African Journal of Academic Publishing in Science and Technology (AJAPST)

Volume 2, Issue 3, 2026
Page No: 22-36
Website: https://easrjournals.com/index.php/AJAPST/index

Exploratory Analysis of Daily Sales, Costs, And Profitability In a Community Pharmacy: A 14-Month Retrospective Study

Asma Agaal[1*], Mansour Essgaer[2], Llahm Ben Dalla[3], Abdulgader Alsharif [4]

[1] Artificial Intelligence Department, Faculty of Technical Sciences, Sabha, Libya.
[2] Artificial Intelligence Department, Faculty of information Technology, University Sebha, Libya
[3] Department of Electric Electronics Engineering, Beyezit University, Ankara Yildirim, Turkiye
[4] Department of Electrical and Electronic Engineering, College of Technical Sciences-Sabha, Sabha, Libya

تحليل استكشافي للمبيعات اليومية والتكاليف والربحية في صيدلية مجتمعية: دراسة استرجاعية لمدة 14 شهرًا

اسمه اعجال[1*]، منصور الصغير[2], للاهم بن دله [3], عبد القادر الشريف[4]

[1] قسم الذكاء الاصطناعي، كلية العلوم التقنية سبها، سبها، ليبيا.
[2] قسم الذكاء الاصطناعي، كلية تقنية المعلومات جامعة سبها، سبها، ليبيا
[3] قسم هندسة الإلكترونيات الكهربائية، جامعة بايزيد، أنقرة يلدريم، تركيا
[4] قسم الهندسة الكهربائية والإلكترونية، كلية العلوم التقنية - سبها، سبها، ليبيا.

[*]Corresponding author: asma.agaal@sebhau.edu.ly

Received: *June 20, 2026*	Accepted: July *26,2026*	Published: August 23, 2026

Abstract:
This study explored daily financial activity in a community pharmacy located in Sabha, Libya, using transaction records collected from September 2022 to October 2023. Exploratory data analysis included descriptive statistics, distribution plots, boxplots, calendar heatmaps, monthly comparisons, chronological plots, and correlation analysis. The initial dataset contained 428 records, of which 396 were retained following preprocessing. Profits, total sales, and total average cost of sales were positively skewed and displayed substantial daily variation. Strong positive correlations were observed between sales and profits (r = 0.86), sales and average cost of sales (r = 0.99), and profits and average cost of sales (r = 0.83). Higher financial activity occurred more frequently toward the end of the working week, although the limited observation period did not permit firm conclusions regarding recurring annual seasonality. The analysis also identified an exceptionally high profit value that required verification against the original records. Overall, the findings demonstrate how exploratory analysis of pharmacy transaction data can support financial monitoring, data-quality improvement, inventory planning, and operational decision-making in community pharmacies.

Keywords: community pharmacy; exploratory data analysis; pharmaceutical sales; profitability; cost of sales; transaction data; retail pharmacy, Libya, Sabha

الملخص
استكشفت هذه الدراسة النشاط المالي اليومي في صيدلية مجتمعية بمدينة سبها الليبية، باستخدام سجلات المعاملات التي جُمعت خلال الفترة من سبتمبر 2022 إلى أكتوبر 2023. شمل تحليل البيانات الاستكشافي الإحصاءات الوصفية، ومخططات التوزيع، ومخططات الصندوق، وخرائط الحرارة التقويمية، والمقارنات الشهرية، والمخططات الزمنية، وتحليل الارتباط. احتوت مجموعة البيانات الأولية على 428 سجلاً، احتُفظ منها بـ 396 سجلاً بعد المعالجة المسبقة. اتسمت الأرباح وإجمالي المبيعات ومتوسط تكلفة المبيعات بانحراف إيجابي، وأظهرت تباينًا يوميًا كبيرًا. لوحظت ارتباطات إيجابية قوية بين المبيعات والأرباح (r = 0.86)، وبين المبيعات ومتوسط تكلفة المبيعات (r = 0.99)، وبين الأرباح ومتوسط تكلفة المبيعات (r = 0.83). لوحظ ارتفاع النشاط المالي بشكل متكرر قرب نهاية أسبوع العمل، على الرغم من أن فترة الملاحظة المحدودة لم تسمح باستخلاص استنتاجات قاطعة بشأن الموسمية السنوية المتكررة. كما كشف التحليل عن قيمة ربح مرتفعة بشكل استثنائي، مما استدعى التحقق منها

بالرجوع إلى السجلات الأصلية. بشكل عام، توضح النتائج كيف يمكن للتحليل الاستكشافي لبيانات معاملات الصيدليات أن يدعم المراقبة المالية، وتحسين جودة البيانات، وتخطيط المخزون، واتخاذ القرارات التشغيلية في صيدليات المجتمع.

الكلمات الدالة: صيدلية مجتمعية؛ تحليل البيانات الاستكشافي؛ مبيعات الأدوية؛ الربحية؛ تكلفة المبيعات؛ بيانات المعاملات؛ صيدلية بيع بالتجزئة، ليبيا، سبها.

1- Introduction

Community pharmacies operate at the intersection of healthcare delivery and retail business. In addition to dispensing medicines, they provide medication-management, preventive-care, and health-promotion services, supported by frequent contact with patients [1]. Maintaining these services depends partly on effective financial and operational management, particularly the management of sales, acquisition costs, discounts, inventory, and profits [2].

Digital point-of-sale systems generate transaction data that can support evidence-based pharmacy management. Previous retail and operations-management research has shown that such data can inform decisions concerning inventory, pricing, resource allocation, and business performance [3][4]. However, aggregated totals and simple averages may conceal short-term variation, skewed distributions, unusual transactions, and calendar-related patterns. Exploratory data analysis (EDA), which combines descriptive statistics with graphical techniques, provides an appropriate framework for identifying these characteristics before confirmatory or predictive analyses are considered [5].

Although retail analytics has received considerable research attention, much of the literature emphasizes forecasting and optimization [6][7]. Comparatively limited evidence describes the joint daily behavior of sales, costs, discounts, and profits in individual community pharmacies, particularly in Libya. Context-specific exploratory studies are therefore needed to clarify how routinely collected pharmacy data can support operational decisions while identifying possible data-quality problems [8].

Accordingly, this study conducts a retrospective exploratory analysis of daily transaction records collected from a community pharmacy in Sabha, Libya, between September 2022 and October 2023. It aims to characterize the distributions and variability of the principal financial variables, identify unusual observations, examine temporal and calendar-related variation, and evaluate relationships among sales, costs, discounts, and profits. The findings are intended to support financial monitoring, inventory planning, and data-quality improvement within the study pharmacy.

The remainder of the paper is organized as follows: Section II reviews the relevant literature; Section III describes the data and exploratory methods; Section IV presents and discusses the findings; and Section V provides the conclusion, limitations, and directions for future research.

2- Literature review

2.1 Community Pharmacy as a Healthcare and Business Setting

Community pharmacies occupy a distinctive position because they combine healthcare delivery with retail and financial operations. In addition to dispensing medicines, community pharmacists increasingly provide medication management, vaccination, preventive care, health education, and support for chronic conditions. Berenbrok et al. [9] demonstrated the accessibility of community pharmacies by showing that Medicare beneficiaries visited community pharmacies more frequently than they encountered primary-care physicians. However, their study focused on healthcare encounters rather than the financial and operational performance of pharmacies.

