Kickstarter Analysis

**Background**

Over $2 billion has been raised using the massively successful crowdfunding service, Kickstarter, but not every project has found success. Of the more than 300,000 projects launched on Kickstarter, only a third have made it through the funding process with a positive outcome.

Getting funded on Kickstarter requires meeting or exceeding the project's initial goal, so many organizations spend months looking through past projects to discover some trick for finding success. For this week, you will organize and analyze a database of 4,000 past projects to uncover any hidden trends.

**Preliminary Analysis**

We were provided a dataset of 4,000 past Kickstarter projects in an attempt to uncover some market trends. From here, we created some additional features such as “Percent Funded” and “Average Donation” to see how much money a campaign made to reach its initial goal and how much each backer paid on average for the project. We then took the “Category and Sub-Category” column and used formulas to split it into two new separate columns to help with further analysis later. The next step was to conduct data analysis to identify trends by utilizing conditional formatting to color code the outcome and percentage funded categories to help visually display the information for analysis.

**Data Analysis**

Initially we were able to create a pivot table to analyze how many campaigns were successful, failed, canceled, or are currently live per category. We went ahead and used that pivot table to construct a stacked pivot chart which displayed the different outcomes by category. Figure 1 below is a stacked pivot chart displaying the different categories outcomes and can be filtered by country as well if we wish to dive a little deeper in where the outcomes are originating from.

Figure 1.

Next, we used the same data that was provided to create a new pivot table which helped with the amount of campaigns that were successful, failed, canceled, and are currently live per sub-category. We then utilized this pivot table to display a stacked column pivot chart which could be filtered by country and parent category. It would be interesting to see how the filters would yield values based off country and then how many were in each parent category. Figure 2 shows this below.

Figure 2.

Finally, we were asked to convert timestamps into normal dates. Therefore, we created two new columns to house the data for the conversions. The “Date Created Conversion” and “Date Ended Conversion” columns were the two created. The goal from here was to construct a pivot table that held a column of state, rows of Date Created Conversion, values based on the count of state, and filters based on parent category and years. This provided an easy way to visualize numerically how many were successful, failed, canceled, and were still live from January to December. This also allows us to look at specific years if we desire and distinguish between parent categories. Figure 3 presents this below in a pivot chart line graph.

Figure 3.

**Conclusions**

Based on the data collected, we can draw a few conclusions from what is displayed about Kickstarter campaigns. The first conclusion that we can draw is from the sub-categories in Figure 2. Here we see that the most successful sub-categories are Documentary (Film & Video), Hardware (Technology), and Rock (Music). Figure 2 shows that these sub-categories had success rates of 100%. In particular, the Rock sub-category had the highest number of successful projects as well out of the ones with a 100% success rate. So, to achieve success in Kickstarter someone would want to investigate the Music category and the Rock sub-category. The next conclusion that we can draw is that of the timing of the project such as what time of the year is most successful. Here we look at Figure 3 to see that the month of May has been the most successful in terms of when to hold a campaign or project. It also shows that it is the month with failure not being at a high occurrence as well. Our third conclusion that we can draw shows which parent category would be best to get involved with. The top percentages of successful categories were Music (77%), Theater (60%), and Film and Video (58%). The rest of the categories did not break 50% success rate. So, from the top percentages Music would seemingly be the best opportunity in order to run a successful campaign and if we were to be more precise, Rock would be the main sub-category with in Music to run a campaign with because of the volume of projects and the 100% success rate of the projects that have been conducted thus far. The most ideal conditions would be to run a campaign or project in May which is centered around Rock Music.

**Limitations**

One limitation of the provided dataset is that it may not accurately represent the data of all Kickstarter campaigns considering that it only represented a percentage of the total number. Another limitation of this dataset is that it did not include all factors that have gone into a campaign. Such factors could be how long the campaign lasted, where the campaign or project was hosted, the monetary value that went into the Kickstarter, the donation values that were made to each specific campaign, and ultimately a scope of maybe narrowing down what exact conclusions we could look for in the analysis of the data. We could also be missing information that would contribute to the analysis and setup of the Kickstarters such as information on the backers, if there are more categories and sub-categories involved, and maybe what were some restrictions that were put on or maybe not put on that could have influenced what backers were available on each of the categories and sub-categories.

**Other Tables/Graphs That Could Be Created**

Other tables and graphs could be created and analyzed from our dataset. A couple ideas and interests that I have are if the staffs were picked or not picked. We could try to analyze a dataset on duration form start to finish or the desired categories and sub-categories to ultimately see which ones were most successful over a longer period. I think that the funding and donation values would be interesting to see as far as concluding on if certain Kickstarters were successful or not based on monetary standing with finances and backing. I think the confidence in the findings would increase if there was more analysis conducted that narrowed down which Kickstarters were more ideal as far as location and financial backing and donations over a lengthy period. I believe it would allow for the successful outcomes to mean more if a Kickstarter had a longer duration of success compared to a Kickstarter with a shorter duration of success.

**Bonus Statistical Analysis**

In the case of successful date, the median summarizes the data more meaningfully compared to the mean because the distribution of the dataset is affected by values that can be considered as outliers on both ends of the dataset spectrum. The same can be said for that of the failed data such that the median summarizes the data more meaningfully than the mean because of its outliers as well. The mean considers outliers which will draw its value away from the middle of the data set where the median will find the middle of the dataset by dispelling and somewhat forgetting about the outliers. There is more variability within successful campaigns based on the data results. It makes sense because successful campaigns have a wide range and variety of backers where the failed campaigns seemingly have a smaller number of backers and not as high of a ceiling. Therefore, in failed campaigns the number of backers is more closely related to each other compared to those of successful campaign backers. The variation is greater in one campaign to another in successful campaigns compared to failed campaigns due to some successful campaigns overachieving from their set goal and a higher value of backers due to the success.