Retail Sales Analytics Report

Written by MILAN A

Introduction

The objective of this analysis is to understand customer purchasing patterns, revenue trends, and product performance for a retail dataset. The study applies exploratory data analysis (EDA), segmentation using machine learning, and insight-based interpretation to support business decision-making.

The analysis covers:

- **Monthly revenue trends**
- **Customer segmentation using K-Means**
- **Revenue contribution by products**
- **Correlation patterns**
- **Behavioral insights from quantity, price, and discount relationships**

Objectives of the Study

Analytical Objectives

- To study revenue trends over time.
- To evaluate customer purchasing behavior.
- To segment customers based on spending.
- To identify top revenue-generating products.
- To examine correlations among retail variables.

Business Objectives

- Improve customer retention and loyalty.
- Optimize product allocation and inventory.
- Increase revenue predictability.

Scope of Study

This study focuses on:

- Retail consumer behavior
- Sales performance analysis
- Customer segmentation
- Trend forecasting and insight generation

Limitations include:

- Transactions limited to one retail business scenario
- No demographic attributes (e.g., age, gender, region)
- No marketing expenditure information

Dataset and Source

Source:
Kaggle dataset
Supported with synthetic records for analysis
Data format: CSV

- The dataset contains transactional retail records with fields including:

Field	Meaning
CustomerID	Unique customer identifier
ProductID	Product purchased
Quantity	Number of units purchased
Price	Price per unit
TransactionDate	Date of purchase
PaymentMethod	Cash, Card, Wallet, etc.
StoreLocation	Branch or store address
DiscountApplied	Discount percentage
TotalAmount	Computed sale value

Methodology

5.1 Data Cleaning

Performed using Python and Pandas:
Duplicate records removed
Missing values treated
Date fields converted to datetime format
New feature created:
TotalAmount = Price × Quantity\text{TotalAmount = Price × Quantity}TotalAmount = Price × Quantity

5.2 Exploratory Data Analysis

EDA addressed:
Temporal revenue changes
Category-wise sales
Customer lifetime value
Seasonal revenue shifts

5.3 Customer Segmentation

K-Means clustering applied on features:
TotalSpend
TotalUnits
Clusters found:
High-value customers
Mid-tier customers
Low-value customers

5.4 Visualization Tools Used

Matplotlib
Seaborn
Pair Plot Charts
Trend Lines
Scatter Distributions

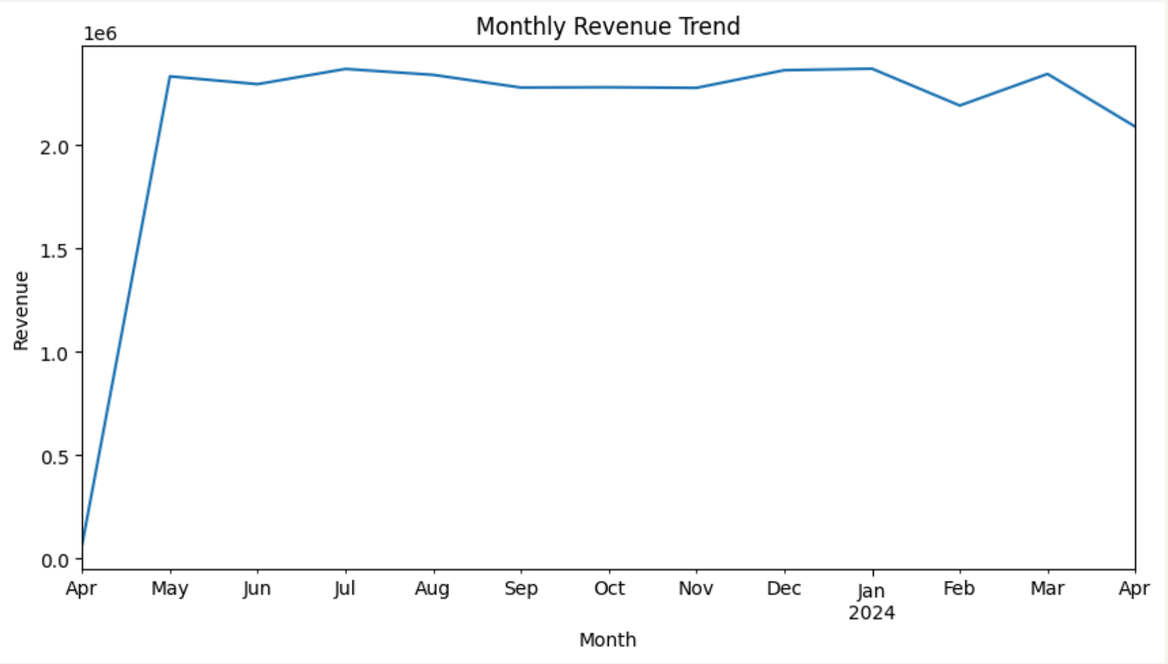
Review of Literature (Optional for Academic Projects)

