The model dataset after merging tables is provided in the ModelAnswerFullDataset.csv file. Note, that some of the column names have been renamed for ease of understanding.

This dataset combined three datasets:

- Reaction
- Content
- Reaction Types

By reading the brief carefully it states that the client wanted to see "An analysis of their content categories showing the top 5 categories with the largest aggregate popularity".

This meant that the client wanted to know which categories of their content had yielded the greatest popularity out of all their content. But how do we quantify popularity? This was explained in the data model.

Popularity is quantified by the "Score" given to each reaction type, as a numeric value.
Therefore, each reaction gives a weighting to how popular a piece of content may become.

To find the categories with the greatest popularity, the data analyst must sum up which content categories have the largest aggregate Score.

Note, to achieve the correct answer the data must have been properly cleaned!

To create the model dataset, you want to start with the Reaction table as your base table. This table shows all of the reactions to particular content IDs. To find out the category of these pieces of content that have been reacted to, we must merge values of the Content dataset to the Reaction table using "VLOOKUP" formula and performing the same action for joining the values of Reaction Type dataset to show the meaningful connection.

To clean things up and to make your data more understandable, we changed some column names and remove unnecessary columns.

To finally calculate the top 5 categories with respect to popularity, we used "SUM IF" formula to add up the scores for each category. After this, sort your data by Score in descending order and take the top 5.

Finally, we get the end result containing the spreadsheet which contains:

1. A cleaned dataset
2. The top 5 categories.