INDIANA UNIVERSITY
BLOOMINGTON

INDIANA UNIVERSITY BICENTENNIAL

INFO-I 590 DATA VISUALIZATION

# Analyzing Retail Sales Performance Across the USA Regions

Final Presentation


- Priyanka Prem Kumar
- Sumedh Ambapkar

# TEAM MEMBERS

**Priyanka Prem Kumar**
prpremk@iu.edu

**Sumedh Ambapkar**
suambapk@iu.edu

# INTRODUCTION

- Analyzes sales data from a U.S. office supply retailer (9,994 transactions, 49 states, 3 product categories).
- Focuses on data-driven decision-making in a competitive retail landscape.
- Dataset includes sales, discounts, profits, shipping details, and customer segments.
- Key focus areas:
  - Product performance (sales vs. profit)
  - Shipping efficiency and delays
  - Customer segment behavior
  - Regional sales trends
- Uses advanced visualizations (contour plots, chord diagrams) for deeper insights.
- Aims to turn complex data into clear, strategic recommendations.

# OBJECTIVE

Our Goal is to:

- Apply descriptive and visual analytics to a U.S. office supply sales dataset.
- Identify high- and low-performing products based on sales and profit.
- Analyze how discounts affect profitability by category and sub-category.
- Evaluate shipping and order timelines to assess operational efficiency.
- Segment customers (Consumer, Corporate, Home Office) by behavior and profitability.
- Assess regional performance through geographic sales and profit analysis.

We aim to:

- Uncover patterns where few products or customers drive most revenue.
- Identify inefficiencies and regional trends for better resource allocation.
- Support strategic decisions in pricing, logistics, marketing, and customer engagement.

# DATASET AND DATA PREPROCESSING

- Dataset includes sales, orders, and shipping data across regions, states, and categories
- Cleaned missing values and standardized date formats for consistency
- Engineered features like shipping delay and extracted order month
- Aggregated sales and profit data for regional and product-level insights
- Extracted time-based features such as year, month, and weekday
- Final dataset prepared for advanced visualizations and modeling

# METHODOLOGY and TOOLS USED

- **Tool**: Used ObservableHQ for interactive D3 visualizations, Power BI for business dashboards, and Python libraries like Plotly, Seaborn, and Matplotlib for advanced analytics and visual storytelling
- **Platform**: Developed and executed the project in Google Colab with integrated notebook support for real-time visualization
- **Data Handling**: Preprocessed and aggregated data using Pandas and NumPy for transformation, cleaning, and feature engineering
- **Visualization Strategy**: Mapped multi-level hierarchies and trends using Sunburst, Treemap, Choropleth maps, and bar charts
- **Objective**: To identify regional and product-level performance patterns based on sales, profit, and shipping delay metrics
- **Design Approach**: Created both static and interactive visuals to compare KPIs across regions, states, and categories
- **Output**: Generated dynamic charts for stakeholder presentations and decision-making dashboards across platforms

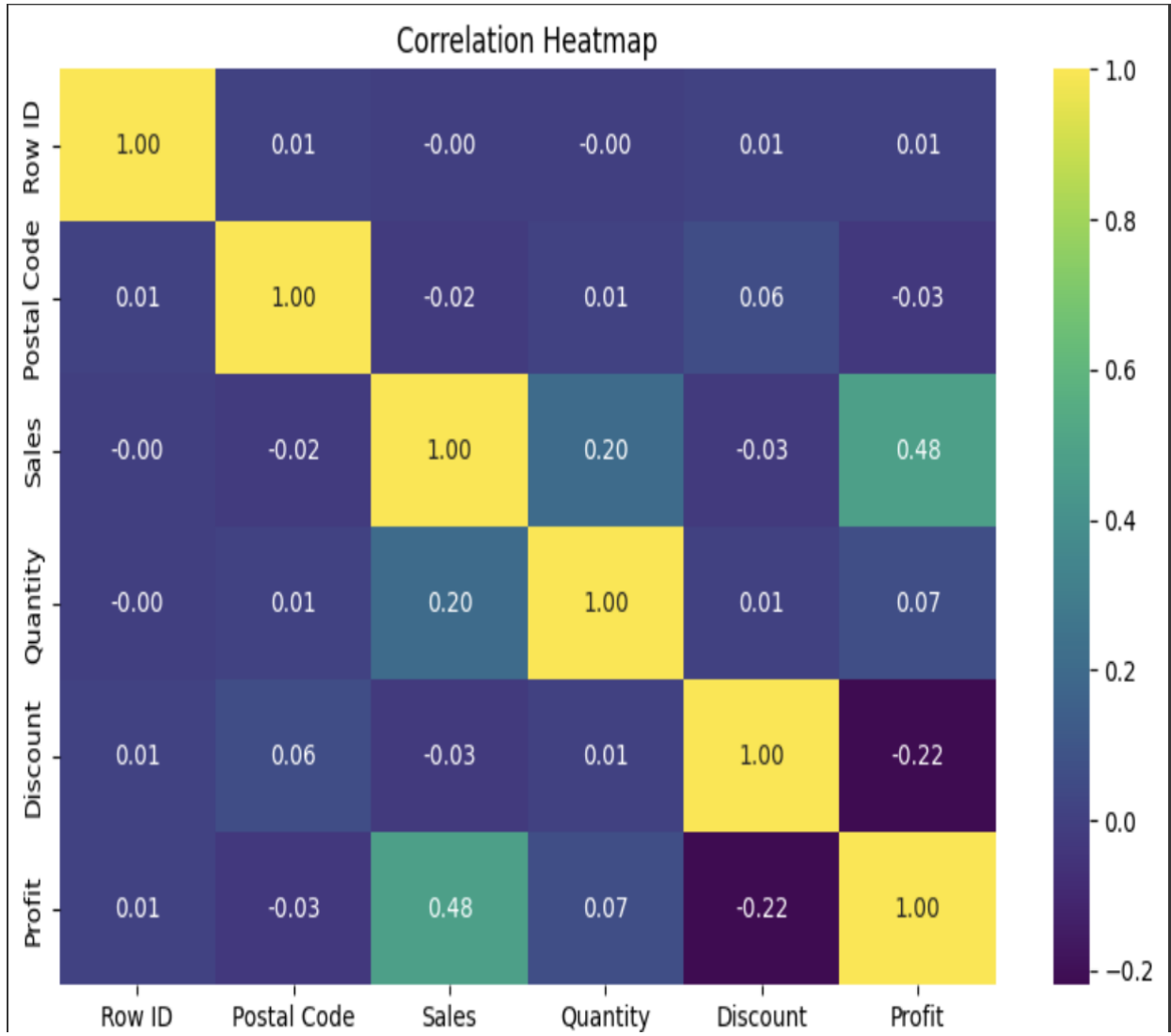
# EDA

➢**Top Metrics Identified:** Sales, profit, discount, quantity, and shipping delay were among the most informative numerical features, influencing regional and category-level performance patterns

