

Excel Challenge

CASE STUDY FUNDRAISING

Orlando Ortega | Data Bootcamp Rice University | 11/1/2020

# Introduction

This report is for Homework 1 of the Data and Visualization Boot Camp At Rice University. For this homework we were tasked to analyze a database of fundraising campaigns to try to understand why some are successful and others are not. We need to analyze the behavior of fundraisings according to Categories and through time. We needed to prepare the data first to create pivot tables appropriate for the analysis and finally discuss any insights that we can have from the behaviors observed.

## Instructions

Below are the instructions provided:

**# Excel Homework: Kickstart My Chart**

**## Background**

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.

**### Before You Begin**

1. Create a new space for this project called `excel-challenge` in either DropBox or Google Drive. **\*\*Do not add this homework to an existing space\*\***.

2. Store your excel workbooks in here and create a sharable link for submission.

**## Instructions**

![Kickstarter Table](Images/FullTable.png)

Using the Excel table provided, modify and analyze the data of 4,000 past Kickstarter projects as you attempt to uncover some market trends.

\* Use conditional formatting to fill each cell in the `state` column with a different color, depending on whether the associated campaign was successful, failed, or canceled, or is currently live.

  \* Create a new column O called `Percent Funded` that uses a formula to uncover how much money a campaign made to reach its initial goal.

\* Use conditional formatting to fill each cell in the `Percent Funded` column using a three-color scale. The scale should start at 0 and be a dark shade of red, transitioning to green at 100, and blue at 200.

  \* Create a new column P called `Average Donation` that uses a formula to uncover how much each backer for the project paid on average.

  \* Create two new columns, one called `Category` at Q and another called `Sub-Category` at R, which use formulas to split the `Category and Sub-Category` column into two parts.

  ![Category Stats](Images/CategoryStats.png)

  \* Create a new sheet with a pivot table that will analyze your initial worksheet to count how many campaigns were successful, failed, canceled, or are currently live per **\*\*category\*\***.

  \* Create a stacked column pivot chart that can be filtered by country based on the table you have created.

  ![Subcategory Stats](Images/SubcategoryStats.png)

  \* Create a new sheet with a pivot table that will analyze your initial sheet to count how many campaigns were successful, failed, or canceled, or are currently live per **\*\*sub-category\*\***.

  \* Create a stacked column pivot chart that can be filtered by country and parent-category based on the table you have created.

\* The dates stored within the `deadline` and `launched\_at` columns use Unix timestamps. Fortunately for us, [there is a formula](https://www.extendoffice.com/documents/excel/2473-excel-timestamp-to-date.html) that can be used to convert these timestamps to a normal date.

  \* Create a new column named `Date Created Conversion` that will use [this formula](https://www.extendoffice.com/documents/excel/2473-excel-timestamp-to-date.html) to convert the data contained within `launched\_at` into Excel's date format.

  \* Create a new column named `Date Ended Conversion` that will use [this formula](https://www.extendoffice.com/documents/excel/2473-excel-timestamp-to-date.html) to convert the data contained within `deadline` into Excel's date format.

  ![Outcomes Based on Launch Date](Images/LaunchDateOutcomes.png)

  \* Create a new sheet with a pivot table with a column of `state`, rows of `Date Created Conversion`, values based on the count of `state`, and filters based on `parent category` and `Years`.

  \* Now create a pivot chart line graph that visualizes this new table.

\* Create a report in Microsoft Word and answer the following questions.

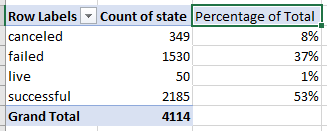
1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?

2. What are some limitations of this dataset?

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

# ANALYSIS

# Count of state of fundraising events



# Count of fundraising events per category and status

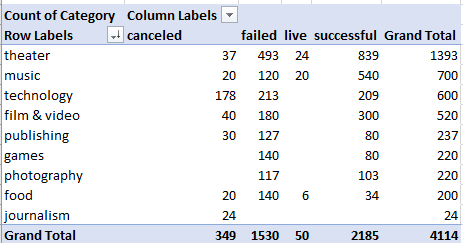
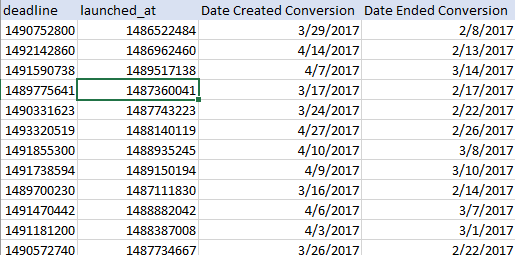


Table showing the distribution of fundraising state per category. The table can be filtered by category and state using the choices in the pull-down menu.

Figure 1. Graph showing the number of total fundraising events per category and their relative split in different outcomes.

