Summary of Kickstarter Data Analysis

# overview

Kickstarter has become a popular crowdfunding platform for creative projects. Over $5 billion US dollars have been pledged to over 190,000 successfully funded Kickstarter projects since its inception. As of December 2020, Kickstarter projects have a success rate of 38.33%. In this analysis I attempt to identify patterns that may contribute to the successful funding of a Kickstarter campaign.

## Scope

There are potentially many factors that contribute to the success of a Kickstarter project. Some of those are qualitative factors like the design of the projects’ Kickstarter web page & video, the types of rewards and structure of the reward tiers, the outreach and marketing done to promote and follow up with potential backers, etc. This analysis will not address any of those factors but will focus entirely on quantitative measures.

# data & methods

I was provided a dataset by BCS that was likely produced by web scraping the Kickstarter website (as Kickstarter’s public APIs seem to be quite limited). I prepared the data and performed the analysis that is documented in the accompanying Excel spreadsheet:

* Analyzed the relationship between the success rate of campaigns, the Kickstarter project category & sub-category (Film & Video, Dance, Photography, etc.) and looked for geographical trends.
* Analyzed the relationship between the success rate of campaigns and their launch dates, looking for seasonal trends.
* Analyzed the relationship between the success rate of campaigns and the amount of their funding goal, looking for successful project size trends.
* Analyzed the relationship between the success rate of campaigns and the number of backers that pledged money to the project, looking for trends in the number of backers.
* Analyzed the relationship between the success rate of campaigns and the day of the week that they were launched, looking for trends indicating one day is more successful than others for launching campaigns.
* Analyzed the relationship between the success rate of campaigns and their average duration to determine whether projects of different lengths have been more successful.
* Analyzed the relationship between the success rate of campaigns and the average amount of donations, to see if donors of successful projects tended to donate more or less, on average.

## Limitations of the Data

I am uncertain of the provenance and accuracy of the original dataset, as I did not collect the data. It had some quality issues. It is clearly not a representative sample of random Kickstarter campaign data. I also wish it would have contained some additional data for analysis.

A bit more on these observations:

* Quality of sample: The historical success rate of Kickstarter projects is 38.33% (as published on <https://www.kickstarter.com/help/stats>). But the overall success rate of this dataset was almost 60%.
* Quality of data: the dataset was missing sub-category values and had some duplicate rows, etc. that I cleaned up & imputed.
* Limited project data: The dataset did not include some of the data that I would have liked to analyze: the number of reward tiers per campaign, for example. According to interviews that I read with successful project owners they claim that the reward structure is important (“Keep it simple.”). I would have liked to analyze whether campaigns with smaller numbers of rewards tended to be more successful but was unable to with this dataset.

## Statistical Analysis

This dataset shows that the number of backers of *successful* projects is quite skewed towards the upper range, and that the median would serve as a much better measure of central tendency (this is not the case with the backer counts of *unsuccessful* projects). As is illustrated in the charts & calculations on the “Statistical Summary” tab of the accompanying Excel spreadsheet, there are 250 outliers: successful projects with more than 350 backers (some of them with many thousands of backers).

The successful campaign data is extremely variable, as evidenced by the huge variance and standard deviation shown on the Statistics Summary excel sheet…. much more so than the unsuccessful project data. This makes sense given that many of the funded projects are huge (have thousands of backers) while the unsuccessful projects never really get off the ground precisely because they can’t attract enough backers (they have 10 on average, with a maximum of 363).

# conclusions

Based on the dataset provided and the visualizations required by the assignment I have drawn the following conclusions about successful Kickstarter campaigns:

1. Bigger is NOT better - smaller projects tend to be much more successful.
   1. Most (82%) of successfully funded projects raised less than $15,000.
2. December is a terrible month to launch a campaign.
   1. The winter months are unpopular times for launching campaigns, and with good reason…. the success rates tend to be quite low. November, December, January & February average the smallest number of successful projects launched of all the months of the year.
3. You are practically guaranteed success if you launch a Product Design, Accessories, Documentary (film), Fiction, Nonfiction, Indie Rock or Rock project.
   1. Of the 513 projects launched in these sub-categories not one of them failed!

But I have a **giant disclaimer**! I don’t believe that the dataset used was a representative random sample of Kickstarter campaign data (see earlier comments in “Data Limitations” section) – therefore *any conclusions drawn from this dataset likely have little connection to reality*. And, as always, ***correlation does not imply causation***….so your mileage may vary when following these suggestions.

## some additional findings

Based on some additional visualization & analysis, I have drawn a few additional conclusions that are worth mentioning:

* Launch your project on a Tuesday to increase your chances of success.
  + Many more successful projects were launched on Tuesdays than any other day of the week.
* A campaign duration of 3-4 weeks is optimal.
  + Most successful campaigns were between 21-30 days
* Funders of successful projects give larger donations.
  + Donors to successful projects gave pledges that were 2 to 3 times larger, on average, than donors of unsuccessful projects.

The supporting analysis & visualizations for these additional conclusions is contained in the “Extra -…” tabs of the accompanying Excel spreadsheet.