Retail Analytics
Group 1
Final Report
Phuong Nguyen, William Fehmi, Qiutong Shi

# Kickstarter: Predicting Project Success on Crowdfunding

## 1. Introduction

Kickstarter is an intersection between a crowdfunding platform and a retailer. As opposed to the stability of a traditional retail platform, Kickstarter presents significant risks in consumer's purchases/backing of products that are not yet produced. If consumers are willing to back a product that does not yet exist, without definite guarantee that it will be produced, then projects that get backed on Kickstarter are likely an indication of future trends. Project categories and types that succeed on Kickstarter can be indicative of where physical and online retailers are lacking in products that currently exist, or where demand is unmet. Small companies and upcoming start-ups often use kickstarter as both a way to gain funding for a project, but also gauge interest in a product/idea/project. Retailers can use kickstarter trends as another avenue to determine what to anticipate in the future.

On an individual level, upcoming Kickstarter project developers would be able to use this information to determine what products they should develop in the future, based on the success rate of others in their field. Using specific terms, as well as catering toward a specific audience may impact the success rate of a project.

The dataset could be used to find what contributes to a "successful" kickstarter versus a "failed" kickstarter. What "keywords" shows up that greenlight to people that this is a company that they want to back?

## 2. Data Summary

We obtained data from a website that has an AI that scrapes information from all of kickstarter's category and subcategory pages monthly, then exports them into CSV and JSON formats and posts them online (Web Robots 2022). The original dataset contains 1,278,789 observations, with over 40 variables of interest.

### 2.1 Variables Description

backers_count – The amount of backers for each kickstarter, or how many individuals have contributed money to the project.

blurb – A brief description of the kickstarter (we could use this information to create a word cloud of what successful vs unsuccessful kickstarters contain)

category – What section the product falls into, examples include “webtoon” “product design” “comedy” “cookbook” “design” etc.

converted_pleged_amount – Pledged amount in USD

Country- Country that kickstarter project creator is from

Country_displayable_name - Full country name that kickstarter campaign is based in

Created_at – date and time of when the project was initially created on Kickstarter

Creator – Creator’s kickstarter user ID

Currency – original currency the project goal was denominated in

Currency_symbol – symbol of the original currency the project goal was denominated in

Currency_trailing_code –code of the original currency the project goal was denominated in

Current_currency – currency the project goal was converted to

Deadline – date and time of when the project will close for donations

Disable_communication – whether or not a project owner disabled communication with their backers

Friends – this column is completely blank

Fx_rate – foreign exchange rate of currency

Goal – fundraising goal of project in USD

Id- kickstarter id of project

Is_backing – this row is blank

Is_starable - whether or not a project can be starred (liked and saved) by users

Is_starred - whether or not a project has been starred (liked and saved) by users (blank)

Launched_at – date and time of when the project was launched for funding

Location – contains the town or city of the project creator

Name – name of project

Permissions - this row completely blank

Photo – asset ID of photo used for campaign

Pledged – amount currently pledged

Profile – profile name & includes project name, last active time, and if profile is active or inactive

Slug- a version of the name with no spaces, used for better coding

Source_url – url for the project's category

Spotlight –after a project has been successful, it is spotlighted on the Kickstarter website

Staff-pick – whether a project was highlighted as a staff_pick when it was launched/live

State - whether a project was successful, failed, canceled, suspending or still live

State_changed_at – date and time of when a project's status was changed (same as the deadline for successful and failed projects)

Static_usd_rate – conversion rate between the original currency and USD

Urls - same as other url just more complex for no reason

Usd_exchange_rate – dynamic usd exchange rate, not static

Usd-pledged – amount of money pledged in USD

Usd_type – international/domestic

### 2.2 Data Cleaning Processes and Rationale

The goal of our data cleaning processes was to ensure that we remove all missing values, reduce the number of variables that are not helpful to us, check for multicollinearity between independent variables, and prepare our data for analysis. First, we removed country_displayable_name, created_at, currency_symbol, creator, currency_trailing_code, launched_at, urls, deadline, current_currency, location, profile, photo, state_changed_at, usd_type, fx_rate, pledged. Country_displayable_name, currency_symbol, currency_trailing_code, current_currency, usd_type, fx_rate, location are all highly correlated and provide little information on the dependent variable, thus we remove them in favor of country. Launched_at, deadline, state_changed_at are variables pertaining to project lengths and launch time, which is beyond the scope of our focus and will be removed. Similarly, profile and photo pertains to image analysis, which is beyond the scope of our focus and thus removed. Because the dataset was pulled from AI web scrapers monthly, a lot of the projects are repeated and thus will not be helpful to us. Based on project IDs, we removed all repeated projects from the dataset.

In checking for multicollinearity, we decided to create a new data frame consisting of non-repeated numeric variables. We then generated a correlation plot (Appendix 6.1) and found that converted_pledged_amount and backers_count are shown to be positively strongly correlated. We would keep that in mind when running regressions. Additionally, we want to understand whether there exists strong associations between categorical variables through the chi squared test. The sets to be tested include: country and currency, spotlight and staff_picked, spotlight and state. The "state" column has four kinds of variables: canceled, live, failed, successful, therefore it is an ordinal variable with a clear ordering of things. However, the other categorical variables remain nominal. Therefore, the test we selected is the Cramer's V test. The Cramer's V value between country and currency is 0.971, indicating that there is a strong association. The test does not return perfect association because some European countries use the same currency. This can be verified in the contingency table (Appendix 6.1). The Cramer's V value between spotlight and staff pick is 0.2535, indicating a weak association. The Cramer's V value between spotlight and state is 1, indicating a perfect association. Only successful projects are spotlight projects. We can remove the spotlight altogether as a decision variable.

We then visualize and explicitly show quantile and outliers of pledged_amouint, goal, and backers count to observe potential outliers (Appendix 6.1). By looking at the box plots, we found that there are many outliers in pledged_amouint, goal, and backers. The outliers would stop us from having useful information on distribution. Therefore we decided to clean the outliers based on IQR criterion.

We run some initial data overview of the variables left to determine their relevance in our analysis. For the variable disable_communication, we found that 99.7% of project owners did not disable communication with their backers, which is to be expected. Thus, this variable will not give us much useful insight and will be removed from our analysis. For the purpose of our evaluation, we are only interested in "Successful" and "Failed." We removed all projects that were "Canceled" and "Live." The initial database had Category as a JSON variable. We extract Category and Subcategory from the variable and create two new columns in the dataset.

The final cleaned dataset contains 192,456 observations and 15 variables, which are backers_count, blurb, category, sub_category, converted_pledged_amount, country, goal, id, is_starrable, name, staff_pick, state, static_usd_rate, usd_pledged.

## 3. Methodologies

### 3.1 Exploratory Data Analysis

Before generating predictive models for whether or not a project will succeed, we conduct some overall exploratory analysis of the cleaned dataset to observe the differences between a failed project and a successful project. We perform a summary of the data (see *Appendix 6.2*). From our dataset, we found that out of 192,546 observations, 77143 projects failed and 115403 were successful, representing a success rate of 59.93%. The mean project fundraising goal is \$50,846, while the mean amount pledged per project is \$14,799. The mean amount pledged per successful project is \$23,982. The mean amount pledged per failed project is \$1,061. The mean number of backers per project is 162.4. The mean number of backers per successful project is 263.05 and the mean number of backers per failed project is 11.74.

To observe the difference, we first look at the difference between project goals in failed and successful projects. Failed projects have a median goal of \$7,500, while successful projects have a median goal of \$3,500 (half of that of failed projects). Unsurprisingly, successful projects tend to have smaller and more realistic goals.

Additionally, the median amount pledged per successful project is \$5,052, while the median amount pledged per failed project is \$59. This number is a little surprising, as the difference is considerably higher than the median amount requested of \$1,740. This may suggest that once successful projects have met their goals, they tend to gain even more funding from Kickstarter's platform because people are more likely to fund a project with a lot of existing fundings.

There are 176 project categories, of which Music is the most common type of projects, followed by Film & Video, Art, and Technology. Technology projects have the highest goals, followed by Food, yet have one of the lowest amounts of fundings on Kickstarter. Games, Comics, Dance, and Design Projects overall receive the greatest amount of funding on average. The projects that are most frequently successful are Comics and Dance, while the least successful projects are Food, Journalism, and Technology. This can be attributed to their large funding goals. Dance and Film & Video attracts the most generous backers, while Comics and Games attracts the most backers but pledge less money to the project.

