<u>**Sales And Revenue Analyst**</u>

D – Discover (Basic Finding)

Initial Questions:

1. What drives sales performance across different dimensions?
2. Which factors most impact revenue generation?
3. Where are the growth opportunities?

Basic Answers / Metrics:

- Monthly sales trend shows clear seasonality with spikes toward year-end.
- **Product category impact**:
  - The technology category contributes disproportionately to sales.
  - Furniture and Office Supplies vary but lag Technology.
- **Regional differences**:
  - The West consistently outperforms all other regions.
- **Customer segment performance**:
  - The Consumer segment drives the largest share of total sales and profit.
- **Discounts and quantity sold**:
  - Higher discounts correlate negatively with profit but drive greater unit volumes.

First Impressions:

- Sales performance is strongly seasonal and driven by customer segment (Consumer), category (Technology), and region (West).
- Discounts appear effective in driving volume but erode profitability.

Key Findings at Discover stage:

- Sales drivers include seasonality, category preference (Technology), regional concentration (West), and the Consumer segment.
- Discounting boosts quantity but at a marginal cost.

I – Investigate (Dig Deeper)

Why do these patterns exist?

- **Seasonality:** Peaks align with retail holidays (November–December), suggesting consumer-driven cycles.

- **Technology dominance:** Technology products (e.g., phones, accessories) may attract repeat purchases, upgrades, and bundling opportunities.
- **Regional strength in the West:** This could reflect population density, customer demographics, or superior delivery logistics.
- **Customer segment:** Consumers show consistent buying behavior, while Corporate and Home Office are more variable.

Variability across dimensions:

- South and Central regions underperform despite a sizable customer base—potential untapped market.
- Furniture category shows volatility, possibly driven by high-ticket purchases and greater sensitivity to economic conditions or discounting.
- **Discount patterns:** Regression analysis confirms diminishing returns—beyond a certain point; deeper discounts reduce overall revenue contribution.

Key Findings at Investigate stage:

- Sales performance patterns are driven by both external factors (holidays, demographics) and internal pricing/promotion strategies.
- Underperforming segments/regions could yield growth with tailored strategies.
- Discounts drive sales quantity but must be balanced against the margin of impact.

V – Validate (Challenge Assumptions)

What could make these conclusions wrong?

- **Data limitations:**
    - No marketing campaign data: cannot confirm causality between discounts and volume.
    - No demographic breakdown: cannot fully explain regional variations without customer profiles.
    - Limited time span: dataset may not reflect longer-term trends (e.g., post-pandemic shifts).
- **Alternative explanations:**
    - West region outperformance could be due to operational efficiencies (e.g., fulfillment centers) rather than consumer demand alone.
    - Technology category success could be amplified by strategic promotions rather than inherent customer preference.

Key Findings at Validate stage:

- The drivers identified are credible but subject to important limitations.
- The analysis would benefit from additional data (e.g., marketing spending, customer demographics) and external benchmarks.

E – Extend (Strategic Application)

Strategic recommendations:

1. Refine discounting strategy:

- Introduce elasticity-based discounting, adjusting discount depth by category and seasonality to protect margins.
- Limit blanket discounts on Technology category, which shows strong performance even without heavy discounting.

2. Unlock regional growth:

- Target South and Central regions with localized promotions and potentially enhanced delivery options.
- Investigate potential logistical or brand awareness issues suppressing sales.

3. Expand Consumer segment engagement:

- Develop loyalty programs for the Consumer segment, which drives volume and profit.
- Use personalized offers to sustain momentum outside peak seasons.

4. Leverage seasonality proactively:

- Forecast inventory and marketing campaigns based on predictable Q4 demand spikes.
- Explore additional promotional periods (e.g., back-to-school) for Office Supplies.

5. Improve data analytics maturity:

- Integrate marketing, logistics, and demographic data to refine predictive models and better understand drivers across regions and categories.

Measuring impact:

- KPIs:
  - Profit per discount dollar by category.
  - Regional sales uplift (South and Central).
  - Repeat purchases from the Consumer segment.

