Excel Homework: Kickstart My Chart

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

The most frequently used Kickstarter campaign category within this dataset is theatre, with 1,393 theatre campaigns created during this time frame. Both within theatre and across the board, plays are the most popular subcategory with 1,066 projects.

Based on the percentage of successful campaigns by month (across all years), May is the most successful launch month with 61% of campaigns ending successfully. By contrast, December is the least successful month with 44% of campaigns ending successfully.

A high average donation amount does not necessarily lead to campaign success. Of the ten projects with the highest average donation amounts, only five reached their funding goal.

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

This dataset does not include all past Kickstarter campaigns – it is a sample of about 4,000 campaigns launched between 2009 and 2017.

The monetary data for each campaign is recorded in different currencies. We cannot meaningfully compare the goal and pledge amounts across different countries without converting these values to a common currency.

As the data only includes the total pledged amount and total number of backers, we can only determine the mean average of backers’ donations. The mean may be affected by outliers e.g. a large number of backers associated with one campaign.

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

* Column chart comparing the success and fail rates (y-axis) for staff picks and non-staff pick campaigns (x-axis), to see if projects picked by Kickstarter staff are more successful than others
* Line graph plotting the relationship between the length of a campaign in days (x-axis) and number of backers (y-axis)
* Treemap chart reflecting the number of campaigns within each category and sub-category.

Bonus Statistical Analysis

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

For successful campaigns, the number of backers includes outliers (two successful campaigns with over 20k backers, while the rest fall between 1 and approx. 8300). In the case of failed campaigns, many campaigns have zero backers. The median would be a more meaningful measure as the mean would be skewed upwards by the outliers for successful projects, and downwards by the zero values for unsuccessful projects.

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

Both the variance and standard deviation are much higher for successful campaigns, which indicates there is more variation in this dataset. This makes sense, as there is wider range in the number of campaign backers for successful projects. On the other hand, many failed campaigns do not have any backers, which means there are many zero values in this dataset and therefore less variance.