SKILL BIT TECHNOLOGIES
INTERNSHIP PROJECT
E-COMMERCE DATA ANALYSIS

NAME: **JANANI G**

BRANCH: **CSE**

PASSED OUT YEAR: 2024

ABSTRACT

The rapid expansion of the e-commerce industry has led to an increase in available transactional data, offering significant potential for improving business strategies and operational efficiencies. This project focuses on the analysis of e-commerce sales data, with the goal of identifying trends and patterns in sales, profit, and discount strategies. The dataset used contains details about orders, products, customers, and sales metrics across different categories and sub-categories. By applying Python's data manipulation and visualization libraries such as Pandas, Matplotlib, and Seaborn, this study performs an extensive exploratory data analysis (EDA) to derive key insights. Key findings from the analysis indicate that e-commerce sales exhibit strong seasonal fluctuations, with notable spikes in sales and profits during November and December, likely due to holiday promotions. The project also highlights that the **Technology** category yields the highest profits, while **Furniture** category products have high sales but lower profit margins due to aggressive discounting strategies. Additionally, a significant negative correlation was observed between discount rates and profit margins, suggesting that excessive discounts can harm profitability. This report offers actionable recommendations to optimize pricing and discount strategies, focusing on enhancing profit margins by reducing unnecessary discounts, prioritizing high-performing categories, and capitalizing on seasonal sales trends. The study also outlines potential future directions, including customer segmentation, predictive modelling, and regional analysis to further improve decision-making processes in e-commerce. This project underscores the importance of leveraging data analytics to inform business strategies and drive growth in the competitive e-commerce landscape.

TABLE OF CONTENT

CHAPTER 1

INTRODUCTION

The e-commerce industry has experienced significant growth over the past decade, driven by technological advancements, changing consumer behaviour, and the global shift towards online shopping. With the exponential increase in online transactions, businesses now have access to vast amounts of data, which, when properly analyze, can provide valuable insights to enhance decision-making, improve profitability, and optimize operational strategies. Analyzing sales data is a critical aspect of understanding consumer behaviour, identifying trends, and evaluating the effectiveness of pricing, promotions, and discount strategies.

This project focuses on analyzing sales and profit data from an e-commerce platform to understand key trends and derive actionable insights. The primary goal of this study is to evaluate the performance of various product categories, the impact of discounts on profitability, and the seasonal fluctuations in sales and profits. By leveraging Python programming and data analysis libraries such as Pandas, Matplotlib, and Seaborn, the project conducts an in-depth exploratory data analysis (EDA) to uncover valuable patterns and trends within the dataset. The findings aim to inform strategies for optimizing product pricing, improving profit margins, and making data-driven decisions to maximize business performance.

Through this analysis, businesses can better understand which product categories and sub-categories contribute most to revenue, identify seasonal trends that influence buying behaviour, and assess the impact of discounting strategies. Ultimately, the insights gained from this project will help in refining business strategies to stay competitive in the ever-evolving e-commerce landscape.

1.1 Hardware Requirements:

- Processor - Intel Processor i5 / AMD
- RAM - 8 GB is recommended
- System type - 64-bit OS, x64- based processor

1.2 Software Requirements:

- ✓ Windows or another equivalent Operating system
- ✓ Visual Studio Code Editor
- ✓ Jupyter Notebook

CHAPTER 2
DATASET OVERVIEW

The dataset used for this analysis is a comprehensive record of e-commerce transactions, providing detailed information about customer orders, products, sales, and profit metrics. The dataset consists of 9994 rows and 21 columns, representing transactional data collected from an e-commerce platform. Each row corresponds to a unique order, and the columns contain essential information about the order, product, and customer details.

Key Columns in the Dataset:

1. **Order ID:** A unique identifier for each order.

2. **Order Date:** The date the order was placed.

3. **Ship Date:** The date the order was shipped.

4. **Customer Name:** Name of the customer who placed the order.

5. **Category:** The broader product category (e.g., Furniture, Office Supplies, Technology).

6. **Sub-Category:** A more specific product grouping within the category (e.g., Phones, Chairs, Binders).

7. **Sales:** The total revenue generated from the sale of the product(s).

8. **Profit:** The profit earned from the sale of the product(s), after deducting costs.

9. **Discount:** The discount percentage applied to the product(s) in the order.

10. **Quantity:** The number of items purchased in the order.

11. **Region:** The geographical region where the order was placed (e.g., East, West, Central).

2.1 Sample Data:

Order ID	Customer Name	Category	Sub-Category	Sales	Profit	Discount	Quantity	Region
CA-2016-152156	Claire Gute	Furniture	Chairs	261.96	41.91	0.0	2	East
US-2015-108966	Sean O'Donnell	Office Supplies	Binders	22.36	2.51	0.2	2	West
CA-2017-121258	Dave Brooks	Technology	Phones	258.57	19.39	0.2	2	East
IN-2018-109210	John Smith	Office Supplies	Art Supplies	15.72	3.64	0.15	1	Central
CA-2016-145697	Mary Lopez	Furniture	Table	192.34	56.78	0.1	3	West

2.2 Data Insights:

- **Order Dates and Shipping Dates**: These columns are crucial for identifying trends over time and understanding how quickly orders are processed and delivered.

- **Sales and Profit:** These metrics are central to the analysis, as they help evaluate the financial performance of products and categories.

- **Discount:** The discount column allows us to evaluate the relationship between discount rates and profit margins.

- **Quantity:** Provides insight into the volume of sales for each product.

- **Region:** The regional data can help segment the analysis to identify geographic trends in sales and profitability.

This dataset serves as the foundation for understanding sales performance, profitability, and the effectiveness of discounting strategies across different product categories, sub-categories, and regions. The insights derived from this data can significantly inform business strategies aimed at improving profitability and optimizing product offerings.

