**Kickstart My Chart**

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   1. Maximum number of successful campaigns are from the areas of theater and music. Possible Explanation: Theatre and music and the most attention drawing areas where masses like to connect.
   2. Campaigns with the goal set to less than $5000 are most likely to be successful. Possible Explanation : Smaller targets are more likely to be achieved.
   3. The month of May observes the maximum number of successful campaigns while most of the campaigns launched in the month of July would fail. Possible Explanation: During summer months people are in the leisure mood, connecting with family as kids come home during summer break, and hence are able to think on issues other than their own everyday stresses at work /school.
2. What are some limitations of this dataset?
   1. The campaign is providing data which is 3 years old and may not be relevant in today’s business environment.
   2. The data has large number of outliers and hence the distribution is highly skewed.
   3. The data set does not identify the techniques used in the various campaigns which could also be the underlying factor of their success or failure.
3. What are some other possible tables and/or graphs that we could create?
   1. Analysis of Average Duration of campaign vs success of campaign
   2. Graph of Region wise successful campaigns
   3. Analysis of Category VS Percent Campaign Funded to identify the most well-funded campaigns

**Bonus Statistical Analysis**

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

Table 1 in Sheet ‘Bonus Statistical Analysis’ reveals that the average number of campaign backers in the successful campaigns is 199.74 while the median value is only 61. In this case the mean is not accurately useful as the skewness of the data is very high at 21.4 with a kurtosis of 583.105. This is an indicator of large number of outliers in the data distribution. Hence, mean would not be an accurate representation of the central tendency and we should select median to be our accurate representation of central value.

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

Variance is a metric which tries to explain how far away a data point is from the middle/central value. The variability of the data set for successful campaigns is 713167.3 while that of the unsuccessful campaigns is 3775.6. Looking at these values one can say that the values of the successful campaigns is not consistent and have a lot of outliers. Hence it can not be used to predict the future success of the campaign. Similarly the variance of unsuccessful campaign even though less than that of the successful campaign, it very high. I think, it point towards the fact that the success or failure of any given campaign is very random and hard to predict.