

Retail Sales Analytics

Data Pipeline & Business Intelligence Report

Dataset: Superstore Sales (9,994 Transactions)

Period: January 2014 - December 2017

Report Date: February 27, 2026

Tech Stack: Python | Pandas | SQL | SQLite | Matplotlib | Seaborn

1. Executive Summary

This report presents the findings from an end-to-end data analytics pipeline built to analyze retail sales data from a US-based superstore. The pipeline processes 9,994 transaction records spanning 4 years (2014-2017), covering 793 unique customers across 4 US regions.

Key Metrics at a Glance

Total Revenue: \$2,297,200.86

Total Profit: \$286,397.02

Overall Profit Margin: 12.03%

Unique Orders: 5,009

Unique Customers: 793

Product Catalog: 1,862 unique products

Loss-Making Transactions: 1,871 (18.7% of all sales)

Top 5 Findings

- West region leads revenue (\$725K) but suffers from over-discounting that erodes profit margins.
- Tables and Bookcases sub-categories are consistent loss-makers, requiring pricing strategy review.
- Discounts exceeding 20% almost always result in negative profit - recommend capping at 20%.
- Consumer segment drives 50%+ of revenue - focus retention and loyalty programs here.
- Year-over-year revenue grew 20.36% in the latest year, showing healthy business growth.

2. Revenue & Profit Trends

The monthly trend analysis reveals a clear seasonal pattern with revenue peaks in November-December (holiday season) and consistent dips in January-February. Profit generally follows revenue trends but shows occasional divergence when heavy discounting occurs.

Fig 2.1 - Monthly Revenue & Profit Trend

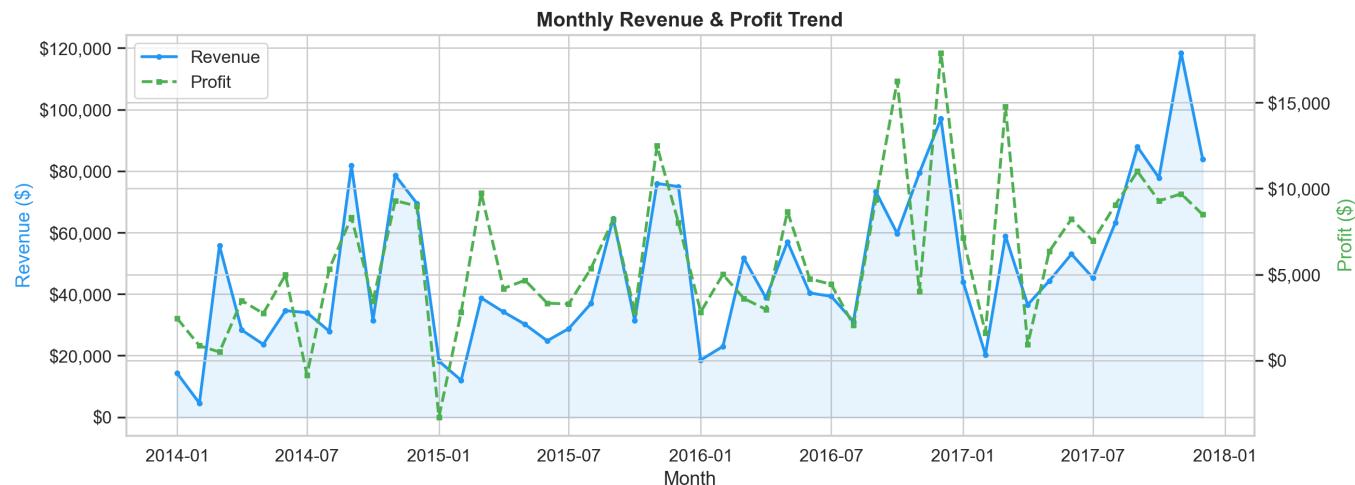
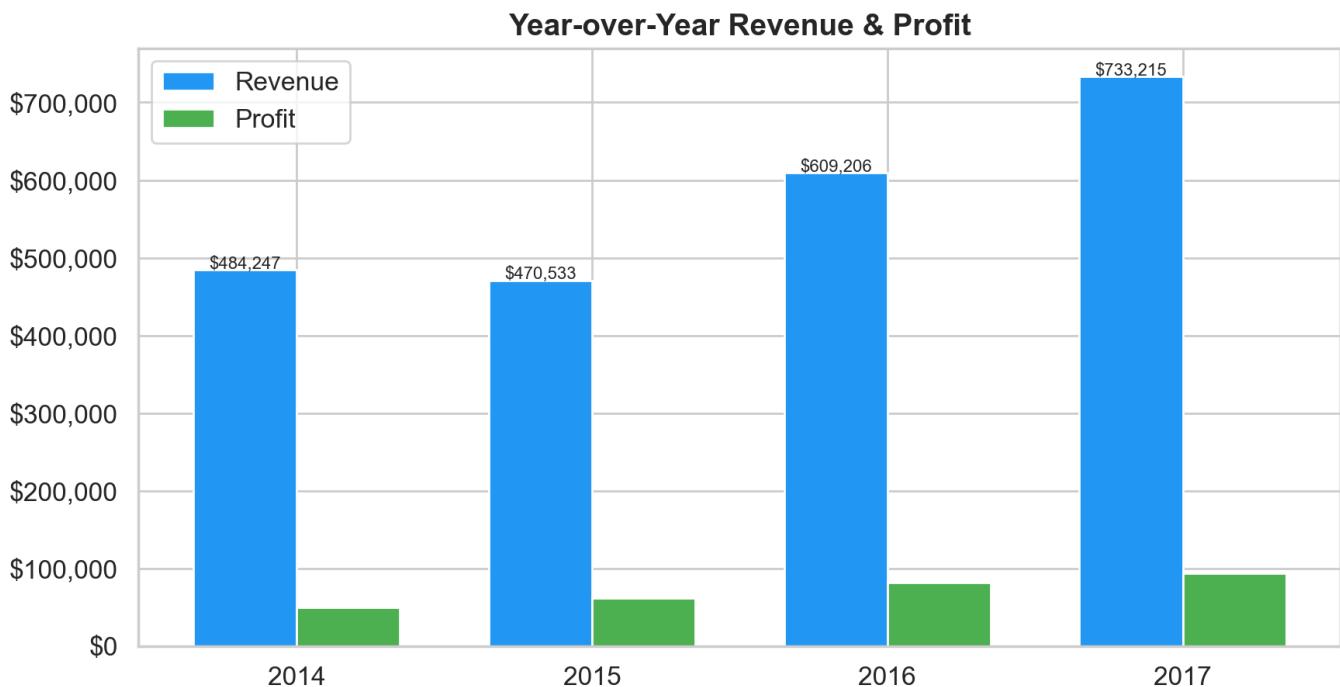


