Kickstarter Analysis Report

By: Patrick Gendotti

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   1. We can multiple conclusions from the Kickstarter campaigns according to this data. They are listed as follows:
      1. The first conclusion we can reach about the Kickstarter campaigns is that the project success and failure rates per category is roughly the same across all categories measured, with the exception of projects involving food or journalism largely failing and projects involving music and film and video (to some extent) commonly succeeding. Otherwise, there seems to be about the same number of successful and failed projects per category. See the chart below for more information.
      2. The second conclusion that we can reach about the Kickstarter campaigns is that certain subcategories show nearly all successes or no successes at all. The subcategories that show all or nearly all successes include: classical music, documentary, electronic music, hardware, indie rock, metal, nonfiction, pop, radio & podcasts, rock, shorts, small batch, space exploration, tabletop games, and television. The subcategories that show none or nearly no successes include: animation, art books, audio, children’s books, drama, faith, fiction, food trucks, gadgets, jazz, mobile games, nature, people, places, restaurants, translations, video games, wearables, web, and world music. See the chart on the next page for more information.
      3. The third conclusion that we can reach about the Kickstarter campaigns is that the platform seems to be getting more popular over time. The line chart below shows the largely increasing number of projects over time, at a rate of about 200 projects per year. However, one possible confounding factor is that this data shows a sample of Kickstarter projects, so unless the yearly sample is proportional to the total project count for each year, this trend could be a result of a skewed data sample.
2. What are some limitations of this dataset?
   1. The biggest limitation of this dataset is the size – there are approximately 4,000 projects included in this sample of all the projects, but there are over 300,000 projects that have been launched on Kickstarter to date. So this sample of the entire population of projects on Kickstarter may not be enough to draw valid conclusions about general trends for projects on the website, especially depending on how this sample was constructed. It is important to use a **representative sample** to get a clear picture of trends or features of an entire population. If the data is not a representative sample, there may be biases or trends that exist in the sample data set that are not present or are different in the overall population.
   2. Another limitation of this dataset is that it only covers the years 2009 to 2017 – and it is currently the year 2020. So there may have been recent events or trends that are not apparent in the dataset because there is no data from the past three years. For example, the data shows that there have been more projects in recent years (e.g. 1225 in 205 and 950 in 2016) as compared to earlier years (e.g. 65 in 2010 and 171 in 2011), which is a trend that could continue into the past few years or change, and if it changes, we could try to determine the reason(s) why. However, with no data from the past three years, we cannot tell how this trend seen in earlier data continues or changes.
3. What are some other possible tables and/or graphs that we could create?
   1. We could create another pivot table and resulting graph that shows the success rates for different project durations – that is, we want to know if projects that have longer durations for support on Kickstarter are more successful overall.
   2. We could look at the success rates for different categories based on country – that is, do different countries seem to support different types of projects more or less than others?
   3. We could look at the average amount pledged per backer for each different subcategory of project – which could be especially useful for project subcategories that show all or nearly all successes or none or nearly no successes.

Bonus Statistical Analysis

The data is better summarized by the median than by the mean because the mean is pulled artificially higher by the maximum number of backers or a few other large projects. It appears that there is more variability in the number of backers for the successful campaigns, which makes sense because there could be a whole range of numbers of backers for successful campaigns, but the number of backers for unsuccessful campaigns will likely cluster around 0 or some other low number because those campaigns could not attract enough backers to make the campaigns successful.

|  |  |  |
| --- | --- | --- |
|  | **Successful** | **Unsuccessful** |
| mean | 194.42517 | 19.53014879 |
| median | 62 | 4 |
| minimum | 1 | 0 |
| maximum | 26457 | 1293 |
| variance | 713167.38 | 4444.760258 |
| st. dev. | 844.49238 | 66.66903523 |