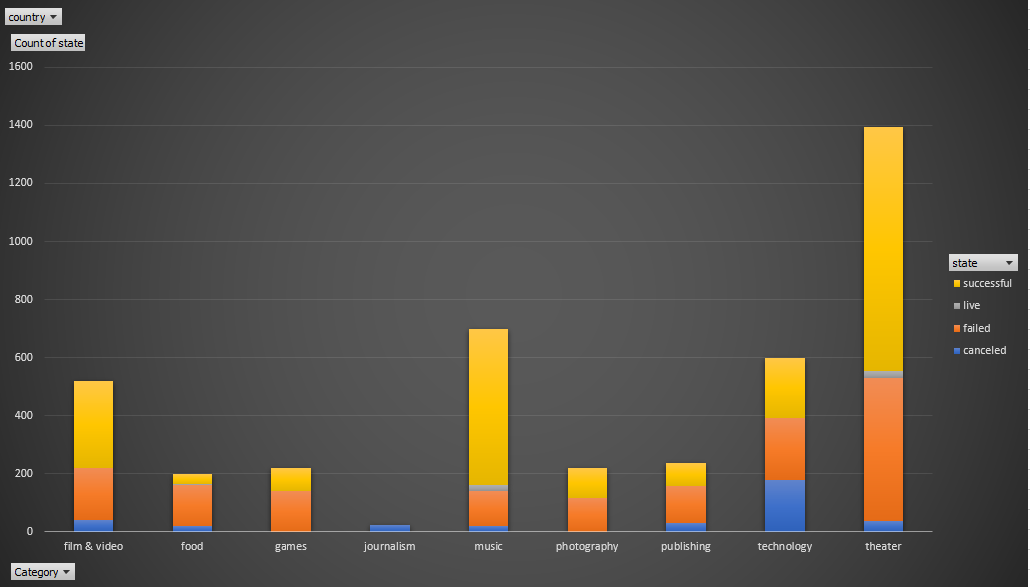
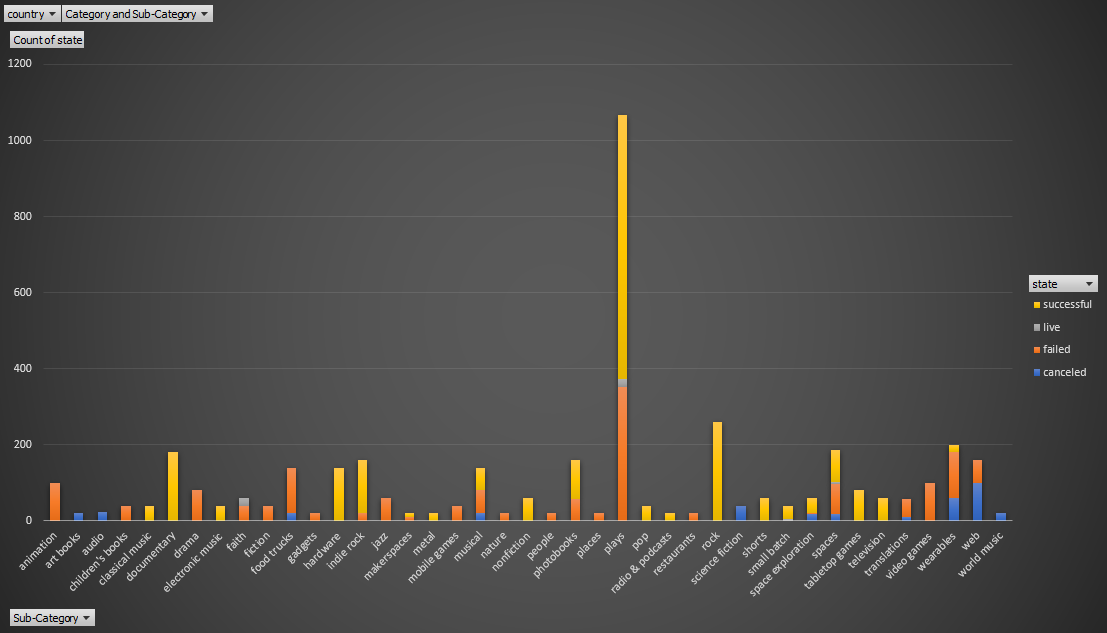
**Excel Homework: Kickstart My Chart – Matthew Taylor**

