ckf
activity_clu
## 3 8 10 3 0 0 0
## activity_cwo activity_fmw activity_kkk activity_kwu activity_sfi
activity_wxe
## 3 0 0 0 0 0 0
## channel_epu channel_ewp channel_fix channel_foo channel_lmk channel_sdd
## 3 0 0 0 0 0 0
## channel_usi origin_ewx origin_kam origin_ldk origin_lxi origin_usa
## 3 1 0 0 0 1 0
## mean_year_price_p1_var mean_year_price_p2_var mean_year_price_p3_var
## 3 0.1815585 0 0
## mean_year_price_p1_fix mean_year_price_p2_fix mean_year_price_p3_fix
## 3 45.31971 0 0
## mean_year_price_p1 mean_year_price_p2 mean_year_price_p3 G.Total_1_yr
## 3 45.50127 0 0 540803.6
## G.Total_6_months G.Total_3_months discount_1 discount_2 discount_3
active_1Y
## 3 0 0 432642.9 0 0 1
## active_6m active_3m
## 3 0 0
```

```

tail(test,1L)

## id cons_12m cons_gas_12m cons_last_month
## 16091 ffebf6a979dd0b17a41076df1057e733 5.068865 0 4.102948
## forecast_cons_12m forecast_discount_energy forecast_meter_rent_12m
## 16091 3.859518 0 2.152105
## forecast_price_energy_p1 forecast_price_energy_p2 forecast_price_pow_p1
## 16091 0.112922 0.09781 40.6067
## has_gas imp_cons margin_gross_pow_ele margin_net_pow_ele nb_prod_act
## 16091 0 2.774006 13.02 13.02 1
## net_margin pow_max churn days_since months_activ months_to_end
## 16091 742.37 34.64 0 4173 130 9
## months_modif_prod months_renewal id2 activity_apd activity_ckf
## 16091 130 4 16091 0 0
## activity_clu activity_cwo activity_fmw activity_kkk activity_kwu
## 16091 0 0 0 0 0
## activity_sfi activity_wxe channel_epu channel_ewp channel_fix channel_foo
## 16091 0 0 0 0 0 0
## channel_lmk channel_sdd channel_usi origin_ewx origin_kam origin_ldk
## 16091 1 0 0 0 0 1
## origin_lxi origin_usa mean_year_price_p1_var mean_year_price_p2_var
## 16091 0 0 0.1224256 0.1023752
## mean_year_price_p3_var mean_year_price_p1_fix mean_year_price_p2_fix
## 16091 0.07221025 40.6067 24.36402
## mean_year_price_p3_fix mean_year_price_p1 mean_year_price_p2
## 16091 16.24268 40.72913 24.46639
## mean_year_price_p3 G.Total_1_yr G.Total_6_months G.Total_3_months
## 16091 16.31489 4702772 2825002 1883792
## discount_1 discount_2 discount_3 active_1Y active_6m active_3m
## 16091 3762218 2260002 1507034 1 1 1

```

Using multinomial logistic regression function from the “nnet” package

Probability model

```

prob.model = multinom(formula =
churn~exp(cons_12m+1)+exp(cons_gas_12m+1)+exp(cons_last_month+1)+
exp(forecast_cons_12m+1)+exp(forecast_discount_energy+1)+exp(forecast_meter_r
ent_12m+1)+forecast_price_energy_p1+forecast_price_energy_p2+forecast_price_p
ow_p1+
has_gas+exp(imp_cons+1)+margin_gross_pow_ele+margin_net_pow_ele+nb_prod_act+n
et_margin+pow_max+days_since+months_activ+months_to_end+months_modif_prod+mon
ths_renewal+
activity_apd+activity_ckf+activity_clu+activity_cwo+activity_fmw+

```

```

        activity_kkk+activity_kwu+activity_sfi+activity_wxe+
channel_ewp+channel_fix+channel_foo+channel_lmk+channel_sdd+channel_usi+
        origin_kam+origin_ldk+origin_lxi+origin_usa+
mean_year_price_p1+mean_year_price_p2+mean_year_price_p3+
        G.Total_1_yr+G.Total_6_months+G.Total_3_months,
data = inTrain)

