# REPORT MODULE 1 CHALLENGE – JUAN SIERRA

Create a report in Microsoft Word, and answer the following questions:

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

1. In aggregate, June and July seem to be the best months to launch a campaign, given the number of successful campaigns during those months
2. The three most popular categories are, in order:
   1. Theater – 34.4% of all campaigns
   2. Film and Video – 17.8% of all campaigns
   3. Music – 17.5% of all campaigns

Combined they account for 69.7% of all campaigns

1. Within Theater, the only subcategory available is “plays”. It clearly dominates all other subcategories, with the 34.4% indicated in the previous point. A distant closest subcategory is “Music-Rock” with 8.5%, followed by “Film & Video / documentary” with 6.0%. All other categories are below 5.0%.
2. Campaigns that have a goal between $20,000 and $50,000 have a greater percentage of successful campaigns.

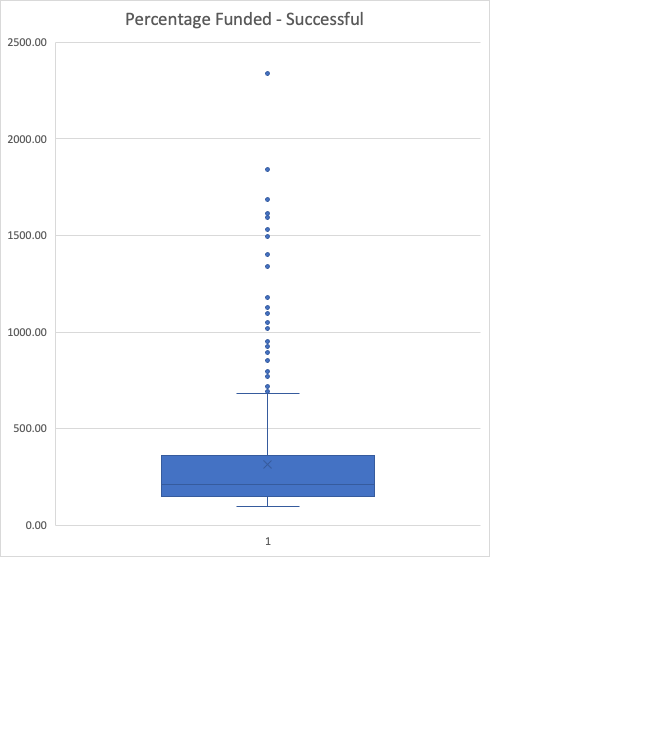
## What are some limitations of this dataset?

1. There is not enough context to determine what causes a campaign to fail or succeed. The columns “staff pick” and “spotlight” may provide partial insights, but do not seem to have a clear correlation
2. It would be beneficial to know the target audience for each campaign and some information on the demographics of the backers.

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

1. Number of outcomes based on goal – clearly shows that the number of campaigns below $10,000 and above $50,000 have the greater number of campaigns, and that successful campaigns tend to fall within $1,000 and $10,000.
2. Percentage Funded of Successful Campaigns: The statistical figures are as follows

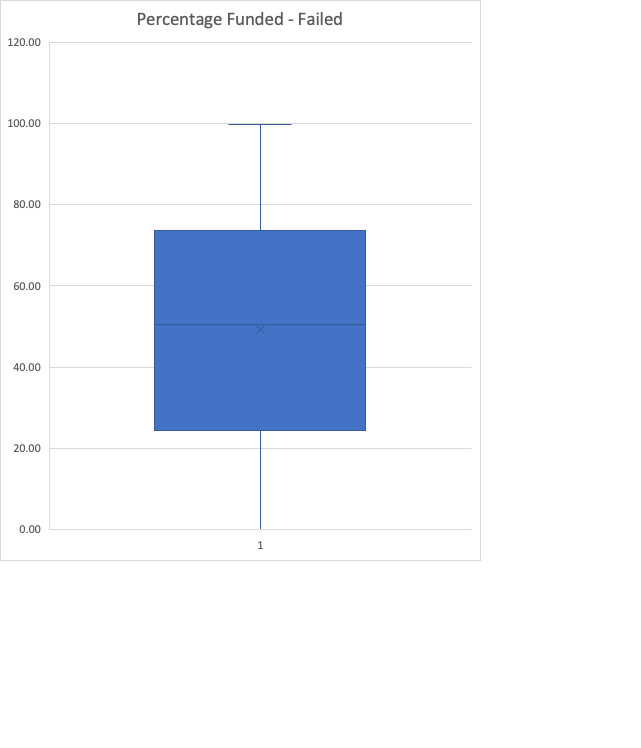
|  |  |
| --- | --- |
| Mean | 317.270805 |
| Median | 212.304348 |
| Min | 100.011506 |
| Max | 2338.83333 |
| Variance | 80728.2114 |
| Sd Deviation | 284.127104 |



