Crowdfunding Analysis Report

* Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
  + Crowdfunding campaigns are helpful in raising large budgets. The data shows 566 successful company outcomes out of 1001 companies.
  + The goal amount differs from pledged. There is a diverse range of success. It is hard to compare the outcomes of the different companies as it is in different currencies.
  + Crowdfunding campaigns are generally very cumbersome as large sets of data are pulled from different sources. It can get hard to keep accurate records if dates, currencies, goals, and other variables are not defined clearly.
* What are some limitations of this dataset?
  + Accuracy: As this large data set. There is a margin for error for recording data accurately. It is hard to organize such large data sets if there in any error in values recorded.
  + Comparing data: As data is different currencies and since there is some campaigns that were canceled, there success/failure across all values in the data set cannot be compared.
  + Dates: The launched date and deadlines are in a format which is hard to read without converted the data to year/month/day. When a person takes a first look at this data, they might be unable to make out the value without the knowledge of how to convert values in excel.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  + Scatter plot on all success, failed, cancelled outcomes based on dates to see outliers.
  + Mean, Median, Mode box plot.
  + Pie chart of percentages of successful, failed, canceled.
  + Table on average donation per category/subcategory
  + Graph representing staff pick/spotlight.

* Use your data to determine whether the mean or the median better summarizes the data.

For successful campaigns: Mean is larger (851.15) than median (201), there is a right skew of date. The outlier (7295) is probably influencing the skewing of data. Median here is less affected by outlier.

For unsuccessful campaigns: Mean (585.2) is larger than median (114.50), also presenting right skewness due to the maximum outlier value (6080).

In both these cases the median is the better summary of data as it less sensitive to outliers.

* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

There is more variability in the successful campaigns. As the variance and standard variation is higher in the successful campaigns compared to the unsuccessful campaigns.

A screenshot of a table

Description automatically generatedA screenshot of a spreadsheet

Description automatically generated



Some of the factors contributing to high variability in successful campaigns can be more resources to get diverse funding from, better marketing, better purpose of the campaigns, more gain to funders, different goals.