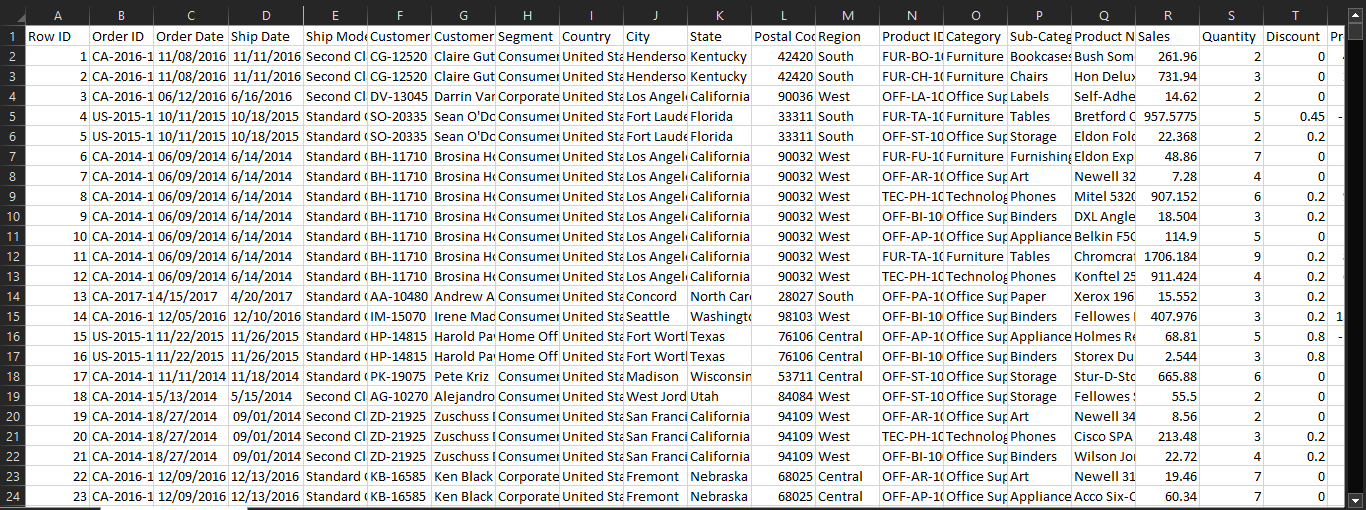
**Exploratory Data Analysis of Superstore Sales Data**

**Objective:**

Analyze the Superstore dataset to uncover insights on:  
 Regional performance  
 Customer segments  
 Product categories & sub-categories  
 Discount policies  
 Shipping efficiency  
…and recommend actionable strategies.