In terms of country, 68.73% of projects come from the U.S. We generate a comparison plot between the 25 countries (*see Appendix 6.2*). Overall, projects from Hong Kong have the highest success rate of 79.01%, receive more per backers and attract a higher pledged amount than other countries.

### 3.2 Logistic Regression

In order to predict whether a project is successful or not, we developed a logistic regression framework to understand the variables' effect. First, we set a random seed and split the training and testing data with 80/20. We then ran a full logistic regression model (Appendix 6.1) on variables backers_count, category, converted_pledged_amount, goal, staff_picked. The full model has a null deviance of 207378, and a residual deviance of 97304. To overcome potential overfitting, we decided to use stepwise selection(Appendix 6.2)to find a simpler model from the library "MASS". After selection, the stepwise model uses variables backers_count, category, converted_pledged_amount, and goal to predict success.

With both models in hand, we tested the true positive rate of each model with the same testing dataset. The full model has an accuracy of 0.8909634. The stepwise model has an accuracy of 0.891275. Although the performance is similar, the stepwise selection reduced the complexity of the model without compromising its accuracy. Therefore, we choose the simpler model, here the final model is returned by the stepwise regression.

The model to predict whether a project is successful or not based on backers_count, category, converted_pledged_amount, goal.

### 3.3 Random Forest

We created a classification tree in R to determine what decisions impact the success rate of kickstarter campaigns, and how they compound on each other. First, we set a random seed then created an 80/20 split for randomized samples of test and training data. We then plot a fitted classification tree to predict "state" (the variable that contains success/fail) using rpart package. Next we used the R output to create a visual classification tree using photoshop (*Appendix 6.3.1*). The fitted tree uses category, sub-category, and staff pick as variables, this is after excluding numerical values such as goal, pledged amount, and backers count, with those in a category within categories 2, which contains categories such as (*Appendix 6.3.2*) with a subcategory in subcategories 3, that is not featured as a staff pick being the most likely to have a successful Kickstarter campaign with a success rate of 71.5% (95/112), meanwhile those in the same category and subcategory but were featured as staff picks had the lowest success rate at 0% (0/15). Those in categories 1, with a subcategory within subcategories 1, that were not a staff pick were the second most successful with a success rate of 69.5% (659/860). All other combinations of categories, subcategories, and staff picks were unlikely to be successful. Based on a confusion matrix of the rpart results (*Appendix 6.3.3*), this classification tree has an accuracy of 81.1%, with a misclassification rate of 18.9%, precision of 92.5%, and recall of 75.4%.

We then ran a Random Forest to predict "state" using the same conditions as the classification tree (*Appendix 6.3.4)*. By creating a random forest of 500 trees we are able to improve the performance of the model to an accuracy of 85.4%, misclassification of 14.6%, precision of 85.5%, and precision of 91.7%, based on calculations done on the random forest's confusion matrix (*Appendix 6.3.5*).

There were limitations and challenges to using the Random Forest method using this dataset. The cleaned dataset contains over 190,000 observations per category, which is above the processing power of our personal computers. This resulted in us having to test smaller randomized samples than initially intended, otherwise the R result would never load. This may negatively impact the data collected through this method due to insufficient sample size relative to the actual dataset.

### 3.4 Word Clouds

To explore the effects of Project Names on the probability of success of a Kickstarter Campaign, we perform text mining analysis on the Name variable in successful projects. Because of the size of the data, we use a random sample of 10% of the dataset. We first remove all special characters from the observations and transform them into a vector with text corpus. We remove all punctuations, numbers, popular stop words such as "the," words that provide no additional meaning to Kickstarter such as "project," stem words, and strip all of the white space from the text corpus. We then begin the analysis process by building a term-document matrix. We generate a frequency table with the top 25 most used words for successful projects, which are Album, New, Film, Pin, Art, Book, Music, Enamel, Short, Card, Game, World,
Help, Debut, Play, First, Make, Record, Seri, Collect, Love, Video, Food, Life, Story. From this, we generate a Word Cloud for 25 most used words in successful projects. We found that successful project names are highly correlated with category, i.e. if a project name clearly states the type of projects that it is, it is more likely to be successful. Adjectives that correlate with successful projects are "New," "Debut," "Short," which suggests that including them might

increase the success rate for each project (although further analysis is required). Similarly, the verbs are “Help,” “Play,” “Make,” “Record,” and “Collect.”

To determine the actual level of influence these words have on the probability of success, further analysis is required. Because the names of projects are highly related to project categories, this variable may not be indicative of whether or not a project will succeed.

## 4. Summary

On an individual level, Kickstarter developers who wish to optimize their success on the platform should focus on development within Kickstarter’s more favorable project categories such as Comics and Dance. Technology, Food, and Journalism developers receive the lowest success rate. Additionally, they should set a realistic goal of around $2,000. Country of origin is also a factor to consider, with projects from Hong Kong having the highest success rate of 79.01%. Furthermore, for projects to receive a high success rate, Kickstarter developers should clearly state the project category/project types in their Project Names. Descriptive adjectives indicating novelty such as “New,” “Debut,” and “Short,” and creative verbs such as “Help,” “Play,” “Make,” “Record,” and “Collect” correlates with successful projects.

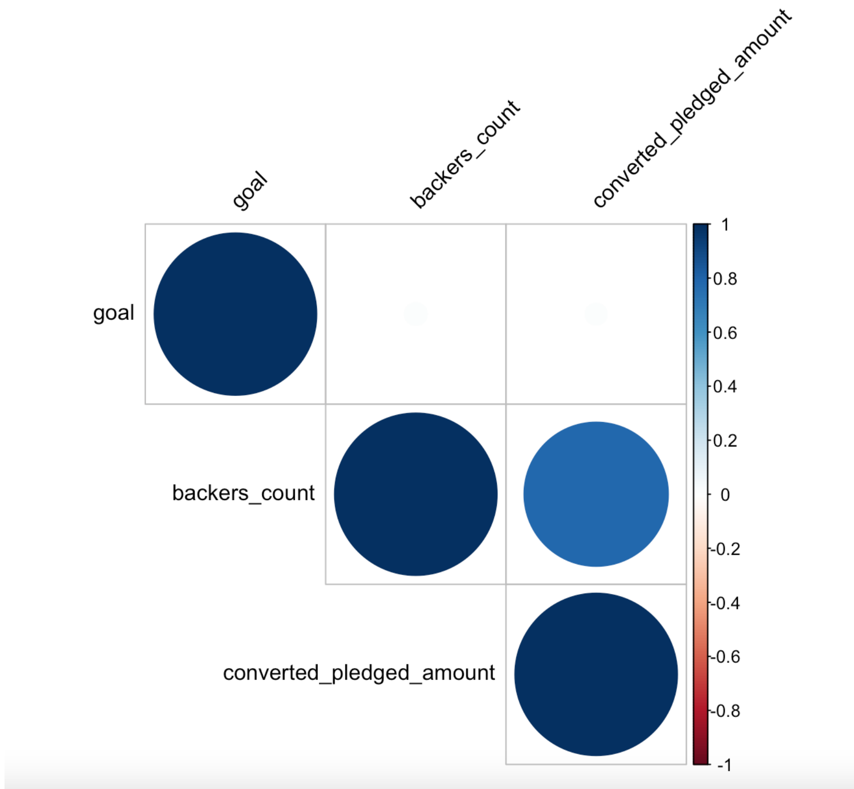
In terms of predicting whether or not a project will succeed, it seems like categories remain the most important variables from both our Logistic Regression Model and the Random Forest model. The final logistic model to predict whether a project is successful or not is based on backers_count, category, converted_pledged_amount, and goal.

