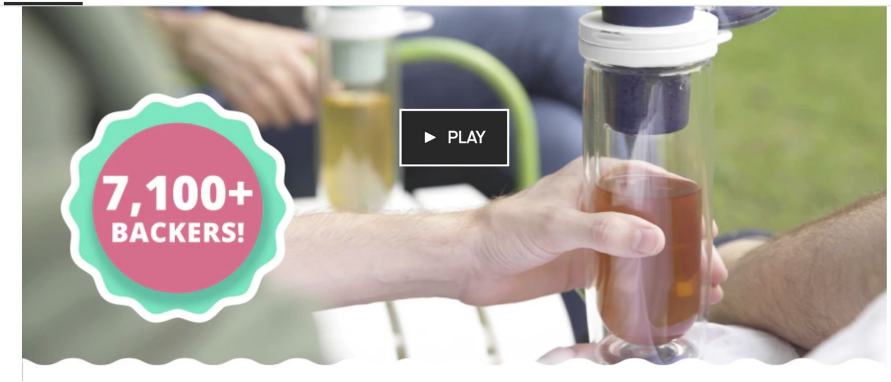
Kickstarter Analysis

An exploration into the influencers of Kickstarter project success

The Laureates: Farzeen Najam, Olivia Olsher, Vincent Liu, Émile Therrien

KICKSTARTER

Campaign | FAQ 1 | Updates 27 | Comments 1,450 | Community | Share this project | Save



Mosi Tea™ - Brew and drink loose-leaf tea anywhere

Project We Love | Durham, NC | Product Design

$458,200
pledged of $15,000 goal

7,114
backers

STORY

RISKS

Story

mosi

Introducing the World's Best Tea Infuser

Support

Pledge $32 or more

EARLY TASTER | MOSI TEA INFUSER

Save 36%! (Est. $50 MSRP)

↳ Receive 1X Mosi Tea Infuser + Standard Sieve

↳ Choose your color at the end of the campaign

↳ Add $5 to pledge to receive a Matcha Sieve

INCLUDES:
- Mosi Tea Infuser

ESTIMATED DELIVERY
Dec 2019

SHIPS TO
Anywhere in the world

2,758 backers

Pledge $64 or more

EARLY TASTER DUO | 2X MOSI TEA INFUSERS

Research Questions

What differences between Kickstarter projects influence their chances of success?

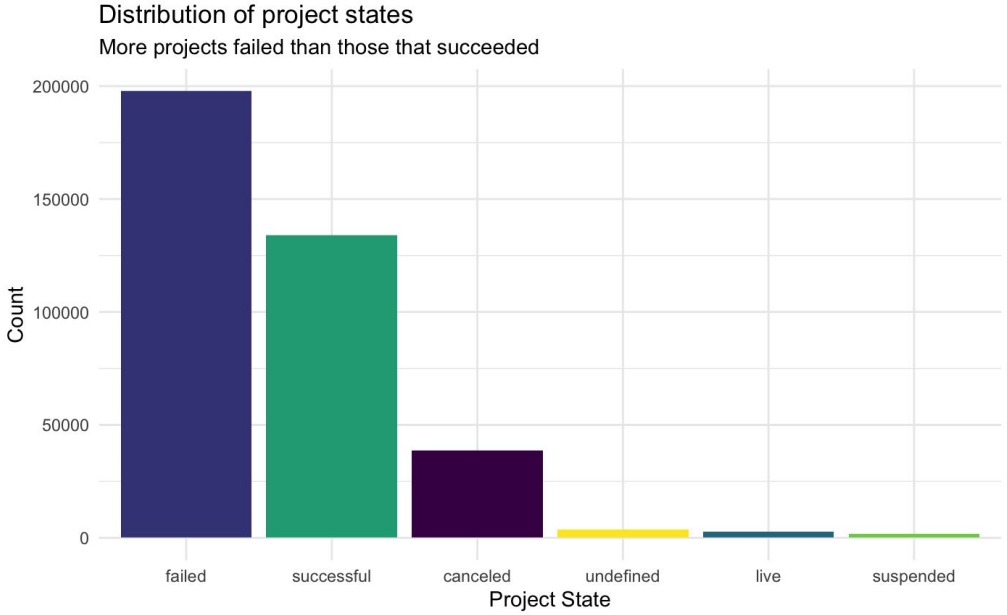
- Is the project goal (in USD) associated with its chance of success?
  - What goal amount is more likely to get funding?
- Is the project category associated with its chance of success?
  - Which category of project is more likely to succeed?

