SUPERSTORE SALES DATA ANALYSIS

REPORT

BRAINWAVE MATRIX SOLUTIONS
DATA SCIENCE/DATA ANALYTICS
MAULI PATEL
MAY 25, 2025

TABLE OF CONTENTS

➢ Introduction

This report presents an in-depth analysis of the Superstore dataset using two approaches:

1. An interactive **dashboard-based analysis**, and
2. A **Python-powered Jupyter Notebook analysis**.

It aims to provide insights into sales performance, category breakdowns, regional impact, and profitability trends.

➢ Internship Details

Field	Information
Intern Name	Mauli Patel
Batch	May 19
Internship Domain	Data Analytics / Business Intelligence
Interest Areas	Data Visualization, Insight Generation, Tableau, Python, Data Analysis, Data Scientist, AI

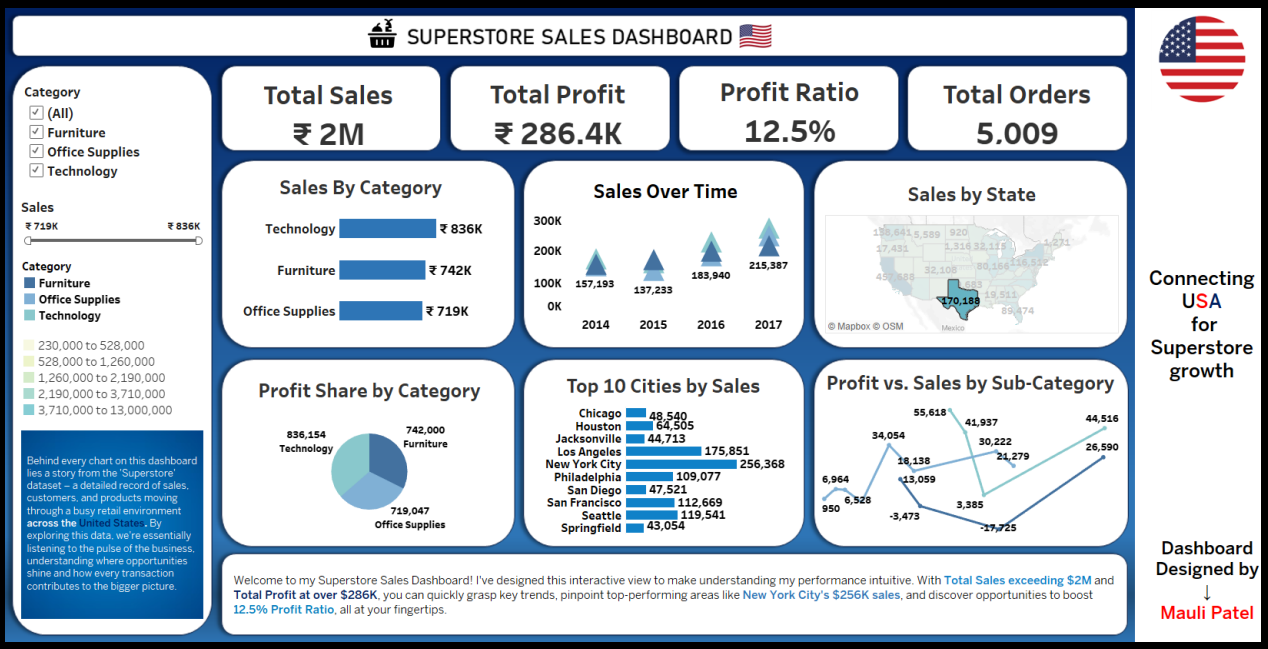
➢ Sales Dashboard Overview

The dashboard presents a visual summary of Superstore's key performance indicators (KPIs) and trends using graphs, maps, and tables.

This dashboard provides an interactive visual overview of Superstore's US sales performance. Key performance indicators include:

- **Total Sales:** ₹2M
- **Total Profit:** ₹286.4K
- **Profit Ratio:** 12.5%

- **Total Orders:** 5,009



➢ Python (Jupyter Notebook) Analysis

The Python analysis file, superstore_analysis.ipynb, includes code-driven exploration of the dataset. Here's a breakdown of that analysis:

✓ Libraries Used

⇒ pandas for data handling

⇒ matplotlib & seaborn for visualization

⇒ numpy for numerical operations

✓ Data Overview

- **Rows:** 9994
- **Columns:** 21

- Features include: Order ID, Sales, Profit, Region, Category, State, etc.