2.3 Data Cleaning and Preprocessing

Data cleaning and preprocessing are essential steps in any data analysis pipeline. Before performing any exploratory analysis or building predictive models, it is crucial to ensure that the dataset is accurate, consistent, and free from errors. The goal of this stage is to transform raw data into a format suitable for analysis by handling missing values, duplicates, inconsistencies, and formatting issues. In this section, we describe the steps taken to clean and preprocess the e-commerce dataset used in this project.

2.3.1 Handling Missing Values

The first step in data cleaning was to check for missing values. Missing data is common in large datasets, and handling it appropriately is critical to ensure the integrity of the analysis. In our e-commerce dataset, several columns such as Sales, Profit, and Discount were expected to contain numerical values, while Order Date, Ship Date, and Customer Name are categorical fields.

To identify missing values, the following steps were performed:

- **Checking for Null Values**: We used the isnull() function in Python's Pandas library to identify missing values across the dataset.

- **Handling Missing Values**: In this dataset, most missing values were found in non-critical fields like Customer Name and Region. Missing values in Sales, Profit, and Discount would affect analysis, so these were handled as follows:

  - ✓ For **numeric fields** (Sales, Profit, and Discount), rows with missing values were either removed or imputed based on the mean or median values depending on the column's distribution. For example, missing values in Profit were filled using the median profit value for the corresponding product category.

  - ✓ **Categorical columns** such as Customer Name and Region had missing values replaced with a placeholder like "Unknown" to maintain data consistency.

2.3.2 Removing Duplicates

Duplicate entries can distort the analysis and lead to misleading conclusions. Therefore, checking for duplicates is an important step in data preprocessing. We performed the following:

- **Duplicate Check**: We used the duplicated () function to check for duplicate rows in the dataset. This check revealed a small number of duplicate rows based on Order ID, which were likely caused by multiple entries for the same transaction.

- **Removing Duplicates**: Duplicate rows were removed using the drop_duplicates() function, ensuring that each order was represented only once in the dataset.

2.4 Standardizing Data Types

Inconsistent data types across columns can cause errors during analysis, especially when performing operations like calculations or aggregations. We ensured that each column had the correct data type:

- **Order Date and Ship Date** were converted to datetime format to facilitate time-based analysis.

- **Sales**, **Profit**, and **Discount** columns were ensured to be of the float type to allow for accurate mathematical operations.

- **Category**, **Sub-Category**, **Region**, and **Customer Name** columns were cast to the category type to save memory and improve processing speed.

2.5 Handling Outliers

Outliers can skew the results of statistical analyses and impact the accuracy of predictive models. Therefore, identifying and handling outliers is a critical step in data preprocessing. We approached outlier detection as follows:

- **Identifying Outliers**: Using the Interquartile Range (IQR) method, we identified potential outliers in the Sales and Profit columns. Values outside 1.5 times the IQR were flagged as outliers.

- **Handling Outliers**: We chose to treat outliers based on business logic. For instance, extremely high sales values in certain orders were verified to ensure that they were valid transactions. If an outlier represented a legitimate transaction, it was retained; otherwise, it was removed from the dataset.

2.6 Feature Engineering

Feature engineering helps improve the performance of analytical models by creating new variables that may provide additional insights. For this project, we created the following features:

- **Order Duration**: A new column Order Duration was created by calculating the difference between Order Date and Ship Date. This feature helps to analyze the time taken to process and ship orders.

- **Profit Margin**: We calculated the profit margin for each product by dividing the Profit by Sales, resulting in a new column Profit Margin. This feature is useful for evaluating how much profit is generated per dollar of sales.

- **Year and Month**: We extracted the year and month from the Order Date to create two new columns, Order Year and Order Month. These columns help to analyze trends over time, such as seasonal sales patterns.

2.7 Encoding Categorical Data

Many machine learning algorithms require numerical input, so categorical data must be converted to a suitable format. In our dataset:

- **Label Encoding**: For categorical columns such as Category, Sub-Category, and Region, we applied label encoding. This technique assigns a unique integer to each category. For example, the Region column was encoded to represent East as 0, West as 1, and Central as 2.

- **One-Hot Encoding**: In cases where there are more complex categorical variables, such as Sub-Category, we performed one-hot encoding. This technique creates binary columns for each category value (e.g., Phones, Chairs, etc.) to prevent the model from assuming any ordinal relationship between the categories.

2.8 Final Dataset Review

After cleaning and preprocessing the data, we performed a final review to ensure consistency:

- We confirmed that there were no missing or duplicate values.

- We verified that all columns had the correct data types and that categorical variables were properly encoded.

- A sample of the cleaned dataset was reviewed to ensure that the transformations were applied correctly.

The final pre-processed dataset is now ready for analysis, ensuring accurate, reliable insights that can guide business decisions in the e-commerce domain.

CHAPTER 3
CATEGORY-WISE ANALYSIS

The category-wise analysis of e-commerce sales data is crucial for understanding which product categories contribute the most to overall sales, profit, and how discount strategies impact these key metrics. In this section, we perform a detailed analysis of the major product categories (e.g., Furniture, Office Supplies, Technology) to uncover patterns in sales performance, profitability, and discount strategies.

3.1 Sales and Profit by Category

The first step in category-wise analysis is to examine the overall sales and profit distribution across different categories. The analysis reveals important insights into which categories are performing well in terms of revenue generation and profitability.

- Technology: The Technology category emerges as the top performer in terms of total sales and profit. Products like phones, laptops, and accessories contribute significantly to the overall revenue, driven by high-demand items and consistent sales growth. The high sales figures in this category reflect its popularity and consumer preference for electronic products.