## # weights:  48 (47 variable)
## initial  value 8367.672764
## final  value 8367.672764
## converged

## "channel_epu" removed; it gives "-Inf" "in coef/std" ratio

coef = summary(prob.model)$coefficients
std  = summary(prob.model)$standard.errors

coef/std

##              (Intercept)                exp(cons_12m + 1)
##              -1.035473e-12                -1.749256e-13
##              exp(cons_gas_12m + 1)          exp(cons_last_month + 1)
##              -7.292759e-14                -1.303192e-13
##              exp(forecast_cons_12m + 1) exp(forecast_discount_energy + 1)
##              -6.168544e-13                -5.361432e-02
##              exp(forecast_meter_rent_12m + 1) forecast_price_energy_p1
##              -2.132133e-13                -5.701656e-13
##              forecast_price_energy_p2        forecast_price_pow_p1
##              -2.664268e-13                -1.001919e-12
##              has_gas                        exp(imp_cons + 1)
##              -2.003129e-13                -3.889209e-13
##              margin_gross_pow_ele            margin_net_pow_ele
##              -4.921986e-13                -4.776141e-13
##              nb_prod_act                    net_margin
##              -5.452450e-14                -1.839932e-13
##              pow_max                        days_since
##              -1.299772e-13                -1.065396e-12
##              months_activ                  months_to_end
##              -1.063925e-12                -1.046803e-12
##              months_modif_prod              months_renewal
##              -8.664102e-13                -8.248966e-13
##              activity_apd                  activity_ckf
##              -9.847100e-14                -4.169589e-12
##              activity_clu                  activity_cwo
##              -1.021853e-11                -4.339934e-12
##              activity_fmww                  activity_kkk
##              -1.007784e-12                -9.983917e-13
##              activity_kwu                  activity_sfi

```

```
##          -4.388568e-13          -1.379964e-11
##          activity_wxe          channel_ewp
##          -2.849893e-12          -5.180213e-12
##          channel_fix          channel_foo
##          -8.811661e-15          -1.632801e-12
##          channel_lmk          channel_sdd
##          -2.394760e-13          -6.946275e-13
##          channel_usi          origin_kam
##          -2.317400e-12          -6.951129e-13
##          origin_ldk          origin_lxi
##          -3.315008e-13          -5.449760e-13
##          origin_usa          mean_year_price_p1
##          -2.657071e-11          -9.972945e-13
##          mean_year_price_p2          mean_year_price_p3
##          -8.684414e-14          -8.136094e-14
##          G.Total_1_yr          G.Total_6_months
##          -1.884830e-14          -2.923456e-15
##          G.Total_3_months
##          -2.872950e-15
```

```
ratio = as.data.frame(coef/std);ratio %>% arrange(desc(coef/std))
```

```
##          coef/std
## G.Total_3_months          -2.872950e-15
## G.Total_6_months          -2.923456e-15
## channel_fix          -8.811661e-15
## G.Total_1_yr          -1.884830e-14
## nb_prod_act          -5.452450e-14
## exp(cons_gas_12m + 1)          -7.292759e-14
## mean_year_price_p3          -8.136094e-14
## mean_year_price_p2          -8.684414e-14
## activity_apd          -9.847100e-14
## pow_max          -1.299772e-13
## exp(cons_last_month + 1)          -1.303192e-13
## exp(cons_12m + 1)          -1.749256e-13
## net_margin          -1.839932e-13
## has_gas          -2.003129e-13
## exp(forecast_meter_rent_12m + 1)          -2.132133e-13
## channel_lmk          -2.394760e-13
## forecast_price_energy_p2          -2.664268e-13
## origin_ldk          -3.315008e-13
## exp(imp_cons + 1)          -3.889209e-13
## activity_kwu          -4.388568e-13
## margin_net_pow_ele          -4.776141e-13
## margin_gross_pow_ele          -4.921986e-13
## origin_lxi          -5.449760e-13
## forecast_price_energy_p1          -5.701656e-13
## exp(forecast_cons_12m + 1)          -6.168544e-13
## channel_sdd          -6.946275e-13
## origin_kam          -6.951129e-13
```

```
## months_renewal -8.248966e-13
## months_modif_prod -8.664102e-13
## mean_year_price_p1 -9.972945e-13
## activity_kkk -9.983917e-13
## forecast_price_pow_p1 -1.001919e-12
## activity_fmw -1.007784e-12
## (Intercept) -1.035473e-12
## months_to_end -1.046803e-12
## months_activ -1.063925e-12
## days_since -1.065396e-12
## channel_foo -1.632801e-12
## channel_usi -2.317400e-12
## activity_wxe -2.849893e-12
## activity_ckf -4.169589e-12
## activity_cwo -4.339934e-12
## channel_ewp -5.180213e-12
## activity_clu -1.021853e-11
## activity_sfi -1.379964e-11
## origin_usa -2.657071e-11
## exp(forecast_discount_energy + 1) -5.361432e-02
```

INTERPRETATION

Variables having Ratio value (coef/std) greater than +2 or less than -2 are highly significant and therefore impact the prediction.

CREATING AN AMOUNT MODEL;

1. Create an “amount.model” (i.e. revenue realized for 1 year).
Note; we could also create amount models for “3 months” and “6 months” but would only be focusing for 1 year.
2. Filter those who were active (i.e in the 1st year) in order to create an amount model.

Filtering out those who were active in the First Year

