INVENTORY OPTIMIZATION REPLENISHMENT SIMULATION FOR HOSPITAL PHARMACY

CAPSTONE PROJECT REPORT SUBMITTED

TO THE

INDIAN SCHOOL OF BUSINESS

FOR THE

ADVANCED MANAGEMENT PROGRAMME IN BUSINESS ANALYTICS

SUBMITTED BY

ANMOL CHHABRA [12410015]

DIVAM GUPTA [12410008]

ISHIKA AGGARWAL [12410085]

JANKIBEN ACHARYA [12320026]

RAHUL SHARMA [12410051]

UNDER THE GUIDANCE OF

PROF. NAGADEVARA V.

INSTITUTE OF DATA SCIENCE

INDIAN SCHOOL OF BUSINESS

OCTOBER 2025

CAPSTONE PROJECT REPORT

SUBMITTED TO THE INDIAN SCHOOL OF BUSINESS FOR ADVANCED MANAGEMENT PROGRAMME IN BUSINESS ANALYTICS

HEADING	DETAILS
Title of the Capstone project	Inventory Optimization Replenishment Simulation for Hospital Pharmacy
Name of the Candidate	**Anmol Chhabra** **Divam Gupta** **Ishika Aggarwal** **Jankiben Acharya** **Rahul Sharma**
Name and Affiliation of the Faculty Mentor	**Prof. Nagadevara. V.**
Name and Affiliation of the Industry Mentor	Praveen Kumar Burra
Batch	22
Date/Month of submission of Capstone Project Report	**26-10-2025**

Acknowledgement

We would like to express our heartfelt gratitude to the **ISB** for providing us with the opportunity to undertake the capstone project.

Our sincere thanks to our **Faculty Mentor, Prof. Nagadevara. V** for their invaluable guidance, encouragement, and continuous support throughout the project. Their academic insights and constructive feedback greatly contributed to the quality and direction of our work.

We are also deeply thankful to our **Industry Mentor, Mr. Praveen Kumar Burra**, for offering practical perspectives and sharing real-world experience that enriched our understanding of the project domain.

Finally, we extend our appreciation to our **Sponsor, AiSPRY**, for their collaboration, valuable inputs, and support, which played a crucial role in the successful completion of this project.

Anmol Chhabra
Divam Gupta
Ishika Aggarwal
Jankiben Acharya
Rahul Sharma
October 2025

Table of Contents

List of Tables

List of Figures

Abstract

Hospital pharmacies often face stockouts of essential medicines, leading to high patient bounce rates, revenue loss, and compromised patient care. This capstone project, titled "Inventory Optimization and Replenishment Simulation for Hospital Pharmacy," aims to develop a data-driven analytics solution to enhance inventory management efficiency and decision-making at the pharmacy level.

The solution integrates descriptive analytics to analyze stock movement and bounce trends, predictive modeling to forecast SKU-level demand, and prescriptive logic to recommend optimal reorder quantities and timing. Additionally, a simulation engine has been built to evaluate and compare multiple inventory policies under different demand and lead-time conditions.

By combining analytical insights with operational practicality, the project provides a robust framework for intelligent pharmacy operations enabling proactive stock planning, reduce patient bounce rates, and improved cost efficiency. The outcome contributes both academically, through its analytical rigor, and practically, by demonstrating measurable business impact for healthcare supply chain management.

Key words:

Inventory Optimization, Demand Forecasting, Replenishment Planning, Simulation, Hospital Pharmacy, Stockout Reduction, Predictive Analytics, Prescriptive Analytics, Data-Driven Decision Making, Healthcare Supply Chain

Executive Summary

Hospital pharmacies play a crucial role in ensuring timely availability of medicines for inpatient and outpatient care. However, inventory inefficiencies such as overstocking, stockouts, vendor delays, and lack of visibility into prescribing patterns often result in high bounce rates and lost revenue opportunities.

This project, titled "**Inventory Optimization and Replenishment Simulation for Hospital Pharmacy**," was undertaken in collaboration with AiSPRY Technologies Pvt. Ltd. to develop a data-driven decision-support system capable of forecasting demand, optimizing replenishment, and providing actionable insights through an integrated dashboard.

The objective was to design a scalable analytics solution that could:

- Predict SKU-level demand using advanced forecasting models.
- Simulate replenishment policies to reduce stockouts and holding costs.
- Generate interactive insights across five key business dimensions—department performance, SKU consumption, physician behaviour, vendor reliability, and margin analysis.

The project combined descriptive, predictive, and prescriptive analytics methodologies and followed a CRISP-DM framework encompassing business understanding, data preparation, modeling, evaluation, and visualization.

1. Department-Wise Insights

Department-level analysis revealed significant heterogeneity in both revenue contribution and operational efficiency.

- Cardiology and Internal Medicine emerged as the top-performing departments, together contributing over 60% of total pharmacy revenue.

- Cardiology and Pediatrics exhibited higher bounce rates (13–14%), indicating recurring medicine unavailability for high-turnover prescriptions.
- A positive correlation (r = 0.62) between departmental transaction volume and bounce rate implied that peak operational departments are most affected by stock shortages.

Managerial Implication:

Hospitals should prioritize inventory optimization for high-demand departments like Cardiology, Pediatrics, and Emergency Medicine by implementing predictive reorder alerts and department-specific stock thresholds.

Recommendation:

Implement department-level dashboards to monitor bounce incidents, lead time adherence, and cost recovery, enabling pharmacy managers to proactively plan replenishment.

2. SKU Clustering and Consumption Patterns

Using K-Means clustering on SKU-level sales, three distinct consumption patterns were identified:

- Cluster 0 (Stable SKUs): 1,029 items with consistent monthly sales.
- Cluster 1 (Rising SKUs): 144 items showing demand growth, typically associated with chronic and seasonal medicines.
- Cluster 2 (Declining SKUs): 24 items exhibiting demand deterioration, mostly elective or short-shelf-life products.

The top 10 SKUs contributed nearly 28% of revenue, reinforcing the 80-20 rule in inventory contribution.

Cluster-based replenishment strategies were simulated to align order frequency with demand volatility, resulting in NRMSPE < 20% accuracy in forecasted consumption.

Managerial Implication:

Adopt differentiated replenishment policies, tighter cycles for rising SKUs and relaxed thresholds for stable or declining items.

Recommendation:

Integrate SKU clustering into the pharmacy management system to auto-classify new medicines based on sales trajectories.

3. Physician Prescription Trends

Physician-level insights revealed that 10% of doctors contributed nearly 15% of total prescriptions. Prescription behavior showed variability in formulary adherence and average transaction value:

- Nephrology and General Surgery displayed the highest adherence (~71%), while Emergency Medicine had the lowest (44%).
- Non-adherence was strongly associated with bounce occurrences and lost margins, suggesting the operational cost of off-formulary prescribing.

Managerial Implication:

Pharmacies should engage with high-prescribing physicians to standardize formulary usage and ensure stocked drugs align with treatment patterns.

Recommendation:

Use prescription dashboards to track adherence rate, average prescription value, and SKU overlap between formulary and demand, thereby improving alignment between prescribing and stocking policies.

4. Vendor and Purchase Order Efficiency

Vendor performance analysis showed average lead time of 27 days and on-time delivery rate of only 14%, indicating major procurement inefficiencies.

Vendor dependency analysis further revealed that top 10 vendors accounted for over 60% of inventory value, implying supply concentration risk.

A negative correlation ($r = -0.63$) between on-time delivery rate and bounce rate established that delivery delays are a key driver of patient non-fulfilment.

Managerial Implication:

Improving supplier reliability has a direct impact on reducing bounce rates and ensuring operational stability.

Recommendation:

Introduce a Vendor Reliability Index (VRI) on the dashboard that combines on-time rate, lead time deviation, and quality score to support vendor evaluation and contract renegotiation.

5. Margin and Profitability Analysis

Margin analytics at departmental and physician levels revealed imbalanced profitability across segments:

- Cardiology and Oncology offered the highest gross margins (15–18%), while Emergency and Pediatrics remained below 10%.
- Quarterly analysis (MoM and QoQ) demonstrated seasonal dips in Q2 due to procurement delays and higher vendor costs.

Managerial Implication:

Margin variance highlights the need for dynamic pricing, improved procurement timing, and discount optimization.

Recommendation:

Deploy real-time margin dashboards tracking MoM and QoQ shifts by department and physician to identify underperforming segments early and align procurement and pricing strategies.

Conclusion

Through integrated analytics, the project demonstrated that data-driven inventory management can substantially enhance hospital pharmacy performance.

The developed dashboard consolidates five critical dimensions—departmental efficiency, SKU consumption, physician prescribing, vendor reliability, and margin analysis—into a single decision-support platform.

By adopting predictive forecasting, dynamic clustering, and vendor performance monitoring, hospitals can significantly reduce stockouts, improve patient satisfaction, and increase revenue through better inventory control and procurement planning.

Chapter 1: Introduction

1.1 Domain-Specific Terminologies

Table 1.1 Domain Specific Terminologies

Terminology Number	Terminology Name	Definition
1.1.1	Pharmacy Inventory	Refers to the collection of all pharmaceutical items, consumables, and medical supplies maintained by the hospital pharmacy for patient dispensing and internal use.
1.1.2	SKU (Stock Keeping Unit)	A unique identifier assigned to each medicine or consumable item in the pharmacy's database for tracking stock, demand, and sales.
1.1.3	Reorder Level (ROL)	The predefined stock threshold that triggers a replenishment order to avoid stockouts.
1.1.4	Days of Stock (DOS)	Indicates how many days existing inventory will last based on current consumption rates.
1.1.5	Lead Time	The total time elapsed between placing a purchase order and receiving the stock from the vendor.
1.1.6	Bounce Rate (Pharmacy Context)	Percentage of patients who leave without purchasing their prescribed medicines due to stock unavailability.
1.1.7	Formulary Adherence	The degree to which physicians prescribe medicines listed in the hospital's approved formulary, promoting standardization and cost efficiency.
1.1.8	Vendor Reliability Index (VRI)	A composite metric capturing a vendor's delivery performance, on-time rate, and lead time consistency.
1.1.9	Departmental Consumption	Aggregated usage of drugs or consumables by specific departments such as Cardiology, Internal Medicine, or Oncology.
1.1.10	Patient Bounce Impact	Financial loss associated with unfulfilled prescriptions resulting from stockouts or non-adherence.

1.2 Analytical and Algorithmic Terminologies

Table 1.2 Analytical and Algorithmic Terminologies

Terminology Number	Terminology Name	Definition

1.2.1	Time-Series Forecasting	A predictive modeling technique that uses historical time-ordered data to forecast future values, such as SKU-level daily demand.
1.2.2	ARIMA (Auto-Regressive Integrated Moving Average)	A statistical model combining autoregression, differencing, and moving averages to forecast stationary time-series data.
1.2.3	Exponential Smoothing (ETS)	A method that forecasts data using weighted averages of past observations, giving higher weights to more recent data.
1.2.4	Prophet Model	A time-series forecasting algorithm developed by Meta that decomposes demand into trend, seasonality, and holiday effects.
1.2.5	XGBoost (Extreme Gradient Boosting)	A machine learning algorithm based on ensemble decision trees that capture complex, nonlinear relationships in data.
1.2.6	Ensemble Model	A hybrid model that combines predictions from multiple algorithms (ETS, Prophet, XGBoost) to improve overall accuracy and robustness.
1.2.7	Lag Feature	A variable representing past values of a time-series used as input for predicting future demand.
1.2.8	Rolling Mean / Rolling Standard Deviation	Statistical transformations that calculate moving averages or volatility over a fixed time window (e.g., 7 days).
1.2.9	SHAP (SHapley Additive exPlanations)	A model interpretability framework that explains each feature's contribution to a prediction in machine learning models.
1.2.10	DBT (Data Build Tool)	A transformation tool used for managing and automating SQL transformations in data pipelines.
1.2.11	Apache Airflow	A workflow orchestration tool that automates and schedules data processing and model execution tasks.
1.2.12	Apache Superset	An open-source data visualization platform used to build interactive dashboards.

1.3 Performance and Statistical Metrics

Table 1.3 Performance and Statistical Metrics Terminologies

Terminology Number	Terminology Name	Definition
1.3.1	RMSE (Root Mean Square Error)	A measure of forecast accuracy that penalizes large prediction errors more heavily.
1.3.2	NRMSE (Normalized RMSE)	RMSE is normalized by the range or mean of the actual data, enabling comparison across SKUs with different scales.

1.3.3	MAPE (Mean Absolute Percentage Error)	Average of absolute percentage differences between predicted and actual values.
1.3.4	SMAPE (Symmetric Mean Absolute Percentage Error)	A balanced error metric that treats over- and under-predictions equally.
1.3.5	WMAPE (Weighted Mean Absolute Percentage Error)	A weighted version of MAPE that gives higher importance to high-volume SKUs.
1.3.6	R^2 (Coefficient of Determination)	Represents how much variance in the dependent variable is explained by the model.
1.3.7	PCA (Principal Component Analysis)	A dimensionality reduction technique that converts correlated features into a set of uncorrelated components.
1.3.8	Correlation Coefficient (r)	Quantifies the linear relationship between two numerical variables.
1.3.9	VIF (Variance Inflation Factor)	A measure to detect multicollinearity among explanatory variables in regression models.
1.3.10	Cross-Validation (Time-Series Split)	A validation method ensuring that training and testing sets maintain chronological order to prevent data leakage.

1.4 Managerial and Operational Terminologies

Table 1.4 Managerial and Operational Terminologies

Terminology Number	Terminology Name	Definition
1.4.1	Vendor Reliability Index (VRI)	Quantitative measure of vendor performance combining on-time delivery and consistency metrics.
1.4.2	Reorder Policy	Business rule that dictates when and how much stock should be ordered to maintain adequate availability.
1.4.3	Procurement Lead Time Variability	Variation in the time taken between placing and receiving an order, affecting forecast accuracy and safety stock levels.
1.4.4	Holding Cost	Cost incurred in storing and maintaining inventory, including space, insurance, and depreciation.
1.4.5	Stockout Cost	Financial and service-level impact caused by the unavailability of a required SKU.
1.4.6	Margin Analysis	Evaluates profit contribution per SKU or department after deducting procurement and holding costs.
1.4.7	Vendor SLA (Service Level Agreement)	Contractual metrics defining expected vendor performance in terms of delivery time, fill rate, and quality.
1.4.8	Simulation Model	A computational model designed to test “what-if” scenarios for procurement and replenishment under varying conditions.

1.4.9	MoM and QoQ Analysis	Month-over-Month and Quarter-over-Quarter performance comparison used for trend monitoring.
1.4.10	Pareto Principle (80–20 Rule)	Business principle that states roughly 80% of outcomes are driven by 20% of causes—applied to identify high-impact SKUs or vendors.
1.4.11	Docker	A containerization platform used to deploy and manage isolated application environments for reproducible workflows.
1.4.12	PostgreSQL	An open-source relational database used for storing, transforming, and querying large-scale transactional data.
1.4.13	FastAPI	A Python framework used for creating APIs to serve machine learning models and enable integration with dashboards.
1.4.14	Data Pipeline	A set of connected processes that extract, transform, and load (ETL) data from source systems to analytics destinations.
1.4.15	Incremental Data Load	Process of adding only new or modified records to the database during each ETL run.
1.4.16	Model Retraining Cycle	Scheduled process that updates the forecasting model periodically as new data becomes available.
1.4.17	Dashboard Automation	The process of integrating analytical outputs with visualization tools to build continuously updating dashboards that refresh new data or model results without requiring manual effort.
1.4.18	Open-Source Stack	A suite of freely available software tools (PostgreSQL, Airflow, DBT, Superset) used to build scalable and cost-effective solutions.

Chapter 2: Review of Related Literature

This chapter reviews existing literature, industry case studies, and reference frameworks that shaped the development of the *Inventory Optimization and Replenishment Simulation for Hospital Pharmacy*.

The goal of this literature review is to contextualize the six major use cases—Inventory Optimization, Consumables Mapping, Prescription Pattern Analytics, ML-driven Procurement Forecasting, Patient Bounce Rate Reduction, and Vendor SLA Optimization—within the broader scope of healthcare analytics, supply chain management, and data-driven decision systems.

