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

Theater, specifically plays, is the largest category of crowdfunding campaigns, four times as large as the next sub-category.

Summer, especially June and July, is the best time to have a successful campaign, but August is bad. Maybe there’s an end-of-summer effect that causes people to give less to crowdfunding campaigns?

Set a crowdfunding campaign in the range of $15,000-$35,000. Campaigns in the $15,000-$35,000 range tend to be more successful, verging on 100% success rates, but even campaigns in the $35,000-$50,000 range have a ~75% success rate. Campaigns in the $5,000-$15,000 range fare 50-50, a roughly even chance of success or failure.

Most campaigns tend to have smaller numbers of backers, in the low hundreds range, so a crowdfunding campaign doesn’t have to reach thousands of donors to be successful. A few hundred is probably sufficient.

* What are some limitations of this dataset?
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
* Does the mean or the median better summarize the data? Why?

I think the median better summarizes the data, although they both provide useful information. The data is right-skewed, so the mean (average) has a significantly higher value (successful 851, failed 586) than the median (successful 201, failed 115). I checked the modes as well (successful 85, failed 1). Given all of that, I think you can pretty safely say that a few massive outliers (maxs of 7295 and 6080) are pulling the data right-wards.

In sum, most campaigns tend to have smaller numbers of backers, with success around 200 and failure around 100. The mean is skewed by outliers.

* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

There’s more variation with the successful campaigns, which makes sense. The variance for failed campaigns is roughly 45% less than the value for successful campaigns. You would expect failed campaigns to have less support, and therefore less variability. More importantly, there are more successful campaigns than unsuccessful campaigns, so you’d expect there to be more total variability, as there are simply more data points with successful campaigns.