Re-Admissions Analysis

Kathleen Sheehan & Abhishek Kanaparthi

November 8, 2017

The following document elucidates about the in-depth analysis of the Re-Admissions Data and tries to find out the most important factors that are contributing for the Re-admission and also for finding patterns in the data, that will help reduce the number of re-admits.

library(readr)  
readmitsV1 <- read\_csv("R:/People's Projects/Kathleen Sheehan Projects/Data/readmitsV1.csv")

Now as the data has been imported, we can go ahead and start working on the dataset with the basic filtering part of it.

optional = c(1,2,4,6,7,8,9,10,18,19,41)  
group1 = c(20,21,22,23,24,25,26,27,28,29,30,31,32,33)  
group2 = c(34,35,36,37,38,39)  
  
rad1= readmitsV1[optional] # Building the dataset with the most important columns only  
rad2= readmitsV1[group1]  
rad3= readmitsV1[group2]  
  
target = readmitsV1[3] # Storing the Target Variable  
  
# Making a new Data Frame with only the desired varirables  
rad.raw = cbind(target,rad1,rad2,rad3)

As the data has been designed now, the next step is to check the structure of the data.

str(rad.raw)

## 'data.frame': 62788 obs. of 32 variables:  
## $ READMIT : chr "NO" "NO" "NO" "NO" ...  
## $ DOMAIN : chr "P152C" "P152W" "P152W" "P152E" ...  
## $ FACILITY : chr "TMC Center" "CHH Center" "DSH Center" "GW Hospital" ...  
## $ HAS\_PCP : chr "NO" "YES" "YES" "YES" ...  
## $ DNR\_DISPLAY : chr "Full Resuscitation" "Do Not Resuscitate" "Full Resuscitation" "Full Resuscitation" ...  
## $ GENDER : chr "Male" "Male" "Male" "Male" ...  
## $ AGE : int 57 86 63 83 68 88 66 86 63 66 ...  
## $ ENC\_TYPE : chr "Inpatient" "Inpatient" "Inpatient" "Inpatient" ...  
## $ FIN\_CLASS : chr "Uninsured" "Medicare Acute Care" "Medicaid Managed Care" "Medicare Acute Care" ...  
## $ ESTIMATED\_LOS : chr "2 to 3 Nights" "2 to 3 Nights" "2 to 3 Nights" "2 to 3 Nights" ...  
## $ LOS : num 5.84 3.76 20.42 8.73 3.76 ...  
## $ DAYS\_FROM\_LAST\_ADMIT : num NA 16.2 40.7 49.2 31.2 ...  
## $ WOUND : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ ADL\_ASSIST\_NEED : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ COG\_IMPAIR : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ COND\_KNOWL\_DEF : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ MENTAL\_HLTH : chr "Yes" "Yes" "No" "Yes" ...  
## $ END\_STG\_CDTN : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ HIGH\_RSK\_MED : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ HOSP\_LAST\_60\_DAYS : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ INCONTINENCE : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ LIVES\_ALONE : chr "Yes" "No" "Yes" "No" ...  
## $ MED\_KNOWL\_DEF : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ MULTI\_DX : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ O2\_ASSIST : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ POLYPHARM : chr "Yes" "Yes" "Yes" "Yes" ...  
## $ MEDS\_COUNT : int 1 2 16 3 9 NA 2 4 10 2 ...  
## $ HAS\_ANTICOAGULANT : chr "NO" "NO" "NO" "NO" ...  
## $ HAS\_ANTICONVULSANT : chr "NO" "NO" "NO" "NO" ...  
## $ HAS\_INSULIN : chr "NO" "NO" "NO" "NO" ...  
## $ HAS\_LOOP\_DIURETIC : chr "NO" "NO" "NO" "NO" ...  
## $ HAS\_NARCOTIC\_ANALGESIC: chr "NO" "NO" "NO" "NO" ...

summary(rad.raw)

