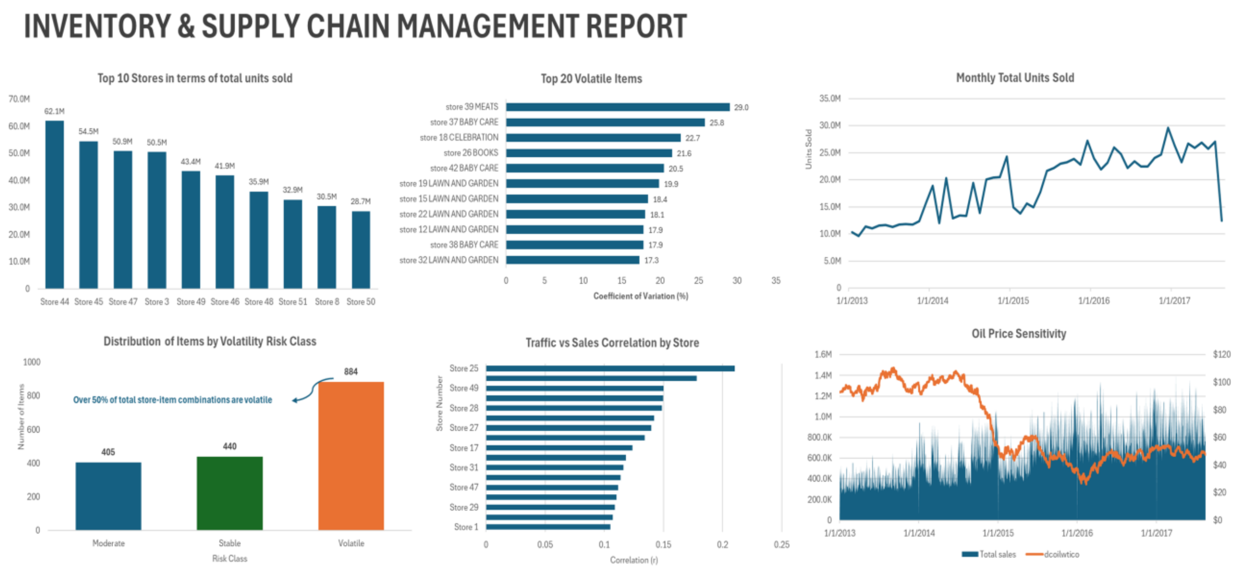
Inventory & Supply Chain Management Project Report: A Decision-Oriented Demand Risk Analysis Using Store Sales Time Series Data



1. Problem: What decision is at stake?

The central decision in this project is how inventory should be allocated, buffered, and governed across stores and product categories under demand uncertainty. Specifically, the business entity must decide whether to continue using largely uniform inventory rules, or to adopt differentiated policies based on demand scale, volatility, seasonality, and store importance.

Over-buffering ties up capital and increases holding costs, while under-buffering, especially in high-contribution stores or volatile categories can create systemic stockout risk. The project therefore reframes demand data as an input into *inventory risk management*, rather than treating forecasting as an end in itself.

2. Data Reality: Limitations, missing data, and noise

The analysis relies on historical sales data sourced from the Kaggle "Store Sales – Time Series Forecasting" dataset. While the data is extensive in temporal coverage and granularity, it however, reflects sales outcomes rather than true demand. This means stockouts, substitution effects, and lost sales are not directly observable. This introduces downward bias in observed demand during constrained periods.

Several operational variables are absent including lead times, supplier constraints, replenishment frequency, and on-hand inventory levels. Traffic data and macroeconomic indicators are available but imperfect proxies for customer intent and purchasing power. Additionally, the dataset includes noise arising from promotions, intermittent demand, and calendar effects that are not fully disaggregated by causal driver.

These limitations mean the analysis cannot claim causal precision. Instead, it operates in a decision-support framework despite noise and incomplete observability.

