Predict 411

AUTO INSURANCE ASSIGNMENT

Zeeshan Latifi

Northwestern University Winter 2018

Introduction:

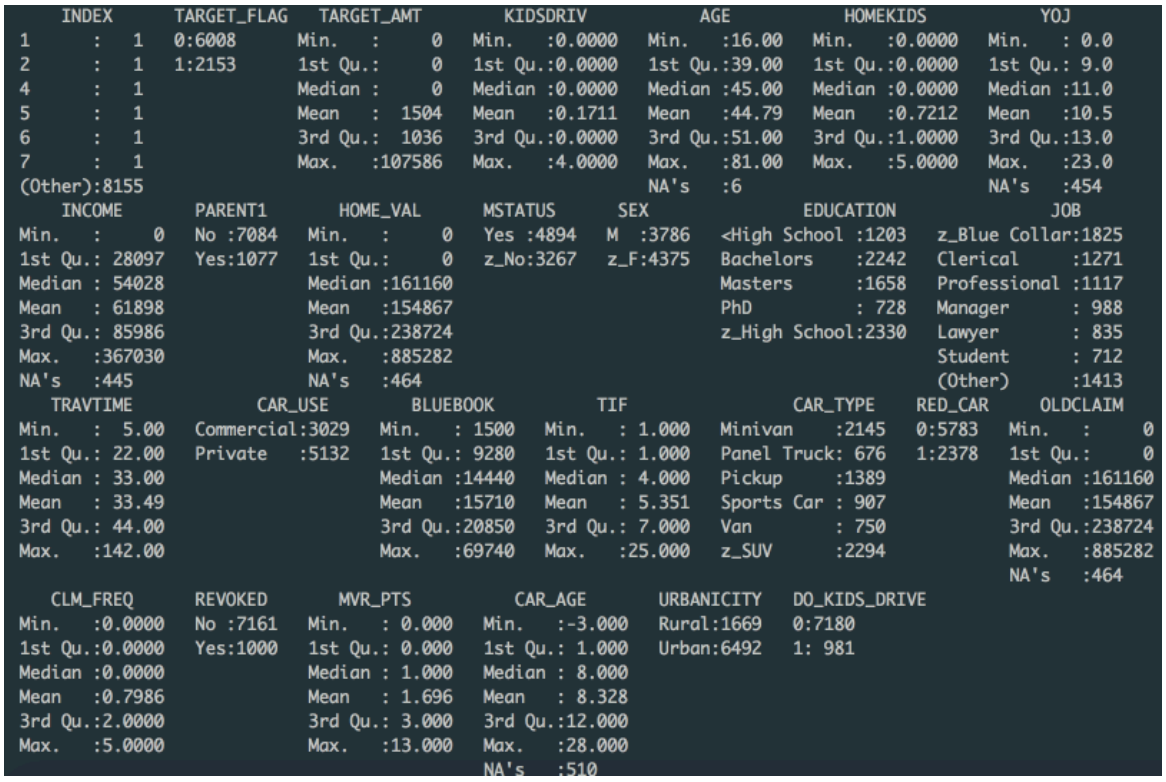
The purpose of this analysis is to determine whether we can predict the average cost a driver will incur if they crash their vehicle. The data set contains approximately 8000 records. Each record represents a customer at an auto insurance company. Each record has two target variables. The first target variable, TARGET_FLAG, is a binary variable 1 or a 0. A "1" means that the driver was in a car crash. A zero means that the driver was not in a car crash. We will predict the average probability that a driver will have a car accident. Once this model is created we will predict the average cost they will incur. This data set contains several categorical variables; because of this, we will utilize logistic regression for our analysis.

Data Exploration:

The data contains 23 variables and 8,161 records of driver data. Figure 1 below highlights the summary of the data for each variable. As you can see there are several variables that contain missing values. For these variables with "NA's" we'll impute values using the simple average method.

As part of our analysis we made sure to look at the values provided and confirm if they made the most sense logically. From here you can make reasonable inferences from our data. The mean income is about $61k dollars. The mean age is 44 with a standard deviation of 8.6 years. The average bluebook value of a car is approximately $14.5k dollars with a standard deviation of about $8k. From a general population perspective, this all can be reasonably understood.

Figure 1



```
     INDEX      TARGET_FLAG   TARGET_AMT        KIDSDRIV          AGE           HOMEKIDS          YOJ
1      :   1   0:6008      Min.   :     0   Min.   :0.0000   Min.   :16.00   Min.   :0.0000   Min.   : 0.0
2      :   1   1:2153      1st Qu.:     0   1st Qu.:0.0000   1st Qu.:39.00   1st Qu.:0.0000   1st Qu.: 9.0
4      :   1               Median :     0   Median :0.0000   Median :45.00   Median :0.0000   Median :11.0
5      :   1               Mean   :  1504   Mean   :0.1711   Mean   :44.79   Mean   :0.7212   Mean   :10.5
6      :   1               3rd Qu.:  1036   3rd Qu.:0.0000   3rd Qu.:51.00   3rd Qu.:1.0000   3rd Qu.:13.0
7      :   1               Max.   :107586   Max.   :4.0000   Max.   :81.00   Max.   :5.0000   Max.   :23.0
(Other):8155                                                  NA's   :6                         NA's   :454
    INCOME       PARENT1       HOME_VAL      MSTATUS      SEX             EDUCATION              JOB
Min.   :     0   No :7084   Min.   :     0   Yes :4894   M  :3786   <High School :1203   z_Blue Collar:1825
1st Qu.: 28097   Yes:1077   1st Qu.:     0   z_No:3267   z_F:4375   Bachelors    :2242   Clerical      :1271
Median : 54028              Median :161160                          Masters      :1658   Professional :1117
Mean   : 61898              Mean   :154867                          PhD          : 728   Manager      : 988
3rd Qu.: 85986              3rd Qu.:238724                          z_High School:2330   Lawyer       : 835
Max.   :367030              Max.   :885282                                               Student      : 712
NA's   :445                 NA's   :464                                                  (Other)      :1413
   TRAVTIME          CAR_USE        BLUEBOOK         TIF              CAR_TYPE    RED_CAR     OLDCLAIM
Min.   :  5.00   Commercial:3029   Min.   : 1500   Min.   : 1.000   Minivan    :2145   0:5783   Min.   :     0
1st Qu.: 22.00   Private   :5132   1st Qu.: 9280   1st Qu.: 1.000   Panel Truck: 676   1:2378   1st Qu.:     0
Median : 33.00                     Median :14440   Median : 4.000   Pickup     :1389            Median :161160
Mean   : 33.49                     Mean   :15710   Mean   : 5.351   Sports Car : 907            Mean   :154867
3rd Qu.: 44.00                     3rd Qu.:20850   3rd Qu.: 7.000   Van        : 750            3rd Qu.:238724
Max.   :142.00                     Max.   :69740   Max.   :25.000   z_SUV      :2294            Max.   :885282
                                                                                                NA's   :464
   CLM_FREQ       REVOKED        MVR_PTS          CAR_AGE        URBANICITY   DO_KIDS_DRIVE
Min.   :0.0000   No :7161   Min.   : 0.000   Min.   :-3.000   Rural:1669   0:7180
1st Qu.:0.0000   Yes:1000   1st Qu.: 0.000   1st Qu.: 1.000   Urban:6492   1: 981
Median :0.0000              Median : 1.000   Median : 8.000
Mean   :0.7986              Mean   : 1.696   Mean   : 8.328
3rd Qu.:2.0000              3rd Qu.: 3.000   3rd Qu.:12.000
Max.   :5.0000              Max.   :13.000   Max.   :28.000
                                             NA's   :510
```

