Rajalakshmi Engineering College(An Autonomous Institution) Rajalakshmi Nagar, Thandalam- 602105

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

AD23632 - Framework for Data Visualization and Analytics

Mini Project: Store Sales Data Analysis

Report submitted by

REGISTRATION NUMBER : 231501008

STUDENT NAME : Afrah M

YEAR : 2023-2027

SUBJECT CODE : AD23632

INDEX

Chapter 1: Abstract

This project presents a comparative analysis of three distinct visual analytics frameworks by developing dashboards on the Sample Superstore dataset. The primary objective is to transform raw sales data into actionable business intelligence using Python's Matplotlib/Streamlit library, Microsoft Power BI, and Tableau. The project involves a comprehensive data preprocessing phase, followed by the design and implementation of three separate dashboards, each tailored to the strengths of the respective tool. By creating visualizations for key performance indicators (KPIs) like sales, profit, and discounts, the project uncovers critical insights into regional performance, product category profitability, and customer segments. The methodology provides a structured approach to evaluate each tool based on its data handling, visualization capabilities, interactivity, and ease of use. The final report culminates in a detailed comparative analysis, offering a clear evaluation of each framework's capabilities in a real-world business analytics scenario and providing recommendations for their optimal use cases.

Chapter 2: Introduction

In the contemporary business landscape, data is a critical asset. However, raw data in its unprocessed form offers little value. The ability to interpret this data, identify patterns, and derive meaningful conclusions is what drives strategic decision-making. This is where data visualization and business intelligence (BI) play a pivotal role. Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces, allowing users to explore complex datasets intuitively.

This project delves into the world of visual analytics by employing three of the industry's most prominent tools and frameworks on the widely-used "Sample Superstore" dataset. The goal is to answer critical business questions such as: Which product categories are most profitable? How do sales trends vary over time and across regions? What is the impact of discount strategies on profitability?

To answer these questions, we will utilize:

- **Microsoft Power BI**: A market-leading business analytics service that provides interactive visualizations with a simple, user-friendly interface.
- **Tableau**: A powerful and flexible data visualization tool known for its stunning and highly customizable charts and graphs.
- **Python (with Matplotlib & Streamlit)**: An open-source, code-driven approach that offers unparalleled customization and integration capabilities for creating bespoke web-based dashboards.

By building a functional dashboard in each of these platforms, this project aims not only to analyze the Superstore data but also to conduct a thorough comparative study of the tools

themselves, evaluating their strengths, weaknesses, and ideal applications.

Chapter 3: Dataset Description

- **Dataset Name**: Sample – Superstore Dataset
- **Source**: Kaggle / Tableau Public Superstore dataset
- **Content**: The dataset contains detailed transactional data for a fictional superstore. It includes approximately 9,994 rows and 21 columns, covering sales, products, customers, and shipping information.

