Ecommerce Sales Analysis Using Data Analytic Tools

Prof. Dr. N. M. Halgare[1], Prof. Dr. Ashwini A. Patil[2], Nikita Dattatry Kotwal[3]

Department of Information Technology

M. S. Bidve Engineering College Latur, India

Email - halgare@gmail.com[1] ,ashwinibiradar29@gmail.com[2],nikitakotwal2004@gmail.com[3]

ABSTRACT—The rapid expansion of e-commerce platforms has resulted in the generation of massive volumes of transactional and customer-related data, creating significant opportunities for data-driven business intelligence and decision-making. Effective analysis of e-commerce sales data is essential for understanding customer behavior, identifying revenue trends, optimizing pricing strategies, and improving overall operational efficiency. This paper presents a comprehensive E-Commerce Sales Analysis system that leverages structured data analytics techniques using SQL, Python, and data visualization tools to extract meaningful insights from large-scale sales datasets.

The proposed system focuses on analyzing key business metrics such as total sales, profit, order volume, discount impact, shipping performance, and time-based sales trends. SQL is employed as the primary tool for data extraction, cleaning, aggregation, and transformation, enabling efficient handling of large relational datasets. Python is utilized for advanced data preprocessing, statistical analysis, and trend evaluation, while visualization tools are used to represent insights through interactive dashboards and charts. The analysis incorporates month-wise and category-wise sales performance, identification of high-revenue and high-profit product, and assessment of delivery efficiency.

By examining sales patterns over time, the system helps in detecting seasonal demand fluctuations and identifying peak and low-performing periods. The study also analyzes the relationship between discounts and profitability, providing valuable insights into pricing effectiveness and margin optimization. Additionally, shipping and delivery performance analysis highlights logistical bottlenecks that impact customer satisfaction and order fulfillment efficiency.

The results demonstrate that structured e-commerce data analysis significantly enhances business visibility and supports strategic decision-making. The system enables organizations to identify profitable product segments, optimize marketing and promotional strategies

Keywords— E-Commerce Sales Analysis, Data Analytics, SQL, Python, Business Intelligence, Data Visualization, Sales Trends, Profit Analysis, Decision Support System.

I. INTRODUCTION

E-commerce has become one of the most significant pillars of the global digital economy, playing a vital role in modern business operations and consumer purchasing behavior. With the rapid growth of internet accessibility and digital payment systems, online retail platforms generate massive volumes of transactional data on a daily basis [1]. The global e-commerce market has witnessed exponential growth over the past decade and is expected to expand further due to increasing consumer reliance on online shopping, mobile commerce, and personalized digital experiences. This rapid growth has led to an increased demand for efficient sales monitoring, performance evaluation, and data-driven decision-making mechanisms.

As competition among e-commerce platforms intensifies, businesses face challenges in understanding customer behavior, identifying profitable products, managing discounts, and optimizing supply chain and logistics operations. Seasonal demand fluctuations, dynamic pricing strategies, delayed deliveries, and ineffective promotional campaigns can significantly impact sales performance and profitability [2]. A single poor sales cycle caused by improper inventory planning, excessive discounting, or delayed shipment can lead to revenue loss and reduced customer satisfaction. Therefore, accurate estimation and continuous monitoring of e-commerce sales performance are essential for sustainable business growth.

In many advanced economies, data-driven analytical systems are used to evaluate sales trends, forecast demand, and optimize marketing strategies based on regional and temporal factors [3]. With the increasing scale of online transactions, traditional manual analysis techniques are no longer sufficient to handle large and complex datasets. Consequently, automated sales analysis and performance monitoring systems have become a necessity rather than an option [4]. An effective analytical framework must consider multiple influencing factors such as sales volume, profit margins, discounts, delivery performance, and time-based trends to support better business decisions [5].

