InsNova Auto Insurance Claim Prediction Variable Selection

Lohit Marla

2023-11-20

```
InsNova.data <- read.csv("data/InsNova_data_2023_train.csv")</pre>
train.data <- InsNova.data
InsNova.val_data <- read.csv("data/InsNova_data_2023_vh.csv")</pre>
test.data <- InsNova.val_data</pre>
nrow(train.data)
## [1] 22619
nrow(InsNova.data)
## [1] 22619
nrow(test.data)
## [1] 22620
nrow(InsNova.val_data)
## [1] 22620
column_names <- c(</pre>
     "gender", "agecat", "engine_type",
    "veh_color", "marital_status", "e_bill", "time_of_week_driven", "high_education_ind", "veh_body"
)
# Convert the selected columns to factors in your data frame
train.data[, column_names] <- lapply(train.data[, column_names], as.factor)</pre>
test.data[, column_names] <- lapply(test.data[, column_names], as.factor)
# Check the data frame structure
train.data$clm <- NULL</pre>
train.data$id <- NULL</pre>
test.data$id <- NULL
str(train.data)
## 'data.frame':
                    22619 obs. of 20 variables:
## $ veh_value
                           : num 0.77 4.45 4.9 0.48 0.85 1.37 4.74 0.41 1.41 3.26 ...
## $ exposure
                           : num 0.445 0.562 0.465 0.271 0.142 ...
                           : Factor w/ 13 levels "BUS", "CONVT", ...: 10 11 11 8 10 10 13 10 10 11 ...
## $ veh_body
## $ veh_age
                          : int 4 1 1 4 4 3 1 4 3 2 ...
## $ gender
                           : Factor w/ 2 levels "F", "M": 2 2 1 2 1 2 2 2 1 1 ...
## $ area
                           : chr "D" "A" "A" "A" ...
                          : Factor w/ 6 levels "1", "2", "3", "4", ...: 3 3 3 4 5 4 2 2 4 2 ...
## $ agecat
## $ engine_type
                           : Factor w/ 4 levels "dissel", "electric", ...: 4 4 4 4 4 2 4 4 4 4 ...
```

```
## $ max power
                          : int 147 158 159 80 126 152 232 106 105 100 ...
## $ driving_history_score: num 67 76 58 72 91 59 61 37 41 99 ...
## $ veh color
                         : Factor w/ 9 levels "black", "blue", ...: 1 8 1 8 8 8 4 1 1 8 ...
                          : Factor w/ 2 levels "M", "S": 2 2 1 2 2 2 1 1 2 2 ...
## $ marital_status
## $ e bill
                          : Factor w/ 2 levels "0", "1": 2 2 2 2 1 2 2 1 1 2 ...
## $ time_of_week_driven : Factor w/ 2 levels "weekday", "weekend": 1 1 1 1 1 1 1 2 1 ...
                         : chr "6pm - 12am" "6am - 12pm" "6pm - 12am" "12pm - 6pm" ...
## $ time driven
## $ trm len
                          : int 6 12 6 12 6 6 6 12 12 6 ...
## $ credit score
                          : num 640 684 654 643 647 ...
## $ high_education_ind : Factor w/ 2 levels "0","1": 2 1 2 1 1 1 1 2 2 1 ...
## $ numclaims
                         : int 0000000000...
                          : num 0000000000...
## $ claimcst0
str(test.data)
                   22620 obs. of 18 variables:
## 'data.frame':
                         : num 3.4 2.55 3.04 2.05 1.93 1.36 1.59 0.84 1.59 4.23 ...
## $ veh_value
## $ exposure
                         : num 0.0763 0.0934 0.1578 0.5607 0.2583 ...
                          : Factor w/ 13 levels "BUS", "CONVT", ...: 11 11 11 7 4 13 10 4 10 11 ...
## $ veh_body
## $ veh_age
                         : int 2 2 2 4 2 3 3 4 2 2 ...
## $ gender
                         : Factor w/ 2 levels "F", "M": 2 1 1 2 2 2 1 2 2 1 ...
                         : chr "B" "A" "E" "C" ...
## $ area
## $ agecat
                          : Factor w/ 6 levels "1", "2", "3", "4", ...: 4 3 4 6 4 4 2 2 6 3 ...
## $ engine_type
                         : Factor w/ 4 levels "dissel", "electric", ...: 4 4 4 1 1 4 4 4 3 1 ...
## $ max_power
                         : int 174 181 136 164 89 236 178 97 126 143 ...
## $ driving_history_score: int 83 65 64 82 48 46 59 57 79 56 ...
## $ veh_color
                         : Factor w/ 9 levels "black", "blue", ...: 1 9 8 4 1 1 8 5 8 1 ...
                         : Factor w/ 2 levels "M", "S": 2 1 2 1 2 2 2 1 1 2 ...
## $ marital status
                         : Factor w/ 2 levels "0", "1": 2 1 2 2 1 1 2 2 2 2 ...
## $ e bill
## $ time_of_week_driven : Factor w/ 2 levels "weekday", "weekend": 1 1 1 1 1 1 1 1 1 1 ...
## $ time driven
                          : chr "6pm - 12am" "12am - 6 am" "12pm - 6pm" "6am - 12pm" ...
                          : int 6 12 12 12 12 12 6 6 12 6 ...
## $ trm_len
## $ credit score
                          : num 648 638 661 648 640 ...
## $ high_education_ind : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
#variable selection for the claimcst0
mlr.full.mod <- lm(claimcst0 ~ . , data = train.data)</pre>
step_model <- step(mlr.full.mod)</pre>
## Start: AIC=317588
## claimcst0 ~ veh_value + exposure + veh_body + veh_age + gender +
      area + agecat + engine_type + max_power + driving_history_score +
##
##
      veh_color + marital_status + e_bill + time_of_week_driven +
##
      time_driven + trm_len + credit_score + high_education_ind +
##
      numclaims
##
                                              RSS
##
                          Df Sum of Sq
                                                     AIC
## - veh body
                          12
                               7759243 2.8216e+10 317570
## - veh_color
                          8 11063200 2.8220e+10 317581
## - agecat
                           5
                               6766486 2.8215e+10 317583
                          3 2186674 2.8211e+10 317584
## - time_driven
                         3 3494139 2.8212e+10 317585
## - engine_type
## - e bill
                         1
                                4575 2.8208e+10 317586
                          1
## - credit_score
                                63717 2.8209e+10 317586
```

```
1
1
## - max power
                                445309 2.8209e+10 317586
                                 729470 2.8209e+10 317587
## - veh value
## - high education ind
                                838606 2.8209e+10 317587
                           1 1814313 2.8210e+10 317587
## - veh_age
                            5 11901945 2.8220e+10 317588
## - area
                          1 2272709 2.8211e+10 317588
## - marital status
## - trm len
                          1 2492129 2.8211e+10 317588
                                         2.8208e+10 317588
## <none>
## - driving_history_score 1 2504441 2.8211e+10 317588
## - time_of_week_driven 1
                                4586574 2.8213e+10 317590
## - exposure
                           1
                                 6093199 2.8215e+10 317591
## - gender
                                 7694815 2.8216e+10 317592
                           1
## - numclaims
                            1 8219918892 3.6428e+10 323370
##
## Step: AIC=317570.3
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
##
       agecat + engine_type + max_power + driving_history_score +
##
       veh_color + marital_status + e_bill + time_of_week_driven +
##
       time_driven + trm_len + credit_score + high_education_ind +
##
       numclaims
##
##
                           Df Sum of Sq
                                                RSS
                                                       AIC
                           8 11026151 2.8227e+10 317563
## - veh_color
                               7078223 2.8223e+10 317566
## - agecat
                           5
                          3 2198902 2.8218e+10 317566
## - time driven
                          3 3685015 2.8220e+10 317567
## - engine_type
## - e_bill 1 5310 2.8216e+10 317568

## - credit_score 1 60305 2.8216e+10 317568

## - veh_value 1 225548 2.8216e+10 317568

## - high_education_ind 1 798850 2.8217e+10 317569
                            1 910006 2.8217e+10 317569
## - max_power
                            1 1297291 2.8218e+10 317569
## - veh_age
                            5 11699108 2.8228e+10 317570
## - area
                           1 2188457 2.8218e+10 317570
## - marital_status
## <none>
                                         2.8216e+10 317570
## - trm len
                                 2536930 2.8219e+10 317570
## - driving_history_score 1 2556217 2.8219e+10 317570
## - time_of_week_driven 1 4638312 2.8221e+10 317572
## - exposure
                            1
                                 6325850 2.8223e+10 317573
                                 7998970 2.8224e+10 317575
## - gender
                           1
## - numclaims
                           1 8219568761 3.6436e+10 323351
##
## Step: AIC=317563.1
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
       agecat + engine_type + max_power + driving_history_score +
       marital_status + e_bill + time_of_week_driven + time_driven +
##
##
       trm_len + credit_score + high_education_ind + numclaims
##
                                                RSS
##
                           Df Sum of Sq
                                                       ATC
                                6766130 2.8234e+10 317559
## - agecat
                           5
                           3 2181655 2.8229e+10 317559
## - time_driven
                          3 3798127 2.8231e+10 317560
## - engine type
## - e bill
                           1
                                 5305 2.8227e+10 317561
                          1 52460 2.8227e+10 317561
## - credit score
```

```
## - veh value
                               208796 2.8227e+10 317561
                           1
## - high_education_ind
                               834368 2.8228e+10 317562
                          1
                               948282 2.8228e+10 317562
## - max power
                              1302323 2.8229e+10 317562
## - veh_age
                           1
                           5 11855498 2.8239e+10 317563
## - area
                         1 2297800 2.8230e+10 317563
## - marital status
## - trm len
                          1 2391266 2.8230e+10 317563
                                        2.8227e+10 317563
## <none>
## - driving_history_score 1
                                2631772 2.8230e+10 317563
## - time_of_week_driven 1
                                4597644 2.8232e+10 317565
## - exposure
                           1
                                6120902 2.8233e+10 317566
                                7956621 2.8235e+10 317567
## - gender
                           1
## - numclaims
                           1 8215986368 3.6443e+10 323340
##
## Step: AIC=317558.5
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
      engine_type + max_power + driving_history_score + marital_status +
##
      e_bill + time_of_week_driven + time_driven + trm_len + credit_score +
##
      high_education_ind + numclaims
##
                                              RSS
##
                          Df Sum of Sq
                                                     ATC:
## - time driven
                              2136681 2.8236e+10 317554
## - engine_type
                          3
                                3686177 2.8238e+10 317555
                              2283 2.8234e+10 317557
58054 2.8234e+10 317557
## - e_bill
                           1
                         1
## - credit score
                          1 119497 2.8234e+10 317557
## - veh value
## - high_education_ind
                               861907 2.8235e+10 317557
                          1
                               948114 2.8235e+10 317557
## - max_power
                           1
## - veh_age
                           1 1026112 2.8235e+10 317557
                           1 2364864 2.8236e+10 317558
## - marital_status
                              2446236 2.8236e+10 317558
## - trm_len
                           1
## - area
                           5 12484072 2.8247e+10 317559
## <none>
                                        2.8234e+10 317559
                                2666866 2.8237e+10 317559
## - driving_history_score 1
## - time_of_week_driven
                               4552911 2.8239e+10 317560
## - exposure
                                6327796 2.8240e+10 317562
                           1
## - gender
                                8123549 2.8242e+10 317563
## - numclaims
                          1 8243824124 3.6478e+10 323351
##
## Step: AIC=317554.2
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
##
      engine_type + max_power + driving_history_score + marital_status +
      e_bill + time_of_week_driven + trm_len + credit_score + high_education_ind +
##
##
      numclaims
##
                                                     AIC
##
                          Df Sum of Sq
                                              RSS
## - engine_type
                              3692794 2.8240e+10 317551
## - e_bill
                                1647 2.8236e+10 317552
                          1
                                53886 2.8236e+10 317552
## - credit_score
                           1
                              118831 2.8236e+10 317552
## - veh_value
                           1
                           1 870324 2.8237e+10 317553
## - high_education_ind
                               940493 2.8237e+10 317553
## - max_power
                          1
                           1 1009574 2.8237e+10 317553
## - veh age
                          1 2378558 2.8239e+10 317554
## - marital status
```

```
## - trm len
                                2447064 2.8239e+10 317554
## <none>
                                        2.8236e+10 317554
## - area
                           5 12569003 2.8249e+10 317554
## - driving_history_score 1
                               2654818 2.8239e+10 317554
## - time_of_week_driven
                           1
                               4511972 2.8241e+10 317556
## - exposure
                           1 6307417 2.8242e+10 317557
## - gender
                                8102974 2.8244e+10 317559
## - numclaims
                           1 8242182977 3.6478e+10 323345
##
## Step: AIC=317551.2
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
##
      max_power + driving_history_score + marital_status + e_bill +
      time_of_week_driven + trm_len + credit_score + high_education_ind +
##
##
      numclaims
##
##
                          Df Sum of Sq
                                               RSS
                                                      AIC
## - e_bill
                                   2407 2.8240e+10 317549
                           1
## - veh value
                                  10442 2.8240e+10 317549
## - credit_score
                                57112 2.8240e+10 317549
                          1
                               471007 2.8240e+10 317550
## - max power
                           1
                           1 698097 2.8241e+10 317550
## - veh_age
## - high_education_ind
                          1 878382 2.8241e+10 317550
                           5 12149103 2.8252e+10 317551
## - area
                           1 2420558 2.8242e+10 317551
## - marital status
## - trm_len
                           1 2453082 2.8242e+10 317551
## <none>
                                        2.8240e+10 317551
## - driving_history_score 1
                                2702971 2.8243e+10 317551
## - time_of_week_driven
                           1
                               4496679 2.8244e+10 317553
## - exposure
                           1
                                6310286 2.8246e+10 317554
## - gender
                           1
                                7678879 2.8248e+10 317555
## - numclaims
                           1 8241458432 3.6481e+10 323341
##
## Step: AIC=317549.2
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
##
      max_power + driving_history_score + marital_status + time_of_week_driven +
##
      trm_len + credit_score + high_education_ind + numclaims
##
##
                          Df Sum of Sq
                                               RSS
## - veh value
                           1
                               10027 2.8240e+10 317547
## - credit_score
                                 57579 2.8240e+10 317547
                          1
## - max power
                               470721 2.8240e+10 317548
                          1
                               715725 2.8241e+10 317548
## - veh age
                           1
                                877634 2.8241e+10 317548
## - high_education_ind
                           1
                           5 12149138 2.8252e+10 317549
## - area
## - marital_status
                           1 2419739 2.8242e+10 317549
## - trm_len
                                2456141 2.8242e+10 317549
                           1
## <none>
                                        2.8240e+10 317549
## - driving_history_score 1
                                2704060 2.8243e+10 317549
## - time_of_week_driven
                                4496281 2.8244e+10 317551
                           1
## - exposure
                           1
                                6315140 2.8246e+10 317552
                                7683149 2.8248e+10 317553
## - gender
                           1
                          1 8241471348 3.6481e+10 323339
## - numclaims
##
## Step: AIC=317547.2
```

```
## claimcst0 ~ exposure + veh_age + gender + area + max_power +
##
       driving_history_score + marital_status + time_of_week_driven +
       trm_len + credit_score + high_education_ind + numclaims
##
##
##
                           Df Sum of Sq
                                                RSS
                                 57514 2.8240e+10 317545
## - credit score
                                  530572 2.8240e+10 317546
## - max power
                            1
                                 878541 2.8241e+10 317546
