

Transforming Grocery Inventory Planning and Decision Intelligence Using Data + AI on a Databricks Lakehouse Architecture.

(<https://grocerydataintelligence.com>)

Brahma Reddy Katam

Technical Lead, Data Engineering and Advanced Computing

Abstract : Grocery stores face a continuous operational challenge in balancing product availability with waste reduction. Fast-moving products may go out of stock before store teams notice, leading to missed sales and poor customer experience. At the same time, fresh products may remain on shelves too long and expire before they are sold, creating food waste and financial loss. Traditional dashboards help visualize inventory and sales trends, but they often require business users to interpret charts manually and decide the next action.

This research presents Grocery Data Intelligence, a Data + AI-driven grocery inventory planning system built on a modern data engineering lakehouse architecture. The proposed system converts raw grocery data into trusted business metrics and action recommendations using a layered Bronze, Silver, Gold, and semantic data design. The application enables users to analyze stockout risk, waste risk, days of supply, reorder needs, promotion opportunities, and donation possibilities through dashboards, natural language questions, inventory storytelling, and action workflows.

The system was designed using Databricks Free Edition as an experimental platform for building a practical Data + AI solution. The architecture demonstrates how clean data engineering foundations can support business-ready insights, natural language analytics, and faster decision-making. Unlike traditional reporting systems, Grocery Data Intelligence moves beyond static dashboards by helping store users understand what is happening, why it matters, and what action should be taken.

This paper explains the problem background, architecture, implementation approach, methodology, expected outcomes, and future enhancements of the

proposed system. The findings show that combining data engineering, trusted metrics, semantic definitions, and AI-powered interaction can help grocery businesses reduce stockouts, reduce waste, improve planning, and move from insight to action faster.

Keywords

Data Engineering, Grocery Analytics, Lakehouse Architecture, Databricks, Data + AI, Inventory Optimization, Stockout Risk, Waste Risk, Semantic Layer, Natural Language Analytics, Decision Intelligence

1. Introduction

Grocery stores operate in a fast-moving environment where customer demand, product freshness, supplier availability, and inventory levels change continuously. A store must ensure that essential products are available when customers need them while also preventing fresh items from expiring on the shelf. This creates a difficult balancing problem for store managers and inventory teams.

When products sell faster than expected, the store may experience stockouts. Stockouts reduce revenue and may also impact customer trust. Customers who repeatedly fail to find the products they need may shift to another store. On the other side, overstocking fresh products can create waste when items expire before being sold. Food waste not only affects store profitability but also creates social and environmental concerns.

Many grocery businesses use dashboards and reports to monitor sales and inventory. These reports are useful, but they often stop at visibility. Business users still need to review charts, compare numbers, identify patterns, and decide what to do next. This creates a delay between insight and action. In daily store operations, even a small delay can result in missed sales, expired inventory, and poor planning.

Recent advancements in data engineering, lakehouse architecture, semantic layers, and AI-powered natural language interaction provide an opportunity to rethink grocery inventory planning. Instead of only showing what happened, a modern system can help users understand

what is happening now, why it matters, and what action should be taken. This research introduces Grocery Data Intelligence, a practical Data + AI application designed to help grocery stores turn raw inventory and sales data into faster business actions.

The proposed system uses a simple but powerful data flow. Raw grocery data is stored in a Bronze layer. Cleaned and standardized data is prepared in a Silver layer. Business-ready metrics are created in a Gold layer. A semantic layer defines consistent business logic for stockout risk, waste risk, days of supply, priority, and recommended actions. The application layer then uses these trusted metrics to power dashboards, natural language questions, inventory stories, and action workflows.

The goal of this paper is to demonstrate how Data + AI can improve grocery inventory planning by reducing the time between data, insight, and action.

2. Literature Review

Inventory planning and demand management have been widely studied across retail, supply chain, and operations research domains. Traditional inventory management systems focus on reorder points, safety stock, demand forecasting, and replenishment cycles. These methods are effective when demand patterns are stable and when data is accurate. However, grocery retail introduces additional complexity because many products are perishable, demand can fluctuate quickly, and shelf-level decisions must often be made daily.

Conventional retail analytics systems usually rely on historical sales, point-of-sale data, inventory snapshots, and manual reporting. Dashboards provide visibility into metrics such as total sales, product movement, current stock, and category performance. While these systems improve awareness, they are limited when business users need direct action recommendations. A dashboard may show that inventory is low, but it may not clearly explain whether the store should reorder immediately, monitor the item, discount it, promote it, or donate it.

