

From Sales to Strategy: A Lightweight Analytical Framework for E-Commerce Performance Evaluation

¹ Ritu Singhal, ² Priya Gupta, ^{3*} Harshita Sharma, ⁴ Deepak Kumar

¹Indraprastha College of Women, University of Delhi, Delhi, India

^{2,3}Atal Bihari Vajpayee School of Management and Entrepreneurship, Jawaharlal Nehru University, Delhi, India

⁴TataConsultancyServices, Boston, MA, USA

¹rsinghal@ip.du.ac.in, ²privagupta@jnu.ac.in, ³harshi60_sme@jnu.ac.in, ⁴deepakkbme@gmail.com

*Corresponding Author: harshi60_sme@jnu.ac.in

Information of Article

Article history:

Received: 17 June 2026

Revised: 29 August 2026

Accepted: 27 September 2026

Online: 20 October 2026

Keywords:

Lightweight Analytics

E-Commerce Performance

Evaluation

Profitability Analysis

Visual Analytics

Strategic Decision Support

ABSTRACT

Modern e-commerce generates large amounts of transactional data, although, surprisingly, often fails to leverage its strategic potential. Typically, companies are focused only on sales quantity metrics but do not ask themselves what specific products or product groups bring their company any money. In this regard, the present paper attempts to fill the existing gap by providing an analytical approach based on easily replicable methodology to convert sales data into strategic information. Using publicly available Amazon Sales Report data from Kaggle, the research applies a set of algorithms in Python via Google Colab platform to convert raw sales numbers into meaningful information that goes far beyond traditional sales reports. Since the data does not provide any information regarding costs incurred in order to make sales, the profit is estimated using a common proxy margin, making it possible to compare SKUs and categories in relative terms, without measuring absolute profits. Visual analytics in the form of monthly charts, category-level analysis, SKU ranking and contribution treemap are employed to provide insights that cannot be gained with conventional dashboarding techniques. According to the results, some of the high-sales SKUs and product categories account for surprisingly small shares of profits. With the emphasis of this revenue-profits divide, the model offers managers and analysts an effective perspective through which to focus product decisions, adjust their marketing spending, and negotiate costs based on margins. The research shows that strategy does not require sophisticated machine learning or costly business intelligence tools but rather intelligent and intelligible analytics on available data.

1. Introduction

Amazon handles millions of transactions daily on its own, leading to large amounts of data that can be analyzed for order information, SKUs, product category, and sales statistics. While useful in itself, the sales performance data often lacks any kind of connection to profit margins and business strategy. What is needed are the points of data that can inform the decision-making process regarding the choice of products to promote, considering costs involved, and addressing specific target markets.

While studies on analytical methods in e-commerce commonly concern themselves with sales forecasting, demand analysis, and dashboard reporting [1], [2], rarely do they attempt to draw the connection between sales information gained via transaction and analyzing strategies for SKUs and product categories, though such an approach is important for choosing profitable products among the highly demanded ones with low margin [3].

This paper seeks to fill the identified gap by introducing a lightweight e-commerce analytics methodology to measure performance. The paper entitled “From Sales to Strategy” describes the process of turning sales numbers into insights and strategies through the use of transparent and understandable analysis. Unlike machine learning algorithms or intensive business intelligence systems, the presented methodology is lightweight, reproducible, and easily accessible as it does not rely on any other resources other than Python code in Google Colab.

Firstly, the inclusion of three different factors within the methodology, namely, revenue, estimated profit, and margin, allows one to conduct a more diverse set of analyses than merely reporting the results of sales. Secondly, the use of lightweight, reproducible analysis via Google Colab through open source Python code distinguishes the study from any existing research methodologies. Thirdly, the specific focus on analyzing the profitability at SKU-level along with overall revenue makes it relevant for managerial purposes. In addition, the theoretical approach that was developed in the course of this research has been constructed in a way to be applicable when there is not enough information regarding the actual cost, thus, allowing relative profits to be calculated rather than absolute ones.

The research demonstrates that it is possible to generate an effective strategy using the combination of limited assets and publicly available data sources. In terms of its contributions from the academic standpoint, the paper enriches the existing literature by introducing a process-based model aimed at transparency instead of black-box solutions [4]. From

a practical standpoint, the study helps business owners and managers understand how it is possible to get below the surface of sales and identify profitable goods and product categories.

Therefore, the purpose of the research is to develop and test an efficient method for analyzing online retail sales through visual analytics in Python. In order to achieve this goal in a systematic way, this study will be conducted within the framework of the following research questions:

RQ1: Do high revenue SKUs rank better on the estimated relative profitability scale than do low revenue SKUs if profitability is measured via an identical margin?

RQ2: How much information can revenue concentration in different categories provide for strategic prioritization in e-commerce performance assessment?

RQ3: Is it possible that simple visualization-based analytics can provide similar insights compared to sophisticated business intelligence solutions?

2. Literature Review

The development of e-commerce has led to an increase in academic research into transaction analysis, customer behavior analysis, and business operational analysis. Earlier approaches for e-commerce analytics are largely descriptive sales analytics and performance dashboards. Later approaches have moved towards advanced analytics including predictive models, customer behavior analysis, and analytics-driven decision support. While there have been many advancements in this direction, the use of analytics in strategic decision-making based on routine sales transactions especially at SKU and product category levels is still an under-researched topic.

2.1 Customer-Centric and Customer Experience Analytics

There is considerable literature on the topic of customer-centric and experience analytics based on unstructured data. Kumar et al. [1] proposed a deep learning hybrid model for dissatisfaction analytics in Amazon review dataset. The paper demonstrates the utility of advanced analytics in quality management and customer experience improvement; however, the approach is computationally expensive. Xu et al. [2] extended the customer-centric analytics by proposing a multimodal model for live-streaming e-commerce platforms. This approach takes advantage of visual and behavioral analysis apart from textual analysis of anchor reputation to predict sales performance and plan inventory.

Recommendation systems are important to promote the level of customer engagement. According to Dhage et al. [4], there is a need for a recommendation system-as-a-service model provided through cloud APIs to ensure scalability and availability for smaller businesses within the e-commerce domain. Although this study contributes towards the process of analytics democratization, its main concern seems to be focused on recommendation success instead of profit-making ability or strategic performance measurement. According to Mohanty et al. [22], the importance of practical analytics as well as intelligent systems in decision-making for e-commerce and services cannot be overstated. Rao et al. [23], in their comparative analysis of analytical approaches to e-commerce, point out that the model selection and interpretability issues are very important for effective business decision-making. Yadav et al. [24], in their research concerning analytics adoption, argue that interpretability and usability issues for managers should not be overlooked. According to Yadav et al. [25], modern analytics frameworks should be sophisticated enough but also transparent for successful decision-making.

