AXON HEALTHCARE

EMR Analytics Capstone

From raw EMR data to a defensible BI story — Excel → SQL → Tableau → Power BI → QA

TEAM MEMBERS

Nagargoje Krushnakant

Saniya Suresh Rajeshirke

Achi Venkata Sunendra

Sanket Dilip Katore

Lakshita Thakur

Ashish Singhdeo

Final Presentation • Team AXON • October 2026

Project Objectives

What this capstone set out to deliver

1 Prepare the data

Clean and validate the 10-table EMR dataset (patients, visits, treatments, labs, billing, insurance claims) in Excel.

2 Build a source of truth

Load the data into MySQL with primary and foreign keys, so every dashboard reads from one database.

3 Build dashboards

Create interactive dashboards in Tableau and Power BI covering visits, treatments, revenue and patient health.

4 Prove the numbers

Reconcile every KPI between SQL, Tableau, Power BI and Excel using QA queries before sign-off.

Project Overview

AXON Healthcare EMR Capstone — one dataset, five stages, one story

10,000	1,000	10	0
Patients	Doctors	Tables	Orphan rows

1 — Excel

Clean the raw export, build a first-pass pivot dashboard

2 — SQL

Load into MySQL under a proper schema with PK/FK

3 — Tableau

Connect to SQL, build interactive dashboards

4 — Power BI

Same dashboards, rebuilt with DAX measures

5 — QA

Reconcile every number across SQL, Tableau, Power BI

Stage 1 — Excel

Work directly on the raw data — no database yet

Cleaning checks (all clean)

- 0 duplicate Visit ID / Treatment ID / Claim ID
- 0 missing values in Visit Status, Diagnosis, Outcome, Costs
- 0 orphan Treatment rows (every one links to a real Visit)
- Costs in normal range ($50–$5,000), no negative values

Expected blanks (not bugs)

- Approved Amount / Processing Days blank for 1,211 Pending claims
- Denial Reason blank for 7,286 non-Denied claims

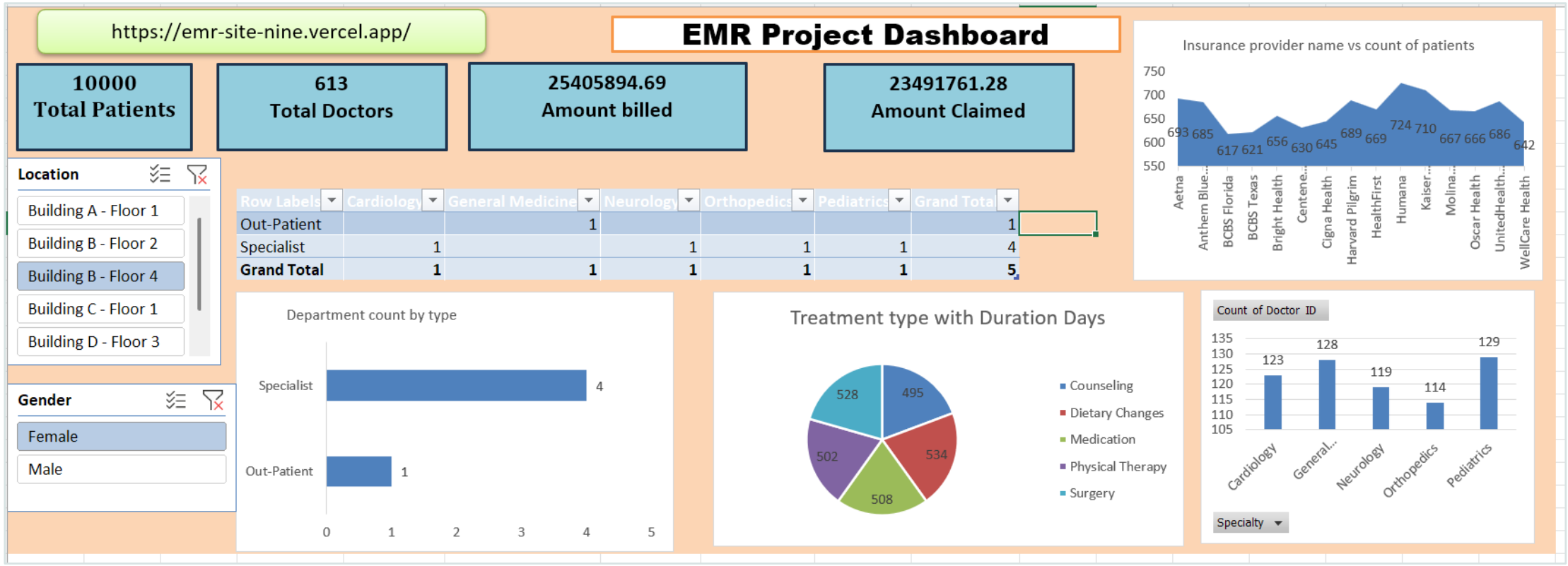
Excel KPI dashboard (Raw_Clinical + Raw_Claims)

