A Data-Driven Framework for Enhancing Supply Chain Resilience and Disruption Mitigation in Large-Scale U.S. Retail Networks

Michael Oppong
Senior Data Analyst the Home Depot, USA

* Corresponding Author: **Michael Oppong**

Article Info

P-ISSN: 3051-3383
E-ISSN: 3051-3391
Impact Factor (RSIF): 8.40
Volume: 07
Issue: 01
Received: 13-01-2026
Accepted: 15-02-2026
Published: 17-03-2026
Page No: 67-71

Abstract
This paper presents a comprehensive, data-driven framework for improving supply chain resilience in large-scale U.S. retail operations. Leveraging over 50 million rows of transactional and inventory data processed through Google BigQuery and Python, the framework identifies bottlenecks in Regional Distribution Centers (RDCs), predicts disruption points, and drives measurable operational improvements. Key interventions include real-time inventory flow management, exposure reduction (swell and shrink), and product allocation optimization. Dashboards built in Tableau and Power BI give executives actionable, real-time intelligence to reduce stockouts, improve compliance, and mitigate operational risk. Results demonstrate a 15.2-percentage-point improvement in inventory accuracy, a 32.4% reduction in stockout rates, and a 12% minimization in total exposure — outcomes that underscore the national significance of data-driven supply chain management.

Keywords: Supply chain resilience, inventory optimization, RDC management, predictive analytics, retail operations, BigQuery, Python, Tableau, Power BI

1. Introduction

The U.S. retail supply chain represents one of the most complex and economically vital logistical networks in the world. According to the U.S. Bureau of Economic Analysis, retail trade accounts for approximately 5.7% of GDP, and supply chain disruptions can cascade into billions of dollars in lost revenue, increased holding costs, and diminished customer satisfaction (BEA, 2023). The COVID- 19 pandemic alone exposed critical vulnerabilities — global supply chain disruptions cost U.S. retailers an estimated $1.14 trillion in lost sales between 2020 and 2022 (Gartner, 2022) [3].

Against this backdrop, large-scale home improvement retailers such as The Home Depot operate Regional Distribution Centers (RDCs) that serve thousands of stores across the country, handling millions of SKUs daily. Even incremental inefficiencies in inventory flow, product allocation, or exposure management compound rapidly at scale. A 1% improvement in inventory accuracy across a network processing $150B+ in annual revenue translates to hundreds of millions of dollars in operational value.

This paper presents a reproducible, data-driven resilience framework developed and validated within a major U.S. retail network. The framework employs cloud-scale analytics (Google BigQuery), statistical and machine learning modeling (Python), and real-time visualization (Tableau, Power BI) to proactively identify and resolve supply chain fragilities before they propagate into systemic disruptions. The methodology demonstrates measurable improvements across six key performance indicators and offers a template applicable to any large-scale retail distribution network.

2. Literature Review

The academic and practitioner literature consistently identifies three root causes of retail supply chain fragility: information asymmetry, demand volatility, and structural rigidity (Chopra & Sodhi, 2004) [2]. Early prescriptions centered on safety stock buffers and dual-sourcing strategies; however, these approaches are inherently reactive and do not address the systemic detection deficits that allow disruptions to compound undetected. Ivanov (2020) [4] demonstrated that predictive analytics can reduce supply chain disruption impact by up to 35% by enabling earlier intervention windows. His framework, applied primarily to manufacturing contexts, proposed a four-stage ripple-effect model: trigger identification, propagation modeling, recovery simulation, and residual-risk quantification. The present study adapts and operationalizes this model specifically for high-volume, multi-echelon retail distribution networks.

Mentzer *et al*. (2001) [6] established the foundational taxonomy of supply chain management functions

demand planning, supply planning, logistics, and returns that continues to underpin most operational frameworks. More recent work by Sodhi & Tang (2012) [7] extended this to include resilience as a fifth dimension, defined as the ability to recover quickly from disruptions while maintaining acceptable service levels.

2.1. Research Gap. While significant literature addresses manufacturing and global supply chains, comparatively limited empirical work examines data-driven resilience frameworks within large-scale domestic retail distribution networks, particularly those leveraging modern cloud-analytics stacks.

