**Part 1:**

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

* Theater (sub-category = plays) had the most crowdfunding projects (34% of total projects) with film & video and music following at 18% of total projects each
  + Theater also only had one sub-category, plays, whereas film & video and music both had 6 sub-categories; therefore, plays had the largest dataset
  + Overall, Theater-Plays had a 54% success rate vs 38% failed
  + Mobile games had the least amount of success compared to all categories
* A goal of $15,000 and $24,999 had the most successful crowdfunding projects
* Overall, crowdfunding projects were more successful than not
* The more backers there are for the crowdfunding project, the more likely it is to be successful
* There is a greater variance in the successful project’s backers count than the failed projects
* Years 2017-2019 had the most successful projects

**What are some limitations of this dataset?**

* We are not able to determine what crowdfunding platform was used for the projects
  + This could help draw conclusions on investigating why a project may be more successful than another if there were more sources or the sources were labeled
* There is a large range of project count by category. For example, audio, metal, radio & podcasts, and world music had less than 10 projects in the dataset; whereas, plays had 344 projects. This could mislead some of the analysis
* There wasn’t an equivalent number of projects evaluated in each year. Range is also quite large with only 2 projects in 2020, but nearly 100 in all the other years

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

* Analyze the percent of successful, failed, canceled by category to determine how each category did overall to others
  + Could answer: What one category more successful overall than another?
  + I analyzed some of this in the purple cells on my spreadsheet
* Analyze the average number of backers for each category and the outcome of the project to see if there’s a correlation between the number of backers and success rate by category (did a bit of this on tab “Additional - Backers Analysis”)
* Staff pick and spotlight could be analyzed to see if there’s any correlation with that and the success of the project
* Analyze percent funded to see if there was a correlation between that and the success of the project (on tab “Additional - % Funded Analysis”)

**Statistical Analysis:**

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

* The median would be better to use in this case since there is very high variability within the data. For successful campaigns, the mean is 851 backers versus the median is 201 and for failed campaigns, the mean is 585 with a median of 114. The variance of 1603373 and 921574, respectively, shows that the data has a large spread from it’s mean value. With the high variation among the data and the skewed data, median is the obvious choice.

**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 successful projects (variance = 1603373 and standard deviation = 1266) than unsuccessful (variance = 921574 and standard deviation = 959). I believe this makes sense because the more backers there are the higher the probability the success of the project.