Crowdfunding Campaign Analysis

University of Toronto

Ahmed Abdelrahman

31 Oct 2022

Contents

1 Data

Data is a sample of 1000 projects sourced from the University of Toronto Bootcamp course: UTOR-VIRT-DATA-PT-10-2022-U-LOLC-MTTH and is saved alongside this report in sheet 1 of the Excel document 'CrowdfundingBook_AA.xlsx'.

2 Results

2.1 Outcome per Category

Information was extracted from the data set to show the total number of outcomes (successful, failed, live, canceled) per parent category, see Figure 1, and the outcome type as a percentage of the total per parent category, see Figure 2.

[Next Page]

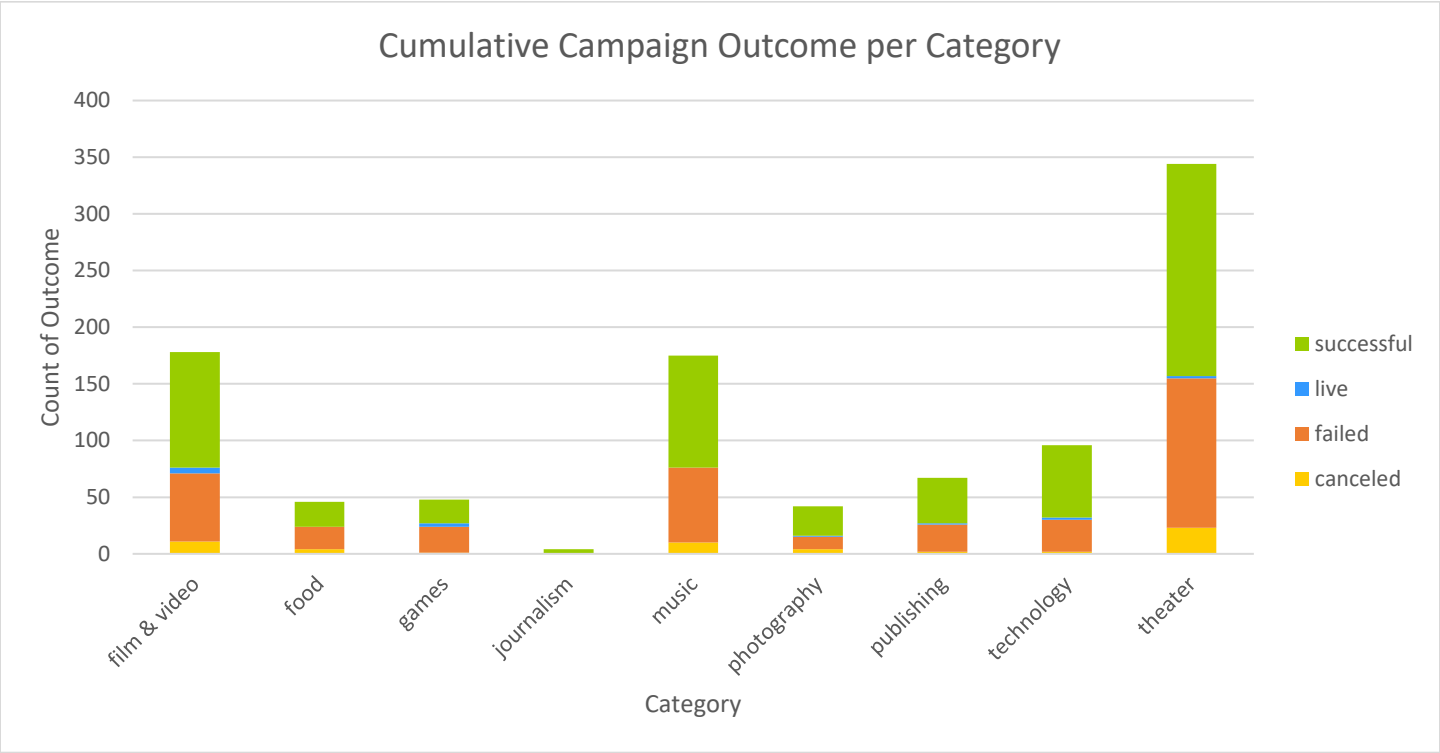


Figure 1: Stacked column plot of the total number of outcomes per parent category:

