**Student:** Ezinne Obayagbona

**Problem Questions - Answers**

**The three conclusions that we can draw about the crowdfunding campaign are:**

1. We have a high successful outcome from the campaign as there were 565 successful outcomes.
2. The highest and most common Parent category and Sub category for crowdfunding is Theater and Sub category respectively. We see from the data that it has the highest count of 344 grand total which includes both failed and successful crowdfunding outcome. We can infer from this that this is a subcategory mostly used to crowdfund a project.
3. July appears to be the most successful month for crowdfunding. From the data, the month of July had the highest successful project of 58 total successful projects. Also from the data it appears that there were higher grand total of crowdfunding campaigns (94) which can give us an idea that June tends to be a busy month for crowdfunding. It can also mean that since there were higher number of campaigns there’s a possibility to have higher number of successful campaigns.

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

1. There is a possibility that some backers have made a pledge but may under deliver on the pledged amount
2. There is no guarantee that the projects that is being crowdfunded will be implemented as expected and will get enough backers. There is a risk of not reaching the funding goal.

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

We can further create a table/ graph on the date of goal completion in correlation to the success rate to see how quick a pledge is made.

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

In my opinion, I feel that the median better summarizes the data. The reason for this is that in crowdfunding projects, generally, there might be skewed distribution being that a few successful projects may have a larger number of backers than others so there will be outliers which can influence the mean calculation pulling towards higher values. On the other hand, using a median shows more representation of the data and provides a more realistic number of backers. It represents the value at which half of the projects have fewer backers.

**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 or unsuccessful campaigns. This makes sense as I explained previously since a larger variance indicates greater variability, while a smaller variance indicates less variability and a more clustered dataset around the mean. Looking at the variance for successful backers it is higher than that of unsuccessful backers so there’s a possibility that there were a few successful campaigns that had higher number of backers than others.