Exploratory Data Analysis (EDA) of Superstore Dataset

Goals of the Project

The objective of this project is to analyze the Superstore dataset to extract valuable insights related to sales performance, customer behavior, and shipping efficiency. The key goals include:

- Understanding sales trends and customer purchasing patterns.
- Identifying top-performing products and customer segments.
- Analyzing shipping performance to optimize delivery efficiency.
- Deriving business insights to improve profitability and operational strategies.

Materials and Methods

❖ **Dataset Description**

The dataset comprises various attributes related to sales, orders, customers, and shipping, including:

- **Order Information:** Order ID, Order Date, Ship Date, Ship Mode
- **Customer Details:** Customer ID, Customer Name, Segment, Region, City, State
- **Product Details:** Product ID, Category, Sub-Category, Product Name, Quantity
- **Sales Data:** Sales, Discount, Profit

❖ **Tools and Libraries Used**

The analysis is conducted using the following tools:

- **Programming Language:** Python
- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn

General Parts

❖ **Data Preparation and Cleaning**

- Loaded the dataset and checked for missing values and duplicates.
- Dropped unnecessary columns like Row ID and Postal Code.
- Converted Order Date and Ship Date into datetime format.
- Filtered out invalid or inconsistent entries.

❖ **Data Exploration**

- Examined summary statistics of sales, profit, and discount distribution.

- Identified the most and least frequently ordered products.

- Analyzed customer order frequency and purchasing behavior.

Project Outcomes and Insights

❖ **Sales Performance**

- **Top Selling Products:** Identified products contributing the most to revenue.

- **Time-Based Trends:** Observed peak sales months and seasonal fluctuations.

- **Customer Segments:** Found the most profitable customer groups.

❖ **Customer Behavior**

- **Returning vs. New Customers:** Assessed repeat customer purchases.

- **Top Spending Customers:** Listed customers with the highest total purchases.

- **Sales by Region:** Identified locations with the highest order volumes.

❖ **Shipping Performance**

- **Delivery Speed Analysis:** Measured shipping efficiency by categorizing delays.

- **Shipping Mode Comparison:** Evaluated the most frequently used and cost-effective shipping modes.

- **Profitability by Shipping Type:** Identified shipping methods yielding the highest profits.

Feature Engineering

To enhance the analysis, the following new features were created:

- **actual_shipping_delay:** Difference between Order Date and Ship Date.

- **profit_margin:** Ratio of profit to sales for each order.

- **order_year, order_month, order_weekday:** Extracted from Order Date for time-series analysis.

Key Questions and Insights to be Addressed:

- Average Discount by Category:

Category	ABG Discount
Furniture	0.169828
Office Supplies	0.160039
Technology	**0.129977**

- **Number of Orders by Ship Mode:**

Ship Mode	Count Of Orders
Standard Class	2994
Second Class	964
First Class	787
Same Day	264

- **Top 10 Customers by Total Sales:**

Customer ID	Sales
AB-10105	12120.595
HL-15040	11713.828
TA-21385	11649.210
SE-20110	10639.654
BS-11365	10351.023
CC-12370	8953.368
GT-14635	8174.868
TB-21400	7291.958
JH-15985	6695.556
ME-17320	6220.506

- **Average Profit by Category:**

Category	Profit
Furniture	6.868578
Office Supplies	18.004928
Technology	81.200896

