

Lab 1 – Task 2: Identify Duplicate Rows

Using `COUNT(DISTINCT name)`

Objective:

Use SQL in BigQuery to determine:

- The **total number of rows** in the products table.
 - The **number of unique product names** using `COUNT(DISTINCT name)`.
-

Steps to Complete the Task:

1. Open a **new query tab** in BigQuery.
2. Paste and run the following SQL query:

SQL

SELECT

`COUNT(*) AS NumberOfRows,`

`COUNT(DISTINCT name) AS NumberofProducts`

`FROM `thelook_ecommerce.products`;`

3. Click **Run** to execute the query.
-

Correct Answers:

Question 1: How many distinct products are returned in total?

-  15010

Question 2: Why is the total in the `NumberOfRows` column different from the total in the `NumberofProducts` column?

-  The count of products has a **DISTINCT** clause or statement.

Feedback if Completed Correctly:

✅ Well done! You correctly used `COUNT(DISTINCT name)` to identify unique product names. This helps detect **duplicate entries**, which is crucial for ensuring data accuracy in analysis.

Feedback if Not Completed Correctly:

⚠️ Double-check your query syntax. Make sure you're using:

- `COUNT(*)` for total rows
 - `COUNT(DISTINCT name)` for unique product names Also, ensure you're querying the correct table: `thelook_ecommerce.products`.
-