Module 1 Challenge

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

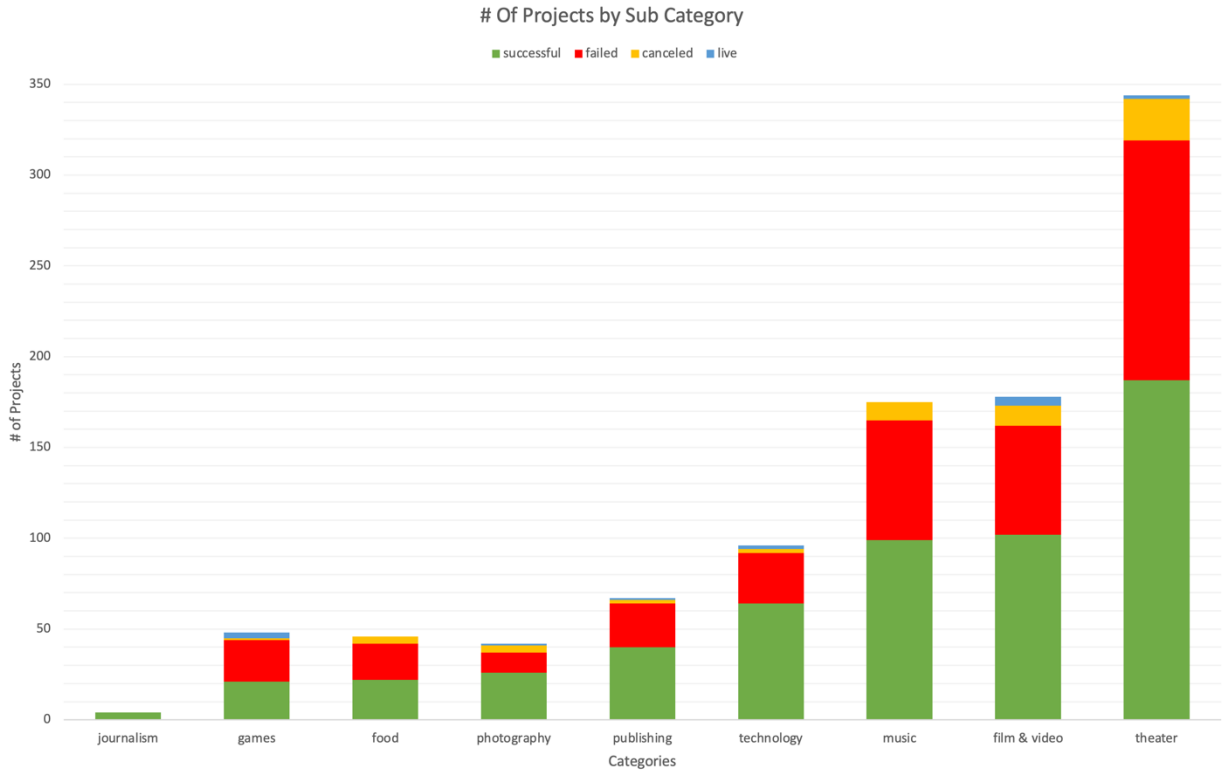


Figure 1 – # of Projects by Category

The first and most apparent conclusion drawn from examining the pivot bar charts of the # of Projects by Category data is that the current dataset shows us a significantly higher number of projects and successes in the sectors of creative media and the arts, namely in Film & Video, Music, and Theatre.

