

# KICKSTARTER

*Supporting Kickstarter management to help creators  
succeed on the platform to drive platform growth*

*24<sup>th</sup> April 2022*



# Executive Summary

<b>Introduction</b>	<ul style="list-style-type: none"><li>• Introduction to Kickstarter's business and slowing growth trend in recent years</li><li>• Kickstarter can do more to help creators increase project success rates to improve our platform uptake and growth</li></ul>
<b>Objective</b>	<ul style="list-style-type: none"><li>• Identifying factors that affect the success of projects to guide creators in marketing and producing better projects that will increase repeat backers and position Kickstarter as a go-to platform for creators with higher chances of success</li></ul>
<b>Dataset</b>	<ul style="list-style-type: none"><li>• The dataset consists of projects from 2021 to 2022</li><li>• Contains the details of each project such as the goal, amount raised, deadline and the state of project</li></ul>
<b>EDA</b>	<ul style="list-style-type: none"><li>• Understanding success rates by category and sub-categories</li><li>• Analysing goal settings and pledge amounts over time as an indicator of project success</li></ul>
<b>Model Development</b>	<ul style="list-style-type: none"><li>• Exploratory data analysis coupled with objective assessment to identify potential relationships such as multicollinearity between variables</li><li>• Oversampling via model weights to account for data imbalance of significantly higher number of successful projects in our dataset</li><li>• Ran random forest, Logistic regression and XGBoost</li></ul>
<b>Model Evaluation</b>	<ul style="list-style-type: none"><li>• XGBoost chosen as the preferred model with <b>ROC of 0.919, accuracy of 0.842</b></li><li>• XGBoost yields <b>low false positive rate of 4.9%</b></li><li>• Goal, Deadline Day (Sat/Mon), Main/sub-category, Launch Month, Launch Day (Sat), Launch Duration, Blurb Sentiments, Project Duration found to be key features</li></ul>
<b>Limitation</b>	<ul style="list-style-type: none"><li>• Previously successful projects are not accounted for as a handful of Kickstarter projects are a continuation of previous projects and many projects are dependent on the success of previous projects due to the pre-existing large pool of trusted backers</li><li>• Project features relating to level of engagement unavailable (e.g. number of updates or responses to comments) reflecting level of engagement of creators with backers which could be another key factor for determining a project's success</li></ul>
<b>Recommendation</b>	<ul style="list-style-type: none"><li>• <b>KickOff</b> – Interactive <b>Tableau dashboard</b> for creators to understand the <b>success rate and funding target of similar projects / projects in the same category and/or subcategory across different geographies</b></li><li>• <b>KickFar</b> – <b>Campaign success prediction dashboard</b> on Heroku allows creators to <b>predict project success</b></li><li>• <b>KickHigh</b> – Helps <b>struggling creators succeed</b> by partnering with equity crowdfunding platforms</li></ul>

# Kickstarter's AON model passes on risks to creators but more can be done to incentivise creators to succeed on the platform

## Kickstarter is seeing slowing growth in recent years...

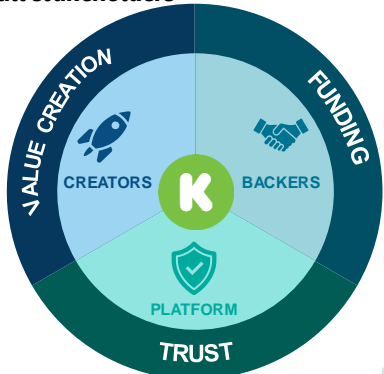
### 1 Business Overview

**About Kickstarter**

- Crowdfunding platform to back projects launched by creators
- **All-or-nothing (AON) funding model**
- Heavy focus on mitigating funding risks for backers

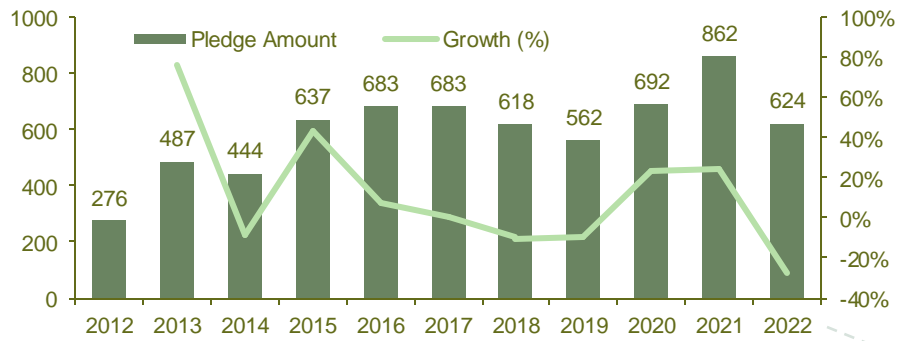
**Associated Concerns**  
 However, while AON shifts risks to the creators, it can **impact platform growth**

*Ecosystem model requires buy-in from all stakeholders*



### 2 Slowing platform growth

Annual amount pledged since 2012 (USDm)



