



Team Kickstarter

Descriptive Statistics and Results from Data Exploration

Allison Bass, Dylan Juarez, Prakash Rao,
Franklin Tan, Alfonso Alday

BANA 5160 | 12-JUL-2023

KICKSTARTER

Hypotheses and Approaches



Key Hypotheses

- a) The most popular categories have the highest success rates
 - b) High backer count on a project acts as social proof, increasing the likelihood of individuals backing the project, regardless of funding goal status
 - c) A higher funding goal commands more pledges
 - d) There are certain times of year where pledge amounts are higher
 - e) As the platform grows, Kickstarter has failed to increase the pledges per project
-



Proposed Approaches

- a) Look at descriptive statistics to see if the categories with the largest number of projects also have the highest proportion of successes
- b) Research whether the number of backers has a strong impact on a project's likelihood of success
- c) Determine if there is a correlation between pledge amount and a project's goal
- d) Look at average pledge amount by month throughout the year
- e) Find the number of pledges per project for every month since Kickstarter's inception (2009) to see if there has been an increase over time

Data Quality Assessment

The starting point for further data exploration, data cleaning, and analysis tasks

#	Column	Non-Null Count	Dtype
0	ID	378661 non-null	int64
1	name	378657 non-null	object
2	category	378661 non-null	object
3	main_category	378661 non-null	object
4	currency	378661 non-null	object
5	deadline	378661 non-null	object
6	goal	378661 non-null	float64
7	launched	378661 non-null	object
8	pledged	378661 non-null	float64
9	state	378661 non-null	object
10	backers	378661 non-null	int64
11	country	378661 non-null	object
12	usd pledged	374864 non-null	float64
13	usd_pledged_real	378661 non-null	float64
14	used_goal_real	378661 non-null	float64

- **Missing Values:** The "Non-Null Count" column reveals the presence of missing values in the dataset, particularly "name" and "usd pledged"
- **Numerical Columns for Analysis:**
 - "goal", "pledged", "backers", "usd pledged", "usd_pledged_real", "used_goal_real")
- **Categorical Columns for Grouping, Aggregation, or Classification Purposes:**
 - "category", "main_category", "currency", "deadline", "launched", "state", "country"

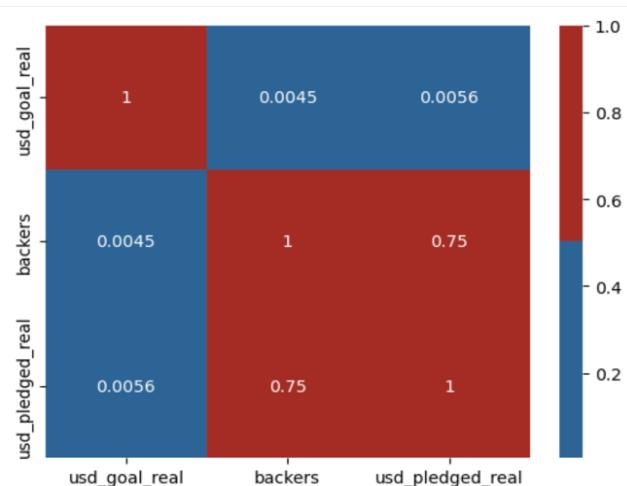
Our next steps involve addressing missing values, addressing categorical variables, scaling or normalizing numeric variables, and definite the target variable.

Multicollinearity

Several features in the Kickstarter dataset are highly correlated

ANALYSIS

1 Kickstarter Correlation Matrix



INSIGHT

There is a strong positive correlation between **backers** and **money pledged**

2 VIF Factor for Features

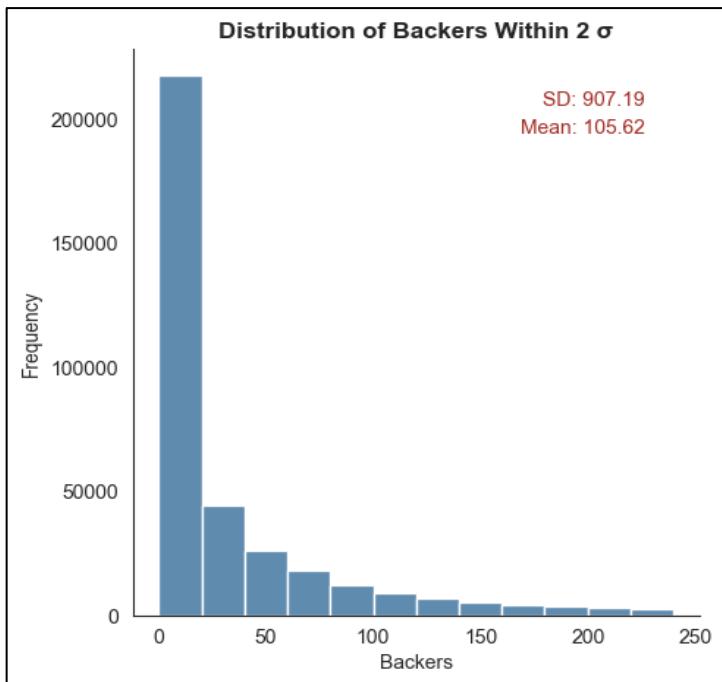
Feature	VIF Factor
year_launched	61468829
year_deadline	61467628
month_deadline	14
month_launched	13
day_launched	5
day_deadline	5
backers	2
usd_pledged_real	2
usd_goal_real	1