2.2 Sales Predictive Analytics and Forecasting Models

Sales forecasting models have gained great interest among the areas of e-commerce analytics research. Tan [3] implemented multiple layers of LSTM forecasting with attention techniques for e-commerce sales forecasting and achieved high predictive accuracy and better interpretability via visualization techniques. Similarly, Chen et al. [11] combined GRU and LightGBM using stacking to predict perishable goods' sales volume to overcome issues related to volatility and expensive inventory costs.

Similarly, Chen et al. [11] combined GRU and LightGBM using stacking to predict perishable goods' sales volume to overcome issues related to volatility and expensive inventory costs. While these studies have shown high accuracy in sales forecasting, their main focus remains on model performance and complexity, rather than management usefulness.

Zhang et al. [5] exposed the problems with static sales models by collecting data regarding sales event dynamics, demonstrating how consumers' behavior during promotion periods can be highly non-stationary. However, even though they emphasized the necessity for an adaptive evaluation framework, no aspect of profit and strategic planning was mentioned in their study.

2.3 Analysis Centering on Profitability and Decision Support Systems

As compared to predictive analysis of sales, research on profitability-oriented analysis is more limited. Rahman [7] highlighted the significant role that decision support systems play with both revenue and profit margin in better resource allocation for online sales. The research concluded that the sales quantity does not guarantee profitability of products, highlighting the importance of taking margins into consideration. Nevertheless, decision support systems are usually integrated into comprehensive software packages and, therefore, less available for SMEs.

Visualization has become an increasingly popular tool to build a bridge between data analysis and decision-making processes. Varsha and Karan [8] stressed the importance of using descriptive analytics and visualization for e-commerce, while Zhang et al. [9] developed the system called PromotionLens for evaluating promotional activities. This proves the importance of visualization in making decisions regarding pricing, promotions, and inventory; however, all those tools are typically stand-alone dashboards.

2.4 Big Data Analytics and System-Level Studies

A number of investigations have explored e-commerce analytics using a big data perspective. Specifically, Kalla [6] gave an extensive review on the effects of big data analytics in transforming e-commerce, including benefits such as operational efficiency, customer experience improvement, and enhanced decision-making. Pan [16] performed a bibliometric analysis based on a collection of 265 papers, which mapped out major research clusters and uncovered gaps in literature relating to big data-based studies in e-commerce. While these studies show the transformative power of big data, there are associated challenges of high costs, complexity, and infrastructure dependency.

2.5 Big Data Analytics and Its Value in E-Commerce

Alrumiah et al. assessed the value of Big Data Analytics (BDA) for the development of e-commerce [18], conducted a review of fifteen studies on the implications of BDA on vendors and consumers. It was found that BDA leads to improved decision-making, personalization, and competitive advantage; at the same time, BDA poses several difficulties in terms of high costs of implementation, complexity of data accumulation, and even adverse consequences for customers. As an illustration of the latter point, the investigation of machine learning in e-commerce presented in [19] by Zineb et al. is worth mentioning. However, in both cases, there is much focus on large-scale data utilization and advanced analysis without practical guidelines about how to evaluate strategies through easier analysis.

A model by Behl et al. [20] for ISM adoption of BDA in e-commerce ventures revealed the key facilitators in this respect to be availability of appropriate data and technical skills among employees, especially in developing countries. These insights highlight the importance of employing methods that combine analytical power with organizational reality.

2.6 Organizational Architecture, Logistics, and Infrastructure

Apart from analytic methods, various studies have examined organizational and architectural issues in e-commerce. For instance, Irawan et al. [13] used the TOGAF enterprise architecture approach for introducing e-commerce into conventional markets by addressing the problem of infrastructure constraints and technological readiness as the most crucial obstacles. In his work, Wu [17] has considered models of e-commerce logistics through applying big data unstructured analysis and inter-organizational cooperation as key components.

Moreover, another trend in the field is connected to cloud engineering and automated processes of e-commerce implementation. According to [21], containerization, orchestration, and automated implementation positively influenced the scalability, robustness, and reactivity of the systems under consideration. Nevertheless, while such techniques increased the reliability of e-commerce platforms, not all organizations would be able to implement them.

2.7 Issues of Security and Trust in an Evolving Context

Security and trust issues still constitute some of the vital issues that need to be considered in modern e-commerce environments. The AIGUARD system [14], developed to detect harmful AI-generated content in the practical context of e-commerce, successfully performed its operation with greater detection accuracy and minimal computational cost. On the other hand, [15] introduced a secure and scalable transaction protocol using cryptographic algorithms such as

ElGamal encryption and AES, which can help ensure the security of e-commerce transactions from any kind of cyber attack [26].

2.8 Gaps and Contributions

In general, current literature shows significant advances in the areas of predictive modeling, recommendation engines, big data analysis, visualization, and system optimization. Nonetheless, many methods focus on forecasting accuracy, customer behavior, or system infrastructures, using highly sophisticated models and requiring significant computing power. The link between regular sales transactions and strategic decision making in terms of SKUs and product categories is not adequately analyzed via simple and explainable analytical techniques.

This gap in literature can be addressed by developing an explainable analytical framework based on simple Python tools.

Through emphasizing the importance of profit-oriented analysis rather than traditional revenue reporting, the framework allows practitioners to discover profit-sales discrepancies and prioritize relevant products for future decisions. In doing so, this paper adds to the body of literature on predictive and big data models while providing e-commerce companies with a solution that can enable them to “move from sales to strategy.” It is critical to state here that the suggested framework is not designed as a replacement of any cost accounting systems or even profitability tools. The framework is seen here more like a preliminary diagnostic framework for practitioners and scholars dealing with public transactional data, which usually lacks cost-level information.

2.9 Comparative Positioning of the Proposed Framework

To make the novelty explicit, Table 1 positions the proposed framework against the main families of profit and performance oriented retail analytics reviewed above. Existing approaches deliver accuracy or depth but at the cost of data requirements, compute, interpretability, or accessibility. Predictive sales models [3], [11] and deep or multimodal customer analytics [1], [2] achieve strong forecasting performance but need large labelled data, GPU scale compute, and specialist skills, and they optimise prediction rather than margin. Decision support systems and BI dashboards [7], [8], [9] foreground profit and visualisation but are usually embedded in licensed enterprise software out of reach for smaller sellers. Big data pipelines [6], [16], [18] add scale but also cost and infrastructure dependency. None offers a transparent, openly reproducible, cost free path from raw transactional exports to SKU and category level profit oriented strategy. That specific combination, profit orientation plus full reproducibility plus zero licensing and infrastructure cost, is the contribution of this paper.

