ISSS616 Applied Statistical Analysis with R
Final Report

Group Project: Understanding and Predicting Success of Kickstarter Projects Across Top 15 Countries

G2 Group 6

Archie Quiambao DOLIT

MIN Xiaoqi

Kevin Magic Rialubin SUNGA

WANG Zhifei

Lynette WONG Yik Lynn

WONG Sook Xian Sophia

6th April 2021

Table of Contents

1. Background

Kickstarter, which was launched in 2009, provides tools and a platform to help creative projects become reality by linking artists and creators with backers who are willing to support their projects. As of April 2021, Kickstarter has launched 513,806 projects with 38.4% of the projects becoming successful and raising US$5.7 billion in total[1].

As a Benefit Corporation, Kickstarter is committed to make a "positive impact on society"[2] and a core part of Kickstarter is its "all-or-nothing" model. Under this model, the funds placed by backers are not released unless a project meets its goals. In this regard, the artists and creators decide on the financial goal and fundraising deadline for their project and unless the project is successfully funded, Kickstarter does not collect any fees.

2. Overall Concept

2.1. Objective

Currently, statistics on Kickstarter projects need to be accessed from different websites with limited data analysis that project creators could rely on for a holistic view of the past projects' performances. To ensure successful fundraising, it is important for creators to understand the underlying variables contributing to a successful project. With this information, they can make more informed decisions when designing the different aspects of their Kickstarter projects. In our preliminary studies, we found that staff picked projects are more likely to be successful. Hence, this study aims to help creators to design a winning project from the following aspects:

1. To have a holistic dashboard where creators can assess the performance of past projects providing both overview and detailed campaign metrics
2. To determine a suitable goal from historical pledged amount
3. To understand the contributing factors of a staff-picked project
4. To fine tune the variables to simulate the possible outcome of their project: successful or failed

2.2. Literature Review

Existing research on predicting the success of Kickstarter or crowdfunding projects have utilized logistic analysis, decision trees and regression analysis as the three main methods for predictive modelling. The inputs to these models are mainly focused on basic project variables, such as goal funding amount, duration of campaign and project category to predict the outcome of the project.

There are researchers who have focused on other factors that are less common project properties. Zhou et al. (2018) used a unimodal theory of persuasion and the results show that the length, readability and tone in the project description are important in predicting the success of the projects. Greenberg et al. (2013) have shown that the quality of the project introduction is one of the contributing factors to successful projects. If a project introduction has included images, videos or other visualization tools, there will be a higher likelihood of funding success. Studies have also shown that the influential power of the project owner can affect the success of the funding, for instance the project owners' influence on social media (Mollick, 2014) and the project's geographic influence where creators are likely to propose ideas that relates to the cultural products of their regions (Agrawal et al., 2011). Another factor, researched by Zvilichovsky et al. (2015), is reciprocity: the phenomenon whereby projects owners who have backed other projects are more likely to succeed with their own projects.

[1] Kickstarter Stats — Kickstarter
[2] Charter — Kickstarter

Ullah et al. (2020) used univariate analysis to compare the means and medians of the variables and then proceed with logistical and regression analysis. The results showed that projects with longer project descriptions, longer campaign durations and higher goal amounts are less likely to succeed. Furthermore, they factored in demographic data about creators which led to their finding that female project creators are more likely to succeed in non-traditional project categories such as comics, film, and technology. However, a limitation with their study is their model does not consider dynamic factors and the data timeframe is only 6 months long. Kuppuswamy and Bayus (2018) showed that there is a high correlation of success rate with the project timeline where backers are likely to pledge with higher amounts at the earlier stage of a project.

As an extension of existing research, our model encompasses a data timeframe of 5 years with both common project properties as well as alternative variables such as staff pick, project and blurb word length and project locations for our predictive modelling. We have integrated these into a R Shiny App for effective visualization of the contributing factors to a successful project as well as the suitable pledge amount. Moreover, our analysis will consolidate different variables into one and emphasize on pledged amount and staff pick prediction, which offers further flexibility on our modelling as compared to other studies.

As an extension of existing research, our model encompasses a data timeframe of 5 years with both common project properties as well as alternative variables such as staff pick, project and blurb word length and project locations for our predictive modelling. We have integrated these into a R Shiny App for effective visualization of the contributing factors to a successful project as well as the suitable pledge amount.

2.3. Analysis Process

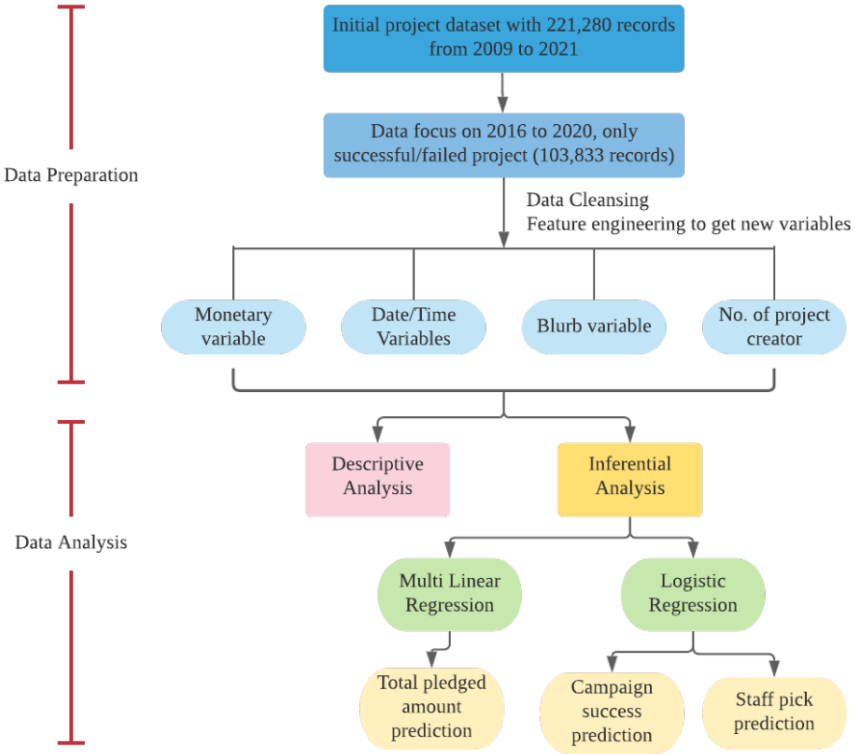


Figure 1 Process Flow of Data Preparation and Data Analysis

3. Data Source and Preparation

3.1. Dataset Used

Our project dataset, taken from webrobots[3], contains monthly Kickstarter projects from 2014 to 2021. The data was processed and combined to create an initial dataset that contains 221,280 records of projects launched from 25 Apr 2009 to 14 Jan 2021, including projects that are successful, failed, cancelled and live (i.e., projects that are still within the funding period as of the date on which the data was scraped).

3.2. Data Cleansing

At the data cleansing stage, Microsoft Excel and JMP Pro were used as the main tools for table merging, JSON reformatting, epoch timestamp transformation and data cleansing. Monthly data extracted from webrobots were combined and filtered for projects that have reached either "successful" or "failed" state with country of origin from one of the top 15 countries. "Live" and "Cancelled" projects are indeterministic in understanding the traits that contribute to a successful project and hence irrelevant to our analysis. Our final dataset contains 103,833 records which accounts for 94.87% of the total completed projects launched during the period.

Standard cleansing steps were carried out to identify variables with a high proportion of missing values and possible outliers.

Missing Values

Table 1 Summary statistics of variables with missing values

Columns	N	N Missing	% Missing	N Categories	Min	Max	Mean	Std Dev
category_parent_id	98,331	5,502	5%	.	1.0000	26.0000	11.6495	5.5630
category_parent_name	98,331	5,502	5%	15	.	.	.	.
creator_slug	54,592	49,241	47%	43,258	.	.	.	.
friends	15	103,818	100%	1	.	.	.	.
is_backing	15	103,818	100%	1	.	.	.	.
is_starred	15	103,818	100%	1	.	.	.	.
permissions	15	103,818	100%	1	.	.	.	.

Summary statistics shown that column *friends*, *is_backing*, *is_starred* and *permissions* have close to 100% missing values and *creator_slug* has half of its values missing. Hence, these 5 columns were removed and will not be considered in the subsequent study.

Column *category_parent_id* and *category_parent_name* have 5% of the values missing. Upon closer investigation, *category_parent_name* was omitted when it has the same name as *category_name*, the granular classification of *category_parent_name*. Hence, we recoded *category_parent_name* into *category_parent_name_recoded* whereby we replaced the missing values with *category_name*.

[3] Kickstarter Datasets – Web Scraping Service (webrobots.io)

Outliers

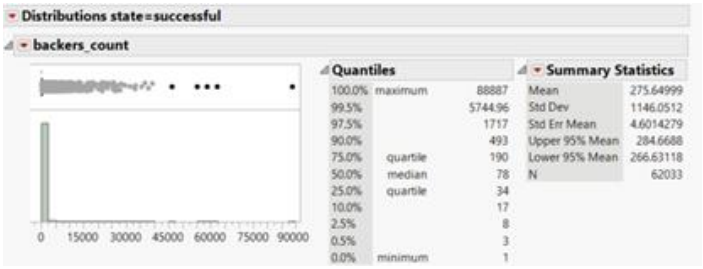


Figure 2 backers_count distribution of successful projects

In graphing the distribution of *backers_count*, we see there are 5 successful projects with more than 450,000 *backers_count*. Most of them are in the Video Game and Product Design categories. We investigated a sample of these projects, and found that they are projects launched by relatively sophisticated start-ups. As such, they remain a crucial part of the study in helping us understand what factors contribute to their success. Hence, they are kept in subsequent analysis and not removed as outliers.

