Determinants of success in Kickstarter Campaigns

Word count: 4527

Candidate ID(s): 38134, 51126

Date: 01/05/2025

<u>Permission to use as an example</u>

I give permission for my assignment to be used as an example in ST312 (in an electronic form).

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Introduction to the dataset and presentation of research question

Crowdfunding platforms like Kickstarter have transformed the entrepreneurial landscape by enabling individuals to raise capital directly from the public to fund a variety of creative endeavours. Many are interested in maximising their chances of launching a successful campaign; as such, our goal was to create a statistical model that could act as a useful tool for said individuals.

We conducted our analysis on a sample of over 200,000 campaigns launched between 2009 and 2024, which we obtained from a web scraping website (Vitulskis & Jonaitis, 2025). As there have only been slightly more than 650,000 Kickstarter campaigns ever (Woodward, 2025), we find our dataset to be a representative sample. It includes information on a variety of project characteristics:

- The project's funding goal (in US dollars)
- The total number of backers behind a project
- A project's category (ranging from photography to food)
- The currency a project was launched in
- Whether the project is a 'staff-pick' (a special designation given to projects the Kickstarter staff finds particularly noteworthy or deserving of extra visibility)
- Whether a video is included in the Kickstarter page
- The campaign's duration and desired duration prior to launch
- The date a campaign was launched (broken down into the year, month and day of the week)
- The number of characters in a project's blurb
- The final outcome of a project (whether it met its goal)
- The total amount pledged to a project (converted to dollars)

Our research is guided by the central question: *what project characteristics are responsible for successful Kickstarter campaigns?* We break this down into two sub questions; what are the characteristics responsible for a project meeting its goal, and what characteristics are responsible for a project being *more* successful than others. Our analysis surprisingly found there was a difference in the answer to each sub question; for example, whilst the inclusion of a video was significant for predicting whether a project met its goal, it was not significant when it came to a project raising more money, relative to others.

Overall, we find there are ways to reliably improve the probability a project meets its goal; including a video, setting a reasonable goal, including a higher number of blurb characters, as well as of course, trying to attract many backers, are essential. However, there does not appear to be strong evidence that optimising the date a project is launched improves the likelihood of success. Modelling the relationship between predictors and the *scale* of success is more difficult, though being a staff pick is highly recommended to improve the probability of launching a large-scale campaign.

1.2 Literature Review

There have been previous statistical analyses conducted about Kickstarter, with the usual goal being to predict successful projects. For example, students using SVMs at the California Institute of Technology (Caltech) in 2013 found that the most important predictors of success[1] were whether the project page included a video or not, the funding goal of a project, the number of projects backed by a creator, as well as the total number of projects created by a creator (Chen, Jones, Kim, & Schlamp, 2013). Another finding was that proxies for a project's social media presence, such as the total number of tweets linking to a project and the view count of YouTube videos about it, were not significant predictors. This somewhat justified our exclusion of factors related to a campaign's social media use, as it is likely quantitative proxies may struggle to capture the true link between social media use and probability of success.[2]

In 2014 Mitra et al. conducted an analysis into the specific language that drives successful campaigns (Mitra & Gilbert, 2014). They suggest the use of phrases that 'reflect the principle of reciprocity', such as '*also receive two*' are significant predictors of success. Whilst our analysis does not go in depth into the language used in campaigns, we found the number of characters in a project's blurb to be an important factor in predicting a project meeting its goal, reflecting the importance of language for campaign success.

Oduro, Yu, & Huang, (2022) employed logistic regression and classification tree methodologies to predict the probability a project meets its goal. They similarly emphasised the predictive value of factors such as funding goal and the inclusion of promotional videos on the project page, which they hypothesised increased engagement and trust among potential backers.

---

[1] Where in this case success meant a project meeting its goal

[2] We further discuss more abstract concepts left out of our model in the Conclusion and Limitations section of this paper

Our analysis differs from existing research in a few aspects. For one, many statistical analyses into Kickstarter have a heavy emphasis on pure prediction of campaign success, which differed from our research aim, which was to build an interpretable model. For example, Etter et al. (2013) used a K-Nearest Neighbours (KNN) classifier as a method to predict the success of a project. However, KNN was unsuitable for our analysis, as it does not provide interpretable measures of feature importance and instead relies purely on distance-based classification without modelling the relationship between individual predictors and the outcome.

