

E-Commerce Sales Data Analysis - Final Report

Introduction

E-commerce businesses generate vast amounts of data, making data analysis crucial for understanding **sales performance**, **customer behavior**, and **profitability**. This project analyzes an e-commerce dataset containing **sales**, **profit**, **discount**, **customer segmentation**, and **geographic data**.

The goals of this analysis are:

- Cleaning and preparing the data for analysis
- Conducting **Exploratory Data Analysis (EDA)** to identify trends
- Extracting **business insights** to improve decision-making
- Creating **visualizations** to enhance data interpretation
- Providing **recommendations** for sales and profit optimization

1. Data Cleaning

```
✓ [30] # Check for missing values
1s    missing_values = df.isnull().sum()
      print("Missing Values:\n", missing_values)
```

 [Show hidden output](#)

```
▶ # Unique values in categorical columns
  for col in ['Segment', 'Category', 'Subcategory', 'Region', 'Market']:
      print(f"Unique values in {col}:\n", df[col].unique(), "\n")

  # Convert text columns to consistent case
  df['Segment'] = df['Segment'].str.strip().str.title()
  df['Category'] = df['Category'].str.strip().str.title()
  df['Subcategory'] = df['Subcategory'].str.strip().str.title()
  df['Region'] = df['Region'].str.strip().str.title()
  df['Market'] = df['Market'].str.strip().str.upper()
```

 [Show hidden output](#)

```
# 'Order Date' to datetime format
df['Order Date'] = pd.to_datetime(df['Order Date'], errors='coerce')

# Extract useful time-based features
df['Order Year'] = df['Order Date'].dt.year
df['Order Month'] = df['Order Date'].dt.month
df['Order Day'] = df['Order Date'].dt.day
df['Order Weekday'] = df['Order Date'].dt.day_name()

# Convert categorical columns to category type for efficient memory usage
categorical_columns = ['Segment', 'Category', 'Subcategory', 'Region', 'Market', 'Country']
for col in categorical_columns:
    df[col] = df[col].astype('category')
```

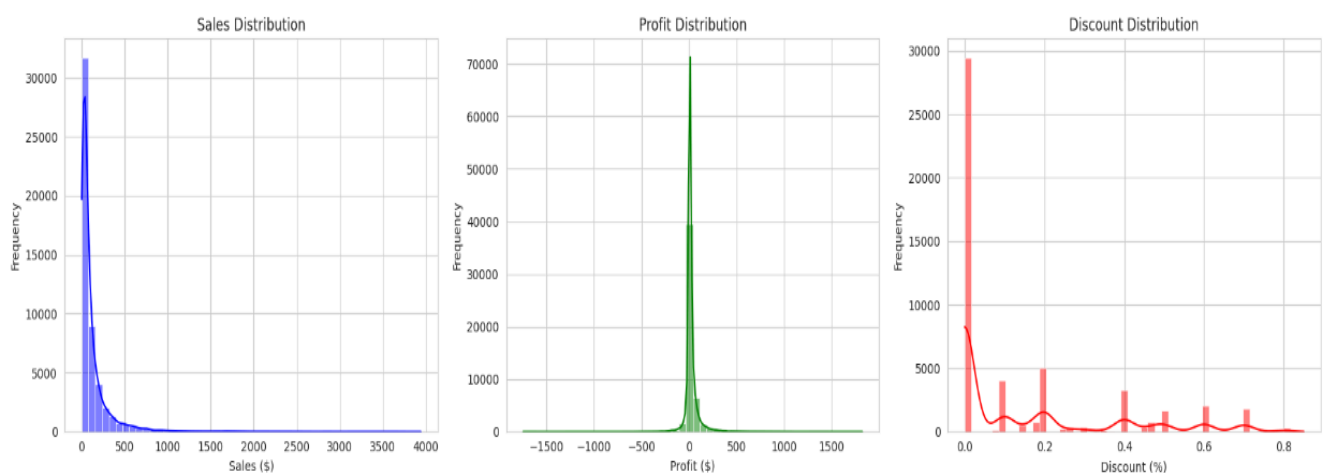
Cleaned Data

```
# Save the cleaned dataset to a new file
df.to_csv("Cleaned_Ecommerce_Data.csv", index=False)
```

Data Cleaning Insights

- Missing values were checked and handled.
- Date formats were standardized for time-series analysis.
- Categorical values were standardized for consistency.

