Reflection Questions Excel

1. Given the data provided, what are three conclusions we can draw about Kickstarter campaigns?

First, film and theater seem to be best fitted to succeed in this environment, and journalism seems least fitted to the point of having nothing but canceled campaigns - though it doesn’t have to do with written word projects not succeeding because publishing has a fair showing. It’s also not true of all entertainment. Another conclusion is that failed projects seem to often come very close to the count of backers that successful projects do, so the determining factor must be average value of donation and/or how many upper outlier backers each state category has. A last conclusion is that on the whole successful campaigns last longer than unsuccessful ones, as evidenced by the extra pivot table and chart I added comparing the duration of campaigns with their goals, filtered by starting year. That speaks for itself – it’s a natural assumption to make even not knowing.

2. What are some limitations of this dataset?

As implied above, it would be useful to compare average donation against the upper outliers of donations and normalize somehow for goal. A few other columns aren’t as manipulated as they could be either, for example I calculated duration of campaign on my own as that’s a more useful metric than starting or ending date alone. It’s also a difficult thing to compare any monetary values because the currency differs widely – there should be a column next to each column of monetary values showing conversion to a common denomination, say US dollars.

3.What are some other possible tables and/or graphs that we could create?

As I did on my final sheet, running and comparing mean/median of the goal dollar amounts is interesting. You might compare the mean/median duration of each campaign, or get those of the average donation per each status. A pivot chart that is filtered based on Year, with an x axis (rows) of campaign goal amount and a y axis (values) of duration of campaign in days and line names (columns) of state of campaign, could be interesting to show how duration and goal amount interact to affect outcome. Filtering by year adds a dimension of trend to that analysis.

Bonus Stats Analysis

1. Does the mean or the median summarize the data more meaningfully?

Here the variability is pretty wild, and the median for both categories is much lower than the mean. There are definitely some upper outliers, some very high counts skewing the numbers, so I choose the median.

2. Is there more variability in successful or unsuccessful campaigns? Does this make sense – why or why not?

There’s much more variability in successful campaigns than unsuccessful ones, as evidenced by the comparative variances of the number of backers (successful campaigns have 188 times the variance of failed ones). In addition, the minimums for each category are almost equal, but the maximums vary widely, showing a greater range of backer count in successful campaigns. A campaign can be entirely funded by one wealthy backer, or funded in a mass of 26,457 small donations. An unsuccessful campaign probably has not found any wealthy backers yet, because once someone wealthy gets involved they tend to make the goal happen one way or another. Collecting tiny sums of money would seem to have a threshold too – at a certain point the count is high enough that you’re likely to get close to your goal. To test that explanation, I ran the average of the goals of successful vs unsuccessful campaigns. Successful campaigns average roughly 20% of the goal of unsuccessful ones, roughly half taking the median. There is more variability in how to compose a lower number, so this supports my theory.