**Assignment 1 [Excel]**

**Saranya P**

<<Data Analytics >>

**Kickstart My Chart**

**Table of Contents**

[**Background** 2](#_Toc16884343)

[**Objective** 2](#_Toc16884344)

[**Instructions** 2](#_Toc16884345)

[**Kickstarter - FullTable** 2](#_Toc16884346)

[**Category Stats** 3](#_Toc16884347)

[**Subcategory Stats** 4](#_Toc16884348)

[**Launch Date** 5](#_Toc16884349)

[**Report** 6](#_Toc16884350)

[**Conclusion** 6](#_Toc16884351)

[***Outcome Based on Category*** 6](#_Toc16884352)

[***Outcome Based on Sub Category*** 7](#_Toc16884353)

[***Outcome Based on Launched Date*** 9](#_Toc16884354)

[***Outcome Based on Goals*** 10](#_Toc16884355)

[**Final Conclusion** 11](#_Toc16884356)

[**Limitations** 11](#_Toc16884357)

[**Other Possibilities** 12](#_Toc16884358)

[**Bonus** 12](#_Toc16884359)

[**Goal Outcomes** 13](#_Toc16884360)

**# Unit 1 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 analyse a database of 4,000 past projects in order to uncover any hidden trends.

# **Objective**

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

# **Instructions**

## **Kickstarter - FullTable**

\* 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 Cancelled, or is currently live.

**<Sara Comment>:**

***The below formatting has been done for the state column.***

* ***Successful - Green Fill with Dark Green Colour Text***
* ***Failed - Red Fill with Dark Red Colour Text***
* ***Cancelled - Yellow with Black Colour Text***
* ***Live - Blue with Black Colour Text***

\* 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.

**<Sara Comment>:**

***Column O “Percent Funded” has been created.***

***Formula Used to derive the Percent Funded column= Pledged/Goal for which the format has been changed to Percentage Style.***

\* 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.

**<Sara Comment>:**

***Conditional Formatting Rule has been applied with a format style as 3-Colour Scale (Red Green Blue) with its corresponding values.***

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

**<Sara Comment>:**

***Column P “Average Donation” has been created.***

***Formula Used to derive the Average Donation column= Pledged/Backers\_count for which the format has been changed to Number Style.***

***For few row values, Backers\_count was zero, hence Divide by Zero rule has been applied for the same.***

\* 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.

**<Sara Comment>:**

***Column Q “Category” and Column R “Sub-Category” has been created.***

***The Column N “Category and Sub-Category” has been converted from Text to Columns (Under Data Tab) by providing the Delimiter as “/” and the Destination as Q2:R4115***

## **Category Stats**

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

**<Sara Comment>:**

***Pivot table has been created in a new sheet “Category Stats Pivot” having campaigns outcomes based on the category.***

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

**<Sara Comment>:**

***Stacked Column Pivot Chart has been created by choosing the Stacked Column pivot chart type in a sheet “Category Stats Pivot”. This Pivot Chart shows the campaigns outcomes based on the category which can be filtered by country.***

## **Subcategory Stats**

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

**<Sara Comment>:**

***Pivot table has been created in a new sheet “Sub Category Stats Pivot” having campaigns outcomes based on the sub category.***

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

**<Sara Comment>:**

***Stacked Column Pivot Chart has been created by choosing the Stacked Column pivot chart type in a sheet “Sub Category Stats Pivot”. This Pivot Chart shows the campaigns outcomes based on the sub category which can be filtered by country.***

\* 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.

**<Sara Comment>:**

***Column S “Date Created Conversion” has been created.***

***Formula Used to derive the Date Created Conversion column =***

***(((launched\_at column /60)/60)/24)+DATE(1970,1,1)***

***The Column has been formatted in 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.

**<Sara Comment>:**

***Column T “Date Ended Conversion” has been created.***

***Formula Used to derive the Date Ended Conversion column =***

***(((deadline column /60)/60)/24)+DATE(1970,1,1)***

***The Column has been formatted in Date format.***

## **Launch Date**

\* 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`.

**<Sara Comment>:**

***Pivot table has been created in a new sheet “Launch Date Outcomes Pivot” having campaigns outcomes based on the Date Created Conversion.***

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

**<Sara Comment>:**

***Line Graph Pivot Chart has been created by choosing the Line with Markers pivot chart type in a sheet “Launch Date Outcomes Pivot”. This Pivot Chart shows the campaigns outcomes based on the Date Created Conversion which can be filtered by parent category and Years.***

\* 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?

