The dataset on crowdfunding shows that about half of campaigns are successful and this rate doesn’t vary much by country, the most popular category of crowdfunding project is theater, and for projects that get funded, most get much more than their goal—the average funding amount is over 300% of the target for successful crowdfunding projects.

This dataset doesn’t give information on backers themselves which limits how deep we can go in the analysis. It’s possible that a dataset that gave information on individual backers could be used to much greater affect. For example, we could see how the pledge amount per backer varies and whether it is skewed positively. This could enable people pursuing crowdfunding to determine if there’s a subset of backer they should try to target to maximize their chance of success.

Another limitation of this dataset is that the subcategories do not capture enough detail. The most popular category of project is “theater,” and one might want to dig into that to understand if there are categories within “theater” that receive more or less funding or are more likely to succeed. However, there is only one subcategory of “theater,” “plays,” which means we gain no further insight into this category by diving to the deeper grain. It would be helpful if this category was better defined. Knowing that every theater project is a play, we should define more subcategories for plays that are dramas, comedies, tragedies, etc.

Another useful graph would be a line graph that tracks success rate over time. The current line graph only allows us to identify trends in success rate in the calendar year, but it’s possible that the success rate has changed over the years and a graph of success rate over time would tell us that. Further, a table that shows success rate by country could show us if there is variation in the probability of success by country. If we then filtered by category we could see if there are preferences that vary by country based on the probability of success.