Empowering Sustainable Practices

-Trase-

Robert Gruber | CIS5860-101 Applied Analytics Project | August 3rd, 2024

Table of Contents

Executive Summary

The agriculture and commodities industry is a fundamental pillar of global trade, producing and distributing essential goods such as soybeans, corn, and meat products. This sector drives economic growth and meets the world's increasing demand for food and agricultural products. However, it faces significant challenges related to sustainability and environmental impact, such as deforestation, habitat loss, and greenhouse gas emissions. These challenges are compounded by the globalized nature of commodity supply chains, making it difficult to trace product origins and ensure responsible sourcing.

The project, aimed to address these challenges by leveraging data from Trase (Transparency for Sustainable Economies) to provide transparency in commodity supply chains. Trase offers a wealth of datasets from various countries and products, yet these datasets are not consolidated into a single format for comprehensive analysis. This project sought to create a centralized repository to examine global trade patterns and trends effectively, empowering researchers and stakeholders to gain actionable insights.

Our primary objectives were to integrate diverse datasets from multiple sources, analyze key drivers of trade dynamics, assess the impact of economic policies and environmental factors on trade, and propose sustainable practices for the industry. I conducted extensive data collection and cleansing, utilizing RStudio for integration and analysis. The data integration process involved creating a unified dataset from over 9 million observations and 111 variables, spanning 10 countries and 38 products. This centralized dataset facilitated detailed analysis of trade volumes, export destinations, economic relationships, and environmental impacts.

One of the significant outcomes of this project was the development and public release of RStudio code, making the data and analytical tools accessible to the broader community, including stakeholders, researchers, and policymakers. This transparency promotes collaborative efforts in analyzing and addressing the environmental impacts of global trade. Additionally, the majority of data files are centralized in google drive to enhance the usability of the code for anyone who interacts with it.

Early exploratory analysis using the integrated dataset revealed critical insights, such as average trade volumes and (FOB) values based on countries, the number of exporters and countries of production, and other key metrics. For instance, I identified significant variations in trade volumes and FOB values across different countries and products, highlighting the need for tailored strategies to address sustainability challenges in specific regions. Additionally, I uncovered patterns in the number of exporters and production countries, providing insights into market dynamics and the concentration of production activities.

These findings offer valuable information for stakeholders to make informed decisions that promote sustainability in global trade. Our recommendations focus on enhancing transparency, optimizing supply chains, and encouraging policies that mitigate environmental risks. By providing a comprehensive understanding of trade patterns and their environmental implications, this project contributes to the ongoing efforts to transition towards more sustainable practices in the agriculture and commodities industry.

Organizational Background

Trase (Transparency for Sustainable Economies) is an innovative initiative developed through a collaboration between the Stockholm Environment Institute (SEI) and Global Canopy (Trase.earth). Trase aims to illuminate the intricate supply chains of agricultural commodities, such as soy, palm oil, beef, and timber, by providing a comprehensive platform that maps the journey of these commodities from their origin to their final destination. The mission of Trase is to enhance transparency within these supply chains to foster sustainable and responsible trade practices. By integrating vast datasets from various sources, Trase creates detailed, high-resolution supply chain maps that reveal the environmental and social impacts associated with commodity trade flows. This information is invaluable for stakeholders, including businesses, governments, and civil society organizations, enabling them to identify and address sustainability risks within their supply chains.

Trase focuses on key commodities that are significant drivers of deforestation and environmental degradation, providing actionable intelligence to guide efforts in mitigating these impacts. The platform's data is accessible to a wide audience, promoting collaboration and informed decision-making across different sectors. This accessibility helps foster a sense of shared responsibility among stakeholders, encouraging more sustainable practices and policies. Trase's work is crucial in promoting transparency and accountability, aiding the transition toward more sustainable economies. By offering robust and comprehensive data, Trase supports initiatives aimed at improving land use practices, protecting biodiversity, and reducing greenhouse gas emissions associated with commodity production and trade.

Business Problem

Global trade has expanded significantly over recent decades, boosting production and export activities across various sectors. However, this growth has raised critical environmental concerns, particularly in regions heavily involved in commodity production such as soy, cocoa, and palm oil. The environmental impacts include widespread deforestation, biodiversity loss, and increased greenhouse gas emissions, all of which adversely affect local communities. For instance, deforestation driven by agricultural expansion can lead to soil degradation, reduced water resources, and diminished agricultural yields, putting strain on local populations who rely on these resources for their livelihoods.

Analyzing the environmental impact of trade is complex due to the fragmented nature of available data. Trase, a prominent platform providing data on supply chains and environmental impacts, does not offer a unified dataset. Instead, it presents highly specific datasets tailored to individual products and their associated environmental factors. This specialization, while useful for detailed analysis of specific commodities, complicates efforts to conduct comprehensive cross-product comparisons. The lack of consistent variables across datasets impedes the ability to track and analyze elements such as overall emissions and deforestation rates comprehensively.

