1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
2. Theatre with a total backer of 264,269 and with an average of average donation value of 69.17; is the most popular crowdfunding category contributing approximately 33% and 36% of total successful and unsuccessful projects respectively. Journalism is the least popular crowdfunding category with total backers of 1,194 with an average of average donation value of 30.49 and 100% of success.
3. Play within the Theatre parent category is the most invested crowdfunding campaign with success rate of approximately 54% and a failure rate of 38% respectively. The total contribution of Play is approximately 33% of total successfully crowdfunding campaign in all countries.
4. Most (763 of 1000) of the crowdfunding campaigns are occurring in the US with an average of average donation value of US$68.59 per campaign. The success rate of crowdfunding campaign is 54.6% ± 2.9% for all countries.
5. What are some limitations of this dataset?

Some of the limitations are:

* All the campaigns are short term campaigns, i.e., campaigns occurring for few weeks. Results could be used for long term (projects/campaigns running for few months to years) crowdfund campaigns.
* Crowdfunding campaign datasets is limited to few countries and therefore it could be considered as a sample dataset not as a population dataset.
* Total projects with goal value ranging from 10,000 to 50,000 are few in comparison with projects with values below 10,000 and above 50,000. The results from the analysis could be biased.

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

Charts:

* Pie Chart: It would provide contribution of each parent category/ sub-category in outcome for each country.
* Box and Whisker: It would help in identifying the outlier for failure, as it would help in further investigation of that crowdfund campaign.

Bonus

1. Use your data to determine whether the mean or the median better summarises the data.

Mean better summarises the data. Median is suitable when the data is evenly spread, however for the above problem data is not evenly distributed. So median results will be skewed.

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

Successfully campaigns have more variability in comparison with unsuccessful campaigns, as the number of outliers is more for the former. It makes senses, as the data is closer to mean for the unsuccessful campaigns (variance) than the successful campaign.