# **Final Data Quality & Testing Report**

## **1. Test Strategy**

Our testing approach covers the entire pipeline, from raw source files to end-user reports in Power BI.

* **Scope of testing:**
  + Source data (CSV/SQL files)
  + Landing & Staging layer
  + 3NF / DWH layer
  + Data Marts (DM)
  + Power BI dashboards
* **Approach:**
  + **Automated SQL checks** for row counts, duplicates, referential integrity, SCD updates.
  + **Manual analysis** in PostgreSQL and Power BI for business rule validations and KPI checks.
  + **Profiling tools** (e.g., ydata-profiling / Pandas) to spot anomalies or distributions.
* **Tools:** PostgreSQL, SQL scripts, Power BI, Python (optional for profiling).

This strategy ensures that data is validated at each transformation step and the final reports reflect accurate, complete, and consistent information.

## **2. Test Cases & Scenarios**

### **2.1 Data Quality Checks**

|  |  |  |  |
| --- | --- | --- | --- |
| **Test Case** | **Steps** | **Expected Result** | **Dimension** |
| Row count consistency | Compare row counts across source → staging → DWH | Counts match (no dropped/extra rows) | Completeness |
| Duplicate detection | Query DISTINCT vs. total count on key columns (e.g., Customer ID) | No duplicates found | Uniqueness |
| Null/invalid values | Check NOT NULL columns (e.g., IDs, Dates) | No NULLs; invalid values flagged | Validity |

### **2.2 ETL Pipeline Validations**

|  |  |  |
| --- | --- | --- |
| **Test Case** | **Steps** | **Expected Result** |
| Transformation check | Verify business rules (e.g., phone concatenation, total cost calculation) | Values match defined transformation rules |
| Join integrity | Run join queries between fact and dimensions | No orphan keys; all FKs resolve |
| Filtering logic | Check applied filters (date ranges, active customers) | Only correct rows included |

### **2.3 Schema Validation**

|  |  |  |
| --- | --- | --- |
| **Test Case** | **Steps** | **Expected Result** |
| Dimension defaults | Inspect each dim table | Default rows (-1, ‘N/A’) exist |
| Column types | Compare ERD vs. actual schema | Data types match (no truncation) |
| Constraints | Check PK/FK constraints | All enforced correctly |

### **2.4 Dashboard Validation (Power BI)**

|  |  |  |
| --- | --- | --- |
| **Test Case** | **Steps** | **Expected Result** |
| KPI correctness | Compare Power BI totals with SQL queries | Values match (e.g., revenue, margin) |
| Filters & slicers | Apply filters in Power BI (date, product, region) | Dashboard updates correctly |
| Drill-down | Test hierarchy navigation (year → quarter → month) | Smooth drill-down with consistent numbers |

## **3. Checklist**

Before delivery, the following items must be verified:

* All ETL jobs ran successfully without errors
* Row counts consistent across all layers
* No duplicates in keys (Customer, Product, etc.)
* Default rows exist in every dimension
* All foreign keys resolve (no orphan facts)
* Power BI visuals match SQL output for KPIs
* Filters and slicers tested and functional
* Documentation updated (ERD, mapping, test report)

## **4. Test Report**

**Execution summary:**

* Data quality checks executed (row counts, nulls, duplicates)
* ETL rules validated (transformations, joins, filters)
* Schema verified (types, defaults, constraints)
* Power BI validated (KPIs, filters, drilldowns)

**Defects found:**

* Minor formatting issues in Power BI visuals (resolved)
* One duplicate row in staging (fixed during ETL script adjustments)

**Final status:**  
 All **critical tests passed**. The solution meets both **business** and **technical** requirements and is ready for delivery.

## **5. Additional Notes**

* **Automated scripts**: Basic SQL queries for row counts and duplicates are reusable for regression testing.
* **Logging**: ETL jobs write logs of inserted/updated/deleted rows for traceability.
* **Future improvement**: Add anomaly detection scripts to flag unusual KPI spikes automatically.

# **Conclusion**

The testing process confirms that the solution is **reliable, maintainable, and trustworthy** from a data quality perspective. Business stakeholders can rely on the dashboards for accurate decision-making.