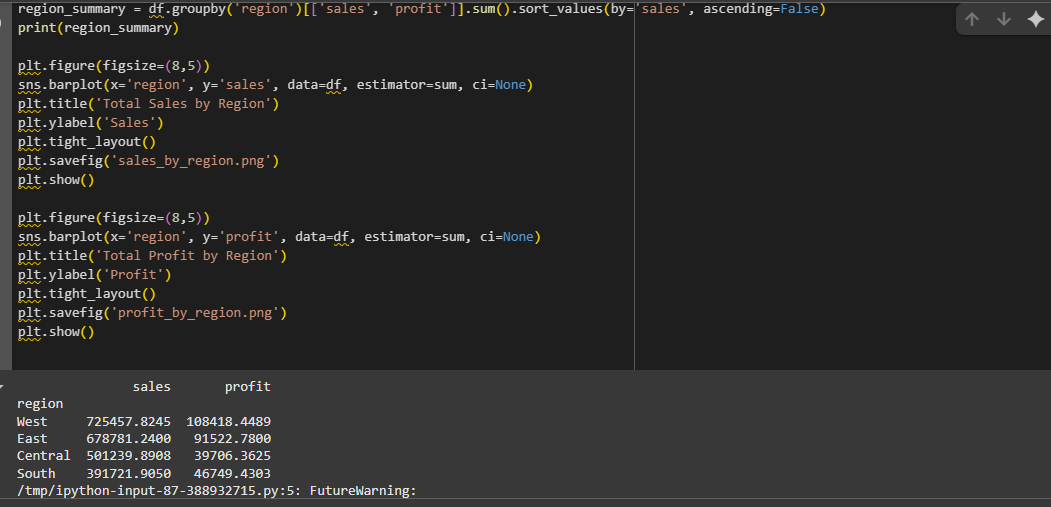
**Dataset Description:**

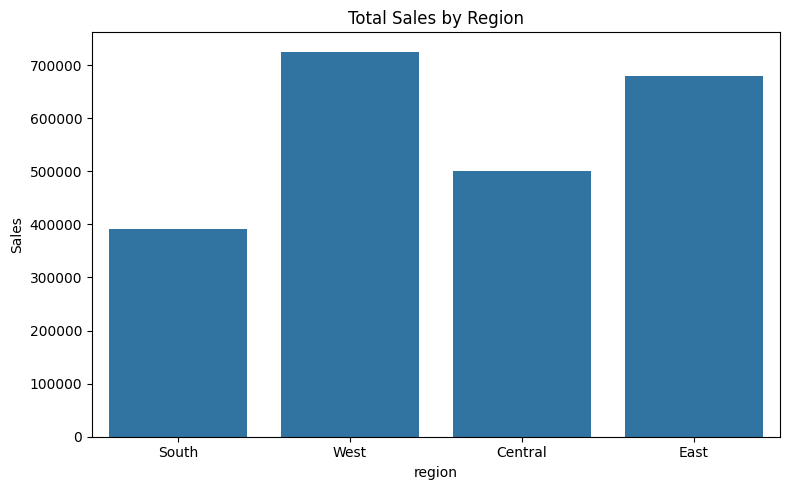
* **Name:** Sample - Superstore.csv
* **Records:** ~10,000 transactions
* **Features:**
  + Order & Ship Dates
  + Customer Name, Segment & Region
  + Product Category & Sub-Category
  + Sales, Quantity, Discount, Profit
* **Period Covered:** Historical transaction data