3.3. Feature Engineering to obtain new variables

Since the dataset is a crawling of the Kickstarter website, many of the fields needed to be processed before they can be used in our analysis. New fields derived can be broadly classified into date/time variable, monetary variable, blurb variable and others.

Date/Time Variables: After converting *launch_at* from epoch format to a more readable UTC format, it was broken down into the corresponding year, month, and day to give us *launch_year, launch_month* and *launch_day*. *Campaign_duration* was derived by calculating the number of days between *launch_at* and *deadline*.

Monetary Variable: Since the projects have different country of origin, some of the goals were set in a local currency instead of a standardised currency. To facilitate easier comparison among projects of different origin and currency, we standardised all variables that are monetary in nature to US dollars.

Title/Blurb Variable: Since sentiment study of text content is not in the scope of this analysis, the content of the blurb is excluded in this study. Instead, length of project name and length of blurb were derived because we have the hypothesis that it might affect the outreach of the campaign and in turn affect the backing. We have also used Google API to detect the language of the project title and description. However, the detected language variables were excluded in the final model since 95% of the blurbs and 91% of the title names were in English.

Others: *num_project_creator* was derived from counting the number of projects from the same *creator_id*. We use this field to proxy the experience level and how sophisticated the creators are. A limitation with the way we calculated num_project_creator variable is that we assumed that the creators have not created projects prior to 2016. However, it is possible for a creator to have created projects in prior years.

The complete list of derived variables, description and calculations can be found in Appendix. Feature Engineering.

4. Data Analysis Descriptive and Inferential Statistics

4.1. Country of Origin

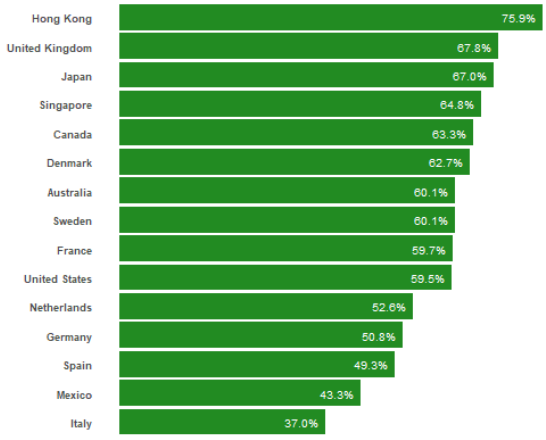


Figure 3 Descriptive Analysis - Success Rate by Country of Origin

We have analysed the past 5 years of data for each of the selected 15 countries. Figure 3 shows the success rate by their origin. It is noted that from 2016 to 2020, the top 3 countries in terms of success rate are Hong Kong (75.9%), United Kingdom (67.8%) and Japan (67.0%). Most of the countries with high success rate during this period are from Asia, such as Hong Kong, Japan, and Singapore.

4.2. Main Category

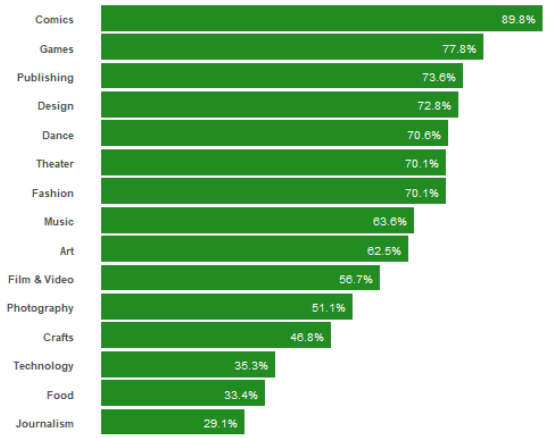


Figure 4 Descriptive Analysis - Success Rate by Category

In terms of projects' parent categories, we note that comics, games, publishing, design and dance have the top 5 success rate. On the other hand, technology, food and journalism are at the bottom with the lowest success rate.

4.3. Campaign Duration

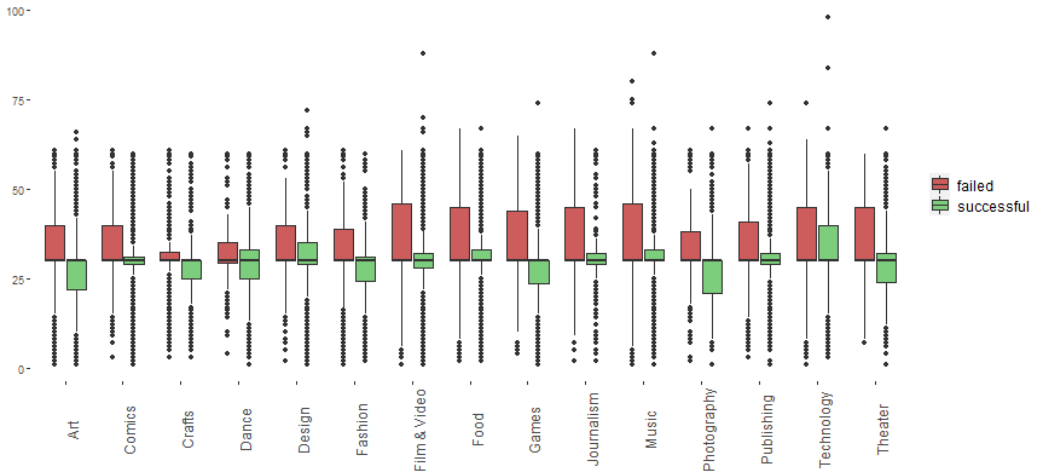


Figure 5 Descriptive Analysis - Boxplot of Campaign Duration by Category

Campaign duration appears to be a factor in determining the success of projects. From the boxplots, we noticed that the majority of the successful projects have a median and 75% quantile duration around 30 days while projects that exceeded 30 days are more likely to fail than projects with shorter durations (Figure 5).

To test for association, we grouped the projects into two groups: projects with campaign duration less than or equal to the median of 30 days and more than 30 days. We then conducted a Chi-Square Test on campaign duration and state of project.

```
> R1 <- c(43031, 25072)
> R2 <- c(19002, 16728)
> rows <- 2
> Matriz <- matrix(c(R1,R2),nrow=rows,byrow=TRUE)
>
> rownames(Matriz) <- c("<= 30d", "> 30d")
> colnames(Matriz) <- c("Success", "Fail")
>
> Matriz
       Success  Fail
<= 30d   43031 25072
> 30d    19002 16728
>
> Result <- chisq.test(Matriz,correct=FALSE)
> Result

        Pearson's Chi-squared test

data:  Matriz
X-squared = 974.98, df = 1, p-value < 2.2e-16

> Result$observed - Result$expected
        Success      Fail
<= 30d  2344.191 -2344.191
> 30d  -2344.191  2344.191
```

Figure 6 Chi-Square Test Result of Campaign Duration by Project Success State

The null hypothesis for this test is there is no association between campaign duration and success or fail while the alternative hypothesis states that there is an association between campaign duration and success or fail. Since the p-value is less than 0.05, we reject the null hypothesis at 5% significance level and conclude that there is an association between campaign duration and success or failure of the project. Furthermore, we find that campaigns with less than or equal to 30 days duration tend to be more successful while campaigns with longer than 30 days duration tend to fail more.

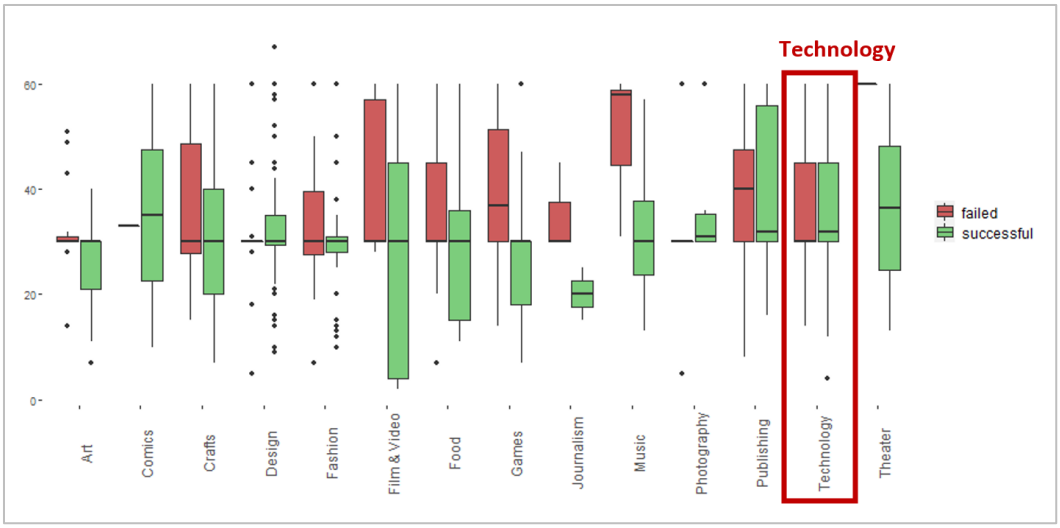


Figure 7 Boxplot of Campaign Duration By project state for Hong Kong

However, this association may not hold for all project categories across the countries. In **Error! Reference source not found.**, campaign duration of Technology projects in Hong Kong showed equal probability of success and failure.