The core objective of e-commerce sales analysis is to maximize revenue and profitability while improving operational efficiency and customer experience. In recent years, data analytics, SQL-based querying, and machine learning techniques have been widely adopted to extract meaningful insights from structured e-commerce datasets. These techniques help organizations identify hidden patterns, detect anomalies, and evaluate the effectiveness of pricing and promotional strategies. Advanced analytical models assist in reducing losses caused by inefficient discount policies, poor inventory planning, and delayed deliveries.

Existing analytical approaches have utilized statistical methods and rule-based reporting systems to evaluate sales performance. Some studies have applied regression-based models to forecast sales trends, while others have used classification techniques to identify high-performing products and customer segments [6]. Machine learning-based approaches such as decision trees, random forests, and neural networks have also been explored for demand forecasting and sales prediction, demonstrating improved accuracy over traditional methods [7]. However, these models often require large datasets and complex implementation, making them less suitable for exploratory business analysis and real-time decision support [8], [9].

Structured data analysis using SQL, combined with Python-based analytics and visualization tools, provides an effective and scalable solution for e-commerce sales evaluation. These techniques enable efficient data aggregation, trend analysis, and KPI computation while maintaining transparency and interpretability of results. Applications of e-commerce sales analytics include revenue forecasting, profit optimization, customer behavior analysis, supply chain monitoring, and performance evaluation of promotional campaigns [10]. Such analytical systems significantly enhance business productivity while reducing manual effort and operational risk.

This research paper analyzes e-commerce sales data using structured data analytics techniques to extract actionable business insights. The remainder of this paper is organized as follows: Section II describes the methodology adopted for e-commerce sales analysis. Section III discusses the analytical techniques and KPIs used for performance evaluation. Sections IV and V present the problem statement and objectives of the study. Section VI provides a comparative analysis of related research works. Finally, Section VII concludes the paper and outlines future research directions.

II. RELATED WORK AND EXISTING SYSTEMS

The analysis of e-commerce sales data has become a critical area of research due to the increasing importance of digital commerce and the availability of large-scale transactional data. Several studies have proposed methods ranging from traditional reporting to advanced predictive analytics. Below is a point-wise summary of significant research contributions.

A. Traditional Reporting System

1. Static Sales Reports

Early e-commerce platforms relied on static reporting using Excel or basic database queries [1].

Reports included total sales, number of orders, and revenue summaries.

Limitations: Only descriptive analytics; no predictive or prescriptive insights; time-consuming manual extraction.

2. Manual KPI Monitoring

Businesses tracked metrics such as average order value, customer acquisition, and monthly revenue manually [2].

Limitations: High risk of human error; inefficient for large datasets; lacked visualization capabilities.

B. SQL-Based Analytical Approaches

3. Relational Database Analytics

Kim et al. (2018) used SQL to extract KPIs like total sales, total profit, order count, and regional performance [3].

Enabled efficient aggregation and historical trend analysis.

Limitation: Requires technical knowledge; primarily historical analysis; limited integration with visualization or predictive models.

4. Advanced Query Techniques

Use of GROUP BY, JOINs, and window functions for time-series sales analysis [4].

Improved analysis of category-wise and month-wise sales patterns.

Limitation: Cannot directly handle unstructured or semi-structured data; does not provide insights into customer segmentation or discount effects.

C. Business Intelligence (BI) Dashboards

5. Interactive Visualization Tools

Tools like Power BI, Tableau, and QlikView have been applied to visualize e-commerce KPIs [5].

Common features: regional sales maps, trend analysis, category-wise revenue, and delivery performance charts.

Limitation: Often dependent on preprocessed datasets; limited real-time analytics; cannot perform advanced predictive modeling.

6. Sales Forecasting Dashboards

Dashboards integrated with statistical forecasting (e.g., ARIMA) for predicting future sales [6].

Advantages: Improved managerial decision-making; early detection of low-performing products.

Limitation: Requires consistent historical data; minimal support for multi-factor analysis like discounts, promotions, and delivery performance.