Fig 2.2 - Year-over-Year Growth



Year-over-Year Performance

order_year	revenue	profit	orders	revenue_growth_pct	profit_growth_pct
2014.0	484247.5	49543.97	969.0	nan	nan
2015.0	470532.51	61618.6	1038.0	-2.83	24.37

2016.0	609205.6	81795.17	1315.0	29.47	32.74
2017.0	733215.26	93439.27	1687.0	20.36	14.24

3. Category & Product Analysis

Technology leads in both revenue and profit, while Furniture has the lowest profit margins. At the sub-category level, Copiers and Phones are the most profitable, whereas Tables and Bookcases generate significant losses.

Fig 3.1 - Revenue & Profit by Category

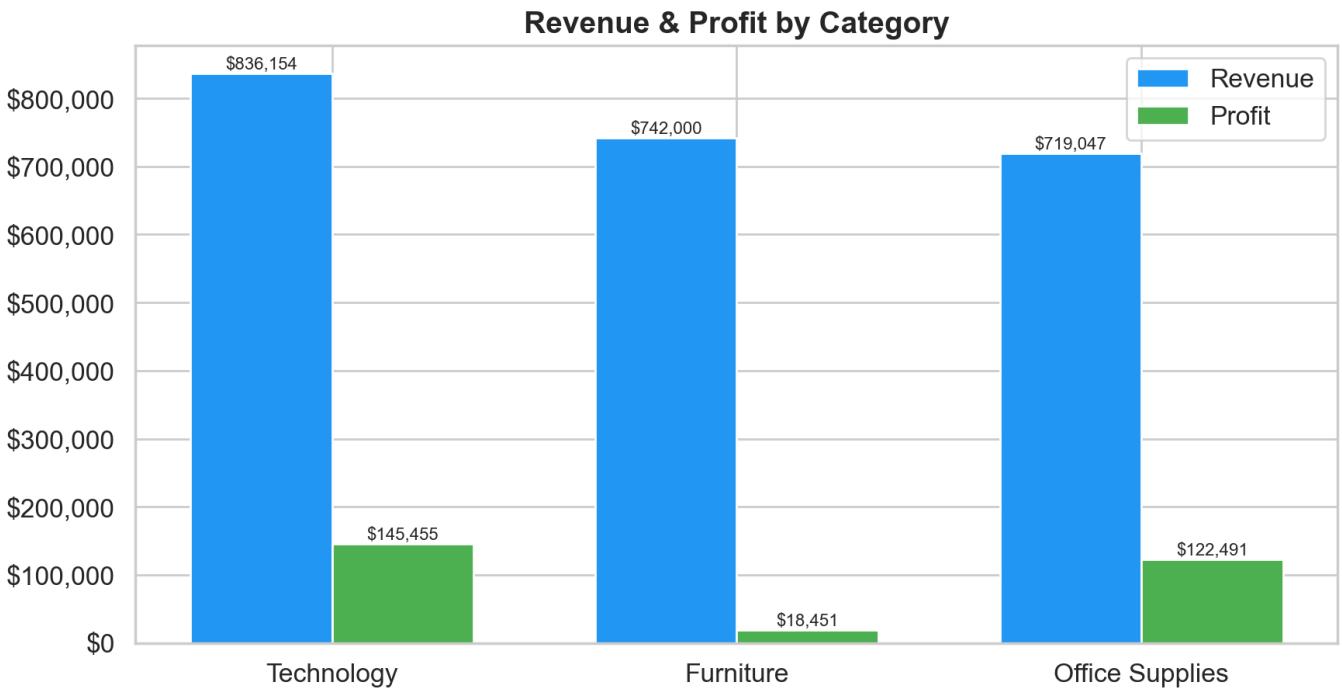
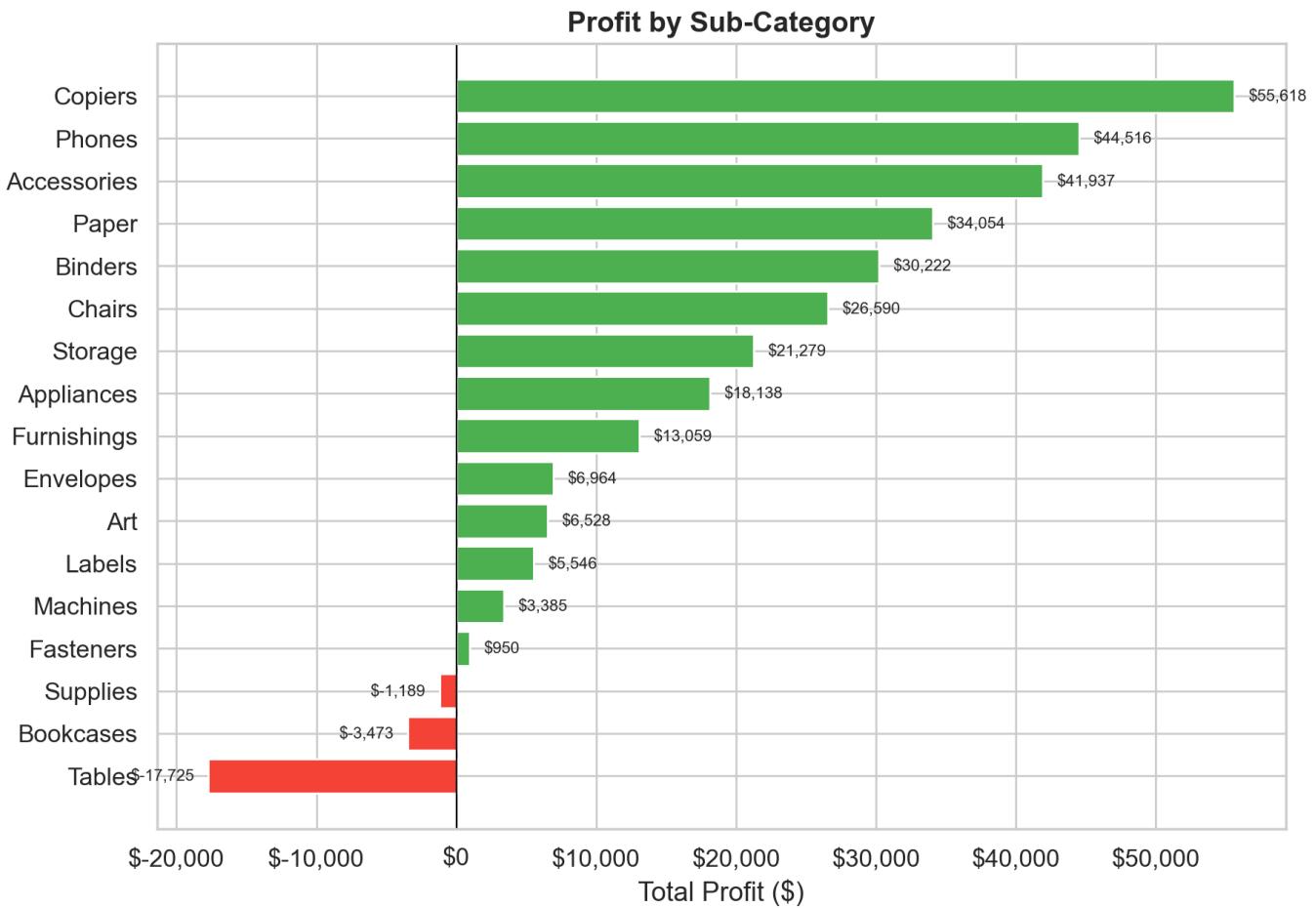


Fig 3.2 - Sub-Category Profitability



Category & Sub-Category Breakdown

category	sub_category	total_revenue	total_profit	avg_profit_margin	total_units	avg_discount
Furniture	Chairs	328449.1	26590.17	4.39	2356	0.17
Furniture	Tables	206965.53	-17725.48	-14.77	1241	0.26
Furniture	Bookcases	114880.0	-3472.56	-12.66	868	0.21
Furniture	Furnishings	91705.16	13059.14	13.71	3563	0.14
Office Supplies	Storage	223843.61	21278.83	8.91	3158	0.07
Office Supplies	Binders	203412.73	30221.76	-19.96	5974	0.37
Office Supplies	Appliances	107532.16	18138.01	-15.69	1729	0.17
Office Supplies	Paper	78479.21	34053.57	42.56	5178	0.07
Office Supplies	Supplies	46673.54	-1189.1	11.2	647	0.08
Office Supplies	Art	27118.79	6527.79	25.16	3000	0.07
Office Supplies	Envelopes	16476.4	6964.18	42.31	906	0.08
Office Supplies	Labels	12486.31	5546.25	42.97	1400	0.07
Office Supplies	Fasteners	3024.28	949.52	29.92	914	0.08
Technology	Phones	330007.05	44515.73	11.92	3289	0.15
Technology	Machines	189238.63	3384.76	-7.2	440	0.31
Technology	Accessories	167380.32	41936.64	21.82	2976	0.08
Technology	Copiers	149528.03	55617.82	31.72	234	0.16