```
z_1 = inTrain %>% filter(active_1Y==1)

amount.model_1 = lm(formula =
G.Total_1_yr~exp(cons_12m+1)+exp(cons_gas_12m+1)+exp(cons_last_month+1)+
exp(forecast_cons_12m+1)+exp(forecast_discount_energy+1)+exp(forecast_meter_r
ent_12m+1)+forecast_price_energy_p1+
```

```

forecast_price_energy_p2+forecast_price_pow_p1+
has_gas+exp(imp_cons+1)+

margin_gross_pow_ele+margin_net_pow_ele+nb_prod_act+net_margin+pow_max+days_s
ince+

months_activ+months_to_end+months_modif_prod+months_renewal+

activity_apd+activity_ckf+activity_clu+activity_cwo+activity_fmw+
activity_kkk+activity_kwu+activity_sfi+activity_wxe+

channel_ewp+channel_fix+channel_foo+channel_lmk+channel_sdd+channel_usi+
origin_kam+origin_ldk+origin_lxi+origin_usa+

mean_year_price_p1+mean_year_price_p1+mean_year_price_p3,
data = z_1)

summary(amount.model_1)

##
## Call:
## lm(formula = G.Total_1_yr ~ exp(cons_12m + 1) + exp(cons_gas_12m +
## 1) + exp(cons_last_month + 1) + exp(forecast_cons_12m + 1) +
## exp(forecast_discount_energy + 1) + exp(forecast_meter_rent_12m +
## 1) + forecast_price_energy_p1 + forecast_price_energy_p2 +
## forecast_price_pow_p1 + has_gas + exp(imp_cons + 1) + margin_gross_pow_ele
+
## margin_net_pow_ele + nb_prod_act + net_margin + pow_max +
## days_since + months_activ + months_to_end + months_modif_prod +
## months_renewal + activity_apd + activity_ckf + activity_clu +
## activity_cwo + activity_fmw + activity_kkk + activity_kwu +
## activity_sfi + activity_wxe + channel_ewp + channel_fix +
## channel_foo + channel_lmk + channel_sdd + channel_usi + origin_kam +
## origin_ldk + origin_lxi + origin_usa + mean_year_price_p1 +
## mean_year_price_p1 + mean_year_price_p3, data = z_1)
##
## Residuals:
## Min 1Q Median 3Q Max
## -3.383e+10 -1.739e+08 3.028e+08 5.818e+08 3.047e+10
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.337e+08 5.156e+08 -0.259 0.795417
## exp(cons_12m + 1) 1.143e+05 2.969e+05 0.385 0.700180
## exp(cons_gas_12m + 1) 1.856e+07 2.028e+05 91.519 < 2e-16 ***
## exp(cons_last_month + 1) 1.683e+07 7.391e+05 22.768 < 2e-16 ***
## exp(forecast_cons_12m + 1) -1.309e+06 1.387e+06 -0.944 0.345169
## exp(forecast_discount_energy + 1) -2.160e-15 1.457e-13 -0.015 0.988172
## exp(forecast_meter_rent_12m + 1) 8.294e+05 7.367e+06 0.113 0.910368
## forecast_price_energy_p1 -4.477e+07 1.673e+09 -0.027 0.978648

```