## - high_education_ind
                            1
## - veh_age
                            1
                                 923707 2.8241e+10 317546
                              12337582 2.8252e+10 317547
## - area
                            5
## - marital_status
                            1
                                2419645 2.8242e+10 317547
                                 2456237 2.8242e+10 317547
## - trm len
                            1
                                         2.8240e+10 317547
## <none>
                                 2707967 2.8243e+10 317547
## - driving_history_score 1
## - time_of_week_driven
                                 4497733 2.8244e+10 317549
                            1
## - exposure
                            1
                                 6307936 2.8246e+10 317550
                                 7856454 2.8248e+10 317551
## - gender
                            1
## - numclaims
                            1 8241656002 3.6482e+10 323337
##
## Step: AIC=317545.2
## claimcst0 ~ exposure + veh_age + gender + area + max_power +
       driving_history_score + marital_status + time_of_week_driven +
##
      trm_len + high_education_ind + numclaims
##
##
                                                       ATC
                               Sum of Sq
                                                RSS
## - max_power
                            1
                                529005 2.8240e+10 317544
## - high_education_ind
                                  879914 2.8241e+10 317544
                            1
                                 921656 2.8241e+10 317544
## - veh_age
                            1
## - area
                            5
                              12353298 2.8252e+10 317545
## - marital_status
                                 2425134 2.8242e+10 317545
                            1
## - trm_len
                            1
                                 2437872 2.8242e+10 317545
## <none>
                                         2.8240e+10 317545
                                 2711393 2.8243e+10 317545
## - driving_history_score 1
                                 4497059 2.8244e+10 317547
## - time_of_week_driven
                            1
## - exposure
                            1
                                 6254225 2.8246e+10 317548
## - gender
                                 7859872 2.8248e+10 317550
                            1
## - numclaims
                            1 8241784621 3.6482e+10 323335
##
## Step: AIC=317543.7
## claimcst0 ~ exposure + veh_age + gender + area + driving_history_score +
      marital_status + time_of_week_driven + trm_len + high_education_ind +
##
      numclaims
##
                                                RSS
                           Df Sum of Sq
                                                       ATC:
                                 874749 2.8241e+10 317542
## - high_education_ind
                            1
                                 997303 2.8241e+10 317542
## - veh_age
                            1
## - marital_status
                            1
                                 2430999 2.8243e+10 317544
                                 2448129 2.8243e+10 317544
## - trm len
## <none>
                                         2.8240e+10 317544
## - driving_history_score 1
                                 2720693 2.8243e+10 317544
                            5
                               13027623 2.8253e+10 317544
## - area
                                4464435 2.8245e+10 317545
## - time of week driven
                            1
## - exposure
                            1
                                 6290348 2.8247e+10 317547
                                 8696659 2.8249e+10 317549
## - gender
```

```
## - numclaims
                           1 8242233035 3.6483e+10 323334
##
## Step: AIC=317542.4
## claimcst0 ~ exposure + veh_age + gender + area + driving_history_score +
      marital_status + time_of_week_driven + trm_len + numclaims
##
##
                          Df Sum of Sq
                                              RSS
                               974784 2.8242e+10 317541
## - veh age
                           1
## - marital_status
                                2406750 2.8244e+10 317542
## <none>
                                       2.8241e+10 317542
## - driving_history_score 1 2720908 2.8244e+10 317543
                           5 12963548 2.8254e+10 317543
## - area
## - trm_len
                           1
                               3534453 2.8245e+10 317543
## - time_of_week_driven
                        1 4479147 2.8246e+10 317544
## - gender
                          1 8695142 2.8250e+10 317547
                           1 13675412 2.8255e+10 317551
## - exposure
## - numclaims
                          1 8241393825 3.6483e+10 323332
##
## Step: AIC=317541.2
## claimcst0 ~ exposure + gender + area + driving_history_score +
##
      marital_status + time_of_week_driven + trm_len + numclaims
##
##
                          Df Sum of Sq
                                              RSS
                                                     ATC
                           1 2393387 2.8245e+10 317541
## - marital status
                                       2.8242e+10 317541
## <none>
## - driving_history_score 1 2708939 2.8245e+10 317541
## - area
                           5 13034811 2.8255e+10 317542
## - trm_len
                          1 3538736 2.8246e+10 317542
## - time_of_week_driven
                        1 4513088 2.8247e+10 317543
## - gender
                          1 9020922 2.8251e+10 317546
                          1 13449328 2.8256e+10 317550
## - exposure
## - numclaims
                          1 8241463433 3.6484e+10 323331
##
## Step: AIC=317541.1
## claimcst0 ~ exposure + gender + area + driving_history_score +
      time_of_week_driven + trm_len + numclaims
##
##
                          Df Sum of Sq
                                              RSS
                                                     ATC
## <none>
                                        2.8245e+10 317541
## - driving_history_score 1
                               2792698 2.8248e+10 317541
## - area
                           5 13052626 2.8258e+10 317542
## - trm len
                           1 3537532 2.8248e+10 317542
                         1 4503205 2.8249e+10 317543
## - time_of_week_driven
                        1 8942152 2.8254e+10 317546
## - gender
## - exposure
                          1 13552487 2.8258e+10 317550
## - numclaims
                          1 8243234764 3.6488e+10 323331
formula(step_model)
## claimcst0 ~ exposure + gender + area + driving_history_score +
      time_of_week_driven + trm_len + numclaims
AIC(step_model)
```

[1] 381733

```
BIC(step_model)
## [1] 381837.4
final_model.forward <- step(mlr.full.mod, direction = "forward")</pre>
## Start: AIC=317588
## claimcst0 ~ veh_value + exposure + veh_body + veh_age + gender +
      area + agecat + engine_type + max_power + driving_history_score +
##
      veh_color + marital_status + e_bill + time_of_week_driven +
##
      time_driven + trm_len + credit_score + high_education_ind +
##
      numclaims
formula(final_model.forward)
## claimcst0 ~ veh_value + exposure + veh_body + veh_age + gender +
      area + agecat + engine_type + max_power + driving_history_score +
##
##
      veh_color + marital_status + e_bill + time_of_week_driven +
##
      time_driven + trm_len + credit_score + high_education_ind +
      numclaims
AIC(final_model.forward)
## [1] 381780
BIC(final_model.forward)
## [1] 382189.3
final_model.backward <- step(mlr.full.mod, direction = "backward")</pre>
## Start: AIC=317588
## claimcst0 ~ veh_value + exposure + veh_body + veh_age + gender +
      area + agecat + engine_type + max_power + driving_history_score +
##
      veh_color + marital_status + e_bill + time_of_week_driven +
##
      time_driven + trm_len + credit_score + high_education_ind +
##
      numclaims
##
##
                         Df Sum of Sq
                                              RSS
                                                     ATC
## - veh_body
                         12 7759243 2.8216e+10 317570
                        8 11063200 2.8220e+10 317581
## - veh color
## - agecat
                         5 6766486 2.8215e+10 317583
                        3 2186674 2.8211e+10 317584
## - time_driven
                      3 3494139 2.8212e+10 317585
## - engine_type
## - e bill
                         1 4575 2.8208e+10 317586
## - credit_score 1 63717 2.8209e+10 317586
## - max_power 1 445309 2.8209e+10 317586
                         1 729470 2.8209e+10 317587
## - veh_value
5 11901945 2.8220e+10 317588
## - area
                        1 2272709 2.8211e+10 317588
## - marital_status
## - trm_len
                 1 2492129 2.8211e+10 317588
                                       2.8208e+10 317588
## <none>
## - driving_history_score 1 2504441 2.8211e+10 317588
## - time_of_week_driven 1 4586574 2.8213e+10 317590
## - exposure
                               6093199 2.8215e+10 317591
```

```
## - gender
                                 7694815 2.8216e+10 317592
## - numclaims
                           1 8219918892 3.6428e+10 323370
##
## Step: AIC=317570.3
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
       agecat + engine_type + max_power + driving_history_score +
##
       veh_color + marital_status + e_bill + time_of_week_driven +
##
      time_driven + trm_len + credit_score + high_education_ind +
##
       numclaims
##
##
                           Df Sum of Sq
                                                RSS
                              11026151 2.8227e+10 317563
## - veh_color
                                 7078223 2.8223e+10 317566
## - agecat
                            5
                           3 2198902 2.8218e+10 317566
## - time_driven
                          3 3685015 2.8220e+10 317567
## - engine_type
                          1 5310 2.8216e+10 317568
1 60305 2.8216e+10 317568
1 225548 2.8216e+10 317568
## - e_bill
## - credit_score
## - veh value
                                798850 2.8217e+10 317569
## - high_education_ind
                           1
                                910006 2.8217e+10 317569
## - max power
                            1
## - veh_age
                           1
                               1297291 2.8218e+10 317569
## - area
                           5 11699108 2.8228e+10 317570
                          1 2188457 2.8218e+10 317570
## - marital_status
                                         2.8216e+10 317570
## <none>
                            1 2536930 2.8219e+10 317570
## - trm len
## - driving_history_score 1
                                 2556217 2.8219e+10 317570
## - time_of_week_driven
                                 4638312 2.8221e+10 317572
                            1
                                 6325850 2.8223e+10 317573
## - exposure
                           1
## - gender
                           1
                                 7998970 2.8224e+10 317575
## - numclaims
                           1 8219568761 3.6436e+10 323351
##
## Step: AIC=317563.1
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
       agecat + engine_type + max_power + driving_history_score +
##
##
       marital_status + e_bill + time_of_week_driven + time_driven +
##
      trm_len + credit_score + high_education_ind + numclaims
##
##
                           Df Sum of Sq
                                                RSS
## - agecat
                           5
                               6766130 2.8234e+10 317559
## - time_driven
                           3
                                 2181655 2.8229e+10 317559
                           3 3798127 2.8231e+10 317560
## - engine_type
                                 5305 2.8227e+10 317561
## - e bill
                           1
                                 52460 2.8227e+10 317561
## - credit score
                           1
                           1 52460 2.8227e+10 317561
1 208796 2.8227e+10 317561
## - veh_value
## - high_education_ind
                           1 834368 2.8228e+10 317562
                                948282 2.8228e+10 317562
## - max_power
                            1
## - veh_age
                            1 1302323 2.8229e+10 317562
## - area
                            5 11855498 2.8239e+10 317563
## - marital_status
                           1
                                2297800 2.8230e+10 317563
                                 2391266 2.8230e+10 317563
## - trm_len
                            1
## <none>
                                         2.8227e+10 317563
## - driving history score 1
                                 2631772 2.8230e+10 317563
## - time_of_week_driven
                            1
                                 4597644 2.8232e+10 317565
## - exposure
                                 6120902 2.8233e+10 317566
```

```
## - gender
                                   7956621 2.8235e+10 317567
## - numclaims
                             1 8215986368 3.6443e+10 323340
##
## Step: AIC=317558.5
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
       engine_type + max_power + driving_history_score + marital_status +
       e_bill + time_of_week_driven + time_driven + trm_len + credit_score +
       high_education_ind + numclaims
##
##
##
                                                           AIC
                             Df Sum of Sq
                                                   RSS
## - time_driven
                                   2136681 2.8236e+10 317554
                                   3686177 2.8238e+10 317555
## - engine_type
                             3
## - e_bill 1 2283 2.8234e+10 317557

## - credit_score 1 58054 2.8234e+10 317557

## - veh_value 1 119497 2.8234e+10 317557

## - high_education_ind 1 861907 2.8235e+10 317557

## - max_power 1 948114 2.8235e+10 317557
## - veh age
                            1 1026112 2.8235e+10 317557
                             1 2364864 2.8236e+10 317558
## - marital_status
                                 2446236 2.8236e+10 317558
## - trm len
                             1
                             5 12484072 2.8247e+10 317559
## - area
## <none>
                                           2.8234e+10 317559
## - driving_history_score 1 2666866 2.8237e+10 317559
## - time_of_week_driven 1 4552911 2.8239e+10 317560
                  1 6327796 2.8240e+10 317562
## - exposure
## - gender
                            1
                                   8123549 2.8242e+10 317563
## - numclaims
                             1 8243824124 3.6478e+10 323351
##
## Step: AIC=317554.2
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
##
       engine_type + max_power + driving_history_score + marital_status +
##
       e_bill + time_of_week_driven + trm_len + credit_score + high_education_ind +
##
       numclaims
##
                            Df Sum of Sq
                                                   RSS
                                                          AIC
                            3 3692794 2.8240e+10 317551
## - engine_type
## - e bill
                            1 1647 2.8236e+10 317552
## - credit_score 1 53886 2.8236e+10 317552

## - veh_value 1 118831 2.8236e+10 317552

## - high_education_ind 1 870324 2.8237e+10 317553
## - max_power
                            1 940493 2.8237e+10 317553
                             1 1009574 2.8237e+10 317553
## - veh age
## - marital_status
                             1 2378558 2.8239e+10 317554
## - trm_len
                            1 2447064 2.8239e+10 317554
## <none>
                                           2.8236e+10 317554
                              5 12569003 2.8249e+10 317554
## - area
## - driving_history_score 1 2654818 2.8239e+10 317554
                             1 4511972 2.8241e+10 317556
## - time_of_week_driven
## - exposure
                             1
                                   6307417 2.8242e+10 317557
## - gender
                             1
                                   8102974 2.8244e+10 317559
## - numclaims
                            1 8242182977 3.6478e+10 323345
## Step: AIC=317551.2
## claimcst0 ~ veh value + exposure + veh age + gender + area +
```

```
##
       max_power + driving_history_score + marital_status + e_bill +
##
       time_of_week_driven + trm_len + credit_score + high_education_ind +
##
       numclaims
##
##
                          Df Sum of Sq
                                                RSS
                                                       AIC
                                   2407 2.8240e+10 317549
## - e bill
                           1
## - veh value
                                  10442 2.8240e+10 317549
                           1
                                  57112 2.8240e+10 317549
## - credit_score
                           1
## - max_power
                           1
                                 471007 2.8240e+10 317550
## - veh_age
                           1
                                698097 2.8241e+10 317550
## - high_education_ind
                           1
                                878382 2.8241e+10 317550
                            5 12149103 2.8252e+10 317551
## - area
                              2420558 2.8242e+10 317551
## - marital_status
                           1
## - trm_len
                                2453082 2.8242e+10 317551
## <none>
                                         2.8240e+10 317551
