Kickstarter Dataset Interpretation

The three most significant conclusions to me from this dataset are thus:

Performing arts are the most successful kickstarters by a longshot. Music has a staggering 77% success rate for kickstarters within the dataset. A little way behind in second place is the theater category with a 60% success rate. Rounding out the top three is film & video with a 58% success rate. These three categories are the only parent categories that have an above 50% success rate and they skew the entirety of the program to an above 50% success rate. The next closest behind them is photography which is down at 47%.

Culinary endeavors are the least successful category. (Not including journalism projects due to all of the journalism kickstarters being cancelled.) This mirrors the difficulty of the food industry amidst traditionally-started businesses as ~80% of restaurants close within five years according to a 2016 CNBC article. The 83% failure rate in kickstarter’s food category is tellingly close to that number with our dataset being around that timeframe as well.

Lastly, kickstarters have become less successful over time. In 2009, 64% of kickstarters succeeded. With the platform climbing to the height of its success rate in 2011 at almost 80%. This success rate dwindled down to 46% by 2015. Despite a slight rally to 50% in 2016 (and 2017’s data being inconclusive as there are still live kickstarters in this dataset), the popularity of the platform as a means of fundraising has drastically declined in later years.

There are some stark limitations to this dataset if it is being considered an evaluation of kickstarter’s viability to begin an endeavor.

The dataset gives no sense of creator’s reputations, resources, or skillsets. Kickstarter has often been used to provide known, capable professionals a means to pursue passion projects. These people typically have extensive experience in their fields, popular followings, and existing capital and connections that are vital to the success of their endeavor. These facets give donors confidence in the project. Confidence establishes actual value and will often be the largest factor in the success or failure of a project.

The dataset measures only if the kickstarter goal was reached, not if the final product was delivered. Funding is simply the first step of a project. An essential piece of context to the dataset is whether or not the goal that was established was an accurate and realistic goal to actually achieving the end result. Some projects that made their funding goal still failed to deliver their final product.

Third, the dataset gives very little sense of marketing work done on a project. Outside of a staff pick or kickstarter spotlight, there is no sense for social media or word-of-mouth awareness campaigns for a project. Even passionate fans of a creator will not donate if they are unaware of the project during its kickstarter campaign.

However, even within this dataset, there are still valuable stones to be turned.

A chart that shows the influence of staff picks and spotlights would be vital. Seeing the trends of a kickstarter having neither, either, or both of these advantages bestowed on them would be mandatory information for myself before I considered kickstarter an option for a project. Seeing the additional marketing is important for a methodology that is heavily awareness-based.

A graph showing the success rate in regard to average donation would be very useful. The presence of an outlier data point like a very large donation that could come from an existing friend or investor would heavily skew the success rate. While we can’t see individual donors, it can usually be inferred on these types of projects that high average numbers are not reflective of what people typically donate.

##Bonus Statistical Analysis Summaries

The means and medians of the data:

For the average kickstarter hopeful or the data-curious, the median is a much more valuable data point, as there were quite a few outlier projects that had goals that were magnitudes higher than the vast majority of projects. That said, if we were looking at the data with any sort of context, it could be very useful to break out the data into narrower ranges that more closely represent a prospective project that is considering kickstarter as an option. In these smaller arrays, the mean would be a much more useful number.

The variance of successful vs unsuccessful campaigns:

The variance is comfortably greater amongst successful campaigns. This makes perfect sense as the projects ranged from a few hundred dollars to tens of thousands. If the pool only takes into account the successful ones, it will include ones that made the very high values and required many backers. Certain projects could succeed with a single donor and others took many thousands. To contrast, if the pool only contains the unsuccessful ones, it will, by nature, trend towards lesser numbers. If all data points are skewed nebulously “lower” with a minimum value of zero, the variance will be compressed towards zero.