D. Machine Learning and Hybrid Approaches

7. Predictive Analytics for Sales Forecasting

Random Forest, Decision Trees, and Linear Regression models applied to predict monthly and seasonal sales [7].

Benefits: Higher accuracy in forecasting; ability to detect patterns not obvious from descriptive stats.

Limitation: Requires large datasets; sensitive to missing or noisy data.

8. Customer Segmentation and Behavior Analysis

K-Means, DBSCAN, and hierarchical clustering applied to segment customers by purchase frequency, order size, and location [8].

Helps businesses target promotions and optimize marketing budgets.

Limitation: Does not directly integrate profitability or supply chain data; clustering quality depends on feature selection.

9. Hybrid SQL + Python Systems

SQL used for data extraction, Python for EDA, statistical evaluation, and visualization [9].

Benefits: End-to-end analytics workflow; handles both descriptive and diagnostic analysis.

Limitation: Mostly offline analysis; limited real-time monitoring capabilities.

10. Discount Impact and Pricing Analysis

Regression and correlation models applied to measure discount effects on sales and profit [10].

Benefits: Identifies optimal discount range for maximizing revenue without sacrificing profit.

Limitation: Requires frequent updates with new transaction data; cannot predict customer response to future discounts.

E. Real-Time and Big Data Analytics

11. Real-Time Analytics Systems

Use of streaming data platforms (Kafka, Spark Streaming) for monitoring live transactions and sales KPIs [11].

Advantages: Immediate insights for promotions and supply chain adjustment.

Limitations: High computational cost; requires sophisticated infrastructure.

12. Integration of Multiple Data Sources

Combining sales, logistics, customer feedback, and product metadata for advanced analysis [12].

Advantages: Holistic view of performance; enables multi-dimensional KPIs.

Limitation: Data heterogeneity complicates integration and consistency.

F. Key Observations from Literature

SQL and BI dashboards are effective for **descriptive analytics** but insufficient for predictive decision-making.

Machine learning and hybrid Python-based systems improve **forecasting and customer segmentation**.

Real-time analytics and multi-source integration are emerging areas for **dynamic KPI monitoring**.

Most existing systems are **fragmented**, focusing on one aspect (sales, profit, or customer behavior) rather than a unified framework.

There is a need for an **end-to-end, scalable, and interactive system** integrating SQL, Python, ML, and BI tools.

III. PROPOSED SYSTEM AND METHODOLOGY

A. System Architecture Overview

The proposed E-Commerce Sales Analysis system adopts a modular and scalable architecture that integrates SQL, Python, and Power BI to deliver a comprehensive analytics framework for large-scale transactional datasets. Such multi-layered analytical architectures are widely used in modern business intelligence systems to support descriptive, diagnostic, and predictive analysis in e-commerce environments [1]. The system ensures seamless data flow across different analytical layers, enabling accurate and timely decision-making.

The SQL data layer forms the backbone of the system and is responsible for structured data extraction, cleaning, aggregation, and computation of key performance indicators. SQL queries are used to compute essential metrics such as total sales, total profit, order volume, discount impact, category-wise revenue, and delivery performance. Additionally, SQL supports time- based aggregations including monthly and seasonal trend analysis, ensuring efficient processing of large relational datasets and maintaining data consistency during analytical operations [2], [3].

The Python analytical layer enables advanced exploratory data analysis and statistical evaluation of sales data. Python libraries such as Pandas and NumPy are used for data preprocessing and transformation, while visualization libraries including

Matplotlib, Seaborn, and Plotly assist in identifying trends, correlations, and anomalies. This layer facilitates customer segmentation using clustering techniques such as K-Means, along with correlation analysis between discounts, revenue, and profit margins, thereby enhancing the analytical depth of the system [4], [5].