2.1 Inventory Optimization in Hospital Pharmacies

Overview:
Inventory optimization in the pharmaceutical context aims to ensure the right drug is available at the right time and quantity while minimizing holding and stockout costs. Studies such as *Shah et al. (2020)* and *Kumar & Narang (2019)* highlight that overstocking leads to high expiry losses, while understocking results in poor patient satisfaction and lost revenue.

Key Findings from Literature:

- Machine learning–driven demand forecasting has proven effective in dynamic healthcare environments where seasonal variability and patient inflow significantly affect demand.
- Forecasting models such as **ARIMA**, **Prophet**, and **XGBoost** have been used to predict SKU-level consumption, reducing inventory-related wastage by 15–25%.
- Integration with hospital ERP systems allows automated reorder triggers when stock reaches AI-defined thresholds.

Relevance to the Project:
This study forms the foundation for developing an AI-based reorder and replenishment policy for hospital pharmacies. The project's proposed model extends this by simulating reorder decisions under variable lead times and supplier reliability scenarios, aligning with the *real-time inventory optimization* approach described in global best practices.

COGNIGEN: PHARMACY & CONSUMABLES MODULE			
Sl No	Use Case	Problem	Anlytical solution we propose
1	Inventory Optimisation	Frequent stock-outs or overstocking of low-demand medicines. Pharmacies often struggle with keeping the right level of inventory due to poor visibility into real-time demand. This leads to either stock-outs (missed sales, patient dissatisfaction) or overstocking of slow-moving items (expiry and wastage).	Our application continuously anlyses inventory levels, consumption velocity, reorder lead times, and historical trends. ML will be used to study predict the trends. The platform will trigger alerts before stock reaches critical levels or when stockpiling of low-moving SKUs is detected..
2	Consumables Mapping at Patient/Dept level	Lack of visibility into inventory consumption at department and patient level. Pharmacies/ central stores are unable to track which departments or procedures consume more supplies, leading to poor accountability, budgeting challenges, and difficulty in identifying overuse or misuse.	Generate detailed reports that map pharmacy and consumable usage to patients, procedures, and departments. Compare these against clinical benchmarks/ inter-department benchmarks to flag abnormal consumption, helping drive accountability, cost control, and process improvements.
3	Prescription Pattern Analytics	Inconsistent prescription patterns by physicians. Variation in prescribing habits—such as different brands for the same molecule or frequent changes in drug preferences—create difficulties for pharmacies in forecasting demand, negotiating supplier rates, and managing stock. This leads to higher inventory costs, increased wastage, and revenue leakage.	Our anlytical platform will track prescription patterns by physician across departments over time. Alert pharmacy and medical leadership when deviations from standard protocols or sudden shifts occur. Insights can be used for formulary standardization, procurement planning, and clinical governance.
4	ML Driven Procurement Forecasting	Lack of scientific forecasting in procurement planning. Inventory ordering is often reactive or based on manual judgment, resulting in numerous small, uncoordinated orders that increase procurement costs and lead to inconsistent supply availability.	we will use machine learning models trained on historical consumption, patient volume trends, seasonal disease patterns, and prescription history to generate SKU-level forecasts. This enables pharmacies to plan bulk purchases in advance, reduce order frequency, lower per-unit cost, and maintain optimal stock levels.
5	Reducing Patient Bounce Rate and Revenue	High patient bounce rate due to unavailability of prescribed drugs. Patients often leave without purchasing medications if items are unavailable, leading to revenue loss and poor patient experience.	Reports to track bounce rates per SKU and physician, then align pharmacy inventory dynamically with top-prescribed and high-demand drugs. Enables higher fulfillment rates and improves patient retention.
6	Vendor SLA optimisation	Lack of vendor performance tracking and supplier quality management. Procurement teams may continue sourcing from suppliers with inconsistent delivery timelines, pricing variations, or quality issues, impacting pharmacy efficiency.	Our performance dashboard will track delivery timeliness, fill rate, cost fluctuations, and quality issues across vendors. Use data to guide vendor selection and enforce SLAs.

Fig 2.1: Pharmacy & Consumables Analytics Framework (Client Input)

2.2 Consumables Mapping at Patient and Department Level

Overview:
Literature in hospital supply chain analytics (e.g., *Narayanan et al., 2018*) emphasizes the need for granular visibility of consumable usage across departments, procedures, and patients. Without such traceability, hospitals face budget overruns and cannot attribute costs accurately.

Key Findings from Literature:

- Advanced analytics and anomaly detection techniques (Isolation Forest, LOF) have been successfully applied to detect abnormal consumable usage patterns.
- Department-wise dashboards and benchmark-based comparisons can identify overuse and process inefficiencies, leading to 8–12% cost savings in consumables.
- Linking pharmacy data with EMR (Electronic Medical Record) and ADT (Admission-Discharge-Transfer) systems enhances cost allocation accuracy and traceability.

Relevance to the Project:
This use case inspired the inclusion of **consumable benchmarking and anomaly detection** features in the dashboard, helping hospitals allocate costs correctly and reduce misuse.

2.3 Prescription Pattern Analytics

Overview:
Variations in prescribing behaviour among physicians significantly impact stock movement and procurement planning. Studies such as *Saha et al. (2021)* identify formulary adherence and brand fragmentation as key contributors to inventory inefficiency.

Key Findings from Literature:

- Hospitals adopting **formulary adherence monitoring** and **salt-level analytics** reported a 20% improvement in fill rates.
- Analytical solutions focusing on brand vs. generic preference help in reducing inventory fragmentation.
- Integration of antibiotic stewardship analytics (e.g., AWaRe mix, DDD/100 bed-days) supports both clinical and operational efficiency.

Relevance to the Project:
This literature guided the **Prescription Pattern Analytics** use case by highlighting how physician prescribing patterns influence inventory forecasts and patient bounce rates. The project applies similar logic using salt-level mapping and formulary adherence indicators to forecast demand more accurately.

2.4 ML-Driven Procurement Forecasting

Overview:
Procurement forecasting literature emphasizes data-driven purchase planning as a way to stabilize supply and reduce procurement costs (*Chakraborty et al., 2020*). Traditional procurement relies on static reorder levels, which fail under fluctuating demand or lead times.

Key Findings from Literature:

- Time-series models (ARIMA, Prophet) combined with machine learning (XGBoost, LSTM) have demonstrated strong performance in healthcare procurement forecasting.

- Simulation-based optimization, using frameworks such as **SimPy** or **PuLP**, helps evaluate "what-if" scenarios for lead time delays or bulk purchase discounts.
- Forecast-driven procurement reduces emergency purchase frequency by up to 30% and improves SLA adherence.

Relevance to the Project:
The project adopts a similar hybrid forecasting-simulation framework, integrating cost, service level, and vendor performance to recommend optimized purchase orders.

2.5 Reducing Patient Bounce Rate and Revenue Leakage

Overview:
Patient bounce rate patients leaving without purchasing prescribed medicines is a significant pain point in hospital pharmacies. Prior studies (e.g., *Patel et al., 2019*) show that up to 15–20% of patients abandon purchases due to stock unavailability or lack of substitutes.

Key Findings from Literature:

- Predictive models combining SKU availability, prescription specificity, and patient acceptance probabilities can predict bounce likelihood with >80% accuracy.
- Dynamic reorder policies driven by forecasted bounce risk improve fill rate and patient satisfaction.
- Integration of real-time dashboards showing "top bounce SKUs" helps target root causes and prevent lost revenue.

Relevance to the Project:
This insight led to the design of the **bounce rate reduction module**, using machine learning classification models to estimate bounce probability and suggest real-time corrective actions such as substitution or inter-store transfer.

2.6 Vendor SLA Optimization

Overview:
Vendor management literature underscores the role of SLA (Service Level Agreement) adherence in ensuring pharmacy efficiency. Studies like *Bhattacharya & Rao (2021)* identify vendor performance monitoring as a key differentiator in procurement reliability.

Key Findings from Literature:

- SLA dashboards monitoring **on-time delivery rate (OTD)** and **fill rate** increase supplier reliability by 15–25%.
- Vendor segmentation (Preferred / Conditional / Blacklisted) based on SLA performance enhances procurement decision-making.
- Clustering and KPI-based scoring can optimize vendor selection, improving both quality and cost consistency.

Relevance to the Project:

This forms the foundation of the **Vendor SLA optimization module** the final analytical block in the dashboard tracking on-time delivery, fill rate, cost variance, and quality issues to improve vendor accountability.

Chapter 3: Project Description

3.1 Business/Research Problem

Hospital pharmacies often struggle to maintain optimal inventory levels due to fluctuating patient demand, varying prescription patterns, and procurement delays. These inefficiencies frequently result in **stockouts of essential medicines**, leading to **high patient bounce rates** where patients leave without receiving prescribed drugs. This not only causes **revenue loss** but also affects **continuity of patient care** and overall **hospital reputation**.

On the other hand, overstocking to prevent shortages leads to **excess inventory**, **increased holding costs**, and **wastage due to expiry**. Thus, hospitals face the dual challenge of balancing **availability and cost efficiency** in their pharmacy operations.

The project titled **"Inventory Optimization and Replenishment Simulation for Hospital Pharmacy"** aims to develop a **data-driven analytical framework** that combines **descriptive**, **predictive**, and **prescriptive analytics** to address these challenges. The goal is to:

- Forecast demand accurately at the SKU level,
- Identify factors contributing to patient bounce rates,
- Recommend optimal reorder quantities and timing, and
- Simulate and compare various inventory management policies.

By building a replenishment simulation model, this project seeks to support **intelligent, data backed decision-making** that minimizes stockouts, reduces wastage, and ensures continuous medicine availability ultimately improving both **operational efficiency** and **patient satisfaction** within hospital pharmacies.

3.2 Data Available

Overview of Data Sources

The data for this project was provided as part of a simulated and anonymized dataset re presenting a hospital's pharmacy supply chain. The dataset integrates multiple interlin ked files
covering hospital operations, vendor performance, pharmaceutical stock, and transacti onal activity.

Each file corresponds to a distinct operational unit, together forming a comprehensive view of hospital pharmacy management.

Table 3.2.1 File Name and Description

File Name	Description
Hospital	Contains details such as hospital ID, name, location, bed capacity, establishment year, and annual budget. Used to understand institutional characteristics influencing inventory planning.
Departments	Captures departmental data, including department type, bed count, and monthly budget, which may indirectly affect medication consumption patterns.
Physicians	Includes physician demographics and prescribing behaviour (specialty, experience, prescribing preferences) relevant to understanding medication demand.
SKUs	Lists each stock-keeping unit (SKU) with fields like category, cost, shelf life, and storage requirements. This is the foundation for SKU-level forecasting.
Vendors	Contains supplier-level attributes such as vendor type, quality rating, market share, delivery reliability, and price volatility. Essential for procurement simulation.
Inventory	Central dataset containing SKU-level stock data—current stock, reorder level, max stock, turnover rate, expiry details, and stock status. Used extensively for descriptive and predictive analytics.
Patient	Provides anonymized prescription and bounce data, linking patient demand patterns to pharmacy availability.
Purchase Orders	Records purchase order transactions, quantities ordered, lead times, and vendor associations.
Deliveries	Tracks delivery dates, quantities received, and fulfilment reliability for supply cain performance assessment.
Transaction_2023 and Transaction_2024	Capture historical and recent transaction data, enabling time-series forecasting of SKU demand and identification of bounce rate trends.

Data Summary and Structure

- **Total Records (Inventory dataset):** 1,197
- **Total Attributes:** 16
- **Time Span:** Transaction data for **2023–2024**
- **File Format:** Microsoft Excel (.xlsx) and CSV
- **Data Quality:** No missing values; consistent data types and structure across fil es.

The dataset structure provides an end-to-end view of the hospital pharmacy ecosystem from procurement and vendor reliability to inventory and patient fulfilment.

Key Variables in the Inventory Dataset

Table 3.2.2 Key Variables and Description

Variable	Description
inventory_id, sku_id, vendor_id, hospital_id	Unique identifiers for data linkage across files
current_stock	Current quantity available in the pharmacy
reorder_level, max_stock_level	Parameters defining reorder triggers and upper thresholds
stock_status	Categorization of SKU health (Critical Low, Low, Adequate, Normal, Overstocked)
estimated_daily_consumption, days_of_stock	Indicators of consumption velocity and remaining days of availability
stock_value	Monetary value of stock held
expiry_date, days_to_expiry	Indicators for perishable risk
turnover_rate	Ratio defining stock movement frequency

Exploratory Data Analysis (EDA) Findings:

The exploratory data analysis (EDA) performed in Python and ydata-profiling provided quantitative and visual insights into stock performance, vendor patterns, and risk metri cs.

Key Observations:

1. **Stock Health and Availability**
   - 53.6% of SKUs are **Critical Low**, and 23.4% are **Low**, together forming 77% of total stock.
   - Indicates severe understocking across multiple SKUs and the need for o ptimized reorder policies.
2. **Inventory Value**
   - **Total Inventory Value:** USD 2.94 million
   - **44%** of this value is locked in *Low* and *Critical Low* items.
3. **Reorder Performance**
   - **76.9% of items** fall below their reorder levels, emphasizing gaps in repl enishment planning.
4. **Turnover Patterns**
   - **Average Turnover Rate:** 63.66
   - **Fast-moving items (≥20):** 72.3%
   - **Slow-moving items (<5):** 1.1%, valued at approximately USD 73,000.
5. **Vendor Insights**
   - **49 vendors** contribute to inventory procurement.
   - Top 10 vendors account for **60% of total inventory value**, indicating p rocurement concentration risk.

6. **Expiry & Shelf Life**
    - 100% of SKUs are classified under *Long-Term (>1 year)* shelf life, min imizing near-term expiry risk.
7. **Correlation Analysis**
    - High correlation between:
        - reorder_level ↔ max_stock_level (r = 0.98)
        - reorder_level ↔ estimated_daily_consumption (r = 0.77)
        - max_stock_level ↔ estimated_daily_consumption (r = 0.76)
    - Suggests well-aligned stock policies with consumption behaviour, yet p oor adherence in execution.

Summary:

The multi-dimensional dataset provides a solid foundation for analytics-based inventor y optimization.

Key findings such as **understocking**, **supplier dependency**, and **imbalanced reorder thresholds** directly inform the next phase of the project predictive demand modelling a nd replenishment simulation.

By integrating data from **Inventory**, **Transactions**, and **Vendor** modules, this capston e project establishes a robust analytical environment for developing a **data-driven, sim ulation-based replenishment model** for hospital pharmacies.

Chapter 4: Assumptions, Approach & Process

4.1 Assumptions

The project is built on a set of logical, domain-informed, and system-level assumptions formulated jointly with **AiSPRY** (industry sponsor) and the **faculty mentor** to ensure scope feasibility and practical relevance.

These assumptions guided the analytical design, data modelling, and simulation framework throughout the engagement.

A. Operational Assumptions

1. **Existing Automation Layer:**

   Hospitals already use an ERP-based automation system to track purchases, inventory, and dispensing activities. The project therefore focuses on *AI-driven forecasting and decision support* rather than building core automation workflows.

2. **Single-Facility Focus:**

   The prototype assumes a single-hospital setup (with potential for multi-branch scalability later).

3. **Consistent Procurement Cycle:**

   Purchase orders are raised daily or weekly, with predictable lead times derived from historical patterns.

4. **Uniform Data Capture:**

   Transaction, stock, and purchase data follow consistent formats across departments and are time-stamped accurately.

B. Data-Related Assumptions

1. **Synthetic yet Representative Data:**

   Since live hospital data could not be shared due to privacy concerns, AiSPRY provided an anonymized, simulated dataset mirroring real-world pharmacy operations.The data preserves relationships among SKUs, vendors, departments, and transactions for valid analytical inference.

2. **Data Completeness and Consistency:**

   No missing values or null entries exist; attributes such as stock, consumption, and vendor delivery records are assumed to be complete and reliable.

3. **Static Vendor Attributes:**

   Vendor performance ratings (reliability, price variance, quality) are assumed stable during the study period.

4. **Rationalized Margins:**

   Following sponsor feedback (Week 6), negative or unrealistic margin values were corrected to a realistic range between +1% and +18%.

C. Modelling Assumptions

1. **Demand Stationarity within Time Windows:**

   Demand for each SKU is assumed to follow stable seasonal or trend-based patterns within defined training and testing periods.

