Supply Chain Analytics Dashboard

Advanced Statistics Project

Sam Collin Allen Joseph Sahaya Jude
Yogesh Muthukumar

Montpellier Business School
MSc Purchasing and Supply Chain Management

December 14, 2025

Abstract

This project aims to design and implement an interactive Supply Chain Analytics Dashboard using Python and Streamlit. The dashboard provides key performance indicators, visualization tools, and analytical insights that help evaluate logistics efficiency, delivery performance, and profit trends across countries and product categories. The project demonstrates practical data analysis and dashboard design skills aligned with advanced statistical concepts in business analytics.

1 Introduction

Modern supply chains generate large volumes of data related to customer orders, deliveries, transportation, warehouse operations, and product performance. Analyzing this data enables organizations to make informed strategic decisions and optimize operational processes. The goal of this project is to develop a web-based dashboard that processes the DataCo Supply Chain dataset and transforms it into interactive, visual insights.

The dashboard assists in:

- Monitoring delivery performance and delays.
- Identifying profitability across product categories.
- Analyzing geographic differences in late shipments.
- Understanding patterns in logistical performance.

2 Literature Background and Research Inspiration

Existing research in supply chain analytics emphasizes the role of data-driven techniques in improving delivery reliability, operational efficiency, and profitability. Carbonneau et al. (2008) demonstrate how analytical and machine learning approaches can be applied to large supply chain datasets to identify patterns related to delivery performance, demand behavior, and operational risk.

Inspired by this literature, the present project adopts an applied analytics approach rather than a full predictive modeling framework. Similar to prior research, the analysis focuses on key performance indicators such as delivery delays, late shipment rates, and profitability across product categories and geographic regions.

While the referenced literature primarily focuses on forecasting accuracy and predictive models, this project adapts the analytical logic to an exploratory and managerial context. The emphasis is placed on descriptive statistics, visualization, and interactive dashboard design to support business decision-making. This approach aligns academic insights with practical supply chain analytics objectives, consistent with the goals of the course.

3 Dataset Description

The dataset used is the *DataCo Supply Chain Dataset*, consisting of over 180,000 rows of shipment, customer, and product data. Key variables include:

- **Days for shipment (scheduled)** – expected delivery timeline.
- **Days for shipping (real)** – actual shipping time.
- **Benefit per order** – profit margin.
- **Customer Country**.
- **Category Name** – product category.

The dataset contains both numerical and categorical variables suitable for descriptive and inferential statistical analysis.

4 Methodology

4.1 Data Cleaning

Data preparation included:

- Removing leading and trailing spaces from column names.

- Handling missing values.
- Ensuring consistent data types.

4.2 Feature Engineering

A new metric, **Delivery Delay**, was created:

$$\text{Delivery\_Delay} = \text{Days for shipment (scheduled)} - \text{Days for shipping (real)}$$

A binary indicator was computed:

$$\text{Is\_Late} = \begin{cases} 1 & \text{if Delivery\_Delay ¡ 0} \\ 0 & \text{otherwise} \end{cases}$$

4.3 Dashboard Implementation

The dashboard was developed using Streamlit and Plotly. Key components include:

- KPI cards for total orders, average profit, and late delivery rate.
- Bar charts for profitability by product category.
- Global choropleth maps for late delivery rates.
- Interactive sidebar filters.

5 Analytical Insights

5.1 Profitability by Category

Profitability varies significantly across categories. Some product groups generate higher margins, while others show lower or volatile performance, supporting strategic sourcing decisions.

5.2 Delivery Performance

Late delivery rates highlight operational weaknesses in specific regions. High-delay regions indicate the need for improved logistics coordination and demand planning.

5.3 Geographical Patterns

Geographical visualization identifies countries with persistent delivery challenges, enabling targeted operational improvements.

6 Advanced Statistical Analysis

6.1 Correlation Analysis

Correlation analysis helps identify relationships between delivery time, profit, and operational metrics.

6.2 Distribution Analysis

Histograms and boxplots reveal skewness and outliers in shipment delays and profitability.

6.3 Hypothesis Testing

An example hypothesis test compares mean delivery delays across product categories to evaluate statistically significant differences.

7 Conclusion

This project integrates supply chain analytics concepts with advanced statistical methods. By grounding the analysis in existing literature and applying it to a real-world dataset, the project demonstrates both academic rigor and practical relevance. The Streamlit dashboard supports data-driven decision-making and illustrates the value of applied analytics in supply chain management.

References

- Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 184(3), 1140–1154.
- DataCo Supply Chain Dataset. Kaggle.
- Streamlit Documentation.
- Plotly Visualization Library.