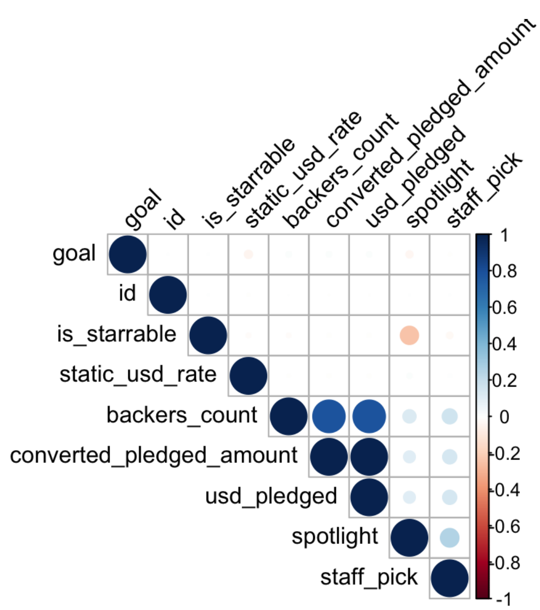
## 5. References

*Kickstarter datasets*. Web Robots. (2022, February 14). Retrieved March 7, 2022, from https://webrobots.io/kickstarter-datasets/

# 6. Appendix

## 6.1 Data Cleaning

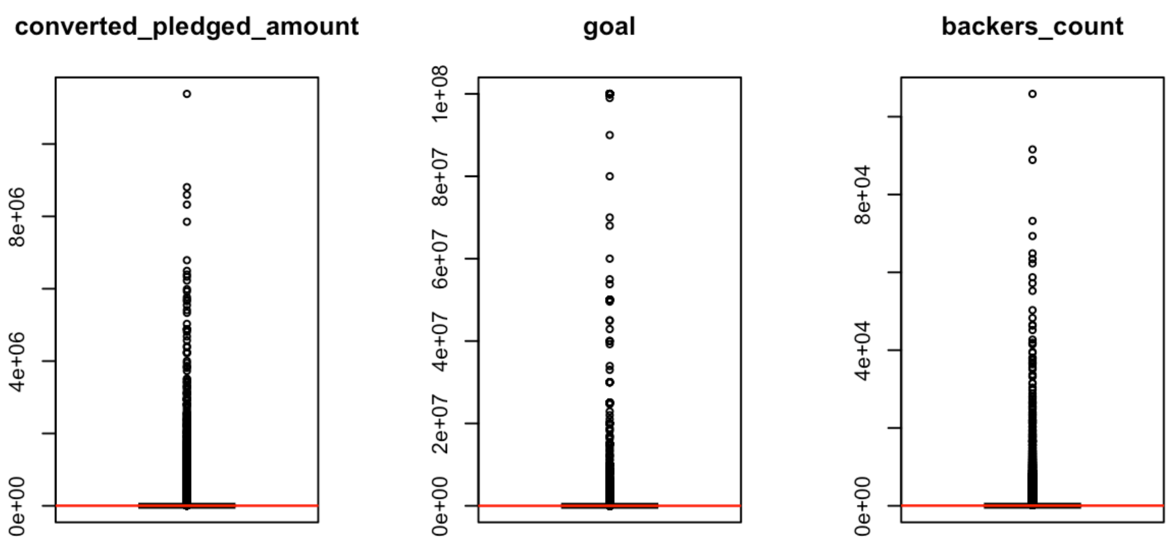


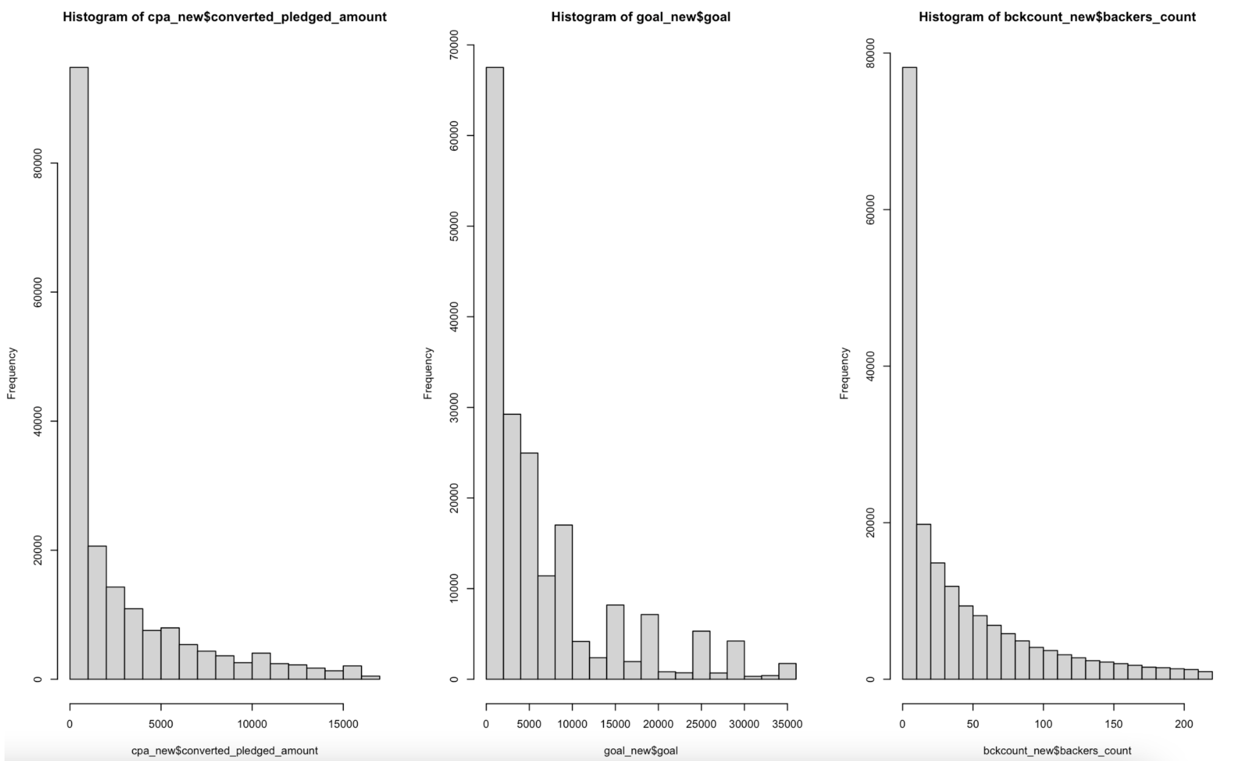

*Correlations*

| | AUD | CAD | CHF | DKK | EUR | GBP | HKD | JPY | MXN | NOK | NZD | PLN | SEK | SGD | USD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AT | 0 | 0 | 0 | 0 | 569 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AU | 5080 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| BE | 0 | 0 | 0 | 0 | 675 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CA | 0 | 10208 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CH | 0 | 0 | 789 | 0 | 39 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| DE | 0 | 0 | 0 | 0 | 4124 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| DK | 0 | 0 | 0 | 923 | 55 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ES | 0 | 0 | 0 | 0 | 2688 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| FR | 0 | 0 | 0 | 0 | 3335 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GB | 0 | 0 | 0 | 0 | 0 | 24094 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GR | 0 | 0 | 0 | 0 | 97 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HK | 0 | 0 | 0 | 0 | 0 | 0 | 1729 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| IE | 0 | 0 | 0 | 0 | 669 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| IT | 0 | 0 | 0 | 0 | 3029 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| JP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 763 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LU | 0 | 0 | 0 | 0 | 72 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MX | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3242 | 0 | 0 | 0 | 0 | 0 | 0 |
| NL | 0 | 0 | 0 | 0 | 1915 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| NO | 0 | 0 | 0 | 0 | 18 | 0 | 0 | 0 | 0 | 488 | 0 | 0 | 0 | 0 | 0 |
| NZ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 930 | 0 | 0 | 0 | 0 |
| PL | 0 | 0 | 0 | 0 | 95 | 0 | 0 | 0 | 0 | 0 | 0 | 65 | 0 | 0 | 0 |
| SE | 0 | 0 | 0 | 0 | 87 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1494 | 0 | 0 |
| SG | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 978 | 0 |
| SI | 0 | 0 | 0 | 0 | 50 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| US | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 143961 |

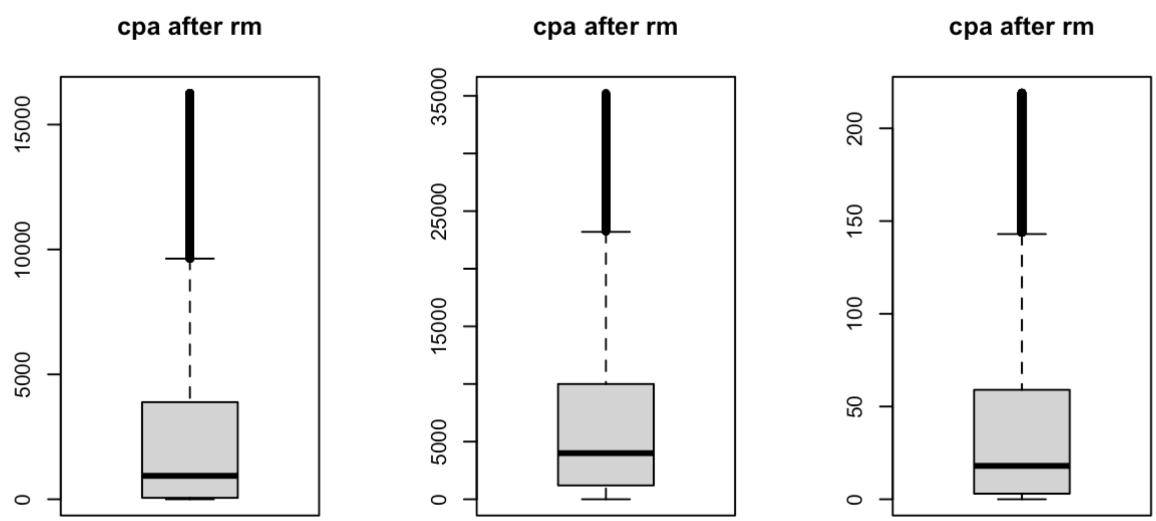
*Contingency table*




*Box plot before handling outlier*




*Distribution after handling outlier*




*Box plot after handling outlier*

## 6.2. Data Exploration

### Data Summary

```
(Other)
                             :192430
        category                sub_category      converted_pledged_amount    country
 music       : 13010                      :102688   Min.   :       0         US     :132353
 film & video: 12866   tabletop games   :  2116   1st Qu.:     136         GB     : 21829
 art         : 10509   hardware         :  2012   Median :    1738         CA     :  9033
 technology  : 10447   comic books      :  1765   Mean   :   14803         AU     :  4490
 publishing  :  9626   web              :  1724   3rd Qu.:    7173         DE     :  3569
 games       :  7747   children's books:  1653   Max.   :11385449         FR     :  2924
 (Other)     :128341   (Other)          : 80588                            (Other): 18348
      goal                  id             is_starrable                         name
 Min.   :        0   Min.   :1.852e+04   Min.   :0     #NAME?                     :    19
 1st Qu.:     1500   1st Qu.:5.364e+08   1st Qu.:0     Debut Album                :     7
 Median :     5000   Median :1.073e+09   Median :0     Home                       :     7
 Mean   :    50846   Mean   :1.073e+09   Mean   :0     A Midsummer Night's Dream:     6
 3rd Qu.:    15000   3rd Qu.:1.609e+09   3rd Qu.:0     Music Video                :     5
 Max.   :100000000   Max.   :2.147e+09   Max.   :0     Reflections                :     5
                                                       (Other)                    :192497
   staff_pick            state        static_usd_rate      usd_pledged          success
 Min.   :0.0000   failed    : 77143   Min.   :0.008706   Min.   :       0   Min.   :0.0000
 1st Qu.:0.0000   successful:115403   1st Qu.:1.000000   1st Qu.:     136   1st Qu.:0.0000
 Median :0.0000                       Median :1.000000   Median :    1740   Median :1.0000
 Mean   :0.1357                       Mean   :1.002498   Mean   :   14799   Mean   :0.5994
 3rd Qu.:0.0000                       3rd Qu.:1.000000   3rd Qu.:    7184   3rd Qu.:1.0000
 Max.   :1.0000                       Max.   :1.716408   Max.   :11385449   Max.   :1.0000
```

| Country | Number of Projects | Success | Success Rate | Number of Backers | Mean USD pledged | USD pledged per backers |
|---|---|---|---|---|---|---|
| HK | 1439 | 1137 | 79.01% | 1439 | 50026.68 | 34.76489229 |
| SI | 23 | 18 | 78.26% | 23 | 75151.738 | 3267.46687 |
| PL | 73 | 51 | 69.86% | 73 | 208224.051 | 2852.38426 |
| GR | 43 | 30 | 69.77% | 43 | 50372.626 | 1171.456419 |
| JP | 637 | 441 | 69.23% | 637 | 32453.842 | 50.94794662 |
| SG | 810 | 539 | 66.54% | 810 | 12311.748 | 15.19968889 |
| GB | 21829 | 14072 | 64.46% | 21829 | 10261.31 | 0.470076962 |
| LU | 64 | 39 | 60.94% | 64 | 13875.661 | 216.8072031 |
| US | 132353 | 80636 | 60.92% | 132353 | 15800.547 | 0.119381858 |
| DK | 873 | 518 | 59.34% | 873 | 10241.432 | 11.73130813 |
| FR | 2924 | 1729 | 59.13% | 2924 | 17216.857 | 5.888117989 |

| | | | | | | |
|---|---|---|---|---|---|---|
| CA | 9033 | 5291 | 58.57% | 9033 | 10135.062 | 1.122003985 |
| NZ | 846 | 472 | 55.79% | 846 | 7689.426 | 9.089156028 |
| SE | 1408 | 785 | 55.75% | 1408 | 15523.516 | 11.02522443 |
| AU | 4490 | 2428 | 54.08% | 4490 | 9511.001 | 2.118263029 |
| DE | 3569 | 1788 | 50.10% | 3569 | 16114.028 | 4.514998039 |
| ES | 2265 | 1118 | 49.36% | 2265 | 12467.698 | 5.504502428 |
| NO | 452 | 223 | 49.34% | 452 | 7405.846 | 16.38461504 |
| IE | 589 | 290 | 49.24% | 589 | 8139.535 | 13.81924448 |
| CH | 705 | 340 | 48.23% | 705 | 22515.811 | 31.93732057 |
| BE | 600 | 288 | 48.00% | 600 | 16945.531 | 28.24255167 |
| NL | 1686 | 805 | 47.75% | 1686 | 15024.424 | 8.911283511 |
| AT | 489 | 211 | 43.15% | 489 | 20567.527 | 42.06038241 |
| MX | 2805 | 1206 | 42.99% | 2805 | 2721.046 | 0.970069875 |
| IT | 2541 | 946 | 37.23% | 2541 | 10563.051 | 4.157044864 |

## 6.2 Logistic Regression

### 6.2.1 Full Model

```
full.model <- glm(success~backers_count+category+converted_pledged_amount+goal+
                  staff_pick, data = train, family = binomial)

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

```
summary(full.model)
```

```
##
## Call:
## glm(formula = success ~ backers_count + category + converted_pledged_amount +
##     goal + staff_pick, family = binomial, data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -8.4904  -0.5466   0.0000   0.3460   8.4904
##
## Coefficients:
##                              Estimate Std. Error  z value Pr(>|z|)
## (Intercept)                -2.968e+00  2.590e-01  -11.457  < 2e-16 ***
## backers_count               4.575e-02  4.060e-04  112.690  < 2e-16 ***
## categoryAcademic            9.326e-01  3.031e-01    3.077 0.002093 **
```