➢ **Temporal Analysis:** Shipping delays peaked during end-of quarter months, with a clear time-to-ship impact on customer satisfaction and profitability metrics

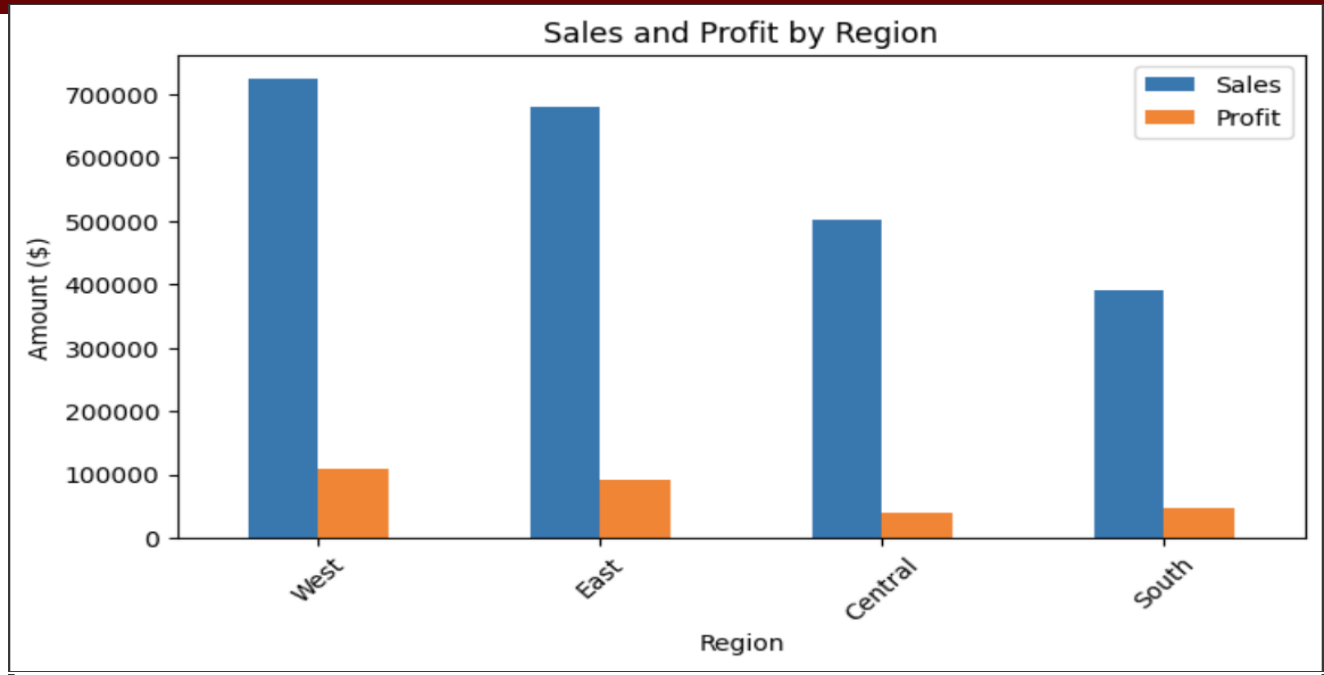
➢**Category-Level Variation:** Technology segment had the highest average profit per order, while Furniture showed higher variance in profit margins due to inconsistent discounting strategies

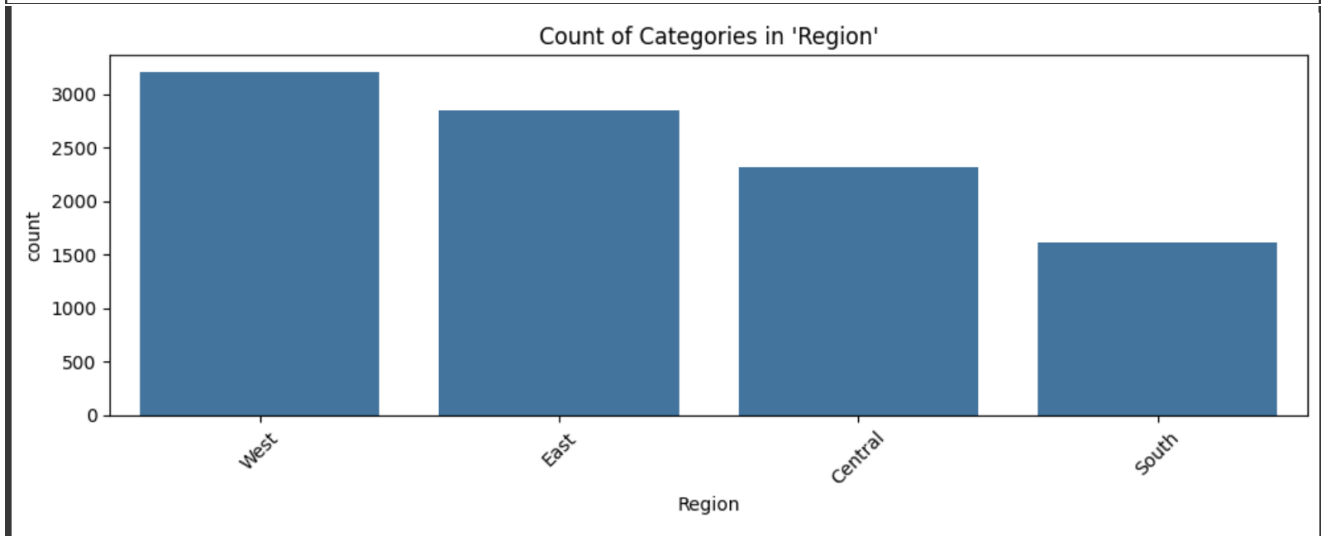
➢**Correlation Structure:** Pairwise correlation matrix revealed:
- Strong positive correlation between sales and profit
- Negative correlation between discount and profit
- Weak/no correlation between quantity and profit, indicating volume ≠ profitability