Risks:

- Misalignment between discounting strategy and customer expectations.
- Regional campaigns may require upfront investment with uncertain ROI.
- Forecasts rely on historical patterns that may change.

Key Findings at Extend stage:

- Actionable strategies exist to protect margin, grow underperforming regions, and deepen Consumer engagement.
- Building predictive capabilities and enriching data will improve precision and decision quality.

Final Recommendations Summary:

Strategic Actions:

- Leverage Q4 seasonal surge through targeted promotions.
- Invest in underperforming regions to unlock growth potential.
- Optimize discount strategy to protect profitability while stimulating demand.
- Focus on stable-performing categories like Technology for year-round revenue.

Refine discount strategies for profitability:
Implement a data-driven, elasticity-based discounting policy that varies discount depth by product category and seasonal timing. For example, limit heavy discounting on Technology, where demand is strong even without promotions, while offering targeted discounts for Furniture to stimulate demand during off-peak periods.

Target South and Central regions with localized strategies:
Analyze underperformance in these regions by investigating customer demographics, brand awareness, and logistics capabilities. Roll out regional marketing campaigns, regional pricing strategies, and assess opportunities for improved shipping and delivery performance.

Build customer loyalty programs:
Focus on the Consumer segment, which consistently contributes the largest share of sales and profits. Offer rewards for repeat purchases, early access to seasonal sales, and personalized promotions to increase retention and lifetime value.

Improve analytics infrastructure:
Integrate broader datasets (e.g., marketing campaign data, logistics performance, customer demographic insights) into a predictive analytics model to continuously optimize revenue generation, regional performance, and category strategy.

Monitor key metrics continuously:
Establish KPIs such as:

- Regional sales uplift in South and Central
- Profit per discount dollar by product category
- Repeat purchase rate for Consumer segment
- Seasonal inventory turnover rates

These metrics will allow ongoing monitoring and optimization of strategy execution.

DIVE Analysis: Customer Behavior Analysis

D - Discover (Basic Finding)

Initial Question: Who are our most valuable customers and what are their basic purchasing patterns?

Basic Answer/Metric: We've segmented customers by total revenue. The 'Very High Value' segment represents customers contributing the most revenue historically. We've also identified the top selling products and categories.

First Impression: A small percentage of customers likely contribute a large portion of the revenue (Pareto principle). There seem to be clear preferences for certain products and categories.

I - Investigate (Dig Deeper)

Why does this pattern exist?

- **Revenue Segmentation:** Why do some customers spend significantly more than others? Factors could include their needs, purchasing power, product preferences, frequency of purchase, or the types of products they buy (high-ticket items).
- **Popular Products/Categories:** Why are certain products/categories more popular? This could be due to market demand, pricing, quality, marketing efforts, or seasonal trends.
- **Purchase Frequency:** Why do some customers buy more often? This could be related to the type of product (consumables vs. durables), their loyalty, or specific needs.

What factors contribute to this?

- Product characteristics (price, type, necessity)
- Customer demographics/needs
- Marketing and promotions
- Seasonality
- Customer experience

How does it vary across dimensions?

- **Geographic:** Do purchasing patterns or customer value vary by region (State, City)? (This was explored in the code if 'State'/'City' were available).
- **Product Category:** Do high-value customers purchase across many categories or specialize?
- **Time:** How do purchasing patterns change over time (seasonality, trends)? (Explored with monthly/weekday sales).

Gemini Prompts Used (Simulated):

- "Based on the Superstore dataset schema (Order ID, Customer ID, Product Name, Sales, etc.), what hypotheses can we form about why some customers have significantly higher total sales?"
- "Suggest additional analysis steps to understand customer purchasing patterns beyond simple frequency and popular items, considering potential time-based factors or product relationships."
- "Given customer data with purchase dates, how can we calculate customer lifespan and use it as an indicator of CLTV?"
- "Explain the concept of RFM analysis and how Recency, Frequency, and Monetary values are calculated from transactional data."