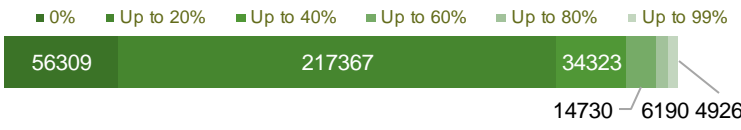
Source: Kickstarter, Statista, Lit Analysis

## ... but more can be done to help creators

### 3 There is huge potential to improve success rate on platform...



Of those that failed, by % funded



**10% - 47% uplift in project success rate**

**Additional opportunity**      **Quick wins**

**Help creators succeed at current risk levels**  
 Kickstarter can retain the AON model but do more *to push near-success projects over the finishing line.*

### 4 ...using a data-driven approach

**Our data-driven approach will deliver the most benefits to creators**

<p><b>Understand Kickstarter's value-add</b></p> <p><i>Identify platform fit for aspiring creators</i></p>	<p><b>Incentivise best practices</b></p> <p><i>Ride on what makes past projects successful</i></p>	<p><b>Promote Kickstarter's growth</b></p> <p><i>Promote security by providing benchmarks at each milestone</i></p>
<b>EDA Dashboard</b>	<b>ML Model + Recommendation</b>	<b>EDA + Funding Guidance</b>

# Understanding our datasets

## Main Projects Dataset

**Information:** 34269 rows x 39 variables  
**Retrieved from:** Web Robots - <https://webrobots.io/kickstarter-datasets/>  
**Retrieved on:** April 16, 2022

Key Variable	Description
Backer Count	Total number of backers that have backed the project
Blurb	Project Description
Category	Main Category and Sub-Category
Converted Pledge Amount	Total pledged amount converted to the default currency of the project from multiple currencies (if any)
Country (Displayable Name)	Country Code and Country Name where the project is in
Created At	UNIX timestamp of the Project Creation
Creator	Details of the Project Creator
Currency (Symbol)	Default Backing Currency of Project
Deadline	UNIX timestamp of the Project Deadline
FX_Rate	Standardised USD:Currency FX Rate
Launched At	UNIX timestamp of the Project Launch
Location	Location details (Lat Long, Country etc.)
(USD) Pledged	Pledge amount in chosen currency and USD
Spotlight	Kickstarter showcases successful projects
Staff Pick	Kickstarter staff choose which projects they like

## Time Series Dataset

**Information:** 600000 rows x 10 variables  
**Retrieved from:** <http://sidekick.epfl.ch/data><sup>1</sup>  
**Retrieved on:** April 22, 2022  
**Additional Pre-Processing done to extract from NumPy array files**

Key Variable	Description
Id	Project id, used as unique identifier
time_interval	Time of the sample (between 0 and 1, relative to the start and end of the campaign)
Pledge_amt	Current amount of pledged money (relative to the campaign's goal, so multiply with the goal to obtain the real amount)
No_backers	Current number of backers (usually not an integer, because the statuses have been resampled to have an equal number of statuses)
Goal	Project goal state: Project final state (1=successfully funded, 0=failed)
Launch_date	Launch date (as a UNIX timestamp)
Deadline	Deadline (as a UNIX timestamp)
Duration	Duration of project in days
actualpledge	Actual pledged amount, already multiplied by campaign goal

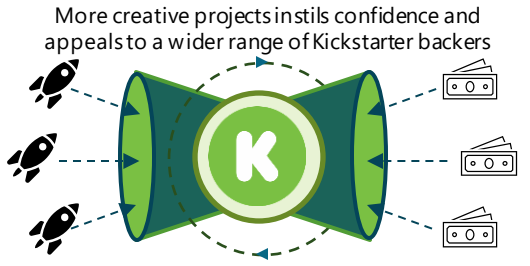
Source: 1. (Etter, Grossglauer & Thiran, 2013)

# Objective: To help creators succeed on Kickstarter and drive the flywheel of growth

## Main Objective: Help Creators Succeed on Kickstarter

**K Kickstarter Fees**

**Fixed:** 5% of total funds raised  
**Payment Processing:** 3 – 5% + \$0.30 per pledge (varies by country)  
**Total:** ~8 – 10% of funds raised

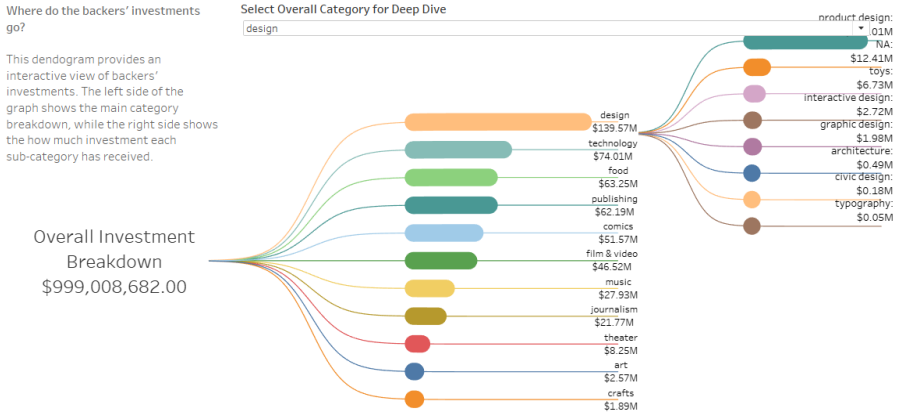


**Management's Priority**  
 Kickstarter is incentivised to help creators succeed on the platform as it drives revenue for the company

