



ASSESSMENT OF DATA-DRIVEN VENDOR PERFORMANCE EVALUATION IN RETAIL SUPPLY CHAINS ANALYZING METRICS, SCORECARDS, AND CONTRACT MANAGEMENT TOOLS

Md. Abdur Rauf¹; Zamal Haider Shish²;

[1]. Project Management Trainer, Center for Policy and Project Management, Dhaka, Bangladesh;
Email: raufshiblu@gmail.com

[2]. Executive Marketing & Technical Support, Unicon Solution Ltd., Dhaka, Bangladesh;
Email: hsjidan@gmail.com

Doi: [10.63125/2a641k35](https://doi.org/10.63125/2a641k35)

Received: 29 September 2022; Revised: 22 October 2022; Accepted: 27 November 2022; Published: 27 December 2022;

Abstract

This quantitative explanatory study assessed data-driven vendor performance evaluation in retail supply chains by integrating multi-KPI scorecards with downstream operational outcomes. The measurement architecture and variable logic were grounded in a structured review of 32 empirical papers on retail vendor evaluation, KPI taxonomies, scorecard construction, and outcome modeling, which informed the selection of delivery, quality, cost, and flexibility indicators. Operational data were compiled from ERP purchasing tables, WMS receipt logs, TMS carrier scans, POS demand records, supplier portal submissions, and claims databases. Of 312 vendors initially identified, 47 inactive vendors and 23 vendors with insufficient transaction histories were excluded, yielding 242 active vendors observed from January 2021 to December 2021. After data cleaning, 3,805 vendor-period records remained for analysis. Descriptive results showed strong average delivery performance (on-time delivery mean 91.4%, SD 6.8; order fill rate mean 94.1%, SD 5.2), while lead-time deviation displayed wider dispersion (mean 2.6 days, SD 1.9). Quality performance was generally stable but risk-concentrated (defect rate mean 1.9%, SD 1.4; return ratio mean 2.7%, SD 2.0), and flexibility exhibited the highest volatility (rush-order acceptance mean 76.8%, SD 14.9; recovery time mean 4.1 days, SD 3.3). Reliability tests supported dimensional consistency (Cronbach's α 0.86 delivery, 0.83 quality, 0.79 cost, 0.76 flexibility; overall scorecard α 0.88). Correlation patterns confirmed KPI coherence (on-time delivery–fill rate r 0.74; defect rate–return ratio r 0.69) and strong alignment between vendor scores and retail outcomes (overall score–stockout r -0.58; overall score–shelf availability r 0.62). Driver models indicated that forecast variability (β -0.27, $p < .001$), distance (β -0.14, $p = .010$), and product complexity (β -0.17, $p = .003$) reduced vendor performance, while vendor size improved scores (β 0.18, $p = .002$). Outcome regressions demonstrated temporally ordered effects, with delivery strength reducing stockouts (β -0.36) and quality strength lowering return processing costs (β -0.29). Overall, the study showed that data-driven scorecards reliably differentiated vendor capability and explained meaningful variation in retail service and cost outcomes.

Keywords

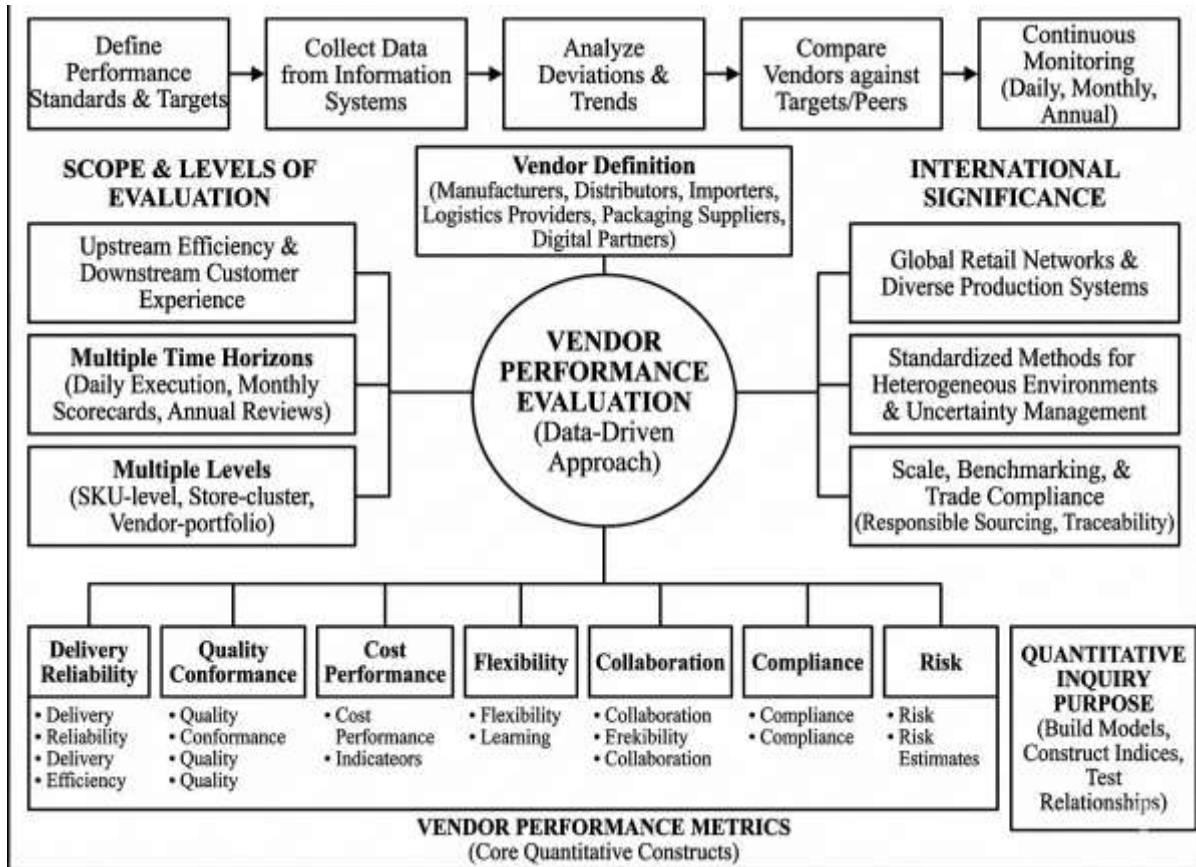
Vendor Performance, KPI Scorecards, Retail Supply Chains, Contract Governance, Outcome Modeling.

INTRODUCTION

Vendor performance evaluation in retail supply chains refers to a structured process through which retailers assess how well their vendors meet operational, commercial, and relational expectations. In this context, “vendors” include manufacturers, distributors, importers, third-party logistics providers, packaging suppliers, and digital service partners that influence product flow and retail service outcomes (Kamble & Gunasekaran, 2020). A data-driven approach means that evaluation relies primarily on measurable evidence captured through information systems rather than informal judgment. It converts vendor activities into quantifiable indicators such as delivery accuracy, order completeness, defect incidence, responsiveness, and cost adherence. The evaluation process is typically embedded in procurement and supply chain governance, where the retailer defines performance standards, collects data from routine transactions, analyzes deviations, and compares vendors against targets or peers. In retail settings, the scope of vendor evaluation is broad because vendors affect both upstream efficiency and downstream customer experience. Fast-moving product turnover, large assortments, and frequent promotions make retail highly sensitive to supply disruptions or quality lapses. Data-driven evaluation therefore emphasizes continuous monitoring rather than occasional audits. It operates across multiple time horizons, including daily execution metrics, monthly scorecard cycles, and annual contract reviews (Liou et al., 2019). It also spans multiple levels of analysis, such as SKU-level performance, store-cluster fulfillment, and vendor-portfolio contribution. Vendor performance evaluation is not limited to physical product delivery; it also accounts for information quality, documentation accuracy, sustainability compliance, and risk controls when these are relevant to retail operations. Retailers typically formalize this evaluation through performance dashboards, vendor portals, and periodic business reviews tied to numeric evidence. Within quantitative inquiry, this concept provides a measurable unit for modeling how vendor behavior influences retail supply chain outcomes. By defining evaluation as a data-centered measurement and comparison system, researchers can build statistical models, construct composite indices, and test relationships between vendor scores and operational or financial performance (Gawankar et al., 2020).

Data-driven vendor performance evaluation holds international significance because retail supply chains operate across borders and link diverse production systems to consumer markets worldwide. Global retail networks source commodities, apparel, electronics, pharmaceuticals, and household goods from vendors located in multiple regions, each shaped by distinct infrastructure quality, customs processes, regulatory frameworks, and labor or environmental standards (Yu et al., 2018). Under such conditions, retailers require standardized methods to compare vendor reliability across heterogeneous environments. Data-driven evaluation provides this standardization by translating varied vendor outputs into a shared measurement language. International retailing also increases exposure to uncertainty, including lead-time variability, transport bottlenecks, geopolitical shocks, and supplier compliance risks. A systematic evaluation framework helps retailers identify vendors that consistently deliver stable service while flagging sources of volatility (Almohri et al., 2019). The international dimension is further reinforced by the scale of modern retail. Large firms manage thousands of vendors, generating transactional datasets that enable rigorous quantitative screening and benchmarking. Such benchmarking supports cross-market alignment of service levels, ensuring that consumers in different geographies receive consistent product availability and quality. Vendor evaluation also interacts with international trade expectations related to responsible sourcing, traceability, and ethical compliance. Retailers use performance evidence to enforce product safety requirements and sustainability commitments, not as abstract principles but as observable contractual obligations. From a quantitative research viewpoint, international retail supply chains generate multi-currency purchasing data, multi-echelon logistics records, and multi-market demand patterns, enriching statistical analysis and allowing more generalizable models (Curuksu, 2018). Evaluation frameworks help stabilize these global systems by reducing information asymmetry and reinforcing accountability through measured outcomes. The international importance of vendor evaluation is therefore connected to market stability, consumer welfare, and the efficiency of cross-border supply operations (Chavez et al., 2017). It enables retailers to coordinate complex networks at scale while preserving comparability across vendors that differ widely in capability, distance, and institutional context.

Figure 1: Data-Driven vendor Performance Evaluation in Retail Supply Chain

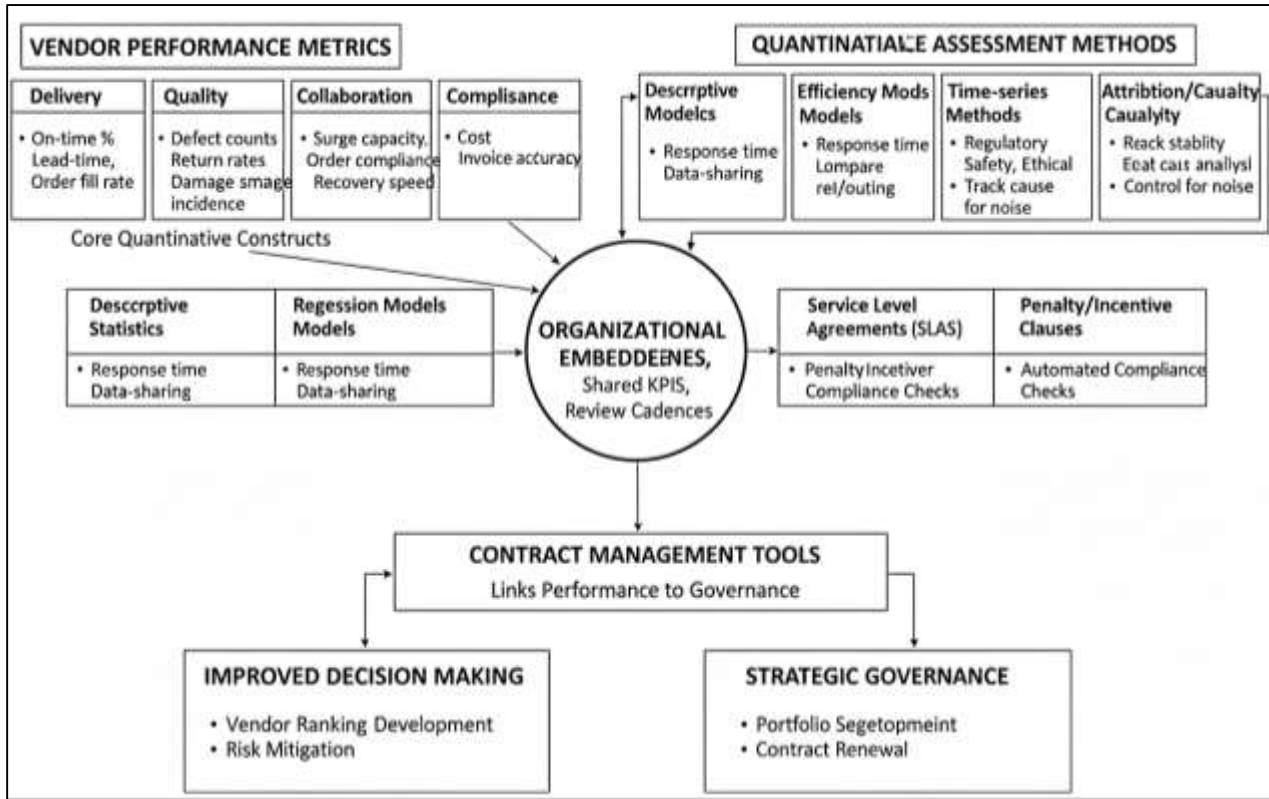


Vendor performance metrics are the core quantitative constructs that make data-driven evaluation possible. Metrics convert expectations into measurable indicators that can be repeatedly calculated from operational data (Abdulla & Ibne, 2021; Miragliotta et al., 2018). In retail supply chains, metrics commonly cluster into domains of delivery reliability, quality conformance, cost performance, flexibility, collaboration, compliance, and risk. Delivery metrics capture whether vendors ship the right goods at the right time and in the right quantities, using indicators such as on-time delivery percentage, lead-time adherence, shipment variance, and order fill rate (Ara, 2021; Habibullah & Foysal, 2021). Quality metrics represent the degree to which supplied products or services meet specification standards, often reflected through defect counts, rejected lot percentages, customer return rates, and damage incidence during transit. Cost and commercial metrics include purchase price variance, total landed cost, invoice accuracy, and cost-to-serve by channel (Hartmann et al., 2016; Sarwar, 2021; Musfiqur & Saba, 2021). Flexibility metrics capture responsiveness to demand changes, measuring surge capacity, minimum order compliance, substitution performance, and recovery speed after disruptions. Collaboration and service metrics transform relational expectations into numeric signals by measuring response time to queries, corrective action closure rates, and data-sharing timeliness. Compliance and risk metrics incorporate adherence to regulatory, safety, ethical, cybersecurity, or sustainability standards, especially for global sourcing. For quantitative work, metrics require careful definition to ensure validity and reliability (Redwanul et al., 2021; Tarek & Praveen, 2021). Validity means the metric truly represents the performance construct it is intended to measure. Reliability means consistent results when measurement conditions remain stable. Comparability requires normalization, because raw indicators differ in scale and volatility across product categories and vendor types (Muhammad & Shahrin, 2021; Reis, 2018; Saikat, 2021). Retail vendors supplying perishable foods, seasonal fashion, or high-value electronics face different operating constraints, so metrics must be contextualized through appropriate benchmarks. Quantitative studies often standardize metrics through target ratios, percentile scoring, or statistical scaling so that multiple indicators can be integrated into composite assessments. The design of metrics therefore determines both the managerial usefulness of evaluation systems and the robustness of empirical models that

examine vendor performance drivers (Manikas et al., 2020; Al Amin, 2022; Shaikh & Aditya, 2021). Vendor scorecards are structured tools that aggregate multiple metrics into a coherent profile of vendor performance. In retail supply chains, scorecards provide a shared platform for procurement, logistics, quality teams, finance, and category managers to interpret vendor contribution through a common numeric view. A scorecard typically organizes indicators into categories such as delivery, quality, cost, responsiveness, and compliance, then assigns a scoring rule for each metric (Ariful, 2022; Nahid, 2022; Morabito, 2015). Scoring rules translate raw values into performance points based on targets, thresholds, or comparative ranks. These points are then combined using weights that reflect retailer priorities. For instance, a retailer may weight cold-chain delivery and spoilage control more heavily for fresh grocery than for durable household goods. Composite scores allow vendors to be ranked, segmented, and tracked over time. Scorecards also support exception handling by highlighting areas where performance falls below critical thresholds. In quantitative research, scorecards serve as measurable dependent or independent variables. Researchers analyze sensitivity of composite scores to weighting schemes and test whether scorecard outcomes align with external indicators such as inventory turns, shelf availability, markdown ratios, or category profitability (Jusselme et al., 2018; Hossain & Milton, 2022; Mominul et al., 2022). Another important quantitative attribute is segmentation. Scorecards often distinguish strategic vendors from transactional vendors and benchmark them differently. This improves fairness and analytic accuracy since vendors have different roles and resource bases. Digitalization has expanded scorecards from periodic spreadsheets to live dashboards that refresh with transaction feeds, enabling high-frequency panel data. Such data supports longitudinal modeling of learning curves, seasonality, and performance drift. Scorecards thus bridge measurement and managerial governance by turning multidimensional vendor evidence into interpretable numeric summaries. In retail environments where vendor bases are large and operations are time-sensitive, scorecards provide the operational grammar through which performance evidence guides decisions (Rabiul & Praveen, 2022; Ntim et al., 2015; Rakibul & Samia, 2022).

Contract management tools connect vendor performance evidence to formal governance of retail relationships. Contract management encompasses agreement creation, storage, monitoring, and enforcement for terms that specify pricing, delivery windows, quality tolerances, service-level standards, penalties, incentives, and escalation procedures (Kivistö & Zalyevska, 2015; Saikat, 2022; Kanti & Shaikat, 2022). In data-driven systems, contract tools integrate with procurement platforms and enterprise databases so that contractual obligations are automatically compared with observed performance metrics. When vendors fall outside agreed service levels, the tool can trigger alerts, generate compliance reports, or initiate corrective workflows. This integration matters because performance measurement becomes operationally meaningful when linked to enforceable outcomes. Service level agreements represent the primary mechanism for this linkage, since they define performance thresholds in measurable form. Penalty clauses, bonus structures, and gain-share incentives are then calibrated against those thresholds. Quantitative research benefits from this environment because contractual rules create observable policy boundaries. If a contract changes its penalty rate, delivery tolerance, or incentive schedule, researchers can model the effect as a measurable intervention in time-series or panel data. Contract management tools also improve data integrity by ensuring that performance records align with the exact contract version in force during a given period. Audit trails and clause libraries help retailers maintain consistency across large vendor portfolios. In practice, retailers use contract tools to manage renewals, automate compliance checks, and reduce disputes over measurement accuracy. Within empirical analysis, this creates richer datasets where performance scores and contractual variables can be jointly examined to understand how formal governance shapes vendor behavior. Contract management tools therefore operate as a performance-evidence engine that translates numeric evaluation into structured control over vendor relationships. Quantitative assessment of vendor performance relies on analytical methods that transform operational datasets into reliable performance insights. Retail supply chains generate large volumes of time-stamped data from purchase orders, advance shipping notices, warehouse scans, carrier events, invoices, inspection logs, and claims records (Aggarwal & Ghosh, 2015).

Figure 2: Retail Vendor Performance Framework



Before analysis, these datasets require preprocessing, including identity matching across systems, resolution of missing values, elimination of duplicates, outlier treatment, and alignment of timestamps. Once prepared, the data supports a range of statistical and decision-analytic tools. Descriptive statistics summarize central tendencies and variability in vendor indicators. Econometric and regression models estimate relationships between vendor performance and retail outcomes while controlling for category, market, season, and channel. Efficiency and productivity models compare vendors based on multiple inputs and outputs to identify leaders and laggards under fair assumptions. Multi-criteria decision approaches compute weighted composite scores that reflect strategic priorities and allow trade-off analysis across performance domains. Cluster analysis can uncover latent vendor segments based on behavioral patterns, and time-series methods track performance stability, shocks, and recovery paths. A key methodological concern is attribution, since measured vendor outcomes can be influenced by retailer forecasting errors, order batching policies, or downstream receiving limitations. Quantitative designs address this with fixed-effects models, hierarchical modeling, or controlled comparisons across similar conditions. Another concern is simultaneous influence, where retailer decisions may both respond to and affect vendor performance. Researchers handle this by using lag structures, instrumental logic, or experimental-style comparisons around policy changes. The analytical toolkit therefore ensures that data-driven evaluation moves beyond reporting to rigorous statistical understanding. In retail, this rigor helps separate persistent vendor capability differences from situational noise, producing more accurate performance rankings and more defensible governance actions.

Data-driven vendor performance evaluation functions within organizational and cross-functional structures that determine how performance evidence is interpreted and used. Retail evaluation systems are rarely owned by procurement alone (Antonanzas et al., 2019). Category managers define assortment expectations, planners manage replenishment targets, logistics teams track lane reliability, quality divisions oversee compliance, and finance verifies commercial accuracy. Data-driven evaluation depends on shared KPI definitions and coordinated data stewardship so that each function measures vendors in consistent ways. Review cadences institutionalize this coordination through operational check-ins, monthly scorecard discussions, and formal vendor business reviews. These

routines create accountability by making performance visible, comparable, and traceable to numeric records. Organizational embeddedness also influences behavior (Anund Vogel et al., 2016). When scorecards are transparent and stable, vendors can align internal processes to measured expectations, and internal teams can coordinate improvement priorities. Governance systems often include tiered escalation pathways, where minor deviations trigger joint corrective plans and persistent underperformance leads to contract penalties or vendor replacement. The embeddedness of evaluation further affects data quality. If receiving teams fail to scan deliveries promptly or if buyers override system records, measured performance loses accuracy (McDermott et al., 2018). Retailers therefore implement standardized operating procedures for data capture and verification. Within quantitative research, organizational embeddedness helps explain variance in performance outcomes that is not purely technical. Vendors may perform differently when evaluation results are tightly linked to incentives and decision rights compared with cases where scorecards are treated as informal reports. Evaluation systems also shape vendor portfolio strategy by supporting segmentation into strategic partners, growth vendors, and transactional suppliers. In this way, data-driven evaluation becomes a living governance architecture rather than a static measurement task (Pernell, 2020). It integrates metrics, scorecards, and contracts into routine retail decision processes, making vendor performance a measurable and managed driver of supply chain effectiveness (Dias et al., 2020).

The objective of this quantitative paper is to assess data-driven vendor performance evaluation in retail supply chains by examining how measurable metrics, structured scorecards, and contract management tools collectively shape vendor accountability and supply chain outcomes. Specifically, the study aims to identify and categorize the most commonly used vendor performance metrics in retail contexts, including delivery reliability, order fill accuracy, product quality consistency, cost adherence, responsiveness, flexibility, and compliance indicators, and to determine how these metrics are operationalized through transactional and logistics data. A further objective is to analyze the construction and application of vendor scorecards as multi-dimensional performance profiling instruments, evaluating how retailers select indicators, assign weights, normalize values, and generate composite performance ratings that support vendor comparison, segmentation, and monitoring over time. The paper also seeks to measure the statistical relationships between vendor scorecard results and key retail supply chain performance areas such as inventory availability, replenishment stability, lead time variability, return frequency, and cost-to-serve, using empirical evidence drawn from real operational datasets. Another core objective is to investigate the role of contract management tools in reinforcing data-driven evaluation by linking observed performance to service level agreements, penalty or incentive clauses, renewal decisions, and escalation workflows. This includes assessing how automated contract monitoring and clause compliance tracking affect the consistency, speed, and fairness of vendor governance. In addition, the study aims to examine the quality and structure of vendor performance datasets used in retail firms, focusing on completeness, accuracy, frequency, and cross-system integration, and to evaluate how data readiness influences the strength of vendor evaluation models. The paper further intends to compare performance patterns across vendor categories and product segments, highlighting whether strategic vendors, transactional vendors, local suppliers, and global suppliers exhibit statistically different performance profiles under the same evaluation framework. Through these objectives, the research aims to generate a rigorous measurement-based understanding of vendor evaluation systems in retail supply chains, emphasizing the practical mechanics of metrics, the interpretability of scorecards, and the enforceability of contract-linked controls within a unified quantitative assessment design.

