Lovensky B. Lubin

April 8, 2021

“Kickstart My Chart” Findings

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

According to the data provided, the top three leading parent categories were: theatre, music, and technology. Of the campaigns that fell into the “theatre” parent category, the majority were under the “plays” sub-category (i.e. Sub-category count). “musical” and “spaces” theatre campaigns, were found in fewer numbers in this dataset. Under the “music” parent category, there was much more variability in sub-categories than the prior. Leading with 1) rock at 260 campaigns, 2) indie rock at 160, 3) faith and jazz: at 60. The jazz sub-category had a success rate of 0% (i.e. every campaign failed in reach their goal). Rock on the other hand, had a success rate of 100%. The “journalism” parent category (which had the fewest amount of campaigns), had the one sub-category: “audio”. 100% of which, were cancelled.

Of the campaigns that were successfully completed, a majority of them were created in May. In the month of December, there were fewer successful accounts created. Which dipped lower than the amount of those campaigns that failed, for that month. Campaigns that were canceled, consistently remained below a count of 50. With slight increases in months of July and November. The majority of theatre campaigns were created in May. While the majority of Music and Technology campaigns were released in January and November, respectively. Lastly, it also appears that the highest number of campaigns were created in 2015.

1. What are some limitations of this dataset? (\*)

Though the data may appear to suggest, that the “plays” sub-category are more popular than “musicals” and “spaces”. It’s important to remember that we are limited in what are data presents. Therefore, the limitation we find for “musicals” and “spaces” can potentially suggest that the formally mentioned sub-categories are in fact more popular and rarely require the need of a campaign. Which made me keenly aware that our results, cannot be applied to larger populations (other than we’re presented here).

Similarly, as we find that Journalism. The least represented parent category in our data. We could make the assumption that Journalism isn’t particularly cared for, by this data set. However, when drilling deeper we see that the sub-category is audio. Which in my opinion, could be fit the description of a podcast. After doing a quick google search, I found that there were several other potential sources to generate funding for those types of endeavors. Which could have led to the low representation of auditory journalism in this data.

In addition to this, it’s also important to note that this data does appear to be dated. With highest number of campaigns created in 2015 and steadily decrease as time goes on. Therefore, it’s difficult to determine whether any insights derived from this dataset can be applied to our current landscape.

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

We could potentially create charts showing how the category distribution changes by country. Which could help us identify what kind of campaigns, creators use Kickstarter for, in each country. I imagine this would be best visually represented in either a stacked bar chart or pie graph. With the conversion done to create both the beginning and end date for each campaign, I believe it would also be interesting to see whether the number of days by “state” change depending on campaign state. Using a histogram or box plot, we can see if there are any differences here.

Bonus Statistical Analysis

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

For those accounts that are successful, I would have to say that the median may more meaningfully reflect the data. Which is due to the potential evidence of outliers (i.e. 2 values over 20,000). I also took note that we also generated a fairly high variance and standard deviation. Another indicator of the possible influence of outliers, was the standard deviation being very far from the mean.

As for those campaigns that were unsuccessful, the mean may be a better indicator of central tendency. There is a more gradual progression in the data, however the better indicator of this would be the fact that the standard deviation is reasonably close to the mean.

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

Looking at the data, I would have to say that there appears to be more variability amongst the successful campaign data. Variance is described as the averaged squared difference from the mean. In comparing the two datasets, it’s evident that the successful campaign data returned a higher variance (of 713167.38). Which does make sense because when looking at the minimum (1) and maximum (26,457) number of backers. This range is much larger, than the unsuccessful campaign data (from 0 to 1,293).