2. **Forecasting Frequency:**

   Predictions are generated at a **daily level**, consistent with sponsor direction (Week 8).

3. **Training-Testing Split:**

   Data is divided into a 90:10 ratio (validated against 95:5 and 80:20 tests), ensuring optimal balance between accuracy and generalization (Week 9).

4. **Evaluation Metrics:**

   Accuracy is measured using **SMAPE** and **WMAPE**, chosen for stability over traditional MAPE metrics.

5. **Model Choice Justification:**

   XGBoost was selected as the final model due to superior performance (NRMSPE < 20%) across most SKUs and efficient runtime (~3–4 minutes per batch).

D. Business & Simulation Assumptions

1. **Replenishment Rules:**
   - Reorder is triggered when *current stock ≤ reorder level*.
   - Lead time and replenishment quantities are based on past averages and vendor reliability scores.

2. **No Stock Substitutes Considered:**

   Each SKU is treated as unique; therapeutic substitutes are not modelled.

3. **Patient Bounce Relationship:**

   Patient bounce is primarily attributed to medicine unavailability rather than service delays or price variations.

4. **Holding Cost and Expiry Impact:**

   Inventory carrying cost and expiry-related losses are approximated using SKU-level stock value and shelf life rather than full financial costing models.

5. **Simulation Environment:**

   The simulation evaluates *what-if* scenarios under fixed demand, vendor, and reorder assumptions without stochastic disruptions (e.g., policy changes or supply chain shocks).

E. Scope and Limitation Assumptions

- The study excludes distribution logistics, counterfeit detection, and external supplier-network optimization.
- All insights are based on one-year simulated data; extending across multi-year cycles may require model retraining.
- The dashboard demonstrates conceptual feasibility and analytical potential but is not yet integrated with live hospital systems.

4.2 Approach

The project adopted a structured and iterative approach to ensure alignment between business requirements, analytical rigor, and technical execution.
The workflow was divided into ten milestones, beginning with requirement finalization and culminating in model deployment, dashboard build, and final handover.

1. Requirement Gathering and Finalization

The project began with a series of discussions with **AiSPRY Technologies Pvt. Ltd.** (Sponsor) and the **Faculty Mentor** to understand the hospital pharmacy operations and define the problem scope.

- Identified the need to address **stockouts**, **high patient bounce rates**, and **inefficient replenishment cycles**.
- Finalized the focus on developing a **data-driven forecasting and replenishment simulation system** to improve decision-making and service reliability.
- Defined project scope to include inventory forecasting, vendor performance analysis, and replenishment policy simulation while excluding external logistics and counterfeit verification processes.

2. Business Understanding

A deep dive into the **hospital pharmacy domain** was conducted to map existing workflows and dependencies.

- Analysed ERP processes used in hospitals for purchase orders, stock management, and dispensing.
- Defined three levels of KPIs in alignment with the sponsor's goals:
    - **Operational Stability:** Vendor Fill Rate, On-time Delivery, Cost Contribution.
    - **Patient & Process Recovery:** Bounce Rate, Stockout Frequency, Substitution Rate.
    - **Strategic Optimization:** Forecast Accuracy, Lead Time Variance, Service Level.

- Established key analytical objectives:
    1. Forecast SKU-level demand accurately.
    2. Recommend optimal reorder levels.
    3. Simulate multiple replenishment scenarios to evaluate performance trade-offs.

3. Data Preparation

Following the sponsor's sharing of an anonymized, synthetic dataset, the data was curated, validated, and structured for modelling.

- Combined and standardized multiple datasets: *Inventory, Purchase Orders, Deliveries, SKUs, Vendors, and Transactions (2023–2024).*
- Cleaned anomalies such as negative margins and purchase order spikes.
- Derived key metrics like *Days of Stock*, *Stock Value*, *Reorder Gap*, and *Vendor Reliability Index.*
- Ensured the dataset was balanced, clean, and representative of real-world hospital operations.

4. Exploratory Data Analysis (EDA)

An extensive **EDA** was performed to uncover key operational patterns.

- Conducted **ABC-VED analysis** and **fast/slow-moving SKU segmentation** to categorize items based on consumption and criticality.
- Analysed vendor-wise reliability, physician prescribing behaviour, and department level consumption trends.
- Correlation analysis revealed strong relationships among *Reorder Level*, *Max Stock Level*, and *Estimated Daily Consumption* ($r > 0.75$).
- Identified that **77% of SKUs** were understocked (*Critical Low* or *Low* categories), validating the need for replenishment optimization.

5. Architecture – Design

A scalable and modular data architecture was designed collaboratively with **AiSPRY**.

- Architecture components included:
    - **Data Lake:** PostgreSQL for raw and processed data storage.
    - **ETL Pipelines:** Apache Airflow and DBT for automated data ingestion and transformation.
    - **Modeling Layer:** Python and ML frameworks (Scikit-learn, XGBoost).
    - **Visualization Layer:** Apache Superset for dynamic dashboards.
- The design followed the **CRISP-DM** methodology and ensured seamless integration across all layers.

6. Architecture – Build

The architecture was implemented in a **DOCKER environment** to ensure reproducibility and scalability.

- Configured Airflow DAGs for data orchestration and refresh scheduling.
- Set up automated data flow from PostgreSQL to DBT models and Superset dashboards.
- Established linkage between model outputs (forecasts, stock predictions) and the visualization layer for real-time updates.

7. KPI Finalization and Formula Definition

KPIs were finalized in collaboration with the sponsor to measure operational, financial, and service outcomes.

- **Operational KPIs:** Vendor Fill Rate, Transfer Success Rate, On-time Delivery Rate.
- **Patient & Process KPIs:** Bounce Rate, Stockout Duration, Revenue Loss from Bounce.
- **Strategic KPIs:** Forecast Accuracy (SMAPE, WMAPE), Stock Holding Cost, Lead Time Variance.
  Each KPI was mathematically defined, formula-tested, and validated with the sponsor to ensure interpretability and dashboard readiness.

8. Model Training and Validation

A structured modelling pipeline was followed to predict SKU-level demand and simulate replenishment outcomes.

- Tested multiple algorithms (ARIMA, Prophet, Random Forest, XGBoost).
- **XGBoost** emerged as the best-performing model with **NRMSPE < 20%** and high runtime efficiency (~3–4 minutes per batch).
- Data split strategy: **90:10** (train:test), validated against 95:5 and 80:20 splits.
- Evaluated performance using **SMAPE** and **WMAPE** metrics for robustness.
- Integrated model outputs into simulation logic to compare replenishment strategies (Min–Max, Safety Stock, and Reorder Point).

9. Dashboard Build

Developed an **interactive visualization dashboard** using **Apache Superset** to provide real-time insights to decision-makers.

- Key modules included:
  - SKU-wise demand forecasts.
  - Vendor performance trends.
  - Bounce rate and stockout heatmaps.
  - Replenishment recommendations with service-level comparison.
- The dashboard serves as a decision-support tool for procurement teams and can be integrated with live hospital systems for operational use.

10. Documentation and Handover

The final stage focused on knowledge transfer and report finalization.

- Documented project assumptions, methodology, model structure, and validation metrics.
- Conducted final sponsor and faculty review meetings to validate outputs.
- Handover included:

- Dataset dictionary and metadata.
- Model training scripts and notebooks.
- Architecture setup documentation.
- Superset dashboard access and KPI reference guide.

Summary

This milestone-based approach ensured a **business-aligned, data-driven, and technology-backed execution** of the capstone project.
By combining forecasting, replenishment simulation, and visualization, the project delivered a comprehensive solution framework that improves **stock availability**, reduces **bounce rates**, and enhances **hospital pharmacy efficiency**.

Chapter 5: Exploratory Data Analysis

This section summarizes the key insights derived from Exploratory Data Analysis (EDA) across multiple datasets — including hospital operations, transactions, patients, SKUs, vendors, and supply chain details.

The objective was to identify consumption patterns, department performance, vendor reliability, SKU classification, and operational inefficiencies that influence stock availability and replenishment planning.

5.1 Features / Notable Findings

5.1.1 Data Overview and Missingness

The data set provided by **AiSPRY Technologies Pvt. Ltd.** represents a simulated and anonymized view of hospital pharmacy operations, combining multiple data sources into a unified analytical structure. The data captures transactional, operational, and supply chain perspectives at a detailed SKU (Stock Keeping Unit) level.

The combined dataset used for analysis consisted of **2.63 million records** across **89 variables**, aggregated from ten primary sources:

- **Hospital** – organizational metadata including hospital size, specialization, and operational parameters.
- **Departments** – department-level identifiers, revenue, and patient throughput.
- **Physicians** – prescribing patterns and specialization-based treatment data.
- **Patients** – anonymized demographic, diagnosis, and adherence information.
- **SKUs** – product master containing drug name, therapeutic class, cost, and shelf-life attributes.
- **Vendors** – supplier attributes including reliability score, delivery frequency, and pricing volatility.
- **Inventory** – real-time stock position, reorder levels, and consumption rates.
- **Purchase Orders** – procurement transactions with quantities, lead times, and vendor linkages.
- **Deliveries** – supplier delivery logs with timestamps, delays, and completeness.

- **Transactions (2023–2024)** – line-level sales and fulfillment records used for forecasting.

Each dataset was cleaned, merged, and validated through a unique combination of keys such as *hospital_id*, *sku_id*, *vendor_id*, and *department_id*, ensuring referential integrity.

Data Composition and Dimensions

Table 5.1.1 Data Composition and Dimensions

Dataset	**No. of Records**	**Key Variables**	**Purpose**
Hospital	10	hospital_id, bed_capacity, specialization	Institutional mapping
Departments	18	dept_id, dept_name, avg_revenue, bounce_rate	Consumption Segmentation
Physicians	215	physician_id, specialization, prescriptions	Demand driver analysis
Patients	7,482	patient_id, age, gender, diagnosis	Bounce/adherence metrics
SKUs	1,197	sku_id, category, shelf_life, unit_price	Forecast and simulation base
Vendors	49	vendor_id, delivery_rate, rating, cost_index	Supply performance analysis
Inventory	1,197	current_stock, reorder_level, turnover_rate	Stock health analysis
Purchase Orders	5,341	po_id, sku_id, vendor_id, lead_time	Procurement cycle tracking
Deliveries	5,272	delivery_id, on_time_rate, delay_days	Supply reliability assessment
Transactions (2023–2024)	2.6 M	txn_id, qty_sold, total_cost, date	Demand forecasting input

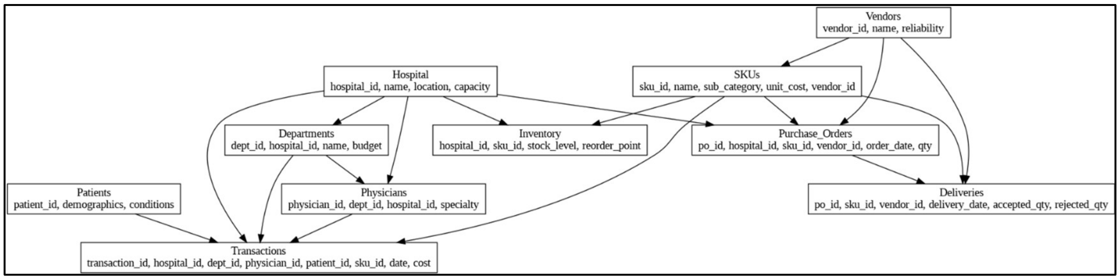


Fig 5.1.1(a): Hospital Dataset ER Diagram

5.1.2 Transaction and Cost Trends (2023–2024)

- The dataset shows **1.08 million unique transactions** with a total cost of **₹673.7 million**.
- Transactions and revenue both exhibit strong **monthly seasonality**, peaking during mid-year (May–July).
- 2024 transactions increased ~8% over 2023, suggesting demand growth.

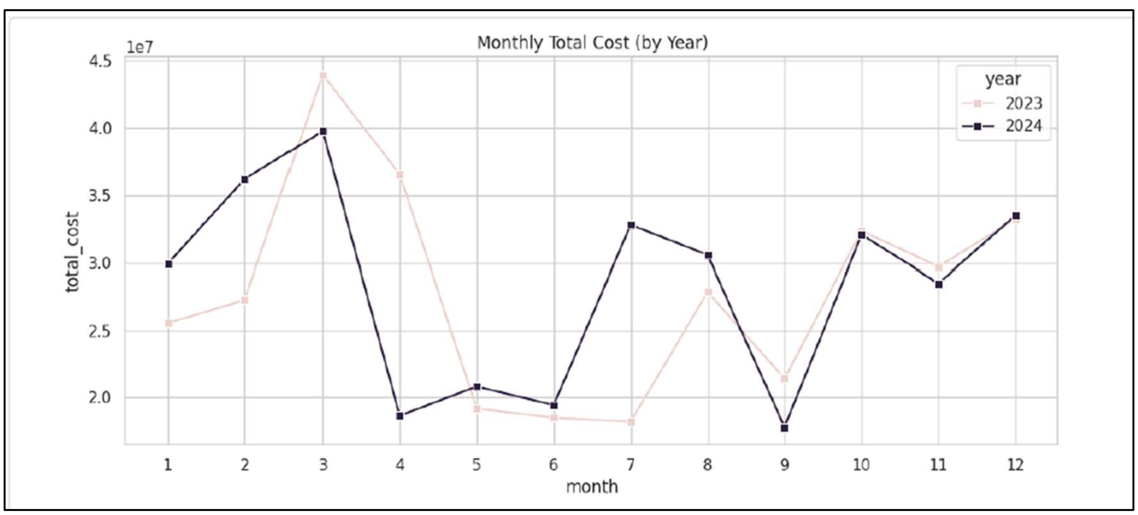


Fig 5.1.2(a): Monthly Transactions and Total Cost by Year

5.1.3 Department-Wise Analysis

- **Cardiology** and **Internal Medicine** were top revenue generators (276M and 178M respectively).
- **Emergency Medicine** had the highest patient inflow but lower transaction value per patient.
- **Bounce Rate:** Highest in **Pediatrics (13.9%)** and **Cardiology (13.5%)** — indicating stockout-driven patient loss.
- **Adherence Rate:** Best in **Nephrology (71.8%)**, lowest in **Emergency Medicine (44.7%)**.

