Module 1 Challenge

Written Report

Based on the crowdfunding database of 1,000 sample projects:

* Most of the campaigns (70%) are categorized under arts and entertainment – namely film & video, music, and theater – and about half of these campaigns (55%) succeeded in meeting or exceeding their initial goal.
* The month in which the project is launched may affect the outcome of the campaign as the most significant percent change in successful campaigns occurred between campaigns that were launched in July and August (29% decrease). Meanwhile, the most significant percent change in unsuccessful campaigns occurred between campaigns that were launched in August and September (34% decrease).
* The projects that had an initial goal between $15,000 and $50,000 are more likely to succeed as more than 65% of the projects within this range met or exceeded their initial goal.

Some limitations of this dataset include:

* The criteria for the “staff\_pick” and “spotlight” columns are not given, only a “True/False” value is given.
* The dataset only includes the number of backers for each project, without specifying the individual contribution amount from each backer.
* Based on the blurb of each project, the parent category and subcategory can be broad or there can be some overlap in the categories.

Other possible tables and graphs that could be created based on this data are:

* A pivot table that gives the average count of backers for successful, failed, canceled, and live projects per category. A stacked-column pivot chart that visualizes this table can also be included.
  + This would give an idea of which categories had more backers as well as the outcome of each project per category based on the backers count.
* A table and graph similar to the crowdfunding goal analysis, can also be created and instead of ranges of the initial goal, the ranges are made according to the average donation for each campaign.
  + This would display if there is a relationship between the average donation to the outcome of the campaigns. It would provide information on whether a certain threshold for the average donation needs to be met or exceeded for a campaign to have a higher likelihood of success.
* A table with a column of ranges of how many days each campaign were held for (0 to 5, 6 to 10, 11 to 15, … , 55 to 60) and columns for number of projects that were successful, failed, and canceled within these ranges. A stacked chart that visualizes this table can also be included.
  + This would provide information on the more frequent duration for launching projects and whether the number of days affect the outcome of the campaign.

Statistical Analysis

Since the data for successful and unsuccessful campaigns are both skewed left, the median is a more appropriate measure of central tendency. The median better summarizes the data because it is less affected by outliers and provides a value that is more representative of the typical number of backers for successful and unsuccessful campaigns.

There is more variability with successful campaigns and this result makes sense because unsuccessful campaigns are more likely to have a lower number of backers, potentially reaching an amount where the project fails to meet its funding goal. In contrast, successful campaigns are more likely to have a greater number of backers and there is no limit to the number of backers the project would have as it has already surpassed a certain threshold for success.