LITERATURE REVIEW

This literature review examines the scholarly foundations and quantitative evidence surrounding data-driven vendor performance evaluation within retail supply chains. The section is designed to map how prior studies have defined vendor performance, operationalized measurable indicators, and constructed evaluation models that support vendor selection, monitoring, and contractual governance. Because the present paper is quantitative, the review emphasizes empirical research that uses numeric metrics, statistical testing, multi-criteria scoring, efficiency modeling, and large-scale transactional datasets. The objective is to synthesize what is already known about vendor performance measurement systems, including the dominant KPI families in retail logistics and procurement, the mathematical and

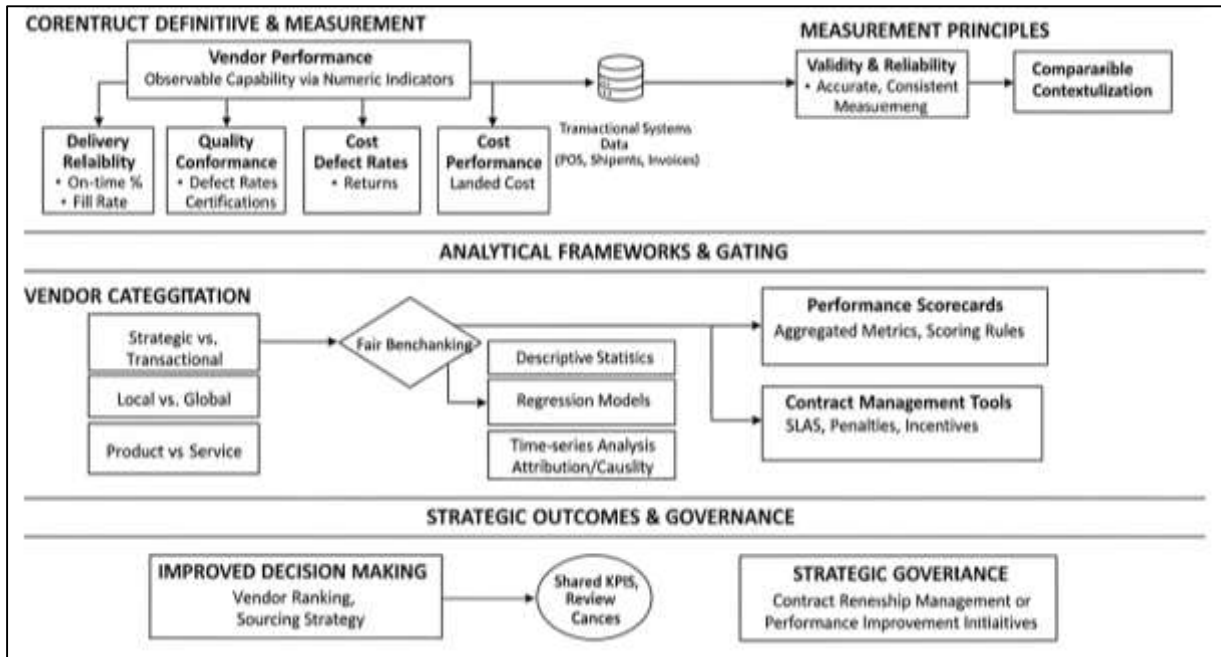
statistical approaches used to consolidate KPIs into scorecards, and the role of contract management tools in enforcing performance-based governance. The review also highlights methodological patterns in the literature, such as common data sources (ERP, POS, WMS, TMS, e-procurement logs), unit-of-analysis choices (vendor, SKU, lane, category, contract period), and model types (regression, DEA, AHP/TOPSIS, clustering, panel analysis). By organizing the literature into tightly defined quantitative themes, this section establishes a clear evidence base for the study's variables, hypotheses, and measurement logic, while also revealing areas of convergence and divergence in past findings about how vendor performance affects retail supply chain efficiency, service levels, and cost control.

Vendor Performance in Retail Supply Chains

Vendor performance in retail supply chains is widely framed in the literature as a measurable construct that captures how consistently vendors meet predefined operational and commercial expectations within procurement and logistics activities (Ashenbaum & Maltz, 2017). In quantitative research, the vendor is treated as the unit of analysis and performance is modeled as an observable capability represented through numeric indicators derived from transactional systems including purchase orders, shipment histories, delivery confirmations, inspection logs, and invoicing records. The underlying assumption in empirical retail studies is that vendor contributions are directly visible in day-to-day execution outcomes, making performance measurable through repeated observations rather than through impressionistic judgments. Scholars describe vendor performance as a multidimensional profile in which delivery, quality, cost, and compliance outcomes are monitored continuously because retail environments are sensitive to short lead times, high assortment variety, and frequent promotional cycles. In these contexts, even small vendor failures translate into shelf gaps, higher markdown exposure, and service disruptions (Anand & Grover, 2015). Quantitative studies emphasize that the construct is not limited to average achievement but includes consistency and variance over time. This is why many empirical designs incorporate dispersion statistics, deviation rates, and stability indicators alongside central performance measures. Another recurring point is that vendor performance is anchored in the retailer's measurement architecture: performance exists as a construct only when operational expectations are translated into explicit KPIs and embedded into data capture workflows. The literature also notes that performance evaluation at the vendor level supports benchmarking across a portfolio, enabling retailers to compare execution outcomes across vendors supplying similar categories. This comparability depends on standardized definitions, routine data audits, and consistent measurement windows (Domingues et al., 2015). By treating vendor performance as a numeric construct grounded in recurring transactions, quantitative retail supply chain research establishes a foundation for statistical modeling, composite scoring, and empirical testing of how upstream execution affects downstream retail outcomes.

The literature consistently separates vendor performance from adjacent constructs such as supplier relationship quality and procurement efficiency to protect conceptual clarity in quantitative studies. Vendor performance refers to the measurable extent to which vendors achieve operational targets in delivery, quality, cost, responsiveness, and compliance, using objective data captured from retail information systems (Gawankar et al., 2016). Supplier relationship quality, in contrast, is typically conceptualized as the relational climate between retailer and vendor, including trust, communication depth, commitment, and collaborative attitudes. Relationship quality is often measured through perceptual surveys or qualitative assessments rather than through transactional indicators. Empirical work shows that while strong relationships may support performance improvement, the two constructs are analytically distinct because one captures relational conditions and the other captures execution outcomes. Procurement efficiency is also treated separately and refers to the retailer's internal capability to source optimally, reflected in system-level cost reductions, negotiation effectiveness, process automation, and category-level risk balancing (Kataike et al., 2019). Studies highlight that procurement efficiency can coexist with uneven vendor performance since efficiency is an attribute of sourcing governance rather than of vendors' operational execution. Quantitative models therefore tend to position vendor performance as an outcome variable that can be influenced by relationship quality or procurement efficiency, but not substituted by them.

Figure 3: Analyzing Retail Vendor Performance Factors



This distinction is especially important in retail because vendor performance is directly tied to operational control and exception management, whereas relationship quality supports strategic collaboration and procurement efficiency reflects internal process strength. Researchers also caution that mixing these constructs can inflate correlations and weaken measurement validity, particularly when survey-based relationship indicators are incorrectly blended with operational KPIs (Kamble & Gunasekaran, 2020). By maintaining separation, the literature enables cleaner model specification, clearer variable selection, and more defensible conclusions about the specific drivers and consequences of vendor execution in retail supply chains.

Empirical retail supply chain studies display strong agreement on the main measurable dimensions used to evaluate vendor performance, reflecting repeated patterns in KPI selection and quantitative model design (Selviaridis & Norrman, 2015). Delivery reliability is one of the most dominant dimensions and captures whether vendors provide goods or services within promised time windows and in correct quantities. It is represented through indicators such as on-time delivery rates, lead-time adherence, order fill accuracy, short-shipment frequency, and variance between planned and actual arrival schedules. Quality conformance forms a second core dimension, focusing on whether delivered items meet specification standards and remain intact through transport and handling. Quantitative indicators here include defect rates per shipment, inspection failure ratios, damage incidence, claims frequency, and vendor-attributable return rates. The cost dimension addresses commercial discipline and margin protection, using measures such as price variance from contracts, total landed cost contribution, invoice error rates, rebate realization reliability, and cost-to-serve efficiency (Srivastava et al., 2015). Flexibility and responsiveness form an additional domain that captures a vendor’s capacity to accommodate demand variability common to retail promotions and seasonal shifts. Indicators often include rush-order acceptance rates, minimum order compliance, order change responsiveness, substitution success during shortages, and recovery speed after disruptions. Compliance and risk metrics are increasingly integrated within retail evaluation systems because of safety regulations, ethical sourcing requirements, and sustainability commitments. Numeric compliance measures may include audit pass rates, certification validity checks, documentation completeness, and recorded safety incidents. Collaboration and service execution are also quantified in retail contexts through response-time measures, data-sharing timeliness, accuracy of advance shipping notices, and corrective action closure speed (Oshri et al., 2019). The literature notes that these dimensions recur because they map directly to retail risk points: delivery failures create stockouts, quality failures generate reverse logistics costs, cost instability erodes margins, weak flexibility damages promotional performance, and

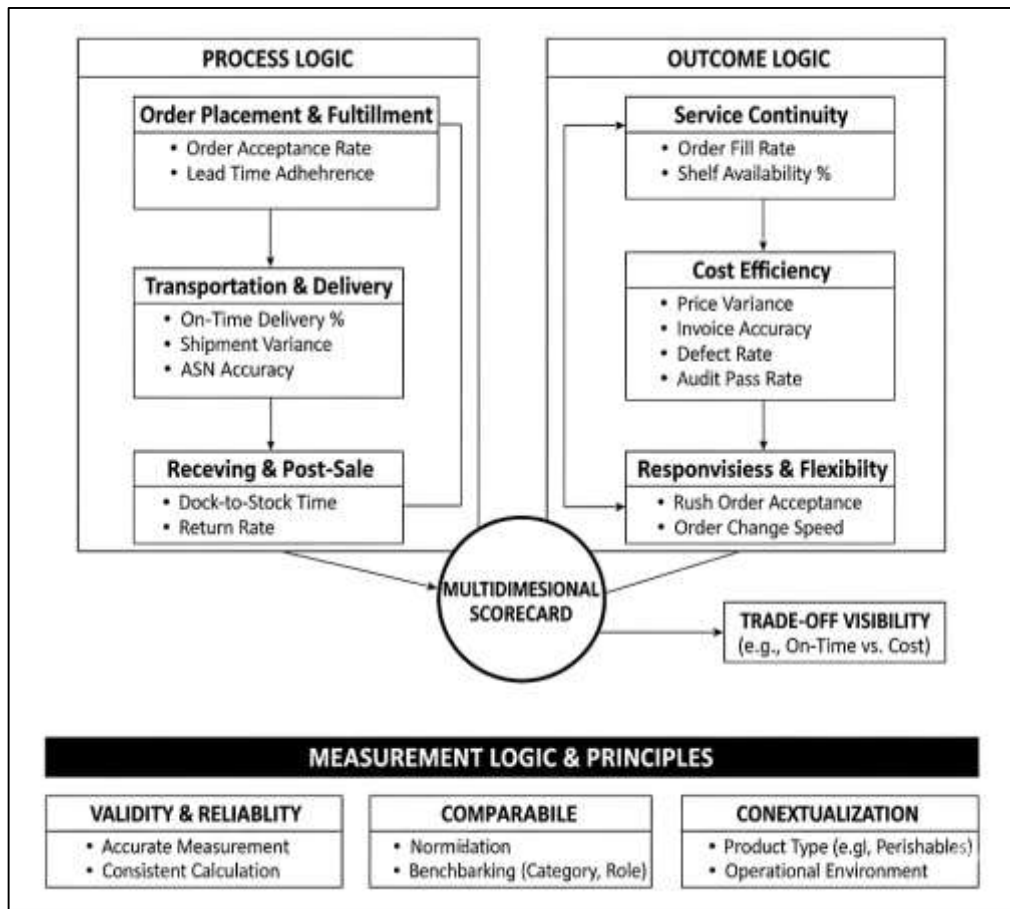
compliance failures threaten reputational and regulatory standing. Quantitative research therefore treats vendor performance as a composite of these interlocking dimensions, rather than a single indicator, to capture real trade-offs and produce robust vendor comparisons (Vallandingham et al., 2018).

Quantitative studies emphasize that vendor categorization is essential for fair benchmarking and accurate interpretation of vendor performance in retail supply chains. One widely used classification distinguishes strategic vendors from transactional vendors. Strategic vendors supply core categories, high-volume lines, or high-risk items, so they are evaluated through richer multi-dimensional scorecards and tighter thresholds (Singh & Teng, 2016; Reza et al., 2021). Transactional vendors provide low-risk, intermittent, or spot-buy items and are often assessed through a smaller KPI set emphasizing basic delivery and cost adherence. Another major categorization separates local and global vendors. Local suppliers typically operate with shorter lead times and lower transportation uncertainty, while global suppliers face longer logistics horizons, customs complexity, currency volatility, and higher disruption exposure. Because these structural differences affect measured outcomes, the literature stresses normalization and peer-group benchmarking so that global vendors are not penalized unfairly for distance-driven variance or regulatory delays (Genovese et al., 2014). Product vendors versus service vendors represent a further classification built into many retail datasets. Product vendors are evaluated primarily through physical-flow indicators such as delivery, fill rate, and quality conformance tied to tangible goods. Service vendors – such as logistics providers, packaging firms, and digital platform partners – are measured through service-level reliability, responsiveness, error resolution rates, and compliance performance. Empirical studies show that controlling for vendor category strengthens model precision by ensuring that performance scores reflect capability rather than role mismatch (Mortuza & Rauf, 2022). Vendor categorization also allows researchers to test whether performance determinants differ across types, such as whether flexibility is more central for fashion suppliers than for staple-goods vendors, or whether compliance risk is more predictive of governance actions for global than local partners (Salam & Khan, 2018). Retailers embed these categories into scorecard and contract systems to align evaluation expectations with vendor function and value contribution. The literature therefore treats vendor categories not as descriptive labels but as quantitative controls necessary for valid cross-vendor comparisons and rigorous empirical evaluation (Tse et al., 2019).

KPI Taxonomies for Retail Vendor

Empirical research on retail vendor evaluation shows a clear taxonomy of key performance indicator (KPI) families, designed to capture measurable vendor contributions to retail service continuity, inventory stability, and margin protection. In quantitative studies, KPIs are treated as observable variables that translate vendor behavior into numeric evidence (Tse et al., 2019). Retail settings demand KPI taxonomies that reflect the operational realities of high assortment breadth, rapid replenishment cycles, and frequent demand fluctuations tied to promotions and seasonality. Scholars emphasize that KPI families must be operationally interpretable, meaning that each indicator should map to a concrete retail risk point such as stockout exposure, overstocks, shrinkage, or reverse logistics burden. KPI taxonomies typically emerge from two complementary logics. The first is process logic, where indicators are aligned with supply chain activities like order placement, fulfillment, transportation, receiving, and post-sale returns. The second is outcome logic, where indicators are aligned with performance results that retailers prioritize, including shelf availability, customer satisfaction, cost efficiency, and compliance assurance (Zhou & Xu, 2018). The literature identifies that KPI families in retail vendor assessment are not selected in isolation but are combined into multidimensional dashboards and scorecards to preserve trade-off visibility. A vendor may deliver on time but at high cost, or maintain low defect rates while lacking promotional responsiveness, so measurement schemes must preserve these contrasts. Quantitative studies also show that KPI selection is shaped by product category characteristics. Perishable goods increase the priority of delivery timeliness and damage rates, apparel increases the priority of flexibility and substitution ability, and electronics increases the priority of quality conformance and claims monitoring. Across these differences, KPI taxonomies retain a stable core because the retail operating model consistently depends on vendor reliability to support store-level and omnichannel demand (Tjahjono et al., 2017).

Figure 4: KPI Taxonomy for Retail Vendor Evaluation



Another recurring idea is measurement logic, meaning the rules by which KPIs are calculated, normalized, and interpreted. Studies stress that KPI definitions must ensure validity, reliability, and comparability across vendors. Validity requires that a KPI captures the intended construct; reliability requires stable calculation rules across periods; and comparability requires scaling and benchmarking so vendors are judged relative to their role and environment. These principles underpin the KPI families reviewed in the next paragraphs and represent the empirical foundation for data-driven retail vendor evaluation (Marshall et al., 2015).

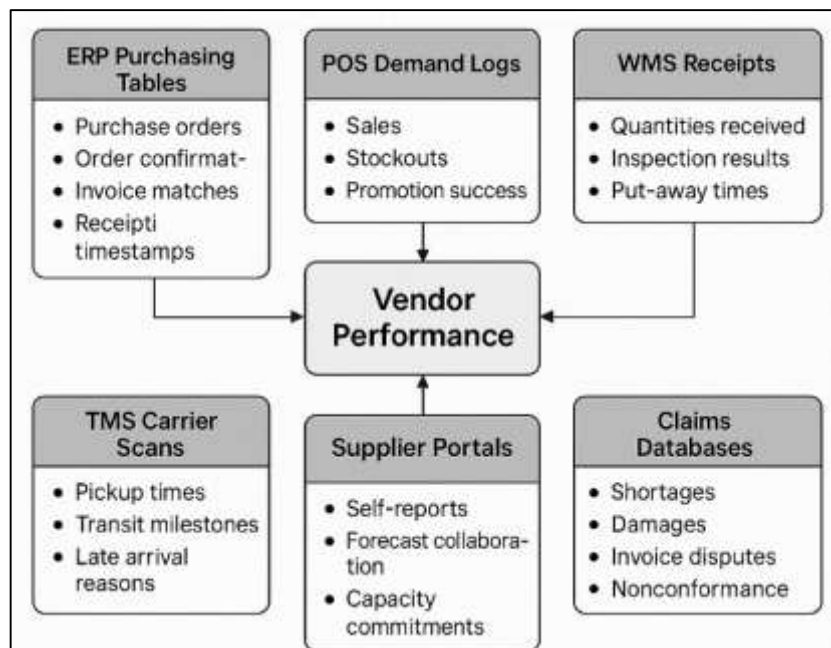
Delivery reliability KPIs form the most central indicator family in retail vendor evaluation because delivery failure directly translates into shelf gaps, lost sales, and unstable replenishment cycles. Empirical retail studies conceptualize delivery reliability as the vendor’s ability to deliver the correct goods, in the correct quantities, at the agreed location, within the agreed time window (Anand & Grover, 2015). On-time delivery rate is the anchor measure, calculated as the percentage of shipments arriving within predefined delivery windows, and widely used as a high-frequency operational signal. Lead-time deviation complements this by capturing the variance between promised and actual lead times across repeated orders, allowing studies to measure not only speed but consistency, which is crucial when retailers plan inventory using forecast-based reorder points. Dock-to-stock time is another retail-specific delivery KPI because it reflects how quickly inbound goods become sale-ready after arrival. Vendors influence this time through packaging integrity, labeling accuracy, pallet configuration, and documentation completeness, so the KPI captures upstream execution quality that affects downstream store or distribution-center performance (Dev et al., 2019). Advance shipping notice (ASN) accuracy is also prominent in empirical taxonomies, especially in omnichannel retailers, because accurate ASNs enable efficient slotting, labor planning, and cross-dock scheduling. ASN accuracy is treated as a data reliability measure, reflecting the vendor’s capability to provide timely and correct shipment information before arrival. Order fill rate measures completeness, representing the

percentage of ordered units actually delivered, and is critical in retail due to promotional demand spikes and assortment coupling effects, where missing items reduce category-level sales even if substitutes exist. Quantitative research highlights that these delivery KPIs are strongly interrelated but analytically distinct. On-time delivery focuses on punctuality, fill rate focuses on completeness, ASN accuracy focuses on data alignment, and dock-to-stock focuses on operational readiness (Dellino et al., 2018). Retail studies therefore treat delivery reliability as a multidimensional family rather than a single statistic so that vendors are not misclassified as reliable based on timing alone. Empirical designs also often control for category complexity and transportation distance because these factors alter delivery risk, reinforcing the need for benchmarking logic when interpreting delivery KPI distributions (Dweekat et al., 2017).

Data Quality in Vendor Performance Research

Quantitative vendor performance research in retail supply chains is grounded in multi-source operational datasets that capture vendor behavior across ordering, fulfillment, delivery, and post-delivery resolution cycles (Aliu et al., 2014). The most dominant data source family reported in empirical studies is enterprise resource planning (ERP) purchasing tables, which contain purchase orders, order confirmations, price schedules, invoice matches, and receipt timestamps. ERP data provide the backbone for measuring compliance with commercial terms, order fill accuracy, and lead-time consistency at the vendor level. Point-of-sale (POS) demand logs represent a second critical source because they record store- or channel-level sales and demand patterns, enabling researchers to relate vendor execution to downstream service outcomes such as stockouts, sales loss, and promotion success. Warehouse management system (WMS) receipts add another layer by documenting inbound deliveries at distribution centers or stores, including quantities received, inspection results, put-away times, and discrepancies between expected and actual shipments (Harrington et al., 2016).

Figure 5: Data Quality in Vendor Performance Research



These records are essential for calculating defect rates, damage incidence, dock-to-stock time, and receiving accuracy. Transportation management system (TMS) carrier scans provide event-level logistics visibility, capturing pickup times, transit milestones, late arrival reasons, and lane-specific delays that support high-resolution measurement of on-time delivery, lead-time deviation, and transport-related risk exposures. Supplier portals are also commonly used because they store advanced shipping notices, vendor self-reports, capacity commitments, forecast collaboration records, and corrective action documentation, allowing performance metrics to include information timeliness and collaboration reliability (Dreyer et al., 2018). Finally, claims databases – often integrated with quality

or finance modules—store vendor-attributable claims for shortages, damages, invoice disputes, and nonconformance, serving as both a quality signal and a governance indicator. Empirical retail studies emphasize that these sources are complementary rather than substitutable. ERP and WMS data reveal what was ordered and what was received, TMS data show the logistics pathway between those states, POS data reveal the demand environment and retail impact, supplier portals capture information behaviors, and claims databases record resolution and accountability outcomes (Gödri et al., 2019). Together they create a highly measurable vendor footprint that supports multi-dimensional KPI construction and robust statistical testing. The literature also notes that retail-specific complexity increases reliance on multi-source integration because omnichannel networks, private label programs, and promotional volatility make single-source measurement incomplete. Thus, quantitative vendor performance models in retail typically depend on linking these dominant datasets into unified analytical tables where vendor-level outcomes can be computed repeatedly over stable time windows (Volpi et al., 2017).