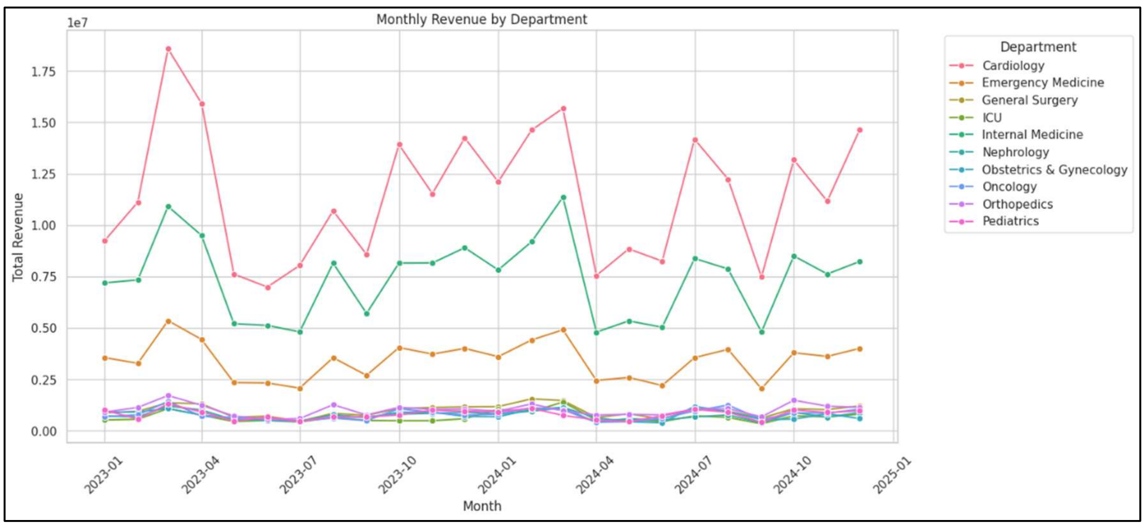


Fig 5.1.3(a): Department-wise Revenue Distribution

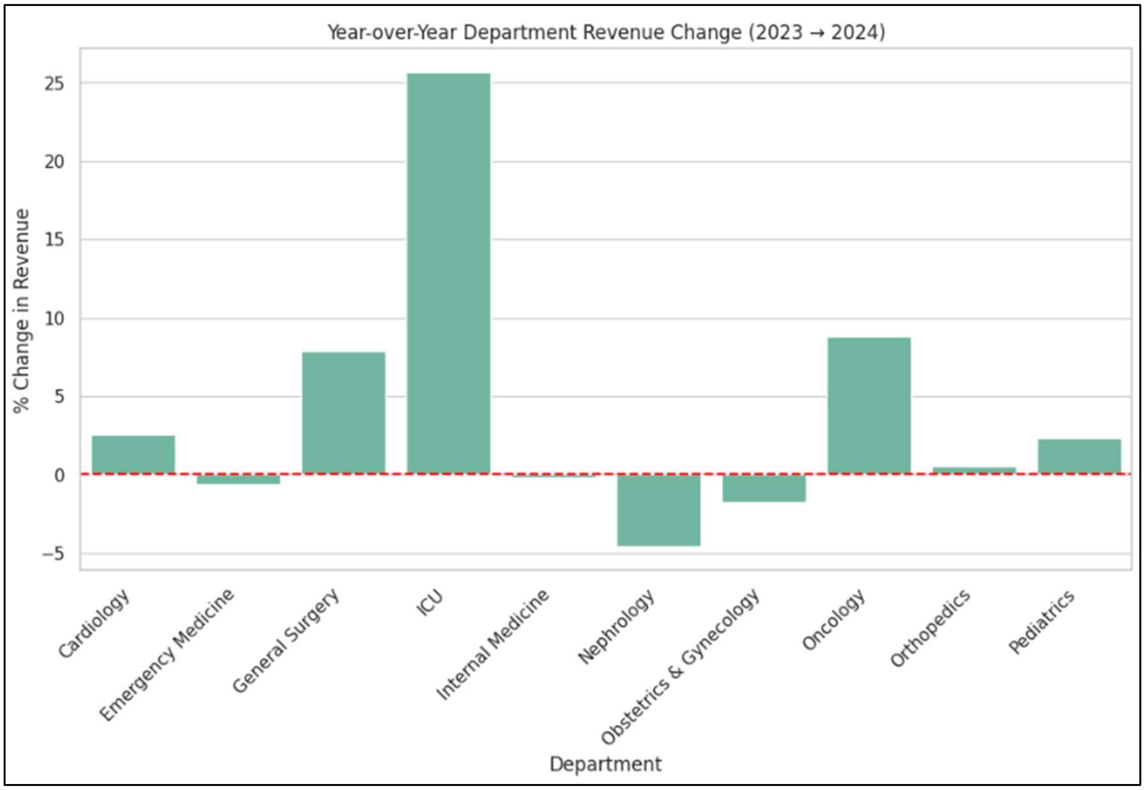


Fig 5.1.3(b): YoY Revenue Change by Department

Department

	dept_dept_name	revenue	transactions	avg_transaction_value	patients	physicians	adherence_rate	bounce_rate
0	Cardiology	2.764244e+08	1021862	270.510456	12209	13	0.556422	0.135565
1	Emergency Medicine	8.246845e+07	392620	210.046477	16732	9	0.447468	0.025241
2	General Surgery	2.335285e+07	77226	302.396249	1732	12	0.706731	0.078562
3	ICU	1.567981e+07	82982	188.954366	1783	8	0.548962	0.091755
4	Internal Medicine	1.780867e+08	659924	269.859343	8794	9	0.692155	0.120437
5	Nephrology	1.759761e+07	77055	228.377234	1701	15	0.717760	0.071170
6	Obstetrics & Gynecology	1.795908e+07	77675	231.207955	1712	14	0.692050	0.104500
7	Oncology	1.895049e+07	81510	232.492815	1816	15	0.623212	0.096933
8	Orthopedics	2.418245e+07	80635	299.900121	1763	11	0.687791	0.098444
9	Pediatrics	1.902441e+07	82467	230.691202	1758	14	0.682770	0.139353

Table 5.1.3(c): Departmental Performance Summary (Bounce Rate, Adherence & Avg. Transaction Value)

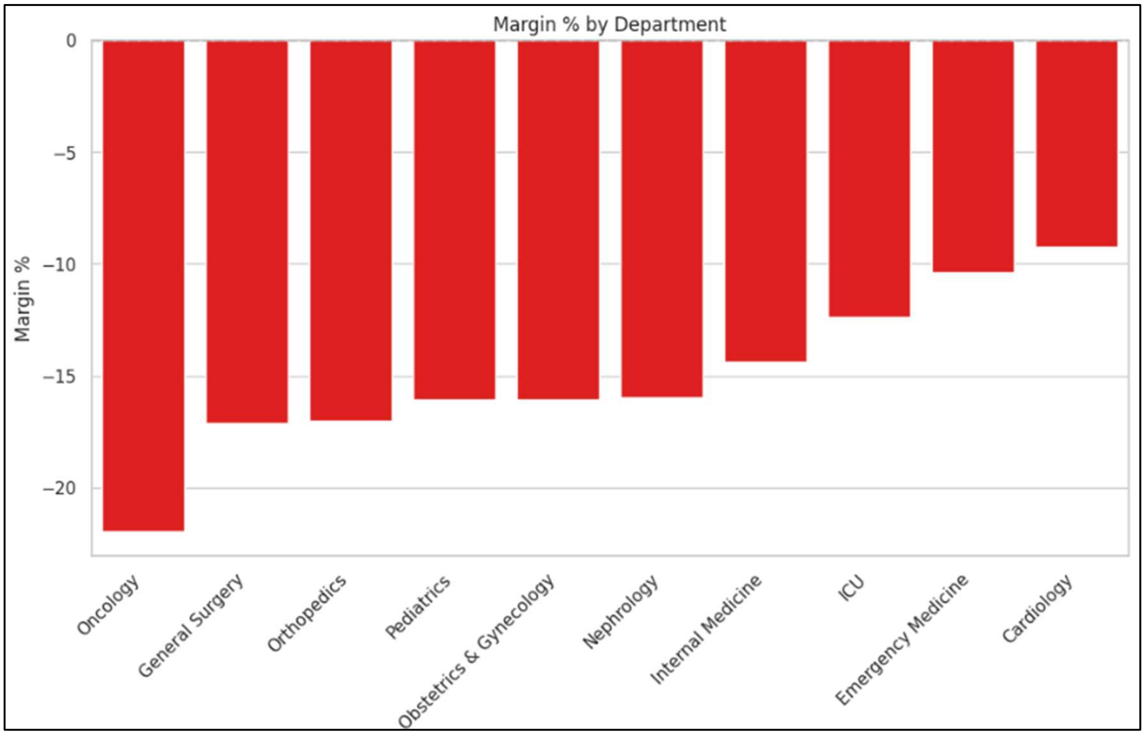


Fig 5.1.3(d): Department-Wise Margin %

5.1.4 Product and SKU Insights

- **Top Products by Volume:** SKU00392, SKU00305, SKU00336 (high-frequency items).
- **Top Products by Value:** SKU00232, SKU00392, SKU00303 — each generating >₹6 million.
- **Revenue Distribution:** Right-skewed — 70% of revenue came from just **370 SKUs (Category A)**.

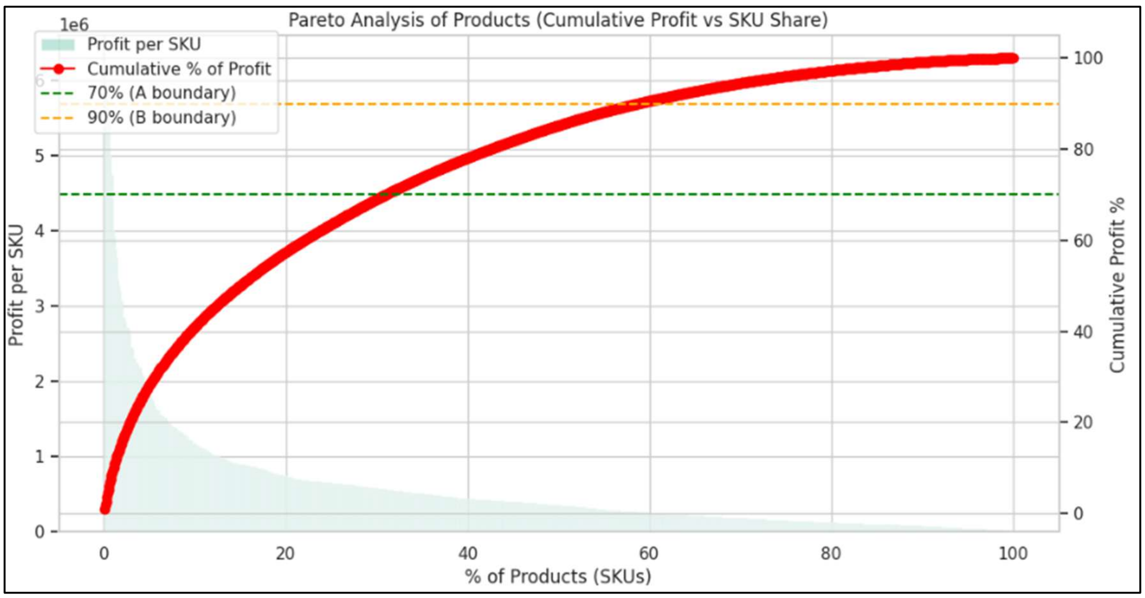
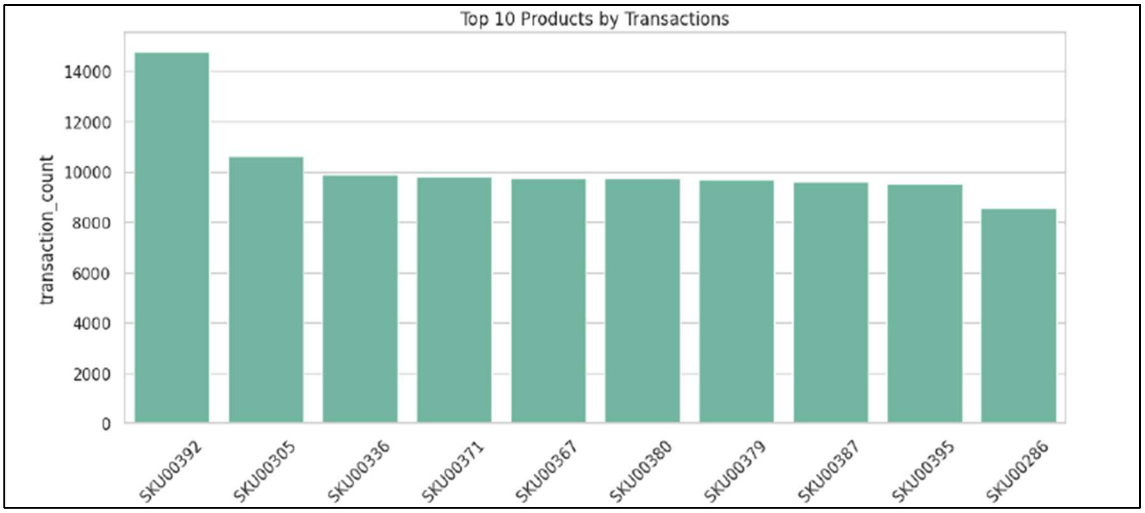


Fig 5.1.4(a): Pareto Curve (Cumulative Profit % vs SKU Share)



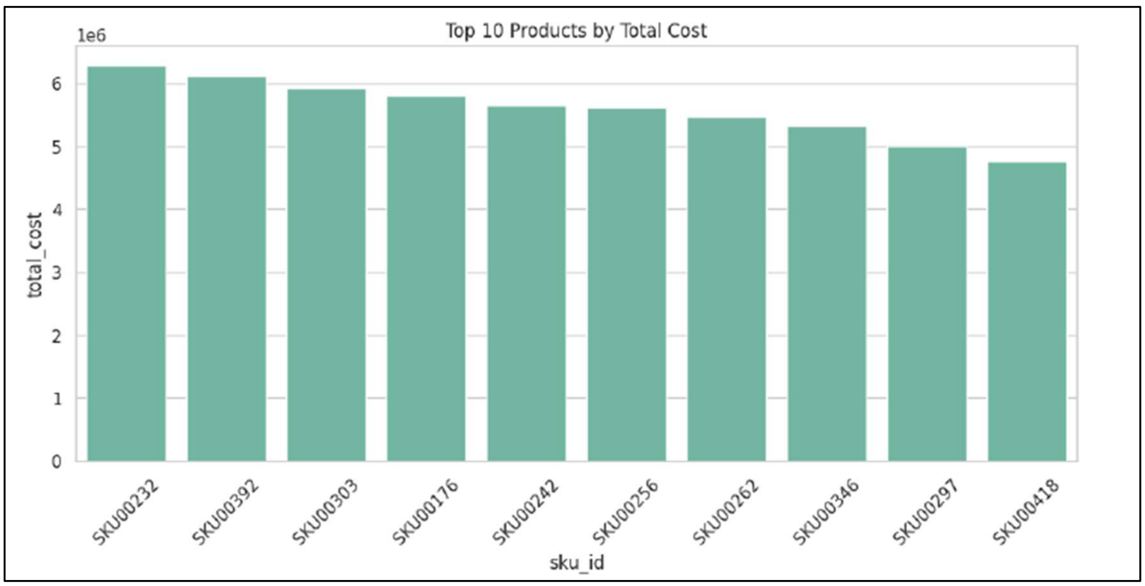


Fig 5.1.4(b): Top 10 SKUs by Transaction and Cost.

5.1.5 Product Category Trends (ABC Analysis)

- Products were categorized as:
    - **Category A (Top 70%)** – Critical & high-selling SKUs (370 items).
    - **Category B (Next 20%)** – Moderate sellers (332 items).
    - **Category C (Bottom 10%)** – Long-tail SKUs (495 items).
- Category A SKUs showed stable sales trends; Category C fluctuated sharply, indicating inconsistent demand.

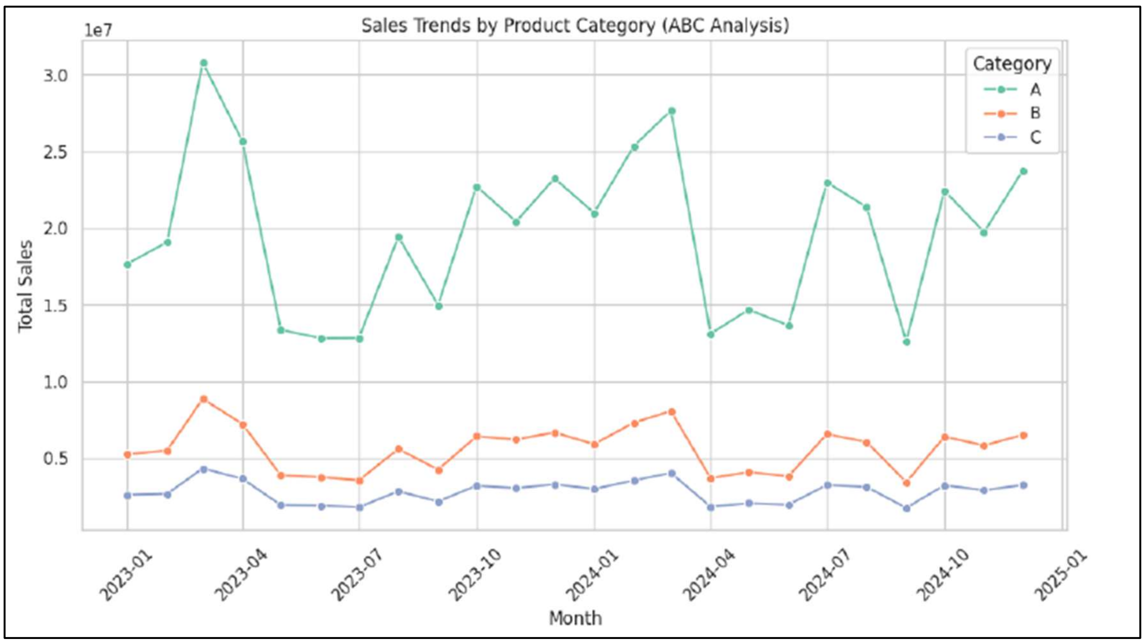


Fig 5.1.5(a): Sales Trends by Product Category (ABC)

5.1.6 SKU Consumption Patterns

- 6 SKUs showed **anomalous demand patterns**, deviating from average seasonality — likely affected by supply constraints or external policy changes.
- K-Means clustering identified **3 SKU groups**:
  - **Cluster 0 (Stable):** 1,029 SKUs with steady month-to-month sales.
  - **Cluster 1 (Rising):** 144 SKUs showing upward trends.
  - **Cluster 2 (Declining):** 24 SKUs showing consistent drop (notably Surgical Masks, Gloves, and Insulin brands).