Data engineering has become an important foundation for modern analytics and AI systems. Lakehouse architecture provides a unified approach where raw data, processed data, analytics, and AI workloads can operate on a single data platform. This reduces data duplication and improves consistency across business use cases. A

layered data architecture also helps separate raw data storage, cleaning, business logic, and consumption layers.

AI and natural language analytics are now changing how business users interact with data. Instead of depending only on predefined dashboards, users can ask questions in plain English and receive contextual answers. This approach can reduce the time required to access insights, especially for non-technical users. However, AI-powered responses are only useful when they are grounded in trusted data. Without a clean and reliable semantic layer, natural language answers may become inconsistent or difficult to trust.

In retail inventory planning, the combination of data engineering and AI can be highly valuable. Clean data pipelines can create reliable metrics such as average daily sales, days of supply, stockout risk, and waste risk. AI-powered interfaces can help users ask questions such as “What should I reorder today?” or “Which fresh products may go to waste?” Action workflows can then convert recommendations into tasks such as reorder plans, promotion plans, discount actions, or donation opportunities.

Existing research and industry practice show that data and AI are most effective when they work together. Data engineering creates trust. AI creates speed and accessibility. Action workflows create business value. This paper builds on this direction by presenting Grocery Data Intelligence as a practical example of decision intelligence for grocery inventory planning.

Use Case: Implementing Data + AI for Grocery Inventory Planning

The use case presented in this research focuses on a common grocery store problem: reducing stockouts and waste while improving daily inventory decisions. A grocery store handles multiple product categories, including fresh produce, packaged goods, dairy items, frozen products, and household essentials. Some products sell quickly and need replenishment. Some products move slowly and may create waste. Some products require monitoring based on days of supply or expiry risk.

Grocery Data Intelligence is designed to help store users move from reporting to action. The system ingests store, product, sales, and inventory data. It then prepares trusted business metrics that help identify product-level risks and recommended actions. For example, a product with low

current stock and high average daily sales may be marked as a reorder candidate. A fresh product with high stock and low sales speed may be marked as a waste-risk item. A product that is close to expiry may be recommended for discount, promotion, or donation.

The application includes dashboards for visibility, an Ask Data + AI experience for natural language questions, an inventory story page for business explanation, and an action flow for reviewing recommended next steps. This makes the solution useful for both technical and non-technical users. Store managers can understand what needs attention without writing SQL or waiting for custom reports.

The use case demonstrates how a focused Data + AI solution can improve operational decision-making in grocery retail. Instead of simply displaying data, the app helps users understand risk, review supporting metrics, and take practical actions.

3. Research Objectives

The primary objective of this research is to design and evaluate a Data + AI-based grocery inventory planning system that helps stores reduce stockouts, reduce waste, and improve action speed.

The first objective is to demonstrate how a lakehouse architecture can support grocery analytics by organizing data into Bronze, Silver, Gold, and semantic layers. This structure helps maintain data quality, traceability, and reusable business logic.

The second objective is to define trusted business metrics for grocery inventory planning. These include current stock, average daily sales, days of supply, stockout risk, waste risk, priority, and recommended action. These metrics form the foundation for dashboards, AI responses, and action workflows.

The third objective is to evaluate how natural language interaction can help business users access insights faster. Instead of requiring users to browse multiple dashboards, the system allows them to ask questions in plain English and receive simple answers based on trusted data.

The fourth objective is to show how recommendation logic can be converted into business actions. The system does not stop at identifying risk. It also suggests next steps such as reorder, monitor, promote, discount, or donate.

The final objective is to provide a practical reference architecture for building Data + AI applications that move beyond reporting and support decision intelligence.

4. Architecture and System Design

The architecture of Grocery Data Intelligence follows a layered data engineering approach designed for clarity, trust, and reusability. The system separates raw data storage, cleaning, business metrics, semantic definitions, application interaction, and action workflows.

The first layer is the raw grocery data layer. This includes stores, products, sales transactions, inventory levels, and product category information. For the demo implementation, sample grocery sales and inventory records were used to simulate real business scenarios.

The Bronze layer stores raw data in Databricks. This layer preserves the original data as received, allowing traceability and future reprocessing. The purpose of the Bronze layer is to maintain a complete record of incoming data before applying transformations.

