Home Work Assignment – 1

**Data Analytics and Visualization**

**Rice University**

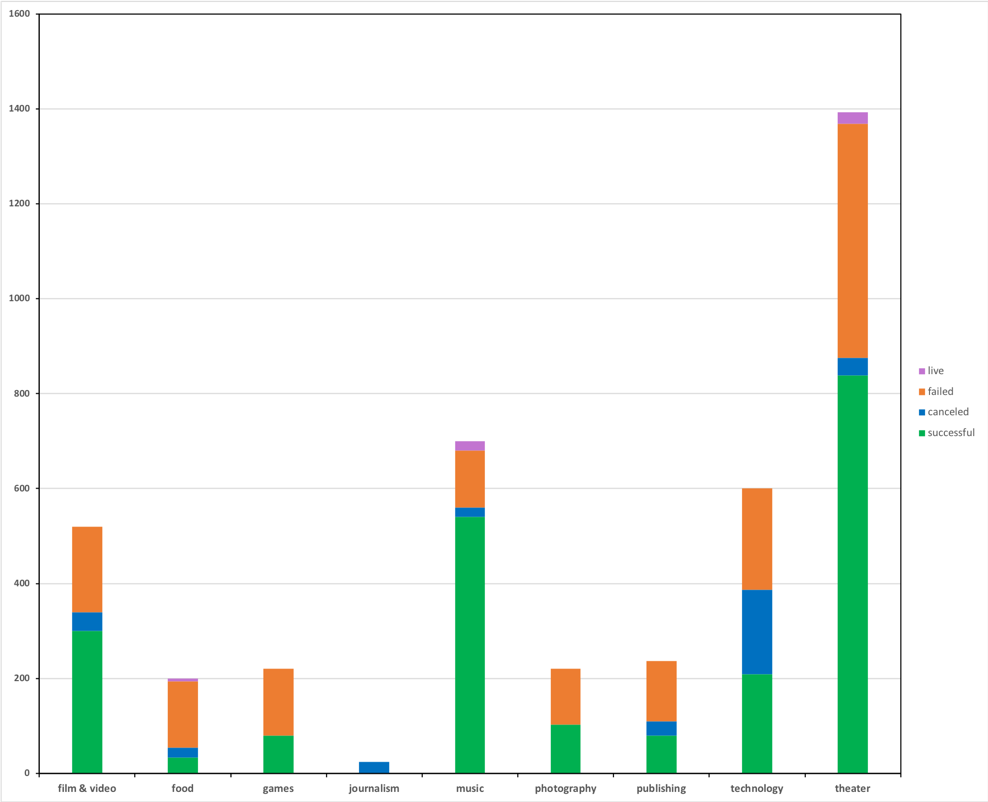
Dr. Sailendra Mahapatra

Q1: What are the three conclusions we can make about Kickstarter campaigns given the provided data?

The answers are based on analysis done in the attached Excel Work book:

