9/13/2021

Rebecca Melo

E**xcel Homework: Kickstart My Chart**

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

When reviewing the first pivot table, the first conclusion I can pull from is that the most successful projects were music, theater, film, and videos. The second conclusion is that the project categories that failed the most were journalism, food, technology, and publishing. When looking at the top successful projects and break them down into the sub-categories, rock and indie rock music was at the top of the list compared to other types of music. For theater, the sub-category documentary was the most successful. We can also conclude that when looking into the top failed projects, Food trucks projects had the lowest success rate in the food category, along with web and wearable being the least successful sub-category out of the technology category.

When reviewing the third pivot table, the third conclusion that I can pull from this table is that the number of successful campaigns started to decline towards the end of the year. Considering all the years, February, May and October are the most successful months to review new launches. However, April through July months shows an increase in failed and canceled campaigns along with displaying a steep drop in terms of successful campaigns.

1. What are some limitations of this dataset?

Some limitations of this data are there is not enough information in some areas of the data provided. For example, we have the number of categories and sub-categories as to which projects were canceled or failed, yet we do not understand why they were not funded over others. If there were more information around how a project gets canceled or why a project was not adequately supported, there would be a better understanding of the causation. In addition, we cannot go further than the sub-category, meaning I cannot precisely know what kind of rock projects were funded. They could be live shows, albums, studio time, equipment, etc., but there is no additional information to determine that. There is also another limitation of data on the demographic of the backers (i.e., age, income, education, region, state, etc.).

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

One table that could be useful to create would be a pivot table comparing the amount of money contributed to each campaign by country to see which categories are the most successful in different countries. We can also filter the data to create a pivot table to compare categories with the largest average donation with the number of backers to determine which categories are easier to fund. A histogram could also be helpful to understand whether there is a correlation between percent funded and successful versus failed campaigns.

**Bonus Statistical Analysis:**

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

The median summarizes the data more meaningfully than the mean because there are outliers that skew the data. The standard deviation is high relative to the mean, determining that the data points significantly differ from the mean. Since the median is where the most data points are located and is closer to the data points than the mean, this gives us a better picture of the data and where most of these campaigns lie.

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

Based on the variance of both successful and unsuccessful campaigns, the variance of successful campaigns is higher. This would make sense because there is a considerable range of backers in the successful campaigns— between 1 and 26,457 backers. The broader scope of backers had many projects that exceeded their goal and made large donations which offset the data and increased variability. The max for failed campaigns is 1,293, which is exceptionally high compared to the number of total failed campaigns (1,533). Out of that, 287 of the failed campaigns had no backers at all. This reduces the variability since most campaigns will fall in the 0-40 range while a small number of campaigns exceed this range.