William Fetter

**Module 1 Challenge**

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
2. When the data is visualized by Parent Category versus Campaign Outcomes, Theater had the overall greatest number of campaigns started.
3. When the data is visualized by Sub-Category versus Campaign Outcomes , Plays had the largest number of successful and failed crowdfunding campaigns.
4. July was the month with the greatest number of successful crowdfunding campaigns created.
5. What are some limitations of this dataset?
   1. A major limitation of the dataset is that the reason for campaign outcome is not provided. While we have enough information to determine the exact number of successful, failed, canceled, and live outcomes, we do not have enough information to determine why this was the case. If we knew why campaigns were more successful or less successful, it may help performing future crowdfunding efforts.
6. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
7. We could create a chart and graph according to “country” of campaign started to analyze differences by individual countries. It may be that different countries had better or worse success with crowdfunding campaigns. Focusing on more successful counties may help to organize future efforts.
8. Using the date created and date ended, we could create a chart and graph according to “length” of campaign versus campaign outcomes. It may be that campaign lengths could be directly correlated to campaign outcomes, and by focusing on those lengths with the highest success may help to plan future crowdfunding campaigns.
9. Summary Statistics Analysis - Use your data to determine whether the mean or the median better summarizes the data.
   1. To understand if the mean or median better summarizes the data we first must determine if the data is normally distributed, or if it is skewed. Since the both mean values are greater than the median values, we know that the data is right-skewed. Additionally, from the calculations or standard deviation and variance, we know that both data sets are significantly spread out from the mean. With the data being skewed, the median values better summarize the data (the mean would better summarize the data if it was normally distributed).
10. Summary Statistics Analysis - Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
    1. If we compare the standard deviation and variance of the successful campaigns (std dev = 1,266 & var = 1,603,374) versus the unsuccessful campaigns (std dev = 960 & var = 921,575), we see that the successful campaigns have a larger data spread and therefore more variability. This makes sense because the successful campaigns have a larger difference between the mean and median, as well as a greater range of data. Therefore we would expect the successful campaigns to be more spread out.