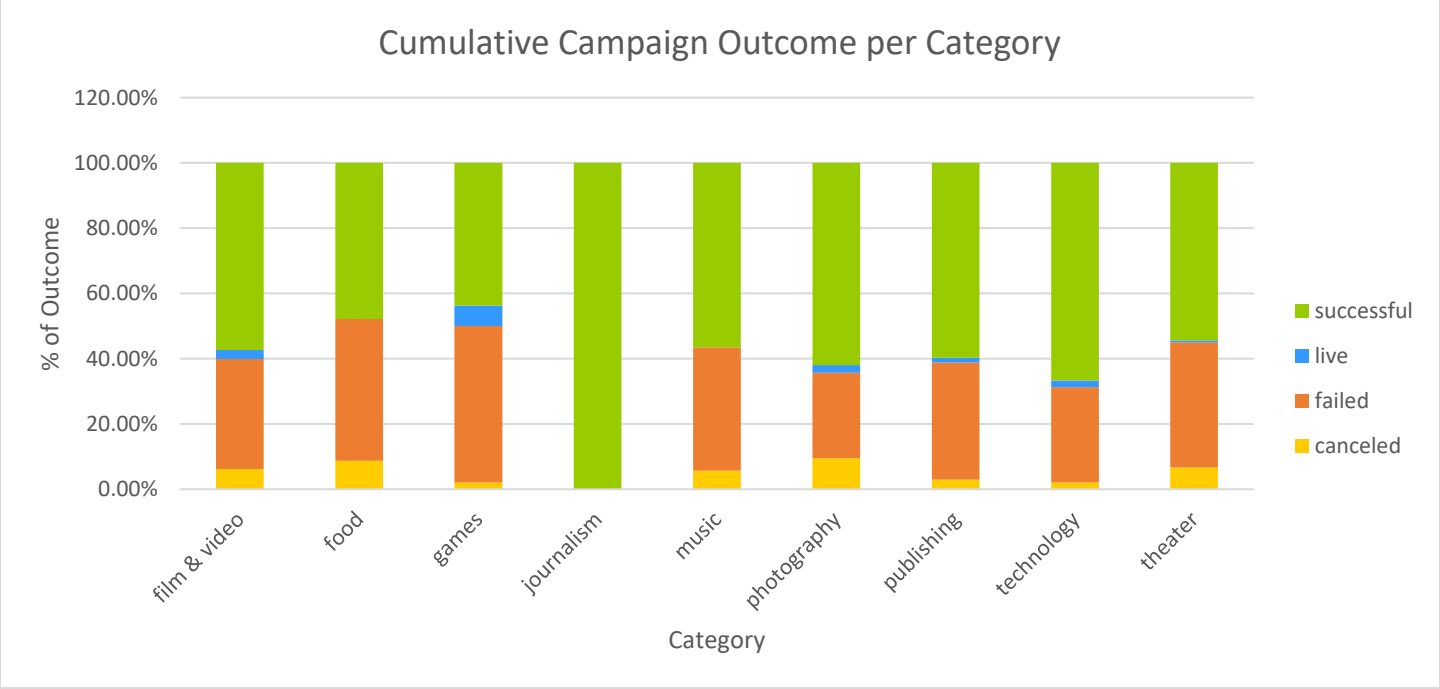


Figure 2: Stacked column plot of the percentage of outcome type of the total outcomes per parent category

2.2 Outcome per sub-Category

Information was extracted from the data set to show the total number of outcomes (successful, failed, live, canceled) per sub-category, see Figure 3, and the outcome type as a percentage of the total per sub-category, see Figure 4.