```
## forecast_price_energy_p2 2.256e+09 9.735e+08 2.318 0.020493 *
## forecast_price_pow_p1 1.667e+07 1.556e+07 1.071 0.284179
## has_gas -4.011e+09 9.352e+07 -42.883 < 2e-16 ***
## exp(imp_cons + 1) -3.635e+07 2.759e+06 -13.174 < 2e-16 ***
## margin_gross_pow_ele 1.498e+06 2.017e+06 0.743 0.457727
## margin_net_pow_ele -1.169e+06 1.651e+06 -0.708 0.478844
## nb_prod_act 7.215e+08 1.970e+07 36.624 < 2e-16 ***
## net_margin -1.431e+05 9.464e+04 -1.512 0.130564
## pow_max -1.116e+06 1.516e+06 -0.736 0.461925
## days_since 4.663e+06 2.220e+06 2.100 0.035709 *
## months_activ -1.427e+08 6.656e+07 -2.144 0.032092 *
## months_to_end -1.672e+08 6.673e+07 -2.505 0.012248 *
## months_modif_prod 8.737e+04 9.756e+05 0.090 0.928642
## months_renewal -2.656e+07 1.209e+07 -2.196 0.028095 *
## activity_apd -2.498e+09 1.372e+08 -18.209 < 2e-16 ***
## activity_ckf -1.041e+08 2.386e+08 -0.436 0.662543
## activity_clu -4.753e+08 2.921e+08 -1.627 0.103783
## activity_cwo -3.803e+07 3.069e+08 -0.124 0.901376
## activity_fmw 1.360e+08 2.199e+08 0.618 0.536346
## activity_kkk -1.484e+09 1.639e+08 -9.054 < 2e-16 ***
## activity_kwu -5.360e+08 2.259e+08 -2.373 0.017671 *
## activity_sfi -1.325e+09 3.652e+08 -3.627 0.000288 ***
## activity_wxe 2.222e+08 2.985e+08 0.744 0.456678
## channel_ewp -3.010e+07 1.269e+08 -0.237 0.812538
## channel_fix 1.556e+09 1.992e+09 0.781 0.434577
## channel_foo -1.896e+08 8.274e+07 -2.291 0.021969 *
## channel_lmk -3.488e+08 9.878e+07 -3.531 0.000415 ***
## channel_sdd -2.249e+08 1.001e+09 -0.225 0.822268
## channel_usi 2.012e+07 1.135e+08 0.177 0.859264
## origin_kam -5.367e+08 3.594e+08 -1.493 0.135375
## origin_ldk -4.479e+08 3.596e+08 -1.246 0.212893
## origin_lxi -6.226e+08 3.575e+08 -1.742 0.081606 .
## origin_usa -8.673e+08 2.022e+09 -0.429 0.668018
## mean_year_price_p1 -1.066e+07 1.487e+07 -0.717 0.473558
## mean_year_price_p3 -3.836e+07 9.104e+06 -4.213 2.54e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.812e+09 on 12016 degrees of freedom
## Multiple R-squared: 0.7232, Adjusted R-squared: 0.7222
## F-statistic: 747.4 on 42 and 12016 DF, p-value: < 2.2e-16

# An Adjusted R-squared of 72.2% means this model explains 72.2% of the
observed variance in the dependent variables by the Independent (Predictor)
variables
```

Create a "discounted_amount.model" for comparison with "amount.model"

```
discounted_amount.model = lm(formula =
discount_1~exp(cons_12m+1)+exp(cons_gas_12m+1)+exp(cons_last_month+1)+
```

```

exp(forecast_cons_12m+1)+exp(forecast_discount_energy+1)+exp(forecast_meter_rent_12m+1)+forecast_price_energy_p1+

forecast_price_energy_p2+forecast_price_pow_p1+
                                has_gas+exp(imp_cons+1)+

margin_gross_pow_ele+margin_net_pow_ele+nb_prod_act+net_margin+pow_max+days_since+

months_activ+months_to_end+months_modif_prod+months_renewal+

activity_apd+activity_ckf+activity_clu+activity_cwo+activity_fmw+

activity_kkk+activity_kwu+activity_sfi+activity_wxe+

channel_ewp+channel_fix+channel_foo+channel_lmk+channel_sdd+channel_usi+
                                origin_kam+origin_ldk+origin_lxi+origin_usa+
                                mean_year_price_p1+discount_2+discount_3,
                                data = z_1)

summary(discounted_amount.model)

##
## Call:
## lm(formula = discount_1 ~ exp(cons_12m + 1) + exp(cons_gas_12m +
## 1) + exp(cons_last_month + 1) + exp(forecast_cons_12m + 1) +
## exp(forecast_discount_energy + 1) + exp(forecast_meter_rent_12m +
## 1) + forecast_price_energy_p1 + forecast_price_energy_p2 +
## forecast_price_pow_p1 + has_gas + exp(imp_cons + 1) + margin_gross_pow_ele
+
## margin_net_pow_ele + nb_prod_act + net_margin + pow_max +
## days_since + months_activ + months_to_end + months_modif_prod +
## months_renewal + activity_apd + activity_ckf + activity_clu +
## activity_cwo + activity_fmw + activity_kkk + activity_kwu +
## activity_sfi + activity_wxe + channel_ewp + channel_fix +
## channel_foo + channel_lmk + channel_sdd + channel_usi + origin_kam +
## origin_ldk + origin_lxi + origin_usa + mean_year_price_p1 +
## discount_2 + discount_3, data = z_1)
##
## Residuals:
## Min 1Q Median 3Q Max
## -2.154e+10 -1.823e+08 1.657e+08 4.327e+08 3.019e+10
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.087e+09 3.574e+08 -3.042 0.002353 **
## exp(cons_12m + 1) -5.298e+04 2.133e+05 -0.248 0.803840
## exp(cons_gas_12m + 1) 1.141e+07 1.590e+05 71.777 < 2e-16 ***
## exp(cons_last_month + 1) 1.160e+07 5.324e+05 21.795 < 2e-16 ***

```

