Module 1 Challenge - Written Report

* + Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
* ***The goal of this analysis*** is to uncover any hidden trends in a crowdfunding platform’s data that can identify data trends for successful projects.
* ***The fist observation is in the definition of the attributes*** such as failed, successful, canceled and live. The attribute such as success is measured by comparing the original goal set by the project in comparison to the value of the pledge received. When the value of the pledge exceeds the initial goal, a success is defined, while the reverse is defined for failed. At the same time, live represents the projects that have been launched in comparison to those that were canceled.
* **The second observation is in the way the graph reports are organized**. The above mentioned set of attributes (failed, successful, live, canceled) can be analyzed from different row labels, namely those that include categories such as including parent categories, sub-categories and associated dates. These reports are trying to draw particular attention to analysis of the above mentioned attributes based on a macro view (parent categories) in comparison to a micro view (sub-categories) and lastly a view point based on time/dates in an attempt to draw attention to the relationship of these factors to each other.
* **The third observation is in the visualization** of the graph types. In the case of the line graph which consist of row labels of dates/months, we can see that the trend for failed, successful, and canceled projects are pretty much within a consistent trend across the time chart. What is observed however is a sharp drop in successful projects occurring around the same time that a rise of failed and canceled projects occur in the month of August, indicating a potential trend in the time series. The stacked-column pivot chart shows that out of all the parent and sub-categories, theatre/plays generate the most successful outcomes, while technology/web has the best ratio of higher successful outcomes compared to lower failed/canceled outcomes, suggesting that this type of trend may be more attractive for a broker/investor who may have a lower risk tolerance.

* + What are some limitations of this dataset?
* As an investor/broker, who is looking for an analysis on successful trend, this dataset is limited to the relationship to the definitions of the data as mentioned above (i.e. success is defined as meeting or exceed the initial goal in comparison to the value of the pledges). Also, while a stacked-pivot chart provides the amount of successful, canceled, and failed/live based on the type of category, it’s not obvious what is the best trend for success. In business, typically investors that are interested in successful trends as more than just meeting the goal in pledges. They want to see things like what is the highest success ratio in comparison to the lowest failed / canceled projects, and match that with the lowest investment or pledge value made, that produced the highest return on investment. This way, you can assess the risk tolerance and best vertical to invest in.
  + What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
* Firstly, it would be useful redefine a new attribute called ‘best investment’. Then define this new attribute as representing the ratio between highest success to lowest failure. An example of this can be that ‘best investment’ would show that “web” sub-category shows that for every 3 successful campaigns only 1 failed, whereas for “plays” every 17 successful campaigns 12 failed. In order to do this, you would need to use a math formula that converted the total number of failed (example cell B1) and total number of successful (example cell C1) to a ratio as follows “=B1/GCD(B1,C1)&":"&C1/GCD(B1,C1)” which will display the ratio to give an alternative perspective in a new table format.
* Use your data to determine whether the mean or the median better summarizes the data.
  + The mean uses the entire data set but, the downside of the mean is its vulnerability to outliers and skewed data. In the case of successful outcomes, we have 3 higher backer\_counts, and we may want to represent the summary to include those outliers. If we were to a) discard those outliers, or b) use the median as a representation of that data, those three higher backers\_counts in the data points are essentially ignored. Just because only three backer\_counts that have higher values, does not mean we don’t care about the data. In the case of failed outcomes, we have 2 higher backer\_counts, and similarly we may want to represent those in our summary.
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
  + When I calculated the upper and lower limit by looking at mean +1SD, mean -1SD, as outlined in the worksheet, it indicated that successful backer\_counts had lower percentage of 67% in comparison to the unsuccessful backer\_counts which showed 89%.