The visualization layer is implemented using Power BI, which transforms analytical outputs into interactive dashboards for effective data interpretation. Power BI dashboards present key business metrics through charts, graphs, maps, and slicers, allowing stakeholders to analyze performance across product categories, regions, customer segments, and time periods. Interactive visualization has been shown to significantly improve managerial insight and decision-making in e-commerce analytics [6].

An integration and workflow layer coordinates the interaction between SQL, Python, and Power BI components. This layer ensures consistency across datasets, automates the end-to-end analysis pipeline, and supports scalability for large datasets and concurrent analytical tasks. Such integrated analytical workflows are essential for maintaining reliability and performance in enterprise-level analytics systems [7].

B. System Workflow and Process Flow

The proposed E-Commerce Sales Analysis system follows a structured multi-stage workflow to ensure accurate and actionable insights. The workflow begins with data acquisition, where sales transactions, customer details, product information, discount rates, shipping types, and order timestamps are imported from CSV files or relational databases. At this stage, the dataset is validated to identify missing values, incorrect formats, and duplicate records, which is a critical step in maintaining data quality [8].

During the data preprocessing stage, missing and inconsistent values are handled using appropriate imputation or elimination techniques. Date attributes are converted into standardized datetime formats to enable time-based analysis. Shipping delays are calculated by determining the difference between actual and scheduled shipment days, and outliers in sales, profit, and discount values are identified and treated to reduce analytical bias [9].Exploratory Data Analysis (EDA) is then conducted using Python to generate descriptive statistics and visual insights. This stage includes analysis of sales distribution, profit variation, discount effectiveness, shipping performance, and regional order trends. Seasonal sales patterns and top-performing products are identified, while customer segmentation categorizes buyers into Consumer, Corporate, and Home Office groups to better understand purchasing behavior [4], [10].

Subsequently, KPI computation and aggregation are performed using SQL queries to calculate total sales, total profit, average order value, order quantity, percentage of late deliveries, and category-wise revenue and profit. Time-based aggregations further highlight monthly and seasonal performance trends, enabling effective evaluation of business growth patterns [2], [11].

The computed analytical results are visualized using interactive Power BI dashboards. These dashboards display sales and profit indicators, monthly sales trends, region-wise and category-wise performance, and shipping efficiency metrics. Interactive filters and slicers allow stakeholders to dynamically explore the dataset and derive customized insights for strategic decision-making [6], [12].

Finally, insight generation and reporting are carried out by interpreting the combined outputs from SQL and Python analysis. This stage identifies high-profit and high-volume product categories, evaluates effective discount ranges, analyzes customer segment contributions, and highlights delivery bottlenecks [13].

C. Predictive and Trend Analysis Methodology

To support proactive decision-making, the system incorporates predictive and trend analysis techniques. Time-series forecasting models such as linear regression and ARIMA are applied to historical sales data to predict future sales trends. These models assist in identifying seasonal peaks and low-demand periods, enabling improved inventory planning and marketing strategy formulation [14].

Discount impact analysis is performed using correlation and regression techniques to assess the influence of discount percentages on profitability and sales volume. This analysis helps businesses determine optimal pricing and promotional strategies that maximize revenue while maintaining profit margins [15].

Customer segmentation is further enhanced using clustering techniques such as K-Means or hierarchical clustering, which group customers based on order frequency, total spending, and product preferences. Such segmentation supports personalized marketing campaigns and improves customer retention strategies [5], [16].

Delivery performance analysis focuses on identifying late deliveries using SQL-based computations and visualizing them through dashboards. This analysis helps optimize shipping methods and logistics planning, ultimately improving customer satisfaction and service quality [17].

D. Data Validation and Reliability Assurance

Data validation is a critical component of the proposed system to ensure analytical accuracy and reliability. The system verifies dataset correctness by validating date formats, removing duplicate records, and handling null values appropriately. Consistency checks ensure that calculated KPIs match aggregated transactional data across different analytical layers [8], [18].

