Predicting The Success of Crowdfunding Campaigns

Lina Arustamyan

May 2021

American University of Armenia

College of Business and Economics

Instructor: Vahram Ghushchyan, Hayk Kamalyan

Course: Capstone

Abstract

In recent years, more and more people started to seek funding for their startups by launching a project in crowdfunding platforms. However, not all campaigns always manage to achieve success (raise a target amount of money), because project creators usually are not informed how to start a campaign, what details about the project they should provide to potential investors, what features are the most important ones, etc. Thus, the main motivation of this research was to show how different features of a campaign can contribute to its success and predict success based on that information.

For that purpose, I scraped one of the most famous crowdfunding platforms, Kickstarter, and was able to collect about 20 features of crowdfunding campaigns. Besides the basic features of a campaign, a popularity factor was used as a factor of success quantified by number of google searches, that was possible to do in Google Trends. Results showed that goal size was the most important feature, followed by the number of reward tiers, duration and google trend searches. All the machine learning models conducted were able to reach more than 70% accuracy with the Random Forest being the best one.

Keywords: Crowdfunding, Kickstarter, Machine learning, Prediction, Success, Google Trends

Contents

Introduction

Crowdfunding platforms have become popular in these days as places where people can create projects and raise funds to transform their ideas into goods, as well as support other projects. Despite the rapid growth in the number of users and projects, the overall success rate of projects has been declining due to a lack of awareness among project developers about what factors lead to project success.

With the dataset scraped from reward-based crowdfunding platform Kickstarter, I firstly determined what makes a crowdfunding project successful and based on that predicted the success of campaign.Successful projects are those ones that were able to reach their funding goal, as Kickstarter follows "All or nothing" principle.

By conducting this analysis, I hope to inform creators and self-driven individuals who want to contribute to the world through their ideas. Especially considering the fact that the crowdfunding culture is underdeveloped in Armenia, I want to raise awareness and provide creators with all the necessary information they should take into consideration before starting a project in crowdfunding platforms.

Literature on modeling.

In general, literature on crowdfunding can be divided into two types. One group of researchers built models using basic information of the campaign.

One of the most popular papers in this area was published by **Mollick(2013)**. The author created a model that included basic campaign features such as the funding goal, the number of supporters, the category, and so on. After that, he measured the correlations between the model's features. One of the most important results was that goal and success rate have a negative relationship, implying that campaigns with higher funding goals have lower success rates. Another result was that the creator's wide social network, such as Facebook, has a positive effect on the campaign's success.

Adebola Lamidi(2017) conducted the analysis by collecting the 17 basic features for 41,965 crowdfunding campaigns from 14 categories for the period 2009-2011. In order to predict the success of crowdfunding campaigns Lamidi used k-Nearest Neighbors (kNN) algorithm, Random Forest and Linear Regression. As a result, Random Forest was the best classifier for the prediction, having the highest level of accuracy of 68%.

Mati Kucz(2018) similarly, using simple campaign features, tried to address the following three questions: what is the most important factor in a Kickstarter project's success, what type of project is most likely to succeed (Art, Photography, Technology, etc.), and are there any discernible distinctions between successful and failed projects. In addition to descriptive statistics, the author used a Random Forest Classifier to forecast campaign success. The funding goal was the most important factor in making a successful kickstarter project, based on the Random Forest model's feature importance. Music, Film Video, Dance, and Comics - many of the arts - were the most common. Finally, there were a number of distinctions between successful and failed projects.

The second group of researchers used neural network and natural language processing approach,social media details to predict the outcome of crowdfunding startup pitches, or used mentioned information combined with basic features of campaigns.

Nam Phan(2018) used data on project and user features, as well as NLP features from sections of the project containing text, to predict the success of crowdfunding campaigns. They chose the LightGBM model because it is fast and has a high-performance gradient boosting framework. For the prediction purposes dataset was separated into training set and testing set using the ratio of 80:20 and the model was applied to three different subsets. One contains only projects + creators features, one contains NLP features and the rest contains combination of all features above. Additionally, SHAP values(as a powerful tool for interpreting tree-based models) were used to calculate feature importance metrics. One of the main findings of this study was that project and creator features have more predictive power than NLP features in predicting the success of Kickstarter project. Moreover, creator's experience with supporting other campaigns was found to be important predictor, based on feature importance analysis.

J.C.Kaminsk(2019) predicted the outcome of crowdfunding campaigns using text,speech, and video metadata in 20,188 crowdfunding campaigns, which was aimed to understand crowdfunding from an investor's perspective.The findings emphasized that positive psychological language is especially important in environments where objective information is limited and where investment preferences are subjective.

In addition to the NLP features used in Phan's analysis such as content or risk section's word count,etc, **Shreyas Devalapurkar(2019)** calculated positivity score and reading ease score of a title to improve the accuracy of prediction models. Applying Linear Support Vector Classification and

Random Forest Classifier models he found that Reading Ease and Positivity of a campaign's title play a role in predicting campaign success, and got 66% accuracy in predicting campaign success or failure.

Another interesting approach to make a prediction was conducted by **L.Castro, T.Couto, N.Felix(2020)**. They proposed a solution that combines classification and regression to produce a prediction in two stages: stage 1 (launch) and stage 2 (implementation). For step one, they created a classifier that could accurately predict campaign performance with 71.2 percent accuracy before it was launched. This model was built using static features and meta classifiers that used text features as input. On stage 2, they provided a classification model based on dynamic features that could achieve 85% accuracy even before a campaign's duration reached 20%, for example. The model was solely based on funding features, showing that the Twitter features had no significant effect on overall accuracy.

Despite of the many studies conducted in this field, there were some important factors that weren't considered in the models, such as information on how many reward tiers creators provide to investors or whether the project is featured by Kickstarter or not, etc.

Moreover, I also used popularity factor as a predictor for success of a campaign to increase the accuracy of my prediction. My goal was to understand how the number of google searches starting from launching till the end of a particular project contribute to it's success.Thus, by using Google Trends I was able to calculate the average number of google searches of the category name of all projects.

Research Limitations

There were several limitations in data collection process, which can be expressed as follows:

- The website from where I scraped the data didn't provide the information on projects that were launched more than 1-2 years ago. Instead, for those projects it provided already scraped data with selected features. Thus, not to use only projects that were launched in years of crisis, I included also the data for 2019 with selected features. As a result, there were features that I scraped for 2020-2021 projects but had to drop them as there were no such variables in 2019 data downloaded from the website.

- Another limitation was with the Google permission to scrape google trend's data. Google limited the number of requests one can send to a server. As a result, the number of observations was reduced, because google trends was one of my main contribution to this topic.

Data and Methodology

The scraped dataset consists of data of 17813 kickstarter campaigns launched between December 2018 and March 2021. The data was scraped from Kicktraq website, which keeps all archived Kickstarter projects.

As my goal is to classify between projects that were able to reach their funding goal and those that weren't (successful and failed projects only), therefore, cancelled,live and suspended projects have been removed from the dataset.

For comparison purposes I took the projects with raised and goal amount of money presented in dollars. Finally, the dataset was formed by 3627 failed projects and 5605 successful projects. In total, 9232 projects were employed, which were mostly from US.

Besides the scraped data available in the website I have also used **Google Trends** as a measure of popularity of campaigns in order to understand whether it has any impact on the success or failure of the projects, the category names of campaigns was queried using google trends.