The project seeks to overcome these limitations by consolidating diverse datasets into a single, cohesive framework. By doing so, it aims to enhance the visibility of trade patterns and their environmental consequences, offering actionable insights for stakeholders committed to promoting sustainable practices. Addressing these issues requires innovative data integration strategies and analytical techniques to ensure that economic growth does not come at the expense of environmental and community health.

Project Goals or Problem Scope

The primary goal of this project is the construction of a robust dataset containing comprehensive data from all countries and products, facilitating easier access for running extensive data analytics and understanding the larger picture within the data. By consolidating this information into a single, unified dataset, the project aims to overcome the limitations posed by the lack of consistent variables across different datasets, which previously hindered comprehensive analysis.

To enhance accessibility and usability, the project utilized Google Drive to store and share the files, eliminating the need for each user to individually download all data. This approach streamlines the process, except for four large CSV files that posed issues when shared through Google Drive. Addressing these challenges was crucial to ensuring the dataset's availability and usability for all potential users.

In addition to building the dataset, the project conducted an exploratory analysis to uncover insights within the newly constructed data. This analysis aimed to identify key trends and patterns, such as average volumes and Free on Board (FOB) values based on countries, the number of exporters, and countries of production. By performing this preliminary analysis, the project aims to better direct future researchers and stakeholders who will utilize the dataset and the accompanying RStudio code when made publicly available. This will enable them to carry out their own analyses and drive forward sustainability and informed decision-making based on comprehensive data insights.

Through this project, the ultimate aim is to promote more sustainable practices in global trade. By providing a comprehensive and accessible dataset, along with the tools and insights necessary for effective analysis, the project seeks to empower researchers, policymakers, and stakeholders to make data-driven decisions that prioritize environmental conservation and socio-economic well-being. This initiative strives to foster a more transparent and sustainable global trade system, where the impacts on local communities and ecosystems are carefully considered and mitigated.

Data Collection

Initially, I identified and reviewed various trade data sources to find information that could support our project's objectives. Trase, a platform offering detailed data on global trade, environmental impacts, and supply chains, was identified as a key resource. The information obtained from Trase included data on production volumes, trade flows, exporter and importer details, economic blocs, and environmental impact metrics. The data obtained from Trase encompassed comprehensive information on various aspects of global trade, including production quantities, trade routes, details about exporting and importing entities, regional economic groupings, and associated environmental impacts. This foundational data allowed us to begin the collection and review process, ensuring that I had a comprehensive view of the trade dynamics across different commodities and regions.

Data was collected from multiple Trase datasets, each containing unique variables specific to certain products and countries. This necessitated a meticulous approach to ensure all relevant data was included. The primary challenge was constructing the code to load all datasets into their appropriate country-specific categories and then combine them into a unified dataset. The data collected encompassed various aspects such as production volumes, trade flows, exporter and importer details, economic blocs, and environmental impact metrics. These elements were crucial for creating a dataset that could support in-depth analysis and provide meaningful insights into global trade and its environmental implications.

Data Source

Trase's datasets are particularly valuable because they cover a broad spectrum of commodities, such as soy, palm oil, beef, and timber, which are key contributors to global trade and have significant environmental impacts. Each dataset includes specific variables pertinent to the commodity in question, ranging from trade volumes and values to detailed environmental indicators like emissions and deforestation rates. This diversity in data types and the level of detail provided makes Trase an invaluable resource for constructing a comprehensive dataset that can support a wide range of analyses.

Data Cleaning

The data cleaning process was critical to ensuring the accuracy and usability of our dataset. Given the variety and volume of data sourced from Trase, I encountered several challenges that required careful attention. These challenges varied across the entire process but included the handling of missing values, standardizing formats, cleaning column names, changing data types, and validating the data uploads. Challenges were dealt with as they arose meticulously to ensure a robust dataset and ensure future analysis.

Handling Missing Values

One of the primary challenges was dealing with missing values. Given the variability in methodologies across the datasets, implementing a single approach, such as using an average mean for imputation, was not feasible. Instead, I worked to address missing values by evaluating each dataset individually and applying context-specific strategies. In some cases, records with missing values were removed if they were deemed non-essential, which helped maintain the overall integrity of the dataset. This approach ensured that the dataset remained as accurate and complete as possible without introducing significant biases. Much of the combined data set contained missing data due to highly unique data sets and variables pertaining to only specific products.

Cleaning Column Names

Column names from the original datasets were often inconsistent or unclear. To address this I undertook a thorough renaming process to establish a standard naming convention, which improved the readability and usability of the dataset. This task included removing any special characters, abbreviations, or ambiguities in the column names, ensuring that they were descriptive yet concise. Cleaning the column names was done through utilizing R-studios library janitor to standardize and ensure consistency. Product names under this category proved very difficult as they varied across different countries and lacked standardization.

Data Type Conversion

Ensuring that all data types were consistent was another crucial step. Numeric fields, in particular, required conversion from text to numeric formats to facilitate accurate analysis. This step was essential for subsequent statistical analysis and visualization, ensuring that operations and calculations performed on the data were accurate and reliable. Another step in relation to this included filtering through fob and volume values to ensure no text data was present that would have disrupted future analysis.

