**Module 1 Challenge**

**Data Analysis report written by God’swill Anyasor**

**Title:** Analysis of Factors Influencing Crowdfunding Reception Trends

**Executive Summary:** This data analysis report aims evaluate crowdfunding reception trends by analysing a dataset of 1000 sample crowdfunding projects. The primary objective is to uncover patterns and trends within this dataset to gain a deeper understanding of the dynamics of crowdfunding reception.

**Introduction**

Crowdfunding has become a popular means of funding creative projects and businesses in their initial stage to avoid resort to bank financing sources (Mora-Cruz & Palos-Sanchez, 2023). To better understand what factors and trends influences crowdfunding reception, an analysis of a dataset containing 1000 sample crowdfunding projects was performed.

**Data Source**

The dataset used for this analysis consists of 1000 sample crowdfunding projects excel file and can be downloaded as a .zip file from <https://static.bc-edx.com/data/dl-1-2/m1/lms/starter/Starter_Code.zip>. The data is sourced from various crowdfunding campaigns and represents a diverse range of projects as prepared by edX Boot Camps LLC. All codes were obtained from https://exceljet.net/formulas.

**Data Analysis Strategy**

The analysis strategy involved the following steps:

1. Creation of a project space called excel-challenge on Google Drive to facilitate data storage and sharing.
2. Downloaded the provided Excel database in .zip format.
3. Utilized the Excel workbook contained in the downloaded folder to analyse sample-project data and discover market trends.
4. Employed conditional formatting tools to identify trends and patterns.
5. Created pivot tables and pivot charts to visualize the data effectively.

**Results**

Data analysis yielded the following insights regarding crowdfunding campaign reception:

1. **Theater Category Success:** Data showed that the theater category, specifically the "plays" subcategory, had the highest number of successful campaigns, surpassing other categories. This indicates the relative success of theater projects in achieving funding goals (Figure 1 and 2; Pivot Table 1 and 2).
2. **Effective Campaign Months:** The analysis revealed that campaigns staged in specific months, namely January, March, June, and June, showed higher success rates. This suggests that campaign strategies should be strategically timed to these months for improved effectiveness.
3. **August Cancellations and Failures:** The dataset showed a notable spike in the number of canceled and failed projects in the month of August, with 8 cancellations and 35 failures. Understanding the reasons behind this August spike is crucial for enhancing project management and identifying potential issues specific to that month (Pivot Table 3).

**Figure 1: Crowdfunding project outcomes by parent category**

**Figure 2: Crowdfunding project outcomes by sub-category**

**Figure 3: Crowdfunding project outcomes by months**

**Limitations of the Dataset**

While the dataset is valuable, it has several limitations:

1. **Lack of Qualitative Data:** The dataset only provides quantitative data on project outcomes. Qualitative data, such as customer feedback, reasons for cancellation, or internal post-mortem reports, is essential for a deeper understanding of project performance.
2. **Ambiguity in Live Projects:** The live category in the dataset lacks information about the status and progress of these projects, making it unclear whether they are on track to success or facing challenges.
3. **Data Quality Concerns:** Without details on data collection and cleaning processes, there may concern about data accuracy and quality.
4. **Unexplained Categorical Variables:** Some variables, such as "staff\_pick" and "spotlight," are represented as Boolean values without clear explanations of their meanings.
5. **Currency Variability:** The dataset includes various currencies (CAD, USD, GBP, EUR, AUD, CHF, and DKK) with lack of currency conversion for financial analysis and comparisons.
6. **Geographical Bias:** The dataset includes projects from specific countries, potentially introducing geographical bias in the sample.

**Other Possible Tables and/or Graphs with Additional Values**

The following tables and graphs are also possible:

1. **Funding Goals and Outcomes:**

Table: Display the average funding goal for successful and failed projects, along with the median and standard deviation.

Graph: A box plot or a grouped bar chart can illustrate the differences in funding goals for different outcomes.

**Additional Value:** Understanding the relationship between funding goals and project outcomes can provide insights into setting realistic goals.

1. **Backer Count Distribution:**

Table: Present the distribution of the number of backers, including key statistics (e.g., mean, median, range).

Graph: A histogram or box plot can visualize the distribution.

**Additional Value:** This table/graph can help explore the typical backer count for projects and identify potential outliers.

1. **Country Funding Performance:**

Table: Show the number of projects, success rates, and average funding amounts for each country.

Graph: A world map or bar chart can illustrate the performance of crowdfunding projects in different countries.

**Additional Value:** This analysis can highlight the geographical patterns in crowdfunding success and identify potentially lucrative markets.