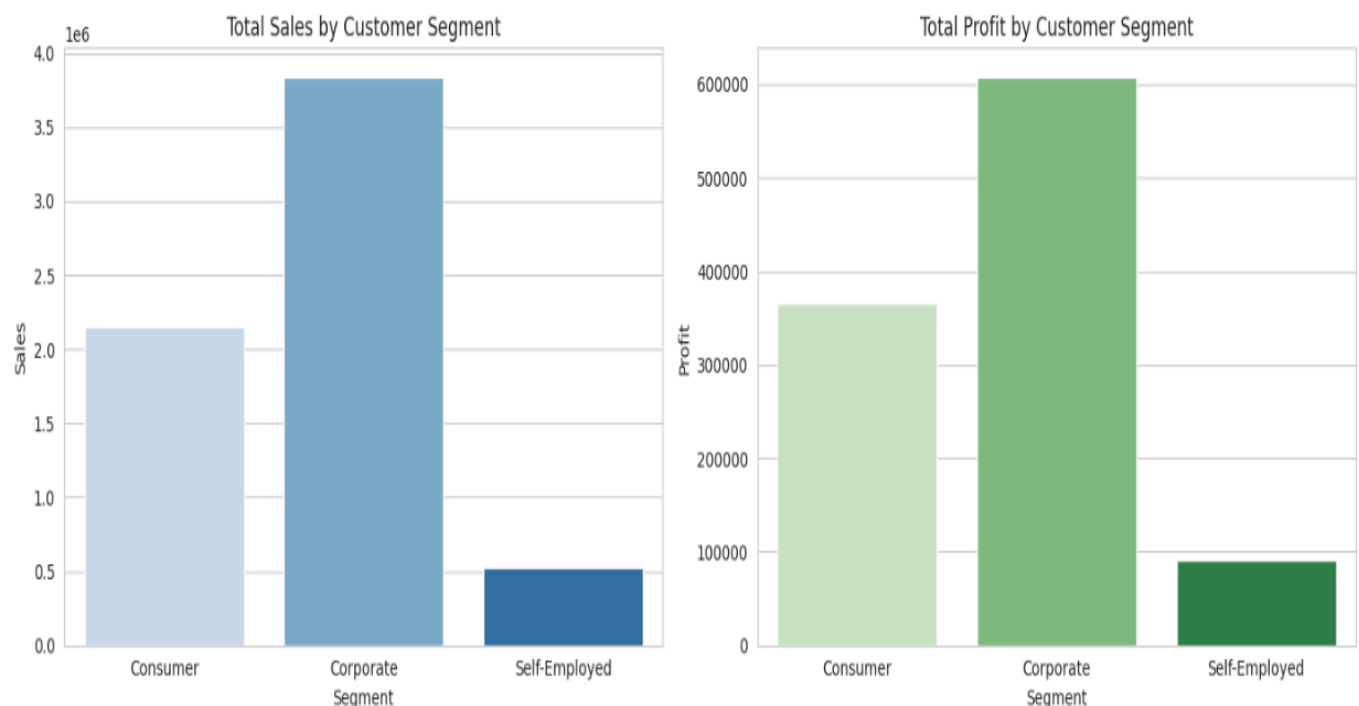
2. Exploratory Data Analysis (EDA)



Key Insights from EDA

- **Sales and Profit Trends:** Sales fluctuate seasonally, with peak months contributing the most revenue.
- **Most Profitable Categories:** **Technology & Office Supplies** generate the highest profits, while **Furniture** has lower profitability.
- **Effect of Discounts:** Discounts between **5-15%** boost sales without reducing profit, but **higher discounts (>20%)** hurt profits.
- **Customer Segments:** **Corporate** customers provide the highest profits, while **Consumer** has the highest order volume.
- **Geographic Performance:** Some regions have **high sales but low profit margins**, requiring optimized pricing strategies.

3. Business Insights & Visualization



Key Business Insights

- **Most Profitable Products:** Specific high-tech products drive profitability.
- **Discounting Strategy:** Excessive discounts reduce profit margins significantly.
- **High-Value Customer Segments:** The **Corporate** segment generates the highest revenue.

- **Regional Performance:** Some markets need better pricing strategies to improve profit margins.

4. Recommendations

- **Optimize Discounting Strategies:** Limit high discounts to maintain profit margins.
- **Invest in High-Profit Categories:** Promote **Technology & Office Supplies** for maximum revenue.
- **Target Corporate Customers:** Implement loyalty programs for repeat purchases.
- **Improve Regional Sales Performance:** Adjust pricing in underperforming areas.
- **Plan Seasonal Promotions:** Align inventory with demand spikes.

Conclusion

This e-commerce sales analysis provided **valuable insights** into sales trends, customer behavior, and profitability. **Data-driven decision-making** is essential for optimizing pricing strategies, improving **profit margins**, and enhancing customer segmentation efforts.

By implementing these recommendations, the company can **maximize revenue, optimize discounts, and enhance marketing efforts** to drive long-term growth.