



## SUPERMARKET BILLING DATA INSIGHTS

<sup>1</sup>P.Srinu, <sup>2</sup>P.Sagar, <sup>3</sup>G.Abhinay kumar, <sup>4</sup>K.Yugendher chari

<sup>1</sup>Assistant Professor, <sup>234</sup>Students

Department of CSE(Data Science)

Siddhartha institute of technology & sciences,narapally

[srinu.p@siddhartha.co.in](mailto:srinu.p@siddhartha.co.in), [23TQ1A6719@siddhartha.co.in](mailto:23TQ1A6719@siddhartha.co.in),  
[23TQ1A6755@siddhartha.co.in](mailto:23TQ1A6755@siddhartha.co.in), [23TQ1A6726@siddhartha.co.in](mailto:23TQ1A6726@siddhartha.co.in)

### ABSTRACT

Supermarket businesses generate a large amount of billing data every day through customer purchases. Analyzing this data can provide useful insights into sales patterns, product demand, and customer purchasing behavior. The main objective of this project titled “Supermarket Billing Data Insights” is to analyze supermarket transaction data and extract meaningful information that can help improve business decisions. In this project, billing data is analyzed using Python and the Pandas library for data manipulation and analysis. The dataset contains information such as product name, quantity, price, total sales, and transaction details. By applying data analysis techniques, the project identifies popular products, total revenue, and sales trends. The analysis helps in understanding which products sell the most, how sales vary over time, and how supermarkets can manage inventory more effectively. The results of this project demonstrate how data analysis can transform raw billing data into useful insights that support better decision-making and improve overall business performance.

### INTRODUCTION

The Supermarket Billing Data Insights project is a data analysis study focused on understanding sales patterns, product demand, and customer purchasing behavior in supermarkets. Modern supermarkets generate a large amount of transactional data through daily billing activities. Each transaction contains valuable information such as product names, quantities purchased, price per item, and total bill amount. When this data is properly analyzed, it can reveal meaningful insights that help businesses improve operations and decision-making. In traditional supermarket environments, billing systems mainly record transactions for accounting purposes, but the stored data is rarely analyzed in depth.



However, with the help of data science tools and techniques, this raw billing data can be transformed into useful information. By analyzing supermarket billing records, it becomes possible to identify sales trends, popular products, seasonal demand patterns, and customer purchasing preferences. This project uses Python programming language and the Pandas library to analyze supermarket billing datasets. Pandas provides powerful tools for data cleaning, processing, and statistical analysis. Using these tools, the project examines different aspects of supermarket sales such as product-wise sales, category-wise revenue, total income generated, and transaction patterns over time. The analysis is further supported by graphical representations such as charts and graphs that make it easier to interpret the results. Visualization helps store managers and analysts quickly identify trends and patterns within the data. Through this data-driven approach, supermarkets can improve inventory planning, optimize product placement, and increase overall business efficiency. The insights obtained from billing data can help managers make better decisions regarding stock management, promotional strategies, and customer service improvements.

## II LITERATURE SURVEY

1. Supermarket Sales Analysis Using Data Analytics This study explains how data analytics techniques can be used to analyze supermarket sales data and identify important business insights. The research focuses on analyzing billing records to understand product demand, customer purchasing patterns, and sales performance. By applying statistical analysis and data visualization techniques, businesses can identify top-selling products and improve inventory management strategies.

2. Retail Sales Prediction Using Machine Learning This research explores the use of machine learning algorithms to predict retail sales trends based on historical sales data. Factors such as product category, price, promotional offers, and seasonal demand are analyzed to forecast future sales performance. The study highlights how predictive analytics can help retailers improve stock management, reduce product shortages, and increase overall profitability.

3. Customer Purchase Behavior Analysis in Supermarkets This paper investigates customer buying behavior using transaction data collected from supermarket billing systems. The study analyzes factors such as product combinations, purchase frequency, and customer preferences. Data analysis techniques help retailers understand which products are frequently purchased together and which categories generate the highest revenue, enabling better marketing and product placement strategies.

