**Project Report  
  
  
Project Title: Visual Analytics Framework using Matplotlib, Power BI, and Tableau**

**Student Name: Devesh D, Afrah M.**

**Roll Number: 231501034, 231501008.**

**Course: Framework for Data and Visual Analytics**

**Subject Code: AD23632**

**Department: ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

**Institution Name: RAJALAKSHMI ENGINEERING COLLEGE**

**Academic Year: 2025-2026**

**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.  
  
  
**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:

1. **Microsoft Power BI**: A market-leading business analytics service that provides interactive visualizations with a simple, user-friendly interface.
2. **Tableau:** A powerful and flexible data visualization tool known for its stunning and highly customizable charts and graphs.
3. **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.  
  
  
**Objectives:**

The primary goals of this project are defined as follows:

* **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.
* **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.
* **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.
* **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.

**Dataset Description:** 1. **Dataset Name:** Sample – Superstore Dataset

2. **Source:** Kaggle / Tableau Public Superstore dataset

3. **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.



**Data Preprocessing & Transformation:**

Before visualization, the dataset was carefully cleaned and prepared to ensure data integrity and consistency across all three platforms. This foundational step is crucial for accurate analysis.

**General Preprocessing Steps:**

* **Column Removal:** The Row ID column was removed as it serves as a unique identifier with no analytical value for aggregation or trend analysis.
* **Data Type Verification:** All columns were checked to ensure they were assigned the correct data type. This is especially important for dates and numerical figures.
* **Column Renaming:** Columns with names like Ship\_Mode were renamed to a more readable format, Ship Mode, for better presentation in visuals.

**Platform-Specific Transformations:**

**1. In Power BI (using Power Query Editor):**

* **Changing Data Types:**
  + Order Date and Ship Date were converted from text to the Date/Time type.
  + Postal Code was converted to Text to prevent erroneous mathematical aggregations.
  + Sales, Profit, Discount, and Quantity were confirmed as Decimal/Number types.
* **Error and Null Value Check:** The dataset was scanned for any null values or inconsistencies before being loaded into the Power BI workspace.

2. **In Tableau:**

* **Data Type Verification**: Upon connecting to the data source, Tableau automatically assigns data types. These were manually verified to ensure Order Date was a date, Region was a geographical role, and numerical fields were correctly identified.
* **Calculated Fields:** Custom calculations for key metrics were created to facilitate analysis, such as Total Profit and Total Sales.

3. **In Python (using Pandas):**

* The raw CSV file was loaded into a Pandas DataFrame. The following code snippets from app.py were used for transformation:
  + Loading Data: df = pd.read\_csv("Sample - Superstore.csv", encoding='latin1')
  + Converting Dates: The to\_datetime function was used to correctly parse the date columns, which is essential for time-series analysis.

df['Order Date'] = pd.to\_datetime(df['Order Date'])

df['Ship Date'] = pd.to\_datetime(df['Ship Date'])

* + Creating New Columns: For deeper analysis in the "Deep Insights" page, a Profit Margin (%) column was engineered.

df['Profit Margin (%)'] = (df['Profit'] / df['Sales']) \* 100

**Visual Analytics using Matplotlib & Streamlit:**This section details the creation of a lightweight, interactive dashboard using an open-source Python stack. This approach showcases 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 static plots and charts.

**Dashboard Architecture (app.py):** The main application is built to provide a high-level overview of the Superstore's performance.

* **KPIs:** At the top, four key metrics provide an instant summary:
  + Total Sales
  + Total Profit
  + Total Orders
  + Average Discount
* **Interactive Filters:** A sidebar was implemented to allow users to filter the entire dashboard by Region and Category, enabling dynamic exploration of the data.
* **Visualizations Grid:** A 4x2 grid of Matplotlib plots was created to display various insights simultaneously.

**Code Walkthrough and Visuals:**

1. **Sales by Category (Bar Chart):** Compares total sales across the three main product categories.

**Python**

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')

1. **Monthly Sales Trend (Line Chart):** Shows sales growth patterns over time.

