Homework #4: Insurance Claim Prediction

Data 621 Business Analytics and Data Mining

Aadi Kalloo, Nathan Lim, Asher Meyers, Daniel Smilowitz, Logan Thomson

Due July 10, 2016

Contents

1 Data Exploration **2**

2 Data Preparation **8**

3 Model Creation **11**

3.1 Multiple Linear Regression 11

3.2 Binary Logistic Regression 14

4 Model Selection & Prediction **23**

Appendix A: Index-wise Results from Predictive Model **24**

Appendix B: R Code **61**

1 Data Exploration

The dataset of interest contains information about customers of an auto insurance company. The dataset has 8161 rows (each representing a customer) and 25 variables. There are 23 predictor variables and 2 response variables: `TARGET_FLAG`, a binary categorical variable representing whether each customer has been in an accident; and `TARGET_AMT`, a numerical variable indicating the cost of a crash that a customer was in. The class of variables read in from the dataset is presented below:

	Class	Levels
TARGET_FLAG	integer	-
TARGET_AMT	numeric	-
KIDSDRIV	integer	-
AGE	integer	-
HOMEKIDS	integer	-
YOJ	integer	-
INCOME	factor	6613
PARENT1	factor	2
HOME_VAL	factor	5107
MSTATUS	factor	2
SEX	factor	2
EDUCATION	factor	5
JOB	factor	9
TRAVTIME	integer	-
CAR_USE	factor	2
BLUEBOOK	factor	2789
TIF	integer	-
CAR_TYPE	factor	6
RED_CAR	factor	2
OLDCLAIM	factor	2857
CLM_FREQ	integer	-
REVOKED	factor	2
MVR_PTS	integer	-
CAR_AGE	integer	-
URBANICITY	factor	2

The very high number of levels for four of the variables (`INCOME`, `HOME_VAL`, `BLUEBOOK`, and `OLDCLAIM`) indicates that these variables are not in fact factors; investigation of the dataset indicates that these are dollar values interpreted as strings due to the presence of dollar signs and commas. The numerical values are extracted for these variables.

Additionally, there are 7 variables with only two levels. These are recast as binary variables as follows:

- `PARENT1`, `MSTATUS`, `RED_CAR`, and `REVOKED`: using 1 to indicate Yes
- `SEX`: using 1 to indicate Male
- `CAR_USE`: using 1 to indicate Commercial
- `URBANICITY`: using 1 to indicate Highly Urban/ Urban

Finally, there are three categorical variables – factors with more than two levels. Dummy variables are created for each of these, as follows:

- `EDUCATION`: using High School as the base case
- `CAR_TYPE`: using Minivan as the base case
- `JOB`: using the blank value as the base case

A summary of each variable is presented below:

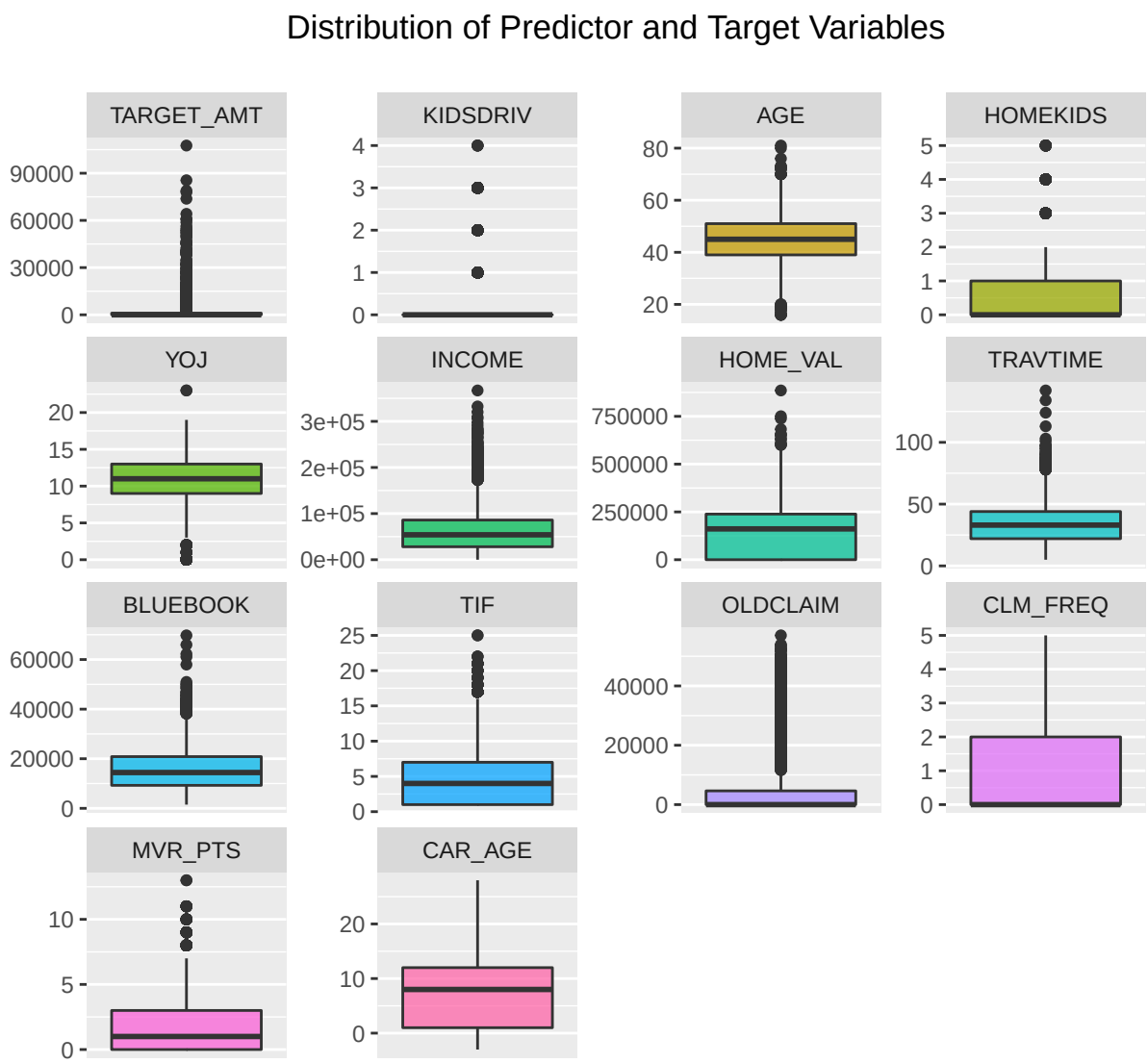
	MEAN	MEDIAN	IQR	SKEW	r_{FLAG}	r_{AMT}	NAs
TARGET_FLAG	0.26	0	1	1.07	1	0.54	0
TARGET_AMT	1504	0	1036	8.71	0.54	1	0
KIDSDRIV	0.17	0	0	3.35	0.09	0.05	0
AGE	44.79	45	12	-0.03	-0.11	-0.05	6
HOMEKIDS	0.72	0	1	1.34	0.11	0.06	0
YOJ	10.5	11	4	-1.2	-0.07	-0.02	454
INCOME	61898	54028	57889	1.19	-0.14	-0.06	445
PARENT1	0.13	0	0	2.17	0.16	0.1	0
HOME_VAL	154867	161160	238724	0.49	-0.18	-0.09	464
MSTATUS	0.6	1	1	-0.41	-0.13	-0.1	0
SEX	0.46	0	1	0.14	-0.02	0.01	0
TRAVTIME	33.49	33	22	0.45	0.05	0.03	0
CAR_USE	0.37	0	1	0.53	0.14	0.1	0
BLUEBOOK	15710	14440	11570	0.79	-0.11	0	0
TIF	5.35	4	6	0.89	-0.08	-0.04	0
RED_CAR	0.29	0	1	0.92	-0.02	0	0
OLDCLAIM	4037	0	4636	3.12	0.14	0.08	0
CLM_FREQ	0.8	0	2	1.21	0.22	0.12	0
REVOKED	0.12	0	0	2.3	0.15	0.06	0
MVR_PTS	1.7	1	3	1.35	0.23	0.14	0
CAR_AGE	8.33	8	11	0.28	-0.11	-0.06	510
URBANICITY	0.8	1	0	-1.46	0.22	0.12	0
HSDropout	0.15	0	0	1.99	0.06	0.04	0
Bachelors	0.27	0	1	1.01	-0.05	-0.02	0
Masters	0.2	0	0	1.48	-0.09	-0.05	0
PhD	0.09	0	0	2.88	-0.06	-0.02	0
Panel_Truck	0.08	0	0	3.03	0	0.04	0
Pickup	0.17	0	0	1.75	0.05	0.02	0
Sports_Car	0.11	0	0	2.47	0.06	0.03	0
Van	0.09	0	0	2.82	0	0.01	0
SUV	0.28	0	1	0.97	0.05	0.01	0
Professional	0.14	0	0	2.11	-0.04	0	0
Blue_Collar	0.14	0	0	2.11	-0.04	0	0
Clerical	0.16	0	0	1.9	0.04	0	0
Doctor	0.03	0	0	5.49	-0.05	-0.03	0
Lawyer	0.1	0	0	2.62	-0.06	-0.03	0
Manager	0.12	0	0	2.32	-0.12	-0.07	0
Home_Maker	0.08	0	0	3.13	0.01	0	0
Student	0.09	0	0	2.92	0.07	0.02	0

From the table above, it is clear that there are four variables with missing values, with the proportion of values missing ranging from less than $< 0.1\%$ to roughly 6.2%; these missing values will need to either be imputed or excluded from the dataset before modeling. The variables exhibit varying levels of skewness, with a few extreme values.