# Date data conversion exercise

Below is an extract of the results obtained converting integers to dates using the formula =(((B2/60)/60)/24)+DATE(1970,1,1).



# Variation of fundraising state through the year

The time stamps from the exercise above were used to construct the following table:

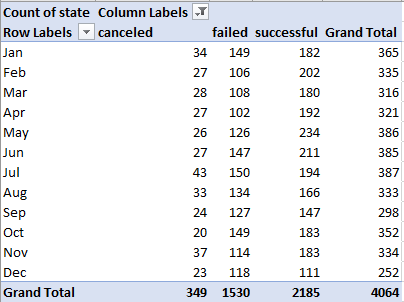


Table that can be filtered by month and fundraising status to obtain the observed success, failure, live or canceled number of fundairings.

Figure 2. Graph showing the number of total fundraising events and their relative split in different outcomes of the fundraising event across the calendar year.

# QUESTIONS

# Mention 3 conclusions that we can draw from the homework

* 1. Most Fundraising events are successful (53%) but the failure rate is significant (37%).
  2. Most fundraisings are in the “theater” category, followed by “music” and “technology”. “Technology” seems to have a high chance of being canceled compared to other categories.
  3. The worst month for successful fundraising in general is December. The highest fundraising activity occurs May through July.

# What are some limitations of this dataset?

* 1. The categories are not very well defined. For example: Is a musical fundraising a theater or a music event? What is technology?
  2. The state “canceled” is ambiguous and although we can assume that these fundraisings were canceled because they were failing it is just an assumption.
  3. The data does not distinguish new fundraising events versus “old” (i.e. traditional) which could also influence the reasons for success or failure.

# What are some other tables or graphs that we could create?

* 1. We can create the same graphs but using normalized data, for example success or failure can be skewed to how ambitious was the goal.
  2. Similarly, a graph of the time spent in the fundraising event versus other parameters may show the correlation (or lack thereof) between time spent in fundraising and rate of success.
  3. We can try to discount holidays (like Christmas, or Thanksgiving) and normalize the data by actual number of days outside these special holidays.
  4. We can use the sub-category qualifier to see if faith-based fundraisings have a higher rate of success than other sub-categories.

# BONUS 1

Distribution of fundraising status according to their fundraising goal amount and calculated percentages.

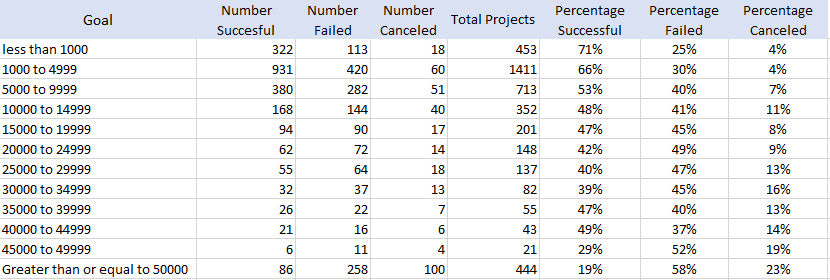
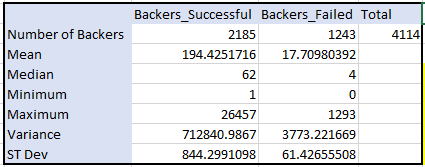


Figure 3. Variation of the percentage of fundraising campaigns that succeed, fail or get canceled against the goal set up at the start.

# BONUS 2

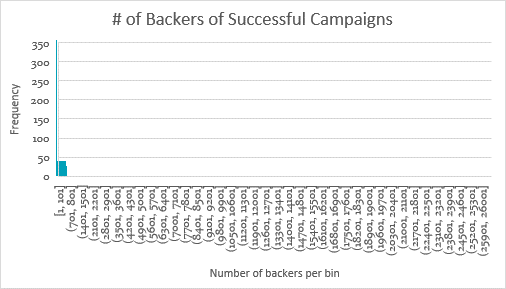
Summary statistics of the number of backers of fundraisings for successful and failed outcomes.



The variability in the number of backers for successful events is greater than for the failed events. This makes sense because successful campaigns will have greater number of backers with some occasional solo big backers of certain causes. The numbers of backers varies from very few that contribute a lot to large numbers that in average contribute less to most successful fundraisings.

# Is the mean or the mean or the median more meaningful?

The histogram below and the statistical summary shows that the number of backers is highly skewed to small number of backers for most fundraising events. The median of a highly skewed distribution is a better representation of the expectation. The arithmetic average (mean) is not a particularly good representation of the behavior of this distribution.



# Is there more variability in the number of backers with successful or unsuccessful events

The variability in the number of backers for successful events is greater than for the failed events. Fundraising campaigns can succeed with large number of small donations or few large donations of sponsors. Unsuccessful events have a smaller number of backers, in general. The variability of the spread of the data is provided by the variance and the standard deviation of the distribution.