**Excel Homework: Kickstart My Chart**

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

* Whether a campaign is successful, failed or cancelled is not dependant on the time of the year the campaign started, although an argument could be made about December being a poor time to start a campaign.
* Campaigns falling under ‘food’ or ‘games’ categories are most likely to fail (70% fail rate for food and 64% fail rate for games) whereas campaigns falling under ‘music’ are least likely to fail (17% fail rate).
* A few Sub-Categories that have had all campaigns fail include animation, children’s books, drama, and gadgets.

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

* It is unknown how much publicity each campaign was raised. This could potentially influence
  + State: more public campaigns would most likely tend to be more successful than less public ones
  + Backers Count: more people would have to see the campaign in order to increase the number of people contributing
  + Percent Funded: having a higher number of people contributing will increase the chance of reaching the goal amount
* It is unknown how realistic each campaign would be to finalise their project
  + Campaigns where their end goal is too unrealistic to achieve may have an effect over the amount of backing they get and ultimately the final state of their campaign
* No information included as to why some campaigns were cancelled
* Average Donation column: only having an average donation amount could affect some analysis

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

* A table/graph to compare
  + Country vs State – compare countries based on the amount of successful, failed and cancelled campaigns
  + Backers Amount vs Category – see which categories are popular among backers
  + Backers Amount vs Percent Funded – see whether percent funded is dependent on the amount of backers
  + Goal Amount vs State – see whether the state of the campaign is dependent on the starting goal
  + Time Elapsed vs State – see whether how long the campaign was active determines result

**Bonus Statistical Analysis**

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

* With this dataset, the median would summarise the data more meaningfully
  + As the maximum is significantly greater than the mean, the mean can be skewed in a way that would not correctly represent the data as well as it could (as the mean is calculated using the summation of the backers count, a large maximum count, or several large amounts, would increase the mean)
  + The median would not be as affected by the large maximum or larger outliers, hence why the median would summarise the data better

**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 in the successful campaigns then the unsuccessful campaigns
  + Successful Variance = 713 167
  + Unsuccessful Variance = 3 776
* This does make sense as the difference between the minimum and maximum number of backers is greater in the successful campaigns than the unsuccessful campaigns, meaning there is more variability in the data
* The greater variance in successful than unsuccessful campaigns can also be attributed to the fact that unsuccessful campaigns are limited by a percent funded outcome being less than 100%, meaning they will generally have a smaller backer count (assuming having more backers will increase the likelihood of a successful campaign)