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 enable analyses of data from 4000 Kickstarter Crowdfunding projects. Formulas were used to calculate data for additional fields to supplement the initial information. Next, the entire dataset was used to create pivot tables and charts displaying the project outcomes based on different criteria and utilizing various filters. These tabular and graphical representations allowed for certain conclusions to be drawn, but also exposed some limitations of the dataset.

The first 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 project counts for those outcomes being in the thousands compared to the canceled and live counts in the hundreds. Slightly more projects were successes rather than failures, but this can’t be assumed for the total population of Kickstarter projects because these outcomes were only separated by approximately fifteen percentage points. With tens of thousands more similar projects not included in this dataset, that percentage point difference could easily be eliminated or reversed. This highlights the limiting nature of the narrow sample size of this dataset compared to the total project population.

When looking at the percentage of project outcome versus the monetary goal range, it can be concluded that there is no discernable correlation between those variables for projects that were either 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. That trend then reverses, which invalidates the aforementioned analysis. In the latter case, the line plot for the canceled data continues to grow with goal range, which indicates a correlation is present.

Another comparison can be made when analyzing project outcome versus time of year. 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. Conversely, the number of failed projects declines in the spring and fall, but has a moderate to significant increase in the summer and winter months. The data for canceled projects was a different story. It can be concluded that the number of canceled projects was constant on average throughout the year. While small peaks and valleys were evident in the data trace, they were roughly of the same magnitude and equalized. All of the above trends were consistent from year to year, regardless of variations in the number of projects.

In addition to the aforementioned limitation regarding the narrow sample size, 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 doesn’t affect the overall distribution of successful to failed projects, arts projects make up a significant portion of both outcomes. The dataset would most likely represent a far more accurate distribution of projects if there weren’t so many sampled from this singular category and sub-category. Additionally, the lack of data for live projects in the majority of categories and sub-categories limits the ability to make any meaningful observations about this project outcome.

Moving beyond these conclusions and limitations, there are definitely areas where additional tables and/or plots would be useful. One instance would be to add a pivot table of project outcome based on the number of project backers, with filters for project category. The backer data would need to be grouped into numerical ranges and added to the initial data. With that completed, a bar chart could be created to visualize whether projects with more backers were more successful than those with fewer. Another addition could be a line chart that compares the project outcome percentage to the average monetary donation. This would be similar to the goal range analysis but would help identify how average donation affects project outcome. Finally, for each of the individual project outcomes, a pie chart could visualize the distribution of project categories. Additionally, for each category within an outcome, a pie chart could serve the same purpose displaying sub-categories. These pie charts would provide a clearer representation of this data compared to the stacked bar charts currently being used.