Table 1: Comparative positioning of the proposed framework against existing approaches

Approach family	Representative work	Primary objective	Profit oriented	Data / compute demand	Interpretability	Cost / accessibility
Predictive forecasting	[3], [11]	Forecast sales	No	High (GPU, large data)	Moderate	Low
Deep multimodal analytics	[1], [2]	Customer sentiment	No	Very high	Low	Low

DSS and BI dashboards	[7], [8], [9]	Report and monitor	Partly	Moderate	High	Low (licensed)
Big data pipelines	[6], [16], [18]	Scale and integrate	No	Very high	Moderate	Low
Proposed framework	This paper	SKU / category profit strategy	Yes	Very low (free Colab)	High	Very low / accessible

3. The Proposed Lightweight Analytical Framework

The proposed architecture (Figure 1) starts with the Amazon Sale Report dataset that consists of raw transactions, such as order IDs, SKUs, sales, amounts, quantities, and shipping details. In the second phase, data preprocessing is conducted to clean the missing values, standardize the date format, and convert numeric variables like quantity and amount into the proper format. Additional features like unit price, derived profit, and margin percentage are also computed at this stage.

In the third phase, the analytical engine in light weight, which has been programmed using Python inside Google Colab, will compute aggregations in various granularities (months, categories, SKUs, and regions), and compute KPIs such as revenue, profit, and margin percentage.

At the fourth phase, descriptive visual analytics techniques can be applied by creating charts showing trends of monthly revenue-profit, category-wise comparisons, ranking of SKUs, proxies for the Price-Qty Elasticity measure, contribution tree map based on category, and revenue per region.

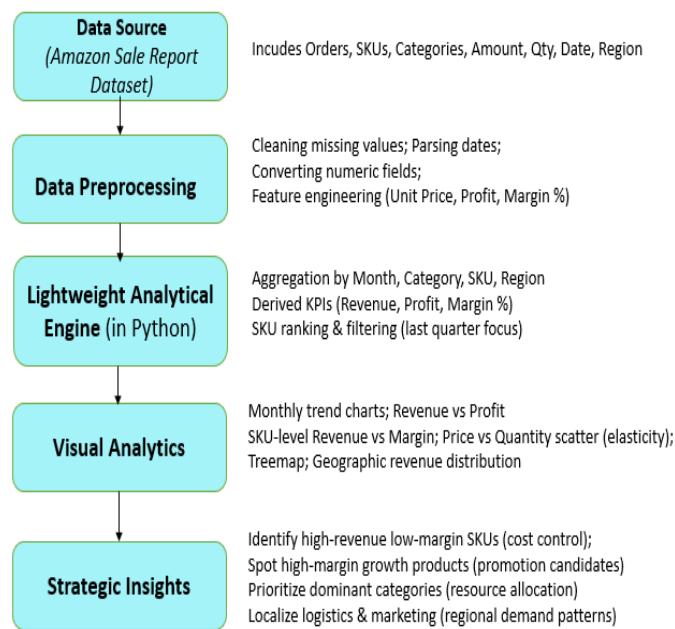


Figure 1: Lightweight Analytical Framework

The last stage is the strategic insights stage. While descriptive reporting will only end up reporting the revenues, this approach goes ahead to identify managerial insights from such reporting. For example, it shows SKUs that generate high revenues but have low margins, those SKUs that generate both high revenues and margins, as well as areas that cluster in terms of their demands. This makes it possible for organizations to progress from sales to strategy. The point which must not be overlooked is that the method has an inherent extensibility to its credit. Once the COGS data becomes accessible via ERP data or supplier data, the substitution of the proxy measure for margin for the per-product margin becomes very easy indeed.

Scalability beyond the present dataset raises two distinct considerations. The first is data volume. The current pipeline relies on pandas based aggregation and Matplotlib and Seaborn visualization within Google Colab, which is well suited to datasets on the order of hundreds of thousands of rows, as used here, but would need to move to chunked processing or a SQL or warehouse backed aggregation layer for datasets in the millions of rows. The second is multi channel applicability. The present dataset is limited to a single sales channel; the raw data already carries a Sales Channel field, so extending the framework to multiple channels would primarily require channel level aggregation and, if channel specific fee structures exist, channel specific margin assumptions rather than a single global margin.

4. Methodology and Experimental Setup

In this paper, we employed a sales dataset for Amazon from Kaggle. The dataset had basic attributes such as order number, date, product code (SKU), product category, sold quantities, the price paid, and shipping details. There were no profit-related attributes in the data; thus, profit was estimated using a simple method where the profit was derived using a generic profit margin on the sales amount. Similar to the conventional methods used by [7] to estimate profits, we used a single margin of 15% for estimating product-level profit based on the sales amount. As explained by [7], the margin served as a proxy variable and was not meant to capture actual profitability at the firm level. Rather, it was used in order to compare relative performance across SKUs and their respective product categories. Even with the simplicity involved, the approach made it possible for us to make some meaningful comparisons across various SKUs and categories. Table 2 below provides the characteristics of the Amazon Sales Report dataset used in this paper.

Table 2: Dataset Overview

Attribute	Description
Data source	Amazon Sales Report dataset (Kaggle)
Number of transactions	~128,000 orders
Time period covered	April 2022 – March 2023
Number of SKUs	~7,000 unique product codes
Number of product categories	9 categories
Geographic granularity	State-level (India)
Key variables	Order ID, date, SKU, category, quantity, amount, state

The Amazon Sales Report dataset was selected because it is one of the few publicly available e-commerce datasets that retains order level granularity down to individual SKUs and product categories, which is a precondition for the SKU level profitability analysis proposed in this paper. Unlike aggregated sales summaries, it reflects the operational structure of an actual online retail seller, including order status, fulfilment type, and shipment geography, which allows the framework to be demonstrated under realistic data constraints rather than idealized conditions.

