Final Project DATA 621

Hector Santana, Mario Pena, Zachary Safir

12/4/2021

Abstract

Credit risk is essential to the modern financial world. It is the backbone for underwriting, loss mitigation, and the overall issuance of debt. Without strict monitoring, policies, and analytics, potential expected losses can be astronomical.

These same stringent policies and procedures can however be too narrow. Traditional models may exclude worthy applicants based on the type of data used. That is why alternative data is being used more heavily. It expands the criteria to potential applicants that would've otherwise been excluded from the process. How does one discern the best set of alternative variables? Here we work to solve this problem by analyzing credit risk data sets posted on Kaggle by the Home Credit Group.

The training data set provides 176 variables to choose from in order to discern which loans will face default issues and which will not. The ability to predict the possibility of default based on a subset of key variables will help underwriters issue secure loans to a wider applicant pool.

Key Words

- Credit Risk
- Predictive Modeling
- Loan Level Analysis
- Structured Credit
- Predictive Analytics

Introduction

Data science has revolutionized the modern world. The ability to extract useful and powerful insights from a data set has become a differentiator for top performers across a variety of industries. With that said, one industry that has lagged behind is finance. There has been a slow adoption of new technologies and techniques to analyze, dissect, and execute on data driven decisions. There is still a dependency on archaic processes, "intuition," and excel based models. One area in particular that has been resistant to change is credit. Loan level analytics, predictive models, and new technologies have only recently been applied to the industry. With the change in approach people have now begun to question whether traditional frameworks are the most operable in today's financial landscape.

That is where our group asked a fundamental question, are credit risk methodologies up to par with today's diverse applicant pool? The answer is no. Traditional models do not capture a large segment of the global population. People that do not have existing credit lines or traditional bank accounts are often excluded from the process even if they have proof of income or savings. That lead us to investigate existing problems and ultimately allowed us to find a Kaggle data set from Home Credit Group that permitted us to tackle this problem.

We set out to see if we could discern the best set of variables to predict whether an applicant would struggle with their loan payments or not. We studied relevant literature, modeling techniques, and read the data dictionary for the respective data set to decide on a best approach to solving this problem.

At the end of the day no one wants to loan money to someone who cannot meet their obligation and thus we set out to see if we can answer this question.

Literature Review

Since the 2008 financial crisis, credit risk management has been at the forefront of finance. From regulators to banks, everyone has been looking for ways to create robust and accurate credit models to prevent the next disaster. (Lu, Zhang, and Li 2019)

More recently financial services firms, including many FinTech companies, have turned their attention to the unbanked community. (Lu, Zhang, and Li 2019) This pool of applicants does not get captured in traditional processes and thus gets excluded from the applicant pool for products such as loans. There has been a heavy investment in developing robust alternative data sets with a wide range of characteristics to meet modeling standards. (Lu, Zhang, and Li 2019)

The fast paced change in the financial landscape brings and ever greater focus on this problem. Preventing model bias and being able to accurately determine whether applicants can meet their loan payments are essential to today's credit risk management process. (Lu, Zhang, and Li 2019)

Methodology

The training set for this data was extremely large. This required significant data exploration and preparation. Factor application to the data frame required deep analysis of the underlying variable values, and transformations that fit the data set.Several variables had ~50% and greater missing values thus requiring their removal from the analysis. Several other variables had ~20% or less missing values and were imputed using the median value. Transformations involved a combination of square root,log, and normalization functions based on the level of skewness in key numerical variables. A number of variables exhibited extreme skew and they needed to be changed in order for the models to interpret less erratic data.

Three models were created:

- A saturated binary logistic regression model
- A logistic regression model parsed for the highest correlating variables and using the stepAIC methodology for variable selection
- A Lasso Logistic Regression model utilizing variable penalization for optimal variable selection.

Experimentation and Results

Based on accuracy and AUC the optimal model was selected. The glmnet model was selected because the algorithm operates to achieve a sparse solution. Simplified models perform better overall and prevent over fitting. Accuracy was only marginally affected for this model versus the saturated logistic regression model. AUC was ~73% and only marginally different for the glmnet model versus the saturated model. The stepAIC model calculated all 0s and was therefore not applicable for this analysis.

Conclusion

A total of 23 variables of the 176 variables were significant to the glmnet model. These variables are alternative data points that can be considered for loan issuance. The credit risk process can be further enhanced with alternative data as traditional models are too narrow in scope.Using a robust modeling process for the size, complexity, and dimensionality of credit risk data is essential to obtaining interpretable results.

Reference the appendix for analytics and the variable list discerned by the model.

References

Roeder, Jan. ALTERNATIVE DATA FOR CREDIT RISK MANAGEMENT: AN ANALYSIS OF THE CURRENT STATE OF RESEARCH. University of Goettingen, Faculty of Business and Economics, Goettingen, Germany. Retrieved From: https://press.um.si/index.php/ump/catalog/view/581/744/1598-3

Friedman, Jerome; Hastie, Trevor; Tibshirani, Rob.(April 29th, 2009) Regularization Paths for Generalized Linear Models via Coordinate Descent. Department of Statistics, Stanford University. Retrieved from: https://hastie.su.domains/Papers/glmnet.pdf

Lu, Tian; Zhang, Yingjie; and Li, Beibei, "The Value of Alternative Data in Credit Risk Prediction: Evidence from a Large Field Experiment" (2019). ICIS 2019 Proceedings. 10. Retrieved From: https://core.ac.uk/download/pdf/301383651.pdf.

Appendix: R Code and Analytics

```r
# import libraries

library(RPostgres)
library(DBI)
library(dbplyr)
library(dplyr)
library(ggplot2)
library(imputeTS)
library(moments)
library(glmnet)
pacman::p_load(MASS,tidyverse,janitor,DataExplorer,knitr,arsenal,kableExtra,car,
               geoR,caret,
               psych,gridExtra,DMwR2,lmtest,pscl,MKmisc,ROCR,survey,stats,
               rstatix,Rcpp,
               corrplot,forecast,cowplot,gridExtra,arsenal,e1071,car)
```

```r
#factor application to training data and removal of problematic variables

for(i in colnames(training[sapply(training, is.numeric)] %>% select(-SK_ID_CURR))){
  if(length(unique(training[,i]))<=5){
    training[,i] = as.factor(training[,i])
  }
}
```

```r
training[sapply(training, is.character)] <- lapply(training[sapply(training,
                                                          is.character)],
                                       as.factor)
training$CNT_CHILDREN = as.factor(training$CNT_CHILDREN)

training = training %>% select(-c('AMT_REQ_CREDIT_BUREAU_DAY',
                                  'AMT_REQ_CREDIT_BUREAU_WEEK',
                                  'AMT_REQ_CREDIT_BUREAU_MON',
                                  'AMT_REQ_CREDIT_BUREAU_QRT',
                                  'AMT_REQ_CREDIT_BUREAU_HOUR',
                                  'AMT_REQ_CREDIT_BUREAU_YEAR' ))
```

```r
#structure of dataset

str(training)
```

```
## 'data.frame':    307511 obs. of  116 variables:
##  $ SK_ID_CURR                  : int  100002 100003 100004 100006 100007 100008 100009 100010 100011
##  $ TARGET                      : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 1 1 ...
##  $ NAME_CONTRACT_TYPE          : Factor w/ 2 levels "Cash loans","Revolving loans": 1 1 2 1 1 1 1 1
##  $ CODE_GENDER                 : Factor w/ 3 levels "F","M","XNA": 2 1 2 1 2 2 1 2 1 2 ...
##  $ FLAG_OWN_CAR                : Factor w/ 2 levels "N","Y": 1 1 2 1 1 1 2 2 1 1 ...
##  $ FLAG_OWN_REALTY             : Factor w/ 2 levels "N","Y": 2 1 2 2 2 2 2 2 2 2 ...
##  $ CNT_CHILDREN                : Factor w/ 15 levels "0","1","2","3",..: 1 1 1 1 1 1 2 1 1 1 ...
##  $ AMT_INCOME_TOTAL            : num  202500 270000 67500 135000 121500 ...
##  $ AMT_CREDIT                  : num  406598 1293503 135000 312683 513000 ...
##  $ AMT_ANNUITY                 : num  24701 35699 6750 29687 21866 ...
##  $ AMT_GOODS_PRICE             : num  351000 1129500 135000 297000 513000 ...
##  $ NAME_TYPE_SUITE             : Factor w/ 8 levels "","Children",..: 8 3 8 8 8 7 8 8 2 8 ...
##  $ NAME_INCOME_TYPE            : Factor w/ 8 levels "Businessman",..: 8 5 8 8 8 5 2 5 4 8 ...
##  $ NAME_EDUCATION_TYPE         : Factor w/ 5 levels "Academic degree",..: 5 2 5 5 5 5 2 2 5 5 ...
##  $ NAME_FAMILY_STATUS          : Factor w/ 6 levels "Civil marriage",..: 4 2 4 1 4 2 2 2 2 4 ...
##  $ NAME_HOUSING_TYPE           : Factor w/ 6 levels "Co-op apartment",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ REGION_POPULATION_RELATIVE  : num  0.0188 0.00354 0.01003 0.00802 0.02866 ...
##  $ DAYS_BIRTH                  : int  -9461 -16765 -19046 -19005 -19932 -16941 -13778 -18850 -20099
##  $ DAYS_EMPLOYED               : int  -637 -1188 -225 -3039 -3038 -1588 -3130 -449 365243 -2019 ...
##  $ DAYS_REGISTRATION           : num  -3648 -1186 -4260 -9833 -4311 ...
##  $ DAYS_ID_PUBLISH             : int  -2120 -291 -2531 -2437 -3458 -477 -619 -2379 -3514 -3992 ...
##  $ OWN_CAR_AGE                 : num  NA NA 26 NA NA NA 17 8 NA NA ...
##  $ FLAG_MOBIL                  : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ FLAG_EMP_PHONE              : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 1 2 ...
##  $ FLAG_WORK_PHONE             : Factor w/ 2 levels "0","1": 1 1 2 1 1 2 1 2 1 1 ...
##  $ FLAG_CONT_MOBILE            : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ FLAG_PHONE                  : Factor w/ 2 levels "0","1": 2 2 2 1 1 2 2 1 1 1 ...
##  $ FLAG_EMAIL                  : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ OCCUPATION_TYPE             : Factor w/ 19 levels "","Accountants",..: 10 5 10 10 5 10 2 12 1 10
##  $ CNT_FAM_MEMBERS             : num  1 2 1 2 1 2 3 2 2 1 ...
##  $ REGION_RATING_CLIENT        : Factor w/ 3 levels "1","2","3": 2 1 2 2 2 2 2 3 2 2 ...
##  $ REGION_RATING_CLIENT_W_CITY : Factor w/ 3 levels "1","2","3": 2 1 2 2 2 2 2 3 2 2 ...
##  $ WEEKDAY_APPR_PROCESS_START  : Factor w/ 7 levels "FRIDAY","MONDAY",..: 7 2 2 7 5 7 4 2 7 5 ...
##  $ HOUR_APPR_PROCESS_START     : int  10 11 9 17 11 16 16 16 14 8 ...
##  $ REG_REGION_NOT_LIVE_REGION  : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ REG_REGION_NOT_WORK_REGION  : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

```
##  $ LIVE_REGION_NOT_WORK_REGION : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ REG_CITY_NOT_LIVE_CITY      : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ REG_CITY_NOT_WORK_CITY      : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 2 1 1 ...
##  $ LIVE_CITY_NOT_WORK_CITY     : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 2 1 1 ...
##  $ ORGANIZATION_TYPE           : Factor w/ 58 levels "Advertising",..: 6 40 12 6 38 34 6 34 58 10 ..
##  $ EXT_SOURCE_1                : num  0.083 0.311 NA NA NA ...
##  $ EXT_SOURCE_2                : num  0.263 0.622 0.556 0.65 0.323 ...
##  $ EXT_SOURCE_3                : num  0.139 NA 0.73 NA NA ...
##  $ APARTMENTS_AVG              : num  0.0247 0.0959 NA NA NA NA NA NA NA NA ...
##  $ BASEMENTAREA_AVG            : num  0.0369 0.0529 NA NA NA NA NA NA NA NA ...
##  $ YEARS_BEGINEXPLUATATION_AVG : num  0.972 0.985 NA NA NA ...
##  $ YEARS_BUILD_AVG             : num  0.619 0.796 NA NA NA ...
##  $ COMMONAREA_AVG              : num  0.0143 0.0605 NA NA NA NA NA NA NA NA ...
##  $ ELEVATORS_AVG               : num  0 0.08 NA NA NA NA NA NA NA NA ...
##  $ ENTRANCES_AVG               : num  0.069 0.0345 NA NA NA NA NA NA NA NA ...
##  $ FLOORSMAX_AVG               : num  0.0833 0.2917 NA NA NA ...
##  $ FLOORSMIN_AVG               : num  0.125 0.333 NA NA NA ...
##  $ LANDAREA_AVG                : num  0.0369 0.013 NA NA NA NA NA NA NA NA ...
##  $ LIVINGAPARTMENTS_AVG        : num  0.0202 0.0773 NA NA NA NA NA NA NA NA ...
##  $ LIVINGAREA_AVG              : num  0.019 0.0549 NA NA NA NA NA NA NA NA ...
##  $ NONLIVINGAPARTMENTS_AVG     : num  0 0.0039 NA NA NA NA NA NA NA NA ...
##  $ NONLIVINGAREA_AVG           : num  0 0.0098 NA NA NA NA NA NA NA NA ...
##  $ APARTMENTS_MODE             : num  0.0252 0.0924 NA NA NA NA NA NA NA NA ...
##  $ BASEMENTAREA_MODE           : num  0.0383 0.0538 NA NA NA NA NA NA NA NA ...
##  $ YEARS_BEGINEXPLUATATION_MODE: num  0.972 0.985 NA NA NA ...
##  $ YEARS_BUILD_MODE            : num  0.634 0.804 NA NA NA ...
##  $ COMMONAREA_MODE             : num  0.0144 0.0497 NA NA NA NA NA NA NA NA ...
##  $ ELEVATORS_MODE              : num  0 0.0806 NA NA NA NA NA NA NA NA ...
##  $ ENTRANCES_MODE              : num  0.069 0.0345 NA NA NA NA NA NA NA NA ...
##  $ FLOORSMAX_MODE              : num  0.0833 0.2917 NA NA NA ...
##  $ FLOORSMIN_MODE              : num  0.125 0.333 NA NA NA ...
##  $ LANDAREA_MODE               : num  0.0377 0.0128 NA NA NA NA NA NA NA NA ...
##  $ LIVINGAPARTMENTS_MODE       : num  0.022 0.079 NA NA NA NA NA NA NA NA ...
##  $ LIVINGAREA_MODE             : num  0.0198 0.0554 NA NA NA NA NA NA NA NA ...
##  $ NONLIVINGAPARTMENTS_MODE    : num  0 0 NA NA NA NA NA NA NA NA ...
##  $ NONLIVINGAREA_MODE          : num  0 0 NA NA NA NA NA NA NA NA ...
##  $ APARTMENTS_MEDI             : num  0.025 0.0968 NA NA NA NA NA NA NA NA ...
##  $ BASEMENTAREA_MEDI           : num  0.0369 0.0529 NA NA NA NA NA NA NA NA ...
##  $ YEARS_BEGINEXPLUATATION_MEDI: num  0.972 0.985 NA NA NA ...
##  $ YEARS_BUILD_MEDI            : num  0.624 0.799 NA NA NA ...
##  $ COMMONAREA_MEDI             : num  0.0144 0.0608 NA NA NA NA NA NA NA NA ...
##  $ ELEVATORS_MEDI              : num  0 0.08 NA NA NA NA NA NA NA NA ...
##  $ ENTRANCES_MEDI              : num  0.069 0.0345 NA NA NA NA NA NA NA NA ...
##  $ FLOORSMAX_MEDI              : num  0.0833 0.2917 NA NA NA ...
##  $ FLOORSMIN_MEDI              : num  0.125 0.333 NA NA NA ...
##  $ LANDAREA_MEDI               : num  0.0375 0.0132 NA NA NA NA NA NA NA NA ...
##  $ LIVINGAPARTMENTS_MEDI       : num  0.0205 0.0787 NA NA NA NA NA NA NA NA ...
##  $ LIVINGAREA_MEDI             : num  0.0193 0.0558 NA NA NA NA NA NA NA NA ...
##  $ NONLIVINGAPARTMENTS_MEDI    : num  0 0.0039 NA NA NA NA NA NA NA NA ...
##  $ NONLIVINGAREA_MEDI          : num  0 0.01 NA NA NA NA NA NA NA NA ...
##  $ FONDKAPREMONT_MODE          : Factor w/ 5 levels "","not specified",..: 4 4 1 1 1 1 1 1 1 1 ...
##  $ HOUSETYPE_MODE              : Factor w/ 4 levels "","block of flats",..: 2 2 1 1 1 1 1 1 1 1 ...
##  $ TOTALAREA_MODE              : num  0.0149 0.0714 NA NA NA NA NA NA NA NA ...
##  $ WALLSMATERIAL_MODE          : Factor w/ 8 levels "","Block","Mixed",..: 7 2 1 1 1 1 1 1 1 1 ...
```

```
##  $ EMERGENCYSTATE_MODE          : Factor w/ 3 levels "","No","Yes": 2 2 1 1 1 1 1 1 1 1 ...
##  $ OBS_30_CNT_SOCIAL_CIRCLE     : num  2 1 0 2 0 0 1 2 1 2 ...
##  $ DEF_30_CNT_SOCIAL_CIRCLE     : num  2 0 0 0 0 0 0 0 0 0 ...
##  $ OBS_60_CNT_SOCIAL_CIRCLE     : num  2 1 0 2 0 0 1 2 1 2 ...
##  $ DEF_60_CNT_SOCIAL_CIRCLE     : num  2 0 0 0 0 0 0 0 0 0 ...
##  $ DAYS_LAST_PHONE_CHANGE       : num  -1134 -828 -815 -617 -1106 ...
##  $ FLAG_DOCUMENT_2              : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_3              : Factor w/ 2 levels "0","1": 2 2 1 2 1 2 1 2 2 1 ...
##  $ FLAG_DOCUMENT_4              : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##   [list output truncated]
```

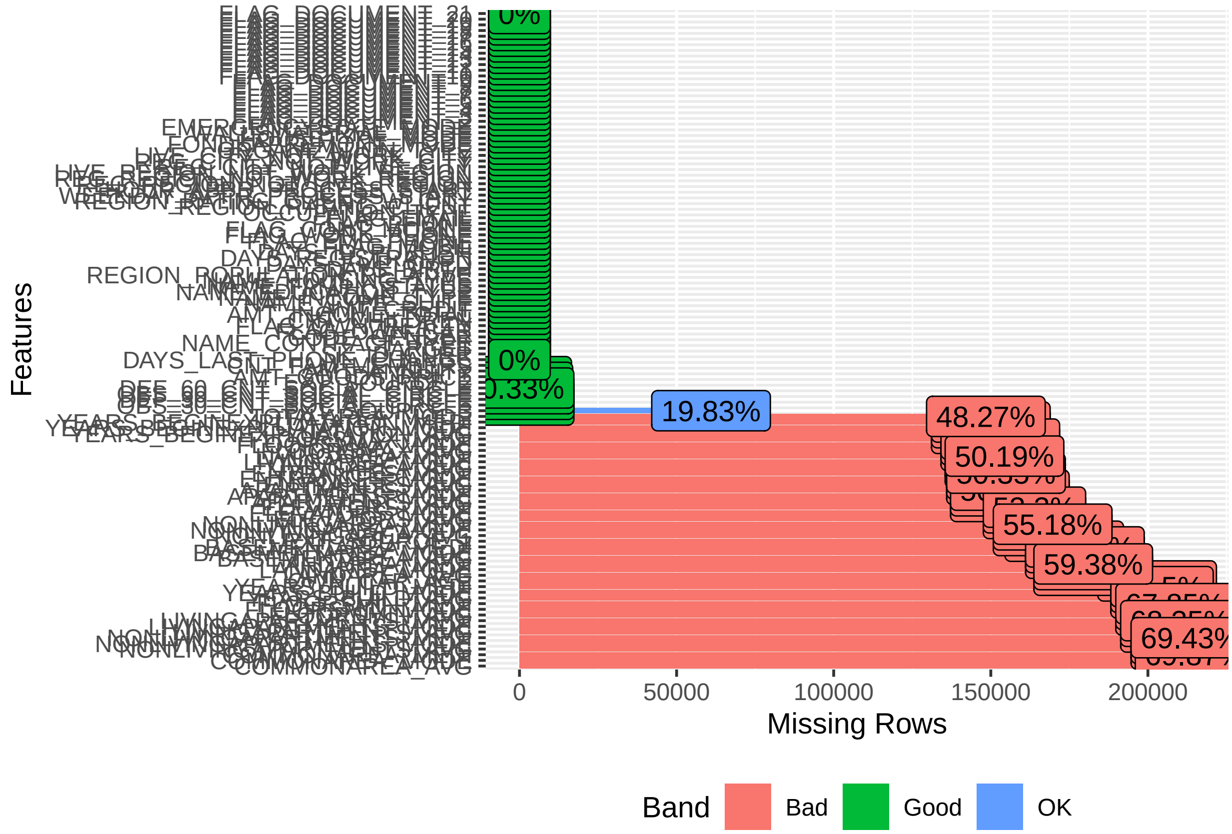
```
#descriptive statistics of dataset

des = describe((training %>% select(where(is.numeric))))
des
```

```
##                              vars      n      mean        sd       min
## SK_ID_CURR                      1 307511 278180.52 102790.18 100002.00
## AMT_INCOME_TOTAL                2 307511 168797.92 237123.15  25650.00
## AMT_CREDIT                      3 307511 599026.00 402490.78  45000.00
## AMT_ANNUITY                     4 307499  27108.57  14493.74   1615.50
## AMT_GOODS_PRICE                 5 307233 538396.21 369446.46  40500.00
## REGION_POPULATION_RELATIVE      6 307511      0.02      0.01      0.00
## DAYS_BIRTH                      7 307511 -16037.00   4363.99 -25229.00
## DAYS_EMPLOYED                   8 307511  63815.05 141275.77 -17912.00
## DAYS_REGISTRATION               9 307511  -4986.12   3522.89 -24672.00
## DAYS_ID_PUBLISH                10 307511  -2994.20   1509.45  -7197.00
## OWN_CAR_AGE                    11 104582     12.06     11.94      0.00
## CNT_FAM_MEMBERS                12 307509      2.15      0.91      1.00
## HOUR_APPR_PROCESS_START        13 307511     12.06      3.27      0.00
## EXT_SOURCE_1                   14 134133      0.50      0.21      0.01
## EXT_SOURCE_2                   15 306851      0.51      0.19      0.00
## EXT_SOURCE_3                   16 246546      0.51      0.19      0.00
## APARTMENTS_AVG                 17 151450      0.12      0.11      0.00
## BASEMENTAREA_AVG               18 127568      0.09      0.08      0.00
## YEARS_BEGINEXPLUATATION_AVG    19 157504      0.98      0.06      0.00
## YEARS_BUILD_AVG                20 103023      0.75      0.11      0.00
## COMMONAREA_AVG                 21  92646      0.04      0.08      0.00
## ELEVATORS_AVG                  22 143620      0.08      0.13      0.00
## ENTRANCES_AVG                  23 152683      0.15      0.10      0.00
## FLOORSMAX_AVG                  24 154491      0.23      0.14      0.00
## FLOORSMIN_AVG                  25  98869      0.23      0.16      0.00
## LANDAREA_AVG                   26 124921      0.07      0.08      0.00
## LIVINGAPARTMENTS_AVG           27  97312      0.10      0.09      0.00
## LIVINGAREA_AVG                 28 153161      0.11      0.11      0.00
## NONLIVINGAPARTMENTS_AVG        29  93997      0.01      0.05      0.00
## NONLIVINGAREA_AVG              30 137829      0.03      0.07      0.00
## APARTMENTS_MODE                31 151450      0.11      0.11      0.00
## BASEMENTAREA_MODE              32 127568      0.09      0.08      0.00
## YEARS_BEGINEXPLUATATION_MODE   33 157504      0.98      0.06      0.00
## YEARS_BUILD_MODE               34 103023      0.76      0.11      0.00
## COMMONAREA_MODE                35  92646      0.04      0.07      0.00
## ELEVATORS_MODE                 36 143620      0.07      0.13      0.00
## ENTRANCES_MODE                 37 152683      0.15      0.10      0.00
```

