E-Commerce Sales & Profit Analysis

Tools Used: Python | SQL | Power BI
Dataset: Sample Superstore Dataset

1. Introduction

This project focuses on analyzing e-commerce sales data to understand sales performance, profitability, discount impact, and regional trends. The goal is to extract meaningful business insights that can support data-driven decision making.

2. Dataset Description

The dataset contains transactional sales data with the following key attributes:

- Ship Mode

- Segment

- Country, City, State, Region

- Category, Sub-Category

- Sales, Quantity, Discount, Profit

The dataset was clean with **no missing values**, making it suitable for direct analysis.

3. Data Preprocessing (Python)

Data preprocessing was performed using Python to ensure accuracy and consistency.

Steps Performed:

- Removed duplicate records

- Verified data types and numerical ranges

- Checked for missing values (none found)

- Prepared data for analysis and visualization

📌 **Screenshot – Data Overview / `df.info()` Output:**

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 13 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Ship Mode     9994 non-null   object
 1   Segment       9994 non-null   object
 2   Country       9994 non-null   object
 3   City          9994 non-null   object
 4   State         9994 non-null   object
 5   Postal Code   9994 non-null   int64
 6   Region        9994 non-null   object
 7   Category      9994 non-null   object
 8   Sub-Category  9994 non-null   object
 9   Sales         9994 non-null   float64
 10  Quantity      9994 non-null   int64
 11  Discount      9994 non-null   float64
 12  Profit        9994 non-null   float64
dtypes: float64(3), int64(2), object(8)
memory usage: 1015.1+ KB
```

4. Exploratory Data Analysis (EDA)

4.1 Category & Sub-Category Analysis

- Identified profit contribution across categories and sub-categories

- Technology category emerged as the most profitable

📌 **Screenshot – Category/Sub-Category Profit Chart:**

Category	Profit
Technology	145454.9481
Office Supplies	122364.6608
Furniture	18421.8137

Sub-Category	
Copiers	55617.8249
Phones	44515.7306
Accessories	41936.6357
Paper	33944.2395
Binders	30228.0003
Chairs	26567.1278
Storage	21278.8264
Appliances	18138.0054
Furnishings	13052.7230
Envelopes	6964.1767
Art	6524.6118
Labels	5526.3820
Machines	3384.7569
Fasteners	949.5182
Supplies	-1189.0995
Bookcases	-3472.5560
Tables	-17725.4811

4.2 Regional Performance Analysis

- Compared sales and profit across regions

- West region showed the strongest performance

📌 **Screenshot – Region-wise Sales/Profit Chart:**