For this analysis our response variable is binary. Therefore, we will be utilizing logistic regression. A driver will either get into an accident (1) or not get an accident (0). Let's begin by looking at the predictor variables. Figure 2 below outlines the age and target amount of cost for our data set. The distribution is very much normal when it comes to age of the driver. However, for the target amount of cost it is very heavily skewed right. This is because most accidents are minor and do not cause much damage. Because of this, we may need to transform our variable using the log function to normalize our data.

Figure 2

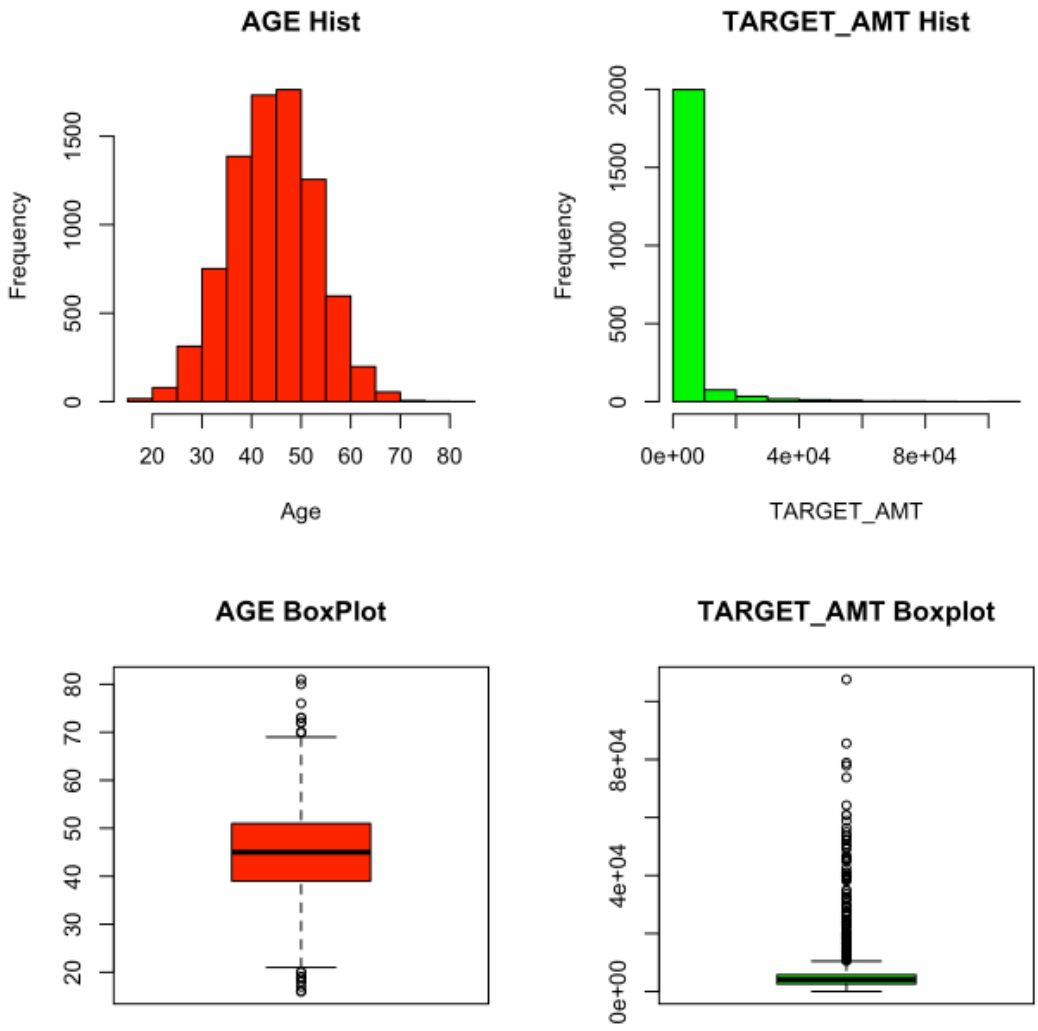
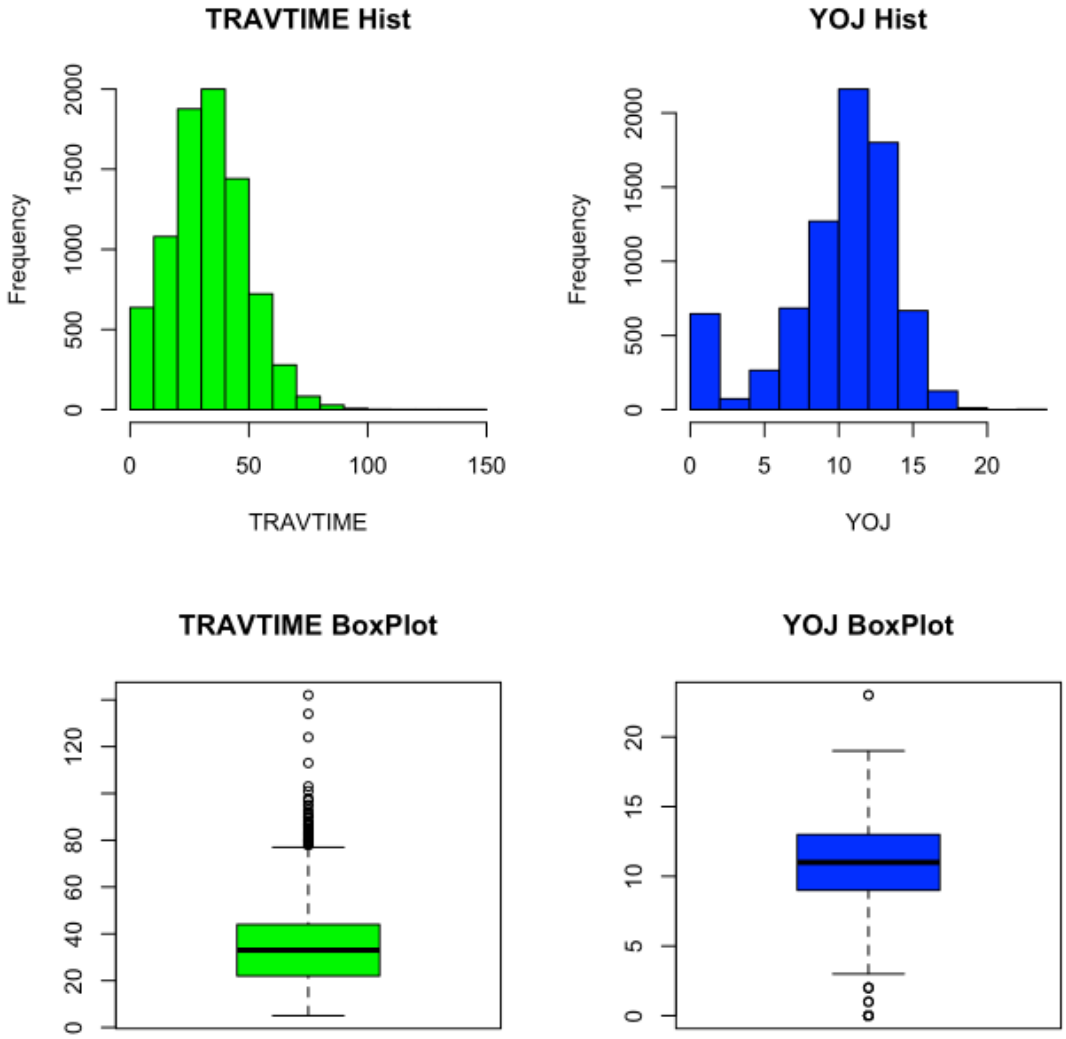