Data Validation

Beyond these core cleaning tasks, I worked to implement validation checks to verify data accuracy and consistency. This included cross-referencing records against known benchmarks and performing consistency checks between related variables. These validation steps helped identify and correct any anomalies or errors that might have persisted despite the initial cleaning efforts. This was achieved through the uploading of data sets through an initial download file path validated for observations and variables present then was converted later on to google drive file paths for better usability.

Data Storage

To facilitate data collection, Google Drive was utilized for storing and sharing files. This approach streamlined the process, allowing team members to access data without the need for individual downloads. However, four large CSV files posed issues when shared through Google Drive, requiring alternative methods for transfer and access. The Google Drive setup was chosen for its convenience and the ability to collaborate seamlessly across different locations. Handling large files necessitated additional strategies, such as segmenting the data into smaller files or using direct download links to ensure smooth access for all team members.

RStudio was the primary tool used for data processing and integration. Scripts were developed to automate the loading and merging of datasets, ensuring consistency and accuracy. These scripts were designed to handle various data formats and structures, making the integration process efficient and reducing the risk of errors. RStudio's comprehensive toolset allowed for detailed data examination, ensuring that the final dataset was both accurate and insightful.

To ensure data security and integrity, regular backups were performed, and version control was maintained throughout the project. This approach minimized the risk of data loss and allowed for easy tracking of changes made to the dataset. Additionally, metadata was meticulously documented and included in the appendix, providing clear context and reference for all variables used in the dataset. This documentation is crucial for future users who will rely on the dataset for their analyses.

Data Analysis

Data Manipulation

To effectively manage the diverse data sources, the initial step involved uploading all relevant datasets into Rstudio. These datasets, sourced from various locations, included comprehensive trade and environmental data. After uploading, the data was consolidated into a unified format to create a combined dataset, which served as the foundation for our analysis.

Moving forwards, I developed a dataset named common_data. This dataset was created by focusing on the key variables extracted from the combined dataset, which included year, country_of_production, port_of_export, exporter, country_of_destination, economic_bloc, product, volume, and fob. By isolating these common variables, it streamlined the dataset to emphasize the most pertinent aspects for our research. This would make the data more usable by excluding the variables not being utilized at the time.

Following the consolidation, an exploratory data analysis was conducted to understand the structure and quality of the data. This included examining the unique values within key columns to identify any anomalies or inconsistencies. Summary statistics for critical variables such as volume and fob were generated to provide an overview of the data's distribution and variability. This initial exploration was essential for detecting outliers and understanding the central tendencies and dispersion within the dataset.

Merging Data Sets

The process of merging datasets was pivotal for creating a unified and comprehensive dataset that would serve as the foundation for our analysis. Initially, I worked to consolidate multiple datasets, each containing various segments of trade and environmental data, into a single, integrated dataset. Rather than merging on specific variables, our approach focused on extracting and harmonizing common variables from the combined dataset. This identified key variables relevant to our research. These variables were essential for providing a coherent view of the data. The next step involved creating a simplified dataset, named common_data, which included these common variables. This simplification was critical to ensure that the data concentrated on the most pertinent aspects while maintaining the integrity of the information.

Descriptive Analysis

With the common_data dataset prepared, descriptive analysis was conducted to understand the dataset's structure and key characteristics. This involved generating summary statistics for critical variables such as volume and fob to gain insights into their distribution and variability. Next I looked at the unique values within key columns, including year, country_of_production, port_of_export, and others, to identify any anomalies or data quality issues. To further explore the data, I created visualizations and summary tables to illustrate trends and patterns. For example, I plotted the distribution of volume and fob over time and analyzed these metrics by country and product. This exploration provided a high-level overview of trade patterns and highlighted significant trends.

Country	Observations	Variables
Brazil	7,125,359	46
Indonesia	1,835,367	60
Argentina	416,088	22
Ecuador	106,359	22
Colombia	43,067	18
Paraguay	30,258	26
Cote D'Ioire	27,636	17
Bolivia	14,793	19
Peru	15,554	11
Ghana	2,111	13

Predictive Analysis

Once the dataset was prepared and descriptive summaries were completed, I focused on the predictive analysis to infer potential trends and future directions. This process was guided by the insights derived from our summary statistics and exploratory data analysis. I analyzed the descriptive statistics to identify patterns and anomalies in the data, such as shifts in trade volumes or variations in FOB values. By examining these trends, I worked to create educated predictions about future developments in trade and environmental impacts. For instance, significant increases in volume for certain products or regions could indicate emerging trade patterns or shifts in production practices.

The insights gained from this preliminary analysis allowed us to anticipate potential changes in trade dynamics and their environmental consequences. For example, a noticeable rise in trade volumes for specific products might suggest a growing demand that could impact resource usage or emissions. Similarly, trends in FOB values could hint at shifts in market value or pricing strategies.