## READMIT DOMAIN FACILITY   
## Length:62788 Length:62788 Length:62788   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## HAS\_PCP DNR\_DISPLAY GENDER AGE   
## Length:62788 Length:62788 Length:62788 Min. : 0.00   
## Class :character Class :character Class :character 1st Qu.: 38.00   
## Mode :character Mode :character Mode :character Median : 58.00   
## Mean : 54.66   
## 3rd Qu.: 72.00   
## Max. :117.00   
##   
## ENC\_TYPE FIN\_CLASS ESTIMATED\_LOS   
## Length:62788 Length:62788 Length:62788   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## LOS DAYS\_FROM\_LAST\_ADMIT WOUND   
## Min. : 0.000 Min. :-52.87 Length:62788   
## 1st Qu.: 1.270 1st Qu.: 5.85 Class :character   
## Median : 2.340 Median : 23.55 Mode :character   
## Mean : 3.859 Mean : 41.44   
## 3rd Qu.: 4.280 3rd Qu.: 62.66   
## Max. :221.170 Max. :179.96   
## NA's :32612   
## ADL\_ASSIST\_NEED COG\_IMPAIR COND\_KNOWL\_DEF   
## Length:62788 Length:62788 Length:62788   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## MENTAL\_HLTH END\_STG\_CDTN HIGH\_RSK\_MED   
## Length:62788 Length:62788 Length:62788   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## HOSP\_LAST\_60\_DAYS INCONTINENCE LIVES\_ALONE   
## Length:62788 Length:62788 Length:62788   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## MED\_KNOWL\_DEF MULTI\_DX O2\_ASSIST   
## Length:62788 Length:62788 Length:62788   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## POLYPHARM MEDS\_COUNT HAS\_ANTICOAGULANT HAS\_ANTICONVULSANT  
## Length:62788 Min. : 0.0 Length:62788 Length:62788   
## Class :character 1st Qu.: 2.0 Class :character Class :character   
## Mode :character Median : 3.0 Mode :character Mode :character   
## Mean : 4.6   
## 3rd Qu.: 6.0   
## Max. :40.0   
## NA's :15144   
## HAS\_INSULIN HAS\_LOOP\_DIURETIC HAS\_NARCOTIC\_ANALGESIC  
## Length:62788 Length:62788 Length:62788   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##

The next step is to convert the ones with the categorical data into factors.

library(tidyverse)

## Loading tidyverse: ggplot2  
## Loading tidyverse: tibble  
## Loading tidyverse: tidyr  
## Loading tidyverse: purrr  
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats  
## lag(): dplyr, stats

library(magrittr)

##   
## Attaching package: 'magrittr'

## The following object is masked from 'package:purrr':  
##   
## set\_names

## The following object is masked from 'package:tidyr':  
##   
## extract

library(dplyr)  
library(plyr)

## -------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

## -------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

## The following object is masked from 'package:purrr':  
##   
## compact

cols=c(1:6,8:10,13:26,28:32)  
rad.raw %<>% mutate\_at(cols, funs(factor(.)))

As we have changed the type of the variables now, we can go ahead and check the datatype now again.

str(rad.raw)

## 'data.frame': 62788 obs. of 32 variables:  
## $ READMIT : Factor w/ 2 levels "NO","YES": 1 1 1 1 2 1 2 2 1 1 ...  
## $ DOMAIN : Factor w/ 3 levels "P152C","P152E",..: 1 3 3 2 1 1 2 2 1 1 ...  
## $ FACILITY : Factor w/ 31 levels "AIK Centers",..: 27 2 7 11 16 16 17 17 19 19 ...  
## $ HAS\_PCP : Factor w/ 2 levels "NO","YES": 1 2 2 2 2 2 2 2 2 2 ...  
## $ DNR\_DISPLAY : Factor w/ 3 levels "Do Not Resuscitate",..: 3 1 3 3 3 1 NA 1 3 3 ...  
## $ GENDER : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 1 2 2 2 1 ...  
## $ AGE : int 57 86 63 83 68 88 66 86 63 66 ...  
## $ ENC\_TYPE : Factor w/ 2 levels "Inpatient","Observation": 1 1 1 1 1 1 1 1 1 1 ...  
## $ FIN\_CLASS : Factor w/ 35 levels "AETNA Insurance",..: 29 15 12 15 28 15 16 15 15 15 ...  
## $ ESTIMATED\_LOS : Factor w/ 5 levels "2 to 3 Nights",..: 1 1 1 1 5 3 5 1 1 1 ...  
## $ LOS : num 5.84 3.76 20.42 8.73 3.76 ...  
## $ DAYS\_FROM\_LAST\_ADMIT : num NA 16.2 40.7 49.2 31.2 ...  
## $ WOUND : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 1 2 ...  
## $ ADL\_ASSIST\_NEED : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ COG\_IMPAIR : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ COND\_KNOWL\_DEF : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ MENTAL\_HLTH : Factor w/ 2 levels "No","Yes": 2 2 1 2 2 1 1 2 2 2 ...  
## $ END\_STG\_CDTN : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ HIGH\_RSK\_MED : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ HOSP\_LAST\_60\_DAYS : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 1 ...  
## $ INCONTINENCE : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ LIVES\_ALONE : Factor w/ 2 levels "No","Yes": 2 1 2 1 1 2 2 1 2 2 ...  
## $ MED\_KNOWL\_DEF : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ MULTI\_DX : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ O2\_ASSIST : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ POLYPHARM : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ MEDS\_COUNT : int 1 2 16 3 9 NA 2 4 10 2 ...  
## $ HAS\_ANTICOAGULANT : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 NA 1 1 1 1 ...  
## $ HAS\_ANTICONVULSANT : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 NA 1 1 1 1 ...  
## $ HAS\_INSULIN : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 NA 1 1 1 1 ...  
## $ HAS\_LOOP\_DIURETIC : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 NA 1 1 1 1 ...  
## $ HAS\_NARCOTIC\_ANALGESIC: Factor w/ 2 levels "NO","YES": 1 1 1 1 1 NA 1 1 1 1 ...