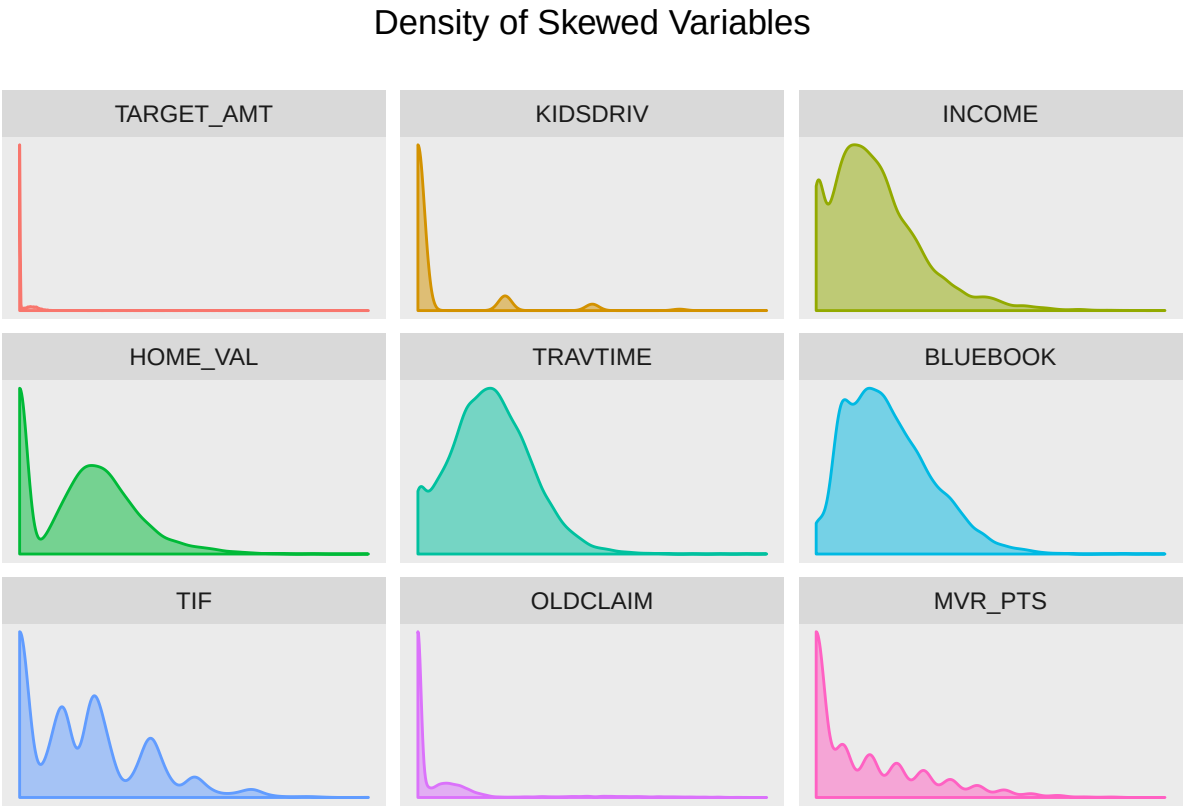
The large number of binary variables in the dataset makes graphical visualization of the distribution of all variables not particularly useful. The proportion of binary variables having a value of 0 or 1 is presented in the table below:

	0	1
TARGET_FLAG	0.74	0.26
PARENT1	0.87	0.13
MSTATUS	0.4	0.6
SEX	0.54	0.46
CAR_USE	0.63	0.37
RED_CAR	0.71	0.29
REVOKED	0.88	0.12
URBANICITY	0.2	0.8
HSDropout	0.85	0.15
Bachelors	0.73	0.27
Masters	0.8	0.2
PhD	0.91	0.09
Panel_Truck	0.92	0.08
Pickup	0.83	0.17
Sports_Car	0.89	0.11
Van	0.91	0.09
SUV	0.72	0.28
Professional	0.86	0.14
Blue_Collar	0.86	0.14
Clerical	0.84	0.16
Doctor	0.97	0.03
Lawyer	0.9	0.1
Manager	0.88	0.12
Home_Maker	0.92	0.08
Student	0.91	0.09

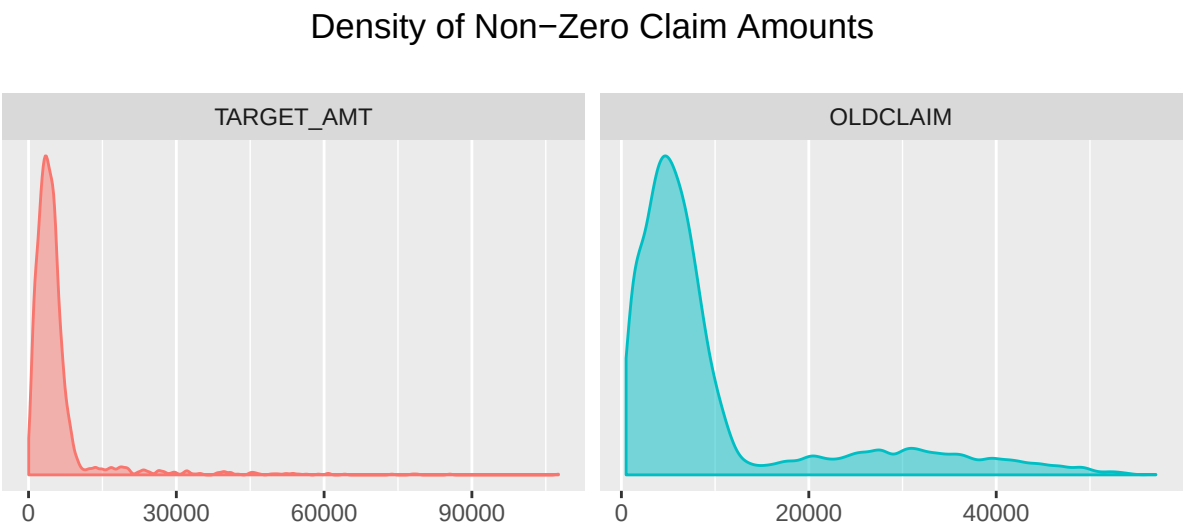
The remaining variables are vizualized below in boxplots:



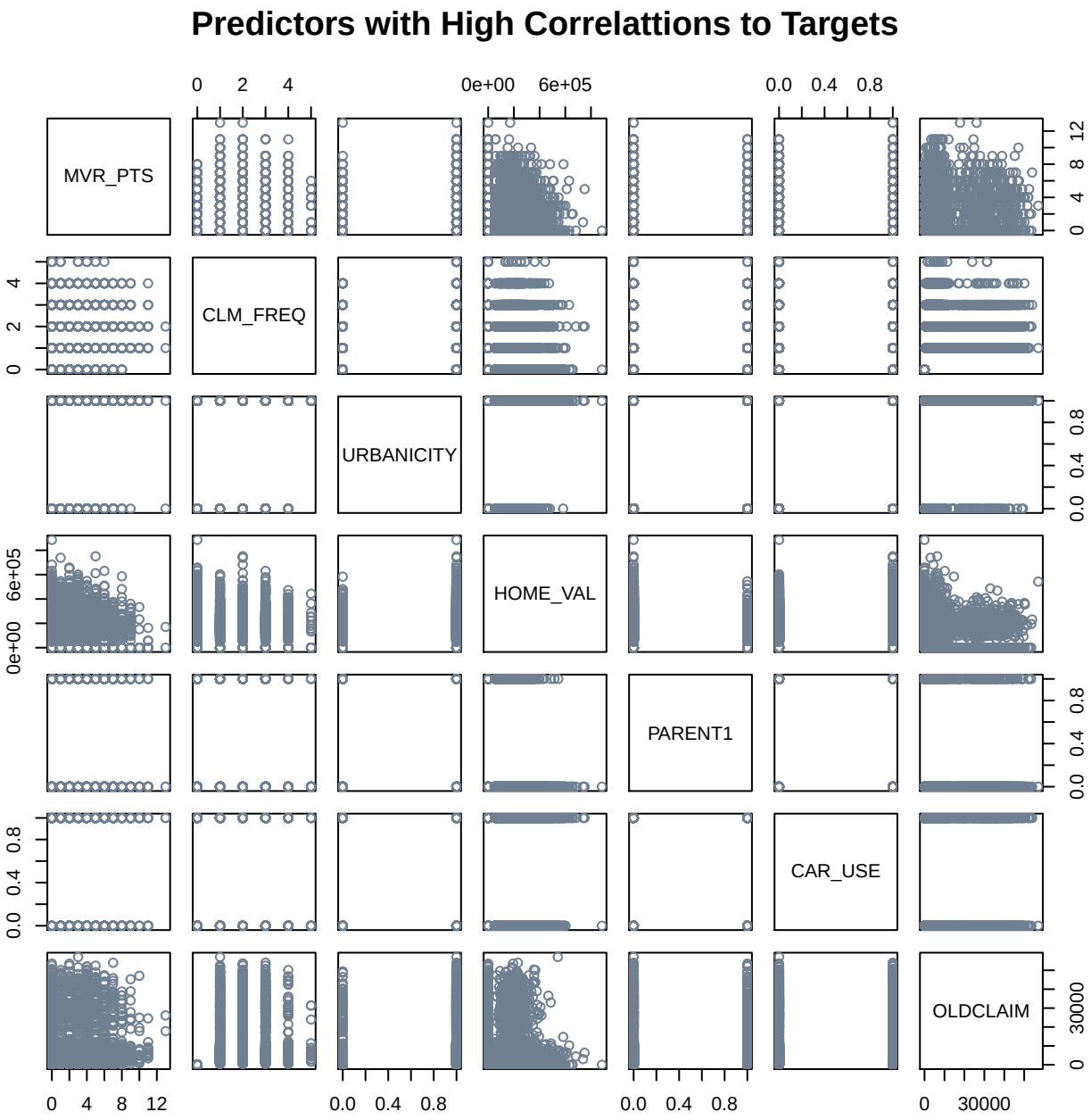
The boxplots illustrate the high skewness of the distributions of the predictors `KIDSDRIV`, `INCOME`, `HOME_VAL`, `TRAVTIME`, `BLUEBOOK`, `TIF`, `OLDCLAIM` and `MVR_PTS`. The target `TARGET_AMT` is also highly skewed – this makes sense, as this value is 0 for any customers without claims. Density plots of these variables are presented below:



