

Over $2 billion has been raised using the massively successful crowdfunding service, Kickstarter, but not every project has found success. Of the more than 300,000 projects launched on Kickstarter, only a third have made it through the funding process with a positive outcome.

Getting funded on Kickstarter requires meeting or exceeding the project's initial goal, so many organizations spend months looking through past projects in an attempt to discover some trick for finding success. For this week's homework, you will organize and analyze a database of 4,000 past projects in order to uncover any hidden trends.

People who back Kickstarter projects are offered tangible rewards or experiences in exchange for their pledges.

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   1. Just over half (53%) of the Kickstarter campaigns reached or exceeded their funding goal.
   2. The most common category within Kickstarter is Theater with 34% of all campaigns.
   3. The most successful category within Kickstarter is music with 77% of campaigns meeting or exceeding their funding goal.
2. What are some limitations of this dataset?
   1. We only have 4,000 of 300,000 projects. I.e. our sample represents only 1.33% of all projects launched in Kickstarter. We should confirm this is truly a random sample before spending a single minute with this dataset.
   2. We are not provided with a table of column definitions.
   3. Would be really beneficial to know the relative value of goods/services rewarded as a percentage of the amount pledged.
   4. Would be really beneficial to know which Kickstarter campaigns had promotional videos and some sort of rating on the production quality of those videos.
   5. Would be better if all data was converted to the same currency.
   6. We have based the bulk of our analysis on the ratio of the pledged amount to the goal, but each individual product/service selected their own goal. I.e. some of the goal values are way too small (ex. $1) or possibly too large (ex. $100M), so we should really filter out some of these products/services with unreasonable goals.
   7. Another data point that would be beneficial is the amount of marketing dollars and time spent on the Kickstarter campaigns.
3. What are some other possible tables and/or graphs that we could create?
   1. I would be interested to see if the “staff\_pick” or “spotlight” columns are correlated with meeting funding goals and I would want to see that based on category and subcategory. I would do this in a pivot table with a 100% stacked pivot chart to see percentages of each group. For example, given the “staff\_pick” is true, what is the probability of success. If the data suggests the staffers have a high success rate of picking winners within a given category or subcategory, this would be a very beneficial piece of information to share with a potential campaign launcher. I.e. have staffer’s take a look at the product/service before launch.
   2. I would also want to run some regression analysis to see which data point best explains the percent funded. I would definitely want to check goal, backers count and average donation.