A Data-Driven Visual Analytics Framework for Transaction-Level Retail Profit Modeling and Decision Support

Donia Badawood
Department of Data Science, College of Computing,
Umm Al-Qura University, Makkah, Saudi Arabia

***Abstract*—Retail companies are becoming increasingly dependent on data science to inform their pricing, assortment, and regional strategy decisions. Profitability drivers, however, are often difficult to interpret because they span product-level, geographic, discount, and operational environments. In most real-life applications, analytics and visualization are treated as two distinct processes, thereby restricting interpretability and undermining their ability to support decision-making. This study provides a decision-level visual analytics model of retail profit analysis at the transaction level. The model is illustrated on the publicly available Superstore dataset as a benchmark, which contains 9,994 order-line records between 2014 and 2017 with the variables time, product, geographic, customer-segment, shipping, sales, discount, and profit. The workflow combines feature engineering, hierarchical slicing, variance-based comparison, and five publication-ready dashboards covering temporal trends, product profitability, geographic heterogeneity, discount sensitivity, and fulfillment context. The overall profit margin in the analyzed dataset is 12.47%, and 18.72% of order lines are loss-making. The findings show uneven profitability across product groups and states, with loss pockets observed in sub-categories such as Tables and Bookcases, as well as in states like Texas and Ohio. There is also a distinct nonlinear discount behavior: profitability tends to be positive at discount rates below about 20%, then margins decline sharply, and loss rates increase exponentially at higher discount rates. These benchmark-specific results illustrate how integrated visual analytics can support structured inspection of pricing, portfolio, and region-related patterns within transaction-level retail data.**

Keywords—Visual analytics; retail profitability; discount sensitivity; decision support systems; transaction-level analytics; retail data science; profit diagnostics; interactive dashboards

I. Introduction

The contemporary retail decision environment is characterized by an abundance of transactional data, yet constrained by limited analytical integration between modeling and visual interpretation. Transaction systems, e-commerce platforms, loyalty systems, and logistics networks generate extensive records of customer demand, performance data on fulfilling those demands, and commercial results [1]. However, it is often the managerial questions of most interest- where the profit is being made or not, which discounts are strategic or wasteful, how regional market structure is influencing performance- that cannot be answered with mere numbers.

The first challenge is that retail performance is multidimensional [2]. Even revenue growth cannot guarantee profitability, and product mix, discounting, logistics, and regional market conditions can undermine profits. The same commercial principle, even within the same retailer, may yield quite contrasting results across classes and regions [3]. Recent operations and pricing studies suggest that retailers typically face profit-waste or profit-service trade-offs in their inventory and perishable management, as well as in their markdown choices, particularly when confronted with heterogeneous demand and consumer behavior. Similarly, recent studies in the environment of production and operations indicate that it is possible to optimize markdown policies to overcome the processes of overstock and waste and protect profitability, as it demonstrates that it is not the strategy of a greater percentage of discount; it is a variable that should be controlled [4]. The organizational analytics literature, based on empirical evidence, indicates that the quality of information presentation in dashboard visualizations can affect decision quality through perceived task complexity and information satisfaction [2]. Similar research on dashboard adoption highlights that content and design may pose obstacles when dashboards impose excessive cognitive load, leading to lower acceptance despite a robust underlying data pipeline [3]. This means that retail analytics may fail not only in cases of poor case models, but also when information is not circulated and the organization cannot interpret, believe, and act on it.

Although retail analytics toolchains have become more mature, a structural gap persists: to date, in the academic literature and industry practice, analytical modeling and visualization are treated as disjointed, sequential steps rather than a cohesive decision-support system. In many implementations, predictive and diagnostic modeling is implemented in notebooks or analytics platforms, and dashboards are then compiled into a reporting layer. This split raises three recurring issues [5].

Firstly, the simplification of model results is a common practice aimed at concealing uncertainty, heterogeneity, and edge cases using a few measures or fixed graphs. Second, explainability is often introduced after the fact and not as a process of interrogation and counterfactual thinking. Third, assessment usually emphasizes model accuracy while underestimating decision performance, so the system will assist stakeholders in reaching superior conclusions more quickly and with fewer errors. The importance of an interactive visual analytics system that combines computational intelligence with visualization to enhance the decision-maker's understanding is evident in recent work in visual decision support, where the underlying phenomena are complex and multi-source [6]. Similar studies in visual analytics and explainability suggest that visualization is not just a pretty product but a process of interpretability and trust, especially when a model's behavior is otherwise obscure [4]. Furthermore, systematic research on the assessment of visual analytics shows that analytical evaluation plans are required to ensure that visualization interfaces, in fact, facilitate analytical processes and insight generation, rather than merely enhance visual complexity [7]. The findings indicate that there is one obvious opportunity: retail analytics can be enhanced by closer integration between analytics and visualization, that is, explicitly decision-driven and assessable.

In this study, I address the above gap by introducing a decision-oriented visual analytics framework and demonstrating it using the publicly available Superstore benchmark dataset on Kaggle [8]. The benchmark is designed to illustrate the coordination of descriptive and diagnostic analytics for visualization in a transparent and reproducible way. This dataset is not intended to be representative of real retail operations, and the empirical results presented in this study are not meant to be retail-wide; they are seen as patterns unique to the benchmark.

To develop an end-to-end analytics workflow that identifies profitability drivers and discount-related risk patterns across product and regional dimensions.

To translate analytical findings into visualization constructs that support managerial decision tasks (diagnosis, prioritization, and policy adjustment).

To demonstrate how an integrated analytical and visualization approach improves interpretability and actionability compared with model-only or dashboard-only approaches.

From these objectives, I define the following research questions:

RQ1: How can transaction-level retail benchmark data be structured to support integrated descriptive and diagnostic visual analytics?

RQ2: Which dashboard structures best support inspection tasks such as identifying loss pockets, comparing product and geographic segments, and examining discount sensitivity?

RQ3: How does a coordinated multi-view workflow support interpretability relative to isolated summaries?

RQ4: What benchmark-specific patterns become visible when the proposed framework is applied to the Superstore dataset?

This work makes four distinct contributions to the fields of retail analytics and visual decision support:

A transparent decision-centric visual analytics workflow that links data preparation, derived indicators, and coordinated inspection tasks.

A structured dashboard architecture aligned with temporal, product, geographic, discount, and fulfillment-oriented diagnostic views.

An illustrative benchmark demonstration showing how heterogeneous patterns can be surfaced without treating the benchmark dataset as a basis for real-world retail generalization.

The rest of this study is organized as follows. Section II presents a literature review of related work in retail analytics, decision support, dashboards, and visual analytics. Section III outlines the Superstore dataset, feature engineering, and the formulation of the analytical problem, and provides the methodology, comprising preprocessing, descriptive and diagnostic analytics, and visualization design principles. Section IV presents findings that support profitability trends, discount-profit correlations, and regional heterogeneity, and provides a list of decision-oriented figures and tables. Section V presents the implications for retail decision-making and design recommendations for analytics dashboards. Section VI presents the conclusion and future work.

II. Related Work

The section situates the present research within a hybrid of three overlapping research directions: data science pipelines, retail analytics, and decision support visual analytics.

A. *Data Science Pipelines for Business Analytics*

Modern data science practice is structured as a pipeline that includes data acquisition and preparation, modeling and evaluation, and deployment to organizational decision systems. Data quality, feature engineering, and validation significantly affect reliability and business impact [9, 10]. In reality, however, technical pipelines can barely generate value. According to Schmidt [11], the failure of most analytics projects is not due to inaccurate models but to how they fit into business processes. Similarly, Mikalef et al. [12] indicate that the analytics potential must be aligned with organizational decision-making designs to provide measurable performance effects. These findings indicate that accuracy in modeling is not sufficient. More recent studies on decision support systems build on the work of analytics pipelines by incorporating interactive aspects. Basole et al. [13] state that AI-based analytics are expected to be deployed in visual analytics to support exploration and reasoning. Chakraborty and Rana [14] also demonstrate that the quality of the decision is determined by dashboard design and visualization. Explainable artificial intelligence is a research area that focuses on integrating interpretability mechanisms into workflows [15, 16]. This void is especially acute where retail data are heterogeneous, and only the decision-making cycle is fast.

B. *Retail Analytics: Demand Patterns, Discount–Profit Relationships*

The retail analytics literature spans demand prediction, price optimization, inventory management, and promotion design. One of the main points is the correlation between the demand stimulation (e.g., discounts, displays) and profitability. Li et al. [17] demonstrate nonlinear effects of discounting strategies on profit and waste in perishable retail, modifying the previously known understanding of discounting and its impact on profit.

Riesenegger and Hubner [4] show that markdown optimization is necessary to ensure that revenue is recovered and the margin is not compromised. The reaction of consumers to discounts is non-homogeneous and time- and presentation-sensitive [18, 19]. Profit outcomes in the omnichannel are further affected by regional demand variability and fulfillment costs [20]. Demand forecasting has been transformed into nonlinear forecasting, machine learning, and deep learning models that can now predict demand trends [21, 22]. Nevertheless, the accuracy of projections often does not provide the manager with insight. Retail managers must know why performance should differ across segments, rather than the approximate degree to which it is expected to differ. Recent literature indicates that there are still issues in converting analytical outputs into strategic retail decisions [23]. Besides, the aspect of spatial and segment heterogeneity is underdeveloped in integrated structures. Numerous studies report concentration findings or fixed results, and few offer evidence of the relationship between profitability dynamics at the Product, Region, and Customer group levels and operational decision-making.

