

## Assignment 1 – Module 1 Challenge - Writeup

Steph Abegg

Due: June 27, 2024 at 11:59 pm

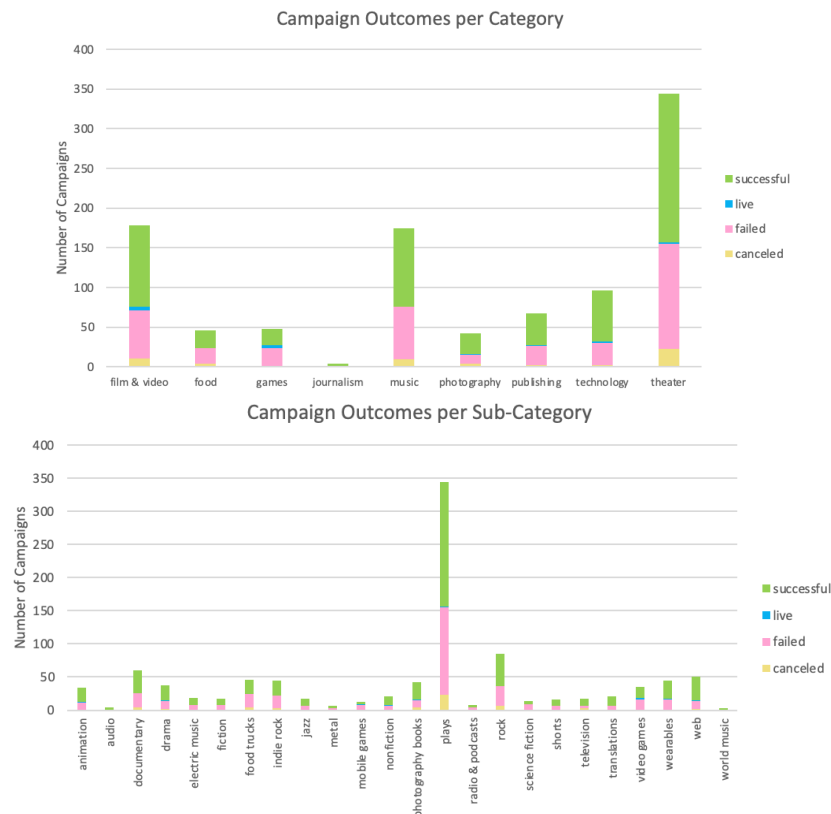
### Excel

See CrowdfundingBook-Abegg.xlsx for the required Excel work. I have pasted in some of my visuals in this document to support my conclusions.

### Outcomes per Category, Sub-Category, Launch Date, Country, Goal Amount

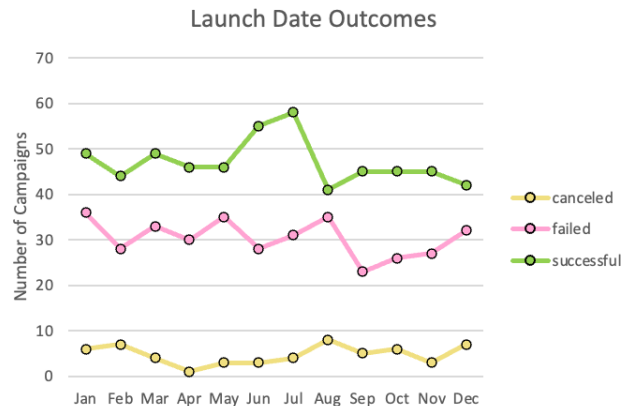
*Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?*

1. Regardless of campaign category, about a third of the campaigns fail (36.4% failed campaigns over all categories) and just over half are successful (56.5% successful campaigns over all categories). The bar charts below illustrate this conclusion.

