**Paola Londono**

**Module 1 Challenge**

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

Looking at the Outcome Count per Parent Category chart, I can observe that about half of them are successful and most of them are in entertainment including “Theater”, “Music” and “Film & Video”. Nevertheless, these 3 categories also have the largest number of failed projects.

Then going to the Outcome Count per Sub-Category Chart, we can see that the most successful Sub-Category is for “Plays”, which at the same time has the most failed outcomes.

The Date Conversion chart shows how the month of December has period from August to December could be considered “low season” for successful outcomes and from January to July could be considered “high season”.

* What are some limitations of this dataset?

I’m able to measure the outcome count per category but not the outcome rate per category, and I think this will provide me with better information about the real projects’ success and fail rates.

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

As mentioned before, I will include an outcome rate per category. I will add the following columns called Canceled Rate, Failed Rate, Successful Rate, and Live Rate. They are calculated by dividing the count outcome per category by the total of each outcome. From there I can create two separate charts, one for the Parent Categories and another for the Sub-Categories.

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

Looking at my Box and Whiskers Boxes, I see how the Means are positioned closer to the middle of the IQR section meanwhile both medians are closer to the Quartile 3 line. Therefore, I would say that mean is a better way to measure to summarize data in this opportunity.

* 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 campaigns because first, the variance is higher that the failed variance; second the standard deviation is also higher for the successful rate; and third, the box plot shows outliers with a max away from the mean higher than the failed box plot numbers.