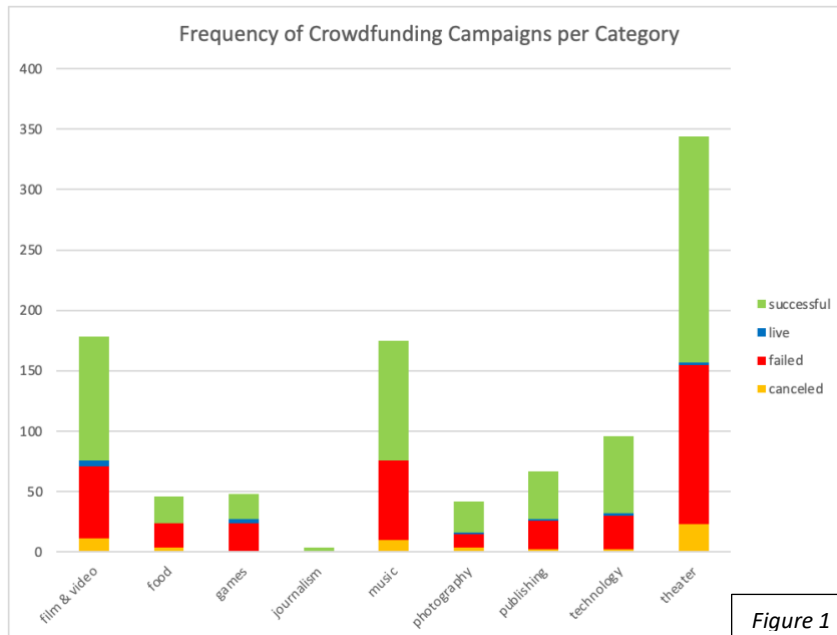


# Module 1 Report

Bilal Abdul Hak

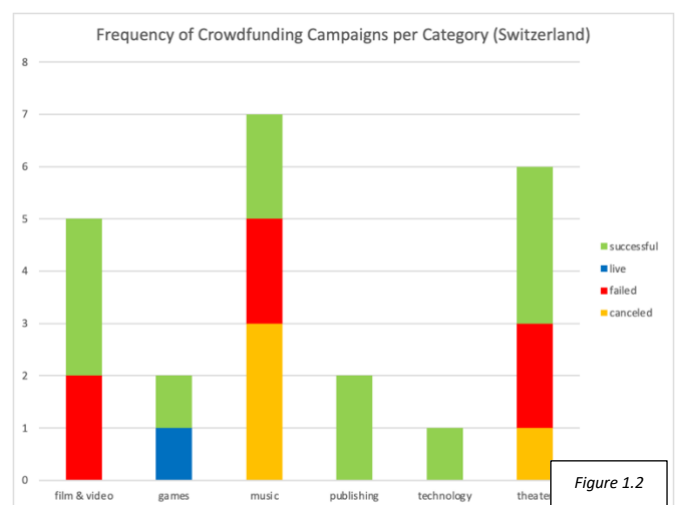
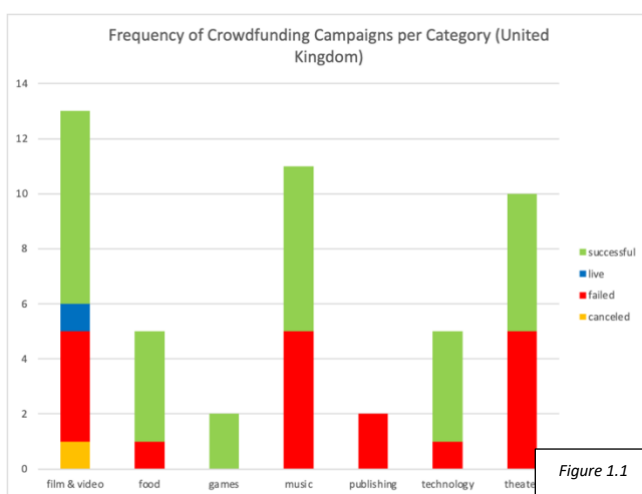
Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

Analysis:



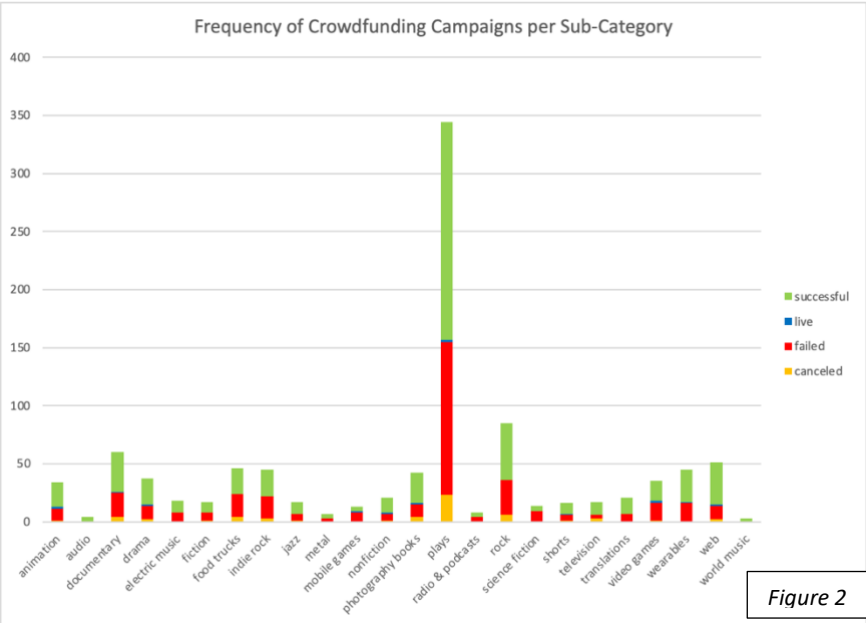
Referring to figure 1, the category with the highest number of crowdfunded campaigns is the *theater* category. Categories *music* and *film & video* are roughly tied for the second highest frequency.