More repeat backers position Kickstarter as a go-to platform for creators with higher chances of success

**Helping creators first will drive the flywheel of growth for Kickstarter**

## Identify Platform Fit: Amount Pledged by Category



## Data proves that success drives future confidence



**Success incentivise project listings**

Over time, projects in the food, publishing, comics and design categories **have become the most common, with the highest success rates**

**Why?** Investment flows tells us which **categories see a strong community of backers** and thus, creators can be comforted by a larger pool of organic viewership / pledges on their project

**Which?** **Greatest investment flow:** Design, Technology, Food  
**Least investment flow:** Theater, Arts and Craft

**Insights** Technology is one of the categories with lowest success rates but the second highest investment flow, which suggests **backers are idea-focused and have a high-risk appetite.**

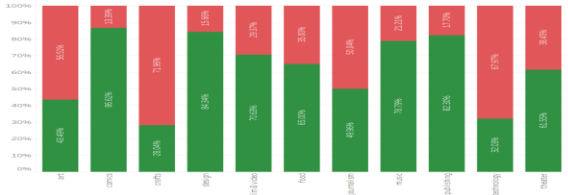
Management can partner with angel investor communities / venture capital funds to bolster seed investments in weaker categories

Link to tableau analysis: <https://public.tableau.com/app/profile/adelyn.koh/viz/ConsultingforAnalytics-Kickstarter2/ProjectOwnerDashboard?publish=yes>

# EDA – Understanding project success rates and making them visible to creators

Success rates can be influenced by multiple factors, but funding target is one major factor

## 1 By Category



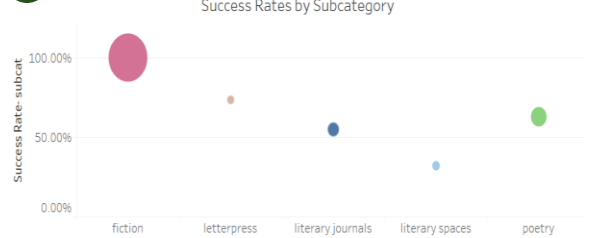
**Highest success** Comics, Design Publishing

**Lowest success** Craft, Technology (Only 1 in 3 succeed)

## 2 By Sub-Category

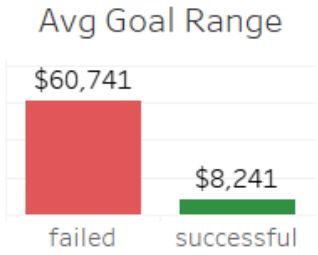


**Insights** Technology projects have the *highest targets*, which makes it difficult for them to reach goals



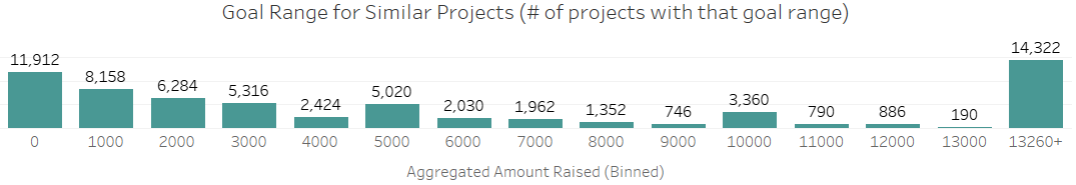
**Insights** Success rates vary greatly within sub-categories, which suggests that there are *multiple factors involved – funding targets, backer demand and competition*

## The importance of goal setting



**Goal Rationalisation**

On average, successful projects had much lower (~7x) goals than failed projects

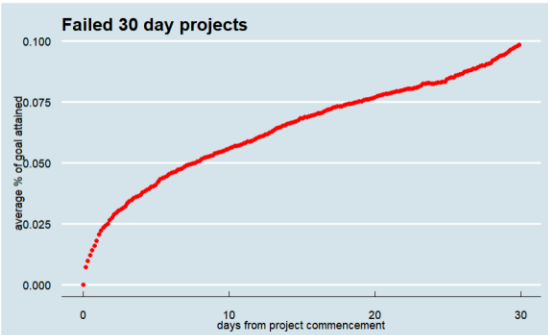
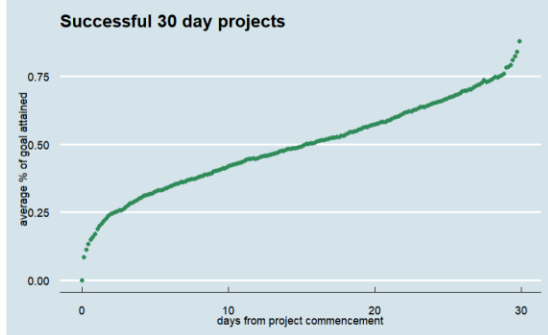


**Cultivating the perception of success** Majority of successful projects (60.4%) set goals that are < 5000 USD. **This suggests that herding and certainty economic factors influences backers to fund projects that are more likely to meet their goals.**