Since `TARGET_AMT` and `OLDCLAIM` have such high concentrations at values of zero, separate plots for these two variables are created with values of zero removed:

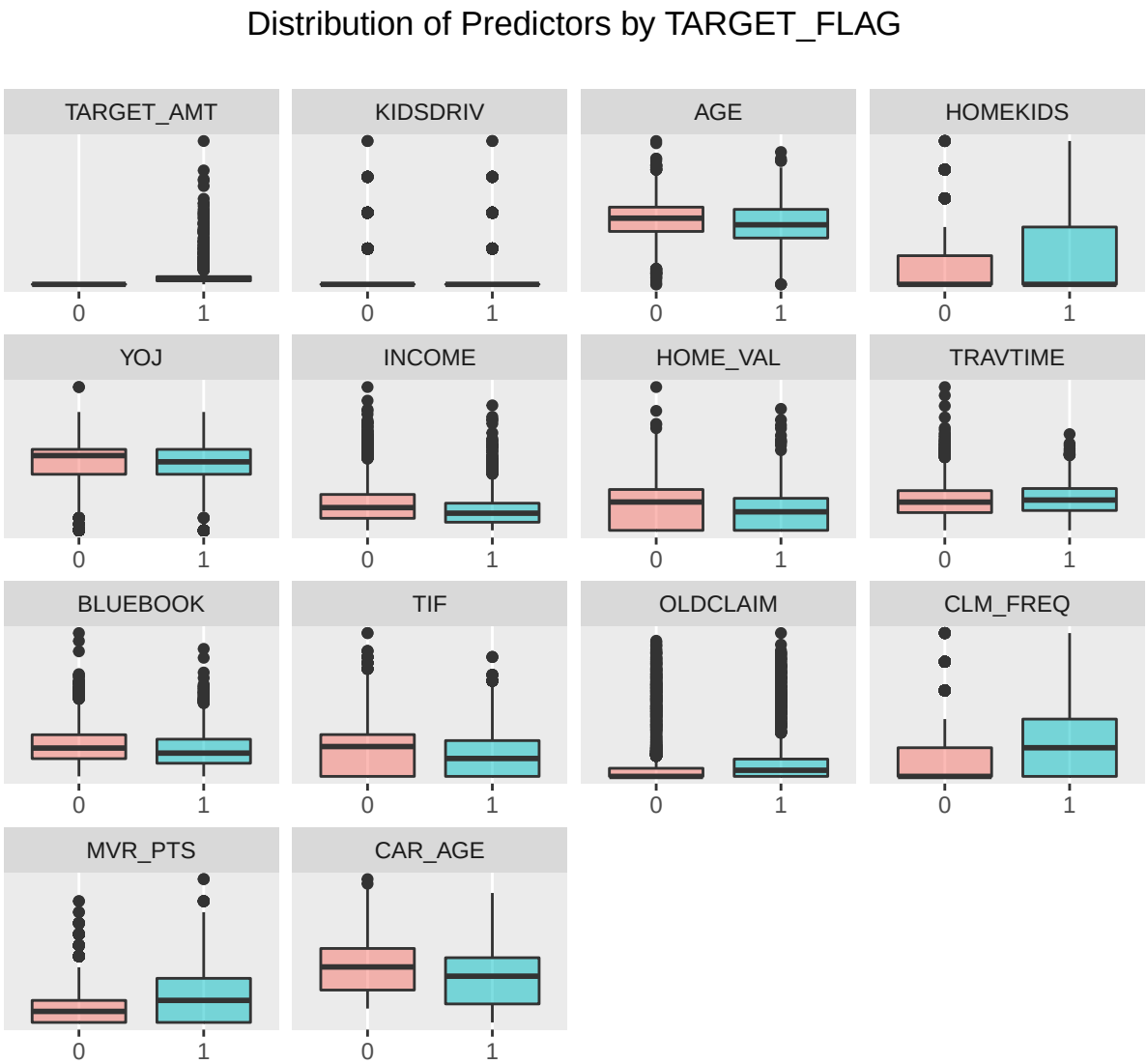


The 8 predictors with the highest correlation to `TARGET_FLAG` and the 8 predictors with the highest correlation to `TARGET_AMT` share 7 predictors. The correlation between these variables is investigated and plotted below:



There appears to be evidence of possible multicollinearity between `HOME_VAL` and `OLDCLAIM`.

Finally, boxplots are also prepared for non-binary variables split by `TARGET_FLAG`:



Interestingly, there are only a few variables with immediately visible difference in median value based on `TARGET_FLAG` values, while the range of predictor values for each flag value differs noticeably.

2 Data Preparation

As stated in Part 1, four predictor variables have missing values. Because these missing values represent, in many cases, a non-menial proportion of the dataset at large, they are imputed so the cases containing missing values can be included in modeling. Due to the skewness illustrated by some of the variables with missing data (`INCOME` and `HOME_VAL`), the median is used to avoid any bias introduced into the mean by the skewness of these variables' distribution. For the remaining two variables, the median is also used for imputation for consistency.

Additionally, there is one instance of a vehicle's `CAR_AGE` being -3. Since this variable represents the age of the car in years, this is a nonsensical value. This instance of this variable is removed before imputation.

A summary of the variables with imputation of median values is presented below:

	MEAN	MEDIAN	IQR	SKEW	r_{FLAG}	r_{AMT}	NAs
TARGET_FLAG	0.26	0	1	1.07	1	0.53	0
TARGET_AMT	1504	0	1036	8.71	0.53	1	0
KIDSDRIV	0.17	0	0	3.35	0.1	0.06	0
AGE	44.79	45	12	-0.03	-0.1	-0.04	0
HOMEKIDS	0.72	0	1	1.34	0.12	0.06	0
YOJ	10.53	11	4	-1.26	-0.07	-0.02	0
INCOME	61469	54028	53597	1.24	-0.14	-0.06	0
PARENT1	0.13	0	0	2.17	0.16	0.1	0
HOME_VAL	155225	161160	233352	0.49	-0.18	-0.08	0
MSTATUS	0.6	1	1	-0.41	-0.14	-0.09	0
SEX	0.46	0	1	0.14	-0.02	0.01	0
TRAVTIME	33.49	33	22	0.45	0.05	0.03	0
CAR_USE	0.37	0	1	0.53	0.14	0.1	0
BLUEBOOK	15710	14440	11570	0.79	-0.1	0	0
TIF	5.35	4	6	0.89	-0.08	-0.05	0
RED_CAR	0.29	0	1	0.92	-0.01	0.01	0
OLDCLAIM	4037	0	4636	3.12	0.14	0.07	0
CLM_FREQ	0.8	0	2	1.21	0.22	0.12	0
REVOKED	0.12	0	0	2.3	0.15	0.06	0
MVR_PTS	1.7	1	3	1.35	0.22	0.14	0
CAR_AGE	8.31	8	8	0.3	-0.1	-0.06	0
URBANICITY	0.8	1	0	-1.46	0.22	0.12	0
HSDropout	0.15	0	0	1.99	0.05	0.03	0
Bachelors	0.27	0	1	1.01	-0.04	-0.02	0
Masters	0.2	0	0	1.48	-0.08	-0.04	0
PhD	0.09	0	0	2.88	-0.07	-0.02	0
Panel_Truck	0.08	0	0	3.03	0	0.03	0
Pickup	0.17	0	0	1.75	0.06	0.02	0
Sports_Car	0.11	0	0	2.47	0.06	0.02	0
Van	0.09	0	0	2.82	0	0.02	0
SUV	0.28	0	1	0.97	0.05	0.01	0
Professional	0.14	0	0	2.11	-0.04	0	0
Blue_Collar	0.14	0	0	2.11	-0.04	0	0
Clerical	0.16	0	0	1.9	0.03	0.01	0
Doctor	0.03	0	0	5.49	-0.06	-0.03	0
Lawyer	0.1	0	0	2.62	-0.06	-0.03	0
Manager	0.12	0	0	2.32	-0.11	-0.06	0
Home_Maker	0.08	0	0	3.13	0.01	-0.01	0
Student	0.09	0	0	2.92	0.08	0.02	0

The creation of additional or combined predictors did not yield any predictors that would improve models, however due to the skewness of some predictors, log transformations were conducted.