At the same time, the dataset carries limitations that bound the generalizability of the findings. It represents a single seller operating on the Amazon.in channel, so patterns observed here may not extend to multi seller marketplaces or other platforms. The catalog is heavily weighted toward apparel categories, with Set and kurta alone accounting for roughly 77 percent of total revenue in the sample, so category level findings should not be read as representative of e-commerce in general, for example electronics or grocery. In addition, about 14 percent of the recorded orders carry a Cancelled status rather than Delivered, so revenue figures computed from raw order amounts can overstate realized sales unless order status is filtered or accounted for explicitly. These constraints are revisited as limitations in Section 6.

The data analysis is conducted using the use of Google Colab and different Python packages such as Pandas, Matplotlib, and Seaborn. Data is also cleaned to remove missing data and format dates and numbers. Further, some additional information like unit price, profit, and profit margins have been calculated for the purpose of analysis. We only focused on those transactions that occurred in the last quarter mentioned in the data to focus the analysis on the most recent transactional period. Products or SKUs are sorted based on their sales performances and compared for their corresponding profitability.

Data visualization is one of the key aspects in any research. Thus, for the purpose of conducting visual analysis, bar charts, scatter plots, line graphs, and tree maps were used to show the relationship between monthly sales, products categories performance, product rankings, and price against quantity patterns.

4.1 Limitations of Profit Estimation

It is essential to note the inherent structure constraint in the profitability estimation methodology used herein. Since there is a constant margin rate of 15% for all SKUs and product categories, profit estimates are a simple linear function of revenue. Essentially, multiplication of all revenue values by a constant scalar does not affect their ordinal ranking. Thus, ranking by revenue and ranking by profit estimation will always be consistent in this case, and any variations that occur between different categories and SKU codes can be solely blamed on the degree of revenue concentration and not on real differences in profitability. Therefore, the process serves more as an analysis tool to diagnose revenue concentration as opposed to measuring actual profitability. This is an intentional approach taken bearing in mind the resource limitations involved with analyzing publicly available transaction data where individual product cost data is lacking. Furthermore, given that the margin rate is merely a multiple of the revenue figures, analyzing on the basis of any other margin rate such as 10% or 20% would yield structurally similar results.

4.2 Sensitivity Analysis of the Margin Assumption

To empirically verify the claim above, profit and category level rankings were recomputed under four margin assumptions: 10 percent, 15 percent (the baseline used throughout the paper), 20 percent, and 25 percent. Table 3 reports the top five categories by estimated profit at each margin level and the Spearman rank correlation between that ranking and the 15 percent baseline.

Table 3 : Category profit rankings under alternative margin assumptions

Margin	Top 5 categories	Correlation
10%	Set, kurta, Western Dress, Top, Ethnic Dress	1.00
15% (baseline)	Set, kurta, Western Dress, Top, Ethnic Dress	1.00
20%	Set, kurta, Western Dress, Top, Ethnic Dress	1.00
25%	Set, kurta, Western Dress, Top, Ethnic Dress	1.00

As anticipated, category and SKU rankings are invariant to the choice of margin percentage, since profit is defined as a fixed linear multiple of revenue. The sensitivity analysis therefore confirms empirically, rather than only asserting, that the qualitative findings reported in Section 5 do not depend on the specific 15 percent figure chosen.

This robustness result should not be read as evidence that profitability itself is insensitive to costs. It shows only that, given a single uniform margin, the choice of that margin's magnitude does not change the ranking. If real margins vary by category or SKU, as they typically do in practice, rankings could diverge materially from the uniform margin case; this distinction is carried forward as a stated limitation and a direction for future work in Section 6.

4.3 Data Preprocessing Details

The data for the Amazon Sales Report collected from Kaggle included close to 128,000 transactions made between April 2022 and March 2023. During the data cleaning process, those records that had missing values in significant columns, such as the date of the order, the category of the product, units sold, and transaction amount, were identified and removed. The columns relating to dates were normalized in order to enable the accurate calculation of the data by months, while columns containing numbers, such as number of units sold and total amount, were converted into appropriate formats. The unit price, estimated profit, and margin percentage were calculated as new variables. The dataset used for the analysis was composed of nearly 7,000 SKU's across 9 product categories, with resolution down to the state level in India. The analysis included only the last quarter's data.

5. Results and Discussion

Figure 2 shows the dynamics of the total revenue and profit estimate on a monthly basis, thus providing an insight into comparing sales performance and profitability dynamics during the research period.

