Bryce Deering / Module 1 Challenge / Data Analytics Bootcamp

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  1. Out of the 1000 total crowdfunding campaigns, 565 or 56.5% of the campaigns were successful compared to 364 or 36.4% that failed. We can infer from this information that most of the campaigns were successful.
  2. The data set also shows that Theater is by far the most prevalent category, coming in at 344 out of the 1000 total.
  3. The data also shows that the Journalism category is the most successful at a rate of 100% but only has 4 total campaigns, which is an extremely small sample size compared to the rest of the data population. The Games category was the least successful at a rate of 44%. It is noted that the Games category includes 48 campaigns out of the 1000 total, which is still relatively small.
* What are some limitations of this dataset?
  1. The dataset includes a variety of limitations that make it hard to draw accurate conclusions about the campaigns. The dataset includes information about the outcome of the campaigns, the number of backers involved, the country where the campaign is located, the length of the campaigns, a “blurb” about the campaign, and the category/sub-category for each campaign. The “blurb” section of the dataset provides relatively limited information about the specified campaign whereas more detailed information about the campaign would be helpful to better understand the nature of the campaign and any potential planning/consideration that went into creating/running the campaign. Additionally, we are given the category and sub-category for each campaign, but that information is relatively useless outside of making correlations about the successfulness of campaigns based on the category/sub-category they fall into. These labels are also rather broad and more specific information about be helpful to further break down the campaigns for more accurate analysis. We are also not given any information about the background of the campaigns outside of the “blurb” and country they are located in. Having more contextual information about the location and size of the involved parties would provide more information related to popularity, the reach each campaign has, and whether the goals are realistic or ambitious.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  1. While we created a table to provide detailed analysis on information related to outcomes and backers count, it would be worthwhile to include information related to donations. We can see the average donation per campaign but it would be useful to provide analysis related to donation data that takes into account category/sub-category, median, max, and min donation amount to locate outliers and better understand causes for success/failure outcomes.
  2. Creating a table/graph that provides information about the average timeline of a campaign, or how much time each campaign had to reach its goal, would be valuable to see if the length of the campaign had a positive or negative correlation to success rate or if the length of the campaigns were correlated with the category/sub-category.