Rebecca Tricker | Excel Homework

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

When looking at successful campaigns per category, while theater-based campaigns have been the most successful, they also have the most failed campaigns. Looking ahead, music-based campaigns might be better to conduct. Past music campaigns have provided a high number of successful campaigns with a low number of failed campaigns.

The sub-category of plays preforms the best in the theater category compared to spaces and musicals.

Campaigns that were successful performed the best during the month of May and performed the worst in the month of December. During the spring and early summer months, the number of unsuccessful campaigns tend to grow, but slow back down slightly in the fall.

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

The total number of projects launch on Kickstarter have been more than 300,000. Looking at the dataset only 4,000 past projects are being analyzed. Since not all the data is given, we could be drawing false conclusions based on the data we were given.

The 4,000 projects that have been analyzed in the dataset could be skewed based on how they were picked from the original 300,000. It would be best if they were randomized, but we aren’t sure that this is the case. More details about how this data was collected would be helpful.

While this dataset has the country of the project, it doesn’t give the specific location. Having specific states listed like in the US, we could look at consumer preference data per state.

1. **What are some other possible tables and/or graphs that we could create?**

One table and graph that could be created is one that focuses on the date created and date ended in order to see if there are any trends in the success of projects compared to how long they run.

A second table that could be created could look at if it was a staff pick and the state of the campaign. If there is a large correlation between unsuccessful campaigns and staff picks being true, Kickstarter might want to look at eliminating staff picks.

A third table and graph that could be created could look at the sum of back counters and the parent category and sub category.

1. **Use your data to determine whether the mean or the median summarizes the data more meaningfully.**

After reviewing the data, I believe that the median summarizes the data more meaningfully. Given our high level of variance for each campaign, using the median will eliminate some of the outliers that are causing the high variance. If we used the mean instead, the outliers would be factored into the data, which could skew it one way or another. The median is less affected by these outliers.

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

Looking at the data, there is more variability with successful campaigns because the variance calculated is much larger than the variance calculated for unsuccessful campaigns. When looking at variance, the larger the number the more variability within the dataset. I believe that this does make sense because the number of backers counted in the successful campaigns varied widely across the board. The numbers were more spread out compared to the unsuccessful campaigns. Successful campaigns created a more widespread dataset.