Similarly, Goode et al. [10] reviewed the expansion of community pharmacy from a product-centred dispensing environment to a broader patient-care setting. The authors highlighted the need to redesign pharmacy workflows, technologies, staffing arrangements, and business models to support this expanded role. Nevertheless, their review did not empirically examine routinely generated financial indicators such as sales revenue, purchasing costs, discounts, or profits.

The ability of a community pharmacy to sustain its healthcare role depends partly on its operational and financial viability. Pharmacy performance is influenced by the relationship between sales revenue, medicine acquisition costs, discounts, transaction volumes, and profit margins [11]. These indicators should not be considered independently because changes in one measure can influence the interpretation of another. For example, an increase in total sales does not necessarily indicate improved profitability when the cost of sold products or the value of discounts increases simultaneously. A multidimensional examination of pharmacy transaction records is therefore necessary to understand the underlying structure of daily business activity [12].

2.2 Data Analytics in Retail Operations

The broader retail literature provides evidence that transaction data can support more informed operational decisions. Fisher and Raman [13] explained how retailers can use routinely collected data to improve inventory management, product assortment, pricing, and overall performance. Their work demonstrated the managerial value of retail data but was based on examples from general retail settings. Its conclusions cannot be transferred directly to community pharmacies without considering medicine expiry, product criticality, regulatory requirements, purchasing-price variation, and the healthcare responsibilities of pharmacists.

Choi et al. [14]reviewed the role of data analytics in operations management and identified applications in inventory management, revenue management, marketing, transportation, supply-chain coordination, and risk analysis. Their review established that operational data can support both descriptive understanding and advanced decision-making. However, much of the literature examined in that review focused on large datasets, predictive analytics, and optimization. Comparatively less attention was given to how smaller; single-site organizations can use exploratory techniques to understand the distributions, relationships, and temporal characteristics of their routinely recorded data.

Another important consideration is that electronically stored operational data may contain errors or inconsistencies. DeHoratius and Raman [15] examined approximately 370,000 inventory records from 37 retail stores and found that 65% were inaccurate. They also showed that inventory-record accuracy differed across stores and product categories. Although their study did not involve community pharmacies, its findings demonstrate that operational databases should not automatically be treated as complete or error-free. Duplicate dates, missing observations, implausible values, and unusual transactions must therefore be examined before financial or managerial conclusions are developed.

2.3 Pharmacy Inventory and Cost Analysis

Previous pharmacy-management studies have frequently concentrated on inventory classification and expenditure control. Devnani et al.[16] , for example, combined ABC and vital–essential–desirable analyses to evaluate medicines held by a tertiary-care pharmacy store. ABC analysis classified medicines according to their contribution to total expenditure, whereas vital–essential–desirable analysis classified them according to their clinical importance. The combined approach enabled managers to identify high-cost and clinically important medicines requiring stricter control.

Although such studies demonstrate the value of analysing pharmacy expenditure, their scope differs from the present investigation. First, they commonly focus on individual medicines or annual inventory expenditure rather than aggregated daily activity. Second, they are frequently conducted in hospital or institutional pharmacies rather than community-pharmacy settings [17]. Third, they aim to classify inventory items instead of examining the joint behaviour of sales, costs, discounts, and profits.

Inventory-based analysis alone may therefore provide an incomplete representation of community-pharmacy operations. Daily profitability can be affected by the total value of sales, the average cost associated with sold items, the most recently recorded purchase price, and discounts applied to transactions. Examining these indicators jointly may reveal whether increases in sales are accompanied by proportional changes in costs and profits. It may also help distinguish genuine high-value business activity from potential recording errors[17].

2.4 Exploratory Analysis of Transaction Data

Exploratory data analysis provides a structured framework for identifying the principal characteristics of operational datasets. The NIST framework [18] describes EDA as a combination of numerical and graphical techniques used to examine distributions, detect unusual observations, evaluate assumptions, identify important variables, and reveal previously unrecognized patterns. Such an approach is particularly relevant to pharmacy financial data because sales, costs, discounts, and profits may demonstrate skewed distributions and may be strongly influenced by a small number of high-value transactions.

Baillie et al. distinguished initial data analysis from exploratory data analysis. Initial analysis involves examining metadata, data completeness, coding consistency, missingness, and implausible values to ensure that the dataset is suitable for further investigation [19]. Exploratory analysis subsequently examines substantive patterns and relationships. The authors emphasized that preprocessing decisions should be transparent, reproducible, and informed by the context in which the data were generated.

This distinction is important when analyzing pharmacy transaction records. Missing values or extreme observations should not be modified or removed without considering their operational meaning. A statistically extreme profit or sales value might represent an extraction error, but it could also correspond to a legitimate bulk transaction, a high-cost medicine, a promotional period, or an unusual purchasing event. Consequently, statistical outlier criteria should be used to flag observations for investigation rather than as automatic deletion rules [20].

Temporal organization adds another dimension to exploratory analysis. Hyndman and Athanasopoulos [21] recommend plotting time-indexed data before undertaking more complex analyses because chronological visualizations can reveal trends, seasonal patterns, changes in variability, structural changes, and unusual observations. Daily transaction records can be grouped by calendar variables such as day, month, and year to investigate whether financial activity changes systematically over time. Time plots, rolling summaries, monthly comparisons, autocorrelation plots, and seasonal subseries plots can therefore complement conventional descriptive statistics.

However, temporal patterns should be interpreted carefully. A difference between months does not necessarily represent stable seasonality, particularly when the observation period covers only slightly more than one year [22]. Similarly, correlation among financial indicators does not demonstrate causation. Relationships involving profit may partly reflect mathematical dependence because profit is calculated from sales and costs. Exploratory findings should therefore be presented as descriptive evidence that generates operational insights and hypotheses for future investigation [23].

2.5 Research Gap and Study Contribution

The literature relevant to community-pharmacy data can be classified into four main areas. Pharmacy-practice studies emphasize accessibility, clinical services, and the changing professional role of pharmacists [24]. Retail-analytics research demonstrates the value of transaction data but is not specific to the regulatory and operational characteristics of pharmacies. Pharmacy inventory studies focus primarily on medicine classification and expenditure control, often in institutional settings. Finally, methodological literature provides guidance for data-quality assessment and exploratory analysis but does not directly describe community-pharmacy financial behaviour [25].

These areas leave a need for context-specific studies that jointly examine the daily financial indicators generated by community pharmacies. Existing research commonly focuses on individual inventory items, aggregated expenditure, clinical services, or predictive modelling. Consequently, the relationships among daily total sales, profits, total discounts, average cost of sales, and sales values based on the latest purchase price remain insufficiently described at the individual community-pharmacy level.

The present study addresses this gap by conducting a retrospective exploratory analysis of daily pharmacy transaction data. The analysis integrates data-quality assessment, robust descriptive statistics, distributional analysis, correlation analysis, and temporal visualization. The year, month, and day variables are used to construct a chronological index and investigate calendar-related variation. Rather than developing predictive models, the study aims to provide a transparent and interpretable description of the distributions, unusual observations, temporal patterns, and relationships within the recorded financial indicators. The findings may provide a foundation for improved cost monitoring, discount assessment, purchasing decisions, and future multi-site investigations.

---

3. Methodology

3.1 Study Design and Setting

This study adopted a retrospective, quantitative, and exploratory design to examine daily pharmaceutical sales transactions recorded by a community pharmacy. Exploratory data analysis (EDA) was used to characterize the distributions of sales, costs, discounts, and profits; identify unusual observations; investigate temporal variations; and examine relationships among the financial variables. The study was descriptive and did not involve forecasting, predictive modeling, or causal inference.

