Data Analysis - Report
Review of sales and profitability

Tatiana Bonoeva

Data and Business Analysis Trainee

Date: 14th of Feb

Contents

Report Summary

The objective of this report is to conduct a comprehensive review of the company's sales and profitability, identify key areas for improvement, and propose ways to enhance sales performance.

The data analysis utilized two cleaned Excel datasets: the "Orders" table, comprising 9994 entries, and the "Returns" table, consisting of 296 entries.

The analytical process began with a high-level descriptive analysis, which provided an initial understanding of the core statistical properties of the numerical variables and helped to identify primary areas for in-depth investigation.

Subsequent analyses focused on profitability, revealing critical weaknesses in current operations. The final phase of the analysis examined product returns, highlighting specific items and locations that were associated with subpar quality or service issues.

Drawing upon the comprehensive analyses conducted, the report outlines 5 recommendations to address the identified challenges.

Data Preprocessing

To begin the analysis, it is convenient to merge the tables. This was accomplished by adding a new column named 'Returns' to the 'Orders' table. The column was populated with two possible values: 'Yes' if the product was returned and 'No' if it was not. This was achieved using the following formula:

```
=IF(ISNA(MATCH(B2;Returns!$A$2:$A$297;0)); "No"; "Yes")
```

Descriptive Analysis

Starting the analysis with basic statistics (totals, averages, minimums, maximums) for sales, quantity, discount, and profit to get a high level of understanding of the data (Table 1). The visual representation of data distributions is in Appendix 1.

	Average	Mode	Median	Standard Deviation	MAX	MIN
	AVERAGE (number1, [number2], ...)	MODE.SNGL (number1, [number2],...)	MEDIAN (number1, [number2],...)	STDEV (number1, [number2],...)	MAX (number1, [number2],...)	MIN (number1, [number2],...)
Sales	229.8032	12.96	54.376	623.422931	22638.48	0.444
Quantity	3.7886	3	3	2.22497616	14	1
Discount	0.1562	0	0.2	0.20650573	0.8	0
Profit	28.6332	0	8.64135	234.327679	8399.976	-6599.978

Table 1. Basic statistics for sales, quantity, discount, and profit.

Based on the calculated statistics values the following summary of each of the parameters can be made:

Sales: The dataset reveals a substantial range in transaction values, with an average sale of approximately $229.80, countered by a prevalent transaction amount at a mode of $12.96. This contrast suggests a heavy volume of low-value sales alongside fewer high-value purchases, as evidenced by a median of $54.38. The significant standard deviation of $623.42 points to a diverse sales profile, extending to a maximum sale of $22,638.48.

Quantity: The average quantity per transaction stands at 3.79, with the most common quantity being 3. This indicates a consistent purchasing pattern. The variation in quantity is moderate, with a standard deviation of 2.22, and a range from 1 to 14 units.

Discount: Discounts show a low average of 0.16 with a minimal standard deviation, reflecting a restrained discounting strategy. The mode at 0 and the median of 0.2 suggest that discounts are sparingly applied, with the highest discount given is 0.8.

Profit: Profits average at $28.63, but the loss recorded at a minimum of-$6599.98 indicates instances of significant negative impact on profitability. The maximum profit of $8399.976 alongside a standard deviation of $234.33 denotes a volatile profit landscape.

Relying on the overall descriptive data analysis the following conclusions can be made:

- The company's revenue largely comes from selling a **high number of items at low prices**.
- Sales and profit data show a mix of frequent small-profit transactions and occasional large, profitable sales.
- There are **loss-leading products** that make the overall profit close to 0.
- The company has a restrained discount strategy.

Trend Analysis

The most critical factor potentially influencing sales success is specific periods. To determine whether these periods significantly impact the company, scatter charts depicting the distribution of sales and profits were constructed (Figure 1).

