Kickstarter Campaign

1. One of the conclusions that can be drawn from the data is that theater, music, and film & video have the highest successful rate. Plays contribute the most to theater success. In music, rock contributes the most to success rate while documentary contributes the most to film & video success rate.

Another conclusion that can be drown from the data is that under technology chances of success, failure or cancelled fall close to each other, which indicate there is almost an equal chance of success, failure, or cancelation. Hardware contributes most to success rate. Space has almost equal chance of success or failure. Web and wearables are some of sub-categories that contribute to failure and cancelation in this category.

Last, success rate vs failure rate as seen on Sheet4 seem to be higher in the first half of the year then declines for the second half of the year. Cancelations are less volatile compared to Success and failure.

2. One limitation of this dataset is that it can be time consuming to manually enter such large amount of data in a spreadsheet. However, it could be possible to extract this data from other sources into excel. Another limitation that was mentioned in class is that spreadsheet use a lot of storage space. This in dataset has a lot of information and could be using a lot of storage space. I’ve used databases before, and I think they could be more efficient handling this kind of data compared to spreadsheet. Database only retrieve what you currently working with.

3. One graph that could be useful for the data is pie chart. A pie chart can visualize state i.e. successful, failed, cancelled and live better than applying different colors like we’ve done on the spreadsheet. Excel tables can also help manage analyze the data much easier. For example, headers have filters that can make sorting of data easier. Alternate shading of data makes it easier to read the dataset. Adding a Total row to Excel Tables automatically gives you the sum of the columns.

Bonus Statistical Analysis

1. Graphs on sheet 7 indicate presence of outliers that could affect the mean for both successful and unsuccessful campaigns, therefore median could help summarize the data more meaningfully than the mean.

2. Variance and standard deviation indicate that successful campaigns are more volatile than unsuccessful campaigns. Also, there is a bigger difference between mean and median for successful campaigns compared to unsuccessful campaigns, which could indicate presence of outliers in successful campaigns that enlarge the mean. This is also evident with the range or difference between minimum and maximum backers. Range for successful campaigns is larger compared to that of unsuccessful campaigns.

1.