# Written Report

**Q1: Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**

A1:   
1. Theater campaigns, specifically plays, are highly popular and demonstrate great success compared to the other categories, with 187 successful campaigns out of 344.  
  
2. On the other hand, journalism has a very low engagement, with a total of 4 projects, all of which were successful.  
  
3. August is a key month - high number of canceled and failed campaigns compared to other months.

**Q2: What are some limitations of this dataset?**

**Q3: What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

A2&3:   
1. **Year over Year Comparison** - Data lacks aggregation of data by year, which is necessary for comparing performance across different years. To analyze trends over time and make year-over-year comparisons, we need to categorize the data into annual segments. We could display this information in pivot table with an appropriate graph.  
  
Pivot Table:  
- Rows: Years extracted from the 'launched\_at' or 'deadline' date fields  
- Columns: Outcomes (successful, failed, canceled, live)  
- Values: Count of projects, sum of pledged amounts, and average donation per year  
  
Chart/Graph:  
- Line or column chart to show the number of successful vs. failed projects over the years  
- Stacked column chart to visualize the breakdown of outcomes (successful, failed, canceled, live) per year  
  
2. **Backer Donation History** - While the number of backers is provided, there is no detail on the level of backer engagement, such as returning backer history. This metric would provide further insight into successful crowding funding campaigns.  
  
Pivot Table:  
- Rows: Backer IDs or names  
- Columns: Project categories  
- Values: Count of projects backed, average, and total donations per backer

Chart/Graph:  
- Heatmap to show which categories attract the most repeat backers  
- Bar chart to display the average donation amount per backer across different categories  
  
3. **Campaign Duration** - While we have the launch and end dates for the campaigns, having a clear, calculated field for the duration of each campaign (in days) would be helpful. Campaign duration can significantly impact success, as longer campaigns may have more time to gather support, while shorter campaigns might require more aggressive marketing.  
  
Pivot Table:  
- Rows: Campaign names  
- Columns: Duration categories (0-30 days, 31-60 days, etc), created by calculating the difference between 'launched\_at' and 'deadline'  
- Values: Count of projects, average pledged amount, and success rate within each duration category  
  
Chart/Graph:  
- Bar chart to compare the success rate of campaigns by duration category  
- Scatter plot with campaign duration on the x-axis and pledged amount on the y-axis to visualize any correlation  
  
4. **Information on Marketing Efforts** - There is no data on the marketing strategies used for these campaigns. Information on promotional efforts, social media presence, or advertising spend can play a significant role in the success of a campaign and would be necessary to understand the factors driving backers to pledge.  
  
Pivot Table:  
- Rows: Campaign names  
- Columns: Marketing metrics (social media shares, advertising spend, etc)  
- Values: Count of projects, success rate, and average pledged amount  
  
Chart/Graph:  
- Bar chart showing success rates of campaigns with different levels of marketing efforts  
- Multi-line chart tracking average pledged amounts against different marketing metrics over time

# Statistical Analysis

**Q1: Use your data to determine whether the mean or the median better summarizes the data.**

A1: The median is better used to initially visualize the data, since it is less affected by outliers and skewed distributions. However, both the mean and the median hold significant value in statistical analysis. Since the mean provides an average that considers every data point, you can determine the nature of the distribution based on the relationship between the mean and the median. If the mean is less than the median, the data suggests a left-skewed distribution; whereas, if the mean is greater than the median, the data suggests a right-skewed distribution. In the Crowdfunding example, the data suggests a right-skewed distribution.

**Q2: Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

A2: The data shows that successful crowdfunding campaigns have a wider range of backers compared to failed ones. This wider range means that there is more variability in successful campaigns, creating a right-skewed distribution. This makes sense because successful campaigns often have qualities that can attract a broader audience, like innovative ideas or effective marketing, leading to some becoming exceptionally popular. These standout campaigns can draw in a large number of backers, acting as outliers and increasing the overall variability. On the other hand, failed campaigns tend to attract fewer backers overall, and without those extreme outliers, they show less variability in the number of backers. This reflects a general pattern where success in crowdfunding is not uniform; a small number of campaigns achieve outstanding success, contributing to the greater variability observed.