4/30/2023

Jeffrey Robertson

[company name]

Crowdfunding Analysis

Written Report

# Analyzing Outcomes of Crowdsourcing by Category

From the data collected from you could conclude that theatre was the most popular type of campaign to be launched at 344 total followed by film & video, then music. However, theatre did not have a higher chance of successful campaigns in comparison to other categories. The table below shows the outcome based on the category.

As is demonstrated, theater is popular and had the highest number of campaigns launch and be successful. However, the data suggest that while theater is the most popular category, it doesn’t succeed at a higher rate than other categories. In comparison to theater success percentage of 54% both film and music. The one drawback from this data is that the data doesn’t give insight as to why the they are not successful. The question could be asked about the average goal of each category compared to the number of campaigns. For some categories there is not enough data to draw conclusions because the number of campaigns is significantly lower than the rest of the categories.

The data could be further broken down by sub-category for each campaign type. From this data we could see which subcategories are popular with fundraisers and which have the most success or failures. The graph below shows that plays is the most popular subcategory of campaigns with 344 total campaigns. The flaw in this data is that there are not an equal number of sub-categories for each category. As the graph below shows, plays is the most popular, but it doesn’t show that it is the only subcategory of theater which is the most popular category.

This data though can be used to compare subcategories within one category to analyze when drilling deeper into a category to analyze which more successful in that category and which is more likely to fail.

A third way to analyze the data would be to look at the outcomes over time. This data would suggest that a campaign is more likely to succeed in July, and not succeed in August. This could be caused by the fact that the number campaigns started in July is higher than any other month.

Some limitations on this data set is that the currency type is not consistent and we don’t have knowledge of the exchange rate to be able to translate the data into a consistent rate. This means that the amount funded for a campaign could be a lot higher than another but be shown as the same because they are in different currencies. There are some challenges with the some categories only having one subcategory which skews the data received to that subcategory.

Some other types of charts that could be used for analysis is to a column chart to analyze where most campaigns are launched and to see if country has an impact on the success of a campaign. Another good method of to have a graph comparing the outcome based off of if they campaign was a staff pick or spotlight to see if either effects the outcome of the campaign.

# Statistical Analysis

The problem with the data is that there is a large variance between the number of backers and the standard deviation of backers. This is more pronounced in for successful campaigns as the outliers push the mean higher than the median. For successful campaigns the mean was 851 which put it in between the second quartile and the third. With the data at hand it would be a better representation of the data to use the median as it is closer to the majority of the data and there is a too many outliers that greatly skew the mean and the variance. The same argument can be made for the unsuccessful campaigns. While the variability isn’t as wide as successful campaigns, there is still a large variability in number of backers for unsuccessful campaigns. The mean is also skewed because of outliers in the data that skew the mean away from the normal distribution model. For this reason it would be better to use median to judge the number of backers for successful and unsuccessful campaigns.