**Written Report/Statistical Analysis**

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

The median better summarizes the data in this set. The median is 201 or 1115 the mean is 851 or 586. Meanwhile the maximum is 7295 or 6080 and the minimum is 16 or 0. Meaning there are as many campaigns who had 16 to 201 backers or 0 to 115 as there is campaigns with 202 to 7295 or 116 to 6080. The variance and deviation are way higher on the side above the median in both instances. This suggests there are major high outliers skew the average taking it farther away from what a typical crowdfunding campaign backing would look like.

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

There is more variability and yes it makes sense. It makes sense because a failing campaign has a limit. The lowest it can go is having 0 backers it cannot go into the negatives. Where a successful campaign has no ceiling of where it can have too many backers. That means a few very successful campaigns can create have a huge backer count and that is going to skew the data to have more variance. Additionally, if the crowdfunding goal is low or someone just donates a ton a campaign can still be successful based with a small number of backers. It cannot be 0 but it can be in the 30’s which is nothing compared to how a successful campaign has the potential to have hundreds more backers than it needs.

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

The number of successful projects hits its peak in June and July.

By far plays are the subcategory that has the most attempts at successful crowdfunding campaigns.

Crowdfunding campaigns that have a goal between 15000 to 34999 had the best success percentage except for a little dip between 25000 and 29999 that still has a higher success rate than any other range excluding 1000 to 4999.

* **What are some limitations of this dataset?**

The limitations of the data set are we have no real context to how the individual campaigns were advertised. We don’t have any idea how many people worked towards advertising the crowdfunding. We don’t have any idea what group has previous experience with crowdfunding versus who is doing this for the first time. We also currently don’t know what subcategories and categories have higher and lower fundraising goals. We could calculate this but we do not currently have it.

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

We could create percentage successful failed graphs for subcategory and category. Right now, we only have evidence of what categories tried crowdfunding more as you would expect with more attempts you would likely get more successful attempts. A percentage would allow up to what the likelihood is a certain category or subcategory would succeed. This could just be put in the pivot table next to

I would also like to see a line graph of percentage success based on month. This would allow us to see which months led to the crowdfunding being successful. Whereas number of successful campaigns is skewed by the fact that more campaigns are attempted in a lot of the months that have the most successful crowdfunding.

A pivot table that shows the average number of backers per successful crowdfunding campaign and number of backers per failed campaign according to the goal. This will allow us to figure out how much the number of backers leads to success or failure. Or is the amount they give more important. It is likely somewhere in the middle but testing that would be helpful.