Logistic & Delivery Time Optimization Report

Project Title: Logistic & Delivery Time Optimization

Prepared By: Hitesh Ananth.UC

Internship Role: Data Analyst

Tools & Technologies: Python, Power BI, Pandas, NumPy

Dataset: Logistic_deliver_data

Abstract:

- Efficient logistics and timely delivery are critical components of modern supply chain management. Delays, high transportation costs, and operational inefficiencies can significantly affect customer satisfaction and business performance.
- This project focuses on analyzing logistics and delivery data to identify key factors influencing delivery time, delays, and shipping costs. Using Python for data preprocessing and Power BI for visualization, the project transforms raw logistics data into actionable insights.
- The outcomes of this analysis help optimize transport mode selection, improve warehouse efficiency, and support data-driven decision-making.

Introduction:

- Logistics plays a vital role in connecting suppliers, warehouses, and customers. As businesses scale, managing delivery timelines and controlling logistics costs become increasingly challenging. Real-world logistics data is often unstructured, inconsistent, and contains missing or incorrect information, making preprocessing an essential step before analysis.
- This project aims to analyze logistics and delivery operations using historical data. By cleaning and processing the data, identifying performance gaps, and visualizing insights, the project provides a structured approach to logistics optimization.

Problem Statement:

Organizations face several challenges in logistics operations, including:

- Frequent delivery delays
- Rising shipping costs
- Inefficient transport mode selection
- Inconsistent warehouse performance
- Lack of data-driven visibility into logistics operations

Without proper data analysis, these issues remain unidentified, leading to poor decision-making. This project addresses these challenges through systematic data analysis and visualization.

Objectives of the Project:

The key objectives of this project are:

- To preprocess and clean raw logistics data for accurate analysis
- To calculate delivery time and identify delayed shipments
- To analyze delivery performance across cities, warehouses, and transport modes
- To identify major reasons for delivery delays
- To visualize key logistics metrics using Power BI
- To provide actionable recommendations for logistics optimization

Dataset Description:

The dataset represents logistics and delivery operations for a supply chain network.

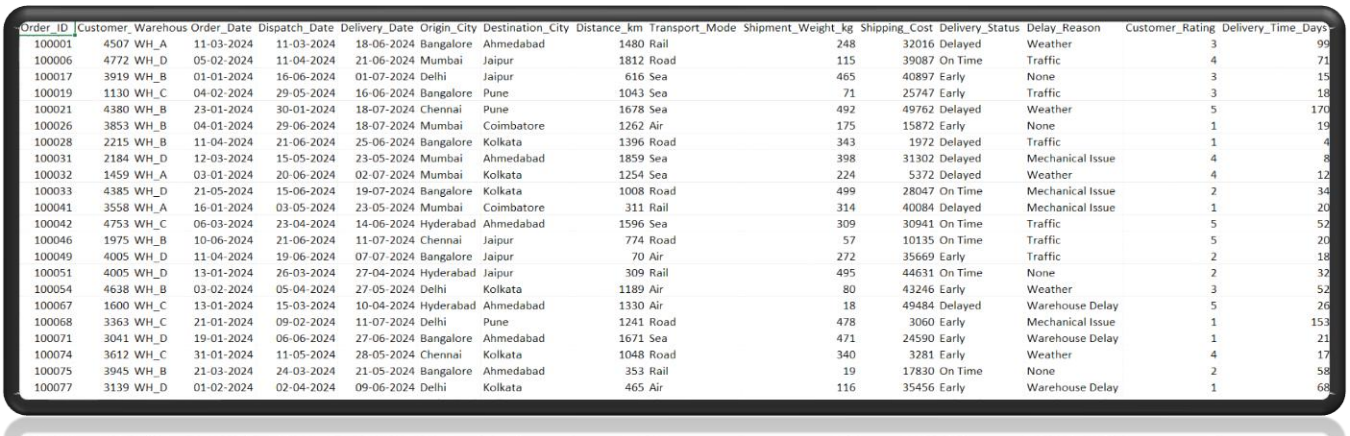
Dataset Characteristics:

- Total records: Approximately 2,500
- Total columns: 15 (additional derived columns added during preprocessing)