Results

The results of our data analysis offer a detailed examination of global trade patterns and their environmental impacts, highlighting critical insights from the common_data dataset. The analysis of Free on Board (FOB) values reveals significant economic disparities. The FOB values ranged from a minimum of 0.0003 to a staggering maximum of 1,150,000,000.0. The first quartile was 965.0, the median was 2276.0, and the mean was 91884.0, with the third quartile at 8823.0. This wide range and high mean value indicate that while many transactions are relatively modest, a few extremely high-value trades significantly skew the average, reflecting substantial economic activities concentrated in a few transactions.

The volume data further emphasizes this variability. With volumes ranging from 0 to 6,617,072, the first quartile was 0, the median was 1, the mean was 305, and the third quartile was 5. The presence of zero values in the lower quartile and a median of just 1 signifies that many trade entries involve very small quantities, whereas the large maximum and mean values suggest that high-volume trades are less frequent but heavily impact overall trade volume.

Our data covers a vast global network, with port of export data spanning 193 countries and exporters represented by over 7,689 locations. The country of destination data includes 235 locations, reflecting the extensive reach of global trade. Economic blocs are well represented with 213 unique entities, indicating a broad participation across different economic groupings. The diversity of products traded—ranging from agricultural commodities like cotton, corn, soy, and sugarcane to animal products such as beef, shrimp, chicken, and pork—illustrates the varied nature of global trade.

Top exporters by volume and FOB values stand out significantly. Bunge Alimentos S/A recorded a total volume of 106,632,853 and a total FOB of 39,254,860,369. Cargill Agricola SA followed with a total volume of 89,341,430 and a total FOB of 33,099,961,254. Raizen Energia S.A reported a total volume of 78,268,541 and a total FOB of 3,637,844,345. These figures underscore the dominance of these exporters in the market. Bunge Alimentos S/A and Cargill Agricola SA, in particular, not only lead in volume but also in the economic value of their exports, indicating their significant influence on global trade dynamics.

The numerical analysis of FOB and volume values, alongside the distribution of exporters and destinations, paints a picture of a highly concentrated and variable trade environment. The data reflects the complexity and scale of international trade, with a few large players driving significant economic impact while a multitude of smaller transactions populate the lower end of the scale. This insight is crucial for understanding the economic and environmental implications of global trade patterns, emphasizing the need for targeted sustainability efforts in high-impact areas.

Limitations

Several limitations emerged that impact the reliability and comprehensiveness of the results. One significant limitation is the lack of methodological consistency across various variables in the dataset. The commodities corn, cotton, and wood pulp, do not have disclosed methodologies for measuring or reporting their volume and FOB (Free on Board) units. This complicates the comparability of data across different countries and commodities, potentially leading to inaccuracies in cross-comparisons.

Additionally, some key commodities lack detailed methodological descriptions, which affects the precision of the data. For instance, commodities like cocoa and coffee are reported without consistent methodological details, making it challenging to interpret the volume and FOB units accurately. This lack of clarity particularly impacts commodities such as soy, where data on exports, production, and deforestation is provided, but without comprehensive methodological explanations.

The variability in volume and FOB units further complicates the analysis. Different metrics, such as million tonnes, tonnes, and thousand tonnes, are used for reporting, which, combined with inconsistent FOB units, hinders the standardization of data and integration efforts. This inconsistency affects the financial metrics and overall analysis of trade values.

Regional disparities also present a challenge. The dataset shows significant differences in data reporting across countries. While some regions provide detailed reports on commodities like soy and beef, others, such as Colombia and Peru, have minimal or inconsistent data on commodities like cocoa and coffee. These discrepancies can introduce biases and limit the generalizability of the analysis, as the dataset may not equally represent all regions or commodities.

Finally, the predictive analysis conducted was based on descriptive summary statistics rather than advanced statistical or machine learning models. This approach provided directional insights but lacked the rigor of more sophisticated predictive techniques. As a result, while the analysis was informative, it may not fully capture the complexities and nuances of the trade and environmental relationships examined.

Implications

The analysis carried out in this project has several important implications for policymakers, stakeholders, and researchers involved in global trade and environmental conservation. By developing a more robust and comprehensive understanding of trade patterns and their environmental impacts, stakeholders can make more informed decisions and implement strategies that balance economic growth with environmental sustainability.

First, the data-driven insights from this analysis can help policymakers create more effective regulations and incentives to promote sustainable trade practices. For instance, understanding the trade volumes and deforestation impacts associated with commodities like soy and beef can guide the formulation of policies that mitigate environmental damage while supporting economic development. Secondly, the identification of key variables and their interactions provides a foundation for future research and analysis. Researchers can build upon this work by incorporating more sophisticated predictive models and addressing the data limitations identified in this study. By enhancing the methodology and expanding the dataset, future analyses can yield even more precise and actionable insights.

Additionally, the lack of methodological consistency across different countries and commodities underscores the need for standardized data collection and reporting practices. Establishing common methodologies will improve the comparability and reliability of data, enabling more accurate cross-country and cross-commodity analyses. This standardization is essential for global efforts to monitor and manage the environmental impacts of trade.

