

Data-Driven Insights into Sales Performance, Customer Loyalty, and Delivery Optimization — GreenCart Ltd.

Executive Summary

By merging customer, product, and sales data, this study provided a unified view of GreenCart's performance. Findings show:

- Top categories: Cleaning, Storage, and Outdoors lead revenue.
- West & South regions (purple, red dots) generate strong weekly revenue.
- Discounts above ~20% lift sales but pose margin risks.
- Gold-tier customers (highest order volume) drive value across all regions.
- High-priced items face the most delivery delays or cancellations.

Targeted promotions, smarter discount policies, and region-specific delivery improvements can drive measurable growth.

1) Type of Analysis

- Exploratory Data Analysis (EDA)
 - Descriptive Analytics (sales, loyalty, discount, delivery performance)
 - Diagnostic Insights (why patterns happen, e.g., delayed high-price deliveries)
-

2) Data Sources & Integration

Datasets

- customer_info.csv — customer profile
- product_info.csv — product details
- sales_data.csv — transaction data

Merging Logic

1. sales_df merged with customer_df on customer_id
 2. Result merged with product_df on product_id
-

3) Tools & Libraries

- Python (EDA & visualization)
 - Pandas, Seaborn, Matplotlib
 - Jupyter Notebook (workflow)
 - VS Code (editing & repo management)
-

4) Data Preparation & Feature Engineering

- Cleaned inconsistent labels (silver → Silver)
- Handled missing data (Unknown for loyalty tier/email)
- Parsed and standardized dates
- Created new fields:

- revenue = quantity × price
- order_week = ISO week from order date
- is_late = delivery beyond expected date
- price_band = Low / Medium / High

[14]

...

	revenue	order_week	price_band	days_to_order	email_domain	is_late
0	117.7500	23	High	423.0	mills-logan.com	False
1	94.6000	23	Medium	140.0	morgan.com	True
2	25.2280	23	Medium	74.0	walters-smith.com	False
3	26.2080	23	High	327.0	gmail.com	False
4	38.0960	23	High	107.0	hotmail.com	True
5	102.3030	23	High	230.0	yahoo.com	True
6	146.4425	23	High	259.0	moore.com	False
7	37.6800	23	Low	325.0	whitehead-hernandez.biz	False
8	72.2160	23	Medium	64.0	herring.com	False
9	26.3160	23	High	293.0	russell.com	True