In the table below, we can see the effect of each of the log transformation on the correlations to the two target variables. For many variables, the transformations do not improve the correlation, but the log transformations of `INCOME` and `HOME_VAL` may fit our models better.

	r_{FLAG}	$\log r_{FLAG}$	r_{AMT}	$\log r_{AMT}$
TARGET_FLAG	1	1	0.53	0.53
TARGET_AMT	0.53	1	1	0.58
KIDSDRIV	0.1	0.11	0.06	0.06
AGE	-0.1	-0.12	-0.04	-0.05
HOMEKIDS	0.12	0.13	0.06	0.07
YOJ	-0.07	-0.08	-0.02	-0.02
INCOME	-0.14	-0.11	-0.06	-0.03
PARENT1	0.16	0.16	0.1	0.1
HOME_VAL	-0.18	-0.15	-0.08	-0.07
MSTATUS	-0.14	-0.14	-0.09	-0.09
SEX	-0.02	-0.02	0.01	0.01
TRAVTIME	0.05	0.06	0.03	0.03
CAR_USE	0.14	0.14	0.1	0.1
BLUEBOOK	-0.1	-0.11	0	0
TIF	-0.08	-0.08	-0.05	-0.05
RED_CAR	-0.01	-0.01	0.01	0.01
OLDCLAIM	0.14	0.24	0.07	0.13
CLM_FREQ	0.22	0.24	0.12	0.13
REVOKED	0.15	0.15	0.06	0.06
MVR_PTS	0.22	0.17	0.14	0.11
CAR_AGE	-0.1	-0.09	-0.06	-0.05
URBANICITY	0.22	0.22	0.12	0.12
HSDropout	0.05	0.05	0.03	0.03
Bachelors	-0.04	-0.04	-0.02	-0.02
Masters	-0.08	-0.08	-0.04	-0.04
PhD	-0.07	-0.07	-0.02	-0.02
Panel_Truck	0	0	0.03	0.03
Pickup	0.06	0.06	0.02	0.02
Sports_Car	0.06	0.06	0.02	0.02
Van	0	0	0.02	0.02
SUV	0.05	0.05	0.01	0.01
Professional	-0.04	-0.04	0	0
Blue_Collar	-0.04	-0.04	0	0
Clerical	0.03	0.03	0.01	0.01
Doctor	-0.06	-0.06	-0.03	-0.03
Lawyer	-0.06	-0.06	-0.03	-0.03
Manager	-0.11	-0.11	-0.06	-0.06
Home_Maker	0.01	0.01	-0.01	-0.01
Student	0.08	0.08	0.02	0.02

3 Model Creation

3.1 Multiple Linear Regression

3.1.1 Full Model

	Estimate	Std. Error	t value	Pr(>|t|)
KIDSDRIV	318	113.1	2.81	0.004964
AGE	5.159	7.066	0.7301	0.4653
HOMEKIDS	78.22	65.36	1.197	0.2314
YOJ	-3.682	15.09	-0.2441	0.8072
INCOME	-0.004656	0.001799	-2.588	0.009662
PARENT1	573.6	202	2.839	0.004532
HOME_VAL	-0.0005489	0.0005908	-0.9292	0.3528
MSTATUS	-571.4	144.9	-3.944	8.075e-05
SEX	365.2	183.8	1.987	0.04694
TRAVTIME	12.05	3.222	3.739	0.0001857
CAR_USE	814.8	162.9	5.002	5.799e-07
BLUEBOOK	0.01446	0.008623	1.676	0.0937
TIF	-48.35	12.18	-3.97	7.246e-05
RED_CAR	-48.02	149.1	-0.3221	0.7474
OLDCLAIM	-0.01066	0.007436	-1.433	0.1518
CLM_FREQ	141.3	55.03	2.568	0.01024
REVOKED	551.4	173.5	3.178	0.00149
MVR_PTS	175	25.92	6.752	1.562e-11
CAR_AGE	-26.89	12.8	-2.101	0.03566
URBANICITY	1650	139.1	11.86	3.427e-32
HSDropout	108.8	171.5	0.6345	0.5258
Bachelors	-197.7	156.7	-1.261	0.2072
Masters	-85.32	230	-0.3709	0.7107
PhD	156.2	290.8	0.5371	0.5912
Panel_Truck	164.3	270.8	0.6066	0.5441
Pickup	342.9	169.4	2.024	0.043
Sports_Car	1018	217.8	4.676	2.979e-06
Van	470	211.4	2.224	0.02619
SUV	748	179.3	4.172	3.055e-05
Professional	80.35	196.1	0.4098	0.682
Clerical	80.59	188.9	0.4266	0.6697
Doctor	-698.7	388.8	-1.797	0.07239
Lawyer	-0.3399	256.8	-0.001324	0.9989
Manager	-802	203.1	-3.949	7.899e-05
Home_Maker	-50.72	260.2	-0.1949	0.8455
Student	-175.4	231.7	-0.7568	0.4492
(Intercept)	-796	463.7	-1.717	0.08606

(Dispersion parameter for gaussian family taken to be 20653324)

Null deviance:	1.806e+11 on 8160 degrees of freedom
Residual deviance:	1.678e+11 on 8124 degrees of freedom

3.1.2 Log-Transformed Full Model

	Estimate	Std. Error	t value	Pr(>|t|)
KIDSDRIV	0.779	0.1249	6.238	4.643e-10
AGE	-0.001144	0.007798	-0.1467	0.8834

	Estimate	Std. Error	t value	Pr(>|t|)
HOMEKIDS	0.09154	0.07213	1.269	0.2044
YOJ	-0.02696	0.01665	-1.619	0.1055
INCOME	-5.625e-06	1.985e-06	-2.834	0.004613
PARENT1	0.9949	0.223	4.462	8.222e-06
HOME_VAL	-2.127e-06	6.52e-07	-3.263	0.001108
MSTATUS	-0.925	0.1599	-5.786	7.494e-09
SEX	0.1936	0.2028	0.9546	0.3398
TRAVTIME	0.0259	0.003556	7.284	3.543e-13
CAR_USE	1.616	0.1798	8.986	3.14e-19
BLUEBOOK	-3.082e-05	9.516e-06	-3.239	0.001206
TIF	-0.103	0.01344	-7.661	2.05e-14
RED_CAR	-0.03608	0.1645	-0.2193	0.8264
OLDCLAIM	-2.989e-05	8.207e-06	-3.642	0.0002724
CLM_FREQ	0.404	0.06073	6.653	3.066e-11
REVOKED	1.958	0.1915	10.23	2.125e-24
MVR_PTS	0.2799	0.02861	9.784	1.75e-22
CAR_AGE	-0.004512	0.01412	-0.3195	0.7494
URBANICITY	3.804	0.1535	24.79	8.396e-131
HSDropout	-0.03817	0.1892	-0.2017	0.8401
Bachelors	-0.8804	0.1729	-5.091	3.64e-07
Masters	-0.9068	0.2539	-3.572	0.0003564
PhD	-0.8084	0.3209	-2.519	0.01177
Panel_Truck	0.5264	0.2988	1.762	0.07817
Pickup	0.8583	0.187	4.591	4.486e-06
Sports_Car	1.845	0.2404	7.677	1.813e-14
Van	0.8601	0.2332	3.688	0.0002278
SUV	1.346	0.1979	6.801	1.112e-11
Professional	-0.1106	0.2164	-0.5112	0.6092
Clerical	0.3074	0.2085	1.474	0.1404
Doctor	-0.9105	0.4291	-2.122	0.03387
Lawyer	-0.1213	0.2834	-0.4279	0.6687
Manager	-1.457	0.2241	-6.5	8.48e-11
Home_Maker	0.113	0.2872	0.3935	0.6939
Student	-0.005919	0.2557	-0.02315	0.9815
(Intercept)	-4.561	0.5117	-8.913	6.054e-19

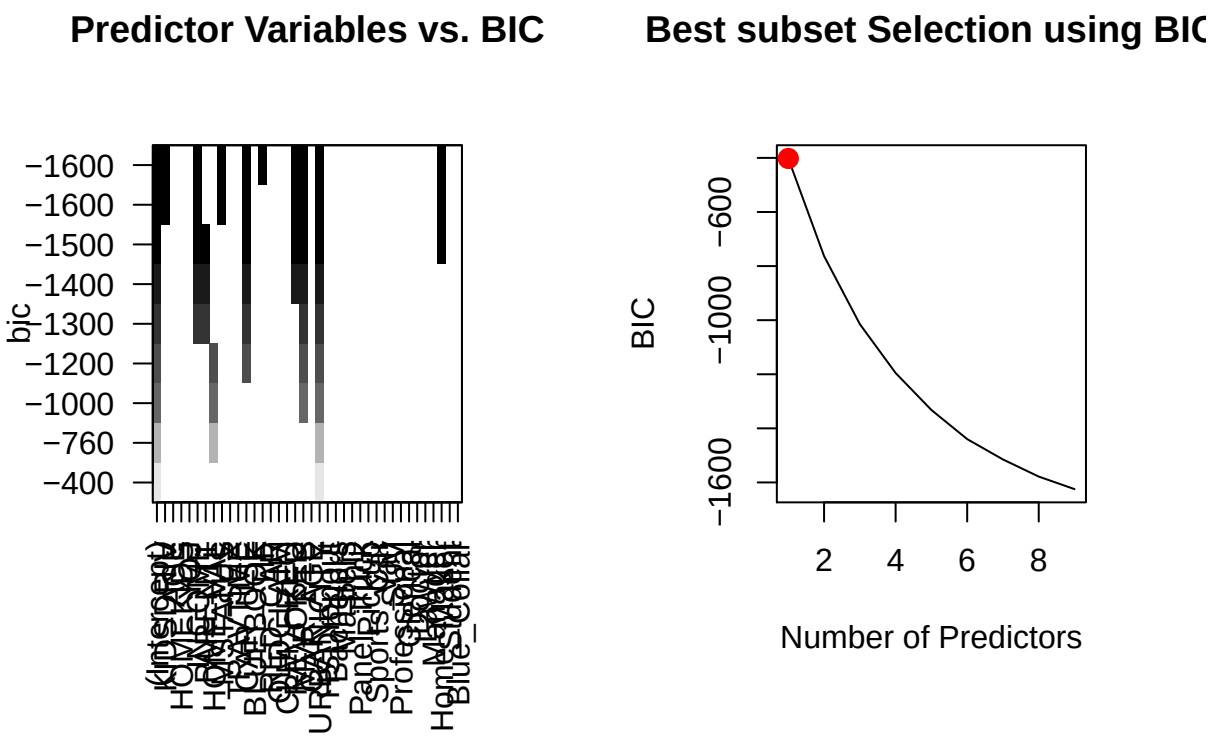
(Dispersion parameter for gaussian family taken to be 25.15463)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	204356 on 8124 degrees of freedom

