[1]

**Analysis of Projects Launched on Kickstarter**

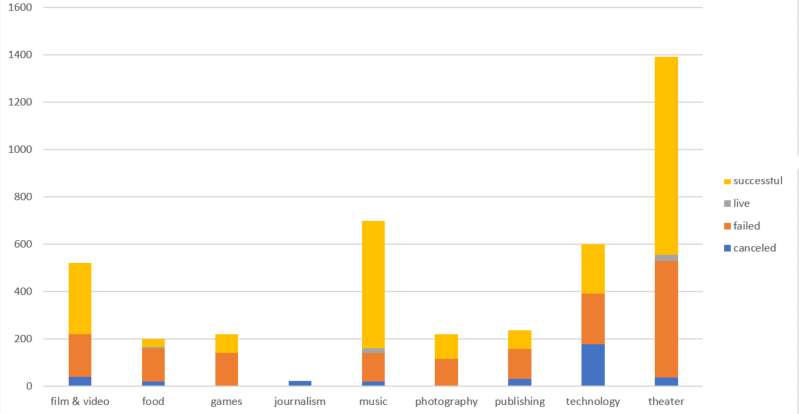
Susan Thomas

UNC-CH Data Analytics Boot Camp

**Conclusions about Kickstarter Campaigns**

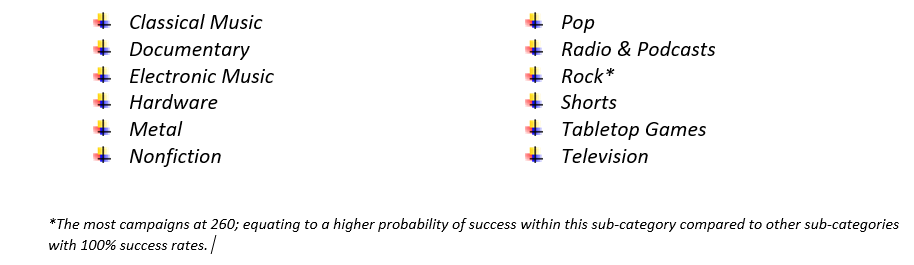
**Category Analysis:**

1. The *journalism* campaigns were all canceled. (100% cancellation rate)
2. Campaigns launched within the *theater* category were more prevalent than any other project with a total number of campaigns being 1,393.
3. Although the *music* category had the second highest amount of campaigns, this category had the most “successful” campaigns and is a more favorable category to submit a project and its outcome be successful.
4. There were few campaigns that were currently “live”; falling within the categories of *theater, music* and *food.* The highest “live” campaigns were within the *theater* and *music* categories which allowed those demonstrating their products to engage with the communities in real-time and potentially have more backers to their project.
5. The *food* campaigns may have lacked favorability and failed as they are not an easy project to reproduce and enjoy in multiple locations other than where produced. This category had a failure rate of 70% compared to the other categories.

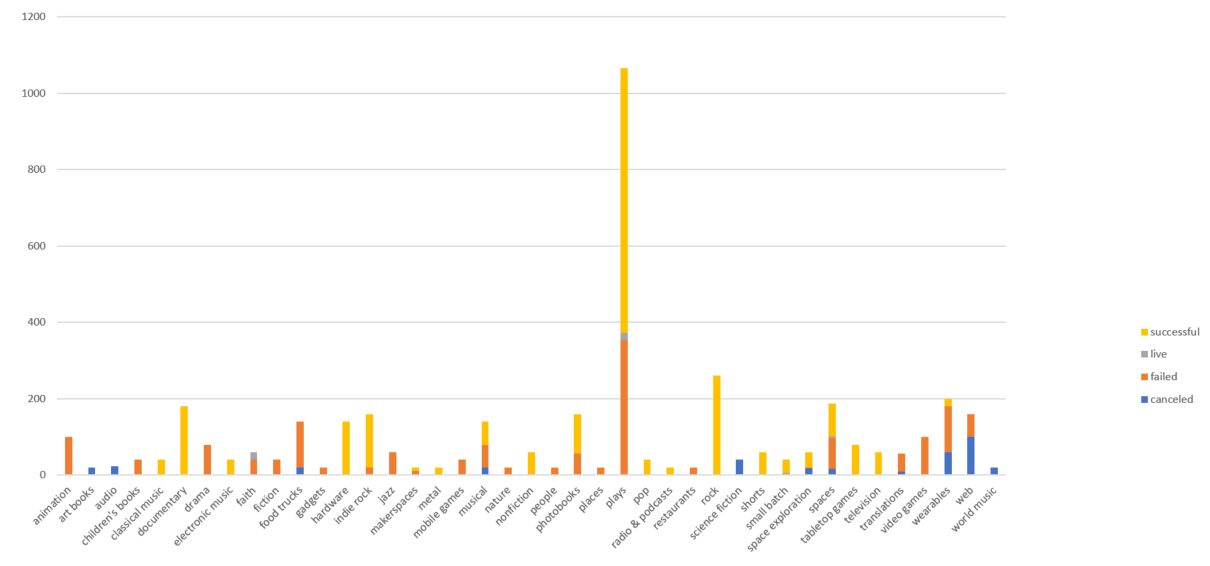
****

**Sub-Category Analysis:**

1. In congruence with its parent category, *plays* have the highest number of total campaigns. The *play* sub-category had the most “failed” campaigns overall but within the *play* sub-category, there were more “successful” campaigns than “failed” and “live” campaigns combined.
2. The following sub-categories have a 100% success rate:

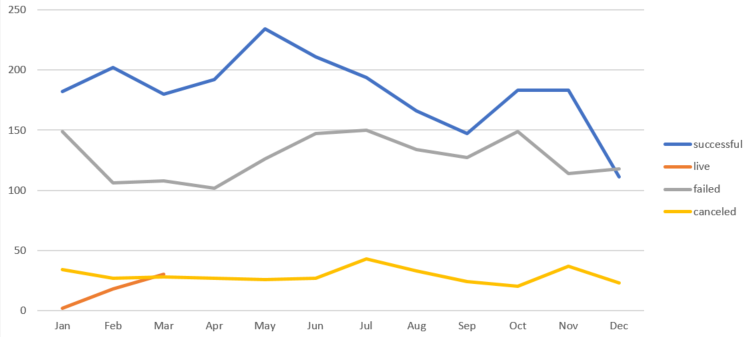
****

1. The sub-categories that had a 100% cancellation rate were *art books, audio, science fiction* and *world music*, with *science fiction* having the most campaigns “canceled”. This would suggest that the *science-fiction* subcategory would have the highest probability of being “canceled” if submitted on Kickstarter.
2. The *animation* sub-category not only had a 100% failure rate of campaigns but also had the most campaigns to fail among other sub-categories that also “failed”.
3. Within the parent category, *food*, which had the highest probability of failure, houses the sub-categories, *food trucks, restaurants*, and *small batch*. The *food trucks* campaigns either “failed” or were “canceled”, the *restaurant* campaigns all “failed” and the *small batch* campaigns were the only “successful” sub-category within the *food* parent category; allowing for 6 “live” campaigns and 34 “successful” campaigns, which contributed to the 17% success rate of *food* campaigns.

****

**Launch Date Outcomes Analysis:**

1. July had the most campaigns in total with 387 campaigns as well as the most “failed” and “canceled” campaigns.
2. May had the most “successful” campaigns with 234 campaigns.
3. December had the least amount of campaigns (252) as well as the lowest number of “successful” campaigns (111).
4. It can be concluded that if one wanted to have a higher likelihood of a successful campaign, they should launch them in May as July and December are far less favorable; potentially due to the holiday and summer season and the market already saturated with established brands during that time.

****

**Limitations of the Dataset**

1. The dataset may not accurately reflect or be able to assist in predicting the likelihood of success for a campaign as the Kickstarter website lists several categories and sub-categories that were not used to analyze.
2. There were current campaigns that were “live” and could’ve potentially been deemed as, “successful”, “failed”, or “canceled” if they were able to be seen through. The “live” campaigns included in the dataset appear to be outliers.
3. This dataset does not include more recent years of data (2018-present) where campaigns could be trending differently due to an increase in willingness to participate in crowdfunding or even to consider the push to support smaller businesses/start-ups.

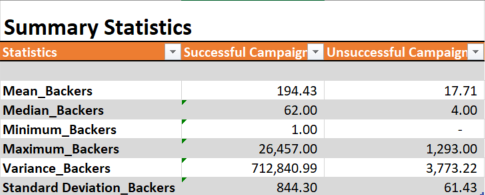
***Considerations:*** It would be interesting to compare this dataset with another crowdfunding company’s data to see if there are any similarities or statically significant differences. It would also be interesting to gain more insight or potentially explore a sub-category for reasoning behind backing a specific project to help predict the likelihood of a campaign’s success, failure or cancellation.

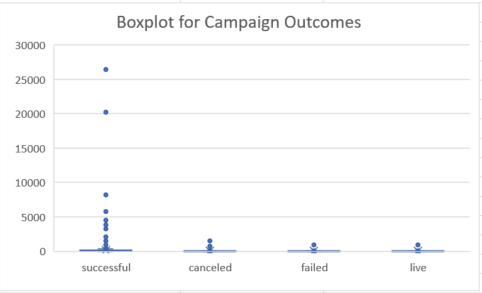
**Other Possible Tables/Graphs**

* This dataset seemed to be skewed more towards determining the success of the campaign based on the backers of the campaign as well as the campaign falling within a certain category. It would be good to create a table comparing the data based on the country of origin and understanding cultural trends around what could influence backers’ decisions.
* Calculating the summary statistics per parent category as well as sub-category to have a deeper understanding/prediction of how a campaign would perform as far as success or failure
* *Boxplot displaying the distribution of the data for the backers of both successful and failed campaigns \**
* A table that includes updated information about the status of the successful campaigns that reached the fundraising and backer goal
* A table or graph that dives deeper in to the time frame that it takes the campaign to be developed or if there is a limit that Kickstarter places on how long the campaign can raise funds; or to make amendments to the fundraising efforts

**BONUS continued on the next page >>**

**\*\*Bonus: Statistical Analysis**



****

Based on the Summary Statistics, I opted to create a boxplot to have a better visualization of which factor, mean or median, would summarize the data more meaningfully. I noticed that there was an extreme difference between the minimum and maximum amount of backers for the unsuccessful campaigns and the minimum and maximum amount of backers for the successful campaigns. The visualization shows outliers that have a large amount of backers for the successful campaigns which is greatly skewing the mean of the data. The remainder of the data points appear to cluster around the median. Due to this reason, the median should give a better summarization of the data than the mean. The outliers also impact the variance of the data as the successful campaign data is significantly varied in comparison with that of the unsuccessful campaign since it is skewed in the direction of the outliers.

[1] Image from Cover Page found at: <https://www.cleanpng.com/png-kickstarter-centauri-saga-crowdfunding-dale-of-mer-2357583/download-png.html>