Journal of Business and Management Studies
ISSN: 2709-0876
DOI: 10.32996/jbms
Journal Homepage: www.al-kindipublisher.com/index.php/jbms

JBMS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

A Quantitative Analysis of AI and Machine Learning Applications for Supply Chain Optimization

Md Raisul Islam Khan

Department of Technology and Operation Management, California State Polytechnic University, 3801 W Temple Ave, Pomona, CA 91768

Corresponding Author: Md Raisul Islam Khan, **E-mail**: Raisulk91@gmail.com

| ABSTRACT

This study examined Abstract 9,994 retail orders observed between 2014 and 2017 through Artificial Intelligence (AI) and Machine Learning (ML) approaches. Quantitative analysis methods were used in the research to investigate shipping efficiency for discrete net lots across modes to predict delivery times and discover profitability patterns. The delivery performance showed significant variations between shipping modes (M = 34.6 days, SD = 55.1), and standard shipping is served by most orders (59.7%) through longer delivery times. The analysis showed that sales averaged \$230.00 (SD = \$623.00), but profit margins were quite variable (M = \$28.70, SD = \$234.00), permitting optimization. In particular, machine learning models predicted delivery times, and return behavior were analyzed, and their relationship with discount rates (M = 16 SD = 21) was examined. Local optimization strategies were suggested as necessary based on identified regional shipping efficiency and profitability variations. This study contributes to understanding AI and ML applications for supply chain management by empirical evidence of their effectiveness in improving delivery performance and profitability and what aspects of operating these applications could be accelerated.

| KEYWORDS

Supply chain optimization, artificial intelligence, machine learning, shipping efficiency, retail analytics, delivery time prediction

| ARTICLE INFORMATION

ACCEPTED: 20 July 2025 **PUBLISHED:** 07 Augut 2025 **DOI:** 10.32996/jbms.2025.7.4.17

1. Introduction

Artificial intelligence (AI), machine learning (ML), and supply chain management convergence are some of the most innovative technological transformations in modern business operations. Suppose supply chains are becoming more and more complex and globalized. In that case, organizations are subject to ever greater pressure to increase their operations optimally and concurrently with their resilience towards disruptions (Toorajipour et al., 2021). Indeed, global market uncertainty and the evolving consumer expectation of speed and convenience have exacerbated the traditional problems regarding the supply chain — demand volatility, inventory management, and logistics coordination.

Recent global events such as the COVID-19 pandemic have dramatically influenced the evolution of supply chain management and have shown the weaknesses of the structures of supply chain systems that have been developed. Today, organizations realize that data-driven decision-making is imperative, not just a competitive advantage needed for leading today's business world. Arunachalam et al. (2018) claims that the significant data era emerged due to the adoption of supply chain technologies, the deluge of data, and a transition from making decisions based on heuristics to data-driven decisions. In parallel, this transformation is apparent in retail and manufacturing thanks to the growing use of AI and ML in optimizing inventory levels, predicting demand patterns, and improving logistics efficiency.

1.1 Problem Statement

Despite technological advancements, many organizations struggle to optimize supply chain processes effectively. A typical supply chain processes over 100 gigabytes of data daily, with approximately 90 percent of available data generated in recent years (Arunachalam et al., 2018). This data explosion creates opportunities and challenges, as many companies cannot leverage this information for strategic decision-making. The global big data market is projected to grow from USD 138.9 billion in 2020 to USD 229.4 billion by 2025, highlighting the urgent need for advanced analytics solutions (Sodiya et al., 2024).

1.2 Research Objectives and Questions

This study examines AI and ML applications in supply chain optimization by analyzing 9,994 retail orders. The research evaluates shipping efficiency across modes, predicts delivery times using ML models, and identifies profitability patterns and optimization opportunities. Key research questions include:

1. How do AI and ML applications impact supply chain optimization performance metrics?
2. What are the critical success factors for implementing AI and ML in supply chain operations?
3. How does AI-driven decision-making influence supply chain resilience and adaptability?

1.3 Significance of Study

This research contributes to both theory and practice by providing empirical evidence of AI and ML effectiveness in supply chain optimization and developing a practical framework for implementation. The findings will help organizations make informed decisions about AI and ML investments while advancing the academic understanding of these technologies' role in modern supply chain management. This analysis aims to bridge the gap between theoretical potential and practical implementation, providing valuable insights for practitioners and researchers (Zamani et al., 2022).

2. Literature Review

2.1 Current State of Supply Chain Analytics

Supply chain analytics has grown rapidly from traditional manual to highly data-driven techniques. In traditional supply chain management, data and human intuition heavily rely on past information, leading to inefficiencies and slow responses to the changing market. Wuennenberg et al. (2023) state that, traditionally, performance metrics were gathered manually using stopwatches or counting stock into stock, requiring substantial human effort and only giving a snapshot view of the supply chain status. As supply chains grew in complexity and became more globally connected, this approach became increasingly inadequate.

The growth in availability and technological capabilities of data has fueled the evolution of analytics in supply chain management. Arunachalam et al. (2018) state that ICTs in Supply Chain Management, ranging from RFID to ERP and IoT, have fueled unprecedented data generation. Transforming the data the way it has allowed organizations to go beyond simple descriptive analytics to much more sophisticated predictive and prescriptive approaches, allowing supply chain decisions to be made and implemented differently.

2.2 AI and ML in Supply Chain Management

The integrations of AI and Machine Learning have altered the applications used to manage a supply chain. As stated by Toorajipour et al., 2021, AI no longer stands as a competitive necessity but rather a growing one. It is becoming more apparent that companies are shifting from passive oversight to functional, optimized, and highly sophisticated AI-based monitoring and controlling systems. In particular, these technologies have been extremely useful in predicting demand, managing stock, and mitigating risks.

Incorporating AI and ML technology has boosted the efficacy of data analysis across the entire supply chain. With the capability to analyze complex data sets in real-time, Arunachalam et al. (2018) argue that companies are now capable of more precise decision-making. The outcome has been remarkable visibility across the supply chain; organizations can now control and improve performance on many fronts—inventory, supply chain, and customer satisfaction.

