

HW 4

Team 2

April 12, 2019

Contents

Model Building	4
MODEL 1	4
MODEL 2	13
MODEL 3	20
MODEL 4	27
MODEL 5	34
MODEL 6	40
Assessing selected models	46
AIC, BIC, Loglik, pseudoR2	47
Choosing the best model and applying it on test data set, saving results as a csv file	47
Assessing the performance of the choosen model: ROC and AUC	48
## Loading required package: broom	
## Loading required package: ggplot2	
## Loading required package: dplyr	
##	
## 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	
## Loading required package: tidyr	
## Loading required package: corrplot	
## corrplot 0.84 loaded	
## Loading required package: randomForest	
## randomForest 4.6-14	
## Type rfNews() to see new features/changes/bug fixes.	
##	
## Attaching package: 'randomForest'	
## The following object is masked from 'package:dplyr':	
##	
## combine	
## The following object is masked from 'package:ggplot2':	
##	
## margin	

```

## Loading required package: olsrr

##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
##     rivers

## Loading required package: caret
## Loading required package: lattice
## Loading required package: fastDummies
## Loading required package: car
## Loading required package: carData

##
## Attaching package: 'car'

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

## Loading required package: pROC
## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var

## Loading required package: pscl

## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis

Logistic models

## 'data.frame':   8161 obs. of  26 variables:
## $ INDEX      : int  1 2 4 5 6 7 8 11 12 13 ...
## $ TARGET_FLAG: int  0 0 0 0 0 1 0 1 1 0 ...
## $ TARGET_AMT : num  0 0 0 0 0 ...
## $ KIDSDRIV   : int  0 0 0 0 0 0 0 1 0 0 ...
## $ AGE        : int  60 43 35 51 50 34 54 37 34 50 ...
## $ HOMEKIDS   : int  0 0 1 0 0 1 0 2 0 0 ...
## $ YOJ        : int  11 11 10 14 NA 12 NA NA 10 7 ...
## $ INCOME     : Factor w/ 6613 levels "", "$0 ", "$1,007 ",...: 5033 6292 1250 1 509 746 1488 315 4765 1
## $ PARENT1    : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 2 1 1 1 1 ...
## $ HOME_VAL   : Factor w/ 5107 levels "", "$0 ", "$100,093 ",...: 2 3259 348 3917 3034 2 1 4167 2 2 ...
## $ MSTATUS    : Factor w/ 2 levels "Yes","z_No": 2 2 1 1 1 2 1 1 2 2 ...
## $ SEX        : Factor w/ 2 levels "M","z_F": 1 1 2 1 2 2 2 1 2 1 ...
## $ EDUCATION  : Factor w/ 5 levels "<High School",...: 4 5 5 1 4 2 1 2 2 2 ...

```

```

## $ JOB      : Factor w/ 9 levels "", "Clerical",...: 7 9 2 9 3 9 9 9 2 7 ...
## $ TRAVTIME : int  14 22 5 32 36 46 33 44 34 48 ...
## $ CAR_USE  : Factor w/ 2 levels "Commercial", "Private": 2 1 2 2 2 1 2 1 2 1 ...
## $ BLUEBOOK : Factor w/ 2789 levels "$1,500 ", "$1,520 ",...: 434 503 2212 553 802 746 2672 701 135 8
## $ TIF      : int  11 1 4 7 1 1 1 1 1 7 ...
## $ CAR_TYPE : Factor w/ 6 levels "Minivan", "Panel Truck",...: 1 1 6 1 6 4 6 5 6 5 ...
## $ RED_CAR  : Factor w/ 2 levels "no", "yes": 2 2 1 2 1 1 1 2 1 1 ...
## $ OLDCLAIM : Factor w/ 2857 levels "$0 ", "$1,000 ",...: 1449 1 1311 1 432 1 1 510 1 1 ...
## $ CLM_FREQ : int  2 0 2 0 2 0 0 1 0 0 ...
## $ REVOKED  : Factor w/ 2 levels "No", "Yes": 1 1 1 1 2 1 1 2 1 1 ...
## $ MVR_PTS  : int  3 0 3 0 3 0 0 10 0 1 ...
## $ CAR_AGE  : int  18 1 10 6 17 7 1 7 1 17 ...
## $ URBANITY : Factor w/ 2 levels "Highly Urban/ Urban",...: 1 1 1 1 1 1 1 1 1 2 ...

##      INDEX      TARGET_FLAG      TARGET_AMT      KIDSDRIV
## Min.      : 1      Min.      :0.0000      Min.      : 0      Min.      :0.0000
## 1st Qu.: 2559      1st Qu.:0.0000      1st Qu.: 0      1st Qu.:0.0000
## Median : 5133      Median :0.0000      Median : 0      Median :0.0000
## Mean    : 5152      Mean    :0.2638      Mean    : 1504      Mean    :0.1711
## 3rd Qu.: 7745      3rd Qu.:1.0000      3rd Qu.: 1036      3rd Qu.:0.0000
## Max.    :10302      Max.    :1.0000      Max.    :107586      Max.    :4.0000
##
##      AGE      HOMEKIDS      YOJ      INCOME
## Min.      :16.00      Min.      :0.0000      Min.      : 0.0      $0      : 615
## 1st Qu.:39.00      1st Qu.:0.0000      1st Qu.: 9.0      : 445
## Median :45.00      Median :0.0000      Median :11.0      $26,840 : 4
## Mean    :44.79      Mean    :0.7212      Mean    :10.5      $48,509 : 4
## 3rd Qu.:51.00      3rd Qu.:1.0000      3rd Qu.:13.0      $61,790 : 4
## Max.    :81.00      Max.    :5.0000      Max.    :23.0      $107,375 : 3
## NA's    :6      NA's    :454      (Other) :7086
##
## PARENT1      HOME_VAL      MSTATUS      SEX      EDUCATION
## No :7084      $0      :2294      Yes :4894      M :3786      <High School :1203
## Yes:1077      : 464      z_No:3267      z_F:4375      Bachelors :2242
##      $111,129 : 3      Masters :1658
##      $115,249 : 3      PhD : 728
##      $123,109 : 3      z_High School:2330
##      $153,061 : 3
##      (Other) :5391
##
##      JOB      TRAVTIME      CAR_USE      BLUEBOOK
## z_Blue Collar:1825      Min.      : 5.00      Commercial:3029      $1,500 : 157
## Clerical :1271      1st Qu.: 22.00      Private :5132      $6,000 : 34
## Professional :1117      Median : 33.00      :5,800 : 33
## Manager : 988      Mean : 33.49      :$6,200 : 33
## Lawyer : 835      3rd Qu.: 44.00      :$6,400 : 31
## Student : 712      Max. :142.00      :$5,900 : 30
## (Other) :1413      : (Other):7843
##
##      TIF      CAR_TYPE      RED_CAR      OLDCLAIM
## Min.      : 1.000      Minivan :2145      no :5783      $0 :5009
## 1st Qu.: 1.000      Panel Truck: 676      yes:2378      $1,310 : 4
## Median : 4.000      Pickup :1389      :$1,391 : 4
## Mean : 5.351      Sports Car : 907      :$4,263 : 4
## 3rd Qu.: 7.000      Van : 750      :$1,105 : 3
## Max. :25.000      z_SUV :2294      :$1,332 : 3
##      (Other):3134

```

```
##      CLM_FREQ      REVOKED      MVR_PTS      CAR_AGE
##  Min.    :0.0000    No :7161    Min.    : 0.000    Min.    : -3.000
##  1st Qu.:0.0000    Yes:1000    1st Qu.: 0.000    1st Qu.:  1.000
##  Median :0.0000                    Median :  1.000    Median :  8.000
##  Mean   :0.7986                    Mean   :  1.696    Mean   :  8.328
##  3rd Qu.:2.0000                    3rd Qu.:  3.000    3rd Qu.:12.000
##  Max.   :5.0000                    Max.   :13.000    Max.   :28.000
##                                     NA's    :510
##
##      URBANICITY
##  Highly Urban/ Urban :6492
##  z_Highly Rural/ Rural:1669
##
##
##
##
##
##
##      INDEX TARGET_FLAG TARGET_AMT  KIDSDRIV      AGE  HOMEKIDS
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0735204  0.0000000
##      YOJ      INCOME      PARENT1  HOME_VAL  MSTATUS      SEX
##  5.5630437  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##  EDUCATION      JOB      TRAVTIME  CAR_USE  BLUEBOOK      TIF
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##  CAR_TYPE  RED_CAR  OLDCLAIM  CLM_FREQ  REVOKED  MVR_PTS
##  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
##  CAR_AGE  URBANICITY
##  6.2492342  0.0000000
```

Data Cleaning

1. Removing \$ sign and commas
2. Cleaning data set from “_z” characters
3. Removing rows with NAs

Transforming TARGET_FLAG, KIDSDRIV and HOMEKIDS to factors.

Model Building

MODEL 1

1. Building logistic regression model using variables without any transformations.

```
model_1<-step(glm(TARGET_FLAG~., training %>% select(-INDEX, -TARGET_AMT), family = "binomial"), direct

## Start:  AIC=5968.06
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + TRAVTIME + CAR_USE +
##      BLUEBOOK + TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ +
##      REVOKED + MVR_PTS + CAR_AGE + URBANICITY

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

##           Df Deviance    AIC
## - TRAVTIME  95   5554.6 5918.6
## - CAR_AGE   29   5447.3 5943.3
```

```

## - YOJ          19    5439.8 5955.8
## - HOMEKIDS     5     5419.9 5963.9
## - SEX          1     5415.3 5967.3
## - INCOME       1     5415.5 5967.5
## <none>         5414.1 5968.1
## - RED_CAR      1     5416.4 5968.4
## - PARENT1      1     5417.1 5969.1
## - HOME_VAL     1     5426.4 5978.4
## - OLDCLAIM     1     5426.9 5978.9
## - EDUCATION    3     5432.6 5980.6
## - AGE          57     5540.7 5980.7
## - BLUEBOOK     1     5432.6 5984.6
## - MVR_PTS     12     5454.8 5984.8
## - KIDSDRIV     4     5443.9 5989.9
## - MSTATUS      1     5440.4 5992.4
## - TIF          22     5488.2 5998.2
## - CLM_FREQ     5     5465.0 6009.0
## - JOB          8     5478.4 6016.4
## - CAR_USE      1     5478.9 6030.9
## - CAR_TYPE     5     5488.9 6032.9
## - REVOKED      1     5483.0 6035.0
## - URBANICITY   1     5856.9 6408.9
##
## Step:  AIC=5918.56
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + CAR_AGE + URBANICITY
##
##           Df Deviance    AIC
## - CAR_AGE   29    5587.4 5893.4
## - YOJ       19    5584.1 5910.1
## - HOMEKIDS   5    5560.6 5914.6
## - INCOME     1    5556.4 5918.4
## <none>       5554.6 5918.6
## - SEX        1    5556.9 5918.9
## - RED_CAR    1    5557.4 5919.4
## - PARENT1    1    5557.4 5919.4
## - EDUCATION  3    5570.3 5928.3
## - HOME_VAL   1    5567.2 5929.2
## - AGE       57    5680.7 5930.7
## - OLDCLAIM   1    5569.3 5931.3
## - BLUEBOOK   1    5569.8 5931.8
## - MVR_PTS   12    5596.1 5936.1
## - KIDSDRIV   4    5582.5 5938.5
## - MSTATUS    1    5580.6 5942.6
## - TIF       22    5629.9 5949.9
## - CLM_FREQ   5    5610.8 5964.8
## - JOB        8    5621.7 5969.7
## - CAR_USE    1    5617.3 5979.3
## - CAR_TYPE   5    5629.8 5983.8
## - REVOKED    1    5624.0 5986.0
## - URBANICITY 1    5974.8 6336.8
##

```

```

## Step: AIC=5893.41
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##     HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##     TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##     MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - YOJ      19   5616.6 5884.6
## - HOMEKIDS   5   5594.1 5890.1
## - INCOME     1   5589.2 5893.2
## <none>      5587.4 5893.4
## - SEX        1   5589.5 5893.5
## - PARENT1    1   5589.7 5893.7
## - RED_CAR    1   5590.2 5894.2
## - HOME_VAL   1   5598.5 5902.5
## - AGE       57   5712.2 5904.2
## - OLDCLAIM   1   5602.3 5906.3
## - BLUEBOOK   1   5602.4 5906.4
## - MVR_PTS    12   5628.2 5910.2
## - EDUCATION   3   5612.4 5912.4
## - KIDSDRIV    4   5616.7 5914.7
## - MSTATUS     1   5616.3 5920.3
## - TIF        22   5663.0 5925.0
## - CLM_FREQ    5   5645.1 5941.1
## - JOB         8   5655.2 5945.2
## - CAR_USE     1   5652.8 5956.8
## - CAR_TYPE    5   5662.2 5958.2
## - REVOKED     1   5658.2 5962.2
## - URBANICITY  1   6007.4 6311.4
##
## Step: AIC=5884.61
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + INCOME + PARENT1 +
##     HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##     TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##     MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - HOMEKIDS    5   5622.3 5880.3
## <none>      5616.6 5884.6
## - PARENT1     1   5618.6 5884.6
## - SEX          1   5618.7 5884.7
## - RED_CAR      1   5619.0 5885.0
## - INCOME       1   5619.9 5885.9
## - HOME_VAL     1   5627.4 5893.4
## - AGE         58   5744.5 5896.5
## - BLUEBOOK     1   5631.1 5897.1
## - OLDCLAIM     1   5631.2 5897.2
## - EDUCATION    3   5641.4 5903.4
## - MVR_PTS     12   5660.0 5904.0
## - KIDSDRIV     4   5645.5 5905.5
## - MSTATUS      1   5645.3 5911.3
## - TIF         22   5694.4 5918.4
## - JOB          8   5677.9 5929.9
## - CLM_FREQ     5   5673.1 5931.1

```

```

## - CAR_USE      1    5684.0 5950.0
## - CAR_TYPE     5    5692.2 5950.2
## - REVOKED      1    5689.0 5955.0
## - URBANICITY   1    6036.6 6302.6
##
## Step: AIC=5880.32
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##      MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF +
##      CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS +
##      URBANICITY
##
##           Df Deviance    AIC
## - SEX      1    5624.2 5880.2
## <none>      5622.3 5880.3
## - RED_CAR   1    5624.7 5880.7
## - INCOME    1    5625.5 5881.5
## - PARENT1   1    5628.4 5884.4
## - HOME_VAL  1    5633.4 5889.4
## - OLDCLAIM  1    5636.1 5892.1
## - BLUEBOOK  1    5637.3 5893.3
## - AGE       58    5755.3 5897.3
## - EDUCATION  3    5646.7 5898.7
## - MVR_PTS   12    5666.5 5900.5
## - MSTATUS   1    5650.6 5906.6
## - KIDSDRIV  4    5660.8 5910.8
## - TIF       22    5699.7 5913.7
## - JOB        8    5683.8 5925.8
## - CLM_FREQ  5    5678.1 5926.1
## - CAR_TYPE  5    5697.4 5945.4
## - CAR_USE   1    5689.6 5945.6
## - REVOKED   1    5693.7 5949.7
## - URBANICITY 1    6042.2 6298.2
##
## Step: AIC=5880.19
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##      MSTATUS + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
##      RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - RED_CAR   1    5625.3 5879.3
## <none>      5624.2 5880.2
## - INCOME    1    5627.4 5881.4
## - PARENT1   1    5630.2 5884.2
## - HOME_VAL  1    5635.1 5889.1
## - OLDCLAIM  1    5638.1 5892.1
## - AGE       58    5758.0 5898.0
## - EDUCATION  3    5648.6 5898.6
## - MVR_PTS   12    5668.1 5900.1
## - BLUEBOOK  1    5646.8 5900.8
## - MSTATUS   1    5652.8 5906.8
## - KIDSDRIV  4    5662.2 5910.2
## - TIF       22    5701.1 5913.1
## - JOB        8    5685.2 5925.2
## - CLM_FREQ  5    5680.2 5926.2

```

```

## - CAR_USE      1    5691.2 5945.2
## - CAR_TYPE     5    5700.0 5946.0
## - REVOKED      1    5695.7 5949.7
## - URBANICITY   1    6043.9 6297.9
##
## Step: AIC=5879.31
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##      MSTATUS + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
##      OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY
##
##              Df Deviance    AIC
## <none>              5625.3 5879.3
## - INCOME           1    5628.6 5880.6
## - PARENT1          1    5631.5 5883.5
## - HOME_VAL         1    5636.0 5888.0
## - OLDCLAIM         1    5639.3 5891.3
## - AGE              58    5758.2 5896.2
## - EDUCATION        3    5649.9 5897.9
## - BLUEBOOK         1    5646.8 5898.8
## - MVR_PTS         12    5669.4 5899.4
## - MSTATUS          1    5653.8 5905.8
## - KIDSDRIV         4    5663.4 5909.4
## - TIF              22    5702.3 5912.3
## - JOB              8    5686.8 5924.8
## - CLM_FREQ         5    5681.3 5925.3
## - CAR_USE          1    5692.4 5944.4
## - REVOKED          1    5696.7 5948.7
## - CAR_TYPE         5    5712.8 5956.8
## - URBANICITY       1    6044.9 6296.9

summary(model_1)

##
## Call:
## glm(formula = TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS +
##      URBANICITY, family = "binomial", data = training %>% select(-INDEX,
##      -TARGET_AMT))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5621  -0.6928  -0.3840   0.5741   3.0962
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.575e+00  1.310e+00  -1.202  0.229283
## KIDSDRIV1      5.495e-01  1.283e-01   4.284  1.84e-05 ***
## KIDSDRIV2      8.263e-01  1.771e-01   4.667  3.06e-06 ***
## KIDSDRIV3      8.989e-01  3.376e-01   2.662  0.007762 **
## KIDSDRIV4     -1.095e+01  5.218e+02  -0.021  0.983255
## AGE18         -1.326e+00  1.913e+00  -0.693  0.488268
## AGE19         -3.497e-02  1.760e+00  -0.020  0.984149
## AGE20         -1.922e+00  1.857e+00  -1.035  0.300674
## AGE21         -3.968e-01  1.492e+00  -0.266  0.790206