4.4. Staff Pick

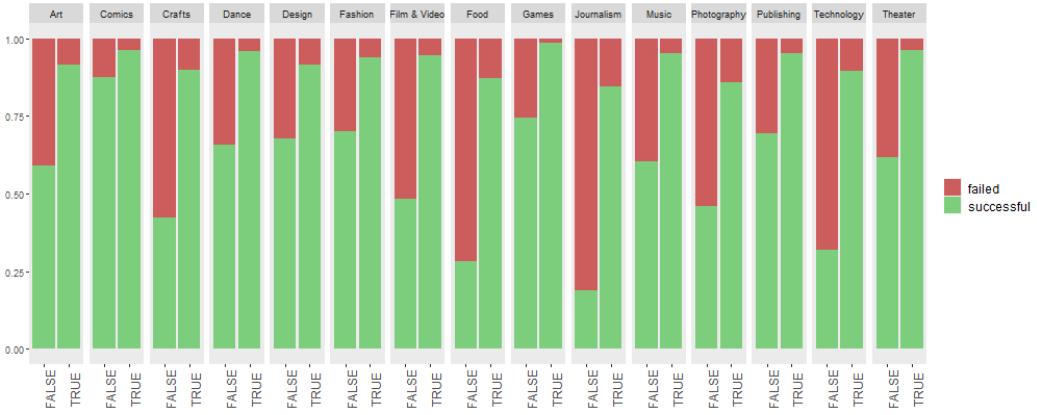


Figure 8 Descriptive Analysis - Project state by Staff Pick

The proportion of successful projects is higher if it was picked by staff (Staff pick = TRUE) across all project categories.

```
                      Australia Canada Denmark France Germany Hong Kong Italy Japan Mexico Netherlands Singapore Spain Sweden United Kingdom  United States
Art - Staff Pick             11     18       4     21      13         2    15    11     42           3         1    11      4            118            460
Comics - Staff Pick          25    135       4      7      13         2    18     7     25           2         3     6      6            199           1148
Crafts - Staff Pick           6      8       3      4      10         0     1     8     37           1         4     0      1             39            120
Dance - Staff Pick            0      2       0      2       1         0     1     0      4           0         0     2      1             20            189
Design - Staff Pick          14     30      15     35      35        17    10     3      4           8         3    13      8             85            314
Fashion - Staff Pick          7     31       9     14      22         2     6     5     13           9         3     5      9             56            238
Film & Video - Staff Pick    12     35       5     34      27         3     7    32    209           6         2    10     12            229           1060
Food - Staff Pick             7     31       6      6      12         3     4     5     15          11         4     3      6             98            649
Games - Staff Pick           24     72       3     57      52         1    22    31     21          11         5    37     18            102            502
Journalism - Staff Pick       4     16       0      6       8         1     0     0     12           4         0     0      5             46            192
Music - Staff Pick           10     32       9     22      31         2    13    10     35           9         4     8     24            168            845
Photography - Staff Pick     11      6       3     24      38         2     9     8     14           7         2     6      7            112            190
Publishing - Staff Pick      68    104       8     52      57         9    22    16     42          12         6    20     19            405           1525
Technology - Staff Pick      29     46      10     35      52        32    17    35      6          20         6    11     13            103            606
Theater - Staff Pick          1      6       0      0       2         0     1     0     13           1         0     1      0             74            248
>
> Result <- chisq.test(Matriz,correct=FALSE)
Warning message:
In chisq.test(Matriz, correct = FALSE) :
  chi-squared approximation may be incorrect
> print(Result)

        Pearson's Chi-squared test

data:  Matriz
X-squared = 1962.4, df = 196, p-value < 2.2e-16
```

```
> Result$observed - Result$expected
                      Australia      Canada    Denmark     France     Germany  Hong Kong       Italy       Japan      Mexico Netherlands   Singapore      Spain     Sweden United Kingdom United States
Art - Staff Pick     -1.9197540 -14.2711602 -0.4570331  3.002613  -8.0439662 -2.2877786  6.76295158  1.3524981  14.2422752  -2.8674865 -1.42598002  3.4963874 -3.503613     13.4007686   -7.480707
Comics - Staff Pick  -3.1629516  64.6541222 -5.7156034 -32.231360 -32.8724058 -7.3466564  0.04458109 -14.0299769 -35.5073021 -10.7901614 -2.28823982 -10.3566487 -10.356649    -29.0092237  128.968486
Crafts - Staff Pick   1.7403536  -2.6398155  1.5305150  -1.933743   3.0617986 -1.4136818 -1.71575711  4.8192160  27.8482706  -0.9345119  3.20015373 -2.4739431 -1.473943      4.5136049  -34.128517
Dance - Staff Pick   -3.9076095  -7.7604919 -1.3480400  -3.443351  -5.3647963 -1.2968486 -1.49131437 -2.9179093  -4.3953882  -1.7746349 -0.73374327 -0.2694850 -1.269485    -11.6362798   47.609377
Design - Staff Pick   3.5445042   3.8840891 11.3930822  20.435357  17.9698693 13.5300538  3.33405073 -4.8073789 -18.4633359   3.2516526  1.03674097  6.9275942  1.927594      0.3515757  -64.315450
Fashion - Staff Pick -0.5511914  12.1385088  6.3950038   3.481091   9.7004612 -0.5060723  1.18570331 -0.6386626  -3.2235204   5.5706380  1.58209070  0.6143736  4.614374     -5.1349731  -35.227825
Film & Video - Staff Pick -17.6239047 -38.99508071 -5.2196003  -7.266487 -21.2520369 -6.8315142 -11.88685626  9.8790930 145.3538816  -7.4536510 -3.56256726 -7.2051499 -5.205150    -10.8372022  -11.893774
Food - Staff Pick     -8.1375865  -6.81091468  0.7778632 -15.086856 -12.6564181 -2.0238278  -5.65103766 -6.3036126 -17.5226749   4.1252882  1.15757110 -5.7916987 -2.791699    -24.5549577  101.270561
Games - Staff Pick     7.1374327  29.88039969 -2.8172175  33.510223  24.5338970 -4.5963105  11.24919293 18.4083013 -15.2287471   3.3418909  1.83366641 27.2064566  8.206457    -34.5205227 -108.145119
Journalism - Staff Pick -1.1749424   3.07394312 -1.7852421  -1.208762  -0.4290546 -0.7174481 -3.29930822 -3.8642583   0.8817832   1.6498078 -0.97171407 -3.0055342  1.994466      4.1033051    4.752959
Music - Staff Pick   -11.5094543 -21.72667179  1.5797079  -7.962952  -4.0350500 -5.1385088 -0.71345119 -6.0616449 -11.2124520  -0.7684858 -0.03889316 -4.4923905 11.507610    -6.1420446   66.714681
Photography - Staff Pick  3.2727902 -13.30115296  0.3342813  13.235895  25.4137586 -0.5644889  4.07348194  2.2299001  -2.6016910   3.4906995  0.54903920  1.5121445  2.512145     49.4399693  -89.596772
Publishing - Staff Pick  26.3716372   0.01998463 -6.3608762  -5.988855 -10.8051499 -4.8155265 -4.54035357 -15.0849347 -47.4373559  -6.9054573 -1.81667948 -4.1771714 -5.177171     67.9738663   18.744043
Technology - Staff Pick  11.0285165   1.11053036  3.8002306   9.965488  22.7276710 26.0356649  5.54219831  21.5802460 -32.6112221  11.8382782  2.62544197  0.5624135  2.562414    -42.4983859  -44.269485
Theater - Staff Pick     -5.1078401  -9.25626441 -2.1070715  -8.508301  -7.9485780 -2.0270561 -2.89408148 -4.5608762  -0.1225211  -1.7738663 -1.14688701 -2.5473482 -3.547348     24.5504996   26.997540
```

Figure 9 Chi-Square Test Result of Main Category by Country of Origin for Staff Picked Projects

From EDA, we observed that the Kickstarter staff appears to exhibit preferences for projects in certain countries for some categories. Conducting a Chi-Square Test, we found that there is significant statistical proof that there association exists between Main Category and Country of Origin for Staff Picked Projects. From the contingency table of observed minus expected number of Staff Picked projects for each Main Category by Country of Origin, we observed that staff tend to pick proportionately more Film & Video projects from Japan and Mexico, and Technology projects from Germany, Hong Kong and Japan.

4.5. Number of projects by creator

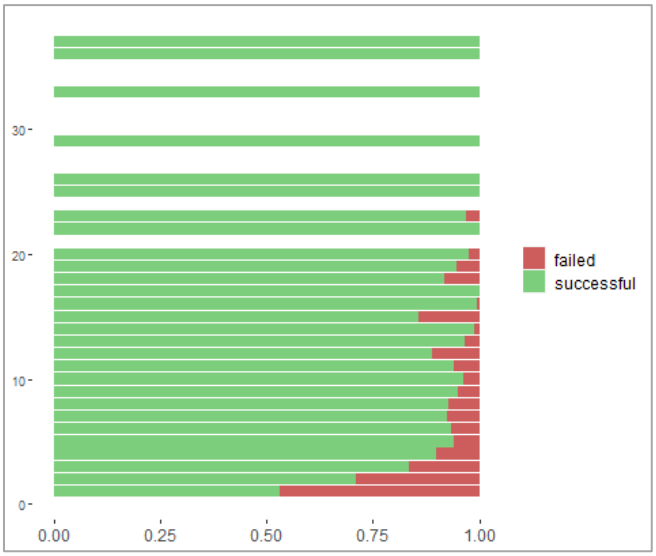


Figure 10 Number of Projects by Creator

The proportion of failure decreases with more projects created by the creator. Creators with higher number of projects created have a higher proportion of successful projects which could be attributed to their track record and reputation and this can be observed in all countries.