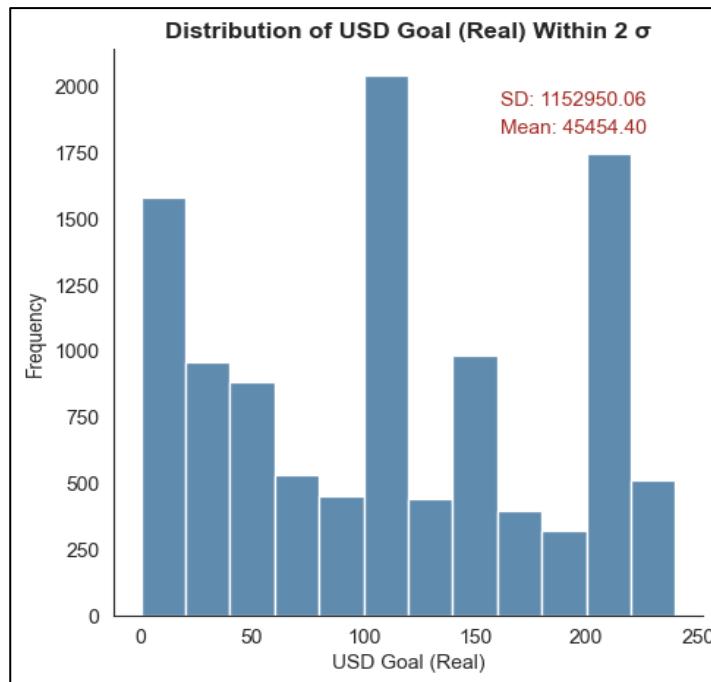
The numerical VIF values greater than 5 suggest **severe multicollinearity**

Distribution of Numeric Columns

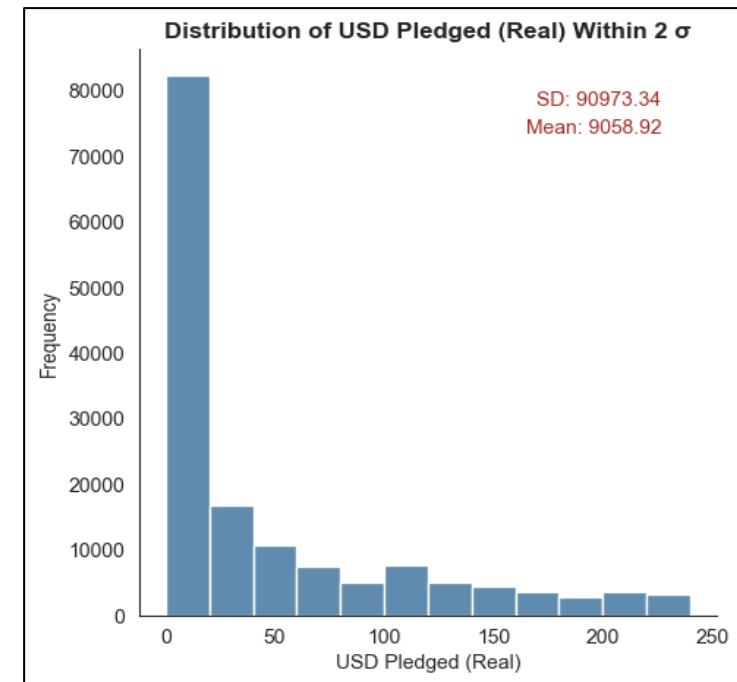
There is significant variability in the number of backers, campaign goals, and amount pledged



Positively skewed, with a few campaigns having a significantly higher number of backers compared to the majority



Wide range of funding targets set by campaigns, ranging from relatively small to very large goals.