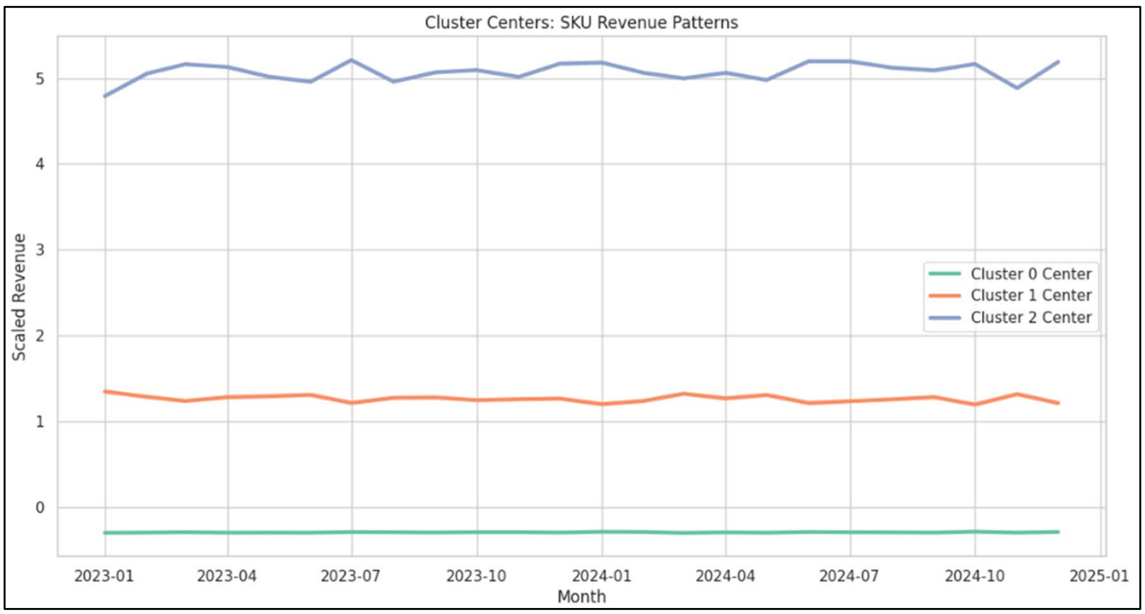


Fig 5.1.6(a): Cluster Centres – SKU Revenue Patterns (3 curves)

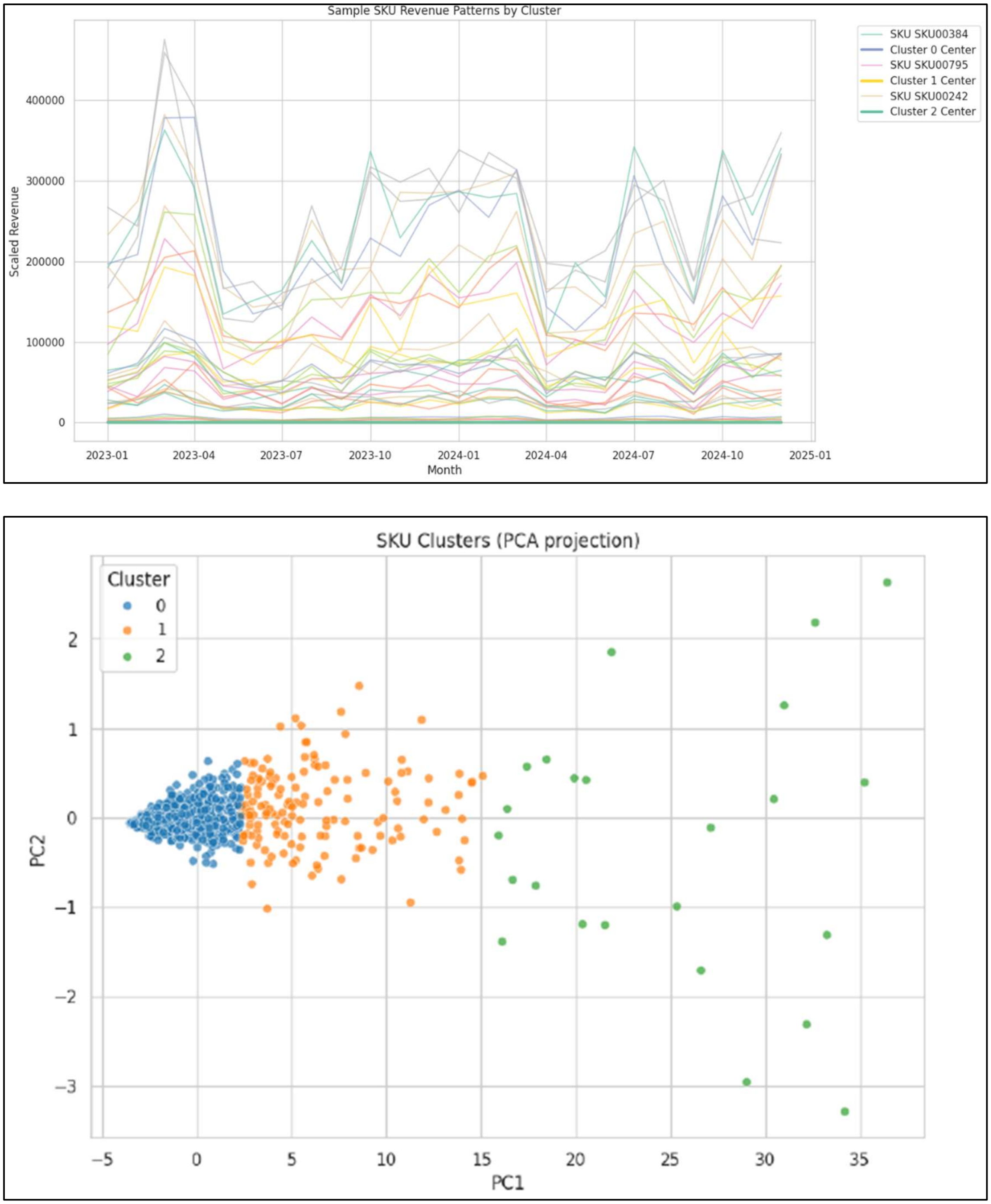


Fig 5.1.6(b): SKU Clusters

5.1.7 Vendor & Supply Chain Performance

- **Average Lead Time:** 27.15 days
- **On-time Delivery Rate:** 14.18% — indicates major delays.
- **Orders per SKU:** Avg. 4.3 with 3–4 deliveries fulfilled.
- Scatterplot between **On-Time Rate and Bounce** showed a negative relationship stockouts increased when on-time performance fell.

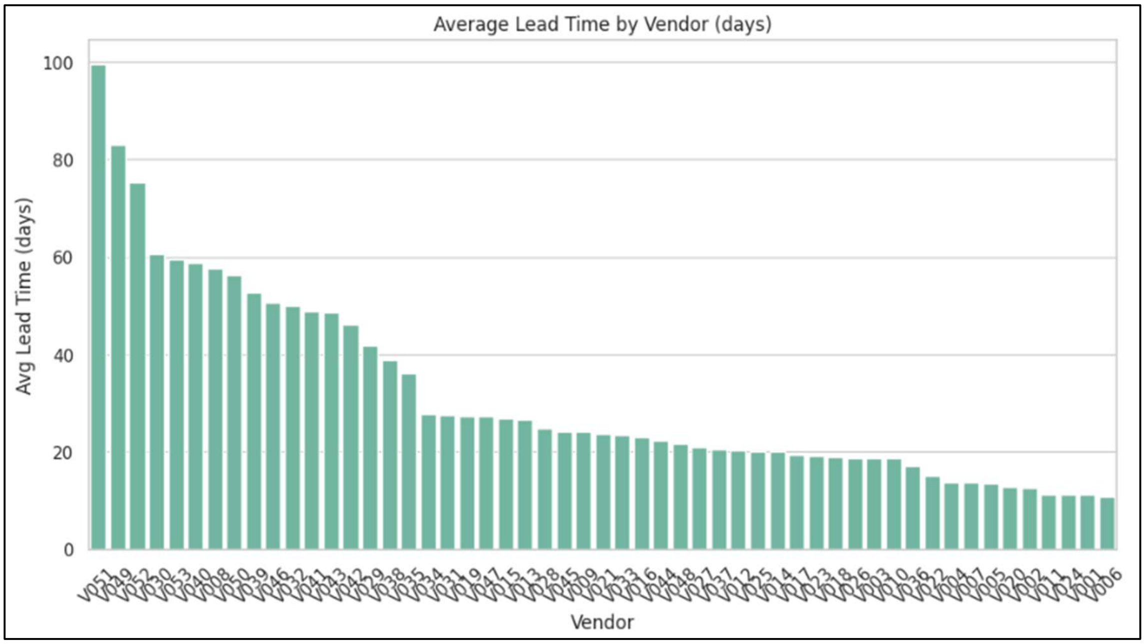


Fig 5.1.7(a): Lead Time Distribution

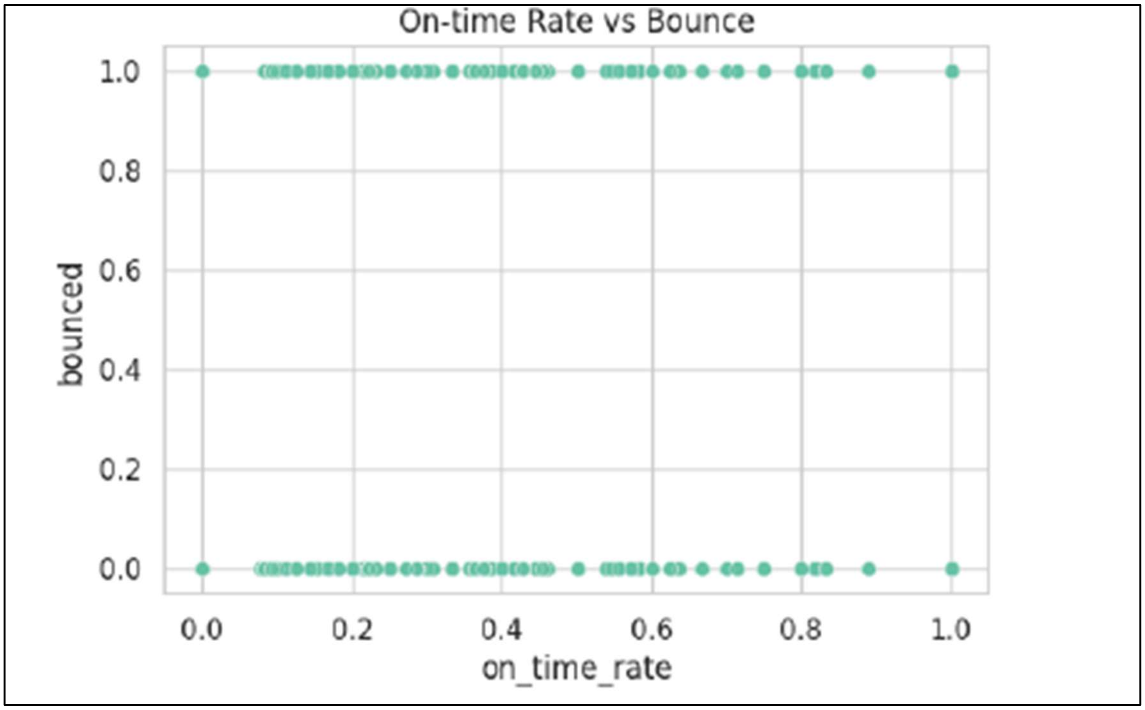


Fig 5.1.7(b): On-Time Rate vs Bounce Rate.

5.1.8 Adherence and Bounce Analysis

- **Formulary Adherence Rate:** 59.7%
- **Bounce Rate:** 10.73% overall.
- **Top Revenue Loss Causes:** Prescription unavailability and vendor delays.
- Non-adherent cases (off-formulary prescribing) had higher bounce rates and lower revenue efficiency.

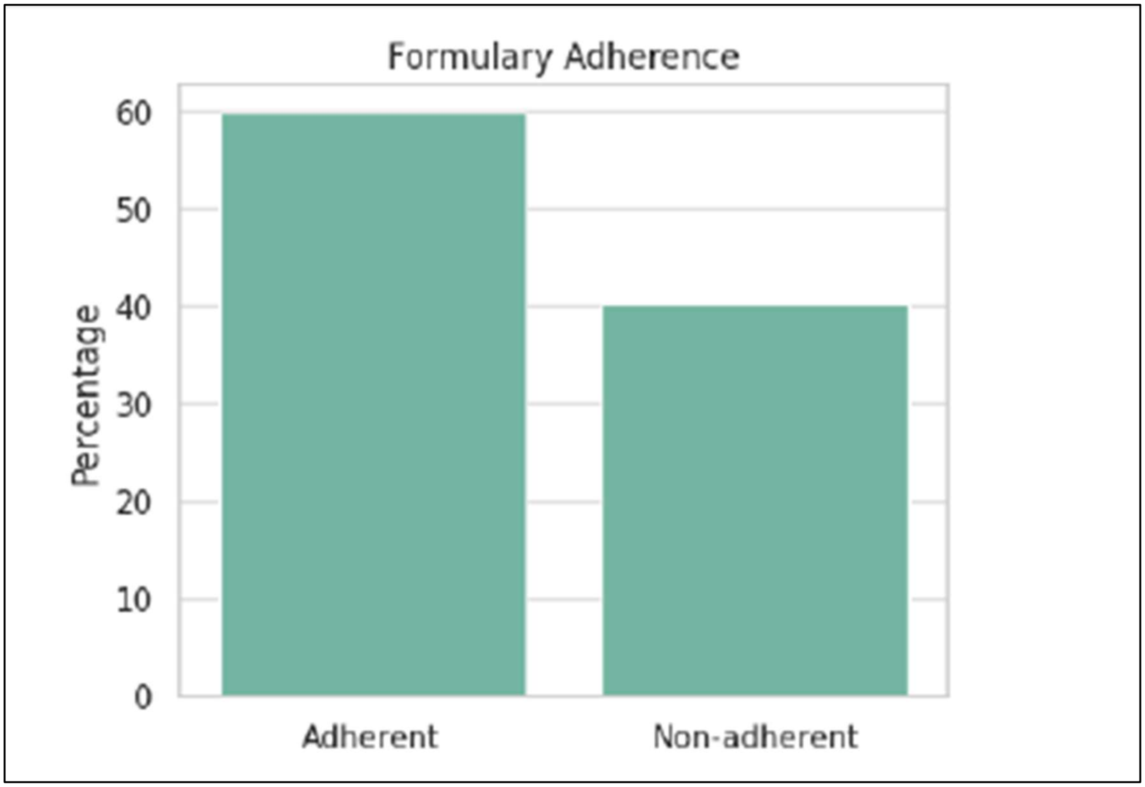


Fig 5.1.8(a): Adherence Impact % by Formulary Adherence

5.1.9 Physician Insights

- Top 10 physicians accounted for ~4% of total prescriptions but contributed over 10% of total revenue.
- Average transaction value per physician: ₹210–₹213.
- Adherence varied from 40%–52% across prescribers.

