Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 09 , September - 2026

Sales Performance Dashboard with Time-Series Forecasting:A Data-Driven Analytics Platform for Business Intelligence

Raj Rahul Pimpalkar[1], Prof. Abhiraj Ubale[2]
[1]Post Graduate Diploma in Data Science & Artificial Intelligence, COEP Technological University, Pune, India
[2]Department of Computer Engineering, COEP Technological University, Pune, India
Corresponding Author: Raj Rahul Pimpalkar

Abstract - Sales organizations generate large volumes of transactional data, but converting those records into timely managerial insight requires an integrated analytical workflow. This paper presents a reproducible sales analytics platform that combines synthetic transaction generation, relational data storage, exploratory analysis, dashboard visualization, and short-horizon time-series forecasting. The implementation processes 17,820 transaction records covering January 2024 to August 2025 across five regions, five product categories, four sales channels, 25 products, and 10 sales representatives. The analytical database stores 17 transaction attributes and is queried through SQL views and aggregations, while Python and Pandas are used for data processing and Matplotlib for publication-quality visualizations. The analysis reports total revenue of ₹626.69 million, total profit of ₹205.98 million, a 32.87% profit margin, and an average customer rating of 4.32/5. Electronics contributes 52.8% of revenue, the North region contributes 25.8%, and the Online channel contributes 39.9%. For forecasting, a transparent decomposition-based method extracts a three-month centered trend, estimates monthly seasonal deviations, extrapolates the recent trend linearly, and recombines trend and seasona lity for six future months. The resulting forecast for September 2025 to February 2026 totals ₹144.24 million. A fixed ±15% uncertainty band is reported as a heur istic range rather than as a statistically estimated confidence interval. The results demonstrate how open-source tools can support an end-to-end sales analytics workflow while also highlighting the limitations of short synthetic time series and the need for out-of-sample validation on real business data.

Keywords - Sales Analytics, Business Intelligence, Time-Series Forecasting, Exploratory Data Analysis, MySQL, Python, Data Visualization, Trend Decomposition, Retail Analytics

1. INTRODUCTION

Sales analytics is increasingly used to convert transactional records into information that supports inventory planning, marketing allocation, pricing decisions, regional strategy, and performance management. The challenge is not only the availability of data but also the ability to organize, analyze, visualize, and forecast it in a single workflow. Manual spreadsheet-based reporting can require repeated data preparation and often separates historical reporting from forward-looking analysis.

This work develops an end-to-end sales analytics prototype using Python, MySQL, and Matplotlib. The system begins with reproducible synthetic transaction generation, stores the resulting records in a relational database, performs multidimensional aggregation and exploratory analysis, generates a five-panel dashboard, and produces a six-month revenue forecast. The design emphasizes transparency so that each stage—from data generation to forecast recomposition—can be inspected and reproduced.

The contribution of the work is therefore practical rather than a claim of a new forecasting algorithm. The study demonstrates how a compact open-source stack can combine database-backed sales analytics with an interpretable forecasting workflow. Because the dataset is synthetic and contains only 20 monthly observations, the results are treated as a prototype demonstration rather than evidence of production forecasting accuracy.

2. RELATED WORK

Dashboard design is a central component of business intelligence because it compresses large quantities of information into a small number of decision-oriented visual elements. Few [1] emphasizes visual hierarchy and the presentation of the most important information in a form that can be monitored quickly. Commercial platforms such as Tableau and Power BI provide mature interactive analytics, whereas Python libraries provide a flexible open-source alternative for custom analytical applications.

Time-series forecasting methods range from classical statistical approaches to machine-learning and deep-learning models. Box and Jenkins [4] established the ARIMA framework for autoregressive and moving-average modeling. Exponential smoothing methods are effective when level, trend, and seasonality can be represented directly [2]. Prophet uses additive trend and seasonal components with an emphasis on practical business forecasting [3]. Tree-based methods such as XGBoost [6] and sequence models such as LSTM [7] can model richer nonlinear relationships when adequate historical data and explanatory features are available.

For the present study, a classical decomposition approach is used because the available series contains only 20 monthly observations and the project emphasizes interpretability. The method is intentionally simple: the trend is smoothed, monthly seasonal deviations are estimated, the recent trend is extrapolated, and the components are recombined. This makes the forecast logic easy to inspect, although it also limits the statistical strength of the resulting predictions.

