

EU estimated agricultural balance sheets

Data source

EU estimated agricultural balance sheets

The **EU Estimated Agricultural Balance Sheets dataset** is published by the Joint Research Centre through the European Union Data Portal, the official provider of publishing services for all EU institutions, bodies, and agencies. This dataset is open-source, and we can rely on it as a trustworthy source of data.

I was inspired to choose this dataset because, since I migrated to Germany a few years ago, I have encountered a wide variety of grocery and dairy products imported from all over Europe. This sparked my curiosity about the main exporting countries for these commodities within Europe and which countries are the leading producers and consumers in specific categories. I anticipate gaining valuable insights from this data.

Data collection method

The data set combines trade data from Eurostat COMEXT database and DG AGRI's short-term outlook for EU agricultural markets to provide historic data on agricultural production, trade and an estimate of apparent use.

DG AGRI, in the Short-term outlook, prepares regularly EU balance sheets for the main agricultural commodities, as well as production, area and yield figures at Member State level

The dataset is updated 3 times per year, at every new release of the DG AGRI Short-term outlook. The latest update is on 13rd Jul 2023.

Data content

This dataset includes annual production and consumption data for 78 agricultural commodities in each of the 27 EU member countries, covering both import and export figures within and outside European countries. The data has been provided annually since 2002.

Link

<https://data.europa.eu/data/datasets/33243e5e-44a1-4b43-9444-31d64dc7921f?locale=en>

https://agridata.ec.europa.eu/extensions/DataPortal/agricultural_markets.html

Demographic indicators by region, subregion and country, annually for 1950-2100

Demographic indicators by region, subregion and country, annually for 1950-2100 is prepared and published by the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. It presents population estimates from 1950 to the present for 237 countries or areas. This dataset is open-source, and we can rely on it as a trustworthy source of data.

Data collection method

The data collection method involves analyzing historical demographic trends by considering data from 1,758 national population censuses conducted between 1950 and 2022, along with information from vital registration systems and 2,890 nationally representative sample surveys. This assessment also includes population projections up to 2100 for global, regional, and national levels.

Data content

This dataset comprises three worksheets: 'Estimates,' 'Medium Variant,' and 'Notes.' Both the 'Estimates' and 'Medium Variant' worksheets contain 65 columns each, which include information about countries, years, populations, fertility, mortality, and migration. The 'Estimates' worksheet presents estimated data for the years 1950 to 2021, while the 'Medium Variant' worksheet provides population projection data for the years 2022 to 2100. The data required from this dataset for this analysis is the annual 'Total Population as of 1 January (thousands)' for each country."

Link

<https://population.un.org/wpp/Download/Standard/MostUsed/>

EU Countries area

EU Countries area data is acquired from Central Intelligence Agency (CIA) website which is a U.S. government agency that provides objective intelligence on foreign countries and global issues to the president, the National Security Council, and other policymakers. This data source is reliable and the data is up to date.

Data collection method

Since 2002, there have been no incidents indicating changes in the land area of each European country. Therefore, the data collection method can remain undisclosed.

Data content

I extracted this dataset from their website and pasted it into a new Excel file. The dataset contains two columns: 'Country Name' and 'Area (square kilometers).'

