Qualitative Data Analysis Report of Kickstarter Funding Behavior

Author: Kimberley Maldonado

Rutgers Data Science Bootcamp

4/10/2021

Introduction

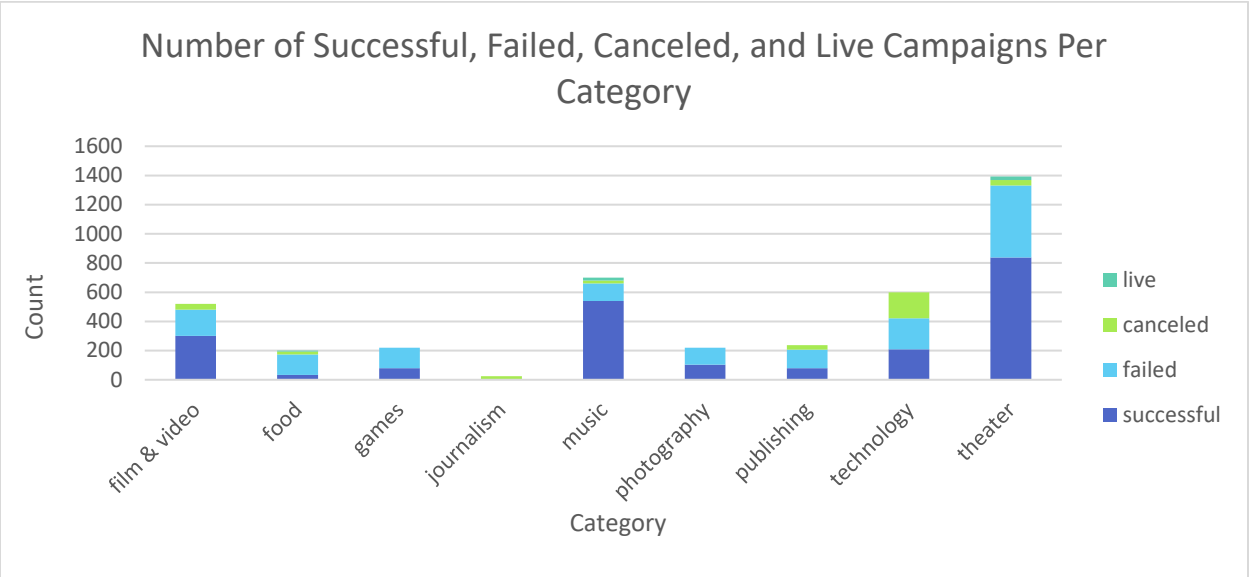
The community-based crowdfunding service Kickstarter allows an individual or group of individuals to receive funding for a project from the general public as opposed to for example, traditional venture capital funding. In this case, success is defined as having met or exceeded the project's initial goal. In order to gain insight into the characteristics of successful campaigns, data was collected from the platform spanning 21 countries. The final result (success or failure) of individual campaigns was monitored alongside several metrics including industry category and sub-category, number of individual backers and the average donation made to each campaign. Using Excel to visualize the relational information provided by the raw data on Sheet 1[1], it is evident that

1. The projects in the music and theater categories were the most successful i.e., benefitted most from crowdfunding on Kickstarter, both in terms of total number of successful campaigns and conditional row percentage based on the number of campaigns in the specific category.
2. The largest number of successful campaigns over a cumulative period from 2009 to 2017 were initiated during the month of May.
3. There is approximately a 30% drop from 51% to 19% between the percentage of successful campaigns in the $40,000 to $44,999 goal range and the percentage of successful campaigns in the $50,000+ goal range.
4. Both the average and median number of backers was higher for successful campaigns than for failed campaigns.

[1] Check Appendix for worksheet descriptions.

