Juan Coronado - Questions

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

* Overall, theater campaigns are the first place in terms of campaigns started. There was a total of 344 (out of 1,000) for this Category, which represents 34% of total campaigns. This is almost double than either the second (film & video) or third place (music), which represent 18% each. This means people have a lot of interest for Theater campaigns.
* From all of the campaigns, more than half (56.5%) were successful, followed by failed at 36.4%, 5.7% at canceled, and 1.4% currently live. In terms of percentage, it seems that Television and Photography books are the sub-categories that fail the most, at 18% and 10% respectively. These should be some of the riskiest campaigns to get involved in.
* January (91) and July (93) seem to be the month in which more campaigns get started. However, this doesn’t necessarily mean they are the most successful. In June the successful rate is 64% for the first place, and both July and September tied with 62% for second and third place. It would be wise to participate in June for a higher % for this campaign to be successful.

What are some limitations of this dataset?

* We only have data at campaign level. Perhaps having higher granularity in different tables could be helpful, such as:
  + Backers’ information – demographic info about who is backing up, by which specific amount, how did they hear about it, etc. We created an avg donation, but perhaps 20% of backers have 80% of amount pledged, for example, or we could see which countries pledge more.
  + Transactions information – date in which a transaction occurred, amount, method, source, etc. We might be able to uncover hidden helpful patterns, such which source is used the most, at what time should publicity be made, etc.

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

* Same tables as we created in the example but including the % per row. This lets you see not just in terms of pure #s, but also %s which can be important.
* Statistical multi variable analysis to analyze which variables are predictors that a campaign will most likely be successful or not.
* If there was a way to connect with external data, possibly with what social media and/or other websites comment about the campaign and do sentiment analysis with this could be interesting to analyze what went good/wrong.