



# E-commerce Sales Analysis with MySQL & Power BI

Happy Lethabo Seoketsa | 20/08/2025

# Introduction

This report analyzes a synthetic e-commerce transactions dataset from Kaggle. MySQL was used for data cleaning and querying, while Power BI applied for visualization.

The goal of this project is to uncover insights about sales patterns, customer behavior, product performance, the impact of discounts on sales, and preferred payment methods.

## ABOUT DATASET

**Source:** Kaggle e-commerce-dataset

**Rows:** 3660

**Columns:** 8

### Column Descriptions:

- **User\_ID:** A unique identifier for each user (e.g., a shortened version of a UUID)
- **Product\_ID:** A unique identifier for each product (e.g., a shortened version of a UUID)
- **Category:** The product category (e.g., Electronics, Clothing, Sports, etc.)
- **Price:** The original price of the product before any discount is applied
- **Discount (%):** The discount percentage applied to the product
- **Final\_Price:** The final price of the product after applying the discount
- **Payment\_Method:** The method used for payment (e.g., Credit Card, UPI, Net Banking)
- **Purchase\_Date:** The date when the transaction occurred, formatted as MM-DD-YYYY

## METHODOLOGY

### MySQL (Data Cleaning and Querying):

#### a. Data Cleaning:

- Explored the dataset using **SELECT \*** and **DESCRIBE** to understand the structure and data types.
- Converted the **Purchase\_Date** column from text to **DATE** format using **STR\_TO\_DATE()** and updated the column type.
- Checked for missing values in all columns to ensure data completeness.
- Renamed columns and standardized data types:
  - Price (Rs.) → price\_rupees (DECIMAL)
  - Discount (%) → discount\_percent (INT)
  - Final\_Price(Rs.) → final\_price\_rupees (DECIMAL)

#### b. Querying:

After cleaning the dataset in MySQL, several queries were performed to better understand the e-commerce data and answer key business questions:

##### 1. Total Transactions

- Counted the total number of transactions: **3,660 transactions**.

##### 2. Top Product Categories

- Identified the most frequently purchased categories along with their transaction counts and percentage of total transactions:

Home & Kitchen (549 transactions) 14.59%  
Books (534 transactions) 14.59%  
Clothing (531 transactions) 14.51%  
Toys (523 transactions) 14.29%  
Sports (520 transactions) 14.21%  
Beauty (505 transactions) 13.8%  
Electronics (498 transactions) 13.61%

**Finding:** Home & Kitchen was the most purchased category, but all categories were balanced with similar transaction counts.

### 3. Top Products

- Determined the best-selling products along with their category and percentage of total transactions

**Finding:** Each product was purchased **only once**, so there were no repeat purchases for any single product.

### 4. Top Customers

- Retrieved the customers with the highest number of transactions.

**Finding:** Each customer made **only one purchase**, indicating that no customer made multiple purchases in this dataset.

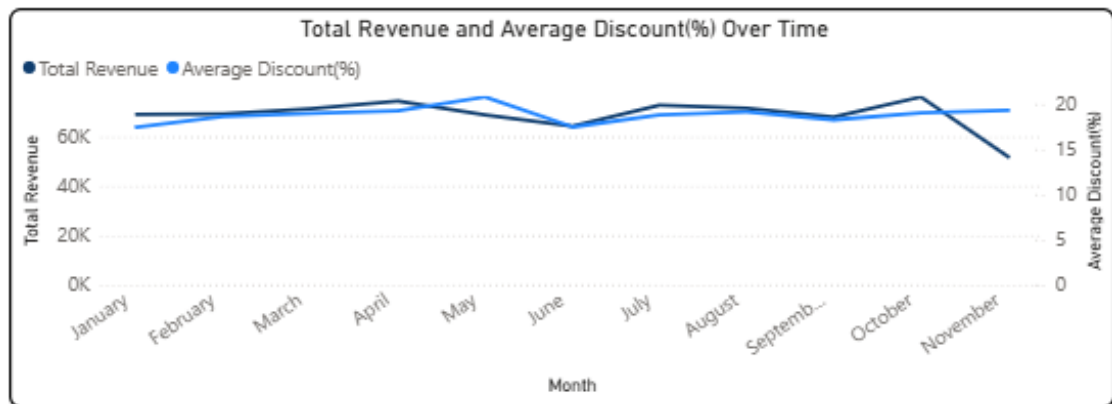
## PowerBI Visualization:

### • Key Visualizations:

- **Total Sales by Category:** Highlights which product categories contribute the most to total sales.
- **Total Revenue and Average Discount (%) Over Time:** Shows revenue trends throughout the year while comparing average discounts.
- **Total Sales by Payment Method:** Displays the distribution of sales across different payment options, helping understand customer payment preferences.
- **Total Revenue by Category:** Provides a view of revenue contribution per category, identifying which categories are not only selling but also profitable.
- **Average Discount (%) by Category:** Shows how discounting strategies vary across categories.

