INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

An End-to-End Fundamental Data Analytics Framework for Enhancing Decision-Making in SMEs

1st Krithik Subramanian
Department of Data Science and Business Analytics
Vels Institute of Scinece, Technology & Advanced Studies
Chennai, India
Krithikvs711@gmail.com

2nd Ms. S. Nithya Priya
Department of Data Science and Business Analytics
Vels Institute of Scinece, Technology & Advanced Studies
Chennai, India
nithya.scs@vistas.ac.in

Abstract

Small and medium-sized businesses generate substantial volumes of sales data; however, this data is often underutilized and not systematically analysed to identify trends or support forecasting. As a result, key decisions related to sales planning and customer management frequently rely on conventional reporting rather than data-driven insights.

This study applies fundamental data analytics techniques to improve sales optimization through customer segmentation and sales forecasting. Simple and practical techniques such as rule-based classification, moving averages, and the use of Python with relevant data analysis libraries can support the complete process of data analysis to ensure data quality and reliable forecasting. Graphical visualisation through Power BI provides clear insights into sales performance, regional patterns, and customer segments.

Overall, the study demonstrates that the effective use of simple and cost-efficient data analytics tools can significantly enhance sales outcomes for small and medium-sized businesses without requiring large investments.

Introduction

Customer behaviour varies across individuals, regions, and time periods, making it essential for businesses to analyse sales data to identify meaningful patterns and trends. However, many small and medium-sized enterprises rely primarily on conventional reporting methods and do not fully utilise their available data for analytical insights and decision-making.

This study is conducted using data from Surag Instruments, a small- to medium-scale trading firm in India that has been operating for over 15 years. The company deals in industrial instruments sourced both internationally and locally and maintains detailed sales records, including invoice dates, customer details, product categories, quantities sold, pricing, and transaction values. Despite the availability of such comprehensive data, it has not been systematically analysed for strategic insights. The dataset used for this study covers the period from 2019 to 2025, with annual sales ranging approximately between ₹10 lakhs and ₹2 crores. Customer identities have been kept confidential to ensure privacy.

Overall, existing research emphasizes that the use of simple and fundamental data analytics techniques can lead to meaningful improvements in decision-making. Such approaches are particularly valuable for small and medium-sized businesses, where resource constraints make cost-effective and practical solutions essential.

Objectives of the Study

1. To support SMEs in making effective, data-driven decisions
2. To apply fundamental data analytics techniques to analyse historical sales data in SMEs
3. To classify customers based on purchasing behaviour and revenue contribution
4. To estimate short-term sales trends using basic analytical methods
5. To present insights through clear graphical visualisations

Scope of the Study

This study is limited to small and medium-sized enterprises (SMEs) and focuses only on the application of fundamental data analytics techniques. It does not involve advanced machine learning models, and emphasizes simple, practical methods that can be easily implemented for business decision-making.

Literature Survey

In recent years, deriving insights from data has become increasingly important in business, particularly for understanding customer behaviour and supporting future planning. The growing emphasis on data-driven decision-making highlights the need for analytical approaches even in small and medium-sized enterprises (SMEs).

According to *Thomas H. Davenport and Jeanne G. Harris (2007)*[1], organizations that effectively apply analytical techniques are more likely to gain a competitive advantage through informed decision-making. Their work indicates that even relatively simple analytical methods, when applied systematically, can significantly enhance the quality and consistency of business decisions.

In the context of customer segmentation, *Konstantinos Tsiptsis and Antonios Chorianopoulos (2011)*[2] suggest that rule-based and behaviour-driven segmentation approaches are particularly suitable for SMEs. These methods are considered practical because they are easy to implement, transparent in logic, and readily interpretable by business users.

Sales forecasting is another key area widely discussed in business analytics literature, as it plays a crucial role in planning and operational control. *Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman (1998)*[3] highlight that simple time-series techniques, such as moving averages and trend analysis, can be effective for short-term forecasting, especially when data availability is limited and patterns remain relatively stable. These approaches are often recommended for smaller organizations due to their simplicity and reliability.

