Module 1 Challenge:

Crowdfunding Analysis

# Prepared by: Andrew Bourgeois; 20230613

# Background

Using the provided data set of 1,000 sample projects from various startup campaigns who sought funding through various crowdfunding sources (eg, Kickstarter, Indiegogo, etc), organize the data in order to analyze and determine any trends which might denote the success or failure of future campaigns.

# Interpretations from Statistics Tab

Q: Which better summarizes the data, the mean or the median?

A: In both “successful” & “failed” datasets, based on the min/max values, & the large variances, the mean is being pulled towards the higher skew, and the median is a better representation of the actual center of the values. In both sets ~ 70% of the data lies below the mean. Additionally, if you review the histogram and box and whisker charts, I added to the Statistics Tab, you clearly see a skew to the lower numbers in the data sets. Even when removing the outliers in the box and whisker. This again supports the median being the more accurate representation of centricity in both data sets.

Q: Is there more variability with successful or unsuccessful campaigns? Does this make sense? Why or why not?

A: Clearly the "successful" campaigns show a larger variance and standard deviation, by almost double, (even with only approx. 55% more successful campaigns compared to failed). Due to the larger range between min/max, the larger data set for "successful", and the nature of the data (all start-up’s goals and products are different) it does make sense you would experience a higher degree of variability. If we were to limit the data sets of each to business type and initial goal, I suspect we would see the variances decrease.

# 3 Conclusions from Data

Given the provided data, what are (3) conclusions we can draw about crowdfunding campaigns?

1. Overall, the full dataset showed more campaigns were “successful” versus “failed” or “canceled”. Although, this was not by much and tended to vary based on the month and year the funding campaign ended. (See Count\_Year tab, CrowdfundingBook\_Bourgeois).
2. We can see from the Goal\_Analysis tab the campaigns with an initial goal of 15000 to 34999 met with highest percentage of success, but it was the campaigns with initial goals ranging from 1000 to 9999 which saw the largest number of successful campaigns.A picture containing text, receipt

   Description automatically generated
3. The largest number of campaigns seeking crowdfunding by category fell under the parent category of “theater” and the subcategory of “plays”, with a total of 344 campaigns, 187 Successfully funded and 2 more currently in the process of seeking funding. The “journalism” category saw the least number of campaigns by far (only 4) but did realize a 100% success rate for funding.

# Some Limitations of the Dataset

1. First limitation I noticed is the currencies are not being converted to a common metric. This changes the dynamic of comparing goal/pledge/avg donation amounts which may appear to be equal between campaigns but are from different countries. For example, a 100 average donation in the US dollar is far different from a 100 average donation in Chinses yuan.
2. Ideally, I would like to see what the disposable income, populations, donor age groups and cultural preferences are for the various countries represented in the data set. It would be nice to see if campaigns in populations with less discretionary income showed a higher failure rate for getting funding. Additionally, does a cultural or age preference cause any funding to move from one category to another in certain populations/countries.
3. Lastly, any donor/company demographics might help uncover additional trending patterns not represented in the current dataset. For instance, are donors from a global pool or restricted to the country of origin for the start-up. Meaning are countries sending their investment dollar sot other countries causing more success in another country.

# Additional Tables/Graphs to Add, and the Value They Could Provide

1. The first additional charts I added were to the “Statistics” tab. I added a histogram and box and whiskers plot for both the “successful” and “failed” data sets. Both these charts help give a visual depiction in the skew of the data set, which helps outline the reasoning behind why the median is a stronger representation of centricity in both datasets.
2. The next table/chart I would add would be one to convert the currency amounts of each campaign to a common metric. This would give a more consistent apples-to-apples representation of the data when comparing volumes.
3. Lastly, a comparison of goal amount versus category/success rate. One would suspect technology funding would typically cost more and require either more donors or a higher per donation amount. It would be interesting to see if these constraints lead to a higher or lower success rate in each category.