

Retail/E-Commerce Data Analytics Report

Hemant

2025-09-23

Table of contents

1	Executive Summary	2
2	Introduction & Business Context	2
3	Project Objectives	3
4	Dataset Description	3
5	Data Preparation & Cleaning	4
6	Exploratory Data Analysis (EDA)	5
6.1	Summary Statistics	5
6.2	Univariate Analysis	6
6.2.1	Distribution of cost, sales, profit, returns, discounts and profit margin. .	6
6.2.2	Most frequent Brands and Product Categories.	7
6.3	Bivariate/Multivariate Analysis	8
6.3.1	Correlation Between Numerical Columns	8
6.3.2	Net Sales & Net Profit by Brand.	8
6.3.3	Net Profit by Sales Channel	9
6.3.4	Discount, Returns, Net Sales and Net Profit by Product Category . . .	10
6.3.5	Impact of Discount & Returns on Net Profit	12
6.3.6	Regional Sales and Profitability differences.	13
6.4	Time-Series Analysis	14
7	Key Insights	14
8	Recommendations	14
9	Conclusion	14

1 Executive Summary

This report explores sales and profitability in the retail and e-commerce sector. The company has enjoyed steady sales growth, but profit margins have not kept pace. Management is particularly concerned about the impact of product returns, aggressive discounting, and uneven regional performance.

Our analysis takes a closer look at these challenges by breaking down profitability across products, geographies, and sales channels, while also examining trends over time. The findings highlight where revenue is leaking and provide practical recommendations the business can act on to strengthen margins and support sustainable growth.

2 Introduction & Business Context

The retail and e-commerce industry is fast-moving and highly competitive. Companies often compete on price and promotions, which can boost sales in the short term but put pressure on profit margins in the long run. Even when sales volumes look healthy, profitability can be eroded by factors such as high return rates, frequent discounting, rising costs, and inconsistent performance across regions and channels.

In this project, we focus on a global retail company that operates across multiple continents, countries, brands, and sales channels. Although the business has achieved steady sales growth, profitability has been inconsistent. Management is particularly concerned about revenue leakage from returns, discounts, and low-margin products.

This analysis is designed to help the company get a clearer view of its sales and cost dynamics. By examining historical data, the project aims to uncover where profitability is being lost, highlight underperforming areas, and identify opportunities to improve margins. The insights generated will not only explain current challenges but also support more informed decisions going forward.

3 Project Objectives

The project aims to:

- Analyze sales and profitability drivers across product categories, brands, and channels.
 - Evaluate the financial impact of returns and discounts on net sales and net profit.
 - Compare regional and geographical performance to highlight markets with high sales but low profitability.
 - Identify trends and seasonality in sales and profit margins across years and quarters.
 - Highlight actionable insights and recommendations to improve overall profitability.
-

4 Dataset Description

- **Source:** (e.g., Kaggle, synthetic dataset, company data)
 - **Rows & Columns:** 3,46,369 rows, 14 columns
 - **Key Columns:**
 1. `id` → Row Number
 2. `dates` → Transaction date
 3. `qtr_year` → Time periods identifiers
 4. `channel` → Sales channel (e.g., Online, Retail store)
 5. `brand` → Product brand
 6. `product_category` → Category of product
 7. `country_name` → Geographical info
 8. `return_amount` → Value of returned products
 9. `discount_amount` → Value of discounts offered
 10. `total_cost` → Total cost of goods sold
 11. `total_sales` → Gross sales
 12. `net_sales` → Sales after returns & discounts
 13. `net_profit` → Profit after deducting costs
 14. `profit_margin` → Profitability ratio (%)
-

5 Data Preparation & Cleaning

The analysis was based on data stored in a **MySQL** relational database. Data from five separate tables was extracted and consolidated into a single sales summary table, providing a comprehensive view of sales, costs, returns, discounts, and profitability across products, channels, and regions.