```
## FLOORSMAX_MODE               38 154491      0.22      0.14      0.00
## FLOORSMIN_MODE               39  98869      0.23      0.16      0.00
## LANDAREA_MODE                40 124921      0.06      0.08      0.00
## LIVINGAPARTMENTS_MODE        41  97312      0.11      0.10      0.00
## LIVINGAREA_MODE              42 153161      0.11      0.11      0.00
## NONLIVINGAPARTMENTS_MODE     43  93997      0.01      0.05      0.00
## NONLIVINGAREA_MODE           44 137829      0.03      0.07      0.00
## APARTMENTS_MEDI              45 151450      0.12      0.11      0.00
## BASEMENTAREA_MEDI            46 127568      0.09      0.08      0.00
## YEARS_BEGINEXPLUATATION_MEDI 47 157504      0.98      0.06      0.00
## YEARS_BUILD_MEDI             48 103023      0.76      0.11      0.00
## COMMONAREA_MEDI              49  92646      0.04      0.08      0.00
## ELEVATORS_MEDI               50 143620      0.08      0.13      0.00
## ENTRANCES_MEDI               51 152683      0.15      0.10      0.00
## FLOORSMAX_MEDI               52 154491      0.23      0.15      0.00
## FLOORSMIN_MEDI               53  98869      0.23      0.16      0.00
## LANDAREA_MEDI                54 124921      0.07      0.08      0.00
## LIVINGAPARTMENTS_MEDI        55  97312      0.10      0.09      0.00
## LIVINGAREA_MEDI              56 153161      0.11      0.11      0.00
## NONLIVINGAPARTMENTS_MEDI     57  93997      0.01      0.05      0.00
## NONLIVINGAREA_MEDI           58 137829      0.03      0.07      0.00
## TOTALAREA_MODE               59 159080      0.10      0.11      0.00
## OBS_30_CNT_SOCIAL_CIRCLE     60 306490      1.42      2.40      0.00
## DEF_30_CNT_SOCIAL_CIRCLE     61 306490      0.14      0.45      0.00
## OBS_60_CNT_SOCIAL_CIRCLE     62 306490      1.41      2.38      0.00
## DEF_60_CNT_SOCIAL_CIRCLE     63 306490      0.10      0.36      0.00
## DAYS_LAST_PHONE_CHANGE       64 307510   -962.86    826.81  -4292.00
##                                      max        range     se
## SK_ID_CURR                     456255.00    356253.00 185.36
## AMT_INCOME_TOTAL            117000000.00 116974350.00 427.61
## AMT_CREDIT                    4050000.00   4005000.00 725.81
## AMT_ANNUITY                    258025.50    256410.00  26.14
## AMT_GOODS_PRICE               4050000.00   4009500.00 666.53
## REGION_POPULATION_RELATIVE          0.07         0.07   0.00
## DAYS_BIRTH                      -7489.00     17740.00   7.87
## DAYS_EMPLOYED                  365243.00    383155.00 254.76
## DAYS_REGISTRATION                   0.00     24672.00   6.35
## DAYS_ID_PUBLISH                     0.00      7197.00   2.72
## OWN_CAR_AGE                        91.00        91.00   0.04
## CNT_FAM_MEMBERS                    20.00        19.00   0.00
## HOUR_APPR_PROCESS_START            23.00        23.00   0.01
## EXT_SOURCE_1                        0.96         0.95   0.00
## EXT_SOURCE_2                        0.85         0.85   0.00
## EXT_SOURCE_3                        0.90         0.90   0.00
## APARTMENTS_AVG                      1.00         1.00   0.00
## BASEMENTAREA_AVG                    1.00         1.00   0.00
## YEARS_BEGINEXPLUATATION_AVG         1.00         1.00   0.00
## YEARS_BUILD_AVG                     1.00         1.00   0.00
## COMMONAREA_AVG                      1.00         1.00   0.00
## ELEVATORS_AVG                       1.00         1.00   0.00
## ENTRANCES_AVG                       1.00         1.00   0.00
## FLOORSMAX_AVG                       1.00         1.00   0.00
## FLOORSMIN_AVG                       1.00         1.00   0.00
## LANDAREA_AVG                        1.00         1.00   0.00
```

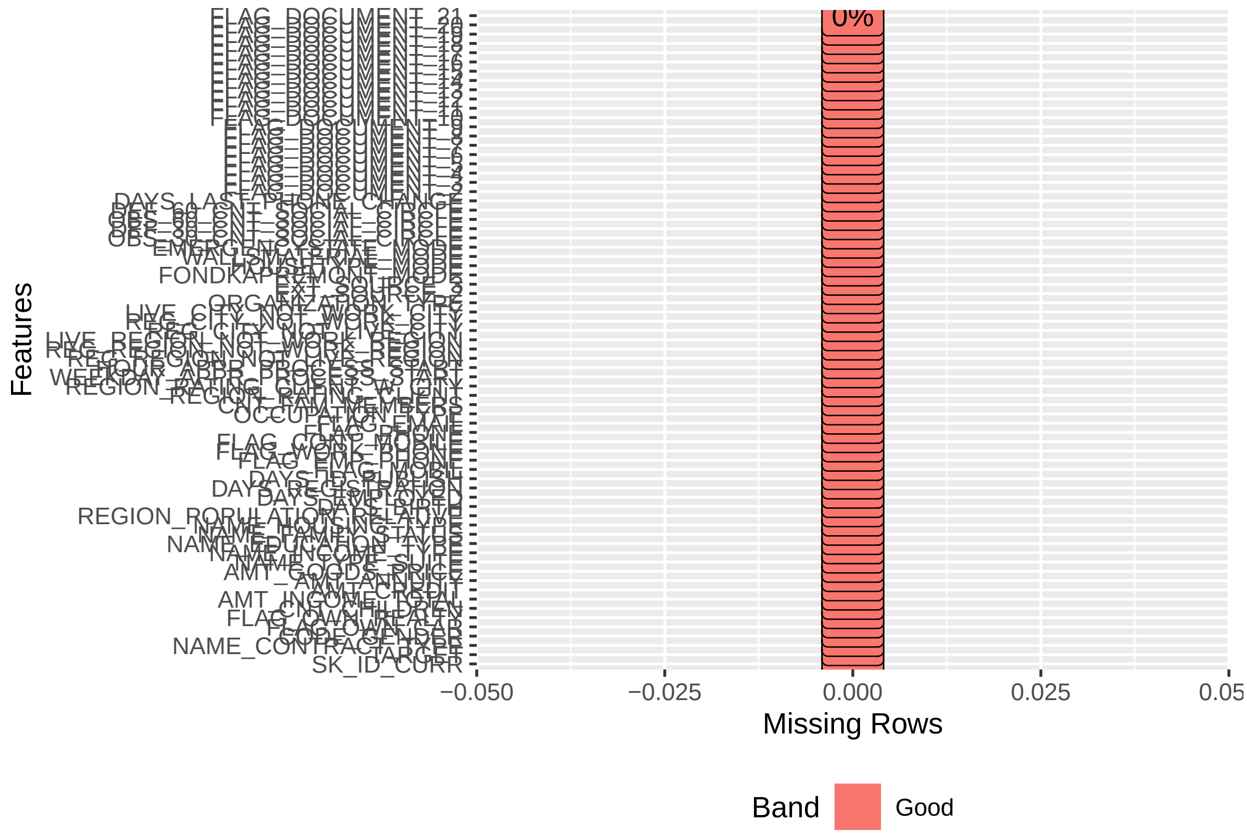
```
## LIVINGAPARTMENTS_AVG                1.00         1.00   0.00
## LIVINGAREA_AVG                      1.00         1.00   0.00
## NONLIVINGAPARTMENTS_AVG             1.00         1.00   0.00
## NONLIVINGAREA_AVG                   1.00         1.00   0.00
## APARTMENTS_MODE                     1.00         1.00   0.00
## BASEMENTAREA_MODE                   1.00         1.00   0.00
## YEARS_BEGINEXPLUATATION_MODE        1.00         1.00   0.00
## YEARS_BUILD_MODE                    1.00         1.00   0.00
## COMMONAREA_MODE                     1.00         1.00   0.00
## ELEVATORS_MODE                      1.00         1.00   0.00
## ENTRANCES_MODE                      1.00         1.00   0.00
## FLOORSMAX_MODE                      1.00         1.00   0.00
## FLOORSMIN_MODE                      1.00         1.00   0.00
## LANDAREA_MODE                       1.00         1.00   0.00
## LIVINGAPARTMENTS_MODE               1.00         1.00   0.00
## LIVINGAREA_MODE                     1.00         1.00   0.00
## NONLIVINGAPARTMENTS_MODE            1.00         1.00   0.00
## NONLIVINGAREA_MODE                  1.00         1.00   0.00
## APARTMENTS_MEDI                     1.00         1.00   0.00
## BASEMENTAREA_MEDI                   1.00         1.00   0.00
## YEARS_BEGINEXPLUATATION_MEDI        1.00         1.00   0.00
## YEARS_BUILD_MEDI                    1.00         1.00   0.00
## COMMONAREA_MEDI                     1.00         1.00   0.00
## ELEVATORS_MEDI                      1.00         1.00   0.00
## ENTRANCES_MEDI                      1.00         1.00   0.00
## FLOORSMAX_MEDI                      1.00         1.00   0.00
## FLOORSMIN_MEDI                      1.00         1.00   0.00
## LANDAREA_MEDI                       1.00         1.00   0.00
## LIVINGAPARTMENTS_MEDI               1.00         1.00   0.00
## LIVINGAREA_MEDI                     1.00         1.00   0.00
## NONLIVINGAPARTMENTS_MEDI            1.00         1.00   0.00
## NONLIVINGAREA_MEDI                  1.00         1.00   0.00
## TOTALAREA_MODE                      1.00         1.00   0.00
## OBS_30_CNT_SOCIAL_CIRCLE          348.00       348.00   0.00
## DEF_30_CNT_SOCIAL_CIRCLE           34.00        34.00   0.00
## OBS_60_CNT_SOCIAL_CIRCLE          344.00       344.00   0.00
## DEF_60_CNT_SOCIAL_CIRCLE           24.00        24.00   0.00
## DAYS_LAST_PHONE_CHANGE              0.00      4292.00   1.49
```

```
# pre-transformation missing data plot

plot_missing(training)
```

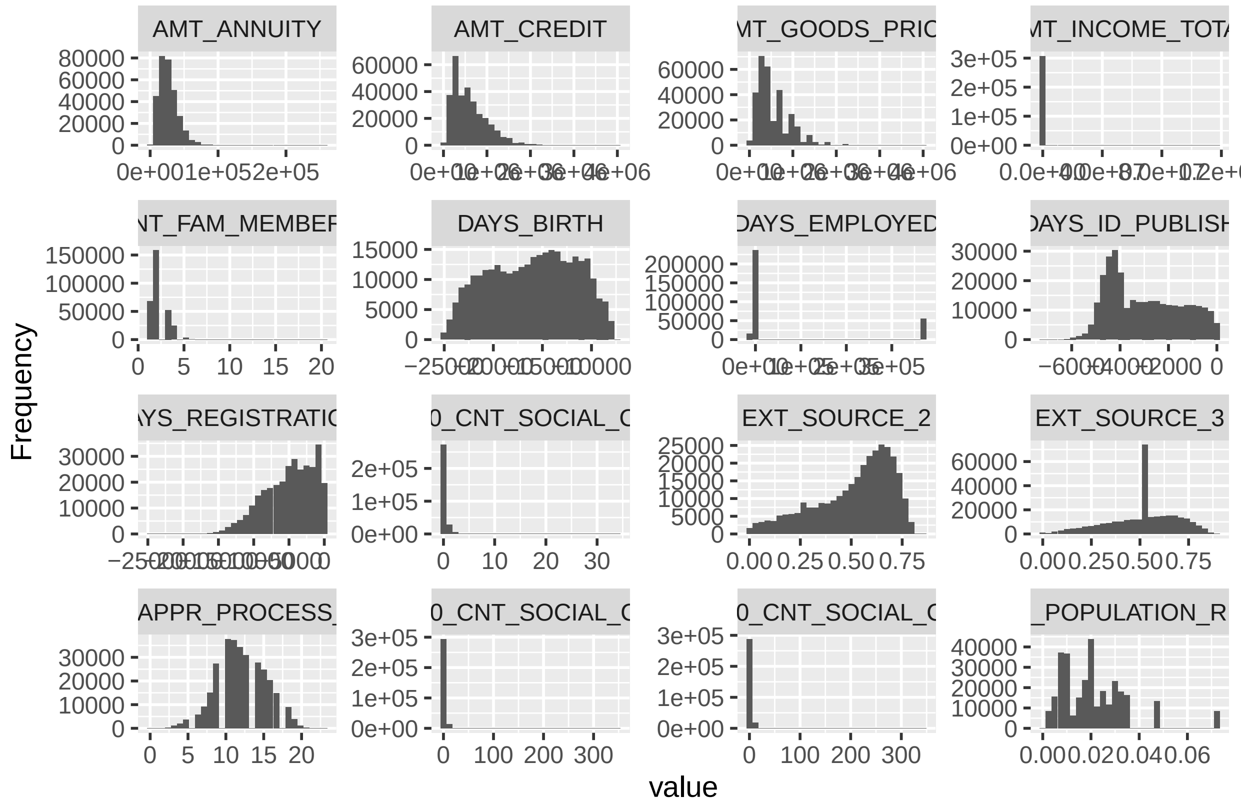



```r
# Clean/remove columns with missing data

training = training[, which(colMeans(!is.na(training)) > 0.6)]
training = imputeTS::na_mean(training,option = "median")
```

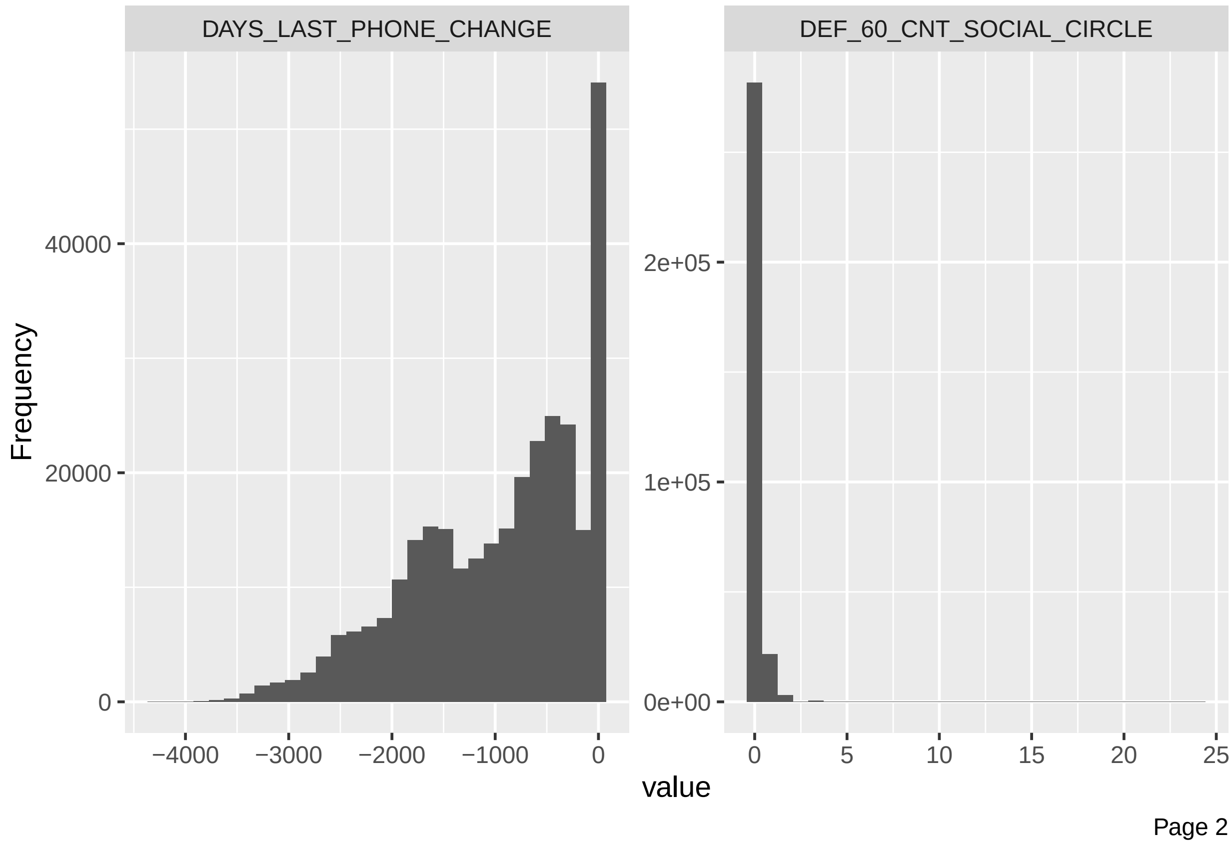
```r
# Post transformation missing data plot

plot_missing(training)
```




```r
# Histogram plot

quant_var <- split_columns(training%>% select(-SK_ID_CURR))
plot_histogram(quant_var$continuous)
```

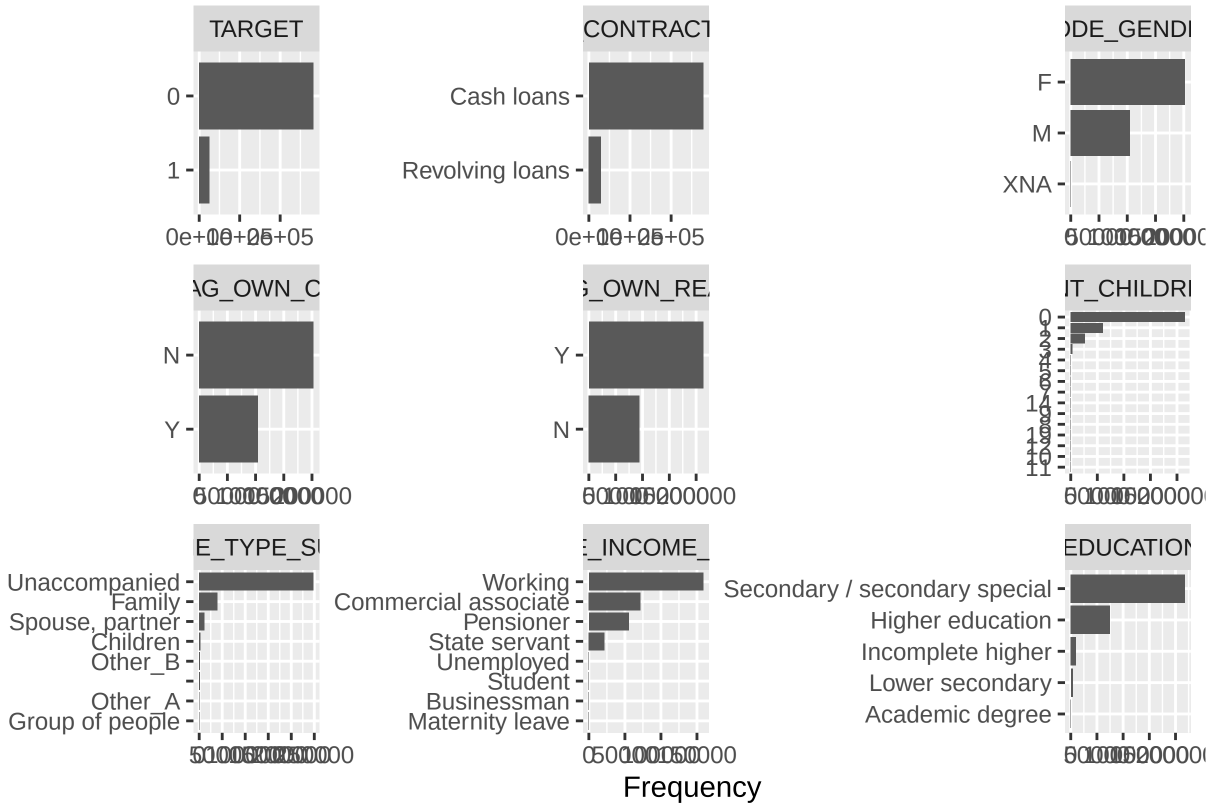

AMT_ANNUITY
AMT_CREDIT
MT_GOODS_PRIC
MT_INCOME_TOTA
NT_FAM_MEMBER
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_ID_PUBLISH
YS_REGISTRATIC
0_CNT_SOCIAL_C
EXT_SOURCE_2
EXT_SOURCE_3
APPR_PROCESS_
0_CNT_SOCIAL_C
0_CNT_SOCIAL_C
_POPULATION_R
Frequency
value

Page 1



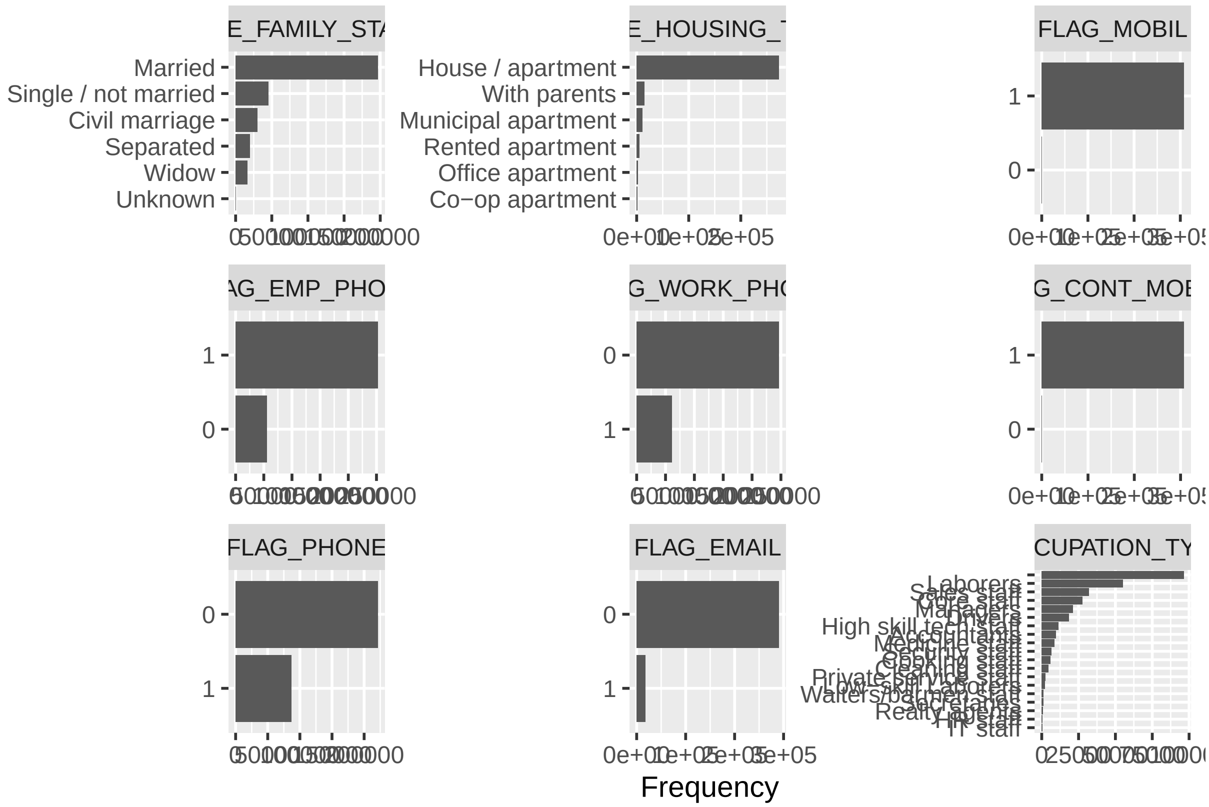