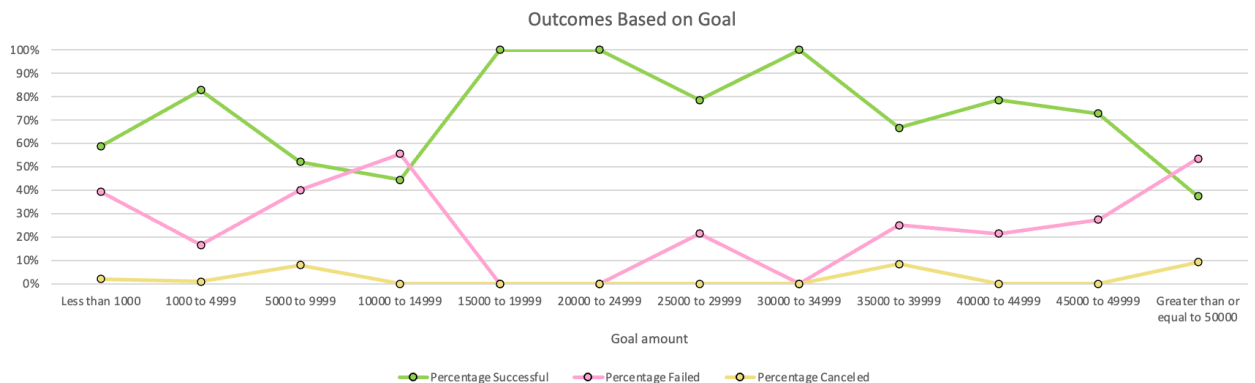


2. About a third (34.4%) of all campaigns are for theater/plays. The bar charts above illustrate this conclusion.

3. The number of successful campaigns peaks in June and July. Interestingly, the number of successful campaigns dips in August while the number of failed and canceled campaigns rises (perhaps campaigners trying to continue the June and July successes longer than funders are willing). The line graph below illustrates this conclusion.



4. Campaign success percent is highest when the goal amount is \$15,000-\$35,000. There are more failed campaigns than successful campaigns when the goal amount is between \$10,000-15,000 (suggesting the goal should have been a bit higher to increase chances of success) or above \$50,000 (suggesting the goal was too high to be reached). The line graph below illustrates this conclusion.



*What are some limitations of this dataset?*

Some limitations of the dataset are:

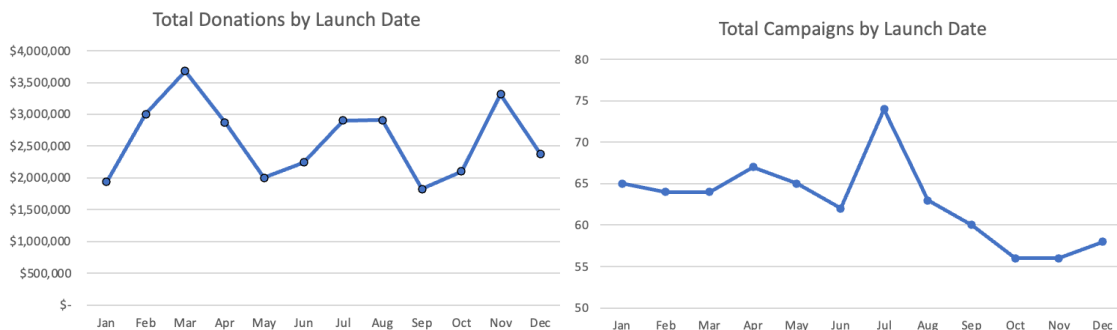
- A large percentage (34.4%) of the campaigns are theater. This biases the results to be fit the trends for theater. To be unbiased, there should be an equal number in each category.
- A large percentage (76.3%) of the campaigns are in the US. This biases the results to be fit the trends for US. To be unbiased, there should be an equal number in each country.

- The donation amounts are in different currencies (i.e. AUD, CAD, CHF, DKK, EUR, GBP, USD). To be directly comparable, these should all be converted to one currency using the currency conversion appropriate to the date of the donation.
- The data is from 2010-2020. Due to changing campaign dynamics, economy, social trends, etc., the results of this analysis should probably not be projected too far outside of this date range.

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

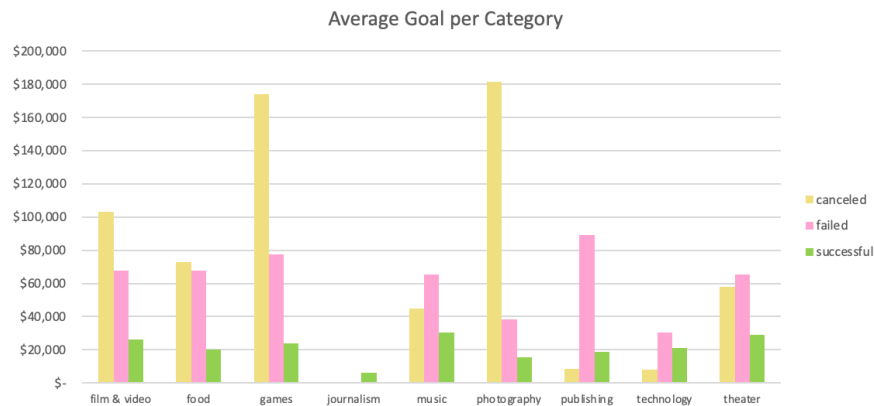
There are several additional pivot tables and corresponding graphs that we could create. I provide some examples below.