3.1.3 Bayesian Information Criteria

```
Reordering variables and trying again:
```

```
[1] 9
```



	Estimate	Std. Error	t value	Pr(>|t|)
BLUEBOOK	-6.701e-05	7.448e-06	-8.997	2.835e-19
(Intercept)	-0.1542	0.1327	-1.161	0.2455

(Dispersion parameter for gaussian family taken to be 32.08896)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	261814 on 8159 degrees of freedom

	Estimate	Std. Error	t value	Pr(>|t|)
log(BLUEBOOK)	-0.9649	0.09934	-9.713	3.485e-22
(Intercept)	7.955	0.9453	8.415	4.588e-17

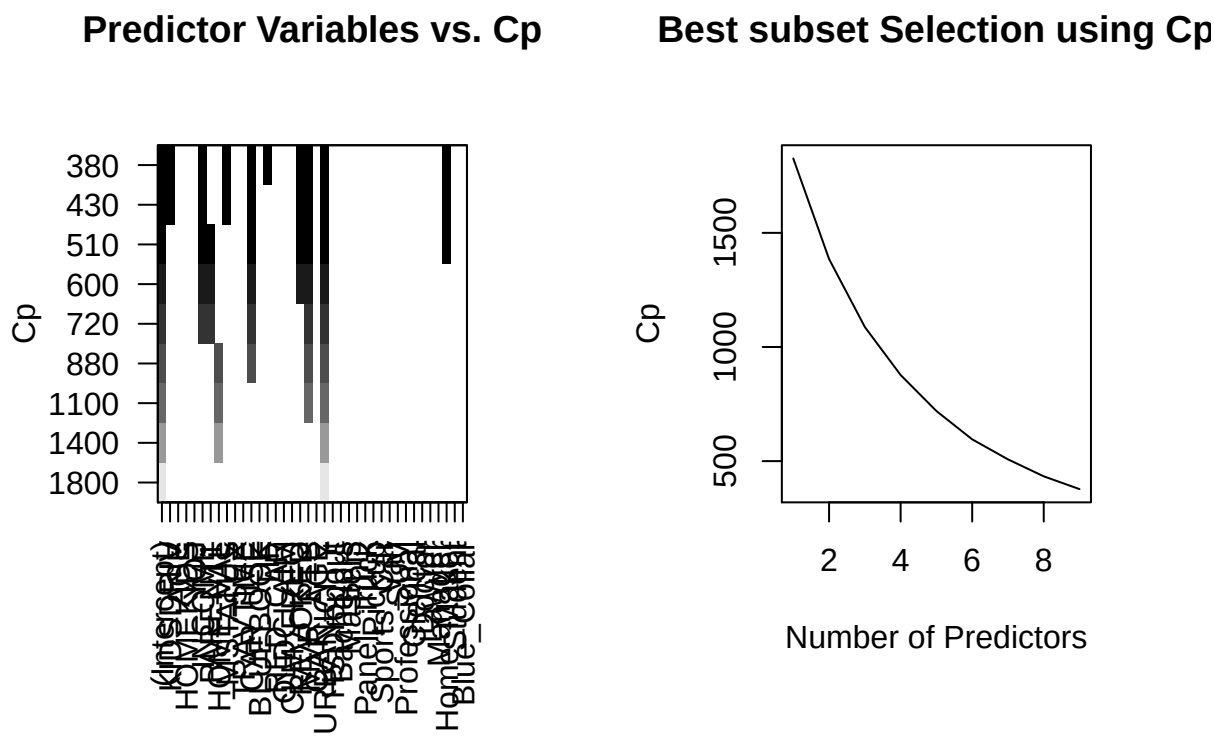
(Dispersion parameter for gaussian family taken to be 32.03687)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	261389 on 8159 degrees of freedom

3.1.4 Mallow's C_p

```
Reordering variables and trying again:
```

```
[1] 9
```



	Estimate	Std. Error	t value	Pr(>|t|)
BLUEBOOK	-6.701e-05	7.448e-06	-8.997	2.835e-19
(Intercept)	-0.1542	0.1327	-1.161	0.2455

(Dispersion parameter for gaussian family taken to be 32.08896)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	261814 on 8159 degrees of freedom

	Estimate	Std. Error	t value	Pr(>|t|)
log(BLUEBOOK)	-0.9649	0.09934	-9.713	3.485e-22
(Intercept)	7.955	0.9453	8.415	4.588e-17

(Dispersion parameter for gaussian family taken to be 32.03687)

Null deviance:	264412 on 8160 degrees of freedom
Residual deviance:	261389 on 8159 degrees of freedom

3.2 Binary Logistic Regression

3.2.1 Full Model

	Estimate	Std. Error	z value	Pr(>|z|)
KIDSDRIV	0.389	0.06119	6.357	2.056e-10

	Estimate	Std. Error	z value	Pr(>|z|)
AGE	-0.001005	0.004018	-0.2502	0.8025
HOMEKIDS	0.04965	0.03712	1.337	0.1811
YOJ	-0.011	0.008585	-1.281	0.2003
INCOME	-3.587e-06	1.078e-06	-3.326	0.0008796
PARENT1	0.3813	0.1095	3.481	0.0004991
HOME_VAL	-1.304e-06	3.418e-07	-3.815	0.0001362
MSTATUS	-0.4937	0.08354	-5.91	3.412e-09
SEX	0.07723	0.1002	0.771	0.4407
TRAVTIME	0.01464	0.001882	7.778	7.362e-15
CAR_USE	0.7769	0.09101	8.537	1.379e-17
BLUEBOOK	-2.074e-05	5.258e-06	-3.945	7.99e-05
TIF	-0.05556	0.007344	-7.566	3.849e-14
OLDCLAIM	-1.397e-05	3.91e-06	-3.574	0.0003515
CLM_FREQ	0.1958	0.02853	6.864	6.685e-12
REVOKED	0.8876	0.09132	9.719	2.495e-22
MVR_PTS	0.113	0.0136	8.308	9.704e-17
CAR_AGE	-0.0005727	0.007548	-0.07587	0.9395
URBANICITY	2.386	0.1129	21.14	3.67e-99
HSDropout	-0.006142	0.09479	-0.0648	0.9483
Bachelors	-0.4135	0.08908	-4.642	3.458e-06
Masters	-0.4523	0.134	-3.375	0.0007376
PhD	-0.3481	0.1715	-2.03	0.04238
Panel_Truck	0.5037	0.1582	3.184	0.001454
Pickup	0.5326	0.09999	5.327	9.999e-08
Sports_Car	1.024	0.1299	7.888	3.077e-15
Van	0.5906	0.1254	4.709	2.494e-06
SUV	0.767	0.1112	6.897	5.296e-12
Professional	-0.07245	0.1109	-0.6536	0.5134
Clerical	0.1324	0.1051	1.26	0.2075
Doctor	-0.5477	0.2599	-2.108	0.03507
Lawyer	-0.01933	0.1516	-0.1275	0.8986
Manager	-0.7582	0.1229	-6.168	6.907e-10
Home_Maker	-0.01721	0.1483	-0.1161	0.9076
Student	-0.07218	0.1282	-0.5628	0.5735
(Intercept)	-3.36	0.2819	-11.92	9.275e-33

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7300 on 8125 degrees of freedom

