Kickstarter Analysis

The purpose of this analysis was to review data for 4000 Kickstarter projects to try and find a trend from the data on what can help a project succeed or fail. The data covers a wide variety of information: timeframe of the project, goal and pledged amount of money, the number of backers, country it took place in, and the category for the projects. Some of these variables will have a greater effect than others on the outcome of the project.

I started off analyzing the data by color coding the outcome of each project (green for successful, red for failed, yellow for cancelled, and blue for live). From there I determined the percentage of funding for each Kickstarter and then color coding the data (0-100% in red, 100-200% in green, and anything above 200% in blue). I also created an average donation column, to see the average amount donated per backer. Next, I split the Category and Sub-Category's into two separate columns. Lastly, the Launched and Deadline dates were in Unix timestamps, so I converted them to normal dates in new columns.

After altering the data, I created multiple pivot charts to create a visual on how different variables effected the outcome of the projects. The first chart looked at the outcome of the projects by category and then filter by country. The second chart was the outcomes by subcategory, filtered by country and the main category. The third chart analyzed the different outcomes by month.

The first conclusion that was made, is that while the category Theatre has the most successful outcomes, the Music category has a higher success rate by about 17% and can be seen in Table 1. In Graph 1, every Kickstarter Project that had a sub-category of Rock was successful. Lastly, in Graph 2 the worst time to start a Kickstarter is in December, as the fail rate is at about 46%.

After analyzing the data, the recommendation would be to have a Theatre or Music Kickstarter with a focus on Rock. The project would also need to be launched in May, as seen in Graph 2, when they have the highest rate to succeed.
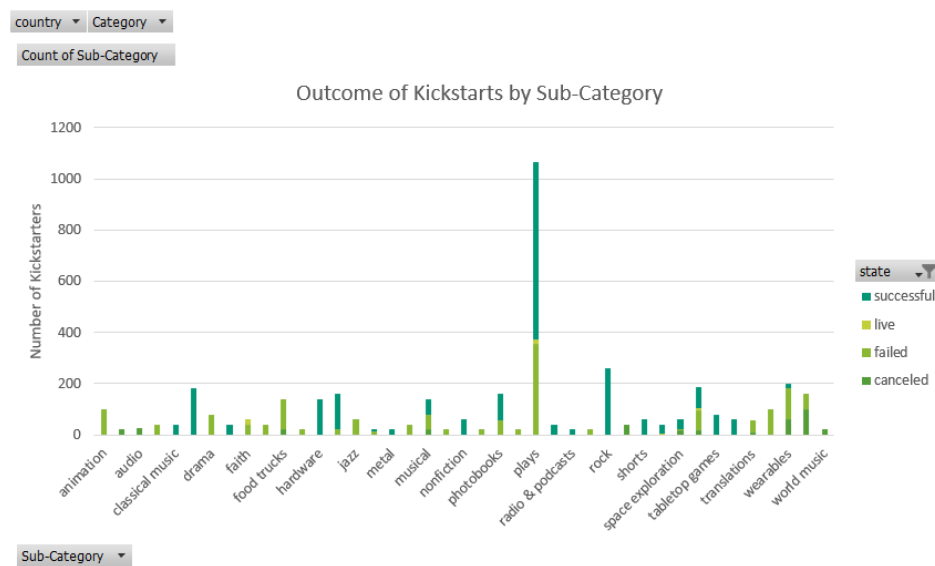
Some limitations of the information provided is that while Theatre does show a greater number of successful projects, it would be better to show the rates of success/failure/etc. to have better analysis of the information. Looking at just the raw numbers of successful projects does not give an accurate description of which Category is the most successful. Another issue is that some of these projects have an extremely low goal of the amount of money needed to be raised. This skews the successful projects numbers higher in some cases. I also found one project that states it is a hundred percent funded, and this is just a second round of funding, so some projects could be counted twice in this data, further skewing the data.

One table that could be created would be the success rate per category and sub-category to have a better understanding of what is more likely to succeed and fail. Another graph to create would be the outcomes and the categories by country, too compare which country has better or worse outcomes for projects, and which category of projects do better in different countries.
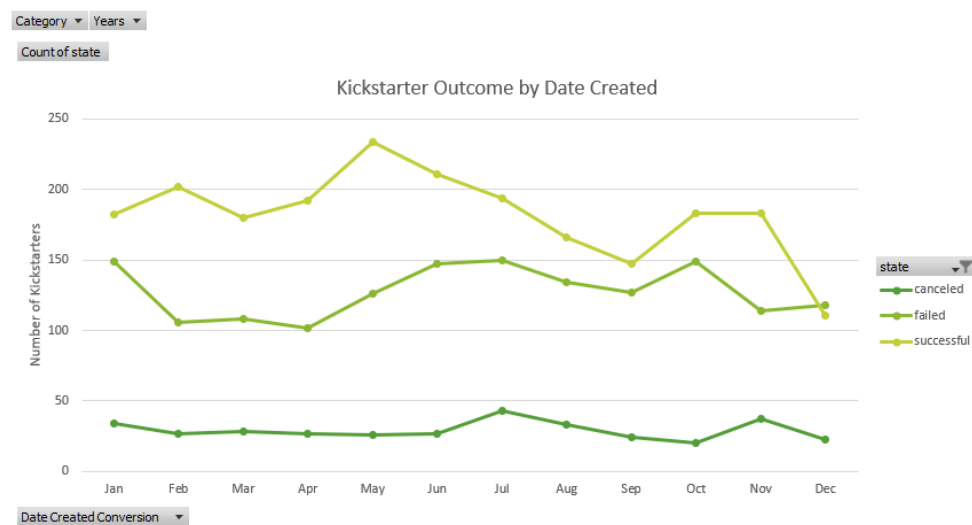
Table 1:

Count of Category	Column Labels					
Row Labels	canceled	failed	live	successful	Grand Total	Success Rate
film & video	40	180		300	520	0.576923077
food	20	140	6	34	200	0.17
games		140		80	220	0.363636364
journalism	24				24	0
music	20	120	20	540	700	0.771428571
photography		117		103	220	0.468181818
publishing	30	127		80	237	0.337552743
technology	178	213		209	600	0.348333333
theater	37	493	24	839	1393	0.6022972
Grand Total	349	1530	50	2185	4114	

Graph 1

country ▾   Category ▾

Count of Sub-Category

Graph 2

Category ▾   Years ▾

Count of state

Date Created Conversion ▾

Bonus 2:

The median would summarize the data better, because there is a large skew for both successful and failed projects. The maximum number for both successful and failed projects is far greater than the median, therefore creating a higher mean.

There is more variability for successful campaigns because there is a higher variance. Yes it does make sense, because if project fails, that means they didn't meet the goal for their project, which would typically mean a smaller number of backers for those projects, creating less variance. Successful projects, would most likely have more backers due to multiple people trying to be part of the project.