

Article

A Big Data Analytics Framework with Interactive Dashboards for Decision-Support in Ecuador's Agricultural Sector

Ashley Aguilar-Serrano ¹, Jean Ávila-Villaprado ¹, Maritza Pinta ² and Bertha Mazon-Olivo ^{2,*}

¹ Department of Information Technology, Facultad de Ingeniería Civil, Universidad Técnica de Machala, 5.5 km Pan-American Av., Machala 070150, Ecuador; aaguilar31@utmachala.edu.ec (A.A.-S.); javila3@utmachala.edu.ec (J.Á.-V.)

² AutoMathTIC Research Group, Facultad de Ingeniería Civil, Universidad Técnica de Machala, 5.5 km Pan-American Av., Machala 070150, Ecuador; mpinta@utmachala.edu.ec

* Correspondence: bmazon@utmachala.edu.ec

Abstract

Ecuador's agricultural sector plays a strategic role in the national economy; however, agricultural data remains fragmented across heterogeneous and isolated sources, limiting integrated analysis and evidence-based decision-making. This study proposes and implements a Big Data analytics framework based on the Medallion architecture and interactive dashboards to integrate, process, and visualize agricultural indicators from INEC, ESPAC, Ecuador Open Data, and FAOSTAT for the 2010–2024 period. The proposed framework adopts the Team Data Science Process (TDSP) methodology and structures workflows into Bronze, Silver, and Gold layers using Databricks for scalable data ingestion, transformation, and dimensional modeling. Interactive dashboards were developed in Tableau Public to support dynamic analysis of agricultural production, trade, producer prices, losses, and producer profiles. A comparative performance evaluation between Databricks Free Edition and Azure Databricks was conducted using SQL analytical workloads and dashboard interaction tests. Results showed that Azure Databricks reduced query execution times by up to 57%, especially in aggregation and join operations. Usability validation with 31 agricultural stakeholders reported high acceptance levels, including a 100% recommendation rate and a data trust score of 4.45/5. The findings demonstrate that scalable and low-cost Big Data technologies can effectively support agricultural digital transformation.

Keywords: Big Data analytics; Medallion architecture; interactive dashboards; decision support systems; digital agriculture; Tableau; Databricks

Academic Editor: Jorge Bernardino,
Le Gruenwald, Elio Masciari and
Laurent D'Orazio

Received: 19 May 2026

Revised: 23 June 2026

Accepted: 26 June 2026

Published: 2 July 2026

Copyright: © 2026 by the author.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Agriculture is a strategic pillar of the global economy, with an estimated value of 4 trillion USD in 2023, highlighting its importance both in terms of production and the economy [1]. In Latin America and the Caribbean, the agricultural sector plays a fundamental role in the provision of food and raw materials; however, it faces growing challenges stemming from the economic slowdown, climate change, and the vulnerability of small-scale producers [2,3].

In Ecuador, this sector takes on an even more significant strategic role; according to the agro-economic sector indicators from the Agricultural Public Information System

(SIPA), in 2025 it contributed 8.83% of the annual real Gross Domestic Product (GDP), establishing itself as one of the country's main drivers of production [4–6]. That same year, this performance was directly reflected in the agricultural export sector: banana exports reached nearly 3.656 billion USD, while cocoa exports recorded extraordinary growth of 182.2%, with an approximate value of 3.351 billion USD [7,8]. Furthermore, in 2025, the agricultural sector recorded solid year-over-year growth of 8.6%, driven by better weather conditions, higher productivity standards, and growing external demand [9].

However, this momentum contrasts with a concerning structural reality regarding information. The Ecuadorian agricultural sector generates data from multiple institutional sources, such as the National Institute of Statistics and Censuses (INEC) through the Continuous Agricultural Area and Production Survey (ESPAC), the Ministry of Agriculture and Livestock (MAG) through the Public Agricultural Information System (SIPA), Ecuador's Open Data Portal, and the Food and Agriculture Organization's (FAO) international FAOSTAT platform. However, this information is scattered across a variety of static formats that hinder its analytical use. Data on production, agricultural losses, foreign trade, producer prices, and producer profiles are not integrated into a single repository, nor are they published in a relational structure that allows for dynamic cross-referencing.

Even though INEC and MAG have portals with agricultural information [5,10], each organizes its data around its own thematic areas: agricultural production in the case of INEC through ESPAC, and prices, foreign trade, and sectoral indicators in the case of MAG through SIPA. This institutional segmentation prevents the cross-analysis of indicators across different domains—such as relating the producer price to the volume exported for the same crop and period—which limits its usefulness for stakeholders with diverse analytical needs. The absence of a unified platform with consolidated sources, therefore, constitutes a specific gap in Ecuador's agricultural information ecosystem.

Although Ecuador has official agricultural data portals, the current platforms have three main limitations. First, they offer limited data integration capabilities, as data on production, prices, foreign trade, losses, and producer profiles are published separately and according to each entity's own criteria. Second, they offer limited support for analyzing correlations between data sources, preventing users from dynamically linking indicators such as farm-gate prices, export volumes, crop yields, and agricultural losses by crop, year, and region. Third, most scalable Big Data and business intelligence solutions require enterprise-level infrastructure or paid licenses, which limits their adoption by universities, public institutions, and small- or medium-sized actors in the agricultural sector in resource-constrained contexts. Therefore, a reproducible, low-cost, and scalable analytical framework is needed to integrate heterogeneous open agricultural data and enable the visualization of agricultural information on dashboards.

In light of this issue, the following research question arises: Is it possible to implement, using open-source or low-cost technologies, a scalable Big Data analytics framework capable of integrating heterogeneous agricultural data sources and supporting dynamic processes in Ecuador's agricultural sector through dashboards that offer acceptable levels of performance and ease of use?

To address this gap, this study implements a Big Data framework based on interactive dashboards to analyze and visualize key indicators of Ecuador's agricultural sector, integrating open data from four official sources—INEC (ESPAC), MAG, Ecuador's Open Data Portal, and FAOSTAT—covering the period 2010–2024, depending on the availability of each source. This constitutes a pilot study, with plans to incorporate data from earlier and later years in future work.

Building on this implementation, the study pursues the following specific objectives: (1) to implement a Medallion-based Big Data framework integrating at least four heterogeneous open agricultural data sources from national and international institutions; (2) to

evaluate the technical performance of the framework by comparing SQL query execution times between Databricks Free Edition and Azure Databricks across standard operation types; (3) to assess the perceived usability of the Tableau Public dashboard through a descriptive perception survey administered to agricultural sector stakeholders, using acceptance rate and Likert score distributions as evaluation criteria; and (4) to document and publish the complete data pipeline in an open repository to ensure transparency and reproducibility.

Beyond the development of visualization dashboards, the main contribution of this study is a reproducible framework for integrating heterogeneous open agricultural data into a unified analytical environment. Unlike existing agricultural dashboards, which typically focus on isolated datasets or predefined descriptive reports tailored to each entity's criteria, the proposed framework combines multiple national and international data sources using a standardized Medallion-based architecture and a dimensional modeling approach. This integration enables the generation of consistent analytical indicators for production, marketing, foreign trade, agricultural losses, producer prices, and producer characteristics. The study also contributes by documenting the complete data analysis workflow in a publicly accessible repository, including datasets and SQL notebooks, thus ensuring transparency and reproducibility. The framework is evaluated using descriptive performance metrics and user-perception measures, demonstrating its applicability for the integration, analysis, and visualization of agricultural data.

For data processing and organization, the Medallion architecture in Databricks is employed to structure the ingestion, transformation, and modeling workflow into Bronze, Silver, and Gold layers through SQL notebooks and automated jobs. The Gold layer supports a dimensional model composed of fact and dimension tables linked by unique identifiers. The resulting data warehouse is integrated with Tableau Public to provide seven thematic dashboards for the analysis of agricultural production, sales, foreign trade, losses, producer prices, and producer profiles.

From a methodological standpoint, the study is guided by the TDSP methodology, selected from among the most recognized reference frameworks for its alignment with collaborative, cloud-based work environments [11]. Used in data analysis and Big Data projects, as well as CRISP-DM and Hefesto [12–14].

Additionally, a comparative performance evaluation is conducted in two dimensions: (1) SQL query execution times between Databricks Free Edition and Azure Databricks; and (2) dashboard performance in Tableau Public, measuring initial load times, response to filters, and interaction between visualizations. This dual evaluation allows for the identification of bottlenecks in both the processing and visualization layers, guiding implementation decisions in resource-constrained contexts.

The main contributions of this work are summarized as follows:

The scientific contribution of this study is methodological, analytical, and practical in nature, rather than algorithmic. The proposed framework formalizes a comprehensive workflow that combines TDSP, the Medallion architecture, dimensional modeling, the integration of open agricultural data, and interactive visualization, all within an agricultural context with limited resources.

- A methodological framework for integrating heterogeneous agricultural open-data sources using TDSP and Medallion architecture.
- A low-cost implementation strategy combining Databricks and Tableau Public for reproducible agricultural analytics and public dissemination.
- The development of interactive dashboards for agricultural analytics.
- A comparative performance evaluation of cloud-based Big Data environments.
- An empirical usability validation involving real agricultural stakeholders.

The remainder of this paper is organized as follows. Section 2 presents the State of the Art. Section 3 describes the methodology, data architecture, and analytical workflow. Section 4 presents the experimental results and dashboard evaluation. Section 5 discusses the findings and implications. Finally, Section 6 concludes the study and outlines future research directions.

2. State of the Art

The State-of-the-Art review was conducted using the scientific literature published between 2018 and 2025, in Spanish and English, related to Big Data, data analytics, and business intelligence applied to the agricultural sector.

The adoption of Big Data technologies and advanced analytics has proven to be an effective way to transform large volumes of heterogeneous data into valuable information for producers and public entities [15–18]. Recent research also highlights the use of Medallion-type architectures, promoted by Databricks, as well as distributed processing technologies and scalable platforms, to structure robust and efficient analytical workflows geared toward handling large volumes of data in the agricultural sector [19–21]. Although the current implementation employs an exploratory subset of Ecuadorian agricultural data, the proposed framework is built upon Big Data-oriented technologies, including Apache Spark version 4.1.0 and Databricks, which are specifically designed for distributed and scalable data processing. Furthermore, the Medallion architecture enables the future incorporation of additional historical records and heterogeneous agricultural data sources as the volume and complexity of available information continue to grow. Likewise, the integration of Big Data with business intelligence and interactive visualization has proven effective in identifying production patterns, reducing risks, and improving business planning [22–24].

To provide a more structured overview, the existing literature can be classified into four thematic areas relevant to this study: (i) agricultural Big Data platforms, (ii) agricultural business intelligence (BI) systems, (iii) agricultural data lake architectures, and (iv) agricultural visualization and decision-making systems.

Regarding agricultural Big Data platforms, Wolfert et al. [15] established a foundational framework identifying how Big Data technologies—combined with IoT and cloud computing—can transform farm management cycles. Their review concluded that, while data infrastructure for agriculture is advancing, challenges related to data governance, ownership, and interoperability remain unresolved. Krisnawijaya et al. [25] conducted a systematic literature review of data analytics platforms for agricultural systems, analyzing 45 primary studies; their findings indicate that most platforms prioritize specific analytical domains (e.g., precision agriculture, yield prediction) and lack support for multi-source, open-data integration. More recently, Chergui and Kechadi [26] reviewed crop management analytics from a Big Data perspective, identifying scalability limitations in production-level deployments and an over-reliance on proprietary infrastructure.