summary(rad.raw)

## READMIT DOMAIN FACILITY HAS\_PCP   
## NO :48676 P152C:17463 SHM Center : 4830 NO :18816   
## YES:14112 P152E:14215 TMC Center : 4484 YES:43972   
## P152W:31110 GW Hospital: 4238   
## SVH Center : 4201   
## MMH Main : 3733   
## CHH Center : 3529   
## (Other) :37773   
## DNR\_DISPLAY GENDER   
## Do Not Resuscitate : 3779 Female:35636   
## Do Not Resuscitate/Comfort Measures ONLY: 3 Male :27152   
## Full Resuscitation :40769   
## NA's :18237   
##   
##   
##   
## AGE ENC\_TYPE   
## Min. : 0.00 Inpatient :49963   
## 1st Qu.: 38.00 Observation:12825   
## Median : 58.00   
## Mean : 54.66   
## 3rd Qu.: 72.00   
## Max. :117.00   
##   
## FIN\_CLASS   
## Medicare Acute Care :16349   
## Managed Care/HMO/PPO (inclds Blue Cross):14056   
## Medicaid Managed Care :10473   
## Medicare Managed Care :10160   
## Uninsured : 3159   
## (Other) : 8585   
## NA's : 6   
## ESTIMATED\_LOS LOS DAYS\_FROM\_LAST\_ADMIT  
## 2 to 3 Nights :29974 Min. : 0.000 Min. :-52.87   
## 3 to 4 Nights : 4885 1st Qu.: 1.270 1st Qu.: 5.85   
## 5 or More Nights : 1760 Median : 2.340 Median : 23.55   
## 5 to More Nights : 35 Mean : 3.859 Mean : 41.44   
## Less than 2 Midnights:26134 3rd Qu.: 4.280 3rd Qu.: 62.66   
## Max. :221.170 Max. :179.96   
## NA's :32612   
## WOUND ADL\_ASSIST\_NEED COG\_IMPAIR COND\_KNOWL\_DEF MENTAL\_HLTH   
## No :41222 No :32642 No :42185 No :38604 No :40632   
## Yes : 5492 Yes :14127 Yes : 4514 Yes : 8164 Yes : 6128   
## NA's:16074 NA's:16019 NA's:16089 NA's:16020 NA's:16028   
##   
##   
##   
##   
## END\_STG\_CDTN HIGH\_RSK\_MED HOSP\_LAST\_60\_DAYS INCONTINENCE LIVES\_ALONE   
## No :43310 No :38315 No :31807 No :40796 No :41913   
## Yes : 3399 Yes : 8454 Yes :14846 Yes : 5894 Yes : 4859   
## NA's:16079 NA's:16019 NA's:16135 NA's:16098 NA's:16016   
##   
##   
##   
##   
## MED\_KNOWL\_DEF MULTI\_DX O2\_ASSIST POLYPHARM MEDS\_COUNT   
## No :38079 No :39041 No :41661 No :25709 Min. : 0.0   
## Yes : 8689 Yes : 7627 Yes : 4924 Yes :21059 1st Qu.: 2.0   
## NA's:16020 NA's:16120 NA's:16203 NA's:16020 Median : 3.0   
## Mean : 4.6   
## 3rd Qu.: 6.0   
## Max. :40.0   
## NA's :15144   
## HAS\_ANTICOAGULANT HAS\_ANTICONVULSANT HAS\_INSULIN HAS\_LOOP\_DIURETIC  
## NO :40411 NO :46982 NO :46796 NO :46613   
## YES : 7233 YES : 662 YES : 848 YES : 1031   
## NA's:15144 NA's:15144 NA's:15144 NA's:15144   
##   
##   
##   
##   
## HAS\_NARCOTIC\_ANALGESIC  
## NO :46196   
## YES : 1448   
## NA's:15144   
##   
##   
##   
##

The next step is to work on using one group of variables alone to do the analysis and then the next.