Although the source data was already clean, basic data validation and standardization were performed to ensure accuracy and consistency:

- Verified that numerical fields (sales, cost, returns, discounts) were consistent with expected business logic.
- Standardized categorical fields such as product categories, brands, channels, and countries.
- Ensured time-related fields (dates, years, quarters) were in proper formats for trend analysis.
- Reconfirmed derived metrics:
 - $\text{Net Sales} = \text{Total Sales} - \text{Returns} - \text{Discounts}$
 - $\text{Net Profit} = \text{Net Sales} - \text{Total Cost}$
 - $\text{Profit Margin (\%)} = (\text{Net Profit} \div \text{Net Sales}) \times 100$

The resulting sales summary table was clean, consistent, and ready for exploratory analysis and reporting.

6 Exploratory Data Analysis (EDA)

6.1 Summary Statistics

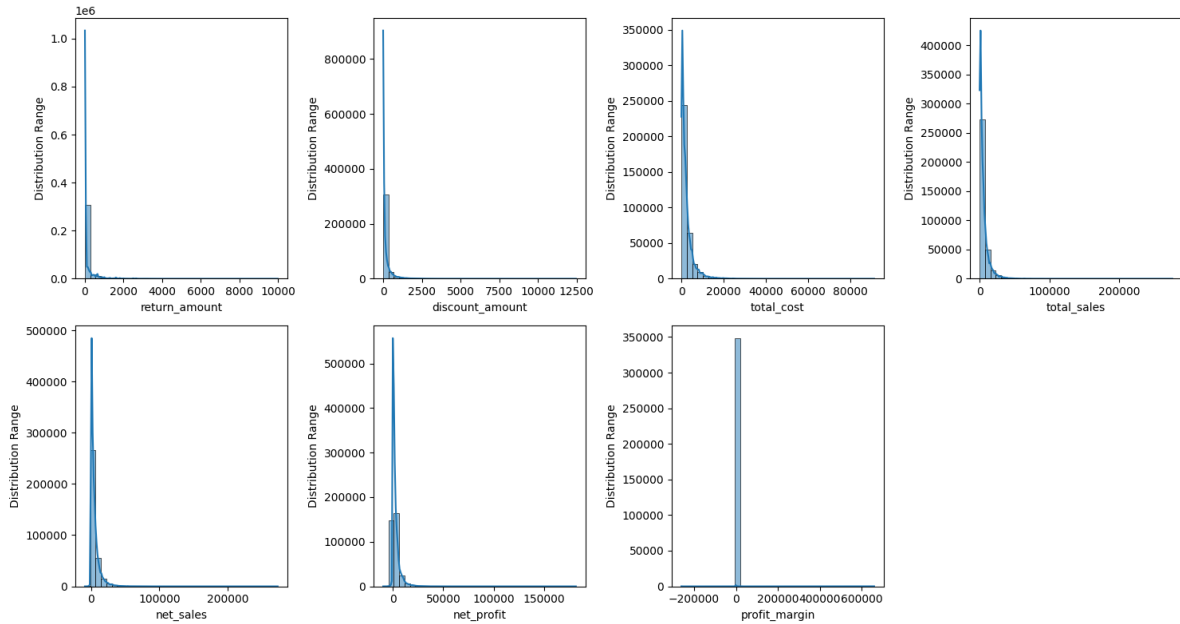
count	348484.00	348484.00	348484.00	348484.00	348484.00	348484.00	348484.00
mean	118.34	153.45	2386.72	5352.14	5080.34	2693.61	51.77
std	371.48	386.31	3118.63	7116.97	6850.34	3896.50	1383.19
min	0.00	0.00	1.92	3.80	-8791.87	-9515.77	-260665.22
25%	0.00	0.00	521.64	1079.91	980.00	457.39	47.23
50%	0.00	15.96	1427.40	3055.20	2899.89	1450.80	49.02
75%	9.99	135.98	2985.60	6730.00	6412.15	3445.84	54.01
max	9996.00	12461.60	91411.84	275901.60	272718.12	181306.28	659350.00
	return_amount	discount_amount	total_cost	total_sales	net_sales	net_profit	profit_margin

Insights :

- Negative Values :
 - Net Profit: Minimum values is (-9515.77), indicating losses. Some products or transactions may be selling at loss, due to high cost.
 - Profit Margin: Here minimum value is (-260665.22%), which indicates that there are some cases where revenue is lower than the cost. It can be due to heavy discount because cost of product is higher.
 - Net Sales : Here minimum values is (-8791.87), which indicates that there is heavy discount or returns.
- Outliers Indicated by High Standard Deviations :
 - Total Cost & Total Sales : The maximum values are (91411.84) and (275901.6) respectively which is significantly higher than the mean (2386.72) and (5352.13), indicating potential premium products.

