



# EXPLORATORY DATA ANALYSIS ON SUPERSTORE DATASET

Project: Practicing SQL & Data Visualization

by: Nabila Salsabillah



[Linkedin](#)



[Github](#)

# PURPOSE

1

Practice SQL querying  
(aggregations, CTEs,  
window functions)

2

Learn to translate raw data  
into business insights

3

Improve visualization  
skills with Tableau

4

Build a portfolio  
project to showcase  
my learning journey

# DATA & TOOLS



## Dataset Used

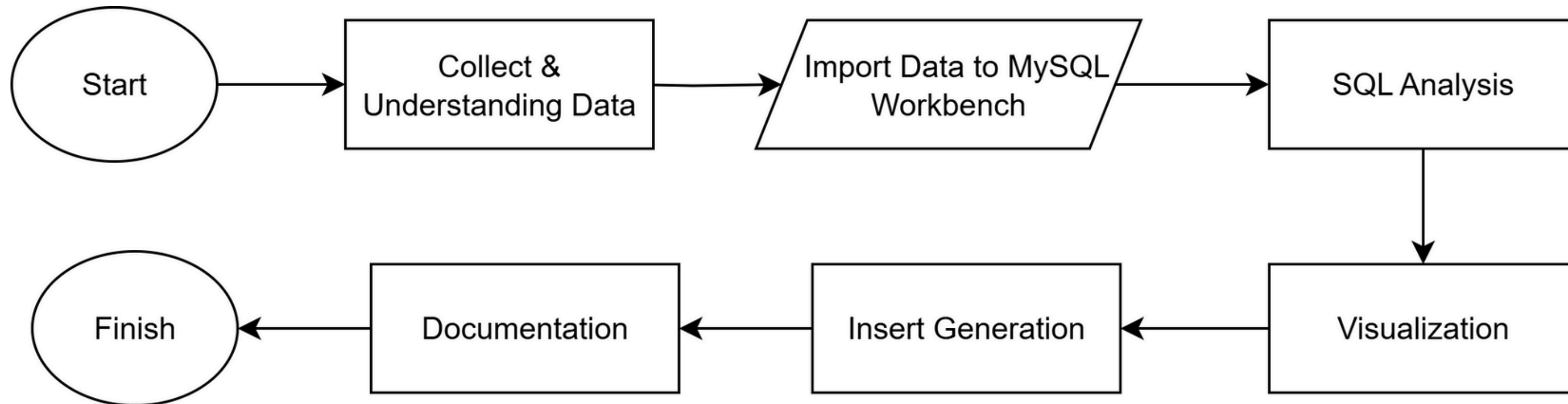
Superstore dataset with 21 columns and 9994 records. Data sourced from Kaggle [Superstore Dataset](#)

## Tools Used

- MySQL Workbench (SQL Querying)
- Tableau (Interactive visualization)
- Notion (Documentation)



# WORKFLOW



# SALES VS PROFIT BY REGION

```
SELECT Region,  
       SUM(Sales) AS Total_Sales,  
       SUM(Profit) AS Total_Profit  
FROM superstore  
GROUP BY Region;
```