## III SYSTEM ANALYSIS

---



System analysis for a supermarket billing data insights system focuses on understanding how billing transactions are generated, processed, and utilized to extract meaningful business insights. The existing billing process captures essential data such as product details, quantity, price, time of purchase, and customer information (if available). This data is often stored in databases but not fully leveraged for decision-making. The analysis identifies key requirements such as efficient data collection, storage, and processing mechanisms to handle large volumes of transactions in real time. It also evaluates challenges like data redundancy, lack of integration, and limited reporting capabilities. The proposed system aims to transform raw billing data into actionable insights by applying data analysis techniques, enabling features like sales trends identification, customer behavior analysis, inventory tracking, and revenue forecasting. Overall, system analysis ensures that the designed solution is scalable, accurate, and capable of supporting business intelligence needs for better operational and strategic decisions

### **Existing System**

The existing system in most supermarkets primarily focuses on basic billing and transaction processing rather than data analysis. It records daily sales through billing software or manual entry, capturing details such as product name, quantity, price, and total bill amount. However, this data is mainly used for generating bills and maintaining simple records, with limited capability for deeper analysis. Reports are often static and generated periodically, making it difficult to track real-time sales trends or customer behavior. Additionally, the system may suffer from issues like data redundancy, lack of integration between billing and inventory systems, and minimal automation. As a result, decision-making relies heavily on manual interpretation, which can be time-consuming and less accurate, limiting the supermarket's ability to optimize sales, manage stock efficiently, and improve overall performance.

### **Disadvantages of Existing System**

- Limited data analysis – only basic billing, no deep insights
- No real-time reporting – delays in understanding sales trends
- Manual data handling – increases chances of human errors
- Data redundancy – same data stored multiple times
- Poor integration – billing, inventory, and customer data not connected

### **Proposed System**

---



The proposed system is designed to enhance the traditional billing process by integrating advanced data analytics and automation capabilities. It captures billing data in real time and stores it in a centralized database, ensuring accuracy and consistency. The system processes this data using analytical techniques to generate meaningful insights such as sales trends, customer purchasing behavior, product performance, and peak shopping hours. It also integrates with inventory management to automatically update stock levels and provide alerts for low or excess inventory. Interactive dashboards and dynamic reports enable quick and informed decision-making. Additionally, the system supports scalability, improved data security, and reduced manual effort through automation. Overall, the proposed system transforms raw billing data into valuable business intelligence, helping supermarkets optimize operations, increase efficiency, and improve customer satisfaction.

### Advantages of Proposed System

- Real-time data processing and instant report generation
- Improved accuracy with reduced human errors
- Centralized data storage for better management
- Advanced analytics for sales trends and customer behavior
- Better inventory management with automatic stock updates

## IV METHODOLOGY

The methodology of the Supermarket Billing Data Insights project explains the steps followed to collect, process, analyze, and visualize supermarket billing data. The main goal of this methodology is to transform raw billing data into meaningful insights that can support better business decisions. The first step in the methodology is data collection. The supermarket billing dataset is collected and stored in a structured format such as a CSV file. The dataset contains important information such as product name, product category, quantity sold, price per unit, and total bill amount. The second step is data preprocessing. In this stage, the dataset is loaded into Python using the Pandas library. The data is cleaned by checking for missing values, removing duplicate records, and organizing the dataset for accurate analysis. The third step is data analysis. Using Pandas functions, the dataset is analyzed to calculate key statistics such as total sales revenue, product-wise sales, and category-wise revenue. This step helps identify top selling products and understand sales performance across different product categories. The fourth step is data visualization. Visualization libraries such as

Matplotlib are used to create graphical representations of the analyzed data. Charts and graphs are generated to display product demand, category sales performance, and overall revenue trends. The final step is result interpretation. The insights obtained from the analysis are interpreted to understand sales patterns and customer purchasing behavior. These insights help supermarket managers make informed decisions regarding inventory management and sales strategies. Through this methodology, the project effectively converts raw supermarket billing data into useful business insights using data analysis technique.