V - Validate (Challenge Assumptions)

What could make this conclusion wrong?

- **Data Accuracy:** Are the 'Sales' figures accurate? Are 'Customer ID's consistently identifying unique customers?
- **Segmentation Method:** Is revenue the best primary metric for segmentation? RFM provides a different perspective.
- **Time Period:** This analysis is based on the available data time frame. Is this representative of long-term behavior? A longer time series would be needed for more robust CLTV or lifespan analysis.
- **RFM Segmentation:** The manual RFM segmentation rules (`rfm_segment` function) are simple examples; real-world RFM requires careful tuning based on business goals and data distribution. The 'Others' category is broad and needs refinement.
- **Causation vs. Correlation:** Identifying patterns doesn't automatically mean causation. E.g., a product selling well might be popular, or it might simply be heavily promoted.

What data limitations exist?

- Lack of external demographic data beyond what's in the dataset (if any beyond State/City).

- Limited information on customer interactions outside of purchases (e.g., website visits, marketing engagement).
- No explicit definition of 'churned' customers, making true customer lifespan and retention rate calculations difficult.
- Market Basket Analysis requires significant computational resources for large datasets and requires careful interpretation of rules.

Are there alternative explanations?

- High sales for a customer could be one large infrequent purchase rather than consistent loyalty. RFM helps distinguish this (low Frequency, high Monetary, potentially high Recency if it was long ago).
- Seasonality might be driven by external factors (holidays, promotions) rather than intrinsic customer behavior alone.

E - Extend (Strategic Application)

What should the business do?

- **Targeted Marketing:** Develop specific campaigns for each RFM segment ('Champions' for loyalty programs, 'At Risk' for re-engagement, 'New Customers' for onboarding).
- **Customer Service:** Prioritize support for high-value segments.
- **Inventory Management:** Ensure popular products and categories are well-stocked, especially during peak seasons.
- **Promotions:** Design promotions based on popular items and potentially bundling products frequently bought together (requires Market Basket Analysis).
- **Sales Strategy:** Focus sales efforts on acquiring customers similar to those in the 'Very High Value' or 'Champions' segments.
- **Website/Store Layout:** Highlight popular categories and suggest related products.

How can we measure impact?

- **Customer Retention Rate:** Track the percentage of customers who make repeat purchases over time, especially within target segments.
- **Segment Migration:** Monitor how customers move between RFM or revenue segments over time (e.g., 'New Customers' becoming 'Loyal').
- **Average Order Value (AOV):** Track if promotions or bundling increases AOV.
- **Purchase Frequency:** Measure if re-engagement campaigns increase the frequency of purchases for 'At Risk' or 'Low Value' customers.
- **Campaign ROI:** Evaluate the revenue generated by targeted marketing campaigns for specific segments.

What are the risks?

- **Alienating Segments:** Over-focusing on high-value segments might neglect the potential growth of medium- or low-value customers.

- **Incorrect Segmentation:** Misclassifying customers can lead to ineffective or counterproductive strategies.
- **Execution Risk:** Marketing or operational teams may face challenges implementing segment-specific strategies.
- **Data Privacy:** Ensure compliance when using customer data for personalization.
- **Ignoring Trends:** Relying solely on historical data may cause the business to miss emerging trends or shifts in customer behavior.

Summary and Recommendations:

Based on the DIVE analysis, we have identified key customer segments based on revenue and RFM metrics. We understand basic purchasing patterns including popular items and seasonality.

Key Recommendations:

Implement RFM-based Marketing: Develop and execute targeted campaigns for 'Champions', 'Loyal Customers', 'New Customers', and 'At Risk' segments.

1. **Enhance High-Value Customer Experience:** Introduce loyalty programs, early access to sales, or personalized service for 'Champions' and 'Very High Value' customers.
2. **Optimize Operations based on Patterns:** Use insights on popular products, categories, and seasonality to refine inventory management, staffing, and promotional calendars.
3. **Further Investigate 'At Risk' and Low-Value Segments:** Conduct deeper analysis or qualitative research to understand why these customers are less engaged and develop tailored strategies for improvement.
4. **Explore Predictive CLTV Modeling:** If more historical data becomes available, consider implementing predictive models for CLTV to forecast future customer value more accurately.
5. **Conduct Market Basket Analysis:** Investigate product co-purchasing patterns to inform bundling, cross-selling recommendations, and store/website layout.

