Samantha McKay

Data Analytics Bootcamp

Crowdfunding Analysis

Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

1. Crowdfunding campaigns with a crowdfunding goal over $50,000 are less likely to succeed than campaigns with a goal less than $50,000.
2. Games and Food are the most likely categories to fail. Journalism and Technology are the most likely categories to succeed.
3. No successful projects were less than 100% funded, but some projects that met the goal for funding still failed. Success is highly correlated with meeting the funding goal.

What are some limitations of this dataset?

1. Each category is not represented equally, in terms of numbers of projects. For instance, journalism has four projects listed, which were all successful, so one could draw the conclusion that journalism projects rarely fail. Whereas theater has a large dataset of 344 total projects listed so the outcomes drawn from the theater data are more likely to represent the realistic probability of success or failure in this category.
2. It would be valuable to know how long the projects took to complete. Projects could have cancelled or failed purely because they were taking too long.
3. I would like to find a way to indicate how realistic the funding goals were in relation to the actual costs of the project. The funding goal might not have been well founded, and that would factor into a project failing.
4. I think that all of these projects were digital, but if they are not digital, it would be helpful to be able to separate digital products from physical products that would be manufactured. Also, it would be helpful to separate goods from services.
5. I would like to the reach of a project. Is this product only useful in one industry or would it have implications across multiple industries. We could also understand the cost per unit of the crowdfunded product. It might be difficult to get a clear value for things like “Grass roots 24/7 attitude” which does not relate to any easily quantified metric.

What are some other possible tables and/or graphs that we could create and what additional value would they provide?

1. We could create a table comparing the average cost of a project that would be filtered by category and sub-category. This would give more context to the success or failure. This table could then be used to compare the average cost of a project to the percentages of outcomes obtained in that category or sub-category.
2. We could create a table showing the minimum number of backers needed for a successful project from each category or subcategory. If possible, it would be interesting to further separate the categories into entertainment, like books, theater and film and video, and needs like food and technology.

The median better summarizes the data. The difference between the mean and the median is over 500. The mean is being pulled higher by a few very high backer counts.

There is more variability with successful campaigns. This does not make sense because more variability would equate to more unpredictability. The maximum number of backers is lower for failed projects, but not as low as I expected. Logistically, fewer backers would indicate a higher probability of failure. It would be useful to know the percent of the goal that was contributed, on average, by each backer. A project could have many backers, but each contributed $1. If the goal was $100, the $1 contributed would be more significant than if the goal was $50,000. It is difficult to find a usable/reliable measure that would be indicative of backer buy in to the project.