Upon initially looking at the graphs created with the provided datasets, an instant conclusion can be drawn that most people are going to find success when crowdfunding to support an artistic endeavor such as theater, music or film/video. Based on the pivot tables created to compare parent and subcategory campaigns to the outcome of the project, there are very large numbers (mostly successes) for the three categories listed prior – with other parent categories like food, games, technology, publishing and photography displaying about half of the traffic as the three main categories.

The pivot table created for showing the successful/failed/canceled outcome of projects based on the time of year also indicated that the big three fields experience the most growth, as well as least failures, throughout the middle months of the year, with very slight increase in failures/cancelations towards the end of the year. While the other parent categories experience the slight bump in successes throughout the middle portion of the year, they also experience much more failures around this time of year, and don’t benefit from the same end-of-year trend that the big three categories do and experience a spike in failed campaigns.

The last major trend to be gleaned from the pivot tables is that most projects are being funded/taking place in North America. When filtering out the other countries on either the parent or subcategory pivot tables, there are hundreds more projects being started in Canada and the US than all the other countries within the dataset. When looking at the outcomes of the different categories for the other countries in the dataset, the same trend can be seen of the most popular projects being those relating to the arts, with projects of these natures receiving not only the most attention, but also the most success from various crowdfunding campaigns.

While the data within this dataset is helpful for determining the success of various crowdfunding projects, I think an extreme limitation is not being able to see more minute details on how or why some projects fail or get cancelled while others succeed. If there was available data to express how each campaign was operating – such as their advertising, their promises of backer benefits provided, or even the length of time for delivering the promised product, then there would be more information to work with as far as what some projects were doing that supported their success, or what some projects were doing that ultimately spelled their end.

For the final statistical calculations at the end of the assignment, there can be useful information derived from both the median and the mode to help summarize the data. While the mean is also just the average of the values being calculated, it can show the user an expected result, where the median will represent the middle of the datapoints when they are organized by increasing numerical value. For example, for the outcomes of successful data, we got a mean of 851 backers, with a median of 201 backers. If we were using the mean to summarize the data here, a crowdfunding project would need around 850 backers to be successful. However, when using the median to summarize the data, most of the projects in the middle of dataset needed only 200 backers to be considered successful. Since this is such a stark difference in numbers of backers only to yield the same result of being a successful project, I would elect to use the median to represent the data here, as it lowers the bar for many projects on terms of them not being reliant on the number of backers to determine their overall project success. The same trend can be seen for failed projects against the number of backers they had.

There is more variability with successful campaigns than unsuccessful campaigns. This makes sense because the volume of successful campaigns was greater than that of the failed campaigns. The variability represented here can even be seen when comparing the data of the median and the modes for the respective data sets. More variance/standard deviations are going to be present with datasets that show high levels of fluctuation with the initial data.