Data Analytics Track

YAT133_GIZ1_DAT1_S4e_DEPI

Detailed Data Analysis Report: Supply Chain Insights

Submitted by:

Mohamed Mahmoud Hassan (Questions, Documentation, Visualization)

Shorouq Mohsen Ibrahim El-Houssiny (Questions, Visualization)

Hamza Abdelaaty Saber (Documentation, Visualization)

Saad Usama Saad (Visualization)

Rana Ismail Helmy (Questions, Data Cleaning)

Habiba Amr AbdelFatah (Visualization, Presentation)

Under the Supervision of:

Eng. Soha Nagy

Table of Content

1. Introduction

In the modern world of commerce, optimizing supply chain management is pivotal for achieving operational efficiency and staying competitive. The ability to fine-tune various stages of the supply chain—from production and manufacturing to shipping and logistics—can drastically influence profitability and customer satisfaction. This report focuses on a detailed analysis of a dataset containing 100 entries with 23 distinct features, each representing critical aspects of the supply chain process.

Our primary goal is to extract valuable insights from this data, addressing essential questions regarding supplier performance, logistics costs, production efficiency, and product quality. The report will also explore challenges faced during the analysis, tools used, and actionable recommendations for enhancing supply chain performance. By answering a set of targeted questions, we aim to provide clarity on areas that need improvement and potential strategies for optimization.

2. The Dataset Overview

2.1. Dataset Structure

The dataset is composed of a broad spectrum of supply chain data, divided into 23 columns. These columns capture the following dimensions:

- **Product Details:** SKU (Stock Keeping Unit), product type, price, availability, and the number of products sold.
- **Revenue & Costing:** Revenue generated, manufacturing costs, shipping costs, and overall costs.
- **Manufacturing Metrics:** Production volumes, manufacturing lead times, inspection results, and defect rates.
- **Shipping and Logistics:** Shipping times, transportation modes, carriers, and routes.
- **Supplier and Location Data:** Supplier names and corresponding geographic locations.
- **Customer Demographics:** Segments represented as "Non-binary," "Female," and "Unknown."

2.2. Initial Observation

Upon an initial examination, a few important observations surfaced:

- **Sales Patterns:** Skincare products exhibited higher sales volumes compared to haircare products, which suggests a higher demand in the market.
- **Defect Rates and Inspection Results:** A varying degree of quality was observed across suppliers, with certain suppliers experiencing more frequent defects or failed inspections.
- **Logistics Inefficiencies:** Shipping costs appeared to be higher for certain routes, particularly those associated with specific transportation modes or supplier locations.

- **Supplier Performance:** Supplier-related performance metrics such as defect rates, inspection results, and manufacturing costs differed across the dataset, highlighting potential areas for deeper investigation.

3. Data Exploration

Before proceeding to the data cleaning and analysis phase, we conducted a thorough exploration of the dataset to identify patterns, inconsistencies, and potential areas for deeper investigation.

3.1. Sales & Revenue Trends:

3.1.1. Product Sales:

The dataset revealed that some products are performing exceptionally well, such as skincare items with higher quantities sold. Conversely, other products with lower sales could be candidates for re-evaluation in terms of market demand.

3.1.2. Revenue Generated:

The revenue data shows a correlation between the number of products sold and revenue, but certain SKUs with fewer sales still yielded high revenue, possibly due to higher price points.

3.2. Supplier Insights:

3.2.1. Supplier Location Impact:

There is a geographical variation in the performance of suppliers. For example, products manufactured by suppliers based in Mumbai seem to have relatively lower defect rates compared to those based in other cities, such as Kolkata.

3.2.2. Supplier Defect Rates:

Supplier defect rates ranged from less than 1% to over 4%, suggesting significant discrepancies in quality control across different suppliers.

3.3. Logistics and Shipping Analysis:

3.3.1. Routes and Costs:

The analysis revealed that some shipping routes are considerably more expensive than others. This is especially noticeable with routes that depend heavily on air transport, which tends to inflate the costs compared to road or rail options.

3.3.2. Shipping Times:

Some suppliers experienced significantly longer shipping times due to reliance on less efficient transportation modes or longer routes. Optimizing shipping processes and choosing better routes could result in faster and more cost-effective deliveries.

4. Tools and Methods Used:

This section outlines the tools and methodologies used during the various stages of data cleaning, analysis, and visualization.

