Data Cleanup Process

Nick (Dataset 1)
- Separated creative_size from user_id via a SPLIT function.
    - *I used the " | " as the delimiter to separate the two to separate columns*

- Renamed certain columns to give more specific information
    - *creative_size → creative_size_px*
    - *gross_cost → gross_cost_USD*

- Changed number formatting for gross_cost_USD to reflect dollar amounts
    - ***Note: If you're using SQL, you need to take out the "$" symbol because importing the dataset as is will treat the values as strings, not integers.***

- Checked for duplicates on campaign_id via conditional formatting.
    - *campaign_id is a unique primary key that defines each ad campaign, so it cannot be allowed to repeat.*

- Checked for blanks and removed the rows with them.
    - *The blanks didn't have enough content or relevant information for our analysis. For example, there'd be missing KPIs like impressions in the original dataset.*

- Checked campaign_id if they were consistent via LEN function.
    - *I had to make sure that we were consistent with the labeling*

- Applied filters onto the column headers and skimmed across the values to identify if there were weird spelling errors
    - *There weren't any misspellings for any of the columns, so I didn't change anything.*

- Replaced "Â»" with "-" from audience_segment with Find & Replace