**Overview and Purpose of the Project**

The project aims to illuminate relevant data trends by analyzing and organizing a crowdfunding database comprising 1,000 campaigns launched at different months within a given year concerning funding goals and the number of backers or donors.

The project will consist of four visuals titled **Outcomes based on launch date, Outcomes based on category, Outcomes based on subcategory, and Outcomes based on initial goal**. Descriptive statistics of the successful campaigns will also be explored.

**Analysis**

Four analyses with corresponding visuals are performed as described below. **Outcomes based on launch date** is constructed by creating a pivot table from the raw dataset Crowdfunding and by graphing a line chart to see the correlation between campaign outcomes and launch dates. Date defined by months is plotted on the x-axis, and the count of total successful, failed, and canceled outcomes are plotted on the y-axis. **Outcomes based on funding goals** obtained by pooling raw data and computing the percentage of successful, failed, and canceled projects on the Y-axis and the range of funding goals on the X-axis by a factor of $5,000. Similarity, **Outcomes based on category**, and **Outcomes based on subcategory**, are analyzed by creating a stacked-column pivot chart from the Crowdfunding dataset. **The results are below.**

**Results**

Based on the analysis of **Outcomes based on launch Date**, July appears to feature the most successful campaigns, while January and August feature the most failed campaigns. One should consider launching a campaign between the months of May-July as the trend of successful campaigns increases within those months.

Based on the analysis of **Outcomes based on goals**, campaigns that aimed to raise between 15K-25K were the most successful (100%), while campaigns that aimed to raise between 1K – 5K were 83% successful. Simultaneously, campaigns that aimed to raise between 10K – 15K were the least successful.

Based on the analysis of **Outcomes based on category and Outcomes based on subcategory, the technology category with a subcategory of web appears to be the most successful, with a ratio of 2:1 between successful (including live) to failed (including canceled) technology campaigns, and a ratio of almost 3:1 between successful and failed web subcategory campaigns. In contrast, food campaigns are the least successful, where more than half of the launched campaigns failed with a ratio of 0.91.**

**Limitations**

Data does not measure or report the staffers hired nor the process behind securing pledges to fund campaigns. Understanding such external factors can help control mediating factors using statistical computations, minimizing the chances of a Type I error.

**Other Possibilities**

Possible graphs that can be helpful are the number of backers based on outcomes, the number of backers based on pledges funds/goals, and outcomes based on country, coupled with outcomes based on category and subcategory. These graphs or analyses can provide further insight into how campaigns perform within each country and the ideal number of backers needed to reach the initial goal within a given time frame.

**Summary statistics**

A descriptive summary of the number of successful campaign backers vs the number of failed campaign backers can be explored for further quantitative metrics.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Outcome** | **Mean** | **Median** | **Min** | **Max** | **Std.Dev** | **Variance** | **Skewness** | **Kurtosis** |
| **Successful** | 851.15 | 201 | 16 | 7295 | 1267.36 | 1606216.6 | 2.18 | 4.97 |
| **Failed** | 585.62 | 114.5 | 0 | 6080 | 961.31 | 924113.46 | 2.70 | 8.80 |

***Table 1****. Descriptive statistics of the number of successful and failed campaign backers*

The Skewness and Kurtosis values of both outcomes [2.18 Skew/ 4.97 Kurt, **successful**: 2.70 Skew/8.80 Kurt, **failed**] indicates that the dataset is not normally distributed but is rather positively skewed with high Kurtosis or a “pointy” curve. Therefore, the **median** would better summarize the data since it is less affected by outliers and skewed data.

The median values of successful and failed outcome backers are 201 and 114.5, respectively.

There is more variability with successful campaigns because the variance (**σ2 )** of successful outcome backers is greater than the variance of failed outcome backers [ 1606216.6 **>** 924113.46 ] because the number of backers that funded successful campaigns varied greatly as opposed to backers that funded failed campaigns.