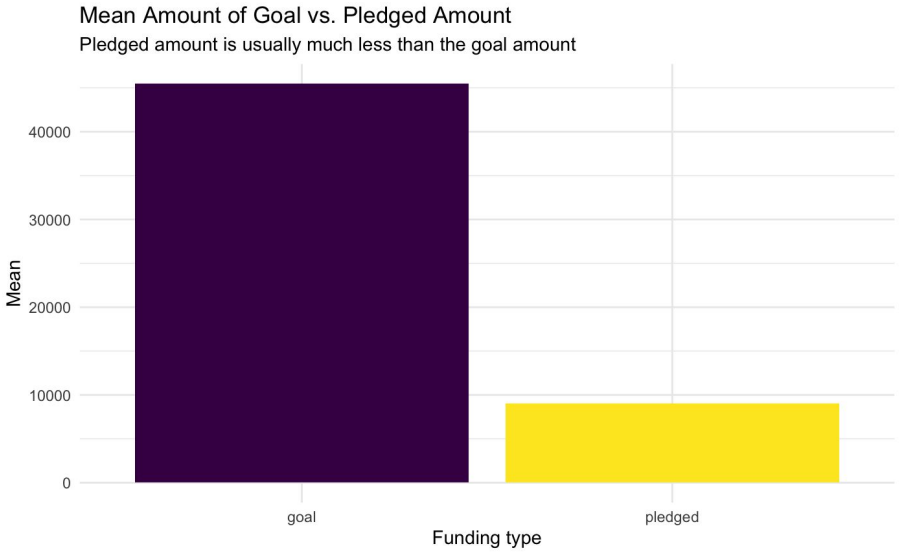
Data Description

- 378,661 observations, where each observation is one Kickstarter project.

- Relevant Variables:
  - state
  - main_category
  - usd_goal_real
  - usd_pledged_real
  - backers
  - usd_goal_real_tier (mutated)
  - success_state (mutated)

EDA


Distribution of project states
More projects failed than those that succeeded
200000
150000
100000
50000
0
Count
failed
successful
canceled
undefined
live
suspended
Project State


Mean Amount of Goal vs. Pledged Amount
Pledged amount is usually much less than the goal amount
40000
30000
20000
10000
0
Mean
goal
pledged
Funding type

Methods

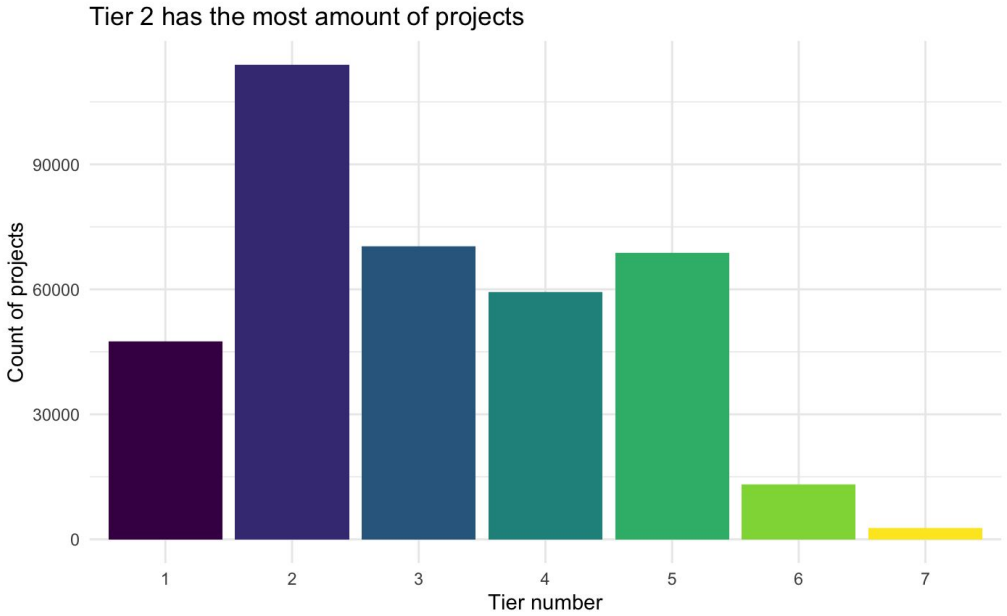
1. Used a χ^2 **test** to determine **independence** of variables such as project goal and success.

2. Used two **Logistic Regression models** to show differences in success between Tiers and differences in success between project categories

Project funding goal and success

We made Tiers to categorize the originally continuous data into discrete buckets as follows:

- Tier 1 < $1,000 (USD)
- Tier 2 ≥ $1,000 and < $5,000
- Tier 3 ≥ $5,000 and < $10,000
- Tier 4 ≥ $10,000 and < $20,000
- Tier 5 ≥ $20,000 and < $100,000
- Tier 6 ≥ $100,000 and < $500,000
- Tier 7 ≥ $500,000



Project funding goal and success

At the α = 0.05 level:

- H_0: Project tiers and success have no relationship
- H_1: Project tiers and success have no relationship

Using a CLT-based approach:

- χ^2 = 19,624, 6 df, p-value < 2.2e-16
- Enough evidence to reject the null that project tiers and success are unrelated

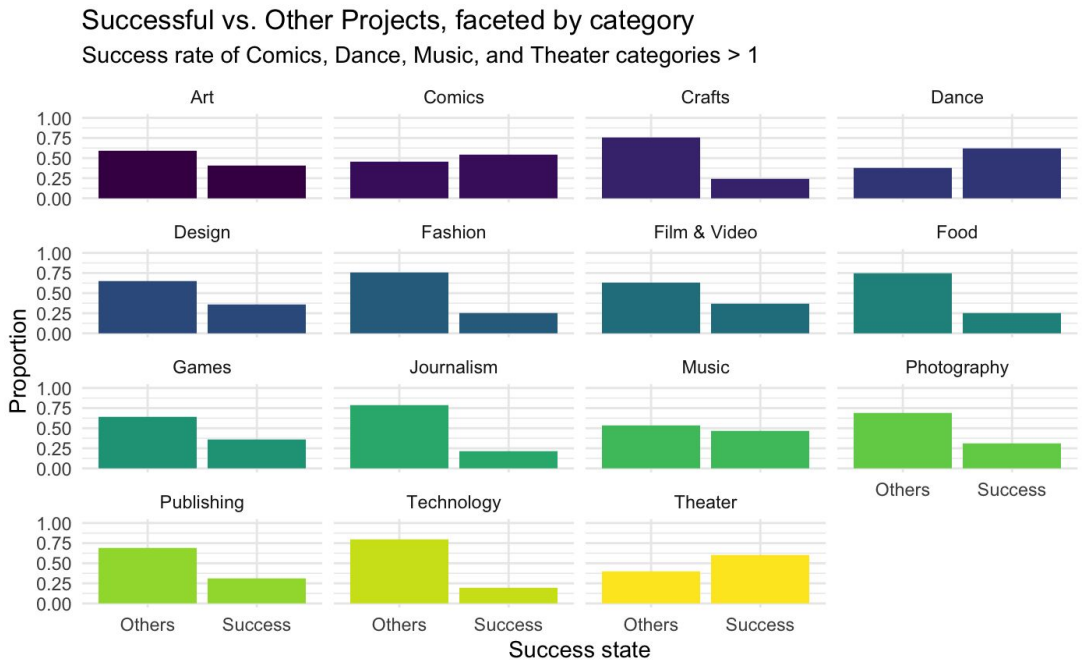
Funding tier success predictor

- Probability of tier success based on logistic regression model:
  - *Tier 1 = 0.51
  - Tier 2 = 0.44
  - Tier 3 = 0.36
  - Tier 4 = 0.32
  - Tier 5 = 0.22
  - Tier 6 = 0.09
  - Tier 7 = 0.03

- Tier success probability decreases with increasing tier number

Highlighted tiers have success probability > 0.50
*Tier 1 as the reference level

Project category and success



At the α = 0.05 level:

- H_0: Project tiers and success have no relationship
- H_1: Project tiers and success have no relationship

Using a CLT-based approach:

- χ^2 = 16,137, 14 df, p-value < 2.2e-16
- Enough evidence to reject the null that project category and success are unrelated

Category success predictor

- Used a logistic regression model to create a function to calculate probabilities of categorical success:
  - *Technology = 0.21
  - Art = 0.41
  - **Comics = 0.54**
  - Crafts = 0.24
  - **Dance = 0.62**
  - Design = 0.35
  - Fashion = 0.24
  - Film & Video = 0.37
  - Food = 0.25
  - Games = 0.36
  - Journalism = 0.21
  - Music = 0.47
  - Photography = 0.31
  - Publishing = 0.31
  - **Theater = 0.60**

Highlighted categories have a success probabilities > 0.50
*Technology used as the reference level

Discussion

Summary of key findings from our research:

- There is enough evidence to suggest that a project's Tier is related to its success rate
    - Tier 1 projects have the highest probability of success
- There is enough evidence to suggest that a project's Tier is related to its success rate
    - Dance was the category most likely to be funded

Discussion

Critiques of our method:

- Logistic regression models assume linearity in the log-odds, which we did not screen for before implementing the model
- “The chi-square test is very sensitive to sample size. With a large enough sample, even trivial relationships can appear to be statistically significant” (University of Utah, 2020)
- The chi-square test can only tell us whether two variables are related to one another. It does not necessarily imply that one variable has any causal effect on the other (University of Utah, 2020)

Source: https://soc.utah.edu/sociology3112/chi-square.php

Discussion

What we would do differently:

- We have neither looked into how long each project was on Kickstarter for funding, nor what time of the year donations may spike

What we would do next:

- Explore whether time of year affects success of projects. (People may be more charitable during the holiday period, and that may lead to more projects being funded)
- Analyze how different projects used marketing campaigns to influence their funding success