We have to fix upon the most important variables before we jump into the logistic regression on the same.

library(devtools)

## Warning: package 'devtools' was built under R version 3.4.2

#install\_github("riv","tomasgreif")  
#install\_github("woe","tomasgreif")  
library(woe)  
library(riv)

## Warning: package 'riv' was built under R version 3.4.2

## Loading required package: MASS

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

## Loading required package: rrcov

## Warning: package 'rrcov' was built under R version 3.4.2

## Loading required package: robustbase

## Warning: package 'robustbase' was built under R version 3.4.2

## Scalable Robust Estimators with High Breakdown Point (version 1.4-3)

## Loading required package: quantreg

## Warning: package 'quantreg' was built under R version 3.4.2

## Loading required package: SparseM

##   
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':  
##   
## backsolve

iv.mult(rad.raw,"READMIT",summary = TRUE, verbose = TRUE)

## Started processing of data frame: rad.raw   
## Calling iv.str for variable: DOMAIN   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: FACILITY   
## Factor: Assuming bad = level 2 and good = level 1

## Warning in iv.str(df, x, y, verbose = verbose): Some group for outcome 0 has zero count. This will result in -Inf or Inf WOE. Replacing - ODDS=1, WoE=0, MIV=0.   
## The bin is either too small or suspiciously predictive.   
## You should fix this before running any model. It does not make any sense to keep WoE = 0 for such bin.

## Information Value 0.08   
## Calling iv.str for variable: HAS\_PCP   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.04   
## Calling iv.str for variable: DNR\_DISPLAY   
## Factor: Assuming bad = level 2 and good = level 1

## Warning in iv.str(df, x, y, verbose = verbose): Some group for outcome 0 has zero count. This will result in -Inf or Inf WOE. Replacing - ODDS=1, WoE=0, MIV=0.   
## The bin is either too small or suspiciously predictive.   
## You should fix this before running any model. It does not make any sense to keep WoE = 0 for such bin.

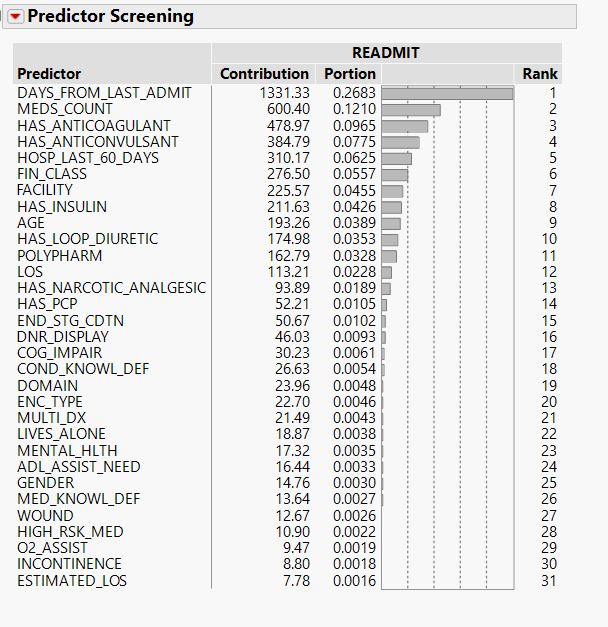
## Information Value 0.01   
## Calling iv.str for variable: GENDER   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.num for variable: AGE   
## Building rpart model  
## Model finished  
## Sending model to tree parser  
## Rules parsed: 3  
## Mapping nodes to data  
## SQL Merge  
## DF Merge  
## Calling iv.str for nodes  
## Information Value 0.05   
## Formatting output  
## Calling iv.str for variable: ENC\_TYPE   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0   
## Calling iv.str for variable: FIN\_CLASS   
## Factor: Assuming bad = level 2 and good = level 1

## Warning in iv.str(df, x, y, verbose = verbose): Some group for outcome 0 has zero count. This will result in -Inf or Inf WOE. Replacing - ODDS=1, WoE=0, MIV=0.   
## The bin is either too small or suspiciously predictive.   
## You should fix this before running any model. It does not make any sense to keep WoE = 0 for such bin.