```r
# Bar Plot

plot_bar(quant_var$discrete)
```

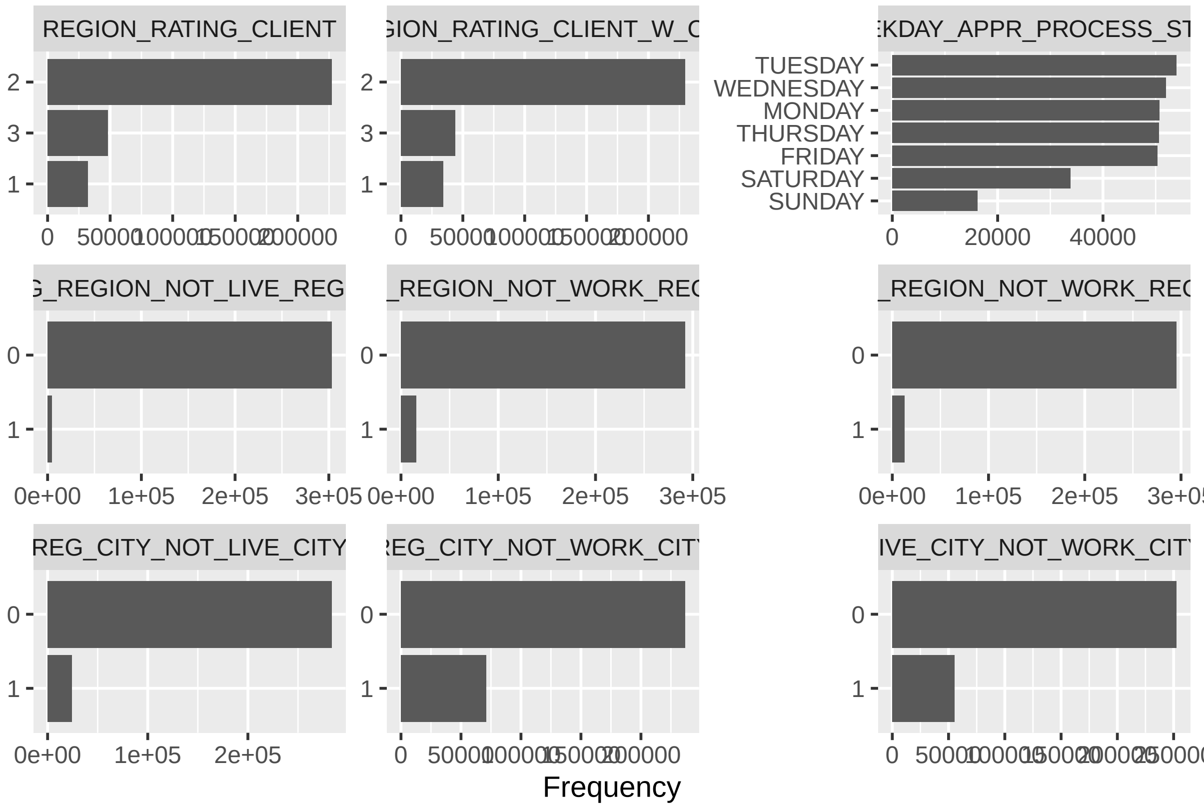
```
## 1 columns ignored with more than 50 categories.
## ORGANIZATION_TYPE: 58 categories
```



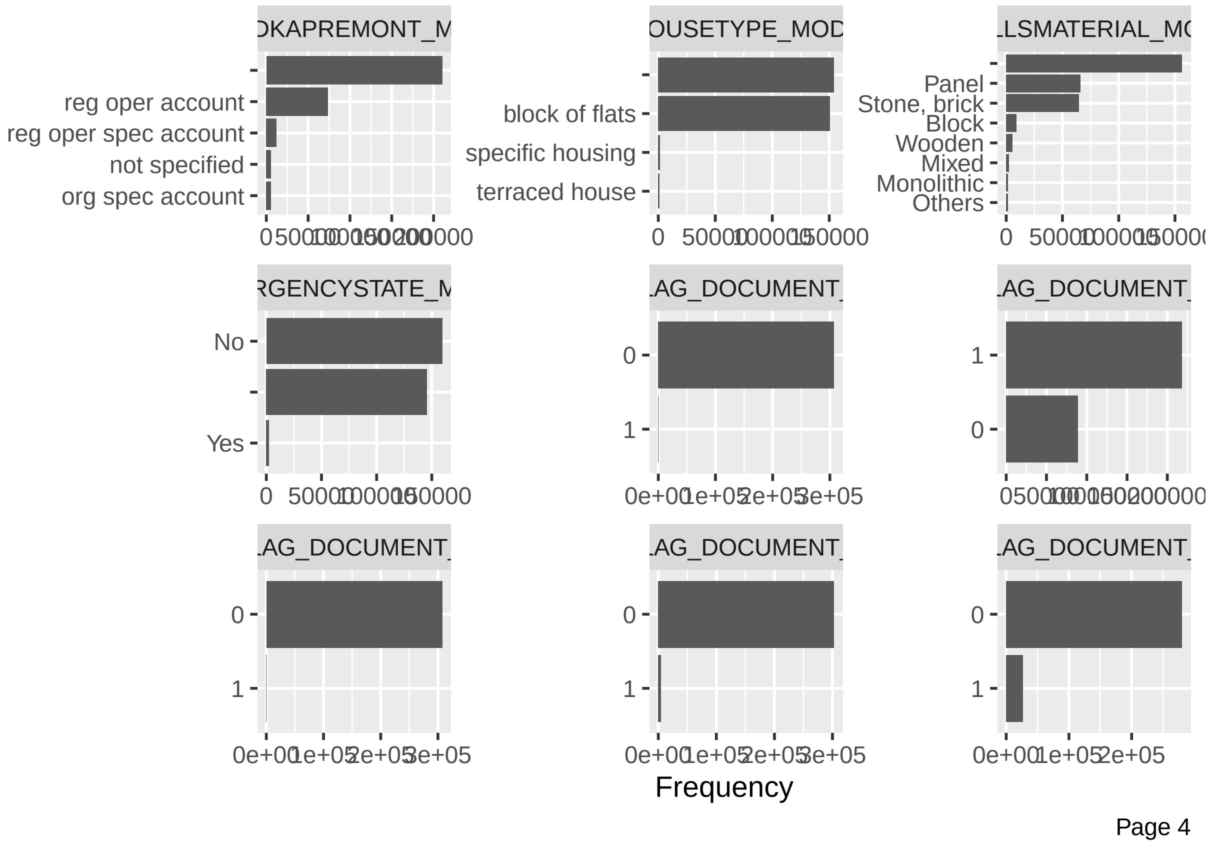
TARGET
0
1
CONTRACT
Cash loans
Revolving loans
DE_GEND
F
M
XNA
AG_OWN_C
N
Y
_OWN_RE
Y
N
IT_CHILDR
E_TYPE_S
Unaccompanied
Family
Spouse, partner
Children
Other_B
Other_A
Group of people
INCOME
Working
Commercial associate
Pensioner
State servant
Unemployed
Student
Businessman
Maternity leave
EDUCATION
Secondary / secondary special
Higher education
Incomplete higher
Lower secondary
Academic degree
Frequency

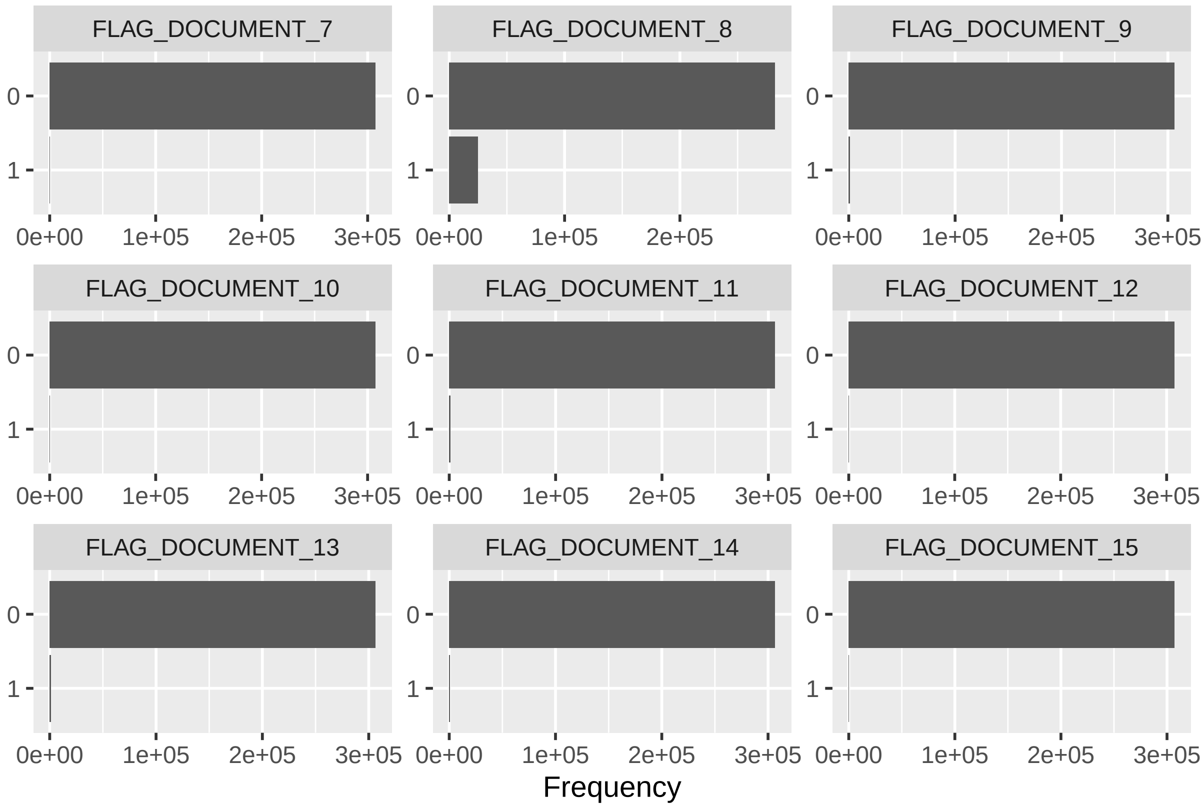
Page 1


E_FAMILY_STA
Married
Single / not married
Civil marriage
Separated
Widow
Unknown
E_HOUSING_T
House / apartment
With parents
Municipal apartment
Rented apartment
Office apartment
Co−op apartment
FLAG_MOBIL
AG_EMP_PHO
G_WORK_PH
G_CONT_MOB
FLAG_PHONE
FLAG_EMAIL
CUPATION_TY
Laborers
High skill tech staff
IT staff
Frequency
Page 2

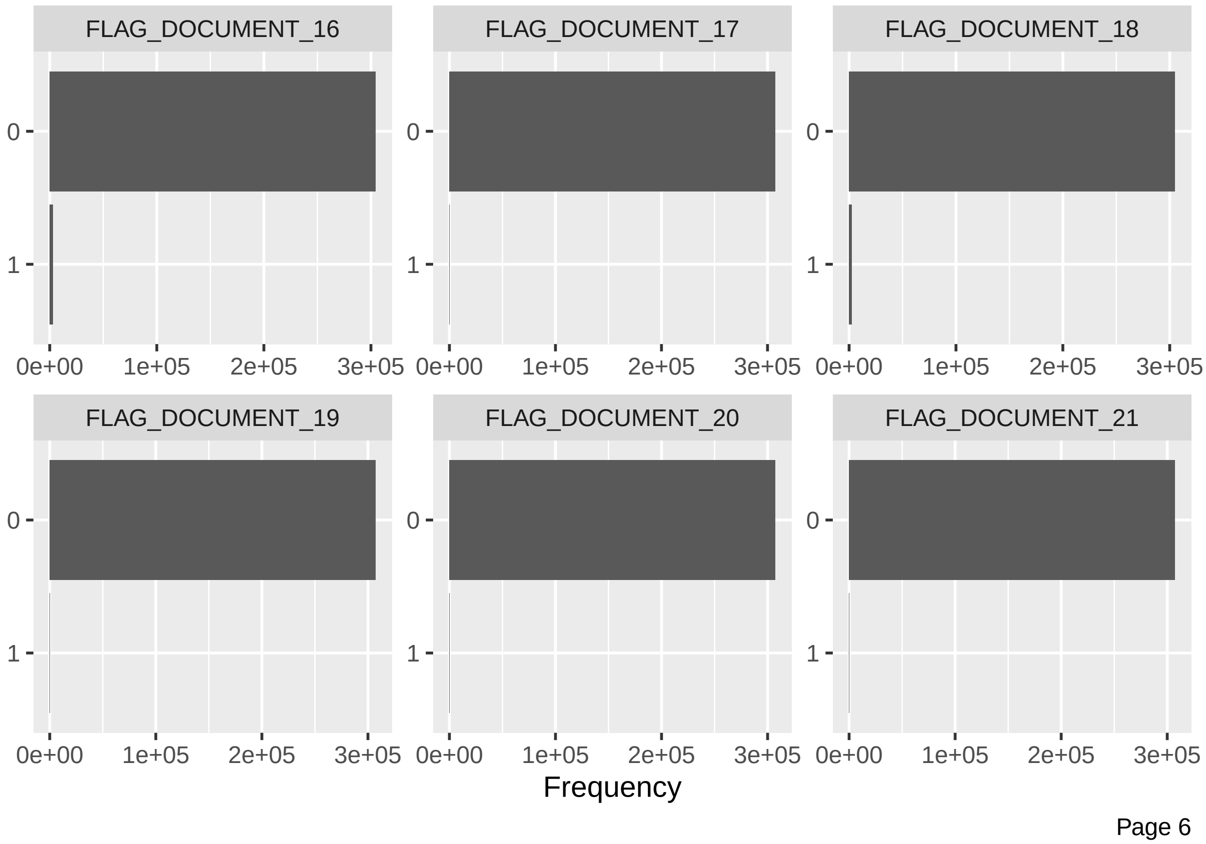

REGION_RATING_CLIENT
GION_RATING_CLIENT_W_C
EKDAY_APPR_PROCESS_ST
TUESDAY
WEDNESDAY
MONDAY
THURSDAY
FRIDAY
SATURDAY
SUNDAY
G_REGION_NOT_LIVE_REG
_REGION_NOT_WORK_REG
_REGION_NOT_WORK_REG
REG_CITY_NOT_LIVE_CITY
REG_CITY_NOT_WORK_CITY
IVE_CITY_NOT_WORK_CITY
Frequency

Page 3

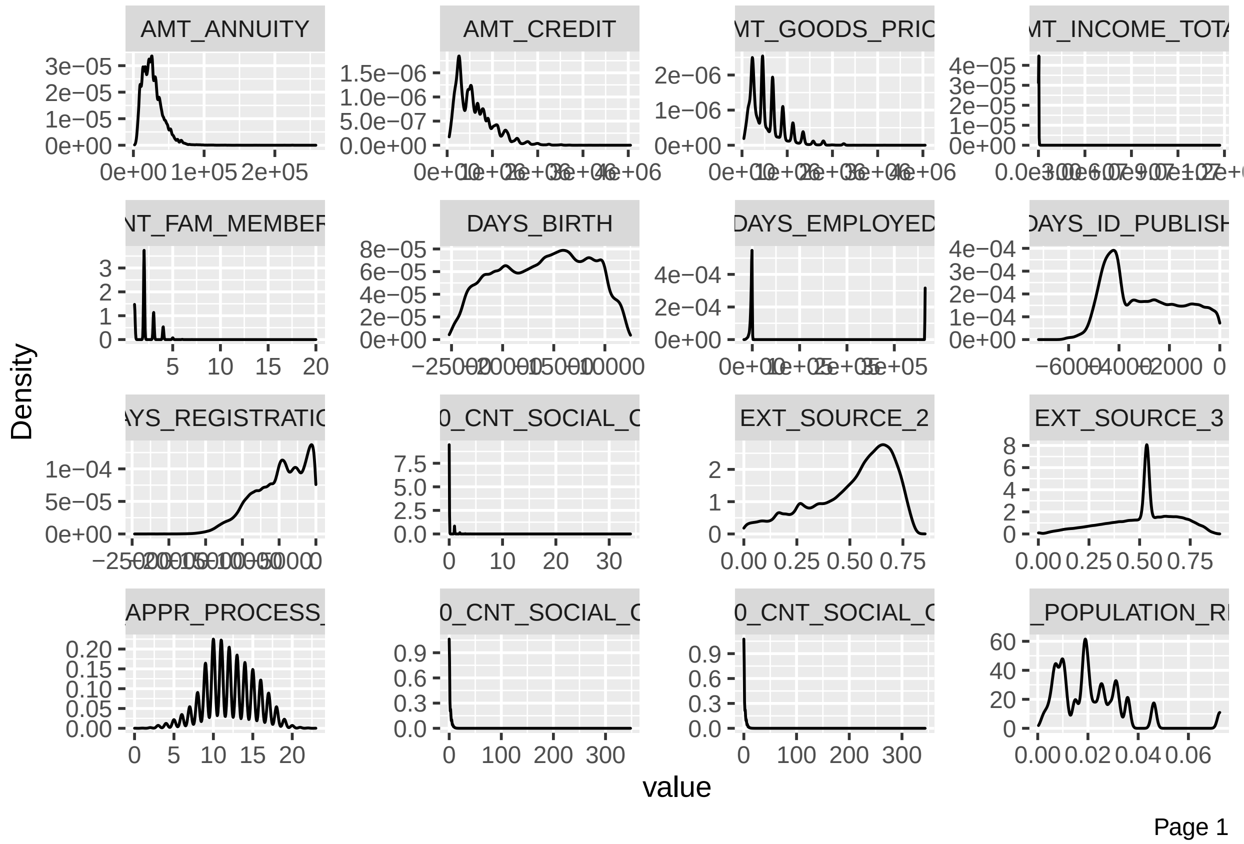

OKAPREMONT_M
reg oper account
reg oper spec account
not specified
org spec account
OUSETYPE_MOD
block of flats
specific housing
terraced house
LLSMATERIAL_MO
Panel
Stone, brick
Block
Wooden
Mixed
Monolithic
Others
RGENCYSTATE_M
No
Yes
LAG_DOCUMENT_
LAG_DOCUMENT_
LAG_DOCUMENT_
LAG_DOCUMENT_
LAG_DOCUMENT_
Frequency
Page 4


FLAG_DOCUMENT_7
FLAG_DOCUMENT_8
FLAG_DOCUMENT_9
FLAG_DOCUMENT_10
FLAG_DOCUMENT_11
FLAG_DOCUMENT_12
FLAG_DOCUMENT_13
FLAG_DOCUMENT_14
FLAG_DOCUMENT_15
0
1
0e+00
1e+05
2e+05
3e+05
Frequency

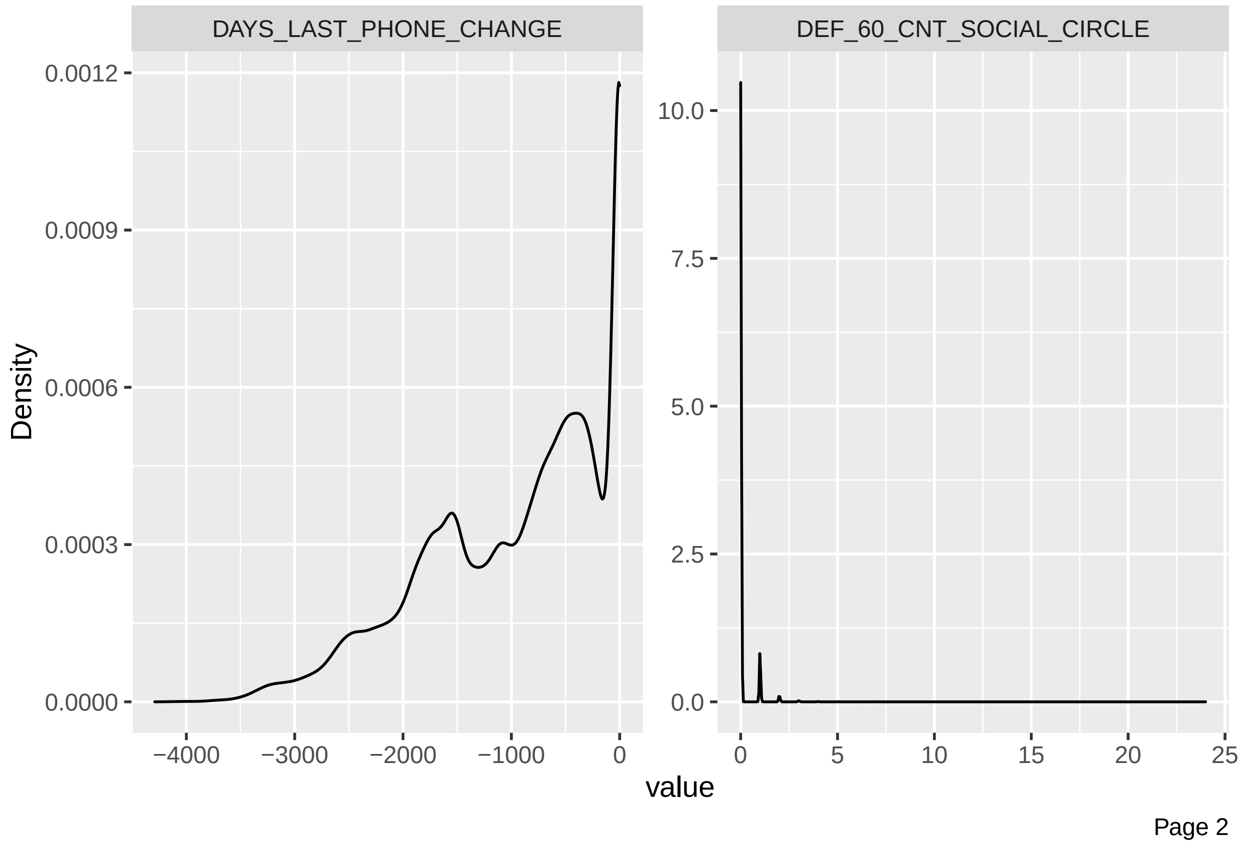
Page 5



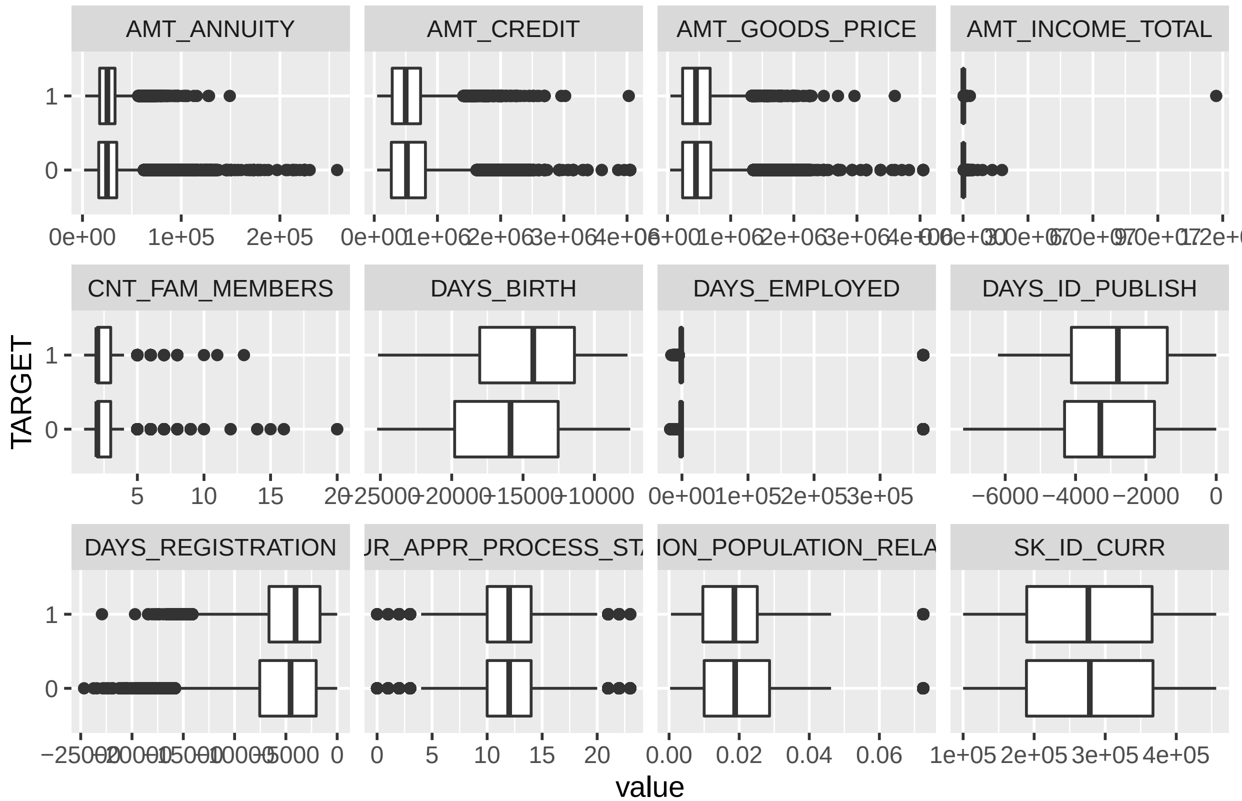

