Brendan Doucette

Crowdfunding Excel Report

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  + When changing the years to single specific years for the outcome by month based, the graphs become a lot more sporadic and produce line charts that show instances where there are more failed projects than successful projects. Then, once you include all years the graph shows a more balanced picture with each month producing more successful projects than failed and canceled projects. The canceled projects are a much smaller sample size compared to the failed and successful projects.
  + I also found when factoring by country on the pivot tables that the US had a lot more crowdfunding projects than any of the other countries. This causes it to have a big impact on the table and charts for the outcomes based on the parent and subcategory. The graph for outcomes based on parent category the US had an almost identical looking chart to the one including all the other countries, whereas the other countries have different looking charts because of the smaller sample size.
  + Lastly, I noticed that the plays subcategory had significantly more crowdfunded projects compared to the other subcategories by a significant amount. The plays subcategory had 344 total projects, meanwhile the other subcategories all have less than 100 total projects respectively. A lot of these projects were about 50/50 in failed vs successfully for each subcategory, with a few having a small sample size to compare.
* What are some limitations of this dataset?
  + One limitation is that the blurb variable does not provide much to help us tell a story with the data set. It is hard to correlate the blurb variable to the other variables given, especially the outcome which was our main tested variable. Maybe it could be cleaned up in a way that the blurb shows more relationships to other data points to be used to tell a story. For example, if the blurb consisted of certain words that were present in other data points so that we can used those values that show similarity to find a new story from the data.
  + Another limitation is that there are a lot of outliers involved in this data set. Especially within the backers count and goal, it is hard to find a correlation between those two and the outcome because of how wide ranged the values are in those variables.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  + Another possible table could be viewing the outcomes based on the Average donation. We can also use a line graph to help visualize this better for us. This can help us see if there is a correlation between increase Average donation and the outcome of the crowdfunded campaign.
  + Another graph we could make would be an area graph of the percent funded based on outcome. We can use this graph to show the difference between how much percent a project needed funding to be successful compared to failing.