Link

<https://www.cia.gov/the-world-factbook/countries/european-union/>

Data profile

Data cleaning

Column name	Changes	Comment / Reason
Country (NUTS Code)	Drop the column	Unnecessary column
Country (Description)	Drop the column	Unnecessary column
Commodity (Primary product)	Rename 'WINE' in the commodity column (588 rows) as 'WINE (VINEYARD/GRAPES FOR WINE)' and 'TOTAL OLIVE OIL' in the commodity column (532 rows) as 'TOTAL OLIVE OIL (OLIVES FOR OIL)' and drop the column	Unnecessary column
Apparent use (THOUSAND TONS)	Drop the column	Unnecessary column
Apparent use (THOUSAND HECTOLITRES)	Drop the column	Unnecessary column
Apparent use (THOUSAND TONS (CARCASS WEIGHT EQ.))	Drop the column	Unnecessary column
Apparent use (THOUSAND TONS (RETAIL WEIGHT EQ.))	Drop the column	Unnecessary column
Apparent use (THOUSAND TONS (MILK EQ.))	Drop the column	Unnecessary column
Apparent per capita use (KG/PERSON)	Drop the column	Unnecessary column
Apparent per capita use (LITRES/PERSON)	Drop the column	Unnecessary column
Apparent per capita use (KG (RETAIL WEIGHT EQ.)/PERSON)	Drop the column	Unnecessary column
Apparent per capita use (KG (CARCASS WEIGHT EQ.)/PERSON)	Drop the column	Unnecessary column
Beginning stocks (THOUSAND TONS)	Drop the column	Unnecessary column
Beginning stocks (THOUSAND HECTOLITRES)	Drop the column	Unnecessary column
Carcass weight (KG/HEAD)	Drop the column	Unnecessary column
Deliveries (THOUSAND TONS)	Drop the column	Unnecessary column
Ending stocks (THOUSAND TONS)	Drop the column	Unnecessary column
Ending stocks (THOUSAND HECTOLITRES)	Drop the column	Unnecessary column
Exports (Extra-EU) (THOUSAND TONS (MILK EQ.))	Drop the column	Unnecessary column
Exports (Intra-EU) (THOUSAND TONS (MILK EQ.))	Drop the column	Unnecessary column
Imports (Extra-EU) (THOUSAND TONS (MILK EQ.))	Drop the column	Unnecessary column
Imports (Intra-EU) (THOUSAND TONS (MILK EQ.))	Drop the column	Unnecessary column
Gross indigenous production (THOUSAND TONS (CARCASS WEIGHT EQ.))	Drop the column	Unnecessary column
Harvested area (THOUSAND HECTARES)	Drop the column	Unnecessary column
Production (THOUSAND HECTOLITRES)	Drop the column	Unnecessary column
Production (THOUSAND TONS (MILK EQ.))	Drop the column	Unnecessary column
Production (from dairy cows) (THOUSAND TONS)	Drop the column	Unnecessary column
Production of primary product (THOUSAND TONS)	Drop the column	Unnecessary column
Self sufficiency (%)	Drop the column	Unnecessary column
Yield (crops) (TONNE/HA)	Drop the column	Unnecessary column
Yield (HECTOLITRES PER HECTARE)	Drop the column	Unnecessary column
Yield (animals) (KG/HEAD)	Drop the column	Unnecessary column
Yield of primary product (TONNE/HA)	Drop the column	Unnecessary column
Yield of secondary product (KG PER TONNE OF PRIMARY PRODUCT)	Drop the column	Unnecessary column

Consistency check

Missing values

1. Remove 392 rows which have '2023' value in the year column.
2. Remove 1,092 rows of 'OTHER DAIRY PRODUCTS' and 'OTHER FRESH DAIRY PRODUCTS'.
3. Impute 0 value for 21 rows in Population (THOUSAND HEAD) and Production (THOUSAND TONS) columns which have country 'Luxembourg', commodity 'RAW MILK'.
4. Impute 0 value for 21 rows in Population (THOUSAND HEAD) column which have country 'Luxembourg', commodity 'DAIRY COWS'.
5. Impute data of 8 rows in Exports (Extra-EU) (THOUSAND TONS), Exports (Intra-EU) (THOUSAND TONS), Imports (Extra-EU) (THOUSAND TONS), and Imports (Intra-EU) (THOUSAND TONS) columns which have country 'Croatia, Cyprus, France, Greece, Italy, Malta, Portugal, Spain', commodity 'ORANGES', year '2022', by calculating average values of the latest 3 years of each mentioned columns.
6. Impute data of 27 rows for 27 countries in Production (THOUSAND TONS) column which have commodity 'RAW MILK', year '2022' by calculating average values of 3 latest years of production (thousand tons) per production (thousand head) multiply by production (thousand head) of year 2022.

Duplicates

No duplicate found.

