*Analysis*

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

Between the years of 2009 – 2017, the most popular subcategory of Kickstarter projects were plays, with a total of 1066 attempted, 65% of which were successful. Plays are included in the Theater parent category, and the 4 parent categories with the most projects are, in order, Theater, Music, Technology, Film & Video. This data applies both when all countries are included in the calculation and also when looking at the United States alone.

A somewhat alarming statistic for the years above is that 37% (1530/4114) of all attempted projects failed. If we include the number that were canceled with the number that failed, the percentage goes up to 46% ([349+1530]/4114), which is nearly half of the attempted projects.

If we look instead at successes, the subcategories for which 100% of the projects were successful include: classical music, documentary, electronic music, hardware, metal, non-fiction, pop, radio & podcasts, rock, shorts, tabletop games, and television. Five of those twelve subcategories are in the parent category of music. Music overall has a 77% success rate (540/700), the highest of any parent category.

Also, with regard to ensuring success, from a quick sort of the data based on backer-count, there seems to be a correlation between the number of backers and the chance of success.

One other interesting thing from the bonus question was to note how the percentage chance of success decreases as goal numbers increase, indicating that the more money a project is trying to raise, the less chance they have of succeeding.

1. What are some limitations of this dataset?

Because the dataset is not current, the “live” category isn’t very meaningful.

I searched the meaning of the columns “staff pick” and “spotlight” to decide whether those are worth analyzing. There seems to be some concern about which projects get the badge of “staff pick”; it sounds very subjective. Because of that, I wouldn’t necessarily spend time analyzing that column with success/fail data.

I was wondering about the accuracy of this data and how it was collected. When I searched that, I found they have a scraper robot which crawls all Kickstarter projects to collect data. They mentioned there could be duplication of data if projects are listed in multiple categories.

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

As I mentioned above in question 1, another table & graph combination that could be interesting would be one which shows the relationship between backer-count and outcome. The theory would be that the projects that have more backers would more likely be successful.

It might also be interesting to subtract the date created from the date ended, calculating the number of days. I’m thinking this would show how many days it took for the project to reach it’s goal, and maybe we would see whether successful projects typically reached their goals in fewer days.