Factory-to-Customer Shipping Route Efficiency Analysis

Author: Rovin Lobo
Email: rovinlobo0099@gmail.com
Internship Organization: Unified Mentor Pvt Ltd
Institution: Presidency University

Abstract—

Efficient logistics and supply chain management are essential components of modern business operations. Organizations must ensure that products are delivered to customers quickly, reliably, and cost-effectively. Delays in transportation or inefficient shipping routes can lead to increased operational costs and reduced customer satisfaction. This project focuses on analyzing the efficiency of factory-to-customer shipping routes using a dataset provided by Nassau Candy Distributor.

The study uses data analytics techniques to examine delivery times, regional shipping performance, product sales distribution, and transportation efficiency. By applying Python-based analytical tools such as Pandas, Matplotlib, and Seaborn, the project identifies patterns in logistics performance and highlights areas where improvements can be made. An interactive dashboard was developed using Streamlit to visualize shipping performance and key operational metrics.

The results of this analysis provide insights into shipping delays, high-performing regions, and product demand trends. These insights can help organizations optimize their logistics networks, improve delivery efficiency, and enhance overall supply chain performance.

Index Terms—
Logistics, Shipping Efficiency, Data Analytics, Supply Chain, Route Optimization.

I. Introduction

In today's highly competitive business environment, efficient supply chain management has become a crucial factor for organizational success. Companies that distribute products across multiple geographic locations must ensure that their logistics and transportation systems operate efficiently. Delays in shipping routes, inefficient delivery planning, and poor logistics coordination can increase operational costs and negatively impact customer satisfaction.The primary objective of logistics management is to ensure that products are delivered from manufacturers or factories to customers in the most efficient manner possible.

Shipping route efficiency plays a significant role in achieving this objective. Organizations must continuously monitor delivery times, transportation performance, and regional distribution patterns in order to identify potential inefficiencies and improve operational performance.

This project focuses on analyzing factory-to-customer shipping efficiency using transactional shipping data from Nassau Candy Distributor. The dataset provides valuable information about order dates, shipping dates, shipping modes, product sales, and regional distribution. By analyzing these variables, it becomes possible to evaluate logistics performance and identify patterns that may indicate inefficiencies in the shipping process.

The project was conducted as part of a **Data Science Internship at Unified Mentor Pvt Ltd**, where the objective was to apply data analysis techniques to real-world logistics data. During the internship, various data analytics tools and programming techniques were used to analyze shipping performance and visualize key operational metrics.

Python was used as the primary programming language for this analysis. Several libraries such as **Pandas, Matplotlib, and Seaborn** were used for data manipulation and visualization. Additionally, an interactive analytics dashboard was developed using **Streamlit**, allowing users to explore shipping performance through dynamic charts and filters.Through the application of data analytics techniques, this project aims to identify delivery delays, analyze regional shipping performance, evaluate different shipping modes, and assess overall route efficiency. The insights generated from this analysis can help organizations improve logistics operations, optimize shipping routes, and enhance overall supply chain efficiency.

II. Objectives

The primary objective of this project is to analyze shipping route efficiency using data analytics techniques. The analysis aims to identify logistics inefficiencies and provide recommendations to improve delivery performance.

The specific objectives of this study include:

- Analyzing shipping performance across different geographic regions.
- Measuring delivery times and identifying delays in the distribution process.
- Evaluating the performance of different shipping modes such as Standard Class and First Class shipping.
- Identifying cities that experience longer shipping durations.
- Analyzing product sales distribution to determine high-demand products.

- Evaluating route efficiency between cities and regions.
- Developing an interactive dashboard to visualize logistics performance.

By achieving these objectives, the project aims to provide meaningful insights that can support improvements in supply chain management and logistics operations.

III. Dataset Description

The dataset used in this project contains shipping and sales transaction records from Nassau Candy Distributor. The dataset includes information about orders, shipping details, product sales, and geographic distribution.The dataset consists of multiple attributes that allow comprehensive analysis of logistics performance. Some of the key attributes included in the dataset are:

- **Order ID:** Unique identifier for each order transaction.
- **Order Date:** The date when the order was placed.
- **Ship Date:** The date when the order was shipped to the customer.
- **Ship Mode:** The method used for shipping, such as Standard Class or First Class.
- **City:** The destination city for the shipment.
- **Region:** The geographic region where the customer is located.
- **Product Name:** The product that was ordered by the customer.
- **Sales:** Total sales value generated from the order.
- **Units:** The number of product units sold.
- **Gross Profit:** The profit generated from the transaction.

These variables enable detailed analysis of logistics efficiency, sales performance, and regional distribution patterns. By analyzing these attributes, it becomes possible to evaluate how effectively products are delivered from the factory to customers.

IV. Methodology

This project follows a structured data analytics methodology consisting of several stages. Each stage focuses on transforming raw data into meaningful insights.

