Kickstarter Sample Data Analysis

# Executive Summary

The given Kickstarter data suggest certain conclusions about factors influencing a successful Kickstarter project. However, there are limitations in the data set that suggest validation of conclusions would be recommended with a more robust sample.

The conclusions drawn are:

* Projects that achieve “Spotlight” status are highly successful
* Month of project launch does not appear to affect chances of success, suggesting that seasonal timing confers no strategic advantage
* A project funding goal of $1,000-$5,000 is more likely to be successful, and chances of success drop off as the funding goal rises

Limitations of Data Set

* Success rate in sample is significantly higher than known population
* Geographic data are limited to country level
* Theater projects are highly over-represented

# Discussion

## Conclusions

1. Spotlight corresponds 1 to 1 with successful projects. There are no successful projects that haven’t been in the spotlight, and all projects that have been in the spotlight are also successful. This points to a strong relationship between “Spotlight” projects, and the likelihood of project success.
2. Seasonality shows tendency for the total count of projects to ramp up in the middle of the year, and then drop off towards the end of the year, but the month of campaign launch does not appear to affect chances of success.
3. Projects aiming to raise less money are more likely to succeed than projects with a high goal. This stands to reason, as it is easier to raise a small amount of money than a more ambitious target. Further, the goals tend to be lower, with about half (1,865) of projects aiming to raise less than $5,000. By way of comparison, 1,483 projects had goals of $10,000 or more.

## Data Set Limitations:

There are a few reasons to question the representativeness of the data set, relative to the entire population of kickstarter projects.

1. More detailed geographical information is needed to identify any regional differences, and/or whether location influences project success.
2. The data set shows an overstated success rate, 53% successful projects in sample, while I expected closer to 33%, as this is the rate in the known population.
3. The single largest category in the approximately 4,000 project sample is Theater with almost 1,400 projects (35%), dwarfing any other single category. Unless it is the case that Kickstarter is particularly focused on theater, the sample is overly influenced by Theater projects.

With this in mind, it is reasonable to validate the conclusions against a more robust sample. Possible improvements could be:

* A larger sample size, or
* A more random sampling technique.

## Additional Tables or Graphs

I found it helpful to visualize the distribution of donor counts broken out by successful compared to failed projects. These both show a heavily skewed distribution, with over 95% of backers per project within one Standard Deviation of the Minimum value.

Based on this finding, I think further analysis should focus on the lower end of the distribution, compared to the total sample.

Secondly, it would be instructive to look at the success rate as a percentage of the total count of projects. This ratio analysis would equalize the differences in counts across categories.