➢ **Total Products Sold by Region:**
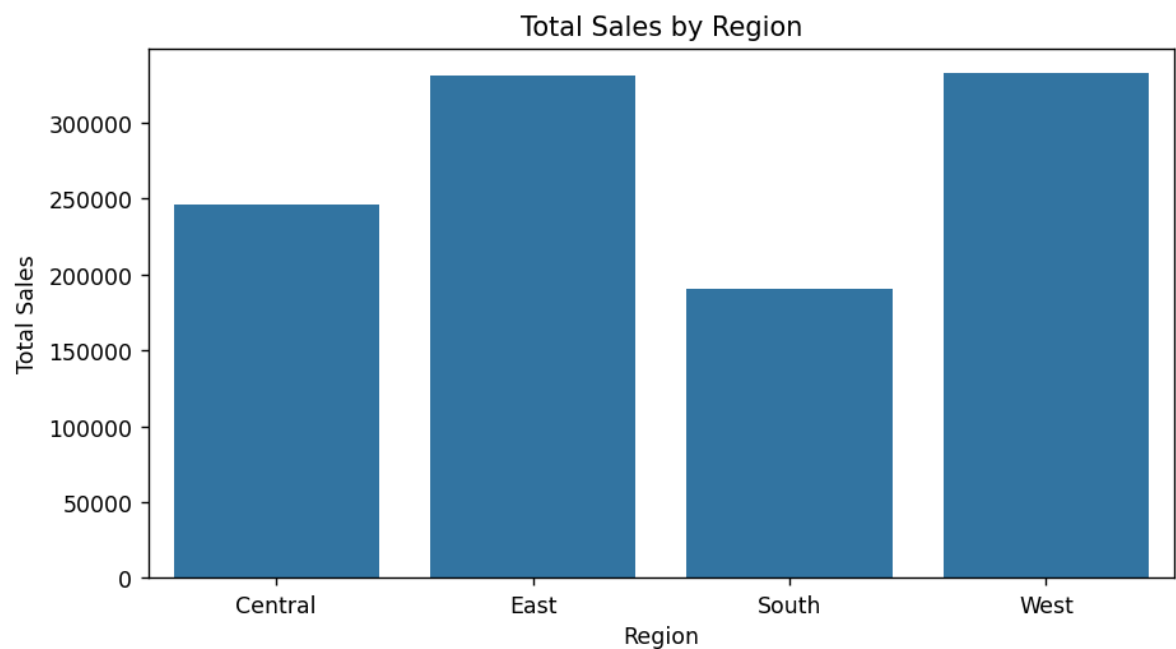
Region   Product Sold

