

Data Case Study

Jessica Martinez

December 2025

1 Executive Summary

Class simulated data analysis project which included cleaning, interpreting and applying data to drive company decision making. The datasets given were incomplete and inconsistent, they required data cleaning and validation before being able to be analyzed. After preparing the data, sales trends and profitability across products, regions, and customers were able to identify opportunities to improve the company's profitability and to help support suggested data-driven business decisions.

2 Introduction

The project included analyzing a set of Excel spreadsheets that contain the company's sales and operations data. This includes customer orders, product pricing, sales regions, product profitability, and sales representative performances. Data given was originally disorganized and contained missing values, duplicates, and inconsistencies. Tariffs were also a factor in decision making as they changed two times during the project, since items were being imported the tariffs increased from 10% to 50% throughout the duration of the project which affected business decisions. Plans for sales growth and increased profitability were decided based on data.

3 Methodology

The analysis began with a comprehensive data quality assessment of multiple Excel datasets, which revealed significant issues including duplicate records, inconsistent geographical information, unclear field definitions, and missing identifiers. Data cleaning steps included removing duplicate orders using unique order IDs, standardized location fields such as city, state, and zip codes, and correcting mislabeled and inaccurate entries. For customer data, duplicate customer IDs without distinguishing attributes were resolved by assigning unique identifiers to maintain the integrity of the record. Additional data reprocessing included renaming columns to reflect business meaning, formatting numerical fields for currency and dates, and implementing formulas to calculate totals and identify anomalies. Once the data was clean and validated, pivot tables and summary calculations were created to evaluate profitability by sales representative, product, region, and customer. This structured approach ensured that subsequent analysis and recommendations were based on accurate and reliable data.

4 Analysis & Insight

Product Profitability

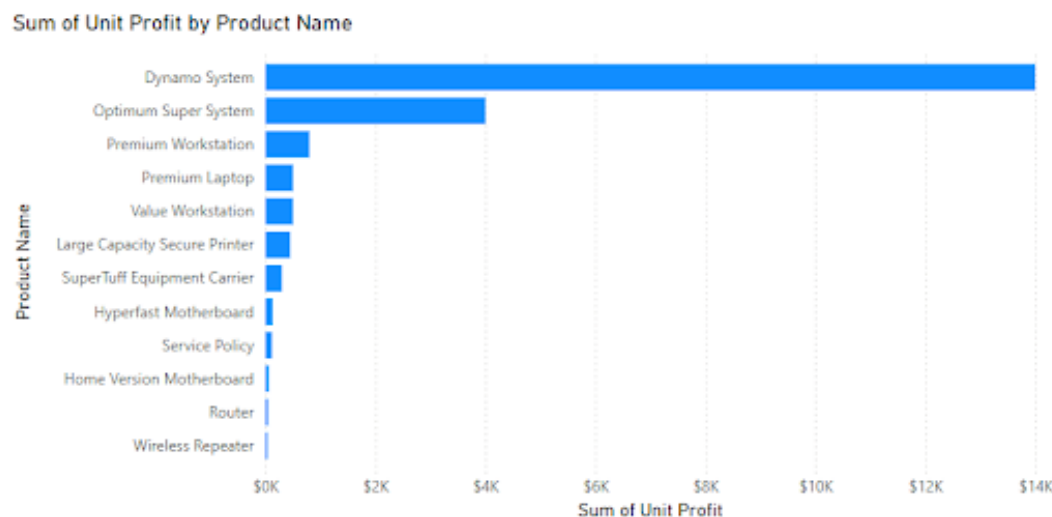


Figure 1: Shows total profit by product, revealing that profitability is highly concentrated in a small number of products. The Dynamo system and Optimum Super System generate a disproportionate share of the total profit.

