Crowdfunding Data Report

Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

1. The most popular month for starting crowdfunding campaigns is July, followed shortly by January. July however has far more successful campaigns than January.
2. June is the best month for starting a successful campaign relative to the total number of campaigns started in that month.
3. The Theater category is the most popular category by a large margin.

What are some limitations of this dataset?

The dataset doesn’t contain any information pertaining to when successful campaigns hit their 100% funded goal, which would be useful for predicting if live campaigns are on pace to meet their funding goal before a deadline compared to the mean or median successful campaign. Additionally, this set has a very large amount of theater, film, and music based campaigns, so any conclusions drawn from the data as a whole may be biased towards campaigns from those categories, which is potentially not applicable to campaigns in other categories.

What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

The most relevant measures of success when it comes analyzing this data would be the outcome of the campaign, and the related metric of percent funded. So-far in this exercise we have analyzed date created and category/sub-category in relation to outcome of the campaign. With that in mind we could also create graphs and tables that incorporate:

* Average Donation
* Backer Count
* Funding Goal
* Staff Pick
* Spotlight

I would be interested in seeing those categories compared to campaign outcome where possible, but percent funded would function better for most of the above statistics. For Staff Pick and Spotlight, since those are fields with a TRUE/FALSE flag, we could make a stacked bar graph to compare outcome counts by TRUE or FALSE. For Average Donation, Backer Count, and Funding Goal, a scatter plot with percent funded on the vertical axis and including a line of best fit makes the most sense for visually presenting that data to find trends.