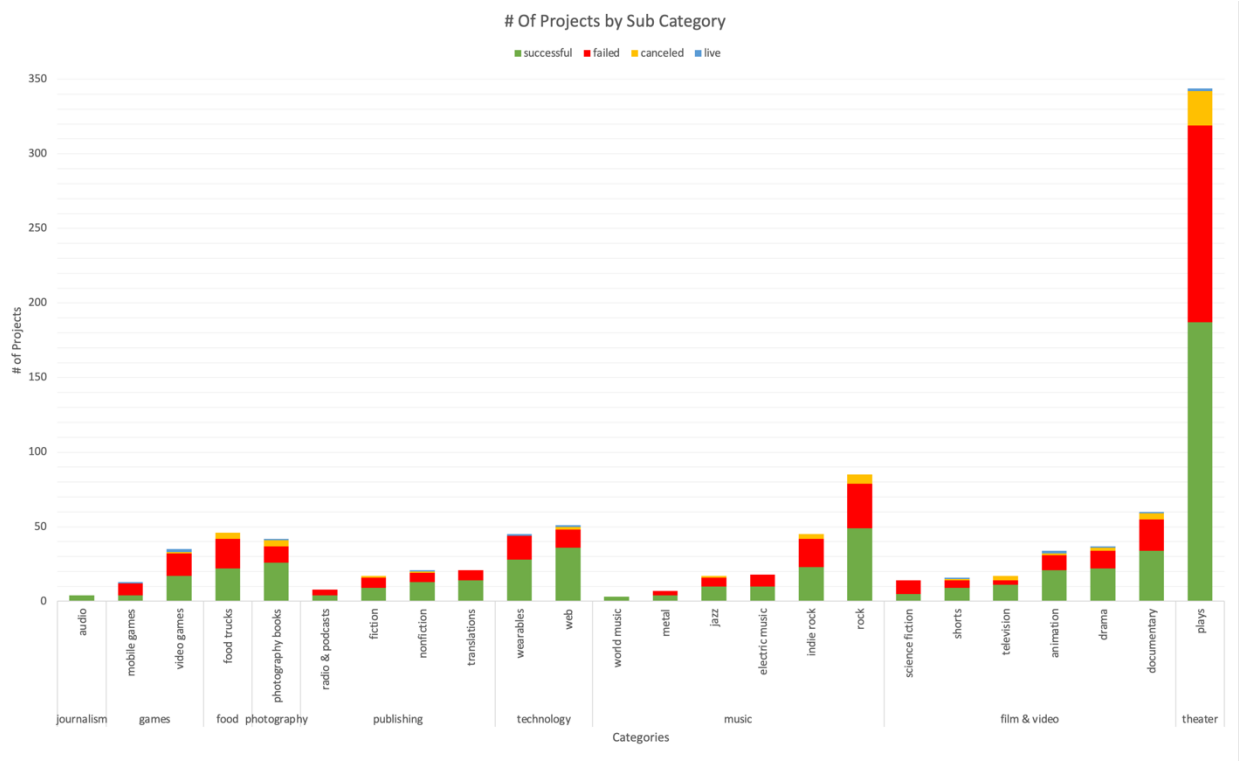
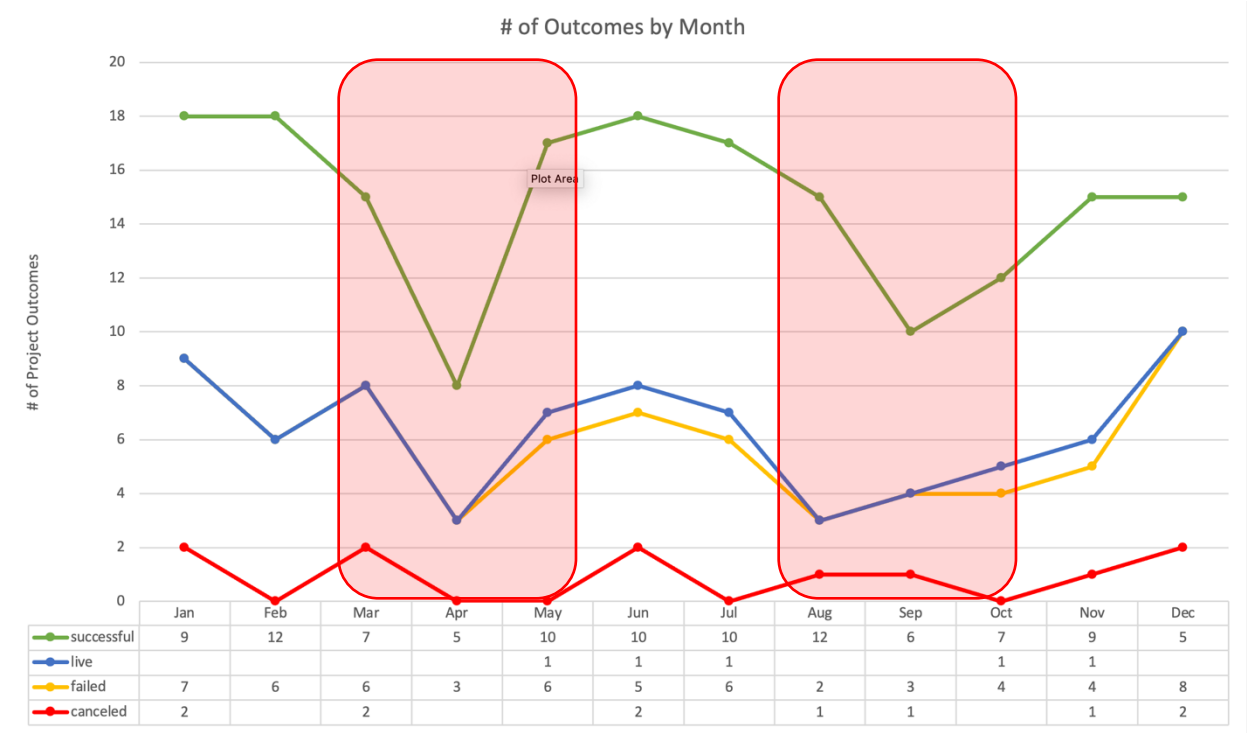