```
# Density Plot

plot_density(quant_var$continuous)
```



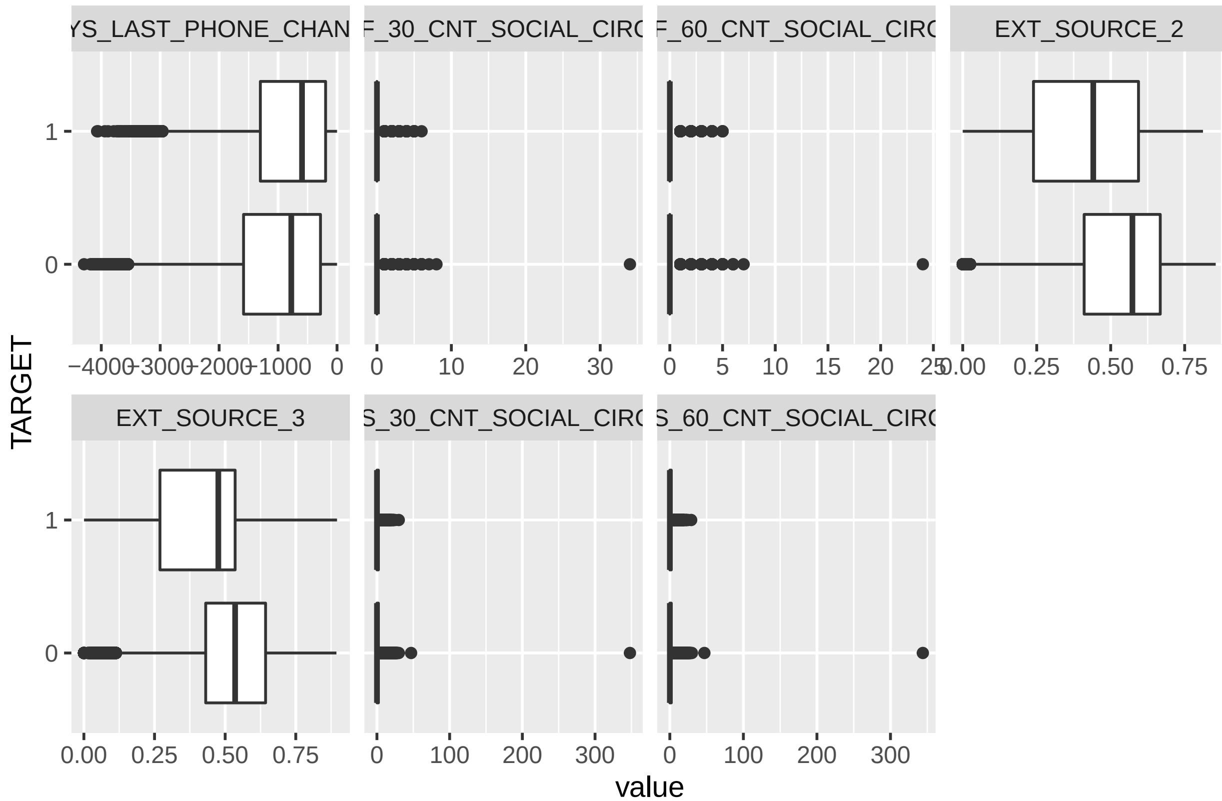
AMT_ANNUITY
AMT_CREDIT
MT_GOODS_PRIC
MT_INCOME_TOTA
NT_FAM_MEMBER
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_ID_PUBLISH
AYS_REGISTRATIO
0_CNT_SOCIAL_C
EXT_SOURCE_2
EXT_SOURCE_3
APPR_PROCESS_
0_CNT_SOCIAL_C
0_CNT_SOCIAL_C
_POPULATION_R
Density
value
Page 1




```r
# Boxplot

plot_boxplot(
  data = training,
  by = "TARGET")
```


AMT_ANNUITY
AMT_CREDIT
AMT_GOODS_PRICE
AMT_INCOME_TOTAL
CNT_FAM_MEMBERS
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_ID_PUBLISH
DAYS_REGISTRATION
SK_ID_CURR
TARGET
value

Page 1



Page 2

```r
# Proportion tables

for( i in colnames(training[sapply(training, is.factor)])){
  print(i)
  print(table(training[,i], TARGET_FLAG=training$TARGET))
}
```

```
## [1] "TARGET"
##    TARGET_FLAG
##          0      1
##   0 282686      0
##   1      0  24825
## [1] "NAME_CONTRACT_TYPE"
##                   TARGET_FLAG
##                         0      1
##   Cash loans       255011  23221
##   Revolving loans   27675   1604
## [1] "CODE_GENDER"
##      TARGET_FLAG
##            0      1
##   F   188278  14170
##   M    94404  10655
##   XNA      4      0
## [1] "FLAG_OWN_CAR"
##    TARGET_FLAG
##          0      1
```

```
##   N 185675  17249
##   Y  97011   7576
## [1] "FLAG_OWN_REALTY"
##    TARGET_FLAG
##          0      1
##   N  86357   7842
##   Y 196329  16983
## [1] "CNT_CHILDREN"
##     TARGET_FLAG
##           0      1
##   0  198762  16609
##   1   55665   5454
##   2   24416   2333
##   3    3359    358
##   4     374     55
##   5      77      7
##   6      15      6
##   7       7      0
##   8       2      0
##   9       0      2
##   10      2      0
##   11      0      1
##   12      2      0
##   14      3      0
##   19      2      0
## [1] "NAME_TYPE_SUITE"
##                   TARGET_FLAG
##                         0      1
##                      1222     70
##   Children           3026    241
##   Family            37140   3009
##   Group of people     248     23
##   Other_A             790     76
##   Other_B            1596    174
##   Spouse, partner   10475    895
##   Unaccompanied    228189  20337
## [1] "NAME_INCOME_TYPE"
##                        TARGET_FLAG
##                              0      1
##   Businessman               10      0
##   Commercial associate   66257   5360
##   Maternity leave            3      2
##   Pensioner              52380   2982
##   State servant          20454   1249
##   Student                   18      0
##   Unemployed                14      8
##   Working               143550  15224
## [1] "NAME_EDUCATION_TYPE"
##                                 TARGET_FLAG
##                                      0      1
##   Academic degree                  161      3
##   Higher education               70854   4009
##   Incomplete higher               9405    872
##   Lower secondary                 3399    417
```

```
##   Secondary / secondary special 198867  19524
## [1] "NAME_FAMILY_STATUS"
##                       TARGET_FLAG
##                             0      1
##   Civil marriage        26814   2961
##   Married              181582  14850
##   Separated             18150   1620
##   Single / not married  40987   4457
##   Unknown                   2      0
##   Widow                 15151    937
## [1] "NAME_HOUSING_TYPE"
##                      TARGET_FLAG
##                            0      1
##   Co-op apartment       1033     89
##   House / apartment   251596  21272
##   Municipal apartment  10228    955
##   Office apartment      2445    172
##   Rented apartment      4280    601
##   With parents         13104   1736
## [1] "FLAG_MOBIL"
##    TARGET_FLAG
##          0      1
##   0      1      0
##   1 282685  24825
## [1] "FLAG_EMP_PHONE"
##    TARGET_FLAG
##          0      1
##   0  52395   2991
##   1 230291  21834
## [1] "FLAG_WORK_PHONE"
##    TARGET_FLAG
##          0      1
##   0 227282  18921
##   1  55404   5904
## [1] "FLAG_CONT_MOBILE"
##    TARGET_FLAG
##          0      1
##   0    529     45
##   1 282157  24780
## [1] "FLAG_PHONE"
##    TARGET_FLAG
##          0      1
##   0 202336  18744
##   1  80350   6081
## [1] "FLAG_EMAIL"
##    TARGET_FLAG
##          0      1
##   0 266618  23451
##   1  16068   1374
## [1] "OCCUPATION_TYPE"
##                        TARGET_FLAG
##                             0     1
##                         90113  6278
##   Accountants            9339   474
```

```
##   Cleaning staff         4206   447
##   Cooking staff          5325   621
##   Core staff            25832  1738
##   Drivers               16496  2107
##   High skill tech staff 10679   701
##   HR staff                527    36
##   IT staff                492    34
##   Laborers              49348  5838
##   Low-skill Laborers     1734   359
##   Managers              20043  1328
##   Medicine staff         7965   572
##   Private service staff  2477   175
##   Realty agents           692    59
##   Sales staff           29010  3092
##   Secretaries            1213    92
##   Security staff         5999   722
##   Waiters/barmen staff   1196   152
## [1] "REGION_RATING_CLIENT"
##    TARGET_FLAG
##          0      1
##   1  30645   1552
##   2 209077  17907
##   3  42964   5366
## [1] "REGION_RATING_CLIENT_W_CITY"
##    TARGET_FLAG
##          0      1
##   1  32513   1654
##   2 211314  18170
##   3  38859   5001
## [1] "WEEKDAY_APPR_PROCESS_START"
##            TARGET_FLAG
##                 0     1
##   FRIDAY    46237  4101
##   MONDAY    46780  3934
##   SATURDAY  31182  2670
##   SUNDAY    14898  1283
##   THURSDAY  46493  4098
##   TUESDAY   49400  4501
##   WEDNESDAY 47696  4238
## [1] "REG_REGION_NOT_LIVE_REGION"
##    TARGET_FLAG
##          0      1
##   0 278462  24392
##   1   4224    433
## [1] "REG_REGION_NOT_WORK_REGION"
##    TARGET_FLAG
##          0      1
##   0 268462  23437
##   1  14224   1388
## [1] "LIVE_REGION_NOT_WORK_REGION"
##    TARGET_FLAG
##          0      1
##   0 271239  23769
##   1  11447   1056
```

```
## [1] "REG_CITY_NOT_LIVE_CITY"
##    TARGET_FLAG
##          0      1
##   0 261586  21886
##   1  21100   2939
## [1] "REG_CITY_NOT_WORK_CITY"
##    TARGET_FLAG
##          0      1
##   0 219339  17305
##   1  63347   7520
## [1] "LIVE_CITY_NOT_WORK_CITY"
##    TARGET_FLAG
##          0      1
##   0 232974  19322
##   1  49712   5503
## [1] "ORGANIZATION_TYPE"
##                         TARGET_FLAG
##                              0     1
##   Advertising              394    35
##   Agriculture             2197   257
##   Bank                    2377   130
##   Business Entity Type 1  5497   487
##   Business Entity Type 2  9653   900
##   Business Entity Type 3 61669  6323
##   Cleaning                 231    29
##   Construction            5936   785
##   Culture                  358    21
##   Electricity              887    63
##   Emergency                520    40
##   Government              9678   726
##   Hotel                    904    62
##   Housing                 2723   235
##   Industry: type 1         924   115
##   Industry: type 10        102     7
##   Industry: type 11       2470   234
##   Industry: type 12        355    14
##   Industry: type 13         58     9
##   Industry: type 2         425    33
##   Industry: type 3        2930   348
##   Industry: type 4         788    89
##   Industry: type 5         558    41
##   Industry: type 6         104     8
##   Industry: type 7        1202   105
##   Industry: type 8          21     3
##   Industry: type 9        3143   225
##   Insurance                563    34
##   Kindergarten            6396   484
##   Legal Services           281    24
##   Medicine               10456   737
##   Military                2499   135
##   Mobile                   288    29
##   Other                  15408  1275
##   Police                  2224   117
##   Postal                  1975   182
```

```
##   Realtor                  354    42
##   Religion                  80     5
##   Restaurant              1599   212
##   School                  8367   526
##   Security                2923   324
##   Security Ministries     1878    96
##   Self-employed          34504  3908
##   Services                1471   104
##   Telecom                  533    44
##   Trade: type 1            317    31
##   Trade: type 2           1767   133
##   Trade: type 3           3131   361
##   Trade: type 4             62     2
##   Trade: type 5             46     3
##   Trade: type 6            602    29
##   Trade: type 7           7091   740
##   Transport: type 1        192     9
##   Transport: type 2       2032   172
##   Transport: type 3       1000   187
##   Transport: type 4       4897   501
##   University              1262    65
##   XNA                    52384  2990
## [1] "FONDKAPREMONT_MODE"
##                         TARGET_FLAG
##                               0      1
##                          192170  18125
##   not specified            5258    429
##   org spec account         5292    327
##   reg oper account        68678   5152
##   reg oper spec account  11288    792
## [1] "HOUSETYPE_MODE"
##                    TARGET_FLAG
##                          0      1
##                     140177  14120
##   block of flats    140053  10450
##   specific housing    1347    152
##   terraced house      1109    103
## [1] "WALLSMATERIAL_MODE"
##                TARGET_FLAG
##                      0      1
##                 142070  14271
##   Block           8603    650
##   Mixed           2123    173
##   Monolithic      1695     84
##   Others          1490    135
##   Panel          61848   4192
##   Stone, brick   60015   4800
##   Wooden          4842    520
## [1] "EMERGENCYSTATE_MODE"
##      TARGET_FLAG
##            0      1
##       132257  13498
##   No  148324  11104
##   Yes   2105    223
```

```
## [1] "FLAG_DOCUMENT_2"
##    TARGET_FLAG
##          0      1
##   0 282677  24821
##   1      9      4
## [1] "FLAG_DOCUMENT_3"
##    TARGET_FLAG
##          0      1
##   0  83658   5513
##   1 199028  19312
## [1] "FLAG_DOCUMENT_4"
##    TARGET_FLAG
##          0      1
##   0 282661  24825
##   1     25      0
## [1] "FLAG_DOCUMENT_5"
##    TARGET_FLAG
##          0      1
##   0 278410  24453
##   1   4276    372
## [1] "FLAG_DOCUMENT_6"
##    TARGET_FLAG
##          0      1
##   0 257115  23318
##   1  25571   1507
## [1] "FLAG_DOCUMENT_7"
##    TARGET_FLAG
##          0      1
##   0 282630  24822
##   1     56      3
## [1] "FLAG_DOCUMENT_8"
##    TARGET_FLAG
##          0      1
##   0 259498  22989
##   1  23188   1836
## [1] "FLAG_DOCUMENT_9"
##    TARGET_FLAG
##          0      1
##   0 281562  24751
##   1   1124     74
## [1] "FLAG_DOCUMENT_10"
##    TARGET_FLAG
##          0      1
##   0 282679  24825
##   1      7      0
## [1] "FLAG_DOCUMENT_11"
##    TARGET_FLAG
##          0      1
##   0 281558  24750
##   1   1128     75
## [1] "FLAG_DOCUMENT_12"
##    TARGET_FLAG
##          0      1
##   0 282684  24825
```

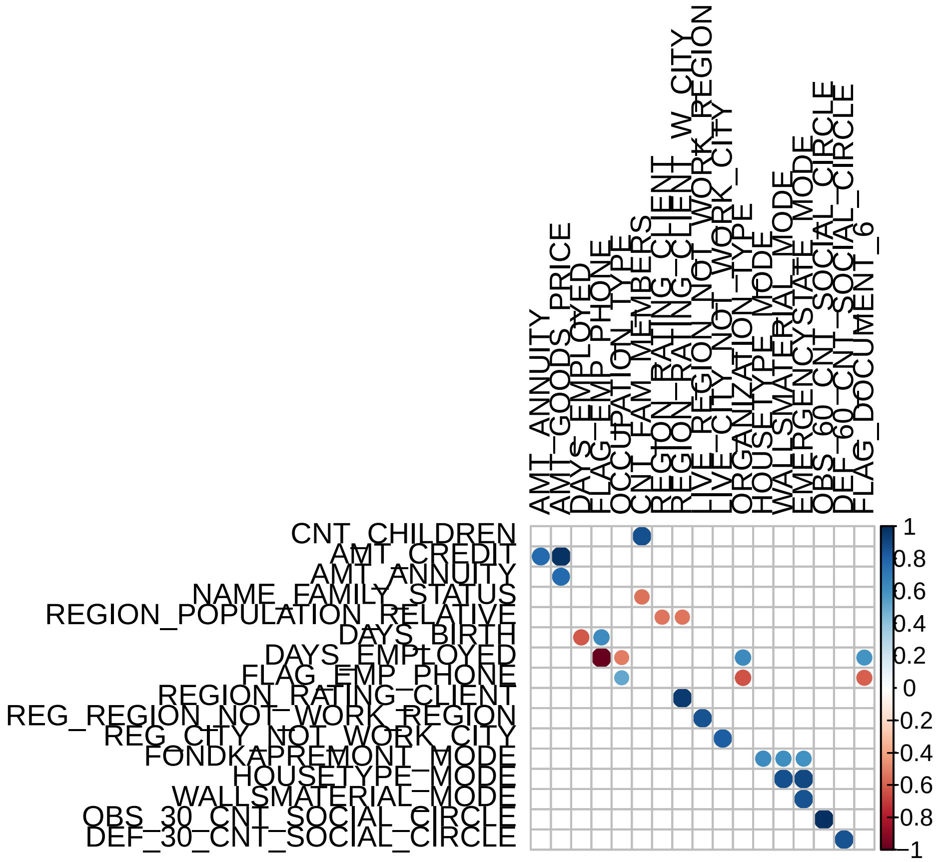
```
##   1       2       0
## [1] "FLAG_DOCUMENT_13"
##    TARGET_FLAG
##          0       1
##   0 281632   24795
##   1   1054      30
## [1] "FLAG_DOCUMENT_14"
##    TARGET_FLAG
##          0       1
##   0 281813   24795
##   1    873      30
## [1] "FLAG_DOCUMENT_15"
##    TARGET_FLAG
##          0       1
##   0 282325   24814
##   1    361      11
## [1] "FLAG_DOCUMENT_16"
##    TARGET_FLAG
##          0       1
##   0 279783   24675
##   1   2903     150
## [1] "FLAG_DOCUMENT_17"
##    TARGET_FLAG
##          0       1
##   0 282606   24823
##   1     80       2
## [1] "FLAG_DOCUMENT_18"
##    TARGET_FLAG
##          0       1
##   0 280328   24683
##   1   2358     142
## [1] "FLAG_DOCUMENT_19"
##    TARGET_FLAG
##          0       1
##   0 282515   24813
##   1    171      12
## [1] "FLAG_DOCUMENT_20"
##    TARGET_FLAG
##          0       1
##   0 282543   24812
##   1    143      13
## [1] "FLAG_DOCUMENT_21"
##    TARGET_FLAG
##          0       1
##   0 282597   24811
##   1     89      14
```

```r
# Correlation Function

corr_simple <- function(data=df,sig=0.5){

  library(corrplot)
  df_cor <- data %>% mutate_if(is.character, as.factor)
  df_cor <- df_cor %>% mutate_if(is.factor, as.numeric)
```

```
  corr <- cor(df_cor)
  corr[lower.tri(corr,diag=TRUE)] <- NA
  corr[corr == 1] <- NA
  corr <- as.data.frame(as.table(corr))

  corr <- na.omit(corr)   #select significant values
  corr <- subset(corr, abs(Freq) > sig)

  corr <- corr[order(-abs(corr$Freq)),]
  print(corr)
  mtx_corr <- reshape2::acast(corr, Var1~Var2, value.var="Freq")


  par( cex= 0.9,mar=c(6,4,6,4) )
  corrplot(mtx_corr, is.corr=FALSE, tl.col="black", na.label=" ",number.cex= 7/ncol(corr))
}
```

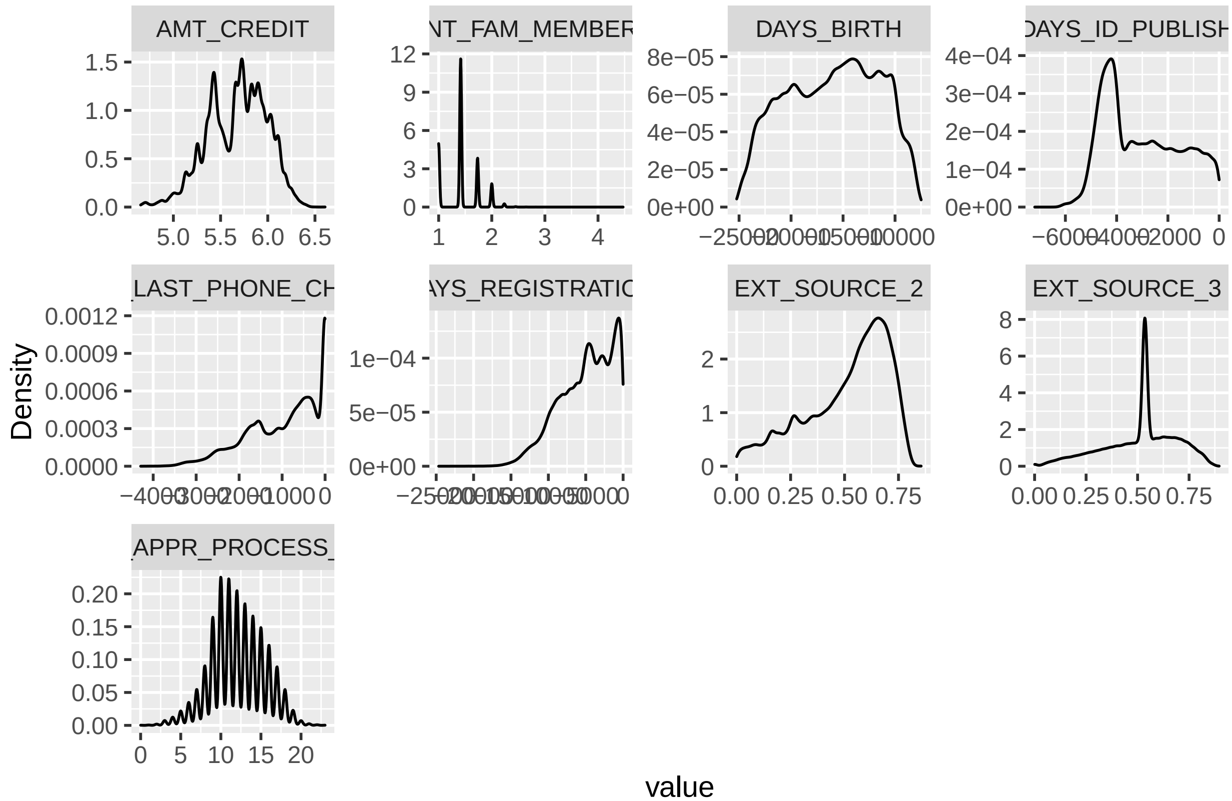
```
#Correlation plot

corr_simple(training)
```

```
##                           Var1                         Var2       Freq
## 1581             DAYS_EMPLOYED                FLAG_EMP_PHONE -0.9997550
## 3455    OBS_30_CNT_SOCIAL_CIRCLE    OBS_60_CNT_SOCIAL_CIRCLE  0.9984913
## 719                 AMT_CREDIT               AMT_GOODS_PRICE  0.9867343
## 2160       REGION_RATING_CLIENT REGION_RATING_CLIENT_W_CITY  0.9508422
## 3239            HOUSETYPE_MODE          EMERGENCYSTATE_MODE  0.9005561
## 1995              CNT_CHILDREN               CNT_FAM_MEMBERS  0.8791749
## 3168            HOUSETYPE_MODE            WALLSMATERIAL_MODE  0.8716528
## 3240        WALLSMATERIAL_MODE          EMERGENCYSTATE_MODE  0.8654099
## 2520 REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION  0.8606268
## 3527    DEF_30_CNT_SOCIAL_CIRCLE    DEF_60_CNT_SOCIAL_CIRCLE  0.8605555
## 2736     REG_CITY_NOT_WORK_CITY      LIVE_CITY_NOT_WORK_CITY  0.8255747
## 720                AMT_ANNUITY               AMT_GOODS_PRICE  0.7748366
## 648                 AMT_CREDIT                   AMT_ANNUITY  0.7701267
## 2792            FLAG_EMP_PHONE             ORGANIZATION_TYPE -0.6308378
## 2788             DAYS_EMPLOYED             ORGANIZATION_TYPE  0.6304793
## 3096        FONDKAPREMONT_MODE                HOUSETYPE_MODE  0.6200414
## 1580                DAYS_BIRTH                FLAG_EMP_PHONE  0.6198877
## 1296                DAYS_BIRTH                 DAYS_EMPLOYED -0.6158642
## 3167        FONDKAPREMONT_MODE            WALLSMATERIAL_MODE  0.6124898
## 3238        FONDKAPREMONT_MODE          EMERGENCYSTATE_MODE  0.6036966
## 3928            FLAG_EMP_PHONE               FLAG_DOCUMENT_6 -0.5977318
## 3924             DAYS_EMPLOYED               FLAG_DOCUMENT_6  0.5974844
## 2003        NAME_FAMILY_STATUS               CNT_FAM_MEMBERS -0.5485502
## 2076 REGION_POPULATION_RELATIVE         REGION_RATING_CLIENT -0.5328765
## 2147 REGION_POPULATION_RELATIVE REGION_RATING_CLIENT_W_CITY -0.5315355
## 1940            FLAG_EMP_PHONE               OCCUPATION_TYPE  0.5158064
## 1936             DAYS_EMPLOYED               OCCUPATION_TYPE -0.5151704
```




