DATA DYNAMO

Quality Control Form

Team Name	GROUP_BY
Team Members	Hena, Jake, Madushan, Houcine
Client Name	Majestic Manuscripts
Date of file received	22/09/2025
Name of file received	Book Sales Records - V2
File Format received	csv
Size of file (KB) received	928
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	**group_by_bsr**

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 -1	**Date format Check 1**	5/31/2019	5/31/2019	*Pass*
SQL-6 -2	**Date format Check 2**	4/7/2019	4/7/2019	*Pass*
SQL-6 -3	**Date format Check 3**	4/10/2019	4/10/2019	*Pass*
SQL-6 -4	**Date format Check 4**	4/21/2019	4/21/2019	*Pass*
SQL-6 -5	**Date format Check 5**	11/17/2019	11/17/2019	*Pass*
SQL-7-1-1	**Eyeball check 1**	1430	1430	*Pass*
SQL-7-1-1	**Eyeball check 2**	2099	2099	*Pass*
SQL-7-1-2	**Eyeball check 3**	690	690	*Pass*
SQL-7-1-3	**Eyeball check 4**	2291	2291	*Pass*
SQL-7-1-4	**Eyeball check 5**	2221	2221	*Pass*
SQL-8	**Count of Null Customer ID**	0	0	*Pass*
SQL- 9	**Checking Expected Range**	0	0	*Pass*
SQL-10	**Category consistency-gender**	0	0	*Pass*

	Timeliness check	Max datedate:12/12/2019 Min date: 01/01/2018	Max datedate:12/12/2019 Min date: 01/01/2018	*Pass*

Example for extra checks: Count of Nulls - {COLUMN NAME} ; Aggregates (SUM/MIN/MAX) checks between Source and Destination

Notes On Quality Checks

Definitions:

SOURCE: A copy of the original file received from the client. This is the starting point for the data before any changes have been made by DATA DYNAMO.

DESTINATON: This is otherwise known as the ‘target’ in source to target mappings. Where the data now resides post extraction – Client Postgres SQL database server,

OUTCOME: The outcome of a quality control test. This indicates that the query executed, or test conducted meets acceptance criteria for the data quality rule. Please see the query notes for the acceptance criteria and data quality aspect tested for in each query.

Result: The output from running the query.

Query Notes:

- **SQL-1:** Count of rows - This is a check for data completeness. This check compares the number of rows in the source data with the destination data to ensure no records were lost or unintentionally added in the data cleansing and migration process.
- **SQL-2:** Count of rows - This is a check for uniqueness quality in the dataset.The count number of distinct rows, was queried to verify that only unique records exist in either source and/or destination. In other words, no hidden duplication.
- **SQL-3:** Count of columns - A comparison column counts between source and destination data done to test the consistency or structural integrity of the

data. A difference in column counts would indicate an addition or loss of structure after cleansing and or migrating.

- **SQL-4:** Sum of column sums **–** For this check, a sum of numeric column was taken giving resulting in one sum value for each row- the resulting sum values are added up for each row. The accuracy characteristic of data quality for this query. Comparison of source and destination totals helps to detects silent loss or corruption of date that would typically go undetected through eyeball check. (e.g truncation of numerical values).

- **SQL-5:** Sum of row sums **–** For this check, a sum of numeric rows was taken and the resulting sum values are added up for each column. The accuracy characteristic of data quality for this query. Comparison of source and destination totals helps to ensure that transformation preserve keep their row level integrity of numerical values.

- **SQL-6:** Time format check- These test checks the validity of 5 dates by checking if the format of the date conforms to that required or accepted by the client in the source and destination structure. Randomness and reproducibility were achieved by using a random function (random()) and bounding the output between the minimum and maximum row id. This allows for the mitigation of human bias from the selection process. This check assesses both accuracy, consistency and validity of values in (5) randomly chosen records all rows. The matching data
-
- **SQL-7:** Eyeball check - An inspection of 5 randomly chosen records is conducted to identify any unintended values, the rules of which may not be easily defined in the table structure. Randomness and reproducibility were achieved by seeding a random function in SQL. See notes on SQL-6 for details on our process of random selection.

- **SQL-8:** Null value check: This test checks for the presence of null values in the source and destination structure. Customer_id was used for the check as it would be critical for sales analysis. Customer_id also should not be null. Although this is not required at this stage, further transformations will be to the data when changing the schema structure to serve analytical processing.

- **SQL-9:** Checking expected range: This test checks if there are any values in the cost of column that are outside of the expected range (>0), a score of 0 means that there is a count of 0 values in that column that lie outside of the expected range. This is essentially a validity check.

- **SQL-10:** Checking Categorical fields: This is a data consistency check on the categories, ensuring that there is no unexpected category.

- **SQL-11:** -Transaction timeliness check: This checks the timeliness of the data. Given that our client is interested in doing analysis on data between years 2018 and 2019. It is important that the data being cleansed is fit for purpose. This test checks the most and least recent dates in the data set for the source and destination data.

Errors during testing:

- Errors during testing: Duplicates found on v1, discussed with stakeholder from mystic manuscripts who supplied us with an updated csv file v2
- Error when importing csv file due to date datatype issues resolved by setting new data types in the creation table (increase character limits and changing some integers to float)
- Issue with float data type rounding error for sum rows and sum columns fixed by updating data types to numeric
- Issue using date/time formats present as yyyy-mm-dd, fixed by keeping all dates as varchar