## Information Value 0.11   
## Calling iv.str for variable: ESTIMATED\_LOS   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0   
## Calling iv.num for variable: LOS   
## Building rpart model  
## Model finished  
## Sending model to tree parser  
## Rules parsed: 3  
## Mapping nodes to data  
## SQL Merge  
## DF Merge  
## Calling iv.str for nodes  
## Information Value 0.04   
## Formatting output  
## Calling iv.num for variable: DAYS\_FROM\_LAST\_ADMIT   
## Building rpart model  
## Model finished  
## Sending model to tree parser  
## Rules parsed: 2  
## Mapping nodes to data  
## SQL Merge  
## DF Merge  
## Calling iv.str for nodes  
## Information Value 0.05   
## Formatting output  
## Calling iv.str for variable: WOUND   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: ADL\_ASSIST\_NEED   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: COG\_IMPAIR   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: COND\_KNOWL\_DEF   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: MENTAL\_HLTH   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: END\_STG\_CDTN   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.05   
## Calling iv.str for variable: HIGH\_RSK\_MED   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.02   
## Calling iv.str for variable: HOSP\_LAST\_60\_DAYS   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.16   
## Calling iv.str for variable: INCONTINENCE   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: LIVES\_ALONE   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: MED\_KNOWL\_DEF   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: MULTI\_DX   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.02   
## Calling iv.str for variable: O2\_ASSIST   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.01   
## Calling iv.str for variable: POLYPHARM   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.05   
## Calling iv.num for variable: MEDS\_COUNT   
## Building rpart model  
## Model finished  
## Sending model to tree parser  
## Rules parsed: 2  
## Mapping nodes to data  
## SQL Merge  
## DF Merge  
## Calling iv.str for nodes  
## Information Value 0.01   
## Formatting output  
## Calling iv.str for variable: HAS\_ANTICOAGULANT   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.08   
## Calling iv.str for variable: HAS\_ANTICONVULSANT   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.08   
## Calling iv.str for variable: HAS\_INSULIN   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.08   
## Calling iv.str for variable: HAS\_LOOP\_DIURETIC   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.08   
## Calling iv.str for variable: HAS\_NARCOTIC\_ANALGESIC   
## Factor: Assuming bad = level 2 and good = level 1   
## Information Value 0.08   
## Preparing summary

## Variable InformationValue Bins ZeroBins Strength  
## 1 HOSP\_LAST\_60\_DAYS 0.1601672203 2 0 Average  
## 2 FIN\_CLASS 0.1053206903 35 18 Average  
## 3 HAS\_ANTICOAGULANT 0.0833399188 2 0 Weak  
## 4 HAS\_NARCOTIC\_ANALGESIC 0.0816526809 2 0 Weak  
## 5 HAS\_LOOP\_DIURETIC 0.0807720985 2 0 Weak  
## 6 HAS\_ANTICONVULSANT 0.0806106983 2 0 Weak  
## 7 HAS\_INSULIN 0.0805543698 2 0 Weak  
## 8 FACILITY 0.0790552636 31 1 Weak  
## 9 POLYPHARM 0.0542811087 2 0 Weak  
## 10 END\_STG\_CDTN 0.0516321951 2 0 Weak  
## 11 AGE 0.0493911436 3 0 Weak  
## 12 DAYS\_FROM\_LAST\_ADMIT 0.0458326704 2 0 Weak  
## 13 HAS\_PCP 0.0421668135 2 0 Weak  
## 14 LOS 0.0378947539 3 0 Weak  
## 15 MULTI\_DX 0.0223880685 2 0 Weak  
## 16 HIGH\_RSK\_MED 0.0153138075 2 0 Wery weak  
## 17 MENTAL\_HLTH 0.0145048737 2 0 Wery weak  
## 18 GENDER 0.0140434623 2 0 Wery weak  
## 19 O2\_ASSIST 0.0138943309 2 0 Wery weak  
## 20 MEDS\_COUNT 0.0113556283 2 0 Wery weak  
## 21 DOMAIN 0.0106646617 3 0 Wery weak  
## 22 ADL\_ASSIST\_NEED 0.0103309465 2 0 Wery weak  
## 23 LIVES\_ALONE 0.0101365192 2 0 Wery weak  
## 24 MED\_KNOWL\_DEF 0.0096264720 2 0 Wery weak  
## 25 COND\_KNOWL\_DEF 0.0095275557 2 0 Wery weak  
## 26 WOUND 0.0088912080 2 0 Wery weak  
## 27 DNR\_DISPLAY 0.0088875317 3 1 Wery weak  
## 28 INCONTINENCE 0.0087392296 2 0 Wery weak  
## 29 COG\_IMPAIR 0.0082400165 2 0 Wery weak  
## 30 ENC\_TYPE 0.0030821354 2 0 Wery weak  
## 31 ESTIMATED\_LOS 0.0006088338 5 0 Wery weak

You can also find the relative significance of the variables from predictor screening in JMP.



