# Kickstarter Data Homework – Excel

## Conclusions from the Data

From the data presented, it can be seen that by volume, “Theater” and “Music” are the most successful Parent Categories of project in reaching their goal for funding. Theater has more successfully funded projects with 839 compared to Music with 540, however, Music has 77.14% of campaigns shown in the data hit their goal (with 120 failed, 20 live and 20 cancelled) compared to Theater’s 60.23% success rate (with 493 failed, 24 live and 37 cancelled). From this we can conclude that Music campaigns are the most likely to succeed in their funding.

If we then drill further down to the Sub-Category data, we learn more about what is likely to make a project successful. We can see that in Music, the following categories have a greater than 85% success rate:

* Classical Music (100% success)
* Electronic Music (100% success)
* Indie Rock (87.5% success)
* Metal (100% success)
* Pop (100% success)
* Rock (100% success)

If someone is looking to create a successful campaign in music, they should prioritise these sub-categories and avoid World Music, Jazz and Faith genres which will not succeed based on this data set, as all have a 0% success rate as of the time this data was gathered.

Based on the date created information available to us, we can see that a Kickstarter campaign is more likely to be successful if launched in the first 6 months of the year. There is a marked decrease in the volume of successful campaigns after June, moving from an average of approximately 200/month in January-June to 164/month in July-December.

## Limitations of the Dataset

This dataset only represents approximately 4,000 of the over 300,000 Kickstarter projects which have been created. This dataset is therefore incomplete without looking at all the data to study if the trends we have identified above would hold.

It is a limitation of the data that it consists of multiple currencies such as Euros and US Dollars, instead of having a standardised unit of currency measurement to compare the campaigns together across the categories.

There are also some columns in this data with titles and impacts which have not been explained or looked at for trends, such as the Staff Pick and Spotlight column. In this data, all rows with Spotlight equals ‘TRUE’ have been successful campaigns, but without knowing what this column represents, we are unable to make an informed comment on the significance amongst the data.

## Other Charts and Tables

Other charts and tables we could create from this data include a breakdown of the outcome of campaigns based on their country of origin over time. It would be interesting to see any observable trends in countries which have become more successful or less successful with Kickstarter campaigns over time.

Having created the average donation column for each campaign, we could also compare average donation size across campaign category and sub-category. We could then introduce the campaign launch date into this data to observe of there have been changes in spending patterns of Kickstarter backers over the years i.e. do backers spend more on average in more recent campaigns or have they become more wary of backing over time and reduced their amount of investment.

Although we do not have the full details of the information on what the “Staff Pick” and “Spotlight” data represents, I believe there is value in mapping the outcomes of campaigns which have been tagged ‘TRUE’ vs ‘FALSE’ in these columns. I would filter this data by country and by Category to find out which country or category produces the most ‘TRUE’ tagged campaigns. This data can be presented back to Kickstarter to help quantify the impacts of these ratings, and also to help identify if there is any trend in the type of campaign likely to be given these pieces of information.

## Bonus: Statistical Analysis

In this data set I would argue that the median more accurately summarises the data set, the mean is 3 times as large as the median and is being affected by some of the outliers at the top end of the dataset with thousands of backers.

From the data we can see there is more variance in the successful campaign, a situation I would argue makes sense. From examining the unsuccessful campaigns, when they fail it is most often to be because they receive single digit or zero backers, and their number of backers does not reach the level of a successful campaign. Successful campaigns demonstrate a significantly higher median number of backers which is in line with my expectations on how these campaigns proceed, more backers, more money and a higher chance of success.