```


## AGE22	-4.770e-01	1.432e+00	-0.333	0.739075
## AGE23	-9.352e-02	1.505e+00	-0.062	0.950467
## AGE24	7.723e-01	1.457e+00	0.530	0.596145
## AGE25	5.064e-01	1.396e+00	0.363	0.716726
## AGE26	1.441e-01	1.353e+00	0.107	0.915167
## AGE27	-6.181e-01	1.329e+00	-0.465	0.641794
## AGE28	-7.625e-01	1.321e+00	-0.577	0.563919
## AGE29	-1.527e+00	1.320e+00	-1.157	0.247371
## AGE30	-9.278e-01	1.318e+00	-0.704	0.481391
## AGE31	-9.272e-01	1.305e+00	-0.710	0.477453
## AGE32	-1.009e+00	1.302e+00	-0.775	0.438217
## AGE33	-1.302e+00	1.298e+00	-1.003	0.315818
## AGE34	-1.424e+00	1.298e+00	-1.097	0.272726
## AGE35	-1.123e+00	1.293e+00	-0.868	0.385255
## AGE36	-1.041e+00	1.292e+00	-0.805	0.420750
## AGE37	-1.161e+00	1.292e+00	-0.899	0.368676
## AGE38	-9.992e-01	1.291e+00	-0.774	0.438954
## AGE39	-1.048e+00	1.290e+00	-0.812	0.416795
## AGE40	-1.526e+00	1.290e+00	-1.183	0.236611
## AGE41	-1.411e+00	1.290e+00	-1.094	0.274149
## AGE42	-1.201e+00	1.289e+00	-0.931	0.351611
## AGE43	-1.452e+00	1.290e+00	-1.125	0.260479
## AGE44	-1.732e+00	1.291e+00	-1.341	0.179824
## AGE45	-1.412e+00	1.289e+00	-1.096	0.273237
## AGE46	-1.481e+00	1.290e+00	-1.148	0.250949
## AGE47	-1.549e+00	1.291e+00	-1.200	0.230318
## AGE48	-1.655e+00	1.291e+00	-1.282	0.199819
## AGE49	-1.599e+00	1.292e+00	-1.238	0.215854
## AGE50	-1.387e+00	1.292e+00	-1.074	0.283013
## AGE51	-1.622e+00	1.295e+00	-1.252	0.210449
## AGE52	-1.295e+00	1.295e+00	-1.000	0.317166
## AGE53	-1.316e+00	1.294e+00	-1.016	0.309501
## AGE54	-1.198e+00	1.297e+00	-0.923	0.355873
## AGE55	-1.469e+00	1.300e+00	-1.130	0.258422
## AGE56	-1.212e+00	1.305e+00	-0.929	0.353020
## AGE57	-5.848e-01	1.299e+00	-0.450	0.652507
## AGE58	-5.293e-01	1.308e+00	-0.405	0.685708
## AGE59	-4.454e-01	1.322e+00	-0.337	0.736223
## AGE60	-4.377e-01	1.316e+00	-0.333	0.739441
## AGE61	-7.757e-01	1.325e+00	-0.585	0.558314
## AGE62	-9.034e-01	1.337e+00	-0.676	0.499242
## AGE63	-8.643e-01	1.347e+00	-0.642	0.521192
## AGE64	-2.867e-01	1.376e+00	-0.208	0.834932
## AGE65	-1.387e+00	1.466e+00	-0.946	0.344266
## AGE66	-1.290e+00	1.446e+00	-0.892	0.372443
## AGE67	-2.105e-01	1.529e+00	-0.138	0.890475
## AGE68	-4.814e-01	1.790e+00	-0.269	0.787987
## AGE69	-6.551e-01	1.759e+00	-0.373	0.709480
## AGE70	-1.400e+01	4.001e+02	-0.035	0.972092
## AGE72	3.295e-01	1.902e+00	0.173	0.862476
## AGE73	-8.550e-01	1.962e+00	-0.436	0.663046
## AGE76	1.745e+01	8.827e+02	0.020	0.984232
## AGE80	-1.279e+01	8.827e+02	-0.014	0.988441
## AGE81	-1.380e+01	8.827e+02	-0.016	0.987524

## INCOME	-2.303e-06	1.284e-06	-1.793	0.072943	.
## PARENT1Yes	2.895e-01	1.167e-01	2.481	0.013103	*
## HOME_VAL	-1.294e-06	3.959e-07	-3.268	0.001083	**
## MSTATUSYes	-5.153e-01	9.588e-02	-5.374	7.70e-08	***
## EDUCATIONHigh School	4.523e-01	9.430e-02	4.797	1.61e-06	***
## EDUCATIONMasters	5.554e-02	1.576e-01	0.352	0.724531	
## EDUCATIONPhD	2.194e-01	2.041e-01	1.075	0.282382	
## JOBBBlue Collar	3.740e-01	2.159e-01	1.732	0.083261	.
## JOBClerical	5.228e-01	2.289e-01	2.284	0.022399	*
## JOBDoctor	-2.656e-01	2.928e-01	-0.907	0.364448	
## JOBHome Maker	3.352e-01	2.404e-01	1.395	0.163124	
## JOBLawyer	2.532e-01	1.945e-01	1.302	0.193004	
## JOBManager	-5.837e-01	2.006e-01	-2.909	0.003623	**
## JOBProfessional	2.210e-01	2.066e-01	1.070	0.284730	
## JOBStudent	2.214e-01	2.487e-01	0.890	0.373252	
## CAR_USEPrivate	-8.182e-01	1.009e-01	-8.105	5.26e-16	***
## BLUEBOOK	-2.507e-05	5.464e-06	-4.589	4.46e-06	***
## TIF10	-7.405e-01	1.307e-01	-5.664	1.48e-08	***
## TIF11	-4.095e-01	2.016e-01	-2.031	0.042218	*
## TIF12	-4.316e-01	4.578e-01	-0.943	0.345758	
## TIF13	-4.232e-01	1.944e-01	-2.176	0.029522	*
## TIF14	-1.066e+00	4.201e-01	-2.537	0.011195	*
## TIF15	1.286e-01	4.847e-01	0.265	0.790791	
## TIF16	-1.205e+00	5.378e-01	-2.241	0.025037	*
## TIF17	-5.741e-01	3.102e-01	-1.851	0.064163	.
## TIF18	-8.457e-01	6.588e-01	-1.284	0.199235	
## TIF19	-1.296e+01	3.611e+02	-0.036	0.971360	
## TIF2	2.763e-01	9.822e-01	0.281	0.778435	
## TIF20	-1.388e+01	3.172e+02	-0.044	0.965090	
## TIF21	2.132e-01	7.520e-01	0.284	0.776786	
## TIF22	-1.460e+01	4.790e+02	-0.030	0.975684	
## TIF25	-1.239e+01	6.029e+02	-0.021	0.983602	
## TIF3	-1.696e-01	1.500e-01	-1.131	0.258253	
## TIF4	-2.484e-01	1.031e-01	-2.410	0.015949	*
## TIF5	-6.463e-01	4.433e-01	-1.458	0.144845	
## TIF6	-4.073e-01	1.014e-01	-4.016	5.92e-05	***
## TIF7	-5.672e-01	1.388e-01	-4.088	4.36e-05	***
## TIF8	5.455e-01	3.939e-01	1.385	0.166070	
## TIF9	-6.856e-01	2.213e-01	-3.098	0.001948	**
## CAR_TYPEPanel Truck	6.873e-01	1.716e-01	4.005	6.20e-05	***
## CAR_TYPEPickup	5.284e-01	1.150e-01	4.593	4.36e-06	***
## CAR_TYPESports Car	1.005e+00	1.258e-01	7.988	1.37e-15	***
## CAR_TYPESUV	7.341e-01	9.938e-02	7.387	1.51e-13	***
## CAR_TYPEVan	6.706e-01	1.404e-01	4.776	1.79e-06	***
## OLDCLAIM	-1.771e-05	4.786e-06	-3.700	0.000215	***
## CLM_FREQ1	6.004e-01	1.161e-01	5.172	2.31e-07	***
## CLM_FREQ2	6.523e-01	1.084e-01	6.016	1.79e-09	***
## CLM_FREQ3	6.929e-01	1.216e-01	5.696	1.22e-08	***
## CLM_FREQ4	8.242e-01	1.985e-01	4.153	3.28e-05	***
## CLM_FREQ5	1.184e+00	6.484e-01	1.825	0.067953	.
## REVOKEDYes	9.096e-01	1.068e-01	8.518	< 2e-16	***
## MVR_PTS1	8.386e-02	1.050e-01	0.798	0.424651	
## MVR_PTS10	1.045e+00	8.799e-01	1.188	0.234804	
## MVR_PTS11	1.817e+00	1.132e+00	1.606	0.108356	

```

## MVR_PTS13          1.552e+01  5.536e+02  0.028 0.977638
## MVR_PTS2           2.448e-01  1.100e-01  2.226 0.025982 *
## MVR_PTS3           3.048e-01  1.185e-01  2.573 0.010084 *
## MVR_PTS4           2.905e-01  1.284e-01  2.262 0.023688 *
## MVR_PTS5           1.689e-01  1.481e-01  1.140 0.254156
## MVR_PTS6           2.389e-01  1.775e-01  1.346 0.178366
## MVR_PTS7           6.357e-01  2.065e-01  3.078 0.002083 **
## MVR_PTS8           1.240e+00  3.194e-01  3.883 0.000103 ***
## MVR_PTS9           1.248e+00  3.999e-01  3.122 0.001799 **
## URBANICITYHighly Urban/ Urban 2.216e+00  1.268e-01 17.474 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 7445.1  on 6447  degrees of freedom
## Residual deviance: 5625.3  on 6321  degrees of freedom
## AIC: 5879.3
##
## Number of Fisher Scoring iterations: 13

```

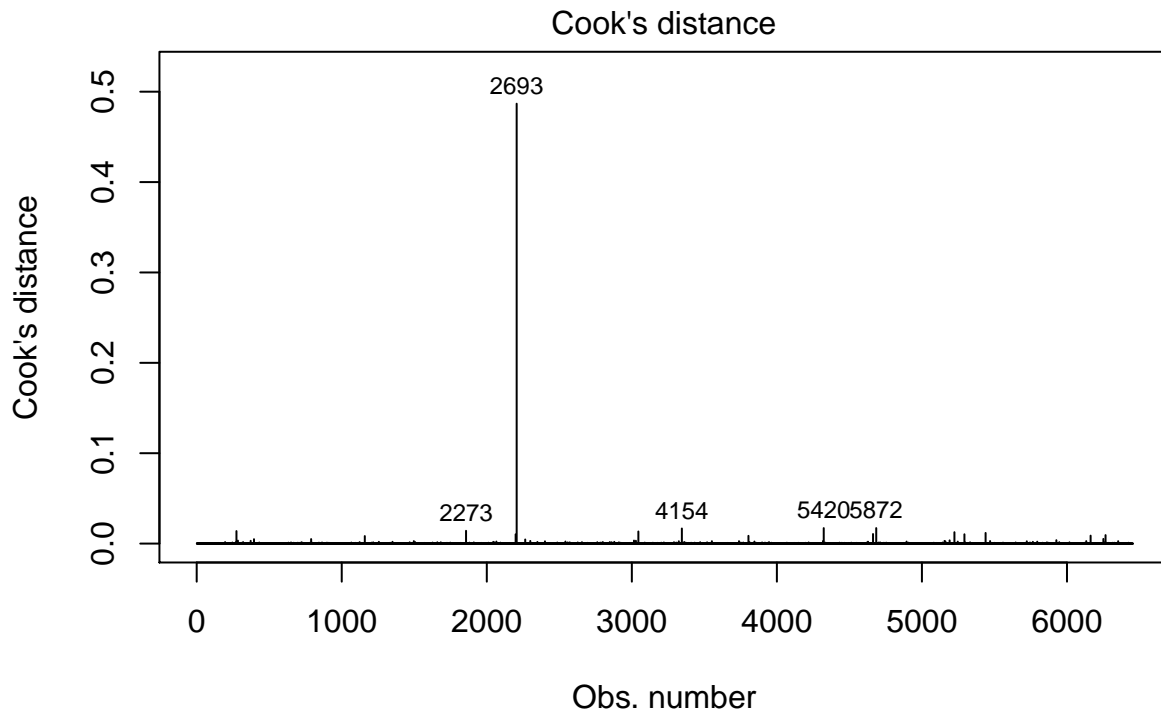
```
vif(model_1)
```

```

##              GVIF Df GVIF^(1/(2*Df))
## KIDSDRIV      1.374893  4      1.040599
## AGE           3.731650 58      1.011417
## INCOME        2.930509  1      1.711873
## PARENT1       1.705264  1      1.305858
## HOME_VAL      2.091848  1      1.446322
## MSTATUS       2.081513  1      1.442745
## EDUCATION     8.075751  3      1.416437
## JOB           23.379420  8      1.217733
## CAR_USE       2.283311  1      1.511063
## BLUEBOOK      1.834626  1      1.354484
## TIF           1.461604 22      1.008663
## CAR_TYPE      2.972147  5      1.115083
## OLDCLAIM      1.941762  1      1.393471
## CLM_FREQ      2.066599  5      1.075290
## REVOKED       1.392372  1      1.179988
## MVR_PTS       1.676936 12      1.021774
## URBANICITY    1.156896  1      1.075591

```

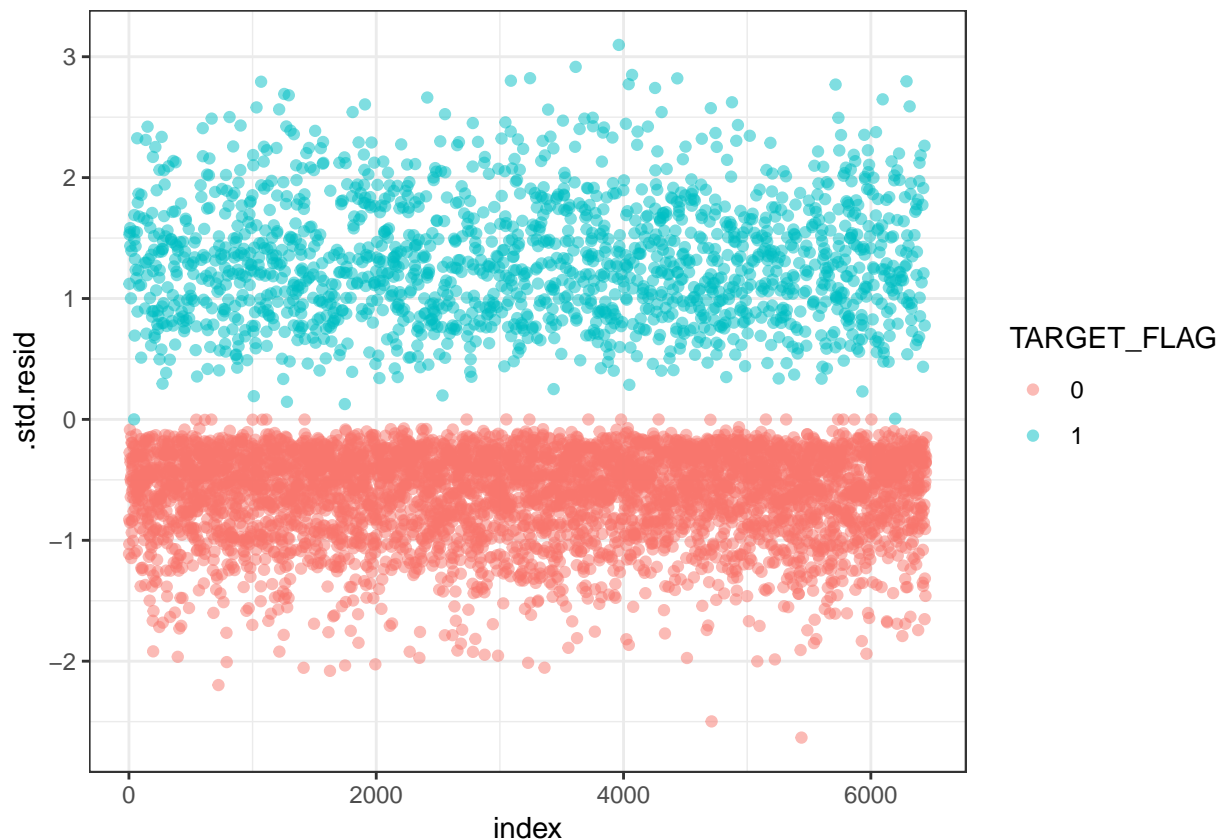
Checking model_1 for the presence of influential values.