About 20 features of a project was scraped,that are presented in Table 1.

Table 1: Features of projects used in the analysis

Feature	Description
category	**15** categories of projects that was formed by combining the subcategories
country	Country of the creator
start date	Launching date of a project
end date	Closing date of a project
googletrends	Average number of google searches of subcategory name of a project from launching till closing
duration	Number of days between starting and closing a project
backer	Number of backers that contributed to a project
pledge_per_backer	Average amount of money pledged by a backer
funded	Amount of money a project received till closing a project
goal	The goal amount that was set by the creator
len_desc	Length of project's description(number of characters)
len_title	Length of project's title(number of characters)
num_update	Number of updates creators posted during a project
num_comment	Number of comments a project received
num_faq	Number of FAQs about a project
num_tiers	Number of reward tiers a creator offers to donors
num_pledge_backers	Overall number of backers that pledged by reward tier
backer_per_tier	Average number of backers per reward tier
featured	Whether the project was featured by Kickstarter or not
num_created_by_owner	Number of projects created by the owner of a project

The next step was to construct prediction models and understand which one is more accurate by using appropriate evaluation metrics. For the prediction purposes dataset was separated into training set and testing set using the ratio of 70:30.

Four main models were used that were further combined with principal component analysis(**PCA**) and **optimized** by finding optimal parameters with **GridSearch**.

Logistic regression

The first model I have looked at is a logistic regression. In order to predict which of two groups a data point belongs to, logistic regression was used as a binary classifier.

The censored parametric regression with logistic distribution assumes that the observed time y_j and the function vector x_j have a linear relationship, which is modeled as follows:

$$y_j = X_i\beta + \sigma\epsilon_j$$

where $\beta = (\beta_1, , \beta_m)^T$ is the coefficient vector, σ is an adjusted parameter, and ϵ_j follows a logistic distribution.

kNN Classifier

A k-nearest neighbors (kNN) classifier is the next model. kNN computes the gap between c and each known campaign c' given a new campaign c, its partial trajectory $M_i(c)_{i\in I}$, and a list of campaigns for which the ending state is known.

$$d_X(c, c') = \sqrt{\sum_{i\in X}(M_i(c) - M_i(c'))^2}$$

Then it chooses top$_{k,I}$(c), the k known campaigns that are the most similar to c in terms of the distance described above, and calculates the likelihood of success $\phi_{kNN}(c, I)$ of c as the average final state of these k closest neighbors.

$$\phi_{kNN}(c, I) = \frac{1}{k} \sum_{c' \in top_{k,I}(c)} F(c')$$

Random Forests

After that, the Random Forest classifier was used. The Random Forest algorithm is a supervised learning algorithm based on classification. It generates a "forest" out of a set of decision trees that are usually trained using the "bagging" process. The bagging approach is based on the premise that combining different learning models increases the overall outcome.

XGBoost

This is a gradient boosting algorithm in its most basic form. It's an ensemble approach that generates multiple decision trees to improve data point classification, similar to Random Forests, except it uses gradient descent (first-order iterative optimization algorithm for finding a local minimum of a differentiable function) to improve the model's efficiency for the data points that are especially difficult to classify.

Performance evaluation

The weighted average **F1 score** was chosen as an evaluation method. The F1 score measures the harmonic mean between **precision and recall**, and is a suitable measure because there is no preference for false positives or false negatives in this case (both are equally bad). The number of positive class predictions that actually belong to the positive class is measured by precision, while number of positive class predictions made out of all positive examples in the dataset is measured by recall.The weighted average was used and both successes and failures was predicted.

Discussion of the results

The analysis was divided into two parts. In the first part I discuss how successful and failed projects differ in terms of the collected features of a project. Then I discuss the procedure of constructing and evaluating a number of prediction models.

Descriptive Statistics

Table 2 contains descriptive statistics on the variables used in the success prediction of crowdfunding campaigns. As can be observed there is a significant variation across all the features explaining success of overall 9232 crowdfunding campaigns. For instance, there are projects with the funding goal being just $ 1 ranging up to projects that set the funding goal to $10 mln. Another example is the popularity factor of the projects, there are topics with almost no popularity in google in contrast with the topics being searched almost 100 times usually during a month.

Table 2: Descriptive Statistics

Features	count	mean	std	min	25%	50%	75%	max
googletrends	9,232	52	24	1	32	53	73	95
duration	9,232	33	13	1	28	30	35	121
backer	9,232	251	1,424	1	10	42	144	88,884
pledge_per_backer	9,232	70	108	1	24	44	77	2,516
funded	9,232	20,829	161,688	1	402	2,209	8,829	11,385,074
goal	9,232	20,270	191,464	1	1,000	4,500	11,000	10,000,000
len_desc	9,232	93	36	1	64	99	126	255
len_title	9,232	34	15	3	22	33	47	72
num_update	9,232	5	6	-	-	3	7	86
num_comment	9,232	42	530	-	-	1	9	34,621
num_faq	9,232	1	3	-	-	-	-	43
num_tiers	9,232	8	6	-	4	7	10	106
num_pledge_backers	9,232	192	839	-	7	36	132	36,160
backer_per_tier	9,232	27	131	-	1	5	16	5,856
num_created_by_owner	9,232	2	4	1	1	1	1	61

Before going to prediction models it was important to understand how successful and failed projects differ and see the distribution of successful and failed projects by year of launching (see Figure 1 and Figure 2).

