END-TO-END HEALTHCARE ANALYTICS WITH SQL & POWER BI

Business Problem

City Hospital aims to deliver high-quality healthcare services while ensuring operational efficiency and financial sustainability. However, the hospital faces several challenges that impact both patient care and resource management. High no-show rates and inactive patients result in underutilized appointment slots and inefficiencies in staff allocation. At the same time, billing and collections processes face delays, partial payments, and low insurance coverage, creating financial leakage. Managing doctor performance and appointment scheduling across multiple specialties is critical to optimize hospital resources and ensure patients receive timely care. Additionally, the hospital needs structured programs to monitor chronic diseases, optimize prescription patterns, and reduce unnecessary diagnostic tests. Without a comprehensive analytics framework, these challenges limit the hospital's ability to make informed decisions and improve overall patient outcomes.

Project Objectives

- Build a complete end-to-end analytics solution to monitor patient engagement, appointment trends, and clinical care patterns.
- Analyse financial performance, including billing, payments, and insurance coverage, to identify revenue leakages and improve collections.
- Evaluate doctor performance, specialty demand, and workload distribution to optimize scheduling and resource allocation.
- Provide actionable insights into chronic disease management, prescription trends, and diagnostic testing to enhance patient care.
- Enable data-driven decision-making through interactive dashboards and visualizations using SQL and Power BI, supporting both operational and strategic initiatives.

Data Generation & Preparation

To enable end-to-end healthcare analytics, a realistic synthetic dataset was generated and systematically prepared for analysis. This dataset models real-world hospital operations, including patient care, clinical activities, and financial transactions, while intentionally incorporating imperfections to reflect practical data challenges.

1. Synthetic Data Generation

The synthetic dataset was **generated using Python with the Faker library**, which allowed for realistic data simulation with controlled imperfections.

- **Patients (1,000 records):** Included demographics, insurance type, last visit date, and status (Active/Inactive). Random missing values and noisy entries (e.g., misspelled names, missing DOB) were introduced to simulate real-world data imperfections.

- **Doctors (100 records):** Captured specialties, hospital affiliation, years of experience, and ratings. Noise was added through missing or extreme values in ratings and experience.
- **Appointments (5,000 records):** Scheduled realistically with links to patients and doctors. Captured appointment type, status (Completed, No-show, Cancelled), duration, and visit purpose.
- **Diagnoses (~3,742 records):** Aligned with **completed appointments only**. Each diagnosis includes condition, severity, ICD-10 codes, and whether additional tests were required.
- **Prescriptions (~3,742 records):** Linked to completed appointments and diagnoses. Included medication, dosage, frequency, and treatment duration.
- **Billing (~3,742 records):** Corresponds to completed appointments. Modelled cost logic based on appointment type, diagnosis severity, and tests. Included insurance coverage, patient payments, payment methods, due dates, and data imperfections such as negative amounts and misspelled statuses.

2. Data Preparation for Analysis

The synthetic data was imported into a MySQL relational database, cleaned, and structured to ensure consistency, accuracy, and relational integrity:

- **Database & Schema Design:** Six interlinked tables—Patients, Doctors, Appointments, Diagnoses, Prescriptions, Billing—with primary and foreign keys.
- **Data Cleaning & Standardization:**
    - *Patients:* Corrected missing DOB, standardized phone numbers, addresses, and insurance type; calculated age groups for segmentation; validated Active/Inactive status based on last visit.
    - *Doctors:* Imputed missing years of experience, cleaned specialties, and validated ratings.
    - *Appointments:* Removed nulls, ensured valid status and type, and extracted monthly and quarterly trends.
    - *Diagnoses:* Standardized diagnosis codes and ensured alignment with appointment and doctor specialty.
    - *Prescriptions:* Cleaned medication names, dosages, and frequencies.
    - *Billing:* Corrected negative amounts, enforced consistency between settled amounts, insurance, and paid values; standardized Payment Status and Payment Method fields.
- **Data Quality Validation:** Cross-checked all tables for logical consistency (e.g., Last Visit after DOB, billing totals matching insurance and patient payments). Standardized categorical fields and validated patient and doctor metrics.

3. Export for Power BI

Once cleaned, the tables were ready for analytics and connected directly to Power BI for visualization:

- Patients
- Doctors
- Appointments
- Diagnoses
- Prescriptions
- Billing

This rigorous data generation and preparation process ensured that the dataset was realistic, relationally consistent, and ready for business intelligence analysis, forming the foundation for SQL-based insights and Power BI dashboards.