Row ID	Order ID	Order Dat	Ship Date	Ship Mod	Customer	Customer	Segment	Country	City	State	Postal Co	Region	Product ID	Category	Sub-Categ	Product N	Sales	Quantity	Discount	Profit
1	CA-2016-1	########	########	Second Cl	CG-12520	Claire Gut	Consumer	United Sta	Henderso	Kentucky	42420	South	FUR-BO-1(	Furniture	Bookcases	Bush Som	261.96	2	0	41.9136
2	CA-2016-1	########	########	Second Cl	CG-12520	Claire Gut	Consumer	United Sta	Henderso	Kentucky	42420	South	FUR-CH-1(	Furniture	Chairs	Hon Delu	731.94	3	0	219.582
3	CA-2016-1	########	6/16/2016	Second Cl	DV-13045	Darrin Van	Corporate	United Sta	Los Angel	California	90036	West	OFF-LA-1(	Office Sup	Labels	Self-Adhe	14.62	2	0	6.8714
4	US-2015-1	########	10/18/201	Standard (	SO-20335	Sean O'Dc	Consumer	United Sta	Fort Laude	Florida	33311	South	FUR-TA-1(	Furniture	Tables	Bretford C	957.5775	5	0.45	-383.031
5	US-2015-1	########	10/18/201	Standard (	SO-20335	Sean O'Dc	Consumer	United Sta	Fort Laude	Florida	33311	South	OFF-ST-10	Office Sup	Storage	Eldon Fol(	22.368	2	0.2	2.5164
6	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	FUR-FU-1(	Furniture	Furnishing	Eldon Exp	48.86	7	0	14.1694
7	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	OFF-AR-1(	Office Sup	Art	Newell 32	7.28	4	0	1.9656
8	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	TEC-PH-1(	Technolog	Phones	Mitel 532(	907.152	6	0.2	90.7152
9	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	OFF-BI-10	Office Sup	Binders	DXL Angle	18.504	3	0.2	5.7825
10	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	OFF-AP-1(	Office Sup	Appliance	Belkin F5(	114.9	5	0	34.47
11	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	FUR-TA-1(	Furniture	Tables	Chromcra	1706.184	9	0.2	85.3092
12	CA-2014-1	########	6/14/2014	Standard (	BH-11710	Brosina Ho	Consumer	United Sta	Los Angel	California	90032	West	TEC-PH-1(	Technolog	Phones	Konftel 25	911.424	4	0.2	68.3568
13	CA-2017-1	4/15/2017	4/20/2017	Standard (	AA-10480	Andrew A	Consumer	United Sta	Concord	North Car	28027	South	OFF-PA-1(	Office Sup	Paper	Xerox 196	15.552	3	0.2	5.4432
14	CA-2016-1	########	########	Standard (	IM-15070	Irene Mac	Consumer	United Sta	Seattle	Washingt	98103	West	OFF-BI-10	Office Sup	Binders	Fellowes I	407.976	3	0.2	132.5922
15	US-2015-1	11/22/201	11/26/201	Standard (	HP-14815	Harold Pa	Home Off	United Sta	Fort Worth	Texas	76106	Central	OFF-AP-1(	Office Sup	Appliance	Holmes Re	68.81	5	0.8	-123.858
16	US-2015-1	11/22/201	11/26/201	Standard (	HP-14815	Harold Pa	Home Off	United Sta	Fort Worth	Texas	76106	Central	OFF-BI-10	Office Sup	Binders	Storex Du	2.544	3	0.8	-3.816
17	CA-2014-1	########	11/18/201	Standard (	PK-19075	Pete Kriz	Consumer	United Sta	Madison	Wisconsin	53711	Central	OFF-ST-10	Office Sup	Storage	Stur-D-Stc	665.88	6	0	13.3176
18	CA-2014-1	5/13/2014	5/15/2014	Second Cl	AG-10270	Alejandro	Consumer	United Sta	West Jord	Utah	84084	West	OFF-ST-10	Office Sup	Storage	Fellowes S	55.5	2	0	9.99
19	CA-2014-1	8/27/2014	########	Second Cl	ZD-21925	Zuschuss I	Consumer	United Sta	San Franci	California	94109	West	OFF-AR-1(	Office Sup	Art	Newell 34	8.56	2	0	2.4824
20	CA-2014-1	8/27/2014	########	Second Cl	ZD-21925	Zuschuss I	Consumer	United Sta	San Franci	California	94109	West	TEC-PH-1(	Technolog	Phones	Cisco SPA	213.48	3	0.2	16.011
21	CA-2014-1	8/27/2014	########	Second Cl	ZD-21925	Zuschuss I	Consumer	United Sta	San Franci	California	94109	West	OFF-BI-10	Office Sup	Binders	Wilson Jo	22.72	4	0.2	7.384

Chapter 4: Objectives

The primary goals of this project are defined as follows:

1. **To perform Exploratory Data Analysis (EDA)**: Conduct a thorough initial investigation of the Superstore dataset to understand its structure, identify key variables, check for anomalies, and prepare it for the visualization phase.
2. **To Build Interactive and Static Dashboards**: Design and implement three distinct, fully functional dashboards using Microsoft Power BI, Tableau, and Python's Matplotlib/Streamlit stack to represent the same underlying data.
3. **To Conduct a Comparative Framework Analysis**: Systematically compare the visualization capabilities, features, ease of use, interactivity levels, and limitations of each tool based on the hands-on development experience.
4. **To Derive Actionable Business Insights**: Extract and clearly articulate meaningful patterns, trends, and correlations from the visualized data to support strategic business decision-making regarding sales, profitability, inventory management, and market focus.

Chapter 5: Methodology

The methodology for this project involved a systematic approach to data handling and visualization across all three platforms.

1. Data Preprocessing & Transformation:
Before visualization, the dataset was carefully cleaned and prepared to ensure data integrity and consistency. This foundational step is crucial for accurate analysis.