**Python**

monthly = df\_filtered.groupby(df\_filtered['Order Date'].dt.to\_period('M'))['Sales'].sum()

axes[1,0].plot(monthly.index, monthly.values, marker='o', color='orange')

axes[1,0].set\_title('Monthly Sales Trend')

1. **Discount vs. Profit (Scatter Plot):** Visualizes the relationship between discounts and profit margins.

**Python**

axes[2,1].scatter(df\_filtered['Discount'], df\_filtered['Profit'], alpha=0.5, color='purple')

axes[2,1].set\_title('Discount vs Profit')

1. **Ship Mode Distribution (Pie Chart):** Highlights customer preferences for shipping methods.

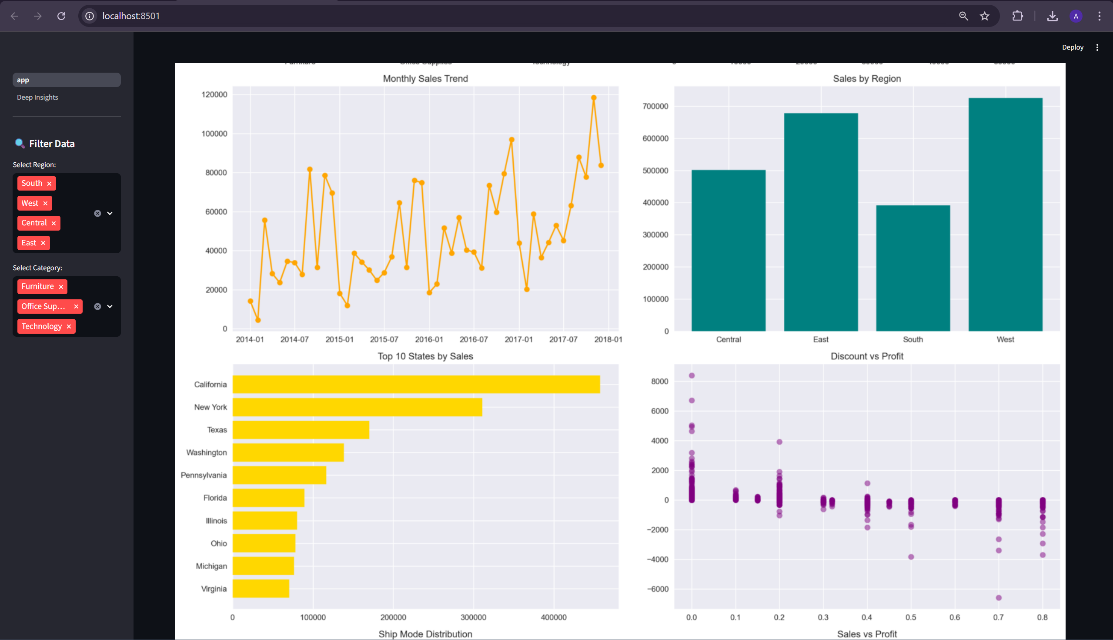
**Python**

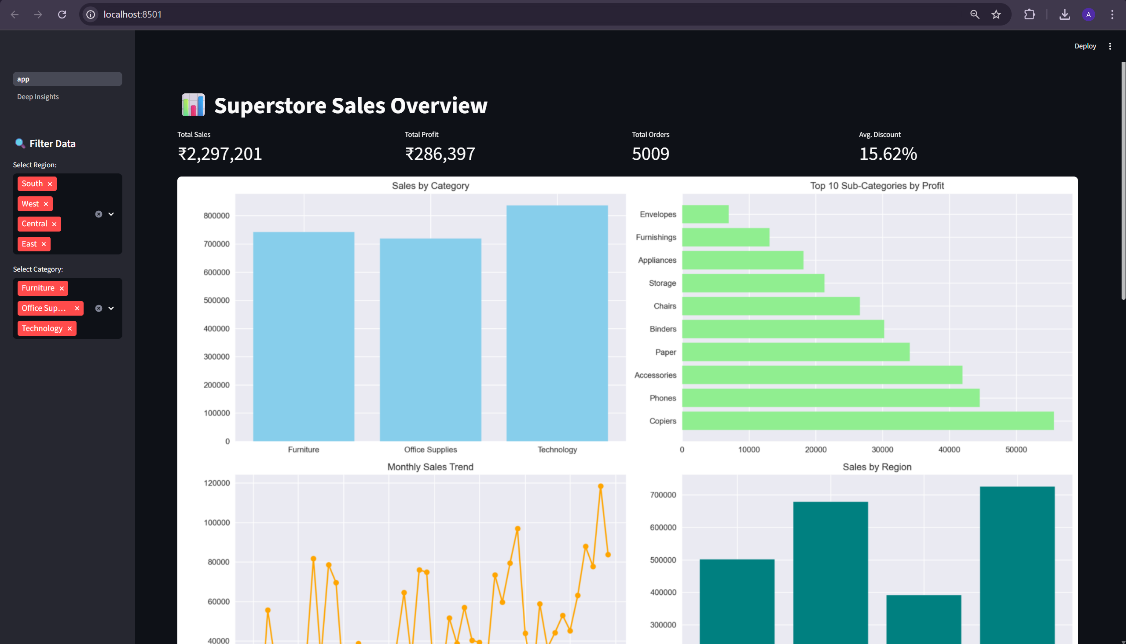
ship = df\_filtered['Ship Mode'].value\_counts()

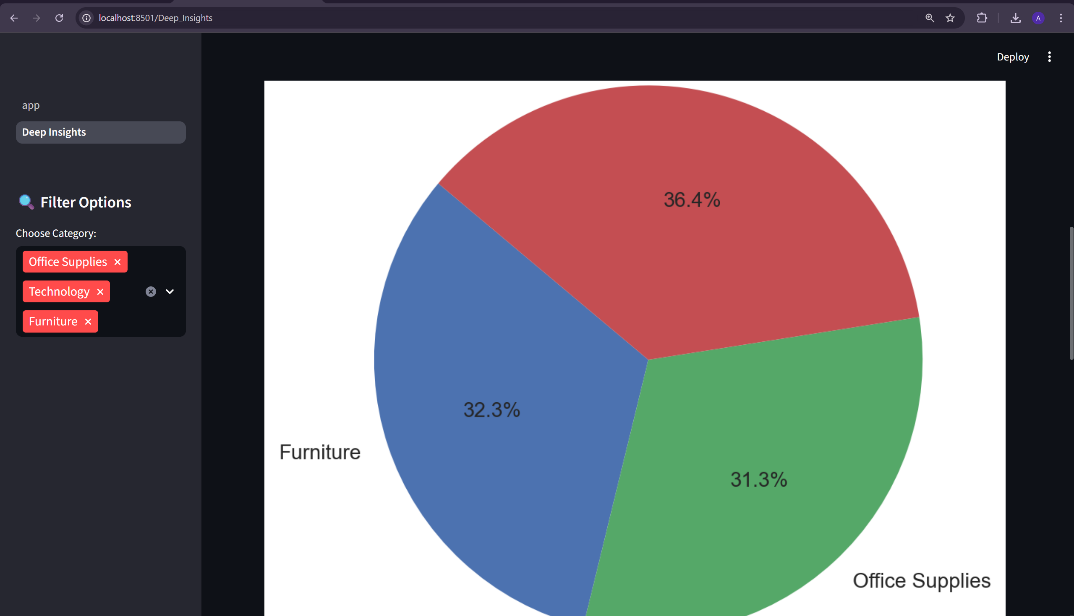
axes[3,0].pie(ship.values, labels=ship.index, autopct='%1.1f%%')

axes[3,0].set\_title('Ship Mode Distribution')

