Ross Teare

12.12.2020

Assignment #1

With the data provided from Kickstarter, there are three takeaways from the data:

1. Out of the nine categories to invest into, "music" is overwhelmingly the most successful, with 77 percent of those invested in the category seeing success. On the reverse, the investment category that saw the least amount of success is "food", with 70 percent of the "food" category seeing failure. The most funded category is "theater", which received over a third of the funded in the data set at 34 percent of the funded.
   1. Looking at the subcategories, the information confirms that the music category is overwhelmingly the most successful, seeing multiple of the subcategories with 100 percent success. However, the highest number of campaigns for a subcategory is "plays", having 26 percent of the breakdown of subcategories with second-highest coming in at 6 percent, rock.
2. When looking at the trends of the launch dates, the projects that failed and are successful run analogous to each other throughout the year except for around the holidays and start of summer. During the holidays, you see the failed projects go up compared to successes. In May, you see the successful projects hit their peak throughout the year during the start of summer.
3. The biggest takeaway that you can decipher from the data is it Kickstarter campaigns that had a fundraising goal set for a lower amount had a higher percentage of success. There was a 71% success rate for Kickstart campaigns with a fundraising goal of less than $1,000 and a 66% success rate for a fundraising goal between $1,000 to $4,999. As for moderate fundraising goals, those between $5,000 and $44,999, the successful vs. failed campaigns stayed about even; however, once the fundraising goal reached above $45,000, there is a drastic decline in successful campaigns and an increase in both failed and canceled campaigns. There was only a 29% success for fundraising goals between $45,000 to $49,999 and a 19% success for campaign goals of greater than $50,000, and a 23% canceled campaign rate. Concluding from this, you can see that the more modest fundraising goals saw a greater chance of success than those that asked for high amounts.

There are many limitations to the dataset. This is just a fraction of the data from Kickstarter and shows more of a success rate than reported. With the limited data provided, there is no way of knowing if this is an accurate and scientific sample of Kickstarter campaigns, leading to skewed analysis and results. Additionally, there is no way to predict the margins of error in the statistical analysis due to this lack of information provided about the sample.

Many additional analyses could be done for this data set. For example, you can look at the success rate of those featured vs. not featured, depicted by a simple pie chart to compare featured vs. no featured. Similarly, you could see if there is relevance to the success rate if it is a staff feature vs. a non-staff feature campaign. Additionally, if there is a linear progression on the amount funded vs. backer count, I think this one would be fascinating to analyze.

Bonus:

Looking at the statistical data for the number of backers, the median is more useful to look at than the mean. There are many outliers for both successful and failed projects that are skewing the data. Additional support for using the median is looking at the Variance. Both successful and failed campaign variances were extremely high, meaning the data set is far from the mean and shows a large variation in the data. Due to this analysis, you can conclude a larger variability with successful campaigns vs. unsuccessful campaigns. Though both data sets have extreme outliers that are well outside the upper bound, the standard deviation and the variance for successful campaigns are much higher, with successful campaign variance = 712840.99 and standard deviation = 844.30 compared to failed campaigns variance = 3773.22 and standard deviation = 61.43. A data set with a higher standard deviation shows that the data is more spread out. This makes sense for failed campaigns that are more likely to have fewer backers due to their inability to receive their fundraising campaign's full goal amount. Where successful campaigns will have a large variance of the number of backers, they could have achieved their fundraising goals from one donner up to 26,457 (in this data set).