The second difference is that most papers focus on predicting success with success defined as a project meeting its goal. We explicitly compare the likelihood a project raises more money or attracts more backers to the probability a project meets its goal. These different outcomes may be of interest to those with different goals with their campaigns; those who want to launch large campaigns that raise a lot of money may differ from those who simply want to reach a more reasonable goal.[3]

2. ANALYSIS

2.1 Data Cleaning

We first removed information about ongoing projects from our dataset, as well as merging projects that were 'cancelled' and 'failed' together.[4] After exploring the columns of our dataset, we found some projects were trivially categorised as 'successful' after setting a goal of $1 and meeting it. As we did not want our model to be skewed by such projects, we removed all projects with target amounts under $50.

We also considered removing very influential points from our dataset, based on Cook's distance criterion. However, we chose to not remove any outliers from our model for two primary reasons. The first is that our dataset had almost 200,000 observations after cleaning, so the effects of outliers would be muted in our models (though to prevent extremely high influence on coefficients we still applied logarithmic transforms to certain numerical predictors). The second is that we believed the information within these outliers was important for our model and interpretation- if a project performs extremely well within a certain category, that should rightfully impact the estimated effect of that category in our model.

---

[3] The median goal in our dataset was $5000 after cleaning

[4] Whilst these aren't exactly the same, we imagined prospective creators would want to similarly avoid their project being cancelled as much as failing.

We built two models to predict whether projects met their goal or not. The first was a logistic regression model, whilst the second was a Classification and Regression Tree (CART) model.

2.2. Logistic Regression

2.2.1 Model Building

We began our analysis with a logistic regression model that included all features[5] in our dataset as predictors for the *state* of a project. After this initial model we decided to remove the predictor *subcategory*[6], as some subcategories only contained one state of project, leading our model to have infinite coefficients for certain predictors (causing a 'perfect separation' problem).

We tested a variety of logistic regression models with different combinations of predictors included. Each model was evaluated via the Bayesian Information Criterion (BIC), which penalised excessive parameters and encouraged easier to understand models. After testing multiple combinations of predictors, we found excluding the day a campaign was launched, as well as the month it was launched, reduced BIC, leading us to exclude them from our model.

We also tested a few potential interactions. After evaluating each with the BIC we only found an interaction between the number of backers and category to be significant. However, we did not test every possible interaction for our logistic regression model, as we focused more on interactions between predictors in our CART model and did not want to risk overfitting for our logistic model.

2.2.2 Model Assumptions

After we fit our final logistic regression model, we checked the generalised variance inflation factor (GVIF) score. No predictor had a value that indicated any collinearity issues (all GVIF values were less than 4), increasing our confidence in the reliability of our model.

---

[5] Other than pledged amount, which prevented our model from converging, as well as a campaign's duration- due to collinearity issues with its 'desired duration'

[6] We still included a project's category, such as 'Fashion', but removed its subcategory, which may be 'Pet Fashion'

2.2.3 Model Results

The logistic regression model we created is provided in the appendix. As expected, the number of backers was a highly significant variable in the model. For example, a project in the comics category, launched in the UK in 2025, with a goal of $1000[7], but only 5 backers, is predicted a 14.3% chance of success by our model. However, if that same project had 15 backers the chance of success rises to 71.5%. This was more than we expected, though on the whole, unsurprising, as most Kickstarter creators likely know that with more backers, their chance of success rises.

Category was also significant. For example, the model predicts that a project in the crafts category (with a few other characteristics[8]) has a 35.2% chance of success, whereas a project with the same features other than category, being in the film category, has an 83% chance of success. The interaction term between *backers_count* and *category* also provides some interesting insight. In a way; this interaction reverses the original influence of a projects' category. For example, whilst a project in the crafts category is predicted to be less successful than other projects when having a low number of backers, its higher interaction coefficient with *backers_count* relative to those other categories reverses this trend, effectively making it so that when a project has a high number of backers, its category becomes less relevant to predicting its success.