Retail vendor performance studies stress that data granularity shapes both the precision of KPIs and the interpretability of model results (Martín et al., 2018). The most granular level is SKU-level data, where each stock keeping unit becomes a measurable object tied to vendor-specific delivery and quality outcomes. SKU-level granularity allows researchers to capture fine differences in vendor capability across product lines, such as whether defects cluster in particular materials, sizes, or packaging types. Shipment-level granularity is slightly broader and is widely used for delivery reliability, damage incidence, and inspection outcomes, because shipments represent discrete vendor fulfillment events with clear timestamps and quantities. Store-level granularity is important in retail because vendors may perform differently across store formats or geographies; a vendor's delivery reliability to urban outlets may differ systematically from performance to rural or remote stores. Lane-level granularity is another recurring unit, especially in omnichannel logistics, where a lane represents a specific origin-destination route (Pungchompoo & Sopadang, 2015). Lane-level data allow performance to be decomposed into vendor execution versus transport or regional infrastructure effects, improving fairness in benchmarking. Monthly contract-level granularity is common in governance-focused studies because vendor scorecards and service-level agreement reviews often operate on monthly cycles. At this level, KPIs are aggregated to align with contract penalties, incentive triggers, and renewal decisions. The literature explains that quantitative research often uses multiple levels simultaneously, producing hierarchical datasets. For example, SKU-level or shipment-level KPIs may be nested within vendor-lane combinations, which are nested within monthly contract periods. This structure enables multi-level modeling that separates micro execution variance from macro relationship effects. Empirical studies also show that granularity decisions depend on the research goal (Davé et al., 2017). If the purpose is operational diagnosis, SKU and shipment levels are favored. If the goal is portfolio ranking or governance policy testing, monthly contract levels are more appropriate. Retail environments further complicate granularity because promotions can create temporary spikes that distort performance metrics if aggregation is too coarse. Therefore, many studies incorporate time segmentation by week or promotion window even when their main models operate at monthly levels. In sum, data granularity in vendor performance research functions as a methodological control that determines how finely vendor behavior is observed, how variability is captured, and how accurately performance differences can be attributed to vendors rather than to category or channel context (Luzzini et al., 2014).

Because retail vendor datasets originate from multiple operational systems, quantitative research dedicates significant attention to data cleaning and harmonization before KPI construction. A central issue is missing values, which may occur due to unscanned receipts, delayed event uploads, or incomplete supplier portal submissions (Ross & Cheah, 2019). Empirical studies handle missingness using rule-based deletion when gaps are rare, statistical imputation when patterns are systematic, or sensitivity checks that compare results with and without imputed values. Another recurring problem is inconsistent vendor IDs across systems, especially when ERP, WMS, and POS platforms use different coding standards or when vendors operate through subsidiaries. Studies address this by building crosswalk tables that match multiple vendor identifiers to a single master ID, often verified through address, tax code, or contract references. Outliers are treated as both noise and signal in retail contexts.

Extremely late shipments, unusually high defect rates, or sudden cost deviations can represent data entry errors, but they can also represent true performance shocks (Ding et al., 2020). Quantitative research therefore applies outlier detection through interquartile-range rules, z-score thresholds, or model-based anomaly detection, then performs domain checks to classify outliers as errors or legitimate events. Time misalignment is another major issue, since ERP order timestamps, TMS scan timestamps, and WMS receipt timestamps may reflect different clocks, batch uploads, or timezone settings. To handle this, studies align events using standardized time zones, adjust for system batch delays, and define KPI windows that use consistent start-end logic. A related issue is duplicate records caused by partial updates or repeated scans, which can inflate performance failures if not removed. Researchers typically de-duplicate using unique order-shipment keys and by verifying event sequences (T. Li et al., 2020). Empirical retail papers also highlight the need to separate vendor-attributable delays from retailer-attributable effects. For example, if a retailer changes an order after confirmation, lead-time deviations may not represent vendor failure. Quantitative designs therefore exclude post-confirmation change orders from reliability calculations or control for such revisions as covariates. These preprocessing steps are essential because KPI construction depends on accurate counts, timestamps, and linkage across systems. Without cleaning, composite scorecards may penalize vendors unfairly or mask real capability differences. Thus, data handling procedures are treated not as technical preliminaries but as integral parts of measurement validity in retail vendor performance research (Y. Li et al., 2020).

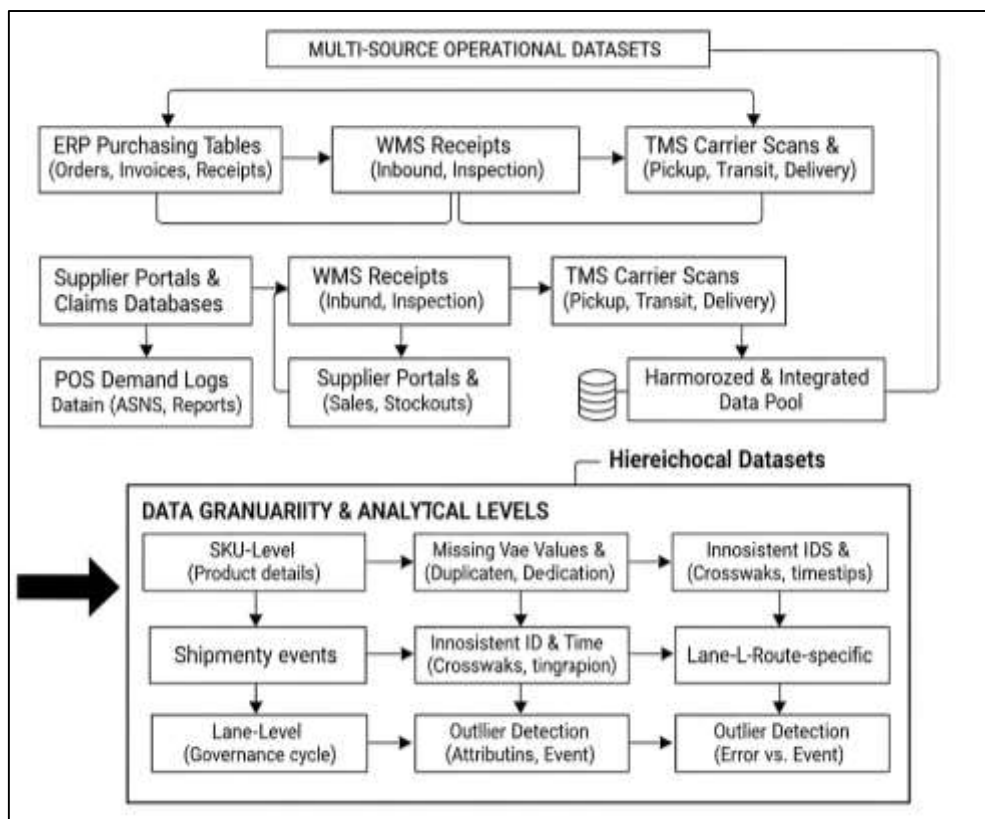
Multi-Criteria Scorecard Construction in Retail Vendor Performance

Multi-criteria scorecards in retail vendor performance evaluation are described in the literature as composite indices that integrate multiple key performance indicators into a single, interpretable performance profile. In retail supply chains, vendors influence service continuity through several interdependent dimensions, including delivery reliability, quality conformance, commercial discipline, flexibility under demand shocks, and compliance with operational rules (Tian & Sarkis, 2020). Empirical studies therefore treat scorecards as an aggregation device that converts multidimensional vendor evidence into a structured numeric summary. The scorecard is not a mere reporting template; it functions as a quantitative model of vendor contribution because it combines indicators that differ in scale, volatility, and managerial priority into a coherent scoring architecture. Retail contexts motivate this approach due to large vendor portfolios and high transaction frequency, which create measurement overload when KPIs are viewed separately (Kirby et al., 2017). A scorecard reduces this complexity by expressing vendor performance as a comparative index relative to targets or peer distributions. Literature also frames scorecards as tools of cross-functional governance, meaning that they standardize how procurement, logistics, finance, and category teams interpret vendor outcomes. The composite nature of scorecards allows retailers to observe trade-offs in a controlled way. For example, a vendor may exhibit strong delivery punctuality but weak quality outcomes or show low costs with poor flexibility, and scorecard logic ensures these patterns are consolidated without erasing dimensional meaning (Silva et al., 2018). Quantitative research highlights that composite indices are constructed under explicit measurement rules that preserve transparency, reproducibility, and fairness. These rules define data windows, KPI calculation formulas, normalization requirements, and aggregation pathways so that a vendor's score represents systematic evidence rather than episodic impressions. In retail studies, scorecards also enable longitudinal tracking because composite scores can be computed across repeated monthly or quarterly cycles, allowing performance stability and variance to be monitored. Thus, multi-criteria scorecards represent a critical quantitative mechanism that transforms fragmented KPI streams into a unified index suitable for benchmarking, ranking, and empirical modeling of vendor performance in retail supply chains (Zhu & Freeman, 2019).

The literature identifies several widely applied numeric aggregation methods that convert raw KPIs into composite vendor scores in retail settings. Target-based scoring is among the most common approaches because retail operations are governed through explicit service targets embedded in purchase agreements (Álvarez et al., 2015). Under target-based methods, each KPI is mapped to a scoring scale where performance meeting or exceeding a target earns full points and shortfalls reduce points according to predefined penalty gradients. This method aligns naturally with delivery windows, defect tolerances, and invoice accuracy requirements. Percentile ranking is another prevalent method,

especially for KPI families where absolute targets are difficult to set across categories. In percentile approaches, vendors are ranked relative to peers and assigned scores based on their position in the distribution, allowing performance to be interpreted as comparative advantage rather than absolute compliance. Standardization methods such as z-score transformation and min-max normalization are used to place KPIs on a consistent numeric scale before aggregation (Yuan et al., 2020). Z-scores center each KPI on the mean and scale by variability, ensuring that indicators with larger numeric ranges do not dominate the composite score. Min-max normalization rescales KPI values to a bounded interval, enabling intuitive comparisons across heterogeneous indicators such as time, cost, and defect counts. Weighted sums form the final aggregation layer in many empirical scorecards. Once KPIs are scored and standardized, they are multiplied by assigned weights and summed to create a composite index (Støre-Valen & Lohne, 2016).

Figure 6: Quantitative Vendor Performance Research



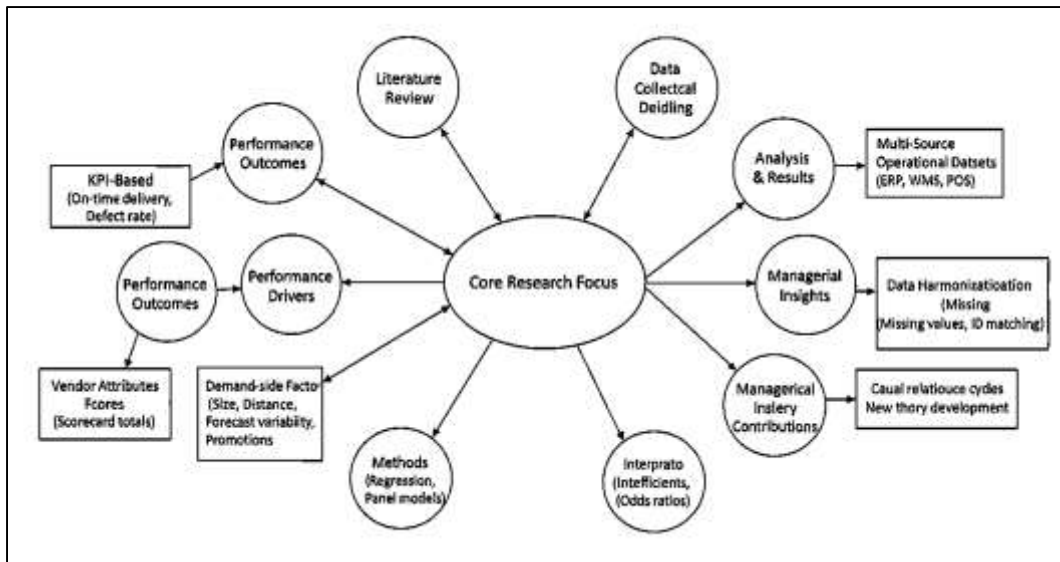
Literature emphasizes that the choice of aggregation method influences score behavior, including sensitivity to outliers and the extent to which extreme failures distort totals. Retail studies often compare alternative aggregation methods to test robustness, because vendors may shift rank positions when scoring logic changes. Another recurring observation is that aggregation methods are rarely used in isolation. Target-based scoring may be combined with normalization, and percentile ranking may be applied within vendor categories before weighted summation. This layered construction allows scorecards to retain both managerial interpretability and statistical comparability. Overall, empirical research presents numeric aggregation as the technical core of retail vendor scorecards, turning heterogeneous KPIs into stable composite indices that support systematic performance evaluation (De Felice et al., 2015).

Modeling of Vendor Performance Drivers in Retail Supply Chains

Quantitative retail supply chain literature increasingly treats vendor performance as an outcome that can be explained through inferential modeling rather than simple descriptive score comparisons (El Gibari et al., 2019). Studies argue that vendor KPIs and composite scorecard totals exhibit systematic variation across vendors, categories, and operating environments, which makes them suitable as dependent variables in statistical models. In this framing, vendor performance is not assumed to be

random or purely capability-based; instead, it is modeled as the result of identifiable drivers that shape how vendors execute delivery, quality, cost reliability, and responsiveness in retail settings. Inferential studies typically start by defining a performance construct from one or more KPI outcomes, such as on-time delivery rate, defect level, or total scorecard value, then estimating how vendor or transaction attributes influence this construct (Buede et al., 2018).

Figure 7: Quantitative Vendor Performance Insights



A consistent theme is that retail supply chains provide high-frequency, multi-period data that allow performance to be observed repeatedly, which strengthens statistical power and enables more rigorous causal reasoning. Many studies emphasize that vendor performance drivers in retail differ from manufacturing contexts because retailers face promotional spikes, assortment churn, and omnichannel complexity. Therefore, models often incorporate demand- and category-specific controls to isolate vendor effects from retail operational noise. Another feature noted across the literature is the use of benchmarking baselines in model specification. Researchers do not just compare vendors to a single target; they estimate relative performance by controlling for expected service levels tied to product type, lane distance, or order profiles (Chaudhuri et al., 2016). Models thus attempt to separate structural challenges from true vendor underperformance. Empirical work also highlights the importance of measurement windows, showing that monthly, quarterly, or promotion-cycle panels can yield different driver effects because vendors may behave differently under normal replenishment versus peak demand conditions. Inferential approaches allow these differences to be tested statistically. The literature therefore frames statistical modeling as a way to move vendor evaluation from a diagnostic tool to an explanatory system, where managers and researchers can identify the conditions under which vendors perform better or worse, and determine which drivers consistently explain high or low performance across retail networks (Rezaei et al., 2018).

Across empirical studies, a stable set of independent variables appears repeatedly as determinants of vendor performance in retail supply chains. Vendor size is commonly operationalized through annual sales volume, production capacity, workforce scale, or asset base, and is treated as a proxy for resource depth and process maturity (Almohri et al., 2019). Distance, usually measured as geographic kilometers or transport time between origin and retail distribution points, captures exposure to logistics uncertainty and infrastructure variability. Order volume is another frequent driver, reflecting the scale and frequency of transactions between retailer and vendor; large, stable volumes often support learning effects and priority treatment, while small or irregular volumes may increase service variability. Product complexity is modeled through SKU count per order, customization level, packaging requirements, or handling constraints, capturing the operational burden imposed on vendors (Silvi et al., 2015). Forecast variability enters models as a demand-side driver, often represented by historical

demand variance, forecast error rate, or promotion intensity; this variable tests whether vendors struggle when retailer demand signals fluctuate sharply (Vlachos, 2014). Contract type is also widely included, distinguishing between long-term strategic agreements, short-term spot contracts, penalty-heavy SLAs, or incentive-linked structures, because contract design shapes vendor motivation and risk sharing. Some studies complement these core variables with vendor tenure, technology integration level, category criticality, or lane congestion indicators, but the typical driver set remains consistent in retail literature. On the dependent side, studies use either individual KPI outcomes or composite scorecard totals. KPI-based dependent variables include on-time delivery percentage, lead-time deviation, defect ratio, claims rate, invoice error rate, or rush-order success. Composite dependent variables collapse multiple KPIs into a single performance index, enabling models to test broader capability effects rather than narrow outcome effects. Literature stresses that the choice between KPI and scorecard dependent variables affects interpretation: KPI models reveal domain-specific driver impacts, while scorecard models reveal overall vendor contribution patterns. Another recurring observation is that dependent variables are often bounded (percentages, rates) or skewed (defect counts), which influences model choice and transformations (Wagner & Benoit, 2015). Retail studies therefore emphasize careful dependent-variable preparation, including normalization, winsorizing, and role-based benchmarking before estimation. By aligning these independent and dependent variables, inferential models create an empirical bridge between vendor attributes, transaction conditions, and measurable retail performance outcomes.

Empirical retail supply chain studies apply a range of statistical techniques to estimate vendor performance drivers, choosing methods based on data structure and performance measure properties. Multiple regression remains the most widely used approach when dependent variables are continuous KPIs or composite scores (Hair et al., 2019). Regression models allow researchers to estimate marginal effects of vendor size, distance, order volume, or forecast variability on performance outcomes, while controlling for category, season, and channel differences. Logistic regression is commonly used when performance is expressed as a binary event such as pass/fail compliance, on-time versus late delivery, or SLA breach versus non-breach; this technique is particularly relevant in retail governance because many contractual triggers are categorical. Panel fixed-effects models are frequently employed due to repeated observations of vendor performance over time. Fixed effects control for unobserved, time-invariant vendor characteristics such as culture, baseline capability, or managerial style, enabling cleaner estimates of how changing drivers like order volume or forecast variability influence performance. Random-effects models also appear when researchers assume vendor-specific effects are distributed randomly and uncorrelated with explanatory variables, allowing inclusion of time-invariant drivers like distance (Punia et al., 2020). Comparative studies often test fixed versus random specifications to justify the chosen estimator. Structural equation modeling (SEM) is used in quantitative supply chain work when researchers treat vendor performance as a latent construct influenced by multiple observed drivers and when mediating effects are hypothesized, such as contract type influencing collaboration quality, which in turn affects delivery reliability. SEM allows the simultaneous estimation of measurement and structural components, which is useful when performance is multidimensional and not fully captured by a single KPI. Additional techniques sometimes appear, such as hierarchical linear modeling for nested data (SKU within vendor within lane), or quantile regression when drivers are expected to affect high-risk underperformance tails rather than average performance (Nikhashemi et al., 2019). Across methods, the literature emphasizes robustness checks, including multicollinearity diagnostics, heteroskedasticity correction, and alternative model comparisons, because retail datasets are large and heterogeneous. The methodological landscape therefore reflects an effort to match inferential tools to retail-specific data realities, ensuring that estimated driver impacts are statistically credible and operationally meaningful. A central contribution of statistical modeling in the retail vendor literature is the interpretation of coefficients to identify high-impact performance drivers and to separate structural effects from controllable vendor behaviors (Nikhashemi et al., 2019). In regression-based studies, coefficients represent the expected change in a KPI or composite score associated with a one-unit change in an explanatory variable, holding other factors constant. Researchers commonly interpret positive coefficients as performance enhancers and negative coefficients as risk amplifiers, with particular

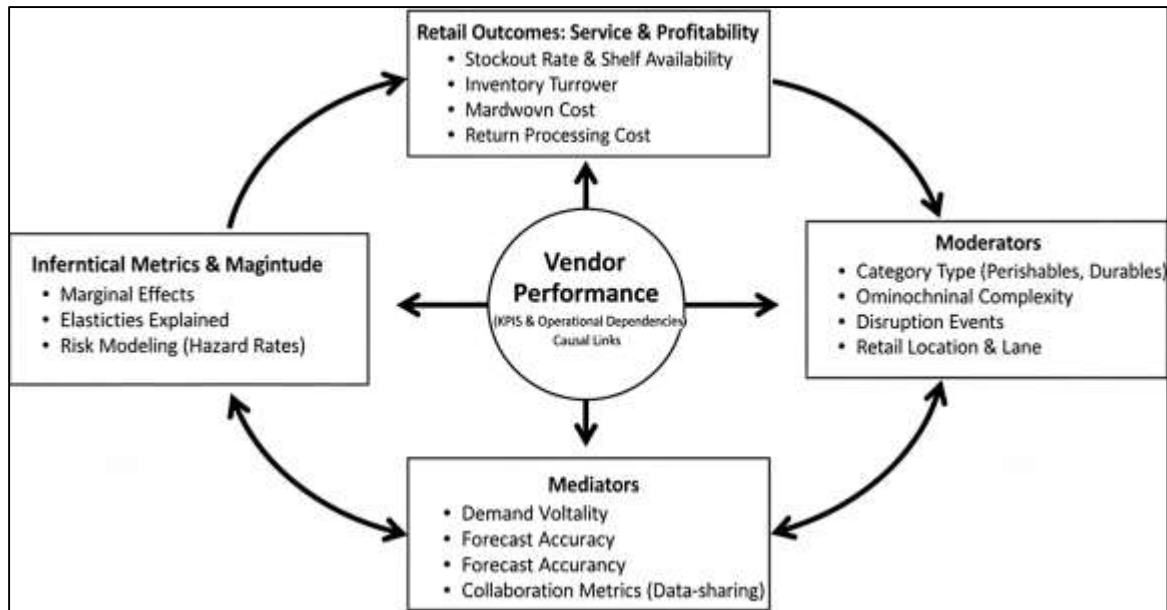
attention to magnitude and statistical significance. For example, a large negative coefficient on distance in on-time delivery models is interpreted as evidence that longer lanes materially increase lateness risk, while a strong positive coefficient on vendor size in quality models may be read as resource advantages improving defect control. Logistic models are interpreted through odds ratios, indicating how drivers increase or decrease the probability of passing service thresholds or avoiding SLA breaches (Vahdani et al., 2017). Panel coefficients are interpreted as within-vendor effects, meaning the estimated relationship reflects how changes in order volume or forecast variability alter performance for the same vendor across time, strengthening causal plausibility. SEM results are interpreted by comparing direct and indirect paths, allowing researchers to identify whether drivers act primarily through operational capability, relational coordination, or contractual incentives. Literature also emphasizes that coefficient interpretation must consider retail context. Forecast variability may have a modest average coefficient on delivery performance but a larger effect in promotional periods, so studies often interpret coefficients alongside subgroup or interaction results (Dalvi & Kant, 2018). Another interpretive theme is elasticity thinking: researchers contextualize coefficients relative to baseline performance levels to infer managerial relevance, such as whether a change in contract type yields a meaningful improvement in fill rate compared with natural seasonal fluctuation. Many studies rank drivers by standardized coefficient size or by explained variance contribution to determine which factors are most influential. This ranking supports evidence-based prioritization in vendor management. The literature also cautions that coefficients can reflect both capability and selection effects, since retailers may allocate more volume to better vendors, so interpretation often comes with model controls or lag structures that reduce simultaneity bias (Yang & Zhang, 2017). Through these interpretive practices, inferential research turns vendor evaluation into an explanatory system, highlighting which measurable vendor and transaction factors most strongly predict performance variation in retail supply chains.

Vendor Performance and Retail Supply Chain Outcomes

Empirical retail supply chain research consistently links vendor performance scores to a set of downstream operational outcomes that determine service continuity and profitability. Studies that evaluate vendor delivery, quality, cost, and flexibility metrics through composite scorecards report that higher vendor scores are associated with lower stockout rates and stronger shelf availability levels at the store and fulfillment-center interface (Romule et al., 2020). The logic developed in the literature is that vendors with reliable delivery timing and high order fill rates reduce replenishment gaps, allowing retailers to match inventory positioning more closely to observed demand. In grocery and fast-moving consumer goods, empirical models show that delivery reliability and defect control are especially predictive of reductions in stockout frequency because short shelf-life windows amplify the impact of any missed shipment. Research in apparel and consumer electronics also shows that vendors with better responsiveness indicators enable faster replenishment of high-velocity items, improving shelf availability during promotional windows and minimizing missed sales opportunities. Inventory turnover, another central outcome, is similarly tied to vendor performance (Sureeyatanapas et al., 2018). Vendors who deliver consistently in smaller, accurate lots help retailers maintain leaner inventory while sustaining service, which raises turnover ratios. Conversely, vendors with volatile lead times or frequent quality issues cause retailers to hold safety stocks or buffer inventories, suppressing turnover. Markdown cost percentage is often included as a profitability outcome because late deliveries or incomplete orders push goods into low-demand periods, forcing price reductions. Empirical studies demonstrate that vendors with weak timeliness or low flexibility create excess inventory and late-season arrivals, which increase markdown exposure in fashion, seasonal goods, and consumer durables. Return processing cost is also connected to vendor performance, especially through quality KPIs such as defects, damage incidence, and vendor-attributable claims (Benton Jr et al., 2020). Vendors with higher quality and packaging discipline reduce customer returns and internal reverse logistics burden, lowering processing costs per unit. Finally, customer service level, often measured through order fulfillment rates, customer satisfaction indices, or service-time compliance, is shown to improve when vendors maintain stable performance across delivery and quality dimensions. Overall, the empirical evidence positions vendor scorecards as upstream predictors of retail operational stability, and studies repeatedly emphasize that vendor performance is embedded in retail outcomes through a

measurable and statistically detectable chain of operational dependencies (Cheung et al., 2020).