According to *Wes McKinney* ***(2017)***[4], Python provides a simple and flexible environment for handling, processing, and analysing structured data through libraries such as Pandas and NumPy. The language's readability and extensive ecosystem make it particularly suitable for beginners as well as small and medium-sized businesses that require cost-effective and easy-to-implement analytical solutions. According to *Danish Mahmud and Md. Zafor Ikbal (2024)*[5], Power BI plays a significant role in transforming raw data into meaningful insights by enabling interactive dashboards, real-time reporting, and integration with advanced analytics tools.

These existing studies support the aim of the present study.

Proposed Methodology

This study adopts a structured and fundamental data analytics approach, focusing on simplicity and clarity, making it suitable for small and medium-sized trading organizations with limited analytical resources.

The process begins with data collection and understanding, using historical sales records that include invoice details, dates, customer information, product categories, quantities, pricing, and total sales values. Initial exploration helps identify the structure of the dataset and detect issues such as missing values, duplicates, and inconsistencies.

Data pre-processing ensures quality and reliability by cleaning and transforming the dataset. Missing values are handled, duplicate records are removed, and irrelevant fields are excluded. Data is standardized and aggregated, while feature engineering is applied to derive key variables such as total customer revenue, purchase frequency, and sales trends.

Customer segmentation is performed using a rule-based approach, classifying customers into high-, medium-, and low-value groups based on purchase value and frequency. Sales forecasting is carried out using a three-year moving average method, which helps estimate short-term trends in a stable environment.

Python and its data analysis libraries support the analytical processes, ensuring accuracy and consistency. Graphical visualisation through Power BI can present insights on customer segments, regional performance, sales trends, and forecasted values, enabling informed decision-making.

Architecture Diagram: The architecture structured into the following components:

Methodology

To ensure a systematic approach, the model is divided into a series of structured phases.

Phase 1: Data Collection and Understanding

Data is central to this study. It is collected from the organization and includes invoice dates, customer details, product categories, quantities, pricing, and total sales values. Initial exploration helps understand the dataset and identify basic issues. The output:

- Organized raw dataset
- Understanding of key variables

Phase 2: Data Pre-processing and Feature Engineering

The collected data is unstructured and requires cleaning. Missing values, duplicates, and inconsistencies are handled, and date formats are standardized. New features such as total customer revenue, purchase frequency, and yearly trends are created. The output:

- Cleaned dataset
- Derived features

Phase 3: Customer Segmentation

Customers are grouped into high-, medium-, and low-value categories using a simple rule-based approach based on purchase value and frequency. The output:

- Customer segments
- Revenue-based classification

Phase 4: Sales Forecasting
Sales data is aggregated yearly. A moving average method is applied to identify trends and estimate future sales values. The output:

- Year-wise trend analysis
- Estimated future sales

Phase 5: Visualization and Interpretation
Graphical visualisations are created to present customer segments, regional performance, and sales trends. These help in better understanding and decision-making. The output:

- Charts and visuals
- Actionable insights

Each phase helps convert raw data into useful insights for better business decisions.

Input

The following attributes are taken as input for the study:

- **Invoice Date:** Indicates the date of each transaction and helps in analysing time-based patterns such as yearly sales trends.
- **Region:** Represents the location of the customer and supports region-wise comparison of sales performance.
- **Customer Name:** Actual name of the customer is not mentioned. Instead the customers are numbers to maintain
- **Product Name:** Denotes the type of product sold and helps in understanding how sales are distributed across different categories.
- **Quantity Sold:** Refers to the number of units sold in each transaction and contributes to the overall sales volume.
- **Unit Price:** Indicates the price per unit and, along with quantity, helps in calculating total sales value.

