Excel-challenge

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

Based on the dataset, one conclusion that we can draw is that a start-up campaign is more likely to be successful than it is to fail, given that a little more than 57% of the crowdfunding campaigns were successful at achieving their goal. In addition, the first half of the year (January 1 to July 31) tends to be the time when crowdfunding campaigns are the most successful. Though, June and July we the two most successful months overall.

Plays were by far the most popular subcategory with 344 total crowdfunding funding campaigns. However, the total number of successful campaign ventures (187) was only slightly above those that either failed or were canceled (154 aggregated). It could be concluded that there is only a marginal chance that a theater related crowdfunding campaign may be a successful endeavor.

Lastly, entertainment related parent categories (film and video, music, and theater) had the most crowdfunding campaigns (697), accounting for nearly 70% of the total (1000) in the dataset. Due to this, if one is trying to enter this space, they should expect to have a lot of competition with those trying to get the same backers for their projects.

What are some limitations of this dataset?

One of the biggest limitations of this data set to account for is that crowdfunding platforms such as *IndieGoGo* and *Kickstarter* were as popular at their respective launch dates as they are today, or even in the years following their initial launches. As a result, the crowdfunding campaigns that started in the most recent years after the launch on these platforms would not have received the same attention as a start-up from later years as crowdfunding platforms became more popular to the public. In short, an early participant in the crowdfunding platform game, that ended up getting canceled or failing, might have been limited by the crowdfunding platforms themselves not having the same acceptance as they would grow to have in later years.

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

An additional graph that would be beneficial would be one that includes how long a crowdfunding campaign was active. A campaign that is open for a longer period of time will have a greater amount of time to receive backers and donations to reach its goal. Not all campaigns were able to be open for the same time frame. Consequently, campaigns that have a short active period may not be as successful at reaching their goal as a campaign that can remain active for a longer period of time. A graph that shows how long a campaign was open in relation to the outcome of that campaign would be useful in determining if there is a correlation between campaign term length and outcome of the campaign.

Moreover, a pie chart that shows how much of the total space each one of the parent categories occupies would be valuable. This type of chart would visually represent how prevalent a category is in relation to the other categories. In other words, this pie chart would show how saturated the crowdfunding market is with a certain crowdfunding category.

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

For both the successful and the failed outcomes the means were vastly greater than the medians. For this dataset the median would be a better representation of the data since there are several outliers in multiple data points that skew the data if the mean was used to summarize the data. For example, one of the crowdfunding campaigns had 6,080 backers and failed while another campaign had 0 backers and also failed. Compare this data with campaigns that had 16 backers and another with 7,295 backers which were both successful in reaching their respective goals. The differences between these low minimums and high maximums present outliers that would drastically impact a mean calculation but would have less of an impact on a median.

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 data, there seems to be more variability with the successful campaigns than with the unsuccessful campaigns. This makes sense because there is a wider range between the maximum and minimum as well as a higher variance. Further, the successful campaigns had more data points to add to the variability of the data.