## Key Insights:

1. **Overall Sales Performance:** Sales stayed above 300 units for most of the year but fell to 245 units in November. Total revenue was above 64,000 Rupees in most months but dropped to 51,915 Rupees in November. The average discount stayed between 17% and 20%, so it doesn't explain the drop. Seasonal changes, supply problems, or marketing issues should be checked to prevent future declines.



2. **Clothing Category Performance:** Clothing has the highest total revenue even though it is only third in total sales. It also has the lowest average discount (17.92%), and its May average discount (%) was at its highest for the clothing category yet total revenue was almost at its lowest, this suggests that customers are willing to pay higher prices for clothing products without needing large discounts.

### E-commerce Sales Performance Analysis

Category

Beauty
Books

Clothing
Electronics

Home & Kitchen
Sports

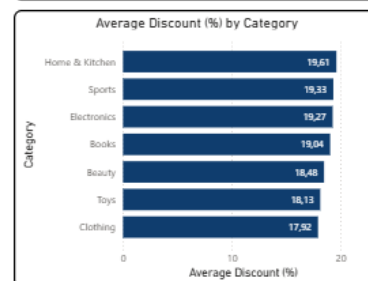
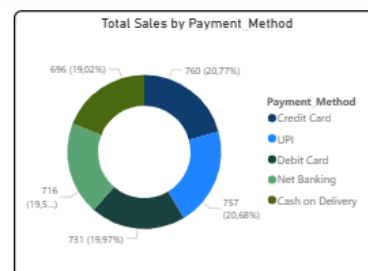
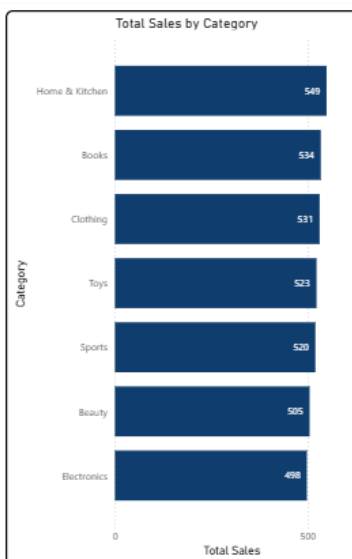
Toys