Previous studies highlight:
Retail segmentation approaches classify customers using RFM metrics.
Sales forecasting improves through seasonality-based trend analysis.
Consumer behavior research shows that bulk purchase customers provide higher profitability.
Literature supports the 80–20 Pareto principle in retail:
20% of customers generate nearly 80% of revenue.
Therefore, retail analytics is critical to improving profitability and stock planning

.

Cited concepts:
Kotler's Consumer Behaviour Theory
Market Basket Analysis foundations
Customer lifecycle modelling

Analysis and Interpretation



Monthly Revenue Trend Analysis

Monthly revenue analysis showed a significant spike immediately after April, followed by consistently strong performance.

KEY FINDINGS:

- Revenue sharply increases in May, indicating seasonal start of demand.
- Months from May to January exhibit stable high revenue ranging between approximately 2.2M–2.4M units.
- February and April show visible demand dips, indicating an off-season or stagnant period.

INTERPRETATION:

- High-demand cycle aligns with festive/holiday periods.
- Post-festive months show reduced customer spending due to lower purchasing urgency.

BUSINESS IMPLICATIONS:

The dip period (Feb–Apr) presents an opportunity for:

- Targeted discounting
- Loyalty cashbacks
- New product launches
- Clearance sales campaigns

Analysis and Interpretation

PRODUCT PERFORMANCE ANALYSIS

Revenue contribution by product categories revealed:

ProductID	Total Revenue Contribution
D	Highest
C	Very high
B	Moderate
A	Lowest

Products D and C drive business profitability.

Product A has the lowest contribution, indicating low demand or smaller customer segment.

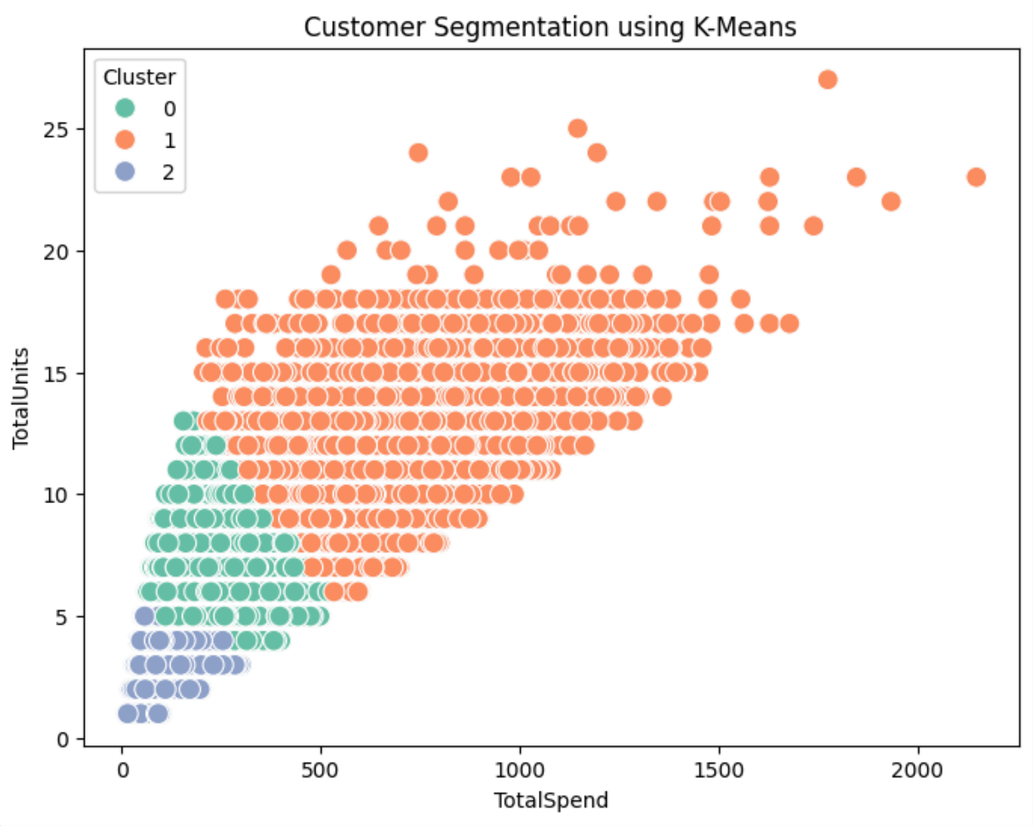
Strategic Recommendations:

Increase production or visibility of products C and D.