```

## exp(forecast_cons_12m + 1) 9.055e+05 9.957e+05 0.909 0.363192
## exp(forecast_discount_energy + 1) -5.005e-15 1.047e-13 -0.048 0.961869
## exp(forecast_meter_rent_12m + 1) -2.617e+07 4.437e+06 -5.899 3.76e-09 ***
## forecast_price_energy_p1 3.878e+09 1.108e+09 3.501 0.000465 ***
## forecast_price_energy_p2 1.405e+09 5.782e+08 2.429 0.015137 *
## forecast_price_pow_p1 9.628e+06 1.118e+07 0.861 0.389002
## has_gas -2.740e+09 6.776e+07 -40.431 < 2e-16 ***
## exp(imp_cons + 1) -2.500e+07 1.984e+06 -12.599 < 2e-16 ***
## margin_gross_pow_ele -9.184e+05 1.419e+06 -0.647 0.517407
## margin_net_pow_ele 2.921e+05 1.182e+06 0.247 0.804847
## nb_prod_act 6.314e+08 1.419e+07 44.499 < 2e-16 ***
## net_margin -1.436e+05 6.803e+04 -2.111 0.034824 *
## pow_max -4.178e+06 1.088e+06 -3.841 0.000123 ***
## days_since 4.993e+06 1.595e+06 3.130 0.001752 **
## months_activ -1.507e+08 4.782e+07 -3.151 0.001633 **
## months_to_end -1.706e+08 4.794e+07 -3.558 0.000375 ***
## months_modif_prod 1.315e+06 6.990e+05 1.882 0.059906 .
## months_renewal -2.574e+07 8.688e+06 -2.963 0.003057 **
## activity_apd -1.876e+09 9.863e+07 -19.016 < 2e-16 ***
## activity_ckf -6.427e+06 1.714e+08 -0.038 0.970085
## activity_clu -3.125e+08 2.098e+08 -1.489 0.136413
## activity_cwo -2.728e+06 2.205e+08 -0.012 0.990128
## activity_fmw 1.112e+08 1.580e+08 0.704 0.481330
## activity_kkk -8.867e+08 1.179e+08 -7.522 5.78e-14 ***
## activity_kwu -2.708e+08 1.623e+08 -1.669 0.095190 .
## activity_sfi -8.476e+08 2.624e+08 -3.230 0.001240 **
## activity_wxe 1.491e+08 2.143e+08 0.696 0.486582
## channel_ewp 1.139e+07 9.116e+07 0.125 0.900552
## channel_fix -7.499e+08 1.431e+09 -0.524 0.600388
## channel_foo -8.533e+07 5.935e+07 -1.438 0.150526
## channel_lmk -2.024e+08 7.100e+07 -2.851 0.004365 **
## channel_sdd -1.206e+08 7.193e+08 -0.168 0.866869
## channel_usi 2.504e+07 8.150e+07 0.307 0.758697
## origin_kam -2.465e+08 2.586e+08 -0.953 0.340449
## origin_ldk -1.230e+08 2.587e+08 -0.476 0.634423
## origin_lxi -2.835e+08 2.571e+08 -1.102 0.270272
## origin_usa -4.211e+08 1.453e+09 -0.290 0.771938
## mean_year_price_p1 2.673e+06 1.064e+07 0.251 0.801568
## discount_2 1.059e+00 7.002e-02 15.128 < 2e-16 ***
## discount_3 -4.909e-01 1.071e-01 -4.584 4.62e-06 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.02e+09 on 12015 degrees of freedom
## Multiple R-squared: 0.7768, Adjusted R-squared: 0.776
## F-statistic: 972.3 on 43 and 12015 DF, p-value: < 2.2e-16

```

Obtained an Adjusted R-squared of 77.6% (This model explains 77.6% of the variance in this dataset)

NOW PREDICTIONS ON THE “TEST” DATA SET USING MODELS CREATED

Using amount.model_1

```
test$prob_predicted = predict(prob.model, test, type = "probs")
```

```
test$revenue_predicted = predict(amount.model_1, test)
```

```
test$score_predicted = test$prob_predicted * test$revenue_predicted
```

```
head(test %>% select(1,64,65,66))
```

```
##                               id prob_predicted revenue_predicted
## 3  0010bcc39e42b3c2131ed2ce55246e3c      0.4981562      386558117
## 4  0010ee3855fdea87602a5b7aba8e42de      0.5000000      21598258569
## 7  0013f326a839a2f6ad87a1859952d227      0.5000000      1442119426
## 8  00184e957277eeef733a7b563fdabd06      0.5000000      317930729
## 16 002d70a2bdf9cf62f10d1efbea890f69      0.5000000      2142844291
## 19 003742573bb97760793ced15e3e11745      0.5000000      -490819165
##      score_predicted
## 3          192566336
## 4          10799129284
## 7           721059713
## 8          158965364
## 16         1071422145
## 19         -245409582
```