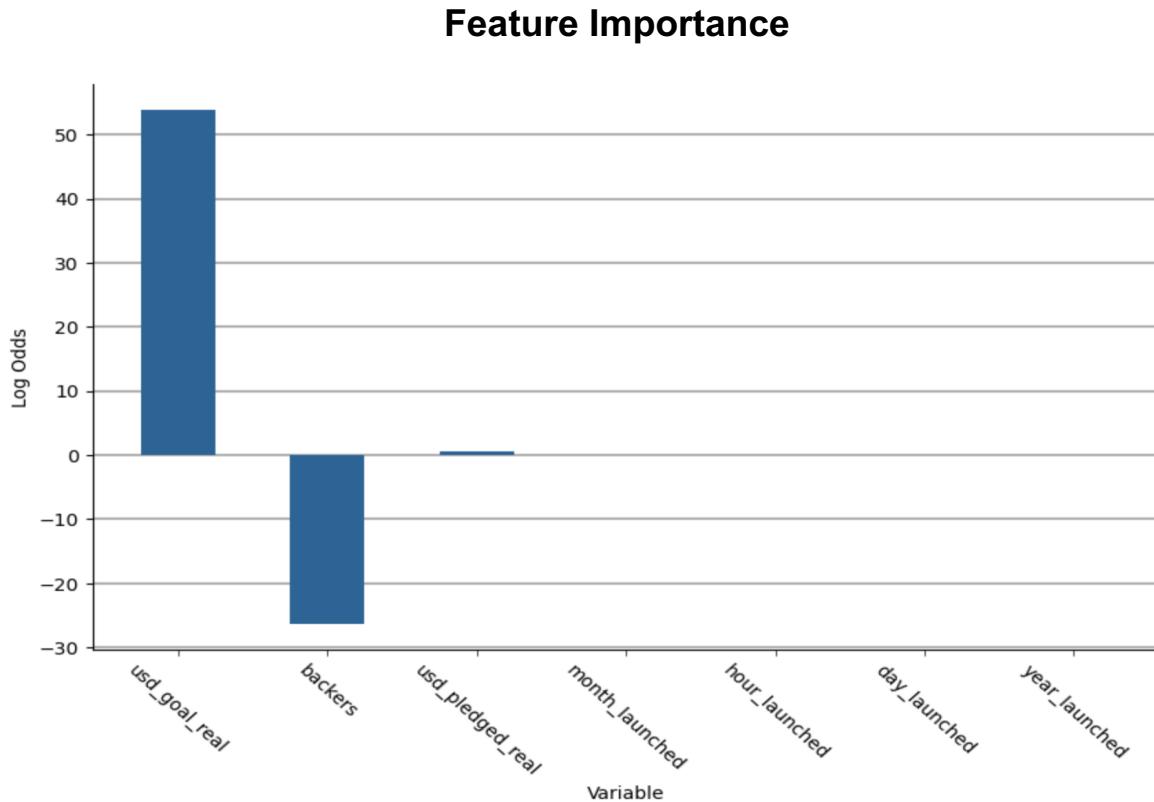


Positively skewed, with a few campaigns having a significantly higher amount of pledged funds compared to the majority.

There are campaigns with both low and high levels of engagement and funding, indicating a diverse range of outcomes.

Feature Importance

Fundraising goals and backers are the most important predictors of project failure



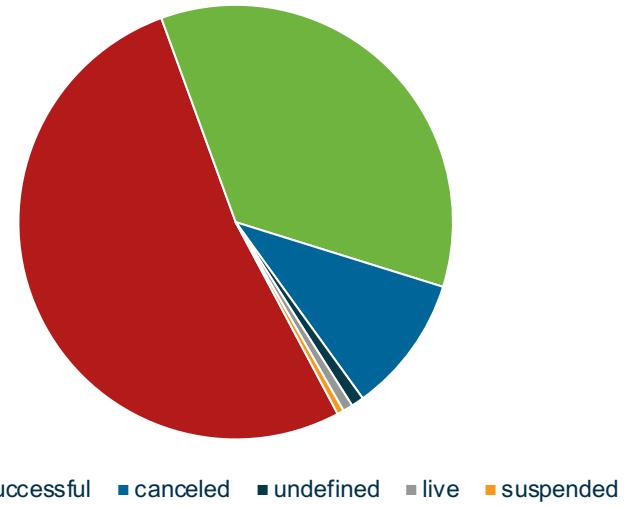
- **usd_goal_real:** Higher campaign goals in terms of USD are more likely to lead to project failure.
- **backers:** An increase in the number of people that contribute to a given crowdfunding project raises the likelihood of its success.

Distribution of Project States

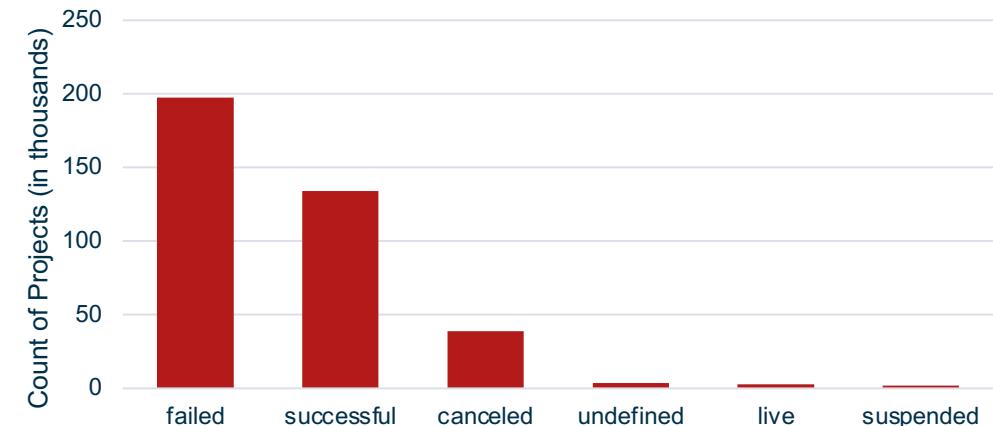
Failures make up the highest proportion of projects

