Research Paper 1.8

Nile Kealey

Chris “Topher” Rutherford

Stephen Ramirez

## **Introduction**

Kickstarter.com is a website which provides an easy way for grassroots projects to receive funding. A project could be a game, a website, an app, a film, a piece of art, etc.

An example of how it works: Suppose I wanted 10,000 dollars to make a movie, but I didn’t have 10,000 dollars. I could create a Kickstarter campaign online, and market it so that people would donate to me. If I reached my goal, I could keep the money and use it for my movie. If not, the money would return to those who donated.

The primary goal of this project is to discover what factors influence the success of a Kickstarter project, such as: the time of year that the campaign starts, how long campaign is open for funding, if certain categories garner more attention/funding, etc. What category has the highest percent of successful campaigns? Which category has the most generous donors?

For the sake of having clear hypotheses, our hypotheses are:

-All times of year have the same success.

-All categories have the same rate of success.

-All categories have the same money-to-donor ratio.

This project won’t necessarily be limited to these questions. More questions, big and small may appear, which may be squashed off the bat, or they may be struggled with for hours, or they may never be satisfied.

## **Plan of Action**

o ***What data is needed to address the research questions and/or test the hypotheses?***

In our dataset, the most useful columns I expect will be the outcomes, the start and end dates, the categories, the goal and pledges in USD, and the number of backers.

o ***How will you use this data to answer the research questions and/or test the hypotheses?***

Questions to answer

*Which time of year is most successful?*

This can be answered by looking at the times of year for the projects. One way to see

the distribution of success is by finding the success percentages for each month. Plotting that might be good.

*Do certain categories get more success/funding?*

To start, we’d look at the categories and find the success percentages for each category. We’d also look at the average amount of money donated per project in each category. Plotting this might also be good!

*Which category has the most generous donors?*

For each category, we’ll find the ratio of the amount of donors to the amount of

money donated. This will let us see which category has the highest rollers.

o ***What needs to be done to prepare the data for this analysis?***

Cleanup, for one. Checking for any ridiculous outliers will be useful.

Creating subsets of the main dataframe which only includes data from one category or

one month, for example, could be useful.

Also, depending on what we end up doing later on, adding new columns for more

precise information, such as the total length of the campaign, and then the length in days (for integer manipulation) will make analyzing the data and finding trends easier.

## 

## **Dataset Description**

The primary dataset being used is a record of approximately 370,000 kickstarter campaigns up through January 2, 2018, which was obtained from Kaggle. The column contains the following info for each campaign: ID, name, category, main category, currency, deadline, launch date, goal, how much was pledged, outcome, how much was pledged in USD (converted from local currency if necessary), and the goal in USD.

Kaggle.com, the source of our dataset, is a popular data science website. It offers a variety of useful resources for data scientists and it is a reputable source for datasets.

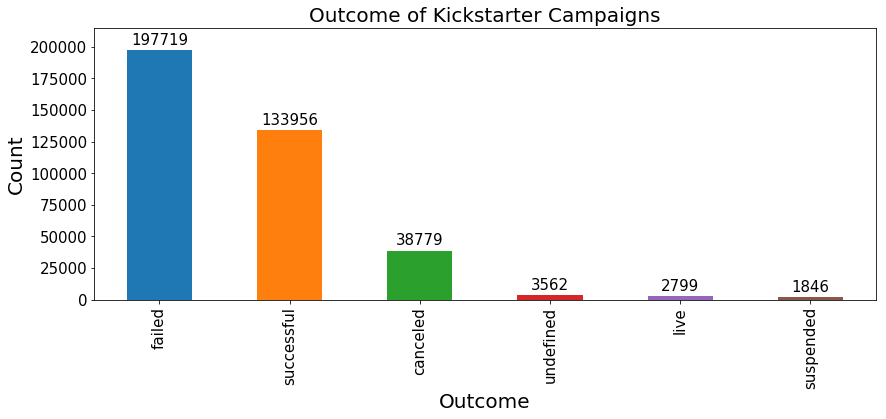
Here is the link to the source: <https://www.kaggle.com/kemical/kickstarter-projects>

Note that there are two datasets in this source. The one we’re using is the more up-to-date one: “ks-projects-201801.csv”

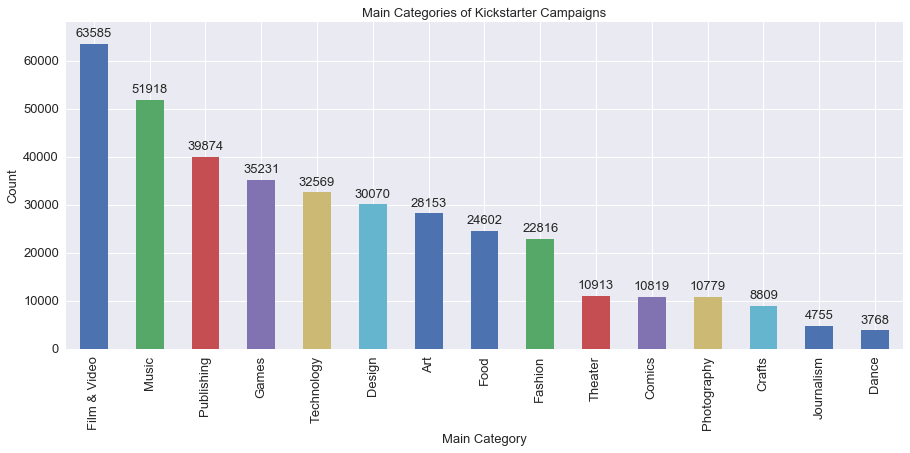
### **Data Processing**

Initial preprocessing includes checking for data that is either missing, or has other issues with it. There are only four campaigns with missing names out of the entire dataset, but there are also about 3,500 campaigns with an “undefined” outcome, almost all of which also have NaNs for the usd\_pledged column, and nonsense (labeled as N,0”) in the country column. The pledged column has values but usd\_pledged does not.