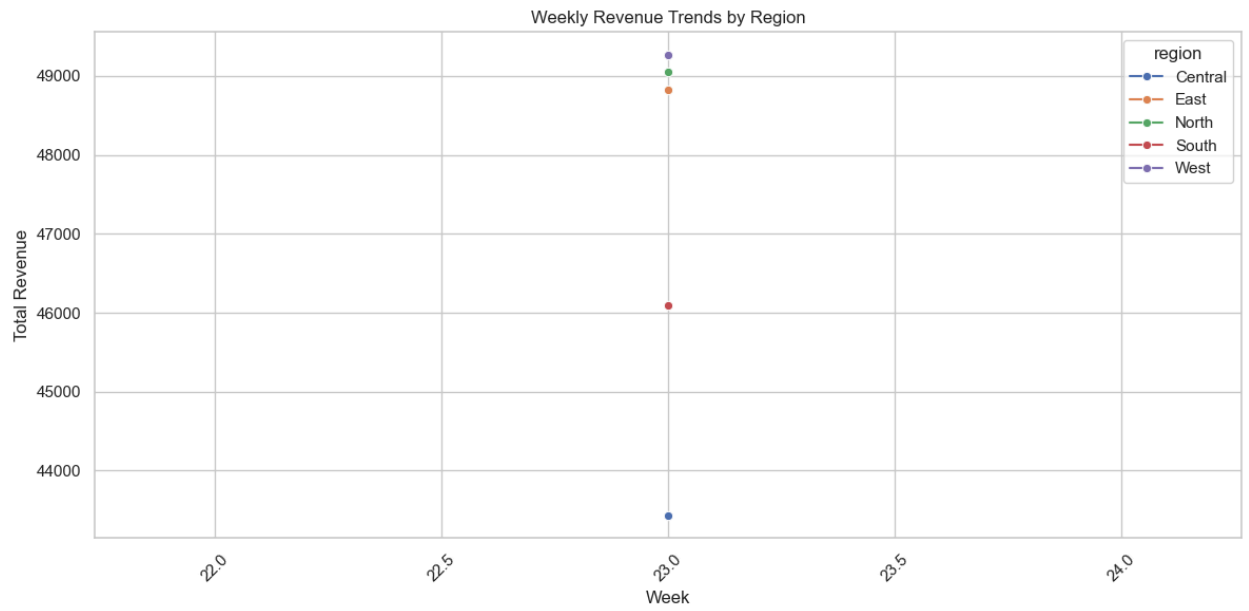
▶ Feature engineering output

5) Exploratory Analysis & Visuals

5.1 Weekly Revenue by Region (Line with Dots)

- West (purple) and South (red) regions generated the highest weekly revenues.

- Central (blue) and North (green) lagged behind.



▶ Figure 1

5.2 Top 5 Product Categories by Revenue (Barplot)

- Cleaning (highest)
- Storage
- Outdoors
- Kitchen

- Personal Care (lowest in top 5)