### System Architecture

The System Architecture of the Supermarket Billing Data Insights project describes the overall structure and workflow of the system used to analyze supermarket billing data. The architecture explains how different components interact with each other to process raw data and generate meaningful insights. A well-designed system architecture ensures that the data flows through different modules efficiently and that each component performs its specific task in the analysis process. The system begins with the data source, which contains the supermarket billing dataset. This dataset includes important information such as product name, product category, quantity purchased, price per unit, and total bill amount. The dataset is usually stored in a CSV (Comma-Separated Values) file, which is a widely used format for storing structured data. The dataset acts as the primary input for the analysis system and provides the necessary information required for performing sales analysis. The next component is the data loading module, where the dataset is imported into the Python environment. The Pandas library is used to read the CSV file and convert it into a Data Frame structure. This structure allows the system to easily access, manipulate, and analyze the data using various data analysis functions provided by Pandas. After the dataset is loaded, the system performs data preprocessing. Data preprocessing is an important step that prepares the dataset for accurate analysis. In this stage, the system checks for missing values, removes duplicate records, and organizes the dataset into a structured format. Data cleaning helps ensure that the dataset is reliable and free from inconsistencies that could affect the results of the analysis.



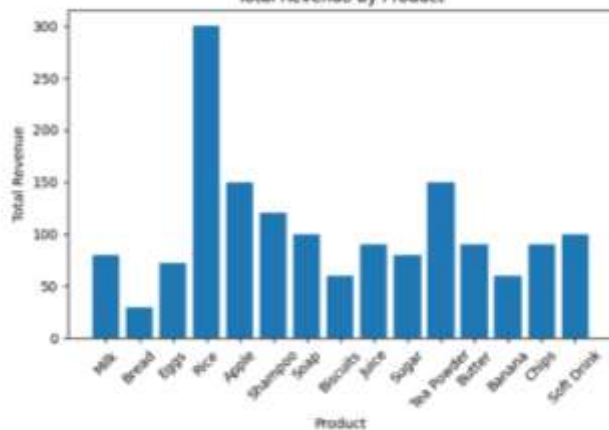
## V RESULTS&OUTPUT

Sample Supermarket Billing Dataset

Product	Category	Quantity	Price	Total
Milk	Dairy	2	40	80
Bread	Bakery	1	30	30
Eggs	Dairy	12	6	72
Rice	Grocery	5	60	300
Apple	Fruits	3	50	150
Shampoo	Personal Care	1	120	120
Soap	Personal Care	4	25	100
Biscuits	Snacks	3	20	60
Juice	Beverages	2	45	90
Sugar	Grocery	2	40	80
Tea Powder	Grocery	1	150	150
Butter	Dairy	1	90	90
Banana	Fruits	6	10	60
Chips	Snacks	3	30	90
Soft Drink	Beverages	2	50	100



Total Revenue by Product





## VI CONCLUSION

The Supermarket Billing Data Insights project demonstrates how data analysis techniques can be applied to retail sales data to extract meaningful information. Supermarkets generate large amounts of billing data through daily transactions, and analyzing this data helps identify important sales patterns and customer purchasing behavior. In this project, supermarket billing data was analyzed using the Python programming language and the Pandas library, which are widely used tools in the field of data science and analytics. The system successfully loaded and processed the dataset, calculated the total revenue generated from sales, identified frequently purchased products, and analyzed category-wise sales performance. These operations help transform raw transaction data into useful business insights. Data visualization techniques were also implemented in the project using graphical tools such as charts and graphs. These visual representations make it easier for users to understand complex sales data and quickly identify patterns and trends. For example, category-wise sales charts help managers see which product categories generate the highest revenue, while product demand charts highlight the most frequently purchased items. The analysis results provide valuable information for supermarket managers and business analysts. By understanding which products and categories contribute most to sales, store managers can improve inventory management, product placement, and stock availability. This information can also help supermarkets reduce product shortages, manage supply more efficiently, and improve customer satisfaction. Furthermore, the insights obtained from this project support data-driven decision-making in retail businesses. Instead of relying on assumptions or manual analysis, store managers can use actual sales data to guide their decisions. This helps improve operational efficiency and allows businesses to respond more effectively to changes in customer demand.

## REFERENCE

- [1] Kumar, R. D., Prudhviraaj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
- [2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.



- [3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
- [4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
- [5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
- [6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0\_79.
- [7] R. D. Kumar, V. N. S. Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
- [8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
- [9] Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
- [10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
- [11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.