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

Crowdfunding campaigns are raising money for a new development program and starting a new small business project. I did analyze 1000 sample projects in late 2000,

The first analysis was category based for the successful and unsuccessful projects. The stacked column chart shows, most campaigns were trialed in the theater part of plays, but only successful 187 and failed 132 campaigns. Also, each field has half successful and half failed except journalism. Only 4 campaigns in journalism were all successful. In top categories like film, music, and theater, music campaigns were less failed percentage.

The second analysis was based on the sub-category of successful and unsuccessful campaigns, 344 campaigns were tried in the theater/plays field, 187 were successful and 132 failed. Also, music is a good choice for the campaign, less failure is happening.

The third analysis is what month campaigns get good results. Most campaigns were successful in July. Fail is comparable. Some campaigns failed more in January and march. The result shows July and august are the best months for getting successful results.

* What are some limitations of this dataset?

1. Datasets are skewed type. So mostly we get the median of the datasets is good for the analysis.
2. The category-based data sets show only the main field comparison only. So, we can’t figure out the actual percentage of successes and failures.
3. The year-based data sets show only what happening in all months. We can’t predict time and days depending on the seasons.

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

1. The option to find a good result based on country. Then we know good campaigns possible in which place to start.
2. Launching dates are another option to find good results. The graphs show what is the best day and time to start campaigns.

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

Both datasets (successful and unsuccessful) show mean is greater than the median, which means a positively skewed distribution. In this case, a median is a good option for analyzing 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?

Successful and failed campaigns’ R^2 value in regression is less than 1 which is 0.0013 and 0.0004, which means weak data value. The relationship between the variable is weak variance. R^2=1 is the best description of the data values.