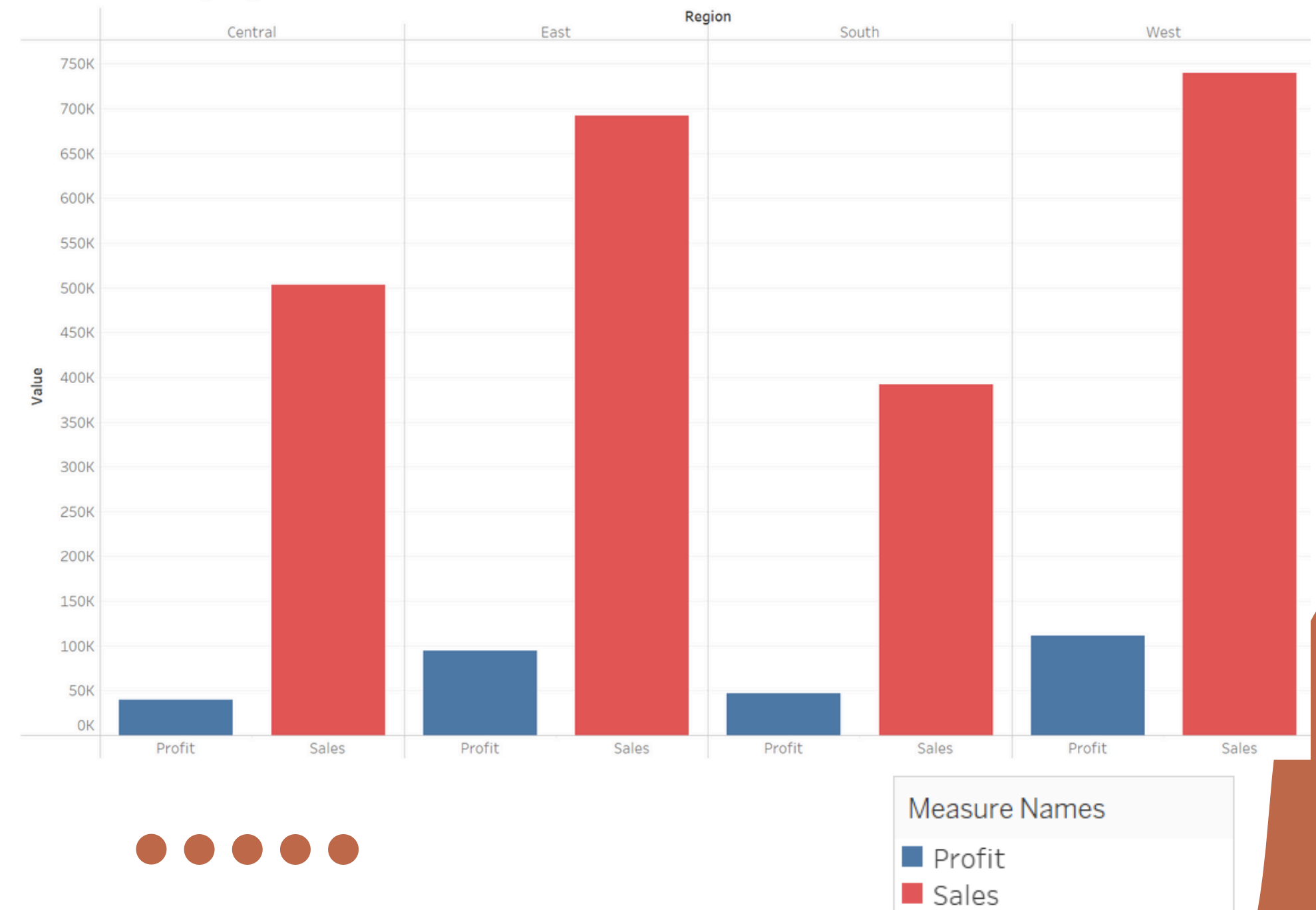
## 💡 Insight

- West region has the strongest sales & profit, even with high discounts.
- Central region struggles — big discounts but lowest profit.