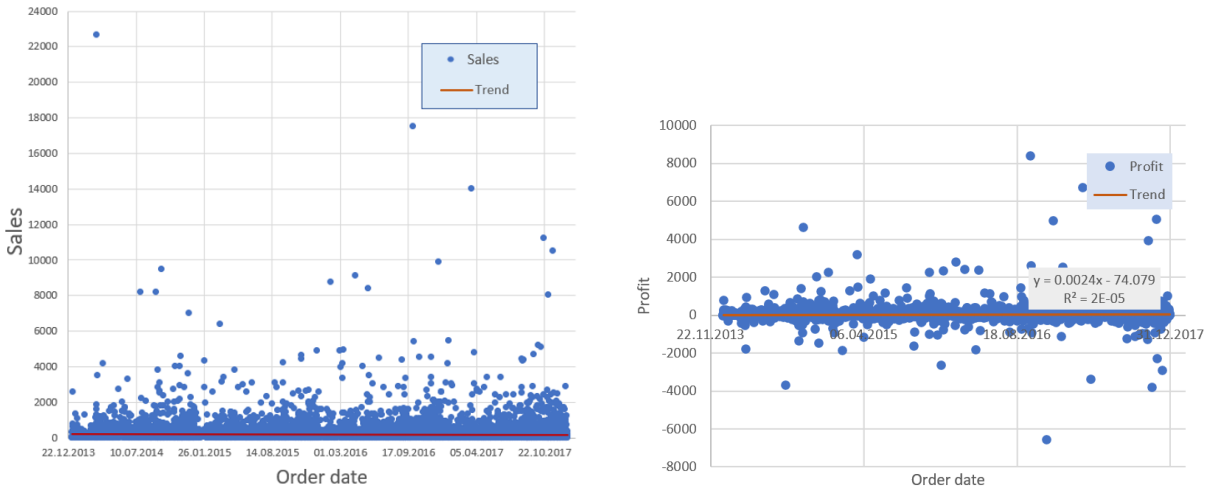


Figure 1. Right: Sales scatter plot over order time. Left: Profit scatter plot over order time.

The performance over time does not visually display any differences. To verify this observation statistically, trends were analyzed using linear regressions, which are represented by red lines on the plots. These trends indicate a straight-line relationship between sales (or profit) and the order date.

Consequently, it can be concluded that **there are no distinct profitable or non-profitable seasons for the company**, as time does not appear to significantly impact these parameters.

Profitability Analysis

The primary focus is to identify the reasons behind negative profits. To achieve this, the dataset was filtered to select only the entries with negative values using the method illustrated in Figure 2.

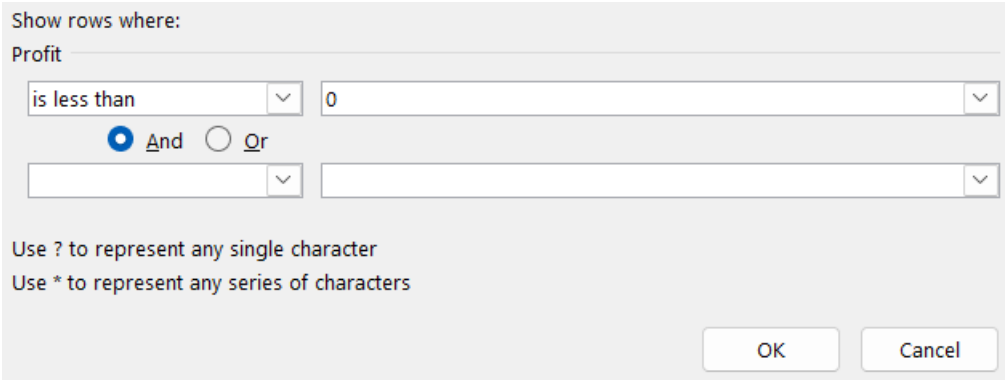


Figure 2. A method for filtering rows in a column.

The total amount of non-profitable sales is equal to 1871 (18% of the total sales) and is calculated by the following formula:

```
=COUNTIF(U8:U9994; "<0")
```

To find categories that do not perform well, a pie chart was chosen as an optimal tool for representing the amount of each category in the total unprofitable outcome.