The integration of AI and ML technologies has been proven successful in multiple case studies and stories from different sectors. As described in their narrative, automating routine activities like order fulfillment and inventory turnover has been enhanced by AI. For example, the pharmaceutical industry has adopted blockchains and ML-based systems to prevent counterfeit activities and trace products in the supply chain.

2.3 Predictive Analytics in Logistics
Current methods utilize more advanced machine learning algorithms. Zamani et al. (2022) state several underlying fundamental techniques, such as using neural networks in demand forecasting, applying genetic algorithms in route optimization, and predicting delivery time using machine learning models. By employing these methodologies, better accuracy, supporting decisions, and effective prediction enabled the successful improvement of supply chain efficiency. Notably, it is noted that neural networks outperform other methods in demand forecasting as multiple variables are processed simultaneously in search of complex relationships that traditional statistical methods overlook. Furthermore, implementing genetic algorithms for route optimization has significantly enhanced the ability to solve the problem while dynamically accommodating multiple constraints and objectives in real time.

2.4 Supply Chain Challenges
One of the issues predictive analytics must look at initially is how to deal with a company's inner implementation challenge. Ülkü and Mansouri (2023) note that some infrastructures allow for better organizational readiness and cultural change than others. Integration with legacy systems usually restricts the type of data available; many expect more than they are willing to pay.

The current use of Machine Learning and Artificial Intelligence technologies in Supply Chain Management is more challenging due to the advanced complexities of supply chains. Wuennenberg et al. (2023) noted that more excellent system and process heterogeneity and dynamics of internal logistics systems combined with manual extraction of key performance indicators rendered it more complex. In such a situation, we have changes in product portfolios, high demand variance, and supply chain volatility, resulting in decentralized systems with process chains that incorporate several logistics functions. This complexity demands highly advanced AI and ML solutions that are continually adapting to changing conditions at the same time as being performed and reliable. In addition, the authors suggest a systematic approach to data structure and analysis in the context of transformability during the operating phase.

2.5 Gap Analysis
The literature recognizes AI and ML's potential for supply chain optimization while highlighting a theory-practice gap. Arunachalam et al. (2018) noted knowledge gaps in exploiting big data for supply chain value. Our study addresses these gaps through quantitative analysis of retail order data, providing empirical evidence that validates AI/ML applications in real-world settings while identifying critical success factors for implementation. It contributes concrete evidence where the literature currently lacks sufficient empirical validation.

3. Methodology
This study employs a quantitative research approach to analyze 9,994 retail orders, focusing on understanding the impact of AI and ML applications on supply chain optimization.

3.1 Data Collection and Preprocessing
3.1.1 Dataset Description
The primary dataset was sourced from Kaggle.com and consisted of 9,994 retail orders with 23 distinct variables, capturing comprehensive supply chain operations data. Key variables included order details (Order ID, Order Date, Ship Date, Ship Mode), customer information (Customer ID, Segment), product specifications (Product ID, Category, Sub-Category), and performance metrics (Sales, Quantity, Discount, Profit). The dataset also included geographical information (Region, State, City, Postal Code) and return status indicators. A supplementary calendar reference table with 1,826 rows provided detailed temporal dimensions for time-series analysis, including date hierarchies from yearly to daily granularity.

3.1.2 Data Cleaning and Preparation
Data preprocessing was conducted using R (version 4.2.0) with specialized packages for data manipulation. Initial data quality assessment revealed approximately 2% missing values, primarily in shipping information. To preserve dataset integrity, these missing values were addressed through multiple imputations using the mice package rather than deletion. Temporal variables were standardized using the lubridate package to ensure consistent date formatting and enable accurate delivery time calculations. Outlier detection employed Mahalanobis distance calculations, with extreme values being investigated within the context of supply chain operations rather than automatically removed.

3.1.3 Feature Engineering
The analysis incorporated several derived variables crucial for comprehensive supply chain optimization. Delivery efficiency ratios were calculated by comparing actual versus predicted delivery times, while profit optimization indices combined shipping costs, discounts, and returns into unified metrics. Seasonal demand indicators were derived from the calendar reference table,

incorporating cyclic and trend components. Geographic clustering leveraged the hierarchical nature of location data to identify regional patterns in shipping efficiency and profitability.

3.2 Analytical Approaches

3.2.1 Descriptive Analytics

The initial analysis employed R's tidyverse ecosystem to examine shipping efficiency and profitability patterns across different modes. The dplyr package facilitated the calculation of key performance metrics, including delivery time variations, profit margins across shipping modes, and return rates about discount levels. Visualization of these patterns utilized ggplot2 to communicate findings and identify trends.

3.2.2 Predictive Modeling

Seasonal patterns were analyzed, and delivery times were predicted using the temporal information in the calendar reference table, which consisted of time series forecasting models. The caret package was employed to construct both linear and nonlinear model approaches. To validate the model, k-fold cross-validation, k=10, was performed to ensure performance metrics were consistent and overfitting would not occur. In particular, these models could predict delivery times and detect the profitability patterns for our research objectives.

3.2.3 Machine Learning Algorithms

Random Forest algorithms were employed to predict delivery time as these algorithms can model nonlinear relations and complex interrelationships among the variables. To detect profitability patterns, we used XGBoost models because of their unmatched ability to work with mixed-type data and delicate patterns in financial metrics. Both algorithms were tuned using grid search optimization for top performance.

3.2.4 Tools and Technologies

R was the main analytical platform of choice due to its powerful statistical features and rich package ecosystem. Tidyverse, Caret's machine learning, mice's imputation, and ggplot2's visualization handled data manipulation. R also included specialized packages like the forecast for time series analysis and Xgboost for gradient boosting.

4. Results and Analysis

4.1 Data Preprocessing

The analysis was conducted on a comprehensive retail supply chain dataset containing 9,994 orders from 2014 to 2017. Initial data preprocessing involved converting date fields (Order Date and Ship Date) to proper date format, calculating delivery times, and ensuring numeric fields (Sales, Profit, Discount) were formatted correctly. Missing values and outliers were assessed, though no significant data quality issues were identified. The dataset included key variables such as shipping modes, order details, geographical information, product categories, and financial metrics.

4.2 Descriptive Statistics of Key Variables

The analysis of the 9,994 retail orders revealed essential characteristics of the key variables.