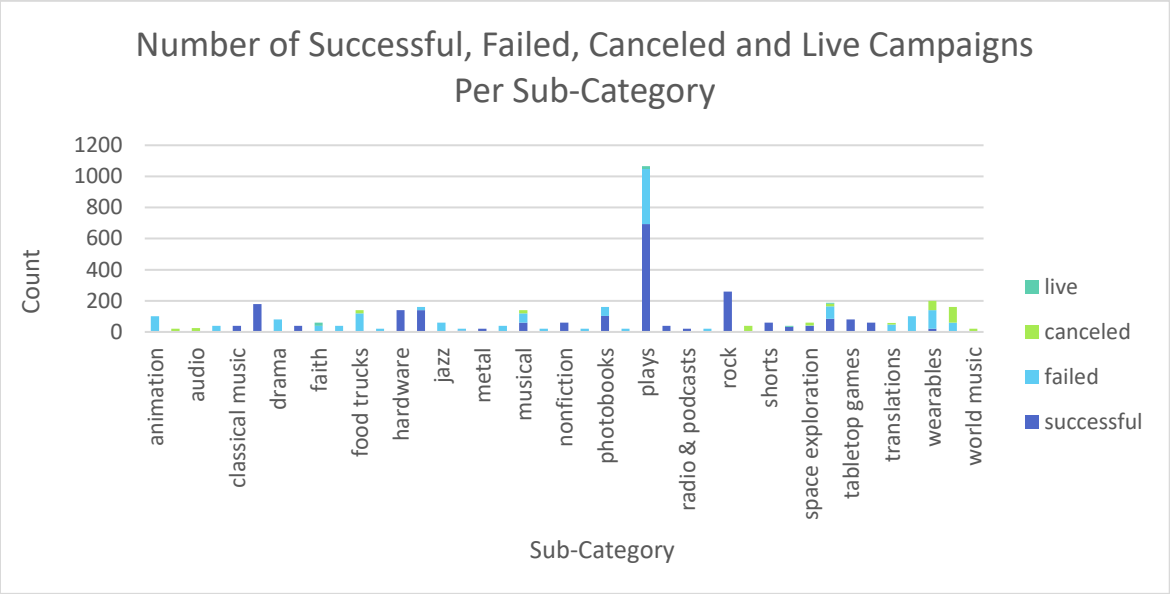
Methods

The pivot table on Sheet 2 describes the number of successful, failed, canceled and live campaigns according to category. Its columns reference the status of the campaign and the rows reference the category pertaining to the campaign. This table is able to be filtered by country.



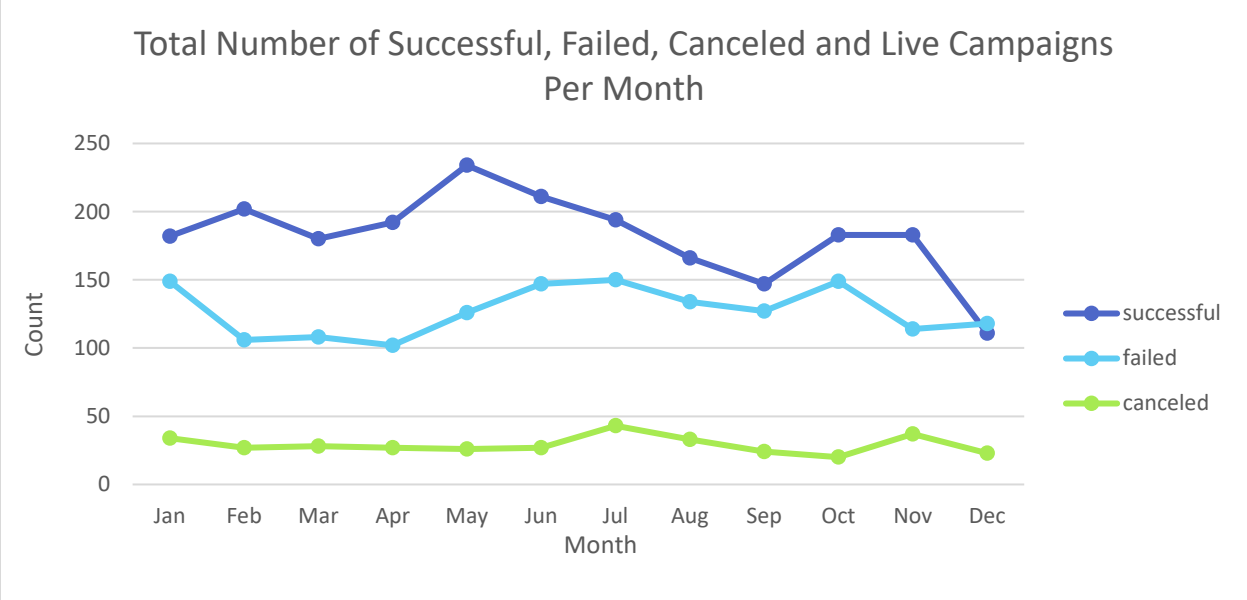
Pivot Chart Corresponding to Sheet 2 Table

The pivot table on Sheet 3 below examines the number of successful, failed canceled and live campaigns according to sub-category. As in Sheet 2, the columns reference the state of the campaign and the rows reference the sub-category. This table has the option to filter by country and main category.