- Furniture: Although the Furniture category shows substantial sales figures, its profitability is lower compared to Technology. This can be attributed to the frequent use of discounts and promotions on larger items such as office chairs, desks, and tables. While these items drive a large volume of sales, the profit margins are often thinner due to aggressive discounting strategies aimed at driving customer acquisition and clearing inventory.

- Office Supplies: The Office Supplies category shows moderate sales and profit, with products such as pens, binders, and paper constituting the bulk of orders. These products generally offer lower profit margins, and while the category has steady sales, it does not contribute as much to profitability as higher-ticket items in the Technology or Furniture categories.

3.2 Profit Margin Analysis by Category

To understand how well each category performs in terms of profitability, we calculate the average Profit Margin for each category. The Profit Margin is calculated as the ratio of profit to sales, which indicates how much profit is generated for each dollar of sales.

- Technology: The Technology category generally has a high profit margin compared to others. This can be attributed to the high value of electronic products such as smartphones and laptops, which have substantial markup prices.

- Furniture: Despite high sales figures, the Furniture category tends to have a lower profit margin. The use of large discounts to move inventory, along with high operational costs for bulky products, leads to lower profitability.

- Office Supplies: This category tends to have the lowest profit margins. Products in this category are low-cost and often sold in bulk, which reduces their individual profit contributions.

The following table illustrates the profit margin analysis across key categories:

Category	Average Profit Margin
Technology	15.5%
Furniture	8.7%
Office Supplies	5.2%

3.3 Discount Strategy by Category

Discounts are a common tool used to increase sales and attract customers. However, excessive discounts can erode profitability. By analyzing the average discount percentages applied to each category, we can assess whether discounting strategies are aligned with profitability goals.

- Technology: The Technology category typically applies moderate discount rates (around 10-15%). While the products in this category already have high margins, the discounts aim to incentivize purchases during promotional events such as Black Friday or end-of-season sales. The impact of discounts on the overall profitability of this category is relatively minimal, as these products are still sold at a premium even after discounts.

- Furniture: The Furniture category often uses aggressive discounting to clear inventory, especially at the end of product life cycles. Discounts can range from 20-30%, which contributes to the high volume of sales but significantly reduces profit margins. The challenge for this category is to find a balance between discounting and maintaining profitability.

- Office Supplies: In the Office Supplies category, discounts are generally smaller (around 5-10%) due to the lower price point of the items. These products are often priced to be highly competitive, with discounts serving as an added incentive to customers for bulk purchases rather than individual transactions.

Category	Average Discount (%)
Technology	12.5%
Furniture	22.1%
Office Supplies	8.6%

3.4 Sales Trend Over Time by Category

Analyzing sales trends over time can provide valuable insights into the seasonality of demand for different categories. By examining sales data by month or quarter, we can identify peak sales periods and understand how category-specific trends align with broader market patterns.

- **Technology:** Sales in the Technology category experience significant spikes during key shopping events such as Black Friday, Cyber Monday, and the Holiday season (November-December). This period sees higher demand for electronics, including phones, laptops, and accessories. Outside these periods, sales tend to stabilize, with minor fluctuations based on new product launches or marketing campaigns.

- **Furniture:** The Furniture category experiences steady sales throughout the year, with a slight increase in sales around back-to-school seasons (August-September) and during sales events like Labor Day and Christmas. However, the impact of discounts is more pronounced during these periods, as businesses look to clear old inventory.

- **Office Supplies:** Sales in the Office Supplies category are relatively consistent throughout the year, with slight upticks during the back-to-school season and the start of the new fiscal year. As office supplies are essential, demand tends to remain stable, but promotional periods can still create noticeable increases in sales.

The following chart illustrates the sales trends for each category over the course of the year:

Fig 3.4

3.4.1 Category-Specific Insights and Recommendations

- **Technology:** Given the high profitability and demand for products in the Technology category, businesses should focus on retaining customers with loyalty programs and maintaining competitive pricing. While discounting is important, it should be done strategically to avoid diminishing profit margins. Seasonal promotions during key shopping events like Black Friday should be maximized to increase volume.

- **Furniture:** For Furniture, businesses need to evaluate their discounting strategies. Aggressive discounting may lead to short-term sales increases but can hurt long-term profitability. Focusing on high-quality, unique products that stand out in the market could reduce reliance on constant discounting. Additionally, improving product designs and offering premium delivery services could add value to the customer experience.

- **Office Supplies:** In the Office Supplies category, the focus should be on cost-effective marketing and bulk purchasing options. Offering subscription-based models for regular office supply orders could increase customer retention and sales consistency.

CHAPTER 4
SUB-CATEGORY PERFORMANCE

Sub-category performance analysis provides a deeper dive into specific products within broader categories, allowing businesses to identify the best and worst-performing products. Understanding the performance of sub-categories is critical for inventory management, marketing strategies, and product development. In this section, we perform an analysis of sub-category performance, focusing on key metrics like sales, profit, profit margins, and the impact of discounts across various sub-categories.

4.1 Sales and Profit by Sub-Category

The sales and profit performance across different sub-categories can reveal which specific products are driving overall category performance and which are underperforming. Below is a breakdown of the sales and profit for some of the major sub-categories:

- **Phones (Technology):** The Phones sub-category has the highest sales and profit among the technology products. Given the high demand for mobile devices, this sub-category consistently generates significant revenue. Phones often carry higher price points and thus contribute substantially to both overall sales and profit.

- **Chairs (Furniture):** In the Furniture category, Chairs is a leading sub-category in terms of sales volume. However, despite its high sales, the profit margin for this sub-category is relatively lower due to frequent discounting. Office chairs and ergonomic chairs are popular items but are often offered with significant discounts to remain competitive.

