**Ainash Toleu – Excel Homework**

**Kickstarter** – is a crowdfunding service that launched more than 300,000 projects and raised over $2 billion.

For this week's homework, I organized and analyzed a selected database of 4,000 projects from launched campaigns on Kickstarter.

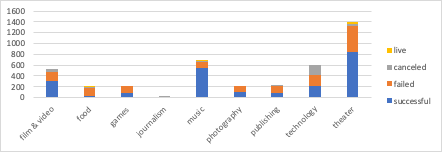
1. **Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**

* From the **Pivot\_Table\_1 tab** we created, it is evident that compared to others,
* “Theater”,
* “Music” and
* “Film & Video”

categories had more successful launched campaigns,

where the **“Music”** category was **the most successful** in all the countries one with 540 successful (120 failed) out of total 700 launches.

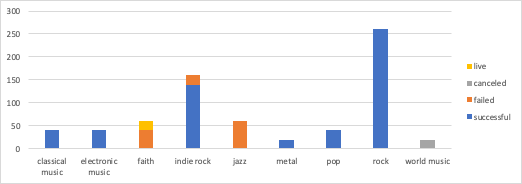
And **the least successful campaigns** were in the **“Food”** category with 140 failed campaigns (34 successful) out of total 200.



* If we analyze the information on a sub-category level (**Pivot\_Table\_2**), then we can see that
  + within the successful “Music” category:
* “classical music”,
* “electronic music”,
* “metal”,
* “pop”,
* “rock” subcategories **had 100% success**,

The “jazz” subcategory failed all 60 out of its launched campaigns and

the “faith” subcategory had 40 failed campaigns with another 20 still live. If previous performance is any indication of past success, we would expect these 20 faith campaigns to also fail.

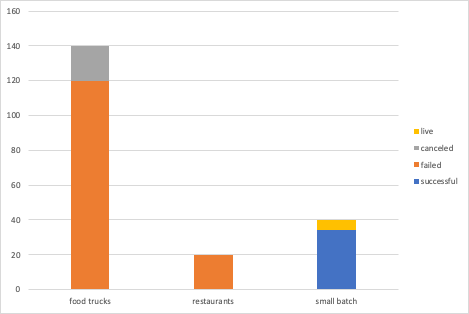


* + From the least successful “Food” category:

“restaurants” subcategory was 100% failed and

“food trucks” subcategory was failed with 120 campaigns out of 140 (20 campaigns are still live). Again, if previous performance is any indication, we would expect these 20 live 'food trucks' campaigns to also fail.

The “Small batch” subcategory showed 34 successful campaigns out of 40 (6 are live).



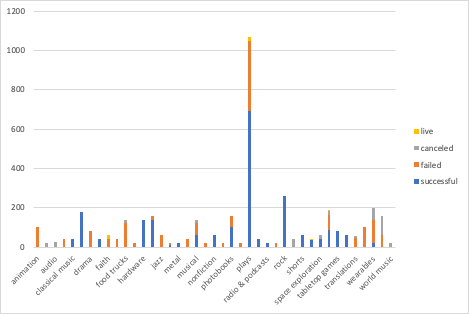
* + by analyzing all subcategories, we can see that:

**100% success in launches** **had subcategories:**

* “classical music”,
* “documentary”,
* “electronic music”,
* “hardware”,
* “metal”,
* “nonfiction”,
* “pop”,
* “radio & podcasts”,
* “rock”,
* “shorts”,
* “tabletop games”,
* “television”

**100% failed campaigns had subcategories:**

* “animation”,
* “children’s books”,
* “drama”,
* “fiction”,
* “gadgets”,
* “jazz”,
* “mobile games”,
* “nature”,
* “people”,
* “places”,
* “restaurants”,
* “video games”



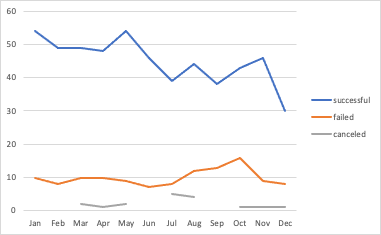
* If we analyze the campaigns by the month (**Pivot\_Table\_3**),
  + we can see that campaigns launched in the month of May had the highest count of successful campaigns during the year, at 234. The month with the highest count of failed campaigns were July (150), closely followed by January (149), and October (149).