A deeper analysis into the parent category data across country showcases that the United Kingdom (country code: GB) and Switzerland (CH) have differing results to the general observation. GB and CH's data are shown below in figures 1.1 and 1.2 respectively



# Module 1 Report

Bilal Abdul Hak



As shown in figure 2, the sub-category *plays* contains the highest number of campaigns by an overwhelming amount. This is a reflection its parent category, *theater's* majority as it is the only subcategory. Following it is *rock* as the second highest sub-category.

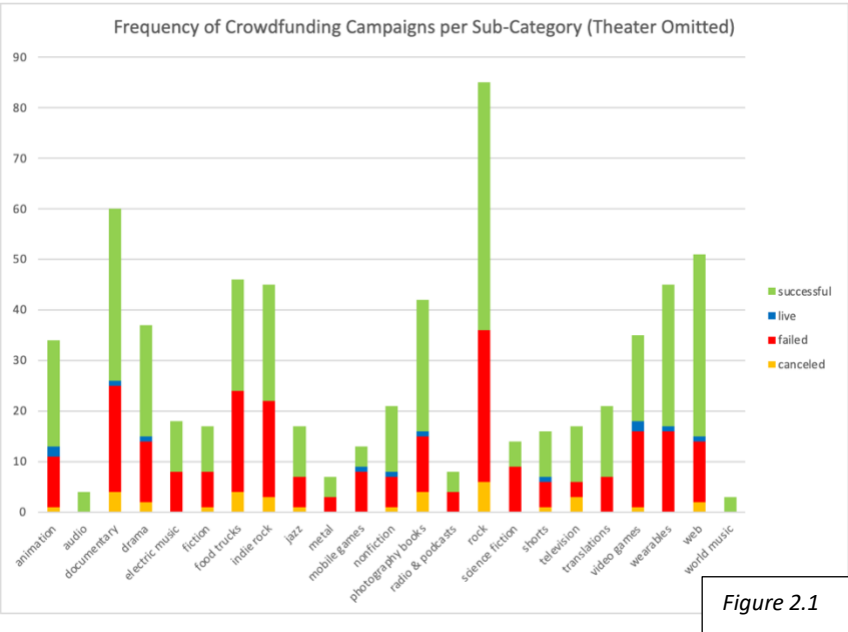


Figure 2.1 omits theater to better analyze the other sub-categories. After *rock*, *documentary* and *web* follow respectively.

A deeper analysis into the sub-category data indicates that *plays* remains with the highest count of crowdfunded campaigns across all countries.