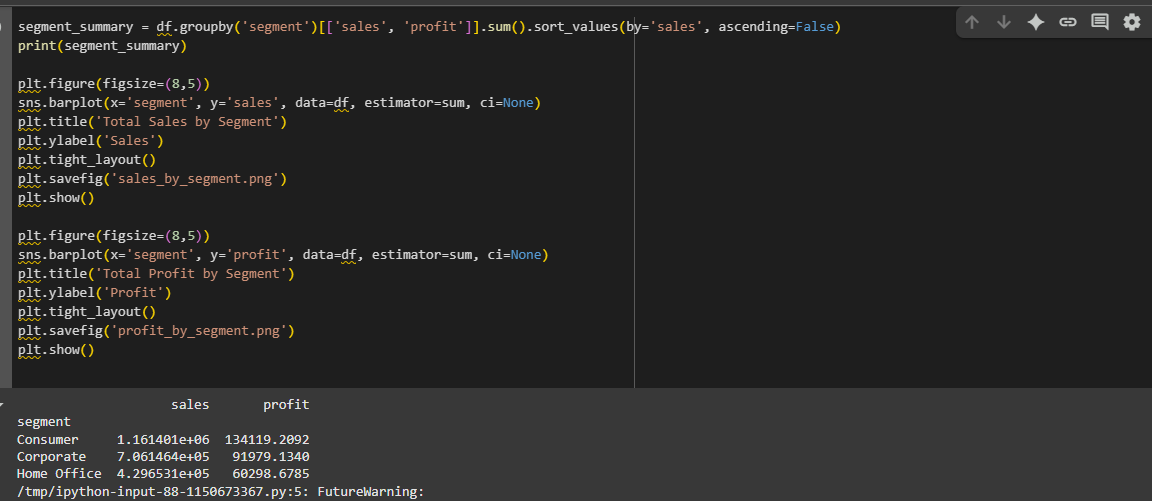
**Business Questions:**

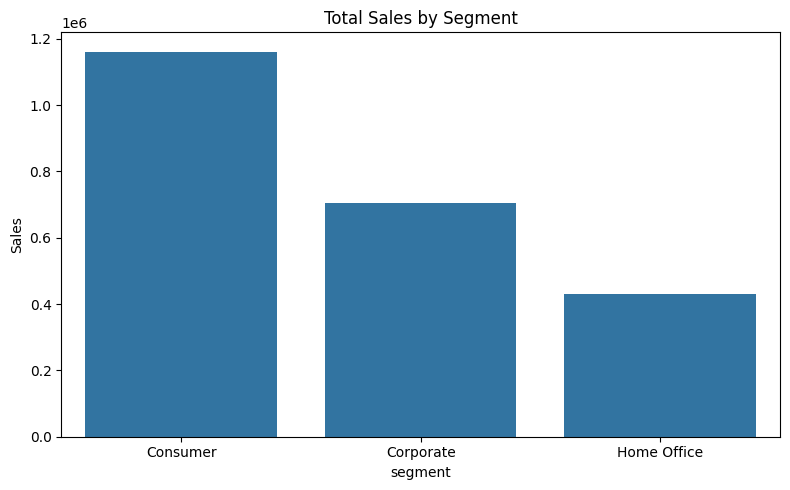
**Which regions generate the highest sales & profit?**



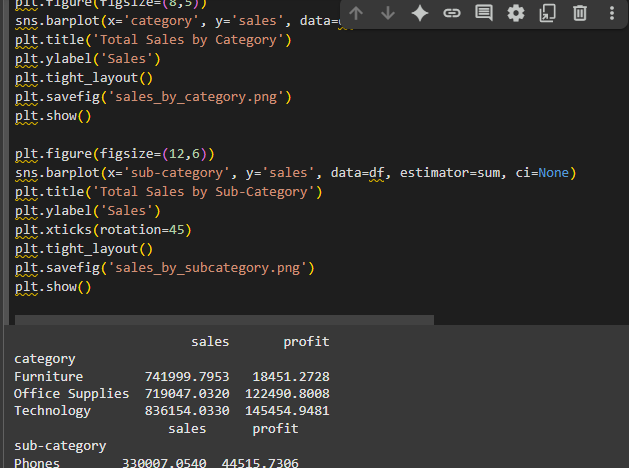


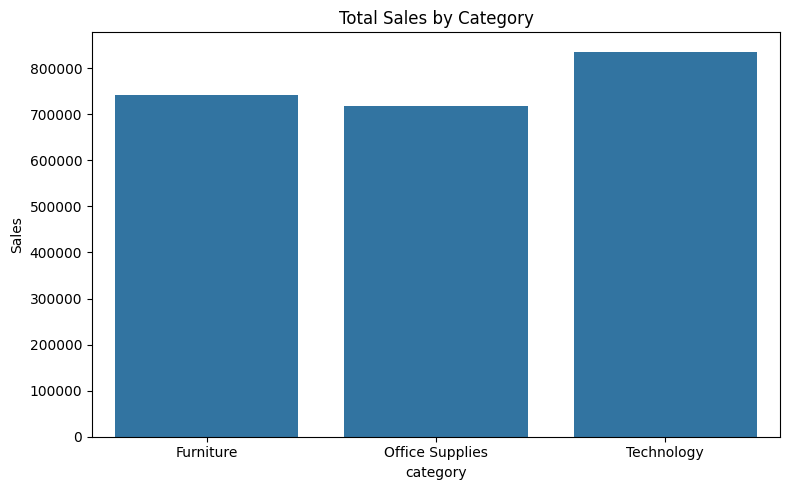
**Which customer segments contribute most to profit?**