The Silver layer cleans, standardizes, and joins the data. Product names, categories, sales dates, store identifiers, and inventory values are prepared for analysis. This layer removes inconsistencies and creates reliable records for downstream calculations.

The Gold layer creates business-ready inventory metrics. These include average daily sales, current stock, days of supply, stockout risk, waste risk, priority level, and recommended action. This layer acts as the main source for business consumption.

The semantic layer sits on top of the Gold layer and defines consistent business meanings. For example, “days of supply” should have the same definition across dashboard, Ask Data + AI, inventory story, and action flow. This avoids confusion and improves trust.

The application layer provides the user experience. It includes dashboard views, natural language question answering, action recommendations, inventory stories, and action review screens.

The action layer converts insight into business activity. Recommended actions include reorder, monitor, promote, discount, or donate. In future versions, an action backend such as Lakebase can store reorder plans, action status, user decisions, and feedback history.

The high-level architecture can be summarized as follows:

Raw Grocery Data

Stores, Products, Sales, Inventory

Bronze Layer

Raw data stored in Databricks

Silver Layer

Cleaned, standardized, and joined data

Gold Layer

Business-ready inventory metrics

Semantic Layer

Standard definitions for stockout risk, waste risk, days of supply, priority, and recommended actions

Grocery Data Intelligence App

Dashboard, Ask Data + AI, Inventory Story, and Action Flow

Lakebase-Ready Action Backend

Reorder plans, task status, user decisions, and feedback history

Business Action

Reorder, Monitor, Promote, Discount, or Donate

5. Implementation

The implementation of Grocery Data Intelligence follows an incremental approach. The first phase focuses on data preparation. Sample grocery data is generated and organized into tables representing stores, products, sales, and inventory. The demo includes multiple product conditions such as low stock, healthy inventory, waste risk, promotion opportunity, and donation opportunity.

The second phase focuses on data transformation. Raw data is cleaned and standardized. Product identifiers, store identifiers, sales quantities, inventory levels, and product categories are prepared for metric calculation. This ensures that the application does not rely on inconsistent or incomplete data.

The third phase creates Gold layer metrics. Average daily sales is calculated using sales history. Days of supply is calculated by comparing current stock with average daily demand. Stockout risk is identified when days of supply is low and sales speed is high. Waste risk is identified when stock is high, sales speed is low, or fresh products are near expiry. Recommended action is assigned based on the risk condition.

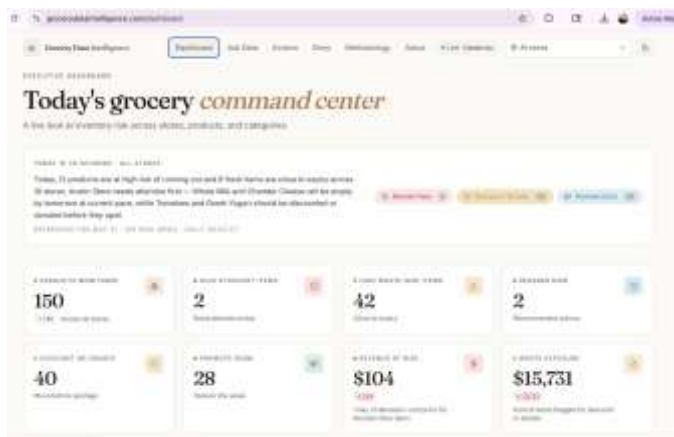
The fourth phase implements the application experience. The dashboard shows product-level metrics, risk indicators, and summary views. The Ask Data + AI experience enables users to ask inventory-related questions in natural language. The story page summarizes the current inventory condition in simple business language. The action flow allows users to review recommended next steps.

The fifth phase improves action clarity. Instead of using a generic "Take Action" button, the system can display specific actions such as Review Reorder Plan, Review Waste Risk Actions, Create Promotion Plan, Create Donation Plan, and Review Monitoring Items. This improves user confidence because the user understands what action is being taken.

Figure 1: High-Level Architecture of Grocery Data Intelligence



This architecture supports the central message of the system: dashboard gives visibility, semantic layer gives trust, action memory gives continuity, and Data + AI helps users move from insight to action faster.



A future implementation phase can integrate Lakebase or a similar transactional backend to store action plans, user confirmations, status updates, and feedback. This will allow the application to move from a demo analytics app to a more complete operational decision system.