In the area of agricultural BI systems, Sanabia-Lizarraga et al. [23] demonstrated the value of Power BI-based dashboards for analyzing agricultural foreign trade in Mexico, using a CRISP-DM workflow with two to three institutional data sources. Although effective for single-domain analysis, the approach is limited to static export data and does not support multidimensional integration across production, prices, and trade variables. Mazon-Olivo and Tinoco Egas [22] analyzed the application of business intelligence in Ecuador's agricultural sector, highlighting the potential of BI tools for cross-domain decision-making, but noting the absence of scalable architectures for data consolidation.

Concerning agricultural data lake architectures, regarding agricultural data lake architectures, a Data Lake architecture is a repository that stores massive amounts of data in its raw, original format (whether structured, images, or audio) very economically, while

a Data Warehouse stores only structured, clean, and preprocessed information to support specific business reports. The fusion of both concepts gives rise to the Data Lakehouse, a modern architecture that combines the flexibility, massive storage capacity, and low cost of a Data Lake with the structure, quality rules, security controls, and ACID transactions of a Data Warehouse on a single, unified platform [27]. Studies such as those of Zaharia et al. [21] proposed the Lakehouse paradigm as an evolution of traditional data warehouses, unifying batch and streaming processing with ACID transaction support through platforms such as Databricks and Delta Lake. This architecture enables the structured layering of data—from raw ingestion to curated analytical models—directly applicable to heterogeneous agricultural data environments. Ouafiq et al. [28] proposed a data lake architecture for smart farming as part of a data migration strategy for Big Data analytics, highlighting the need for metadata management and data quality controls when integrating IoT-generated agricultural data.

Finally, in the area of agricultural visualization and decision-making systems, Chamorro-Padial et al. [20] reviewed agricultural data-sharing practices and identified interactive visualization as a critical enabler for multi-stakeholder decision-support. Their analysis indicated that most existing systems restrict data access to single-institutions.

Table 1 summarizes the key attributes of representative works in these four categories, comparing data sources, technical architectures, and identified limitations.

Table 1. Comparative summary of related works by thematic category.

Category	Study	Data Source	Technical Architecture	Limitations
Agricultural Big Data Platform	Wolfert et al. [15]	IoT sensors, farm records	IoT + Cloud (conceptual)	No open-data integration; limited to primary production
	Krisnawijaya et al. [25]	Multiple agricultural datasets	Review of 45 platforms (SLR)	Fragmented scope; no multi-source open-data integration
	Chergui and Kechadi [26]	Crop, weather, soil records	Distributed Big Data	Proprietary infrastructure; limited scalability for developing countries
Agricultural BI System	Sanabria-Lizarraga et al. [23]	2–3 institutional sources (Mexico)	CRISP-DM/Power BI	Single-domain; no production or loss data; focused on foreign trade data.
	Mazon-Olivo and Tinoco Egas [22]	INEC, MAG (Ecuador)	BI (conceptual)	No scalable Big Data architecture; limited source integration
Agricultural Data Lake Architecture	Zaharia et al. [21]	Enterprise/streaming data	Lakehouse (Databricks, Delta Lake)	Focused on enterprise contexts; no agricultural application
	Ouafiq et al. [28]	IoT sensors (Smart farming)	Data lake + migration strategy	Limited to IoT data; no open institutional data integration
Agricultural Visualization and Decision-Making	Chamorro-Padial et al. [20]	Institutional agricultural data	Data sharing frameworks	Mostly single-institutions; limited dynamic filtering and interactivity

However, most of the solutions identified in the literature are developed in contexts with substantial technological infrastructure and budgets, which limits their replicability in environments such as Ecuador's, where public institutions, universities, and small producers can hardly afford enterprise licenses for Big Data or business intelligence platforms.

Despite these limitations, various national and international government agencies have developed statistical portals, open-data repositories, and interactive dashboards that serve as important resources for agricultural analysis. Notable among these are the National Institute of Statistics and Censuses (INEC) through the Continuous Agricultural Area and Production Survey (ESPAC), the Ministry of Agriculture and Livestock (MAG) through the Public Agricultural Information System (SIPA), the Ecuador Open Data Portal, and the international FAOSTAT platform of the Food and Agriculture Organization (FAO). These tools allow users to consult indicators related to production, foreign trade, prices, and agricultural yields; however, the data is distributed across different environments, with limited capabilities for integration, relational analysis, and dynamic visualization within a single platform.

To select technological tools, a comparative analysis was conducted based on criteria such as analytical capabilities, computing power, storage, cost, and dashboard publishing. Tables 2 and 3 present the comparative analysis of the tools used in this study. Table 2 compares Big Data processing platforms Databricks Free Edition and Azure Databricks, while Table 3 compares the visualization platforms Tableau Public and Tableau. Databricks is used for data processing and dimensional modeling, whereas Tableau Public is used for interactive visualization and public dissemination of results.

Table 2. Big Data Analytics Tools (Databricks Free Edition vs. Azure Databricks).

Criteria	Metric/Scale	[29] Databricks Free Edition	[30] Azure Databricks (Paid)
Cost and Licensing			
Licensing cost	USD/month	\$0 (fair-use quota; non-commercial use only)	Pay-per-use DBU ~\$0.40/hr (all-purpose) + Azure VM costs (billed separately)
Trial/initial credits	USD one-time	None (always free, no credits)	\$200 Azure credit (valid 30 days)
Compute and Scalability			
Compute type	Compute options	Serverless only (no custom clusters, no GPU)	Full cluster management (custom sizes, GPU, Photon engine)
Max concurrent job tasks	Number of tasks/account	5 concurrent job tasks per account	Unlimited (by budget/quota)
Max active pipelines	Number of pipelines/type	1 active pipeline per pipeline type	Unlimited (by budget)
SQL Warehouse size	Cluster size	1 warehouse, max 2X-Small cluster size	All sizes (2X-Small to 5X-Large)
GPU support	Yes/No	No	Yes (ML Runtime + GPU clusters)
Storage and Data Access			
Storage limit	GB/configuration	Limited DBFS (Unity Catalog; no custom storage locations)	Unlimited (Azure Blob/ADLS)
Native read connectors	Source types count	8 formats (CSV, JSON, Delta, Parquet, XML, Kafka, MySQL, PostgreSQL)	20+ sources (enterprise, Big Data, streaming, APIs)
Write/export formats	Export formats count	9 formats (CSV, JSON, Delta, Parquet, XML, Kafka, MySQL, PostgreSQL, SQL Server)	All formats + streaming outputs
AI and Analytics Capabilities			
ML support level	0 = None, 1 = Limited, 2 = Advanced, 3 = Full	1—Limited (Genie AI assist; model serving with limits; no GPU serving)	3—Full (MLflow, AutoML, ML Runtime, GPU clusters)

Delta Lake/ACID transactions	Yes/Partial/No	Partial (Unity Catalog only; custom storage not supported)	Yes—Full Support (Delta Lake + Unity Catalog)
Feature Checklist (Quantitative)			
Spark full support (RDD, Scala, R)	Yes/No	No (R, Scala, Spark RDD not supported)	Yes (full Spark engine)
SSO/SCIM access control	Yes/No	No (email OTP/Google/Microsoft only)	Yes (Azure AD/Entra ID, full RBAC)
SLA/support coverage	SLA/No SLA	No SLA (no support; fair-use policy applies)	Yes (enterprise SLA + Databricks support)
Photon engine (query acceleration)	Yes/No	No	Yes
Total capabilities supported (score)	Score/10	5/10 (Spark Connect, ANSI SQL, Serverless SQL WH, Delta tables, Genie AI assist)	10/10 (Spark full, MLflow, Delta Lake, Unity Catalog, Photon, GPU, AutoML, SSO, SLA, streaming)
Access Control and Administration			
Access control level	0 = None, 1 = Basic, 2 = Advanced	1—Basic (Unity Catalog limited; no SSO, no SCIM, no private networking)	2—Advanced (Azure AD/Entra ID, SSO, full RBAC, Unity Catalog)
Commercial use allowed	Yes/No	No (non-commercial only)	Yes
Suitability for This Study			
Selection score	2 = Selected, 1 = Partial use, 0 = Not selected	2—Selected (ETL and dimensional modeling pipeline)	1—Partial use (benchmark comparison only)
Rationale	Justification	Zero cost; serverless compute; no infrastructure mgmt; adequate for study-scale ETL; non-commercial context fits academic research	Superior performance (Photon + custom clusters); pay-per-use cost limits sustained use in public sector/university

Table 3. Big Data Analytics Tools (Tableau Public vs. Tableau Paid).

Criteria	Metric/Scale	[31] Tableau Public	[32] Tableau (Paid Standard)
Cost and Licensing			
Licensing cost	USD/user/month (billed annually)	\$0 (free; all content published publicly)	Viewer: \$15 Explorer: \$42 Creator: \$75 (billed annually)
Free trial	Days	None (always free)	14-day free trial (no credit card required)
Storage and Data Access			
Row limit per data source	Million rows	15 M rows or 1 GB (whichever comes first)	No row limit (depends on cloud infrastructure)
Total account storage	GB	10 GB per account	Depends on the Tableau Cloud/Server configuration
Native read connectors	Source types count	~7 sources (CSV, Excel, JSON, Google Sheets, OData, PDF, spatial files) No live DB connections	15+ sources (cloud DBs, MySQL, PostgreSQL, Big Query, Redshift, Snowflake...) Live connections supported
Export/write formats	Supported formats	CSV, PDF, image (local export only; no embedding)	CSV, PDF, image, Tableau extract (.hyper); embedded analytics supported
Dashboard and Visualization Capabilities			

Dashboard creation	Yes/Limited/No	Yes—Fully interactive and dynamic dashboards, advanced chart types	Yes—Full (advanced dashboards, data modeling, Tableau Prep)
Public dashboard publishing	Public/Private/None	Public only (all visualizations publicly accessible; no privacy control)	Private (Tableau Server/Cloud required for sharing)
Offline work/local save	Yes/No	No (internet required; no local. twbx save)	Yes (Tableau Desktop; local. twbx save)
Dashboard embedding	Yes/No	No (public iFrame only; no auth control)	Yes (Tableau Embedded Analytics)
AI-driven analytics	0 = None, 1 = Basic, 2 = Advanced	1—Basic (no ML/statistical modeling)	2—Advanced (Explain Data, predictions, Tableau Pulse AI)
Feature Checklist (Quantitative)			
Access control/role-based permissions	Yes/No	No (public-only; no access control or privacy mechanisms)	Yes (Creator/Explorer/Viewer roles)
Live database connections	Yes/No	No (file-based extracts only)	Yes (live + extract modes)
Data preparation tool bundled	Yes/No	No (basic file import only)	Yes (Tableau Prep Builder bundled with Creator)
Private/secure sharing	Yes/No	No (all content public)	Yes (Tableau Cloud/Server with authentication)
Total capabilities supported (score)	Score/8	4/8 (Interactive dashboards, advanced charts, free public sharing, Tableau Public profile)	8/8 (dashboards, advanced charts, live connections, Tableau Prep, RBAC, private sharing, embedding, Pulse AI)
Suitability for This Study			
Selection score	2 = Selected, 1 = Partial use, 0 = Not selected	2—Selected (interactive dashboard publishing for study)	0—Not selected (licensing cost prohibitive for public institutions)
Rationale	Justification	Zero cost; public sharing; sufficient interactivity for agricultural stakeholders; 10 GB/15 M rows adequate for study-scale data	Requires per-user annual license (\$15–\$75/user/mo); not viable for university/public sector budgets; private sharing adds infrastructure cost

3. Materials and Methods

3.1. Methodological Approach

The research was conducted using a descriptive–applicative approach, focused on analyzing information from Ecuador’s agricultural sector over the past decade, using Big Data tools and interactive visualization. The source data were obtained from open-data websites in Ecuador, which were processed and organized to provide useful information for decision-making by producers, local and national governments, and export associations, among others.

