Predictive Bed Occupancy & Patient Flow Optimization Using Hospital Inpatient Data

Hospitals often struggle with inefficient utilization of beds and critical resources despite having adequate physical capacity. This project focuses on a large hospital ecosystem handling diverse admission types, including emergency, urgent, elective, and newborn cases. While patient inflow remained high, beds, diagnostic equipment, and rooms frequently stayed idle due to bottlenecks in earlier stages of patient movement such as delayed discharges, documentation issues, housekeeping turnaround, and administrative clearances.

The absence of predictive insights into patient length of stay and future bed availability forced operations teams to rely on reactive decisions. This led to delayed admissions, emergency congestion, and suboptimal resource utilization, directly impacting both patient experience and hospital revenue.

Objectives

Predict Patient LOS

Predict patient length of stay (LOS) using historical and clinical indicators.

Identify Delays

Identify delays caused by operational and administrative inefficiencies.

Quantify Bed Occupancy

Quantify bed occupancy on any selected date.

Forecast Bed Occupancy

Forecast bed occupancy for the next 7 days based on predicted admissions and discharges.

Enable Proactive Planning

Enable proactive planning for admissions, staffing, and housekeeping.

Reduce Idle Bed Time

Reduce idle bed time and improve patient throughput.

Dataset Overview

Source & Provider

- Source: Hospital Inpatient Discharges (SPARCS), De-Identified
- Provider: New York State Department of Health (NYSDOH)
- Data Access: Public API using the sodapy library

Time Range & Size

- Time Range: Year 2012
- Size: ~250,000 records with ~28 core attributes

Key Data Coverage (summarized)

- Patient demographics (age group, gender)
- Admission details (emergency, urgent, elective)
- Diagnosis and procedure categories
- Severity of illness and mortality risk
- Facility and service area information
- Length of stay (LOS), total cost, and total charges

Key Fields Used for Analysis & Limitations

Key Fields Used for Analysis

- Facility identifiers
- Age
- Diagnosis categories
- Medical/surgical classification
- Severity of illness
- Length of stay
- Total cost
- Total charges

Limitations

The source data did not include exact admission or discharge dates; only the year was provided. Date-level timelines were simulated and enhanced for downstream modeling and forecasting.

The cleaned and standardized dataset was stored in a PostgreSQL (NeonDB) database for scalable querying and analysis.

Tools Used

Programming & Analysis

Python (Pandas, NumPy)

Machine Learning

Scikit-learn (Gradient Boosting Regressor)

Database

PostgreSQL (NeonDB)