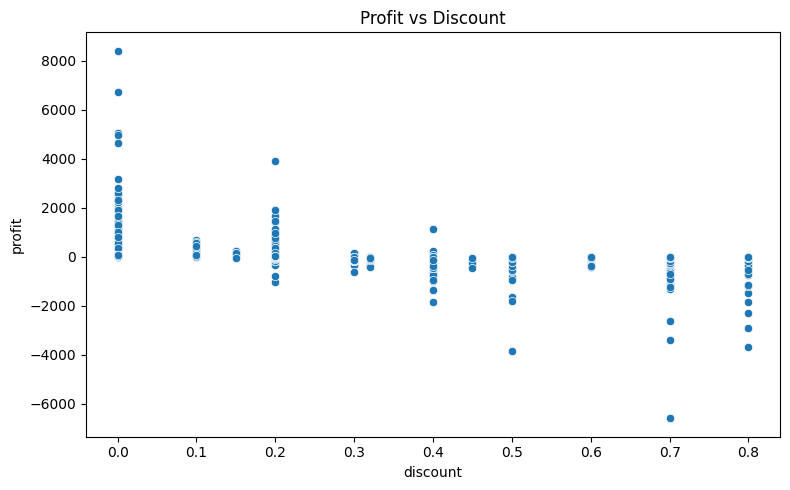
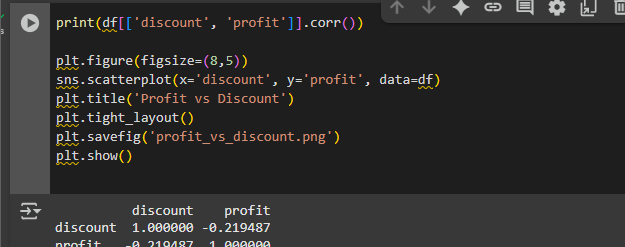