- The dataset provides information for 378,662 projects from 2009 – 2017.
- By distributing the projects into states we find that 52.22% of projects started fail.
- Further consolidation of states into successful or failed, which is coherent with the business model, reveal that Kickstarter's problem is more serious.
- Failed + Canceled + Suspended projects represent **62.94% of the projects.**
- Successful projects represent **35.38% of the projects.**
- Undefined and live projects represent **1.68% of the projects.**

Project States as Percent of Total



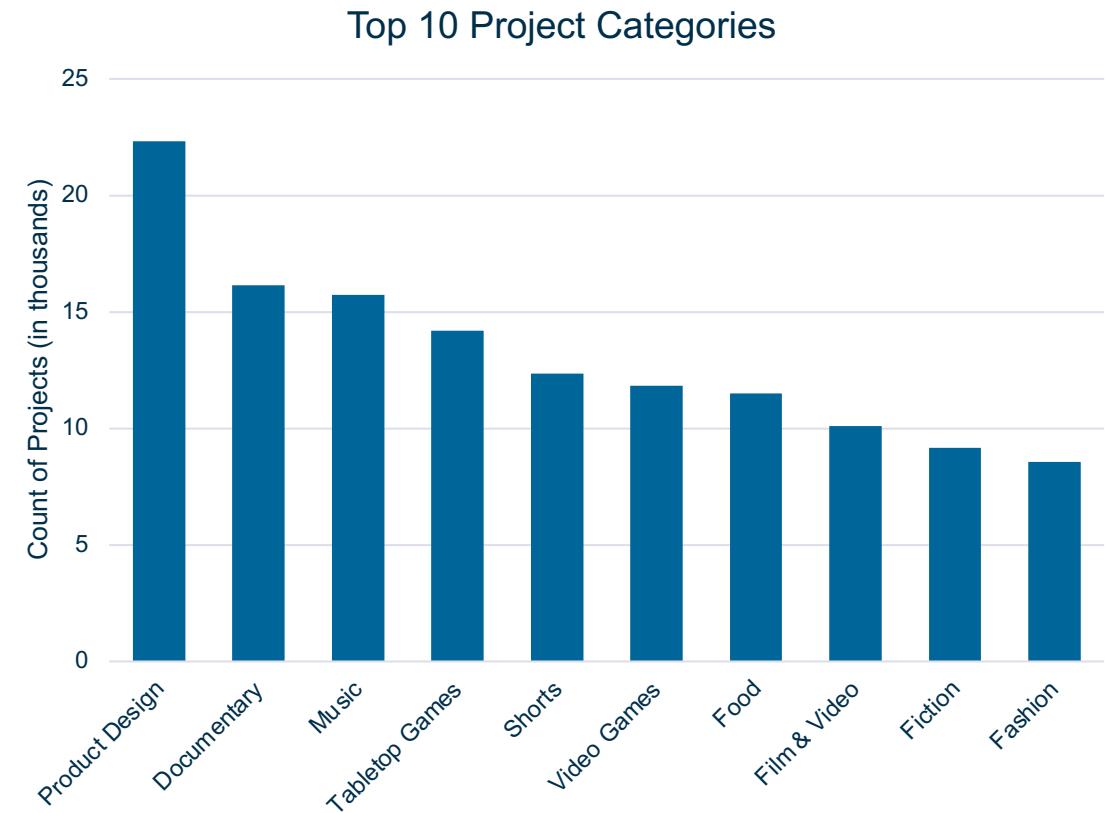
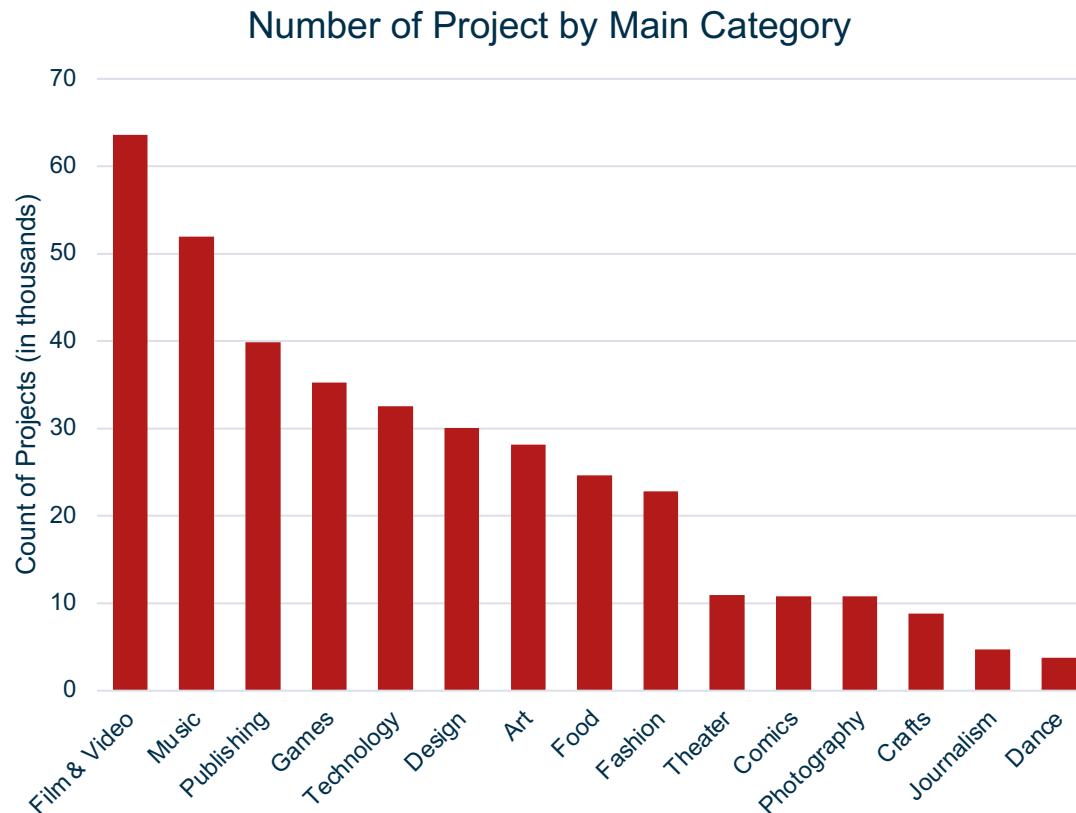
Distribution of Project States