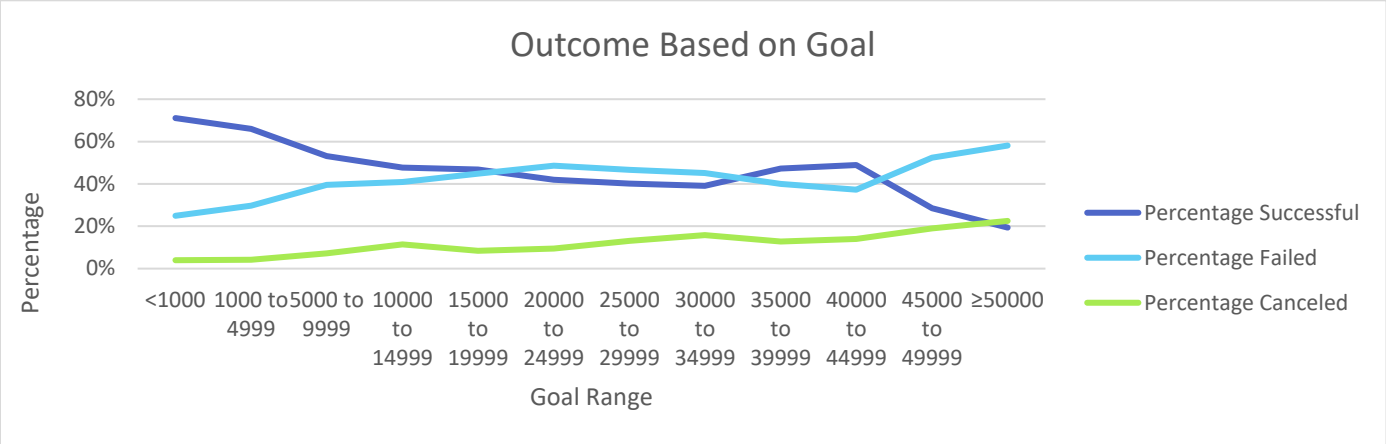
Pivot Chart Corresponding to Sheet 3 Table

The pivot table found on Sheet 4 counts the total number of successful, failed, canceled, and live campaigns taking place each month of year from 2009 to 2017. The columns reference the status of the campaign while the rows reference the date in which the campaign launched. This table is filterable by year and by category.



Pivot Chart Corresponding to Sheet 4 Table

The Sheet 5 table separates the campaign goal amounts in Sheet 1, Column D into dollar ranges and counts the number of successful, failed and canceled campaigns in each one. A total is calculated across each status within the range. A percentage is computed by dividing the number of successful, failed and canceled campaigns by the total in the range, with formatting done on the cells to reflect a percentage.



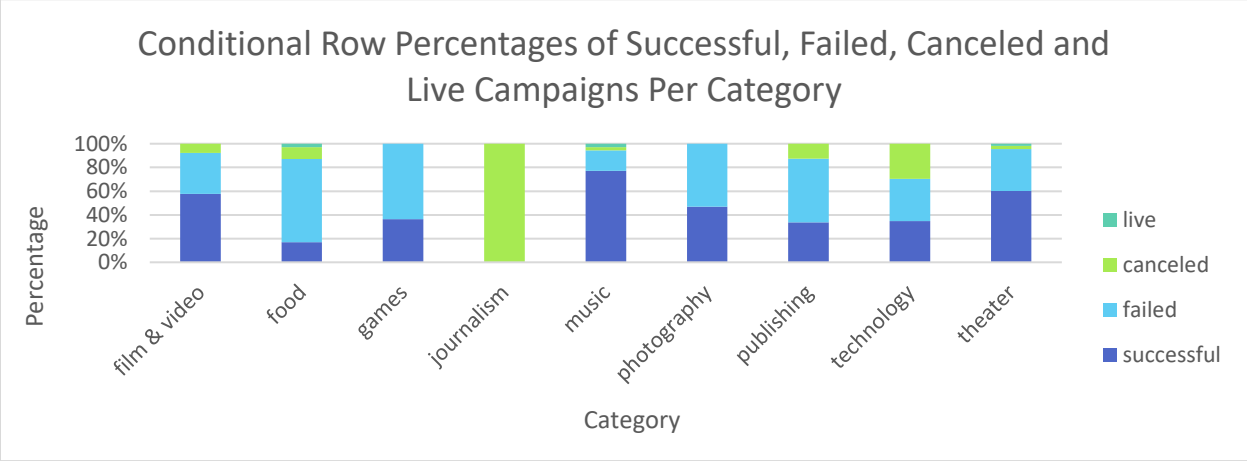
Line Chart Corresponding to Sheet 5 Table

In addition to the visualizations provided by the bar and line charts in Sheets 2 to 5, the summary statistics provided on Sheet 6 is useful for further analysis and investigation into the underlying patterns in the data. For this, the mean, median, minimum, maximum, variance and standard deviation figures are common barometers used to describe the data's behavior.

