Monash University: Data Analytics Boot Camp

Module 01 Challenge

Report on Crowdfunding Campaigns

*Student: Greg Presneill*

# Questions & Answers

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

1. Overall, based on this data set Crowdfunding Campaigns succeed ‘more often than not’. In other words, successful campaigns outnumbered unsuccessful (failed or cancelled) campaigns when considered across all Countries and all Categories of campaign.
2. The most popular topic for Crowdfunding across all Countries is ‘plays’.
3. Crowdfunding is most active in the US (comprising over 75% of the reported Campaigns).

## What are some limitations of this dataset?

* We can’t easily compare the financial investment across campaigns.
  + Although we were able to calculate an ‘Average Donation’ amount for each Campaign, and we know the ‘Currency’ type for each, the data set doesn’t include the currency conversion factor relevant at the time (say, an average currency conversion factor over the duration of each campaign). That makes it difficult to convert all donation amounts to a single currency, say USD, for comparison across campaigns.
* We can’t easily compare the Crowdfunding campaign success rates by Country.
  + The majority (> 75%) of the reported Campaigns are US–based. At face value that might suggest the US population is more active in this space / more entrepreneurial. However, If the data set included each country’s population size per year, then we’d be able to make a comparison of the number of Campaigns per head of population, and so be able to determine whether or not there is a similar level of Crowdfunding occurring in each country.
  + Similarly, if we could determine the number of Campaigns per head of population, we could better compare the success rate of Crowdfunding campaigns across different countries.
  + Alternatively, we could consider normalising the data, so we are comparing relative proportions of successful/live/failed/cancelled Campaigns across the various Countries, and not just the raw numbers for each of those Outcomes from each Country.

* It’s not clear whether the source of the data is from the general population, or not.
  + There seems to be a high proportion of Crowdsourcing campaigns for ‘artistic’ pursuits, with the ‘Theatre’ category having the highest number of campaigns for almost all Countries. So, is the information derived from the general population of each country, or just the ‘Arts community’, for example?
  + At face value, it appears there could be some seasonality in the data, with an apparent spike of ‘Success’ in July followed by increased ‘Failed’ and ‘Cancelled’ numbers in August. That could imply the data is skewed by the large proportion of US–based data, as those figures could align with the start and end of the US summer holiday period.

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

* We could also calculate the length (in days) of each Crowdfunding campaign, and introduce another Stacked Bar Chart to investigate whether shorter or longer campaigns were beneficial to the Outcome.
* Similarly, we could introduce another Stacked Bar Chart to investigate whether having more backers was generally beneficial to the Outcome.
* As mentioned above, if we normalised the ‘Outcome’ data from each Country, then we could create a Stacked Bar Chart that showed a comparison of the various Outcome proportions by Country. In that way, could consider if one Country tends to run more ‘Successful’ Crowdfunding campaigns than others, for example.
* A ‘treemap’ visualisation might be a better way to show the relative different Outcomes by Country, as the Category/Sub-Categories form a hierarchy that could be ‘drilled into’ (instead of, or in addition to, the separate ‘Outcome by Category’ and ‘Outcome by Sub-Category’ pivot charts).