*Francesca Palik*

*Module 1 Challenge*

*4/20/2023*

***Crowdfunding Report –*** ***Part 1***

***Given the provided data, what are three conclusions you can draw about crowdfunding campaigns?***

1. *First, and foremost, it appears that crowdfunding, as reflected in this sampling, may be a uniquely western concept, and even more pointedly, an American one. Our data is comprised from seven countries, the US making up 76% of that. The remaining countries, with the exception of China, are part of western Europe—or a direct descendant of western Europe in the case of Canada and Australia. It is noteworthy that the bulk of our project funding is from that of a capitalist economy, the US, and then followed by countries of a more socialistic nature. China, traditionally defined as communist, has the least representation. This makes sense in light of the wealth disparity between the rich and the poor and the abundance of abject poverty and homelessness in wealthy capitalistic economies with a popular political aversion to government subsidy/influence in the welfare of its citizens. Plainly said, the value and need for ‘charity’ arises from this disparity, i.e., an abundance of discretionary spending among a few amidst the presence of those with little to nothing. On the opposite end of the spectrum, a communist-built economy may not inspire nor manifest the need for grassroots charity among its people. Not that charitable means nor opportunity wouldn’t exist, but that it may not be as greatly relied upon nor developed over time as it is in an economy such as the US. With this in mind, it is fitting that China represents only 2.3% of the data. With unlimited time, further analysis could be done on the total amount pledged stratified by country which could lead to a rather simplistic, yet perhaps interesting discussion of ‘cultural generosity’ in relationship to a country’s known ‘standard of living ‘. In this case, the country’s standard of living for the given timeframe of the crowdfunding initiative would have to be aggregated to the current datafile.*
2. *A noticeable uptick in crowdfunding success starts in May, peaks in July, and peters out in August. This could make sense in light of the fact that our data is largely dominated by a country, the US, whose warmer months of leisure, travel and spending fall between May and August. These months, not only of warmer temps but also of longer days, allow for the expanded use of outdoor venues especially utilized by the more popular categories of theatre, music and food. They also represent a time prime for family vacation and increased extracurricular activity as often children and young adults are off from school. The winding down of summer, the start of school, and the change in weather could explain the sharp descent in success beginning in August.*
3. *According to the data, the overall ‘buzz’ or ‘popularity’ of a project does not seem to inspire a greater level of generosity among participants. Note, the average donation between successful projects and that of failed projects is comparable. This can be found by comparing the weighted average of total donations of successful projects with that of failed. Specifically, the grand total value of money raised for successful projects divided by the grand total number of pledgers of successful projects is 59.13, as compared to that for failed is 57.23. Now, bear with me, this brings up the topic of currency conversion, which is absolutely essential for this calculation to carry any true integrity and ‘publish worthiness’. I offer my simplified, and thus inaccurate, calculation – falsely assuming all currency the same, as analogous to what the numbers are telling us if a true currency conversion were included. If time allowed, that essential piece would be more than do-able. But for the overall objective and timeline of the assignment, I still felt it valuable to offer this ‘not so integrous’ discovery, again believing it to be analogous to the result begotten by the more accurate one with currency conversion. All to say, with our given data, we could possibly investigate and toss around ideas, albeit rudimentary and maybe even cynical, about the generosity of human nature, and the larger influence of the state of the economy over one’s decision of how much to donate.*

***What are some limitations of this dataset?***

*From the outset, this data looked, shall I say, a bit ‘wonky’, perhaps ‘jumbled’. For instance, I could not make sense of the company/blurb connection to the category & subcategory. Note, Harris Group (ID 5), heralds a ‘blurb’ of ‘Open-source optimizing database’. This text seems of no relationship to the category/subcategory of ‘theatre/plays’ it falls under. In a professional setting, when given data that appears to have no rhyme nor reason in content, it is a best practice to validate the file with its source (i.e., to simply ask ‘Is this ‘actually’ the file you intended me to work with?’). A dataset of curious, unexpected or incongruent content could be evidence of possible file corruption, an incorrect data aggregation or even a flawed data extraction process. Ensuring your file is legit at the outset can save both you and your client unnecessary time and expense in completion of the analytic assignment.*

*As well, the data lacks a quality of homogeneity and focus. Comparing crowdfunding success across varying countries of differing economies and cultural norms does not make for an effective sample on which to evaluate the success of crowdfunding enterprises. The absence of any detailed demographic data only adds to this. For instance, there is no individual, nor even group, pledge participant data like age, income, interests, etc. All would be helpful to provide meaningful insights into business intelligence and overall campaign performance assessments.*

*Not to mention, the sample population is small, and purportedly a ‘sample’ of ‘sample projects’, rendering any trends noted to be misleading or completely unreliable. Clearly, this sample is most likely not ‘representative’ of the larger populations of countries it includes. Statistical sampling methodology is a vastly complex subject based on mathematics as well as rigorous academic standards, and there is no reference to that which was applied to this particular dataset.*

*In conclusion, the data is grossly limited in its real-world analytic utility, but does indeed make for an extremely effective and interesting first assignment in the GWU Data Analytics Bootcamp.*

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

* *Box and Whiskers*

*The ‘box and whiskers’ chart would be a prime choice for analysis. As the lines extending parallel from the boxes, known as the “whiskers”, would effectively indicate variability outside the upper and lower quartiles. Outliers could be plotted as individual dots in-line with whiskers. With our data’s large variability and presence of outliers, a vertical ‘box and whiskers’ chart would serve us well in our analysis.*

* *The Pie Chart*

*I offer this up as an example of a chart type that would contribute little, if anything, to our crowdfunding analysis. Although it can be prepared with alluringly colorful segments to make an effective illustration of ‘parts’ to the whole, the prominent ‘parts’ of our whole are just too plentiful, namely ‘Parent Category’ and ‘Sub-Category’, to be able to provide the substantiveness needed for meaningful insight or actionable intelligence. With 9 categories, and even worse 23 subcategories, our pie would be divided up into piecemeal slivers. Perhaps using ‘Country’ would make for some interesting analysis, and yet the entirety of the pie would be largely the US, with the remaining six slices razor thin. In conclusion, the pie chart could be better used to illustrate the level of crowdfunding popularity among countries overall, but still not the best choice for our crowdfunding analysis as a whole.*

***Part 2***

***Use your data to determine whether the mean or the median better summarizes the data.***

*I would favor the median over the mean when summarizing this data. The variance of our data is quite large as evidenced by our values of 1,603,374 and 921,575 for successful and failed campaigns respectively. Note, the median serves as a better indicator when the distribution is either skewed or there are outliers present. We see not only from the high variances, but also in the large magnitude of range (i.e., the difference between the min and max values for successful and failed campaigns individually) that the presence of outliers is likely. A box and whiskers plot would effectively confirm this, as well as additional calculations of standard deviation and z-scores.*

**Use your data to determine if there is more variability with successful or unsuccessful campaigns? Does this make sense? Why or why not?**

*The variability between successful and unsuccessful campaigns is comparable (as referenced by the close variance values noted above). This could make sense if one considers that the structure, management and marketing of a crowdfunding campaign likely has a greater influence on the campaign’s success than merely the popularity of the search category it falls under. And yet, in light of the full extent of data limitations noted heretofore, I am hesitant to make even this assessment.*