These input attributes form the base for further analysis, including total sales calculation, feature creation, customer segmentation, and sales forecasting.

Pseudo code and Implementation

START

Load historical sales dataset

// Data Pre-processing

Remove duplicate records

Handle missing values

Convert invoice date to a standard date format

// Data Analysis

Compute Total Sales = Quantity Sold × Unit Price

Group data by Customer Name

Calculate Total Revenue for each customer

Determine Purchase Frequency for each customer

Aggregate data to obtain yearly sales totals

// Customer Segmentation

FOR each customer

IF Total Revenue ≥ High Value Threshold

Assign to High-Value Segment

ELSE IF Total Revenue ≥ Medium Value Threshold

Assign to Medium-Value Segment

ELSE

Assign to Low-Value Segment

END FOR

// Sales Forecasting

Apply moving average method to yearly sales data

Generate predicted sales values

// Output

Save processed results

Create charts and reports for analysis

END

The system is implemented using Python for data analysis and Power BI for visualization. Feature engineering techniques are applied to compute total sales, customer-wise revenue, purchase frequency, and yearly sales figures. Customer segmentation is carried out using a rule-based approach, where customers are classified into high-value, medium-value, and low-value categories based on predefined revenue thresholds. Sales forecasting is achieved by aggregating historical data on a yearly basis and applying a moving average technique to identify trends and estimate future sales.

Output

The outputs obtained are as expected and include both numerical results and graphical representations. Following outputs to support analysis and decision-making:

Customer Segmentation Summary: Shows the classification of customers into high-value, medium-value, and low-value groups based on their purchasing behaviour and revenue contribution.

Year-wise Sales Trend Analysis: Presents sales performance over different years, helping to understand trends, growth patterns, and possible variations.

Forecasted Sales Values: Provides estimated future sales figures using the moving average method, supporting planning and target setting.

Region-wise and Customer-wise Sales Reports: Displays how sales are distributed across regions and among customers, helping to assess performance and identify opportunities for improvement.

Interactive Visuals and Charts: Graphical representations created using Power BI that present data through charts and graphs for easier understanding and better insights.

Results and Discussion

The results of the study show that sales analysis in SMEs can be effectively done using simple data analytics techniques. By working with historical data, raw transaction details are converted into meaningful information that supports better business decisions.

Customer segmentation groups customers into high-, medium-, and low-value categories based on their contribution. It is observed that a smaller set of customers contributes a larger share of revenue, while many others contribute less. This helps the business understand where to focus more attention and effort.

Sales trend analysis shows changes in performance over the years, helping to identify patterns and variations. The moving average method gives a simple estimate of future sales, which can support planning and setting realistic targets.

Region-wise and customer-wise analysis highlights differences in performance across areas and customer groups. Graphical representations make the results easier to understand and help in quickly identifying trends and patterns.

Overall, the study shows that even simple and basic data analysis methods can provide economical, useful insights and improve decision-making in SMEs without the need for complex tools

Recommendations

- SMEs can use simple data analytics techniques with existing employees, without investing in costly tools.
- Regular analysis and forecasting can improve sales planning and overall performance.

Screenshots

FULL DATASET:

	InvoiceDate	Region	CustomerName	...	Quantity	UnitPrice	SalesAmount
0	2019-01-15	South India	Customer 1	...	5	8000	40000
1	2019-06-10	West India	Customer 2	...	3	12000	36000
2	2020-02-20	South India	Customer 1	...	6	15000	90000
3	2020-09-18	North India	Customer 3	...	4	10000	40000
4	2021-03-25	East India	Customer 4	...	7	9000	63000
5	2021-11-12	South India	Customer 2	...	5	13000	65000
6	2022-04-30	West India	Customer 5	...	8	8500	68000
7	2022-10-14	North India	Customer 3	...	6	16000	96000
8	2023-05-05	South India	Customer 1	...	4	11000	44000
9	2023-12-22	East India	Customer 4	...	3	9500	28500
10	2024-01-19	South India	Customer 2	...	5	14000	70000