## - driving_history_score 1
                                2702971 2.8243e+10 317551
## - time_of_week_driven
                                4496679 2.8244e+10 317553
                            1
## - exposure
                                 6310286 2.8246e+10 317554
                                7678879 2.8248e+10 317555
## - gender
                            1
## - numclaims
                            1 8241458432 3.6481e+10 323341
##
## Step: AIC=317549.2
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
       max_power + driving_history_score + marital_status + time_of_week_driven +
       trm_len + credit_score + high_education_ind + numclaims
##
##
##
                           Df
                              Sum of Sq
                                                RSS
                                 10027 2.8240e+10 317547
## - veh_value
                           1
## - credit_score
                                  57579 2.8240e+10 317547
                           1
## - max_power
                           1
                                470721 2.8240e+10 317548
                                715725 2.8241e+10 317548
## - veh_age
                            1
## - high_education_ind
                            1
                                877634 2.8241e+10 317548
## - area
                            5 12149138 2.8252e+10 317549
                                2419739 2.8242e+10 317549
## - marital_status
                            1
## - trm len
                                2456141 2.8242e+10 317549
## <none>
                                        2.8240e+10 317549
## - driving_history_score 1
                                2704060 2.8243e+10 317549
## - time_of_week_driven
                                4496281 2.8244e+10 317551
                            1
## - exposure
                            1
                                 6315140 2.8246e+10 317552
## - gender
                                7683149 2.8248e+10 317553
                           1
## - numclaims
                           1 8241471348 3.6481e+10 323339
##
## Step: AIC=317547.2
## claimcst0 ~ exposure + veh_age + gender + area + max_power +
       driving_history_score + marital_status + time_of_week_driven +
##
       trm_len + credit_score + high_education_ind + numclaims
##
##
                             Sum of Sq
                                                RSS
                                                       AIC
## - credit_score
                            1
                                 57514 2.8240e+10 317545
                                 530572 2.8240e+10 317546
## - max_power
                            1
                                878541 2.8241e+10 317546
## - high_education_ind
                            1
                                923707 2.8241e+10 317546
## - veh_age
                            1
## - area
                            5 12337582 2.8252e+10 317547
## - marital status
                           1 2419645 2.8242e+10 317547
```

```
## - trm len
                                2456237 2.8242e+10 317547
                                         2.8240e+10 317547
## <none>
                                2707967 2.8243e+10 317547
## - driving_history_score 1
                                4497733 2.8244e+10 317549
## - time_of_week_driven
                           1
## - exposure
                           1
                                6307936 2.8246e+10 317550
                                7856454 2.8248e+10 317551
## - gender
                           1
## - numclaims
                           1 8241656002 3.6482e+10 323337
##
## Step: AIC=317545.2
## claimcst0 ~ exposure + veh_age + gender + area + max_power +
       driving_history_score + marital_status + time_of_week_driven +
##
       trm_len + high_education_ind + numclaims
##
                          Df Sum of Sq
##
                                                RSS
                                                       AIC
                                529005 2.8240e+10 317544
## - max_power
                           1
## - high_education_ind
                           1
                                879914 2.8241e+10 317544
                               921656 2.8241e+10 317544
## - veh_age
                           1
## - area
                           5 12353298 2.8252e+10 317545
## - marital_status
                           1 2425134 2.8242e+10 317545
## - trm len
                           1
                              2437872 2.8242e+10 317545
## <none>
                                        2.8240e+10 317545
## - driving_history_score 1
                                2711393 2.8243e+10 317545
                                4497059 2.8244e+10 317547
## - time_of_week_driven
                           1
## - exposure
                                6254225 2.8246e+10 317548
                           1
                                7859872 2.8248e+10 317550
## - gender
                           1
## - numclaims
                           1 8241784621 3.6482e+10 323335
##
## Step: AIC=317543.7
## claimcst0 ~ exposure + veh_age + gender + area + driving_history_score +
##
       marital_status + time_of_week_driven + trm_len + high_education_ind +
##
       numclaims
##
                          Df Sum of Sq
##
                                               RSS
                                                       AIC
                               874749 2.8241e+10 317542
## - high_education_ind
                           1
## - veh_age
                           1
                                 997303 2.8241e+10 317542
## - marital_status
                                2430999 2.8243e+10 317544
                           1
## - trm len
                           1 2448129 2.8243e+10 317544
## <none>
                                         2.8240e+10 317544
## - driving_history_score 1
                               2720693 2.8243e+10 317544
## - area
                           5 13027623 2.8253e+10 317544
## - time of week driven
                              4464435 2.8245e+10 317545
                           1
## - exposure
                                6290348 2.8247e+10 317547
                           1
                                8696659 2.8249e+10 317549
## - gender
                           1
## - numclaims
                           1 8242233035 3.6483e+10 323334
## Step: AIC=317542.4
## claimcst0 ~ exposure + veh_age + gender + area + driving_history_score +
       marital_status + time_of_week_driven + trm_len + numclaims
##
##
##
                          Df Sum of Sq
                                               RSS
                           1
                                 974784 2.8242e+10 317541
## - veh_age
                                2406750 2.8244e+10 317542
## - marital status
                           1
## <none>
                                        2.8241e+10 317542
## - driving history score 1
                                2720908 2.8244e+10 317543
```

```
## - area
                           5 12963548 2.8254e+10 317543
                           1 3534453 2.8245e+10 317543
## - trm len
## - time_of_week_driven
                         1 4479147 2.8246e+10 317544
                               8695142 2.8250e+10 317547
## - gender
                           1
## - exposure
                           1 13675412 2.8255e+10 317551
## - numclaims
                           1 8241393825 3.6483e+10 323332
## Step: AIC=317541.2
## claimcst0 ~ exposure + gender + area + driving_history_score +
       marital_status + time_of_week_driven + trm_len + numclaims
##
##
##
                          Df Sum of Sq
                                               RSS
## - marital_status
                                2393387 2.8245e+10 317541
## <none>
                                        2.8242e+10 317541
## - driving_history_score 1
                                2708939 2.8245e+10 317541
## - area
                           5 13034811 2.8255e+10 317542
## - trm_len
                           1 3538736 2.8246e+10 317542
## - time_of_week_driven
                           1 4513088 2.8247e+10 317543
                               9020922 2.8251e+10 317546
## - gender
                           1
## - exposure
                           1 13449328 2.8256e+10 317550
## - numclaims
                           1 8241463433 3.6484e+10 323331
## Step: AIC=317541.1
## claimcst0 ~ exposure + gender + area + driving_history_score +
       time_of_week_driven + trm_len + numclaims
##
##
##
                          Df Sum of Sq
                                               RSS
                                                       AIC
                                         2.8245e+10 317541
## <none>
## - driving_history_score 1
                                2792698 2.8248e+10 317541
## - area
                           5 13052626 2.8258e+10 317542
## - trm_len
                           1
                               3537532 2.8248e+10 317542
## - time_of_week_driven
                           1
                                4503205 2.8249e+10 317543
## - gender
                                8942152 2.8254e+10 317546
## - exposure
                           1 13552487 2.8258e+10 317550
                            1 8243234764 3.6488e+10 323331
## - numclaims
formula(final_model.backward)
## claimcst0 ~ exposure + gender + area + driving_history_score +
       time_of_week_driven + trm_len + numclaims
AIC(final_model.backward)
## [1] 381733
BIC(final_model.backward)
## [1] 381837.4
final_model.both <- step(mlr.full.mod, direction = "both")</pre>
## Start: AIC=317588
## claimcst0 ~ veh_value + exposure + veh_body + veh_age + gender +
       area + agecat + engine_type + max_power + driving_history_score +
##
       veh_color + marital_status + e_bill + time_of_week_driven +
##
      time driven + trm len + credit score + high education ind +
##
      numclaims
```

```
##
##
                                               RSS
                                                      ATC
                          Df Sum of Sq
## - veh body
                         12 7759243 2.8216e+10 317570
                         8 11063200 2.8220e+10 317581
## - veh_color
## - agecat
                          5 6766486 2.8215e+10 317583
## - time driven
                         3 2186674 2.8211e+10 317584
                         3 3494139 2.8212e+10 317585
## - engine_type
                          1 4575 2.8208e+10 317586
1 63717 2.8209e+10 317586
1 445309 2.8209e+10 317586
## - e bill
## - credit_score
## - max_power
## - veh_value
                          1 729470 2.8209e+10 317587
                               838606 2.8209e+10 317587
## - high_education_ind
                           1
## - veh_age
                           1 1814313 2.8210e+10 317587
                           5 11901945 2.8220e+10 317588
## - area
## - marital_status
                          1 2272709 2.8211e+10 317588
## - trm_len
                                2492129 2.8211e+10 317588
                                        2.8208e+10 317588
## <none>
## - driving_history_score 1
                                2504441 2.8211e+10 317588
                                4586574 2.8213e+10 317590
## - time_of_week_driven
                           1
## - exposure
                           1
                                6093199 2.8215e+10 317591
## - gender
                           1
                                7694815 2.8216e+10 317592
## - numclaims
                           1 8219918892 3.6428e+10 323370
##
## Step: AIC=317570.3
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
      agecat + engine_type + max_power + driving_history_score +
##
      veh_color + marital_status + e_bill + time_of_week_driven +
      time_driven + trm_len + credit_score + high_education_ind +
##
##
      numclaims
##
##
                          Df Sum of Sq
                                               RSS
## - veh_color
                           8
                              11026151 2.8227e+10 317563
## - agecat
                           5 7078223 2.8223e+10 317566
                         3 2198902 2.8218e+10 317566
## - time_driven
## - engine_type
                           3 3685015 2.8220e+10 317567
                          1 5310 2.8216e+10 317568
## - e bill
## - credit score
                          1
                                60305 2.8216e+10 317568
## - veh_value
                               225548 2.8216e+10 317568
                           1
## - high_education_ind
                                798850 2.8217e+10 317569
                           1
## - max_power
                           1
                                910006 2.8217e+10 317569
## - veh age
                           1 1297291 2.8218e+10 317569
## - area
                           5 11699108 2.8228e+10 317570
## - marital status
                           1
                                2188457 2.8218e+10 317570
## <none>
                                        2.8216e+10 317570
                                2536930 2.8219e+10 317570
## - trm_len
                           1
                                2556217 2.8219e+10 317570
## - driving_history_score 1
## - time_of_week_driven
                           1
                                4638312 2.8221e+10 317572
## - exposure
                           1
                                6325850 2.8223e+10 317573
## - gender
                           1
                                7998970 2.8224e+10 317575
## + veh_body
                          12
                                7759243 2.8208e+10 317588
## - numclaims
                           1 8219568761 3.6436e+10 323351
## Step: AIC=317563.1
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
```

```
##
       agecat + engine_type + max_power + driving_history_score +
##
      marital_status + e_bill + time_of_week_driven + time_driven +
##
       trm_len + credit_score + high_education_ind + numclaims
##
##
                          Df Sum of Sq
                                               RSS
                                6766130 2.8234e+10 317559
## - agecat
                                2181655 2.8229e+10 317559
## - time driven
                            3
                                3798127 2.8231e+10 317560
## - engine_type
                           3
                           1 5305 2.8227e+10 317561
1 52460 2.8227e+10 317561
## - e_bill
## - credit_score
                           1 208796 2.8227e+10 317561
## - veh_value
                               834368 2.8228e+10 317562
## - high_education_ind
                            1
## - max_power
                            1
                                948282 2.8228e+10 317562
                            1 1302323 2.8229e+10 317562
## - veh_age
## - area
                            5 11855498 2.8239e+10 317563
## - marital_status
                           1
                              2297800 2.8230e+10 317563
                                2391266 2.8230e+10 317563
## - trm_len
                            1
## <none>
                                         2.8227e+10 317563
                                2631772 2.8230e+10 317563
## - driving_history_score 1
## - time_of_week_driven
                            1
                                4597644 2.8232e+10 317565
## - exposure
                           1
                                6120902 2.8233e+10 317566
## - gender
                                7956621 2.8235e+10 317567
                           8 11026151 2.8216e+10 317570
## + veh_color
                          12
                                7722194 2.8220e+10 317581
## + veh body
## - numclaims
                           1 8215986368 3.6443e+10 323340
## Step: AIC=317558.5
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
       engine_type + max_power + driving_history_score + marital_status +
       e_bill + time_of_week_driven + time_driven + trm_len + credit_score +
##
##
      high_education_ind + numclaims
##
##
                              Sum of Sq
                                                RSS
                                                       AIC
                                2136681 2.8236e+10 317554
## - time_driven
                            3
## - engine type
                            3
                                3686177 2.8238e+10 317555
## - e bill
                                   2283 2.8234e+10 317557
                           1
## - credit score
                                 58054 2.8234e+10 317557
## - veh_value
                                119497 2.8234e+10 317557
                            1
## - high_education_ind
                                861907 2.8235e+10 317557
                           1
## - max_power
                                948114 2.8235e+10 317557
                            1
## - veh age
                           1 1026112 2.8235e+10 317557
## - marital status
                                2364864 2.8236e+10 317558
                            1
                                2446236 2.8236e+10 317558
## - trm_len
                            1
## - area
                               12484072 2.8247e+10 317559
                                         2.8234e+10 317559
## <none>
                                2666866 2.8237e+10 317559
## - driving_history_score 1
## - time_of_week_driven
                            1
                                4552911 2.8239e+10 317560
## - exposure
                                6327796 2.8240e+10 317562
## - gender
                           1