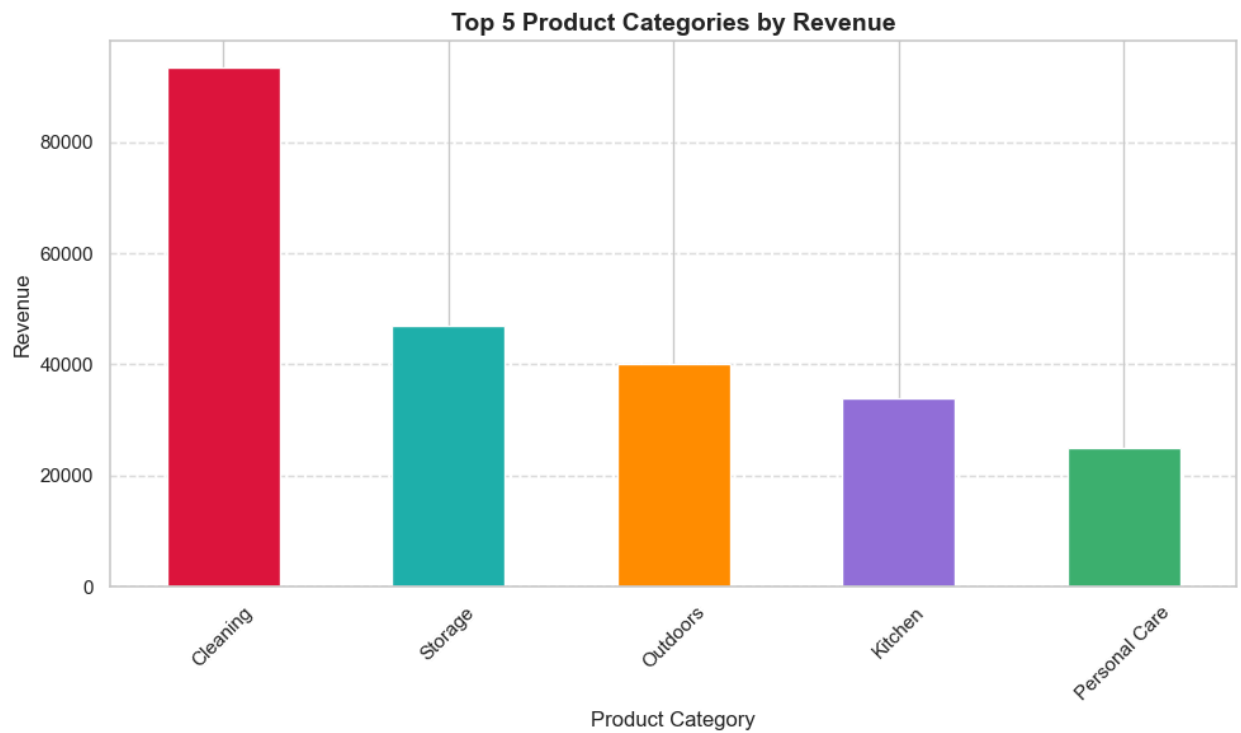
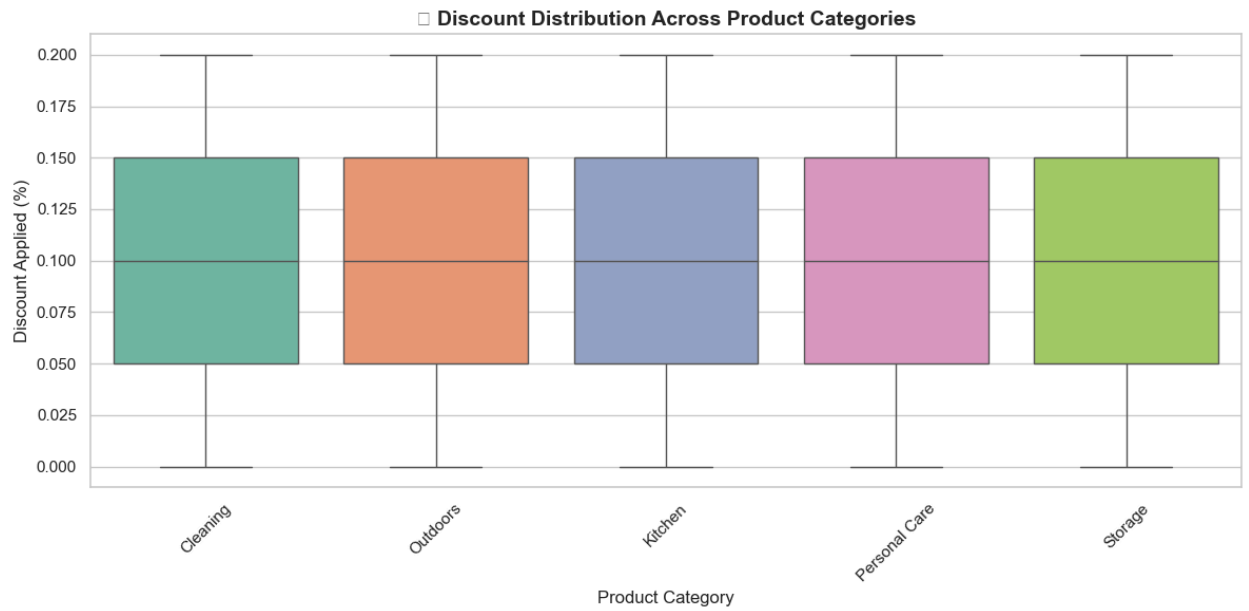


Figure 2

5.3 Discount Distribution Across Categories (Boxplot)

- Discounts vary by category:
 - Cleaning (green), Outdoors (orange), Kitchen (purple), Personal Care (pink), Storage (lemon green).

- Categories like Cleaning & Storage showed wider discount ranges.



▶ Figure 3

5.4 Correlation: Revenue, Discount, Quantity (Heatmap)

- Positive correlation: discount $\uparrow \rightarrow$ quantity \uparrow .

- Revenue benefits, but excessive discounts threaten profit margins.

