Module 1 Report: Crowdfunding Campaigns (CCs)

By Karoly Burgyan

Note: For ease of reading, the “content creator”, “celebrity”, “organization”, or company that is under the column heading “name” (in the dataset) will be referred to simply as “company”.

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

1. By looking at the **Outcome By Category** graph and **Outcome By Sub-Category** graph, theaters and plays are the most common categories and sub-categories respectively for **crowdfunding campaigns** (**CCs**).
2. As the dataset and graphs currently are, they appear to indicate that the most successful category and sub-category that has been attempted beyond just a handful of times (i.e. e.g. excluding journalism, audio, and world music) is technology and web respectively especially when they crowdfund in June and in any of the countries evaluated with the exception of Denmark (as shown in the **Outcome By Category** graph, **Outcome By Sub-Category** graph and the **Outcome By Month** graph).
3. In reference to the **Outcome Percentage By Goal** (understanding that an assumption has been made that all the pledges and goals of the data are in the same currency (more on this later)), it seems that setting a goal between 15000 and 25000 has some special quality and is associated with success but this quality, if it exists, can’t be determined by the dataset and the graphs as they are. And the total number of CCs in this goal range are quite small so this may be a case, statistically, of noise over signal that caused the zero fail and zero cancellation counts within this goal range.
4. Extra: By looking at the **Outcome Percentage By Goal** graph (and if we assume “live” is not a *final* outcome) cancelation is the least likely *final* outcome and success is the most likely *final* outcome for CCs.

What are some limitations of this dataset?

1. As the dataset and its (the dataset’s) processing currently stands, the data gleaned from the goal and pledge columns that are used for the **Outcome Percentage By Goal** graph have been assumed to be in the same currency but this assumption seems dubious given that a currency type was listed in the same row as the pledge and goal of each CC.
2. As the dataset and its (the dataset’s) processing currently stands, it’s not clear whether the dataset has been **cleaned** of duplicate rows. Duplicate rows would bias the data towards those rows.
3. As the dataset and its (the dataset’s) processing currently stands, it’s not clear whether the dataset shows multiple **unique** CC event rows for a *single* company. If there are multiple unique CC event rows for a *single* company, then data analysis could be done to determine whether 1) the CC type or 2) the nature of the company is more influential in causing a particular outcome (i.e. do people pledge because they enjoy the CC type or do they pledge because they believe in the cause/vision of the company). And so, a company understanding that the success of its (the company’s) CC may be less determined on the type of CC and more determined by its (the company’s) nature/vision and knowing this could change how the company will proceed with “generating buzz” for its (the company’s) new product.
4. As the dataset and its (the dataset’s) processing currently stands, there has been no analysis of the “blurb”, “staff\_pick”, “spot\_light”, or “name” data that was also provided which may provide insight into the outcome of the CC types.
5. As the dataset and its (the dataset’s) processing currently stands, there has been no analysis on the signal-to-noise ratio on the data. And so, confidence in the conclusions stand on less strong grounds.

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

1. Graph the **Outcome Percentage By Goal** graph under a *new* assumption and an *new* additional level of process. The new assumption would be that the “pledge” and “goal” data listed in the origin dataset are **not in the same currency** across the different rows but, instead, the currency type of the “pledge” and “goal” of a row corresponds to the currency type stated under the “currency” column in the dataset. So, under this assumption a “new additional level of processing” is required to convert all the rows’ “pledges” and “goals” to the same currency. And then create a **New Outcome Percentage By Goal** graph.
2. On the **Outcome Percentage By Goal** graph, a filter for category and sub-categories might reveal a goal range for technology and web that produces the most success for technology and web CCs.
3. A graph showing the ***Number* of CCs By Outcome Percentage By Month And Year** could help determine if CCs were more successful if there were fewer CCs being launched that month (presumably because it gave investors/customers fewer CCs to choose to invest in, thereby enabling a higher percentage of CCs to meet their goals). –And so, one could conclude that launching one’s CC in a month with fewer competitor CC launches increases one’s odds of success.
4. Create **new** graphs of the same types that have already been done but with data that has been checked and cleaned of any duplicate rows. This would increase the confidence in the findings of the resulting **new** graph versions.
5. Create a graph of **Outcome By Spotlight**, **Outcome By Staff Pick**, and **if** there are multiple **unique** CCs for a *single* company then **Outcome By Company**. This would help determine if there is a connection between 1) outcome and spotlight, 2) outcome and staff pick, and 3) outcome and nature/vision of company, respectively.