# **Report**

## **Conclusion**

### ***Outcome Based on Category***

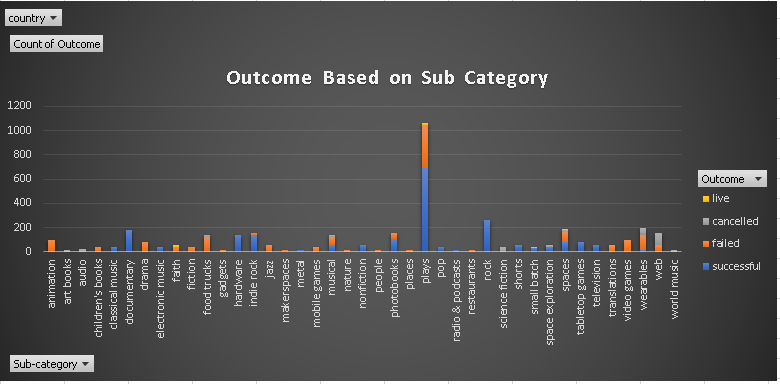
The First Graph (***Outcome Based on Category***) describes about the relationship between the project category and its outcome to determine which category has the highest and least Successful, Failed, Cancelled and Live rates.

From the Category Stats Pivot and Bar Graph, we can conclude that

* The **Music** Category (77%) had the highest success rates followed by theatre (60%) and Film & Video (58%).
* The **Food** (70%), **Games** (64%) and **publishing** (54%) had the highest failed rates (Lowest success rates).
* The **Journalism** Category based projects gets cancelled 100% and the technology related projects get cancelled of about 30%.
* The chances of projects in Kickstarter going live is very less.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category Name** | **% Successful** | **% Failed** | **% Cancelled** | **% Live** |
| film & video | 58% | 35% | 8% | 0% |
| food | 17% | 70% | 10% | 3% |
| games | 36% | 64% | 0% | 0% |
| journalism | 0% | 0% | 100% | 0% |
| music | 77% | 17% | 3% | 3% |
| photography | 47% | 53% | 0% | 0% |
| publishing | 34% | 54% | 13% | 0% |
| technology | 35% | 36% | 30% | 0% |
| theatre | 60% | 35% | 3% | 2% |

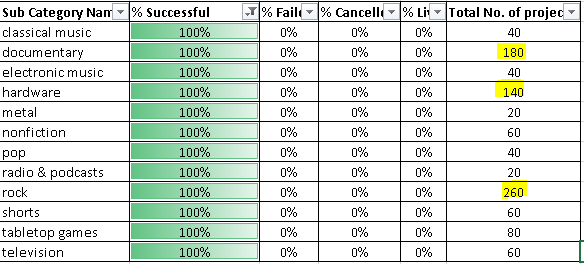
### ***Outcome Based on Sub Category***



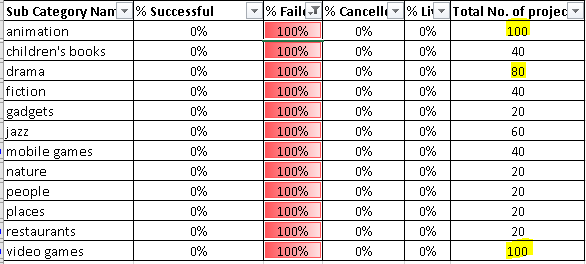
The Second Graph (***Outcome Based on Sub Category***) describes about the relationship between the project sub category and its outcome to determine which sub category has the highest and least Successful, Failed, Cancelled and Live rates.

From the Sub Category Stats Pivot and Graph, we can conclude that

* The below Sub Categories had the 100% Success rates, but **Rock** had the highest number of projects (260) followed by Documentary (180) and Hardware (140).



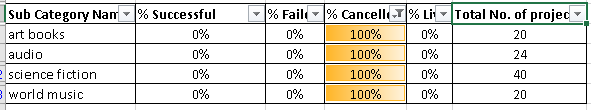
* The below Sub Categories had the 100% Failure rates, but Animation, Video Games and Drama had the highest number of projects.



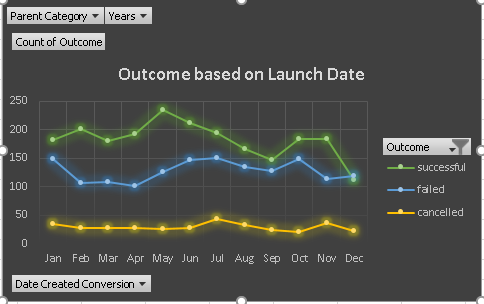
* The Sub Category - **Plays** had the highest number of projects (1066) but the success rate was 65% and the failure rate was 33%.