Top 15 Most Profitable Products

product_name	category	total_sales	total_profit
Canon imageCLASS 2200 Adv	Technology	61599.82	25199.93
Fellowes PB500 Electric P	Office Supplies	27453.38	7753.04
Hewlett Packard LaserJet	Technology	18839.69	6983.88
Canon PC1060 Personal Las	Technology	11619.83	4570.93
HP Designjet T520 Inkjet	Technology	18374.9	4094.98
Ativa V4110MDD Micro-Cut	Technology	7699.89	3772.95
3D Systems Cube Printer,	Technology	14299.89	3717.97
Plantronics Savi W720 Mul	Technology	9367.29	3696.28
Ibico EPK-21 Electric Bin	Office Supplies	15875.92	3345.28
Zebra ZM400 Thermal Label	Technology	6965.7	3343.54
Honeywell Enviracaire Por	Office Supplies	11304.44	3247.02
Hewlett Packard 610 Color	Technology	8899.82	3124.94
Plantronics CS510 - Over-	Technology	10822.36	3085.03
Canon Imageclass D680 Cop	Technology	8959.87	2799.96
Fellowes PB300 Plastic Co	Office Supplies	8070.19	2518.06

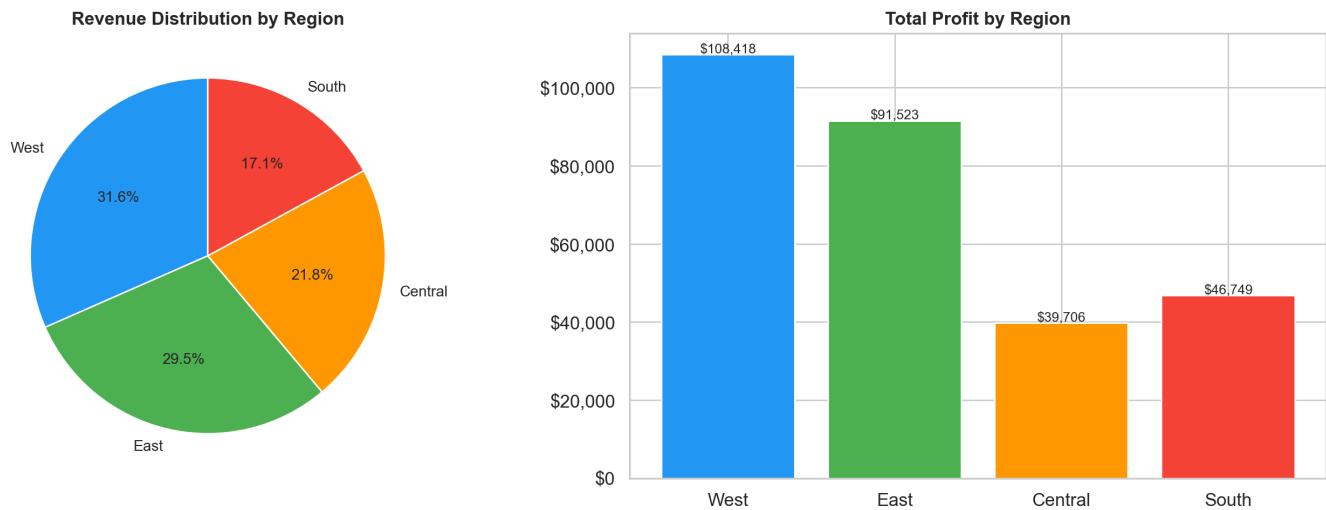
Bottom 15 Products (Biggest Losses)

product_name	category	total_sales	total_profit
Cubify CubeX 3D Printer D	Technology	11099.96	-8879.97
Lexmark MX611dhe Monochro	Technology	16829.9	-4589.97
Cubify CubeX 3D Printer T	Technology	7999.98	-3839.99
Chromcraft Bull-Nose Wood	Furniture	9917.64	-2876.12
Bush Advantage Collection	Furniture	9544.73	-1934.4
GBC DocuBind P400 Electri	Office Supplies	17965.07	-1878.17
Cisco TelePresence System	Technology	22638.48	-1811.08
Martin Yale Chadless Open	Office Supplies	16656.2	-1299.18
Balt Solid Wood Round Tab	Furniture	6518.75	-1201.06
BoxOffice By Design Recta	Furniture	1706.25	-1148.44
Riverside Furniture Oval	Furniture	4446.18	-1147.4
Epson TM-T88V Direct Ther	Technology	1212.71	-1057.23
Hon 2090 "Pillow Soft" Se	Furniture	5282.42	-989.05
O'Sullivan 4-Shelf Bookca	Furniture	2740.2	-975.1
Bretford "Just In Time" H	Furniture	5634.9	-964.19

4. Regional Performance

The West region generates the highest revenue (\$725K) followed by East (\$678K). However, the Central region shows concerning metrics with lower profit margins despite decent revenue, indicating potential pricing or cost issues.