$n(\text{TARGET_FLAG} \sim \text{KIDSDRIV} + \text{AGE} + \text{INCOME} + \text{PARENT1} + \text{HOME_VAL} + \text{MSTATUS})$

```
## # A tibble: 5 x 27
##   .rownames TARGET_FLAG KIDSDRIV AGE  INCOME PARENT1 HOME_VAL MSTATUS
##   <chr>      <fct>      <fct> <fct> <dbl> <fct>      <dbl> <fct>
## 1 2273      0          0      16      0 Yes          0 No
## 2 2693      0          4      59    79178 No        290382 Yes
## 3 4154      1          0      20     4209 No          80790 No
## 4 5420      1          0      18      479 No           0 Yes
## 5 5872      0          0      18     6284 Yes           0 No
## # ... with 19 more variables: EDUCATION <fct>, JOB <fct>, CAR_USE <fct>,
## #   BLUEBOOK <dbl>, TIF <fct>, CAR_TYPE <fct>, OLDCLAIM <dbl>,
## #   CLM_FREQ <fct>, REVOKED <fct>, MVRPTS <fct>, URBANICITY <fct>,
## #   .fitted <dbl>, .se.fit <dbl>, .resid <dbl>, .hat <dbl>, .sigma <dbl>,
## #   .cooks_d <dbl>, .std.resid <dbl>, index <int>

## Warning: Removed 3 rows containing missing values (geom_point).
```



```
## # A tibble: 1 x 27
##   .rownames TARGET_FLAG KIDSDRIV AGE INCOME PARENT1 HOME_VAL MSTATUS
##   <chr>      <fct>      <fct> <fct> <dbl> <fct>      <dbl> <fct>
## 1 4958      1          0      44      0 No          0 Yes
## # ... with 19 more variables: EDUCATION <fct>, JOB <fct>, CAR_USE <fct>,
## #   BLUEBOOK <dbl>, TIF <fct>, CAR_TYPE <fct>, OLDCLAIM <dbl>,
## #   CLM_FREQ <fct>, REVOKED <fct>, MVRPTS <fct>, URBANICITY <fct>,
## #   .fitted <dbl>, .se.fit <dbl>, .resid <dbl>, .hat <dbl>, .sigma <dbl>,
## #   .cooksdi <dbl>, .std.resid <dbl>, index <int>
```

Eliminating the row from training data set with influential value.

```
training_no_infl <- training %>%
  filter(!(INDEX==4958 & AGE==44))
```

MODEL 2

2. Building a model with removed influential values.

```
model_2 <- step(glm(TARGET_FLAG ~ ., training_no_infl %>% select(-INDEX, -TARGET_AMT), family = "binomial"))
```

```
## Start: AIC=5968.06
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##   HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + TRAVTIME + CAR_USE +
##   BLUEBOOK + TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ +
##   REVOKED + MVRPTS + CAR_AGE + URBANICITY
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```

##          Df Deviance    AIC
## - TRAVTIME 95    5554.6 5918.6
## - CAR_AGE  29    5447.3 5943.3
## - YOJ      19    5439.8 5955.8
## - HOMEKIDS  5    5419.9 5963.9
## - SEX       1    5415.3 5967.3
## - INCOME    1    5415.5 5967.5
## <none>      5414.1 5968.1
## - RED_CAR   1    5416.4 5968.4
## - PARENT1   1    5417.1 5969.1
## - HOME_VAL  1    5426.4 5978.4
## - OLDCLAIM  1    5426.9 5978.9
## - EDUCATION 3    5432.6 5980.6
## - AGE       57    5540.7 5980.7
## - BLUEBOOK  1    5432.6 5984.6
## - MVR_PTS   12    5454.8 5984.8
## - KIDSDRIV  4    5443.9 5989.9
## - MSTATUS   1    5440.4 5992.4
## - TIF       22    5488.2 5998.2
## - CLM_FREQ  5    5465.0 6009.0
## - JOB       8    5478.4 6016.4
## - CAR_USE   1    5478.9 6030.9
## - CAR_TYPE  5    5488.9 6032.9
## - REVOKED   1    5483.0 6035.0
## - URBANICITY 1    5856.9 6408.9
##
## Step:  AIC=5918.56
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + CAR_AGE + URBANICITY
##
##          Df Deviance    AIC
## - CAR_AGE  29    5587.4 5893.4
## - YOJ      19    5584.1 5910.1
## - HOMEKIDS  5    5560.6 5914.6
## - INCOME    1    5556.4 5918.4
## <none>      5554.6 5918.6
## - SEX       1    5556.9 5918.9
## - RED_CAR   1    5557.4 5919.4
## - PARENT1   1    5557.4 5919.4
## - EDUCATION 3    5570.3 5928.3
## - HOME_VAL  1    5567.2 5929.2
## - AGE       57    5680.7 5930.7
## - OLDCLAIM  1    5569.3 5931.3
## - BLUEBOOK  1    5569.8 5931.8
## - MVR_PTS   12    5596.1 5936.1
## - KIDSDRIV  4    5582.5 5938.5
## - MSTATUS   1    5580.6 5942.6
## - TIF       22    5629.9 5949.9
## - CLM_FREQ  5    5610.8 5964.8
## - JOB       8    5621.7 5969.7
## - CAR_USE   1    5617.3 5979.3
## - CAR_TYPE  5    5629.8 5983.8

```

```

## - REVOKED      1    5624.0 5986.0
## - URBANICITY   1    5974.8 6336.8
##
## Step:  AIC=5893.41
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - YOJ      19    5616.6 5884.6
## - HOMEKIDS   5    5594.1 5890.1
## - INCOME     1    5589.2 5893.2
## <none>      5587.4 5893.4
## - SEX        1    5589.5 5893.5
## - PARENT1    1    5589.7 5893.7
## - RED_CAR    1    5590.2 5894.2
## - HOME_VAL   1    5598.5 5902.5
## - AGE       57    5712.2 5904.2
## - OLDCLAIM   1    5602.3 5906.3
## - BLUEBOOK   1    5602.4 5906.4
## - MVR_PTS    12    5628.2 5910.2
## - EDUCATION   3    5612.4 5912.4
## - KIDSDRIV   4    5616.7 5914.7
## - MSTATUS     1    5616.3 5920.3
## - TIF        22    5663.0 5925.0
## - CLM_FREQ    5    5645.1 5941.1
## - JOB         8    5655.2 5945.2
## - CAR_USE     1    5652.8 5956.8
## - CAR_TYPE    5    5662.2 5958.2
## - REVOKED     1    5658.2 5962.2
## - URBANICITY  1    6007.4 6311.4
##
## Step:  AIC=5884.61
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - HOMEKIDS   5    5622.3 5880.3
## <none>      5616.6 5884.6
## - PARENT1    1    5618.6 5884.6
## - SEX        1    5618.7 5884.7
## - RED_CAR    1    5619.0 5885.0
## - INCOME     1    5619.9 5885.9
## - HOME_VAL   1    5627.4 5893.4
## - AGE       58    5744.5 5896.5
## - BLUEBOOK   1    5631.1 5897.1
## - OLDCLAIM   1    5631.2 5897.2
## - EDUCATION   3    5641.4 5903.4
## - MVR_PTS    12    5660.0 5904.0
## - KIDSDRIV   4    5645.5 5905.5
## - MSTATUS     1    5645.3 5911.3

```

```

## - TIF          22    5694.4 5918.4
## - JOB           8    5677.9 5929.9
## - CLM_FREQ      5    5673.1 5931.1
## - CAR_USE        1    5684.0 5950.0
## - CAR_TYPE       5    5692.2 5950.2
## - REVOKED        1    5689.0 5955.0
## - URBANICITY     1    6036.6 6302.6
##
## Step:   AIC=5880.32
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##             MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF +
##             CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS +
##             URBANICITY
##
##           Df Deviance    AIC
## - SEX          1    5624.2 5880.2
## <none>          5622.3 5880.3
## - RED_CAR       1    5624.7 5880.7
## - INCOME        1    5625.5 5881.5
## - PARENT1       1    5628.4 5884.4
## - HOME_VAL      1    5633.4 5889.4
## - OLDCLAIM      1    5636.1 5892.1
## - BLUEBOOK     1    5637.3 5893.3
## - AGE          58    5755.3 5897.3
## - EDUCATION     3    5646.7 5898.7
## - MVR_PTS      12    5666.5 5900.5
## - MSTATUS       1    5650.6 5906.6
## - KIDSDRIV      4    5660.8 5910.8
## - TIF          22    5699.7 5913.7
## - JOB           8    5683.8 5925.8
## - CLM_FREQ      5    5678.1 5926.1
## - CAR_TYPE       5    5697.4 5945.4
## - CAR_USE        1    5689.6 5945.6
## - REVOKED        1    5693.7 5949.7
## - URBANICITY     1    6042.2 6298.2
##
## Step:   AIC=5880.19
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##             MSTATUS + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
##             RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - RED_CAR       1    5625.3 5879.3
## <none>          5624.2 5880.2
## - INCOME        1    5627.4 5881.4
## - PARENT1       1    5630.2 5884.2
## - HOME_VAL      1    5635.1 5889.1
## - OLDCLAIM      1    5638.1 5892.1
## - AGE          58    5758.0 5898.0
## - EDUCATION     3    5648.6 5898.6
## - MVR_PTS      12    5668.1 5900.1
## - BLUEBOOK     1    5646.8 5900.8
## - MSTATUS       1    5652.8 5906.8
## - KIDSDRIV      4    5662.2 5910.2

```



```

## - TIF          22    5701.1 5913.1
## - JOB           8    5685.2 5925.2
## - CLM_FREQ      5    5680.2 5926.2
## - CAR_USE       1    5691.2 5945.2
## - CAR_TYPE      5    5700.0 5946.0
## - REVOKED       1    5695.7 5949.7
## - URBANICITY    1    6043.9 6297.9
##
## Step:  AIC=5879.31
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##      MSTATUS + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
##      OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY
##
##              Df Deviance    AIC
## <none>                5625.3 5879.3
## - INCOME             1    5628.6 5880.6
## - PARENT1            1    5631.5 5883.5
## - HOME_VAL           1    5636.0 5888.0
## - OLDCLAIM           1    5639.3 5891.3
## - AGE                58    5758.2 5896.2
## - EDUCATION          3    5649.9 5897.9
## - BLUEBOOK           1    5646.8 5898.8
## - MVR_PTS           12    5669.4 5899.4
## - MSTATUS            1    5653.8 5905.8
## - KIDSDRIV           4    5663.4 5909.4
## - TIF                22    5702.3 5912.3
## - JOB                 8    5686.8 5924.8
## - CLM_FREQ           5    5681.3 5925.3
## - CAR_USE            1    5692.4 5944.4
## - REVOKED            1    5696.7 5948.7
## - CAR_TYPE           5    5712.8 5956.8
## - URBANICITY         1    6044.9 6296.9

```

```
summary(model_2)
```

```

##
## Call:
## glm(formula = TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 +
##     HOME_VAL + MSTATUS + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##     TIF + CAR_TYPE + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS +
##     URBANICITY, family = "binomial", data = training_no_infl %>%
##     select(-INDEX, -TARGET_AMT))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5621  -0.6928  -0.3840   0.5741   3.0962
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.575e+00  1.310e+00  -1.202  0.229283
## KIDSDRIV1      5.495e-01  1.283e-01   4.284  1.84e-05 ***
## KIDSDRIV2      8.263e-01  1.771e-01   4.667  3.06e-06 ***
## KIDSDRIV3      8.989e-01  3.376e-01   2.662  0.007762 **
## KIDSDRIV4     -1.095e+01  5.218e+02  -0.021  0.983255
## AGE18         -1.326e+00  1.913e+00  -0.693  0.488268

```

## AGE19	-3.497e-02	1.760e+00	-0.020	0.984149
## AGE20	-1.922e+00	1.857e+00	-1.035	0.300674
## AGE21	-3.968e-01	1.492e+00	-0.266	0.790206
## AGE22	-4.770e-01	1.432e+00	-0.333	0.739075
## AGE23	-9.352e-02	1.505e+00	-0.062	0.950467
## AGE24	7.723e-01	1.457e+00	0.530	0.596145
## AGE25	5.064e-01	1.396e+00	0.363	0.716726
## AGE26	1.441e-01	1.353e+00	0.107	0.915167
## AGE27	-6.181e-01	1.329e+00	-0.465	0.641794
## AGE28	-7.625e-01	1.321e+00	-0.577	0.563919
## AGE29	-1.527e+00	1.320e+00	-1.157	0.247371
## AGE30	-9.278e-01	1.318e+00	-0.704	0.481391
## AGE31	-9.272e-01	1.305e+00	-0.710	0.477453
## AGE32	-1.009e+00	1.302e+00	-0.775	0.438217
## AGE33	-1.302e+00	1.298e+00	-1.003	0.315818
## AGE34	-1.424e+00	1.298e+00	-1.097	0.272726
## AGE35	-1.123e+00	1.293e+00	-0.868	0.385255
## AGE36	-1.041e+00	1.292e+00	-0.805	0.420750
## AGE37	-1.161e+00	1.292e+00	-0.899	0.368676
## AGE38	-9.992e-01	1.291e+00	-0.774	0.438954
## AGE39	-1.048e+00	1.290e+00	-0.812	0.416795
## AGE40	-1.526e+00	1.290e+00	-1.183	0.236611
## AGE41	-1.411e+00	1.290e+00	-1.094	0.274149
## AGE42	-1.201e+00	1.289e+00	-0.931	0.351611
## AGE43	-1.452e+00	1.290e+00	-1.125	0.260479
## AGE44	-1.732e+00	1.291e+00	-1.341	0.179824
## AGE45	-1.412e+00	1.289e+00	-1.096	0.273237
## AGE46	-1.481e+00	1.290e+00	-1.148	0.250949
## AGE47	-1.549e+00	1.291e+00	-1.200	0.230318
## AGE48	-1.655e+00	1.291e+00	-1.282	0.199819
## AGE49	-1.599e+00	1.292e+00	-1.238	0.215854
## AGE50	-1.387e+00	1.292e+00	-1.074	0.283013
## AGE51	-1.622e+00	1.295e+00	-1.252	0.210449
## AGE52	-1.295e+00	1.295e+00	-1.000	0.317166
## AGE53	-1.316e+00	1.294e+00	-1.016	0.309501
## AGE54	-1.198e+00	1.297e+00	-0.923	0.355873
## AGE55	-1.469e+00	1.300e+00	-1.130	0.258422
## AGE56	-1.212e+00	1.305e+00	-0.929	0.353020
## AGE57	-5.848e-01	1.299e+00	-0.450	0.652507
## AGE58	-5.293e-01	1.308e+00	-0.405	0.685708
## AGE59	-4.454e-01	1.322e+00	-0.337	0.736223
## AGE60	-4.377e-01	1.316e+00	-0.333	0.739441
## AGE61	-7.757e-01	1.325e+00	-0.585	0.558314
## AGE62	-9.034e-01	1.337e+00	-0.676	0.499242
## AGE63	-8.643e-01	1.347e+00	-0.642	0.521192
## AGE64	-2.867e-01	1.376e+00	-0.208	0.834932
## AGE65	-1.387e+00	1.466e+00	-0.946	0.344266
## AGE66	-1.290e+00	1.446e+00	-0.892	0.372443
## AGE67	-2.105e-01	1.529e+00	-0.138	0.890475
## AGE68	-4.814e-01	1.790e+00	-0.269	0.787987
## AGE69	-6.551e-01	1.759e+00	-0.373	0.709480
## AGE70	-1.400e+01	4.001e+02	-0.035	0.972092
## AGE72	3.295e-01	1.902e+00	0.173	0.862476
## AGE73	-8.550e-01	1.962e+00	-0.436	0.663046