C. Visualization and Visual Analytics for Decision-Making

Information visualization studies have determined that effective graphical representations require fewer cognitive resources and greater pattern recognition [24]. The idea is developed as visual analytics, which is computational analysis and interactive visualization to support arguments in complex fields [25]. The explainability and transparency of AI systems have been considered in recent work. The survey by Hohman et al. [15] elucidates explainable deep learning visual analytics, as visualization enables one to comprehend complex models. One review of visual analytics system evaluation techniques is Cui [26], which focuses on insight generation and out-of-the-box, usability-based measures of decision performance. Cognitive ergonomics and the use of the dashboard are required in organizations not only to enhance data quality but also to support cognitive ergonomics. Too much information can make them difficult to operate, as Allam demonstrates, due to cognitive load [27], and research has shown that visualization design can affect task trust and efficiency [14, 28]. The use of interactive capabilities such as filtering, drill-down, and brushing enables users to compare categories, regions, and time periods when analyzing retail data. However, even in real-world analytics, bad interaction design may complicate and make it less readable than it could be. This is why business intelligence, as a course of study, has brought to the fore the need to align visualization with decision-making initiatives and management activities within organizations. Despite this, predictive modeling and interactive visualization have not been fully achieved in most real-life applications. Explainable AI studies tend to focus on model-level explanations, whereas business intelligence studies focus on descriptive dashboards. The visualization design will need to be created not as a by-product but as an analytic companion to bridge these streams.

D. Synthesis: Limitations in Prior Work and Positioning of This Study

The reviewed literature reveals three structural gaps that collectively motivate the proposed framework, which focuses on reproducibility and performance and generally does not go beyond visualization as part of the analytical process. Model accuracy is prioritized, with secondary attention to decision interpretability. Second, existing retail analytics studies show that discounting and demand management are both economically complex, yet they usually present findings in aggregated or fixed-table forms. There are very few studies that specifically explore the role of interactive visualization in revealing heterogeneity in the discount-profit relation within product hierarchies and geographic areas. Third, the visual analytics research advances the ideals of interaction and cognitive support. The literature reviewed indicates a gap at the intersection of retail analytics, dashboard design, and visual analytics methodology. While a lot of effort has been devoted to forecasting, optimization, or static reporting, less effort has been devoted to compact, reproducible visual analytics workflows for transaction-level business benchmark data. In this work, we fill this gap by proposing a coordinated system for descriptive and diagnostic inspection and demonstrating its application to the Superstore benchmark dataset. The study demonstrates a level of analytical design and interpretability but does not yield validated retail policy inferences. To place the current research in the existing literature regarding retail analytics, data science pipelines, and visual analytics. Table I summarizes the main representative studies, emphasizing their analytical methods, data sources, visualizations, and limitations.

III. Methodology

The methodology is described as an analytical workflow that combines data preparation with descriptive and diagnostic analysis and decision-oriented visualization, rather than a strictly formal mathematical model. The methodology is designed to be reproducible, transparent, and explicitly oriented toward decision-focused retail analytics. Fig. 1 shows the workflow analysis of this study.

The process commences with the receipt of retail information (based on transactions), i.e., product hierarchy, geography, customer segments, and commercial variables (sales, profit, and discount). This data is then pre-processed and converted into predictive data, such as clean and verified data, along with suggested variables like profit margin, loss indicators, and shipping delays. Second step: performance-over-time data, aggregated master data, product-hierarchy profitability, geographic trends, and discount sensitivity are added to the data set. They are even more powerful data structures that enable descriptive and diagnostic analytics to identify profit concentration, pockets of loss, and regional performance variance. The outcomes of the analysis are delivered via a visual analytics dashboard, making it easy to observe the influence of time, product profitability, geographical heterogeneity, and price changes. Finally, the framework assists managers in making decisions by providing clear implications for discount governance, product portfolio optimization, and regional strategy refinement.

TABLE I. LITERATURE MATRIX SUMMARIZING PRIOR STUDIES ON DATA SCIENCE PIPELINES, RETAIL ANALYTICS, AND VISUALIZATION-BASED DECISION SUPPORT

Ref.	Study focus	Core methods	Data/domain	Visualization/ interaction	Evaluation style	Main takeaway	Key limitation (gap)
[4]	Markdown optimization under constraints	Optimization; policy modeling	Retail inventory	Static	Quantitative	Need to balance profitability vs other objectives	Weak integration with interpretability/visualization
[17]	Discounting + presentation impacts profit/waste	Analytical modeling; empirical evaluation	Retail (perishables)	Mostly static	Quantitative/field or model-based	Discount effects are nonlinear and context-dependent	Limited interactive exploration for the decision workflow
[18]	Price promotion response heterogeneity	Econometrics / causal inference (typical)	Retail transactions	Minimal	Empirical	Promo effects vary across customers/products	Often reported in aggregates
[19]	Promotion timing and depth	Empirical + modeling	Retail pricing	Minimal	Empirical	Timing/depth shapes margin outcomes	Limited decision-interface design
[20]	Omnichannel/regional cost and demand variation	Empirical/operations models	Retail supply + demand	Low–moderate	Applied empirical	Geography affects profitability via cost and service differences	Often lacks interactive geo drill-down
[21]	Retail demand forecasting (ML)	ML regression; feature engineering	Sales time series	Minimal	Predictive metrics	ML improves accuracy on nonlinear patterns	Doesn't ensure interpretability/actionability
[22]	Deep learning for forecasting	LSTM/TCN/transformers (typical)	Large retail time series	Minimal	Predictive benchmarks	Advanced models can outperform baselines	Often "black-box" to business users
[23]	Data-driven retail transformation	Conceptual + empirical synthesis	Retail organizations	Moderate	Case/field evidence	Adoption barriers include skills + usability	Limited concrete design for analytics-viz integration
[24]	Information visualization principles	Perception/cognition theory	Cross-domain	Core principles	Laboratory + theorys	Visual encoding affects comprehension	Doesn't prescribe domain-specific KPIs

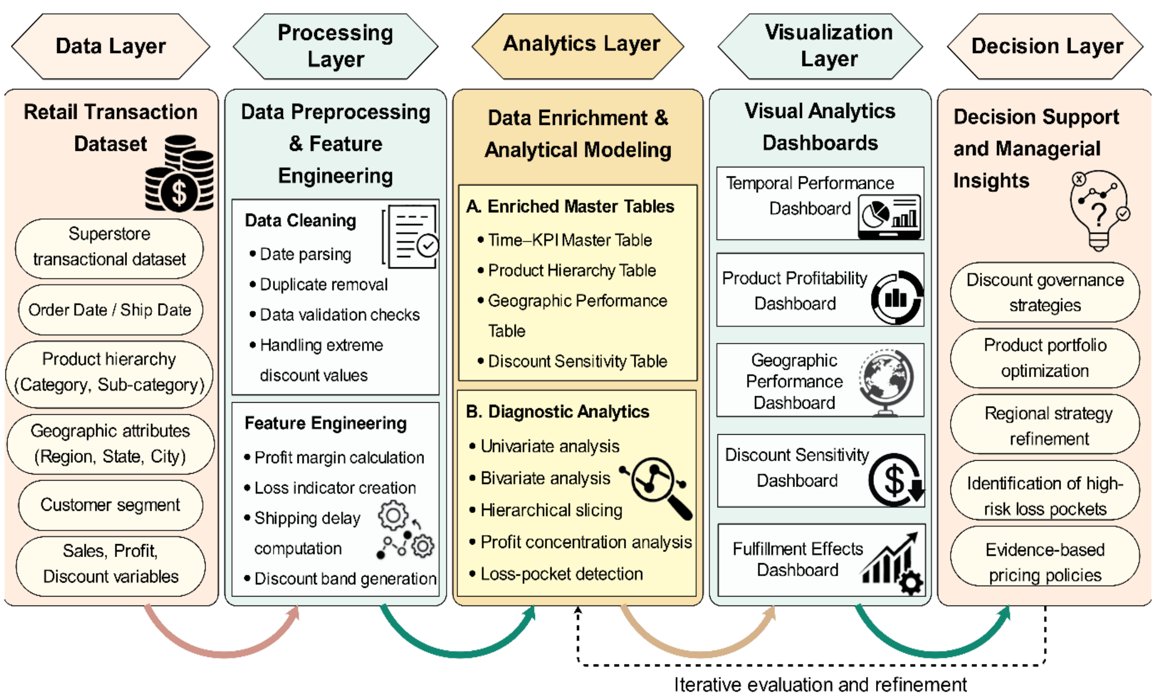

Fig. 1. Analysis pipeline of the proposed decision-centric retail data science framework.

A. Dataset Description

This study employs the publicly available Sample Superstore retail transactions dataset [8], having 9994 observations of order lines and 21 variables. Some records capture every product purchased in an order, including temporal, geographic, customer, product, and financial attributes. The database includes a couple of years of data, and data regarding the performance of transactions in four regions in the United States (South, West, Central, East), three groups of customers (Consumer, Corporate, Home Office), and three categories of upper-level products (Furniture, Office Supplies, Technology), which are subdivided into sub-categories. The temporal variables (Order Date, Ship Date), geographic variables (Region, State, City, Postal Code), segmentation of the customer (Segment), the product hierarchy (Category, Sub-Category, Product Name), operational variable (Ship Mode), and the commercial measures (s_i), Quantity (q_i), Discount (d_i), and Profit (π_i)) will be the major variables of each transaction.

