

Data Cleaning Documentation:

1. Merging and Location Field Update
 - Added the *Location* field from Dataset1-Brightspace to Dataset1-Modified.
 - Used *gross_cost* from both datasets as a matching reference:
 - Sorted *gross_cost* in both datasets.
 - Copied matched rows into a new worksheet.
 - Implemented conditional checks to validate that *gross_cost* values aligned row-by-row.
 - Renamed the *Location* field to *City* for consistency.
2. Data Filtering
 - Filtered and removed rows where *gross_cost* was marked as N/A.
 - Renamed 'Tri-Cities, Tennessee-Virginia' to 'Tennessee' due to lack of data for the Tennessee region.
 - Removed rows with zero clicks but non-zero conversions.
3. Column Splitting
 - Used Text to Columns to split the *creative_size.userid* field into two separate columns:
 - Creative_size
 - user.id
4. Data Standardization
 - Applied the TRIM function in Excel to the newly split *user.id* column to remove leading and trailing spaces.
 - Designated *user.id* as the primary key for both Dataset 1 and Dataset 2 to enable consistent merging and reference.
5. Character Cleaning
 - Used Find & Select to replace all special characters:
 - Replaced all occurrences of accented characters (e.g., Â) and guillemets (») with appropriate standard characters.
 - Replaced '> ' (with space) with '>' to maintain delimiter consistency.
6. Audience Segment Parsing
 - Used Text to Columns with '>' as the delimiter to split *audience_segment* into multiple distinct columns for better segmentation analysis.
7. Missing Value Handling
 - Encapsulated all calculation formulas with IFERROR to mitigate DIV errors