Furthermore, dashboard accuracy is verified by cross-validating Power BI visualizations against SQL query results. Filters and slicers are tested for correctness and responsiveness, ensuring reliable and consistent reporting for business stakeholders [12].

IV. EXPERIMENTAL SETUP AND DATA COLLECTION

The experimental evaluation of the proposed E-Commerce Sales Analysis system was designed to assess its effectiveness in extracting meaningful business insights from large-scale transactional data. The evaluation focuses on three primary research questions: how effectively the system processes and cleans real-world e-commerce datasets, how accurately it computes key performance indicators (KPIs) relevant to business decision-making, and how efficiently it visualizes analytical results to support managerial insights.

A. Dataset Construction

The experimental dataset used in this study consists of structured e-commerce sales data obtained from publicly available benchmark datasets and simulated retail transaction records commonly used in academic research [1], [2]. The dataset includes approximately 50,000 sales records spanning multiple product categories, customer segments, and geographic regions. Each record contains attributes such as order ID, order date, customer segment, product category, product price, discount percentage, sales value, profit, shipping type, scheduled delivery days, and actual delivery days.

To ensure comprehensive analysis, the dataset was organized across multiple dimensions, including temporal (month-wise and year-wise sales), categorical (product category and sub-category), geographical (region and country), and behavioral (customer segment and order frequency). The dataset also includes discount and shipping attributes, enabling analysis of promotional effectiveness and delivery performance. Table III presents representative attributes and sample records used during experimentation.

B. Development Environment

The E-Commerce Sales Analysis system was implemented and evaluated in a controlled computational environment to ensure consistency and reproducibility of results. SQL-based operations were executed using MySQL 8.0, while Python-based analysis was performed using Python version 3.11 with libraries including Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn [3]. Power BI Desktop was used for dashboard creation and visualization of analytical outcomes.

All experiments were conducted on a system configured with an Intel Core i7 processor, 16 GB RAM, and Windows 11 operating system. Data processing and visualization tasks were executed locally to maintain stable performance conditions. Logging mechanisms were implemented to record query execution times, preprocessing duration, and dashboard rendering latency to evaluate system efficiency.

C. Experimental Procedure

The experimental evaluation followed a structured multi-phase procedure conducted over a four-week period. In the first phase, baseline data validation and preprocessing were performed. Raw sales data was examined for missing values, duplicate records, and inconsistent formats. Date attributes were standardized, numerical attributes were normalized where required, and derived metrics such as delivery delay were computed.

In the second phase, SQL-based analysis was conducted to compute core business KPIs including total sales, total profit, average order value, order volume, category-wise revenue, discount impact, and percentage of late deliveries. These KPIs served as baseline indicators for business performance evaluation.

The third phase involved exploratory and statistical analysis using Python. Sales trends over time were analyzed to identify seasonal patterns and demand fluctuations. Correlation analysis was performed to evaluate the relationship between discounts and profitability. Customer segmentation was carried out using clustering techniques to identify purchasing behavior patterns across different customer groups.

In the final phase, analytical outputs were visualized using interactive Power BI dashboards. These dashboards displayed sales trends, profit distribution, regional performance, shipping efficiency, and customer segmentation insights. Stakeholders evaluated dashboard usability, clarity, and responsiveness through controlled interaction sessions.

D. Evaluation Metrics and Analysis

The system performance was evaluated using both quantitative and qualitative metrics. Quantitative evaluation included accuracy of KPI computation, execution time for SQL queries, preprocessing duration in Python, and dashboard rendering time. Qualitative evaluation focused on insight clarity, interpretability of visualizations, and usefulness of analytical results for business decision-making.

Statistical analysis of experimental results was conducted using Python libraries to compute descriptive statistics and validate consistency across analytical layers. Cross-verification between SQL outputs and Power BI visualizations ensured correctness and reliability of reported metrics. Experimental controls such as fixed dataset versions and consistent system configurations were maintained throughout the evaluation to ensure reproducible and unbiased results [4], [5].