                                8123549 2.8242e+10 317563
                                6766130 2.8227e+10 317563
## + agecat
                           5
                           8 10714058 2.8223e+10 317566
## + veh_color
                                8030741 2.8226e+10 317576
## + veh_body
                          12
## - numclaims
                          1 8243824124 3.6478e+10 323351
##
```

```
## Step: AIC=317554.2
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
      engine_type + max_power + driving_history_score + marital_status +
      e_bill + time_of_week_driven + trm_len + credit_score + high_education_ind +
##
##
      numclaims
##
                          Df Sum of Sq
                              3692794 2.8240e+10 317551
## - engine_type
                          1 1647 2.8236e+10 317552
1 53886 2.8236e+10 317552
## - e_bill
                       1 53886 2.8236e+10 317552
1 118831 2.8236e+10 317552
1 870324 2.8237e+10 317553
1 940493 2.8237e+10 317553
## - credit_score
## - veh_value
## - high_education_ind
## - max_power
                          1 1009574 2.8237e+10 317553
## - veh_age
## - marital_status
                          1 2378558 2.8239e+10 317554
                           1 2447064 2.8239e+10 317554
## - trm_len
## <none>
                                        2.8236e+10 317554
## - area
                           5 12569003 2.8249e+10 317554
## - driving_history_score 1 2654818 2.8239e+10 317554
## - time_of_week_driven 1 4511972 2.8241e+10 317556
## - exposure
                         1 6307417 2.8242e+10 317557
## + time driven
                         3 2136681 2.8234e+10 317559
## - gender
                          1 8102974 2.8244e+10 317559
                          5 6721156 2.8229e+10 317559
## + agecat
                         8 10698660 2.8225e+10 317562
## + veh color
## + veh body
                        12 8039348 2.8228e+10 317572
## - numclaims
                          1 8242182977 3.6478e+10 323345
## Step: AIC=317551.2
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
##
      max_power + driving_history_score + marital_status + e_bill +
##
      time_of_week_driven + trm_len + credit_score + high_education_ind +
##
      numclaims
##
##
                          Df Sum of Sq
                                              RSS
                                                     AIC
                              2407 2.8240e+10 317549
## - e bill
                          1
## - veh value
                          1
                                 10442 2.8240e+10 317549
## - credit_score
                          1
                                57112 2.8240e+10 317549
                               471007 2.8240e+10 317550
## - max power
                          1
                          1 698097 2.8241e+10 317550
## - veh_age
## - high_education_ind 1 878382 2.8241e+10 317550
                          5 12149103 2.8252e+10 317551
## - area
## - marital_status
                          1 2420558 2.8242e+10 317551
## - trm_len
                                2453082 2.8242e+10 317551
                           1
## <none>
                                        2.8240e+10 317551
                                2702971 2.8243e+10 317551
## - driving_history_score 1
## - time_of_week_driven 1
                               4496679 2.8244e+10 317553
                           3
## + engine_type
                                3692794 2.8236e+10 317554
## - exposure
                         1
                                6310286 2.8246e+10 317554
## - gender
                          1
                                7678879 2.8248e+10 317555
                         3 2143297 2.8238e+10 317555
## + time_driven
                         5 6612904 2.8233e+10 317556
## + agecat
## + veh_color
                         8 10805453 2.8229e+10 317559
## + veh body
                          12 8205296 2.8232e+10 317569
```

```
## - numclaims
                          1 8241458432 3.6481e+10 323341
##
## Step: AIC=317549.2
## claimcst0 ~ veh_value + exposure + veh_age + gender + area +
      max_power + driving_history_score + marital_status + time_of_week_driven +
##
      trm_len + credit_score + high_education_ind + numclaims
##
##
                         Df Sum of Sq
## - veh value
                          1
                              10027 2.8240e+10 317547
## - credit_score
                         1
                                57579 2.8240e+10 317547
## - max_power
                         1
                               470721 2.8240e+10 317548
                              715725 2.8241e+10 317548
## - veh_age
                          1
## - high_education_ind
                       1
                              877634 2.8241e+10 317548
## - area
                          5 12149138 2.8252e+10 317549
## - marital_status
                         1 2419739 2.8242e+10 317549
## - trm_len
                          1
                               2456141 2.8242e+10 317549
## <none>
                                       2.8240e+10 317549
## - driving_history_score 1 2704060 2.8243e+10 317549
## - time_of_week_driven
                          1 4496281 2.8244e+10 317551
## + e bill
                          1
                                2407 2.8240e+10 317551
## + engine_type
                          3 3693554 2.8236e+10 317552
## - exposure
                         1 6315140 2.8246e+10 317552
                         1 7683149 2.8248e+10 317553
## - gender
                         3 2142533 2.8238e+10 317553
## + time driven
## + agecat
                         5 6609912 2.8233e+10 317554
## + veh color
                         8 10805531 2.8229e+10 317557
## + veh_body
                        12 8205164 2.8232e+10 317567
                          1 8241471348 3.6481e+10 323339
## - numclaims
##
## Step: AIC=317547.2
## claimcst0 ~ exposure + veh_age + gender + area + max_power +
##
      driving_history_score + marital_status + time_of_week_driven +
##
      trm_len + credit_score + high_education_ind + numclaims
##
##
                         Df Sum of Sq
                                             RSS
                                                    AIC
## - credit_score
                          1
                             57514 2.8240e+10 317545
## - max power
                              530572 2.8240e+10 317546
## - high_education_ind
                          1
                              878541 2.8241e+10 317546
                              923707 2.8241e+10 317546
## - veh_age
                          1
## - area
                          5 12337582 2.8252e+10 317547
## - marital_status
                          1 2419645 2.8242e+10 317547
## - trm len
                               2456237 2.8242e+10 317547
                          1
                                       2.8240e+10 317547
## <none>
## - driving_history_score 1
                             2707967 2.8243e+10 317547
## - time_of_week_driven 1
                               4497733 2.8244e+10 317549
                               10027 2.8240e+10 317549
## + veh_value
                          1
## + e_bill
                          1
                                 1993 2.8240e+10 317549
## - exposure
                               6307936 2.8246e+10 317550
                         1
## + engine_type
                          3
                               3585730 2.8236e+10 317550
## + time_driven
                          3
                               2142371 2.8238e+10 317551
## - gender
                               7856454 2.8248e+10 317551
                          1
                         5 6574176 2.8233e+10 317552
## + agecat
## + veh color
                         8 10802060 2.8229e+10 317555
                         12 7605434 2.8232e+10 317565
## + veh body
```

```
## - numclaims
                           1 8241656002 3.6482e+10 323337
##
## Step: AIC=317545.2
## claimcst0 ~ exposure + veh_age + gender + area + max_power +
       driving_history_score + marital_status + time_of_week_driven +
##
       trm len + high education ind + numclaims
##
##
                           Df Sum of Sq
                                               RSS
                                                       ATC
## - max power
                           1
                               529005 2.8240e+10 317544
                           1 879914 2.8241e+10 317544
## - high_education_ind
## - veh_age
                          1 921656 2.8241e+10 317544
                          5 12353298 2.8252e+10 317545
## - area
## - marital_status
                          1 2425134 2.8242e+10 317545
                          1 2437872 2.8242e+10 317545
## - trm_len
## <none>
                                         2.8240e+10 317545
## - driving_history_score 1 2711393 2.8243e+10 317545
## - time_of_week_driven 1 4497059 2.8244e+10 317547
## + credit_score 1 57514 2.8240e+10 317547
## + veh_value 1 9962 2.8240e+10 317547
## + e_bill 1 2420 2.8240e+10 317547
## - exposure
                          1 6254225 2.8246e+10 317548
## + engine_type
                         3 3589134 2.8236e+10 317548
                       3 2138034 2.8238e+10 317550
## + time_driven
                          1 7859872 2.8248e+10 317550
## - gender
                          5 6579343 2.8233e+10 317550
## + agecat
## + veh color
                          8 10793775 2.8229e+10 317553
## + veh_body
                         12 7602892 2.8232e+10 317563
## - numclaims
                           1 8241784621 3.6482e+10 323335
##
## Step: AIC=317543.7
## claimcst0 ~ exposure + veh_age + gender + area + driving_history_score +
##
       marital_status + time_of_week_driven + trm_len + high_education_ind +
##
       numclaims
##
##
                           Df Sum of Sq
                                               RSS
                                                       AIC
## - high_education_ind
                          1 874749 2.8241e+10 317542
## - veh age
                          1
                                997303 2.8241e+10 317542
## - marital_status
                           1 2430999 2.8243e+10 317544
## - trm len
                            1 2448129 2.8243e+10 317544
## <none>
                                         2.8240e+10 317544
## - driving_history_score 1 2720693 2.8243e+10 317544
                            5 13027623 2.8253e+10 317544
## - area
## - time_of_week_driven 1 4464435 2.8245e+10 317545
                  1 529005 2.8240e+10 317545
## + max_power
                          1 69583 2.8240e+10 317546
1 55947 2.8240e+10 317546
1 1474 2.8240e+10 317546
## + veh_value
## + credit_score
## + e_bill
## - exposure
                          1 6290348 2.8247e+10 317547
## + engine_type
                          3 2990399 2.8237e+10 317547
                       3 2131382 2.8238e+10 317548
5 6543859 2.8234e+10 317548
## + time_driven
## + agecat
## - gender
                          1 8696659 2.8249e+10 317549
## + veh color
                          8 10811398 2.8230e+10 317551
                           12 7707366 2.8233e+10 317561
## + veh body
```

```
## - numclaims
              1 8242233035 3.6483e+10 323334
##
## Step: AIC=317542.4
## claimcst0 ~ exposure + veh_age + gender + area + driving_history_score +
      marital_status + time_of_week_driven + trm_len + numclaims
##
##
                         Df Sum of Sq
                               974784 2.8242e+10 317541
## - veh age
                          1
## - marital status
                               2406750 2.8244e+10 317542
                                       2.8241e+10 317542
## <none>
## - driving_history_score 1 2720908 2.8244e+10 317543
                          5 12963548 2.8254e+10 317543
## - area
                             3534453 2.8245e+10 317543
## - trm len
                          1
                          1 874749 2.8240e+10 317544
## + high_education_ind
## + max_power
                          1 523841 2.8241e+10 317544
                          1 4479147 2.8246e+10 317544
## - time_of_week_driven
                          1 71434 2.8241e+10 317544
## + veh_value
                               57305 2.8241e+10 317544
900 2.8241e+10 317544
## + credit score
                         1
## + e_bill
                         1
                          3 2997354 2.8238e+10 317546
## + engine type
## + time_driven
                         3 2139918 2.8239e+10 317547
## + agecat
                         5 6570221 2.8235e+10 317547
                         1 8695142 2.8250e+10 317547
## - gender
                          8 10848091 2.8230e+10 317550
## + veh color
                         1 13675412 2.8255e+10 317551
## - exposure
## + veh body
                        12 7657139 2.8234e+10 317560
## - numclaims
                         1 8241393825 3.6483e+10 323332
## Step: AIC=317541.2
## claimcst0 ~ exposure + gender + area + driving_history_score +
##
      marital_status + time_of_week_driven + trm_len + numclaims
##
##
                         Df Sum of Sq
                                             RSS
                                                    AIC
## - marital_status
                               2393387 2.8245e+10 317541
## <none>
                                       2.8242e+10 317541
## - driving_history_score 1
                             2708939 2.8245e+10 317541
## - area
                          5 13034811 2.8255e+10 317542
## - trm len
                          1
                             3538736 2.8246e+10 317542
                             974784 2.8241e+10 317542
## + veh age
                          1
                          1 852230 2.8241e+10 317542
## + high_education_ind
## + max power
                          1 598309 2.8242e+10 317543
## - time_of_week_driven
                          1 4513088 2.8247e+10 317543
                             105876 2.8242e+10 317543
## + veh value
                          1
## + credit_score
                               55072 2.8242e+10 317543
                          1
## + e_bill
                               35921 2.8242e+10 317543
                          1
                          3 2893481 2.8239e+10 317545
## + engine_type
                          3 2118532 2.8240e+10 317545
## + time_driven
## + agecat
                         5 6398466 2.8236e+10 317546
## - gender
                         1 9020922 2.8251e+10 317546
                          8 10881276 2.8231e+10 317548
## + veh_color
## - exposure
                         1 13449328 2.8256e+10 317550
                        12 7762038 2.8235e+10 317559
## + veh_body
## - numclaims
                         1 8241463433 3.6484e+10 323331
##
```

```
## Step: AIC=317541.1
## claimcst0 ~ exposure + gender + area + driving_history_score +
       time_of_week_driven + trm_len + numclaims
##
##
                           Df Sum of Sq
                                                RSS
## <none>
                                          2.8245e+10 317541
                               2393387 2.8242e+10 317541
## + marital_status 1
## - driving_history_score 1 2792698 2.8248e+10 317541
## - area
                            5 13052626 2.8258e+10 317542
## - trm_len
                           1 3537532 2.8248e+10 317542
## + veh_age
                           1 961421 2.8244e+10 317542
## + high_education_ind 1 828517 2.8244e+10 317542 
## + max_power 1 604025 2.8244e+10 317543
## - time_of_week_driven 1 4503205 2.8249e+10 317543
                      1 103231 2.8245e+10 317543
1 60398 2.8245e+10 317543
1 32286 2.8245e+10 317543
## + veh_value
## + credit_score
## + e_bill
## + engine_type
                          3 2927479 2.8242e+10 317545
## + time_driven
                          3 2131661 2.8243e+10 317545
                        5 6464700 2.8238e+10 317546
## + agecat
                           1 8942152 2.8254e+10 317546
## - gender
## + veh_color
                          8 10989222 2.8234e+10 317548
                           1 13552487 2.8258e+10 317550
## - exposure
                         12 7655956 2.8237e+10 317559
## + veh body
                    1 8243234764 3.6488e+10 323331
## - numclaims
formula(final_model.both)
## claimcst0 ~ exposure + gender + area + driving_history_score +
       time_of_week_driven + trm_len + numclaims
AIC(final_model.both)
## [1] 381733
BIC(final_model.both)
## [1] 381837.4
#We have choosen the model with the less BIC and AIC values and the variables selected are as follows
#exposure + gender + area + driving_history_score +
# time_of_week_driven + trm_len + numclaims
#variable selection for the numclaims
InsNova.data <- read.csv("data/InsNova_data_2023_train.csv")</pre>
train.data <- InsNova.data
InsNova.val_data <- read.csv("data/InsNova_data_2023_vh.csv")</pre>
test.data <- InsNova.val_data</pre>
nrow(train.data)
## [1] 22619
nrow(InsNova.data)
## [1] 22619
```

```
nrow(test.data)
## [1] 22620
nrow(InsNova.val_data)
## [1] 22620
column names <- c(</pre>
     "gender", "agecat", "engine_type",
    "veh_color", "marital_status", "e_bill", "time_of_week_driven", "high_education_ind", "veh_body"
)
# Convert the selected columns to factors in your data frame
train.data[, column_names] <- lapply(train.data[, column_names], as.factor)</pre>
test.data[, column_names] <- lapply(test.data[, column_names], as.factor)</pre>
# Check the data frame structure
train.data$clm <- NULL
train.data$claimcst0 <- NULL
train.data$id <- NULL</pre>
test.data$id <- NULL</pre>
str(train.data)
## 'data.frame':
                   22619 obs. of 19 variables:
## $ veh value
                         : num 0.77 4.45 4.9 0.48 0.85 1.37 4.74 0.41 1.41 3.26 ...
