**Hannah Bridges 9/9/20**

**Excel Challenge Reflection**

1. Given the provided data, what are three conclusions we can draw about Kickstarter Campaigns?
   1. Campaigns in entertainment categories (music, film, theater) were more likely to be successful than other campaign types. Music was the most successful Kickstarter category during this time period.
   2. Campaigns with lower funding goals are much more likely to be successful. Past a goal of $15000 the goal amount seems to have a smaller effect on outcome.
   3. Overall Kickstarter use appears to have peaked in 2015. Because the line chart does not include live campaigns it is unclear whether 2017 will continue the downward trend seen between 2015 and 2016.
2. What are some limitations of this dataset?
   1. The goal and raised amounts are listed in multiple currencies, this is likely skewing the goal related data in unknown ways.
   2. This data is from only one crowdsourcing platform. There are many other such platforms that are potentially attracting different types of fundraisers and contributors. For this reason, assumptions made by analyzing this data may not be representative of crowdsourcing in general.
   3. This data set does not include whether campaigns offered incentives for donating. Often campaigns use crowdsourcing platforms to encourage preordering of products. It is possible some campaign categories are more likely to do this than others thereby artificially inflating their success rates.
3. What are some other possible tables and/or graphs that we could create?
   1. I would be interested in comparing goal amount by campaign type and subtype (likely in a line chart) to see if certain categories generally have lower goal amounts and thus are more likely to be successful for that reason. This might explain the greater success rate seen in the music category.
   2. You could also create a pivot table to look at cancelled campaigns. Specifically, I would look at how long they were live for and what percentage of their goal they achieved. It seems likely that many cancelled campaigns should be included in failed campaigns, especially if they were far from their funding goal after a significant period of being live.
   3. Additionally one could determine the time live (number of days between the launch and deadline) and see if the length of that period is correlated with the likelihood of success or failure.

**Summary Statistics Bonus Reflection**

1. Use your data to determine whether the mean or the median summarizes the data more meaningfully:
   * For both the Successful and Unsuccessful Campaigns the median better summarizes the data. This is because both data sets have several large outliers in the upper data range. I confirmed this using both Box and Whisker plots and by calculating the mean of the data if 20% of the outlier data is removed. Calculating the Trimmean using 20% in excel showed a number much closer to the Median then the original Mean.
2. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
   * There was significantly more variability in Successful campaigns then Unsuccessful campaigns. This is seen both in the higher variance (712841 vs 3773) and the larger standard deviation (844 vs 61) of Successful Campaigns.
   * It makes logical sense that the Successful Campaigns would have more variance simply because there is a wide range of goal amounts and assuming higher goal amounts result in more backers you would expect to see a much higher upper limit of backers. With unsuccessful campaigns there is a similar range in goal amounts (there are fewer unsuccessful low goal amounts but similar distribution overall) but since none of these campaigns reached their goal it makes sense that there would be lower overall numbers and fewer high outliers.