Success and Failure of Kickstarter Projects

Jess Tillis

Link to code on Github

Background on Dataset

- 379,000 projects launched between 2009-2018
- Sample of included features:
  - Name of project
  - Category
  - Date project launched
  - Deadline for project
  - Goal amount (in desired currency and USD)
  - Country of project
- Outcome of interest: whether or not the project gets fully funded ("successful") or not ("failed")

Research Questions

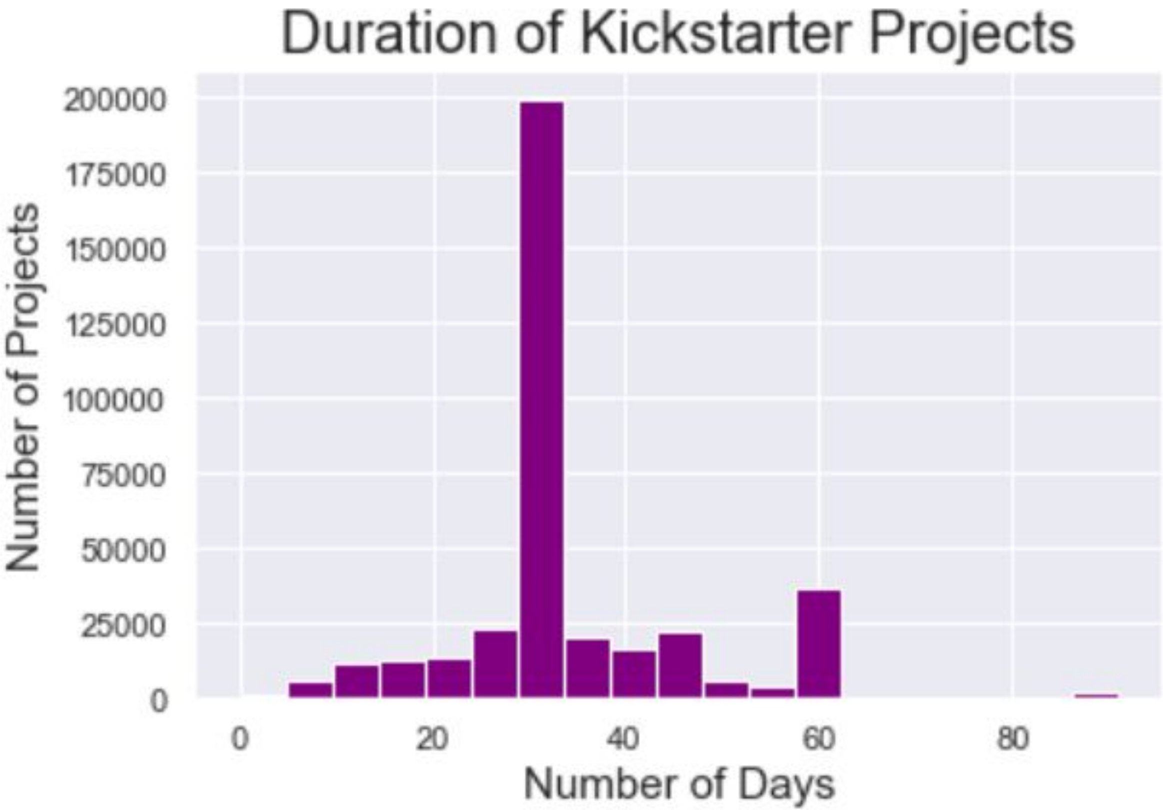
How accurately can one predict the end status of a Kickstarter project?

Which features most explain whether or not a Kickstarter project will be successfully funded?