32.7%	Visit Completion Rate
49.6%*	Treatment Success Rate
60.8%	Claim Approval Rate
9.2%	Claim Denial Rate
41.2%	Insurance Coverage Rate

**Excel's formula excludes Ongoing treatments from the denominator — reconciled to 33.0% in Tableau/Power BI (see QA slide).*

Excel Dashboard

EMR Project Dashboard — built directly on the raw data with slicers



KPI cards, pivot table and charts with Location and Gender slicers — the first-pass dashboard before the data moved to SQL.

Stage 2 — SQL

Load into MySQL under a proper schema — the new source of truth

Department	**Doctor**	**Patient**	**Visit**	**Treatment**
Lab Test	**Billing**	**Insurance Provider**	**Insurance Policy**	**Insurance Claims**

Load order: Department → Doctor → Patient → Visit → Treatment / Lab Test / Billing → Insurance Provider → Insurance Policy → Insurance Claims

10	14	0	1
Tables loaded	FK relationships verified	Orphan rows found	Bug found & fixed

The bug: Policy Number had 59 duplicate values in Insurance Policy (two different patients' policies collided on the same generated number) — fixed by adding a surrogate key, Policy Row ID.

All KPIs

Every metric tracked on the Healthcare Analysis Dashboard

10,000	10,000	32.7%	33.0%	$25.41M
Total Patients	Total Visits	Completion Rate	Success Rate	Total Revenue

$7.06M	$14.49M	2K	0.28	80.35%
Outstanding Amount	Insurance Covered	Partially Paid Bills	Outstanding %	Chronic Rate

Supporting breakdowns

Revenue by Specialty: General Medicine $5.5M • Cardiology $5.3M • Pediatrics $5.2M • Neurology $4.7M • Orthopedics $4.6M

Patients by Gender: Female 34.0% • Male 32.8% • Other 33.1%

Patients by Age Group: 41–60 (2.8K) • 21–40 (2.8K) • 61–80 (2.8K) • 0–20 (1.0K) • 80+ (0.6K)

Stage 3 — Tableau

Connected live to MySQL — never to the Excel file

What we built

- Live connection to MySQL (axon_healthcare)
- Data model: 10 tables, 14 relationships
- Calculated fields: Completion Rate, Success Rate, Age Group
- Slicers: Year of Visit Date, Department, Specialty
- Dashboard actions: click any chart to filter the rest