# EDA

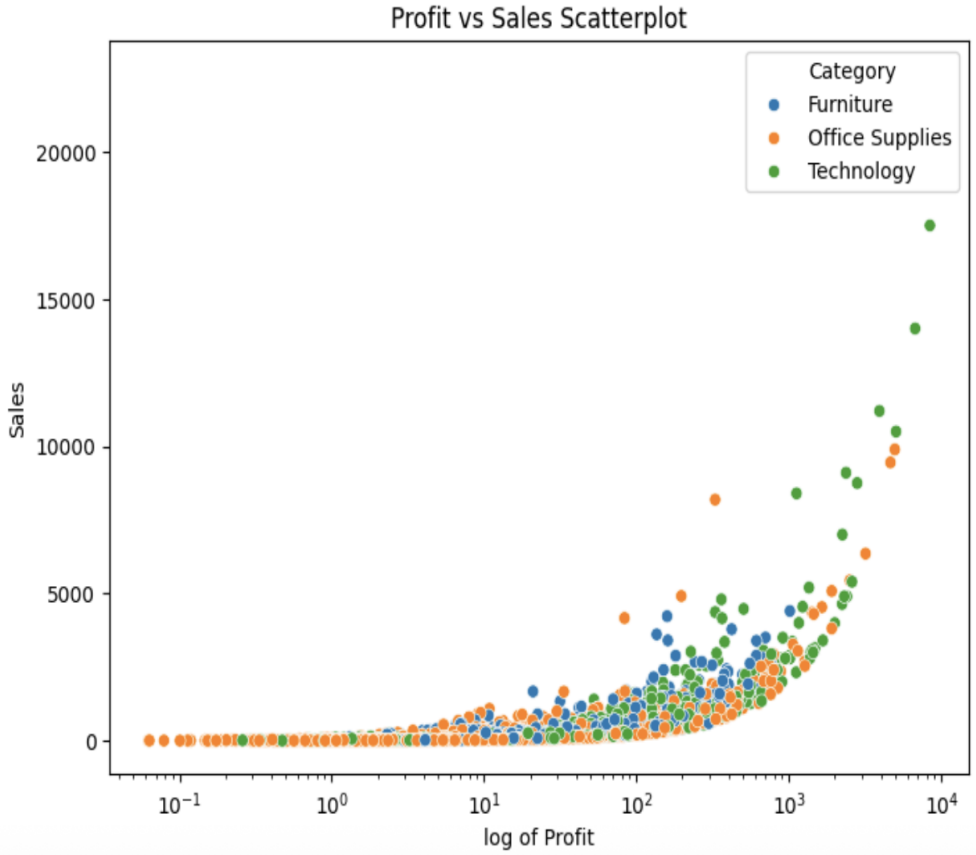
- **Regional Trends:**
  - Western region dominated in total sales and average order profit
  - Central region had higher shipping delays and lower profit per sale
  - States like California and New York consistently outperformed others in both volume and margin
- **Outlier Detection:**
  - Data revealed significant profit outliers in the Technology category
  - Shipping delays exceeding 15 days were isolated and flagged for operational review
- **Missing Data Handling:**
  - Minimal missingness observed; date fields and state data were cleaned and standardized using pd.to_datetime and string normalization
- **Feature Importance for Modeling:**
  - Delay, discount, region, and category were prioritized for predictive modeling and visualization pipelines due to their high variance and influence



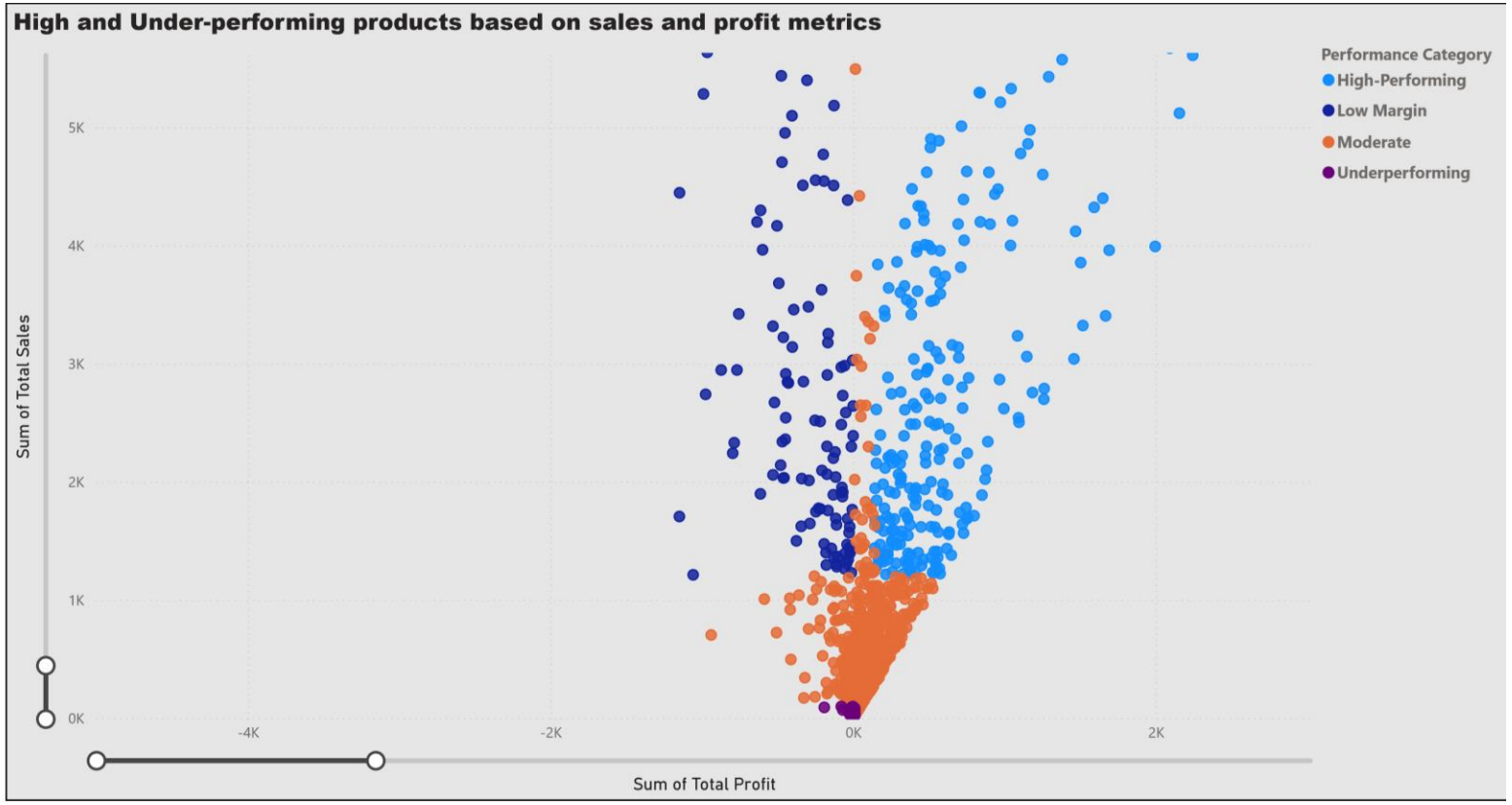
# Utilizing descriptive and visual analytics to analyze a U.S. office supply retail dataset.