For businesses and trade organizations, the insights derived from this analysis can inform strategic decisions and operational improvements. Companies involved in the production and export of commodities can use the findings to optimize their supply chains, reduce environmental impacts, and enhance their sustainability credentials. This can lead to increased market competitiveness and compliance with evolving environmental regulations. Lastly, the engagement of various stakeholders, including producers, exporters, and policymakers, is crucial for the continuous improvement of data quality and the refinement of analytical models. Collaboration and information sharing can address gaps in data and methodology, fostering a more transparent and sustainable trade environment.

Recommendations

These results imply several important considerations for further analysis. The wide variation in FOB and volume values suggests that there are significant differences in trade value and quantity among different countries and commodities. This variation can be leveraged to identify key trends and patterns in global trade, such as which countries are the largest exporters of specific commodities and how trade volumes have changed over time.

The extensive range of years covered in the dataset allows for longitudinal studies to assess trends and shifts in trade practices and their environmental impacts over nearly two decades. Analyzing data from 2004 to 2022 can reveal how trade policies, economic conditions, and environmental regulations have influenced trade flows and sustainability.

The diversity of exporters and destinations highlights the complexity of global trade networks. By examining the relationships between different exporters, ports, and destination countries, it is possible to uncover the dynamics of supply chains and identify potential bottlenecks or areas for improvement in trade efficiency. Understanding the distribution of economic blocs and their involvement in trade can also provide insights into regional trade agreements and their effects on global commerce.

The identification of top exporters by volume and FOB can help in pinpointing major players in the market and understanding their influence on global trade. This information can be used to focus more detailed analyses on the practices and strategies of these key exporters, potentially revealing best practices or areas for policy intervention to enhance sustainability.

The average volume and FOB data by country and product can guide targeted investigations into specific commodities and regions. For instance, the high average volume of sugarcane and wood pulp exports can prompt deeper exploration into the environmental impacts of these industries and the sustainability of their supply chains. Similarly, the significant FOB values for countries like Ghana and Paraguay suggest a need to understand the economic implications of their trade activities and the potential for economic development through sustainable trade practices.

References

- *Trase - insights and analysis on commodity trade sustainability - trase*. trase.earth. (n.d.). https://trase.earth/

Variables Table

No.	Variable Name	Data Type	Description
1	year	integer	Year of data record.
2	country_of_production	string	Country where the product was produced.
3	port_of_export	string	Port from which the product was exported.
4	exporter	string	Name of the exporter.
5	exporter_group	string	Group or category of the exporter.
6	country_of_destination	string	Country where the product was exported to.
7	economic_bloc	string	Economic bloc associated with the country of production.
8	volume	numeric	Volume of the product exported.
9	fob	numeric	Free on board (FOB) value of the exported product.
10	product	string	Type of product.
11	country	string	Country name.
12	biome	string	Biome associated with the production area.
13	product_type	string	Type or category of the product.
14	forest_500_soy	boolean	Indicator for soy production in forests identified as '500'.
15	zero_deforestation_argentina_soy	boolean	Indicator for zero deforestation commitment for soy in Argentina.
16	province_of_production	string	Province where the product was produced.
17	department_of_production	string	Department where the product was produced.
18	co2_emissions_soy_deforestation_5_year_total_exposure	numeric	CO2 emissions from soy-related deforestation over 5 years.
19	land_use	string	Type of land use

			associated with production.

20	soy_deforestation_5_year_total_exposure	numeric	Total exposure to soy-related deforestation over 5 years.
21	province_of_production_trase_id	integer	Unique identifier for province of production in Trase.
22	department_of_production_trase_id	integer	Unique identifier for department of production in Trase.
23	zero_deforestation_bolivia_soy	boolean	Indicator for zero deforestation commitment for soy in Bolivia.
24	municipality_of_production	string	Municipality where the product was produced.
25	importer_group	string	Group or category of the importer.
26	co2_gross_emissions_soy_deforestation_5_year_total_exposure	numeric	Gross CO2 emissions from soy-related deforestation over 5 years.
27	municipality_of_production_trase_id	integer	Unique identifier for municipality of production in Trase.
28	importer	string	Name of the importer.
29	port	string	Port of importation.
30	coffee_bean	boolean	Indicator for coffee bean production.
31	water_shortage_index	numeric	Index indicating water shortage severity.
32	department	string	Department name.
33	municipality_of_export	string	Municipality where the product was exported from.
34	department_trase_id	integer	Unique identifier for department in Trase.
35	zero_deforestation_cote_divoire_cocoa	boolean	Indicator for zero deforestation commitment for cocoa in Côte d'Ivoire.
36	logistics_hub	string	Hub for logistical operations.
37	cocoa_deforestation_15_years_total_exposure	numeric	Total exposure to cocoa-related deforestation over 15 years.
38	province	string	Province name.