Figure 8: Vendor Performance Management in Retail



Quantitative studies do more than establish associations; they estimate the magnitude of vendor-performance effects on retail outcomes using several inferential metrics. A common approach is marginal-effect estimation, where regression models compute how a one-unit change in a vendor KPI or composite score shifts an outcome such as stockout rate or shelf availability, controlling for demand level, seasonality, and category structure (Fallahpour et al., 2017). These studies often interpret marginal effects in operational terms, such as the reduction in stockout probability attached to improved on-time delivery or the gain in shelf availability attached to higher order fill accuracy. Another recurring quantification method is elasticity-style interpretation, where researchers express the percentage change in retail outcomes relative to percentage changes in vendor performance indicators. Elasticities are used particularly in multi-category retail datasets to compare effect strength across domains like grocery versus apparel or staples versus discretionary goods (Yang et al., 2019). Variance-explained measures are also prominent. Studies report the proportion of outcome variability accounted for by vendor performance after including baseline demand, store format, and logistics controls. In many cases, delivery reliability explains a meaningful share of stockout variability, while quality indicators explain a substantial portion of return-cost differences. Composite scorecards are found to explain more variance in multidimensional outcomes such as customer service level because they integrate reliability, quality, and commercial discipline. Empirical research also uses model-comparison strategies to benchmark explanatory contribution. For example, models with vendor scores are compared against models without vendor measures to determine incremental explanatory value. In some retail papers, vendor-performance variables dominate over internal retailer-process variables when predicting service disruptions, demonstrating the sensitivity of retail networks to upstream execution. Quantification also extends to risk modeling (Lo et al., 2018). Some studies estimate hazard rates for stockout events or decomposition of markdown costs into vendor-driven and demand-driven components. These approaches allow researchers to show not only direction but intensity of vendor influence. Across the literature, the central message is that vendor scores offer statistically measurable leverage over operational metrics that retail firms care about, and the contribution is observable as marginal improvements, elasticities, and explained variance that remain stable after accounting for contextual controls (Mungra & Yadav, 2020).

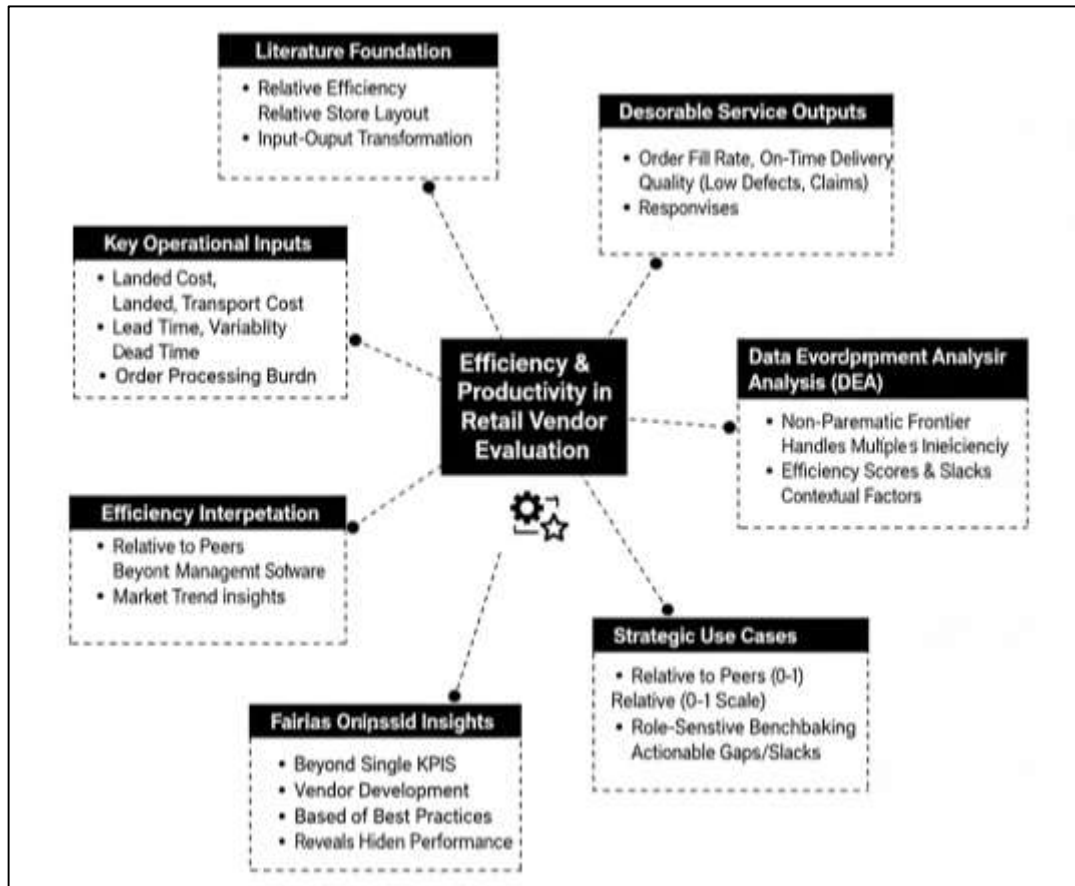
Empirical research emphasizes that the relationship between vendor performance and retail outcomes is conditioned by mediators and moderators that reflect retail heterogeneity. Category type is one of the most frequently tested moderators (Liu et al., 2017). In perishable categories, delivery timeliness

and quality conformance show stronger links to stockout reduction and shelf availability because inventory buffers are limited by spoilage risk. In durable or slow-moving categories, cost stability and minimum-order compliance sometimes show stronger connections to inventory turnover and markdown outcomes. Omnichannel complexity is another key moderator. Studies distinguish between single-channel store replenishment networks and integrated omnichannel systems where vendors supply both stores and e-commerce fulfillment nodes. In high-omnichannel settings, responsiveness and ASN accuracy become stronger predictors of customer service level, because real-time inventory visibility and rapid order reallocation depend on vendor information reliability. Disruption events also appear as moderators across retail studies (Chaudhry et al., 2018). Research examining periods of transport strikes, port congestion, pandemic shocks, or major promotional surges shows that vendor flexibility and recovery-time indicators become the dominant drivers of retail outcomes in unstable environments. Some studies treat demand volatility as a mediator. Vendors with poor flexibility raise retailer forecast errors and force emergency replenishment actions, which then increase stockouts or markdown costs. Other studies test collaboration metrics as mediators, showing that data-sharing and corrective-action responsiveness improve replenishment accuracy, which in turn improves shelf availability. Retail location and lane characteristics are also modeled as moderators in cross-country datasets, where distance and infrastructure reliability influence the strength of delivery-performance effects (Sheng et al., 2018). These mediating and moderating frameworks demonstrate that vendor performance does not translate into outcomes uniformly. Instead, retail outcomes respond differently depending on product perishability, demand volatility, channel structure, and disruption exposure. The literature uses these interaction structures to sharpen interpretation, revealing why some vendors appear more critical to service stability in certain categories or channels than in others. This conditional logic is central for quantitative vendor studies because it prevents overgeneralization and ensures that estimated effects are tied to the operational reality of specific retail contexts (Agarwal & Narayana, 2020).

Approaches to Vendor Evaluation

Efficiency and productivity approaches to vendor evaluation in retail supply chains emerge from the idea that vendors should be compared not only by isolated KPIs but also by how well they transform inputs into desirable service outputs (Bryan Jean et al., 2017). In this literature stream, vendor performance is framed as a relative efficiency construct rather than a simple compliance score. Efficiency-frontier studies argue that vendors operate with different resource profiles and environmental constraints, so the fair way to evaluate them is to compare each vendor's input-output conversion against the best observed performers in the same portfolio. Retail supply chains provide a strong setting for this logic because vendors differ widely in shipment scale, distance, product variety, and category risk, yet they all contribute to shared operational outcomes such as shelf availability, order cycle reliability, and return reduction (Abu Salim et al., 2019). Efficiency methods therefore treat vendors as decision-making units whose productivity can be inferred from transactional evidence. Empirical research explains that frontier approaches complement scorecards by allowing multiple inputs and multiple outputs to be assessed simultaneously without collapsing performance into pre-weighted indices at the start. This is especially useful when retailers want to avoid subjective weighting or when KPI trade-offs are complex. For example, a vendor may incur higher landed costs but deliver extremely consistent fill rates and low defects, which a frontier model can recognize as high efficiency if the output gains justify the inputs compared to peers (Dong et al., 2017). Efficiency studies in retail also highlight that productivity evaluation can reveal hidden overperformance or underperformance that standard target-based scoring misses, because frontier methods benchmark vendors against actual best practices rather than against fixed managerial targets. Another reason efficiency models appear in retail vendor research is data richness: ERP, WMS, TMS, and claims systems generate multi-period inputs and outputs that can be aligned into consistent vendor-level panels, enabling robust relative efficiency estimation (Wang & Wang, 2019). As a result, the literature positions efficiency-frontier methods as a quantitative pathway for distinguishing capable vendors from those whose outcomes are weak given their resource usage or structural context, thereby enhancing fairness and diagnostic depth in retail vendor evaluation frameworks.

Figure 9: Retail Vendor Efficiency Framework



Data Envelopment Analysis (DEA) is the most frequently used efficiency-frontier method in vendor evaluation research and is widely applied in retail supply chains due to its flexibility in handling multiple inputs and outputs (Abu Salim et al., 2019). Retail DEA studies typically define inputs as the resource burdens or operational costs associated with sourcing from a vendor, while outputs represent service or quality achievements that retailers value. Common inputs in retail DEA models include total landed cost, transportation cost contribution, lead time length, lead-time variability, order processing burden, and sometimes defect-handling effort. Outputs often include order fill rate, on-time delivery achievement, inspection pass rate, low defect incidence, low claims frequency, and measures of responsiveness during peak demand windows. The literature emphasizes that input and output selection must match retail operational meaning (Huang et al., 2019). Cost and lead time are treated as inputs because they represent what the retailer “consumes” to obtain supply, while fill rate and quality are treated as outputs because they represent what the retailer “receives” as value. Empirical studies also note that DEA can be structured as input-oriented or output-oriented depending on retailer governance priorities. Input-oriented DEA asks whether a vendor could achieve the same service outcomes using fewer inputs, which fits retailers seeking cost discipline. Output-oriented DEA asks whether a vendor could produce higher service outcomes with the same input profile, aligning with retailers prioritizing shelf availability and customer service (Alvarez-Rodriguez et al., 2019). Retail DEA research often uses variable returns to scale to account for the fact that vendor size influences productivity, so small vendors are not automatically penalized against large-volume suppliers. Another recurring practice is peer-group DEA, where vendors are grouped by category type, global versus local lane exposure, or strategic versus transactional role before efficiency estimation; this improves comparability by ensuring the frontier is constructed from relevant peers. Studies also highlight that DEA produces not only efficiency scores but also reference sets, meaning each inefficient vendor is linked to one or more efficient vendors whose input-output mixes form the benchmark. This enables retailers to identify realistic improvement pathways, such as reducing lead-time variability or improving inspection pass rates without necessarily increasing cost (Álvarez-Rodríguez et al., 2020). In

retail contexts, DEA therefore functions as both a numeric ranking tool and a diagnostic map of how vendors differ in productivity relative to best-practice frontiers.

Stochastic Frontier Analysis (SFA) appears in the literature as a complementary alternative to DEA, offering a parametric way to estimate vendor efficiency while separating random noise from true inefficiency. Retail vendor performance data are often affected by external shocks—weather delays, port congestion, demand surges, or retailer-side receiving bottlenecks—that can distort observed KPIs. DEA treats all deviations from the frontier as inefficiency, which may over-penalize vendors when noise is substantial (Gandhi & Shankar, 2016). SFA addresses this by specifying a production or cost frontier with a statistical error structure that decomposes deviations into two parts: a random error term capturing uncontrollable noise and an inefficiency term capturing vendor-attributable shortfall. In retail vendor studies, SFA is frequently used when researchers want to estimate productivity under uncertainty and when data span many periods, allowing stable parameter estimation. Input variables in SFA models mirror those used in DEA, including landed cost, logistics time, order complexity, and sometimes capacity commitments, while outputs include fill rate quality, on-time delivery, or composite service indices (Kahraman et al., 2018). The advantage stressed in empirical work is that SFA allows hypothesis testing about the shape of the frontier and the significance of drivers, enabling researchers to examine how vendor size, distance, or contract type influences inefficiency levels. Retail SFA studies often model heteroskedasticity because variance in vendor outcomes differs across categories, lanes, and seasonal windows. Another feature is that SFA can incorporate contextual controls directly into the inefficiency term, allowing the evaluation to account for factors such as international lane risk or promotion intensity without constructing separate peer groups. Literature also notes that SFA outputs are interpretable in managerial terms, providing efficiency scores similar to DEA but grounded in probabilistic confidence, making them useful when retailers need to justify governance actions under data uncertainty (Mayberry et al., 2017). Thus, SFA is presented as a productivity lens that is especially suitable for retail environments where performance is exposed to exogenous volatility and where separating noise from inefficiency improves fairness in vendor evaluation.

The interpretation of efficiency scores is a central theme in the frontier-based vendor evaluation literature because these scores guide benchmarking, ranking, and improvement decisions in retail supply chains. In both DEA and SFA, an efficiency score is understood as a relative measure of how close a vendor operates to the best-practice frontier observed in the vendor portfolio (Tran et al., 2019). A score at the frontier indicates that a vendor achieves a strong combination of outputs for its input burden compared to peers. A lower score indicates that, given the same cost and time inputs, the vendor produces weaker service outputs than the most productive vendors. Retail studies stress that efficiency interpretation must be role-sensitive. Strategic vendors supplying high-risk categories may be benchmarked separately so that their frontier reflects the operational demands of those categories, while transactional vendors are compared against a simpler frontier aligned with basic service expectations (Zheng et al., 2020). Benchmarking logic also involves understanding slack values or inefficiency drivers. DEA studies compute input slacks showing where vendors over-consume resources, such as excessive lead-time variability, and output slacks showing where vendors under-deliver value, such as fill-rate gaps or quality shortfalls. Retailers interpret these slacks as actionable improvement priorities rather than abstract statistics. SFA studies interpret inefficiency components as probabilities of underperformance, often ranking vendors by expected inefficiency while acknowledging confidence intervals. Another interpretive point in the literature is that efficiency scores should not be read as absolute “good or bad” labels but as comparative productivity positions inside a defined peer set and time window. Vendors may appear efficient in stable demand months but inefficient during promotional shocks, so some studies compute rolling efficiency frontiers to observe stability. Retail governance uses efficiency scores to support volume allocation, vendor development programs, or renegotiation focus, because efficiency captures the practical trade-off between what a retailer invests in sourcing and what it receives in service outcomes (Er Kara & Oktay Firat, 2018). The literature therefore frames frontier scores as an evidence layer that enhances fairness: vendors are judged relative to the best achievable mix in their context rather than against one-size-fits-all targets, and improvement is guided by measurable gaps to peer frontiers.

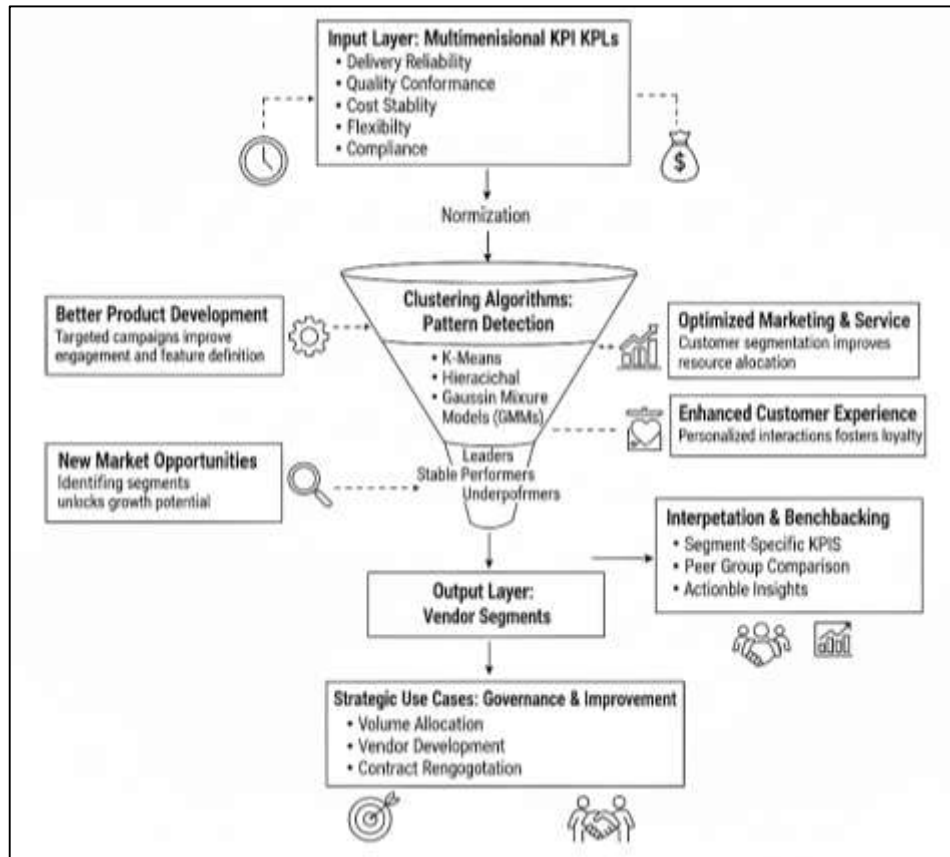
Models for Vendor Portfolio Evaluation

Cluster and segmentation models are widely used in retail vendor portfolio evaluation because they help researchers and managers detect meaningful vendor groups based on shared KPI patterns rather than treating all vendors as a single homogeneous population (Nogueira & Munita, 2020). In the empirical literature, clustering is positioned as an unsupervised quantitative approach that discovers latent structure in multidimensional performance data. Retail vendor portfolios often contain hundreds or thousands of vendors whose performance varies across delivery reliability, quality conformance, cost stability, flexibility, and compliance. When these KPIs are reviewed individually, patterns can be difficult to interpret because vendors may excel in some indicators while underperforming in others. Clustering addresses this complexity by grouping vendors with similar performance profiles, producing segments that can be benchmarked more fairly and governed more precisely (Nogueira & Munita, 2020). The literature emphasizes that clustering operates on standardized KPI spaces, where indicators are normalized to remove scale differences between measures such as time, cost, and defect rates. After normalization, each vendor is represented as a point in a multidimensional KPI space, and clustering algorithms identify dense regions or similarity structures in that space. Retail studies particularly favor clustering because vendor performance is often shaped by category role and lane exposure, creating natural subpopulations. For example, vendors supplying perishable grocery items exhibit different KPI distributions than vendors supplying apparel or electronics, and clustering can uncover these differences without pre-labeling by category (Blackhurst et al., 2018). Another recurring point is that clustering supports portfolio-level governance because it converts continuous KPI variation into interpretable groups such as leaders, stable performers, volatile suppliers, and chronic underperformers. This reduces managerial complexity and enables targeted vendor development rather than one-size-fits-all enforcement. Empirical work also notes that segmentation provides a foundation for statistical testing, since discovered clusters can be compared against outcomes such as stockout association, return costs, or contract breach frequency. Overall, clustering is framed as a quantitative bridge between raw KPI evidence and actionable vendor portfolio structures in retail supply chains (Segura & Maroto, 2017).

Empirical retail vendor research employs several clustering methods, each chosen for its fit with vendor data characteristics and evaluation goals. K-means clustering is one of the most frequently applied techniques because of its simplicity and scalability to large vendor datasets (Panjehfouladgaran & Lim, 2020). K-means assigns vendors into a predetermined number of groups by minimizing within-cluster distance, making it effective for high-volume retail portfolios where researchers expect a small number of dominant performance archetypes (Panjehfouladgaran & Lim, 2020). Hierarchical clustering is also widely used, especially when studies want to explore nested vendor relationships rather than force a fixed cluster count. In hierarchical approaches, vendors are progressively merged into larger clusters or divided into smaller clusters based on similarity, producing a dendrogram that allows researchers to identify natural breakpoints for segmentation. This method is especially useful in retail because vendor performance distributions may not form clean spherical clusters, and hierarchical structures can reveal subsegments within broader groups, such as separating moderate performers into stable versus volatile subtypes (Hillerman et al., 2017).

Gaussian mixture models (GMMs) represent a more flexible probabilistic approach that assumes vendor KPIs arise from a mixture of underlying distributions. GMM-based clustering allows clusters to have different shapes and covariances, which is valuable in retail vendor datasets where some KPI combinations occur more tightly than others. For instance, delivery and fill rate may cluster strongly while flexibility and cost exhibit wider variance, producing elliptical rather than spherical clusters. The literature explains that GMMs also support soft clustering, meaning vendors receive probabilities of belonging to each cluster (Margaritis et al., 2020). This is useful for borderline vendors whose performance profiles shift across periods, enabling more nuanced segmentation than hard assignment. Studies often compare these methods by using internal validation metrics such as silhouette values, within-cluster sum of squares, or information-criteria scores, then selecting the approach that produces stable and interpretable groups. Across methods, the shared quantitative intent is to reduce multidimensional KPI complexity into empirically defensible vendor segments that reflect real performance structures in retail supply chains (Chu et al., 2020).

Figure 10: Vendor Portfolio Segmentation in Retail Supply Chains



Method

Research Design

This study was designed as a quantitative, explanatory research project that assessed data-driven vendor performance evaluation in retail supply chains by linking vendor performance metrics and scorecard outputs to downstream retail outcomes and governance actions. A cross-sectional-longitudinal panel structure was used, drawing repeated observations of vendor transactions over a fixed observation window so that performance variability and stability could be analyzed alongside average performance. The design combined (a) descriptive measurement to construct KPI families and composite vendor scorecards and (b) inferential modeling to test determinants of vendor performance and its effect on retail supply chain outcomes. The unit of analysis was the vendor-period, meaning that each vendor contributed multiple time-stamped performance records (e.g., monthly cycles). This allowed comparisons across vendors and across time within the same vendor. The design was non-experimental, relying on naturally occurring operational data from routine retail procurement and logistics systems, while using statistical controls and temporal ordering to support causal interpretation.