3. METHODOLOGY

3.1 System Architecture

The proposed workflow is organized into data generation, storage, analysis, visualization, and forecasting stages. Figure 1 shows the complete flow. Raw or simulated data is prepared in Python and exported to CSV, loaded into MySQL, and then queried for analytical aggregation. Python-based EDA and visualization operate on the aggregated or transaction-level data, while the forecasting module consumes monthly revenue values and produces a six-month forecast.

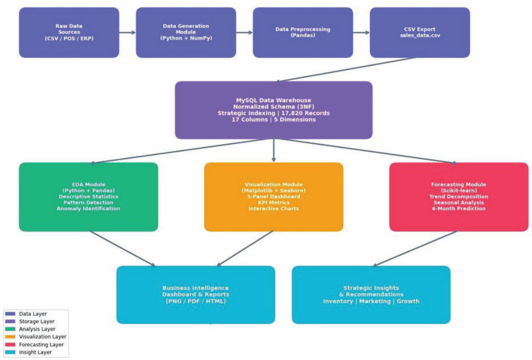

Fig. 1. System architecture of the sales analytics and forecasting platform.

3.2 Data Generation and Preprocessing

The dataset contains 17,820 synthetic transactions from January 2024 through August 2025. A fixed random seed of 42 is used for reproducibility. Five regions are represented with weighted probabilities: North 25%, South 20%, East 20%, West 20%, and Central 15%. The product catalog contains 25 products distributed across Electronics, Clothing, Home & Kitchen, Sports, and Books. Four sales channels and ten sales representatives are included.

Business patterns are deliberately embedded in the generator. Transactions receive a 1.4× multiplier during Q4, a 0.75× multiplier during Q2, and a 1.3× weekend multiplier. Pricing, quantity, discount, cost, profit, and customer rating are generated from probabilistic rules. The resulting CSV is validated for data types, date fields, derived columns, and missing values before database loading.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 09 , September - 2026

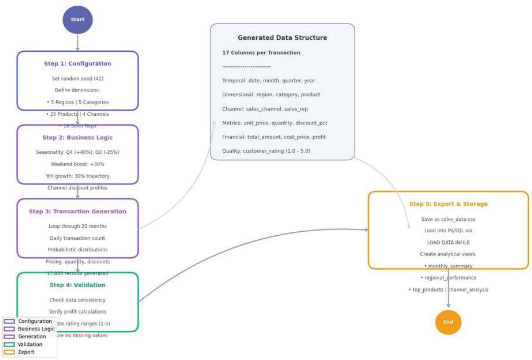

Fig. 2. Data generation and validation process.

3.3 Database and Analytical Layer

The implementation uses a MySQL database with a transactions table containing 17 attributes: transaction identifier, date, month, quarter, year, region, category, product, sales channel, sales representative, unit price, quantity, discount percentage, transaction amount, cost price, profit, and customer rating. Indexes are defined on time and frequently filtered dimensions. Four analytical views—monthly summary, regional performance, top products, and channel analysis—are used to simplify repeated dashboard queries.

The thesis also includes an entity-relationship representation with a transaction fact-like table and dimension entities for time, region, category, product, channel, and sales representative (Fig. 3). In the implementation description, however, the project states that the 17 attributes are stored in a single transactions table. The figure is therefore treated here as a conceptual analytical model rather than as a claim that the deployed database physically contains separate dimension tables.

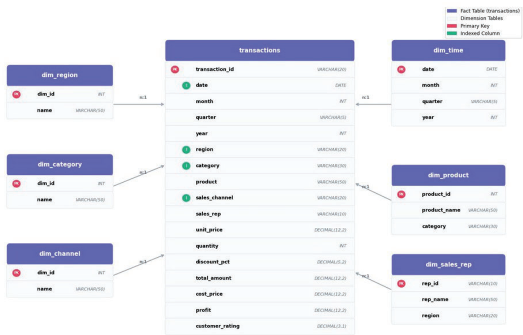

Fig. 3. Conceptual entity-relationship representation used for sales analysis.

3.4 Forecasting Method

Monthly revenue is first aggregated from the transaction data. The forecasting procedure then follows five steps. First, a three-month centered rolling mean is calculated to smooth short-term fluctuations and extract the trend. Second, monthly seasonal effects are estimated as the mean deviation between observed revenue and the extracted trend for each calendar month. Third, the slope of the most recent three trend values is used for linear extrapolation. Fourth, the future trend is recombined with the corresponding seasonal component. A floor equal to 50% of the historical mean is applied to avoid unrealistically low forecasts.

The final step applies a symmetric ±15% uncertainty band to each point forecast. Because this band is fixed rather than estimated from a sampling distribution or forecast-error model, it is reported in this paper as a heuristic uncertainty range, not as a formal 95% statistical confidence interval.