Source: Kickstarter, Lit Analysis, Team's tableau dashboard

# EDA – Understanding project success rates and making them visible to creators

## Pledged amounts over funding duration



**Initial 5 day period is crucial to success**

**Context**  
▼

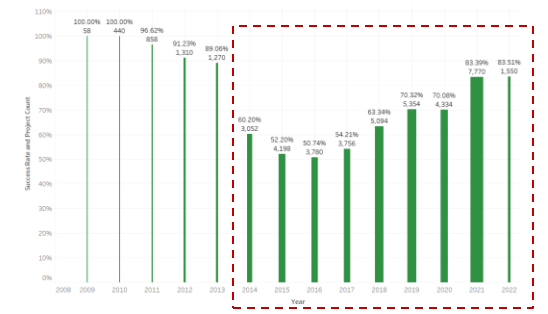
Successful projects raises **10x more** pledges as a percentage of their overall goal. **First 5 days indicates fundraising traction.**

**Action**

Management can **upsell spotlight** to creators

## Our goal is to make it easily visible to them what targets made past projects successful

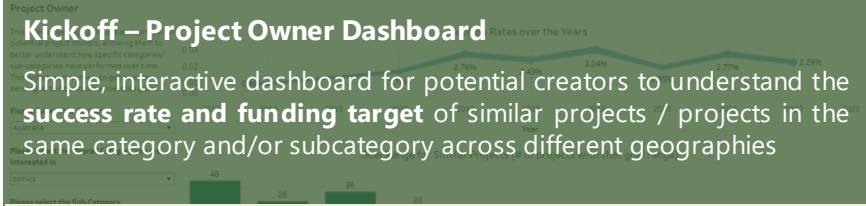
### Success rates (%) and No. of Projects over time



Source: Kickstarter

### Guidance is important to creators

Number of projects picked up in 2014 and success rates grew steadily over time since then. This coincides with the release of the creator's handbook in 2014, which suggests that **creators constantly seek and adopt advice to make their fundraising successful**



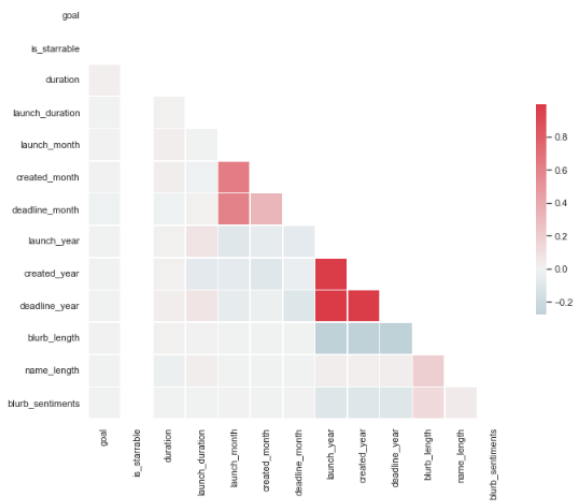
**Why?**

While creators should think about funding targets based on their own needs, this dashboard **helps in goal-rationalisation and budgeting.**

Link to Kickoff: <https://public.tableau.com/app/profile/adelyn.koh/viz/ConsultingforAnalytics-Kickstarter2/ProjectOwnerDashboard>

# Model Development

## Multicollinearity



- Severe Collinearity**  
created\_year, deadline\_year and launch\_year
- Some Collinearity**  
created\_month, deadline\_month and launch\_month
- Decision**  
Projects can be listed up to 60 days, thus, the month information is critical and will be kept.

- We will thus be using the following features for our models**
- Goal
  - Parent category
  - Child category
  - Campaign Duration
  - Launch Duration
  - Launch Month
  - Created Month
  - Deadline Month
  - Created Year
  - Launch Day
  - Created Day
  - Deadline Day
  - Country
  - Blurb Length
  - Name Length
  - Blurb Sentiment Score

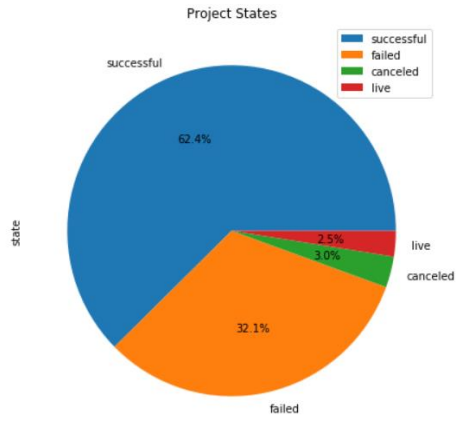
Source: Dataset used excludes certain countries such as Russia due to lack of availability

## Scaling and Oversampling the data

### 1 Scaling/ Standardize of Data

As we are utilising PCA analysis for dimensionality reduction, we need to scale the data to standardize columns prior to model fitting.

### 2 Oversampling of Data (Assigning different weights instead of SMOTE)



**2x**

The number of successful projects is twice of that of the number of failed projects<sup>1</sup>.

