Data Quality Report – QuickBite ETL Pipeline

# 1. Duplicate Check

No duplicate 'order\_id' values were found in the 'fact\_order' table.

# 2. Null Value Summary

All null values in critical columns have been addressed in the ETL process. Below are the results:

|  |  |
| --- | --- |
| Column | Null Count |
| order\_id | 0 |
| customer\_id | 0 |
| order\_date | 0 |
| quantity | 0 |
| price\_per\_item | 0 |

# 3. Data Cleaning Fixes

- Removed rows where 'order\_id', 'order\_date', or 'price\_per\_item' were null.  
- Removed rows where 'quantity' was 0 or less.  
- Replaced missing 'customer\_id' values with 'unknown'.  
- Converted 'order\_date' to timestamp format.  
- Recalculated 'total\_price' if missing using 'quantity × price\_per\_item'.  
- Partitioned data by 'year' and 'month' for optimization.

# 4. Order Status Distribution

Refer to the query below to fetch counts:

SELECT order\_status, COUNT(\*) AS count  
FROM fact\_order  
GROUP BY order\_status  
ORDER BY count DESC;