Joseph Arambulo

Challenge 1 Report

Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

1. Most crowdfunding campaigns are related to the arts. Performing arts/Theater is at the top of the list, followed by Film & Video and Music which are almost even. Other art related campaign category includes Publishing, Photography, and, potentially, Journalism.
2. Overall, there’s a downward trend regarding art related crowdfunding campaign but there are distinctions. For example – when viewed altogether, the six categories mentioned above shows a downward trend in both success and failure from 2010-2020. When analysed individually, Theatre campaigns show an upward trend in successful outcomes and a slight downward trend in failed outcomes; Music campaigns’ successful outcomes have plateaued, but its failed outcomes show a significant downward trend; and both successful and failed outcomes show a downward trend for Film & Video campaigns.
3. It seems that crowdfunding campaigns related to the Arts gets more attention than others, but that does not mean that their success rate is higher than other types of crowdfunding campaigns. Glancing at graph “Crowdfunding Campaigns Outcome per Category”, Technology campaigns seem to have a higher success to failure ratio.

What are some limitations of this dataset?

* Considering that the dataset spans from 2010-2020, the sample size is small. It averages at 100 data per year; a larger dataset is better for increasing confidence in analysis.
* More information (lack of metadata) about the different columns would be great as well. For example, it would be great to know what the “staff\_pick” and “spotlight” data represents, and if it could be useful for analysis.
* Better organization of the “category&sub-category” column would make the dataset better. Several categories are too similar to each other, so much so that they could be sub-categories in a bigger category grouping.

What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

* Graphs that consider Average Donation/pledged/backers\_count would be interesting to see. We could analyse which category of campaign generates funding efficiently, i.e. if there are cases where a campaign succeeds but it had lower than average backers\_count.
* Observing which type of campaign receives more donations (Average Donation) would help entrepreneurs decide if going through crowdfunding is a viable option for obtaining funding for their product.
* Hypothetically, if the column “spotlight” pertains to how upfront or close to the first page a campaign can be found in a crowdfunding website, then it would be interesting to analyse how much a campaign getting “spotlighted” attributed to the success of a campaign. If there is a relationship, the crowdfunding company could advertise, and monetise, the fact that getting “spotlighted” could increase chances of success.

Statistical Analysis

* In this dataset’s case, the mean better summarizes the data. The median, the number in the middle of the data, of either outcome’s backer count is less meaningful since there’s a considerable difference in sample sizes (“successful” outcome’s sample size is 565 and “failed” outcome’s sample size is 364). Using mean, the average of the data (in this case, the average amount of backers), might balance out that fact.
* Regarding variability, “successful” and “failed” outcomes are somewhat even. About 85% and 88% of the data for both outcomes, respectively, falls within one standard deviation of their mean. This makes sense because “backers\_count” only measure how many people “I want to support this campaign”, it does not reveal how much people pledged towards the campaign. There might be cases where a successful campaign has a smaller than average amount of backer, but each backer pledged significant amounts of money to reach the goal.