**Module 1 Challenge**

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

1. Journalism has a low number of new products, but it has a 100% success rate, indicating that there may be a loyal and supportive audience for this type of content.
2. Theatre has a high number of crowdfunding campaigns, suggesting a demand for this type of content, but it only has a 54% success rate, indicating that it may be a competitive market or that there are other factors that contribute to success beyond demand.
3. Crowdfunding for plays is more prevalent than for television, indicating that theatre may be a more popular medium for crowdfunding than television.

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

However, there are limitations to this dataset that should be considered. Firstly, the data only covers a specific time period (2010-2020) and may not accurately reflect current trends. Secondly, the dataset includes a limited number of variables and may not capture all the factors that contribute to campaign success or failure, such as timing of product release. Finally, campaigns that are staff picked or featured in a spotlight have a higher chance of success, which may skew the results and make it difficult to generalise findings to all crowdfunding campaigns.

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

Two additional tables/graphs that could be created to gain insights into crowdfunding campaigns are:

1. A histogram showing the distribution of goal amounts for successful and failed campaigns: This type of graph would provide information on the distribution of goal amounts for both successful and failed campaigns. It would help to identify any patterns or trends in the data that suggest certain goal amounts are more likely to result in a successful campaign.
2. A bar chart showing the top categories and sub-categories for successful and failed campaigns: This type of graph would provide insights into which categories and sub-categories of campaigns are most successful on the platform. It would help to identify any patterns or trends in the data that suggest certain types of campaigns are more likely to result in a successful campaign.

However, it's important to note that there may be limitations to this dataset, such as missing or incomplete data, bias in the sample of campaigns, and potential confounding variables that may affect the success of a campaign. Additionally, the conclusions drawn from these tables and graphs should be interpreted with caution and not used to make definitive claims about crowdfunding campaigns.

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

We may compare the mean and median figures for both successful and unsuccessful campaigns to see which best summarises the data—the mean or the median. It is possible that certain extreme values (outliers) are distorting the mean if the mean is noticeably greater or smaller than the median. The median might be a better summary statistic in these circumstances.

The median and mean number of backers for successful campaigns are 201.5 and 852.38 respectively. This shows that some efforts with a sizable following are having an impact on the mean. The median might be a more useful summary statistic in this situation.

The median and mean number of backers for unsuccessful campaigns are 115 and 587.23, respectively. This shows that some campaigns with a sizable following may also be having an impact on the mean. The median could potentially serve as a superior summary statistic in this situation.

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

We may examine the standard deviation and variance figures for both groups of campaigns to see if there is more variability with successful or unsuccessful campaigns. The variance and standard deviation quantify how widely distributed or variable the data are. Higher values for the standard deviation and variance imply that the data is more variable.

The variance is 1608214.66 and the standard deviation is 1268.15 for effective campaigns. The variance is 925716.28 and the standard deviation is 962.14 for unsuccessful campaigns. This shows that the data for effective campaigns is more variable.

This may be understandable given that successful campaigns typically have a larger number of supporters, and that this number might fluctuate greatly depending on the popularity and appeal of the campaign. The average number of backers for unsuccessful campaigns may be lower, which may lead to less variability.