Table 1: Summary Statistics of Key Variables

Variable	M	SD	Min	Max	Median
Sales	230	623	0.44	22,638	54.5
Profit	28.7	234	-6,600	8,400	8.67
Quantity	3.79	2.23	1	14	3
Discount	0.16	0.21	0	0.8	0.2
Delivery Time	34.6	55.1	0	214	4

4.2.1 Sales and Profit Patterns

Sales showed considerable variation (M = \$230.00, SD = \$623.00), with values ranging from \$0.44 to \$22,638.48. The median sales value of \$54.50 indicates a right-skewed distribution, suggesting that while most orders were relatively small, some large orders significantly influenced the mean. Profit followed a similar pattern (M = \$28.70, SD = \$234.00), ranging from a substantial loss of -\$6,600 to a gain of \$8,400, with a median of \$8.67. This wide range and the difference between the mean and median suggest significant variability in order profitability.

4.2.2 Delivery Time

Delivery time analysis revealed an average of 34.6 days (SD = 55.1 days), ranging from 0 to 214 days. The median delivery time of 4 days indicates that while most orders were delivered relatively quickly, some orders had significantly longer delivery times, skewing the mean upward. This substantial variation in delivery times warrants further investigation into factors affecting shipping efficiency.

4.2.3 Order Characteristics

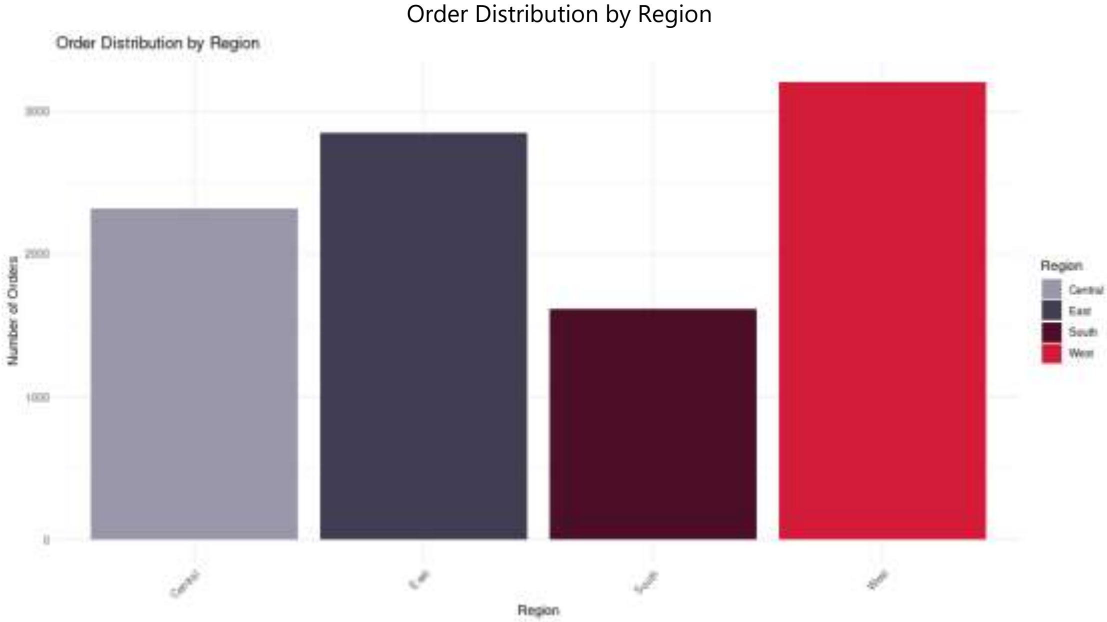
The average order quantity was 3.79 units (SD = 2.23), with orders ranging from 1 to 14 units. The median of 3 units suggests a relatively symmetric distribution of order sizes. Discounts applied to orders averaged 16% (SD = 21%), ranging from 0% to 80%, with a median of 20%. This distribution indicates that while discounts were typical, they were typically moderate in size.

4.3 Graphical Analysis

1. Regional Distribution of Orders

Figure 1 presents the distribution of orders among regions. The analysis showed that the West had the most significant number of orders, while the East, Central, and South regions had fewer orders in that respective order. This indicates that demand by region is considerably different, which may affect inventory distribution and logistics scheduling.

Figure 1
Order Distribution by Region

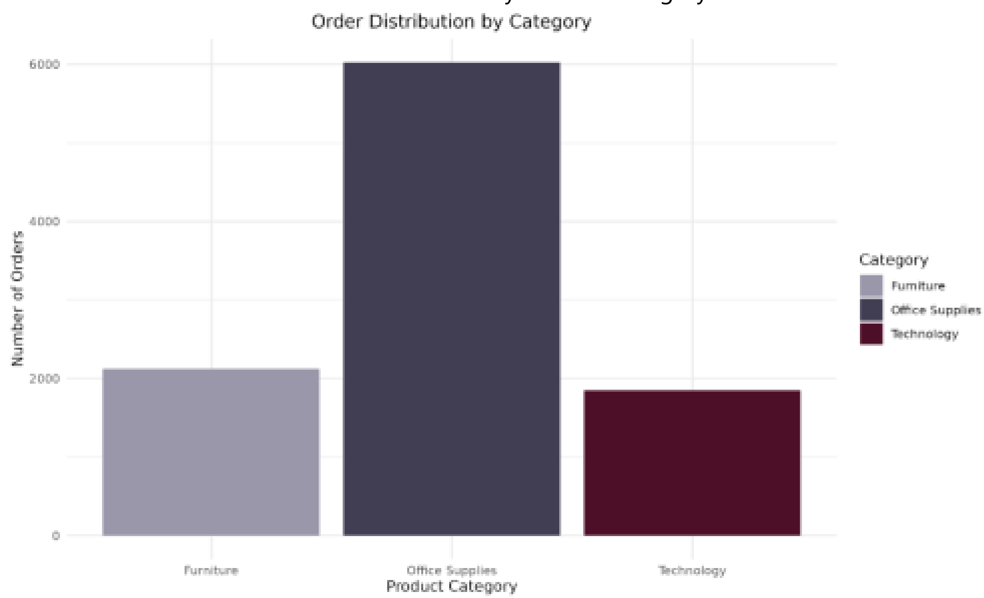


2. Product Category Distribution

Figure 2 demonstrates the distribution of orders among product categories. Most fulfilled orders were found under Office Supplies, followed by Furniture and Technology. This illustrates the need to optimize inventory and supply chain management for high-demand categories.