```
## categoryAccessories         2.030e+01  3.306e+02    0.061 0.951040
## categoryAction              7.123e-01  3.300e-01    2.158 0.030913 *
## categoryAnimals             6.873e-01  3.679e-01    1.868 0.061749 .
## categoryAnimation           9.891e-01  2.827e-01    3.499 0.000468 ***
## categoryAnthologies         1.302e+00  3.077e-01    4.231 2.33e-05 ***
## categoryApparel             2.026e+01  2.988e+02    0.068 0.945949
## categoryApps                7.282e-01  2.892e-01    2.518 0.011798 *
## categoryArchitecture        3.955e-01  3.432e-01    1.152 0.249170
## categoryart                 2.162e+00  2.603e-01    8.305  < 2e-16 ***
## categoryArt                 2.046e+01  3.564e+02    0.057 0.954220
## categoryArt Books           2.344e+00  2.914e-01    8.045 8.65e-16 ***
## categoryAudio               1.208e+00  2.934e-01    4.116 3.85e-05 ***
## categoryBacon               9.204e-01  3.687e-01    2.496 0.012553 *
## categoryBlues               1.617e+00  3.723e-01    4.344 1.40e-05 ***
## categoryCalendars           1.359e+00  3.496e-01    3.886 0.000102 ***
## categoryCamera Equipment    1.327e-01  4.246e-01    0.313 0.754650
## categoryCandles             5.860e-01  3.061e-01    1.915 0.055545 .
## categoryCeramics            1.499e+00  2.993e-01    5.008 5.51e-07 ***
## categoryChildren's Books    1.967e+01  3.470e+02    0.057 0.954792
## categoryChildrenswear       4.700e-01  3.390e-01    1.386 0.165661
## categoryChiptune            3.033e+00  7.684e-01    3.947 7.91e-05 ***
## categoryCivic Design        4.610e-01  4.280e-01    1.077 0.281438
## categoryClassical Music     2.848e+00  2.694e-01   10.573  < 2e-16 ***
## categoryComedy              1.674e+00  2.655e-01    6.307 2.84e-10 ***
## categoryComic Books         1.927e+01  3.312e+02    0.058 0.953593
## categorycomics              2.546e+00  2.697e-01    9.440  < 2e-16 ***
## categoryComics              1.928e+01  8.467e+02    0.023 0.981838
## categoryCommunity Gardens   9.706e-01  3.239e-01    2.997 0.002726 **
## categoryConceptual Art      1.670e+00  2.707e-01    6.171 6.79e-10 ***
## categoryCookbooks           4.120e-01  3.044e-01    1.353 0.175915
## categoryCountry & Folk      1.955e+01  3.266e+02    0.060 0.952266
## categoryCouture             8.708e-01  3.937e-01    2.212 0.026987 *
## categorycrafts              1.910e+00  2.637e-01    7.244 4.36e-13 ***
## categoryCrafts              2.082e+01  5.167e+02    0.040 0.967852
## categoryCrochet             1.402e+00  4.535e-01    3.091 0.001994 **
## categorydance               2.419e+00  2.763e-01    8.757  < 2e-16 ***
## categoryDance               2.047e+01  3.251e+02    0.063 0.949780
## categorydesign              1.347e+00  2.654e-01    5.075 3.88e-07 ***
## categoryDesign              2.328e+01  3.898e+02    0.060 0.952379
## categoryDigital Art         1.723e+00  2.739e-01    6.291 3.15e-10 ***
## categoryDIY                 9.246e-01  2.762e-01    3.348 0.000814 ***
## categoryDIY Electronics     8.714e-01  3.118e-01    2.795 0.005196 **
## categoryDocumentary         1.978e+01  2.268e+02    0.087 0.930512
## categoryDrama               2.072e+00  2.738e-01    7.567 3.82e-14 ***
## categoryDrinks              1.060e+00  2.785e-01    3.807 0.000140 ***
## categoryElectronic Music    1.311e+00  2.764e-01    4.744 2.09e-06 ***
## categoryEmbroidery          9.170e-01  4.659e-01    1.968 0.049060 *
## categoryEvents              8.479e-01  2.857e-01    2.967 0.003003 **
## categoryExperimental        1.677e+00  2.950e-01    5.687 1.29e-08 ***
## categoryFabrication Tools   4.903e-01  4.820e-01    1.017 0.309050
## categoryFaith               1.397e+00  2.924e-01    4.778 1.77e-06 ***
```

```
## categoryFamily              8.171e-01  3.778e-01    2.163 0.030560 *
## categoryFantasy             1.460e+00  3.416e-01    4.274 1.92e-05 ***
## categoryFarmer's Markets    9.243e-01  3.464e-01    2.668 0.007622 **
## categoryFarms               8.968e-01  2.942e-01    3.049 0.002299 **
## categoryfashion             2.215e+00  2.621e-01    8.451  < 2e-16 ***
## categoryFashion             1.995e+01  1.351e+03    0.015 0.988216
## categoryFestivals           2.052e+00  3.008e-01    6.821 9.07e-12 ***
## categoryFiction             2.056e+01  2.270e+02    0.091 0.927841
## categoryfilm & video        1.929e+00  2.600e-01    7.418 1.19e-13 ***
## categoryFilm & Video        2.069e+01  1.073e+03    0.019 0.984613
## categoryFine Art            1.585e+00  2.816e-01    5.627 1.84e-08 ***
## categoryFlight              9.770e-02  4.600e-01    0.212 0.831820
## categoryfood                1.881e-01  2.628e-01    0.716 0.474180
## categoryFood                2.207e+01  3.022e+02    0.073 0.941762
## categoryFood Trucks         1.606e-01  2.959e-01    0.543 0.587192
## categoryFootwear            8.751e-02  3.112e-01    0.281 0.778549
## categoryGadgets             8.540e-01  3.051e-01    2.799 0.005127 **
## categorygames               1.814e+00  2.626e-01    6.908 4.93e-12 ***
## categoryGames               1.945e+01  1.040e+03    0.019 0.985086
## categoryGaming Hardware     7.108e-01  3.097e-01    2.295 0.021724 *
## categoryGlass               1.219e+00  4.238e-01    2.875 0.004039 **
## categoryGraphic Design      1.949e+00  2.702e-01    7.213 5.48e-13 ***
## categoryGraphic Novels      5.025e+00  3.765e-01   13.349  < 2e-16 ***
## categoryHardware            5.869e-01  3.393e-01    1.729 0.083743 .
## categoryHip-Hop             1.208e+00  2.797e-01    4.319 1.56e-05 ***
## categoryHorror              1.470e+00  2.846e-01    5.165 2.40e-07 ***
## categoryIllustration        2.027e+01  2.929e+02    0.069 0.944835
## categoryImmersive           1.818e+00  4.275e-01    4.252 2.12e-05 ***
## categoryIndie Rock          2.031e+01  3.465e+02    0.059 0.953251
## categoryInstallations       1.802e+00  3.228e-01    5.582 2.37e-08 ***
## categoryInteractive Design  2.486e-01  3.242e-01    0.767 0.443289
## categoryJazz                1.789e+00  2.756e-01    6.493 8.39e-11 ***
## categoryJewelry             1.296e+00  2.699e-01    4.803 1.56e-06 ***
## categoryjournalism         -1.744e-01  3.032e-01   -0.575 0.565295
## categoryJournalism          2.703e+00  2.851e-01    9.483  < 2e-16 ***
## categoryKids                1.171e+00  3.673e-01    3.188 0.001431 **
## categoryKnitting            1.819e+00  4.344e-01    4.187 2.82e-05 ***
## categoryLatin               9.525e-01  4.705e-01    2.024 0.042930 *
## categoryLetterpress         1.581e+00  4.684e-01    3.374 0.000740 ***
## categoryLiterary Journals   1.898e+00  3.055e-01    6.213 5.19e-10 ***
## categoryLiterary Spaces     1.518e+00  3.735e-01    4.063 4.84e-05 ***
## categoryLive Games          9.859e-01  3.024e-01    3.261 0.001112 **
## categoryMakerspaces         4.627e-01  6.086e-01    0.760 0.447106
## categoryMetal               1.580e+00  3.042e-01    5.194 2.05e-07 ***
## categoryMixed Media         2.221e+00  2.718e-01    8.170 3.09e-16 ***
## categoryMobile Games       -1.143e-03  2.978e-01   -0.004 0.996938
## categoryMovie Theaters     -5.291e-02  5.224e-01   -0.101 0.919322
## categorymusic               2.089e+00  2.602e-01    8.029 9.81e-16 ***
## categoryMusic               1.986e+01  5.480e+02    0.036 0.971096
## categoryMusic Videos        1.858e+00  2.999e-01    6.194 5.88e-10 ***
## categoryMusical             1.868e+00  2.867e-01    6.517 7.16e-11 ***
```