Thus, **oversampling is necessary** to ensure both failed and successful projects are equally represented in the model and will be less biased towards predicting successful projects. (we assigned different weights)



# Choosing and evaluating the best model – XGBoost

## XGBoost is chosen as the preferred model

Method	ROC-AUC	Precision	Recall
Logistic Regression	0.89	0.77	0.80
Random Forests	0.88	0.79	0.81
XGBoost	0.919	0.919	0.835

XGBoost outperforms random forest and logistic regression models

Method	Training Accuracy	Validation Accuracy
Random Forest	0.95	0.83
XGBoost	0.92	0.83

Random Forest suffers from slightly more severe overfitting

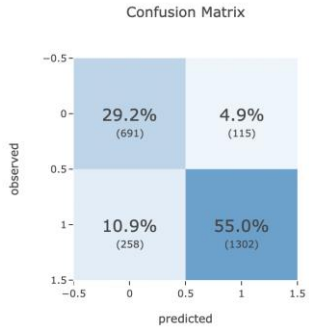
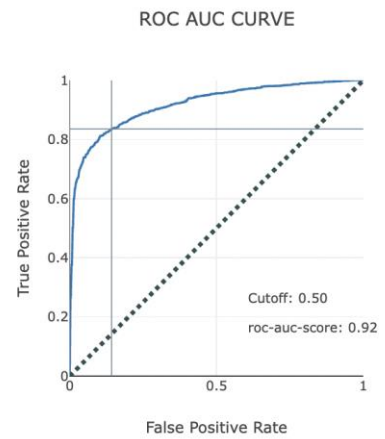
- ✓ XGBoost emphasises on functional space over hyperparameters in model optimization which is more **efficient**
- ✓ XGBoost is **good for unbalanced dataset**
- ✓ XGBoost is **less likely to overfit**

**XGBoost** is the model of choice

## Evaluating XGBoost on Test Set

Metrics	Scores
Accuracy	0.842
Precision	0.919
Recall	0.835
ROC-AUC	0.919

The ROC plot shows a curve that is close to the top-left corner, indicating a relatively good performance.



**What we prioritise**

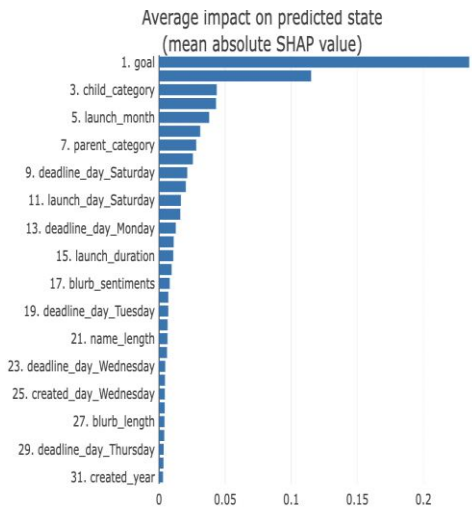
**Minimising False Positive (%)**  
Inaccurately predicts failed projects as successful ones

**Observations**

% of false positives is **small at 4.9% (115)**

# Which features are the most important in project success?

## Key Features



Feature	SHAP Rank
Goal	1
Main category/Sub-category	2/4
Launch Month	3
Deadline (Sat)	5
Launch Day (Sat)	6
Deadline (Mon)	7
Launch Duration	8
Blurb Sentiments	9
Deadline (Tues)	10

**What is SHAP Plot?**

Shows the effect a single feature has on the predictions made by the model

**Model Explainer**

We created a [model explainer](#) for the Kickstarter data scientist to iterate on and propose future recommendations to the management

Link to model explainer: <https://kickstarter-explain-e2s6bu6tja-uc.a.run.app/>

## Insights

Why do we think it's important (Intuition)



**Goal Amount is the most important feature** – makes intuitive sense as campaign success is decided by whether pledge amount meets the goal

**2**

Categorical Preference

Categories such as Games/ Design, attract more backers than categories such as Crafts/ Journalism as they likely appeal to a wider range of audiences

**3**

Launch Timing/ Day

Product releases are generally time sensitive. i.e., a video game might attract more backers if launched during certain seasons such as holidays

**4**

Project Descriptions

A well-crafted blurb might drum up more interest in a project and lead to more backer interest.

How can management act on these insights?

Educate creators on how they can utilize the dashboard by inputting variables such as launch dates to view the success rates of other projects with the same variables

# KickFar Initiative – Allowing creators to predict campaign success

We have designed a predictive dashboard with our model to provide actionable insights to creators...