6.2 Univariate Analysis

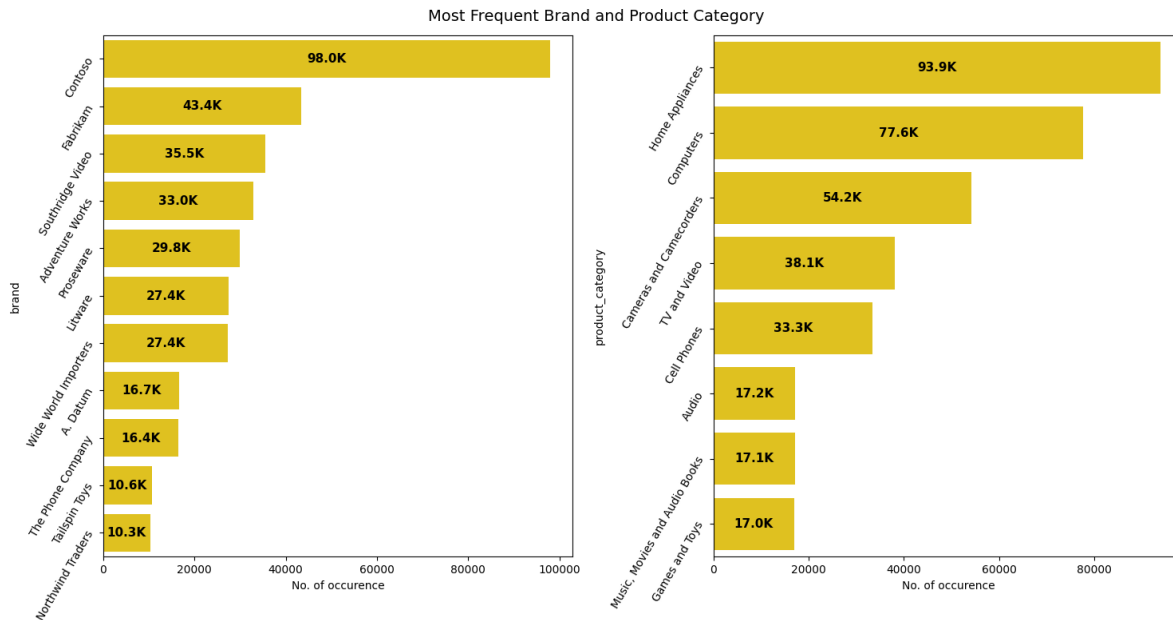
6.2.1 Distribution of cost, sales, profit, returns, discounts and profit margin.



Insights:

- **Return Amount:** Predominantly low with occasional high-value returns, highlighting specific products that may require targeted interventions to minimize revenue loss.
- **Discount Amount:** Sparingly applied and typically modest, discounts are managed to preserve profit margins while supporting selective promotional activities.
- **Total Cost:** Mostly moderate with a few substantial outliers, emphasizing the need to control high-cost items for optimizing overall profitability.
- **Total Sales / Net Sales:** Characterized by frequent low-value transactions and infrequent large sales events, reflecting a sales model driven by volume with occasional high-impact sales.
- **Net Profit:** Concentrated on modest gains with rare but significant profit spikes, underscoring the critical impact of high-margin transactions for overall financial performance.
- **Profit Margin:** Exhibits high variability with many low or negative margins offset by a few exceptionally profitable transactions, indicating areas for focused margin enhancement and risk management.