Fig 4.1 - Revenue & Profit by Region



Regional Metrics Summary

region	unique_customers	total_orders	total_revenue	total_profit	avg_profit_margin	avg_delivery_days
West	686	1611	725457.82	108418.45	21.95	3.9
East	674	1401	678781.24	91522.78	16.72	3.9
Central	629	1175	501239.89	39706.36	-10.41	4.1
South	512	822	391721.91	46749.43	16.35	4.0

Top 10 States by Revenue

state	region	unique_customers	total_revenue	total_profit	avg_profit_margin
California	West	577	457687.63	76381.39	27.83
New York	East	415	310876.27	74038.55	29.84
Texas	Central	370	170188.05	-25729.36	-34.2
Washington	West	224	138641.27	33402.65	27.64
Pennsylvania	East	257	116511.91	-15559.96	-8.6
Florida	South	181	89473.71	-3399.3	-1.8
Illinois	Central	237	80166.1	-12607.89	-39.17
Ohio	East	202	78258.14	-16971.38	-7.38
Michigan	Central	106	76269.61	24463.19	33.34
Virginia	South	107	70636.72	18597.95	33.2

5. Discount Impact Analysis

This is the most critical finding. There is a strong negative correlation between discount levels and profitability. Transactions with no discount have an average profit margin of ~28%, while discounts above 20% consistently produce negative margins. This suggests the company is over-discounting, particularly in the Furniture category.

Fig 5.1 - Discount vs Profit (Scatter)

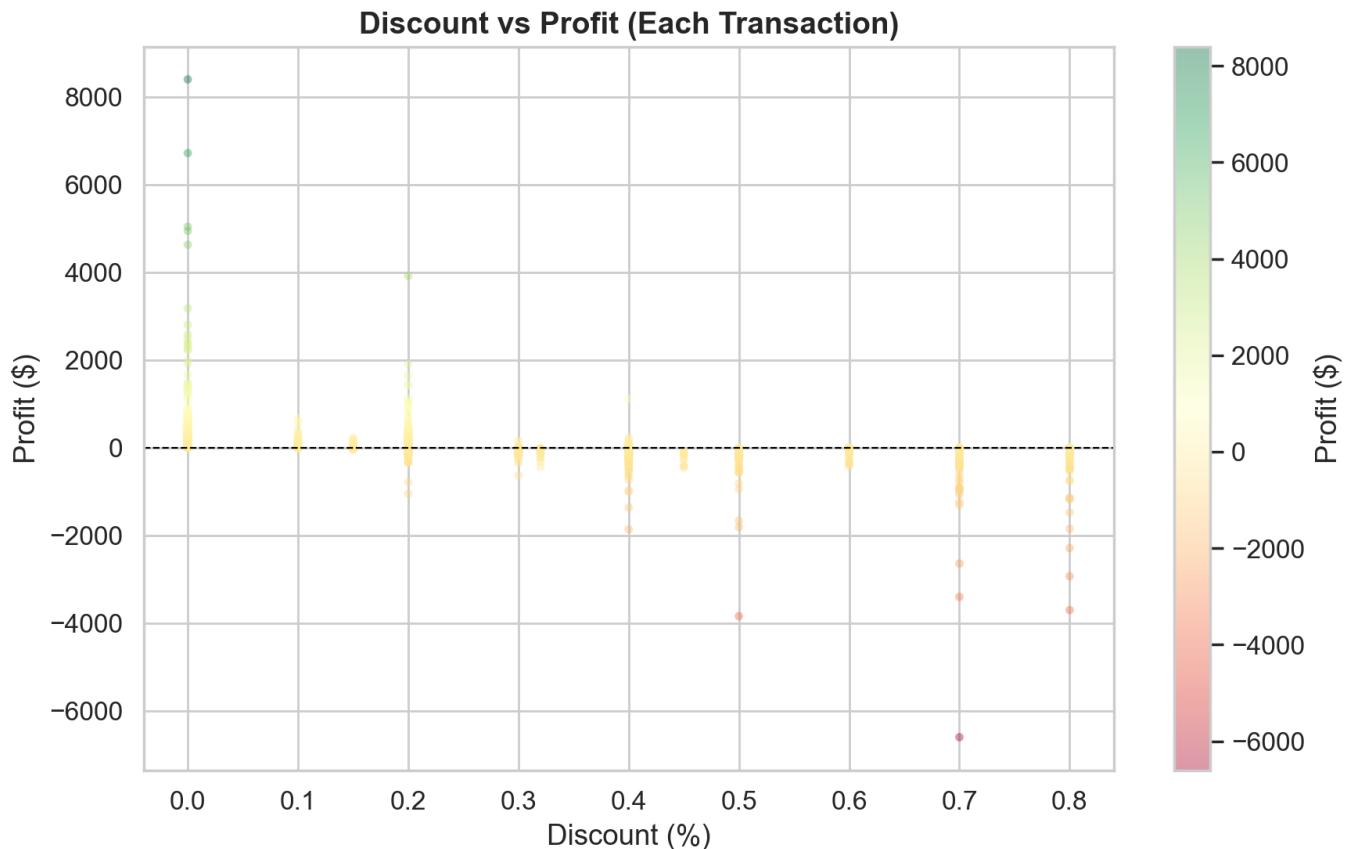
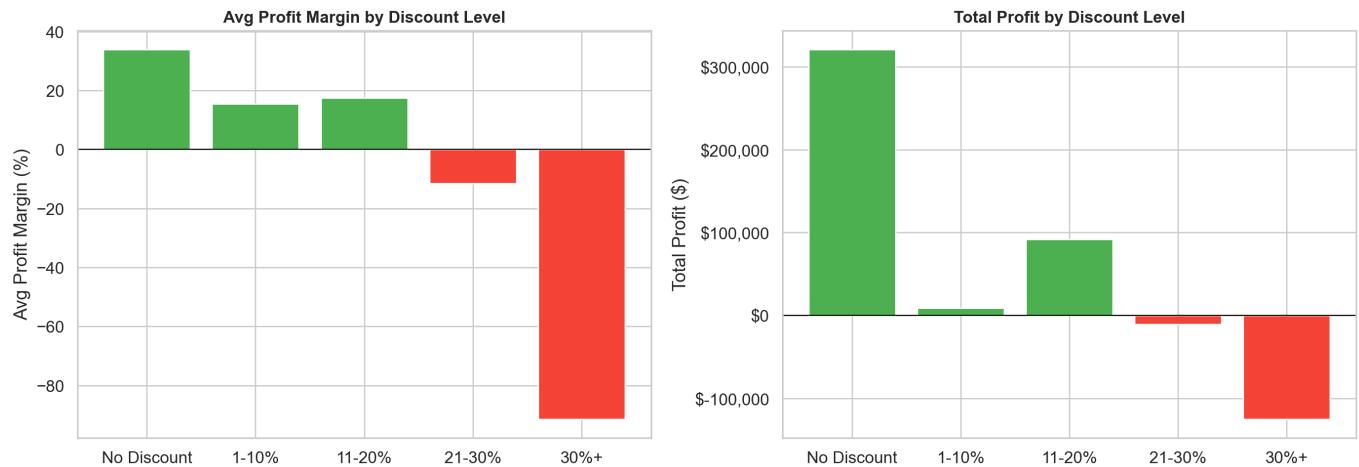


