

INDEX **BOX**

# IndexBox AI Platform

## Metadata

v. 1.0.2018

# Table of Contents

<b>1. INTRODUCTION</b>	<b>3</b>
<b>2. PRODUCT COVERAGE</b>	<b>3</b>
<b>3. DATA COVERAGE</b>	<b>4</b>
<b>4. RAW, MIRROR AND NORMALIZED DATA</b>	<b>6</b>
<b>5. UNITS OF MEASUREMENT</b>	<b>7</b>
<b>6. REGIONAL COVERAGE</b>	<b>7</b>
<b>7. RESEARCH METHODOLOGY AND AI PLATFORM</b>	<b>9</b>
<b>8. DATA SOURCES</b>	<b>10</b>

## 1. Introduction

This IndexBox AI Platform Data Service provides access to key data regarding key market indicators for a diverse range of products, categorised by region and country. The primary aim of this service is to help your company source market data, perform basic market analysis and prepare customized market reports, with a view to further expanding your business.

Here the latest data on consumption, production, imports and exports is presented, as well as, where applicable, data with regard to harvested areas, the number of producing animals, yield figures and supplementary indicators (per capita consumption, net exports etc.). The data is presented for the period from 2007 to 2016 and the market size is forecast for the immediate term.

The Analyze Data section includes a global market map, by region or by country; current trends are also visualized, as well as the rankings and share distribution of the key market indicators.

The Report Writing section gives access to prepared textual descriptions regarding the trend pattern of the key market indicators and a basic finished report on the market being researched can be downloaded; the IndexBox AI data platform can be used as a final completed market report, or as a data-based background for further more comprehensive research.

## 2. Product coverage

Products in the system are identified according to an HS classifier (Harmonized system); this is an international commodity classificatory developed by The World Customs Organization (WCO). The system covers about 5,000 commodity groups, each identified by a six-digit code. More than 179 countries and economies use the system. All the countries using this system collect and process international trade data in the given product classification, therefore this information can be compiled and used to analyse global markets.

The products in the system are defined in line with HS 2007 and comprise six-digits: this is to ensure data compatibility and the integrity of the dynamic series. In several cases, more general products comprising 4-digits are additionally reviewed. The product item 'poultry' for example (which includes all types of poultry meat) is defined under the code HS 0207, while the data for different types of poultry meat is accessible under their respective code (e.g., 020724 *Meat and edible offal; of turkeys, not cut in pieces, fresh or chilled*). The HS code is listed next to the product name in the product selection menu. Products can be searched for by using either the name of the product or its HS code.

### 3. Data coverage

The system presents data on the following key market indicators:

Factor Label	Factor Code	Definition
Consumption Volume	CONSQNT	estimated amount of goods available for consumption in the particular market (country, region or the world). 'CONSUMPTION VOLUME' refers to consumption in physical terms. It is calculated as follows: $CONSQNT = PRODQNT + IMPVAL - EXPVAL$
Consumption Value	CONSVL	'CONSUMPTION VALUE' refers to consumption in value terms Depending on whether it is being calculated on a country/global level, and on the data available, it can be estimated as follows: 1) Basic calculation in value terms: $CONSVL = PRODVAL + IMPVAL - EXPVAL$ 2) Estimate based on consumption volume and export/import prices: $CONSVL = CONSQNT * EXPRICE$
Production Volume	PRODQNT	estimated amount of goods produced in the particular market (country, region or the world). 'PRODUCTION VOLUME' refers to production in physical terms. For crop commodities, the term production data refers to the actual harvested crop from the areas under seed. This figure does not accommodate any potential crop losses. Production also includes the amount of the crop product sold in the market (marketed production) and the quantities consumed or used by the producers (auto-consumption). When the production data available is based on a crop cultivation period that is divided between two successive calendar years, it is standard practice to use the production data of the year that generated the bulk of production. For meat and other livestock commodities, data figures are typically based on total domestic production; figures also include data from both commercial and farm slaughter. The data is expressed in terms of dressed carcass weight, excluding offal and slaughter fats.
Production Value	PRODVAL	'PRODUCTION VALUE' refers to production in value terms Depending on whether it is being calculated on a country/global level and on the data available, it can be estimated as follows: 1) In producer prices: $PRODVAL = PRODQNT * PRODPRICE$ 2) In export prices: $PRODVAL = PRODQNT * EXPRICE$
Import Volume	IMPQNT	the volume of industry goods supplied from foreign countries to be sold on the domestic market. Both physical and value terms can apply; 'Import Volume' refers to imports in physical terms.
Import Value	IMPVAL	refers to imports in value terms. Basically, data on the import value is based on the CIF (Incoterms 2010) price, which typically includes the costs of international delivery and insurance.
Export Volume	EXPQNT	the volume or value of industry goods sold to foreign customers. Both physical and value terms can apply. 'Export Volume' refers to exports in physical terms.
Export Value	EXPVAL	refers to exports in value terms. Basically, data on the export value is based on the FOB (Incoterms 2010) price, which typically includes the costs only to the port of departure, unlike the CIF price used in the import data. Therefore, the reported imports and exports values usually do not coincide accurately.
Import Price	IMPRICE	The importer price means the average unit value of imported goods on a CIF basis, calculated as follows: $IMPRICE = IMPVAL / IMPQNT$
Export Price	EXPRICE	The exporter price in the IndexBox AI data platform means the average unit value of exported goods on an FOB basis, calculated as follows: $EXPRICE = EXPVAL / EXPQNT$
Producer Price	PRODPRICE	average unit value of goods, usually recorded on the first commercial transaction. The average producer prices are typically recorded by the statistical authorities in the reporting countries in local currencies and then converted to US dollars.

