Assignment #01: Crow Funding

Due Date: Sun, 10/2/2022

Student: Han Le

1/ Question: Given the provided data, what are three conclusions we can draw about crowdfunding campaigns?

The data gives us 1,000 samples of crowdfunding campaigns which include information of goal, pledge amount, the date that the campaign was launched, and closed, number of backer, and the category & sub-category each campaign used for fundraising.

The first pivot table/chart is showing the correlation between the outcome and the category the campaign used. Based on the table and chart, the successful rate is relatively high compared to the failed rate (56.5% vs 36.4%) which mainly comes from theatre (33%) and followed by film and video (18%) and music (18%). Out of the campaigns, journalism has the highest successful rate at 100% (with small sample), followed by technology (67%) then photography (62%). Out of the campaigns, game has the highest failed rate at 48%, followed by music and theater at 38%

The second pivot table and chart are showing the correlation between outcome and the subcategories used for the campaigns in the 1,000 samples. Based on the information, audio and world music campaigns are 100% successful (with small sample). It is followed translation at 67% then by animation, nonfiction, photography books and wearables at 62%. Science fiction has the highest failed rate at 64%

The third chart and table are showing the correlation between outcome and the month when the campaigns started. It looks like July has the most successful campaigns at 58 and January has the most failed campaigns at 36. August has the least successful campaigns at 41 and Sep has the least failed campaign at 23.It also looks like those more successful campaigns occurred from Jan to Jul compared to August to December

2/ Question: What are some limitations of this dataset?

Even though that the sheet includes $1,000 sample, but the number of samples is not enough to represent the whole population. Some categories and subcategories are relatively small (3,4) so it is tricky to come up with a precise conclusion.

In addition, appx 76% of the data comes from US and the rest comes from foreign countries making the analysis become tricky. The collected data could be inconsistent as different countries have different demographics, populations, incomes, etc. which may lead to different result for the campaigns. Also, the data did not mention the community (income, demographic etc) that the campaigns were made accessible to. This could factor into the pledged amount.

To my understanding, the pledge is what the donor promise to give which is not the actual amount that is raised for the campaigns so the final information could be different. This is based on my own experience from my job. I'd prefer to have data for the actual raised amounts for the campaigns so we can have a good comparison.

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

We could think about table or chart showing the length of the campaign and successful rate. The length of the campaign could tell us some information. It could show if the campaign last longer, more money can be raised/is pleaded leading the successful rate could be higher or vice versa.

In addition, we could think of a table or chart showing the goal ((threshold) compared with the successful rate and campaign type. If goal falls within tier, the successful rate could be higher. If the goal is too high (not unattainable), the successful goal could be lower

Bonus

1/Use your data to determine whether the mean or the median summarizes the data more meaningfully.

The median represents better than mean when the distribution of data values is skewed or when there are clear outliers.

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

There is more variability with successful campaigns compared to the failed campaigns based on the standard deviation (1,267.37 vs 961.31), The higher the standard deviation is, the more the data spreads over a wide area. Backers in crowdfunding are the people who pledge money, provide support to bring the campaign to life. Failed campaigns have very low number of backers or no backers at all and the variation is small. On the other hand, successful campaigns have very high numbers of backers to donate, and the variation is high.