4.1. Python for Data Cleaning

For the data cleaning process, we utilized **Python** in a Google Colab environment, which allowed seamless access to the dataset stored in Google Drive. The primary library used was **Pandas**, which is highly effective for handling and manipulating large datasets. The following steps detail the method we used to import and explore the data:

4.2. Tableau for Data Visualization

Tableau was chosen for creating interactive and insightful visualizations. Tableau's intuitive drag-and-drop interface enabled us to build dashboards that answered key business questions and provided actionable insights. The tool was invaluable for exploring the relationships between various metrics, such as costs, defect rates, and shipping times.

5. Data Cleaning

5.1. Key Steps in the process:

5.1.1. Mounting Google Drive:

We mounted Google Drive to access the dataset directly in the Colab environment.

5.1.2. Reading the Dataset:

Using **`pd.read_csv()`**, we loaded the dataset into a Pandas DataFrame, which allowed us to perform further cleaning and manipulation.

5.1.3. Initial Exploration:

By using **`print(df.head())`**, we displayed the first few rows of the dataset, providing an initial overview of the data's structure, including columns, data types, and any visible inconsistencies.

5.2. Cleaning and Data Issues Identified:

5.2.1. Inconsistent Entries:

During the initial exploration, it became evident that some supplier names and product types had inconsistent entries (e.g., variations in spelling and formatting). These were standardized to ensure accurate analysis.

5.2.2. Missing Values:

Some columns, particularly "Inspection Results," contained missing or pending entries, which were flagged for future consideration.

This process laid the foundation for the subsequent steps in our analysis, ensuring that the dataset was properly cleaned and ready for further investigation.

6. Core Questions Explored

We designed the analysis around a set of key questions that target the main aspects of the supply chain. Below are the expanded versions of these questions, focusing on uncovering detailed insights.

6.1. Shipping Costs & Routes

Question: How do shipping costs vary across different routes?

Analysis: By analyzing the shipping routes (e.g., Route A, Route B, etc.) and correlating them with shipping costs, we found that routes involving air transport (Route C) were significantly more expensive than those using road or rail. Optimizing shipping routes by reducing reliance on air transport could lead to substantial cost savings.

6.2. Defect Rates per Supplier

Question: Which suppliers exhibit the highest defect rates, and how does this affect product quality?

Analysis: Suppliers such as Supplier 5, based in Kolkata, showed consistently higher defect rates, suggesting that their quality control processes are less efficient. These findings emphasize the need to either renegotiate terms with underperforming suppliers or consider alternative options.

6.3. Product Availability & Order Quantities

Question: How does product availability impact the quantity of orders placed?

Analysis: Products with higher availability levels generally had larger order quantities, which aligns with expectations. This finding underscores the importance of maintaining adequate stock levels to meet customer demand without delays.

6.4. Supplier Issues (Defects & Inspection Failures)

Question: Which suppliers are facing the most issues, either in terms of defects or failed inspections?

Analysis: Supplier 1 had a disproportionately high number of inspection failures compared to others, which suggests deeper problems with quality assurance. Addressing these issues would involve tightening the inspection process and possibly re-evaluating contracts with problematic suppliers.

6.5. Production Volume vs. Order Quantities

Question: Are production volumes sufficient to meet the demand reflected in order quantities, or do we need to scale production?

Analysis: In some cases, such as with skincare products, production volumes exceeded demand, leading to excess stock. For other products, particularly in the haircare line, production volumes were slightly lower than the order quantities, indicating a potential risk of stockouts.

6.6. Cost by Location

Question: How do costs vary depending on the geographic location of suppliers, and are there inefficiencies in storage or shipping?

Analysis: A significant insight revealed that products frequently ordered for delivery in Delhi were stored in Chennai, leading to inflated transportation costs. By optimizing storage locations to be closer to primary markets, the company could save on logistics expenses.

6.7. Average Defect Rates per Supplier

Question: What are the average defect rates per supplier?

Analysis: Suppliers 2 and 5 had below-average defect rates, indicating stronger quality control measures. Suppliers with consistently low defect rates could be rewarded with larger contracts, while those with higher rates would require intervention to improve product quality.

6.8. Manufacturing Costs vs. Selling Price

Question: What are the average manufacturing costs compared to the selling price of each product?

