Data 621 - HW4

Devin Teran, Atina Karim, Tom Hill, Amit Kapoor

5/2/2021

Contents

Overview

In this homework assignment, you will explore, analyze and model a data set containing approximately 8000 records representing a customer at an auto insurance company. Each record has two response variables. The first response variable, TARGET_FLAG, is a 1 or a 0. A "1" means that the person was in a car crash. A zero means that the person was not in a car crash. The second response variable is TARGET_AMT. This value is zero if the person did not crash their car. But if they did crash their car, this number will be a value greater than zero.

Your objective is to build multiple linear regression and binary logistic regression models on the training data to predict the probability that a person will crash their car and also the amount of money it will cost if the person does crash their car. You can only use the variables given to you (or variables that you derive from the variables provided). Below is a short description of the variables of interest in the data set:

Response Variables:

VARIABLE NAME	DEFINITION	THEORETICAL EFFECT
TARGET_FLAG	Was Car in a crash? 1=YES 0=NO	None
TARGET_AMT	If car was in a crash, what was the cost	None

Explanatory Variables:

VARIABLE NAME	DEFINITION	THEORETICAL EFFECT
AGE	Age of Driver	Very young people tend to be risky. Maybe very old people also.
BLUEBOOK	Value of Vehicle	Unknown effect on probability of collision, but probably effect the payout if there is a crash
CAR_AGE	Vehicle Age	Unknown effect on probability of collision, but probably effect the payout if there is a crash
CAR_TYPE	Type of Car	Unknown effect on probability of collision, but probably effect the payout if there is a crash
CAR_USE	Vehicle Use	Commercial vehicles are driven more, so might increase probability of collision
CLM_FREQ	# Claims (Past 5 Years)	The more claims you filed in the past, the more you are likely to file in the future
EDUCATION	Max Education Level	Unknown effect, but in theory more educated people tend to drive more safely
HOMEKIDS	# Children at Home	Unknown effect
HOME_VAL	Home Value	In theory, home owners tend to drive more responsibly
INCOME	Income	In theory, rich people tend to get into fewer crashes
JOB	Job Category	In theory, white collar jobs tend to be safer
KIDSDRIV	# Driving Children	When teenagers drive your car, you are more likely to get into crashes
MSTATUS	Marital Status	In theory, married people drive more safely
MVR_PTS	Motor Vehicle Record Points	If you get lots of traffic tickets, you tend to get into more crashes
OLDCLAIM	Total Claims (Past 5 Years)	If your total payout over the past five years was high, this suggests future payouts will be high
PARENT1	Single Parent	Unknown effect
RED_CAR	A Red Car	Urban legend says that red cars (especially red sports cars) are more risky. Is that true?
REVOKED	License Revoked (Past 7 Years)	If your license was revoked in the past 7 years, you probably are a more risky driver.
SEX	Gender	Urban legend says that women have less crashes then men. Is that true?
TIF	Time in Force	People who have been customers for a long time are usually more safe.
TRAVTIME	Distance to Work	Long drives to work usually suggest greater risk
URBANICITY	Home/Work Area	Unknown
YOJ	Years on Job	People who stay at a job for a long time are usually more safe

Data Exploration

```
## Rows: 8,161
## Columns: 26
## $ INDEX       <int> 1, 2, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15, 16, 17, 19, 20, 2~
## $ TARGET_FLAG <int> 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1~
## $ TARGET_AMT  <dbl> 0.000, 0.000, 0.000, 0.000, 0.000, 2946.000, 0.000, 4021.0~
## $ KIDSDRIV    <int> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ AGE         <int> 60, 43, 35, 51, 50, 34, 54, 37, 34, 50, 53, 43, 55, 53, 45~
## $ HOMEKIDS    <int> 0, 0, 1, 0, 0, 1, 0, 2, 0, 0, 0, 0, 0, 0, 0, 3, 0, 3, 2, 1~
## $ YOJ         <int> 11, 11, 10, 14, NA, 12, NA, NA, 10, 7, 14, 5, 11, 11, 0, 1~
## $ INCOME      <chr> "$67,349", "$91,449", "$16,039", "", "$114,986", "$125,301~
## $ PARENT1     <chr> "No", "No", "No", "No", "No", "Yes", "No", "No", "No", "No~
## $ HOME_VAL    <chr> "$0", "$257,252", "$124,191", "$306,251", "$243,925", "$0"~
## $ MSTATUS     <chr> "z_No", "z_No", "Yes", "Yes", "Yes", "z_No", "Yes", "Yes",~
## $ SEX         <chr> "M", "M", "z_F", "M", "z_F", "z_F", "z_F", "M", "z_F", "M"~
## $ EDUCATION   <chr> "PhD", "z_High School", "z_High School", "<High School", "~
## $ JOB         <chr> "Professional", "z_Blue Collar", "Clerical", "z_Blue Colla~
## $ TRAVTIME    <int> 14, 22, 5, 32, 36, 46, 33, 44, 34, 48, 15, 36, 25, 64, 48,~
## $ CAR_USE     <chr> "Private", "Commercial", "Private", "Private", "Private", ~
## $ BLUEBOOK    <chr> "$14,230", "$14,940", "$4,010", "$15,440", "$18,000", "$17~
## $ TIF         <int> 11, 1, 4, 7, 1, 1, 1, 1, 1, 7, 1, 7, 7, 6, 1, 6, 6, 7, 4, ~
## $ CAR_TYPE    <chr> "Minivan", "Minivan", "z_SUV", "Minivan", "z_SUV", "Sports~
## $ RED_CAR     <chr> "yes", "yes", "no", "yes", "no", "no", "no", "yes", "no", ~
## $ OLDCLAIM    <chr> "$4,461", "$0", "$38,690", "$0", "$19,217", "$0", "$0", "$~
## $ CLM_FREQ    <int> 2, 0, 2, 0, 2, 0, 0, 1, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 2~
## $ REVOKED     <chr> "No", "No", "No", "No", "Yes", "No", "No", "Yes", "No", "N~
## $ MVR_PTS     <int> 3, 0, 3, 0, 3, 0, 0, 10, 0, 1, 0, 0, 3, 3, 3, 0, 0, 0, 0, ~
## $ CAR_AGE     <int> 18, 1, 10, 6, 17, 7, 1, 7, 1, 17, 11, 1, 9, 10, 5, 13, 16,~
## $ URBANICITY  <chr> "Highly Urban/ Urban", "Highly Urban/ Urban", "Highly Urba~
```

There are 8161 observation in the training dataset having 21 feature variables and 2 target variables.

```
##   INDEX TARGET_FLAG TARGET_AMT KIDSDRIV AGE HOMEKIDS YOJ   INCOME PARENT1
## 1     1           0          0        0  60        0  11  $67,349      No
## 2     2           0          0        0  43        0  11  $91,449      No
## 3     4           0          0        0  35        1  10  $16,039      No
## 4     5           0          0        0  51        0  14                No
## 5     6           0          0        0  50        0  NA $114,986      No
## 6     7           1       2946        0  34        1  12 $125,301     Yes
##   HOME_VAL MSTATUS SEX     EDUCATION           JOB TRAVTIME    CAR_USE BLUEBOOK
## 1       $0    z_No   M           PhD  Professional       14    Private  $14,230
## 2 $257,252    z_No   M z_High School z_Blue Collar       22 Commercial  $14,940
## 3 $124,191     Yes z_F z_High School      Clerical        5    Private   $4,010
## 4 $306,251     Yes   M  <High School z_Blue Collar       32    Private  $15,440
## 5 $243,925     Yes z_F           PhD        Doctor       36    Private  $18,000
## 6       $0    z_No z_F     Bachelors z_Blue Collar       46 Commercial  $17,430
##   TIF   CAR_TYPE RED_CAR OLDCLAIM CLM_FREQ REVOKED MVR_PTS CAR_AGE
## 1  11    Minivan     yes   $4,461        2      No       3      18
## 2   1    Minivan     yes       $0        0      No       0       1
## 3   4      z_SUV      no  $38,690        2      No       3      10
## 4   7    Minivan     yes       $0        0      No       0       6
## 5   1      z_SUV      no  $19,217        2     Yes       3      17
## 6   1 Sports Car      no       $0        0      No       0       7
##            URBANICITY
```

```
## 1 Highly Urban/ Urban
## 2 Highly Urban/ Urban
## 3 Highly Urban/ Urban
## 4 Highly Urban/ Urban
## 5 Highly Urban/ Urban
## 6 Highly Urban/ Urban
```

```
##      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   Length:8161       
##  1st Qu.:39.00   1st Qu.:0.0000   1st Qu.: 9.0   Class :character  
##  Median :45.00   Median :0.0000   Median :11.0   Mode  :character  
##  Mean   :44.79   Mean   :0.7212   Mean   :10.5                     
##  3rd Qu.:51.00   3rd Qu.:1.0000   3rd Qu.:13.0                     
##  Max.   :81.00   Max.   :5.0000   Max.   :23.0                     
##  NA's   :6                        NA's   :454                      
##    PARENT1            HOME_VAL           MSTATUS              SEX           
##  Length:8161        Length:8161        Length:8161        Length:8161       
##  Class :character   Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character  
##                                                                             
##                                                                             
##                                                                             
##                                                                             
##   EDUCATION             JOB               TRAVTIME        CAR_USE         
##  Length:8161        Length:8161        Min.   :  5.00   Length:8161       
##  Class :character   Class :character   1st Qu.: 22.00   Class :character  
##  Mode  :character   Mode  :character   Median : 33.00   Mode  :character  
##                                        Mean   : 33.49                     
##                                        3rd Qu.: 44.00                     
##                                        Max.   :142.00                     
##                                                                           
##    BLUEBOOK              TIF           CAR_TYPE           RED_CAR         
##  Length:8161        Min.   : 1.000   Length:8161        Length:8161       
##  Class :character   1st Qu.: 1.000   Class :character   Class :character  
##  Mode  :character   Median : 4.000   Mode  :character   Mode  :character  
##                     Mean   : 5.351                                        
##                     3rd Qu.: 7.000                                        
##                     Max.   :25.000                                        
##                                                                           
##    OLDCLAIM            CLM_FREQ        REVOKED             MVR_PTS      
##  Length:8161        Min.   :0.0000   Length:8161        Min.   : 0.000  
##  Class :character   1st Qu.:0.0000   Class :character   1st Qu.: 0.000  
##  Mode  :character   Median :0.0000   Mode  :character   Median : 1.000  
##                     Mean   :0.7986                      Mean   : 1.696  
##                     3rd Qu.:2.0000                      3rd Qu.: 3.000  
##                     Max.   :5.0000                      Max.   :13.000  
```

```
##
##     CAR_AGE        URBANICITY
##  Min.   :-3.000   Length:8161
##  1st Qu.: 1.000   Class :character
##  Median : 8.000   Mode  :character
##  Mean   : 8.328
##  3rd Qu.:12.000
##  Max.   :28.000
##  NA's   :510
```

There are several recurring issues with some columns: all columns containing money amounts have incomptaible punctuation and characters. Also, categorical variables neeed to be changed to factors and their factor names edited for intelligibility.

```
##   TARGET_FLAG       TARGET_AMT        KIDSDRIV           AGE
##  Min.   :0.0000   Min.   :     0   Min.   :0.0000   Min.   :16.00
##  1st Qu.:0.0000   1st Qu.:     0   1st Qu.:0.0000   1st Qu.:39.00
##  Median :0.0000   Median :     0   Median :0.0000   Median :45.00
##  Mean   :0.2638   Mean   :  1504   Mean   :0.1711   Mean   :44.79
##  3rd Qu.:1.0000   3rd Qu.:  1036   3rd Qu.:0.0000   3rd Qu.:51.00
##  Max.   :1.0000   Max.   :107586   Max.   :4.0000   Max.   :81.00
##                                                     NA's   :6
##     HOMEKIDS           YOJ           INCOME       PARENT1       HOME_VAL
##  Min.   :0.0000   Min.   : 0.0   Min.   :     0   No :7084   Min.   :     0
##  1st Qu.:0.0000   1st Qu.: 9.0   1st Qu.: 28097   Yes:1077   1st Qu.:     0
##  Median :0.0000   Median :11.0   Median : 54028              Median :161160
##  Mean   :0.7212   Mean   :10.5   Mean   : 61898              Mean   :154867
##  3rd Qu.:1.0000   3rd Qu.:13.0   3rd Qu.: 85986              3rd Qu.:238724
##  Max.   :5.0000   Max.   :23.0   Max.   :367030              Max.   :885282
##                   NA's   :454    NA's   :445                 NA's   :464
##  MSTATUS    SEX                    EDUCATION             JOB
##  No :3267   F:4375   Bachelors           :2242   Blue Collar :1825
##  Yes:4894   M:3786   High School         :2330   Clerical    :1271
##                      Less than High School:1203   Professional:1117
##                      Masters             :1658   Manager     : 988
##                      PhD                 : 728   Lawyer      : 835
##                                                  Student     : 712
##                                                  (Other)     :1413
##     TRAVTIME           CAR_USE        BLUEBOOK         TIF
##  Min.   :  5.00   Commercial:3029   Min.   : 1500   Min.   : 1.000
##  1st Qu.: 22.00   Private   :5132   1st Qu.: 9280   1st Qu.: 1.000
##  Median : 33.00                     Median :14440   Median : 4.000
##  Mean   : 33.49                     Mean   :15710   Mean   : 5.351
##  3rd Qu.: 44.00                     3rd Qu.:20850   3rd Qu.: 7.000
##  Max.   :142.00                     Max.   :69740   Max.   :25.000
##
##         CAR_TYPE    RED_CAR       OLDCLAIM        CLM_FREQ      REVOKED
##  Minivan     :2145   no :5783   Min.   :    0   Min.   :0.0000   No :7161
##  Panel Truck: 676   yes:2378   1st Qu.:    0   1st Qu.:0.0000   Yes:1000
##  Pickup      :1389              Median :    0   Median :0.0000
##  Sports Car : 907              Mean   : 4037   Mean   :0.7986
##  SUV         :2294              3rd Qu.: 4636   3rd Qu.:2.0000
##  Van         : 750              Max.   :57037   Max.   :5.0000
##
##     MVR_PTS          CAR_AGE                     URBANICITY
```

```
##  Min.   : 0.000   Min.   :-3.000   Highly Rural/ Rural:1669
##  1st Qu.: 0.000   1st Qu.: 1.000   Highly Urban/ Urban:6492
##  Median : 1.000   Median : 8.000
##  Mean   : 1.696   Mean   : 8.328
##  3rd Qu.: 3.000   3rd Qu.:12.000
##  Max.   :13.000   Max.   :28.000
##                   NA's   :510
```

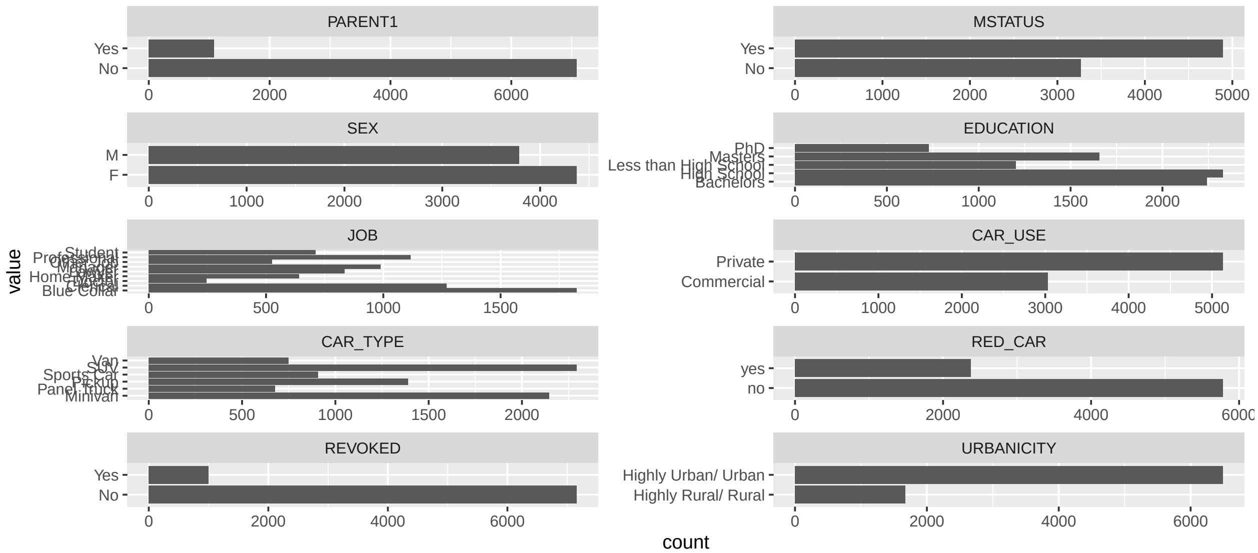
The fixed dataframe now only includes columns that are numeric or factors. Car age appears to have some values less than 1, including a negative values. These will be changed to the mode of 1.

Categorical variables

```
## [1] "PARENT1"
## [1] "No"  "Yes"
## [1] "MSTATUS"
## [1] "No"  "Yes"
## [1] "SEX"
## [1] "F" "M"
## [1] "EDUCATION"
## [1] "Bachelors"             "High School"           "Less than High School"
## [4] "Masters"               "PhD"
## [1] "JOB"
## [1] "Blue Collar"  "Clerical"     "Doctor"       "Home Maker"   "Lawyer"
## [6] "Manager"      "Other Job"    "Professional" "Student"
## [1] "CAR_USE"
## [1] "Commercial" "Private"
## [1] "CAR_TYPE"
## [1] "Minivan"     "Panel Truck" "Pickup"      "Sports Car"  "SUV"
## [6] "Van"
## [1] "RED_CAR"
## [1] "no"  "yes"
## [1] "REVOKED"
## [1] "No"  "Yes"
## [1] "URBANICITY"
## [1] "Highly Rural/ Rural" "Highly Urban/ Urban"
```

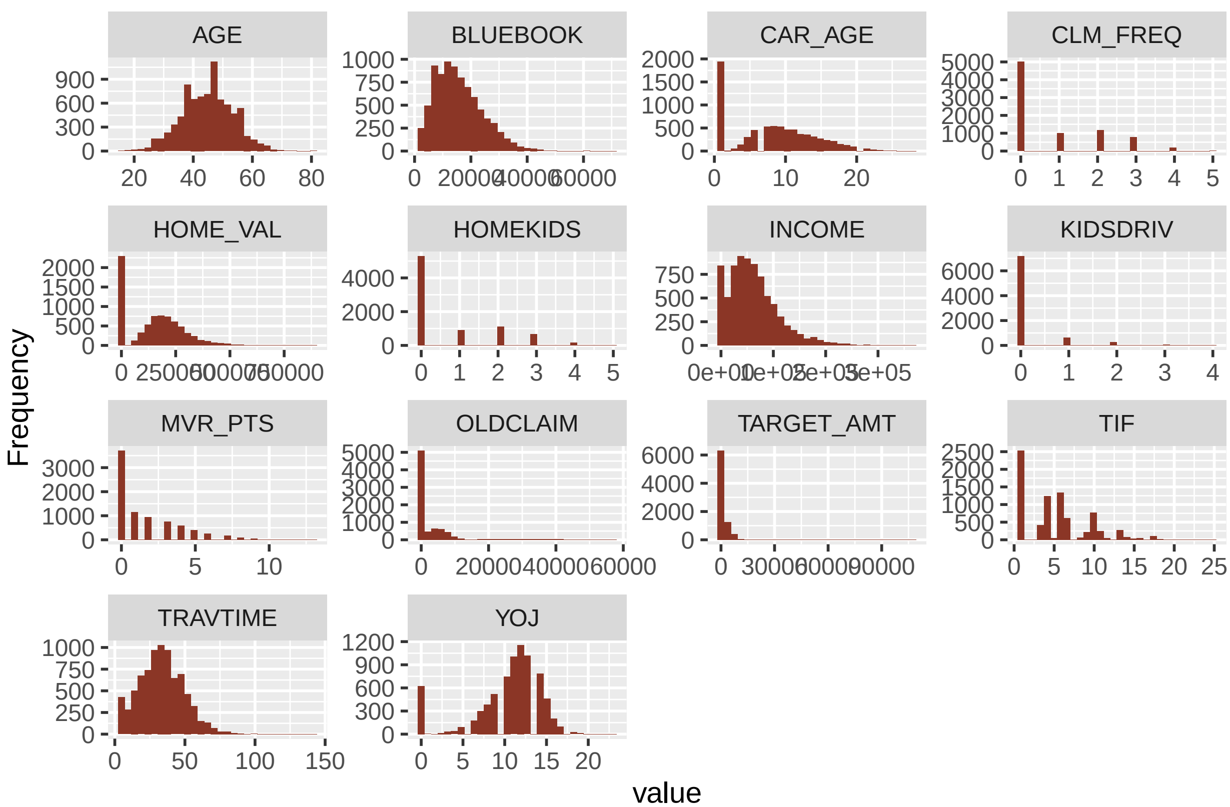
Looking at categorical variables, most of the columns are binary.

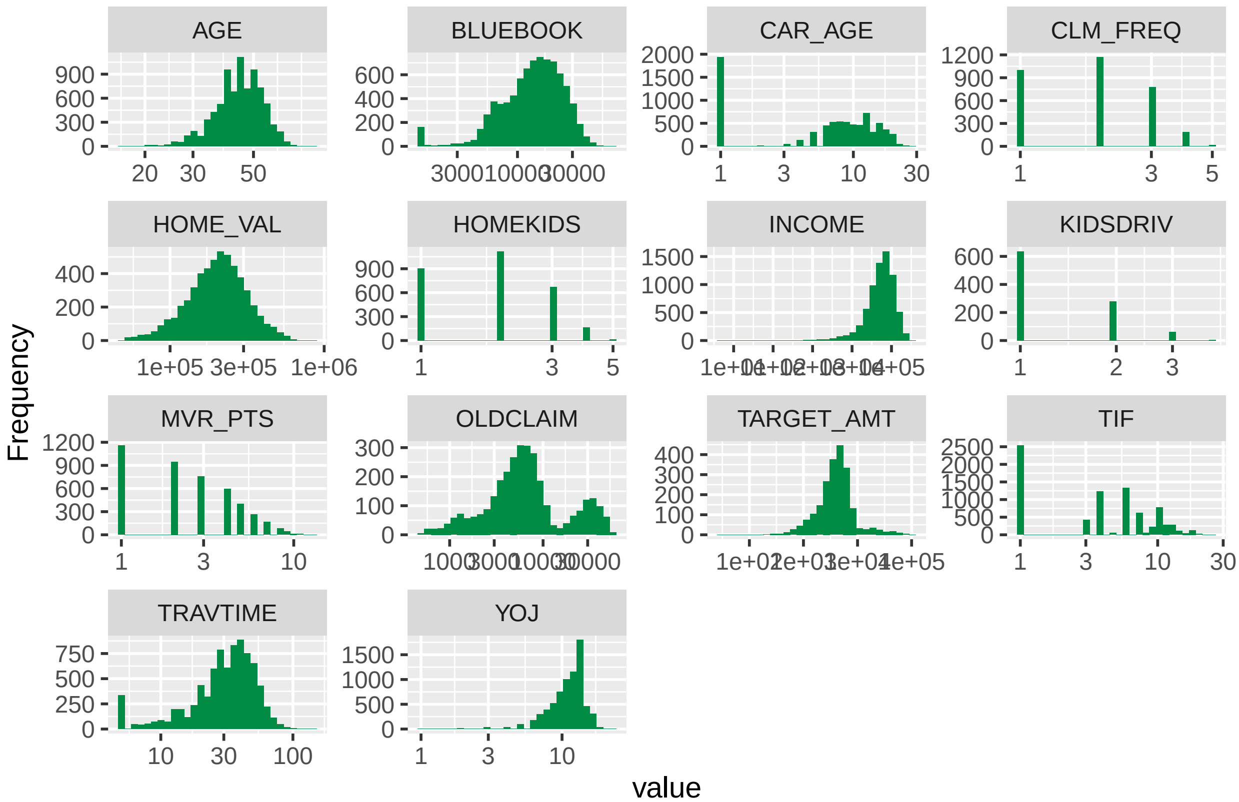
Below graphs shows the distribution of all categorical predictors.



Numeric Variables

Below 2 graphs shows the distribution of numeric variables. The red graphs are on normal scale and the green ones are on log10 scale. Many numeric variables feature the value of zero as a mode.