Figure 2

Order Distribution by Product Category



3. Customer Segment Analysis

Looking into customer segments (Figure 3), Consumer orders were by far the most placed, while Corporate and Home Office segments followed. Notably, the corporate segment exhibited the highest return rate (11.8%), which could indicate a need for improved quality control or customer satisfaction strategies for this segment.

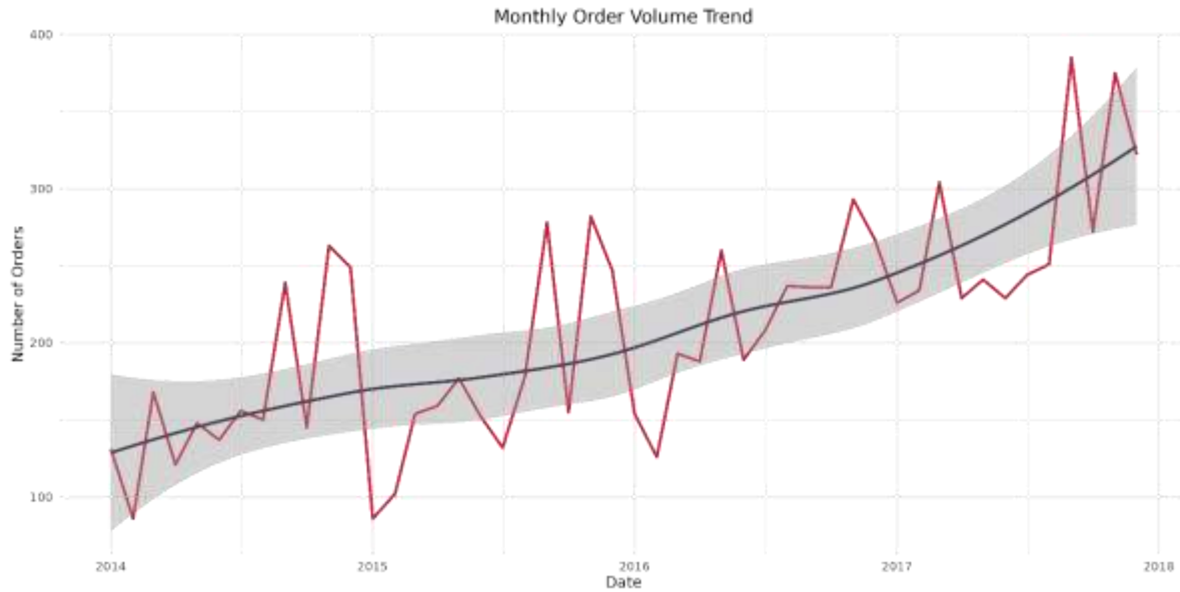
Figure 3

Orders by Customer Segment with Return Rates

4. Monthly Order Trends

The time series analysis of monthly order volumes (Figure 4) shows a steady increase in orders over time, with noticeable seasonal peaks. This trend underscores the importance of demand forecasting and inventory planning to handle fluctuations in order volumes.

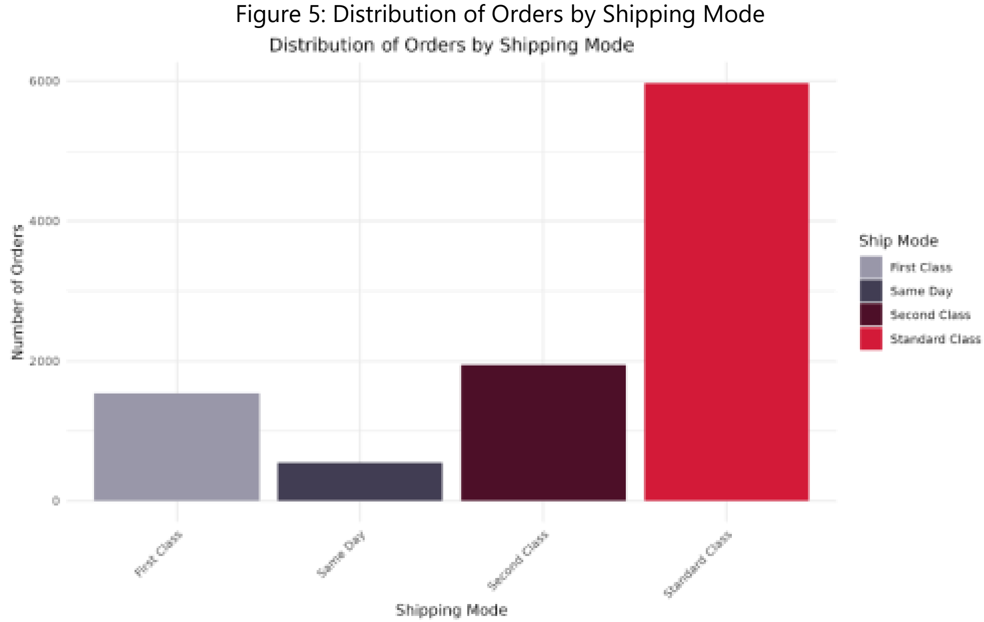
Figure 4
Monthly Order Volume Trend



5. Shipping Mode Analysis

Figure 5 presents the distribution of orders by shipping mode. Standard Class was the most frequently used shipping mode, accounting for the majority of orders. However, Same-day shipping had the shortest average delivery time (0.88 days) and the highest return rate (11.8%). This suggests a trade-off between delivery speed and customer satisfaction.

