Jorge Arriola Villafuerte

Data Analytics Bootcamp

University of Toronto School of Continuing Studies

Homework #1

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

*1) Regarding pivot table #1 – Number of live, canceled, failed & successful campaigns per category*

‘Theater’ and ‘music’ stand out as the two categories with the highest number of Kickstarter requests (1,393 and 700 respectively). These are also the categories with the highest ratio of successful campaigns vs. failed & canceled campaigns. ‘Music’ campaigns have 3.86 times more successful campaigns than failed & cancelled ones, while ‘theater’ campaigns have 1.58 times more successful campaigns than failed & cancelled ones.

On the contrary, there are categories that are unpopular in Kickstarter, or that do not have a high ratio of success. Journalism is the category with the lowest number of total campaigns (24) as well as the only campaign with no successful campaigns. The category with the second lowest performance is ‘food’. Only 200 Kickstarter requests exist in the dataset and they have a ratio of 0.21 successful campaigns vs. failed & cancelled ones which means that nearly 4 out of 5 campaigns fail or are cancelled.

*2) Regarding pivot table #2 – Number of live, canceled, failed & successful campaigns per subcategory.*

‘Plays’ is, by far, the most popular subcategory of Kickstarter campaigns. There are 1,066 Kickstarter campaigns for plays (taking into account all four states: successful, failed, cancelled and live ones) in the dataset we are analyzing. The subcategory with the second highest number of campaigns is ‘rock’, with 260 campaigns. ‘Plays’ has 4.1 more Kickstarter campaigns than ‘rock’.

*3) Regarding pivot table #3 – Number of canceled, failed and successful campaigns per start month of campaign.*

December has the lowest number of campaigns initiated in a month. December also has the lowest number of successful campaigns, both in absolute terms, as well as in relation to the number of campaigns started.

i.e. December only had 111 successful campaigns. The lowest record of all 12 months. Only 44% of initiated campaigns in December were successful. This is also the lowest ratio of successful vs. initiated campaigns.

On the contrary, May is the month with the highest number of campaigns initiated in a month. May is also the month with the highest number of successful campaigns, both in absolute terms, as well as in relation to the number of campaigns started.

i.e. December had 235 successful campaigns. The highest record of all 12 months. 61% of campaigns initiated in December were successful. This is also the highest ratio of successful vs. initiated campaigns.

**What are some limitations of this dataset?**

* It would be useful to know how many views each Kickstarter campaign’s webpage got. It is sound to assume that the number of backers is impacted by how well a campaign’s webpage is promoted, and a proxy for this indicator is the number of views per webpage. Our assessment is incomplete without this piece of information.
* The dataset gives us a brief description of the initiatives. However, when you request funds in Kickstarter, you also need to describe a series of benefits that backers may potentially receive if the initiative being funded is successful. The degree of thought put into developing these benefits & how attractive these are to backers will also impact whether campaigns are successful or not. This is a criterion we are not able to assess with the current dataset.
  + This is more of a qualitative criterion. To make it a quantitative one, a Yes/No column indicating whether the person running the campaign has filled out the ‘description’ section of their Kickstarter webpage, or whether they have added ‘supporting options’ based on benefits received may help us filter out those campaigns that individuals created, but where creators did not complete all the steps.

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

The exercises have allowed us to evaluate category, subcategory and start date as criteria to determine which Kickstarter campaigns are the most successful. The dataset provides even more criteria to take into consideration:

* Table & line charts for each ‘state’ of campaign per country: Are certain countries more willing to contribute to Kickstarter campaigns? Maybe we will find patterns across regions (i.e. North America, Europe).
* We can add a new column where we subtract ‘launch date’ from ‘deadline’ to determine the period of time that each campaign was available. A bar chart may allow us to observe how long successful campaigns are open for donations.
  + To achieve this, we may need to group the ‘duration of campaign’ in ranges. i.e. 1-2 weeks, 3-4 weeks, 5-6 weeks.
* A bar chart to compare the number of successful campaigns per goal set.
  + To achieve this, we may need to convert all the datapoints under ‘goal’ to the same currency. As well, we may need to group the goals in ranges. i.e. $0 - $5,000 / $5,001 - $10,000