```r
# skewness transformation

for(i in colnames(training[sapply(training, is.numeric)] %>%
                  select(-SK_ID_CURR))){
  if ((skewness(training[,i])>.5) & (skewness(training[,i])<1)){
    training[,i] = sqrt(training[,i])
  } else if ((skewness(training[,i])>=1) & (skewness(training[,i])<1.3)){
    training[,i] = log10(training[,i])
  } else if (skewness(training[,i])>=1.3){
    training[,i] = ifelse(skewness(training[,i])>0,1/training[,i],
                          1/(max(training[,i]+1)-training[,i]))}
  }
```

```r
# Post skewness density plot

plot_density(training%>% select(-SK_ID_CURR))
```




```
# Model Construction

set.seed(120)


glm_train <- createDataPartition(training$TARGET, p=.8, list=FALSE, times = 1)
glm_test <- training[-glm_train,]
glm_train <- training[glm_train,]

str(glm_train)
```

```
## 'data.frame':    246009 obs. of  71 variables:
##  $ SK_ID_CURR                   : int  100002 100003 100004 100006 100007 100008 100009 100010 100011
##  $ TARGET                       : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 1 1 ...
##  $ NAME_CONTRACT_TYPE           : Factor w/ 2 levels "Cash loans","Revolving loans": 1 1 2 1 1 1 1 1
##  $ CODE_GENDER                  : Factor w/ 3 levels "F","M","XNA": 2 1 2 1 2 2 1 2 1 2 ...
##  $ FLAG_OWN_CAR                 : Factor w/ 2 levels "N","Y": 1 1 2 1 1 1 2 2 1 1 ...
##  $ FLAG_OWN_REALTY              : Factor w/ 2 levels "N","Y": 2 1 2 2 2 2 2 2 2 2 ...
##  $ CNT_CHILDREN                 : Factor w/ 15 levels "0","1","2","3",..: 1 1 1 1 1 1 2 1 1 1 ...
##  $ AMT_INCOME_TOTAL             : num  4.94e-06 4.94e-06 4.94e-06 4.94e-06 4.94e-06 ...
##  $ AMT_CREDIT                   : num  5.61 6.11 5.13 5.5 5.71 ...
##  $ AMT_ANNUITY                  : num  4.05e-05 4.05e-05 4.05e-05 4.05e-05 4.05e-05 ...
##  $ AMT_GOODS_PRICE              : num  2.85e-06 2.85e-06 2.85e-06 2.85e-06 2.85e-06 ...
##  $ NAME_TYPE_SUITE              : Factor w/ 8 levels "","Children",..: 8 3 8 8 8 7 8 8 2 8 ...
##  $ NAME_INCOME_TYPE             : Factor w/ 8 levels "Businessman",..: 8 5 8 8 8 5 2 5 4 8 ...
##  $ NAME_EDUCATION_TYPE          : Factor w/ 5 levels "Academic degree",..: 5 2 5 5 5 5 2 2 5 5 ...
```

```
## $ NAME_FAMILY_STATUS         : Factor w/ 6 levels "Civil marriage",..: 4 2 4 1 4 2 2 2 2 4 ...
## $ NAME_HOUSING_TYPE          : Factor w/ 6 levels "Co-op apartment",..: 2 2 2 2 2 2 2 2 2 2 ...
## $ REGION_POPULATION_RELATIVE : num  53.2 53.2 53.2 53.2 53.2 ...
## $ DAYS_BIRTH                 : int  -9461 -16765 -19046 -19005 -19932 -16941 -13778 -18850 -20099 -
## $ DAYS_EMPLOYED              : num  -0.00157 -0.00157 -0.00157 -0.00157 -0.00157 ...
## $ DAYS_REGISTRATION          : num  -3648 -1186 -4260 -9833 -4311 ...
## $ DAYS_ID_PUBLISH            : int  -2120 -291 -2531 -2437 -3458 -477 -619 -2379 -3514 -3992 ...
## $ FLAG_MOBIL                 : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ FLAG_EMP_PHONE             : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 1 2 ...
## $ FLAG_WORK_PHONE            : Factor w/ 2 levels "0","1": 1 1 2 1 1 2 1 2 1 1 ...
## $ FLAG_CONT_MOBILE           : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ FLAG_PHONE                 : Factor w/ 2 levels "0","1": 2 2 2 1 1 2 2 1 1 1 ...
## $ FLAG_EMAIL                 : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ OCCUPATION_TYPE            : Factor w/ 19 levels "","Accountants",..: 10 5 10 10 5 10 2 12 1 10 .
## $ CNT_FAM_MEMBERS            : num  1 1.41 1 1.41 1 ...
## $ REGION_RATING_CLIENT       : Factor w/ 3 levels "1","2","3": 2 1 2 2 2 2 2 3 2 2 ...
## $ REGION_RATING_CLIENT_W_CITY: Factor w/ 3 levels "1","2","3": 2 1 2 2 2 2 2 3 2 2 ...
## $ WEEKDAY_APPR_PROCESS_START : Factor w/ 7 levels "FRIDAY","MONDAY",..: 7 2 2 7 5 7 4 2 7 5 ...
## $ HOUR_APPR_PROCESS_START    : int  10 11 9 17 11 16 16 16 14 8 ...
## $ REG_REGION_NOT_LIVE_REGION : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ REG_REGION_NOT_WORK_REGION : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ LIVE_REGION_NOT_WORK_REGION: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ REG_CITY_NOT_LIVE_CITY     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ REG_CITY_NOT_WORK_CITY     : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 2 1 1 ...
## $ LIVE_CITY_NOT_WORK_CITY    : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 2 1 1 ...
## $ ORGANIZATION_TYPE          : Factor w/ 58 levels "Advertising",..: 6 40 12 6 38 34 6 34 58 10 ...
## $ EXT_SOURCE_2               : num  0.263 0.622 0.556 0.65 0.323 ...
## $ EXT_SOURCE_3               : num  0.139 0.535 0.73 0.535 0.535 ...
## $ FONDKAPREMONT_MODE         : Factor w/ 5 levels "","not specified",..: 4 4 1 1 1 1 1 1 1 1 ...
## $ HOUSETYPE_MODE             : Factor w/ 4 levels "","block of flats",..: 2 2 1 1 1 1 1 1 1 1 ...
## $ WALLSMATERIAL_MODE         : Factor w/ 8 levels "","Block","Mixed",..: 7 2 1 1 1 1 1 1 1 1 ...
## $ EMERGENCYSTATE_MODE        : Factor w/ 3 levels "","No","Yes": 2 2 1 1 1 1 1 1 1 1 ...
## $ OBS_30_CNT_SOCIAL_CIRCLE   : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
## $ DEF_30_CNT_SOCIAL_CIRCLE   : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
## $ OBS_60_CNT_SOCIAL_CIRCLE   : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
## $ DEF_60_CNT_SOCIAL_CIRCLE   : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
## $ DAYS_LAST_PHONE_CHANGE     : num  -1134 -828 -815 -617 -1106 ...
## $ FLAG_DOCUMENT_2            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_3            : Factor w/ 2 levels "0","1": 2 2 1 2 1 2 1 2 2 1 ...
## $ FLAG_DOCUMENT_4            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_5            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_6            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_7            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_8            : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 2 1 1 1 ...
## $ FLAG_DOCUMENT_9            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_10           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_11           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_12           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_13           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_14           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 2 1 1 1 ...
## $ FLAG_DOCUMENT_15           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_16           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_17           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ FLAG_DOCUMENT_18           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

```
##  $ FLAG_DOCUMENT_19           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_20           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_21           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