Bar Chart for showcasing distribution if sales and profit over different region

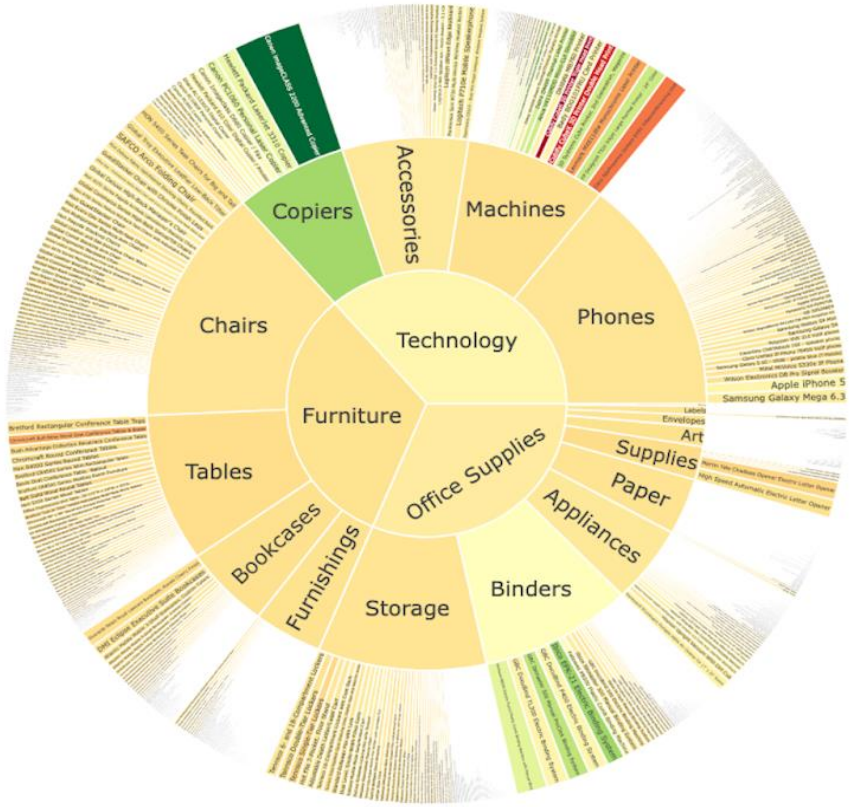



Scatterplot showcasing distribution of sales and profit in the dataset over different categories