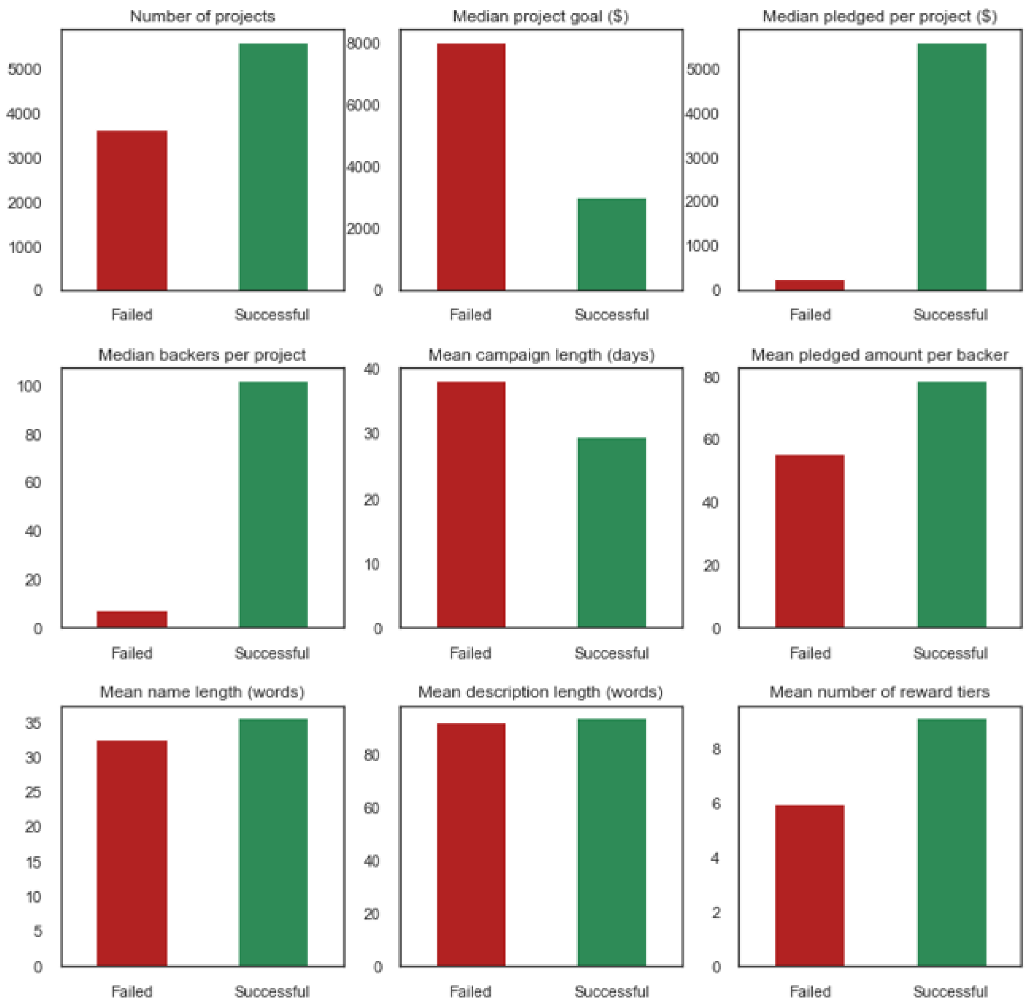

Figure 1: Differences between successful and failed projects

- Unsurprisingly, successful projects tend to have smaller goals - the median amount sought by successful projects is smaller than half that of failed projects.

- The median amount pledged per successful project is significantly higher than the median amount required, indicating that successful projects are more likely to receive additional funding.
- Successful projects have slightly shorter durations.
- Average title and description lengths are very similar between failed and successful projects.
- Successful projects tend to have more reward tiers suggested.
- Successful projects tend to have considerably more updates, comments, faqs, which means that projects that provide more information about their projects attract more investors.
- Owners of the projects that already have an experience with launching crowdfunding campaigns tend to be more successful.
- The differences in the google trend searches per project are more surprising. Failed projects have slightly more searches on average. One of the explanations can be that people want to fund more unique projects (for the features not shown in FIgure 1 see Appendix).

The proportion of successful projects is the highest in 2021 of about 80%, while for example in 2020 the number of successful and failed projects are almost equal.

What types of projects do people launch and which are more successful? (see Figure 3)

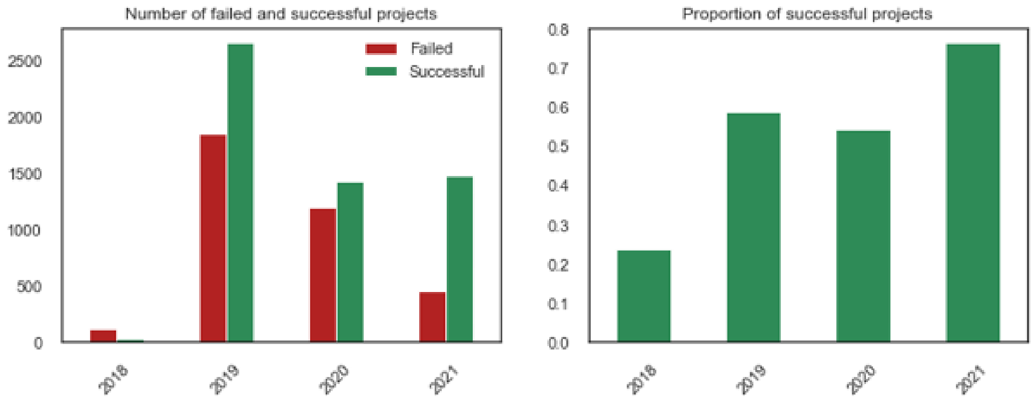

Figure 2: Successful and failed projects by year of launching

- There are 15 project categories, of which comics is the most common, followed by games and art, while the most successful projects are **games and art** projects.
- Technology projects have the highest goals by far, followed by food (e.g. funding for restaurants), with other categories generally much smaller in terms of their funding goals.
- However, in terms of the median amount currently pledged, technology ventures rank towards the bottom of the leaderboard, indicating that projects with smaller targets are more likely to succeed.
- Games, dance and design projects obtain the greatest amount of funding, on average (median).
- Comics and games tend to attract the most backers, but each backer tends to pledge less.
- Design and technology tend to attract the most generous backers.
- Comics and art projects suggest more reward tiers.
- Comics and games project owners put more updates about the project

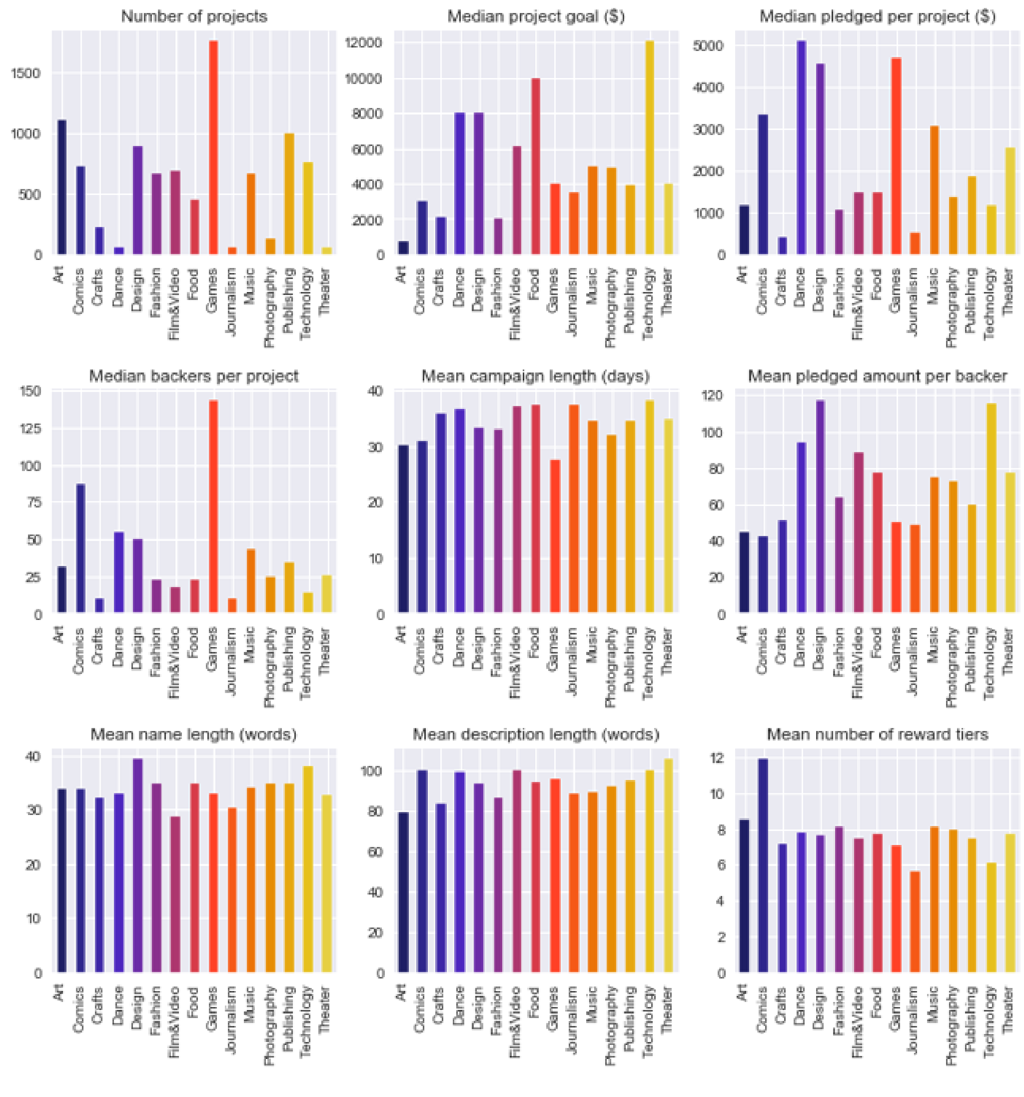

Figure 3: Feature performances by categories of the projects

process, which increase the interest of investors, which can be observed by their comments

• Journalism and fashion projects are the most searchable (for the features not shown in Figure 3 see Appendix).