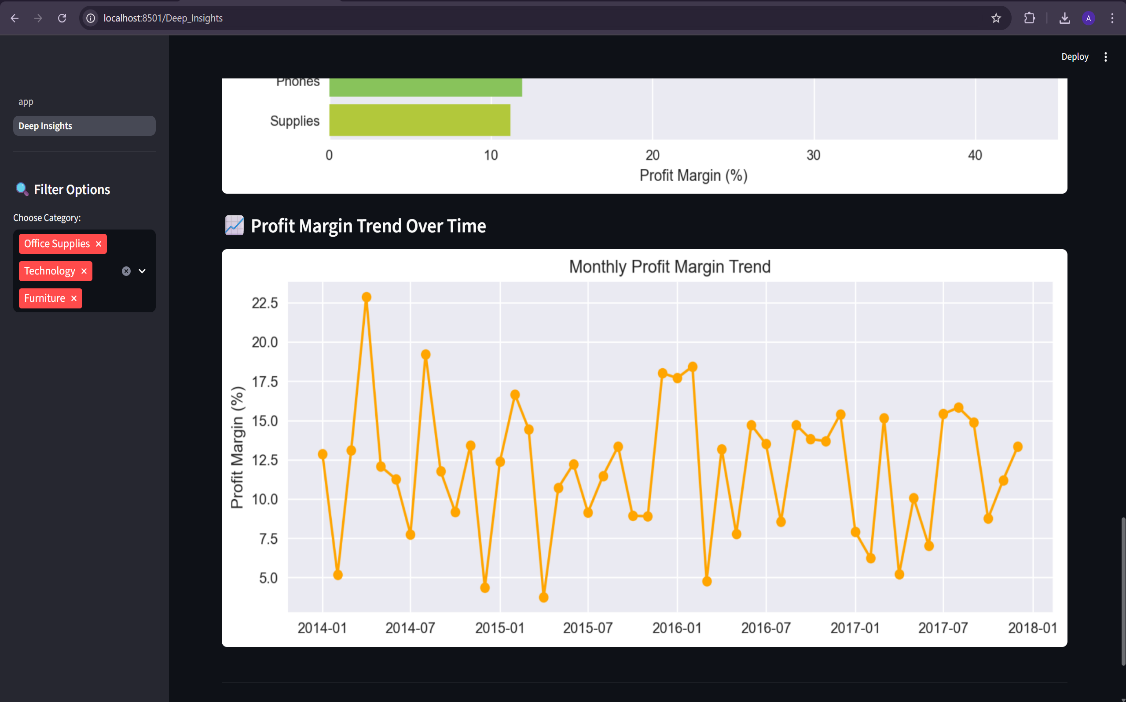
Output:











**Dashboard in Power BI:**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:**

* **Theme:** A professional dark gradient background was used to make the visuals stand out.
* **Color Palette:** A consistent theme of light-blue and orange tones was applied for a uniform look.
* **Layout:** A balanced 3x3 grid was used to arrange the visuals, ensuring a harmonious and intuitive user experience.

**Dashboard Components:**

**1. KPI Cards:** Four prominent KPI cards were placed at the top of the dashboard for an instant summary of the most critical metrics:

* **Total Sales:** $2.30M
* **Total Profit:** $286.40K
* **Total Orders:** 5.009K
* **Average Discount:** 15.6% These cards were formatted with gradient shading and bold typography for emphasis.

**2. Detailed Visual Breakdown:**

* **Sales by Category (Bar Chart):** This visual was used to compare total sales across Office Supplies, Furniture, and Technology, quickly identifying the highest revenue-generating category.
* **Sales by Year (Line Chart):** This chart displays the sales growth trend from 2014 to 2017, providing insight into the business's year-over-year performance.
* **Profit vs. Discount (Scatter Plot):** This crucial visual reveals the impact of discounting on profitability, showing a clear trend where higher discounts often lead to reduced or negative profit.
* **Sales by Segment (Pie Chart):** This chart shows the distribution of sales among Consumer, Corporate, and Home Office customers, highlighting the most significant customer segment.
* **Sales by Region (Bar Chart):** Displays and compares performance across the four regional markets: West, East, Central, and South.
* **Treemap (Sales by Category and Sub-Category):** This provides a hierarchical view, illustrating which sub-categories contribute the most to each main category's sales.