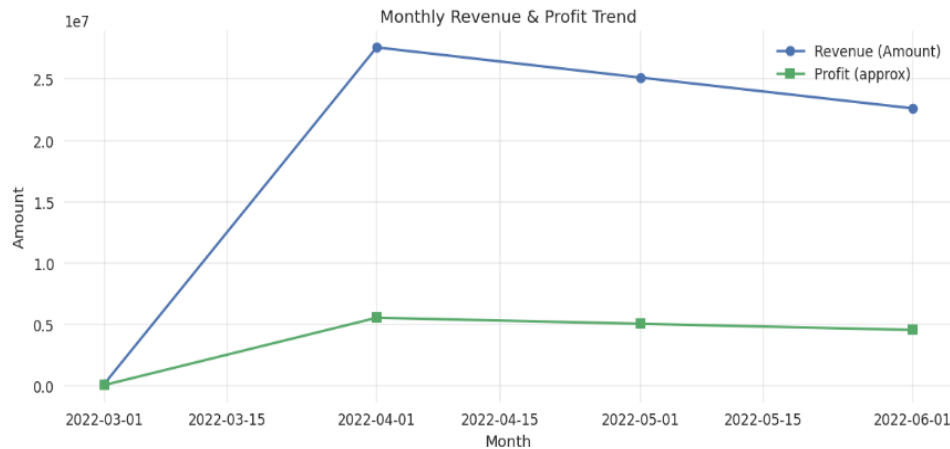


Figure 2: Monthly Revenue and Profit Trend

Figure 3:
Revenue vs Profit
by Category

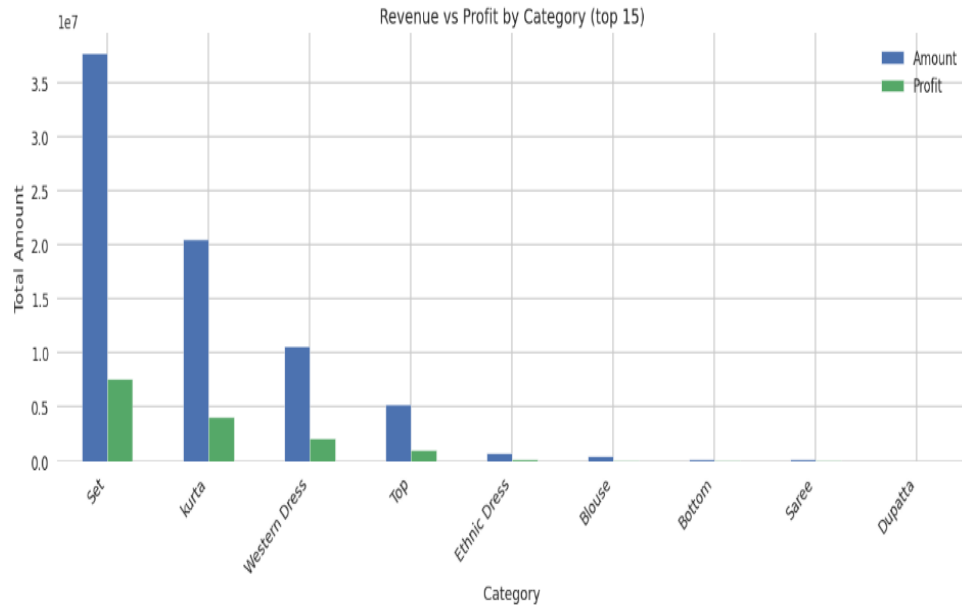


Figure 3 provides a chance for a more detailed analysis of category performance according to the total revenue and profit estimate. Despite high sales volume of such categories as Electronics or Apparel, the profit from their sales may differ significantly. Such discrepancies indicate that sales figures alone are not enough to measure company performance.

By adding the profitability component into the equation, the figure allows for a shift towards a more business-oriented approach. For example, the results will provide insight into which categories are responsible for generating profits and which generate sales mainly. These findings can then be applied to improve decision-making when planning marketing budgets or negotiating with suppliers. In summary, the information presented above implies that revenue as well as profit should be taken into account before increasing any investments in product categories.

Figure 4 underlines the practical significance of conducting performance analysis based on individual SKU. Those SKUs which yielded good sales but had a poor margin were the ones where improvements can be made in terms of better cost control or discount management. In comparison, those SKUs which yielded both sales and margins in abundance were selected as SKUs worth expanding on commercially. Additionally, those SKUs which were not doing well sales-wise but still managed to yield a lot of margins were identified through this analysis as SKUs worth promoting aggressively.

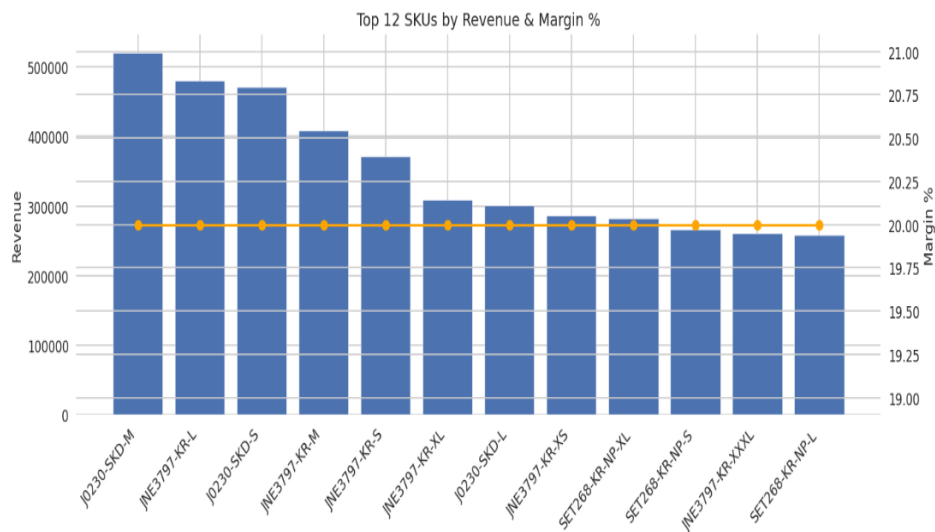


Figure 4: Top SKUs in the Last Quarter

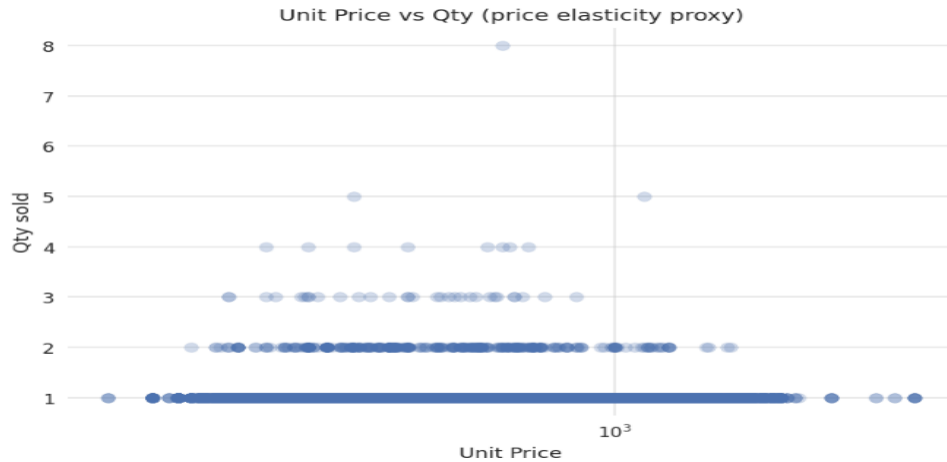


Figure 5: Unit Price vs Quantity Sold

Figure 5 demonstrates the association between the unit price and the quantity sold per SKU, thus providing an indication of price elasticity and demand trends. As observed from the chart above, there exists an inverse relationship, whereby an increase in unit price will result in a low quantity being sold, indicating that SKUs have a price elastic nature. There also exist certain SKUs, which are considered to be luxury items and are still in high demand even when their prices rise; thus, price inelastic.