## AGE76	1.745e+01	8.827e+02	0.020	0.984232	
## AGE80	-1.279e+01	8.827e+02	-0.014	0.988441	
## AGE81	-1.380e+01	8.827e+02	-0.016	0.987524	
## INCOME	-2.303e-06	1.284e-06	-1.793	0.072943	.
## PARENT1Yes	2.895e-01	1.167e-01	2.481	0.013103	*
## HOME_VAL	-1.294e-06	3.959e-07	-3.268	0.001083	**
## MSTATUSYes	-5.153e-01	9.588e-02	-5.374	7.70e-08	***
## EDUCATIONHigh School	4.523e-01	9.430e-02	4.797	1.61e-06	***
## EDUCATIONMasters	5.554e-02	1.576e-01	0.352	0.724531	
## EDUCATIONPhD	2.194e-01	2.041e-01	1.075	0.282382	
## JOBBBlue Collar	3.740e-01	2.159e-01	1.732	0.083261	.
## JOBClerical	5.228e-01	2.289e-01	2.284	0.022399	*
## JOBDoctor	-2.656e-01	2.928e-01	-0.907	0.364448	
## JOBHome Maker	3.352e-01	2.404e-01	1.395	0.163124	
## JOBLawyer	2.532e-01	1.945e-01	1.302	0.193004	
## JOBManager	-5.837e-01	2.006e-01	-2.909	0.003623	**
## JOBProfessional	2.210e-01	2.066e-01	1.070	0.284730	
## JOBStudent	2.214e-01	2.487e-01	0.890	0.373252	
## CAR_USEPrivate	-8.182e-01	1.009e-01	-8.105	5.26e-16	***
## BLUEBOOK	-2.507e-05	5.464e-06	-4.589	4.46e-06	***
## TIF10	-7.405e-01	1.307e-01	-5.664	1.48e-08	***
## TIF11	-4.095e-01	2.016e-01	-2.031	0.042218	*
## TIF12	-4.316e-01	4.578e-01	-0.943	0.345758	
## TIF13	-4.232e-01	1.944e-01	-2.176	0.029522	*
## TIF14	-1.066e+00	4.201e-01	-2.537	0.011195	*
## TIF15	1.286e-01	4.847e-01	0.265	0.790791	
## TIF16	-1.205e+00	5.378e-01	-2.241	0.025037	*
## TIF17	-5.741e-01	3.102e-01	-1.851	0.064163	.
## TIF18	-8.457e-01	6.588e-01	-1.284	0.199235	
## TIF19	-1.296e+01	3.611e+02	-0.036	0.971360	
## TIF2	2.763e-01	9.822e-01	0.281	0.778435	
## TIF20	-1.388e+01	3.172e+02	-0.044	0.965090	
## TIF21	2.132e-01	7.520e-01	0.284	0.776786	
## TIF22	-1.460e+01	4.790e+02	-0.030	0.975684	
## TIF25	-1.239e+01	6.029e+02	-0.021	0.983602	
## TIF3	-1.696e-01	1.500e-01	-1.131	0.258253	
## TIF4	-2.484e-01	1.031e-01	-2.410	0.015949	*
## TIF5	-6.463e-01	4.433e-01	-1.458	0.144845	
## TIF6	-4.073e-01	1.014e-01	-4.016	5.92e-05	***
## TIF7	-5.672e-01	1.388e-01	-4.088	4.36e-05	***
## TIF8	5.455e-01	3.939e-01	1.385	0.166070	
## TIF9	-6.856e-01	2.213e-01	-3.098	0.001948	**
## CAR_TYPEPanel Truck	6.873e-01	1.716e-01	4.005	6.20e-05	***
## CAR_TYPEPickup	5.284e-01	1.150e-01	4.593	4.36e-06	***
## CAR_TYPESports Car	1.005e+00	1.258e-01	7.988	1.37e-15	***
## CAR_TYPESUV	7.341e-01	9.938e-02	7.387	1.51e-13	***
## CAR_TYPEVan	6.706e-01	1.404e-01	4.776	1.79e-06	***
## OLDCLAIM	-1.771e-05	4.786e-06	-3.700	0.000215	***
## CLM_FREQ1	6.004e-01	1.161e-01	5.172	2.31e-07	***
## CLM_FREQ2	6.523e-01	1.084e-01	6.016	1.79e-09	***
## CLM_FREQ3	6.929e-01	1.216e-01	5.696	1.22e-08	***
## CLM_FREQ4	8.242e-01	1.985e-01	4.153	3.28e-05	***
## CLM_FREQ5	1.184e+00	6.484e-01	1.825	0.067953	.
## REVOKEDYes	9.096e-01	1.068e-01	8.518	< 2e-16	***

```
## MVR_PTS1          8.386e-02  1.050e-01  0.798 0.424651
## MVR_PTS10         1.045e+00  8.799e-01  1.188 0.234804
## MVR_PTS11         1.817e+00  1.132e+00  1.606 0.108356
## MVR_PTS13         1.552e+01  5.536e+02  0.028 0.977638
## MVR_PTS2          2.448e-01  1.100e-01  2.226 0.025982 *
## MVR_PTS3          3.048e-01  1.185e-01  2.573 0.010084 *
## MVR_PTS4          2.905e-01  1.284e-01  2.262 0.023688 *
## MVR_PTS5          1.689e-01  1.481e-01  1.140 0.254156
## MVR_PTS6          2.389e-01  1.775e-01  1.346 0.178366
## MVR_PTS7          6.357e-01  2.065e-01  3.078 0.002083 **
## MVR_PTS8          1.240e+00  3.194e-01  3.883 0.000103 ***
## MVR_PTS9          1.248e+00  3.999e-01  3.122 0.001799 **
## URBANICITYHighly Urban/ Urban 2.216e+00  1.268e-01  17.474 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 7445.1 on 6447 degrees of freedom
## Residual deviance: 5625.3 on 6321 degrees of freedom
## AIC: 5879.3
##
## Number of Fisher Scoring iterations: 13
vif(model_2)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## KIDSDRIV  1.374893  4      1.040599
## AGE       3.731650 58      1.011417
## INCOME    2.930509  1      1.711873
## PARENT1   1.705264  1      1.305858
## HOME_VAL  2.091848  1      1.446322
## MSTATUS   2.081513  1      1.442745
## EDUCATION 8.075751  3      1.416437
## JOB       23.379420  8      1.217733
## CAR_USE   2.283311  1      1.511063
## BLUEBOOK  1.834626  1      1.354484
## TIF       1.461604 22      1.008663
## CAR_TYPE  2.972147  5      1.115083
## OLDCLAIM  1.941762  1      1.393471
## CLM_FREQ  2.066599  5      1.075290
## REVOKED   1.392372  1      1.179988
## MVR_PTS   1.676936 12      1.021774
## URBANICITY 1.156896  1      1.075591
```

Model is not improved significantly after removing influential value as it usually the case in smaller data sets.

MODEL 3

3. Building a model with removed JOB variable to avoid multicollinearity.

```
model_3<-step(glm(TARGET_FLAG~., training %>% select(-INDEX, -TARGET_AMT, -JOB), family = "binomial"), c

## Start:  AIC=6016.43
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
```

```

##      HOME_VAL + MSTATUS + SEX + EDUCATION + TRAVTIME + CAR_USE +
##      BLUEBOOK + TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ +
##      REVOKED + MVR_PTS + CAR_AGE + URBANICITY

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

##           Df Deviance    AIC
## - TRAVTIME  95   5621.7 5969.7
## - CAR_AGE   29   5512.2 5992.2
## - YOJ       19   5497.2 5997.2
## - HOMEKIDS   5   5484.5 6012.5
## - SEX        1   5479.6 6015.6
## <none>           5478.4 6016.4
## - PARENT1    1   5481.0 6017.0
## - RED_CAR     1   5481.1 6017.1
## - INCOME      1   5482.1 6018.1
## - HOME_VAL    1   5488.2 6024.2
## - OLDCLAIM    1   5492.3 6028.3
## - AGE        57   5606.1 6030.1
## - BLUEBOOK    1   5497.1 6033.1
## - KIDSDRIV    4   5505.9 6035.9
## - MVR_PTS     12   5522.6 6036.6
## - MSTATUS      1   5506.9 6042.9
## - EDUCATION    3   5512.4 6044.4
## - TIF         22   5551.3 6045.3
## - CLM_FREQ     5   5531.2 6059.2
## - CAR_TYPE     5   5547.8 6075.8
## - REVOKED      1   5550.9 6086.9
## - CAR_USE      1   5589.9 6125.9
## - URBANICITY   1   5898.4 6434.4
##
## Step:  AIC=5969.73
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + CAR_AGE + URBANICITY
##
##           Df Deviance    AIC
## - CAR_AGE   29   5655.2 5945.2
## - YOJ       19   5644.2 5954.2
## - HOMEKIDS   5   5628.0 5966.0
## <none>           5621.7 5969.7
## - SEX        1   5623.8 5969.8
## - PARENT1    1   5624.0 5970.0
## - RED_CAR     1   5624.8 5970.8
## - INCOME      1   5626.2 5972.2
## - HOME_VAL    1   5631.5 5977.5
## - BLUEBOOK    1   5637.2 5983.2
## - OLDCLAIM    1   5637.6 5983.6
## - AGE        57   5750.1 5984.1
## - KIDSDRIV    4   5647.7 5987.7
## - MVR_PTS     12   5667.0 5991.0
## - EDUCATION    3   5653.3 5995.3
## - MSTATUS      1   5650.1 5996.1
## - TIF         22   5696.0 6000.0

```

```

## - CLM_FREQ      5    5680.1 6018.1
## - CAR_TYPE      5    5691.2 6029.2
## - REVOKED       1    5694.2 6040.2
## - CAR_USE       1    5734.4 6080.4
## - URBANICITY    1    6017.8 6363.8
##
## Step:  AIC=5945.23
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - YOJ      19    5677.9 5929.9
## - HOMEKIDS   5    5662.2 5942.2
## - SEX        1    5657.0 5945.0
## - PARENT1    1    5657.1 5945.1
## <none>      5655.2 5945.2
## - RED_CAR    1    5658.2 5946.2
## - INCOME     1    5659.7 5947.7
## - HOME_VAL   1    5663.7 5951.7
## - AGE       57    5782.3 5958.3
## - BLUEBOOK   1    5670.4 5958.4
## - OLDCLAIM   1    5671.2 5959.2
## - KIDSDRIV   4    5682.2 5964.2
## - MVR_PTS   12    5699.7 5965.7
## - MSTATUS    1    5686.4 5974.4
## - TIF       22    5730.1 5976.1
## - CLM_FREQ   5    5714.9 5994.9
## - EDUCATION  3    5713.9 5997.9
## - CAR_TYPE   5    5724.3 6004.3
## - REVOKED    1    5728.8 6016.8
## - CAR_USE    1    5768.6 6056.6
## - URBANICITY 1    6050.8 6338.8
##
## Step:  AIC=5929.87
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - HOMEKIDS   5    5683.8 5925.8
## - SEX        1    5679.5 5929.5
## - PARENT1    1    5679.5 5929.5
## <none>      5677.9 5929.9
## - RED_CAR    1    5680.6 5930.6
## - HOME_VAL   1    5687.2 5937.2
## - INCOME     1    5687.5 5937.5
## - BLUEBOOK   1    5693.2 5943.2
## - OLDCLAIM   1    5693.6 5943.6
## - AGE       58    5807.6 5943.6
## - KIDSDRIV   4    5704.7 5948.7
## - MVR_PTS   12    5723.5 5951.5

```

```

## - MSTATUS      1    5708.1 5958.1
## - TIF          22    5754.5 5962.5
## - CLM_FREQ     5     5736.3 5978.3
## - EDUCATION    3     5733.9 5979.9
## - CAR_TYPE     5     5749.1 5991.1
## - REVOKED      1     5752.7 6002.7
## - CAR_USE      1     5792.7 6042.7
## - URBANICITY   1     6071.3 6321.3
##
## Step:  AIC=5925.78
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##      MSTATUS + SEX + EDUCATION + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
##      RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - SEX      1    5685.2 5925.2
## <none>      5683.8 5925.8
## - RED_CAR   1    5686.5 5926.5
## - PARENT1   1    5689.5 5929.5
## - INCOME    1    5693.0 5933.0
## - HOME_VAL  1    5693.3 5933.3
## - OLDCLAIM  1    5698.6 5938.6
## - BLUEBOOK  1    5699.5 5939.5
## - AGE      58    5820.1 5946.1
## - MVR_PTS  12    5730.2 5948.2
## - MSTATUS   1    5713.4 5953.4
## - KIDSDRIV  4    5720.6 5954.6
## - TIF      22    5759.9 5957.9
## - CLM_FREQ  5    5741.5 5973.5
## - EDUCATION 3    5739.2 5975.2
## - CAR_TYPE  5    5754.5 5986.5
## - REVOKED   1    5757.7 5997.7
## - CAR_USE   1    5798.0 6038.0
## - URBANICITY 1    6077.2 6317.2
##
## Step:  AIC=5925.18
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##      MSTATUS + EDUCATION + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
##      RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - RED_CAR   1    5686.8 5924.8
## <none>      5685.2 5925.2
## - PARENT1   1    5690.8 5928.8
## - INCOME    1    5694.2 5932.2
## - HOME_VAL  1    5694.5 5932.5
## - OLDCLAIM  1    5700.1 5938.1
## - BLUEBOOK  1    5708.1 5946.1
## - AGE      58    5822.2 5946.2
## - MVR_PTS  12    5731.3 5947.3
## - MSTATUS   1    5715.1 5953.1
## - KIDSDRIV  4    5721.6 5953.6
## - TIF      22    5760.9 5956.9
## - CLM_FREQ  5    5743.1 5973.1

```

```

## - EDUCATION      3    5741.1 5975.1
## - CAR_TYPE       5    5758.2 5988.2
## - REVOKED        1    5759.1 5997.1
## - CAR_USE        1    5799.5 6037.5
## - URBANICITY     1    6079.0 6317.0
##
## Step:  AIC=5924.76
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##      MSTATUS + EDUCATION + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
##      OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY
##
##              Df Deviance    AIC
## <none>              5686.8 5924.8
## - PARENT1          1    5692.6 5928.6
## - HOME_VAL         1    5695.9 5931.9
## - INCOME           1    5696.0 5932.0
## - OLDCLAIM         1    5701.8 5937.8
## - BLUEBOOK         1    5708.1 5944.1
## - AGE              58    5822.5 5944.5
## - MVR_PTS          12    5733.2 5947.2
## - MSTATUS          1    5716.5 5952.5
## - KIDSDRIV         4    5723.3 5953.3
## - TIF              22    5762.7 5956.7
## - CLM_FREQ         5    5744.6 5972.6
## - EDUCATION        3    5742.7 5974.7
## - REVOKED          1    5760.6 5996.6
## - CAR_TYPE         5    5774.6 6002.6
## - CAR_USE          1    5801.2 6037.2
## - URBANICITY       1    6080.2 6316.2
summary(model_3)

##
## Call:
## glm(formula = TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + EDUCATION + CAR_USE + BLUEBOOK + TIF +
##      CAR_TYPE + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY,
##      family = "binomial", data = training %>% select(-INDEX, -TARGET_AMT,
##      -JOB))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4796  -0.7011  -0.3980   0.5782   3.0296
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.173e+00  1.302e+00  -0.901  0.367673
## KIDSDRIV1      5.518e-01  1.278e-01   4.319 1.57e-05 ***
## KIDSDRIV2      7.855e-01  1.756e-01   4.472 7.74e-06 ***
## KIDSDRIV3      8.567e-01  3.333e-01   2.571 0.010149 *
## KIDSDRIV4     -1.098e+01  5.402e+02  -0.020 0.983784
## AGE18         -1.381e+00  1.920e+00  -0.719 0.471849
## AGE19         -4.135e-01  1.731e+00  -0.239 0.811176
## AGE20         -1.995e+00  1.865e+00  -1.070 0.284732
## AGE21         -7.374e-01  1.494e+00  -0.494 0.621488

```


## AGE22	-5.465e-01	1.439e+00	-0.380	0.704119
## AGE23	-1.805e-01	1.510e+00	-0.120	0.904863
## AGE24	7.748e-01	1.466e+00	0.529	0.597092
## AGE25	4.397e-01	1.402e+00	0.314	0.753750
## AGE26	1.939e-02	1.363e+00	0.014	0.988653
## AGE27	-6.171e-01	1.338e+00	-0.461	0.644598
## AGE28	-8.490e-01	1.330e+00	-0.638	0.523321
## AGE29	-1.613e+00	1.329e+00	-1.214	0.224819
## AGE30	-9.934e-01	1.326e+00	-0.749	0.453643
## AGE31	-9.831e-01	1.314e+00	-0.748	0.454412
## AGE32	-1.069e+00	1.310e+00	-0.816	0.414449
## AGE33	-1.381e+00	1.307e+00	-1.056	0.290755
## AGE34	-1.463e+00	1.307e+00	-1.119	0.263106
## AGE35	-1.168e+00	1.302e+00	-0.897	0.369933
## AGE36	-1.082e+00	1.302e+00	-0.831	0.405895
## AGE37	-1.229e+00	1.301e+00	-0.945	0.344852
## AGE38	-1.029e+00	1.300e+00	-0.791	0.428854
## AGE39	-1.121e+00	1.299e+00	-0.862	0.388430
## AGE40	-1.575e+00	1.299e+00	-1.213	0.225292
## AGE41	-1.452e+00	1.300e+00	-1.117	0.263795
## AGE42	-1.250e+00	1.299e+00	-0.962	0.335895
## AGE43	-1.525e+00	1.299e+00	-1.174	0.240560
## AGE44	-1.818e+00	1.301e+00	-1.398	0.162137
## AGE45	-1.457e+00	1.298e+00	-1.122	0.261867
## AGE46	-1.549e+00	1.299e+00	-1.193	0.233031
## AGE47	-1.631e+00	1.300e+00	-1.254	0.209822
## AGE48	-1.753e+00	1.300e+00	-1.348	0.177645
## AGE49	-1.680e+00	1.302e+00	-1.291	0.196851
## AGE50	-1.464e+00	1.301e+00	-1.125	0.260396
## AGE51	-1.741e+00	1.304e+00	-1.335	0.181803
## AGE52	-1.398e+00	1.304e+00	-1.072	0.283528
## AGE53	-1.441e+00	1.304e+00	-1.106	0.268873
## AGE54	-1.272e+00	1.306e+00	-0.974	0.330044
## AGE55	-1.590e+00	1.309e+00	-1.215	0.224254
## AGE56	-1.312e+00	1.314e+00	-0.999	0.317856
## AGE57	-6.650e-01	1.308e+00	-0.508	0.611106
## AGE58	-6.342e-01	1.317e+00	-0.481	0.630178
## AGE59	-5.458e-01	1.331e+00	-0.410	0.681769
## AGE60	-5.798e-01	1.325e+00	-0.438	0.661646
## AGE61	-8.358e-01	1.332e+00	-0.628	0.530263
## AGE62	-1.046e+00	1.344e+00	-0.778	0.436430
## AGE63	-1.023e+00	1.356e+00	-0.755	0.450492
## AGE64	-3.411e-01	1.380e+00	-0.247	0.804829
## AGE65	-1.485e+00	1.472e+00	-1.009	0.312817
## AGE66	-1.432e+00	1.448e+00	-0.989	0.322702
## AGE67	-4.860e-01	1.554e+00	-0.313	0.754436
## AGE68	-6.005e-01	1.785e+00	-0.337	0.736490
## AGE69	-9.884e-01	1.751e+00	-0.564	0.572429
## AGE70	-1.460e+01	4.058e+02	-0.036	0.971292
## AGE72	4.786e-01	1.923e+00	0.249	0.803404
## AGE73	-9.673e-01	1.971e+00	-0.491	0.623496
## AGE76	1.729e+01	8.827e+02	0.020	0.984370
## AGE80	-1.282e+01	8.827e+02	-0.015	0.988415
## AGE81	-1.379e+01	8.827e+02	-0.016	0.987535