Results

From the pivot charts denoting the number of successful, failed, canceled and live campaigns per category and sub-category, it is clear that the music and theater categories and the play sub-category have had the highest number of successful campaigns. However, while it is useful to have a count of successful campaigns, it is also valuable to note the success rate in relation to the total number of campaigns within its own specific category or sub-category. This means taking a conditional percentage by dividing the number of successful campaigns in "music", for example, and dividing it by the total number of music campaigns. This can be done with a pivot table with the values being displayed as a percent of the row total as follows:

Count of state	**Column Labels**				
Row Labels	**successful**	**failed**	**canceled**	**live**	**Grand Total**
film & video	57.69%	34.62%	7.69%	0.00%	100.00%
food	17.00%	70.00%	10.00%	3.00%	100.00%
games	36.36%	63.64%	0.00%	0.00%	100.00%
journalism	0.00%	0.00%	100.00%	0.00%	100.00%
music	77.14%	17.14%	2.86%	2.86%	100.00%
photography	46.82%	53.18%	0.00%	0.00%	100.00%
publishing	33.76%	53.59%	12.66%	0.00%	100.00%
technology	34.83%	35.50%	29.67%	0.00%	100.00%
theater	60.23%	35.39%	2.66%	1.72%	100.00%
Grand Total	**53.11%**	**37.19%**	**8.48%**	**1.22%**	**100.00%**



Pivot Chart Corresponding to Sheet 9 Table

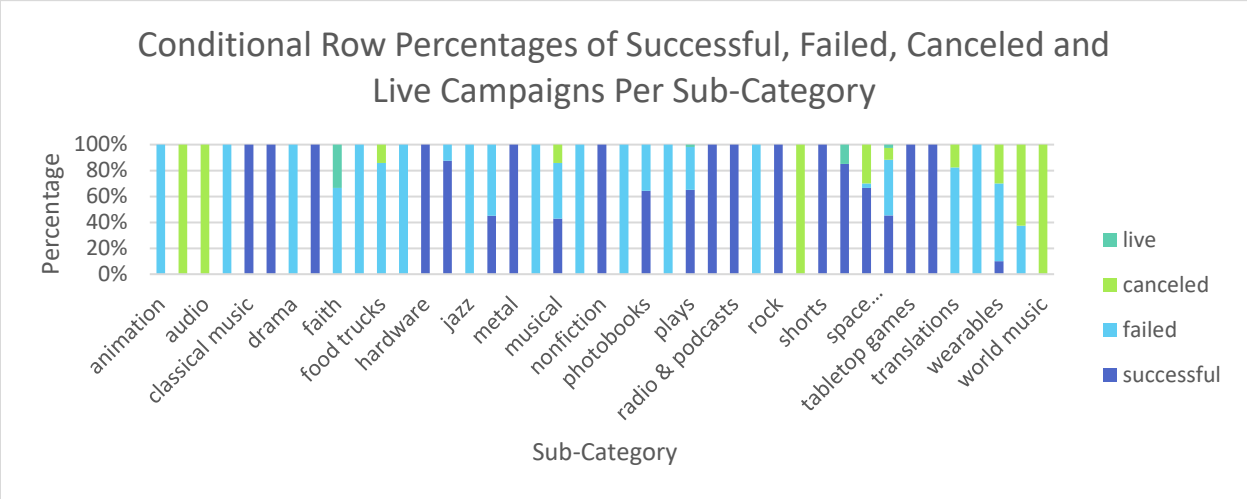
From this, we can gather that though theater had the highest number of successful campaigns, music had the highest percentage of successful campaigns.