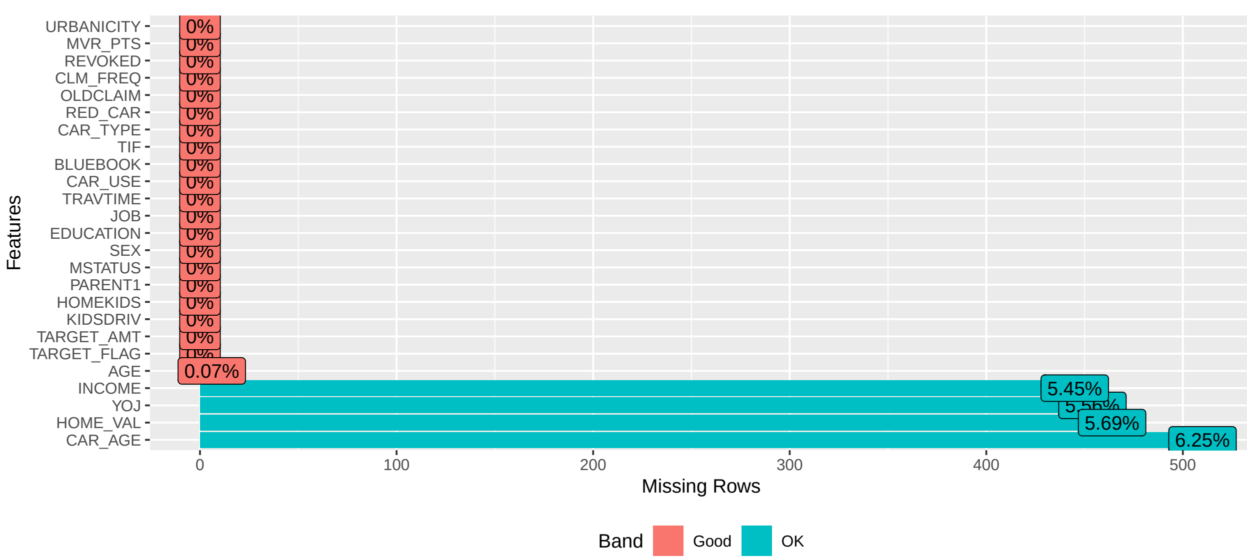




Missing Values

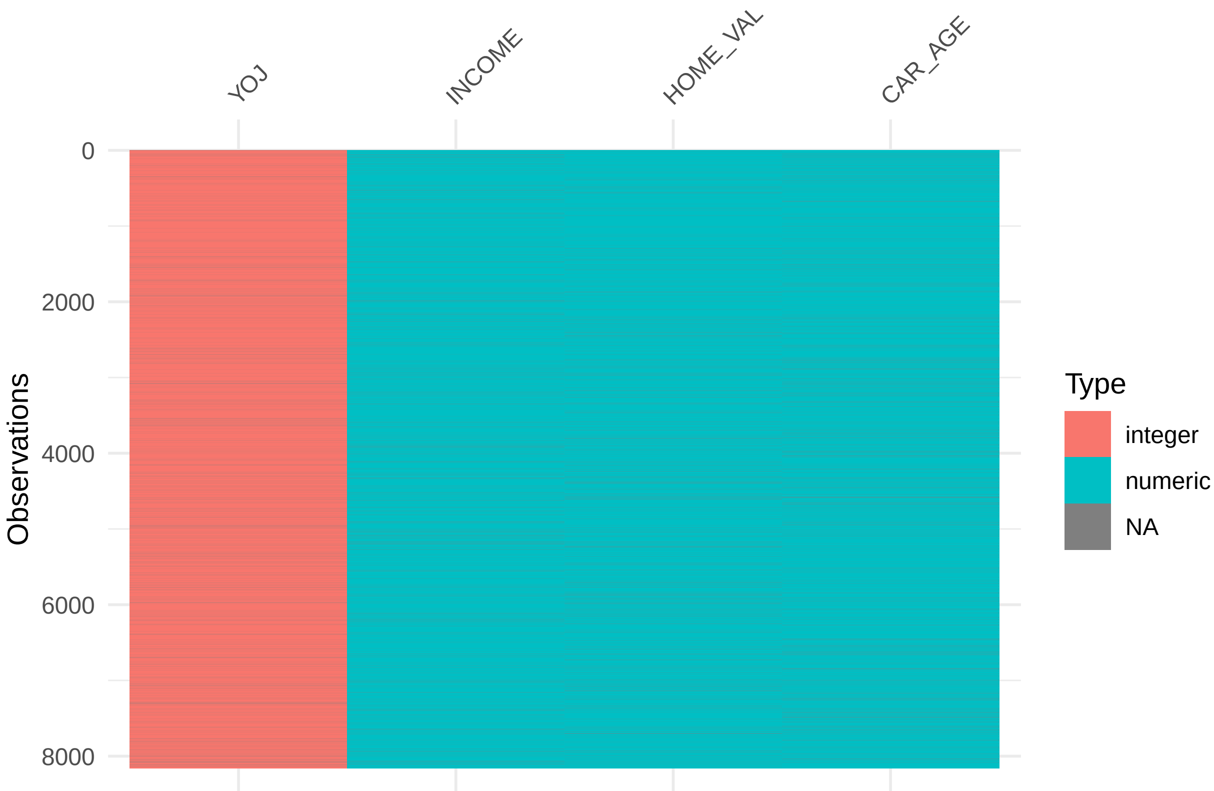
Here are columns having missing values coded as NA:

```
##    AGE YOJ INCOME HOME_VAL CAR_AGE
## 1    6 454    445      464     510
```




```
## TARGET_FLAG  TARGET_AMT    KIDSDRIV         AGE    HOMEKIDS         YOJ
##       0.000       0.000       0.000       0.001       0.000       0.056
##      INCOME     PARENT1    HOME_VAL     MSTATUS         SEX   EDUCATION
##       0.055       0.000       0.057       0.000       0.000       0.000
```

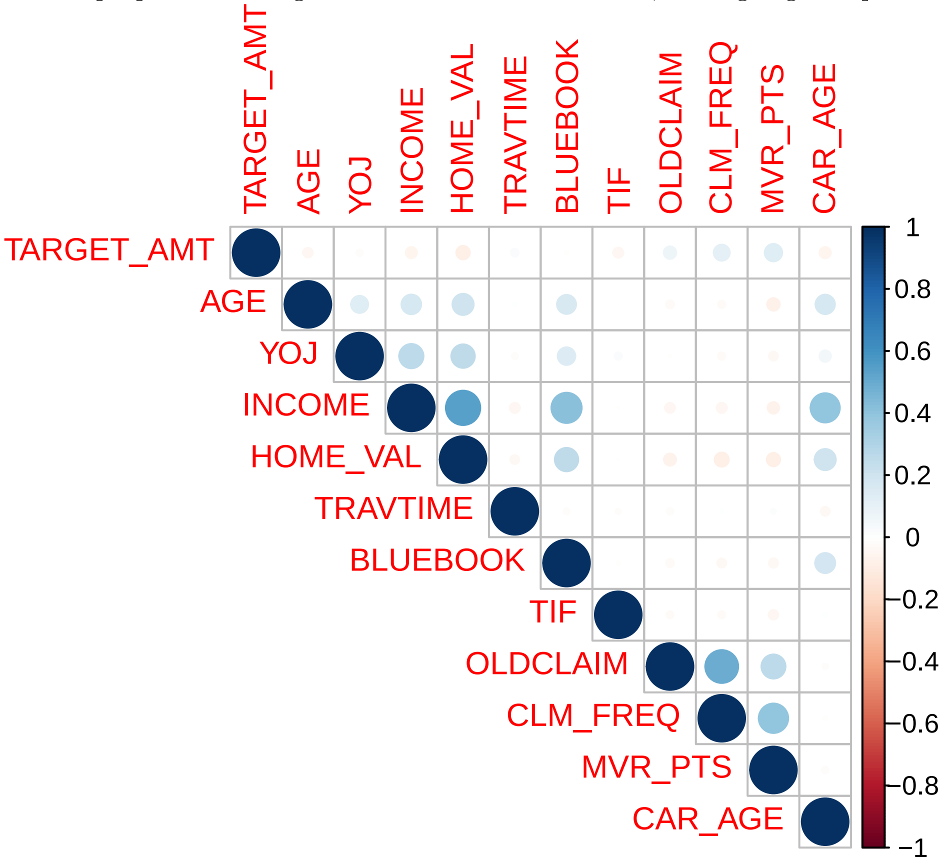
```
##          JOB    TRAVTIME     CAR_USE    BLUEBOOK         TIF    CAR_TYPE
##        0.000       0.000       0.000       0.000       0.000       0.000
##      RED_CAR    OLDCLAIM    CLM_FREQ     REVOKED     MVR_PTS     CAR_AGE
##        0.000       0.000       0.000       0.000       0.000       0.062
##   URBANICITY
##        0.000
```



Four variables have missing values, however there doesn't appear to be a pattern and it's safe to assume they're missing at random.

Correlation

For the purposes of seeing correlation between variables, we're going to replace NA values with the median.



It's clear there are some positive correlations between the following variables:
* **Income** & **Home value**: 0.54
* **Income** & **Bluebook**: 0.42
* **Income** & **Car age**: 0.39
* **Claim Frequency** & **Old claims**: 0.50
* **Claim Frequence** & **MVR_PTS**:0.39

Data Preparation

Removing TARGET_FLAG

Our multiple linear regression model will be predicting the amount of money someone receives if they crash, so we will be removing the variable *TARGET_FLAG*

Handling Missing Data - Multiple Linear Regression

For the multiple linear regression, we're going to assume that the NULL values will take the median value for the variable.

Transforming Variables - Multiple Linear Regression

There some variables that are not normally distributed so we're going to try using a log transformation later to see if that creates a better model. For a few variables with values, 0, we added 1 to avoid negative infinity when taking the log of those variables. This will not alter our modeling results significantly.

Zeroes in Home Value

It seems from the histogram above, that the mode of the variable HOME_VAL is 0. Given that, the distribution seems normal if we remove 0s and that the difference between 0 and the number that appears next on the axis is significant, we are assuming that 0 indicates missing values for HOME_VAL. Therefore, we will convert 0s to NAs in HOME_VAL prior to imputing missing values for Binary Logistic Regression Model 3 below.

Addressing Zeroes using Binning

The histograms for several variables indicate that there many with an overrepresentation of 'zero' values. Some of the worst offenders include CAR_AGE, HOME_VAL, HOMEKIDS, KIDSDRIV, OLDCLAIM, TIF, and YOJ. INCOME also has many 'zero' or very low values, and also similar to CAR_AGE and HOME_VAL because, omitting zero, the rest of the distributions appear to be skewed, approximately normal distributions. To avoid problems with interpretation, the 4th model will consider these continuous variables as categorical variables defined as a number range.

```
##   TARGET_FLAG       TARGET_AMT          AGE            INCOME       PARENT1
##  Min.   :0.0000   Min.   :     0   Min.   :16.00   Min.   :     0   No :7084
##  1st Qu.:0.0000   1st Qu.:     0   1st Qu.:39.00   1st Qu.: 28097   Yes:1077
##  Median :0.0000   Median :     0   Median :45.00   Median : 54028
##  Mean   :0.2638   Mean   :  1504   Mean   :44.79   Mean   : 61898
##  3rd Qu.:1.0000   3rd Qu.:  1036   3rd Qu.:51.00   3rd Qu.: 85986
##  Max.   :1.0000   Max.   :107586   Max.   :81.00   Max.   :367030
##                                    NA's   :6       NA's   :445
##  MSTATUS    SEX                     EDUCATION              JOB
##  No :3267   F:4375   Bachelors           :2242   Blue Collar :1825
##  Yes:4894   M:3786   High School         :2330   Clerical    :1271
##                      Less than High School:1203   Professional:1117
##                      Masters             :1658   Manager     : 988
##                      PhD                 : 728   Lawyer      : 835
##                                                  Student     : 712
##                                                  (Other)     :1413
##     TRAVTIME            CAR_USE        BLUEBOOK          CAR_TYPE
##  Min.   :  5.00   Commercial:3029   Min.   : 1500   Minivan    :2145
##  1st Qu.: 22.00   Private   :5132   1st Qu.: 9280   Panel Truck: 676
##  Median : 33.00                     Median :14440   Pickup     :1389
##  Mean   : 33.49                     Mean   :15710   Sports Car : 907
##  3rd Qu.: 44.00                     3rd Qu.:20850   SUV        :2294
##  Max.   :142.00                     Max.   :69740   Van        : 750
##
##  RED_CAR       CLM_FREQ       REVOKED       MVR_PTS
##  no :5783   Min.   :0.0000   No :7161   Min.   : 0.000
##  yes:2378   1st Qu.:0.0000   Yes:1000   1st Qu.: 0.000
##             Median :0.0000              Median : 1.000
##             Mean   :0.7986              Mean   : 1.696
##             3rd Qu.:2.0000              3rd Qu.: 3.000
##             Max.   :5.0000              Max.   :13.000
##
##               URBANICITY     CAR_AGE_BIN        HOME_VAL_BIN   HAS_HOME_KIDS
##  Highly Rural/ Rural:1669   New     :1938   Zero       :2294   Has kids:2872
##  Highly Urban/ Urban:6492   Like New:  66   $0-$50k    :   0   No kids :5289
##                             Average :3775   $50k-$150k :1274
##                             Old     :1872   $150k-$250k:2445
##                             NA's    : 510   Over $250k :1684
```

```
##                                             NA's       : 464
##
##            HAS_KIDSDRIV    OLDCLAIM_BIN            TIF_BIN
##  Has kids driving: 981   Zero    :5009   Zero            :   0
##  No kids driving :7180   $0-$3k  : 584   Less than 1 year:2533
##                          $3k-$6k : 970   1-4 years       :1672
##                          $6k-$9k : 720   4-7 years       :2013
##                          Over $9k: 878   Over 7 years    :1943
##
##
##                 YOJ_BIN
##  Zero               : 625
##  Less than 10 years :2313
##  Between 10-15 years:4425
##  Over 15 years      : 344
##  NA's               : 454
##
##
##   TARGET_FLAG TARGET_AMT AGE INCOME PARENT1 MSTATUS SEX             EDUCATION
## 1           0          0  60  67349      No      No   M                   PhD
## 2           0          0  43  91449      No      No   M           High School
## 3           0          0  35  16039      No     Yes   F           High School
## 4           0          0  51     NA      No     Yes   M Less than High School
## 5           0          0  50 114986      No     Yes   F                   PhD
## 6           1       2946  34 125301     Yes      No   F             Bachelors
##            JOB TRAVTIME    CAR_USE BLUEBOOK   CAR_TYPE RED_CAR CLM_FREQ REVOKED
## 1 Professional       14    Private    14230    Minivan     yes        2      No
## 2  Blue Collar       22 Commercial    14940    Minivan     yes        0      No
## 3     Clerical        5    Private     4010        SUV      no        2      No
## 4  Blue Collar       32    Private    15440    Minivan     yes        0      No
## 5       Doctor       36    Private    18000        SUV      no        2     Yes
## 6  Blue Collar       46 Commercial    17430 Sports Car      no        0      No
##   MVR_PTS          URBANICITY CAR_AGE_BIN HOME_VAL_BIN HAS_HOME_KIDS
## 1       3 Highly Urban/ Urban         Old         Zero       No kids
## 2       0 Highly Urban/ Urban         New   Over $250k       No kids
## 3       3 Highly Urban/ Urban     Average   $50k-$150k      Has kids
## 4       0 Highly Urban/ Urban     Average   Over $250k       No kids
## 5       3 Highly Urban/ Urban         Old  $150k-$250k       No kids
## 6       0 Highly Urban/ Urban     Average         Zero      Has kids
##      HAS_KIDSDRIV OLDCLAIM_BIN          TIF_BIN             YOJ_BIN
## 1 No kids driving      $3k-$6k     Over 7 years Between 10-15 years
## 2 No kids driving         Zero Less than 1 year Between 10-15 years
## 3 No kids driving     Over $9k        1-4 years  Less than 10 years
## 4 No kids driving         Zero        4-7 years Between 10-15 years
## 5 No kids driving     Over $9k Less than 1 year                <NA>
## 6 No kids driving         Zero Less than 1 year Between 10-15 years
```

Build Models

Model1

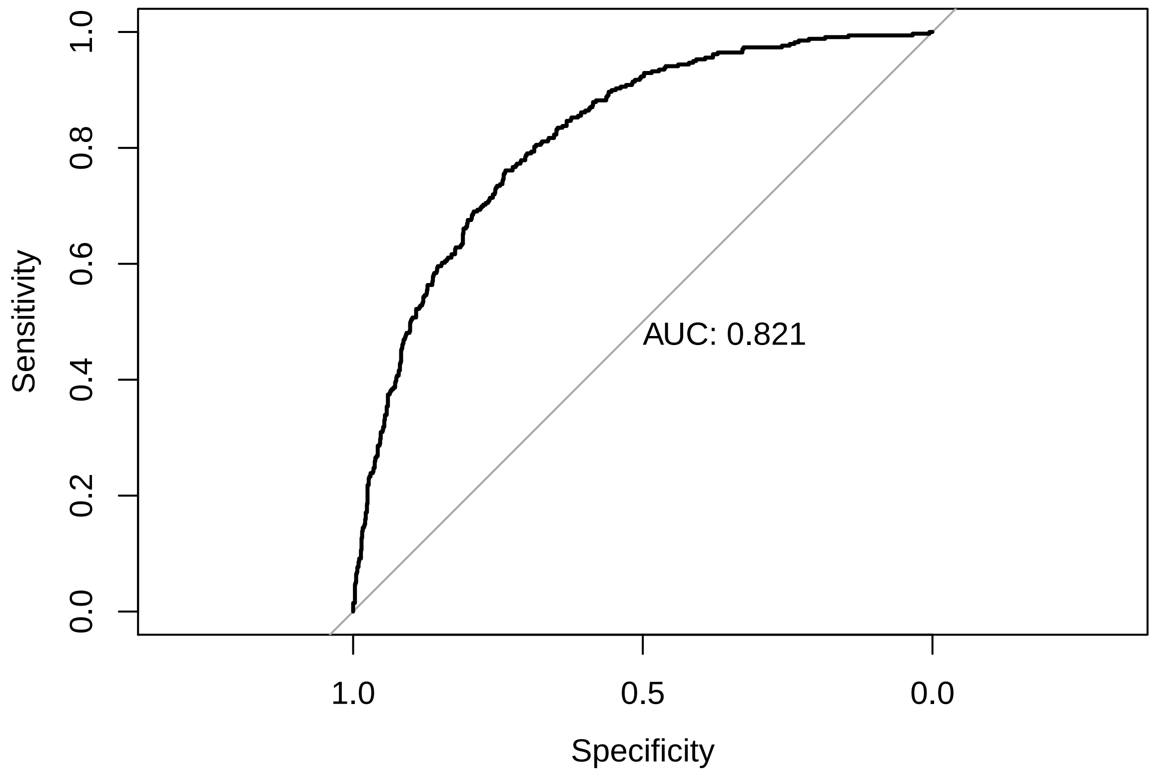
The first model to consider includes all given variables and does not impute any values.

```
##
```

```
## Call:
## glm(formula = TARGET_FLAG ~ . - TARGET_AMT, family = "binomial",
##     data = insurance_fix)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5843  -0.7124  -0.3998   0.6195   3.1633
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -2.881e+00  3.199e-01  -9.005  < 2e-16 ***
## KIDSDRIV                        3.385e-01  6.908e-02   4.900 9.57e-07 ***
## AGE                            -3.665e-03  4.531e-03  -0.809 0.418503
## HOMEKIDS                        3.349e-02  4.176e-02   0.802 0.422588
## YOJ                            -1.071e-02  9.589e-03  -1.117 0.263837
## INCOME                         -2.988e-06  1.260e-06  -2.371 0.017738 *
## PARENT1Yes                      4.337e-01  1.225e-01   3.541 0.000398 ***
## HOME_VAL                       -1.301e-06  3.899e-07  -3.337 0.000848 ***
## MSTATUSYes                     -4.389e-01  9.666e-02  -4.541 5.61e-06 ***
## SEXM                            1.914e-01  1.241e-01   1.543 0.122880
## EDUCATIONHigh School            3.716e-01  1.020e-01   3.645 0.000268 ***
## EDUCATIONLess than High School  3.724e-01  1.306e-01   2.852 0.004342 **
## EDUCATIONMasters                2.887e-02  1.607e-01   0.180 0.857462
## EDUCATIONPhD                    2.617e-01  2.054e-01   1.274 0.202597
## JOBClerical                     2.052e-01  1.193e-01   1.720 0.085428 .
## JOBDoctor                      -5.011e-01  3.136e-01  -1.598 0.110084
## JOBHome Maker                  -8.529e-02  1.750e-01  -0.487 0.625972
## JOBLawyer                      -1.923e-02  2.126e-01  -0.090 0.927939
## JOBManager                     -8.826e-01  1.595e-01  -5.534 3.13e-08 ***
## JOBOther Job                   -3.071e-01  2.117e-01  -1.450 0.146938
## JOBProfessional                -1.066e-01  1.360e-01  -0.784 0.433062
## JOBStudent                     -1.370e-01  1.497e-01  -0.915 0.359966
## TRAVTIME                        1.562e-02  2.118e-03   7.374 1.66e-13 ***
## CAR_USEPrivate                 -8.256e-01  1.040e-01  -7.935 2.10e-15 ***
## BLUEBOOK                       -2.101e-05  5.885e-06  -3.570 0.000357 ***
## TIF                            -5.318e-02  8.241e-03  -6.453 1.10e-10 ***
## CAR_TYPEPanel Truck             6.097e-01  1.807e-01   3.374 0.000740 ***
## CAR_TYPEPickup                  5.246e-01  1.136e-01   4.619 3.85e-06 ***
## CAR_TYPESports Car              1.128e+00  1.450e-01   7.784 7.05e-15 ***
## CAR_TYPESUV                     8.518e-01  1.241e-01   6.866 6.59e-12 ***
## CAR_TYPEVan                     6.335e-01  1.421e-01   4.460 8.21e-06 ***
## RED_CARyes                     -1.227e-01  9.685e-02  -1.267 0.205139
## OLDCLAIM                       -1.180e-05  4.375e-06  -2.698 0.006977 **
## CLM_FREQ                        1.953e-01  3.183e-02   6.136 8.46e-10 ***
## REVOKEDYes                      8.644e-01  1.035e-01   8.354  < 2e-16 ***
## MVR_PTS                         1.143e-01  1.528e-02   7.485 7.16e-14 ***
## CAR_AGE                        -7.075e-03  8.448e-03  -0.837 0.402334
## URBANICITYHighly Urban/ Urban   2.313e+00  1.241e-01  18.640  < 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: 5764.7  on 6410  degrees of freedom
##   (1713 observations deleted due to missingness)
## AIC: 5840.7
##
## Number of Fisher Scoring iterations: 5
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 862 188
##          1  77 149
##
##                Accuracy : 0.7923
##                  95% CI : (0.769, 0.8143)
##     No Information Rate : 0.7359
##     P-Value [Acc > NIR] : 1.650e-06
##
##                   Kappa : 0.4026
##
##  Mcnemar's Test P-Value : 1.406e-11
##
##             Sensitivity : 0.9180
##             Specificity : 0.4421
##          Pos Pred Value : 0.8210
##          Neg Pred Value : 0.6593
##              Prevalence : 0.7359
##          Detection Rate : 0.6755
##    Detection Prevalence : 0.8229
##       Balanced Accuracy : 0.6801
##
##        'Positive' Class : 0
##
```



Model2

The second model imputes values using the 'mice' library using classification and regression trees. We will use glm.mids() that applies glm() to a multiply imputed data set.

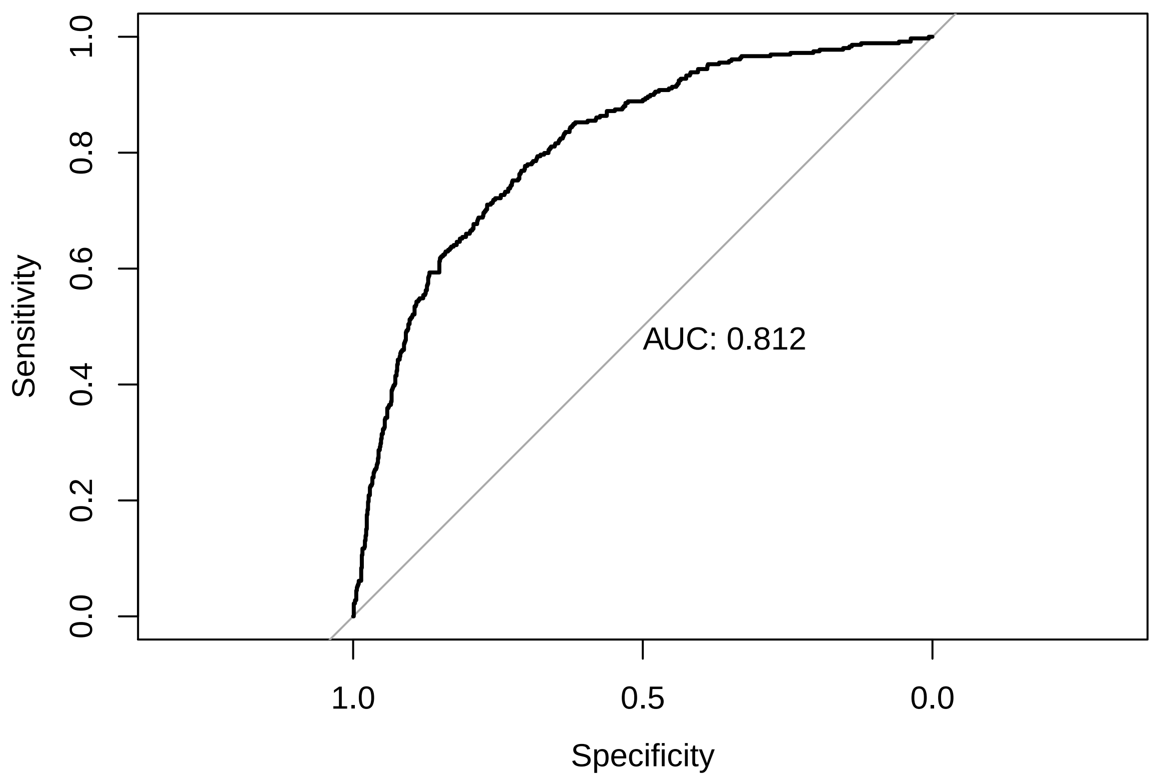
```
##
##  iter imp variable
##   1   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
##   2   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
##   3   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
##   4   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
##   5   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
```

```
## call :
## glm.mids(formula = TARGET_FLAG ~ . - TARGET_AMT, family = "binomial",
##     data = insurance_impute)
##
## call1 :
## mice(data = insurance_fix, m = 1, method = "cart")
##
## nmis :
## TARGET_FLAG  TARGET_AMT    KIDSDRIV         AGE    HOMEKIDS         YOJ
##           0           0           0           6           0         454
##      INCOME     PARENT1    HOME_VAL     MSTATUS         SEX   EDUCATION
##         445           0         464           0           0           0
##         JOB    TRAVTIME     CAR_USE    BLUEBOOK         TIF    CAR_TYPE
##           0           0           0           0           0           0
##     RED_CAR    OLDCLAIM    CLM_FREQ     REVOKED     MVR_PTS     CAR_AGE
##           0           0           0           0           0         510
##  URBANICITY
```

```
##           0
##
## analyses :
## [[1]]
##
## Call:  glm(formula = formula, family = family, data = complete(data,
##     i))
##
## Coefficients:
##                   (Intercept)                       KIDSDRIV
##                    -2.896e+00                      3.840e-01
##                           AGE                       HOMEKIDS
##                    -6.800e-04                      5.566e-02
##                           YOJ                         INCOME
##                    -1.784e-02                     -3.413e-06
##                    PARENT1Yes                       HOME_VAL
##                     3.802e-01                     -1.293e-06
##                    MSTATUSYes                           SEXM
##                    -4.818e-01                      8.755e-02
##          EDUCATIONHigh School  EDUCATIONLess than High School
##                     3.765e-01                      3.506e-01
##              EDUCATIONMasters                   EDUCATIONPhD
##                     1.187e-01                      2.530e-01
##                   JOBClerical                      JOBDoctor
##                     9.534e-02                     -7.712e-01
##                JOBHome Maker                      JOBLawyer
##                    -1.305e-01                     -2.040e-01
##                    JOBManager                   JOBOther Job
##                    -8.666e-01                     -3.031e-01
##               JOBProfessional                     JOBStudent
##                    -1.459e-01                     -1.525e-01
##                      TRAVTIME                  CAR_USEPrivate
##                     1.462e-02                     -7.552e-01
##                      BLUEBOOK                            TIF
##                    -2.042e-05                     -5.558e-02
##          CAR_TYPEPanel Truck                 CAR_TYPEPickup
##                     5.559e-01                      5.547e-01
##           CAR_TYPESports Car                    CAR_TYPESUV
##                     1.023e+00                      7.681e-01
##                   CAR_TYPEVan                      RED_CARyes
##                     6.174e-01                     -1.227e-02
##                      OLDCLAIM                       CLM_FREQ
##                    -1.378e-05                      1.965e-01
##                    REVOKEDYes                        MVR_PTS
##                     8.870e-01                      1.133e-01
##                       CAR_AGE   URBANICITYHighly Urban/ Urban
##                    -5.686e-03                      2.391e+00
##
## Degrees of Freedom: 8160 Total (i.e. Null);  8123 Residual
## Null Deviance:       9418
## Residual Deviance: 7292  AIC: 7368
```

```
## Confusion Matrix and Statistics
##
```

```
##           Reference
## Prediction   0   1
##          0 878 190
##          1  70 136
##
##                Accuracy : 0.7959
##                  95% CI : (0.7727, 0.8177)
##     No Information Rate : 0.7441
##     P-Value [Acc > NIR] : 8.412e-06
##
##                   Kappa : 0.3905
##
##  Mcnemar's Test P-Value : 1.582e-13
##
##             Sensitivity : 0.9262
##             Specificity : 0.4172
##          Pos Pred Value : 0.8221
##          Neg Pred Value : 0.6602
##              Prevalence : 0.7441
##          Detection Rate : 0.6892
##    Detection Prevalence : 0.8383
##       Balanced Accuracy : 0.6717
##
##        'Positive' Class : 0
##
```



