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

* Approximately 53% of the campaigns are successful, 37% have failed, 8% have cancelled and 2% are still live.
* Majority of the campaigns are in the Theater Category and Plays Sub-Category.
* May is the peak season for the successful campaigns.

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

* **Manual and Tedious Functionality**
  + Data must be entered manually which is prone to errors. It is quite easy to change the contents of the cell by copy and pasting the wrong information and entering the wrong contents for the specific cell. Such errors may distort the results for data analysis that leads to bad business decisions.
  + Excel does not sort data automatically, need to carefully select data to sort data. For example, if you select only one column to sort data, it will re-arrange the order for that column, but the data on the other columns do not gets solved.
  + As you have more data, it may be time consuming and tedious to manually input the data.
* **Low Coverage of Operations**
  + Number of rows in the spreadsheet are limited so if there are more data for the coming years, excel cannot hold all the data.
  + Operations of filtered datasets are limited and requires many steps to filter the exact information you need.
  + Limited formulae for data analysis. Formulae such as MAXIF AND MINIF are not available.
  + Limited available memories.
  + As there are more data, Excel slows down when performing actions like filtering, sorting, and creating pivot tables.
* **Lack of Data Integrity**
  + Using Excel as the primary database in a workplace lacks the data integrity as there are countless errors in manual inputs, copy and pasting, flawed formula, and various inputs from various users.
* **No Audit Trail**
  + When an error occurs, there is no audit trail of who inputted, deleted, and/or made changes to the document. This will make it challenging to determine where the root cause of the error is. In some cases, if you have large sets of data, it will be time consuming to find which cell contains the errors.
* **Requires Knowledge in Excel when Performing Data Analysis.** 
  + For the people who does not have prior experience in using Excel, they may need to learn all the function, syntax and formula of Excel before inputting data and conducting data analysis. This may take some time for the individual to understand.

1. **What are some other possible tables and/or graphs that we could create?**
   * Average Donation per Category and Sub-Category
   * % Funded per Category and Sub-Category
   * % Funded by Country
   * Average Donations by Country
   * Launch Month by Country
   * # of Days of Campaign by Category and Sub-Category

**Bonus Statistical Analysis**

|  |  |  |
| --- | --- | --- |
|  | **Successful** | **Failed** |
| **Mean** | 194.43 | 17.71 |
| **Median** | 62.00 | 4.00 |
| **Minimum Value** | 1.00 | 0.00 |
| **Maximum Value** | 26457.00 | 1293.00 |
| **Variance (Population)** | 712840.99 | 3773.22 |
| **Standard Deviation (Population)** | 844.30 | 61.43 |
| **Variance (Sample)** | 713167.38 | 3775.69 |
| **Standard Deviation (Sample)** | 844.49 | 61.45 |

1. **Use your data to determine whether the mean or the median summarizes the data more meaningfully.**
   * The mean is the average number of backers for the campaigns. This mean figure from the above table will determine how many backers I will usually have for the campaign if it is a success or a failure.
   * The median is the middle figure of all the samples in the successful and failed campaigns.
   * These figures are useful because it tells us what is normal in each situation. For example, the organizers will know if the project is a success or failure given how many backers they receive for the project.
   * It is helpful in determining one set of data against the entire set of data. It benchmarks on how well the campaign is doing against the perceived average.
2. **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 campaigns compared to the failed campaigns and this makes sense because:
     + The range of value is a lot higher in the successful campaign are hence becomes more spread out from the mean.
       - The range of the successful campaign is 1-26457 backers to a mean of 194.43.
       - The range of the failed campaign is 0 to 1293 backers to a mean of 17.71.