**Kickstarter Campaign Analysis and Visualization Report**

**Author:** Leonard Paul-Kamara

**Introduction**

This project analyzes kickstarter campaigns hosted on the Kickstarter website, [www.kickstarter.com](http://www.kickstarter.com). The data included in this analysis are of campaigns from various countries across the globe, spanning between 5/17/2009 to 3/15/2017.

The kickstarter website describes itself as a platform that “help bring creative projects to life”. Launched on 4/28/2009, it boasts of having successfully funded over 186,621 projects with over $5B pledged by 18M people. It is now a for-profit company since 2015, when they transformed into a Public Benefit Corporation.

Project ideas are launched, hosted and made public via the website where a goal or target is set. A kickstarter project will be labelled successful when pledged amounts meet the minimum requirement, i.e., the goal.

**Method**

**Data**

The data used in this analysis is in a spreadsheet containing 4114 records of kickstarter campaigns. The headers and their respective columns are seen in the table below:

*Table (i) Headers and their respective columns*

|  |  |
| --- | --- |
| **Column** | **Header** |
| A | id |
| B | Name |
| C | blurp |
| D | Goal |
| E | Pledge |
| F | state |
| G | country |
| H | currency |
| I | deadline |
| J | Launched\_at |
| K | Staff\_pick |
| L | Backers count |
| M | spotlight |
| N | Category and subcategory |

**ANALYSIS**

To analyze the data, tasks were broken down into the following:

1. Highlight the state column with a different color, depending on whether an associated campaign was successful, failed, canceled or is live during the time interval covered. To accomplish this task, conditional formatting was used to distinguish the various states from each other.

*Table (ii) Color code of categories in state column*

|  |  |
| --- | --- |
| **State** | **Color** |
| Successful | Green |
| Failed | Red |
| Cancelled | Orange |
| Live | Blue |

1. Calculate Percent Funded in column O.

This was calculated by the following expression:

|  |  |
| --- | --- |
| Percent Funded = (Pledge / goal) \* 100 | *equation (i)* |

1. Highlight Percent Funded with a 3-color scale. This was done using conditional formatting.

*Table (iii) Color scale used in Percent Funded column*

|  |  |  |  |
| --- | --- | --- | --- |
| Range | Minimum | Middle | Maximum |
| Value | 0 | 100 | 200 |
| Color | Bark Red | Green | Blue |

1. Calculate Average Donation to identify how much each backer paid on average. It was calculated as seen below:

|  |  |
| --- | --- |
| Average Donation = Pledge / backers count | *equation (ii)* |

1. Create two new columns, Category at Q and Sub-Category at R. This was achieved by using excel formulas to split the Category and Subcategory column (N). The formulas are shown below:

*Table (iv) Expressions used to split Category and sub-category*

|  |  |  |
| --- | --- | --- |
| **Operation** | **Expression** |  |
| Select Category | LEFT(N2,FIND("/",N2)-1) | *equation (iii)* |
| Select Sub-Category | RIGHT(N2,LEN(N2)-FIND("/",N2)) | *equation (iv)* |

1. Create a pivot table to count number of campaigns in the different categories of State. This was done in a new sheet labelled Pivot chart Category. A stacked pivot chart was then created from the table. Country filter is included in the sheet.

|  |
| --- |
|  |
| *Figure (i) Stacked Pivot Chart of Categories and their respective state* |

1. Create a pivot table to count number of campaigns in the different sub-categories of State. This was done in a new sheet labelled Pivot chart Sub-Category. A stacked pivot chart was then created from the table. Country and Category are included in the sheet.

|  |
| --- |
|  |
| *Figure (ii) Stacked Pivot Chart of Sub-Categories and their respective state* |

1. Create two new columns named ‘Date Created Conversion’ and Date Ended Conversion. The columns contain date values converted from the launched\_at and deadline columns, respectively. The following link shows steps taken to convert the Unix timestamps to Excel Dates:

|  |  |
| --- | --- |
| Link: | <https://www.extendoffice.com/documents/excel/2473-excel-timestamp-to-date.html> |

1. Create 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`. This was done on a new sheet labelled “launched\_date\_outcome”. A pivot line chart was also created to visualize Count of State for the period represented in the data.

|  |
| --- |
|  |
| *Figure (iii) Count of State across periods in data* |

**Conclusions**

The following conclusions are made based on the data and its subsequent analysis:

1. The dates analyzed range from 5/17/2009 to 3/15/2017, which is about 8 years.
2. The most successful state is theater category with 839 campaigns while publishing and games are the least successful with 80 campaigns each. On the other hand, theatre tops the number of failed state with 493 campaigns.
3. Worth noting is the canceled state of technology category with a total of 178 canceled campaigns. It has the highest cancelation per total campaign. This might indicate how difficult it can be for a tech-startup campaign to survive beyond the launch date.

**Limitations of Dataset**

The following limitations were observed in this dataset:

1. The data is clearly a subset/sample of the total number of campaigns hosted by Kickstarter. Any deduction made from the analysis might not represent campaigns not included in this dataset. In effect, the above analysis is limited to just what is available in the sheet analyzed.
2. Detailed data cleaning and wrangling is needed to effectively perform any form of data analysis. One example of manipulation will be to standardize the currency column by converting all currencies to Dollars.

**Recommended Charts**

1. A scatter plot to observe relationship between goal and pledged amounts.
2. A scatter plot to observe correlation between goal and backer’s count.
3. A pie chart to observe proportional representations of all the categories in “state” based on percent funded.
4. A box plot to reveal spread of the data, outliers and other central tendencies.
5. A histogram to observe the distribution of pledges.

**Bonus Statistical Analysis**

A sheet labelled “Bonus Stats Analysis” was created comprising of summary statistics of successful campaign backers and unsuccessful campaign backers. A copy of the table is seen below:

*Table (v) Summary Statistics of Successful and Unsuccessful Campaigns backers*

|  |
| --- |
|  |
| |  |  |  | | --- | --- | --- | | **Backer's** | **Successful Campaigns** | **Unsuccessful Campaigns** | | **Count** | 424819 | 38427 | |  |  |  | | **Statistics** | **Values** | | | **Mean** | 194.0 | 20.0 | | **Median** | 62.0 | 3.0 | | **Minimum** | 1.0 | 0.0 | | **Maximum** | 26457.0 | 1501.0 | | **Variance** | 712841.0 | 5512.2 | | **Standard Deviation** | 844.3 | 74.2 | |

As observed above, the median is the most appropriate statistics to summarize the data more meaningfully. The median is more of a robust statistic than the mean. Unlike the mean, the median is not influenced by the spread of the data which sometimes contains outliers.

In our case, a histogram chart of both groups will be skewed as there is great difference between the mean and median. Hence, the mean would not be the ideal statistics as it changes with spread while the median remains robust and unchanged.

Also, it can be observed that there is more variability in the “successful Campaigns” category. Assuming the mean is a measure of central tendency, variability measured by the variance is greater in the mentioned category than in “Unsuccessful Campaigns”.