Distribution of Categories

Kickstarter has 15 main categories and ~159 categories

- The main category with the most projects is Film & Video
- The category with the most projects is Product Design, which has a 57% failure rate.

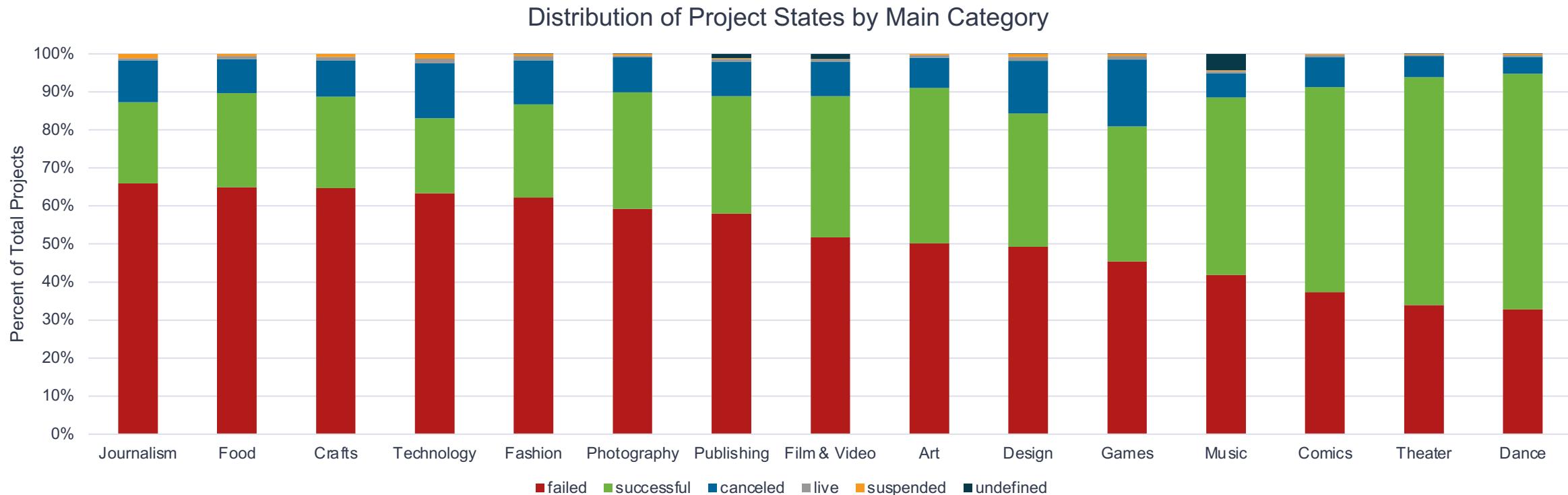


Project States by Category

Categories with fewer projects tend to have a higher success rate

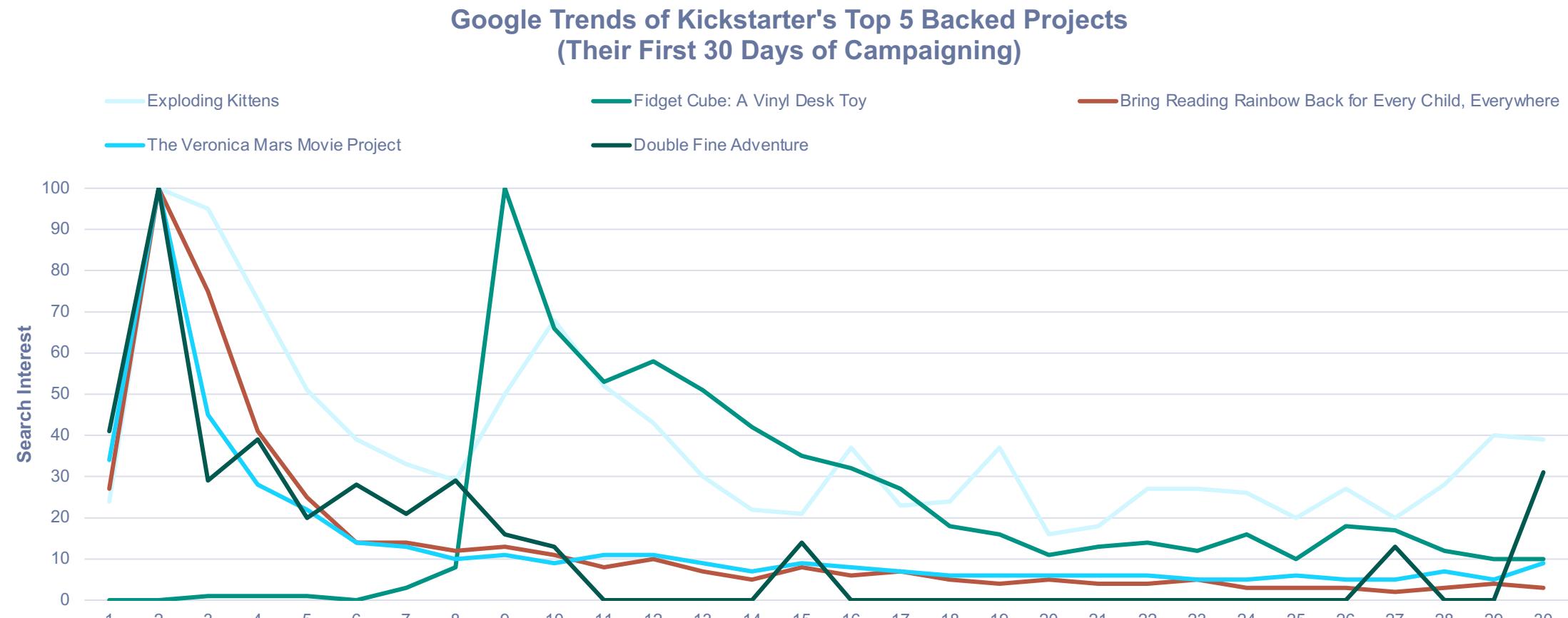
The most successful Main Categories are Dance (62.05% SX), Theater (59.87% SX), and Comics (54% SX).

