

SQL Assessment Summary Report

Isaac Hafdallah
Data science Working Student Position at KoRo

May 31, 2025

1 Environment Setup

All data loading and querying were implemented in Google Cloud Platform's BigQuery:

- BigQuery UI and bq CLI for dataset and table management
- Google Cloud Storage used to stage CSV files
- Standard SQL (BigQuery) for all transformations and analyses

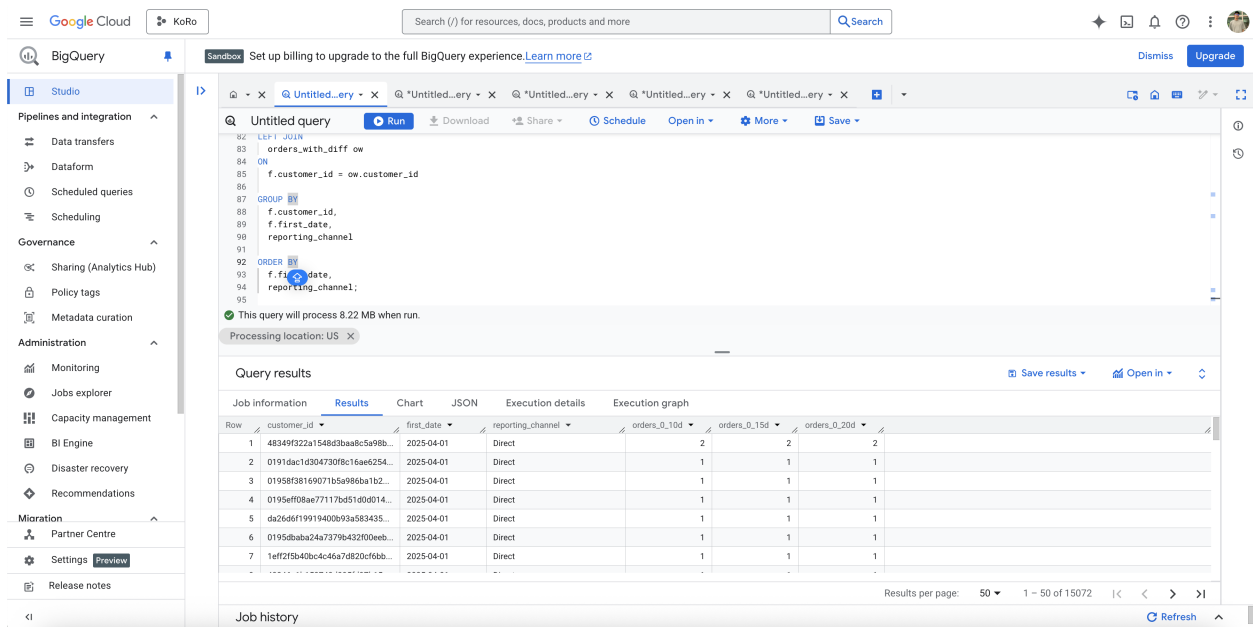


Figure 1: Google Cloud Environment Setup

2 Data-Cleaning Tips

- Filter null or malformed values
- Convert dates safely
- Remove exact duplicates
- Bucket missing values
- Join tables for enriched context

3 Task 1: Customer Order Activity in Early Lifecycle

3.1 Approach

- Cleaned and de-duplicated the `orders` table, filtering out nulls.
- Computed each customer's first purchase date
- Joined back to calculate `days_since_first` and aggregated counts in 0–10, 0–15, and 0–20 day windows
- Joined in `marketing_sources` to break down results by `reporting_channel`, labeling missing channels as “Unknown.”

3.2 Findings

- The vast majority of repeat orders occur within the first 10 days
- Channels “Direct” and “Email” drove the stickiest early activity, with majority of their orders in the 0–10 day window.

4 Task 2: Product Popularity by Country

4.1 Approach

- Cleaned and de-duplicated `orders`, then joined to `product_universal` to map each SKU to its main category
- Aggregated total orders and distinct SKU counts
- Computed order counts per product and top 5 best-sellers.
- Computed the bottom 5 least popular SKUs.

4.2 Findings

- The “Beverages” category led in total orders across most countries, followed by “Snacks”.
- The top 5 products per country accounted for 20–30% of that country's total orders.
- The bottom 5 least-ordered products often logged only a single sale

5 Task 3: First Orders and New Customer Share by Day

5.1 Approach

- Cleaned and deduplicated `orders`, parsed dates
- Joined marketing data to assign `reporting_channel`, defaulting nulls to “Unknown.”
- Flag each customer's first order.
- Aggregated daily metrics
- Unioned overall daily results with a breakdown by `reporting_channel`.

5.2 Findings

- Daily total orders around 400; new-customer orders comprised around 90% of volume on most days.
- On April 1, 2025, majority of Direct orders and Email orders were first-time purchases