Key Attributes:

- Order_ID
- Customer_ID
- Warehouse_ID
- Origin_City
- Destination_City
- Order_Date
- Dispatch_Date
- Delivery_Date
- Distance_km
- Transport_Mode
- Shipment_Weight_kg
- Shipping_Cost
- Delivery_Status

- Delay_Reason
- Customer_Rating



Order_ID	Customer_	Warehous	Order_Date	Dispatch_Date	Delivery_Date	Origin_City	Destination_City	Distance_km	Transport_Mode	Shipment_Weight_kg	Shipping_Cost	Delivery_Status	Delay_Reason	Customer_Rating	Delivery_Time_Days
100001	4507	WH_A	11-03-2024	11-03-2024	18-06-2024	Bangalore	Ahmedabad	1480	Rail	248	32016	Delayed	Weather	3	99
100006	4772	WH_D	05-02-2024	11-04-2024	21-06-2024	Mumbai	Jaipur	1812	Road	115	39087	On Time	Traffic	4	71
100017	3919	WH_B	01-01-2024	16-06-2024	01-07-2024	Delhi	Jaipur	616	Sea	465	40897	Early	None	3	15
100019	1130	WH_C	04-02-2024	29-05-2024	16-06-2024	Bangalore	Pune	1043	Sea	71	25747	Early	Traffic	3	18
100021	4380	WH_B	23-01-2024	30-01-2024	18-07-2024	Chennai	Pune	1678	Sea	492	49762	Delayed	Weather	5	170
100026	3853	WH_B	04-01-2024	29-06-2024	18-07-2024	Mumbai	Coimbatore	1262	Air	175	15872	Early	None	1	19
100028	2215	WH_B	11-04-2024	21-06-2024	25-06-2024	Bangalore	Kolkata	1396	Road	343	1972	Delayed	Traffic	1	4
100031	2184	WH_D	12-03-2024	15-05-2024	23-05-2024	Mumbai	Ahmedabad	1859	Sea	398	31302	Delayed	Mechanical Issue	4	8
100032	1459	WH_A	03-01-2024	20-06-2024	02-07-2024	Mumbai	Kolkata	1254	Sea	224	5372	Delayed	Weather	4	12
100033	4385	WH_D	21-05-2024	15-06-2024	19-07-2024	Bangalore	Kolkata	1008	Road	499	28047	On Time	Mechanical Issue	2	34
100041	3558	WH_A	16-01-2024	03-05-2024	23-05-2024	Mumbai	Coimbatore	311	Rail	314	40084	Delayed	Mechanical Issue	1	20
100042	4753	WH_C	06-03-2024	23-04-2024	14-06-2024	Hyderabad	Ahmedabad	1596	Sea	309	30941	On Time	Traffic	5	52
100046	1975	WH_B	10-06-2024	21-06-2024	11-07-2024	Chennai	Jaipur	774	Road	57	10135	On Time	Traffic	5	20
100049	4005	WH_D	11-04-2024	19-06-2024	07-07-2024	Bangalore	Jaipur	70	Air	272	35669	Early	Traffic	2	18
100051	4005	WH_D	13-01-2024	26-03-2024	27-04-2024	Hyderabad	Jaipur	309	Rail	495	44631	On Time	None	2	32
100054	4638	WH_B	03-02-2024	05-04-2024	27-05-2024	Delhi	Kolkata	1189	Air	80	43246	Early	Weather	3	52
100067	1600	WH_C	13-01-2024	15-03-2024	10-04-2024	Hyderabad	Ahmedabad	1330	Air	18	49484	Delayed	Warehouse Delay	5	26
100068	3363	WH_C	21-01-2024	09-02-2024	11-07-2024	Delhi	Pune	1241	Road	478	3060	Early	Mechanical Issue	1	153
100071	3041	WH_D	19-01-2024	06-06-2024	27-06-2024	Bangalore	Ahmedabad	1671	Sea	471	24590	Early	Warehouse Delay	1	21
100074	3612	WH_C	31-01-2024	11-05-2024	28-05-2024	Chennai	Kolkata	1048	Road	340	3281	Early	Weather	4	17
100075	3945	WH_B	21-03-2024	24-03-2024	21-05-2024	Bangalore	Ahmedabad	353	Rail	19	17830	On Time	None	2	58
100077	3139	WH_D	01-02-2024	02-04-2024	09-06-2024	Delhi	Kolkata	465	Air	116	35456	Early	Warehouse Delay	1	68