Analysis: In some cases, particularly for higher-end products, manufacturing costs were disproportionately high compared to their selling price. This suggests that either the production process needs optimization to reduce costs, or pricing strategies need revisiting to ensure profitability.

6.9. Fail Inspection Rate per Supplier

Question: What is the fail inspection rate per supplier?

Analysis: Supplier 4 had the highest inspection failure rate, which exceeded 15%. This may indicate systematic issues with their production process, necessitating a review of their manufacturing practices and potentially stricter inspections.

6.10. Manufacturing Costs and Lead Time per Supplier

Question: What is the manufacturing cost and lead time for each supplier?

Analysis: Supplier 6 had the longest lead times and highest manufacturing costs. This suggests inefficiencies in their processes, potentially caused by outdated technology or poor workflow management, indicating that production improvements or alternative suppliers may need to be considered.

6.11. Stock Levels per Product

Question: What are the stock levels for each product?

Analysis: Certain products, especially high-demand items, were frequently out of stock, while others were overstocked. This suggests a need for improved inventory management, such as implementing just-in-time (JIT) inventory systems or better demand forecasting methods.

6.12. Production Volumes vs. Order Quantities

Question: How do production volumes compare with order quantities?

Analysis: Production volumes generally matched order quantities, but in some cases, there were overproductions, leading to surplus stock. This could be addressed by aligning production schedules more closely with actual demand and making real-time adjustments based on order patterns.

6.13. Revenue per Supplier and Product Type

Question: What is the total and average revenue generated by each supplier across different product types?

Analysis: Supplier 5 generated the highest revenue across the board, particularly in the skincare category. This suggests that Supplier 5 should be prioritized for high-volume contracts, while other suppliers may need to focus on product categories where they perform better.

6.14. Products Sold by Customer Demographics

Question: What are the products sold based on customer demographics?

Analysis: Analysis of sales by demographic data indicated that younger customers (ages 18-30) were more likely to purchase skincare products, while older customers preferred haircare products. This insight can help tailor marketing strategies and product development to target specific demographic groups effectively.

7. Dashboards Construction via Tableau

In **Tableau**, a powerful data visualization tool, we created a series of interactive dashboards to present insights gained from the dataset. These dashboards were designed to allow users to explore the data dynamically, filtering by various dimensions to reveal deeper insights.

7.1. Common Filters Used

To enhance user experience and facilitate targeted analyses, we implemented several common filters across all dashboards. These filters included:

7.1.1. Location:

This filter allows users to focus on specific geographic regions, such as Mumbai, Delhi, or Kolkata. By filtering based on location, users can assess how various metrics—like shipping costs, defect rates, and production efficiencies—differ across regions.

7.1.2. Supplier Name:

By enabling filtering by individual suppliers, users can quickly examine the performance metrics associated with each supplier. This includes critical insights into defect rates, shipping times, and manufacturing costs, providing a clear comparison of suppliers' performances.

7.1.3. Product Type:

Users can filter the data based on different product types, such as haircare or skincare products. This filter allows stakeholders to identify trends and patterns within specific categories, which can be useful for targeted marketing strategies and inventory planning.

7.2. Dashboard Breakdown

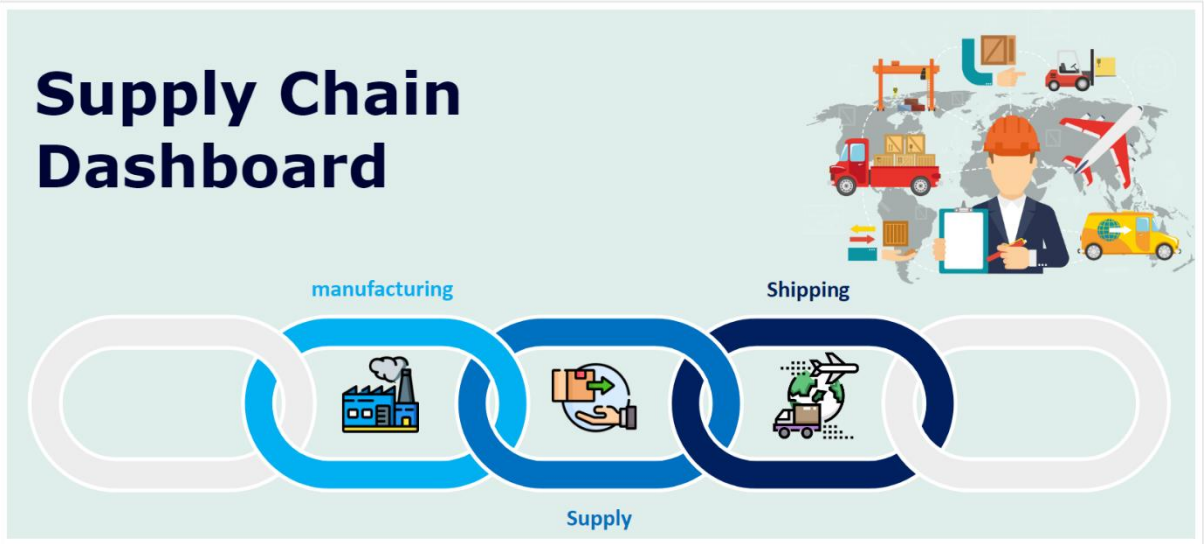
The dashboards were categorized into three distinct areas, each focusing on a specific aspect of the supply chain with a landing page containing hyperlinked icons for each single dashboard of the following:

7.2.1. Home Dashboard

Purpose: This Dashboard, as shown below, acts as a landing page for the whole dashboards where it is crucial for linking and accessing the three dashboards (Manufacturing, Supply, Shipping).

Key Metrics:

- Manufacturing Icon: a navigator to the Manufacturing Dashboard.
- Supply Icon: a navigator to the Supply Dashboard.
- Shipping Icon: a navigator to the Shipping Dashboard.



7.2.2. Manufacturing Dashboard:

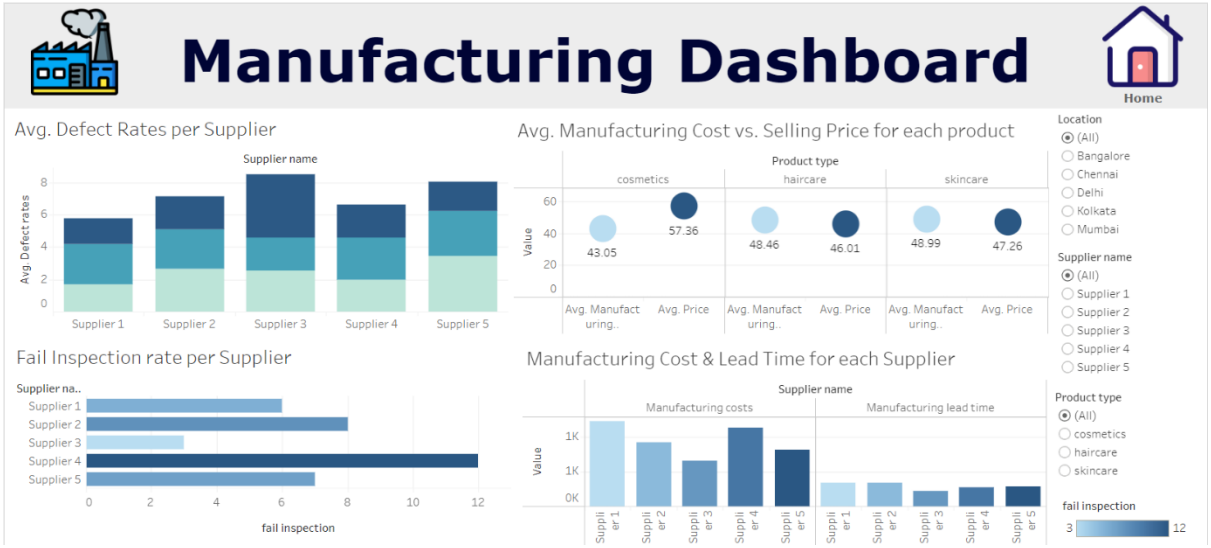
Purpose: This dashboard presents metrics crucial for assessing manufacturing efficiency and product quality.

Key Metrics:

- **SKU**: Identifies individual products and their characteristics.

- **Production Volume:** Reflects how much product is manufactured over a specific period.
- **Manufacturing Lead Time:** Shows the time taken from production start to finish.
- **Manufacturing Cost:** Breaks down the costs incurred during the manufacturing process.
- **Inspection Results:** Displays the outcomes of quality inspections.
- **Defect Rates:** Measures the percentage of defective items produced.