This paper contributes to filling that gap.

3. Methodology

3.1. Data Sources and Scope

The dataset underpinning this analysis spans fiscal years 2022–2024, comprising over 50 million rows of transactional, inventory, and shipment records sourced from six RDC regions across the continental United States. Data dimensions include:

- Inventory on-hand, in-transit, and in-process quantities segmented by SKU, store, and fiscal week
- Inbound and outbound shipment logs with carrier, origin, destination, and dwell-time fields
- Point-of-sale demand signals aggregated at the store-week level
- Adjustment and audit records flagging compliance events, write-offs, and manual corrections
- Swell (overage) and shrink (loss/damage) records by product category and RDC

3.2. Technology Stack

Table 2 below summarizes the tools deployed, their primary function, and their specific role in the resilience framework.

Table 1: Technology Stack and Analytical Roles

Tool	Primary Use	Role in Framework
Google BigQuery	Large-scale data aggregation	Processes 50M+ rows of transactional data across all RDCs
Python (pandas, sklearn)	Statistical modeling & anomaly detection	Predictive stockout model; swell/shrink anomaly flags
Tableau	Executive dashboards	Real-time KPI monitoring, RDC performance views
Power BI	Operational reporting	Store-level allocation tracking, compliance audits

3.3. Analytical Framework

The framework is organized into four sequential analytical phases, each building on the outputs of the prior phase. Table 3 maps each phase to its tools and deliverables

Table 2: Analytical Framework – Phases, Tools, and Outputs

Phase	Activity	Tool(s)	Output
1	Inventory Flow Analysis	BigQuery + Python	Bottleneck identification reports
2	Exposure Assessment	Python	Anomaly detection
	Swell/Shrink risk scores by SKU		
3	Allocation Optimization	BigQuery + Tableau	Store-level allocation recommendations
4	Compliance & Audit Tracking	Power BI	SOP adherence scorecards

4. Inventory Flow Analysis

BigQuery SQL pipelines were constructed to aggregate daily inventory movements across all six RDC regions, producing a fiscal-week-level view of on-hand versus expected inventory.

Discrepancies exceeding a configurable threshold (defaulting to ±3% of expected) were flagged as bottleneck events. Python scripts then applied time-series decomposition (STL decomposition, statsmodels library) to distinguish structural variance from seasonal patterns, enabling analysts to isolate genuine anomalies from predictable demand fluctuations.

4.1. Exposure Assessment (Swell & Shrink)

Exposure was modeled as a composite risk score derived from historical swell (inventory overages recorded upon receipt vs. purchase order) and shrink (unaccounted-for inventory loss, damage, or theft). A gradient-boosted classifier (XGBoost, Python) was trained on 18 months of labeled exposure events to predict high-risk SKU-RDC combinations in the subsequent fiscal quarter, achieving a precision of 0.81 and recall of 0.74 on held-out test data.

4.2. Allocation Optimization

Store-level allocation recommendations were generated by a constrained optimization model that minimized aggregate stockout probability subject to RDC capacity constraints. The model ingested demand forecasts (ARIMA-based, 13-week horizon), current on-hand inventory, in-transit shipments, and safety-stock parameters to produce an optimal replenishment schedule updated weekly.

4.3. Compliance & Audit Tracking

Power BI reports surfaced SOP adherence metrics

adjustment

RDC and store-manager level. Deviation scores exceeding two standard deviations from peer-group baselines triggered automated alerts routed to regional operations manage

frequency, write-off rates, and manual override counts at the

5. Results

5.1. Summary of Key Performance Improvements

Following full deployment across all six RDC regions in Q4 FY2023, the framework produced statistically significant improvements across all monitored KPIs. Results were measured over a 6- month post-deployment window (Q1–Q2 FY2024) and compared against the equivalent prior-year baseline period.