Figure 3 below showcases travel time driver information for travel time and the years they've been working at the same job. Logically, these imply that the longer you've been at a job the safer you are. In addition, the further the driving distance, the more chances of having an accident.

Figure 3



As shown, you can see that travel time is skewed right, this is due to the fact that many people tend to live closer to work. Years on the job data also makes sense, most people that don't like the job will leave hence the high turnover early on. The rest of the data is pretty normally distributed.

Figure 4

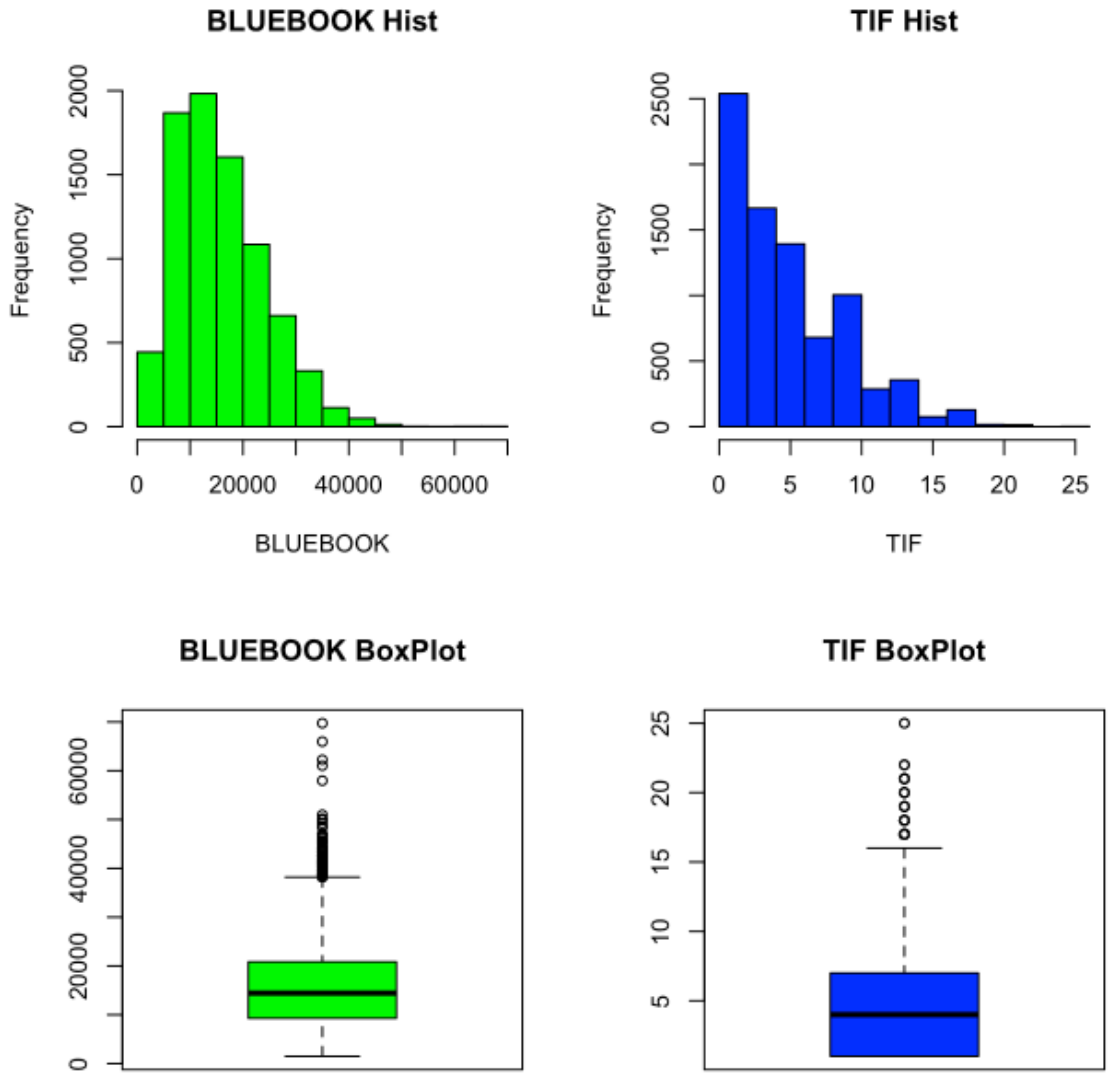
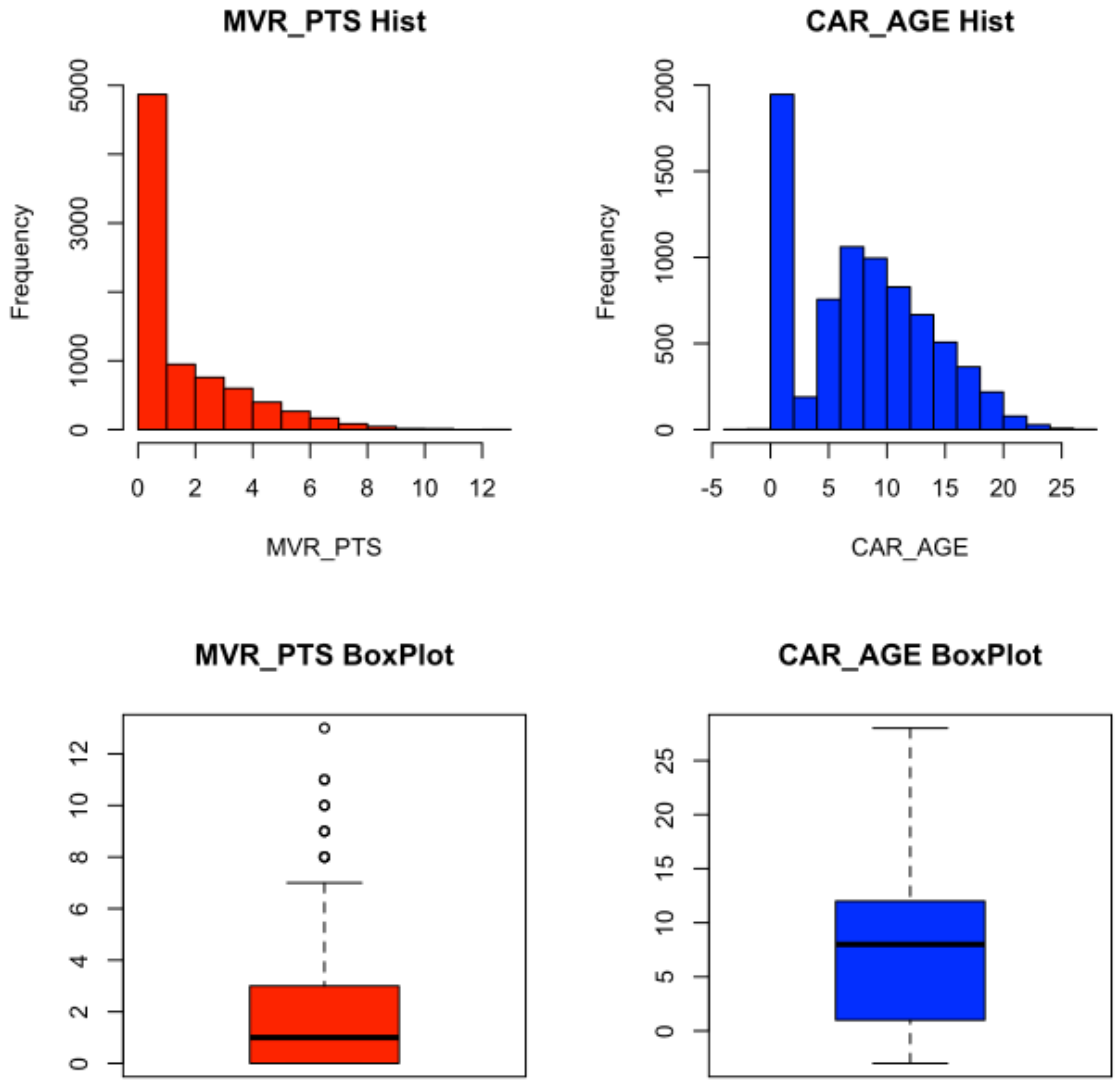


