Data Quality Report
Database: mcqueen_db
Tables: Car, Customer, Employee, Date, Sales
Focus: Post-partitioning validation of Sales table
Date: 2025-08-23
Author: Rolando Abagon Jr., Jovic Salmeron, Gretchen Contillo

Summary

Metric	Status	Notes
Partitioning Implemented	☑ Yes	Sales partitioned (SaleID)
Materialized View	☑ Created	top10_highest_sales confirmed working
Data Integrity	☑ Verified	No major integrity issues found
Referential Integrity	☑ Validated	Foreign keys match with dimension tables
Row Count Consistency	☑ Checked	Pre- and post-partition row counts match
Data Duplicates	⊘ None	No duplicate SaleID detected
Null Values	☑ Checked	No Null values
Outliers	⊘ None	No significant outliers in Sale_amount
Timeliness	☑ Correct	Data mapped to correct partitions

1. Partitioning Check (Sales)

Check	Result
Partitioning key	SaleID
Type of partitioning	RANGE on SaleID
Partition counts	3 range partitions (p1, p2, p3)
Partition coverage	All sales record
Query performance	Improved on any sales record queries

Showing Partition via Show Create Table

Showing Partition via Information Schema

2. Row Count Consistency

Table	Pre-Partition Count	Post-Partition Count	Match
Sales	100 rows	100 rows	☑ Yes

Run:
SELECT COUNT(*) FROM sales;
Result: Total count from all partitions

3. Referential Integrity

FK Relationship	Check	Result
sales.customerid → customer.customerid	☑ Exists	All valid
sales.employeeid → employee.employeeid	☑ Exists	All valid
sales.carid → car.carid	☑ Exists	All valid
sales.dateid → date.dateid	☑ Exists	All valid

Run:
SELECT COUNT(*) FROM sales s
LEFT JOIN customer c ON s.customerid = c.customerid
WHERE c.customerid IS NULL;
Result: 0

4. Uniqueness Check

Column	Check	Result
saleid	No duplicates	☑ Pass

Run:
SELECT saleid, COUNT(*)
FROM sales
GROUP BY saleid
HAVING COUNT(*) > 1;
Result: 0 rows

5. Null Value Check

Table	Column	Null Count	Notes
Sales	carid	0	OK
Sales	customerid	0	OK
Sales	dateid	0	OK
Sales	employeeid	0	OK
Sales	profit	0	OK
Sales	quantity_sold	0	OK
Sales	sale_amount	0	OK

Run:
SELECT
 SUM(CASE WHEN carid IS NULL THEN 1 ELSE 0 END) AS car_nulls,
 SUM(CASE WHEN customerid IS NULL THEN 1 ELSE 0 END) AS customer_nulls,
 SUM(CASE WHEN dateid IS NULL THEN 1 ELSE 0 END) AS date_nulls,
 SUM(CASE WHEN employeeid IS NULL THEN 1 ELSE 0 END) AS employee_nulls,
 SUM(CASE WHEN profit IS NULL THEN 1 ELSE 0 END) AS profit_nulls,
 SUM(CASE WHEN quantity_sold IS NULL THEN 1 ELSE 0 END) AS quantity_sold_nulls,
 SUM(CASE WHEN sale_amount IS NULL THEN 1 ELSE 0 END) AS amount_nulls
FROM sales;
Result: 0

6. Sale Amount Range/Outlier Check

Metric	Value
Min Sale	₱ 839,854.00
Max Sale	₱ 2,188,598.00
Average	₱ 1,539,710.72
Outliers Found?	⊘ No extreme values found

Run:
SELECT MIN(sale_amount), MAX(sale_amount), AVG(sale_amount) FROM sales;
Result: No outliers

7. Timeliness Check (Partition Alignment)

Partition	SaleID	Records (Sale_amount)	In Correct Partition?
p1	24	971,088	☑ Yes
p2	63	1,290,739	☑ Yes

Sample check:
SELECT sum(Sale_amount) as `Total Sales`
FROM sales
WHERE saleId BETWEEN '0' AND '50';
Result: Matched sum of sales of partition 1

SaleID	DateID	CarID	CustomerID	EmployeeID	Sale_amount	Profit	Quantity_sold
0	94	53	67	14	1065831	132766.9	2
1	99	46	65	75	1740101	105981.6	3
2	78	5	93	84	1264623	129175.1	2
3	3	7	61	51	2149183	199073.2	1
4	91	56	9	10	1828168	150295.9	1
5	16	35	79	81	937745	101779.5	3
6	48	76	67	37	1481963	141337.1	3
7	12	89	14	83	1702193	195908.2	3
8	55	57	29	52	1250936	104945	2
9	53	93	12	40	1636197	151636.9	3
10	47	19	87	60	1334542	75686.32	1
11	55	12	95	94	995517	86873.48	1
12	7	75	71	71	1966794	163174.3	1
13	45	85	96	54	1661528	227312.3	3
14	36	76	39	45	907890	54800.29	3
15	19	84	61	28	1246145	167890	2
16	47	14	97	35	1966573	211229.6	2
17	98	79	78	86	1976430	225846.5	1

8. Materialized View Check

View	Status	Notes
top10_highest_sales	☑ Working	Correct rows, sorted by sale amount

Run:
SELECT * FROM top10_highest_sales;
Result: Correct sorting and data consistency

Sale ID	Date Sold	Customer Name	Employee Name	Car Model	Sale Amount
54	2/20/2020	Ramon Del Rosario	Justine Reyes	Ranger	2188598
20	2/21/2020	Ana Garcia	Maria Lopez	Everest	2178551
71	2/12/2020	Ramon Flores	Angelo Mendoza	Ranger	2163066
51	2/12/2020	Jose Bautista	Carlos Bautista	Everest	2157497
38	1/19/2020	Carlos Cruz	Angelo Mendoza	CR-V	2153592
3	1/4/2020	Ana Garcia	Luz Flores	Everest	2149183
61	1/22/2020	Ligaya Lopez	Carlos Bautista	Hilux	2110175
60	1/17/2020	Rosario Reyes	Angelo Mendoza	Montero Sport	2096105
45	3/17/2020	Andres Torres	Maria Lopez	Mirage	2089766
18	3/25/2020	Rosario Bautista	Trishia Torres	Fortuner	2077154

Final Recommendation

- ☑ No critical data issues found
- 🔁 Schedule regular DQ checks (daily/weekly)
- 💾 Consider automating the above queries for routine monitoring
- 📌 Partition maintenance (add new partition if records reached 1 million rows)