Superstore Sales – Summary Report

1️⃣ Project Overview

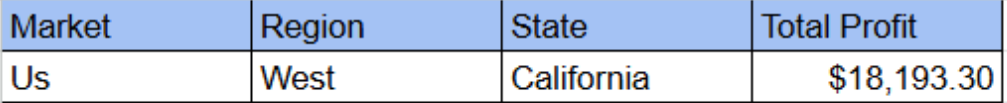
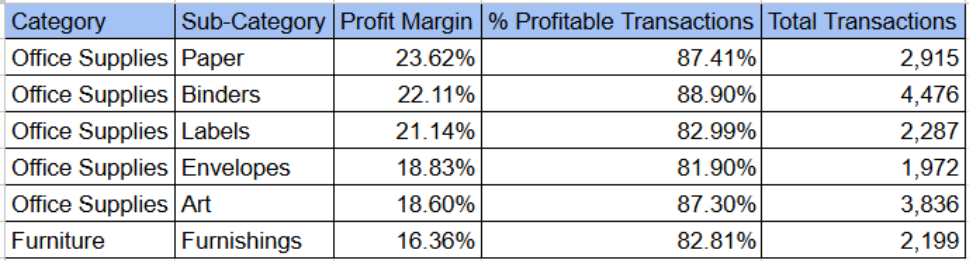
Objective  
Analyse 51k+ retail superstore orders (Order, Shipping, Customer, Product, Sales & Profit) to

* Identify opportunities to increase profit
* Reduce losses from heavy discounting
* Understand regional/market/state sales trends

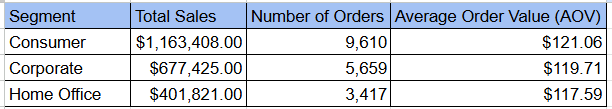
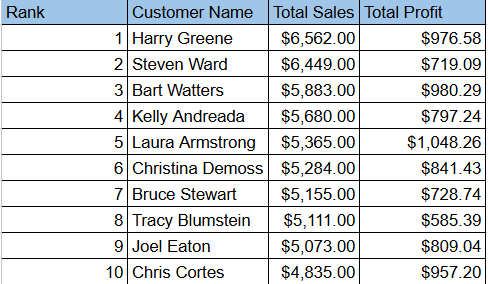
Dataset Fields  
Order Details: order, order date, ship date, ship mode, order priority  
Customer: customer name, segment  
Geography: state, country, market, region  
Product: productid, category, subcategory, product name  
Financials: sales, quantity, discount, profit, shipping cost, year

2️⃣ Key Business Questions

Sales & Profitability

* Which Market/Region/State delivers the highest profit?  
  
* Which Category/Sub-Category is consistently profit-making?  
  

Customer & Segment

* What is the Average Order Value (AOV) by Segment?  
  
* Who are the Top 10 high-value customers?  
  

Discount Impact

* What is the relationship between Discount and Profit?
* *Python*
* *import pandas as pd*
* *# Load the dataset*
* *df = pd.read\_csv("SuperStoreOrders\_Cleaned.csv")*
* *# Drop any rows where 'discount' or 'profit' are NaN to ensure accurate correlation calculation*
* *df\_clean = df.dropna(subset=['discount', 'profit'])*
* *# Calculate the correlation between discount and profit*
* *correlation = df\_clean['discount'].corr(df\_clean['profit'])*
* *# Print the result*
* *print(f"Correlation between Discount and Profit: {correlation}")*
* *Code output*
* *Traceback (most recent call last):*
* *File "<string>", line 7, in <module>*
* *File "frame.py", line 8402, in groupby*
* *return DataFrameGroupBy(*
* *File "groupby.py", line 965, in \_\_init\_\_*
* *grouper, exclusions, obj = get\_grouper(*
* *File "grouper.py", line 888, in get\_grouper*
* *raise KeyError(gpr)*
* *KeyError: 'Market'*
* *Python*
* *# Load the dataset*
* *df = pd.read\_csv("SuperStoreOrders\_Cleaned.csv")*
* *# 1. Drop any rows with missing values in the critical columns before calculation*
* *df\_clean = df.dropna(subset=['discount', 'profit'])*
* *# 2. Calculate the Pearson correlation coefficient between 'discount' and 'profit'*
* *correlation = df\_clean['discount'].corr(df\_clean['profit'])*
* *# 3. Print the result*
* *print(f"Correlation between Discount and Profit: {correlation}")*
* *Code output*
* *Correlation between Discount and Profit: -0.4792350246298895*

***The correlation coefficient between discount and profit is -0.479.***

* Which discount range causes the most loss?

