Data Quality Detective Challenge – Final Report

Backend Engineering Perspective

Organization: MedTrack Ghana
Role: Junior Backend Developer
Dataset: Patient Appointment CSV

---

1. Identified Data Quality Issues

Accuracy

- *Issue:* Same patient (*Ama Serwa*) has conflicting phone numbers (244789012 vs 0244789012).

- *Impact:* SMS reminders may fail or reach the wrong number.

Completeness

- *Issue:* Missing patient name for PatientID P004.

- *Impact:* Check-in confusion; billing systems may reject incomplete records.

Consistency

- *Issue:* Doctor names recorded inconsistently (Dr. Osei vs dr. osei).

- *Impact:* Incorrect reports and duplicated doctor records.

Timeliness

- *Issue:* Mixed and ambiguous appointment date formats (2025-10-15, 15/10/2025, 10/16/2025).

- *Impact:* Reminders may be sent on incorrect dates.

Validity

- *Issue:* Invalid phone number formats not conforming to Ghanaian standards.

- *Impact:* SMS gateway rejects messages.

Uniqueness

- *Issue:* Duplicate patient records (e.g., P001, P002).

- *Impact:* Inflated reports; duplicate or missed billing.

2. Business Impact Assessment

Data Quality Issue	Operational Problem	Most Affected Function
Accuracy	SMS failures, missed appointments	Operations
Completeness	Billing rejections, identification issues	Clinical, Finance
Consistency	Duplicate doctor counts, faulty reports	Operations
Timeliness	Wrong reminder schedules	Operations
Validity	SMS gateway rejections	Operations
Uniqueness	Duplicate counts and billing errors	Finance, Operations

3. Recommended Solutions (Top 3 Critical Issues)

A. Invalid / Incorrect Phone Numbers

- **Technical Fix:** Input validation (regex), normalization script, block invalid saves.

- **Owner:** Backend Engineering Team.

- **Verification:** Regex audit, SMS delivery success metrics, gateway error logs.

B. Duplicate Patient Records

- **Technical Fix:** Uniqueness constraints, deduplication script, master patient index.

- **Owner:** Data Engineering (with Operations review).

- **Verification:** Patient count reconciliation; no duplicate PatientIDs; billing cross-checks.

C. Inconsistent / Invalid Appointment Dates

- **Technical Fix:** Enforce ISO 8601 (YYYY-MM-DD), reject ambiguous formats, standardize historical data.

- **Owner:** Backend Engineering + DBA.

- **Verification:** Date-format audits; reminder scheduling tests; on-time SMS confirmation.

4. Backend Perspective: Biggest Risk of Poor Data Consistency

Unreliable system behavior due to conflicting source data.

- Backend logic depends on uniform data; inconsistencies break conditionals and queries.

- Leads to skipping/duplicating reminders, billing failures, and unreliable APIs.

- Increases technical debt and production incidents, slowing development and scaling.

- Ultimately undermines automation and system trust.

Conclusion

The dataset contains critical data quality violations directly responsible for SMS failures, inaccurate reporting, and billing issues. Implementing backend validation, standardization, and deduplication, paired with clear ownership and verification, will significantly improve system reliability, operational efficiency, and trust in MedTrack Ghana's platform.