Quality Control Form

Team Name	CROSS_JOIN
Team Members	Hansel, Gus, Konrad, Tye, Thomas
Client Name	Mystic Manuscripts
Date of file received	22/09/2025
Name of file received	Books Sales Records – V2
File Format received	.csv
Size of file (KB) received	931kb
Encoding of the file	utf-8
Recorded Number of Columns	36
Recorded Number of Rows	3480
Name of Schema of Destination Table	**All_2509**
Name of Destination Table	**cross_join**

ID of SQL or Checked Rows	Description of Check	Result in SOURCE	Result in DESTINATION	OUTCOME
SQL-1	**Count of rows**	3480	3480	
SQL-2	**Count of Distinct Rows**	3480	3480	
SQL-3	**Count of Columns**	36	36	
SQL-4	**Sum of column Sums**	11680962.79	11680963.46	-2.83e-10% difference due to rounding errors in different programs
SQL-5	**Sum of Row Sums**	11680962.79	11680963.46	-2.83e-10% difference due to rounding errors in different programs
SQL-6	**Date format Check**	ID 564 Date 4/21/2018 ID 2516 Date 7/16/2019 ID 1899 Date 5/19/2019	ID 564 Date 4/21/2018 ID 2516 Date 7/16/2019 ID 1899 Date 5/19/2019	

		ID 3174 Date 9/28/2019 ID 993 Date 7/14/2018	ID 3174 Date 9/28/2019 ID 993 Date 7/14/2018	
SQL-7	**Eyeball Check**	1667 2048 2416 434 88	1667 2048 2416 434 88	
SQL-8	Customer with most total expenditure	Customer ID = B07FWWBYQ Total spent = 7300 INR	Customer ID = B07FWWBYQ Total spent = 7300 INR	
SQL-9	Average profit percentage per sale	16.66	16.66	
SQL-10	Count of Nulls	0	0	
SQL-11	Number of morning purchases	1107	1107	
SQL-12	Numbers of sales per weekday	3 572 7 525 4 515 1 496 6 467 5 458 2 447	3 572 7 525 4 515 1 496 6 467 5 458 2 447	

For Query 6, Date Type formatting in PostgeSQL for importing and storing is YYYY/MM/DD. Data type for Date has been changed to VarChar to stay consistent with the original CSV.

Results in DESTINATION use column names from original source, changes in DESTINATION column names as follows:

ID -> id

Purchase-date -> purchase_date

Date -> ddmmyyyy

Time -> time_of_purchase

Quarter -> quarter_of_purchase

Qtr -> qtr

Year -> year_purchase

Month -> month_of_purchase

Date_1 -> day_of_month

Special Day -> special_day

Online Sale Offers -> online_sale_offers

Day -> day_of_week

Weekend -> weekend

Morning -> morning

Afternoon -> afternoon

Evening -> evening

Night -> night

Gender -> isfemale

Customer ID -> customer_id

Gender F=1665 M=1815 -> gender_fm

Product-Name -> product_name

Item-Status -> item_status

Quantity -> quantity

Currency -> currency

Item Price -> item_price

Shipping-Price -> shipping_price

Ship-City -> ship_city

Ship-State -> ship_state

Ship-Postal-Code -> ship_postal_code

Category -> category

Total_ amount -> total_amount

Author -> author

Publication -> publication_name

Profit Percentage (%) -> profit_percentage

Profit (INR) -> profit_inr

Cost Price -> cost_price

These changes were made to match the standard formatting typically used in SQLpostgres for best practice.

Column Format Discrepancies

'Date'/'ddmmyyyy' – Inconsistent format for the month, some single digit observations contain a 0 before while others do not.

'Product-Name'/'product_name' - No consistent format for titles or capitalisation. Occasional inconsistent use of abbreviations.

'Ship-City'/'ship_city' - Frequent inconsistent capitalisation.

'Ship-State'/'ship_state' - Frequent inconsistent capitalisation.

'Author'/'author' - No consistent format for author name, inconsistent capitalisation.

'Publication'/'publication' - No consistent format for publication, inconsistent capitalisation.