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 13, 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 Dataset Description

The dataset used is the *DataCo Supply Chain Dataset*, consisting of 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.

3 Methodology

3.1 Data Cleaning

Data preparation included:

- Removing leading/trailing spaces from column names.
- Handling missing values.
- Ensuring consistent data types (numeric vs categorical).

3.2 Feature Engineering

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

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

A binary flag was also computed:

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

3.3 Dashboard Implementation

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

- KPI cards for total orders, average profit, and late delivery rate.
- Bar charts displaying profit by product category.
- Geographic heatmap visualizing late deliveries globally.
- Sidebar filters for category selection.

4 Analytical Insights

4.1 Profitability by Category

The dashboard shows variations in profitability across categories. Some categories generate consistently higher margins, while others perform poorly. This insight helps in portfolio optimization and procurement planning.

4.2 Delivery Performance

Late shipment percentage is a crucial metric for supply chain reliability. Countries with the highest late delivery rates require:

- better logistics partnerships,
- improved demand planning,
- stronger inventory positioning.

4.3 Geographical Patterns

Through the choropleth map, countries with significant operational challenges were identified. This can support:

- geographic risk assessment,
- transportation mode adjustments,
- targeted service-level improvements.

5 Advanced Statistical Analysis

5.1 Correlation Analysis

A correlation matrix can reveal how variables like profit, days to ship, and sales value relate. For example:

- Strong correlations may indicate dependency.
- Weak correlations suggest independent operational factors.

5.2 Distribution Analysis

Shipment delays and profit margins often follow skewed distributions. Histograms and boxplots help identify:

- outliers,
- right or left skewness,
- clusters in performance.

5.3 Hypothesis Testing

Example analysis:

$$H_0 : \text{Mean delay for Category A} = \text{Mean delay for Category B}$$

Using a t-test, we determine whether delays statistically differ between categories.

6 Conclusion

This project successfully integrates supply chain analytics concepts with advanced statistical methods. The Streamlit dashboard offers a user-friendly environment to explore metrics, identify patterns, and support decision-making. The use of feature engineering, visualization, and statistical analysis demonstrates practical business analytics skills relevant to supply chain management.

References

- DataCo Supply Chain Dataset. Kaggle.
- McKinsey Global Supply Chain Analytics Reports.

- Streamlit Documentation.
- Plotly Visualization Library.