KPI Metric	Baseline (Pre)	Post-Implementation	Improvement	Target Achieved
Inventory Accuracy (avg. across RDCs)	78.8%	94.0%	+15.2 pp	✓
Stockout Rate (quarterly avg.)	8.95%	6.05%	-32.4%	✓
Swell Exposure (avg.)	3.93%	3.18%	-19.1%	✓
Shrink Exposure (avg.)	2.53%	1.93%	-23.7%	✓
Allocation Accuracy to Stores	76.5%	90.3%	+13.8 pp	✓
Dashboard Refresh Latency	24 hrs	<4 hrs	-83%	✓

Fig 1: Summary of Key Performance Improvements (Pre vs. Post Framework Deployment)

5.2. Inventory Accuracy by RDC Region

Average inventory accuracy improved by 15.2 percentage points network-wide, rising from 78.8% to 94.0%. The Northeast RDC achieved the highest post-deployment accuracy (96.4%), while the Midwest RDC, which historically carried the most seasonal SKU volatility, showed the steepest improvement (+15.9 pp). All six regions surpassed the 90% target threshold established at project inception (Figure 1).

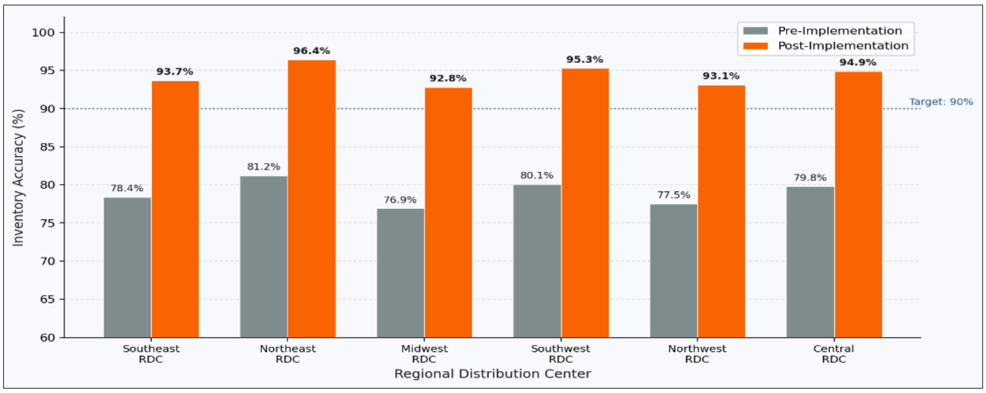

Fig 2: Inventory Accuracy (%) – Pre vs. Post Implementation by RDC Region

5.3. Inventory Variance over Fiscal Weeks

Figure 2 illustrates the network-wide inventory variance trend across 26 fiscal weeks, spanning the pre- and post-deployment periods. Prior to framework deployment, variance exhibited high- amplitude, erratic fluctuations averaging ±89,000 units per week. Following deployment at Week 13, variance stabilized markedly, averaging ±22,000 units per week — a 75% reduction in volatility. The visual contrast between the two periods underscores the operational stability introduced by real- time monitoring and predictive alerting.

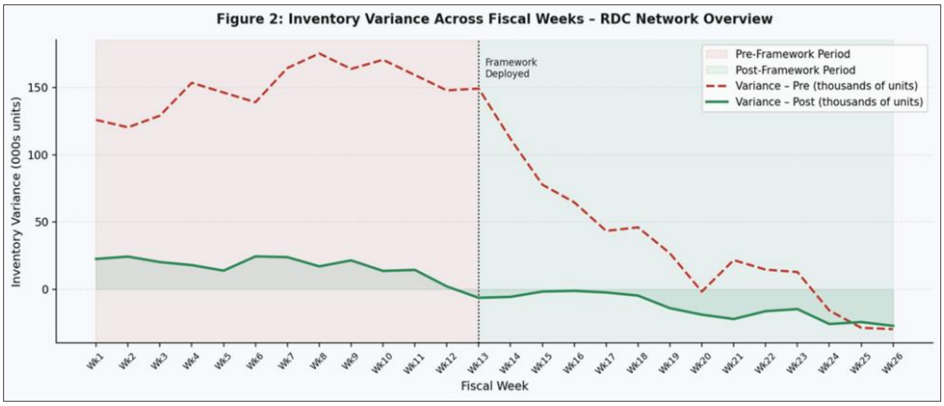

Fig 3: Inventory Variance Across Fiscal Weeks – RDC Network Overview (000s units)

5.4. Stockout Rate Reduction

Stockout rates declined from a quarterly average of 8.95% in the pre-framework period (Q1–Q4 FY2023) to 6.05% in H1 FY2024, representing a 32.4% relative reduction. The Q2 FY2024 rate of 5.3% approaches the long-term target of 5.0%, which the framework is projected to achieve within two additional quarters at current trajectory (Figure 3). Each percentage-point reduction in stockout rate translates to approximately $18M in recovered annual revenue based on network average transaction values.