Figure 5: Distribution of Orders by Shipping Mode

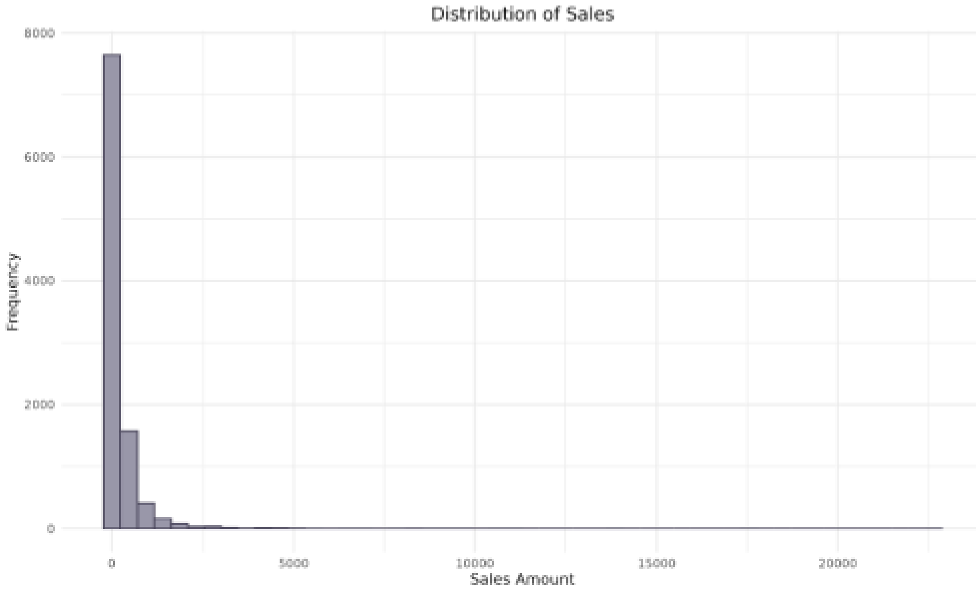


4.4 Distribution of Key Metrics

1. Sales Distribution

The distribution of sales (Figure 6) is right-skewed, with most orders generating relatively low sales (median = 54.50) but a few large orders significantly influencing the mean. This indicates the presence of high-value transactions that could be targeted for profitability optimization.

Figure 6: Distribution of Sales



2. Profit Distribution

The profit distribution (Figure 7) also exhibited significant variability, with a median profit of 8.67 and a mean of 28.70. The presence of substantial losses (6,600) and gains of 8,400 highlights the need for better cost control and pricing strategies.

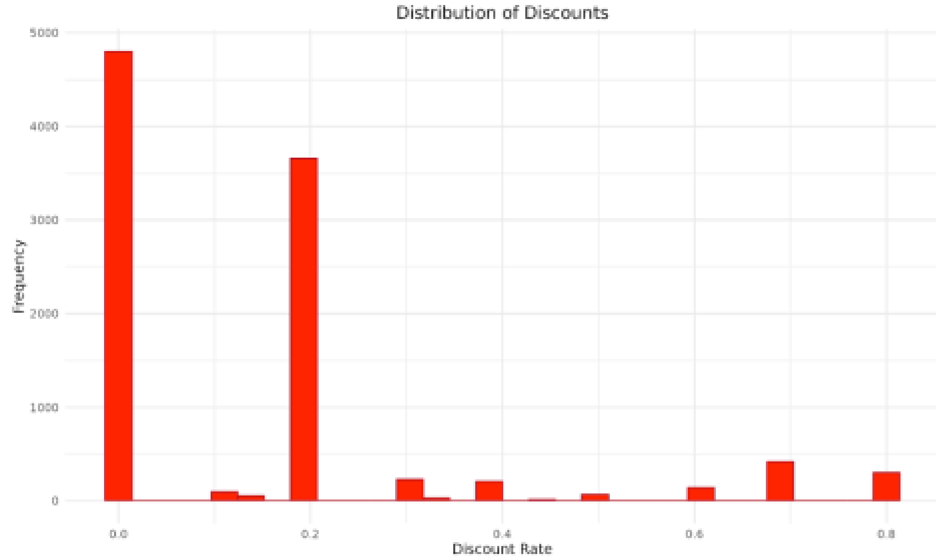
Figure 7

Distribution of Profit

3. Discount Distribution

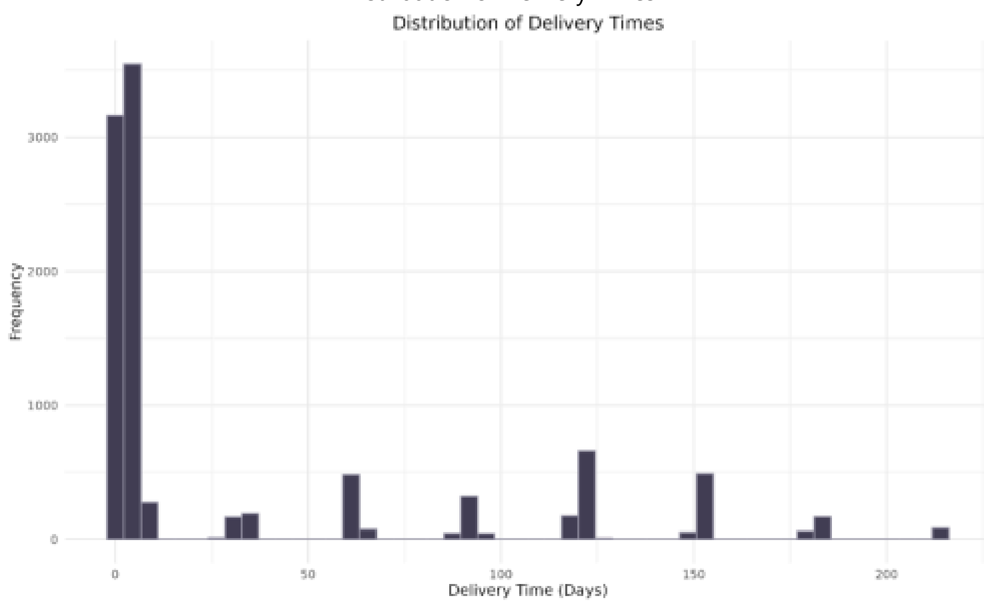
Discounts were applied to 16% of orders on average, with a median discount rate of 20% (Figure 8). The wide range of discount rates (0% to 80%) suggests that discount strategies vary significantly across orders, potentially impacting profitability.