Figure 4 above illustrates the bluebook value of cars and the time they have been on the same insurance. These both make sense because many people don't own cars valued above $60,000. The TIF also is telling, as most drivers will switch insurance providers if they find a better deal, hence the high turnover early on here as well.

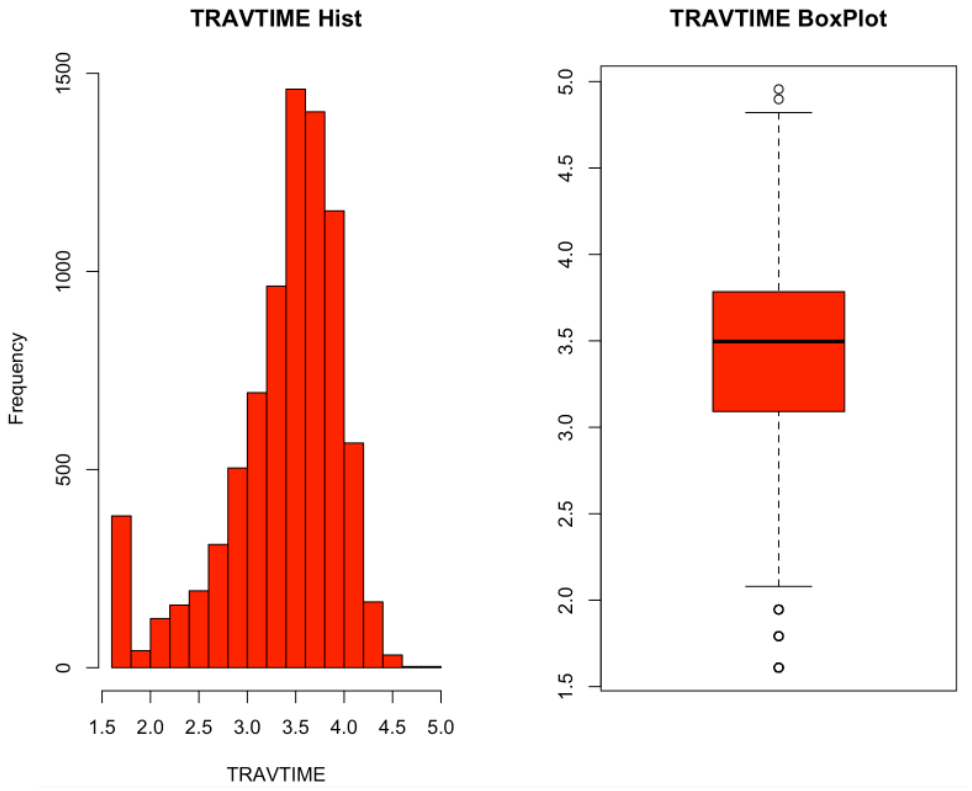
Figure 5 below shows the distribution of motor vehicle record points and the age of the car. Both of these have high counts early on, this is due to most drivers buying newer cars and newer cars are found on the road as well as most drivers have 1 or 2 points on their records. Very few will have high points on their records.

Figure 5



Because some of our data is skewed right some transformations would be required. For one, we'll use the log transformation of travel time as seen below in Figure 6.

