**Crowdfunding Data Analysis**

**Conclusions**

Analysis of the crowdfunding data provided leads us to the following conclusions: (1) The proportion of successful to failed to canceled projects does not vary particularly widely across parent categories except for journalism projects, which showed a 100% success rate. This, of course, may be due to the relatively small showing of journalism projects in the data set. Games appear to show the highest failure rate, while technology appears to show the highest success rate next to journalism. It appears that the only parent categories with a success rate less than 50% are food, games and publishing. Interestingly parent categories representing media projects, namely theater, music and film & video are drastically overrepresented in comparison to other parent categories. (2) When considering sub-categories, it seems that science fiction films and mobile games have by far the highest failure rates, with the only other sub-categories showing success rates less than 50% are video games and food trucks. (3) There appears to be a fairly consistent month-to-month success-to-failure ratio when considering time series data, with a clear spike in successes at the end of calendar year Q2 / beginning of calendar year Q3, which may indicate a seasonal trend of increased cash flow for investors corresponding with that spike.

**Limitations and Additional Direction**

The primary limitation of this dataset is that it is unclear where these data originated. For a true comparison of categories and sub-categories across regions and over time, we would need to source reliable data from a well diversified sampling of crowdfunding platforms. Without knowing variations in investor demographics and platform mechanics, it is impossible to draw generalizable conclusions from these data.

In the context of the summaries presented in the accompanying Excel workbook, it would be prudent to conduct further graphical analysis. The present discussion relied upon stacked-to-100 bar charts to compare proportional success and failure rates across categories, which is immeasurably more useful than the visualizations shown by default. Furthermore, the present study did not consider geographical trends because it would have required a much more robust analysis, likely with map-based visualizations in order to draw useful conclusions. Additionally, the seasonality trends discussed above should be further explored by visualizing not only aggregated month-over-month data, but rather by considering such trends over a number of years in which the crowdfunding platform has seen similar volumes of activity.

**Statistical Analysis**

The summary statistics table presented in the accompanying Excel workbook indicates that backer counts for both successful and failed campaigns are heavily right-skewed, with successful campaigns showing a greater number of projects that acquire large numbers of investors, which pulls its mean backer count farther to the right of its median backer count than we see for failed campaigns. In general, the media is a better measure of central tendency for skewed data, because it presents a more accurate picture of where the bulk of the population falls, without being overly influenced by extreme outliers. The higher variability seen in successful campaigns, as suggested above, is to be expected because of the attention that projects widely considered to be a good idea can achieve through viral word of mouth and social media mechanisms. Therefore, it is not surprising that more successful campaigns have huge numbers of backers in comparison to failed campaigns.