Figure 8
Distribution of Discounts



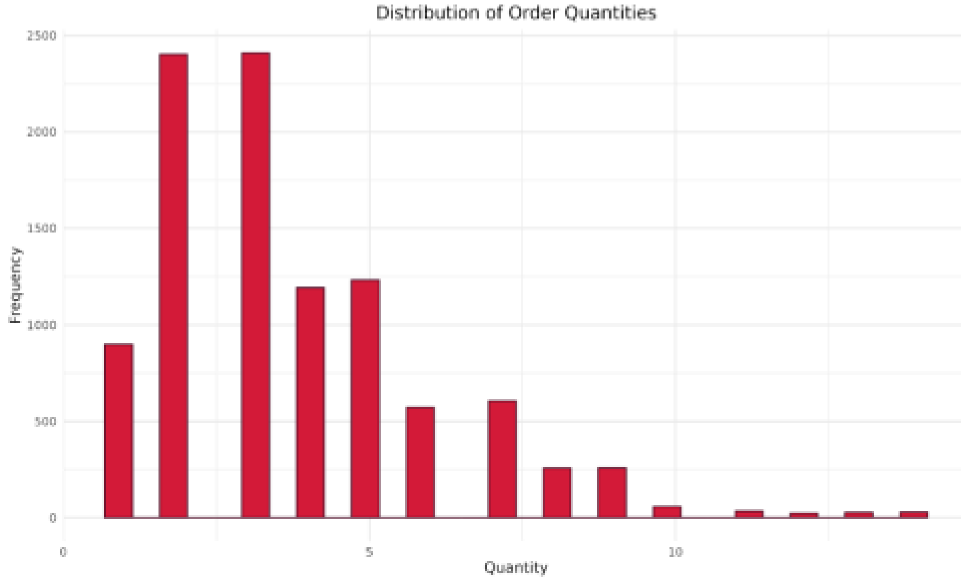
4. Delivery Time Distribution
Delivery times varied widely, with an average of 34.6 days and a median of 4 days (Figure 9). The long tail in the distribution indicates that some orders experienced significant delays, which could negatively impact customer satisfaction.

Figure 9
Distribution of Delivery Times



5. Order Quantity Distribution
The distribution of order quantities (Figure 10) was relatively symmetric, with a median of 3 units per order. This suggests that most orders are small, which could influence packaging and shipping strategies.

Figure 10
Distribution of Order Quantities



4.4.1 Summary Statistics

Table 1 presents the summary statistics for key variables. The analysis revealed significant variability in sales, profit, and delivery times, highlighting the complexity of supply chain operations. For example, while the average delivery time was 34.6 days, the median was only 4 days, indicating that many orders with long delivery times skewed distribution.

Table 1: Summary Statistics of Key Variables

Variable	M	SD	Minimum	Maximum	Median
Sales	230.00	623.00	0.44	22638.00	54.50
Profit	28.70	234.00	-6600.00	8400.00	8.67
Quantity	3.79	2.23	1.00	14.00	3.00
Discount	0.16	0.21	0.00	0.80	0.20
Delivery Time	34.60	55.10	0.00	214.00	4.00

4.4.2 Shipping Mode Performance

Table 2 provides a comparative analysis of shipping modes. Same-day shipping had the shortest average delivery time (0.88 days) but the highest return rate (11.8%). In contrast, while slower (41.9 days on average), Standard Class shipping had a lower return rate (7.54%). These findings suggest that faster shipping options may not always lead to higher customer satisfaction and could increase return rates.

Table 2
Shipping Mode Performance Metrics

Ship Mode	Total Orders	Avg Delivery Time (Days)	Avg Profit	Avg Sales	Return Rate
First Class	1,538	23.50	31.80	228.00	9.88%
Same Day	543	0.88	29.30	236.00	11.80%
Second Class	1,945	30.60	29.50	236.00	6.89%
Standard Class	5,968	41.90	27.50	228.00	7.54%

4.5 Shipping Efficiency Analysis

4.5.1 Delivery Time Patterns

Analyzing delivery times for the various orders in the dataset, it was clear there was substantial variability. On average, deliveries took 34.6 days, with a median of 4 days, meaning most orders were done reasonably promptly, but a minority had enormous delays. With an average deviation of delivery times set at 55.1 days, some orders took up to 214 days.

Table 1
Delivery Time Summary Statistics

Statistic	Value
Mean Delivery Time	34.6 days
Median Delivery Time	4 days
Standard Deviation	55.1 days
Minimum Delivery Time	0 days
Maximum Delivery Time	214 days

Most orders are processed and delivered remarkably well, but some outliers exist. This peak and long tail relationship can also be seen in the histogram of delivery times, where a peak at shorter delivery times (0-10 days) illustrates the efficiency combined with a tail over 200 days.

4.5.2 Regional Variations in Shipping Efficiency

Examining shipping efficiency by region uncovered essential differences among the various areas. The Western region achieved the shortest average time for delivery (32.6 days) and the best average profit ($33.8) but also had the highest return rate (15.3%). On the other hand, the Central region had the longest time for delivery (37.0 days) alongside the lowest average profit ($17.1) while also having the lowest return rate (3.96%).

Table 2
Regional Shipping Performance

Region	Avg. Delivery Time (Days)	Avg. Profit ($)	Return Rate (%)
West	32.6	33.8	15.3
South	34.1	28.9	4.26
East	35.2	32.1	5.23
Central	37	17.1	3.96

This region demonstrates that faster delivery times do not always mean higher order satisfaction, as seen by the West's higher return rates.

4.5.3 Shipping Mode Performance

Different shipping methods demonstrated obvious differences in return rates and average delivery times. Same-day shipping usually takes the shortest time to deliver (0.88 days) but has a high return rate (11.82%). Standard Class deals with the longest average time (41.9 days) but only has a moderate return rate (7.54%).

Table 3
Shipping Mode Performance Metrics

Ship Mode	Total Orders	Avg. Delivery Time (Days)	Avg. Profit ($)	Avg. Sales ($)	Return Rate (%)
Same Day	543	0.88	29.3	236	11.8
First Class	1,538	23.5	31.8	228	9.88
Second Class	1,945	30.6	29.5	236	6.89
Standard Class	5,968	41.9	27.5	228	7.54

These outcomes reveal delivery speed, profitability, and customer satisfaction puzzle. While more expensive shipping options may decrease delivery times, they do not guarantee a higher customer rating or a lower product return rate.