To guide the stages of data integration, processing, modeling, and visualization, the TDSP methodology was adopted as a structured framework. Regarding data organization, the Medallion architecture was implemented, which establishes three hierarchical levels: Bronze, intended for storing data in its original state; Silver, corresponding to cleaned and structured data; and Gold, oriented toward data prepared for analysis and visual representation.

Data analysis adopted the Team Data Science Process (TDSP) methodology, as shown in Figure 1, because it aligns with the Big Data environment due to its iterative and

collaborative approach, based on three main phases: business understanding and data acquisition and understanding, modeling, and deployment.

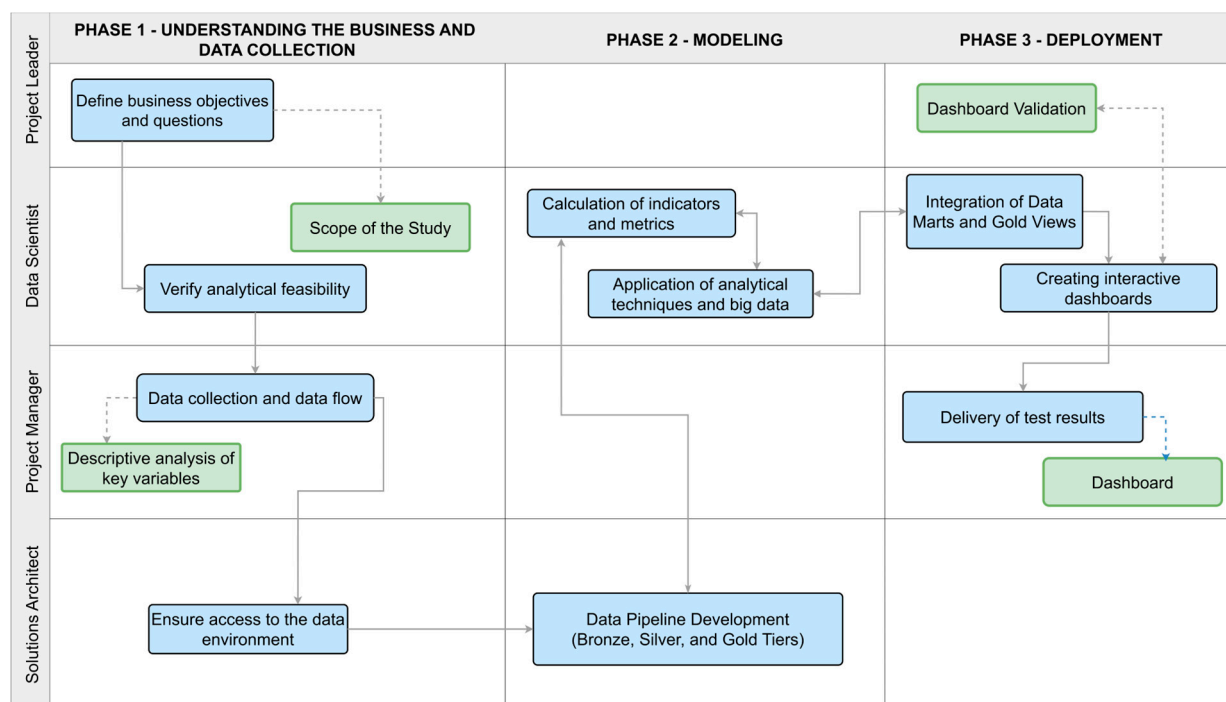


Figure 1. Model of the phases of the TDSP methodology.

3.2. Application of the TDSP Methodology

3.2.1. Business Understanding and Data Acquisition

Business Understanding

This study focuses on Ecuador's agricultural sector, which is a fundamental pillar of the national economy due to its contribution to production, employment, and agricultural exports. However, managing the sector's large volume of data presents challenges not only due to its dispersion and lack of integration but also due to the limited application of Big Data and predictive analytics technologies that would enable correlations between production and commercial indicators. Furthermore, there remains a lack of an approach that further drives agricultural digitization and promotes the use of data as a strategic resource to improve analysis and decision-making by producers, public entities, and other sector stakeholders.

Based on these needs, and to guide the analysis of Ecuador's agricultural sector, business questions were defined to guide the construction of the dashboards; these are detailed in the following section, along with their dimensions and indicators.

Data Acquisition and Understanding

The data forming the basis for the analysis comes from official sources and freely accessible institutional websites dedicated to collecting agricultural statistics from Ecuador over various time periods. Among the main sources are the following: INEC through ESPAC, which provided data on planted and harvested areas, yields, and losses; MAG [10], with information on prices and foreign trade; FAOSTAT [33], for international comparative data; and Ecuador Open Data [29], which provided records such as producer prices and domestic market data.

The datasets were analyzed to identify key variables and, based on the defined indicators, establish the formulas used to calculate the KPIs. Additionally, business questions

aligned with the main pillars of the agricultural sector were established, which allowed for the structuring of the dashboards and the determination of their dimensions and indicators. Table 4 presents the relationship between dashboards, dimensions, KPIs, and business questions that guide the analysis.

Table 4. Definition of Dimensions and KPIs by Dashboard.

Dashboard	Dimensions	Business Questions	KPIs + Formula
Agricultural Production	Year, Product	- Which provinces register the highest production in tons per year?	Production = Σ Production
		- In which provinces is the highest sales volume concentrated in tons per year?	Sales = Σ Sales
		- Which provinces obtain the highest yield (t/ha) per year and product?	Yield = Production/Harvested Area
Flowers	Year, Species	- Which species registers the highest production of cut stems per year?	Production = Σ Cut stems
		- How has the yield (cut stems/hectares) evolved by species and year?	Yield = Cut stems/Harvested area
Production and losses	Year, Crop, Cause	- How has the lost area (hectares) evolved by cause and year?	Lost Area (Ha.) = Σ Lost Area (Ha.)
Producer Profile	Year, Range ha	- How is the participation of Agricultural Production Units distributed by gender, education level, ethnicity, and age of the producer according to the size of the production unit?	Participation (%) = (Producers per category/Total producers) $\times$ 100
Producer prices	Year, Month, Product	- How has the weighted average producer price (USD) varied by product and year?	Average price = Σ Weighted price
Foreign Trade (Ecuador Overview)	Year, Product	- What are the export volumes (metric tons) and export value (thousand USD) by product and year?	Exports = Σ Export value
		- What are the import volumes (metric tons) and import value (thousand USD) by product and year?	Imports = Σ Import value
		- What is the trade balance in volume (metric tons) and value (thousand USD) by product and year?	Trade Balance = Σ Export Value – Σ Import Value
FAOSTAT Foreign Trade by Partner Country	Year, Country, Partner, Product	- How have the export volume (metric tons) and value (thousand USD) evolved by product, partner country, and year?	Exports = Σ Export value
		- How have the import volume (metric tons) and value (thousand USD) evolved by product, partner country, and year?	Imports = Σ Import value
		- How has the trade balance in volume (metric tons) and value (thousand USD) evolved by product, partner country, and year?	Trade Balance = Σ Export Value – Σ Import Value

These expressions enable the standardization of data processing, ensure the consistency of results, and facilitate the interpretation of the presented indicators.

3.2.2. Modeling

In this phase, data modeling was carried out using the Medallion architecture on the Databricks Platform, structured in three layers: Bronze, Silver, and Gold, as illustrated in Figure 2. This approach allowed for the progressive organization and transformation of data from selected sources, including INEC, ESPAC, Ecuador Open Data, and FAOSTAT, which were initially ingested in formats such as Excel and CSV.

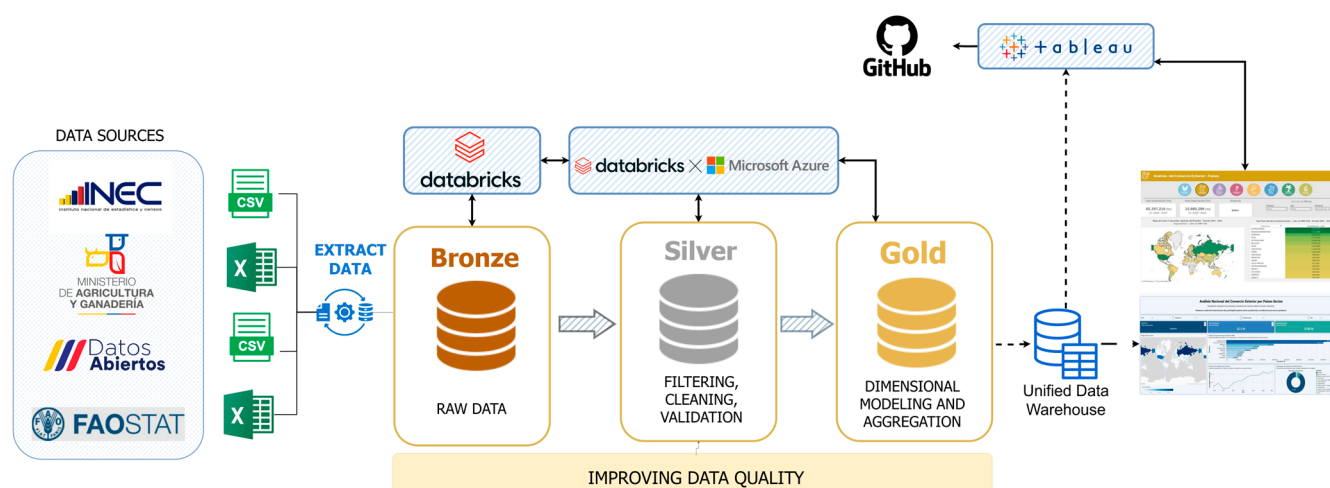


Figure 2. Medallion-based data analysis architecture. The Medallion architecture in Databricks, structured in three layers: Bronze (ingestion), Silver (transformation), and Gold (analytical modeling).

Following the principles of the Medallion architecture, data processing was organized into three sequential layers, which are described below.

Bronze Layer: Data Ingestion

The Bronze layer represents the ingestion stage of the Medallion architecture. Its purpose is to store data from different agricultural information sources in their original structure.

For the Databricks implementation, the `medallion_lab` catalog was created with three schemas: Bronze, Silver, and Gold. The datasets obtained from official sources, including INEC, MAG, Open Data Ecuador, and FAOSTAT, were collected in the formats provided by each institution, primarily CSV and Microsoft Excel (XLSX) files. These files were subsequently uploaded to the Bronze schema through Unity Catalog. Each dataset was stored as an independent table using the `_raw` suffix as a naming convention.

At this stage, no business transformations, data integration processes, or dimensional modeling activities were performed. The Bronze layer acts as a repository of raw data, preserving the original structure, attributes, and values provided by each source. This approach ensures data traceability and provides a reliable foundation for subsequent cleaning, harmonization, and analytical modeling processes in the Silver and Gold layers.