The data covered the period from September 1, 2022, to October 31, 2023 in a community pharmacy located in Sabha, Libya, . This period provided approximately 14 months of transaction records, allowing daily, monthly, and year-specific patterns to be examined. To preserve commercial confidentiality, the identity and exact location of the pharmacy were not disclosed.

3.2 Data Source and Unit of Analysis

The data were obtained retrospectively from the pharmacy’s electronic sales-recording system. The original database contained 428 records. Each record represented the aggregated sales activity recorded for a particular day. No patient-level, prescription-level, or personally identifiable information was included.

The primary unit of analysis was therefore the pharmacy-day. The financial values were reported in Dainar Libya. Before analysis, the year, month, and day variables were combined to construct a complete calendar-date variable. The constructed dates were used to arrange the observations chronologically and examine temporal patterns.

Following data-quality assessment and preprocessing, 396 observations were included in the final exploratory analysis. The number of excluded observations and the reasons for their exclusion were documented to support transparency and reproducibility.

3.3 Study Variables

The dataset contained five continuous financial variables and three categorical calendar variables. The definitions and measurement types of the variables are presented in Table 1.

Table 1. Description of the study variables

Variable	Description	Type
Profits	Profit generated from pharmaceutical sales during the recorded period	Numerical (continuous)
Total last purchase sales price	Total value calculated using the most recent purchase price of the products sold	Numerical (continuous)
Total average cost of sales	Total cost of the products sold, calculated using their average acquisition cost	Numerical (continuous)
Total discounts	Total monetary value of discounts applied to sales	Numerical (continuous)
Total sales	Total revenue generated from pharmaceutical sales	Numerical (continuous)
Year	Calendar year of the sales record	Categorical (nominal)
Month	Calendar month of the sales record	Categorical (nominal)
Day	Calendar day of the sales record	Categorical (nominal)

The calendar variables were treated as categorical when comparing sales activity across years, months, or days. However, they were also combined into a chronological date variable for temporal visualization.

When total sales were greater than zero, the daily profit margin was calculated as [26]:

$$\textbf{Profit=Daily profit}\backslash\textbf{Total daily sales*100}$$

Profit margin was used as a descriptive measure of profitability relative to sales revenue. Because profit may be calculated from sales and cost information, relationships involving profit were interpreted as accounting associations rather than independent causal relationships.

3.4 Data Quality Assessment and Preprocessing

Data preprocessing was performed before the statistical analysis. The assessment addressed duplicate records, invalid dates, missing values, inconsistent data types, impossible financial values, and potentially influential observations.

First, the year, month, and day fields were converted into a valid date format. The resulting dates were checked for chronological consistency, repeated dates, missing dates, and observations outside the specified study period. When more than one record represented the same date, the source records were reviewed to determine whether they were genuine separate entries or accidental duplicates. Genuine records for the same day were aggregated only when all financial variables shared the same measurement basis [27].

Second, all financial variables were converted to numerical format. Currency symbols, commas, spaces, and other non-numerical characters were removed where necessary. Values were checked for negative sales, costs, or discounts. Any value that was inconsistent with the accounting definition of its variable was verified against the original record before correction or exclusion[28].

Third, missingness was calculated for each variable and for the dataset overall. Missing values were not automatically replaced with the mean because mean imputation can artificially reduce variability and alter correlations. When the proportion of missing observations was small, an available-case approach was used, meaning that each analysis included records with valid values for the variables involved. The number of observations contributing to each analysis was reported where it differed from the final sample size [29].

Potential outliers were initially identified using boxplots and the interquartile-range rule [24]:

$$\textbf{Lower boundary} = Q_1 - 1.5(IQR)$$
$$\textbf{Upper boundary} = Q_3 + 1.5(IQR)$$

where (Q_1) and (Q_3) represent the first and third quartiles, respectively, and (IQR=Q_3-Q_1). Observations outside these boundaries were treated as potentially unusual rather than automatically removed. Each flagged value was reviewed in relation to its date and the other financial variables. An observation was corrected or excluded only when it could reasonably be attributed to duplication, data-entry error, an invalid date, or another documented recording problem. Legitimate high-sales or high-profit days were retained because they represented meaningful operational variation.

A sensitivity analysis was conducted, where appropriate, by repeating key descriptive summaries and correlations with and without highly influential observations. This made it possible to determine whether the main findings were driven by a small number of extreme values.

3.5 Exploratory Data Analysis

The exploratory analysis consisted of univariate, bivariate, and temporal analyses.

3.5.1 Univariate Analysis

The distribution of each continuous variable was summarized using the number of valid observations, mean, standard deviation, median, interquartile range, minimum, maximum, and quartiles. The median and interquartile range were emphasized when distributions were strongly skewed. Histograms, density plots, and boxplots were used to assess distributional shape, variability, skewness, and extreme observations. For variables with highly right-skewed distributions, logarithmic visualizations were considered to improve graphical interpretation. Any transformation was used only for visualization and did not replace the original monetary values in the descriptive tables. Frequencies and percentages were calculated for the year, month, and day variables. The completeness of daily observations within the study period was also examined [30].

3.5.2 Bivariate Analysis

Relationships among profits, total sales, total discounts, total average cost of sales, and total last purchase sales price were examined using scatterplots and correlation analysis. Pearson’s correlation coefficient was considered when two variables displayed an approximately linear relationship without substantial skewness or influential outliers. Spearman’s rank correlation coefficient was used as the principal measure when distributions were skewed, relationships were monotonic but not linear, or extreme observations were present. Correlation coefficients were presented using a correlation matrix and heatmap. The strength and direction of the relationships were interpreted descriptively. Correlations were not interpreted as evidence that one financial variable caused changes in another. Particular caution was exercised when interpreting correlations between profit, sales, and costs because these variables may be linked through the pharmacy’s accounting calculations [31].

3.5.3 Temporal Analysis

Daily values of total sales, profits, costs, and discounts were plotted chronologically to identify changes, clusters of high or low activity, gaps, and unusual transaction days. Monthly summaries were then calculated to reduce short-term daily variation and provide a clearer description of changes over the observation period. For each calendar month, the following measures were calculated where applicable:

- number of recorded operating days;
- total sales;
- average and median daily sales;
- total profit;
- average and median daily profit;
- total cost of sales;
- total discounts; and
- average profit margin.

Monthly totals were interpreted alongside the number of recorded operating days because months with more observations could naturally produce larger totals. Average and median daily values were therefore used when comparing months with unequal numbers of recorded days [32].The study period covered approximately 14 months. Consequently, recurring annual seasonality could not be established reliably from this dataset. Month-to-month patterns were reported as patterns observed during the study period rather than as evidence of stable seasonal behavior.

3.6 Data Presentation

The results were presented using descriptive tables and figures. Summary-statistics tables were used to report the central tendency and dispersion of the continuous variables. Histograms and boxplots described their distributions, while scatterplots and a correlation heatmap illustrated relationships among the financial measures. Line charts were used to display daily and monthly changes in sales, costs, discounts, and profits. Bar charts were used to compare monthly and yearly summaries. All figures included clearly labeled axes, measurement units, observation periods, and explanatory captions [33].