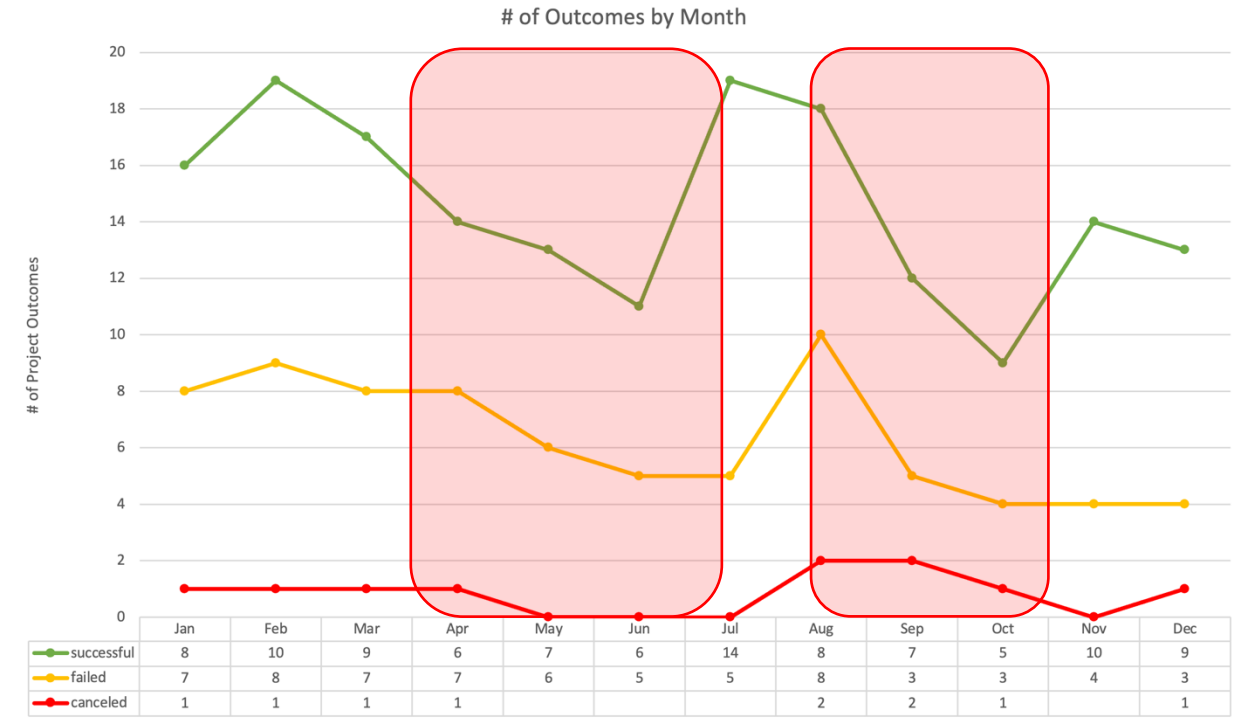
Figure 2 - # of Projects by Sub-Category

Additionally, the pivot bar charts of the # of Projects by Sub-Category data also indicates that although Film & Video and Music offer a variety of subcategories, with 5-6 different options thus collectively adding up to their large number of projects, the topics of Theatre is currently limited to only plays. Despite this, Its noteworthy that the Theatre sector on this crowd funding site still seems to consistently have a higher number of both successful and failed projects which suggests that consumers of the crowdfunding platform are already primarily/ largely seeking Theatre-based content, as it appears to be the most popular category based on the data by a large margin.