Central   4521

East   5232

South   3142

West   6149

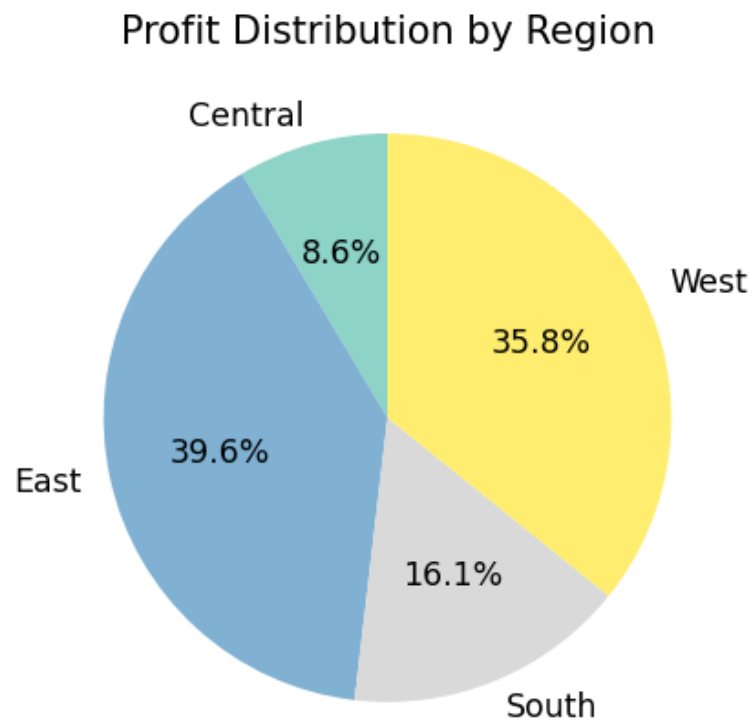➢ **Top 5 Products by Quantity Sold:**

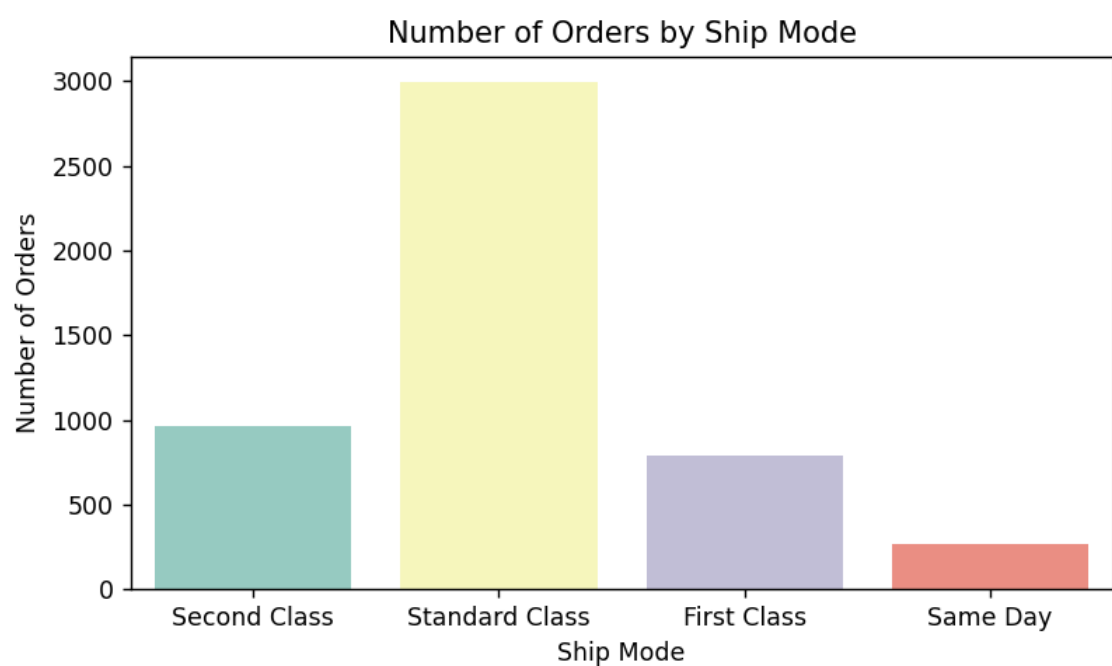| Product Name	Quantity
Staples	441
Avery Non-Stick Binders	47
Situations Contoured Folding Chairs, 4/Set	40
Dual Level, Single-Width Filing Carts	40
KI Adjustable-Height Table	40

Visualization:
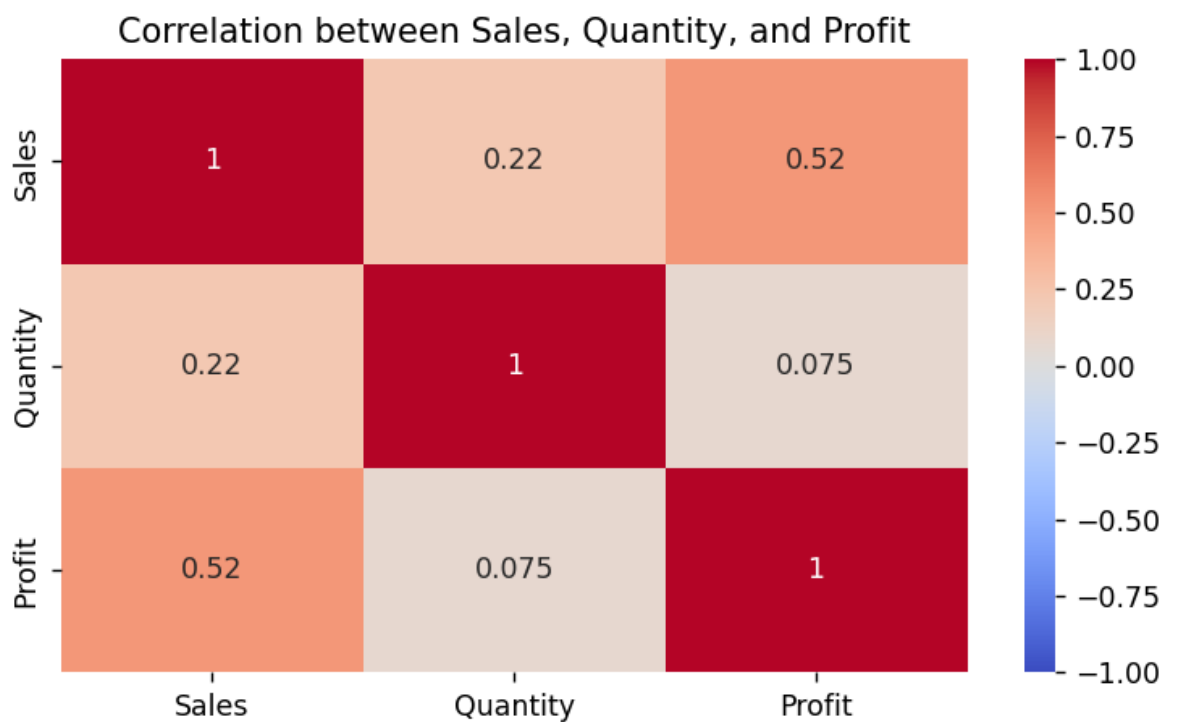
➢ **Total Sales by Region:**