... to predict the success of their campaigns before launch

**Predictions**

### Actionable Predictions

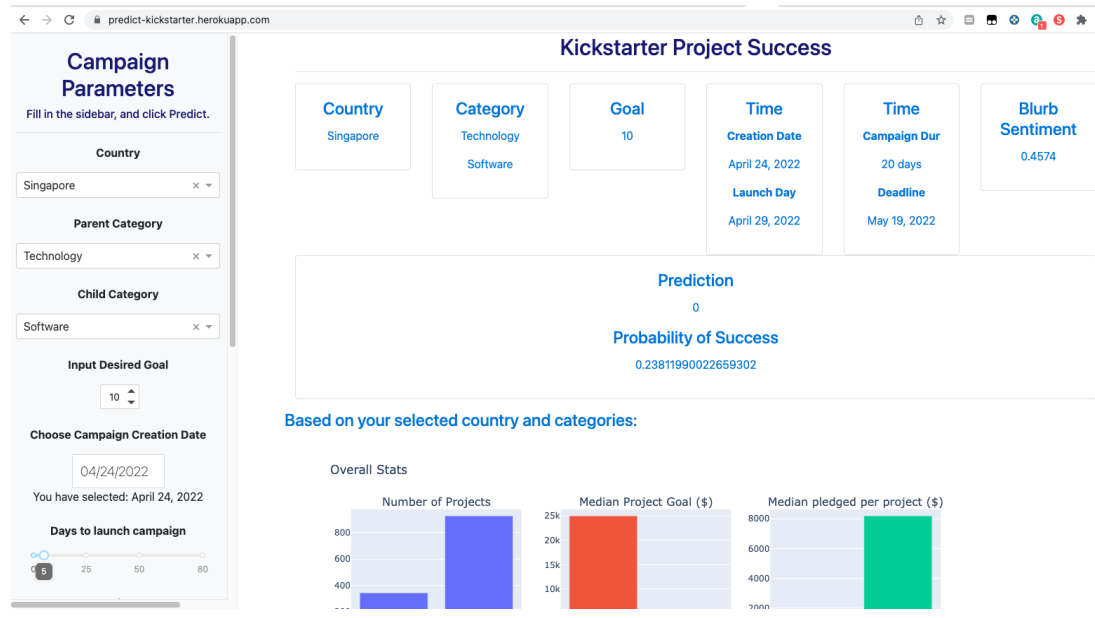
<b>Fast</b>	Minimal load time for creators to tweak their campaign parameters
<b>Actionable</b>	Campaigners can immediately see the probability of success, and tweak their campaign parameters
<b>Insights</b>	Provides basic insights on all projects in their selected country and categories.

**KIV**

### Future recommendations

<b>Fast</b>	Improve load speeds by upgrading to better servers
<b>Actionable</b>	Provide creators with actual recommendations to improve their campaign parameters
<b>Risk</b>	Incorporate forecasting models to forecast pledge growths over time to minimize risk

Built using Python Dash, Hosted on Heroku



Link to predictive model dashboard: <https://predict-kickstarter.herokuapp.com/>

# KickHigh – the all-round solution to combat high targets and stretch goals

## Problem: Stretch Goals & High Targets

**1 Mechanics of stretch goals**

Project meets fundraising target easily and **rides on the hype** to achieve stretch goals

Project inputs stretch goals in comments to deliver **additional value**

Project inputs "official" funding target to achieve **bare minimum MVP**

Stretch goals is an organic creation by the Kickstarter community to get around AON model and achieve minimum funding

**2 Why are stretch goals risky**

<b>Lemon Effect</b>	Creators know more than backers and may deliver products that are lower-than-expected quality if projects only meets minimum funding
<b>Project Complexity</b>	Stretch goals make projects difficult to complete and creators may fail to honour initial promises to backers

Source: Kickstarter

## Solution: Equity Crowdfunding Partnership

**3 Mechanics of equity crowdfunding**

Why did Kickstarter not do it organically?

1. Regulatory Complexity
2. Financial Complexity

**4 Kickstarter can partner and earn through project referrals**

Equity Crowdfunding Platforms	Private Market Equity Platforms

IDENTIFY	REFER	WIN-WIN
Kickstarter can screen for <b>high potential, high target and stretch goals</b> projects	Kickstarter can refer these projects onto equity crowdfunding platforms on a <b>revenue sharing basis</b>	<b>Kickstarter earns referral revenue</b> while <b>creators</b> are more likely to <b>attain sustainable funding</b>

# Conclusions

## Our Winning Strategy

### KickOff

*Dashboard for interested creators-to-be*

Help creators-to-be decide whether Kickstarter is the right platform to raise funds and rationalise fundraising targets

### KickFar

*Predicting campaign success for creators*

Help creators understand their probability of campaign success from campaign parameters

### KickHigh

*Project creation & goal guidance for creators*

Help to direct creators who struggle to raise funds or whose needs extend beyond small raises to equity crowdfunding platforms

## Limitations and Recommendations

### 1 Limitations

#### Reiterated Projects

- A handful of Kickstarter projects are a continuation of previous projects, and are creating new and improved versions of a product
- These projects' success are highly dependent on the success of previous projects and already have a large pool of trusted backers

#### Level of Engagement

- Features such as the FAQs displayed, number of updates or responding to comments are unavailable and hard to collect
- Reflects the level of engagement by the creator which is crucial in building confidence and trust among the backers

### 2 Model lifecycle and management

#### Objective

- Capture new data for continuous learning
- Retrain models so they continually adapt to the dynamically changing customer segments of the company

#### Benefits (Contextualised)

- Prioritise projects that have high chance of success but might fail due to lack of visibility
- Track effectiveness of creator's success rate

#### Future considerations

- Adopt the use of alternative data – competitor data, credit card data (when signing up as a member), geospatial data to further segment the customers by income and location

# Appendix - More Initiatives – Improvements to Project Creation Dashboard

We can further improve the predictive dashboard by actively recommending creators how to improve their campaigns..