Model 3

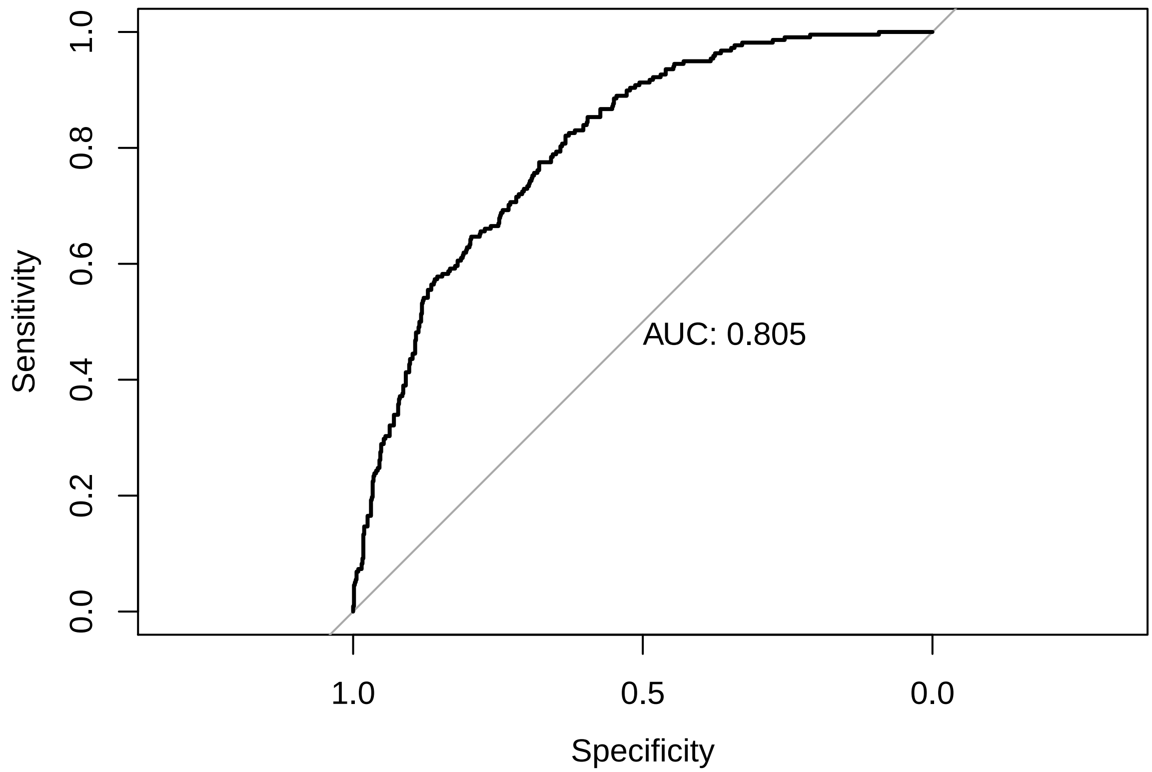
Now we will replicate the model above to see if our assumption about treating 0s in HOME_VAL as missing data, yields a better model fit.

```
##
##  iter imp variable
##   1   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
##   2   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
##   3   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
##   4   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
##   5   1  AGE  YOJ  INCOME  HOME_VAL  CAR_AGE
```

```
## call :
## glm.mids(formula = TARGET_FLAG ~ . - TARGET_AMT, family = "binomial",
##     data = insurance_impute2)
##
## call1 :
## mice(data = insurance_fix2, m = 1, method = "cart")
##
## nmis :
## TARGET_FLAG  TARGET_AMT    KIDSDRIV         AGE    HOMEKIDS         YOJ
##           0           0           0           6           0         454
##      INCOME     PARENT1    HOME_VAL     MSTATUS         SEX   EDUCATION
##         445           0        2758           0           0           0
##         JOB    TRAVTIME     CAR_USE    BLUEBOOK         TIF    CAR_TYPE
##           0           0           0           0           0           0
##     RED_CAR    OLDCLAIM     CLM_FREQ     REVOKED     MVR_PTS     CAR_AGE
##           0           0           0           0           0         510
##  URBANICITY
##           0
##
## analyses :
## [[1]]
##
## Call:  glm(formula = formula, family = family, data = complete(data,
##     i))
##
## Coefficients:
##                  (Intercept)                      KIDSDRIV
##                   -2.920e+00                     3.863e-01
##                          AGE                      HOMEKIDS
##                   -2.083e-03                     5.737e-02
##                          YOJ                        INCOME
##                   -1.598e-02                    -5.084e-06
##                   PARENT1Yes                      HOME_VAL
##                    3.585e-01                    -4.278e-08
##                   MSTATUSYes                          SEXM
##                   -6.449e-01                     7.930e-02
##          EDUCATIONHigh School  EDUCATIONLess than High School
##                    4.095e-01                     3.924e-01
##              EDUCATIONMasters                  EDUCATIONPhD
##                    9.530e-02                     2.425e-01
##                  JOBClerical                     JOBDoctor
##                    9.797e-02                    -7.434e-01
```

```
##                 JOBHome Maker                          JOBLawyer 
##                    -1.180e-01                         -2.033e-01 
##                    JOBManager                       JOBOther Job 
##                    -8.532e-01                         -2.962e-01 
##               JOBProfessional                         JOBStudent 
##                    -1.489e-01                         -5.974e-02 
##                      TRAVTIME                     CAR_USEPrivate 
##                     1.462e-02                         -7.546e-01 
##                      BLUEBOOK                                TIF 
##                    -1.992e-05                         -5.572e-02 
##           CAR_TYPEPanel Truck                    CAR_TYPEPickup 
##                     5.418e-01                          5.527e-01 
##           CAR_TYPESports Car                        CAR_TYPESUV 
##                     1.028e+00                          7.653e-01 
##                   CAR_TYPEVan                         RED_CARyes 
##                     6.128e-01                         -4.897e-03 
##                      OLDCLAIM                           CLM_FREQ 
##                    -1.395e-05                          1.989e-01 
##                    REVOKEDYes                            MVR_PTS 
##                     8.933e-01                          1.138e-01 
##                       CAR_AGE  URBANICITYHighly Urban/ Urban 
##                     3.030e-04                          2.396e+00 
## 
## Degrees of Freedom: 8160 Total (i.e. Null);  8123 Residual
## Null Deviance:       9418 
## Residual Deviance: 7307  AIC: 7383
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction   0   1
##          0 666 116
##          1  53  58
##                                           
##                Accuracy : 0.8108          
##                  95% CI : (0.7835, 0.8359)
##     No Information Rate : 0.8052          
##     P-Value [Acc > NIR] : 0.3547          
##                                           
##                   Kappa : 0.3009          
##                                           
##  Mcnemar's Test P-Value : 1.849e-06       
##                                           
##             Sensitivity : 0.9263          
##             Specificity : 0.3333          
##          Pos Pred Value : 0.8517          
##          Neg Pred Value : 0.5225          
##              Prevalence : 0.8052          
##          Detection Rate : 0.7458          
##    Detection Prevalence : 0.8757          
##       Balanced Accuracy : 0.6298          
##                                           
##        'Positive' Class : 0               
## 
```



Model 4

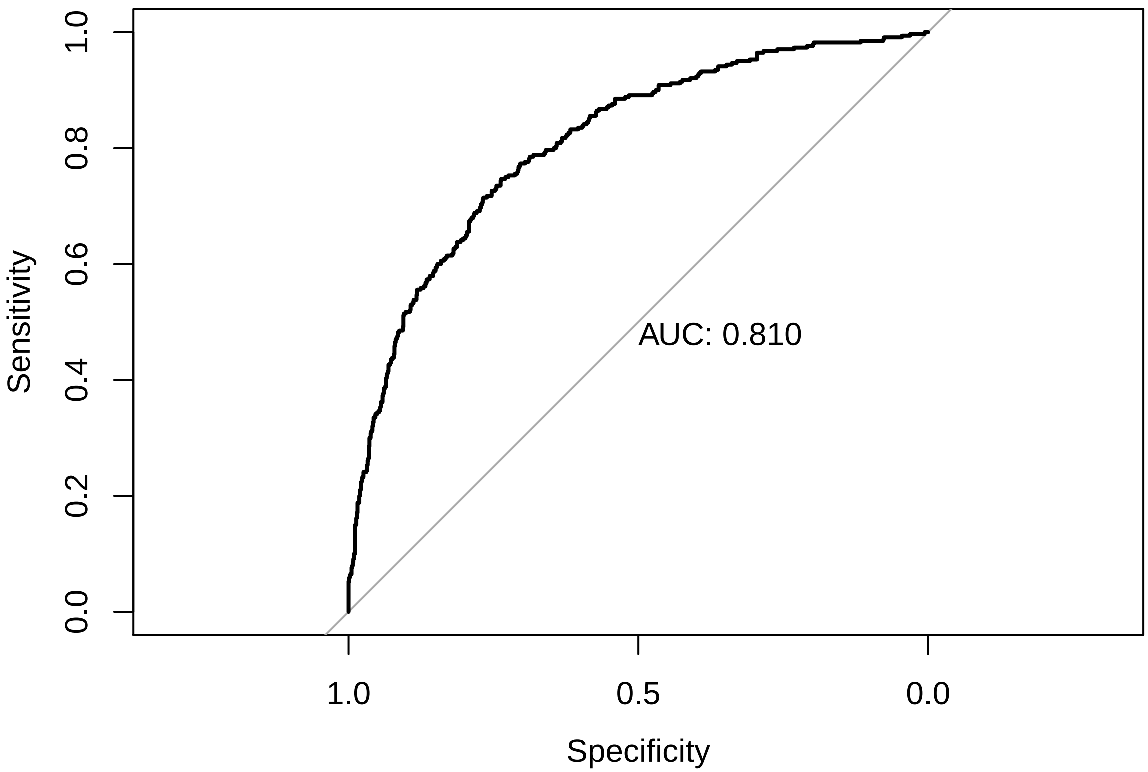
```
##
## Call:
## glm(formula = TARGET_FLAG ~ . - TARGET_AMT, family = "binomial",
##     data = insurance_bins)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4626  -0.7053  -0.3955   0.6199   3.1398
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.797e+00  3.584e-01  -5.013 5.36e-07 ***
## AGE                          -2.185e-03  4.754e-03  -0.459 0.645876
## INCOME                       -2.814e-06  1.344e-06  -2.094 0.036240 *
## PARENT1Yes                    2.826e-01  1.374e-01   2.057 0.039716 *
## MSTATUSYes                   -4.613e-01  1.046e-01  -4.408 1.04e-05 ***
## SEXM                          1.923e-01  1.249e-01   1.540 0.123660
## EDUCATIONHigh School          3.623e-01  1.022e-01   3.545 0.000393 ***
## EDUCATIONLess than High School  3.819e-01  1.300e-01   2.937 0.003312 **
## EDUCATIONMasters             -5.378e-04  1.664e-01  -0.003 0.997421
## EDUCATIONPhD                  2.007e-01  2.092e-01   0.959 0.337374
## JOBClerical                   1.937e-01  1.213e-01   1.597 0.110252
## JOBDoctor                    -4.930e-01  3.153e-01  -1.564 0.117906
## JOBHome Maker                -2.461e-01  1.915e-01  -1.285 0.198816
## JOBLawyer                    -6.033e-03  2.145e-01  -0.028 0.977560
## JOBManager                   -8.712e-01  1.609e-01  -5.413 6.18e-08 ***
## JOBOther Job                 -3.073e-01  2.131e-01  -1.442 0.149177
```

```
## JOBProfessional               -9.770e-02  1.369e-01  -0.714 0.475349
## JOBStudent                    -4.025e-01  1.690e-01  -2.381 0.017254 *
## TRAVTIME                       1.617e-02  2.135e-03   7.572 3.66e-14 ***
## CAR_USEPrivate                -8.233e-01  1.048e-01  -7.855 4.00e-15 ***
## BLUEBOOK                      -2.099e-05  5.904e-06  -3.555 0.000378 ***
## CAR_TYPEPanel Truck            6.416e-01  1.818e-01   3.530 0.000415 ***
## CAR_TYPEPickup                 5.401e-01  1.141e-01   4.734 2.21e-06 ***
## CAR_TYPESports Car             1.113e+00  1.460e-01   7.625 2.43e-14 ***
## CAR_TYPESUV                    8.572e-01  1.249e-01   6.864 6.72e-12 ***
## CAR_TYPEVan                    6.329e-01  1.429e-01   4.428 9.51e-06 ***
## RED_CARyes                    -1.138e-01  9.730e-02  -1.170 0.242142
## CLM_FREQ                       5.041e-02  5.036e-02   1.001 0.316827
## REVOKEDYes                     8.822e-01  1.024e-01   8.619  < 2e-16 ***
## MVR_PTS                        9.784e-02  1.588e-02   6.163 7.15e-10 ***
## URBANICITYHighly Urban/ Urban  2.289e+00  1.249e-01  18.321  < 2e-16 ***
## CAR_AGE_BINLike New           -1.338e-01  3.469e-01  -0.386 0.699741
## CAR_AGE_BINAverage            -1.262e-01  8.393e-02  -1.503 0.132808
## CAR_AGE_BINOld                -1.346e-01  1.290e-01  -1.044 0.296614
## HOME_VAL_BIN$50k-$150k        -3.229e-01  1.266e-01  -2.551 0.010744 *
## HOME_VAL_BIN$150k-$250k       -3.035e-01  1.089e-01  -2.787 0.005324 **
## HOME_VAL_BINOver $250k        -5.742e-01  1.330e-01  -4.316 1.59e-05 ***
## HAS_HOME_KIDSNo kids          -2.294e-01  1.149e-01  -1.996 0.045923 *
## HAS_KIDSDRIVNo kids driving   -4.551e-01  1.114e-01  -4.085 4.41e-05 ***
## OLDCLAIM_BIN$0-$3k             4.055e-01  1.614e-01   2.513 0.011983 *
## OLDCLAIM_BIN$3k-$6k            3.729e-01  1.479e-01   2.522 0.011683 *
## OLDCLAIM_BIN$6k-$9k            5.461e-01  1.555e-01   3.512 0.000445 ***
## OLDCLAIM_BINOver $9k           3.841e-02  1.549e-01   0.248 0.804231
## TIF_BIN1-4 years              -2.044e-01  9.180e-02  -2.226 0.025982 *
## TIF_BIN4-7 years              -4.302e-01  8.854e-02  -4.859 1.18e-06 ***
## TIF_BINOver 7 years           -5.787e-01  9.156e-02  -6.320 2.62e-10 ***
## YOJ_BINLess than 10 years     -5.332e-01  1.659e-01  -3.214 0.001307 **
## YOJ_BINBetween 10-15 years    -5.828e-01  1.605e-01  -3.631 0.000282 ***
## YOJ_BINOver 15 years          -3.052e-01  2.154e-01  -1.417 0.156469
## ---
## 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: 5718.0  on 6399  degrees of freedom
##   (1713 observations deleted due to missingness)
## AIC: 5816
##
## Number of Fisher Scoring iterations: 5
```

This and the consequent model considers all binned variables plus old variables.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 862 196
##          1  65 167
##
##                Accuracy : 0.7977
```

```
##                  95% CI : (0.7747, 0.8193)
##     No Information Rate : 0.7186
##     P-Value [Acc > NIR] : 4.259e-11
##
##                   Kappa : 0.438
##
##  Mcnemar's Test P-Value : 8.499e-16
##
##             Sensitivity : 0.9299
##             Specificity : 0.4601
##          Pos Pred Value : 0.8147
##          Neg Pred Value : 0.7198
##              Prevalence : 0.7186
##          Detection Rate : 0.6682
##    Detection Prevalence : 0.8202
##       Balanced Accuracy : 0.6950
##
##        'Positive' Class : 0
##
```



Model 5

The next model provides a combination of imputation and binning.

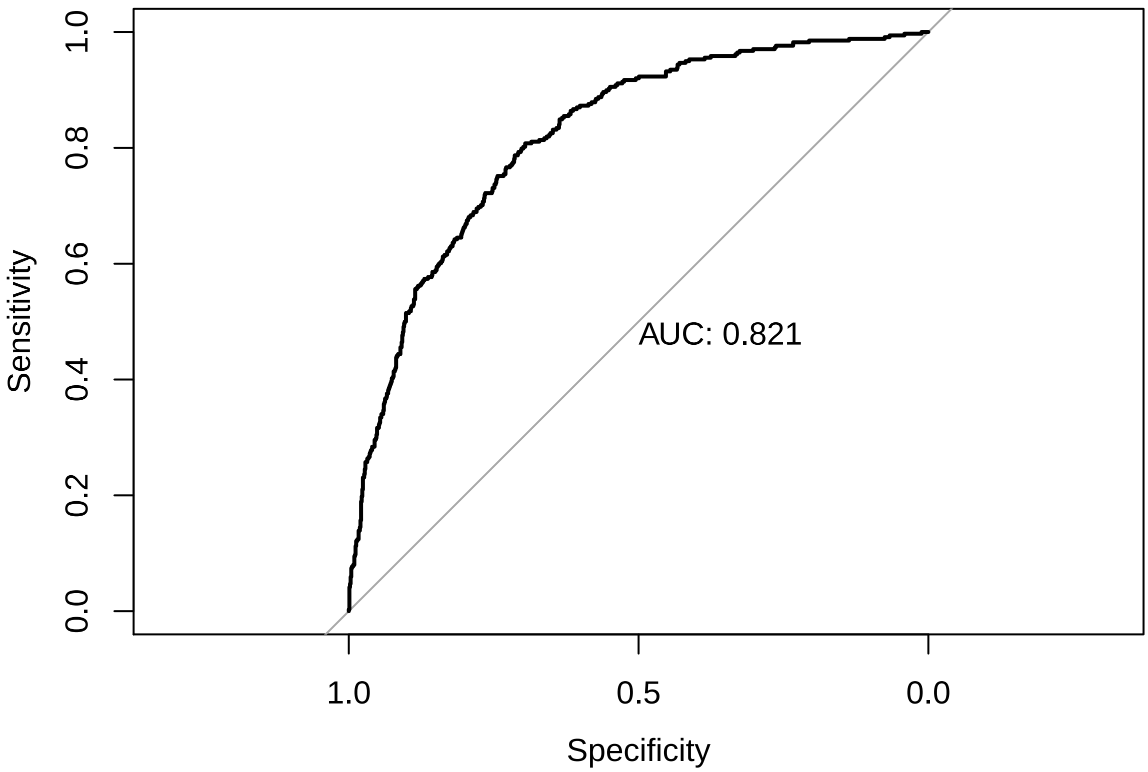
```
##
##  iter imp variable
##   1   1  AGE  INCOME  CAR_AGE_BIN  HOME_VAL_BIN  YOJ_BIN
##   2   1  AGE  INCOME  CAR_AGE_BIN  HOME_VAL_BIN  YOJ_BIN
##   3   1  AGE  INCOME  CAR_AGE_BIN  HOME_VAL_BIN  YOJ_BIN
##   4   1  AGE  INCOME  CAR_AGE_BIN  HOME_VAL_BIN  YOJ_BIN
```

```
##   5   1  AGE  INCOME  CAR_AGE_BIN  HOME_VAL_BIN  YOJ_BIN
## call :
## glm.mids(formula = TARGET_FLAG ~ . - TARGET_AMT, family = "binomial",
##     data = insurance_binned_impute)
##
## call1 :
## mice(data = insurance_bins, m = 1, method = "cart")
##
## nmis :
##   TARGET_FLAG    TARGET_AMT           AGE        INCOME       PARENT1
##             0             0             6           445             0
##       MSTATUS           SEX     EDUCATION           JOB      TRAVTIME
##             0             0             0             0             0
##       CAR_USE      BLUEBOOK      CAR_TYPE       RED_CAR      CLM_FREQ
##             0             0             0             0             0
##       REVOKED       MVR_PTS    URBANICITY   CAR_AGE_BIN  HOME_VAL_BIN
##             0             0             0           510           464
## HAS_HOME_KIDS  HAS_KIDSDRIV  OLDCLAIM_BIN       TIF_BIN       YOJ_BIN
##             0             0             0             0           454
##
## analyses :
## [[1]]
##
## Call:  glm(formula = formula, family = family, data = complete(data,
##     i))
##
## Coefficients:
##                    (Intercept)                              AGE
##                     -1.734e+00                       -7.178e-04
##                         INCOME                       PARENT1Yes
##                     -3.449e-06                        2.461e-01
##                     MSTATUSYes                             SEXM
##                     -5.170e-01                        9.158e-02
##           EDUCATIONHigh School  EDUCATIONLess than High School
##                      3.891e-01                        3.798e-01
##               EDUCATIONMasters                     EDUCATIONPhD
##                      1.073e-01                        2.039e-01
##                    JOBClerical                        JOBDoctor
##                      8.246e-02                       -7.537e-01
##                 JOBHome Maker                        JOBLawyer
##                     -2.709e-01                       -2.062e-01
##                     JOBManager                     JOBOther Job
##                     -8.592e-01                       -3.156e-01
##                JOBProfessional                       JOBStudent
##                     -1.531e-01                       -3.632e-01
##                       TRAVTIME                   CAR_USEPrivate
##                      1.488e-02                       -7.493e-01
##                       BLUEBOOK              CAR_TYPEPanel Truck
##                     -2.023e-05                        5.765e-01
##                 CAR_TYPEPickup               CAR_TYPESports Car
##                      5.616e-01                        1.011e+00
##                    CAR_TYPESUV                      CAR_TYPEVan
##                      7.750e-01                        6.148e-01
```

```
##                      RED_CARyes                          CLM_FREQ
##                      -3.817e-03                         5.084e-02
##                      REVOKEDYes                           MVR_PTS
##                       8.913e-01                         9.843e-02
##  URBANICITYHighly Urban/ Urban              CAR_AGE_BINLike New
##                       2.369e+00                         1.287e-01
##              CAR_AGE_BINAverage                    CAR_AGE_BINOld
##                      -6.374e-02                        -7.366e-02
##         HOME_VAL_BIN$50k-$150k        HOME_VAL_BIN$150k-$250k
##                      -3.077e-01                        -2.663e-01
##         HOME_VAL_BINOver $250k            HAS_HOME_KIDSNo kids
##                      -5.013e-01                        -2.195e-01
##    HAS_KIDSDRIVNo kids driving              OLDCLAIM_BIN$0-$3k
##                      -5.669e-01                         3.926e-01
##            OLDCLAIM_BIN$3k-$6k             OLDCLAIM_BIN$6k-$9k
##                       3.579e-01                         4.999e-01
##           OLDCLAIM_BINOver $9k                TIF_BIN1-4 years
##                       2.028e-02                        -1.924e-01
##               TIF_BIN4-7 years             TIF_BINOver 7 years
##                      -4.310e-01                        -5.888e-01
##      YOJ_BINLess than 10 years      YOJ_BINBetween 10-15 years
##                      -5.673e-01                        -6.194e-01
##           YOJ_BINOver 15 years
##                      -4.101e-01
##
## Degrees of Freedom: 8160 Total (i.e. Null);  8112 Residual
## Null Deviance:       9418
## Residual Deviance: 7250  AIC: 7348
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 889 186
##          1  74 150
##
##                Accuracy : 0.7998
##                  95% CI : (0.777, 0.8213)
##     No Information Rate : 0.7413
##     P-Value [Acc > NIR] : 4.533e-07
##
##                   Kappa : 0.4146
##
##  Mcnemar's Test P-Value : 5.822e-12
##
##             Sensitivity : 0.9232
##             Specificity : 0.4464
##          Pos Pred Value : 0.8270
##          Neg Pred Value : 0.6696
##              Prevalence : 0.7413
##          Detection Rate : 0.6844
##    Detection Prevalence : 0.8276
##       Balanced Accuracy : 0.6848
##
```

```
##        'Positive' Class : 0
##
```



Multiple Linear Regression

Model 1

Below code shows output for preliminary regression modelling insurance payout given that a claim has been predicted. R-squared values are very low, but this assumes that a correct prediction from the binary logistic model has been made.

```
##
## Call:
## lm(formula = TARGET_AMT ~ ., data = mlr_crash)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -9657  -3165  -1474    574  76279
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    4.075e+03  1.809e+03   2.253   0.0244 *
## KIDSDRIV                      -1.771e+02  3.556e+02  -0.498   0.6185
## AGE                            5.833e-01  2.351e+01   0.025   0.9802
## HOMEKIDS                       2.752e+02  2.295e+02   1.199   0.2306
## YOJ                            1.917e+01  5.463e+01   0.351   0.7256
## INCOME                        -1.510e-02  7.821e-03  -1.930   0.0537 .
## PARENT1Yes                    -9.951e+01  6.469e+02  -0.154   0.8778
## HOME_VAL                       2.230e-03  2.268e-03   0.984   0.3255
## MSTATUSYes                    -1.387e+03  5.662e+02  -2.450   0.0144 *
```

```
## SEXM                            1.816e+03  7.167e+02   2.534   0.0114 *
## EDUCATIONHigh School           -8.578e+02  5.772e+02  -1.486   0.1374
## EDUCATIONLess than High School -1.712e+02  7.149e+02  -0.239   0.8108
## EDUCATIONMasters                6.457e+02  1.048e+03   0.616   0.5380
## EDUCATIONPhD                    2.938e+03  1.282e+03   2.293   0.0220 *
## JOBClerical                    -1.143e+03  6.452e+02  -1.772   0.0766 .
## JOBDoctor                      -3.784e+03  1.998e+03  -1.894   0.0584 .
## JOBHome Maker                  -1.046e+03  9.995e+02  -1.047   0.2954
## JOBLawyer                      -6.243e+02  1.323e+03  -0.472   0.6370
## JOBManager                     -1.788e+03  1.042e+03  -1.716   0.0864 .
## JOBOther Job                   -4.589e+02  1.304e+03  -0.352   0.7250
## JOBProfessional                 7.702e+02  7.712e+02   0.999   0.3181
## JOBStudent                     -1.059e+03  8.089e+02  -1.309   0.1905
## TRAVTIME                        4.108e+00  1.234e+01   0.333   0.7393
## CAR_USEPrivate                 -2.737e+02  5.849e+02  -0.468   0.6399
## BLUEBOOK                        1.486e-01  3.376e-02   4.402 1.14e-05 ***
## TIF                            -5.847e+00  4.695e+01  -0.125   0.9009
## CAR_TYPEPanel Truck            -2.619e+02  1.053e+03  -0.249   0.8036
## CAR_TYPEPickup                  3.003e+02  6.627e+02   0.453   0.6505
## CAR_TYPESports Car              1.951e+03  8.262e+02   2.361   0.0183 *
## CAR_TYPESUV                     1.657e+03  7.363e+02   2.251   0.0245 *
## CAR_TYPEVan                    -2.228e+02  8.588e+02  -0.259   0.7953
## RED_CARyes                     -3.138e+02  5.511e+02  -0.569   0.5692
## OLDCLAIM                        5.024e-02  2.528e-02   1.987   0.0471 *
## CLM_FREQ                       -2.048e+02  1.749e+02  -1.171   0.2416
## REVOKEDYes                     -1.259e+03  5.850e+02  -2.152   0.0315 *
## MVR_PTS                         8.937e+01  7.564e+01   1.182   0.2375
## CAR_AGE                        -9.797e+01  4.878e+01  -2.009   0.0447 *
## URBANICITYHighly Urban/ Urban   5.991e+01  8.182e+02   0.073   0.9416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7586 on 1665 degrees of freedom
##   (450 observations deleted due to missingness)
## Multiple R-squared:  0.04273,    Adjusted R-squared:  0.02145
## F-statistic: 2.009 on 37 and 1665 DF,  p-value: 0.000334
```

The R^2 value is very low, around 4%, and many of the variables are not significant.

Model 2

Using our log transformation on certain variables, the results are slightly worse.