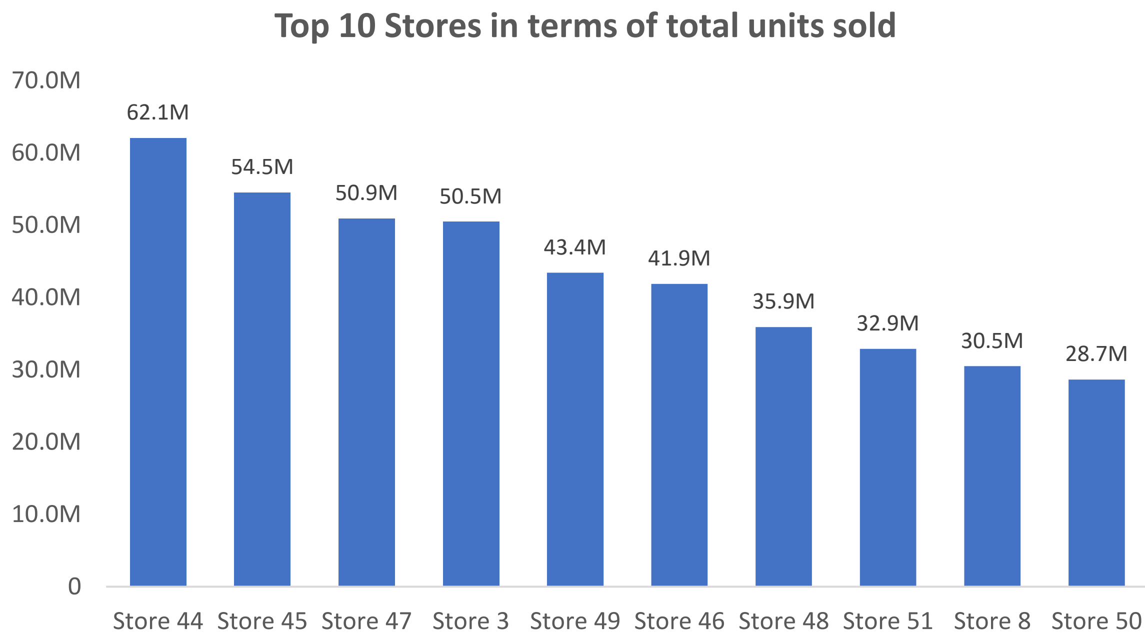
3. Method: Why this approach?

Rather than deploying a single predictive model, the project uses a multi-angle analytical approach focused on inventory relevance. Time-series aggregation, volatility measurement, and segmentation were chosen because inventory decisions depend less on point forecasts and more on distributional properties of demand, that is, scale, variance, and temporal concentration.

Key techniques include:

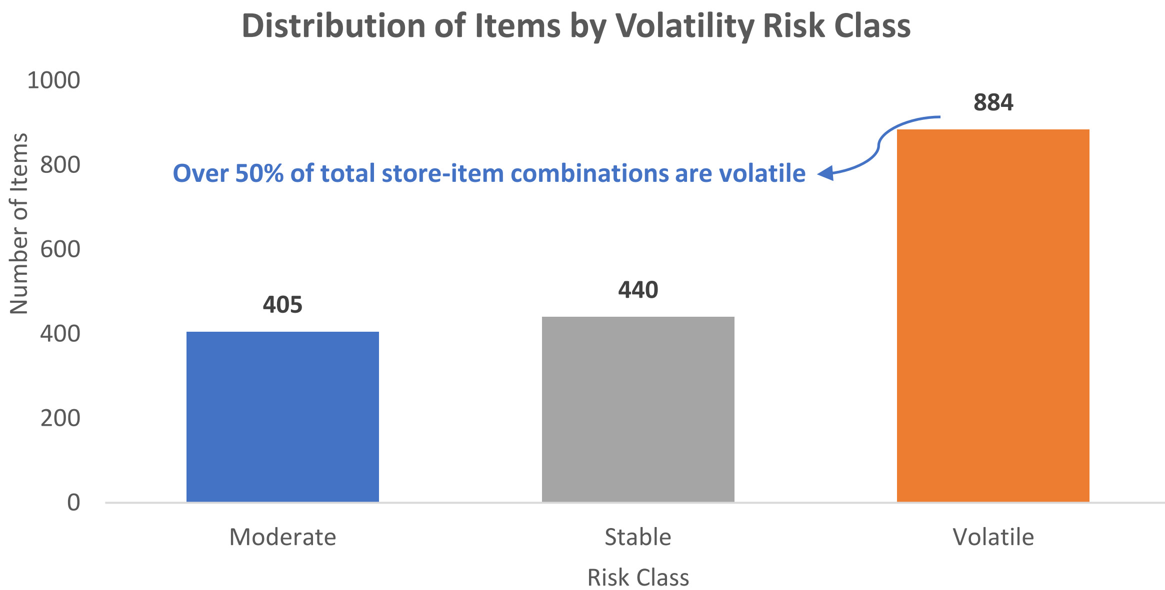
- Store-level demand aggregation to identify inventory-critical locations.
- Monthly time-series analysis to capture seasonality and growth.
- Coefficient of variation to quantify forecast risk and safety stock needs.
- Risk segmentation to enable differentiated inventory policy.
- Correlation analysis to evaluate commonly assumed demand drivers such as traffic and macro factors.