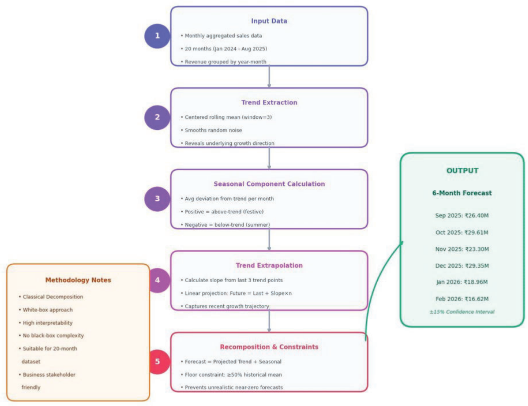

Fig. 4. Forecasting methodology flowchart.

4. IMPLEMENTATION

The system is implemented with Python 3.x using Pandas and NumPy for data engineering, Matplotlib for visualization, and Scikit-learn for supporting metrics and feature-processing utilities. MySQL 8.0 or MariaDB 10.6 is used for relational storage. The development environment is a standard workstation with at least 8 GB RAM and a modern Intel Core i5/AMD Ryzen 5 class processor.

The visualization module produces five analytical views: monthly revenue and profit trend, regional revenue distribution, category revenue and profit comparison, channel revenue comparison, and top-product ranking. These views are assembled into a static dashboard image. Although the project objective originally describes an interactive dashboard, the implemented version is static; interactive filtering and drill-down are identified as future work.

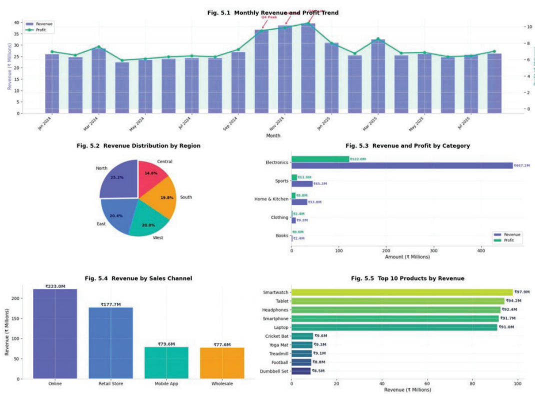

Fig. 5. Five-panel sales performance dashboard generated from the analytical dataset.

5. RESULTS AND DISCUSSION

5.1 Dataset and Key Performance Indicators

The generated dataset contains 17,820 transactions and 44,326 units sold. Total revenue is ₹626.69 million and total profit is ₹205.98 million, producing a profit margin of 32.87%. The average customer rating is 4.32/5.0. The results are internally consistent with the transaction-level aggregates reported in the project.

Table 1. Dataset summary statistics.

Metric	Value
Transactions	17,820
Period	Jan. 2024 – Aug. 2025
Revenue	₹626.69 M
Profit	₹205.98 M
Profit margin	32.87%
Units sold	44,326
Products	25
Sales representatives	10
Average rating	4.32 / 5.0

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 09 , September - 2026

5.2 Regional, Category, and Channel Analysis

The North region records the highest revenue at ₹161.93 million, representing 25.8% of total revenue. Central contributes the lowest share at 14.9%. Electronics is the dominant category with ₹331.10 million, or 52.8% of total revenue, followed by Sports at ₹162.24 million. Online sales contribute ₹249.95 million, or 39.9%, making Online the largest sales channel.

Table 2. Regional performance.

Region	Revenue (₹M)	Share	Profit (₹M)
North	161.93	25.8%	53.21
South	125.19	20.0%	41.14
East	124.46	19.9%	40.91
West	121.59	19.4%	39.97
Central	93.52	14.9%	30.75

Table 3. Category-wise revenue and profit.

Category	Revenue (₹M)	Share	Profit (₹M)
Electronics	331.10	52.8%	108.87
Sports	162.24	25.9%	53.35
Home & Kitchen	102.29	16.3%	33.63
Clothing	22.30	3.6%	7.33
Books	8.74	1.4%	2.87

Table 4. Sales channel performance.

Channel	Revenue (₹M)	Share	Transactions
Online	249.95	39.9%	7,128
Retail Store	199.96	31.9%	5,346
Mobile App	94.13	15.0%	2,673
Wholesale	82.65	13.2%	2,673

5.3 Forecast Results

The decomposition model generates forecasts for September 2025 through February 2026. The six-month total is ₹144.24 million and the mean monthly forecast is ₹24.04 million. October and December are projected at approximately ₹29–30 million, reflecting the modeled festive-season behavior, while January and February decline to approximately ₹19 million and ₹16.62 million.

