Julianne Lo

UNC Chapel Hill Data Analytics Boot Camp

Module 1 Challenge

15Jun2023

Crowdfunding Data Analysis Summary Report

Results and Discussion

According to the current dataset between 2010 and 2020, there are total of 1000 campaigns in Crowdfunding, among which there are 565 successful, 364 failed, 14 live, and 57 canceled campaigns (Table 1). The mean of backer counts for both successful and failed campaign outcomes are 851.1 and 585.6 respectively, while the median for successful and failed are 201 and 114.5 respectively (Table 1). Since the means encompass several outliers in the box plots and cannot be excluded due to no known causes, they are more suitable for representing the center tendency of data population compared to the median (Figure 1).

The variance of backer counts from successful campaigns is greater than that from failed (Table 1). It is possible because in any successful campaign with any amount of goal and pledged funds, there are going to be various amounts of backers investing. On the other hand, it is worth looking into which categories and or subcategories contribute to the variation.

Among 1000 campaign outcomes, the top 3 contributors by categories are Theater, Film& video, and Music respectively with Journalism scoring the lowest (Figure 2). With Theater scoring the number 1 in the category, its subcategory, Plays contributes 344 overall campaigns: 187 successful and 132 failed (Figure 3).

When looking at campaign outcomes versus months, June and July have the most successful campaigns (Figure 4). By looking at campaign outcomes based on goals, 100% successful campaigns have goals set among these three ranges: 15000 to 19999, 20000 to 24999, and 30000 to 34999. When campaign goal is ≥ 50000 or between 10000 to 14999, the failed campaigns outrated its successful counterpart (Figure 5).

The dataset does not present a lot of data points considering its date range between 2010 and 2020 and from different countries, after factoring in pandemics in 2020. This can be one of the limitations of this dataset because larger sample sizes represent population and assist decision making better. Another limitation could be unable to understand additional information of the data in column staff\_pick and spotlight. Perhaps these two binary data have impacts on backer counts, name or companies of campaigns.

To further decipher this dataset, additional analyses are suggested, graphs such as line charts to compare successful, failed, live, and canceled campaigns from different countries by years, months and quarters. Scatter plots are useful to identify correlations between variables such as percent funded and categories. With more datapoint in the future, analysis such as histogram can be incorporated and utilized in machine learning to predict campaign outcomes for higher successful rates.

Table 1

A picture containing text, screenshot, number, font

Description automatically generated

Figure 1

A picture containing text, screenshot, number, diagram

Description automatically generated

Figure 2

A picture containing text, screenshot, plot, diagram

Description automatically generated

Figure 3

A picture containing text, screenshot, plot, diagram

Description automatically generated

Figure 4

A picture containing text, diagram, plot, line

Description automatically generated

Figure 5

A picture containing line, plot, diagram, text

Description automatically generated