## 🎯 Recommendation

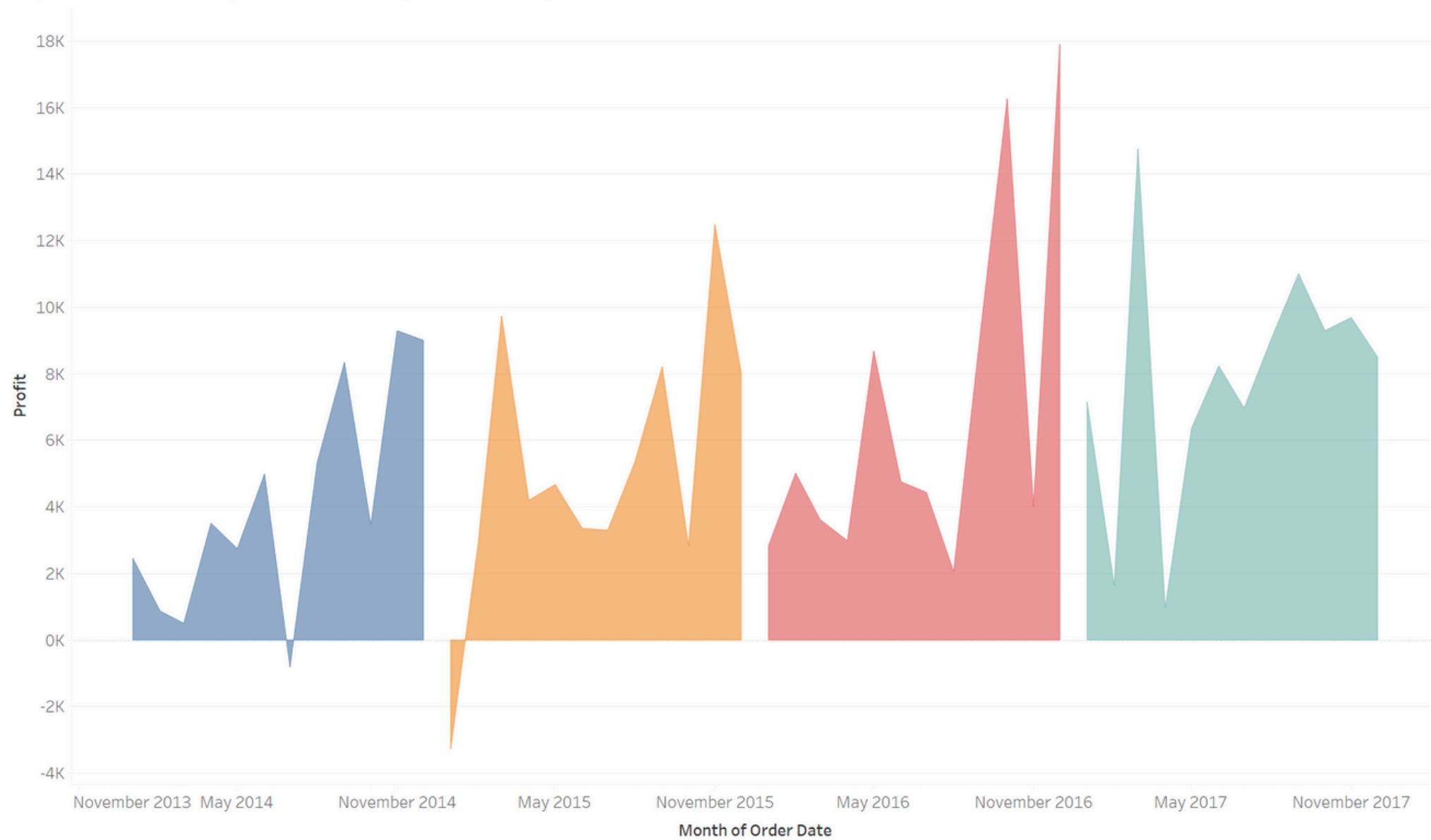
- Benchmark West's discount strategy.
- Reevaluate Central's pricing & discount approach.

Sales vs Profit by Region



# MONTHLY PROFIT TREND

Superstore Monthly Profit Trend (2014–2017)



YEAR(Order Date)

- 2014
- 2015
- 2016
- 2017



```
SELECT DATE_FORMAT(order_date_new, '%Y-%m') AS Month,
       SUM(Profit) AS Total_Profit
FROM superstore
GROUP BY Month
ORDER BY Month;
```



## Insight

- Profit fluctuates throughout 2014–2017.
- 2 negative months: July 2014 & January 2015.
- Year-end promotions do not always guarantee higher profit.

