Emily Rusin

Challenge 1: Crowdfunding Analysis

Conclusions: 1. When reviewing a database of 1,000 sample Crowdfunding campaigns, the data was broken into 9 categories: film & video, food, games, journalism, music, photography, publishing, technology, and theater. The most popular category was theater with 34% of the campaigns. It had the most successful and the most failed campaigns. Film & video was the next category with about 18% and music was a close third with 17%. Despite these categories being the most popular with the most campaigns, the journalism category had a 100% outcome success rate, though it had a small sample (4 campaigns). Technology campaigns were the second most successful with 66%.

Conclusions from second pivot table: When reviewing the breakdown of the campaigns by the 24 subcategories, play campaigns accounted for the most of the crowdfunding efforts. This was consistent across the 7 countries where these campaigns originated. The subcategory of plays had the most campaigns, and the most successes and failures. For every subcategory, there were more successful campaigns than failed campaigns. Audio and World Music had 100% success rates, however they both had very small samples. The country of the United Stated had the highest number of campaigns (76%).

Conclusions from third pivot table: With all data categories and years aggregated, May to July is were the peak months for successful campaigns. The campaign successes lulled during September to December. January, May and August had the most failed campaigns. Almost every category seemed to peak during warm or summer months.

Limitations: The data has some limitations in that the description is very short and cannot be understood therefore it is hard to decipher what the campaign was for. Another problem is that the goal, pledged amount, and average donation are in the currency of the country, therefore its not a uniform or accurate comparison.

Additional tables/graphs: It would be useful to create a column by categories/subcategories to show the percent of failed/successful campaigns. Creating this metric and then graphing it would help identify which campaign categories and subcategories are most likely to succeed based on past results. I think it would help to understand this data if we graphed how many days a campaign was active vs. its success rate. Does a campaign get more backers if its open longer? Or does that not impact the number of backers and its more dependent upon other variables. We could see if the distribution is symmetrical.

Statistical Analysis

The median is a better summarization of the both the failed and successful datasets. For Successful campaigns, the range in values is between 16 and 7295. When the data is graphed in a bar chart, it is observed that the number of backers has a gradual increase to the median (201), but the data has a more steep increase afterwards. When the frequency of occurrences is graphed, the distribution is not symmetrical.   
  
For the failed campaigns, there are a number of outliers in the data set and as with the successful dataset, the backers is relatively flat but has a steep increase after the median. This is also not a symmetrical dataset.

There is more variability in the successful campaigns. I think this makes sense because it is a much larger dataset so there are more data points and more room for outliers to exist. The range is larger as is the variance in the successful campaign backers dataset.