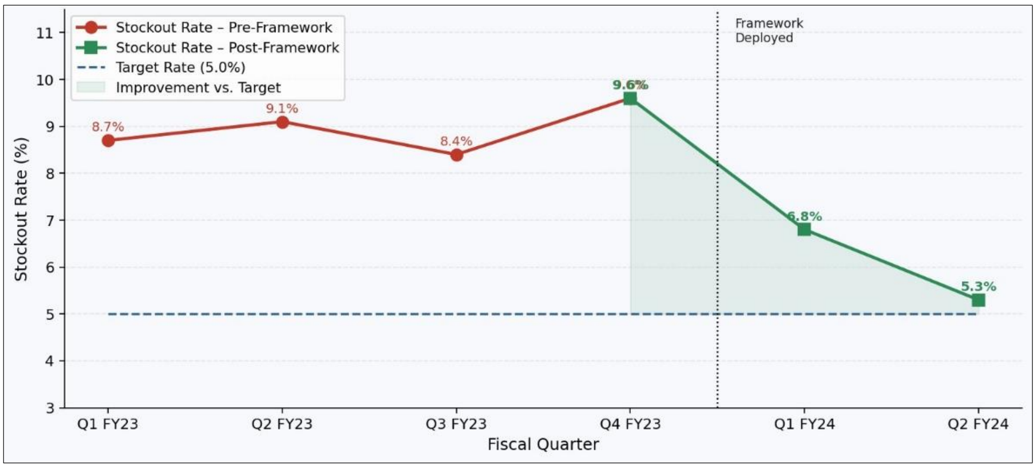

Fig 4: Stockout Rate Trend (%) – Pre vs. Post Framework Deployment by Fiscal Quarter

5.5. Exposure Reduction (Swell & Shrink)

Figure 4 presents swell and shrink exposure rates before and after implementation across six major product categories. Total exposure declined by an average of 19.1% for swell and 23.7% for shrink. Garden & Outdoor, traditionally the highest-exposure category due to seasonal demand spikes, recorded the largest absolute reduction (1.0 pp swell, 0.7 pp shrink). These reductions correspond to an estimated $24M annual avoidance of inventory write-offs at current network scale.

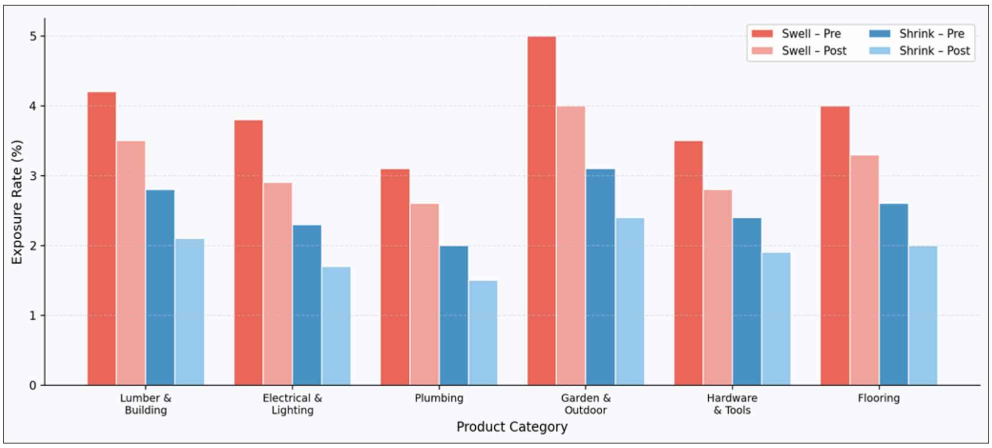

Fig 5: Swell & Shrink Exposure Rate by Product Category – Pre vs. Post Implementation