4.6. Correlation study on numerical variables

A correlation study was performed on continuous variables to identify variables that have strong linear correlation with one another. We observed a strong linear correlation in the following 4 groups of variables see Figure 11. Within the groups, many of the variables can be derived from one another. Hence, it is not necessary to keep all of them.

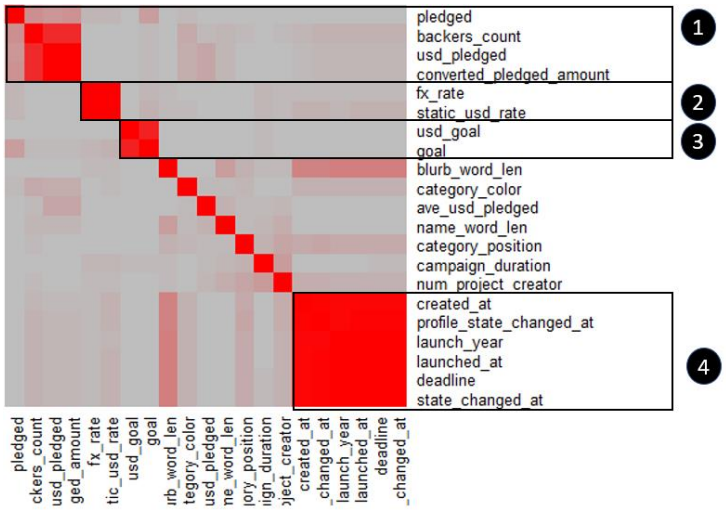


Figure 11 Correlation Study on numerical variables

4.7. Final Dataset

Our final dataset used for predictive modelling consists of the following variables.

For the MLR model on predicting total pledged amount, the model contained the variables seen in Table 2.

Table 2 Final Dataset to predict total pledged amount

	Variable	Description
Dependent Variable	usd_pledged	
Independent Variable	category_parent_name_recode	Main Category of the Project
	location_country / location_expanded_country	Country Code of Country of Origin / Country Name
	campaign_duration	Duration of Project in days
	launch_month	Month on which the project was launched
	launch_day	Day of the week on which the project was launched
	backers_count	Number of Backers at campaign completion
	name_word_len	Number of words in the Project Name
	blurb_word_len	Number of words in the Blurb
	staff_pick	True - Someone on the Kickstarter team likes the project and the project is being featured at “Staff Picks” and “New & Noteworthy” projects False – Project is not picked by the Kickstarter team to be featured
	num_project_creator	Number of Projects created by the Creator

For the Logistic Regression model predicting success, the model contained the variables seen in Table 3. For the staff_pick model, the same variables were used with the exception of staff_pick as the independent variable.

Table 3 Final dataset to predict success state and staff pick

	Variable	Description
Dependent Variable	State	Successful – Pledged Amount has reached or exceeded the Goal upon the End of Funding Period Failed – Pledged Amount does not meet the Goal
Independent Variable	category_parent_name_recode	Main Category of the Project
	location_country / location_expanded_country	Country Code of Country of Origin
	campaign_duration	Duration of Project in days
	launch_month	Month on which the project was launched
	launch_day	Day of the week on which the project was launched
	num_project_creator	Number of Projects created by the Creator
	name_word_len	Number of words in the Project Name
	blurb_word_len	Number of words in the Blurb
	staff_pick	True - Someone on the Kickstarter team likes the project and the project is being featured at “Staff Picks” and “New & Noteworthy” projects False – Project is not picked by the Kickstarter team to be featured
	usd_goal	Goal Amount converted to USD

5. Data Analysis - Predictive Modelling

After performing Exploratory Descriptive Analysis on the dataset, regression models were built to run our predictive modelling. In this analysis, we predict the pledged amount using multiple linear regression while we predict the probability of a project being Staff Picked and probability of campaign success using logistic regression. This helps to address the key concern of creators at different stages of their Kickstarter projects. The success rate model can also be useful to backers who may want to assess the likelihood of a project succeeding before becoming too invested in the project.

1. MLR to predict total pledged amount (can be used as a proxy for fundraising goal)
2. Logit to predict Staff Pick
3. Logit
4. to predict Success Rate

5.1. Multiple Linear Regression

Before campaign launch, creators need to decide on the optimal amount of money to set for as their goal. This needs to cover the expenses needed to execute the project and at the same time should be as realistic as possible because Kickstarter will release the funds only when goal has been met. Creators may select variables that they have visibility into as inputs to generate a MLR model to predict total pledged amount based on historical project data. After the launch of the campaign, when creators have more information on ongoing variables such as the number of backers and staff pick, the same approach can be adopted to derive a new prediction model with such variables as additional inputs. This would help creators decide if they should continue with the campaign, cancel their campaign or finetune certain aspects to save or increase their project’s likelihood of success.

$$Y = \beta_{0} + \sum \beta_{i} . X_{i}$$

β_0 : intercept
β_i : Co-efficient
X_i : Explanatory variables

We derived our model for predicting total pledged amount by first fitting the explanatory variables to ordinary least square MLR model. Next, we ran stepwise regression and the model with smallest AIC is chosen as the best model.

5.1.1 Final Model and Result

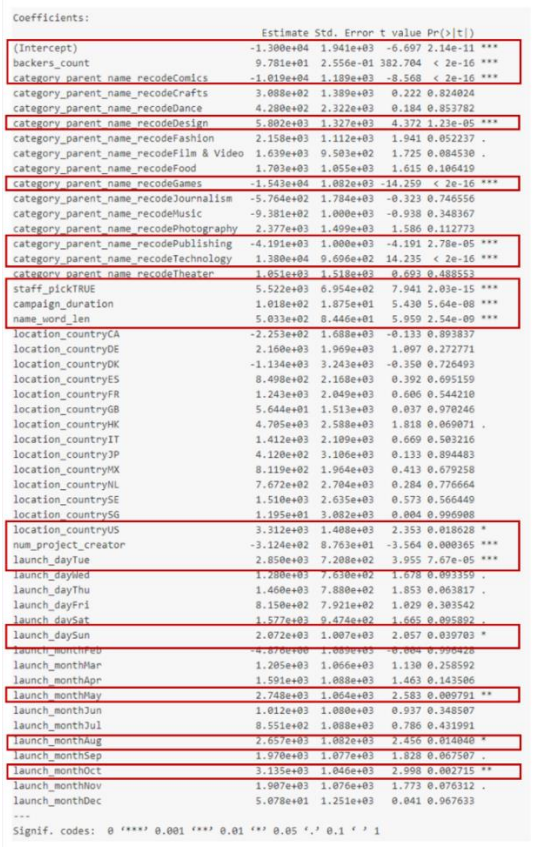
```
lm(formula = usd_pledged ~ backers_count + category_parent_name_recode +
    staff_pick + campaign_duration + name_word_len + location_country +
    num_project_creator + launch_day + launch_month, data = mlr_data)
```

Figure 12 MLR model to predict total pledged amount

```
Residual standard error: 71450 on 103782 degrees of freedom
Multiple R-squared:  0.6033,    Adjusted R-squared:  0.6031
F-statistic:  3156 on 50 and 103782 DF,  p-value: < 2.2e-16
```