**Interactivity:** The dashboard is highly interactive. Users can click on a category in one chart (e.g., "Technology"), and all other visuals on the page will automatically cross-filter to show data relevant only to that category. Slicers for Region and Year were also added to allow for easy drill-down analysis.



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

**Design & Layout:**

* Theme: A soft pastel background was chosen for a clean and professional appearance.
* Color Scheme: A consistent palette of blues and oranges was used to easily differentiate between regions and categories across different charts.
* Layout: The dashboard was structured to guide the user from high-level KPIs at the top to more granular breakdowns below.

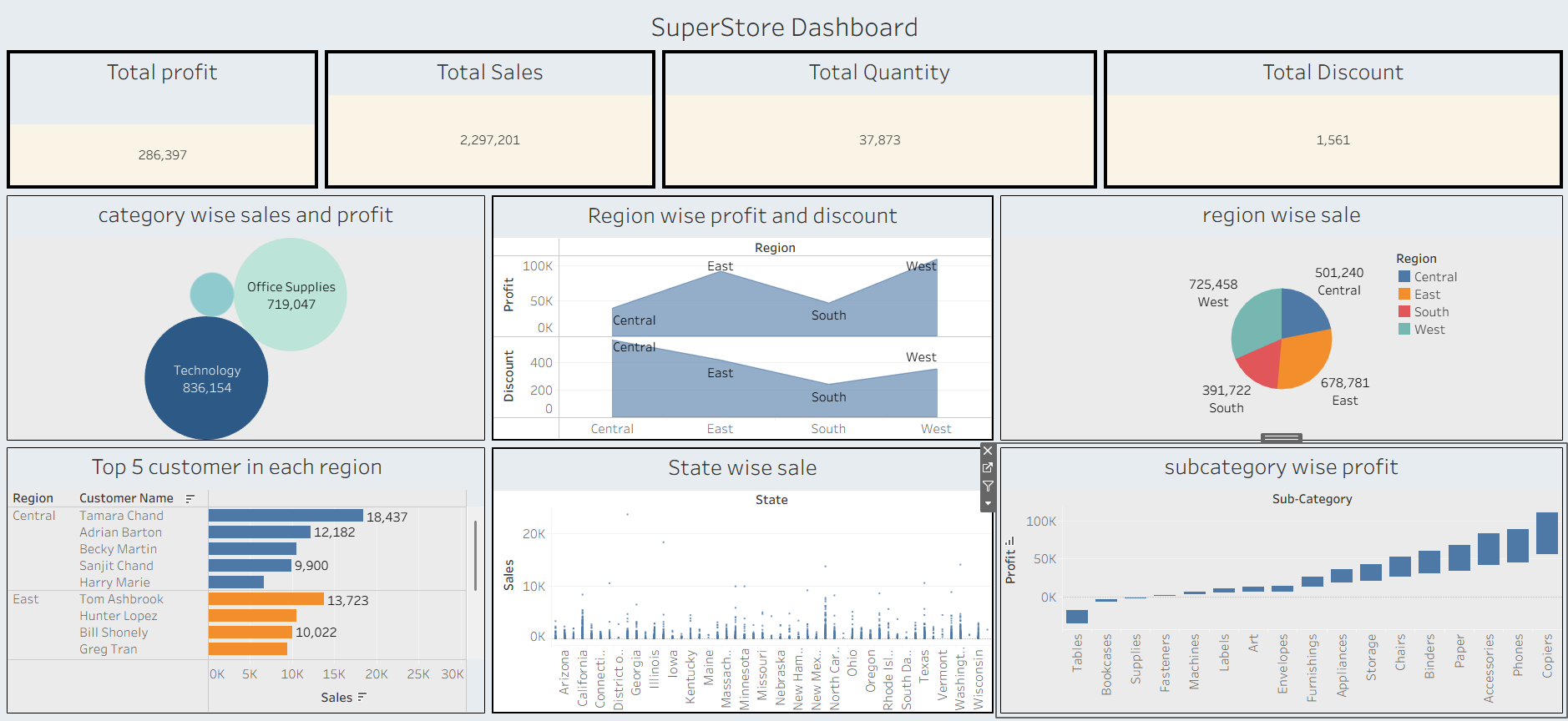
**Dashboard Components:**

**1. KPI Cards**: Four primary KPIs were displayed to provide an at-a-glance view of overall performance:

* Total Profit: $286,397
* Total Sales: $2,297,201
* Total Quantity: 37,873
* Total Discount: $1,561

2. **Detailed Visual Breakdown:** The dashboard was constructed by combining several worksheets, each containing a specific visual.