## $ exposure
                          : num 0.445 0.562 0.465 0.271 0.142 ...
## $ veh_body
                          : Factor w/ 13 levels "BUS", "CONVT", ...: 10 11 11 8 10 10 13 10 10 11 ...
## $ veh_age
                          : int 4 1 1 4 4 3 1 4 3 2 ...
## $ gender
                         : Factor w/ 2 levels "F", "M": 2 2 1 2 1 2 2 2 1 1 ...
                          : chr "D" "A" "A" "A" ...
## $ area
## $ agecat
                          : Factor w/ 6 levels "1", "2", "3", "4", ...: 3 3 3 4 5 4 2 2 4 2 ...
                          : Factor w/ 4 levels "dissel", "electric", ...: 4 4 4 4 4 2 4 4 4 4 ...
## $ engine_type
                          : int 147 158 159 80 126 152 232 106 105 100 ...
## $ max_power
## $ driving_history_score: num 67 76 58 72 91 59 61 37 41 99 ...
## $ veh_color
                          : Factor w/ 9 levels "black", "blue", ...: 1 8 1 8 8 8 4 1 1 8 ....
## $ marital_status
                          : Factor w/ 2 levels "M", "S": 2 2 1 2 2 2 1 1 2 2 ...
## $ e_bill
                          : Factor w/ 2 levels "0","1": 2 2 2 2 1 2 2 1 1 2 ...
## $ time_of_week_driven : Factor w/ 2 levels "weekday", "weekend": 1 1 1 1 1 1 1 2 1 ...
## $ time_driven
                          : chr "6pm - 12am" "6am - 12pm" "6pm - 12am" "12pm - 6pm" ...
## $ trm len
                          : int 6 12 6 12 6 6 6 12 12 6 ...
## $ credit_score
                          : num 640 684 654 643 647 ...
## $ high education ind : Factor w/ 2 levels "0","1": 2 1 2 1 1 1 1 2 2 1 ...
## $ numclaims
                          : int 0000000000...
str(test.data)
                   22620 obs. of 18 variables:
## 'data.frame':
## $ veh_value
                          : num 3.4 2.55 3.04 2.05 1.93 1.36 1.59 0.84 1.59 4.23 ...
## $ exposure
                          : num   0.0763   0.0934   0.1578   0.5607   0.2583   ...
## $ veh_body
                          : Factor w/ 13 levels "BUS", "CONVT", ...: 11 11 11 7 4 13 10 4 10 11 ...
## $ veh age
                         : int 2 2 2 4 2 3 3 4 2 2 ...
## $ gender
                         : Factor w/ 2 levels "F", "M": 2 1 1 2 2 2 1 2 2 1 ...
## $ area
                          : chr "B" "A" "E" "C" ...
## $ agecat
                          : Factor w/ 6 levels "1","2","3","4",..: 4 3 4 6 4 4 2 2 6 3 ...
```

```
## $ engine_type
                          : Factor w/ 4 levels "dissel", "electric", ...: 4 4 4 1 1 4 4 4 3 1 ...
                          : int 174 181 136 164 89 236 178 97 126 143 ...
## $ max_power
## $ driving_history_score: int 83 65 64 82 48 46 59 57 79 56 ...
                          : Factor w/ 9 levels "black", "blue", ...: 1 9 8 4 1 1 8 5 8 1 ...
## $ veh_color
## $ marital status
                          : Factor w/ 2 levels "M", "S": 2 1 2 1 2 2 2 1 1 2 ...
## $ e_bill
                          : Factor w/ 2 levels "0", "1": 2 1 2 2 1 1 2 2 2 2 ...
## $ time of week driven : Factor w/ 2 levels "weekday", "weekend": 1 1 1 1 1 1 1 1 1 1 ...
                          : chr "6pm - 12am" "12am - 6 am" "12pm - 6pm" "6am - 12pm" ...
## $ time driven
## $ trm len
                          : int 6 12 12 12 12 12 6 6 12 6 ...
## $ credit_score
                          : num 648 638 661 648 640 ...
## $ high_education_ind : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
mlr.full.mod.numclaims <- lm(numclaims ~ . , data = train.data)
step_model <- step(mlr.full.mod.numclaims)</pre>
## Start: AIC=-57948.47
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
      area + agecat + engine_type + max_power + driving_history_score +
      veh_color + marital_status + e_bill + time_of_week_driven +
##
      time_driven + trm_len + credit_score + high_education_ind
##
##
                          Df Sum of Sq
                                          RSS
                                                 AIC
                                0.3783 1738.0 -57960
## - veh_color
                                0.0270 1737.7 -57954
## - engine type
                           3
                           5
                                0.4334 1738.1 -57953
## - area
## - time driven
                           3
                                0.2286 1737.8 -57951
## - e_bill
                           1
                                0.0027 1737.6 -57950
## - credit score
                                0.0080 1737.6 -57950
                           1
## - veh_age
                           1 0.0110 1737.6 -57950
## - max_power
                           1 0.0155 1737.6 -57950
## - veh value
                           1 0.0281 1737.7 -57950
## - marital status
                           1 0.0456 1737.7 -57950
## - time_of_week_driven
                           1 0.0970 1737.7 -57949
## - driving_history_score 1
                                0.1453 1737.8 -57949
## - high_education_ind
                                0.1489 1737.8 -57949
                           1
## <none>
                                       1737.6 -57948
## - gender
                           1
                                0.1601 1737.8 -57948
## - veh_body
                          12
                                2.0629 1739.7 -57946
## - agecat
                           5
                                3.0581 1740.7 -57919
                                5.3467 1743.0 -57881
## - trm_len
                           1
## - exposure
                               18.7545 1756.4 -57708
##
## Step: AIC=-57959.55
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
      area + agecat + engine_type + max_power + driving_history_score +
##
      marital_status + e_bill + time_of_week_driven + time_driven +
##
      trm_len + credit_score + high_education_ind
##
                          Df Sum of Sq
                                          RSS
## - engine_type
                                0.0243 1738.0 -57965
## - area
                           5
                                0.4279 1738.4 -57964
## - time_driven
                           3
                              0.2274 1738.2 -57963
## - e_bill
                           1 0.0020 1738.0 -57962
```

1 0.0084 1738.0 -57961

- credit_score

```
## - veh age
                                 0.0113 1738.0 -57961
                                 0.0153 1738.0 -57961
## - max_power
                            1
## - veh value
                                 0.0284 1738.0 -57961
## - marital_status
                                 0.0421 1738.0 -57961
                            1
## - time_of_week_driven
                            1
                                 0.0987 1738.1 -57960
## - driving history score 1
                                 0.1425 1738.2 -57960
## - high education ind
                                 0.1499 1738.2 -57960
## <none>
                                        1738.0 -57960
## - gender
                                 0.1604 1738.2 -57959
                            1
## - veh_body
                           12
                                 2.0499 1740.0 -57957
## - agecat
                            5
                                 3.0559 1741.1 -57930
                                 5.3119 1743.3 -57893
## - trm len
                            1
## - exposure
                            1
                                18.7140 1756.7 -57719
##
## Step: AIC=-57965.23
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
       area + agecat + max_power + driving_history_score + marital_status +
##
##
       e_bill + time_of_week_driven + time_driven + trm_len + credit_score +
##
       high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                  AIC
## - area
                                 0.4320 1738.5 -57970
## - time_driven
                            3
                                 0.2277 1738.3 -57968
## - e bill
                                 0.0022 1738.0 -57967
                            1
## - credit score
                            1
                                 0.0084 1738.0 -57967
## - veh_age
                            1
                                 0.0111 1738.0 -57967
## - max_power
                                 0.0155 1738.0 -57967
                            1
                                 0.0280 1738.1 -57967
## - veh_value
                            1
                               0.0425 1738.1 -57967
## - marital_status
                            1
## - time_of_week_driven
                            1
                                 0.0987 1738.1 -57966
## - driving_history_score
                            1
                                 0.1428 1738.2 -57965
## - high_education_ind
                            1
                                 0.1486 1738.2 -57965
## <none>
                                        1738.0 -57965
                                 0.1607 1738.2 -57965
## - gender
                            1
## - veh_body
                           12
                                 2.0584 1740.1 -57962
                            5
                                 3.0490 1741.1 -57936
## - agecat
## - trm len
                            1
                                 5.3180 1743.3 -57898
## - exposure
                            1
                                18.7214 1756.8 -57725
##
## Step: AIC=-57969.61
## numclaims ~ veh value + exposure + veh body + veh age + gender +
##
       agecat + max_power + driving_history_score + marital_status +
       e_bill + time_of_week_driven + time_driven + trm_len + credit_score +
##
##
       high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                   AIC
## - time_driven
                                 0.2189 1738.7 -57973
                                 0.0017 1738.5 -57972
## - e_bill
## - credit_score
                                 0.0070 1738.5 -57972
                            1
## - veh_age
                            1
                                 0.0105 1738.5 -57971
                                 0.0132 1738.5 -57971
## - max_power
                            1
## - veh_value
                            1
                               0.0276 1738.5 -57971
## - marital_status
                            1 0.0423 1738.5 -57971
## - time of week driven
                            1 0.1047 1738.6 -57970
```

```
## - gender
                                 0.1402 1738.6 -57970
## - driving_history_score 1
                                 0.1435 1738.6 -57970
## <none>
                                        1738.5 -57970
## - high_education_ind
                                 0.1546 1738.6 -57970
                            1
## - veh_body
                           12
                                 2.0624 1740.5 -57967
                            5
## - agecat
                                 3.1297 1741.6 -57939
## - trm_len
                            1
                                 5.3168 1743.8 -57903
## - exposure
                            1
                                18.6922 1757.2 -57730
##
## Step: AIC=-57972.76
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
       agecat + max_power + driving_history_score + marital_status +
##
##
       e_bill + time_of_week_driven + trm_len + credit_score + high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                  AIC
## - e_bill
                                 0.0018 1738.7 -57975
## - credit_score
                                 0.0076 1738.7 -57975
                            1
## - veh age
                                 0.0099 1738.7 -57975
                                 0.0130 1738.7 -57975
## - max_power
                            1
## - veh value
                            1
                                 0.0281 1738.7 -57974
## - marital_status
                            1
                              0.0414 1738.7 -57974
## - time_of_week_driven
                            1 0.1054 1738.8 -57973
                              0.1385 1738.8 -57973
## - gender
                            1
## - driving_history_score 1
                                 0.1451 1738.8 -57973
## <none>
                                        1738.7 -57973
## - high_education_ind
                            1
                                 0.1570 1738.8 -57973
## - veh_body
                                 2.0669 1740.8 -57970
                           12
## - agecat
                            5
                                 3.1347 1741.8 -57942
                                 5.3223 1744.0 -57906
## - trm_len
                            1
## - exposure
                            1
                                18.6800 1757.4 -57733
##
## Step: AIC=-57974.74
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
       agecat + max_power + driving_history_score + marital_status +
##
       time_of_week_driven + trm_len + credit_score + high_education_ind
##
                           Df Sum of Sq
                                           RSS
## - credit_score
                                 0.0077 1738.7 -57977
                            1
## - veh_age
                                 0.0092 1738.7 -57977
                            1
                                 0.0131 1738.7 -57977
## - max_power
                            1
## - veh value
                            1
                                 0.0287 1738.7 -57976
## - marital_status
                                 0.0414 1738.7 -57976
                            1
## - time_of_week_driven
                            1
                                 0.1053 1738.8 -57975
                                 0.1382 1738.8 -57975
## - gender
                            1
## - driving_history_score
                            1
                                 0.1453 1738.8 -57975
                                        1738.7 -57975
## <none>
## - high_education_ind
                            1
                                 0.1573 1738.8 -57975
## - veh_body
                           12
                                 2.0667 1740.8 -57972
## - agecat
                            5
                                 3.1337 1741.8 -57944
## - trm_len
                            1
                                 5.3206 1744.0 -57908
                                18.6787 1757.4 -57735
## - exposure
                            1
## Step: AIC=-57976.64
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
```

```
##
       agecat + max_power + driving_history_score + marital_status +
##
       time_of_week_driven + trm_len + high_education_ind
##
##
                           Df Sum of Sq
                                            RSS
                                                   AIC
## - veh_age
                                 0.0091 1738.7 -57979
                                 0.0130 1738.7 -57978
## - max power
                            1
                                 0.0284 1738.7 -57978
## - veh value
                            1
                                 0.0416 1738.7 -57978
## - marital status
                            1
## - time_of_week_driven
                            1
                                 0.1053 1738.8 -57977
## - gender
                            1
                                 0.1379 1738.8 -57977
                                 0.1456 1738.8 -57977
## - driving_history_score 1
                                         1738.7 -57977
## <none>
                                 0.1571 1738.8 -57977
## - high_education_ind
                            1
                                 2.0654 1740.8 -57974
## - veh_body
                           12
## - agecat
                            5
                                 3.1343 1741.8 -57946
## - trm_len
                            1
                                 5.3356 1744.0 -57909
                                18.8303 1757.5 -57735
## - exposure
                            1
##
## Step: AIC=-57978.52
## numclaims ~ veh_value + exposure + veh_body + gender + agecat +
##
       max_power + driving_history_score + marital_status + time_of_week_driven +
##
       trm_len + high_education_ind
##
                                            RSS
##
                           Df Sum of Sa
## - max power
                                 0.0122 1738.7 -57980
## - veh value
                            1
                                 0.0198 1738.7 -57980
## - marital_status
                                 0.0417 1738.7 -57980
                            1
                                 0.1050 1738.8 -57979
## - time_of_week_driven
                            1
                                 0.1421 1738.8 -57979
## - gender
                            1
## - driving_history_score
                                 0.1453 1738.8 -57979
                            1
## <none>
                                         1738.7 -57979
## - high_education_ind
                            1
                                 0.1565 1738.8 -57978
## - veh_body
                           12
                                 2.0670 1740.8 -57976
                            5
                                 3.1427 1741.8 -57948
## - agecat
## - trm len
                            1
                                 5.3372 1744.0 -57911
                            1
                                18.8215 1757.5 -57737
## - exposure
##
## Step: AIC=-57980.36
## numclaims ~ veh_value + exposure + veh_body + gender + agecat +
##
       driving_history_score + marital_status + time_of_week_driven +
##
       trm len + high education ind
##
##
                           Df Sum of Sq
                                            RSS
## - veh_value
                                 0.0217 1738.7 -57982
                                 0.0411 1738.8 -57982
## - marital_status
                            1
