**Crowdfunding DATA**

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

1. In General terms, campaigns have the following outcome rates:

|  |  |  |  |
| --- | --- | --- | --- |
| **canceled** | **failed** | **live** | **successful** |
| **5.70%** | **36.40%** | **1.40%** | **56.50%** |

1. Theater, Film & video, Music and Technology, hold the 80% of the total campaigns, so we can focus on increasing the success rate of these categories in order to improve our results.



* 1. Plays (in theater) and Rock in Music are the most numerous campaigns and have the following success rate:



1. The best months to create a new campaign are **July, June and March.** Thus, the worst months to create a campaign are **August, December and February.** We calculated the difference between the Successful and Failed outcomes (Successful – Failed) and these are the results.

** **

* What are some limitations of this dataset?

1. The **blurb** field is not very clear so no further analysis can be done on this field and I consider this information as relevant.
2. Is not real time information so is biased to the downloaded date.
3. **Data Integrity:** Excel's can sometimes lead to errors, especially with large datasets where it's easy to make mistakes when entering or manipulating data.
4. **Memory:** large datasets can consume a significant amount of memory (RAM). With only 1000 rows, is not an issue, but with larger datasets, you might encounter errors or performance issues related with RAM.

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

1. Many campaigns on this dataset, have more than 200% of Percentage funded, so the table could be **Percentage funded vs categories.**
2. **Pledged sum. vs category/sub-category:** This table would provide the total amount of money donated per category in order to visualize which one has the bigger quantity.
3. Money should be standardized in this case to USD.
4. Total duration of campaign: this value should be added to the dataset in order to make comparisons between other fields; for ex. **Duration (days) vs. Pledged sum. ($).**

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

The mean is a better value to summarize the data in both cases because there are many outliers up the upper whisker, and the median is a very low value very close to the lower whisker.

 

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

Our values of variance and standard deviation are higher by almost 60% on the successful campaigns and it does make sense because there’s a higher number of successful cases in this study, with a higher number of samples, there is more chances for us to observe more variability.