TASK 2

We have taken the dataset Superstore.csv and used the tableau tool for data visualization.

This dataset contains transactional sales data, ideal for analysis of business performance, shipping timelines, and regional trends. Below is a breakdown of the columns included:

Column Name	Data Type	Description
Order Date	Date/Time	The date when the customer placed the order.
Ship Date	Date/Time	The date when the order was shipped to the customer.
Sales	Numerical	The total dollar amount of the sale.
Quantity	Numerical	The number of units sold in the order.
Discount	Numerical	The discount applied to the order (0–1, where 0.2 = 20% off).
Profit	Numerical	The profit earned from the order (Sales - Costs).
Category	Categorical	High-level product category (e.g., Furniture, Office Supplies).
Sub-Category	Categorical	More detailed product classification.
Region	Categorical	The geographical region where the sale was made.
Segment	Categorical	Customer segment (e.g., Consumer, Corporate, Home Office).
State	Categorical	The U.S. state where the customer is located.
City	Categorical	The city where the customer is located.
Ship Mode	Categorical	The type of shipping selected (e.g., First Class, Standard).

The dataset captures information on customer purchases, including order dates, shipment timelines, product categories, and financial metrics like sales and profit. It reflects realistic business scenarios and can be used for tasks like:

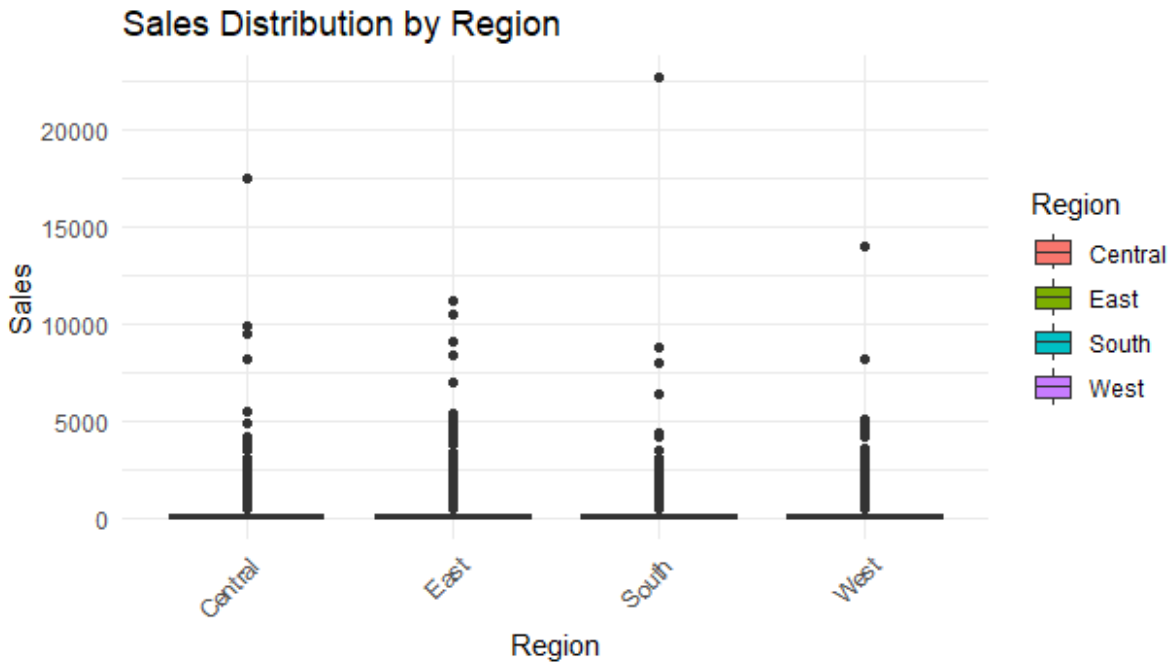
- Sales forecasting
- Customer segmentation
- Shipping performance analysis
- Profitability analysis

- Data visualization practice

Here are some ideas for how to use this dataset:

- **Profitability Analysis:** Identify which product categories and regions are most profitable.
- **Shipping Efficiency:** Analyze the gap between Order Date and Ship Date to evaluate logistics.
- **Geographical Trends:** Visualize sales by State, City, or Region on a map.
- **Predictive Modeling:** Build models to predict Profit or Ship Mode based on other features.
-

```
# Install ggplot2 if you haven't already

# install.packages("ggplot2")

library(ggplot2)



# Load your dataset

data <- read.csv("your_dataset.csv")



# Create a boxplot: Sales by Region

ggplot(data, aes(x = Region, y = Sales, fill = Region)) +

  geom_boxplot() +

  theme_minimal() +

  labs(title = "Sales Distribution by Region",

       x = "Region",

       y = "Sales") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Possible Extended Visualizations

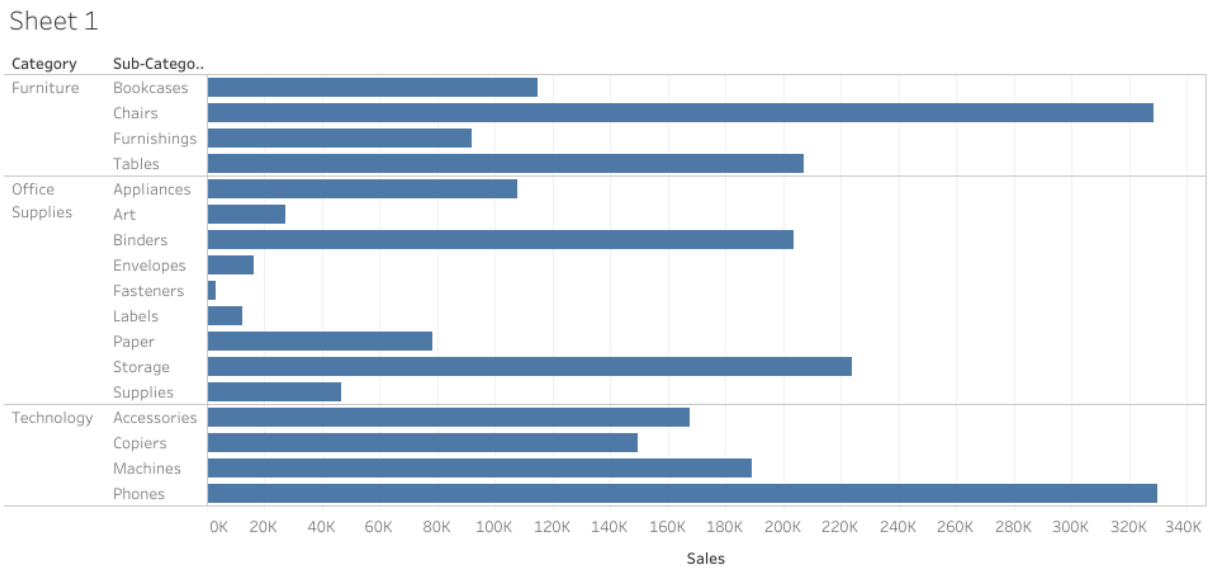
1. **Profit Trends Over Time**
   Line plot of total monthly or quarterly profits.
2. **Geographical Heatmap**
   State-wise sales and profit analysis.
3. **Correlation Matrix**
   Explore relationships between numerical variables (Sales, Quantity, Discount, Profit).
4. **Top Categories by Profit**
   Bar chart of sub-categories ranked by total profit.

The “right” chart really depends on what you're trying to analyze! Here are a few common scenarios:

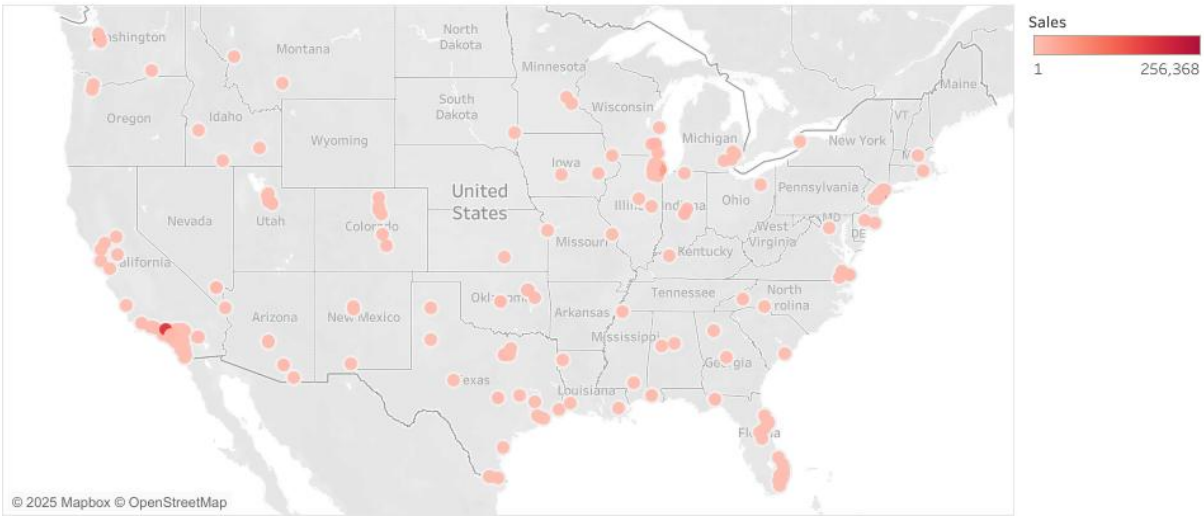
1. **Sales or Profit over time →**
   □ **Line Chart** (good for trends).
2. **Comparison between categories →**