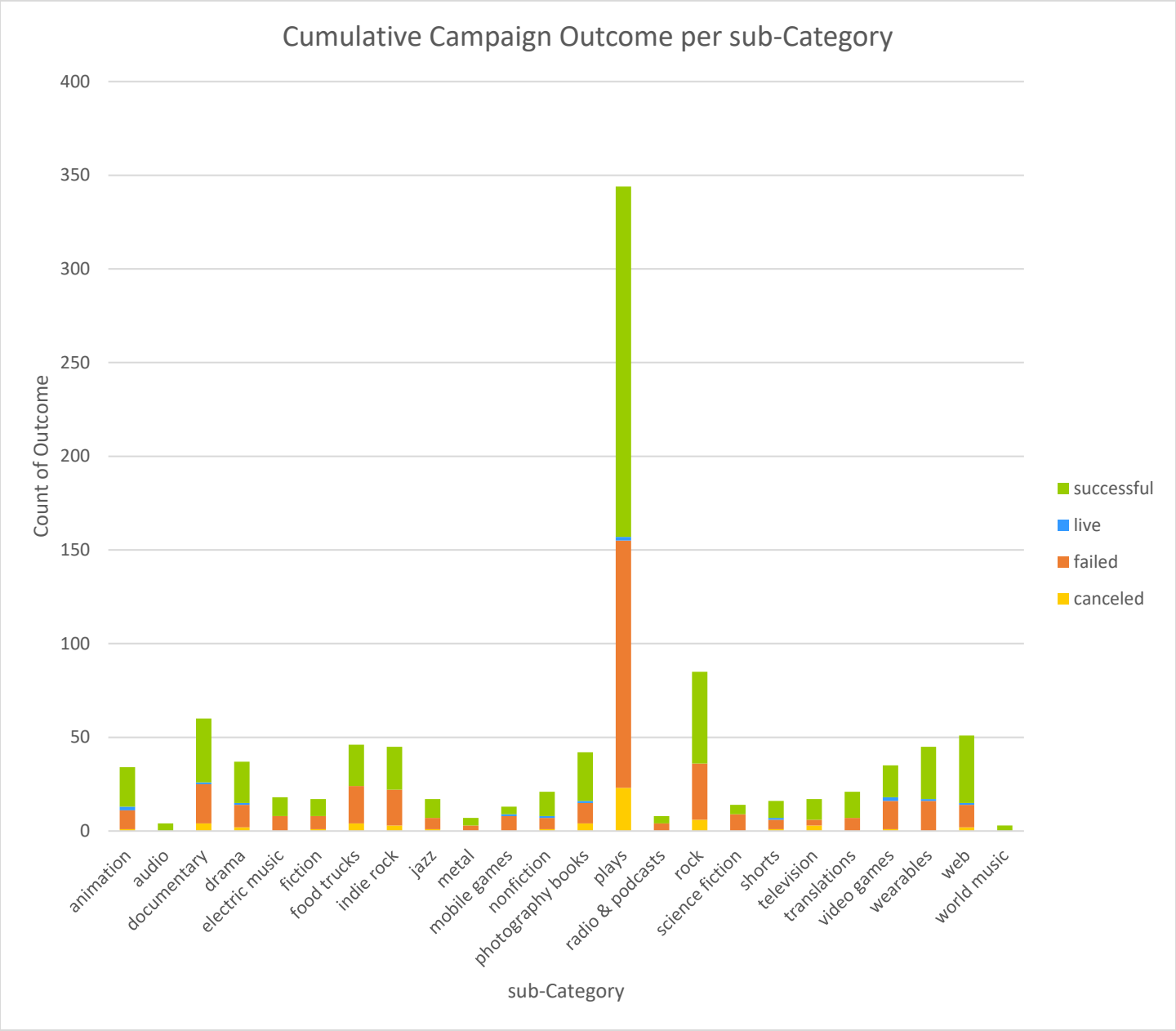


Figure 3: Stacked column plot of the total number of outcomes per sub-category

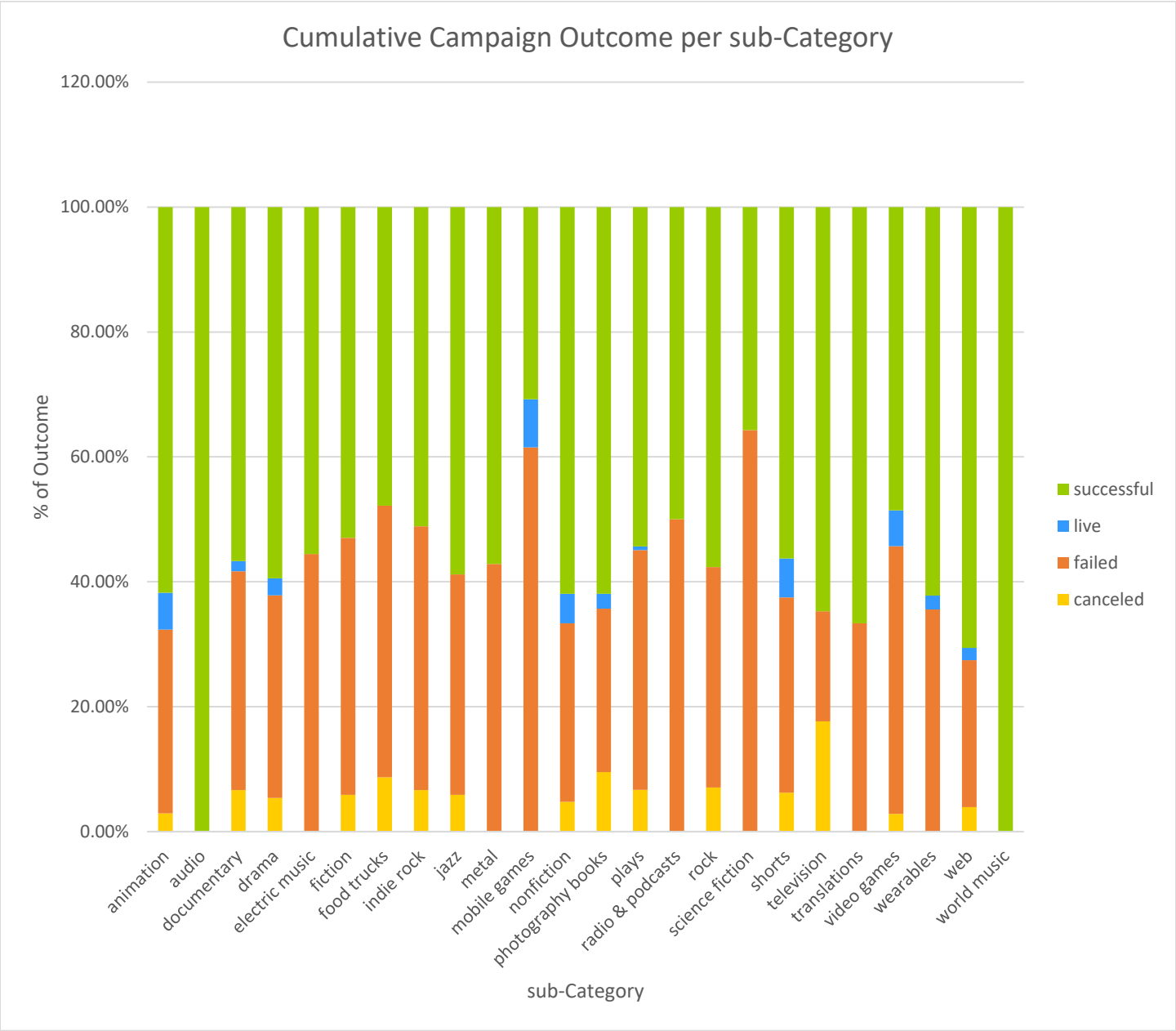


Figure 4: Stacked column plot of the percentage of outcome type of the total outcomes per sub-category

2.3 Outcome per Launch Month

Information was extracted to show the outcome per launch month of the campaign.