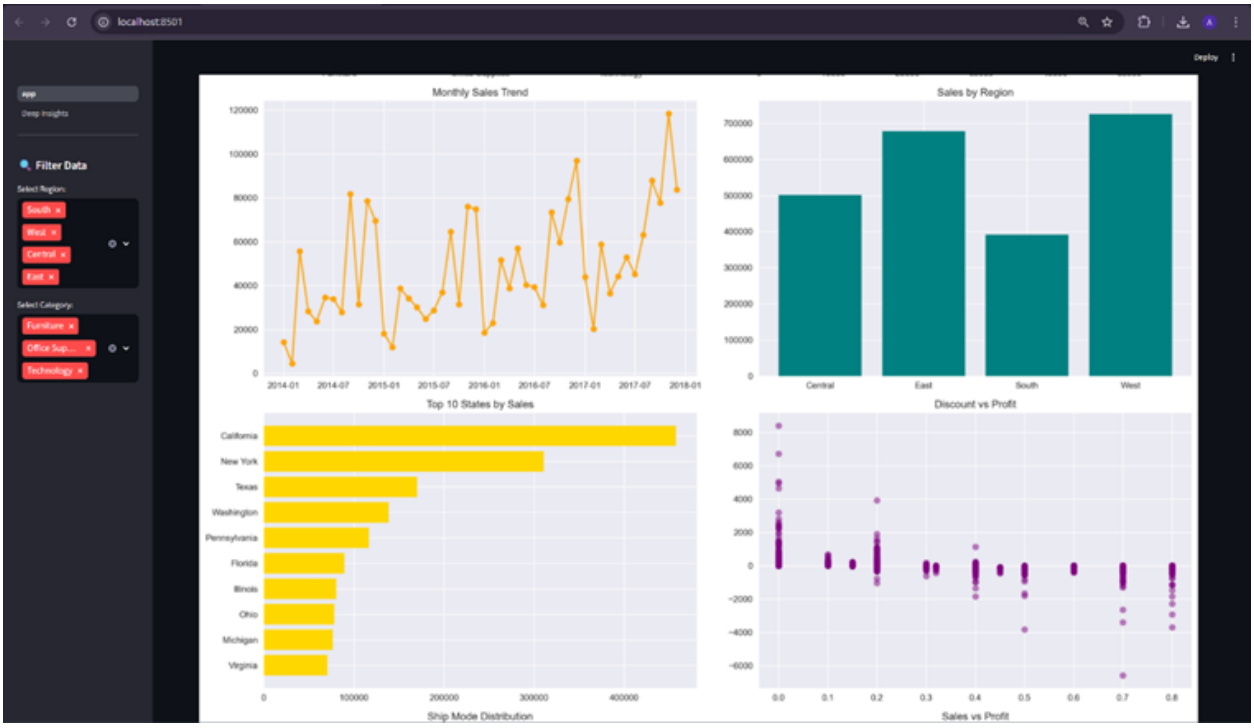
- **General Steps**: The Row ID column was removed. Data types for all columns were verified and corrected (e.g., dates, numerical figures). Column names were renamed for readability (e.g., Ship_Mode to Ship Mode).
- **Platform-Specific Transformations**:
    - **Power BI**: Using the Power Query Editor, Order Date and Ship Date were converted to Date/Time types, and Postal Code was set to Text to prevent incorrect aggregations.
    - **Tableau**: Data types were verified upon connection, and calculated fields for key metrics like Total Profit and Total Sales were created.
    - **Python (Pandas)**: The CSV was loaded into a DataFrame. The to_datetime function was used for date columns. A new Profit Margin (%) column was engineered for deeper analysis.