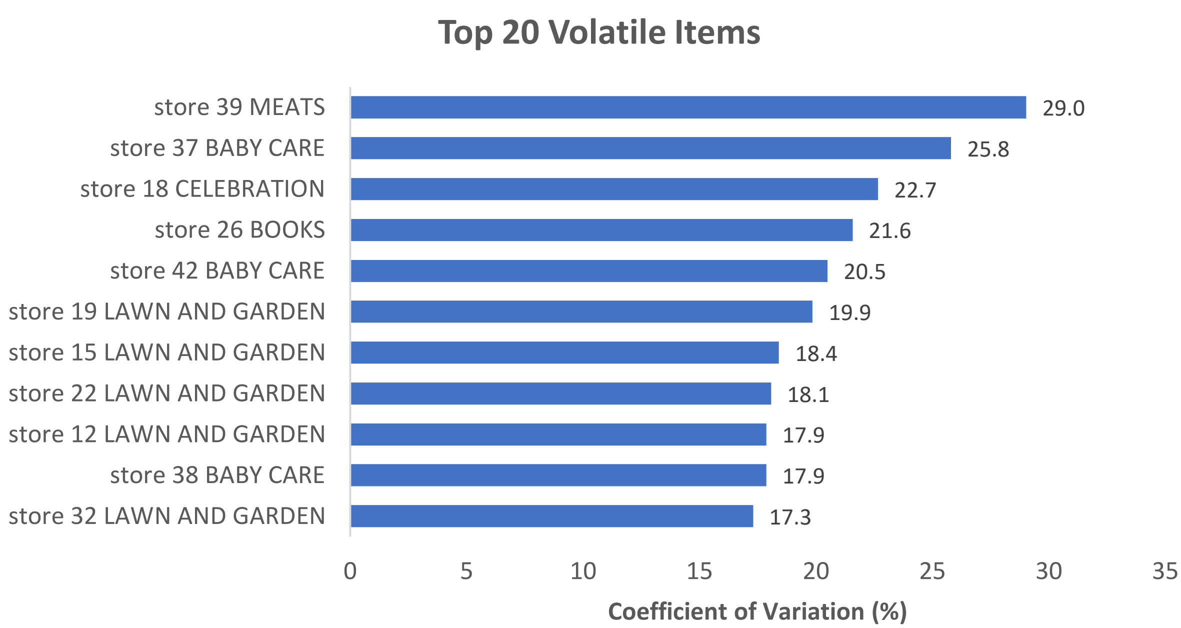
4. Results: What changed?



First, demand is shown to be highly concentrated at the store level. The top ten stores each sold between approximately 28 million and 62 million units over the observed period, with the single highest-performing store exceeding 62 million units. In contrast, the long tail of stores contributes materially less on an individual basis. This concentration implies that inventory failures in a small subset of locations can impact a disproportionate share of total system sales.

