

## Introduction

Exploratory Data Analysis (EDA) is a crucial step in understanding data structures and trends. This project conducts an EDA on supermarket retail sales data, analyzing key metrics, possible market trends, and profitability patterns. The goal is to generate a technical report providing stakeholders with actionable insights for strategic decision-making and business optimization.

## Overview of the Dataset

The dataset contains 9,994 rows and 13 columns. It includes categorical and numerical data. Key variables include Sales, Quantity, Discount, and Profit, alongside Product Category, Sub-Category, Region, and Segment.

## Initial Insight of the Dataset at First Glance

A key observation from the **Profit** column reveals the presence of negative values, which are closely correlated with applied discounts. Specifically, discount rates of 30% or higher consistently result in negative profits. However, some instances with a 20% discount also exhibit losses, to a lesser extent. Notably, the **Furniture** and **Office Supplies** categories contain a substantial number of transactions yielding no profit. This trend suggests these categories may be subject to higher discount rates, potentially impacting overall profitability.

## Data Cleaning Process and Preparation

1. Checked for null values, but none were found.
2. Checked for duplicates, 17 were identified and removed, leaving us with 9,977 rows.
3. Checked for the data types to see if they correspond.
4. The tool used for this analysis is Microsoft Excel.

## Descriptive Statistics

Some analysis of key financial metrics like Profit, Discount, and Quantity using Excel's Descriptive Statistics tool.

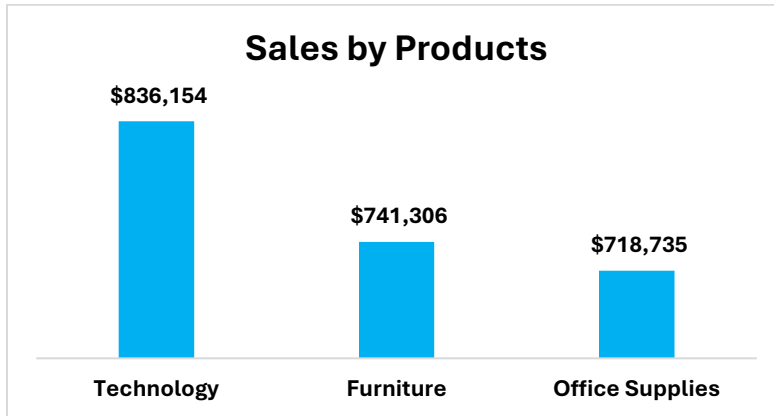
	Profit Analysis	Quantity Analysis	Discount Analysis
<b>Mean Profit</b>	\$28.69	Mean Quantity 3.79	Mean discount 0.16
<b>Standard Deviation</b>	\$234.46	2.23	0.21
<b>Median</b>	\$8.67	8.67	0.2
<b>Skewness</b>	7.55	1.28	1.68
<b>Range</b>	\$14,999.95 Minimum Profit is - \$6,599.98 (a substantial loss) Maximum Profit is \$8,399.98	Minimum: 1 item Maximum: 14 items. Range: 13 signifies customers purchase in relatively small quantities.	Minimum discount is 0% (no discount) Maximum discount is 0.8.
<b>Kurtosis</b>	396.52	1.98	2.41
<b>Mode</b>	0	3	0

This data reveals a complex retail environment with significant profit variability, ranging from substantial losses (-\$6,600) to high gains (\$8,400). While customers typically purchase small quantities (average 3.79 items), discounts vary from 0% to 80%. The highly skewed profit distribution (7.55) and high kurtosis (396.52) indicate extreme variations in transaction profitability.

