Comprehensive Supply Chain Analytics: From Product Revenue to Customer Insights

Prepared by: **Oussama KASSIMI**

Montpellier Business School

Contents

Executive Summary

This report presents a data-driven analysis of a supply chain system to identify performance patterns and optimization opportunities. The project explored product profitability, supplier reliability, logistics cost efficiency, and customer segmentation using Python analytics and interactive dashboards. Five research questions (RQ1–RQ5) guided the analysis to reveal actionable business insights.

Methodology

The dataset contained diverse operational data collected across several functional areas of the supply chain, including:

- **SKU-level sales data:** representing quantities sold, product categories, and associated revenue.
- **Supplier performance metrics:** such as defect rates, manufacturing costs, and delivery lead times.
- **Logistics data:** encompassing transportation modes, carriers, routes, and average shipping costs.
- **Customer demographics:** detailing gender distribution, regional segmentation, and purchasing behavior.

Data Preparation and Cleaning

Before analysis, the dataset underwent several preprocessing steps. Missing or inconsistent values were handled using imputation or removal, and duplicate rows were eliminated to maintain accuracy. Numeric variables were standardized to enable fair comparison between suppliers and logistics modes.

Analytical Tools and Framework

The data analysis was conducted using Python and its analytical ecosystem:

- **pandas** for data manipulation and aggregation,
- **matplotlib** and **seaborn** for descriptive and statistical visualization,
- **plotly** for interactive and dynamic charting.

Dashboard Development

To make the analysis interactive, a **Streamlit dashboard** was developed. It allows real-time exploration of data through filters (by product type, supplier, or region) and integrates both matplotlib and plotly charts for dynamic visualization. Users can interact with plots, compare suppliers, and export filtered datasets.

Analytical Objectives

The overall goal of the methodology was to transform raw operational data into actionable insights that support decision-making in:

- Product strategy — identifying profitable categories;
- Supplier evaluation — highlighting inefficiencies;
- Logistics optimization — balancing cost and speed;
- Customer relationship management — focusing on recurring sales and segmentation.

RQ1: Which Product Categories Contribute Most to Total Revenue?

Skincare products generated the highest total revenue ($241,628), followed by Haircare ($174,455) and Cosmetics ($161,521). This shows that skincare represents the company's strongest product line and should remain a strategic focus area.

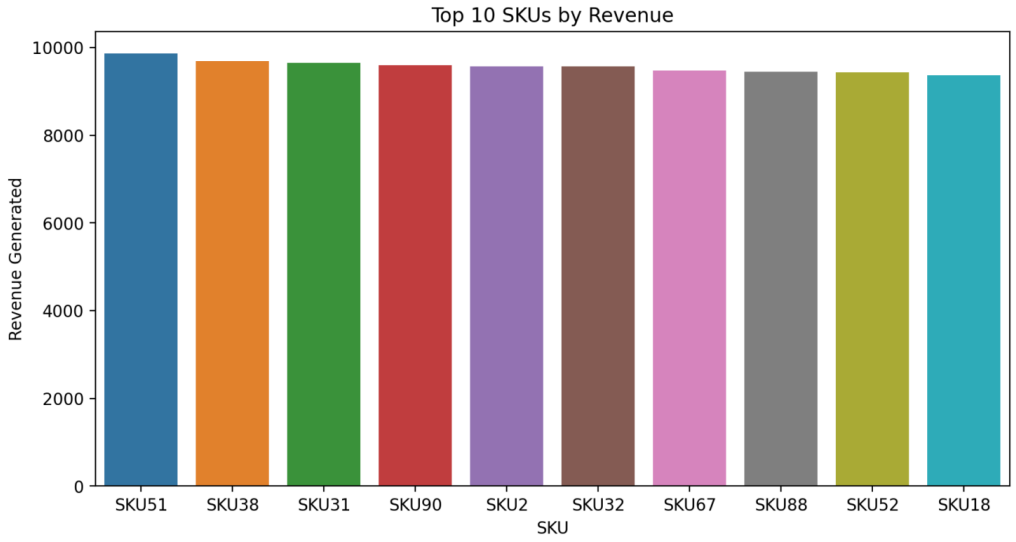

Figure 1: Top 10 SKUs by Revenue

Interpretation: Skincare dominates the revenue mix, while haircare and cosmetics provide steady but secondary contributions.

RQ2: Supplier Performance and Defect Rates

Supplier analysis revealed that Supplier 3 and Supplier 5 have higher defect rates and longer lead times, while Supplier 1 performs best overall. The scatterplot between defect rate and lead time shows a weak positive relationship, indicating that slower suppliers often deliver lower-quality output.