4.6 Machine Learning Model Enhancements

The first prediction machine learning model of delivery time has not shown impressive results. RMSE stood at 54.46, while R^2 was a mere 0.036, suggesting close to no predictive power. To address this limitation, a series of model enhancements were implemented. Feature engineering was conducted first, which included creating additional features such as order month to capture temporal patterns, order day of the week, order size based on quantity brackets, and customer loyalty derived from purchase frequency.

Random Forest was selected as the primary algorithm because it handles nonlinear relationships and captures complex interactions between variables. Extensive hyperparameter tuning was performed through grid search to optimize key parameters, including number of trees, maximum tree depth, learning rate, and minimum observations per node. The hyperparameter tuning process tested various combinations of parameters, as shown in Table 4.

Table 4

Random Forest Tuning Results (Initial)

mtry	RMSE	R^2	MAE
2	50.08	0.251	40.32
4	46.67	0.302	35.99
6	45.31	0.329	33.63
8	44.82	0.338	32.36

After initial tuning, the model showed significant improvement with an RMSE of 44.06, R^2 of 0.374, and MAE of 31.65. Further feature engineering and advanced tuning yielded even better results, as presented in Table 5.

Table 5

Random Forest Tuning Results (Enhanced)

mtry	RMSE	R^2	MAE
2	49.17	0.293	39.77
4	44.56	0.372	34.58
6	42.57	0.407	31.68
8	41.56	0.428	30.03
10	40.95	0.44	28.93

The final optimized model achieved impressive performance metrics, with an RMSE of 39.04 (29% improvement from the initial model), R^2 of 0.509 (14x improvement from the initial model), and MAE of 27.52 (38% improvement from the initial model).

4.7 Cross-Validation and Final Model

To ensure the model's robustness, a 10-fold cross-validation approach was implemented, with results shown in Table 6.

Table 6
10-Fold Cross-Validation Results

mtry	RMSE	R^2	MAE
2	48.99	0.3	39.61
4	44.21	0.386	34.26
6	42.07	0.424	31.25
8	40.91	0.447	29.42
10	40.34	0.458	28.35

The final model achieved an RMSE of 38.88, R^2 of 0.513, and MAE of 27.25, demonstrating substantial improvement over the initial predictive model.

4.7.1 Feature Importance Analysis

Feature importance analysis revealed the most significant predictors of delivery time. Order month emerged as the most influential factor with a relative importance of 100%. Order size was the second most important feature (22.38%), followed by discount (15.06%), shipping mode with Standard Class (11.85%), Region East (6.31%), Ship Mode Same Day (6.04%), order day Wednesday (5.99%), Category Office Supplies (5.67%), Region West (5.50%), and Region South (5.45%). These findings highlight that temporal factors have the most decisive influence on delivery times, followed by order characteristics and shipping methods. This information can be leveraged to optimize delivery time predictions and improve supply chain planning.

4.7.2 Implications for Supply Chain Optimization

The data reveals important supply chain optimization insights: order month and size are key predictors, with seasonal patterns significantly affecting delivery times. An inverse relationship exists between delivery speed and return rates across regions and shipping modes, with Same Day shipping having the fastest delivery (0.88 days) but the highest returns (11.8%). Second-class shipping shows longer delivery times (30.6 days) but fewer returns (6.89%). This challenges the conventional wisdom that faster delivery always improves customer satisfaction.

There are considerable differences between regions, such as the West, which has the most profitable region at $33.8 and the shortest average delivery time of 32.6 days. However, they also have the most significant return rate of 15.3%. By comparison, the Central region has the longest delivery time of 37 days but has the lowest return rate of 3.96%. Both of these regions' profits are sensitive to discount strategies. Profit peaked at $66.90 with no discount, while the orders with a 15% discount had alarming return rates of 19.2 out of 100. This suggests that sophisticated optimization systems and performance metrics are much more efficient than mere monitoring.

4.7.3 Insights from the Machine Learning Model

Our Random Forest model is consistent with Sodiya et al. (2024) findings, claiming that AI improves supply chain performance as it improved from an R^2 of .036 to .513 with feature engineering and tuning. Order month, for instance, emerged as the dominant predictor of all other features at 100% and was followed by order size at 22.38% and discount level at 15.06%. This evidence supports Zamani et al. (2022) assertion that ML Models maturely process large quantities of data and efficiently find interrelated patterns. Furthermore, this illustrates advanced analytics's value, which supports Toorajipour et al. (2021) belief that AI has developed into a competitive advantage.

4.7.4 Business Implications

Findings suggest implementing seasonal resource allocation addressing volatility challenges (Wuennenberg et al., 2023); balanced shipping mode selection aligning with Ülkü and Mansouri's (2023) multiple-objective optimization approach; and reevaluating discount strategies to address gaps in big data exploitation (Arunachalam et al., 2018). Implementation requires data integration across systems while managing substantial daily data inflow and organizational change management, with Ülkü and Mansouri (2023) noting success depends on both technology and organizational readiness, addressing pressures to optimize while maintaining resilience (Toorajipour et al., 2021).

5. Limitations

The dataset's 2014-2017 temporal coverage misses recent disruptions like COVID-19, and limited product information restricts analysis of shipping performance factors, aligning with Arunachalam et al.'s (2018) observations about significant data exploitation challenges. Despite improvement, the model only explained half the variance in delivery times (R^2 = 0.513), reflecting theory-practice gaps and supporting Wuennenberg et al.'s (2023) observations about increasing complexity in deriving performance indicators.

The retail-focused context limits generalizability to other supply chain environments, reflecting Zamani et al.'s (2022) identified challenge of translating AI/ML potential into real-world applications. These limitations highlight the need for research with more diverse datasets, advanced modeling, and cross-industry applications.

5.1 Strategic Recommendations

Organizations should prioritize implementing enhanced delivery time prediction models, leveraging AI as a competitive advantage (Toorajipour et al., 2021). They should also develop differentiated shipping strategies based on product category, order value, and region, building on Zamani et al.'s (2022) work on genetic algorithms for route optimization. A systematic review of discount strategies should address the relationship between discount levels and return rates, responding to gaps in exploiting big data for supply chain value (Arunachalam et al., 2018).