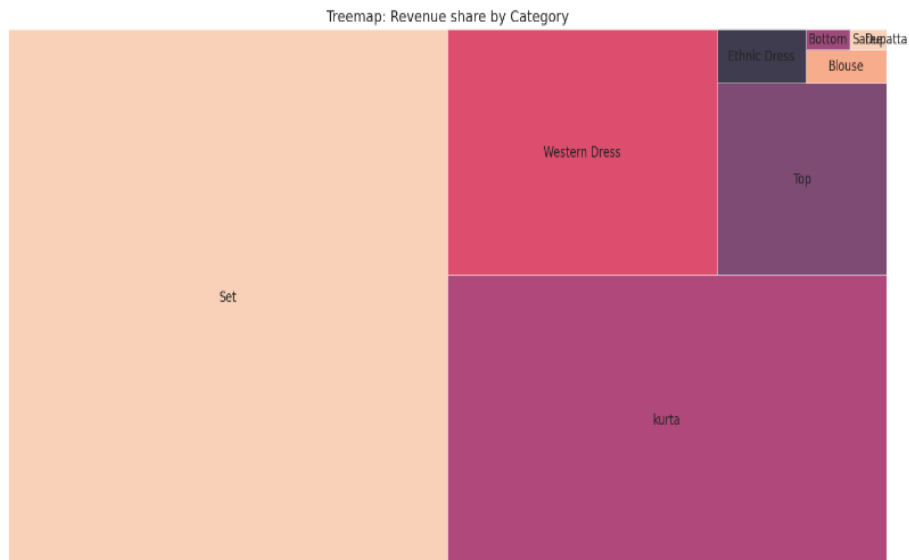


Figure 6: Treemap of Category Contribution

The treemap depicted in Figure 6 is helpful for showing the part played by each of the categories in relation to their contribution to profit. The amount of space occupied by the category in the diagram reveals the percentage of contribution made to total sales. In view of this, from Figure 6, it turns out that there are only a few categories contributing significantly, whereas a great many do not make any serious contributions at all. These results correspond to the Pareto principle, which says that it takes just a few categories to achieve success in any business. It is important to focus on successful products rather than waste time on insignificant ones.

All these results are reflected in Figure 7 which provides information regarding the distribution of total revenue and the role played by different regions in meeting demand for e-commerce products. It appears that urbanized areas become the major sources of demands, which can probably be explained by high purchasing power and availability of these regions. In addition to learning the places where the transactions should take place, the data can be helpful for deciding what steps should be taken strategically. In essence, the findings show a shift in strategy from the existing generalized way to a regionalized way considering profitability and new markets.

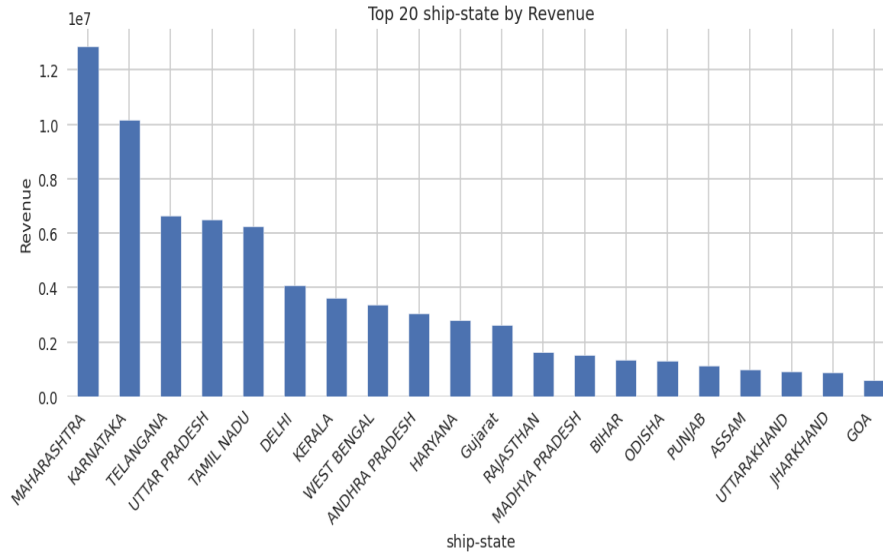


Figure 7: Geographic (Statewise) Revenue Distribution

A correlation heatmap displays connections between key numerical factors in the Amazon Sales Report dataset, including Amount (revenue), Qty (quantity), unit price, estimated profit, and margin_pct (profit margin). As shown in Figure 8 below, the data set shows extremely strong positive correlations between amount, profit, and unit price, with correlation factors approaching unity. It is natural to observe such a tendency since, at higher unit prices, revenue increases; and assuming the margins remain constant, so does profit.

At the same time, the quantity variable demonstrates a much lower correlation with the financial factors (both with amount and with profit). Moreover, it shows minimal connections with the unit price and the margin percentage variables. Thus, the number of units sold seems to be of little importance from a purely financial standpoint since the revenue depends on the pricing policy implemented by a company.

As can be seen from the figure, the profit margin variable demonstrates extremely low correlation with other financial variables. In other words, the profit margin is not significantly dependent on any other financial factors. It may imply that there is significant room for improvement in terms of increasing business performance through pricing policies.

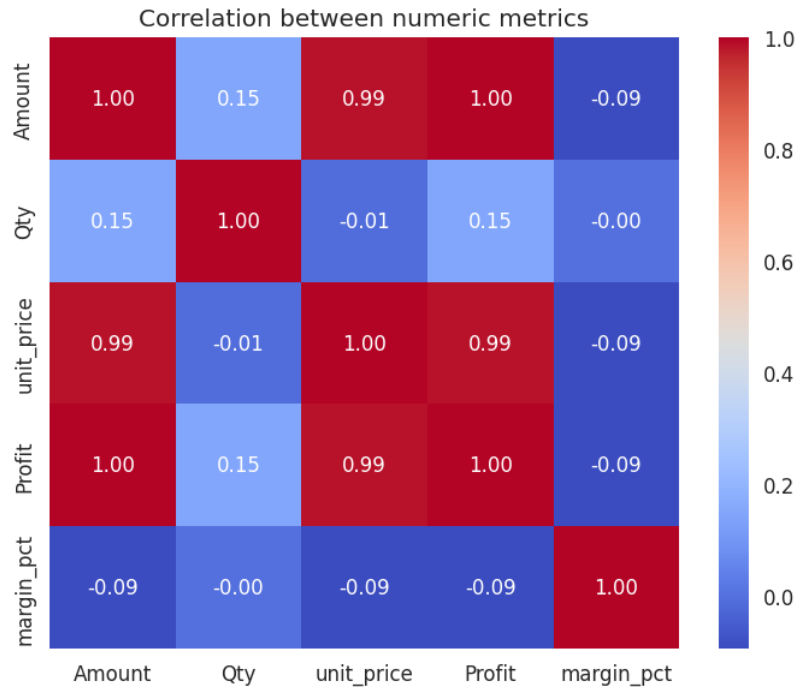


Figure 8: Correlation heatmap between numeric variables

The correlation that has been observed between income and profit is due to the use of constant margin methodology adopted in the present study, as opposed to the observation of an empirical relationship. The heatmap is presented to show the ability of the proposed Python framework to perform effective analysis and visualizations of datasets. Even if the assumptions of the dataset used may have been too simplistic, the correlation matrix has helped to show how effective the use of open source Python software is for identifying meaningful relationships within finance and providing insights into pricing and margin.

