Leo Wheeler

Data Analytics Bootcamp

16 Dec. 2024

Module 1 Challenge Report

The data used for this report was the crowdfunding dataset provided to me. This dataset contains a list of crowdfunding projects with their:

* Goal
* Amount raised
* Success/failure of project
* Number of backers
* And several more (less important) information

1. Categories
   1. From the graph in the “parentcat” sheet, it is clear that most of the projects for this are either theater, film & video, and music (referred to as ‘these categories’), with all other categories barely combining to equal music. These categories also had a much higher rate of success than the smaller ones. This must be taken with a little hesitation, given that more opportunities can create more success. In line with that thought, these categories have more failures than the others as well, so it may just be an issue of volume rather than category.
2. Sub-Categories
   1. When you jump to the ‘subcat’ sheet, it becomes immediately clear that plays are the number one crowdfunded project here. Plays seem to have a high failure rate, but when compared to the others, the gap between success and failure is larger for plays than a lot of other categories.
3. Success vs. Failure
   1. There are more successful projects than failed projects overall. This does not mean that any given project is more likely to succeed, rather that from this dataset, we have more successful projects to look at. Over time, the successes and failures seem to follow the same curve respective to each other, which could suggest that its more likely to succeed than fail, but I am not sold.
4. Data limitations/Suggestions
   1. Without an understanding of where the data comes from and what has been omitted, it can be difficult to draw any conclusions. There were also a couple of columns that we didn’t even look at for this project. Regressions and other such analysis could yield better results in this case than pivot tables alone. Running a regression on project success and categories, for example, could tell us a little bit about any correlation between these that might not be easily spotted.