6. Data Collection and Preparation

Data collection is a critical step in the Grocery Data Intelligence system because the quality of inventory recommendations depends on the accuracy and consistency of the underlying data. For the demo, sample data was created to represent common grocery business entities.

The main data entities include stores, products, sales, and inventory. Store data contains store-level identifiers and location information. Product data contains product names, categories, freshness indicators, and product attributes. Sales data contains transaction-level or aggregated product movement information. Inventory data contains current stock levels, expiry-related information, and available quantity.

The data preparation process begins with validation. Records are checked for missing product identifiers, invalid sales quantities, missing inventory values, and inconsistent category names. Duplicate records are identified and removed where required. Date fields are standardized to ensure consistent time-based calculations.

The next step is standardization. Product categories are normalized so that similar category labels do not create confusion in reporting. Sales and inventory values are converted into consistent numeric formats. Product and store identifiers are aligned across tables to support joins and aggregations.

Once cleaned and standardized, the data is prepared for metric generation. Average daily sales, days of supply, and product risk indicators are calculated from the prepared data. These metrics become the foundation of the Gold layer and semantic layer.

This approach highlights the importance of data engineering in AI-powered applications. AI can help users ask questions and receive recommendations, but reliable answers require clean, structured, and trusted data.

7. Methodology: Data + AI Algorithms

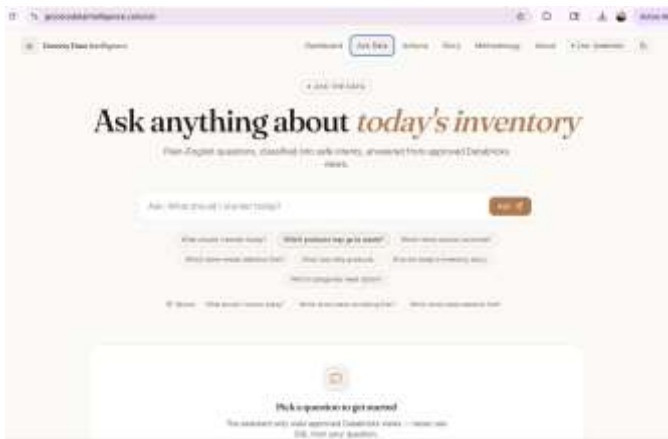
The methodology used in Grocery Data Intelligence combines rule-based business logic, data engineering transformations, semantic metric definitions, and natural language interaction.

The first part of the methodology is inventory metric calculation. Average daily sales is calculated from historical sales records. Current stock is obtained from inventory data. Days of supply is calculated by dividing current stock by average daily sales. This metric helps estimate how many days the product can remain available before stock runs out.

The second part is risk classification. A product is marked as stockout risk when its days of supply is low and sales movement is strong. A product is marked as waste risk when inventory is high, sales movement is weak, or freshness-related factors indicate possible expiry. Products with balanced stock and demand are marked as healthy.

The third part is action recommendation. Based on the product condition, the system recommends an action. Low-stock fast-moving products are recommended for reorder. Slow-moving fresh products may be recommended for promotion, discount, or donation. Products with uncertain behavior may be marked for monitoring.

The fourth part is semantic consistency. Each metric and recommendation is defined in a consistent way so that dashboard, Ask Data + AI, story page, and action flow all use the same business logic. This improves trust and reduces confusion.



The fifth part is natural language interaction. The Ask Data + AI experience allows users to ask questions in plain English. The system maps these questions to trusted inventory metrics and returns simple answers with supporting data. For example, when a user asks, “What should I reorder today?” the system identifies reorder candidates and provides the product list, priority, and suggested next action.