... through our all-improved project creation dashboard

Spotlight

## Spotlight

<b>Data Used</b>	Historical Pledged Amounts Over Time
<b>Action</b>	Upsell carousel / banner ads on platform for projects that have high potential but need the extra push
<b>Risk</b>	Promote lemon projects

FC

## Funding Comparison

<b>Data Used</b>	Goal Range for Similar Projects (Tableau)
<b>Action</b>	Besides the funding calculator, creators can compare funding targets of past successful projects to rationalise goals
<b>Risk</b>	Creators may feel limited by the guidance

RTA

## Reward Tier Analysis

<b>Data Used</b>	Average Amount Pledged per Backer for the Category
<b>Action</b>	Besides pledge breakdown, creators can compare pledges with previous projects to boost marketing efforts
<b>Risk</b>	Additional fundraising stress leads to increased project cancellation rates

Consult

## Specialist Consultation

<b>Data Used</b>	Model
<b>Action</b>	Kickstarter specialists can help to breakdown creator's plans into bite-size goals and optimise fundraising using model guidance
<b>Risk</b>	Each project require a different, tailored solutions

✓ Offer recommendations to creators to improve success rates

✓ Provide KPI comparison against previous projects

*PMO tool*

Problem Analysis
Exploratory Data Analysis
Model Development
Model Evaluation
Conclusions

14

# Appendix – Project Creation Current Interface

## Funding goal

Set an achievable goal that covers what you need to complete your project.

Funding is all-or-nothing. If you don't meet your goal, you won't receive any money.

Goal amount

SS	0
----	---



[Use our calculator](#) to estimate total costs, including taxes and fees.

## Pledge Breakdown






The pledge breakdown at the top of the Fulfillment dashboard will help you to gain a clearer understanding of the final number of backers on your campaign page, and how this relates to fulfillment. You'll see the total pledges on the campaign, including the number of pledges with rewards and the number of pledges without a reward.

If your campaign had any dropped pledges, a count will appear alongside the other pledge numbers. If there are no dropped pledges however, this information will not be displayed.

# Appendix – Project Creation Current Interface

## Project description

Describe what you're raising funds to do, why you care about it, how you plan to make it happen, and who you are. Your description should tell backers everything they need to know. If possible, include images to show them what your project is all about and what rewards look like. [Read more about telling your story](#)

Headline  **B** *i*    

Write about your project like you're explaining it to a friend...



## Allows individuals to look at more detailed analysis of the trained model

### Model Explainer

Positive class: 1

Feature Importances | Classification Stats | Individual Predictions | What if... | Feature Dependence | Decision Trees

**Global cutoff**

Cutoff prediction probability:

Cutoff percentile of samples:

**Model performance metrics**

metric	Score
accuracy	0.842
precision	0.919
recall	0.835
f1	0.875
roc_auc_score	0.919
pr_auc_score	0.959

**Confusion Matrix**  
How many false positives and false negatives?

		Predicted	
		0	1
Observed	0	29.2% (691)	4.9% (118)
	1	10.9% (258)	55.0% (1302)

**Individual xgboost decision trees predicting state**

**Average impact on predicted state (mean absolute SHAP value)**

Feature	Impact
1. goal	High
3. child_category	Medium
5. launch_month	Medium
7. parent_category	Medium
9. deadline_day_Saturday	Medium
11. launch_day_Saturday	Medium
13. deadline_day_Monday	Medium
15. launch_duration	Medium
17. blurp_sentiment	Medium
19. deadline_day_Tuesday	Medium
21. name_length	Medium
23. deadline_day_Wednesday	Medium
25. created_day_Wednesday	Medium
27. blurp_length	Medium
29. deadline_day_Thursday	Medium
31. created_year	Medium

Feature Importances | Classification Stats | Individual Predictions | What if... | Feature Dependence | Decision Trees

**Select Index**  
Select from list or pick at random

2019 | Random Index

Observed state: 0 | Range: probability

Predicted probability range:

**Prediction**

Index: 2019

label	probability
0	3.3 %
1*	96.7 %

\* indicates observed label

**Contributions Plot**  
How has each feature contributed to the prediction?

Index: 2019 | Depth: | Sorting: High to Lo

Contribution to prediction probability = 96.7%

**Partial Dependence Plot**  
How does the prediction change if you change one feature?

Feature: goal | Index: 2019

**Shap Dependence**  
Relationship between feature value and SHAP value

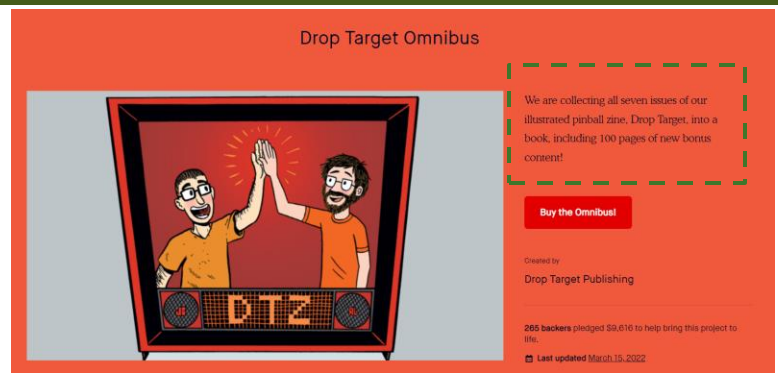
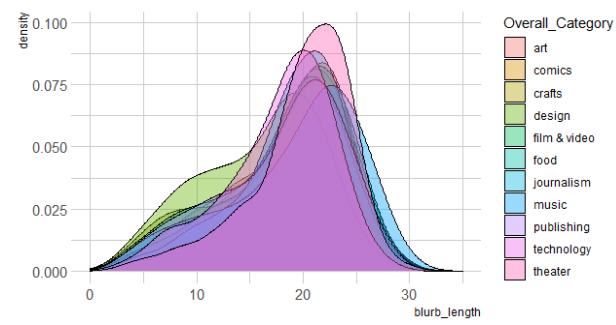
Feature: goal | Color feature: deadline\_day\_Sun | Index: Select...

