Module 1 Assignment

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

1. First Conclusion
   1. The Arts Categories (film & video, music, and theater) make up approximately 70% of the crowdfunding campaigns of this data set. Drawing a conclusion that the ‘Arts’ use crowdfunding campaigns more than the other industries.
   2. With ‘Plays’ being the most likely to start a crowdfunding campaign at approximately 34% of all campaigns.
2. Second Conclusion-
   1. Approximately 57% of crowdfunding campaigns are successful (according to this data set). With no real outliers to this overall percentage of success/failure. Technology does seem to have a slightly better success rate than most other campaigns, with a success rate of approximately 66%.
      1. We cannot use Journalism as a true success rate (100%), as there were only 4 crowdfunding campaigns within this category. I would need to see more data in this category before claiming it as an outlier to the data set and that ‘journalism’ is the best when it comes to crowdfunding campaigns.
3. Third Conclusion
   1. In recent years: 2018-2020 the best months to start a campaign is:
      1. January: 15 for 26
      2. May: 10 for 14
      3. October: 12 for 16
      4. Novembers: 10 for 12
   2. The worst month to start a campaign is:
      1. July: 4 for 12 (over the lifetime of this data set it would be considered one of the best months to start campaign)

What are some limitations of this data set?

* The data set is relatively ‘old’ when you think about the fact there were only 2 crowdfunding campaigns in this data set, from 2020. This could be due to ‘COVID-19’ if it is truly looking at actual crowdfunding campaigns. If I was giving this information to a group, I would want to make sure it is known that this data could be considered ‘out-of-date.’ Over half of all the data is from 2010-2015, making it almost a decade old.
  + It would be more helpful to view data from 2021-2022 to see, if anything has changed, due to the world moving into a more technological world, due to COVID-19 as well.
* Another limitation would be if we tried to look deeper into each category and why it was successful or not. There is more data that we need to help make better conclusions before giving advice to a business on whether they should pursue crowdfunding.
  + Was the food truck Mexican, Italian, Greek, American, Desserts?
  + What genre was the best for plays? Comedy? Romantic? Historical? Musical?

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

I feel like there is numerous

* + Length of Crowdfunding Campaigns
    - This would give us a decent snapshot into how duration affected these campaigns.
  + Number of backers
    - This would help us gain more insight into how these campaigns became successful.
      * Shorter Campaign and Small Number of backers, likely had funding or knew who the donors were ahead of starting a crowdfunding campaign.
    - Number of backers vs. goal of failed campaigns. We’d be able to get insight into if the goals set we’re not realistic to begin with (I.e. some campaigns failed with 1000+ backers).
  + Spotlight
    - This would give us major insight into how important it is to be spotlighted and how it affected your success rate.
    - We could then compare being spotlighted to the number of backers, to see if that helped gain more backers for these crowd funding campaigns.

Statistical Analysis

*Use your data to determine whether the mean or the median better summarizes the data.*

At first, I was going to say mean because that is what I always use. However, I feel like median is a better tool for this kind of data due to the numerous outliers on both sides of success vs. failure. We have over 500 data points for success and 300-plus data points for failed campaigns.

The higher number of backers is going to skew our data set tremendously and not give us an identity of how many backers you should have for a successful vs. failed campaign. We must account for successful campaigns, going WAY over their goal which then in turn skews our average of an average campaign.

The median accounts for that allows us to see that campaigns are most successful when they cross 200 backers and more likely to fail when below 115 backers.

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

By looking at the numbers, the technical answer would be to say successful campaigns have more variability than unsuccessful campaigns. Successful campaigns have a higher standard deviation than unsuccessful campaigns and it also has an extremely higher variance.

It does make sense to me, as successful campaigns can keep collecting backers, even after they pass their goal which in turns skews the data. You’ve also got data points with extremely low backers that are still successful due to passing the goal. This is why we see almost double the variance number when comparing successful to failed campaigns (1606110.97 vs. 840721.10)