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

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

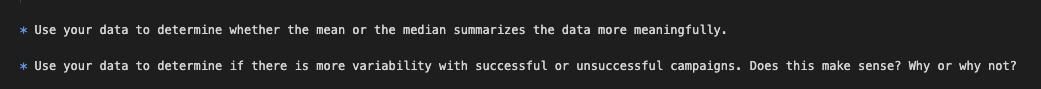
2. What are some limitations of this dataset?

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

Kickstarter Campaign Analysis:

1. Three conclusions about Kickstarter campaigns:
   1. The theater category (specifically, the sub-category “plays”) have the most overall campaigns, and the most successful campaigns, three times that of the next most successful sub-category (rock).
   2. The most successful campaigns are launched in May, while the least successful campaigns are launched in December.
   3. Based on historical trends, the campaign categories of technology, publishing, photography, games, and food have failed more often than they have been successful.
2. Limitations:
   1. This database (the raw data provided) contains information only for 4,114 past projects, when >300,000 projects have actually been launched on Kickstarter. So the data is a small sample of the entire portfolio of projects.
   2. 410 projects (~10% of the selected projects for data analysis) had 0 backers (no funding at all).
   3. A relatively large number of projects (3,038 of 4,114) were from the US, perhaps there could have been more data from other countries.
   4. Column H in the raw data shows the campaigns were in different currencies, when showing the average donation in column P, exchange rates to one target currency were not taken into account, so it would be difficult to compare average donations without first converting to one common currency.
3. Other possible tables/graphs:
   1. Average donation by category covering 2009-2017 (bar chart).
   2. Pie chart of successful campaigns, % for each country.
   3. Bar chart - average length of time for a campaign by category (find the difference in days between “Date Ended Conversion” and “Date Created Conversion”, then find the average over 8 years for each category).
   4. Scatter plot to show pledged vs. goal to analyze if there was a relationship between what the goal was and what was pledged.
   5. Tree map to show count of campaigns by country over the course of 8 years.
   6. Of the campaigns that were >100% funded, showing the average % over 100% by category using a bar chart, perhaps to highlight elevated interest in a particular category.

Bonus Questions Analysis:



1. When looking at the successful campaign data in terms of a frequency of distribution, the mean (194) is greater than the median (62), and the median (62) is greater than the mode (27). Thus, there is a positive (right) skew in the data, the “tail” is leading to the right, so the outliers are on the right side of the data.

Mean > Median > Mode.

Same for the failed campaigns data.

The more skewed the data, the greater the difference between median and mean, so the **median** would be preferred in this case to summarize the data (mean is preferred if there is a normal bell curve to the data, where mean = median).

1. There is more variability with successful campaigns because the variance and standard deviation are relatively large, indicating that the data points are very far from the mean and from one another, more than that of the failed campaigns. This makes sense, because in real life, Kickstarter campaigns are risky and success is dependent on a lot of factors such as interest in a particular category, time of year, perceived value to the consumer, etc.