01 – Excel Homework: Kickstart My Chart

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?

A broad conclusion we could draw from the data is that campaigns with lower goals, tend to have a much higher success rate than those with higher goals. Campaigns with a goal of $15,000 or lower, succeed at a rate of 60%. Conversely, campaigns with a goal of over $15,000 fail at a rate of 50%. Going even further, campaigns on the very high end (over $100k), succeed at a rate of only 13.7% while failing or being cancelled at a rate of 85%.

The first thing that jumped out to me when looking at the data was that campaigns in the arts (music, film, theater) not only comprise the majority of campaigns (63.5%), but they get funded at a rate 2.5x of all other categories (65% to 27%). This excludes the publishing/radio category as that seems to be an outlier in that it has very few campaigns, but 100% of them have been funded. Nonetheless, I would be more inclined to include this category in the arts than not. This would drive those numbers up even higher. Also, plays comprise the highest percentage of total campaigns by a huge margin. They encompass nearly 26% of all campaigns on Kickstarter. There is no other sub-category that covers more than 6%. Plays also are successful at a rate of 65%, which is high for such a large category.

The final conclusion I drew from the data is that technology has the highest number of backers than any other category. They cover 37.7% of all backers on Kickstarter. Within that category, the vast majority are for hardware technology. Of the technology backers, almost 77% are for hardware items. Also, the technology backers average donation is higher than anyone else at about $133.

While plays encompass the largest amount of campaigns and get funded at the highest rates, technology campaigns seem to appeal to the most people. There are almost 3x as many backers for technology campaigns than plays, and they donate at a rate about 2x as much.

1. What are some limitations of this dataset?

The first limitation I noticed was there is no information about cancelled campaigns. There is probably much you could learn from knowing why a campaign was cancelled. There is a pretty substantial number of cancelled campaigns. Knowing what caused them could help prevent that in the future.

The data is also not very current. The most recent dates are from May of 2017. I believe that Kickstarter has grown a considerable amount since then. The category breakdown could have changed over time as well as the total number of campaigns per year. The number of campaigns seems to have risen year over year, most significantly from 2013 to 2014.

Also, the ‘blurb’ data could be helpful to see a more granular view below the sub-category. However, this data is all written as free form text which makes it almost impossible to gather anything from it. Having that data could help provide insight into which areas of each category do better than others. For example, Technology Hardware is an extremely broad category/sub-category. It could be helpful to know what kinds of specific hardware seems to be the most popular.

1. What are some other possible tables and/or graphs that we could create?

There were a few other tables and graphs I created for my analysis along with the ones I already had made. The first was a pivot table showing the number of backers and the average donation for each category and sub-category. This allowed me to see which categories had the most backers as well as the highest average donations. This also would be easy to see in a stacked column graph as well.

I also added a column with either true or false if the goal was greater than or equal to $15k. I used that to create a pivot table that showed the success and failures for campaigns based on the total goal amount.

BONUS

For both the successful and failed campaigns, the mean summarizes the data better than the median. The median value for each seems to be very low. Even removing the outlier amounts at the top still gives an average closer to the mean than the median.

There is way more variance in successful campaigns rather than failed ones. This makes sense to me because in failed campaigns, the backer numbers will almost always be lower, that’s why they fail. About 72% of all failed campaigns have 10 or fewer backers and almost 19% have no backers at all. There will always be a significant portion of failed campaigns that receive little or no backers and by the nature failing the rest will have lower numbers of backers as well. This means the variance in failed campaigns would likely always be less than in successful ones.