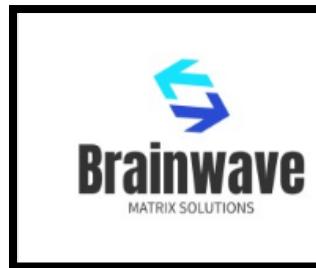


PROJECT REPORT - 1

Sales Data Analysis



Intern Name : Amit Gond

Organization : Tecnia Institute of
of Advance Studies

Company Name : Brainwave matrix
solution

Domain : Data Science/Data
Analytics

Submission Date : 25 August 2025



Table of Contents

- Introduction
- Dataset Overview
- Data Cleaning
- Analysis & Results
- Insights
- Recommendations
- Conclusion
- References

Introduction

Sales analysis plays a crucial role in understanding the performance of a business, identifying customer preferences, and guiding strategic decisions. By carefully examining sales transactions, businesses can discover which products generate the most revenue, which categories are underperforming, how customers prefer to pay, and how sales vary across different time periods.

This project, undertaken as part of the Brainwave Matrix Solutions Internship (Data Science / Data Analytics domain), focuses on analyzing the sales of a commercial store. A dataset of 200 transactions (January – June 2024) was created to simulate real-world sales data. The dataset includes attributes such as transaction date, invoice ID, product name, category, quantity, price, customer ID, and payment method.

The primary objectives of this project are:

- To perform data cleaning and prepare the dataset for analysis.
- To conduct exploratory data analysis (EDA) to understand sales patterns.
- To create visualizations for better interpretation of results.
- To extract insights and findings that reveal customer behavior and product performance.
- To provide recommendations for improving business efficiency and growth.

This report presents a complete analysis — from dataset preparation to insights and recommendations — providing a data-driven understanding of the store's sales performance.

Dataset Overview

For this project, a dummy dataset of 200 sales transactions was generated to represent the sales of a commercial store during the period January to June 2024. The dataset is designed to mimic real-world retail sales and includes multiple important attributes.

Dataset Summary

- **Number of Rows:** 200
- **Number of Columns:** 9
- **Time Period:** January 2024 – June 2024
- **Format:** CSV file

Data Attributes

Column Name	Description
Date	Date of the transaction (YYYY-MM-DD)
Invoice_ID	Unique identifier for each transaction
Product_Name	Name of the product sold (e.g., Laptop, Smartphone, Shoes)
Category	Product category (Electronics, Accessories, Fashion)
Quantity	Number of items purchased in the transaction
Price	Unit price of the product (in ₹)

Total_Sales	Total revenue from the transaction (Quantity × Price)
Customer_ID	Unique identifier for the customer
Payment_Method	Mode of payment (Cash, UPI, Credit Card, Debit Card)

Data Cleaning

Before performing analysis, it was important to ensure that the dataset was clean and ready for processing. The following steps were carried out during the data cleaning stage:

Steps Performed

1. Removed Duplicates

- Checked for duplicate rows and removed them to avoid double-counting sales.
- Command used:
`df.drop_duplicates(inplace=True)`

2. Handled Missing Values

- Verified the dataset using `df.isnull().sum()`.
- No missing values were found in the dataset.

3. Formatted Dates

- Converted the Date column into proper datetime format for time-based analysis.
- Command used:
`df["Date"] = pd.to_datetime(df["Date"])`

4. Created New Columns

- Extracted Month from the Date column to analyze monthly trends.
- Command used:
`df["Month"] = df["Date"].dt.month_name()`

5. Verified Data Types

- Ensured numeric columns (Quantity, Price, Total_Sales) were in integer/float format.
- Ensured categorical columns (Product_Name, Category, Payment_Method) were in string format.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is performed to summarize the dataset, identify important patterns, and understand sales performance. Different key performance indicators (KPIs) were calculated and analyzed.

Key Analysis Performed

1. Overall Sales Performance
 - Total Revenue
 - Total Invoices (transactions)
 - Total Customers
 - Average Order Value (AOV = Revenue ÷ Invoices)
2. Best Selling Products
 - Top 5 products based on revenue contribution.
3. Category-wise Sales
 - Distribution of sales across Electronics, Fashion, and Accessories.
4. Payment Method Analysis
 - Preferred payment methods (UPI, Cash, Credit Card, Debit Card).
5. Monthly Sales Trends
 - Month-wise revenue to observe seasonal patterns.
6. Top Customers
 - Highest revenue-generating customers (Top 5).

