

GROUP 2 HW4: Insurance - Data 621 Assignment 4

GROUP 2 MEMBERS: Banu Boopalan, Gregg Maloy, Alexander Moyse, Umais Siddiqui

11-12-2024

Contents

Assignment Introduction	1
Data Exploration	1

Assignment Introduction

We have read in the 2 datasets provided as part of this project from Github.

Our goal for this project is to analyze 8000 records of a customer of an insurance company.

Once have fit the data, we can then run and test on the evaluation dataset provided, we will analyze model validity and performance and report model that can be selected as the best fit for the data. Shown below is an image all the variables and their definitions.

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

Data Exploration

In this section we will explore the data and look at the number of columns , column data types and gain a better understanding of what we are dealing with . First lets go with the basics.

Explore DataSet & Review Missing Data

```
## 'data.frame':   8161 obs. of  26 variables:
## $ INDEX       : int  1 2 4 5 6 7 8 11 12 13 ...
## $ TARGET_FLAG: int  0 0 0 0 0 1 0 1 1 0 ...
## $ TARGET_AMT  : num  0 0 0 0 0 ...
## $ KIDSDRIV    : int  0 0 0 0 0 0 0 1 0 0 ...
## $ AGE         : int  60 43 35 51 50 34 54 37 34 50 ...
## $ HOMEKIDS    : int  0 0 1 0 0 1 0 2 0 0 ...
## $ YOJ         : int  11 11 10 14 NA 12 NA NA 10 7 ...
## $ INCOME      : chr  "$67,349" "$91,449" "$16,039" "" ...
## $ PARENT1     : chr  "No" "No" "No" "No" ...
## $ HOME_VAL    : chr  "$0" "$257,252" "$124,191" "$306,251" ...
## $ MSTATUS     : chr  "z_No" "z_No" "Yes" "Yes" ...
```

```

## $ SEX      : chr  "M" "M" "z_F" "M" ...
## $ EDUCATION : chr  "PhD" "z_High School" "z_High School" "<High School" ...
## $ JOB       : chr  "Professional" "z_Blue Collar" "Clerical" "z_Blue Collar" ...
## $ TRAVTIME  : int   14 22 5 32 36 46 33 44 34 48 ...
## $ CAR_USE   : chr  "Private" "Commercial" "Private" "Private" ...
## $ BLUEBOOK : chr  "$14,230" "$14,940" "$4,010" "$15,440" ...
## $ TIF       : int   11 1 4 7 1 1 1 1 7 ...
## $ CAR_TYPE  : chr  "Minivan" "Minivan" "z_SUV" "Minivan" ...
## $ RED_CAR   : chr  "yes" "yes" "no" "yes" ...
## $ OLDCLAIM  : chr  "$4,461" "$0" "$38,690" "$0" ...
## $ CLM_FREQ  : int   2 0 2 0 2 0 0 1 0 0 ...
## $ REVOKED   : chr  "No" "No" "No" "No" ...
## $ MVR_PTS   : int   3 0 3 0 3 0 0 10 0 1 ...
## $ CAR_AGE   : int   18 1 10 6 17 7 1 7 1 17 ...
## $ URBANICITY : chr  "Highly Urban/ Urban" "Highly Urban/ Urban" "Highly Urban/ Urban" "Highly Urban/ Urban"

```

```
## [1] 2141 26
```

```

##      INDEX      TARGET_FLAG  TARGET_AMT      KIDSDRIV      AGE
## Min.      : 3      Mode:logical  Mode:logical  Min.      :0.0000  Min.      :17.00
## 1st Qu.: 2632  NA's:2141      NA's:2141      1st Qu.:0.0000  1st Qu.:39.00
## Median : 5224                        Median :0.0000  Median :45.00
## Mean      : 5150                        Mean      :0.1625  Mean      :45.02
## 3rd Qu.: 7669                        3rd Qu.:0.0000  3rd Qu.:51.00
## Max.      :10300                       Max.      :3.0000  Max.      :73.00
##                                     NA's      :1
##      HOMEKIDS      YOJ      INCOME      PARENT1
## Min.      :0.0000  Min.      : 0.00  Length:2141  Length:2141
## 1st Qu.:0.0000  1st Qu.: 9.00  Class :character  Class :character
## Median :0.0000  Median :11.00  Mode  :character  Mode  :character
## Mean      :0.7174  Mean      :10.38
## 3rd Qu.:1.0000  3rd Qu.:13.00
## Max.      :5.0000  Max.      :19.00
##                                     NA's      :94
##      HOME_VAL      MSTATUS      SEX      EDUCATION
## Length:2141      Length:2141      Length:2141      Length:2141
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##      JOB      TRAVTIME      CAR_USE      BLUEBOOK
## Length:2141      Min.      : 5.00  Length:2141      Length:2141
## Class :character  1st Qu.: 22.00  Class :character  Class :character
## Mode  :character  Median : 33.00  Mode  :character  Mode  :character
##                                     Mean      : 33.15
##                                     3rd Qu.: 43.00
##                                     Max.      :105.00
##
##      TIF      CAR_TYPE      RED_CAR      OLDCLAIM
## Min.      : 1.000  Length:2141      Length:2141      Length:2141
## 1st Qu.: 1.000  Class :character  Class :character  Class :character
## Median : 4.000  Mode  :character  Mode  :character  Mode  :character

```

```

## Mean      : 5.245
## 3rd Qu.: 7.000
## Max.      :25.000
##
##      CLM_FREQ      REVOKED      MVR_PTS      CAR_AGE
## Min.      :0.000  Length:2141  Min.      : 0.000  Min.      : 0.000
## 1st Qu.:0.000  Class :character  1st Qu.: 0.000  1st Qu.: 1.000
## Median :0.000  Mode  :character  Median : 1.000  Median : 8.000
## Mean      :0.809      Mean      : 1.766  Mean      : 8.183
## 3rd Qu.:2.000      3rd Qu.: 3.000  3rd Qu.:12.000
## Max.      :5.000      Max.      :12.000  Max.      :26.000
##                                     NA's      :129
##      URBANICITY
## Length:2141
## Class :character
## Mode  :character
##
##
##
##

```