1. Total donation by month line graph. Since the donations are in different currencies, either the currencies all need to be converted to a single currency using the currency conversion appropriate to the date of the donation, or a separate graph should be made for each country. Since most of the data is in USD, the line graph below shows the total donations by launch date in the US only. It seems that donations follow somewhat of a sinusoidal pattern with a 4 month period, peaking around March (before taxes!), July (summer optimism), and November (pre-holiday cheer), even though the total number of campaigns does not (graph on the right) This information could be useful to campaigners, as running campaigns around March, July, and November when donors seem more generous.

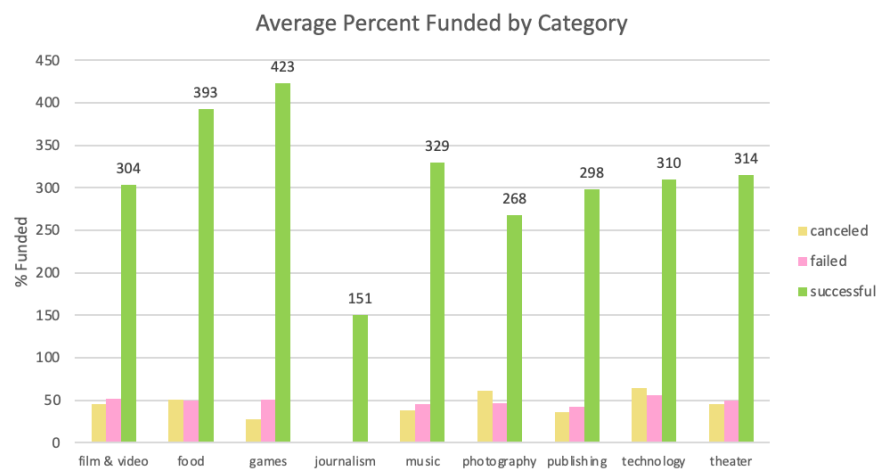


2. Average goal by category or subcategory bar chart, separated by outcome. This can be used to show which goals are too high and which categories are setting too high of goals. The bar chart below (for US only to keep the currency all in USD) shows that the average goal for the successful campaigns is lower than the average goal for failed and cancelled campaigns. Also, for

example, high average goals for the cancelled campaigns for photography and games suggest that these goals were set too high.

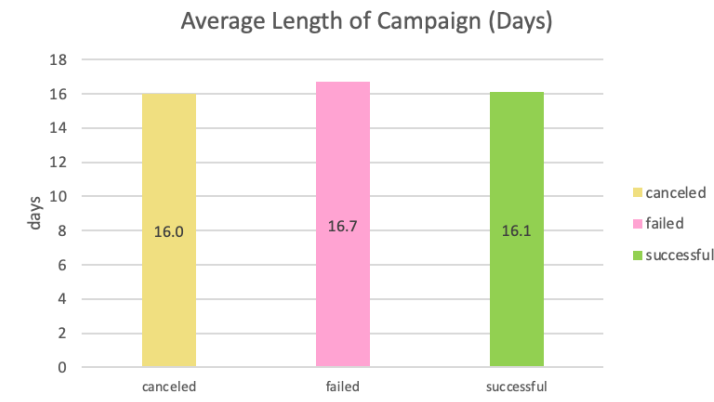


3. Average Percent funded by category or subcategory bar chart. This can be used to show which categories of campaigns are most successful in achieving funding. The bar chart below shows that the average percent funded for the successful campaigns is much higher than the average percent funded for failed and cancelled campaigns, suggesting that funding is a main reason why campaigns fail or get cancelled. Also, for example, we see that some categories like games and food are better at achieving higher percent funding on their goals than categories like journalism and photography. This could be due to different goal amounts or more successful campaigning.

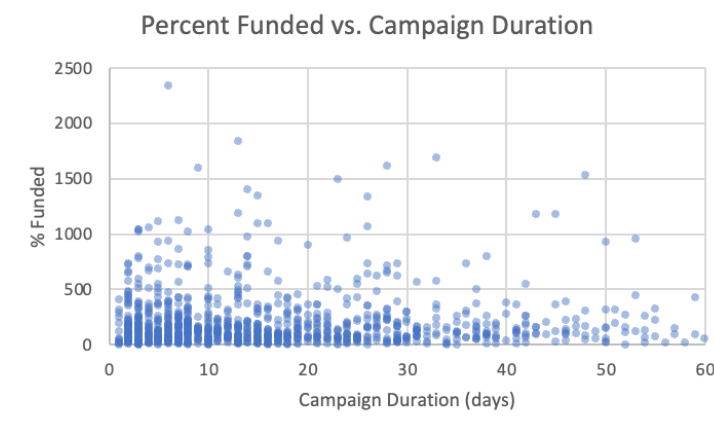


4. Average Length of campaign (Date Ended – Date Created) by outcome bar chart. This could be used to show the ideal length of a campaign leading to a successful campaign. The following bar

chart shows that the average duration of campaigns (whether successful, failed, or cancelled) is around 16 days. So duration of campaign does not seem to have too much influence on success.