# RESULTS - Identifying High-Performing and Underperforming Products Using Sales and Profit Metrics

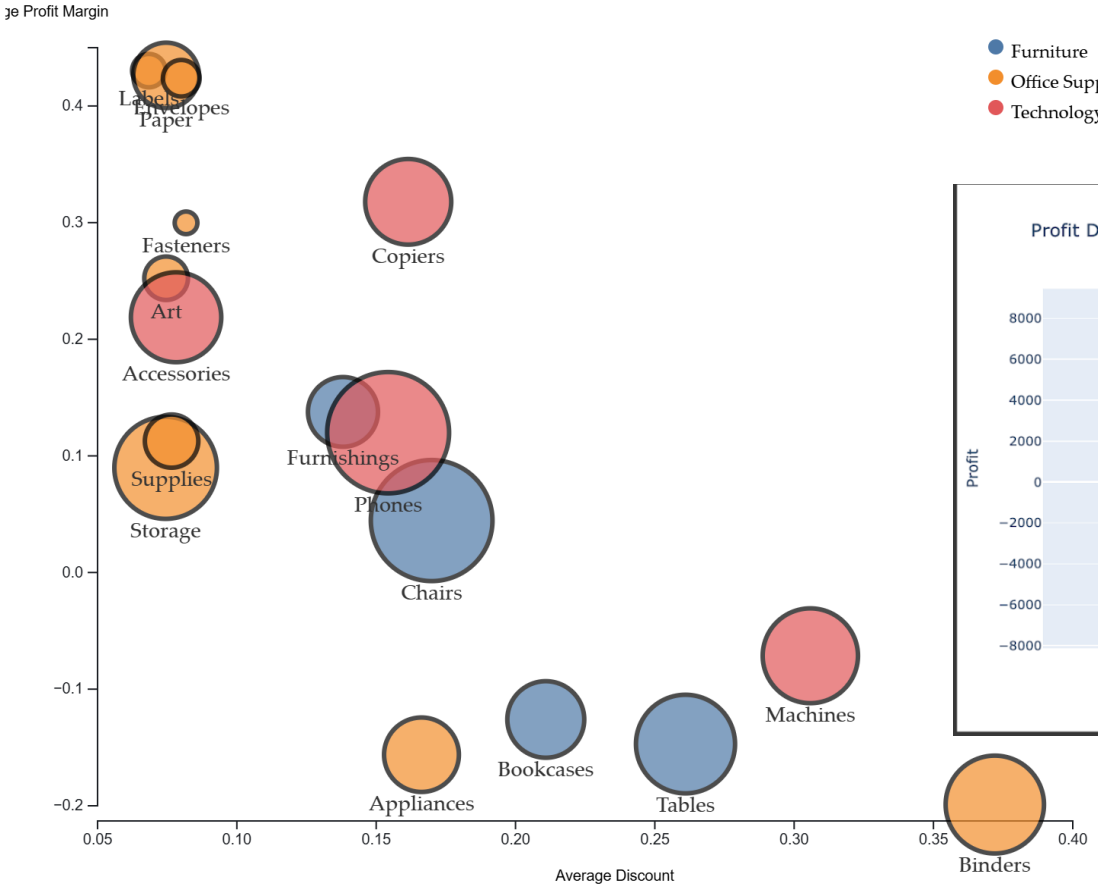


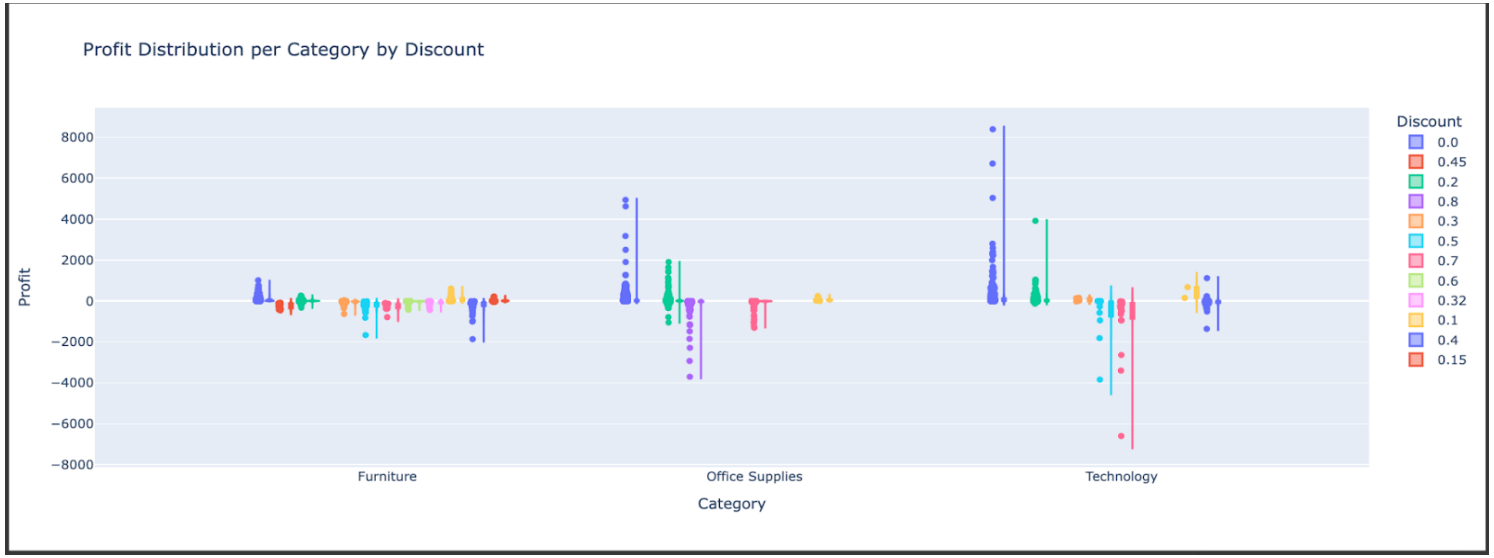

Scatterplot visualizing product performance, with total sales and total profit, products were color-coded based on their performance category