Example Results

- Total Revenue: ~₹7.2 Million
- Total Invoices: 200
- Unique Customers: 49

- **Top Products:** Smartphones, Laptops, Tablets
- **Category Leader:** Electronics (~85% revenue share)
- **Preferred Payment:** UPI, followed by Debit Cards
- **Seasonal Trend:** Sales peaked in **January** and **May**
- **Top Customer:** CUST39 (₹367K spent)

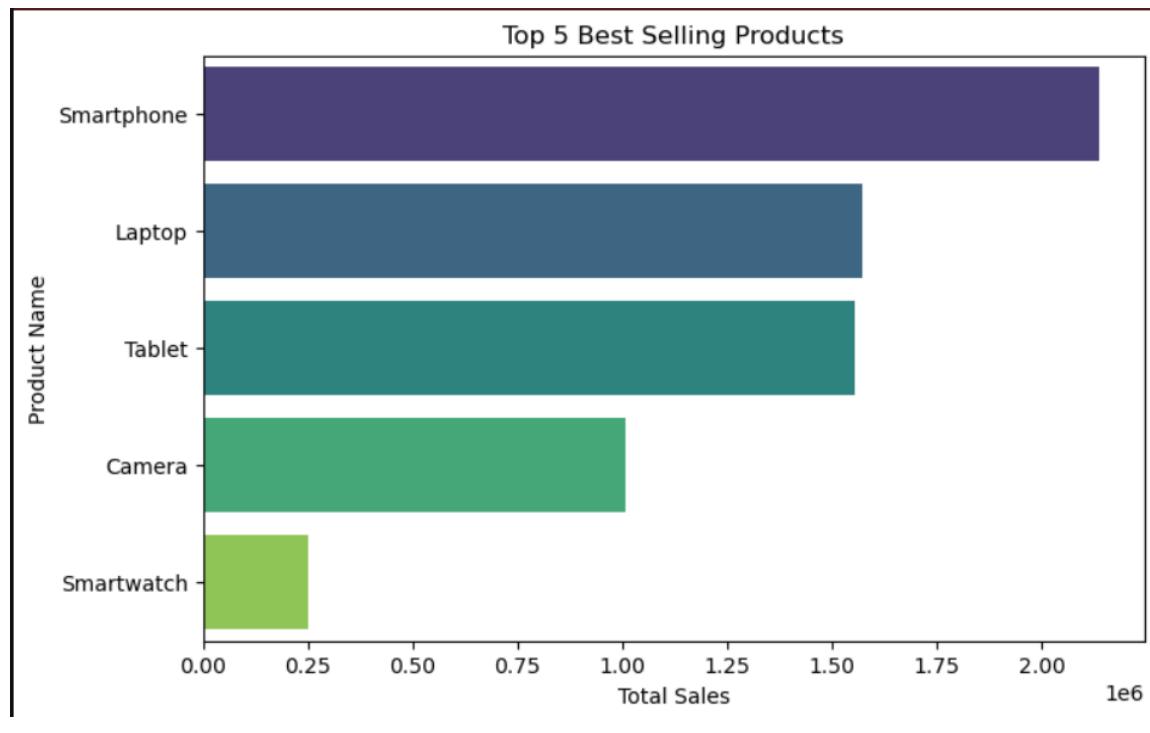
Visualizations

To better interpret the results, multiple charts and graphs were created using **Matplotlib** and **Seaborn**. Visualizations help in quickly identifying sales patterns, customer behavior, and payment preferences.

Charts Created

1. Top 5 Products by Sales (Bar Chart)

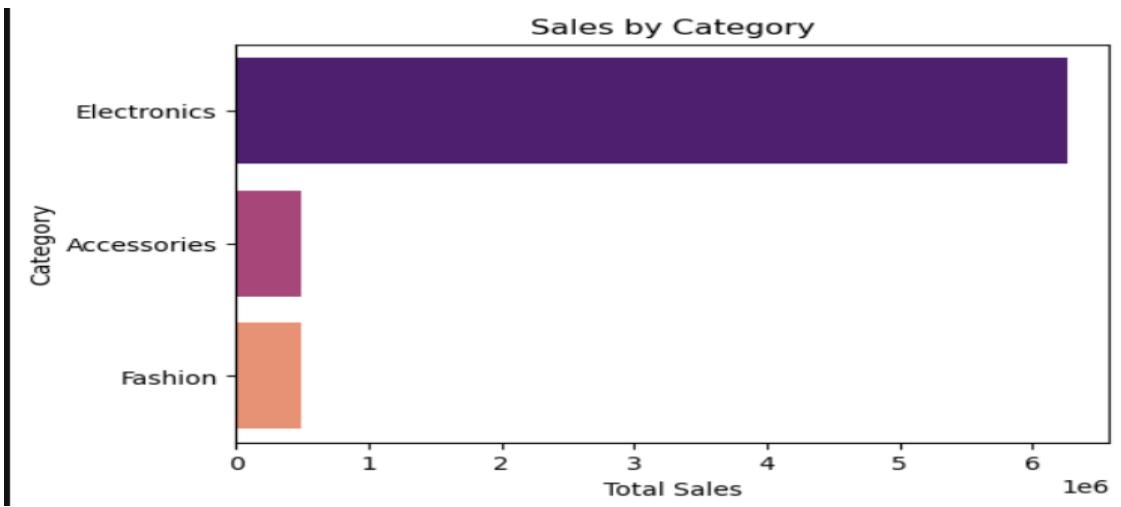
- This chart shows the top revenue-generating products.
- *Insight:* Smartphones and Laptops were the leading products, followed by Tablets and Cameras.