Longer-term recommendations include supply chain restructuring based on regional performance patterns, investment in advanced analytics capabilities, and strategic partnerships with high-performing logistics providers, addressing. Sodiya et al. (2024) suggest growth in the value of big data globally and a sophisticated need for analytics in supply chain management. Adoption needs to be gradual, as noted by Ülkü and Mansouri 2023 with the remark on technological infrastructure, organizational readiness, and culture change required to successfully adopt AI and ML into supply chain functionality.

5.2 Future Research Directions

Future research can investigate the correlation between the speed of delivery and the rate of product returns and other possible explanations as well as moderating variables. This correlates with the research problem on the influence of Artificial Intelligence-activated decision-making on supply chain resilience and flexibility. More granular geographical analysis and incorporation of product characteristics would address the gap Arunachalam et al. (2018) identified regarding what is needed to exploit big data at the supply chain level.

Methodological improvements could include advanced time series techniques, causal inference methods, and multi-objective optimization approaches based on Zamani et al.'s (2022) discussion of current methodologies for predictive analytics in logistics. Additional customer data, external factor information, and extended longitudinal data would enhance future analyses, potentially addressing what Wuennenberg et al. (2023) characterized as the increasing complexity of modern supply chains.

6. Conclusion

This analysis of 9,994 retail orders revealed significant insights into shipping efficiency, profitability, and machine-learning applications in supply chain optimization. The substantial improvement in predictive modeling (from R^2 of 0.036 to 0.513) demonstrates the value of advanced analytics in supply chain management, addressing the research objective of evaluating the implementation and effectiveness of AI and ML applications in this domain.

Key findings include the counter-intuitive relationship between delivery speed and return rates, the critical role of discount strategies in profitability, and the dominant influence of temporal factors on delivery performance. These insights contribute to theory and practice by providing empirical evidence of AI and ML effectiveness in supply chain optimization, as outlined in the study's significance statement.

Organizations can leverage these findings to make more informed decisions about AI and ML investments while advancing academic understanding of these technologies' role in modern supply chain management. As global market uncertainty and evolving consumer expectations continue to exacerbate traditional supply chain challenges, as noted in the introduction, the ability to predict, adapt, and optimize through data-driven approaches will become increasingly central to competitive success.

Funding: This research received no external funding
Conflicts of Interest: The author declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers
Orcid ID: 0009-0006-2069-0303

References

[1] Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges, and implications for practice. *Transportation Research Part E: Logistics and Transportation Review*, *114*, 416–436. https://doi.org/10.1016/j.tre.2017.04.001

[2] Bala, P. K. (2012). Improving inventory performance with clustering-based demand forecasts. *Journal of Modelling in Management*, *7*(1), 23–37. https://doi.org/10.1108/17465661211208794

[3] Culot, G., Podrecca, M., & Nassimbeni, G. (2024). Artificial intelligence in supply chain management: A systematic literature review of empirical studies and research directions. *Computers in Industry*, *162*, 104132–104132. https://doi.org/10.1016/j.compind.2024.104132

[4] Kavaliauskiene, I. M., Taşkın, K., Ghorbani, S., Činčikaitė, R., & Kačenauskaitė, R. (2022). Reviewing the Applications of Neural Networks in Supply Chain: Exploring Research Propositions for Future Directions. *Information*, *13*(5), 261. https://doi.org/10.3390/info13050261

[5] Khedr, A. M., & Sheeja , R. S. (2024). Enhancing supply chain management with deep learning and machine learning techniques: A review. *Journal of Open Innovation Technology Market and Complexity*, *10*(4), 100379–100379. ScienceDirect. https://doi.org/10.1016/j.joitmc.2024.100379

[6] Lee, I., & Mangalaraj, G. (2022). Big data analytics in supply chain management: A systematic literature review and research directions. *Big Data and Cognitive Computing*, *6*(1), 1–29. MDPI. https://doi.org/10.3390/bdcc6010017

[7] Liu, K. Y. (2022). *Supply chain analytics*. Springer International Publishing. https://doi.org/10.1007/978-3-030-92224-5

[8] Mallesham, G. (2022). The role of AI and ML in revolutionizing supply chain management. *International Journal of Scientific Research and Management*, *10*(06), 918–928. https://doi.org/10.18535/ijsrm/v10i6.ec05

[9] Mishra, D., Dwivedi, Y. K., Rana, N. P., & Hassini, E. (2019). Evolution of supply chain ripple effect: A bibliometric and meta-analytic view of the constructs. *International Journal of Production Research*, *59*(1), 129–147. https://doi.org/10.1080/00207543.2019.1668073

[10] Smyth, C., Dennehy, D., Wamba, S. F., Scott, M., & Harfouche, A. (2024). Artificial intelligence and prescriptive analytics for supply chain resilience: a systematic literature review and research agenda. *International Journal of Production Research*, 1–25. https://doi.org/10.1080/00207543.2024.2341415

[11] Sodiya, E. O., Jacks, B. S., Ugwuanyi, E. D., Adeyinka, M. A., Umoga, U. J., Daraojimba, A. I., & Lottu, O. A. (2024). Reviewing the role of AI and machine learning in supply chain analytics. *GSC Advanced Research and Reviews*, *18*(2), 312–320. https://doi.org/10.30574/gscarr.2024.18.2.0069

[12] Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., & Fischl, M. (2021). Artificial intelligence in supply chain management: A systematic literature review. *Journal of Business Research*, *122*(1), 502–517. ScienceDirect. https://doi.org/10.1016/j.jbusres.2020.09.009

[13] Ülkü, M. A., & Mansouri, B. (2023). Supply chain analytics: Overview, emerging issues, and research outlook. *Springer EBooks*, 1–25. https://doi.org/10.1007/978-3-030-89822-9_80-1

[14] Wuennenberg, M., Muehlbauer, K., Fottner, J., & Meissner, S. (2023). Towards predictive analytics in internal logistics – an approach for the data-driven determination of key performance indicators. *CIRP Journal of Manufacturing Science and Technology*, *44*, 116–125. https://doi.org/10.1016/j.cirpj.2023.05.005

[15] Zamani, E. D., Smyth, C., Gupta, S., & Dennehy, D. (2022). Artificial intelligence and big data analytics for supply chain resilience: A systematic literature review. *Annals of Operations Research*, *327*(2). https://doi.org/10.1007/s10479-022-04983-y