```
##
## Call:
## lm(formula = TARGET_AMT ~ ., data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8045  -3199  -1526    438  99546
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -9715.099   4630.184  -2.098   0.0360 *
## KIDSDRIV                        -186.329    320.282  -0.582   0.5608
```

```
## AGE                               544.526    882.174   0.617   0.5371
## HOMEKIDS                          187.340    209.948   0.892   0.3723
## YOJ                                 8.150     61.050   0.133   0.8938
## INCOME                             22.840     96.307   0.237   0.8126
## PARENT1Yes                        331.308    588.943   0.563   0.5738
## HOME_VAL                           58.650     38.287   1.532   0.1257
## MSTATUSYes                       -868.702    509.343  -1.706   0.0882 .
## SEXM                             1212.639    630.947   1.922   0.0547 .
## EDUCATIONHigh School             -457.376    505.973  -0.904   0.3661
## EDUCATIONLess than High School     51.500    635.038   0.081   0.9354
## EDUCATIONMasters                  548.316    883.446   0.621   0.5349
## EDUCATIONPhD                     1658.219   1088.609   1.523   0.1278
## JOBClerical                       -85.075    581.159  -0.146   0.8836
## JOBDoctor                       -2759.504   1870.439  -1.475   0.1403
## JOBHome Maker                     -73.493    941.671  -0.078   0.9378
## JOBLawyer                        -249.977   1173.707  -0.213   0.8314
## JOBManager                      -1310.356    904.347  -1.449   0.1475
## JOBOther Job                     -529.041   1140.250  -0.464   0.6427
## JOBProfessional                   509.067    684.161   0.744   0.4569
## JOBStudent                        317.311    799.632   0.397   0.6915
## TRAVTIME                          -51.921    299.067  -0.174   0.8622
## CAR_USEPrivate                   -345.492    522.462  -0.661   0.5085
## BLUEBOOK                         1398.356    328.055   4.263 2.11e-05 ***
## TIF                               -14.903     42.536  -0.350   0.7261
## CAR_TYPEPanel Truck               -29.775    881.064  -0.034   0.9730
## CAR_TYPEPickup                   -136.236    596.552  -0.228   0.8194
## CAR_TYPESports Car               1011.268    735.029   1.376   0.1690
## CAR_TYPESUV                       677.040    643.223   1.053   0.2927
## CAR_TYPEVan                       135.500    762.155   0.178   0.8589
## RED_CARyes                       -192.707    497.240  -0.388   0.6984
## OLDCLAIM                            7.773     67.902   0.114   0.9089
## CLM_FREQ                          -67.375    232.751  -0.289   0.7722
## REVOKEDYes                       -765.210    422.770  -1.810   0.0704 .
## MVR_PTS                           126.448     70.048   1.805   0.0712 .
## CAR_AGE                          -380.023    263.152  -1.444   0.1489
## URBANICITYHighly Urban/ Urban      31.111    755.064   0.041   0.9671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7695 on 2115 degrees of freedom
## Multiple R-squared:  0.02941,    Adjusted R-squared:  0.01244
## F-statistic: 1.732 on 37 and 2115 DF,  p-value: 0.004147
```

Model 3: Backwards Elimination

Now let's use backwards elimination to remove some of variables that are not significant.

```
##
## Call:
## lm(formula = TARGET_AMT ~ ., data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8045  -3199  -1526    438  99546
##
```

```
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                   -9715.099   4630.184  -2.098   0.0360 *  
## KIDSDRIV                       -186.329    320.282  -0.582   0.5608    
## AGE                             544.526    882.174   0.617   0.5371    
## HOMEKIDS                        187.340    209.948   0.892   0.3723    
## YOJ                               8.150     61.050   0.133   0.8938    
## INCOME                           22.840     96.307   0.237   0.8126    
## PARENT1Yes                      331.308    588.943   0.563   0.5738    
## HOME_VAL                         58.650     38.287   1.532   0.1257    
## MSTATUSYes                     -868.702    509.343  -1.706   0.0882 .  
## SEXM                           1212.639    630.947   1.922   0.0547 .  
## EDUCATIONHigh School           -457.376    505.973  -0.904   0.3661    
## EDUCATIONLess than High School   51.500    635.038   0.081   0.9354    
## EDUCATIONMasters                548.316    883.446   0.621   0.5349    
## EDUCATIONPhD                   1658.219   1088.609   1.523   0.1278    
## JOBClerical                     -85.075    581.159  -0.146   0.8836    
## JOBDoctor                     -2759.504   1870.439  -1.475   0.1403    
## JOBHome Maker                   -73.493    941.671  -0.078   0.9378    
## JOBLawyer                      -249.977   1173.707  -0.213   0.8314    
## JOBManager                    -1310.356    904.347  -1.449   0.1475    
## JOBOther Job                   -529.041   1140.250  -0.464   0.6427    
## JOBProfessional                 509.067    684.161   0.744   0.4569    
## JOBStudent                      317.311    799.632   0.397   0.6915    
## TRAVTIME                        -51.921    299.067  -0.174   0.8622    
## CAR_USEPrivate                 -345.492    522.462  -0.661   0.5085    
## BLUEBOOK                       1398.356    328.055   4.263 2.11e-05 ***
## TIF                             -14.903     42.536  -0.350   0.7261    
## CAR_TYPEPanel Truck             -29.775    881.064  -0.034   0.9730    
## CAR_TYPEPickup                 -136.236    596.552  -0.228   0.8194    
## CAR_TYPESports Car             1011.268    735.029   1.376   0.1690    
## CAR_TYPESUV                     677.040    643.223   1.053   0.2927    
## CAR_TYPEVan                     135.500    762.155   0.178   0.8589    
## RED_CARyes                     -192.707    497.240  -0.388   0.6984    
## OLDCLAIM                          7.773     67.902   0.114   0.9089    
## CLM_FREQ                        -67.375    232.751  -0.289   0.7722    
## REVOKEDYes                     -765.210    422.770  -1.810   0.0704 .  
## MVR_PTS                         126.448     70.048   1.805   0.0712 .  
## CAR_AGE                        -380.023    263.152  -1.444   0.1489    
## URBANICITYHighly Urban/ Urban    31.111    755.064   0.041   0.9671    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 7695 on 2115 degrees of freedom
## Multiple R-squared:  0.02941,    Adjusted R-squared:  0.01244 
## F-statistic: 1.732 on 37 and 2115 DF,  p-value: 0.004147
```

```
## 
## Call:
## lm(formula = TARGET_AMT ~ KIDSDRIV + AGE + HOMEKIDS + YOJ + INCOME + 
##     PARENT1 + HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + TRAVTIME + 
##     CAR_USE + BLUEBOOK + TIF + CAR_TYPE + RED_CAR + CLM_FREQ + 
##     REVOKED + MVR_PTS + CAR_AGE + URBANICITY, data = mlr_crash_transf)
## 
```

```
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8055  -3195  -1534    449  99520
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -9703.231   4627.944  -2.097   0.0361 *
## KIDSDRIV                        -186.712    320.190  -0.583   0.5599
## AGE                              543.441    881.917   0.616   0.5378
## HOMEKIDS                         187.371    209.899   0.893   0.3721
## YOJ                                8.449     60.979   0.139   0.8898
## INCOME                            22.822     96.285   0.237   0.8127
## PARENT1Yes                       328.742    588.379   0.559   0.5764
## HOME_VAL                          58.642     38.278   1.532   0.1257
## MSTATUSYes                      -869.123    509.211  -1.707   0.0880 .
## SEXM                            1213.494    630.756   1.924   0.0545 .
## EDUCATIONHigh School            -457.887    505.835  -0.905   0.3655
## EDUCATIONLess than High School    51.393    634.890   0.081   0.9355
## EDUCATIONMasters                 543.613    882.285   0.616   0.5379
## EDUCATIONPhD                    1652.076   1087.033   1.520   0.1287
## JOBClerical                      -82.867    580.703  -0.143   0.8865
## JOBDoctor                      -2765.994   1869.144  -1.480   0.1391
## JOBHome Maker                    -69.836    940.909  -0.074   0.9408
## JOBLawyer                       -242.197   1171.465  -0.207   0.8362
## JOBManager                     -1307.098    903.688  -1.446   0.1482
## JOBOther Job                    -522.305   1138.465  -0.459   0.6464
## JOBProfessional                  511.708    683.613   0.749   0.4542
## JOBStudent                       319.696    799.174   0.400   0.6892
## TRAVTIME                         -52.423    298.965  -0.175   0.8608
## CAR_USEPrivate                  -347.085    522.155  -0.665   0.5063
## BLUEBOOK                        1398.320    327.978   4.263  2.1e-05 ***
## TIF                              -14.956     42.524  -0.352   0.7251
## CAR_TYPEPanel Truck              -33.151    880.365  -0.038   0.9700
## CAR_TYPEPickup                  -137.900    596.236  -0.231   0.8171
## CAR_TYPESports Car              1012.421    734.788   1.378   0.1684
## CAR_TYPESUV                      676.299    643.040   1.052   0.2930
## CAR_TYPEVan                      135.417    761.977   0.178   0.8590
## RED_CARyes                      -194.931    496.745  -0.392   0.6948
## CLM_FREQ                         -46.161    140.797  -0.328   0.7431
## REVOKEDYes                      -756.269    415.397  -1.821   0.0688 .
## MVR_PTS                          128.158     68.418   1.873   0.0612 .
## CAR_AGE                         -379.748    263.080  -1.443   0.1490
## URBANICITYHighly Urban/ Urban     31.696    754.871   0.042   0.9665
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7693 on 2116 degrees of freedom
## Multiple R-squared:  0.02941,    Adjusted R-squared:  0.0129
## F-statistic: 1.781 on 36 and 2116 DF,  p-value: 0.003007
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ KIDSDRIV + AGE + HOMEKIDS + INCOME +
##     PARENT1 + HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + TRAVTIME +
```

```
##     CAR_USE + BLUEBOOK + TIF + CAR_TYPE + RED_CAR + CLM_FREQ +
##     REVOKED + MVR_PTS + CAR_AGE + URBANICITY, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8028  -3203  -1530    439  99526
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -9802.39    4571.21  -2.144   0.0321 *
## KIDSDRIV                       -190.69     318.83  -0.598   0.5498
## AGE                             565.15     867.68   0.651   0.5149
## HOMEKIDS                        193.93     204.45   0.949   0.3430
## INCOME                           30.91      76.57   0.404   0.6865
## PARENT1Yes                      329.39     588.22   0.560   0.5756
## HOME_VAL                         58.81      38.25   1.538   0.1243
## MSTATUSYes                     -860.73     505.48  -1.703   0.0888 .
## SEXM                           1215.25     630.48   1.927   0.0541 .
## EDUCATIONHigh School           -456.40     505.60  -0.903   0.3668
## EDUCATIONLess than High School   57.35     633.28   0.091   0.9278
## EDUCATIONMasters                544.42     882.06   0.617   0.5372
## EDUCATIONPhD                   1651.22    1086.76   1.519   0.1288
## JOBClerical                     -81.44     580.48  -0.140   0.8884
## JOBDoctor                     -2766.26    1868.71  -1.480   0.1389
## JOBHome Maker                   -71.81     940.58  -0.076   0.9392
## JOBLawyer                      -244.04    1171.12  -0.208   0.8350
## JOBManager                    -1307.12     903.48  -1.447   0.1481
## JOBOther Job                   -524.53    1138.09  -0.461   0.6449
## JOBProfessional                 508.91     683.16   0.745   0.4564
## JOBStudent                      321.71     798.86   0.403   0.6872
## TRAVTIME                        -53.43     298.81  -0.179   0.8581
## CAR_USEPrivate                 -344.52     521.71  -0.660   0.5091
## BLUEBOOK                       1400.31     327.59   4.275    2e-05 ***
## TIF                             -15.01      42.51  -0.353   0.7241
## CAR_TYPEPanel Truck             -39.29     879.05  -0.045   0.9644
## CAR_TYPEPickup                 -138.62     596.07  -0.233   0.8161
## CAR_TYPESports Car             1008.47     734.06   1.374   0.1696
## CAR_TYPESUV                     676.28     642.89   1.052   0.2929
## CAR_TYPEVan                     129.97     760.79   0.171   0.8644
## RED_CARyes                     -195.58     496.61  -0.394   0.6938
## CLM_FREQ                        -46.05     140.76  -0.327   0.7436
## REVOKEDYes                     -753.35     414.77  -1.816   0.0695 .
## MVR_PTS                         128.13      68.40   1.873   0.0612 .
## CAR_AGE                        -380.42     262.97  -1.447   0.1482
## URBANICITYHighly Urban/ Urban    32.33     754.68   0.043   0.9658
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7691 on 2117 degrees of freedom
## Multiple R-squared:  0.0294, Adjusted R-squared:  0.01335
## F-statistic: 1.832 on 35 and 2117 DF,  p-value: 0.002154
```

```
##
## Call:
```

```
## lm(formula = TARGET_AMT ~ KIDSDRIV + AGE + HOMEKIDS + INCOME +
##     PARENT1 + HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + TRAVTIME +
##     CAR_USE + BLUEBOOK + TIF + CAR_TYPE + RED_CAR + CLM_FREQ +
##     REVOKED + MVR_PTS + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8029  -3200  -1530    442  99526
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -9767.63    4497.57  -2.172   0.0300 *
## KIDSDRIV                        -191.06     318.64  -0.600   0.5488
## AGE                              563.91     866.99   0.650   0.5155
## HOMEKIDS                         193.78     204.37   0.948   0.3432
## INCOME                            30.97      76.54   0.405   0.6858
## PARENT1Yes                       329.24     588.07   0.560   0.5756
## HOME_VAL                          58.77      38.23   1.537   0.1244
## MSTATUSYes                      -859.38     504.37  -1.704   0.0886 .
## SEXM                            1214.56     630.13   1.927   0.0541 .
## EDUCATIONHigh School            -456.51     505.48  -0.903   0.3666
## EDUCATIONLess than High School    57.49     633.13   0.091   0.9277
## EDUCATIONMasters                 544.35     881.85   0.617   0.5371
## EDUCATIONPhD                    1651.00    1086.49   1.520   0.1288
## JOBClerical                      -83.04     579.13  -0.143   0.8860
## JOBDoctor                      -2764.75    1867.94  -1.480   0.1390
## JOBHome Maker                    -71.56     940.34  -0.076   0.9393
## JOBLawyer                       -244.07    1170.84  -0.208   0.8349
## JOBManager                     -1305.71     902.66  -1.447   0.1482
## JOBOther Job                    -523.68    1137.64  -0.460   0.6453
## JOBProfessional                  508.32     682.86   0.744   0.4567
## JOBStudent                       318.99     796.14   0.401   0.6887
## TRAVTIME                         -54.22     298.16  -0.182   0.8557
## CAR_USEPrivate                  -344.51     521.58  -0.661   0.5090
## BLUEBOOK                        1400.54     327.47   4.277 1.98e-05 ***
## TIF                              -14.97      42.49  -0.352   0.7246
## CAR_TYPEPanel Truck              -38.22     878.48  -0.044   0.9653
## CAR_TYPEPickup                  -138.32     595.89  -0.232   0.8165
## CAR_TYPESports Car              1008.24     733.87   1.374   0.1696
## CAR_TYPESUV                      676.31     642.74   1.052   0.2928
## CAR_TYPEVan                      130.50     760.51   0.172   0.8638
## RED_CARyes                      -195.48     496.49  -0.394   0.6938
## CLM_FREQ                         -45.73     140.53  -0.325   0.7449
## REVOKEDYes                      -752.87     414.51  -1.816   0.0695 .
## MVR_PTS                          128.21      68.36   1.875   0.0609 .
## CAR_AGE                         -380.35     262.91  -1.447   0.1481
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7689 on 2118 degrees of freedom
## Multiple R-squared:  0.0294, Adjusted R-squared:  0.01382
## F-statistic: 1.887 on 34 and 2118 DF,  p-value: 0.001515

##
```

```
## Call:
## lm(formula = TARGET_AMT ~ KIDSDRIV + AGE + HOMEKIDS + INCOME +
##     PARENT1 + HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE +
##     BLUEBOOK + TIF + CAR_TYPE + RED_CAR + CLM_FREQ + REVOKED +
##     MVR_PTS + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7928  -3193  -1536    437  99511
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -9919.35    4418.51  -2.245   0.0249 *
## KIDSDRIV                        -190.38     318.54  -0.598   0.5501
## AGE                              561.46     866.69   0.648   0.5172
## HOMEKIDS                         193.67     204.33   0.948   0.3433
## INCOME                            30.55      76.49   0.399   0.6896
## PARENT1Yes                       332.46     587.67   0.566   0.5716
## HOME_VAL                          58.96      38.20   1.543   0.1229
## MSTATUSYes                      -860.93     504.18  -1.708   0.0879 .
## SEXM                            1212.02     629.83   1.924   0.0544 .
## EDUCATIONHigh School            -453.99     505.17  -0.899   0.3689
## EDUCATIONLess than High School    59.11     632.92   0.093   0.9256
## EDUCATIONMasters                 542.00     881.56   0.615   0.5387
## EDUCATIONPhD                    1647.94    1086.12   1.517   0.1293
## JOBClerical                      -81.79     578.96  -0.141   0.8877
## JOBDoctor                      -2761.12    1867.40  -1.479   0.1394
## JOBHome Maker                    -74.69     939.97  -0.079   0.9367
## JOBLawyer                       -239.16    1170.26  -0.204   0.8381
## JOBManager                     -1301.37     902.14  -1.443   0.1493
## JOBOther Job                    -517.79    1136.92  -0.455   0.6488
## JOBProfessional                  508.69     682.70   0.745   0.4563
## JOBStudent                       322.09     795.78   0.405   0.6857
## CAR_USEPrivate                  -348.16     521.08  -0.668   0.5041
## BLUEBOOK                        1398.46     327.19   4.274    2e-05 ***
## TIF                              -14.75      42.47  -0.347   0.7284
## CAR_TYPEPanel Truck              -39.82     878.24  -0.045   0.9638
## CAR_TYPEPickup                  -136.54     595.68  -0.229   0.8187
## CAR_TYPESports Car              1009.62     733.66   1.376   0.1689
## CAR_TYPESUV                      673.92     642.46   1.049   0.2943
## CAR_TYPEVan                      133.45     760.16   0.176   0.8607
## RED_CARyes                      -197.06     496.30  -0.397   0.6914
## CLM_FREQ                         -46.24     140.47  -0.329   0.7421
## REVOKEDYes                      -751.98     414.39  -1.815   0.0697 .
## MVR_PTS                          128.03      68.34   1.873   0.0611 .
## CAR_AGE                         -381.09     262.82  -1.450   0.1472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7688 on 2119 degrees of freedom
## Multiple R-squared:  0.02938,    Adjusted R-squared:  0.01427
## F-statistic: 1.944 on 33 and 2119 DF,  p-value: 0.001059

##
```

```
## Call:
## lm(formula = TARGET_AMT ~ KIDSDRIV + AGE + HOMEKIDS + PARENT1 +
##     HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##     TIF + CAR_TYPE + RED_CAR + CLM_FREQ + REVOKED + MVR_PTS +
##     CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7925  -3197  -1545    443  99526
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -9694.85    4381.75  -2.213   0.0270 *
## KIDSDRIV                        -185.98     318.29  -0.584   0.5591
## AGE                              564.77     866.48   0.652   0.5146
## HOMEKIDS                         192.47     204.26   0.942   0.3462
## PARENT1Yes                       326.40     587.36   0.556   0.5785
## HOME_VAL                          59.53      38.17   1.560   0.1190
## MSTATUSYes                      -866.79     503.87  -1.720   0.0855 .
## SEXM                            1214.06     629.69   1.928   0.0540 .
## EDUCATIONHigh School            -457.37     505.00  -0.906   0.3652
## EDUCATIONLess than High School    39.79     630.95   0.063   0.9497
## EDUCATIONMasters                 551.82     881.04   0.626   0.5312
## EDUCATIONPhD                    1658.08    1085.60   1.527   0.1268
## JOBClerical                      -97.88     577.44  -0.170   0.8654
## JOBDoctor                      -2783.28    1866.21  -1.491   0.1360
## JOBHome Maker                   -292.97     764.65  -0.383   0.7017
## JOBLawyer                       -254.76    1169.38  -0.218   0.8276
## JOBManager                     -1308.39     901.79  -1.451   0.1470
## JOBOther Job                    -521.56    1136.66  -0.459   0.6464
## JOBProfessional                  502.63     682.39   0.737   0.4615
## JOBStudent                       129.67     633.27   0.205   0.8378
## CAR_USEPrivate                  -337.81     520.33  -0.649   0.5163
## BLUEBOOK                        1408.77     326.11   4.320 1.63e-05 ***
## TIF                              -15.27      42.44  -0.360   0.7191
## CAR_TYPEPanel Truck              -30.76     877.77  -0.035   0.9721
## CAR_TYPEPickup                  -125.32     594.89  -0.211   0.8332
## CAR_TYPESports Car              1007.17     733.49   1.373   0.1699
## CAR_TYPESUV                      682.65     641.96   1.063   0.2877
## CAR_TYPEVan                      139.30     759.87   0.183   0.8546
## RED_CARyes                      -199.44     496.16  -0.402   0.6878
## CLM_FREQ                         -46.11     140.44  -0.328   0.7427
## REVOKEDYes                      -752.76     414.30  -1.817   0.0694 .
## MVR_PTS                          126.28      68.18   1.852   0.0642 .
## CAR_AGE                         -380.99     262.76  -1.450   0.1472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7686 on 2120 degrees of freedom
## Multiple R-squared:  0.02931,    Adjusted R-squared:  0.01466
## F-statistic:     2 on 32 and 2120 DF,  p-value: 0.0007551
```

```
##
## Call:
```

```
## lm(formula = TARGET_AMT ~ KIDSDRIV + AGE + HOMEKIDS + PARENT1 +
##     HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##     TIF + CAR_TYPE + RED_CAR + REVOKED + MVR_PTS + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7934  -3210  -1541    443  99469
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -9717.22    4380.30  -2.218   0.0266 *
## KIDSDRIV                        -187.09     318.20  -0.588   0.5566
## AGE                              560.79     866.21   0.647   0.5174
## HOMEKIDS                         192.96     204.21   0.945   0.3448
## PARENT1Yes                       327.70     587.22   0.558   0.5769
## HOME_VAL                          59.68      38.16   1.564   0.1180
## MSTATUSYes                      -868.05     503.75  -1.723   0.0850 .
## SEXM                            1215.55     629.54   1.931   0.0536 .
## EDUCATIONHigh School            -455.67     504.87  -0.903   0.3669
## EDUCATIONLess than High School    42.80     630.75   0.068   0.9459
## EDUCATIONMasters                 546.87     880.72   0.621   0.5347
## EDUCATIONPhD                    1655.60    1085.35   1.525   0.1273
## JOBClerical                      -98.34     577.31  -0.170   0.8648
## JOBDoctor                      -2814.40    1863.41  -1.510   0.1311
## JOBHome Maker                   -294.97     764.46  -0.386   0.6996
## JOBLawyer                       -238.46    1168.08  -0.204   0.8383
## JOBManager                     -1296.96     900.93  -1.440   0.1501
## JOBOther Job                    -517.27    1136.35  -0.455   0.6490
## JOBProfessional                  503.33     682.25   0.738   0.4607
## JOBStudent                       131.22     633.12   0.207   0.8358
## CAR_USEPrivate                  -335.11     520.16  -0.644   0.5195
## BLUEBOOK                        1409.19     326.04   4.322 1.62e-05 ***
## TIF                              -15.71      42.41  -0.370   0.7111
## CAR_TYPEPanel Truck              -29.39     877.58  -0.033   0.9733
## CAR_TYPEPickup                  -127.40     594.74  -0.214   0.8304
## CAR_TYPESports Car              1000.79     733.08   1.365   0.1723
## CAR_TYPESUV                      683.27     641.82   1.065   0.2872
## CAR_TYPEVan                      143.65     759.59   0.189   0.8500
## RED_CARyes                      -202.05     495.99  -0.407   0.6838
## REVOKEDYes                      -754.20     414.19  -1.821   0.0688 .
## MVR_PTS                          119.70      65.16   1.837   0.0663 .
## CAR_AGE                         -385.16     262.40  -1.468   0.1423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7685 on 2121 degrees of freedom
## Multiple R-squared:  0.02926,    Adjusted R-squared:  0.01507
## F-statistic: 2.062 on 31 and 2121 DF,  p-value: 0.0005236
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ KIDSDRIV + AGE + HOMEKIDS + PARENT1 +
##     HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##     CAR_TYPE + RED_CAR + REVOKED + MVR_PTS + CAR_AGE, data = mlr_crash_transf)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7929  -3210  -1538    442  99523
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -9820.06    4370.60  -2.247   0.0248 *
## KIDSDRIV                       -186.88     318.14  -0.587   0.5570
## AGE                             563.78     866.00   0.651   0.5151
## HOMEKIDS                        192.10     204.16   0.941   0.3468
## PARENT1Yes                      332.02     586.99   0.566   0.5717
## HOME_VAL                         59.66      38.15   1.564   0.1180
## MSTATUSYes                     -859.82     503.16  -1.709   0.0876 .
## SEXM                           1216.81     629.40   1.933   0.0533 .
## EDUCATIONHigh School           -457.18     504.75  -0.906   0.3652
## EDUCATIONLess than High School   41.75     630.61   0.066   0.9472
## EDUCATIONMasters                542.75     880.47   0.616   0.5377
## EDUCATIONPhD                   1653.85    1085.12   1.524   0.1276
## JOBClerical                    -104.83     576.93  -0.182   0.8558
## JOBDoctor                     -2798.33    1862.52  -1.502   0.1331
## JOBHome Maker                  -294.00     764.30  -0.385   0.7005
## JOBLawyer                      -232.83    1167.74  -0.199   0.8420
## JOBManager                    -1294.47     900.72  -1.437   0.1508
## JOBOther Job                   -520.50    1136.08  -0.458   0.6469
## JOBProfessional                 499.74     682.04   0.733   0.4638
## JOBStudent                      134.49     632.93   0.212   0.8317
## CAR_USEPrivate                 -323.75     519.14  -0.624   0.5329
## BLUEBOOK                       1409.68     325.97   4.325  1.6e-05 ***
## CAR_TYPEPanel Truck             -22.29     877.19  -0.025   0.9797
## CAR_TYPEPickup                 -125.55     594.59  -0.211   0.8328
## CAR_TYPESports Car              997.34     732.87   1.361   0.1737
## CAR_TYPESUV                     680.38     641.64   1.060   0.2891
## CAR_TYPEVan                     146.89     759.39   0.193   0.8466
## RED_CARyes                     -200.26     495.87  -0.404   0.6864
## REVOKEDYes                     -751.21     414.03  -1.814   0.0698 .
## MVR_PTS                         120.49      65.11   1.851   0.0644 .
## CAR_AGE                        -384.91     262.35  -1.467   0.1425
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7683 on 2122 degrees of freedom
## Multiple R-squared:  0.0292, Adjusted R-squared:  0.01547
## F-statistic: 2.127 on 30 and 2122 DF,  p-value: 0.0003608
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ KIDSDRIV + AGE + HOMEKIDS + PARENT1 +
##     HOME_VAL + MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK +
##     CAR_TYPE + REVOKED + MVR_PTS + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7921  -3209  -1542    438  99449
```