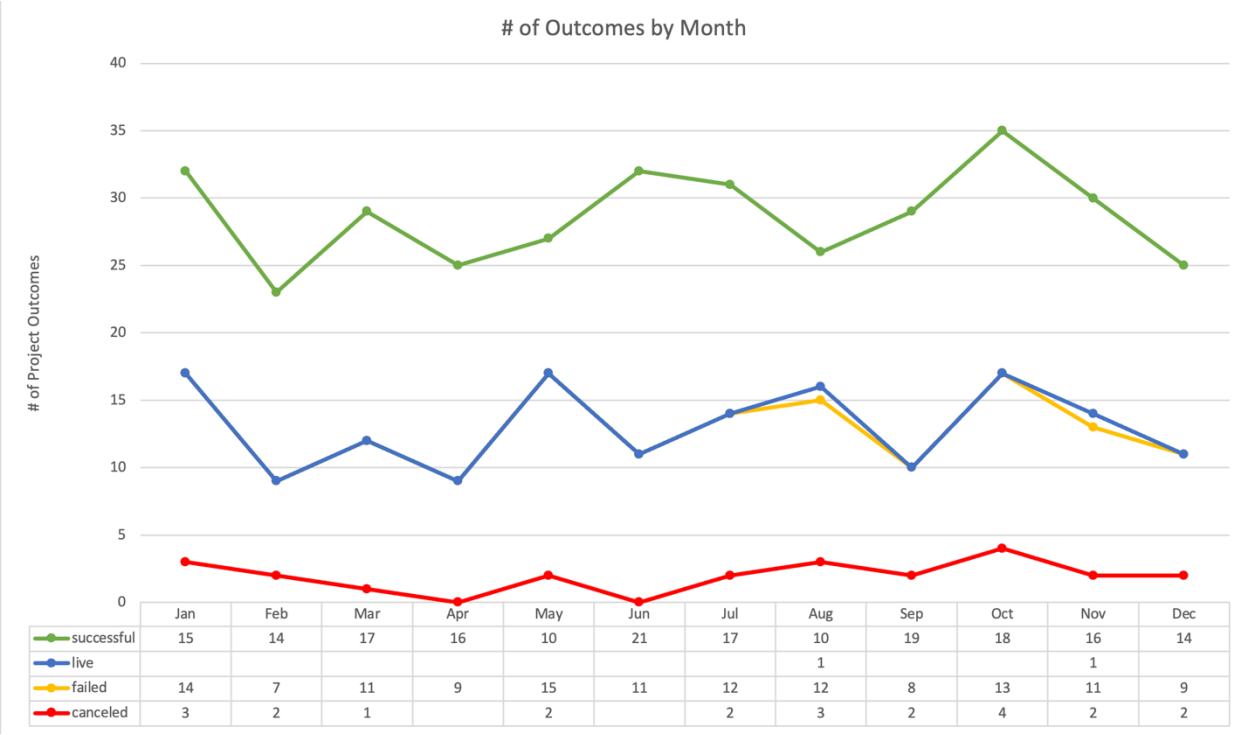
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
successful	9	12	7	5	10	10	10	12	6	7	9	5
live					1	1	1			1	1	
failed	7	6	6	3	6	5	6	2	3	4	4	8
canceled	2		2			2		1	1		1	2

Figure 3 - Film & Video Sub-Category on the # of Outcomes by Month



	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
successful	8	10	9	6	7	6	14	8	7	5	10	9
failed	7	8	7	7	6	5	5	8	3	3	4	3
canceled	1	1	1	1				2	2	1		1

Figure 4 – Music Sub-Category on the # of Outcomes by Month



	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
successful	15	14	17	16	10	21	17	10	19	18	16	14
live								1			1	
failed	14	7	11	9	15	11	12	12	8	13	11	9
canceled	3	2	1		2		2	3	2	4	2	2

Figure 5 - Theater Sub-Category on the # of Outcomes by Month

Interestingly, when examining the Pivot Line Chart of the # of Outcomes by Month data, we notice that while Film & Video and Music experience slight variability in success rates, seemingly influenced by seasonal factors with dips in early to mid-summer and fall periods. The theatre sector seems to maintain a remarkable balance and consistency in both successes and failures without much of a disparity or difference between different months. This points to Theater being a comparatively more secure category, with its projects showing greater resilience to seasonal effects and popular culture.

What are some limitations of this dataset?

The most significant limitation of this dataset is the lack of information on individual donations and their repeatability. Specifically, the adding of data about the mode or median amounts donated would allow for greater insight on the donation habits for each Project. Relying solely on the average values of the donations can be very misleading, as a few large contributions would skew the data and give an inaccurate story. Without insight into the median or mode donation amounts, we cannot determine whether the majority of project funding comes from many small donations or is disproportionately influenced by a few large contributions. This gap in the data makes it difficult to assess and see any potentially crucial patterns which would allow for actionable decisions as per the expected stakeholders wishes which based on the scenario would be to find out how to maximize the chance of your own project succeeding in reaching its initial Goal.