Figure 6

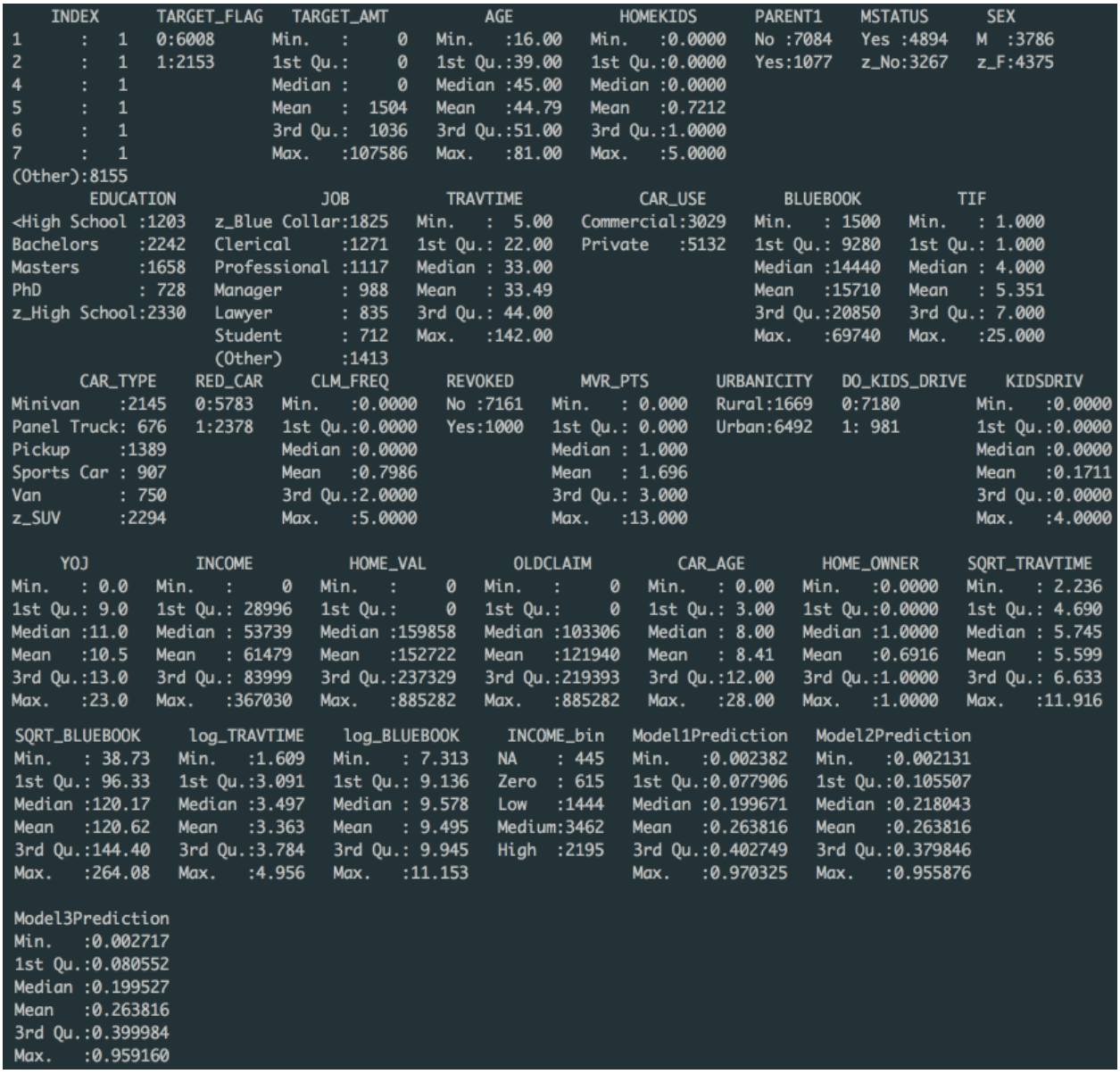


Data Preparation:

To being our data preparation, we'll impute the values needed for our missing variables. To do this, we'll use the simple average method and just replace our missing values with NA's.

To start, we need to impute the missing values in our data. In order to this we've using the MICE package in R; specifically, the PMM (Predictive Mean Matching) method. PMM only works on continuous variables therefore, we broke up our variables by their categories and then ran the imputation function. After rejoining our variables together, we see the results below in Figure 7. No NA's remain within our data set. It should be mentioned that every transformation, categorization, or imputation were also replicated on our test data set. This is to ensure proper predictability when deploying our model.

Figure 7

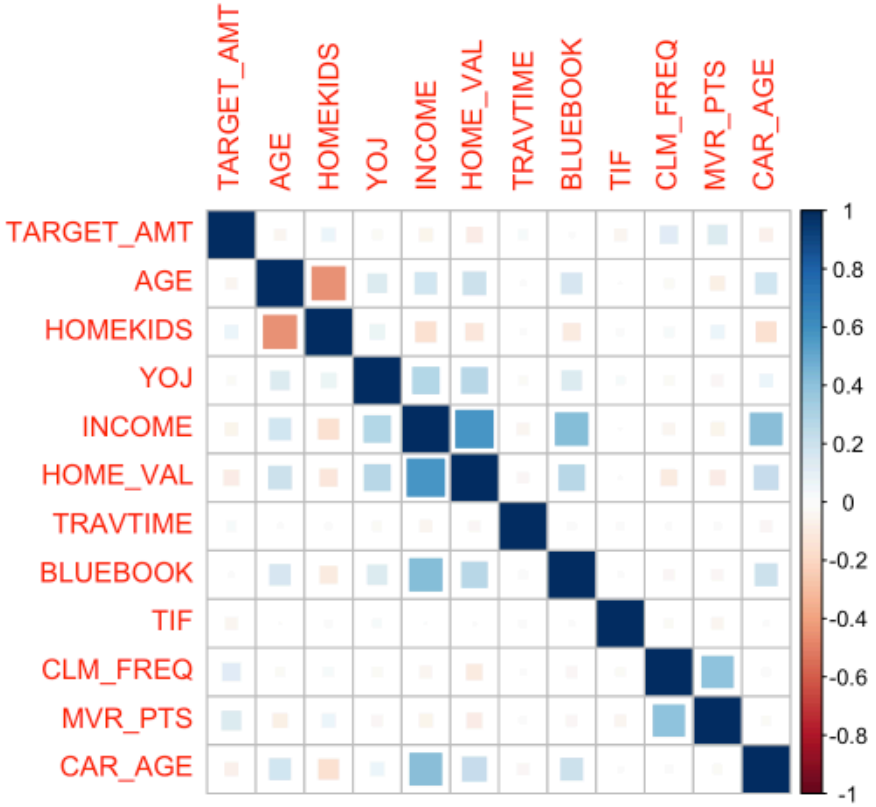


```
     INDEX      TARGET_FLAG   TARGET_AMT          AGE            HOMEKIDS      PARENT1    MSTATUS      SEX
1      :   1   0:6008      Min.   :     0   Min.   :16.00   Min.   :0.0000   No :7084   Yes :4894   M  :3786
2      :   1   1:2153      1st Qu.:     0   1st Qu.:39.00   1st Qu.:0.0000   Yes:1077   z_No:3267   z_F:4375
4      :   1               Median :     0   Median :45.00   Median :0.0000
5      :   1               Mean   :  1504   Mean   :44.79   Mean   :0.7212
6      :   1               3rd Qu.:  1036   3rd Qu.:51.00   3rd Qu.:1.0000
7      :   1               Max.   :107586   Max.   :81.00   Max.   :5.0000
(Other):8155
        EDUCATION                JOB          TRAVTIME          CAR_USE        BLUEBOOK          TIF
<High School :1203   z_Blue Collar:1825   Min.   :  5.00   Commercial:3029   Min.   : 1500   Min.   : 1.000
Bachelors    :2242   Clerical      :1271   1st Qu.: 22.00   Private   :5132   1st Qu.: 9280   1st Qu.: 1.000
Masters      :1658   Professional  :1117   Median : 33.00                     Median :14440   Median : 4.000
PhD          : 728   Manager       : 988   Mean   : 33.49                     Mean   :15710   Mean   : 5.351
z_High School:2330   Lawyer        : 835   3rd Qu.: 44.00                     3rd Qu.:20850   3rd Qu.: 7.000
                     Student       : 712   Max.   :142.00                     Max.   :69740   Max.   :25.000
                     (Other)       :1413
       CAR_TYPE    RED_CAR     CLM_FREQ       REVOKED        MVR_PTS       URBANICITY   DO_KIDS_DRIVE    KIDSDRIV
Minivan    :2145   0:5783   Min.   :0.0000   No :7161   Min.   : 0.000   Rural:1669   0:7180        Min.   :0.0000
Panel Truck: 676   1:2378   1st Qu.:0.0000   Yes:1000   1st Qu.: 0.000   Urban:6492   1: 981        1st Qu.:0.0000
Pickup     :1389            Median :0.0000              Median : 1.000                             Median :0.0000
Sports Car : 907            Mean   :0.7986              Mean   : 1.696                             Mean   :0.1711
Van        : 750            3rd Qu.:2.0000              3rd Qu.: 3.000                             3rd Qu.:0.0000
z_SUV      :2294            Max.   :5.0000              Max.   :13.000                             Max.   :4.0000

      YOJ           INCOME          HOME_VAL         OLDCLAIM          CAR_AGE        HOME_OWNER     SQRT_TRAVTIME
Min.   : 0.0   Min.   :     0   Min.   :     0   Min.   :     0   Min.   : 0.00   Min.   :0.0000   Min.   : 2.236
1st Qu.: 9.0   1st Qu.: 28996   1st Qu.:     0   1st Qu.:     0   1st Qu.: 3.00   1st Qu.:0.0000   1st Qu.: 4.690
Median :11.0   Median : 53739   Median :159858   Median :103306   Median : 8.00   Median :1.0000   Median : 5.745
Mean   :10.5   Mean   : 61479   Mean   :152722   Mean   :121940   Mean   : 8.41   Mean   :0.6916   Mean   : 5.599
3rd Qu.:13.0   3rd Qu.: 83999   3rd Qu.:237329   3rd Qu.:219393   3rd Qu.:12.00   3rd Qu.:1.0000   3rd Qu.: 6.633
Max.   :23.0   Max.   :367030   Max.   :885282   Max.   :885282   Max.   :28.00   Max.   :1.0000   Max.   :11.916

 SQRT_BLUEBOOK     log_TRAVTIME    log_BLUEBOOK     INCOME_bin   Model1Prediction   Model2Prediction
Min.   : 38.73   Min.   :1.609   Min.   : 7.313   NA    : 445   Min.   :0.002382   Min.   :0.002131
1st Qu.: 96.33   1st Qu.:3.091   1st Qu.: 9.136   Zero  : 615   1st Qu.:0.077906   1st Qu.:0.105507
Median :120.17   Median :3.497   Median : 9.578   Low   :1444   Median :0.199671   Median :0.218043
Mean   :120.62   Mean   :3.363   Mean   : 9.495   Medium:3462   Mean   :0.263816   Mean   :0.263816
3rd Qu.:144.40   3rd Qu.:3.784   3rd Qu.: 9.945   High  :2195   3rd Qu.:0.402749   3rd Qu.:0.379846
Max.   :264.08   Max.   :4.956   Max.   :11.153                 Max.   :0.970325   Max.   :0.955876

 Model3Prediction
 Min.   :0.002717
 1st Qu.:0.080552
 Median :0.199527
 Mean   :0.263816
 3rd Qu.:0.399984
 Max.   :0.959160
```

