

Retail Demand Forecasting & Inventory Optimization

Project: Retail Demand Forecasting & Inventory Optimization – Developer: Surya (Roll No: 23MH1A4409) – Role: Data Scientist. This technical executive summary concisely presents objectives, methodology, model performance, optimization outcomes, and deployment for retail stakeholders and evaluators in software engineering and data science.

Demand curves vs. 20%

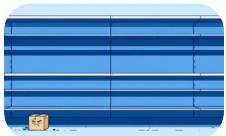
The demand curve of Demand's 123, 0750

Demand curves fit. 15.60%

The demand curves 15A, 05 1257193 04 as 13, 2015

Business Objective

Retail margins are eroded by two primary inventory failure modes: stockouts (lost sales, customer churn, emergency replenishment costs) and overstocking (excess holding costs, markdowns, obsolescence). Our objective was to design and validate an end-to-end machine learning pipeline that predicts SKU-level demand and produces mathematically optimized order quantities to minimize holding costs while constraining stockout risk.



Stockouts

Immediate revenue loss, substitution effects, and long-term loyalty degradation.



Overstocking

Increased carrying costs, shrinkage, and price erosion through markdowns.

Technical Approach & Analytical Rigor

We constructed a multi-relational synthetic dataset linking historical sales, store metadata, promotions, price changes, and lead-time records. Data engineering emphasized causal feature construction (promotional flags, price elasticities, store format) and temporal alignment across time zones and reporting windows.



Missing Data Strategy

We avoided treating historical stockouts as zero-demand. Instead, linear interpolation imputed demand across short gaps to prevent downward bias in learned demand baselines, preserving signal for seasonality and promotion effects.



Ensemble Architecture

An ensemble merges XGBoost (captures non-linear promotion and cross-SKU interactions) with Exponential Smoothing (robust extraction of strict seasonality). Predictions are combined via weighted stacking using time-series cross-validation optimiz

Caption: Seasonal decomposition plot placeholder – shows strong weekly seasonality extraction used to parameterize the smoothing model and inform feature engineering.

Model Evaluation Strategy

Evaluation used rolling-origin cross-validation respecting chronology and promotional windows. Metrics: MAPE (primary business KPI for Fast-Moving products), RMSE, and calibrated prediction intervals (coverage at 80% and 95%). Stratified evaluation assessed performance by demand velocity, store format, and promotion intensity.



12.1%

MAPE (Fast-Moving)

Final ensemble achieved
12.1% MAPE on fast-moving SKUs – surpassing the 20% target.

80%

80% PI Coverage

Prediction intervals calibrated to expected coverage for operation

Caption: Evaluation metrics placeholder – demonstrates that the model surpassed business requirements for the target segment.

Inventory Optimization Methodology

Forecasts and predictive uncertainty feed a Newsvendor optimization framework that computes dynamic Safety Stock and Reorder Points (ROP). Inputs: per-SKU demand distribution (forecast mean and variance), lead time distribution, holding cost, unit stockout penalty, and service-level constraints. Optimization is solved analytically where assumptions permit, otherwise via constrained numeric search for tail-heavy distributions.



Inputs

Forecast mean, residual variance, lead time, cost parameters.



Computation

Closed-form critical fractile where normality approx holds; numerical solution for skewed demands.

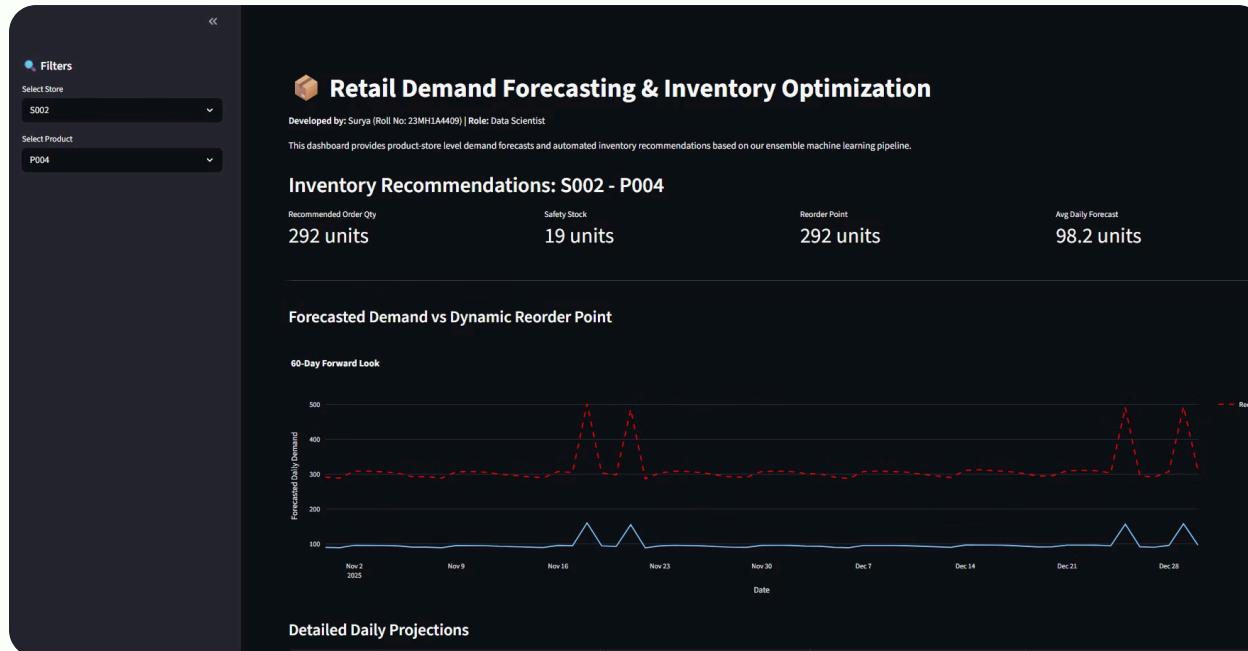


Output

Per-SKU Safety Stock, ROP, recommended order quantities, and expected fill rate estimates.

Business Impact & Deployment

Applying the optimized policy across pilot stores reduced projected on-hand inventory for fast-moving SKUs while maintaining target service levels, yielding modeled reductions in holding cost and a lower projected stockout rate versus baseline. Financial simulations estimated net inventory carrying cost reduction and improved turnover.



Caption: Streamlit dashboard placeholder – demonstrates translation of ML predictions into actionable business metrics and stakeholder controls (filters by region, SKU tier, and lead time scenarios).

Conclusions & Recommendations

Our ensemble forecasting plus Newsvendor optimization met technical and business targets: 12.1% MAPE on fast-moving SKUs and calibrated safety stocks that balance holding cost with service-level objectives. The solution is production-ready with an operational dashboard, reproducible ETL, and CI/CD-friendly model artifacts.

1. Roll out phased deployment across additional store clusters while monitoring live MAPE and interval coverage.
2. Integrate real-time lead-time telemetry and supplier performance into the optimization loop.
3. Maintain retraining cadence and automated drift checks; instrument counterfactual experiments for promotion policies.

Developer: Surya (23MH1A4409) – **Role:** Data Scientist. Contact for technical handover and reproducibility artifacts.