## Key Insights Derived and Analysis

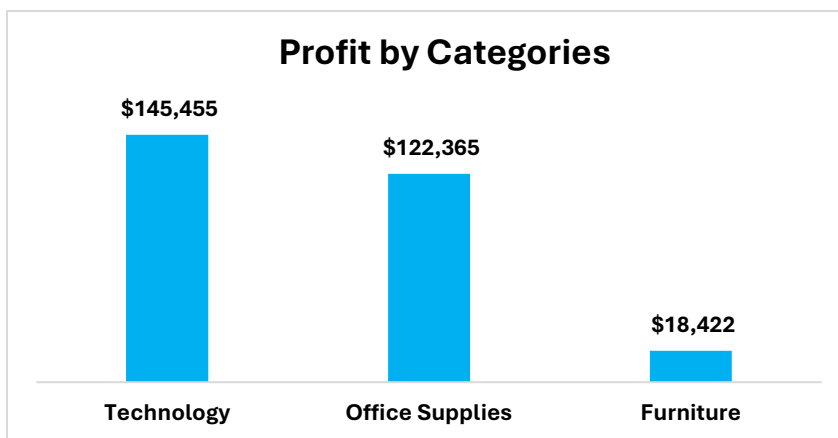
### 1. Top Selling Products by Sales

Technology produces the highest sales among the product categories.

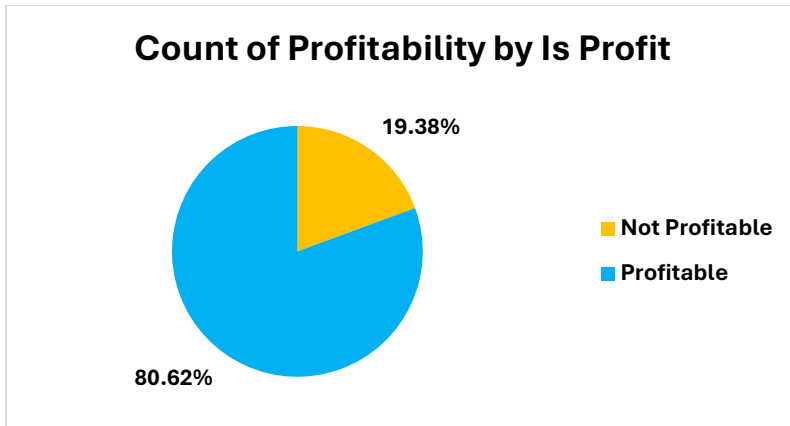


### 2. Most Profitable Categories

Technology leads in profitability, followed by Office Supplies, Furniture records the lowest profit. This could indicate potential issues such as high cost or heavy discount.



A new column was inserted and **IF** function was used to classify transactions as profitable or not, showing 81% of sales are profitable and 19% are not, highlighting the need to address unprofitable sales through a pie chart.



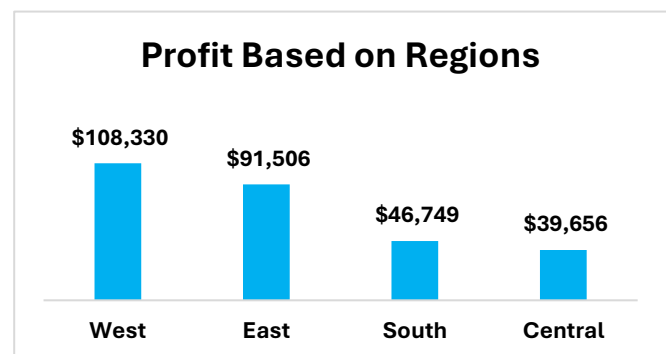
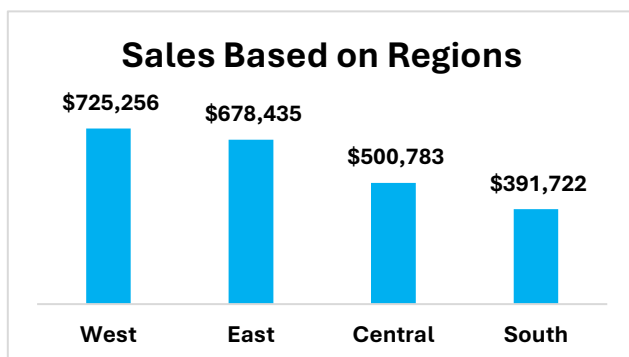
### Impact of Discount on Profit

Excessive discounts, exceeding 30%, often lead to negative profits, while 20% discounts can also result in losses, indicating a significant erosion of profit margins.