- Despite their success, these Main Categories represent 1%, 2.88%, & 2.86% of the total number of projects.



Analyzing the First 30 Days of Kickstarter's Most Backed Projects

Early bird gets the worm and early interest is crucial to generating momentum.

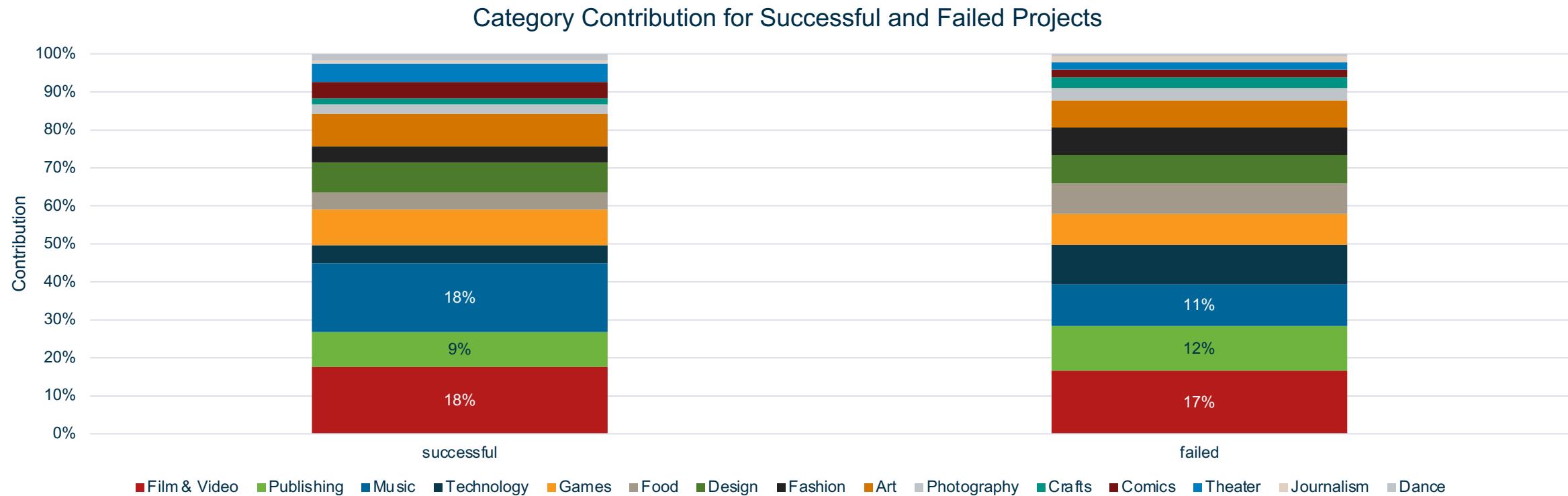


Early engagement and marketing efforts are crucial for creating awareness, building initial momentum, and sustaining interest over time.