Analysis for just the Rad.raw

The further analysis can be done with the usage of the important variables or can also be done with the help of some of the groups of variables to help understand the importance of the variables in a step by step manner, rather than in a whole manner.

The following predictors are choosen just to check a model, which can be developed with the variables with the least number of missing values, just to develop the simplest prototype of the model in case we finalise the most important predictors after the screening analysis.

predictors = c(1, 2:4, 6:11)  
logit1 = glm(data = rad.raw[predictors], rad.raw$READMIT~., family = binomial(link = "logit"))  
summary(logit1)

##   
## Call:  
## glm(formula = rad.raw$READMIT ~ ., family = binomial(link = "logit"),   
## data = rad.raw[predictors])  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5654 -0.7527 -0.6369 -0.4187 2.7623   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error  
## (Intercept) -2.837e+00 2.703e+02  
## DOMAINP152E 7.430e-02 2.565e-01  
## DOMAINP152W 4.168e-01 2.540e-01  
## FACILITYCHH Center -1.690e-01 5.739e-02  
## FACILITYCRM Center -4.035e-01 7.270e-02  
## FACILITYCRMB-Magnolia 3.048e-01 1.257e+00  
## FACILITYCSR-Cornerstone -1.102e+01 1.592e+02  
## FACILITYDHL Hospital -4.254e-01 2.608e-01  
## FACILITYDSH Center -1.383e-01 5.748e-02  
## FACILITYDSHB Psych -1.250e+01 3.782e+02  
## FACILITYERM Center -9.901e-02 2.576e-01  
## FACILITYFDR Center -3.056e-03 2.644e-01  
## FACILITYGW Hospital -1.431e-01 6.837e-02  
## FACILITYHHN Center -1.107e-01 6.676e-02  
## FACILITYIVM Center -3.728e-01 7.091e-02  
## FACILITYLWR Main 2.701e-01 8.442e-02  
## FACILITYMHH Heart 1.246e-01 2.657e-01  
## FACILITYMMC Center 6.509e-01 2.543e-01  
## FACILITYMMH Main 1.831e-02 6.683e-02  
## FACILITYNNM Center -5.603e-01 9.152e-02  
## FACILITYNWT Hospital -1.112e-01 2.553e-01  
## FACILITYNWTB Psych 6.728e-01 5.856e-01  
## FACILITYPRM Center -3.344e-01 7.155e-02  
## FACILITYRSM Center -4.492e-01 7.313e-02  
## FACILITYSHM Center -2.721e-01 5.399e-02  
## FACILITYSTB Center 6.082e-01 2.542e-01  
## FACILITYSTM Center -1.893e-02 2.643e-01  
## FACILITYSVH Center -3.109e-01 5.540e-02  
## FACILITYTMC Center 4.914e-01 2.536e-01  
## FACILITYTMCB- Psych NA NA  
## FACILITYTVH Hospital -3.576e-01 6.830e-02  
## FACILITYVHM Center NA NA  
## FACILITYWRM Center -1.140e-01 7.232e-02  
## HAS\_PCPYES 5.604e-01 2.453e-02  
## GENDERMale 2.082e-01 1.996e-02  
## AGE 4.273e-03 6.649e-04  
## ENC\_TYPEObservation 1.954e-01 2.860e-02  
## FIN\_CLASSAETNA MCR -3.675e-01 5.783e+02  
## FIN\_CLASSATTORN.ACCOUNT -1.982e-01 2.813e+02  
## FIN\_CLASSBCTX MCR 1.116e-02 4.371e+02  
## FIN\_CLASSBlue Cross of Texas -1.770e-01 2.887e+02  
## FIN\_CLASSCommercial Insurance -4.988e-01 2.703e+02  
## FIN\_CLASSHumana MCR Manage Care -3.199e-01 3.078e+02  
## FIN\_CLASSIndigent 1.281e+00 2.703e+02  
## FIN\_CLASSLiability 1.930e-01 2.703e+02  
## FIN\_CLASSManaged Care/HMO/PPO (inclds Blue Cross) 1.866e-01 2.703e+02  
## FIN\_CLASSMedicaid 1.218e+00 2.703e+02  
## FIN\_CLASSMedicaid Managed Care 8.639e-01 2.703e+02  
## FIN\_CLASSMedicaid Pending 1.262e+00 2.703e+02  
## FIN\_CLASSMedicare 9.711e+00 2.185e+02  
## FIN\_CLASSMedicare Acute Care 6.351e-01 2.703e+02  
## FIN\_CLASSMedicare Managed Care 5.203e-01 2.703e+02  
## FIN\_CLASSMedicare Psych Inpatient 1.857e+00 2.703e+02  
## FIN\_CLASSMedicare Rehab Inpatient 8.167e-01 2.703e+02  
## FIN\_CLASSMolina Health Care MMP -1.561e-01 3.453e+02  
## FIN\_CLASSMolina Healthcare of Texas Inc -1.510e-01 3.778e+02  
## FIN\_CLASSOther -2.442e-01 4.371e+02  
## FIN\_CLASSOutreach Client Billing 7.268e-03 2.703e+02  
## FIN\_CLASSSelf Pay -4.063e-02 2.821e+02  
## FIN\_CLASSSpecial Program 6.011e-01 2.703e+02  
## FIN\_CLASSSuperior -1.332e-01 3.078e+02  
## FIN\_CLASSTexas Healthsprings -3.643e-02 3.784e+02  
## FIN\_CLASSTexas Healthsprings Star Plus -1.931e-01 2.889e+02  
## FIN\_CLASSTRICARE/Champus 3.284e-01 2.703e+02  
## FIN\_CLASSUninsured 1.089e+00 2.703e+02  
## FIN\_CLASSUnited Healthcare -1.522e-01 3.444e+02  
## FIN\_CLASSUnited Healthcare Medicare -1.018e-01 3.782e+02  