### 3. Sales and Profit by Region

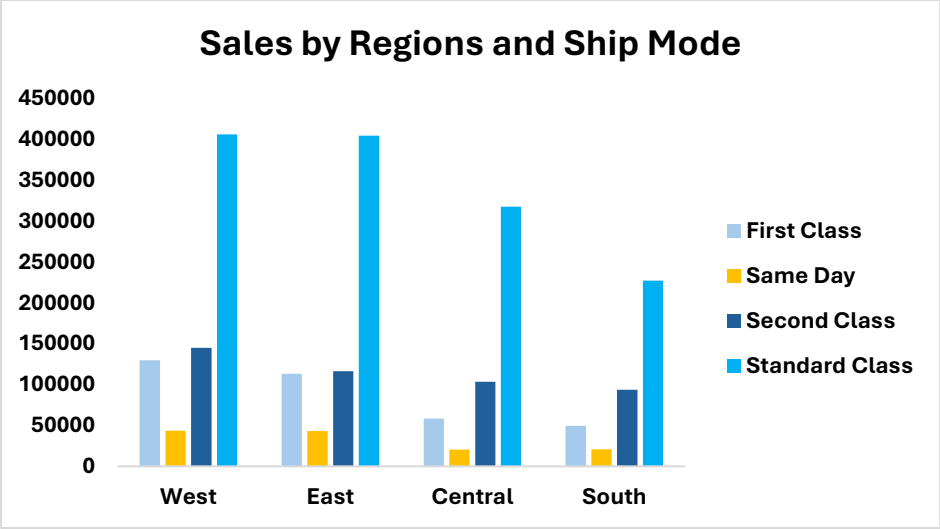
The **West** and **East** regions led in profitability, which also means they had the highest sales.

The **South** records the lowest sales, while Central recorded the lowest Profit



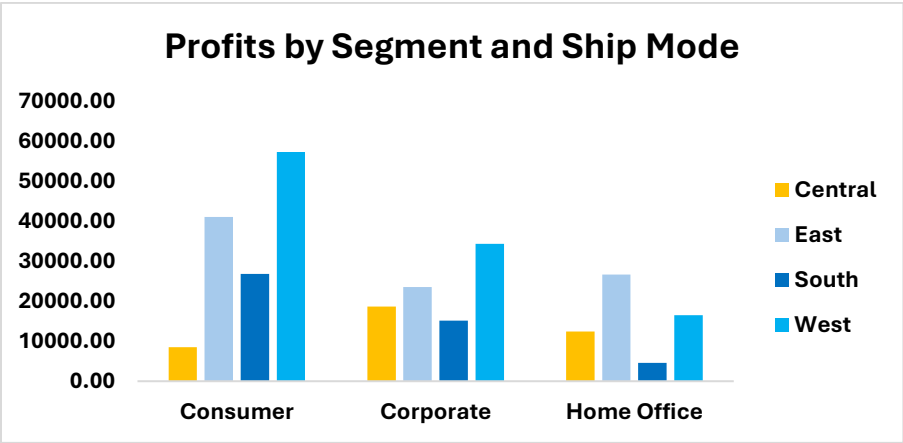
### 4. Preference for Shipping Mode by Regions

The Standard Class seems to be the preferred shipping mode across all the regions.



**5. Segment Performance Across Regions**

The West region appears to be the most profitable overall, particularly in the Consumer and Corporate segments. The East region follows closely, especially in the Home Office segment. Meanwhile, the Central and South regions exhibit comparatively lower profit margins across all segments.



## Recommendations

1. **Discount Strategy Optimization:** Implement a 25% discount ceiling across all categories, develop a dynamic pricing model for Furniture and Office Supplies, and create a data-driven discount framework.
2. **Regional Market Development:** The strategy involves focusing on growth initiatives in the South region, utilizing successful West and East practices, designing region-specific marketing campaigns, and improving operational efficiency in the Central region.
3. **Segment Based Marketing:** Create targeted marketing campaigns for South and West regions, create specialized B2B programs for Central, optimize product mix based on segment performance data, and implement loyalty programs for customer retention.
4. **Product Category Enhancement:** Expand Technology's product range, review Furniture's pricing and cost structure, and analyze successful Office Supplies transactions to maximize performance.

## Conclusion

The analysis suggests profit optimization through strategic discount management and regional market development, reducing unprofitable sales by 19%. Combining regional specialization, segment-specific marketing, and optimized discount strategies could drive growth in underperforming areas.