* **Create a report in Microsoft Word and answer the following questions.**

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

* Total projects related to theater have the higher number of campaigns. Plays dominate the number of campaigns. Moreover, total projects on theater have more successful campaigns in comparison to the other categories except for music. Under the category of music, the rock genre dominated the number of campaigns and with 100% of success. Interestingly, the category that raised more money and had the higher number of backers is technology.
* The majority of the projects (88%) are coming from USA followed by Great Britain.
* Projects with a high monetary amount of funding goal tended to fail in reaching their goals in comparison to those projects that were gathering small amount of money. Similar trend was observed in projects that were cancelled. Further studies to the data set revealed that theater and music had more projects with goals ranging from <$1,000 to $20,000 which correlate with high success. However, projects related with technology had more projects with goals higher than $60,000; 65% of the projects on technology failed or were canceled.

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

The dataset does not represent the total number of projects launched on Kickstarter and 4000 projects is not a representative sample size to explain trends related with Kickstarter. Example: According to studies the categories film and video, music, and games have raised the most money. This data set is telling us that technology have raised more money.

**What are some other possible tables and/or graphs that we could create?**

1. A table presenting the percentage of successful, failed, and canceled campaigns per category as counting the number of campaigns does not give me a clear idea of the success of a specific category. Same table can be created for sub-categories. A stacked column bar graph with percentage as the Y axis could be employed to represent the data. The Y axis is normalized to 100% and the bars are easier to understand.
2. Table presenting the total money raised by categories and subcategories. Two graphs can be create showing amount of money raised by categories or subcategories.
3. Tables exhibiting the number of backers per category and sub-category. Two stacked column graphs (successful, failed and canceled) could have been created to present the data visually.
4. X and Y graphs number of backers v.s. pledges. To determine if the number of backer correlate with the amount of money raised.
5. A line chart that graphs goal range in the x-axis and number of projects in the y-axis. Lines represent a different category. This graph will allow us to see how many projects per category falls within the set goal range.

**Bonus:**

As I explained above in conclusion, this data show that projects with a high monetary amount of funding goal tended to fail in reaching their goals in comparison to those projects that were gathering small amount of money.

**Bonus statistical analysis**

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

Since the variance is high this is an indication that the data points are very spread out from the mean. Both data set are left skewed. Relative standard deviations for successful and unsuccessful campaigns are 434% and 347% respectively. For those reasons, the median is a better representation of the data.

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

Variability is high as the standard deviation is higher than the mean. Relative Standard deviation for successful campaign is 434% showing more variability than unsuccessful campaigns (347%).

Because we are analyzing the total number of backers from different categories which are independent from each other and each project had a different set funding goal this variability is expected. Seventy two percent of the unsuccessful data corresponded to campaign with 0 to 19 backers whereas successful campaigns data where more scattered within 1 to 130 campaigns. It makes sense that there is more variability in the successful campaign data. Eliminating outliers probably would have make the data less variable.