Features I Engineered

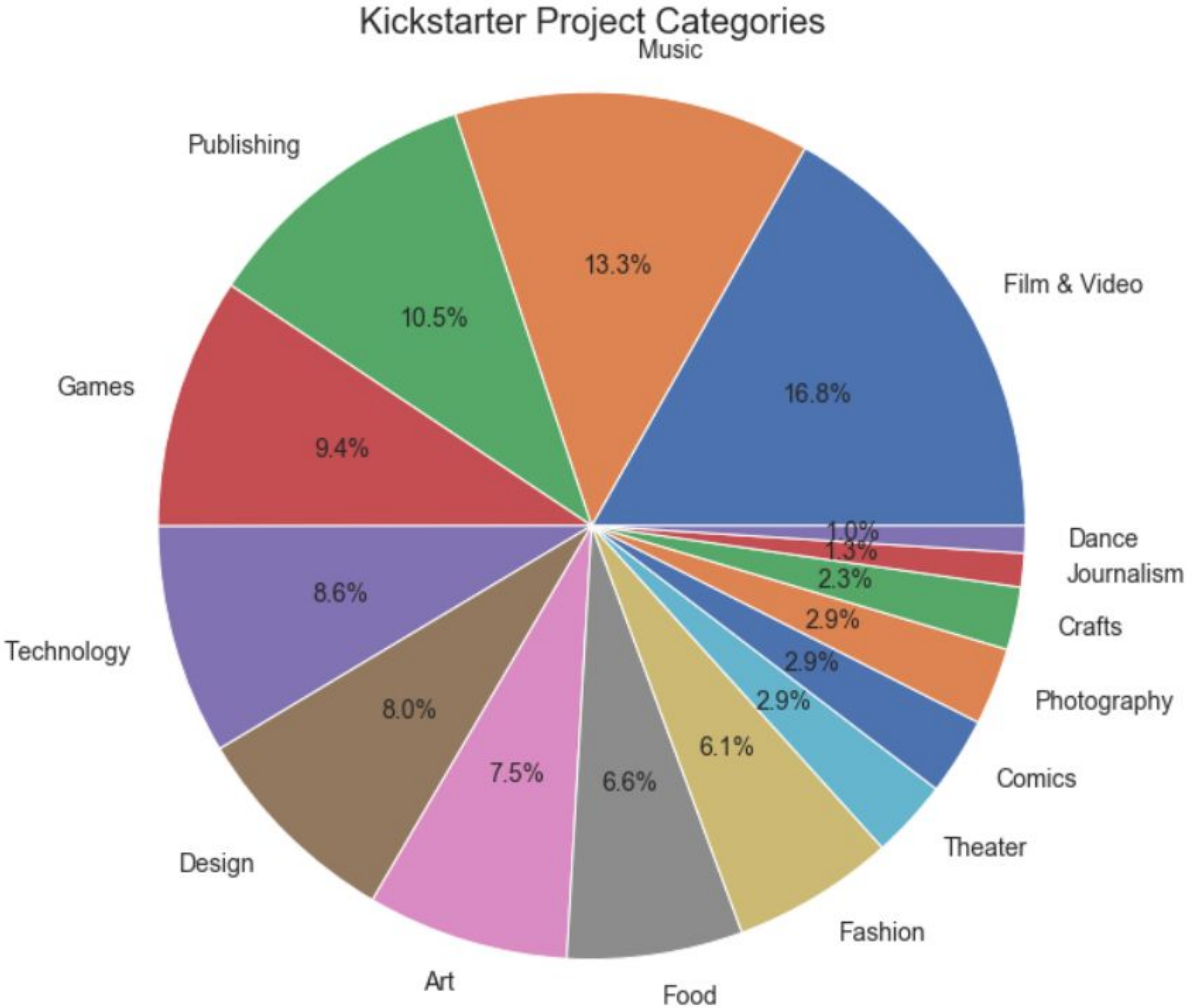
- Launch month of the project
- Duration of the project
- Created dummies for project category, launch month, and project country