6.2.2 Most frequent Brands and Product Categories.

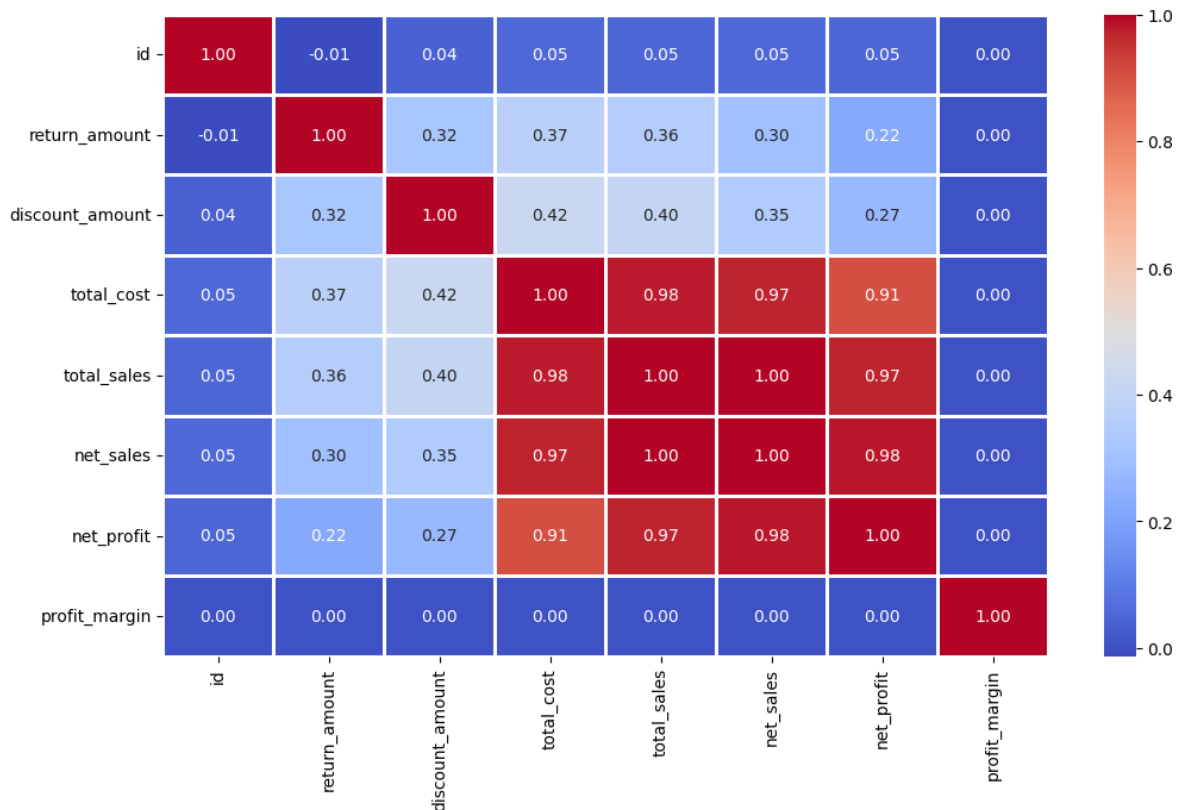


Insights :

- Brand (Left Plot):
 - Contoso is the leading brand with 97,986 transactions, far ahead of others.
 - Secondary brands like Fabrikam (43,429) and Southridge Video (35,481) also have a strong presence.
 - Brands such as Northwind Traders (10,307) and Tailspin Toys (10,610) show very limited transactions.
- Product Category (Right Plot):
 - Home Appliances (93,925) and Computers (77,630) are the most frequently sold categories.
 - Cameras & Camcorders (54,215) and TV & Video (38,096) form the mid-tier categories.
 - Audio (17,214), Music, Movies & Audiobooks (17,105), and Games & Toys (16,971) have the least frequency.

