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

Based off of the category pivot chart and table, the category with the highest chance of success on Kickstarter is Music. Out of the 700 Music Kickstarter campaigns in the dataset, 540 were successful in meeting or exceeding their funding goal. Within this category, the majority of the subcategories are very successful, with the exclusion of Faith, Jazz, and World Music.

The second most successful category was Theater. The Theater category also made up a huge portion of the total dataset at approximately 34% of the sample. Within this category, the most successful subcategory was Plays.

I found it incredibly interesting that two of the most successful categories featured artistic fields and specifically fields that heavily rely on live performance. As mentioned above, Plays were the most successful subcategory of Theater. Skimming through the dataset looking exclusively at the Music category, I noticed that some of the Kickstarter campaigns were for future tours or live shows. Preliminarily, it appears that a Kickstarter campaign that leads to an event a donor could attend may find a higher probability of success.

Conversely, some of the less successful categories were in fields that I would more generally expect on Kickstarter. For example, the Technology category had a very high rate of canceled campaigns. Looking at the subcategories, it appears that this is attributable to three subcategories: Space Exploration, Wearables and Web. Technology is also the category with the highest number of campaigns after the two most successful categories talked about above.

The last insight I had was that it appears December is a terrible month for Kickstarter campaigns in terms of finding success. It is the only month on our pivot chart in which the number of failed campaigns surpasses the number of successful campaigns.

1. What are some limitations of this dataset?

The first limitation I can think of is that Kickstarter campaigns have the ability to feature rewards for their backers. Rewards and reward levels are set by the campaign team and I assume can vary wildly based on the scope of the project and the desired outcomes. The dataset that we have does not feature any information about rewards or reward levels. I think that having this information could greatly improve the insights we’re able to achieve.

The dataset also skews very heavily towards the United States, so the conclusions that we draw might only be applicable to the United States.

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

In terms of analyzing this dataset to determine factors that could contribute to a Kickstarter campaign’s success, I would first want to look at how average donation vs. number of donors correlate to success rates. Which of the two has a stronger correlation? Would it be better to focus on gaining a large number of donors who contribute smaller amounts? Or would it be better to focus on getting larger donations from a smaller pool of donors? This could dramatically impact how a team approaches the campaign.

I would also like to see a line chart that tracks the success or failure of particular categories over the years. This would allow us to see trends in what categories are finding success on the Kickstarter platform.

Lastly, it appears that having a campaign featured in Kickstarter’s Spotlight might be a contributing factor towards success. It would be very beneficial to have a simple pivot chart or table showing whether this spotlight feature can contribute towards a project’s success.

**Bonus Analysis**

1. Use your data to determine whether the mean or the median summarizes the data more meaningfully.

In both instances, the data for successful and failed campaigns feature some pretty significant outliers. In particular, the data for successful campaigns has at least one incredibly large outlier. This means that the data is not going to follow a bell curve and would be quite skewed. Because of this, the median would be more meaningful when summarizing the data.

1. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

There is more variability with successful campaigns. The standard deviation, which represents the spread of the data, is much higher for successful campaigns than for failed ones. This makes sense due to the incredibly large outliers mentioned above. In this instance, success is simply any campaign that has met or exceeded its funding goal. The amount a campaign possibly exceeded the goal is not factored in, which could definitely cause more variability.