

Supply Chain Analytics Project – Data Overview & Exploration

Overview of the Project:

The "Dynamic Supply Chain Logistics Analytics" project aims to **analyze** and **optimize supply chain operations using a large-scale logistics dataset**. This dataset reflects real-world supply chain activities including production, shipping, supplier performance, inventory, and customer behavior. The goal is to use **Python** and **MySQL** to extract, clean, and analyze data, then communicate findings via **dashboards** and **reporting**.

This project is conducted as part of the Data Analysis Diploma's final requirement and is meant to simulate a real freelance or industry project scenario where business decisions depend on data-driven insights.

Problem Statement:

Supply chain systems are **increasingly complex**, with issues such as **delayed shipments, inefficient supplier management, poor inventory forecasting, and rising logistics costs**. Without a data-driven approach, businesses struggle to understand **where inefficiencies lie**.

Project Goals:

- Identify key performance trends in logistics and inventory data.
- Analyze supplier performance and risk.
- Forecast demand and shipping delays using historical data.
- Build actionable dashboards for decision-makers.
- Gain hands-on experience in end-to-end data projects using MySQL and Python.

Dataset Columns:

COLUMN NAME	DESCRIPTION
TIMESTAMP	The date and time at which a logistics event or data point was recorded. This acts like the heartbeat of the entire system, allowing us to track events over time.
VEHICLE_GPS_LATITUDE	The real-time latitude location of the vehicle involved in delivery or shipping. Used for tracking movements.
VEHICLE_GPS_LONGITUDE	The corresponding longitude for the GPS coordinate — helps create a full map of the journey.
FUEL_CONSUMPTION_RATE	Measures how much fuel the vehicle is using. Useful for cost and sustainability analysis.
ETA_VARIATION_HOURS	The difference between actual and expected arrival time, measured in hours. Key to evaluating delivery performance.
TRAFFIC_CONGESTION_LEVEL	Indicates how bad traffic was during the delivery — essential for route optimization.

WAREHOUSE_INVENTORY_LEVEL	It shows how fully the warehouse was at that time — critical for demand planning and restocking decisions.
LOADING_UNLOADING_TIME	Time spent loading or unloading goods. High times might point to inefficiencies at the dock.
HANDLING_EQUIPMENT_AVAILABILITY	Whether tools like forklifts and cranes were available — important for warehouse efficiency.
ORDER_FULFILLMENT_STATUS	Whether a customer order was fulfilled on time or not. Affects satisfaction and KPIs.
WEATHER_CONDITION_SEVERITY	Describes how extreme the weather was — storms, snow, rain, etc. This influences safety and delays.
PORT_CONGESTION_LEVEL	How busy the port was — a major factor in global shipping delays.
SHIPPING_COSTS	The total cost of moving goods from origin to destination — critical for financial optimization.
SUPPLIER_RELIABILITY_SCORE	Rates how dependable a supplier is based on past performance — higher = more trustworthy.
LEAD_TIME_DAYS	Time in days between placing and receiving an order. Vital for supply chain planning.
HISTORICAL_DEMAND	How many units were previously requested — helps with forecasting future needs.
IOT_TEMPERATURE	IoT sensor data reporting the temperature during shipping — essential for perishable goods.
CARGO_CONDITION_STATUS	A status update on whether the cargo arrived in good condition or not.
ROUTE_RISK_LEVEL	How risky the transportation route was, considering weather, traffic, theft, etc.
CUSTOMS_CLEARANCE_TIME	Time spent waiting for customs approval. Bottlenecks here can slow down international shipments.
DRIVER_BEHAVIOR_SCORE	Score based on the driver's performance — speed, braking, safety compliance, etc.
FATIGUE_MONITORING_SCORE	IoT-driven estimate of driver tiredness — higher fatigue increases risk.
DISRUPTION_LIKELIHOOD_SCORE	Predicts the chances of a supply chain disruption. Preventative planning relies on this.
DELAY_PROBABILITY	Probability that a shipment will be delayed based on all real-time and historical factors.
RISK_CLASSIFICATION	Categorize the shipment or route into risk levels (e.g., Low, Medium, High).
DELIVERY_TIME_DEVIATION	The actual difference between scheduled and actual delivery time. Like ETA variation, but from a scheduling standpoint.

Task Assignment & Roles

Team Member	Main Task	%	Secondary task	%
Ziad Khaled Masoud	Build Data Model, Data Cleaning and Preprocessing	100%	Visualization Dashboard and Final Presentation	Not started
Abdelrhman Ahmed	Analysis Questions Phase	100%	Visualization Dashboard and Final Presentation	Not started
Sama Mahmoud	Analysis Questions Phase	100%	Forecasting Questions Phase	Pending
Seif El-Deen Ahmed	Forecasting Questions Phase	In progress	Analysis Questions Phase	Pending
Mohamed Hymeda	Forecasting Questions Phase	85%	Analysis Questions Phase	Not started
Renah Romany	Visualization Dashboard and Final Presentation	Not started	Build Data Model, Data Cleaning and Preprocessing	100%
Mohamed Rafaat	Visualization Dashboard and Final Presentation	Not started	Forecasting Questions Phase	Pending