Next we also created some flags within our data set to ensure accuracy and real life applicability. If the car age variable is less than 0, we created a flag to mark it as 0; a car cannot have an age of less than 0. We then looked to normalize some of our variables that were heavily skewed. Travel time and the car bluebook value were the transformed via log and square root. These values are also present in the Figure above. In addition, we categorized our driver income variable into 5 bins: NAs, No Income, Low, Medium, and High. No variables were combined for this analysis. Figure 8 below showcases the correlation between our response and predictor variables. As shown, there is a negative correlation between driver age and

number of kids. This makes sense as older people tend to have more children. Home value also has a very positive correlation to income, which logically also makes sense.

Figure 8



Build Linear Regression Models:

From our cleansed data, we now begin building an optimal model for predicting our response variable: target flag. In our first try we used all the variables provided within our analysis to produce a model that predicts our expected results of approximately 0.27 for target flag and 1540 for target amount. Figure 9 below provides the summary of our first model. We used stepwise for this model. The model with the AIC value at its lowest 7359.4. Here, we can also see that the model intercept is negative. Although, this is the case, the rest of the variables make sense logically. The bluebook value is negative, this is because the higher the value of the car, the safer the driver tends to drive. TIF, marital status, and car type all play a big role in this model.

Figure 9

```
Call:
glm(formula = TARGET_FLAG ~ BLUEBOOK + TRAVTIME + KIDSDRIV +
    URBANICITY + CLM_FREQ + REVOKED + MVR_PTS + TIF + EDUCATION +
    MSTATUS + PARENT1 + CAR_USE + CAR_TYPE + JOB + INCOME_bin +
    HOME_VAL, family = binomial(), data = finalset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6048  -0.7159  -0.3999   0.6127   3.1231

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)            -3.687e+00  3.033e-01 -12.158  < 2e-16 ***
BLUEBOOK               -2.272e-05  4.699e-06  -4.835 1.33e-06 ***
TRAVTIME                1.480e-02  1.884e-03   7.855 4.00e-15 ***
KIDSDRIV                4.255e-01  5.514e-02   7.717 1.19e-14 ***
URBANICITYUrban         2.407e+00  1.132e-01  21.265  < 2e-16 ***
CLM_FREQ                1.484e-01  2.553e-02   5.813 6.13e-09 ***
REVOKEDYes              7.369e-01  8.038e-02   9.168  < 2e-16 ***
MVR_PTS                 1.078e-01  1.357e-02   7.944 1.96e-15 ***
TIF                    -5.501e-02  7.343e-03  -7.491 6.82e-14 ***
EDUCATIONBachelors     -3.817e-01  1.112e-01  -3.434 0.000595 ***
EDUCATIONMasters       -3.057e-01  1.629e-01  -1.876 0.060613 .
EDUCATIONPhD           -2.420e-01  1.965e-01  -1.231 0.218221
EDUCATIONz_High School  1.232e-02  9.686e-02   0.127 0.898811
MSTATUSz_No             4.822e-01  7.847e-02   6.145 7.99e-10 ***
PARENT1Yes              4.633e-01  9.432e-02   4.912 9.00e-07 ***
CAR_USEPrivate         -7.530e-01  9.185e-02  -8.198 2.45e-16 ***
CAR_TYPEPanel Truck     6.105e-01  1.509e-01   4.046 5.22e-05 ***
CAR_TYPEPickup          5.600e-01  1.008e-01   5.558 2.73e-08 ***
CAR_TYPESports Car      9.579e-01  1.077e-01   8.896  < 2e-16 ***
CAR_TYPEVan             6.491e-01  1.220e-01   5.319 1.04e-07 ***
CAR_TYPEz_SUV           7.233e-01  8.609e-02   8.402  < 2e-16 ***
JOBClerical             4.220e-01  1.963e-01   2.150 0.031543 *
JOBDoctor              -4.047e-01  2.664e-01  -1.519 0.128702
JOBHome Maker           5.562e-02  2.169e-01   0.256 0.797645
JOBLawyer               1.172e-01  1.687e-01   0.695 0.487348
JOBManager             -5.616e-01  1.710e-01  -3.285 0.001021 **
JOBProfessional         1.648e-01  1.781e-01   0.925 0.354828
JOBStudent              1.036e-01  2.218e-01   0.467 0.640284
JOBz_Blue Collar        3.147e-01  1.853e-01   1.699 0.089395 .
INCOME_bin2             6.796e-01  1.845e-01   3.684 0.000230 ***
INCOME_bin3             1.009e-01  1.484e-01   0.680 0.496248
INCOME_bin4             3.769e-02  1.356e-01   0.278 0.781015
INCOME_bin5            -2.878e-01  1.497e-01  -1.923 0.054503 .
HOME_VAL               -1.239e-06  3.241e-07  -3.822 0.000132 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9418.0  on 8160  degrees of freedom
Residual deviance: 7291.4  on 8127  degrees of freedom
AIC: 7359.4

Number of Fisher Scoring iterations: 5
```

