**Module 1 - Challenge  
  
  
Written Report (20 points)**

**States limitations of the dataset and suggestions for additional tables of graph (10 points)  
Create a report in Microsoft Word, and answer the following questions (20 Points):**

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?  
   We can conclude from the data:  
   1) From 1000 Crowdfunding campaigns around Australia, Canada, China, Denmark, Great Britain, Italy and the US, more than half are successful at 56.5%.  
   2) The top successful parent categories campaigns are Technology, Photography and Journalism. All three has success rate over 60%.  
   3) In Pivot Table 4, we can see overall more campaigns were successful than canceled or failed.
2. What are some limitations of this dataset?  
   Some limitation of this dataset is the number of campaigns for each country. It is heavily dominated by the US campaigns with Italy, Great Britain (GB) and Australia represented equally. There is an assumption that these countries have the same representation outside of the 1000 campaigns followed. Which they may not. Also, another limitation is there are 7 countries versus funding campaigns around the world.
3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?  
   Some possible tables can be created are: Success Rate which would be Number of Successful Outcomes over Number of Campaigns by Month or Year and Unsuccessful Campaigns, by combining Failed and Canceled campaigns and creating a percentage over the total submissions.

**Statistical Analysis**

|  |  |  |
| --- | --- | --- |
| **Statistics** | **Successful** | **Failed** |
| Mean | 851.15 | 585.62 |
| Median | 201.0 | 114.5 |
| Minimum Backers | 16 | 0 |
| Standard Deviation | 1266.24 | 959.99 |
| Variance | 1603373.73 | 924113.45 |
| Mode | 85 | 1 |

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

The data is better summarized by the median. With the mean and median, for both, successful and failed campaigns, being very different, the data has a skewed distribution. With the mean > median> mode, for both of these outcomes, they have a left skewed distribution. With a skewed distribution there are many outliers in the data. Using the mean would count the outliers and give the perception of a much higher number of backers to a campaign for successful and failed campaigns. Also, the large variance (1603373.73; 924113.45) and standard deviation (1266.24; 959.99) from the mean also shows the disparity between the average and where most of the data falls in. The mode better represents the number of backers for the selected outcome.

|  |  |  |
| --- | --- | --- |
| **Statistics** | **Successful** | **Failed** |
| Mean | 851.15 | 585.62 |
| Median | 201.0 | 114.5 |
| Minimum Backers | 16 | 0 |
| Standard Deviation | 1266.24 | 959.99 |
| Variance | 1603373.73 | 924113.45 |
| Quartile 1 | 128 | 38 |
| Quartile 3 | 1280 | 787 |
| Inter Quartile Range | 1152 | 749 |
|  |  |  |

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

There is more variability with the outcome of successful. With calculating the Interquartile Range, we can see the outcome of successful (1152) has a larger range than the outcome of failed (749). This makes sense because the successful campaigns hit their goals, with supporters pledging different amounts of funds. These goals vary in range and therefor the amount pledged and how many backers to reach the goal also vary. With the minimum numbers of backers for a campaign to reach the successful outcome in this data was 16 and the highest was 7295. With the failed outcome there is less variability because they did not reach their goals and the numbers of backers who pledged were low.