

Project E26: WFP – FAOSTAT

Analyzing Impact of Food Crisis, Case of Cereals

Project repository: <https://github.com/OjohD/WFP---FAOSTAT>

Ulrich Ojoh Dioni
Ron-Aran Paju

Background

Several components and factors that contribute to crop productivity are evaluated periodically to characterize consumers' food availability and their impact on the world's population. But, not until early 2020 with the outbreak of the COVID-19 pandemic crisis, and more recently with the war between Russia and Ukraine, a rise in commodity prices reshaped by structural changes in demand from strong recovery of industries and the world's trade specifically impacted the COVID-19 pandemic have seen its repercussions in every aspect of life.

In spite multiple alerts indicted by the World Food Programme and the Food and Agriculture Organization of the United Nations that leads international efforts to defeat hunger in the world most especially in sensitive areas and cohorts, in the advent of non resolution of the ongoing geopolitics and economic crisis occurring in distinct parts of the world, there would be more and more lower global supply of food stock especially cereals, due to but not limited to complicated transporting commodities over the black sea, but also prizes increases in fertilizers, energy and productivity resources.

At the local level, elected committees develop meaningful strategies and partnership compromises to overcome these observed shortages to maintain social balance. But the adopted resolutions don't come cheap in the long run and the crystallization of social constraints trends regression and could strive to protest until something different is done for a change. However, knowing beforehand answers to some key questions: when would this happen; which sector of social life would be affected the most; What supply shortages influence the rarity of others; could stand without reasons to doubt what strategies decisions makers and commissioners must put to play to insure security in the supply of grains cereals, regulate local commercial practices to ensure equal accessibility and affordability to all.

Business goals

Since the WFP and FAO host a publicly accessed database website that collects worldwide data cash crops production per country, import and export quantity, and per countries world's countries population, we intend to explore the data available on cereals (listed over eight categories of products) which is a global trend at the top of the world's food export market of which shortages in supply in many countries in the world poses a major concern at the highest geopolitical and economical stage.

After a brief analysis of cereals global supply chain situation, and corresponding some findings with the latest global economic prospects report of June 2022 by the World Bank, we would narrow our analysis to our local environment and hypothesize on the stakes in upcoming years in shortages in supply, typically in consumer prices. We expect to see strong relations with consumer prices within export and import ratios, since there are specific countries known to be world's producers of cereals. Cereals are staple crops, which justifies concern in its supply shortage for an extended period. Moreover, cereals can be stored in silos for longer periods which makes it important to evaluate its availability, affordability, transportability in terms of imported quantities. Moreover, cereals constitute the base processing product for several other staple crops derivatives such as flour, bread etc. having them in sufficient stock to guarantee food security for a given period. This urges our final business goal as to predict the price of an example cereals derivative good in the local market.

Business success criteria

Our model prediction will be evaluated using test data for the year 2022 already available on the world bank and the food and agriculture programme database websites, and a subjective conclusion would be drawn from the study and understanding the direction of future research goals in the topic. These organizations also provide comprehensive visualization tools for their data, and this provides us with the possibility to challenge our implementation to their results while serving as a guideline. Also, by correlating our findings with the media forecasts as reported to date on the global food crisis.

Situation Assessment

Inventory of resources

Dataset: Crop and livestock products statistics are recorded for over 173 products with the objective to comprehensively cover production of all primary crops for all countries and regions in the world. FAOSTAT database provides clear datasets varying from crop production to producer market price by

countries per year, export and import quantities, population per country, and consumer prices. Terminology defined in the corresponding section.

Terminology: alongside the dataset is provided comprehensive descriptive material for each features and collection aggregates.

Hardware: computation processing will be performed to assess the collected data, and mobile phone communication to stay updated with team members.

Software: “jupyter notebook” will be used as a data analysis tool on a primary research pattern ranging from data preparation, exploration, representation, and data mining. Online resources as search engines for search of support materials. “Github” for our collaborative online development tool and for versioning control, and mobile communication application.

Requirements, assumptions, and constraints

The dataset availability is poised on effective enrollment of each member state and effective update of the country’s record for every specific year on each crop or relevant features. This directs our observation as not all countries began submitting their data to FAOSTAT the same year. These disparities could constrain the research on the fact that similar categories of data may produce opposite results.

In such a case, crop data has been recorded from 1961 to 2020 for some countries, but the price data is available starting 2000, while population data is recorded from 1950 to 2021. This certifies our assumption that not all countries would have available indicators for every feature at the same time. Thus, we would consider some dropouts for some given features provided it is not relevant to replace such missing values by some specific projections due to the nature of the indicators.