When is the best time to launch a campaign?

To present a few more interesting findings I generated two more variables showing project's launching month and day. As a result, the following findings were obtained (see Figure 4).

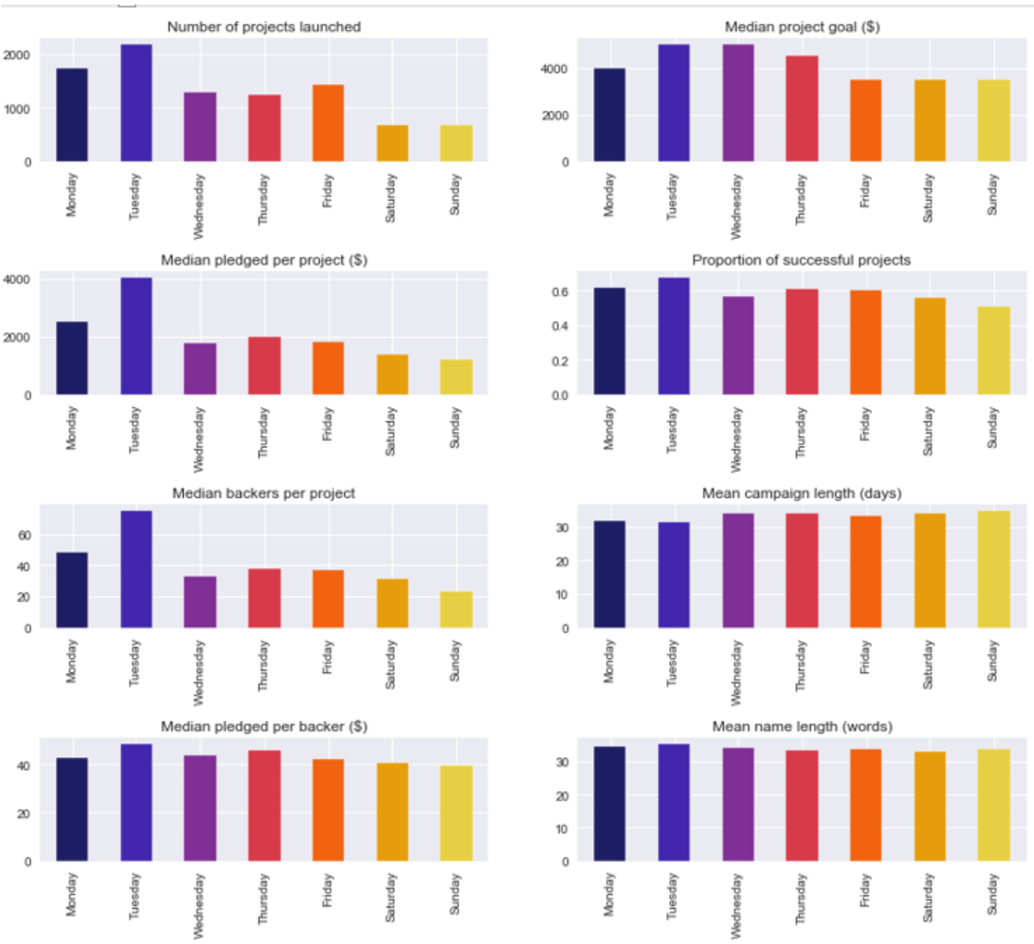

Figure 4: Feature performances by day of launching

• **Tuesday** appears to be the best day to launch a project. It is the most popular launch day, and has the highest proportion of successful projects, the most backers, the highest median amount pledged per backer, and the highest median pledge amount overall.

- Weekends are the least popular days to launch a project, attract less money, have fewer backers, receive smaller pledges per backer, and are slightly less successful.

- The most popular month to launch a project is January, and the least common is October.

- Median goal sizes are roughly similar throughout most of the year, but smaller for projects launched in January.

- The best month to launch in is **February**, which has the second highest median amount pledged per project, the highest success rate and the highest number of backers per project (see Appendix).

Prediction Models

Some features were retained for descriptive purposes, but in order to use machine learning models they need to be dropped. This includes datetime features, features that are related to outcomes (e.g. the amount pledged and the number of backers, etc.) rather than related to the basic features of the project itself (e.g. category, goal, length of description),etc.

After dropping the variables, the data was checked for multicollinearity and distribution of variables.

Except for title length and googletrends, most continuous numerical features are heavily positively skewed (see Appendix). For the first few models, these features were not log-transformed because this was not a problem for some machine learning models. Models were then rerun with log-transformed data to see if this improved model accuracy.

Multicollinearity was investigated by looking at correlations between predictor attributes, as this can cause problems with certain models. Correlation matrix below shows that it is not an issue:

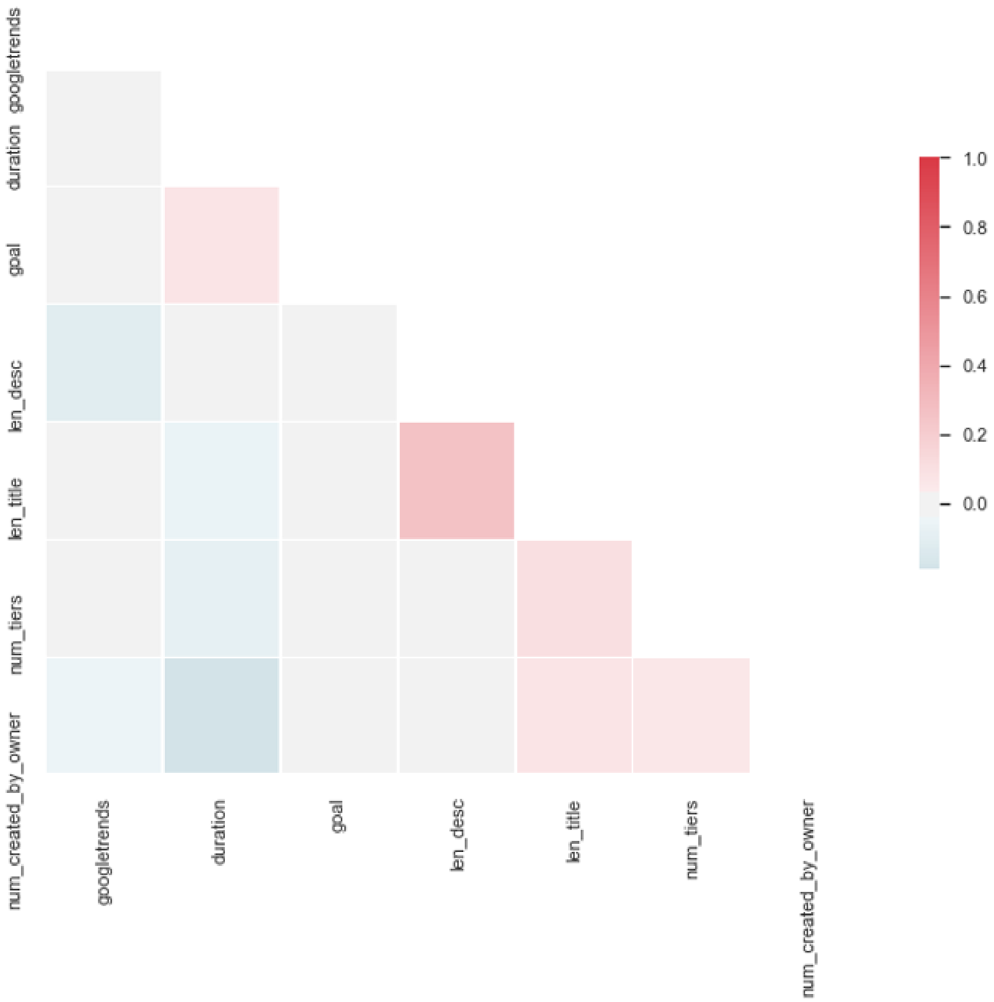

Figure 5: Correlation matrix

Before proceeding with prediction models I run simple linear regression(see Python output in Appendix) to understand which factors have a significance impact on success and use them in prediction models. After dropping not significant variables the following model was left.

Table 3: OLS results

	OLS
Features	**Coefficients**
const	0.813
googletrends	-0.0009
duration	-0.0091
goal	-1.263E-07
len_title	0.0019
num_tiers	0.0156
num_created_by_owner	0.0047
category_Comics	-0.0654
category_Crafts	-0.1993
category_Design	-0.107
category_Fashion	-0.1194
category_Film&Video	-0.1559
category_Food	-0.243
category_Journalism	-0.1901
category_Photography	-0.1782
category_Publishing	-0.1445
category_Technology	-0.2455
category_Theater	-0.0184
featured	0.3076
N	9232
R^2	0.25

Note: All coefficients are significant with p<0.01, except categoryTheater

The results are somewhat similar with the findings from descriptive part. For example, funding goal, duration have a negative relationship with success rate, while number of created projects by owner positively contribute to a success.

Four different machine learning models for classification with the additional improvements were applied to the data, such as **PCA and parameter optimization using GridSearch**, in order to create a model to classify projects into successes and failures. Below are the results (evaluation metrics) of models applied with a splited data by 70:30 proportion.

Table 4: Performance of Prediction Models

Models	Precision		Recall		F1 score	
	Train	*Test*	*Train*	*Test*	*Train*	*Test*
Model 1: Logistic regression	76.11%	75.69%	76.32%	76.07%	75.91%	75.56%
Model 2: Logistic regression with PCA and parameter optimization	76.10%	75.88%	76.32%	76.17%	76.10%	75.95%
Model 3: Random Forests	81.19%	74.06%	81.28%	74.48%	81.04%	74.09%
Model 4: XGBoost	88.61%	74.72%	88.61%	75.09%	88.53%	74.77%
Model 5: Logistic regression with log-transformed data	77.14%	76.50%	77.34%	76.82%	77.13%	76.52%
Model 6: Random Forests without PCA	**96.23%**	**77.84%**	**96.19%**	**78.12%**	**96.18%**	**77.74%**
Model 7: XGBoost without PCA	80.41%	77.57%	80.50%	77.87%	80.23%	77.50%
Model 8: kNN neighbors	77.04%	75.71%	77.24%	76.07%	76.90%	75.71%
Model 9: kNN neighbors without PCA	77.79%	75.66%	77.95%	76.03%	77.59%	75.65%

Note: Starting from Model 5 log-transformed data of positively skewed variables was applied

Final model evaluation and interpretation

Each model was able to achieve an accuracy of 73-78% after parameter optimization. The final chosen model is the tuned **Random Forest model**, which had the highest test set weighted average F1 score of 0.777.

Surprisingly, the model performed worse at predicting failures(64% of the time) than successes (78% of the time), with a true negative rate that was lower than true positive rate. In other words, it categorized a large number of failed projects as successes while classifying a small number of successful projects as failures. Perhaps the factors that cause a project to fail are more likely to be beyond the the data, such as poor marketing, not relevant updates, or not keeping contact with potential investors.

Because PCA was not used, it was possible to plot feature importance (see Appendix). It shows that goal size is the most important feature, followed by the number of reward tiers, duration and google trend searches.

Conclusion and Recommendations

The main goal of this research was to understand what factors lead to the success of crowdfunding campaigns and based on that information predict success of campaigns. For that purpose I analyzed about 20 features of more than 9000 crowdfunding campaigns. By using collected data I was able to show how different features of a campaign contribute to its success and based on the results constructed several machine learning models to predict the success of campaigns. Among those models **Random Forest combined with PCA** proved to be the best one by providing the highest accuracy(average F1 score) of around 78%. However it was better on predicting successes rather than failures. Additionally, goal size, number of reward tiers, duration and google trend searches were the the most important features by looking at feature importance scores.

Both OLS regression and descriptive part of research showed that there are features that have a positive effect on the success of projects and features that rather lead them to failure. Among the factors that had a **positive effect** on success rate and/or the amount of money received are:

- Having smaller project goals
- Being chosen as a featured
- Having shorter duration
- Comics, art and games projects were the most successful
- Suggesting more reward tiers attracts more funding
- Having experience with launching a crowdfunding campaign (number of created projects)

• Providing more updates, FAQs increases the interest in projects which lead to more comments and more funding.

Factors that had a **negative effect** on success rate are:

• Large goals

• Longer duration lead to failure

• Food and technology projects were the least succesful

• Surprisingly, failed projects had more google trend searches.

Besides the key conclusions, there were also some **random interesting findings** that are also very important to know before launching a project.

• **Tuesday** was the most common and successful day to launch a project.

• Weekends were the least popular days to launch a project(they attract less money, have fewer backers, receive smaller pledges per backer, and are slightly less successful).

• While the most popular month to launch a project was January, the most successful one was **February**.

All above mentioned conclusions can be viewed as **recommendations** for future project creators and taken into consideration before starting a new campaign.

References

[1] Adebola Lamidi. *"Predicting the success of Kickstarter campaigns"*. September 20, 2017

[2] Ethan R. Mollick. *"The Dynamics of Crowdfunding: Determinants of Success and Failure"*. SSRN Electronic Journal 29(1). April 2013.

[3] Jermain C. Kaminski et al. *"Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals"*. Springer Science+Business Media, LLC, part of Springer Nature 2019. March 13, 2019.

[4] Kartik Sawhney et al. *"Using Language to Predict Kickstarter Success"*. Stanford University, Department of Computer Science. 2016.

[5] Laura Lewis. *"Using Language to Predict Kickstarter Success"*. April 9, 2019.