## FIN\_CLASSUnited Healthcare Texas Star -3.189e-01 3.763e+02  
## FIN\_CLASSVeterans Fee Services -1.675e-01 3.452e+02  
## FIN\_CLASSWELLMED -2.806e-01 2.517e+02  
## FIN\_CLASSWorkers Compensation 2.128e-01 2.703e+02  
## ESTIMATED\_LOS3 to 4 Nights 5.759e-03 3.836e-02  
## ESTIMATED\_LOS5 or More Nights -5.724e-02 6.697e-02  
## ESTIMATED\_LOS5 to More Nights -1.419e+00 6.094e-01  
## ESTIMATED\_LOSLess than 2 Midnights -5.323e-02 2.461e-02  
## LOS 9.310e-03 1.772e-03  
## z value Pr(>|z|)   
## (Intercept) -0.010 0.99163   
## DOMAINP152E 0.290 0.77206   
## DOMAINP152W 1.641 0.10074   
## FACILITYCHH Center -2.946 0.00322 \*\*   
## FACILITYCRM Center -5.551 2.85e-08 \*\*\*  
## FACILITYCRMB-Magnolia 0.242 0.80848   
## FACILITYCSR-Cornerstone -0.069 0.94480   
## FACILITYDHL Hospital -1.631 0.10289   
## FACILITYDSH Center -2.407 0.01610 \*   
## FACILITYDSHB Psych -0.033 0.97362   
## FACILITYERM Center -0.384 0.70074   
## FACILITYFDR Center -0.012 0.99078   
## FACILITYGW Hospital -2.093 0.03637 \*   
## FACILITYHHN Center -1.658 0.09727 .   
## FACILITYIVM Center -5.257 1.46e-07 \*\*\*  
## FACILITYLWR Main 3.200 0.00138 \*\*   
## FACILITYMHH Heart 0.469 0.63918   
## FACILITYMMC Center 2.560 0.01047 \*   
## FACILITYMMH Main 0.274 0.78416   
## FACILITYNNM Center -6.122 9.24e-10 \*\*\*  
## FACILITYNWT Hospital -0.435 0.66320   
## FACILITYNWTB Psych 1.149 0.25060   
## FACILITYPRM Center -4.674 2.96e-06 \*\*\*  
## FACILITYRSM Center -6.143 8.12e-10 \*\*\*  
## FACILITYSHM Center -5.040 4.66e-07 \*\*\*  
## FACILITYSTB Center 2.393 0.01671 \*   
## FACILITYSTM Center -0.072 0.94291   
## FACILITYSVH Center -5.613 1.99e-08 \*\*\*  
## FACILITYTMC Center 1.938 0.05261 .   
## FACILITYTMCB- Psych NA NA   
## FACILITYTVH Hospital -5.236 1.64e-07 \*\*\*  
## FACILITYVHM Center NA NA   
## FACILITYWRM Center -1.576 0.11503   
## HAS\_PCPYES 22.850 < 2e-16 \*\*\*  
## GENDERMale 10.433 < 2e-16 \*\*\*  
## AGE 6.426 1.31e-10 \*\*\*  
## ENC\_TYPEObservation 6.834 8.27e-12 \*\*\*  
## FIN\_CLASSAETNA MCR -0.001 0.99949   
## FIN\_CLASSATTORN.ACCOUNT -0.001 0.99944   
## FIN\_CLASSBCTX MCR 0.000 0.99998   
## FIN\_CLASSBlue Cross of Texas -0.001 0.99951   
## FIN\_CLASSCommercial Insurance -0.002 0.99853   
## FIN\_CLASSHumana MCR Manage Care -0.001 0.99917   
## FIN\_CLASSIndigent 0.005 0.99622   
## FIN\_CLASSLiability 0.001 0.99943   
## FIN\_CLASSManaged Care/HMO/PPO (inclds Blue Cross) 0.001 0.99945   
## FIN\_CLASSMedicaid 0.005 0.99640   
## FIN\_CLASSMedicaid Managed Care 0.003 0.99745   
## FIN\_CLASSMedicaid Pending 0.005 0.99627   
## FIN\_CLASSMedicare 0.044 0.96454   
## FIN\_CLASSMedicare Acute Care 0.002 0.99813   
## FIN\_CLASSMedicare Managed Care 0.002 0.99846   
## FIN\_CLASSMedicare Psych Inpatient 0.007 0.99452   
## FIN\_CLASSMedicare Rehab Inpatient 0.003 0.99759   
## FIN\_CLASSMolina Health Care MMP 0.000 0.99964   
## FIN\_CLASSMolina Healthcare of Texas Inc 0.000 0.99968   
## FIN\_CLASSOther -0.001 0.99955   
## FIN\_CLASSOutreach Client Billing 0.000 0.99998   
## FIN\_CLASSSelf Pay 0.000 0.99989   
## FIN\_CLASSSpecial Program 0.002 0.99823   
## FIN\_CLASSSuperior 0.000 0.99965   
## FIN\_CLASSTexas Healthsprings 0.000 0.99992   
## FIN\_CLASSTexas Healthsprings Star Plus -0.001 0.99947   
## FIN\_CLASSTRICARE/Champus 0.001 0.99903   
## FIN\_CLASSUninsured 0.004 0.99678   
## FIN\_CLASSUnited Healthcare 0.000 0.99965   
## FIN\_CLASSUnited Healthcare Medicare 0.000 0.99979   
## FIN\_CLASSUnited Healthcare Texas Star -0.001 0.99932   
## FIN\_CLASSVeterans Fee Services 0.000 0.99961   
## FIN\_CLASSWELLMED -0.001 0.99911   
## FIN\_CLASSWorkers Compensation 0.001 0.99937   
## ESTIMATED\_LOS3 to 4 Nights 0.150 0.88067   
## ESTIMATED\_LOS5 or More Nights -0.855 0.39270   
## ESTIMATED\_LOS5 to More Nights -2.328 0.01990 \*   
## ESTIMATED\_LOSLess than 2 Midnights -2.163 0.03052 \*   
## LOS 5.255 1.48e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 66912 on 62781 degrees of freedom  
## Residual deviance: 64368 on 62708 degrees of freedom  
## (6 observations deleted due to missingness)  
## AIC: 64516  
##   
## Number of Fisher Scoring iterations: 12

