Excel-Challenge

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

* The entertainment industry, which include film & video, music and theater categories, account for most of the campaigns out of all the other categories. We can see this in the Categories pivot chart, which displays the following:
  + Film & Video had a total of 178 campaigns
  + Music had a total of 175 campaigns
  + Theater had a total of 344 campaigns

In comparison, all other categories had less than 100 total campaigns

* Aligned with this, the most prominent sub-category is the plays category, which sits within the theater category. By analyzing the sub-categories pivot chart we can see that the plays sub-category had a total of 344 campaigns while all other subcategories had under 100 total campaigns.
* The most successful months seem to be the summer months of June and July as those are the months with the most successful campaigns with 55 in June and 58 in July. All other months had under 50 successful campaigns.

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

* The data set is heavily based on US data. For example, there were 763 total campaigns in the US, but under 50 campaigns for each of the other countries represented. Therefore, it is difficult to compare data based on country. I think it could be useful to separately analyze US data and “International” data.
* The run times of the campaigns vary greatly. For example, one campaign ran for 1 day and was deemed as a failed campaign while other campaign ran for over a month and was successful.
* It is unclear what the staff picks and spotlight columns represent.

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

* I think it would be valuable to see the range of campaign run time and compare it to the outcome. It would be interesting to see if there’s a correlation between longer run time and success of campaigns. To demonstrate a possible relationship I would recommend creating a scatter plot with run time on the X-Axis and outcome as the Y-Axis.
* It would also be interesting to analyze if there’s a correlation between goal amount and outcome. Are lower goals more likely to be successful? Again, since we are trying to demonstrate a relationship, I recommend creating a scatter plot with the goal amount on the Y-Axis and the Outcome on the X-Axis.

Statistical Analysis

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

* When trying to summarize data, the mean/average will always be a better choice than the median because it takes into account all of the data versus the middle point of 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?

* There is more variance within successful campaigns, which makes sense given there were a lot more successful campaigns than there were unsuccessful campaigns.