[6] L.Castro, T.Couto, N.Felix. *"Success Prediction of Crowdfunding Campaigns: A Two-Phase Modeling"*. International Journal of Web Information Systems. July 2020

[7] Mati Kucz. *"Crowdfunding Success — Chance or Strategy?"*. October 2, 2018

[8] Nam Phan. *"Predicting the success of Kickstarter campaigns"*. October 18, 2018

[9] Nihit Desai et al. *"Plead or Pitch? Predicting the Performance of Kickstarter Projects"*. Department of Computer Science, Stanford University. 2015

[10] Riley Predum. *"Predicting Kickstarter Campaign Success with Gradient Boosted Decision Trees: A Machine Learning Classification Problem"*. February 2019

[11] Shreyas Devalapurkar. *"Solving Kickstarter: Predicting Campaign Success with Support Vector Machines"*. June 5, 2015

[12] Vincent Etter et al. *"Launch Hard or Go Home!Predicting the Success of Kickstarter Campaigns"*. School of Computer and Communication Sciences. 2013

[13] Yan Li et al. *"Project Success Prediction in Crowdfunding Environments"*. February 2016

[14] `kicktraq.com`

[15] `kickstarter.com`

[16] `https://trends.google.com/`

Appendix

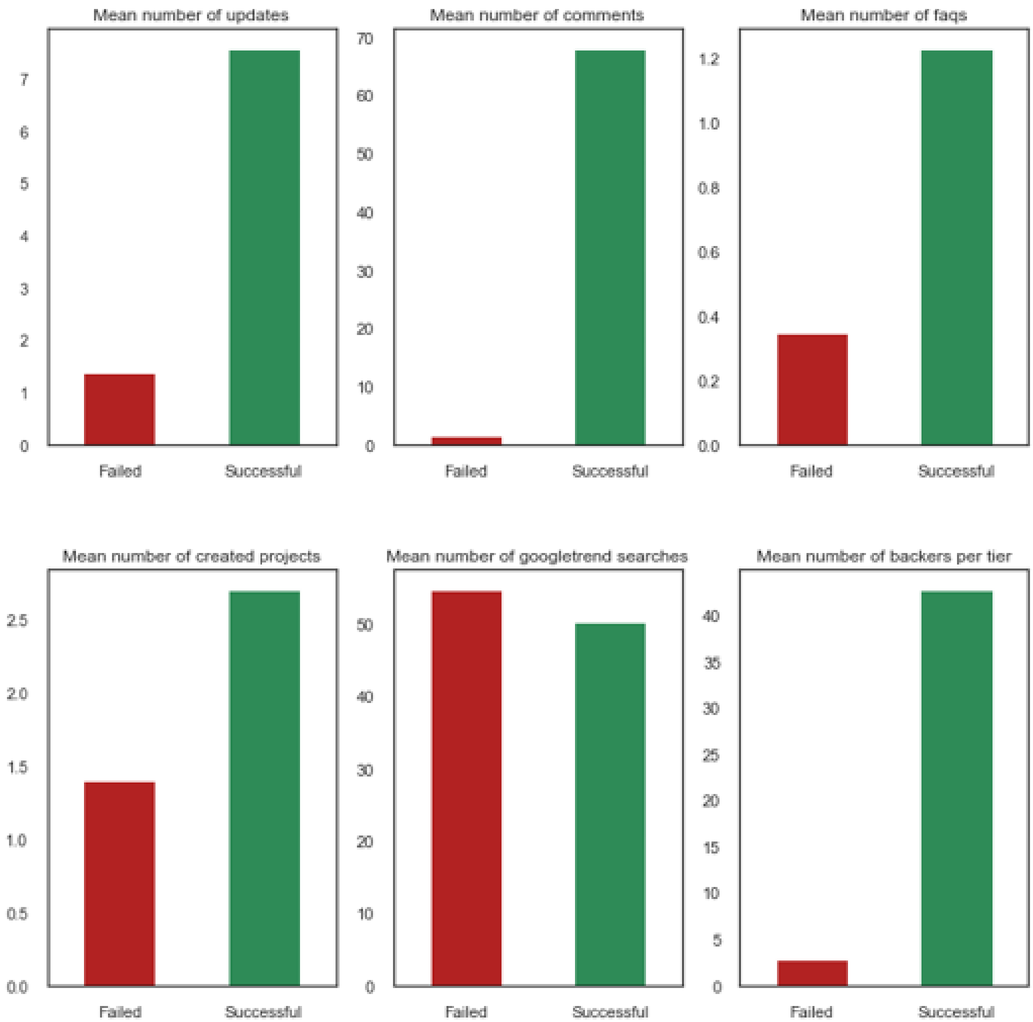

Figure 6: Differences between successful and failed projects (continued)