* + if we look at our the most and least successful categories (“Music” and “Food”),

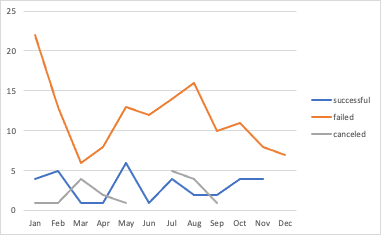
we can see that for “Music”

* the most successful months were January and May (54 successful campaigns each)
* the least successful month was October with 16 failed campaigns



for “Food” category:

* the most successful month was May, with 6 successful campaigns
* the least successful month was January, with 22 failed campaigns



Across the various categories, it seems that campaigns launched in the month of May had the most success.

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

When looking at the "state" column, there are 4 different types of campaigns: campaigns that were successful, failed, canceled, or is currently live. The "canceled" category raises some questions, because it is not determined why those campaigns were canceled. Would cancellation mean a failure because the campaign organizer lost determination, or perhaps a success because the organizer was able to meet their financial goals with donations outside of Kickstarter? Additional information would be needed to better analyze this data.

The launch and deadline dates provided were unix timestamps and not very user friendly. Providing the dates in a standard MM/DD/YYYY format would have been more straightforward and avoid additional calculations.

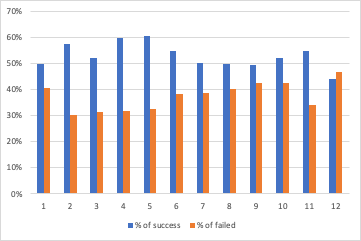
The category and sub-category would ideally be separated, and we wouldn't have to perform additional work to separate those using the "/" sign. The situation would have been more complex if the "/" sign wasn't used consistently to separate the categories and sub-categories.

Individual donation-level data would add a new perspective and complexity to the results, since we could then determine if the campaigns were successful because of a few individuals' large donations, or a 'team-effort' approach with most of the donations falling within a small range.

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

* When analyzing campaigns by month, we could analyze the data by % of successful and failed campaigns in each month.

It would then be evident that the most successful launch months were May (61%) and April (60%). The worst month was December with 47% of campaigns launched failed.



Looking at the data on relative, rather than absolute terms (**Pivot\_Table\_3**) adds additional analysis opportunities.

* Furthermore, we could analyze successful and failed campaigns by country.

Then we would see that “Music” (for example) category had success performance not only in overall countries, but almost in all of the countries:

* IT, AT, AU, DK, SE, ES had 100% success
* FR, CA, GB and US had overall success over 70%
* And only DE had 50%

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **music** | **540** | **77.14%** | **120** | **17.14%** |
| US | 490 | 77.04% | 110 | 17.30% |
| GB | 25 | 73.53% | 7 | 20.59% |
| CA | 10 | 83.33% | 1 | 8.33% |
| FR | 4 | 80.00% |  | 0.00% |
| ES | 3 | 100.00% |  | 0.00% |
| SE | 2 | 100.00% |  | 0.00% |
| DE | 2 | 50.00% | 2 | 50.00% |
| DK | 1 | 100.00% |  | 0.00% |
| AU | 1 | 100.00% |  | 0.00% |
| AT | 1 | 100.00% |  | 0.00% |
| IT | 1 | 100.00% |  | 0.00% |

* We don’t analyze the data given by year, were there any successful years compared to others or not? Presumably, campaign success and the overall economic conditions would correlate positively, and campaigns launched in 2009 would have less favorable outcomes than campaigns launched more recently.
* In our analysis, we don’t consider the actual amount of money that was raised for the campaigns. Since the amounts were in currencies other than $USD, additional calculations to convert these amounts to $USD would have helped analyze the data from a different perspective. Additionally, we calculate, but don't analyze how many people participated in the campaigns including the average donation per person, per campaign.
* Analysis could be performed on the date period between the launch and the end date of the campaign. Presumably, campaigns with longer periods of time between the launch and deadline date would have higher success rates than those with shorter dates. A longer period to allow donations would allow for more time and people to donate.

**Bonus Statistical Analysis**

1. **Use your data to determine whether the mean or the median summarizes the data more meaningfully.**

**Mean** – average, take all the values, add them up and divide by the number of values we have.

**Median** – this is the value in the middle, takes all the values, sort them from smallest to largest and then find the value in the middle of the list.

According to **Bonus Statistical Analysis** table:

* average number of backers for a successful campaign: 194.43;
* average number of backers for a failed campaign: 17.71;
* the median number of backers for a successful campaign: 62;
* median number of backers for a failed campaign: 4

By given data we can say that **Median** summarizes the data more meaningfully. It shows us the average of backers in all the campaigns.