Factor Label	Factor Code	Definition
Producer Price Index	PRODINDEX	<p>an index illustrating the price change for a certain period against a basic period. It can be calculated as follows:</p> <ol style="list-style-type: none"> <li>1) A ratio between the current-year price and the basic-year price. Illustrate year-to-year change:  <math>PRODINDEX_n = PRODPRICE_n / PRODPRICE_{basic}</math></li> <li>2) A ratio between the current-year price and the one-year-past price. Illustrate year-to-year change:  <math>PRODINDEX_n = PRODPRICE_n / PRODPRICE_{n-1}</math></li> </ol>
Harvested Area	HARVAREA	<p>A concept relevant for crop commodities. Data refers to the area where the crop is produced and collected; it does not allow for areas which suffered crop failure, or losses, or that did not produce a harvest yield. Data is usually net for temporary crops and occasionally gross for permanent crops. The gross area will include uncultivated patches, footpaths, ditches, headlands, shoulders, shelterbelts, etc. All harvest yield figures are considered (should an area under seed produce more than one consecutive harvest a year); on the contrary, the area harvested will be recorded only once in the case of successive crop gathering during the year from the same standing crops. In terms of mixed and associated crops, the area sown relating to each crop should be reported separately. When the mixed crop in question refers to particular crops, generally grains, it is recommended to consider it as if it were a single crop; therefore, the area sown is recorded only for the crop reported.</p>
Producing Animals	PRODSTOCK	<p>A concept relevant for meat and other livestock commodities. The number of animals such as cattle/pig/sheep/poultry etc, which are kept on the holding or otherwise for agricultural production.</p> <p>With regard to products from slaughtered animals, e.g. meat, offals, raw fats, fresh hides and skins, PRODSTOCK refers to the number of slaughtered animals. All data shown relates to the total meat production from both commercial and farm slaughter.</p> <p>With regard to products from live animals, which include milk, eggs, honey, beeswax and fibres of animal origin, PRODSTOCK refers to the live producing population.</p>
Yield	YIELD	<p>For crop products, this refers to the harvested production per unit of harvested area.</p> <p>For livestock products, it refers to the production per unit of livestock (live or slaughtered, according to a particular product).</p> <p>Typically, yield data is not recorded but obtained by dividing the production data by the data on the harvested area or producing animals:  <math>YIELD = PRODQNT / HARVAREA</math> or <math>YIELD = PRODQNT / PRODSTOCK</math></p>
Population	POPL	<p>the total number of humans currently living in a particular country, region or in the world.</p>
Per Capita Consumption	PERCAP	<p>average consumption per person within a population. Basically, this factor is used to indicate how popular a product is in a particular country, as compared with other countries or the world average; this also may indicate a degree of market saturation. Both physical and value terms can apply.  <math>PERCAP = CONSQNT / POPL</math> or <math>PERCAP = CONSVAL / POPL</math></p>
Net Exports Volume	NETEXQNT	<p>Net exports is the difference between a country's total exports minus its total imports. Both physical and value terms can apply. In physical terms, it refers to Net Exports Volume:  <math>NETEXQNT = EXPQNT - IMPQNT</math></p>
Net Exports Value	NETEXVAL	<p>Net exports in value terms:  <math>NETEXVAL = EXPVAL - IMPVAL</math></p>

## 4. Raw, mirror and normalized data

The system uses the following data categories:

Raw data	Data obtained directly from a primary source. With regard to import/export, this also means that this is 'direct' data, i.e. the data reported directly from a particular country.
Mirror data	<p>This data refers to import/export only. This data reflects imports/exports for a particular country, but the data is obtained by 'mirroring' trade flow data reported by a country's trade partners. I.e., the difference between direct and mirror data is as follows:</p> <p>Direct data: COUNTRY_A reported imports 1000t of avocados from COUNTRY_B in 2010. We simply use this figure in a dataset on imports of avocados to COUNTRY_A.</p> <p>Mirror data: COUNTRY_C did not report any avocado imports from COUNTRY_B in 2010, so the direct data is '0'. But when we check exports of avocados, we find that COUNTRY_B reported 200t of avocados exported to COUNTRY_C. Therefore, we go back to data normalization and make an imputation of the '200t' figure as imports of COUNTRY_C. Thus, COUNTRY_C direct data for 2010 is '0', but the mirror data is '200'. This figure is used when normalizing the datasets.</p> <p>Such discrepancies may occur for various reasons, primarily poor data collection in COUNTRY_C, mistakes while the data is being reported, or differences in trade systems. Please see a further explanation below.</p>
Normalized data	Data obtained after combining raw and mirror data and performing IB AI algorithms to eliminate any data anomalies and complete missing data.

The performance of the IndexBox AI Platform is assumed as follows. Immediately after being received from a raw data source, the data is subject to normalization, reassessment and elimination of anomalous values, in order to obtain a coherent and consistent statistical picture of production and trade worldwide. The AI-powered system of data normalization and improvement is based on a smart application of statistical methods, machine learning and data mining, which, as an integrated algorithm, was developed by our data scientists and constitutes the intellectual property of IndexBox, Inc.

Raw data, in most cases, contains numerous omissions and anomalous values, so while being processed by the IndexBox AI Platform, it is likely to change significantly; in many datasets, more than 80% of figures are re-estimated or imputed. Further features of data visualization and data-driven market analysis, which are available for users of the system via the web interface, are based on solely the improved data resulting from the IndexBox AI Platform.

In addition to the numerous data gaps and anomalous import/export values, all the raw data on global trade is subject to inconsistencies as a result of the bilateral asymmetries in the trade flows. This means that the import/export data from one country to another does not always coincide with the mirror data of the partner country; these contradictions are caused by fundamental factors such as:

- Imports are reported in CIF-type values and exports are reported in FOB-type values;
- The time lag between exports and imports; e.g., goods leaving COUNTRY\_A in 2012 might only reach COUNTRY\_B in 2013;
- Goods passing through third countries;
- Goods entering customs warehousing for several months;

- Goods being classified differently;
- Countries having different trade systems (General versus Special Trade System);
- Goods passing through industrial processing zones may or may not be recorded by the exporting country;
- Human error in completing and submitting data 'on site', before it is compiled in the statistical databases.

These inconsistencies cannot be eliminated or tracked separately; together, however, they affect the quality of any international trade analysis. We use our AI platform to ensure and improve data consistency as much as possible, so that the data forms a consistent picture of the market on both a regional and global scale. Using the AI platform, we can work effectively with data gaps and anomalous values; we can also ensure that the data obtained is as relevant as possible to complete the marketing analysis.

## 5. Units of measurement

The choice of units is available on the download data page. Unless specified otherwise, the absolute values in physical terms are presented in units relevant to the product type (units of weight, volume or number of items). In monetary terms, the indicators are presented at the current U.S. dollar rate (nominal values for each particular year, not adjusted to inflation). The calculated units are presented in the derived units for this product type, or appear as dimensionless quantities.

## 6. Regional coverage

The AI Platform contains statistical data for 200+ countries worldwide, depending on a particular product and data availability. These countries can also be presented within a certain region, according to the following list. The list of countries is compiled based on the official UN classificatory of countries and regions; the mention or omission of a country from the list of countries, or within a certain region, cannot be interpreted as neither an IndexBox statement in support of any party regarding the issue of any disputed territory, nor as an IndexBox statement on the recognition or non-recognition of any sovereignty over any territory; nor does it contain any opinion whatsoever regarding the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitations of its frontiers or boundaries.

Africa	Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cabo Verde, Cameroon, Central African Republic, Chad, Comoros, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Djibouti, Egypt, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea-Bissau, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Morocco, Mozambique, Namibia, Niger, Nigeria, Réunion, Rwanda, Saint Helena, Ascension and Tristan da Cunha, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, Sudan, Swaziland, Tanzania, Togo, Tunisia, Uganda, Western Sahara, Zambia, Zimbabwe
--------	--