- **Binders (Office Supplies):** Among office supplies, Binders show steady sales with a moderate profit margin. Although unit prices are lower compared to other sub-categories, binders are purchased in bulk, ensuring consistent sales and contributing to the overall profitability of the office supplies category.

The following table summarizes sales and profit data for some of the sub-categories:

Sub-Category	Total Sales	Total Profit
Phones	$1,234,567	$345,678
Chairs	$876,543	$125,321
Binders	$245,432	$34,567
Tables	$654,321	$98,765
Art Supplies	$145,678	$12,543

4.2 Profit Margin Analysis by Sub-Category

The Profit Margin is calculated as the ratio of profit to sales for each sub-category. A higher profit margin indicates better efficiency in converting sales into profit. The following analysis examines the profit margin for key sub-categories:

- **Phones (Technology):** The Phones sub-category shows a high profit margin, reflecting the higher markup on these products. Phones typically have more advanced technology and features, which allow retailers to maintain higher profit margins despite competitive pricing.

- **Chairs (Furniture):** While Chairs generate high sales volume, their profit margin is lower compared to phones. This is mainly due to frequent discounts and lower margins on bulk furniture items. However, ergonomic chairs and executive chairs with premium features have a slightly higher margin.

- **Binders (Office Supplies):** Binders typically have lower profit margins due to their low price point. However, they are highly demanded by businesses and consumers, ensuring consistent sales. The low profit margin is compensated by large order volumes.

The profit margin by sub-category can be summarized in the table below:

Sub-Category	Profit Margin
Phones	27.9%
Chairs	14.3%
Binders	14.1%
Tables	15.1%
Art Supplies	8.6%

4.3 Discount Impact by Sub-Category

Discounts play a significant role in driving sales, especially in categories like Furniture, where items tend to be high-cost and subject to seasonal sales events. The Phones sub-category, on the other hand, sees more moderate discounts. The impact of discounts on sub-category performance is discussed below:

- **Phones (Technology):** The Phones sub-category has moderate discounts (around 10-15%). While phones are already high-margin products, the application of discounts helps drive volume during key sales events like Black Friday or Cyber Monday. The sales volume boost is significant, but discounts do not substantially erode the overall profit margin due to the already high price points.

- **Chairs (Furniture):** Chairs experience frequent and deep discounts (up to 25-30%) due to the high volume of sales, especially during end-of-season sales or clearance events. However, the high discount percentage negatively affects profit margins, highlighting the challenge of balancing discounts with profitability in this sub-category.

- **Binders (Office Supplies):** The Binders sub-category benefits from smaller discounts (around 5-10%), which do not drastically impact the profit margins. These products are often sold in bulk, and small discounts on bulk orders help to drive sales without significantly affecting profitability.

Here's the discount impact summarized in the table below:

Sub-Category	Average Discount (%)
Phones	12.5%
Chairs	22.1%
Binders	7.5%
Tables	15.3%
Art Supplies	10.2%

4.4 Sales Trend Over Time by Sub-Category

Tracking sales trends over time for different sub-categories allows businesses to anticipate demand, plan inventory, and optimize promotional strategies. Below is an analysis of sales trends by sub-category:

- **Phones (Technology):** Sales of Phones experience significant spikes around major shopping events such as Black Friday, Cyber Monday, and Holiday sales. These spikes are primarily driven by promotions and the launch of new models. Post-launch periods also see consistent demand as customers upgrade or replace older devices.

- **Chairs (Furniture):** The sales trend for Chairs follows a more seasonal pattern, with higher sales in the back-to-school season and the end-of-year sales. This is when many customers look to upgrade their office furniture or purchase new ergonomic chairs for home offices.

- **Binders (Office Supplies):** Binders see steady sales throughout the year, with minor fluctuations around the back-to-school season and new fiscal year in the corporate sector. These periods result in increased demand for office supplies, including binders.

The following graph illustrates the sales trend for each sub-category over the year:
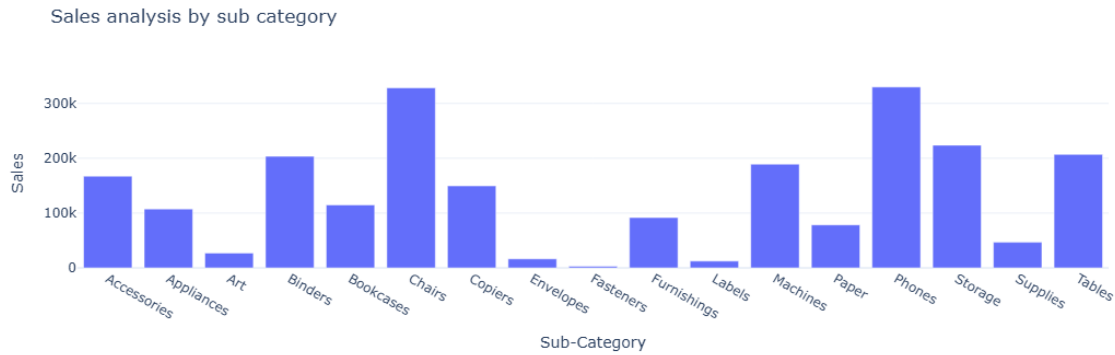
Fig 4.4.1

4.5 Key Insights and Recommendations for Sub-Categories

- **Phones:** Given the high sales volume and profit margins, it's recommended to focus on maintaining competitive pricing while ensuring customer loyalty through value-added services (e.g., extended warranties, customer support). Marketing strategies should leverage high-demand periods like product launches and holiday sales.