```
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -9915.44    4363.36  -2.272   0.0232 *
## KIDSDRIV                        -184.70     318.03  -0.581   0.5615
## AGE                              578.39     865.07   0.669   0.5038
## HOMEKIDS                         192.45     204.12   0.943   0.3459
## PARENT1Yes                       333.44     586.86   0.568   0.5700
## HOME_VAL                          59.81      38.14   1.568   0.1170
## MSTATUSYes                      -860.83     503.05  -1.711   0.0872 .
## SEXM                            1104.16     564.11   1.957   0.0504 .
## EDUCATIONHigh School            -450.55     504.38  -0.893   0.3718
## EDUCATIONLess than High School    48.79     630.25   0.077   0.9383
## EDUCATIONMasters                 548.71     880.18   0.623   0.5331
## EDUCATIONPhD                    1666.91    1084.42   1.537   0.1244
## JOBClerical                      -97.36     576.52  -0.169   0.8659
## JOBDoctor                      -2807.36    1862.02  -1.508   0.1318
## JOBHome Maker                   -292.72     764.14  -0.383   0.7017
## JOBLawyer                       -234.95    1167.50  -0.201   0.8405
## JOBManager                     -1300.32     900.42  -1.444   0.1489
## JOBOther Job                    -535.77    1135.23  -0.472   0.6370
## JOBProfessional                  502.88     681.86   0.738   0.4609
## JOBStudent                       129.31     632.67   0.204   0.8381
## CAR_USEPrivate                  -327.47     518.96  -0.631   0.5281
## BLUEBOOK                        1412.50     325.83   4.335 1.53e-05 ***
## CAR_TYPEPanel Truck              -34.26     876.52  -0.039   0.9688
## CAR_TYPEPickup                  -129.40     594.40  -0.218   0.8277
## CAR_TYPESports Car              1000.28     732.69   1.365   0.1723
## CAR_TYPESUV                      688.83     641.18   1.074   0.2828
## CAR_TYPEVan                      142.73     759.17   0.188   0.8509
## REVOKEDYes                      -748.97     413.91  -1.809   0.0705 .
## MVR_PTS                          119.73      65.07   1.840   0.0659 .
## CAR_AGE                         -383.29     262.26  -1.461   0.1440
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7682 on 2123 degrees of freedom
## Multiple R-squared:  0.02912,    Adjusted R-squared:  0.01586
## F-statistic: 2.196 on 29 and 2123 DF,  p-value: 0.0002469
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ KIDSDRIV + AGE + HOMEKIDS + HOME_VAL +
##     MSTATUS + SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK + CAR_TYPE +
##     REVOKED + MVR_PTS + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8001  -3182  -1544    429  99508
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -9645.55    4336.73  -2.224   0.0262 *
## KIDSDRIV                        -177.34     317.72  -0.558   0.5768
```

```
## AGE                             522.09     859.24   0.608   0.5435
## HOMEKIDS                        244.05     182.77   1.335   0.1819
## HOME_VAL                         59.37      38.13   1.557   0.1196
## MSTATUSYes                    -1008.08     431.08  -2.338   0.0195 *
## SEXM                           1101.07     563.99   1.952   0.0510 .
## EDUCATIONHigh School           -443.57     504.15  -0.880   0.3791
## EDUCATIONLess than High School   50.65     630.14   0.080   0.9359
## EDUCATIONMasters                533.07     879.61   0.606   0.5446
## EDUCATIONPhD                   1656.38    1084.09   1.528   0.1267
## JOBClerical                     -96.79     576.43  -0.168   0.8667
## JOBDoctor                     -2822.82    1861.53  -1.516   0.1296
## JOBHome Maker                  -291.98     764.02  -0.382   0.7024
## JOBLawyer                      -211.22    1166.57  -0.181   0.8563
## JOBManager                    -1282.01     899.70  -1.425   0.1543
## JOBOther Job                   -519.14    1134.67  -0.458   0.6473
## JOBProfessional                 518.77     681.18   0.762   0.4464
## JOBStudent                      126.15     632.54   0.199   0.8419
## CAR_USEPrivate                 -322.17     518.79  -0.621   0.5347
## BLUEBOOK                       1415.19     325.74   4.345 1.46e-05 ***
## CAR_TYPEPanel Truck             -44.90     876.18  -0.051   0.9591
## CAR_TYPEPickup                 -133.98     594.25  -0.225   0.8216
## CAR_TYPESports Car             1007.48     732.47   1.375   0.1691
## CAR_TYPESUV                     690.69     641.06   1.077   0.2814
## CAR_TYPEVan                     134.90     758.92   0.178   0.8589
## REVOKEDYes                     -754.18     413.75  -1.823   0.0685 .
## MVR_PTS                         120.97      65.02   1.860   0.0630 .
## CAR_AGE                        -379.67     262.15  -1.448   0.1477
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7680 on 2124 degrees of freedom
## Multiple R-squared:  0.02898,    Adjusted R-squared:  0.01618
## F-statistic: 2.264 on 28 and 2124 DF,  p-value: 0.0001746
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ AGE + HOMEKIDS + HOME_VAL + MSTATUS +
##     SEX + EDUCATION + JOB + CAR_USE + BLUEBOOK + CAR_TYPE + REVOKED +
##     MVR_PTS + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8078  -3178  -1530    459  99524
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -9136.73    4239.16  -2.155   0.0312 *
## AGE                              400.50     831.03   0.482   0.6299
## HOMEKIDS                         189.67     154.61   1.227   0.2200
## HOME_VAL                          59.59      38.12   1.563   0.1181
## MSTATUSYes                     -1006.39     431.00  -2.335   0.0196 *
## SEXM                            1106.98     563.80   1.963   0.0497 *
## EDUCATIONHigh School            -436.94     503.93  -0.867   0.3860
## EDUCATIONLess than High School    51.04     630.04   0.081   0.9354
```

```
## EDUCATIONMasters                511.06     878.58   0.582   0.5608
## EDUCATIONPhD                   1645.52    1083.74   1.518   0.1291
## JOBClerical                     -88.08     576.12  -0.153   0.8785
## JOBDoctor                     -2799.95    1860.77  -1.505   0.1325
## JOBHome Maker                  -279.85     763.59  -0.366   0.7140
## JOBLawyer                      -190.63    1165.80  -0.164   0.8701
## JOBManager                    -1314.95     897.62  -1.465   0.1431
## JOBOther Job                   -510.27    1134.37  -0.450   0.6529
## JOBProfessional                 510.66     680.91   0.750   0.4534
## JOBStudent                      132.06     632.35   0.209   0.8346
## CAR_USEPrivate                 -335.48     518.16  -0.647   0.5174
## BLUEBOOK                       1409.23     325.51   4.329 1.57e-05 ***
## CAR_TYPEPanel Truck             -51.81     875.95  -0.059   0.9528
## CAR_TYPEPickup                 -139.97     594.06  -0.236   0.8138
## CAR_TYPESports Car             1016.08     732.19   1.388   0.1654
## CAR_TYPESUV                     699.27     640.78   1.091   0.2753
## CAR_TYPEVan                     143.98     758.63   0.190   0.8495
## REVOKEDYes                     -765.21     413.21  -1.852   0.0642 .
## MVR_PTS                         120.13      64.99   1.848   0.0647 .
## CAR_AGE                        -374.75     261.96  -1.431   0.1527
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7679 on 2125 degrees of freedom
## Multiple R-squared:  0.02883,    Adjusted R-squared:  0.01649
## F-statistic: 2.337 on 27 and 2125 DF,  p-value: 0.0001215
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ HOMEKIDS + HOME_VAL + MSTATUS + SEX +
##     EDUCATION + JOB + CAR_USE + BLUEBOOK + CAR_TYPE + REVOKED +
##     MVR_PTS + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8151  -3184  -1523    459  99553
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -7876.66    3336.17  -2.361   0.0183 *
## HOMEKIDS                        160.41     142.16   1.128   0.2593
## HOME_VAL                         60.39      38.08   1.586   0.1129
## MSTATUSYes                     -987.07     429.06  -2.301   0.0215 *
## SEXM                           1132.80     561.15   2.019   0.0436 *
## EDUCATIONHigh School           -436.65     503.84  -0.867   0.3862
## EDUCATIONLess than High School   58.18     629.75   0.092   0.9264
## EDUCATIONMasters                528.50     877.68   0.602   0.5471
## EDUCATIONPhD                   1672.96    1082.05   1.546   0.1222
## JOBClerical                    -107.99     574.53  -0.188   0.8509
## JOBDoctor                     -2761.44    1858.72  -1.486   0.1375
## JOBHome Maker                  -263.45     762.69  -0.345   0.7298
## JOBLawyer                      -163.91    1164.27  -0.141   0.8881
## JOBManager                    -1310.21     897.41  -1.460   0.1444
## JOBOther Job                   -506.24    1134.14  -0.446   0.6554
```

```
## JOBProfessional                  522.82     680.32   0.768   0.4423
## JOBStudent                       129.66     632.22   0.205   0.8375
## CAR_USEPrivate                  -331.96     518.02  -0.641   0.5217
## BLUEBOOK                        1432.35     321.90   4.450 9.04e-06 ***
## CAR_TYPEPanel Truck              -68.44     875.11  -0.078   0.9377
## CAR_TYPEPickup                  -139.29     593.95  -0.235   0.8146
## CAR_TYPESports Car              1045.93     729.43   1.434   0.1517
## CAR_TYPESUV                      731.48     637.17   1.148   0.2511
## CAR_TYPEVan                      139.25     758.43   0.184   0.8543
## REVOKEDYes                      -757.71     412.84  -1.835   0.0666 .
## MVR_PTS                          119.52      64.97   1.840   0.0660 .
## CAR_AGE                         -374.56     261.91  -1.430   0.1528
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7678 on 2126 degrees of freedom
## Multiple R-squared:  0.02873,    Adjusted R-squared:  0.01685
## F-statistic: 2.419 on 26 and 2126 DF,  p-value: 8.117e-05
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ HOMEKIDS + HOME_VAL + MSTATUS + SEX +
##     EDUCATION + JOB + BLUEBOOK + CAR_TYPE + REVOKED + MVR_PTS +
##     CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8303  -3189  -1522    430  99678
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -8109.987   3315.783  -2.446   0.0145 *
## HOMEKIDS                         156.882    142.037   1.105   0.2695
## HOME_VAL                          61.055     38.057   1.604   0.1088
## MSTATUSYes                      -991.309    428.947  -2.311   0.0209 *
## SEXM                            1125.415    560.949   2.006   0.0450 *
## EDUCATIONHigh School            -433.541    503.748  -0.861   0.3895
## EDUCATIONLess than High School   -39.142    611.076  -0.064   0.9489
## EDUCATIONMasters                 528.314    877.554   0.602   0.5472
## EDUCATIONPhD                    1680.341   1081.838   1.553   0.1205
## JOBClerical                     -274.495    512.351  -0.536   0.5922
## JOBDoctor                      -3026.832   1811.747  -1.671   0.0949 .
## JOBHome Maker                   -452.065    703.509  -0.643   0.5206
## JOBLawyer                       -419.505   1093.662  -0.384   0.7013
## JOBManager                     -1497.993    848.098  -1.766   0.0775 .
## JOBOther Job                    -596.233   1125.255  -0.530   0.5963
## JOBProfessional                  348.994    623.820   0.559   0.5759
## JOBStudent                        80.140    627.392   0.128   0.8984
## BLUEBOOK                        1446.868    321.058   4.507 6.95e-06 ***
## CAR_TYPEPanel Truck              121.834    823.083   0.148   0.8823
## CAR_TYPEPickup                    -7.405    557.076  -0.013   0.9894
## CAR_TYPESports Car              1029.205    728.861   1.412   0.1581
## CAR_TYPESUV                      723.797    636.964   1.136   0.2559
## CAR_TYPEVan                      263.545    733.104   0.359   0.7193
```

```
## REVOKEDYes                      -747.782    412.490  -1.813   0.0700 .
## MVR_PTS                          122.581     64.785   1.892   0.0586 .
## CAR_AGE                         -376.213    261.859  -1.437   0.1509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7677 on 2127 degrees of freedom
## Multiple R-squared:  0.02854,    Adjusted R-squared:  0.01712
## F-statistic:   2.5 on 25 and 2127 DF,  p-value: 5.655e-05

##
## Call:
## lm(formula = TARGET_AMT ~ HOMEKIDS + HOME_VAL + MSTATUS + SEX +
##     EDUCATION + BLUEBOOK + CAR_TYPE + REVOKED + MVR_PTS + CAR_AGE,
##     data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -8096  -3207  -1527    378 100059
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -8094.118   3197.807  -2.531   0.0114 *
## HOMEKIDS                        142.794    141.266   1.011   0.3122
## HOME_VAL                         55.348     34.571   1.601   0.1095
## MSTATUSYes                     -910.378    410.811  -2.216   0.0268 *
## SEXM                           1133.015    552.890   2.049   0.0406 *
## EDUCATIONHigh School           -427.565    474.626  -0.901   0.3678
## EDUCATIONLess than High School  -70.085    567.078  -0.124   0.9017
## EDUCATIONMasters                 29.144    556.698   0.052   0.9583
## EDUCATIONPhD                    552.367    780.792   0.707   0.4794
## BLUEBOOK                       1433.180    313.328   4.574 5.06e-06 ***
## CAR_TYPEPanel Truck             245.320    787.406   0.312   0.7554
## CAR_TYPEPickup                   -0.581    554.157  -0.001   0.9992
## CAR_TYPESports Car              952.486    727.235   1.310   0.1904
## CAR_TYPESUV                     664.736    635.728   1.046   0.2959
## CAR_TYPEVan                     281.676    720.588   0.391   0.6959
## REVOKEDYes                     -681.358    411.172  -1.657   0.0976 .
## MVR_PTS                         127.543     64.525   1.977   0.0482 *
## CAR_AGE                        -365.036    261.332  -1.397   0.1626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7675 on 2135 degrees of freedom
## Multiple R-squared:  0.02527,    Adjusted R-squared:  0.01751
## F-statistic: 3.256 on 17 and 2135 DF,  p-value: 7.297e-06

##
## Call:
## lm(formula = TARGET_AMT ~ HOMEKIDS + HOME_VAL + MSTATUS + SEX +
##     BLUEBOOK + CAR_TYPE + REVOKED + MVR_PTS + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7893  -3212  -1557    410 100200
```

```
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -8654.216   3095.823  -2.795  0.00523 **
## HOMEKIDS              135.427    140.861   0.961  0.33645
## HOME_VAL               58.351     34.436   1.694  0.09032 .
## MSTATUSYes           -963.913    407.931  -2.363  0.01822 *
## SEXM                 1116.708    552.118   2.023  0.04324 *
## BLUEBOOK             1452.995    311.031   4.672 3.18e-06 ***
## CAR_TYPEPanel Truck   314.423    783.769   0.401  0.68834
## CAR_TYPEPickup         -2.979    553.431  -0.005  0.99571
## CAR_TYPESports Car    959.028    725.935   1.321  0.18661
## CAR_TYPESUV           638.714    634.744   1.006  0.31441
## CAR_TYPEVan           336.811    717.794   0.469  0.63895
## REVOKEDYes           -697.721    410.676  -1.699  0.08947 .
## MVR_PTS               129.059     64.464   2.002  0.04541 *
## CAR_AGE              -226.895    209.010  -1.086  0.27779
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7671 on 2139 degrees of freedom
## Multiple R-squared:  0.02439,    Adjusted R-squared:  0.01846
## F-statistic: 4.114 on 13 and 2139 DF,  p-value: 9.195e-07
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ HOMEKIDS + HOME_VAL + MSTATUS + SEX +
##     BLUEBOOK + REVOKED + MVR_PTS + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7506  -3167  -1547    392 100397
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7682.75    2396.42  -3.206  0.00137 **
## HOMEKIDS      126.65     140.53   0.901  0.36755
## HOME_VAL       59.05      34.40   1.717  0.08621 .
## MSTATUSYes   -948.00     407.02  -2.329  0.01995 *
## SEXM          666.22     335.54   1.986  0.04721 *
## BLUEBOOK     1410.39     255.13   5.528 3.63e-08 ***
## REVOKEDYes   -695.80     409.88  -1.698  0.08973 .
## MVR_PTS       128.90      64.30   2.005  0.04512 *
## CAR_AGE      -217.32     208.65  -1.042  0.29775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7667 on 2144 degrees of freedom
## Multiple R-squared:  0.02321,    Adjusted R-squared:  0.01957
## F-statistic: 6.369 on 8 and 2144 DF,  p-value: 3.381e-08
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ HOME_VAL + MSTATUS + SEX + BLUEBOOK +
##     REVOKED + MVR_PTS + CAR_AGE, data = mlr_crash_transf)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7364  -3150  -1572    412 100285
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7400.44    2375.76  -3.115  0.00186 **
## HOME_VAL       57.02      34.32   1.661  0.09682 .
## MSTATUSYes   -914.64     405.32  -2.257  0.02413 *
## SEXM          637.15     333.97   1.908  0.05655 .
## BLUEBOOK     1395.31     254.57   5.481 4.73e-08 ***
## REVOKEDYes   -677.87     409.37  -1.656  0.09790 .
## MVR_PTS       130.71      64.27   2.034  0.04209 *
## CAR_AGE      -227.51     208.34  -1.092  0.27495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7667 on 2145 degrees of freedom
## Multiple R-squared:  0.02284,    Adjusted R-squared:  0.01966
## F-statistic: 7.164 on 7 and 2145 DF,  p-value: 1.71e-08
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ HOME_VAL + MSTATUS + SEX + BLUEBOOK +
##     REVOKED + MVR_PTS, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7435  -3176  -1595    386 100375
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7489.85    2374.46  -3.154  0.00163 **
## HOME_VAL       55.56      34.30   1.620  0.10540
## MSTATUSYes   -887.80     404.59  -2.194  0.02832 *
## SEXM          653.55     333.65   1.959  0.05026 .
## BLUEBOOK     1358.16     252.30   5.383 8.12e-08 ***
## REVOKEDYes   -682.24     409.37  -1.667  0.09575 .
## MVR_PTS       133.92      64.20   2.086  0.03711 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7667 on 2146 degrees of freedom
## Multiple R-squared:  0.0223, Adjusted R-squared:  0.01957
## F-statistic: 8.158 on 6 and 2146 DF,  p-value: 9.631e-09
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ MSTATUS + SEX + BLUEBOOK + REVOKED +
##     MVR_PTS, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7042  -3176  -1561    401 100457
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7646.12    2373.39  -3.222  0.00129 **
## MSTATUSYes   -510.63     331.01  -1.543  0.12306
## SEXM          652.64     333.77   1.955  0.05067 .
## BLUEBOOK     1400.78     251.02   5.580  2.7e-08 ***
## REVOKEDYes   -710.83     409.15  -1.737  0.08247 .
## MVR_PTS       128.56      64.14   2.004  0.04516 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7670 on 2147 degrees of freedom
## Multiple R-squared:  0.02111,    Adjusted R-squared:  0.01883
## F-statistic: 9.258 on 5 and 2147 DF,  p-value: 9.836e-09
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ SEX + BLUEBOOK + REVOKED + MVR_PTS,
##     data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7317  -3180  -1617    423 100195
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8002.80    2362.86  -3.387  0.00072 ***
## SEXM          645.74     333.85   1.934  0.05322 .
## BLUEBOOK     1411.85     251.00   5.625  2.1e-08 ***
## REVOKEDYes   -690.94     409.08  -1.689  0.09136 .
## MVR_PTS       129.43      64.16   2.017  0.04378 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7672 on 2148 degrees of freedom
## Multiple R-squared:  0.02002,    Adjusted R-squared:  0.0182
## F-statistic: 10.97 on 4 and 2148 DF,  p-value: 8.306e-09
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ SEX + BLUEBOOK + MVR_PTS, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7181  -3173  -1607    348 100329
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8153.34    2362.20  -3.452 0.000568 ***
## SEXM          648.01     333.99   1.940 0.052483 .
## BLUEBOOK     1412.22     251.11   5.624 2.11e-08 ***
## MVR_PTS       131.00      64.18   2.041 0.041360 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 7676 on 2149 degrees of freedom
## Multiple R-squared:  0.01872,    Adjusted R-squared:  0.01735
## F-statistic: 13.66 on 3 and 2149 DF,  p-value: 7.883e-09
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ BLUEBOOK + MVR_PTS, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7511  -3151  -1545    328 100673
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8251.14    2363.18  -3.492  0.00049 ***
## BLUEBOOK     1453.68     250.36   5.806 7.33e-09 ***
## MVR_PTS       130.32      64.22   2.029  0.04256 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7681 on 2150 degrees of freedom
## Multiple R-squared:  0.017,  Adjusted R-squared:  0.01609
## F-statistic: 18.59 on 2 and 2150 DF,  p-value: 9.889e-09
```

Model 4: Forward Elimination

Now let's use forward addition to add of variables one at a time.

```
##
## Call:
## lm(formula = TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7181  -3173  -1607    348 100329
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8153.34    2362.20  -3.452 0.000568 ***
## BLUEBOOK     1412.22     251.11   5.624 2.11e-08 ***
## MVR_PTS       131.00      64.18   2.041 0.041360 *
## SEXM          648.01     333.99   1.940 0.052483 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7676 on 2149 degrees of freedom
## Multiple R-squared:  0.01872,    Adjusted R-squared:  0.01735
## F-statistic: 13.66 on 3 and 2149 DF,  p-value: 7.883e-09
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX + MSTATUS,
##     data = mlr_crash_transf)
##
## Residuals:
```

```
##    Min     1Q Median     3Q    Max
##  -6912  -3152  -1537    329 100585
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7813.51    2372.55  -3.293  0.00101 **
## BLUEBOOK     1401.56     251.14   5.581  2.7e-08 ***
## MVR_PTS       130.20      64.16   2.029  0.04256 *
## SEXM          654.74     333.93   1.961  0.05004 .
## MSTATUSYes   -492.51     331.00  -1.488  0.13691
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7674 on 2148 degrees of freedom
## Multiple R-squared:  0.01973,    Adjusted R-squared:  0.0179
## F-statistic: 10.81 on 4 and 2148 DF,  p-value: 1.127e-08
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX + MSTATUS +
##     HOME_VAL, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7317  -3147  -1567    342 100494
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7643.27    2373.65  -3.220   0.0013 **
## BLUEBOOK     1357.01     252.40   5.376 8.43e-08 ***
## MVR_PTS       135.73      64.22   2.113   0.0347 *
## SEXM          655.60     333.78   1.964   0.0496 *
## MSTATUSYes   -887.17     404.76  -2.192   0.0285 *
## HOME_VAL       58.03      34.28   1.693   0.0907 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7670 on 2147 degrees of freedom
## Multiple R-squared:  0.02104,    Adjusted R-squared:  0.01876
## F-statistic: 9.227 on 5 and 2147 DF,  p-value: 1.057e-08
```

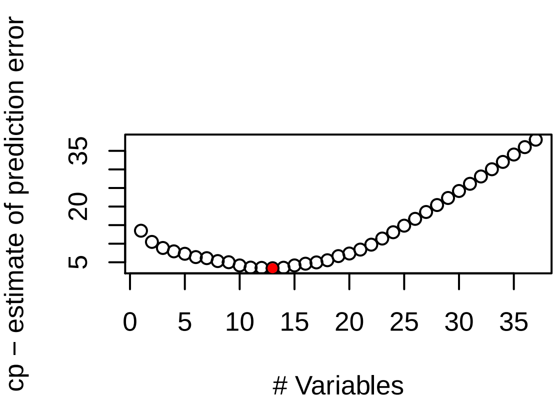
```
##
## Call:
## lm(formula = TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX + MSTATUS +
##     HOME_VAL + REVOKED, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7435  -3176  -1595    386 100375
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7489.85    2374.46  -3.154  0.00163 **
## BLUEBOOK     1358.16     252.30   5.383 8.12e-08 ***
## MVR_PTS       133.92      64.20   2.086  0.03711 *
```

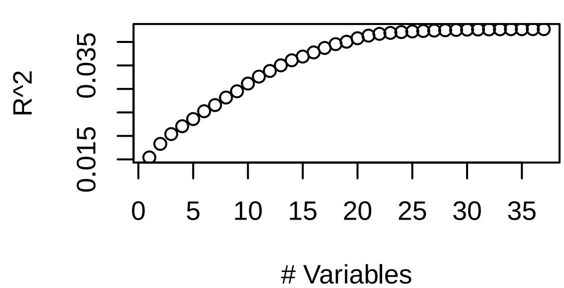
```
## SEXM          653.55     333.65   1.959  0.05026 .
## MSTATUSYes   -887.80     404.59  -2.194  0.02832 *
## HOME_VAL       55.56      34.30   1.620  0.10540
## REVOKEDYes   -682.24     409.37  -1.667  0.09575 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7667 on 2146 degrees of freedom
## Multiple R-squared:  0.0223, Adjusted R-squared:  0.01957
## F-statistic: 8.158 on 6 and 2146 DF,  p-value: 9.631e-09
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX + MSTATUS +
##     HOME_VAL + REVOKED + CAR_AGE, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7364  -3150  -1572    412 100285
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7400.44    2375.76  -3.115  0.00186 **
## BLUEBOOK     1395.31     254.57   5.481 4.73e-08 ***
## MVR_PTS       130.71      64.27   2.034  0.04209 *
## SEXM          637.15     333.97   1.908  0.05655 .
## MSTATUSYes   -914.64     405.32  -2.257  0.02413 *
## HOME_VAL       57.02      34.32   1.661  0.09682 .
## REVOKEDYes   -677.87     409.37  -1.656  0.09790 .
## CAR_AGE      -227.51     208.34  -1.092  0.27495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7667 on 2145 degrees of freedom
## Multiple R-squared:  0.02284,    Adjusted R-squared:  0.01966
## F-statistic: 7.164 on 7 and 2145 DF,  p-value: 1.71e-08
```

Model 5: Picking the best model using Leaps

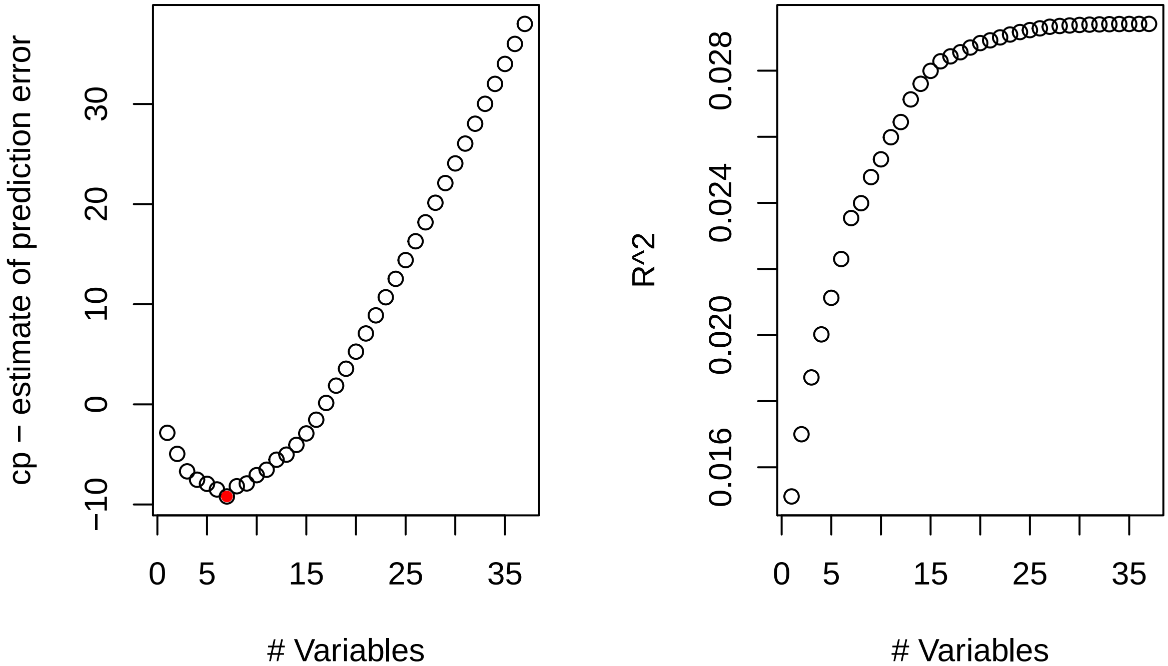
The function, *regsubsets()*, will go through iterations to find the best model using parameters = 1,2,3,4,… n. Here we see the model with 13 variables (represented by the red dot) had the lowest cp, which indicates the best model. The R^2 remains to be around 3.5% from about 13 variables and higher, which is extremely low.





Model 6:

Using the regsubsets function and our data that includes log transformations, we see it suggests a model with 7 variables is best look at the cp value.



Using the transformed variables, we will choose the model that has 7 parameters since the R^2 value doesn't change by much as the number of parameters increases. This gives us the following equation:

```
##     (Intercept)      MSTATUSYes    EDUCATIONPhD       JOBDoctor      JOBManager
##    4857.7855103    -866.2249453    2008.6181953   -3283.3214513   -1358.0216839
## JOBProfessional        BLUEBOOK         CAR_AGE
##    1083.6185705       0.1127877     -67.5694404
```

```
##
## Call:
## lm(formula = TARGET_AMT ~ MSTATUS + JOB + BLUEBOOK + CAR_AGE +
##     EDUCATION, data = mlr_crash_transf)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -7308  -3123  -1531    374 100678
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       -5467.5     2656.6  -2.058   0.0397 *
## MSTATUSYes                         -491.1      334.2  -1.470   0.1418
## JOBClerical                        -306.4      510.7  -0.600   0.5486
## JOBDoctor                         -2863.7     1806.9  -1.585   0.1131
## JOBHome Maker                      -710.4      681.5  -1.042   0.2973
## JOBLawyer                          -605.8     1087.2  -0.557   0.5774
## JOBManager                        -1531.3      845.0  -1.812   0.0701 .
## JOBOther Job                       -449.7     1104.0  -0.407   0.6838
## JOBProfessional                     316.3      622.3   0.508   0.6112
## JOBStudent                         -279.7      573.6  -0.488   0.6258
```

```
## BLUEBOOK                          1342.2      268.7   4.996 6.33e-07 ***
## CAR_AGE                           -439.1      261.4  -1.680   0.0932 .
## EDUCATIONHigh School              -539.8      502.5  -1.074   0.2829
## EDUCATIONLess than High School   -116.7      609.6  -0.191   0.8482
## EDUCATIONMasters                   534.5      877.3   0.609   0.5424
## EDUCATIONPhD                      1618.9     1080.7   1.498   0.1343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7687 on 2137 degrees of freedom
## Multiple R-squared:  0.02142,    Adjusted R-squared:  0.01455
## F-statistic: 3.118 on 15 and 2137 DF,  p-value: 4.575e-05
```

Model 7

For this model, we used the log transformation of the response variable and a combination of predictors. Here is the model that yielded the best results:

```
##
## Call:
## lm(formula = log(TARGET_AMT) ~ MSTATUS + SEX + BLUEBOOK + CLM_FREQ +
##     MVR_PTS + EDUCATION, data = mlr_crash_transf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7062 -0.4084  0.0422  0.4048  3.2688
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     6.78059    0.25943  26.136  < 2e-16 ***
## MSTATUSYes                     -0.07614    0.03488  -2.183   0.0292 *
## SEXM                            0.05556    0.03503   1.586   0.1128
## BLUEBOOK                        0.15326    0.02712   5.652  1.8e-08 ***
## CLM_FREQ                       -0.02297    0.01457  -1.577   0.1150
## MVR_PTS                         0.01766    0.00705   2.505   0.0123 *
## EDUCATIONHigh School            0.06214    0.04575   1.358   0.1745
## EDUCATIONLess than High School  0.06322    0.05455   1.159   0.2466
## EDUCATIONMasters                0.08379    0.05693   1.472   0.1412
## EDUCATIONPhD                    0.13885    0.08042   1.726   0.0844 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.804 on 2143 degrees of freedom
## Multiple R-squared:  0.0251, Adjusted R-squared:  0.02101
## F-statistic: 6.131 on 9 and 2143 DF,  p-value: 1.473e-08
```

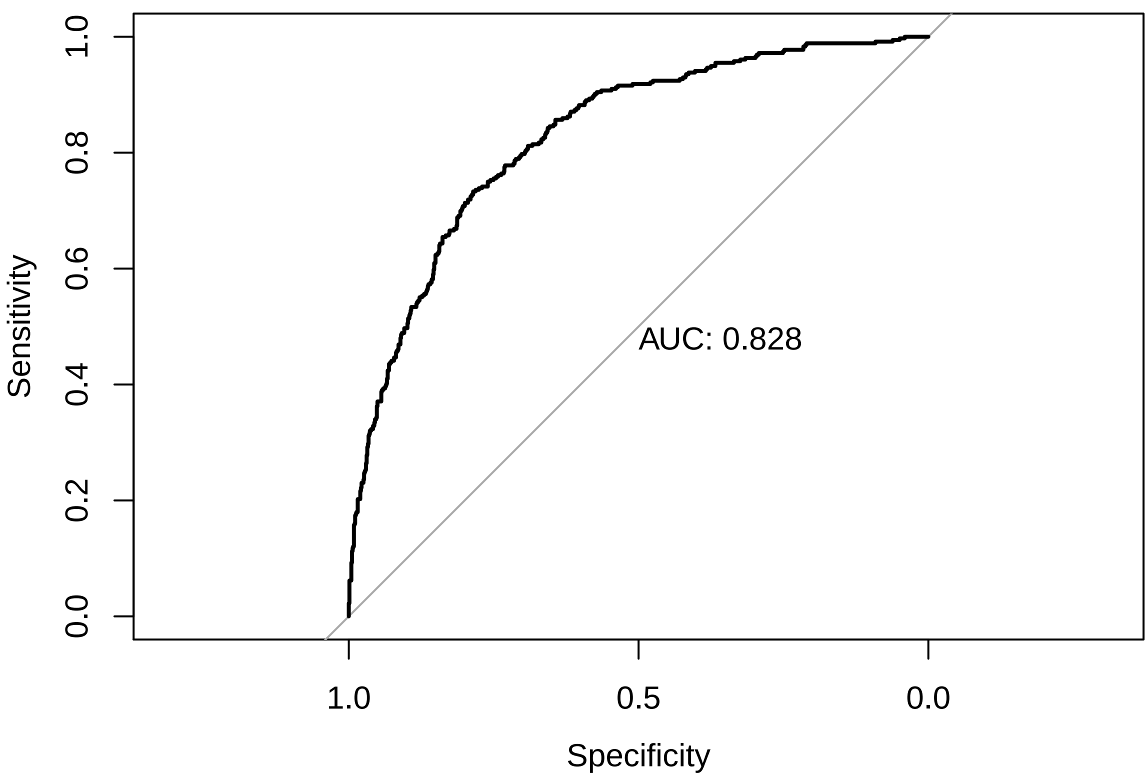
Select Models & Prediction

Binary Logistic Regression

Based on the peformance diagnostics, model 4 or our binned model performs the best. AIC is 5816 and here are the other performance diagnostics:

```
## Confusion Matrix and Statistics
##
```

```
##           Reference
## Prediction   0   1
##          0 880 195
##          1  85 134
##
##                Accuracy : 0.7836
##                  95% CI : (0.7602, 0.8058)
##     No Information Rate : 0.7457
##     P-Value [Acc > NIR] : 0.0008298
##
##                   Kappa : 0.3587
##
##  Mcnemar's Test P-Value : 7.318e-11
##
##             Sensitivity : 0.9119
##             Specificity : 0.4073
##          Pos Pred Value : 0.8186
##          Neg Pred Value : 0.6119
##              Prevalence : 0.7457
##          Detection Rate : 0.6801
##    Detection Prevalence : 0.8308
##       Balanced Accuracy : 0.6596
##
##        'Positive' Class : 0
##
```

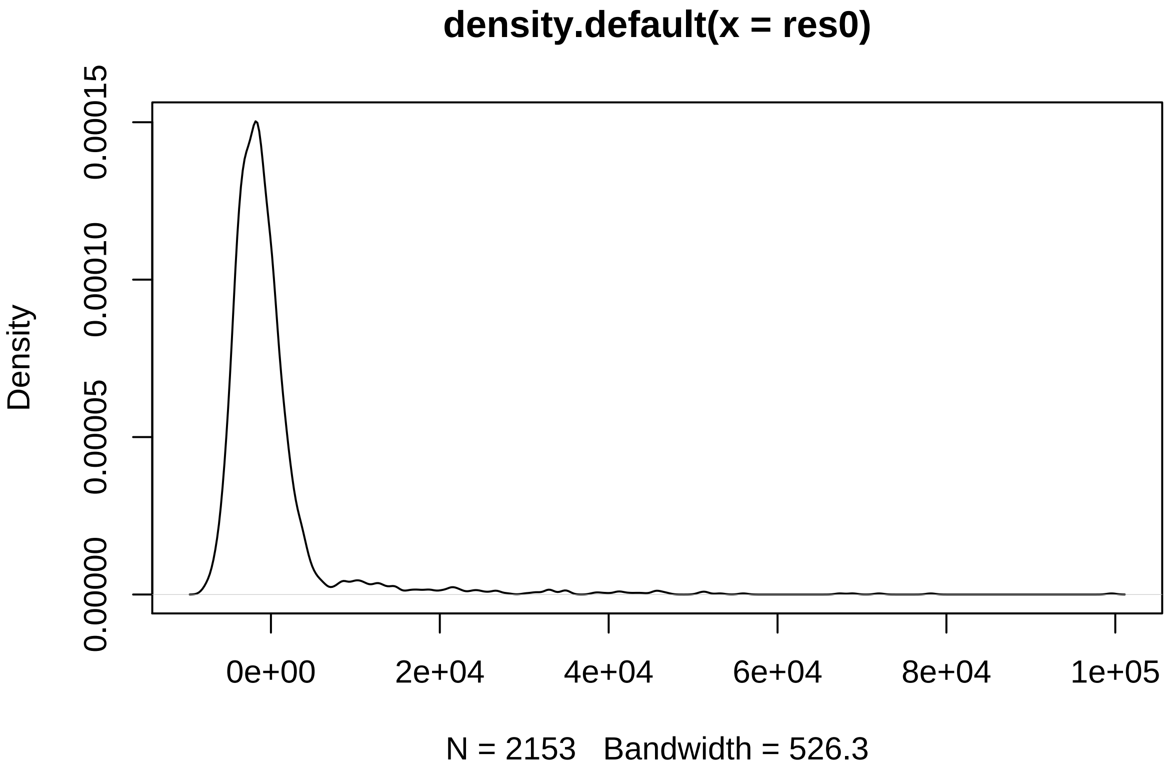


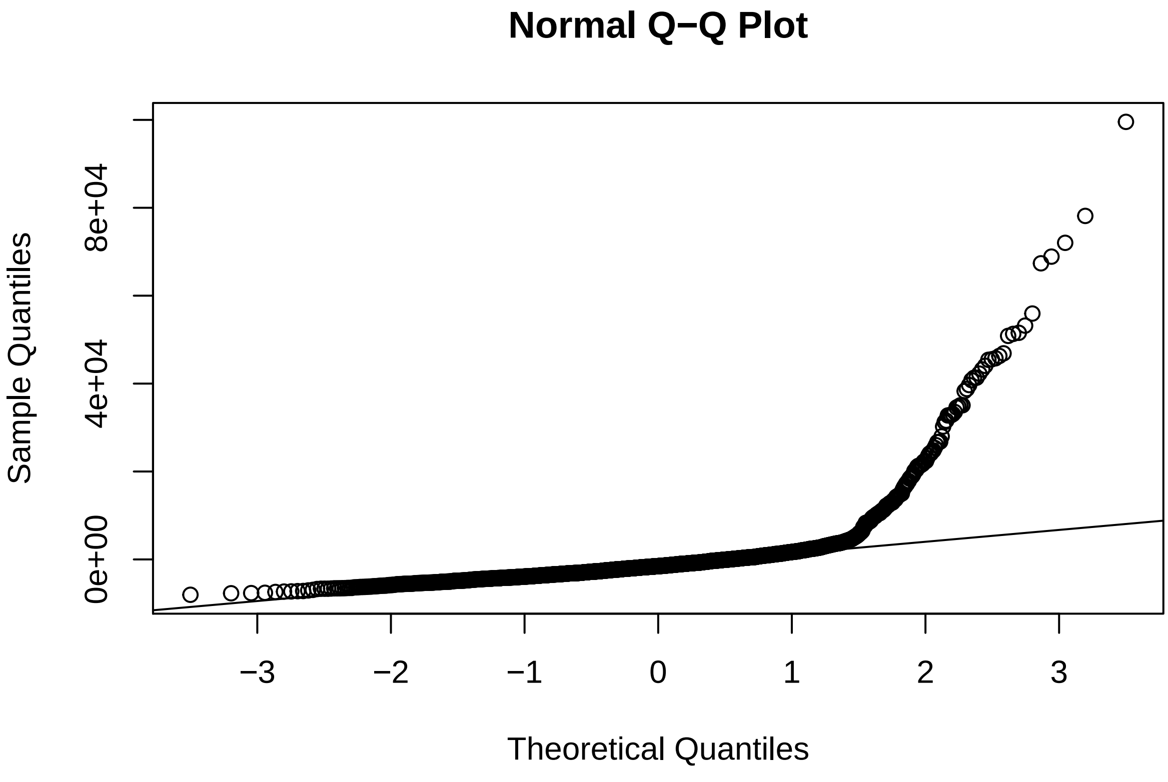
Multiple Linear Regression

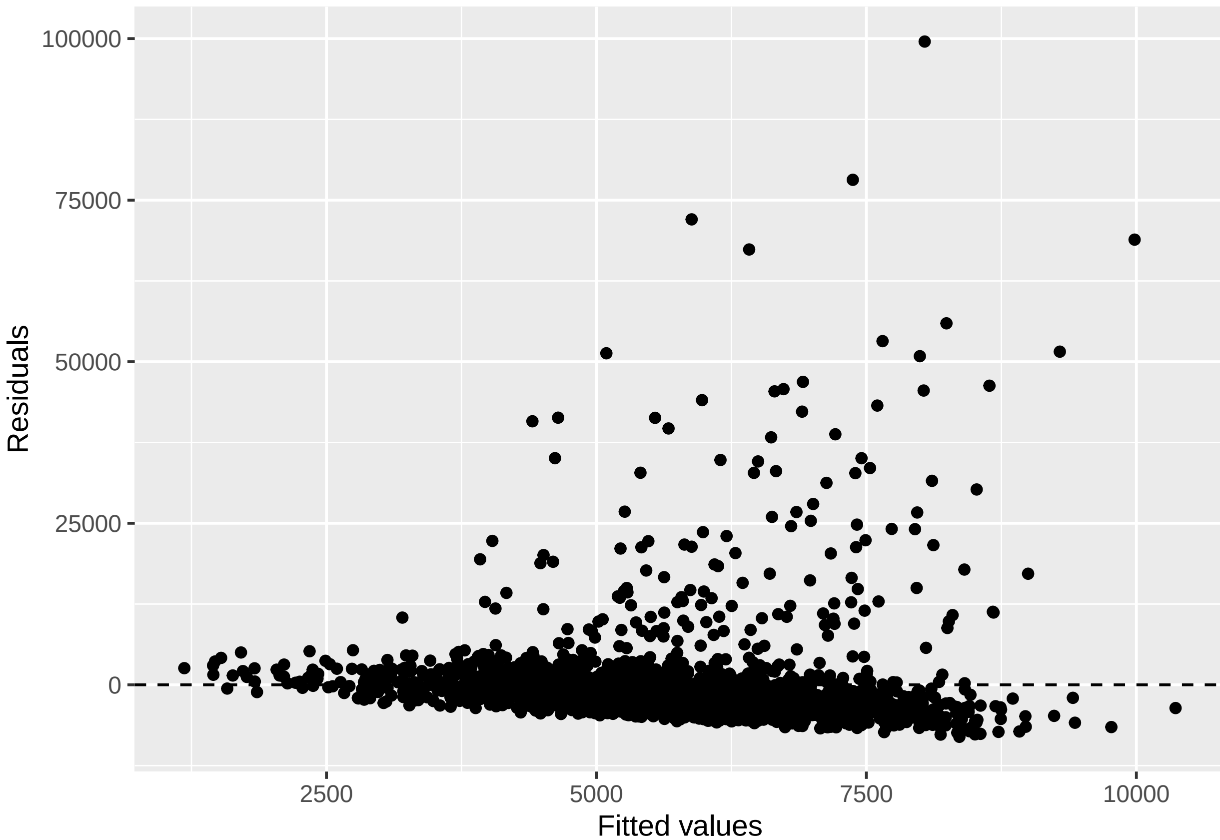
We will look at the diagnostic plot for the two models that had the highest adjusted r^2. Particularly model 1(with all variables minus TARGET_FLAG) and model 7 (log of response variable and a combination of predictors).

Model 1

Model 1 had an adjusted r^2 of 0.02145 and is significant. Here is the diagnostic plot for model 1



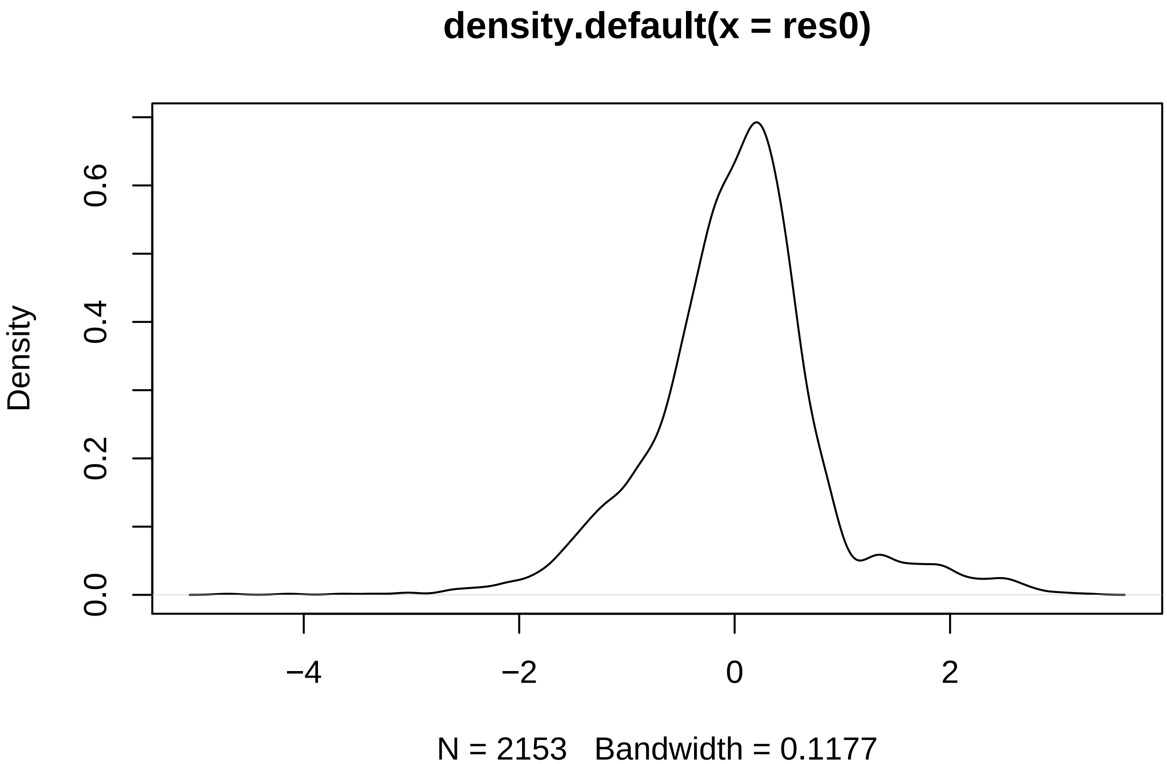


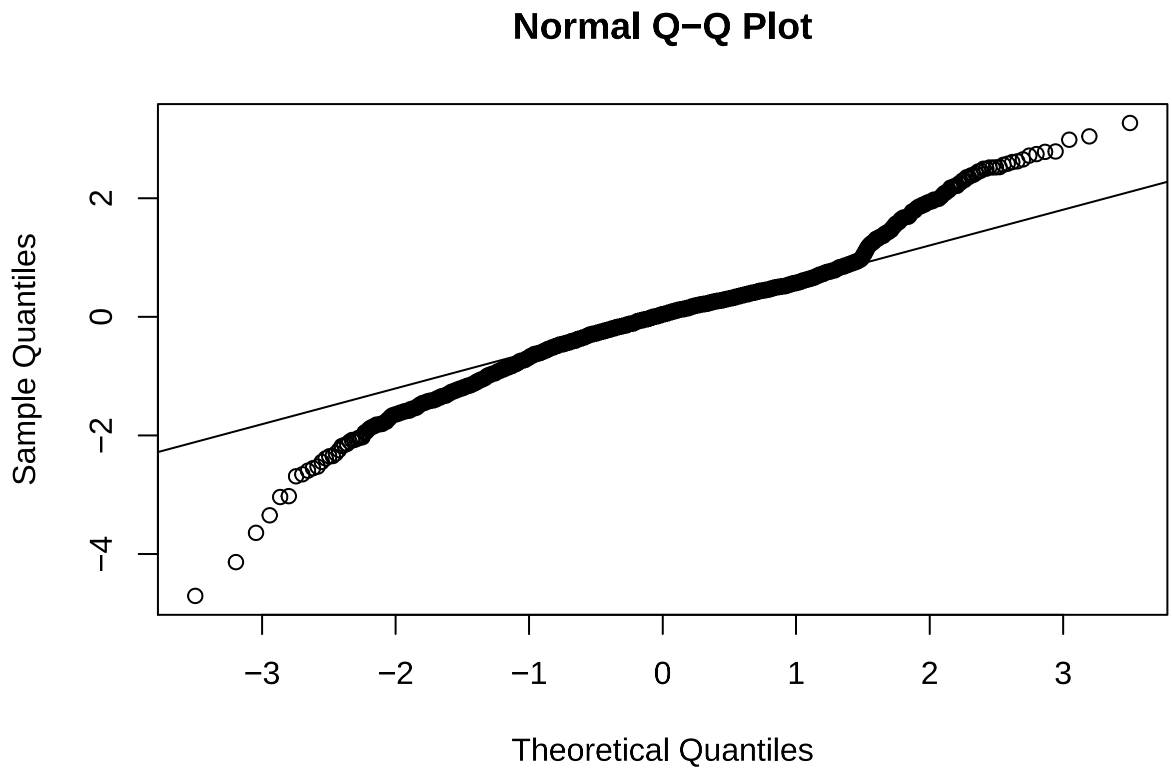


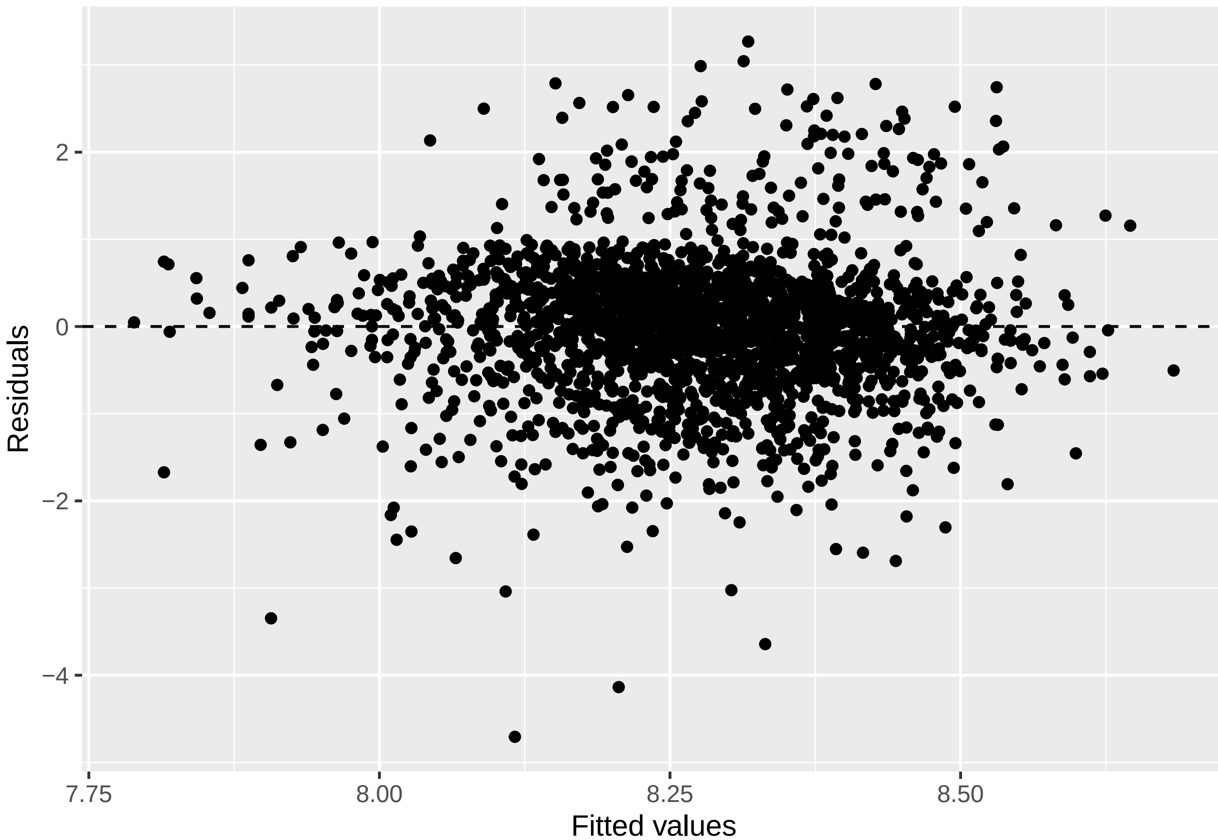
The density plot seems skewed and the qq plot deviates quite a bit.

Model 7

Model 7 had an adjusted r^2 of 0.02158 and is significant


density.default(x = res0)
Density
0.0
0.2
0.4
0.6
−4
−2
0
2
N = 2153 Bandwidth = 0.1177


Normal Q−Q Plot
Sample Quantiles
−4
−2
0
2
−3
−2
−1
0
1
2
3
Theoretical Quantiles



The density and qqplot for model 7 seem somewhat normally distributed. The residual plot indicates homoscedasticity.

Prediction

