Excel Homework: Kickstart My Chart report

Jeongdae (JD) Kwak

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

Theater is the most popular category of the Kickstarter campaigns, along with plays being the most popular sub-category of the campaigns and journalism being the least popular category of campaigns.

May seems to be the month that has the highest success rate of campaigns, whereas December being the worst because there are more failed campaigns present compared to the successful campaigns.

The higher the amount needed for goal, the lower the success rate. This is present on the line graph. There is a clear trend of more expensive campaign having much lower rate of success compared to the less-costly campaigns.

1. *What are some limitations of this dataset?*

It is hard to tell if there is any correlation in successfulness of the campaign with spotlight being TRUE or false, and with staff pick being TRUE or FALSE. It is also tough to see the average numbers (or central tendency) for each category’s goal number to see if there is a trend in the price of the pledge money within categories without making a separate statistical analysis table on another spreadsheet.

1. *What are some other possible tables and/or graphs that we could create?*

Some possible tables can be made to help clear some things mentioned above. A bar graph showing the mean, median, and standard deviation as a 2D clustered bar graph comparing the goal value for each category, or sub-category. Another table could show the difference between the number of counts in successful, failed, and cancelled campaigns depending on if the campaign were on a spotlight or not. Same goes for count table in status of the campaign that was staff picked or not.

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

Median summarizes the data more because for both successful campaign and failed campaign data set, data are heavily skewed to right. Mean of the data is used when the data is spread more normally, whereas median of the data is used when the data is skewed towards one direction.

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

More variability is shown in successful campaign data compared to failed campaign data. This makes sense because of few reasons. For a campaign to be successful, it needs to meet the requirement money, which typically requires a lot of people supporting the campaign (also known as the backers). When campaign fails, that means it did not meet the required money, which can correlate to not having enough backers. Thus, making the failed campaigns to generally have smaller amount of backers compared to the successful campaigns. When campaign becomes successful on Kickstarter, the campaign starter also adds extra perks when more money is reached (this is typically true when it comes to game-related Kickstarter by adding an additional side-story quest, or putting extra scene in a game, etc.). This tends to gather more backers to an already successful campaign, and have them donate more money towards it.