

Data Cleaning

campaign_export_csv

- Using the split column function in the home tab, I split the campaign column by each occurrence of the delimiter “_”. This split the column into 12 new columns, then canceled the ‘changed type’ step which automatically detects datatypes.
- After that, I renamed the newly created columns to match those in the marketing_BI_mapping_table_v2.xlsx.
- I made sure all columns have the correct data type, and changed spend from the whole number data type to fixed decimal place.
- I ensured that the columns did not contain null, or empty cell values by using column profiling, column distribution, and column quality functions in the view tab.
- For the campaign_creation_date column, I divided it into 3 columns year, month, and day using the extract function in the add column tab. This is because it was stored in the format “ddmmyyy” which could not be converted directly to date.
- Using the merge column in the transform tab I joined the 3 newly created columns separating each by the delimiter “-”.
- I converted the new joined column from datatype text to datatype date. Then replaced this as the campaign_creation_date column and remove the other from the dataset.

transaction_csv and website_tracking_checkout_csv

- I filtered the medium column in the website_tracking_checkout table to contain only roles whose medium is ads. This resulted in 366 rows.
- Using merge queries in the home tab I merged the transaction and website_tracking_checkout into a new table named transaction_by_ads.
- This I did by using inner join to join both tables referencing the transaction_ID column in both. This gave a total of 366 rows.
- I did this to select just transactions that occurred due to the ads run by XYZ Limited.
- Using the split column function in the home tab, I split the campaign column by each occurrence of the delimiter “_”. This split the column into 12 new columns, then canceled the ‘changed type’ step which automatically detects datatypes.
- After that, I renamed the newly created columns to match those in the marketing_BI_mapping_table_v2.xlsx.
- I made sure all columns have the correct data type.
- Ensured that the columns did not contain null, or empty cell values by using column profiling, column distribution, and column quality functions in the view tab.

- For the campaign_creation_date column, I divided it into 3 columns year, month, and day using the extract function in the add column tab. This is because it was stored in the format “ddmmyyy” which could not be converted directly to date.
- Using the merge column function in the transform tab I joined the 3 newly created columns separating each by the delimiter “-”.
- I converted the new joined column from datatype text to datatype date. Then replace this as the campaign_creation_date column and remove the other from the dataset.
- I changed the date_num column to a text datatype and then to a date datatype.
- I changed customer_id and is_chinese_language columns to a text datatype
- I changed revenue and cogs columns to fixed decimal place datatype.
- This gave me 29 columns and 366 rows.

website_tracking_visits

- I filtered the medium column in the table to contain only roles whose medium is ads.
- Using the split column function in the home tab, I split the campaign column by each occurrence of the delimiter “_”. This split the column into 12 new columns, then canceled the ‘changed type’ step which automatically detects datatypes.
- After that, I renamed the newly created columns to match those in the marketing_BI_mapping_table_v2.xlsx.
- I ensured that the columns did not contain null, or empty cell values by using column profiling, column distribution, and column quality functions in the view tab.
- For the campaign_creation_date column, I divided it into 3 columns year, month, and day using the extract function in the add column tab. This is because it was stored in the format “ddmmyyy” which could not be converted directly to date.
- Using the merge column in the transform tab to join the 3 newly created columns separating each by the delimiter “-”.
- I converted the new joined column from datatype text to datatype date. Then replace this as the campaign_creation_date column and remove the other from the dataset.
- I ensured all columns had the correct datatype.