Figure 13 MLR Model Result

5.1.2 Parameter Estimates



```
Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                           -1.300e+04  1.941e+03  -6.697 2.14e-11 ***
backers_count                          9.781e+01  2.556e-01 382.704  < 2e-16 ***
category_parent_name_recodeComics     -1.019e+04  1.189e+03  -8.568  < 2e-16 ***
category_parent_name_recodeCrafts      3.088e+02  1.389e+03   0.222 0.824024
category_parent_name_recodeDance       4.280e+02  2.322e+03   0.184 0.853782
category_parent_name_recodeDesign      5.802e+03  1.327e+03   4.372 1.23e-05 ***
category_parent_name_recodeFashion     2.158e+03  1.112e+03   1.941 0.052237 .
category_parent_name_recodeFilm & Video 1.639e+03  9.503e+02   1.725 0.084530 .
category_parent_name_recodeFood        1.703e+03  1.055e+03   1.615 0.106419
category_parent_name_recodeGames      -1.543e+04  1.082e+03 -14.259  < 2e-16 ***
category_parent_name_recodeJournalism -5.764e+02  1.784e+03  -0.323 0.746556
category_parent_name_recodeMusic      -9.381e+02  1.000e+03  -0.938 0.348367
category_parent_name_recodePhotography 2.377e+03  1.499e+03   1.586 0.112773
category_parent_name_recodePublishing -4.191e+03  1.000e+03  -4.191 2.78e-05 ***
category_parent_name_recodeTechnology  1.380e+04  9.696e+02  14.235  < 2e-16 ***
category_parent_name_recodeTheater     1.051e+03  1.518e+03   0.693 0.488553
staff_pickTRUE                         5.522e+03  6.954e+02   7.941 2.03e-15 ***
campaign_duration                      1.018e+02  1.875e+01   5.430 5.64e-08 ***
name_word_len                          5.033e+02  8.446e+01   5.959 2.54e-09 ***
location_countryCA                    -2.253e+02  1.688e+03  -0.133 0.893837
location_countryDE                     2.160e+03  1.969e+03   1.097 0.272771
location_countryDK                    -1.134e+03  3.243e+03  -0.350 0.726493
location_countryES                     8.498e+02  2.168e+03   0.392 0.695159
location_countryFR                     1.243e+03  2.049e+03   0.606 0.544210
location_countryGB                     5.644e+01  1.513e+03   0.037 0.970246
location_countryHK                     4.705e+03  2.588e+03   1.818 0.069071 .
location_countryIT                     1.412e+03  2.109e+03   0.669 0.503216
location_countryJP                     4.120e+02  3.106e+03   0.133 0.894483
location_countryMX                     8.119e+02  1.964e+03   0.413 0.679258
location_countryNL                     7.672e+02  2.704e+03   0.284 0.776664
location_countrySE                     1.510e+03  2.635e+03   0.573 0.566449
location_countrySG                     1.195e+01  3.082e+03   0.004 0.996908
location_countryUS                     3.312e+03  1.408e+03   2.353 0.018628 *
num_project_creator                   -3.124e+02  8.763e+01  -3.564 0.000365 ***
launch_dayTue                          2.850e+03  7.208e+02   3.955 7.67e-05 ***
launch_dayWed                          1.288e+03  7.630e+02   1.678 0.093359 .
launch_dayThu                          1.460e+03  7.880e+02   1.853 0.063817 .
launch_dayFri                          8.150e+02  7.921e+02   1.029 0.303542
launch_daySat                          1.577e+03  9.474e+02   1.665 0.095892 .
launch_daySun                          2.072e+03  1.007e+03   2.057 0.039703 *
launch_monthFeb                       -4.870e+00  1.089e+03  -0.004 0.996428
launch_monthMar                        1.205e+03  1.066e+03   1.130 0.258592
launch_monthApr                        1.591e+03  1.088e+03   1.463 0.143506
launch_monthMay                        2.748e+03  1.064e+03   2.583 0.009791 **
launch_monthJun                        1.012e+03  1.080e+03   0.937 0.348507
launch_monthJul                        8.551e+02  1.088e+03   0.786 0.431991
launch_monthAug                        2.657e+03  1.082e+03   2.456 0.014040 *
launch_monthSep                        1.970e+03  1.077e+03   1.828 0.067507 .
launch_monthOct                        3.135e+03  1.046e+03   2.998 0.002715 **
launch_monthNov                        1.907e+03  1.076e+03   1.773 0.076312 .
launch_monthDec                        5.078e+01  1.251e+03   0.041 0.967633
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 14 MLR Model Result Parameter Estimation

5.1.3 Model Interpretation

Our MLR model has an adjusted R-square of 0.6031 which we consider to be moderately good at explaining the total pledged amount given the above set of independent variables. Other than Description Length (*blurb_word_len*), the rest of the variables are significant and went into the model.

A closer look at the granular level of categorical variables, we noticed that not all values are significant at a 5% significance level. For example, within Main Category (*category_parent_name_recode),* Comics, Design, Games, Publishing and Technology have p-value < 0.05, and have a significant impact on total pledged amount whereas the other categories are not statistically significant at an alpha of 0.05.

Our base pledged amount is a negative value of -13,000.

Number of Backers (*backers_count*) : +9.781, each additional backer count contributes $9.78 increment to pledged amount. Intuitively, a creator receives more funds with increasing number of backers.

Main Category : Among the significant categories, Design and Technology have positive coefficients, indicating a positive relationship with total pledged amount. Projects in these two categories are likely to receive more funding. In addition, from the descriptive study, we observed that they are more likely to get staff picked, contributing to their likelihood of success. On the other hand, Comics, Games and Publishing have negative coefficients, indicating a negative relationship with total pledged amount. Creators might

want to review their project thoroughly and decide on a realistic goal should they decide to launch project in these categories.

Staff Pick : +5,522, when a project is Staff Pick, it brings around an additional value of $5k to total pledged amount. Staff pick is an important element that creators should strike to achieve. Being featured by staff helps the project to gain higher visibility from potential backers. This may be contributing to the higher pledged amount and higher likelihood of success.

Campaign Duration: +100, as campaign duration gets longer, the number of days the creator has to gather funds increases. Hence, resulting in a positive relationship. However, the boxplot from our descriptive analysis indicated that the impact diminishes beyond 30 days. This indicates that a positive linear relationship might not be the best model to proxy the impact for campaign period beyond 30 days.

Launch Month and Launch Day: launching the campaign during May or Oct appears to bring around a higher pledged amount. Campaigns launched on Tuesday and Sunday are favored. We link this to pre-holiday and weekend effect. However, to understand the cause behind this, more data-points are required to conduct tests for various hypotheses before a more conclusive result can be obtained.

With the help of our Shiny App, creators can indicate preset values for certain variables, as well as fine-tune the threshold for variables that allows changes and receive instant feedback on the predicted result of total pledged amount.

5.2. Logistic Regression

In the Logit model, project main category has been preset under the assumption that creators have decided on the project content. We will use projects in the Arts category for the past 5 years as an example to illustrate the regression results for both staff pick and success state prediction models.

5.2.1 Staff Pick

The logistic regression model for predicting the probability of being picked by staff uses the following independent variables with a fixed main category.

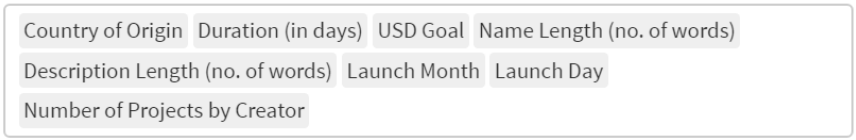


Figure 15 Explanatory Variables of Logistic Model – Staff Pick

Depending on the available information that the creator has, independent variables can be adjusted accordingly. After fitting in the independent variables, the model is run and generates a result summary as shown below.

```
Call:
glm(formula = staff_pick ~ ., family = binomial(link = "logit"),
    data = logit_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9907  -0.4244  -0.3721  -0.3116   2.8943

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)          -3.231e+00  3.828e-01  -8.439  < 2e-16 ***
location_countryCA   -5.186e-01  3.916e-01  -1.324 0.185451
location_countryDE    3.531e-01  4.230e-01   0.835 0.403811
location_countryDK    4.894e-01  6.077e-01   0.805 0.420642
location_countryES    4.843e-02  4.383e-01   0.110 0.912014
location_countryFR    5.477e-01  3.851e-01   1.422 0.155008
location_countryGB    6.130e-01  3.247e-01   1.888 0.059032 .
location_countryHK   -4.723e-01  7.831e-01  -0.603 0.546429
location_countryIT    6.081e-01  4.127e-01   1.473 0.140632
location_countryJP    1.802e+00  4.611e-01   3.908 9.31e-05 ***
location_countryMX    1.345e+00  3.541e-01   3.797 0.000147 ***
location_countryNL   -3.790e-01  6.645e-01  -0.570 0.568413
location_countrySE    4.133e-01  6.063e-01   0.682 0.495457
location_countrySG   -1.318e+00  1.053e+00  -1.251 0.210861
location_countryUS    4.698e-01  3.137e-01   1.498 0.134189
campaign_duration    -1.213e-03  3.313e-03  -0.366 0.714378
usd_goal             -6.499e-08  1.779e-07  -0.365 0.714823
name_word_len         1.041e-02  1.592e-02   0.654 0.513057
blurb_word_len        1.958e-02  6.624e-03   2.955 0.003123 **
launch_monthFeb      -1.005e-01  1.886e-01  -0.533 0.594181
launch_monthMar      -8.937e-02  1.833e-01  -0.488 0.625812
launch_monthApr       4.447e-02  1.819e-01   0.244 0.806893
launch_monthMay      -2.277e-01  1.871e-01  -1.217 0.223670
launch_monthJun      -3.736e-02  1.824e-01  -0.205 0.837748
launch_monthJul      -8.865e-02  1.829e-01  -0.485 0.627799
launch_monthAug      -1.134e-01  1.839e-01  -0.617 0.537329
launch_monthSep       2.823e-02  1.809e-01   0.156 0.876018
launch_monthOct       2.246e-01  1.687e-01   1.331 0.183056
launch_monthNov       2.776e-01  1.691e-01   1.642 0.100678
launch_monthDec      -2.380e-01  2.345e-01  -1.015 0.309993
launch_dayTue         2.114e-01  1.240e-01   1.704 0.088300 .
launch_dayWed        -3.288e-02  1.341e-01  -0.245 0.806358
launch_dayThu         1.609e-01  1.308e-01   1.230 0.218693
launch_dayFri        -2.669e-01  1.419e-01  -1.879 0.060188 .
launch_daySat        -4.348e-01  1.788e-01  -2.431 0.015044 *
launch_daySun        -2.756e-01  1.840e-01  -1.498 0.134111
num_project_creator -4.002e-02  1.955e-02  -2.047 0.040681 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5256.0  on 10059  degrees of freedom
Residual deviance: 5113.9  on 10023  degrees of freedom
AIC: 5187.9