DIVE Analysis: Product Performance Analyst

D - Discover (Basic Finding)

Initial Question: Which products/categories perform best and what drives product profitability?

Basic Answer/Metric:

The initial analysis of product performance reveals that Technology and Furniture subcategories generated the highest total sales over the past four years, with Technology leading in both sales and profitability. In contrast, Furniture, despite sustained sales growth, experienced fluctuating profitability and a decline in 2022. This divergence suggests an increase in the cost of goods sold or margin pressure within the Furniture category, contributing to its relatively low compared to Office Supplies and Technology. At the same time, Office Supplies showed a steady upward trend in profitability, reinforcing its position as a stable contributor.

A sharp increase in profitability was observed in Copiers starting in 2021, along with Accessories and Phones, which consistently contributed strong revenue under the Technology category.

These patterns clearly indicate that high sales volume does not always correlate with high profitability. In fact, some high revenue categories, such as Furniture are operating at a loss, highlighting the need to evaluate both sales and profitability metrics together.

First Impression:

Based on our initial findings, the top three subcategories with the highest profitability in 2022 were Copiers, Accessories, and Phones. All of these fall under the Technology category, which not only drives the most revenue but also generates the strongest profit margins.

In contrast, Binders, despite being the most frequently ordered and heavily discounted subcategory, consistently produced the largest negative profitability. This suggests that frequent discounting, high order volume, or low pricing efficiency in Binders is significantly affecting and eroding its margin.

I - Investigate (Dig Deeper)

- Product Profitability Analysis:
    - Phones and Accessories show healthy profit margins, with average profitability well above 20%.
    - Tables and Chairs, although high in sales, are associated with negative profits, indicating over-discounting or high costs.
- Category Performance Comparison:
    - Technology is the top-performing category, delivering both high sales and positive profit margins.
    - Furniture, despite strong sales, lags behind with negative profitability — making it a priority for cost or pricing reevaluation.
    - Office Supplies performs steadily with modest sales and moderate profit margins.
- Discount Impact on Product Success:
  As discount levels increase, average profitability declines significantly:
    - Low discounts (≤10%) yield the highest profit margins, about 22%.
    - Medium discounts (10–20%) result in negative margins, about −6.6%.
    - High discounts (>20%) are associated with the worst margins, about −12.5%.

V - Validate (Challenge Assumptions)

Limitation to our assumptions:

- Sales are assumed to reflect product performance success
    - In reality, high sales volumes do not always translate to profitability. Products with strong sales may still operate at a loss due to excessive costs, discounts, or inefficiencies.

- Profit is assumed to be a reliable indicator of product health
    - The analysis relies on a pre-calculated profit field, but key financial components such as cost of goods sold, overhead, or returns are not available. This limits the accuracy of profit-based assessments and may misrepresent true product performance.
- Discounts are assumed to directly reduce profit margins
    - While higher discounts often correlate with lower margins, some discounts may be strategic, such as to increase volume, clear excess inventory, or support customer acquisition, and could have long-term value beyond short-term margin loss.
- Pricing behavior is assumed to be consistent across segments and customers
    - Pricing at the subcategory is not provided, so we cannot determine whether special offers, segment-specific deals, or bundling strategies were applied. As a result, important pricing-related insights may be overlooked or misinterpreted.

Confidence in findings:

- High confidence in the finding that high discounts erode profit margins and contribute significantly to negative profits, particularly for specific subcategories.
- Moderate to high confidence in the identification of key profitable and unprofitable categories/subcategories/segments.

