**Juan Camarillo**

**Module 1 Challenge Kickstarter Campaign**

**Overview**

The intent of this data analysis is to examine the results of our client’s Kickstarter project for the play *Fever*. The goal is to provide insight on the success or failure of the potential project based on the multiple factors provided in the data set.

**Background**

The dataset provided spans 8 years (2010-2017) and includes 4225 projects across 9 categories and 41 subcategories. I focused on the Kickstarter campaign currently being launched for the play *Fever*, and so the analysis concentrates on the theatre category and play subcategory.

**Analysis**

The analysis was performed in two phases, based on launch dates and funding goals respectively.

**Phase 1: Launch Dates**

Using the data provided I analyzed the time of year that campaigns were historically launched and compared their success and failure rates, determine the best time to launch the clients Kickstarter campaign. The dataset was filtered by the Theater category and arranged by outcomes, identifying successful, failed, and cancelled projects.

Chart, line chart

Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parent Category | theater |  |  |  |
| Years | (All) |  |  |  |
|  |  |  |  |  |
| **Count of outcomes** | **Column Labels** |  |  |  |
| **Row Labels** | **canceled** | **failed** | **successful** | **Grand Total** |
| Jan | 7 | 33 | 56 | 96 |
| Feb | 3 | 39 | 71 | 113 |
| Mar | 3 | 33 | 56 | 92 |
| Apr | 2 | 40 | 71 | 113 |
| May | 3 | 52 | 111 | 166 |
| Jun | 4 | 49 | 100 | 153 |
| Jul | 1 | 50 | 87 | 138 |
| Aug | 4 | 47 | 72 | 123 |
| Sep | 4 | 34 | 59 | 97 |
| Oct |  | 50 | 65 | 115 |
| Nov | 3 | 31 | 54 | 88 |
| Dec | 3 | 35 | 37 | 75 |
| **Grand Total** | **37** | **493** | **839** | **1369** |

As noted in the graph, the results indicate that May, June, and July had the highest total project launches and the highest number of successful projects. The months with the lowest chance of success are November and December, indicating that the end of the year and the Holiday season may play a role in how much support a project will see.

**Challenge:**

The biggest challenge faced during the analysis of this dataset revolved around the dates associated with each project. *Deadline* and *Launched\_At* dates are formatted as Unix timestamps, which required conversion using the formula:=(((CELL/60)/60)/24)+DATE(1970,1,1)

**Phase 2: Funding Goals**

The dataset for funding goals was analyzed by investigating the initial requested amount for each project and comparing the success or failure within predetermined ranges. Using the number of results against each outcome and dividing by the total projects for that range, I was able to calculate the success and failure percentage rate for comparison.

Table

Description automatically generated

Chart, line chart

Description automatically generated

The smaller the requested amount, the higher the probability of success. There are two exceptions to that, with the 35,000 range and the 40,000 range both showing a 67% success rate, however, with 6 and 3 projects in each range respectively, these numbers are outliers within the dataset and are less reliable.

**Challenge:**

There were three different formula setups that needed to be used, depending on the cost range. All of the formulas limit the count for the results by focusing only on the “plays” category and then the related outcome:

*=COUNTIFS(Kickstarter!R:R,"plays",Kickstarter!F:F,"successful")*

However, an additional set of modifiers need to be added for the range. For the “less than 1000”, the formula was relatively simple:

*=COUNTIFS(Kickstarter!R:R,"plays",Kickstarter!F:F,"successful",Kickstarter!D:D,"<1000")*

Other ranges (excluding the 50,000 and above range) require two amount modifies, one for the minimum amount of the range and one for the maximum:

*=COUNTIFS(Kickstarter!R:R,"plays",Kickstarter!F:F,"successful",Kickstarter!D:D,">=1000",Kickstarter!D:D,"<=4999")*

The final range set requests that the field include amounts “greater than 50,000”, however this request is not inclusive of the 50,000 amount itself and would miss several datapoints, the formula is modified to read “50,000 or above”:

*=COUNTIFS(Kickstarter!R:R,"plays",Kickstarter!F:F,"successful",Kickstarter!D:D,">=50000")*

**Conclusions**

***What are two conclusions you can draw about the Theater Outcomes by Launch Date?***

The results indicate that May, June, and July had the highest total project launches and the highest number of successes. The months with the lowest chance of success are November and December, indicating that the end of the year, and the Holiday season may play a role in how much support a project will see.

**What can you conclude about the Outcomes based on Goals?**

The smaller the requested amount, the higher the probability of success. There are two exceptions to that, with the 35,000 range and the 40,000 range both showing a 67% success rate, however, with only 6 and 3 projects in each range respectively, these numbers are outliers within the dataset.

**What are some limitations of this dataset?**

The small number of data points in some areas (such as the outcomes based on goal ranges), which yielded a sampling that is not large enough to reliably draw conclusions from compared to the other data points.

Secondly, additional data categories could be included that would expand the understanding of how pledges are being made, such as the different backer tiers for each project and the number of pledges per tier.

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

There are several different ways the information could be analyzed, including looking at the length of time a project is active, the impact of spotlight and staff picks on outcomes.