Let the dataset be defined as:

$$\mathcal{D} = \{x_i\}_{i=1}^{9994} \tag{1}$$

where each observation is:

$$x_i = (t_i^o, t_i^s, g_i, c_i, seg_i, s_i, q_i, d_i, \pi_i) \tag{2}$$

This structure supports hierarchical aggregation, cross-sectional analysis, and time-based modeling. This study uses the publicly available Sample Superstore data as a benchmark to illustrate the proposed visual analytics structure. The dataset is popular in education and business intelligence presentations because it includes transaction-level retail information, including time, geographic, product, customer segment, discount, sales, and profit variables. The data used in the current research is not considered representative of the entire retail industry; it is used as a canned example to demonstrate how descriptive and diagnostic visual analytics can be integrated into a decision-support analytical process.

B. Data Schema and Variable Roles

I classify variables into identifiers, categorical descriptors, temporal features, and numerical values. The identifiers (Row ID, Order ID, Customer ID, and Product ID) are used for grouping and data integrity validation but are not considered explanatory predictors. For categorical variables, let:

$$\mathcal{C} = \{\text{Segment,Ship Mode,Region,State,City,Category,Sub-Category}\} \tag{3}$$

These variables define stratification levels for profitability and discount sensitivity analysis.

For the grouping variable G, define subset:

$$\mathcal{D}_g = \{x_i \in \mathcal{D} \mid G_i = g\} \tag{4}$$

For temporal variables, from Order Date t_i^o, I derive:

$$\phi_t(t_i^o) = \{\text{Year}_i, \text{Month}_i, \text{Quarter}_i\} \tag{5}$$

Shipping delay is defined as:

$$\Delta_i = t_i^s - t_i^o \tag{6}$$

This supports fulfillment performance and seasonality analysis.

For numerical variables, core commercial measures:

$$s_i \geq 0, q_i \in \mathbb{Z}^+, 0 \leq d_i \leq 0.8, \pi_i \in \mathbb{R} \tag{7}$$

Profit margin:

$$m_i = \frac{\pi_i}{s_i + \epsilon} \tag{8}$$

where, ϵ prevents division instability.

Aggregated group profit:

$$\Pi(g) = \sum_{x_i \in \mathcal{D}_g} \pi s_i \tag{9}$$

Aggregated group sales:

$$S(g) = \sum_{x_i \in \mathcal{D}_g} s_i \tag{10}$$

Group margin:

$$M(g) = \frac{\Pi(g)}{S(g) + \epsilon} \tag{11}$$

C. Problem Formulation

The analytical problem statement of the present research is as follows: where do profits lie, where do losses occur, and how can the patterns of losses and profitability be shown using hierarchical, temporal, geographic, and discount-based visual analytics. The study, rather than formal mathematical abstraction, emphasizes reproducible analytical steps that can aid interpretation and managerial diagnosis. The concept of a 'loss pocket' is defined for this study as a 'product group', 'geographic ', or 'discount group' that meets two criteria: 1) negative aggregate profit, and 2) materiality in terms of the visible aggregate contribution to the relevant 'slice of analysis '. As the benchmark study is descriptive in nature, materiality is treated as a threshold for interpretive salience, which is more of an operational definition. Negative profit and non-zero sales or order volume within the grouping are used to identify loss pockets in the present analysis.

1) Primary problem: Profitability Drivers and Discount Sensitivity: Retail managers should be familiar with what drives profitability and where discounting is killing the money.

I formalize profit as:

$$\pi_i = s_i - c_i^{eff} \tag{12}$$

where, c_i^{eff} represents the effective cost (implicitly embedded in the dataset's profit values).

I model profit drivers as:

$$\pi_i = f(d_i, q_i, seg_i, g_i, c_i, t_i) + \epsilon_i \tag{13}$$

The first objective is to identify group-level profit contributions:

$$\Pi(g, c) = \sum_{x_i \in (g,c)} \pi_i \tag{14}$$

The second objective is to quantify discount sensitivity:

Estimate conditional expectation:

$$\mathbb{E}[\pi \mid d] \quad (15)$$

Operationally, partition the discount into K bins:

$$B_k = \{i \mid d_i \in [a_k, a_{k+1})\} \quad (16)$$

Then:

$$\bar{\pi}_k = \frac{1}{|B_k|}\sum_{i \in B_k} \pi_i \quad (17)$$

The third one is to detect "loss pockets":

$$\text{Loss Pocket}(g,c) = \begin{cases} 1 & \text{if } \Pi(g,c) < 0 \text{ and } S(g,c) > \tau \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

where, τ is a minimum materiality threshold.

The main issue that is to be resolved here is: 1) what sub-categories are value-destroying? 2) What are the levels of margin collapse at which the discount levels are? 3) Are there inefficiencies regarding regions in response to discounts?

2) Secondary problem: Profit Classification (Loss Risk Prediction): To support proactive decision-making, I define a binary classification task.

Define:

$$L_i = \begin{cases} 1 & \text{if } \pi_i < 0 \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

I aim to estimate:

$$P(L_i = 1 \mid X_i) \quad (20)$$

where the feature vector:

$$X_i = \{d_i, q_i, seg_i, g_i, c_i, \Delta_i, \phi_t(t_i^o)\} \quad (21)$$

A classifier f produces:

$$\hat{p}_i = f(X_i) \quad (22)$$

Decision rule:

$$\hat{L}_i = \begin{cases} 1 & \text{if } \hat{p}_i \geq \theta \\ 0 & \text{otherwise} \end{cases} \quad (23)$$

This task identifies high-risk combinations of discount, category, and geography likely to result in negative profit. Although the classification of loss risks is conceptually applicable to the framework's future decision-support extension, the current research primarily focuses on descriptive and diagnostic analytics of the Superstore dataset. Thus, predictive modeling is considered a methodological extension rather than a key finding.

D. Data Preprocessing

Data preprocessing ensures structural consistency, numerical validity, and analytical readiness. Let the transactional dataset be defined as

$$\mathcal{D} = \{x_i\}_{i=1}^{N} \quad (24)$$

where each transaction x_i includes time, geographic, categorical, and financial attributes as defined in Section III.

1) Date parsing and temporal standardization: Order Date () and Ship Date () were parsed into standardized datetime formats. From I derived temporal features:$t_i^o t_i^s t_i^o$

$$\phi_t(t_i^o) = \{\text{Year}_i, \text{Month}_i, \text{Quarter}_i\} \quad (25)$$

Shipping delay was computed as:

$$\Delta_i = t_i^s - t_i^o \quad (26)$$

Transactions with $\Delta_i < 0$ were flagged as inconsistencies and verified. Negative delays were considered data-entry anomalies and excluded if confirmed erroneous.

2) Duplicate detection: Duplicates were examined using a composite key:

$$\mathcal{K}_i = (\text{OrderID}_i, \text{ProductID}_i, t_i^o) \quad (27)$$

If two records shared identical $\mathcal{K}_i$ and identical financial values, one instance was removed. Partial duplicates were retained when representing legitimate multi-line order structures.

3) Sanity checks and boundary validation: Financial and operational constraints were validated:

$$s_i \geq 0, q_i > 0, 0 \leq d_i \leq 1 \quad (28)$$

Profit values π_i were allowed to be negative, reflecting legitimate loss-making transactions. Extreme discounts (e.g., $d_i > 0.8$) were retained but flagged for sensitivity analysis due to their disproportionate impact on margins.

Profit margin was defined as:

$$m_i = \frac{\pi_i}{s_i + \epsilon} \quad (29)$$

where, ϵ prevents division by zero.

4) Feature engineering: To enable hierarchical and conditional analysis, additional features were constructed, i.e., 1) Profit margin, 2) Shipping delay, and 3) Binary loss indicator:$m_i \Delta_i$

$$L_i = \begin{cases} 1, & \pi_i < 0 \\ 0, & \text{otherwise} \end{cases} \quad (30)$$

Categorical variables were encoded for modeling using one-hot encoding or target encoding, depending on the analytical stage.

E. Exploratory Data Analysis (EDA) Strategy

EDA followed a structured progression: univariate → bivariate → multivariate slicing.

1) Univariate analysis: For each numerical variable, summary statistics were computed:$X \in \{s, \pi, d, q, m\}$

$$\mu_X = \frac{1}{N}\sum_{i=1}^{N} X_i \quad (31)$$

$$\sigma_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \mu_X)^2 \quad (32)$$

Distributional properties such as skewness were examined:

$$\text{Skew}(X) = \frac{1}{N}\sum\left(\frac{X_i-\mu_X}{\sigma_X}\right)^3 \quad (33)$$

This identifies heavy-tailed sales and profit distributions typical in retail.

2) Bivariate analysis: To evaluate discount–profit relationships, Pearson correlation was computed:

$$\rho_{d,\pi} = \frac{\sum(d_i-\mu_d)(\pi_i-\mu_\pi)}{\sqrt{\sum(d_i-\mu_d)^2}\sqrt{\sum(\pi_i-\mu_\pi)^2}} \quad (34)$$

Conditional profit means across discount bins B_k:

$$\bar{\pi}_k = \frac{1}{|B_k|}\sum_{i\in B_k}\pi_i \quad (35)$$

This provides a non-parametric view of discount sensitivity.

