Written Report

Question1:

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

• Crowdfunding projects in the theater group were the most successful compared to other groups, and the Play-related campaigns were the most successful

• Mid-sized crowdfunding campaigns have the highest relative success rates (targets between 15,000 and 25,000)

•The success of crowdfunding campaigning is largely variable and dependent on those who pledge and the popularity of the project funded.

Question 2:

What are some limitations of this dataset?

First, I found that the data is not normally distributed and has many outliers, we need more samples to generalize the dataset to reduce the error. On top of that, this dataset comes from multiple countries and their currencies, in which case I think an additional column of data is needed to normalize these currencies to one currency and compare the results.

Question 3:

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

* 1. To perform outlier testing, you can first find outliers by using a box chart.
  2. In addition to this you can also create tables and charts by excluding outliers.
  3. Create a chart that shows the results of the quarterly and yearly campaigns. In addition, a bar chart showing the distribution of project status (e.g., completed, in progress, delayed, cancelled) will provide an overview of the status of the project. This can help identify any bottlenecks or issues that are affecting the progress of the project.
  4. Create charts and tables by comparing campaign success or failure in different countries.
  5. Create a comparison table or chart showing average contributions across campaign results

Statistical Analysis

|  |
| --- |
| 1. Use your data to determine whether the mean or the median better summarizes the data. |
| The Medium best summarizes the data for both successful and failed campaigns. Depend on the box plot, we can find there are too many outliers in both failed and successful data, that will affect the mean value higher which can't be good representations of these data. |
|  |
|  |
| 2. 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’ dataset. This makes sense because the successful campaigns’ dataset size is bigger than failed. According to the data, the minimum of successful dataset is lower, and the maximum is higher. |