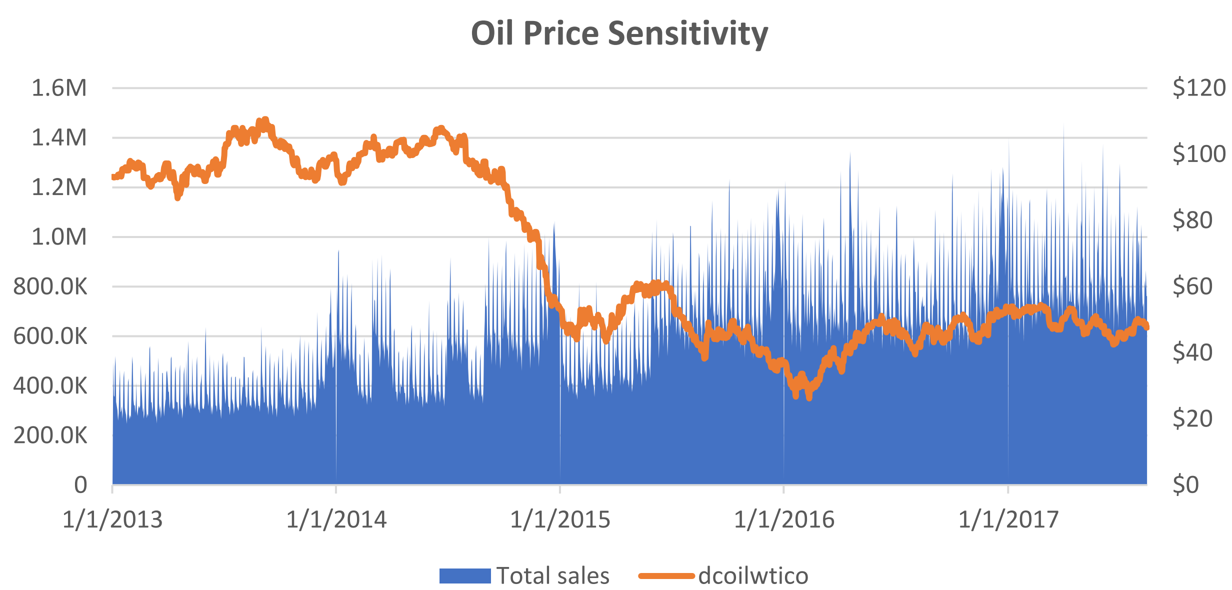
At the temporal level, total monthly units sold increased steadily from approximately 10–12 million units per month in 2013 to regular peaks exceeding 25–29 million units per month by 2016–2017. End-of-year seasonality is pronounced, with December consistently representing the highest demand month. For example, December sales rose from roughly 15.8 million units in 2013 to nearly 29.6 million units in 2016, representing almost a 90% increase over the period. This sustained growth and predictable seasonality demonstrate that static safety stock levels would have become increasingly misaligned with actual demand over time.



