**Week 1 Homework - Excel**

*Q1. Given the provided data, what are three conclusions we can draw about crowdfunding campaigns?*

First of all, the theatre received the highest funding, therefore contains the highest number of successful, canceled, and failed campaigns.

Plays, which are part of the theatre category, have the highest success rate.

The highest number of projects were launched in July, which also showed to be one of the months with the highest success rate.

*Q2. What are some limitations of this dataset?*

The project presents a wide amount of data, however, it does not take into consideration external factors, such as the power of marketing which would help projects to be more successful.

In addition, it did not consider socio-economical factors such as the wealth of those funding the projects. For example, theatre is known to be popular among the most prestigious and elite within our society, and therefore those who can financially contribute to projects without suffering.

*Q3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?*

It would be interesting to analyse which types of challenges were more prone to failure and which were more prone to success. The success rate could be defined by the type of challenge and not the category.

We could do a pivot table with the type of challenge vs the number of successful/failed, with categories below.

It would so be interesting to check whether the currency per project had an impact on the overall success rate too. For example, would be nice if most projects with USD as currency were more or less successful compared to the ones with GBP.

Lastly, it would be interesting to see what the budget between each category is. For example, with theatre, the most funded and successful category, what is the goal range compared to journalism which had the least amount of projects.

*Q4. Use your data to determine whether the mean or the median summarizes the data more meaningfully.*

Usually, the mean is used to measure tendency as it uses all values in the data set to give you an average, however, since the dataset in this document is wide and skewed, the median does summarize the data more meaningfully.

*Q5. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?*

Such a high standard deviation and variability mean that the values within the range are not consistent. Despite the variability with unsuccessful campaigns being lower by 25% circa compared to the successful campaigns, and therefore less variable within the range, it is not a good enough value to define it as not variable. If the values within the range were similar to each other, then looking at the variability would make sense. But in this case, it is not accurate.