Matt Kennedy

Data Analytics Bootcamp

14 September 2019

Kickstarter Crowdfunding Analysis & Observations

The purpose of this assignment was to use Excel formulas, pivot tables, and pivot charts to perform analyses of data from 4000 Kickstarter Crowdfunding projects. Formulas were used to calculate additional data for additional fields to supplement the initially provided information. These values were then used to create pivot tables and charts to display the project outcomes for different criteria and using different filters. These tabular and graphical representations allowed for certain conclusions to be drawn, but also exposed some limitations from the dataset.

The first of several conclusions that can be drawn from the data is that the vast majority of Kickstarter Crowdfunding projects were either successes or failures. This is proven by the projects counts for those project states being in the thousands compared to the canceled and live counts in the hundreds. There is a slight edge towards more projects being successful rather than failing, but it can’t be assumed because there are close in overall percentage of projects in the sample (~15 percentage points). With tens of thousands more similar projects not included in this sample, that close gap of 15 percentage points could easily be closed up and flipped in the other direction, which highlights this limitation of the dataset.

When looking at the percentage of project state versus the stated goal range, it can be concluded that there is no discernable correlation for projects that were successes or failures. However, it is clear that more projects in the higher goal ranges were canceled than those in the lower ranges. Justifying the former case, the line plots for the successful and failed projects do initially show a consistent trend of decreasing and increasing respectively. However, the traces then change directions, which invalidates that initial analysis. The line plot for the canceled data continues to grow, so that does give a clear indication that a conclusion can be drawn.

A similar situation to the goal range scenario above is that of the project states related to month for all of the years included in the data. It is evident that the number of successful projects grows in the spring and fall, but experiences a moderate to significant decline in the summer and winter months. On the other hand, the number of failed projects declined in the spring and fall, but had a moderate to significant increase in the summer and winter months. The data for canceled projects was a different story though because it can be concluded that the number of these projects was constant on average for each month. While small peaks and valleys were evident in the data trace, they were roughly of the same magnitude and evened out in the end. These trends were roughly consistent for the individual year data, with expected variations in the number of projects.

In addition to the aforementioned limitation regarding the distribution of successful and failed projects, a few further limitations are evident. The first comes in the form of a bias in the sampled data towards projects in the arts, especially in the sub-category of plays. While it roughly doesn’t affect the overall distribution of successful to failed projects, it does make up for a significant number of projects in both categories. The dataset would present a far more accurate distribution of projects if there weren’t so many sampled from this singular sub-category. Additionally, that lack of data for live projects in the majority of categories and sub-categories limits any ability to make any meaningful decisions from that data.

Moving past the conclusions and limitations, there are definitely areas where additional tables and plots would be useful. One instance would be to add a pivot table of project state based on the number of backers for the project, with filters for category. However, before doing this, an additional column would need to be added to the initial data tab to group the backer data into specified ranges. With that completed, a bar chart could be created to visualize the data to find whether a project with more backers was more successful than one with fewer backers. Another possible option would be to include a line chart that compares the monetary ranges of the average donation to the percentage of project state. This would be similar to the already performed goal range analyses but would help identify a potential meaningful relationship. Finally, for each of the individual project states, a pie chart could be used to show the distribution of categories and sub-categories included in two of the already present pivot tables. These pie charts would provide a different perspective on the data compared to the stacked bar charts currently being used, as the latter can be a little difficult to view at times.