With such results, one can clearly understand how each factor i.e. pricing, sales volume, and margins impacts the overall success of e-commerce operations. However, the results shows that strategic considerations require a review of financial aspects which go beyond simple sales measurement. Therefore, these conclusions confirm that the assumptions of the proposed study framework are true, and the framework itself will help in making more reasonable decisions by means of analysis.

The above-presented research shows that despite its simplicity, this framework can successfully incorporate sales-related descriptive analytics and develop strategies based on these. As the introduction of margin analysis proves, one should pay attention not to sales measurement anymore, but focus on strategic consideration.

Overall, the combination of these results shows that sales volume is not enough to make a conclusion about the strategic performance of e-commerce operations. Volume of revenues is not the only factor determining priority strategy; concentration trends, geography of demand, and price sensitivity all help to create a bigger picture of performance. From a management point of view, this analysis provides three practical takeaways: Firstly, products with high revenues but low margins need to be considered for re-pricing or restructuring of costs; secondly, those products which show good revenues as well as profits should be targeted for growth through promotion; and lastly, geographic clustering of demand allows location-based inventory and marketing strategies. Taken together, these three takeaways demonstrate how simple visualizations can turn data into strategic decisions.

Beyond the qualitative narrative above, the framework yields concrete concentration indicators that quantify the revenue profit divide referenced throughout this section. In the analyzed data, the top 20 percent of SKUs by revenue (1,427 of 7,139 SKUs) account for approximately 72 percent of total revenue, while reaching 80 percent of total revenue requires about 28 percent of SKUs. This is a somewhat less extreme concentration than the classic 80/20 rule, and reporting the exact figure, rather than only invoking the Pareto principle qualitatively, gives managers an actionable target for how much of the catalog can realistically be deprioritized. As a basic indicator of the framework's lightweight claim, the full pipeline, from data cleaning through all eight visualizations, runs on the complete dataset of roughly 119,000 valid transaction rows in Google Colab using only pandas, Matplotlib, and Seaborn, with no paid infrastructure or proprietary software. All together, the eight visualisations discussed in this section are positive answers to the three research questions posed at the beginning of this paper: revenue-driven SKU rankings are a valid proxy for profitability (RQ1), revenue concentration in categories and geographies clearly represents an actionable prioritisation framework (RQ2), and visualisations created in Python can indeed reveal strategic insights of equivalent value to BI software (RQ3).

6. Conclusion

The present research paper has suggested a lightweight analytical approach to evaluating the performance of an e-commerce business through analyzing the example of Amazon Sales Report Dataset. With the help of proper data processing, calculations of profitability measures, and visualizations of individual SKU performance, this research has revealed how companies can move from simple sales reporting to more advanced types of analytics. It has become clear that sales volumes are not directly proportional to the levels of profitability, that only a few products and categories generate significant revenue and profitability, and that demand is highly concentrated on geographical grounds. As seen, such an approach based on rather simple analyses can contribute to decision-making on pricing and promotions, among other things. In terms of theory, this research adds to the body of knowledge regarding decision support systems and lightweight analytics in that visual analytics that are both explanatory and descriptive can be used to produce strategic outputs without employing any sort of predictive model or robust infrastructure. Rather than being seen as competition, the framework serves as an introduction into the world of data-driven strategy.

There are some limitations associated with this research paper that need to be mentioned. To start with, profitability measures have been calculated via a unified profit margin ratio because cost-level data is unavailable. In addition, the study relies on the analysis of just one Amazon Sales Report Dataset that has been made publicly available, which could be seen as a potential weakness as it is not evident whether the results are applicable elsewhere. This concern is compounded by the fact that the dataset covers a single seller on a single sales channel within one national market over a one year window, with a catalog skewed toward apparel, so the framework's applicability to other product mixes, sellers, or platforms remains to be tested empirically rather than assumed.

Lastly, the research is purely descriptive, and it does not prove the causality among the variables considered. Nevertheless, the framework proposed proves useful in comparative analysis and strategic insight generation.

The conclusions of this research validate the underlying assumption of the use of the operational sales data that has undergone an analysis and evaluation process and becomes the basis of informed decisions. Going beyond simple tracking of revenues, the proposed model takes into account the profits made in categories and individual SKUs, thus making it possible to connect the dots between operational data and strategy. Being designed with simplicity in mind, the model relies solely on open-source libraries of Python, which makes its implementation feasible and easily replicable by other researchers and practitioners.

The framework can be further enhanced in terms of forecasting using time-series approaches, taking into account cost/margin data and performing clustering analysis. Also, the use of the analytical process in containerized computing platforms like Docker would provide a level of scalability as well as facilitate near-real time decision-making. Concretely, this would involve containerizing the preprocessing and aggregation steps so that the pipeline can be scheduled to run on a recurring basis against a growing transactional database, with the visualization layer served through a lightweight dashboard rather than regenerated manually in a notebook, which is the natural path from the present descriptive script to a deployable multi channel monitoring tool.

While the current paper is purely descriptive, it shows that effective strategic analysis is not always necessarily complicated; on the contrary, a straightforward and systematic analytical process can aid in shifting from sales-oriented

analysis to strategic decision-making. Future studies should use actual cost of goods sold (COGS) figures obtained from ERP or supplier reports rather than the proxy margin for conducting profit analysis. The concept validation of this study could be improved if performed through testing with other e-commerce sites like Flipkart and/or Shopify, and geographically beyond India.

Several concrete extensions follow from the limitations discussed above. First, future work should replace the uniform proxy margin with actual cost of goods sold obtained from ERP or supplier records, which as shown in Section 4.2 would allow margins to vary by category or SKU rather than being held constant, producing a genuine profitability ranking rather than a revenue concentration diagnostic. Second, the current descriptive layer could be extended with predictive components, for example short term SKU level demand or profit forecasting, building on approaches such as the LSTM based forecasting of Tan [3] or the GRU and LightGBM stacking of Chen et al. [11], positioned as a complementary forecasting layer on top of the present diagnostic framework rather than a replacement for it. Third, validation should be extended to multi-channel and larger scale data, following the scalability discussion in Section 3, and to other platforms and geographies as already noted above. Finally, the present script based pipeline could be developed into a lightweight, deployable decision support interface, so that practitioners can move from a one time analysis notebook to a recurring monitoring tool.