3.7 Statistical Software and Reproducibility

Data preparation, statistical analysis, and visualization were performed using software version Python 3.13. The principal packages used were pandas, NumPy, Matplotlib, Seaborn, and SciPy.A reproducible analytical workflow was maintained by preserving the original data separately from the cleaned analytical dataset. All cleaning decisions, excluded records, derived variables, and graphical procedures were documented. A fixed significance level was not emphasized because the study was primarily exploratory and descriptive [34].

4. Result and Discussion

This section presents the findings of the exploratory analysis of the pharmacy's daily financial activity. The analysis focuses on data quality, descriptive distributions, temporal variation, and relationships among profits, total sales, and total average cost of sales.

4.1 Data Quality and Analytical Sample

The original dataset contained 428 records covering the period from September 1, 2022, to October 31, 2023. Approximately 2.3% of the data points were missing, principally in the profits, total average cost of sales, and total sales variables. Following the preprocessing procedures described in Section III, including missing-value treatment and screening of potentially influential observations using the interquartile-range criterion, 396 records were retained in the cleaned analytical dataset.

The exploratory assessment identified several unusually high financial values. These observations were examined using histograms and boxplots before subsequent analyses were conducted. Extreme observations were interpreted cautiously because they could represent genuine high-activity days, data-entry errors, inconsistent measurement units, or accounting adjustments.

The maximum recorded profit was particularly unusual relative to the distributions of profits and sales. Consequently, results based on the mean and standard deviation were interpreted alongside the median and interquartile range, which are less sensitive to extreme observations.

4.2 Descriptive Statistics

Table 2 presents the descriptive statistics for the three principal financial variables before the exclusion of potentially influential observations. Year, month, and day were not summarized using means and standard deviations because they were used primarily as calendar identifiers.

Table 2. Descriptive statistics of the principal daily financial variables

Variable	N	Mean	SD	Minimum	Q1	Median	Q3	Maximum
Profits	428	322.75	5,019.70	0.00	40.11	66.46	105.89	103,921.24
Total average cost of sales	428	174.14	119.59	5.35	90.51	155.79	219.72	978.84
Total sales	428	257.23	171.99	15.00	133.75	233.25	328.12	1,369.75

Note: SD = standard deviation; Q1 = first quartile; Q3 = third quartile. Monetary values are expressed in Dinar Libya.

Daily profit displayed the greatest variability. Its mean was 322.75, whereas its median was only 66.46. The considerable difference between the mean and median, together with the standard deviation of 5,019.70, indicates that the arithmetic mean was strongly influenced by an extreme upper observation. The middle 50% of profit observations ranged from 40.11 to 105.89, suggesting that profit on a typical day was substantially lower than the reported mean.

The maximum profit of 103,921.24 was more than 1,500 times the median and substantially exceeded the maximum recorded daily sales value of 1,369.75. If profits and sales were measured for the same reporting period and in the same currency, such a result would be inconsistent with a conventional calculation of daily profit. This observation may represent a data-entry error, a different unit of measurement, an accounting adjustment, or the aggregation of transactions from different periods. It should therefore be verified against the original pharmacy

record. Until it is reconciled, the median and interquartile range provide more reliable descriptions of routine profitability than the mean and standard deviation.
Total sales showed less extreme variation. Mean daily sales were 257.23, while median daily sales were 233.25. The central 50% of sales observations ranged from 133.75 to 328.12. The difference between the mean and median indicates moderate positive skewness, with a smaller number of high-sales days increasing the mean.
Total average cost of sales had a mean of 174.14 and a median of 155.79. Its middle 50% ranged from 90.51 to 219.72. The relatively modest difference between the mean and median suggests that the central cost distribution was more stable than the profit distribution, although the maximum of 978.84 indicates that several high-cost days occurred.
The mean and median values of total sales were higher than the corresponding cost values. This difference is consistent with the operation of a business that generates a positive gross margin. However, it should not be interpreted as net profitability because other expenses, such as salaries, rent, utilities, taxes, wastage, and administrative costs, were not included in the dataset.

4.3 Distributional Characteristics and Influential Observations

The histograms and boxplots demonstrated that profits, total sales, and total average cost of sales were positively skewed (Figure 1). Most observations were concentrated at lower or moderate values, while a smaller number of observations formed extended upper tails.
Profit exhibited the most pronounced skewness and contained one exceptionally influential upper value. Total sales and total average cost of sales also contained observations above the upper boxplot whiskers, but these were considerably less extreme than the maximum profit value.
The distributional findings demonstrate the importance of presenting robust summary measures. In highly skewed financial data, the mean may not represent the experience of a typical operating day. Therefore, the median and interquartile range should be considered alongside the mean and standard deviation when evaluating pharmacy performance.
Observations beyond 1.5 times the interquartile range are statistical outliers but are not necessarily erroneous. A high-sales day may represent a genuine increase in demand, a bulk purchase, a public holiday effect, or another operational event.

