Payson Cushman

3/20/23

Module One Challenge

Crowdfunding Report

* Create a report in Microsoft Word, and answer the following questions:
  + Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  + What are some limitations of this dataset?
  + What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

Although there may be others, I drew three conclusions from this data. First, summer months are a good time to start a crowdfunding campaign and that June, in particular, had a high number of successful cases started during the month. Second, Kickstarter looks to be a good place for plays and theater to raise money. And lastly, games, in particular mobile games, have not been particularly successful reaching their goal on Kickstarter.

However, his data is only based on category, subcategory, and date created. This is helpful to find some trends, but not necessarily the whole story especially since Kickstarter is all about raising money. Thus, it may be helpful to dive deeper into the data that revolves around currency and the number of people donating that currency. Looking at data like the average donation, the backers count, number of pledges, and the funding goal would be ideal places to investigate.

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

For successful cases, the median is a better measure because it describes the central tendency more accurately. In the successful instances, there are 403 cases below the mean and only 162 above it. So obviously, there are some larger ones skewing the mean. The same goes for the unsuccessful cases, although not quite as pronounced. There are 253 cases below the mean and only 111 above it. In conclusion, the median more accurately summarizes the data for both successful and unsuccessful cases alike.

There is more variance in the successful campaigns. There was a standard deviation of 1267 in successful campaigns to only to 961 in unsuccessful ones. This makes a lot of sense since there is a higher likelihood of success if the goal is small since there is less money to raise. In these cases, there are likely to be less backers or at least less are needed to be successful. This creates a large set of successful campaigns with a smaller number of backers. At the same time, the more backers a campaign has, the more likely it is to be successful since more backers would in most cases lead to more money raised. This creates a large data set with a high number of backers. Thus, because there is both a large data set of successful campaigns with a small number of backers and a large data set of successful campaigns with a high number of backers, there is a high variance in the number of backers for successful campaigns.

In the cases of the failed campaigns, there is still plenty of variance; just less. This adds up because, although there a plenty of cases where there are little or no backers (64 cases of less than 24) there are less cases where there are a high number backers and still fail (4 cases of more than 4500). Therefore, there is almost a top end in the number of backers in failed campaigns and thus less variance overall.