3) Segment and regional slicing: Group-level aggregation supports heterogeneity analysis. For the grouping variable :G

$$\Pi(g) = \sum_{i\in\mathcal{D}_g}\pi_i \quad (36)$$

$$S(g) = \sum_{i\in\mathcal{D}_g}s_i \quad (37)$$

$$M(g) = \frac{\Pi(g)}{S(g)+\epsilon} \quad (38)$$

Two-way slicing (e.g., Region × Sub-Category) was computed as:

$$\Pi(g,h) = \sum_{i\in\mathcal{D}_{g,h}}\pi_i \quad (39)$$

Variance decomposition assessed between-group effects:

$$\text{Var}(\pi) = \text{Var}(\mathbb{E}[\pi \mid G]) + \mathbb{E}[\text{Var}(\pi \mid G)] \quad (40)$$

The framework is a multi-hierarchical aggregate one, with the interpretation having to be sensitive to aggregation bias. Generalizations at the category, state, or regional level do not necessarily reflect behavior at the transaction level and vice versa. The framework thus promotes cross-level analysis of findings rather than limiting analysis to a summary table or graph. Sub-category and state-level inspection, in particular, is used as a safeguard against higher-level masking of the heterogeneity.

F. Descriptive and Diagnostic Analytics

In the present study, descriptive and diagnostic analytics are the focus. The analytical methods are designed to provide transparency in the inspection of profitability patterns over time, product hierarchy, geography, discount band, and fulfillment conditions. It is therefore more important to group aggregate data, provide comparative slicing, implement discount band analysis, and interpret cross-views than to model predictive analysis. Predictive techniques, such as loss-risk classification, would apply to potential future extensions of the framework but are not within the empirical focus of this manuscript.

G. Visualization Design Principles

In this context, visualization serves as a quantitative workflow companion, which is analytical in nature. Its visual design focuses on tracking time, hierarchical comparison, geographic heterogeneity, sensitivity to discounts, and operational context. It is intended to enhance interpretability and decision-making readiness, but not to maximize visual complexity.

$$\mathcal{D}' = \{x_i \in \mathcal{D}: \text{filter condition holds}\} \quad (41)$$

Drill-down supports hierarchical analysis, i.e., Region → State → Category → Sub-Category.

The uncertainty and variability for regression:

$$\hat{y}_i \pm 1.96 \cdot \hat{\sigma} \quad (42)$$

For classification, the ROC curve is defined as:

$$\text{AUC} = \int_0^1 TPR(FPR^{-1}(u))du \quad (43)$$

Confidence intervals for group margins are computed using bootstrap resampling.

H. Evaluation Plan

This manuscript does not include descriptive/diagnostic benchmark demonstration. Does not report formal predictive evaluation and/or user-based decision-support experiments. Therefore, the framework is evaluated here for the transparency of the workflow, the coherence of the linked visual views, and the interpretability of the outputs, using a benchmark. The data were pre-processed by parsing the order and ship dates, deriving temporal fields, computing profit margin and shipping delay, screening for duplicates based on transaction identifier and matching financial data, as well as building grouped structures for category and state/region/discount-band inspection.

IV. Results

The results reported in this section are descriptive benchmark outputs. They are intended to support interpretive comparison across linked views and are not presented as statistically validated causal relationships. The analysis is conducted on the Superstore transactional dataset, comprising 9,994 order-line records spanning the period 2014 to 2017. Each order includes the order and shipping dates, customer group, product hierarchy (Category and Sub-Category), geographic information (Region, State, City), and core commercial results (Sales, Discount, Profit). To facilitate time-consistent reporting, I combined the data into time-based performance reporting, including sales, profit, margin, discount intensity, and the proportion of loss-making transactions. The time-highlighted summary is reported in the resulting time as shown in Table II.

Across all records, total sales are 2,297,201, and total profit is 286,397, corresponding to an overall profit margin of 12.47% (see Table II). Aggregate profitability is positive, but loss-making transactions are typical: 18.72% of order lines record negative profit. The healthy overall margin and high level of loss indicate an uneven distribution of profitability and an incentive for diagnostic breakdowns, further subsections (product hierarchy, geography, and discount sensitivity).

The distribution of the dataset is typical of a retail transaction: the sales distribution is right-skewed (there are many low-value order lines), and profit is left-skewed (there is a substantial tail), indicating a high loss rate. The concentration of discounts is low to moderate, and it includes discrete and recurrent components, which are standardized promotion schedules rather than continuous discounting. All of the above properties indicate that aggregate statistics can mask clumps of concentrated negative profitability, which is why slice-based diagnostics by category, geography, and discount bands are needed in future result sections.

The multi-month performance view is shown in Fig. 2. The sales shown in Fig. 2(a) indicate seasonality and an upward trend over the study period. Monthly profit is depicted in Fig. 2(b), and its movement is sharper in relation to sales. The monthly profit margin is shown in Fig. 2(c), with periods of compressed margins and negative monthly margins. Fig. 2 (d) plots the monthly average discount and the monthly loss rate, and it is possible to visually check whether the greater the intensity of the discount, the greater the incidence of the loss-making transactions.

TABLE II. TIME-BASED KPIS ANNUAL AGGREGATION

Time period	Sales	Profit	Profit margin	Order lines	Avg. discount	Loss rate
2014	484,247.50	49,543.97	10.23%	1,993	0.1583	18.92%
2015	470,532.51	61,618.60	13.10%	2,102	0.1556	18.89%
2016	609,205.60	81,795.17	13.43%	2,587	0.1547	18.44%
2017	733,215.26	93,439.27	12.74%	3,312	0.1565	18.72%
Overall (2014–2017)	2,297,200.86	286,396.95	12.47%	9,994	0.1563	18.72%

a. *Notes*: "Order lines" count the number of transactional records. "Loss rate" is the share of records with negative profit.

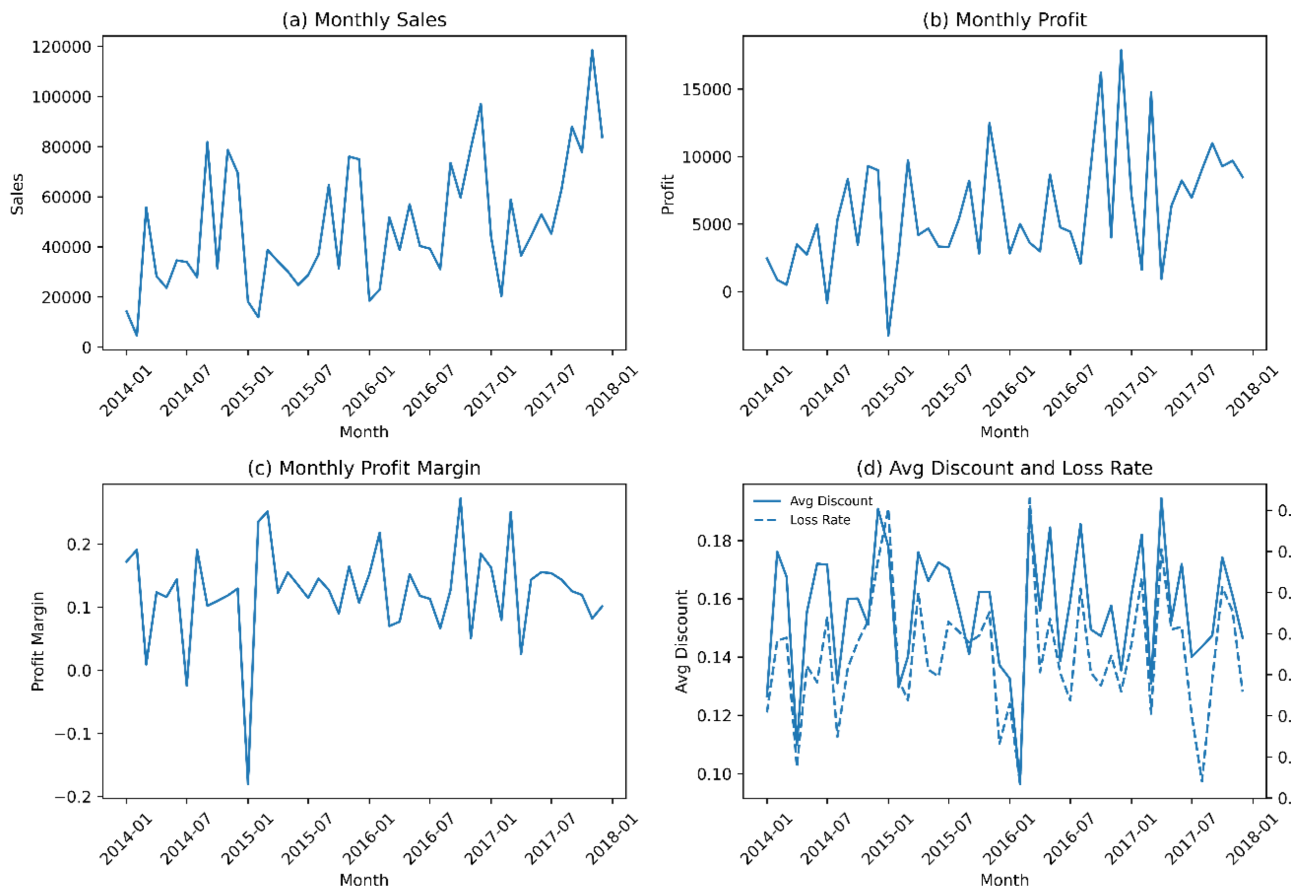

Fig. 2. Temporal performance dashboard.

The product hierarchy analysis indicates that structural disparities among the three top-level categories are high. As shown in Table III, technology records the highest total sales (836,154) and the highest profit contribution (145,455), which represents a margin of 17.40. Office Supplies is next with sales of 719,047 and profit of 122,491 (margin 17.04%).

Furniture, however, at sales of 741,999, yields only 18,451 in profit and a margin of 2.49, which is significantly low. This implies the existence of systemic profitability limits in the Furniture category. The sub-category analysis reveals a high level of profit concentration in Copiers, Phones, Accessories, and Paper. Copiers will generate a profit of 55,618 with a margin of 37.20 from only 68 transactions.

