**Module 1 Challenge-Crowdfunding**

**Written Report**

After creating visualizations of the data provided, it is possible to draw several conclusions about how the success rate of crowdfunding campaigns is affected by category, subcategory, and date. When viewing the categorical breakdown of the data, it appears that theater, music, and film & video were the most popular areas, with theater at roughly the same rate as the 2 trailing categories combined. However, when leveraging the rate of success against the number of crowdfunding initiatives per category, theater crowdfunding initiatives were only successful about 54% of the time. This is the 3rd lowest success rate by category-and music and film & video are trailing directly behind. In fact, the least popular category (journalism) was successful 100% of the time-but only ran 4 campaigns.

When broken into sub-categories, the data becomes slightly less clear. While every theater category data point is located within the “plays” sub-category, the music and film & video categories each contain 6 sub-categories. When viewing this on a bar graph, it appears that plays were the most popular category by a wide margin. Every other sub-category is largely equal, with “rock” taking a slight lead-but the number of crowdfunding initiatives in that category does not equal even a third of the total for “plays”.

The line graph displaying the success, failure, and cancellation rates over the course of the year yields different insights. From this visualization, it appears that summer is the most successful season for crowdfunding initiatives-it also appears to be the most popular. The month of May starts a positive trend upward, with July being the most successful month of the year. Success rates appear to plummet in August, however-according to the data, this is the least successful month to launch a campaign.

While the data provides valuable insights into crowdfunding initiatives, it is not without limitations. For example, while the data addresses 9 categories, it seems largely skewed towards 2-3 areas. It would be helpful to have a more representative sample of different kinds of campaigns, especially in areas like journalism and technology where the success rates seemed relatively high. It would also be interesting to dive deeper into the process of obtaining funding-were certain categories more popular or successful due to the crowdsourcing method or the demographics of the backers involved? There seems to be some variability in the average donation rates, leading one to believe that certain campaigns may have sought or attracted wealthier backers.

There are a few additional analytical and visualization techniques that could help bring a fuller perspective to this data. For example-there are 2 columns we did not work with, Staff Pick and Spotlight. If these are methods to promote the initiative on the crowdfunding site, it would be interesting to see if applying these strategies yielded more successful results. It would also be enlightening to break the data down by country-perhaps different regions contribute more regularly or give more to individual campaigns. All in all, a larger and more diverse sample size would paint a clearer picture of crowdfunding initiatives.

**Statistical Analysis Statement**

Based on the box and whisker plots for both successful and failed crowdfunding initiatives, it appears that the median better summarizes the data in both categories. Most of the campaigns (both successful and failed) had less than 1,000 backers. However, the campaigns that had more than 1,000 often had many times more, as many as 7,295 successful and 6,080 failed. These outlier values skewed the rest of the data, making the mean a poor summarization of the information provided. The median of both categories was much more representative of the majority of the campaigns in the spreadsheet.

Successful campaigns appear to have more variability than failed campaigns. This makes sense for 2 reasons: first, there are more data points to compare with a successful outcome, giving it a wider array of potential data. Secondly, every successful outcome had more than a few backers-the minimum data point for this outcome was 16 but it is by no means the norm. Alternately, failed outcomes could have only a few backers-in fact, 30 different campaigns with that outcome had 10 or less backers. This keeps the window of possibility smaller, resulting in less variance.