   □ **Bar Chart** (e.g., Sales by Category, Profit by Region).
3. **Profit vs. Discount relationship →**
   □ **Scatter Plot** (to spot correlation patterns).
4. **Market share or distribution →**
   □ **Pie Chart** (for proportions — e.g., Sales share by Segment).
5. **State or Region-based data →**
   □ **Map/Choropleth** (geographical insight).

6. **Sales/Profit contribution across multiple dimensions →**
   Treemap (for a hierarchy view, like Category → Sub-Category → Sales).

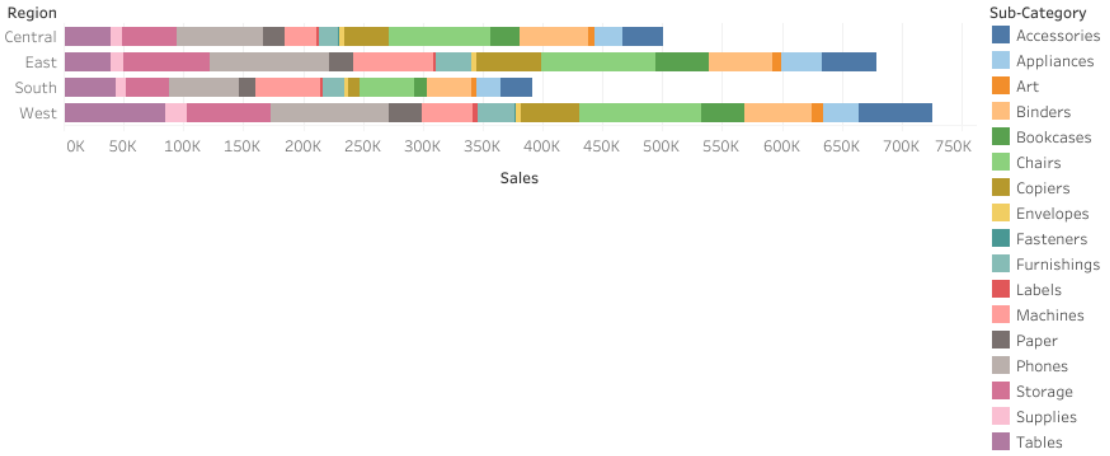
Profit over Time	**Line Chart**	One line per region, max 3–5. Use subtle color contrast.
Sales Distribution by State	**Map (Single Color Gradient)**	Use one color (e.g., Blue) with intensity for sales volume.
Discount vs. Profit Correlation	**Scatter Plot**	Use small dots, neutral colors, size by quantity if needed.
Top 10 Products by Sales	**Simple Bar Chart**	Sort descending. Avoid bright colors; use soft grayscale.
Shipping Mode Share	**Donut Chart (if <4 categories)**	Limit to 2–3 colors only.



Sheet 3



Sheet 5



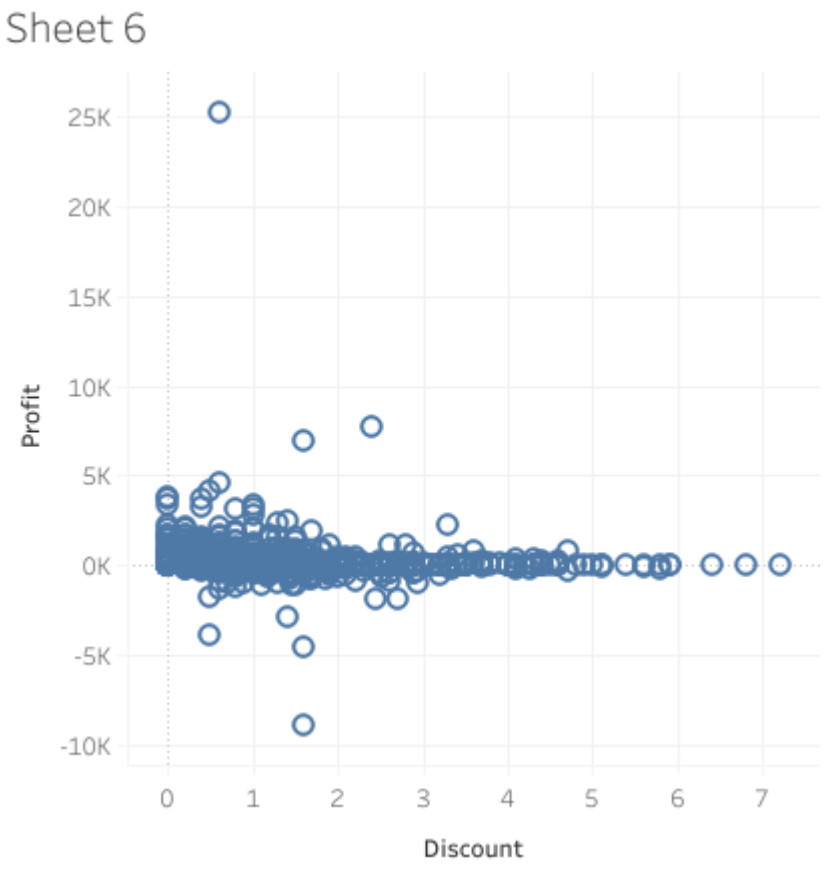
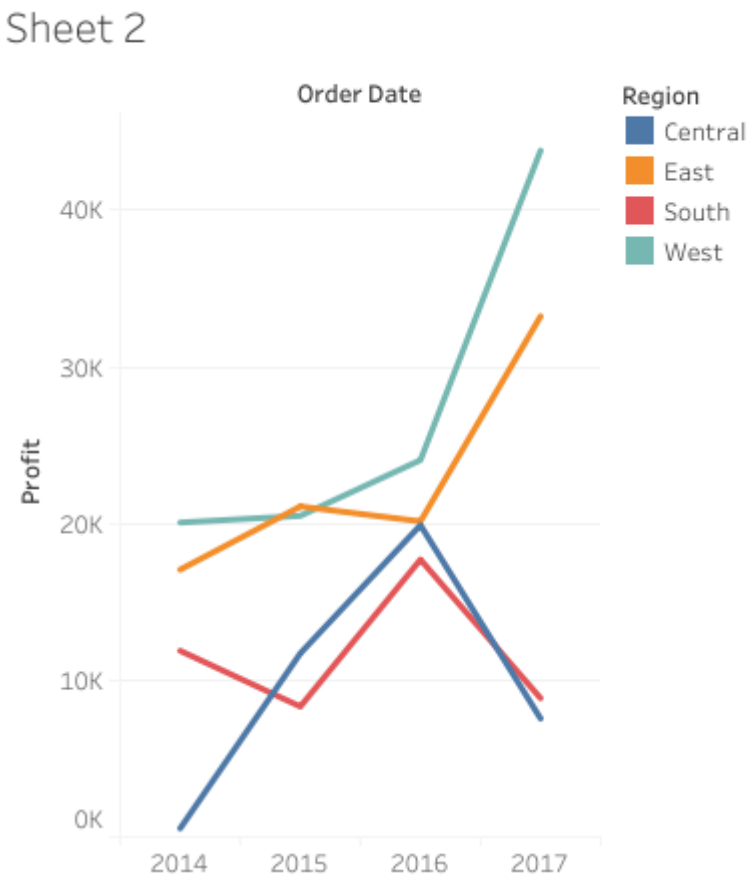


Chart Type	Data Example	Context to Add	How to Apply in Tableau
□ **Sales by Category (Bar Chart)**	Sales grouped by `Category`	"This shows which product categories drive total sales and help identify focus areas for marketing."	Use **Title:** “Which categories drive our revenue?” Add **Tooltip:** `Category, Sales, % of Total`