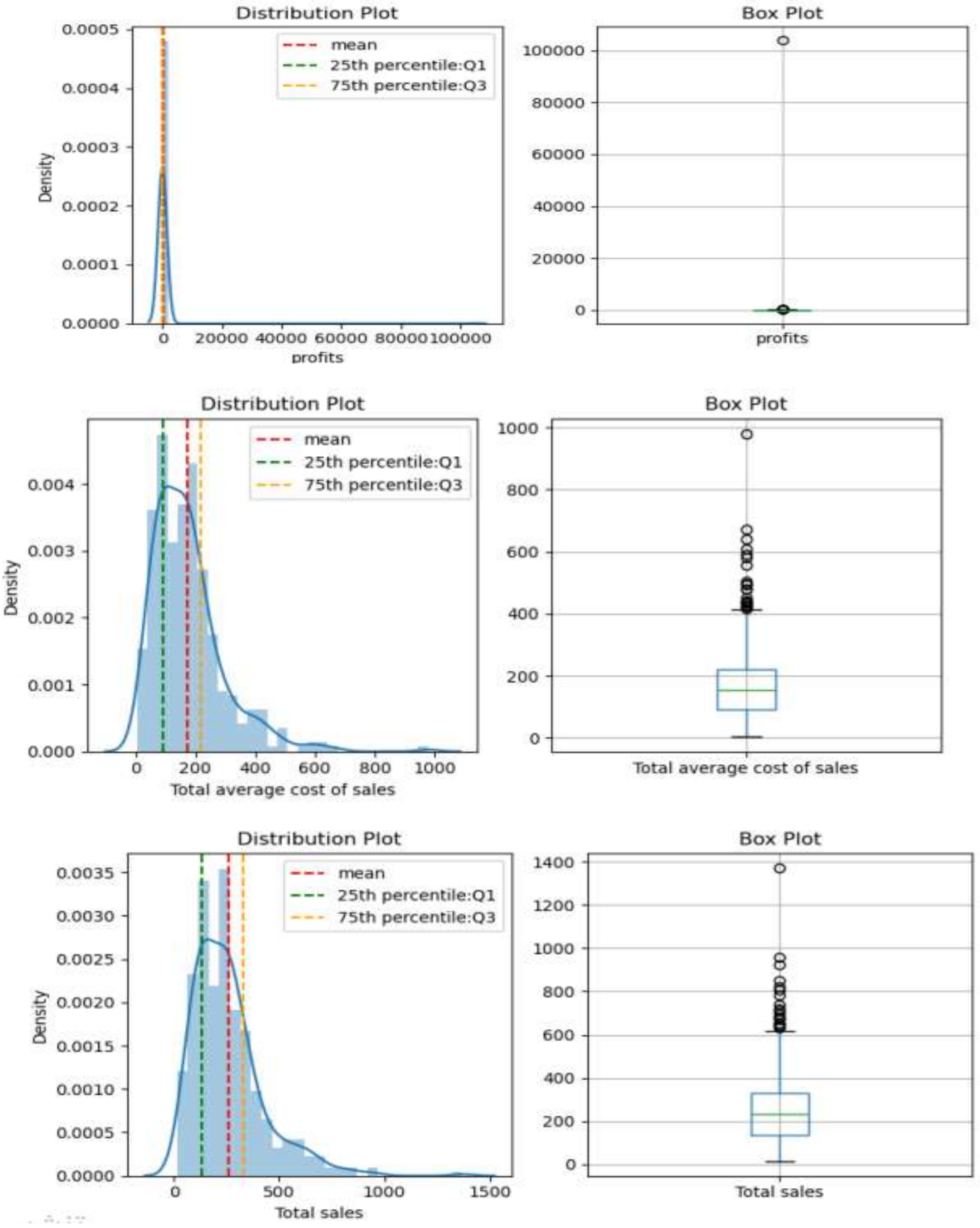

Figure 1. Distributions and boxplots of the principal daily financial variables before outlier treatment: profits, total average cost of sales, and total sales

4.4 Calendar Patterns

Calendar heatmaps were used to examine the distribution of daily profits, sales, and costs throughout the study period (Figure 2). The visualizations suggested that higher sales and profits occurred more frequently toward the end of the working week, particularly from Thursday to Saturday. Lower values appeared more frequently on Mondays and Tuesdays.

Total average cost of sales followed a broadly similar weekly pattern, although the differences were less pronounced than those observed for sales and profits. This correspondence is operationally reasonable because an increase in sales activity generally increases the value of the products sold.

Several factors may explain the apparent end-of-week increase, including customer purchasing routines, prescription-refill patterns, pharmacy opening schedules, and changes in demand before weekends. However, the dataset did not include customer counts, opening hours, prescription information, product availability, or holiday indicators. Therefore, the analysis could identify the temporal concentration of high values but could not determine its underlying cause.

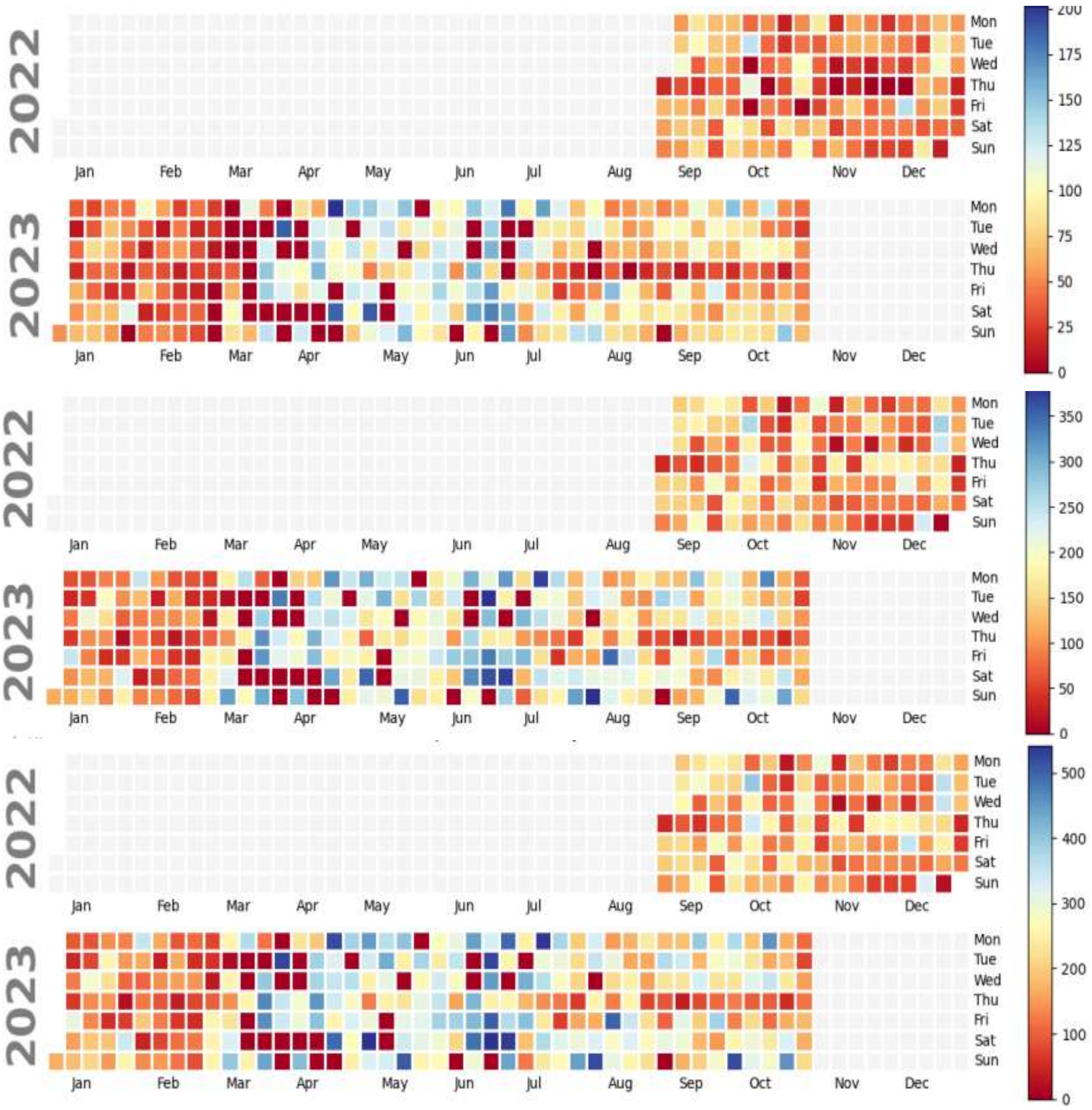

Figure 2. Calendar heatmaps of daily profits, total average cost of sales, and total sales during the study period

4.5 Monthly Variation

The monthly boxplots demonstrated differences in both the central tendency and variability of the three financial measures (Figure 3). Profit medians were comparatively higher during March and several mid-year months, including May, June, and July. April displayed a noticeable reduction in median profit. Lower profit medians were also observed during some later months, although the extent of these differences varied between the calendar heatmaps and monthly boxplots.

Total sales demonstrated a pattern broadly similar to that of profits. Higher sales medians and wider interquartile ranges were observed during parts of January–March and May–July, while April displayed a distinct reduction.

The wider interquartile ranges during the higher-sales months indicate that daily activity was not only greater but also more variable.

Monthly variation in total average cost of sales was less pronounced. January, February, and March generally showed higher median costs and somewhat wider distributions, whereas April and several later months showed lower median values. The alignment between higher sales and higher costs supports the expectation that the cost of products sold increases with pharmacy activity.

These monthly differences should not be described as established annual seasonality. The dataset covered approximately 14 months and did not contain enough complete annual cycles to establish that the same patterns recur from year to year. The observed differences could also reflect unequal numbers of recorded operating days, public holidays, medicine availability, local disease patterns, or economic conditions.