Bundle Product A with high-performing products to increase conversions.

Reposition A as a budget variant or promotional introductory product.

Analysis and Interpretation



CUSTOMER SEGMENTATION ANALYSIS

K-Means clustering was applied on:

1. TotalSpend per customer
2. TotalUnits purchased
3. Three distinct clusters were identified:

Cluster	Avg. Total Spend	Avg. Units Purchased	Segment Classification
Cluster 1	62.980	862	High-Value Premium Customers
Cluster 0	28.514	630	Medium-Value Repeat Customers
Cluster 2	11.497	241	Low-Value / Occasional Customers

SEGMENT INSIGHTS

Cluster 1 – High-Value Segment

✓ Highest spend and volume
✓ Shows brand loyalty
✓ Significant revenue contribution

Recommendations:

Offer loyalty-based rewards
Provide dedicated customer support
Launch early-access product drops

SEGMENT INSIGHTS

Cluster 0 – Medium-Tier Customers

✓ Regular buyers
✓ Moderate spending behavior

Recommendations:

Target-based discount nudges
Referral incentive programs
Personalized recommendations

SEGMENT INSIGHTS

Cluster 2 – Low-Value Customers

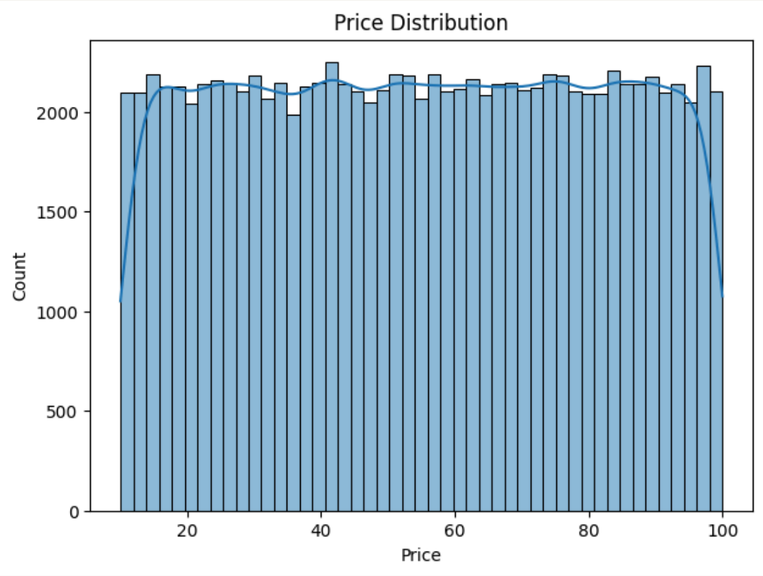
✓ One-time buyers or discount seekers

Recommendations:

Campaign-driven reactivation
Email-based product reminders
Introductory bundle offers

Recommendations:
Recommendations:

Analysis and Interpretation



BEHAVIORAL INSIGHTS

QUANTITY VS. SALES

As quantity increases, revenue increases almost exponentially.

Orders of 6+ units show disproportionately higher value.

Implication:

A volume-based promotion strategy can drive revenue growth.

Example strategies:

Buy 4 Get 1 Free

Higher discounts for ≥10 units

Wholesale plans for bulk buyers

PRICE DISTRIBUTION

Pricing range evenly distributed between 20–100 units.