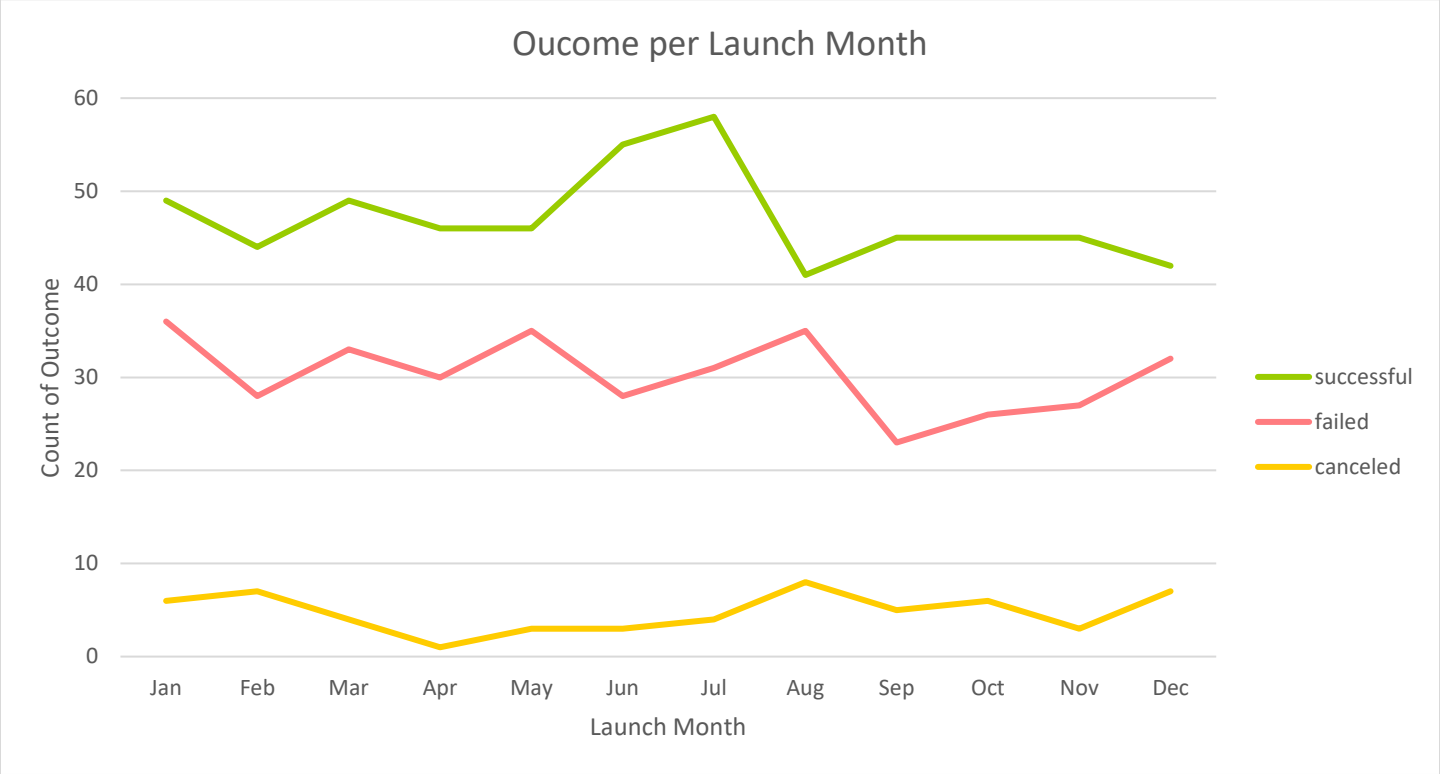


Figure 5: Line plot showing the number of each outcome type (excluding live) per launch month

2.4 Outcome based on Target Funding (Goal)

Information was extracted to show the outcome based on the target funds for the campaign.