A similar process can be done to the sub-category table.

country	(All)
Category	(All)

Count of state	Column Labels				
Row Labels	successful	failed	canceled	live	Grand Total
animation	0.00%	100.00%	0.00%	0.00%	100.00%
art books	0.00%	0.00%	100.00%	0.00%	100.00%
audio	0.00%	0.00%	100.00%	0.00%	100.00%
children's books	0.00%	100.00%	0.00%	0.00%	100.00%
classical music	100.00%	0.00%	0.00%	0.00%	100.00%
documentary	100.00%	0.00%	0.00%	0.00%	100.00%
drama	0.00%	100.00%	0.00%	0.00%	100.00%
electronic music	100.00%	0.00%	0.00%	0.00%	100.00%
faith	0.00%	66.67%	0.00%	33.33%	100.00%
fiction	0.00%	100.00%	0.00%	0.00%	100.00%
food trucks	0.00%	85.71%	14.29%	0.00%	100.00%
gadgets	0.00%	100.00%	0.00%	0.00%	100.00%
hardware	100.00%	0.00%	0.00%	0.00%	100.00%
indie rock	87.50%	12.50%	0.00%	0.00%	100.00%
jazz	0.00%	100.00%	0.00%	0.00%	100.00%
makerspaces	45.00%	55.00%	0.00%	0.00%	100.00%
metal	100.00%	0.00%	0.00%	0.00%	100.00%
mobile games	0.00%	100.00%	0.00%	0.00%	100.00%
musical	42.86%	42.86%	14.29%	0.00%	100.00%
nature	0.00%	100.00%	0.00%	0.00%	100.00%
nonfiction	100.00%	0.00%	0.00%	0.00%	100.00%
people	0.00%	100.00%	0.00%	0.00%	100.00%
photobooks	64.38%	35.63%	0.00%	0.00%	100.00%
places	0.00%	100.00%	0.00%	0.00%	100.00%
plays	65.10%	33.11%	0.00%	1.78%	100.00%
pop	100.00%	0.00%	0.00%	0.00%	100.00%
radio & podcasts	100.00%	0.00%	0.00%	0.00%	100.00%
restaurants	0.00%	100.00%	0.00%	0.00%	100.00%
rock	100.00%	0.00%	0.00%	0.00%	100.00%
science fiction	0.00%	0.00%	100.00%	0.00%	100.00%
shorts	100.00%	0.00%	0.00%	0.00%	100.00%
small batch	85.00%	0.00%	0.00%	15.00%	100.00%
space exploration	66.67%	3.33%	30.00%	0.00%	100.00%
spaces	45.45%	42.78%	9.09%	2.67%	100.00%

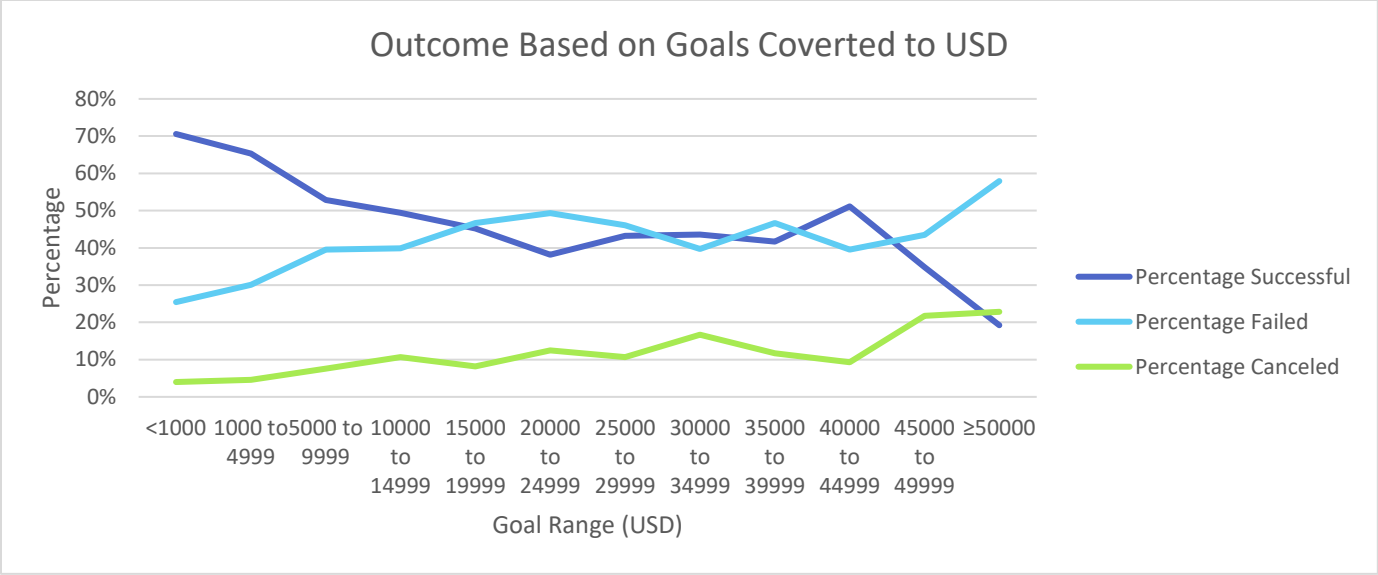
tabletop games	100.00%	0.00%	0.00%	0.00%	100.00%
television	100.00%	0.00%	0.00%	0.00%	100.00%
translations	0.00%	82.46%	17.54%	0.00%	100.00%
video games	0.00%	100.00%	0.00%	0.00%	100.00%
wearables	10.00%	60.00%	30.00%	0.00%	100.00%
web	0.00%	37.50%	62.50%	0.00%	100.00%
world music	0.00%	0.00%	100.00%	0.00%	100.00%
Grand Total	**53.11%**	**37.19%**	**8.48%**	**1.22%**	**100.00%**