```
str(glm_test)
```

```
## 'data.frame':    61502 obs. of  71 variables:
##  $ SK_ID_CURR                 : int  100019 100021 100023 100024 100032 100039 100044 100051 100054
##  $ TARGET                     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ NAME_CONTRACT_TYPE         : Factor w/ 2 levels "Cash loans","Revolving loans": 1 2 1 2 1 1 1 1 1
##  $ CODE_GENDER                : Factor w/ 3 levels "F","M","XNA": 2 1 1 2 2 2 2 2 1 1 ...
##  $ FLAG_OWN_CAR               : Factor w/ 2 levels "N","Y": 2 1 1 2 1 2 1 1 1 1 ...
##  $ FLAG_OWN_REALTY            : Factor w/ 2 levels "N","Y": 2 2 2 2 2 1 2 2 2 2 ...
##  $ CNT_CHILDREN               : Factor w/ 15 levels "0","1","2","3",..: 1 2 2 1 2 2 1 1 1 1 ...
##  $ AMT_INCOME_TOTAL           : num  4.94e-06 4.94e-06 4.94e-06 4.94e-06 4.94e-06 ...
##  $ AMT_CREDIT                 : num  5.48 5.43 5.74 5.63 5.51 ...
##  $ AMT_ANNUITY                : num  4.05e-05 4.05e-05 4.05e-05 4.05e-05 4.05e-05 ...
##  $ AMT_GOODS_PRICE            : num  2.85e-06 2.85e-06 2.85e-06 2.85e-06 2.85e-06 ...
##  $ NAME_TYPE_SUITE            : Factor w/ 8 levels "","Children",..: 3 8 8 8 3 8 8 8 8 8 ...
##  $ NAME_INCOME_TYPE           : Factor w/ 8 levels "Businessman",..: 8 8 5 8 8 2 8 8 8 5 ...
##  $ NAME_EDUCATION_TYPE        : Factor w/ 5 levels "Academic degree",..: 5 5 2 5 5 5 5 5 5 2 ...
##  $ NAME_FAMILY_STATUS         : Factor w/ 6 levels "Civil marriage",..: 4 2 4 2 2 2 2 1 2 2 ...
##  $ NAME_HOUSING_TYPE          : Factor w/ 6 levels "Co-op apartment",..: 5 2 2 2 2 2 2 2 2 2 ...
##  $ REGION_POPULATION_RELATIVE : num  53.2 53.2 53.2 53.2 53.2 ...
##  $ DAYS_BIRTH                 : int  -8728 -9776 -11348 -18252 -15948 -11694 -21077 -9827 -20121 -124
##  $ DAYS_EMPLOYED              : num  -0.00157 -0.00157 -0.00157 -0.00157 -0.00157 ...
##  $ DAYS_REGISTRATION          : num  -3494 -4143 -1021 -298 -5782 ...
##  $ DAYS_ID_PUBLISH            : int  -1368 -2427 -3964 -1800 -3153 -3557 -4270 -2380 -3283 -2576 ...
##  $ FLAG_MOBIL                 : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ FLAG_EMP_PHONE             : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ FLAG_WORK_PHONE            : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 2 2 1 ...
##  $ FLAG_CONT_MOBILE           : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 1 ...
##  $ FLAG_PHONE                 : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 2 1 2 1 ...
##  $ FLAG_EMAIL                 : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ OCCUPATION_TYPE            : Factor w/ 19 levels "","Accountants",..: 10 10 5 10 10 6 6 1 13 5 ..
##  $ CNT_FAM_MEMBERS            : num  1 1.73 1.41 1.41 1.73 ...
##  $ REGION_RATING_CLIENT       : Factor w/ 3 levels "1","2","3": 3 2 2 2 2 2 2 2 2 2 ...
##  $ REGION_RATING_CLIENT_W_CITY: Factor w/ 3 levels "1","2","3": 3 2 2 2 2 2 2 2 2 2 ...
##  $ WEEKDAY_APPR_PROCESS_START : Factor w/ 7 levels "FRIDAY","MONDAY",..: 3 2 2 1 3 5 1 7 7 5 ...
##  $ HOUR_APPR_PROCESS_START    : int  6 10 12 13 10 10 10 12 16 9 ...
##  $ REG_REGION_NOT_LIVE_REGION : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ REG_REGION_NOT_WORK_REGION : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ LIVE_REGION_NOT_WORK_REGION: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ REG_CITY_NOT_LIVE_CITY     : Factor w/ 2 levels "0","1": 2 2 1 1 1 2 1 1 1 1 ...
##  $ REG_CITY_NOT_WORK_CITY     : Factor w/ 2 levels "0","1": 2 2 1 1 2 2 2 2 2 1 ...
##  $ LIVE_CITY_NOT_WORK_CITY    : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 2 2 2 1 ...
##  $ ORGANIZATION_TYPE          : Factor w/ 58 levels "Advertising",..: 5 8 29 43 17 43 56 6 6 40 ...
##  $ EXT_SOURCE_2               : num  0.347 0.684 0.587 0.113 0.541 ...
##  $ EXT_SOURCE_3               : num  0.679 0.535 0.478 0.535 0.659 ...
##  $ FONDKAPREMONT_MODE         : Factor w/ 5 levels "","not specified",..: 1 1 1 4 1 1 1 1 1 1 ...
##  $ HOUSETYPE_MODE             : Factor w/ 4 levels "","block of flats",..: 1 1 1 2 1 1 1 1 1 2 ...
##  $ WALLSMATERIAL_MODE         : Factor w/ 8 levels "","Block","Mixed",..: 1 1 1 7 1 1 1 1 1 6 ...
##  $ EMERGENCYSTATE_MODE        : Factor w/ 3 levels "","No","Yes": 1 1 1 2 1 1 1 1 1 2 ...
##  $ OBS_30_CNT_SOCIAL_CIRCLE   : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
##  $ DEF_30_CNT_SOCIAL_CIRCLE   : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
##  $ OBS_60_CNT_SOCIAL_CIRCLE   : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
##  $ DEF_60_CNT_SOCIAL_CIRCLE   : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
##  $ DAYS_LAST_PHONE_CHANGE     : num  -925 -2811 -1850 -296 -2 ...
##  $ FLAG_DOCUMENT_2            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_3            : Factor w/ 2 levels "0","1": 1 1 2 1 2 1 2 2 2 2 ...
##  $ FLAG_DOCUMENT_4            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_5            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_6            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_7            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_8            : Factor w/ 2 levels "0","1": 2 1 1 1 1 2 1 1 1 1 ...
##  $ FLAG_DOCUMENT_9            : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_10           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_11           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_12           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_13           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_14           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_15           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_16           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_17           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_18           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 ...
##  $ FLAG_DOCUMENT_19           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_20           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ FLAG_DOCUMENT_21           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

```r
glm_mod1 = glm(TARGET ~., data = (glm_train%>%select(-CNT_CHILDREN)),
               family = binomial(link='logit'))
summary(glm_mod1)
```

```
##
## Call:
## glm(formula = TARGET ~ ., family = binomial(link = "logit"),
##     data = (glm_train %>% select(-CNT_CHILDREN)))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5947  -0.4339  -0.3183  -0.2314   3.2331
##
## Coefficients: (9 not defined because of singularities)
##                                                  Estimate Std. Error z value
## (Intercept)                                    -2.005e+01  3.400e+02  -0.059
## SK_ID_CURR                                     -7.820e-09  7.479e-08  -0.105
## NAME_CONTRACT_TYPERevolving loans              -6.444e-02  6.662e-02  -0.967
## CODE_GENDERM                                    3.760e-01  2.007e-02  18.733
## CODE_GENDERXNA                                 -9.869e+00  1.611e+02  -0.061
## FLAG_OWN_CARY                                  -2.796e-01  1.833e-02 -15.251
## FLAG_OWN_REALTYY                                6.036e-02  1.768e-02   3.415
## AMT_INCOME_TOTAL                                       NA         NA      NA
## AMT_CREDIT                                      2.590e-01  2.797e-02   9.257
## AMT_ANNUITY                                            NA         NA      NA
## AMT_GOODS_PRICE                                        NA         NA      NA
## NAME_TYPE_SUITEChildren                         3.775e-01  1.652e-01   2.285
## NAME_TYPE_SUITEFamily                           3.424e-01  1.468e-01   2.332
## NAME_TYPE_SUITEGroup of people                  2.441e-01  3.003e-01   0.813
```

```
## NAME_TYPE_SUITEOther_A                          2.483e-01  2.037e-01   1.219
## NAME_TYPE_SUITEOther_B                          4.854e-01  1.725e-01   2.814
## NAME_TYPE_SUITESpouse, partner                  2.961e-01  1.508e-01   1.963
## NAME_TYPE_SUITEUnaccompanied                    3.757e-01  1.454e-01   2.585
## NAME_INCOME_TYPECommercial associate            8.924e+00  1.007e+02   0.089
## NAME_INCOME_TYPEMaternity leave                 1.260e+01  1.007e+02   0.125
## NAME_INCOME_TYPEPensioner                      -8.977e-01  1.504e+02  -0.006
## NAME_INCOME_TYPEState servant                   8.897e+00  1.007e+02   0.088
## NAME_INCOME_TYPEStudent                        -1.360e+00  1.275e+02  -0.011
## NAME_INCOME_TYPEUnemployed                      1.631e+00  1.504e+02   0.011
## NAME_INCOME_TYPEWorking                         9.031e+00  1.007e+02   0.090
## NAME_EDUCATION_TYPEHigher education             1.429e+00  7.204e-01   1.984
## NAME_EDUCATION_TYPEIncomplete higher            1.528e+00  7.213e-01   2.118
## NAME_EDUCATION_TYPELower secondary              1.811e+00  7.229e-01   2.506
## NAME_EDUCATION_TYPESecondary / secondary special  1.731e+00  7.202e-01   2.403
## NAME_FAMILY_STATUSMarried                      -1.770e-01  2.519e-02  -7.025
## NAME_FAMILY_STATUSSeparated                     5.594e-02  4.003e-02   1.397
## NAME_FAMILY_STATUSSingle / not married          1.896e-02  3.453e-02   0.549
## NAME_FAMILY_STATUSUnknown                      -9.015e+00  2.194e+02  -0.041
## NAME_FAMILY_STATUSWidow                        -8.910e-02  4.867e-02  -1.831
## NAME_HOUSING_TYPEHouse / apartment             -1.315e-02  1.282e-01  -0.103
## NAME_HOUSING_TYPEMunicipal apartment            1.660e-01  1.340e-01   1.238
## NAME_HOUSING_TYPEOffice apartment              -1.557e-01  1.562e-01  -0.997
## NAME_HOUSING_TYPERented apartment               3.673e-02  1.382e-01   0.266
## NAME_HOUSING_TYPEWith parents                   3.653e-02  1.314e-01   0.278
## REGION_POPULATION_RELATIVE                             NA         NA      NA
## DAYS_BIRTH                                      2.475e-05  2.659e-06   9.307
## DAYS_EMPLOYED                                          NA         NA      NA
## DAYS_REGISTRATION                               1.120e-05  2.475e-06   4.528
## DAYS_ID_PUBLISH                                 4.890e-05  5.433e-06   9.000
## FLAG_MOBIL1                                     9.186e+00  3.247e+02   0.028
## FLAG_EMP_PHONE1                                -1.414e+00  1.153e+00  -1.226
## FLAG_WORK_PHONE1                                1.332e-01  2.031e-02   6.558
## FLAG_CONT_MOBILE1                              -2.479e-01  1.875e-01  -1.322
## FLAG_PHONE1                                    -9.935e-02  1.907e-02  -5.210
## FLAG_EMAIL1                                    -5.129e-02  3.386e-02  -1.515
## OCCUPATION_TYPEAccountants                     -2.448e-01  5.971e-02  -4.100
## OCCUPATION_TYPECleaning staff                   1.965e-01  6.163e-02   3.189
## OCCUPATION_TYPECooking staff                    9.632e-02  5.575e-02   1.728
## OCCUPATION_TYPECore staff                      -9.922e-02  3.940e-02  -2.518
## OCCUPATION_TYPEDrivers                          1.882e-01  3.734e-02   5.039
## OCCUPATION_TYPEHigh skill tech staff           -1.047e-01  4.971e-02  -2.105
## OCCUPATION_TYPEHR staff                         8.703e-02  1.963e-01   0.443
## OCCUPATION_TYPEIT staff                        -1.430e-01  2.079e-01  -0.688
## OCCUPATION_TYPELaborers                         1.033e-01  2.766e-02   3.734
## OCCUPATION_TYPELow-skill Laborers               3.083e-01  7.263e-02   4.244
## OCCUPATION_TYPEManagers                        -4.230e-02  3.976e-02  -1.064
## OCCUPATION_TYPEMedicine staff                  -7.569e-02  6.749e-02  -1.121
## OCCUPATION_TYPEPrivate service staff           -9.221e-02  9.646e-02  -0.956
## OCCUPATION_TYPERealty agents                    2.233e-02  1.605e-01   0.139
## OCCUPATION_TYPESales staff                      5.538e-02  3.292e-02   1.682
## OCCUPATION_TYPESecretaries                      9.259e-02  1.274e-01   0.727
## OCCUPATION_TYPESecurity staff                   2.086e-01  5.843e-02   3.570
## OCCUPATION_TYPEWaiters/barmen staff             9.105e-03  1.087e-01   0.084
```

```
## CNT_FAM_MEMBERS                                       1.076e-01  3.697e-02   2.910
## REGION_RATING_CLIENT2                                -3.193e-01  1.214e-01  -2.631
## REGION_RATING_CLIENT3                                -2.877e-01  1.244e-01  -2.313
## REGION_RATING_CLIENT_W_CITY2                          4.283e-01  1.173e-01   3.651
## REGION_RATING_CLIENT_W_CITY3                          5.440e-01  1.215e-01   4.479
## WEEKDAY_APPR_PROCESS_STARTMONDAY                     -6.041e-02  2.701e-02  -2.236
## WEEKDAY_APPR_PROCESS_STARTSATURDAY                   -6.267e-02  3.021e-02  -2.074
## WEEKDAY_APPR_PROCESS_STARTSUNDAY                     -7.920e-02  3.869e-02  -2.047
## WEEKDAY_APPR_PROCESS_STARTTHURSDAY                   -7.184e-03  2.679e-02  -0.268
## WEEKDAY_APPR_PROCESS_STARTTUESDAY                     2.888e-02  2.619e-02   1.103
## WEEKDAY_APPR_PROCESS_STARTWEDNESDAY                   4.056e-05  2.660e-02   0.002
## HOUR_APPR_PROCESS_START                              -7.364e-03  2.505e-03  -2.939
## REG_REGION_NOT_LIVE_REGION1                          -1.638e-01  1.066e-01  -1.536
## REG_REGION_NOT_WORK_REGION1                           1.132e-01  1.134e-01   0.999
## LIVE_REGION_NOT_WORK_REGION1                         -1.660e-01  1.128e-01  -1.472
## REG_CITY_NOT_LIVE_CITY1                               1.832e-01  3.970e-02   4.615
## REG_CITY_NOT_WORK_CITY1                              -3.188e-02  4.421e-02  -0.721
## LIVE_CITY_NOT_WORK_CITY1                              3.845e-02  4.288e-02   0.897
## ORGANIZATION_TYPEAgriculture                          1.218e-01  2.381e-01   0.512
## ORGANIZATION_TYPEBank                                -1.845e-01  2.473e-01  -0.746
## ORGANIZATION_TYPEBusiness Entity Type 1              -3.312e-03  2.321e-01  -0.014
## ORGANIZATION_TYPEBusiness Entity Type 2              -1.913e-02  2.293e-01  -0.083
## ORGANIZATION_TYPEBusiness Entity Type 3               1.276e-01  2.260e-01   0.564
## ORGANIZATION_TYPECleaning                             3.821e-01  3.148e-01   1.214
## ORGANIZATION_TYPEConstruction                         2.740e-01  2.300e-01   1.191
## ORGANIZATION_TYPECulture                             -7.805e-02  3.462e-01  -0.225
## ORGANIZATION_TYPEElectricity                         -1.795e-01  2.712e-01  -0.662
## ORGANIZATION_TYPEEmergency                           -9.735e-02  2.929e-01  -0.332
## ORGANIZATION_TYPEGovernment                          -3.003e-02  2.300e-01  -0.131
## ORGANIZATION_TYPEHotel                               -1.762e-01  2.723e-01  -0.647
## ORGANIZATION_TYPEHousing                             -5.820e-02  2.399e-01  -0.243
## ORGANIZATION_TYPEIndustry: type 1                     1.855e-01  2.528e-01   0.734
## ORGANIZATION_TYPEIndustry: type 10                   -4.768e-01  5.380e-01  -0.886
## ORGANIZATION_TYPEIndustry: type 11                    2.474e-02  2.391e-01   0.103
## ORGANIZATION_TYPEIndustry: type 12                   -6.392e-01  3.864e-01  -1.654
## ORGANIZATION_TYPEIndustry: type 13                    1.585e-01  4.565e-01   0.347
## ORGANIZATION_TYPEIndustry: type 2                    -2.070e-01  3.066e-01  -0.675
## ORGANIZATION_TYPEIndustry: type 3                     1.786e-01  2.346e-01   0.761
## ORGANIZATION_TYPEIndustry: type 4                     1.330e-01  2.601e-01   0.511
## ORGANIZATION_TYPEIndustry: type 5                    -2.836e-01  2.924e-01  -0.970
## ORGANIZATION_TYPEIndustry: type 6                    -9.432e-02  4.917e-01  -0.192
## ORGANIZATION_TYPEIndustry: type 7                    -9.352e-02  2.549e-01  -0.367
## ORGANIZATION_TYPEIndustry: type 8                     7.569e-01  6.934e-01   1.092
## ORGANIZATION_TYPEIndustry: type 9                    -2.879e-01  2.394e-01  -1.202
## ORGANIZATION_TYPEInsurance                           -8.936e-02  3.144e-01  -0.284
## ORGANIZATION_TYPEKindergarten                        -2.361e-02  2.326e-01  -0.101
## ORGANIZATION_TYPELegal Services                       5.791e-01  3.260e-01   1.776
## ORGANIZATION_TYPEMedicine                            -4.824e-02  2.326e-01  -0.207
## ORGANIZATION_TYPEMilitary                            -5.220e-01  2.495e-01  -2.092
## ORGANIZATION_TYPEMobile                              -2.924e-01  3.492e-01  -0.837
## ORGANIZATION_TYPEOther                                2.904e-02  2.281e-01   0.127
## ORGANIZATION_TYPEPolice                              -4.434e-01  2.535e-01  -1.749
## ORGANIZATION_TYPEPostal                               1.587e-01  2.425e-01   0.654
## ORGANIZATION_TYPERealtor                              7.005e-01  2.997e-01   2.337
```

```
## ORGANIZATION_TYPEReligion                         4.747e-02  5.322e-01   0.089
## ORGANIZATION_TYPERestaurant                       1.876e-01  2.423e-01   0.775
## ORGANIZATION_TYPESchool                          -1.354e-01  2.317e-01  -0.584
## ORGANIZATION_TYPESecurity                         9.694e-03  2.392e-01   0.041
## ORGANIZATION_TYPESecurity Ministries             -3.804e-01  2.559e-01  -1.486
## ORGANIZATION_TYPESelf-employed                    1.990e-01  2.265e-01   0.879
## ORGANIZATION_TYPEServices                         1.447e-02  2.577e-01   0.056
## ORGANIZATION_TYPETelecom                          1.562e-02  2.925e-01   0.053
## ORGANIZATION_TYPETrade: type 1                    1.371e-01  3.091e-01   0.444
## ORGANIZATION_TYPETrade: type 2                   -3.326e-01  2.506e-01  -1.327
## ORGANIZATION_TYPETrade: type 3                    1.902e-01  2.353e-01   0.808
## ORGANIZATION_TYPETrade: type 4                   -7.520e-01  7.729e-01  -0.973
## ORGANIZATION_TYPETrade: type 5                    7.196e-02  6.544e-01   0.110
## ORGANIZATION_TYPETrade: type 6                   -1.890e-01  3.080e-01  -0.614
## ORGANIZATION_TYPETrade: type 7                    1.638e-01  2.301e-01   0.712
## ORGANIZATION_TYPETransport: type 1               -5.714e-01  4.531e-01  -1.261
## ORGANIZATION_TYPETransport: type 2               -9.786e-02  2.435e-01  -0.402
## ORGANIZATION_TYPETransport: type 3                5.837e-01  2.456e-01   2.377
## ORGANIZATION_TYPETransport: type 4                1.117e-01  2.319e-01   0.482
## ORGANIZATION_TYPEUniversity                      -1.221e-01  2.682e-01  -0.455
## ORGANIZATION_TYPEXNA                              8.480e+00  1.116e+02   0.076
## EXT_SOURCE_2                                     -2.099e+00  3.986e-02 -52.664
## EXT_SOURCE_3                                     -2.715e+00  4.242e-02 -64.011
## FONDKAPREMONT_MODEnot specified                  -1.033e-02  6.300e-02  -0.164
## FONDKAPREMONT_MODEorg spec account               -8.180e-02  6.738e-02  -1.214
## FONDKAPREMONT_MODEreg oper account               -8.342e-03  2.600e-02  -0.321
## FONDKAPREMONT_MODEreg oper spec account          -9.868e-02  4.730e-02  -2.086
## HOUSETYPE_MODEblock of flats                      1.540e-02  6.583e-02   0.234
## HOUSETYPE_MODEspecific housing                    2.163e-01  1.177e-01   1.837
## HOUSETYPE_MODEterraced house                      6.807e-02  1.353e-01   0.503
## WALLSMATERIAL_MODEBlock                          -2.355e-02  7.603e-02  -0.310
## WALLSMATERIAL_MODEMixed                           5.757e-02  1.078e-01   0.534
## WALLSMATERIAL_MODEMonolithic                     -2.318e-01  1.417e-01  -1.635
## WALLSMATERIAL_MODEOthers                          1.535e-01  1.216e-01   1.263
## WALLSMATERIAL_MODEPanel                          -7.987e-02  6.222e-02  -1.284
## WALLSMATERIAL_MODEStone, brick                    6.770e-02  6.127e-02   1.105
## WALLSMATERIAL_MODEWooden                          9.543e-02  7.932e-02   1.203
## EMERGENCYSTATE_MODENo                            -1.051e-01  5.186e-02  -2.026
## EMERGENCYSTATE_MODEYes                           -8.482e-02  1.008e-01  -0.842
## OBS_30_CNT_SOCIAL_CIRCLE                                 NA         NA      NA
## DEF_30_CNT_SOCIAL_CIRCLE                                 NA         NA      NA
## OBS_60_CNT_SOCIAL_CIRCLE                                 NA         NA      NA
## DEF_60_CNT_SOCIAL_CIRCLE                                 NA         NA      NA
## DAYS_LAST_PHONE_CHANGE                            8.374e-05  1.046e-05   8.009
## FLAG_DOCUMENT_21                                  1.453e+00  7.915e-01   1.836
## FLAG_DOCUMENT_31                                  3.938e-01  6.533e-02   6.028
## FLAG_DOCUMENT_41                                 -9.286e+00  6.904e+01  -0.134
## FLAG_DOCUMENT_51                                  3.657e-01  8.395e-02   4.356
## FLAG_DOCUMENT_61                                  3.317e-01  7.577e-02   4.377
## FLAG_DOCUMENT_71                                 -4.339e-03  7.374e-01  -0.006
## FLAG_DOCUMENT_81                                  1.629e-01  7.092e-02   2.297
## FLAG_DOCUMENT_91                                  1.383e-01  1.511e-01   0.915
## FLAG_DOCUMENT_101                                -1.026e+01  1.302e+02  -0.079
## FLAG_DOCUMENT_111                                -1.964e-01  1.471e-01  -1.335
```

```
## FLAG_DOCUMENT_121                              -8.519e+00  3.247e+02  -0.026
## FLAG_DOCUMENT_131                              -8.035e-01  2.035e-01  -3.949
## FLAG_DOCUMENT_141                              -6.744e-01  2.123e-01  -3.176
## FLAG_DOCUMENT_151                              -8.462e-01  3.452e-01  -2.452
## FLAG_DOCUMENT_161                              -5.623e-01  9.895e-02  -5.683
## FLAG_DOCUMENT_171                              -9.089e-01  7.307e-01  -1.244
## FLAG_DOCUMENT_181                              -4.931e-01  1.012e-01  -4.873
## FLAG_DOCUMENT_191                              -6.690e-01  3.965e-01  -1.687
## FLAG_DOCUMENT_201                               2.936e-01  3.315e-01   0.886
## FLAG_DOCUMENT_211                               1.970e-01  3.414e-01   0.577
##                                                Pr(>|z|)
## (Intercept)                                    0.952984
## SK_ID_CURR                                     0.916719
## NAME_CONTRACT_TYPERevolving loans              0.333422
## CODE_GENDERM                                    < 2e-16 ***
## CODE_GENDERXNA                                 0.951141
## FLAG_OWN_CARY                                   < 2e-16 ***
## FLAG_OWN_REALTYY                               0.000638 ***
## AMT_INCOME_TOTAL                                     NA
## AMT_CREDIT                                      < 2e-16 ***
## AMT_ANNUITY                                          NA
## AMT_GOODS_PRICE                                      NA
## NAME_TYPE_SUITEChildren                        0.022330 *
## NAME_TYPE_SUITEFamily                          0.019684 *
## NAME_TYPE_SUITEGroup of people                 0.416272
## NAME_TYPE_SUITEOther_A                         0.222928
## NAME_TYPE_SUITEOther_B                         0.004888 **
## NAME_TYPE_SUITESpouse, partner                 0.049623 *
## NAME_TYPE_SUITEUnaccompanied                   0.009743 **
## NAME_INCOME_TYPECommercial associate           0.929408
## NAME_INCOME_TYPEMaternity leave                0.900438
## NAME_INCOME_TYPEPensioner                      0.995236
## NAME_INCOME_TYPEState servant                  0.929622
## NAME_INCOME_TYPEStudent                        0.991492
## NAME_INCOME_TYPEUnemployed                     0.991347
## NAME_INCOME_TYPEWorking                        0.928569
## NAME_EDUCATION_TYPEHigher education            0.047308 *
## NAME_EDUCATION_TYPEIncomplete higher           0.034149 *
## NAME_EDUCATION_TYPELower secondary             0.012214 *
## NAME_EDUCATION_TYPESecondary / secondary special 0.016241 *
## NAME_FAMILY_STATUSMarried                      2.14e-12 ***
## NAME_FAMILY_STATUSSeparated                    0.162275
## NAME_FAMILY_STATUSSingle / not married         0.583013
## NAME_FAMILY_STATUSUnknown                      0.967231
## NAME_FAMILY_STATUSWidow                        0.067132 .
## NAME_HOUSING_TYPEHouse / apartment             0.918292
## NAME_HOUSING_TYPEMunicipal apartment           0.215571
## NAME_HOUSING_TYPEOffice apartment              0.318816
## NAME_HOUSING_TYPERented apartment              0.790380
## NAME_HOUSING_TYPEWith parents                  0.781059
## REGION_POPULATION_RELATIVE                           NA
## DAYS_BIRTH                                      < 2e-16 ***
## DAYS_EMPLOYED                                        NA
## DAYS_REGISTRATION                              5.97e-06 ***
```

```
## DAYS_ID_PUBLISH                                   < 2e-16 ***
## FLAG_MOBIL1                                      0.977434
## FLAG_EMP_PHONE1                                  0.220360
## FLAG_WORK_PHONE1                                 5.46e-11 ***
## FLAG_CONT_MOBILE1                                0.186067
## FLAG_PHONE1                                      1.89e-07 ***
## FLAG_EMAIL1                                      0.129847
## OCCUPATION_TYPEAccountants                       4.13e-05 ***
## OCCUPATION_TYPECleaning staff                    0.001430 **
## OCCUPATION_TYPECooking staff                     0.084023 .
## OCCUPATION_TYPECore staff                        0.011799 *
## OCCUPATION_TYPEDrivers                           4.68e-07 ***
## OCCUPATION_TYPEHigh skill tech staff             0.035280 *
## OCCUPATION_TYPEHR staff                          0.657543
## OCCUPATION_TYPEIT staff                          0.491498
## OCCUPATION_TYPELaborers                          0.000189 ***
## OCCUPATION_TYPELow-skill Laborers                2.19e-05 ***
## OCCUPATION_TYPEManagers                          0.287410
## OCCUPATION_TYPEMedicine staff                    0.262088
## OCCUPATION_TYPEPrivate service staff             0.339105
## OCCUPATION_TYPERealty agents                     0.889343
## OCCUPATION_TYPESales staff                       0.092525 .
## OCCUPATION_TYPESecretaries                       0.467312
## OCCUPATION_TYPESecurity staff                    0.000357 ***
## OCCUPATION_TYPEWaiters/barmen staff              0.933262
## CNT_FAM_MEMBERS                                  0.003609 **
## REGION_RATING_CLIENT2                            0.008509 **
## REGION_RATING_CLIENT3                            0.020745 *
## REGION_RATING_CLIENT_W_CITY2                     0.000261 ***
## REGION_RATING_CLIENT_W_CITY3                     7.50e-06 ***
## WEEKDAY_APPR_PROCESS_STARTMONDAY                 0.025328 *
## WEEKDAY_APPR_PROCESS_STARTSATURDAY               0.038041 *
## WEEKDAY_APPR_PROCESS_STARTSUNDAY                 0.040630 *
## WEEKDAY_APPR_PROCESS_STARTTHURSDAY               0.788581
## WEEKDAY_APPR_PROCESS_STARTTUESDAY                0.270135
## WEEKDAY_APPR_PROCESS_STARTWEDNESDAY              0.998783
## HOUR_APPR_PROCESS_START                          0.003291 **
## REG_REGION_NOT_LIVE_REGION1                      0.124479
## REG_REGION_NOT_WORK_REGION1                      0.317858
## LIVE_REGION_NOT_WORK_REGION1                     0.141124
## REG_CITY_NOT_LIVE_CITY1                          3.94e-06 ***
## REG_CITY_NOT_WORK_CITY1                          0.470809
## LIVE_CITY_NOT_WORK_CITY1                         0.369868
## ORGANIZATION_TYPEAgriculture                     0.608944
## ORGANIZATION_TYPEBank                            0.455712
## ORGANIZATION_TYPEBusiness Entity Type 1          0.988616
## ORGANIZATION_TYPEBusiness Entity Type 2          0.933504
## ORGANIZATION_TYPEBusiness Entity Type 3          0.572445
## ORGANIZATION_TYPECleaning                        0.224908
## ORGANIZATION_TYPEConstruction                    0.233696
## ORGANIZATION_TYPECulture                         0.821629
## ORGANIZATION_TYPEElectricity                     0.508061
## ORGANIZATION_TYPEEmergency                       0.739638
## ORGANIZATION_TYPEGovernment                      0.896098
```

```
## ORGANIZATION_TYPEHotel                             0.517504
## ORGANIZATION_TYPEHousing                           0.808305
## ORGANIZATION_TYPEIndustry: type 1                  0.463066
## ORGANIZATION_TYPEIndustry: type 10                 0.375425
## ORGANIZATION_TYPEIndustry: type 11                 0.917590
## ORGANIZATION_TYPEIndustry: type 12                 0.098083 .
## ORGANIZATION_TYPEIndustry: type 13                 0.728364
## ORGANIZATION_TYPEIndustry: type 2                  0.499662
## ORGANIZATION_TYPEIndustry: type 3                  0.446601
## ORGANIZATION_TYPEIndustry: type 4                  0.609050
## ORGANIZATION_TYPEIndustry: type 5                  0.332146
## ORGANIZATION_TYPEIndustry: type 6                  0.847878
## ORGANIZATION_TYPEIndustry: type 7                  0.713728
## ORGANIZATION_TYPEIndustry: type 8                  0.275035
## ORGANIZATION_TYPEIndustry: type 9                  0.229198
## ORGANIZATION_TYPEInsurance                         0.776254
## ORGANIZATION_TYPEKindergarten                      0.919176
## ORGANIZATION_TYPELegal Services                    0.075680 .
## ORGANIZATION_TYPEMedicine                          0.835686
## ORGANIZATION_TYPEMilitary                          0.036396 *
## ORGANIZATION_TYPEMobile                            0.402411
## ORGANIZATION_TYPEOther                             0.898668
## ORGANIZATION_TYPEPolice                            0.080259 .
## ORGANIZATION_TYPEPostal                            0.512987
## ORGANIZATION_TYPERealtor                           0.019419 *
## ORGANIZATION_TYPEReligion                          0.928930
## ORGANIZATION_TYPERestaurant                        0.438588
## ORGANIZATION_TYPESchool                            0.558943
## ORGANIZATION_TYPESecurity                          0.967678
## ORGANIZATION_TYPESecurity Ministries               0.137205
## ORGANIZATION_TYPESelf-employed                     0.379554
## ORGANIZATION_TYPEServices                          0.955230
## ORGANIZATION_TYPETelecom                           0.957413
## ORGANIZATION_TYPETrade: type 1                     0.657345
## ORGANIZATION_TYPETrade: type 2                     0.184417
## ORGANIZATION_TYPETrade: type 3                     0.418846
## ORGANIZATION_TYPETrade: type 4                     0.330584
## ORGANIZATION_TYPETrade: type 5                     0.912443
## ORGANIZATION_TYPETrade: type 6                     0.539381
## ORGANIZATION_TYPETrade: type 7                     0.476686
## ORGANIZATION_TYPETransport: type 1                 0.207248
## ORGANIZATION_TYPETransport: type 2                 0.687837
## ORGANIZATION_TYPETransport: type 3                 0.017477 *
## ORGANIZATION_TYPETransport: type 4                 0.630107
## ORGANIZATION_TYPEUniversity                        0.648994
## ORGANIZATION_TYPEXNA                               0.939453
## EXT_SOURCE_2                                        < 2e-16 ***
## EXT_SOURCE_3                                        < 2e-16 ***
## FONDKAPREMONT_MODEnot specified                    0.869779
## FONDKAPREMONT_MODEorg spec account                 0.224770
## FONDKAPREMONT_MODEreg oper account                 0.748285
## FONDKAPREMONT_MODEreg oper spec account            0.036958 *
## HOUSETYPE_MODEblock of flats                       0.814981
## HOUSETYPE_MODEspecific housing                     0.066196 .
```

```
## HOUSETYPE_MODEterraced house                       0.614869
## WALLSMATERIAL_MODEBlock                          0.756762
## WALLSMATERIAL_MODEMixed                          0.593266
## WALLSMATERIAL_MODEMonolithic                     0.101997
## WALLSMATERIAL_MODEOthers                         0.206676
## WALLSMATERIAL_MODEPanel                          0.199256
## WALLSMATERIAL_MODEStone, brick                   0.269166
## WALLSMATERIAL_MODEWooden                         0.228926
## EMERGENCYSTATE_MODENo                            0.042781 *
## EMERGENCYSTATE_MODEYes                           0.400007
## OBS_30_CNT_SOCIAL_CIRCLE                               NA
## DEF_30_CNT_SOCIAL_CIRCLE                               NA
## OBS_60_CNT_SOCIAL_CIRCLE                               NA
## DEF_60_CNT_SOCIAL_CIRCLE                               NA
## DAYS_LAST_PHONE_CHANGE                           1.15e-15 ***
## FLAG_DOCUMENT_21                                 0.066298 .
## FLAG_DOCUMENT_31                                 1.66e-09 ***
## FLAG_DOCUMENT_41                                 0.893009
## FLAG_DOCUMENT_51                                 1.32e-05 ***
## FLAG_DOCUMENT_61                                 1.20e-05 ***
## FLAG_DOCUMENT_71                                 0.995305
## FLAG_DOCUMENT_81                                 0.021635 *
## FLAG_DOCUMENT_91                                 0.359969
## FLAG_DOCUMENT_101                                0.937213
## FLAG_DOCUMENT_111                                0.181959
## FLAG_DOCUMENT_121                                0.979071
## FLAG_DOCUMENT_131                                7.85e-05 ***
## FLAG_DOCUMENT_141                                0.001491 **
## FLAG_DOCUMENT_151                                0.014221 *
## FLAG_DOCUMENT_161                                1.32e-08 ***
## FLAG_DOCUMENT_171                                0.213529
## FLAG_DOCUMENT_181                                1.10e-06 ***
## FLAG_DOCUMENT_191                                0.091538 .
## FLAG_DOCUMENT_201                                0.375883
## FLAG_DOCUMENT_211                                0.563887
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 138034  on 246008  degrees of freedom
## Residual deviance: 123730  on 245832  degrees of freedom
## AIC: 124084
##
## Number of Fisher Scoring iterations: 11
```

```
df_cor <- training %>% mutate_if(is.character, as.factor)
df_cor <- df_cor %>% mutate_if(is.factor, as.numeric)
corr <- cor(df_cor)
```

```
## Warning in cor(df_cor): the standard deviation is zero
```

```r
corr[lower.tri(corr,diag=TRUE)] <- NA

corr[corr == 1] <- NA
corr <- as.data.frame(as.table(corr))
corr <- na.omit(corr)
corr <- subset(corr, abs(Freq) > .5)
corr <- corr[order(-abs(corr$Freq)),]
gmmod2_traing <- glm_train %>% select("SK_ID_CURR","TARGET",corr$Var1)
```

```r
glm_mod2 = glm(TARGET ~., data = gmmod2_traing, family = binomial(link='logit'))
summary(glm_mod2)
```

```
##
## Call:
## glm(formula = TARGET ~ ., family = binomial(link = "logit"),
##     data = gmmod2_traing)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9914  -0.4457  -0.3861  -0.3280   2.7749
##
## Coefficients:
##                                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         -1.833e+00  6.891e-02 -26.596  < 2e-16
## SK_ID_CURR                          -2.138e-08  7.248e-08  -0.295 0.768013
## REGION_RATING_CLIENT2                4.623e-01  3.131e-02  14.767  < 2e-16
## REGION_RATING_CLIENT3                8.308e-01  3.438e-02  24.168  < 2e-16
## HOUSETYPE_MODEblock of flats        -1.253e-01  5.327e-02  -2.353 0.018640
## HOUSETYPE_MODEspecific housing       1.750e-01  1.083e-01   1.616 0.106061
## HOUSETYPE_MODEterraced house        -5.312e-02  1.265e-01  -0.420 0.674432
## WALLSMATERIAL_MODEBlock             -4.419e-02  7.005e-02  -0.631 0.528156
## WALLSMATERIAL_MODEMixed             -4.495e-02  1.017e-01  -0.442 0.658387
## WALLSMATERIAL_MODEMonolithic        -4.072e-01  1.364e-01  -2.985 0.002838
## WALLSMATERIAL_MODEOthers             7.477e-02  1.160e-01   0.645 0.519095
## WALLSMATERIAL_MODEPanel             -1.905e-01  5.573e-02  -3.419 0.000629
## WALLSMATERIAL_MODEStone, brick      -2.944e-02  5.468e-02  -0.538 0.590350
## WALLSMATERIAL_MODEWooden             1.664e-01  7.352e-02   2.263 0.023626
## REG_REGION_NOT_WORK_REGION1          5.127e-02  3.379e-02   1.517 0.129180
## REG_CITY_NOT_WORK_CITY1              1.849e-01  1.823e-02  10.144  < 2e-16
## CNT_CHILDREN1                       -1.650e-02  1.948e-02  -0.847 0.397054
## CNT_CHILDREN2                       -3.838e-02  2.699e-02  -1.422 0.154989
## CNT_CHILDREN3                        3.372e-02  6.331e-02   0.533 0.594338
## CNT_CHILDREN4                        2.688e-01  1.677e-01   1.603 0.108921
## CNT_CHILDREN5                       -3.302e-01  4.641e-01  -0.711 0.476817
## CNT_CHILDREN6                        1.318e+00  5.303e-01   2.486 0.012927
## CNT_CHILDREN7                       -9.230e+00  7.304e+01  -0.126 0.899447
## CNT_CHILDREN8                       -9.321e+00  1.393e+02  -0.067 0.946639
## CNT_CHILDREN10                      -9.268e+00  1.387e+02  -0.067 0.946710
## CNT_CHILDREN11                       1.396e+01  1.970e+02   0.071 0.943512
## CNT_CHILDREN12                      -8.470e+00  1.393e+02  -0.061 0.951510
## CNT_CHILDREN14                      -8.819e+00  1.393e+02  -0.063 0.949508
## CNT_CHILDREN19                      -9.423e+00  1.391e+02  -0.068 0.945990
## FLAG_EMP_PHONE1                      3.359e-02  2.844e-02   1.181 0.237506
```

```
## FONDKAPREMONT_MODEnot specified         -4.529e-03  6.106e-02  -0.074 0.940871
## FONDKAPREMONT_MODEorg spec account      -9.379e-02  6.561e-02  -1.429 0.152871
## FONDKAPREMONT_MODEreg oper account       1.655e-02  2.510e-02   0.659 0.509702
## FONDKAPREMONT_MODEreg oper spec account -7.272e-02  4.597e-02  -1.582 0.113654
## DAYS_BIRTH                               5.791e-05  2.337e-06  24.778  < 2e-16
## NAME_FAMILY_STATUSMarried               -2.333e-01  2.413e-02  -9.669  < 2e-16
## NAME_FAMILY_STATUSSeparated             -3.871e-02  3.659e-02  -1.058 0.290091
## NAME_FAMILY_STATUSSingle / not married  -3.683e-02  2.857e-02  -1.289 0.197268
## NAME_FAMILY_STATUSUnknown               -9.489e+00  1.384e+02  -0.069 0.945343
## NAME_FAMILY_STATUSWidow                 -1.662e-01  4.534e-02  -3.666 0.000247
##
## (Intercept)                             ***
## SK_ID_CURR
## REGION_RATING_CLIENT2                   ***
## REGION_RATING_CLIENT3                   ***
## HOUSETYPE_MODEblock of flats            *
## HOUSETYPE_MODEspecific housing
## HOUSETYPE_MODEterraced house
## WALLSMATERIAL_MODEBlock
## WALLSMATERIAL_MODEMixed
## WALLSMATERIAL_MODEMonolithic            **
## WALLSMATERIAL_MODEOthers
## WALLSMATERIAL_MODEPanel                 ***
## WALLSMATERIAL_MODEStone, brick
## WALLSMATERIAL_MODEWooden                *
## REG_REGION_NOT_WORK_REGION1
## REG_CITY_NOT_WORK_CITY1                 ***
## CNT_CHILDREN1
## CNT_CHILDREN2
## CNT_CHILDREN3
## CNT_CHILDREN4
## CNT_CHILDREN5
## CNT_CHILDREN6                           *
## CNT_CHILDREN7
## CNT_CHILDREN8
## CNT_CHILDREN10
## CNT_CHILDREN11
## CNT_CHILDREN12
## CNT_CHILDREN14
## CNT_CHILDREN19
## FLAG_EMP_PHONE1
## FONDKAPREMONT_MODEnot specified
## FONDKAPREMONT_MODEorg spec account
## FONDKAPREMONT_MODEreg oper account
## FONDKAPREMONT_MODEreg oper spec account
## DAYS_BIRTH                              ***
## NAME_FAMILY_STATUSMarried               ***
## NAME_FAMILY_STATUSSeparated
## NAME_FAMILY_STATUSSingle / not married
## NAME_FAMILY_STATUSUnknown
## NAME_FAMILY_STATUSWidow                 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

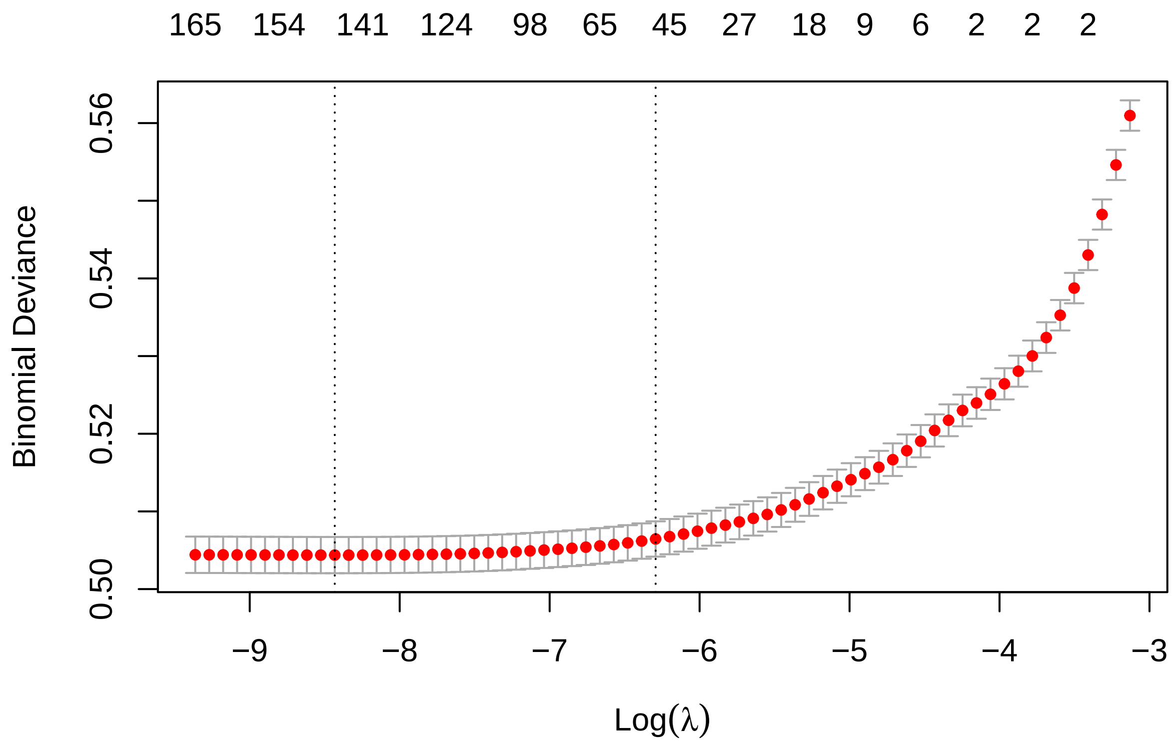
```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 138034  on 246008  degrees of freedom
## Residual deviance: 134984  on 245969  degrees of freedom
## AIC: 135064
##
## Number of Fisher Scoring iterations: 10
```