```
## categoryNarrative Film      3.983e+00  2.870e-01   13.879  < 2e-16 ***
## categoryNature              4.946e-01  3.136e-01    1.577 0.114802
## categoryNonfiction          1.994e+01  3.459e+02    0.058 0.954041
## categoryPainting            2.449e+00  2.735e-01    8.952  < 2e-16 ***
## categoryPeople              1.182e+00  2.901e-01    4.075 4.60e-05 ***
## categoryPerformance Art     1.769e+00  2.752e-01    6.426 1.31e-10 ***
## categoryPerformances        2.254e+00  2.743e-01    8.216  < 2e-16 ***
## categoryPeriodicals         6.636e-01  3.048e-01    2.177 0.029465 *
## categoryPet Fashion         4.785e-01  4.827e-01    0.991 0.321524
## categoryPhoto               6.291e-01  3.679e-01    1.710 0.087262 .
## categoryPhotobooks          1.539e+00  2.778e-01    5.540 3.02e-08 ***
## categoryphotography         1.123e+00  2.652e-01    4.233 2.30e-05 ***
## categoryPhotography         2.066e+01  5.740e+02    0.036 0.971285
## categoryPlaces              1.307e+00  3.049e-01    4.285 1.83e-05 ***
## categoryPlaying Cards       1.900e+01  2.546e+02    0.075 0.940521
## categoryPlays               2.263e+00  2.678e-01    8.451  < 2e-16 ***
## categoryPoetry              1.777e+00  2.676e-01    6.639 3.16e-11 ***
## categoryPop                 2.650e+00  2.679e-01    9.891  < 2e-16 ***
## categoryPottery             1.657e+00  4.589e-01    3.612 0.000304 ***
## categoryPrint               1.286e+00  2.897e-01    4.437 9.13e-06 ***
## categoryPrinting            6.165e-01  4.222e-01    1.460 0.144191
## categoryProduct Design      2.189e+01  1.480e+02    0.148 0.882410
## categoryPublic Art          3.240e+00  2.733e-01   11.853  < 2e-16 ***
## categorypublishing          2.122e+00  2.610e-01    8.130 4.30e-16 ***
## categoryPublishing          1.938e+01  9.497e+02    0.020 0.983714
## categoryPunk                1.869e+00  3.021e-01    6.188 6.09e-10 ***
## categoryPuzzles             1.149e+00  3.685e-01    3.119 0.001817 **
## categoryQuilts              1.311e+00  6.395e-01    2.050 0.040412 *
## categoryR&B                 1.054e+00  2.985e-01    3.532 0.000412 ***
## categoryRadio & Podcasts    1.409e+00  2.778e-01    5.074 3.89e-07 ***
## categoryReady-to-wear       6.616e-01  3.143e-01    2.105 0.035313 *
## categoryResidencies         3.091e+00  4.683e-01    6.600 4.10e-11 ***
## categoryRestaurants         5.075e-01  2.886e-01    1.759 0.078635 .
## categoryRobots              9.690e-01  3.240e-01    2.990 0.002787 **
## categoryRock                2.039e+01  2.642e+02    0.077 0.938475
## categoryRomance             1.125e+00  3.367e-01    3.340 0.000838 ***
## categoryScience Fiction     1.253e+00  2.822e-01    4.439 9.03e-06 ***
## categorySculpture           1.486e+00  2.743e-01    5.416 6.09e-08 ***
## categoryShorts              2.080e+01  3.465e+02    0.060 0.952140
## categorySmall Batch         1.178e+00  2.772e-01    4.250 2.14e-05 ***
## categorySocial Practice     2.173e+00  5.577e-01    3.897 9.75e-05 ***
## categorySoftware            1.848e-01  2.758e-01    0.670 0.502991
## categorySound               2.345e-01  2.998e-01    0.782 0.434076
## categorySpace Exploration   1.208e+00  3.837e-01    3.148 0.001642 **
## categorySpaces              7.561e-01  3.049e-01    2.480 0.013138 *
## categoryStationery          1.444e+00  3.495e-01    4.132 3.60e-05 ***
## categoryTabletop Games      1.817e+01  1.791e+02    0.101 0.919202
## categoryTaxidermy           1.694e+00  1.244e+00    1.362 0.173169
## categorytechnology          5.679e-01  2.614e-01    2.173 0.029794 *
## categoryTechnology          1.949e+01  7.922e+02    0.025 0.980376
## categoryTelevision          5.893e-01  3.240e-01    1.819 0.068924 .
```

```
## categoryTextiles           1.348e+00  3.471e-01    3.884 0.000103 ***
## categorytheater            2.483e+00  2.651e-01    9.367  < 2e-16 ***
## categoryTheater            2.077e+01  6.609e+02    0.031 0.974933
## categoryThrillers          1.532e+00  3.033e-01    5.053 4.36e-07 ***
## categoryToys               1.438e+00  3.890e-01    3.697 0.000218 ***
## categoryTranslations       5.923e-01  4.157e-01    1.425 0.154256
## categoryTypography         1.038e+00  5.521e-01    1.881 0.060003 .
## categoryVegan              5.833e-01  3.121e-01    1.869 0.061646 .
## categoryVideo              4.618e-01  3.190e-01    1.448 0.147729
## categoryVideo Art          1.238e+00  3.641e-01    3.401 0.000672 ***
## categoryVideo Games        1.949e+01  2.343e+02    0.083 0.933691
## categoryWearables         -1.060e+00  2.946e-01   -3.597 0.000322 ***
## categoryWeaving            1.989e+00  4.548e-01    4.373 1.23e-05 ***
## categoryWeb                4.408e-01  2.738e-01    1.610 0.107371
## categoryWebcomics          1.436e+00  2.776e-01    5.173 2.30e-07 ***
## categoryWebseries          2.659e+00  2.805e-01    9.481  < 2e-16 ***
## categoryWoodworking        1.218e+00  2.819e-01    4.321 1.55e-05 ***
## categoryWorkshops          1.917e+00  3.817e-01    5.020 5.15e-07 ***
## categoryWorld Music        1.337e+00  2.818e-01    4.745 2.08e-06 ***
## categoryYoung Adult        1.047e+00  3.049e-01    3.435 0.000593 ***
## categoryZines              1.948e+00  3.192e-01    6.100 1.06e-09 ***
## converted_pledged_amount   3.893e-05  2.448e-06   15.899  < 2e-16 ***
## goal                      -8.896e-06  7.086e-08 -125.555  < 2e-16 ***
## staff_pick                 8.824e-02  3.560e-02    2.479 0.013187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 207378  on 154035  degrees of freedom
## Residual deviance:  97304  on 153856  degrees of freedom
## AIC: 97664
##
## Number of Fisher Scoring iterations: 18
```

### 6.2.2 Stepwise Selection Process

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

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

```
step.model <- full.model %>% stepAIC(trace = FALSE)
```

```
## 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: algorithm did not converge