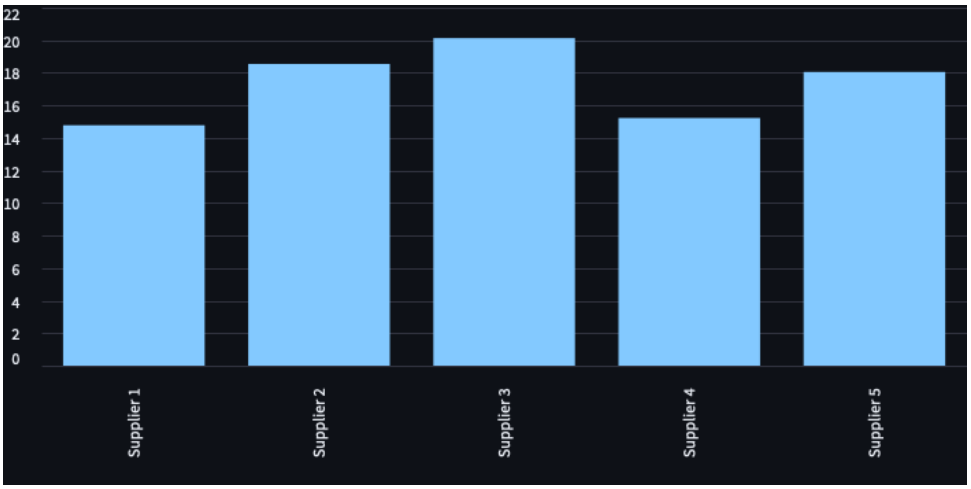

Figure 2: Supplier Defect Rate vs. Lead Time

Interpretation: Supplier 1 demonstrates the highest efficiency and consistency. Suppliers 3 and 5 require performance improvement initiatives or contractual renegotiation.

RQ3: Shipping Analysis (Costs, Times, and Modes)

The logistics analysis showed that Sea transport is the most cost-effective mode (avg. $417), while Air transport is the most expensive (avg. $561). Bubble charts revealed that longer shipping times correlate with higher costs, particularly for Air and Rail.

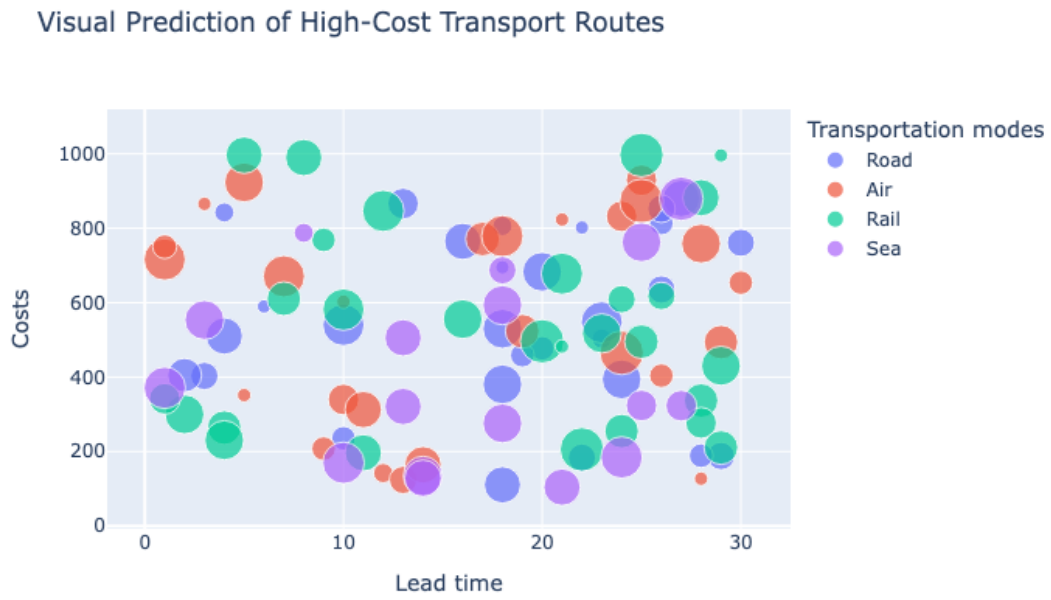

Figure 3: Shipping Cost vs. Lead Time by Transport Mode

Interpretation: Air transport provides speed but lacks cost efficiency. Sea and Rail modes offer a better cost–time balance for long-term planning.

RQ4: Customer Segmentation and Recurring Sales

Customer analysis revealed that the "Unknown" demographic segment produced the highest total sales volume, possibly due to unclassified online purchases. Male customers, however, recorded the highest average revenue per sale ($16.87), while Female customers led in total transaction count.