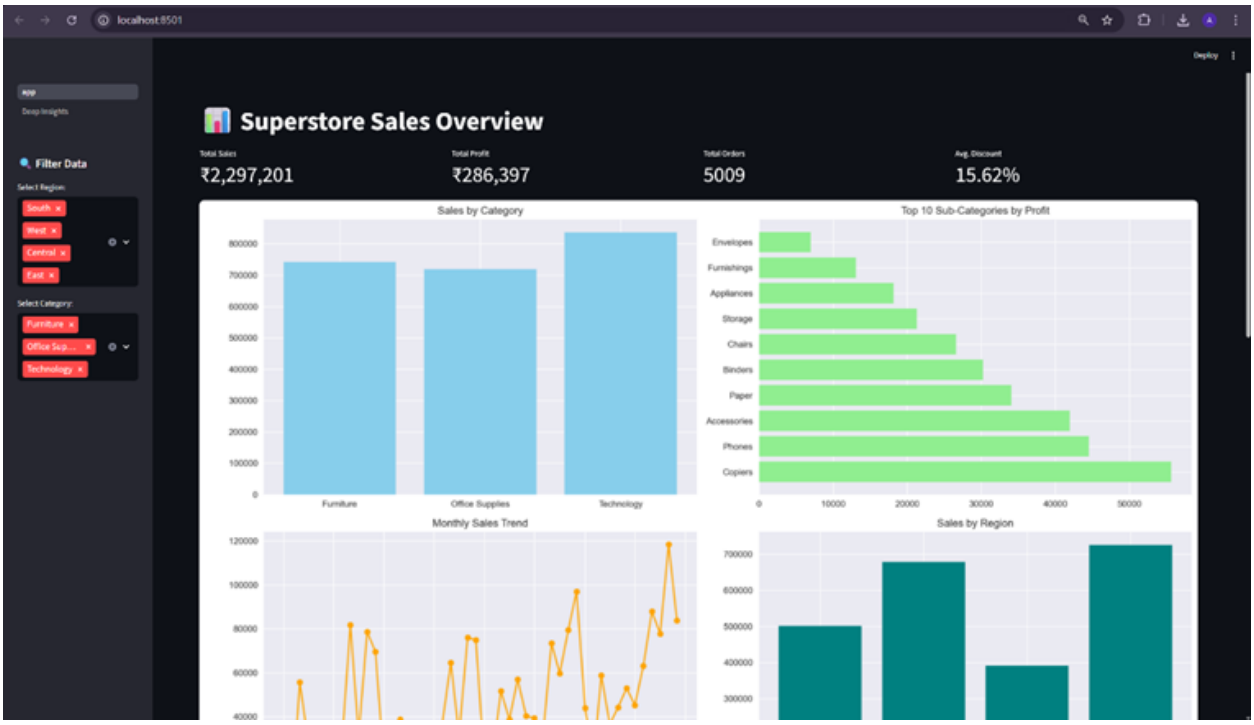
2. Dashboard Development:
Separate dashboards were designed and implemented in each tool, focusing on presenting the same core KPIs and insights through the unique features of each platform.
3. Comparative Analysis:
The final step involved evaluating each framework based on the development experience, focusing on ease of use, data handling, visualization capabilities, and interactivity.

Chapter 6: Python Implementation (Matplotlib & Streamlit)

This section details the creation of a lightweight, interactive dashboard using an open-source Python stack, showcasing a code-driven alternative to proprietary BI tools.

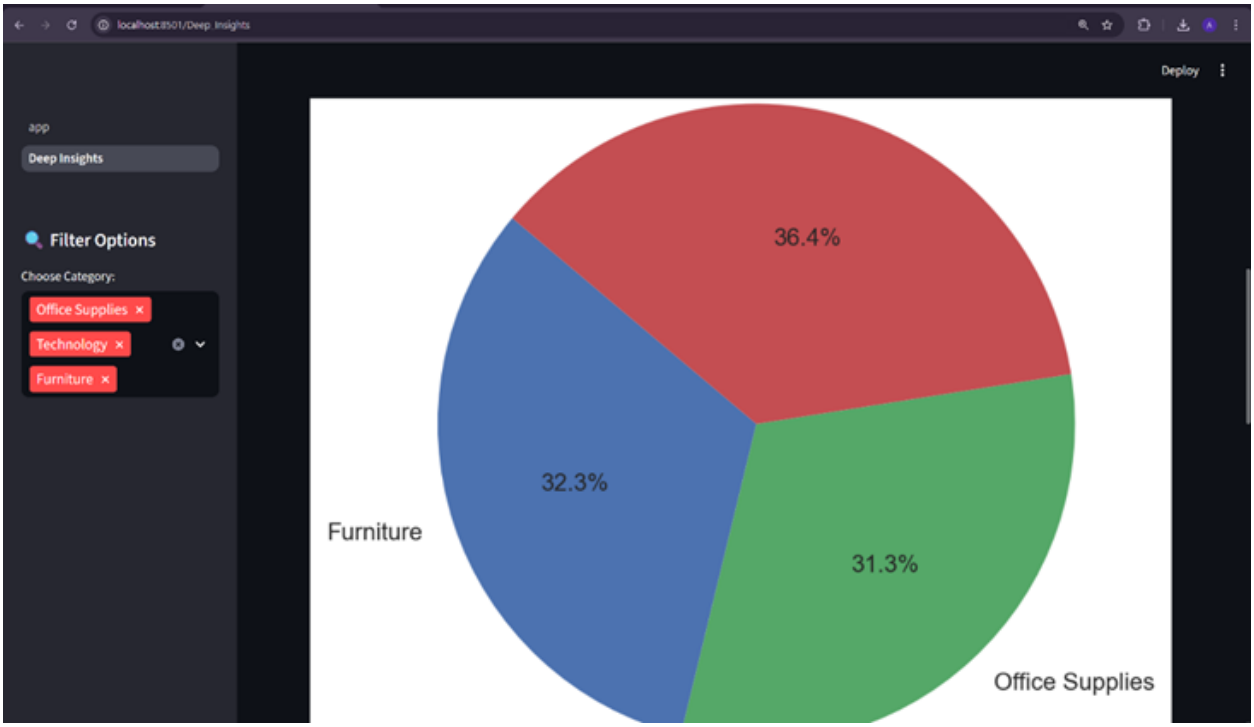
- **Framework Overview**:
  - **Streamlit**: Used to build and serve the interactive web application interface.
  - **Pandas**: Used for data manipulation and preparation.
  - **Matplotlib & Seaborn**: Used as the core libraries for generating plots and charts.
- **Dashboard Architecture**: The main application provides a high-level overview of performance with four key metrics: Total Sales, Total Profit, Total Orders, and Average Discount. Interactive sidebar filters for Region and Category were implemented.
- **Visualizations**: A grid of Matplotlib plots was created to display various insights, including Sales by Category (Bar Chart), Monthly Sales Trend (Line Chart), Discount vs. Profit (Scatter Plot), and Ship Mode Distribution (Pie Chart).