Number of Fisher Scoring iterations: 7
```

Figure 16 Logistic Model - Parameter Estimates Staff Pick

The coefficients estimate and respective p-value at the last column indicates the direction of the change in predicted outcome and whether the variable is significant in determining the target variable. Positive coefficients indicate a one unit increase in the variable will lead to an increase in the probability of being staff picked or success rate. If the respective p-value is less than 0.05, this variable is a significant determinant in the prediction. The difference between null deviance and residual deviance (seen at the bottom of summary) indicates how good the model fit is. The bigger the difference, the better the model is. Finally, lower AIC values also indicate a better-fit model.

From the regression summary above, we can see that only Japan, Mexico, blurb word length, launched on Saturday and number of projects by creators are the variables that are significant in contributing to the probability of being staff picked within the Art category.

By running logistic regression on all the main categories and identifying the significant variables as shown in the Table 4, we have several observations. Firstly, countries like Japan and Mexico have a higher chance of being picked by the staff across most of the categories. Secondly, variables such as campaign duration, name length of the project and number of projects by creator are the top 3 variables that are significant in determining the probability of being staff picked. Lastly, projects launched during Friday and Saturday are generally more likely to be picked by the staff across all categories.

Table 4 Significant variables (with dummy for categorical variables levels) – Staff Pick

		comics	games	publishing	design	dance	theater	fashion	music	art	film&video	photography	crafts	technology	food	journalism
location_country	CA	√						√			√					
	DE		√		√			√	√		√					
	DK				√			√	√						√	
	ES													√		
	FR		√	√	√			√			√	√				
	GB	√		√				√	√		√				√	
	HK		√	√							√					
	IT													√		
	JP		√	√				√	√	√	√		√	√	√	
	MX	√	√					√	√	√	√		√	√		
	NL										√				√	
	SE		√	√				√	√		√					
	SG								√							
	US	√	√				√	√			√				√	
campaign_duration		√	√	√			√		√		√	√	√		√	√
usd_goal		√	√												√	
name_word_len		√		√				√	√		√	√	√	√	√	√
blurb_word_len			√	√					√	√		√			√	√
launch_month	feb										√					
	mar		√						√		√		√			
	apr															
	may	√									√		√			
	jun								√							
	jul															
	aug						√									
	sep															
	oct		√					√								
	nov		√	√												
	dec		√		√	√										
launch_day	tue		√	√							√			√	√	
	wed															√
	thu	√														
	fri	√	√	√	√		√				√			√		√
	sat	√	√	√	√		√		√	√	√		√	√		√
	sun	√	√	√		√			√		√			√		
num_project_creator		√	√	√	√	√		√	√	√		√		√		

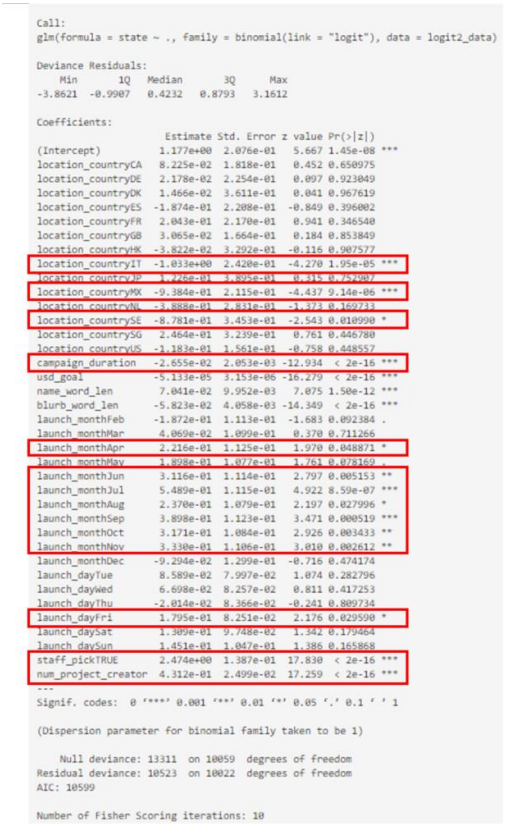
5.2.2 Campaign State

The logistic regression model for predicting campaign state (success or failure) uses the following independent variables with a fixed main category.

Country of Origin | Duration (in days) | USD Goal | Name Length (no. of words)
Description Length (no. of words) | Launch Month | Launch Day | Number of Projects by Creator

Figure 17 Explanatory Variables of Logistic Model – Campaign State

Depending on the available information that the creator has, the independent variables will be adjusted. After fitting in the independent variables, the model is run and will give the regression summary as shown below.



```
Call:
glm(formula = state ~ ., family = binomial(link = "logit"), data = logit2_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.8621  -0.9907   0.4232   0.8793   3.1612

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          1.177e+00  2.076e-01   5.667 1.45e-08 ***
location_countryCA   8.225e-02  1.818e-01   0.452 0.650975
location_countryDE   2.178e-02  2.254e-01   0.097 0.923049
location_countryDK   1.466e-02  3.611e-01   0.041 0.967619
location_countryES  -1.874e-01  2.208e-01  -0.849 0.396002
location_countryFR   2.043e-01  2.170e-01   0.941 0.346540
location_countryGB   3.065e-02  1.664e-01   0.184 0.853849
location_countryHK  -3.822e-02  3.292e-01  -0.116 0.907577
location_countryIT  -1.033e+00  2.420e-01  -4.270 1.95e-05 ***
location_countryJP   1.225e-01  3.895e-01   0.315 0.752807
location_countryMX  -9.384e-01  2.115e-01  -4.437 9.14e-06 ***
location_countryNL  -3.888e-01  2.831e-01  -1.373 0.169733
location_countrySE  -8.781e-01  3.453e-01  -2.543 0.010990 *
location_countrySG   2.464e-01  3.239e-01   0.761 0.446780
location_countryUS  -1.183e-01  1.561e-01  -0.758 0.448557
campaign_duration   -2.655e-02  2.053e-03 -12.934  < 2e-16 ***
usd_goal            -5.133e-05  3.153e-06 -16.279  < 2e-16 ***
name_word_len        7.041e-02  9.952e-03   7.075 1.50e-12 ***
blurb_word_len      -5.823e-02  4.058e-03 -14.349  < 2e-16 ***
launch_monthFeb     -1.872e-01  1.113e-01  -1.683 0.092384 .
launch_monthMar      4.069e-02  1.099e-01   0.370 0.711266
launch_monthApr      2.216e-01  1.125e-01   1.970 0.048871 *
launch_monthMay      1.898e-01  1.077e-01   1.761 0.078169 .
launch_monthJun      3.116e-01  1.114e-01   2.797 0.005153 **
launch_monthJul      5.489e-01  1.115e-01   4.922 8.59e-07 ***
launch_monthAug      2.370e-01  1.079e-01   2.197 0.027996 *
launch_monthSep      3.898e-01  1.123e-01   3.471 0.000519 ***
launch_monthOct      3.171e-01  1.084e-01   2.926 0.003433 **
launch_monthNov      3.330e-01  1.106e-01   3.010 0.002612 **
launch_monthDec     -9.294e-02  1.299e-01  -0.716 0.474174
launch_dayTue        8.589e-02  7.997e-02   1.074 0.282796
launch_dayWed        6.698e-02  8.257e-02   0.811 0.417253
launch_dayThu       -2.014e-02  8.366e-02  -0.241 0.809734
launch_dayFri        1.795e-01  8.251e-02   2.176 0.029590 *
launch_daySat        1.305e-01  9.748e-02   1.342 0.179464
launch_daySun        1.451e-01  1.047e-01   1.386 0.165868
staff_pickTRUE       2.474e+00  1.387e-01  17.830  < 2e-16 ***
num_project_creator  4.312e-01  2.499e-02  17.259  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13311  on 10059  degrees of freedom
Residual deviance: 10523  on 10022  degrees of freedom
AIC: 10599