1. **Currency Conversion:**

Table: Create a table that converts all funding amounts to a common currency (e.g., USD) for easy cross-project comparisons.

Graph: A scatterplot or bar chart can visualize funding amounts in a standardized currency.

**Additional Value:** This allows for fair comparisons between projects with different original currencies.

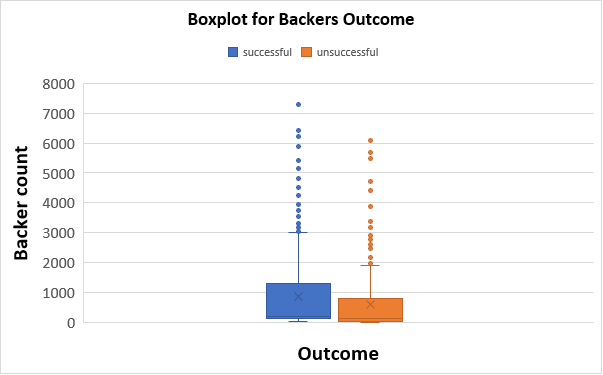
By considering these additional tables and graphs, we can gain a more comprehensive understanding of crowdfunding reception trends and facilitate data-driven decisions.

**Statistical analysis**

The new created worksheet title “Number of backers\_outcome” shows the columns for the number of backers of successful campaigns and unsuccessful campaigns (see CrowdfundingBook.xlsx).

**Table 1: Summary statistics of the number of backers and outcome of crowdfunding campaigns**

|  |  |  |
| --- | --- | --- |
| **Summary Statistics** | | |
| **Statistics** | **Successful campaigns** | **Unsuccessful campaigns** |
| Total backers | 480898 | 213164 |
| Mean number of backers | 851.15 | 585.62 |
| Median number of backers | 201 | 114.5 |
| Minimum number of backers | 16 | 0 |
| Maximum number of backers | 7295 | 6080 |
| Variance of the number of backers | 1603373.73 | 921574.68 |
| Standard deviation of the number of backers | 1266.24 | 959.99 |



**Figure 4: Comparative boxplot crowdfunding backers’ outcome**

* **Determine whether the mean or median better summarizes the data.**

To determine whether the mean or median better summarizes the data, the Table 1 above was evaluated as follows:

1. For successful campaigns, the mean number of backers is 851.15, and the median number of backers is 201.
2. For unsuccessful campaigns, the mean number of backers is 585.62, and the median number of backers is 114.5.
3. In both successful and unsuccessful campaigns, the mean is higher than the median. This suggests that the data is right-skewed, with some projects having many backers, which drives up the mean (also see Figure 4). As a measure of central tendency, the median is less affected by outliers (Mccluskey et al., 2007). Therefore, the median better summarizes the data for both successful and unsuccessful campaigns.

* **Determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

To determine if there more variability with successful or unsuccessful campaigns using the Table above, it was evaluated as follows:

1. Variance measures how data points vary from the mean (Mccluskey et al., 2007). Successful campaigns have a higher variance (1,603,373.73) compared to unsuccessful campaigns (921,574.68). This means that the number of backers for successful campaigns is more spread out, indicating greater variability.
2. The standard deviation provides a measure of how spread out data points are around the mean (Mccluskey et al., 2007). Successful campaigns have a higher standard deviation (1,266.24) compared to unsuccessful campaigns (959.99). This means there is more variability in the number of backers for successful campaigns.

* **Does this make sense?** Yes, it makes sense.
* **Why or why not?**

Successful campaigns often have a wider range of backers because some projects can attract a high number of backers, potentially due to factors like popularity, marketing, or unique high product offerings. Unsuccessful campaigns, on the other hand, may have less variability because they did not attract as many backers perhaps due to lower-quality projects (Mollick, 2013).

**REFERENCES**

Exceljet (2023). 500 Excel formulas. <https://exceljet.net/formulas>. Accessed on October 17, 2023.

Mccluskey, A., Mb, B., Frca, C., Ghaaliq, A., & Mb, L. (2007). *Statistics II: Central tendency and spread of data*. https://doi.org/10.1093/bjaceaccp/mkm020.

Mollick, E. (2013). *The dynamics of crowdfunding: An exploratory study*. *Journal of Business Venturing* 29(2014) 1-16. https://doi.org/10.1016/j.jbusvent.2013.06.005

Mora-Cruz, A., & Palos-Sanchez, P. R. (2023). Crowdfunding platforms: a systematic literature review and a bibliometric analysis. *International Entrepreneurship and Management Journal*, *19*, 1257–1288. https://doi.org/10.1007/s11365-023-00856-3