

KickStarter Dataset

Project Overview: KickStarter is a crowdfunding platform for startups to raise funds. We are interesting in exploring this dataset to see what kinds of projects can possibly get more funding.

URL: <https://www.kaggle.com/kemical/kickstarter-projects>

Features: ID , name , category , main_category , currency , deadline , goal , launched , pledged , state , backers , country , usd pledged, usd_pledged_real, usd_goal_real

Questions we would like to explore:

1. Which industries are more likely to attract investors? Within the top industries, which specific domains are more likely to attract investors?
 - a. How many industries only have a small amount of subdomain that make up for the lion's share of success?
2. Which features are strongly correlated to the funding statuses?
3. What's an ideal funding timeline? If we increase the funding timeline, can we attract more fundings?
 - a. Duration between deadline and launch
4. What's the idea goal for each category? Is there a goal that works best for all categories?
5. What's the percentage of project that failed of all projects?
6. Which matters more? A lot of backers with small pledged amount or fewer backers but with bigger pledged amount?
7. What are the top successful, failed, canceled, live categories?
8. Why do projects get cancelled?
9. What's the best month to launch a project?
10. Is there a threshold number of backers beyond which projects tend to be successful?
11. Is there a correlation with currency or countries?
12. Year to year comparison
 - a. What industries are trending upwards / downwards across the years?
 - b. General comparisons on success rates, amounts request / goals reached

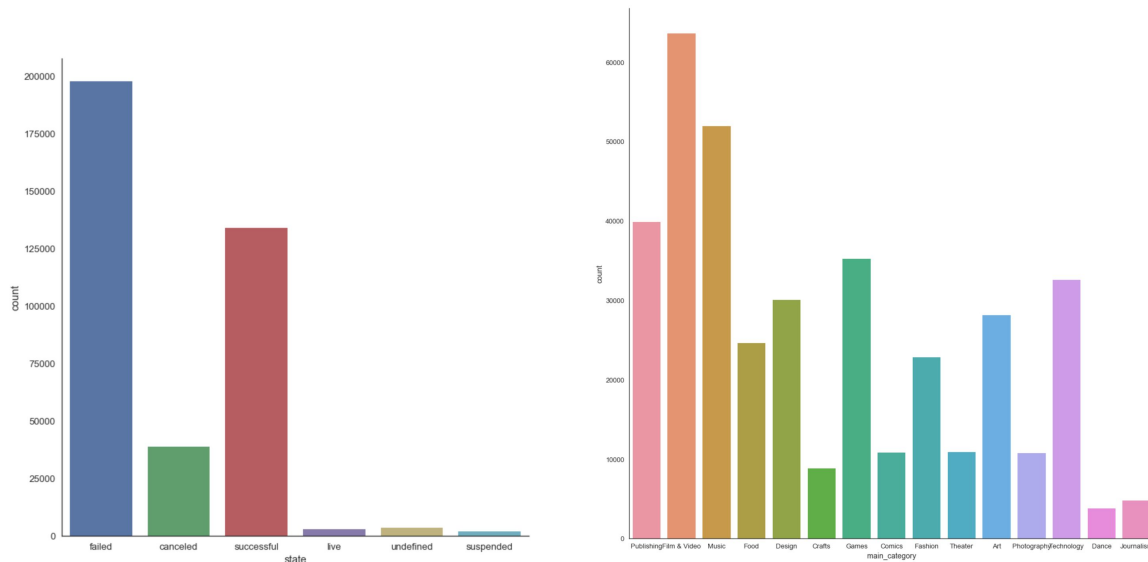
Initial Discovery

(Brief summary of numerical data values)

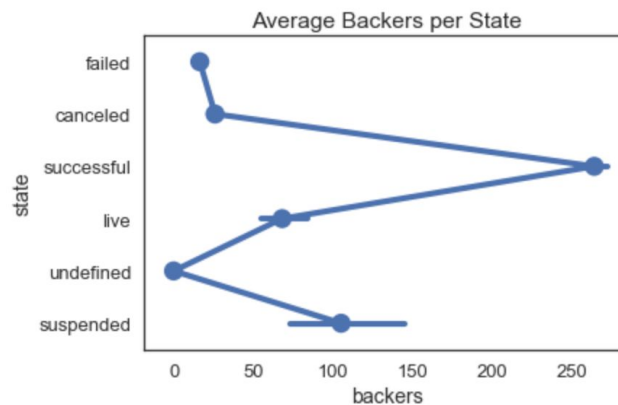
	ID	goal	pledged	backers	usd pledged	usd_pledged_real	usd_goal_real
count	378661.00	378661.00	378661.00	378661.00	374864.00	378661.00	378661.00
mean	1074731191.99	49080.79	9682.98	105.62	7036.73	9058.92	45454.40
std	619086204.32	1183391.26	95636.01	907.19	78639.75	90973.34	1152950.06
min	5971.00	0.01	0.00	0.00	0.00	0.00	0.01
25%	538263516.00	2000.00	30.00	2.00	16.98	31.00	2000.00
50%	1075275634.00	5200.00	620.00	12.00	394.72	624.33	5500.00
75%	1610148624.00	16000.00	4076.00	56.00	3034.09	4050.00	15500.00
max	2147476221.00	100000000.00	20338986.27	219382.00	20338986.27	20338986.27	166361390.71

Evaluation of Kickstarter project states & main categories

(Data shows most projects fail, with most being in the film & video category and music second)



Graph below shows the impact the amount of backers may have on the success of a project



Plan of Action:

1. Data Preparation: We will explore the data to exclude records with missing values, exceptions, and outliers. We will separate the data into “successful”, “canceled”, and “failed” subsets and examine each variable. We can normalize the data if necessary.
2. Variable Analysis and Summary: We will examine the features to determine which have more influence on the fundraising outcome with feature engineering and feature selection
3. Hypothesis and Questions: We will utilize the graphs and analysis to address the questions raised earlier in this proposal.
4. Compare with secondary data set: We will use the data set from this project as a secondary data set to explore more features.

https://data.world/rdowns26/kickstarter-campaigns/workspace/file?filename=kickstarter_data_with_features.csv