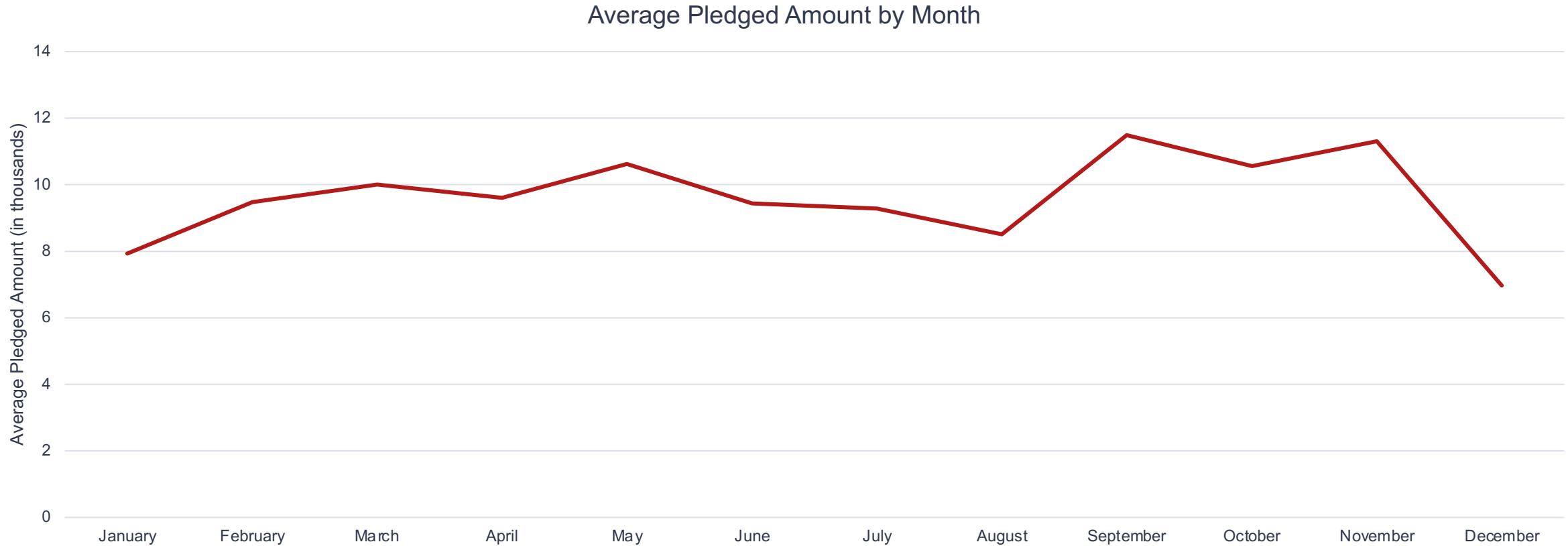
Category Contribution

Film & Video, Publishing, and Music make up the largest proportion of successes and failures