Profit by Product

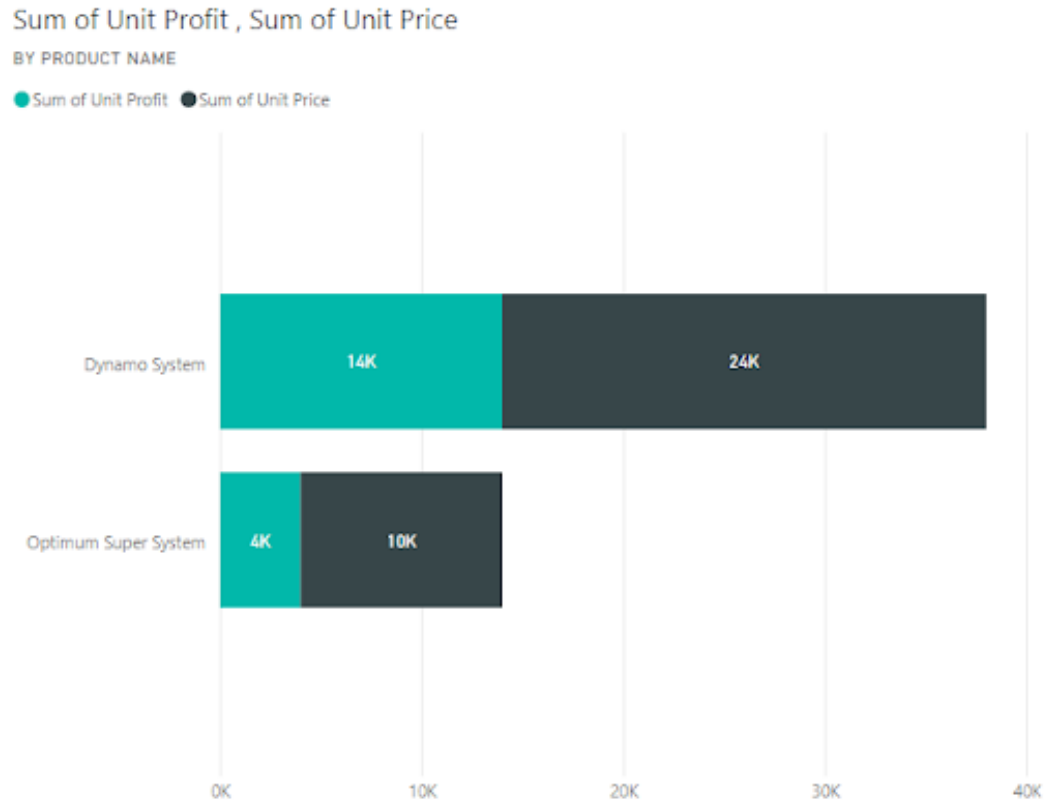


Figure 2: The comparison between the two best selling products highlight the relationship between revenue (unit price) and profit by product. While both the Dynamo system and Optimum Super System generate a high revenue, the Dynamo System produces a higher profit, which indicates a stronger margin and higher profitability.

Profit by Region

Total Profit by State

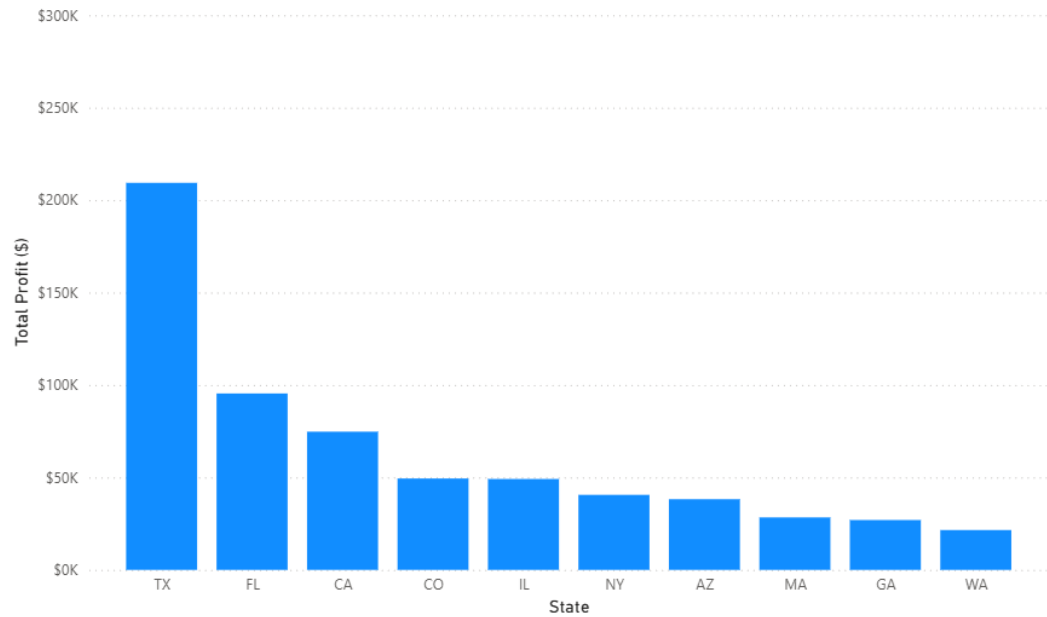


Figure 3: This chart shows total profit by state, revealing a strong geographical concentration of profitability. Texas generates significantly more profit than all of the other regions, followed by Florida and California. This indicates an opportunity to prioritize high-performing regions.