**Which product categories and sub-categories perform best?**

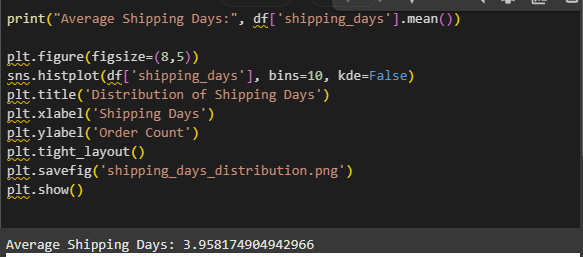




**How do discounts impact profitability?**



**How efficient is the shipping process?**



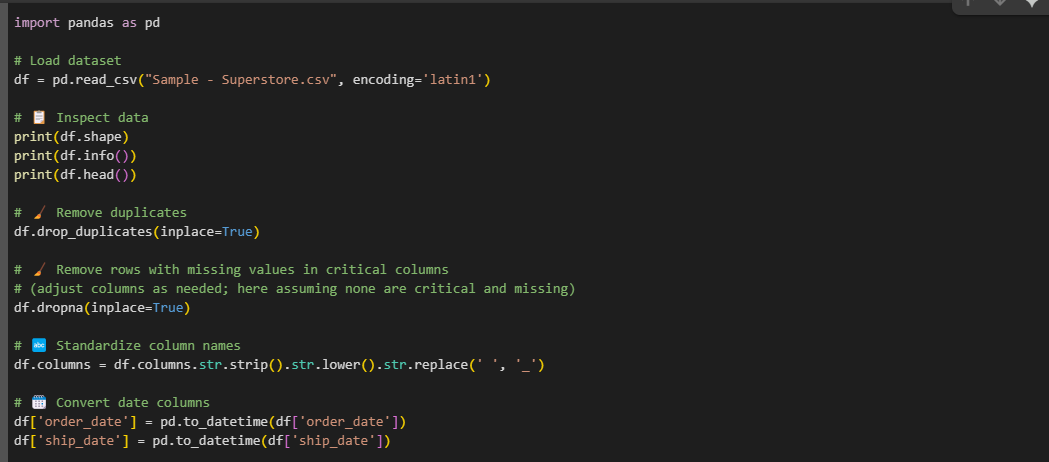
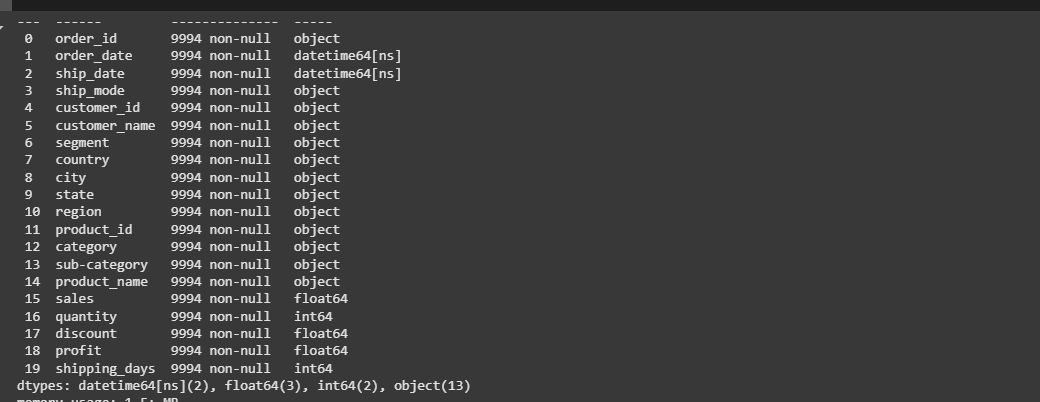
**Data Cleaning:**

**What it does:**

Removes duplicates  
 Drops rows with missing values  
 Standardizes column names to snake\_case  
 Converts order\_date & ship\_date to datetime  
 Computes shipping\_days  
 Drops irrelevant columns like row\_id & postal\_code  
 Saves a clean version as Superstore\_Cleaned.csv

**Steps:**

* **Removed missing & duplicate records**
  + Checked for and eliminated rows with missing critical fields.
  + Dropped exact duplicate rows to avoid over-representing any transaction.
* **Converted date columns**
  + Converted Order Date and Ship Date columns from text to proper datetime format.
  + Enabled calculation of derived fields like shipping duration.
* **Standardized column names**
  + Renamed columns to lowercase.
  + Replaced spaces with underscores for easier code handling.
  + Example: Order Date → order\_date, Ship Date → ship\_date.
* **Removed irrelevant columns**
  + Dropped columns that are not meaningful for business analysis, such as:
    - Row ID
    - Postal Code

  Loaded the data → Cleaned it → Removed irrelevant columns → Converted dates → Standardized names → Added shipping\_days.

dataset has:  
 **9,994 rows × 20 columns**  
 order\_date & ship\_date are proper datetimes  
 shipping\_days is calculated  
 row\_id & postal\_code removed  
 column names are clean (lowercase + underscores)

**Superstore EDA — Results**

**Region-wise Sales & Profit**

| **Region** | **Sales** | **Profit** |
| --- | --- | --- |
| **West** | 725,457.82 | 108,418.45 |
| **East** | 678,781.24 | 91,522.78 |
| **Central** | 501,239.89 | 39,706.36 |
| **South** | 391,721.90 | 46,749.43 |

📌 *Insight: West region leads both in sales and profit; South region is weakest.*

**Segment-wise Sales & Profit**

| **Segment** | **Sales** | **Profit** |
| --- | --- | --- |
| **Consumer** | 1,161,401.00 | 134,119.21 |
| **Corporate** | 706,146.40 | 91,979.13 |
| **Home Office** | 429,653.10 | 60,298.68 |

📌 *Insight: Consumer segment contributes the most to sales and profit; Home Office is smallest.*

