2024/04/29 Jinseo Baek

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

* From the Parent Category, we can analyze that Film & Video has the highest demand. Therefore, it carries a higher risk rate than others but also presents higher opportunities for success.
* Regarding Sub-Category, plays show a significant demand. However, mobile games and science fiction have higher failure rates than success rates, making crowdfunding for them too risky.
* Regarding the Month, June and July exhibit the highest success rates. Considering the higher demand during these months, the likelihood of failure is lower.

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

* Differences in currency could lead to inaccuracies.
* Because the data is from previous years, there might be issues with data quality.
* The varying project durations make it difficult to standardize the data.

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

* Pie charts: These can illustrate the proportion of different outcomes between Failures and Successes, making it easier to visualize the overall distribution.
* Time Series Analysis: Time series tables allow us to visualize trends over time, making it easier to observe long-term changes.

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

Mean vs Median:

* For successful campaigns, the average number of backers is 851, and the median is 201.
* For unsuccessful campaigns, the average number of backers is 565, and the median is 118966.
* In both cases, the median better summarizes the data. This is because the mean can be heavily influenced by extreme values, such as outliers, while the median provides a more robust measure of central tendency.

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

Variability:

1. For successful campaigns, the standard deviation is approximately 1267.
2. For unsuccessful campaigns, the standard deviation is approximately 919.37.
3. There is more possibility to have variability in successful campaigns, as it shows the higher standard deviation compared to unsuccessful campaigns based on the number that I calculated. This might seem counterintuitive at first, as one might expect unsuccessful campaigns to have more variability. However, successful campaigns can vary widely in terms of their backers count, ranging from very few to extremely high numbers. Therefore, it makes sense that successful campaigns would exhibit more variability.