                                 0.1058 1738.8 -57981
## - time_of_week_driven
                            1
## - gender
                            1
                                 0.1409 1738.8 -57981
                                 0.1455 1738.9 -57980
## - driving_history_score 1
## <none>
                                        1738.7 -57980
## - high_education_ind
                            1
                                 0.1568 1738.9 -57980
                                 2.0552 1740.8 -57978
                           12
## - veh_body
## - agecat
                            5
                                 3.1430 1741.8 -57950
## - trm len
                            1
                                 5.3364 1744.0 -57913
## - exposure
                            1 18.8213 1757.5 -57739
```

```
##
## Step: AIC=-57982.08
## numclaims ~ exposure + veh_body + gender + agecat + driving_history_score +
       marital_status + time_of_week_driven + trm_len + high_education_ind
##
##
##
                           Df Sum of Sq
                                           RSS
                                                   ATC
                                 0.0411 1738.8 -57984
## - marital status
                            1
                                 0.1067 1738.8 -57983
## - time_of_week_driven
                            1
## - gender
                            1
                                 0.1440 1738.9 -57982
## - driving_history_score 1
                                 0.1442 1738.9 -57982
## <none>
                                        1738.7 -57982
                                 0.1576 1738.9 -57982
## - high_education_ind
                            1
## - veh_body
                           12
                                 2.0379 1740.8 -57980
                                 3.1318 1741.9 -57951
## - agecat
                            5
## - trm_len
                            1
                                 5.3348 1744.1 -57915
## - exposure
                            1
                                18.8233 1757.6 -57741
##
## Step: AIC=-57983.54
## numclaims ~ exposure + veh_body + gender + agecat + driving_history_score +
       time_of_week_driven + trm_len + high_education_ind
##
##
                           Df Sum of Sq
## - time_of_week_driven
                                 0.1065 1738.9 -57984
                            1
## - gender
                                 0.1451 1738.9 -57984
                            1
## - driving_history_score 1
                                 0.1468 1738.9 -57984
## <none>
                                        1738.8 -57984
## - high_education_ind
                                 0.1591 1738.9 -57983
                            1
## - veh_body
                           12
                                 2.0414 1740.8 -57981
## - agecat
                            5
                                 3.1375 1741.9 -57953
## - trm_len
                            1
                                 5.3332 1744.1 -57916
## - exposure
                            1
                               18.8049 1757.6 -57742
##
## Step: AIC=-57984.16
## numclaims ~ exposure + veh_body + gender + agecat + driving_history_score +
##
       trm_len + high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                   AIC
                                 0.1447 1739.0 -57984
## - gender
                            1
## - driving_history_score 1
                                 0.1458 1739.0 -57984
                                        1738.9 -57984
## <none>
## - high education ind
                                 0.1581 1739.0 -57984
                            1
## - veh body
                           12
                                 2.0433 1740.9 -57982
## - agecat
                            5
                                 3.1287 1742.0 -57953
## - trm_len
                                 5.3180 1744.2 -57917
                            1
## - exposure
                            1
                               18.8121 1757.7 -57743
##
## Step: AIC=-57984.28
## numclaims ~ exposure + veh_body + agecat + driving_history_score +
##
       trm_len + high_education_ind
##
                           Df Sum of Sq
                                           RSS
                                                   AIC
## - driving_history_score 1
                                 0.1465 1739.2 -57984
## <none>
                                        1739.0 -57984
## - high education ind
                            1
                                 0.1594 1739.2 -57984
```

```
## - veh_body
               12
                                2.0682 1741.1 -57981
                        5 3.2078 1742.2 -57953
## - agecat
## - trm len
                         1 5.3144 1744.3 -57917
                          1 18.7694 1757.8 -57743
## - exposure
## Step: AIC=-57984.37
## numclaims ~ exposure + veh_body + agecat + trm_len + high_education_ind
                       Df Sum of Sq
##
                                      RSS
                                              AIC
## <none>
                                    1739.2 -57984
## - high_education_ind 1 0.1593 1739.3 -57984
## - veh_body 12 2.0602 1741.2 -57982
                     5 3.2039 1742.4 -57953
## - agecat
## - trm_len
                      1 5.3123 1744.5 -57917
                  1 18.7551 1757.9 -57744
## - exposure
formula(step_model)
## numclaims ~ exposure + veh_body + agecat + trm_len + high_education_ind
AIC(step_model)
## [1] 6207.571
BIC(step_model)
## [1] 6384.155
final_model.forward.numclaims <- step(mlr.full.mod.numclaims, direction = "forward")</pre>
## Start: AIC=-57948.47
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
      area + agecat + engine_type + max_power + driving_history_score +
      veh_color + marital_status + e_bill + time_of_week_driven +
##
      time_driven + trm_len + credit_score + high_education_ind
formula(final_model.forward.numclaims)
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
      area + agecat + engine_type + max_power + driving_history_score +
##
      veh_color + marital_status + e_bill + time_of_week_driven +
      time_driven + trm_len + credit_score + high_education_ind
AIC(final_model.forward.numclaims)
## [1] 6243.471
BIC(final model.forward.numclaims)
## [1] 6644.798
final model.backward.numclaims <- step(mlr.full.mod.numclaims, direction = "backward")
## Start: AIC=-57948.47
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
      area + agecat + engine_type + max_power + driving_history_score +
##
##
      veh_color + marital_status + e_bill + time_of_week_driven +
##
      time_driven + trm_len + credit_score + high_education_ind
##
##
                          Df Sum of Sq
                                         RSS
                                                AIC
```

```
## - veh_color
                           8
                                0.3783 1738.0 -57960
                                0.0270 1737.7 -57954
## - engine_type
                           3
## - area
                           5
                                0.4334 1738.1 -57953
## - time_driven
                           3
                                0.2286 1737.8 -57951
## - e bill
                           1
                                0.0027 1737.6 -57950
                          1 0.0080 1737.6 -57950
## - credit score
## - veh age
                          1 0.0110 1737.6 -57950
## - max_power
                          1 0.0155 1737.6 -57950
## - veh_value
                           1 0.0281 1737.7 -57950
## - marital_status
                          1 0.0456 1737.7 -57950
## - time_of_week_driven
                           1 0.0970 1737.7 -57949
                                0.1453 1737.8 -57949
## - driving_history_score 1
## - high_education_ind
                           1
                                0.1489 1737.8 -57949
## <none>
                                       1737.6 -57948
## - gender
                                0.1601 1737.8 -57948
                           1
## - veh_body
                          12
                                2.0629 1739.7 -57946
                           5
                                3.0581 1740.7 -57919
## - agecat
## - trm len
                                5.3467 1743.0 -57881
                               18.7545 1756.4 -57708
## - exposure
                           1
## Step: AIC=-57959.55
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
      area + agecat + engine_type + max_power + driving_history_score +
##
      marital_status + e_bill + time_of_week_driven + time_driven +
##
##
      trm_len + credit_score + high_education_ind
##
##
                          Df Sum of Sq
                                          RSS
                                0.0243 1738.0 -57965
## - engine_type
                                0.4279 1738.4 -57964
## - area
## - time_driven
                           3
                                0.2274 1738.2 -57963
## - e_bill
                           1
                                0.0020 1738.0 -57962
## - credit_score
                           1
                                0.0084 1738.0 -57961
## - veh_age
                                0.0113 1738.0 -57961
                                0.0153 1738.0 -57961
## - max_power
                           1
## - veh value
                           1
                                0.0284 1738.0 -57961
                           1 0.0421 1738.0 -57961
## - marital_status
## - time of week driven
                           1 0.0987 1738.1 -57960
## - driving_history_score 1
                                0.1425 1738.2 -57960
## - high_education_ind
                                0.1499 1738.2 -57960
                                       1738.0 -57960
## <none>
## - gender
                                0.1604 1738.2 -57959
                           1
## - veh body
                          12
                                2.0499 1740.0 -57957
## - agecat
                           5
                                3.0559 1741.1 -57930
## - trm_len
                                5.3119 1743.3 -57893
                           1
## - exposure
                           1
                               18.7140 1756.7 -57719
##
## Step: AIC=-57965.23
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
      area + agecat + max_power + driving_history_score + marital_status +
##
      e_bill + time_of_week_driven + time_driven + trm_len + credit_score +
##
      high_education_ind
##
##
                          Df Sum of Sq
                                          RSS
                                                 ATC
## - area
                           5 0.4320 1738.5 -57970
```

```
## - time driven
                                0.2277 1738.3 -57968
## - e_bill
                                0.0022 1738.0 -57967
                           1
## - credit score
                                0.0084 1738.0 -57967
## - veh_age
                                0.0111 1738.0 -57967
                           1
## - max power
                           1
                                0.0155 1738.0 -57967
                          1 0.0280 1738.1 -57967
## - veh value
## - marital_status
                          1 0.0425 1738.1 -57967
                         1 0.0987 1738.1 -57966
## - time_of_week_driven
## - driving_history_score 1
                               0.1428 1738.2 -57965
## - high_education_ind 1
                                0.1486 1738.2 -57965
## <none>
                                       1738.0 -57965
                                0.1607 1738.2 -57965
## - gender
                           1
## - veh_body
                          12
                                2.0584 1740.1 -57962
## - agecat
                           5
                                3.0490 1741.1 -57936
## - trm_len
                                5.3180 1743.3 -57898
                           1
## - exposure
                           1
                               18.7214 1756.8 -57725
##
## Step: AIC=-57969.61
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
      agecat + max_power + driving_history_score + marital_status +
##
      e_bill + time_of_week_driven + time_driven + trm_len + credit_score +
##
      high_education_ind
##
                          Df Sum of Sa
##
                                          RSS
## - time driven
                                0.2189 1738.7 -57973
## - e bill
                           1
                                0.0017 1738.5 -57972
## - credit_score
                                0.0070 1738.5 -57972
                           1
## - veh_age
                           1
                               0.0105 1738.5 -57971
                          1 0.0132 1738.5 -57971
## - max_power
## - veh_value
                          1 0.0276 1738.5 -57971
## - marital_status
                           1 0.0423 1738.5 -57971
## - time_of_week_driven
                           1
                             0.1047 1738.6 -57970
## - gender
                                0.1402 1738.6 -57970
                                0.1435 1738.6 -57970
## - driving_history_score 1
## <none>
                                       1738.5 -57970
                                0.1546 1738.6 -57970
## - high_education_ind
                           1
## - veh body
                          12
                                2.0624 1740.5 -57967
## - agecat
                           5
                                3.1297 1741.6 -57939
## - trm_len
                           1
                                5.3168 1743.8 -57903
## - exposure
                           1
                               18.6922 1757.2 -57730
## Step: AIC=-57972.76
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
      agecat + max_power + driving_history_score + marital_status +
##
      e_bill + time_of_week_driven + trm_len + credit_score + high_education_ind
##
##
                          Df Sum of Sq
                                          RSS
## - e_bill
                                0.0018 1738.7 -57975
## - credit_score
                                0.0076 1738.7 -57975
                           1
## - veh_age
                           1
                                0.0099 1738.7 -57975
                               0.0130 1738.7 -57975
## - max_power
                           1
## - veh value
                          1 0.0281 1738.7 -57974
## - marital_status
                           1 0.0414 1738.7 -57974
## - time of week driven
                           1 0.1054 1738.8 -57973
```

```
## - gender
                                 0.1385 1738.8 -57973
## - driving_history_score 1
                                 0.1451 1738.8 -57973
                                        1738.7 -57973
## - high_education_ind
                                 0.1570 1738.8 -57973
                           1
## - veh_body
                           12
                                 2.0669 1740.8 -57970
                           5
## - agecat
                                 3.1347 1741.8 -57942
## - trm_len
                           1
                                 5.3223 1744.0 -57906
## - exposure
                            1
                                18.6800 1757.4 -57733
##
## Step: AIC=-57974.74
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
       agecat + max_power + driving_history_score + marital_status +
##
##
       time_of_week_driven + trm_len + credit_score + high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                  AIC
## - credit_score
                            1
                                 0.0077 1738.7 -57977
                                 0.0092 1738.7 -57977
## - veh_age
                            1
## - max power
                                 0.0131 1738.7 -57977
                                 0.0287 1738.7 -57976
## - veh_value
                            1
## - marital status
                            1
                                0.0414 1738.7 -57976
## - time_of_week_driven
                            1
                              0.1053 1738.8 -57975
## - gender
                                0.1382 1738.8 -57975
                            1
## - driving_history_score 1
                                0.1453 1738.8 -57975
## <none>
                                        1738.7 -57975
## - high_education_ind
                            1
                                0.1573 1738.8 -57975
## - veh body
                           12
                                 2.0667 1740.8 -57972
## - agecat
                           5
                                 3.1337 1741.8 -57944
## - trm_len
                            1
                                 5.3206 1744.0 -57908
                                18.6787 1757.4 -57735
## - exposure
                            1
##
## Step: AIC=-57976.64
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
       agecat + max_power + driving_history_score + marital_status +
##
       time_of_week_driven + trm_len + high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                  ATC
## - veh age
                                0.0091 1738.7 -57979
## - max_power
                                 0.0130 1738.7 -57978
                            1
                                 0.0284 1738.7 -57978
## - veh_value
                            1
## - marital_status
                                0.0416 1738.7 -57978
                            1
                                 0.1053 1738.8 -57977
## - time_of_week_driven
                            1