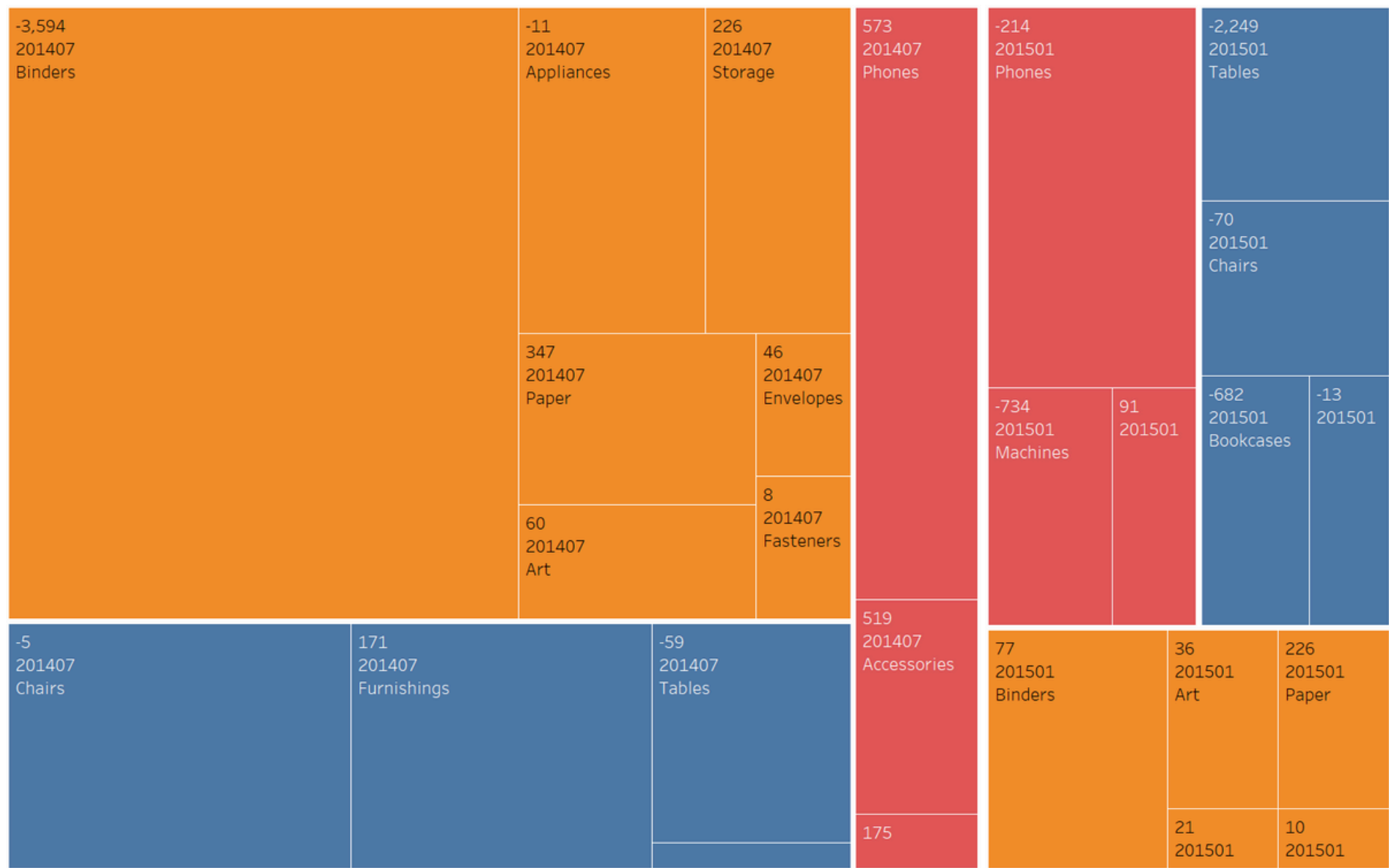


## Recommendation

- Reevaluate year-end promotion strategies.
- Analyze year-over-year trends to optimize campaign performance.

# LOSS-MAKING MONTHS BREAKDOWN

Profit and Discount Subcategories in Loss Months



```
SELECT DATE_FORMAT(order_date_new, '%Y-%m') AS Month,
       Category, `Sub-Category`,
       ROUND(SUM(Profit),2) AS Total_Profit,
       ROUND(SUM(Discount),2) AS Total_Discount
FROM superstore
```

...

- Insight**
- Binders (Office Supplies) caused the biggest loss in July 2014 (-3,594).
  - Tables (Furniture) caused the biggest loss in January 2015 (-2,249).
  - High discounts reduced margins, even in high-order sub-categories.

- Recommendation**
- Implement data-driven, targeted discounting instead of blanket promotions.
  - Conduct margin analysis before launching large promotions, especially in critical months.