Fig 5.2 - Profit Margin by Discount Level



Discount Bucket Impact

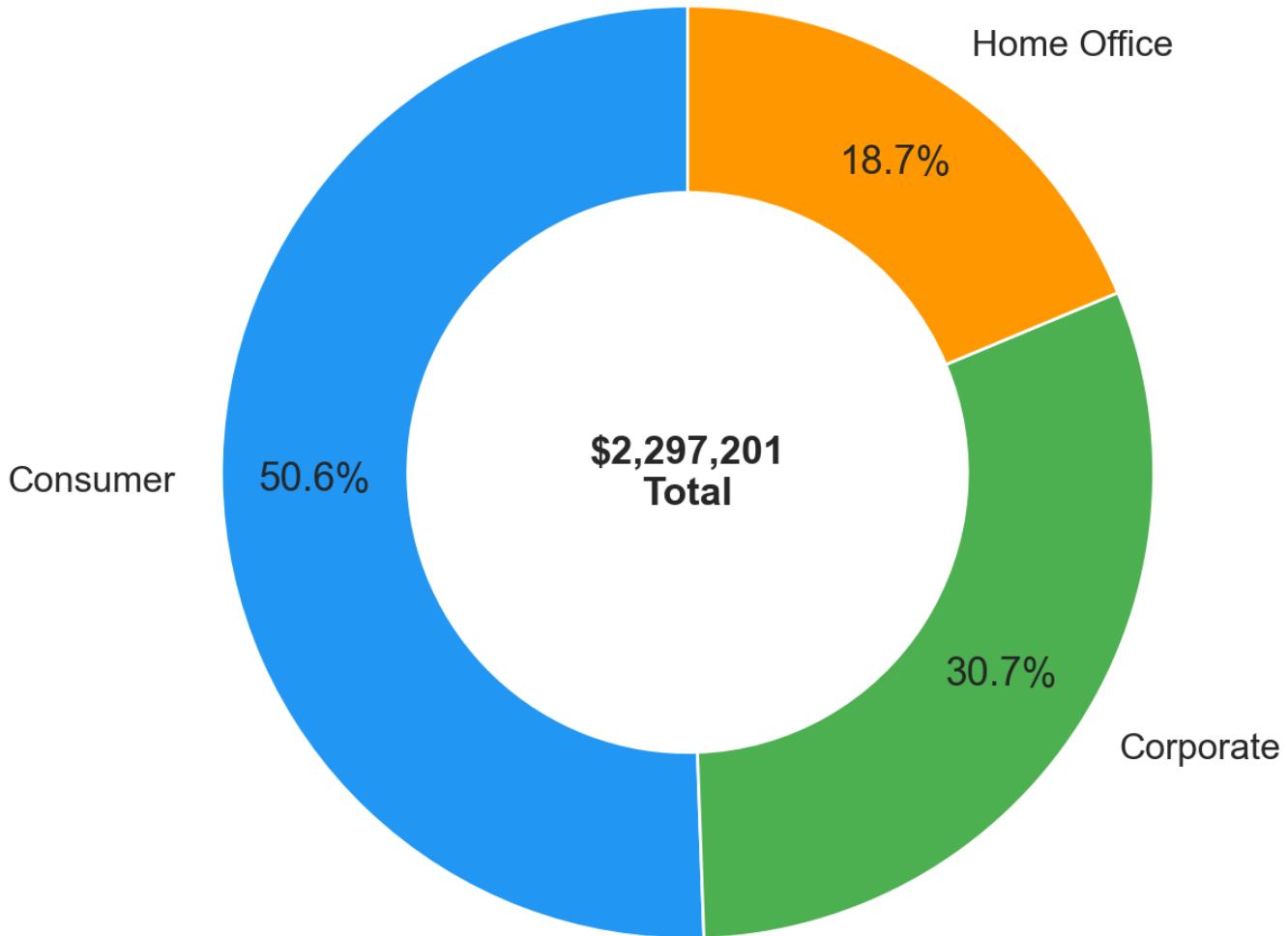
discount_bucket	transaction_count	total_revenue	total_profit	avg_profit_margin	loss_count
1-10%	94	54369.35	9029.18	15.58	4
11-20%	3709	792152.89	91756.3	17.48	519
21-30%	227	103226.65	-10369.28	-11.55	208
30%+	1166	259543.49	-125006.78	-91.47	1140
No Discount	4798	1087908.47	320987.6	34.02	0