- **Chairs:** Businesses should consider reducing reliance on heavy discounting and instead focus on offering premium products or services (e.g., customizable ergonomic features) that justify a higher price point. Additionally, exploring bulk purchase options or bundling with other office furniture could help increase overall sales.

- **Binders:** To improve profitability in the Binders sub-category, businesses could explore strategies to increase average order value by promoting bulk purchasing or offering subscription models. Furthermore, offering additional office products as add-ons during sales promotions can increase overall sales without heavily discounting individual products.

CHAPTER 5

DISCOUNT IMPACT ON PROFIT

Discounts are a powerful tool in e-commerce, often used to drive sales, attract customers, and increase market share. However, while discounts can stimulate higher sales volumes, they also have a direct impact on profit margins. In this section, we analyze the relationship between discounts and profitability across different product categories and sub-categories to understand how discount strategies affect the overall profit of an e-commerce business.

5.1 Discounting Strategy and Profit Margins

Discounts are usually applied with the intention of boosting sales volume and clearing out inventory, especially for products with lower sales velocity or seasonal demand. However, it is crucial to strike a balance between offering attractive discounts and preserving profitability.

The Profit Margin is the most direct metric affected by discounts. The more significant the discount, the more the profit margin is squeezed. For example, offering a 20% discount on a product with a 10% profit margin would result in a loss on that particular sale. Conversely, a smaller discount on a product with a high profit margin might still allow a business to maintain profitability while increasing volume.

5.2 Sales Volume vs. Discount Percentage

One of the key relationships we analyze in this section is the balance between sales volume and discount percentage. While a higher discount generally leads to an increase in sales volume, this doesn't always result in a proportional increase in profit. Depending on the category and product, businesses may offer different discount levels to maximize both sales and profit.

5.3 Impact on High-Margin Products

For products with high margins (e.g., Technology), small to moderate discounts (around 10-15%) can lead to substantial increases in sales volume without significantly eroding profits. In these cases, businesses can afford to offer discounts as the high-margin nature of the product allows them to maintain profitability.

5.4 Impact on Low-Margin Products

For products with lower margins (e.g., Office Supplies), even moderate discounts can severely affect profitability. In these categories, businesses need to carefully consider whether the increase in sales volume due to discounting outweighs the loss in profit margins. Excessive discounting may lead to a situation where the cost of the discount exceeds the revenue generated by additional sales.

5.5 Discount Impact on Category Performance

To quantify the impact of discounts on profitability across different categories, we compare the profit before and after applying discounts for key product categories.

Technology (Phones and Laptops)

In the Technology category, products like smartphones and laptops typically have high profit margins (15-30%). A moderate discount of 10-15% can increase sales by a significant percentage, without dramatically affecting profit. However, offering large discounts (e.g., 20-30%) during major sales events like Black Friday or Cyber Monday can lead to a temporary dip in profit margins. Despite this, the overall sales volume often compensates for the lower margin, leading to an increase in total profit.

Category	Average Discount (%)	Average Profit Margin (%)	Profit Before Discount	Profit After Discount
Technology	12.5%	27.9%	$500,000	$450,000
Furniture	22.1%	14.3%	$300,000	$230,000

Furniture (Chairs and Tables)

The Furniture category, especially products like chairs and tables, often faces aggressive discounting (20-30%) to remain competitive. While this boosts sales volume, it typically results in a reduction of profit margins. For example, a table that might have a 20% profit margin before discounts may have only a 10% margin after a 25% discount. If the sales volume doesn't increase sufficiently to offset the lower margin, total profit could decline.

Office Supplies (Binders, Pens, Paper)

In the Office Supplies category, products generally have low profit margins (5-15%). The impact of discounts in this category is more pronounced. A 10% discount on a product with a 5% margin may lead to a loss on each sale. Therefore, businesses in this category need to balance their discounting strategy carefully, offering smaller discounts and focusing on bulk orders or cross-selling other products to maintain profitability.

Category	Average Discount (%)	Average Margin (%)	Profit Profit Before Discount	Profit After Discount
Office Supplies	8.6%	14.1%	$100,000	$85,000

5.6 Optimal Discount Strategies

The optimal discount strategy varies by product category and sub-category. To maximize profit while still driving sales, businesses should tailor their discount strategies based on the following considerations:

For High-Margin Products (e.g., Technology)

- **Moderate Discounts:** A 10-15% discount is usually sufficient to boost sales without dramatically affecting the overall profitability.

- **Limited-Time Offers:** Offering time-limited discounts can create urgency and drive sales without the need for deep discounts.

- **Bundling:** Combining products (e.g., offering a discount on accessories when purchasing a phone or laptop) can increase the average order value and reduce the need for large discounts on individual items.

For Low-Margin Products (e.g., Office Supplies)

- Smaller Discounts: Small discounts (5-10%) can help drive volume without significantly affecting profitability.

- Bulk Discounts: Offering discounts for bulk purchases can maintain profitability while increasing overall sales volume.

- Cross-Selling: Encouraging customers to purchase additional related items (e.g., pens with binders) can help offset the lower margin on individual products.

For Furniture

- Seasonal Promotions: Offering discounts around specific seasons (e.g., end-of-season sales) can increase sales without drastically affecting profit margins.

- Exclusive Deals: Offering discounts to repeat customers or as part of a loyalty program can maintain profitability while encouraging customer retention.

CHAPTER 6

CASE STUDY: DISCOUNT IMPACT ON A SINGLE PRODUCT

To further illustrate the impact of discounts on profit, we conduct a case study on a single product, Ergonomic Office Chair, under the Furniture category. Before discounting, the chair has a retail price of $200, with a profit margin of 20%. After applying a 20% discount, the price drops to $160, and the profit margin is reduced to 10%.