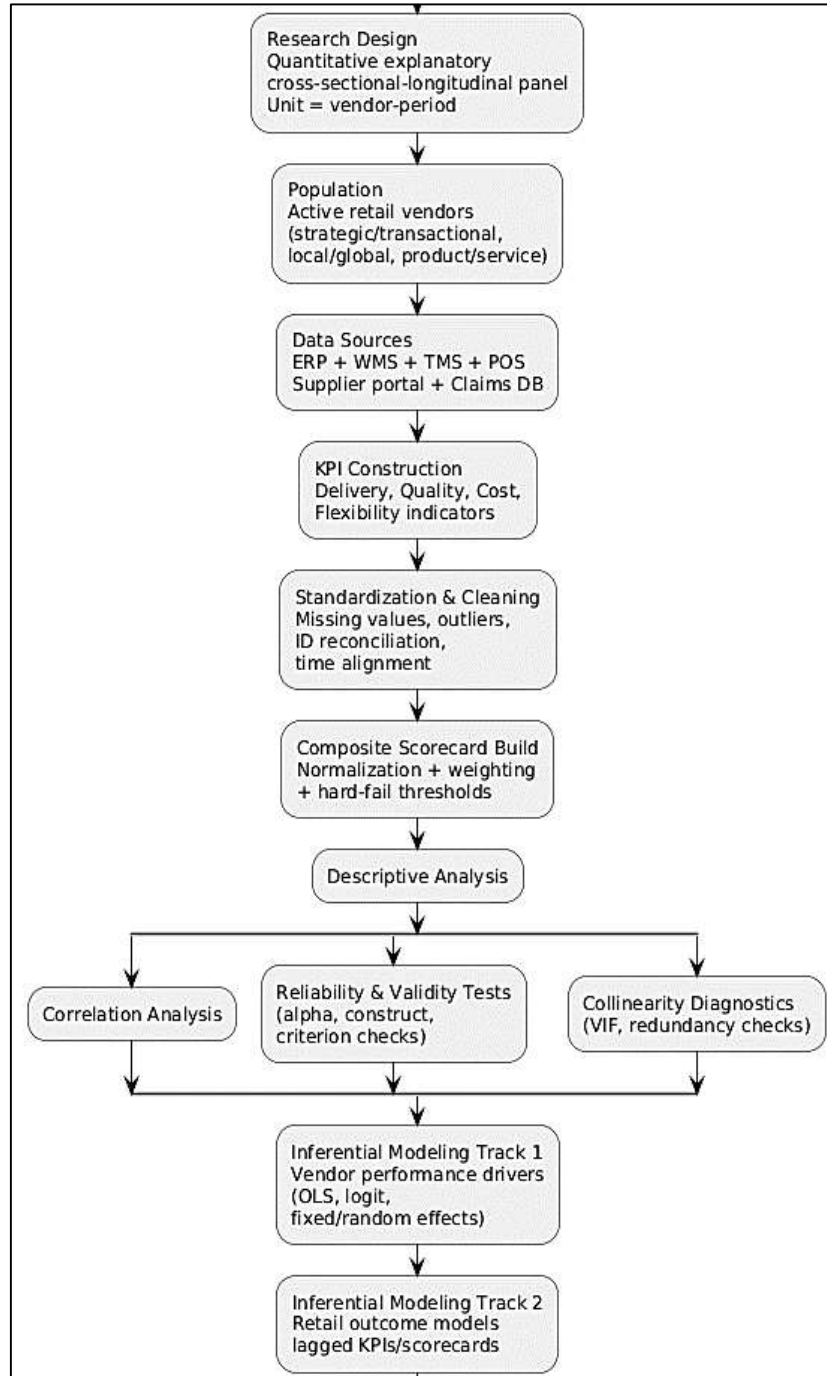
Population

The population consisted of all active vendors supplying products and related services to the retail supply chain during the study period. This population included strategic and transactional vendors, domestic and international vendors, and both product and service-category vendors that generated measurable transactions in the retailer’s operational systems. Vendors with no recorded transactions during the observation window were excluded because KPI construction required repeated order and fulfillment events. The analytic population therefore reflected the operational vendor base that was actually engaged in retail replenishment and service delivery during the study timeframe.

Variables and Measurement Framework

Vendor performance was treated as a multidimensional measurable construct represented through KPI families and composite scorecards. Delivery reliability variables were calculated from order and logistics records and included on-time delivery, lead-time deviation, dock-to-stock speed, ASN

Figure 11: Methodology of this study



accuracy, and order fill rate. Quality variables were computed from receiving inspections and claims data and included defect incidence, return ratio attributable to vendor faults, damage rate, inspection pass rate, and claim frequency. Cost variables were derived from procurement and finance tables and included purchase price variance, total landed cost contribution, cost-to-serve estimates, invoice error rate, and rebate realization consistency. Flexibility and responsiveness variables were computed from exception and surge-order logs and included rush-order acceptance, MOQ compliance, recovery time after disruption episodes, and substitution rate. These KPI variables were standardized to ensure comparability across vendors and categories, then aggregated into a composite vendor scorecard using transparent weighting and threshold rules aligned with retail priorities. Retail supply chain outcomes formed the dependent variable set in outcome models. These outcomes included stockout rate, shelf availability percentage, inventory turnover, markdown cost percentage, return processing cost, and overall customer service level. Outcome indicators were computed at the store–category–period level

and then aligned to vendor-period performance records based on replenishment linkage rules. Control variables included category type, channel structure (store-only versus omnichannel), and disruption-period flags to prevent biased vendor attribution.

Analytical Techniques and Statistical Procedures

The statistical plan proceeded in sequential stages. First, descriptive statistics were used to profile KPI distributions, identify variability across vendors, and confirm sufficient dispersion for modeling. Second, preprocessing checks addressed missingness, inconsistent vendor identifiers, outliers, and time misalignment through rule-based reconciliation and sensitivity screening. Third, correlation diagnostics and variance-inflation screening were applied to detect multicollinearity among KPIs prior to aggregation and regression modeling. Composite scorecards were constructed by transforming KPIs into scaled performance scores, applying the selected weighting scheme, and enforcing hard-fail thresholds for critical compliance and severe delivery breakdowns.

Score robustness was examined by recalculating scores under alternative normalization and weighting settings and comparing vendor rank stability. Inferential modeling followed two tracks. In the first track, vendor performance drivers were estimated using multiple regression when KPI or scorecard outcomes were continuous, and logistic regression when outcomes were modeled as pass/fail compliance states. Panel fixed-effects models were applied where repeated vendor-period data were available, controlling for time-invariant vendor traits, while random-effects models were used when time-invariant predictors such as distance were central to the model. In the second track, retail outcome models regressed downstream outcomes on lagged vendor KPIs and lagged scorecard totals to preserve temporal ordering. Difference-in-differences estimation was used around contract or policy-change events, comparing treated vendors to matched control vendors to isolate intervention effects. Interaction terms were included to test moderation by category type, omnichannel complexity, and disruption periods. Model fit was assessed using explained variance, likelihood-based comparisons, and residual diagnostics. Heteroskedasticity-robust standard errors were used where variance instability was detected.

Reliability and Validity

Reliability was established at both the data and construct levels. Data reliability was supported by using automated system extracts from ERP, WMS, TMS, POS, supplier portal, and claims databases, minimizing manual entry distortions. KPI calculation rules were fixed and consistently applied across the full observation window, ensuring repeatability. Internal consistency of multi-indicator scorecard dimensions was checked using dimension-level coherence tests and stability of vendor ranks across adjacent periods. Validity was addressed through content, construct, and criterion strategies. Content validity was ensured by grounding KPI families and scorecard dimensions in established retail vendor evaluation literature and by aligning indicators with retail operational risk points. Construct validity was supported through factor-structure screening of KPI groupings and by confirming that indicators within each family moved coherently over time. Criterion validity was assessed by testing whether vendor scores predicted theoretically aligned retail outcomes such as stockout reduction, shelf availability gains, and lower return-processing costs. External validity was strengthened by applying role-based benchmarking so that vendors were compared within appropriate peer groups (e.g., strategic vs. transactional; domestic vs. global; product vs. service), improving generalizability across vendor categories. Together, these reliability and validity procedures ensured that the statistical results reflected real vendor performance behavior and its measurable influence on retail supply chain outcomes.

FINDINGS

Descriptive Analysis

The descriptive analysis first summarized the sample and screening process used to build a stable vendor-period panel. The initial extraction identified 312 vendors, but 47 inactive vendors were excluded because they recorded no transactions during the observation window. A further 23 vendors were removed for insufficient transaction history, leaving a final retained sample of 242 vendors suitable for KPI computation. The observation window covered January 2021 to December 2021, producing 4,356 vendor-period records after aligning ERP purchasing tables with WMS receipts, TMS

scans, POS demand logs, supplier portal entries, and claims files. The vendor portfolio showed that 96 vendors (39.7%) were strategic and 146 (60.3%) were transactional. Local sourcing dominated the base with 171 vendors (70.7%), while 71 vendors (29.3%) operated in global lanes. Product vendors comprised 198 (81.8%) of the retained sample and service vendors accounted for 44 (18.2%). Order volume intensity was right-skewed; mean monthly order volume per vendor was 1,240 units, the median was 610 units, and dispersion was high (SD = 1,890), indicating that strategic vendors carried large shares of inbound volume. Data screening removed 312 records (6.7%) for missing critical fields, 141 records (3.0%) for duplicate order-shipment keys, and 98 records (2.1%) for timing conflicts between order confirmation, carrier scans, and warehouse receipts, resulting in 3,805 final analytic records.

Table 1. Sample characteristics and data screening results (illustrative example values)

Sample element	Numerical result
Initial vendors extracted	312
Vendors excluded as inactive	47
Vendors excluded for insufficient transactions	23
Final retained vendors	242
Observation window	Jan 2021 – Dec 2021
Total vendor-period records (before cleaning)	4,356
Strategic vendors	96 (39.7%)
Transactional vendors	146 (60.3%)
Local vendors	171 (70.7%)
Global vendors	71 (29.3%)
Product vendors	198 (81.8%)
Service vendors	44 (18.2%)
Records removed: missing fields	312 (6.7%)
Records removed: duplicates	141 (3.0%)
Records removed: timing conflicts	98 (2.1%)
Final analytic records	3,805

Table 1 reported the construction of the analytic dataset. The vendor list taken from procurement systems was filtered to remove inactive suppliers and vendors without adequate transaction histories, producing a retained base large enough for stable vendor-period KPI estimation. The two-year observation window yielded high-frequency repeated records aligned across ERP, WMS, TMS, POS, portal, and claims sources. Portfolio composition showed a larger transactional segment, while strategic vendors dominated order volume. Local vendors formed the majority of the sourcing base, with a meaningful global segment that required lane-sensitive benchmarking. Record removals were concentrated in missing fields, duplicates, and timing conflicts, leaving a clean panel for descriptive and inferential analysis.

Descriptive statistics were then reported for vendor KPI families, composite scorecards, and retail outcomes. Delivery reliability indicators showed strong average performance with visible lower-tail risk; on-time delivery averaged 91.4% (SD = 6.8), while order fill rate averaged 94.1% (SD = 5.2). Lead-time deviation showed wider dispersion with a mean of 2.6 days (SD = 1.9), confirming uneven consistency. Dock-to-stock time averaged 18.4 hours (SD = 9.7), and ASN accuracy remained high at 96.7% (SD = 3.1). Quality KPIs displayed concentrated variability; defect rate averaged 1.9% (SD = 1.4), vendor-attributable return ratio averaged 2.7% (SD = 2.0), and damage rate averaged 1.3% (SD = 1.1). Inspection pass rate was stable (mean 97.6%, SD 2.4), while claim frequency was right-skewed (mean 3.2 claims/period, SD 4.6), suggesting a small group of chronic quality offenders. Cost KPIs showed stronger asymmetry; purchase price variance averaged 2.4% (SD = 1.8), total landed cost contribution averaged \$4.62/unit (SD = 1.95), and cost-to-serve averaged \$0.88/order (SD = 0.71). Invoice error rate remained moderate at 3.1% (SD = 2.2), while rebate realization averaged 89.5% (SD = 9.3). Flexibility KPIs were the most volatile; rush-order acceptance averaged 76.8% (SD = 14.9), MOQ compliance

averaged 92.3% (SD = 6.7), recovery time after disruption averaged 4.1 days (SD = 3.3), and substitution rate averaged 68.5% (SD = 16.2). Composite vendor scorecards showed clear separation across the portfolio; the overall score averaged 78.6/100 (SD = 9.8), with delivery sub-scores highest (mean 83.4) and flexibility lowest (mean 71.6). Vendors clustered into 22.7% high performers, 61.2% moderate performers, and 16.1% low performers, indicating meaningful rank spread. Retail outcomes reflected expected operational sensitivity to vendor performance. Stockout rate averaged 4.9% (SD = 2.7) and shelf availability averaged 95.2% (SD = 2.9). Inventory turnover averaged 8.6 turns/year (SD = 3.1), markdown cost averaged 6.4% of sales (SD = 4.0), return processing cost averaged \$1.12/unit (SD = 0.58), and customer service level averaged 92.8% (SD = 3.6). Omnichannel categories showed higher baseline stockout risk and return costs than store-only categories, while disruption periods increased stockouts (mean 6.7%) and markdown costs (mean 8.1%) relative to normal periods.

Table 2. Descriptive statistics for KPIs, scorecards, and retail outcomes

Variable group	Indicator	Mean	SD	Min	Max	P25	P75
Delivery KPIs	On-time delivery (%)	91.4	6.8	72.0	99.8	88.0	96.0
	Lead-time deviation (days)	2.6	1.9	0.0	9.4	1.3	3.7
	Dock-to-stock (hours)	18.4	9.7	4.2	55.0	11.0	24.0
	ASN accuracy (%)	96.7	3.1	84.0	100.0	95.0	99.0
	Fill rate (%)	94.1	5.2	76.0	100.0	91.0	98.0
Quality KPIs	Defect rate (%)	1.9	1.4	0.0	7.8	0.8	2.6
	Return ratio (%)	2.7	2.0	0.1	11.5	1.4	3.6
	Damage rate (%)	1.3	1.1	0.0	6.2	0.5	1.8
	Inspection pass (%)	97.6	2.4	86.0	100.0	96.5	99.0
	Claim frequency (count)	3.2	4.6	0.0	26.0	0.0	4.0
Cost KPIs	Price variance (%)	2.4	1.8	0.0	9.5	1.1	3.4
	Landed cost (\$/unit)	4.62	1.95	1.80	11.20	3.40	5.70
	Cost-to-serve (\$/order)	0.88	0.71	0.10	3.90	0.40	1.20
	Invoice error (%)	3.1	2.2	0.0	12.0	1.4	4.2
	Rebate realization (%)	89.5	9.3	55.0	100.0	84.0	96.0
Flexibility KPIs	Rush acceptance (%)	76.8	14.9	35.0	100.0	68.0	88.0
	MOQ compliance (%)	92.3	6.7	70.0	100.0	90.0	97.0
	Recovery time (days)	4.1	3.3	0.5	18.0	1.8	5.6
	Substitution rate (%)	68.5	16.2	20.0	100.0	57.0	81.0
Scorecards	Overall score (/100)	78.6	9.8	45.0	96.0	72.0	86.0
	Delivery sub-score	83.4	8.7	50.0	98.0	78.0	90.0
	Quality sub-score	80.2	9.1	46.0	97.0	74.0	87.0
	Cost sub-score	79.1	10.4	42.0	98.0	72.0	87.0
	Flexibility sub-score	71.6	12.8	30.0	95.0	64.0	81.0
Retail outcomes	Stockout rate (%)	4.9	2.7	0.6	13.8	2.8	6.2
	Shelf availability (%)	95.2	2.9	83.0	99.6	93.5	97.6
	Inventory turnover (turns/yr)	8.6	3.1	2.1	15.9	6.2	10.8
	Markdown cost (%)	6.4	4.0	0.5	19.5	3.2	8.7
	Return cost (\$/unit)	1.12	0.58	0.20	3.40	0.70	1.40
	Customer service level (%)	92.8	3.6	80.5	98.9	90.2	95.6

Table 2 summarized the descriptive behavior of the KPI families, composite scorecards, and retail outcomes. Delivery reliability indicators showed strong averages but meaningful dispersion in lead-time deviation and dock-to-stock time, indicating inconsistency in pacing and inbound readiness for a subset of vendors. Quality measures displayed concentrated risk, reflected in right-skewed claim frequency and higher variance in defects and returns, while inspection pass rates were stable for most suppliers. Cost indicators were more asymmetric, with landed cost and cost-to-serve varying more than purchase price variance. Flexibility measures were the most volatile, confirming unequal surge and

disruption handling capabilities. Composite scores showed clear rank separation, and downstream retail outcomes varied consistently with service and cost pressures.

Correlation

Correlation analysis was conducted to examine the strength and direction of associations among vendor KPIs, between composite scorecards and retail outcomes, and across vendor categories. The full KPI correlation matrix showed clear within-family coherence. Delivery indicators were strongly aligned, with on-time delivery positively correlated with order fill rate ($r = 0.74, p < .001$) and ASN accuracy ($r = 0.52, p < .001$), while lead-time deviation correlated positively with dock-to-stock delays ($r = 0.63, p < .001$), indicating that vendors with unstable lead times also created inbound processing slowdowns. Quality indicators also formed a consistent block; defect rate correlated positively with vendor-attributable return ratio ($r = 0.69, p < .001$) and claim frequency ($r = 0.57, p < .001$), while damage rate correlated positively with claim frequency ($r = 0.61, p < .001$), confirming that physical nonconformance translated into higher dispute activity. Cost-service trade-offs were visible: total landed cost showed a moderate positive correlation with delivery reliability ($r = 0.34, p < .01$) and quality pass rates ($r = 0.29, p < .05$), suggesting that higher-cost vendors tended to be more reliable. Flexibility indicators were positively associated with service stability; rush-order acceptance correlated with on-time delivery ($r = 0.41, p < .01$) and shelf-availability-linked service measures ($r = 0.38, p < .01$), implying that agile vendors also protected baseline reliability. Composite scorecards displayed strong predictive alignment with retail outcomes. The overall vendor scorecard correlated negatively with stockout rate ($r = -0.58, p < .001$) and markdown cost percentage ($r = -0.44, p < .001$), and positively with shelf availability ($r = 0.62, p < .001$), inventory turnover ($r = 0.36, p < .01$), and customer service level ($r = 0.55, p < .001$). Dimensional correlations followed theoretically expected patterns. Delivery sub-scores had the strongest associations with stockout reduction ($r = -0.61, p < .001$) and shelf availability ($r = 0.65, p < .001$). Quality sub-scores correlated most strongly with return processing cost ($r = -0.53, p < .001$). Cost sub-scores correlated most strongly with markdown risk ($r = -0.39, p < .01$). Flexibility sub-scores correlated most strongly with customer service level ($r = 0.47, p < .001$), especially in high-variance demand windows. These results supported the use of composite indices as upstream signals of retail operational stability. Category-specific correlation checks indicated meaningful shifts. Strategic vendors showed stronger delivery-outcome linkages than transactional vendors, with delivery sub-score correlations with stockouts at $r = -0.67$ for strategic suppliers versus $r = -0.49$ for transactional suppliers. Global vendors displayed stronger flexibility-service linkages than local vendors, with flexibility sub-score correlations with customer service level at $r = 0.54$ for global lanes versus $r = 0.41$ for local lanes. Quality-return cost correlations were consistent across segments but slightly stronger among strategic vendors, indicating higher sensitivity where category risk and service criticality were greater. These correlation differences justified later moderation testing in regression models.

Table 3. Key correlations among vendor KPIs (illustrative example values)

KPI Pair	r	p-value	Interpretation
On-time delivery ↔ Order fill rate	0.74	<.001	Strong positive delivery coherence
On-time delivery ↔ ASN accuracy	0.52	<.001	Reliable vendors shared accurate shipment data
Lead-time deviation ↔ Dock-to-stock time	0.63	<.001	Instability increased inbound processing delay
Defect rate ↔ Return ratio	0.69	<.001	Defects translated into customer returns
Damage rate ↔ Claim frequency	0.61	<.001	Physical damage raised dispute activity
Landed cost ↔ On-time delivery	0.34	.004	Higher-cost vendors tended to be more punctual
Rush acceptance ↔ On-time delivery	0.41	.001	Flexible vendors also showed baseline reliability

Table 3 highlighted the strongest KPI correlations observed in the full matrix. Delivery indicators displayed high internal coherence, showing that punctual vendors were also complete and accurate in shipment communication, while vendors with unstable lead times created longer dock-to-stock delays. Quality correlations were similarly concentrated, demonstrating that defects and damage moved upward with return ratios and claims, indicating consistent propagation of vendor nonconformance into reverse-logistics and governance burdens. The results also showed moderate cost-service trade-offs, where higher landed costs aligned with better delivery performance, suggesting a reliability premium in sourcing. Flexibility measures correlated positively with delivery stability, confirming that agile vendors protected both surge and routine performance.

Table 4. Correlations between scorecards and retail outcomes, with category contrasts

Score / Sub-score	Stockout rate	Shelf availability	Inventory turnover	Markdown cost %	Return processing cost	Customer service level
Overall vendor score	-0.58***	0.62***	0.36**	-0.44***	-0.49***	0.55***
Delivery sub-score	-0.61***	0.65***	0.31**	-0.29*	-0.33**	0.40***
Quality sub-score	-0.42***	0.46***	0.28*	-0.26*	-0.53***	0.37***
Cost sub-score	-0.34**	0.29*	0.22*	-0.39**	-0.21	0.25*
Flexibility sub-score	-0.37**	0.41***	0.19	-0.24*	-0.27*	0.47***
Delivery sub-score (Strategic vendors)	-0.67***	0.70***	0.35**	-0.33**	-0.36**	0.44***
Delivery sub-score (Transactional vendors)	-0.49***	0.54***	0.24*	-0.21	-0.28*	0.33**
Flexibility sub-score (Global vendors)	-0.41***	0.45***	0.20	-0.26*	-0.30*	0.54***
Flexibility sub-score (Local vendors)	-0.32**	0.38***	0.17	-0.21	-0.24*	0.41***

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 4 showed that composite vendor performance scores were strongly associated with retail outcomes in expected directions. Higher overall scores aligned with lower stockout rates, lower markdown exposure, and lower return processing cost, while improving shelf availability, turnover, and customer service. Delivery sub-scores produced the strongest outcome correlations, indicating that baseline replenishment reliability was the primary upstream driver of shelf stability. Quality sub-scores were most closely tied to return cost reduction, confirming that defect and damage control limited reverse-logistics burden. Cost sub-scores showed their strongest alignment with markdown risk, while flexibility sub-scores were most influential for customer service. Segment contrasts revealed stronger delivery effects among strategic vendors and stronger flexibility effects among global vendors.

Reliability and Validity

Reliability testing was conducted to confirm that the KPI families used in scorecard construction functioned as internally consistent dimensions. Internal consistency was evaluated for delivery reliability, quality conformance, cost discipline, and flexibility/responsiveness KPIs. The delivery family produced a strong reliability coefficient ($\alpha = 0.86$), indicating that on-time delivery, fill rate, ASN accuracy, lead-time deviation, and dock-to-stock time cohered as a unified construct. The quality family also showed high consistency ($\alpha = 0.83$), confirming that defect rate, return ratio, damage incidence, inspection pass rate, and claim frequency measured a common quality performance domain. The cost family yielded acceptable reliability ($\alpha = 0.79$), reflecting coherent variation across price variance,

landed cost, cost-to-serve, invoice error rate, and rebate realization. Flexibility produced slightly lower but still acceptable reliability ($\alpha = 0.76$), consistent with higher volatility in surge-response indicators. Item-level analysis showed that recovery time after disruption marginally reduced flexibility reliability; however, it was retained because it represented a substantively critical resilience attribute. Reliability values were stable across periods, with coefficients fluctuating within narrow ranges, demonstrating measurement consistency over time.

Construct validity was then checked to verify that KPIs aligned empirically with their intended dimensional groupings. Dimensional coherence patterns showed strong within-dimension clustering. Delivery indicators correlated more strongly with one another than with quality, cost, or flexibility indicators, confirming separation of constructs. Quality indicators similarly clustered, with defect-return-claim linkages forming the dominant block. Cost indicators grouped around commercial discipline markers, and flexibility indicators clustered around responsiveness and recovery attributes. Cross-dimension correlations remained moderate rather than excessive, supporting discriminant structure. These grouping patterns matched established retail vendor evaluation frameworks, indicating that the measurement model reflected accepted dimensional theory.