E - Extend (Strategic Application)

- Optimal Discount Strategy: We analyzed total profit and average profit margin across different discount ranges to identify ranges that might balance sales volume and profitability. This revealed that high discount levels are generally harmful to profitability, particularly for certain product groups.
- Targeted Discounting: We identified specific subcategories with significant negative profit (like 'Tables') and analyzed the discount ranges contributing most to these losses. This allowed for recommendations on where to reduce or eliminate high discounts.

Summary and Recommendations:

Overall Conclusion:

- The comprehensive analysis of the Superstore data using the DIVE framework has revealed key insights into sales performance, profitability, and the impact of pricing strategies, specifically discounting.
- The analysis clearly demonstrates a strong negative correlation between discount levels and profit margin. High discounts are a significant driver of negative profitability, particularly for certain product subcategories like 'Tables' and 'Bookcases'.
- While some subcategories and segments, such as 'Technology' category, 'Phones' subcategory, 'Consumer' segment, are generally profitable, even within these areas, high discounts can erode margins and lead to losses on individual orders.

Key Recommendations:

- Review and Optimize Discounting Strategy: Implement a more targeted and data-driven discounting strategy. Significantly reduce or eliminate high discounts, such as >40-50%, for subcategories that consistently show negative profit at these levels, especially 'Tables', 'Bookcases', and 'Supplies'. Analyze profitability within specific discount ranges for other subcategories to identify levels that maximize total profit while maintaining acceptable margins.
- Address Unprofitable Products/Subcategories: Investigate the underlying cost structure of products within subcategories with significant negative profit. If cost of goods sold or operational/shipping costs are high, explore options to reduce these costs, adjust pricing upwards, even without discounts, or consider discontinuing consistently unprofitable items that cannot be made profitable.
- Implement Targeted Pricing and Promotions: Tailor discounting and pricing strategies to different customer segments based on their observed price sensitivity and profitability. Consider loyalty programs or value-based pricing for high-value repeat customers rather than relying solely on broad discounts.
- Enhance Data Collection: To conduct more in-depth analysis, collect and integrate additional data, particularly on cost of goods sold, and shipping costs. This will enable more accurate calculation of net profit and identification of cost drivers.

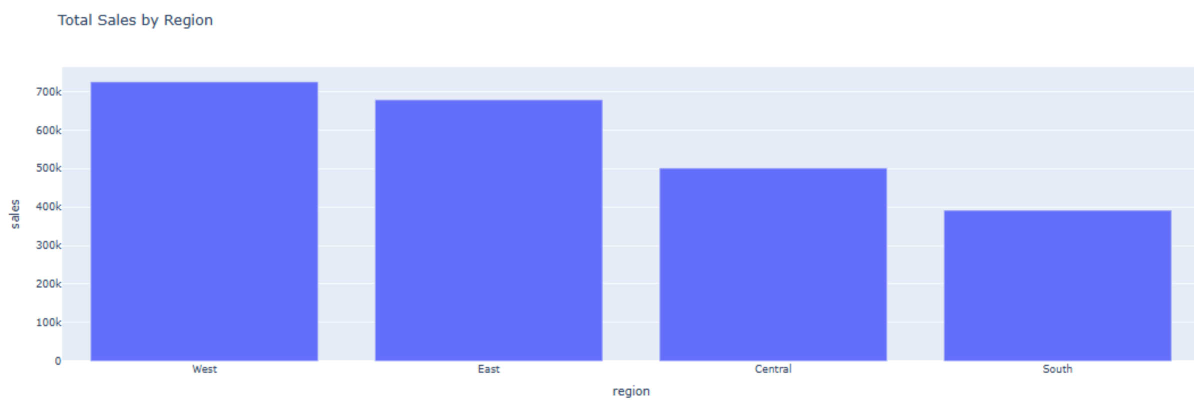
By implementing these recommendations, the business can move towards a more profitable sales operation, optimize pricing strategies to enhance overall profitability, and make data-driven decisions to ensure long-term success.

DIVE Analysis: Regional/Operational analyst

D - Discover (Basic Finding)

The regions in descending order of importance is West, East, Central and South.

the higher sales in the West seem to be driven by a combination of factors: more orders, a larger quantity of items sold, and higher profitability per order and per item.



