Brandon Valmont

Rice Data Analytics Homework 1

1. **Given the provided data, what are three conclusions we can draw about Kickstarter campaigns.?** Based on the provided data, music is a category that has the best chance of becoming successful in a Kickstarter campaign as the music category had a 77.1% success rate. The data also seems to show that the beginning of November through December is not the best time to start a campaign due to number of successful campaigns decreasing tremendously and the number of failed campaigns increasing; perhaps due to the holiday season. It also appears that goals with a range of 1,000 to 4,999 has the best chance of being successful compared with other goal numbers. Another indicator of the success of a campaign is the number of backers. In this data set, the average number of backers for a successful campaign was 194.43 compared with an average number of backers of 17.71 for failed campaigns.
2. **What are some limitations of this dataset?** There were more than 300,000 projects launched on Kickstarter. The dataset only sampled 4,000 past projects. It may be possible that this sample of data does not accurately reflect the 300,000+ past projects. A larger sample may have been more appropriate for the dataset given. Success in the dataset is defined as reaching or exceeding a goal. There have been plenty of projects that have reached its goal but has ultimately been known as a failure. The definition of success itself could have been drilled down a bit further.
3. **What are some other possible tables and/or graphs that we could create?** A graph showing the number of backers compared to the outcome or state of a campaign would have clearly shown that if a significant number of backers are not behind a project, that project will likely fail.

**Bonus Question:**

Use your data to determine whether the mean or the median summarizes the data more meaningfully.

Based on my observation, it would seem the median summarizes the data more meaningfully due to the standard deviation and variance of the number of backers are high. The median would be a more meaningful number for finding the middle in this case.