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

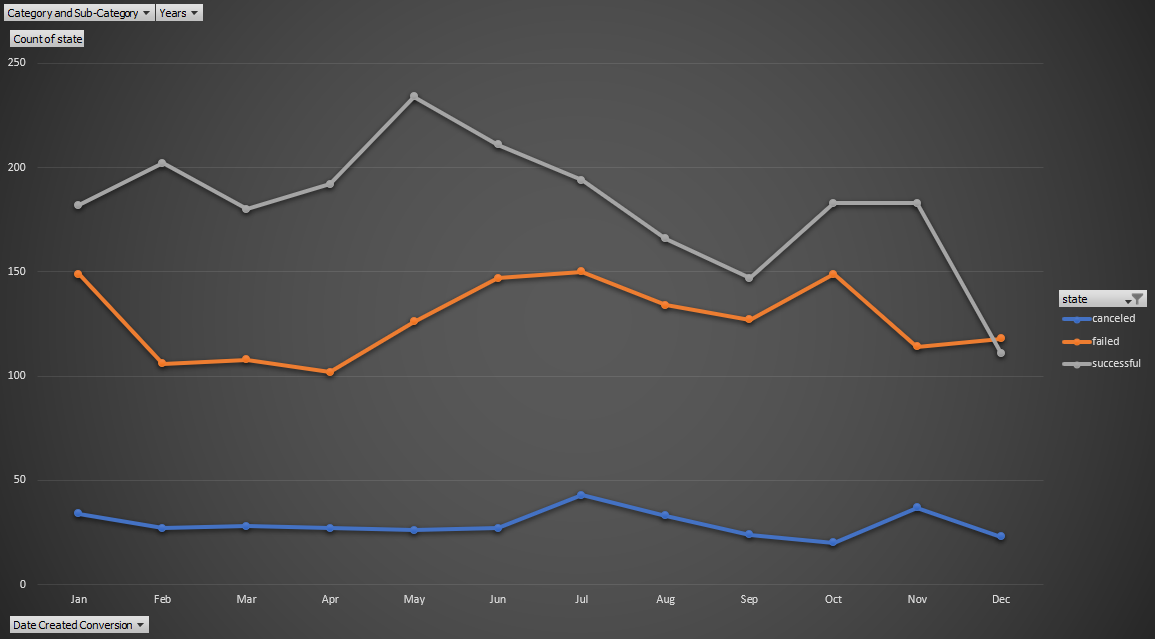
That the most popular projects on Kickstarter are part of the Theatre category with 1393 projects out of the sample of 4114 belonging to this category.



The dominance of Theatre projects is largely attributed to the popularity of the underlying sub-category Plays. The number of projects in this sub-category was 1066 which makes up 76.5% of the overall Theatre category.



During the period the sample covers, the number of successful projects tends to decrease with time with a sharp fall between November and December from 183 successful projects to 111.



1. What are the limitations of this dataset?

This dataset only has 4000 entries, it’s difficult to say if this is representative of all projects on Kickstarter, as more than 300,000 projects have been launched in total, we are barely analysing over 1% of the total.

We have also been told that roughly only a third of projects have been successful whereas in our dataset the rate of success is 53.1% which strengthens the idea that this dataset is not in fact representative of the overall population.

The data also seems skewed towards theatre projects, more specifically, plays. There are 41 unique sub-categories yet plays alone take up 25.9% of all projects in the sample. It could be more useful to re-run the analysis with these data points removed or filtered out.

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

Another useful table we could create would be to see what category or sub-category of projects are most popular in certain regions and how successful they are. This could tell us which location it would be most beneficial to launch certain projects in.

We could create a new column telling us total time project was live based on the launched at and deadline dates, and then analyse if the length of time the project was available affected the amount of funding it received and whether it was successful or not.

**Bonus Statistical Analysis Questions**

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

For both datasets, there are large maximum values which are contributing to raising the arithmetic mean, for this reason I would suggest the **median** is a better measure of the average value.

Based on the Q3+1.5IQR rule any successful project with a number of backers over 309 could be considered an outlier. (confirming the hypothesis that the maximum value in the dataset of 26,457 is an outlier).

Based on the Q3+1.5IQR rule any failed project with a number of backers over 29 backers could be considered an outlier. (confirming the hypothesis that the maximum value in the dataset of 1,293 is an outlier).

**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 the successful campaigns than with failed with a variance of 712,840 compared to 3,773. I believe that this result does make sense, it's more likely that if a project is unsuccessful it will have a lower count of backers and many failed projects will have 0 backers, the combination of these reasons will contribute to the data points (number of backers) to be closer together and cause lower variance.

Successful projects however could have many times more backers as the projects gain popularity and momentum. Successful projects will all have different goal amounts, average donation sizes, and over-subscriptions meaning the number of backers they ultimately have will be very variable and as such it makes sense for the measure of variance in these data points to be higher when compared to the failed or unsuccessful projects.