predictedProb = predict(logit1,type = "response")  
library(InformationValue)  
optCutOff <- optimalCutoff(rad.raw$READMIT, predictedProb)[1]   
optCutOff

## [1] 0.002792422

The next sep is to check the accuracy of the model.

target = ifelse(rad.raw$READMIT == "YES", 1, 0)  
print(paste("Mis-Class error is :",misClassError(target, predictedProb, threshold = optCutOff)))

## [1] "Mis-Class error is : 0.7736"

#plotROC(target, predictions)  
sensitivity(target, predictedProb, threshold = optCutOff)

## [1] 0.9996457

specificity(target, predictedProb, threshold = optCutOff)

## [1] 0.002198209

targetRef = rad.raw$READMIT[!is.na(rad.raw$FIN\_CLASS)]  
targetRefNum = ifelse(targetRef == "YES", 1, 0)  
targetPredicted = ifelse(predictedProb > optCutOff, 1, 0)  
  
library(caret)  
library(e1071)  
confusionMatrix(data = logit1$y, reference = targetRefNum, positive = "1")

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1  
## 0 48670 0  
## 1 0 14112  
##   
## Accuracy : 1   
## 95% CI : (0.9999, 1)  
## No Information Rate : 0.7752   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 1   
## Mcnemar's Test P-Value : NA   
##   
## Sensitivity : 1.0000   
## Specificity : 1.0000   
## Pos Pred Value : 1.0000   
## Neg Pred Value : 1.0000   
## Prevalence : 0.2248   
## Detection Rate : 0.2248   
## Detection Prevalence : 0.2248   
## Balanced Accuracy : 1.0000   
##   
## 'Positive' Class : 1   
##