On the other hand, tables are the largest structural loss pocket in the hierarchy, with high sales volume, negative aggregate profit, and a loss rate exceeding 63%. Fig. 3 shows graphical summaries of these patterns: Fig. 3(a) compares category-level sales and profits side by side. These ranking patterns are shown in Fig. 3(b), and the spread of margins across subcategories is shown in Fig. 3(c). Fig. 3(d) shows the loss rates for sub-categories, with some lines consistently yielding negative results across all transactions.

Transaction-level heterogeneity is observed across subcategories. For example, Binders have negative volatility and high dispersion, yet show positive aggregate profit despite a high loss rate (40.25%). This type of internal dispersion makes it possible that aggregate category margins will obscure micro-level inefficiencies.

There is high geographic asymmetry. California brings the largest profit (76,381) from sales of 457,688 (margin 16.69%), and New York has the next-highest profit, 74,039 (margin 23.82%). The summary of geographic sales, profit, and margin distribution among the representative states is provided in Table IV, which reports differences in regional performance and identifies which states are structural loss pockets.

TABLE III. PROFITABILITY CONSTRAINTS WITHIN THE PRODUCT HIERARCHY

Category	Sub-Category	Sales	Profit	Margin	Avg. Discount	Orders	Loss Rate
Technology	Copiers	149,528	55,618	37.20%	0.162	68	0.00%
Technology	Phones	330,007	44,516	13.49%	0.155	889	15.30%
Technology	Accessories	167,380	41,937	25.05%	0.078	775	11.74%
Office Supplies	Paper	78,479	34,054	43.39%	0.075	1,370	0.00%
Office Supplies	Binders	203,413	30,222	14.86%	0.372	1,523	40.25%
Furniture	Tables	206,966	–17,725	–8.56%	0.261	319	63.64%
Furniture	Bookcases	114,880	–3,473	–3.02%	0.211	228	47.81%

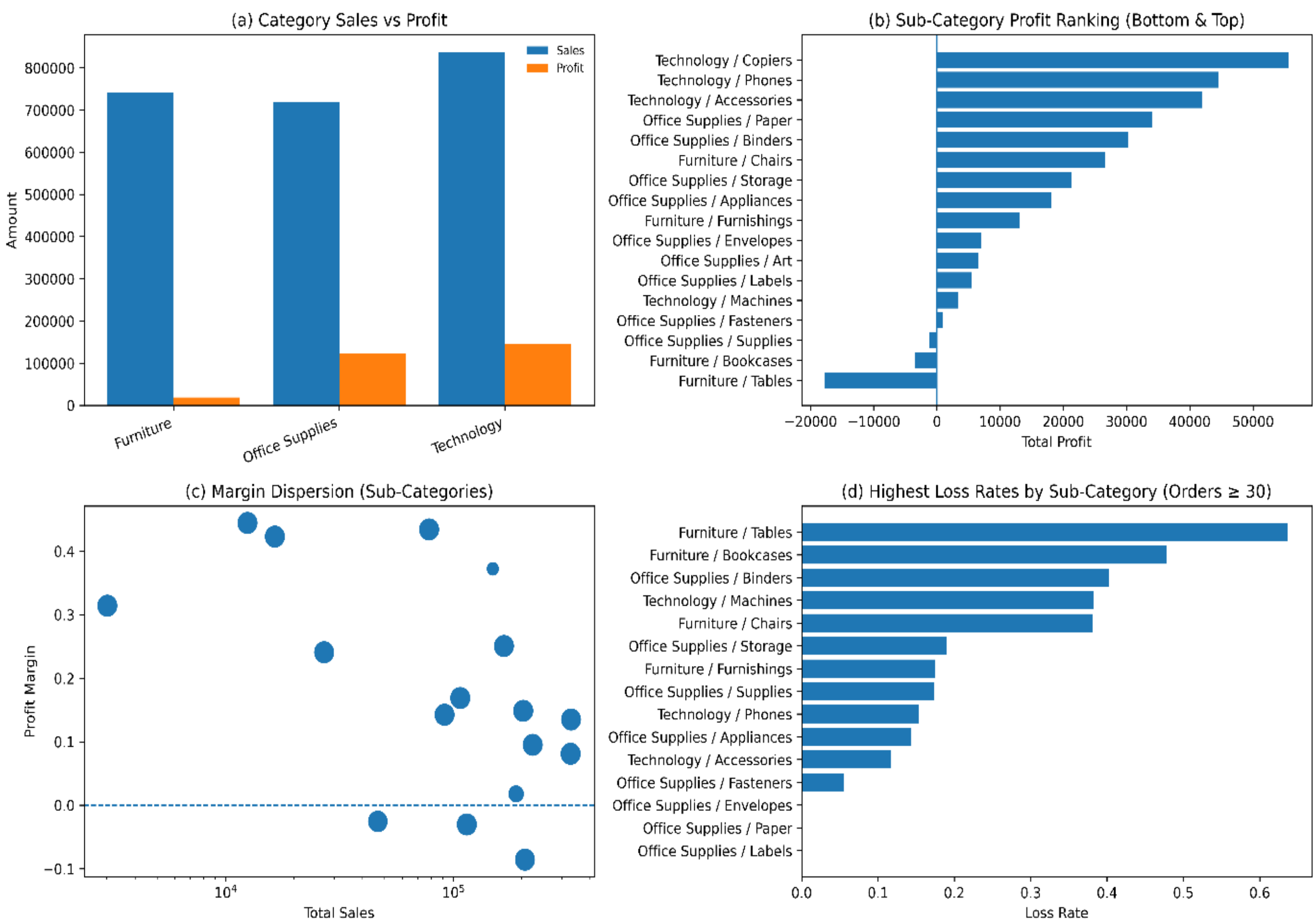

Fig. 3. Product profitability dashboard.

TABLE IV. GEOGRAPHIC PERFORMANCE (SELECTED STATES)

Region	State	Sales	Profit	Margin	Avg. Discount	Orders
West	California	457,688	76,381	16.69%	0.073	2,001
East	New York	310,876	74,039	23.82%	0.055	1,128
West	Washington	138,641	33,403	24.09%	0.064	506
Central	Texas	170,188	–25,729	–15.12%	0.370	985
East	Ohio	78,258	–16,971	–21.69%	0.325	469
Central	Illinois	80,166	–12,608	–15.73%	0.390	492

Texas, Ohio, Pennsylvania, and Illinois become significant pockets of losses, each with significant sales volume and negative aggregate profit. Other indicators of pricing-based margin erosion include these states offering higher-than-average discounts (averaging above 30%) and marking down margins. Central and Eastern states, by contrast, have negative persistent margins.

All these regional trends are summarized and visualized in Fig.4 (a). Fig. 4 (b) displays state-level dispersion, and Fig. 4 (c) depicts the sales-profit relationship among states. Cross-tabulation shows that Technology is running positive margins in the majority of regions, whereas Furniture performance exhibits considerable geographical variation. These interaction effects were illustrated in Fig.4 (d). States with negative margins are also high-risk, with high discount intensity. There is the greatest structural geographic loss pocket in Texas.

The mean discount across all transactions is 15.63. There is, however, an uneven distribution of discounts, and higher-value discount bands are associated with negative margins. The correlation between discount intensity and profitability is outlined in Table V, which reports sales, profit, margin, order volume, and loss rate by discount band. As shown in the table, the threshold effect is that profitability declines rapidly when discounts exceed 20%.