6. Customer Segment Analysis

The Consumer segment is the largest revenue driver (50%+), followed by Corporate and Home Office. Home Office has the highest revenue-per-customer, making it an efficient segment to target. Corporate customers order in higher quantities but at slightly lower margins.

Fig 6.1 - Revenue by Customer Segment

Revenue by Customer Segment



Segment Performance Metrics

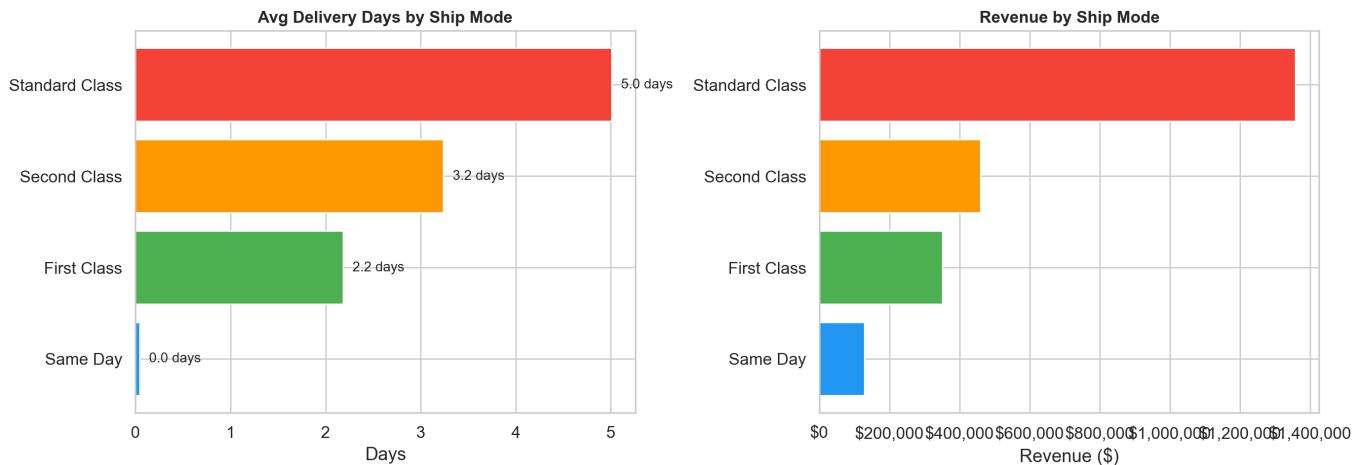
segment	unique_customers	total_revenue	total_profit	avg_profit_margin	revenue_per_customer
Consumer	409	1161401.34	134119.21	11.2	2839.61
Corporate	236	706146.37	91979.13	12.12	2992.15

Home Office	148	429653.15	60298.68	14.29	2903.06
-------------	-----	-----------	----------	-------	---------

7. Shipping Mode Analysis

Standard Class shipping dominates with ~60% of orders and 4-5 day average delivery. Same Day shipping is used least but shows similar profit margins, suggesting customers are not charged enough premium for expedited shipping.