```
tail(test %>% select(1,64,65,66))
```

```
##                               id prob_predicted revenue_predicted
## 16083 ffcddcc071bb55e276f9dbcaa5241c0a      0.5          6238101
## 16084 ffcfa2e614242678d1330d052dab8cd9      0.5          1719732285
## 16086 ffd619330ca153638aca1f0c03dda2a2      0.5          710510478
## 16087 ffd81a648a4394d296acca247eb81019      0.5          1032240000
## 16090 ffeb3f48e6728e2f688073b77ad8703      0.5          -554597478
## 16091 ffeb6a979dd0b17a41076df1057e733      0.5          -68968989
##      score_predicted
## 16083          3119050
## 16084          859866143
## 16086          355255239
## 16087          516120000
## 16090         -277298739
## 16091         -34484494
```

```
summary(test$prob_predicted)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.4982  0.5000   0.5000   0.5000  0.5000   0.5000
```

This model predicts that there is a 50-50% chance/probability for everyone in this dataset to churn.

```
summary(test$revenue_predicted)
```

```
##      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
## -5.332e+09 -5.520e+08 -2.802e+08  7.213e+08  2.770e+08  5.383e+10
```

```
summary(test$score_predicted)
```

```
##      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
## -2.666e+09 -2.760e+08 -1.401e+08  3.606e+08  1.385e+08  2.692e+10
```

The mean value obtained in the "score_predicted" is important managerially because it tells the manager that the expected revenue in the future is equal to the mean value obtained above.

Applying the 20% discount proposed (i.e. 0.8) (i.e. Using the discounted model). A discount will be applied to everyone in the test dataset because everyone has a 50-50% chance of churning.

```
test$discounted_revenue_predicted = predict(discounted_amount.model,test)
```

```
test$discounted_score_predicted = test$prob_predicted *
test$discounted_revenue_predicted
```

Comparing "Scores Predicted" From the "amount.model" and "discounted_amount.model"

```
head(test %>% select(1,64,66,68))
```

```
##              id prob_predicted score_predicted
## 3  0010bcc39e42b3c2131ed2ce55246e3c      0.4981562      192566336
## 4  0010ee3855fdea87602a5b7aba8e42de      0.5000000      10799129284
## 7  0013f326a839a2f6ad87a1859952d227      0.5000000       721059713
## 8  00184e957277eef733a7b563fdabd06      0.5000000      158965364
## 16 002d70a2bdf9cf62f10d1efbea890f69      0.5000000      1071422145
## 19 003742573bb97760793ced15e3e11745      0.5000000     -245409582
## discounted_score_predicted
## 3              271422562
## 4             13240242368
## 7             413562697
## 8             144467014
## 16            629398627
## 19           -139309229
```

```
tail(test %>% select(1,64,66,68))
```

```
##              id prob_predicted score_predicted
## 16083 ffcddcc071bb55e276f9dbcaa5241c0a      0.5      3119050
## 16084 ffcfa2e614242678d1330d052dab8cd9      0.5      859866143
## 16086 ffd619330ca153638aca1f0c03dda2      0.5      355255239
## 16087 ffd81a648a4394d296acca247eb81019      0.5      516120000
```

```
## 16090 ffebf3f48e6728e2f688073b77ad8703      0.5      -277298739
## 16091 ffebf6a979dd0b17a41076df1057e733      0.5      -34484494
##      discounted_score_predicted
## 16083      64492740
## 16084      494395592
## 16086      301900519
## 16087      347013965
## 16090      -202922137
## 16091      73988771
```