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

Considering this challenge situation, the initial charts do not provide any highly useful or actionable insight for the stakeholders, as they do not address the core questions at hand. In this situation, the question being asked is why certain projects are more prone to succeed or fail and if there are any markers or trends we can use to predict and replicate those successes. Considering this, the primary important metric to look at would be probabilities or ratios of Success by Category and even Month. To look at this Id create a Pivot Chart looking at the Outcome Ratios by Sub Category and display the probabilities of success, failure, cancellation, or live status. This would allow stakeholders to visualize and compare the likelihoods of the various outcomes on a Simple 100% Stacked Chart (find below). Along with a simple Baseline added to the Average Success Rate, this chart would provide a clear visualization of which categories are more likely to succeed than others, which is in this case are Audio, World Music, Translations, and Web.

Count of outcome	Column Labels				
Row Labels	successful	failed	canceled	live	Grand Total
journalism	**100.00%**	**0.00%**	**0.00%**	**0.00%**	**100.00%**
audio	100.00%	0.00%	0.00%	0.00%	100.00%
games	**43.75%**	**47.92%**	**2.08%**	**6.25%**	**100.00%**
mobile games	30.77%	61.54%	0.00%	7.69%	100.00%
video games	48.57%	42.86%	2.86%	5.71%	100.00%
food	**47.83%**	**43.48%**	**8.70%**	**0.00%**	**100.00%**
food trucks	47.83%	43.48%	8.70%	0.00%	100.00%
photography	**61.90%**	**26.19%**	**9.52%**	**2.38%**	**100.00%**
photography books	61.90%	26.19%	9.52%	2.38%	100.00%
publishing	**59.70%**	**35.82%**	**2.99%**	**1.49%**	**100.00%**
radio & podcasts	50.00%	50.00%	0.00%	0.00%	100.00%
fiction	52.94%	41.18%	5.88%	0.00%	100.00%
nonfiction	61.90%	28.57%	4.76%	4.76%	100.00%
translations	66.67%	33.33%	0.00%	0.00%	100.00%
technology	**66.67%**	**29.17%**	**2.08%**	**2.08%**	**100.00%**
wearables	62.22%	35.56%	0.00%	2.22%	100.00%
web	70.59%	23.53%	3.92%	1.96%	100.00%
music	**56.57%**	**37.71%**	**5.71%**	**0.00%**	**100.00%**
world music	100.00%	0.00%	0.00%	0.00%	100.00%
metal	57.14%	42.86%	0.00%	0.00%	100.00%
jazz	58.82%	35.29%	5.88%	0.00%	100.00%
electric music	55.56%	44.44%	0.00%	0.00%	100.00%
indie rock	51.11%	42.22%	6.67%	0.00%	100.00%
rock	57.65%	35.29%	7.06%	0.00%	100.00%
film & video	**57.30%**	**33.71%**	**6.18%**	**2.81%**	**100.00%**
science fiction	35.71%	64.29%	0.00%	0.00%	100.00%
shorts	56.25%	31.25%	6.25%	6.25%	100.00%
television	64.71%	17.65%	17.65%	0.00%	100.00%
animation	61.76%	29.41%	2.94%	5.88%	100.00%
drama	59.46%	32.43%	5.41%	2.70%	100.00%
documentary	56.67%	35.00%	6.67%	1.67%	100.00%
theater	**54.36%**	**38.37%**	**6.69%**	**0.58%**	**100.00%**
plays	54.36%	38.37%	6.69%	0.58%	100.00%
Grand Total	**56.50%**	**36.40%**	**5.70%**	**1.40%**	**100.00%**

