**HOMEWORK 01 – EXCEL**

Main Data

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

First, theater is the most popular category for Kickstarter campaigns, and also the most successful. Breaking it down to sub-categories, plays in particular have the most Kickstarter campaigns. However, it should be noted that while theater (plays) are most popular and successful, music is actually more successful if looking at the proportion of successes out of total music related campaigns. Even more specifically this is driven by two subcategories of music; rock music and classical music have 100 percent success rates. Technology seems to be the most variable category, having almost equal successes, failures, and cancelations. When accounting for proportion (percentage), technology is also the category with the most failures. Second with respect to country, the U.S. has the most Kickstarter campaigns (overall and most successful, unsuccessful, etc.). Third, in regard to time of year, Kickstarter campaigns all together appear to be somewhat more successful during the summer (around May/June), and are the least successful during the winter (around December/January).

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

There seems to be an overrepresentation of Kickstarter campaigns in the U.S., thus, this data may not be representative of other countries. The data also seem to be largely driven by the values for theater due to the sheer number of campaigns in this category. This heavy influence can be seen once you start to look more closely at the categories individually. For example, while Kickstarter campaigns overall appear to be somewhat more successful in the summer, technology does not seem to show a clear pattern of success based on time of year, games actually appear to be more successful during the winter and spring months, and least successful during the summer, and music is relatively steady with a slight decline in success in winter. Thus, the main limitation is that the aggregate findings may be skewed because of the larger numbers represented by plays (theater), and these findings may not necessarily represent non-U.S. countries.

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

To account for the overrepresentation of number of theater related Kickstarter campaigns, another table/graph that could be examined would be one showing percentages of failed/successful campaigns per each country and per each category/sub-category. By examining the percentages at each of these levels, it would take into account proportions as opposed to raw totals, which are skewed towards theater. Second, another way to account for the overrepresentation of Kickstarter campaigns in the U.S. would be to combine non-US countries into one group and to create a table/graph that compares this new combined group directly to the U.S. This provides more statistical power for the non-U.S. countries and may show findings that are more representative outside of the U.S. Finally, if the data are NOT truly reflective of Kickstarter campaigns globally, meaning that this data do not represent non-U.S. countries, non-U.S. countries could be removed as outliers, and new tables/graphs examining the state of campaigns by category could be examined for just the U.S thereby providing a more direct and less noisy examination of data for the country that seems to be best represented.

Bonus Statistical Analysis

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

The median summarizes that data more meaningfully because there are some large outliers in the number of backers for both successful and unsuccessful Kickstarter campaigns. For example, looking at the successful campaigns, a couple of projects had over 20,000 backers, but most campaigns had well under 5,000 backers. The mean is more easily skewed by outliers. Given that the variance and standard deviation for both successful and unsuccessful campaigns are quite high, suggesting the prevalence of outliers and lots of variability, the median is probably the more accurate measure of central tendency.

**2. 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 with the successful Kickstarter campaigns as compared to the unsuccessful ones. This makes sense when considering outliers and also the sheer number of backers drawn to each type of campaign. The successful campaigns have much larger outliers (i.e., reaching over 20,000 backers for at least two campaigns, while the largest outlier for failed campaigns is around 1,300 backers). This is a big discrepancy. A possibility is that even though many successful campaigns bring in around several thousand backers, a few types of successful campaigns may draw in an extremely high number of backers, creating much more variance. On the other hand, failed campaigns seem to have a ceiling point/cutoff point; a point at which they are not likely to draw in more backers. I would say this is around 1,300 or so backers. This makes sense given that failed campaigns would be ones that are not attractive or enticing to as many people. Further, when looking at successful campaigns, ignoring the few that draw in over 20,000 backers, it appears that many of these have between close to 1,000 and 4,00 backers. Looking at the failed campaigns, many fall below 200 backers. Thus, successful campaigns bring in a higher number of the backers overall, and, some of these bring in an unusually high number of backers, whereas failed campaigns seem to reach a limit of backers (that is much lower). This creates more variability in the number of backers for successful campaigns as numbers have more room to vary, whereas the numbers for failures seem to be limited to a point below ~1,300 (giving much less room for variation).