```
fit_2 = stepAIC(glm_mod2,direction="both",trace=FALSE)
summary(fit_2)
```

```
##
## Call:
## glm(formula = TARGET ~ REGION_RATING_CLIENT + HOUSETYPE_MODE +
##     WALLSMATERIAL_MODE + REG_REGION_NOT_WORK_REGION + REG_CITY_NOT_WORK_CITY +
##     DAYS_BIRTH + NAME_FAMILY_STATUS, family = binomial(link = "logit"),
##     data = gmmod2_traing)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7361  -0.4456  -0.3858  -0.3286   2.7810
##
## Coefficients:
##                                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        -1.804e+00  4.833e-02 -37.320  < 2e-16
## REGION_RATING_CLIENT2               4.599e-01  3.126e-02  14.711  < 2e-16
## REGION_RATING_CLIENT3               8.284e-01  3.435e-02  24.118  < 2e-16
## HOUSETYPE_MODEblock of flats       -1.258e-01  5.283e-02  -2.381 0.017286
## HOUSETYPE_MODEspecific housing      1.766e-01  1.082e-01   1.632 0.102690
## HOUSETYPE_MODEterraced house       -5.386e-02  1.261e-01  -0.427 0.669322
## WALLSMATERIAL_MODEBlock            -4.301e-02  6.945e-02  -0.619 0.535713
## WALLSMATERIAL_MODEMixed            -4.543e-02  1.015e-01  -0.447 0.654560
## WALLSMATERIAL_MODEMonolithic       -4.119e-01  1.361e-01  -3.027 0.002469
## WALLSMATERIAL_MODEOthers            7.806e-02  1.156e-01   0.676 0.499334
## WALLSMATERIAL_MODEPanel            -1.911e-01  5.495e-02  -3.478 0.000505
## WALLSMATERIAL_MODEStone, brick     -2.877e-02  5.412e-02  -0.532 0.594958
## WALLSMATERIAL_MODEWooden            1.674e-01  7.324e-02   2.285 0.022316
## REG_REGION_NOT_WORK_REGION1         5.166e-02  3.378e-02   1.529 0.126146
## REG_CITY_NOT_WORK_CITY1             1.891e-01  1.804e-02  10.479  < 2e-16
## DAYS_BIRTH                          5.855e-05  1.903e-06  30.764  < 2e-16
## NAME_FAMILY_STATUSMarried          -2.357e-01  2.402e-02  -9.814  < 2e-16
## NAME_FAMILY_STATUSSeparated        -3.987e-02  3.656e-02  -1.091 0.275443
## NAME_FAMILY_STATUSSingle / not married -3.488e-02  2.830e-02  -1.232 0.217822
## NAME_FAMILY_STATUSUnknown          -6.474e+00  3.087e+01  -0.210 0.833923
## NAME_FAMILY_STATUSWidow            -1.724e-01  4.509e-02  -3.823 0.000132
##
## (Intercept)                        ***
## REGION_RATING_CLIENT2              ***
## REGION_RATING_CLIENT3              ***
## HOUSETYPE_MODEblock of flats       *
## HOUSETYPE_MODEspecific housing
## HOUSETYPE_MODEterraced house
## WALLSMATERIAL_MODEBlock
## WALLSMATERIAL_MODEMixed
```

```
## WALLSMATERIAL_MODEMonolithic          **
## WALLSMATERIAL_MODEOthers
## WALLSMATERIAL_MODEPanel               ***
## WALLSMATERIAL_MODEStone, brick
## WALLSMATERIAL_MODEWooden              *
## REG_REGION_NOT_WORK_REGION1
## REG_CITY_NOT_WORK_CITY1               ***
## DAYS_BIRTH                            ***
## NAME_FAMILY_STATUSMarried             ***
## NAME_FAMILY_STATUSSeparated
## NAME_FAMILY_STATUSSingle / not married
## NAME_FAMILY_STATUSUnknown
## NAME_FAMILY_STATUSWidow               ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 138034  on 246008  degrees of freedom
## Residual deviance: 135010  on 245988  degrees of freedom
## AIC: 135052
##
## Number of Fisher Scoring iterations: 7
```

```r
x <- model.matrix(TARGET~., glm_train)[,-1]
y <- glm_train$TARGET


set.seed(123)
cv.lasso <- cv.glmnet(x, y, alpha = 1, family = "binomial")
plot(cv.lasso)
```




```
cv.lasso$lambda.min
```

```
## [1] 0.0002175897
```

```
cv.lasso$lambda.1se
```

```
## [1] 0.001848976
```

```
coef(cv.lasso, cv.lasso$lambda.min)
```

```
## 200 x 1 sparse Matrix of class "dgCMatrix"
##                                                          s1
## (Intercept)                                   -1.052129e+00
## SK_ID_CURR                                     .
## NAME_CONTRACT_TYPERevolving loans             -2.009080e-01
## CODE_GENDERM                                   3.662168e-01
## CODE_GENDERXNA                                -1.923957e-01
## FLAG_OWN_CARY                                 -2.692423e-01
## FLAG_OWN_REALTYY                               5.331462e-02
## CNT_CHILDREN1                                  2.873773e-02
## CNT_CHILDREN2                                  9.410856e-03
## CNT_CHILDREN3                                  .
## CNT_CHILDREN4                                  1.506870e-01
## CNT_CHILDREN5                                 -3.167644e-01
```

```
## CNT_CHILDREN6                                          1.175353e+00
## CNT_CHILDREN7                                         -9.598411e-01
## CNT_CHILDREN8                                          .
## CNT_CHILDREN9                                          .
## CNT_CHILDREN10                                         .
## CNT_CHILDREN11                                         4.228448e+00
## CNT_CHILDREN12                                         .
## CNT_CHILDREN14                                        -4.974581e-01
## CNT_CHILDREN19                                        -1.622989e-01
## AMT_INCOME_TOTAL                                       .
## AMT_CREDIT                                             2.277735e-01
## AMT_ANNUITY                                            .
## AMT_GOODS_PRICE                                        .
## NAME_TYPE_SUITEChildren                                1.922974e-03
## NAME_TYPE_SUITEFamily                                  .
## NAME_TYPE_SUITEGroup of people                         .
## NAME_TYPE_SUITEOther_A                                -3.693646e-02
## NAME_TYPE_SUITEOther_B                                 1.118664e-01
## NAME_TYPE_SUITESpouse, partner                        -2.644327e-02
## NAME_TYPE_SUITEUnaccompanied                           3.296225e-02
## NAME_INCOME_TYPECommercial associate                  -1.853955e-02
## NAME_INCOME_TYPEMaternity leave                        3.214440e+00
## NAME_INCOME_TYPEPensioner                              .
## NAME_INCOME_TYPEState servant                         -5.690250e-02
## NAME_INCOME_TYPEStudent                               -1.421899e+00
## NAME_INCOME_TYPEUnemployed                             2.307774e+00
## NAME_INCOME_TYPEWorking                                8.243564e-02
## NAME_EDUCATION_TYPEHigher education                   -1.014613e-01
## NAME_EDUCATION_TYPEIncomplete higher                   .
## NAME_EDUCATION_TYPELower secondary                     2.699475e-01
## NAME_EDUCATION_TYPESecondary / secondary special  2.043229e-01
## NAME_FAMILY_STATUSMarried                             -1.646319e-01
## NAME_FAMILY_STATUSSeparated                            2.653582e-02
## NAME_FAMILY_STATUSSingle / not married                 .
## NAME_FAMILY_STATUSUnknown                              .
## NAME_FAMILY_STATUSWidow                               -8.850975e-02
## NAME_HOUSING_TYPEHouse / apartment                    -4.052257e-02
## NAME_HOUSING_TYPEMunicipal apartment                   1.167156e-01
## NAME_HOUSING_TYPEOffice apartment                     -1.525146e-01
## NAME_HOUSING_TYPERented apartment                      .
## NAME_HOUSING_TYPEWith parents                          .
## REGION_POPULATION_RELATIVE                             .
## DAYS_BIRTH                                             2.330191e-05
## DAYS_EMPLOYED                                          .
## DAYS_REGISTRATION                                      1.049513e-05
## DAYS_ID_PUBLISH                                        4.604994e-05
## FLAG_MOBIL1                                            .
## FLAG_EMP_PHONE1                                        .
## FLAG_WORK_PHONE1                                       1.179342e-01
## FLAG_CONT_MOBILE1                                     -2.254145e-01
## FLAG_PHONE1                                           -9.003652e-02
## FLAG_EMAIL1                                           -3.771003e-02
## OCCUPATION_TYPEAccountants                            -2.318166e-01
## OCCUPATION_TYPECleaning staff                          1.559453e-01
```

```
## OCCUPATION_TYPECooking staff                       6.741613e-02
## OCCUPATION_TYPECore staff                         -1.125774e-01
## OCCUPATION_TYPEDrivers                             1.735727e-01
## OCCUPATION_TYPEHigh skill tech staff              -1.038528e-01
## OCCUPATION_TYPEHR staff                            .
## OCCUPATION_TYPEIT staff                           -6.436511e-02
## OCCUPATION_TYPELaborers                            8.909635e-02
## OCCUPATION_TYPELow-skill Laborers                  2.888274e-01
## OCCUPATION_TYPEManagers                           -4.064118e-02
## OCCUPATION_TYPEMedicine staff                     -7.772085e-02
## OCCUPATION_TYPEPrivate service staff              -6.590116e-02
## OCCUPATION_TYPERealty agents                       .
## OCCUPATION_TYPESales staff                         4.245298e-02
## OCCUPATION_TYPESecretaries                         1.948446e-02
## OCCUPATION_TYPESecurity staff                      1.756127e-01
## OCCUPATION_TYPEWaiters/barmen staff                .
## CNT_FAM_MEMBERS                                    4.650531e-02
## REGION_RATING_CLIENT2                              .
## REGION_RATING_CLIENT3                              .
## REGION_RATING_CLIENT_W_CITY2                       9.807056e-02
## REGION_RATING_CLIENT_W_CITY3                       2.425792e-01
## WEEKDAY_APPR_PROCESS_STARTMONDAY                  -4.736813e-02
## WEEKDAY_APPR_PROCESS_STARTSATURDAY                -4.742228e-02
## WEEKDAY_APPR_PROCESS_STARTSUNDAY                  -5.833469e-02
## WEEKDAY_APPR_PROCESS_STARTTHURSDAY                 .
## WEEKDAY_APPR_PROCESS_STARTTUESDAY                  2.734024e-02
## WEEKDAY_APPR_PROCESS_STARTWEDNESDAY                .
## HOUR_APPR_PROCESS_START                           -6.777278e-03
## REG_REGION_NOT_LIVE_REGION1                       -5.140850e-02
## REG_REGION_NOT_WORK_REGION1                        .
## LIVE_REGION_NOT_WORK_REGION1                      -3.878176e-02
## REG_CITY_NOT_LIVE_CITY1                            1.542643e-01
## REG_CITY_NOT_WORK_CITY1                            .
## LIVE_CITY_NOT_WORK_CITY1                           3.905087e-03
## ORGANIZATION_TYPEAgriculture                       8.113322e-02
## ORGANIZATION_TYPEBank                             -1.623662e-01
## ORGANIZATION_TYPEBusiness Entity Type 1           -2.409183e-03
## ORGANIZATION_TYPEBusiness Entity Type 2           -2.334452e-02
## ORGANIZATION_TYPEBusiness Entity Type 3            1.020766e-01
## ORGANIZATION_TYPECleaning                          2.924495e-01
## ORGANIZATION_TYPEConstruction                      2.430867e-01
## ORGANIZATION_TYPECulture                           .
## ORGANIZATION_TYPEElectricity                      -1.321855e-01
## ORGANIZATION_TYPEEmergency                        -3.431364e-02
## ORGANIZATION_TYPEGovernment                       -2.937487e-02
## ORGANIZATION_TYPEHotel                            -1.278945e-01
## ORGANIZATION_TYPEHousing                          -3.962424e-02
## ORGANIZATION_TYPEIndustry: type 1                  1.309167e-01
## ORGANIZATION_TYPEIndustry: type 10                -2.747923e-01
## ORGANIZATION_TYPEIndustry: type 11                 .
## ORGANIZATION_TYPEIndustry: type 12                -4.921045e-01
## ORGANIZATION_TYPEIndustry: type 13                 2.128244e-02
## ORGANIZATION_TYPEIndustry: type 2                 -1.351979e-01
## ORGANIZATION_TYPEIndustry: type 3                  1.421834e-01
```

```
## ORGANIZATION_TYPEIndustry: type 4            7.171730e-02
## ORGANIZATION_TYPEIndustry: type 5           -2.248036e-01
## ORGANIZATION_TYPEIndustry: type 6            .
## ORGANIZATION_TYPEIndustry: type 7           -6.323587e-02
## ORGANIZATION_TYPEIndustry: type 8            5.202209e-01
## ORGANIZATION_TYPEIndustry: type 9           -2.793160e-01
## ORGANIZATION_TYPEInsurance                  -2.465321e-02
## ORGANIZATION_TYPEKindergarten               -1.430705e-02
## ORGANIZATION_TYPELegal Services              4.632717e-01
## ORGANIZATION_TYPEMedicine                   -4.994890e-02
## ORGANIZATION_TYPEMilitary                   -4.814768e-01
## ORGANIZATION_TYPEMobile                     -1.996029e-01
## ORGANIZATION_TYPEOther                       .
## ORGANIZATION_TYPEPolice                     -3.919621e-01
## ORGANIZATION_TYPEPostal                      1.022029e-01
## ORGANIZATION_TYPERealtor                     6.096131e-01
## ORGANIZATION_TYPEReligion                    .
## ORGANIZATION_TYPERestaurant                  1.473024e-01
## ORGANIZATION_TYPESchool                     -1.258366e-01
## ORGANIZATION_TYPESecurity                    .
## ORGANIZATION_TYPESecurity Ministries        -3.301662e-01
## ORGANIZATION_TYPESelf-employed               1.760219e-01
## ORGANIZATION_TYPEServices                    .
## ORGANIZATION_TYPETelecom                     .
## ORGANIZATION_TYPETrade: type 1               3.383333e-02
## ORGANIZATION_TYPETrade: type 2              -3.593391e-01
## ORGANIZATION_TYPETrade: type 3               1.231077e-01
## ORGANIZATION_TYPETrade: type 4              -4.317348e-01
## ORGANIZATION_TYPETrade: type 5               .
## ORGANIZATION_TYPETrade: type 6              -1.229339e-01
## ORGANIZATION_TYPETrade: type 7               1.288481e-01
## ORGANIZATION_TYPETransport: type 1          -4.030503e-01
## ORGANIZATION_TYPETransport: type 2          -7.213222e-02
## ORGANIZATION_TYPETransport: type 3           5.380731e-01
## ORGANIZATION_TYPETransport: type 4           7.240349e-02
## ORGANIZATION_TYPEUniversity                 -8.063212e-02
## ORGANIZATION_TYPEXNA                         .
## EXT_SOURCE_2                                -2.104646e+00
## EXT_SOURCE_3                                -2.695965e+00
## FONDKAPREMONT_MODEnot specified              .
## FONDKAPREMONT_MODEorg spec account          -4.665433e-02
## FONDKAPREMONT_MODEreg oper account           .
## FONDKAPREMONT_MODEreg oper spec account     -6.993994e-02
## HOUSETYPE_MODEblock of flats                 .
## HOUSETYPE_MODEspecific housing               1.672609e-01
## HOUSETYPE_MODEterraced house                 1.381942e-02
## WALLSMATERIAL_MODEBlock                     -5.596635e-02
## WALLSMATERIAL_MODEMixed                      .
## WALLSMATERIAL_MODEMonolithic                -2.296216e-01
## WALLSMATERIAL_MODEOthers                     5.286349e-02
## WALLSMATERIAL_MODEPanel                     -1.320982e-01
## WALLSMATERIAL_MODEStone, brick               6.780698e-03
## WALLSMATERIAL_MODEWooden                     2.580507e-02
## EMERGENCYSTATE_MODENo                       -4.611869e-02
```