This dashboard , as shown below, allows stakeholders to evaluate the efficiency of the manufacturing process and identify areas needing improvement, particularly in quality control.



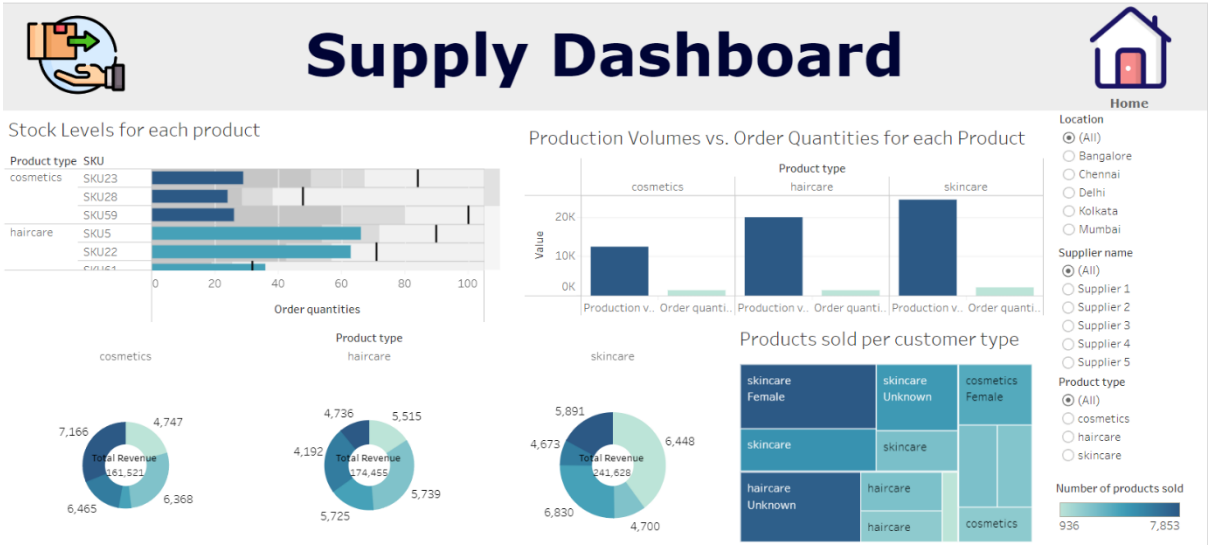
7.2.3. Supply Dashboard:

Purpose: Focused on inventory management and supplier performance, this dashboard helps in assessing the supply chain's effectiveness.

Key Metrics:

- **Stock Levels:** Displays the current inventory of products.
- **Order Quantities:** Indicates the quantities of products ordered by customers.
- **Production Volumes:** Shows how much of each product is being produced.
- **Products Sold:** Tracks sales performance across different products.

This dashboard aids in inventory planning, helping businesses avoid stockouts and ensuring optimal stock levels are maintained.



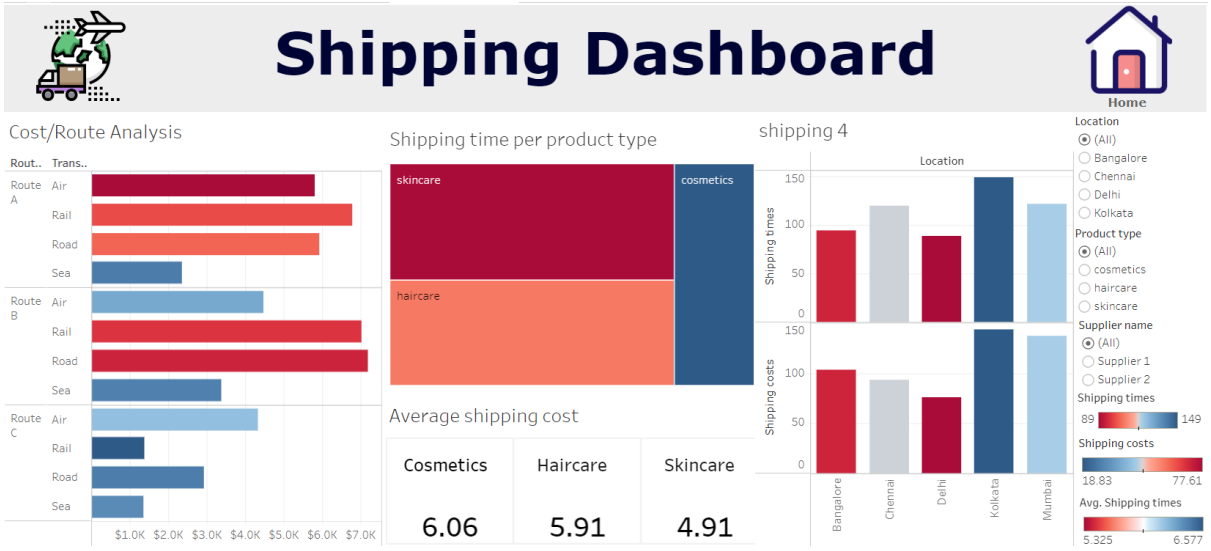
7.2.4. Shipping Dashboard:

Purpose: This dashboard analyzes logistics efficiency, including shipping practices and costs.

Key Metrics:

- **Transportation Modes:** Displays the types of transportation used (e.g., air, road, rail).
- **Routes:** Lists the shipping routes taken.
- **Shipping Costs:** Details the costs associated with shipping.
- **Shipping Carriers:** Shows which carriers are being used for deliveries.
- **Shipping Time:** Measures the duration taken for products to reach customers.

The shipping dashboard, as shown below, is crucial for understanding logistical challenges and exploring opportunities for cost reduction and improved delivery times.



8. Challenges Faced During Analysis

As with any data analysis project, several challenges arose throughout the process. Addressing these challenges was essential for ensuring accurate insights and reliable recommendations.

8.1. 8.1 Data Quality Issues

8.1.1. Inconsistent Data Entries:

Minor inconsistencies were found in supplier names, product types, and other categoricalvariables. These inconsistencies could potentially skew results if not corrected.

8.1.2. Missing Values:

The "Inspection Results" column contained "Pending" entries, which posed a challenge in measuring defect rates comprehensively. These entries needed careful consideration in the analysis to avoid misleading conclusions.

8.1.3. Outlier Detection

Outliers in manufacturing costs and defect rates were identified. Distinguishing between genuine data points and entry errors required additional investigation, impacting the time spent on the analysis.

8.1.4. Tool Limitations

While both Python and Tableau are powerful tools, they come with limitations. For instance, Tableau can sometimes struggle with extremely large datasets, leading to performance issues. Adequate hardware and data preparation strategies were necessary to mitigate these issues.

9. Recommendations Based on Analysis

Based on the comprehensive analysis conducted through the dashboards and the insights gained, several actionable recommendations can be made to enhance supply chain performance:

9.1. Optimize Shipping Costs

Recommendation: Given the significant variation in shipping costs across different routes, it is recommended to conduct a detailed review of shipping practices. This may involve renegotiating contracts with carriers or exploring alternative routes to reduce expenses. For example, utilizing more cost-effective modes of transport for certain routes could result in significant savings.

9.2. Supplier Performance Management

Recommendation: Suppliers with high defect rates should be subject to quality improvement initiatives. This could involve conducting workshops on quality control practices or implementing stricter inspection protocols. Additionally, consider diversifying the supplier base to mitigate risks associated with over-reliance on underperforming suppliers.

9.3. Enhance Inventory Management

Recommendation: Establishing more robust inventory management practices is crucial to aligning stock levels with actual demand. Implementing just-in-time inventory systems could help in reducing excess stock while ensuring that high-demand products are readily available.

9.4. Improve Geographic Cost Efficiency

Recommendation: Address the inefficiencies identified with geographic disparities in storage and shipping costs. For example, if products ordered for Delhi are stored in Chennai, consider relocating inventory closer to key markets to optimize logistics and reduce transportation costs.

10. Conclusion

This comprehensive data analysis report has explored key performance indicators across various facets of the supply chain, utilizing tools such as Python and Tableau to derive insights from a rich dataset. By addressing vital questions related to shipping costs, supplier performance, product availability, and production capacity, we have highlighted areas for improvement and optimization.

The recommendations provided herein serve as a roadmap for enhancing operational efficiency, reducing costs, and improving overall supply chain effectiveness. By implementing these strategies, organizations can better position themselves to respond to market demands and navigate the complexities of supply chain management in today's dynamic business landscape.