Excel Challenge Report

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

1. With 2185 successful campaigns on Kickstarter and 1879 unsuccessful campaigns (adding together failed and canceled campaigns), we could conclude from this that a Kickstarter campaign is more likely to be successful than unsuccessful.
2. As the subcategory with the most successful campaigns are “Plays”, we could conclude that Kickstarter users enjoy going to see plays and are willing to fund them.
3. It is a bad idea to start a campaign in December, as it is the only month where the “failed” campaign total is higher than the “successful” total.

# What are some limitations of this dataset?

We have been provided with just over 4000 campaigns; Kickstarter likely has many more that could drastically change the results that we were able to reach.

The newest campaign was created in 2017, Kickstarter has likely developed a lot more mainstream use since 2017 and could have a wider spread of both users and types of campaigns since this time, which would change the outcomes that we have.

We are unsure about a lot of the details about the backers, having demographic information could change the perspective on what we conclude from the data, i.e., if they are younger, they might be more interested in technology or less interested in some types of music.

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

Like the category and subcategory table/graphs we created, we could table out “country” vs “state”, to find where most campaigns are created.

By adding “Years” to the rows in our line graph we created, we can get an overview of the uptake of Kickstarter over time and whether a campaign has become more or less likely to be successful over the platform’s lifetime.

Using the “staff pick” and/or “spotlight” columns in the data, you could have one of them vs “state” to try and draw a conclusion as to whether giving a staff pick or spotlighting a campaign makes it more likely to succeed.

Bonus

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

Median would be the better measure of central tendency; this is due to a few large outliers in both datasets. This can be determined by viewing the “count” of backers\_count, which we can see that most campaigns that have a low backer count and few that have a very large amount. This can be easily seen through a visualisation.

# 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 in successful campaigns from the calculations. This could be expected due to the higher number of successful campaigns when compared to unsuccessful campaigns.

Another factor of this higher variability would be due to the range of the successful dataset being much larger than the unsuccessful set.