

Hickathon #4 - Description of the datasets

1. Supply chain sales dataset (train-data.csv)

This is the main dataset you were given for sales/demand forecasting. It is called **train-data.csv** on HFactory.

The dataset contains information on where the sale was made, what type of product was sold, what type of customer purchased the product (with different levels of segmentation/hierarchy), the current lifecycle status of the product and the number of sales made, aggregated on a monthly basis into four “month” variables. Variables with “proxy” were anonymised for privacy concerns.

The “month” variables (*Month 1*, *Month 2*, *Month 3*, *Month 4*) don’t represent the same month for each row. To get the exact month/year for each row, you can look at the *Date* variable. (See the description of *Month 1* below for an example)

Variable	Description
Region	Region where the sales were made
Country	Country where the sales were made
Site	The depot distribution center where finished are stocked and shipped to customers
Operations	Higher level to Region
Zone	Sub-level to Region
Cluster	?
Reference proxy	Lowest-level in the product segmentation hierarchy (sub-level to Strategic Product family proxy)
Strategic Product Family proxy	Sub-level to Reference proxy (second lowest product description hierarchy)
Product Line proxy	Sub-level to Division proxy
Division proxy	Highest level of product hierarchy
Customer Persona proxy	Segmentation/type of customer
Product Life cycl status	Current status of the product (EOL: End of life, NPI: new product introduction, NaN: Active lifecycle)

Date	Gives information in which “quarter”
Month 1	Number of sales made the first month of Date (for example, it represents the number of sales made in September 2020 when Date is sept-dec 2020)
Month 2	Number of sales made the second month of Date (for example, it represents the number of sales made in October 2020 when Date is sept-dec 2020)
Month 3	Number of sales made the third month of Date
Month 4	Number of sales made the four month of Date

Note: The following datasets include variables you can add to the main supply chain sales dataset to predict sales. These datasets contain information either on a monthly or yearly level and on a global or country-level. These datasets come from external sources and might need some additional cleaning.

2. Global Supply Chain Pressure Index (GSCPI_data.csv)

Source: <https://www.newyorkfed.org/research/policy/gscpi#/overview>
(Federal Reserve Bank of New York)

Description of the index:

The global supply chain pressure index (GSCPI) integrates transportation cost data and manufacturing indicators to provide a gauge of global supply chain conditions on a monthly level since January 1998.

A dataset with twenty-seven variables was used to compute this index: three country-specific supply chain variables for each of the economies in our sample (the euro area, China, Japan, South Korea, Taiwan, the U.K., and the U.S.), the two global shipping rates, and the four price indices summarizing airfreight costs between the U.S., Asia, and Europe. All these variables are corrected for demand effects to the greatest possible extent.

3. Logistics Performance Index and Population Data (LPIextend.csv)

The Logistics Performance Index reflects the perceptions of a country's logistics performance on a yearly basis. Each variable in the dataset represents the score/rank of a country in 2023.

You can find more variables related to the logistics performance of a country in 2023 here: <https://lpi.worldbank.org/international/tracking-data>

Variable	Description
Country	Country

LPI Score	Logistic Performance Index score
LPI Grouped Rank	LPI Score rank of a country
Customs Score	Efficiency of the customs clearance process
Customs Grouped Rank	Customs Score rank of a country
Infrastructure Score	Quality of the transport-related infrastructure
Infrastructure Grouped Rank	Infrastructure score rank of a country
International Shipments Score	Ease of arranging competitively priced international shipments
International Shipments Grouped Rank	International Shipments Score rank of a country
Logistics Competence and Quality Score	Quality of logistics services
Logistics Competence and Quality Grouped Rank	Logistics Competence and Quality Score rank of a country
Timeliness Score	Frequency with which shipments reach the consignee
Timeliness Grouped Rank	Timeliness Score rank of a country
Tracking and Tracing Score	Ability to track and trace consignments

Source: <https://worldpopulationreview.com/countries>

Here is a description of every variable in the population/density dataset.

Variable	Description
population (2023)	Total population in 2023
Country	Country
area	Total area (in square kilometers)
landAreaKm	Total area excluding area under inland water bodies, national claims to

	continental shelf, and exclusive economic zones (in square kilometers)
unMember	True if a country is a United Nations member, False if it isn't
netChange	?
growthRate	Growth rate of the population
worldPercentage	Percentage of the total world population
density	A measurement of population per unit land area.
densityMi	?
rank	Rank of a country based on the total population in 2023

4. World Bank Economic indicators (worldbank_economic_data.csv)

Source: <https://datatopics.worldbank.org/world-development-indicators/themes/economy.html>
(World Bank)

This dataset contains some of the economic indicators provided by the world bank. The indicators were computed yearly.

Variable	Description
Country	Country
Year	Year
Agriculture, forestry, and fishing, value added (annual % growth)	An economic indicator that measures the annual rate of growth in the value added by the agriculture, forestry, and fishing sectors of an economy
Exports of goods and services (annual % growth)	Annual growth rate of exports of goods and services based on constant local currency
Final consumption expenditure (annual % growth)	Annual percentage growth of general government final consumption expenditure based on constant local currency (see below for a definition on final consumption expenditure)
GDP (current US\$)	Annual sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products (in dollars)
Gross capital formation (annual % growth)	Annual growth rate of gross capital formation based on constant local currency (see below for a definition on gross capital formation)

Imports of goods and services (annual % growth)	Annual growth rate of imports of goods and services based on constant local currency.
Industry (including construction), value added (annual % growth)	Annual growth rate for industrial (including construction) value added based on constant local currency. It comprises value added in mining, manufacturing (also reported as a separate subgroup), construction, electricity, water, and gas.
Manufacturing, value added (annual % growth)	Annual growth rate for manufacturing value added based on constant local currency.
Services, value added (annual % growth)	Annual growth rate for value added in services based on constant local currency. It includes value added in wholesale and retail trade (including hotels and restaurants), transport, and government, financial, professional, and personal services such as education, health care, and real estate services.

Final consumption expenditure: Expenditure incurred by resident institutional units (Households, NPISHs, sector of general government) on goods or services that are used for the direct satisfaction of individual needs or wants or the collective needs of members of the community.

Gross capital formation: Sum of all goods produced and services in an economy over a specific period, minus the value of goods and services used up in production. It represents the net increase in physical or financial assets for a country during a particular time frame.

5. World Bank Inflation (worldbank_inflation_data.csv)

Source: <https://www.worldbank.org/en/research/brief/inflation-database>
(World Bank)

This dataset contains two inflation measurements provided by the World Bank given on a monthly level.

Variable	Description
Country	Country
Year-Month	Year/month of the measurement
Energy Price Index	Average change in the prices of energy goods and services over a one-year period
Headline Consumer Price Index	Raw inflation figure reported through the Consumer Price Index (CPI) that is released monthly