Number of Fisher Scoring iterations: 10
```

Figure 18 Logistic Model - Parameter Estimates - Campaign State

Similar to the approach used in interpretation for staff pick, different project categories yield different regression results. Again, we run logistic regression on all the main categories and identify the significant variables as shown in the Table 5 below.

Table 5 Significant variables (with dummy for categorical variables levels) - Campaign State

		comics	games	publishing	design	dance	theater	fashion	music	art	film&video	photography	crafts	technology	food	journalism
location_country	CA								√							
	DE		√	√											√	
	DK										√					
	ES	√		√												
	FR	√									√					
	GB			√					√		√					
	HK			√	√									√		
	IT	√		√			√	√	√	√	√	√	√	√		
	JP	√											√			
	MX	√	√	√	√		√	√	√	√		√	√	√		
	NL			√					√		√					
	SE						√			√					√	
	SG													√		
	US			√										√		
campaign_duration		√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
usd_goal		√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
name_word_len		√	√	√	√			√	√	√	√	√	√	√	√	√
blurb_word_len			√	√	√	√	√	√	√	√	√	√	√		√	
launch_month	feb			√					√		√				√	
	mar			√					√		√				√	
	apr		√	√			√			√		√			√	
	may		√	√											√	
	jun		√	√	√					√				√		√
	jul		√	√	√					√					√	
	aug		√	√	√					√						
	sep		√	√	√			√		√			√		√	
	oct		√	√	√			√	√	√	√	√			√	
	nov		√	√	√				√	√	√			√	√	√
	dec				√						√		√			
launch_day	tue		√		√			√			√			√	√	
	wed													√		
	thu														√	
	fri			√						√				√	√	
	sat			√					√		√			√		
	sun	√		√	√			√	√					√		
staff_pickTRUE		√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
num_project_creator		√	√	√	√	√	√	√	√	√		√	√	√	√	√

We observe that countries like Italy and Mexico have higher significance level across most of the project categories, however the coefficient for the different categories are negative for these two countries as shown in the example in Figure 18. Moreover, factors such as campaign duration, goal amount, word length of the project name, blurb word length, number of projects per creator and whether the project is picked by the staff are significant variables in determining the success of the project. These are consistent with our initial exploratory data analysis mentioned in “Descriptive Analysis”. Lastly, projects launched in October and November has higher probability of success.

6. Conclusion and Future Work

Through our study, we have developed an app which can help Kickstarter creators by providing a holistic dashboard to assess the performance of past projects, support in determining their suitable goal by predicting the pledge amount using multiple linear regression and finally deeper understanding of different factors of staff-picked and successful projects using logistics regression.

The input variables we used in our models are mainly quantitative. However, from our literature review, we acknowledge that qualitative factors are also critical to campaign success. This is an area of improvement for our study to increase accuracy of the models.

Some of the future works that can be done are:

- Text Analytics or text data mining to derive additional information from project name and project description
- For the variable 'Number of Projects created by the Creator', we should have taken into account the total number of Kickstarter projects launched by a creator instead of counting the number of projects the creator created during the 5-year period of the dataset. Further study should also be undertaken on certain project categories (e.g. games) to analyse its impact on success if a project has a successive series of projects.
- Need to consider training, validation and testing data

Additional information could be gathered:

- Social analytics to consider the impact of online ads particularly targeted and hyper personalized marketing campaigns on different social media platforms
- Backers' demographics including age, spending behaviour, gender, location, socioeconomic income which can be used for clustering and profiling
- Results of search engine optimization for the Kickstarter campaign webpage including layout, UI/UX design, backlinks and use of videos and images to entice possible backers

7. References:

- Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, *29*(1), 1–16. https://doi.org/10.1016/j.jbusvent.2013.06.005
- Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M.,…Han, J. (2011). Challenges and Opportunities with Big Data 2011–1. http://docs.lib.purdue.edu/cctech/1/.
- Greenberg, M. D., Pardo, B., Hariharan, K., & Gerber, E. (2013).Crowdfunding support tools: predicting success & failure. In
- CHI'13 Extended Abstracts on Human Factors in Computing Systems (pp. 1815–1820). http://dl.acm.org/citation.cfm?id=2468682
- Zvilichovsky, D., Inbar, Y., & Barzilay, O. (2015). Playing both sides of the market: success and reciprocity on crowdfunding platforms. http://papers.ssrn.com/abstract=2304101.
- Kuppuswamy, Venkat, and Barry L. Bayus. 2018. Crowdfunding creative ideas: The dynamics of project backers. In The Economics of Crowdfunding. Cham: Palgrave Macmillan, pp. 151–82.
- Ullah, S., & Zhou, Y. (2020). Gender, Anonymity and Team: What Determines Crowdfunding Success on Kickstarter. *Journal of Risk and Financial Management*, *13*(4), 80–. https://doi.org/10.3390/jrfm13040080
- Zhou, M., Lu, B., Fan, W., & Wang, G. (2018). Project description and crowdfunding success: an exploratory study. *Information Systems Frontiers*, *20*(2), 259–274. https://doi.org/10.1007/s10796-016-9723-1

8. Appendix

Data Preparation Change Log

No	Name	Issue Identified	Action
1	Unexpected JSON fields in csv file	There are columns remain embedded in JSON format	JSON transformation on fields encoded in JSON format
2	Timestamps are in epoch format	Epoch timestamp are not interpretable directly	Convert timestamp to UTC
3	Recode parent category	Missing parent category	Fill up with category name
4	Currency fields	Base to a standardised currency - USD	Multiply the amount by FX
5	New Fields	Existing fields cannot be used directly	Derive new fields

Feature Engineering – Derived Variables

Derived Variables	Description	Calculations
campaign_duration	Duration of Project in days	'deadline_utc' less 'launched_at_utc'
num_project_creator	Number of Projects created by the Creator	Count of number of projects attributed to the same 'creator_id'
ave_usd_pledged	Average Amount pledged by each Backer (as a proxy)	'usd_pledged' / 'backers_count'
launch_year	Year on which the project was launched	Year based on 'launched_at_utc'
launch_month	Month on which the project was launched	Month based on 'launched_at_utc'
launch_day	Day of the week on which the project was launched	Day based on 'launched_at_utc'
name_word_len	Number of words in the Project Name	Count of number of words in 'name'
name_detected_language	A guess of the language of the Project Name. NA if the language could not reliably be determined	
blurb_word_len	Number of words in the Blurb	Count of number of words in 'blurb'
blurb_detected_language	A guess of the language of the Blurb. NA if the language could not reliably be determined	
usd_goal	Goal Amount converted to USD	'goal' * 'static_usd_rate'

Chi-square test details

Chi-Square Test of Association between Country of Origin and Parent Category for Staff Picked Projects

```
> # Chi-Square Test of Association
>
> # H0: no association exists between Country of Origin and Parent Category for Staff Picked Projects
> # H1: association exists between Country of Origin and Parent Category for Staff Picked Projects
>
> R1 <- c(11,18,4,21,13,2,15,11,42,3,1,11,4,118,460)
> R2 <- c(25,135,4,7,13,2,18,7,25,2,3,6,6,199,1148)
> R3 <- c(6,8,3,4,10,0,1,8,37,1,4,0,1,39,120)
> R4 <- c(0,2,0,2,1,0,1,0,4,0,0,2,1,20,189)
> R5 <- c(14,30,15,35,35,17,10,3,4,8,3,13,8,85,314)
> R6 <- c(7,31,9,14,22,2,6,5,13,9,3,5,9,56,238)
> R7 <- c(12,35,5,34,27,3,7,32,209,6,2,10,12,229,1060)
> R8 <- c(7,31,6,6,12,3,4,5,15,11,4,3,6,98,649)
> R9 <- c(24,72,3,57,52,1,22,31,21,11,5,37,18,102,502)
> R10 <- c(4,16,0,6,8,1,0,0,12,4,0,0,5,46,192)
> R11 <- c(10,32,9,22,31,2,13,10,35,9,4,8,24,168,845)
> R12 <- c(11,6,3,24,38,2,9,8,14,7,2,6,7,112,190)
> R13 <- c(68,104,8,52,57,9,22,16,42,12,6,20,19,405,1525)
> R14 <- c(29,46,10,35,52,32,17,35,6,20,6,11,13,103,606)
> R15 <- c(1,6,0,0,2,0,1,0,13,1,0,1,0,74,248)
>
> rows <- 15
> Matriz <- matrix(c(R1,R2,R3,R4,R5,R6,R7,R8,R9,R10,R11,R12,R13,R14,R15),nrow=rows,byrow=TRUE)
>
> rownames(Matriz) <- c("Art - Staff Pick",
+                       "Comics - Staff Pick",
+                       "Crafts - Staff Pick",
+                       "Dance - Staff Pick",
+                       "Design - Staff Pick",
+                       "Fashion - Staff Pick",
+                       "Film & Video - Staff Pick",
+                       "Food - Staff Pick",
+                       "Games - Staff Pick",
+                       "Journalism - Staff Pick",
+                       "Music - Staff Pick",
+                       "Photography - Staff Pick",
+                       "Publishing - Staff Pick",
+                       "Technology - Staff Pick",
+                       "Theater - Staff Pick")
> colnames(Matriz) <- c("Australia",    "Canada",       "Denmark",      "France",       "Germany",      "Hong Kong",    "Italy",        "Japan",
       "Mexico",       "Netherlands",  "Singapore",    "Spain",        "Sweden",       "United Kingdom",       "United States")
> Matriz
```

	Australia	Canada	Denmark	France	Germany	Hong Kong	Italy	Japan	Mexico	Netherlands	Singapore	Spain	Sweden	United Kingdom	United States
Art - Staff Pick	11	18	4	21	13	2	15	11	42	3	1	11	4	118	460
Comics - Staff Pick	25	135	4	7	13	2	18	7	25	2	3	6	6	199	1148
Crafts - Staff Pick	6	8	3	4	10	0	1	8	37	1	4	0	1	39	120
Dance - Staff Pick	0	2	0	2	1	0	1	0	4	0	0	2	1	20	189
Design - Staff Pick	14	30	15	35	35	17	10	3	4	8	3	13	8	85	314
Fashion - Staff Pick	7	31	9	14	22	2	6	5	13	9	3	5	9	56	238
Film & Video - Staff Pick	12	35	5	34	27	3	7	32	209	6	2	10	12	229	1060
Food - Staff Pick	7	31	6	6	12	3	4	5	15	11	4	3	6	98	649
Games - Staff Pick	24	72	3	57	52	1	22	31	21	11	5	37	18	102	502
Journalism - Staff Pick	4	16	0	6	8	1	0	0	12	4	0	0	5	46	192
Music - Staff Pick	10	32	9	22	31	2	13	10	35	9	4	8	24	168	845
Photography - Staff Pick	11	6	3	24	38	2	9	8	14	7	2	6	7	112	190
Publishing - Staff Pick	68	104	8	52	57	9	22	16	42	12	6	20	19	405	1525
Technology - Staff Pick	29	46	10	35	52	32	17	35	6	20	6	11	13	103	606
Theater - Staff Pick	1	6	0	0	2	0	1	0	13	1	0	1	0	74	248

```
>
> Result <- chisq.test(Matriz,correct=FALSE)
Warning message:
In chisq.test(Matriz, correct = FALSE) :
  Chi-squared approximation may be incorrect
> print(Result)