For Model 2 we used automated variable selection. To do this, we used the regsubset() function in R and obtained a score for each variable. Based on the output in R, we were able to determine that the best model could be built using the following predictor variables:

- BLUEBOOK
- TRAVTIME
- URBANCITY
- REVOKED
- MVR_PTS
- TIF
- PARENT1
- CARUSE
- HOMEVAL

Figure 10 below outlines the summary statistics for Model 2.

Figure 10

```
Call:
glm(formula = TARGET_FLAG ~ BLUEBOOK + TRAVTIME + URBANICITY +
    REVOKED + MVR_PTS + TIF + PARENT1 + CAR_USE + HOME_VAL, family = binomial(),
    data = finalset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4983  -0.7500  -0.4721   0.7139   2.9961

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.078e+00  1.478e-01 -14.061  < 2e-16 ***
BLUEBOOK        -3.908e-05  3.678e-06 -10.626  < 2e-16 ***
TRAVTIME         1.464e-02  1.813e-03   8.075 6.73e-16 ***
URBANICITYUrban  2.155e+00  1.079e-01  19.982  < 2e-16 ***
REVOKEDYes       7.558e-01  7.736e-02   9.770  < 2e-16 ***
MVR_PTS          1.526e-01  1.234e-02  12.366  < 2e-16 ***
TIF             -5.259e-02  7.100e-03  -7.407 1.30e-13 ***
PARENT1Yes       7.741e-01  7.932e-02   9.760  < 2e-16 ***
CAR_USEPrivate  -8.986e-01  5.932e-02 -15.149  < 2e-16 ***
HOME_VAL        -3.004e-06  2.439e-07 -12.315  < 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: 9418.0  on 8160  degrees of freedom
Residual deviance: 7722.1  on 8151  degrees of freedom
AIC: 7742.1

Number of Fisher Scoring iterations: 5
```

As shown above, the model provides an AIC value of 7742.1. This score is higher than our first model. Further review also shows that this model may be a bit counterintuitive. The model here implies that the probability of getting into an accident is negative, however, the lowest this value should be is 0 (logically speaking). That being said, further analysis must be done for models 1 and 2. The remainder of the coefficients make logical sense here.

Model 3 was based on user selection. We selected user variables that we believed would provide the best predictive model. We also used log transformed variables for travel time and bluebook values. This is because in our initial data exploration, these values were heavily skewed. For this model, we utilized forward selection. For this model, we also observed a counterintuitive intercept value. Further analysis of the 3 models will need to be done to prove our models. The observed AIC value for the third model was the lowest yet at 7387.3. This could be indicative of the best fit model.

Figure 11

```
Call:
glm(formula = TARGET_FLAG ~ AGE + log_TRAVTIME + log_BLUEBOOK +
    DO_KIDS_DRIVE + URBANICITY + CLM_FREQ + REVOKED + MVR_PTS +
    TIF + EDUCATION + MSTATUS + PARENT1 + CAR_USE + CAR_TYPE +
    JOB + HOME_OWNER, family = binomial(), data = finalset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4665  -0.7165  -0.4068   0.6286   3.1653

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.350618   0.606093  -2.228 0.025854 *
AGE                  -0.003265   0.003682  -0.887 0.375137
log_TRAVTIME          0.398221   0.051231   7.773 7.66e-15 ***
log_BLUEBOOK         -0.356306   0.054594  -6.527 6.73e-11 ***
DO_KIDS_DRIVE1        0.679188   0.087698   7.745 9.59e-15 ***
URBANICITYUrban       2.357318   0.111601  21.123  < 2e-16 ***
CLM_FREQ              0.155606   0.025465   6.110 9.93e-10 ***
REVOKEDYes            0.739669   0.080202   9.223  < 2e-16 ***
MVR_PTS               0.113981   0.013523   8.429  < 2e-16 ***
TIF                  -0.054170   0.007316  -7.404 1.32e-13 ***
EDUCATIONBachelors   -0.496597   0.106839  -4.648 3.35e-06 ***
EDUCATIONMasters     -0.425993   0.159110  -2.677 0.007421 **
EDUCATIONPhD         -0.482383   0.191296  -2.522 0.011680 *
EDUCATIONz_High School -0.011854   0.094351  -0.126 0.900016
MSTATUSz_No           0.466788   0.081616   5.719 1.07e-08 ***
PARENT1Yes            0.410699   0.099681   4.120 3.79e-05 ***
CAR_USEPrivate       -0.761520   0.091637  -8.310  < 2e-16 ***
CAR_TYPEPanel Truck   0.508634   0.143760   3.538 0.000403 ***
CAR_TYPEPickup        0.567401   0.100153   5.665 1.47e-08 ***
CAR_TYPESports Car    0.950884   0.107832   8.818  < 2e-16 ***
```

```
CAR_TYPEVan             0.627210   0.121538   5.161 2.46e-07 ***
CAR_TYPEz_SUV           0.722788   0.085481   8.456  < 2e-16 ***
JOBClerical             0.599381   0.192776   3.109 0.001876 **
JOBDoctor              -0.369231   0.264161  -1.398 0.162188
JOBHome Maker           0.625116   0.193277   3.234 0.001219 **
JOBLawyer               0.181747   0.168402   1.079 0.280478
JOBManager             -0.478069   0.169613  -2.819 0.004823 **
JOBProfessional         0.251173   0.176921   1.420 0.155698
JOBStudent              0.484905   0.205643   2.358 0.018374 *
JOBz_Blue Collar        0.426132   0.184072   2.315 0.020612 *
HOME_OWNER             -0.309057   0.080287  -3.849 0.000118 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9418.0  on 8160  degrees of freedom
Residual deviance: 7325.3  on 8130  degrees of freedom
AIC: 7387.3

Number of Fisher Scoring iterations: 5
```