Figure 6 Outcome Ratio by Sub Category Table

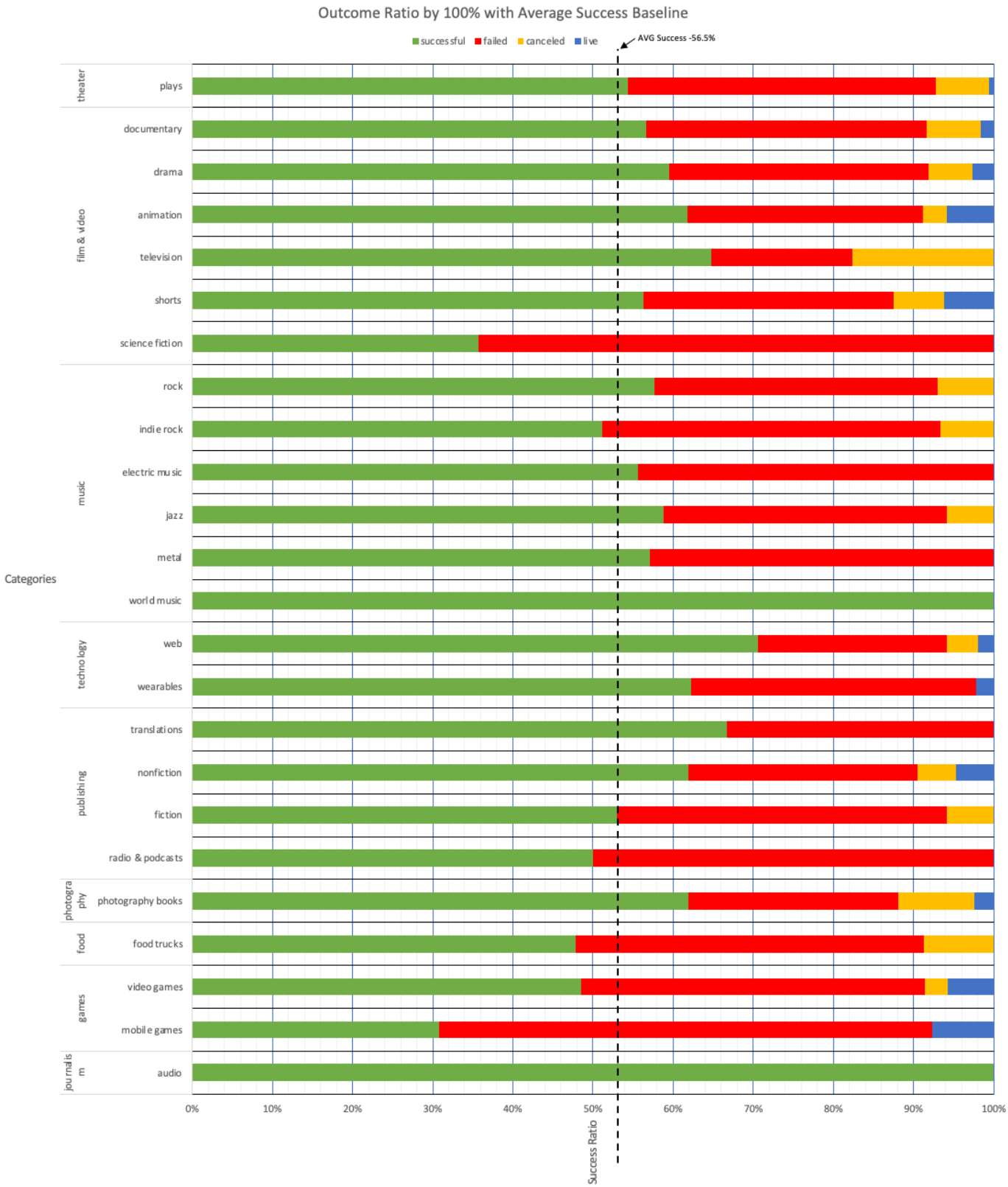


Figure 7 Outcome Ratios by 100% with Average Success Baseline

Success vs Fail Ratio by Sub-Category

	successful Ratio	failed Ratio	Success vs Fail	Probability to Succeed	Average Funding Goal	Average Pledged
audio	100.00%	0.00%	100.00%		$ 6,425.00	$ 9,044.00
world music	100.00%	0.00%	100.00%		$ 48,100.00	$ 75,187.67
translations	66.67%	33.33%	66.67%		$ 37,171.43	$ 39,878.14
web	70.59%	23.53%	66.44%		$ 39,109.80	$ 43,045.94
wearables	62.22%	35.56%	60.84%		$ 26,284.44	$ 36,408.49
metal	57.14%	42.86%	57.14%		$ 10,028.57	$ 40,169.29
animation	61.76%	29.41%	56.31%		$ 50,835.29	$ 50,018.35
nonfiction	61.90%	28.57%	56.01%		$ 58,095.24	$ 53,785.57
electric music	55.56%	44.44%	55.56%		$ 42,238.89	$ 43,397.11
jazz	58.82%	35.29%	55.36%		$ 28,794.12	$ 32,409.76
drama	59.46%	32.43%	54.64%		$ 37,710.81	$ 32,668.78
photography books	61.90%	26.19%	54.54%		$ 32,183.33	$ 29,141.21
rock	57.65%	35.29%	53.58%		$ 38,631.76	$ 42,395.99
television	64.71%	17.65%	53.29%		$ 40,741.18	$ 43,317.12
documentary	56.67%	35.00%	51.94%		$ 56,443.33	$ 40,813.05
plays	54.36%	38.37%	50.41%		$ 45,459.01	$ 45,823.33
radio & podcasts	50.00%	50.00%	50.00%		$ 34,400.00	$ 35,697.50
fiction	52.94%	41.18%	49.83%		$ 56,723.53	$ 52,782.88
shorts	56.25%	31.25%	49.22%		$ 48,468.75	$ 53,488.75
indie rock	51.11%	42.22%	47.70%		$ 50,628.89	$ 45,279.47
video games	48.57%	42.86%	44.41%		$ 42,405.71	$ 32,198.60
food trucks	47.83%	43.48%	43.67%		$ 41,767.39	$ 37,721.28
science fiction	35.71%	64.29%	35.71%		$ 54,735.71	$ 39,979.50
mobile games	30.77%	61.54%	28.40%		$ 105,676.92	$ 68,374.31