Data Collection and Loading

The dataset was imported into a Python environment using the Pandas library. The CSV file containing shipping records was loaded into a DataFrame for analysis.

Data Cleaning

Data cleaning is an important step in ensuring the accuracy of analysis results. During this stage, date columns were converted into appropriate datetime formats. Shipping durations were calculated by determining the difference between the order date and ship date.Unrealistic shipping durations were normalized to ensure that extreme values did not distort the analysis.

Feature Engineering

Several new variables were created to enhance the analysis:

- **Shipping Days:** Number of days between order date and ship date.
- **Profit Margin:** Ratio of gross profit to total sales.
- **Sales per Unit:** Average revenue generated per product unit sold.

These additional features help measure logistics efficiency and product profitability.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) was performed to understand data distributions and identify patterns. Various visualizations were created to analyze regional sales, shipping times, and product demand.

Visualization and Dashboard Development

Graphs and charts were generated using Matplotlib and Seaborn libraries to visualize key logistics metrics. An interactive web dashboard was built using Streamlit, allowing users to explore shipping performance through filters and dynamic visualizations.

V. Data Analysis

Several analytical approaches were used to examine the shipping performance of the distribution network.

Shipping Time Analysis

Shipping duration was calculated by subtracting the order date from the ship date. This metric helps identify delivery delays and evaluate how efficiently products are transported to customers.

Regional Sales Analysis

Sales data was grouped by region to analyze geographic distribution of revenue. This analysis helps identify regions that contribute the most to overall sales.

Shipping Mode Analysis

Different shipping modes were analyzed to determine which transportation methods deliver products faster. This helps businesses evaluate the effectiveness of different delivery strategies.

Product Performance Analysis

Sales data was grouped by product name to identify top-performing products. This analysis helps businesses understand product demand patterns and prioritize inventory management.

City-Level Delay Analysis

Shipping times were analyzed at the city level to detect locations where deliveries take longer than expected. Identifying these cities helps organizations focus on improving logistics in problematic areas.

Route Efficiency Analysis

Shipping routes were analyzed by combining city and region information. This allows evaluation of transportation efficiency across different delivery routes.

VI. Results and Insights

The analysis generated several insights related to logistics performance and sales distribution.First, the analysis revealed that shipping times vary across different regions. Some regions consistently show longer delivery times, suggesting potential inefficiencies in transportation routes.

Second, shipping mode analysis indicates that faster shipping methods provide significantly shorter delivery durations compared to standard shipping options.Third, city-level analysis highlighted certain cities where delivery delays occur more frequently. These locations may require improved logistics planning or better transportation infrastructure.Fourth, product sales analysis showed that a small number of products contribute a large portion of the total sales revenue. This indicates that focusing inventory management on these products may improve operational efficiency.Finally, route efficiency analysis identified specific city-to-region combinations where shipping times are longer than expected. Optimizing these routes could improve delivery performance.

VII. Recommendations

Based on the findings of the analysis, several recommendations can be made to improve logistics efficiency.First, organizations should evaluate shipping routes that consistently show longer delivery times and consider optimizing these routes to reduce delays.Second, businesses should prioritize faster shipping modes for high-priority orders or regions where delays are common.Third, logistics teams should focus on improving delivery performance in cities where shipping delays occur frequently.Fourth, inventory management strategies should focus on high-performing products to ensure that popular products are readily available for distribution.Finally, organizations should implement continuous monitoring of logistics metrics using dashboards and automated analytics systems.

VIII. Conclusion

This project demonstrates the value of data analytics in improving logistics and supply chain management. By analyzing shipping data, organizations can identify inefficiencies in delivery routes, understand product demand patterns, and improve overall distribution performance.The insights generated from this analysis highlight the importance of monitoring shipping performance across different regions and transportation methods. With the help of interactive dashboards and data-driven decision-making, businesses can optimize their logistics networks and enhance customer satisfaction.Future work could extend this analysis by incorporating additional variables such as transportation cost, warehouse locations, and seasonal demand patterns. Integrating machine learning techniques could also help predict delivery delays and optimize route planning.

References

1. Pandas Development Team, “Pandas: Python Data Analysis Library,” Available: https://pandas.pydata.org/
2. Streamlit Inc., “Streamlit Documentation,” Available: https://docs.streamlit.io/
3. W. McKinney, Python for Data Analysis, O’Reilly Media.

4. Matplotlib Development Team, “Matplotlib Documentation,” Available: https://matplotlib.org/
5. Seaborn Documentation, Available: https://seaborn.pydata.org/
6. GeeksforGeeks, “Introduction to Pandas in Python,” Available: https://www.geeksforgeeks.org/pandas/introduction-to-pandas-in-python/
7. DataCamp, “Streamlit Tutorial,” Available: https://www.datacamp.com/tutorial/streamlit