* **Category-wise Sales and Profit (Bubble Chart):** This chart effectively compares sales (represented by bubble size) and profit (represented by color intensity) across the major product categories.
* **Region-wise Profit and Discount (Area Chart):** This visual shows how profit and discount levels vary together across different regions, helping to identify regional pricing strategies.
* **Region-wise Sales (Pie Chart):** Displays the simple sales distribution across the Central, East, South, and West regions.
* **Top 5 Customers in Each Region (Bar Chart):** A more advanced chart that highlights the top-performing customers within every region, useful for sales and marketing focus.
* **Subcategory-wise Profit (Bar Chart):** Identifies the most and least profitable product subcategories, pinpointing specific items like Copiers and Phones as high performers.

**Interactivity:** The Tableau dashboard is rich with interactive features. Hovering over any data point reveals a detailed tooltip with precise figures. Furthermore, filters for region and category were implemented, allowing users to dynamically slice the data. Using a region on the map as a filter, for example, updates all other charts to reflect data for only that selected region.  
  
  
  
  
  
  


**Insights & Observations:**The analysis across all three dashboards consistently revealed the following key business insights:

1. **Technology is the Most Profitable Category:**
   * Observation: While Office Supplies generated high sales volume, the Technology category yielded the strongest profits. The "Profit by Sub-Category" charts in all three dashboards clearly show that Copiers and Phones are the most profitable items.
   * Business Implication: The company should focus marketing efforts and ensure healthy inventory levels for high-margin technology products.
2. **The West Region is the Top Performer:**
   * Observation: The "Sales by Region" charts consistently show the West region as the best-performing area in terms of both sales and profit.
   * Business Implication: Business strategies that are successful in the West could potentially be replicated in other regions. The Central region, which often shows low profitability, may require a strategic review.
3. **High Discounts Erode Profit Margins:**
   * Observation: The scatter plots visualizing Profit vs. Discount in all three dashboards show a clear negative correlation. Discounts above a certain threshold (around 20-30%) consistently lead to financial losses on sales.
   * Business Implication: A revised discount strategy is urgently needed. The company should consider setting caps on discounts for certain product categories or regions to protect profitability.
4. **The Consumer Segment Drives Sales:**
   * Observation: The "Sales by Segment" pie charts revealed that the Consumer segment accounted for over half of the total sales.
   * Business Implication: This segment is the backbone of the business. Marketing campaigns, loyalty programs, and customer service efforts should be heavily focused on retaining and growing this customer base.
5. **Standard Shipping is Dominant:**
   * Observation: Standard Class shipping was the preferred method for approximately 60% of all orders.
   * Business Implication: Logistics and supply chain operations must be optimized for Standard Class shipping to ensure efficiency and customer satisfaction. Offering small incentives to use this cost-effective method could further improve margins.

**Results:** This project successfully achieved all its stated objectives. Three fully functional analytics dashboards were developed using Matplotlib, Power BI, and Tableau, each providing a valid and insightful view of the Superstore dataset. These dashboards effectively translate complex transactional data into clear, digestible, and interactive visualizations, proving the value of visual analytics. The KPIs and detailed charts provide a holistic view of business performance, highlighting areas of strength (e.g., West region, Technology category) and opportunities for strategic improvement (e.g., discount policy, Central region profitability). The side-by-side development and comparison process has provided a robust framework for evaluating these tools for future projects.  
  
  
 **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.

**Future Work:**

* **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.

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.