Risks and contingencies

The foreseeable risk that may arise in the course of our work is the potential loss of work resources that may include datasets (which would take a considerable amount of time to constitute), progress register of analysis and future tasks. For this reason, we will keep track of all progress and resources using Github as our collaborative development platform and versioning control and maintain our work integrity. We organize our work in objectives related sub-tasks with peer tasks reviewing, to allow flexible work integration and full participation of each team member.

Terminology

WFP: is the World Food Programme, the leading humanitarian organization saving lives and changing lives, delivering food assistance in emergencies and working with communities to improve nutrition and build resilience.

FAO/FAOSTAT: is the Food and Agriculture Organization of the United Nations that leads international efforts to defeat hunger in the world. s

Producer prices: are prices received by farmers for primary crops as collected at the point of initial sale (prices paid at the farm-gate). Units: USD/tonne of fresh product.

Consumer indices: measure the price change between the current and reference periods of the average basket of goods and services purchased by households.

consumer prices: they are prices received by farmers for primary crops, live animals and livestock primary products as collected at the point of initial sale (prices paid at the farm-gate). Annual data are provided from 1991, while monthly data from January 2010 for 180 countries and 212 products.

Producer price index: is the index of agricultural producer prices that measures the average annual change over time in the selling prices received by farmers (prices at the farm-gate or at the first point of sale).

Crops primary: Cereals, and others are expressed in terms of area harvested, production quantity and yield. Cereals: Area and production data on cereals relate to crops harvested for dry grain only.

Trade: the database includes all food and agricultural products imported/exported annually by all the countries in the world.

Annual population: time series data on population, by sex and urban/rural. The series consist of both estimates and projections for different periods as available from the original sources, namely: 1. Population data refers to the World Population Prospects: The 2019 Revision from the UN Population Division. 2. Urban/rural population data refers to the World Urbanization Prospects: The 2018 Revision from the UN Population Division.

Data understanding

Considering our desired objective as to predict the consumer prices subjected to cereals stock availability related to economic and geopolitics crisis, we are downloading the dataset from FAOSTAT <https://www.fao.org/faostat/en/#data> following the period from 1991 to 2020 and partly available for 2021 and 2022. Cereals of interest we will use for analysis are barley, maize (corn), millet, oats, quinoa, rice, rye, sorghum, and wheat

expressed in terms of production quantity, area harvested, stock, yield, consumer and producer prices, export and import quantities.

Gathering data

. Outline data requirements

Country: UN FAO member states countries arranged in alphabetical order in the dataset as per country per year per primary crops (list of cereals categories listed above), it is interesting to group countries by highest producers and exporters with influence on the global market, middle producers for those that would tend to export less and whose production is directed to mainly local consummation and non producer whose dependence are directed to the market availability.

Year: availability of the cereals data collection record years for the period of 1991 to 2020 since all nations did not submit production data at the same time.

Production quantity: it covers the period of 1991 to 2020 and represents the annual collection of products to area harvested agricultural production.

Import/Export quantity/value: in addition to food aid, with cross border trade flow, it represents the amount of food and agricultural products imported/exported annually by all the countries in the world.

Producer prices: these are prices received by farmers for primary crops, products as collected at the point of initial sale. The annual data are provided from 1991 to 2020. Producer Price would include an index representing agricultural producer prices that measures the average annual change over time in the selling prices received by farmers.

. Verify data availability

From the FAOSTAT portal, we can refine our selection as per our needs. We choose from Crops and livestock products under the option trade, then select the list of countries, the record years 1962 to 2020, the list of cereals in the primary crops category and the export and export features we are interested in.

The screenshot displays the FAOSTAT database portal interface with four main filter sections:

- COUNTRIES:** Includes a search bar with the text "Filter results e.g. afghanistan". A list of countries is shown with radio buttons: Afghanistan, Albania, Algeria, Angola, Antigua and Barbuda, and Argentina. Below the list are "Select All" and "Clear All" buttons.
- ELEMENTS:** Includes a search bar with the text "Filter results e.g. Import quantity". A list of elements is shown with radio buttons: Import Quantity, Import Value, Export Quantity, and Export Value. Below the list are "Select All" and "Clear All" buttons.
- ITEMS:** Includes a search bar with the text "Filter results e.g. abaca, manila hemp, raw". A list of items is shown with radio buttons: Lucerne (alfalfa) meal and pellets, Maize (corn) (which is selected with a checkmark), Malt, whether or not roasted, Mangoes, guavas and mangosteens, and Margarine and shortening. Below the list are "Select All" and "Clear All" buttons.
- YEARS:** Includes a search bar with the text "Filter results e.g. 2020". A list of years is shown with radio buttons: 2020, 2019, 2018, 2017, 2016, and 2015. Below the list are "Select All" and "Clear All" buttons.

