**Crowdfunding book questions**

1. The category “theatre” has the highest count (187) of successful campaigns
2. The sub-category “plays” has the highest count (187) of successful campaigns
3. The highest number of successful campaigns started in the month “July”
4. This dataset is limited as it largely just looks at the "count" of outcomes. For example, Theatre has the highest count of successful campaigns but also the highest count of failed campaigns. Since Theatre has the largest overall number of campaigns, comparing outcomes between categories solely based on count outcome can be misleading.
5. I would add extra columns on the original data set to calculate percentages of cancelled, failed, live and successful outcomes. These percentages can be calculated by dividing the count of each outcome by the grand total of outcomes. The create pivot tables and charts to visualise; Category vs Percentage of Outcome, Sub-category vs Percentage of Outcome and Date created vs Percentage of outcome. This would offer insights into how category, sub-category, and date created can influence the success rate which is not skewed by different grand totals in each row label.

**Statistical Questions (also included in a text box on relevant excel sheet)**

1. Since the means are considerably larger than the medians, it suggests that this data contains extremely high values and outliers which are skewing the mean upwards. Therefore the median be a better measure of the central tendency and summary of the data as it is less affected by extreme values.
2. The data shows that successful campaigns exhibit higher variability. This does make sense because successful campaigns had a range of different funding goals, requiring a varying number of backers to meet each unique goal. On the other hand, failed campaigns may have more consistent backers counts, reflecting their inability to attract significant support regardless of the funding goal.