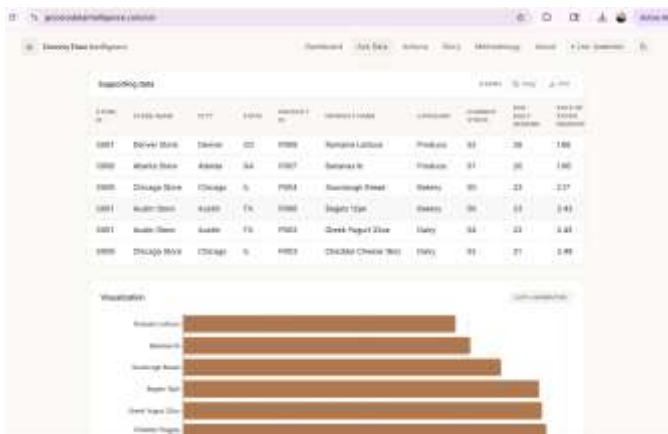
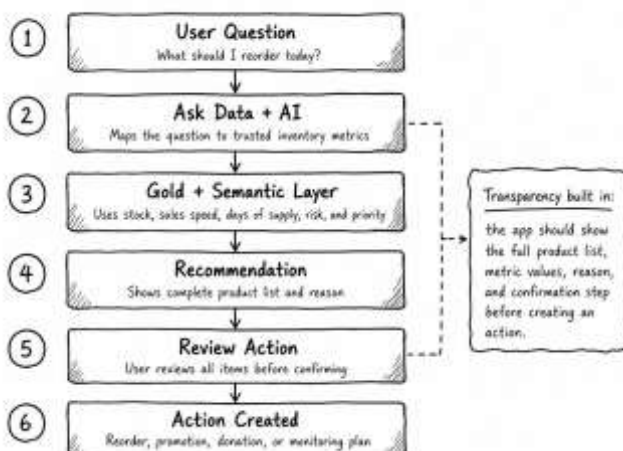


Figure 2: Data + AI Recommendation and Action Flow



The methodology demonstrates how practical AI does not always require complex models at the beginning. In many

business scenarios, clean data, trusted metrics, simple risk logic, and user-friendly AI interaction can produce meaningful value.

8. System Integration and Deployment

The Grocery Data Intelligence system is designed as a web-based application powered by Databricks-backed data. The integration strategy focuses on keeping the data layer trusted and the application layer simple.

At the data layer, Databricks is used to organize raw, cleaned, and business-ready data. The Bronze, Silver, and Gold approach makes the pipeline easy to understand and extend. Business metrics are prepared in the Gold layer and exposed to the application for reporting, storytelling, and question answering.

At the application layer, the web interface presents information through dashboards, natural language responses, and action review screens. The user does not need to understand the underlying tables or SQL logic. Instead, the app presents insights in business language.

The Ask Data + AI layer acts as a bridge between business questions and trusted metrics. The system is designed so that answers are grounded in the Gold and semantic layers instead of being generated randomly. This improves transparency and confidence.

The action layer is designed to support future operational workflows. In the current demo, actions can be represented as recommendations or mock action plans. In a production version, a transactional backend can store user-created actions, reorder tasks, donation plans, and feedback status.

This integration approach demonstrates how a Data + AI application can combine analytics, natural language interaction, and business actions in a single user experience.

9. Results and Expected Outcomes

The expected outcome of Grocery Data Intelligence is improved decision speed and better inventory awareness for grocery store teams. Instead of reviewing multiple reports, users can quickly identify which products need attention and why.

One expected result is reduced stockout risk. By calculating days of supply and average daily sales, the

The main limitation of the current system is that it is a demo built using sample data. Real-world deployment would require integration with point-of-sale systems, supplier systems, inventory systems, and store operations workflows. However, the architecture is designed to support such extensions.

Conclusion

This paper presented Grocery Data Intelligence, a Data + AI-driven grocery inventory planning system built on a lakehouse architecture. The system addresses a practical retail problem: reducing stockouts and waste by converting raw grocery data into trusted metrics and recommended actions.

The research demonstrates how data engineering and AI can work together to support faster business decisions. The Bronze layer stores raw data, the Silver layer cleans and standardizes data, the Gold layer prepares business metrics, and the semantic layer provides consistent definitions. The application then uses these trusted metrics to power dashboards, natural language questions, inventory storytelling, and action workflows.

The project shows that the future of analytics is not limited to reporting. Data should help users understand what is happening, why it matters, and what action should be taken. Grocery Data Intelligence follows this direction by moving from dashboard visibility to decision support.

The findings indicate that Data + AI solutions can help grocery stores plan better, reduce stockouts, reduce waste, and act faster. The approach can also be extended to other retail and supply chain domains where timely action is critical.

Future Work

Future versions of Grocery Data Intelligence can include advanced demand forecasting using machine learning models. Forecasting can improve reorder recommendations by considering seasonality, promotions, holidays, weather, and local demand patterns.