➢ **Profit by Region:**

Profit Distribution by Region

➢ **Number of Orders by Ship Mode:**

Number of Orders by Ship Mode

➢ **Monthly Sales Trend:**

Monthly Sales Trend

➢ **Correlation between Sales, Quantity, and Profit:**
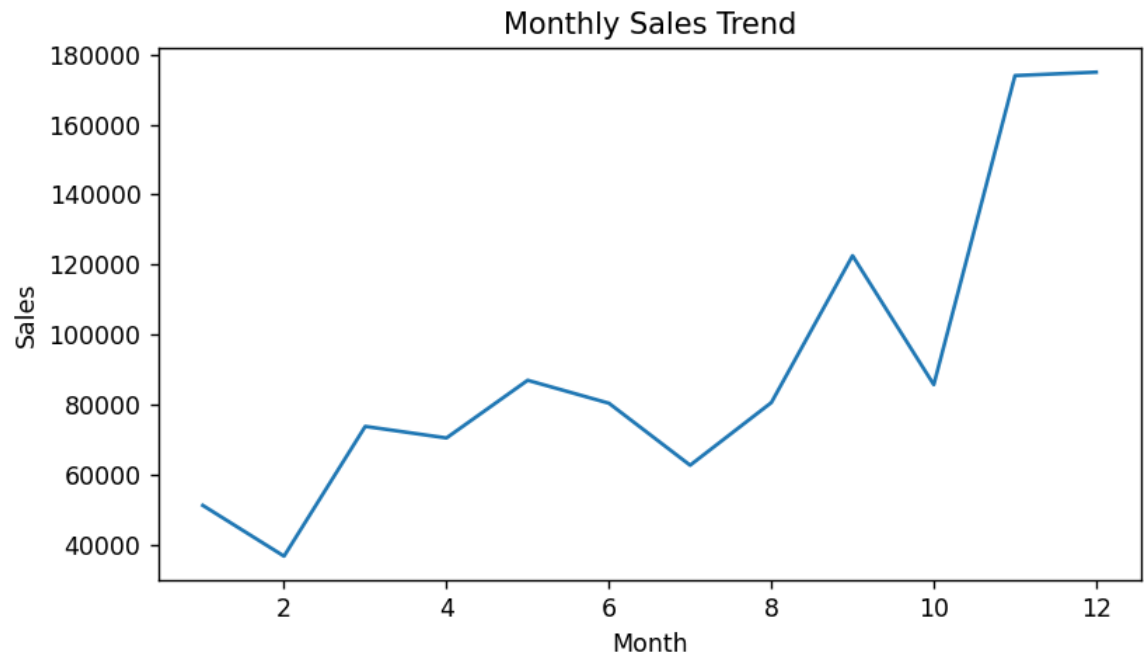
Correlation between Sales, Quantity, and Profit

Conclusion

This analysis provided critical insights into sales performance, customer behavior, and shipping efficiency. Businesses can use these insights to enhance their marketing strategies, improve shipping logistics, and boost overall profitability.

- Gained hands-on experience in working with real-world datasets using Python.

- Learned how to clean, preprocess, and analyze large datasets efficiently.

- Developed expertise in feature engineering to extract meaningful insights.

- Strengthened data visualization skills using Seaborn and Matplotlib.

- Understood the significance of EDA in making data-driven business decisions.

Future Scope

- Implement predictive modeling for future sales forecasting.

- Develop an interactive dashboard for real-time analysis.

- Expand dataset integration with external factors like competitor pricing and market trends.