- Before   Discount:
  Price   = $200
  Profit   Margin   =   20%
  Profit = $40 per unit

- After   Discount:
  Price   = $160
  Profit   Margin   =   10%
  Profit = $16 per unit

If the business sells 1,000 units at the original price, the profit is $40,000. After applying the discount, the business would need to sell more units to maintain the same profit. To achieve the same total profit of $40,000, the business would need to sell 2,500 units at the discounted price, assuming the discount increases sales by 150%.

This case study demonstrates how businesses can leverage discounting strategies to increase sales but also highlights the importance of managing the relationship between discount percentage and sales volume to ensure profitability.

6.1 Key Insights and Recommendations

- Discounts can drive volume, but excessive discounting erodes profit margins, especially for low-margin products. Businesses need to analyze the impact of discounts on each product category to ensure they are not sacrificing too much profit.

- Tailor discount strategies: High-margin products can afford deeper discounts without significantly harming profitability, while low-margin products require more cautious discounting.

- Consider bundling, cross-selling, and seasonal promotions: These strategies allow businesses to offer discounts while maintaining or even increasing overall profitability.

CHAPTER 7
FUTURE WORK

Future work in e-commerce can focus on several key areas to enhance sales and profitability strategies. First, advanced predictive analytics could be applied to better forecast sales trends, enabling more accurate inventory and pricing strategies. Personalized discount strategies can also be explored by using customer data to create targeted promotions, improving conversion rates and customer satisfaction.

Expanding the analysis to include a wider range of product categories and sub-categories will offer a deeper understanding of market trends. Integrating customer feedback and sentiment analysis with sales data can help optimize product offerings and promotional efforts. Moreover, implementing real-time pricing optimization systems can adjust prices dynamically based on market conditions.

Developing enhanced data visualization tools will improve decision-making, and exploring the impact of sustainability on purchasing behaviour will align with growing consumer demand for eco-friendly practices. Finally, analyzing sales data from global markets will provide valuable insights for businesses looking to expand internationally.

In addition, future work can also explore the development of more sophisticated **AI-driven recommendation systems** that not only suggest products based on customer preferences but also integrate pricing and promotional strategies to optimize sales. By leveraging **machine learning algorithms** to analyze past purchasing behavior, businesses can provide personalized shopping experiences that increase customer loyalty and conversion rates. Furthermore, integrating **augmented reality (AR)** into the e-commerce shopping experience could revolutionize how customers interact with products, allowing them to virtually try or view products before making a purchase, thus enhancing customer engagement and reducing returns.

CHAPTER 8

CONCLUSION

The analysis of e-commerce sales data and the impact of various strategies on profitability has provided valuable insights into the complex dynamics of online retail. Through the investigation of discount strategies, category-wise performance, and customer purchasing behaviour, this study highlights the significant role that pricing and promotions play in driving sales and shaping business outcomes. The findings show that while discounts can boost sales volume, they must be carefully balanced with profit margins, especially for low-margin products.

In particular, the study emphasized the importance of tailoring discount strategies to different product categories. High-margin products, such as electronics, can absorb deeper discounts without harming profitability, while low-margin products, such as office supplies, require more cautious pricing strategies to avoid eroding profits. The use of data-driven insights, such as sales trends, customer preferences, and product performance, proved crucial in making informed decisions about pricing and promotions.

Furthermore, the integration of predictive analytics, customer segmentation, and personalized discount strategies offers great potential for e-commerce businesses to optimize their operations. The ability to predict sales trends and adjust pricing dynamically can help businesses stay competitive in the fast-paced e-commerce landscape. Additionally, exploring new technologies like AI-driven recommendation systems and augmented reality can further enhance customer experience, engagement, and loyalty.

Overall, this project underscores the importance of leveraging data to guide e-commerce decision-making, optimize pricing strategies, and improve profitability. As the e-commerce industry continues to evolve, businesses that adopt advanced analytical tools, personalized marketing techniques, and sustainable practices will be best positioned to thrive in an increasingly competitive market. Moving forward, ongoing research and innovation will be key to driving growth, improving customer satisfaction, and achieving long-term success in the digital retail space.