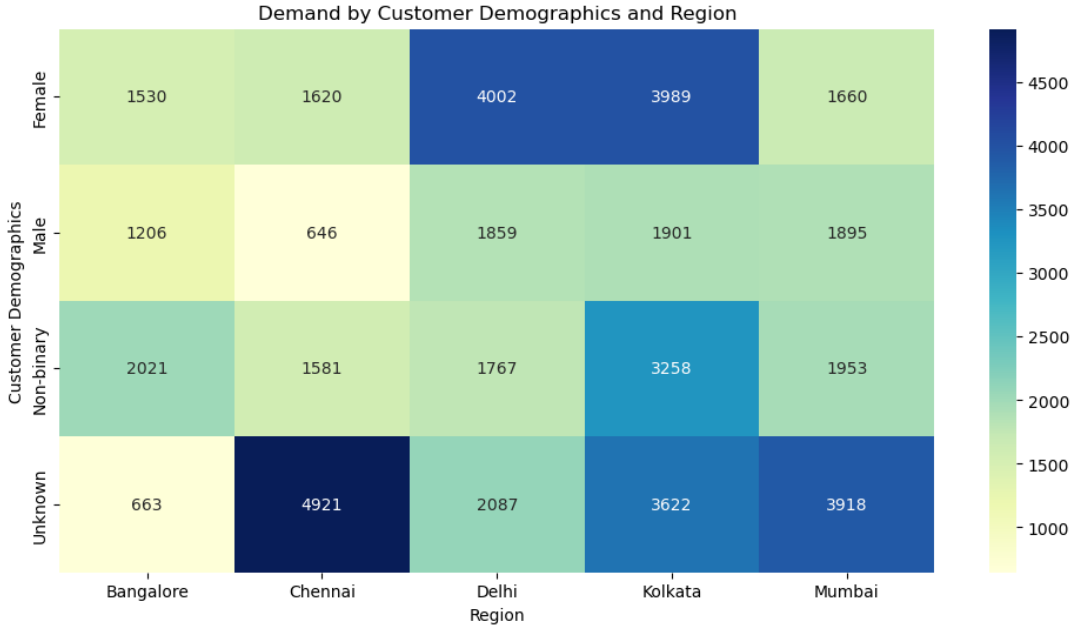

Figure 4: Demand by Customer Demographics and Region

Interpretation: Female customers are more numerous, while male customers are more profitable. Regional heatmaps show strong demand in Delhi and Kolkata.

RQ5: Correlation Between Cost, Defects, and Lead Time

Correlation analysis revealed a mild positive relationship between cost and lead time ($r = 0.24$), suggesting that longer shipments increase costs. Defect rates were largely independent, implying that quality differences are supplier-specific rather than process-driven.