Other significant predictors were currency, desired duration, goal, inclusion of a video, and blurb characters. A higher goal predicted a lower chance of success, likely swayed by a few huge and unrealistic goals set for some poorly thought-out projects (one creator in Italy set a goal of over 100 million dollars for their film, though only raised 1 dollar). Longer desired durations also predicted a lower chance of success, perhaps suggesting shorter campaigns are more likely to meet their goal. Additionally, including more blurb characters increased the chances of success, though the maximum number of characters in our dataset was only 196[9] and the coefficient for the inclusion of an additional character in our model was very small (0.003) suggesting a limit to the importance (which would later be confirmed in our CART model).

Most surprisingly, a project being designated a staff pick in this model predicted a *lower* chance

---

[7] As well as a video and a desired campaign duration of 2 weeks

[8] Such as having 10 backers, being launched in 2025 and having a video

[9] A limit that cannot be exceeded on the Kickstarter website

of success. We explored why this could be the case by conducting a separate logistic regression without the number of backers as a predictor. In this regression, being a staff pick *increased* the chances of success, suggesting the impact of being a staff pick was directly connected with the number of backers, a fact we confirm in a later section. This led us to believe the importance of being a staff pick is that it encourages more backers; though if a project has many backers and is not a staff pick, being a staff pick does not increase the probability of success (and in fact, may do the opposite).

The final logistic regression model, when evaluated on a confusion matrix, had a 94% prediction accuracy. However, this was heavily reliant on the effect of the number of backers; excluding it as a predictor reduced prediction accuracy to 74%. To further investigate possible interactions at play, we fitted a CART model.

2.3 Classification Trees (CART)

We fitted our tree using the *rpart* package in R, to predict project success, similar to our logistic model.

2.3.1 Model Assumptions

A key strength of the CART model is that it does not rely on many statistical assumptions typically required in parametric models, such as linearity, normality of residuals, or homoscedasticity.[10] This makes it particularly suitable for analysing complex real-world data, where relationships between variables may be nonlinear or interaction based.

Additionally, the CART model has a few advantages over our logistic model. For one, CART models are naturally robust to outliers since splits are determined by threshold-based rules rather than being overly sensitive to extreme values. This provides a good comparison to our logistic model, which may have been swayed by the projects with extreme characteristics in our dataset. Furthermore, unlike black-box machine learning models, CART produces a transparent, rule-based structure that is easy to visualise and interpret - aligning well with our primary aim of analysing project characteristics rather than purely predicting outcomes.

[10] Though our logistic model also did not make these assumptions, this aspect of CART models was very useful in our model built for predicting pledged amount

2.3.2 Model Building

CART models are highly dependent on the choice of their optimal complexity parameter (CP), a regularisation term that plays a critical role in cost-complexity pruning (CCP)-a pruning technique proposed by Breiman, L., Friedman, J. et. al in 1984 for building regression trees. CCP balances model fit and complexity by penalising trees for each additional split that does not significantly reduce prediction error. In this framework, the CP controls the trade-off between interpretability and predictive accuracy; where a lower CP value leads to a more complex tree, which can be a difficult to interpret model, and increases the risk of overfitting. (Of course, a too high CP value can increase the risk of *underfitting*).

We initially considered selecting the CP based on the minimum cross-validation error (min xerror), which aims to maximise the predictive performance. However, this resulted in an extremely small CP value (approximately 0.0001), yielding an overly complicated tree, that conflicted with our project aim of easy interpretability. We also explored the 1-Standard Error (SE) rule (Breiman et al. 1984) which selects the simplest model within one standard error of the minimum cross-validation error. Yet, due to our large dataset (~200,000 observations), cross-validation was highly stable, and the standard error was very small—causing the 1-SE rule to select the same complex tree as the min xerror method.

