Exploratory Data Analysis (EDA) Report

# Dataset: Superstore Sales

This project involved the analysis of the Superstore Sales dataset, which contains 9,994 records and 21 columns.   
The objective of this EDA was to uncover key business insights and identify patterns related to sales, profit, discount, product categories, customer segments, and geographical regions using Python's Pandas, Seaborn, and Matplotlib libraries.

# Step-by-Step Analysis

## 1. Data Loading and Inspection

- Loaded the dataset and examined data types and basic statistics.  
- Dataset contains 15 categorical features and 6 numerical ones (3 integer, 3 float).  
- No missing values or duplicates were found.

## 2. Data Cleaning

- Removed unnecessary columns: 'Row ID' and 'Postal Code'.  
- Converted 'Order Date' and 'Ship Date' to datetime format.  
- Created two new features: 'Order Month' and 'Order Year' to enable time-based analysis.

## 3. Univariate Analysis

- Bar plots were used to analyze total sales and profit by 'Category' and 'Sub-Category'.  
 - Technology had the highest profit margin.  
 - Furniture showed high sales but often negative profit.  
 - A pie chart was used to display segment distribution, showing the 'Consumer' segment accounted for the highest percentage of orders (~52%).  
 - Count plot on 'Ship Mode' showed 'Standard Class' was the most used shipping method.

## 4. Bivariate Analysis

- **Scatter Plot:** ***Sales vs Profit***  
 - Showed many high-sales orders resulting in negative profit.  
 - Technology had clusters in high profit zones.  
 - Furniture consistently underperformed in profitability.

- **Scatter Plot: *Discount vs Profit*** - Discounts above 20% were mostly linked with negative profit.  
 - Even low discounts in Furniture failed to generate significant profit.

**- Bar Plot: *Region-wise Sales and Profit***  
 - West Region: Highest in both sales and profit.  
 - Central Region: High sales but lowest profit — possibly due to high discount rates.  
  
**- Line Plot: *Year-wise Sales Trend***  
 - Sales increased steadily from 2014 to 2017, showing growth in business performance.

## 5. Correlation Heatmap

- A heatmap of the correlation matrix revealed:  
 - Moderate positive correlation between Sales and Profit.  
 - Strong negative correlation between Discount and Profit.  
 - Weak or no correlation between Quantity and other variables.

# Key Business Insights

1. Large orders do not guarantee profitability. In fact, some lead to substantial losses.  
2. Technology is the most profitable category; Furniture needs review.  
3. Discounts should be optimized. More than 20% discount often results in losses.  
4. Central region offers high discounts but fails to convert sales into profit.  
5. Sales are rising year-over-year, reflecting a growing customer base.

# Conclusion

This EDA provides critical insights for strategic business decisions. The company should evaluate its pricing strategy, optimize discount policies, and reassess its Furniture product line.   
The regional performance and shipping modes should also be reviewed to boost profitability across segments.