In this project I had been tasked with Data collection, Cleaning, Transformation and Visualization. Below I have explained what I did in each task and completed my project (Client Brief Has Been Attached in the Evaluation Portfolio Folder):  
  
**Data Collection:**

The client has sent through:

* 7 data sets - each data set contains different columns and values
* A data model - this shows the relationships between all of the data sets, as well as any links that you can use to merge tables.

I Collected the Data based on:

* The brief carefully states that the client wanted to see “An analysis of their content categories showing the top 5 categories with the largest popularity”.
* As explained in the data model, popularity is quantified by the “Score” given to each reaction type.
* I therefore need data showing the content ID, category, content type, reaction type, and reaction score.
* So, to figure out popularity, I’ll have to add up which content categories have the largest score.

**Cleaning of datasets:**

* Removing rows that have values which are missing.
* Changing the data type of some values within a column, and
* Removing columns which are not relevant to this task.

**Transformation:**

Create a final data set by merging your three tables together

* By using the Reaction table as base table, then joining the relevant columns from the Content data set, and then the Reaction Types data set.
* This was done by using the “VLookUp” formula.

Figure out the Top 5 performing categories

* Add up the total scores for each category.
* This was done by using the “SumIf” formula.

**Data Visualization:**

* By figuring out the top 5 best performing categories. We can visualize the Aggregated Score by the top 5 best performing categories:

