UCB-BEL-DATA-PT-02-2020-U-C/tree/master/01-Excel/Hw

Report on Kickstarter campaign

Misha Berkowitz

February 25, 2020

**Question #1: Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**

**Conclusion #1:**

The most common terminal states for Kickstarter campaigns are success (53.1%) and failure (37.2%).

**Conclusion #2:**

Though Kickstarter campaigns originate from all over the world, the country with the most campaigns is the United States (73.8%), followed by Great Britain (14.7%), Canada (3.5%), and Australia (1.8%). From this, we can draw the conclusion that Kickstarter campaigns are most prevalent in countries whose populations primarily speak English.

**Conclusion #3:**

Most campaigns that fail don’t even come close to succeeding: fewer than 1% of failed campaigns (12 out of 1530) secured even 70% funding, and the mean funding level for unsuccessful projects was 1.2%.

**Question #2: What are some limitations of this data set?**

We do not have access to potentially informative data about funding; for example, we are unable to see when funding came in and in what amounts. Having access to that data would enable us to better understand the cadence of funding within and across different campaigns and might allow us to predict whether campaigns that are successful tend to be successful right from the start, or whether funding trickles in more gradually. A less robust but simpler version of this would be to capture the date at which projects reached any number of funding milestones (e.g., 10% Fund Date; 50% Fund Date; 100% Fund Date).

A legend or “read me” would also be helpful for explaining what some of the data signify. For instance, country codes can be guessed but are not given explicitly. It’s also not clear what “live” means, or when the data were pulled. Finally, we have no idea what the “Staff Pick” column means.

**Question #3: What are some other possible tables and/or graphs that we could create?**

It would be interesting to change the stacking format of the pivot charts to be “100% stacked,” which might allow users to more easily visualize the *relative frequency* that each category ended up in each state, rather than the *count*.

**Bonus Statistical Analysis:**

It seems like the median summarizes the data more meaningfully than the mean (no pun intended), particularly because the data are bound at the bottom by 0. For similar reasons, there is much more variance with successful campaigns; across the board, the data in *unsuccessful* campaigns tend to gravitate toward 0, which necessarily reigns in the variability in the number of backers.