Data Analysis

library(Amelia)

## Loading required package: Rcpp

## ##   
## ## Amelia II: Multiple Imputation  
## ## (Version 1.8.0, built: 2021-05-26)  
## ## Copyright (C) 2005-2021 James Honaker, Gary King and Matthew Blackwell  
## ## Refer to http://gking.harvard.edu/amelia/ for more information  
## ##

#import libraries  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.0 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(tidyr)  
library(dplyr)  
library(forecast)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library(broom)  
library(ggplot2)  
library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(mltools)

##   
## Attaching package: 'mltools'

## The following object is masked from 'package:tidyr':  
##   
## replace\_na

library(data.table)

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

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

library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

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

library(fitdistrplus)

## Loading required package: MASS

##   
## Attaching package: 'MASS'

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

## Loading required package: survival

##   
## Attaching package: 'survival'

## The following object is masked from 'package:caret':  
##   
## cluster

library(Hmisc)

## Loading required package: Formula

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':  
##   
## src, summarize

## The following objects are masked from 'package:base':  
##   
## format.pval, units

library(readr)  
data2 <- read\_csv("data.csv", col\_types = cols(enddate = col\_date(format = "%d/%m/%Y"),   
 startdate = col\_date(format = "%d/%m/%Y")))  
  
data2$tokenNum = ifelse(is.na(data2$tokenNum),ave(data2$tokenNum, FUN = function(x) mean(x, na.rm = 'TRUE')),data2$tokenNum)  
  
data\_x = data2 %>%  
 dplyr::mutate(logtokennumber = log(tokenNum),   
 categorycount = sapply(as.list(strsplit(categories,",")),length),   
 duration = as.numeric(enddate - startdate),  
 currencycount = sapply(as.list(strsplit(acceptingCurrency,",")),length)  
 )

sapply(data\_x,function(x) sum(is.na(x)))

## id success tokenNum teamSize   
## 0 0 0 0   
## country categories overallrating offered\_ownership   
## 0 0 0 541   
## enddate startdate tokenName tokenPrice   
## 0 5 32 181   
## tokenType platform acceptingCurrency softcap   
## 458 131 227 29   
## hardcap whitepaper video socialMedia   
## 29 29 29 29   
## logtokennumber categorycount duration currencycount   
## 0 0 5 0

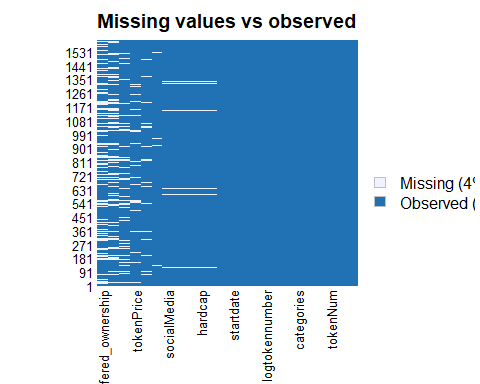
sapply(data\_x, function(x) length(unique(x)))

## id success tokenNum teamSize   
## 1606 2 481 50   
## country categories overallrating offered\_ownership   
## 117 696 41 98   
## enddate startdate tokenName tokenPrice   
## 372 372 1493 1422   
## tokenType platform acceptingCurrency softcap   
## 44 64 235 3   
## hardcap whitepaper video socialMedia   
## 3 3 3 5   
## logtokennumber categorycount duration currencycount   
## 481 16 144 12

missmap(data\_x, main = "Missing values vs observed")

## Warning: Unknown or uninitialised column: `arguments`.  
  
## Warning: Unknown or uninitialised column: `arguments`.

## Warning: Unknown or uninitialised column: `imputations`.



data\_x = data\_x[!is.na(data\_x$video),]  
data\_x = data\_x[!is.na(data\_x$duration),]  
sapply(data\_x,function(x) sum(is.na(x)))

## id success tokenNum teamSize   
## 0 0 0 0   
## country categories overallrating offered\_ownership   
## 0 0 0 530   
## enddate startdate tokenName tokenPrice   
## 0 0 32 178   
## tokenType platform acceptingCurrency softcap   
## 454 129 222 0   
## hardcap whitepaper video socialMedia   
## 0 0 0 0   
## logtokennumber categorycount duration currencycount   
## 0 0 0 0

modeldata = data\_x %>%  
 dplyr::select(-c(offered\_ownership, tokenName, tokenType, platform, acceptingCurrency, categories, tokenPrice, startdate, enddate,country ))  
modeldata = modeldata[!is.infinite(rowSums(modeldata)),]

model <- glm(success ~.,family=binomial(link='logit'),data=modeldata)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(model)

##   
## Call:  
## glm(formula = success ~ ., family = binomial(link = "logit"),   
## data = modeldata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.52380 -0.78653 0.01956 0.71573 2.23556   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.656e+00 5.319e-01 -4.993 5.94e-07 \*\*\*  
## id -5.969e-05 1.492e-04 -0.400 0.6891   
## tokenNum 4.386e-13 2.217e-13 1.978 0.0479 \*   
## teamSize 2.074e-02 1.140e-02 1.819 0.0690 .   
## overallrating 1.184e+00 1.336e-01 8.868 < 2e-16 \*\*\*  
## softcap -1.905e-01 1.573e-01 -1.211 0.2260   
## hardcap 9.920e-02 1.842e-01 0.538 0.5903   
## whitepaper -9.378e-01 3.834e-01 -2.446 0.0144 \*   
## video 3.346e-01 1.665e-01 2.009 0.0445 \*   
## socialMedia 4.474e-01 7.515e-02 5.954 2.62e-09 \*\*\*  
## logtokennumber -8.305e-03 1.826e-02 -0.455 0.6492   
## categorycount -7.271e-02 3.910e-02 -1.860 0.0629 .   
## duration -1.851e-03 2.462e-03 -0.752 0.4521   
## currencycount 4.201e-02 4.893e-02 0.859 0.3905   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1848.7 on 1393 degrees of freedom  
## Residual deviance: 1224.2 on 1380 degrees of freedom  
## AIC: 1252.2  
##   
## Number of Fisher Scoring iterations: 14

anova(model, test="Chisq")

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: success  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 1393 1848.7   
## id 1 0.046 1392 1848.7 0.829324   
## tokenNum 1 265.343 1391 1583.3 < 2.2e-16 \*\*\*  
## teamSize 1 142.216 1390 1441.1 < 2.2e-16 \*\*\*  
## overallrating 1 156.849 1389 1284.3 < 2.2e-16 \*\*\*  
## softcap 1 3.747 1388 1280.5 0.052904 .   
## hardcap 1 0.092 1387 1280.4 0.761288   
## whitepaper 1 6.782 1386 1273.7 0.009207 \*\*   
## video 1 3.493 1385 1270.2 0.061643 .   
## socialMedia 1 41.111 1384 1229.0 1.438e-10 \*\*\*  
## logtokennumber 1 0.127 1383 1228.9 0.721383   
## categorycount 1 3.589 1382 1225.3 0.058154 .   
## duration 1 0.414 1381 1224.9 0.520107   
## currencycount 1 0.737 1380 1224.2 0.390517   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

library(pscl)

## Classes and Methods for R developed in the  
## Political Science Computational Laboratory  
## Department of Political Science  
## Stanford University  
## Simon Jackman  
## hurdle and zeroinfl functions by Achim Zeileis

pR2(model)

## fitting null model for pseudo-r2

## llh llhNull G2 McFadden r2ML r2CU   
## -612.0890773 -924.3625504 624.5469462 0.3378258 0.3611113 0.4916315