Data Access

REST APIs via sodapy

Visualization

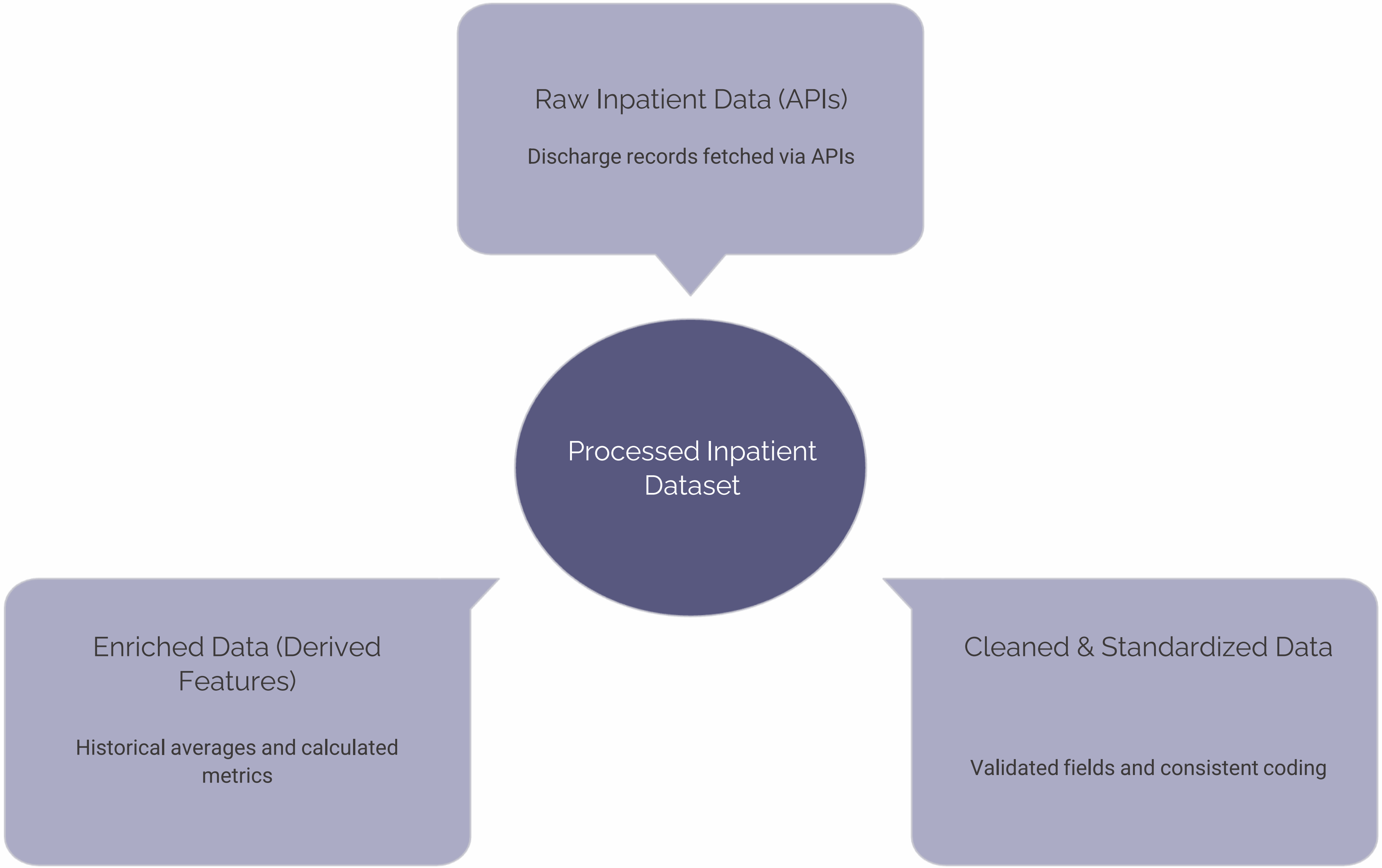
Plotly, Matpltlib, Power BI

Environment

Google Colab

Methodology: Data Processing

The analysis followed a structured end-to-end analytics workflow. Raw inpatient discharge data was fetched via APIs, cleaned, standardized, and enriched with derived features such as historical average LOS by diagnosis and facility.



Methodology: Machine Learning Models

Two machine learning models were developed.

1

Enhanced Gradient Boosting Model

Trained using rich historical features (including cost, charges, severity, and diagnosis patterns) to accurately predict length of stay and quantify delays between expected and actual LOS. These delays were used as a proxy for operational inefficiencies.

2

Admission-Time LOS Model

Built using only features available at patient entry. This enabled early estimation of discharge timelines without relying on post-admission financial data.

Predicted LOS values were combined with admission dates to estimate discharge dates.

Using these timelines, bed occupancy was calculated for any target date and forecasted for the next seven days. The final outputs were presented through dashboards and trend visualizations designed for operational decision-makers.

Key Insights

- Length of stay is strongly influenced by severity of illness, diagnosis category, and facility-level historical patterns.
- Respiratory distress syndrome and administrative/social diagnosis categories consistently show longer LOS.
- A significant portion of delays occur in medical cases with no major procedures, indicating non-clinical causes such as paperwork, payment, or discharge coordination.
- Elderly patients (70+) and emergency or urgent admissions contribute disproportionately to bed blocking.
- Newborn cases involving very low birth weight and major procedures experience the longest delays due to extended clinical monitoring and specialized care needs.

Key Insights & Recommendations

Financial Impact of Delays

Delay-driven bed blocking resulted in an estimated ₹30.23 million opportunity cost in the year of 2012.

Analysis of excess Length of Stay (LOS), calculated as the difference between predicted and actual discharge days and multiplied by cost per bed-day, revealed a total delay-related financial impact.

High-complexity procedures contribute disproportionately to delays. Kidney Transplants: ₹1,10,000 & Peripheral Vascular Bypass: ₹1,00,000 contributing in delay-driven opportunity cost.

Recommendations

- Use admission-time LOS predictions to plan discharges proactively rather than reactively.
- Prioritize operational reviews for medical cases with unusually long stays and no critical procedures.
- Introduce process tracking for administrative and social diagnosis categories to reduce non-clinical delays.
- Align housekeeping, transport, and discharge documentation schedules with predicted discharge timelines.
- Use bed occupancy forecasts in daily operational meetings to manage staffing and admission planning.

Business Impact

By leveraging predictive LOS modeling and bed occupancy forecasts, the hospital can significantly reduce idle beds and optimize resource utilization, recovering over ₹30M in potential opportunity costs.

The analysis highlights operational bottlenecks in discharge, documentation, and patient flow, enabling targeted process improvements that increase bed turnover. It also supports better management of emergency and urgent admissions, improving patient throughput and experience, while providing facility- and procedure-specific insights for informed decision-making and strategic resource allocation.