Silver Layer: Data Cleaning, Harmonization, and Dimensional Modeling

The Silver layer was implemented through seven SQL notebooks executed sequentially. Each notebook reads the raw files from the Bronze layer and generates cleaned datasets in the Silver schema using the `_clean` suffix. This layer was responsible for both data cleaning and standardization, as well as the construction of the relational model subsequently used for analysis.

The transformation process focused on variable harmonization and the normalization of equivalent entities originating from different sources. Common attributes were consolidated into reusable dimensions, resulting in a total of 25 dimensions, including year, product, province, region, and producer. Each dimension was normalized and assigned a unique, sequentially generated identifier, which was used to consistently relate information across the different tables of the model.

To integrate information from multiple sources, a semantic harmonization process was carried out to unify different terms representing the same business concept. For example, “Banano” (ESPAC) and “Banana” (FAOSTAT) were consolidated into a single record within the product dimension. Based on these standardized dimensions, fact tables and foreign key relationships were defined, forming the relational model used for analysis. The complete structure of this relational model is presented in Figure 3.

Once the dimensional structure was defined, the following data cleaning and standardization techniques were applied:

- (a) **Data type conversion.** Data type conversion. Numeric identifiers stored as text were converted to integers using `TRY_CAST(REGEXP_REPLACE(column, '[,\.\.]', '') AS INT)`, removing formatting characters and performing safe type conversion. This regex-based pattern was applied to fact tables; for dimension tables, `REPLACE` was used equivalently to strip potential formatting characters before casting.
- (b) **Handling of regional numeric formats.** Quantitative columns were processed using the custom user-defined function (UDF) `clean_number()`, which standardizes numeric values from heterogeneous source formats by removing symbols, handling decimal separators, eliminating thousands of separators, and safely converting the resulting values to `DOUBLE`.

```
CREATE OR REPLACE FUNCTION clean_number(x STRING)
RETURNS DOUBLE
RETURN
TRY_CAST(
CASE
WHEN x IS NULL OR TRIM(x) IN ('', '.', '- ', 'N/A') THEN '0'
ELSE
REGEXP_REPLACE(
REGEXP_REPLACE(
REPLACE(REPLACE(x, '%', ''), ',', '.'),
'(?<=\d),(?=\d{3}\b)', ''),
'(?<=\d)\.(?=\d{3}\b)', '')
END AS DOUBLE
);
```

- (c) **Standardization of geographic names and countries.** Province names were normalized using `CASE WHEN UPPER(TRIM(province))` for the 24 provinces of Ecuador. Trading partner countries were translated from Spanish to English using `CASE WHEN LIKE` expressions; for example, “PAÍSES BAJOS” was transformed into “Netherlands”. Additionally, diacritical characters were removed using `TRANSLATE()` to facilitate matching with ISO codes.
- (d) **Handling of missing data.** Records containing `NULL` values in foreign keys required for the dimensional model were excluded using `WHERE IS NOT NULL` conditions. No imputation techniques were applied. Quantitative columns were processed through the `clean_number()` UDF, which performs safe type conversion and standardizes invalid or non-processable values according to predefined rules.

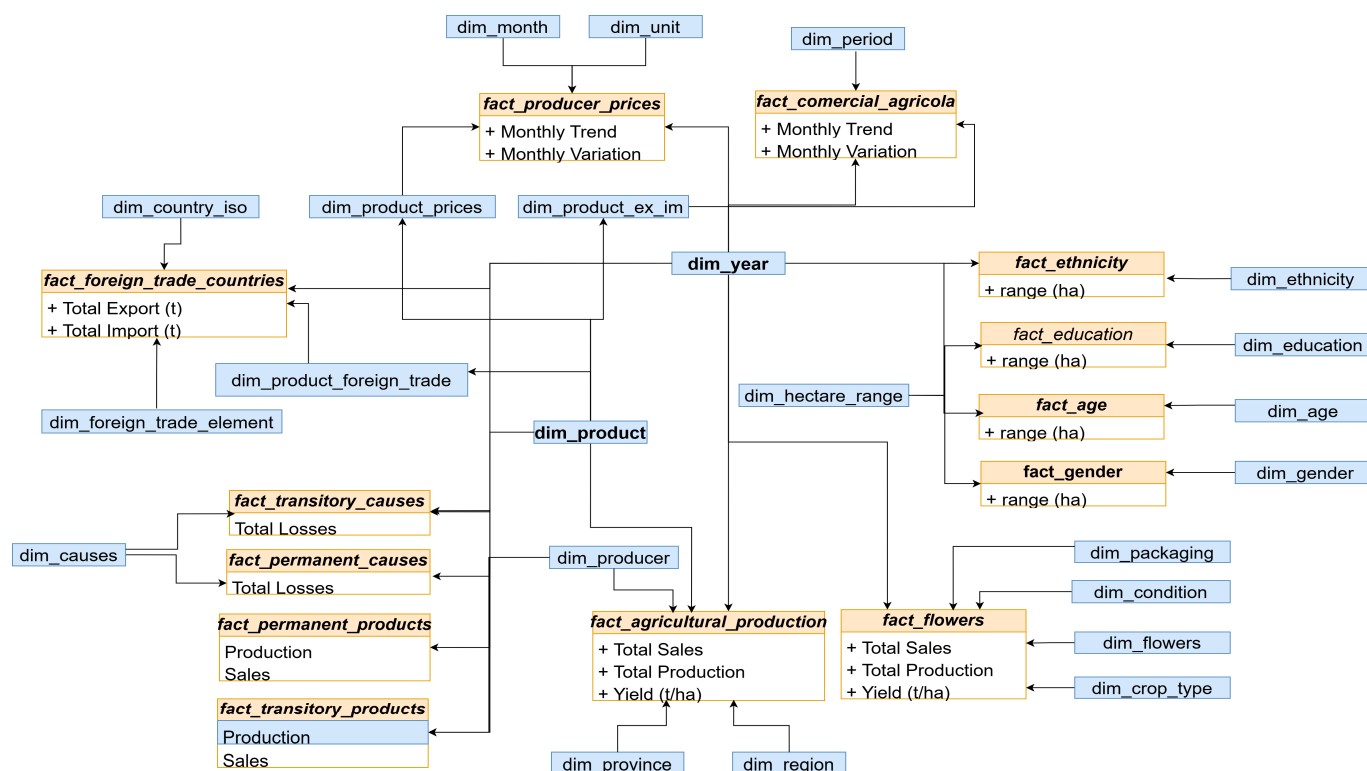


Figure 3. Dimensional Data Model.

Gold Layer: Analytical Data Modeling

The Gold layer corresponds to the final stage of the Medallion architecture, where the information previously integrated, cleansed, and standardized in the Silver layer is materialized into datasets optimized for analysis. Based on the dimensional structure defined during the transformation and cleansing stage, the data warehouse was built through the integration of fact tables and their associated dimensions. As a result, several specialized data marts were generated to represent the main areas of the Ecuadorian agricultural sector, including agricultural production (*fact_produccion_agricola*), the floriculture sector (*fact_flores*), producer prices (*fact_precios_productor*), and foreign trade by partner country (*fact_comercio_exterior_paises*).

Additionally, in those areas where it was appropriate, key performance indicators (KPIs) were incorporated as derived attributes within the fact tables. Among the most relevant are agricultural yield, expressed in tons per harvested hectare, and floriculture yield, expressed in stems per cultivated hectare, among other indicators aimed at sector-specific analysis.

The resulting data warehouse was designed to be consumed by business intelligence tools. The Gold layer was used as a data source for Tableau Public, where seven interactive dashboards were developed, as well as for Databricks native dashboards.

3.2.3. Deployment

Once the views are consolidated in the Gold layer, they are consumed directly in Databricks using native dashboards. Each Data Mart is integrated into the dashboards via SQL queries on the Gold views, which allow the various widgets and visualizations to be dynamically fed. As a result, interactive dashboards are built that can automatically reflect any updates to the data.

However, it is important to note that Databricks has certain limitations regarding dashboard publishing, as access is restricted to users within the same account or who are authenticated within the environment. Therefore, there is no native option for generating

fully public dashboards in the Free Edition. For this reason, tools like Tableau offer greater capabilities for the open dissemination of information, allowing for free publication and access by external users.

To ensure the transparency and reproducibility of the proposed system, both the data sources used, and the dashboards developed are available online.

On the one hand, the datasets used to build the data marts and dashboards were published in a public repository, facilitating their consultation, validation, and reuse by researchers, students, and stakeholders in the agricultural sector. Datalink: https://github.com/Jean66534/data_repository_ec_agro.

On the other hand, the interactive dashboards were deployed on the Tableau Public platform, allowing open access to the analytical visualizations. This promotes the dissemination of the results and encourages the practical use of the information to support decision-making. Link to the dashboard: <https://public.tableau.com/shared/CB66JQ2CT>.

4. Results

The results of this study are presented in two complementary sections. The Section 4.1 focuses on the visualization and analytical capabilities of the dashboards developed from the agricultural data lake across different Big Data platforms. This evaluation examines the implementation of the dashboards, the visualization features provided by each platform, and the strengths and limitations identified during their use. Furthermore, the agricultural business insights derived from the dashboards published in Tableau Public are presented, illustrating the descriptive analyses that can be performed using the available visualizations. The Section 4.2 provides a comparative assessment of the selected Big Data platforms in terms of performance and cost, enabling an evaluation of their suitability for supporting agricultural data processing, analysis, and visualization tasks.

4.1. Results of the Agricultural Analysis Using Dashboards

Three platforms were evaluated for the development and visualization of agricultural dashboards: Databricks Free Edition, Azure Databricks Free Trial, and Tableau Public. This section analyzes the visualization capabilities, strengths, and limitations of each platform and subsequently presents the agricultural business insights obtained from the dashboards published in Tableau Public, which was selected as the final dissemination platform.

4.1.1. Visualization in Databricks Free Edition

Figure 4 shows the dashboard developed in Databricks Free Edition, which demonstrates an intuitive working environment, particularly for data integration and transformation (ETL) processes, facilitating the implementation of analytical architectures. However, during development, limitations associated with the use of computing resources were identified: the free version only offers a preconfigured 2X-Small serverless environment, without the ability to manage multiple clusters, which restricts processing. Regarding dashboard creation, the tool incorporates innovative features such as AI-assisted code and chart generation (Genie Code), although it still has limitations in the variety of visualizations and in advanced features like predictive analytics, which are currently in beta. Likewise, one of the main limitations lies in dashboard publishing, which is primarily geared toward closed collaborative environments, making public sharing more difficult. Additionally, the Free Edition of Databricks is designed mainly for learning and experimentation and therefore includes restrictions on resource availability and platform usage. Computational resources may become temporarily unavailable when usage limits are reached, while prolonged inactivity can result in the removal of inactive accounts. Moreover, the use of Databricks in production environments with private data and enterprise

requirements generally requires paid subscriptions, leading to additional licensing and infrastructure costs. Overall, Databricks positions itself as a robust solution for data analysis in business contexts, prioritizing security, scalability, and data integration capabilities over advanced visualization features.