**Dependence plot for goal**

# Appendix – Additional Analysis (Textual Analysis)

**Hypothesis:** There are specific keywords that make a project listing more appealing / less appealing than others

Probability Density Curve of Projects by Blurb Length based on Category



**Short and Sweet:** Most projects have less than 35 words in their blurb

	word	success	success_rank	fail	failed_rank
1	album	1462	1	316	13
2	book	1143	2	454	6
3	world	959	3	504	4
4	music	884	4	524	3
5	story	811	5	240	25
6	series	803	6	297	15
7	love	731	7	372	9
8	graphic	682	8	28	590
9	life	679	9	449	7
10	film	664	10	262	20

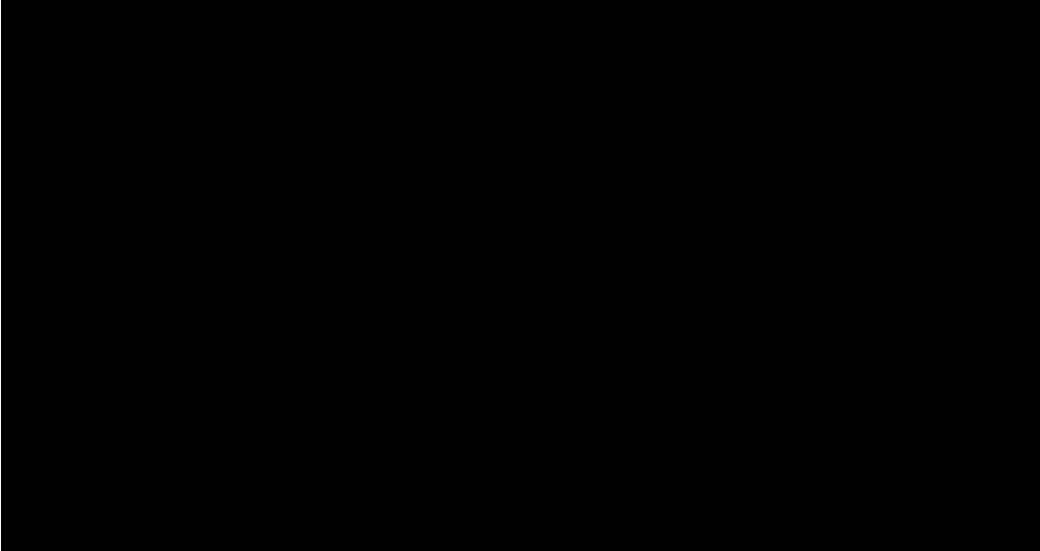
**We analysed unigram frequency of titles between successful and failed projects to notice any difference:**

1. Most words appear in equivalent proportions for both successful and failed projects (on a ranking basis)
2. Graphic stood out to be a special word as more successful project had that – but upon further interpretation, it was associated with graphic novel subcategory, which had an extremely high success rate

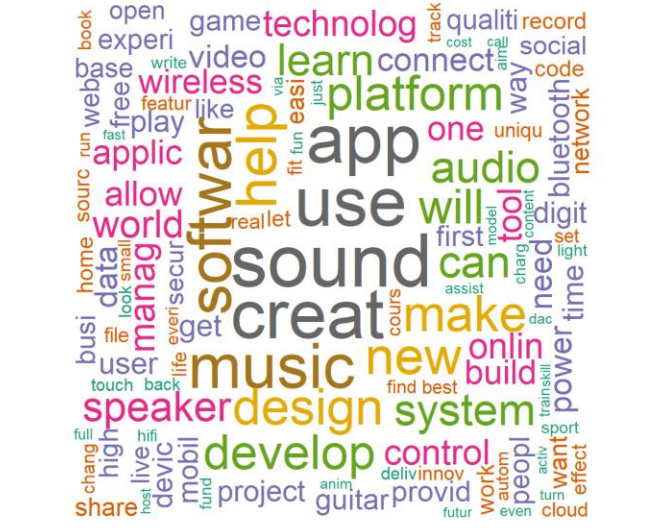
# Appendix – Additional Analysis (Textual Analysis)

**Hypothesis:** There are specific keywords that make a project listing more appealing / less appealing than others

## Bigram Analysis



## Wordcloud



We also did Hierarchical Clustering, K-Means, K-Medoids (removed from code as there were no insights) but ultimately determined that there were **no meaningful relationship between specific words and success of the project**. In fact, many words were project-specific and clusters did not form according to categories / subcategories.