Given these limitations, we manually selected a CP value that offered a better balance between interpretability and predictive accuracy. As Breiman et al. (1984) emphasise, the choice of CP should be guided by practical considerations, with ours being to build an understandable model. We therefore used an initial CP of 0.0001 to allow full tree growth, printed the cross-validation table, and examined a range of CP values to find an appropriate trade-off. A CP of 0.01 was found to yield a manageable tree structure with a cross-validation error around 0.14[11]. This allowed us to retain key decision splits while discarding branches with minimal impact, producing a final model that is both interpretable and robust. Additionally, consistent with our logistic model, we removed subcategory as a predictor, as it complicated model interpretation without substantially improving predictive accuracy.

2.3.3 Model Results

With our final classification model, we achieved an accuracy of 94%, which is reasonably consistent with the cross-validation estimate and suggests good generalisation performance. As seen in the confusion matrix in Table 1, the test accuracy we have obtained is quite high, which

[11] Table 1A in the Appendix shows the different CP values tested

increases our confidence in the later variable importance table.

Table 1 Confusion Matrix CART Model

	Reference	
Prediction	**No**	**Yes**
No	14453	395
Yes	1792	22016

Our CART model confirmed a number of findings from our logistic model. For one, predictors such as the number of backers, category, inclusion of a video, as well as funding goal were once again highly significant. Additionally, the model confirms how the day or month a project was launched in were much less significant than other predictors.

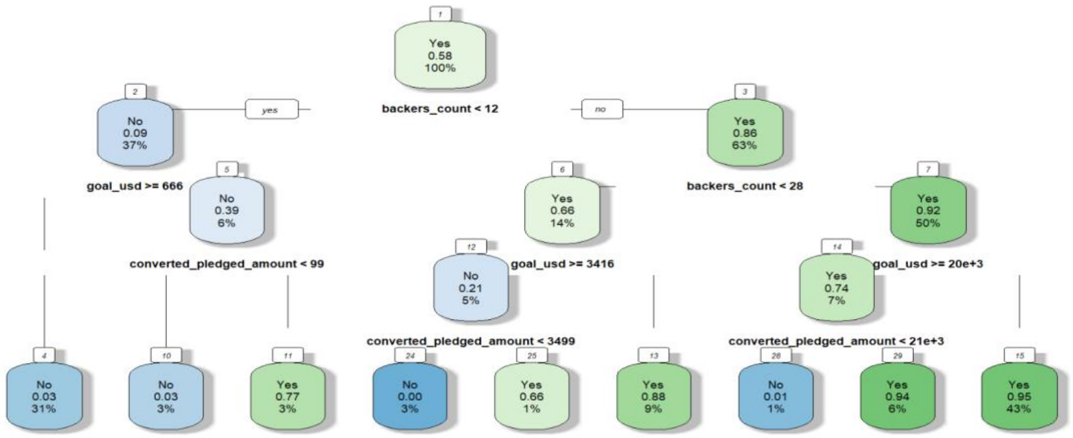
The CART model's inclusion of a variable importance table (Table 2) provided additional insights. For one, the currency or location a project is launched in is much less important for project success compared to factors such as inclusion of a video or funding goal. Additionally, both duration variables (campaign duration and desired project duration) emerged as important predictors again. Lastly, staff pick does not appear a significant predictor in our CART model, in a way confirming our logistic model's finding that when the number of backers is included in a model, being a staff pick does not affect the probability of success.

Table 2 Variable Importance from our CART model

Variable	Variable Importance
`backers_count`	48 898.57
`converted_pledged_amount`	44 540.30
`goal_usd`	8735.44
`video`	5000.71
`desired_duration`	2949.18
`campaign_duration`	1951.27
`category`	616.92
`year_launched`	412.44
`country_displayable_name`	24.76
`currency`	17.48
`blurb_characters`	0.22

