**Kickstart My Chart**

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

Not all categories are equal here.

Globally, there are more theater projects proposed than any other categories, and so they have the largest numbers of successful projects. Almost 1400 theater projects were on Kickstarter in our dataset. However, its success rate still trails other categories.

In terms of success rate, music has the best success rate by far at 77%. Still, we have to be wary of the genres of music, because even in this category there are many failed projects. Genres like faith, jazz, and world music do not have a single successful attempt in our entire dataset.

Lastly, there appears to be a relationship between success rate and seasonality. The lowest success rate appears to be in December across the 4,000 samples, where there were even more failed projects than successful projects. We still have to keep in mind, though, that the least amount of projects were released in December. The most successful number of projects peaked in May and the surrounding Summer months.

1. **What are some limitations of this dataset?**

One limitation is that we are purely focused on the bottom line here and not the level of time and effort that were put in each project. Anyone with a hobby can post a pet project on Kickstarter. There are many minimal effort projects that were thrown out there and ended up with no funding whatsoever. This could dilute the impact of that category. On the other hand, there are many projects where the proposing organization put much time and initial investment into the research and marketing materials. Perhaps they have promotional videos or prototypes of the product/project already. If I was running this research for a serious organization looking for their best entry into Kickstarter’s platform, I would try my best to separate these groups.

Another limitation I notice is separating “failed” and “cancelled” into different categories. From the eye of the proposing organization, the goal is to get successful funding. I do not know if they care if the projects are going to get cancelled or fail. They are, in many ways, effectively the same. It would be much easier to review success rate, rather than volume as we are doing now, if we only reviewed success vs. fail.

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

I would propose to review the relationship between goal amount and categories/sub-categories. This is because we are mainly conducting our research on volume right now. Music seems to have the best success rate. But what if that was because the average cost to produce a music album is far less than the cost to produce a film or to publish a book? Maybe the average ask is less and so it is easy to fulfill them? I would take a look at a Pivot Table reviewing the relationship between cost/goal and categories.

Another element we could look at is the relationship between categories and average donation. These categories cater to different groups of audiences. If there is a relationship here, it could influence our decision-making. It would be interesting to see if people are spending more on documentary films (largest sub-category of films) for projects they care about or on a technology product.

**Bonus Statistical Analysis**

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

While both calculations have their shortcomings here, I would argue that the median is a more effective way of summarizing this data. I arrived at this conclusion mainly due to the aggressive variance in the dataset.

For the successful campaigns, there are projects with only 1 backers and projects ranging up to 26,457 backers. With such a big swing in variability and distribution, the mean is impacted too much by outliers on both ends of the range.

When we look at the unsuccessful campaigns, the shortcoming of calculating the mean is even more apparent. There are projects that almost became successful but did not, where they amass hundreds of followers. But those projects are outliers. The reality of the situation is that most failed and cancelled projects have no followers. With a distribution that skews so much towards zero, mean would not be appropriate.

The median, on the other hand, paints a much more realistic picture of how many people are actually backing the campaigns, right in between the high visibility groups and the ones with lower numbers of backers. We have 62 backers in a typical successful campaign, and 3 backers in an unsuccessful campaign which is much closer to the mode as well.

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 in successful campaigns. This makes sense as successful campaigns tend to have outliers with high visibility projects. This means a greater range and a greater variance.

Unsuccessful campaigns, as discussed previously, tend to have either 0 or 1 backers so there is not a lot of variability there. The variability comes from large outliers that almost became successful. I chose to include cancelled projects in this category as well because the numbers are similar enough. Calculations with and without cancelled projects are also in line with each other. Also in the eye of the researcher, I believe success rate is the main concern, rather than whether the project is failed or cancelled.