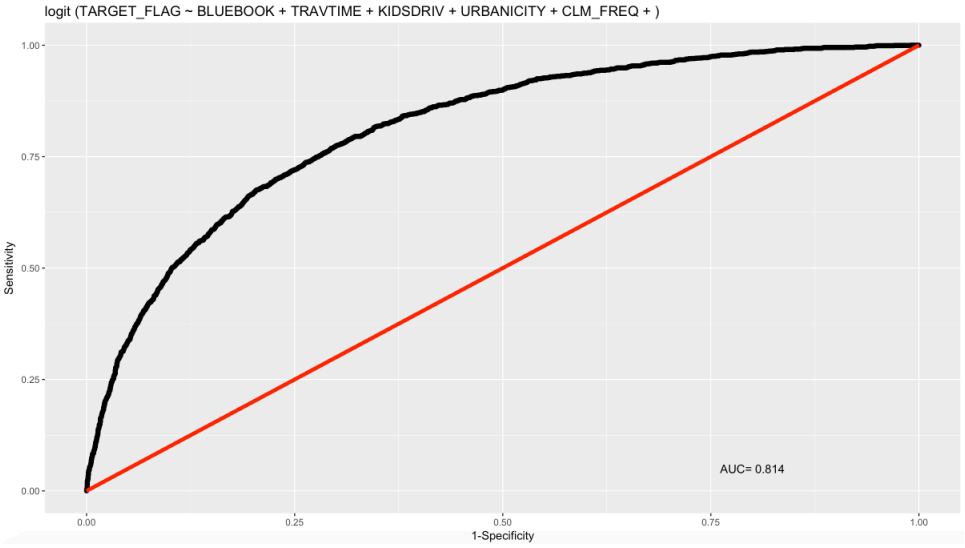
Select Best Model and Stand Alone Scoring:

To asses our model, we will use the AIC, Log Likelihood, and KS Statistic values. The table below outlines the model metrics. Figure 12 below provides these values as well as the ROC curves for each model.

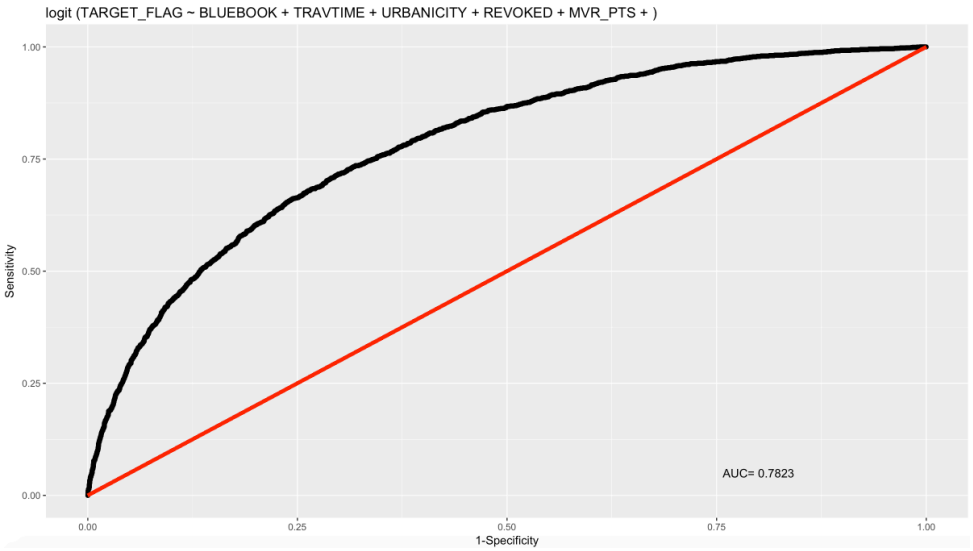
Figure 12

Model	AIC	Log Likelihood	KS Statistic
Model 1	7359.365	7291.365	0.4725
Model 2	7742.057	7722.057	0.4163
Model 3	7387.331	7325.331	0.4693

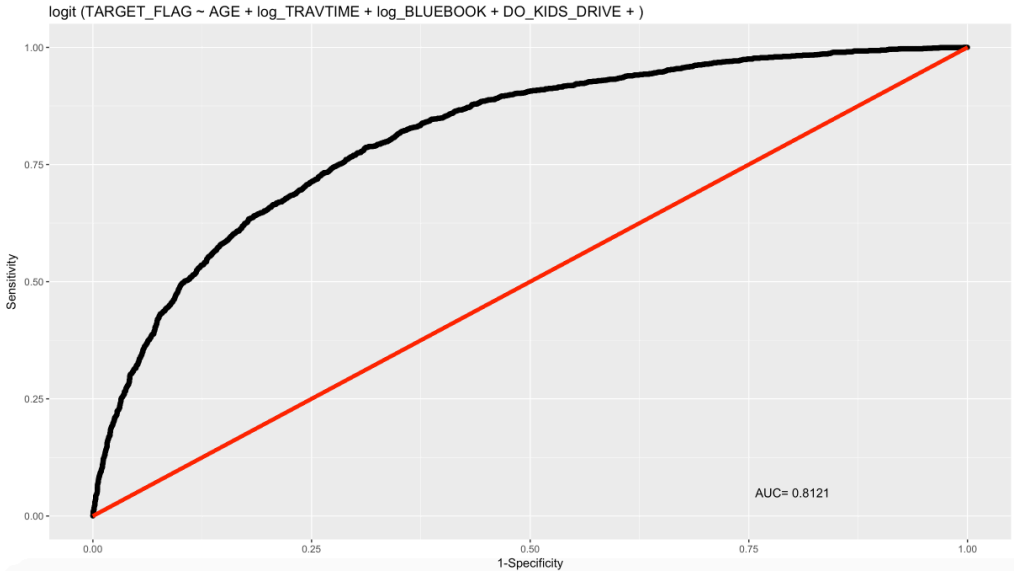
Model 1



Model 2



Model 3



Based on the values above the best model is either model 1 or model 3. This can be inferred by the AIC, as the two lowest values are listed here. The log likelihood for Model 2 is the highest of the bunch. This indicates that this model may be incomplete; there may be other variables that must be added to make this the most optimal predictive model.

As we mentioned earlier, all of our models seem a bit counterintuitive. There is a significant value for our beta parameter (intercept). This may be due to bias of an omitted variable, sometimes contained within the analysis or sometimes outside of our model. However, for the purposes of our analysis, we have decided to keep these models due to their ability to accurately predict values using our test data set.

For our scoring step, we used the predict() function in R to find the probability of one getting into an accident based on the variables provided. Based on our initial assessment, the expected value for our probability is 27%. Based on Model 3, our probability was 27.07%. After creating our scoring step in R, we used the bluebook value of the vehicle by car type. This is used to predict the target amount. Our expected value here should be approximately $1,540 of incurred expenses. Based on Model 3, our expected value is $1,544.57. These values prove that Model 3 is the most optimal model for our analysis.