**Title:** Module 1 Challenge - Excel - Week 1

**Course**: Data Analytics Boot Camp

**Author**: Dr Murali Chidambaram Veerabahu

# Introduction

This ‘ReadMe’ file is the solves the assignment for Module 1 of the Data Analytics Bootcamp on Excel. This should be read in conjunction with Excel file ‘CrowdfundingBook-Solved’ submitted with this file.

The assignment requirements has been listed below and explanation has been provided on how each question has been addressed. The assignment questions are detailed in blue and the answers below the questions. The written report for this assignment is also included below with the appropriate sub heading. Please note campaigns have also been referred to as Projects in this assignment report.

# Assignment Requirements

## Conditional Formatting (10 points)

### Conditional formatting is applied appropriately to the outcome column (5 points)

Four colours have been used to show the different outcomes using conditional formatting.

* “Successful” is shown in Green
* “Failed” is shown in Red
* “Live” is shown in Yellow
* “Cancelled” is shown in Grey

### Conditional formatting is applied appropriately to the percent funded column (5 points)

Three Colour Scale was used to Conditionally format “Percent Funded” column with Dark Red at 0%, Dark Green at 100% and Dark Blue at 200%

## Column Creation (10 points)

### Six new columns were correctly created for:

* ***percent funded*** - “Percent Funded created left of “Outcome” Column by dividing “Pledged” by “Goal” and formatting for Percentage.
* ***average donation*** *-* “Average Funded” was created right of the “backers\_count”. The formula =IFERROR was used to remove the indivisible error of zero divided by zero.
* ***category*** *-*  The category and sub-category is differentiated using “/“ in the data. The location of the “/“ was found using the =SEARCH function. Knowing the differentiation character point the =LEFT function was used to extract the category.
* ***sub-category*** *-* The differentiation point was identified using the =SEARCH function using functions =LEN and =RIGHT the “Sub-category” was extracted.
* **Date Created Conversion** - The Unix Date in “Launched at” field was converted to normal date to the field “Date Created Conversion” using the formula =(((<UnixDate>/60)/60/24) + DATE(1970,1,1) 1
* **Date Ended Conversion** – The Unix Date in “deadline” field was converted to normal date to the field “Date Ended Conversion” using the formula =(((<UnixDate>/60)/60/24) + DATE(1970,1,1) 1

## Pivot Tables and Stacked Column Charts (15 points)

### Correctly created a pivot table that counts how many campaigns were "successful," "failed," "canceled," or are currently "live" per category (7.5 points)

The pivot table showing outcome counts by category is shown in worksheet “Outcome by Category”. The field of category was placed in ‘Rows’ and the field of outcome was placed in both ‘Values’ and ‘Columns’ in the Pivot Table fields.

### Correctly created a stacked column pivot chart that can be filtered by country (7.5 points)

A stacked column chart was created to the right of the pivot table in the “Outcome by Category” worksheet.

## Pivot Tables and Line Graphs (15 points)

### Correctly created a pivot table with a column of outcome, rows of Date Created Conversion, values based on the count of outcome, and filters based on parent category and Years (7.5 points)

The Pivot Table was created in the ‘Outcome by Time’ Worksheet. The Campaigns were divided by month created with an option to be filtered by category and year.

### Correctly created a pivot chart line graph (7.5 points)

A Line Chart was created using the pivot table in ‘Outcome by Time’ to display the tabular information in a graphical form.

## Written Report (20 points)

### Presents a cohesive written analysis that:

### Draws three conclusions from the data (10 points)

### States limitations of the dataset and suggestions for additional tables of graph (10 points)

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

1. Crowd sourced projects have just over 50% chance of successfully funded and majority of projects stem from the United States. (See Stacked Bar Chart in “Outcome by Category” Worksheet). This maybe because many crowd source platforms may have developed in the US and probably is a more popular way to fund projects.
2. The largest sub-category is for plays under the parent category of theatre. More than one third of all projects are plays and they have just over 50% chance of being successfully funded. (See Stacked Bar Chart in “Outcome by Sub-category” worksheet). This perhaps reveals the poor mainstream funding available for the Arts sector and artists have turned to alternative funds like crowd sourcing to fulfil their dreams..
3. Without any detailed analysis but looking at the trends over time (See Line Graph in “Outcome over Time”) projects which seem to have a high success are journalism (4 of 4) and technology (64 of 96) albeit the number of journalism projects were very small.

