Haley Armenta

Module 1 Challenge

Dec. 10, 2024

Given the data, what are three conclusions that we can draw about crowdfunding campaigns?

* Based on the data, we can conclude that theater campaigns have been the most successful, followed by film & video, and then music. Additionally, successful campaigns tend to be funded beyond 100% of their goal. Lastly, the best time to run a crowdfunding campaign is typically in mid-year, with June and July being the most successful months.

What are some limitations of this dataset?

* One limitation of the dataset is the presence of outliers, where some failed projects receive little to no backing and have very low funding percentages. On the other hand, successful campaigns often have large numbers of backers, sometimes exceeding 100% of their funding goal, which can skew the data.

What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

* Other useful visualizations for this dataset could include a scatter plot to examine the relationship between the number of backers and the percentage of funding for each campaign, helping identify patterns or correlations. A pie chart could also be useful to show the distribution of campaign success by category (theater, film & video, music), providing a clear overview of the proportions of success in each category. Additionally, a box plot could help visualize the distribution of funding goals across all campaigns, highlighting the range, median, and any potential outliers. These visuals would add more insight into funding trends, campaign distribution, and the variance in campaign success.

**Analysis**

Use your data to determine whether the mean or the median better summarizes the data.

* To figure out whether the mean or the median does a better job summarizing the data, I compared how each one reflects the distribution of backers for both successful and failed campaigns. The mean is sensitive to outliers, so if there are campaigns with a huge number of backers, the mean could get pulled higher. The median is less affected by outliers and represents the middle value, so it tends to be a more reliable measure when the data isn’t evenly distributed. I also looked at the variance and standard deviation—if these are large, it suggests a lot of variability, which might mean there are outliers. In that case, the median could be a better choice for summarizing the data.

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

* To figure out whether there’s more variability with successful or unsuccessful campaigns, I’d compare the variance and standard deviation for both groups. Variability refers to how spread out the data is, and higher variance or standard deviation means more variability. If the variance and standard deviation are larger for unsuccessful campaigns, that would mean there’s more spread in the number of backers for those campaigns—some might have a few backers, while others could have a lot. On the other hand, if successful campaigns have larger variance, it could mean there are a few really big campaigns with lots of backers, but most are more modest.
* It makes sense to look at this because variability gives me insight into the consistency or unpredictability of campaign outcomes. If there’s high variability, it suggests that some campaigns are either overachieving or failing dramatically. If the variability is low, it could mean that the backer numbers are more predictable or consistent across campaigns, whether they succeed or fail.