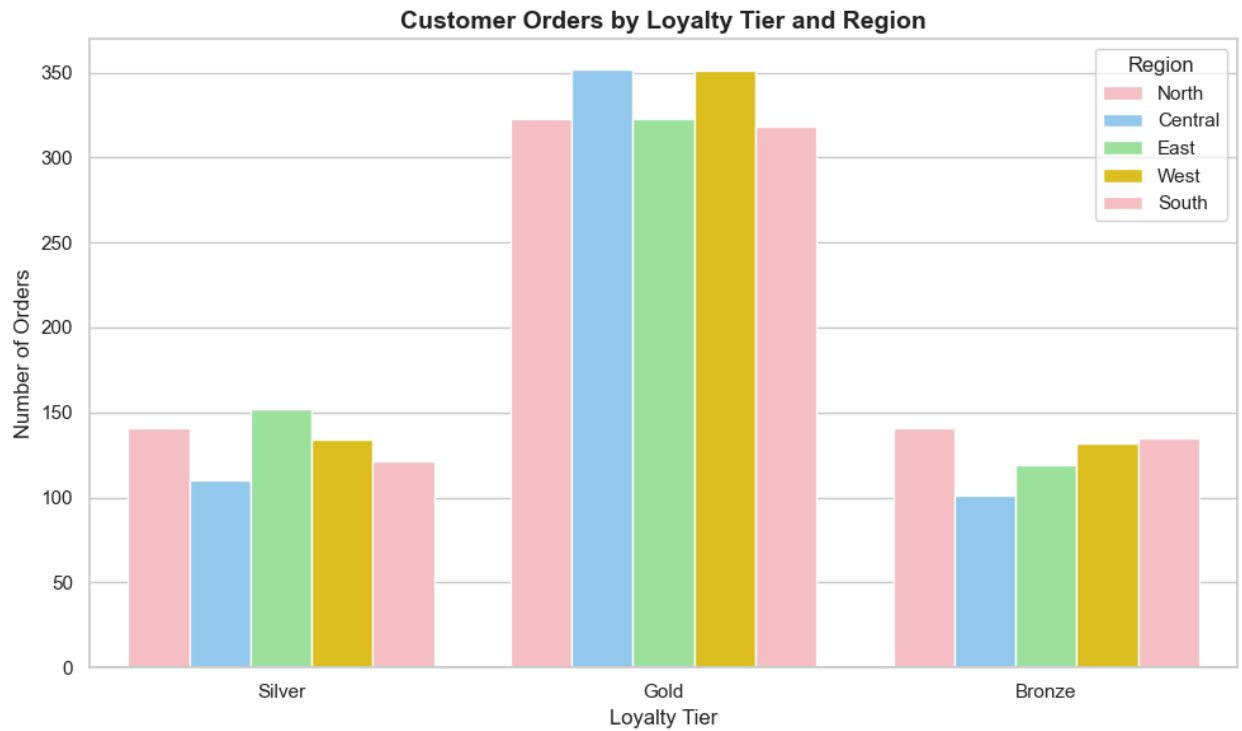


▶ Figure 4

5.5 Customer Orders by Loyalty Tier & Region (Countplot)

- Gold-tier customers consistently ordered the most.
- Silver and Bronze trailed lower.

- Regional colors: North (pink), Central (blue), East (green), West (yellow), South (pink).