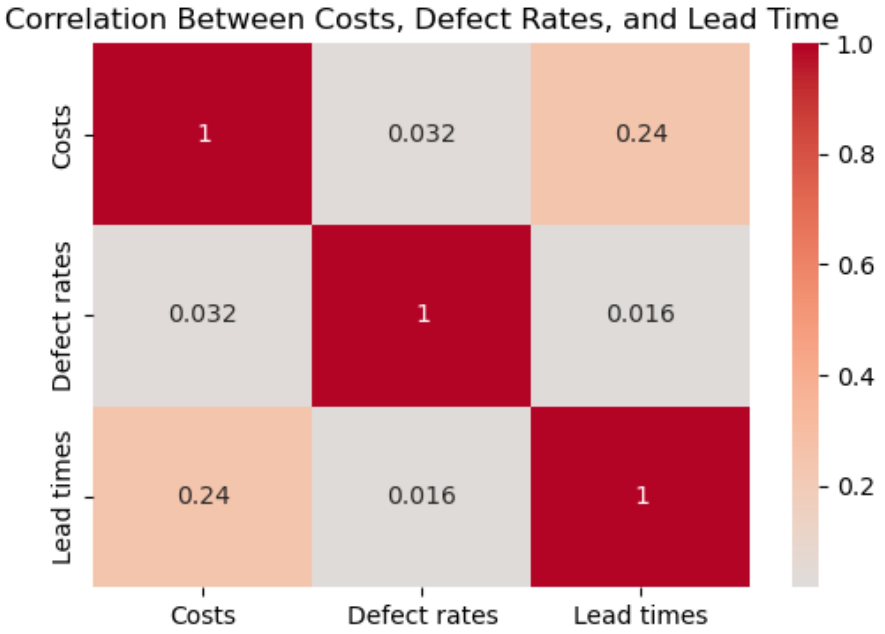

Figure 5: Correlation Matrix: Cost, Defects, and Lead Time

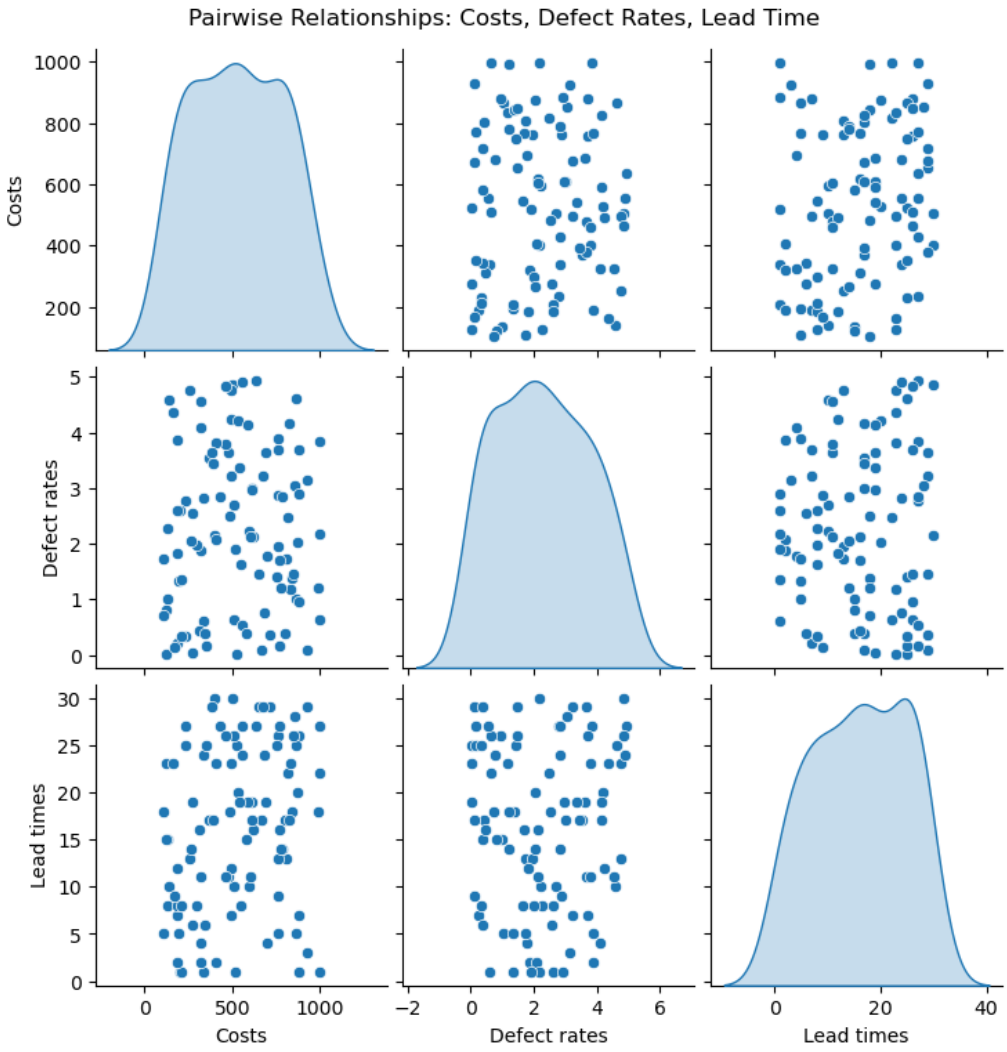

Figure 6: Pairwise Relationships Between Key Supply Chain Metrics

Interactive Dashboard Benefits

Interactive dashboards accelerate cycle time analysis and decision-making. Features such as filtering, tooltips, zooming, and dynamic linking enable real-time exploration of supplier, product, and logistics data.

Managerial Insights and Recommendations

- **Product Strategy:** Prioritize skincare for revenue growth.
- **Supplier Management:** Reassess Suppliers 3 and 5 due to reliability concerns.
- **Logistics Optimization:** Use sea and rail transport to minimize cost.
- **Customer Segmentation:** Target male customers for premium pricing, female for volume sales.
- **Data Management:** Improve classification of "Unknown" customers to refine marketing strategies.
- **Analytics Adoption:** Deploy interactive dashboards for continuous performance tracking.

Conclusion

This analysis provided a clear understanding of the company's supply chain performance from multiple perspectives: products, suppliers, logistics, and customers. Skincare products emerged as the most profitable category, confirming their central role in driving revenue. Supplier evaluation highlighted significant quality and reliability differences — particularly for Suppliers 3 and 5 — which should be addressed through performance improvement initiatives. Logistics analysis revealed that sea and rail transport offer the best balance between cost and efficiency, while air transport should be reserved for urgent deliveries. Customer analysis showed that the "Unknown" demographic segment contributes heavily to revenue, suggesting the need to strengthen data collection on customer profiles. The correlation study confirmed that costs, lead times, and defect rates interact weakly but still influence each other in complex ways. Overall, the combination of data analytics and interactive dashboards has proven essential for identifying inefficiencies and guiding strategic decisions. Future work should focus on predictive analytics and automation to further improve responsiveness and supply chain resilience.

End of Report