Alex Zapuchlak

01/09/2021

Kickstarter Dataset Analysis

In looking at the raw data and the tables we generated, there were some trends and learnings to take away from this dataset.

From looking at the Amount of Campaigns by State & Sub-Category graph, it was easy to identify that Plays drive the most interest from Kickstarter content creators and generate the highest successful campaigns in quantity. It seems that the Kickstarter community has a large interest in Theatre, specifically and vastly with plays. So much so that that sub-category held over 25% of the total amount of campaigns. For reference, the second largest sub-category was Rock music holding 6.3%.

Secondly, Kickstarter has a high success rate in allowing inventors and creators reach their goals. Only about 37% of campaigns that launch fail, and there is over a 50% chance that any given campaign will succeed. Looking into why, it seems that time could be a possible factor, having trends over time of high success in the spring and lower success in the winter.

Lastly, there is evidence in the data that campaigns with lower goals are more successful in accomplishing success. This may come as no surprise, but it is interesting that 67% of campaigns with a goal of less than $5000 resulted in success. Moving forward, there’s an initial drop of success rate between $5,000 and $10,000, but after that results stay within a close range up until the higher end of the goal amounts only showing a significant drop around $40,000. It seems there are niches for low-goal campaigns, mid-goal campaigns, and high goal campaigns and where you see significant changes is in the gaps between low to mid to high.

As with many databases and analysis reports, there were some limitations to this dataset. The first one that I noticed was in the state column and in the way the states were broken out. From an analysis perspective, it is hard to fully understand the story of the data without knowing full details. Why did the canceled campaigns get canceled? Was it because they were not generating enough buyers and the creators backed out early? Did they realize it was not a realistic investment? It is hard to compare the percentages of canceled campaigns with successful and failed campaigns without having any context.

Another limitation to the data is that the dates in which the campaigns ran went from 2009 to 2017. The internet has changed a lot in that time frame and with Kickstarter being an online based crowdfunding platform, it has surely been impacted by the expansion of technology. Some things that might have changed could be the number of people who know what Kickstarter is, the amount of platforms to share campaign information through, and perhaps consumers, or “backers”, having more apprehension of investing money into a product through the internet. Furthermore, one main way the database is categorized and broken down is by category and subcategory, many of which have also gone through substantial changes between 2009 and 2017. The nature of Kickstarter is fueled by people’s interests and hobbies, and because of social media providing access to an infinite amount of content, it become easier to distribute and find good content. Especially with this data containing categories like “Technology”, “Film & Video”, “Music”, and “Games”, all of which have drastically evolved over time and widely increased opportunities for both creators and consumers. To back these theories up I turned to the data and it did not let me down. Between 2009 and 2011 there were a total of 250 campaigns in this dataset for Kickstarter, while there were over 2,300 listed between 2015 and 2017. While the comparison of percentages between successful and failed campaigns remained similar, the amount of data within each time frame could skew any analysis relative to time.

Based off the limitations and thorough analysis of the data, I noticed there was a lot more opportunity for data comparison and had some ideas about ways to improve the data and visualizations that are a part of this assignment. The first being a graph or table that breaks down success and failure rates, backer counts, and then comparing both with the years the campaign ran. Alternatively, creating a few year-over-year graphs state statuses by month could be beneficial in analyzing these same trends and could be a viable option as well. To be able to notice trends in certain categories or sub-categories based on yearly trends would be interesting and beneficial to customers/clients of Kickstarter.

Additionally, it would be beneficial to have a percentage breakout of State data by Sub-Category data, similar to what was created in the bonus table which looked at state percentages by goal amounts. The table we created was beneficial to analysis, but limited takeaways and possibly could skew perception of success. For example, in the Amount of Campaigns by State and Sub-Category graph the viewer’s attention instantly goes to “Plays” because it has the highest number of campaigns, however, they might not notice that 100% of the classical music campaigns ran have been successful. Seeing a breakdown by percentage would likely say a lot about the nature of the campaigns and their likelihood of driving success.

In conclusion, this was an interesting set of data to examine and analyze. Diving down deep into the data, I was able to notice some key trends like