

# OVERVIEW

This dataset captures a comprehensive set of logistics and supply-chain operations collected from a logistics network in Southern California. It spans January 2021 to January 2024 and covers transportation, warehouse management, route planning, and real-time monitoring. Records are hourly, reflecting conditions in urban areas and transport corridors with high traffic and dynamic operational challenges.

## SOURCES & MODALITY

- Data sources: GPS tracking systems, IoT sensors, warehouse management systems (WMS), and external data providers.
- Transportation modes: trucks, drones, and rail.
- Privacy: data is anonymized and processed to preserve utility for analysis.

## COVERAGE

- Temporal window (declared): January 2021 – January 2024
- Temporal window (auto-detected from file): 2021-01-01 to 2024-08-29
- Geospatial: Southern California (urban areas and transport corridors)
- Resolution: Hourly observations

## FILE & LICENSE

- File: `dynamic_supply_chain_logistics_dataset.csv`
- License: CC0 (Public Domain)
- Expected update frequency: Not specified
- Tags: logistics; supply chain; transportation; IoT; routing; risk management; warehouse; Southern California

# FEATURES (INPUTS)

1. Timestamp — datetime (hourly). Date/time record of the observation.
2. Vehicle GPS Latitude — float (degrees). Vehicle position latitude.
3. Vehicle GPS Longitude — float (degrees). Vehicle position longitude.
4. Fuel Consumption Rate — float (liters/hour). Vehicle fuel consumption rate.
5. ETA Variation (hours) — float (hours). Estimated vs actual arrival time difference.
6. Traffic Congestion Level — float (0–10). Congestion affecting the route.
7. Warehouse Inventory Level — float (units). Current inventory level.
8. Loading/Unloading Time — float (hours). Time for loading/unloading.
9. Handling Equipment Availability — int/binary (0/1). Forklifts/equipment availability.
10. Order Fulfillment Status — int/binary (0/1). Whether order was fulfilled on time.
11. Weather Condition Severity — float (0–1). Weather severity affecting operations.
12. Port Congestion Level — float (0–10). Congestion level at port.
13. Shipping Costs — float (USD). Costs associated with shipping operations.
14. Supplier Reliability Score — float (0–1). Supplier reliability indicator.
15. Lead Time (days) — float (days). Average supplier lead time.
16. Historical Demand — float (units). Historical demand for services.
17. IoT Temperature — float (°C). Temperature from IoT sensors.
18. Cargo Condition Status — int/binary (0/1). Cargo condition (0=poor, 1=good).
19. Route Risk Level — float (0–10). Risk level for a logistics route.
20. Customs Clearance Time — float (hours/days). Time to clear customs (dataset-specific).
21. Driver Behavior Score — float (0–1). Driving-pattern-based behavior indicator.
22. Fatigue Monitoring Score — float (0–1). Driver fatigue indicator.

# TARGET VARIABLES (LABELS)

A) Disruption Likelihood Score — float (0–1). Likelihood of disruption.

B) Delay Probability — float (0–1). Probability of shipment delay.

C) Risk Classification — categorical {Low Risk, Moderate Risk, High Risk}. Risk class label.

D) Delivery Time Deviation — float (hours). Deviation from expected delivery time.

## NOTES

- Data is hourly and multi-source; feature scales vary (e.g., 0–1 vs 0–10 vs USD).
- Consider explicit type casting: timestamps → datetime; binary flags → {0,1}; categorical labels → ordered categories.
- For modeling, standardize/normalize numeric features and encode categoricals (e.g., one-hot/target encoding).
- Check missingness and outliers per feature; impute or filter based on task.
- When using geospatial fields, validate lat/lon ranges and consider map-matching or region filters.