```
region_sales, region_profit

(Region
 Central    500782.8528
 East       678435.1960
 South      391721.9050
 West       725255.6365
 Name: Sales, dtype: float64,
 Region
 Central     39655.8752
 East        91506.3092
 South       46749.4303
 West       108329.8079
 Name: Profit, dtype: float64)
```
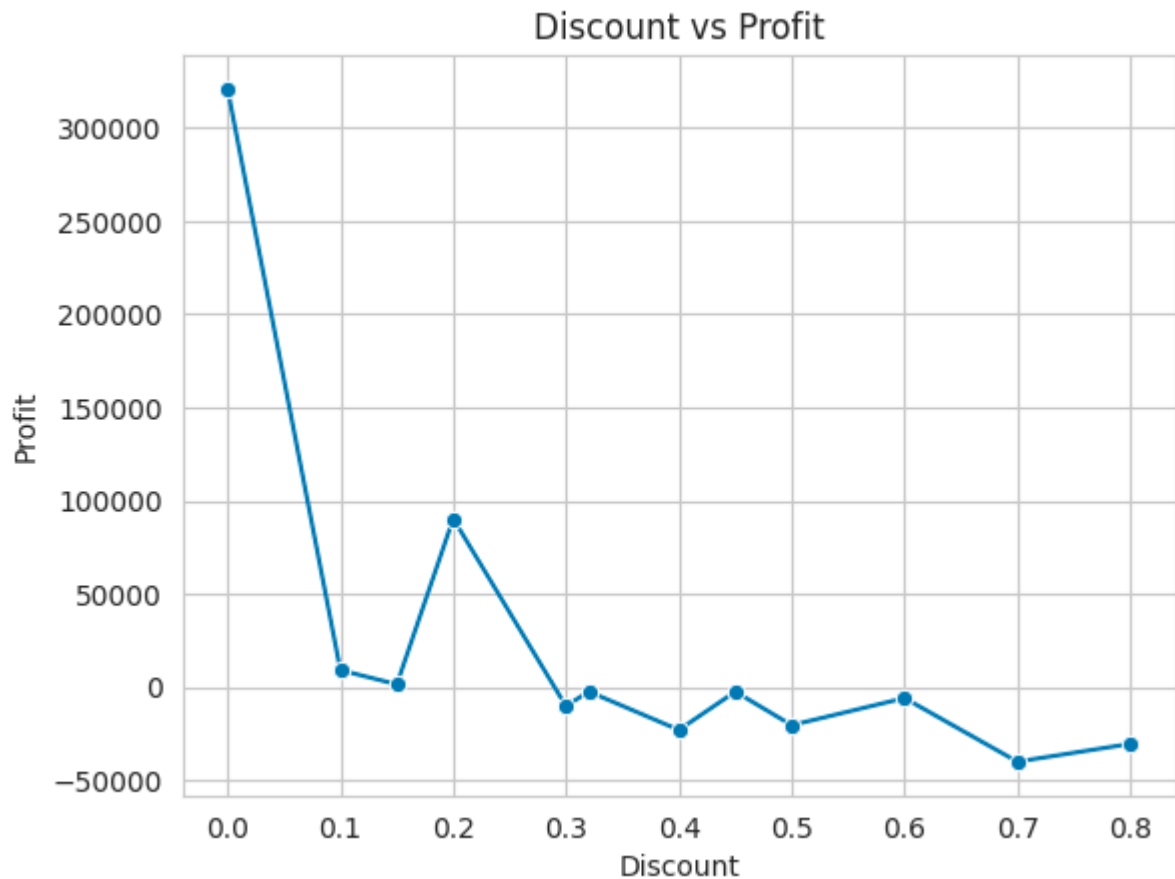
4.3 Discount vs Profit Analysis

- Analyzed how different discount levels impact profit

- Higher discounts were associated with reduced overall profitability

📌 **Screenshot – Discount vs Profit Plot:**

Discount vs Profit

4.4 Segment & Shipping Mode Analysis

- Consumer segment contributed the highest profit

- Shipping mode analysis provided operational insights

📌 **Screenshot – Segment / Ship Mode Analysis:**

Segment	Profit
Consumer	134007.4413
Corporate	91954.9798
Home Office	60279.0015

Ship Mode	Profit
First Class	48953.6561
Same Day	15871.8869
Second Class	57446.6516
Standard Class	163969.2280

4.5 Sales vs Profit Relationship

- High sales do not always translate into high profit

- Discounts and cost factors play a significant role

📌 **Screenshot – Sales vs Profit Scatter Plot:**
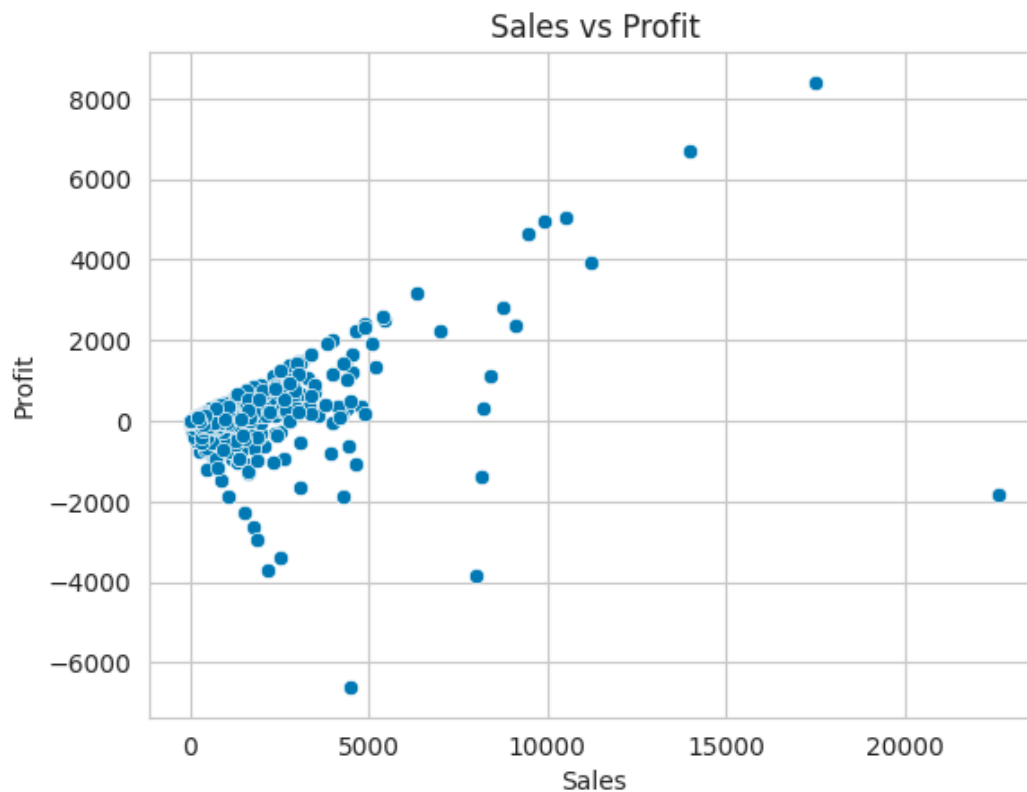
5. SQL Analysis

SQL was used to perform structured business queries such as:

- Total sales and profit calculation

- Category-wise and segment-wise profit analysis

- Discount impact on profitability

SQL Query Results:

-- Total Sales & Profit

SELECT

   sum(sales) AS total_sales,

      sum(profit) AS total_profit

FROM superstore_data

total_sales	total_profit
2296195.5903	286241.4226

-- Profit by Category

SELECT category, sum(profit) AS total_profit

FROM superstore_data

GROUP BY category

ORDER BY total_profit DESC;

	category	total_profit
1	Technology	145454.9481
2	Office Supplies	122364.6608