Figure 1: FAOSTAT database portal with selected crops production per country per year for statistical analysis .

. Define selection criteria

The dataset for the selected features are available for download as CSV file template. The data is clean and can be imported in “jupyter notebook” as our data-mining platform. We perform a first hand exploration insight to reveal the potential mining resources we need to carry on the research.

	Area	Wheat
0	China	1.123065
1	China, mainland	1.123037
2	India	0.777602
3	USSR	0.719910
4	United States of America	0.583258
5	Russian Federation	0.506518
6	France	0.356978
7	Canada	0.271003
8	Germany	0.217194
9	Pakistan	0.210391
10	Australia	0.204280

	Area	Rice
0	China	1.950158
1	China, mainland	1.932494
2	India	1.407477
3	Indonesia	0.537731
4	Bangladesh	0.411034
5	Viet Nam	0.352931
6	Thailand	0.286471
7	Myanmar	0.237133
8	Philippines	0.146066
9	Japan	0.113161
10	Brazil	0.110848

Figure 2: A preview of the world’s top 10 producers of wheat (left) and rice (right) for the period of 1991 to 2020, aggregate is average for 100 million tonnes.

. Describing data

The FAOSTAT provides a dataset in “.csv” file format for download. Our selection comprises eight categories of cereals features that are grouped by countries for an average production covering the period of 1991 to 2020. It appeared reasonable for us to choose the period with most countries having enough data on every category selected. It is confirmed after exploration that not all countries provided data concerning cereal production at the same period. Consequently, some time series would be dropped to maintain data integrity for projections.

. Exploring data

From 1991 to 2020, the average production is recorded for 133 countries over 193 who are members. This is an indication that our data have missing values to deal with.

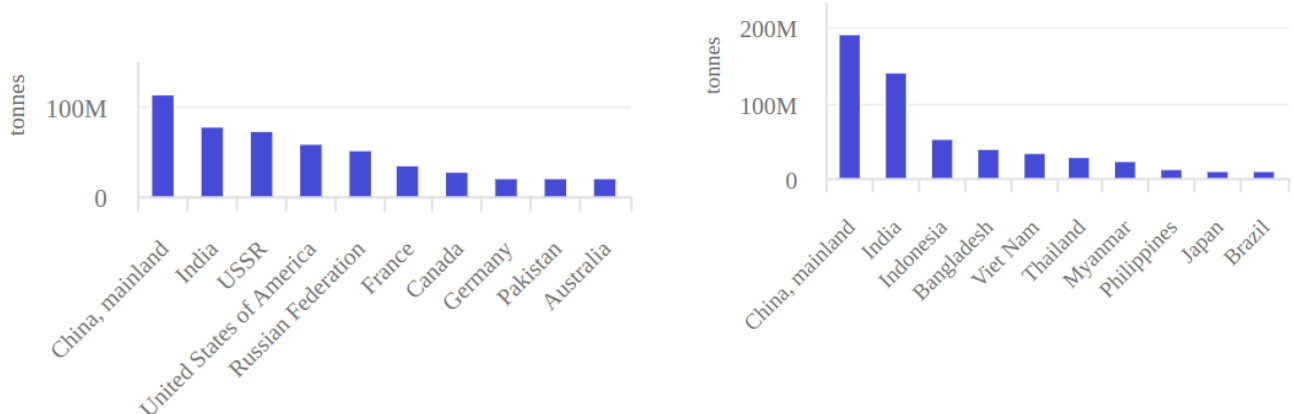


Figure 3 : World's top 10 countries producers of wheat (left) and rice (rice) for the period of 1991 to 2020 for 100 million tonnes.

. Verifying data quality

The first insight from the data representation highlights the feasibility of our primary goal in this project workflow, which is to report the world's best producers of cereals and target countries export and import rate.

Project planning

Find suitable and relevant data from all the available data sets (5 hours)

Look through the data (5 hours)

Take the parts that are relevant for the project (20 hours)

Construct a unique dataset with correspondent features useful into our prediction goal since they are independent pieces on separate products. (30 hours)

Predict the price for 2023 (30 hours)

Firstly we want to find all the datasets that interest us, so we are searching for databases that include information about food prices and shortages in supplies around the world since the global pandemic Covid-19 has had an impact on prices and food supplies. Then we select out the data that interests us, and features we need to use when predicting the outcome for the upcoming year. When we have downloaded our necessary datasets (they come as CSV files format), we look through them and see if everything is as we expected and want it to be. Our plan is to firstly analyze the production/yield related features and evaluate the market prices per export/import rates, and finally predict the food shortage indices per country and market prices including fertilizers where applicable. We plan to build a regression model that predicts the potential shortage of food for the upcoming year.