Andrew Koller

Module 1 Challenge

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

1. The majority (55.3%) of crowdfunding campaigns occur within the $1000 to $9999 range. After that, crowdfunding campaigns tend to exist in the >=$50,000 range (30.9%).
2. Theater crowdfunding campaigns are the most popular type of campaign, and account for 34.4% of total campaigns. This is followed by film & video, which accounts for 17.8% of total campaigns, and music, which accounts for 17.5% of total campaigns.
3. Campaigns in the $10,000 to $14,999 range were most likely to fail (56%), while campaigns in the $15,000 to $19,999, $20,000 to $24,999, and $30,000 to $34,999 were most likely to succeed (100%).

What are some limitations of this dataset?

1. While country data is provided, it would be beneficial to provide more localized geographic data, perhaps on a state-by-state basis, or city by city basis etc. Knowing campaign success rates by localized regions might help more geographically dependent categories such as theater, music, or food, for example.
2. It would have been nice to see website page interaction metrics for the “staff\_pick” and “spot\_light” data. From a top-down perspective, one might be able to determine that “staff\_pick” and “spot\_light” help drive traffic to certain campaigns, but it would be interesting to see exactly how that promotion is affecting campaigns on a case-by-case basis.
3. In some crowdfunding campaigns, a return on investment is offered for donations. For example, if one chooses to fund a theater play, they may receive tickets to the show. Or if one funds a music-oriented campaign, they may gain access to the music being produced. Usually, one must donate a set amount to earn this return on investment. Thus, it would be interesting to see:
   1. How many crowdfunding campaigns offered a return on investment.
   2. How required donation amounts (to receive a return on investment) affect average donation amounts.

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

1. Average Donation amount could be applied to many data sets:
   1. Average donation in relation to parent and sub-categories (bar graph)
   2. Average donation in relation to temporal data (line graph)
   3. Average donation in relation to success rate (bar graph)
   4. Average donation in relation to geographical data (bar graph)
   5. Comparison of average successful goal to average failed goal (line graph)

**Statistical Analysis**

|  |  |
| --- | --- |
| Successful | |
| **Mean** | 851 |
| **Median** | 201 |
| **Min** | 16 |
| **Max** | 7295 |
| **Variance** | 1603374 |
| **Std. Dev.** | 1266.244 |

|  |  |
| --- | --- |
| Failed | |
| **Mean** | 586 |
| **Median** | 114.5 |
| **Min** | 0 |
| **Max** | 6080 |
| **Variance** | 921574.7 |
| **Std. Dev.** | 960.0 |

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

* It is better to use the median in this case because the median is less influenced by extreme values, unlike the mean. The mean is much higher than the median because there are a few extreme outliers that skew the mean.

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

* The “Successful” data set has a higher variance and standard deviation; therefore, we could say that the “successful” campaign has more variability. Whether or not this makes sense is a matter of context. Statistically speaking, yes, it makes sense due to the variance and std. deviation. From a practical standpoint, however, it is unclear. Successful campaigns being more variable seems odd – one would think that success is marked by consistency, not inconsistency. The only thing that might be considered is that successful campaigns contain viral or more popular campaigns – if a campaign takes off in popularity, it’s likely to garner an extremely high volume of support. Failed campaigns don’t have an inverse equivalent where an extremely unpopular campaign lowers the average. In this context, then, viral crowdfunding campaigns are anomalous occurrences that skew the mean positively.