Tools and Technologies Used:

- **Python** – Primary tool for data preprocessing
- **Pandas** – Data manipulation and cleaning
- **NumPy** – Numerical operations
- **Google Colab** – Development and execution environment
- **Power BI** – Data visualization and dashboard creation

Data Preprocessing Methodology:

Data preprocessing was performed using Python to ensure data accuracy and reliability.

Steps Performed:

1. **Data Loading and Inspection**
    - Imported the dataset using Pandas
    - Examined dataset structure using info() and head()
2. **Data Type Conversion**
    - Converted date columns (Order_Date, Dispatch_Date, Delivery_Date) into datetime format
3. **Handling Missing Values**
    - Filled missing values in Delay_Reason with “None”
    - Replaced missing Customer_Rating values with the median
    - Removed records with missing critical date information
4. **Duplicate Removal**
    - Identified and removed duplicate records to prevent data distortion
5. **Date Validation**
    - Ensured logical date sequence:
      Delivery_Date ≥ Dispatch_Date ≥ Order_Date
6. **Feature Engineering**

    - Created Delivery_Time_Days to calculate actual delivery duration
    - Created Is_Delayed flag to identify delayed shipments
7. **Outlier Inspection**
    - Analyzed numerical columns using summary statistics

After preprocessing, a clean and structured dataset was prepared for analysis.

Exploratory Data Analysis (EDA):

Exploratory analysis was conducted to identify patterns and trends in logistics performance.

Key Areas of Analysis:

- Average delivery time by destination city
- Delivery performance across transport modes
- Shipping cost variation by distance category
- Distribution of delivery status (Early, On Time, Delayed)
- Analysis of delay reasons
- Warehouse-wise delivery performance

EDA helped uncover inefficiencies and performance variations across logistics operations.

Dashboard and Visualization (Power BI):

An interactive Power BI dashboard was developed to visualize key logistics metrics.

Dashboard Components:

- **KPI Cards**
    - Total Orders
    - Average Delivery Time
    - Delay Percentage
- **Visual Analysis**
    - Average delivery time by destination city
    - Delivery performance by transport mode
    - Shipping cost analysis by distance and transport mode
    - Delay reason distribution
    - Delivery status overview
- **Interactive Slicers**
    - Warehouse ID
    - Transport Mode
    - Delivery Status

The dashboard enables users to filter and analyze logistics data dynamically.

Key Findings:

- Delivery time varies significantly across destination cities
- Certain transport modes lead to higher delivery times and costs
- Mechanical issues, warehouse delays, traffic, and weather are major causes of delays
- Performance inconsistencies exist across warehouses
- Most deliveries are completed on time or early, but delayed orders still affect reliability

Recommendations:

Based on the analysis, the following recommendations are proposed:

- Optimize transport mode selection based on distance and delivery urgency
- Improve warehouse operational efficiency through process standardization
- Implement proactive vehicle maintenance to reduce mechanical delays
- Redesign delivery routes for high-delay cities
- Use dashboards for continuous performance monitoring

Business Impact:

The insights generated from this project help organizations:

- Reduce delivery delays
- Control shipping costs
- Improve logistics planning
- Enhance customer satisfaction
- Enable data-driven operational decisions

Conclusion:

- This project demonstrates the importance of data preprocessing and visualization in logistics optimization. By transforming raw logistics data into a clean and structured format, meaningful insights were derived using Power BI dashboards. The analysis highlights key inefficiencies and provides actionable solutions to improve delivery performance, reduce delays, and optimize costs. Overall, the project showcases how data analytics can significantly enhance logistics decision-making.

Future Scope:

- Integration of real-time shipment tracking data
- Predictive modeling for delivery delay prediction
- Machine learning-based cost optimization
- Automation of data pipelines for live dashboards

Reference

- https://github.com/uchitesh8-web/Logistic-Delivery-Time-Optimization