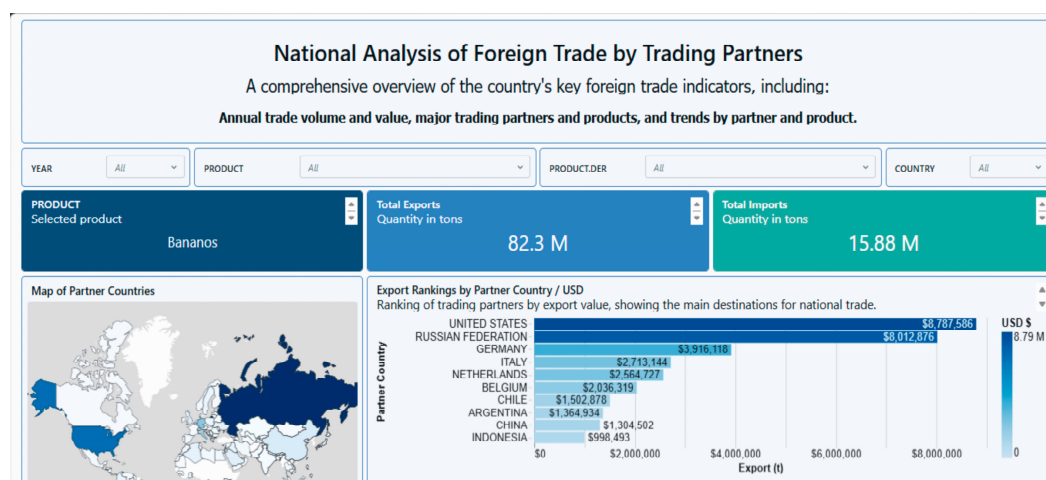


Figure 4. Dashboard in Databricks Free Edition—National Analysis of Foreign Trade by Partner Countries.

4.1.2. Visualization Using the Azure Databricks Free Trial

In the case of the Azure Databricks trial version runtime 17.3, greater complexities were identified in the initial setup, particularly when connecting to storage services such as Azure Data Lake Storage Gen2, where credentials must be managed using mechanisms like Access Keys or Shared Access Signature (SAS) token Tokens (more tightly controlled permissions), as well as the assignment of roles and permissions at the user level within the Microsoft Azure environment.

While this version allows for the creation and management of clusters, it has significant limitations regarding visualization capabilities in trial environments or free versions: it does not allow for the creation of dashboards or the use of integrated graphical tools; these features may be limited or not fully available. It also offers support features based on artificial intelligence agents (Genie Code). In this context, its use is primarily geared toward data processing, which may limit its application in scenarios focused on communicating results. Furthermore, access to advanced functionalities, larger-scale resources, and enterprise-oriented capabilities requires migration to paid subscription plans, which introduces additional operational and infrastructure costs.

4.1.3. Visualization in Tableau Public

Dashboards developed in Tableau stand out for their high level of customization and their focus on the user experience (Figure 5), allowing for the creation of interactive, dynamic, and visually appealing visualizations. The tool is intuitive and facilitates data exploration, as well as the publication and sharing of dashboards in public environments, which represents a significant advantage over other platforms.

Furthermore, Section 3.2.3 (Deployment) provides access to the Tableau Public environment where the dashboards are hosted, enabling readers to interact with the visualizations presented in this study and explore the complete collection of publicly available agricultural dashboards.

However, limitations were identified in its free version, such as restrictions on connecting to external data sources, especially in Big Data environments, where it is not pos-

sible to establish direct integrations with platforms like Databricks or cloud services, being limited mainly to local files or basic repositories.

Likewise, performance may be affected during the loading or publishing of dashboards. In this context, Tableau positions itself as a robust tool for visualizing and communicating results, although with limitations regarding the direct handling of large volumes of data in its public version.

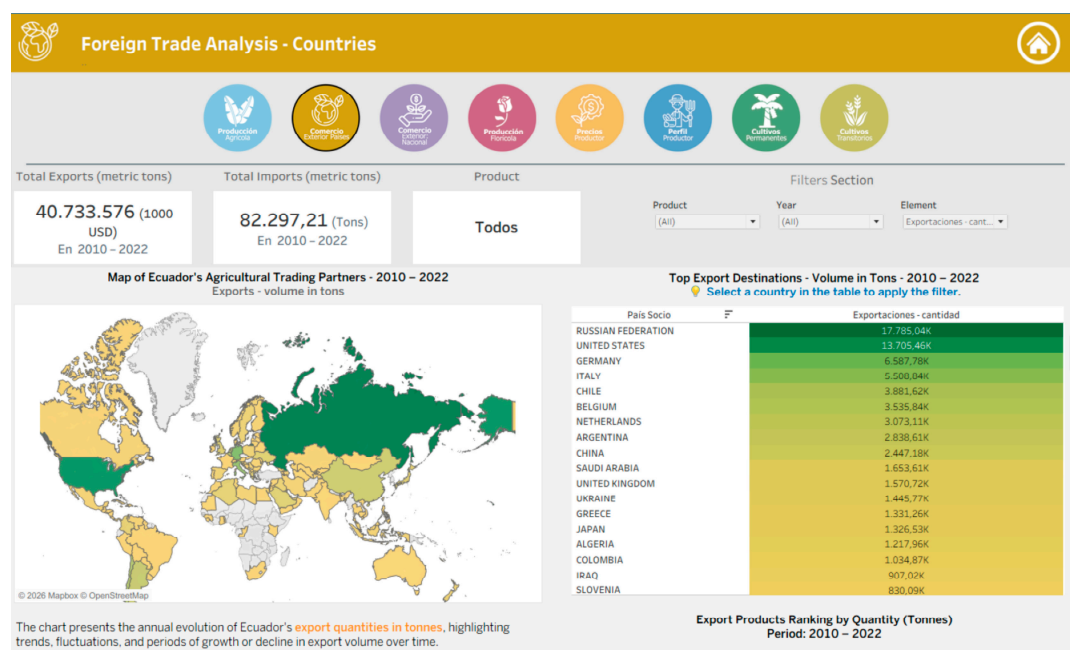


Figure 5. Dashboards in Tableau Public—Analysis of Foreign Trade by Partner Countries.

4.1.4. Agricultural Business Insights Derived from the Dashboards Published on Tableau Public

The dashboards were designed primarily to provide descriptive agricultural information through the integration and visualization of data from multiple sources within a single platform. They allow users to explore production, sales, yields, agricultural losses, producer prices, foreign trade, and producer profiles over time, facilitating the identification of patterns and trends that are difficult to observe when data is dispersed across different institutions such as INEC-ESPAC, MAG-SIPA, Ecuador Open Data, and FAOSTAT. Additionally, the foreign trade module includes time-series forecasts for selected indicators, specifically annual exports, trade balance by quantity (tons), and trade balance by value (USD).

These projections were generated using the Exponential Smoothing forecasting model built into Tableau, which provides an estimate of future trends based on historical data, specifically from the 2010–2024 period.

Table 5 summarizes the main categories of information derived from the analytical dashboards: (i) the identification of the provinces and crops with the highest production and yield levels; (ii) the determination of the crops and underlying causes associated with the most significant agricultural losses; (iii) the analysis of products exhibiting the greatest variation in producer prices; (iv) the characterization of the main export products and trading partner countries according to their monetary value and traded volume; and (v) the description of producer profiles based on gender, age, educational attainment, ethnicity, and farm size.

Taken together, these findings provide descriptive and diagnostic evidence that may support the identification of relevant patterns, trends, and disparities for decision-making

in the agricultural sector. However, given the observational and descriptive nature of the analysis, the results should not be interpreted as evidence of causal relationships among the variables examined.

Table 5. Agricultural Insights Supported by the Final Dashboards.

Dashboard	Analytical Insight Supported	Relevant Agricultural Information Identified
Agricultural Production	Allows analysis of agricultural production and sales through filters by Province, Year, and Product, enabling spatial and temporal queries on production volume, sales, and yield. Identify of the provinces and crops with the highest productive performance, informing territorial planning and resource allocation decisions.	With the filters Product = Banana, Year = 2024, and Province = (All) applied, total national production reached 7.56 million tonnes, with sales of 7.22 million tonnes, indicating a high commercialization rate for the crop. Los Ríos emerges as the leading producing province, with 3 million tonnes (3011.20 K), a yield of 44 t/ha, and sales of 2943.11 K tonnes, underscoring its central role in the banana supply chain.
Flowers	Allows identification of flower species with the highest production of cut stems and yield through Year and Flower Species filters, enabling comparative analyses of productive performance across species, as well as the historical evolution of yield for a selected species.	With the filters Year = 2024 and Flower Species = Rosa applied, the total cut stems reached 15,533.47 M (Tm), yielding 2.90 M (T/ha) in 2024. In the species ranking, Rosa shows by far the highest production of cut stems (close to 15 B), although Clavel records the highest relative yield (29.70 M Tallos/Ha) compared to Rosa's 2.90 M. The historical evolution of Rosa's yield (2014–2024) reveals a notable peak in 2019 (8.32 M Tallos/Ha), followed by a sharp decline in 2020 (2.75 M) and subsequent fluctuations between 2.56 M and 5.66 M, with 2024 closing at 2.90 M, suggesting variable productive performance possibly linked to external shocks.
Production and Losses	Allows identification of the main causes of agricultural area loss in permanent and transitory crops and their evolution over time through Year and Product filters. Combining these filters enables comparative queries on the intensity of loss by cause and crop.	With the filters Year = 2024 and Product = All applied, the total cumulative production of permanent crops between 2014 and 2024 reached 223.37 million tons, with sales of 15.23 million tons in 2024. Sugarcane shows the highest yield (83.54 T/ha). Regarding losses, pests are the main cause of damage, affecting 15,594 hectares, followed by other causes (10,909 ha) and drought (7612 ha).
Producer Profile	Allows characterization of agricultural producers through the Year filter, segmenting information by gender, education level, age group, ethnicity, and the size of the Agricultural Production Unit (UPA).	With the filter Year = 2024 applied, Ecuadorian agricultural producers are predominantly male (75.02%) compared to 24.98% female. Female participation decreases as UPA size increases: women represent 46.40% of producers on farms of less than 1 hectare, but only 11.30% on units of 200 hectares or more, evidencing a gender gap associated with land tenure. The dominant age group is producers aged 65 and over (32.90%), followed by those aged 55–64 (23.63%), reflecting an aging producer base.
Producer Prices	Allows the analysis of the temporal evolution of weighted-average producer prices, filtered by Year and Product, at both annual and monthly levels, in total value (USD) and per-kilogram price (USD/kg). This enables monitoring price trends and identifying relevant seasonal variations.	With the filters Product = Banano Cavendish and Year = 2025 applied, the weighted-average price stood at \$1.30 per kilogram and \$42.26 per box (a unit of approximately 43 lb). The annual series of the weighted price in USD shows an upward trend, from \$6.22 in 2013 to \$7.17 in 2025, with a slight decline in 2017 (\$6.16) following a 2015 peak (\$6.55). At the monthly level, during 2025, the weighted price increased from \$6.85 in January to \$7.25 from February onward, remaining stable through May, suggesting an upward price consolidation for Banano Cavendish in the short term.
Foreign Trade (Ecuador Overview)	Allows the analysis of Ecuador's agricultural foreign trade through filters by Product and Year, evaluating the evolution of exports and imports in both volume (tonnes) and value (USD), as well as a ranking of products according to their contribution to the trade balance.	With the filters Product = All and Year = 2025 applied, total exports reached 44,480 million tonnes (USD 53,457.95 M), while total imports reached 42,536 million tonnes (USD 57,220.98 M), indicating a positive physical trade balance but a deficit in value. In the product ranking, palm oil leads with 8527 M tonnes exported, followed by pineapple (7846 M) and broccoli (6985 M), while orito shows the lowest relative trade balance. In value terms, orito (\$8633 M) and soluble coffee (\$6592 M) are the products with the largest economic contribution to exports.
FAOSTAT Foreign Trade by Partner Country	Allows the analysis of Ecuador's agricultural foreign trade by partner country through filters by Product, Year, and Element (export or import, in quantity or value), including a geographic representation of the main destinations.	With the filters Product = All, Year = 2022, and Element = Exports – quantity applied; total exports amounted to 7380.50 K tonnes, with a trade balance of 5715,652 tonnes and 3760,512 thousand USD. The Russian Federation is the leading trading partner with 1603.74 K tonnes, followed by the United States (1408.03 K) and the Netherlands (527.52 K), evidencing a high concentration of agricultural trade in a few destinations. Banana remains the dominant product in the export ranking, far exceeding cocoa beans (391.56 K) as well as coffee extracts and cocoa derivatives, confirming its role as the country's leading agricultural export commodity.