6.3 Bivariate/Multivariate Analysis

6.3.1 Correlation Between Numerical Columns

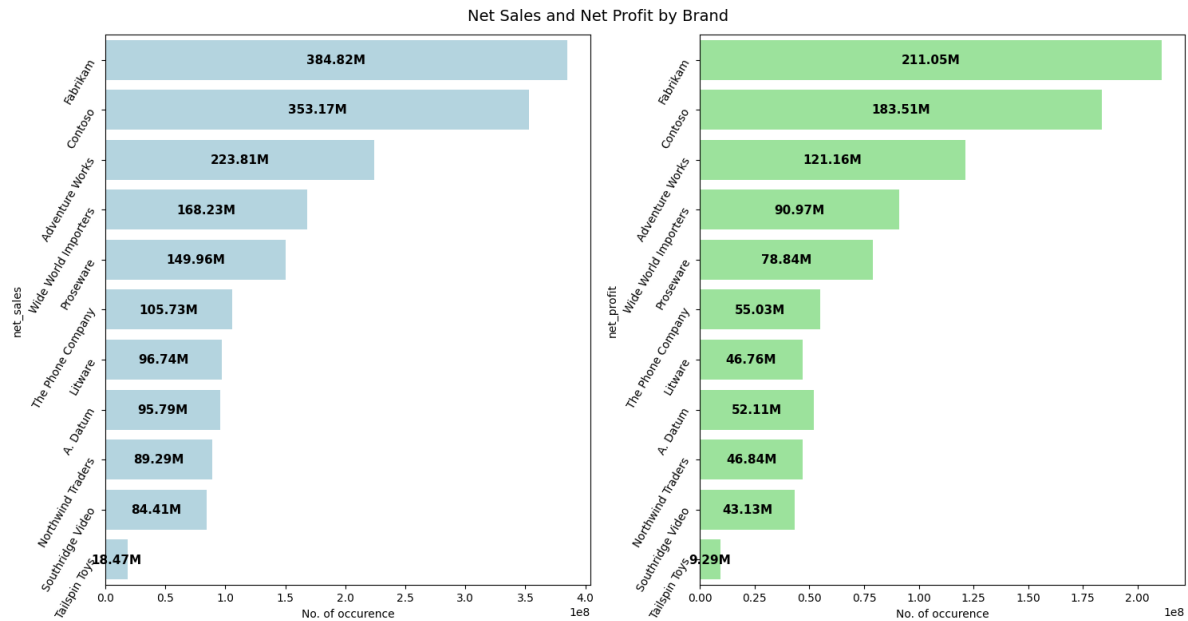


Correlation Insights :

- Net Sales has weak correlation with Return Amount (0.30) and Discount Amount (0.35), which indicates that there is no significant impact of discounts and returns on sale. May be impact is seasonal, during holidays.
- There is very strong correlation between Total Cost and Total Sales (0.98), Indicating that profit margin will stay about the same.
- There is very strong positive correlation between Net Sales and Net Profit which means if sales is increasing, net profit also tends to increase proportionally. Indicating company effectively converting sales into profit.

6.3.2 Net Sales & Net Profit by Brand.

Q: Which brands contribute most to net sales and net profit?



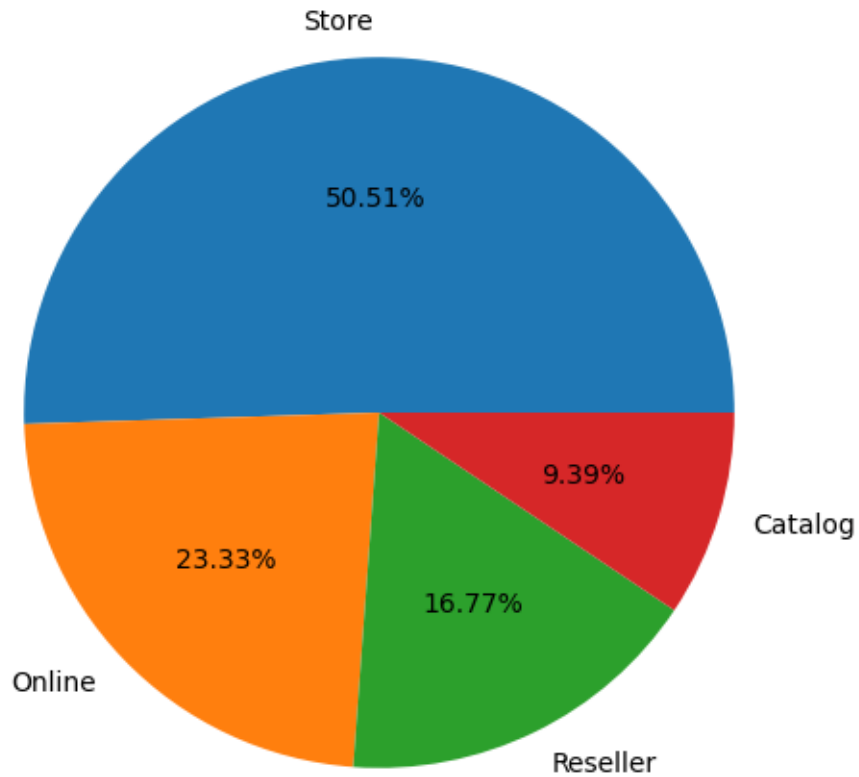
Insights :