```
## EMERGENCYSTATE_MODEYes                             .
## OBS_30_CNT_SOCIAL_CIRCLE                          .
## DEF_30_CNT_SOCIAL_CIRCLE                          .
## OBS_60_CNT_SOCIAL_CIRCLE                          .
## DEF_60_CNT_SOCIAL_CIRCLE                          .
## DAYS_LAST_PHONE_CHANGE                             7.927012e-05
## FLAG_DOCUMENT_21                                   1.067657e+00
## FLAG_DOCUMENT_31                                   2.361435e-01
## FLAG_DOCUMENT_41                                  -6.273560e-01
## FLAG_DOCUMENT_51                                   1.902798e-01
## FLAG_DOCUMENT_61                                   1.565892e-01
## FLAG_DOCUMENT_71                                   .
## FLAG_DOCUMENT_81                                   .
## FLAG_DOCUMENT_91                                   .
## FLAG_DOCUMENT_101                                 -7.943294e-01
## FLAG_DOCUMENT_111                                 -2.401909e-01
## FLAG_DOCUMENT_121                                  .
## FLAG_DOCUMENT_131                                 -6.792310e-01
## FLAG_DOCUMENT_141                                 -5.534288e-01
## FLAG_DOCUMENT_151                                 -6.578386e-01
## FLAG_DOCUMENT_161                                 -5.216450e-01
## FLAG_DOCUMENT_171                                 -5.031908e-01
## FLAG_DOCUMENT_181                                 -4.544391e-01
## FLAG_DOCUMENT_191                                 -4.687441e-01
## FLAG_DOCUMENT_201                                  8.891515e-02
## FLAG_DOCUMENT_211                                  9.428466e-02
```

```
coef(cv.lasso, cv.lasso$lambda.1se)
```

```
## 200 x 1 sparse Matrix of class "dgCMatrix"
##                                                              s1
## (Intercept)                                       -2.806670e-01
## SK_ID_CURR                                         .
## NAME_CONTRACT_TYPERevolving loans                 -1.922436e-01
## CODE_GENDERM                                       3.192566e-01
## CODE_GENDERXNA                                     .
## FLAG_OWN_CARY                                     -2.038436e-01
## FLAG_OWN_REALTYY                                   .
## CNT_CHILDREN1                                      .
## CNT_CHILDREN2                                      .
## CNT_CHILDREN3                                      .
## CNT_CHILDREN4                                      .
## CNT_CHILDREN5                                      .
## CNT_CHILDREN6                                      .
## CNT_CHILDREN7                                      .
## CNT_CHILDREN8                                      .
## CNT_CHILDREN9                                      .
## CNT_CHILDREN10                                     .
## CNT_CHILDREN11                                     .
## CNT_CHILDREN12                                     .
## CNT_CHILDREN14                                     .
## CNT_CHILDREN19                                     .
## AMT_INCOME_TOTAL                                   .
## AMT_CREDIT                                         6.427235e-02
```

```
## AMT_ANNUITY                                             .
## AMT_GOODS_PRICE                                         .
## NAME_TYPE_SUITEChildren                                 .
## NAME_TYPE_SUITEFamily                                   .
## NAME_TYPE_SUITEGroup of people                          .
## NAME_TYPE_SUITEOther_A                                  .
## NAME_TYPE_SUITEOther_B                                  .
## NAME_TYPE_SUITESpouse, partner                          .
## NAME_TYPE_SUITEUnaccompanied                            .
## NAME_INCOME_TYPECommercial associate                    .
## NAME_INCOME_TYPEMaternity leave                         .
## NAME_INCOME_TYPEPensioner                               .
## NAME_INCOME_TYPEState servant                          -4.343503e-02
## NAME_INCOME_TYPEStudent                                 .
## NAME_INCOME_TYPEUnemployed                              1.127655e+00
## NAME_INCOME_TYPEWorking                                 1.090112e-01
## NAME_EDUCATION_TYPEHigher education                    -1.792103e-01
## NAME_EDUCATION_TYPEIncomplete higher                    .
## NAME_EDUCATION_TYPELower secondary                      4.755390e-02
## NAME_EDUCATION_TYPESecondary / secondary special  1.257300e-01
## NAME_FAMILY_STATUSMarried                              -9.126911e-02
## NAME_FAMILY_STATUSSeparated                             .
## NAME_FAMILY_STATUSSingle / not married                  .
## NAME_FAMILY_STATUSUnknown                               .
## NAME_FAMILY_STATUSWidow                                 .
## NAME_HOUSING_TYPEHouse / apartment                     -6.148316e-03
## NAME_HOUSING_TYPEMunicipal apartment                    .
## NAME_HOUSING_TYPEOffice apartment                       .
## NAME_HOUSING_TYPERented apartment                       .
## NAME_HOUSING_TYPEWith parents                           .
## REGION_POPULATION_RELATIVE                              .
## DAYS_BIRTH                                              1.948549e-05
## DAYS_EMPLOYED                                           .
## DAYS_REGISTRATION                                       5.799131e-06
## DAYS_ID_PUBLISH                                         3.398483e-05
## FLAG_MOBIL1                                             .
## FLAG_EMP_PHONE1                                         .
## FLAG_WORK_PHONE1                                        3.218888e-02
## FLAG_CONT_MOBILE1                                       .
## FLAG_PHONE1                                            -3.876485e-02
## FLAG_EMAIL1                                             .
## OCCUPATION_TYPEAccountants                             -7.214276e-02
## OCCUPATION_TYPECleaning staff                           .
## OCCUPATION_TYPECooking staff                            .
## OCCUPATION_TYPECore staff                              -9.194963e-02
## OCCUPATION_TYPEDrivers                                  1.066439e-01
## OCCUPATION_TYPEHigh skill tech staff                   -6.683599e-03
## OCCUPATION_TYPEHR staff                                 .
## OCCUPATION_TYPEIT staff                                 .
## OCCUPATION_TYPELaborers                                 5.553182e-02
## OCCUPATION_TYPELow-skill Laborers                       1.907327e-01
## OCCUPATION_TYPEManagers                                 .
## OCCUPATION_TYPEMedicine staff                           .
## OCCUPATION_TYPEPrivate service staff                    .
```

```
## OCCUPATION_TYPERealty agents                          .
## OCCUPATION_TYPESales staff                            1.245754e-02
## OCCUPATION_TYPESecretaries                            .
## OCCUPATION_TYPESecurity staff                         2.844960e-02
## OCCUPATION_TYPEWaiters/barmen staff                   .
## CNT_FAM_MEMBERS                                       .
## REGION_RATING_CLIENT2                                 .
## REGION_RATING_CLIENT3                                 .
## REGION_RATING_CLIENT_W_CITY2                          .
## REGION_RATING_CLIENT_W_CITY3                          1.130147e-01
## WEEKDAY_APPR_PROCESS_STARTMONDAY                      .
## WEEKDAY_APPR_PROCESS_STARTSATURDAY                    .
## WEEKDAY_APPR_PROCESS_STARTSUNDAY                      .
## WEEKDAY_APPR_PROCESS_STARTTHURSDAY                    .
## WEEKDAY_APPR_PROCESS_STARTTUESDAY                     .
## WEEKDAY_APPR_PROCESS_STARTWEDNESDAY                   .
## HOUR_APPR_PROCESS_START                              -2.716607e-03
## REG_REGION_NOT_LIVE_REGION1                           .
## REG_REGION_NOT_WORK_REGION1                           .
## LIVE_REGION_NOT_WORK_REGION1                          .
## REG_CITY_NOT_LIVE_CITY1                               1.236092e-01
## REG_CITY_NOT_WORK_CITY1                               .
## LIVE_CITY_NOT_WORK_CITY1                              .
## ORGANIZATION_TYPEAgriculture                          .
## ORGANIZATION_TYPEBank                                 .
## ORGANIZATION_TYPEBusiness Entity Type 1               .
## ORGANIZATION_TYPEBusiness Entity Type 2               .
## ORGANIZATION_TYPEBusiness Entity Type 3               4.715546e-02
## ORGANIZATION_TYPECleaning                             .
## ORGANIZATION_TYPEConstruction                         1.174336e-01
## ORGANIZATION_TYPECulture                              .
## ORGANIZATION_TYPEElectricity                          .
## ORGANIZATION_TYPEEmergency                            .
## ORGANIZATION_TYPEGovernment                           .
## ORGANIZATION_TYPEHotel                                .
## ORGANIZATION_TYPEHousing                              .
## ORGANIZATION_TYPEIndustry: type 1                     .
## ORGANIZATION_TYPEIndustry: type 10                    .
## ORGANIZATION_TYPEIndustry: type 11                    .
## ORGANIZATION_TYPEIndustry: type 12                    .
## ORGANIZATION_TYPEIndustry: type 13                    .
## ORGANIZATION_TYPEIndustry: type 2                     .
## ORGANIZATION_TYPEIndustry: type 3                     .
## ORGANIZATION_TYPEIndustry: type 4                     .
## ORGANIZATION_TYPEIndustry: type 5                     .
## ORGANIZATION_TYPEIndustry: type 6                     .
## ORGANIZATION_TYPEIndustry: type 7                     .
## ORGANIZATION_TYPEIndustry: type 8                     .
## ORGANIZATION_TYPEIndustry: type 9                    -6.085933e-02
## ORGANIZATION_TYPEInsurance                            .
## ORGANIZATION_TYPEKindergarten                         .
## ORGANIZATION_TYPELegal Services                       .
## ORGANIZATION_TYPEMedicine                             .
## ORGANIZATION_TYPEMilitary                            -1.738150e-01
```

```
## ORGANIZATION_TYPEMobile                                .
## ORGANIZATION_TYPEOther                                 .
## ORGANIZATION_TYPEPolice                               -4.904423e-02
## ORGANIZATION_TYPEPostal                                .
## ORGANIZATION_TYPERealtor                               .
## ORGANIZATION_TYPEReligion                              .
## ORGANIZATION_TYPERestaurant                            .
## ORGANIZATION_TYPESchool                                .
## ORGANIZATION_TYPESecurity                              .
## ORGANIZATION_TYPESecurity Ministries                   .
## ORGANIZATION_TYPESelf-employed                         1.236608e-01
## ORGANIZATION_TYPEServices                              .
## ORGANIZATION_TYPETelecom                               .
## ORGANIZATION_TYPETrade: type 1                         .
## ORGANIZATION_TYPETrade: type 2                        -8.696439e-02
## ORGANIZATION_TYPETrade: type 3                         .
## ORGANIZATION_TYPETrade: type 4                         .
## ORGANIZATION_TYPETrade: type 5                         .
## ORGANIZATION_TYPETrade: type 6                         .
## ORGANIZATION_TYPETrade: type 7                         .
## ORGANIZATION_TYPETransport: type 1                     .
## ORGANIZATION_TYPETransport: type 2                     .
## ORGANIZATION_TYPETransport: type 3                     3.001306e-01
## ORGANIZATION_TYPETransport: type 4                     .
## ORGANIZATION_TYPEUniversity                            .
## ORGANIZATION_TYPEXNA                                   .
## EXT_SOURCE_2                                          -2.130511e+00
## EXT_SOURCE_3                                          -2.577454e+00
## FONDKAPREMONT_MODEnot specified                        .
## FONDKAPREMONT_MODEorg spec account                     .
## FONDKAPREMONT_MODEreg oper account                     .
## FONDKAPREMONT_MODEreg oper spec account                .
## HOUSETYPE_MODEblock of flats                          -5.329646e-03
## HOUSETYPE_MODEspecific housing                         .
## HOUSETYPE_MODEterraced house                           .
## WALLSMATERIAL_MODEBlock                                .
## WALLSMATERIAL_MODEMixed                                .
## WALLSMATERIAL_MODEMonolithic                           .
## WALLSMATERIAL_MODEOthers                               .
## WALLSMATERIAL_MODEPanel                               -9.647601e-02
## WALLSMATERIAL_MODEStone, brick                         .
## WALLSMATERIAL_MODEWooden                               .
## EMERGENCYSTATE_MODENo                                 -4.293573e-02
## EMERGENCYSTATE_MODEYes                                 .
## OBS_30_CNT_SOCIAL_CIRCLE                               .
## DEF_30_CNT_SOCIAL_CIRCLE                               .
## OBS_60_CNT_SOCIAL_CIRCLE                               .
## DEF_60_CNT_SOCIAL_CIRCLE                               .
## DAYS_LAST_PHONE_CHANGE                                 5.259883e-05
## FLAG_DOCUMENT_21                                       .
## FLAG_DOCUMENT_31                                       1.608374e-01
## FLAG_DOCUMENT_41                                       .
## FLAG_DOCUMENT_51                                       .
## FLAG_DOCUMENT_61                                       .
```

```
## FLAG_DOCUMENT_71                                             .
## FLAG_DOCUMENT_81                                             .
## FLAG_DOCUMENT_91                                             .
## FLAG_DOCUMENT_101                                            .
## FLAG_DOCUMENT_111                                            .
## FLAG_DOCUMENT_121                                            .
## FLAG_DOCUMENT_131                                           -6.454024e-02
## FLAG_DOCUMENT_141                                            .
## FLAG_DOCUMENT_151                                            .
## FLAG_DOCUMENT_161                                           -1.823875e-01
## FLAG_DOCUMENT_171                                            .
## FLAG_DOCUMENT_181                                           -1.546132e-01
## FLAG_DOCUMENT_191                                            .
## FLAG_DOCUMENT_201                                            .
## FLAG_DOCUMENT_211                                            .
```

```
lasso.model <- glmnet(x, y, alpha = 1, family = "binomial",
                      lambda = cv.lasso$lambda.1se,exact=FALSE)

# Make prediction on test data
x.test <- model.matrix(TARGET ~., glm_test)[,-1]
probabilities <- lasso.model %>% predict(newx = x.test,type='response')
predicted.classes <- ifelse(probabilities >= 0.5, 1, 0)
# Model accuracy
observed.classes <- glm_test$TARGET
glmnet_accuracy <- mean(predicted.classes == observed.classes)

## Mod1 accuracy
glm1_probs <- glm_mod1 %>% predict(glm_test,type='response')
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from a rank-deficient fit may be misleading
```

```
glm1_predicted.classes <- ifelse(glm1_probs >= 0.5, 1, 0)
glm1observed.classes <- glm_test$TARGET
glmmod1_accuracy <- mean(glm1_predicted.classes == glm1observed.classes)


## Mod2 accuracy
glm2_probs <- fit_2 %>% predict(glm_test,type='response')
glm2_predicted.classes <- ifelse(glm2_probs >= 0.5, 1, 0)
glm2observed.classes <- glm_test$TARGET
glmmod2_accuracy <- mean(glm2_predicted.classes == glm2observed.classes)
```

```
confusionMatrix(data=as.factor(predicted.classes), glm_test$TARGET)
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction     0     1
##          0 56525  4947
##          1    12    18
```

```
##
##                Accuracy : 0.9194
##                  95% CI : (0.9172, 0.9215)
##     No Information Rate : 0.9193
##     P-Value [Acc > NIR] : 0.4684
##
##                   Kappa : 0.0062
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.999788
##             Specificity : 0.003625
##          Pos Pred Value : 0.919524
##          Neg Pred Value : 0.600000
##              Prevalence : 0.919271
##          Detection Rate : 0.919076
##    Detection Prevalence : 0.999512
##       Balanced Accuracy : 0.501707
##
##        'Positive' Class : 0
##
```

```r
confusionMatrix(data=as.factor(glm1_predicted.classes), glm_test$TARGET)
```

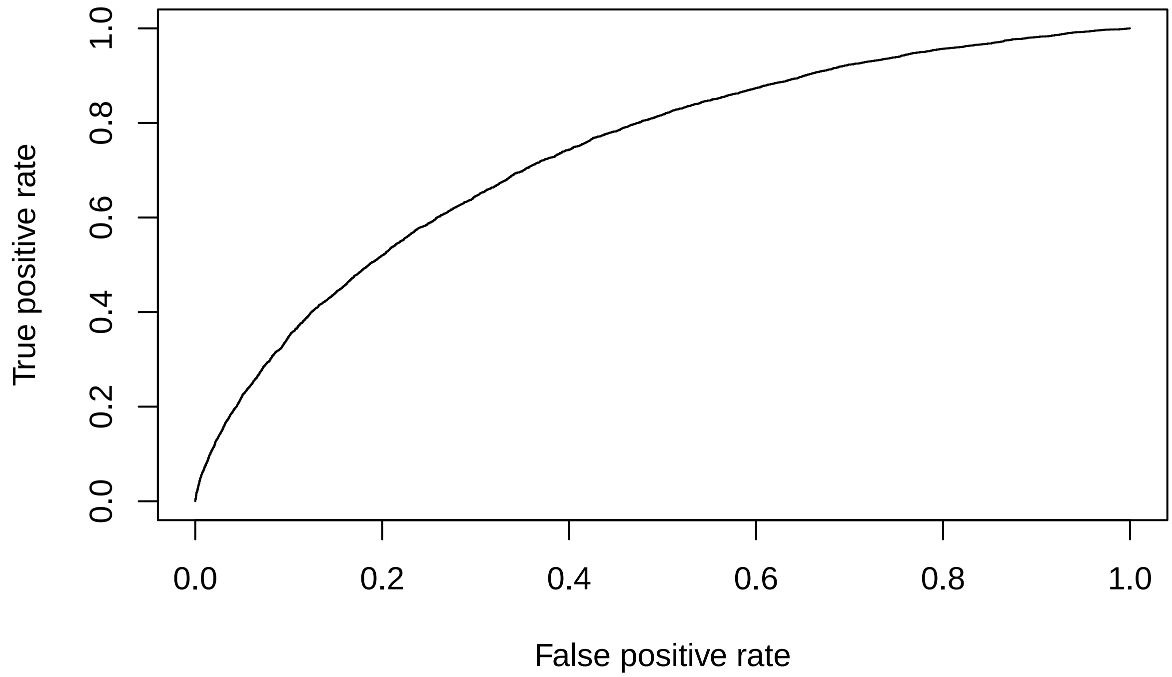
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction     0     1
##          0 56493  4912
##          1    44    53
##
##                Accuracy : 0.9194
##                  95% CI : (0.9172, 0.9216)
##     No Information Rate : 0.9193
##     P-Value [Acc > NIR] : 0.4507
##
##                   Kappa : 0.0179
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.99922
##             Specificity : 0.01067
##          Pos Pred Value : 0.92001
##          Neg Pred Value : 0.54639
##              Prevalence : 0.91927
##          Detection Rate : 0.91856
##    Detection Prevalence : 0.99842
##       Balanced Accuracy : 0.50495
##
##        'Positive' Class : 0
##
```

```r
confusionMatrix(data=as.factor(glm2_predicted.classes), glm_test$TARGET)
```

```
## Warning in confusionMatrix.default(data = as.factor(glm2_predicted.classes), :
## Levels are not in the same order for reference and data. Refactoring data to
## match.
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction     0     1
##          0 56537  4965
##          1     0     0
##
##                Accuracy : 0.9193
##                  95% CI : (0.9171, 0.9214)
##     No Information Rate : 0.9193
##     P-Value [Acc > NIR] : 0.5038
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 1.0000
##             Specificity : 0.0000
##          Pos Pred Value : 0.9193
##          Neg Pred Value :    NaN
##              Prevalence : 0.9193
##          Detection Rate : 0.9193
##    Detection Prevalence : 1.0000
##       Balanced Accuracy : 0.5000
##
##        'Positive' Class : 0
##
```

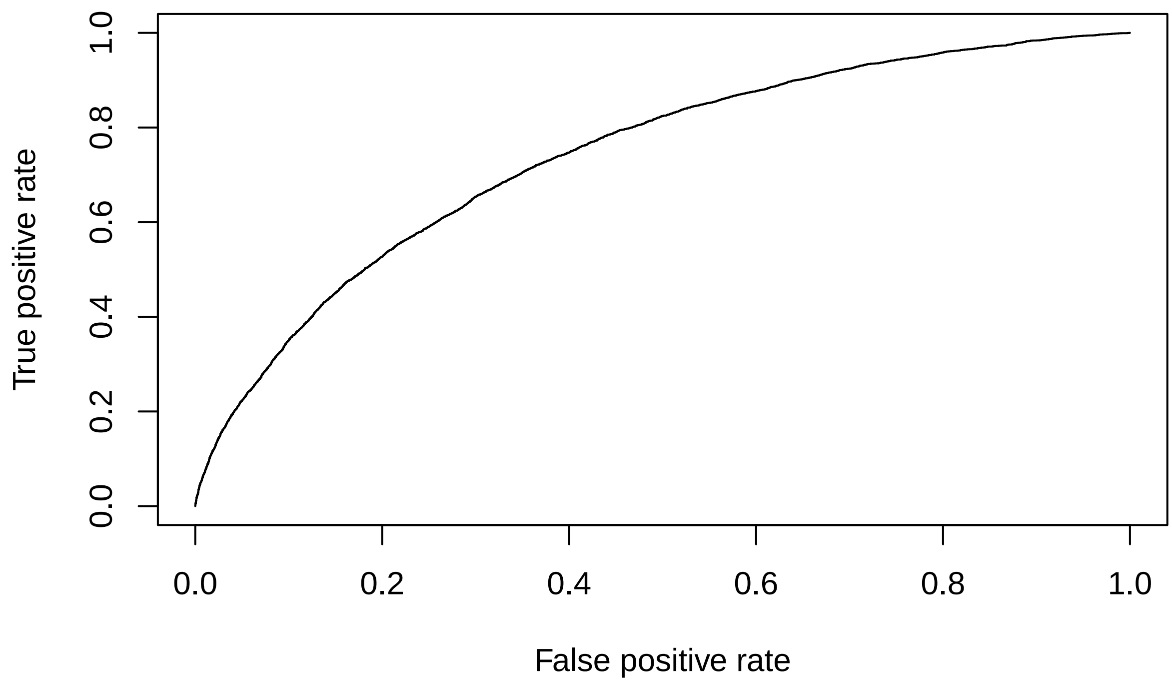
```r
glm_pred_one <- prediction(probabilities, glm_test$TARGET)
glm_perf_one <- performance(glm_pred_one, measure = "tpr", x.measure = "fpr")
glm_auc_mod1 <- performance(glm_pred_one, measure = "auc")
glm_auc_mod1 <- glm_auc_mod1@y.values[[1]]
plot(glm_perf_one)
```




```
print(glm_auc_mod1)
```

```
## [1] 0.7360613
```

```
glm_pred_two <- prediction(glm1_probs, glm_test$TARGET)
glm_perf_two <- performance(glm_pred_two, measure = "tpr", x.measure = "fpr")
glm_auc_mod2 <- performance(glm_pred_two, measure = "auc")
glm_auc_mod2 <- glm_auc_mod2@y.values[[1]]
plot(glm_perf_two)
```




```
print(glm_auc_mod2)
```

```
## [1] 0.7398852
```