* Following three conclusions can be drawn from the data:

1. Based on data represented in Pivot Table -1 sheet, it can be observed that the most popular Parent Category of Crowdfunding projects is ‘theater’, followed by ‘film & video’ and ‘music’. This may be because a lot of people are present at once for theater so campaign can target more people at the same time.
2. Based on the data represented in Pivot Table -3 sheet, it can be observed that Jun-Jul are the most favorable months of the year for having successful campaigns. This could be because it is summer time and lot of people are out and about so it is more favorable for the campaign.
3. Based on the data represented in Goal Analysis sheet, it can be observed that campaigns with a goal between the range 15,000 to 35,000 are most successful.

* Following are the limitations of the dataset:
  1. Dataset has only limited sample size. It does not represent different countries equally and hence can’t be used across the globe. For e.g. US has 763 campaigns listed whereas China has only 23.
  2. The currency is not standard throughout the dataset which could affect the analysis of the data.
  3. Data about the background of backers, source (online transfers, checks, etc) of collected funds is not provided which could affect the outcome of the campaign majorly.
  4. There are outliers present in the data (based on analysis done in Central Tendency sheet), hence analysis would be difficult.
* Following are the suggestions for additional tables of graph:
  1. Determine how many campaigns were successful or failed based on the duration (Date Ended conversion – Date Created conversion) of each campaign. This will help understand the favorable average duration of a campaign to be successful.
  2. Removing the outliers and then doing analysis with Pivot Tables, graphs and Central Tendency.
  3. Creating graph and analysis based on Individual Country to understand the trend of campaigns country wise.

**Statistical Analysis**

|  |  |
| --- | --- |
| Successful Campaign | Failed Campaign |
| |  |  | | --- | --- | | Mode | 85 | | Mean | 851.1469 | | Median | 201 | | Minimum | 16 | | Maximum | 7295 | | Variance | 1603373.7 | | Std Deviation | 1266.2439 | |  |  | | First Quartile | 127.5 | | Second Quartile | 201 | | Third Quartile | 1288.5 | | Inter Quartile Range | 1161 | | Lower Boundary | -1614 | | Upper Boundary | 3030 | | |  |  | | --- | --- | | Mode | 1 | | Mean | 661.0533 | | Median | 114.5 | | Minimum | 0 | | Maximum | 6080 | | Variance | 921574.7 | | Std Deviation | 959.9868 | |  |  | | First Quartile | 38 | | Second Quartile | 114.5 | | Third Quartile | 789.5 | | Inter Quartile Range | 751.5 | | Lower Boundary | -1089.25 | | Upper Boundary | 1916.75 | |
| As we can see that in both cases, Mean is greater than Median and Mode, the data is Right Skewed. Also as observed in the box plot(refer Central Tendency sheet in excel), there are a lot of outliers which drastically affect the calculation of the mean in the data. Hence, the Median will better summarize the data. | |