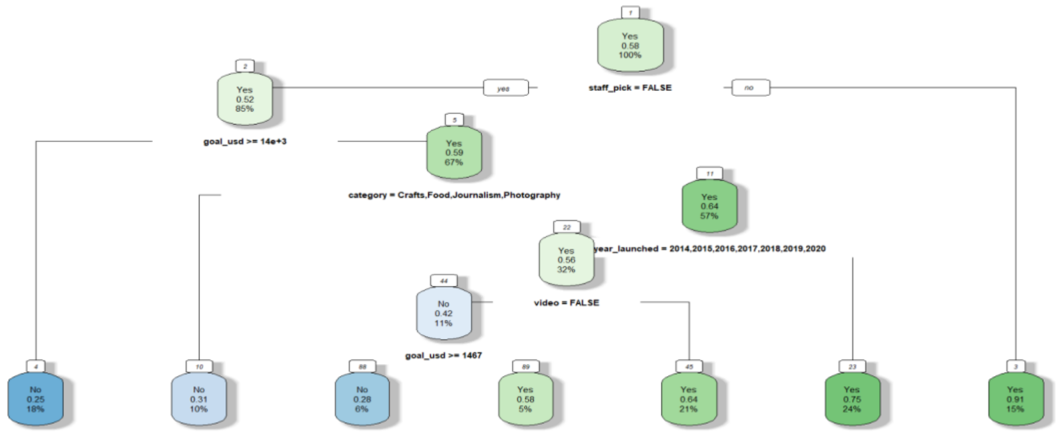
The CART model also provided us with classification trees from which we can infer potential interactions at play. Figure 1 neatly highlights some of these; for example, campaigns with fewer than 12 backers have a higher chance of failure when they have a goal of more than $666, with only 3% of such projects predicted to succeed. Even for those with 12-27 backers, if they have a goal of more than $3,416 they only have a 21% chance of success- though if the goal is less than this, the model predicts an 88% chance of success. This demonstrates the (almost obvious) relationship between *goal* and *backers_count*; loftier goals require more backers to meet.

Figure 1 Pruned Tree on all predictors



Additionally, Figure 2 was created by excluding backers count and pledged amount from our model, in order to gain greater insight into less significant predictors. This tree should be analysed with more caution, as the accuracy of the model fell to 72% after ignoring the two main predictors. It demonstrates that being a staff pick *does* increase the probability of success to 91%; once again suggesting the benefit of being a staff pick is linked to gaining more backers or raising more money, something we confirm in a later section. It also demonstrates how categories such as Crafts and Photography predict a lower chance of success, though the structure of the tree suggests this mostly applies to projects with goals of less than $14,000.

Figure 2 Pruned Tree on all predictors except backers count and pledged amount



2.3.4 Summary of classification models

To conclude our section on classification models; both our CART and logistic models agree on the key factors for Kickstarter success. For prospective creators, including a video is highly recommended; perhaps signalling a minimum level of commitment. Whilst the category a project is launched in *does* affect its probability of success, our logistic model suggests that this is mainly for smaller projects; if a project has enough backers our logistic model suggests the effect of category wears off, and for our tree, the effect of category is important only for lower goals.

Both models also suggest that creators need not worry too much about the day or month of their project launch, as these factors are much less influential than other predictors. Finally, although year of launch is a weaker predictor, our models suggest a modest trend towards increasing project success rates over time.

2.4 Characteristics responsible for *great* success

We also investigated whether there was a different relationship between the characteristics responsible for a project meeting its goal, and for a project to be 'more successful' relative to other projects, that is, raising more money, or attracting many more backers than other projects.

We first attempted to model this relationship with a linear regression. However, the residuals (specifically the residuals vs fitted plot) consistently showed evidence of non-linearity, even after applying logarithmic transforms and including polynomial terms. As a result, we decided to model the relationship between project characteristics and pledged amount with a regression tree with the same complexity parameter used in our CART model. This is due to the aforementioned lack of assumptions in trees, relative to parametric models.

2.4.1 Pledged Amount as Outcome

The model built for pledged amount provided some additional insight into the factors that drive a campaign's financial success. We excluded backers count as a predictor, due to its obvious influence on pledged amount.[12] Once again, a campaign's goal was a very significant predictor for the amount raised; in our linear model this was a negative relationship, though our tree heavily suggested the relationship was non-monotonic, making specific recommendations difficult. In contrast to our original classification model; factors such as month launched, and staff pick were identified as key predictors. Causation is difficult to understand in the effect of the month a project is launched. Our linear model had not found month launched to be a significant predictor, making it likely that the relationship between pledged amount and month is non-linear and possibly interaction based (for example, with category). However, our regression tree was not able to pick up such an interaction, making it hard to give specific advice for the best month to launch a project.

