# Quality Control Form

| Team Name | Select |
|---|---|
| Team Members | Bekim Gerguri, Ewen Henderson, Tom Hill, Cipri Sebastiao, Aiden Tempest |
| Client Name | Mystic Manuscripts |
| Date of file received | 22/9/2025 |
| Name of file received | Book Sales Records - V2 |
| File Format received | csv |
| Size of file (KB) received | 946 |
| Encoding of the file | ISO-8859 |
| Recorded Number of Columns | 36 |
| Recorded Number of Rows | 3481 (including headers) |
| Name of Schema of Destination Table | all_2509 |
| Name of Destination Table | select_test |

| ID of SQL or Checked Rows | Description of Check | Result in SOURCE | Result in DESTINATION | OUTCOME |
|---|---|---|---|---|
| SQL-1 | Count of rows | 3480 | 3480 | *PASS* |
| SQL-2 | Count of Distinct Rows | 3480 | 3480 | *PASS* |
| SQL-3 | Count of Columns | 36 | 36 | *PASS* |
| SQL-4 | Sum of column Sums | 11680962.79 | 11680962.79 | *PASS* |
| SQL-5 | Sum of Row Sums | 11680962.79 | 11680962.79 | *PASS* |
| SQL-6 | Date format Check<br><br>*Compare 5 randomly chosen values from Source, check date matches in Destination.* | **614**: 4/30/2018<br>**1223**: 1/19/2019<br>**1749**: 5/7/2019<br>**2174**: 6/8/2019<br>**3098**: 9/11/2019 | **614**: 04/30/2018<br>**1223**: 01/19/2019<br>**1749**: 05/07/2019<br>**2174**: 06/08/2019<br>**3098**: 09/11/2019 | *PASS* |
| SQL-7 | Eyeball Check<br><br>*Inspect 5 randomly chosen records from Source* | **862**<br>**1397**<br>**1414**<br>**2429**<br>**2605** | Same records, different DATE format due to SQL defaults. | *PASS* |
| Additional SQL-8 | Weekend Check Checking whether all Day = 6 or 7 values correspond to Weekend values being 1. | 648 | 648 | *PASS* |

| Additional SQL-9 | Cost and Profit Check Checking both profit and cost of items are positive values. | 0 | 0 | PASS |
|---|---|---|---|---|

| Old column names | Modified column names |
|---|---|
| ID | id |
| Purchase-date | purchase_date |
| Date | date_mmddyyyy |
| Time | time_stamp |
| Quarter | quarter |
| Qtr | qtr |
| Year | year_num |
| Month | month_num |
| Date_1 | day_num |
| 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 | gender_int |
| Customer ID | customer_id |
| Gender<br>F=1665<br>M=1815 | gender |
| 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 |
| Profit Percentage (%) | profit_percentage |
| Profit (INR) | profit_inr |
| Cost Price | cost_price |

## Additional Notes:

- Purchase_date is stored as VARCHAR to allow a successful import into DBeaver(SQL)
- Within the DESTINATION, all dates are stored as YYYY-MM-DD, due to SQL defaults
- The id has been serialised (in numeric order and unique values). This will ensure all ids can never be duplicated and will allow you to link the table to any future table that utilises the same ids.
- SQL-8 shows the data has passed the quality checks, but this does indicate there are errors within the source data (648 rows have weekend = 0 when the day_of_week is either 6 - Saturday or 7 - Sunday).
- SQL-9 is used to check if there are any negative values in cost_price & price_inr to which we found none.