3	Furniture	18421.8137

-- Profit by Segment

SELECT segment,

  SUM(profit) AS total_profit

FROM superstore_data

GROUP BY segment

ORDER BY total_profit DESC;

	segment	total_profit
1	Consumer	134007.4413
2	Corporate	91954.9798
3	Home Office	60279.0015

-- Discount Impact

SELECT discount, SUM(profit) AS total_profit

FROM superstore_data

GROUP BY discount

ORDER BY discount;

|	discount	total_profit
1	0.0	320844.4073
2	0.1	9029.177
3	0.15	1418.9915
4	0.2	90306.607
5	0.3	-10357.2186
6	0.32	-2391.1377
7	0.4	-23057.0504
8	0.45	-2493.1111
9	0.5	-20506.4281
10	0.6	-5944.6552
11	0.7	-40075.3569
12	0.8	-30532.8022

6. Power BI Dashboard

An interactive Power BI dashboard was created to visualize key business metrics.

Dashboard Features:

- KPI Cards: Total Sales, Total Profit, Quantity

- Charts:

  - Profit by Category

  - Sales by Region

  - Profit by Segment

  - Discount vs Profit

- Filters:

  - Region

  - Category

  - Segment

  - Ship Mode

📌 **Screenshot – Power BI Dashboard:**
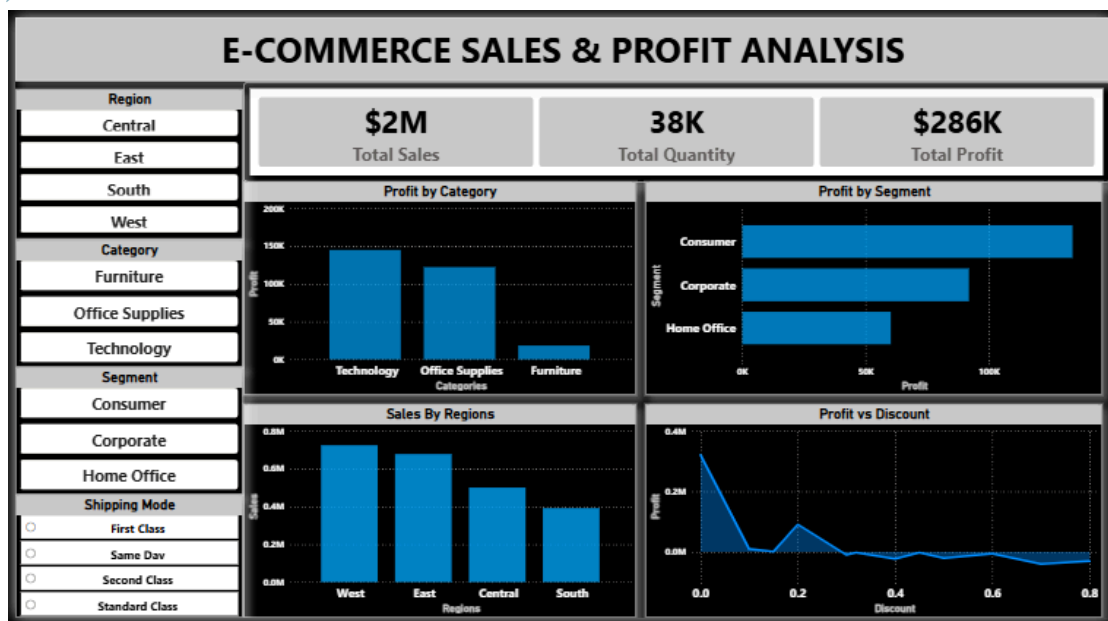
7. Key Insights

- Technology category generates the highest profit

- High discounts negatively impact profitability

- West region performs best in terms of sales and profit

- Consumer segment is the most profitable

- High sales volume does not guarantee high profit

8. Conclusion

This project demonstrates end-to-end data analytics using Python, SQL, and Power BI. The analysis highlights critical business insights related to profitability, discount strategy, and regional performance, showcasing practical data analytics skills suitable for entry-level roles.

9. Skills Demonstrated

- Data preprocessing and EDA using Python

- Business analysis using SQL

- Data visualization and dashboarding with Power BI

- Insight-driven decision support