Exercise One Report

Three possible conclusions that can be made from the data set:

* Overall campaigns had a success rate of 57% when removing the live campaigns since is still uncertain how those campaigns will turn out. Campaigns were most likely to be successful from June through November with a success rate of 59%. The exception to this was the month of August which had the lowest success rate overall at only 49%. If August is removed, the months of June through November have a combined success rate of 61%. The months from December through May have a combined success rate of 55%. This tells people looking to start a campaign that it is best to start a campaign between the months of June through November as long as one avoids starting the campaign in August. The two best months overall to start as campaign would be June or July with a success rate of 64% and 62%.
* The categories of theatre and film and video were the two categories with the highest number of campaigns. However, they were not the most successful campaigns with success rates at 57% and 59%. The most successful campaign category by a large amount, albeit with a very small sample size was journalism with a success rate of 100%. The second most successful campaign was technology with a 68% success rate while also having a much higher sample size than photography. The worst performing categories were games and food with games having a success rate of 47% and food at 48%.
* There was not a major difference between the campaigns that were spotlight and the ones that were not. The campaigns that were spotlight had a success rate of 57% and the ones that were not spotlight had a success rate of 55%.

One of the limitations of this data set is that it is only considering if a campaign is successful or not, and not including other potential factors such as goal size and amount pledged. One campaign could raise substantially more money and fail than another campaign that succeeded if that campaign had a higher goal. Another limitation is that the sample size varies significantly between categories. The results say that the most successful campaign is journalism with a 100% success rate, however this is misleading as less than one percent of the data are campaigns pertaining to journalism. Conversely, over 30% of the data are campaigns pertaining to theater.

Some other potential tables and graphs that could be created would be looking at the pledge and goal amounts and seeing how that would pertain to the number of successes. A pivot table could also be made of the spotlight and staff picks to see how impactful those are on the success rate.

Analysis

When looking at the number of backers for successful and unsuccessful campaigns, it would be best to use the median to summarize the data. Looking at a histogram of the number of backers, it can be seen that the data is skewed to the right. The side-by-side box plot of backers shows that there are several outliers above the upper fence. When data is skewed and has outliers, it will pull the mean away from the center making the median a more accurate measure. The mean should be used when data is normal and free of outliers. The variance is greater in the successful campaigns than in the unsuccessful campaigns. This makes sense given that the campaigns that are successful have a wider range of outcomes. This can be seen by subtracting the minimum from the maximum in both successful and unsuccessful campaigns.