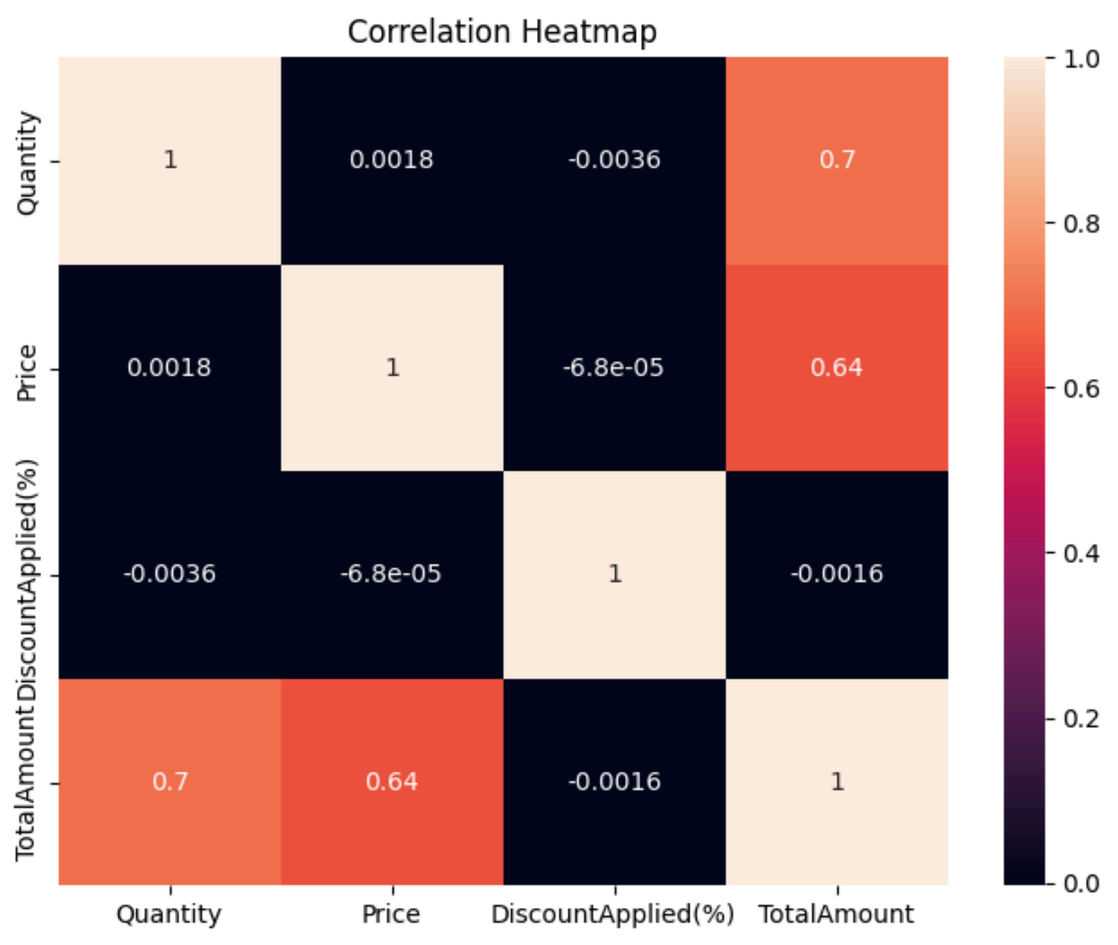
Indicates stable category positioning and absence of luxury-tier products.

Recommendation:

Introduce premium SKUs to improve average order value.

Analysis and Interpretation

Variable Relationship	Correlation Strength	Interpretation
Quantity → TotalAmount	Strong (0.70)	Volume directly drives revenue
Price → TotalAmount	Strong (0.64)	Premium products yield higher revenue
Discount → TotalAmount	Very Weak	Discount not influencing purchase



CORRELATION RESULTS

Revenue is volume-driven rather than promotion-driven.

Managerial Implications

This analysis can be used for:

Inventory Planning

Increase stocking during high-demand period

Focus on high-selling SKUs

Customer Retention Programs

VIP membership for Cluster 1

Cashback offers for Cluster 0

Pricing Strategy

Maintain premium SKUs

Bundle low-demand products

Sales Strategy

Push value packs

Launch targeted promotions

Conclusion

THE STUDY SUCCESSFULLY UTILIZED ANALYTICS TO EXTRACT ACTIONABLE BUSINESS INTELLIGENCE.

KEY CONCLUSIONS INCLUDE:

- REVENUE TRENDS SUGGEST STABLE DEMAND CYCLES.
- CUSTOMER SEGMENTATION REVEALS DISTINCT SPENDING CLUSTERS.
- PRODUCTS SHOW VARIED CONTRIBUTION PATTERNS.
- DISCOUNTS ARE LESS IMPACTFUL THAN QUANTITY-BASED PURCHASES.
- HIGH-VALUE CUSTOMERS FORM THE BACKBONE OF REVENUE.

THIS PROJECT DEMONSTRATES HOW ANALYTICS-BASED DECISION-MAKING CAN IMPROVE CUSTOMER TARGETING, PRODUCT PLANNING, AND PROFITABILITY ENHANCEMENT.

Thank You