Category

☐ Furniture

☐ Office Supplies

☐ Technology

Order Date (Month / Year) (group)

☐ (All)

☒ 201407

☒ 201501

☐ A



# DELIVERY STATUS BY MONTH & CATEGORY

```
WITH shipping_status AS (  
  SELECT Category,  
         DATE_FORMAT(order_date_new, '%Y-%m') AS order_month,  
         CASE  
           WHEN DATEDIFF(ship_date_new, order_date_new) <= 4 THEN  
             'Tepat Waktu'  
           ELSE 'Terlambat'  
         END AS status  
  ...  
)
```

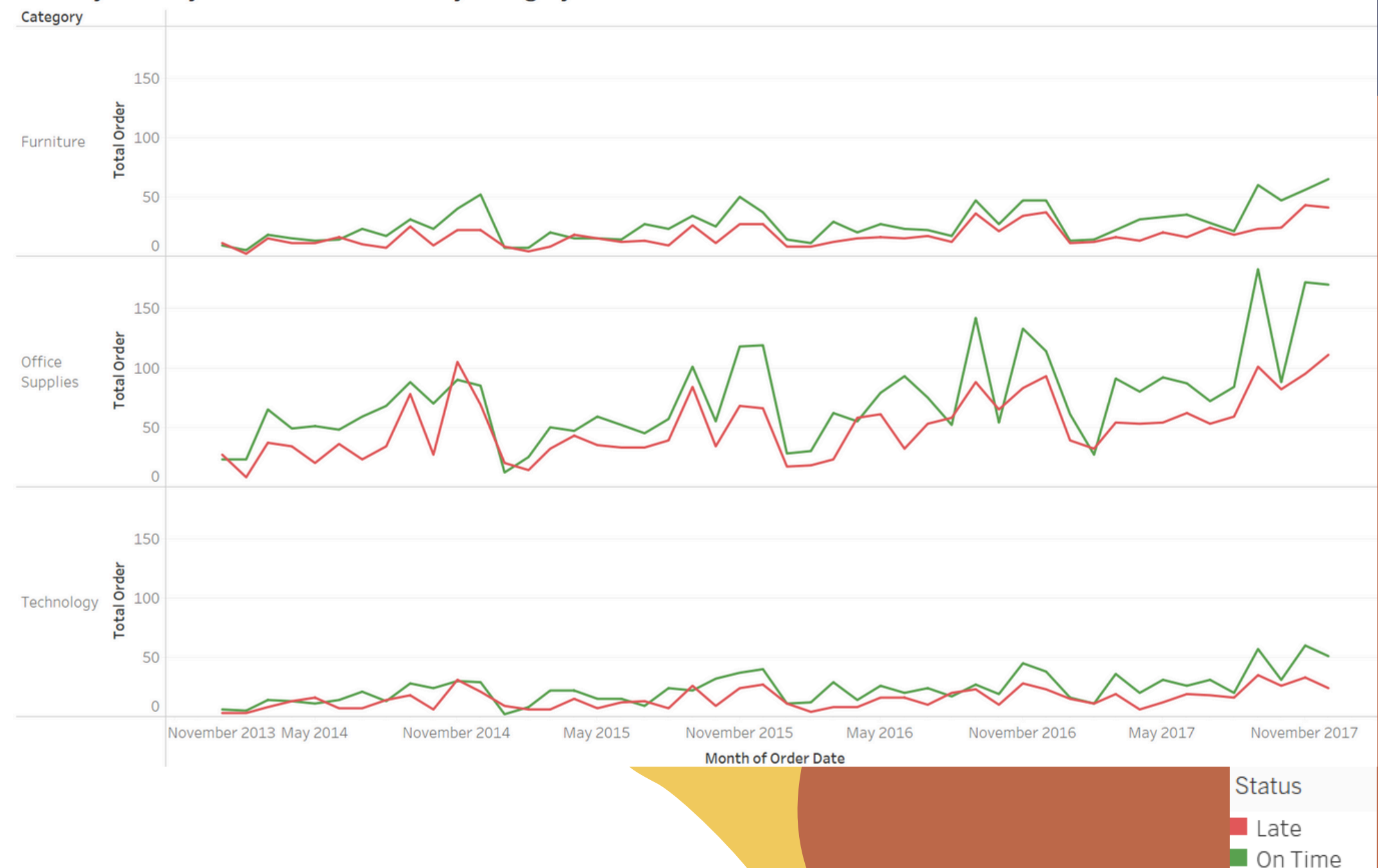
## 💡 Insight

- Delivery delays across all categories (Furniture, Office Supplies, Technology).
- Office Supplies saw a spike in delays (2015–2017) → possible supply chain issues.