Power BI Dashboards – Insights & Recommendations

Dashboard 1: Patient & Appointment Trends

Key Insights

- The hospital serves 1,000 patients, of which 994 have at least one appointment.
- Active patients: 597; Inactive patients: 397 → ~40% inactive, highlighting re-engagement opportunities.
- Appointment completion rate: 75%, with 747 no-shows, indicating operational inefficiencies and underutilized doctor capacity.
- Yearly appointments: Stable in 2023 (2,114) and 2024 (2,121), lower in 2025 (765) due to partial-year data.
- Appointment types distribution: Routine Check-ups (34%), Follow-ups (32%), Emergencies (27%) → well-balanced service utilization.
- Patient demographics: Gender is balanced (51% male, 49% female). Age distribution shows a skew toward older patients: 66+ (32%), 18–35 (21%), 36–50 (19%).
- Insurance coverage: Medicare (31%), Private (29%), None (30%), Unknown (10%) → significant uninsured patient population.

Business Challenges

- High no-show rate reduces operational efficiency.
- Substantial inactive patient base (~40%).
- High proportion of uninsured patients increases financial risk.

Recommendations

- Implement automated appointment reminders via SMS, email, or app notifications to reduce no-shows.
- Launch patient re-engagement campaigns targeting inactive patients (e.g., health check packages or wellness programs).
- Explore partnerships with insurers or introduce affordable care plans for uninsured patients.

Dashboard 2: Financial Performance & Billing Insights

Key Insights

- Total billed: ₹13.93M; Total settled: ₹10.31M → ~26% unpaid dues (₹3.62M).
- Insurance contribution: ₹5.41M; patient out-of-pocket payments: ₹4.90M.
- Yearly billing: 2023 (₹5.91M), 2024 (₹5.87M), 2025 partial (₹2.15M).
- Payment status: 54% partially paid, 37% fully paid, 8% pending → indicates affordability gaps or claim delays.

- Most common payment methods: Credit Card (1,343), Online (1,045), Cash (1,037), Not applicable (308, insurance-covered).
- Insurance vs. Paid amounts indicate that coverage is insufficient in several cases, increasing patient liability.

Business Challenges

- High outstanding dues and partial payments cause cash flow issues.
- Low insurance coverage for many patients increases financial risk.
- Delays or inefficiencies in payment collection affect revenue stability.

Recommendations

- Strengthen collections via automated reminders and flexible payment plans.
- Negotiate better insurance coverage agreements to reduce out-of-pocket expenses.
- Incentivize online and cashless payment options for faster settlements.

Dashboard 3: Doctor, Diagnosis & Prescription Overview

Key Insights

- Total doctors: 100; Total revenue: ₹10.31M.
- Top specialty: Cardiology (₹1.99M revenue).
- Average doctor experience: 23 years; average rating: 3.87/5.
- Doctor performance: Top revenue doctors differ from top appointment volume doctors → volume vs revenue misalignment.
- Common diagnoses: Asthma (381), Diabetes (335), Migraine (296), Hypertension (276), Thyroid Disorder (198) → chronic diseases dominate.
- Additional tests required for 48% of diagnoses, indicating higher diagnostic workload and patient costs.
- Prescription trends: Most frequent medications – Levothyroxine (198), Metformin (172), aligned with chronic conditions; majority prescribed for 15–30 days; daily dosage trends: Once a day (21,435 units), Twice a day (17,667 units).

Business Challenges

- Misalignment between appointment volume and revenue generation among doctors.
- Chronic diseases dominate, requiring long-term management and monitoring.
- High dependency on additional tests increases patient cost and hospital workload.

Recommendations

- Implement chronic disease management programs (e.g., diabetes, hypertension, asthma) to reduce readmissions.
- Review doctor performance metrics to balance appointment volume and revenue targets.
- Optimize diagnostic testing protocols to avoid unnecessary tests while ensuring quality care.
- Introduce preventive health packages to reduce chronic disease progression.

Combined Insights & Strategic Recommendations

Key Observations

1. **Operational Gaps:** High no-show rates (~15%) and inactive patients (~40%) reduce efficiency.
2. **Financial Leakages:** ₹3.62M unpaid dues (26%) with majority partially paid bills; low insurance coverage.
3. **Chronic Disease Burden:** Asthma, Diabetes, Hypertension, and Migraine dominate patient load.
4. **Doctor Performance Misalignment:** Appointment volume does not directly correlate with revenue.

Strategic Recommendations

- **Patient Engagement:** Digital reminders, re-engagement campaigns, wellness initiatives.
- **Financial Optimization:** Strengthen collections, expand insurance coverage, promote cashless and digital payments.
- **Chronic Care Programs:** Long-term disease management plans, regular follow-ups, preventive care packages.
- **Resource & Performance Management:** Align doctor workload with revenue goals; optimize diagnostic test utilization.

Conclusion

This end-to-end healthcare analytics project demonstrates how a structured, relational dataset combined with SQL analysis and Power BI visualizations can provide actionable insights across patient engagement, financial performance, and clinical operations. By identifying operational inefficiencies, financial leakages, and chronic disease patterns, the hospital can make data-driven decisions to improve patient care, optimize resource utilization, and enhance revenue management. The dashboards and recommendations form a foundation for ongoing monitoring, strategic planning, and continuous improvement in healthcare delivery.