

BigQuery/Dataform Developer for Ecommerce

4th Question. Provide a short overview on how you verified the resulting tables from tasks 1-3 produce correct data.

1. For question 1 - ga4_stg01_sessions

- a. Verified that the number of rows in the table matched the unique session_id values in the source dataset. Ensuring that each session_id appears exactly once in the resulting table. Compare row counts between the raw data and the resulting table to ensure no data loss.
- b. Checked that fields like session_start_timestamp and session_end_timestamp accurately reflected the first and last event timestamps for each session. Verified that all required fields (session_id, user_pseudo_id, device, date, etc.) are correctly populated.
- c. Also checked if the is_session_engaged flag is set correctly by verifying that sessions meet at least one engagement condition
- d. Calculated each aggregate fields like number of sessions per user, average session duration, etc., with result table values
- e. Compared values for fields like source, medium, and campaign against the raw event_params fields from raw table to confirm accuracy.

2. For question 2 - ga4_stg02_pages

- a. Checked that each row in the resulting table corresponding to a unique combination of date and page_url from the raw table and ensuring that the resulting table contains all unique page URLs and dates present from the raw data table.
- b. Cross checked the total_time_on_page by calculating engagement_time_msec from raw data for a few random sample data sets.
- c. Manually validated entrances column values from the result table by counting the occurrences of page_view events with entrance parameters and also manually checked the exits column values for few sample datasets

3. For question 3 - shopify_orderlines_stg02_with_collection

- a. Checked all product_id values were mapped to collections correctly by joining with the 3 tables source_shopify_collects, source_shopify_custom_collections, and source_shopify_smart_collection and ensured that each order line appears exactly once in the resulting table and that multiple collections were concatenated in alphabetical order using STRING_AGG function and also manually checking the titles and orderlines names manually randomly with the raw data tables for few samples.
- b. Also thoroughly checked the resulting table maintained the same number of rows as source_shopify_orderlines_stg01 table
- c. Checked the randomly selecting any order_id details with the raw table that all the fields like date, product_id, collections, order_line_names match with their respective raw tables.