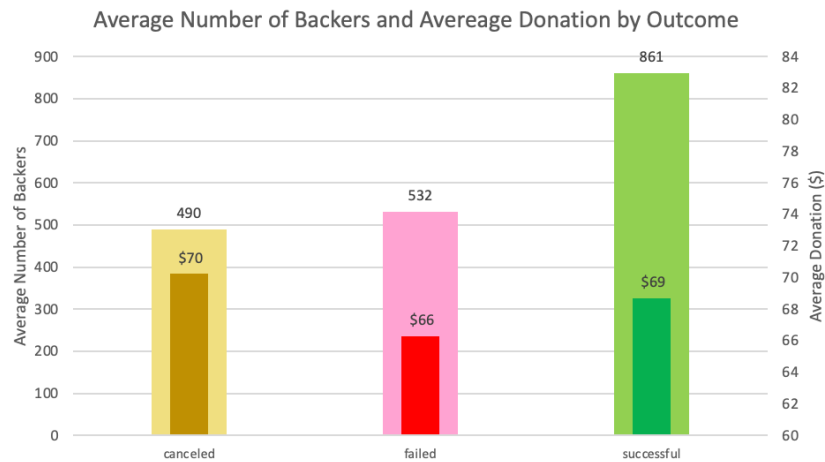


- Length of campaign (Date Ended – Date Created) by Percent funded scatterplot. The scatterplot below shows no obvious pattern between percent funding and campaign duration, yet another indicator that duration of campaign does not seem to have too much influence on success.



- Average number of backers and average donation by outcome bar chart. This bar chart (shown below) shows that successful campaigns have an average of 861 backers, compared to lower numbers for failed (532) and canceled (490) campaigns. The average donation per backer is

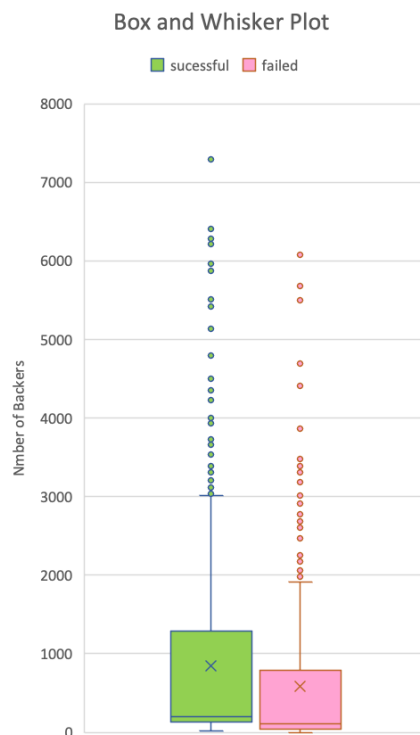
relatively consistent whether the campaign is successful, fails, or is cancelled, indicating that attracting more backers is of key importance to a successful campaign.



## Backers Count

*Use your data to determine whether the mean or the median better summarizes the data.*

The following box and whisker plot for the number of backers shows that the data is positive-, or right-skewed, meaning that the data has a larger tail. This is created by data points of high numbers of backers for some campaigns. On the box and whisker plot below, these are the outliers that lie above the  $Q3 + (1.5 \times IQR)$  Boundary.



The mean ("x" on the box and whisker plot) of positively-skewed data is greater than the median (horizontal line on the box and whisker plot). For distributions that have outliers or are skewed, the median is usually the preferred measure of central tendency because the median is more resistant to outliers than the mean.

*Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?*

There is more variation in the number of backers for successful campaigns (standard deviation 1267 backers) than unsuccessful campaigns (standard deviation 961 backers). This makes sense, as one would expect that, in general, unsuccessful campaigns are correlated with a lower number of backers (and hence lower standard deviation), whereas successful campaigns are correlated with a higher number of backers *in addition to* the lower-backer number cases where individual backers contribute large sums of funding (hence resulting in higher spread of the number of backers and a higher standard deviation).