# Predicting Kickstarter Project Success or Failure

By: Leshauna Hartman, Rachel Thomas, Tanya Visser, and Fardin Hafiz

## **1. Introduction**

A head with gears in the brain

Description automatically generatedKickstarter is a crowd-sourcing platform that allows creators of all sorts share their prospective work to attract community funding. It launched in 2009 and has since become a Public Benefit Corporation that has funded more than 270,000 projects and raised more than $8 billion dollars. Kickstarter creators post their ideas or projects on the website and share it with their friends and supporters, who, in turn, share the project in their networks, and so on. More than 24 million people from all over the world have helped fund Kickstarter campaigns. Projects cover a wide range of categories including art, publishing, design, and technology. “Kickstarter lifts the creative class, gives people the tools to pursue daring ideas on their own terms, and helps creators build communities around their work” ([www.kickstarter.com](http://www.kickstarter.com)).

We are interested in predicting Kickstarter project success or failure and identifying the characteristics of a project that most influence this outcome. This has practical implications as creators can tweak their projects to reflect a higher chance of success. Additionally, backers can make better decisions about which projects to fund. This saves time, money, and resources for both the creators and the backers.

Kickstarter is an “all or none” funding scheme, meaning if a campaign does not raise the entirety of its goal funding, it gets zero funding, all the money pledged everts to the donors, and the project fails.

## **2. SMART Questions**

Our research questions were designed using the SMART framework – Specific, Measurable, Achievable, Relevant, and Time-bound. This analysis seeks to answer the following:

1. Which variables most influence success or failure?
2. Can a logistic regression model accurately predict the success or failure of a Kickstarter campaign?
3. Can a Decision Tree accurately predict the success or failure of a Kickstarter campaign?
4. What are the top five (5) categories with the highest number of successes?
5. What percentage of all campaigns were successful compared to failed?

Through these questions, we aim to determine if Kickstarter campaign outcomes are predictable and identify which features/attributes contribute to the success of a campaign.

## **3. Dataset Description & Preparation**

**Dataset Overview**

The dataset was sourced from Kaggle and contains 378,661 observations with 15 variables. These variables include:

* **backers:** The total number of backers who supported a project.
* **currency:** The currency in which the project was originally launched.
* **country:** The country from which the project was launched. (*Categorical*)
* **main\_category:** The primary category of the project (e.g., Music, Technology).
* **state:** The final status of the project, indicating whether it was successful, failed, or canceled.
* **usd\_pledged\_real:** The total amount of money pledged to a project in USD.
* **usd\_goal\_real:** The funding goal set by the project creators in USD.

**Data Preprocessing/Cleaning**

The dataset was sourced from Kaggle.com. It was subsetted to include only campaigns that were successful or failed – rows reflecting any other campaign states were removed. This reduced the dataset to approximately 331,000 observations. A variable called Duration was created by finding the difference between the “launched” and “deadline” variables. “main\_category”, “currency”, “state” and “country” variable were converted to categorical data type. A subset called kickstarter\_final was created to include “main\_category”, “currency”, “state”, “backers”, “country”, “usd\_pledged\_real”, “usd\_goal\_real”, and “Duration.

## **4. Literature Review (if applicable)**

Online crowdfunding platforms are an increasingly popular way for ordinary people to finance a wide variety of projects ranging from creative arts to healthcare support. In a 2016 study, Ethan Mollick of the Wharton School at the University of Pennsylvania, reported that each dollar given to projects via Kickstarter resulted in a mean of $2.46 in additional revenue (though this was not spread evening though categories). He also reported that Kickstarter projects had resulted in more than 5,000 ongoing full time jobs besides those of the creators, and more than 160,000 temporary positions. The successful campaign also resulted in more than 2,600 patent applications (Mollick, 2016). However, as more campaign have been launched, there has been an observed decrease in success rate, suspected to be due to

## **5. Exploratory Data Analysis**

- Descriptive Statistics

- Plots and analysis

**A red and green graph

Description automatically generated**

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

|  |  |
| --- | --- |
| **Top 5 Categories with the Highest Number of Successful Projects:** | |
| **Main Category** | **Count** |
| Music | 24197 |
| Film & Video | 23623 |
| Games | 12518 |
| Publishing | 12300 |
| Art | 11510 |

Identifying the top five categories with the most successful projects, the table above highlights the areas where Kickstarter campaigns tend to achieve the most success. These categories are Music, Film & Video, Games, Publishing, and Art, with Music accounting for the greatest number of successful projects with a value of 24,197, followed closely by Film and Video with 23,623 successful projects.

To make these insights more visually appealing and intuitive, these values can be represented in a bar chart as seen below:

## **A graph of a number of backers Description automatically generatedA graph of a bar chart Description automatically generated with medium confidence**

## **6. Modeling Techniques & Evaluation**

- *How did you select and determine the correct model to answer your question?*

*- Build Models*

*- Analyze/ Evaluate Models using evaluation metrics?*

**Logistic Regression**

## **7.Discussion**

*- Interpret results*

*- What Predictions Can You Make from Your Models? Examples?*

*- How good is your model?*

- *How reliable are your results?*

\*\* LIMITATIONS:

*- What additional information or analysis might improve your model results or work to control limitations?*

## **8. Conclusion.**

- Summary

- How Do These Answer the Research Questions?

## **9. References (APA Style)**

Mollick, Ethan R., Containing Multitudes: The Many Impacts of Kickstarter Funding (July 11, 2016). Available at SSRN: <https://ssrn.com/abstract=2808000> or [http://dx.doi.org/10.2139/ssrn.2808000](https://dx.doi.org/10.2139/ssrn.2808000)