Payment Method

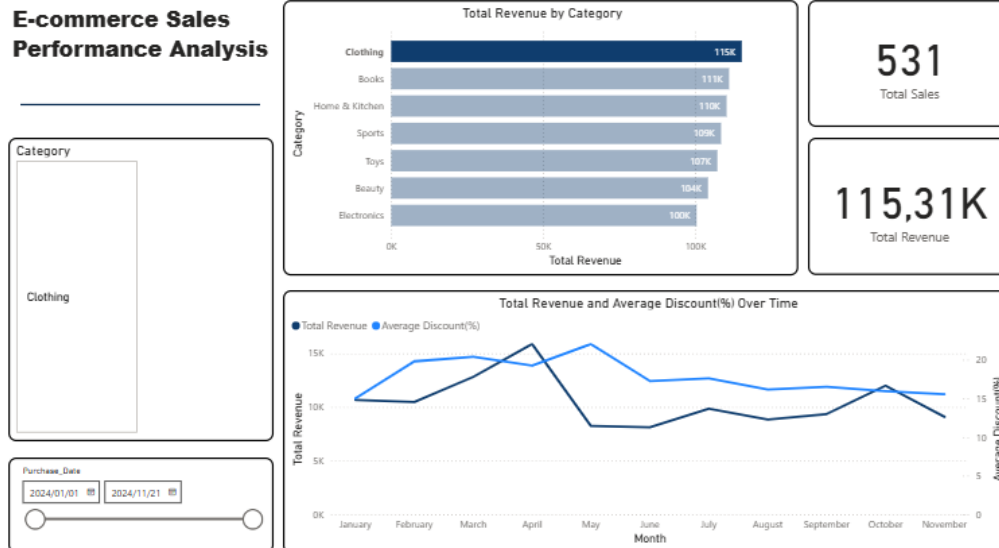
All

Purchase Date

All

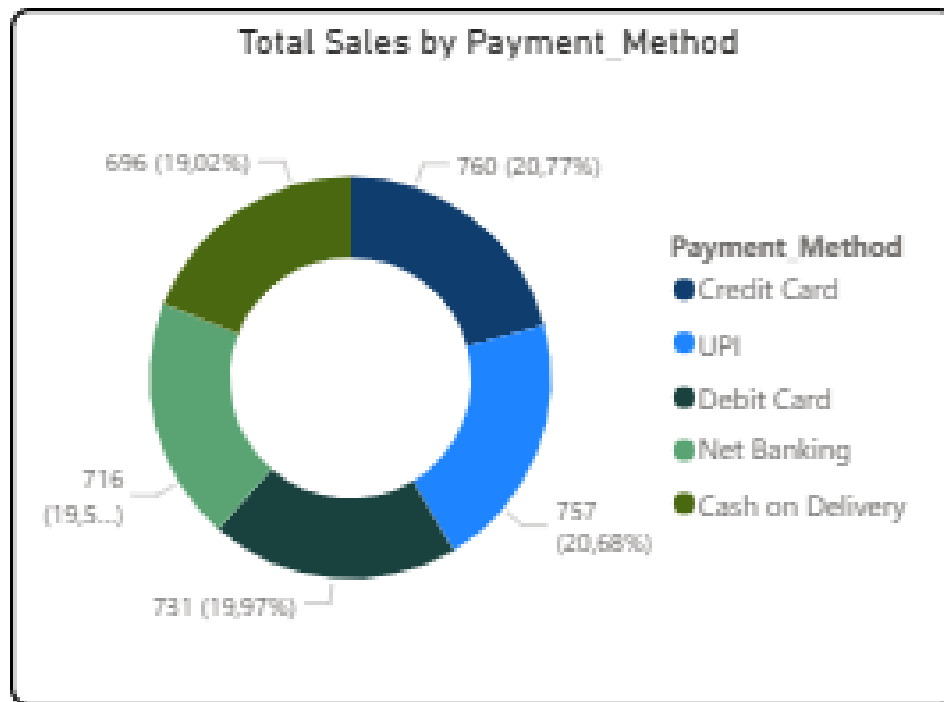


From May (22% discount) to November (15% discount), the shop steadily reduced discounts, yet clothing revenue stayed strong. This shows that heavy discounting is not the main driver of clothing sales.



Since discounts do not strongly impact clothing revenue, the shop could **reduce unnecessary discounts** to protect profit margins. Instead, focus on **product quality, and seasonal promotions** to sustain and grow revenue.

3. **Payment Method Overview:** Online payment accounts for about 80% of all sales made, this shows customers trust the platform, but cash on delivery is still remains important, which accounts for almost 20% of the total sales, even though it can be because some customers don't have access to digital payments, it can cause the business, because some customers can refuse to pay upon delivery, maybe because they not happy with the product, or for some other reason, but the business will still be accountable for delivery fees.



To reduce risks of Cash on delivery payments without removing it as a payment option, the store must try to encourage online payment, maybe by offering rewards like a discount.

## Conclusion

The analysis shows that overall sales and revenue stayed strong throughout the year, although there was a noticeable dip in November that cannot be explained by discount changes alone. Clothing stands out as the highest revenue-generating category despite having relatively lower discounts, suggesting customers are willing to pay for higher-priced products in this segment. This indicates an opportunity to optimize discount strategies further to maximize revenue.

Payment method analysis highlights that most customers prefer digital payments such as credit cards, UPI, and debit cards, while only less than 20% of sales were through Cash on Delivery. Since COD comes with higher risks of making a loss, and delayed cash flow as the funds will still need to be deposited to the business account, the business could encourage customers to shift toward online payment methods through incentives or small discounts.

Overall, the findings suggest that focusing on optimizing discount levels, especially in clothing, and reducing reliance on cash on delivery could improve profitability and operational efficiency for the shop going forward.