## INCOME	-3.493e-06	1.155e-06	-3.025	0.002489	**
## PARENT1Yes	2.796e-01	1.157e-01	2.417	0.015667	*
## HOME_VAL	-1.166e-06	3.854e-07	-3.024	0.002493	**
## MSTATUSYes	-5.190e-01	9.451e-02	-5.492	3.97e-08	***
## EDUCATIONHigh School	5.868e-01	8.709e-02	6.738	1.61e-11	***
## EDUCATIONMasters	-1.118e-02	1.060e-01	-0.105	0.916021	
## EDUCATIONPhD	2.318e-02	1.491e-01	0.155	0.876471	
## CAR_USEPrivate	-8.690e-01	8.181e-02	-10.621	< 2e-16	***
## BLUEBOOK	-2.479e-05	5.429e-06	-4.567	4.96e-06	***
## TIF10	-7.309e-01	1.297e-01	-5.636	1.74e-08	***
## TIF11	-3.995e-01	2.004e-01	-1.993	0.046230	*
## TIF12	-5.099e-01	4.523e-01	-1.127	0.259551	
## TIF13	-4.031e-01	1.929e-01	-2.090	0.036633	*
## TIF14	-1.071e+00	4.153e-01	-2.580	0.009884	**
## TIF15	1.702e-01	4.784e-01	0.356	0.721951	
## TIF16	-1.131e+00	5.376e-01	-2.104	0.035354	*
## TIF17	-5.244e-01	3.089e-01	-1.698	0.089540	.
## TIF18	-9.397e-01	6.514e-01	-1.443	0.149097	
## TIF19	-1.309e+01	3.656e+02	-0.036	0.971439	
## TIF2	2.843e-01	9.823e-01	0.289	0.772237	
## TIF20	-1.372e+01	3.206e+02	-0.043	0.965867	
## TIF21	3.935e-01	7.528e-01	0.523	0.601148	
## TIF22	-1.443e+01	4.807e+02	-0.030	0.976043	
## TIF25	-1.234e+01	5.933e+02	-0.021	0.983407	
## TIF3	-1.374e-01	1.484e-01	-0.926	0.354450	
## TIF4	-2.458e-01	1.026e-01	-2.396	0.016559	*
## TIF5	-6.470e-01	4.452e-01	-1.453	0.146152	
## TIF6	-3.982e-01	1.007e-01	-3.954	7.69e-05	***
## TIF7	-5.494e-01	1.379e-01	-3.983	6.79e-05	***
## TIF8	5.174e-01	3.851e-01	1.344	0.179061	
## TIF9	-6.776e-01	2.202e-01	-3.077	0.002094	**
## CAR_TYPEPanel Truck	5.887e-01	1.626e-01	3.622	0.000293	***
## CAR_TYPEPickup	4.725e-01	1.123e-01	4.208	2.57e-05	***
## CAR_TYPESports Car	9.844e-01	1.236e-01	7.966	1.64e-15	***
## CAR_TYPESUV	7.360e-01	9.807e-02	7.504	6.18e-14	***
## CAR_TYPEVan	6.336e-01	1.377e-01	4.601	4.20e-06	***
## OLDCLAIM	-1.815e-05	4.738e-06	-3.830	0.000128	***
## CLM_FREQ1	6.127e-01	1.153e-01	5.315	1.07e-07	***
## CLM_FREQ2	6.682e-01	1.078e-01	6.198	5.73e-10	***
## CLM_FREQ3	7.011e-01	1.208e-01	5.803	6.52e-09	***
## CLM_FREQ4	8.281e-01	1.960e-01	4.226	2.38e-05	***
## CLM_FREQ5	9.451e-01	6.459e-01	1.463	0.143383	
## REVOKEDYes	9.164e-01	1.057e-01	8.667	< 2e-16	***
## MVR_PTS1	1.018e-01	1.045e-01	0.975	0.329607	
## MVR_PTS10	1.179e+00	8.835e-01	1.334	0.182075	
## MVR_PTS11	1.925e+00	1.137e+00	1.693	0.090464	.
## MVR_PTS13	1.550e+01	5.567e+02	0.028	0.977784	
## MVR_PTS2	2.642e-01	1.092e-01	2.419	0.015580	*
## MVR_PTS3	3.154e-01	1.176e-01	2.683	0.007293	**
## MVR_PTS4	3.168e-01	1.274e-01	2.487	0.012874	*
## MVR_PTS5	2.383e-01	1.471e-01	1.620	0.105318	
## MVR_PTS6	2.152e-01	1.761e-01	1.222	0.221747	
## MVR_PTS7	6.504e-01	2.048e-01	3.175	0.001499	**
## MVR_PTS8	1.248e+00	3.163e-01	3.946	7.94e-05	***

```
## MVR_PTS9          1.255e+00  3.985e-01  3.150 0.001635 **
## URBANICITYHighly Urban/ Urban 2.141e+00  1.264e-01  16.943 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 7445.1 on 6447 degrees of freedom
## Residual deviance: 5686.8 on 6329 degrees of freedom
## AIC: 5924.8
##
## Number of Fisher Scoring iterations: 13
```

```
vif(model_3)
```

```
##          GVIF Df GVIF^(1/(2*Df))
## KIDSDRIV  1.360126  4      1.039196
## AGE       3.228207 58      1.010154
## INCOME    2.362897  1      1.537172
## PARENT1   1.700422  1      1.304002
## HOME_VAL  2.010175  1      1.417807
## MSTATUS   2.048182  1      1.431147
## EDUCATION 1.846946  3      1.107666
## CAR_USE    1.516373  1      1.231411
## BLUEBOOK  1.818495  1      1.348516
## TIF        1.416742 22      1.007949
## CAR_TYPE   2.535545  5      1.097507
## OLDCLAIM   1.938003  1      1.392122
## CLM_FREQ   2.045672  5      1.074196
## REVOKED    1.386616  1      1.177547
## MVR_PTS    1.658826 12      1.021312
## URBANICITY 1.141031  1      1.068191
```

MODEL 4

4. Taking log 10 of INCOME (1 unit increase in income has a consistent increase in the income variable).

```
training$INCOME<- log10(training$INCOME+1)
```

```
model_4<-step(glm(TARGET_FLAG~., training %>% select(-INDEX, -TARGET_AMT), family = "binomial"), direct.
```

```
## Start: AIC=5966.49
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
## HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + TRAVTIME + CAR_USE +
## BLUEBOOK + TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ +
## REVOKED + MVR_PTS + CAR_AGE + URBANICITY
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
##          Df Deviance    AIC
## - TRAVTIME  95  5552.7 5916.7
## - CAR_AGE   29  5446.1 5942.1
## - YOJ       19  5427.0 5943.0
## - HOMEKIDS   5  5418.3 5962.3
## - SEX        1  5413.7 5965.7
## <none>       5412.5 5966.5
```

```

## - RED_CAR      1    5414.8 5966.8
## - PARENT1      1    5415.5 5967.5
## - INCOME       1    5415.5 5967.5
## - OLDCLAIM     1    5425.7 5977.7
## - EDUCATION    3    5429.8 5977.8
## - AGE          57    5538.2 5978.2
## - HOME_VAL     1    5426.5 5978.5
## - BLUEBOOK     1    5431.2 5983.2
## - MVR_PTS     12    5453.6 5983.6
## - KIDSDRIV     4    5442.2 5988.2
## - MSTATUS      1    5439.6 5991.6
## - TIF          22    5486.3 5996.3
## - CLM_FREQ     5    5463.9 6007.9
## - JOB          8    5480.9 6018.9
## - CAR_USE      1    5478.0 6030.0
## - CAR_TYPE     5    5486.7 6030.7
## - REVOKED      1    5481.7 6033.7
## - URBANICITY   1    5854.8 6406.8
##
## Step:  AIC=5916.75
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + CAR_AGE + URBANICITY
##
##           Df Deviance    AIC
## - CAR_AGE   29   5585.9 5891.9
## - YOJ       19   5570.4 5896.4
## - HOMEKIDS   5   5558.7 5912.7
## <none>       5552.7 5916.7
## - SEX       1   5555.0 5917.0
## - PARENT1   1   5555.5 5917.5
## - RED_CAR   1   5555.5 5917.5
## - INCOME    1   5556.4 5918.4
## - EDUCATION  3   5567.2 5925.2
## - AGE       57   5678.0 5928.0
## - HOME_VAL   1   5567.1 5929.1
## - OLDCLAIM   1   5567.8 5929.8
## - BLUEBOOK   1   5568.2 5930.2
## - MVR_PTS   12   5594.7 5934.7
## - KIDSDRIV   4   5580.5 5936.5
## - MSTATUS    1   5579.4 5941.4
## - TIF        22   5627.8 5947.8
## - CLM_FREQ   5   5609.6 5963.6
## - JOB        8   5624.4 5972.4
## - CAR_USE    1   5616.1 5978.1
## - CAR_TYPE   5   5627.5 5981.5
## - REVOKED    1   5622.5 5984.5
## - URBANICITY 1   5972.4 6334.4
##
## Step:  AIC=5891.94
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +

```

```

##      MVR_PTS + URBANICITY
##
##      Df Deviance    AIC
## - YOJ      19  5603.4 5871.4
## - HOMEKIDS   5  5592.5 5888.5
## <none>      5585.9 5891.9
## - SEX        1  5587.9 5891.9
## - PARENT1    1  5588.2 5892.2
## - RED_CAR    1  5588.7 5892.7
## - INCOME     1  5589.2 5893.2
## - AGE       57  5710.0 5902.0
## - HOME_VAL   1  5598.6 5902.6
## - BLUEBOOK   1  5601.1 5905.1
## - OLDCLAIM   1  5601.1 5905.1
## - MVR_PTS   12  5627.1 5909.1
## - EDUCATION  3  5609.5 5909.5
## - KIDSDRIV   4  5615.1 5913.1
## - MSTATUS    1  5615.4 5919.4
## - TIF       22  5661.3 5923.3
## - CLM_FREQ   5  5644.2 5940.2
## - JOB        8  5657.9 5947.9
## - CAR_USE    1  5651.9 5955.9
## - CAR_TYPE   5  5660.2 5956.2
## - REVOKED    1  5657.0 5961.0
## - URBANICITY 1  6005.4 6309.4
##
## Step:  AIC=5871.39
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + INCOME + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + URBANICITY
##
##      Df Deviance    AIC
## - HOMEKIDS   5  5609.3 5867.3
## - PARENT1    1  5605.3 5871.3
## <none>      5603.4 5871.4
## - SEX        1  5605.5 5871.5
## - RED_CAR    1  5605.9 5871.9
## - HOME_VAL   1  5618.0 5884.0
## - OLDCLAIM   1  5618.3 5884.3
## - BLUEBOOK   1  5618.3 5884.3
## - INCOME     1  5619.9 5885.9
## - AGE       58  5734.2 5886.2
## - MVR_PTS   12  5644.8 5888.8
## - KIDSDRIV   4  5631.6 5891.6
## - EDUCATION  3  5629.7 5891.7
## - MSTATUS    1  5632.5 5898.5
## - TIF       22  5678.8 5902.8
## - CLM_FREQ   5  5660.3 5918.3
## - JOB        8  5674.4 5926.4
## - CAR_USE    1  5669.5 5935.5
## - CAR_TYPE   5  5678.3 5936.3
## - REVOKED    1  5675.6 5941.6
## - URBANICITY 1  6024.7 6290.7

```

```

##
## Step: AIC=5867.3
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##      MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF +
##      CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS +
##      URBANICITY
##
##           Df Deviance    AIC
## - SEX      1  5611.2 5867.2
## <none>      5609.3 5867.3
## - RED_CAR   1  5611.8 5867.8
## - PARENT1   1  5615.4 5871.4
## - OLDCLAIM  1  5623.3 5879.3
## - HOME_VAL  1  5624.2 5880.2
## - BLUEBOOK  1  5624.7 5880.7
## - INCOME    1  5625.5 5881.5
## - MVR_PTS   12  5651.4 5885.4
## - AGE       58  5745.1 5887.1
## - EDUCATION  3  5635.3 5887.3
## - MSTATUS   1  5637.7 5893.7
## - KIDSDRIV  4  5647.4 5897.4
## - TIF       22  5684.2 5898.2
## - CLM_FREQ  5  5665.6 5913.6
## - JOB       8  5680.6 5922.6
## - CAR_USE   1  5675.4 5931.4
## - CAR_TYPE  5  5683.7 5931.7
## - REVOKED   1  5680.5 5936.5
## - URBANICITY 1  6030.4 6286.4
##
## Step: AIC=5867.17
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
##      MSTATUS + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
##      RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY
##
##           Df Deviance    AIC
## - RED_CAR   1  5612.4 5866.4
## <none>      5611.2 5867.2
## - PARENT1   1  5617.1 5871.1
## - OLDCLAIM  1  5625.3 5879.3
## - HOME_VAL  1  5625.9 5879.9
## - INCOME    1  5627.4 5881.4
## - MVR_PTS   12  5653.0 5885.0
## - EDUCATION  3  5637.3 5887.3
## - AGE       58  5748.0 5888.0
## - BLUEBOOK  1  5634.6 5888.6
## - MSTATUS   1  5639.9 5893.9
## - KIDSDRIV  4  5648.8 5896.8
## - TIF       22  5685.6 5897.6
## - CLM_FREQ  5  5667.7 5913.7
## - JOB       8  5682.3 5922.3
## - CAR_USE   1  5677.0 5931.0
## - CAR_TYPE  5  5686.3 5932.3
## - REVOKED   1  5682.5 5936.5
## - URBANICITY 1  6032.1 6286.1

```

```
##
## Step: AIC=5866.37
## TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 + HOME_VAL +
## MSTATUS + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
## OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY
##
##          Df Deviance    AIC
## <none>          5612.4 5866.4
## - PARENT1      1   5618.5 5870.5
## - OLDCLAIM      1   5626.6 5878.6
## - HOME_VAL      1   5626.9 5878.9
## - INCOME        1   5628.6 5880.6
## - MVR_PTS      12   5654.5 5884.5
## - AGE          58   5748.2 5886.2
## - BLUEBOOK      1   5634.6 5886.6
## - EDUCATION      3   5638.7 5886.7
## - MSTATUS        1   5641.0 5893.0
## - KIDSDRIV       4   5650.1 5896.1
## - TIF           22   5687.0 5897.0
## - CLM_FREQ       5   5668.9 5912.9
## - JOB            8   5683.8 5921.8
## - CAR_USE        1   5678.3 5930.3
## - REVOKED        1   5683.6 5935.6
## - CAR_TYPE       5   5699.5 5943.5
## - URBANICITY     1   6033.2 6285.2

summary(model_4)

##
## Call:
## glm(formula = TARGET_FLAG ~ KIDSDRIV + AGE + INCOME + PARENT1 +
## HOME_VAL + MSTATUS + EDUCATION + JOB + CAR_USE + BLUEBOOK +
## TIF + CAR_TYPE + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS +
## URBANICITY, family = "binomial", data = training %>% select(-INDEX,
## -TARGET_AMT))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5947  -0.6899  -0.3827   0.5591   3.0126
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.113e+00  1.317e+00  -0.845  0.397844
## KIDSDRIV1      5.473e-01  1.285e-01   4.260  2.05e-05 ***
## KIDSDRIV2      8.176e-01  1.773e-01   4.612  3.99e-06 ***
## KIDSDRIV3      9.142e-01  3.388e-01   2.699  0.006965 **
## KIDSDRIV4     -1.099e+01  5.259e+02  -0.021  0.983324
## AGE18         -1.194e+00  1.916e+00  -0.623  0.533091
## AGE19         -3.177e-02  1.747e+00  -0.018  0.985489
## AGE20         -1.891e+00  1.846e+00  -1.024  0.305645
## AGE21         -2.492e-01  1.496e+00  -0.167  0.867762
## AGE22         -3.800e-01  1.432e+00  -0.265  0.790689
## AGE23         -9.819e-03  1.504e+00  -0.007  0.994790
## AGE24          8.938e-01  1.457e+00   0.613  0.539675
## AGE25          5.915e-01  1.398e+00   0.423  0.672330
```