```
## predicted_flag_bin
##    0    1
## 5337 1111
```

```
## predicted_amt2
##                0 236.563937331378 236.583059129324 236.586911059253
##             7050                1                1                1
## 236.588374348008 236.618567800217 236.639886586829 236.650517024196
##                1                1                1                1
## 236.666228109109 236.680518823897 236.693942533297 236.694517055668
##                1                1                1                1
## 236.709888189348  236.71217556486  236.73197315084 236.733494351473
##                1                1                1                1
## 236.739001222665 236.746711192303 236.768811369856 236.782809601628
##                1                1                1                1
## 236.786029097308 236.793057169133 236.795152892136 236.813629150445
##                1                1                1                1
## 236.815397804792 236.853934318856 236.859249537539 279.623178522203
##                1                1                1                1
## 305.746405082491 324.062675345335 342.466540605108 365.482545811332
##                1                1                1                1
## 380.696781460324    386.850134888 416.101141301053 417.654884091719
##                1                1                1                1
```

```
## 428.276728495385 454.327928747418 498.729564409365 532.402168591273
##                1                1                1                1
## 549.358172460297 552.447608109474 561.498115567589 564.841591548708
##                1                1                1                1
## 581.662040648804 589.206087835678 593.797036191213 604.602223795272
##                1                1                1                1
## 612.095641399929 616.770647501632 619.976025006175 621.326527968014
##                1                1                1                1
## 627.498811976283 635.087741416078 635.097988983583 635.128373653145
##                1                1                1                1
## 638.214441507918 650.492859416711 650.592606486795 665.828997663872
##                1                1                1                1
## 665.858552136897 665.859632555789 666.005199885929 668.878874949939
##                1                1                1                1
## 672.013618417714 679.506844369862 696.418555724399 696.487285255517
##                1                1                1                1
## 696.510450635901 696.521528399944  696.56621943551 711.698116301255
##                1                1                1                1
## 711.767478701776 711.773049594627 711.829819140254 711.836399885403
##                1                1                1                1
## 711.846014360691 711.857234536491 713.396267794833 717.881088658769
##                1                1                1                1
## 727.098551517454 727.187035715922 727.243678065853 734.899133938319
##                1                1                1                1
## 739.413558516032  740.93791775457 742.556011953446 742.561518824638
##                1                1                1                1
##  757.78595868735 757.791082253525 757.793668656989 757.914883470377
##                1                2                1                1
## 762.375507775788 765.448741659593  766.98923820669 770.132090165183
##                1                1                1                1
## 773.106592718168 773.140196883411 773.223174210687 776.218177350495
##                1                1                1                1
## 776.310321826384 782.380506274276 788.480884174377  788.53254536989
##                1                1                1                1
## 788.543360464157 791.423245850874 793.193303017467  803.69298074658
##                1                1                1                1
## 803.705324037088 803.772596824387 803.834440472624 803.896398041239
##                1                1                1                1
## 812.866656987298 812.988745351009  814.48134047314 816.065198072525
##                1                1                1                1
## 819.059562233149  819.11912505081 819.122152236013 819.150632835081
##                1                1                1                1
## 819.173606562957 819.184628525794 819.238733728532 823.848861793891
##                1                1                1                1
## 825.179311495583 829.876019253761 834.394194495739 834.399509714423
##                1                1                1                1
## 834.404882844985 834.481023752944 834.492734609204 834.499762681029
##                1                1                1                1
## 835.973875207211 839.082283697642 843.681908521329 845.110015679882
##                1                1                1                1
## 849.771464331554 849.820189843277 851.348932838298 854.421150760023
##                1                1                1                1
## 855.917981339914 858.965512680623 862.135445971087 864.976980158608
##                1                1                1                1
```

```
##  865.10127205592 865.115263742756 865.121467727824 865.121652384724
##                1                1                1                1
## 865.131209398807 865.131273855621 868.140503253403 868.182848343612
##                1                1                1                1
##  877.41811866818 880.435073686818 880.454238180581 880.537215507857
##                1                1                1                1
## 885.004761167618  885.05494342316  886.53773896243 889.759126579838
##                1                1                1                1
##  895.68736105638 895.703306712431 895.713995061676 895.731653570868
##                1                1                1                1
## 895.735320393224 895.743364427128 895.744870411699 895.764241774032
##                1                1                1                1
## 895.808683245212 895.811461301184 895.836530752063 897.290000279666
##                1                1                1                1
## 898.848297747712 898.867669110045 901.718626120345 903.305070123194
##                1                1                1                1
## 903.364818048427 903.481399540946  910.88950908903 911.098235057438
##                1                1                2                1
## 911.104822347522 911.106328332093 911.157797875098 914.191905506406
##                1                1                1                1
## 915.694563088469 918.713549373296 918.717216195652 924.867094453481
##                1                1                1                1
## 926.250775356916 926.286284027808 926.307602814421 926.418869088256
##                1                1                1                1
## 926.446028359652 926.455251309471 926.469002087528 926.479746237979
##                2                1                1                1
## 926.486333528063 926.501590290692 926.505640433583 938.632159831474
##                1                1                1                1
## 941.650136699158 941.659893586203  953.97557593766 957.005640421685
##                1                1                1                1
## 957.051652334249 957.106084605644 957.123302333096 957.202470426259
##                1                1                1                1
##  958.56120907451 959.966400417066 963.199626086428 967.824683845759
##                1                1                1                1
## 972.301396218978 972.418616690681  972.45247042031 976.985526115508
##                1                1                1                1
## 984.687037254324 986.274874232725 987.592475776771 987.729438033248
##                1                1                1                1
## 987.729636246211 987.737595780236 987.776593807652 987.780700707683
##                1                1                1                1
## 987.783543235987 987.783734888496 989.143994737379 989.178679321514
##                1                1                1                1
## 989.291643232429 990.720888721753 992.245964894418  995.38949918588
##                1                1                1                1
## 999.937037248375 1002.90734726223 1003.01348996989 1003.08076275719
##                1                1                1                1
## 1003.11367802901 1003.16905530315 1004.55140011352 1006.23847995284
##                1                1                1                1
## 1009.21537725678 1010.73090341097 1013.75101935544 1018.28222889203
##                1                1                1                1
## 1018.31627383901 1020.02651905871 1021.59551142094 1023.03722739646
##                1                1                1                1
## 1024.39001773381  1024.4477339577   1026.144555468  1029.0835493501
##                1                1                1                1
```

```
## 1029.14886816818 1032.07259829204 1032.08659652381 1033.60583219617
##                1                1                1                1
## 1033.66412294243 1033.67070368758 1036.75355204667 1036.76589533718
##                1                1                1                1
## 1046.01702681792 1047.37557381366 1047.44108428426 1047.49342737758
##                1                1                1                1
##  1049.0007026293 1049.03785406294  1053.6050681493 1053.62551993052
##                1                1                1                1
##  1056.6743656033 1064.25817213488 1064.28343323827 1075.16461933853
##                1                1                1                1
## 1076.67120635479 1079.62265724048 1079.68367680196 1082.67018858119
##                1                1                1                1
## 1084.31292010481 1085.64026468188 1085.67596457012  1088.9081742774
##                1                1                1                1
## 1090.19780008952 1090.39696082339 1094.90899021499 1095.09108217811
##                1                1                1                1
## 1099.69145335642   1101.081715005 1107.27697274115 1110.23557844368
##                1                1                1                1
## 1110.26659620546   1110.344557342 1111.74792845597 1114.93355172832
##                1                1                1                1
## 1120.90818098729 1122.48506842673 1122.52589231631  1124.0817166433
##                1                1                1                1
## 1125.63152275077 1125.64772451614 1127.22797974998 1128.59778961734
##                1                1                1                1
## 1133.28221854182 1133.33399993742 1134.87899395409 1136.41101435667
##                1                1                1                1
## 1137.88696693168 1137.90589051667 1139.38001003846 1142.57031543727
##                1                1                1                1
## 1147.06214915697 1147.09335857126 1147.14639855565 1148.65468976945
##                1                1                1                1
## 1150.24963423915 1156.22348210718 1156.26152794079 1160.83633898776
##                1                1                1                1
## 1170.03678231644 1171.67887486088 1171.72620810111 1174.75330287666
##                1                1                1                1
## 1180.78215797361 1186.92304091508 1186.94785469179  1190.1092387098
##                1                1                1                1
## 1191.57437573988 1196.22045391584 1199.22323841977 1202.23406047061
##                1                1                1                1
## 1203.72322988646 1203.77387511989  1211.4651386912 1211.49234721887
##                1                1                1                1
## 1214.55038180125 1217.52196388651  1217.5342427202 1217.68512526903
##                1                1                1                1
##  1219.0521640605 1220.71772691576 1225.27615152495  1225.2963964376
##                1                1                1                1
##  1226.7653209408 1228.32076239793 1232.79658600476  1235.9608904905
##                1                1                1                1
## 1237.37139705174 1242.09366494125 1243.60197180627 1245.24310418983
##                1                1                1                1
## 1246.68691588836 1246.71354989366 1246.74734571141 1249.87620598306
##                1                1                1                1
## 1252.99710017576 1254.31988871341 1255.91994153323 1262.07608823294
##                1                1                1                1
## 1269.75651449832 1271.15848700316 1272.66488870455 1272.82690655788
##                1                1                1                1
```

```
## 1274.26240436166 1274.35874094153 1277.38842212054 1280.36127627719
##                1                1                1                1
## 1280.36531942448 1283.46417789643 1289.58557555519 1291.08543986522
##                1                1                1                1
## 1295.79038330981 1295.81765629429 1297.21488810283 1297.30239925784
##                1                1                1                1
## 1301.82330331504 1301.89614088541 1303.42214852028 1303.45746553866
##                1                2                1                1
## 1303.48912227962 1306.50133660195 1311.00487463297  1311.1270274535
##                1                1                1                1
## 1311.20829192763 1314.11607639544 1314.18139521352 1320.27829794378
##                1                1                1                1
## 1321.78177983541 1326.47931168034 1327.90684431653 1327.98190480559
##                1                1                1                1
## 1328.06683610208 1329.40031298898 1331.06031149632 1331.07538360205
##                1                1                1                1
## 1334.06701894746 1340.20993359027 1341.61082567623 1343.19000703611
##                1                1                1                1
## 1344.89873972631 1349.58740345058  1350.8606428249 1350.97853864949
##                1                1                1                1
## 1350.99166333093 1351.05467955745  1353.8956392226 1353.99246582495
##                1                1                1                1
##  1355.6348058231 1357.12418210752 1358.49571137296 1358.49956330289
##                1                1                1                1
## 1360.06893853498 1361.69367793581 1361.71220999533 1363.21100910253
##                1                1                1                1
## 1364.71950718489 1366.21760284058 1370.86773704639 1372.37765727142
##                1                1                1                1
## 1373.78615195157 1375.42152178977 1376.94880149604  1376.9513878995
##                1                1                1                1
## 1386.25822372304 1387.72183955249 1393.89582986896 1396.93980830769
##                1                1                1                1
## 1406.07892715997 1406.10873730716 1409.14729315177 1412.27443402535
##                1                1                1                1
## 1412.33399684301 1413.78513564131 1416.88734014668 1416.89538418058
##                1                1                1                1
## 1417.04031907647  1418.4084805475 1419.95892563415 1421.43487820917
##                1                1                1                1
## 1422.99620940736 1426.04316622569 1426.04329342139 1426.20006007769
##                1                1                1                1
## 1427.49608716096 1428.98747532647 1430.60215881231 1430.61006043446
##                1                1                1                1
## 1433.73974501569 1436.73400180087 1436.77253831493 1438.31455606266
##                1                1                1                1
## 1441.22949579799 1441.28247132556 1441.53075565708 1442.88417908665
##                1                1                1                1
## 1444.27171192001 1446.00493086759 1447.48183666581 1450.53652327403
##                1                2                1                1
## 1456.61905101243  1458.2436053057 1459.76448937439 1462.73880027487
##                1                1                1                1
## 1469.04683026942 1470.47196815465 1478.16645056367 1479.84139398105
##                1                1                1                1
## 1481.19373699172 1484.18810115235  1484.2374164025 1484.39746609994
##                1                1                1                1
```

```
## 1488.86310725405 1491.84210028099 1495.08328548438 1496.58056339095
##                1                1                1                1
## 1496.60171125666 1498.08044291704 1499.55366667684 1501.12064715797
##                1                1                1                1
## 1501.17855503437 1501.21366518418  1508.8468448778 1510.45430863931
##                1                1                1                1
## 1511.98114056823 1513.34974981661 1513.38506683499  1514.8240539953
##                1                1                1                1
##  1514.8812708676 1524.06044743458 1527.15544743167 1531.84639918942
##                1                1                1                1
## 1533.42025878569 1536.33062965696 1536.33994576226 1537.86424054399
##                1                1                1                1
## 1537.92272294275 1539.49322128955 1548.57773143399 1548.63202129363
##                1                1                1                1
## 1550.16368475638 1553.23893052127 1554.75810173682 1556.34113502664
##                1                1                1                1
## 1557.80050951136 1562.29551937295 1564.01474008876 1573.03265934371
##                1                1                1                1
## 1576.14184922508 1576.28252756018 1577.78151831989 1579.32196562623
##                1                1                1                1
##  1582.2669137062 1583.86311459609 1588.49835546611  1589.9548094831
##                1                1                1                1
## 1590.02802992333 1591.51044259274 1594.58810292884 1594.68829098796
##                1                1                1                1
##  1597.6869614008 1602.39108010067 1609.93702590622 1611.42134264613
##                1                1                1                1
## 1611.54133528683 1617.65708366265 1617.69000593007 1623.78614903038
##                1                1                1                1
## 1625.18683424776 1628.31461410053 1631.39911740088 1632.81785964854
##                1                1                1                1
## 1632.87468710604 1632.88000232472 1638.98063589899 1639.05912275231
##                1                1                1                1
## 1639.06867277078 1652.73993873163  1654.3548785496 1658.95588216217
##                1                1                1                1
## 1662.09822888211 1663.57492781175 1666.52944112625 1666.58500305729
##                1                1                1                1
## 1666.62353957135 1666.66177705746 1668.15676693055 1669.69486872669
##                1                1                1                1
## 1671.23000081435 1674.24387007794 1674.27588375874 1675.77544904081
##                1                1                1                1
## 1675.79266022333 1678.88874063931 1682.02304332535 1685.08183753767
##                1                1                1                1
## 1686.53024752076 1694.18360767022 1695.76234170343 1695.77341946748
##                1                1                1                1
## 1701.82937111246  1707.9539882669 1708.09772903082 1708.13335141479
##                1                1                1                1
## 1708.17929887054 1711.06879239491 1712.56138751704 1714.05481349368
##                1                1                1                1
## 1717.30013499983  1718.7281001973 1723.31784093825 1723.33037588127
##                1                1                1                1
## 1726.48734763365 1727.94581916531 1729.38899842959 1729.42266050671
##                1                1                1                1
## 1730.97164841439 1738.58988319798 1738.77290662329 1743.25241226854
##                1                1                1                1
```

```
##  1751.0240515476 1761.62739884332 1766.20513035802 1766.21284032766
##                1                1                1                1
## 1766.22664690693 1770.83632764561 1770.83891404907 1770.86891584877
##                1                1                1                1
## 1770.93550673826 1772.37000298945 1773.81811394459 1773.87488349021
##                1                1                1                1
## 1777.00620990731 1777.05777794734 1778.49765387403 1780.11906006611
##                1                2                1                1
## 1783.17322795317 1787.70467794786 1789.20980823582 1789.23449481684
##                1                1                1                1
## 1789.30619405644 1790.80088555954 1790.86215423473 1799.89778780008
##                1                1                1                1
## 1800.15373329564 1803.00316256037 1804.68956050187 1807.56671073573
##                1                1                1                1
## 1807.60963034831 1810.63082016674 1812.09449390806 1812.12595899652
##                1                1                1                1
## 1812.26936614684 1821.34981749278 1821.43006600483 1827.49785570177
##                1                1                1                1
## 1827.57765688715 1829.11575868329 1830.64835360824 1835.20336706924
##                1                1                1                1
##  1836.7966474914 1838.30050614312  1838.4375242008 1839.91182837949
##                1                1                1                1
## 1844.38752479063 1846.00430465745 1849.06676569213 1850.56272227102
##                1                1                1                1
## 1850.64790969235 1852.11702584805 1861.37898763912 1861.39683780082
##                1                1                1                1
## 1861.40455476607 1864.39955136094 1870.42076116788 1870.44051540008
##                1                1                1                1
##  1873.5989149287 1873.60170820073 1875.14658195938 1881.46955686396
##                1                1                1                1
## 1887.39671092161 1895.04734225585  1898.0397589922 1899.56494297546
##                1                1                1                1
##  1901.0539207388 1901.13006164676 1907.29747121809 1917.94505311642
##                1                1                1                1
## 1918.06885433327 1924.22601699506  1931.7068996567 1933.29158134284
##                1                1                1                1
## 1939.40216345666 1954.86455624147 1957.76330075636 1962.47879013844
##                1                1                1                1
## 1965.48822640479 1967.21070190598 1971.68499012551 1979.21962970567
##                1                1                1                1
## 1988.44202491317 1990.03822580306 1990.03930622196 1990.06867492944
##                1                1                1                1
##  1991.6061377464 2002.20689863866 2003.86557581833 2008.35659178891
##                1                1                1                1
## 2014.52702920341 2019.16323612127 2020.66695892168 2020.69411819308
##                1                1                1                1
## 2020.79505939514 2022.18208698995 2023.70019954761 2023.84228537026
##                1                1                1                1
## 2025.37488029521 2029.97398639773 2036.07012949804 2037.47038849177
##                1                1                1                1
## 2039.04234401755 2040.67644832929 2040.69410683848 2042.16411176056
##                1                1                1                1
## 2049.82816680421 2051.27988012488 2054.22151832232 2060.52272711362
##                1                1                1                1
```

```
## 2063.54728472645  2063.6219619105 2066.74671457806 2068.10867206402
##                1                1                1                1
## 2077.38930832636 2078.90044250362 2081.90329802482 2081.90348967733
##                1                1                1                1
## 2101.98556054598 2103.50752503356 2118.62747712987 2118.87410652013
##                1                1                1                1
## 2120.29113591464 2123.35632576454 2123.47038531041 2129.47184022718
##                1                1                1                1
## 2132.62120802334  2134.1275958707 2135.58753069112  2135.6673318765
##                1                1                1                1
## 2137.21910651127 2140.14209682544 2141.84537188519 2146.28956051206
##                1                1                1                1
## 2150.88791980928 2158.62777515168 2161.72417441587  2161.8573984555
##                1                1                1                1
## 2163.22253274133 2172.29127323101 2172.49835080309  2180.0623379877
##                1                1                1                1
## 2181.44601889113 2183.01943770566 2183.22924409296  2184.5696911398
##                1                1                1                1
## 2189.27664646548 2201.49886935102  2204.5838567869 2209.09823801083
##                1                1                1                1
## 2212.17240335683 2213.71348964236 2213.75992777857 2219.86532186941
##                1                1                1                1
## 2222.86690575436 2222.95202215843 2229.06626454967 2232.20056723571
##                1                1                1                1
##  2233.6693645432 2235.28068677958 2238.25480602755 2241.23873140331
##                1                1                1                1
## 2244.46664185399 2253.50626931035 2258.30633515976 2273.61702764664
##                1                1                1                1
##  2276.5396773516 2282.69342230475 2284.34942692108 2290.42391062558
##                1                1                1                1
## 2291.88607838233 2301.13202184008 2304.18897600357 2304.26784572675
##                1                1                1                1
##  2307.2092343598 2316.51905510788 2319.59638137971 2321.14240001407
##                1                1                1                1
## 2324.19860782293 2333.23671407865 2333.29398886283 2337.93727376676
##                1                1                1                1
## 2339.42791512697 2339.43031687354 2344.01639079213 2350.17734118042
##                1                1                1                1
## 2362.48879436566 2367.10423764971 2368.71708108671 2370.11431289524
##                1                1                1                1
## 2377.71031376065  2382.4033612414 2382.42886557154 2390.06315972426
##                1                1                1                1
## 2393.09455375641 2393.11089093793 2399.21278181821 2400.70455285358
##                1                1                1                1
## 2403.71310689849 2408.30923673208 2411.48525207407 2422.16196222178
##                1                1                1                1
## 2426.77632508693 2429.76556568138 2429.87531075458  2437.4525944529
##                1                1                1                1
## 2439.13327865464 2448.32201178502 2449.75549207414 2451.20423612149
##                1                1                1                1
## 2462.12512299662 2469.75278060299  2471.3736680739 2472.69222181176
##                1                1                1                1
## 2481.86462598107 2482.15739447959 2486.55284237867 2498.72702208545
##                1                1                1                1
```

```
## 2505.00119835038 2506.42891609413 2508.09062745462 2509.53645103424
##                1                1                1                1
##  2512.8609471785 2514.17853486851 2515.75770968346  2517.3268936982
##                1                1                1                1
## 2526.46948772032 2532.47557552281 2537.04886536916 2537.14359624851
##                1                1                1                1
## 2544.81076297723 2546.28707903702 2549.42474297252 2552.43676564234
##                1                1                1                1
## 2561.65550057242 2566.32207933521 2567.79549474751 2569.41379581496
##                1                1                1                1
## 2575.55739235559 2578.59170685547 2578.59258040579  2587.6987922067
##                1                1                1                1
## 2589.26948220601 2592.23663783896 2593.80082504806 2599.97556171917
##                1                1                1                1
## 2599.99594904358 2609.13765429932 2612.38429668247   2615.390378037
##                1                1                1                1
## 2620.05411472215 2626.08277816659 2643.01195637781 2655.20252906732
##                1                1                1                1
## 2656.64405339034 2658.35391574018 2659.67943309305 2664.37999278115
##                1                1                1                1
## 2667.54581846752 2675.12430278482 2676.60861130427 2678.09354592033
##                1                1                1                1
## 2678.10828396179 2678.12784697663 2681.14581795735  2682.8266938116
##                1                1                1                1
## 2690.43410650531  2691.9691748247 2699.59012423683 2701.05784112544
##                1                1                1                1
## 2707.27499217057  2708.8460802558 2716.42315708554  2716.4555048306
##                1                1                1                1
## 2725.62316830361 2725.76013056008 2733.30960733669 2742.56151366344
##                1                1                1                1
## 2760.97840454669 2767.09961709788 2771.64105855075 2773.15770781965
##                1                1                1                1
## 2774.72727470424 2776.35030125194 2777.75987635098 2788.51467483093
##                1                1                1                1
## 2797.72151379718 2803.73804081968 2806.80265614721 2809.97164275942
##                1                1                1                1
## 2817.52335111033 2823.93916657529 2826.78891571762 2839.12158122539
##                1                1                1                1
## 2842.18968454741 2843.80821221893 2858.98131076372 2880.50827374979
##                1                1                1                1
## 2883.50151523086  2883.5558050905 2884.93391510108 2886.60493557126
##                1                1                1                1
## 2888.05208047856 2889.60764913138 2891.20721781568 2895.62993869924
##                1                1                1                1
## 2897.32359928371 2898.86487722175 2900.41898258582 2903.54485793293
##                1                1                1                1
## 2906.69226445078 2912.68594597049 2915.78283525717 2917.22713763616
##                1                1                1                1
## 2917.33449617886 2920.44954121565 2923.35473927999 2926.46221630822
##                1                1                1                1
## 2932.46733150241 2937.22662926345 2941.80967599684 2948.00492673738
##                1                1                1                1
## 2951.07163778791 2955.59220823152 2955.63561852456 2958.70169714084
##                1                1                1                1
```

```
## 2958.72909732102 2964.76904539807 2964.79055583719 2984.76555253598
##                1                1                1                1
## 2989.31626674037 3004.63411615476 3016.79536939306  3023.0029634241
##                1                1                1                1
## 3039.72068730237 3042.93093924385 3046.03271118793  3047.6063861273
##                1                1                1                1
## 3050.50733417579 3053.50316162517 3056.81625293672 3064.29518162915
##                1                1                1                1
## 3064.39518503137 3065.88333848514 3067.51372671825 3069.12435276763
##                1                1                1                1
## 3073.67371669228 3081.22163102515 3081.30137429865 3085.80322057956
##                1                1                1                1
## 3088.87884964947 3092.06593621058 3093.56657536661 3096.40246027118
##                1                1                1                1
## 3102.76562008616 3111.88524038041 3114.97935161112  3130.2806142796
##                1                1                1                1
## 3133.29282860193 3137.95182459077 3160.99771788361 3174.80903608294
##                1                1                1                1
##  3180.8672687658 3194.67867146501 3199.34998989472 3205.31322817502
##                1                1                1                1
##  3213.0639056073 3217.64404708902 3226.84770991338 3240.73055740287
##                1                1                1                1
## 3243.73985825307 3249.76379621726 3251.31132083618 3255.89949113574
##                1                1                1                1
## 3266.54619948375 3274.32063203485 3277.41904318013  3283.5485981352
##                1                1                1                1
## 3285.01182454985 3288.19086707705 3292.76328337307 3305.08840927892
##                1                1                1                1
## 3305.12555215946 3306.59985633815 3317.24948515389 3331.00367053013
##                1                1                1                1
## 3335.67422841854 3337.16492557995 3338.70196217326 3341.89548706776
##                1                1                1                1
## 3366.30150479334 3380.08461155038  3403.1979627381 3409.20346779776
##                1                1                1                1
## 3409.24986258019 3426.08245675783 3429.18017734969 3430.70745705596
##                1                1                1                1
## 3436.71682881094 3442.97080285902 3445.98574599657 3447.69219719822
##                1                1                1                1
## 3449.13099270601 3452.11491808178 3458.25226794153  3464.4056213692
##                1                1                1                1
## 3464.46815557086 3468.97365390874 3472.25011343469 3473.63372332086
##                1                1                1                1
## 3481.29099120146  3482.7048655571 3484.38486152339 3489.01396942561
##                1                1                1                1
## 3490.45676582003 3498.09397344487   3498.214854194 3502.76846801648
##                1                1                1                1
## 3502.78322690758 3521.18461357449 3521.20245674058 3524.30960015522
##                1                1                1                1
## 3550.08860389005 3564.01732133102 3573.31200551656 3580.76567965053
##                1                1                1                1
## 3582.46390864838 3583.91549522403 3586.94795445901  3590.1975173099
##                1                1                1                1
## 3596.18523666467 3600.76959773013 3603.63581849866 3603.80055098053
##                1                1                1                1
```