□ **Profit Over Time (Line Chart)**	Profit by `Order Date (Month)`	"This shows how profits fluctuate month to month, helping spot seasonality or irregular spikes."	Add **Reference Line**: Avg Profit Add note: "Q4 spikes likely due to holiday sales"
□ **Sales by State (Map)**	Sales by `State`	"This map highlights geographical distribution, so regional sales teams can adjust focus."	Use **Legend Title:** “Total Sales per State” Simplify to 1-color gradient.
□ **Discount vs Profit (Scatter Plot)**	`Discount` vs `Profit` for each order	"This plot reveals the impact of discounting on profitability, especially for over-discounted orders."	Add annotation: "Negative profits for discounts >30%"

Chart Type	Data Example	Context to Add	How to Apply in Tableau
□ **Top 10 Products by Sales (Bar)**	Sales by `Product Name` (Top 10)	"This list identifies the top-selling products for inventory and marketing priorities."	**Chart Title:** “Top 10 Revenue-Generating Products” **Label Bars:** Sales Value



Takeaways for Your Superstore Data

1. **Sales by Category**
   - T*akeaway:* "Technology accounts for 45% of total sales — it outperforms Furniture and Office Supplies."
2. **Profit Over Time**
   - *Takeaway:* "Profit dips sharply in Q2 2020 — likely due to high discounts."

3. **Discount vs Profit Scatter Plot**
    - *Takeaway:* "Orders with discounts above 30% often generate negative profit."
4. **Sales by State (Map)**
    - *Takeaway:* "California is the largest contributor to sales, followed by New York and Texas."