V. RESULTS AND DISCUSSION

The results obtained from the proposed E-Commerce Sales Analysis system demonstrate its effectiveness in extracting meaningful insights from structured transactional data. Multiple analytical experiments were conducted using the prepared sales dataset to evaluate system accuracy, computational efficiency, and the usefulness of generated insights for business decision-making. The results validate the system's capability to analyze large-scale e-commerce data and transform it into actionable intelligence.

A. Sales Performance Analysis Results

The system successfully computed and analyzed key sales performance indicators including total sales, total profit, order volume, and average order value. The results revealed clear variations in sales performance across product categories and customer segments. Certain categories consistently contributed higher revenue, while others generated lower profit margins due to excessive discounting. Month-wise analysis highlighted seasonal demand fluctuations, with peak sales observed during promotional and festive periods. These findings confirm the system's ability to accurately capture temporal and categorical sales trends.

B. Discount and Profitability Evaluation

The analysis of discount strategies demonstrated a strong relationship between discount percentage and profitability. While moderate discounts led to increased sales volume, excessive discounts significantly reduced profit margins. Correlation and regression analysis revealed that higher discounts do not always translate into higher profitability, emphasizing the importance of optimized pricing strategies. These results provide valuable insights for designing data-driven promotional campaigns that balance sales growth and profit sustainability.

C. Customer Segment and Regional Insights

Customer segmentation analysis showed distinct purchasing behaviors among Consumer, Corporate, and Home Office segments. The Consumer segment contributed the highest order volume, whereas the Corporate segment generated higher average order values. Regional analysis further indicated uneven sales distribution across geographical locations, with certain regions consistently outperforming others. These insights help organizations identify high-value customer groups and focus marketing efforts on profitable regions.

D. Shipping and Delivery Performance Analysis

The evaluation of shipping performance revealed that a measurable percentage of orders were delivered later than the scheduled delivery time. Late deliveries were more frequent for specific shipping types and regions, indicating potential logistical inefficiencies. Visualization of delivery delays enabled easy identification of bottlenecks affecting customer satisfaction. These results highlight the importance of delivery performance monitoring in improving operational efficiency and customer experience.

E. Dashboard Visualization and Usability Assessment

The Power BI dashboards effectively presented complex analytical results through interactive visualizations. Key performance indicators, trend charts, category-wise comparisons, and delivery metrics were displayed in an intuitive manner. Users were able to dynamically filter data based on time period, product category, region, and customer segment, enabling rapid exploration of insights. The visual clarity and responsiveness of the dashboards enhanced interpretability and supported informed managerial decision-making.

Fig 5.1 Ecommerce Sales Dashboard

F. Discussion

The experimental results confirm that structured e-commerce sales analysis using SQL, Python, and visualization tools significantly improves business insight generation. The system performs well in identifying sales trends, profitability drivers, customer behavior patterns, and operational inefficiencies. However, the quality of insights depends on data completeness and accuracy. Missing values, inconsistent records, or limited historical data may affect analytical outcomes. Despite these limitations, the results validate the feasibility and effectiveness of the proposed analytical framework for real-world e-commerce environments.

G. Generated Analytical Output

Based on the experimental evaluation, the system successfully generated analytical outputs in the form of KPI summaries, trend graphs, category-wise and region-wise comparisons, customer segmentation reports, and shipping performance dashboards. These outputs collectively support strategic decision-making related to pricing, inventory planning, marketing optimization, and logistics management. The results demonstrate that the proposed E-Commerce Sales Analysis system provides a reliable and scalable solution for modern data-driven retail businesses.