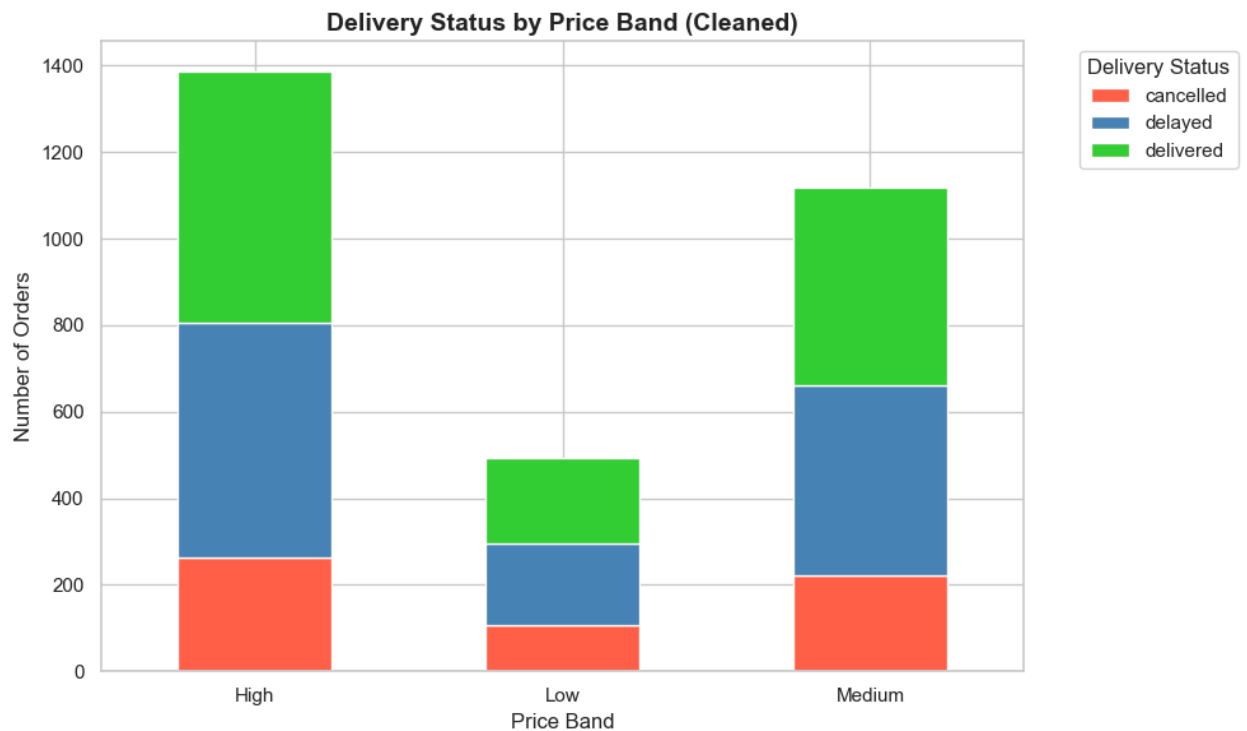


▶ Figure 5

5.6 Delivery Status by Price Band (Stacked Bar)

- High-priced products = more delayed (blue) or canceled (orange).

- Medium- and Low-priced items more likely to be delivered (green).



▶ Figure 6

6) Business Questions & Insights

Q1. Which categories drive revenue?

- Cleaning, Storage, Outdoors dominate revenue mix.

Q2. Do discounts boost performance?

- Yes, discounts beyond ~20% drive volume, but with profit risks.

Q3. Who are the most valuable customers?

- Gold-tier customers lead in order volume and revenue contribution.

Q4. Which products face delivery challenges?

- High-priced items most often delayed or canceled.

Q5. Which regions are strongest?

- West & South deliver the highest weekly revenues.
-

7) Recommendations

- Category Promotions: Prioritize Cleaning, Storage, Outdoors.
 - Discount Strategy: Cap working discounts at ~20–25% to balance sales & margin.
 - Loyalty Optimization: Enhance Gold-tier perks, create upgrade incentives for Silver.
 - Delivery Improvement: Focus on reliability for high-priced products; strengthen logistics in weaker regions.
-

8) Limitations

- Some legend/region colors duplicated (e.g., South & North both pink).
 - “Unknown” imputations may reduce segmentation accuracy.
-

9) Conclusion

Merging three datasets (customer_info.csv, product_info.csv, sales_data.csv) provided actionable insights into sales drivers, discount impacts, loyalty value, and delivery efficiency. Results highlight Gold-tier strength, category leaders (Cleaning/Storage/Outdoors), and delivery risks for expensive items.

Targeted promotions, smarter discounts, and stronger delivery logistics can significantly improve GreenCart’s growth and customer satisfaction.