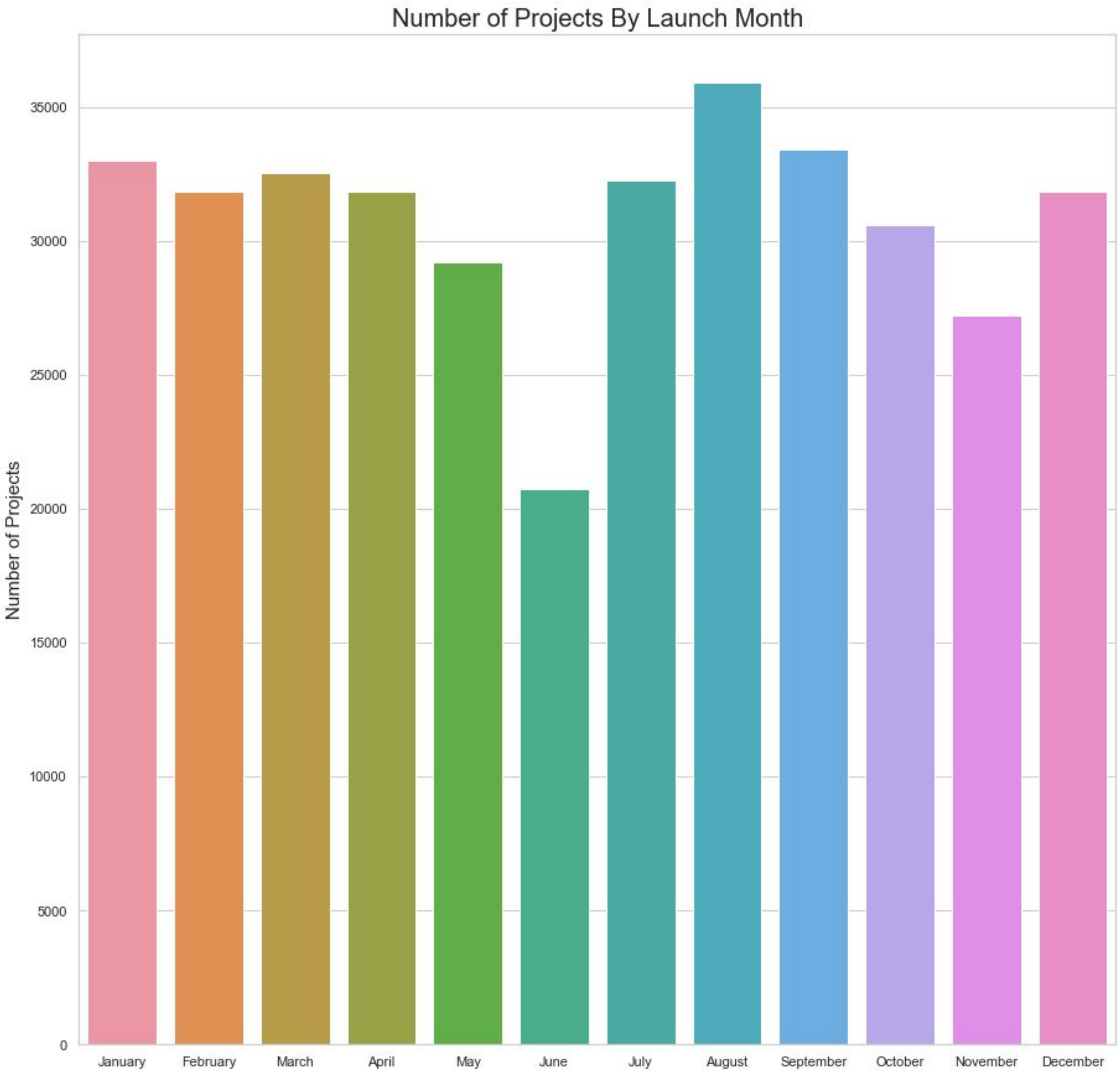
Data Analysis and Visualization



All but three of the projects in the dataset lasted 100 days or fewer. The binsize here is 5 days. The spikes in the data are at the 1-month and 2-month marks, the most popular lengths of time for Kickstarter projects