4.2. Comparative Analysis of Performance and Costs Among Big Data Platforms

To assess SQL query performance on the implemented dimensional model, the fact_production_agricola table was used as the baseline dataset. This table comprises

19,345 observations obtained from the ESPAC/INEC national agricultural survey for the 2014–2024 period.

To evaluate platform performance under a larger data volume, a synthetic dataset was generated from the original records by performing a cross join with the month dimension (`dim_mes_clean`). Through this procedure, each annual observation was replicated once for every month of the year, thereby increasing the number of records while preserving the underlying structure and distribution of the original data. To maintain consistency between the annual and monthly levels of aggregation, the numerical variables used to calculate the key performance indicators were proportionally adjusted by dividing their annual values by 12. For instance, annual production expressed in tonnes per year was converted into estimated monthly production expressed in tonnes per month. This data expansion strategy enabled the execution of comparable SQL workloads across the evaluated platforms under a more computationally demanding scenario, while maintaining a direct relationship with the original agricultural dataset.

Performance tests were run on both data scales in Databricks Free Edition (Serverless Starter Warehouse) and Azure Databricks (Single Node + Photon). Five types of SQL queries representative of analytical workloads were evaluated: `SELECT`, `COUNT()`, `WHERE`, `GROUP BY`, and `JOIN` between the fact table and its dimensions. Each query was executed ten times per platform and scale, and the mean and standard deviation of the execution times were calculated using Python version 3.12.3.

4.2.1. SQL Query Performance

This section summarizes the execution times obtained for each query type, platform, and data scale. The results are expressed as the mean $\pm$ standard deviation (SD), calculated from ten repetitions per experimental scenario.

The results in Table 6 show that Azure Databricks outperformed Databricks Free Edition across all query types evaluated. This difference is most evident at the 232,140-row scale, where the workload is more representative of real-world analytical scenarios and initialization effects have less of an impact on execution times.

Azure Databricks recorded the shortest times for all queries. The most notable improvements were observed in `WHERE`, `GROUP BY`, and `JOIN` operations, with reductions of approximately 29%, 20%, and 26%, respectively, compared to Databricks Free Edition. These results demonstrate greater efficiency in processing analytical queries on the implemented dimensional model.

In terms of stability, Azure Databricks maintained low standard deviations across all tests, while Databricks Free Edition exhibited greater variability, especially at smaller scales. Furthermore, the increase from 19,345 to 232,140 records did not result in proportional increases in response times, indicating adequate scalability on both platforms, though with a consistent advantage for Azure Databricks in terms of performance and stability.

Table 6. SQL query execution times: Mean $\pm$ SD (s), n = 10 repetitions per test.

Test Type	Databricks Free Edition (Serverless Starter Warehouse)	Databricks Free Edition (Serverless Starter Warehouse)	Azure Databricks (Single Node + Photon)	Azure Databricks (Single Node + Photon)
	19,345 Records	232,140 Records	19,345 Records	232,140 Records
Query	Mean $\pm$ SD (s)	Mean $\pm$ SD (s)	Mean $\pm$ SD (s)	Mean $\pm$ SD (s)
SELECT	4.16 $\pm$ 9.59	0.67 $\pm$ 0.10	0.61 $\pm$ 0.27	0.47 $\pm$ 0.08
COUNT	1.18 $\pm$ 0.34	0.75 $\pm$ 0.15	0.65 $\pm$ 0.09	0.60 $\pm$ 0.05
WHERE	0.89 $\pm$ 0.18	0.65 $\pm$ 0.15	0.46 $\pm$ 0.11	0.46 $\pm$ 0.13
GROUP BY	1.47 $\pm$ 2.17	0.74 $\pm$ 0.16	0.61 $\pm$ 0.06	0.59 $\pm$ 0.03
JOIN	1.67 $\pm$ 1.06	1.02 $\pm$ 0.12	0.81 $\pm$ 0.07	0.76 $\pm$ 0.05

4.2.2. Dashboard Interaction Performance

Response times for three dashboard interaction scenarios measured in Databricks Free Edition (DFE) and Tableau Public (TP). These tests evaluate the visualization layer independently from the SQL processing layer, capturing end-to-end latency experienced by users during exploratory analysis.

As shown in Table 7, DFE provides faster response times than TP for all three interaction scenarios, with the most significant differences in filter application (1.08 s vs. 2.41 s) and click interaction (0.68 s vs. 2.44 s). The initial dashboard load time is comparable between platforms (4.09 s vs. 4.80 s). The higher latency observed in Tableau Public is partially attributable to network overhead, server-side rendering of public dashboards, and the lack of direct live connections to the data source in the free tier, which are factors expected in its public-sharing deployment context.

Table 7. Dashboard interaction response times (single measurement per platform).

Dashboard Test	Databricks Free Edition	Tableau Public
Initial dashboard load	4.09 s	4.80 s
Applying filters	1.08 s	2.41 s
Interaction (click)	0.68 s	2.44 s

Although Databricks demonstrated superior performance in dashboard rendering, Tableau Public was selected as the final dissemination platform because it provides a more suitable environment for public access to analytical results and facilitates access for all stakeholders within the agricultural sector. Its cloud-based architecture enables rapid dashboard publication, supports interactive visualizations without requiring user authentication, and incurs no additional licensing or hosting costs. Consequently, the platform was chosen by prioritizing accessibility, ease of deployment, and free dissemination of agricultural information over the higher visualization performance observed in the evaluated Big Data platforms.

Cost Analysis

From an economic perspective, Databricks Free Edition involves no licensing or infrastructure costs, making it a viable option for learning, exploratory testing, and initial validation of analytical models. However, Databricks also offers pay-as-you-go plans designed for production environments, which operate under a pay-as-you-go billing model based on Databricks Units (DBU). As a reference, the cost is estimated to be approximately between USD 0.15 and USD 0.70 per DBU/hour, depending on the plan and service configuration [34].

On the other hand, Azure Databricks operates under a pay-as-you-go model, also based on the consumption of Databricks Units (DBU) and cloud infrastructure resources. As a reference, the cost of the service is estimated at approximately USD 0.40 per DBU/hour, to which are added Microsoft Azure infrastructure costs, such as virtual machines, storage, and networking, with approximate values ranging from USD 0.25 to USD 0.40 per hour, depending on the cluster configuration. Azure Databricks offers advanced features such as the Photon engine, process automation, and scalability, making it more suitable for production environments and large-scale data analysis [30].

Tableau offers a free version, such as Tableau Public, which involves no costs but has limitations in terms of privacy and features. For more advanced uses, it offers a per-user subscription model. As a reference, costs for the Standard edition range from approximately USD 15 to USD 75 per user per month for the Viewer, Explorer, and Creator roles, respectively, which allow for everything from report visualization to the creation and full management of dashboards [35].

4.3. Validation of the Tableau Public Dashboard Through a User Survey

The evaluation of the Big Data-based analytics dashboard published on Tableau Public for Ecuador's agricultural sector was conducted using a survey designed to measure perceptions of usability and user experience.

The metrics evaluated aligned with UI/UX design best practices for analytical dashboards and included ease of learning, navigability, organization, visual clarity, visual design, ease of task execution, perceived performance, data reliability, and recommendation of the dashboard [36]. Table 8 presents the questionnaire design and the evaluation metrics used in the survey applied to users of the analytical dashboard. A five-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree), widely used in perception assessment, was employed for the measurement, allowing user ratings to be transformed into quantifiable data for analysis [37].

The survey's target audience consisted of users in the agricultural sector involved in production, marketing, and the analysis of agricultural information. A diverse range of profiles was considered, including producers, technicians, and managers who use data for monitoring, evaluation, and decision-making support in production contexts.

The sample comprised 31 participants selected through a non-probability purposive sampling strategy based on their involvement in agricultural production, technical management, or the use of institutional agricultural data in Ecuador. Given the exploratory and applied nature of the study, this approach facilitated the inclusion of stakeholders with relevant expertise and practical experience. The evaluation results are reported using descriptive statistics, including acceptance rates and Likert-scale response distributions, which are appropriate for assessing the perceived usability and acceptance of the proposed Big Data and Business Intelligence (BI) platform. Reliability and inferential statistical analyses were not conducted. The objective of the evaluation was not to test hypotheses or generalize findings to a broader population but rather to obtain qualitative and quantitative feedback from domain-relevant users regarding the platform's functionality, clarity, and practical usefulness.

Table 8. Questionnaire design based on UI/UX metrics for analytical dashboards.

Metric	ID	Question
Ease of learning	P1	The dashboard is easy to understand from the very first time.
Navigability	P2	Scrolling and navigating between the different sections of the dashboard are intuitive.
Information organization	P3	The information is structured in a clear and orderly manner.
	P4	The information presented is useful for analyzing agricultural indicators and supporting decision-making.
Visual clarity	P5	The visualization of graphs and tables facilitates understanding of the presented data.
Visual design	P6	The use of colors, typography, and graphic elements contributes to an appropriate visual experience.
Ease of task execution	P7	The dashboard allows you to apply filters or explore information easily.
Performance	P8	The dashboard provides adequate response times during user interaction.
Data reliability	P9	The data displayed on the dashboard inspires confidence regarding its accuracy and timeliness.
Recommendation	P10	I would consider recommending the use of this dashboard for agricultural and agro-export analysis.

Feedback from these users enabled an assessment of the usefulness, accessibility, and clarity of the Big Data-based analytical dashboard, based on the sector's actual needs and its contribution to more efficient, data-driven agricultural management.

Survey Results

This section presents the results obtained from the survey administered to a non-random sample of 31 stakeholders in the Ecuadorian agricultural sector who interacted with the Big Data-based analytical dashboard.

For each item, the relative frequencies of responses, the rejection rates for scales 1, 2, and 3, and the acceptance rates for scales 4 and 5 were analyzed. Additionally, the average score per question was calculated based on the weighted total of responses and the number of participants ($n = 31$).

The results show high overall acceptance of Dashboard in terms of usability, clarity of information, and support for decision-making, reflecting a positive user experience and a design aligned with UI/UX criteria.

Figure 6 shows high acceptance of Dashboard, with values close to or equal to 90% for most of the evaluated items, indicating a favorable user experience. Usability and clarity criteria show acceptance levels between 80% and 90%, while the recommendation to use Dashboard reaches 100% acceptance, confirming its usefulness for agricultural analysis. The lowest value corresponds to performance (P8) at 77.4%, possibly influenced by external factors such as the user's internet connection and the performance of the Tableau Public server.

