**Excel-Challenge Homework**

**Introduction**

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.

**Preliminary Analysis**

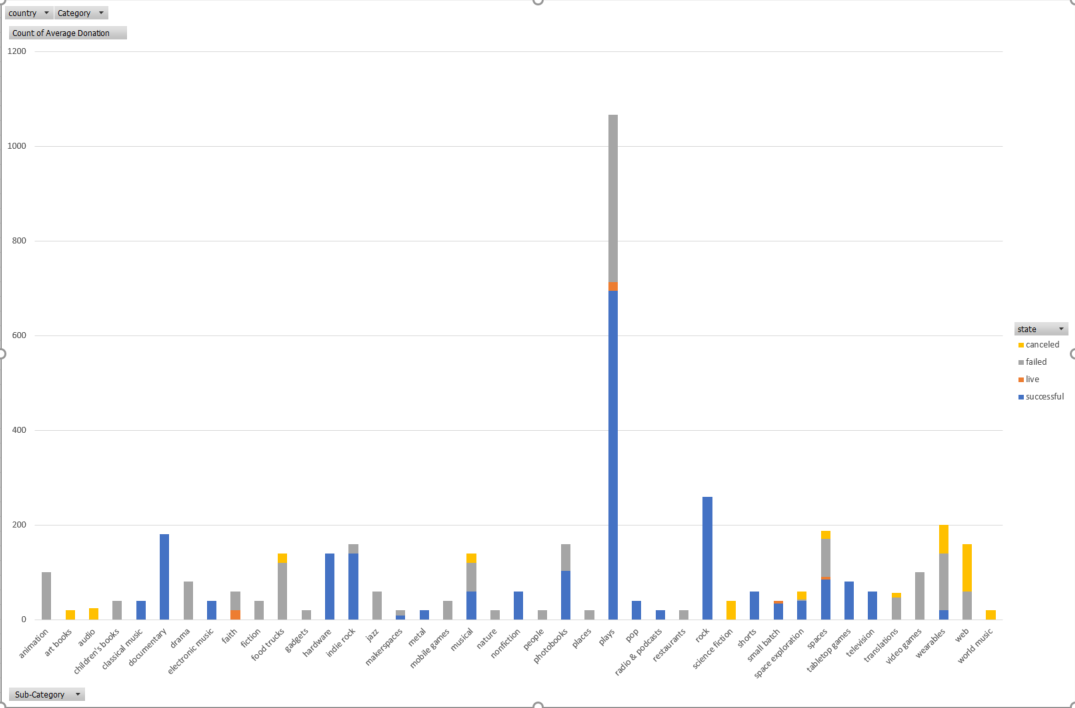
We were provided with 4,000 past Kickstarter projects as an attempt to analyze the market trends. In this table additional features to include, “Percent Funded” and “Average Donation”. And two separate columns were created to show the “Category” and the “Sub-Category”, to summarize and to prepare the data for further analysis.

**Data Analysis:**

1.To begin with, after scrutinizing the relationship between the Parent category and its sub-category, it is easy to determine which category had the highest success rate. We can draw a conclusion that major chunk of Kickstarter projects is towards the entertainment industry as opposed to writing, food, or other industries. From the Chart 1 below, the category that had the highest success was Theater, followed by music and Film & Video respectively.

Chart 1 Stacked Column Chart representing the different outcomes by category

2.Secondly after analyzing how these projects break down by sub category, we can see that Rock(Music), Drama (Film& Video), Hardware(Technology) had a 100% success. And also when compared to the others, Indie Rock(Music) has more success rates and less failures. Chart2, below shows the chart of sub category.



*Chart 2, Stacked Column Chart representing the different outcomes by Sub-category*

3.Thirdly, based on the analysis of what time of the year (month) in which the project was conducted, it very well shows the projects held during the beginning of the year are more successful than the one conducted in the later part of the year. Projects held during the month of May are more successful than the ones in December.

*Chart 3, Line chart of the outcome by month*

**Limitations of the dataset**

The dataset used only represented about 1% of the total number of Kickstarter campaigns conducted, and may not accurately represent the data of all Kickstarter campaigns.

If we had more data on the geographic location (Country, State) of the projects being conducted we can do more analysis based on the same, as many Kickstarter filters who can launch what kind of project based on the geographic limitations.

Lack of information about the Backers also restricts us to do more analysis on which type of Backers and also what genre of Kickstarter campaigns fascinates them to donate more.

**Other possible table/graphs:**

We can a create a table on most successful categories/sub categories vs the most unsuccessful categories.

Table on the country and average donation amount.

Average amount of money raised by successful projects.

How hard do unsuccessful projects fail.

Average goal amount sought by successful and unsuccessful campaigns.

goal vs pledged amount to make an analysis on how each project made it to be a successful one.

