**Written Report (20 points)**

|  |
| --- |
| Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?   1. According to the pivot table and the chart based on parent category, theater emerges as the category with the highest number of participating backers in crowdfunding, totaling 344 out of 1000. The second most popular categories are the music and film & video. 2. Among the sub-categories, play stands out as the dominant sub-category and the scale is much larger than other sub-categories. 3. The country where crowdfunding is most active is the United States with the highest number of participating backers, with 796 out of 1000. |
| What are some limitations of this dataset?   1. Theater only has play as it’s sub-category whereas other parent categories have several sub-categories. For this reason, the scale of play is too large to be compared to other sub-categories, which causes play to be an outlier. 2. Sub-categories corresponding to each category are too limited. For example, there is only one sub-category in theater and food, respectively. 3. There is no demographic information about backers such as age, gender, or occupation. If there is such information, it will help to identify potential target audiences. |
| What are some other possible tables and/or graphs that we could create, and what additional value would they provide?   1. A graph showing changes in backers according to year for each category. 2. A table or graph showing the length of time each campaign runs, starting from its launch to its conclusion. 3. Additional values: demographic information about backers, the source from which the backers know about the campaigns. |

**Statistical Analysis (20 points)**

|  |
| --- |
| Use your data to determine whether the mean or the median better summarizes the data.  In this case, the mean better summarizes the backers’ data because the data is widely ranged and there are no significant outliers. |
| Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?  The data of successful campaigns has more variability because the standard deviation is greater than that for unsuccessful campaigns. I think it makes sense because there might be many underlying aspects and efforts that attribute to the successful campaigns. |