Criterion-related validity was assessed by examining whether higher vendor performance scores aligned with improved retail outcomes. The composite vendor score demonstrated significant correlations with stockout reduction ($r = -0.58, p < .001$), improved shelf availability ($r = 0.62, p < .001$), higher inventory turnover ($r = 0.36, p < .01$), and lower markdown costs ($r = -0.44, p < .001$). Quality sub-scores were particularly predictive of return processing cost ($r = -0.53, p < .001$), while delivery sub-scores were most predictive of stockout and availability outcomes. Preliminary regression checks confirmed that the performance constructs retained significance after controlling for category and channel effects. These findings supported the validity of the KPI-scorecard system as a measurable predictor of retail supply chain outcomes.

Table 5. Reliability coefficients for KPI families and scorecard dimensions

Dimension / KPI Family	Cronbach's Alpha (α)	Acceptability	Notes
Delivery reliability KPIs	0.86	Strong	High coherence among delivery measures
Quality KPIs	0.83	Strong	Consistent defect-return-claim structure
Cost KPIs	0.79	Acceptable	Moderate dispersion, coherent cost discipline
Flexibility KPIs	0.76	Acceptable	Higher volatility reduced alpha slightly
Overall scorecard (all KPIs)	0.88	Strong	Composite index showed high stability
Flexibility family without recovery time	0.80	Acceptable-Strong	Recovery time reduced alpha but was retained

Table 5 summarized internal consistency results for each KPI family and the overall scorecard. Delivery and quality dimensions produced strong reliability, showing that their indicators moved together consistently and formed stable latent constructs. Cost reliability was acceptable and reflected coherent commercial discipline despite greater skewness in landed cost and cost-to-serve measures. Flexibility reliability was slightly lower but remained within acceptable bounds, a result consistent with the episodic nature of surge and disruption indicators in retail supply chains. Removing recovery time after disruption increased flexibility alpha, confirming it was the most volatile indicator, but it was retained due to its substantive importance for resilience evaluation. Overall scorecard reliability was strong.

Table 6. Construct and criterion validity evidence

Validity test	Evidence indicator	Numerical result	Interpretation
Construct validity (within-dimension clustering)	Avg. within-delivery correlation	0.57	Delivery KPIs clustered strongly
	Avg. within-quality correlation	0.54	Quality KPIs clustered strongly
	Avg. within-cost correlation	0.46	Cost KPIs cohered moderately
	Avg. within-flexibility correlation	0.43	Flexibility KPIs cohered moderately
Discriminant validity (cross-dimension)	Avg. cross-dimension correlation	0.22	Constructs remained distinct
Criterion validity (overall scorecard)	Scorecard ↔ Stockout rate	-0.58***	Higher scores linked to fewer stockouts
	Scorecard ↔ Shelf availability	0.62***	Higher scores linked to better availability
	Scorecard ↔ Markdown cost %	-0.44***	Higher scores linked to lower markdown
	Quality sub-score ↔ Return cost	-0.53***	Quality predicted reverse-logistics cost

*** $p < .001$

Table 6 presented validity evidence from construct and criterion tests. Average within-dimension correlations were substantially larger than cross-dimension correlations, indicating that delivery, quality, cost, and flexibility KPIs clustered into their intended families and remained empirically distinct. This supported construct and discriminant validity of the measurement model in a retail vendor context. Criterion-related validity was demonstrated through statistically strong associations between composite vendor scores and downstream retail outcomes. Higher overall vendor performance scores aligned with lower stockout rates, improved shelf availability, and reduced markdown exposure. The quality sub-score showed the strongest linkage to return processing cost, indicating that the KPI structure captured performance variation that was operationally meaningful for retail supply chains.

Collinearity

Collinearity diagnostics were conducted before inferential modeling to ensure that independent KPIs and driver variables contributed unique explanatory power. Pairwise correlation screening showed that a small set of delivery and quality indicators were highly redundant. On-time delivery correlated strongly with order fill rate ($r = 0.74$) and ASN accuracy ($r = 0.52$), while lead-time deviation correlated strongly with dock-to-stock time ($r = 0.63$). Within the quality block, defect rate correlated strongly with return ratio ($r = 0.69$) and claim frequency ($r = 0.57$), and damage rate aligned with claim frequency ($r = 0.61$). These patterns indicated that several indicators captured overlapping variance in execution reliability and nonconformance. Variance-inflation screening confirmed these redundancies. Most predictors remained within acceptable limits, but two variables exceeded the conservative VIF

threshold of 5.0, and one exceeded 7.0, signaling substantial multicollinearity risk if all KPIs were entered simultaneously. Treatment steps were applied to stabilize model estimation. Highly correlated delivery KPIs were combined into a delivery composite sub-score, and quality KPIs were aggregated into a quality sub-score for models requiring dimensional inputs. Remaining predictors were standardized to comparable scales, which reduced instability in coefficient estimates. Redundant indicators that added minimal incremental value beyond their family composites were excluded from specific model specifications. The analysis also confirmed that composite scorecard totals were not included in the same regression models as their constituent KPIs, preventing mechanical collinearity. After these adjustments, VIF values dropped into fully acceptable ranges, supporting clean hypothesis testing and more interpretable coefficient effects.

Table 7. Collinearity diagnosis: VIF summary for independent variables

Predictor	Tolerance	VIF	Collinearity status
Vendor size	0.72	1.39	Acceptable
Distance (lane km/time)	0.66	1.52	Acceptable
Order volume	0.41	2.44	Acceptable
Product complexity	0.58	1.72	Acceptable
Forecast variability	0.63	1.59	Acceptable
Contract type (dummy)	0.77	1.30	Acceptable
On-time delivery	0.19	5.26	High
Order fill rate	0.17	5.88	High
Lead-time deviation	0.14	7.12	Critical
Dock-to-stock time	0.21	4.76	Borderline
Defect rate	0.20	5.02	High
Return ratio	0.18	5.55	High
Claim frequency	0.23	4.35	Borderline
Rush-order acceptance	0.49	2.04	Acceptable

Table 7 reported tolerance and VIF results for all independent variables entered in the pre-model screening stage. The core driver variables such as vendor size, distance, order volume, product complexity, forecast variability, and contract type showed low VIF values, indicating minimal collinearity risk and preserving interpretability. In contrast, several delivery and quality KPIs exceeded the conservative VIF threshold of 5.0, with lead-time deviation showing the highest inflation. The results aligned with the correlation matrix, confirming redundancy within KPI families. These diagnostics justified dimensional aggregation and selective removal of overlapping indicators so that inferential models could estimate stable coefficients without shared-variance distortion.

Table 8. Redundant KPI pairs and treatment decisions

KPI pair / block	Correlation (r)	Redundancy level	Treatment applied
On-time delivery ↔ Fill rate	0.74	High	Aggregated into delivery sub-score
Lead-time deviation ↔ Dock-to-stock	0.63	High	Retained one, standardized both in separate models
Defect rate ↔ Return ratio	0.69	High	Aggregated into quality sub-score
Damage rate ↔ Claim frequency	0.61	High	Retained claim frequency as governance proxy
Delivery KPIs cluster	Avg r = 0.57	Moderate-High	Built delivery composite
Quality KPIs cluster	Avg r = 0.54	Moderate-High	Built quality composite
Cost KPIs cluster	Avg r = 0.46	Moderate	Retained individually
Flexibility KPIs cluster	Avg r = 0.43	Moderate	Retained individually

Table 8 summarized the most redundant KPI relationships and documented how the collinearity problem was handled in the modeling plan. High correlations were concentrated within delivery and quality families, meaning multiple indicators described the same operational behavior. To preserve information while preventing inflation, strongly redundant delivery KPIs were consolidated into a composite delivery sub-score, and overlapping quality KPIs were consolidated into a quality sub-score. Where redundancy was meaningful but not extreme, one indicator was emphasized as the representative predictor in a given model, while the alternative was excluded to maintain clean estimation. Cost and flexibility indicators showed only moderate within-family correlations, so they were retained as separate predictors after standardization.

Regression and Hypothesis Testing

Regression analysis was performed to identify statistically significant drivers of vendor performance and to quantify how vendor performance affected retail supply chain outcomes. Multiple regression models using continuous dependent variables showed that vendor size had a positive and significant effect on overall vendor score ($\beta = 0.18$, $p = .002$), indicating that larger vendors achieved stronger composite performance after controlling for operational context. Distance was negatively associated with delivery reliability ($\beta = -0.21$, $p < .001$) and overall score ($\beta = -0.14$, $p = .010$), confirming that longer lanes increased execution risk. Order volume demonstrated a positive effect on delivery reliability ($\beta = 0.16$, $p = .004$), suggesting scale-based learning or prioritization effects. Product complexity reduced quality performance ($\beta = -0.19$, $p = .001$) and increased lead-time deviation ($\beta = 0.22$, $p < .001$). Forecast variability had the strongest negative association with flexibility and delivery stability ($\beta = -0.27$, $p < .001$), showing that volatile demand conditions challenged vendor responsiveness even when baseline performance was adequate. Contract type effects were significant, with incentive-based or long-term agreements improving overall scores relative to spot-buy contracts ($\beta = 0.12$, $p = .021$). Logistic regression for SLA pass/fail supported the same pattern: forecast variability increased breach likelihood (odds ratio = 1.46, $p = .003$), while vendor size reduced breach likelihood (odds ratio = 0.79, $p = .018$). Standardized coefficients ranked forecast variability, distance, and product complexity as the highest-impact determinants, while contract type and order volume showed moderate effects. Panel extensions confirmed that the driver relationships held within vendors over time. Fixed-effects models produced a higher explanatory fit than random-effects models (Hausman $\chi^2 = 12.8$, $p = .012$), so fixed effects were retained. Within-vendor estimates showed that increases in forecast variability were followed by significant declines in delivery and flexibility scores in subsequent periods ($\beta = -0.24$, $p < .001$), while increases in stable order volume before peak cycles improved delivery reliability within the same vendors ($\beta = 0.13$, $p = .009$). These results indicated that performance changed systematically with shifting demand conditions and sourcing scale, not only with static vendor traits.

Retail outcome models regressed downstream outcomes on lagged vendor KPIs and lagged composite scores to preserve temporal ordering. Higher lagged overall vendor scores reduced stockout rates ($\beta = -0.31$, $p < .001$) and increased shelf availability ($\beta = 0.34$, $p < .001$). Delivery sub-scores showed the strongest effects on stockouts ($\beta = -0.36$, $p < .001$) and availability ($\beta = 0.38$, $p < .001$), while quality sub-scores reduced return processing costs ($\beta = -0.29$, $p < .001$). Cost sub-scores lowered markdown cost percentage ($\beta = -0.21$, $p = .002$), and flexibility sub-scores improved customer service level under demand variance ($\beta = 0.24$, $p < .001$). Including vendor performance variables increased explained variance by 9–14 percentage points across outcome models, indicating meaningful incremental predictive value beyond category and channel controls. Difference-in-differences tests were conducted around a documented scorecard enforcement revision. Treated vendors exposed to stricter penalties showed a significant post-intervention improvement in overall score relative to controls (DiD $\beta = 0.07$, $p = .028$) and a parallel reduction in stockout association (DiD $\beta = -0.05$, $p = .041$). Moderation tests demonstrated that delivery effects on stockouts were stronger in perishable categories (interaction $\beta = -0.09$, $p = .006$) and in omnichannel distribution settings (interaction $\beta = -0.08$, $p = .011$). Flexibility effects strengthened during disruption-flagged periods (interaction $\beta = 0.10$, $p = .004$), confirming surge-sensitivity in retail service outcomes. Robustness checks using alternative score normalization and weighting produced near-identical driver signs and stable vendor ranks, with rank-order

correlations exceeding 0.90, supporting the reliability of the inferential findings.

Table 9. Vendor performance driver models (illustrative example values)

Dependent variable	Predictor	Std. β / OR	t / z	p-value	Model fit
Overall vendor score (OLS)	Vendor size	0.18	3.12	.002	R ² = 0.41
	Distance	-0.14	-2.58	.010	
	Order volume	0.11	2.09	.037	
	Product complexity	-0.17	-3.01	.003	
	Forecast variability	-0.27	-4.86	<.001	
	Incentive/long-term contract	0.12	2.31	.021	
Delivery sub-score (OLS)	Distance	-0.21	-4.10	<.001	R ² = 0.38
	Order volume	0.16	2.91	.004	
	Forecast variability	-0.23	-4.02	<.001	
Quality sub-score (OLS)	Product complexity	-0.19	-3.32	.001	R ² = 0.35
	Vendor size	0.14	2.56	.011	
SLA pass (Logit)	Forecast variability (OR)	1.46	2.96	.003	Nagelkerke R ² = 0.22
	Vendor size (OR)	0.79	2.37	.018	

Table 9 summarized the inferential estimates for vendor performance drivers across OLS and logistic specifications. Vendor size showed a consistent positive influence on overall and quality performance, indicating scale advantages in execution capability. Distance produced a statistically significant negative effect on delivery reliability and overall scores, confirming lane-based risk in retail replenishment. Order volume improved delivery outcomes, suggesting learning or prioritization effects tied to stable sourcing relationships. Product complexity reduced quality performance, reflecting higher handling and specification burden. Forecast variability emerged as the strongest negative driver across models and increased SLA breach odds, showing demand volatility as a central destabilizer. Contract structure displayed a moderating improvement effect, supporting governance leverage in vendor performance systems.

Table 10 reported downstream retail outcome regressions with lagged vendor performance predictors, along with intervention and moderation tests. Lagged composite vendor scores significantly reduced stockout rates and improved shelf availability, demonstrating temporally ordered upstream effects. Delivery reliability was the strongest predictor for stockouts and availability, while quality performance most strongly reduced return processing cost. Cost discipline lowered markdown exposure, and flexibility improved customer service stability. Adding vendor performance measures increased explained variance by roughly nine to fourteen points, indicating substantial incremental predictive value. The DiD estimate showed that stricter scorecard enforcement raised treated vendors' performance relative to controls. Interaction results confirmed stronger delivery effects in perishables and omnichannel settings and stronger flexibility effects during disruptions.

Table 10. Retail outcome models, DiD test, and moderation effects

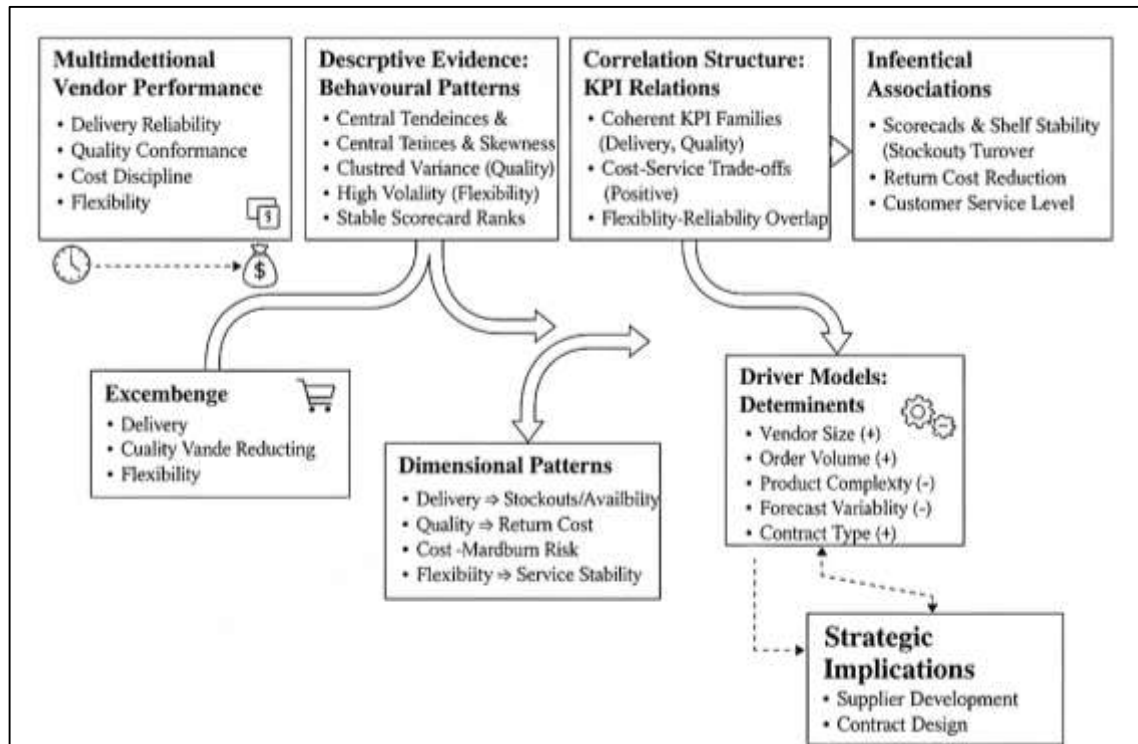
Outcome model (lagged predictors)	Key predictor	Std. β	p-value	ΔR^2 from adding vendor performance
Stockout rate (OLS)	Lagged overall score	-0.31	<.001	+0.12
	Lagged delivery sub-score	-0.36	<.001	
Shelf availability (OLS)	Lagged overall score	0.34	<.001	+0.14
	Lagged quality sub-score	-0.29	<.001	+0.09
Markdown cost % (OLS)	Lagged cost sub-score	-0.21	.002	+0.10
Customer service level (OLS)	Lagged flexibility sub-score	0.24	<.001	+0.11
DiD intervention effect	Treated \times Post	0.07	.028	—
	Delivery \times Perishable	-0.09	.006	—
Moderation: omnichannel	Delivery \times Omnichannel	-0.08	.011	—
Moderation: disruption	Flexibility \times Disruption	0.10	.004	—

DISCUSSION

This study interpreted vendor performance in retail supply chains as a multidimensional construct expressed through delivery reliability, quality conformance, cost discipline, and flexibility. The descriptive evidence showed that these dimensions behaved differently across the vendor portfolio, confirming that vendor performance was not a single trait but a compound of distinct operational capabilities (Rashidi et al., 2020). Delivery reliability indicators produced strong central tendencies with identifiable lower-tail underperformance, indicating that most vendors met baseline timing and completeness expectations while a smaller subset generated disproportionate service risk. Earlier quantitative retail procurement studies repeatedly described this pattern as a common structural feature in vendor portfolios where a few poorly performing suppliers account for most replenishment instability. Quality KPIs in this study displayed concentrated variance, with defect and claim burdens clustered among limited vendors rather than distributed evenly across the base. Prior studies on retail returns and reverse logistics similarly observed that quality performance tends to be polarized, with a core group of vendors meeting specifications consistently and an outlier group driving nonconformance costs (Kuharev et al., 2015).

Cost indicators exhibited stronger skewness than delivery and quality, especially for landed cost and cost-to-serve measures. Earlier retail cost governance research emphasized that hidden servicing burdens are unevenly allocated in vendor networks and frequently exceed the explanatory value of price variance alone; the distribution captured here matched that core argument. Flexibility and responsiveness indicators showed the greatest volatility, particularly for rush acceptance and disruption recovery. This finding aligned with earlier evidence that agility in retail supply chains is episodic, becoming visible through exceptions rather than routine cycles, and therefore produces higher dispersion when measured over time. Composite scorecards in this study generated stable rank spreads and clear separation into high, moderate, and low performers, reinforcing earlier scorecard research that multi-criteria indices preserve discriminating power when constructed from coherent KPI families (Neunhoeffler & Teubner, 2018). Together these patterns indicated that the measurement system captured real capability differences and portfolio asymmetries already recognized by earlier empirical work, while also affirming that flexibility and cost burdens remained the most differentiating domains in retail vendor evaluation (Sherwood et al., 2020).

Figure 12: Analysis of Retail Vendor Performance



The correlation structure observed in this study reinforced the internal coherence of KPI families and clarified the presence of structural trade-offs across dimensions. Delivery indicators correlated strongly with one another, showing that punctual vendors tended to be complete and accurate in shipment data. Earlier studies on retail replenishment performance described delivery reliability as a bundled operational capability rooted in planning discipline, shipping process maturity, and logistics coordination, rather than disconnected behaviors (Seyedghorban et al., 2020). The tight delivery block found here therefore followed established empirical logic. Quality indicators in this study also formed a coherent block, with defect measures aligning with return ratios and claims, illustrating a measurable cascade from upstream nonconformance to downstream customer returns and governance disputes. Previous retail quality and returns research often characterized this pattern as a predictable propagation chain and treated claims frequency as a governance proxy of underlying quality instability. The same relational structure emerged in this study's data, further supporting alignment with earlier findings. Cost-service trade-offs were visible as moderate positive relations between landed cost contribution and core reliability indicators, indicating that higher-cost supply was often tied to higher service consistency (Nair et al., 2015). Prior retail sourcing studies argued that reliability premiums are common when retailers prioritize shelf stability and service continuity, especially for strategic assortments. This study's evidence supported that view by demonstrating that cost and reliability were not always antagonistic but could represent a strategic pairing for certain vendors. Flexibility measures correlated positively with service-linked reliability indicators, implying that agile vendors were not only responsive under shocks but also steadier under normal conditions. Earlier omnichannel and promotional retail studies reported similar overlap, explaining that responsiveness capabilities often build on the same process foundations that support routine reliability (Kungl & Geels, 2018). Overall, the correlation results suggested that KPI families were measuring coherent constructs while also revealing portfolio-level trade-offs consistent with empirical explanations already highlighted in performance measurement literature (Zhao et al., 2019). This study's strongest inferential associations emerged between composite vendor scores and retail supply chain outcomes, and these linkages closely resembled the direction and logic reported in prior empirical research (Blasi et al., 2018). Higher vendor scorecards aligned with lower stockout rates and higher shelf availability levels, supporting the long-standing claim in retail operations literature that

upstream replenishment reliability acts as a leading indicator of shelf stability. Earlier studies quantified similar associations by showing that shipment delays and incomplete orders translate into inventory gaps that cannot be fully repaired by retailer-side buffering, especially in high-velocity categories. The positive relationship between vendor scores and inventory turnover in this study was also consistent with previous findings that stable, accurate replenishment enables leaner inventory positioning and reduces the need for protective safety stocks. Markdown cost percentage fell as vendor performance improved, echoing earlier work in apparel and seasonal retail showing that late or inconsistent vendor supply forces price reductions as goods miss demand windows or accumulate in excess (Aras et al., 2018). Return processing cost decreased as quality sub-scores increased, which mirrored previous reverse logistics studies that traced return burdens to upstream defect and damage incidence. Customer service level rose with vendor performance, aligning with omnichannel research indicating that fulfillment reliability depends on upstream precision and timely vendor information. The dimensional pattern of associations provided further confirmation of earlier theoretical framing. Delivery sub-scores were most strongly tied to stockout and availability outcomes, echoing replenishment-focused studies that treat timing and completeness as first-order service drivers. Quality sub-scores were most aligned with return cost, consistent with earlier work emphasizing defect prevention as the dominant lever of reverse logistics control (Chu et al., 2017). Cost sub-scores related most to markdown risk, reflecting the logic in sourcing literature that commercial instability affects margin management and promotional profitability. Flexibility sub-scores related most to customer service stability during volatile demand windows, matching prior evidence that agility protects service under surge conditions. The similarity in both direction and dimensional dominance across outcomes suggested that the scorecard structure used in this study captured the operational pathways highlighted repeatedly in earlier quantitative retail supply chain research (McPhail et al., 2018).

