Module 1 Challenge

Will Davis | [william.c.davis@nasa.gov](mailto:william.c.davis@nasa.gov)

-----

**Conditional Formatting (10 points)**

* Conditional formatting is applied appropriately to the outcome column (5 points)
* Conditional formatting is applied appropriately to the percent funded column (5 points)

See excel file in submittal.

**Column Creation (10 points)**

* Six new columns were correctly created for:
  + percent funded
  + average donation
  + category
  + sub-category
  + Date Created Conversion
  + Date Ended Conversion

See excel file in submittal.

**Pivot Tables and Stacked Column Charts (15 points)**

* Correctly created a pivot table that counts how many campaigns were "successful," "failed," "canceled," or are currently "live" per category (7.5 points)
* Correctly created a stacked column pivot chart that can be filtered by country (7.5 points)

See excel file in submittal.

**Pivot Tables and Line Graphs (15 points)**

* Correctly created a pivot table with a column of outcome, rows of Date Created Conversion, values based on the count of outcome, and filters based on parent category and Years (7.5 points)
* Correctly created a pivot chart line graph (7.5 points)

See excel file in submittal.

**Written Report (20 points)**

* Presents a cohesive written analysis that:
  + Draws three conclusions from the data (10 points)
  + States limitations of the dataset and suggestions for additional tables of graph (10 points)

Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

* Conclusion 1: For all categories, the typical success rate is usually in the 50-60% range. The only exception to this is categories with very low counts (e.g., journalism = 4).
* Conclusion 2: For all subcategories, much like the observation for categories, the typical success rate is usually in the 50-60% range. The only exception to this is subcategories with very low counts (e.g., audio = 4, world music = 3).
* Conclusion 3: Campaigns occur year-round, and typically less than 10% are cancelled.

What are some limitations of this dataset?

* Data includes campaigns with large range of goals.
* Data is sourced from various countries, each with different currencies that don’t exchange 1 to 1. This may be mixing data in a way that doesn’t lead to clear observations. Examples:
  + For USD, min = $100, max = $199,000
  + For EUR, min = €100, max = €196700

What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

* Option 1: Build a table which compares number of days campaign was active with duration of campaign; look at trends for each outcome type (e.g., successful, failed, cancelled).
  + Why? May indicate the optimal duration for keeping a campaign open. Helps determine when to open and close campaign.
* Option 2: Build a table to view outcome by country.
  + Why: May indicate if country where campaign occurred has influence on outcome rates. May show certain categories/subcategories are more/less successful in specific countries. Helps determine where to launch campaigns.
* Option 3: Build a table showing average donation by category/subcategory.
  + Why? May show a relationship between the type of activity and level of donation. Helps with approach for fundraising and how much to ask for.
* Option 4: Build a table showing goal by category/subcategory.
  + Why? May show there is a certain amount when campaigns are more/less likely to succeed. Helps with realistic goal setting.

**Crowdfunding Goal Analysis (10 points)**

* Computed calculations of percentages for projects that were successful, failed, or were canceled per goal range (5 points)
* Created a line chart showing the relationship between the goal’s amount and its chances at success, failure, or cancellation (5 points)

See excel file in submittal.

**Statistical Analysis (20 points)**

* Computed calculations of the mean, median, min, max, variance, and stdev using Excel formulas (15 points)

See excel file in submittal.

* A brief and compelling justification of whether the mean or median better summarizes the data (5 points)

|  |  |  |
| --- | --- | --- |
|  | Successful | Failed |
| Mean | 851.1 | 585.6 |
| Median | 201.0 | 114.5 |
| Minimum | 16.0 | 0.0 |
| Maximum | 7295.0 | 6080.0 |
| Variance | 1606216.6 | 924113.5 |
| Standard Dev | 1266.2 | 960.0 |

The variance for both successful and failed campaigns is quite high, indicating there is quite a wide dispersion of values. As a result, the mean may not be the best way to summarize the data. In this case, the median would be a more reasonable value to use to minimize the impact of outliers.