Table 3 Variable Importance with Converted Pledged Amount as Outcome

Variable	Variable Importance
goal_usd	13,485,910,079
month_launched	5,927,553,896
staff_pick	2,777,340,290
year_launched	2,469,814,123
category	2,368,251,155
campaign_duration	1,481,888,474
desired_duration	1,481,888,474
blurb_characters	1,087,925,649
country_displayable_name	17,568,302
currency	16,731,716

[12] Which we confirmed with an earlier model.

2.4.2 Number of Backers as Outcome

For the model with *backers_count* as the outcome[13], goal once again emerged a significant predictor, though again, this was a non-linear relationship. The year of launch was also significant, which may reflect the increasing ease of gaining followers on Kickstarter as the platform matures and becomes more widely recognized. The month of launch also again emerged significant; however, the underlying mechanism driving this relationship remains unclear.

Table 4 Variable Importance with Number of Backers as Outcome

Variable	Variable Importance
goal_usd	2.490×10^{14}
category	1.117×10^{14}
year_launched	5.365×10^{13}
month_launched	4.570×10^{13}
country_displayable_name	4.494×10^{13}
staff_pick	4.179×10^{13}
currency	3.958×10^{13}
campaign_duration	3.327×10^{13}
desired_duration	3.327×10^{13}
blurb_characters	1.421×10^{13}
day_launched	3.937×10^{12}

2.4.3 Comparison

Interestingly, though inclusion of a video was highly influential in predicting project success, it did not emerge as a key predictor in the regression trees for predicting pledged amount or number of backers. This could suggest the inclusion of a video may help projects appear credible enough to meet their funding goal but does not substantially influence the *scale* of their success. In a way, including a video could be necessary for meeting the funding goal, but not sufficient for widespread success.

It is also interesting to note that the substantial difference in Root Mean Square Error (RMSE) between our two regression tree models reflects the differing levels of variability and predictability inherent in the two outcome variables. The RMSE for *backers_count* was considerably lower (~741) than for *converted_pledged_amount* (~101,479 USD), suggesting that the number of backers is a more stable and predictable outcome based on project characteristics. In contrast, predicting the exact amount pledged is inherently more difficult, likely due to the highly skewed distribution of funding amounts on Kickstarter, where a small number of projects raise exceptionally large sums, creating greater variability and making accurate prediction more

[13] For this model we excluded pledged amount as a predictor, due to its obvious influence again.

challenging.

3 CONCLUSIONS

3.1 Summary of findings

Our analysis found that the number of backers and the pledged amount were by far the most important predictors of project success across both logistic regression and CART models (as expected). For predictors set prior to campaign launch, the presence of a promotional video and a reasonable funding goal emerged as the most significant factors, with loftier goals making success less likely. Other ways to improve success include setting shorter campaign durations and using higher numbers of blurb characters, both suggested by our logistic model (though the coefficients of both predictors were small). Additionally, whilst the category a project is launched in *is* significant to predicting campaign success, the impact of category varies depending on the goal set and the number of backers attracted to a project. Higher goals are less dependent on the category of the project and if a project attracts enough backers, its category becomes less relevant to predicting its success.

Modelling factors such as pledged amount and the number of backers directly, we found that although video presence was influential in determining whether a project met its goal, it had little impact on the magnitude of success, suggesting its effect is primarily threshold-based. Creators are definitely advised to include a video on their page but should not expect it alone to make them stand out. Furthermore, we found that being a staff pick is highly correlated with attracting more backers or raising more money, though if a project already has a high number of backers or has raised a lot of money, it is unlikely to improve chances of success (and our logistic model suggests it might even do the opposite). Finally, the day of launch failed to be a very significant predictor in any of our models, and whilst the month launched *was* significant for the scale of success, we were unable to understand the mechanism of its relationship with success (it was unlikely to be linear).