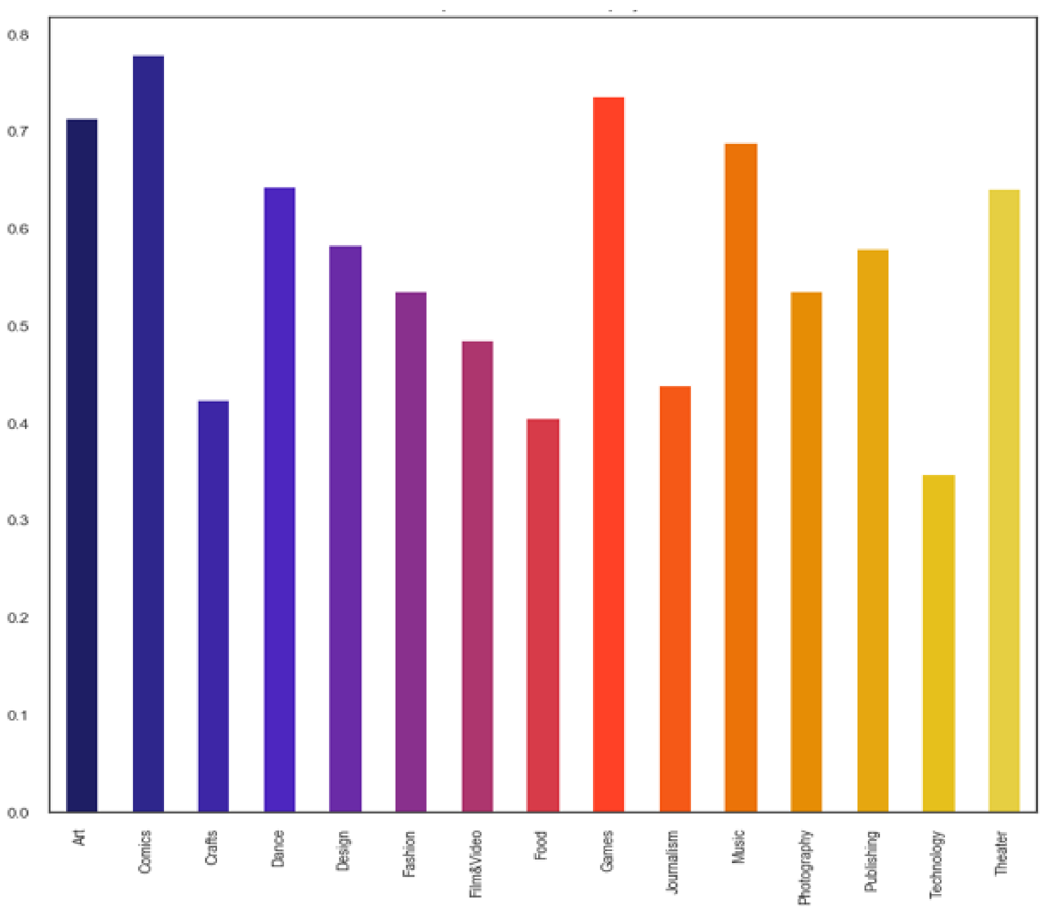

Figure 7: Proportion of successful projects by category

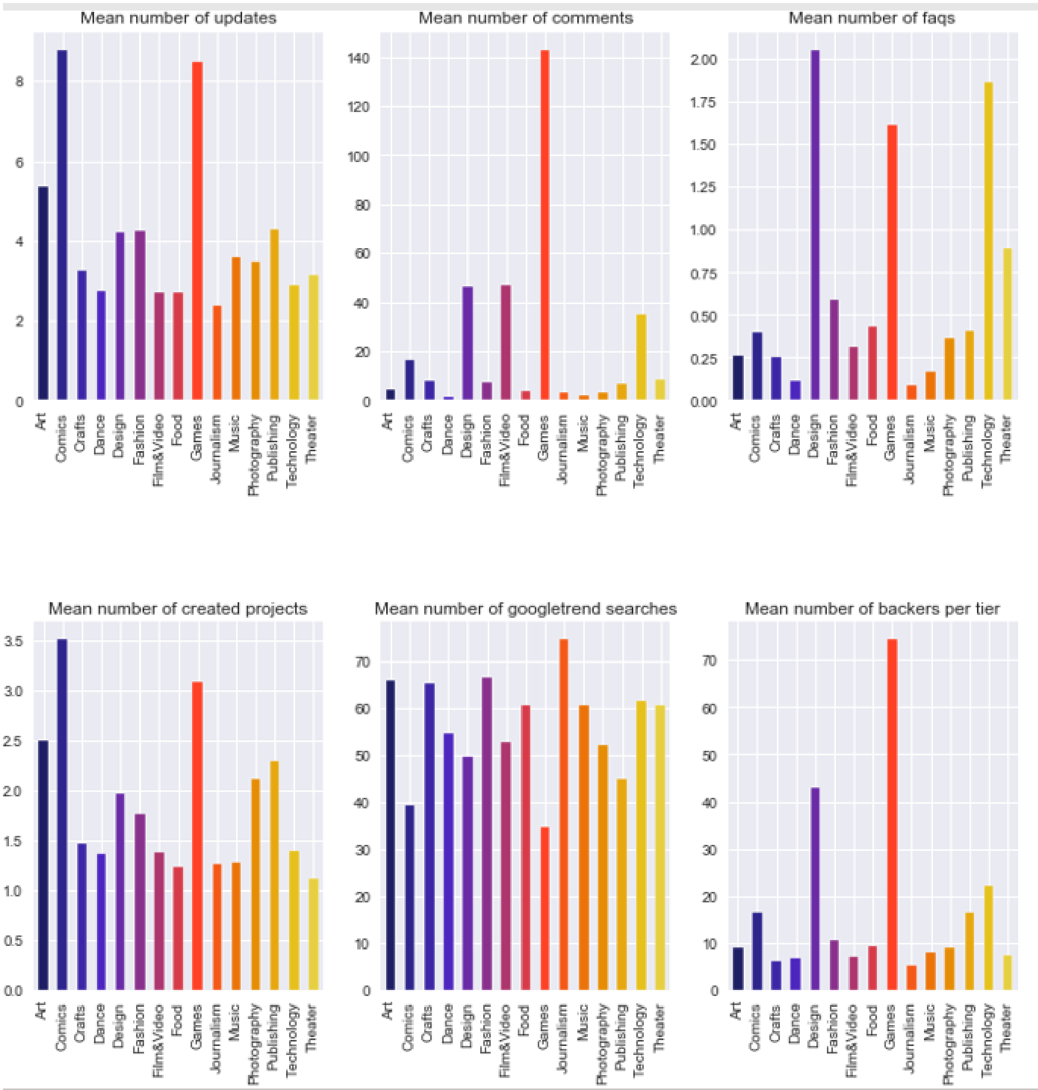

Figure 8: Feature performances by categories of the projects (continued)