Success vs Fail Ratio by Category

	successful Ratio	failed Ratio	Success vs Fail	Probability to Succeed	Average Funding Goal	Average Pledged
journalism	100.00%	0.00%	100.00%		$ 6,425.00	$ 9,044.00
technology	66.67%	29.17%	63.89%		$ 33,097.92	$ 39,934.64
publishing	59.70%	35.82%	57.03%		$ 48,359.70	$ 47,012.34
photography	61.90%	26.19%	54.54%		$ 32,183.33	$ 29,141.21
music	56.57%	37.71%	53.34%		$ 40,150.29	$ 42,743.41
film & video	57.30%	33.71%	52.15%		$ 49,127.53	$ 42,191.44
theater	54.36%	38.37%	50.41%		$ 45,459.01	$ 45,823.33
food	47.83%	43.48%	43.67%		$ 41,767.39	$ 37,721.28
games	43.75%	47.92%	40.10%		$ 59,541.67	$ 41,996.19

Figure 8 Success vs Fail Ratios by Category/ Sub-Category

Furthermore, comparing this Success Probability data with the average initial funding goal and average pledged amounts could reveal patterns that contribute to a project's success. For example, analyzing whether successful projects tend to have lower initial funding goals could indicate a safer target strategy that leads to higher success rates. This approach would allow stakeholders to gauge the financial feasibility of different project types on the platform, e.g. Even though Journalism – Audio has an extremely high success rate (100%), it is likely partly due to the fact that the Initial Goals and Avg Pledged amounts have consistently been comparatively extremely low.