```
## 3628.33201831929 3632.93643264489 3651.41220401283 3652.89289728416
##                1                1                1                1
## 3656.16042400983 3658.99694200662 3669.79353042398 3680.59182440091
##                1                1                1                1
##   3685.092206689 3686.67808250712 3689.64244486804 3700.32847112106
##                1                1                1                1
## 3712.58176100913  3720.3974518939  3764.9104252821 3781.77472545698
##                1                1                1                1
## 3786.27909411969 3787.90175809398 3790.79902513344 3793.85265308377
##                1                1                1                1
## 3830.72176053978 3835.19208382036 3841.50548694547 3844.40194408485
##                1                1                1                1
## 3856.81206483264 3881.21587902462 3881.29934703236 3901.12898261161
##                1                1                1                1
## 3919.56865557022 3927.31131837133 3928.84011857417 3936.41379125133
##                1                1                1                1
## 3951.79081195792 3965.46265244113 3991.71820864109 3994.67757566092
##                1                1                1                1
## 4034.39353252313 4042.25132527137 4048.40580181375 4056.19143369101
##                1                1                1                1
## 4058.94821404787 4072.92878390171 4075.94561069527 4080.34459730965
##                1                1                1                1
## 4086.70301642832 4123.44739120527 4146.34510206584 4163.17364444059
##                1                1                1                1
## 4181.56586395888 4186.23593116683 4189.28770385227 4210.70174562324
##                1                1                1                1
## 4247.68282198251 4253.82478443151 4256.72833409953 4265.98067319486
##                1                1                1                1
## 4284.24864351166  4284.3681454719 4311.90187859343  4311.9861209965
##                1                1                1                1
## 4324.13581233523 4338.07555173904 4354.92307395065 4370.19604767258
##                1                1                1                1
##   4389.949159474 4431.58358253917 4445.17383944665 4491.30104854334
##                1                1                1                1
## 4535.53864927245 4612.21010817592 4616.94523807711 4633.76447134217
##                1                1                1                1
##  4644.4754246343 4656.79129863827 4670.49953655875 4673.67541816012
##                1                1                1                1
## 4679.66516222063 4710.38434769361 4737.87264445433 4757.97637471879
##                1                1                1                1
##  4764.0976299658  4828.4676972326 4836.07642425837 4875.91984915979
##                1                1                1                1
## 4882.01383142232  4977.1085269479 5081.15205703001 5211.53770115873
##                1                1                1                1
## 5220.89997068822 5249.89471250412 5407.72535576869 5447.52350856724
##                1                1                1                1
## 5524.12232155285 5697.44338458309 5968.56129888332 5997.84034678921
##                1                1                1                1
##  6152.4616121133 6958.67142067345
##                1                1
```

Code Appendix

```
knitr::opts_chunk$set(echo=FALSE, error=FALSE, warning=FALSE, message=FALSE)


# Libraries


library(stringr)
library(tidyr)
library(DataExplorer)
library(dplyr)
library(visdat)
library(pROC)
library(mice)
library(corrplot)
library(MASS)
library(caret)
library(e1071)
library(rbin)

library(GGally)
library(ggplot2)
library(readr)
library(reshape2)
library(purrr)
library(leaps)

set.seed(2012)


# training data
insurance <- read.csv('https://raw.githubusercontent.com/hillt5/DATA_621/master/HW4/insurance_training_
# test data
insurance_test <- read.csv('https://raw.githubusercontent.com/hillt5/DATA_621/master/HW4/insurance_trai
glimpse(insurance)
head(insurance)
summary(insurance)
insurance_fix <- dplyr::select(insurance, -INDEX)

insurance_fix$HOME_VAL <- substr(insurance_fix$HOME_VAL, 2, nchar(insurance_fix$HOME_VAL)) # remove the
insurance_fix$HOME_VAL <- as.numeric(str_remove_all(insurance_fix$HOME_VAL, "[[:punct:]]")) # remove th

insurance_fix$BLUEBOOK<- substr(insurance_fix$BLUEBOOK , 2, nchar(insurance_fix$BLUEBOOK ))
insurance_fix$BLUEBOOK<- as.numeric(str_remove_all(insurance_fix$BLUEBOOK,"[[:punct:]]"))

insurance_fix$INCOME  <- substr(insurance_fix$INCOME, 2, nchar(insurance_fix$INCOME))
insurance_fix$INCOME <- as.numeric(str_remove_all(insurance_fix$INCOME, "[[:punct:]]"))

insurance_fix$OLDCLAIM <- substr(insurance_fix$OLDCLAIM, 2, nchar(insurance_fix$OLDCLAIM))
insurance_fix$OLDCLAIM <- as.numeric(str_remove_all(insurance_fix$OLDCLAIM, "[[:punct:]]"))
```

```
insurance_fix$MSTATUS = as.factor(str_remove(insurance_fix$MSTATUS, 'z_')) #several variables have a a
insurance_fix$PARENT1 = as.factor(str_remove(insurance_fix$PARENT1, 'z_'))
insurance_fix$EDUCATION = str_replace(insurance_fix$EDUCATION, '<', 'Less than ') #change < to less tha
insurance_fix$SEX= as.factor(str_remove(insurance_fix$SEX, 'z_'))
insurance_fix$EDUCATION = as.factor(str_remove(insurance_fix$EDUCATION, 'z_'))
insurance_fix$JOB[insurance_fix$JOB == ""] <- 'Other Job' #recode blank spaces as 'Other Job'
insurance_fix$JOB = as.factor(str_remove(insurance_fix$JOB, 'z_'))
insurance_fix$CAR_USE = as.factor(str_remove(insurance_fix$CAR_USE, 'z_'))
insurance_fix$CAR_TYPE = as.factor(str_remove(insurance_fix$CAR_TYPE, 'z_'))
insurance_fix$URBANICITY = as.factor(str_remove(insurance_fix$URBANICITY, 'z_'))
insurance_fix$REVOKED = as.factor(str_remove(insurance_fix$REVOKED, 'z_'))
insurance_fix$RED_CAR = as.factor(str_remove(insurance_fix$RED_CAR, 'z_'))

summary(insurance_fix)

insurance_fix$CAR_AGE[insurance_fix$CAR_AGE <1] <- 1
cat_cols = c()
j <- 1
for (i in 4:ncol(insurance_fix)) {
  if (class((insurance_fix[,i])) == 'factor') {
      print(names(insurance_fix[i]))
      print(levels(insurance_fix[,i]))
      cat_cols[j]=names(insurance_fix[i])
      j <- j+1
  }

}



ins_fact <-  insurance_fix[cat_cols]
ins_factm <- melt(ins_fact, measure.vars = cat_cols, variable.name = 'metric', value.name = 'value')

ggplot(ins_factm, aes(x = value)) +
  geom_bar() +
  scale_fill_brewer(palette = "Set1") +
  facet_wrap( ~ metric, nrow = 5L, scales = 'free') + coord_flip()
plot_histogram(insurance_fix, geom_histogram_args = list("fill" = "tomato4"))

plot_histogram(insurance_fix, scale_x = "log10", geom_histogram_args = list("fill" = "springgreen4"))
# check columns having missing values
insurance_fix %>% summarise_all(funs(sum(is.na(.)))) %>% select_if(~any(.)>0)
plot_missing(insurance_fix)

round(colSums(is.na(insurance_fix))/nrow(insurance_fix),3)

vis_dat(insurance_fix %>% dplyr:: select(YOJ, INCOME, HOME_VAL, CAR_AGE))



numer_data <- insurance_fix[,c('TARGET_AMT','AGE','YOJ','INCOME','HOME_VAL','TRAVTIME','BLUEBOOK','TIF'

AGE_MEDIAN <- median(filter(insurance_fix,AGE > 0)$AGE)
```

```
INCOME_MEDIAN <- median(filter(insurance_fix,INCOME > 0)$INCOME)
YOJ_MEDIAN <- median(filter(insurance_fix,YOJ > 0)$YOJ)
HOME_VAL_MEDIAN <- median(filter(insurance_fix,HOME_VAL > 0)$HOME_VAL)
CAR_AGE_MEDIAN <- median(filter(insurance_fix,CAR_AGE > 0)$CAR_AGE)


numer_data <- numer_data %>% dplyr::mutate(AGE = replace_na(AGE,AGE_MEDIAN),
                             INCOME = replace_na(INCOME,INCOME_MEDIAN),
                             YOJ = replace_na(YOJ,YOJ_MEDIAN),
                             HOME_VAL = replace_na(HOME_VAL,HOME_VAL_MEDIAN),
                             CAR_AGE = replace_na(CAR_AGE,CAR_AGE_MEDIAN))
corrplot(cor(numer_data),type="upper")

mlr_crash <- subset(filter(insurance_fix,TARGET_FLAG==1),select = -c(TARGET_FLAG))

mlr_crash_fix_na <- mlr_crash

AGE_MEDIAN <- median(filter(mlr_crash_fix_na,AGE > 0)$AGE)
INCOME_MEDIAN <- median(filter(mlr_crash_fix_na,INCOME > 0)$INCOME)
YOJ_MEDIAN <- median(filter(mlr_crash_fix_na,YOJ > 0)$YOJ)
HOME_VAL_MEDIAN <- median(filter(mlr_crash_fix_na,HOME_VAL > 0)$HOME_VAL)
CAR_AGE_MEDIAN <- median(filter(mlr_crash_fix_na,CAR_AGE > 0)$CAR_AGE)

mlr_crash_fix_na <- mlr_crash_fix_na %>% dplyr::mutate(AGE = replace_na(AGE,AGE_MEDIAN),
                             INCOME = replace_na(INCOME,INCOME_MEDIAN),
                             YOJ = replace_na(YOJ,YOJ_MEDIAN),
                             HOME_VAL = replace_na(HOME_VAL,HOME_VAL_MEDIAN),
                             CAR_AGE = replace_na(CAR_AGE,CAR_AGE_MEDIAN))
mlr_crash_transf <- mlr_crash_fix_na
mlr_crash_transf$AGE <- log(mlr_crash_transf$AGE)
mlr_crash_transf$BLUEBOOK <- log(mlr_crash_transf$BLUEBOOK)
mlr_crash_transf$CAR_AGE <- log(mlr_crash_transf$CAR_AGE + 1)
mlr_crash_transf$HOME_VAL <- log(mlr_crash_transf$HOME_VAL + 1)
mlr_crash_transf$INCOME <- log(mlr_crash_transf$INCOME + 1)
mlr_crash_transf$OLDCLAIM <- log(mlr_crash_transf$OLDCLAIM + 1)
mlr_crash_transf$TRAVTIME <- log(mlr_crash_transf$TRAVTIME)

insurance_fix2 <- insurance_fix
insurance_fix2$HOME_VAL <-ifelse(insurance_fix2$HOME_VAL == 0, NA, insurance_fix2$HOME_VAL)
insurance_bins <- insurance_fix %>%
  mutate(CAR_AGE_BIN=cut(CAR_AGE, breaks=c(-Inf, 1, 3, 12, Inf), labels=c("New","Like New","Average", '
  mutate(HOME_VAL_BIN=cut(HOME_VAL, breaks=c(-Inf, 0, 50000, 150000, 250000, Inf), labels=c("Zero", "$0
  mutate(HAS_HOME_KIDS = as.factor(case_when(HOMEKIDS == 0 ~ 'No kids', HOMEKIDS > 0 ~ ('Has kids')))) 
  mutate(HAS_KIDSDRIV = as.factor(case_when(KIDSDRIV == 0 ~ 'No kids driving', KIDSDRIV > 0 ~ 'Has kids
  mutate(OLDCLAIM_BIN =cut(OLDCLAIM, breaks=c(-Inf, 0, 3000, 6000, 9000, Inf), labels=c("Zero","$0-$3k"
  mutate(TIF_BIN =cut(TIF, breaks=c(-Inf, 0, 1, 4, 7, Inf), labels=c("Zero","Less than 1 year", "1-4 ye
  mutate(YOJ_BIN =cut(YOJ, breaks=c(-Inf, 0, 10, 15, Inf), labels=c("Zero","Less than 10 years", 'Betwe
  dplyr::select(-c(CAR_AGE, HOME_VAL, HOMEKIDS, KIDSDRIV, OLDCLAIM, TIF, YOJ)) #drop the binned feature

summary(insurance_bins)
head(insurance_bins)
```

```
insurance_logistic_model <- glm(insurance_fix, family = 'binomial', formula = TARGET_FLAG~.-TARGET_AMT)

summary(insurance_logistic_model)

get_cv_performance <- function(data_frame, model, split = 0.8) {  ### input is dataframe for partitioni
  n <- ncol(data_frame) #number of columns in original dataframe
  train_control <- trainControl(method="repeatedcv", number=10, repeats=3)
  trainIndex <- createDataPartition(data_frame[,n], p=split, list=FALSE)
  data_train <- data_frame[trainIndex,]
  data_test <- data_frame[-trainIndex,]

  x_test <- data_test[,2:n] #explanatory variables
  y_test <- data_test[,1]  #response variable
  predictions <- predict(model, x_test, type = 'response')

  return(confusionMatrix(data = (as.factor(as.numeric(predictions>0.5))), reference = as.factor(y_test)

  return(plot(roc(y_test, predictions),print.auc=TRUE))

}
get_roc <- function(data_frame, model, split = 0.8) {  ### input is dataframe for partitioning, model a
  n <- ncol(data_frame) #number of columns in original dataframe
  train_control <- trainControl(method="repeatedcv", number=10, repeats=3)
  trainIndex <- createDataPartition(data_frame[,n], p=split, list=FALSE)
  data_train <- data_frame[trainIndex,]
  data_test <- data_frame[-trainIndex,]

  x_test <- data_test[,2:n] #explanatory variables
  y_test <- data_test[,1]  #response variable
  predictions <- predict(model, x_test, type = 'response')
  return(plot(roc(y_test, predictions),print.auc=TRUE))

}

get_cv_performance(insurance_fix, insurance_logistic_model)
get_roc(insurance_fix, insurance_logistic_model)



insurance_impute <- mice(insurance_fix, method = 'cart', m = 1)

imputed_lm <- glm.mids(data = insurance_impute, formula = TARGET_FLAG ~.-TARGET_AMT, family = 'binomial


imputed_lm


get_cv_performance(insurance_fix, imputed_lm$analyses[[1]])
get_roc(insurance_fix, imputed_lm$analyses[[1]])



insurance_impute2 <- mice(insurance_fix2, method = 'cart', m = 1)
```

```
imputed_lm2 <- glm.mids(data = insurance_impute2, formula = TARGET_FLAG ~.-TARGET_AMT, family = 'binomia
imputed_lm2
get_cv_performance(insurance_fix2, imputed_lm2$analyses[[1]])
get_roc(insurance_fix2, imputed_lm2$analyses[[1]])



binned_lm <- glm(data = insurance_bins, formula = TARGET_FLAG ~.-TARGET_AMT, family = 'binomial')


summary(binned_lm)


get_cv_performance(insurance_bins, binned_lm)
get_roc(insurance_bins, binned_lm)

insurance_binned_impute <- mice(insurance_bins, method = 'cart', m = 1)

binned_imputed_lm <- glm.mids(data = insurance_binned_impute, formula = TARGET_FLAG ~.-TARGET_AMT, famil


binned_imputed_lm


get_cv_performance(insurance_bins, binned_imputed_lm$analyses[[1]])
get_roc(insurance_bins, binned_imputed_lm$analyses[[1]])

mlr<- lm(TARGET_AMT ~ . ,data=mlr_crash)
summary(mlr)
mlr<- lm(TARGET_AMT ~ . ,data=mlr_crash_transf)
summary(mlr)
mlr1 <- lm(TARGET_AMT ~ . ,data=mlr_crash_transf)
summary(mlr1)
mlr2 <- update(mlr1,TARGET_AMT~. - OLDCLAIM)
summary(mlr2)
mlr3 <- update(mlr2,TARGET_AMT~. - YOJ)
summary(mlr3)
mlr4 <- update(mlr3,TARGET_AMT~. - URBANICITY)
summary(mlr4)
mlr5 <- update(mlr4,TARGET_AMT~. - TRAVTIME)
summary(mlr5)
mlr6 <- update(mlr5,TARGET_AMT~. - INCOME)
summary(mlr6)
mlr7 <- update(mlr6,TARGET_AMT~. - CLM_FREQ)
summary(mlr7)
mlr8 <- update(mlr7,TARGET_AMT~. - TIF)
summary(mlr8)
mlr9 <- update(mlr8,TARGET_AMT~. - RED_CAR)
summary(mlr9)
mlr10 <- update(mlr9,TARGET_AMT~. - PARENT1)
summary(mlr10)
mlr11 <- update(mlr10,TARGET_AMT~. - KIDSDRIV)
summary(mlr11)
```

```
mlr12 <- update(mlr11,TARGET_AMT~. - AGE)
summary(mlr12)
mlr13 <- update(mlr12,TARGET_AMT~. - CAR_USE)
summary(mlr13)
mlr14 <- update(mlr13,TARGET_AMT~. - JOB)
summary(mlr14)
mlr15 <- update(mlr14,TARGET_AMT~. - EDUCATION)
summary(mlr15)
mlr16 <- update(mlr15,TARGET_AMT~. - CAR_TYPE)
summary(mlr16)
mlr17 <- update(mlr16,TARGET_AMT~. - HOMEKIDS)
summary(mlr17)
mlr18 <- update(mlr17,TARGET_AMT~. - CAR_AGE)
summary(mlr18)
mlr19 <- update(mlr18,TARGET_AMT~. - HOME_VAL)
summary(mlr19)
mlr20 <- update(mlr19,TARGET_AMT~. - MSTATUS)
summary(mlr20)
mlr21 <- update(mlr20,TARGET_AMT~. - REVOKED)
summary(mlr21)
mlr22 <- update(mlr21,TARGET_AMT~. - SEX)
summary(mlr22)
mlr_fwd <- lm(TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX ,data= mlr_crash_transf)
summary(mlr_fwd)

mlr_fwd <- lm(TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX + MSTATUS ,data= mlr_crash_transf)
summary(mlr_fwd)

mlr_fwd <- lm(TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX + MSTATUS + HOME_VAL,data= mlr_crash_transf)
summary(mlr_fwd)

mlr_fwd <- lm(TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX + MSTATUS + HOME_VAL + REVOKED,data= mlr_crash_trans
summary(mlr_fwd)

mlr_fwd <- lm(TARGET_AMT ~ BLUEBOOK + MVR_PTS + SEX + MSTATUS + HOME_VAL + REVOKED + CAR_AGE,data= mlr_c
summary(mlr_fwd)
mlr_full <- regsubsets(TARGET_AMT ~ . ,data=mlr_crash, nvmax=NULL)
mlr_summary<- summary(mlr_full)
par(mfrow=c(2,2))
plot(mlr_summary$cp,xlab = "# Variables", ylab = "cp - estimate of prediction error")
points(13,mlr_summary$cp[13],pch=20,col="red")
plot(mlr_summary$rsq,xlab = "# Variables", ylab = "R^2")

mlr_full_transf <- regsubsets(TARGET_AMT ~ . ,data=mlr_crash_transf, nvmax=NULL)
mlr_summary_transf <- summary(mlr_full_transf)

par(mfrow=c(1,2))
plot(mlr_summary_transf$cp,xlab = "# Variables", ylab = "cp - estimate of prediction error")
points(7,mlr_summary_transf$cp[7],pch=20,col="red")
plot(mlr_summary_transf$rsq,xlab = "# Variables", ylab = "R^2")
coef(mlr_full,7)
model_6 <- lm(TARGET_AMT ~ MSTATUS +JOB+ BLUEBOOK + CAR_AGE+EDUCATION, data = mlr_crash_transf)
summary(model_6)
```

```
model_log <- lm(log(TARGET_AMT) ~ MSTATUS+SEX+ BLUEBOOK + CLM_FREQ + MVR_PTS+EDUCATION, data = mlr_crash
summary(model_log)

get_cv_performance(insurance_bins, binned_lm)
get_roc(insurance_bins, binned_lm)
res0 <- resid(mlr)
plot(density(res0))
qqnorm(res0)
qqline(res0)
ggplot(data = mlr, aes(x = .fitted, y = .resid)) +
  geom_jitter() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
res0 <- resid(model_log)
plot(density(res0))
qqnorm(res0)
qqline(res0)
ggplot(data = model_log, aes(x = .fitted, y = .resid)) +
  geom_jitter() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
insurance_fix3 <- dplyr::select(insurance_test, -INDEX)

insurance_fix3$HOME_VAL <- substr(insurance_fix3$HOME_VAL, 2, nchar(insurance_fix3$HOME_VAL)) # remove
insurance_fix3$HOME_VAL <- as.numeric(str_remove_all(insurance_fix3$HOME_VAL, "[[:punct:]]")) # remove

insurance_fix3$BLUEBOOK<- substr(insurance_fix3$BLUEBOOK , 2, nchar(insurance_fix3$BLUEBOOK ))
insurance_fix3$BLUEBOOK<- as.numeric(str_remove_all(insurance_fix3$BLUEBOOK,"[[:punct:]]"))

insurance_fix3$INCOME  <- substr(insurance_fix3$INCOME, 2, nchar(insurance_fix3$INCOME))
insurance_fix3$INCOME <- as.numeric(str_remove_all(insurance_fix3$INCOME, "[[:punct:]]"))

insurance_fix3$OLDCLAIM <- substr(insurance_fix3$OLDCLAIM, 2, nchar(insurance_fix3$OLDCLAIM))
insurance_fix3$OLDCLAIM <- as.numeric(str_remove_all(insurance_fix3$OLDCLAIM, "[[:punct:]]"))

insurance_fix3$MSTATUS = as.factor(str_remove(insurance_fix3$MSTATUS, 'z_')) #several variables have a
insurance_fix3$PARENT1 = as.factor(str_remove(insurance_fix3$PARENT1, 'z_'))
insurance_fix3$EDUCATION = str_replace(insurance_fix3$EDUCATION, '<', 'Less than ') #change < to less t
insurance_fix3$SEX= as.factor(str_remove(insurance_fix3$SEX, 'z_'))
insurance_fix3$EDUCATION = as.factor(str_remove(insurance_fix3$EDUCATION, 'z_'))
insurance_fix3$JOB[insurance_fix3$JOB == ""] <- 'Other Job' #recode blank spaces as 'Other Job'
insurance_fix3$JOB = as.factor(str_remove(insurance_fix3$JOB, 'z_'))
insurance_fix3$CAR_USE = as.factor(str_remove(insurance_fix3$CAR_USE, 'z_'))
insurance_fix3$CAR_TYPE = as.factor(str_remove(insurance_fix3$CAR_TYPE, 'z_'))
insurance_fix3$URBANICITY = as.factor(str_remove(insurance_fix3$URBANICITY, 'z_'))
insurance_fix3$REVOKED = as.factor(str_remove(insurance_fix3$REVOKED, 'z_'))
insurance_fix3$RED_CAR = as.factor(str_remove(insurance_fix3$RED_CAR, 'z_'))
insurance_fix3$CAR_AGE[insurance_fix3$CAR_AGE <1] <- 1
insurance_bins2 <- insurance_fix3 %>%
  mutate(CAR_AGE_BIN=cut(CAR_AGE, breaks=c(-Inf, 1, 3, 12, Inf), labels=c("New","Like New","Average", '
  mutate(HOME_VAL_BIN=cut(HOME_VAL, breaks=c(-Inf, 0, 50000, 150000, 250000, Inf), labels=c("Zero", "$0-
```

```
  mutate(HAS_HOME_KIDS = as.factor(case_when(HOMEKIDS == 0 ~ 'No kids', HOMEKIDS > 0 ~ ('Has kids')))) %
  mutate(HAS_KIDSDRIV = as.factor(case_when(KIDSDRIV == 0 ~ 'No kids driving', KIDSDRIV > 0 ~ 'Has kids
  mutate(OLDCLAIM_BIN =cut(OLDCLAIM, breaks=c(-Inf, 0, 3000, 6000, 9000, Inf), labels=c("Zero","$0-$3k"
  mutate(TIF_BIN =cut(TIF, breaks=c(-Inf, 0, 1, 4, 7, Inf), labels=c("Zero","Less than 1 year", "1-4 yea
  mutate(YOJ_BIN =cut(YOJ, breaks=c(-Inf, 0, 10, 15, Inf), labels=c("Zero","Less than 10 years", 'Betwe
  dplyr::select(-c(CAR_AGE, HOME_VAL, HOMEKIDS, KIDSDRIV, OLDCLAIM, TIF, YOJ)) #drop the binned features

mlr_crash2 <- subset(filter(insurance_fix2,TARGET_FLAG==1),select = -c(TARGET_FLAG))
mlr_crash_fix_na2 <- mlr_crash2
AGE_MEDIAN <- median(filter(mlr_crash_fix_na2,AGE > 0)$AGE)
INCOME_MEDIAN <- median(filter(mlr_crash_fix_na2,INCOME > 0)$INCOME)
YOJ_MEDIAN <- median(filter(mlr_crash_fix_na2,YOJ > 0)$YOJ)
HOME_VAL_MEDIAN <- median(filter(mlr_crash_fix_na2,HOME_VAL > 0)$HOME_VAL)
CAR_AGE_MEDIAN <- median(filter(mlr_crash_fix_na2,CAR_AGE > 0)$CAR_AGE)

mlr_crash_fix_na2 <- mlr_crash_fix_na2 %>% dplyr::mutate(AGE = replace_na(AGE,AGE_MEDIAN),
                             INCOME = replace_na(INCOME,INCOME_MEDIAN),
                             YOJ = replace_na(YOJ,YOJ_MEDIAN),
                             HOME_VAL = replace_na(HOME_VAL,HOME_VAL_MEDIAN),
                             CAR_AGE = replace_na(CAR_AGE,CAR_AGE_MEDIAN))
mlr_crash_transf2 <- mlr_crash_fix_na2
mlr_crash_transf2$AGE <- log(mlr_crash_transf2$AGE)
mlr_crash_transf2$BLUEBOOK <- log(mlr_crash_transf2$BLUEBOOK)
mlr_crash_transf2$CAR_AGE <- log(mlr_crash_transf2$CAR_AGE + 1)
mlr_crash_transf2$HOME_VAL <- log(mlr_crash_transf2$HOME_VAL + 1)
mlr_crash_transf2$INCOME <- log(mlr_crash_transf2$INCOME + 1)
mlr_crash_transf2$OLDCLAIM <- log(mlr_crash_transf2$OLDCLAIM + 1)
mlr_crash_transf2$TRAVTIME <- log(mlr_crash_transf2$TRAVTIME)

predicted_amt <- predict(model_log, insurance_bins2)
predicted_amt2 = predicted_amt
predicted_amt2[] = 0

predicted_flag = predict(binned_lm, insurance_bins2, type = "response")
predicted_flag_bin = ifelse(predicted_flag > 0.5, 1, 0)

for (i in 1:length(predicted_amt)) {
  if(predicted_flag_bin[i] == 0 | is.na(predicted_flag_bin[i])) {
    predicted_amt2[i] = 0
  } else {
    predicted_amt2[i] = predicted_amt[i]
  }

}
table(predicted_flag_bin)
table(predicted_amt2)
```