DOM TEAM

DATASTORM 2025
Round 2

FMCG SALES DEMAND FORECASTING & ANALYTICS PLATFORM

A Data-driven Solution to the Demand Illusion

The Core Problem: Censored Demand

The Optimization Dilemma: Retailers struggle to balance Stock-outs (lost revenue) vs. Overstocking (holding costs).

The "Blind Spot": Censored Demand

- When a product is Out-of-Stock, sales drop to zero, but true demand is unrecorded (censored).
- Traditional models treat these zeros as "No Demand," leading to systematic underestimation of future sales.

Our Solution: The 2-Stage AI Pipeline

Goal: Reconstruct the "truth" of customer demand to provide unbiased forecasts.

Stage 1: Latent Demand Reconstruction (The "Repair" Phase)

- Action: Treat stock-out days as missing data.
- Model: XGBoost Regressor trained only on "in-stock" days to impute the true potential demand.

Stage 2: Precision Forecasting & Lead Time Prediction

- Action: Final models trained on the complete, unbiased "Repaired Dataset."

Result: Accurate demand forecasts and operational lead time predictions.

Key Differentiators & Strategic Impact

Strategic Impact: Value across the Triple Bottom Line:

- Economic: Capture lost revenue; optimize inventory holding costs (8-14 day horizon).
- Social & Environmental: Reduce food waste (SDG 12); improve consumer access.

Key Differentiators & Strategic Impact

What Makes DataStorm Unique?

- Censorship-Awareness: Explicitly handles Censored Demand using statistical recovery.
- Granular Precision: Forecasts at the SKU-Store-Day level, not broad averages.
- External Integration: Actively incorporates Temperature, Holidays, and Promotional data.
- Modern Tech Stack: Built on FastAPI (Backend) and Next.js 14 (Frontend) for scalability and superior UX.

ANALYZING THE DEMAND ILLUSION

Insight 1: Target Variable Skewness: units_sold is highly right-skewed. Action: Apply Log-Transformation (Log(1+x)) to stabilize the distribution for model training.

Insight 2: Temporal Patterns: Robust weekly seasonality (sales spikes on weekends). Action: weekday, is_holiday are critical features.

ANALYZING THE DEMAND ILLUSION

Insight 3: Censored Demand Proof: Visual evidence of a "triangle cut-off" in Stock on Hand vs. Units Sold plot, confirming sales are capped by supply at low inventory levels. Action: Only train Stage 1 on sufficient-stock days.

Insight 4: Lead Time Uniformity: Lead time is discrete/multi-modal, with no variance across different cities or major suppliers. Action: Strategy shifts to a Classification Problem focused on Temporal/SKU features, not Geography/Supplier ID.

MODEL PERFORMANCE

Stage 2: Precision Forecasting

Horizon	MSE	RMSE	MAE	R^2
T+1	696.6390	26.3939	16.7954	0.6486
T+7	712.4112	26.6910	16.9176	0.6433
T+14	728.0537	26.9825	17.0651	0.6379

Stage 2: Lead Time Prediction

Horizon	MSE	RMSE	MAE	R^2
Lead Time Validation	4.0517	2.0129	1.6247	0.5312

Solution Architecture & Technology Stack

Architecture: Three-tier, Cloud-Native Application centered around the 2-Stage AI Pipeline.

Data Flow: Data Ingestion -> Data Engineering Pipeline (MLOps) -> AI Pipeline Execution -> Backend Services -> Frontend Visualization.

Backend: FastAPI (Python)

Frontend: Next.js 14 (React)

MLOps: Docker & Git/DVC

Frontend Dashboard Key Features (UX)