Inconsistent data

1. Remove 1,743 rows which have country 'EU'.
2. Create new 27 rows for 27 countries and impute data in Exports (Extra-EU) (THOUSAND TONS), Exports (Intra-EU) (THOUSAND TONS), Imports (Extra-EU) (THOUSAND TONS), and Imports (Intra-EU) (THOUSAND TONS) columns which have commodity 'FRESH ORANGES', year '2022', by calculating average values of the latest 3 years of each mentioned columns.
3. Create new 27 rows for 27 countries and impute data in Exports (Extra-EU) (THOUSAND TONS), Exports (Intra-EU) (THOUSAND TONS), Imports (Extra-EU) (THOUSAND TONS), and Imports (Intra-EU) (THOUSAND TONS) columns which have commodity 'ORANGE JAMS', year '2022', by calculating average values of the latest 3 years of each mentioned columns.
4. Create new 27 rows for 27 countries and impute data in Exports (Extra-EU) (THOUSAND TONS), Exports (Intra-EU) (THOUSAND TONS), Imports (Extra-EU) (THOUSAND TONS), and Imports (Intra-EU) (THOUSAND TONS) columns which have commodity 'ORANGE JUICES', year '2022', by calculating average values of the latest 3 years of each mentioned columns.
5. Create new 27 rows for 27 countries and impute data in Exports (Extra-EU) (THOUSAND TONS), Exports (Intra-EU) (THOUSAND TONS), Imports (Extra-EU) (THOUSAND TONS), and Imports (Intra-EU) (THOUSAND TONS) columns which have commodity 'ORANGE MIXTURES', year '2022', by calculating average values of the latest 3 years of each mentioned columns.
6. Create new 19 rows for 19 countries (except Croatia, Cyprus, France, Greece, Italy, Malta, Portugal, Spain) and impute data in Exports (Extra-EU) (THOUSAND TONS), Exports (Intra-EU) (THOUSAND TONS), Imports (Extra-EU) (THOUSAND TONS), and Imports (Intra-EU) (THOUSAND TONS) columns which have commodity 'ORANGES', year '2022', by calculating average values of the latest 3 years of each mentioned columns.
7. Create new 27 rows for 27 countries with year '2002' and impute 0 value in Exports (Extra-EU) (THOUSAND TONS), Exports (Intra-EU) (THOUSAND TONS), Imports (Extra-EU) (THOUSAND TONS), and Imports (Intra-EU) (THOUSAND TONS) columns which have commodity 'EXTRA VIRGIN AND

VIRGIN OTHER THAN LAMPANTE OLIVE OIL, EXTRA VIRGIN OLIVE OIL, LAMPANTE OLIVE OIL, OTHER OLIVE OIL, TOTAL OLIVE OIL (OLIVES FOR OIL), VIRGIN OTHER THAN LAMPANTE OLIVE OIL'.

8. Create new 27 rows for 27 countries with year '2003' and impute 0 value in Exports (Extra-EU) (THOUSAND TONS), Exports (Intra-EU) (THOUSAND TONS), Imports (Extra-EU) (THOUSAND TONS), and Imports (Intra-EU) (THOUSAND TONS) columns which have commodity 'EXTRA VIRGIN AND VIRGIN OTHER THAN LAMPANTE OLIVE OIL, EXTRA VIRGIN OLIVE OIL, LAMPANTE OLIVE OIL, OTHER OLIVE OIL, TOTAL OLIVE OIL (OLIVES FOR OIL), VIRGIN OTHER THAN LAMPANTE OLIVE OIL'.
9. Create new 27 rows for 27 countries with year '2022' and impute data in Exports (Extra-EU) (THOUSAND TONS), Exports (Intra-EU) (THOUSAND TONS), Imports (Extra-EU) (THOUSAND TONS), and Imports (Intra-EU) (THOUSAND TONS) columns which have commodity 'EXTRA VIRGIN AND VIRGIN OTHER THAN LAMPANTE OLIVE OIL, EXTRA VIRGIN OLIVE OIL, LAMPANTE OLIVE OIL, OTHER OLIVE OIL, TOTAL OLIVE OIL (OLIVES FOR OIL), VIRGIN OTHER THAN LAMPANTE OLIVE OIL' by calculating average 3 latest years values.
10. Remove 6,006 data in the Production (THOUSAND TONS) column of commodity type 'DAIRY'.
11. Remove 566 data in the Production (THOUSAND TONS) column of commodity 'APPLES'.
12. Remove 160 data in the Production (THOUSAND TONS) column of commodity 'ORANGES'.
13. Remove 176 data in the Production (THOUSAND TONS) column of commodity 'PEACHES AND NECTARINES'.
14. Remove 513 data in the Production (THOUSAND TONS) column of commodity 'TOTAL OLIVE OIL (OLIVES FOR OIL)'.
15. Remove 351 data in the Production (THOUSAND TONS) column of commodity 'TOMATOES'.
16. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'BOVINE'.
17. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'PIGS'.
18. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'SHEEP AND GOATS'.
19. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'BARLEY'.
20. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'DURUM WHEAT'.
21. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'GRAIN MAIZE'.
22. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'OATS'.
23. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'OTHER CEREALS'.
24. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'RYE'.
25. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'SOFT WHEAT'.
26. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'SORGHUM'.
27. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'TOTAL CEREALS'.
28. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'TRITICALE'.
29. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'RAPESEED'.
30. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'SOYA BEAN'.
31. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'SUNFLOWER'.
32. Remove 567 data in the Production (THOUSAND TONS) column of commodity 'TOTAL OILSEEDS'.