Australia and Oceania	American Samoa, Australia, Cook Islands, Fiji, French Polynesia, Guam, Kiribati, Marshall Islands, Micronesia (Federated States of), Nauru, New Caledonia, New Zealand, Niue, Northern Mariana Islands, Palau, Papua New Guinea, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu, Vanuatu, Wallis and Futuna Islands
Central Asia	Kazakhstan, Kyrgyzstan, Mongolia, Tajikistan, Turkmenistan, Uzbekistan
Eastern Asia	China, China Hong Kong SAR, China Macao SAR, Democratic People's Republic of Korea, Japan, South Korea, Taiwan (Chinese)
Europe (outside the EU)	Albania, Andorra, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Channel Islands, Faroe Islands, Georgia, Gibraltar, Greenland, Holy See, Iceland, Isle of Man, Liechtenstein, Macedonia, Moldova, Monaco, Montenegro, Norway, Russia, San Marino, Serbia, Switzerland, Ukraine
European Union	Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom
Latin America and the Caribbean	Netherlands Antilles (former), Anguilla, Antigua and Barbuda, Argentina, Aruba, Bahamas, Barbados, Belize, Bermuda, Bolivia, Brazil, British Virgin Islands, Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Curaçao, Dominica, Dominican Republic, Ecuador, El Salvador, Falkland Islands (Malvinas), French Guiana, Grenada, Guadeloupe, Guatemala, Guinea, Guyana, Haiti, Honduras, Jamaica, Martinique, Mexico, Montserrat, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Saint Maarten (Dutch part), Suriname, Trinidad and Tobago, Turks and Caicos Islands, United States Virgin Islands, Uruguay, Venezuela
Middle East	Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Palestine, Qatar, Saudi Arabia, Syrian Arab Republic, Turkey, United Arab Emirates, Yemen
Northern America	Canada, Saint Pierre and Miquelon, USA
South-Eastern Asia	Brunei Darussalam, Cambodia, Indonesia, Lao People's Democratic Republic, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, Viet Nam
Southern Asia	Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka

The data is available both by country and by region; the global total figure can also be accessed. These options can be chosen via the data download interface.

## 7. Research methodology and AI platform

During the process of this research, we combine the accumulated expertise of our analysts and the capabilities of artificial intelligence. The AI-based platform, developed by our data scientists, constitutes the key working tool for analysts, empowering them to discover deep insights from basic market data.

The AI platform's algorithm is based on the following steps:

1. Data collection
2. Data cleaning, categorization and normalization
3. Data-driving decision making

### 1. Data Collection

A company-developed 'robot' is used at the data collection and processing stage; it receives the bulk of the data from a variety of sources that provide official statistical information, such as the UN, the World Bank, industry association sites and commercial databases which contain company data. However, over a third of the data from these sources has to be discarded due to its inherent unreliability: even data from the most reliable of sources frequently contains distortions and omissions.

### 2. Data Cleaning, Categorization and Normalization

Even if the data was obtained from official sources, which should be considered as reliable, it sometimes lacks consistency and completeness. The company's AI platform eliminates any data anomalies and omissions using the 'smart' system of mathematical and statistical tools developed by our experts, in combination principles of machine learning. As a rule, the official world trade data presents the most problems. If necessary, the 'robot' calculates the mirror data for each value, processing in total over 1 million data values for each trade code. In addition, the script completes any mid-series gaps or omissions using an average value, the average rates of growth and the construction of a regression model. In some cases, a regression model is used to extrapolate the data series.

The auxiliary indicators are then calculated and their initial values are evaluated and assessed for accuracy. This is achieved by calculating the confidence intervals using standard deviation and percentiles. Should the value appear as an anomaly, then the program repeats the entire recovery cycle and recalculates the values, depending on the figures that have been changed. It then performs the preparatory calculations and completes any series of missing values, using auxiliary factors and the data available used to calculate them. Using machine learning, data accuracy is constantly increasing. At the same time, our analysts moderate the AI platform results to ensure their consistency and economical sense behind the data.

### **3. Data-Driving Forecasting and Decision Making**

The use of the artificial intelligence platform makes it possible to source actionable insights and generate data-driven decisions to further expand your business. Smart extrapolation is used to develop forecasts; this process incorporates the current and projected average annual rates of growth and regression modeling methods for the projected indicators; it also avoids stating obviously irrelevant values. The advantage of these solutions lies in the fact that they are based on hard and reliable data, obtained during the multi-cycle processes of the AI platform, which is constantly being updated and improved.

### **8. Data sources**

Raw data is collected from a wide range of relevant sources, which could be combined or used for data verification, depending on data availability, product and geographical scope. This list includes, but is not limited to, the following sources:

- 1) UN Comtrade Database <https://comtrade.un.org>
- 2) UN Statistical Database <http://data.un.org>
- 3) UN Food And Agriculture Organization Statistical Database  
<http://www.fao.org/faostat/en/#data>
- 4) WorldBank Databank <https://data.worldbank.org/>
- 5) The national statistical agencies and databanks for 40+ countries worldwide

Raw data is then processed with the algorithms of the AI Platform in order to normalize the data and improve its consistency and reliability, before being presented to users of the Platform.

Since various countries have adopted different systems for product statistical classification, data set compatibility is ensured by establishing conformity between the national classifiers and the HS 2007 classifier, the core for generating the product list within the AI Platform.