Eric Johnson

10/01/2023

**Crowdfunding Report**

**Summary:**

I analyzed data from a sample of 1,000 crowdfunding campaigns from 2010-2020 from several categories and searched for trends in success, failures, time periods, and locations. Below are my findings.

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

1. **Successes and failures:** 
   1. Theater campaigns were the most successful in terms of total number but as a percentage they less often than average.
      1. Theater campaigns accounted for 34% of total campaigns.
      2. Theater campaigns succeeded 54% of the time, compared to an average of 57%.
      3. Theater campaigns cost on average $45,459.01.
      4. On average, theater campaigns funded at 101% of their goal.
   2. Games were the least successful as campaigns were expensive and failed at a higher rate than those of other categories.
      1. Games only accounted for 5% of total campaigns.
      2. Games succeeded only 44% of the time.
      3. Games failed 48% of the time.
      4. Games were the only category to fail more often than they succeeded.
      5. On average, games only met 71% of their goal, compared to 97% on average.
      6. On average, games cost $59,541.67 per campaign, compared to an average of $43,983.01.
2. **Time Period:**
   1. In general, campaigns began to succeed more as the decade went on. 2017 – 2019 saw an increase in successful campaigns and a decrease in failed campaigns.
      1. 2017-2019 averaged 67% success rates compared to 57% from 2010-2019.
      2. 2010-2016 averaged 54% success rates.
   2. 2016 had a higher-than-average number of failed campaigns.
      1. 2016 had only 49 successful campaigns with an average of 50% success rate, compared to an average of 56.5 in 2010-2019 (excluded 2020 since there were only two campaigns for that year).
3. **Location:**
   1. The United States had the highest total campaigns with both the most successful and failed campaigns.
      1. The US had 436 successful campaigns, more than any other country’s total campaigns.
      2. The US had 274 failed campaigns, more than any other country’s total campaigns.
   2. China had the fewest successful campaigns and the most canceled. Canda had fewer successful campaigns as a percentage of total.
      1. China had 23 total campaigns, 4 canceled, 6 failed, and 12 were successful.
      2. Canada had 44 total campaigns, 2 canceled, 19 failed, and 22 were successful.

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

1. This is a small sample of 1,000 crowdfunded campaigns. Smaller samples can skew results.
2. This set of data may be from one source(?) This could introduce bias if not counting crowdfunding from all available sources.
3. Certain specific categories of campaigns had very little data. For example, there were only 4 campaigns for journalism out of the 1,000 campaigns evaluated. This small sample size can provide misleading results.
4. With a small sample size there may be sampling bias in the population, especially given the number of campaigns in various countries and the many countries not represented.
5. There may be additional selection biases depending on the crowdfunding platform. Which countries is it available in? In which languages? Further research would need to be done on the source of the data.
6. There were only two campaigns listed for 2020, so while ten years of data were evaluated, the most recent year evaluated may have been missing information.

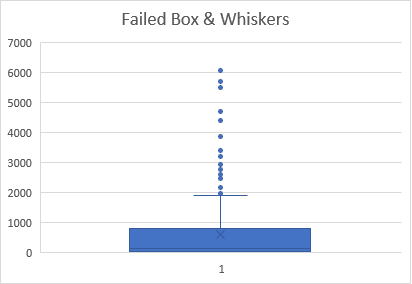
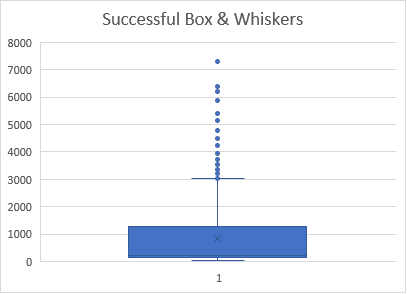
**What are some possible tables and/or graphs that we could create, and what additional value would they provide:**

1. An additional evaluation could be done to see if Staff Picks and/or spotlights impacted success rates.
2. Another evaluation could consider the time scale between the date created and the deadline dates to see if the time window for a given campaign made a difference on the outcome.
3. Additional evaluations could be made within the sub-categories for more insights. For instance, looking at which sub-categories within film & video were most successful.
4. Analysis could be run to determine if there is a statistically significant correlation between the goal amount and the success of campaigns.

**Statistical Analysis:**

**Question:** Use your data to determine whether the mean or the median better summarize the data.

**Answer:** The median better summarizes the data in this example. This is because both successful and failed campaigns have high variability in the number of backers. For successful campaigns the mean is 851.15 while the median is 201. The standard deviation is 1,267.37. The variance measures the distance from the mean and in this case it’s quite high. So, the median better represents because the variance is high. The graph below shows several data points outside the interquartile range.



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

**Answer:** The standard deviation for the number of backers in the successful campaigns is 1,267.37 compared to 961.31. The successful campaigns have more variability. And yes, I think it makes sense. This may be due to the number of backers increasing the likelihood of a successful campaign. I would hypothesize that as the number of backers increases, the more likely a campaign is to succeed, and with failed campaigns will tend to have fewer backers, that means the data range will be lower, closer to zero, and less varied.