Excel-Challenge Report

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

* The outcome of a campaign to be successful, cancelled or failed relies a lot on how many backers they are able to receive. The more backers the campaign is able to get the more likely their project will be successful and the less backers a campaign receives the less likely it will be successful.
* If you wanted to start a campaign I would suggest it to be in the category of theatre and more specifically “plays” as theatre had the most successful campaigns and plays was the most successful sub-category.
* Campaigns seem to have the most success during the early months of the year and peak around spring time and then we start to see a drop-off of successful campaigns and an increase in failed ones around fall season.

What are some limitations of this dataset?

* Some limitations of this dataset is that there are more data for some campaign categories than others (example. There is 1393 campaigns for theatre and only 24 for journalism), this skews the data and favours theatre campaigns and does not let us to make accurate inferences on some campaigns.
* The same limitation applies for countries as some countries have a lot more campaigns than others which makes it hard to make accurate observations as total campaigns for each country is not equal. The country that has the most campaigns is going to have the most succesfull and failed campaigns because it has the most data. Would the data remain true that US had the most succesfull campaigns if all the countries had the same amount of total campaigns??

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

* We could use pie charts for each state/outcome of the campaign and primary categories. This would allow us to visualize what percent of each category makes up the campaigns outcome.

**Bonus statistic analysis**

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

* I would use the median to summarize the data because there are outliers in the data and outliers have a greater influence on than mean than the median. The outliers will skew the mean higher or lower which then would not accurately represent 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?

* There is more variability with successful campaigns because more backers are involved in successful campaigns and there is more data for successful campaigns. If we looked at the correlation between successful campaigns and amount of backers it would show that the more backers there are the more likely the campaigns would be successful and vice versa for unsuccessful campaigns (the less backers the less likely the campaign will be successful). So since succesfull campaigns tend to have more backers than unsuccessful campaigns it would make sense why there is more variability with successful ones because there’s more data points included and there’s a larger range of backers involved with successful ones than unsuccessful ones.