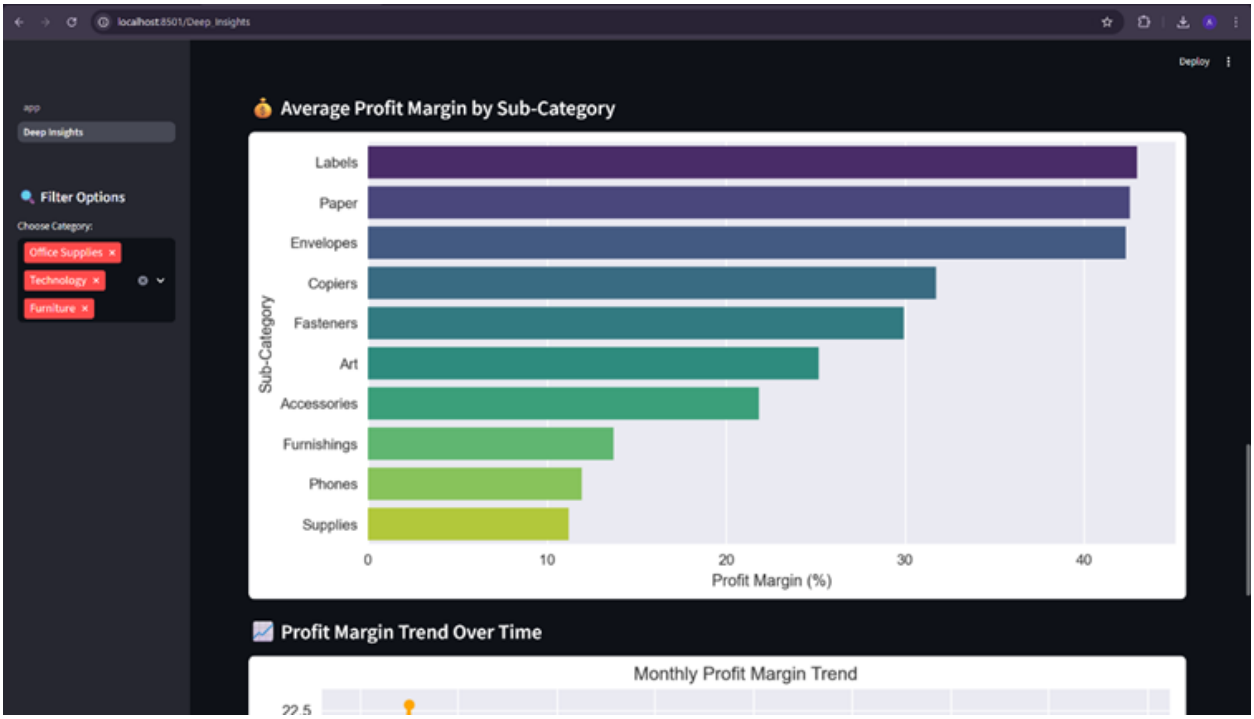


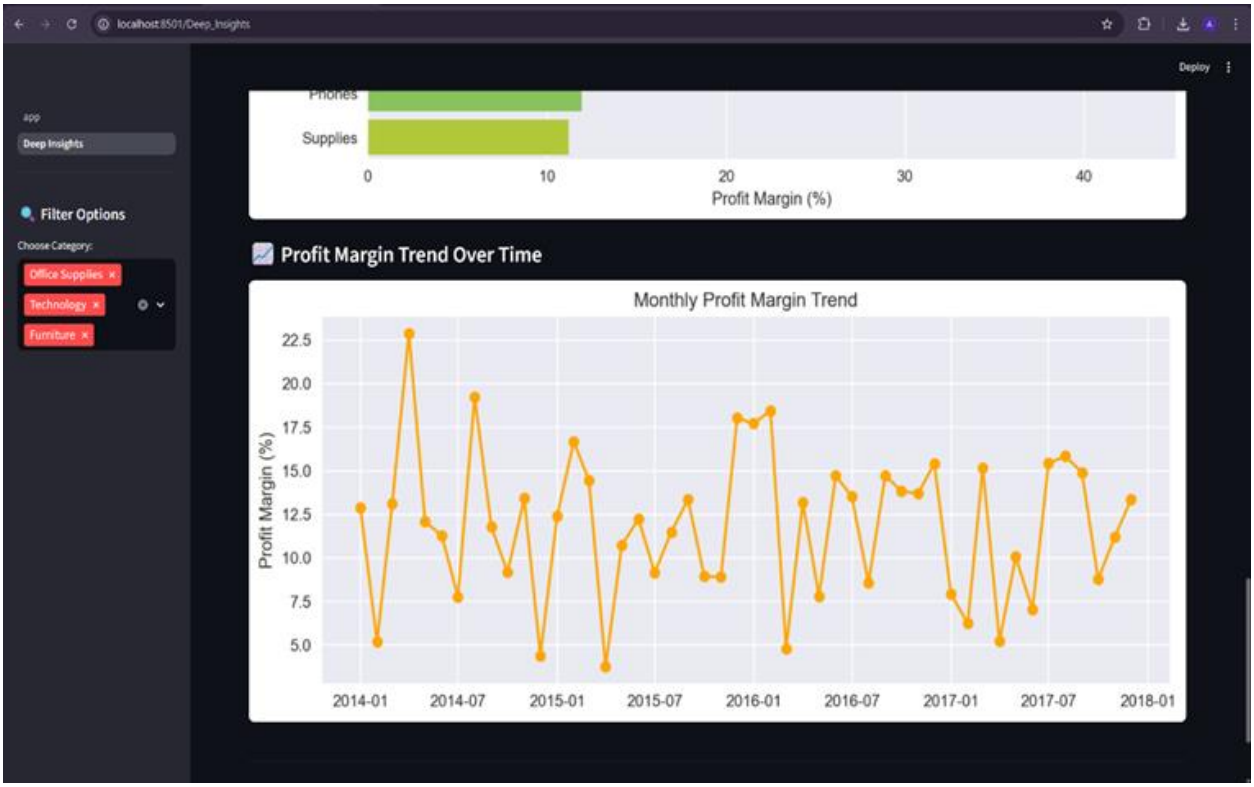

localhost:8501
Deploy
app
Deep Insights
Filter Data
Select Region:
South
West
Central
East
Select Category:
Furniture
Office Sup...
Technology
Superstore Sales Overview
Total Sales
₹2,297,201
Total Profit
₹286,397
Total Orders
5009
Avg. Discount
15.62%
Sales by Category
Furniture
Office Supplies
Technology
Top 10 Sub-Categories by Profit
Envelopes
Furnishings
Appliances
Storage
Chairs
Binders
Paper
Accessories
Phones
Copiers
Monthly Sales Trend
Sales by Region


localhost:8501
Deploy
app
Deep Insights
Filter Data
Select Region:
South
West
Central
East
Select Category:
Furniture
Office Sup...
Technology
Top 10 States by Sales
California
New York
Texas
Washington
Pennsylvania
Florida
Illinois
Ohio
Michigan
Virginia
Discount vs Profit
Ship Mode Distribution
Standard Class
59.7%
5.4%
Same Day
15.4%
First Class
19.5%
Second Class
Sales vs Profit


localhost:8501/Deep_Insights
app
Deep Insights
Filter Options
Choose Category:
Furniture
Office Sup...
Technology
Sales
Quantity
Discount
Profit
Profit Margin (%)