## 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
```

```
coef(step.model)
```

```
##                (Intercept)              backers_count
##              -2.966505e+00               4.595230e-02
##           categoryAcademic        categoryAccessories
##               9.312302e-01               1.931386e+01
##             categoryAction            categoryAnimals
##               7.103480e-01               6.842589e-01
##          categoryAnimation        categoryAnthologies
##               9.930490e-01               1.306950e+00
##            categoryApparel               categoryApps
##               1.930420e+01               7.237470e-01
##       categoryArchitecture                categoryart
##               3.952575e-01               2.161083e+00
##                categoryArt          categoryArt Books
##               1.947802e+01               2.349677e+00
##              categoryAudio              categoryBacon
##               1.212213e+00               9.214716e-01
##              categoryBlues          categoryCalendars
##               1.615198e+00               1.363835e+00
##   categoryCamera Equipment            categoryCandles
##               1.261150e-01               5.872234e-01
##           categoryCeramics   categoryChildren's Books
##               1.496918e+00               1.871770e+01
##      categoryChildrenswear           categoryChiptune
##               4.677318e-01               3.050255e+00
##       categoryCivic Design    categoryClassical Music
##               4.597170e-01               2.847506e+00
##             categoryComedy        categoryComic Books
##               1.672111e+00               1.832583e+01
##             categorycomics             categoryComics
##               2.543128e+00               1.832772e+01
##  categoryCommunity Gardens     categoryConceptual Art
```

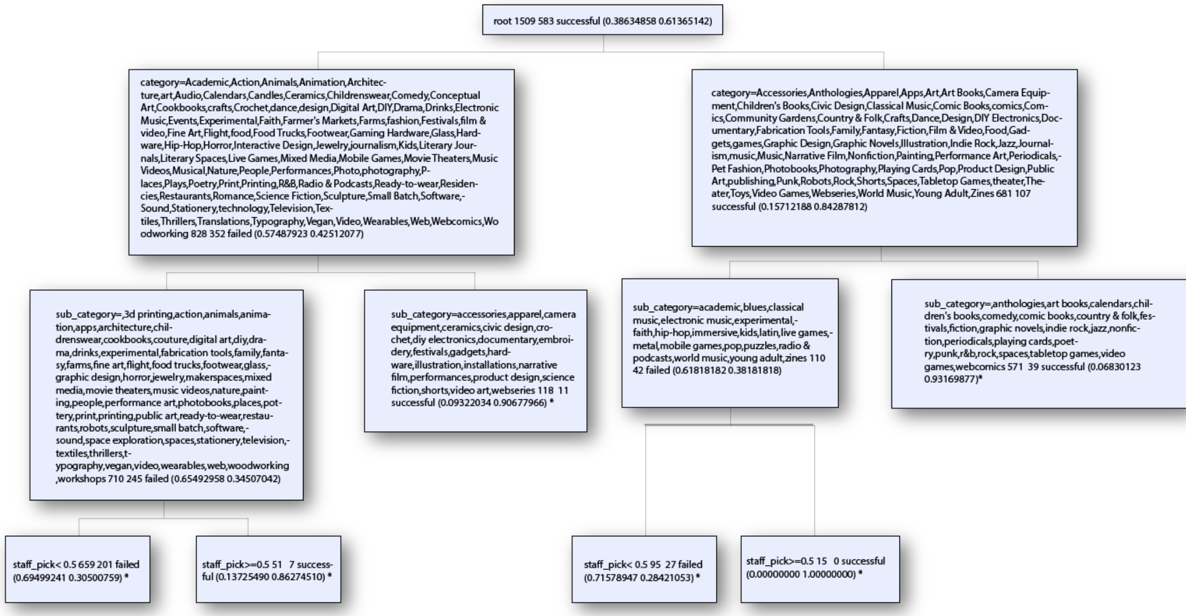
```
##             9.739571e-01             1.672961e+00
##        categoryCookbooks     categoryCountry & Folk
##             4.205342e-01             1.858750e+01
##          categoryCouture            categorycrafts
##             8.708358e-01             1.910578e+00
##           categoryCrafts           categoryCrochet
##             1.985329e+01             1.402801e+00
##            categorydance             categoryDance
##             2.424482e+00             1.956798e+01
##           categorydesign            categoryDesign
##             1.346864e+00             2.234257e+01
##      categoryDigital Art               categoryDIY
##             1.721651e+00             9.239068e-01
##  categoryDIY Electronics      categoryDocumentary
##             8.705477e-01             1.883556e+01
##            categoryDrama            categoryDrinks
##             2.070814e+00             1.057923e+00
##  categoryElectronic Music      categoryEmbroidery
##             1.310428e+00             9.222296e-01
##           categoryEvents      categoryExperimental
##             8.484303e-01             1.680079e+00
## categoryFabrication Tools             categoryFaith
##             4.870069e-01             1.392403e+00
##           categoryFamily           categoryFantasy
##             8.129843e-01             1.456360e+00
##  categoryFarmer's Markets             categoryFarms
##             9.260712e-01             8.978848e-01
##          categoryfashion           categoryFashion
##             2.212736e+00             1.896598e+01
##       categoryFestivals           categoryFiction
##             2.052482e+00             1.959631e+01
##     categoryfilm & video       categoryFilm & Video
##             1.927642e+00             1.972256e+01
##         categoryFine Art            categoryFlight
##             1.586326e+00             9.879228e-02
##             categoryfood              categoryFood
##             1.883626e-01             2.110974e+01
##      categoryFood Trucks          categoryFootwear
##             1.621389e-01             8.746375e-02
##          categoryGadgets             categorygames
##             8.479930e-01             1.810753e+00
##            categoryGames    categoryGaming Hardware
##             1.847399e+01             7.084527e-01
##            categoryGlass     categoryGraphic Design
##             1.214199e+00             1.948052e+00
##   categoryGraphic Novels          categoryHardware
##             5.023387e+00             5.803880e-01
##          categoryHip-Hop            categoryHorror
##             1.206437e+00             1.469356e+00
##     categoryIllustration         categoryImmersive
##             1.930202e+01             1.825271e+00
```

```
##          categoryIndie Rock       categoryInstallations
##                1.935956e+01                1.806780e+00
## categoryInteractive Design                categoryJazz
##                2.467324e-01                1.786176e+00
##             categoryJewelry         categoryjournalism
##                1.293614e+00               -1.651884e-01
##          categoryJournalism                categoryKids
##                2.703822e+00                1.168850e+00
##           categoryKnitting               categoryLatin
##                1.817990e+00                9.510469e-01
##         categoryLetterpress  categoryLiterary Journals
##                1.587871e+00                1.905961e+00
##     categoryLiterary Spaces          categoryLive Games
##                1.525585e+00                9.858165e-01
##         categoryMakerspaces               categoryMetal
##                4.764454e-01                1.585009e+00
##         categoryMixed Media        categoryMobile Games
##                2.220349e+00               -2.114662e-03
##      categoryMovie Theaters               categorymusic
##               -5.799914e-02                2.087104e+00
##               categoryMusic        categoryMusic Videos
##                1.891626e+01                1.856354e+00
##             categoryMusical      categoryNarrative Film
##                1.866623e+00                3.981067e+00
##              categoryNature          categoryNonfiction
##                4.967091e-01                1.897727e+01
##            categoryPainting              categoryPeople
##                2.447082e+00                1.184098e+00
##     categoryPerformance Art        categoryPerformances
##                1.768458e+00                2.263204e+00
##         categoryPeriodicals         categoryPet Fashion
##                6.697603e-01                4.775444e-01
##               categoryPhoto          categoryPhotobooks
##                6.311526e-01                1.543095e+00
##         categoryphotography         categoryPhotography
##                1.124123e+00                1.971088e+01
##              categoryPlaces       categoryPlaying Cards
##                1.308792e+00                1.803824e+01
##               categoryPlays               categoryPoetry
##                2.262488e+00                1.779170e+00
##                 categoryPop              categoryPottery
##                2.646528e+00                1.667764e+00
##               categoryPrint             categoryPrinting
##                1.287516e+00                6.167681e-01
##      categoryProduct Design          categoryPublic Art
##                2.095323e+01                3.241461e+00
##          categorypublishing          categoryPublishing
##                2.121765e+00                1.844187e+01
##                categoryPunk              categoryPuzzles
##                1.868963e+00                1.143158e+00
##              categoryQuilts                 categoryR&B
```

