

The ModelAnswerFullDataset.csv file contains the merged dataset obtained after consolidating multiple tables. It's important to note that some column names have been modified for enhanced comprehension.

This composite dataset stems from the integration of three primary datasets: Reaction, Content, and Reaction Types.

Upon careful examination of the client's request, it was discerned that they sought an "analysis of their content categories showcasing the top 5 categories with the highest cumulative popularity". In essence, the client aimed to ascertain which categories of their content garnered the most significant popularity among all their content. The metric for popularity was elucidated in the data model.

Popularity is quantified through the "Score" assigned to each type of reaction, denoted as a numerical value. Consequently, each reaction contributes a weightage to the potential popularity of a piece of content. To identify the categories with the highest popularity, the data analyst must aggregate the Scores for each content category.

It's crucial to underscore that for an accurate outcome, the data must have undergone thorough cleaning and preparation.

To construct the model dataset, the Reaction table serves as the foundational table. This table elucidates all reactions corresponding to specific content IDs. To ascertain the category of the content that has been reacted to, a left join is performed between the Content table and the Reaction table, merging on the "Content ID" column. This resulting table is denoted as "A".

Subsequently, the Reaction Types table is merged with the "A" table. This facilitates the inclusion of the popularity score for each piece of content based on the type of reaction. This merger can be executed by joining the Reaction Types table onto "A" as a left join, using the column that describes the reaction "Type". This culminates in the final dataset, designated as "B".

For the sake of clarity and improved data comprehensibility, certain column names can be modified, and extraneous columns can be removed.

To ultimately determine the top 5 categories in terms of popularity, these transactions must be grouped by Category, and the Scores for each Category should be aggregated. Following this, the data should be sorted by Score in descending order, and the top 5 categories should be selected.