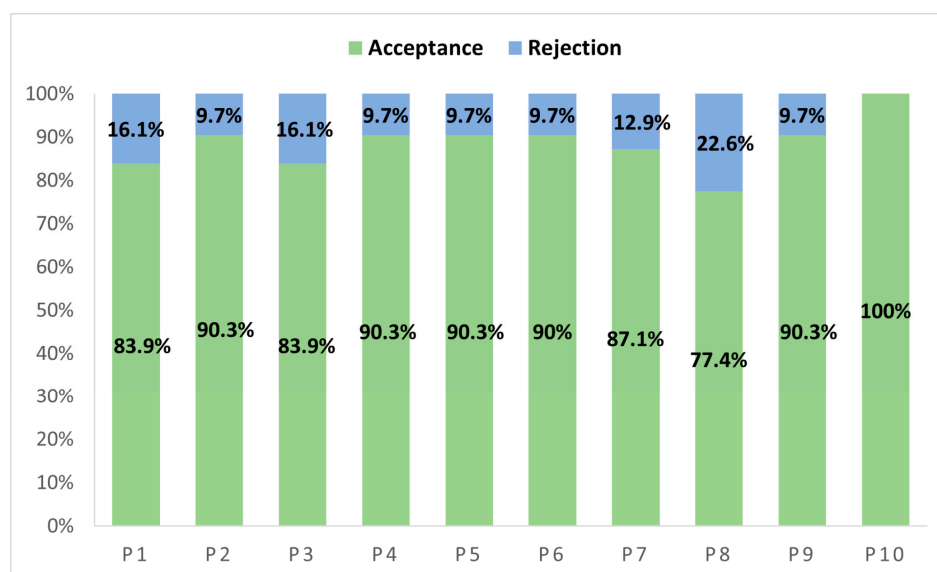


Figure 6. User perceptions of the Ecuadorian Agricultural Dashboard.

Figure 7 shows high UI/UX evaluation averages ($\geq 4.06/5$). Recommendation (4.94) and Data Trust (4.45) stand out as the highest-rated metrics. Ease of Task Execution and Visual Design score 4.39, followed by Visual Clarity (4.35), Navigability (4.32), and Information Organization (4.29). The lowest score corresponds to Performance (4.06), suggesting potential technical improvements.

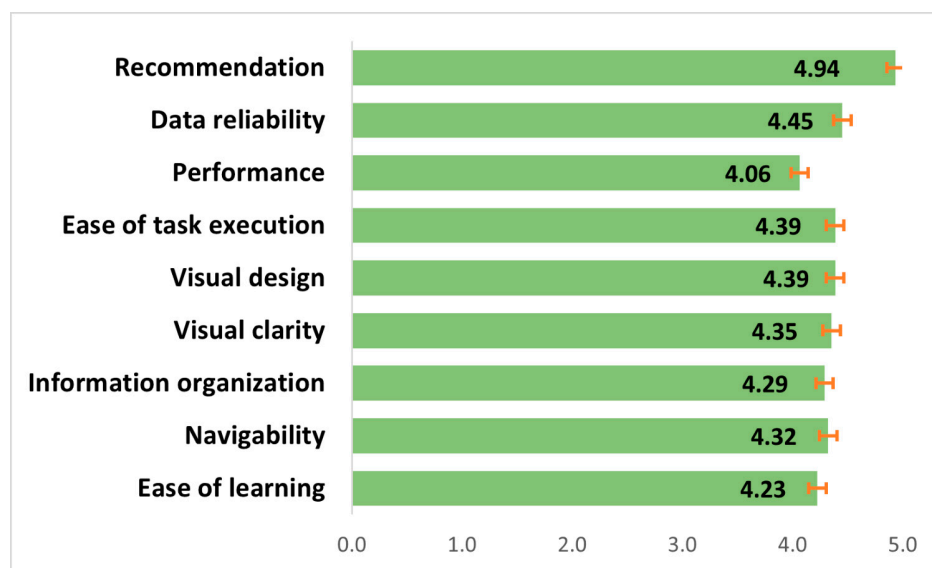


Figure 7. Average Likert-scale score by evaluated metric.

5. Discussion

Analysis of the Results Obtained

The results show that the implementation of a system based on Big Data and interactive dashboards allows for overcoming the limitations of data dispersion and heterogeneity in the agricultural sector. Compared to previous studies, such as Lokhande and Cravero [16,19], which highlight the use of advanced analytics to optimize the sector's analysis and production, this work broadens the approach by integrating multiple open-data sources under a Medallion architecture, covering everything from ingestion to interactive visualization.

Likewise, studies such as that by Sanabia-Lizarraga [23] demonstrate the value of business intelligence tools for improving data interpretation. However, unlike these approaches, this study incorporates a scalable architecture in Databricks, which allows for the management of larger volumes of information and ensures control over processing.

It is important to note that Azure Databricks stands out for its efficiency in large-scale data processing, while Tableau excels in usability and interactivity. The integration of both tools, together with the Medallion architecture and the TDSP methodology, enables the structuring of a complete flow of ingestion, transformation, modeling, and visualization, ensuring data quality, consistency, and traceability.

To compare this proposal with existing work and platforms, various agricultural visualization solutions were analyzed based on criteria such as visualization tools, data architecture, geographic coverage, number of integrated sources, range of crops, UI/UX design, interactivity, navigability by user profile, scalability in Big Data environments, and public availability. In general, existing proposals have limitations in terms of source integration, analytical coverage, and dynamic exploration capabilities, with a predominance of static approaches or those focused on specific indicators.

In contrast, this proposal integrates multiple national and international sources (INEC, ESPAC, Ecuador Open Data, and FAOSTAT), covers all registered crops, and offers interactive dashboards designed for different user profiles, providing a more comprehensive, scalable, and accessible solution.

Table 9 presents a comparative analysis of the results obtained in this study and those of previous studies, allowing for an assessment of the progress made and the main limitations identified in the various studies.

Table 9. Comparative analysis of existing agricultural dashboards.

Authors	Evaluation Criteria					
	Tools for Visualization	Data Architecture	Coverage/Country	Number of Data Sources	Number of Crops	Availability
Sanabia-Lizarraga [23]	Power BI	CRISP-DM/Power BI data model	Mexico	2–3 institutional	Subsectors of agricultural exporters MX	Yes (link public)
INEC [5]	Tables/PDF	Not specified	Ecuador	1 (ESPAC)	52 agricultural	Yes (INEC Portal)
SIPA-MAG [6,10]	Tables/Web Portal	Not specified	Ecuador	2–3 (MAG, BCE)	Selected crops	Yes (SIPA Portal)
Lokhande [16]	Not applicable	Big Data + Data Mining	India	1 (agricultural legacy)	Precision generic agricultural	No
Proposal of this study	Tableau Public	Medallion architecture for Big Data analysis and visualization in the cloud with Databricks.	Ecuador (National and International)	4 (INEC, ESPAC, Open Data Ecuador, FAOSTAT)	All crops (permanent and temporary in Ecuador)	Yes (public)

6. Conclusions

This study proposed and validated a scalable Big Data Analytics framework for the integration, processing, and visualization of heterogeneous agricultural datasets in Ecuador through low-cost digital technologies. The framework combined the Team Data Science Process (TDSP), a Medallion-based architecture implemented in Databricks, and in-

teractive dashboards developed in Tableau Public to support agricultural decision-making.

The results show that the proposed analytical environment successfully integrated multiple open-data sources, including INEC-ESPAC, MAG-SIPA, Ecuador Open Data, and FAOSTAT. This integration enabled multidimensional analysis of agricultural production, trade, producer prices, agricultural losses, and producer profiles within a unified analytical ecosystem.

From a technical perspective, the benchmarking analysis confirmed the advantages of optimized cloud-native environments for analytical workloads. Azure Databricks achieved lower execution times than Databricks Free Edition, particularly in aggregation and relational operations, demonstrating the scalability potential of distributed Big Data architectures. In contrast, Tableau Public provided an accessible and user-friendly environment for the dissemination and interactive visualization of analytical results.

The comparative analysis of the evaluated platforms indicates that tool selection depends largely on the application context, performance requirements, and scalability needs. Databricks Free Edition proved suitable for academic settings, low-cost prototypes, and exploratory analytics. In contrast, Azure Databricks offered better performance for complex analytical workloads, making it more appropriate for enterprise environments requiring greater processing capacity, security, scalability, and integration capabilities. Similarly, Tableau Public represents a viable alternative for publicly sharing dashboards and analytical outputs. However, enterprise scenarios that demand high availability, advanced integrations, secure data governance, and large-scale analytical management require the adoption of commercial or licensed platforms such as Azure Databricks and Tableau.

The usability validation of the Tableau dashboard, conducted with 31 agricultural stakeholders, demonstrated the practical feasibility of the proposed Big Data Analytics framework. This was reflected in high user acceptance, a 100% recommendation rate, and a data trust score of 4.45/5. These findings highlight the potential of scalable and low-cost Big Data technologies to strengthen evidence-based agricultural decision-making in resource-constrained contexts.

The main contribution of this study lies in the integration of open agricultural data, scalable cloud-based processing, dimensional modeling, interactive analytics, and usability validation into a reproducible and low-cost framework tailored to developing-country contexts. Unlike fragmented or static analytical solutions reported in previous studies, the proposed approach provides an end-to-end analytical ecosystem capable of supporting agricultural digital transformation initiatives.

Despite the positive results, several limitations should be acknowledged. First, the datasets used in this study cover the period 2010–2024 and were selected as part of a pilot initiative aimed at assessing the feasibility of integrating heterogeneous agricultural data sources in Ecuador and evaluating cloud-based Big Data platforms for agricultural analytics. This period was determined by the availability of data from the consulted institutional sources. A key finding of this exploratory effort was that each data source follows its own organizational criteria, formats, and publication standards, depending on the institution responsible for data dissemination. This structural heterogeneity required data reprocessing during the construction of the Silver layer and represents a significant challenge for scaling the Big Data and Business Intelligence (BI) platform to support additional historical and future periods. However, addressing these integration and standardization issues constitutes an important future research direction for expanding the platform's temporal coverage.

Second, the contribution of this study is methodological, analytical, and applied rather than algorithmically innovative. Established methodologies, including TDSP, Medal-

lion architecture, and dimensional modeling, were intentionally integrated to address a specific gap in Ecuador's agricultural information ecosystem: the absence of a unified platform capable of correlating data from multiple institutions through shared dimensions such as year and crop type, given that each institution publishes data within its own analytical domain.

Third, usability validation was conducted using descriptive statistics to assess the accessibility and ease of use of the proposed framework, which was consistent with the scope of the study. Reliability analyses and inferential statistical methods were not incorporated and may be considered in future research. Future research may also focus on the automation of ETL processes, the integration of artificial intelligence models, including generative AI for natural language querying, predictive analytics, and materialized views, to further enhance the platform's performance, scalability, analytical capabilities, and accessibility in resource-constrained institutional environments.

Overall, the findings demonstrate that low-cost Big Data technologies and interactive analytics platforms can significantly contribute to the digital transformation of agriculture and the graphical visualization of data with limited resources. The proposed framework successfully addressed a specific gap in Ecuador's agricultural information ecosystem by integrating fragmented institutional data sources into a unified, scalable, and publicly accessible analytics platform. The results confirm that methodological rigor, open-source tools, and structured data architectures can, together, deliver significant analytical value even in contexts lacking enterprise-level infrastructure, providing a reproducible model applicable to similar agricultural environments in developing countries.