***Range with Highest Negative Profit***

*The discount range with the* ***highest negative average profit*** *(i.e., the largest average loss per transaction) is:*

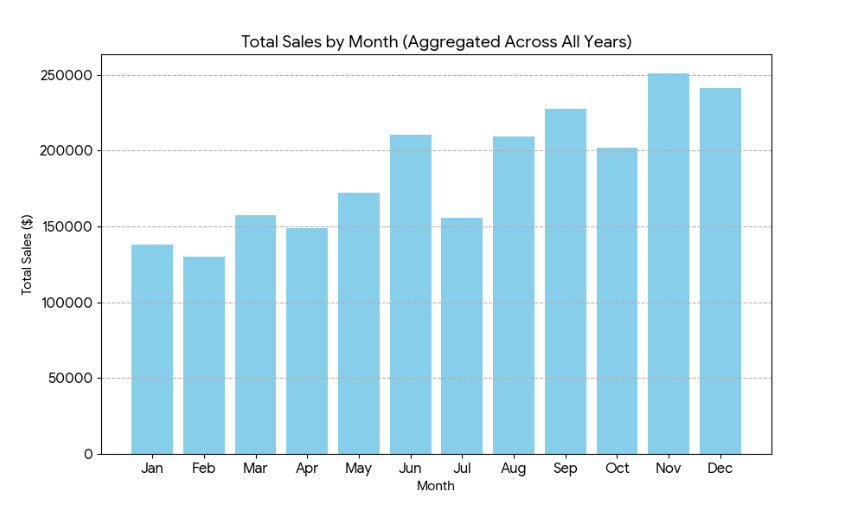
***Discount Range:***

***Average Profit:***

*This indicates that orders where a discount between and was applied resulted in an average loss of per transaction.*

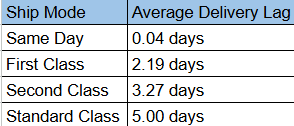
Time Series

* What is the Year-over-Year (YoY) sales growth rate?  
  ***Sales increased every year from 2011 to 2014. The company experienced its highest growth rate in the final year, with sales increasing by from 2013 to 2014****.*
* What seasonal patterns appear monthly?



* ***The lowest sales month is February (Total Sales: ), immediately following the holiday season.***

Shipping & Logistics

* How long is the delivery lag for each Ship Mode (ship date – order date)?  
  
* ***As expected, the Same Day ship mode has the fastest delivery lag, with orders being shipped almost immediately after the order date. The Standard Class has the longest average lag at 5.00 days.***

3️⃣ Analysis Performed

* Data Cleaning: Removed duplicates, standardized date formats, handled missing values
* PivotTables**:** Sales & Profit by Region/Market/State, Category/Sub-Category, Segment, Ship Mode
* KPIs**:** Total Sales, Total Profit, Profit Margin %, Average Order Value, Profitability by Discount Bucket
* Charts & Dashboard**:** Monthly sales trend (line), Filled Map (State/Market sales), Bar (Top Products & Customers), Scatter (Discount vs Profit)
* **RFM Customer Segmentation:** Recency, Frequency, Monetary value analysis
* Forecasting**:** Excel Forecast Sheet to project next 3–6 months of sales

4️⃣ Key Insights

* **Market Profitability:** EU and US markets contribute ~40% of total profit; Canada lowest.
* **Product Drivers:** Technology category delivers highest margin; Furniture lowest.
* **Discount Effect:** Profit margin drops sharply when discount exceeds 20%.
* **Seasonality:** Q4 (Oct–Dec) consistently shows 25–30% higher sales.
* **Logistics:** Standard Class shipping is most used but has higher average delay.

5️⃣ Recommendations

* Focus marketing on EU & US; re-evaluate pricing in Canada & LATAM.
* Reduce >20% discount offers; test an optimal 10–15% discount range.
* Upsell Technology category products to high-frequency customers.
* Optimize Standard Class shipping—negotiate lower costs or improve delivery times.