I - Investigate (Dig Deeper)

Three factors that could drive this difference are:

1. Geographic Influence: Sales performance varies significantly by geographic location. (We've already seen some evidence of this with the region analysis).
2. Product Category/Subcategory Influence: Certain product categories or subcategories have significantly higher sales than others.
3. Discount Impact: The presence or amount of a discount influences sales volume.

V - Validate (Challenge Assumptions)

Based on below, direct bigquery vs AI generated analysis are the same.

Bigquery SQL

Row	region	total_sales	total_profit
1	West	725457.8245000...	108418.4489000...
2	East	678781.2399999...	91522.77999999...
3	Central	501239.890800001	39706.36250000...
4	South	391721.9050000...	46749.43030000...

Gemini created on Colab Prompt: Analyse sales and profit by region and arrange in descendi

```
region_sales_profit = df.groupby('region').agg(
    total_sales=('sales', 'sum'),
    total_profit=('profit', 'sum')
).reset_index()

region_sales_profit = region_sales_profit.sort_values(by='total_sales', ascending=False)

display(region_sales_profit)
```

index	region	total_sales	total_profit
3	West	725457.8245	
1	East	678781.24	
0	Central	501239.8908	
2	South	391721.905	

Show 25 per page

Next steps: Generate code with region_sales_profit | View recommended plots | New interactive sheet

ng order of sales	

E - Extend (Strategic Application)

Growth opportunities:

- Central Region: This region stands out as having the lowest total sales and total profit, and even a negative average profit margin. The profit per order and profit per quantity are also significantly lower than other regions. This suggests a significant opportunity for improvement in the Central region. Strategies could focus on increasing sales volume, improving pricing or cost structures to increase profit margins, or analyzing specific products or customer segments within this region that are performing poorly.
- South Region: While performing better than the Central region, the South has the second-lowest total sales. There might be opportunities to increase sales volume in this region to bring it closer to the performance of the East and West.
- East and West Regions: These regions are performing well in terms of total sales and profit. Opportunities here might involve sustaining this performance, exploring ways to increase the average profit margin further (especially in the East compared to the West), or identifying successful strategies from these regions that could be replicated in the Central and South.

Overall, the most significant growth opportunities appear to lie in improving the performance of the Central and South regions, particularly addressing the profitability issues in the Central region.

KPI to measure success and right thresholds

If the strategy is to improve performance in the Central and South regions, a key KPI to measure success would be Average Profit Margin in the Central Region.

Currently, the average profit margin in the Central region is negative (-0.104). A suitable threshold could be to aim for a positive average profit margin, perhaps at least matching the average profit margin of the South region (0.164) as a first step, or even striving to reach the levels of the East or West regions in the long term.

Another important KPI could be Total Sales in the Central Region. The current total sales in the Central region are significantly lower than the East and West. A threshold could be set to increase total sales by a certain percentage within a specific timeframe, or to reach a certain proportion of the sales in the East or West regions.

RISKS:

Focusing on improving performance in the Central and South regions, while a potential growth opportunity, also comes with certain risks:

1. Resource Allocation Risk: Shifting focus and resources (like marketing budget, sales team efforts, etc.) to the Central and South regions might dilute efforts in the already high-performing East and West regions, potentially slowing down their growth or even negatively impacting their current performance.
2. Difficulty in Addressing Underlying Issues: The reasons for lower sales and profitability in the Central and South regions might be deeply rooted (e.g., market saturation, strong local competition, unfavorable economic conditions, or specific customer preferences that are hard to shift). Investing heavily without a clear understanding or ability to address these fundamental issues could lead to limited or no improvement.
3. Poor Return on Investment (ROI): There's a risk that the investment in improving these regions might not yield a proportional increase in sales or profit, resulting in a poor return on the resources invested.
4. Opportunity Cost: By focusing on the lower-performing regions, you might be missing out on opportunities to further capitalize on the success in the East and West regions, which could potentially offer higher and faster returns.