**Category-wise Sales & Profit**

| **Category** | **Sales** | **Profit** |
| --- | --- | --- |
| **Furniture** | 741,999.79 | 18,451.27 |
| **Office Supplies** | 719,047.03 | 122,490.80 |
| **Technology** | 836,154.03 | 145,454.95 |

📌 *Insight: Technology is the most profitable; Furniture has high sales but very low profit.*

**Top 5 Sub-Categories by Sales**

| **Sub-Category** | **Sales** | **Profit** |
| --- | --- | --- |
| **Phones** | 330,007.05 | 44,515.73 |
| **Chairs** | 328,449.10 | 26,590.17 |
| **Storage** | 223,843.61 | 21,278.83 |
| **Tables** | 206,965.53 | -17,725.48 |
| **Binders** | 203,412.73 | 30,221.76 |

📌 *Insight: Phones & Chairs dominate sales; Tables incur losses despite high sales.*

**Discount vs Profit**

Correlation between **Discount** and **Profit**:  
-0.219 → Higher discounts are associated with lower profits.

📌 *Insight: Keep discounts below 30% where possible to protect margins.*

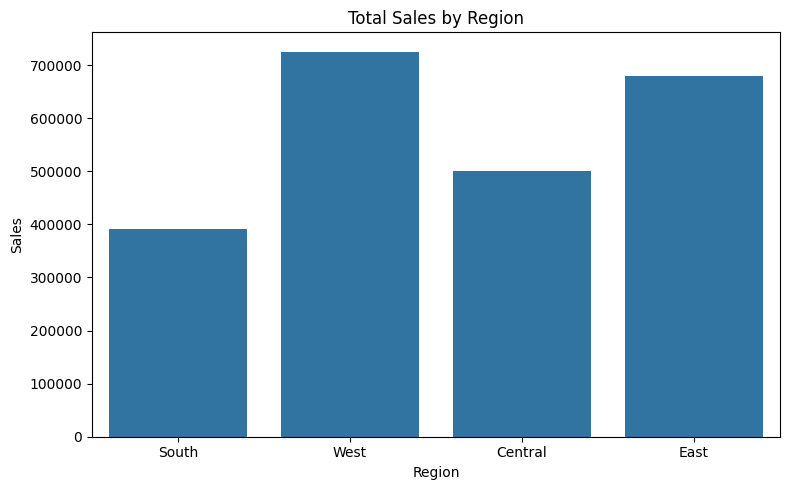
**Shipping Efficiency**

Average shipping time: **~3.96 days**

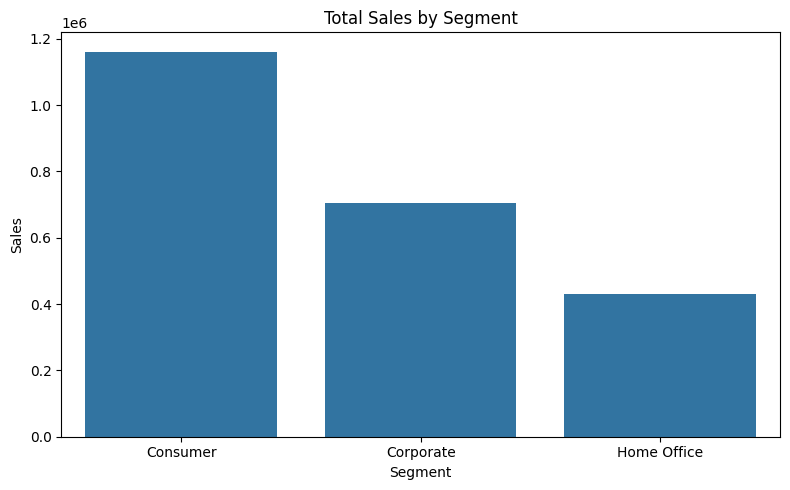
📌 *Insight: Shipping performance is good and should be maintained.*