Sunburst Chart to visualize Sales and Profit analysis by Product Hierarchy

# Evaluating the impact of discounts on profitability across product categories.

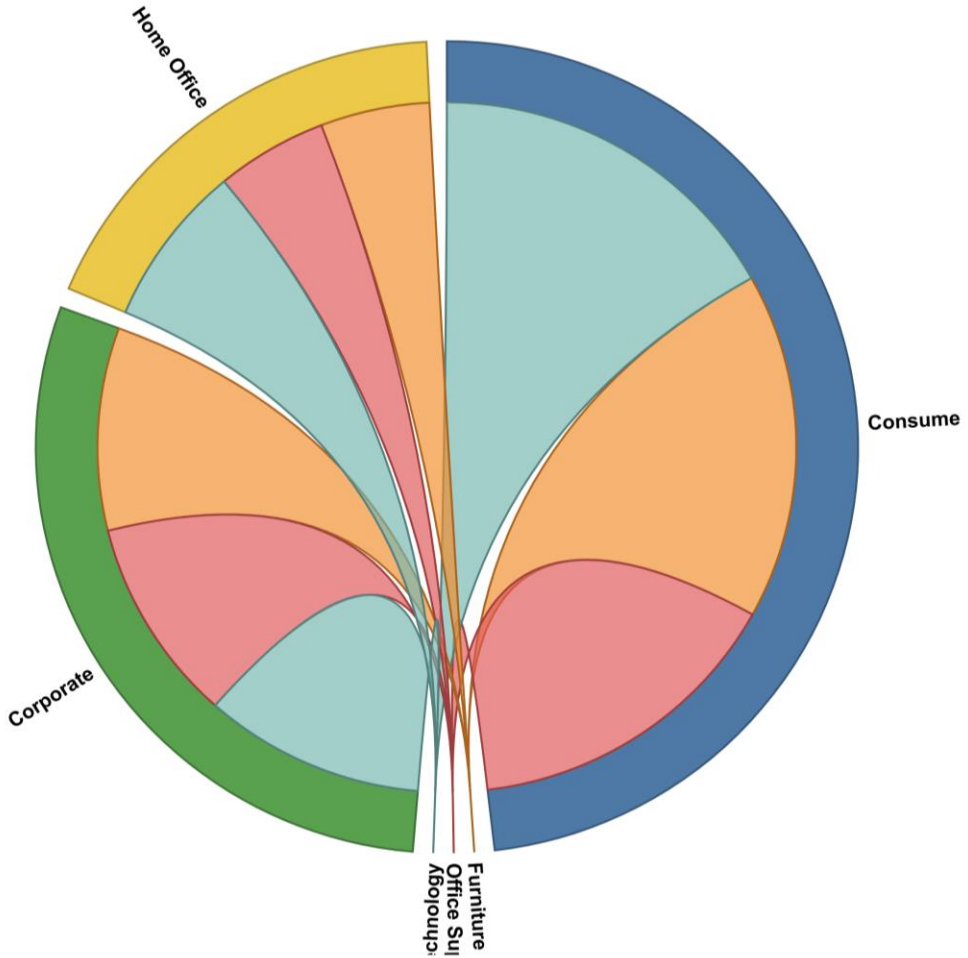


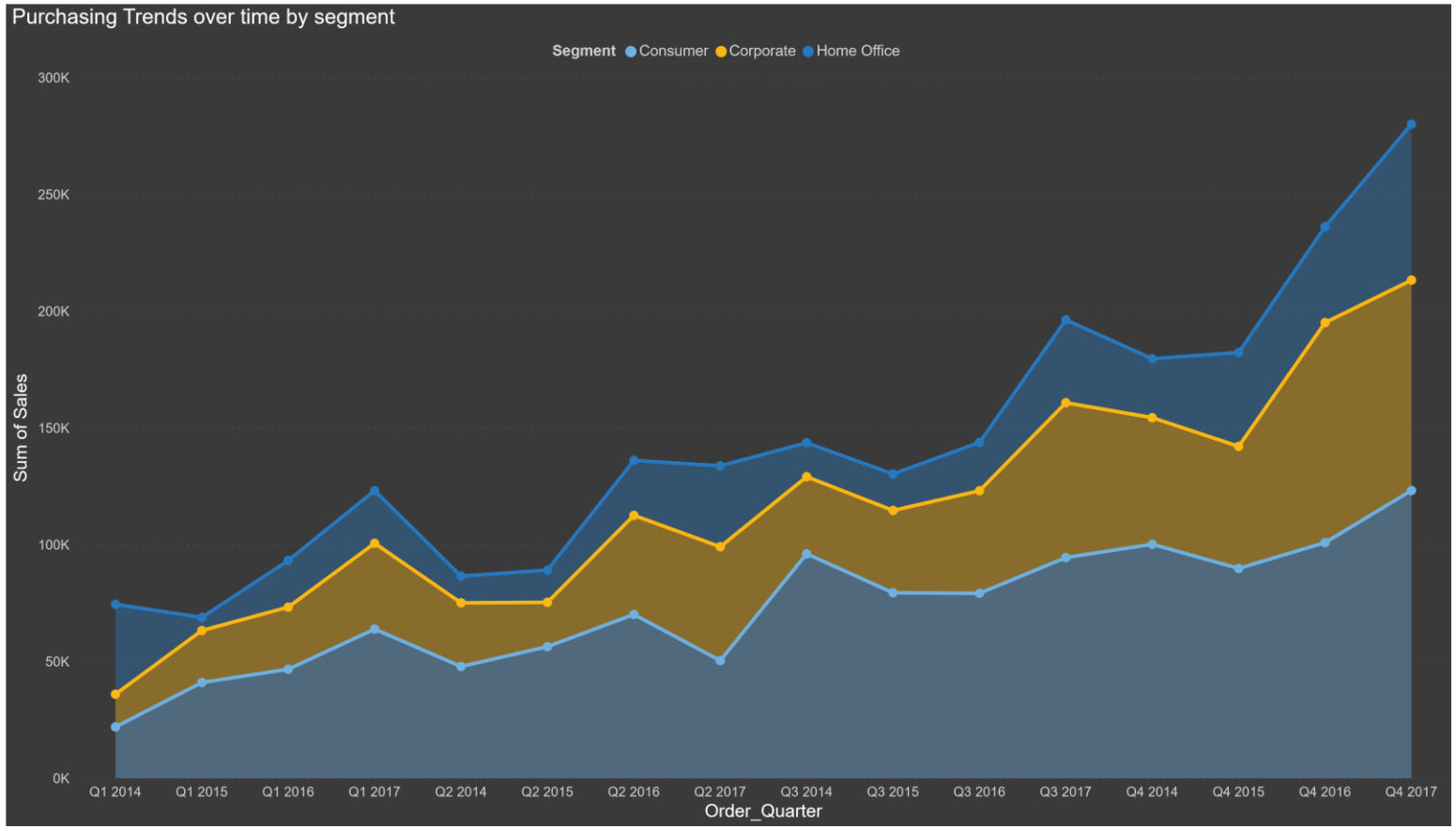




**Bubble chart showing the relationship between average discount and average profit margin, with bubble size representing total sales and color indicating category**