## AGE26	2.345e-01	1.354e+00	0.173	0.862511	
## AGE27	-4.888e-01	1.330e+00	-0.368	0.713246	
## AGE28	-6.565e-01	1.323e+00	-0.496	0.619752	
## AGE29	-1.452e+00	1.321e+00	-1.099	0.271698	
## AGE30	-8.230e-01	1.319e+00	-0.624	0.532541	
## AGE31	-7.986e-01	1.306e+00	-0.611	0.540961	
## AGE32	-9.193e-01	1.303e+00	-0.706	0.480318	
## AGE33	-1.224e+00	1.299e+00	-0.942	0.345975	
## AGE34	-1.319e+00	1.299e+00	-1.015	0.310127	
## AGE35	-1.017e+00	1.294e+00	-0.786	0.432042	
## AGE36	-9.516e-01	1.293e+00	-0.736	0.461876	
## AGE37	-1.055e+00	1.293e+00	-0.816	0.414751	
## AGE38	-9.038e-01	1.292e+00	-0.700	0.484236	
## AGE39	-9.405e-01	1.291e+00	-0.728	0.466408	
## AGE40	-1.415e+00	1.291e+00	-1.096	0.272909	
## AGE41	-1.305e+00	1.291e+00	-1.011	0.312207	
## AGE42	-1.086e+00	1.290e+00	-0.842	0.400023	
## AGE43	-1.372e+00	1.291e+00	-1.062	0.288034	
## AGE44	-1.644e+00	1.292e+00	-1.272	0.203296	
## AGE45	-1.317e+00	1.290e+00	-1.021	0.307221	
## AGE46	-1.381e+00	1.291e+00	-1.070	0.284693	
## AGE47	-1.449e+00	1.292e+00	-1.122	0.261940	
## AGE48	-1.548e+00	1.292e+00	-1.198	0.230811	
## AGE49	-1.504e+00	1.293e+00	-1.163	0.244982	
## AGE50	-1.292e+00	1.293e+00	-0.999	0.317705	
## AGE51	-1.539e+00	1.296e+00	-1.187	0.235113	
## AGE52	-1.188e+00	1.296e+00	-0.917	0.359346	
## AGE53	-1.217e+00	1.295e+00	-0.940	0.347401	
## AGE54	-1.113e+00	1.298e+00	-0.858	0.391085	
## AGE55	-1.365e+00	1.301e+00	-1.050	0.293924	
## AGE56	-1.121e+00	1.306e+00	-0.858	0.390816	
## AGE57	-4.778e-01	1.300e+00	-0.368	0.713153	
## AGE58	-4.273e-01	1.309e+00	-0.326	0.744059	
## AGE59	-3.399e-01	1.323e+00	-0.257	0.797262	
## AGE60	-3.026e-01	1.317e+00	-0.230	0.818303	
## AGE61	-6.454e-01	1.326e+00	-0.487	0.626443	
## AGE62	-7.947e-01	1.338e+00	-0.594	0.552509	
## AGE63	-7.529e-01	1.348e+00	-0.558	0.576548	
## AGE64	-1.654e-01	1.377e+00	-0.120	0.904404	
## AGE65	-1.256e+00	1.472e+00	-0.853	0.393388	
## AGE66	-1.192e+00	1.447e+00	-0.824	0.409841	
## AGE67	-1.133e-01	1.531e+00	-0.074	0.940980	
## AGE68	-4.026e-01	1.785e+00	-0.226	0.821533	
## AGE69	-5.196e-01	1.758e+00	-0.296	0.767599	
## AGE70	-1.390e+01	4.002e+02	-0.035	0.972284	
## AGE72	4.773e-01	1.904e+00	0.251	0.802011	
## AGE73	-6.282e-01	1.981e+00	-0.317	0.751186	
## AGE76	1.757e+01	8.827e+02	0.020	0.984116	
## AGE80	-1.271e+01	8.827e+02	-0.014	0.988514	
## AGE81	-1.374e+01	8.827e+02	-0.016	0.987579	
## INCOME	-1.570e-01	3.916e-02	-4.009	6.11e-05	***
## PARENT1Yes	2.899e-01	1.168e-01	2.482	0.013072	*
## HOME_VAL	-1.379e-06	3.655e-07	-3.772	0.000162	***
## MSTATUSYes	-5.021e-01	9.335e-02	-5.379	7.49e-08	***

## EDUCATIONHigh School	4.669e-01	9.303e-02	5.018	5.21e-07	***
## EDUCATIONMasters	3.463e-02	1.579e-01	0.219	0.826436	
## EDUCATIONPhD	1.341e-01	1.999e-01	0.671	0.502464	
## JOBBBlue Collar	3.952e-01	2.150e-01	1.838	0.066026	.
## JOBClerical	5.533e-01	2.255e-01	2.453	0.014156	*
## JOBDoctor	-2.720e-01	2.925e-01	-0.930	0.352383	
## JOBHome Maker	2.722e-02	2.563e-01	0.106	0.915421	
## JOBLawyer	2.657e-01	1.939e-01	1.370	0.170616	
## JOBManager	-5.714e-01	2.001e-01	-2.855	0.004298	**
## JOBProfessional	2.325e-01	2.061e-01	1.128	0.259343	
## JOBStudent	-8.699e-02	2.641e-01	-0.329	0.741848	
## CAR_USEPrivate	-8.120e-01	1.011e-01	-8.035	9.38e-16	***
## BLUEBOOK	-2.509e-05	5.387e-06	-4.659	3.18e-06	***
## TIF10	-7.222e-01	1.308e-01	-5.519	3.40e-08	***
## TIF11	-3.977e-01	2.021e-01	-1.968	0.049037	*
## TIF12	-4.311e-01	4.572e-01	-0.943	0.345735	
## TIF13	-4.114e-01	1.948e-01	-2.112	0.034729	*
## TIF14	-1.046e+00	4.202e-01	-2.490	0.012761	*
## TIF15	1.253e-01	4.847e-01	0.259	0.795946	
## TIF16	-1.191e+00	5.362e-01	-2.221	0.026351	*
## TIF17	-5.705e-01	3.110e-01	-1.834	0.066590	.
## TIF18	-8.591e-01	6.579e-01	-1.306	0.191642	
## TIF19	-1.296e+01	3.631e+02	-0.036	0.971526	
## TIF2	2.396e-01	9.849e-01	0.243	0.807796	
## TIF20	-1.384e+01	3.169e+02	-0.044	0.965166	
## TIF21	2.270e-01	7.533e-01	0.301	0.763127	
## TIF22	-1.461e+01	4.782e+02	-0.031	0.975625	
## TIF25	-1.236e+01	6.024e+02	-0.021	0.983631	
## TIF3	-1.651e-01	1.503e-01	-1.099	0.271826	
## TIF4	-2.361e-01	1.032e-01	-2.288	0.022154	*
## TIF5	-6.153e-01	4.468e-01	-1.377	0.168513	
## TIF6	-3.967e-01	1.015e-01	-3.907	9.34e-05	***
## TIF7	-5.553e-01	1.389e-01	-3.997	6.41e-05	***
## TIF8	5.517e-01	3.946e-01	1.398	0.162023	
## TIF9	-6.994e-01	2.219e-01	-3.152	0.001623	**
## CAR_TYPEPanel Truck	6.802e-01	1.716e-01	3.963	7.39e-05	***
## CAR_TYPEPickup	5.391e-01	1.151e-01	4.684	2.81e-06	***
## CAR_TYPESports Car	9.990e-01	1.262e-01	7.916	2.45e-15	***
## CAR_TYPESUV	7.433e-01	9.958e-02	7.465	8.34e-14	***
## CAR_TYPEVan	6.678e-01	1.404e-01	4.756	1.98e-06	***
## OLDCLAIM	-1.789e-05	4.798e-06	-3.728	0.000193	***
## CLM_FREQ1	6.099e-01	1.163e-01	5.246	1.55e-07	***
## CLM_FREQ2	6.576e-01	1.086e-01	6.056	1.40e-09	***
## CLM_FREQ3	6.892e-01	1.217e-01	5.661	1.51e-08	***
## CLM_FREQ4	8.389e-01	1.987e-01	4.221	2.43e-05	***
## CLM_FREQ5	1.149e+00	6.531e-01	1.760	0.078480	.
## REVOKEDYes	9.098e-01	1.070e-01	8.506	< 2e-16	***
## MVR_PTS1	7.678e-02	1.052e-01	0.730	0.465307	
## MVR_PTS10	1.064e+00	8.816e-01	1.207	0.227275	
## MVR_PTS11	1.827e+00	1.134e+00	1.611	0.107167	
## MVR_PTS13	1.556e+01	5.509e+02	0.028	0.977470	
## MVR_PTS2	2.424e-01	1.100e-01	2.203	0.027576	*
## MVR_PTS3	3.070e-01	1.187e-01	2.586	0.009720	**
## MVR_PTS4	2.984e-01	1.285e-01	2.322	0.020234	*

```
## MVR_PTS5          1.747e-01  1.485e-01  1.177 0.239248
## MVR_PTS6          2.258e-01  1.782e-01  1.267 0.205099
## MVR_PTS7          6.224e-01  2.067e-01  3.012 0.002597 **
## MVR_PTS8          1.206e+00  3.208e-01  3.759 0.000171 ***
## MVR_PTS9          1.175e+00  4.015e-01  2.926 0.003434 **
## URBANICITYHighly Urban/ Urban 2.226e+00  1.272e-01  17.493 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 7445.1 on 6447 degrees of freedom
## Residual deviance: 5612.4 on 6321 degrees of freedom
## AIC: 5866.4
##
## Number of Fisher Scoring iterations: 13
```

```
vif(model_4)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## KIDSDRIV      1.375126  4      1.040622
## AGE           3.668116 58      1.011267
## INCOME        2.683910  1      1.638264
## PARENT1       1.702396  1      1.304759
## HOME_VAL      1.808496  1      1.344803
## MSTATUS       1.967872  1      1.402809
## EDUCATION     7.392410  3      1.395718
## JOB           38.972352  8      1.257252
## CAR_USE       2.283576  1      1.511151
## BLUEBOOK      1.790810  1      1.338212
## TIF           1.459261 22      1.008626
## CAR_TYPE      2.970088  5      1.115005
## OLDCLAIM      1.946561  1      1.395192
## CLM_FREQ      2.058156  5      1.074850
## REVOKED       1.395272  1      1.181216
## MVR_PTS       1.682928 12      1.021926
## URBANICITY    1.157788  1      1.076006
```

MODEL 5

5. Using derived variable - wealth accumulation which is INCOME/AGE and removing INCOME to avoid multicollinearity.

```
training$INCOME<- as.integer(training$INCOME)
training$AGE<- as.integer(training$AGE)
training$wealth_acc<-training$INCOME/training$AGE
model_5<-step(glm(TARGET_FLAG~., training %>% select(-INDEX, -TARGET_AMT, -INCOME), family = "binomial")

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Start: AIC=5966.96
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + PARENT1 + HOME_VAL +
## MSTATUS + SEX + EDUCATION + JOB + TRAVTIME + CAR_USE + BLUEBOOK +
## TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
## MVR_PTS + CAR_AGE + URBANICITY + wealth_acc
```

```

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
##
##           Df Deviance    AIC
## - TRAVTIME  95   5665.7 5917.7
## - CAR_AGE   29   5556.7 5940.7
## - HOMEKIDS   5   5532.8 5964.8
## - RED_CAR    1   5526.7 5966.7
## <none>      5525.0 5967.0
## - SEX       1   5527.2 5967.2
## - PARENT1    1   5529.5 5969.5
## - AGE        1   5531.1 5971.1
## - YOJ        20  5572.0 5974.0
## - OLDCLAIM   1   5539.5 5979.5
## - EDUCATION  3   5544.8 5980.8
## - BLUEBOOK   1   5544.0 5984.0
## - wealth_acc 1   5544.3 5984.3
## - KIDSDRIV   4   5552.3 5986.3
## - MSTATUS    1   5546.9 5986.9
## - HOME_VAL   1   5547.5 5987.5
## - MVR_PTS    12  5575.0 5993.0
## - TIF        22  5597.8 5995.8
## - CLM_FREQ   5   5578.2 6010.2
## - JOB        8   5592.6 6018.6
## - CAR_USE    1   5589.2 6029.2

```

```

## - REVOKED      1    5595.9 6035.9
## - CAR_TYPE     5    5609.7 6041.7
## - URBANICITY   1    5964.6 6404.6
##
## Step:  AIC=5917.7
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + PARENT1 + HOME_VAL +
##      MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF +
##      CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS +
##      CAR_AGE + URBANICITY + wealth_acc
##
##           Df Deviance    AIC
## - CAR_AGE    29    5697.1 5891.1
## - HOMEKIDS     5    5673.4 5915.4
## <none>          5665.7 5917.7
## - RED_CAR      1    5668.1 5918.1
## - SEX           1    5669.2 5919.2
## - PARENT1      1    5670.0 5920.0
## - AGE          1    5671.2 5921.2
## - YOJ         20    5715.1 5927.1
## - EDUCATION    3    5682.0 5928.0
## - BLUEBOOK     1    5682.2 5932.2
## - OLDCLAIM     1    5682.3 5932.3
## - wealth_acc   1    5684.4 5934.4
## - KIDSDRIV     4    5691.9 5935.9
## - MSTATUS      1    5687.2 5937.2
## - HOME_VAL     1    5689.4 5939.4
## - MVR_PTS     12    5715.1 5943.1
## - TIF         22    5739.9 5947.9
## - CLM_FREQ     5    5723.8 5965.8
## - JOB          8    5737.5 5973.5
## - CAR_USE      1    5727.9 5977.9
## - REVOKED      1    5737.8 5987.8
## - CAR_TYPE     5    5750.9 5992.9
## - URBANICITY   1    6082.0 6332.0
##
## Step:  AIC=5891.08
## TARGET_FLAG ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + PARENT1 + HOME_VAL +
##      MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF +
##      CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS +
##      URBANICITY + wealth_acc
##
##           Df Deviance    AIC
## - HOMEKIDS     5    5705.7 5889.7
## <none>          5697.1 5891.1
## - RED_CAR      1    5699.5 5891.5
## - SEX           1    5700.2 5892.2
## - PARENT1      1    5700.7 5892.7
## - AGE          1    5703.1 5895.1
## - YOJ         20    5746.1 5900.1
## - BLUEBOOK     1    5713.2 5905.2
## - OLDCLAIM     1    5713.9 5905.9
## - wealth_acc   1    5716.1 5908.1
## - KIDSDRIV     4    5724.0 5910.0
## - HOME_VAL     1    5718.3 5910.3

```

```

## - EDUCATION      3    5724.4 5912.4
## - MSTATUS        1    5721.0 5913.0
## - MVR_PTS       12    5745.5 5915.5
## - TIF            22    5772.2 5922.2
## - CLM_FREQ       5    5757.1 5941.1
## - JOB            8    5770.0 5948.0
## - CAR_USE        1    5761.5 5953.5
## - REVOKED        1    5770.7 5962.7
## - CAR_TYPE       5    5781.7 5965.7
## - URBANICITY     1    6113.7 6305.7
##
## Step:  AIC=5889.73
## TARGET_FLAG ~ KIDSDRIV + AGE + YOJ + PARENT1 + HOME_VAL + MSTATUS +
##      SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK + TIF + CAR_TYPE +
##      RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED + MVR_PTS + URBANICITY +
##      wealth_acc
##
##              Df Deviance    AIC
## <none>              5705.7 5889.7
## - RED_CAR          1    5708.1 5890.1
## - SEX              1    5708.6 5890.6
## - AGE              1    5711.1 5893.1
## - YOJ             20    5755.3 5899.3
## - PARENT1         1    5717.9 5899.9
## - OLDCLAIM        1    5721.5 5903.5
## - BLUEBOOK        1    5722.3 5904.3
## - wealth_acc      1    5725.6 5907.6
## - MSTATUS         1    5726.2 5908.2
## - HOME_VAL        1    5727.7 5909.7
## - EDUCATION        3    5732.9 5910.9
## - MVR_PTS         12    5755.2 5915.2
## - KIDSDRIV         4    5742.5 5918.5
## - TIF             22    5780.1 5920.1
## - CLM_FREQ         5    5765.0 5939.0
## - JOB             8    5779.0 5947.0
## - CAR_USE          1    5770.3 5952.3
## - REVOKED          1    5778.2 5960.2
## - CAR_TYPE         5    5790.0 5964.0
## - URBANICITY       1    6121.9 6303.9

```

```
summary(model_5)
```

```

##
## Call:
## glm(formula = TARGET_FLAG ~ KIDSDRIV + AGE + YOJ + PARENT1 +
##      HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##      TIF + CAR_TYPE + RED_CAR + OLDCLAIM + CLM_FREQ + REVOKED +
##      MVR_PTS + URBANICITY + wealth_acc, family = "binomial", data = training %>%
##      select(-INDEX, -TARGET_AMT, -INCOME))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5528  -0.7011  -0.3973   0.5915   3.1138
##
## Coefficients:

