**Data Analytics Boot Camp – February 2023 Intake**

**Module 1 Challenge – Crowdfunding Data Analysis**

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

*Looking at the data and the first bar graph (which compares parent categories to outcomes), one can conclude that the success rate for crowdfunding projects is just over 50% with the majority of projects being focused in the film & video, music, and theatre category. The success rate of 50% - 60% is evident across all countries being analysed.*

*Secondly, referring to the bar graph which compares sub-categories and their outcomes, it can be deduced that the plays sub-category has the majority of crowdfunding projects. This sub-category is the highest across all countries.*

*Finally, looking at the third line graph which shows the outcomes of crowdfunding projects in comparison to time (months in a year), an obvious relationship can be drawn between successful and failed/cancelled crowdfunding projects. As the number of failed/cancelled projects increases, the successful projects drop – basically, an inverse relationship. This is visible in the July-August months where a drop in successful campaigns is due to an increase in cancelled/failed campaigns.*

What are some limitations of this dataset?

*All that the data shows are number of crowdfunding projects, their outcomes, and the pledges that were made. Apart from analysing success rate, we are not able to fully illustrate the reason why some have been successful and the others unsuccessful, especially the lack of data to show what factors come into play for the various categories of the campaigns.*

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

*I would create a graph to see the success rate of campaigns in relation to the staff pick or spotlight field. I assume that staff picks and spotlight campaigns would have a better rate of success due to the exposure and visibility to potential backers. This assumption can be tested by creating such a graph.*

*Additionally, a graph to illustrate the relationship between goal vs. pledged and aggregate donation to the campaigns would shed some light whether the more expensive campaigns have a higher rate of failure or not. Again, assuming microeconomic principles are valid in the crowdfunding space, if it is cheap to back, there will be more backers.*

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

*Looking at the mean, median, and mode values, the data is right skewed/biased and the mean is not the appropriate parameter to summarise the data. In this case, the median better summarises the data.*

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

*As the variance value is higher for backers of successful campaigns, there is more variance from the mean for successful campaigns. Yes, it does make sense to me – in order for a successful campaign, the number of backers need to be above a certain amount, the goal. There is no limit to the maximum number of backers for a successfully campaign. In other words, there is no upper limit which can lead to outliers in the data. One such example is the “Odom Inc - Managed bottom-line architecture” campaign which had 1040% backing rate. These types of outliers skew the data. However, for an unsuccessful campaign, the number of backer’s is limited between 0 and the goal amount (0%-99%) and thus their freedom to get to extremes is limited.*