Jack Wall

10/01/2023

Challenge 1 Report

Three conclusions we can draw about crowdfunding campaigns are that theaters and plays have a higher-than-average amount of crowdfunding campaigns, audio journalism has the highest success rate of crowdfunding campaigns, and the most successful campaigns occur during the beginning of summer.

Some limitations of this dataset are that we don’t know what resources are being used to crowdfund or what the reach of each campaign is which would impact the total funds raised. Another limitation is that these campaigns are being completed in different countries with different currencies which have different exchange rates which fluctuate by the minute so for example any campaigns raising 100 USD would have raised more than a campaign that raised 100 DKK. Another limitation is that the sample sizes are not nearly the same, with audio campaigns having a much smaller sample size and while it had a 100% success rate might not have had the same success rate with a larger sample size of campaigns.

Other tables/graphs we could create would be looking at sub-category by parent category to see which campaigns were most successful by sub-category to determine which of each sub-category is most successful to determine which sub-category would be most successful to focus resources on and to research why some sub-categories were least successful. Another graph we could create is successful and failed campaigns by goal to determine if there is bias in the data as campaigns with lower goals might have a higher success rate than campaigns with higher goals and you could filter down the data by parent and sub-category to determine which campaigns had higher goals that were successful to determine if some categories were more successful if they had higher goals.

When reviewing the statistical importance of the median and mean through a box and whisker plot we see that there are a number of outliers in both data sets which would skew the mean, meaning the median would be the more significant value to summarize the data.

There is a larger variation in the successful data set as there are more successful campaigns with several large outliers in the data. This makes sense with a larger dataset and when the lower and upper bounds have a large gap between them.