Statistical Analysis -

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

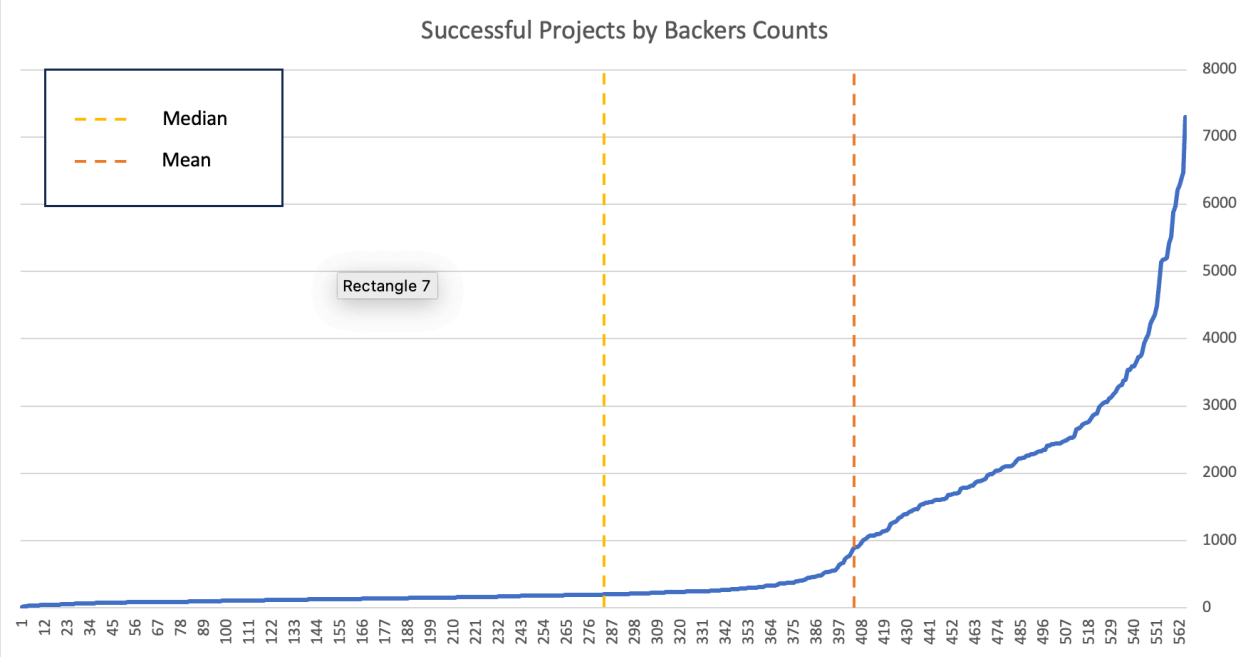


Figure 9 Successful Projects by Backers Counts

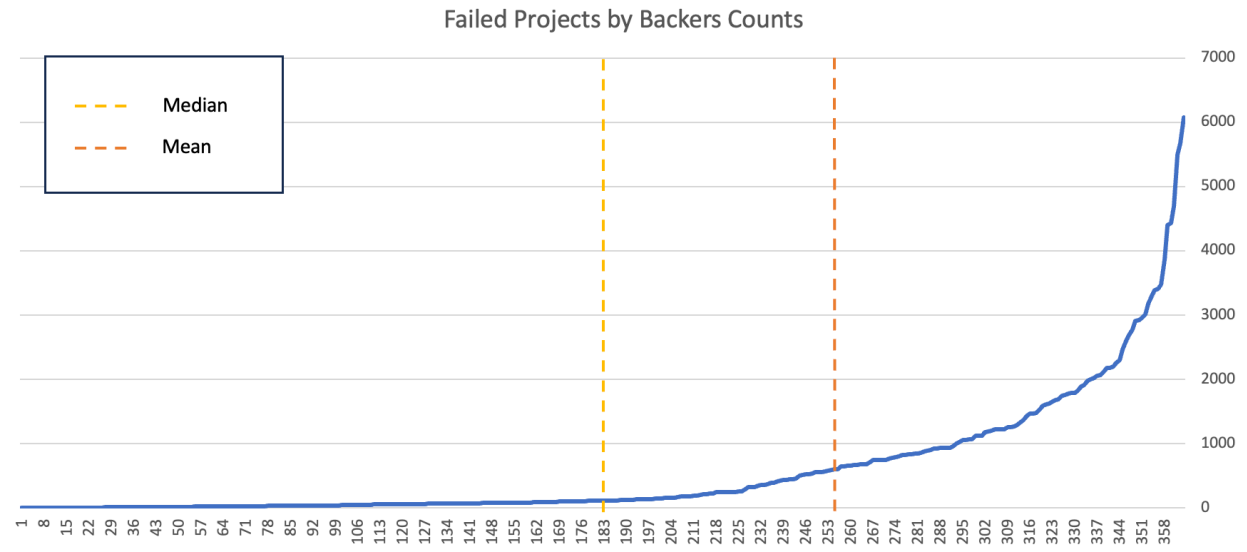


Figure 10 Failed Projects by Backers Counts

Considering the line graphs depicting the data of Successful and Failed backers are extremely positively skewed due to the high level of variance in the dataset, the mean better summarizes the data set than the median. This is clearly seen in the charts above, as the medians in the data set completely miss the areas where the data shows significant change. Though this usually would suggest that the data has a lower amount of outlier data points, in this case, it is more likely due to the variance between categories and the factors affecting each of their popularities and initial goal amounts on the platform. Thus, I believe these chart data sets would only provide truly actionable insights if categorized by category or subcategory.

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

Central Tendency - Range - Variance of Successful and Failed Projects

	Mean	Median	Maximum	Minimum	Variance	Standard Deviation
successful	851.1469	201	7295	16	1606216.594	1267.366006
failed	585.6154	114.5	6080	0	924113.455	961.3081998

Figure 11 Central Tendency - Range - Variance of Successful and Failed Projects

The data in the table depicting variance and standard deviation clearly points to the fact that successful project data has a much higher amount of variability. This is evident in the variance values, which are nearly double for successful projects compared to failed projects, with the standard deviation following a similar trend (see above). This is most likely due to the criteria for categorizing projects as successful or not: everything under 100% pledged is designated as failed, while everything above is marked as successful. As a result, the variance of failed projects is limited to that threshold, while the variance of successful projects is unrestricted and depends solely on the maximum level of interest a single project can attract—something that is very hard, if not impossible, to calculate precisely.