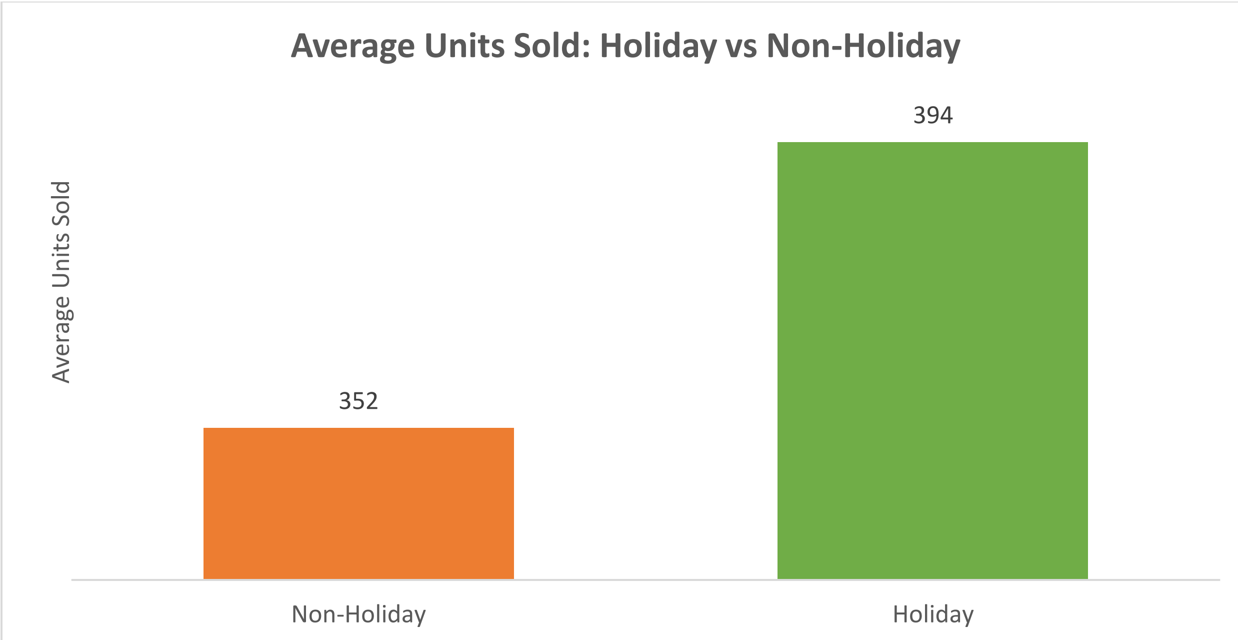


Demand volatility analysis further reveals that risk is both widespread and severe. Out of approximately 1,729 store–item combinations, 884 (over 51%) were classified as volatile, compared with 440 stable and 405 moderate-risk items. High-volatility items exhibited coefficients of variation ranging from approximately 9 to 29, indicating that demand variability often exceeded average demand by an order of magnitude. Volatility was especially concentrated in Lawn and Garden and Baby Care categories, but was also observed in high-volume categories such as Meats, confirming that scale alone does not mitigate forecast risk.

Correlation analysis between traffic and sales produced uniformly low coefficients across stores, with most correlations clustering between 0.10 and 0.15, and the highest observed values remaining below 0.25. This indicates that traffic explains only a small fraction of sales variance, undermining its usefulness as a primary inventory signal. Similarly, comparison of total sales with oil price movements showed no stable or actionable relationship, revealing limited sensitivity of short-term inventory demand to macroeconomic price fluctuations within the observed period.



Finally, holiday impact analysis quantified a clear but bounded demand uplift. Average daily units sold increased from approximately 352 units on non-holidays to nearly 394 units on holidays, representing an uplift of roughly 12%. While modest at the daily level, this effect compounds during peak seasonal periods, materially increasing inventory stress during already high-demand windows.



Collectively, these results demonstrate that demand risk is concentrated, growing, volatile, and weakly linked to intuitive proxies, reinforcing the need for differentiated, dynamic inventory policies grounded in empirical demand behavior rather than uniform assumptions.

5. Risk: What if this is wrong?

The primary risk is over-interpreting historical patterns as stable structural truths. Changes in consumer behavior, supply chain configuration, or pricing strategy could alter demand dynamics in ways not captured by the historical data. Additionally, unobserved stockouts may mask true demand volatility.

Another risk lies in misclassification: items labeled as volatile may stabilize under different merchandising or promotional strategies, while stable items may become volatile under new market conditions. Finally, correlation-based analyses do not imply causation, and decision-makers must avoid attributing intent or mechanism where only association is measured.

To mitigate these risks, the findings should be treated as decision priors rather than fixed rules. Inventory policies should be adaptive, periodically revalidated, and monitored using leading indicators such as service levels, stockout frequency, and inventory turns. The value of this analysis lies in enabling faster, more informed adjustments when reality diverges from expectation.