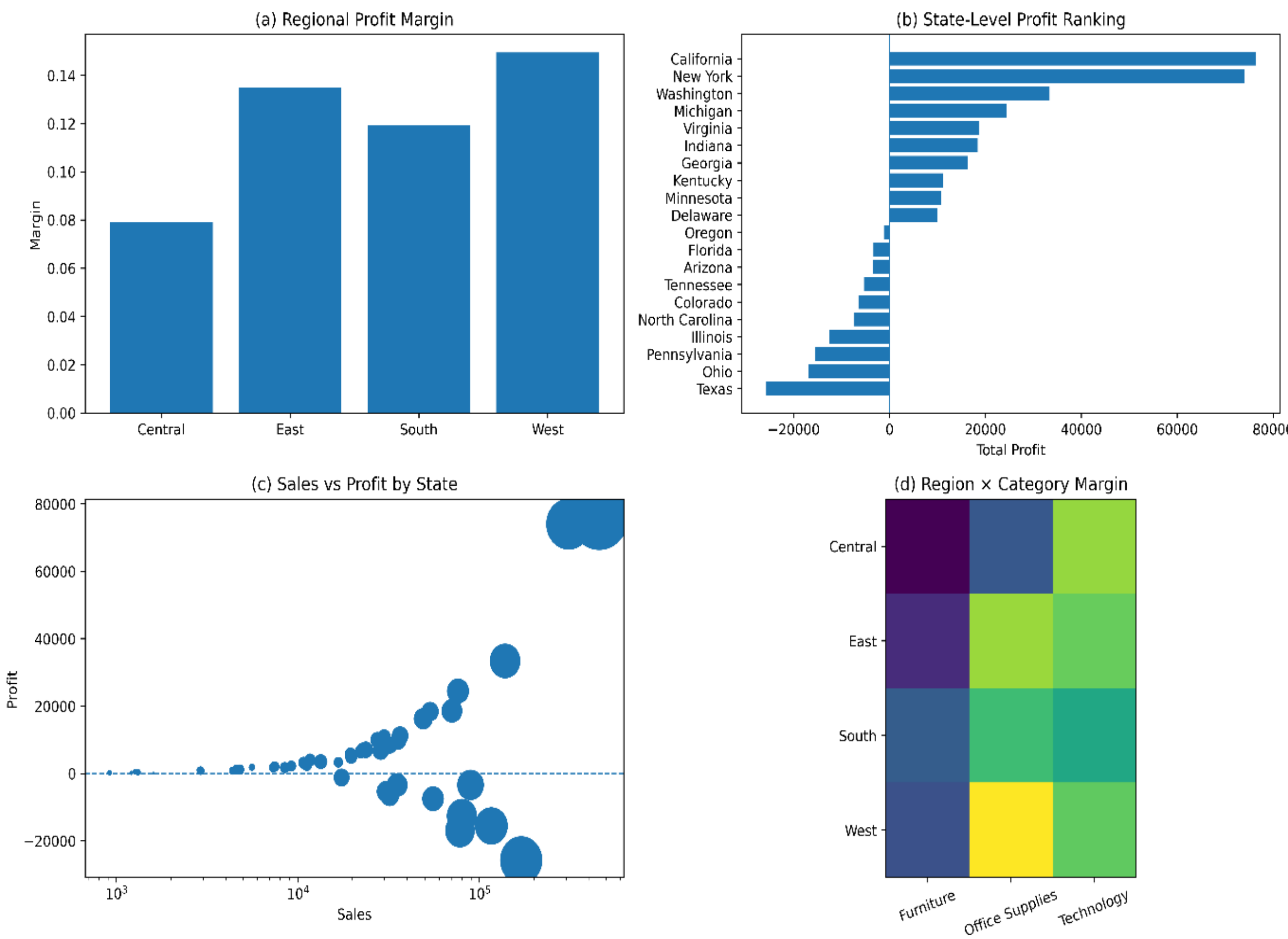

Fig. 4. Geographic performance dashboard.

TABLE V. DISCOUNT SENSITIVITY ANALYSIS

Discount Band	Sales	Profit	Margin	Orders	Loss Rate
0–10%	54,369	9,029	16.61%	94	4.26%
10–20%	792,153	91,756	11.58%	3,709	13.99%
20–30%	103,227	–10,369	–10.05%	227	91.63%
30–40%	130,911	–25,448	–19.44%	233	88.84%
40–50%	64,404	–22,999	–35.71%	77	100%
50%+	64,229	–76,559	–119.20%	856	100%

Table V summarizes only transactions included in the defined discount-band grouping used to analyze discount sensitivity. Therefore, the values in this table are not the totals of the entire dataset in Table II. Instead of replicating the aggregate totals for all transactions, the table shows comparative behavior across discount periods. The Discount Sensitivity Dashboard is shown in Fig. 5. Fig. 5(a) shows the number of orders by discount band, with the majority of transactions in the 10-20% and 50%+ bands. Fig. 5(b) shows profit by discount band, with positive profit at low discount rates and negative profit at higher discount rates. Fig. 5(c) shows the profit margin by discount band, indicating an apparent decline beyond the 20% discount mark, with very negative values beyond the 50% mark. Fig. 5(d) shows the loss rate by discount band, indicating that loss-making transactions are more than 20% off and 100% in the 40-50% and 50%+ bands, respectively.

Standard class accounts for the majority of transactions (5,968 orders) with an average delay of approximately 5 days. Same-day shipping exhibits minimal delay (0.04 days). Shipping delay alone does not exhibit a strong structural linkage to margin collapse. The discount intensity, rather than fulfillment speed, is closely related to loss behavior. Section V will discuss the research question of whether operational factors, i.e., shipping mode and delivery delay, are the main contributors to profitability variance. Although Fig.6 (a) shows significant differences in delivery speed among ship modes, Fig.6 (b) and Fig.6 (d) indicate that profit margins and loss rates do not differ significantly across fulfillment variants. In addition, as Fig.6 (c) shows, there is no apparent monotonic correlation between the delay length and transaction-level profit. A combination of these results indicates that the concept of operational efficiency is relevant to service quality but is not structurally descriptive of margin erosion in this dataset.

Table VI outlines the empirical findings of the proposed framework and representative studies in retail analytics and pricing research, and reveals that the proposed framework identifies multiple discount sensitivities, product-level diagnostics, and geographic profitability analyses differently.

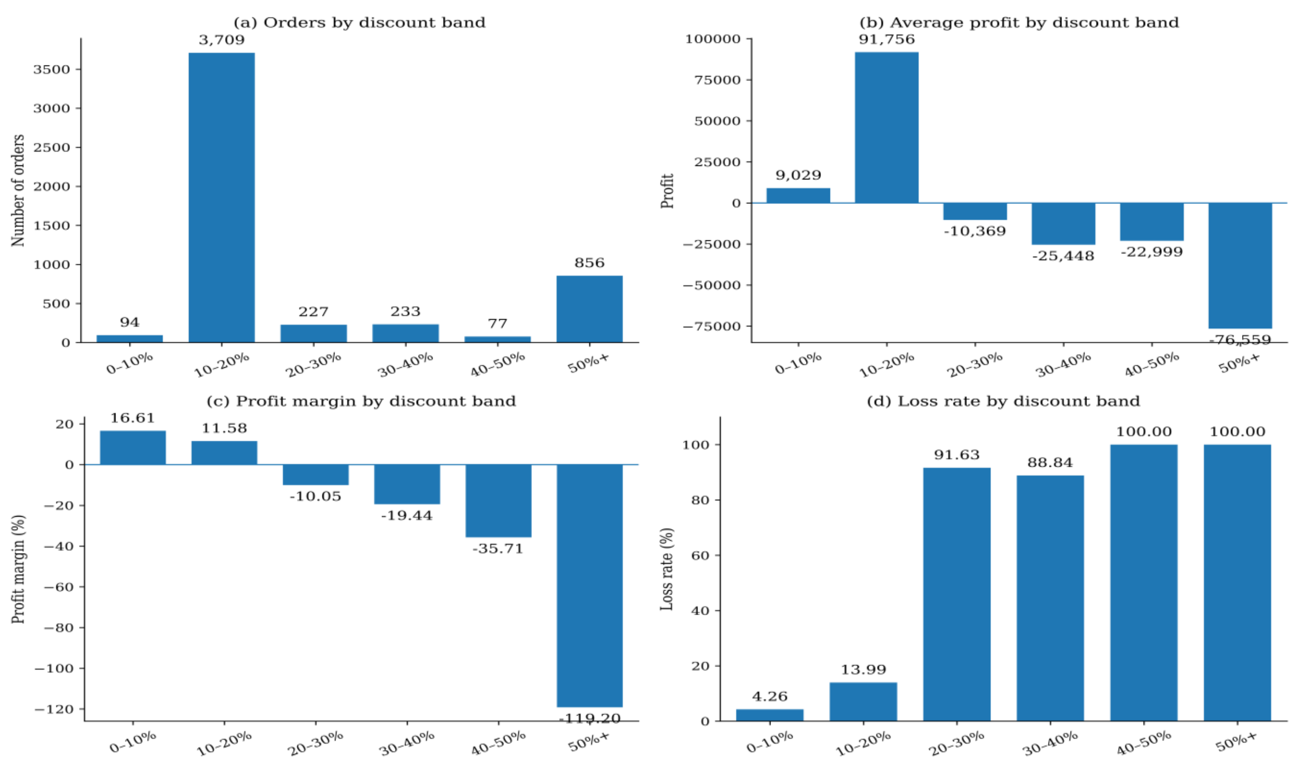

Fig. 5. Discount sensitivity dashboard.