**Violin plot indicating profit distribution per Category, by discount**

# Segmenting customers (Consumer, Corporate, Home Office) to understand behavior and purchasing trends.

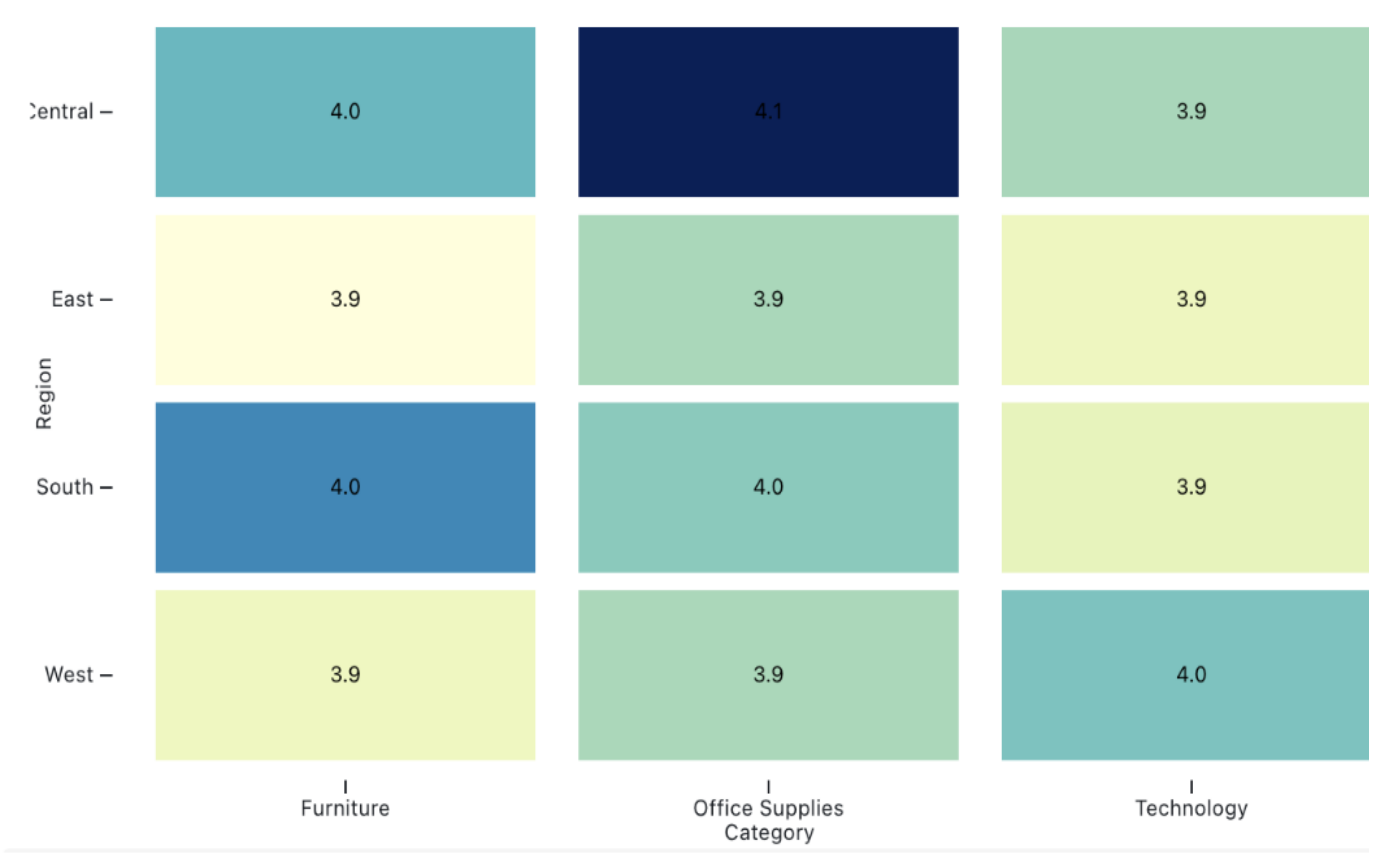


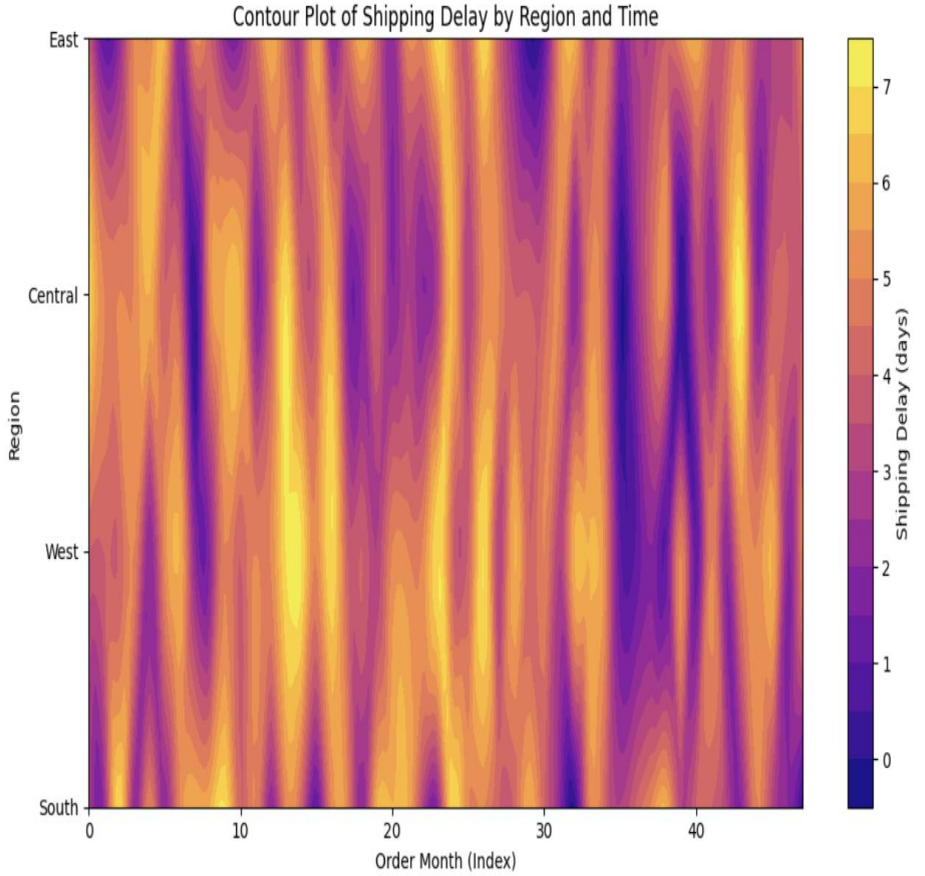

Chord diagram visualizing the relationship between customer segments and product categories through arcs and ribbons

Stacked area chart displaying quarterly sales trends over time from Q1 2014 to Q4 2017, divided by customer segments

