

Data Quality Assessment Report

Source Dataset

PatientID	PatientName	PhoneNumber	AppointmentDate	DoctorName	PaymentStatus
P001	Kwame Mensah	0244123456	2025-10-15	Dr. Osei	Paid
P002	Ama Serwa	244789012	15/10/2025	dr. osei	paid
P001	Kwame Mensah	0244123456	2025-10-20	Dr. Adjei	Pending
P003	Kofi Annan	0244567890	10/16/2025	Dr. Osei	Failed
P004		0555234567	2025-10-17	Dr. Mensah	Paid
P002	Ama Serwa	0244789012	2025-10-15	Dr. Osei	Paid

This report assesses the quality of patient appointment data across key data quality dimensions, evaluates the business impact of identified issues, and recommends corrective actions to improve data reliability and operational efficiency.

Task 1: Identified Data Quality Issues

Accuracy: Incorrect phone numbers and inconsistent date representations were identified. For example, the phone number for Patient P002 appears as both 244789012 and 0244789012. Additionally, the appointment date 10/16/2025 conflicts with other October 2025 records.

Completeness: One record is missing a patient name, making patient identification difficult. Duplicate records without unique visit identifiers also reduce record completeness.

Validity: Multiple format violations exist, including inconsistent phone number formats, mixed date formats, inconsistent capitalization of doctor names, and payment status values.

Consistency: The same patients and doctors appear with different formats and values across records, leading to inconsistent representations of the same entities.

Timeliness: Duplicate and conflicting records suggest delayed or repeated data entry, resulting in outdated appointment information being used operationally.

Uniqueness: Duplicate patient and appointment records exist, particularly for Patient IDs P001 and P002, without a unique appointment identifier.

Task 2: Business Impact Assessment

Accuracy and validity issues lead to SMS reminder failures, incorrect appointment scheduling, and unreliable billing and reporting. Completeness and consistency issues increase the risk of patient misidentification and inaccurate clinical records. Timeliness and uniqueness issues cause duplicate billing, inflated patient counts, and poor operational planning. Operations is most affected by scheduling and reminder failures, Finance by incorrect billing and reports, and Clinical functions by patient identification risks.

Task 3: Recommended Solutions

The most critical issue is inconsistent and invalid date formats. This should be addressed by enforcing a single ISO date format (YYYY-MM-DD) at data entry and database levels, implemented by application developers or data engineers and verified through format validation checks.

Phone number accuracy should be improved by enforcing standardized Ghana phone number validation rules and cleaning existing data. This responsibility lies with the data steward or operations IT team and can be verified by successful SMS delivery rates.

Duplicate records should be prevented by implementing unique constraints and a master patient index. This should be handled by the database administrator and data governance team and verified through post-cleanup duplicate detection checks.

Task 4: Risk of Poor Data Consistency

The biggest risk of poor data consistency is loss of trust in operational and financial reports. When decision-makers cannot rely on dashboards and summaries, organizations resort to manual processes, leading to inefficiency, delayed decisions, and increased operational costs.