Table 5. Six-month sales forecast.

Month	Forecast (₹M)	Lower (₹M)	Upper (₹M)
Sep 2025	26.40	22.44	30.37
Oct 2025	29.61	25.17	34.05
Nov 2025	23.30	19.81	26.80
Dec 2025	29.35	24.95	33.76
Jan 2026	18.96	16.12	21.81
Feb 2026	16.62	14.13	19.11
Total	144.24	—	—

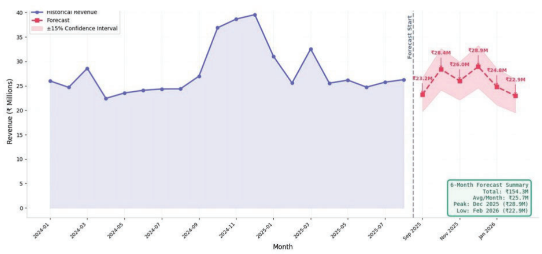

Fig. 6. Six-month sales forecast with the reported ±15% heuristic uncertainty band.

5.4 Business Implications

Five practical implications emerge from the analysis. First, Electronics is highly concentrated and therefore requires both continued investment and diversification into other categories. Second, the lower contribution of Central indicates a need to investigate regional demand and market coverage. Third, the Mobile App channel represents a potential growth area because its revenue share is below Online and Retail Store. Fourth, the recurring seasonal pattern can be used to align inventory and marketing activity with expected demand. Fifth, the 4.32/5 customer rating can be used as a supporting signal in customer-facing campaigns.

These implications should be interpreted as decisions suggested by the synthetic dataset rather than validated business findings. A real deployment would require external factors, actual customer behavior, product availability, promotions, and operational constraints.

6. LIMITATIONS

The principal limitation is the short historical window of 20 monthly observations. A 12-month seasonal period leaves relatively few repeated seasonal cycles for reliable estimation. The dataset is also synthetic, so the embedded Q4, Q2, weekend, and growth patterns are partly design assumptions rather than naturally observed business behavior. The forecasting model assumes a stable seasonal pattern and a locally linear trend and does not use exogenous variables such as marketing campaigns, competitor pricing, macroeconomic conditions, or supply disruptions.

The current study also does not perform a proper out-of-sample or rolling-origin forecast evaluation. Consequently, no empirical MAPE, MAE, RMSE, or sMAPE should be interpreted as validated predictive accuracy. The ±15% range is a heuristic uncertainty band. Future work should use a real or public sales dataset, establish a temporal hold-out set, compare the decomposition model with SARIMA/ETS/Prophet and feature-based machine-learning models, and report forecast errors using the same evaluation protocol.

7. CONCLUSION AND FUTURE WORK

This paper presents an end-to-end sales analytics workflow that integrates synthetic data generation, MySQL-based storage, exploratory analysis, dashboard visualization, and interpretable time-series forecasting. On 17,820 generated transactions, the system identifies a total revenue of ₹626.69 million, a 32.87% profit margin, Electronics as the largest revenue category, North as the leading region, and Online as the leading channel. The decomposition model produces a six-month forecast of ₹144.24 million for September 2025 to February 2026.

The main value of the system is reproducibility and transparency rather than algorithmic novelty. The implementation shows how a small open-source stack can support the complete path from transactional data to business-oriented insight. Future enhancements include integration with real POS/ERP data, statistically validated forecasting comparisons, interactive Streamlit or Plotly Dash deployment, anomaly detection, RFM customer segmentation, MLOps automation, and natural-language analytics.

ACKNOWLEDGMENT

The author gratefully acknowledges Prof. Abhiraj Ubale, Department of Computer Engineering, COEP Technological University, Pune, for guidance and feedback throughout the project. The author also acknowledges the faculty and peers who supported the academic work and the development of the analytical prototype.

REFERENCES

[1] S. Few, Information Dashboard Design: The Effective Visual Communication of Data, O'Reilly Media, 2006.

[2] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed., OTexts, 2021.

[3] S. J. Taylor and B. Letham, "Forecasting at Scale," The American Statistician, vol. 72, no. 1, pp. 37–45, 2018.

[4] G. E. P. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, 1970.

[5] R. Kimball and M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd ed., Wiley, 2013.

[6] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. KDD, pp. 785–794, 2016.

[7] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.