1.00 0.20 -0.03 0.48 0.00
0.20 1.00 0.01 0.07 -0.01
-0.03 0.01 1.00 -0.22 -0.86
0.48 0.07 -0.22 1.00 0.22
0.00 -0.01 -0.86 0.22 1.00
Sales
Quantity
Discount
Profit
Profit Margin (%)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75


localhost:8501/Deep_Insights
Deploy
app
Deep Insights
Filter Options
Choose Category:
Office Supplies
Technology
Furniture
36.4%
32.3%
31.3%
Furniture
Office Supplies


localhost:8501/Deep_Insights
Deploy
app
Deep Insights
Filter Options
Choose Category:
Office Supplies
Technology
Furniture
Average Profit Margin by Sub-Category
Labels
Paper
Envelopes
Copiers
Fasteners
Art
Accessories
Furnishings
Phones
Supplies
Sub-Category
0
10
20
30
40
Profit Margin (%)
Profit Margin Trend Over Time
Monthly Profit Margin Trend
22.5


localhost:8501/Deep_Insights
Deploy
app
Deep Insights
Filter Options
Choose Category:
Office Supplies
Technology
Furniture
Phones
Supplies
0
10
20
30
40
Profit Margin (%)
Profit Margin Trend Over Time
Monthly Profit Margin Trend
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
Profit Margin (%)
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01