VI. CONCLUSION AND FUTURE SCOPE

This paper presented the design and implementation of an E-Commerce Sales Analysis system aimed at extracting actionable business insights from large-scale transactional data. By integrating structured query processing using SQL, advanced data analysis with Python, and interactive visualization through Power BI, the proposed system provides a comprehensive framework for analyzing sales performance, customer behavior, discount effectiveness, and delivery efficiency. The system enables organizations to transform raw sales data into meaningful indicators that support informed strategic decision-making.

The experimental results demonstrate that the proposed approach effectively identifies sales trends, profitable product categories, high-value customer segments, and operational bottlenecks related to logistics and shipping. Time-based analysis revealed seasonal demand patterns, while correlation analysis highlighted the impact of discounts on profitability. The use of interactive dashboards improved data interpretability and enabled stakeholders to dynamically explore key performance metrics across different dimensions such as region, category, and customer segment. Overall, the system enhances business visibility and supports data-driven decision processes in modern e-commerce environments.

Although the proposed system delivers reliable and insightful analytical outcomes, there remains scope for further enhancement. Future work may incorporate predictive analytics and machine learning models to forecast sales demand, customer churn, and inventory requirements. Real-time data processing using streaming frameworks can further improve responsiveness and support dynamic decision-making. Additionally, integrating unstructured data sources such as customer reviews, feedback, and social media sentiment can enhance understanding of customer preferences and market trends.

Further extensions may include the deployment of advanced recommendation systems for personalized product suggestions and dynamic pricing strategies based on customer behavior. Scalability can be improved by adopting cloud-based architectures and distributed data processing frameworks. Incorporating automated data quality assessment and anomaly detection mechanisms can also enhance system robustness. With these enhancements, the proposed E-Commerce Sales Analysis framework can evolve into a fully intelligent business intelligence system capable of supporting strategic, tactical, and operational decisions in real-world e-commerce platforms.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to the project guide for her valuable guidance, constant encouragement, and insightful suggestions throughout the development of this project and research work. The author is also thankful to the faculty members of the Department of Information Technology for their support and cooperation. Special thanks are extended to the institution for providing the necessary resources and environment to successfully complete this work.

REFERENCES

[1] Laudon, K. C., & Traver, C. G., *E-Commerce: Business, Technology, Society*, 16th ed., Pearson Education, 2020.

[2] Turban, E., Outland, J., King, D., Lee, J. K., Liang, T.-P., & Turban, D. C., *Electronic Commerce 2018: A Managerial and Social Networks Perspective*, Springer, 2018.

[3] Kim, Y., Kim, S., & Lee, J., "An SQL-based analytical approach for large-scale e-commerce transaction data," *International Journal of Database Management Systems*, vol. 10, no. 2, pp. 1–15, 2018.

[4] Han, J., Kamber, M., & Pei, J., *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2012.

[5] Chen, H., Chiang, R. H. L., & Storey, V. C., "Business intelligence and analytics: From big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012.

[6] Few, S., *Information Dashboard Design: Displaying Data for At-a-Glance Monitoring*, 2nd ed., Analytics Press, 2013.

[7] Hastie, T., Tibshirani, R., & Friedman, J., *The Elements of Statistical Learning*, 2nd ed., Springer, 2009.

[8] Provost, F., & Fawcett, T., *Data Science for Business*, O'Reilly Media, 2013.

[9] Waller, M. A., & Fawcett, S. E., "Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management," *Journal of Business Logistics*, vol. 34, no. 2, pp. 77–84, 2013.

[10] Gandomi, A., & Haider, M., "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.

[11] Kreps, J., Narkhede, N., & Rao, J., "Kafka: A distributed messaging system for log processing," *Proc. NetDB*, pp. 1–7, 2011.

[12] Microsoft, *Power BI Documentation*, Microsoft Corporation,2023.[Online].Available:https://learn.microsoft.com/power-bi

[13] Chopra, S., & Meindl, P., *Supply Chain Management: Strategy, Planning, and Operation*, 7th ed., Pearson, 2019.