- Net Sales by Brand (Left Plot):
 - Identifies the brand Fabrikam (384.8M) as the primary driver of Sales, followed by Contoso (353.2M).
 - Fabrikam is more prominent among customers, It can be due to better quality, reputation or marketing.
 - Fabrikam have better pricing strategy, product innovation, distribution or brand loyalty.
- Net Profit by Brand (Right Plot):
 - Fabrikam is the most profitable brand with net profit around 211M million, leading the pack by a significant margin, followed by Contoso with 183.5M net profit.
 - The lowest profit contributors such as Southridge Video and Tailspin Toys are below 10 million net profit, suggesting limited profitability or potential areas for strategic review.

6.3.3 Net Profit by Sales Channel

Q: Which sales channels are the most and least profitable?

Net Profit share by Channel

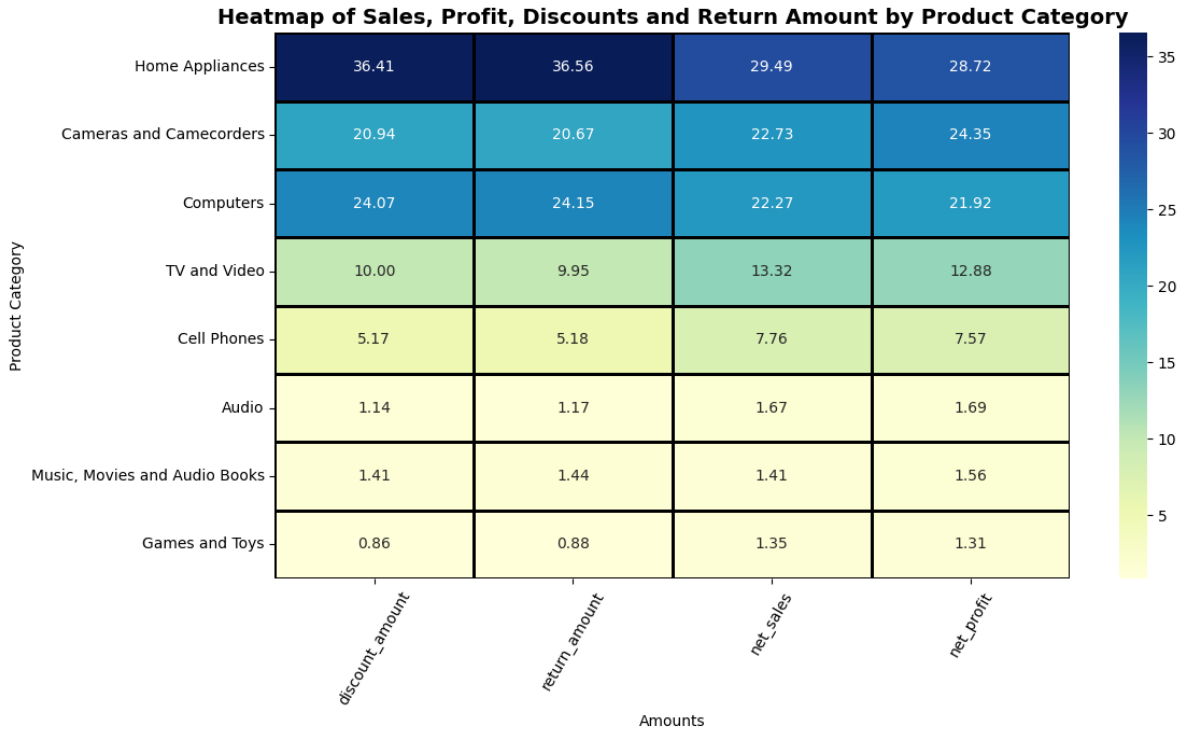


Insights :

- Store channel is generating most of the profit which is above 50% of Net Profit.
- Lowest profit is generated through catalog channel (9.39%).
- Likely due to higher sales volume despite having the smallest profit margin.

6.3.4 Discount, Returns, Net Sales and Net Profit by Product Category

Q: What is the percentage contribution of discount amount, return amount, net sales, and net profit by each product category to the total figures?