Chapter 7: Power BI Dashboard

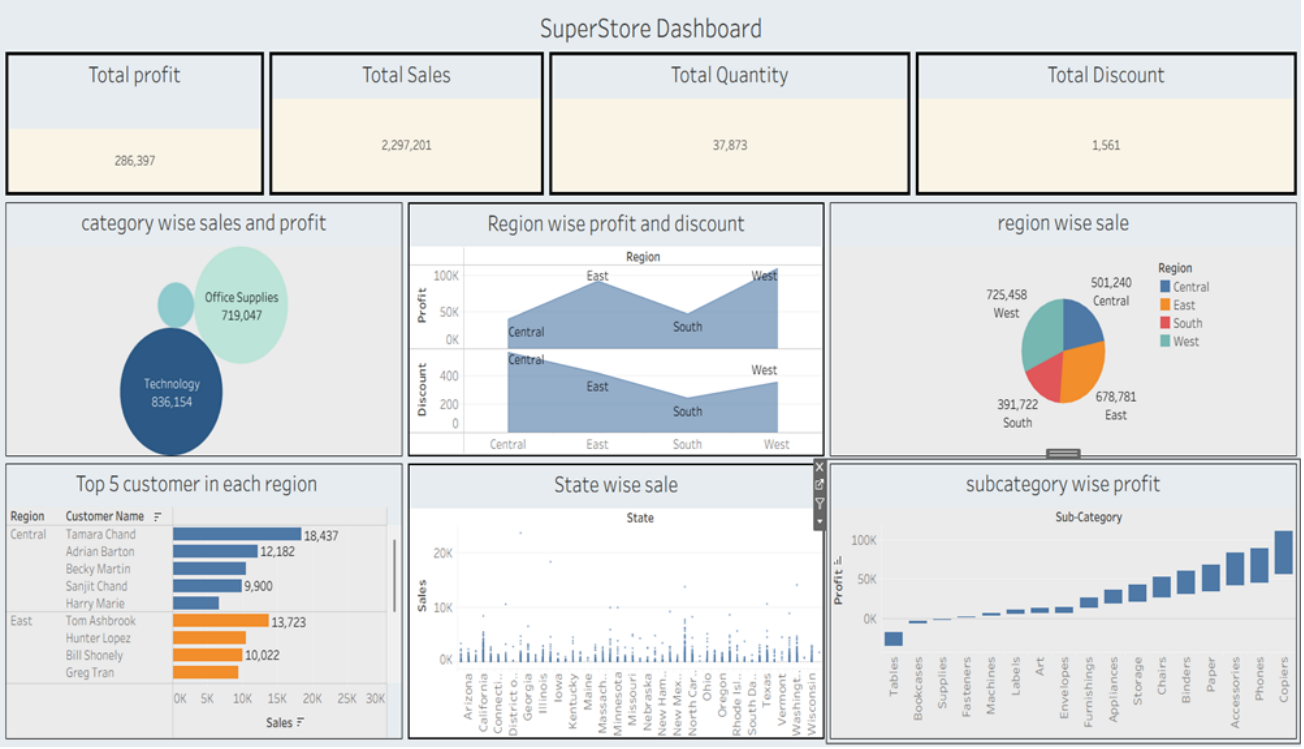
A comprehensive and fully interactive dashboard was created in Microsoft Power BI to provide a 360-degree view of the Superstore's business performance.

- **Design & Layout**: A professional dark theme was used with a consistent light-blue and orange color palette. A balanced 3x3 grid was used to arrange visuals for an intuitive user experience.
- **Dashboard Components**:
  - **KPI Cards**: Prominent cards at the top displayed Total Sales, Total Profit, Total Orders, and Average Discount.
  - **Visual Breakdown**: The dashboard included Sales by Category (Bar Chart), Sales by Year (Line Chart), Profit vs. Discount (Scatter Plot), Sales by Segment (Pie Chart), Sales by Region (Bar Chart), and a Treemap for Sales by Sub-Category.
- **Interactivity**: The dashboard is highly interactive. Clicking on a data point in one visual (e.g., a category) cross-filters all other visuals on the page. Slicers for Region and Year allow for easy drill-down analysis.

Chapter 8: Tableau Dashboard

A visually engaging and analytical dashboard was developed in Tableau, focusing on interactivity and deep-dive analysis capabilities.