39	canton	string	Canton name.
40	parish	string	Parish name.
41	province_trase_id	integer	Unique identifier for province in Trase.
42	canton_trase_id	integer	Unique identifier for canton in Trase.
43	parish_trase_id	integer	Unique identifier for parish in Trase.
44	forest_500_palm_oil	boolean	Indicator for palm oil production in forests identified as '500'.
45	zero_deforestation_indonesia_palm_oil	boolean	Indicator for zero deforestation commitment for palm oil in Indonesia.
46	kabupaten_of_production	string	Kabupaten where the product was produced.
47	mill	string	Name of the mill.
48	mill_group	string	Group or category of the mill.
49	refinery	string	Name of the refinery.
50	refinery_group	string	Group or category of the refinery.
51	country_of_first_import	string	Country of first import.
52	oil_palm_ha	numeric	Area of oil palm production in hectares.
53	palm_oil_deforestation_10_year_total_exposure	numeric	Total exposure to palm oil-related deforestation over 10 years.
54	kabupaten_of_production_trase_id	integer	Unique identifier for kabupaten of production in Trase.
55	kabupaten_of_production_node_sub_type	string	Sub-type of kabupaten of production in Trase.
56	country_of_production_trase_id	string	Unqiue Trase ID
57	country_of_first_import_trase_id	integer	Unique identifier for country of first import in Trase.
58	country_of_wood_production	string	Country of wood production.
59	province_of_wood_production	string	Province of wood production.
60	pulp_wood_source	string	Source of pulp wood.

61	wood_supplier	string	Name of the wood supplier.
62	wood_supplier_corporate_group	string	Corporate group of the wood supplier.
63	pulp_mill	string	Name of the pulp mill.
64	pulp_mill_corporate_group	string	Corporate group of the pulp mill.
65	zero_deforestation_indonesia_wood_pulp	boolean	Indicator for zero deforestation commitment for wood pulp in Indonesia.
66	pulp_grade	string	Grade or quality of pulp.
67	province_of_pulp_export	string	Province where pulp is exported from.
68	importer_corporate_group	string	Corporate group of the importer.
69	country_of_importer	string	Country of the importer.
70	annual_woodpulp_deforestation	numeric	Annual deforestation related to wood pulp production.
71	burned_area	numeric	Area burned.
72	co2_total_emissions_wood_pulp_exposure	numeric	Total CO2 emissions related to wood pulp exposure.
73	concession_area	numeric	Area under concession.
74	concession_deforestation	numeric	Deforestation in concession areas.
75	deforestation_on_peat	numeric	Deforestation occurring on peatlands.
76	peat_area	numeric	Area of peatlands.
77	remaining_forest	numeric	Remaining forest area.
78	wood_pulp_deforestation_10_year_total_exposure	numeric	Total exposure to wood pulp-related deforestation over 10 years.
79	wood_pulp_production_volume_m3	numeric	Volume of wood pulp production in cubic meters.
80	province_of_wood_production_trase_id	integer	Unique identifier for province of wood production in Trase.
81	wood_supplier_trase_id	integer	Unique identifier for wood supplier in Trase.
82	co2_emissions_wood_pulp	numeric	CO2 emissions from

	_burned_peat_exposure		burned wood pulp on peatlands.
83	co2_emissions_wood_pulp_subsidence_co2_exposure	numeric	CO2 emissions from subsidence related to wood pulp production.
84	co2_gross_emissions_wood_pulp_deforestation_exposure	numeric	Gross CO2 emissions from wood pulp-related deforestation.
85	co2_net_emissions_wood_pulp_deforestation_exposure	numeric	Net CO2 emissions from wood pulp-related deforestation.
86	forest_500_beef	boolean	Indicator for beef production in forests identified as '500'.
87	zero_deforestation_paraguay_beef	boolean	Indicator for zero deforestation commitment for beef in Paraguay.
88	district_of_export	string	District from which the product was exported.
89	cattle_deforestation_5_year_total_exposure	numeric	Total exposure to cattle-related deforestation over 5 years.
90	co2_emissions_cattle_deforestation_5_year_total_exposure	numeric	CO2 emissions from cattle-related deforestation over 5 years.
91	zero_deforestation_paraguay_corn	boolean	Indicator for zero deforestation commitment for corn in Paraguay.
92	zero_deforestation_paraguay_soy	boolean	Indicator for zero deforestation commitment for soy in Paraguay.
93	zero_deforestation_brazil_beef	boolean	Indicator for zero deforestation commitment for beef in Brazil.
94	state_of_production	string	State where the product was produced.
95	co2_gross_emissions_cattle_deforestation_5_year_total_exposure	numeric	Gross CO2 emissions from cattle-related deforestation over 5 years.
96	co2_net_emissions_cattle_deforestation_5_year_total_exposure	numeric	Net CO2 emissions from cattle-related deforestation over 5 years.
97	state_of_production_trase_id	integer	Unique identifier for state of production in Trase.
98	state	string	State name.
99	municipality	string	Municipality name.
100	co2_emissions_soy_defore	numeric	Total CO2 emissions from

	station_5_year_total_from_soy_feed_expo		soy-related deforestation over 5 years specifically attributed to soy feed exposure.

101	soy_deforestation_5_year_total__from_soy_feed_exposure	numeric	Total exposure to soy-related deforestation over 5 years specifically attributed to soy feed exposure.
102	soy_feed_area	numeric	Area used for soy feed production.
103	soy_feed_volume	numeric	Volume of soy feed produced.
104	state_trase_id	string	Unique identifier for the state in Trase.
105	municipality_trase_id	string	Unique identifier for the municipality in Trase.
106	product_descr	boolean	Boolean indicator (true/false) for the presence of a detailed description of the product.
107	zero_deforestation_brazil_cocoa	string	Indicator for zero deforestation commitment for cocoa in Brazil.
108	co2_emissions_soy_deforestation_5_year_sum_risk	numeric	Sum of CO2 emissions from soy-related deforestation over 5 years, considering risk factors.
109	soy_deforestation_5_year_sum_risk	numeric	Sum of exposure to soy-related deforestation over 5 years, considering risk factors.
110	zero_deforestation_brazil_soy	string	Indicator for zero deforestation commitment for soy in Brazil.
111	co2_net_emissions_soy_deforestation_5_year_total_exposure	numeric	Net CO2 emissions from soy-related deforestation over 5 years, considering total exposure.

Methodology Table

Country	Commodity	Volume Unit	FOB Unit
Argentina	Corn	No methodology	No methodology
	Cotton	No methodology	No methodology
	Soy	(soy exports) Million tonnes (soy production) Million tonnes	No methodology
	WoodPulp	No methodology	No methodology

Country	Commodity	Volume Unit	FOB Unit
Bolivia	Soy	(soy exports) Million tonnes (soy production) Million tonnes	No methodology

Country	Commodity	Volume Unit	FOB Unit
Brazil	Beef	(Beef and live cattle exports) Million tonnes (Number of cattle) Million tonnes	No methodology
	Chicken	(Chicken exports) Million tonnes (Soy feed consumed by exported chickens) Million tonnes	No methodology
	Cocoa	No methodology	No methodology
	Coffee	No methodology	No methodology
	Corn	No methodology	No methodology
	Cotton	No methodology	No methodology
	Palm Kernel	No methodology	No methodology
	Palm Oil	No methodology	No methodology
	Pork	(Pork exports) tonnes (soy feed consumed (tonnes) to produce	No methodology

		exported pork	
	Soy	(soy area) million hectares planted (soy yield) tonnes per hectare (Soy deforestation) million hectares	No methodology
	Sugar Cane	No methodology	No methodology
	Wood Pulp	No methodology	No methodology

Country	**Commodity**	**Volume Unit**	**FOB Unit**
Colombia	Beef	No methodology	No methodology
	Cocoa	No methodology	No methodology
	Coffee	(coffee exports) million tonnes	No methodology
	Palm Kernel	No methodology	No methodology
	Palm Oil	No methodology	No methodology
	WoodPulp	No methodology	No methodology

Country	**Commodity**	**Volume Unit**	**FOB Unit**
Cote d'Ivoire	Cocoa	Trase exported volume(1,000 tonnes)	FOB(millions)

Country	**Commodity**	**Volume Unit**	**FOB Unit**
Ecuador	Cocoa	No methodology	No methodology
	Shrimp	(Shrimp exports) tonnes	No methodology

Country	**Commodity**	**Volume Unit**	**FOB Unit**
Ghana	Cocoa	No methodology	No methodology

Country	**Commodity**	**Volume Unit**	**FOB Unit**
Indonesia	Cocoa	No methodology	No methodology

	Coffee	No methodology	No methodology
	Palm Oil	Total refined palm oil (RPO) and crude palm oil (CPO) production (CPO equivalents***, million tonnes*) Total RPO and CPO exports (CPO equivalents, million tonnes*)	No methodology
	Shrimp	Shrimp exports (thousand tonnes*) Farmed shrimp production (thousand tonnes)	No methodology
	Wood Pulp	Pulp production (million tonnes*) Pulp exports (million tonnes*)	No methodology

Country	**Commodity**	**Volume Unit**	**FOB Unit**
Paraguay	Beef	Beef exports (million tonnes*)	No methodology
	Corn	Corn exports (million tonnes)	No methodology
	Soy	Soy exports (million tonnes*)	No methodology

Country	**Commodity**	**Volume Unit**	**FOB Unit**
Peru	Cocoa	No methodology	No methodology
	Coffee	No methodology	No methodology
	Shrimp	No methodology	No methodology

Yearly Trends by Country (Total Volume/FOB)

year	country_of_production	total_volume	total_fob
2013	BRAZIL	83989951.5	37554107942
2013	COLOMBIA	1040640.7	3420898064
2013	PERU	299562.3	954216922
2013	ECUADOR	169473.6	1276078486
2014	BRAZIL	89052622.6	38743801693
2014	PARAGUAY	10880615.2	5992365313
2014	COLOMBIA	1019145.8	3296278520
2014	PERU	266756.8	1127628958
2014	ECUADOR	253913.9	2460725550
2015	BRAZIL	385268032.1	61900873764
2015	ARGENTINA	74245778.1	25119797216
2015	PARAGUAY	11037627.2	4585460117
2015	COLOMBIA	1311575.4	3380033250
2015	ECUADOR	327491.1	2520331504
2015	PERU	279290.6	1008504065
2015	INDONESIA	125340.9	1166819187
2016	BRAZIL	404113435.6	64867566420
2016	ARGENTINA	85907702.4	25943213222
2016	PARAGUAY	11051719.5	4566682441
2016	COTE D'IVOIRE	1486402.9	4275430374
2016	COLOMBIA	1292831.0	3243705287
2016	ECUADOR	350769.9	2901977002

