Daniel Freda December 12, 2020

Excel-Challenge: Kickstart My Chart

1. **Given the provide data, what are three conclusions we can draw about Kickstarter campaigns?**
2. The greatest total number of project campaigns—represented within this dataset—fall within the parent category of theater with a grand total of 1,393 campaigns. On a more granular level, the sub-category of plays—within the theater category—accounts for 1,066 out of the respective 1,393 grand total. However, one should not presume a direct correlation between this relatively heightened frequency in total volume of campaigns with that of campaign success rate. For instance, the music category represents the second highest percentage of total campaigns with 700, but the greatest success rate (when excluding live projects). The related campaigns are significantly more likely to be successful with a success rate of approximately seventy-nine percent compared to the sixty-one percent success rate realized by theater campaigns. Therefore, there are variables (not represented in this dataset) relevant to the underlying nature of theater production—and likely industry itself, that causes campaigns within this category to be the most frequently occurring, but an organization is more likely to realize a positive outcome within the Kickstarter paradigm when launching a campaign for a music project.
3. Projects with goal amounts of less than $1,000 and $1,000 to $4,999 are the most likely to realize positive outcomes with seventy-one and sixty-six percent meeting and/or exceeding initial goal amounts, respectively. An analysis that increments goal amounts into approximated $5,000 intervals evidences that projects with goal amounts of $45,000 to $49,999 and greater than $50,000 see percentage successful rates of twenty-eight and nineteen percent, respectively. The percentage of successful campaigns declines gradually from $5,000 to $34,999 with a slight uptick within the range of $35,000 to $44,999. There is a steep decline in the likelihood of a project campaign satisfying the requirements necessary to be categorized as successful when greater than this amount. However, there is not enough information available within this dataset to posit any definitive conclusions regarding these results.
4. This dataset contains Kickstarter data ranging from the years 2009 through 2017. In this time, there has been a grand total of 4,064 campaigns. There was very significant increase in total campaigns beginning in 2014 (976 compared to 274 in the prior year). Total number of campaigns peaked in 2015 with 1,225 before declining to 950 in 2016. An analysis of outcome data focusing on months—regardless of year—evidences little variance in results; however, there is noticeable decline in December.
5. **What are some limitations of this dataset?**

An organization that desires to utilize this dataset as a means of predictive/descriptive analysis will find it inexhaustive and insufficient—most noticeably, in its deficiency regarding detailed information regarding specific and detailed information regarding the nature of the projects themselves: as well as, that relating to the individuals/groups that initiated said projects, and scope of the project. For example, based exclusively on the data provided in this dataset, it is not possible to conclude with a sufficient degree of certainty why some categories of project campaigns are more likely to succeed relative to others. Essentially, the dataset is not robust enough in certain regards—additional categorical and descriptive information would be relevant in formulating an extensive predictive analysis.

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

There are other potential graphs that may be useful in an analysis of this dataset:

1. A pie chart that visualizes the proportionality of associated project categories as a percentage of total project campaigns.
2. A scatter plot graph that demonstrates the correlation between backer count relative to campaign success and/or failure.
3. A box & whisker chart that separates a sorted data set of successful and/or failed campaigns into equal-sized fragments: such as quartiles or percentiles, while using the interquartile to determine outliers and the best measurement of central tendency.

**Bonus Analysis**

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

In this instance, the median (“middle value”) is the preferred, and more accurate, measurement of central tendency because it is not influenced by extreme values known as outliers. For example, according to the statistical analysis, there is a sum total of 424,819 backers across 2,185 successful campaigns with a mean (arithmetic average) of 194.43 and median of 62. The mean value is significantly greater than the median (“middle value” of the dataset) due to the frequency of extreme outliers such as the maximum value of backers (26,457).

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. We use variance to determine the spread of the data as it considers and describes the distance of the values of in the dataset relative to the mean of the data. This makes sense as successful campaigns are more likely attract extreme values corresponding to the relative number of backers, and the number of backers can vary more significantly. Failed campaigns tend to have very low numbers of backers with few exceptions.