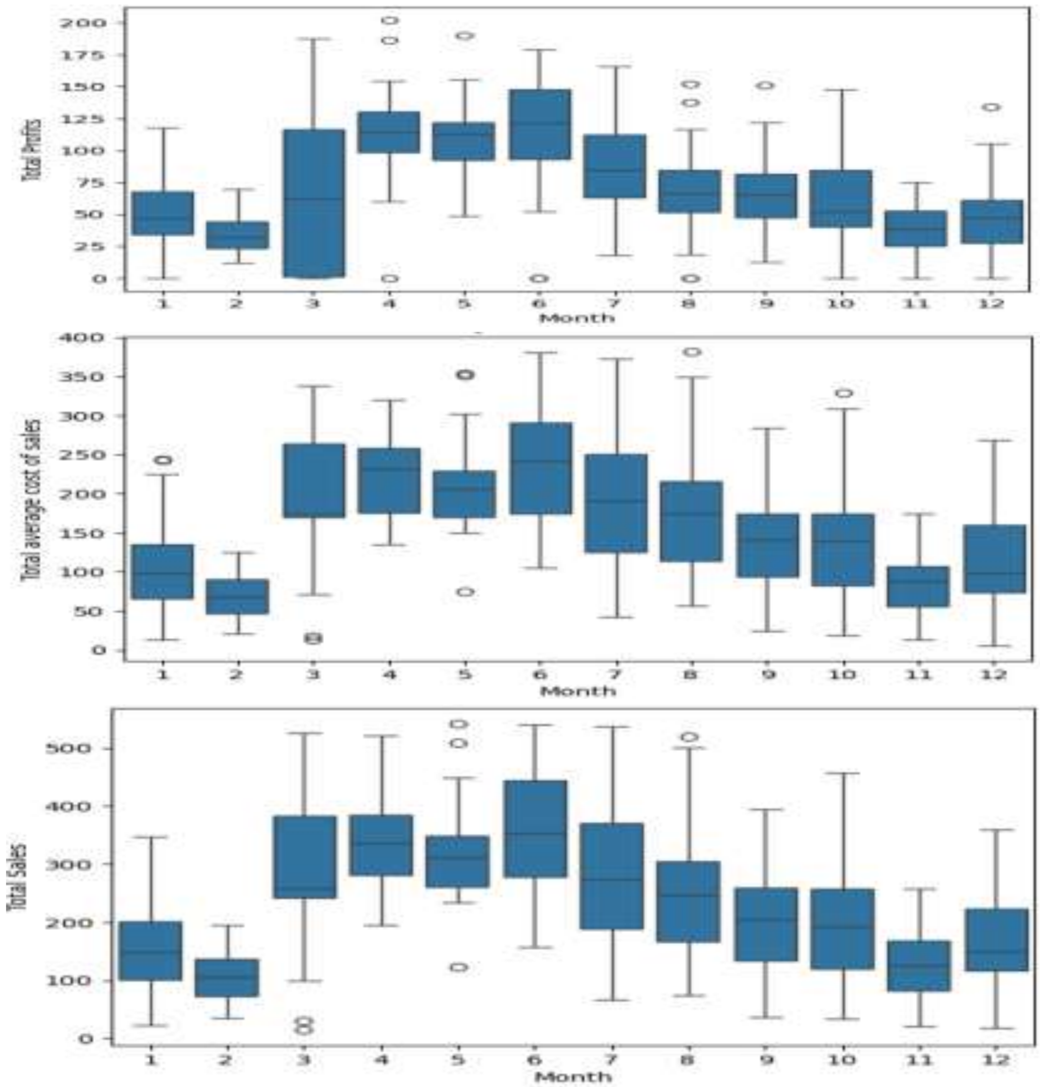

Figure 3. Monthly boxplots of daily profits, total average cost of sales, and total sales in the cleaned analytical dataset

4.6 Relationships Among Sales, Costs, and Profits

The correlation analysis demonstrated strong positive relationships among the three principal financial variables (Figure 4). Total sales and profits had a correlation coefficient of 0.86, indicating that days with higher sales were generally associated with higher profits.

The strongest relationship was observed between total sales and total average cost of sales, with a correlation coefficient of 0.99. This near-proportional relationship suggests that increases in sales were accompanied by corresponding increases in the cost of the products sold. While this relationship is financially plausible, its magnitude may also indicate that the two variables were derived from closely related accounting calculations. Their definitions and calculation procedures should therefore be clearly reported.

Profits and total average cost of sales had a correlation coefficient of 0.83. This positive association should not be interpreted as evidence that higher costs directly produce higher profits. A more plausible explanation is that sales activity acts as a common underlying factor. High-activity days generate greater revenue, require a greater value of products to be sold, and consequently generate higher gross profit.

The scatterplot of total sales against profits supported the correlation results by showing a clear upward relationship. The color gradient representing total average cost of sales indicated that costs generally increased with sales. However, some variation in profits remained among observations with similar sales values, possibly reflecting differences in product margins, acquisition costs, or discounts.

Because these results are correlational, they do not demonstrate causation. In addition, profit is likely calculated partly from sales and costs. The high correlations may therefore reflect accounting relationships rather than statistically independent associations.

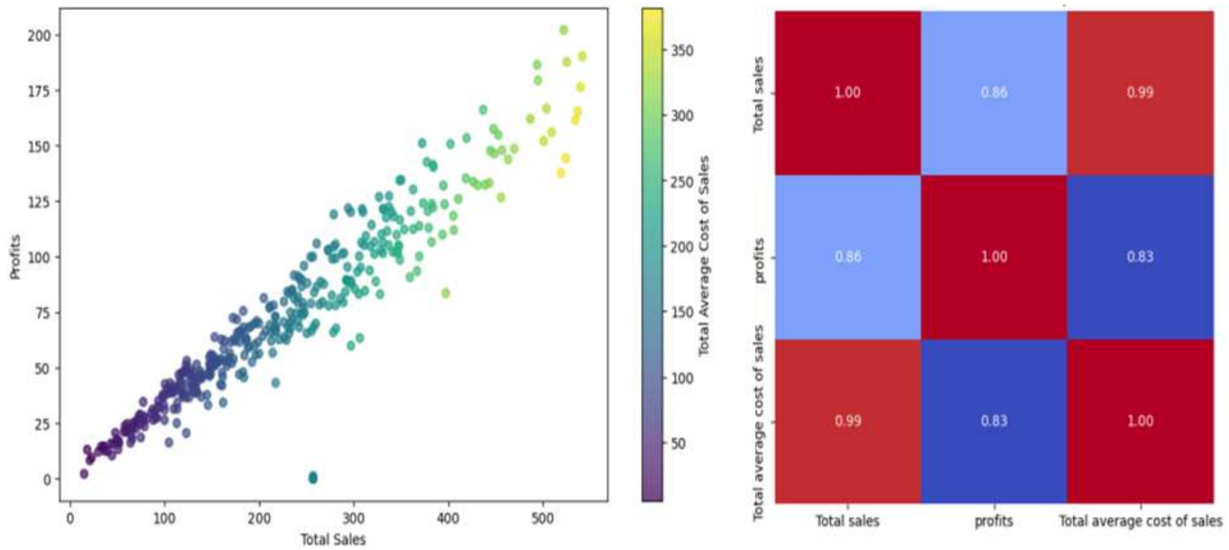

Figure 4. Relationship between total sales, profits, and total average cost of sales: scatterplot of total sales against profits colored by cost of sales, and Pearson correlation matrix

4.7 Daily Variation Across the Study Period

The chronological plots showed substantial day-to-day variation in total sales and profits (Figure 5). These two measures generally moved in the same direction, consistent with their strong positive correlation. Total average cost of sales displayed comparatively less variation but increased during many of the same high-activity periods. During the observed portion of 2022, the three variables fluctuated without a clear sustained upward or downward direction. The 2023 series contained several clusters of high sales, costs, and profits. A possible increase in sales and profits was visible during parts of 2023, but short-term fluctuations were more prominent than any overall trend. The evidence was therefore insufficient to conclude that the pharmacy experienced sustained growth during the study period.