Author Contributions: Conceptualization, A.A.-S., J.Á.-V., B.M.-O., and M.P.; Methodology, A.A.-S., J.Á.-V., and B.M.-O.; Software, J.Á.-V.; Validation, B.M.-O., and M.P.; Investigation, A.A.-S. and J.Á.-V.; Resources, B.M.-O. and M.P.; Data curation, A.A.-S. and J.Á.-V.; Writing—original draft, A.A.-S.; Writing—review and editing, B.M.-O.; Supervision, B.M.-O. and M.P.; Visualization, A.A.-S. and B.M.-O.; Project administration, B.M.-O.; Funding acquisition, B.M.-O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Universidad Técnica de Machala, AutoMathTIC Research Group, Research Project with Resolution No. 0244-2024-CU-SO-13: “Adopción de las Tecnologías de Inteligencia Artificial e Internet de las Cosas en el Sector Agropecuario de la Provincia de El Oro”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: A written informed consent form was not required. However, the survey form designed in Microsoft Forms to evaluate the UI/UX of the proposed dashboard for the Ecuadorian agricultural sector, developed using Big Data and Business Intelligence and published in Tableau, included a specific question requesting participants' informed consent prior to completing the survey. Participation was voluntary, and no personal data or information that could identify the respondents was collected. Only evaluations, perceptions, and comments related to the dashboard's usability, user experience, and usefulness were gathered.

Data Availability Statement: The original data used in this study are publicly available from official open-data sources, including INEC, ESPAC, Ecuador Open Data, and FAOSTAT. These datasets correspond to the Ecuadorian agricultural sector for the 2010–2024 period and can be accessed through the respective official institutional websites. The integrated and preprocessed datasets generated for the purposes of this study, including cleaned and transformed data used for analysis and visualization, are available from the corresponding author upon reasonable request. The original data presented in the study are openly available at https://github.com/Jean66534/data_repository_ec_agro (accessed on 22 June 2026).

Acknowledgments: We thank the Universidad Técnica de Machala for funding this research work. The authors acknowledge the use of artificial intelligence tools during the preparation of this manuscript. ChatGPT (OpenAI- GPT-5.5) was used to support SQL script refinement, including SQL syntax consultation, error correction, code indentation, and refactoring guidance. Google Gemini 3.5 was used to provide guidance on the design of the graphical abstract. Claude Sonnet 4.6 (Anthropic) was used to improve writing clarity and correct grammatical errors. All AI-generated suggestions were critically reviewed, verified, and validated by the authors, who assume full responsibility for the final content of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

TDSP	Team Data Science Process
KPI	Key Performance Indicator
SQL	Structured Query Language
UI/UX	User Interface/User Experience
GDP	Gross Domestic Product
INEC	National Institute of Statistics and Censuses
MAG	Ministerio de Agricultura y Ganadería del Ecuador
FAO	Food and Agriculture Organization
FAOSTAT	Food and Agriculture Organization Statistical Database
ESPAC	Encuesta de Superficie y Producción Agropecuaria Continua del Ecuador
SIPA	Sistema de Información Pública Agropecuaria del Ecuador
DFE	Databricks Free Edition
TP	Tableau Public

References

1. FAO. *Gross Domestic Product and Agriculture Value Added 2014–2023*; FAO: Rome, Italy, 2025. <https://doi.org/10.4060/cd5490en>.
2. Perego, V.M.E.; Morris, M.; Sebastian, A.R. *Future Foodscapes: Re-Imagining Agriculture in Latin America and the Caribbean*; World Bank Publications: Washington, DC, USA, 2020. <https://doi.org/10.1596/34812>.
3. NU. CEPAL Preliminary Overview of the Economies of Latin America and the Caribbean, 2024. Available online: <https://www.cepal.org/en/publications/81105-preliminary-overview-economies-latin-america-and-caribbean-2024> (accessed on 8 October 2025).
4. Pacheco, J.; Ochoa-Moreno, W.-S.; Ordoñez, J.; Izquierdo-Montoya, L. Agricultural Diversification and Economic Growth in Ecuador. *Sustainability* **2018**, *10*, 2257. <https://doi.org/10.3390/su10072257>.
5. INEC Agricultural Statistics. Available online: <https://www.ecuadorencifras.gob.ec/estadisticas-agropecuarias-2/> (accessed on 8 October 2025).
6. Public Agricultural Information System of Ecuador—SIPA. Available online: <https://sipa.agricultura.gob.ec/index.php> (accessed on 16 April 2026).
7. Ministry of Production, Foreign Trade and Investments—Ecuador. Available online: <https://www.produccion.gob.ec/boletines-de-comercio-exterior-2025/> (accessed on 13 December 2025).
8. Escalante Pineda, M.E.; Urbina Bustos, S.S.; Banderas Benítez, V.E.; Farinango Salazar, R.A.; Sotomayor Cabrera, K.K. Analysis of the productive structure of the Ecuadorian economy: Exports from the Agricultural Sector. *Soc. Technol.* **2021**, *4*, 380–398. <https://doi.org/10.51247/st.v4i3.144>.
9. Economic Information. Available online: <https://contenido.bce.fin.ec/documentos/informacioneconomica/PublicacionesGenerales/IndiceEEE.htm> (accessed on 6 May 2026).
10. MAG Agricultural Production Figures. Available online: <https://sipa.agricultura.gob.ec/index.php/sipa-estadisticas/tablerodinamico/cifras-agroproductivas> (accessed on 13 December 2025).
11. Microsoft Team Data Science Process (TDSP): Lifecycle and Templates. Available online: <https://github.com/Azure/Microsoft-TDSP> (accessed on 14 October 2025).
12. Foroughi, F.; Luksch, P. Data Science Methodology for Cybersecurity Projects. *arXiv* **2018**, arXiv:1803.04219.

13. Schröer, C.; Kruse, F.; Gómez, J.M. A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Comput. Sci.* **2021**, *181*, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>.
14. Peña Ramírez, C.; Espinoza, L.; Forcael, E.; Garcés, G.; Cea, C. Leveraging Business Intelligence for Strategic Growth: A Case Study of Small and Medium-Sized Enterprises (SMEs) in the Construction Industry. *Ingeniare Rev. Chil. Ing.* **2024**, *32*, 39. <https://doi.org/10.4067/S0718-33052024000100240>.
15. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.-J. Big Data in Smart Farming—A Review. *Agric. Syst.* **2017**, *153*, 69–80. <https://doi.org/10.1016/j.agsy.2017.01.023>.
16. Lokhande, S.A. Effective Use of Big Data in Precision Agriculture. In Proceedings of the 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 5–7 March 2021; pp. 312–316.
17. Chen, Y.; Li, C.; Wang, H. Big Data and Predictive Analytics for Business Intelligence: A Bibliographic Study (2000–2021). *Forecasting* **2022**, *4*, 767–786. <https://doi.org/10.3390/forecast4040042>.
18. Reddy, G.T.; Reddy, M.P.K.; Lakshmana, K.; Kaluri, R.; Rajput, D.S.; Srivastava, G.; Baker, T. Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access* **2020**, *8*, 54776–54788. <https://doi.org/10.1109/ACCESS.2020.2980942>.
19. Cravero, A.; Pardo, S.; Galeas, P.; López Fenner, J.; Caniupán, M. Data Type and Data Sources for Agricultural Big Data and Machine Learning. *Sustainability* **2022**, *14*, 16131. <https://doi.org/10.3390/su142316131>.
20. Chamorro-Padial, J.; Virgili-Gomá, J.; Gil, R.; Teixidó, M.; García, R. Agriculture Data Sharing Review. *Heliyon* **2025**, *11*, e41109. <https://doi.org/10.1016/j.heliyon.2024.e41109>.
21. Zaharia, M.; Ghodsi, A.; Xin, R.; Armbrust, M. Lakehouse: A New Generation of Open Platforms That Unify Data Warehousing and Advanced Analytics. In Proceedings of the 11th Annual Conference on Innovative Data Systems Research (CIDR '21), Online, 11–15 January 2021.
22. Mazon-Olivo, B.; Tinoco Egas, R.M. Chapter 7: Business Intelligence in the Agricultural Sector. In *Agricultural Data Analysis*; Ediciones UTMACH, Universidad Técnica de Machala: Machala, Ecuador, 2018; pp. 246–278, ISBN 978-9942-24-120-7.
23. Sanabria-Lizarraga, K.G.; Carballo-Mendivil, B.; Arellano-González, A.; Bueno-Solano, A. Business Intelligence for Agricultural Foreign Trade: Design and Application of Power BI Dashboard. *Sustainability* **2024**, *16*, 9576. <https://doi.org/10.3390/su16219576>.
24. Boza, S.Z.; Herrera, X.E.; Reyes, P.S.A.; Silva, A.M. Agricultural innovation systems: A look at the situation of the Ecuadorian agricultural sector. *Rev. Científica Ecociencia* **2021**, *8*, 237–254. <https://doi.org/10.21855/ecociencia.80.647>.
25. Krisnawijaya, N.K.; Tekinerdogan, B.; Catal, C.; van der Tol, R. Data Analytics Platforms for Agricultural Systems: A Systematic Literature Review. *Comput. Electron. Agric.* **2022**, *195*, 106813. <https://doi.org/10.1016/j.compag.2022.106813>.
26. Chergui, N.; Kechadi, M.T. Data Analytics for Crop Management: A Big Data View. *J. Big Data* **2022**, *9*, 123. <https://doi.org/10.1186/s40537-022-00668-2>.
27. What Is a Data Lakehouse? Available online: <https://www.databricks.com/blog/what-is-data-lakehouse> (accessed on 12 June 2026).
28. Ouafiq, E.M.; Saadane, R.; Chehri, A.; Wahbi, M. Data Lake Conception for Smart Farming: A Data Migration Strategy for Big Data Analytics. In *Proceedings of the Human Centred Intelligent Systems*; Zimmermann, A., Howlett, R.J., Jain, L.C., Eds.; Springer Nature: Singapore, 2022; pp. 191–201.
29. Databricks Free Edition Limitations | Databricks on AWS. Available online: <https://docs.databricks.com/aws/en/getting-started/free-edition-limitations> (accessed on 15 April 2026).
30. Azure Databricks Pricing | Microsoft Azure. Available online: <https://azure.microsoft.com/en-us/pricing/details/databricks/> (accessed on 15 April 2026).
31. Tableau Desktop. Available online: <https://www.tableau.com/es-es/products/techspecs> (accessed on 15 April 2026).
32. Tableau Public FAQ. Available online: https://help.tableau.com/current/pro/desktop/en-us/public_faq.htm (accessed on 15 April 2026).
33. FAOSTAT. Available online: <https://www.fao.org/faostat/en/#home> (accessed on 13 December 2025).
34. Databricks Pricing: Flexible Plans for Data and AI Solutions. Available online: <https://www.databricks.com/product/pricing> (accessed on 22 January 2026).
35. Buy Tableau. Available online: <https://buy.tableau.com> (accessed on 15 April 2026).

36. Perales-Domínguez, C.; Sánchez-Calle, J.E.; Lévano-Rodríguez, D.; Gallegos-Carrillo, K. Methodologies for the construction of business intelligence solutions. *Rev. Científica Sist. Informática* **2024**, *4*, e612. <https://doi.org/10.51252/rcsi.v4i1.612>.
37. Kusmaryono, I.; Wijayanti, D.; Maharani, H.R. Number of Response Options, Reliability, Validity, and Potential Bias in the Use of the Likert Scale Education and Social Science Research: A Literature Review. *Int. J. Educ. Methodol.* **2022**, *8*, 625–637. <https://doi.org/10.12973/ijem.8.4.625>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.