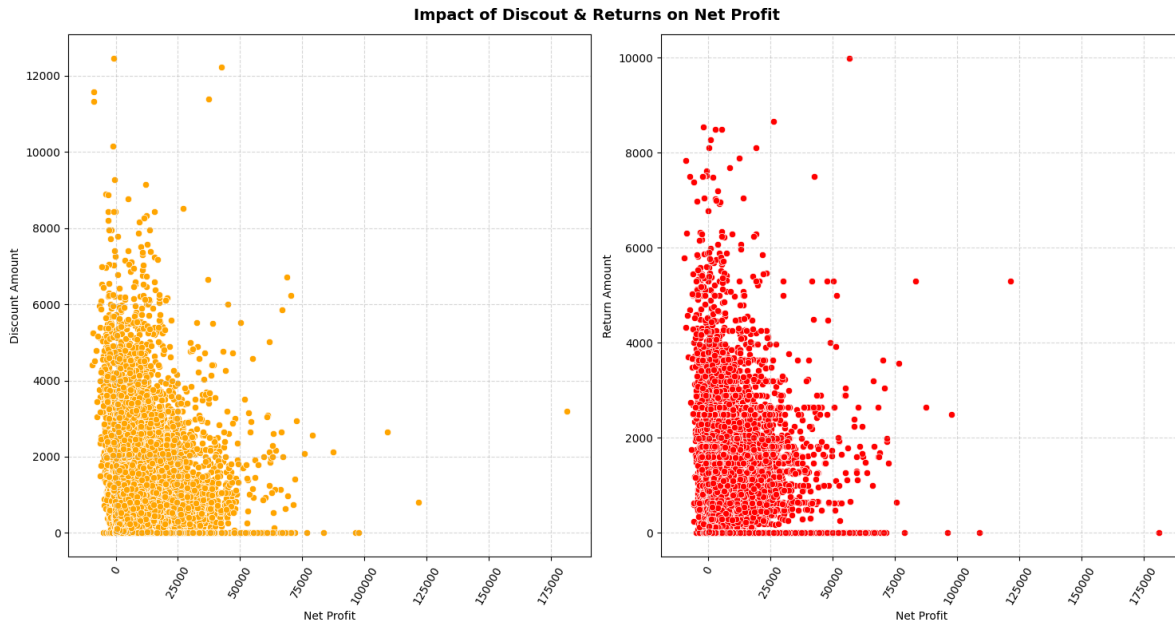
Insights :

- Discount Amount :
 - Home Appliances has highest discount among all product category (36.41%), followed by Computers (24.07%).
 - Discount is very high in Home Appliances category due to high sales price per unit.
- Return Amount :
 - Home Appliances has highest returns among all product category (36.56%), followed by Computers (24.15%).
 - It may be due to product quality issue like defects, malfunctioned or unmet customer expectations.
 - Some categories are more susceptible to return fraud or policy abuse.
- Net Sales :
 - Home Appliances has highest sales among all product category (29.49%), followed by Cameras and Camcorders (22.73%).
 - This is due to popularity among customers.
 - May warrant focused marketing efforts or investment.
- Net Profit :

- Home Appliances product category is generating the highest profit (28.72%), followed by Cameras and Camcorders (24.35%).
- Likely due to higher sales volume because when compare to profit margin there is no significant difference.

6.3.5 Impact of Discount & Returns on Net Profit

Q: How discount amount and return amount impact profitability?



