Excel Homework Analysis Steven Gaetz

According to the data, the Theater category had the greatest number of counts total. Theater had about twice the number of campaigns than the next populous category, which was music. Theater also had the greatest number of successful campaigns, followed by Music, and then Film and Video. Theater also had the most failed campaigns, followed by Technology. Technology had the greatest number of canceled campaigns. Journalism was the only category with only canceled campaigns. About 1000 campaigns in the Theater category were Plays, nearly 700 of which were successful and about 350 of which were failures.

Looking at the counts of state over time, the number of successful campaigns spike in May and trend downwards for the rest of the year with a small upswing in the months of October and November. The number of canceled campaigns remains steady over the whole year. Failed campaigns tend to increase from April to July and overtake the number of successes in the month of December. December also has the least number of total campaigns, while May, June, and July have the highest total number of campaigns. Looking at outcomes based on goal amounts, the number of successful campaigns decreases as the goal amount increases, from 71% for less than 1000 down to 19% for greater or equal to 50000.

A limitation of this data set is that it is an observation study and not a randomized experiment, therefore statistical inferences of cause-and-effect relationships cannot be drawn. From the data, it is also impossible to tell why a campaign has failed, other than that it did not reach its goal.

Another graph we could make, which I have done, is the distribution of the number of backers by state. This shows how right skewed this data is, especially the successful campaigns. This means there are very few counts of campaigns with high numbers of backers. Another graph I made is a box plot of the number of backers by success or failure. It is hard to even see the interquartile range, however. I made this because the median of each seemed too small, but the box plots supported the results of this.

In the bonus statistical analysis part of the homework, the median summarizes the data more meaningfully because the data are so skewed to the right. Since there are a high number of outliers, they are driving the average up for both the successful and for the failed campaigns.

There is more variability with successful campaigns. This makes sense because there are lower numbers of backers for failed campaigns and there is a greater range for the number of backers of successful campaigns. This is clearly illustrated by the graph of the distribution on my Distribution worksheet.