Icon

Description automatically generated

Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?

Theater, Music, and Film and video are the top revenue-generating campaigns on Kickstarter, given the data provided. However, when the sub-category tab plays, rock and documentary were the most successful. Based on the launch date, May is the month that has the highest success rate.

What are some limitations of this dataset?

There is no information about the Kickstarter owner, and such has history, past campaigns, and success rates of any previous campaigns. Additionally, we do not have information about the backers of the projects and if that impacts the success or failure rate. Finally, it would be beneficial to know the age, gender, and time of donation to find more commonalities.

What are some other possible tables and graphs that we could create?

A pivot table about the spotlight feature and how that influences success: also, we could have done a bar chart to show the success versus failure of staff pick. A pivot table to provide the length of time a Kickstarter campaign was live could be represented. A bar chart with the average donation by state and staff pick could be presented.

# Bonus Statistical Analysis

**Use your data to determine whether the mean or the median summarizes the data more meaningfully.**

The median summarizes the data meaningful in both the successful and failure Kickstarter campaigns. To explain that statement, beginning with the successful campaigns, there were 194 campaigns, 62 median campaigns. The average number of backers triple in comparison to the median number of backers. If this information were displayed in a histogram chart, the mean and median data would be right-skewed. Therefore, there is no center to the data. The majority is of the backers are around 100 per campaign. With the range of the backers being between 1, which is the minimum, and the maximum is 26,457, that is a broad span. This is confirmed by the variance of 712,841. The standard deviation for the number of backers is 844.3, also pointing to the spread.

The average failed backers is four times larger than the median. The failed number of backers also is right-skewed due to the spread. The minimum number is 0, and the maximum is 1,293. The increase in maximum pushes out the span of numbers. The majority for this dataset is 20, and the variance is 3,773. The standard deviation of the dataset is 61.4, like the successful dataset indicates a wide spread.

Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why ore why not?

The analysis between the two datasets indicates that successful campaigns have more variables. A pattern was seen within the failed dataset, a low turnout of support, low or no backers. The successful backer data indicated more backers supported the successful campaigns. Some of the campaigns within the successful dataset had a low backer count due to a backer’s pledge amount, but that was the majority within the data.