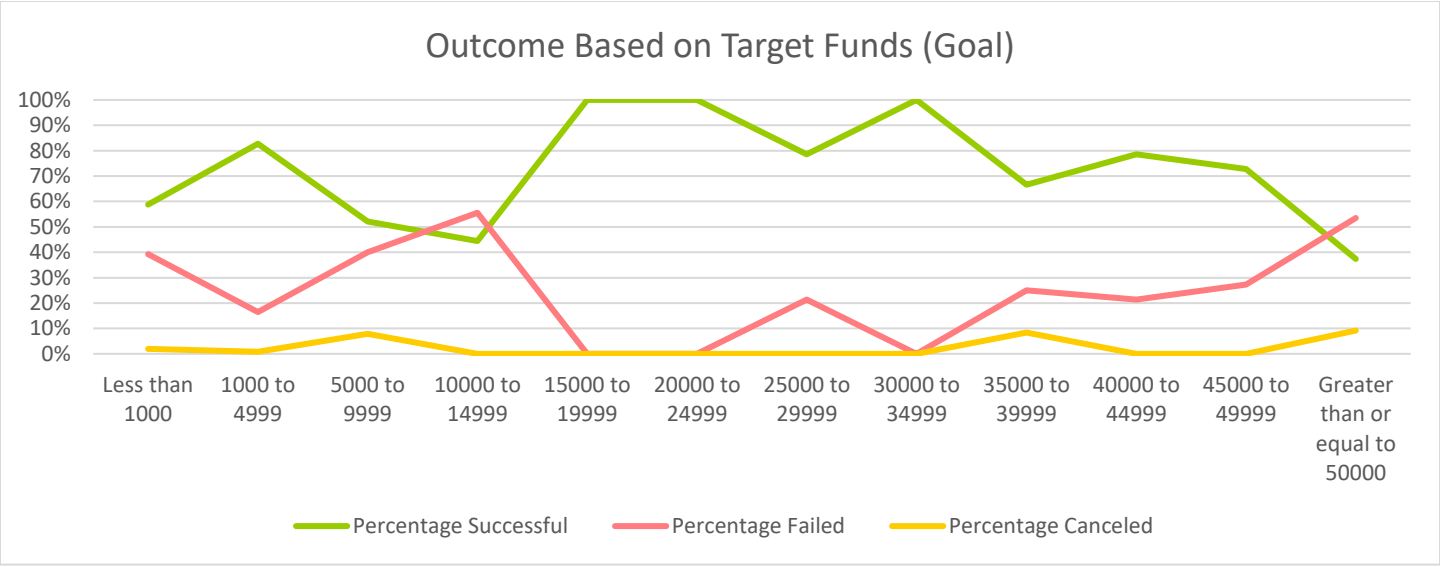


Figure 6: Line plot showing the percentage of each outcome type (excluding live) per Goal range

2.5 Statistical Analysis

The number of backers ('backers_count') for successful and failed campaigns was statistically analyzed, see Table 1, and plotted as a bar chart of raw data and as a line plot of the normal distribution, see Figure 7 and Figure 8 respectively.

Table 1: Statistical Analysis for the Number of Backers

'backers_count' Statistic	**Successful**	**Failed**
population (total)	565	364
mean	851.15	585.62
median	201.00	114.50
min	16	0
max	7,295	6,080
variance	1,606,216.59	924,113.45
std. dev	1,267.37	961.31

[Next Page]