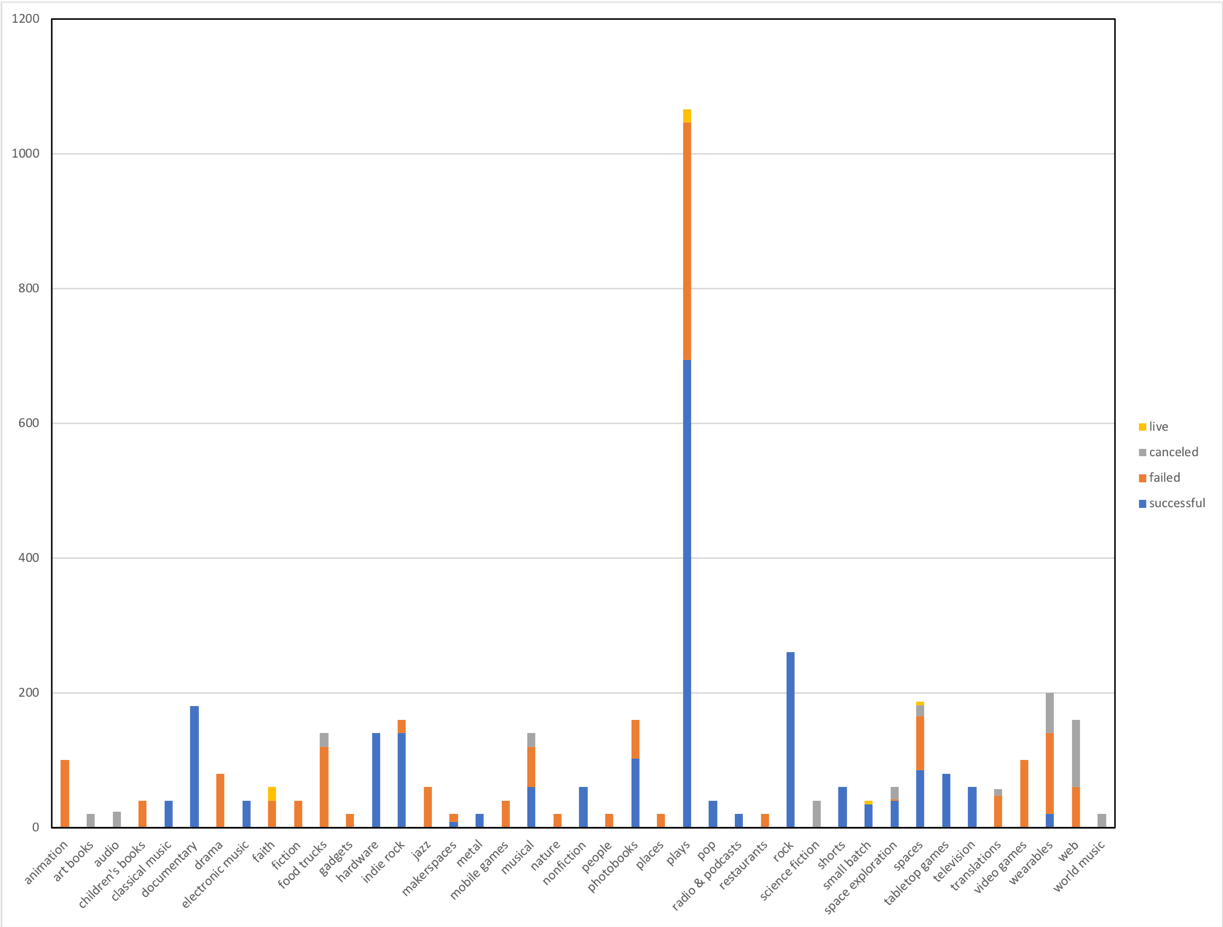
1. In the category class, the “theater” projects were the most successful among others (Total 839 successes which is the most in numbers, however, “music” with the second highest success cases (540) has the most success in percentage (+77%). Please see below the success/canceled/failure rate of different categories and a plot of completion status.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Row Labels** | **successful** | **canceled** | **failed** | **live** | **Grand Total** | **Success%** | **Canceled%** | **Failed%** |
| film & video | 300 | 40 | 180 |  | 520 | 57.6923077 | 7.692308 | 34.615385 |
| food | 34 | 20 | 140 | 6 | 200 | 17 | 10 | 70 |
| games | 80 |  | 140 |  | 220 | 36.3636364 | 0 | 63.636364 |
| journalism |  | 24 |  |  | 24 | 0 | 100 | 34.615385 |
| music | 540 | 20 | 120 | 20 | 700 | 77.1428571 | 2.857143 | 17.142857 |
| photography | 103 |  | 117 |  | 220 | 88.034188 | 0 | 53.181818 |
| publishing | 80 | 30 | 127 |  | 237 | 33.7552743 | 12.65823 | 53.586498 |
| technology | 209 | 178 | 213 |  | 600 | 34.8333333 | 29.66667 | 35.5 |
| theater | 839 | 37 | 493 | 24 | 1393 | 60.22972 | 2.656138 | 35.391242 |
| **Grand Total** | **2185** | **349** | **1530** | **50** | **4114** |  |  |  |



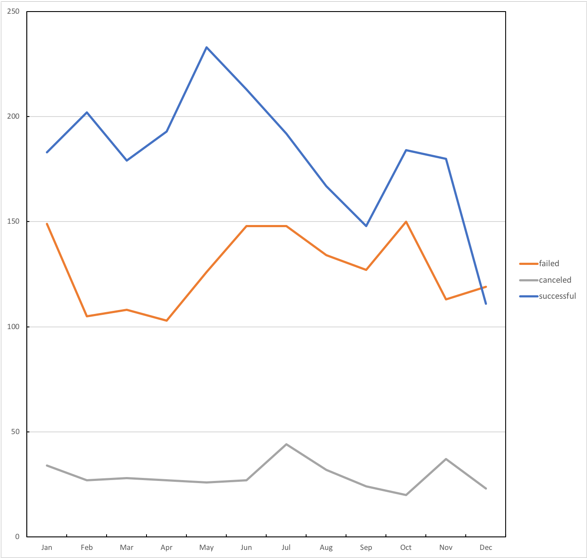
Category completion Status

1. In the sub-category, “rock” has a 100% success rate, however the total projects were only 260. The “plays” with most successes (639 out of a total of 1066) also achieved the second maximum success rate (+65%) among others. “art book”, “audios” and “world music” were canceled all together (cancel rate of 100%). Please see the plot of sub category completion status.