Declaration of usage of generative AI and AI-assisted technology in the writing process

The authors used generative AI tools in order to enhance the quality and readability of language while preparing this paper. Following the use of this service, the author(s) edited and revised the paper accordingly.

References

1. Kumar, R., Mukherjee, S., & Choudhary, D. (2025). Uncovering the roots of customer dissatisfaction via Amazon reviews: a hybrid ensemble-deep learning approach for E-commerce quality management. *Annals of Operations Research*, 353(2), 545-574.
2. Xu, W., Cao, Y., & Chen, R. (2024). A multimodal analytics framework for product sales prediction with the reputation of anchors in live streaming e-commerce. *Decision Support Systems*, 177, 114104..
3. Tan, Y. (2025). E-commerce Sales Prediction and Analysis Based on Multi-layer LSTM and Interactive Visualization Techniques. *International Journal of High Speed Electronics and Systems*, 2540898.
4. Dhage, R., Nehete, S., Hon, S., Patankar, T., & Kale, L. (2023, August). Implementation of recommendation system's service model using amazon E-commerce dataset. In *International Conference on ICT for Sustainable Development* (pp. 415-426). Singapore: Springer Nature Singapore.
5. Zhang, J., Yu, Q., Chen, Y., Zhou, G., Liu, Y., Sun, Y., ... & Yu, H. (2024, July). An E-Commerce Dataset Revealing Variations during Sales. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1162-1171).
6. Kalla, D. (2024). *Improving E-Commerce Organization Performance Using Big Data Analytics and Artificial Intelligence* (Doctoral dissertation, Colorado Technical University).
7. Rahman, M. M. (2024). Data-Driven Decision-Making Through Customer Relationship Management: A Systematic Literature Review In Modern Enterprises. *American Journal of Advanced Technology and Engineering Solutions*, 4(03), 30-59.
8. Varsha, P. S., & Karan, A. (2023). Descriptive analytics and data visualization in e-commerce. In *Handbook of Big Data Research Methods* (pp. 86-104). Edward Elgar Publishing.
9. Zhang, C., Wang, X., Zhao, C., Ren, Y., Zhang, T., Peng, Z., ... & Li, Q. (2022). Promotionlens: Inspecting promotion strategies of online e-commerce via visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 29(1), 767-777.
10. Sardar, M. S. A. Z. (2023). *Amazon Sales Report* [Dataset]. Kaggle. <https://www.kaggle.com/datasets/mdsazzatsardar/amazonsalesreport>
11. Chen, Y., Xie, X., Pei, Z., Yi, W., Wang, C., Zhang, W., & Ji, Z. (2024). Development of a time series e-commerce sales prediction method for short-shelf-life products using GRU-LightGBM. *Applied Sciences*, 14(2), 866.
12. Huacre, K., Daza, A., & Mejia, E. (2021). E-commerce system for the sales process in the company: A systematic review. *Journal of System and Management Sciences*, 11(4), 113-126.

13. Irawan, B. H., Prihadi, D., Rahmatika, D. N., Nugroho, C., & Waskita, J. (2024). Application Development Model E-Commerce in Traditional Markets Using the TOGAF Framework. *JOIV: International Journal on Informatics Visualization*, 8(3), 1253-1259.
14. Zhang, W., Li, W., Rao, X., Zou, L., Luo, X., Zhuang, C., ... & Zheng, B. (2025, July). AIGuard: A Benchmark and Lightweight Detection for E-commerce AIGC Risks. In *Findings of the Association for Computational Linguistics: ACL 2025* (pp. 12437-12450).
15. Shyaa, G. S., & Al-Zubaidie, M. (2023). Utilizing trusted lightweight ciphers to support electronic-commerce transaction cryptography. *Applied Sciences*, 13(12), 7085.
16. Pan, C. L., Liu, Y., & Pan, Y. C. (2022). Research on the status of e-commerce development based on big data and Internet technology. *International Journal of Electronic Commerce Studies*, 13(2), 27-47.
17. Wu, P. J., & Lin, K. C. (2018). Unstructured big data analytics for retrieving e-commerce logistics knowledge. *Telematics and Informatics*, 35(1), 237-244.
18. Alrumiah, S. S., & Hadwan, M. (2021). Implementing big data analytics in e-commerce: Vendor and customer view. *Ieee Access*, 9, 37281-37286.
19. Zineb, E. F., Najat, R. A. F. A. L. I. A., & Jaafar, A. B. O. U. C. H. A. B. A. K. A. (2021). An intelligent approach for data analysis and decision making in big data: a case study on e-commerce industry. *International Journal of Advanced Computer Science and Applications*, 12(7).
20. Behl, A., Dutta, P., Lessmann, S., Dwivedi, Y. K., & Kar, S. (2019). A conceptual framework for the adoption of big data analytics by e-commerce startups: a case-based approach. *IseB* 17 (2): 285–318.
21. Sekar, S. P. (2025). Deployment Automation for Scalable and Resilient E-Commerce Platforms: A Retail Sector Perspective. *Journal of Computer Science and Technology Studies*, 7(3), 683-689.
22. Mohanty, S. N., Chatterjee, J. M., Jain, S., Elngar, A. A., & Gupta, P. (Eds.). (2020). *Recommender system with machine learning and artificial intelligence: practical tools and applications in medical, agricultural and other industries*. John Wiley & Sons.
23. Rao, K. S., Gopathoti, S., Ramakrishna, A., Gupta, P., Potluri, S., & Reddy, G. S. (2023, October). Assessing Machine Learning Algorithms for Customer Segmentation: A Comparative Study. In *International Semantic Intelligence Conference* (pp. 221-229). Singapore: Springer Nature Singapore.
24. Yadav, P., Gupta, P., Rai, P., Naik, N., & Kasip, K. (2023). Exploring the factors influencing the adoption and continuous engagement in unlocking the potential of technology driven chatbots in banking and financial institutions. *Engineered Science*, 28, 1054.
25. Yadav, P., Gupta, P., Sijariya, R., & Sharma, Y. (2024). Artificial intelligence in risk management. *Artificial intelligence for risk mitigation in the financial industry*, 1-25.
26. Gupta, J., Nayyar, A., & Gupta, P. (2015). Security and privacy issues in internet of things (IoT). *International Journal of Research in Computer Science*, 2, 18-22.