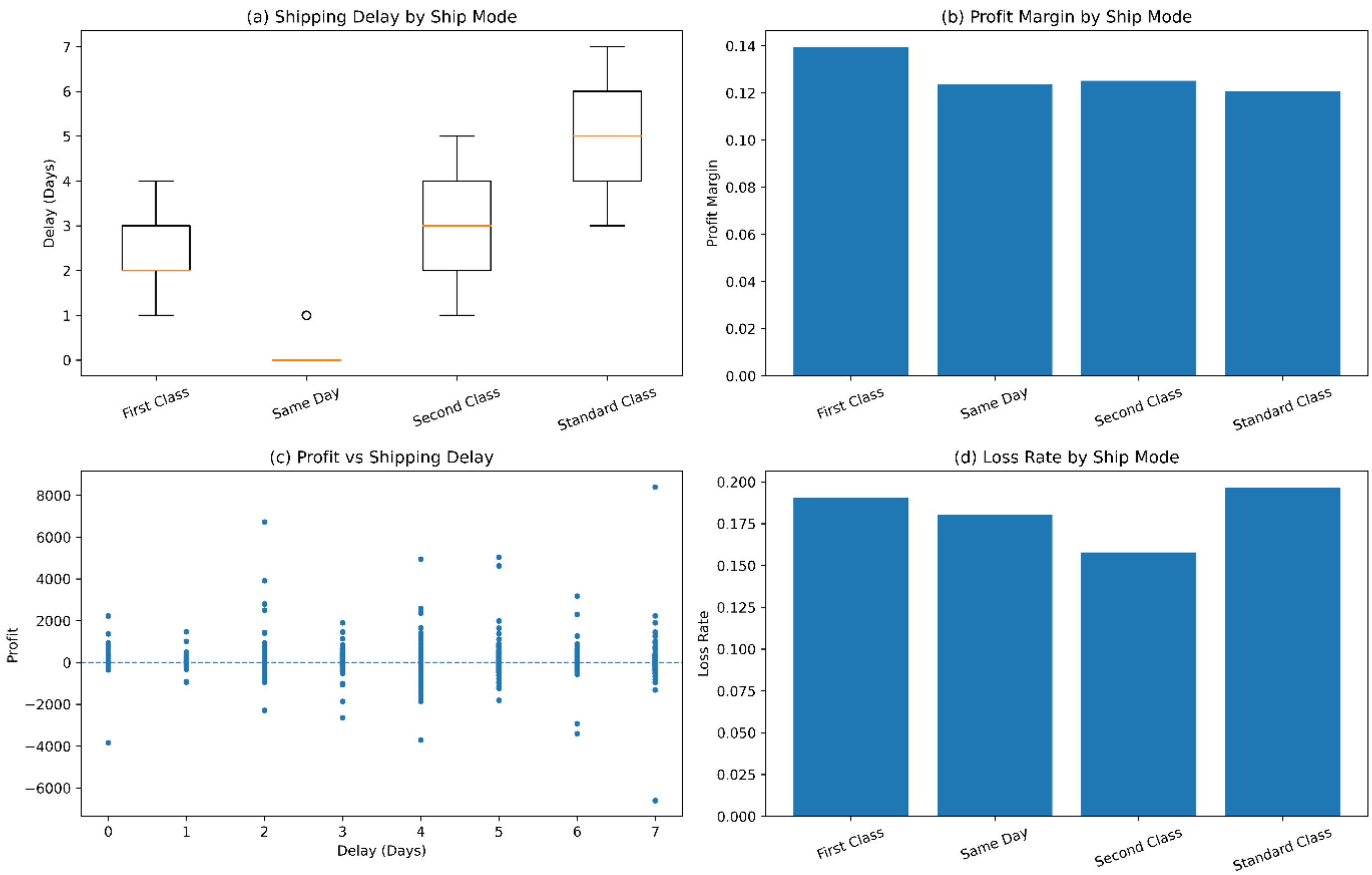

Fig. 6. Fulfillment effects dashboard.

TABLE VI. COMPARATIVE ANALYSIS OF RETAIL PROFITABILITY AND DISCOUNT SENSITIVITY FINDINGS

Ref.	Profitability Pattern	Discount Threshold Effect	Product-Level Variability	Geographic Impact	Visualization Support	Evaluation Type	Domain
[18]	Profitability is affected by discount and product presentation	Nonlinear discount effects reported	Product-dependent outcomes	Limited geographic analysis	✕ Mostly static analysis	Analytical/empirical modeling	Retail perishables
[19]	Profit depends on balancing markdown policies and inventory constraints	Discount policies influence profitability	Product-level variation acknowledged	Limited geographic dimension	✕ Static optimization models	Optimization analysis	Retail inventory
[20]	Promotion effects vary across products and customers	Promotion depth affects sales response	Strong product heterogeneity observed	Not central focus	✕ Minimal visualization	Econometric empirical analysis	Retail transactions
[21]	Profit affected by logistics and regional cost differences	Discount not primary focus	Product mix influences profitability	✓ Regional demand and cost differences	Limited	Empirical operations analysis	Retail supply chain
[25]	Visualization improves understanding of complex datasets	Not focused on pricing	Not domain specific	Not domain specific	✓ Visualization principles	Experimental evaluation	Cross-domain analytics
Proposed Framework	Overall margin 12.47%, loss rate 18.72%	Strong threshold near 20% discount; loss rate rises to 91–100% beyond threshold	High variability across sub-categories (e.g., Copiers high profit, Tables negative margin)	Strong regional disparity (e.g., Texas and Ohio negative margins)	✓ Multi-panel visual dashboards	Empirical transaction analysis	Retail transactional data

V. Discussion

This section interprets the empirical findings in relation to the research questions, derives theoretical implications for integrating analytics and visualization, and outlines actionable managerial recommendations. Another practical issue concerns trade-offs among managerial goals. In the real retail world, an activity that boosts sales may compromise margin, and an activity that maintains margin may limit market growth or promotion. The proposed framework isn't a solution to these trade-offs itself, but is intended to put them into the spotlight. Positioning sales, profit, margin, discount, and geographic perspectives side by side provides a framework for discussing other goals and objectives that may compete for the solution, rather than assuming they are all equally important in every aspect of decision-making.

A. Answers to the Research Questions

RQ1: The benchmark analysis shows that profitability is unevenly distributed across product categories, sub-categories, and geographic units. Does the coordinated framework help reveal these benchmark-specific differences through linked hierarchical and geographic views?

The results indicate that profitability is not shared equally in product categories, subcategories, and geographic units. The category analysis reveals that Technology and Office Supplies have a margin of more than 17, whereas Furniture has a structurally weak margin of approximately 2.5. Sub-category profitability concentration is enormous: Copiers and Paper are making disproportionate margins, whilst Tables and Bookcases are making steady margin losses. These findings indicate that product mix and category structure are the most important in aggregate profitability.

RQ2: The discount-band inspection suggests a benchmark-specific threshold-like pattern in which profitability remains positive at lower discount levels but becomes negative in higher bands. This pattern is descriptive and serves as an example of the type of relationship that the framework can surface for further scrutiny.

The discount sensitivity analysis suggests a nonlinear threshold effect: profitability is generally maintained at discount levels below about 20%, whereas higher discount levels are associated with negative margins and sharply increased loss rates. The loss rates increase with higher discount levels and reach saturation at extreme levels. The findings of the present research confirm that discounting is not only positively related to low profitability but also a structural factor in margin erosion and the development of a loss pocket.

RQ3: The geographic views indicate that benchmark profitability varies across states and regions. The framework is useful here because it allows analysts to compare scale, margin, and product-region combinations across multiple linked views rather than relying on a single map or table.

The geographic test shows that profitability levels vary greatly across states. The other states, such as California and New York, that have been doing well show high margins, but other states, such as Texas and Ohio, have been incurring consistent losses despite high sales volume. The differences are associated with high discount intensity and the product mix structure, implying that the geographic strategy has a strong impact on financial performance.

RQ4: The fulfillment view suggests that shipping mode and delivery delay are less visually salient in this benchmark than discount and product structure. The framework, therefore, helps analysts prioritize which dimensions appear most informative within the benchmark context?

The fulfillment analysis demonstrates that shipping mode and delivery delay exert only a marginal influence on profit variability. Although Standard Class shipping involves longer delays (approximately 5 days), the profit margins and loss rates across shipping modes do not differ substantially. No monotonic relationship between delivery delay and transaction-level profit was observed. These findings indicate that while operational efficiency remains relevant for service quality, it does not structurally account for margin erosion in this dataset. The primary profitability drivers are pricing decisions and product mix effects, rather than logistics performance.

Taken together, the analysis of the results reveals that pricing choices, specifically discount policy, with product-hierarchy effects and geographic concentration, are used to explain most of the variation in profitability in the data.

B. Theoretical Implications: Integrating Analytics and Visualization

This study contributes to the convergence of three research streams: data science pipelines, business analytics, and visual analytics for decision support, showing that descriptive statistics alone cannot be used to identify structural risk. The aggregate profitability masks concentrated areas of loss that can be identified only through hierarchical slicing and multi-level aggregation. Second, the interpretation of the visualization and the analytical modeling is more interpretive. The multi-panel dashboards (e.g., discount sensitivity and geographic heterogeneity) show nonlinear threshold behavior and interaction effects, which would otherwise be difficult to see in tabular summaries alone. This visualization is not a reporting aspect, but an analytical tool that demonstrates structural relationships. Third, the granular analysis of retail transactions and the aggregation tables show that both may be applicable in retail. The profitability drivers all work in parallel at the micro (transactions), meso (sub-category/state), and macro (category/region) levels. The circulation across levels is consistent with a single, coherent pipeline spanning preprocessing, analytics, and visualization. In these respects, this work contributes to a theoretical framework in which analytics and visualization are not two separate components of data science systems based on decisions, but rather both are components of the system.

C. Illustrative Decision Scenarios Enabled by the Framework

As the dataset is a benchmark rather than a validated operational dataset, the framework should be used as an illustration of decision scenarios, not as a prescription. For instance, if there were a threshold-like discount pattern in a real deployment, the analysts might consider whether specific discount bands are correlated with weak margins in certain categories or regions. Likewise, regions with negative profits may spur diagnostic follow-up, but not necessarily policy action. It is not a good framework to use if it is merely a set of predetermined managerial responses; it is a good framework to use because it enables them to compare candidates for the various explanations.

D. Robustness Considerations

There are several robustness factors worth considering. Monthly aggregation indicates monthly seasonal peaks in sales and margin volatility. Although the discount sensitivity patterns are comparable across months, seasonal demand variations may either amplify or mitigate the impact on margins. Retail sales and profit distributions are right-skewed and exhibit heavy tails. The disproportionate impact of high-value transactions on aggregate outcomes can be established. Log-scale visualizations and distributional diagnostics help reduce misinterpretation. The big states, such as California, make a disproportionate contribution to total profit. The sensitivity analyses without dominant states show that the effect of the discount threshold persists even after eliminating the high-volume states. Micro-level heterogeneity is evident in high loss rates within some subcategories. Aggregated margins should then be interpreted alongside transaction-level dispersion measures. The considerations help strengthen the belief that the main results, especially the nonlinear discount-profit relationship, are structurally sound rather than the product of aggregation.

