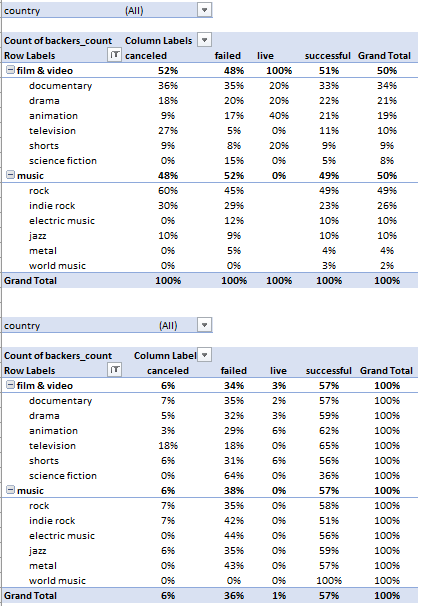
Question: Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

1. Among all categories, Theater, Film & Video and Music are the most promising groups that have the most number of projects. In particular, 344 projects are Theater, 178 for Film & Video and 175 for Music, considerably higher than other categories. For example, the 4th highest category, Technology, has 96 projects throughout 10 years. Additionally, the 3 mentioned categories are also reasonably safe for backers to invest in with acceptable successful rate. The average success rate for 1000 projects is 57%, which is slightly higher than Theater’s of 54% and equal to 57% of both Film & Video and Music categories.

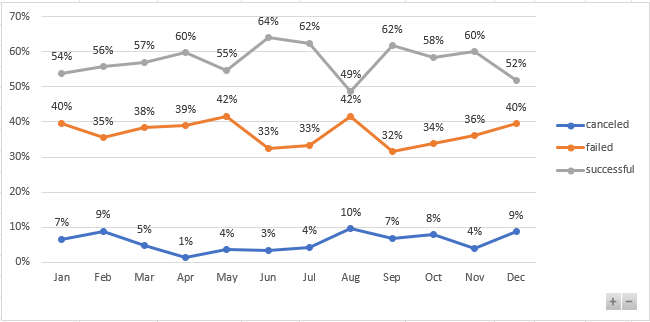
Table

Description automatically generated

2. Between two similar categories of Music and Film & Video, it is more reasonable for backers to invest in Rock Music or Documentary Film & Video. Each category has 6 sub-categories and Rock and Documentary are ranked the highest among all sub-categories. In detail, 48% of Music projects are Rock that has a moderate success rate of 58%. Similarly, Documentary dominates other sub-categories by accounting for 34% Film & Video projects and has 57% chance of success.



3. Projects created in August and December are the riskiest ones. The fail rate of these two months are the highest in comparison to other months, 42% for August and 40% for December. In addition, cancel rate of these months are peaks of the year. Therefore, August and December have the lowest success rate.



Question: What are some limitations of this dataset?

* Any fields with monetary values (Goal, Pledged) are unusable due to mixed currency when it comes to addressing the question of “how much”. They need to be recorded in uniform currency, USD for example, to be valid for data consumption
* Date should be recorded in date and time format with associated time zone. Launch and close time are important factors for investors, especially those who live in different time zones.
* Category and Subcategory should not stay in the same column. This may cause confusion and raise unnecessary work for users to clean up data. Additionally, fields that contain longer text occupy more capacity and therefore can affect loading time and performance of the dataset when it grows over time.

Question: What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

* After converting monetary fields to one currency, we can create a scatterplot between Goal and Pledged to detect any linear regression. Does higher Goal drive higher Pledged or higher Goal drive low Pledged. We can also use the linear model to predict Pledged based on Goal if it generates insignificant errors
* We can add a line graph showing Average Donation by Created Date for each category. The graph means to evaluate which category attracts more high-value backers over time.