Highest Average Volumes By Port of Export

port_of_export	avg_volume	avg_fob
MACEIO	135073.58	6544268.70
RECIFE	36187.84	2540046.98
SUAPE	35276.16	1829732.97
	28923.63	0.00
PORTO VELHO	27629.00	1510464.00
UNKNOWN PORT ARGENTINA	17394.20	4907836.98
IGUAZU	15356.20	10370083.79
BAGE	12151.77	550666.82
PONTA PORA	10821.68	544124.16
URUGUAIANA	8329.26	900071.38
SANTOS	8229.42	1426192.35
PORTO MURTINHO	8194.97	2034669.45
JAGUARAO	6937.60	354116.10
UNKNOWN PORT BRAZIL	6644.27	2555455.32
GUAJARA-MIRIM	5691.84	401421.08
MANAUS	5444.44	2143351.21
CACERES	5376.82	1149647.65
ALMEIRIM	4733.21	2822927.60
SALVADOR	4664.45	1792611.39
PACARAIMA	4638.21	440377.57
QUARAI	4601.14	197745.20
KENDAWANGAN	4510.29	2591162.18

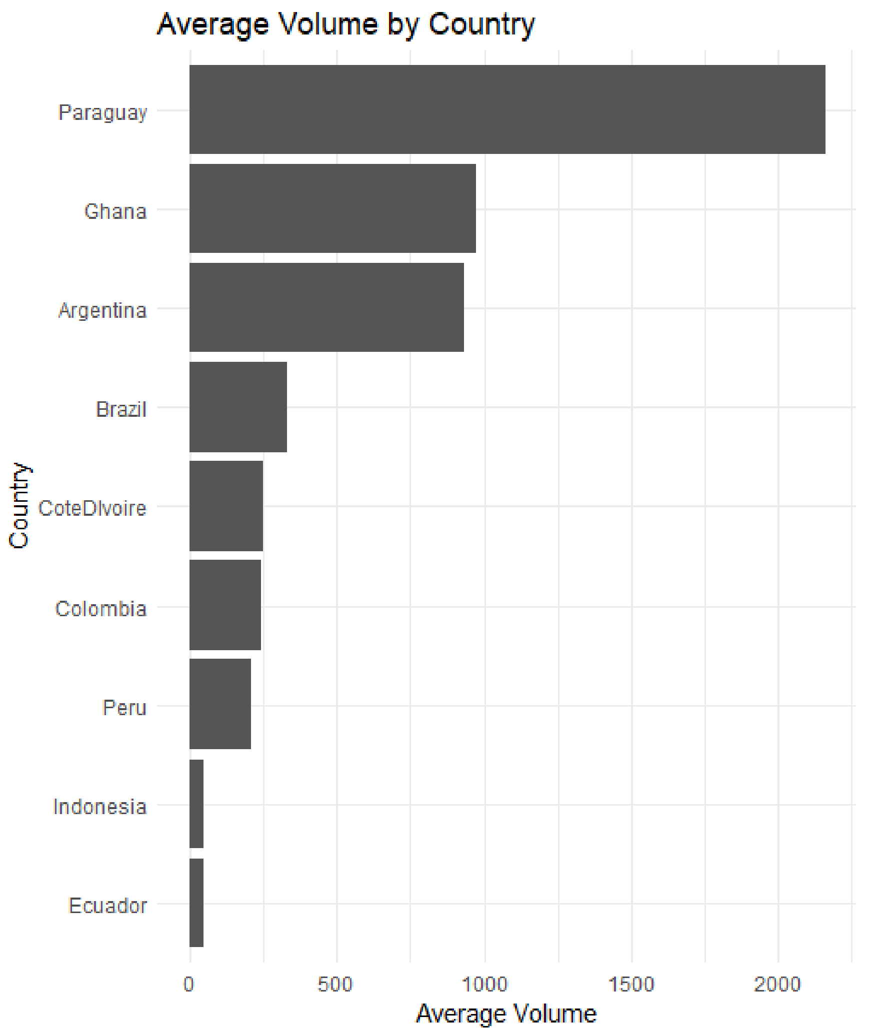
Avg Volume and FOB by country

TestingSet.R | avg_by_country

Filter

	country	avg_volume	avg_fob
1	Argentina	935.72	280172.28
2	Brazil	328.81	79179.35
3	Colombia	246.26	729868.71
4	CoteDIvoire	249.80	603442.73
5	Ecuador	44.23	259808.10
6	Ghana	973.23	4276585.99
7	Indonesia	47.99	40753.50
8	Paraguay	2158.70	954343.11
9	Peru	206.50	707067.68

Average Volume By Country



Average FOB By Country