Another future enhancement is integration with real point-of-sale and inventory management systems. This

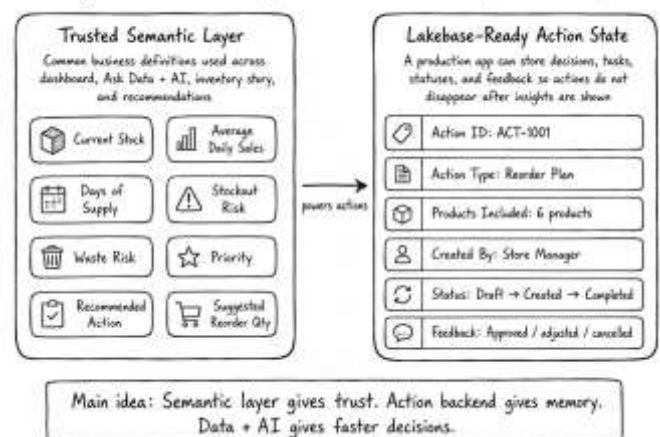
would allow the application to operate on live store data and provide near-real-time recommendations.

The action workflow can also be expanded. A transactional backend can store reorder plans, promotion plans, donation plans, status updates, and user feedback. This would turn the application into a complete decision and action system.

The Ask Data + AI experience can be improved further by adding conversational memory and role-based responses. Store managers, inventory planners, and regional leaders may ask different types of questions. The system can adapt responses based on user role and business context.

The semantic layer can be formalized with stronger metric governance. This would ensure that definitions such as stockout risk, waste risk, days of supply, and priority remain consistent across all dashboards, AI responses, and reports.

Figure 4: Semantic Layer and Lakebase-Ready Action Design



Future work can also include sustainability metrics. The system can track estimated waste avoided, donation value, and environmental impact. This would make the solution useful not only for profitability but also for social responsibility.

Case Study Summary

Grocery Data Intelligence was developed as a practical Data + AI demo to explore how grocery stores can move from raw data to faster business actions. The project focuses on inventory planning, stockout risk, waste risk, reorder recommendations, and action workflows.

The case study demonstrates how a lakehouse-based data architecture can support grocery analytics. Raw data is

stored in the Bronze layer, cleaned data is prepared in the Silver layer, and business-ready metrics are created in the Gold layer. A semantic layer helps keep metric definitions consistent across dashboards, Ask Data + AI, and recommended actions.

The application provides an intuitive user experience through dashboards, natural language questions, inventory story pages, and action review flows. The goal is to help store users make decisions without depending on manual report interpretation.

The project shows that Data + AI can help grocery businesses act faster. It also demonstrates the importance of clean data engineering foundations in building trustworthy AI-powered business applications.

Overall, Grocery Data Intelligence serves as a reference use case for data engineers, retail analysts, and AI practitioners interested in building practical decision intelligence systems.

Author Biography / About the Author

Brahma Reddy Katam is a data engineering professional, researcher, and technology enthusiast with experience in building scalable data platforms, analytics systems, and AI-driven solutions. His areas of interest include data engineering, cloud data architecture, Databricks, PySpark, SQL, lakehouse design, semantic analytics, and practical AI applications.

Brahma focuses on simplifying complex data systems and making data more useful for business users through clean engineering, intuitive design, and action-oriented analytics. He actively contributes to the data engineering community through articles, research papers, proof-of-concept applications, learning platforms, and practical Data + AI projects.

Grocery Data Intelligence is part of his ongoing exploration into how Data + AI can move organizations beyond traditional dashboards toward intelligent decision systems that help users understand, decide, and act faster.

References

1. Armbrust, M., et al. (2021). Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics. Proceedings of the VLDB Endowment, 14(12), 3183–3196.
2. Databricks, Inc. Delta Lake Documentation. Available at: <https://delta.io>
3. Databricks, Inc. Databricks Free Edition Documentation. Available at: <https://docs.databricks.com>
4. Databricks, Inc. Databricks Apps Documentation. Available at: <https://docs.databricks.com>
5. Databricks, Inc. Lakebase Postgres Documentation. Available at: <https://docs.databricks.com>
6. Databricks, Inc. AI/BI Genie Documentation. Available at: <https://docs.databricks.com>
7. Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Wiley.
8. Chopra, S., & Meindl, P. (2019). Supply Chain Management: Strategy, Planning, and Operation. Pearson.
9. Silver, E. A., Pyke, D. F., & Thomas, D. J. (2016). Inventory and Production Management in Supply Chains. CRC Press.
10. Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach. Pearson.
11. Provost, F., & Fawcett, T. (2013). Data Science for Business. O'Reilly Media.
12. Few, S. (2012). Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press.