*What are some limitations of this dataset?*

We are defining success as - a project which has been able to fully fund itself, however a projects delivery should also be considered. Meaning did the project group deliver the product or service in the specified time. The credibility of the project owners would also be a useful element to capture, meaning if they have successfully completed previous projects. More information about the backers would also be useful, so certain projects which may be of interest to the backer may be canvased individually to increase successful funding success.

*What are some other possible tables and/or graphs that we could create, and what additional value would they provide?*

It would be good to understand outcome of sub-categories over the different years. We are able to see outcome in general over each year in the ‘Outcome by Time’ worksheet, however it would be more useful to understand a trend of successful sub-categories over the many years. A graph depicting the success and failure of campaigns by sub category over the 11 years of data would be useful to see. The macro-economic effects on funding can be studied using this data, as many projects may have failed because the country may have gone through an economic recession or an anomaly like COVID.

An analysis of the funding goal of projects and their success rates would be an interesting area to look at. This information may help manage the expectations of project owners and how much they could possibly raise form the crowd source platform. Eg: The analysis of historical data may say that theatrical plays with a funding goal of over $100,000 was rarely successful, while $50,000 has been. This can help project owners plan their funding strategy.

The duration in which the project was open in the crowd source platform and its success rate. This will help to know how long a project should typically be kept open for on the crowdsourcing platform to optimise success.

## Crowdfunding Goal Analysis (10 points)

### Computed calculations of percentages for projects that were successful, failed, or were canceled per goal range (5 points)

The table with the specified goal ranges was created in the ‘Goal Analysis’ Worksheet. The relevant values were calculated using the =COUNTIFS function and these values were used to calculate the percentages.

### Created a line chart showing the relationship between the goal’s amount and its chances at success, failure, or cancellation (5 points)

A Line Chart was created under the table in the ‘Goal Analysis’ worksheet to graphically represent the tabular data.

## Statistical Analysis (20 points)

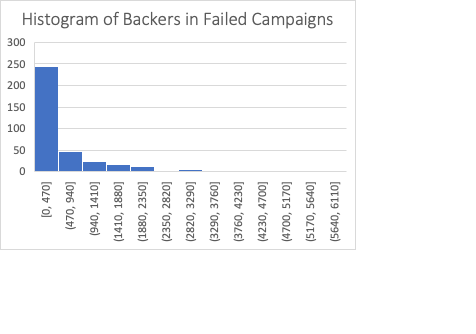
### Computed calculations of the mean, median, min, max, variance, and stdev using Excel formulas (15 points)

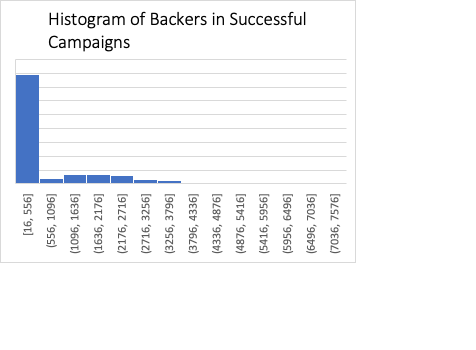
The data table for analysis was created in the new worksheet called ‘Backers Analysis’. The table was created copying filtered data from the main data table in the ‘Crowdfunding’ worksheet. The functions used for the calculations were =AVERAGE, =MEDIAN, =MIN, =MAX, =VAR & =STDEV

### A brief and compelling justification of whether the mean or median better summarizes the data (5 points)

The mean is much greater than the median for both successful and failed campaigns. The minimum and maximum are also show a large range. To graphically show this a histogram and box plot for both successful and failed campaigns has been shown in the excel worksheet ‘backers Analysis’ (Copied below). The histogram shows skewed data to the right. In the Box and Whisker plot the Interquartile range is also closer to the minimum. All of this shows that the data is not normally distributed and hence the median would provide a better summary of the data.

To answer the question around variability between successful and failed campaigns, both the variance and standard deviation for successful campaigns is greater compared to failed campaigns. While both successful and unsuccessful campaigns have outliers looking at the box and whisker plot over all what can be determined form all the data in ‘Backers Analysis’ worksheet is that there are more backers in successful campaigns compared to unsuccessful ones. This would be self-explanatory assuming that the average donation by each backer in both successful and unsuccessful groups is similar.





A graph of a box plot

Description automatically generated

**Reference**

1. Productivity Boosters. How to convert between date and Unix timestamp in Excel? *How to convert between date and Unix timestamp in Excel?* https://www.extendoffice.com/documents/excel/2473-excel-timestamp-to-date.html.