Seasonality

The dataset revealed the existence of seasonality among contributions

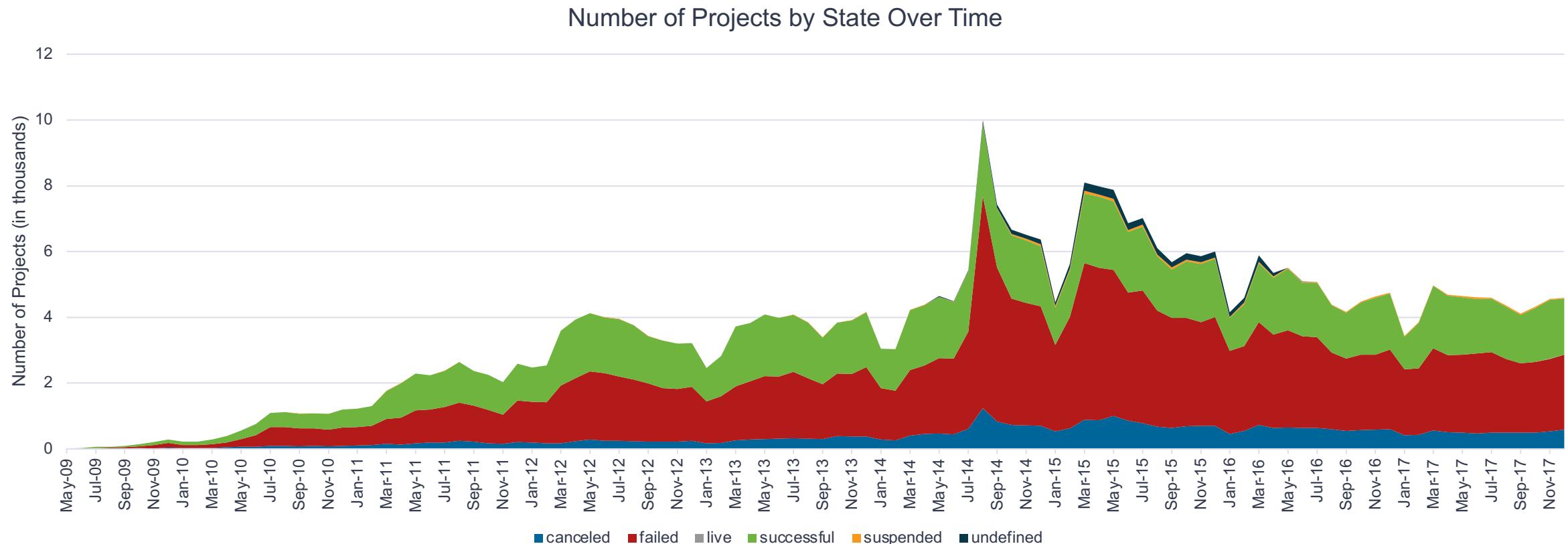


Summer and Winter experience a dip in contributions; This behavior remains steady across the 9 years of the dataset.

Changes in State Over Time

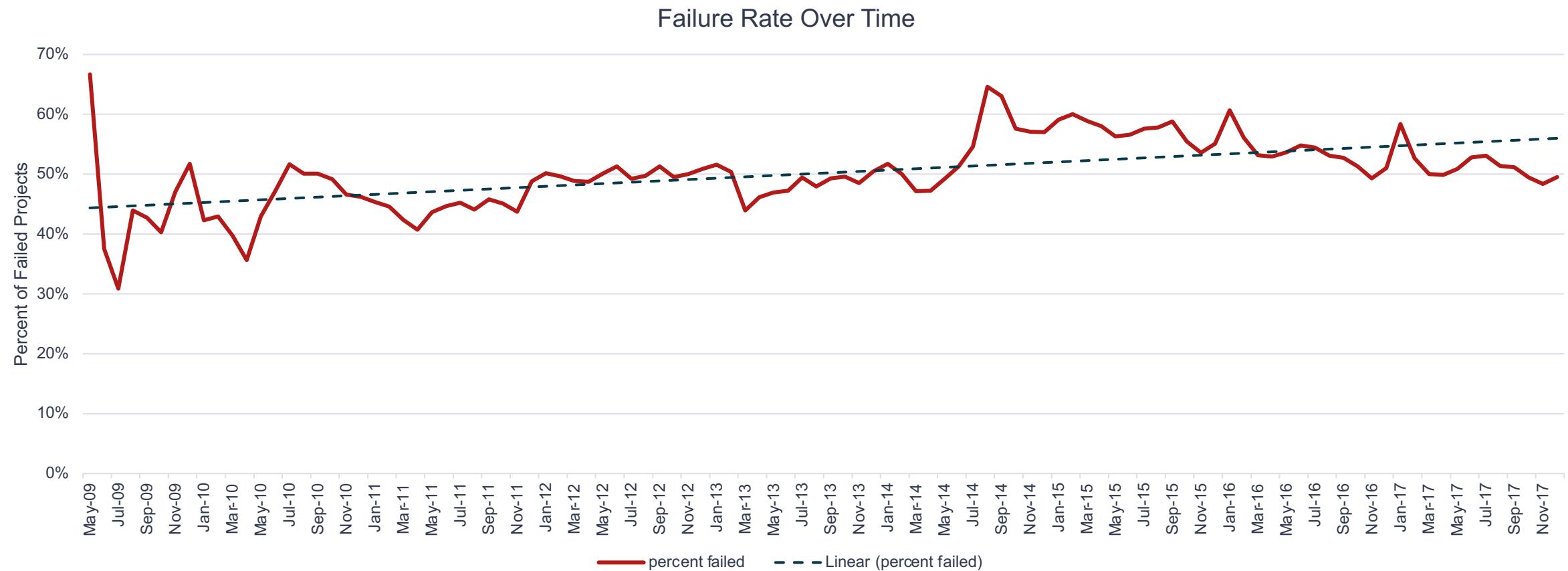
The platform has seen growth in the number of projects since 2009

- The platform's growth peaked in September 2014
- The project state distribution remains consistent across the board.



Failure Rate Over Time

As the platform grows, the proportion of failures has increased



Pledged Amount vs. Funding Goal

There does not seem to be a relationship between pledge amount and funding goal



The scatterplot indicates a high failure rate among KS projects, possibly to ambitious goal setting, and indicates the presence of additional factors influencing pledged amounts.