**Kickstarter Project Success Analysis**

**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.

*Source: Excel Homework:* [*Kickstart My Chart ReadMe*](https://smu.bootcampcontent.com/SMU-Coding-Bootcamp/SMU-DAL-DATA-PT-11-2019-U-C/tree/master/02-Homework/Instructions)

Every project creator sets their project's funding goal and deadline. If people like the project, they can pledge money to make it happen. If the project succeeds in reaching its funding goal, all backers' credit cards are charged when time expires. Funding on Kickstarter is all-or-nothing. If the project falls short of its funding goal, no one is charged.

If a project is successfully funded, Kickstarter applies a 5% fee to the funds collected.

Backers pledge money for different reasons. Some backers are rallying around their friends' projects. Some are supporting people they've long admired. Many are just inspired by a new idea. Others are inspired by a project's rewards — a copy of what's being made, a limited edition, or a custom experience related to the project.

Project creators keep 100% ownership of their work. And Kickstarter cannot be used to offer financial returns or equity, or to solicit loans. Some projects that are funded on Kickstarter may go on to make money, but backers are supporting projects to help them come to life, not financially profit.

*Resource(s):* [*Kickstarter and Taxes*](https://www.kickstarter.com/help/taxes)*,* [*Kickstarter Basics*](https://help.kickstarter.com/hc/en-us/articles/115005028514-What-are-the-basics-)

**Decompose the Ask**

…so many organizations spend months looking through past projects in an attempt to **discover some trick for finding success**. …organize and analyze a database of 4,000 past (Kickstarter) projects in order to **uncover any hidden trends**.”

* Over $2 billion has been raised (using the Kickstarter) crowdfunding service
* not every project has found success
* …more than 300,000 projects launched… only a third (~100,000) have made it through the **funding process** with a **positive outcome; state =? success (see Data Source)**
  + Getting funded on Kickstarter requires **meeting or exceeding the project's initial goal**

*Need to define ‘positive outcome’.*

*Organization finding projects that should be successful? Yes*

*Projects that are likely to be funded? Yes*

*positive outcome = getting funded  
- Is the given 4,000 record database a significant sample?* ***Will need to use statistics***

***Restatement of the problem:*** *Are there statistically significant trends among past projects in the given dataset (4,000+ records) that are helpful when predicting project success or failure?*

*Need to consider if the likelihood of going ‘live’ is a factor for success definition and to what extent. That is to say, is it only important to meet the funding goal, or is the possibility of going ‘live’ also important?*

**Identify Data Sources**

Given: StarterBook.xlsx  
<https://smu.bootcampcontent.com/SMU-Coding-Bootcamp/SMU-DAL-DATA-PT-11-2019-U-C/tree/master/02-Homework/01-Excel/Instructions>

**Define Strategy and Metrics**

Given: See Data Source

**Build Data Retrieval Plan**

Given: See Data Source

**Retrieve the Data**

Given: See Data Source

**Assemble and Clean**

Given: See Data Source

**Analyze for Trends**

* [done] Use conditional formatting to fill each cell in the **state** column with a different color, depending on whether the associated campaign was
  + successful, **[green]**
  + failed, or **[dark red]**
  + canceled, or is [yellow]
  + currently live. **[dark blue]**
* [done] 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.
* [done] 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.
* [done] Create a new column P called **Average Donation** that uses a formula to uncover how much each backer for the project paid on average.
* [done] 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.  
  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* [done] 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*.
* [done] Create a stacked column pivot chart that can be filtered by country based on the table you have created.  
  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* [done] 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*.
* [done] Create a stacked column pivot chart that can be filtered by country and parent-category based on the table you have created.  
    
  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* [done] Create a new column named **Date Created Conversion** that will use this formula to convert the data contained within **launched\_at** into Excel's date format.
* [done] Create a new column named **Date Ended Conversion** that will use this formula to convert the data contained within **deadline** into Excel's date format.  
    
  *\* The dates stored within the deadline and launched\_at columns use Unix timestamps. Use this formula to convert: =(((J2/60)/60)/24)+DATE(1970,1,1)*  
  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* [done] 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.
* [done] Now create a pivot chart line graph that visualizes this new table.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[done] Bonus

* Create a new sheet with 8 columns:
  + Goal
  + Number Successful
  + Number Failed
  + Number Canceled
  + Total Projects
  + Percentage Successful
  + Percentage Failed
  + Percentage Canceled
* In the Goal column, create 12 rows with the following headers:
  + Less than 1000
  + 1000 to 4999
  + 5000 to 9999
  + 10000 to 14999
  + 15000 to 19999
  + 20000 to 24999
  + 25000 to 29999
  + 30000 to 34999
  + 35000 to 39999
  + 40000 to 44999
  + 45000 to 49999
  + Greater than or equal to 50000
* Using the COUNTIFS() formula, count how many successful, failed, and canceled projects were created with goals within the ranges listed above. Populate the **Number Successful, Number Failed,** and **Number Canceled** columns with this data.
* Add up each of the values in the **Number Successful, Number Failed,** and **Number Canceled** columns to populate the **Total Projects** column. Then, using a mathematical formula, find the percentage of projects that were successful, failed, or canceled per goal range.
* Create a line chart that graphs the relationship between a goal's amount and its chances at success, failure, or cancellation.

**Acknowledge Limitations**

What are some limitations of this dataset?

**Determining a trend may be challenging given the sample size**

The 4,114 dataset is a 1.37% sample of the more than 300,000 projects. R Squared calculations were never larger than .0206 when comparing; state to date-created, state to category, and state to sub-category. Significance F and P-values for the same are all 0 or lower.

So with low R Squared numbers, and with low Significance F and P-values, this dataset is not likely to prove useful when trying to explain variation between a project’s state relative to other independent variables, but this dataset is significant and can lead to insights that will drive decision on how to invest in Kickstarter projects.

**Make the Call or Tell the Story**

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

* Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?  
  + Theater projects, Music projects, and Film & Video projects are the most likely to achieve their funding goal, with 3% of Music projects and 2% of Theater projects going live.
  + Plays overwhelmingly receive the largest amount of pledges with close to 67% meeting their target. Honorable mention: All projects in the *small batch(food)* sub-category met their pledge goal with 15% of them going live. The *Faith(music)* sub-category is the only other group with a high go live rate at 33%  
      
    If the goal is to meet your pledge target, then choose a play. Consider choosing a faith music or small batch food if you are interested in achieving success in a more niche area.
  + Project funding achievement across all years is highest in the spring, starting in March and peaking in May before taking a sharp decline until September.   
      
    So an April project launch is ideal as trends show Kickstarter funding is at a sharp increase.
* What are some other possible tables and/or graphs that we could create?