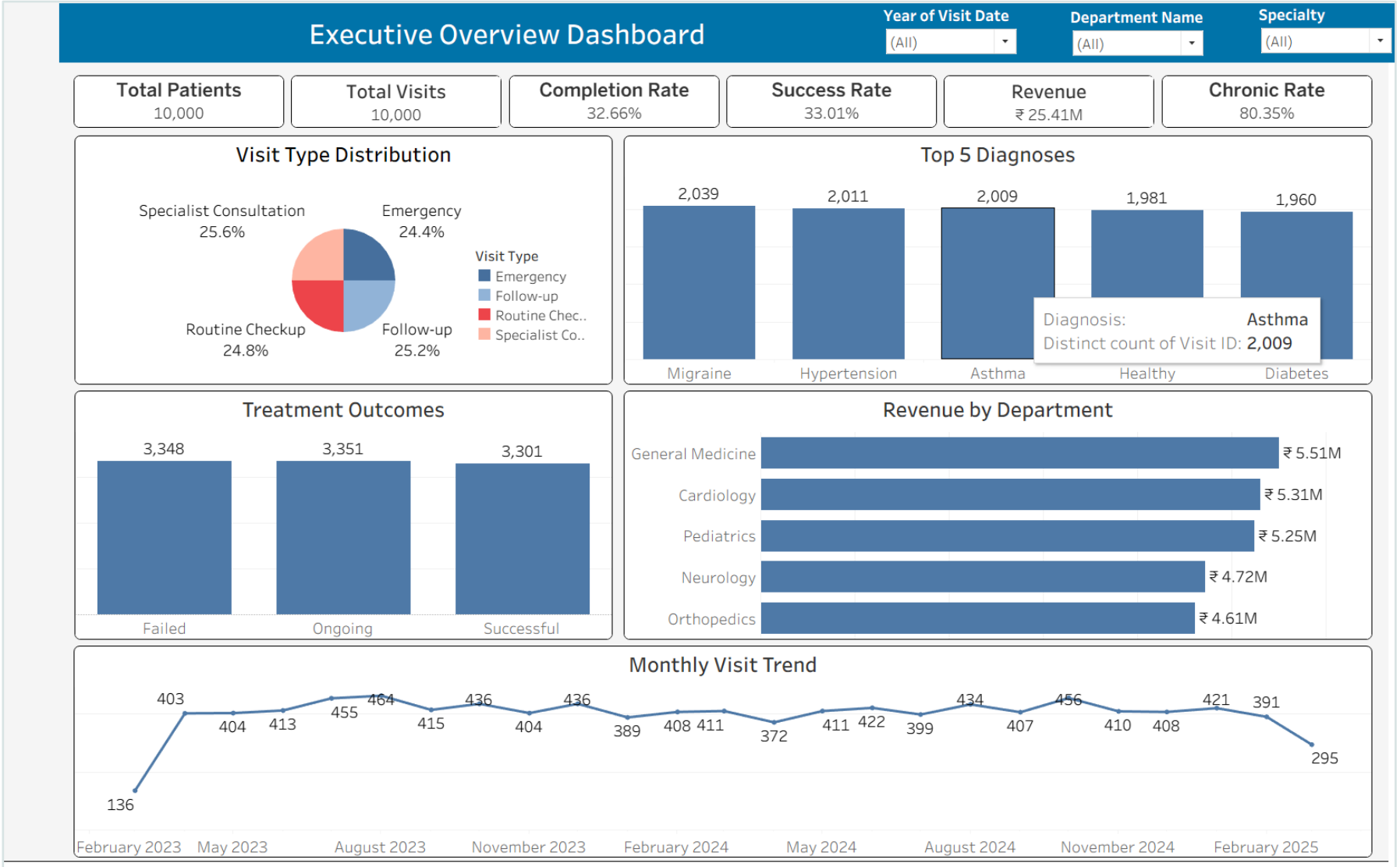
Issue caught in review

The first data model canvas had duplicate copies of tables (lab_test1, patient1, insurance_claims1/2, insurance_provider1) — each created by re-dragging a table instead of drawing a second relationship line.

Fix: one box per table (10 total), extra relationships drawn as lines between existing boxes. This prevents double-counting in KPIs.

Tableau Dashboard

Executive Overview — connected live to MySQL



Stage 4 — Power BI

Same source, same KPIs — rebuilt with DAX measures

```
Visit Completion Rate =
DIVIDE(
  CALCULATE(COUNTROWS(Visit),
    Visit[Visit Status] = "Completed"),
  COUNTROWS(Visit)
)

Success Rate =
DIVIDE(
  CALCULATE(COUNTROWS(Treatment),
    Treatment[Outcome] = "Successful"),
  COUNTROWS(Treatment), 0
)
```