* The below Sub Categories projects gets cancelled 100%.



### ***Outcome Based on Launched Date***

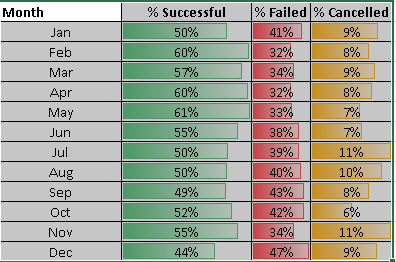


The Third Graph (***Outcome Based on Launch Date***) describes about the relationship between the year period (Month) in which the project was launched and its Outcome.

From the Launch Date Pivot and Line Graph, we can conclude that

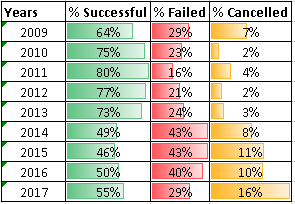
**Month Wise: (Considering all the Years)**

* **May Month** (61%) had the highest number of successful outcomes followed by Feb (60%), April (60%) Month.
* **December Month** (44%) had the highest number of failure outcomes followed by Sept (43%)
* July & Nov Month (11%) had the highest number of cancelled outcomes.

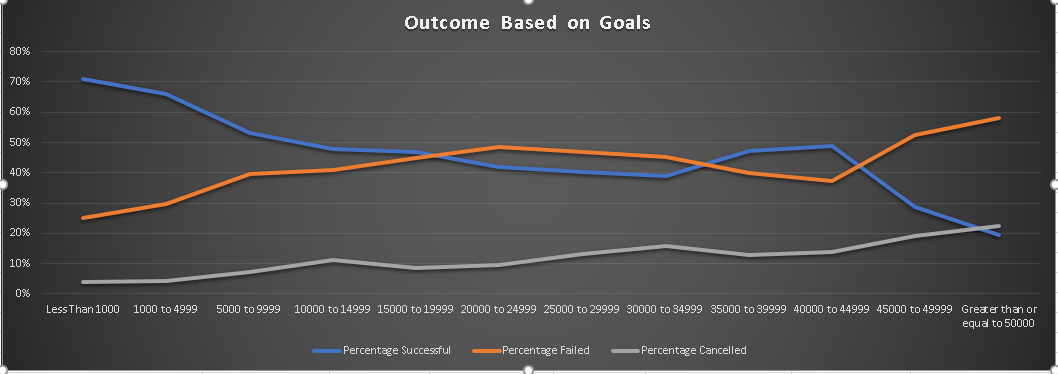


**Year Wise:**

* The projects launched in the first half of the year had the highest success rates compared to the latter half.
* The projects getting cancelled in the latter half years is more.



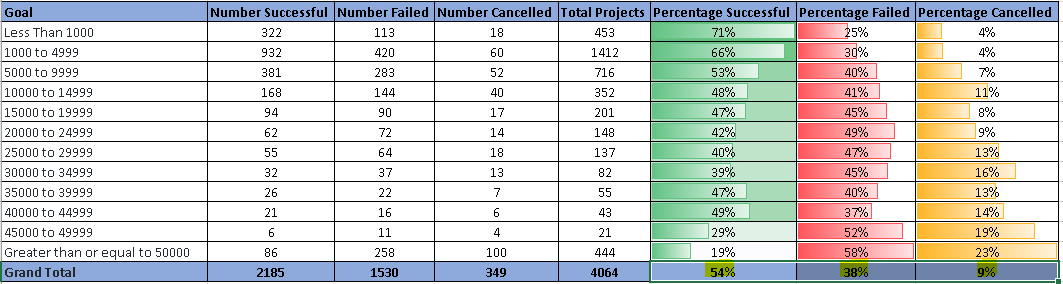
### ***Outcome Based on Goals***



The Fourth Graph (***Outcome Based on Goals***) describes the outcomes based on the project goals.

From the Outcome Based on Goals Data and Line Graph, we can infer that

* The Kickstarter projects are having more successful rates (54%) than being failed (38%) or getting cancelled (9%).
* Goals less than **1000 Range** had the highest successful rates (71%) followed by 1000-4999 Range (66%) and 5000-9999 Range (53%).
* Goals in **>=50000 Range** had the highest Failed Rates (58%) and also the highest rate in getting cancelled (23%).
* Goals from the Range 1000 to 44999 seems to be an ideal goal range since the Success Rate appears to be consistency.