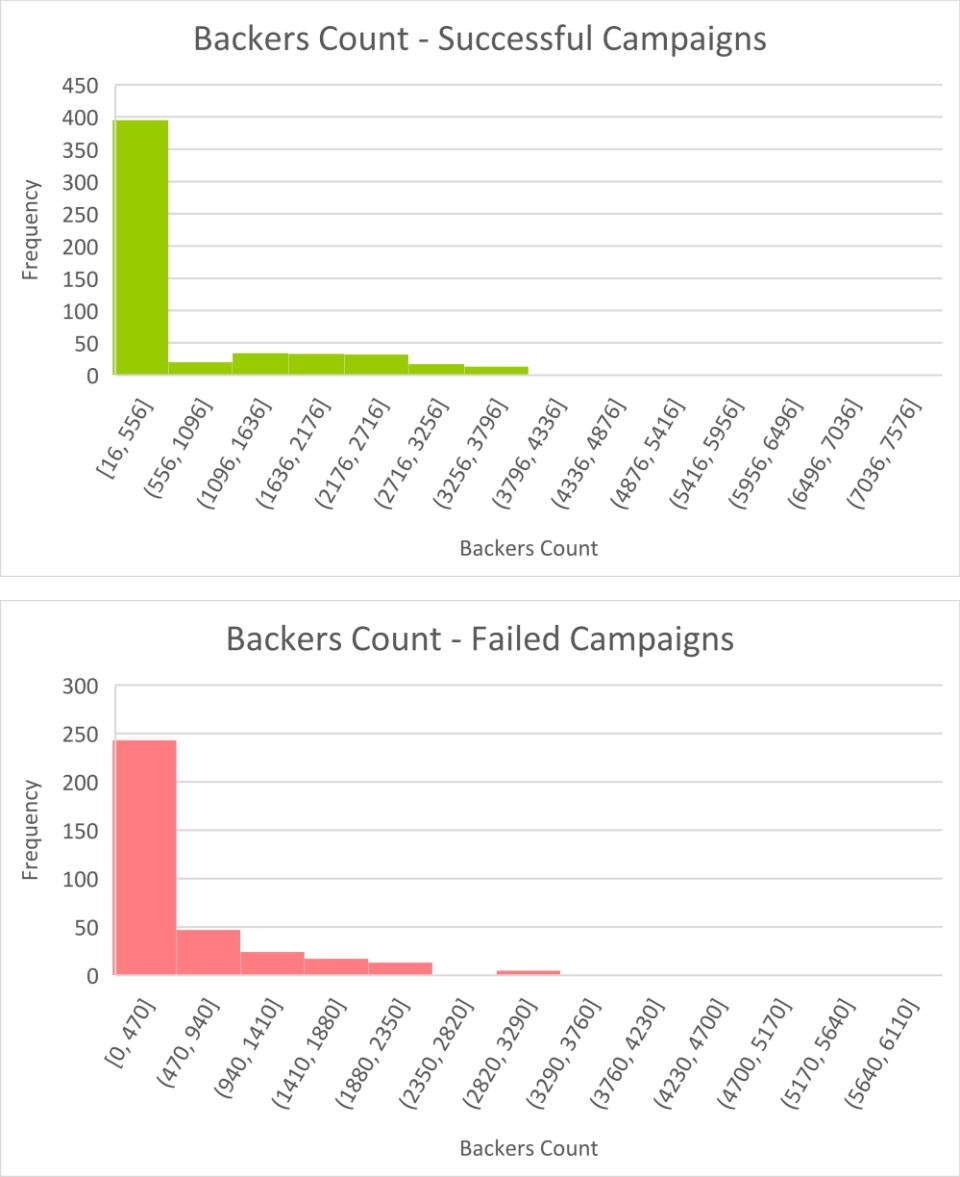
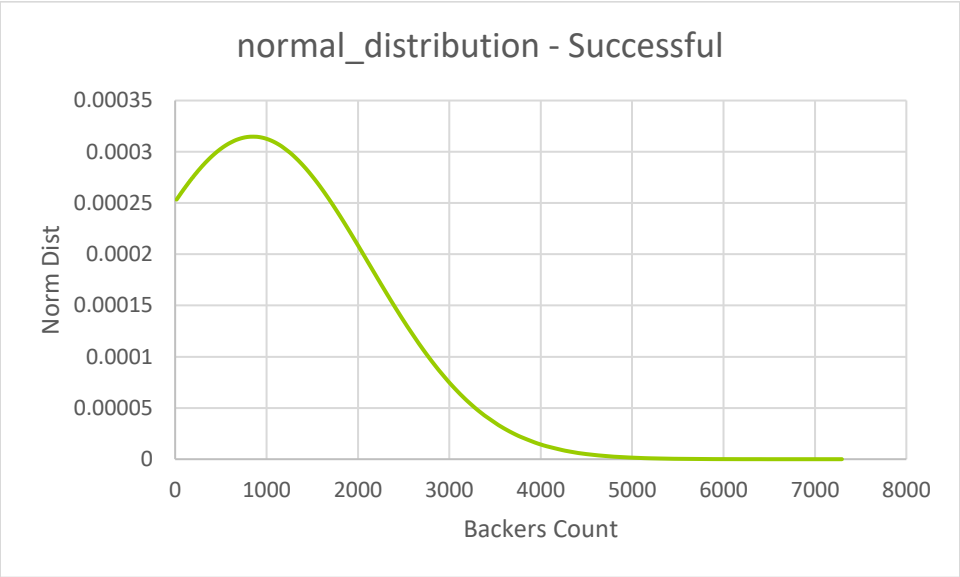


Figure 7: Bar charts of raw data of successful and failed campaigns per goal range