Fig.1: Cleaned and Pre-processed Dataset Sample

Year-wise Sales Data:

	Year	Sales
0	2019	1050000
1	2020	1200000
2	2021	1450000
3	2022	1700000
4	2023	2100000
5	2024	2850000
6	2025	2400000

Fig 2**:** Year-wise Sales Analysis (Python Output)

Fig 3: Power BI output year-wise sales

Fig 4: Using Power BI

Year	Sales	3-Year Moving Total	3-Year Moving Average
2019	950000	-	-
2020	1450000	4080000	1360000
2021	1680000	5050000	1683333
2022	1920000	5750000	1916667
2023	2150000	6920000	2300000
2024	2850000	7400000	2466667
2025	2400000	-	-

Fig 5: Trend Analysis (Python output)

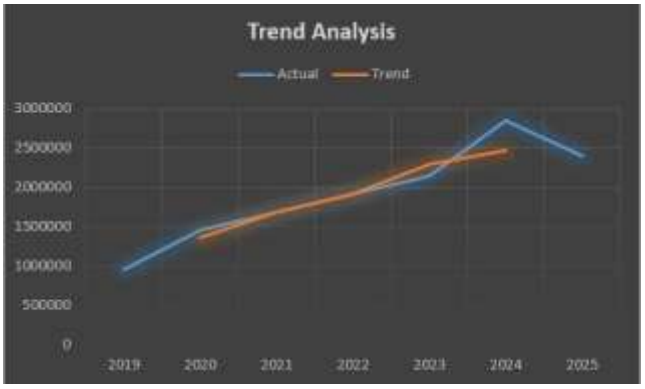

Fig 6: Power BI output

Region-wise Sales:

	Region	Sales
0	North	750000
1	South	950000
2	East	600000
3	West	550000

Fig 7: Region-wise sales (Python output)

```
Segment Count:

Segment
Low Value       18
High Value      10
Medium Value    10
Name: count, dtype: int64


Segment Revenue:

Segment
High Value      1345000
Low Value        487000
Medium Value     725000
Name: Sales, dtype: int64
```

Fig 8: Customer Segmentation (Python output)

Fig 9: Customer Segmentation (Power BI output)

```
FORECASTED SALES (2026):

Year
2026    2466667

Name: ForecastedSales, dtype: float64
```

Fig 10: Forecast (Python output)

Conclusion

This study shows that SMEs must treat their data as a valuable resource. By studying their data regularly, businesses can better understand customer behaviour, improve sales strategies, and make informed decisions. Customer segmentation helps identify key customers, while forecasting supports better planning and target setting.

The study demonstrates an end-to-end use of fundamental data analytics, where raw data is collected, processed, analysed, and converted into meaningful insights. Even simple techniques can effectively transform data into useful information without the need for complex methods. Tools like Python and visualisation platforms help present results clearly and support decision-making.

Data analytics, which combines programming, statistics, and analytical thinking, is becoming an essential requirement for organizations and their employees. Overall, the study highlights that effective use of data can help SMEs improve performance, plan better, and grow in a structured manner.

References

[1] Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business School Press.

[2] Tsiptsis, K., & Chorianopoulos, A. (2011). *Data mining techniques in CRM: Inside customer segmentation*. Wiley.

[3] Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). Wiley.

[4] McKinney, W. (2017). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media.

[5] Mahmud, D., & Ikbal, M. Z. (2024). Power BI and data analytics in financial reporting: A review of real-time dash boarding and predictive business intelligence tools. *International Journal of Scientific Interdisciplinary Research, 5*(2), 125–157.

[6] Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

[7] Kotler, P., & Keller, K. L. (2016). *Marketing management* (15th ed.). Pearson Education.

[8] Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). *Data mining for business analytics: Concepts, techniques, and applications in Python* (1st ed.). Wiley.