Excel Challenge Reflection

1. Dataset Conclusions

Throughout this exercise, there were three different sets of graphs and tables created, each offering its own conclusions that can be taken from the data set. The first set takes a look at the number of successful, live, cancelled and failed kick-starters by their parent category. The second looks at the same information, except it tallies sub-category instead of parent category. The third shows how successful kick-starters are by the month that they were started in.

Looking at the first dataset, the thing that stands out the most is that theater-based kick-starters are the most popular programs seeking public funding, with over a quarter of the programs being theater-based. This also looks to be a more successful category then most of its other counterparts, which could be the cause of the popularity. Another successful category appears to be music, especially when looking at the categories of film & video and technology, which have a similar popularity. This data set also is able to be filtered by country, which shows that the vast majority (89%) of programs are based out of the United States and Great Britain. Based on this information, theater-based programs in the US and Great Britain are the most likely to seek public funding, as well as being able to receive a high success rate. However, more information is needed to see why it is more popular even though music-based programs tend to be more successful.

The second data set shows very similar information to the first, except it focuses on the programs sub-category rather than the parent category. The number of sub-categories is much larger than the number of parent categories, at forty-one to eight, respectively. The only thing this table really shows is that plays are the most popular type of sub-category, which makes sense because it is the most popular sub-category in the most popular parent category. Even digging a little deeper, the data set shows that some sub-categories only succeeded while others only failed. This does not give us enough information to give us any kind of accurate conclusion.

The third set of tables and graphs looks at a different perspective of original data set, as it looks at success rates by the month a kick-starter program was created. For the most part, the data says that there is no significant change when looking, in general, at the months a program was created. However, when filtering by different years, there are some noticeable trends that are worth exploring. One such trend is that there is an explosion of kick-starter programs in 2014, especially between March and May of that year. This dataset gives ideas of where to start looking when wanting to find reasons why kick-starters became popular.

Overall, these three graphs gave us some good information about the original dataset that we can draw conclusions from. The first created dataset shows that kick-starter programs tend to be from the United States or Great Britain and tend to be theater-based programs. The second dataset shows that sub-category is not a particularly helpful tool when looking for information about kick-starter programs. The third dataset helps find useful trends in the popularity of kick-starter programs.

2. Limitations

One of the major limitations of this dataset as a whole is that it includes multiple countries and multiple currencies. Without converting each currency to a standard, there is a lot of information missed out on, such as average goals of, or amount pledged to, programs by category. The main limitation, however, may be that this data set gives more questions than it answers, at best it gives us a place to start to look for the answers to the questions it creates.

3. Other Graphs & Tables

There are many different perspectives that we can take while looking at this entire dataset. While this exercise primarily focused on the success of programs by category and sub-category, it did start in on another way to look at the data by looking at the months kick-starter programs were created. We could also look at the length of the programs by category to find out how long programs tend to run on kick-starter. We could look at the number of backers to each type of category to find out which categories are most popular to the public, or the percentage each backer donated related to the goal of the program filtered by category to see if some programs are heavily backed by a few, or backed by many.

4. Statistical Analysis

When comparing the central tendencies of the backer counts of successful and failed kick-starters, a few things stand out immediately. The first is that there is a wide range of data points for both sets of data. Successful programs run from one backer to over twenty-six thousand, compared to failed programs running from no backers to nearly thirteen thousand. Because of the range in data points, it leads the standard deviations and variances of the data sets to be quite large, leading to the conclusion that median is a much better way to analyze this data sets as opposed to the mean as there is likely to be a large number of outliers for both sets of data.

When looking at the variance of failed and successful campaigns, it is quite obvious that there is much more variance in successful programs compared to failed, with a respective variance of 712841 to 3773. This makes sense due to the much wider range of backers of successful campaigns, but, as well as found in previous analysis, there is a larger number of successful campaigns overall when compared to failed programs leading to a greater chance of variance in the campaigns.