## 🎯 Recommendation

- Conduct seasonal & demand analysis to anticipate surges.
- Improve inventory planning & distribution for delay-prone categories.

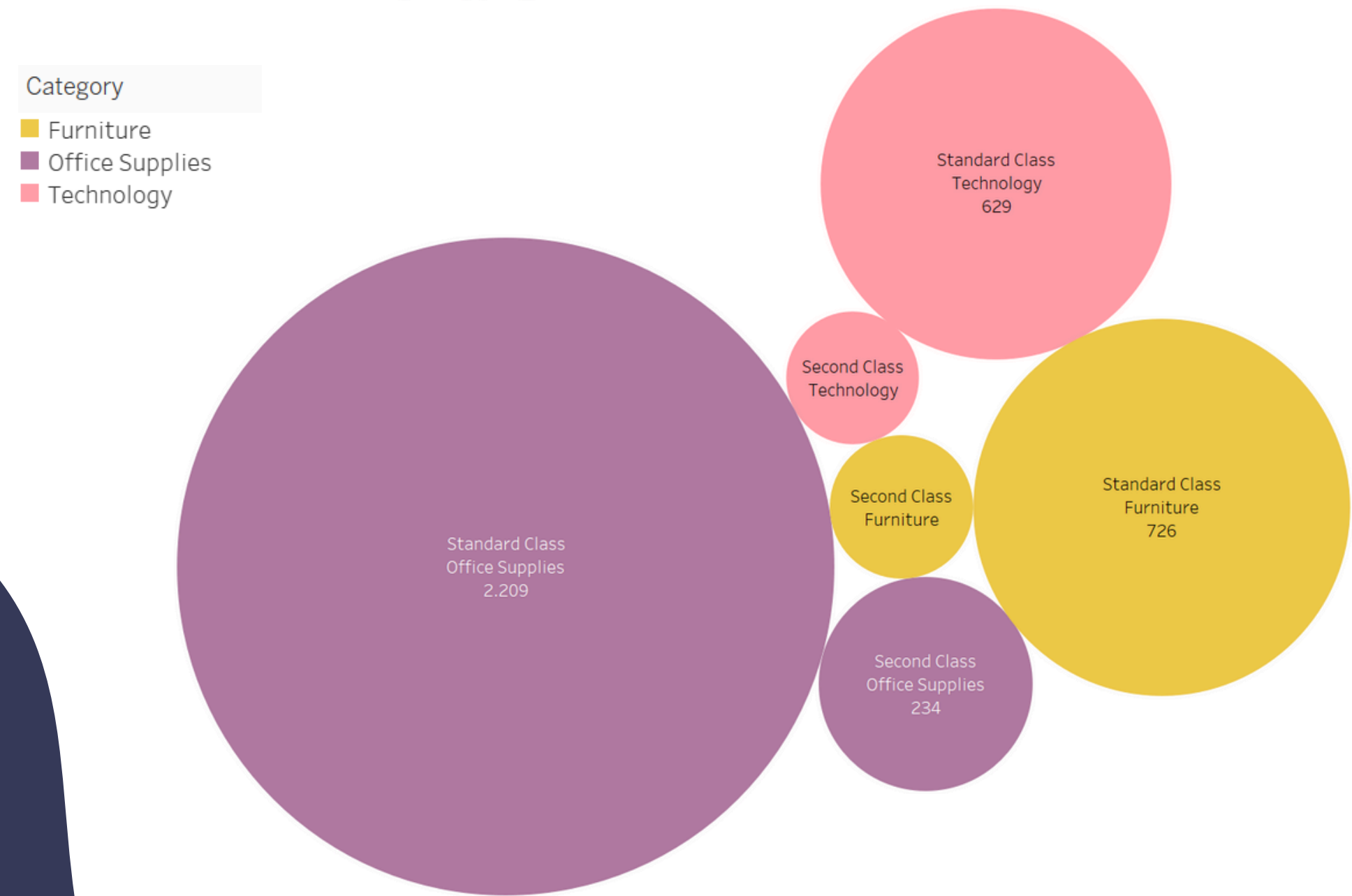
Monthly Delivery Performance Trends by Category





# DELIVERY DELAYS BY SHIPPING MODE

Distribution of Late Deliveries by Shipping Mode



```
WITH shipstatus AS (  
  SELECT 'Ship Mode', Category,  
         CASE WHEN DATEDIFF(ship_date_new, order_date_new) <= 4  
              THEN 'On Time' ELSE 'Late' END AS status  
  FROM superstore  
  ...
```

- Insight**
- Most delays in Standard Class, especially in Office Supplies.
  - Second Class had fewer delays; First Class & Same Day recorded no delays.
- Recommendation**
- Review Standard Class performance: improve SLAs & delivery processes.
  - Reallocate time-sensitive products to faster shipping modes.