Table 20: Table continues below

KIDSDRIV	AGE	HOMEKIDS	YOJ	INCOME	PARENT1	HOME_VAL	MSTATUS	SEX
1.346	1.467	2.178	1.454	2.482	1.932	1.866	2.055	2.961

Table 21: Table continues below

TRAVTIME	CAR_USE	BLUEBOOK	TIF	OLDCLAIM	CLM_FREQ	REVOKED	MVR_PTS
1.039	2.414	2.195	1.01	1.648	1.465	1.316	1.161

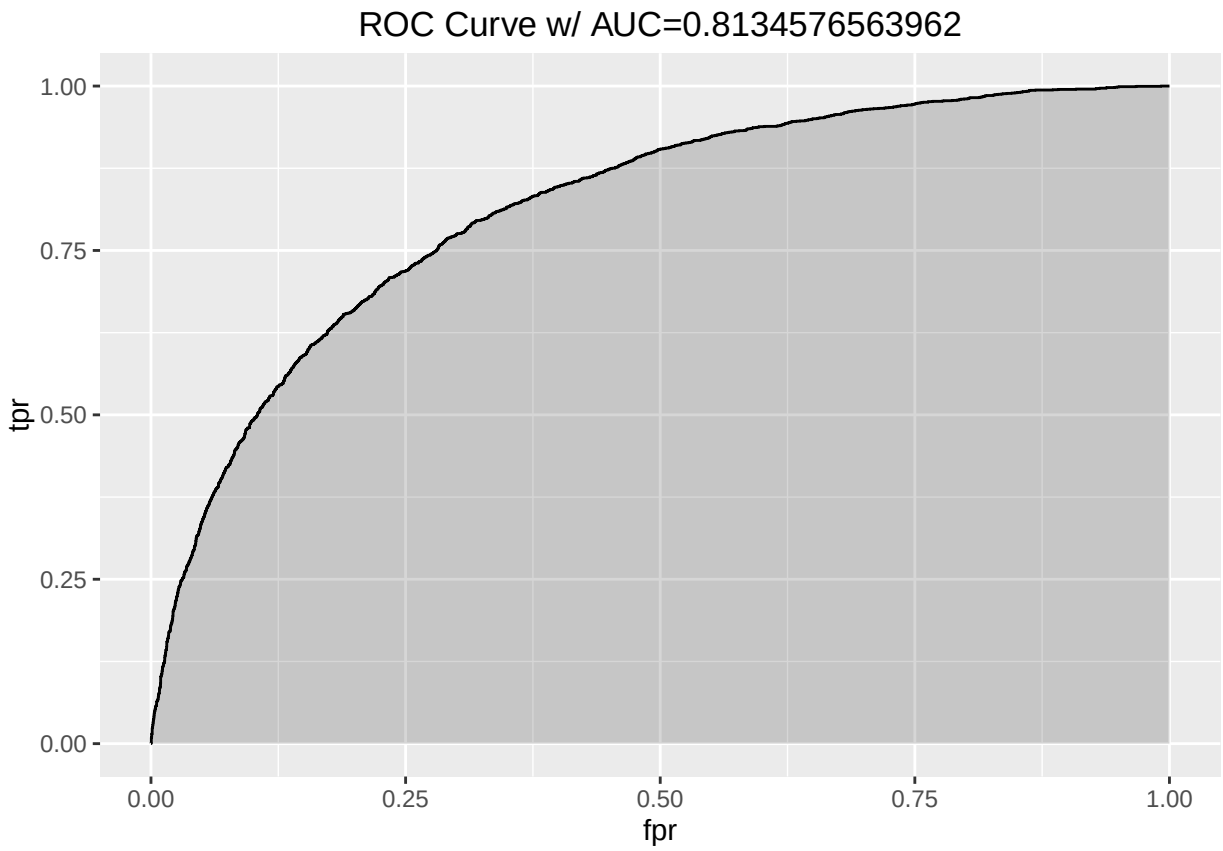
Table 22: Table continues below

CAR_AGE	URBANICITY	HSDropout	Bachelors	Masters	PhD	Panel_Truck
2.018	1.142	1.461	1.821	3.199	2.335	2.412

Table 23: Table continues below

Pickup	Sports_Car	Van	SUV	Professional	Clerical	Doctor	Lawyer
1.818	2.2	1.652	3.081	1.653	1.781	1.493	2.218

Manager	Home_Maker	Student
1.425	1.937	1.753



	Estimate	Std. Error	t value	Pr(>|t|)
KIDSDRIV	318.6	113.1	2.817	0.004862
AGE	5.203	7.064	0.7366	0.4614
HOMEKIDS	78.08	65.35	1.195	0.2322
YOJ	-3.722	15.08	-0.2467	0.8051
INCOME	-0.004656	0.001799	-2.588	0.009661
PARENT1	574.2	202	2.842	0.004487
HOME_VAL	-0.0005446	0.0005906	-0.9221	0.3565
MSTATUS	-571.1	144.9	-3.942	8.131e-05
SEX	337.3	162.1	2.08	0.03752
TRAVTIME	12.04	3.222	3.736	0.0001881
CAR_USE	814.7	162.9	5.001	5.82e-07
BLUEBOOK	0.01453	0.008619	1.686	0.09185

	Estimate	Std. Error	t value	Pr(>|t|)
TIF	-48.33	12.18	-3.968	7.295e-05
OLDCLAIM	-0.01067	0.007436	-1.436	0.1512
CLM_FREQ	141.1	55.03	2.565	0.01033
REVOKED	551.5	173.5	3.178	0.001487
MVR_PTS	175	25.92	6.752	1.562e-11
CAR_AGE	-26.93	12.8	-2.105	0.03536
URBANICITY	1649	139	11.86	3.509e-32
HSDropout	108.8	171.5	0.6348	0.5256
Bachelors	-198.2	156.7	-1.265	0.2059
Masters	-85.26	230	-0.3707	0.7109
PhD	156.3	290.7	0.5376	0.5909
Panel_Truck	163.5	270.7	0.604	0.5458
Pickup	342.6	169.4	2.022	0.04318
Sports_Car	1019	217.8	4.681	2.904e-06
Van	469.8	211.3	2.223	0.02625
SUV	750	179.2	4.185	2.882e-05
Professional	79.6	196.1	0.406	0.6848
Clerical	80.72	188.9	0.4273	0.6692
Doctor	-700.6	388.8	-1.802	0.07157
Lawyer	-0.5333	256.8	-0.002077	0.9983
Manager	-803	203	-3.955	7.723e-05
Home_Maker	-50.41	260.2	-0.1937	0.8464
Student	-176.8	231.7	-0.7629	0.4455
(Intercept)	-799.6	463.5	-1.725	0.08457

Table 26: Fitting linear model: TARGET_AMT ~ . - RED_CAR - Blue_Collar

Observations	Residual Std. Error	R^2	Adjusted R^2
8161	4544	0.07074	0.06674

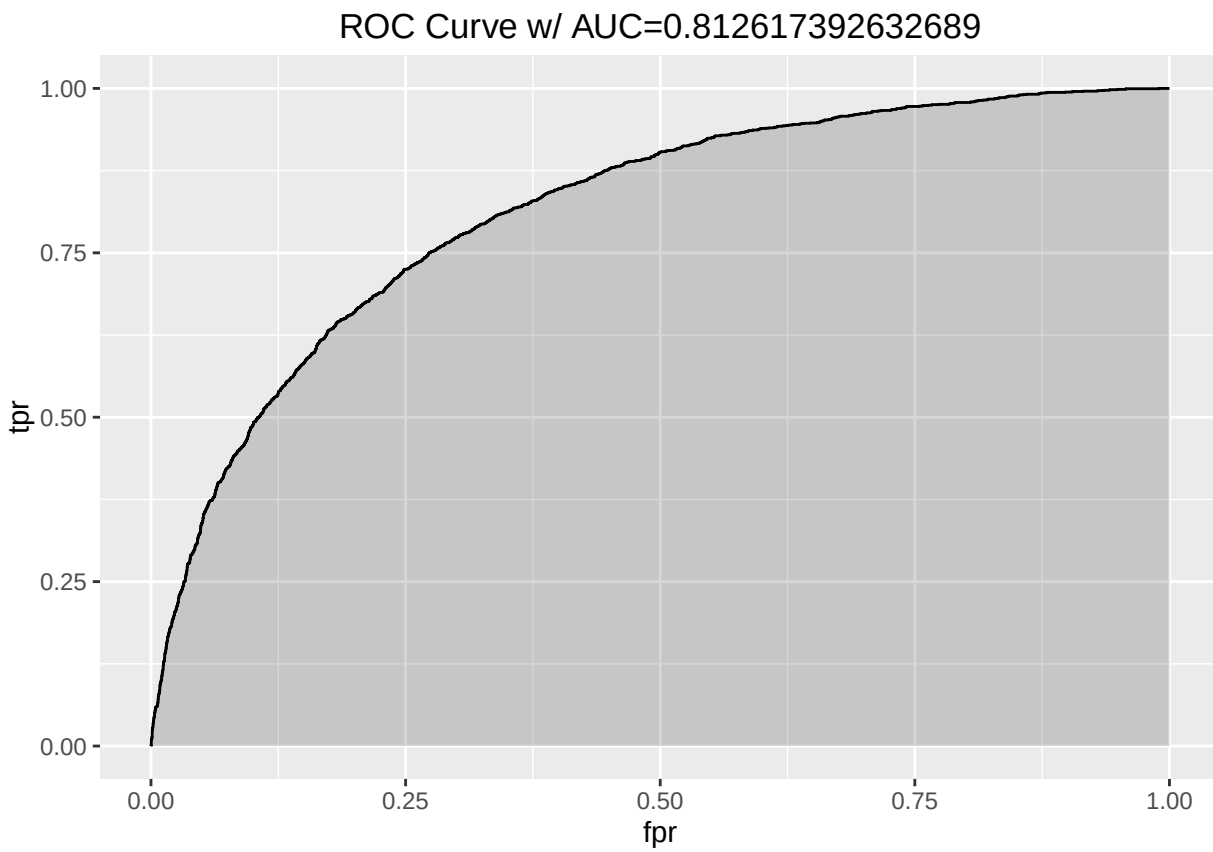
3.2.2 Reduced Model

	Estimate	Std. Error	z value	Pr(>|z|)
KIDSDRIV	0.4208	0.05505	7.645	2.097e-14
INCOME	-3.943e-06	9.833e-07	-4.01	6.065e-05
PARENT1	0.4713	0.09395	5.017	5.253e-07
HOME_VAL	-1.322e-06	3.306e-07	-3.998	6.38e-05
MSTATUS	-0.466	0.07875	-5.918	3.259e-09
TRAVTIME	0.01454	0.001878	7.742	9.749e-15
CAR_USE	0.7758	0.07156	10.84	2.176e-27
BLUEBOOK	-2.303e-05	4.698e-06	-4.902	9.466e-07
TIF	-0.05488	0.007327	-7.491	6.859e-14
OLDCLAIM	-1.388e-05	3.902e-06	-3.556	0.0003771
CLM_FREQ	0.1953	0.02847	6.861	6.834e-12
REVOKED	0.8924	0.0911	9.796	1.17e-22
MVR_PTS	0.1141	0.01356	8.416	3.904e-17
URBANICITY	2.373	0.1124	21.12	5.556e-99
Bachelors	-0.4557	0.07519	-6.061	1.357e-09
Masters	-0.5044	0.09108	-5.537	3.071e-08
PhD	-0.5374	0.1343	-4.002	6.292e-05
Panel_Truck	0.5901	0.1422	4.149	3.345e-05
Pickup	0.5433	0.09803	5.542	2.988e-08
Sports_Car	0.9732	0.1064	9.149	5.757e-20
Van	0.634	0.1193	5.312	1.083e-07