Goal vs Status

Average project duration

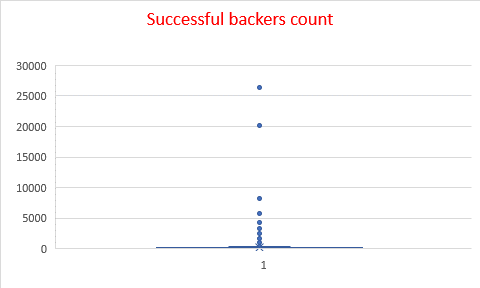
**Bonus Mean and Median:**

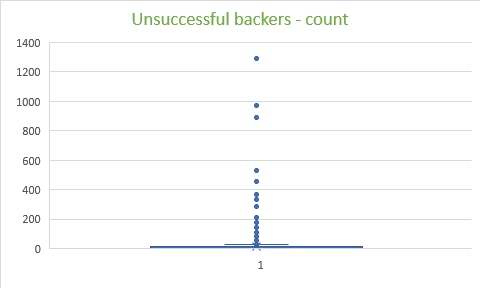
|  |  |  |
| --- | --- | --- |
| **Column1** | **Successful Campaings** | **Unsuccessful Campaigns** |
| **Mean number of backers** | 194.43 | 17.71 |
| **Median number of backers** | 62 | 4 |
| **Minimum number of backers** | 1 | 0 |
| **Maximum number of backers** | 26457 | 1293 |
| **Variance** | 712840.99 | 3773.22 |
| **Standard Deviation** | 844.30 | 61.43 |

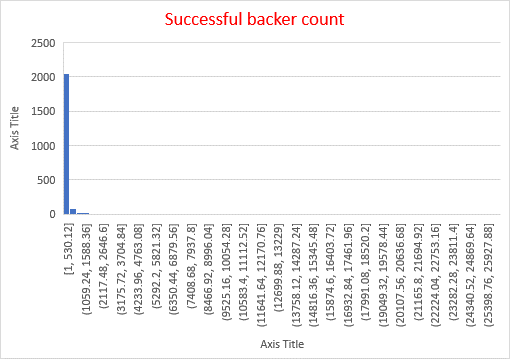
|  |  |  |
| --- | --- | --- |
| **Summary by Mean** | **Successful Campaigns** | **Unsuccessful Campaigns** |
| Mean +1SD | 1038.72 | 79.14 |
| Mean -1sd | -649.87 | -43.72 |

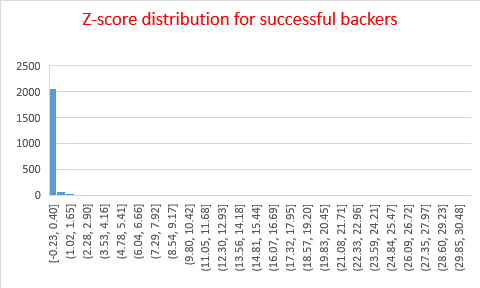
|  |  |  |
| --- | --- | --- |
| **Statistics** | **Successful** | **Unsuccessful** |
| **First Quartile:** | **29** | **1** |
| **Median:** | **62** | **4** |
| **Second Quartile:** | **62** | **4** |
| **Third Quartile:** | **141** | **141** |
| **Interquartile Range:** | **112** | **140** |
|  |  |  |
|  |  |  |
|  |  |  |
| **Lower Boundary** | **-139** | **-209** |
| **Upper Boundary** | **309** | **351** |

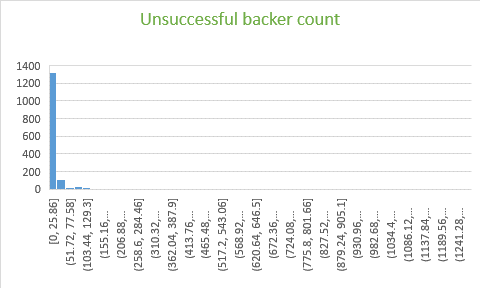
I would like to conclude that **Median** summarizes the data more meaningfully in both the successful and unsuccessful campaigns.

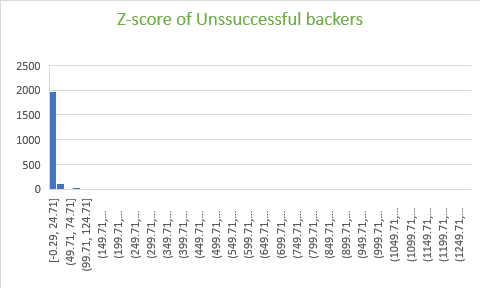












Based on the calculations of mean, z-score and the outlier results of both successful and unsuccessful campaigns, it shows that unsuccessful campaigns have more variability than unsuccessful campaigns.