Abrupt low values and discontinuities were also observed. These may indicate pharmacy closure days, missing transactions, stock shortages, or changes in the recording process. Since contextual variables were unavailable, the temporal plots could identify when unusual behavior occurred but could not explain why it occurred.

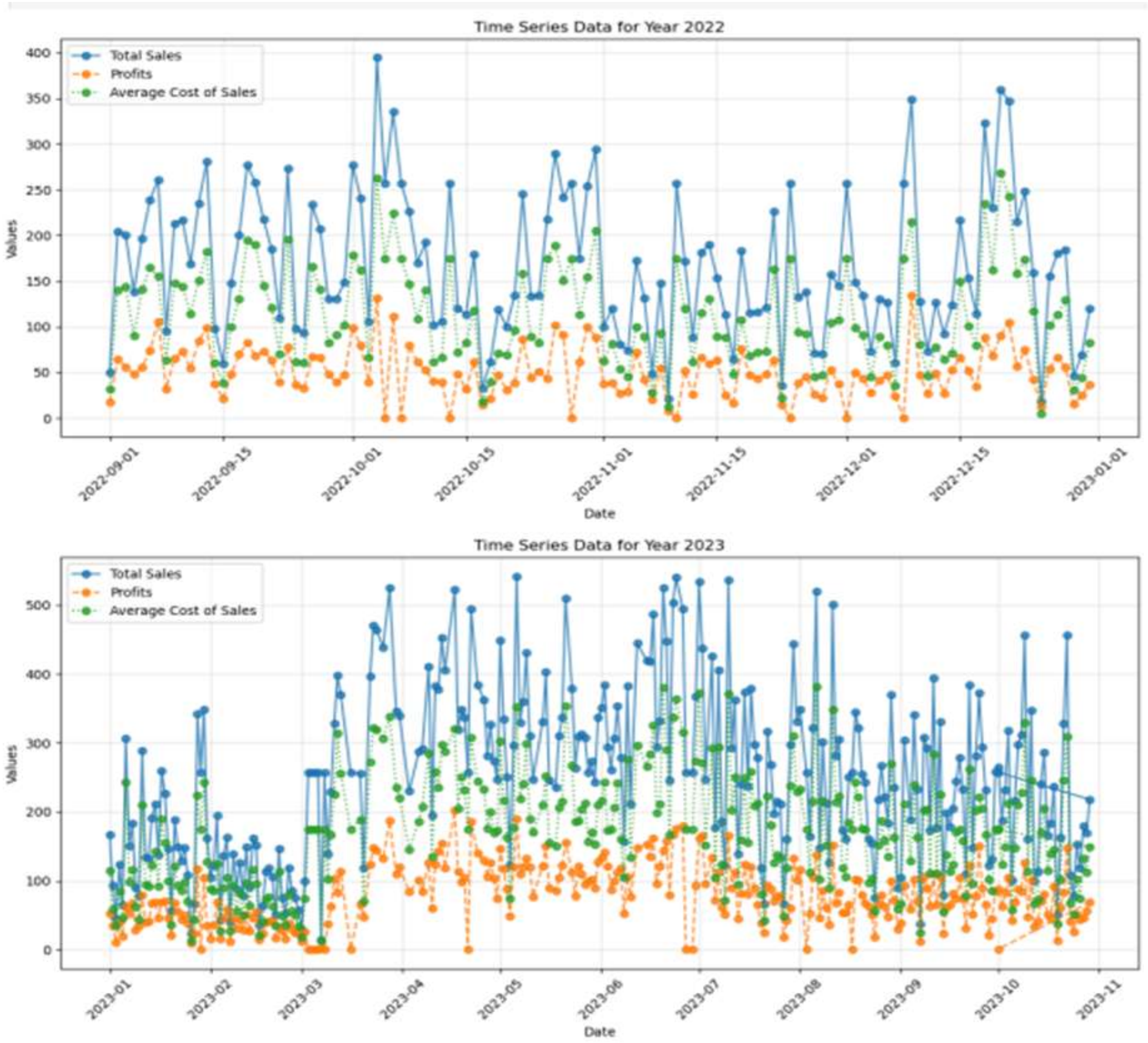

Figure 5. Daily total sales, profits, and total average cost of sales during 2022 and 2023.

4.8 Operational Implications

The exploratory findings have several implications for pharmacy management. First, the highly skewed profit distribution demonstrates that routine performance should not be assessed using the arithmetic mean alone. Median profit, interquartile range, and verified profit margin would provide a more representative description of typical operations.

Second, the strong relationship between sales and costs indicates that revenue growth should be evaluated alongside the acquisition cost of the products sold. An increase in revenue does not necessarily result in a proportional improvement in profitability when product costs and discounts also increase.

Third, the apparent concentration of higher activity toward the end of the week may be relevant to staffing and inventory decisions. If confirmed using a longer observation period, the pharmacy could increase monitoring of high-demand products and adjust staff coverage before the busiest days. However, operational changes should not be based solely on the current dataset because the weekly pattern may be influenced by unmeasured factors.

Finally, the analysis demonstrates the value of routine data-quality monitoring. Financial observations that exceed predefined thresholds should be automatically flagged and compared with the original transaction records. Such a procedure would help distinguish genuine high-activity days from accounting or data-entry errors.

4.9 Limitations

This study has several limitations. The data were obtained from a single community pharmacy, limiting the generalizability of the findings to other pharmacies and geographical settings. The study covered approximately 14 months, which was insufficient to establish recurring annual seasonality or long-term changes.

The dataset included aggregated daily financial measures but did not contain product categories, quantities sold, customer numbers, prescription types, supplier information, opening hours, stock availability, holidays, or promotional activities. The analysis could therefore describe observed patterns and associations but could not determine their causes.

Despite these limitations, the exploratory analysis provides a structured description of the pharmacy's daily financial activity. It identifies highly skewed financial distributions, substantial short-term variation, possible calendar-related differences, and strong relationships among sales, costs, and profits. These findings provide a foundation for improved data-quality practices and evidence-informed pharmacy operations without relying on predictive modeling.

6. Conclusion :

This study explored approximately 14 months of community pharmacy sales data and identified positively skewed distributions, substantial daily variation, and strong relationships among sales, costs, and profits. Higher activity was more frequently observed toward the end of the week, although the limited study period did not support firm conclusions about seasonality. The findings emphasize using robust measures such as medians, interquartile ranges, and profit margins, while verifying extreme values against original records. Although limited to one pharmacy and a small set of aggregated variables, the study demonstrates how exploratory analysis can support financial monitoring, inventory planning, data-quality improvement, and operational decision-making. Future studies should use longer, multi-pharmacy datasets and incorporate product, customer, prescription, and inventory information.

References:

[1] M. A. Assab et al., "Financial indicators utilization among community pharmacists: A comprehensive study for pharmacy management," PLoS One, vol. 19, 2024, doi: 10.1371/journal.pone.0299798.

[2] M. A. Assab et al., "Assessing pharmacists' awareness of financial indicators in community pharmacy management: A cross-sectional study," Heliyon, vol. 10, 2024, doi: 10.1016/j.heliyon.2024.e33338.