# Analyzing order and shipping timelines to assess operational efficiency

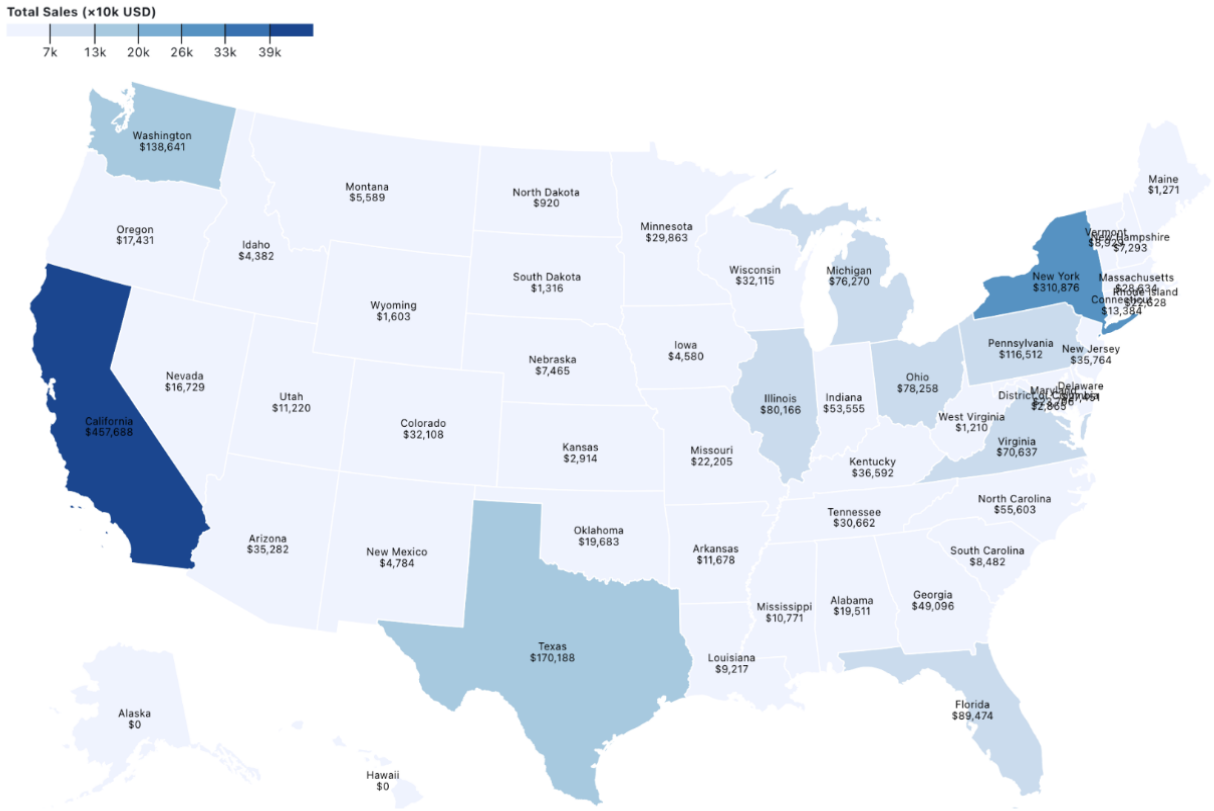



Heatmap showcasing the delay of shipment in each region.

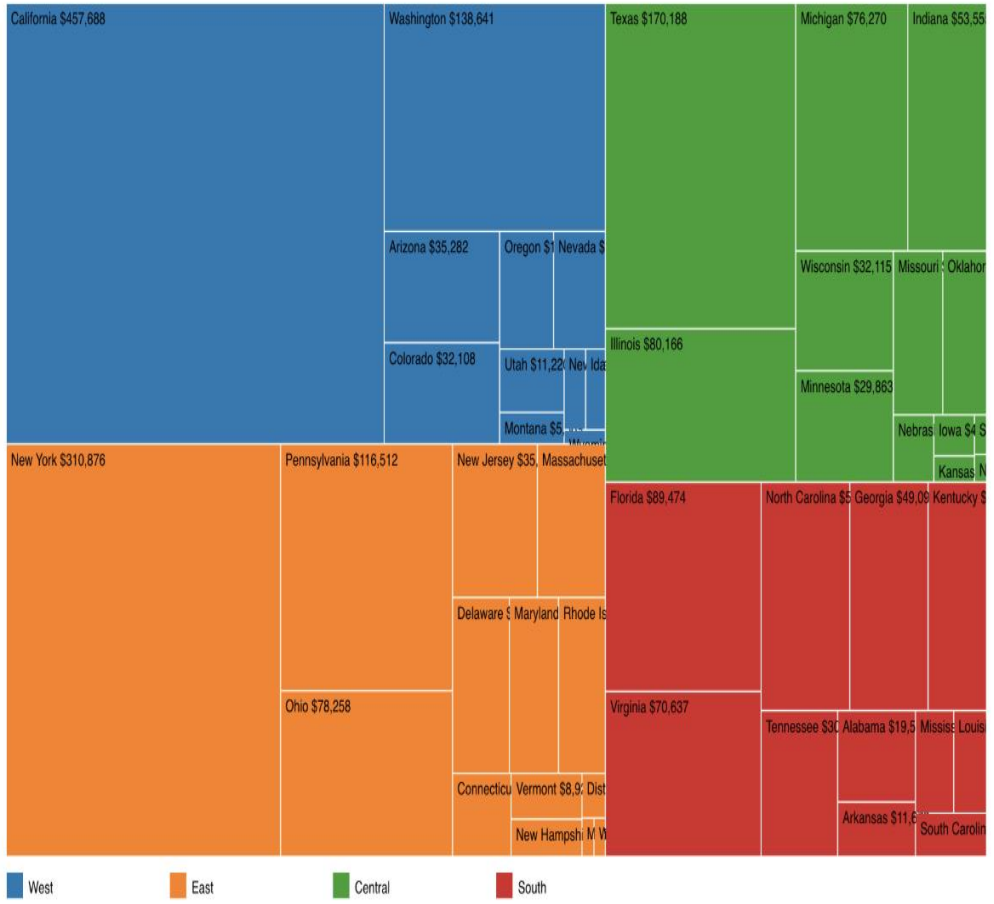



Counter Plot of Shipping Delays by region and time

# Examining regional sales patterns across U.S. states and regions to uncover geographic insights.




Choropleth map showcase the sales in each state in US




TreeMap focusing on sales per state and distributing the graph based on region

# LIMITATIONS

- **Limited customer detail**: Dataset lacks demographic and behavioral data, restricting segmentation.
- **No returns or satisfaction data**: Omits returns, refunds, and customer feedback, which may affect profitability.
- **Geographic granularity**: Insights limited to state-level; more detailed location data could improve targeting.
- **Data quality assumptions**: Relies on clean, complete data—missing or incorrect timestamps could skew results.
- **Scope for future work**:
  - Integrate external data (e.g., competitor pricing, economic indicators)
  - Analyze trends across multiple years for temporal depth

# CONCLUSION

- **Product performance**: High-tech items (e.g., copiers, machines) are top performers; accessories and supplies often underperform.
- **Discounting impact**: High discounts significantly reduce profit margins—especially in steady-demand categories like Binders.
- **Operational inefficiencies**: Notable shipping delays in the Central region and specific categories signal fulfillment issues.
- **Customer segmentation**: Consumer segment leads in volume; Corporate and Home Office segments show untapped potential.
- **Regional performance**: Sales are concentrated in a few states—highlighting both strengths and market expansion opportunities.
- **Visualization tools**: Used advanced visuals (e.g., chord diagrams, contour plots, heatmaps) to reveal hidden patterns.
- **Business impact**: Insights support better decisions in pricing, marketing, logistics, and customer targeting.
- **Outcome**: Delivered a visually driven dashboard and report to guide strategic planning.

# Thank you!