```

##	Estimate	Std. Error	z value	Pr(> z)	
## (Intercept)	-3.033e+00	3.848e-01	-7.881	3.25e-15	***
## KIDSDRIV1	4.875e-01	1.220e-01	3.996	6.45e-05	***
## KIDSDRIV2	7.463e-01	1.714e-01	4.354	1.34e-05	***
## KIDSDRIV3	8.140e-01	3.372e-01	2.414	0.015764	*
## KIDSDRIV4	-1.377e+01	4.862e+02	-0.028	0.977410	
## AGE	1.536e-02	6.657e-03	2.307	0.021037	*
## YOJ1	-1.459e+01	3.162e+02	-0.046	0.963183	
## YOJ10	-9.776e-01	2.267e-01	-4.312	1.62e-05	***
## YOJ11	-1.100e+00	2.186e-01	-5.031	4.87e-07	***
## YOJ12	-1.082e+00	2.170e-01	-4.988	6.11e-07	***
## YOJ13	-1.060e+00	2.202e-01	-4.814	1.48e-06	***
## YOJ14	-1.020e+00	2.262e-01	-4.508	6.54e-06	***
## YOJ15	-9.603e-01	2.450e-01	-3.920	8.85e-05	***
## YOJ16	-8.601e-01	2.748e-01	-3.130	0.001750	**
## YOJ17	-5.508e-01	3.389e-01	-1.625	0.104141	
## YOJ18	-1.191e+00	6.053e-01	-1.967	0.049178	*
## YOJ19	7.440e-01	7.395e-01	1.006	0.314360	
## YOJ2	-1.468e+00	8.821e-01	-1.664	0.096136	.
## YOJ23	-1.344e+01	5.997e+02	-0.022	0.982118	
## YOJ3	-1.221e+00	5.798e-01	-2.106	0.035173	*
## YOJ4	-1.908e+00	5.940e-01	-3.213	0.001314	**
## YOJ5	-1.104e+00	3.639e-01	-3.035	0.002408	**
## YOJ6	-1.159e+00	2.965e-01	-3.907	9.34e-05	***
## YOJ7	-1.286e+00	2.651e-01	-4.851	1.23e-06	***
## YOJ8	-9.498e-01	2.515e-01	-3.776	0.000159	***
## YOJ9	-9.609e-01	2.353e-01	-4.084	4.44e-05	***
## PARENT1Yes	3.981e-01	1.139e-01	3.495	0.000474	***
## HOME_VAL	-1.661e-06	3.583e-07	-4.636	3.55e-06	***
## MSTATUSYes	-4.264e-01	9.381e-02	-4.545	5.50e-06	***
## SEXM	2.116e-01	1.253e-01	1.689	0.091165	.
## EDUCATIONHigh School	4.600e-01	9.139e-02	5.033	4.82e-07	***
## EDUCATIONMasters	-2.753e-02	1.562e-01	-0.176	0.860080	
## EDUCATIONPhD	5.441e-02	1.983e-01	0.274	0.783793	
## JOBBBlue Collar	3.897e-01	2.127e-01	1.832	0.066973	.
## JOBClerical	5.850e-01	2.228e-01	2.626	0.008643	**
## JOBDoctor	-2.152e-01	2.901e-01	-0.742	0.458240	
## JOBHome Maker	2.312e-01	2.437e-01	0.949	0.342761	
## JOBLawyer	3.142e-01	1.917e-01	1.639	0.101271	
## JOBManager	-6.023e-01	1.980e-01	-3.042	0.002354	**
## JOBProfessional	2.462e-01	2.042e-01	1.206	0.227795	
## JOBStudent	1.755e-01	2.522e-01	0.696	0.486549	
## CAR_USEPrivate	-7.944e-01	9.985e-02	-7.956	1.78e-15	***
## BLUEBOOK	-2.361e-05	5.861e-06	-4.029	5.60e-05	***
## TIF10	-7.214e-01	1.289e-01	-5.596	2.19e-08	***
## TIF11	-3.297e-01	1.982e-01	-1.664	0.096173	.
## TIF12	-5.506e-01	4.582e-01	-1.202	0.229488	
## TIF13	-4.611e-01	1.934e-01	-2.384	0.017134	*
## TIF14	-1.009e+00	4.063e-01	-2.485	0.012961	*
## TIF15	1.497e-01	4.739e-01	0.316	0.752119	
## TIF16	-1.195e+00	5.420e-01	-2.204	0.027509	*
## TIF17	-6.415e-01	3.094e-01	-2.073	0.038129	*
## TIF18	-6.087e-01	6.107e-01	-0.997	0.318871	
## TIF19	-1.302e+01	3.624e+02	-0.036	0.971333	

```

## TIF2          1.603e-01  9.820e-01  0.163 0.870301
## TIF20        -1.347e+01  3.331e+02 -0.040 0.967745
## TIF21         1.039e-01  7.542e-01  0.138 0.890453
## TIF22        -1.442e+01  4.767e+02 -0.030 0.975864
## TIF25        -1.235e+01  6.019e+02 -0.021 0.983636
## TIF3         -2.139e-01  1.483e-01 -1.442 0.149196
## TIF4         -2.519e-01  1.022e-01 -2.466 0.013666 *
## TIF5         -6.027e-01  4.382e-01 -1.376 0.168942
## TIF6         -3.830e-01  1.001e-01 -3.824 0.000131 ***
## TIF7         -5.755e-01  1.364e-01 -4.218 2.47e-05 ***
## TIF8          5.367e-01  3.939e-01  1.362 0.173067
## TIF9        -6.971e-01  2.193e-01 -3.179 0.001480 **
## CAR_TYPEPanel Truck  6.369e-01  1.809e-01  3.520 0.000432 ***
## CAR_TYPEPickup      5.455e-01  1.140e-01  4.784 1.72e-06 ***
## CAR_TYPESports Car   1.114e+00  1.466e-01  7.597 3.03e-14 ***
## CAR_TYPESUV          8.405e-01  1.253e-01  6.706 2.00e-11 ***
## CAR_TYPEVan          6.247e-01  1.434e-01  4.357 1.32e-05 ***
## RED_CARyes         -1.516e-01  9.836e-02 -1.541 0.123258
## OLDCLAIM          -1.862e-05  4.754e-06 -3.917 8.96e-05 ***
## CLM_FREQ1          6.197e-01  1.147e-01  5.402 6.61e-08 ***
## CLM_FREQ2          6.811e-01  1.071e-01  6.359 2.03e-10 ***
## CLM_FREQ3          6.695e-01  1.204e-01  5.560 2.69e-08 ***
## CLM_FREQ4          8.542e-01  1.968e-01  4.340 1.43e-05 ***
## CLM_FREQ5          1.131e+00  6.597e-01  1.714 0.086478 .
## REVOKEDYes         9.056e-01  1.055e-01  8.587 < 2e-16 ***
## MVR_PTS1          6.696e-02  1.038e-01  0.645 0.519044
## MVR_PTS10         8.875e-01  8.725e-01  1.017 0.309039
## MVR_PTS11         1.951e+00  1.090e+00  1.790 0.073485 .
## MVR_PTS13         1.517e+01  5.697e+02  0.027 0.978756
## MVR_PTS2          2.283e-01  1.089e-01  2.097 0.036032 *
## MVR_PTS3          2.969e-01  1.171e-01  2.536 0.011208 *
## MVR_PTS4          3.215e-01  1.268e-01  2.535 0.011238 *
## MVR_PTS5          2.208e-01  1.467e-01  1.506 0.132084
## MVR_PTS6          2.927e-01  1.752e-01  1.671 0.094743 .
## MVR_PTS7          7.388e-01  2.031e-01  3.637 0.000276 ***
## MVR_PTS8          1.288e+00  3.137e-01  4.104 4.05e-05 ***
## MVR_PTS9          1.284e+00  3.979e-01  3.226 0.001255 **
## URBANICITYHighly Urban/ Urban 2.172e+00  1.246e-01 17.435 < 2e-16 ***
## wealth_acc        3.425e+00  8.834e-01  3.877 0.000106 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 7445.1 on 6447 degrees of freedom
## Residual deviance: 5705.7 on 6356 degrees of freedom
## AIC: 5889.7
##
## Number of Fisher Scoring iterations: 13
vif(model_5)

##          GVIF Df GVIF^(1/(2*Df))
## KIDSDRIV 1.208927 4      1.024000
## AGE      3.213013 1      1.792488

```

```
## YOJ          4.651641 20          1.039179
## PARENT1      1.648777  1          1.284047
## HOME_VAL     1.778284  1          1.333523
## MSTATUS      2.023930  1          1.422649
## SEX          3.612611  1          1.900687
## EDUCATION    7.210922  3          1.389948
## JOB          30.089479  8          1.237089
## CAR_USE      2.269510  1          1.506489
## BLUEBOOK     2.154222  1          1.467727
## TIF          1.281940 22          1.005661
## CAR_TYPE     6.564468  5          1.207035
## RED_CAR      1.843016  1          1.357577
## OLDCLAIM     1.922974  1          1.386714
## CLM_FREQ     1.983235  5          1.070872
## REVOKED      1.376466  1          1.173229
## MVR_PTS      1.468636 12          1.016143
## URBANICITY   1.133679  1          1.064743
## wealth_acc   4.346726  1          2.084880
```

MODEL 6

6. Building a model on a data set where bins are created for the following continuous variables: AGE, YOJ, INCOME, TRAVTIME, BLUEBOOK, TIF, OLDCLAIM, CAR_AGE, HOME_VAL and converting them to factors.

```
## 'data.frame':   6448 obs. of  35 variables:
## $ INDEX       : Factor w/ 8161 levels "1","100","1000",...: 1 1132 2898 5531 430 508 602 690 778 872 ...
## $ TARGET_FLAG : Factor w/ 2 levels "0","1": 1 1 1 2 2 1 2 1 1 2 ...
## $ TARGET_AMT  : Factor w/ 1949 levels "0","1005.214379",...: 1 1 1 612 460 1 1581 1 1 83 ...
## $ KIDSDRIV    : Factor w/ 5 levels "0","1","2","3",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ AGE         : int  45 28 20 19 19 35 38 28 40 38 ...
## $ HOMEKIDS     : Factor w/ 6 levels "0","1","2","3",...: 1 1 2 2 1 1 1 1 1 1 ...
## $ YOJ         : int  4 4 3 5 3 19 7 17 4 4 ...
## $ INCOME       : int  67349 91449 16039 125301 62978 106952 77100 52642 59162 130795 ...
## $ PARENT1     : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 1 1 1 1 1 ...
## $ HOME_VAL     : int  0 257252 124191 0 0 0 0 209970 180232 0 ...
## $ MSTATUS      : Factor w/ 2 levels "No","Yes": 1 1 2 1 1 1 1 2 2 1 ...
## $ SEX         : Factor w/ 2 levels "F","M": 2 2 1 1 1 2 1 1 2 2 ...
## $ EDUCATION    : Factor w/ 4 levels "Bachelors","High School",...: 4 2 2 1 1 1 3 3 1 4 ...
## $ JOB         : Factor w/ 9 levels "", "Blue Collar",...: 8 2 3 2 3 8 6 8 7 1 ...
## $ TRAVTIME     : int  10 19 47 43 31 45 12 33 22 63 ...
## $ CAR_USE      : Factor w/ 2 levels "Commercial","Private": 2 1 2 1 2 1 2 2 1 1 ...
## $ BLUEBOOK     : int  14230 14940 4010 17430 11200 18510 18300 22420 17600 28340 ...
## $ TIF         : int  3 1 18 1 1 21 1 21 21 20 ...
## $ CAR_TYPE     : Factor w/ 6 levels "Minivan","Panel Truck",...: 1 1 5 4 5 6 4 1 6 2 ...
## $ RED_CAR      : Factor w/ 2 levels "no","yes": 2 2 1 1 1 1 1 1 2 2 ...
## $ OLDCLAIM     : int  4461 0 38690 0 0 0 0 0 5028 0 ...
## $ CLM_FREQ     : Factor w/ 6 levels "0","1","2","3",...: 3 1 3 1 1 1 1 1 3 1 ...
## $ REVOKED      : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 2 1 ...
## $ MVR_PTS      : Factor w/ 13 levels "0","1","10","11",...: 7 1 7 1 1 2 1 1 7 7 ...
## $ CAR_AGE      : int  12 3 4 28 3 11 5 3 30 4 ...
## $ URBANICITY   : Factor w/ 2 levels "Highly Rural/ Rural",...: 2 2 2 2 2 1 2 1 2 2 ...
## $ HOME_VAL_BINS: Factor w/ 2 levels "<1000","5000+": 1 2 2 1 1 1 1 2 2 1 ...
## $ AGE_BINS     : Factor w/ 6 levels "<20","20-30",...: 4 2 2 1 1 3 3 2 4 3 ...
```



```

## $ INCOME_BINS : Factor w/ 8 levels "<1000","1000-20000",...: 7 8 2 4 7 3 7 6 6 4 ...
## $ YOJ_BINS : Factor w/ 5 levels "<5","10-15","15-20",...: 1 1 1 5 1 3 5 3 1 1 ...
## $ TRAVTIME_BINS: Factor w/ 6 levels "<5","25-45","45-65",...: 4 4 3 2 2 3 4 2 4 3 ...
## $ BLUEBOOK_BINS: Factor w/ 7 levels "11500-21500",...: 1 1 2 1 2 1 1 3 1 3 ...
## $ TIF_BINS : Factor w/ 5 levels "1-5","10-15",...: 1 1 3 1 1 4 1 4 4 4 ...
## $ OLDCLAIM_BINS: Factor w/ 7 levels "<500","1000-1500",...: 6 1 6 1 1 1 1 1 6 1 ...
## $ CAR_AGE_BINS : Factor w/ 6 levels "1-5","10-15",...: 2 1 1 5 1 2 6 1 5 1 ...
## - attr(*, "na.action")= 'omit' Named int 4 5 7 8 21 29 45 46 49 54 ...
## ..- attr(*, "names")= chr "4" "5" "7" "8" ...

model_6<-step(glm(TARGET_FLAG~., training_BINS %>% select(-INDEX,-TARGET_AMT, -AGE, -YOJ, -TRAVTIME, -BLUEBOOK_BINS, -TIF_BINS, -OLDCLAIM_BINS, -CAR_AGE_BINS), weights=weights, method="glm", family="binomial", link="logit", na.action="omit", na.rm=TRUE))

## Start: AIC=5901.86
## TARGET_FLAG ~ KIDSDRIV + HOMEKIDS + PARENT1 + MSTATUS + SEX +
## EDUCATION + JOB + CAR_USE + CAR_TYPE + RED_CAR + CLM_FREQ +
## REVOKED + MVR_PTS + URBANICITY + HOME_VAL_BINS + AGE_BINS +
## INCOME_BINS + YOJ_BINS + TRAVTIME_BINS + BLUEBOOK_BINS +
## TIF_BINS + OLDCLAIM_BINS + CAR_AGE_BINS
##
## Df Deviance AIC
## - CLM_FREQ 4 5719.2 5895.2
## - HOMEKIDS 5 5722.6 5896.6
## - OLDCLAIM_BINS 5 5724.9 5898.9
## - YOJ_BINS 4 5724.7 5900.7
## - RED_CAR 1 5719.2 5901.2
## <none> 5717.9 5901.9
## - PARENT1 1 5720.7 5902.7
## - CAR_AGE_BINS 5 5729.1 5903.1
## - SEX 1 5721.2 5903.2
## - BLUEBOOK_BINS 6 5732.1 5904.1
## - EDUCATION 3 5726.7 5904.7
## - TIF_BINS 4 5734.1 5910.1
## - TRAVTIME_BINS 5 5736.4 5910.4
## - HOME_VAL_BINS 1 5735.2 5917.2
## - MSTATUS 1 5736.5 5918.5
## - INCOME_BINS 7 5754.2 5924.2
## - KIDSDRIV 4 5749.9 5925.9
## - AGE_BINS 5 5753.7 5927.7
## - MVR_PTS 12 5776.5 5936.5
## - REVOKED 1 5778.0 5960.0
## - JOB 8 5794.9 5962.9
## - CAR_TYPE 5 5789.3 5963.3
## - CAR_USE 1 5785.0 5967.0
## - URBANICITY 1 6151.3 6333.3
##
## Step: AIC=5895.21
## TARGET_FLAG ~ KIDSDRIV + HOMEKIDS + PARENT1 + MSTATUS + SEX +
## EDUCATION + JOB + CAR_USE + CAR_TYPE + RED_CAR + REVOKED +
## MVR_PTS + URBANICITY + HOME_VAL_BINS + AGE_BINS + INCOME_BINS +
## YOJ_BINS + TRAVTIME_BINS + BLUEBOOK_BINS + TIF_BINS + OLDCLAIM_BINS +
## CAR_AGE_BINS
##
## Df Deviance AIC
## - HOMEKIDS 5 5724.0 5890.0
## - YOJ_BINS 4 5726.1 5894.1

