

PROJECT REPORT

Key Findings

Sales and Profit Structure

- The top 10 most profitable products contribute only **5.1%** to total positive profit, indicating a *diversified profit base*. In contrast, the top 10 loss-making products are responsible for **18.8%** of total losses, showing **high loss concentration**.
- The business operates on a *high volume of low-value sales*. The profit distribution is centered near zero with a significant negative tail, indicating that *unprofitable transactions* are a frequent and systemic occurrence.
- A *positive correlation* exists between sales and profit, but profit variance increases with sales value (*heteroscedasticity*), meaning **high-value transactions carry higher risk**.

Segment and Category Performance

- The Consumer segment buying Office Supplies generates the highest *absolute profit* (**\$134k**), while the Home Office segment buying Office Supplies is the most *efficient*, with the highest *profit margin* (**21.7%**).
- A data-driven role for each category was identified: **Office Supplies** is the high-efficiency volume driver (low price, high margin), **Technology** is the absolute profit driver (high price, moderate margin), and **Furniture** is an inefficient revenue driver (high price, low margin).

Logistics and Regional Performance

- Standard Class shipping is the dominant method by volume and total profit in all regions. No significant correlation was found between *shipping urgency* and *per-order profitability*.
- The **West** region is the top performer, and national performance is highly dependent on **California**, the most profitable state.
- Two distinct state-level models were observed: a *low-volume/high-margin* model in smaller markets and a *high-volume/lower-margin* model in larger ones.

Discount Impact

- A **strong negative correlation** exists between the Discount level and Total Profit. A profitability "**tipping point**" was identified at a **~20% discount**, beyond which transactions are predominantly unprofitable.
- The most aggressive proportional discounts are applied to the "Binders," "Chairs," and "Bookcases" sub-categories, directly linking the **Furniture** category to *margin erosion*.

Temporal Trends

- Significant business activity in the dataset is confined to the **2014-2017** period.
 - A **strong seasonal pattern** exists, with sales and profit consistently peaking in **Q4** and reaching a trough in Q1.
 - *Year-over-year sales growth* accelerated annually, while profit growth fluctuated, indicating a trade-off between **margin expansion (2015)** and **sales growth prioritization (2016-2017)**.
-

4. Technical Challenges and Solutions

- *Logical data errors*, such as negative sales prices and invalid shipping dates, were corrected by taking the absolute value or nullifying the records.
 - *Erroneous outlier dates* (1999, 2029) were handled by programmatically filtering the dataset to the primary period of activity (2014-2017) before temporal analysis.
 - *Inconsistent categorical data* in the State column was standardized using string stripping and a replacement map, then validated against a canonical list.
 - The original Ship Mode column was found to be *unreliable*, which justified the engineering of a more robust Shipping Urgency feature.
 - An incorrect formula for the Original Price feature was identified and corrected through **visual validation**, highlighting the importance of an iterative analytical workflow.
-

5. Data-Driven Conclusions

1. The discount policy is the single largest driver of unprofitability. Discounts over **20%** consistently lead to net losses.

2. The **Furniture** category is a systemic underperformer, characterized by high price points but the lowest profit margins.
3. The company's financial stability comes from the high-efficiency, high-volume sale of **Office Supplies**.
4. The **Technology** category, while having moderate margins, is the key to generating large absolute profit from high-value sales.