Customer Concentration

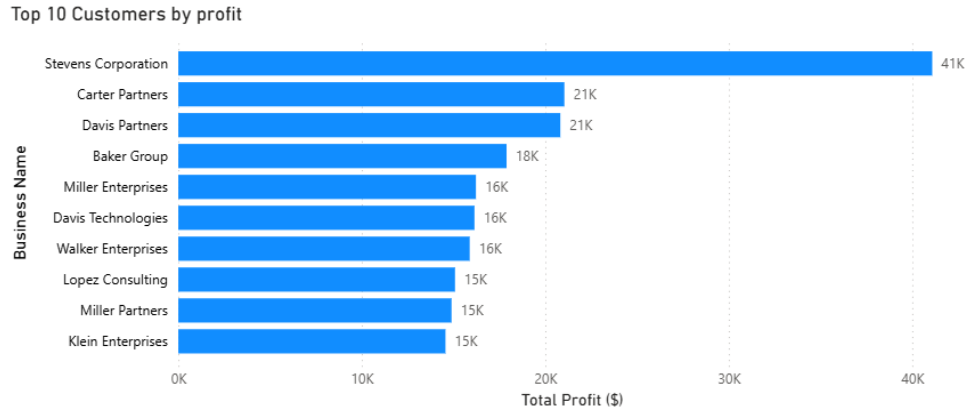


Figure 4: This chart shows the top 10 customers by total profit, revealing a high concentration of profitability among a small group of clients. Steven Corporation alone generates nearly twice the profit of the next highest customer. This highlights both strategic account opportunities but also a customer concentration risk.

5 Business Decisions

Based on the analysis of product profitability, regional performance, and customer concentration, strategic business decisions are recommended to improve overall company profitability and operational efficiency.

The first decision is to prioritize high-margin products, resources, and sales efforts should be focused on most profitable products, particularly the Dynamo System and Optimum Super System, which generate a high share of the total profit. These products should receive priority in marketing, inventory, and sales incentives. Under performing products should be evaluated for cost reductions, repositioning, or potential discontinuation if they consistently fail to generate meaningful profit compared to other products.

The second decision aligns sales strategy with profit, not volume. Sales performance metrics should shift from unit volume and revenue based targets to profit based incentives. This ensures that sales efforts are able to promote high margin products rather than simply high volume sales, which could contribute to limited profitability. Compensation structures and performance evaluations should reward the contribution to profitability rather than the total sales volume alone.

The third decision is to implement regional growth prioritization. Given the strong profitability concentration in specific regions, such as Texas followed by

Florida and California, the company should prioritize investment and expansion efforts in high performing markets. This could include targeted marketing campaigns, increased sales coverage, and localized promotion. Under performing regions should be assessed for operational inefficiencies, pricing strategy gaps, or market fit issues.

The fourth decision is to develop strategic account management for high value customers. High value clients should receive dedicated relationship management, customized offerings, and retention strategies to protect revenue stability while also reducing the dependency risk through broader customer diversification efforts.

6 Impacts & Learnings

flushleft This project demonstrated how transforming unreliable and unorganized raw data into clean structured datasets enables accurate analysis and meaningful decision making. Through systematic data validation and anomaly detection, significant financial distortion caused by duplicates and inconsistent records was able to be identified and corrected. This improved the reliability of profitability reporting. The resulting insight highlighted key profit drivers across products, regions, and customers, which allowed for targeted strategic recommendations rather than generalized revenue growth approaches. Additionally, the project reinforced the importance of strong data governance practices, standardization of data entry, and ongoing quality controls to prevent future errors. Overall, the analysis emphasized that high quality data is foundational to effective business strategy and the focused insight driven decisions can significantly improve operational performance.