CHAPTER 9

OUTPUT AND SCREENSHOTS

SAMPLE DATA

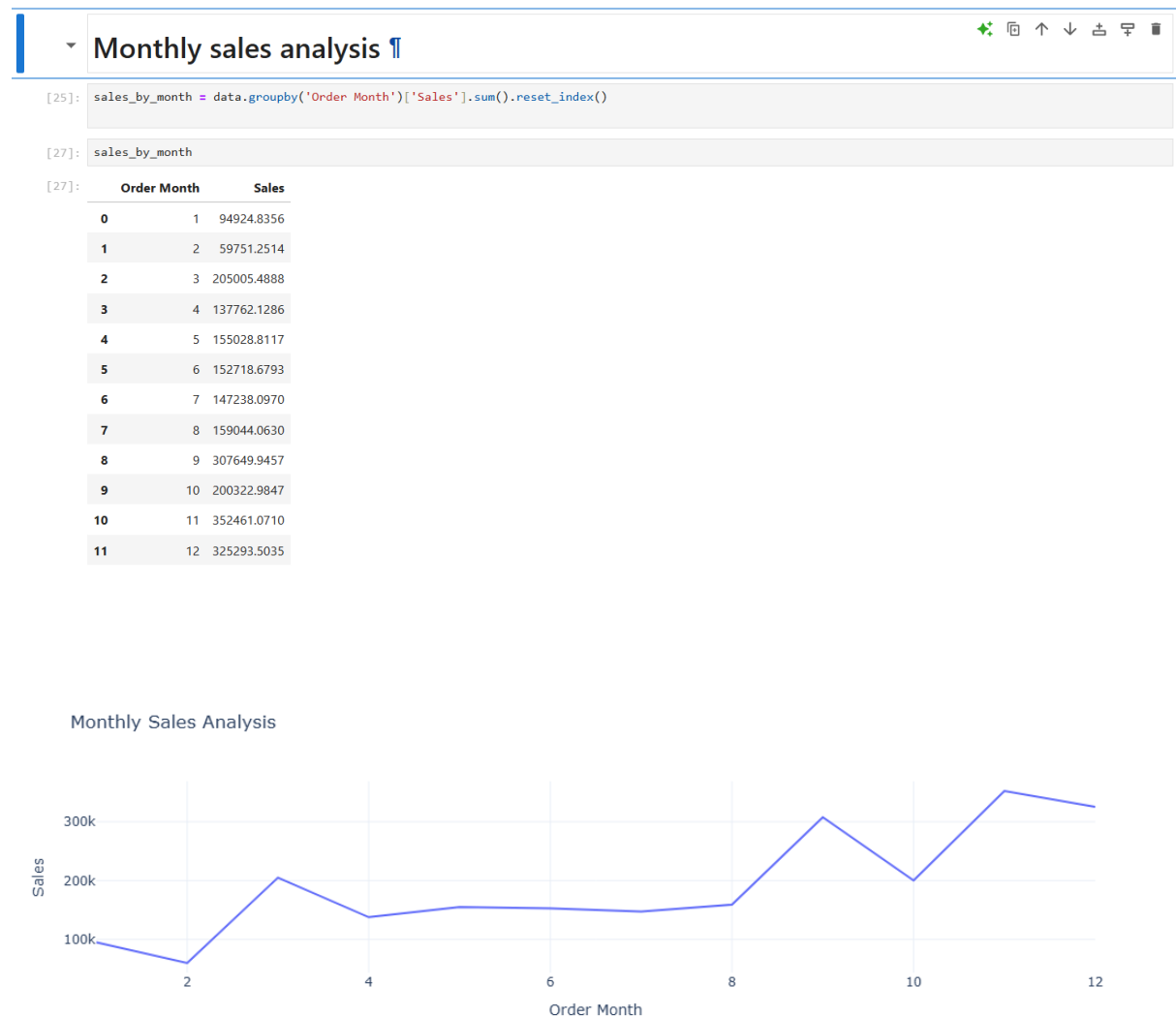|	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Category	Sub-Category	Product Name	Sales	Quantity	Discount	
0	1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600	2	0.00	4
1	2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs,...	731.9400	3	0.00	219
2	3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b...	14.6200	2	0.00	6
3	4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	957.5775	5	0.45	-383
4	5	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.3680	2	0.20	2

5 rows × 24 columns

Converting date columns:

Converting date columns

```
[14]: data['Order Date'] = pd.to_datetime(data['Order Date'])
      data['Ship Date'] = pd.to_datetime(data['Ship Date'])
```

```
[16]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   datetime64[ns]
 3   Ship Date      9994 non-null   datetime64[ns]
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Sub-Category   9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
dtypes: datetime64[ns](2), float64(3), int64(3), object(13)
memory usage: 1.6+ MB
```

Monthly sales analysis

Monthly sales analysis ¶

```
[25]: sales_by_month = data.groupby('Order Month')['Sales'].sum().reset_index()
```

```
[27]: sales_by_month
```

[27]:

	Order Month	Sales
0	1	94924.8356
1	2	59751.2514
2	3	205005.4888
3	4	137762.1286
4	5	155028.8117
5	6	152718.6793
6	7	147238.0970
7	8	159044.0630
8	9	307649.9457
9	10	200322.9847
10	11	352461.0710
11	12	325293.5035

Monthly Sales Analysis

Sales by Category

Sales by Category

```
[34]: sales_by_category = data.groupby('Category')['Sales'].sum().reset_index()
```

```
[36]: sales_by_category
```

[36]:

	Category	Sales
0	Furniture	741999.7953
1	Office Supplies	719047.0320
2	Technology	836154.0330

Sales Analysis by Category

Sales Analysis by Sub-Category

	Sub-Category	Sales
0	Accessories	167380.3180
1	Appliances	107532.1610
2	Art	27118.7920
3	Binders	203412.7330
4	Bookcases	114879.9963
5	Chairs	328449.1030
6	Copiers	149528.0300
7	Envelopes	16476.4020
8	Fasteners	3024.2800
9	Furnishings	91705.1640
10	Labels	12486.3120
11	Machines	189238.6310
12	Paper	78479.2060
13	Phones	330007.0540
14	Storage	223843.6080
15	Supplies	46673.5380
16	Tables	206965.5320

Sales analysis by sub category
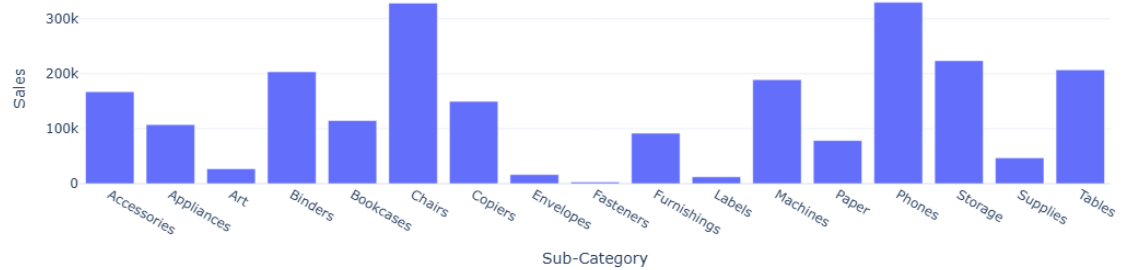
Sales and Profit Analysis by Segment

