

Dataset 4 – Supply Chain & Inventory Data (DC–SKU–Week)

Problem Statement: Assess supply chain performance: forecast accuracy, service levels, lost sales and the interaction between capacity, lead time and inventory decisions.

Grain (each row represents): One row per distribution center × SKU × week.

Notes: Designed with missing values, spikes in lost sales and duplicate rows to replicate real supply chain analytics challenges.

Data Dictionary:

Column Name	Type	Example	Description
week_start	Date (week start)	2024-03-11	Start date of the week for supply chain data (first half of 2024).
dc_id	String	DC_007	Distribution center identifier.
region	Category	North	Region where the DC is located.
sku	Category	Heineken_330_Can	Stock-keeping unit (e.g. Heineken_330_Can, Heineken_650_Bottle).
forecast_demand_cases	Numeric (integer)	800	Forecasted demand for that DC/SKU/week in cases. Contains some missing values.
actual_demand_cases	Numeric (integer)	780	Actual demand observed for that DC/SKU/week in cases.
opening_inventory_cases	Numeric (integer)	950	Inventory level (cases) at the start of the week.
production_cases	Numeric (integer)	520	Production volume allocated to this DC/SKU in the week (cases).

inbound_cases	Numeric (integer)	300	Inbound shipments received from breweries or other DCs during the week (cases).
closing_inventory_cases	Numeric (integer)	990	Inventory at the end of the week after satisfying demand and adding production/inbound.
lost_sales_cases	Numeric (integer)	20	Estimated lost sales in cases due to stock-outs or insufficient supply. Includes some large outliers.
service_level_pct	Numeric (float)	0.97	Service level for the week = $1 - (\text{lost_sales_cases} / \text{actual_demand_cases})$. Recomputed for outlier weeks.
avg_lead_time_days	Numeric (float)	5.2	Average replenishment lead time in days for that DC/SKU. Contains some missing values.
capacity_utilization_pct	Numeric (float)	0.82	Capacity utilization of the DC (0–1+). Near 1 indicates the DC is running close to full capacity. Contains some missing values.
expedite_flag	Numeric (0/1)	1	1 if shipments were expedited due to high capacity utilization and low closing inventory, else 0.