```
summary(test$discounted_revenue_predicted)
```

```
##      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
## -3.708e+09 -4.214e+08 -1.519e+08  5.768e+08  2.350e+08  5.902e+10
```

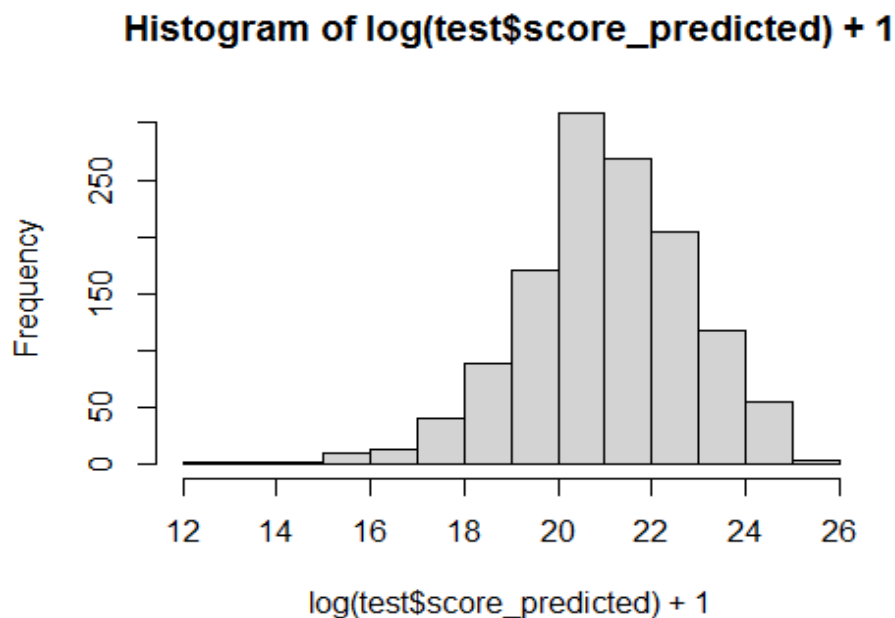
```
summary(test$discounted_score_predicted)
```

```
##      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
## -1.854e+09 -2.107e+08 -7.595e+07  2.884e+08  1.175e+08  2.951e+10
```

The mean value obtained in the "discounted_score_predicted" is important managerially; it tells the manager that expected revenue in the future is the mean value.

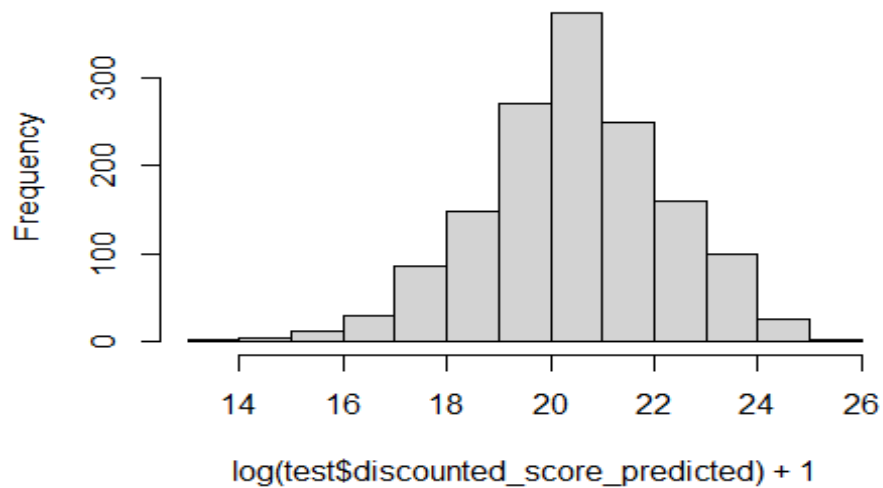
Comparing Histograms of "score_predicted" and that of "discounted_score_predicted"

```
hist(log(test$score_predicted)+1)
```



```
hist(log(test$discounted_score_predicted)+1)
```

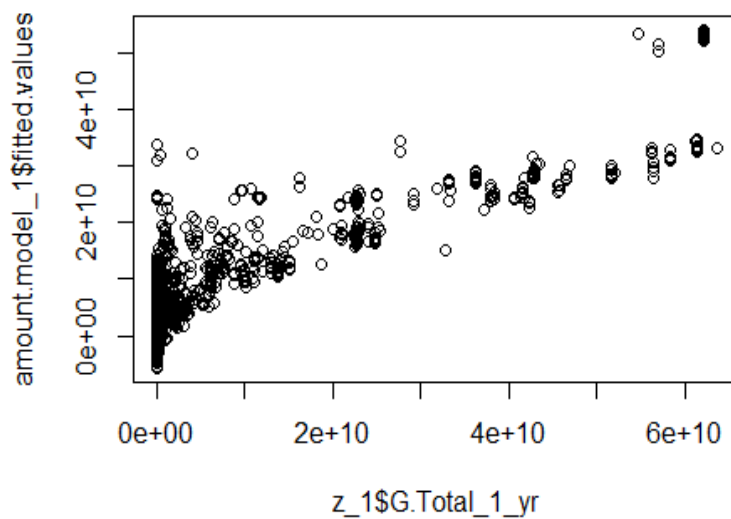
Histogram of $\log(\text{test\$discounted_score_predicted})$



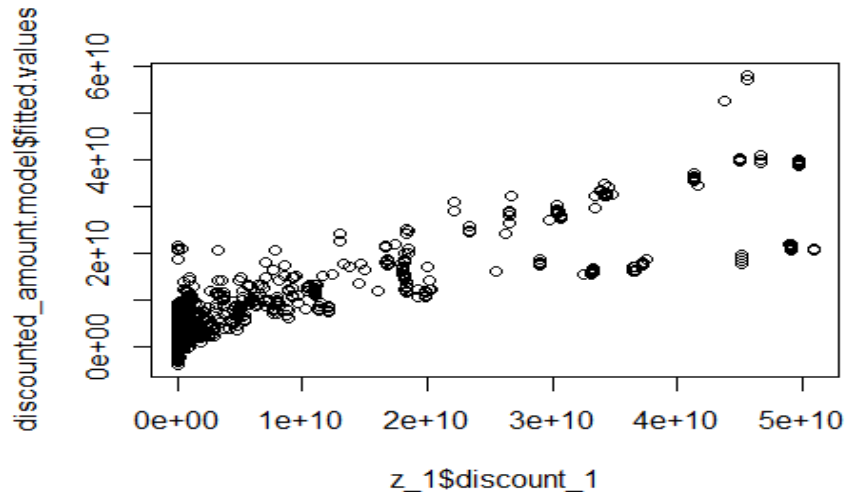
Comparing fitted values of the “amount_model” and that of the “discounted_amount.model”

Note; they look similar, just like the histograms of the scores predicted above.