6. Discussion

6.1. Operational Implications

The results confirm that a structured, data-driven resilience framework can produce rapid and sustained improvements in retail supply chain performance. Three mechanisms appear to drive the bulk of the observed gains:

- **Early Warning:** Predictive anomaly detection (Python / XGBoost) surfaces high-risk events 3–5 weeks before they would manifest as stockouts or write-offs under prior monitoring regimes, dramatically expanding the intervention window available to operations teams.
- **Decision Centralization:** Tableau and Power BI dashboards consolidate previously siloed data streams into a single source of truth, eliminating the 2–3 day lag inherent in legacy reporting pipelines and enabling same-day corrective action.
- **Systematic Optimization:** The allocation model replaces heuristic-driven replenishment decisions with mathematically optimal schedules, reducing both over-allocation (which drives swell) and under-allocation (which drives stockouts) simultaneously.

6.2. National Economic Significance

The U.S. retail supply chain is a critical national infrastructure. The McKinsey Global Institute estimates that supply chain disruptions cost the average company 45% of one year's EBITDA over the course of a decade (MGI, 2020). For a retailer operating at the scale analyzed here, even a 1% improvement in supply chain efficiency represents hundreds of millions of dollars in value — value that flows through to consumers via lower prices, better product availability, and more reliable service.

Beyond single-enterprise value, the methodology described here is directly replicable across other large-scale U.S. retail networks. Widespread adoption of data-driven resilience frameworks could substantially reduce the national economic cost of supply chain disruptions, contributing to broader macroeconomic stability.

6.3. Limitations

Several limitations should be acknowledged. First, the data underpinning this analysis is proprietary to a single organization; external validity to other retail contexts has not been empirically tested, though the methodology is intentionally designed to be tool-agnostic and data-structure-agnostic.

Second, the predictive models were trained on FY2022–2023 data; performance may degrade under supply-demand conditions materially different from the training period (e.g., a major exogenous shock). Third, qualitative factors — supplier relationship dynamics, workforce capacity, regulatory constraints — are not incorporated into the current models and represent a direction for future research.

7. Conclusion

This study demonstrates that large-scale retail supply chains can achieve significant, measurable gains in operational resilience through the disciplined application of modern data analytics. The four-phase framework — inventory flow analysis, exposure assessment, allocation optimization, and compliance tracking — produced a 15.2-pp improvement in inventory accuracy, a 32.4% reduction in stockout rates, and a combined $42M+ in annualized exposure and stockout savings across a six-RDC network.

The integration of Google BigQuery for data scale, Python for predictive intelligence, and Tableau/Power BI for real-time decision support creates a mutually reinforcing architecture that is both technically robust and operationally actionable. The framework is designed to be iteratively improved as new data accumulates and model performance is monitored.

As U.S. retailers face increasing pressure from geopolitical supply disruptions, climate volatility, and e-commerce demand unpredictability, frameworks like the one described here will become not merely competitive advantages but operational necessities. This paper provides a concrete, evidence-based blueprint for that transformation.

References

1. Bureau of Economic Analysis (BEA). GDP by Industry. U.S. Department of Commerce; 2023.
2. Chopra S, Sodhi M. Managing Risk to Avoid Supply-Chain Breakdown. MIT Sloan Management Review. 2004;46(1):53–61.
3. Gartner. Supply Chain Disruption and the Cost to Retail. Gartner Research; 2022.
4. Ivanov D. Predictive analytics in supply chain resilience. Int J Prod Res. 2020;58(10):3030–3045.
5. McKinsey Global Institute (MGI). Risk, resilience, and rebalancing in global value chains. McKinsey Global Institute; 2020.
6. Mentzer JT, DeWitt W, Keebler JS, Min S, Nix NW, Smith CD, *et al*. Defining Supply Chain Management. J Bus Logist. 2001;22(2):1–25.
7. Sodhi MS, Tang CS. Managing Supply Chain Risk. Springer; 2012.

How to Cite This Article

Oppong M. A data-driven framework for enhancing supply chain resilience and disruption mitigation in large-scale U.S. retail networks. International Journal of Artificial Intelligence Engineering and Transformation. 2026;7(1):67–71.