Together, these findings provide some evidence-based insights for campaign creators aiming for Kickstarter success.

3.2 Limitations & Suggestions for future research

A key limitation of our analysis is the unavailability of certain important variables. In particular, we did not have access to creator-specific data, such as the number of projects created by each creator, which prior research (Chen et al., 2013) identified as important predictors of success. Incorporating such variables could further refine predictive models of Kickstarter outcomes.

It should also be noted that some important factors to be considered in campaign strategy are especially difficult to model with a statistical analysis. For example, the Kickstarter team themselves emphasise the importance of a 'visually appealing campaign page' and 'an effective marketing and promotion strategy' (Smithwrick, 2025). Chen et al. showed that proxies for factors like social media presence[14] may fail to appear significant in a statistical model (Chen, Jones, Kim, & Schlamp, 2013), though this does not mean marketing and social media use is not important for project success. Instead, a more abstract concept like 'social media quality' may be the underlying factor behind successful campaigns. Future research could therefore benefit from incorporating qualitative methods to better understand these intangible drivers of success.

Additionally, our analysis was not explicitly focused on causation, making our recommendations to creators less robust. For example, it seems unlikely that just including a poorly made video would significantly increase the chances of success- perhaps it is the case that projects that have a lot of effort put into them also commonly create promotional videos and it is not necessarily the videos themselves that drive success.

Finally, due to data limitations, we did not examine project dynamics over time. With access to detailed project lifecycle data (e.g., daily funding progress or update frequency), time-series or survival analysis methods could be implemented to model not only final success but also the evolving probability of success throughout a campaign's duration.

[14] Like YouTube view count and number of tweets about a project

4 BIBLIOGRAPHY AND APPENDIX

Works Cited

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (2017). *Classification and Regression Trees.* New York: Chapman and Hall.

Chen, K., Jones, B., Kim, I., & Schlamp, B. (2013). *Kickpredict: Predicting Kickstarter success.*

Etter, V., Grossglauser, M., & Thiran, P. (2013). Launch hard or go home!: predicting the success of kickstarter campaigns. *Association for Computing Machinery*, 177-182.

Mitra, T., & Gilbert, E. (2014). The language that gets people to give: phrases that predict success on kickstarter. *Association for Computing Machinery*, 49-61.

Oduro, M. S., Yu, H., & Huang, H. (2022). Predicting the Entrepreneurial Success of Crowdfunding Campaigns Using Model-Based Machine Learning Methods. *International Journal of Crowd Science*, 7-16.

Smithwrick, B. (2023, October 25). *10 Ways To Level Up Your Campaign Page*. Retrieved from https://updates.kickstarter.com/level-up-your-campaign-in-2024/?ref=learn

Smithwrick, B. (2025, April 13). *The Secrets to Creating Award-Winning Kickstarter Campaigns*. Retrieved from https://updates.kickstarter.com/the-secrets-to-creating-award-winning-kickstarter-campaigns/

Vitulskis, T., & Jonaitis, P. (2025, April 11). *Kickstarter Datasets*. Retrieved from Web Robots: https://webrobots.io/kickstarter-datasets/

Woodward, M. (2025, April 11). *Search Logistics*. Retrieved from KICKSTARTER STATS & FACTS 2025: EVERYTHING YOU NEED TO KNOW : https://www.searchlogistics.com/learn/statistics/kickstarter-stats-facts/

Appendix

The code for our project is in 3 separate RMD files on the submission page. These include ST312_Cleaning.RMD, which was the first code we wrote, followed by ST312_models_logistic_and_linear, where the code for the logistic models and linear model we fitted are kept. Finally, 312_script_CART_final.RMD includes all the code for the regression trees we fitted.

