Module 1 Challenge Written Answers

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

* The three parent categories that see the most crowdfunding campaigns are (in descending order of campaign counts) theater, film and video, and music.
* Except for games, every parent category sees more successful campaigns than failed ones, with journalism, photography, and technology leading in the highest ratios of successful over failed campaigns. It should be noted, however, that journalism only has a total of 4 campaigns in the data set.
* There is no significant variation in the number of campaigns from month to month, but June, July, and September see the highest ratios of successful over failed campaigns.

What are some limitations of this dataset?

* The 1000 sample projects are not equally distributed among the 7 countries in the data set. For instance, 763 campaigns are located in the US. This implies that any conclusion drawn could have regional bias.
* The initial funding goal and the amount pledged are based on local currency, making it difficult to directly compare them for campaigns from different countries due to the lack of currency conversions.

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

* How well a funding project is marketed and how wide an audience it reaches on a crowdfunding platform can directly impact its outcome. As such, a pivot table can be created to evaluate how many campaigns were successful, failed, canceled, or are currently live depending on if it has True or False in the spotlight column.
* The duration of a crowdfunding campaign can determine if it reaches its initial goal. Hence, the statistics of campaign duration can be computed to shed light on its correlation with campaign outcome for each category.

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

* For either successful or failed campaigns, the distribution of backers is skewed and there are a number of data points above Quartile 3 + 1.5 x IQR, indicating outliers. As a result, 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?

* The standard deviation is lower for unsuccessful campaigns, indicating less variability. This makes sense to an extent because unsuccessful campaigns tend to have comparatively fewer backers, as evidenced by lower mean, median, and difference between minimum and maximum number of backers compared to successful campaigns. With fewer supporters overall, comparatively less variation results.