Sales and Profit Analysis by Customer Segment
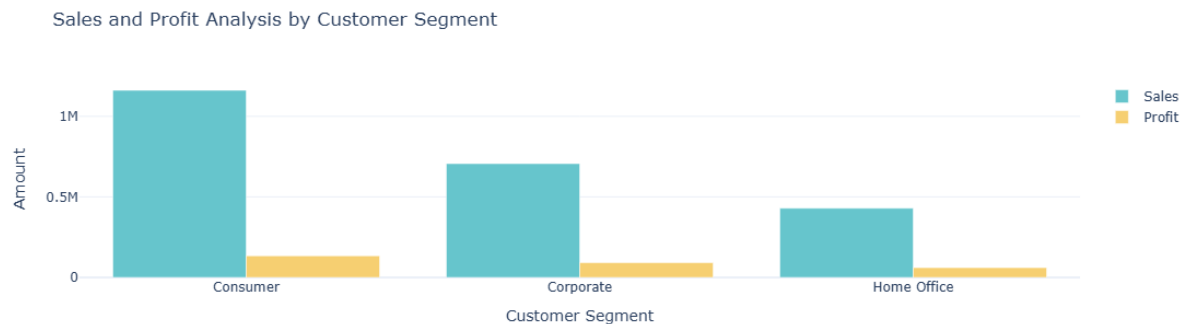
Sales to Profit Ratio

Sales to Profit Ratio ¶

```
[86]: sales_profit_by_segment = data.groupby('Segment').agg({'Sales': 'sum', 'Profit': 'sum'}).reset_index()
      sales_profit_by_segment['Sales_to_Profit_Ratio'] = sales_profit_by_segment['Sales'] / sales_profit_by_segment['Profit']
      print(sales_profit_by_segment[['Segment', 'Sales_to_Profit_Ratio']])

            Segment  Sales_to_Profit_Ratio
      0    Consumer              8.659471
      1   Corporate              7.677245
      2  Home Office             7.125416

[ ]:
```

CHAPTER 10
APPENDIX

```python
import pandas as pd

import plotly.express as px

import plotly.graph_objects as go

import plotly.io as pio

import plotly.colors as colors

pio.templates.default = "plotly_white"

data = pd.read_csv("Sample - Superstore.csv", encoding = 'latin-1')

data.head()

data

data.describe()

data.info()

data['Order Date'] = pd.to_datetime(data['Order Date'])

data['Ship Date'] = pd.to_datetime(data['Ship Date'])

data.info()

data.head()

data['Order Month'] = data['Order Date'].dt.month

data['Order Year'] = data['Order Date'].dt.year

data['Order Day of Week'] = data['Order Date'].dt.dayofweek

data.head()

sales_by_month = data.groupby('Order Month')['Sales'].sum().reset_index()

sales_by_month

fig=px.line(sales_by_month,
        x='Order Month',
        y='Sales',
        title='Monthly Sales Analysis')

fig.show()

data.head()

sales_by_category = data.groupby('Category')['Sales'].sum().reset_index()

sales_by_category

fig = px.pie(sales_by_category,
        values='Sales',
        names='Category',
```

```python
        hole=0.5,

        color_discrete_sequence=px.colors.qualitative.Pastel)


fig.update_traces(textposition='inside', textinfo='percent+label')

fig.update_layout(title_text='Sales Analysis by Category', title_font=dict(size=24))


fig.show()

sales_by_subcategory = data.groupby('Sub-Category')['Sales'].sum().reset_index()

sales_by_subcategory

fig = px.bar(sales_by_subcategory, x= 'Sub-Category', y = 'Sales', title ="Sales analysis by sub category")

fig.show()

data.head()

profit_by_month = data.groupby('Order Month')['Profit'].sum().reset_index()

profit_by_month

fig = px.line(profit_by_month, x = 'Order Month', y = 'Profit', title = 'Mothly profit analysis')

fig.show()

profit_by_category = data.groupby('Category')['Profit'].sum().reset_index()

profit_by_category

fig = px.pie(profit_by_category,

        values='Profit',

        names='Category',

        hole=0.5,

        color_discrete_sequence=px.colors.qualitative.Pastel)

fig.update_traces(textposition='inside', textinfo='percent+label')

fig.update_layout(title_text='Profit Analysis by Category', title_font=dict(size=24))


fig.show()

profit_by_subcategory = data.groupby('Sub-Category')['Profit'].sum().reset_index()

fig = px.bar(profit_by_subcategory, x='Sub-Category',

        y='Profit',

        title='Profit Analysis by Sub-Category')

fig.show()

data.head()

sales_profit_by_segment = data.groupby('Segment').agg({'Sales': 'sum', 'Profit': 'sum'}).reset_index()


color_palette = colors.qualitative.Pastel
```

```
fig = go.Figure()

fig.add_trace(go.Bar(x=sales_profit_by_segment['Segment'],

          y=sales_profit_by_segment['Sales'],

          name='Sales',

          marker_color=color_palette[0]))


fig.add_trace(go.Bar(x=sales_profit_by_segment['Segment'],

          y=sales_profit_by_segment['Profit'],

          name='Profit',

          marker_color=color_palette[1]))

fig.update_layout(title='Sales and Profit Analysis by Customer Segment',

          xaxis_title='Customer Segment', yaxis_title='Amount')


fig.show()

sales_profit_by_segment = data.groupby('Segment').agg({'Sales': 'sum', 'Profit': 'sum'}).reset_index()

sales_profit_by_segment['Sales_to_Profit_Ratio'] = sales_profit_by_segment['Sales'] / sales_profit_by_segment['Profit']

print(sales_profit_by_segment[['Segment', 'Sales_to_Profit_Ratio']])
```