Data Analytics Bootcamp Challenge #1: Crowdfunding data

**Liat Sayfan; December 2022**

—----------------------------------------------------------------------------------------------------------------------------------------

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
   1. Crowdfunding campaigns that focus on the arts (Music, theater, and film) are the most prevalent.
   2. Excluding “photography” which only have 4 campaigns, the most successful campaigns are “technology” ones with 67% success rate (64 out of 96 campaigns were successful). The least successful campaigns are “games” with only 44% success rate.
   3. Across the years, more successful campaigns happened in the months of June and July. The prevalence of failed campaigns is quite consistent over the months.
2. What are some limitations of this dataset?
   1. We can only calculate the average donation per person, but we don’t know the spread of the distribution of donations per each crowdfunding campaign. The average might be influenced by one or more outliers – i.e., extreme donation(s). For example, someone might have placed a $1,000 donation whereas all other donations were around $10. Thus, the average will not be a good way to represent a typical donation. More sample statistics would have been helpful – mode, median, standard deviation and range of contributions.
   2. If we want to have a better sense of crowdfunding efforts globally, we need data from more countries. The number of countries represented in the dataset is quite limited. There are a few European countries, but not a lot. Latin America is not represented and there are no countries from Africa.
   3. The currency used differ by country, so we cannot make valid comparisons between countries with regard to funds requested or received. For example, we cannot answer the question whether theater requests for funds are higher in the US vs. China because we don’t know the exchange rate at the time of each crowdfunding campaign.
   4. The dataset doesn’t include information on how long it took to get to 100% of the requested amount (for successful campaign). It would have been interesting to explore which campaigns get to their target faster.
   5. It would be interesting to know how media push relates to the outcome -Is the number of posts/pushes of the campaign on social media relates to the outcome? Does the platform used to advertise makes a difference (FB, Instagram, twitter, etc.)? Is the number of views of the campaign posts associated with the outcome? These kinds of data are not in the dataset. Therefore we don’t have enough information on what would help in making a campaign successful.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
   1. It would be interesting to see the distribution of funds that exceed the requested amount, by category and subcategory, to investigate which types of campaign get more enthusiastic responses. This kind of information will help people to set up a successful campaign.
   2. We can plot the average target amount and the average amount raised by category and subcategory to see which campaigns requested more funds and which once received more funds.
   3. It would have been interesting to see how long it takes for a campaign to get to the requested amount (by days or hours) - If this kind of information was available (see answer 2d above).
4. Number of Backer (part 2 of the assignment):
   1. In both the successful and the unsuccessful campaigns, the median is closer to the minimum value than the maximum one which means that the distribution is skewed, so there are probably outlier(s). Also looking at the standard deviation - the max value way higher than 3 standard deviations, which means that there are outliers in the distribution.
   2. There is more variability within the successful campaigns than the unsuccessful ones. This makes sense as the successful campaigns should have more backers in general and the spread of the number of backers could vary more. The unsuccessful one generally less backers, and the spread should be smaller, since many probably didn't receive many backers at all

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |  |
|  |
|  |
|  |
|  |