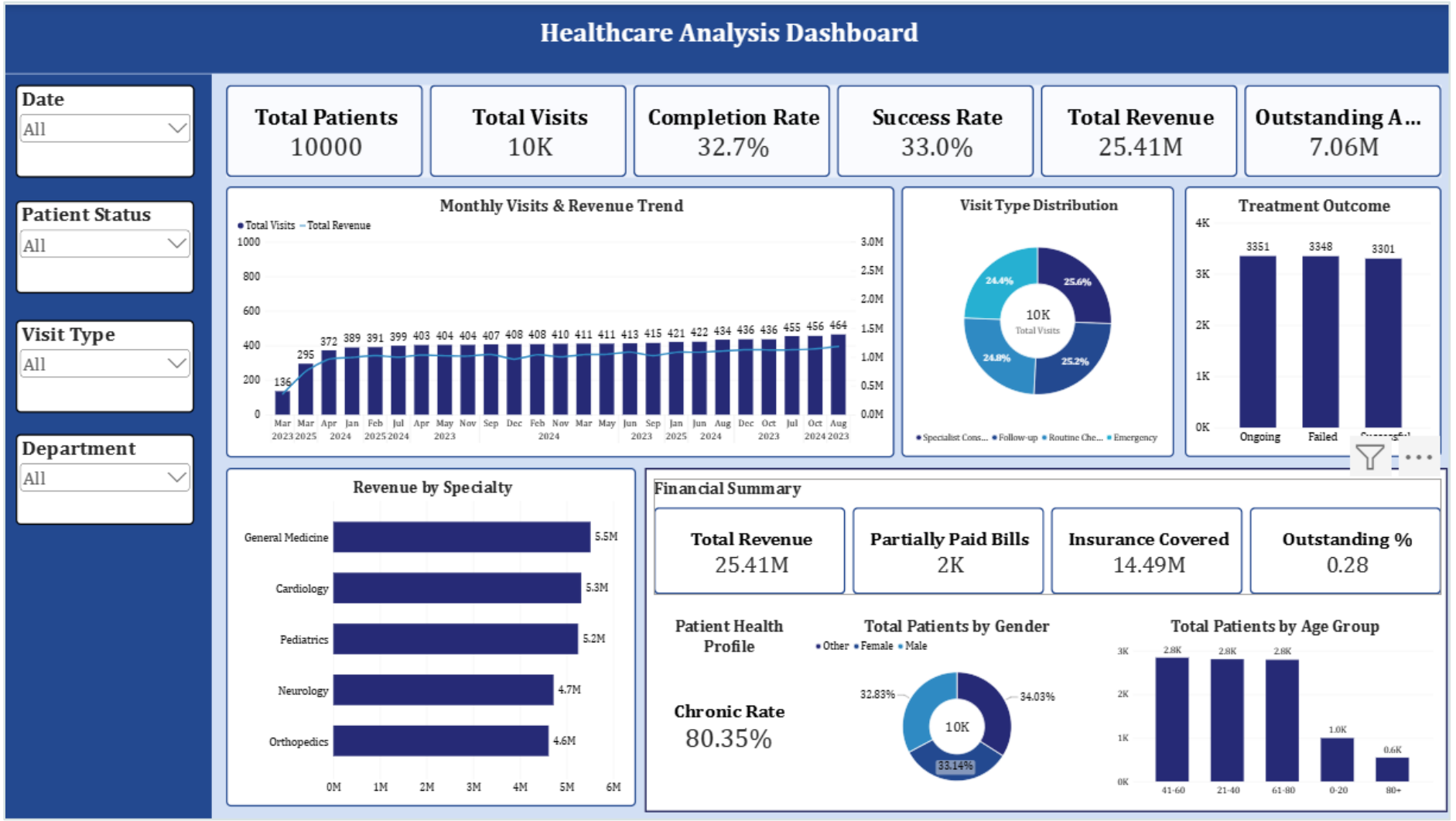
Build notes

- MySQL connector + relationships from the schema
- Slicers: Date, Patient Status, Visit Type, Department
- Rates formatted as Percentage (no ×100 in DAX)
- Success Rate uses Outcome, not Status

Two bugs caught during the build: a 3266% display (percentage format applied on top of ×100) and a 0% Success Rate (filtering the Status column, which never contains "Successful").