Data profiling

The data is presented on a per-commodity basis, and after the data cleaning process, the commodities can be categorized into five groups based on their common column names.;

1. A group with data in exports and imports (thousand tons) columns

Commodity (Type)	Commodity	Exports (Extra-EU) (THOUSAND TONS)	Exports (Intra-EU) (THOUSAND TONS)	Imports (Extra-EU) (THOUSAND TONS)	Imports (Intra-EU) (THOUSAND TONS)
APPLES	APPLE JUICE	✓	✓	✓	✓
APPLES	APPLES	✓	✓	✓	✓
APPLES	DRIED APPLES	✓	✓	✓	✓
APPLES	FRESH APPLES	✓	✓	✓	✓
ORANGES	FRESH ORANGES	✓	✓	✓	✓
ORANGES	ORANGE JAMS	✓	✓	✓	✓
ORANGES	ORANGE JUICES	✓	✓	✓	✓
ORANGES	ORANGE MIXTURES	✓	✓	✓	✓
ORANGES	ORANGES	✓	✓	✓	✓
PEACHES AND NECTARINES	DRIED PEACHES AND NECTARINES	✓	✓	✓	✓
PEACHES AND NECTARINES	FRESH PEACHES AND NECTARINES	✓	✓	✓	✓
PEACHES AND NECTARINES	PEACHES AND NECTARINES	✓	✓	✓	✓
PEACHES AND NECTARINES	PREPARED PEACHES AND NECTARINES	✓	✓	✓	✓
TOMATOES	DRIED TOMATOES	✓	✓	✓	✓
TOMATOES	FRESH TOMATOES	✓	✓	✓	✓
TOMATOES	FROZEN TOMATOES	✓	✓	✓	✓
TOMATOES	KETCHUP	✓	✓	✓	✓
TOMATOES	PEELED AND UNPEELED TOMATOES	✓	✓	✓	✓
TOMATOES	TOMATO JUICE	✓	✓	✓	✓
TOMATOES	TOMATO PASTE	✓	✓	✓	✓
TOMATOES	TOMATOES	✓	✓	✓	✓
DAIRY	ACIDIFIED MILK	✓	✓	✓	✓
DAIRY	BUTTER (80-90% FAT)	✓	✓	✓	✓
DAIRY	BUTTERMILK	✓	✓	✓	✓
DAIRY	CREAM FOR DIRECT CONSUMPTION	✓	✓	✓	✓
DAIRY	DRINKING MILK	✓	✓	✓	✓
DAIRY	FRESH DAIRY PRODUCTS	✓	✓	✓	✓
DAIRY	SKIMMED MILK POWDER	✓	✓	✓	✓
DAIRY	TOTAL BUTTER	✓	✓	✓	✓
DAIRY	TOTAL CHEESE	✓	✓	✓	✓
DAIRY	WHEY POWDER	✓	✓	✓	✓
DAIRY	WHOLE MILK POWDER	✓	✓	✓	✓
SPECIALISED CROPS	EXTRA VIRGIN AND VIRGIN OTHER THAN LAMPANTE OLIVE OIL	✓	✓	✓	✓
SPECIALISED CROPS	EXTRA VIRGIN OLIVE OIL	✓	✓	✓	✓
SPECIALISED CROPS	LAMPANTE OLIVE OIL	✓	✓	✓	✓
SPECIALISED CROPS	OTHER OLIVE OIL	✓	✓	✓	✓
SPECIALISED CROPS	TOTAL OLIVE OIL (OLIVES FOR OIL)	✓	✓	✓	✓
SPECIALISED CROPS	VIRGIN OTHER THAN LAMPANTE OLIVE OIL	✓	✓	✓	✓
CEREAL	BARLEY	✓	✓	✓	✓
CEREAL	DURUM WHEAT	✓	✓	✓	✓
CEREAL	GRAIN MAIZE	✓	✓	✓	✓
CEREAL	OATS	✓	✓	✓	✓
CEREAL	OTHER CEREALS	✓	✓	✓	✓
CEREAL	RYE	✓	✓	✓	✓
CEREAL	SOFT WHEAT	✓	✓	✓	✓
CEREAL	SORGHUM	✓	✓	✓	✓
CEREAL	TOTAL CEREALS	✓	✓	✓	✓
CEREAL	TRITICALE	✓	✓	✓	✓
OILSEED	RAPESEED	✓	✓	✓	✓
OILSEED	SOYA BEAN	✓	✓	✓	✓
OILSEED	SUNFLOWER	✓	✓	✓	✓
OILSEED	TOTAL OILSEEDS	✓	✓	✓	✓

4. A group with data in population (thousand head) column

Commodity (Type)	Commodity	Population (THOUSAND HEAD)
MEAT	BREEDING SOWS	✓
MEAT	DAIRY COWS	✓
MEAT	GOATS	✓
MEAT	OTHER BOVINE	✓
MEAT	OTHER PIGS	✓
MEAT	SHEEP	✓
MEAT	SUCKLER COWS	✓