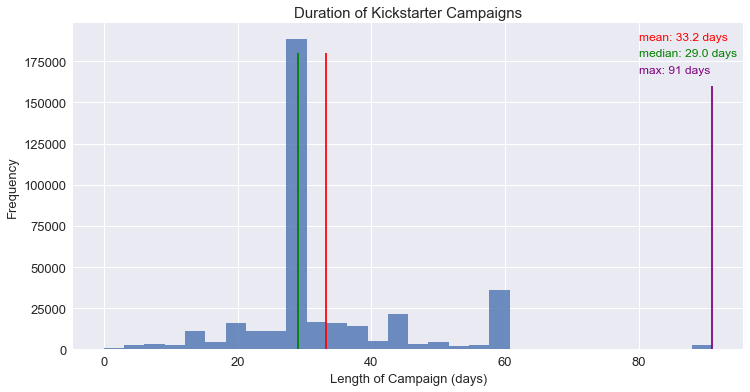
“Outcome of Kickstarter Campaigns” shows us the counts of the outcomes for our full data set, and it gives us an overview of the shape of the data.The majority of Kickstarter campaigns have been unsuccessful, but a solid chunk are successful anyway.



“Main Categories of Kickstarter Campaigns” is another graph made to give more sense of the distribution of campaign categories. Some categories have many more projects than others. Film & Video is the most popular of all ! I’ll have a lot of competition if I am ever going to make Citizen Kane 2 !



“Duration of Kickstarter Campaigns” provides a histogram of how long each campaign lasted in days. People seem to prefer either sticking to the default 30 days, or taking advantage of the maximum 60 day limit. The 91 day campaign was created before this limit was put in place by Kickstarter.



“Kickstarter Success Rates By Category” aids us in answering the question of whether or not certain categories are more likely to be successful. It looks like our hypothesis that they’re all the same is likely incorrect, since some categories are more than twice the success percentage of another. Theater and dance have the highest success rates. I wonder why.

## 

## 

## 

## **Next Steps**

The next steps are to keep playing around with the data and looking for more questions to ask, and following each question down until satisfied. We could dig deeper and deeper, until we can’t dig no more! Then after we feel like we played around with that data, we could move onto something new, or seek out more data, like something that we don’t already have in our dataset.

OUR INDIVIDUAL THOUGHTS:

### Chris

Thoughts: initial visualization can include bar charts to visualize the most popular categories and main categories, how long campaigns last, the outcomes of the campaigns, etc.

Questions: Are the most popular (main) categories also the most successful? Film and video has the most campaigns, but only a 37% success rate, is there a subcategory that’s drawing more attention to this category as a whole? Was there a very successful film or video that convinced people that any old movie pitch could become a success tothrough kickstarter?

Comments: possible addition of features: length of campaign (already done by subtracting launch date from deadline and extracting the day from the datetime object)

One hot encoding for outcomes of campaigns?

Most campaigns end in failure

Kickstarter seemed to add a limit of 60 days for campaigns at some point, noting that campaigns lasting longer tend to end in failure. About 4700 campaigns last longer than 60 days.

### Nile

Thoughts:It’s really hilarious to look at some of the most UN-successful kickstarters, like the ones with zero backers that demand millions of dollars! And maybe the topic is just so ridiculous that there never was a chance of success. That’s just funny to me, like how they went through the trouble of making a ridiculous kickstarter just for fun.

Some projects might have a lot of backers and a lot of money pledged, which makes sense. But what about projects with many backers but not as much money pledged? And vice versa? It could be found that some backers for different things and categories give different amounts of donations. It could be interesting. Example, A kickstarter which benefits the wealthy somehow might have higher donations per backer, since more people can afford it. It could be the other way around though, because wealthy people wouldn’t spend money on something like a kickstarter.

Questions: How many kickstarters were a joke? What categories have higher average donation sizes?

Comments: There are some very successful kickstarters for games and “fun” things. One of the most successful things ever was this card game called “Exploding Kittens” which I have actually played before. I looked through a few successful kickstarters, and it appears as though successful kickstarters are almost always using past popularity to get rolling. For example, Exploding kittens was apparently made by the people/person behind that webcomic “the oatmeal” which is already popular on its own.

Kickstarter seems like a quirky and unique enough thing that it might be best looked at on a case-by-case basis. Every project might be very different from the next, and some might have special success or unsuccess stories. There’s not a lot of time to look at every single kickstarter, but I looked at some of the most successful ones and one or two random ones, and one kickstarter is a often a wildly different idea from the next. Even with thousands of them, they seem like snowflakes.

### Stephen

Thoughts: I like Chris’s idea about initial visuals and agree. I also think we should describe what would be considered successful such as just meeting the goal or blowing past the goal in record time. Also if the campaign did surpass the goal, by what percent.

Questions: Is there a campaign category that is better suited for a certain time of year?

Comments: I wonder if there is a correlation between words used and overall success. Maybe we can run a quick analysis on title length and also words in common amongst good and bad outcomes.