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

A. Based on Category, Crowdfunding is more successful with Theatre the most (187) and Journalism the least (4).

Please refer the table of successful campaigns based on category in decreasing order.

|  |  |
| --- | --- |
| **Category** | **Successful** |
| Theatre | 187 |
| Film & Video | 102 |
| Music | 99 |
| Technology | 64 |
| Publishing | 40 |
| Photography | 26 |
| Food | 22 |
| Games | 21 |
| Journalism | 4 |

B. Based on goals, we concluded that campaigns are more successful where goals are in the range 1000-5000.And the campaigns are most failed where the goal is above or equal to 50000. Please refer the table of successful & failed campaigns based on Goals.

|  |  |  |
| --- | --- | --- |
| **Goal** | **Number Successful** | **Number Failed** |
| Less than 1000 | 30 | 20 |
| 1000 to 4999 | 191 | 38 |
| 5000 to 9999 | 164 | 126 |
| 10000 to 14999 | 4 | 5 |
| 15000 to 19000 | 10 | 0 |
| 20000 to 24999 | 7 | 0 |
| 25000 to 29999 | 11 | 3 |
| 30000 to 34999 | 7 | 0 |
| 35000 to 39999 | 8 | 3 |
| 40000 to 44999 | 11 | 3 |
| 45000 to 49999 | 8 | 3 |
| Greater than or equal to 50000 | 114 | 163 |

C.Most of the crowd funding campaigns were held in US, that’s why we can see the most successful as well as failed goals here country wise.

2.What are some limitations of this dataset?

A. We see that the data is more US centric. We can collect more data from other countries to make it more helpful as to what all places we can do more campaigns.

B. We see that we had a limited data based on years (2010-2020). We also don’t have any data after 2020, so if we have more recent data, our analysis could give better results.

C. We just have the backers count and its not helping us to know who exactly pledged the most. If we know that, it can help in the future campaigns.

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

A. Outcome can be revisited, based on percentage of goal. For example, now we are simply calling the campaign failed just because it is less or not equal to the pledged amount. We can classify it further giving the range.

B. We can create a outcome based table with campaign launched date and deadline date to find out how long a campaign needs to run to make it a successful one.

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

I Would say the median better summarizes the dataset. Because:The main purpose of the crowdfunding dataset is to determine the outcome based on successful or failed pledges. As we can see the difference between the values of minimum and maximum successful or failed pledges. So, the Mean of the data is distorted by the outlier’s values. And in this case, Median should be helpful as it is giving us the overall picture of the whole dataset.

|  |  |  |
| --- | --- | --- |
|  | 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 |
|  |  |  |

So to conclude, “The mean of a dataset is the sum of values divided by the number of values and the Median is the value in the middle of dataset. The mean is typically better when the data follow a symmetric distribution. When the data are skewed, the median is more useful because the mean will be distorted by outliers*.*”

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

Yes, there is more variability with successful or unsuccessful campaigns. As, variance is a measure of dispersion that considers the spread of all data points in a data set. A high variance indicates that the data points are very spread out from the mean, and from one another.

|  |  |  |
| --- | --- | --- |
|  | 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 |