	Estimate	Std. Error	z value	Pr(>|z|)
SUV	0.7166	0.08519	8.412	4.016e-17
Manager	-0.7376	0.1065	-6.929	4.236e-12
(Intercept)	-3.389	0.1712	-19.8	2.884e-87

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7313 on 8137 degrees of freedom

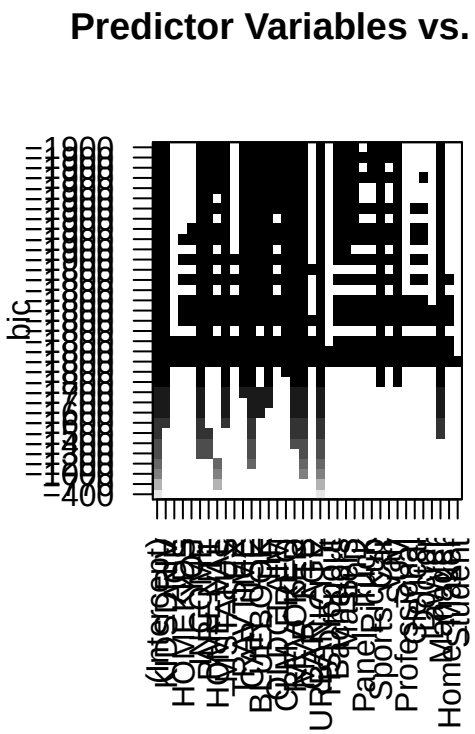


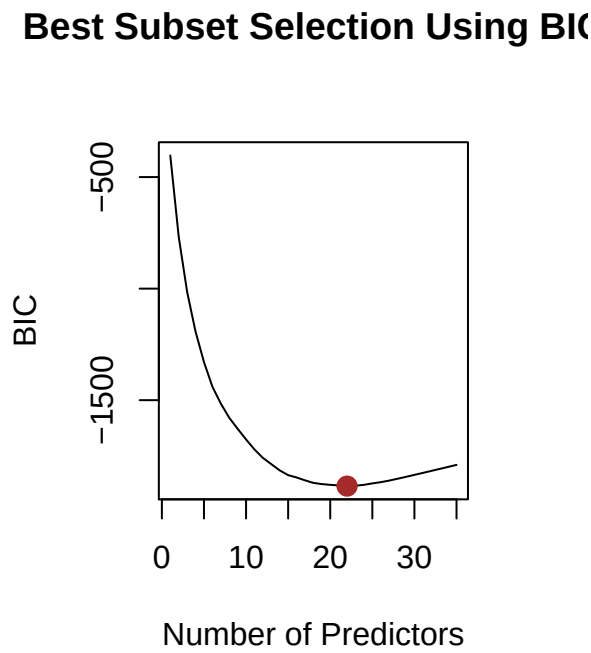
	Estimate	Std. Error	t value	Pr(>|t|)
KIDSDRIV	383.6	102	3.761	0.0001705
INCOME	-0.005582	0.001312	-4.254	2.123e-05
PARENT1	660.3	176.2	3.747	0.00018
MSTATUS	-586.9	119.1	-4.927	8.522e-07
SEX	374.4	146.3	2.56	0.01049
TRAVTIME	11.94	3.217	3.711	0.0002077
CAR_USE	887.3	114	7.787	7.735e-15
BLUEBOOK	0.015	0.00719	2.086	0.03703
TIF	-46.64	12.15	-3.837	0.0001255
CLM_FREQ	108.6	48.78	2.227	0.02596
REVOKED	458.8	154.8	2.964	0.003041
MVR_PTS	175.2	25.77	6.799	1.129e-11
CAR_AGE	-33.85	10.04	-3.371	0.0007525
URBANICITY	1627	137	11.87	3.098e-32
Sports_Car	874	203.2	4.301	1.723e-05
SUV	602.5	163.1	3.694	0.0002226
Manager	-821.3	161	-5.103	3.426e-07
(Intercept)	-462	274.9	-1.681	0.09287

Table 30: Fitting linear model: TARGET_AMT ~ KIDSDRIV + INCOME + PARENT1 + MSTATUS + SEX + TRAVTIME + CAR_USE + BLUEBOOK + TIF + CLM_FREQ + REVOKED + MVR_PTS + CAR_AGE + URBANICITY + Sports_Car + SUV + Manager

Observations	Residual Std. Error	R^2	Adjusted R^2
8161	4545	0.06836	0.06642

3.2.3 Best Subsets



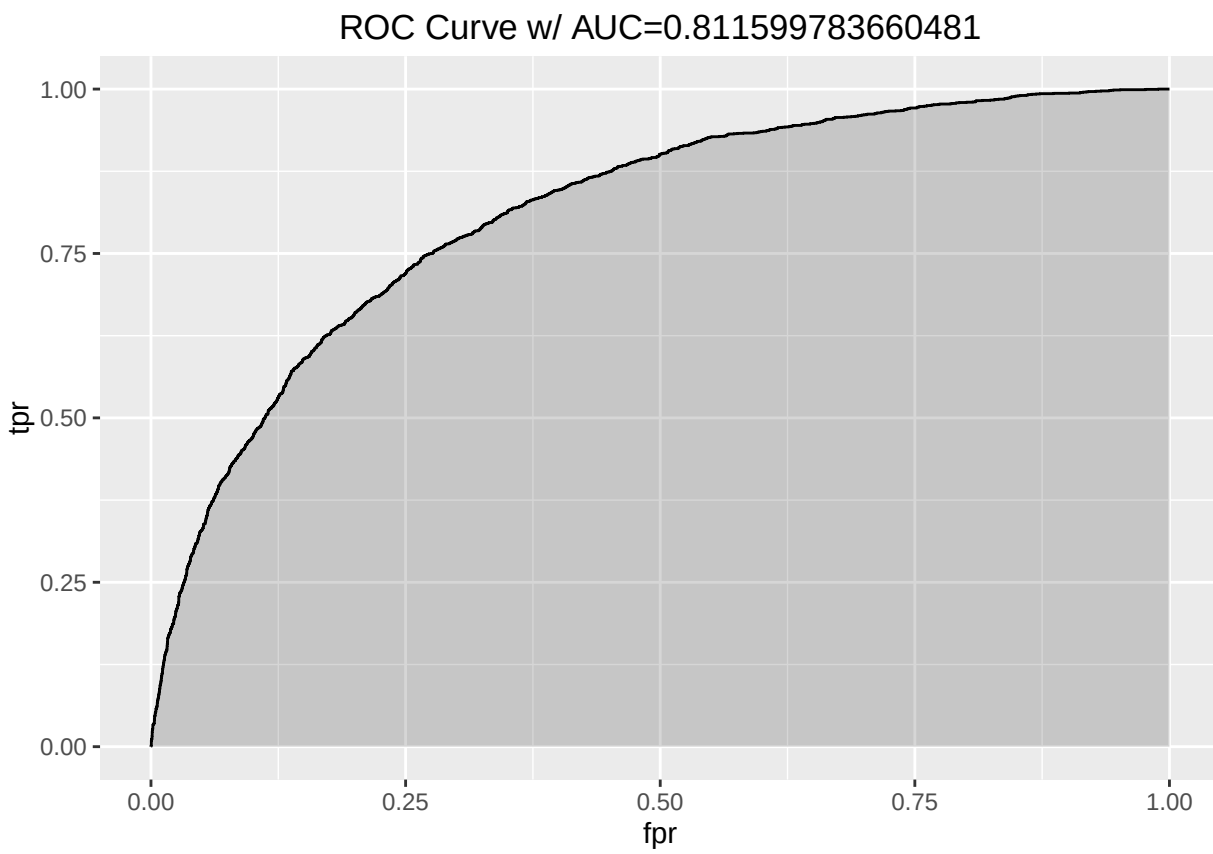


	Estimate	Std. Error	z value	Pr(>|z|)
KIDSDRIV	0.4208	0.05494	7.659	1.869e-14
INCOME	-4.075e-06	9.836e-07	-4.143	3.432e-05
PARENT1	0.463	0.09375	4.939	7.852e-07
HOME_VAL	-1.299e-06	3.308e-07	-3.926	8.628e-05
MSTATUS	-0.4725	0.07873	-6.002	1.95e-09
TRAVTIME	0.01453	0.001874	7.752	9.06e-15
CAR_USE	0.8795	0.06694	13.14	1.965e-39
BLUEBOOK	-1.548e-05	4.281e-06	-3.617	0.0002982
TIF	-0.05494	0.007327	-7.498	6.468e-14
OLDCLAIM	-1.416e-05	3.901e-06	-3.631	0.0002825
CLM_FREQ	0.1987	0.02843	6.988	2.778e-12
REVOKED	0.8913	0.09103	9.791	1.229e-22
MVR_PTS	0.1141	0.01354	8.425	3.596e-17
URBANICITY	2.37	0.1124	21.09	1.065e-98
Bachelors	-0.4615	0.07516	-6.141	8.22e-10
Masters	-0.4694	0.09048	-5.188	2.122e-07
PhD	-0.5011	0.1336	-3.751	0.0001763