```
plot(z_1$G.Total_1_yr, amount.model_1$fitted.values)
```



```
plot(z_1$discount_1, discounted_amount.model$fitted.values)
```



Testing statistical significance

Since differences in mean vales obtained from "score_predicted" and "discounted_score_predicted" seem not to be much; check if it is statistically significant by doing a t-test.

```
t.test(test$score_predicted, test$discounted_score_predicted, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  test$score_predicted and test$discounted_score_predicted
## t = 1.6304, df = 7786.9, p-value = 0.05153
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -650269.8      Inf
## sample estimates:
## mean of x mean of y
## 360635254 288407164
```

since the p-value obtained (0.05153) is greater than 0.05 (i.e. 95% C.I); we fail to reject the null hypothesis; because there is no statistically significant difference between the means of the two scores predicted (i.e. "score_predicted" and "discounted_score_predicted").

Variable Importance from the “caret package”

```
varImp(prob.model) %>% arrange(desc(Overall))
```

```
##                                Overall
## exp(forecast_discount_energy + 1) 2.538858e-16
## G.Total_1_yr                      8.388073e-26
## G.Total_6_months                  2.381104e-26
## G.Total_3_months                  1.524776e-26
## days_since                        1.789516e-31
## exp(cons_12m + 1)                 2.784812e-32
## net_margin                        1.824007e-32
## exp(cons_last_month + 1)          8.654673e-33
## exp(cons_gas_12m + 1)             6.155093e-33
## exp(forecast_cons_12m + 1)        5.789588e-33
## months_activ                      5.359887e-33
## forecast_price_pow_p1             3.792261e-33
## mean_year_price_p1                3.785315e-33
## months_modif_prod                 3.319117e-33
## margin_gross_pow_ele              1.810506e-33
## pow_max                           1.778248e-33
## margin_net_pow_ele                1.756500e-33
## exp(imp_cons + 1)                 1.564435e-33
## exp(forecast_meter_rent_12m + 1)  1.283889e-33
## mean_year_price_p2                8.981176e-34
## months_to_end                     6.046893e-34
## mean_year_price_p3                5.402142e-34
## months_renewal                    4.754810e-34
## nb_prod_act                       1.202191e-34
## origin_lxi                        3.932543e-35
## channel_foo                       3.748796e-35
## origin_kam                        2.705763e-35
## origin_ldk                        2.029322e-35
## has_gas                           1.679843e-35
## channel_lmk                       1.242093e-35
## forecast_price_energy_p1          1.184686e-35
## activity_apd                      9.511602e-36
## channel_usi                       7.746191e-36
## channel_ewp                       5.332262e-36
## forecast_price_energy_p2          4.523449e-36
## activity_kkk                      2.251800e-36
## activity_fmw                      1.107886e-36
## activity_ckf                      1.080864e-36
## activity_kwu                      1.044835e-36
## activity_clu                      6.665327e-37
## activity_cwo                      6.214967e-37
## activity_wxe                      5.944752e-37
## activity_sfi                      4.683744e-37
## channel_sdd                      7.205759e-38
## channel_fix                      1.801440e-38
## origin_usa                       1.801440e-38
```

This proves that price sensitivity is not the main driver for customer attrition rather forecasted discount and the total revenue obtainable for the year is.

CONCLUSION

The strategy suggested by the SME division to offer a 20% discount to all customers targeted might not be optimal.