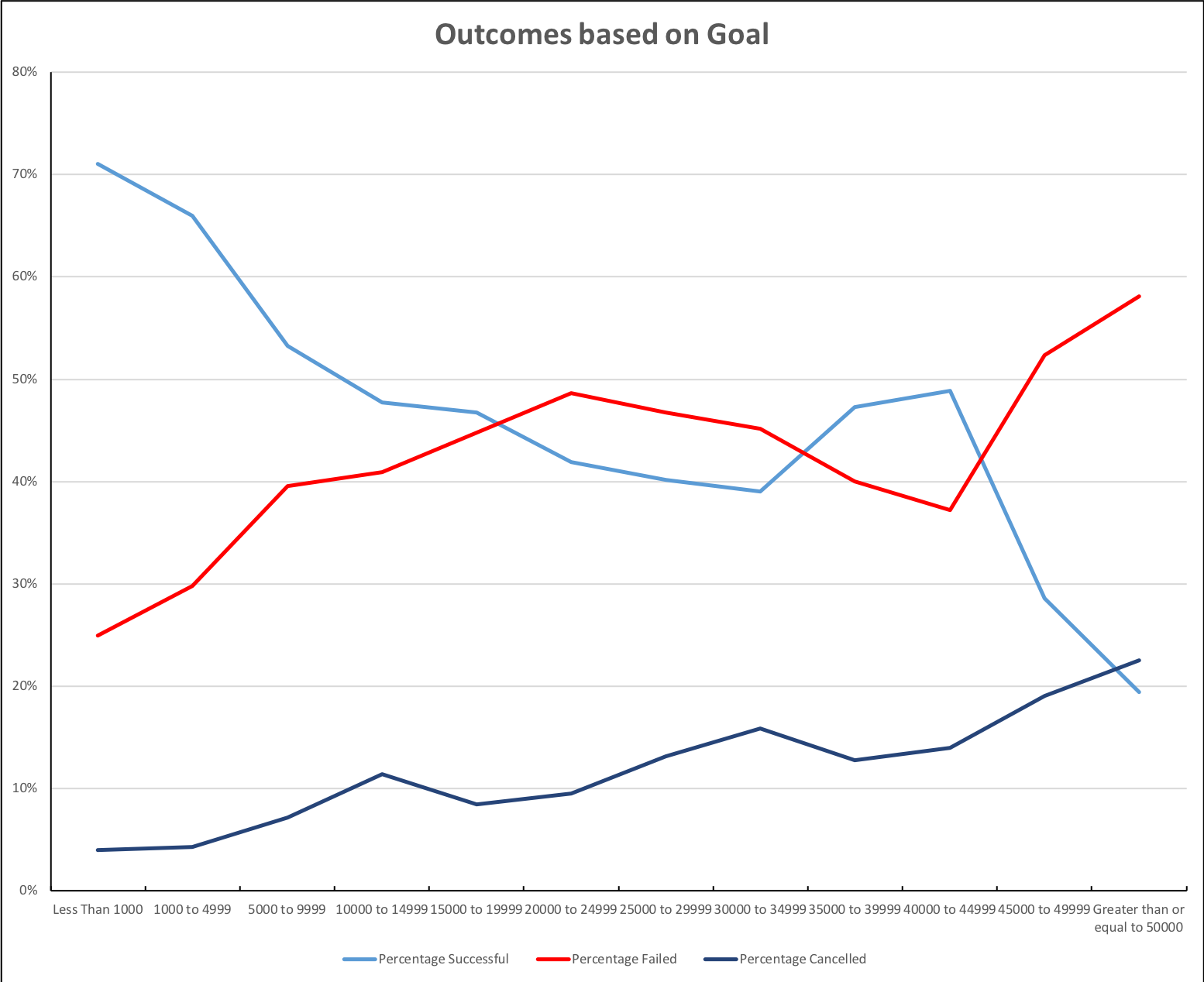
Sub category Completion Status

The most successful months of the launched date year were between April to July probably coinciding with a lag time of three months of year beginning activities leading to launching of the projects, and/or may be attributed to pleasant weather of spring and summer which might have played role in positive response from the audiences.



Launch date Completion Status

1. The most success comes from projects having less than 1000 in goal (71% success), followed by goal between 1000 to 4,999). The most cancelation and failure occurred with the projects with a goal of more than 50,000 (58% and 23%). The next highest in the above rates occurred for goal of 45,000 to 49,999 (52% failure and 19% cancelation. Please see the plot of outcomes out of different goal amounts.



Ranges of Goal

Q2: What are some of the limitations of this dataset?

The selective category/sub categories may not be appropriately selected keeping in mind of an audience of diverse cultures. For example, the video may not be providing actual information – rather a presentation of abstract data.

The role of currency vis a vis cost of living in different places along with cultural preferences may have affected the outcomes.

Q3: What are some other possible tables/graphs that we could create?

* A scattered graph with product name, goal, pledged amount and completion status would further differentiate the broad spectrum of preferential choices.
* All the above plots also could have been generated to analyze the trend in to further much deeper levels. For example, average donation, country/regions, completion status, etc.
* The achievements or failures can be better differentiated with a line chart of country, currency, goal, pledged and maximum achievements/failures (for comparing whether success, cancel or failure rates were a dominant outcome).
* The “Live” data may be also used to the completion rate percentages as a line graph to demarcate the non-completion rate in this analysis.