ISyE 6740 – Summer 2025

Project Proposal

Team Member Names: Phat Dat Tieu

Project Title: What makes a successful Kickstarter campaign

Please include (at least) the following sections.

Problem Statement

(Optional) Data Source

Methodology

Evaluation and Final Results

Evaluation and final results : not required for project proposal, but please include the metrics you expect to evaluate.

Write a proposal for topic “What makes a successful Kickstarter campaign ?”, including the following sections:

Introduction / Background

Objectives / Purpose

Data: data size 250,000 x 16, can be up to 250,000 x 20, note that country is not an important factor for this task

Scope: will not cover rare categories (<2% of total populations)

Methodology / Approach: give some suggestions to perform the task, and expected metrics to evaluate

Expected Outcome

Significant / Impact

Some requirements: minimum 3 pages, or at least 1200 words.

Make the approach less about prediction, and more feature selection, while keeping other phases the same. Some variables can be considered:

Columns:

1. backer counts
2. description
3. category
4. pledge amount
5. created time (datetime)
6. creator
7. currency
8. deadline (datetime)
9. goal (fund goal)
10. is disliked
11. is liked
12. is starrable
13. launched time (datetime)
14. Name
15. percent funded
16. prelaunch activated (true / false)
17. is spotlight
18. is staff-pick
19. has video (true / false)

Can add more features if needed, and data is labeled by state (successful / failed)

ISyE 6740 – Summer 2025

Project Proposal

Team Member Names: Phat Dat Tieu

# Project title: What makes a successful Kickstarter campaign?

## Introduction

Since its launch in 2009, Kickstarter has become one of the most well-known crowdfunding platforms, helping creators to release products across a wide range of categories, including art, design, technology, games, and more. With over hundreds of thousands of campaigns launched and billions of dollars raised to date, Kickstarter has established itself as a go-to site for turning creative ideas into funded realities. However, it’s never easy to be successful on Kickstarter. In fact, a significant proportion of campaigns fail to meet their funding target, resulting in zero funding due to the platform’s all-or-nothing model.

Many creators have attempted to determine the “formula” for a successful campaign — from optimal funding goals, engaging videos to the careful launch timing. While anecdotal observations and informal strategy guides exist, few researches have systematically analyzed the comprehensive mix of campaign attributes that greatly affect the success using large-scale data.

This proposal suggests a data-driven study to investigate which features are most effectively associated with successful Kickstarter campaigns. The findings can guide creators and platform developers in prioritizing key factors during campaign planning and optimization.

## Objective

The primary objective of this project is to **identify and analyze key features** that are associated with successful Kickstarter campaigns. Rather than predicting outcomes, the goal is to understand **which factors matter most** and **why**, through feature selection and interpretability-focused techniques. Some s**pecific objectives include:**

* **Perform exploratory data analysis (EDA)** to understand patterns of success across categories, time periods, and other campaign metadata.
* **Visualize and interpret relationships** between key features and campaign outcomes.
* **Select and rank important features** that influence campaign success, using both statistical tests and model-based importance analysis.
* **Evaluate robustness** of selected features using multiple methods (e.g., permutation, ablation analysis).

## Scope

We will analyze the data set of nearly 260,000 campaigns with 40 original variables, from 2009 to 2025 June, taken from latest dataset on [Web Robots](https://webrobots.io/kickstarter-datasets/) website. One can use more noble approach to gather data from well-established sources like Kaggle or scientific researches. Unfortunately, [Kaggle most popular dataset](https://www.kaggle.com/datasets/kemical/kickstarter-projects) last updated 7 years ago, and IEEE papers need authors’ permission to work with.

Back Web Robots dataset, each campaign is classified as “failed”, “successful” or “on going”. This analysis will **exclude “on going” products and geographic analysis**, as America-based products hold nearly 63% of whole population, while more than half of listed countries only accounts for less than 1% each. Additionally, **rare campaign categories** representing fewer than 100 items will be excluded to reduce noise and ensure adequate sample sizes.