The code is also on this GitHub page https://github.com/codingfishguy/ST312/tree/main

Table 1A Logistic Regression Model:

Term	Estimate	p.value
(Intercept)	3.5118	0.0000
log(backers_count + 1)	3.3744	0.0000
categoryComics	0.7813	0.0008
categoryCrafts	-1.1445	0.0000
categoryDance	0.2343	0.5455
categoryDesign	0.3808	0.0624
categoryFashion	0.6828	0.0001
categoryFilm & Video	1.8595	0.0000
categoryFood	-0.0771	0.6756
categoryGames	0.8411	0.0000
categoryJournalism	-0.0843	0.7646
categoryMusic	0.6534	0.0000
categoryPhotography	0.0035	0.9859
categoryPublishing	0.3557	0.0254
categoryTechnology	1.8006	0.0000
categoryTheater	0.6235	0.0059
currencyCAD	0.0881	0.3039
currencyCHF	0.8421	0.0000
currencyDKK	0.0171	0.9219
currencyEUR	0.0460	0.5594

currencyGBP	0.1099	0.1528
currencyHKD	0.9436	0.0000
currencyJPY	0.6607	0.0000
currencyMXN	-0.3526	0.0006
currencyNOK	0.3890	0.1105
currencyNZD	-0.0133	0.9388
currencyPLN	-0.1156	0.8089
currencySEK	-0.5030	0.0003
currencySGD	0.3841	0.0197
currencyUSD	0.3083	0.0000
staff_pickTRUE	-0.1188	0.0011
log(goal_usd)	-1.8116	0.0000
videoTRUE	0.3470	0.0000
year_centred	0.0492	0.0000
desired_duration	-0.0003	0.0000
blurb_characters	0.0034	0.0000
log(backers_count + 1):categoryComics	-0.6109	0.0000
log(backers_count + 1):categoryCrafts	0.0018	0.9802
log(backers_count + 1):categoryDance	0.0487	0.6926
log(backers_count + 1):categoryDesign	-0.4630	0.0000

log(backers_count + 1):categoryFashion	-0.2404	0.0000
log(backers_count + 1):categoryFilm & Video	-0.3331	0.0000
log(backers_count + 1):categoryFood	-0.1795	0.0008
log(backers_count + 1):categoryGames	-0.4829	0.0000
log(backers_count + 1):categoryJournalism	-0.2024	0.0204
log(backers_count + 1):categoryMusic	-0.0348	0.4723
log(backers_count + 1):categoryPhotography	-0.1875	0.0021
log(backers_count + 1):categoryPublishing	-0.1642	0.0013
log(backers_count + 1):categoryTechnology	-0.6117	0.0000
log(backers_count + 1):categoryTheater	-0.0649	0.3771

Table 2A CP values and its Cross-Validation Error

Index	CP	nsplit	rel error	xerror	xstd
1	0.719 81	0	1.0000	1.0000	0.002 99
2	0.031 96	1	0.2802	0.2803	0.001 95
3	0.018 14	3	0.2163	0.2185	0.001 75
4	0.017 26	5	0.1800	0.1762	0.001 58
5	0.010 91	7	0.1455	0.1429	0.001 44
6	0.007 94	8	0.1346	0.1301	0.001 38
7	0.007 32	9	0.1266	0.1119	0.001 28
8	0.006 53	11	0.1120	0.1034	0.001 23
9	0.005 72	13	0.0989	0.0916	0.001 16
10	0.005 32	16	0.0818	0.0801	0.001 09
11	0.004 70	18	0.0711	0.0713	0.001 03
12	0.002 87	20	0.0617	0.0646	0.000 98
13	0.002 82	24	0.0502	0.0549	0.000 91
14	0.001 69	26	0.0446	0.0470	0.000 84
15	0.001 38	28	0.0412	0.0405	0.000 78
16	0.001 08	30	0.0384	0.0370	0.000 75
17	0.001 06	32	0.0363	0.0342	0.000 72
18	0.000 87	34	0.0341	0.0321	0.000 70
19	0.000 73	36	0.0324	0.0301	0.000 68
20	0.000 70	38	0.0309	0.0299	0.000 67
21	0.000 63	40	0.0295	0.0293	0.000 67
22	0.000 62	44	0.0270	0.0285	0.000 66
23	0.000 59	46	0.0258	0.0278	0.000 65