Written Report

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

Here are three conclusions we can draw about crowdfunding campaigns:

**Category Impacts Success Rates:** There are significant differences in success rates between project categories. Theater has the highest number of unsuccessful projects 132 failed and 23 cancelled, while Film & Video and theater have the most successful ones This suggests that the type of project significantly influences its chance of being funded.

**Possible Seasonal Trends:** There might be a slight trend in the number of campaigns launched each month While the overall numbers are similar, further analysis is needed to see if launching campaigns in specific months impacts success rates.

**Live Campaigns are a Small Minority:** Only a small portion of the campaigns (14 out of 1000) are currently listed as "live" on the platform. This could indicate several possibilities, such as:

* Campaigns might launch and quickly succeed or fail.
* The data might not capture the entire lifecycle of campaigns (e.g., some might be ongoing but not listed as "live").

2- What are some limitations of this dataset?

The dataset has some limitations that hinder drawing more definitive conclusions:

* **Sample Size:** With only 1,000 projects, the data might not be representative of the entire crowdfunding landscape.
* **Timeframe:** The timeframe for which the data was collected is missing. This makes it difficult to identify trends over time.
* **Platform Specificity:** The data likely comes from a single platform. Results may not be generalizable to other crowdfunding platforms.
* **Project Details:** The data lacks information about individual projects, such as funding goals, target audience, or marketing strategies. This makes it difficult to pinpoint factors contributing to project success.

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

Here are some other tables and graphs you could create to gain a deeper understanding of the data:

* **Project Subcategory vs. Outcome:** This table would break down the data by subcategory (e.g., animation, documentary) within each category. This could reveal success rate variations within broader categories.
* **Funding Goal vs. Outcome:** This table would show the relationship between a project's funding goal and its outcome. This could help identify the optimal funding goal range for crowdfunding success.
* **Location vs. Outcome:** This table would show the relationship between a project's location and its outcome. This could reveal any geographical trends in crowdfunding success rates.
* **Time to Completion vs. Outcome:** This table would analyze the time it takes a project to reach its funding goal and its outcome. This could help identify the optimal campaign duration.
* **Line Chart of Monthly Success Rates:** This chart would allow you to visualize success rates over the course of a year, highlighting any potential seasonal trends.

**Statistical Analysis**

Use your data to determine whether the mean or the median better summarizes the data. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

For successful campaigns, the median (560.78) better summarizes the data due to outliers affecting the mean (575.94). This is because the presence of a very high value (1425) can significantly inflate the mean. The median, which is the 'middle' value, is less affected by extreme values.

Similarly, for unsuccessful campaigns, the median (97) better summarizes the data than the mean (148.25) due to the presence of a high value (452).

We can use standard deviation to measure how spread out the data is. A higher standard deviation indicates more variability.