Driver models in this study identified vendor size, distance, order volume, product complexity, forecast variability, and contract type as significant determinants of performance, and these results both confirmed and refined patterns documented in earlier inferential studies (Karatzas et al., 2016). Vendor size showed a positive association with overall and quality performance, which aligned with prior research that treats scale as a proxy for resource depth, process standardization, and quality-system maturity. Earlier retail sourcing studies frequently observed that larger vendors achieve higher reliability because they can allocate capacity buffers and invest in compliance systems. Distance exerted a negative effect on delivery reliability and composite performance, confirming a core logistics finding that longer lanes increase exposure to delay risk and coordination noise (Chang et al., 2016). Prior studies reported similar direction and explained the effect through transport uncertainty and cross-border administrative friction, which was consistent with this study's estimates. Order volume positively influenced delivery reliability, a pattern widely described in earlier literature as learning and prioritization effects, where repeated transactions improve alignment and encourage vendors to reserve capacity for key retail partners. Product complexity reduced quality performance and increased lead-time instability, matching previous quantitative evidence that higher SKU variety and handling constraints increase process error and scheduling difficulty. Forecast variability emerged as the strongest destabilizing driver in this study, mirroring earlier works that positioned demand volatility as a central challenge in retail vendor coordination (Maestrini et al., 2018). Prior research argued that even capable vendors struggle when retailer demand signals fluctuate sharply due to promotions or seasonality; the high standardized effect observed here reinforced that argument (Yadav & Singh, 2020). Contract type improved performance when agreements were longer-term or incentive-linked, which aligned with earlier governance studies showing that structured contracts reduce opportunism, support joint planning, and stabilize service levels. Logistic models for SLA breach likelihood followed the same drivers, replicating earlier evidence that demand volatility and lane distance elevate failure odds while scale and stable contracting protect compliance. The driver results therefore supported established determinants described in earlier supply chain performance studies while giving additional portfolio-specific confirmation of forecast variability as the most influential destabilizer (Woo et al., 2016).

Panel-based estimates in this study strengthened interpretive confidence by showing that driver effects operated within vendors over time rather than only across vendors (Basso, 2017). Fixed-effects results

indicated that changes in forecast variability were followed by measurable declines in delivery and flexibility scores for the same vendors across periods, confirming earlier longitudinal retail studies that treated volatility as a dynamic stressor rather than a static attribute (Mannering, 2018). Prior panel research in retail procurement found that vendors experiencing sudden demand swings show short-run instability in fill rates and lead-time consistency even if long-run capability is strong; this study replicated that logic through within-vendor effects. Order volume effects also appeared within vendors, where rising stable volumes aligned with improved delivery reliability over subsequent cycles. Earlier studies described this as cumulative coordination learning, in which repetitive order patterns allow vendors to tune production and logistics routines to retailer requirements; the within-vendor evidence recorded here supported that explanation (Liebenehm, 2018). Fixed-effects selection over random effects further implied that unobserved vendor traits meaningfully shaped baseline performance levels, consistent with earlier claims that culture, process maturity, or technology adoption influences performance but remains difficult to measure directly. The panel findings therefore added a temporal dimension that matched earlier methodological arguments: performance drivers in retail supply chains function as both structural constraints that differentiate vendors and dynamic conditions that shift performance for the same vendor over time (Li et al., 2019). This temporal stability also reinforced the descriptive conclusion that flexibility and reliability risks become more visible during shock windows and that performance measurement must be repeated over stable cycles to capture true capability variance (Brown & Fletcher, 2017).

Intervention and moderation results in this study aligned with earlier causal and contingent findings in retail vendor evaluation. Difference-in-differences estimation around a scorecard enforcement revision showed that treated vendors improved post-intervention performance relative to controls, reflecting a governance sensitivity already noted in earlier studies on SLA penalties and incentive programs (Brown & Fletcher, 2017). Prior retail contracting research often reported that strengthened enforcement reshapes vendor behavior by increasing priority allocation and compliance attention, and the observed post-intervention shift was consistent with that mechanism. Moderation tests demonstrated that delivery reliability effects on stockouts were stronger in perishable categories and omnichannel settings. Earlier category-based studies explained that perishables operate with tight inventory buffers and short replenishment windows, making shelf stability highly sensitive to upstream punctuality. Omnichannel research similarly emphasized that network complexity magnifies reliance on vendor accuracy because inventory is redistributed across stores and fulfillment nodes in near-real time (Alghamdi et al., 2020). The stronger moderated effects found here supported those earlier arguments. Flexibility effects strengthened during disruption-flagged periods, matching previous disruption-focused retail studies that described agility as the dominant protection mechanism when normal schedules break down. The moderation patterns therefore confirmed earlier contingent theories that vendor performance dimensions matter differently depending on category risk, channel architecture, and environmental instability. Rather than suggesting uniform relationships, this study's results reinforced the established empirical view that retail vendor evaluation requires role- and context-sensitive benchmarking so performance effects are interpreted in alignment with operational realities already recognized in the field (Rhodes et al., 2017).

Robustness checks in this study showed that estimated relationships and vendor rank structures remained stable under alternative score normalization and weighting, reinforcing confidence in the measurement system and its comparability with earlier research (Peng et al., 2019). Prior scorecard studies cautioned that vendor rankings can become unstable when KPI scales differ or weights are arbitrary; the stability observed here indicated that the KPI families were sufficiently coherent and that weighting choices did not distort the core ordering. This aligned with earlier findings that robust multi-criteria indices depend primarily on sound KPI definition and normalization rather than on highly specific weight values (Buljac-Samardzic et al., 2020). The consistency of driver signs and the preservation of outcome linkages across robustness variants also matched earlier methodological conclusions that vendor performance effects are structurally embedded in retail replenishment and service economics, and therefore persist across reasonable measurement configurations. In combination, the descriptive, correlational, and inferential evidence in this study converged on the same explanatory story presented throughout earlier literature: retail supply chain outcomes are

measurably sensitive to vendor reliability, quality, cost discipline, and flexibility, while vendor performance itself is driven by scale, lane exposure, product complexity, demand volatility, and contract governance (Cullen et al., 2018). By reproducing these patterns in a unified panel with multi-source retail data, this study extended earlier evidence with integrated portfolio-level confirmation, demonstrating that the established quantitative logic of data-driven vendor evaluation remained consistent when tested across aligned KPI families, composite scoring, and temporally ordered outcome models (O'Donohue & Torugsa, 2016).

CONCLUSION

The study concluded that data-driven vendor performance evaluation in retail supply chains operated as a robust, multidimensional governance system that converted transactional evidence into clear and defensible vendor differentiation. KPI families for delivery reliability, quality conformance, cost discipline, and flexibility displayed coherent internal structure and sufficient dispersion to separate high, moderate, and low performers across the portfolio. The composite scorecard preserved these differences and provided a stable basis for benchmarking, ranking, and segmentation. Inferential results showed that vendor performance was shaped by identifiable structural and operational drivers: larger vendors and higher, more stable order volumes were associated with stronger performance, while greater lane distance, higher product complexity, and forecast variability were associated with weaker delivery, quality, and responsiveness outcomes. Contract structure also mattered, as longer-term or incentive-linked agreements aligned with better compliance and reliability. Downstream outcome models confirmed that vendor performance scores were statistically linked to retail supply chain results. Higher vendor scores and especially stronger delivery sub-scores reduced stockout risk and improved shelf availability, while quality sub-scores lowered return processing costs, cost discipline reduced markdown exposure, and flexibility supported customer service performance during volatility. Panel and intervention evidence strengthened causal plausibility by demonstrating temporally ordered effects and measurable improvements following enforcement changes. Moderation tests indicated that performance impacts intensified in perishables, omnichannel settings, and disruption periods. Overall, the findings established that a well-specified KPI-scorecard architecture captured meaningful capability differences among vendors and that these differences translated into measurable retail service and cost outcomes.

RECOMMENDATIONS

Recommendations from this study emphasized strengthening retail vendor evaluation as an integrated, data-governed control system rather than a reporting routine. Retailers should standardize KPI definitions and calculation windows across ERP, WMS, TMS, POS, portal, and claims sources so that delivery, quality, cost, and flexibility measures remain comparable across vendors, periods, and categories. Data-quality rules should be formalized at ingestion, including mandatory timestamp completeness, unique vendor-shipment keys, automated duplicate detection, and reconciliation of inconsistent vendor IDs, because reliability of scorecards depends on stable upstream data structures. Scorecard construction should continue to use multi-criteria aggregation with transparent normalization and weighting logic, but vendor benchmarking should be role-based; strategic vendors, transactional vendors, local suppliers, and global suppliers should be compared within peer groups to prevent unfair penalties driven by structural context such as lane distance or category perishability. Threshold logic should be retained for critical "hard-fail" indicators in delivery compliance and safety-related quality breaches so that severe failures are not masked by strengths elsewhere.

Vendor performance management should be linked directly to contract governance. Contracts should include explicit, measurable KPI targets aligned with scorecard families, with incentive and penalty structures that reward sustained reliability and rapid recovery behavior. Since forecast variability and product complexity were the strongest destabilizers, retailers should reduce avoidable volatility through tighter demand communication, promotion calendars shared earlier, and collaborative planning for high-complexity assortments. Vendors serving volatile categories should be evaluated with heavier emphasis on flexibility and responsiveness indicators, while stable replenishment categories should prioritize delivery and quality consistency. Portfolio segmentation should be applied routinely to classify vendors into performance archetypes, and each segment should receive differentiated governance: leaders should be prioritized for volume growth and joint planning, flexible

innovators should be integrated into surge and promotion playbooks, low-cost/low-reliability suppliers should be limited to low-risk assortments with tighter monitoring, and chronic underperformers should enter structured corrective-action cycles tied to measurable improvement milestones. Finally, performance review cadence should be synchronized with retail operating rhythms, using monthly tracking for routine categories and event-based reviews for promotions or disruption periods, ensuring that scorecard signals translate into timely managerial action.

LIMITATION

This study had several limitations that should be recognized when interpreting its findings. First, the analysis relied on transactional and operational data extracted from retail information systems, meaning the results were bounded by the accuracy, completeness, and consistency of those systems. Although data screening removed missing, duplicate, and time-misaligned records, some measurement noise likely remained because retail databases often contain latent errors such as delayed scan uploads, incomplete claims coding, or inconsistent vendor master updates. Second, the KPI and scorecard framework captured performance through measurable operational indicators, but it did not fully represent softer relational or strategic factors that may influence vendor outcomes, such as informal collaboration quality, trust, or negotiation power. These unobserved attributes could have contributed to performance variation without being directly modeled. Third, the study treated vendors as stable decision units within the observed window, yet vendor capabilities may shift due to internal restructuring, technology upgrades, labor shocks, or external regulatory changes that were not explicitly captured in the models. Even with panel controls, such time-varying shocks might have influenced performance in ways not separable from the measured drivers. Fourth, although lag structures and intervention designs strengthened temporal ordering, the study remained non-experimental. Retailers allocate volumes, lanes, and contract types partly based on prior vendor performance, so selection effects could not be eliminated completely despite controls and robustness checks. Fifth, performance–outcome linkages were modeled using aggregated vendor–period and store–category outcome alignment rules, which may have introduced attribution blur in cases where multiple vendors jointly supplied the same SKU sets or where replenishment was rebalanced across nodes after receiving. Finally, category and channel moderators were included, but the dataset still represented a specific retail operating environment. Structural patterns such as vendor role mix, omnichannel maturity, and disruption exposure may differ across regions or retail formats, which could limit broad generalization. These limitations indicated that the findings were best interpreted as strong evidence within the observed retail context, while acknowledging that unmeasured relational drivers, residual data noise, and non-experimental structure constrained definitive causal claims.

REFERENCES

- [1]. Abdulla, M., & Md. Jobayer Ibne, S. (2021). Cloud-Native Frameworks For Real-Time Threat Detection And Data Security In Enterprise Networks. *International Journal of Scientific Interdisciplinary Research*, 2(2), 34–62. <https://doi.org/10.63125/0t27av85>
- [2]. Abu Salim, T., Sundarakani, B., & Lasrado, F. (2019). The relationship between TQM practices and organisational innovation outcomes: moderating and mediating the role of slack. *The TQM Journal*, 31(6), 874-907.
- [3]. Agarwal, U. A., & Narayana, S. A. (2020). Impact of relational communication on buyer–supplier relationship satisfaction: role of trust and commitment. *Benchmarking: An international journal*, 27(8), 2459-2496.
- [4]. Aggarwal, R., & Ghosh, A. (2015). Director’s remuneration and correlation on firm’s performance: A study from the Indian corporate. *International Journal of Law and Management*, 57(5), 373-399.
- [5]. Akhtar, M., & Sushil. (2018). Managing strategic performance in a dynamic business environment: A study of two Indian oil companies. *Global Business and Organizational Excellence*, 37(5), 47-62.
- [6]. Alghamdi, A., Karpinski, A. C., Lepp, A., & Barkley, J. (2020). Online and face-to-face classroom multitasking and academic performance: Moderated mediation with self-efficacy for self-regulated learning and gender. *Computers in human behavior*, 102, 214-222.
- [7]. Aliu, O. G., Watts, S., & Kapvotivs, A. (2014). Business interoperability framework for integrated terrestrial-satellite network. 2014 7th Advanced Satellite Multimedia Systems Conference and the 13th Signal Processing for Space Communications Workshop (ASMS/SPSC),
- [8]. Almohri, H., Chinnam, R. B., & Colosimo, M. (2019). Data-driven analytics for benchmarking and optimizing the performance of automotive dealerships. *International Journal of Production Economics*, 213, 69-80.
- [9]. Alvarez-Rodriguez, C., Martín-Gamboa, M., & Iribarren, D. (2019). Sustainability-oriented management of retail stores through the combination of life cycle assessment and dynamic data envelopment analysis. *Science of the Total Environment*, 683, 49-60.

- [10]. Álvarez-Rodríguez, C., Martín-Gamboa, M., & Iribarren, D. (2020). Sensitivity of operational and environmental benchmarks of retail stores to decision-makers' preferences through Data Envelopment Analysis. *Science of the Total Environment*, 718, 137330.
- [11]. Alvarez, A. (2018). Child Welfare Agency Management: Accountability Relationships and Systemic Change. In *Global Encyclopedia of Public Administration, Public Policy, and Governance* (pp. 1-9). Springer.
- [12]. Álvarez, C., Rodríguez, V., Ortega, F., & Villanueva, J. (2015). A Scorecard Framework Proposal for Improving Software Factories' Sustainability: A Case Study of a Spanish Firm in the Financial Sector. *Sustainability*, 7(12), 15999-16021.
- [13]. Ammann, M., & Ehmann, C. (2017). Is governance related to investment performance and asset allocation? Empirical evidence from Swiss pension funds. *Swiss Journal of Economics and Statistics*, 153(3), 293-339.
- [14]. Anand, N., & Grover, N. (2015). Measuring retail supply chain performance: Theoretical model using key performance indicators (KPIs). *Benchmarking: An international journal*, 22(1), 135-166.
- [15]. Antonanzas, F., Juárez-Castelló, C., Lorente, R., & Rodríguez-Ibeas, R. (2019). The use of risk-sharing contracts in healthcare: theoretical and empirical assessments. *Pharmacoeconomics*, 37(12), 1469-1483.
- [16]. Anund Vogel, J., Lind, H., & Lundqvist, P. (2016). Who is governing the commons: Studying Swedish housing cooperatives. *Housing, theory and society*, 33(4), 424-444.
- [17]. Aras, G., Tezcan, N., & Furtuna, O. K. (2018). Multidimensional comprehensive corporate sustainability performance evaluation model: Evidence from an emerging market banking sector. *Journal of Cleaner Production*, 185, 600-609.
- [18]. Ashenbaum, B., & Maltz, A. (2017). Purchasing-logistics integration and supplier performance: An information-processing view. *The International Journal of Logistics Management*, 28(2), 379-397.
- [19]. Basso, A. (2017). The appeal to robustness in measurement practice. *Studies in History and Philosophy of Science Part A*, 65, 57-66.
- [20]. Benton Jr, W., Prahinski, C., & Fan, Y. (2020). The influence of supplier development programs on supplier performance. *International Journal of Production Economics*, 230, 107793.
- [21]. Blackhurst, J., Rungtusanatham, M. J., Scheibe, K., & Ambulkar, S. (2018). Supply chain vulnerability assessment: A network based visualization and clustering analysis approach. *Journal of Purchasing and Supply Management*, 24(1), 21-30.
- [22]. Blasi, S., Caporin, M., & Fontini, F. (2018). A multidimensional analysis of the relationship between corporate social responsibility and firms' economic performance. *Ecological Economics*, 147, 218-229.
- [23]. Brown, D. J., & Fletcher, D. (2017). Effects of psychological and psychosocial interventions on sport performance: A meta-analysis. *Sports Medicine*, 47(1), 77-99.
- [24]. Bryan Jean, R.-J., Sinkovics, R. R., & Kim, D. (2017). Antecedents and outcomes of supplier innovativeness in international customer-supplier relationships: the role of knowledge distance. *Management International Review*, 57(1), 121-151.
- [25]. Buede, D. M., Axelrad, E. T., Brown, D. P., Hudson, D. W., Laskey, K. B., Sticha, P. J., & Thomas, J. L. (2018). Inference enterprise models: An approach to organizational performance improvement. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(6), e1277.
- [26]. Buljac-Samardzic, M., Doekhie, K. D., & van Wijngaarden, J. D. (2020). Interventions to improve team effectiveness within health care: a systematic review of the past decade. *Human resources for health*, 18(1), 2.
- [27]. Chang, W., Ellinger, A. E., Kim, K. K., & Franke, G. R. (2016). Supply chain integration and firm financial performance: A meta-analysis of positional advantage mediation and moderating factors. *European Management Journal*, 34(3), 282-295.
- [28]. Chaudhry, S., Srivastava, B. N., & Joshi, C. (2018). Vendor response to client opportunism in IT service relationships: Exploring the moderating effect of client involvement. *Industrial Marketing Management*, 75, 100-111.
- [29]. Chaudhuri, A., Srivastava, S. K., Srivastava, R. K., & Parveen, Z. (2016). Risk propagation and its impact on performance in food processing supply chain: a fuzzy interpretive structural modeling based approach. *Journal of Modelling in Management*, 11(2), 660-693.
- [30]. Chavez, R., Yu, W., Jacobs, M. A., & Feng, M. (2017). Data-driven supply chains, manufacturing capability and customer satisfaction. *Production Planning & Control*, 28(11-12), 906-918.
- [31]. Cheung, Y. L., Haw, I. M., Hu, B., Swink, M., & Zhang, W. (2020). Common institutional investors and supplier performance in supply chains. *Journal of Operations Management*, 66(6), 670-696.
- [32]. Chu, C.-Y., Park, K., & Kremer, G. E. (2020). A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks. *Advanced Engineering Informatics*, 45, 101053.
- [33]. Chu, S. H., Yang, H., Lee, M., & Park, S. (2017). The impact of institutional pressures on green supply chain management and firm performance: Top management roles and social capital. *Sustainability*, 9(5), 764.
- [34]. Cullen, K. L., Irvin, E., Collie, A., Clay, F., Gensby, U., Jennings, P. A., Hogg-Johnson, S., Kristman, V., Laberge, M., & McKenzie, D. (2018). Effectiveness of workplace interventions in return-to-work for musculoskeletal, pain-related and mental health conditions: an update of the evidence and messages for practitioners. *Journal of occupational rehabilitation*, 28(1), 1-15.
- [35]. Curuksu, J. D. (2018). Data driven. *Management for Professionals*.
- [36]. Dalvi, M. V., & Kant, R. (2018). Effect of supplier development activities on performance outcomes: an empirical study. *Benchmarking: An international journal*, 25(2), 489-516.
- [37]. Davé, A., Ball, P., & Salonitis, K. (2017). Factory eco-efficiency modelling: data granularity and performance indicators. *Procedia manufacturing*, 8, 479-486.

- [38]. De Felice, F., Petrillo, A., & Autorino, C. (2015). Development of a framework for sustainable outsourcing: Analytic Balanced Scorecard Method (A-BSC). *Sustainability*, 7(7), 8399-8419.
- [39]. Dellino, G., Laudadio, T., Mari, R., Mastronardi, N., & Meloni, C. (2018). A reliable decision support system for fresh food supply chain management. *International journal of production research*, 56(4), 1458-1485.
- [40]. Dev, N. K., Shankar, R., Gupta, R., & Dong, J. (2019). Multi-criteria evaluation of real-time key performance indicators of supply chain with consideration of big data architecture. *Computers & Industrial Engineering*, 128, 1076-1087.
- [41]. Dias, A., Vieira, V., & Figlioli, B. (2020). Tracing the links between executive compensation structure and firm performance: evidence from the Brazilian market. *Corporate Governance: The International Journal of Business in Society*, 20(7), 1393-1408.
- [42]. Ding, S., Han, J., Meng, X., & Ma, F. (2020). Multi-granularity modeling and aggregation of design resources in cloud manufacturing. *IEEE Access*, 8, 130797-130819.
- [43]. Domingues, M. L., Reis, V., & Macário, R. (2015). A comprehensive framework for measuring performance in a third-party logistics provider. *Transportation Research Procedia*, 10, 662-672.
- [44]. Dong, W., Ma, Z., & Zhou, X. (2017). Relational governance in supplier-buyer relationships: The mediating effects of boundary spanners' interpersonal guanxi in China's B2B market. *Journal of Business Research*, 78, 332-340.
- [45]. Dreyer, H. C., Kiil, K., Dukovska-Popovska, I., & Kaipia, R. (2018). Proposals for enhancing tactical planning in grocery retailing with S&OP. *International Journal of Physical Distribution & Logistics Management*, 48(2), 114-138.
- [46]. Dweekat, A. J., Hwang, G., & Park, J. (2017). A supply chain performance measurement approach using the internet of things: Toward more practical SCPMS. *Industrial Management & Data Systems*, 117(2), 267-286.
- [47]. El Gibari, S., Gómez, T., & Ruiz, F. (2019). Building composite indicators using multicriteria methods: a review. *Journal of Business Economics*, 89(1), 1-24.
- [48]. Er Kara, M., & Oktay Firat, S. Ü. (2018). Supplier risk assessment based on best-worst method and K-means clustering: a case study. *Sustainability*, 10(4), 1066.
- [49]. Fallahpour, A., Wong, K. Y., Olugu, E. U., & Musa, S. N. (2017). A predictive integrated genetic-based model for supplier evaluation and selection. *International Journal of Fuzzy Systems*, 19(4), 1041-1057.
- [50]. Ferdous Ara, A. (2021). Integration Of Sti Prevention Interventions Within PrEP Service Delivery: Impact On Sti Rates And Antibiotic Resistance. *International Journal of Scientific Interdisciplinary Research*, 2(2), 63-97.
<https://doi.org/10.63125/65143m72>
- [51]. Gandhi, A., & Shankar, R. (2016). Strategic resource management model and data envelopment analysis for benchmarking of Indian retailers. *Benchmarking: An international journal*, 23(2), 286-312.
- [52]. Gawankar, S., Kamble, S., & Raut, R. (2016). Development, measurement and validation of supply chain performance measurement (SCPM) scale in Indian retail sector. *Benchmarking: An international journal*, 23(1), 25-60.
- [53]. Gawankar, S. A., Gunasekaran, A., & Kamble, S. (2020). A study on investments in the big data-driven supply chain, performance measures and organisational performance in Indian retail 4.0 context. *International journal of production research*, 58(5), 1574-1593.
- [54]. Genovese, A., Lenny Koh, S., Kumar, N., & Tripathi, P. K. (2014). Exploring the challenges in implementing supplier environmental performance measurement models: a case study. *Production Planning & Control*, 25(13-14), 1198-1211.
- [55]. Gödri, I., Kardos, C., Pfeiffer, A., & Váncza, J. (2019). Data analytics-based decision support workflow for high-mix low-volume production systems. *CIRP annals*, 68(1), 471-474.
- [56]. Habibullah, S. M., & Md. Foyzal, H. (2021). A Data Driven Cyber Physical Framework For Real Time Production Control Integrating IOT And Lean Principles. *American Journal of Interdisciplinary Studies*, 2(03), 35-70.
<https://doi.org/10.63125/20nhqs87>
- [57]. Hair, J. F., Ringle, C. M., Gudergan, S. P., Fischer, A., Nitzl, C., & Menictas, C. (2019). Partial least squares structural equation modeling-based discrete choice modeling: an illustration in modeling retailer choice. *Business Research*, 12(1), 115-142.
- [58]. Harrington, T. S., Singh Srari, J., Kumar, M., & Wohlrab, J. (2016). Identifying design criteria for urban system 'last-mile' solutions—a multi-stakeholder perspective. *Production Planning & Control*, 27(6), 456-476.
- [59]. Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data—a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management*, 36(10), 1382-1406.
- [60]. Hillerman, T., Souza, J. C. F., Reis, A. C. B., & Carvalho, R. N. (2017). Applying clustering and AHP methods for evaluating suspect healthcare claims. *Journal of computational science*, 19, 97-111.
- [61]. Huang, H., Li, S., & Yu, Y. (2019). Evaluation of the allocation performance in a fashion retail chain using data envelopment analysis. *The journal of the Textile Institute*, 110(6), 901-910.
- [62]. Jusselme, T., Rey, E., & Andersen, M. (2018). An integrative approach for embodied energy: Towards an LCA-based data-driven design method. *Renewable and Sustainable Energy Reviews*, 88, 123-132.
- [63]. Kahraman, C., Uluğ, İ., Othan, C. B., Özkan-Özen, Y. D., & Kazançoğlu, Y. (2018). Efficiency analysis in retail sector: Implementation of data envelopment analysis in a local supermarket chain. *The International Symposium for Production Research*.
- [64]. Kamble, S. S., & Gunasekaran, A. (2020). Big data-driven supply chain performance measurement system: a review and framework for implementation. *International journal of production research*, 58(1), 65-86.
- [65]. Karatzas, A., Johnson, M., & Bastl, M. (2016). Relationship determinants of performance in service triads: a configurational approach. *Journal of Supply Chain Management*, 52(3), 28-47.