### **Final Conclusion**

Analysing the 4000 data sets of Kickstarter projects, we cannot exactly predict the accurate outcomes based on the Category, Sub-Category, Launched Date and Goals. May be assuming the subset of the entire data, we can make some legitimately assertions about the successful outcomes. The Category **Music** had the highest success outcomes projects. The sub category **rock** had the maximum success rates with the most successful projects compared to the other sub categories. All the projects under Rock sub category achieved the accomplishment. **May** is considered to be the best month to launch the projects since had the highest number of successful outcomes. Finally, the Kickstarter data concludes that projects with **smaller goals** had the most successful outcomes and also Goals from the Range 1000 to 44999 seems to be an ideal goal range since the Success Rate appears to be consistency and optimal.

## **Limitations**

* The **amount of data** provided is very limited and only 1% of data is provided of all the Kickstarter project. Hence it is difficult to perform an accurate analysis of the outcome of the projects.
* There might be more number of **categories and sub-categories** which is not incorporated in our analysis.
* Proper Information about **backers** is not provided and having the count alone we won’t able to judge the outcomes.
* The data can be analysed more if they would have given **state wise** projects in order to see in which state has the highest project outcomes and based on the that we can distribute the projects accordingly.
* **Difficulty levels** of the projects should have been considered.

## **Other Possibilities**

* May be we could have considered the **time duration** between a launch date and its end date for the outcome predictions. We could have created a graph between the duration of a project and its outcome to analyse if the durability affects in any way for the success rates.
* We could have considered the **country** data to determine the outcomes of success and failure rates with respect to country. More accurate I would say if with the states name provided the corresponding state data is given.
* And also the other data like staff picked, average donations or blurb information would have yield more accurate analysis.

# **Bonus**

\* Create a new sheet with 8 columns:

\* `Goal`

\* `Number Successful`

\* `Number Failed`

\* `Number Cancelled`

\* `Total Projects`

\* `Percentage Successful`

\* `Percentage Failed`

\* `Percentage Cancelled`

\* 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

**<Sara Comment>:**

***The above columns and Row with headers has been created in the Sheet “Bonus”.***

## **Goal Outcomes**

\* Using the `COUNTIFS()` formula, count how many successful, failed, and Cancelled projects were created with goals within the ranges listed above. Populate the `Number Successful`, `Number Failed`, and `Number Cancelled` columns with this data.

**<Sara Comment>:**

***The Successful, Failed and Cancelled projects were created within the goal ranges using the below formula:***

***Successful: COUNTIFS('Kickstart Data'!D2:D4115,"<1000",'Kickstart Data'!F2:F4115,"successful")***

***Failed: COUNTIFS('Kickstart Data'!D2:D4115,"<1000",'Kickstart Data'!F2:F4115,"failed")***

***Cancelled: COUNTIFS('Kickstart Data'!D2:D4115,"<1000",'Kickstart Data'!F2:F4115,"cancelled")***

***Note:***

***D:D Range is the Goal Column***

***F:F Range is the State Column***

***Similarly, for the other ranges the condition in the above formula changes.***

\* Add up each of the values in the `Number Successful`, `Number Failed`, and `Number Cancelled` columns to populate the `Total Projects` column. Then, using a mathematical formula, find the percentage of projects that were successful, failed, or Cancelled per goal range.

**<Sara Comment>:**

***The Total Projects and the Percentage of Projects were derived using the below formula:***

***Total Projects = Sum(B3:D3)***

***B3:D3 ->*** `Number Successful`, `Number Failed`, and `Number Cancelled`

**Percentage Successful = Number Successful/Total Projects**

**Percentage Failed = Number Failed/Total Projects**

**Percentage Cancelled = Number Cancelled/Total Projects**

***The format has been changed to Percentage Style.***

\* Create a line chart that graphs the relationship between a goal's amount and its chances at success, failure, or cancellation.

**<Sara Comment>:**

***A line chart has been created in the Sheet “Bonus” which shows the relationship between a goals amount and its Outcome percentage.***

# **Submission**

\* To submit your homework, upload the solution and files to a GitHub repo, Dropbox, or Google Drive and submit the link to <https://bootcampspot.com/>.

The link to the Excel Assignment is in the below link.

< <https://drive.google.com/drive/folders/1dlOSMryRfkCaF_5F1MGj2FGpAOYEVWRN>>