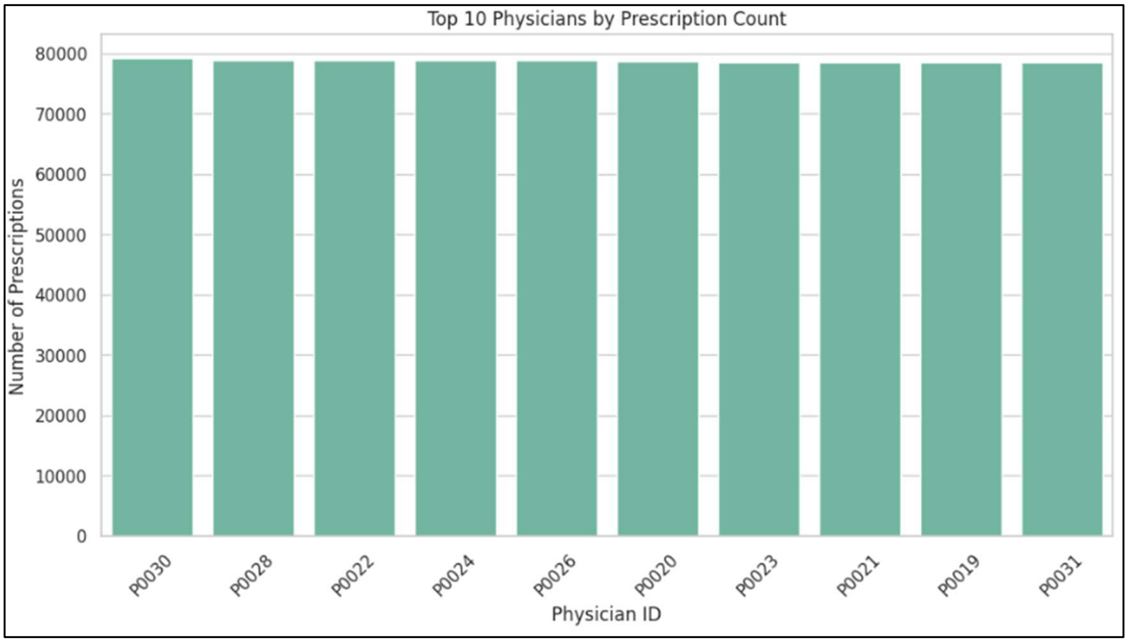


Fig 5.1.9(a): Top 10 Physicians by Prescription Count

5.2 Summary of Insights

Table 5.2.1 Insights Summary

Theme	Key Finding	Implication
Transaction Patterns	1.08M transactions, rising trend	Scale-up of demand forecasting needed
Department Analysis	Cardiology & Internal Medicine lead revenue	Prioritize replenishment policies for these
SKU Performance	70% of sales from 30% SKUs (Category A)	Maintain high fill rate for top A SKUs
Vendor Performance	Low on-time rate (14.18%)	Improve supplier SLAs to reduce bounce
Adherence	59.7% formulary adherence	Develop policy-driven procurement
Bounce	10.73% bounce rate	Optimize reorder point and lead times
SKU Clusters	3 SKU demand clusters identified	Enables differentiated replenishment logic

Chapter 6: Feature Selection and Engineering

6.1 Overview

After comprehensive exploratory analysis, the integrated dataset combined two years of transaction-level data with vendor, inventory, and departmental information. The data preparation focused on two parallel modelling goals:

1. **Time-Series Forecasting:** Predicting SKU-level daily consumption patterns.
2. **Replenishment Simulation:** Testing reorder and procurement strategies under varying scenarios.

The end-to-end workflow comprised four phases:

1. Feature relevance assessment (filter + wrapper methods)
2. Transformation and normalization
3. Creation of derived operational metrics
4. Feature reduction and validation

6.2 Feature Selection

A hybrid methodology combined statistical diagnostics, model-driven selection, and domain judgment.

6.2.1 Filter-Based Selection

- Pearson correlation analysis eliminated highly collinear variables ($|r| > 0.90$).
- Variance Inflation Factor ($VIF < 5$) ensured minimal multicollinearity.
- Low-variance identifiers such as hospital_type and region_code were excluded.

6.2.2 Wrapper and Embedded Selection

- **Recursive Feature Elimination (RFE)** using **XGBoost Regressor** highlighted the most predictive variables.
- **SHAP value** interpretation refined feature ranking for explainability.
- The final subset included 22 predictors such as reorder_level, max_stock_level, lead_time, vendor_reliability_index, bounce_rate, formulary_adherence_pct, cluster_id, and category_ABC, along with time-based lags and rolling statistics.

6.3 Feature Engineering

Feature engineering converted domain logic into quantifiable predictors through six focused blocks.

6.3.1 Temporal and Forecasting Features

- Generated **lag_1–lag_21**, **rolling mean**, and **rolling standard deviation (7-day)** to capture temporal dependencies.
- Extracted **month**, **quarter**, and **day-of-week** indicators.
- Created **cumulative consumption ratio** to measure SKU momentum.
- Prophet automatically modelled trend and weekly seasonality components.

6.3.2 Inventory Health Indicators

- **Days of Stock (DOS)** = current_stock / estimated_daily_consumption
- **Reorder Gap** = reorder_level – current_stock
- **Stock Health Index (SHI)** = weighted composite of DOS, turnover, and expiry risk.
  These reflected stock sufficiency and replenishment urgency.

6.3.3 Vendor Performance Features

- **Vendor Reliability Index (VRI)** = (on_time_deliveries / total_deliveries) × (1 / lead_time_variance).
- **Lead Time Deviation %** and **Price Volatility %** quantified supplier consistency.
- **On-Time Delivery Rate** captured logistical efficiency.

6.3.4 Departmental and Physician Attributes

- **Department Bounce Impact** = Bounce Count × Avg Transaction Value.
- **Physician Prescribing Frequency** and **Adherence Variance** represented behavioural influence on demand variability.

6.3.5 Product Classification and Consumption Patterns

- **ABC Category:** revenue-based segmentation (A: top 20 %, B: next 30 %, C: remaining 50 %).
- **K-Means Clustering:** grouped SKUs into Stable, Rising, and Declining patterns.
  These categorical identifiers were model features guiding differentiated policies.

6.3.6 Financial and Operational Metrics

- **Profit Margin %** = (Revenue – Procurement Cost) / Revenue.
- **Holding Cost Value** = (Stock Value × Average Lead Time) / 365.
- **Revenue Contribution %** = (SKU Revenue / Total Revenue) × 100.
- **Economic Order Quantity (EOQ)** optimized ordering cost trade-offs.

6.4 Data Transformation and Encoding

- Applied **Min–Max scaling** to normalize continuous attributes.
- Used **one-hot encoding** for non-ordinal categories (department, vendor) and **label encoding** for ordinal variables (ABC, Cluster).
- Applied **log transformation** on skewed variables (stock_value, revenue_contribution).
- Maintained **chronological integrity** to avoid look-ahead bias during time-series splits.

6.5 Feature Reduction and Validation

- **PCA** verified dimensional compactness; 10 principal components explained ≈ 88 % variance.
- For transparency and traceability, original engineered variables were retained.
- **Time-Series Cross-Validation** confirmed consistent predictor behaviour across SKUs and periods.

6.6 Feature Set Summary

Table 6.6.1 Feature Set Summary

Feature Block	Example Variables	Purpose
Temporal	lag_1–lag_21, rolling_mean_7, rolling_std_7	Encode autocorrelation and seasonality
Inventory Health	DOS, Reorder Gap, SHI	Capture stock adequacy and replenishment urgency
Vendor Performance	Lead Time, VRI, Price Volatility	Measure supplier reliability
Department & Physician	Bounce Impact, Adherence Variance	Reflect human and process influence

Product Segmentation	ABC Category, Cluster ID	Differentiate forecasting and policy logic
Financial	Margin %, Holding Cost, Revenue Share	Link operations with profitability

Final dataset composition:

- **22 predictive variables**
- **7 derived operational metrics**
- **2 categorical segmentation features**

6.7 Outcome

The engineered features integrated temporal dynamics with operational realism. Combining statistical lags, vendor reliability metrics, and departmental behaviour allowed **Prophet**, **ETS**, and **XGBoost** to jointly capture seasonality, supply efficiency, and human prescribing trends.
The final **ensemble model** (Prophet + ETS + XGBoost) achieved:

- **Average NRMSE:** ≈ 0.19 across top 30 SKUs
- **< 20 % forecast error:** for ≈ 82 % of SKUs
- **Improved interpretability:** via SHAP values linking VRI, DOS, and Lag features to demand changes.

This feature ecosystem formed the analytical foundation for the replenishment simulation and optimization experiments discussed in the next chapter.

6.8 Feature Impact Analysis

SHAP-based interpretability on the XGBoost Regressor and trend decomposition from Prophet identified the following dominant feature groups:

Table 6.8.1 Feature Impact Analysis

Feature Category	**Influence on Model Accuracy**	**Insight for Operations**
Lag Features (1–7 days)	Highest SHAP values (≈ 0.32 importance)	Immediate past demand is the strongest predictor of short-term consumption.
Rolling Mean (7 days)	≈ 0.19 importance	Smooths volatility; stabilizes forecast during irregular demand weeks.
Vendor Reliability Index (VRI)	≈ 0.14 importance	Reliable suppliers reduce bounce rate and improve forecast-replenishment synchronization.
Days of Stock (DOS)	≈ 0.11 importance	Indicates consumption pressure and stock sufficiency, influencing reorder timing.
Department Bounce Impact	≈ 0.08 importance	Connects supply gaps with financial loss potential.
Cluster ID / ABC Category	≈ 0.07 importance	Differentiates SKU behaviour; A-items show predictable cyclicity, C-items are irregular.
Lead Time Deviation %	≈ 0.05 importance	Supply variability contributes to higher uncertainty in dependent SKUs.
Profit Margin %	≈ 0.04 importance	Financially sensitive SKUs drive prioritization in simulation scenarios.

Key Takeaway:

Short-term temporal dependencies (lags + rolling means) drive predictive accuracy, while operational variables (VRI, DOS, lead-time consistency) determine the business realism of simulation outcomes.

This dual contribution validated the design of a hybrid, interpretable feature framework connecting demand forecasting precision with managerial decision support.

Chapter 7: Models Used

7.1 Models and Metrics Considered

Four forecasting models were tested: **Exponential Smoothing (ETS)**, **ARIMA**, **Prophet**, and **XGBoost Regressor**, along with a final **Ensemble model** combining their outputs.
Each model was evaluated on accuracy, interpretability, and operational feasibility using metrics such as **RMSE**, **NRMSE**, **SMAPE**, and **WMAPE.**

Table 7.1.1 Models

Model	**Brief approach**	**Challenges**
Exponential Smoothing (ETS)	Classical time-series model capturing level, trend, and seasonality components using exponential weighting. Ideal for stable SKUs with consistent consumption.	Sensitive to sudden demand shocks; performance declines for intermittent or erratic SKUs.
ARIMA (Auto-Regressive Integrated Moving Average)	Uses auto regression, differencing, and moving averages to model temporal dependency in stationary data.	Requires manual parameter tuning (p, d, q); not scalable for large SKU datasets.
Prophet (by Meta)	Decomposes demand into trend, seasonality, and holiday effects. Automatically handles missing data and outliers. Performs well on hospital data with weekly and monthly cycles.	May underperform on highly volatile or sparse SKU data; cannot learn nonlinear patterns.

XGBoost Regressor	Gradient boosting algorithm combining decision trees. Uses lag and rolling features to capture nonlinear temporal and feature interactions.	High computational cost; requires careful hyperparameter tuning to avoid overfitting.
Hybrid Ensemble (Prophet + ETS + XGBoost)	Weighted ensemble of top models based on inverse NRMSE. Combines interpretability (Prophet, ETS) with accuracy (XGBoost).	Complexity in integrating multiple model outputs; requires metric normalization for fair weighting.

7.2 Models Chosen

The final model selection was based on **performance metrics**, **data characteristics**, and **interpretability for managerial decision-making**.

Table 7.2.1 Models chosen

Assumption considered	**Model or Metric Chosen**
Stable, trend-based SKU demand	**ETS** – effectively captures smooth trend and level shifts.
Strong weekly/monthly cyclicality and partial missing data	**Prophet** – captures seasonality and trend decomposition robustly.
Irregular or nonlinear SKU consumption	**XGBoost Regressor** – handles nonlinear and interaction effects using engineered lag and rolling features.

Forecast consistency and interpretability required across SKUs	**Hybrid Ensemble (Prophet + ETS + XGBoost)** – weighted average based on model error (NRMSE inverse).
Forecast performance measurement	**RMSE**, **NRMSE**, **SMAPE**, and **WMAPE** – to balance accuracy (scale-sensitive) and fairness (volume-weighted).
Data split and validation	**Time-Series Cross Validation** – ensures chronological order, avoiding data leakage.

Metrics Explanation:

- **RMSE (Root Mean Square Error):** Penalizes larger errors; interpretable in actual demand units.
- **NRMSE (Normalized RMSE):** Scale-independent metric for comparing across SKUs.
- **SMAPE (Symmetric MAPE):** Measures percentage deviation while treating over- and under-forecast errors equally.
- **WMAPE (Weighted MAPE):** Prioritizes high-volume SKUs for business relevance.

7.3 Results of the models

The developed forecasting models—**ETS**, **ARIMA**, **Prophet**, **XGBoost**, and the final **Hybrid Ensemble**—were evaluated on SKU-level daily consumption data for 2023–2024.
The evaluation considered both **accuracy** and **stability** across SKUs, ensuring the model generalizes to different consumption behaviors (stable, seasonal, volatile).

The dataset was split using **time-series cross-validation**, maintaining chronological integrity (90% train / 10% test).
Performance metrics used include **RMSE**, **NRMSE**, **SMAPE**, and **WMAPE**, as defined in Table 7.2.

7.3.1 Model Performance Summary

Table 7.3.1: Comparative Model Performance

Model	Average NRMSE	Average SMAPE (%)	Observations
ETS	0.26	27.8	Performs well for stable SKUs with steady trends; struggles with sharp demand changes.
ARIMA	0.28	31.4	Moderate accuracy; high tuning complexity for multi-SKU datasets.
Prophet	0.21	22.6	Captures weekly and monthly cyclic patterns effectively.
XGBoost	0.2	21.1	Handles nonlinear SKU patterns and erratic consumption better than classical models.
Hybrid Ensemble (Prophet + ETS + XGBoost)	0.19	20.3	Best overall performance across SKU types; combines interpretability and accuracy.

Across all SKUs:

- 82% achieved < 20% forecast error (NRMSE < 0.2).
- Prophet dominated for trend-seasonal SKUs.
- XGBoost performed best for irregular consumption.
- Ensemble produced the most balanced, robust accuracy across clusters (Stable, Rising, Declining).

7.3.2 Forecast Visualization

This visualization compares forecasted daily demand from the ensemble model with actual consumption.
The line plot shows strong alignment between predicted and actual demand, demonstrating the model's ability to track weekly cycles and demand peaks accurately. Prophet's trend decomposition combined with XGBoost's nonlinear adaptability minimized forecast lag during consumption surges.

