

Dataset Selection Document & Report

Business Area and Why We Chose It

- **Business Area:** E-commerce (Online Shopping)
- **Reason:** This type of data is great for showing off important analysis skills. We can easily see which customers are the best, how sales change over time, and what products people like most.

Data Source

- **File Name:** `ecommerce_dataset_updated.csv`
- **Data Source:** [Dataset](#)

What the Data Contains

The data has 3,660 unique purchase records, where each record is one product bought by a customer. The key information includes:

- Who bought the product (`User_ID`)
- What product was bought (`Product_ID` and `Category`)
- How much does it cost before and after the discount
- How the customer paid (`Payment_Method`)
- When the purchase happened (`Purchase_Date`)

Main Questions We Want to Answer

1. **Who are our best customers?** We want to split customers into different groups based on their spending and buying habits.
2. **Which products sell best?** We will see which categories bring in the most money and how much we have to discount them.
3. **When do people buy the most?** We will look for busy months and days of the week.
4. **How does the payment method change how much people spend?** We want to know if one payment type leads to bigger shopping carts.

Analysis Report (Top 5 Insights)

Here are the 5 key findings from the E-commerce data analysis.

Insight 1: Customer Value Groups (RFM Analysis)

Clear Statement: We have 233 **Champion Customers** who are the most valuable to the business because they have bought recently and spent the most money. The group of **Lost Customers** who spent little and have not bought recently is 215 people.

| Customer Group | Number of People |
|---------------------------|------------------|
| Champion Customers | 233 |
| Lost Customers | 215 |
| Other Segment (Core Base) | 3,212 |

Analytical Approach: We used RFM (Recency, Monetary) analysis. We sorted customers based on how recently they bought, how often they bought, and how much they spent. We used the highest score (4-Recency, 4-Monetary) to define "Champions" and the lowest score (1-Recency, 1-Monetary) to define "Lost" customers.

Business Implication:

- **Keep Champions Happy:** Give special loyalty gifts or early access to sales to the 233 Champion Customers.
- **Win Back Lost Customers:** The 215 Lost Customers need a special offer to bring them back. This is a big chance to earn back lost money.

Insight 2: Top Products Sell with Normal Discounts

Clear Statement: **Clothing** brings in the most sales money (Rs. 115,314.84), just above **Books** (Rs. 111,149.35). All of the top 5 product categories use almost the same average discount, around **18% to 20%**.

| Category | Total Sales Money (Rs.) | Average Discount |
|----------------|-------------------------|------------------|
| Clothing | 115,314.84 | 17.92% |
| Books | 111,149.35 | 19.04% |
| Home & Kitchen | 110,328.08 | 19.61% |

| | | |
|--------|------------|--------|
| Sports | 108,518.79 | 19.33% |
| Toys | 107,289.69 | 18.13% |

Analytical Approach: We added up the `sales_revenue` for each product `category` and calculated the average discount applied (`discount_pct`) for the best-selling groups.

Business Implication:

- **Check Profits:** Since all top categories have similar discount levels, the company should confirm that **Clothing**'s high sales volume still results in good profit after costs.
- **Discount Strategy:** Increasing discounts on smaller categories might be needed to help them catch up to **Clothing** and **Books**.

Insight 3: Sunday is the Busiest Shopping Day

Clear Statement: **Sunday** is the single busiest day for customer orders with 566 purchases. **Saturday** is the slowest day with only 477 orders.

| Day of Week | Total Orders |
|-------------|--------------|
| Sunday | 566 |
| Monday | 522 |
| Tuesday | 542 |
| Wednesday | 509 |
| Thursday | 525 |

| | |
|----------|-----|
| Friday | 519 |
| Saturday | 477 |

Analytical Approach: We grouped and counted all orders by the `day_of_week` column that was created from the `purchase_date`.

Business Implication:

- **Staffing & Inventory:** Make sure there are enough workers and stock ready for the busy Sunday shopping rush.
- **Weekend Focus:** The pattern shows customers are less active on Saturday. The marketing team can use Saturday for special sales or promotions to shift some of the Sunday rush and increase weekend sales overall.

Insight 4: Cash on Delivery Drives the Biggest Carts

Clear Statement: Purchases made using **Cash on Delivery (COD)** have the highest average value (Rs. 210.36) compared to all digital payment methods. This means customers are more willing to commit to bigger purchases when they can pay upon delivery.

| Payment Method | Average Sales Money (Rs.) |
|------------------|---------------------------|
| Cash on Delivery | 210.36 |
| Debit Card | 209.80 |
| Credit Card | 205.29 |
| Net Banking | 205.03 |
| UPI | 204.34 |

Analytical Approach: We calculated the Average Sales Revenue (AOV) by taking the average of the `sales_revenue` for each `payment_method`.

Business Implication:

- **Protect COD:** Since COD leads to the highest average spending, the company must always make sure this payment option is easy to use and available for customers.
 - **Incentivize Digital:** The difference in spending is small. The company could try small rewards (like a 1% discount) for digital payments to move high-value shoppers to digital methods, which cost the company less money to process.
-

Insight 5: October is the Best Month for Sales Revenue

Clear Statement: **October** is the month with the highest overall sales revenue in the dataset (Rs. 76,034.51), followed by **April** (Rs. 74,365.13).

| Month | Total Sales Money (Rs.) | Total Orders |
|---------|-------------------------|--------------|
| 2024-10 | 76,034.51 | 362 |
| 2024-04 | 74,365.13 | 362 |
| 2024-07 | 72,798.23 | 346 |

Analytical Approach: We grouped transactions by the `purchase_month` feature and summed the `sales_revenue` and `total_orders`.

Business Implication:

- **Plan Ahead:** Marketing and supply chain teams should plan their biggest efforts and stock for October to make the most of this expected surge in demand.
- **Look for Causes:** The business should research what happened in October (like a special holiday, big marketing event, or seasonal demand) and use those learnings to improve sales in other months.