# TOP CUSTOMERS OVERVIEW

```
SELECT `Customer ID`, `Customer Name`,  
       COUNT(DISTINCT `Order ID`) AS Total_Orders,  
       ROUND(SUM(Sales),2) AS Total_Sales,  
       ROUND(SUM(Profit),2) AS Total_Profit  
FROM superstore  
GROUP BY `Customer ID`, `Customer Name`  
ORDER BY Total_Sales DESC;
```

- 💡 Insight
- Sean Miller has the highest sales but low profit.
  - Other top customers contribute balanced sales & profit.
- 🎯 Recommendation
- Identify factors affecting low profitability in high-sales customers.
  - Prioritize high-margin customers for loyalty programs.

Sales vs Profit by Customer Name

Customer Name	Sales	Profit
Sean Miller	25,043	-1,981
Tamara Chand	19,052	8,981
Raymond Buch	15,117	6,976
Tom Ashbrook	14,596	4,704
Adrian Barton	14,474	5,445
Ken Lonsdale	14,175	807
Sanjit Chand	14,142	5,757
Hunter Lopez	12,873	5,622
Sanjit Engle	12,209	2,651
Christopher Conant	12,129	2,177

# CASE STUDY: SEAN MILLER

Customer Who Has High Sales But Low Profit ●●●●●

```
SELECT Category, `Sub-Category`,
       ROUND(SUM(Sales),2) AS Total_Sales,
       ROUND(SUM(Profit),2) AS Total_Profit
FROM superstore
WHERE `Customer Name` = 'Sean Miller'
GROUP BY Category, `Sub-Category`
ORDER BY Total_Sales;
```

- 💡 **Insight**
- Purchases concentrated in low-margin categories → drives high sales but low profit.
  - Discounts were not excessively high, suggesting category mix is the key issue.