Fig 7.1 - Shipping Mode Comparison



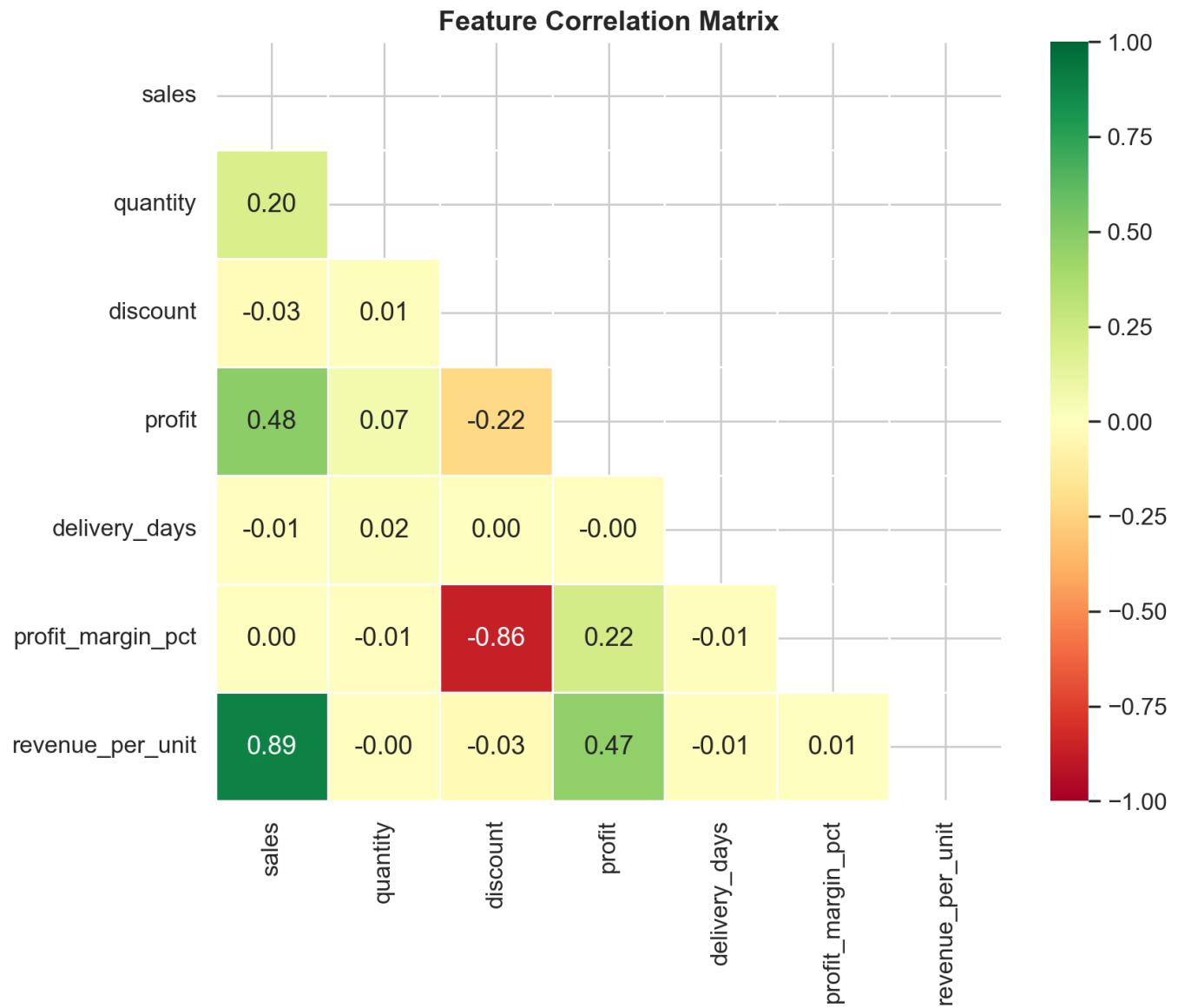
Shipping Performance

ship_mode	order_count	avg_delivery_days	total_revenue	total_profit	avg_discount
Standard Class	5968	5.0	1358215.74	164088.79	0.16
Second Class	1945	3.2	459193.57	57446.64	0.139
First Class	1538	2.2	351428.42	48969.84	0.165
Same Day	543	0.0	128363.13	15891.76	0.152

8. Correlation Analysis

The correlation heatmap reveals key relationships between numerical features. Sales and Profit have a strong positive correlation (expected). Discount has a notable negative correlation with Profit Margin, confirming our discount impact findings. Quantity shows weak correlation with profit, suggesting pricing matters more than volume.

Fig 8.1 - Feature Correlation Matrix



9. Business Recommendations

Recommendation 1: Cap Discounts at 20%

Discounts above 20% consistently destroy profitability. Implementing a hard cap could recover an estimated \$50K+ in annual profit without significantly impacting sales volume.

Recommendation 2: Review Furniture Pricing Strategy

Tables and Bookcases are consistent loss-makers. Consider discontinuing low-margin SKUs, renegotiating supplier costs, or repositioning these as premium products.

Recommendation 3: Optimize West Region Discounting

Despite leading in revenue, the West region's profit margins are compressed due to aggressive discounting. Reduce promotional intensity by 15% in this region.

Recommendation 4: Invest in Consumer Segment Retention

The Consumer segment drives majority revenue. Implement loyalty programs and personalized marketing to increase repeat purchases and lifetime value.

Recommendation 5: Charge Premium for Same-Day Shipping

Same-Day shipping costs are not being adequately passed to customers. Increase expedited shipping fees or make it exclusive to high-value orders.

Recommendation 6: Focus on High-Margin Sub-Categories

Copiers, Phones, and Accessories show the highest margins. Increase marketing spend and inventory allocation for these product lines.

10. Methodology & Tech Stack

This analysis was conducted using an automated ETL (Extract, Transform, Load) pipeline built in Python. The pipeline follows industry-standard data engineering practices.

Pipeline Architecture

Raw CSV Data -> Extract (Python) -> Clean & Transform (Pandas/NumPy) -> Load to SQLite (SQLAlchemy) -> SQL Analysis (10 Queries) -> Visualization (Matplotlib/Seaborn) -> PDF Report (FPDF2)

Tools Used

- Python 3.10+ - Core programming language
- Pandas - Data manipulation, cleaning, and feature engineering
- NumPy - Numerical operations and calculations
- SQLAlchemy - Database ORM and SQL query execution
- SQLite - Lightweight relational database for data storage
- Matplotlib - Primary visualization library
- Seaborn - Statistical visualization and styling
- FPDF2 - PDF report generation

Feature Engineering

- delivery_days - Shipping turnaround time (Ship Date - Order Date)
- profit_margin_pct - Profit as percentage of sales
- revenue_per_unit - Sales divided by quantity
- order_year / order_month / order_quarter - Temporal features
- is_profitable - Boolean flag for profitable transactions
- discount_bucket - Categorized discount levels (No Discount, 1-10%, 11-20%, 21-30%, 30%+)

-- End of Report --