Brian Leitz

Data Analytics Boot Camp

Module 1 Excel Challenge

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
   1. Campaigns with smaller goals, under $10,000 or $5,000 for example are more likely to have a high success rate compared to campaigns with larger goals.
   2. The bigger the campaign goal does not necessarily mean the success rate will be lower, however, we can see from the data that campaigns with goals over 50,000 have a higher failure rate and are inconsistent in reaching that goal.
2. What are some limitations of this dataset?
   1. The biggest limitation of this data is that it is fairly unbalanced which creates some difficulty when drawing conclusions in our analysis. We can observe that the majority of campaign goals are under $10,000 with their being 597 projects with campaign goals under $10,000 and 389 projects with campaign goals above $10,000. There are however only 84 projects with a campaign goal between 10,000 & 50,000 and 305 projects with campaign goals over $50,000. The high number of campaign goals over $50,000 creates an unbalance, or skews, the data as well as the tables created to further analyze or visualize the data.
   2. The data given is in my opinion, surface level data, with regards to each campaign and the details of each campaign. We do not have any info on individual donors or how many campaigns have large donors that represent a large portion of the amount pledged that could be skewing the data.
   3. We can also observe in analyzing the outcomes by category that Journalism has little to no campaigns and Food, Games, along with Photography have a small amount of campaigns.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   1. I believe that a pie chart would be very handy in being able to better visualize the split of outcomes between categories as well as the dollar amount windows that we created in our Crowdfunding goal analysis. This would allows us to visualize the percentage split and better observe outliers or unbalances in the data.
   2. Due to the small size of campaigns with goals between $10,000 & $50,000, I believe that a stacked bar chart would be a great visual for this data. With a stacked bar chart, we can see whether we are analyzing outcomes by category, by goal amount, or subcategory where the unbalances are and visualize the success and failure rates for each said category.

Statistical Analysis Questions

1. Use your data to determine whether the mean or the median better summarizes the data.
   1. I believe that the mean better summarizes the data due to the amount of high dollar donation or pledge campaign amounts. This better reflects how many campaigns have high pledged amounts.
2. Use your data to determine whether if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
   1. There is more variability in successful campaigns, I believe this does make sense because these campaigns were successful and higher dollar pledge amounts or outliers have a higher probability of showing in successful campaigns then failed campaigns. There is also roughly 200 more successful campaigns versus failed, which means there is a higher probability of Variance in the successful campaigns due to the varying and high pledge goals for certain campaigns.