Power BI Dashboard

Healthcare Analysis — Executive Overview



Stage 5 — QA & Reconciliation

Every number matched across tools before sign-off

KPI	Excel	Tableau	Power BI
Total Patients	10,000	10,000	10,000
Visit Completion Rate	32.7%	32.66%	32.7%
Treatment Success Rate	49.6%*	33.01%	33.0%
Total Revenue	—	25.41M	25.41M
Chronic Rate	—	80.35%	80.35%

Discrepancy found and explained

*Excel showed 49.6% because 3,301 successful ÷ (3,301 successful + 3,348 failed) leaves out the 3,351 Ongoing treatments. Tableau and Power BI divide by all 10,000 treatments (33.0%), matching the BRD definition. Standardized on the all-treatments version.

SQL Queries

QA validation run against MySQL before sign-off

```sql
-- 1. Count Validation

SELECT COUNT(*) FROM Patient;
```

COUNT(*)
10000

```sql

SELECT COUNT(*) FROM Visit;

```

COUNT(*)
10000

03 — Consistency Check

```sql
-- 3. Consistency Check
SELECT v.Visit_ID, p.Patient_ID
FROM Visit v
LEFT JOIN Patient p
  ON v.Patient_ID = p.Patient_ID
WHERE p.Patient_ID IS NULL;
-- returns 0 rows
```

04 — Duplicate Records Check

```sql
-- 4. Duplicate Records Check
SELECT Patient_ID, COUNT(*)
FROM Patient
GROUP BY Patient_ID
HAVING COUNT(*) > 1;
-- returns 0 rows
```

05 — Aggregation Check

```sql
-- 5. Aggregation Check
SELECT SUM(TreatmentCost)
FROM Treatments;
SELECT AVG(Age) FROM Patient;
-- compare vs Power BI / Tableau
```

AVG(Age)
48.9383

06 — Performance Check

```sql
-- 6. Performance Check
-- EXPLAIN ANALYZE
SELECT * FROM Visit
WHERE Visit_Date
  BETWEEN '2023-01-01'
  AND '2023-12-31';
```

Visit_ID	Patient_ID	Doctor_ID	Visit_Date
1	236037	35654	2023-08-29 00:00:00
5	860284	96922	2023-09-10 00:00:00
8	359681	83241	2023-10-29 00:00:00
14	489581	31901	2023-08-14 00:00:00
18	169909	91480	2023-12-26 00:00:00
20	113803	23724	2023-10-26 00:00:00
21	893024	45097	2023-06-13 00:00:00

Result: all 10 tables match source row counts exactly, zero orphan rows across 14 foreign keys, zero duplicate keys after fixing Insurance Policy (Policy Number had 59 duplicates — resolved with a surrogate key, Policy Row ID).

Challenges Faced

What went wrong along the way, and how we fixed it

Challenge	How we solved it
59 duplicate Policy Numbers in Insurance Policy	Added a surrogate key (Policy Row ID) as the primary key
Duplicate table copies on the Tableau canvas	One box per table (10 total); extra links drawn as relationships
Power BI showed 3266% and a 0% Success Rate	Removed the extra ×100 and pointed the measure at the Outcome column
MySQL error 1146 and Power BI invalid credentials	Re-ran the schema, then fixed the MySQL login for the connector
Excel Success Rate (49.6%) differed from Tableau and Power BI (33.0%)	Traced it to the denominator and standardized on all treatments

Conclusion

One dataset, four tools, one set of numbers

Delivered end to end

Raw Excel data became a MySQL database, then Tableau and Power BI dashboards, checked by QA queries.

SQL as the source of truth

Both BI tools read from MySQL, so the same KPIs came out the same in each tool.

QA caught real issues

Duplicate keys, duplicate tables, formula bugs and a definition mismatch were all found before sign-off.

What the data says

- 32.7% of visits are completed
- 33.0% of treatments succeed
- 80.35% of patients are chronic
- $7.06M of $25.41M revenue is outstanding

Thank You

AXON Healthcare EMR Capstone — complete end to end

Questions?