Pivot Chart Corresponding to Sheet 10 Table

From the cumulative monthly chart from Sheet 4, it can be stated that campaigns started in May were most successful across the 2009 to 2017 time period.

The chart displaying success trends across several goal ranges from Sheet 5 could be improved by assuring that all campaign goals are expressed in one uniform currency, the same way other technical disciplines require consistent usage of units. This is because the non-converted currencies increase the subjectivity of the necessarily independent goal range variable, even if most of the data entries are in USD.



Line Chart Corresponding to Sheet 8 Table

The chart above displays the same information as Sheet 5, except all currencies have been converted to USD using an exchange rate table located on Sheet 7. Though the overall trends in the percentages of successful, failed and canceled goal ranges are similar, there are key nuances in the behavior that should be noted. For example, it would appear that in Sheet 5 between 35000-39999 and 45000-49999, the percentage successful is clearly higher than the percentage failed. However, in the converted chart above, the percentage failed is higher than the percentage successful in the 35000-39999 goal range, reducing the successful range between 40000-44999 and 45000-49999. Both charts display an approximately 30% drop in success between the 40000-44999 range and the 50000+ range.

The summary statistics on Sheet 6 show that because the mean and median number of backers in successful and failed campaigns are **not** approximately equal (194.4 ≈/ 62 backers in successful campaigns and 17.7 ≈/ 4 backers in failed campaigns), the data does not reflect a normal, or symmetrical, distribution, but a skewed one, indicating the presence of outliers. Outliers highly impact the mean computation and often produces an inaccurate portrait of the data's behavior. Therefore, in this case, the median summarizes the "center" of the data better than the mean. There is additionally more variability in the number of backers, as described by the variance and standard deviation, for successful campaigns than for failed ones. This makes sense considering that a successful campaign could have been completely funded by as few as one person or as many as 26,457 backers, but the maximum number of backers in the failed campaigns was only 1293 with a minimum of zero backers.

Conclusions

The fact that music and theater were the most successful categories both in terms of number of successful campaigns and percent of successful campaigns indicates that there may be a stronger culture or emphasis on art and creativity amongst those who use Kickstart to fund their campaigns than in other disciplines. Additionally, that fact that May was the most fruitful month for successful campaigns may coincide with other important dates out the year, including the onset of Spring and college graduations. The drop in successful campaigns past a certain goal range threshold indicates that the more expensive a campaign is, the more work is needed to justify the monetary demands of that project and convince a larger number of people to donate. Ultimately, successful campaigns have strength in numbers, and it is a campaign's ability to reach a larger number of people which will decide its outcome.

Appendix

Sheet 1: Raw data

Sheet 2: Pivot table of number of successful, failed, canceled and live campaigns per category with pivot bar chart

Sheet 3: Pivot table of number of successful, failed, canceled and live campaigns per sub-category with pivot bar chart

Sheet 4: Pivot table of total number of successful, failed, canceled and live campaigns per month with pivot line graph

Sheet 5: Outcome based on goal table plus line graph

Sheet 6: Summary statistics (preliminary computations plus final summary statistics)

Sheet 7: Converted goal table plus reference table (conversion chart)

Sheet 8: Outcome based of converted goal table plus line graph

Sheet 9: Pivot table of conditional row percentages of successful, failed, canceled and live campaigns per category with pivot bar chart

Sheet 10: Pivot table of conditional row percentages of successful, failed, canceled and live campaigns per sub-category with pivot bar chart

Sheet 11: Number of backers versus average donation scatter plot (not mentioned in report)