**Report on Crowdfunding**

CONCLUSIONS

1. Based on the Percentage of Successful Projects in relation to their Funding Goals, Projects become noticeably less likely to succeed if their Goals exceed $35000, with anything higher than $50000 being more likely to fail than to succeed (%53 vs. %37). In contrast, anything LOWER than $15000 also has a success rate lower than 60%, save for projects between $1000 and $4999 with a respectable 83% success rate. Therefore, it seems that a project is most likely to succeed if its Goal is between $15000 and $34999. Note that there are vastly more projects with goals between $1000 - $4999 and $5000 - $9999 (and to a lesser extent projects with goals less than $1000) compared to Projects of other Goal ranges, so there is the issue of inconsistent sample sizes potentially skewing the data
2. Based on the Parent Column Chart sheet, most projects in the dataset are based in the US. The “shape” of the chart when filtered to only show projects from the US is quite similar to its shape when no filter is applied. This also applies to the chart in the Sub Column Chart Sheet. Therefore, it is safe to assume that this dataset is most representative of an American audience
3. The Parent Categories with the highest rates of success according to the Parent Column Chart sheet are Technology, Film & Video, and Music. While Technology technically has the highest success rate at 65.22% success/34.78% fail, Film & Video and Music are not far behind with success rates of 64.96% success/35.04% fail and 64.23% success/35.77% fail respectively. This, combined with the fact that they have a higher number of Projects overall then Technology is why they are included. Note that these success rates are based on the totals of each project of each category AFTER SUBTRACTING canceled and live projects from them.

Within the above-mentioned Parent Categories, here are the Sub Categories with the highest success rates (with the amount of projects for each sub category accounted for). Sub categories shown in **bold** have more projects representing them then the other one listed

* 1. Technology: Web (75% success/25% fail)
  2. Film & Video: Television (78.57% success/ 21.42% fail), **Animation** (67.74% success/ 32.25% fail)
  3. Music: **Rock** (62.03% success/37.97% fail), Jazz (62.5% success/37.5% fail)

Based on these results, projects of any one of these Sub Categories are more likely to be successful overall

LIMITATIONS

One thing this dataset cannot properly address are Rewards and Stretch Goals. Rewards are what they sound like, rewards given to those who donate to a project, which can be anything from early access to the product, access to special merchandise, etc. Stretch goals are extra features that are promised to be added should the projects exceed its funding goals at certain points. While these both offer a proverbial carrot to dangle in front of an audience to make sure they donate, the nature of Rewards and Stretch Goals vary wildly, often being extra specific to the project they come from. If we were to include these in our dataset, we would have to spend extra time narrowing them down into more broad categories, which given how project specific Rewards and Stretch Goals can be would not be an easy task. As such, we elected not to include Rewards and Stretch Goals in our dataset to avoid such overcomplication.

Another Limitation comes from the projects listed as “canceled”. The dataset does not mention WHY these projects were canceled, and even if they did it is doubtful that they would be of much use in a dataset where canceled projects make such a small contribution.

In relation to our 3rd conclusion, one limitation is that our data does not account for combinations of Sub Categories. After all, if there were a project for an “Animated Drama”, which Sub Category would that go under?

OTHER TABLES AND GRAPHS

The success rates seen in our 3rd conclusion were all done by hand. However, it would have been much more efficient if we used a modified version of Parent Column Chart and Sub Column Chart that made those calculations for us.

As for new charts, one suggestion is a Vertical Bar Chart where x represents Parent Categories, and y represents the average number of backers for each Parent Categories. We figure that this would be a good way to gauge which industries attract the most backers. We might also have a variation of this where instead of average backers, the y axis represent the average donation for each Parent Categories. If we put these two charts side by side, then that would give us a good idea of the relationship between Average number of Backers and the Average Donation for each Primary Category

STATISTICAL ANALYSIS OF NUMBER OF BACKERS SHEET

It appears as though the Median is the best way to find summarize the data present on Number of Backers. The deciding factor is that, upon sorting the Backer Count columns for both Successful and Failed Projects, we noticed that in both cases a larger proportion of projects had less than 1000 backers, creating noticeable skewness. Since the Median better represents skewed data, we believe that to be a better representation of the data.

As for the Variance and Standard Deviation, we found that it was a fair amount larger for Successful Projects then Failed ones. Since there are a higher number of Successful Projects listed in the Number of Backer sheet AND that the range from Minimum to Maximum is slightly higher for Successful Projects, we find that this checks out