Research Paper Title:

Exploratory Analysis of Shipping Lead Times and Delay Patterns in a National Distribution Network

Done By:

Indranil Sanyal (UMID30012679568)

1. Introduction

Having efficient logistics processes is essential for national distributors, as it directly impacts customer satisfaction and scalability. Nassau Candy Distributor is a national distributor that distributes products from the manufacturing divisions to customers in various regions of the United States. Although the company has access to very rich order and shipment data, logistics analysis has not been able to provide insights into route-level and geographic efficiency.

The objective of this research work is to examine the performance of shipping lead times with respect to historical order data to determine delay patterns and inefficiencies. The analysis will be done using exploratory data analysis (EDA) techniques with the help of Power BI.

2. Dataset Description

The data set consists of transactional data for shipping, with the following key features:

- Order information and dates (Order Date, Ship Date)
- Geographic information (Region, State, City, Postal Code)
- Operational information (Ship Mode, Division)
- Financial information (Sales, Cost, Gross Profit)
- Derived variable: Shipping Lead Time (Ship Date - Order Date)

The data set allows analysis to be performed at the order, route, region, and shipping mode levels.

3. Methodology

3.1 Data Preparation

- Converted Order Date and Ship Date to date format
- Computed Shipping Lead Time in days
- Identified extreme lead-time values spanning several years

3.2 Outlier Handling

- Exploratory data analysis showed unusually high lead times for shipping (above 900 days), suggesting data anomalies such as late data entry or closing of old orders. Rather than removing the data, a domain threshold was used for analysis.
- Lead times above 30 days were considered outliers for analysis
- Original data was kept to maintain data integrity

3.3 Key Metrics

- Average Lead Time
- Delay Frequency (%): Percentage of shipments exceeding an acceptable lead-time threshold
- Route Volume: Number of orders per route
- Geographic Aggregations: State and region level averages

4. Exploratory Data Analysis (EDA) Findings

4.1 Lead Time Distribution

- There is a strong right skew in shipping lead times
- The 25th percentile of raw lead time was determined to be an extreme value (over 1,000 days), which supported the presence of outliers
- After adjusting for realistic constraints, most shipments fell within a short delivery window

4.2 Route-Level Performance

- Significant variation exists across division-to-state routes
- Certain routes consistently show higher average lead times and higher delay frequency
- High-volume routes with persistent delays indicate structural inefficiencies rather than random variation

4.3 Geographic Bottlenecks

- State-level analysis reveals clusters of delayed shipments in specific regions
- These bottlenecks suggest infrastructure, distance, or coordination-related challenges

4.4 Ship Mode Comparison

- Shipping performance varies notably across ship modes
- Faster ship modes show lower delay frequency but are not always optimally utilized
- Standard shipping dominates volume but contributes disproportionately to delays

5. Key Insights

- Delays are not evenly distributed, a small set of routes and areas is responsible for the majority of delayed shipments
- Average lead time is not a reliable measure, delay frequency is a better indicator of reliability
- Geographic hotspots of delays indicate location-related operational constraints
- Anomalies in the data have a large impact on KPIs if not addressed by effective data preprocessing

6. Business Recommendations

- Route-Level Monitoring
    - Execute ongoing tracking for high-delay routes
    - Highest priority for correction on high-volume delayed routes
- Geographic Intervention
    - Review regional infrastructure or partner quality in bottleneck states
    - Explore new fulfilment locations for high-delay regions
- Shipping Mode Optimization
    - Reassess ship mode assignments based on reliability, not just cost
    - Execute conditional upgrades for routes with high delay frequencies
- Data Governance Improvements
    - Enforce validation rules for shipment dates
    - Auto-flag extreme lead times

7. Conclusion

This study demonstrates how exploratory data analysis and interactive dashboards can transform raw shipment data into actionable logistics intelligence. By shifting from reactive issue resolution to data-driven decision-making, national distributors can improve reliability, reduce costs, and enhance customer satisfaction.