                                 0.1379 1738.8 -57977
## - gender
                            1
## - driving_history_score 1
                                 0.1456 1738.8 -57977
## <none>
                                        1738.7 -57977
                                 0.1571 1738.8 -57977
## - high_education_ind
                           1
                           12
                                 2.0654 1740.8 -57974
## - veh_body
## - agecat
                            5
                                 3.1343 1741.8 -57946
## - trm_len
                           1
                                 5.3356 1744.0 -57909
## - exposure
                            1 18.8303 1757.5 -57735
## Step: AIC=-57978.52
## numclaims ~ veh_value + exposure + veh_body + gender + agecat +
##
      max_power + driving_history_score + marital_status + time_of_week_driven +
##
      trm_len + high_education_ind
```

```
##
##
                           Df Sum of Sq
                                            RSS
                                                   ATC
## - max power
                                 0.0122 1738.7 -57980
## - veh_value
                                 0.0198 1738.7 -57980
                            1
## - marital_status
                            1
                                 0.0417 1738.7 -57980
## - time_of_week_driven
                                 0.1050 1738.8 -57979
                            1
## - gender
                            1
                                 0.1421 1738.8 -57979
## - driving_history_score 1
                                 0.1453 1738.8 -57979
## <none>
                                         1738.7 -57979
## - high_education_ind
                            1
                                 0.1565 1738.8 -57978
## - veh_body
                           12
                                 2.0670 1740.8 -57976
                            5
                                 3.1427 1741.8 -57948
## - agecat
## - trm_len
                            1
                                 5.3372 1744.0 -57911
## - exposure
                            1
                                18.8215 1757.5 -57737
##
## Step: AIC=-57980.36
  numclaims ~ veh_value + exposure + veh_body + gender + agecat +
       driving_history_score + marital_status + time_of_week_driven +
##
       trm_len + high_education_ind
##
##
                           Df Sum of Sq
                                            RSS
                                                   AIC
## - veh_value
                                 0.0217 1738.7 -57982
                                 0.0411 1738.8 -57982
## - marital_status
                            1
                                 0.1058 1738.8 -57981
## - time_of_week_driven
                            1
## - gender
                            1
                                 0.1409 1738.8 -57981
## - driving_history_score 1
                                 0.1455 1738.9 -57980
## <none>
                                         1738.7 -57980
                                 0.1568 1738.9 -57980
## - high_education_ind
                            1
                                 2.0552 1740.8 -57978
## - veh_body
                           12
## - agecat
                            5
                                 3.1430 1741.8 -57950
## - trm_len
                            1
                                 5.3364 1744.0 -57913
## - exposure
                            1
                                18.8213 1757.5 -57739
##
## Step: AIC=-57982.08
## numclaims ~ exposure + veh_body + gender + agecat + driving_history_score +
       marital_status + time_of_week_driven + trm_len + high_education_ind
##
##
##
                           Df Sum of Sq
                                            RSS
## - marital status
                                 0.0411 1738.8 -57984
                            1
## - time_of_week_driven
                                 0.1067 1738.8 -57983
                            1
                                 0.1440 1738.9 -57982
## - gender
                            1
## - driving_history_score 1
                                 0.1442 1738.9 -57982
## <none>
                                         1738.7 -57982
## - high_education_ind
                                 0.1576 1738.9 -57982
                            1
## - veh_body
                           12
                                 2.0379 1740.8 -57980
## - agecat
                            5
                                 3.1318 1741.9 -57951
## - trm_len
                            1
                                 5.3348 1744.1 -57915
## - exposure
                            1
                                18.8233 1757.6 -57741
##
## Step: AIC=-57983.54
## numclaims ~ exposure + veh_body + gender + agecat + driving_history_score +
##
       time_of_week_driven + trm_len + high_education_ind
##
##
                           Df Sum of Sq
                                            RSS
                                                   AIC
```

```
## - time_of_week_driven
                                0.1065 1738.9 -57984
                           1
                                0.1451 1738.9 -57984
## - gender
                           1
## - driving_history_score 1
                                0.1468 1738.9 -57984
## <none>
                                       1738.8 -57984
                                0.1591 1738.9 -57983
## - high_education_ind
                          1
## - veh body
                          12
                                2.0414 1740.8 -57981
## - agecat
                          5
                              3.1375 1741.9 -57953
## - trm len
                           1
                               5.3332 1744.1 -57916
## - exposure
                              18.8049 1757.6 -57742
##
## Step: AIC=-57984.16
## numclaims ~ exposure + veh_body + gender + agecat + driving_history_score +
      trm_len + high_education_ind
##
##
                          Df Sum of Sq
                                          RSS
                                                 AIC
## - gender
                           1
                                0.1447 1739.0 -57984
## - driving_history_score 1
                                0.1458 1739.0 -57984
## <none>
                                       1738.9 -57984
## - high_education_ind
                                0.1581 1739.0 -57984
                           1
## - veh body
                          12
                                2.0433 1740.9 -57982
## - agecat
                           5
                                3.1287 1742.0 -57953
## - trm len
                           1
                               5.3180 1744.2 -57917
## - exposure
                           1 18.8121 1757.7 -57743
##
## Step: AIC=-57984.28
## numclaims ~ exposure + veh_body + agecat + driving_history_score +
      trm_len + high_education_ind
##
##
##
                          Df Sum of Sq
                                          RSS
                                                 AIC
## - driving_history_score 1
                                0.1465 1739.2 -57984
## <none>
                                       1739.0 -57984
## - high_education_ind
                           1
                                0.1594 1739.2 -57984
## - veh_body
                          12
                                2.0682 1741.1 -57981
                           5
                                3.2078 1742.2 -57953
## - agecat
## - trm_len
                           1
                                5.3144 1744.3 -57917
                           1 18.7694 1757.8 -57743
## - exposure
##
## Step: AIC=-57984.37
## numclaims ~ exposure + veh_body + agecat + trm_len + high_education_ind
##
##
                       Df Sum of Sq
                                       RSS
                                              AIC
## <none>
                                    1739.2 -57984
## - high_education_ind 1
                             0.1593 1739.3 -57984
## - veh_body
                   12
                           2.0602 1741.2 -57982
## - agecat
                             3.2039 1742.4 -57953
                        5
## - trm_len
                             5.3123 1744.5 -57917
                        1
## - exposure
                        1
                            18.7551 1757.9 -57744
formula(final_model.backward.numclaims)
## numclaims ~ exposure + veh_body + agecat + trm_len + high_education_ind
AIC(final model.backward.numclaims)
```

[1] 6207.571

```
BIC(final_model.backward.numclaims)
## [1] 6384.155
final_model.backward.numclaims <- step(mlr.full.mod.numclaims, direction = "both")
## Start: AIC=-57948.47
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
      area + agecat + engine_type + max_power + driving_history_score +
##
##
      veh_color + marital_status + e_bill + time_of_week_driven +
##
      time_driven + trm_len + credit_score + high_education_ind
##
                          Df Sum of Sq
##
                                         RSS
                                                AIC
## - veh_color
                                0.3783 1738.0 -57960
## - engine_type
                           3
                                0.0270 1737.7 -57954
## - area
                           5
                               0.4334 1738.1 -57953
## - time driven
                           3
                               0.2286 1737.8 -57951
                          1 0.0027 1737.6 -57950
## - e_bill
## - credit_score
                         1 0.0080 1737.6 -57950
                          1 0.0110 1737.6 -57950
## - veh_age
                         1 0.0155 1737.6 -57950
## - max_power
## - veh_value
                         1 0.0281 1737.7 -57950
## - marital_status
                         1 0.0456 1737.7 -57950
## - time_of_week_driven
                          1 0.0970 1737.7 -57949
## - driving_history_score 1 0.1453 1737.8 -57949
## - high_education_ind 1 0.1489 1737.8 -57949
## <none>
                                      1737.6 -57948
                               0.1601 1737.8 -57948
## - gender
                          1
                                2.0629 1739.7 -57946
## - veh_body
                          12
## - agecat
                          5
                                3.0581 1740.7 -57919
## - trm len
                               5.3467 1743.0 -57881
                          1
## - exposure
                           1
                               18.7545 1756.4 -57708
##
## Step: AIC=-57959.55
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
      area + agecat + engine_type + max_power + driving_history_score +
##
      marital_status + e_bill + time_of_week_driven + time_driven +
##
      trm_len + credit_score + high_education_ind
##
##
                          Df Sum of Sq
                                         RSS
## - engine_type
                               0.0243 1738.0 -57965
## - area
                           5
                                0.4279 1738.4 -57964
                               0.2274 1738.2 -57963
## - time_driven
                          3
## - e_bill
                           1
                               0.0020 1738.0 -57962
## - credit_score
                          1
                               0.0084 1738.0 -57961
                               0.0113 1738.0 -57961
## - veh_age
                           1
## - max_power
                           1
                               0.0153 1738.0 -57961
## - veh value
                          1 0.0284 1738.0 -57961
## - marital status
                          1 0.0421 1738.0 -57961
## - time_of_week_driven
                          1 0.0987 1738.1 -57960
## - driving_history_score 1
                               0.1425 1738.2 -57960
## - high_education_ind 1
                               0.1499 1738.2 -57960
## <none>
                                      1738.0 -57960
## - gender
                               0.1604 1738.2 -57959
                           1
```

```
## - veh body
                          12
                                2.0499 1740.0 -57957
                                0.3783 1737.6 -57948
## + veh_color
                           8
## - agecat
                                3.0559 1741.1 -57930
                           5
## - trm_len
                           1
                                5.3119 1743.3 -57893
## - exposure
                              18.7140 1756.7 -57719
##
## Step: AIC=-57965.23
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
       area + agecat + max_power + driving_history_score + marital_status +
##
       e_bill + time_of_week_driven + time_driven + trm_len + credit_score +
##
       high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                  AIC
## - area
                                0.4320 1738.5 -57970
## - time_driven
                           3
                                0.2277 1738.3 -57968
## - e_bill
                           1
                                0.0022 1738.0 -57967
                                0.0084 1738.0 -57967
## - credit_score
                           1
## - veh age
                                0.0111 1738.0 -57967
                                0.0155 1738.0 -57967
## - max_power
                           1
## - veh value
                           1
                                0.0280 1738.1 -57967
## - marital_status
                           1
                              0.0425 1738.1 -57967
## - time_of_week_driven
                                0.0987 1738.1 -57966
                           1
## - driving_history_score 1
                                0.1428 1738.2 -57965
                                0.1486 1738.2 -57965
## - high education ind
## <none>
                                        1738.0 -57965
## - gender
                           1
                                0.1607 1738.2 -57965
## - veh_body
                          12
                                2.0584 1740.1 -57962
## + engine_type
                            3
                                0.0243 1738.0 -57960
## + veh_color
                           8
                                0.3756 1737.7 -57954
## - agecat
                           5
                              3.0490 1741.1 -57936
## - trm_len
                           1
                               5.3180 1743.3 -57898
## - exposure
                              18.7214 1756.8 -57725
##
## Step: AIC=-57969.61
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
      agecat + max_power + driving_history_score + marital_status +
##
       e_bill + time_of_week_driven + time_driven + trm_len + credit_score +
##
      high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                  ATC
                                0.2189 1738.7 -57973
## - time driven
## - e bill
                                0.0017 1738.5 -57972
                           1
                                0.0070 1738.5 -57972
## - credit_score
                           1
## - veh_age
                                0.0105 1738.5 -57971
                           1
## - max_power
                           1
                                0.0132 1738.5 -57971
## - veh_value
                                0.0276 1738.5 -57971
                           1
## - marital_status
                            1
                                0.0423 1738.5 -57971
## - time_of_week_driven
                                0.1047 1738.6 -57970
## - gender
                            1
                                0.1402 1738.6 -57970
## - driving_history_score 1
                                0.1435 1738.6 -57970
                                       1738.5 -57970
## <none>
## - high education ind
                           1
                                0.1546 1738.6 -57970
## - veh body
                           12
                                2.0624 1740.5 -57967
## + area
                           5
                                0.4320 1738.0 -57965
```

```
3
                                0.0284 1738.4 -57964
## + engine_type
                                0.3698 1738.1 -57958
## + veh_color
                           8
## - agecat
                           5 3.1297 1741.6 -57939
## - trm_len
                           1
                                5.3168 1743.8 -57903
## - exposure
                              18.6922 1757.2 -57730
##
## Step: AIC=-57972.76
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
       agecat + max_power + driving_history_score + marital_status +
##
       e_bill + time_of_week_driven + trm_len + credit_score + high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                  AIC
## - e_bill
                                 0.0018 1738.7 -57975
## - credit_score
                           1
                                 0.0076 1738.7 -57975
## - veh_age
                                 0.0099 1738.7 -57975
                            1
## - max_power
                            1
                                0.0130 1738.7 -57975
## - veh_value
                                0.0281 1738.7 -57974
                            1
## - marital status
                            1 0.0414 1738.7 -57974
                                0.1054 1738.8 -57973
## - time_of_week_driven
                            1
## - gender
                            1
                                 0.1385 1738.8 -57973
## - driving_history_score 1
                                0.1451 1738.8 -57973
## <none>
                                        1738.7 -57973
## - high_education_ind
                                 0.1570 1738.8 -57973
