The geographic distribution is concentrated in the US with 765 of 1,000 campaigns (76.5%).

The most common type of crowdfunding campaign is theater plays, occupying more than a third of the dataset (34.4%). This is followed by film and video (17.8%) and music (17.5%) campaigns. This may imply that entertainers are more aware of this platform as a funding source.

The least common crowdfunding campaign type is journalism audio, with only 0.04%, however all of these achieved their goal. This is an abnormally small number of campaigns, but it could indicate that backers are interested in supporting this category.

To improve the data, given that theater is the most popular category, it’s sub-categories could be separated from “play” into more specific ones such as: tragedy, comedy, history, musical, etc. This could help further understand the backers’ preferences.

The blurb column doesn’t provide a lot of information. This could be improved using tags of frequent terminology for the creators to further categorize the campaigns and see patterns in the user’s interests.

The engagement of the campaign could be useful. This can be measured on the backers’ side by tracking the amount of people that visited the campaign webpage and potentially the likes, and comments. On the creator’s side, how much feedback and updates they gave over the development of the campaign. These factors might play a role in the success rate.

With the current data I would like to add:

* A pie chart with the outcomes without categories, to represent the outcome rates in a more straight-forward way.
* The “staff pick” and “spotlight” features compared to the outcomes, to see how effective these are.
* The duration between creation and end of every campaign in relation to the success rate, to find which campaigns have a better chance of success.
* The data contains different currencies, these need to be equated to a specific currency (possibly USD) to better categorize the funding goals, pledges, and average donations.

We analyzed the number of backers versus the success rate. We found a big gap where the mean is greater than the median, in both the successful and failed campaigns. This indicates that the data is skewed to the right, with a few instances having too many backers to trust this information fully. For this reason, the median is a better representation of the data. The median tells us that in general, successful campaigns have 175.5% more backers.

There is a very high variability in the backers-success analysis, slightly higher for successful campaigns but this is negligeable considering the mean and median higher in a similar proportion. This means that variability This might be telling us an incomplete story, the analysis could go further by creating ranges to separate the campaigns in scale of backers.