# Purpose

This report is being submitted as Homework assignment 1 of the Rice Data Analytics Bootcamp. The general purpose of the assignment was to analyze the data of 4,000 past Kistarter projects. An Excel file with the data to be analyzed was provided. After utilizing various methods available within Excel the following questions were to be answered.

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

### What are some limitations of this dataset?

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

# Process

After downloading the Excel file the file was reviewed and a plan for cleaning of the data was established. The following steps were taken to better organize and present the data.

* Column headings were capitalized
* Excel’s PROPER() Function was utilized to capitalize text in the Outcome, Category and Sub-Category Columns.
* Text to Columns function was used to separate the Category and Sub-Category column into two columns (Parent Category, Sub-Category)
* Date conversation was utilized to translate the date stamps provided into a usable format for later data analysis.

# Results

## The following sheets were generated within the Excel File and can be viewed.

### Outcome\_Category\_Pivot - Pivot table summarizing project outcome by category and stacked column pivot chart based on this table.

### Outcome\_Sub\_Cat\_Pivot – Pivot table summarizing project outcome by sub-category and stacked column pivot chart based on the table.

### Date\_Created\_Pivot – Pivot table summarizing the date a project was created and the project outcome and a line graph based on the table.

## Questions Answered

### Given the provided data, three conclusion that can be draw are as follows.

* Projects in the parent category Music and the sub-category Rock within this parent category appear to have the highest level of success.
* Projects started in May have the highest level of success.
* Projects in the parent category Food and the sub-category Food Trucks within this parent category appear to have the highest level of failure.

### The dataset provided is sample of 4,000 projects out of the stated 300,000 projects launched on Kickstarter, a more complete analysis would result by having the full 300,000 dataset.

### Additional tables and graphs that can be generated would be tables based on the following.

* Staff picks versus outcome
* Spotlight versus outcome

# Bonus Section

## Further analysis was conducted by producing additional sheets as follows.

### Bonus – This sheet is made up of a data table showing the Goal amount and the corresponding outcome count. Percentage of outcomes was then calculated and a line graph illustrating this data was generated.

### Bonus\_Stat – This sheet is made up of a data table the Backer count of each project separated into Successful and Failed. Using Excel MEAN, MEDIAN, MIN, MAX, VAR.P, and STDEV.P were utilized to generate a table showing the outcomes of these functions. In addition a boxplot was generated to better visualize the datset, it should be noted that due to the wide variance of data, the y-axis max value was limited in order to better visualize the IQR and where the mean value was. The previously mentioned function were used to answer the following questions in a table provided on this sheet.

* The mean number of backers.
* The median number of backers.
* The minimum number of backers.
* The maximum number of backers.
* The variance of the number of backers.
* The standard deviation of the number of backers.

In both cases it is determined that the median of backer better summarizes the data more meaningfully.

### Bonus (USD) – It was noticed that the goal and pledge data given were in various international currencies so additional columns were added to the original dataset converting the currencies to USD using exchange rate data on the sheet Exchange\_Rate\_Data. The dates utilized for Goal Amount are the Date Created and for the Pledge Amount the Date Ended was used. This chart and Graph represents the same data that was requested on the Bonus sheet but in currency normalized to USD.

### Avg\_Donation – It was decided to look at the Average Donation amount, for the sheet the normalized to USD amounts were utilized. A Line Graph visualizing the data was generate and appears to suggest that a project with an average donation amount of between $40-100 have a higher rate of success.