Crowd funding Solution

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

Looking at the parent category, when you look at each outcome, successful, failed, live and cancelled, we can see that overall the campaign was successful in most of the categories with exceptions from the games which has failed at 23 and are successful at 21. Cancel and live are relatively low in all the parent categories.

The same trend or picture can be seen when you look at the sub-category and outcome. We can see that successful outcomes were greater in all the categories with the highest plays recording the highest at 187 and word music recording the lowest at 3. Which is in line with the given observation under the parent category.

When we look at the categories from month to month, we can see that successful and failed outcomes have a similar trend from January through to April, from there we see successful outcomes remaining the same from April to May then increasing from May to July and then reducing again in August and remains almost at the same level from then onwards.

However, the failed outcome continues to show the same up and down trend through to September where it goes to its lowest level and then begins to increase from then onwards. The canceled outcome remains relatively low throughout all the months.

1. What are some limitations of this dataset?

The limitations include

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

The other table that would be added is a table that shows the data on country and pledges and outcomes. This would have helped to look at which country had the highest successes, failed and canceled. I would also like at the scatter plot with pledges and outcomes to see if there is a relationship between the two.

I would also look at the relationship between country and outcome to see if there could be a story to tell there, and then draw a conclusion.

Last Question

Here the median is better because this data is skewed to the left and therefore, the median is the best method of measure of central tendency

Given the statistics above, there is large variability in the data, we can see that the median is 201 and the variance of either the population or the sample is above 1603373 and 1606216 respectively.