Part 1.

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

* The data shows that the crowdfunding related to music, film & video and theater have the most successful chances, compared to other parent category outcome, especially the sub-category of plays.
* Regarding to the Campaigns analysis by date, that shows second quarter of year contains more chance to success. And clients can launch their products and do their marketing campaigns in Summertime, starting from May to July.
* If the goal was set around $20000, clients would get higher percentage successful rate.

1. What are some limitations of this dataset?

* Small sample size, 1000 dataset may not be representative of the whole real-world and may not provide enough information.
* Outdated data, some datasets are from 2010, and the data from 20 years may not show the actual value or result that we need to match the recent needs.

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

Pie chart, a pie chart can be used to compare values across different categories, e.g., Parents category or Sub-category.

Part 2.

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

Table 1:

|  |  |  |
| --- | --- | --- |
|  | **Successful Values** | **Failed Values** |
| **Mean** | 851 | 586 |
| **Median** | 201 | 114.5 |

Table 2:

|  |  |  |
| --- | --- | --- |
| **10% rule Central tendency** | | |
| **Difference** | 650 | 471.5 |
| **10% of Smaller** | 20.1 | 11.45 |

Chart 1:

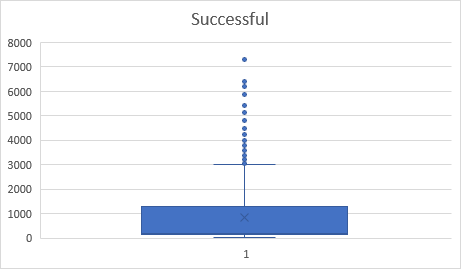
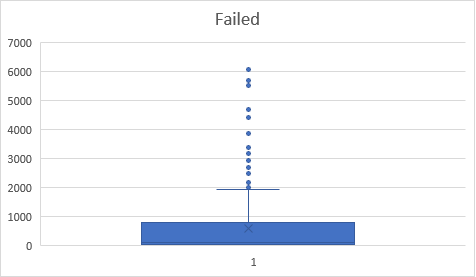


Chart 2:



Regarding to the chart 1 and 2 above, which shows there are outliers in the dataset of successful and unsuccessful campaigns.

Additionally, Central tendency 10% rule is used to decide whether the mean or median is the preferred measure. Since the difference in Table 1 between the mean and median is relatively greater than 10% of small, and the mean is higher than the median in both values of “Successful” and “Failed”, which suggests that there are some high backers count in the dataset, which could be affecting the mean. Therefore, in this case, the median may be the preferred measure of central tendency since it is less affected by the outliers.

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

It does make sense by compare the data from different time period. But at the same time, the dataset is kind of outdated, we need something close to recent, for example, we should use the data from pass 10 years.