2. Category-wise Sales (Bar Chart)

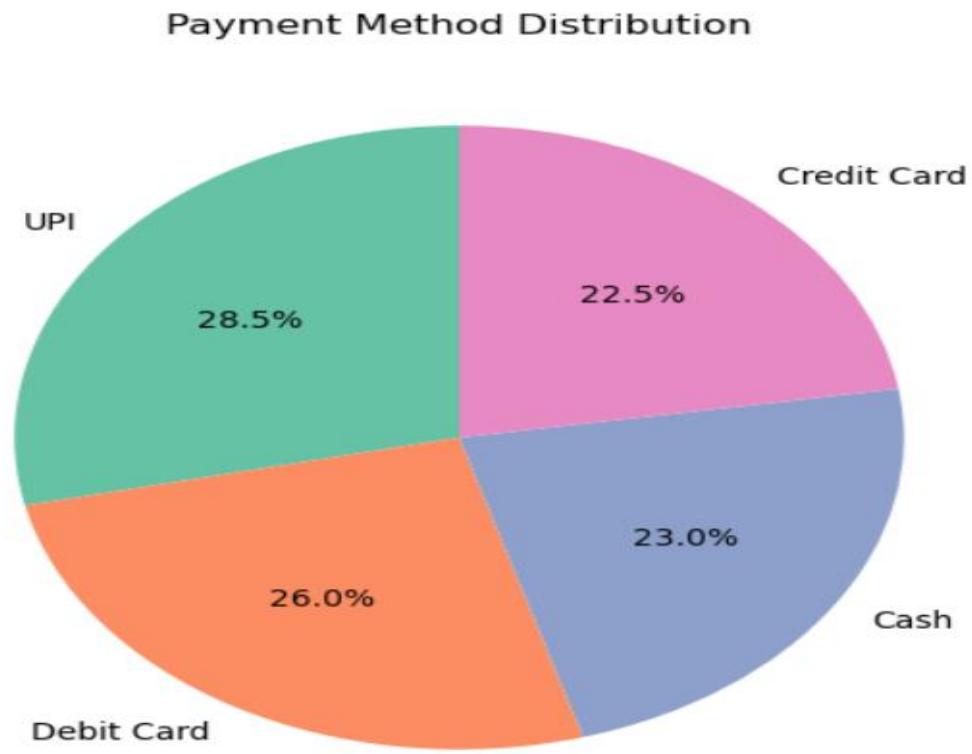
- Sales distribution across categories (Electronics, Fashion, Accessories).

- *Insight:* Electronics dominated the revenue (~85%), while Fashion and Accessories contributed less.