The median in this case is the better representation of the percentage funded for successful campaigns, since large outliers skew the data. The Median in this case is 212% of the campaign’s goal.

1. Percentage Funded of Failed Campaigns: the statistical figures are as follows

|  |  |
| --- | --- |
| Mean | 49.286742 |
| Median | 50.5519206 |
| Min | 0 |
| Max | 99.6835443 |
| Variance | 853.228433 |
| Sd Deviation | 29.2100742 |



In this case the Mean (i.e., Average) is a good representation of the percentage funded for failed campaigns, with 50.55% of the campaign’s goal.

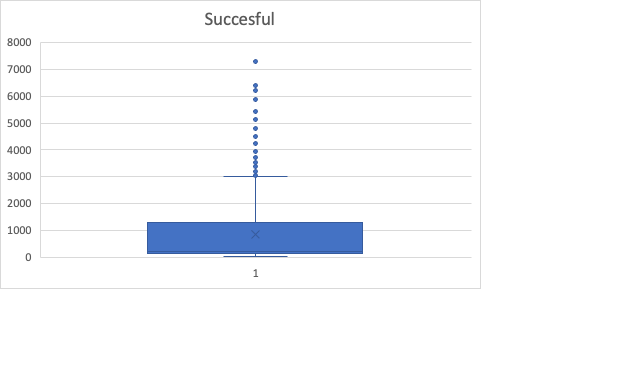
1. The campaigns are mostly run in the United States, with 76.3% of all campaigns run in this country.

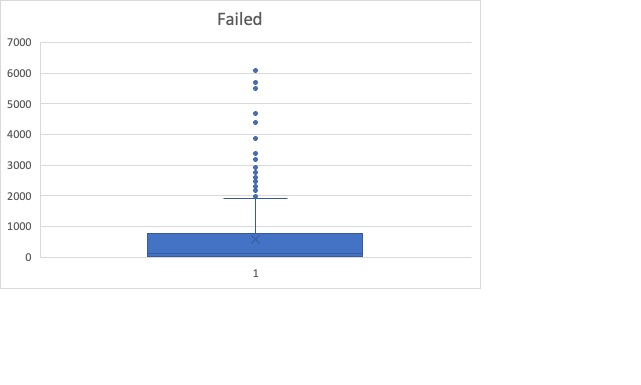
**STATISTICAL ANALYSIS**

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

1. The median is a better representation of the data, since there are large outliers that skew the data. This is evidenced by the large standard deviation compared to the mean of each data set, and also in the whisker plots that show the large number of datasets that fall outside of the 1,5 x IQR rule.

|  |  |  |
| --- | --- | --- |
|  | Successful | Failed |
| Mean | 851.146903 | 585.615385 |
| Median | 201 | 114.5 |
| Minimum | 16 | 0 |
| Maximum | 7295 | 6080 |
| Variance | 1606216.59 | 924113.455 |
| Standard Deviation | 1267.36601 | 961.3082 |
| Mode | 85 | 1 |





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

1. There is more variability for successful campaigns. This makes sense, because there is the probability that a wider range of backers will participate in a successful campaign than an unsuccessful one.