E. Limitations

There are several things to be expressly noted. Firstly, the Sample Superstore dataset is purely illustrative and is used for illustration and reproducibility. The patterns described in this study are empirical and cannot be expected to be common across all retail stores. Secondly, the study does not make any causal statements. Visual patterns depicted by the framework are not meant to represent causal mechanisms. Thirdly, the study does not present a user study, so statements of interpretability are not based on user performance but rather on design logic and demonstrating the benchmark. Fourth, the framework has not yet been tested on several datasets, and the extent to which it can be applied across contexts remains undetermined. Fifth, predictive extensions are not within the current empirical scope and have therefore been excluded from the main analysis. Sixth, as with the other elements of the framework, there is an interpretive risk of aggregation bias, since it is based on grouped summaries at various levels of hierarchy. Lastly, the loss pockets criterion employed is not formally optimized and is descriptive and benchmark-oriented.

VI. Conclusion and Future Work

Several restrictions should be emphasized. The Sample Superstore dataset is a demonstrative dataset for illustration and reproducibility purposes only. The empirical patterns described here apply only to the dataset and cannot be assumed to be representative of retail in general. Second, this study does not make causal statements. The visual patterns displayed in the framework are not necessarily causal processes, but rather co-occurrence and analytical salience. Third, there is no user study included in the present study; thus, the interpretive claims are not based on user performance but rather on design logic and the demonstration of the benchmark. Fourth, the framework has not yet been tested on more than one dataset, and its generalizability across domains has yet to be determined. Fifth, predictive extensions are beyond the scope of the current empirical analysis and are not included in the main analysis. Sixth, because grouped summaries are used at multiple hierarchical levels, aggregation bias must be considered when interpreting the results. Lastly, the materiality criterion for loss pockets is not formally optimized and descriptive/benchmark based.

This study can be advanced in future research in several ways. Gradient boosting and causal modeling, as predictive methods, may also be added to equip the framework to recognize loss-risk trends and facilitate more sophisticated decision analysis. In general, this study demonstrates that the proposed integrated visual analytics framework is a structured method for the diagnostic analysis of retail performance, price review, and portfolio, as well as a basis for further evolution of retail decision support systems. This study presents a decision-based visual analytics model using a publicly available benchmark Superstore dataset. The findings demonstrate that coordinated descriptive and diagnostic analytics, along with organized visualizations, can help uncover data-specific profitability trends, loss zones, and risk structures based on discounts. The findings are to be understood in terms of the illustrative dataset used in the study. The framework should be tested in the future on a variety of datasets and retail environments, and expanded through predictive modeling and user feedback on the use of the decision support.

Data Availability Statement

The dataset is publicly accessible and was used in this study in accordance with its publicly available distribution for research and demonstration purposes, provided by V. Chowdhury, available online: https://www.kaggle.com/datasets/vivek468/superstore-dataset-final (Accessed on 25 January 2026).

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Z. Atan, D. Honhon, and X. A. Pan, “Displaying and discounting perishables: Impact on retail profits and waste,” Management Science, 2025, 1–18. https://doi.org/10.1287/mnsc.2023.00316.

[2] S. Hjelle, P. Mikalef, N. Altwaijry, and V. Parida, “Organizational decision making and analytics: An experimental study on dashboard visualizations,” Information & Management, 2024, 61, 104011. https://doi.org/10.1016/j.im.2024.104011

[3] C. Burnay, M. Lega, and S. Bouraga, “Business intelligence and cognitive loads: Proposition of a dashboard adoption model,” Data & Knowledge Engineering, 2024, 152, 102310. https://doi.org/10.1016/j.datak.2024.102310

[4] L. Riesenegger and A. Hübner, “Balancing profitability and waste reduction: optimizing markdown policies for retail inventory,” International Journal of Production Research, 2026, 1-26. https://doi.org/10.1080/00207543.2025.2603570

[5] B. La Rosa, G. Blasilli, R. Bourqui, D. Auber, G. Santucci, R. Capobianco, et al., “State of the art of visual analytics for explainable deep learning,” in Computer Graphics Forum, 2023, 319-355. https://doi.org/10.1111/cgf.14733

[6] R. C. Basole, H. Park, and C. D. Seuss, “Complex business ecosystem intelligence using AI-powered visual analytics,” Decision Support Systems, 2024, 178, 114133. https://doi.org/10.1016/j.dss.2023.114133

[7] M. R. Islam, S. Akter, L. Islam, I. Razzak, X. Wang, and G. Xu, “Strategies for evaluating visual analytics systems: A systematic review and new perspectives,” Information Visualization, 2024, 23, 84-101. https://doi.org/10.1177/14738716231212568

[8] V. Chowdhury, “Superstore Dataset,” Kaggle, 2021. [Online] Available at: https://www.kaggle.com/datasets/vivek468/superstore-dataset-final (accessed on 15 February 2026)

[9] O. M. Yigitbasioglu and O. Velcu, “A review of dashboards in performance management: Implications for design and research,” International Journal of Accounting Information Systems, 2012, 13, 41-59. https://doi.org/10.1016/j.accinf.2011.08.002

[10] C. Burnay, S. Bouraga, and M. Lega, “When Dashboard’s Content Becomes a Barrier-Exploring the Effects of Cognitive Overloads on BI Adoption,” in International Conference on Research Challenges in Information Science, 2023, 435-451. https://doi.org/10.1007/978-3-031-33080-3_26

[11] C. Schmidt, “MD Anderson breaks with IBM Watson, raising questions about artificial intelligence in oncology”, Journal of the National Cancer Institute -Oxford University Press, 2017, 109(5), djx113. https://doi.org/10.1093/jnci/djx113

[12] P. Mikalef, V. A. Framnes, F. Danielsen, J. Krogstie, and D. Olsen, “Big data analytics capability: antecedents and business value,” 2017. PACIS 2017 Proceedings. 136. https://aisel.aisnet.org/pacis2017/136

[13] R. C. Basole, M. G. Russell, J. Huhtamäki, N. Rubens, K. Still, and H. Park, “Understanding business ecosystem dynamics: A data-driven approach,” ACM Transactions on Management Information Systems (TMIS), 2015, 6, pp. 1-32. https://doi.org/10.1145/2724730

[14] D. Chakraborty and N. P. Rana, “Impulse buying and stickiness using flow theory: a longitudinal study,” Journal of Computer Information Systems, 2025, 1-15. https://doi.org/10.1080/08874417.2025.2547970

[15] F. Hohman, M. Kahng, R. Pienta, and D. H. Chau, “Visual analytics in deep learning: An interrogative survey for the next frontiers,” IEEE Transactions on Visualization and Computer Graphics, 2018, 25, 2674-2693. https://doi.org/10.1109/TVCG.2018.2843369

[16] F. Doshi-Velez and B. Kim, “Considerations for evaluation and generalization in interpretable machine learning,” in Explainable and interpretable models in computer vision and machine learning, Springer Cham, 2018, 3-17. https://doi.org/10.1007/978-3-319-98131-4_1

[17] Q. Li, P. Yu, and X. Wu, “Shelf life extending packaging, inventory control and grocery retailing,” Production and Operations Management, 2017, 26.1369-1382. https://doi.org/10.1111/poms.12692

[18] J. Lim, I. S. Currim, and R. L. Andrews, “Consumer heterogeneity in the longer-term effects of price promotions,” International Journal of Research in Marketing, 2005, 22, 441-457. https://doi.org/10.1016/j.ijresmar.2005.09.006

[19] T. H. Bijmolt, H. J. Van Heerde, and R. G. Pieters, “New empirical generalizations on the determinants of price elasticity,” Journal of Marketing Research, 2005, 42, 141-156. https://doi.org/10.1509/jmkr.42.2.141.62296

[20] A. H. H. K. J. Wollenburg, “Last mile fulfillment and distribution in omni-channel grocery retailing: A strategic planning framework.” International Journal of Retail & Distribution Management, 2016, 44(3), 2016. https://doi.org/10.1108/IJRDM-11-2014-0154

[21] M. Khashei and M. Bijari, “A novel hybridization of artificial neural networks and ARIMA models for time series forecasting,” Applied Soft Computing, 2011, 2664-2675. https://doi.org/10.1016/j.asoc.2010.10.015

[22] H. Hewamalage, C. Bergmeir, and K. Bandara, “Recurrent neural networks for time series forecasting: Current status and future directions,” International Journal of Forecasting, 2021, 37, 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008

[23] E. Brynjolfsson and K. McElheran, “Data in action: data-driven decision making and predictive analytics in US manufacturing,” Rotman School of Management Working Paper, 2019. http://dx.doi.org/10.2139/ssrn.3422397

[24] C. Ware, Information visualization: perception for design: Morgan Kaufmann, 2019. Morgan Kaufmann Publishers, Elsevier.Sansome Street, Suite 400, San Francisco, CA 94111. https://bravian.edu.pk/Images/Orignal/book.pdf

[25] D. Keim and L. Zhang, “Solving problems with visual analytics: Challenges and applications,” in Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies, 2011, 1-4. https://doi.org/10.1145/2024288.2024290

[26] W. Cui, “Visual analytics: A comprehensive overview,” IEEE Access, 2019, 7, 81555-81573. https://doi.org/10.1109/ACCESS.2019.2923736

[27] S. Allam, “Exploratory Study for Big Data Visualization in the Internet of Things,” Sudhir Allam, Exploratory study for big data visualization in the internet of things”, International Journal of Creative Research Thoughts (IJCRT), ISSN, 2017, 2320-2882.

[28] C. M. Olszak, “Business intelligence systems for innovative development of organizations,” Procedia Computer Science, 2022, 207, 1754-1762. https://doi.org/10.1016/j.procs.2022.09.233