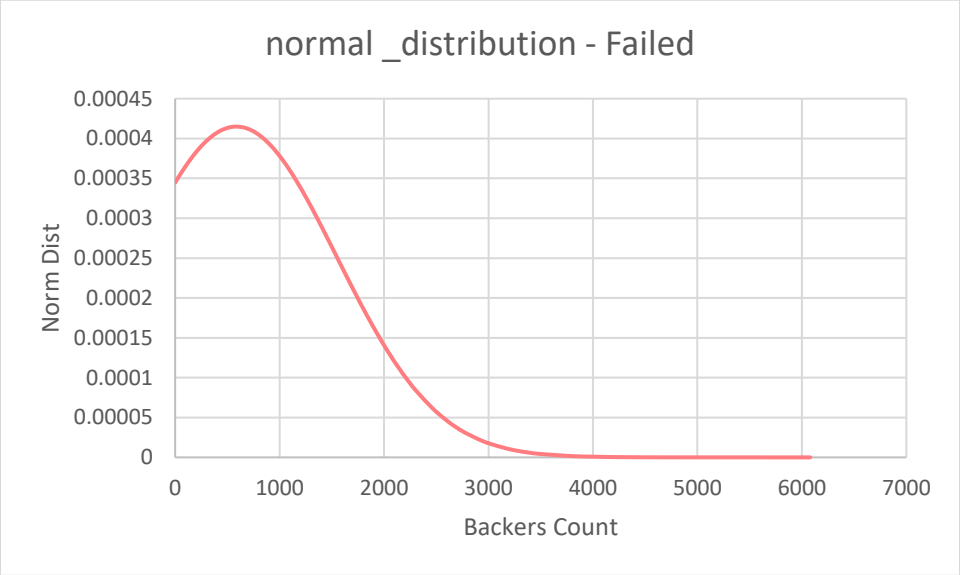


Figure 8: Line plot of normal distribution of successful and failed campaigns per goal range

3 Conclusions

3.1 Main Conclusions

1. The largest number of campaigns seem to be related to pop culture type categories represented in the top three categories: 'film& video', 'music', and 'theater'. The highest number being in the 'theater' and specifically the subcategory of 'plays'.
2. The success rate varies approximately between 45-60% in all categories, with the lowest being in games at 43.75% success rate and the highest in technology at 6.67%. However, the population per category seems to vary significantly, and without further sampling it might be difficult to determine a strong relationship between category and success rate. (Note: Journalism with 100% success rate was excluded due to having a significantly low number or campaigns).
3. The data suggests that campaigns launched in the summer months of June and July have overall a higher chance for success.
4. The data shows that the majority of campaigns have funding goals in the three ranges '$1,000 - $5,000' '$5,000 - $10,000', and 'greater than $50,000'. The data suggests that among these 3 top packed categories, success rate is significantly higher for the campaigns with funding in the range of '$1,000 - $5,000'.

3.2 Limitations of Data Set

- The data does not breakdown campaign by the host website. The host websites could have significantly varying popularity, and this could prove useful when deciding where to launch the campaign.
- The data also does not include any information about the funding that was spent on advertising the campaigns, which can also affect the outreach.

3.3 Additional Potentially Valuable Relationships

- Creating a plot that shows the percentage funded for successful projects in ranges (similar to the goals plot in Figure 6) would help uncover which projects are the most over-funded, which can be used as a further indicator of the popularity of campaigns. The relationship can be made against category, sub-category, launch date, country, etc.
- Creating plots that show the success/failure rate with projects which were featured in SPOTLIGHT or STAFF PICK could help uncover the potential effect of having such a flag on the performance of the campaign.
- Creating plots that show the relationship between funding goal vs SPOTLIGHT/STAFF PICK feature and category vs SPOTLIGHT/STAFF PICK feature could help uncover the potential of gaining such a flag for a new project.

3.4 Notes on Stat Analysis

- The mean is not a good indicator to summarize the backers count data as due to the shape of the normal distribution. The data is skewed and not mirrored around the mean, in this case the median could be a better stat to summarize the data.
- The standard deviation indicates that there is more variability within the successful campaigns which makes sense because the population is larger.