Week 1 Homework: Excel Challenge

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

Conclusion 1: At the category level, it can be concluded that groups involved in the arts seem to be most popular. The 3 largest successful categories are: theatre, music, and film & video. Although the technology category had more campaigns than the film & video category, its success rate was lower. Journalism has by far the least number of campaigns altogether of which all were cancelled. All categories except the top 3 mentioned previously have a greater categorical failure rate than success rate.

Conclusion 2: At the Sub-category level, it is clear to see that plays make up the majority of successful campaigns. This sub-category falls under the most successful category of theatre. We can see more detail at the sub-category level which groups contribute to the success or failure of the main category. A lot of sub-categories such as documentary, hardware and rock music have a 100% success rate. Others such as animation, drama and video games have a 100% failure rate. It is clear to see the distribution of successes and failures at the sub-category level, giving us more insight as to what people are willing to fund and what makes the primary categories more appealing to society.

Conclusion 3: Over time, there seems to be a correlation between the success and failure of campaigns. In the first quarter of the year there is a negative correlation. As more projects succeed, the less projects fail. Although there can be many reasons for this, one possibility could be the willingness of people to contribute to causes close to them after the holiday season and starting the new year off by helping an organisation of their choice. The second quarter sees a spike in successful and failed campaigns with all three states spiking positively and decreasing in the last quarter.

1. What are some limitations of this dataset?

One major limitation is that we don’t know the amount each individual backer has pledged. This could give us a greater understanding of how much a backer is willing to contribute to a campaign and possibly accept the hypothesis of greater backers means more pledges.

Another limitation to this data set is not knowing if all categories and sub-categories was included. This means that some information that was potentially critical to the analysis of this data may have been absent, leaving analysis not true and incomplete. A guarantee that all data was present would provide certainty that this data set is a fair representation of categories and campaign successes/failures.

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

One graph that would be interesting to see is how much all the countries involved contributed. This will give us an idea of where the Kickstarter programme is popular and how to improve it in other countries to make its campaigns more successful.

One table that would help fundraisers would be to see how much their goal was exceeded by. This would help in future campaigns as they would be able to determine how much people are actually willing to contribute to certain campaigns. They can increase their goals accordingly and possibly get an increase in number of backers and/or an increase in pledges.

Bonus Statistical Analysis

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

I think the mean summarizes the data more meaningfully. The reason I believe this is the case has to do with the fact that this data set is quite large. The mean averages out the number of backers, which would take care of outliers better than just looking at one number in the middle of the data set. There could be a big jump between numbers and the mean would give us a better representation of the groups (filed and successful).

1. 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 a greater variance in the successful campaigns. This makes sense as there is a larger sample size of campaigns and a greater number of backers overall. The way that the backers are distributed between the different campaigns are naturally going to be more varied due to the data set being significantly larger.