Module 1 Challenge Hok Yin Cheung

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

1. Among all Parent Categories, Journalism (100%) has the highest successful rate. Technology (67%), Photography (62%), Publishing (60%) range from 60% to 70%. Film & video (57%), Music (57%), Theater (54%) range from 50% to 60%. Food (48%) and Game (44%) range from 40% to 50%. Based on the success rate above, Categories Journalism, Technology, Photography, and Publishing are more likely to yield successful projects. Note that Journalism has only 4 groups, despite its 100% success rate; its sample size is relatively smaller than other categories.
2. Among all sub-categories, audio and world music (100%) had the highest successful rate. Web has 71%, other sub-categories are between 30%-70%. Audio belongs to the Journalism category; world music belongs to Music category and Web belongs to Technology category. These three sub-categories more likely to yield successful project. Despite 100% successful rate, audio (4 sample size) and world music (3 sample size) are relatively smaller than Web and other categories.
3. Launching projects in June and July has higher successful rate among 12 months. January, May and August have higher failed rates among 12 months. August also has the lowest successful rate and highest cancellation rate.

Q: What are some limitations of this dataset?

1. Database sample size: The current database consists of 1,000 sample; it might not represent the whole population of crowdfunding categories.
2. Outside effect: Economic downturns might affect people supporting projects.

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

Beside line graph and bar graphs, we can also use pie chart to provide a clear visual between the relation of outcome rate and categories. Box and whisker plots and scatter plots might also help to provide a better picture of distribution and correlation between backers and outcomes.

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

I think mean is better to summarize the data. With the given data for successful campaigns – Mean 851, Median 201, Min 16, Max 7295. And unsuccessful campaigns – Mean 586, Median 115, Min 0, Max 6080. Comparing both median and mean, the median of these two datasets is somewhat skewed toward Minimum and mean counts every value in the dataset. Therefore, I think mean is better.

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

In backers vs. campaigns outcome, there is more variability with successful campaigns. This means some successful campaigns have huge number of backers and some might have a few to reach the goal. It makes sense that successful campaigns usually have more backers to increase the possibility of reaching the goal.