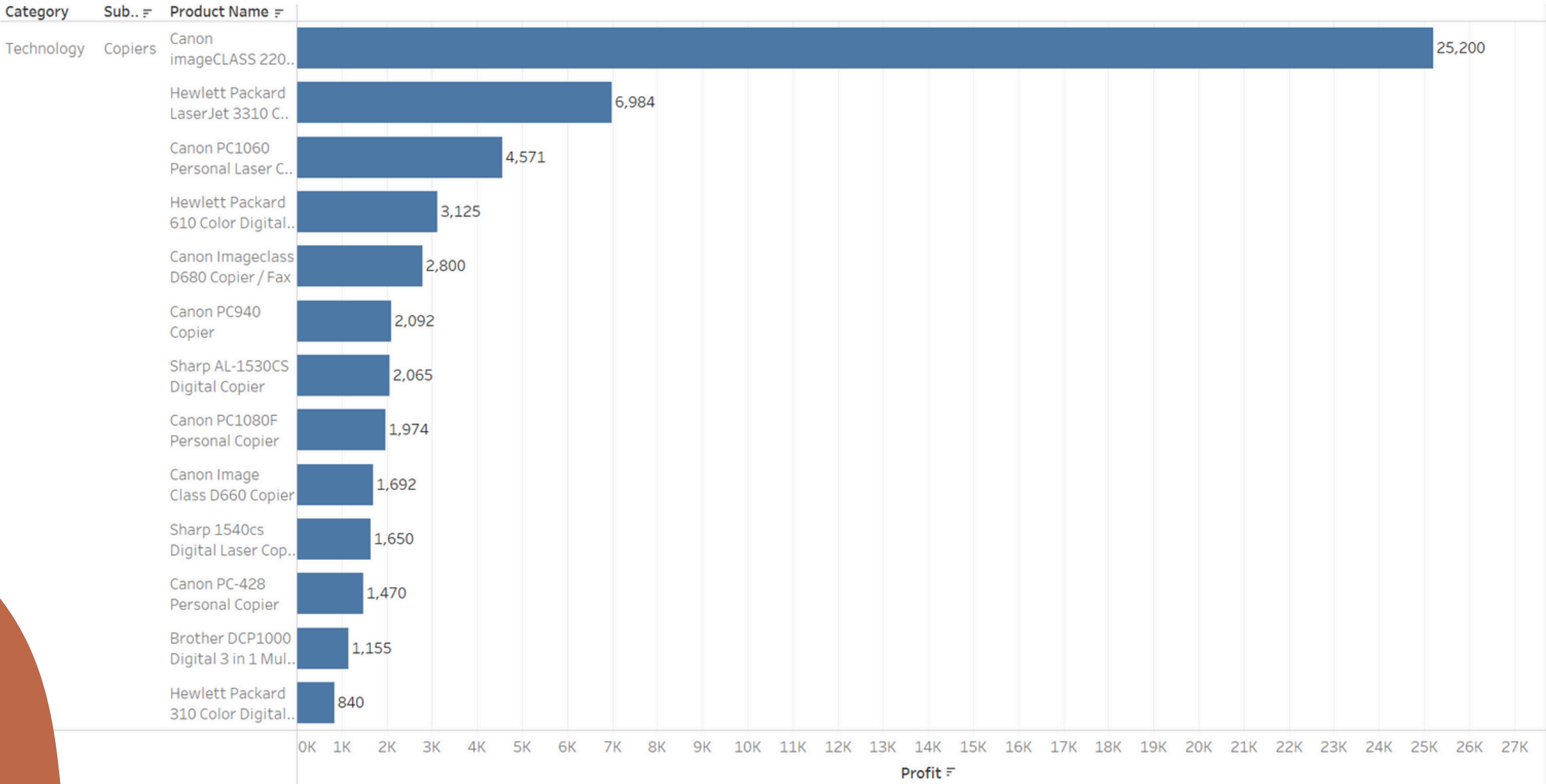
- 🎯 **Recommendation**
- Encourage purchases in higher-margin categories via personalized offers.
  - Review pricing strategies for frequently purchased low-margin items.


Customer Name	Category	Sub-Category	Sales	Profit
Sean Miller	Furniture	Furnishings	680	55
	Office	Art	8	1
	Supplies	Binders	100	-83
		Fasteners	19	6
		Paper	89	31
		Storage	663	-166
		Supplies	3	-1
	Technology	Accessories	22	4
		Machines	23,460	-1,828




# TOP PROFIT CONTRIBUTING PRODUCTS

Top Profit-Contributing Products in the Copiers Sub-Category



-  **Insight**
- Copiers is the most profitable sub-category overall.
  - Canon imageCLASS dominates profits (significantly higher than other products).
  - Large profit gap between top and mid-performing products.

-  **Recommendation**
- Prioritize marketing & inventory for high-profit products like Canon imageCLASS.
  - Explore opportunities to boost mid-performing products to balance profit contribution.

```
WITH Subcat_Profit AS (  
  SELECT `Sub-Category`,  
         ROUND(SUM(Profit),2) AS SubTotal_Profit,  
         RANK() OVER(ORDER BY SUM(Profit) DESC) AS Rank_Profit  
  FROM superstore  
  GROUP BY `Sub-Category`  
  ...
```

# KEY TAKEAWAYS

**West region = most profitable  
(despite discounts)**

**Discounts ≠ guaranteed profit  
(Central proves this)**

**Profit fluctuates significantly,  
especially year-end**

**Standard Class = most delays  
(esp. Office Supplies)**

**Copiers (Canon imageCLASS)  
= biggest profit driver**





# WHAT'S NEXT??



## Recommendations:

1. **Revisit discount strategies (targeted, data-driven).**
2. **Optimize shipping for Standard Class & high-value products.**
3. **Focus on high-performing products & develop mid-tier performers.**
4. **Next Steps:**
5. **Build predictive models for profit & shipping delay.**
6. **Segment customers for personalized discounts & delivery options.**
7. **Automate dashboards for real-time performance monitoring.**

# THANK YOU

For Viewing My Project



[Linkedin](#)



[Github](#)