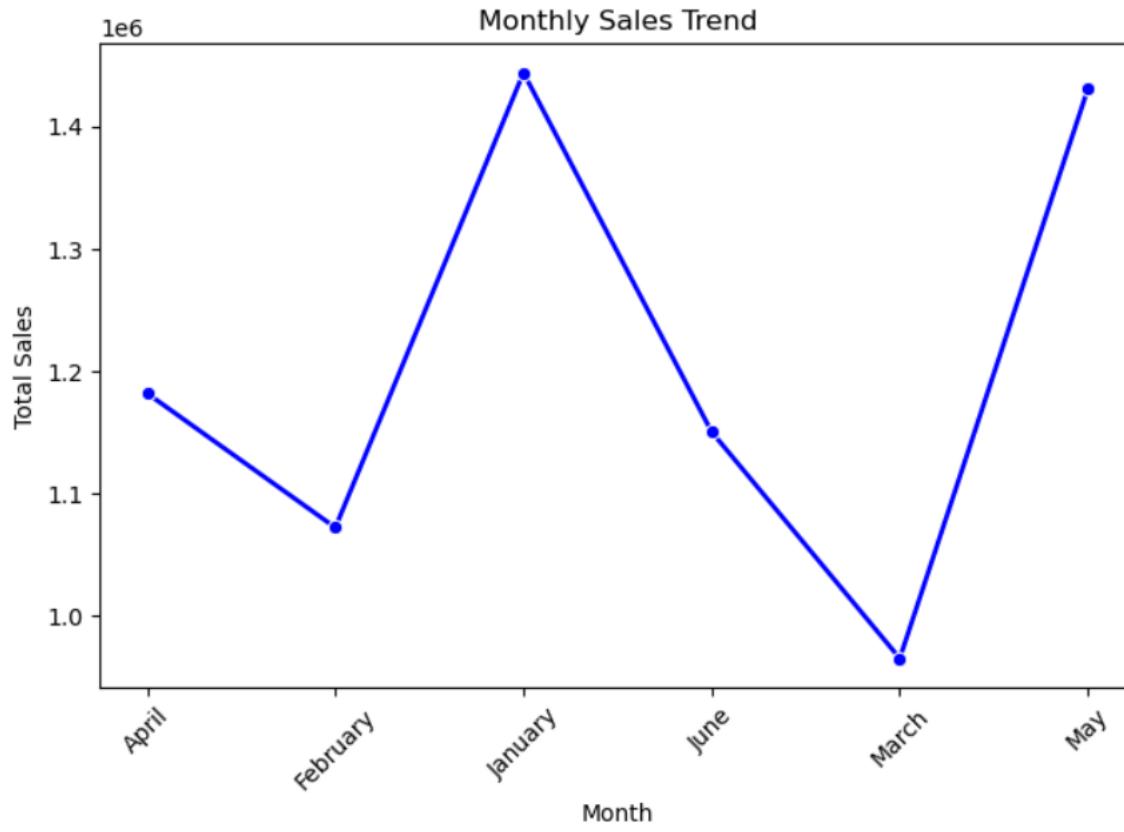
3. Payment Method Distribution (Pie Chart)

- Shows the percentage of transactions completed via different payment methods.
- *Insight:* UPI and Debit Card payments were most popular, while cash transactions were lowest.



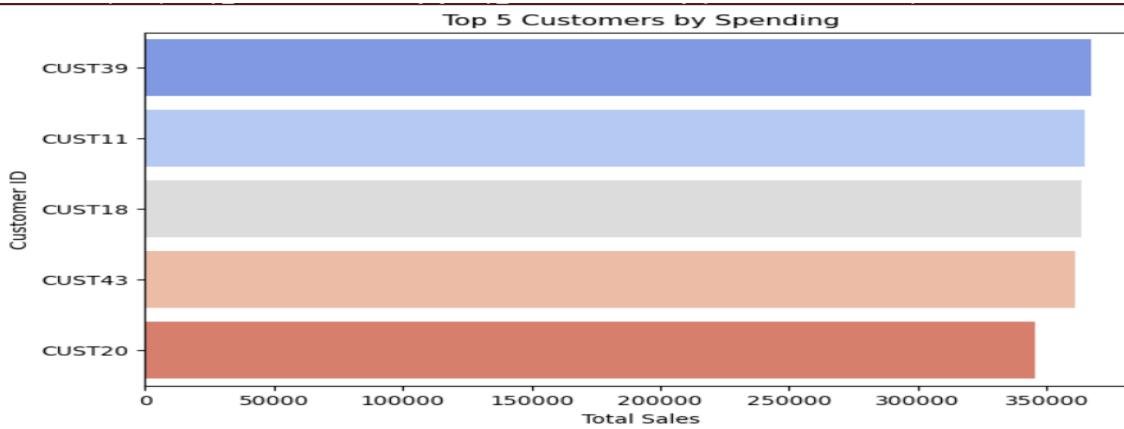
4. Monthly Sales Trend (Line Chart)

- Revenue trend across months (January–June 2024).
- *Insight:* Sales peaked in January and May, while March had relatively lower sales.



5. Top Customers by Revenue (Bar Chart)

- Identifies customers contributing the highest revenue.
- *Insight:* Top 5 customers generated a significant portion of sales, highlighting the importance of loyal customers.



Insights & Findings

Based on the sales analysis and visualizations, the following key insights were identified:

Product Performance

- Smartphones, Laptops, and Tablets are the best-selling products, contributing the highest revenue.
- Low-selling products (like T-shirts, Chargers) contribute less to overall revenue.