# Module 1 Report

Bilal Abdul Hak

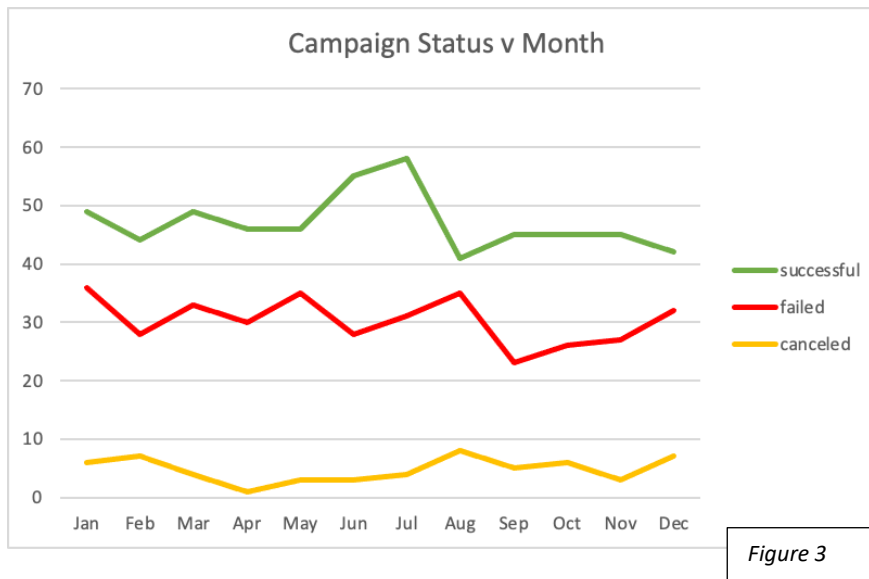
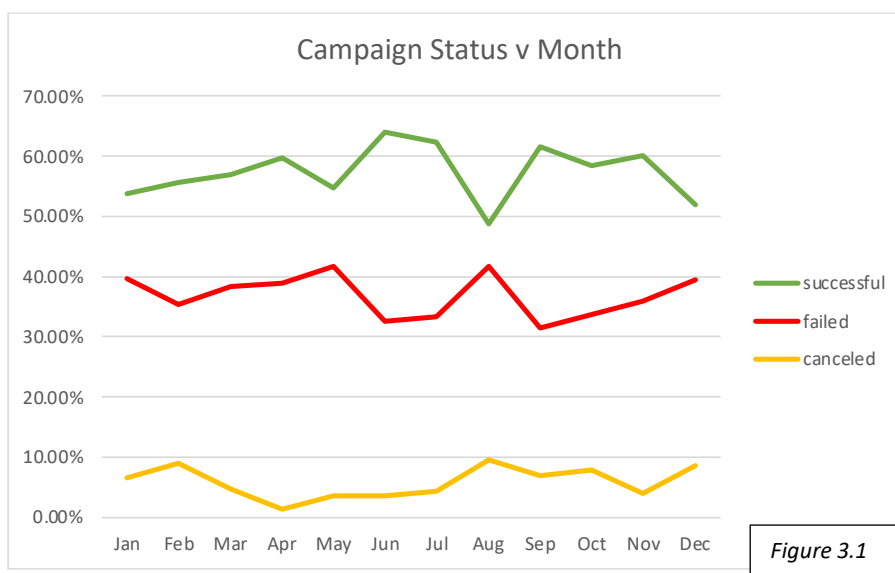


Figure 3 depicts a relationship between campaign status and month of launch; this was compiled from data of the years 2010-2020. An overall analysis of figure 3 indicates that the frequency of success is at its highest in July but quickly drops to its lowest the month after in August.

Correspondingly, both the number of failure and canceled statuses rise to their apex counts.



However, as shown below in figure 3.1, the likelihood of success always higher (albeit sometimes slightly) across the months. This chart reveals that the highest likelihood of success is during June.

A deeper analysis of the data source of this chart indicates that each year, when observed individually, is starkly different to the above overall analysis; the chance for success remains at roughly 50 percent each year.

# Module 1 Report

Bilal Abdul Hak

## Conclusions:

1. *Theater* has the highest number of campaigns among the categories except
2. *Plays*, the only sub-category for *theater*, is the highest sub-category in every country.
3. Overall, the likelihood for a crowdfunded campaign to succeed is at its highest in June; although the year-to-year proportions of success differ, at least half of the campaigns do succeed.

## What are some limitations of this dataset?

1. The dataset does not contain any data regarding crowdfunding campaigns of 2021 and 2022, and we only have two rows of data for 2020 (i.e. two campaigns)
2. There is little metadata to ascertain how valid this sample is. The prompt referred to this as a dataset of sample projects, but there is no way of seeing whether the dataset is representative of the population of crowdfunded campaigns and randomly sampled.
  - a. We do not know where this is sourced, and whether it only came from campaigns of, say, a single website. If it did, then it could not be representative of crowdfunding campaigns on other sites.
  - b. Nevertheless, each year before 2020 has around 90-100 sample projects, so there may have been an attempt at fidelity across the years; in this regard, perhaps stronger conclusions can be made only with respect to the years 2010-2019.
  - c. The lack of metadata means each variable is an assumption of what the data might mean; for example, the *staff\_pick* variable might mean campaigns that the staff of the crowdfunding service have endorsed.
3. The data contains campaigns only from the 7 countries. In addition, a great majority of the campaigns (76.3% of the entire set) is from the U.S. only. This might be because the U.S. does indeed have the most campaigns, but without knowledge of the metadata, this cannot be said with certainty. Regardless, it is possible that conclusions from this set are skewed towards American campaigns.
4. The dataset incorporates different currencies into the dataset without normalizing them. Statistics such as average donation may have to be normalized, but percent funded may serve that function.

# Module 1 Report

Bilal Abdul Hak

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

1. Tables and charts analyzing the relationship between outcome status and a few numeric statistics like funding goal could help understand trends behind what levels of funding might be successful.
  - a. In the place of outcome status, the *Percent Funded* variable can be used to understand how much of a success or failure a campaign is.
  - b. The difference between the launch date and the deadline could serve as a noteworthy variable to research. Namely, the relationship between it and the percent funded may be a point of research. This could be done with a line graph.
  - a. A normalized statistic of backer support with funding goal may be of use. Perhaps backers per 100 units\* of the funding goal and percent funded may allow forecasting how much support is needed for success. This could also be done with a line graph.
2. A bar chart depicting the relationship between *staff\_pick/spotlight* and *outcome* may provide marketing insight. This is under the assumption that spotlight refers to some sort of advertising exposure.
3. A stacked histogram showing the dispersion of funding goals and colored by outcome would visualize the how the data is dispersed and the levels of success found at each goal range.

\*donation currency would need to be normalized