Joshua Acampado

September 15. 2020

Data Science Bootcamp

Kickstarter Data Report

At a glance at the data, we are able to discern three conclusions; on a global scale, Kickstarters that are tied to music or theater are most successful, within those categories, Kickstarters tied to rock or plays, respectively, are most successful, and Kickstarters that are launched within the April to July are most successful. Upon a further analysis, we are able to see that there are more successful ‘theater’ projects, however music has a better ratio of ‘successful’ to ‘total’ projects. While there were 839 successful projects within the category of ‘theater’, that is in relation to a total of 1,393 projects resulting in a 60.23% success rate. In comparison, there were 540 successful projects in relation to the 700 total, resulting in a 77.14% success rate for projects within the ‘music’ category. From this data, one could say that starting a project within the ‘music’ category will result in the highest chance for success. When looking at the ‘sub-categories’ of the ‘music’, the data shows high success rates within several fields. Sub-categories such as ‘classical music’, ‘electronic music’, ‘indie rock’, ‘metal’, ‘pop’, and ‘rock’ all have a success rate of at least 87.5%. Conversely, ‘sub-categories’ such as ‘faith’, ‘jazz’, have a 0% success rate based upon this data. When looking at the ‘sub-categories’ of ‘theater’, the data shows that only one sub-category, ‘plays’, has a success rate that is over 50%, while the other two fall below a 50% success rate. From this comparison, it is more favorable to start a project tied to the ‘music’ category, as opposed to the other categories presented in this data set. Looking over to the outcomes based on launch date, we previously stated that projects launched from April to July are the most successful. Within this range, we see that May is the most favorable month with a 60.62% success rate when the grand total of projects launched take count of successful, failed, and canceled projects. In conclusion, projects that are connected to music or theater are most successful, more sub-categories within music are the most successful, and projects launched in May have the most favorable result.

While this data set yields a lot of information, a few of its limitations are the lacks information regarding the marketing of these projects, the inability to see if outliers in donation amount or just individual donation amounts in general, and the timeline of the donations. Marketing heavily impacts the accessibility of the product, often influencing the success of the project. An individual is unable to tell the demographic of people that contributed to these projects. It could be almost entirely funded within a community or a town, or it could have been a project that garnered national, if not global, attention. By analyzing the location of the project to the average distance of a person who donates, one could come to a clearer conclusion on how to develop a successful project. With regards to donations outliers, it would be important to see if a very large donation or a very small donation skews the average amount donated. This segues into how it is important to see the timeline of when donations came in. If the project had sudden spikes in donations towards the goal, it would influence how one would interpret the data. For example, if all of the successful projects within the music category were a result of sudden large donations, it would be flawed to suggest that music projects are the most reliable. Likewise, if it was seen that a project had a steady flow of donations over a course of time, that would represent genuine public interest rather than a sudden success. If a team were to seriously investigate what makes a successful Kickstarter, these variables would need to be taken into account.

Although there is more that could be sought out, the already presented data could be visualized further to help inform more decisions and conclusions. Looking at the other headers within the data, it would be interesting to see the following tables/graphs;

* Correlation of the state of project to if it is either in ‘spotlight’ or ‘staff\_pick’
  + Seeing if there is an influence where being in the ‘spotlight’ or ‘staff\_pick’ categories results in success
* Correlation of the goal of the project in relation to the state of the project
  + Try to determine an optimal range when trying to setting a goal for a project
    - Does too high of a goal result in more failed projects?
    - Does too low of a goal result in more successful projects?
* Correlation of Average Donation to Category/Sub-Category
  + Are individuals willing to donate more towards a goal?
* Correlation of Average Donation to State
  + Does a higher average donation impact the state or is there no correlation?
* Correlation of Category/Sub-Category to Country and their subsequent success rates in those countries
  + Are certain projects more popular in certain countries? Within those countries, what is most successful?

As a whole, this data set provides ample information to aid in analyzing the results of various projects in different states. While the additional information that was previously mentioned would aid in further decision making, we are able to make confident conclusions from our current visualizations and can continue to do so the with information that is already presented.