We will not discuss the following topics: external marketing efforts (e.g., social media virality), effects of visual contents (photo, videos), full-text natural language processing of long descriptions, real-time user interaction (e.g., backer behavior during live campaigns). Instead, the scope will focus on features available **at or just before launch**, to focus on what creators can influence during campaign preparation.

## Methodology

The analysis will emphasize **feature selection, interpretation, and comparative importance**, not prediction.

### 1. Data Preprocessing and **Exploratory Data Analysis (EDA)**

* Encode categorical variables using techniques like one-hot encoding or target encoding.
* Analyze success rate trends across categories, funding goals, campaign duration, launch dates (seasonality),...
* Compare distributions of backer counts, goal amounts, and video usage across successful vs. failed campaigns.
* Use correlation matrices and visual tools (heatmap, histograms, boxplots) to understand variable distributions and relations.

### 2. **Feature Selection and Ranking**

Several approaches and their metrics will be used to identify which features matter most:

#### **Univariate Feature Tests**

#### Chi-square tests for categorical variables, ANOVA and t-tests for numeric features, and mutual information scores for mixed data types

#### **Recursive Feature Elimination (RFE)**

#### Logistic regression and random forest models are used to iteratively eliminate the least useful features and identify the optimal feature subset.

#### **Model-Based Feature Importance**

#### Train tree-based models (e.g., Random Forest, XGBoost) and extract Gini importance or SHAP (SHapley Additive exPlanations) values for interpretability.

#### **Permutation Importance and Ablation Analysis**

#### Randomly shuffle feature columns and evaluate impact on outcome-related metrics to identify high-leverage variables. Compare model accuracy or AUC before and after removing specific features to assess their real-world value.

### 3. **Interaction and Correlation Analysis**

* Use partial dependence plots and SHAP interaction values to understand how pairs of features influence outcomes.
* Check for collinearity (e.g., between goal, pledge, percent\_funded) and consolidate where needed.

### 4. **Interpretation and Synthesis**

* Visualize the most impactful features and generate interpretive narratives for each (e.g., "having a video increases likelihood of success by X%").
* Cluster features by thematic groups: presentation (e.g., video, description length), social proof (e.g., backer count), timing, etc.

## Expected Outcome

By the end of this project, we expect to deliver:

* **A ranked list of features** most closely associated with campaign success, supported by statistical and model-based evidence.
* **Visual and explanations** of how top features influence success, including charts, graphs, and interpretable thresholds (e.g., “campaigns with goals under $10,000 and a video have a 60% higher success rate”).
* **Actionable recommendations** for Kickstarter creators based on top factors — e.g., “Use a video”, “Launch during mid-week”, “Avoid overly ambitious funding goals without a pre-launch”.
* **Feature importance matrix or dashboard** (optional) for platform or business use.

This will provide a grounded understanding of which elements significantly contribute to campaign success—and how creators can use these insights proactively.

## Potential Impact

This research offers significant impact for both **Kickstarter developers, creators** and the **crowdfunding ecosystem**:

* **For Creators**: A feature-based success guide empowers creators to prioritize high-impact campaign components, improve planning, and reduce the risk of failure.
* **For Kickstarter (Platform-Level)**: Understanding feature salience can guide UI/UX enhancements, offer targeted suggestions during campaign setup, and improve creator support tools.
* **For Researchers and Analysts**: The project advances data science methods for feature selection and interpretability within real-world business platforms.
* **For Educators and Entrepreneurs**: Provides a data-driven reference for teaching best practices in crowdfunding, digital entrepreneurship, and product marketing.

By centering on feature analysis rather than prediction, the study aims to offer interpretable, actionable, and trustworthy insights—not just probabilities, but priorities.

Scope can add more what is covered, and can add some detailed metrics to evaluate our methodology