Basic Information

```
In [11]: print("Shape:", data.shape)
```

```
Shape: (9994, 21)
```

```
In [12]: print("Columns:", data.columns.tolist())
```

```
Columns: ['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode', 'Customer ID', 'Customer Name', 'Segment', 'Countr
y', 'City', 'State', 'Postal Code', 'Region', 'Product ID', 'Category', 'Sub-Category', 'Product Name', 'Sales', 'Quantit
y', 'Discount', 'Profit']
```

```
In [55]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 24 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   datetime64[ns]
 3   Ship Date      9994 non-null   datetime64[ns]
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Sub-Category   9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
 21  Month_Year     9994 non-null   period[M]
 22  Month          9994 non-null   object
 23  Day            9994 non-null   object
dtypes: datetime64[ns](2), float64(3), int64(3), object(15), period[M](1)
memory usage: 1.8+ MB
```

➢ Missing Values

- Checked using isnull().sum()
- Result: No missing values in the dataset

```
Missing Values:
 Row ID           0
Order ID          0
Order Date        0
Ship Date         0
Ship Mode         0
Customer ID       0
Customer Name     0
Segment           0
Country           0
City              0
State             0
Postal Code       0
Region            0
Product ID        0
Category          0
Sub-Category      0
Product Name      0
Sales             0
Quantity          0
Discount          0
Profit            0
dtype: int64
Duplicates: 0
```

- **Profit by Region**

Region	Profit Trend
West	Highest profit
Central	Low/negative

Insight: Business is more profitable in the West region.

- **Ship Mode Analysis**

Ship Mode	Sales Share
Standard Class	Highest
Same Day	Lowest

➢ Data Analysis and Visualization

Total Sales and Profit Overview(Key Metrices)

KPI	Value
Total Sales	₹2,000,000
Total Profit	₹286,400
Profit Ratio	12.5%
Total Orders	5,009

Sales by Category

Category	Sales (₹)
Technology	836K
Furniture	742K
Office Supplies	719K

Interpretation: Technology leads in overall sales, followed by Furniture and Office Supplies.

Sales Over Time(Sales Trend (2014–2017))

Year	Sales (₹)
2014	157,193
2015	137,233
2016	183,940

Year	Sales (₹)
2017	215,387

Interpretation: Sales show an upward trend, peaking in 2017.

State-Level Sales Distribution(Sales by State)

Top States	Sales (₹)
Texas	170,189
California	256,368
New York	175,851

Interpretation: California is the top-performing state in sales, followed by New York and Texas.

Profit Share by Category

Category	Profit Contribution
Technology	₹836,154
Furniture	₹742,000
Office Supplies	₹719,000

Interpretation: Technology not only leads in sales but also significantly contributes to profits.

Top 10 Cities by Sales

City	Sales (₹)
New York City	256,368
Los Angeles	175,851
Seattle	119,541
San Francisco	112,669
Philadelphia	109,077
Houston	64,505
Chicago	48,540
San Diego	47,521
Jacksonville	44,713
Springfield	43,054

Interpretation: New York City and Los Angeles dominate city-level sales.

Top Cities by Sales

City	Sales (₹)
New York City	256,368
Los Angeles	175,851
Seattle	119,541

Profit by Sub-Category

Sub-Category	Profit (₹)
Phones	55,618
Chairs	-17,725
Binders	18,138
Tables	-3,473
Copiers	41,937
Accessories	44,516

Interpretation:

- **Phones** and **Accessories** are highly profitable.
- **Chairs** and **Tables** show negative profit, indicating a need for strategic review.

➢ **Key Insights & Interpretation**

- The business has consistent growth from 2014 to 2017.

- Technology category shows maximum sales and profitability.
- New York City is the highest-grossing city.
- Sub-categories such as Chairs and Tables are operating at a loss and need attention.
- The overall profit margin of **12.5%** is healthy, with potential for optimization in low-performing sub-categories.
- **Top Categories:** Technology leads in both analyses
- **Loss Drivers:** Chairs, Tables, and excessive discounts
- **Regional Trend:** Western states drive most profits
- **Top Cities:** New York City, LA, Seattle
- **Consistent Growth:** Sales trend improves yearly
- **Data Health:** No missing values, ideal for ML if needed later

➢ **Conclusion**

This dual-method analysis (Dashboard + Python) provides a comprehensive view of Superstore's performance.

Dashboard offers visual summaries for fast decision-making, while Python allows deeper, code-driven insights.

This approach enables stakeholders to:

- Identify profitable regions and products
- Cut losses in low-margin categories
- Focus on customer-rich locations like California and NYC
- Improve pricing/discount strategies