	Estimate	Std. Error	z value	Pr(>|z|)
Pickup	0.4033	0.09127	4.419	9.909e-06
Sports_Car	0.8892	0.1035	8.591	8.592e-18
Van	0.4308	0.1083	3.978	6.936e-05
SUV	0.627	0.08135	7.708	1.28e-14
Manager	-0.7177	0.1063	-6.753	1.447e-11
(Intercept)	-3.412	0.1704	-20.02	3.875e-89

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7330 on 8138 degrees of freedom



3.2.4 Asher's Models

	Estimate	Std. Error	z value	Pr(>|z|)
KIDSDRIV	0.389	0.06119	6.357	2.056e-10
AGE	-0.001005	0.004018	-0.2502	0.8025
HOMEKIDS	0.04965	0.03712	1.337	0.1811
YOJ	-0.011	0.008585	-1.281	0.2003
INCOME	-3.587e-06	1.078e-06	-3.326	0.0008796
PARENT1	0.3813	0.1095	3.481	0.0004991
HOME_VAL	-1.304e-06	3.418e-07	-3.815	0.0001362
MSTATUS	-0.4937	0.08354	-5.91	3.412e-09
SEX	0.07723	0.1002	0.771	0.4407
TRAVTIME	0.01464	0.001882	7.778	7.362e-15
CAR_USE	0.7769	0.09101	8.537	1.379e-17
BLUEBOOK	-2.074e-05	5.258e-06	-3.945	7.99e-05

	Estimate	Std. Error	z value	Pr(>|z|)
TIF	-0.05556	0.007344	-7.566	3.849e-14
OLDCLAIM	-1.397e-05	3.91e-06	-3.574	0.0003515
CLM_FREQ	0.1958	0.02853	6.864	6.685e-12
REVOKED	0.8876	0.09132	9.719	2.495e-22
MVR_PTS	0.113	0.0136	8.308	9.704e-17
CAR_AGE	-0.0005727	0.007548	-0.07587	0.9395
URBANICITY	2.386	0.1129	21.14	3.67e-99
HSDropout	-0.006142	0.09479	-0.0648	0.9483
Bachelors	-0.4135	0.08908	-4.642	3.458e-06
Masters	-0.4523	0.134	-3.375	0.0007376
PhD	-0.3481	0.1715	-2.03	0.04238
Panel_Truck	0.5037	0.1582	3.184	0.001454
Pickup	0.5326	0.09999	5.327	9.999e-08
Sports_Car	1.024	0.1299	7.888	3.077e-15
Van	0.5906	0.1254	4.709	2.494e-06
SUV	0.767	0.1112	6.897	5.296e-12
Professional	-0.07245	0.1109	-0.6536	0.5134
Clerical	0.1324	0.1051	1.26	0.2075
Doctor	-0.5477	0.2599	-2.108	0.03507
Lawyer	-0.01933	0.1516	-0.1275	0.8986
Manager	-0.7582	0.1229	-6.168	6.907e-10
Home_Maker	-0.01721	0.1483	-0.1161	0.9076
Student	-0.07218	0.1282	-0.5628	0.5735
(Intercept)	-3.36	0.2819	-11.92	9.275e-33

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7300 on 8125 degrees of freedom

Table 35: Table continues below

KIDSDRIV	AGE	HOMEKIDS	YOJ	INCOME	PARENT1	HOME_VAL	MSTATUS	SEX
1.346	1.467	2.178	1.454	2.482	1.932	1.866	2.055	2.961

Table 36: Table continues below

TRAVTIME	CAR_USE	BLUEBOOK	TIF	OLDCLAIM	CLM_FREQ	REVOKED	MVR_PTS
1.039	2.414	2.195	1.01	1.648	1.465	1.316	1.161

Table 37: Table continues below

CAR_AGE	URBANICITY	HSDropout	Bachelors	Masters	PhD	Panel_Truck
2.018	1.142	1.461	1.821	3.199	2.335	2.412

Table 38: Table continues below

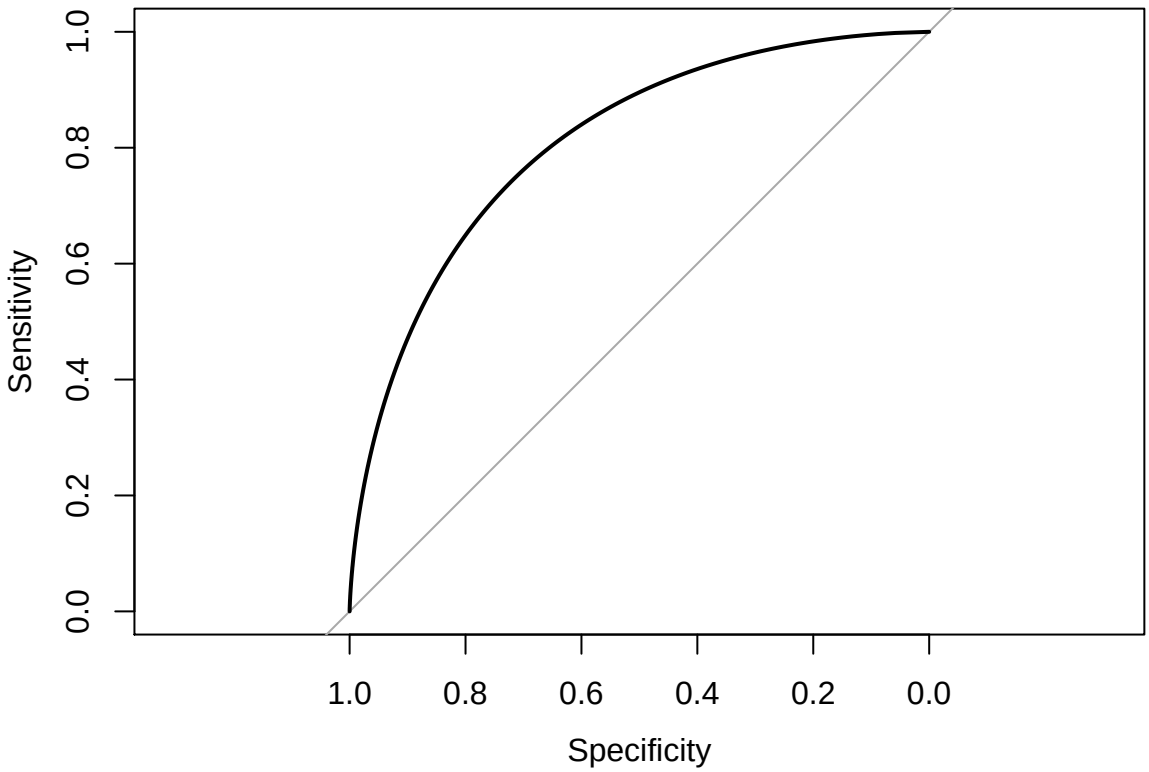
Pickup	Sports_Car	Van	SUV	Professional	Clerical	Doctor	Lawyer
1.818	2.2	1.652	3.081	1.653	1.781	1.493	2.218

Manager	Home_Maker	Student
1.425	1.937	1.753

	Estimate	Std. Error	z value	Pr(>|z|)
KIDSDRIV	0.4208	0.05505	7.645	2.097e-14
INCOME	-3.943e-06	9.833e-07	-4.01	6.065e-05
PARENT1	0.4713	0.09395	5.017	5.253e-07
HOME_VAL	-1.322e-06	3.306e-07	-3.998	6.38e-05
MSTATUS	-0.466	0.07875	-5.918	3.259e-09
TRAVTIME	0.01454	0.001878	7.742	9.749e-15
CAR_USE	0.7758	0.07156	10.84	2.176e-27
BLUEBOOK	-2.303e-05	4.698e-06	-4.902	9.466e-07
TIF	-0.05488	0.007327	-7.491	6.859e-14
OLDCLAIM	-1.388e-05	3.902e-06	-3.556	0.0003771
CLM_FREQ	0.1953	0.02847	6.861	6.834e-12
REVOKED	0.8924	0.0911	9.796	1.17e-22
MVR_PTS	0.1141	0.01356	8.416	3.904e-17
URBANICITY	2.373	0.1124	21.12	5.556e-99
Bachelors	-0.4557	0.07519	-6.061	1.357e-09
Masters	-0.5044	0.09108	-5.537	3.071e-08
PhD	-0.5374	0.1343	-4.002	6.292e-05
Panel_Truck	0.5901	0.1422	4.149	3.345e-05
Pickup	0.5433	0.09803	5.542	2.988e-08
Sports_Car	0.9732	0.1064	9.149	5.757e-20
Van	0.634	0.1193	5.312	1.083e-07
SUV	0.7166	0.08519	8.412	4.016e-17
Manager	-0.7376	0.1065	-6.929	4.236e-12
(Intercept)	-3.389	0.1712	-19.8	2.884e-87

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	9418 on 8160 degrees of freedom
Residual deviance:	7313 on 8137 degrees of freedom




```
Call:
roc.formula(formula = class ~ logit, data = predflag, auc = TRUE,     plot = TRUE, smooth = TRUE)

Data: logit in 6008 controls (class 0) < 2153 cases (class 1).
Smoothing: binormal
Area under the curve: 0.8084
```

4 Model Selection & Prediction

Appendix A: Index-wise Results from Predictive Model

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Index	Predicted Probability	Predicted Classifcation	Predicted Cost
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

Appendix B: R Code