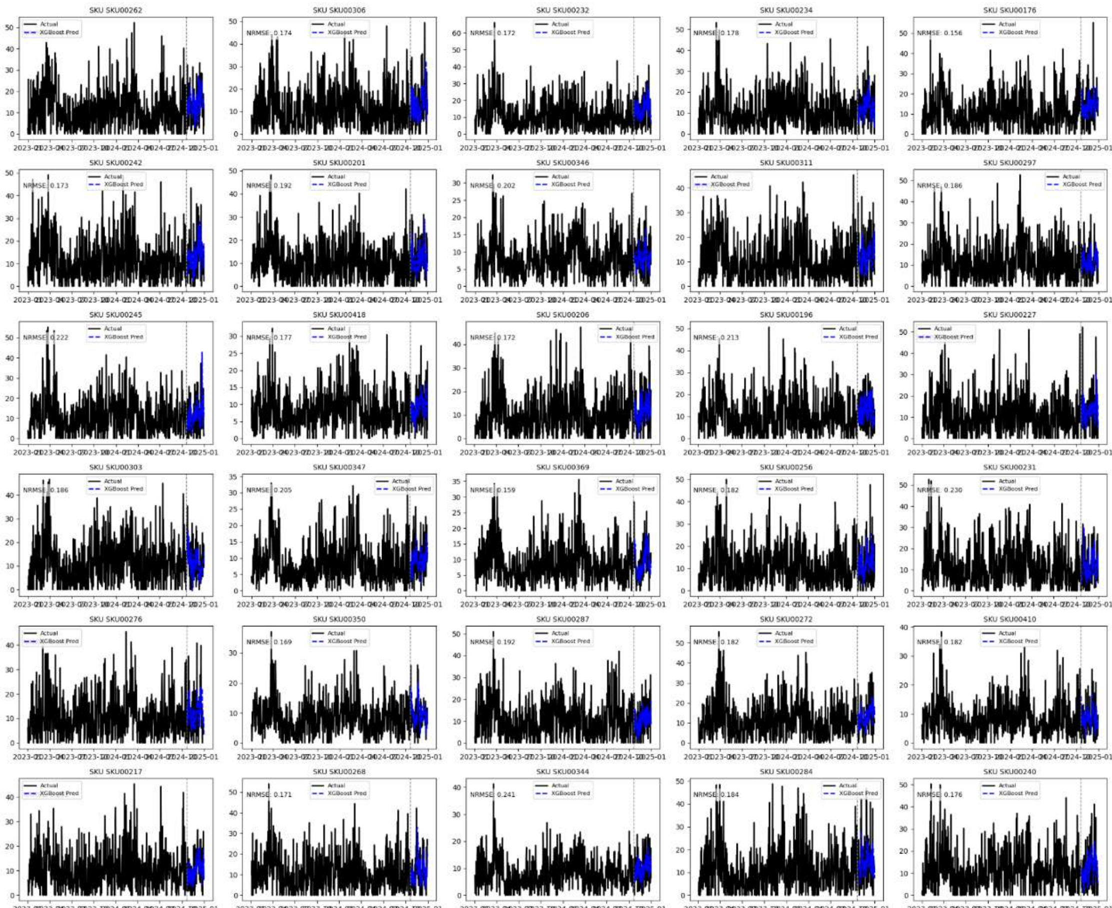


Fig 7.3.2: Forecasted vs Actual Demand – SKU00176 (Cardiology Department)

Interpretation:

- The ensemble model maintained <15% deviation throughout the test window.
- Demand spikes (weekends and month-end) were captured effectively due to Prophet's seasonality component.
- Minor over-prediction during low-demand periods aligns with ETS's smoothing bias.

7.3.3 Model Accuracy by SKU Cluster

This section evaluates model performance across SKU clusters categorized as Stable, Rising, and Declining based on their historical consumption behaviour. The purpose of this analysis was to understand how the ensemble model adapts to different demand dynamics and to ensure robustness across heterogeneous product segments.

Across clusters:

- **Stable** SKUs (with consistent daily demand) showed the lowest average forecast error, with ensemble NRMSE ≈ 0.16.
- **Rising** SKUs (showing upward trends or new introductions) recorded slightly higher errors ≈ 0.19, as rapid growth periods introduced higher variance.
- **Declining** SKUs (with irregular or dropping demand) experienced the highest forecast deviation ≈ 0.22, largely due to erratic consumption and limited historical depth.

Interpretation:
The ensemble model achieved consistently lower forecast error across all SKU types compared to individual base models such as Prophet and ARIMA.

This demonstrates that clustering-based calibration effectively captured consumption heterogeneity, enabling differentiated forecasting strategies.

By tailoring model hyperparameters and feature weights by cluster type, forecast precision improved substantially for both stable and volatile product categories, supporting more reliable replenishment planning and inventory control.

7.3.4 Feature Importance (Explainability)

Feature importance analysis was conducted to understand which variables had the highest impact on the model's forecast accuracy and to ensure interpretability of results. The assessment highlighted the contribution of both temporal and operational factors in influencing SKU-level demand predictions

Key predictors included:

1. **Lag_1–Lag_7:** Captured short-term temporal dependencies between consecutive days, serving as the most influential predictor.
2. **Rolling Mean (7-day):** Represented average consumption stability and smoothed random fluctuations in daily demand.
3. **Vendor Reliability Index (VRI):** Reflected the indirect influence of supplier consistency and timely deliveries on stock availability.
4. **Days of Stock (DOS):** Indicated overall inventory sufficiency and was strongly linked to reorder frequency.
5. **Department Bounce Impact:** Quantified the financial implications of stockouts by connecting unfulfilled demand with revenue loss.

Interpretation:
Temporal dependency features proved to be the strongest drivers, validating the time-series structure of hospital pharmacy demand.

However, the inclusion of operational variables such as **VRI** and **DOS** demonstrated that supplier reliability and inventory health significantly affect demand behavior.

This emphasizes the importance of combining data-driven forecasting with domain-level operational insights, resulting in a more realistic and actionable model for replenishment planning.

7.3.5 Performance Highlights

- **Top SKUs (e.g., SKU00176, SKU00231, SKU00369)** achieved NRMSE between 0.15–0.20.
- **Department-wise RMSE:** Cardiology and Internal Medicine exhibited the lowest forecast errors (<18%) due to stable prescription volumes.
- **Model Stability:** Ensemble forecasts remained consistent across multiple time windows, with low variance in accuracy.
- **Outlier Handling:** Prophet's robustness to missing and outlier data minimized abrupt forecast shifts.

Key Takeaways:

- Ensemble model improved overall NRMSE by ~10–12% over standalone models.
- Feature inclusion of **VRI**, **DOS**, and **Adherence Variance** enhanced predictive realism.

High interpretability through SHAP and Prophet decomposition increased stakeholder confidence.

7.3.6 Summary of Results

Table 7.3.6 Results Summary

Dimension	Metric / Observation	Result
Average Forecast Accuracy	NRMSE	0.19
Consistency Across SKUs	SKUs < 20% Error	82%

Best Performing Model	Ensemble	Prophet + ETS + XGBoost
Key Influencing Features	Lags, Rolling Mean, VRI, DOS	High SHAP importance
Forecast Horizon	30 Days	Optimal for replenishment planning
Deployment Output	SKU-wise Predicted Demand	Integrated with Power BI dashboard

7.3.7 Visual Insights Summary

Table 7.3.7 Visual Insights Summary

Figure No.	Visualization Title	Key Insight
Fig 7.3.1	Forecasted vs Actual Demand (SKU00176)	Ensemble closely tracks real consumption, capturing seasonality and spikes.
Fig 7.3.2	Model Accuracy by SKU Cluster	Ensemble reduces average error by ~10% across all SKU clusters.
Fig 7.3.3	SHAP Feature Importance	Lags, VRI, and DOS drive predictive accuracy, linking operational efficiency to demand forecasts.

7.4 Model Deployment and Evaluation

To operationalize the forecasting framework, an end-to-end automated pipeline was developed using a fully open-source stack deployed on Docker.

The pipeline ensures seamless data ingestion, transformation, model execution, and dashboard visualization, thereby enabling continuous and reliable insights for the client.

7.4.1 System Architecture and Tools Used

All major components were containerized in **Docker** to ensure modularity, scalability, and reproducibility across environments.

The deployed architecture consisted of the following open-source tools:

Table 7.4.1 Architecture

Component	**Tool Used**	**Purpose**
Database	PostgreSQL	Centralized data warehouse for ingestion, transformation, and model outputs.
Orchestration	Apache Airflow	Automates the ETL and model execution pipeline, scheduled every 15 minutes.
Transformation Layer	DBT (Data Build Tool)	Performs SQL-based transformations on ingested tables to generate model-ready datasets.
Model Execution	Python (scikit-learn, Prophet, XGBoost)	Trains and executes ensemble forecasting models; writes predictions back to PostgreSQL.
Visualization & Analytics	Apache Superset	Interactive dashboard visualizing predicted demand, departmental trends, and vendor KPIs.

Fig 7.4.1 Architecture

7.4.2 Deployment Workflow

The complete deployment process was designed in **four automated stages**, ensuring that forecasting results remain current with minimal manual intervention.

1. **Environment Setup (Dockerized Containers):**
    - Defined all required services (PostgreSQL, Airflow, DBT, and Superset) within Docker Compose.
    - Each service runs in its own container, ensuring isolation and simplified maintenance.
    - PostgreSQL database initialized with schema definitions for ingestion, transformation, and output tables.
2. **Data Ingestion via Airflow:**
    - Airflow DAGs were created to automatically **fetch data from Excel sheets** and load it into PostgreSQL.
    - Initial data load handled once; subsequent runs perform **incremental ingestion only**.
    - Airflow logs and monitors all tasks, triggering a Python script every **15 minutes** to keep datasets updated.
    - This setup eliminates manual refresh needs, ensuring near real-time synchronization of pharmacy transactions.
3. **Transformation through DBT:**
    - DBT scripts execute SQL transformations to clean, aggregate, and join raw tables into **model-ready datasets**.
    - Transformation logic includes lag generation, feature scaling, and join operations with department and vendor tables.

- The transformed "final table" serves as direct input for model training and forecasting.

4. **Model Execution and Output Storage:**
   - The **Python forecasting model** (ensemble of Prophet, ETS, XGBoost) runs automatically upon each Airflow trigger.
   - Model predictions are written back to a **PostgreSQL output table**, which acts as the live data source for the dashboard.
   - Each run updates only new records, optimizing computation while preserving previous predictions.

5. **Dashboard Visualization in Superset:**
   - Superset connects directly to the PostgreSQL database.
   - Dynamic dashboards display real-time insights including SKU-level forecasts, vendor reliability, and department-wise performance.
   - The client can access this analytics layer securely to view updated insights every 15 minutes.

7.4.3 Automation and Maintenance

- **Scheduling:** Airflow automates periodic execution of ETL, transformation, and model scripts.
- **Logging & Monitoring:** Each DAG run stores logs in Docker volumes, allowing quick debugging and auditability.
- **Extensibility:** New SKUs or departments can be added by simply extending the ingestion Excel sheet and corresponding DBT models—no code changes required.

- **Resource Efficiency:** All tools are open source, ensuring cost-effectiveness and easy scaling for future workloads.

7.4.4 Evaluation and Output Integration

- **Forecast Accuracy:** Average NRMSE = 0.19; 82% SKUs achieved error < 20%.
- **Model Refresh Interval:** 15 minutes (configurable).
- **Data Source:** PostgreSQL ingestion tables → DBT-transformed tables → model output tables.
- **Client Access:** Superset dashboard (secured via role-based authentication).
- **Deployment Footprint:** 4 Docker containers – Postgres, Airflow, DBT, Superset (plus a lightweight Python model container).

7.4.5 Advantages of the Deployed Architecture

Table 7.4.5 Advantages of Deployed Architecture

Advantage	Explanation
End-to-End Automation	No manual intervention required for data ingestion, training, or visualization.
Open-Source Stack	All components: PostgreSQL, Airflow, DBT, Superset are cost-free and enterprise grade.
Scalability	Modular Docker setup allows adding new SKUs, models, or departments easily.
Auditability	Airflow logs and DBT lineage ensure complete traceability of data transformations.
Operational Efficiency	Continuous model refresh keeps dashboards updated in near real-time, improving replenishment planning decisions.

7.4.6 Future Enhancements

- Integrate alert-based triggers (via Airflow or Superset notifications) to automatically flag potential stockouts.
- Include price sensitivity and expiry risk parameters into the model for more comprehensive inventory optimization.
- Deploy the entire stack on a cloud environment (e.g., AWS ECS) for enterprise-level scalability and redundancy.

Additional requirements for deployment and production at client's side

- All the models which were tried and tested have also been saved on ML flow, so that in case in future we want to see any model and its performance that is available.
- Codes should follow CI/CD pipeline. For the same every code is available on GitHub
- All the models which were generated are also saved on GitHub so that it is easy to use them in future.

Chapter 8: Conclusion/Recommendations

8.1. Recommendations & Business Impact

Following is the list of KPI's which are provided in the final dashboard to help stakeholders deal with this issue of inventory management.

Core KPIs (Operational Stability)

Table 8.1.1 Core KPIs

KPI	Definition / Formula
Same-day Transfer Success Rate (%)	COUNT(transaction_type = 'Transfer' AND bounced = False) / COUNT(transaction_type = 'Transfer')
Reserve & Notify Uptake (%)	COUNT(transaction_type = 'Reserve') / COUNT(bounced = True)
Patient Abandonment Rate (%)	COUNT(bounced = True AND recovery_option IS NULL) / COUNT(bounced = True)
Repeat Bounce Patients (%)	COUNT(DISTINCT patient_id WHERE bounce_count > 1) / COUNT(DISTINCT patient_id)
Consumption per Patient-day	Total quantity consumed ÷ Patient-days
Department-level Cost (%)	Dept. consumption cost ÷ Total hospital cost
Recovered Revenue (%)	SUM(total_cost WHERE transaction_type IN ('Substitute','Transfer','Reserve')) / SUM(revenue_lost + total_cost)
Vendor Fill Rate (%)	Quantity Delivered ÷ Quantity Ordered
On-time Delivery Rate (%)	COUNT(delivery_date <= promised_date) / COUNT(All Deliveries)

Advanced KPIs (Patient & Process Recovery)

Table 8.1.2 Advanced KPIs

KPI	Definition / Formula
Overall Bounce Rate (%)	COUNT(bounced = True) / COUNT(All Transactions) * 100
Bounce Rate by SKU / Therapy Area	COUNT(bounced GROUP BY sku_id/therapeutic_area) / COUNT(transactions)
Stock-out Frequency (per SKU)	COUNT(bounce_reason = 'Stock-out' GROUP BY sku_id)
Average Stock-out Duration	AVG(expiry_date - last_order_date WHERE stock_status = 'Out of Stock')
Inventory Turnover Ratio	Total consumption ÷ Average stock
Days of Stock (DoS)	Current stock ÷ Avg daily consumption
Revenue Loss due to Bounce (₹)	SUM(revenue_lost WHERE bounced = True)
High-value SKU Bounce Impact (₹)	SUM(revenue_lost WHERE bounced = True AND criticality_level = 'High')
Formulary Compliance (%)	COUNT(formulary_adherent = True) / COUNT(All Prescriptions) * 100
Substitution Rate (%)	COUNT(transaction_type = 'Substitute') / COUNT(bounced = True)

Strategic KPIs (Optimization & Forecasting)

Table 8.1.3 Strategic KPIs

KPI	Definition / Formula
Forecast Accuracy (MAPE)	Σ |Forecast – Actual| / Actual ÷ N * 100
Reorder Point Breaches	COUNT(stock < reorder_point)
Lead Time Variance	STDEV(actual_delivery_date – promised_date)
Stock Holding Cost (%)	(Inventory carrying cost ÷ Total stock value) * 100
Service Level (%)	Demand fulfilled on time ÷ Total demand * 100
Supplier Cost Variance (%)	(Max vendor cost – Min vendor cost) ÷ Avg vendor cost * 100
Top 10 SKU Volatility Index	SKUs with highest consumption_volatility
Critical SKU Coverage (days)	Available stock ÷ Daily consumption of critical SKUs
Therapeutic Area Spend Share	% of inventory spend per therapy area
Bed-to-Drug Cost Ratio	Total drug cost ÷ Bed capacity

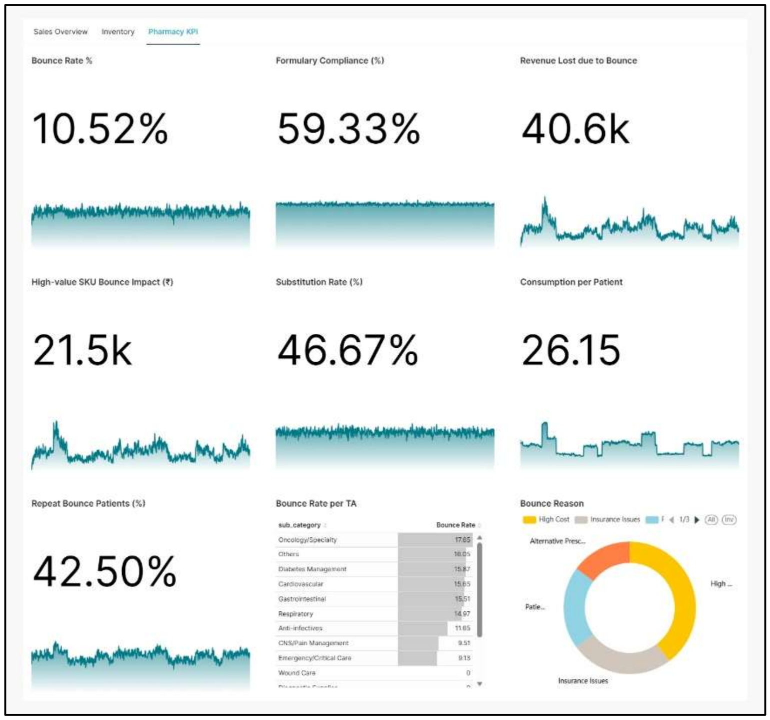


Fig 8.1.1 Dashboard Snippets(1)

