

Name:Diane ISHIMWE

EMail:diane.ishimwe@amalitech.com

Specialization:Python Backend

Data Quality Detective Challenge

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

Task 1: Data Quality Issues Identified

1. Accuracy Violation

Example:

Phone number 244789012

The phone number is factually incorrect because it is missing the leading zero required in Ghanaian phone numbers.

2. Completeness Violation

Example:

P004,, 0555234567, 2025-10-17, Dr. Mensah, Paid

The patient name field is missing.

3. Consistency Violation

Examples:

Doctor names appear as: Dr.Osei and dr.osie

Payment status appears as:Paid and paid

Date formats are mixed:2025-10-15,15/10/2025 and 10/16/2025

4. Timeliness Violation

Example:

Date formats are inconsistent which may cause appointment scheduling systems to interpret dates wrongly.

5. Validity Violation

Examples:

Phone number 244789012 does not follow 10-digit Ghana phone format.

Date format 10/16/2025 is not standardized.

6. Uniqueness Violation

Example:

Patient ID P001 and P002 appear multiple times.

This creates duplicate appointment records.

Task 2: Business Impact Assessment

Data Issue	Operational Problem	Business Function Affected
Invalid phone numbers	SMS appointment reminders fail	Operations
Missing patient name	Difficult patient identification and billing	Clinical and Finance
Inconsistent doctor names	Incorrect reporting and analytics	Operations
Mixed date formats	Wrong appointment scheduling	Operations
Invalid formats	System parsing and SMS gateway errors	IT and Operations
Duplicate records	Double counting and possible double billing	Finance

Task 3: Recommended Solutions for Critical Issues

1. Phone Number Validation

Technical Solution

- Implement regex validation for Ghanaian numbers: Must be 10 digits, Must start with 0.
- Add frontend and backend validation.

Responsible Roles

- Backend Developer
- QA Engineer

Verification

- Test SMS sending functionality.
- Attempt inserting invalid numbers (system should reject).

2. Data Standardization

Technical Solution

- Convert all dates to ISO format (YYYY-MM-DD).
- Standardize doctor names (capitalize first letters).
- Normalize payment status values to: Paid, Pending, Failed

Responsible Roles

- Backend Developer
- Database Administrator

Verification

- Run data cleaning script.
- Check report grouping accuracy.

3. Duplicate Record Prevention

Technical Solution

- Add database constraints: Unique constraint on PatientID + AppointmentDate.
- Implement deduplication script.

Responsible Roles

- Database Administrator
- Backend Developer

Verification

- Try inserting duplicate records.
- Run duplicate detection queries.

Task 4: Biggest Risk of Poor Data Consistency (Specialization Perspective)

From a developer perspective, the biggest risk of poor data consistency is unreliable system behavior.

Poor consistency can lead to:

- Incorrect analytics and reporting.
- Failed automation processes such as SMS reminders.
- Billing errors.
- Increased debugging and maintenance cost.

Maintaining consistent data ensures system reliability, accuracy, and trustworthiness.