David Roth Module 1 Challenge

Create a report in Microsoft Word, and answer the following questions:

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

What are some limitations of this dataset?

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

* Presents a cohesive written analysis that:
  + Draws three conclusions from the data (10 points)
  + States limitations of the dataset and suggestions for additional tables of graph (10 points)

Use your data to determine whether the mean or the 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?

* Computed calculations of the mean, median, min, max, variance, and stdev using Excel formulas (15 points)
* A brief and compelling justification of whether the mean or median better summarizes the data (5 points)

David Roth Module 1 Challenge

The data gathered and compiled in this report gives way to a number of conclusions about the nature of these crowdfunding campaigns. It can be seen that theater has the greatest number of successful campaigns (as well as campaigns in general.) We can also gather that successful campaigns outnumbered the failed ones in every month; however, the gap between the numbers of successful and failed campaigns is closest in the month of August; no month had fewer successful campaigns relative to the number of failed ones. Furthermore, it appears that the percentage of successful campaigns and the percentage of failed campaigns are almost identical inverses of each other. The percentage of successful campaigns have a downward trend (the percentage of failed campaigns have an upward trend, conversely) when the amounts range from $1,000 to $4,999, and from $30,000 to $34,999.

These graphs and representations have their limitations. They do not readily show the percentage of successful, failed, live, or cancelled campaigns on the basis of category. Although one of the stacked bar charts does reveal a representation of the **number** of campaign outcomes, it can be somewhat misleading; it can perhaps appear that theater had the highest percentage of failed campaigns because the bar demonstrating the number of failed campaigns is larger than the others, for instance. While having a chart to count numbers is helpful, it would be additionally beneficial to include a graph for percentages, although that too has its disadvantages (it would be unfair to compare percentages of successful campaigns in music, which had dozens of times more total campaigns than journalism, which only had four total campaigns). There is also no way to see the rate of success in campaigns based on the length of time allotted for each campaign. Some campaigns were able to be successful in a day, others were not able to achieve success in a matter of weeks. If there was a graph to analyze this aspect; that could help in setting more efficient deadlines for campaigns to run their course. Another additional aid would be to display a visual for the average donation amount for each campaign, and break that down into the type of campaign as well as the outcome, so they can put more of their efforts into the campaigns that will receive more donations and have a better chance to be successful.

There seems to be more variability in the number of backers of the successful campaigns. This would make more sense because there were more successful campaigns than failed ones, therefore making it likely that there would be more variance in the numbers.. It also makes sense that under general circumstances it might require more backers to donate to these campaigns in order for each of them to reach their thresholds. 7% of successful campaigns and 9% of failed ones have a number of backers that exceeds the outlier, which is enough to cause the mean to deviate from the majority of the data. On both sets of numbers, the median is the better metric to provide a sense of where most of the data can be found.