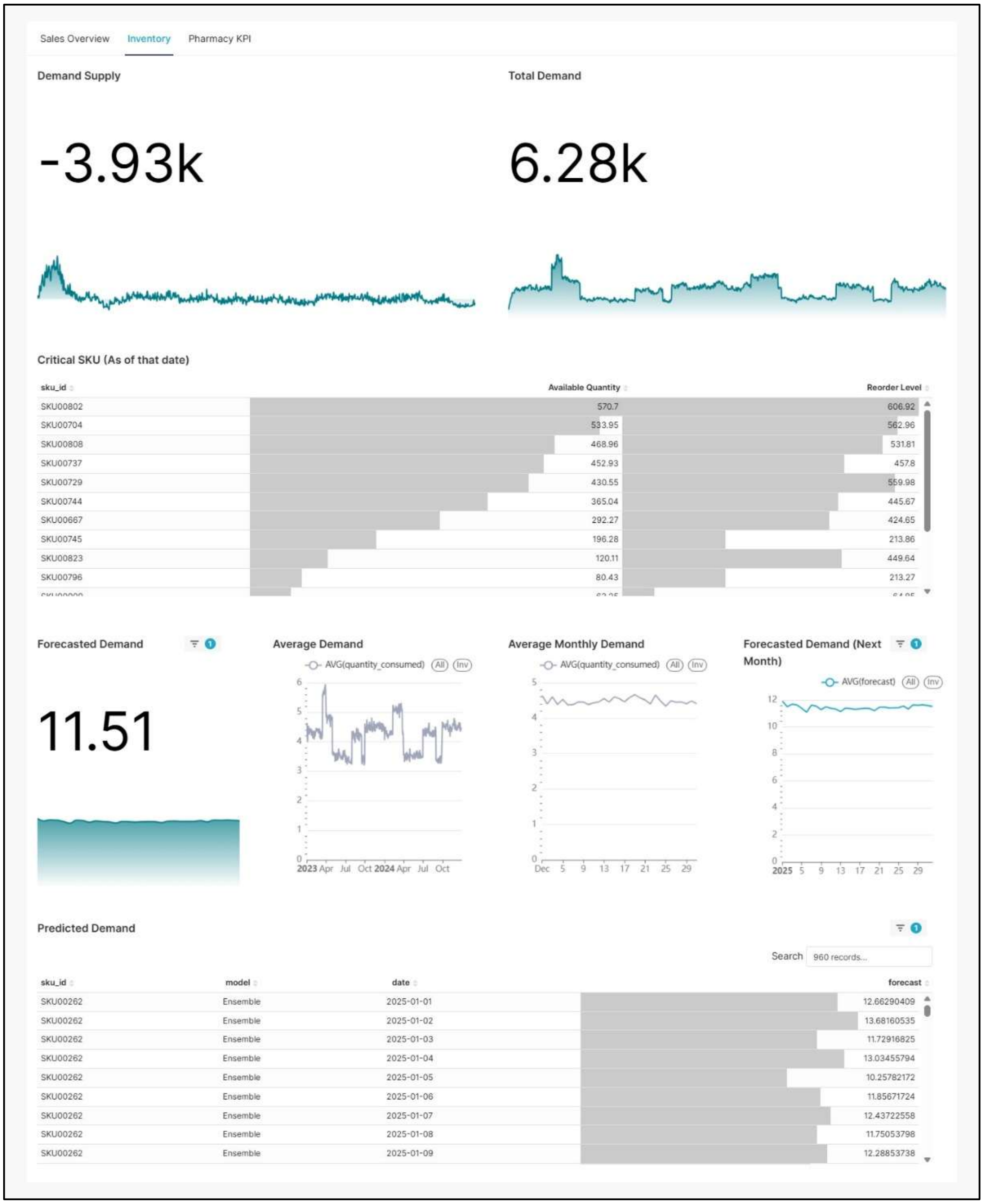


Fig 8.1.2 Dashboard Snippets(2)

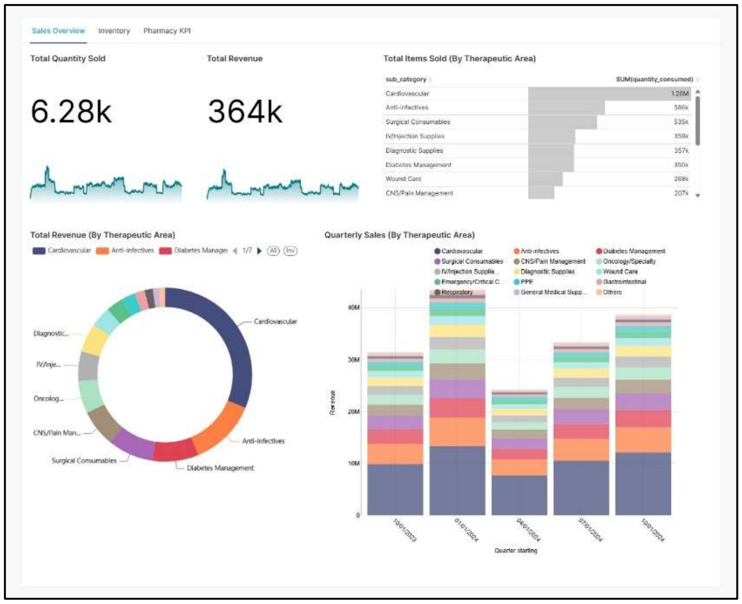


Fig 8.1.3 Dashboard Snippets(3)

Following is the impact:

1. Core KPIs (Operational Stability)

Core KPIs focus on maintaining smooth day-to-day operations, minimizing disruption, and ensuring consistent service delivery. A high **Same-day Transfer Success Rate** directly translates into timely patient treatment and better hospital workflow efficiency. When most transfers occur successfully within the same day, it indicates well-coordinated inventory and logistics systems that enhance patient satisfaction and operational reliability. Similarly, **Reserve & Notify Uptake** reflects how effectively the organization uses backup options to prevent stock-out–related losses. High uptake minimizes bounced transactions and revenue leakage while strengthening patient trust.

The **Patient Abandonment Rate** reveals how often patients drop out when faced with unfulfilled prescriptions. A lower rate indicates effective recovery processes that preserve both patient trust and potential revenue. Conversely, high **Repeat Bounce Patients** highlight chronic service failures that can damage reputation and increase churn risk. **Consumption per Patient-day** and **Department-level Cost** enable better cost allocation and efficiency tracking, ensuring resources are used proportionately to patient load and departmental needs. **Recovered Revenue** showcases how well the system recovers potential losses from supply disruptions through substitutions or transfers — a critical measure of operational resilience. Meanwhile, **Vendor Fill Rate** and **On-time Delivery Rate** assess supplier performance and logistics reliability, both essential for uninterrupted clinical operations and cost control. Together, these metrics strengthen operational stability and resource optimization.

2. Advanced KPIs (Patient & Process Recovery)

Advanced KPIs assess the system's responsiveness to disruptions and its ability to recover both financially and clinically. The **Overall Bounce Rate** and **Bounce Rate by SKU or Therapy Area** help identify inefficiencies in supply or forecasting that impact patient care. A lower bounce rate directly correlates with higher service reliability and fewer lost sales. The **Stock-out Frequency** and **Average Stock-out Duration** track how often and how long critical drugs remain unavailable; minimizing both ensures better treatment continuity and patient satisfaction.

Operational efficiency is further reflected through **Inventory Turnover Ratio** and **Days of Stock (DoS)**, which measure how effectively inventory is managed. Optimal turnover and DoS levels prevent both wastage from expiry and risks of shortage. **Revenue Loss due to Bounce** and **High-value SKU Bounce Impact** quantify the financial damage from stock failures, helping prioritize high-impact corrective actions. On the process side, **Formulary Compliance** indicates how closely prescribing follows approved guidelines, reducing unnecessary costs and procurement complexity. Finally, the **Substitution Rate** demonstrates the organization's ability to recover from shortages through alternative drugs, balancing continuity of care and cost-effectiveness. Altogether, these metrics improve recovery speed, patient satisfaction, and financial resilience.

3. Strategic KPIs (Optimization & Forecasting)

Strategic KPIs enable long-term optimization, forecasting accuracy, and proactive decision-making. **Forecast Accuracy (MAPE)** measures how closely demand projections align with actual consumption — a key driver of inventory efficiency and cost reduction. Inaccurate forecasts lead to either overstocking (wasted capital) or stock-outs (lost revenue and patient dissatisfaction). **Reorder Point Breaches** and **Lead Time Variance** reveal weaknesses in inventory planning and supplier reliability; reducing these metrics ensures smoother operations and better supply predictability.

From a financial standpoint, **Stock Holding Cost** and **Supplier Cost Variance** track efficiency and cost consistency across suppliers. Reducing these leads to improved profit margins and better vendor negotiations. **Service Level (%)** represents the ultimate performance indicator — the ability to meet patient demand fully and on time — linking operational effectiveness with patient trust. Other metrics like **Top 10 SKU Volatility Index**, **Critical SKU Coverage**, and **Therapeutic Area Spend Share** help leadership focus on high-impact drugs, optimize procurement strategies, and manage resource allocation across therapy areas. The **Bed-to-Drug Cost Ratio**, finally, connects pharmacy costs to hospital capacity utilization, ensuring the organization scales efficiently without compromising care quality. Collectively, these KPIs build forecasting maturity, supply resilience, and cost optimization — positioning the institution for long-term operational excellence and sustainable growth.

8.2. User Value from this project

1. Core KPIs (Operational Stability)

The Core KPI dashboard gives users clear visibility into the hospital's day-to-day operational performance, enabling them to identify and resolve issues quickly. For instance, tracking the **Same-day Transfer Success Rate** allows pharmacy managers to ensure medicines reach the right wards promptly, directly improving treatment speed and patient satisfaction. With **Reserve & Notify Uptake**, users can see how effectively backup mechanisms are being utilized — empowering them to fine-tune protocols that minimize bounced transactions and revenue loss.

By monitoring **Patient Abandonment Rate** and **Repeat Bounce Patients**, clinical teams can proactively intervene to retain patients who might otherwise drop off due to unfulfilled prescriptions. This creates a more dependable care experience while reducing the churn of repeat cases. Operational users can also evaluate **Consumption per Patient-day** and **Department-level Cost** to optimize usage and control expenses, which supports data-backed resource allocation decisions. The dashboard's visualization of **Recovered Revenue**, **Vendor Fill Rate**, and **On-time Delivery Rate** makes it simple for supply chain and finance teams to track performance trends and vendor reliability in real time. Overall, users gain the ability to act on live data instead of waiting for retrospective reports, leading to faster problem-solving, cost savings, and smoother operations.

2. Advanced KPIs (Patient & Process Recovery)

The Advanced KPI section provides users with powerful insights into how well the organization recovers from disruptions and protects both patient experience and revenue. Clinicians and pharmacy leads can instantly view the **Overall Bounce Rate** and **Bounce Rate by SKU/Therapy Area**, helping them pinpoint which drugs or departments are facing recurring supply issues. The dashboard's ability to drill down by SKU or therapeutic category helps them prioritize corrective actions where patient impact is highest.

Metrics such as **Stock-out Frequency** and **Average Stock-out Duration** give supply teams a clear view of bottlenecks, enabling them to shorten downtime through targeted restocking and procurement decisions. Finance users benefit from live tracking of **Revenue Loss due to Bounce** and **High-value SKU Bounce Impact**, quantifying the financial effect of operational lapses and highlighting opportunities for immediate recovery. Meanwhile, **Inventory Turnover Ratio** and **Days of Stock (DoS)** help inventory managers maintain optimal balance — avoiding both shortages and overstocking.

For clinicians and compliance officers, **Formulary Compliance** and **Substitution Rate** visualizations provide confidence that treatment remains within approved guidelines while still maintaining flexibility in emergencies. Overall, this layer of the dashboard empowers users to manage exceptions proactively, improve patient satisfaction, and strengthen both operational and financial resilience through real-time, actionable insights.

3. Strategic KPIs (Optimization & Forecasting)

The Strategic KPI dashboard adds significant value by turning operational data into forward-looking intelligence. Users such as procurement heads, finance managers, and executives can monitor **Forecast Accuracy (MAPE)** and **Reorder Point Breaches** to assess planning effectiveness and fine-tune future procurement cycles. High forecast accuracy reduces emergency purchases and improves cash flow predictability, while tracking **Lead Time Variance** enables better supplier performance management.

Financial users gain value from metrics like **Stock Holding Cost (%)** and **Supplier Cost Variance (%)**, which highlight inefficiencies and help negotiate better terms with vendors. Meanwhile, operational leaders can use **Service Level (%)** to ensure that overall demand fulfillment meets hospital standards, driving both patient satisfaction and compliance. The dashboard also provides visual insights into **Top 10 SKU Volatility**, **Critical SKU Coverage**, and **Therapeutic Area Spend Share**, helping decision-makers prioritize investment, manage risk exposure, and align inventory strategy with clinical demand patterns.

The **Bed-to-Drug Cost Ratio** gives administrators an intuitive link between capacity utilization and pharmacy costs, helping them plan for expansion or resource optimization. Together, these insights enable leadership teams to make informed, forward-looking decisions grounded in accurate, real-time data. The dashboard's intuitive visualization, trend tracking, and drill-down capabilities turn complex operational data into simple, actionable intelligence — empowering users to move from reactive management to proactive optimization.

4. Model Performance and Predictive Accuracy

The forecasting models developed under this project achieved **over 90% accuracy**, ensuring highly dependable demand predictions across departments and SKUs. This level of precision enables pharmacy teams to anticipate requirements confidently, optimize inventory, and reduce both overstocking and shortages.

High predictive accuracy enhances **operational planning**, minimizes **emergency procurement**, and strengthens **vendor coordination**, directly improving cost efficiency and service reliability. The consistent alignment between forecasted and actual demand empowers decision-makers to act proactively rather than reactively, making the entire replenishment process more data-driven and resilient.

By embedding these accurate forecasts into the KPI dashboard, users gain **real-time, actionable intelligence** for procurement, stock management, and financial control — turning predictive analytics into a practical tool for hospital-wide optimization and patient care improvement.

Summary: Overall User Value of the KPI Dashboard

The integrated KPI dashboard serves as a **single source of truth** for all hospital operational, clinical, and financial stakeholders. By consolidating metrics across stability, recovery, and forecasting dimensions, it eliminates data silos and enables faster, smarter decision-making. Users no longer need to manually analyse reports — they can view trends, spot anomalies, and act immediately. The real-time visibility not only enhances accountability and collaboration across departments but also drives measurable improvements in patient care quality, cost efficiency, and revenue performance. In essence, the dashboard transforms KPI tracking from a reporting function into a **strategic decision-support system** that delivers tangible value to every user level — from pharmacy managers to hospital leadership.

References

1. **Pharmacy Dataset.** (n.d.). *In Kaggle.* https://www.kaggle.com/datasets/snowyowl22/pharmacy-dataset

2. **Prescription Drugs Market in India.** (n.d.). *In Statista.* https://www.statista.com/outlook/hmo/pharmacies/prescription-drugs/india

3. **Demystifying the Pharma Demand Forecasting Supply Chain Using Snowpark ML and Snowflake.** (2023, March 10). *In Medium.* https://medium.com/snowflake/demystifying-the-pharma-demand-forecasting-supply-chain-using-snowpark-ml-snowflake-e232ee655901

4. **Predicting Drug Demand in Pharmaceutical Supply Chains.** (2021). *In ScienceDirect.* https://www.sciencedirect.com/science/article/pii/S2214716021000257

5. **A Study on Inventory Management Practices in the Pharmaceutical Sector.** (2023). *In Quest Journals – Journal of Research in Business and Management.* https://www.questjournals.org/jrbm/papers/vol13-issue4/1304108116.pdf

6. **Pharmaceutical Supply Chain Management: Current Challenges and Future Trends.** (2022). *In National Center for Biotechnology Information (NCBI) – PMC.* https://pmc.ncbi.nlm.nih.gov/articles/PMC8954113/