        Pearson's Chi-squared test

data:  Matriz
X-squared = 1962.4, df = 196, p-value < 2.2e-16
```

```
> Result$observed
```

	Australia	Canada	Denmark	France	Germany	Hong Kong	Italy	Japan	Mexico	Netherlands	Singapore	Spain	Sweden	United Kingdom	United States
Art - Staff Pick	11	18	4	21	13	2	15	11	42	3	1	11	4	118	460
Comics - Staff Pick	25	135	4	7	13	2	18	7	25	2	3	6	6	199	1148
Crafts - Staff Pick	6	8	3	4	10	0	1	8	37	1	4	0	1	39	120
Dance - Staff Pick	0	2	0	2	1	0	1	0	4	0	0	2	1	20	189
Design - Staff Pick	14	30	15	35	35	17	10	3	4	8	3	13	8	85	314
Fashion - Staff Pick	7	31	9	14	22	2	6	5	13	9	3	5	9	56	238
Film & Video - Staff Pick	12	35	5	34	27	3	7	32	209	6	2	10	12	229	1060
Food - Staff Pick	7	31	6	6	12	3	4	5	15	11	4	3	6	98	649
Games - Staff Pick	24	72	3	57	52	1	22	31	21	11	5	37	18	102	502
Journalism - Staff Pick	4	16	0	6	8	1	0	0	12	4	0	0	5	46	192
Music - Staff Pick	10	32	9	22	31	2	13	10	35	9	4	8	24	168	845
Photography - Staff Pick	11	6	3	24	38	2	9	8	14	7	2	6	7	112	190
Publishing - Staff Pick	68	104	8	52	57	9	22	16	42	12	6	20	19	405	1525
Technology - Staff Pick	29	46	10	35	52	32	17	35	6	20	6	11	13	103	606
Theater - Staff Pick	1	6	0	0	2	0	1	0	13	1	0	1	0	74	248

```
> Result$expected
```

	Australia	Canada	Denmark	France	Germany	Hong Kong	Italy	Japan	Mexico	Netherlands	Singapore	Spain	Sweden	United Kingdom	United States
Art - Staff Pick	12.919754	32.271176	4.457033	17.997387	21.043966	4.287779	8.237048	9.647502	27.757725	5.867487	2.4259800	7.503613	7.503613	104.59923	467.4807
Comics - Staff Pick	28.162952	70.345888	9.715603	39.231360	45.872406	9.346656	17.955419	21.029977	60.507302	12.790161	5.2882398	16.356649	16.356649	228.00922	1019.0315
Crafts - Staff Pick	4.259646	10.639816	1.469485	5.933743	6.938201	1.413682	2.715757	3.180784	9.151729	1.934512	0.7998463	2.473943	2.473943	34.48640	154.1285
Dance - Staff Pick	3.907610	9.760492	1.348040	5.443351	6.364796	1.296849	2.491314	2.917909	8.395388	1.774635	0.7337433	2.269485	2.269485	31.63628	141.3906
Design - Staff Pick	10.455496	26.115911	3.606918	14.564643	17.030131	3.469946	6.665949	7.807379	22.463336	4.748347	1.9632590	6.072406	6.072406	84.64842	378.3154
Fashion - Staff Pick	7.551191	18.861491	2.604996	10.518909	12.299539	2.506072	4.814297	5.638663	16.223520	3.429362	1.4179093	4.385626	4.385626	61.13497	273.2278
Film & Video - Staff Pick	29.623905	73.995081	10.219600	41.266487	48.252037	9.831514	18.886856	22.120907	63.646118	13.453651	5.5625673	17.205150	17.205150	239.83720	1071.8938
Food - Staff Pick	15.137586	37.810915	5.222137	21.086856	24.656418	5.023828	9.651038	11.303613	32.522675	6.874712	2.8424289	8.791699	8.791699	122.55496	547.7294
Games - Staff Pick	16.862567	42.119600	5.817218	23.489777	27.466103	5.596311	10.750807	12.591699	36.228747	7.658109	3.1663336	9.793543	9.793543	136.52052	610.1451
Journalism - Staff Pick	5.174942	12.926057	1.785242	7.208762	8.429055	1.717448	3.299308	3.864258	11.118217	2.350192	0.9717141	3.005534	3.005534	41.89669	187.2470
Music - Staff Pick	21.509454	53.726672	7.420292	29.962952	35.035050	7.138509	13.713451	16.061645	46.212452	9.768486	4.0388932	12.492390	12.492390	174.14204	778.2853
Photography - Staff Pick	7.727210	19.301153	2.665719	10.764105	12.586241	2.564489	4.926518	5.770100	16.601691	3.509301	1.4509608	4.487855	4.487855	62.56003	279.5968
Publishing - Staff Pick	41.628363	103.980015	14.360876	57.988855	67.805150	13.815527	26.540354	31.084935	89.437356	18.905457	7.8166795	24.177171	24.177171	337.02613	1506.2560
Technology - Staff Pick	17.971483	44.889470	6.199769	25.034512	29.272329	5.964335	11.457802	13.419754	38.611222	8.161722	3.3745580	10.437586	10.437586	145.49839	650.2695
Theater - Staff Pick	6.107840	15.256264	2.107071	8.508301	9.948578	2.027056	3.894081	4.560876	13.122521	2.773866	1.1468870	3.547348	3.547348	49.44950	221.0025

```
> Result$observed - Result$expected
```

	Australia	Canada	Denmark	France	Germany	Hong Kong	Italy	Japan	Mexico	Netherlands	Singapore	Spain	Sweden	United Kingdom	United States
Art - Staff Pick	-1.9197540	-14.27117602	-0.4570331	3.002613	-8.0439662	-2.2877786	6.76295158	1.3524981	14.2422752	-2.8674865	-1.42598002	3.4963874	-3.503613	13.4007686	-7.480707
Comics - Staff Pick	-3.1629516	64.65411222	-5.7156034	-32.231360	-32.8724058	-7.3466564	0.04458109	-14.0299769	-35.5073021	-10.7901614	-2.28823982	-10.3566487	-10.356649	-29.0092237	128.968486
Crafts - Staff Pick	1.7403536	-2.63981553	1.5305150	-1.933743	3.0617986	-1.4136818	-1.71575711	4.8192160	27.8482706	-0.9345119	3.20015373	-2.4739431	-1.473943	4.5136049	-34.128517
Dance - Staff Pick	-3.9076095	-7.76049193	-1.3480400	-3.443351	-5.3647963	-1.2968486	-1.49131437	-2.9179093	-4.3953882	-1.7746349	-0.73374327	-0.2694850	-1.269485	-11.6362798	47.609377
Design - Staff Pick	3.5445042	3.88408916	11.3930822	20.435357	17.9698693	13.5300538	3.33405073	-4.8073789	-18.4633359	3.2516526	1.03674097	6.9275942	1.927594	0.3515757	-64.315450
Fashion - Staff Pick	-0.5511914	12.13850884	6.3950038	3.481091	9.7004612	-0.5060723	1.18570331	-0.6386626	-3.2235204	5.5706380	1.58209070	0.6143736	4.614374	-5.1349731	-35.227825
Film & Video - Staff Pick	-17.6239047	-38.99508071	-5.2196003	-7.266487	-21.2520369	-6.8315142	-11.88685626	9.8790930	145.3538816	-7.4536510	-3.56256726	-7.2051499	-5.205150	-10.8372022	-11.893774
Food - Staff Pick	-8.1375865	-6.81091468	0.7778632	-15.086856	-12.6564181	-2.0238278	-5.65103766	-6.3036126	-17.5226749	4.1252882	1.15757110	-5.7916987	-2.791699	-24.5549577	101.270561
Games - Staff Pick	7.1374327	29.88039969	-2.8172175	33.510223	24.5338970	-4.5963105	11.24919293	18.4083013	-15.2287471	3.3418909	1.83366641	27.2064566	8.206457	-34.5205227	-108.145119
Journalism - Staff Pick	-1.1749424	3.07394312	-1.7852421	-1.208762	-0.4290546	-0.7174481	-3.29930822	-3.8642583	0.8817832	1.6498078	-0.97171407	-3.0055342	1.994466	4.1033051	4.752959
Music - Staff Pick	-11.5094543	-21.72667179	1.5797079	-7.962952	-4.0350500	-5.1385088	-0.71345119	-6.0616449	-11.2124520	-0.7684858	-0.03889316	-4.4923905	11.507610	-6.1420446	66.714681
Photography - Staff Pick	3.2727902	-13.30115296	0.3342813	13.235895	25.4137586	-0.5644889	4.07348194	2.2299001	-2.6016910	3.4906995	0.54903920	1.5121445	2.512145	49.4399693	-89.596772
Publishing - Staff Pick	26.3716372	0.01998463	-6.3608762	-5.988855	-10.8051499	-4.8155265	-4.54035357	-15.0849347	-47.4373559	-6.9054573	-1.81667948	-4.1771714	-5.177171	67.9738663	18.744043
Technology - Staff Pick	11.0285165	1.11053036	3.8002306	9.965488	22.7276710	26.0356649	5.54219831	21.5802460	-32.6112221	11.8382782	2.62544197	0.5624135	2.562414	-42.4983859	-44.269485
Theater - Staff Pick	-5.1078401	-9.25626441	-2.1070715	-8.508301	-7.9485780	-2.0270561	-2.89408148	-4.5608762	-0.1225211	-1.7738663	-1.14688701	-2.5473482	-3.547348	24.5504996	26.997540