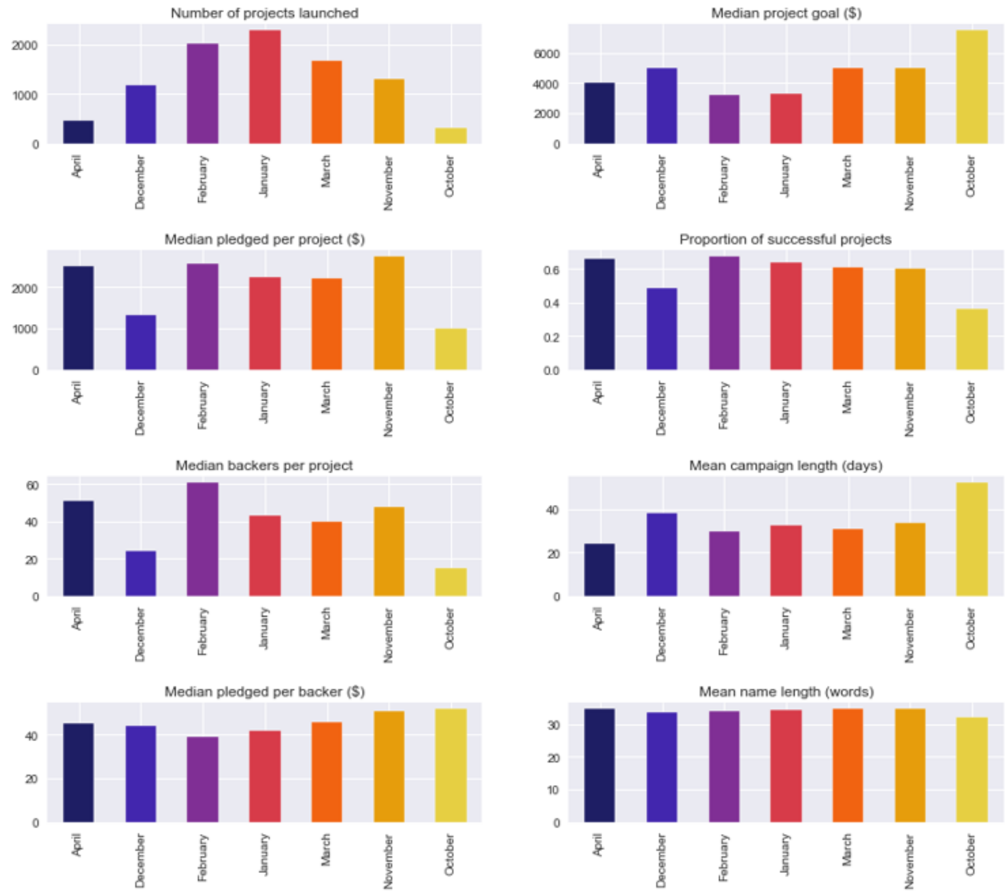

Figure 9: Feature performances by month of launching

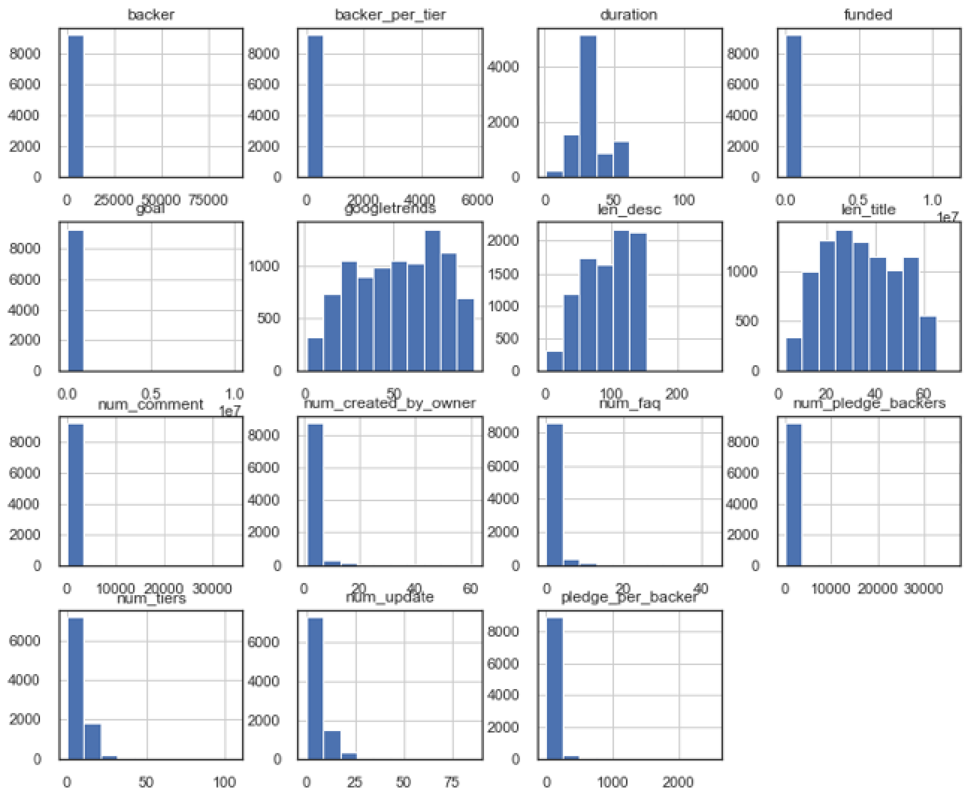

Figure 10: Distribution of the variables

Table 5: OLS output

OLS Regression Results

Dep. Variable:	status_Funding Successful	R-squared:	0.249
Model:	OLS	Adj. R-squared:	0.247
Method:	Least Squares	F-statistic:	169.5
Date:	Thu, 08 Apr 2021	Prob (F-statistic):	0.00
Time:	11:14:26	Log-Likelihood:	-5163.2
No. Observations:	9232	AIC:	1.036e+04
Df Residuals:	9213	BIC:	1.050e+04
Df Model:	18		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8130	0.020	40.337	0.000	0.773	0.852
googletrends	-0.0009	0.000	-4.791	0.000	-0.001	-0.001
duration	-0.0091	0.000	-25.665	0.000	-0.010	-0.008
goal	-1.263e-07	2.32e-08	-5.454	0.000	-1.72e-07	-8.09e-08
len_title	0.0019	0.000	6.555	0.000	0.001	0.003
num_tiers	0.0156	0.001	20.176	0.000	0.014	0.017
num_created_by_owner	0.0047	0.001	4.593	0.000	0.003	0.007
category_Comics	-0.0654	0.018	-3.667	0.000	-0.100	-0.030
category_Crafts	-0.1993	0.029	-6.867	0.000	-0.256	-0.142
category_Design	-0.1070	0.016	-6.677	0.000	-0.138	-0.076
category_Fashion	-0.1194	0.018	-6.529	0.000	-0.155	-0.084
category_Film&Video	-0.1559	0.018	-8.688	0.000	-0.191	-0.121
category_Food	-0.2430	0.021	-11.376	0.000	-0.285	-0.201
category_Journalism	-0.1901	0.057	-3.339	0.001	-0.302	-0.078
category_Photography	-0.1782	0.039	-4.619	0.000	-0.254	-0.103
category_Publishing	-0.1445	0.015	-9.370	0.000	-0.175	-0.114
category_Technology	-0.2455	0.018	-14.020	0.000	-0.280	-0.211
category_Theater	-0.0184	0.059	-0.313	0.754	-0.133	0.097
featured_Featured	0.3076	0.014	22.129	0.000	0.280	0.335

Omnibus:	3653.643	Durbin-Watson:	2.000
Prob(Omnibus):	0.000	Jarque-Bera (JB):	565.662
Skew:	-0.265	Prob(JB):	1.47e-123
Kurtosis:	1.910	Cond. No.	2.58e+06

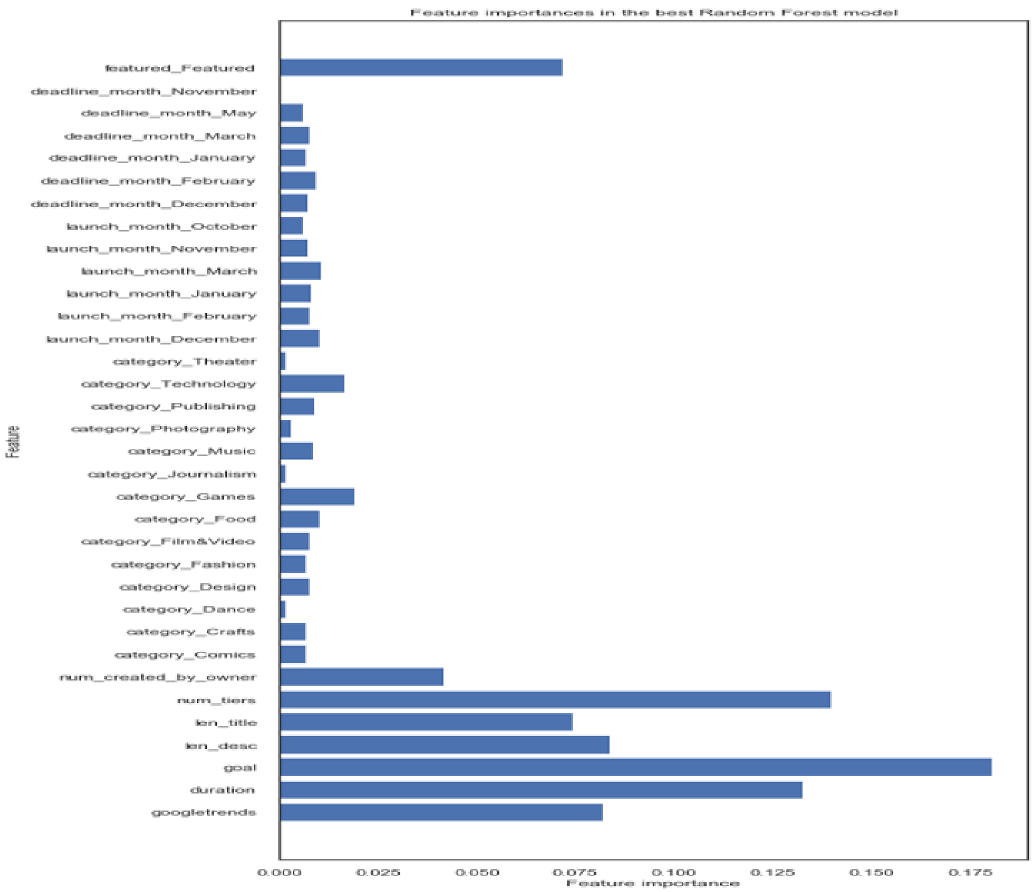

Figure 11: Feature importance