- **Design & Layout**: A soft pastel theme was chosen for a clean and professional appearance, with a consistent color scheme of blues and oranges. The layout guides the user from high-level KPIs to granular breakdowns.
- **Dashboard Components**:
  - **KPI Cards**: Four primary KPIs were displayed: Total Profit, Total Sales, Total Quantity, and Total Discount.
  - **Visual Breakdown**: The dashboard combined several worksheets, including a Category-wise Sales and Profit (Bubble Chart), Region-wise Profit and Discount (Area Chart), Top 5 Customers in Each Region (Bar Chart), and Subcategory-wise Profit (Bar Chart).
- **Interactivity**: The dashboard is rich with interactive features. Hovering over any data point reveals a detailed tooltip. Filters for region and category were implemented, allowing users to dynamically slice the data.



Chapter 9: Analysis & Findings

This project successfully achieved all its stated objectives. Three fully functional analytics dashboards were developed, proving the value of visual analytics in translating complex data into clear, digestible, and interactive visualizations. The analysis across all three dashboards consistently revealed the following key business insights:

- **Technology is the Most Profitable Category**: While Office Supplies had high sales volume, Technology (specifically Copiers and Phones) yielded the strongest profits.
- **The West Region is the Top Performer**: The West region consistently led in both sales and profit, suggesting that successful strategies from this region could be replicated elsewhere.
- **High Discounts Erode Profit Margins**: A clear negative correlation was observed between discounts and profit. Discounts above 20-30% consistently led to financial losses.
- **The Consumer Segment Drives Sales**: This segment accounted for over half of the total sales, identifying it as the backbone of the business.
- **Standard Shipping is Dominant**: Approximately 60% of all orders used Standard Class shipping, highlighting an area for logistical optimization.

Chapter 10: Conclusion

Through this project, we have explored how different visualization tools and frameworks can be leveraged to extract meaningful intelligence from a standard business dataset. The hands-on experience of building dashboards in Matplotlib, Power BI, and Tableau provided invaluable insights into the practical challenges and advantages of each platform, from data preparation to final presentation.

It became evident that while GUI-based tools like Power BI and Tableau democratize data analysis and accelerate the path from data to insight, programmatic libraries like Matplotlib offer ultimate flexibility for bespoke solutions. This project not only enhanced technical skills in these platforms but also strengthened analytical thinking by focusing on the 'so what?' behind the data. Ultimately, this project underscores the critical role of visual analytics in modern business, enabling organizations to move beyond raw data and make agile, informed, and data-driven decisions.

Chapter 11: Future Scope

- **Predictive Analytics**: The historical sales data could be used to build a forecasting model to predict future sales, which could then be integrated into the dashboards.
- **Advanced Interactivity**: More advanced features like drill-through pages in Power BI or set actions in Tableau could be implemented for even deeper analysis.

- **Real-Time Data**: The dashboards could be connected to a live database to reflect real-time sales data, turning them into operational monitoring tools.

Chapter 12: Appendix (Code Snippets)

Key Python code snippets used for data transformation and visualization.

1. Data Loading and Preparation (using Pandas)

```
import pandas as pd

# Loading Data
df = pd.read_csv("Sample - Superstore.csv", encoding='latin1')

# Converting Dates
df['Order Date'] = pd.to_datetime(df['Order Date'])
df['Ship Date'] = pd.to_datetime(df['Ship Date'])

# Creating New Columns
df['Profit Margin (%)'] = (df['Profit'] / df['Sales']) * 100
```

2. Visualization Examples (using Matplotlib)

```
# Sales by Category (Bar Chart)
cat_sales = df_filtered.groupby('Category')['Sales'].sum()
axes[0,0].bar(cat_sales.index, cat_sales.values, color='skyblue')
axes[0,0].set_title('Sales by Category')

# Monthly Sales Trend (Line Chart)
monthly = df_filtered.groupby(df_filtered['Order Date'].dt.to_period('M'))['Sales'].sum()
axes[1,0].plot(monthly.index.to_timestamp(), monthly.values, marker='o', color='orange')
axes[1,0].set_title('Monthly Sales Trend')

# Discount vs. Profit (Scatter Plot)
axes[2,1].scatter(df_filtered['Discount'], df_filtered['Profit'], alpha=0.5, color='purple')
axes[2,1].set_title('Discount vs Profit')

# Ship Mode Distribution (Pie Chart)
ship = df_filtered['Ship Mode'].value_counts()
axes[3,0].pie(ship.values, labels=ship.index, autopct='%1.1f%%')
axes[3,0].set_title('Ship Mode Distribution')
```