**Visualizations:**

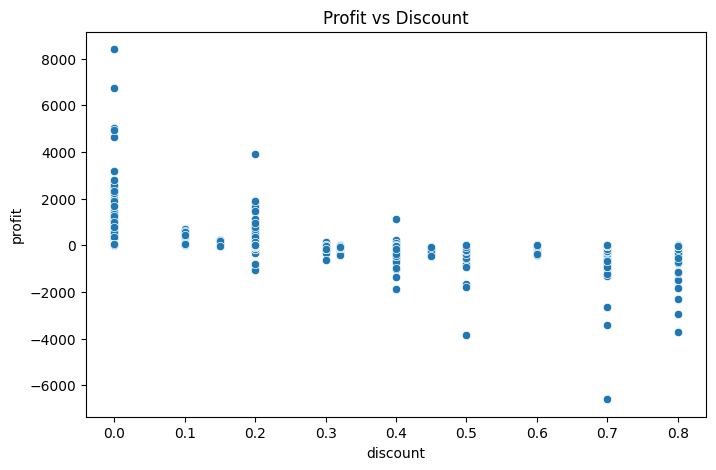
* Barplots:
  + Sales & Profit by Region



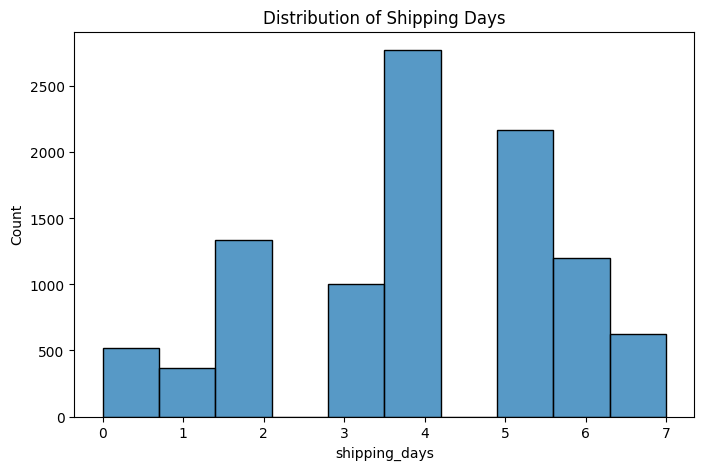
* + Sales & Profit by Segment



* Scatterplot: Profit vs Discount



* Histogram: Shipping Days Distribution



**Insights**

West region has the highest sales & profit.  
 Consumer segment contributes the most to profit.  
 Technology category performs best; Furniture struggles.  
Higher discounts (>30%) reduce profit.  
 Average shipping time ~3.96 days — acceptable.

**Recommendations**

**Regional Strategies**

Continue to **invest in West & East regions**, which already perform strongly.  
 **Focus on improving sales & profitability in the South region**, which currently lags behind.  
 Investigate what works well in the West/East and replicate best practices in the South.

**Customer Segments**

Maintain strong engagement with the **Consumer segment**, as it’s the most profitable.  
 Design targeted promotions & outreach for the **Home Office segment**, which is underperforming.

**Product Categories**

Reassess the **Furniture category**, which shows high sales but low/negative profits.

* Consider optimizing costs, renegotiating supplier contracts, or adjusting prices.  
   Promote high-margin **Technology products** more aggressively.  
   Reduce focus (or redesign pricing) for loss-making **sub-categories** like Tables & Bookcases.

**Discount Policy**

Discounts above ~30% lead to significant profit erosion — cap discounts at **20– 30% max**.  
 Introduce smarter, targeted discounting instead of blanket discounts.

**Logistics**

Maintain current shipping efficiency (~3–4 days on average), which aligns with customer expectations.  
Continue monitoring shipping times & look for incremental improvements.

**Tools & Skills Applied**

* **Data Cleaning & Preprocessing:** pandas
* **Analysis & Aggregation:** pandas
* **Visualization:** matplotlib, seaborn
* **Statistical Insights:** correlation analysis

Skills: Data cleaning, EDA, descriptive statistics, business insights, storytelling with data.