[3] M. Essgaer, A. Agaal, A. Abbas, and R. Al Mamlook, "Profit Forecasting for Daily Pharmaceutical Sales Using Traditional, Shallow, and Deep Neural Networks: A Case Study from Sabha City, Libya".

[4] M. Essgaer, A. Agaal, and M. Matoug, "Integrating SHapley Additive ExPlanations and Hyperparameter Tuning in Weather-Informed Electricity Load Forecasting: A Case Study of Northern Benghazi Power Plant," in International Conference on AI: Current Research, Industry Trends, and Innovations, Springer, 2025, pp. 166–187.

[5] A. Alsharif, A. Agaal, A. A. Ahmed, M. Khaleel, T. M. A. Ali, and K. Qasim, "The Role of Artificial Intelligence Tools in Enhancing Sustainability in Scientific Research," Open Eur. J. Eng. Sci. Res., pp. 24–39, 2025.

[6] A. Agaal, A. Mohammed, A. Ahessin, and R. Abdelhamed, "Beyond Accuracy: A SHAP and LIME-based Explain Ability Framework for Pharmacy Profit Prediction," in 2026 IEEE 5th International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), IEEE, 2026, pp. 1–8.

[7] A. Agaal, M. Essgaer, H. M. Farkash, and Z. A. Othman, "Data-driven insights for informed decision-making: applying LSTM networks for robust electricity forecasting in Libya," arXiv Prepr. arXiv2507.01034, 2025.

[8] J. Maggu, H. Gorantla, D. Singh, and R. Malik, "ProWAR: A profit-oriented approach to weighted association rule mining for retail analytics," in AI and Sustainable Transformations, CRC Press, 2025, pp. 260–266.

[9] L. A. Berenbrok, N. Gabriel, K. C. Coley, and I. Hernandez, "Evaluation of frequency of encounters with primary care physicians vs visits to community pharmacies among Medicare beneficiaries," JAMA Netw. Open, vol. 3, no. 7, p. e209132, 2020.

[10] J.-V. Goode, J. Owen, A. Page, and S. Gatewood, "Community-based pharmacy practice innovation and the role of the community-based pharmacist practitioner in the United States," Pharmacy, vol. 7, no. 3, p. 106, 2019.

[11] C. Piquer-Martinez et al., "Integration of community pharmacy in primary health care: the challenge," Res. Soc. Adm. Pharm., vol. 18, no. 8, pp. 3444–3447, 2022.

[12] S. H. Chalasani, J. Syed, M. Ramesh, V. Patil, and T. M. P. Kumar, "Artificial intelligence in the field of pharmacy practice: A literature review," Explor. Res. Clin. Soc. Pharm., vol. 12, p. 100346, 2023.

[13] S. Thomas, "A Study on the Role of Data Governance in Enhancing Big Data Analytics Precision and Decision-Making in the Retail Industry," 2025, University of the Cumberlands.

[14] Y. Yin, F. Chu, A. Dolgui, T. C. E. Cheng, and M. Zhou, "Big data analytics in production and distribution management," 2022, Taylor & Francis.

[15] K. Tomokazu, “Tutorial Review: Exploratory Data Analysis with R as a Novel Framework for Seismic Data Interpretation,” Sci, vol. 8, no. 4, p. 81, 2026.
[16] U. Kadel, “Inventory Control of Medicine in Private Hospital Pharmacy in Nepal: A Case Study of IFCH Hospital Pharmacy,” 2026.
[17] R. M. Rosenfeld et al., “Lifestyle interventions for treatment and remission of type 2 diabetes and prediabetes in adults: a clinical practice guideline from the American College of Lifestyle Medicine,” Am. J. Lifestyle Med., vol. 19, no. 2_suppl, pp. 10S-131S, 2025.
[18] M. Baillie, S. Le Cessie, C. O. Schmidt, L. Lusa, M. Huebner, and T. G. “Initial D. A. of the S. Initiative, “Ten simple rules for initial data analysis,” 2022, Public Library of Science San Francisco, CA USA.
[19] A. Hubai, S. Szabó, and B. Zaválnij, “Exploratory Data Analysis and Searching Cliques in Graphs,” Algorithms, vol. 17, no. 3, p. 112, 2024.
[20] S. R. Chaganti, Transactional to Transformational Marketing in Pharma: The Science of Why and the Art of How. BSP Books, 2023.
[21] A. R. Fisher, T. S. Hodgdon, and M. G. R. Lewis, “Time-series forecasting methods: A review,” 2024.
[22] X. Xu and A. G. Sam, “Environmental sustainability and business profitability: Profiling winners and losers with machine learning,” Bus. Strateg. Environ., vol. 34, no. 5, pp. 5205–5239, 2025.
[23] C. J. O. Trejo-Pech, K. L. DeLong, and R. Johansson, “How does the financial performance of sugar-using firms compare to other agribusinesses? An accounting and economic profit rates analysis,” Agric. Financ. Rev., vol. 83, no. 3, pp. 453–477, 2023.
[24] L. A. Moura Gomes, A. M. Gomes Costa, S. M. R. Steurbaut, H. D. Mota Filipe, S. N. Leite, and F. da P. C. Alves da Costa Azevedo e Silva, “A qualitative study to explore global trends in clinical pharmacy practice,” J. Eval. Clin. Pract., vol. 31, no. 4, p. e70104, 2025.
[25] V. Leivaditis, C. Ntais, J. Fanourgiakis, F. Mulita, and N. Kontodimopoulos, “International experience in the management of pharmaceutical expenditure: A narrative literature review,” Br. J. Hosp. Med., vol. 86, no. 3, pp. 1–19, 2025.
[26] R. Desai, “Optimizing pharmacy operations: Strategies for efficiency and quality care,” Int. J. Res. Innov. Pharm. Pract., vol. 1, no. 1, 2024.
[27] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” Glob. Transitions Proc., vol. 3, no. 1, pp. 91–99, 2022.
[28] K. Paraskevas and T. Christos, “Data preprocessing and feature engineering for data mining: Techniques, tools, and best practices,” Ai, vol. 6, no. 10, p. 257, 2025.
[29] M. Bilal, G. Ali, M. W. Iqbal, M. Anwar, M. S. A. Malik, and R. A. Kadir, “Auto-prep: efficient and automated data preprocessing pipeline,” Ieee Access, vol. 10, pp. 107764–107784, 2022.
[30] M. Tessler, “Univariate Analysis: Variance, Variables, Data, and Measurement,” in Social Science Research in the Arab World and Beyond: A Guide for Students, Instructors and Researchers, Springer, 2022, pp. 19–50.
[31] D. Sandilands, “Bivariate analysis,” in Encyclopedia of quality of life and well-being research, Springer, 2024, pp. 481–483.
[32] M. Efatmaneshnik, N. Chitsaz, and L. Qiao, “Decision making for multi-objective problems: Mean and median metrics,” Syst. Eng., vol. 26, no. 6, pp. 814–829, 2023.
[33] J. DHANNYA, A. N. RAJAN, C. K. SHAHNAZEER, and I. MISHRA, DATA VISUALIZATION USING PYTHON: NUMPY, PANDAS, MATPLOTLIB AND SEABORN. Vaagai international publishing house, 2025.
[34] S. S. Qudratovich, “DATA VISUALIZATION IN PYTHON,” Eurasian J. Math. Theory Comput. Sci., vol. 4, no. 10, pp. 15–22, 2024.