5. A group with data in population and production (thousand head) columns

Commodity (Type)	Commodity	Population (THOUSAND HEAD)	Production (THOUSAND TONS)
RAW MILK	RAW MILK	✓	✓

Variables and data types

Variables	Data Types			
	Time-variant / -invariant	Structured / Unstructured	Qualitative / Quantitative	Qualitative: Nominal / Ordinal Quantitative: Discrete / Continuous
Country	Time-invariant	Unstructured	Qualitative	Nominal
Commodity	Time-invariant	Unstructured	Qualitative	Nominal
Commodity (Type)	Time-invariant	Unstructured	Qualitative	Nominal
Year	Time-variant	Structured	Quantitative	Discrete
Exports (Extra-EU) (THOUSAND TONS)	Time-variant	Unstructured	Quantitative	Continuous
Exports (Intra-EU) (THOUSAND TONS)	Time-variant	Unstructured	Quantitative	Continuous
Exports (Extra-EU) (THOUSAND HECTOLITRES)	Time-variant	Unstructured	Quantitative	Continuous
Exports (Intra-EU) (THOUSAND HECTOLITRES)	Time-variant	Unstructured	Quantitative	Continuous
Imports (Extra-EU) (THOUSAND TONS)	Time-variant	Unstructured	Quantitative	Continuous
Imports (Intra-EU) (THOUSAND TONS)	Time-variant	Unstructured	Quantitative	Continuous
Imports (Extra-EU) (THOUSAND HECTOLITRES)	Time-variant	Unstructured	Quantitative	Continuous
Imports (Intra-EU) (THOUSAND HECTOLITRES)	Time-variant	Unstructured	Quantitative	Continuous
Live exports (Extra-EU) (THOUSAND TONS (CARCASS WEIGHT EQ.))	Time-variant	Unstructured	Quantitative	Continuous
Live exports (Intra-EU) (THOUSAND TONS (CARCASS WEIGHT EQ.))	Time-variant	Unstructured	Quantitative	Continuous
Live imports (Extra-EU) (THOUSAND TONS (CARCASS WEIGHT EQ.))	Time-variant	Unstructured	Quantitative	Continuous
Live imports (Intra-EU) (THOUSAND TONS (CARCASS WEIGHT EQ.))	Time-variant	Unstructured	Quantitative	Continuous
Meat exports (Extra-EU) (THOUSAND TONS (CARCASS WEIGHT EQ.))	Time-variant	Unstructured	Quantitative	Continuous
Meat exports (Intra-EU) (THOUSAND TONS (CARCASS WEIGHT EQ.))	Time-variant	Unstructured	Quantitative	Continuous
Meat imports (Extra-EU) (THOUSAND TONS (CARCASS WEIGHT EQ.))	Time-variant	Unstructured	Quantitative	Continuous
Meat imports (Intra-EU) (THOUSAND TONS (CARCASS WEIGHT EQ.))	Time-variant	Unstructured	Quantitative	Continuous
Net production (THOUSAND HEAD)	Time-variant	Unstructured	Quantitative	Discrete
Net production (THOUSAND TONS (CARCASS WEIGHT EQ.))	Time-variant	Unstructured	Quantitative	Continuous
Population (THOUSAND HEAD)	Time-variant	Unstructured	Quantitative	Discrete
Production (THOUSAND TONS)	Time-variant	Unstructured	Quantitative	Continuous

Limitations

An estimation of the annual consumption by Member States was unavailable after Eurostat stopped collecting and publishing balance sheets. This dataset partially addresses this gap by providing a basic estimate of apparent use, without detailing specific categories such as food, feed, or processing, nor accounting for losses or changes in stocks. This apparent use cannot be taken at face value as the consumption level, because the calculation relies on trade statistics between EU Member States which are less reliable than extra-EU trade figures.

Ethics

This data is open-source and contains no privacy or personal data, so there are no ethical issues with this data.

Questions to explore

1. In the category of tomatoes, which products see the highest trade volumes in terms of both exports and imports?
2. Among EU countries, which ones lead in tomato exports?
3. What are the primary factors that lead European Union countries to import tomatoes from outside the EU?
4. Is the export of Livestock in live form greater than in the form of meat products?
5. Among EU countries, which one has the largest population of Livestock?
6. What are the primary commodities imported from countries outside the EU?
7. Do EU countries consider importing commodities from within the EU rather than from outside the EU?
8. Do all commodity exports show an annual increase?