# Module 1 Challenge Report

### Angie Foust – November 11, 2023

## Observations & Conclusions

After working with this data several observations jumped out at me. First, the theatre category had the largest number of projects, so the graph on the Count by Cat tab could easily be misread and lead a viewer to think theatre projects also had the highest percentage of successful outcomes. However, the technology category actually had the highest percentage of successful outcomes. The highest percentage of failed outcomes went to the games category. While I feel that the categories are a limitation (further discussion below), it is quite feasible for someone to be deciding between technology and games, which is actual mobile and video games. This data could push them away from the games category, which is likely highly saturated.

When looking at the data by month, the number of successful projects begins to rise in May and peaks in July. During that same time frame, the number of failed projects dipped and then begins a slow rise into July. Ultimately June had the highest difference between successful vs. failed/canceled projects, while August had the smallest. Based on this observed trend, it could be recommended that a project begin crowdfunding in May.

Projects with goals between 15,000 and 35,000 had a higher average percentage (95%) of successful outcomes in general, compared to project goals under 15,000 (60%) and over 35,000 (64%). The projects with the highest goals (over 50,000) had the highest percentage of failed/canceled projects. This leads to the conclusion that project goals are likely to be the most successful when set between 15,000 and 35,000.

## Limitations of the Data

This data set has several limitations. First, only 7 countries are represented and the currencies are different. The exchange rates between those countries are likely different as well, and may not give an accurate comparison. If someone is looking at a global project, the data may not provide as broad of a view as is needed. Conversely, if someone is looking at project that will be local (within country) only, the rest of the data may be obsolete.

This data also seems to be heavily in the fine arts industry (film/video, music, photography, theatre) and may not be a good representation of other project areas. In addition, some areas, in particular Journalism, were not well represented. If the company/individual is in need of a true comparison between different project areas/industries, this data is not going to give an accurate picture.

## Other Tables/Graphs Recommended

As part of my analysis, I found it useful to look at the percentage of successful vs. failed outcomes by category. The data as written only provided a count, so I created an additional table showing percentages of each outcome by category. It would also be interesting to see more data pulled in to counter the limitations I mentioned- additional countries, converting the currencies to equivalent amounts for comparison, a larger representation of different industries, and more equitable representation (i.e. a similar number of projects for each parent category).

## Statistical Analysis Questions

Because there is quite a large range in the number of backers for both outcomes (successful and failed), the projects with a high number of backers skew the mean. In this case, 70% and 71% of the data in the failed and successful categories respectively fall below the mean. Since the median is truly in the middle, it is a better representation than the mean.

There is more variance in the successful outcomes. This makes sense as the minimum number of backers in the successful campaigns was 16, and 26 of the failed projects had backers in the single digits; two campaigns had 0 backers.