```
##             1.312914e+00            1.051689e+00
##   categoryRadio & Podcasts      categoryReady-to-wear
##             1.413281e+00            6.612235e-01
##       categoryResidencies       categoryRestaurants
##             3.095341e+00            5.074247e-01
##           categoryRobots            categoryRock
##             9.722615e-01            1.942445e+01
##          categoryRomance    categoryScience Fiction
##             1.125959e+00            1.251038e+00
##        categorySculpture           categoryShorts
##             1.489091e+00            1.985117e+01
##       categorySmall Batch    categorySocial Practice
##             1.176921e+00            2.179814e+00
##         categorySoftware            categorySound
##             1.810827e-01            2.430523e-01
##  categorySpace Exploration            categorySpaces
##             1.209640e+00            7.549638e-01
##        categoryStationery     categoryTabletop Games
##             1.446098e+00            1.719850e+01
##        categoryTaxidermy        categorytechnology
##             1.693169e+00            5.651492e-01
##        categoryTechnology        categoryTelevision
##             1.855833e+01            5.853594e-01
##         categoryTextiles          categorytheater
##             1.349635e+00            2.483271e+00
##          categoryTheater         categoryThrillers
##             1.982076e+01            1.529295e+00
##             categoryToys      categoryTranslations
##             1.436017e+00            5.927574e-01
##        categoryTypography            categoryVegan
##             1.051035e+00            5.838680e-01
##            categoryVideo         categoryVideo Art
##             4.637572e-01            1.237936e+00
##       categoryVideo Games         categoryWearables
##             1.843406e+01           -1.066625e+00
##          categoryWeaving              categoryWeb
##             1.989780e+00            4.402251e-01
##        categoryWebcomics         categoryWebseries
##             1.436630e+00            2.656786e+00
##       categoryWoodworking          categoryWorkshops
##             1.219098e+00            1.916139e+00
##       categoryWorld Music        categoryYoung Adult
##             1.335368e+00            1.047527e+00
##            categoryZines   converted_pledged_amount
##             1.962536e+00            3.907837e-05
##                   goal
##            -8.932700e-06
```

## 6.3 Random Forest

### 6.3.1 Decision Tree Results




### 6.3.2 Decision Tree Categories

Categories 1:

Academic,Action,Animals,Animation,Architecture,art,Audio,Calendars,Candles, Ceramics,Childrenswear,Comedy,Conceptual Art,Cookbooks,crafts,Crochet,dance ,design,Digital Art,DIY,Drama,Drinks,Electronic Music,Events,Experimental, Faith,Farmer's Markets,Farms,fashion,Festivals,film & video,Fine Art,Flight,food, Food Trucks,Footwear,Gaming Hardware,Glass,Hardware,Hip-Hop,Horror, Interactive Design,Jewelry,journalism,Kids,Literary Journals,Literary Spaces, Live Games,Mixed Media,Mobile Games,Movie Theaters,Music Videos,Musical,Nature, People,Performances,Photo,photography,Places,Plays,Poetry,Print,Printing,R&B, Radio & Podcasts,Ready-to-wear,Residencies,Restaurants,Romance,Science Fiction, Sculpture,Small Batch,Software,Sound,Stationery,technology,Television, Textiles,Thrillers,Translations,Typography,Vegan,Video,Wearables,Web,Webcomics, Woodworking

Categories 2:

Accessories,Anthologies,Apparel,Apps,Art,Art Books,Camera Equipment, Children's Books,Civic Design,Classical Music,Comic Books, comics,Comics,Community Gardens,Country & Folk,Crafts,Dance,Design,DIY Electronics,Documentary,Fabrication Tools,Family,Fantasy,Fiction, Film & Video,Food,Gadgets,games,Graphic Design,Graphic Novels,Illustration, Indie Rock,Jazz,Journalism,music,Music,Narrative Film,Nonfiction ,Painting,Performance Art,Periodicals,Pet Fashion,Photobooks,Photography, Playing Cards,Pop,Product Design,Public Art,publishing,Punk,Robots,Rock, Shorts,Spaces,Tabletop Games,theater,Theater,Toys,Video Games,Webseries, World Music,Young Adult,Zines

Subcategories 1

3d printing,action,animals,animation,apps,architecture,childrenswear,cookbooks, couture,digital art,diy,drama,drinks,experimental,fabrication tools,family, fantasy,farms,fine art,flight,food trucks,footwear,glass,graphic design, horror,jewelry,makerspaces,mixed media,movie theaters,music videos, nature,painting,people,performance art,photobooks,places,pottery,print,printing, public art,ready-to-wear,restaurants,robots,sculpture,small batch,software,sound, space exploration,spaces,stationery,television,textiles,thrillers,typography,vegan, video,wearables,web,woodworking,workshops

Subcategories 2

accessories,apparel,camera equipment,ceramics,civic design,crochet, diy electronics,documentary,embroidery,festivals,gadgets,hardware,illustration, installations,narrative film,performances,product design,science fiction,shorts, video art,webseries

Subcategories 3

academic,blues,classical music,electronic music,experimental, faith,hip-hop,immersive,kids,latin,live games,metal,mobile games, pop,puzzles,radio & podcasts,world music,young adult,zines

Subcategories 4

anthologies,art books,calendars,children's books,comedy,comic books, country & folk,festivals,fiction,graphic novels,indie rock,jazz,nonfiction,periodicals, playing cards,poetry,punk,r&b,rock,spaces,tabletop games,video games,webcomics

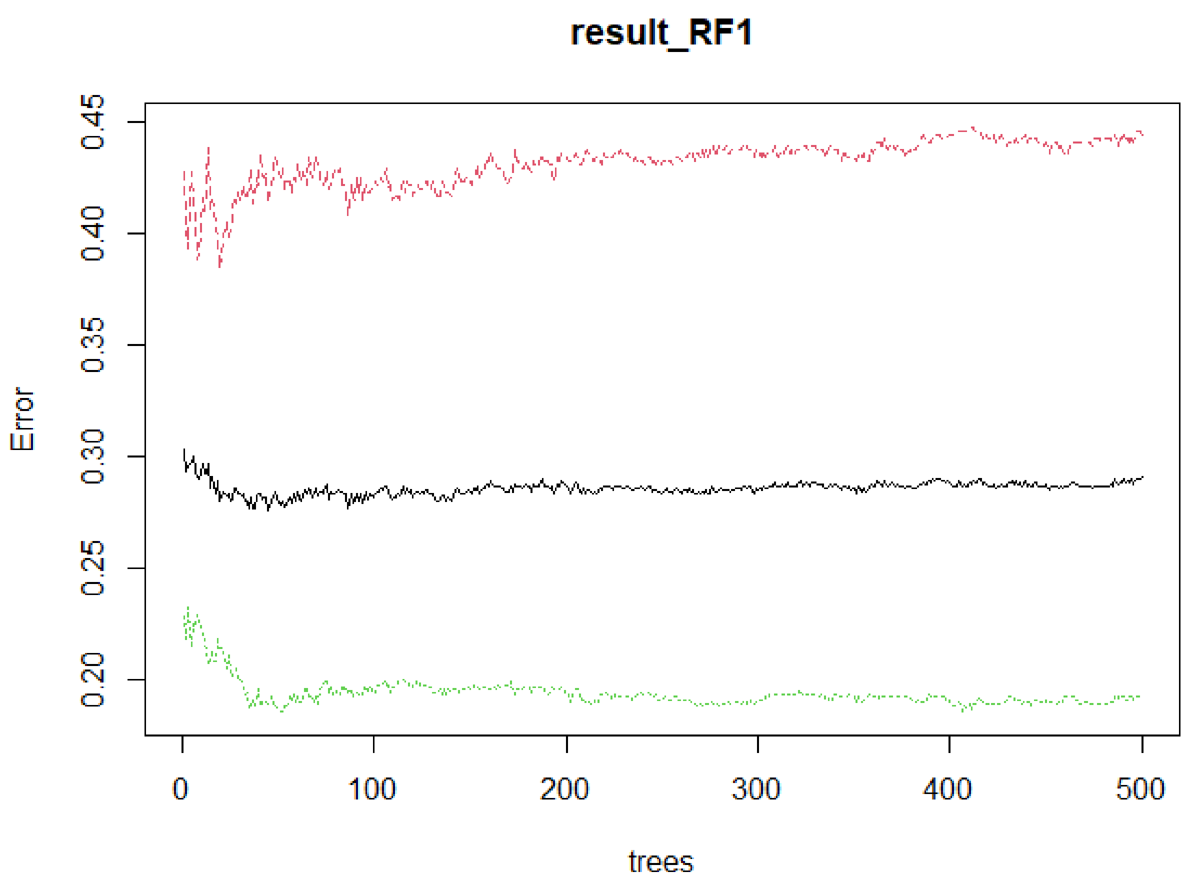
### 6.3.3 Confusion Matrix Classification Tree Results

```
> print(paste("Accuracy:",Accuracy))
[1] "Accuracy: 0.811133200795229"
> print(paste("Misclassification Error:",Mis_error))
[1] "Misclassification Error: 0.188866799204771"
> print(paste("Precision:",Precision))
[1] "Precision: 0.924503311258278"
> print(paste("Recall:",Recall))
[1] "Recall: 0.75377969762419"
> print(yhat.table)
            yhat.result
             failed successful
  failed        526         57
  successful    228        698
```

### 6.3.4 Random Forest Results

```
Call:
 randomForest(formula = as.factor(state) ~ . - id - goal - converted_pledged_amoun
t -      backers_count - static_usd_rate - usd_pledged - success,      data = Data
_Te, mtry = 3, importance = TRUE)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 3

        OOB estimate of  error rate: 29.09%
Confusion matrix:
           failed successful class.error
failed        324        259   0.4442539
successful    180        746   0.1943844
```




**6.3.5 Random Forest Confusion Matrix**

```
[1] ----Class
> print(yhat_RF1.table)
            yhat_RF1.result
             failed successful
  failed        439        144
  successful     77        849
> #Evaluate quality of prediction
> Accuracy1=(yhat_RF1.table[1,1]+yhat_RF1.table[2,2])/nrow(Data_Te)
> Mis_error1=1-Accuracy1
> Precision1=yhat_RF1.table[2,2]/(yhat_RF1.table[1,2]+yhat_RF1.table[2,2])
> Recall1=yhat_RF1.table[2,2]/(yhat_RF1.table[2,1]+yhat_RF1.table[2,2])
> print(paste("Accuracy:",Accuracy1))
[1] "Accuracy: 0.853545394300862"
> print(paste("Misclassification Error:",Mis_error1))
[1] "Misclassification Error: 0.146454605699138"
> print(paste("Precision:",Precision1))
[1] "Precision: 0.854984894259819"
> print(paste("Recall:",Recall1))
[1] "Recall: 0.916846652267819"
```

### 6.4 Text Mining