Analysis Report

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
   * Based on the provided data in pivot 1, theater owns most success and failure among all parent category and journalism has least counts of success in its outcome and percentage of success is 100%
   * Based on the provided data in pivot 2, plays are the only sub-category in theater, so it has same counts of outcomes as theater. It stands out significantly compared to other sub-categories. Among all sub-categories, science fiction is the only sub-category that has a higher failure rate than success rate.
   * Based on provided data in pivot 2, success rate among the years ranges from 40% to 60%, failure rate ranges from 20% to 40%, and cancellation rate ranges from 0 to 10%.
2. What are some limitations of this dataset?
   * These data don’t tell us how good each category performs. As long as their funding passes their goal, it is considered successful, but these data doesn’t provide any information on how much funding is short or over for each category.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   * To tell each categories’ funding outcome against their goal, we can use combo chart. With the sum of goals under each category plotted in line chart type and the sum of pledge funding plotted in stacked area, we can visualize how each category beats or fails its objective funding.
4. Use your data to determine whether the mean or the median better summarizes the data.
   * Because variance is extremely high and median is significantly less than mean of success and failure, we can conclude that the data is heavily skewed toward the minimum point (to the left). In this case, use of median to summarize the data is more valuable than use of mean.
5. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
   * Based on the data, I believe a successful campaign has higher variability than unsuccessful campaign because both variance and standard deviation of success campaign are higher than of failure campaign. This makes sense to me because when we look at the difference on maximum and minimum value of both successful and unsuccessful campaigns, we can tell that the successful campaign has greater difference than unsuccessful campaign as well as both data are positively skewed.