Nefertiti Muhammad

05/01/2024

Module 1: Crowdfunding Book

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
   1. Regardless of the category, crowdfunding campaigns have a high likelihood of failure. In the Parent Category and Sub-Category pivot charts, the ratio between the different outcomes can be compared with each category in the stacked columns. While calculating the percentage of failure per category would be more accurate, the stacked column charts clearly show that failure rates hover around 50% for most campaigns.
   2. Crowdfunding is a popular platform for art related campaigns (film & video, music, and theater) and make up the majority of total campaigns. Specifically, campaigns related to play productions make up the majority. This can be seen in the Parent Category and Sub-Category pivot charts.
   3. The Launch Date vs Outcome pivot line graph depicts the relationship between campaign outcome and launch date, aggregating data from ~2010 to ~2020. Across that time span, there is a spike in successful campaigns in June and July, before falling back down in August through December. There is also a slight dip in failed campaigns during June and July, before rising back up in August. This mirrors the trend in successful campaigns, suggesting that summer may be an optimal time to launch a campaign in order to increase chances of success. Despite the trends shown in the graph, statistical analysis is needed to definitively say that there is or is not a best time of the year to launch a campaign.
2. What are some limitations of this dataset?
   1. The Category pivot charts show the ratio of the different outcomes within each category but using a “100% Stacked Column”, or calculating the percent outcomes per category, would allow us to evaluate whether campaigns from certain categories have a higher or lower likelihood of success. The original chart only provides absolute values, where for instance, plays make up the majority of crowdfunding campaigns. However, its ratio between successful and failed campaigns are similar to the other categories despite having more total campaigns.
   2. The Launch Date vs Outcome pivot line graph would be more informative with statistical analysis (standard deviation, t-test…etc), where the trends in the graph could be determined as significant or not.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   1. A heatmap of the campaign outcomes (success and failure) against continuous variables from the raw data (goal, average donation, backer count,…etc) could show what the most significant variables are that affect campaign outcomes. I would hope to see high-intensity clusters around a few variables and from that a more detailed analysis could be made with charts and graphs.
   2. I am curious about the effect that campaign length has on the outcome. There could be an optimal amount of time needed to reach the funding goal. Calculating the difference between the date ended and date created would give the length of the campaigns. Generating a pivot chart clustered column, comparing the lengths of failed and successful campaigns vs campaign categories, would show whether or not there is an optimal timeframe for campaigns to reach their funding goals. Standard deviation and a t-test would be incorporated into the plot (error bars) to determine if any trends are significant.
4. Use your data to determine whether the mean or the median better summarizes the data.
   1. Both the data for successful and failed campaigns skew right, where the mean is significantly larger than the median. Higher values towards the right pull the mean away from the median, overestimating the central point of the data. The median better summarizes the centrality of the data.
5. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
   1. The standard deviation is 149% of the mean for successful campaigns and 164% of the mean for failed campaigns. There is slightly more variation in the failed campaigns but it may not be significant. The rest of the analysis shows this variation. Both data sets skew right with a number of values significantly larger than the median and mean, the maximum values being around 36 times and 53 times more than the median. The variance is also high in comparison to the mean for both data sets.