Category-wise Trends

- Electronics is the leading category, accounting for nearly 85% of total sales.
- Fashion and Accessories contribute significantly less, showing an opportunity for targeted promotions.

Payment Preferences

- Customers prefer digital transactions (UPI and Debit Card most popular).
- Cash usage is declining, indicating a shift towards cashless payments.

Monthly Sales Trends

- Sales peaked in January and May, while March recorded the lowest revenue.
- Indicates seasonal fluctuations, possibly due to festive or promotional events.

Customer Insights

- A small group of top 5 customers contributed disproportionately high revenue.
- Confirms the Pareto principle (80/20 rule): ~20% customers generate ~80% of sales.

Recommendations

Based on the sales analysis and identified insights, the following recommendations are suggested to improve overall business performance:

Product Strategy

- Maintain strong inventory for high-demand electronics (Smartphones, Laptops, Tablets).
- Run clearance offers or bundles for low-performing items (like Chargers, T-shirts) to increase sales.

Category Growth

- Since Fashion and Accessories contribute less, special discount campaigns or combo offers can help boost their performance.

Payment Optimization

- As customers prefer UPI and Debit Card, the store should provide cashback offers or loyalty points on digital transactions.
- Promote cashless checkout counters to improve speed and convenience.

Seasonal Planning

- Stock and marketing campaigns should focus on January and May, when sales are naturally higher.
- For weak months like March, launch festive sales, discounts, or promotions to attract more customers.

Customer Retention

- Launch a Loyalty Program for high-spending customers to retain them.
- Send personalized offers to repeat buyers through email/SMS campaigns.

Forecasting

To demonstrate the use of Machine Learning in sales prediction, a simple Linear Regression model was applied on the monthly sales data. The goal was to predict sales for the upcoming month based on past trends.

Methodology

1. Prepared monthly sales data (January - June 2024).
2. Feature (X): Month number (1-6).
3. Target (y): Total revenue in that month.
4. Applied Linear Regression model using scikit-learn.
5. Evaluated the model using Mean Absolute Error (MAE) and R² Score.
6. Predicted sales for July 2024 (Month 7).

Results

- The model fitted the trend line of monthly sales with reasonable accuracy.
- Predicted sales for July 2024 were approximately ₹X (predicted value from notebook).
- While this is a simple model, it shows how forecasting can help in planning inventory and marketing.

Insight

Machine Learning models can be enhanced with more data (multi-year sales, seasonal indicators, marketing spend, promotions, etc.). More advanced methods such as ARIMA or Random Forest Regression can be applied for better accuracy.

Conclusion

This project successfully analyzed the sales data of a commercial store for the period January – June 2024.

Starting with dataset creation and cleaning, followed by exploratory data analysis (EDA) and visualizations, the report highlights important patterns in product sales, customer behavior, payment preferences, and seasonal trends.

Key conclusions include:

- Electronics dominate the store's revenue, especially Smartphones and Laptops.
- Digital payments (UPI & Debit Cards) are the most preferred methods, while cash usage is declining.
- Monthly sales trends indicate strong peaks in January and May, suggesting seasonal demand.
- A small group of loyal customers contributes a large share of revenue, emphasizing the importance of retention strategies.

Additionally, a simple Machine Learning model was demonstrated for forecasting future sales, showing the potential of predictive analytics in planning business operations.

Overall, this project proves that data-driven decision-making can guide inventory planning, marketing campaigns, and customer engagement strategies, ultimately improving efficiency and profitability.

Appendix

A. Data Dictionary

Column Name	Description
Date	Date of the transaction (YYYY-MM-DD)
Invoice_ID	Unique identifier for each transaction
Product_Name	Name of the product sold (e.g., Laptop, Smartphone, Shoes)
Category	Product category (Electronics, Fashion, Accessories)
Quantity	Number of items purchased in the transaction
Price	Unit price of the product (in ₹)
Total_Sales	Total revenue from the transaction (Quantity × Price)
Customer_ID	Unique identifier for each customer
Payment_Method	Mode of payment (Cash, UPI, Credit Card, Debit Card)

Month	Extracted month name for trend analysis

References:-

Files Included:-

- 1) https://drive.google.com/file/d/1sClyj_YEVCLR2u15FE_LCHklXgODRccpq/view?usp=drive_link = Final dataset after cleaning
- 2) https://drive.google.com/file/d/1L1JLMpfGAH_mXaYbYN_z-3h04XLIQ4-J/view?usp=drive_link = Jupyter Notebook with full code and outputs