```

```

## - RED_CAR      1    5720.5 5894.5
## <none>          5719.2 5895.2
## - PARENT1      1    5722.0 5896.0
## - CAR_AGE_BINS  5    5730.4 5896.4
## - SEX          1    5722.5 5896.5
## - BLUEBOOK_BINS 6    5733.6 5897.6
## - EDUCATION    3    5728.0 5898.0
## - TIF_BINS     4    5735.6 5903.6
## - TRAVTIME_BINS 5    5737.8 5903.8
## - HOME_VAL_BINS 1    5736.6 5910.6
## - MSTATUS      1    5738.0 5912.0
## - INCOME_BINS  7    5756.1 5918.1
## - KIDSDRIV     4    5751.5 5919.5
## - AGE_BINS     5    5755.2 5921.2
## - OLDCLAIM_BINS 6    5764.0 5928.0
## - MVR_PTS      12   5777.5 5929.5
## - REVOKED      1    5779.1 5953.1
## - JOB          8    5796.2 5956.2
## - CAR_TYPE     5    5790.6 5956.6
## - CAR_USE      1    5786.5 5960.5
## - URBANICITY   1    6152.9 6326.9
##
## Step:  AIC=5889.98
## TARGET_FLAG ~ KIDSDRIV + PARENT1 + MSTATUS + SEX + EDUCATION +
##      JOB + CAR_USE + CAR_TYPE + RED_CAR + REVOKED + MVR_PTS +
##      URBANICITY + HOME_VAL_BINS + AGE_BINS + INCOME_BINS + YOJ_BINS +
##      TRAVTIME_BINS + BLUEBOOK_BINS + TIF_BINS + OLDCLAIM_BINS +
##      CAR_AGE_BINS
##
##           Df Deviance   AIC
## - YOJ_BINS      4    5731.1 5889.1
## - RED_CAR       1    5725.3 5889.3
## <none>           5724.0 5890.0
## - SEX          1    5727.0 5891.0
## - CAR_AGE_BINS  5    5735.5 5891.5
## - EDUCATION    3    5732.4 5892.4
## - BLUEBOOK_BINS 6    5738.6 5892.6
## - PARENT1      1    5730.9 5894.9
## - TIF_BINS     4    5740.2 5898.2
## - TRAVTIME_BINS 5    5742.9 5898.9
## - HOME_VAL_BINS 1    5741.6 5905.6
## - MSTATUS      1    5742.0 5906.0
## - INCOME_BINS  7    5760.5 5912.5
## - AGE_BINS     5    5764.3 5920.3
## - OLDCLAIM_BINS 6    5769.0 5923.0
## - KIDSDRIV     4    5765.4 5923.4
## - MVR_PTS      12   5782.9 5924.9
## - REVOKED      1    5783.5 5947.5
## - CAR_TYPE     5    5794.8 5950.8
## - JOB          8    5801.4 5951.4
## - CAR_USE      1    5791.3 5955.3
## - URBANICITY   1    6157.0 6321.0
##
## Step:  AIC=5889.1

```

```

## TARGET_FLAG ~ KIDSDRIV + PARENT1 + MSTATUS + SEX + EDUCATION +
##      JOB + CAR_USE + CAR_TYPE + RED_CAR + REVOKED + MVR_PTS +
##      URBANICITY + HOME_VAL_BINS + AGE_BINS + INCOME_BINS + TRAVTIME_BINS +
##      BLUEBOOK_BINS + TIF_BINS + OLDCLAIM_BINS + CAR_AGE_BINS
##
##           Df Deviance   AIC
## - RED_CAR      1   5732.4 5888.4
## <none>           5731.1 5889.1
## - SEX          1   5734.1 5890.1
## - CAR_AGE_BINS  5   5742.7 5890.7
## - BLUEBOOK_BINS 6   5745.3 5891.3
## - EDUCATION     3   5739.6 5891.6
## - PARENT1       1   5738.3 5894.3
## - TIF_BINS      4   5746.9 5896.9
## - TRAVTIME_BINS 5   5749.9 5897.9
## - HOME_VAL_BINS 1   5748.2 5904.2
## - MSTATUS       1   5749.9 5905.9
## - INCOME_BINS   7   5770.8 5914.8
## - AGE_BINS      5   5772.4 5920.4
## - OLDCLAIM_BINS 6   5775.3 5921.3
## - KIDSDRIV      4   5772.6 5922.6
## - MVR_PTS      12   5789.7 5923.7
## - REVOKED       1   5791.4 5947.4
## - JOB           8   5807.5 5949.5
## - CAR_TYPE      5   5802.8 5950.8
## - CAR_USE       1   5799.4 5955.4
## - URBANICITY    1   6164.2 6320.2
##
## Step:  AIC=5888.36
## TARGET_FLAG ~ KIDSDRIV + PARENT1 + MSTATUS + SEX + EDUCATION +
##      JOB + CAR_USE + CAR_TYPE + REVOKED + MVR_PTS + URBANICITY +
##      HOME_VAL_BINS + AGE_BINS + INCOME_BINS + TRAVTIME_BINS +
##      BLUEBOOK_BINS + TIF_BINS + OLDCLAIM_BINS + CAR_AGE_BINS
##
##           Df Deviance   AIC
## - SEX          1   5734.3 5888.3
## <none>           5732.4 5888.4
## - CAR_AGE_BINS  5   5743.9 5889.9
## - BLUEBOOK_BINS 6   5746.3 5890.3
## - EDUCATION     3   5741.0 5891.0
## - PARENT1       1   5739.7 5893.7
## - TIF_BINS      4   5748.3 5896.3
## - TRAVTIME_BINS 5   5751.2 5897.2
## - HOME_VAL_BINS 1   5749.3 5903.3
## - MSTATUS       1   5751.1 5905.1
## - INCOME_BINS   7   5771.9 5913.9
## - AGE_BINS      5   5773.4 5919.4
## - OLDCLAIM_BINS 6   5776.7 5920.7
## - KIDSDRIV      4   5773.7 5921.7
## - MVR_PTS      12   5790.9 5922.9
## - REVOKED       1   5792.4 5946.4
## - JOB           8   5809.0 5949.0
## - CAR_TYPE      5   5804.6 5950.6
## - CAR_USE       1   5800.5 5954.5

```

```
## - URBANICITY      1   6165.3 6319.3
##
## Step:   AIC=5888.28
## TARGET_FLAG ~ KIDSDRIV + PARENT1 + MSTATUS + EDUCATION + JOB +
##      CAR_USE + CAR_TYPE + REVOKED + MVR_PTS + URBANICITY + HOME_VAL_BINS +
##      AGE_BINS + INCOME_BINS + TRAVTIME_BINS + BLUEBOOK_BINS +
##      TIF_BINS + OLDCLAIM_BINS + CAR_AGE_BINS
##
##              Df Deviance   AIC
## <none>              5734.3 5888.3
## - CAR_AGE_BINS      5   5745.9 5889.9
## - EDUCATION         3   5742.8 5890.8
## - PARENT1           1   5741.4 5893.4
## - BLUEBOOK_BINS     6   5752.6 5894.6
## - TIF_BINS          4   5750.2 5896.2
## - TRAVTIME_BINS     5   5753.2 5897.2
## - HOME_VAL_BINS     1   5751.1 5903.1
## - MSTATUS           1   5753.3 5905.3
## - INCOME_BINS       7   5774.5 5914.5
## - AGE_BINS          5   5776.4 5920.4
## - OLDCLAIM_BINS     6   5778.9 5920.9
## - KIDSDRIV          4   5775.2 5921.2
## - MVR_PTS           12   5792.4 5922.4
## - REVOKED           1   5794.5 5946.5
## - JOB               8   5810.5 5948.5
## - CAR_USE           1   5802.3 5954.3
## - CAR_TYPE          5   5815.7 5959.7
## - URBANICITY        1   6167.2 6319.2
```

```
summary(model_6)
```

```
##
## Call:
## glm(formula = TARGET_FLAG ~ KIDSDRIV + PARENT1 + MSTATUS + EDUCATION +
##      JOB + CAR_USE + CAR_TYPE + REVOKED + MVR_PTS + URBANICITY +
##      HOME_VAL_BINS + AGE_BINS + INCOME_BINS + TRAVTIME_BINS +
##      BLUEBOOK_BINS + TIF_BINS + OLDCLAIM_BINS + CAR_AGE_BINS,
##      family = "binomial", data = training_BINS %>% select(-INDEX,
##      -TARGET_AMT, -AGE, -YOJ, -TRAVTIME, -BLUEBOOK, -OLDCLAIM,
##      -CAR_AGE, -TIF, -HOME_VAL, -INCOME))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5417  -0.7048  -0.3981   0.6101   2.9265
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.47488    0.43057  -5.748 9.04e-09 ***
## KIDSDRIV1      0.52031    0.12475   4.171 3.04e-05 ***
## KIDSDRIV2      0.80656    0.17430   4.627 3.70e-06 ***
## KIDSDRIV3      0.96750    0.33412   2.896 0.003784 **
## KIDSDRIV4     -13.05643   292.81781  -0.045 0.964435
## PARENT1Yes     0.30441    0.11413   2.667 0.007649 **
## MSTATUSYes    -0.43023    0.09805  -4.388 1.14e-05 ***
## EDUCATIONHigh School  0.28995    0.10092   2.873 0.004064 **
```

## EDUCATIONMasters	-0.02865	0.16711	-0.171	0.863858	
## EDUCATIONPhD	0.02190	0.21118	0.104	0.917394	
## JOBBBlue Collar	0.36954	0.21242	1.740	0.081915	.
## JOBClerical	0.55765	0.22569	2.471	0.013478	*
## JOBDoctor	-0.14131	0.28983	-0.488	0.625855	
## JOBHome Maker	0.02082	0.25992	0.080	0.936158	
## JOBLawyer	0.34810	0.19213	1.812	0.070028	.
## JOBManager	-0.55791	0.19886	-2.806	0.005022	**
## JOBProfessional	0.26733	0.20416	1.309	0.190409	
## JOBStudent	-0.24326	0.27438	-0.887	0.375315	
## CAR_USEPrivate	-0.81455	0.09979	-8.163	3.27e-16	***
## CAR_TYPEPanel Truck	0.52180	0.17670	2.953	0.003147	**
## CAR_TYPEPickup	0.47742	0.11783	4.052	5.08e-05	***
## CAR_TYPESports Car	0.96280	0.12426	7.749	9.29e-15	***
## CAR_TYPESUV	0.72548	0.09874	7.347	2.02e-13	***
## CAR_TYPEVan	0.61664	0.13695	4.502	6.72e-06	***
## REVOKEDYes	0.71777	0.09170	7.827	4.98e-15	***
## MVR_PTS1	0.09838	0.10347	0.951	0.341683	
## MVR_PTS10	0.88767	0.82232	1.079	0.280375	
## MVR_PTS11	2.27770	1.09482	2.080	0.037486	*
## MVR_PTS13	14.92806	313.76010	0.048	0.962053	
## MVR_PTS2	0.28320	0.10827	2.616	0.008906	**
## MVR_PTS3	0.31999	0.11659	2.745	0.006060	**
## MVR_PTS4	0.33598	0.12604	2.666	0.007686	**
## MVR_PTS5	0.24591	0.14687	1.674	0.094063	.
## MVR_PTS6	0.32690	0.17473	1.871	0.061357	.
## MVR_PTS7	0.80027	0.20094	3.983	6.81e-05	***
## MVR_PTS8	1.38478	0.31406	4.409	1.04e-05	***
## MVR_PTS9	1.26570	0.39717	3.187	0.001439	**
## URBANICITYHighly Urban/ Urban	2.20281	0.12430	17.722	< 2e-16	***
## HOME_VAL_BINS5000+	-0.39792	0.09747	-4.082	4.46e-05	***
## AGE_BINS20-30	-0.34432	0.11037	-3.120	0.001809	**
## AGE_BINS30-40	-0.53465	0.11450	-4.670	3.02e-06	***
## AGE_BINS40-50	0.05676	0.13656	0.416	0.677669	
## AGE_BINS50-60	0.01095	0.35156	0.031	0.975161	
## AGE_BINS60-70	-11.98257	535.41123	-0.022	0.982145	
## INCOME_BINS1000-20000	-0.56757	0.16335	-3.475	0.000512	***
## INCOME_BINS100000-120000	-1.20507	0.23912	-5.040	4.66e-07	***
## INCOME_BINS120000+	-1.26013	0.23690	-5.319	1.04e-07	***
## INCOME_BINS20000-40000	-0.82333	0.19122	-4.306	1.66e-05	***
## INCOME_BINS40000-60000	-0.74864	0.19952	-3.752	0.000175	***
## INCOME_BINS60000-80000	-0.86174	0.20890	-4.125	3.71e-05	***
## INCOME_BINS80000-100000	-1.18133	0.22549	-5.239	1.61e-07	***
## TRAVTIME_BINS25-45	0.38692	0.26036	1.486	0.137263	
## TRAVTIME_BINS45-65	0.52378	0.26609	1.968	0.049017	*
## TRAVTIME_BINS5-25	0.17065	0.26348	0.648	0.517189	
## TRAVTIME_BINS65-85	0.21226	0.31641	0.671	0.502337	
## TRAVTIME_BINS85-105	-0.11805	0.41511	-0.284	0.776127	
## BLUEBOOK_BINS1500-11500	0.27412	0.08472	3.235	0.001215	**
## BLUEBOOK_BINS21500-31500	-0.14954	0.11328	-1.320	0.186815	
## BLUEBOOK_BINS31500-41500	-0.05066	0.19620	-0.258	0.796234	
## BLUEBOOK_BINS41500-51500	0.11412	0.47755	0.239	0.811127	
## BLUEBOOK_BINS51500-61500	1.08353	1.43642	0.754	0.450651	
## BLUEBOOK_BINS61500-71500	1.17139	1.33939	0.875	0.381807	

```
## TIF_BINS10-15          -0.36951    0.40439   -0.914  0.360858
## TIF_BINS15-20          -0.06061    0.08625   -0.703  0.482247
## TIF_BINS20-25          -0.26606    0.08088   -3.290  0.001004 **
## TIF_BINS5-10           -0.37804    0.14352   -2.634  0.008439 **
## OLDCLAIM_BINS1000-1500  0.62472    0.21268    2.937  0.003310 **
## OLDCLAIM_BINS1500-2000  0.50691    0.25669    1.975  0.048290 *
## OLDCLAIM_BINS2000-2500  0.94970    0.25954    3.659  0.000253 ***
## OLDCLAIM_BINS2500-3000  0.09229    0.26552    0.348  0.728159
## OLDCLAIM_BINS3000+      0.44105    0.07787    5.664  1.48e-08 ***
## OLDCLAIM_BINS500-1000   0.36142    0.26817    1.348  0.177746
## CAR_AGE_BINS10-15       0.07794    0.15327    0.508  0.611103
## CAR_AGE_BINS15-20      -0.49564    0.26284   -1.886  0.059332 .
## CAR_AGE_BINS20-25      -0.16757    0.35600   -0.471  0.637857
## CAR_AGE_BINS25+        -0.10616    0.08158   -1.301  0.193146
## CAR_AGE_BINS5-10       -0.24499    0.11198   -2.188  0.028683 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 7445.1 on 6447 degrees of freedom
## Residual deviance: 5734.3 on 6371 degrees of freedom
## AIC: 5888.3
##
## Number of Fisher Scoring iterations: 12
```

```
vif(model_6)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## KIDSDRIV      1.299621  4      1.033302
## PARENT1       1.664139  1      1.290015
## MSTATUS       2.220832  1      1.490246
## EDUCATION     11.311217  3      1.498252
## JOB           57.363514  8      1.287997
## CAR_USE       2.279000  1      1.509636
## CAR_TYPE      3.388980  5      1.129814
## REVOKED       1.031809  1      1.015780
## MVR_PTS       1.425387 12      1.014878
## URBANICITY    1.133109  1      1.064476
## HOME_VAL_BINS 2.005902  1      1.416299
## AGE_BINS      1.576845  5      1.046596
## INCOME_BINS   7.691790  7      1.156878
## TRAVTIME_BINS 1.073954  5      1.007160
## BLUEBOOK_BINS 2.514857  6      1.079882
## TIF_BINS      1.053493  4      1.006535
## OLDCLAIM_BINS 1.400568  6      1.028471
## CAR_AGE_BINS  2.431326  5      1.092910
```

Assessing selected models

Cleaning Test Data (test data set should have same variables, order of variables, format of variables and type of variables as a training set for model testing).

```
##      INDEX TARGET_FLAG TARGET_AMT    KIDSDRIV      AGE    HOMEKIDS
```

```
##          0          0          0          0          0          0
##      YOJ      INCOME      PARENT1      HOME_VAL      MSTATUS      SEX
##          0          0          0          0          0          0
##  EDUCATION      JOB      TRAVTIME      CAR_USE      BLUEBOOK      TIF
##          0          0          0          0          0          0
##      CAR_TYPE      RED_CAR      OLDCLAIM      CLM_FREQ      REVOKED      MVR_PTS
##          0          0          0          0          0          0
##      CAR_AGE      URBANICITY
##          0          0
```

AIC, BIC, Loglik, pseudoR2

```
##          AIC          BIC      loglik  pseudoR2
## model_1 5879.306 6739.290 -2812.653 0.2444238
## model_2 5879.306 6739.290 -2812.653 0.2444238
## model_3 5924.762 6730.574 -2843.381 0.2361692
## model_4 5866.371 6726.355 -2806.186 0.2461612
## model_5 5889.726 6512.706 -2852.863 0.2336221
## model_6 5888.275 6409.683 -2867.138 0.2297874
```

Choosing the best model and applying it on test data set, saving results as a csv file

Model_1 was selected as the best model based on AIC, BIC, Loglik and pseudoR2

Assessing the performance of the chosen model: ROC and AUC