                            1
                                 2.0669 1740.8 -57970
## - veh body
                           12
## + time_driven
                            3
                                0.2189 1738.5 -57970
## + area
                            5
                                0.4232 1738.3 -57968
                            3
                                0.0288 1738.7 -57967
## + engine_type
                            8
## + veh_color
                                0.3686 1738.3 -57962
                           5
                                3.1347 1741.8 -57942
## - agecat
## - trm_len
                           1
                                5.3223 1744.0 -57906
## - exposure
                            1
                                18.6800 1757.4 -57733
##
## Step: AIC=-57974.74
## numclaims ~ veh_value + exposure + veh_body + veh_age + gender +
##
       agecat + max_power + driving_history_score + marital_status +
##
       time_of_week_driven + trm_len + credit_score + high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
## - credit score
                                0.0077 1738.7 -57977
                                0.0092 1738.7 -57977
## - veh_age
                            1
                                 0.0131 1738.7 -57977
## - max power
                           1
## - veh value
                                0.0287 1738.7 -57976
                            1
## - marital_status
                            1
                                0.0414 1738.7 -57976
## - time_of_week_driven
                                0.1053 1738.8 -57975
                            1
## - gender
                            1
                                0.1382 1738.8 -57975
## - driving_history_score
                                 0.1453 1738.8 -57975
                            1
## <none>
                                        1738.7 -57975
## - high_education_ind
                                 0.1573 1738.8 -57975
## + e_bill
                                 0.0018 1738.7 -57973
                            1
## - veh_body
                           12
                                 2.0667 1740.8 -57972
## + time_driven
                                0.2189 1738.5 -57972
                           3
## + area
                           5
                                0.4228 1738.3 -57970
## + engine_type
                          3
                                0.0289 1738.7 -57969
## + veh color
                           8
                                0.3679 1738.3 -57964
```

```
## - agecat
                           5
                                3.1337 1741.8 -57944
                                5.3206 1744.0 -57908
## - trm_len
                           1
                           1 18.6787 1757.4 -57735
## - exposure
##
## Step: AIC=-57976.64
## numclaims ~ veh value + exposure + veh body + veh age + gender +
      agecat + max_power + driving_history_score + marital_status +
      time_of_week_driven + trm_len + high_education_ind
##
##
##
                          Df Sum of Sq
                                          RSS
## - veh_age
                                0.0091 1738.7 -57979
                                0.0130 1738.7 -57978
## - max_power
                           1
## - veh_value
                           1
                                0.0284 1738.7 -57978
## - marital_status
                           1 0.0416 1738.7 -57978
                           1 0.1053 1738.8 -57977
## - time_of_week_driven
## - gender
                           1
                              0.1379 1738.8 -57977
## - driving_history_score 1
                                0.1456 1738.8 -57977
## <none>
                                       1738.7 -57977
## - high_education_ind
                                0.1571 1738.8 -57977
                           1
## + credit score
                           1
                                0.0077 1738.7 -57975
## + e_bill
                           1
                                0.0019 1738.7 -57975
## - veh body
                          12
                                2.0654 1740.8 -57974
## + time_driven
                          3
                                0.2195 1738.5 -57973
                           5
                                0.4213 1738.3 -57972
## + area
## + engine_type
                         3 0.0290 1738.7 -57971
## + veh_color
                         8 0.3683 1738.3 -57965
                         5
## - agecat
                                3.1343 1741.8 -57946
## - trm_len
                          1
                                5.3356 1744.0 -57909
                           1 18.8303 1757.5 -57735
## - exposure
##
## Step: AIC=-57978.52
## numclaims ~ veh_value + exposure + veh_body + gender + agecat +
##
      max_power + driving_history_score + marital_status + time_of_week_driven +
##
      trm_len + high_education_ind
##
                          Df Sum of Sq
##
                                          RSS
                                                 AIC
## - max power
                              0.0122 1738.7 -57980
## - veh_value
                                0.0198 1738.7 -57980
                           1
## - marital_status
                           1
                               0.0417 1738.7 -57980
                           1 0.1050 1738.8 -57979
## - time_of_week_driven
                                0.1421 1738.8 -57979
## - gender
## - driving_history_score 1
                                0.1453 1738.8 -57979
## <none>
                                       1738.7 -57979
                                0.1565 1738.8 -57978
## - high_education_ind
                           1
                                0.0091 1738.7 -57977
## + veh_age
                           1
                                0.0076 1738.7 -57977
## + credit_score
                           1
## + e_bill
                           1
                                0.0012 1738.7 -57977
## - veh_body
                          12
                                2.0670 1740.8 -57976
## + time_driven
                          3
                                0.2189 1738.5 -57975
                         5
## + area
                                0.4209 1738.3 -57974
                        3 0.0287 1738.7 -57973
8 0.3687 1738.3 -57967
## + engine_type
## + veh_color
## - agecat
                         5 3.1427 1741.8 -57948
                          1 5.3372 1744.0 -57911
## - trm len
```

```
## - exposure
                           1 18.8215 1757.5 -57737
##
## Step: AIC=-57980.36
## numclaims ~ veh_value + exposure + veh_body + gender + agecat +
      driving_history_score + marital_status + time_of_week_driven +
##
      trm len + high education ind
##
##
                          Df Sum of Sq
                                          RSS
                                                 ATC
## - veh_value
                                0.0217 1738.7 -57982
## - marital_status
                           1
                                0.0411 1738.8 -57982
## - time_of_week_driven
                           1
                                0.1058 1738.8 -57981
## - gender
                                0.1409 1738.8 -57981
                           1
## - driving_history_score 1
                                0.1455 1738.9 -57980
## <none>
                                       1738.7 -57980
## - high_education_ind
                                0.1568 1738.9 -57980
                           1
## + max_power
                           1
                                0.0122 1738.7 -57979
                                0.0083 1738.7 -57978
## + veh_age
                           1
## + credit score
                                0.0075 1738.7 -57978
                                0.0013 1738.7 -57978
## + e_bill
                           1
## - veh body
                          12
                                2.0552 1740.8 -57978
## + time_driven
                         3
                              0.2187 1738.5 -57977
## + area
                          5 0.4188 1738.3 -57976
                          3 0.0289 1738.7 -57975
## + engine type
                           8
                                0.3685 1738.3 -57969
## + veh color
                           5 3.1430 1741.8 -57950
## - agecat
## - trm len
                           1
                                5.3364 1744.0 -57913
## - exposure
                           1 18.8213 1757.5 -57739
##
## Step: AIC=-57982.08
## numclaims ~ exposure + veh_body + gender + agecat + driving_history_score +
##
      marital_status + time_of_week_driven + trm_len + high_education_ind
##
##
                          Df Sum of Sq
                                          RSS
                                                 AIC
                                0.0411 1738.8 -57984
## - marital_status
                           1
## - time of week driven
                           1
                                0.1067 1738.8 -57983
                                0.1440 1738.9 -57982
## - gender
                           1
## - driving_history_score 1
                                0.1442 1738.9 -57982
## <none>
                                       1738.7 -57982
## - high_education_ind
                                0.1576 1738.9 -57982
                           1
## + veh_value
                                0.0217 1738.7 -57980
                           1
## + max_power
                           1
                                0.0142 1738.7 -57980
## + credit score
                                0.0073 1738.7 -57980
                           1
                                0.0031 1738.7 -57980
## + e_bill
                           1
## + veh_age
                                0.0008 1738.7 -57980
                           1
## - veh_body
                          12
                                2.0379 1740.8 -57980
                                0.2201 1738.5 -57979
## + time_driven
                           3
## + area
                           5
                                0.4186 1738.3 -57978
                          3 0.0289 1738.7 -57976
## + engine_type
## + veh_color
                         8 0.3679 1738.4 -57971
## - agecat
                           5
                                3.1318 1741.9 -57951
## - trm_len
                               5.3348 1744.1 -57915
                           1
## - exposure
                           1 18.8233 1757.6 -57741
##
## Step: AIC=-57983.54
```

```
## numclaims ~ exposure + veh_body + gender + agecat + driving_history_score +
##
       time_of_week_driven + trm_len + high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                  ATC
## - time_of_week_driven
                                0.1065 1738.9 -57984
                                0.1451 1738.9 -57984
## - gender
                            1
## - driving_history_score 1
                                0.1468 1738.9 -57984
## <none>
                                        1738.8 -57984
## - high_education_ind
                                0.1591 1738.9 -57983
                           1
## + marital_status
                            1
                                0.0411 1738.7 -57982
## + veh_value
                           1
                                0.0217 1738.8 -57982
                                0.0136 1738.8 -57982
## + max_power
                           1
## + credit_score
                           1
                                0.0075 1738.8 -57982
## + e_bill
                           1
                                0.0030 1738.8 -57982
                                0.0008 1738.8 -57982
## + veh_age
                           1
## - veh_body
                           12
                                2.0414 1740.8 -57981
                         3
## + time_driven
                                0.2192 1738.5 -57980
## + area
                           5
                                0.4184 1738.3 -57979
## + engine_type
                           3
                                0.0293 1738.7 -57978
## + veh color
                           8
                                0.3646 1738.4 -57972
## - agecat
                           5
                                3.1375 1741.9 -57953
## - trm len
                           1
                               5.3332 1744.1 -57916
## - exposure
                           1 18.8049 1757.6 -57742
##
## Step: AIC=-57984.16
## numclaims ~ exposure + veh_body + gender + agecat + driving_history_score +
       trm_len + high_education_ind
##
##
##
                           Df Sum of Sq
                                           RSS
                                                  AIC
## - gender
                                0.1447 1739.0 -57984
                            1
## - driving_history_score 1
                                0.1458 1739.0 -57984
## <none>
                                        1738.9 -57984
## - high_education_ind
                                0.1581 1739.0 -57984
                                0.1065 1738.8 -57984
## + time_of_week_driven
                            1
## + marital status
                            1
                                0.0409 1738.8 -57983
                                0.0225 1738.9 -57982
## + veh_value
                           1
## + max power
                           1
                                0.0145 1738.9 -57982
## + credit_score
                                0.0075 1738.9 -57982
                           1
## + e_bill
                                0.0030 1738.9 -57982
                           1
                                0.0010 1738.9 -57982
## + veh_age
                           1
                                2.0433 1740.9 -57982
## - veh body
                          12
## + time driven
                           3
                                0.2200 1738.7 -57981
## + area
                           5
                                0.4243 1738.5 -57980
## + engine_type
                           3
                                0.0295 1738.8 -57979
## + veh_color
                           8
                                0.3662 1738.5 -57973
                           5
                                3.1287 1742.0 -57953
## - agecat
## - trm_len
                           1
                                5.3180 1744.2 -57917
## - exposure
                           1 18.8121 1757.7 -57743
##
## Step: AIC=-57984.28
## numclaims ~ exposure + veh_body + agecat + driving_history_score +
##
       trm_len + high_education_ind
##
##
                           Df Sum of Sq
                                           RSS
                                                  AIC
```

```
## - driving_history_score 1 0.1465 1739.2 -57984
                                     1739.0 -57984
## <none>
## - high education ind 1 0.1594 1739.2 -57984
## + gender
                          1 0.1447 1738.9 -57984
## + time_of_week_driven 1 0.1061 1738.9 -57984
## + marital_status 1 0.0421 1739.0 -57983
## + veh value
                        1 0.0257 1739.0 -57983
                         1 0.0133 1739.0 -57982
## + max_power
## + credit_score
                        1 0.0071 1739.0 -57982
## + e_bill
                         1 0.0025 1739.0 -57982
## + veh_age
                         1 0.0005 1739.0 -57982
                       12
                              2.0682 1741.1 -57981
## - veh_body
                       3 0.2182 1738.8 -57981
## + time_driven
                        5 0.4034 1738.6 -57980
## + area
                       3 0.0296 1739.0 -57979
8 0.3669 1738.7 -57973
## + engine_type
## + veh_color
                        5 3.2078 1742.2 -57953
## - agecat
## - trm len
                        1 5.3144 1744.3 -57917
                        1 18.7694 1757.8 -57743
## - exposure
## Step: AIC=-57984.37
## numclaims ~ exposure + veh_body + agecat + trm_len + high_education_ind
##
                         Df Sum of Sq
                                        RSS
                                               AIC
## <none>
                                      1739.2 -57984
## - high_education_ind
                         1
                              0.1593 1739.3 -57984
## + driving_history_score 1
                              0.1465 1739.0 -57984
## + gender
                          1
                              0.1453 1739.0 -57984
## + time_of_week_driven 1 0.1051 1739.1 -57984
## + marital_status 1 0.0447 1739.1 -57983
## + veh_value
                         1 0.0243 1739.2 -57983
                        1 0.0135 1739.2 -57983
## + max_power
## + credit_score
                        1 0.0074 1739.2 -57982
                        1 0.0027 1739.2 -57982
## + e_bill
## + veh age
                         1
                              0.0004 1739.2 -57982
                      12 2.0602 1741.2 -57982
3 0.2197 1739.0 -57981
## - veh_body
## + time driven
## + area
                        5 0.4039 1738.8 -57980
                      3 0.0299 1739.1 -57979
## + engine_type
                        8 0.3640 1738.8 -57973
## + veh_color
## - agecat
                        5 3.2039 1742.4 -57953
## - trm len
                              5.3123 1744.5 -57917
                          1
## - exposure
                          1 18.7551 1757.9 -57744
formula(final_model.backward.numclaims)
## numclaims ~ exposure + veh_body + agecat + trm_len + high_education_ind
AIC(final_model.backward.numclaims)
## [1] 6207.571
BIC(final_model.backward.numclaims)
```

[1] 6384.155

#numclaims ~ exposure + veh_body + agecat + trm_len + high_education_ind