Crowdfunding Report

By: Maisie DeGraaf

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

1. Plays have by far the greatest number of crowdfunding campaigns within the past 10 years, accounting for 34 percent of all subcategories.
2. In terms of goals, having a goal greater than or equal to 50,000 can increase the percentage of failure more than half likely to fail for the campaign. The data set for this is large enough to compare to successful, and the only other data set that failed surpasses successful is between 10,000 and 14,999, but this data set has a very small number to compare, vs 50,000 or more.
3. The two most popular months to run a campaign is January and July, with July campaigns being more likely to be successful vs January.

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

Some limitations include the currency is in different currencies, which makes the goal vs pledged numbers inaccurate, as you cannot compare a goal in GBP vs a goal in DKK unless these numbers are converted to something standard. It is also difficult to tell reasoning for a failed campaign. Perhaps the individuals abandoned or forgot the campaign or had next to no advertising. Maybe some were already successful businesses previously, therefore had a wider audience to campaign to. It is also difficult when some subcategories or even parent categories have much more cases and data available to make assumptions about vs other categories, which may only have 3 or 4 campaigns in the past years. With this limited amount of data vs categories with hundreds of cases, it can be difficult to compare and get a fairer and accurate comparison.

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

Some other possible graphs we could create would be a box and whisper plot along with calculating the IQR, lower bound, and upper bound. After calculating this information, it shows that a lot of campaigns may be outliers, which can affect data. So, identifying these outliers can help show how many stay within the normal variance and which ones do not. One graph indicates the goal and how many with certain percentages reached their goal vs failed, but also indicating how much was pledged, and converting the currencies to a standard currency rather than separated currencies could help assist in seeing the true picture on how much people asked vs how much they were pledged in the different outcomes. This is especially helpful for failed campaigns, as it can show how many failed with zero donations vs how many were close to their goal and failed.

# Question 4: Use your data to determine whether the mean or the median better summarizes the data.

The median I think is much better than the mean to calculate and represent this data. This data is not a typical set, and using the IQR and upper bound calculations, each set has dozens of larger outliers. These larger outliers can skew data when calculating via the mean average. We also see many very low levels in the failed outcomes as well, which can affect the overall mean average. It also has many high-level outliers, which makes it difficult to calculate by taking all numbers and dividing.

# Question 5: 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 a higher IQR and a higher standard deviation for successful campaigns. This makes sense, since the criteria to be successful for a campaign may very a lot. Some campaigns may have a target of 500 or 50,000. But the common denominator for all of these is that they passed and could have continued to surpass the number of backers they accumulated. As for failed, they all share one thing and that is that they did not reach their goal. So therefore, they are more likely to almost always be a lower number. Large parts of the data indicate that many campaigns that failed had 0 backers. Calculating the 1st Quartile indicates the percentage if within 38, while the 1st Quartile for successful is 127.5. This shows that many cases fail within a similar range, which is a very low range.