**Excel Homework: Kickstart My Chart**

Over $2 billion has been raised using the massively successful crowdfunding service, Kickstarter, but not every project has found success. Of the more than 300,000 projects launched on Kickstarter, only a third have made it through the funding process with a positive outcome.

Getting funded on Kickstarter requires meeting or exceeding the project's initial goal, so many organizations spend months looking through past projects in an attempt to discover some trick for finding success. For this week's homework, you will organize and analyze a database of 4,000 past projects in order to uncover any hidden trends.

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

Kickstarter projects are more focused on theater, music and technology (34%,17% and 15% respectively). The least used categories to look for funding on this platform are Journalism, and food (0.6% and 4.9% respectively).

Successful projects in Kickstarter correspond to theater and music (38% and 25% respectively). Failed projects are concentrated in theater and technology (32% and 14%).

All projects related to Journalism in the sample analyzed where canceled (24 projects, 0.6% of total) which could indicate that Kickstarter is not an adequate platform to fund this category.

Categories with the higher success rates are music (77% success, vs 14% failure), theater (60% success, vs 35% failure) and film & video (58% success, vs 35% failure) which can lead us to think that people are more likely to back projects related to Arts and entertainment. Subcategory Plays is the highest ranking in successful projects but also on submitted and failures.

Technology projects have 34.8% success rate, 35.5% failure rate and 29.7% cancellation rate which is interesting because even though a project has almost the same rate for success or failure (virtually a coin toss), it appears that funding seekers are just as likely to cancel the projects.

When you see data in a monthly array it becomes clear that success rate for projects is higher in the first half of the year (with the max value during Feb) and has the lowest value during Dec. If you were to start looking to fund a project with Kickstarter, you should do it between Feb and June and avoid the Q4 holidays.

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

We are looking at a small sample (4k vs 300k, ~0.15%) with no information to determine if this is representative of the population. Since we have no info on how it was obtained, we should be careful on drawing conclusions because we don’t know what type of bias it could have (i.e. only 24 Journalism cancelled projects appear on the base but maybe there are 100k Journalism projects that got successful funding in the rest of the population).

Other limitations are that we don’t know nothing around the projects to assess if backers are the only determinant factor for a project to be successful (i.e. if projects had a submission data quality or completeness qualification).

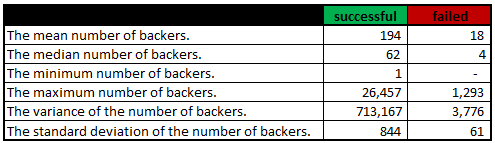
1. ***What are some other possible tables and/or graphs that we could create?***

I created tables and graphs to look at the data in percentages to make comparison between categories easier. We could have also analyzed the country of origin for the projects to look for different trends.

Using the creation and End dates we could have calculated the project funding duration and group them in intervals to analyze them. My hypothesis is that projects with lunger duration are probably more likely to get funding vs those that have shorter funding time. Also, we may be able to find an optimum funding period (i.e. projects with very long funding times may lose backers interest.

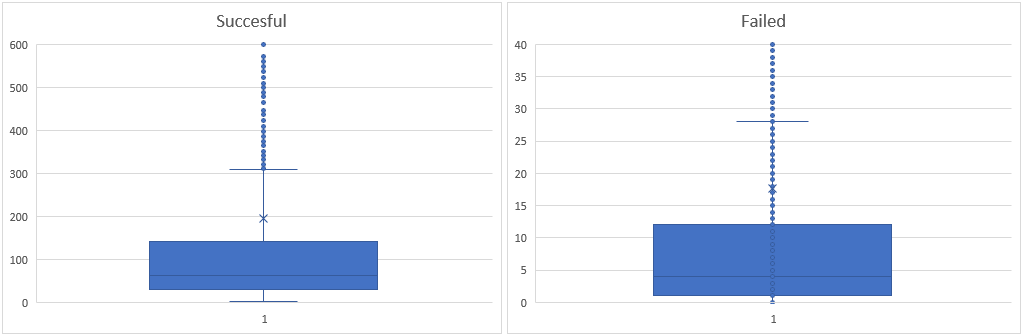
**Bonus Statistical Analysis**

If one were to describe a successful crowdfunding campaign, most people would use the number of campaign backers as a metric of success. One of the most efficient ways that data scientists characterize a quantitative metric, such as the number of campaign backers, is by creating a summary statistics table.

****

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

For this dataset, the median is the more reliable metric to summarize the data. Standard deviation for the data set is too high which is why using the mean will not describe our data accurately. This is visually explained in the below charts.

****

1. ***Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?***

There is more variability in successful campaigns, making them harder to predict. Failed campaigns have a lot less variability (success campaigns σ = 844 vs failed σ = 61). All failed campaigns have something in common, a very low number of backers (median = 4) and not reaching the funding goal vs successful campaigns that have no “roof” for the criteria (i.e. biggest successful campaign by funded percentage had a 2,260,300% of funding) introducing a lot of noise and variability to the data.