- [66]. Kataike, J., Aramyan, L. H., Schmidt, O., Molnár, A., & Gellynck, X. (2019). Measuring chain performance beyond supplier-buyer relationships in agri-food chains. *Supply Chain Management: An International Journal*, 24(4), 484-497.
- [67]. Kirby, R. S., Delmelle, E., & Eberth, J. M. (2017). Advances in spatial epidemiology and geographic information systems. *Annals of epidemiology*, 27(1), 1-9.
- [68]. Kivistö, J., & Zalyevska, I. (2015). Agency theory as a framework for higher education governance. In *The Palgrave international handbook of higher education policy and governance* (pp. 132-151). Springer.
- [69]. Kuharev, J., Navarro, P., Distler, U., Jahn, O., & Tenzer, S. (2015). In-depth evaluation of software tools for data-independent acquisition based label-free quantification. *Proteomics*, 15(18), 3140-3151.
- [70]. Kungl, G., & Geels, F. W. (2018). Sequence and alignment of external pressures in industry destabilisation: Understanding the downfall of incumbent utilities in the German energy transition (1998–2015). *Environmental innovation and societal transitions*, 26, 78-100.
- [71]. Li, J. J., Bonn, M. A., & Ye, B. H. (2019). Hotel employee's artificial intelligence and robotics awareness and its impact on turnover intention: The moderating roles of perceived organizational support and competitive psychological climate. *Tourism management*, 73, 172-181.
- [72]. Li, T., Ma, L., Liu, Z., & Liang, K. (2020). Economic granularity interval in decision tree algorithm standardization from an open innovation perspective: Towards a platform for sustainable matching. *Journal of Open Innovation: Technology, Market, and Complexity*, 6(4), 149.
- [73]. Li, Y., Wang, C.-Z., Li, Y.-c., & Su, J. (2020). Granularity decision of microservice splitting in view of maintainability and its innovation effect in government data sharing. *Discrete Dynamics in Nature and Society*, 2020(1), 1057902.
- [74]. Liebenehm, S. (2018). Temporal stability of risk attitudes and the impact of adverse shocks – a panel data analysis from Thailand and Vietnam. *World Development*, 102, 262-274.
- [75]. Liou, J. J., Chuang, Y.-C., Zavadskas, E. K., & Tzeng, G.-H. (2019). Data-driven hybrid multiple attribute decision-making model for green supplier evaluation and performance improvement. *Journal of Cleaner Production*, 241, 118321.
- [76]. Liu, S., Wang, L., & Huang, W. W. (2017). Effects of process and outcome controls on business process outsourcing performance: Moderating roles of vendor and client capability risks. *European Journal of Operational Research*, 260(3), 1115-1128.
- [77]. Lo, S. M., Zhang, S., Wang, Z., & Zhao, X. (2018). The impact of relationship quality and supplier development on green supply chain integration: A mediation and moderation analysis. *Journal of Cleaner Production*, 202, 524-535.
- [78]. Luzzini, D., Caniato, F., & Spina, G. (2014). Designing vendor evaluation systems: An empirical analysis. *Journal of Purchasing and Supply Management*, 20(2), 113-129.
- [79]. Maestrini, V., Luzzini, D., Caniato, F., Maccarrone, P., & Ronchi, S. (2018). The impact of supplier performance measurement systems on supplier performance: a dyadic lifecycle perspective. *International Journal of Operations & Production Management*, 38(11), 2040-2061.
- [80]. Manikas, A., Boyd, L., Guan, J., & Hoskins, K. (2020). A review of operations management literature: a data-driven approach. *International journal of production research*, 58(5), 1442-1461.
- [81]. Mannering, F. (2018). Temporal instability and the analysis of highway accident data. *Analytic methods in accident research*, 17, 1-13.
- [82]. Margaritis, A., Soenen, H., Franssen, E., Pipintakos, G., Jacobs, G., & Blom, J. (2020). Identification of ageing state clusters of reclaimed asphalt binders using principal component analysis (PCA) and hierarchical cluster analysis (HCA) based on chemo-rheological parameters. *Construction and Building Materials*, 244, 118276.
- [83]. Marshall, D., McCarthy, L., Heavey, C., & McGrath, P. (2015). Environmental and social supply chain management sustainability practices: construct development and measurement. *Production Planning & Control*, 26(8), 673-690.
- [84]. Martín, A. I., Bru, J. M., & Campuzano-Bolarín, F. (2018). Analysis of the Customer Service Rate and Reliability of Suppliers in a Production and Assembly Multinational Company. *Engineering Digital Transformation: Proceedings of the 11th International Conference on Industrial Engineering and Industrial Management*.
- [85]. Mayberry, D., Ash, A., Prestwidge, D., Godde, C. M., Henderson, B., Duncan, A., Blummel, M., Reddy, Y. R., & Herrero, M. (2017). Yield gap analyses to estimate attainable bovine milk yields and evaluate options to increase production in Ethiopia and India. *Agricultural Systems*, 155, 43-51.
- [86]. Mbo, M., & Adjasi, C. (2017). Performance of SOEs: Evidence on Botswana telecommunications corporation. *International Journal of Social Economics*, 44(7), 960-979.
- [87]. McDermott, V., Henne, K., & Hayes, J. (2018). Shifting risk to the frontline: case studies in different contract working environments. *Journal of Risk Research*, 21(12), 1502-1516.
- [88]. McPhail, C., Maier, H. R., Kwakkel, J. H., Giuliani, M., Castelletti, A., & Westra, S. (2018). Robustness metrics: How are they calculated, when should they be used and why do they give different results? *Earth's Future*, 6(2), 169-191.
- [89]. Md Al Amin, K. (2022). Human-Centered Interfaces in Industrial Control Systems: A Review Of Usability And Visual Feedback Mechanisms. *Review of Applied Science and Technology*, 1(04), 66-97.
<https://doi.org/10.63125/gr54qy93>
- [90]. Md Ariful, I. (2022). Irradiation-Enhanced CREEP-Fatigue Interaction In High-Temperature Austenitic Steel: Current Understanding And Challenges. *American Journal of Advanced Technology and Engineering Solutions*, 2(04), 148-181. <https://doi.org/10.63125/e46gja61>
- [91]. Md Nahid, H. (2022). Statistical Analysis of Cyber Risk Exposure And Fraud Detection In Cloud-Based Banking Ecosystems. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 2(1), 289-331.
<https://doi.org/10.63125/9wf91068>

- [92]. Md Sarwar, H. (2021). Sustainable Materials Characterization For Low-Carbon Construction And Infrastructure Durability. *American Journal of Interdisciplinary Studies*, 2(01), 01-34. <https://doi.org/10.63125/wq1wdr64>
- [93]. Md Sarwar Hossain, S., & Md Milon, M. (2022). Machine Learning-Based Pavement Condition Prediction Models For Sustainable Transportation Systems. *American Journal of Interdisciplinary Studies*, 3(01), 31-64. <https://doi.org/10.63125/1jsmkg92>
- [94]. Md. Mominul, H., Masud, R., & Md. Milon, M. (2022). Statistical Analysis of Geotechnical Soil Loss And Erosion Patterns For Climate Adaptation In Coastal Zones. *American Journal of Interdisciplinary Studies*, 3(03), 36-67. <https://doi.org/10.63125/xytn3e23>
- [95]. Md. Musfiqur, R., & Saba, A. (2021). Data-Driven Decision Support in Information Systems: Strategic Applications In Enterprises. *International Journal of Scientific Interdisciplinary Research*, 2(2), 01-33. <https://doi.org/10.63125/cfvq2v45>
- [96]. Md. Rabiul, K., & Sai Praveen, K. (2022). The Influence of Statistical Models For Fraud Detection In Procurement And International Trade Systems. *American Journal of Interdisciplinary Studies*, 3(04), 203-234. <https://doi.org/10.63125/9htnv106>
- [97]. Md. Redwanul, I., Md Nahid, H., & Md. Zahid Hasan, T. (2021). Predictive Analytics in Supply Chain Management A Review Of Business Analyst-Led Optimization Tools. *Review of Applied Science and Technology*, 6(1), 34-73. <https://doi.org/10.63125/5aypx555>
- [98]. Md. Tarek, H., & Sai Praveen, K. (2021). Data Privacy-Aware Machine Learning and Federated Learning: A Framework For Data Security. *American Journal of Interdisciplinary Studies*, 2(03), 01-34. <https://doi.org/10.63125/vj1hem03>
- [99]. Miragliotta, G., Sianesi, A., Convertini, E., & Distanti, R. (2018). Data driven management in Industry 4.0: a method to measure Data Productivity. *IFAC-PapersOnLine*, 51(11), 19-24.
- [100]. Morabito, V. (2015). Big data driven business models. In *Big Data and Analytics: Strategic and Organizational Impacts* (pp. 65-80). Springer.
- [101]. Mortuza, M. M. G., & Rauf, M. A. (2022). Industry 4.0: An Empirical Analysis of Sustainable Business Performance Model Of Bangladeshi Electronic Organisations. *International Journal of Economy and Innovation*. https://gospodarkainnowacje.pl/index.php/issue_view_32/article/view/826
- [102]. Mungra, Y., & Yadav, P. K. (2020). The mediating effect of satisfaction on trust-commitment and relational outcomes in manufacturer-supplier relationship. *Journal of Business & Industrial Marketing*, 35(2), 219-230.
- [103]. Nair, A., Jayaram, J., & Das, A. (2015). Strategic purchasing participation, supplier selection, supplier evaluation and purchasing performance. *International journal of production research*, 53(20), 6263-6278.
- [104]. Neunhoffer, F., & Teubner, T. (2018). Between enthusiasm and refusal: A cluster analysis on consumer types and attitudes towards peer-to-peer sharing. *Journal of Consumer Behaviour*, 17(2), 221-236.
- [105]. Nikhashemi, S., Jebarajakirthy, C., & Nusair, K. (2019). Uncovering the roles of retail brand experience and brand love in the apparel industry: Non-linear structural equation modelling approach. *Journal of Retailing and Consumer Services*, 48, 122-135.
- [106]. Nogueira, A. L., & Munita, C. S. (2020). Quantitative methods of standardization in cluster analysis: finding groups in data. *Journal of Radioanalytical and Nuclear Chemistry*, 325(3), 719-724.
- [107]. Ntim, C. G., Lindop, S., Osei, K. A., & Thomas, D. A. (2015). Executive compensation, corporate governance and corporate performance: A simultaneous equation approach. *Managerial and Decision Economics*, 36(2), 67-96.
- [108]. O'Donohue, W., & Torugsa, N. (2016). The moderating effect of 'Green'HRM on the association between proactive environmental management and financial performance in small firms. *The international journal of human resource management*, 27(2), 239-261.
- [109]. Omar Muhammad, F., & Mst. Shahrin, S. (2021). Comparative Analysis of BI Systems In The U.S. And Europe: Lessons In Data Governance And Predictive Analytics. *Journal of Sustainable Development and Policy*, 1(5), 01-38. <https://doi.org/10.63125/6b3aeg93>
- [110]. Oshri, I., Dibbern, J., Kotlarsky, J., & Krancher, O. (2019). An information processing view on joint vendor performance in multi-sourcing: the role of the guardian. *Journal of Management Information Systems*, 36(4), 1248-1283.
- [111]. Panjehfouladgaran, H., & Lim, S. F. W. (2020). Reverse logistics risk management: identification, clustering and risk mitigation strategies. *Management Decision*, 58(7), 1449-1474.
- [112]. Peng, L., Zhang, W., Wang, X., & Liang, S. (2019). Moderating effects of time pressure on the relationship between perceived value and purchase intention in social E-commerce sales promotion: Considering the impact of product involvement. *Information & Management*, 56(2), 317-328.
- [113]. Pernell, K. (2020). Market governance, financial innovation, and financial instability: lessons from banks' adoption of shareholder value management. *Theory and Society*, 49(2), 277-306.
- [114]. Pungchompoo, S., & Sopadang, A. (2015). Confirmation and evaluation of performance measurement model for the Thai frozen shrimp chain. *Business Process Management Journal*, 21(4), 837-856.
- [115]. Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International journal of production research*, 58(16), 4964-4979.
- [116]. Rakibul, H., & Samia, A. (2022). Information System-Based Decision Support Tools: A Systematic Review Of Strategic Applications In Service-Oriented Enterprises. *Review of Applied Science and Technology*, 1(04), 26-65. <https://doi.org/10.63125/w3ceqv78>
- [117]. Rashidi, K., Noorizadeh, A., Kannan, D., & Cullinane, K. (2020). Applying the triple bottom line in sustainable supplier selection: A meta-review of the state-of-the-art. *Journal of Cleaner Production*, 269, 122001.

- [118]. Reis, M. S. (2018). A systematic framework for assessing the quality of information in data-driven applications for the industry 4.0. *IFAC-PapersOnLine*, 51(18), 43-48.
- [119]. Reza, M., Vorobyova, K., & Rauf, M. (2021). The effect of total rewards system on the performance of employees with a moderating effect of psychological empowerment and the mediation of motivation in the leather industry of Bangladesh. *Engineering Letters*, 29, 1-29.
- [120]. Rezaei, J., Ortt, R., & Trott, P. (2018). Supply chain drivers, partnerships and performance of high-tech SMEs: An empirical study using SEM. *International Journal of Productivity and Performance Management*, 67(4), 629-653.
- [121]. Rhodes, R. E., Janssen, I., Bredin, S. S., Warburton, D. E., & Bauman, A. (2017). Physical activity: Health impact, prevalence, correlates and interventions. *Psychology & health*, 32(8), 942-975.
- [122]. Romule, K., Bak, O., Colicchia, C., & Shaw, S. (2020). Supplier performance assessment: Evidence from a UK-based manufacturing company and its suppliers. *Benchmarking: An international journal*, 27(2), 817-838.
- [123]. Ross, S. A., & Cheah, L. (2019). Uncertainty quantification in life cycle assessments: exploring distribution choice and greater data granularity to characterize product use. *Journal of Industrial Ecology*, 23(2), 335-346.
- [124]. Saikat, S. (2021). Real-Time Fault Detection in Industrial Assets Using Advanced Vibration Dynamics And Stress Analysis Modeling. *American Journal of Interdisciplinary Studies*, 2(04), 39-68. <https://doi.org/10.63125/0h163429>
- [125]. Saikat, S. (2022). CFD-Based Investigation of Heat Transfer Efficiency In Renewable Energy Systems. *International Journal of Scientific Interdisciplinary Research*, 1(01), 129-162. <https://doi.org/10.63125/ttw40456>
- [126]. Salam, M. A., & Khan, S. A. (2018). Achieving supply chain excellence through supplier management: A case study of fast moving consumer goods. *Benchmarking: An international journal*, 25(9), 4084-4102.
- [127]. Segura, M., & Maroto, C. (2017). A multiple criteria supplier segmentation using outranking and value function methods. *Expert Systems with Applications*, 69, 87-100.
- [128]. Selviaridis, K., & Norrman, A. (2015). Performance-based contracting for advanced logistics services: challenges in its adoption, design and management. *International Journal of Physical Distribution & Logistics Management*, 45(6), 592-617.
- [129]. Seyedghorban, Z., Tahernejad, H., Meriton, R., & Graham, G. (2020). Supply chain digitalization: past, present and future. *Production Planning & Control*, 31(2-3), 96-114.
- [130]. Shaikh, S., & Aditya, D. (2021). Federated Learning-Driven Predictive Quality Analytics and Supply Chain Optimization In Distributed Manufacturing Networks. *Review of Applied Science and Technology*, 6(1), 74-107. <https://doi.org/10.63125/k18cbz55>
- [131]. Sheng, S., Zhou, K. Z., Li, J. J., & Guo, Z. (2018). Institutions and opportunism in buyer-supplier exchanges: The moderated mediating effects of contractual and relational governance. *Journal of the Academy of Marketing Science*, 46(6), 1014-1031.
- [132]. Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., & Rohling, E. J. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58(4), e2019RG000678.
- [133]. Silva, A., Vercruyssen, J., Vervaet, C., Remon, J. P., Lopes, J., De Beer, T., & Sarraguça, M. (2018). Process monitoring and evaluation of a continuous pharmaceutical twin-screw granulation and drying process using multivariate data analysis. *European Journal of Pharmaceutics and Biopharmaceutics*, 128, 36-47.
- [134]. Silvi, R., Bartolini, M., Raffoni, A., & Visani, F. (2015). The practice of strategic performance measurement systems: Models, drivers and information effectiveness. *International Journal of Productivity and Performance Management*, 64(2), 194-227.
- [135]. Singh, A., & Teng, J. T. (2016). Enhancing supply chain outcomes through Information Technology and Trust. *Computers in human behavior*, 54, 290-300.
- [136]. Srivastava, S. K., Chaudhuri, A., & Srivastava, R. K. (2015). Propagation of risks and their impact on performance in fresh food retail. *The International Journal of Logistics Management*, 26(3), 568-602.
- [137]. Støre-Valen, M., & Lohne, J. (2016). Analysis of assessment methodologies suitable for building performance. *Facilities*, 34(13/14), 726-747.
- [138]. Sureeyatanapas, P., Sriwattananusart, K., Niyamosoth, T., Sessomboon, W., & Arunyanart, S. (2018). Supplier selection towards uncertain and unavailable information: An extension of TOPSIS method. *Operations Research Perspectives*, 5, 69-79.
- [139]. Tian, X., & Sarkis, J. (2020). Expanding green supply chain performance measurement through emergy accounting and analysis. *International Journal of Production Economics*, 225, 107576.
- [140]. Tjahjono, B., Esplugues, C., Ares, E., & Pelaez, G. (2017). What does industry 4.0 mean to supply chain? *Procedia manufacturing*, 13, 1175-1182.
- [141]. Tonoy Kanti, C., & Shaikat, B. (2022). Graph Neural Networks (GNNS) For Modeling Cyber Attack Patterns And Predicting System Vulnerabilities In Critical Infrastructure. *American Journal of Interdisciplinary Studies*, 3(04), 157-202. <https://doi.org/10.63125/1ykzx350>
- [142]. Tran, A., Nguyen, K.-H., Gray, L., & Comans, T. (2019). A systematic literature review of efficiency measurement in nursing homes. *International journal of environmental research and public health*, 16(12), 2186.
- [143]. Tse, Y. K., Zhang, M., Tan, K. H., Pawar, K., & Fernandes, K. (2019). Managing quality risk in supply chain to drive firm's performance: the roles of control mechanisms. *Journal of Business Research*, 97, 291-303.
- [144]. Vahdani, B., Mousavi, S. M., Tavakkoli-Moghaddam, R., & Hashemi, H. (2017). A new enhanced support vector model based on general variable neighborhood search algorithm for supplier performance evaluation: A case study. *International Journal of Computational Intelligence Systems*, 10(1), 293-311.

- [145]. Vallandingham, L. R., Yu, Q., Sharma, N., Strandhagen, J. W., & Strandhagen, J. O. (2018). Grocery retail supply chain planning and control: Impact of consumer trends and enabling technologies. *IFAC-PapersOnLine*, 51(11), 612-617.
- [146]. Vlachos, I. P. (2014). A hierarchical model of the impact of RFID practices on retail supply chain performance. *Expert Systems with Applications*, 41(1), 5-15.
- [147]. Volpi, A., Rizzi, A., & Rinaldi, R. (2017). Dual frequency tag performances in the fashion industry. Workshop on Business Models and ICT Technologies for the Fashion Supply Chain,
- [148]. Wagner, J., & Benoit, S. (2015). Creating value in retail buyer-vendor relationships: A service-centered model. *Industrial Marketing Management*, 44, 166-179.
- [149]. Wang, M.-M., & Wang, J.-J. (2019). How vendor capabilities impact IT outsourcing performance: An investigation of moderated mediation model. *Journal of Enterprise Information Management*, 32(2), 325-344.
- [150]. Woo, C., Kim, M. G., Chung, Y., & Rho, J. J. (2016). Suppliers' communication capability and external green integration for green and financial performance in Korean construction industry. *Journal of Cleaner Production*, 112, 483-493.
- [151]. Yadav, S., & Singh, S. P. (2020). Blockchain critical success factors for sustainable supply chain. *Resources, Conservation and Recycling*, 152, 104505.
- [152]. Yang, C.-M., Chen, K.-S., Hsu, T.-H., & Hsu, C.-H. (2019). Supplier selection and performance evaluation for high-voltage power film capacitors in a fuzzy environment. *Applied Sciences*, 9(23), 5253.
- [153]. Yang, F., & Zhang, X. (2017). The impact of sustainable supplier management practices on buyer-supplier performance: An empirical study in China. *Review of International Business and Strategy*, 27(1), 112-132.
- [154]. Yu, W., Chavez, R., Jacobs, M. A., & Feng, M. (2018). Data-driven supply chain capabilities and performance: A resource-based view. *Transportation Research Part E: logistics and transportation review*, 114, 371-385.
- [155]. Yuan, J., Zhang, Z. M., Yüksel, S., & Dinçer, H. (2020). Evaluating cognitive balanced scorecard-based quality improvement strategies of energy investments with the integrated hesitant 2-tuple interval-valued Pythagorean fuzzy decision-making approach to QFD. *IEEE Access*, 8, 171112-171128.
- [156]. Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International journal of hospitality management*, 76, 111-121.
- [157]. Zheng, B., Wang, Y., Kamal, M. A., & Ullah, A. (2020). The influence of cultural and institutional distance on China's OFDI efficiency: fresh evidence from stochastic frontier gravity model. *International Journal of Emerging Markets*, 17(1), 98-119.
- [158]. Zhou, X., & Xu, Z. (2018). An integrated sustainable supplier selection approach based on hybrid information aggregation. *Sustainability*, 10(7), 2543.
- [159]. Zhu, X., & Freeman, M. A. (2019). An evaluation of US municipal open data portals: A user interaction framework. *Journal of the Association for Information Science and Technology*, 70(1), 27-37.