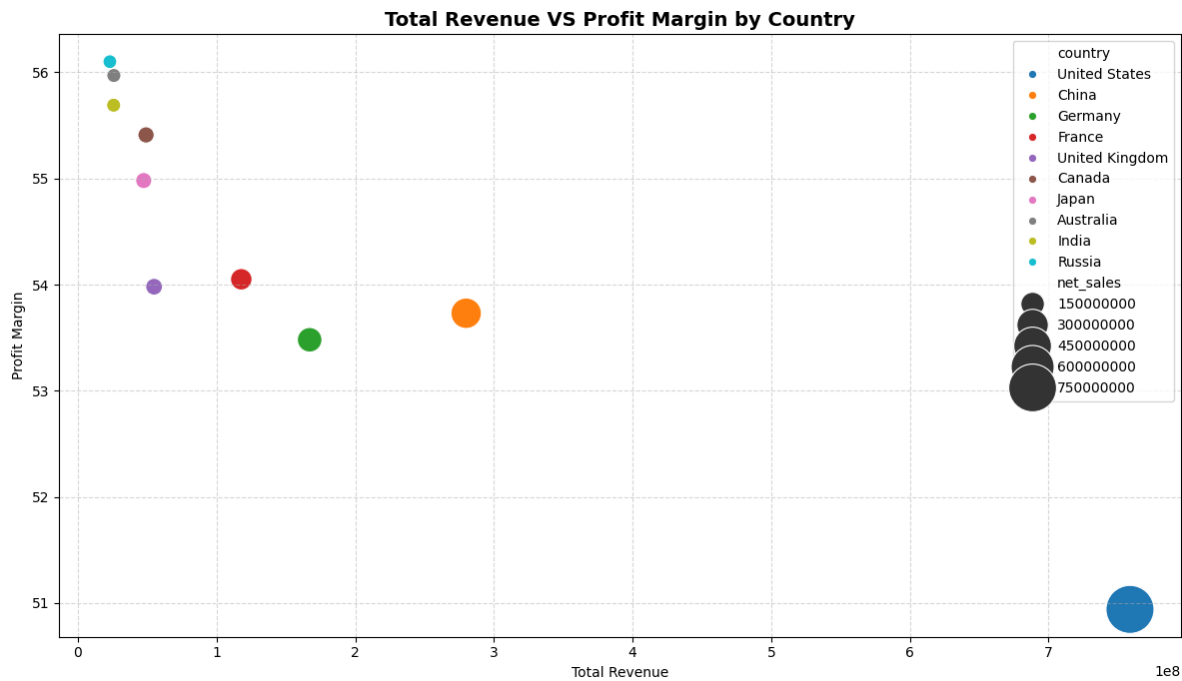
Insights :

- Net Profit vs. Discount Amount (Left Plot):
 - Higher discount amounts generally correlate with lower or negative net profits, indicating discounts can erode profitability.
 - Most transactions cluster around low discount amounts and moderate net profits, suggesting careful discount use supports profitability.
 - Large net profits rarely coincide with high discount amounts, implying that top profits come from low or no discount sales.
- Net Profit vs. Return Amount (Right Plot):
 - Increasing return amounts are associated with reduced net profits, showing returns negatively impact profitability.
 - Most data points show low returns and low to moderate net profit, highlighting that returns are typically small but impactful.

- High net profits are uncommon alongside large returns, underlining the detrimental effect of returns on overall profit.

6.3.6 Regional Sales and Profitability differences.

Q: How do regional variations influence total sales and net profit performance, and what factors contribute to these differences across geographic markets?



Insights :

- **High Sales \neq High Profitability :** United States has highest net sales, but lowest profit margin. This could be due to intense competition, higher cost and aggressive discounting.
- **Cost Structures & Business Matter :** High operational costs, supply chain challenges, pricing strategies, or local regulation can reduce the profit margins even where sales are strong.
- **Consistent Business Model :** This small range of profit margin indicate that the company has fairly consistent pricing and cost model accross these countries.

6.4 Time-Series Analysis

- Quarterly and yearly trends in sales, profit, returns, discounts.
- Seasonal patterns.

(Add charts + 1–2 lines of insight for each.)

7 Key Insights

Summarize your most important findings, e.g.:

- Electronics category has highest sales but lowest profit margin due to high returns.
 - North America contributes 40% of sales but only 20% of profit.
 - Profit margins are shrinking year-over-year despite higher sales.
-

8 Recommendations

Based on the analysis, the company should:

1. Reduce discounting on low-margin categories.
 2. Investigate reasons for high return rates in Electronics.
 3. Focus on expanding in regions with high margins (e.g., Asia-Pacific).
 4. Optimize inventory for seasonal demand.
-

9 Conclusion

- Restate the business problem and objectives.
 - Highlight the value of insights generated.
 - Suggest next steps (e.g., predictive modeling for sales forecasting, customer segmentation for targeted marketing).
-

10 Appendix (Optional)

- Additional plots.
- SQL queries, Python code snippets.
- Link to full GitHub repository.