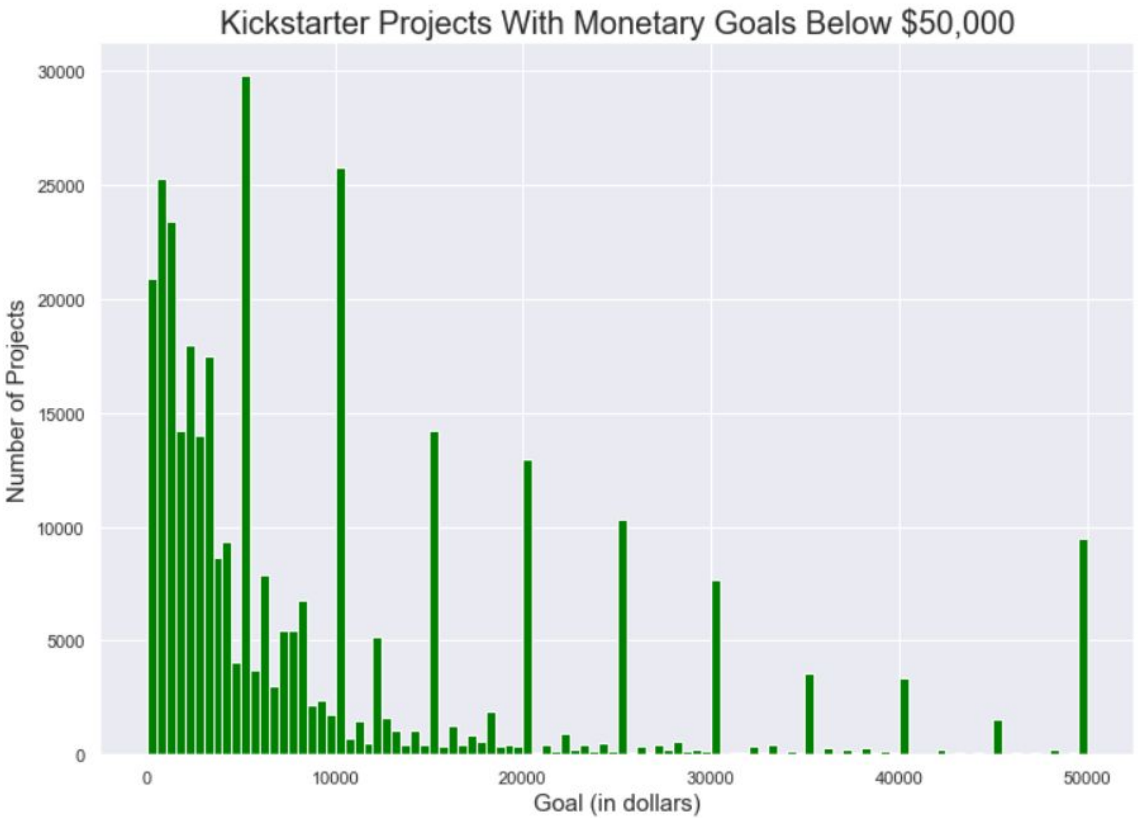
There are 15 different Kickstarter project categories. The most popular is film & video, and the least popular are dance and journalism.





This is the distribution of projects based on launch month. The most popular months to launch Kickstarter projects is late summer, and the least common month to launch is June.


Monetary Goals For All Kickstarter Projects
Number of Projects
140000
120000
100000
80000
60000
40000
20000
0
0
50000
100000
150000
200000
250000
300000
Goal (in dollars)



Ninety percent of projects launched had a monetary goal of $50,000 or less. The binsize here is $500. Most of the projects are concentrated under $10,000. There are spikes in the data at the 10s of thousands.

Selecting a Model

Models Considered

Model	Predictive Power on Test Data
Naive Bayes	65.19%
k-NN	63.81%
Logistic Regression	63.96%
Lasso Regression	66.99%
Ridge Regression	63.96%
Decision Tree	63.96%
Random Forest	64.31%

Landing on Lasso Regression

- Chosen for both the highest predictive power (67%) and potential for explanatory power
- By looking further into the coefficients, it was clear that the following features played the biggest role in predicting whether or not a Kickstarter project is successful:
  - Category of project
  - Country of project

(Slightly) Improving Random Forest

After using lasso regression to eliminate the less significant features, I included only the relevant features in the input data and re-ran Random Forest.

65.12%

Practical Use for Model

- Kickstarter can increase the likelihood that projects launched on their website will be successful
- Entrepreneurs and creatives who will launch a project on Kickstarter can improve the likelihood that their project is successful

Practical Use for Model - Finding Project Success

Feature	Percent Influence on Project Success
Category - Dance	104%
Category - Theater	102%
Category - Comics	75%
Category - Music	58%
Country/Territory - Hong Kong	38%

Practical Use for Model - Avoiding Project Failure

Feature	Percent Influence on Project Failure
Country - Italy	86%
Country - Mexico	64%
Country - Austria	63%
Category - Crafts	60%
Category - Journalism	58%

Shortcomings of the Model

- Predictive power (67%) could be higher
- Needs further analysis into how each of the most significant features impacts the outcome variable