	COUNT	%
	COUNTIF(O:O; "Category Name")	Category/Total
Furniture	714	38
Office Supplies	886	47
Technology	271	14
Total	1871	100

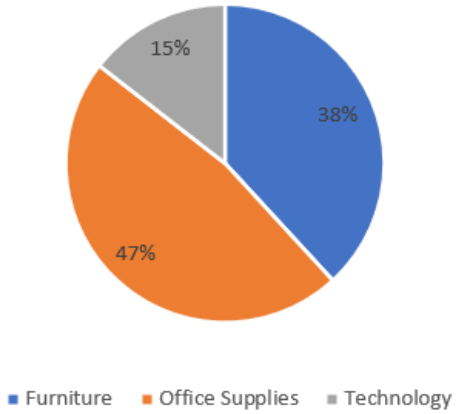
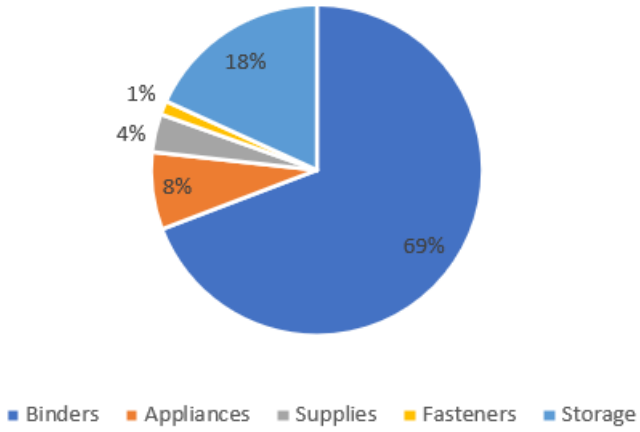


Figure 3. Right: Pie chart of percentages of non-profitable categories. Left: Proportions calculation.

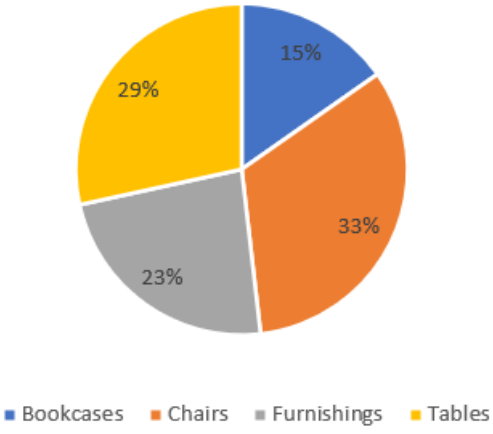
From Figure 3 two categories- Office Supplies and Furniture can be selected for further investigation having the highest amount of non-profitable sales 38% and 47% respectively.

	COUNTIF(P:P; "Sub - Category Name")
Binders	613
Appliances	67
Supplies	33
Fasteners	12
Storage	161
Total	886



a) Pie chart of non-profitable sub-categories of the Office Supplies category

	COUNTIF(P:P; "Sub -Category Name")
Bookcases	109
Chairs	235
Furnishings	167
Tables	203
Total	714



b) Pie chart of non-profitable sub-categories of the Furniture category

Figure 4. Right: Pie charts of non-profitable sub-categories. Left: Proportions calculation for the pie charts.

Relying on the achieved visual data representation (Figure 4) sub-category **Binders shows a highly unsuccessful performance** by 613. The subcategories of Furniture show almost equal numbers around 25% of the subcategories.

Also, it's important to analyze the performance by region. The quicker method – Pivot table was chosen to provide the analytical insight (Figure 5).

Count of City	Column Labels				
Row Labels	Central	East	South	West	Grand Total
Arizona				85	85
California				105	105
Colorado				66	66
Connecticut		2			2
Delaware		2			2
Florida			126		126
Illinois	255				255
Maryland		2			2
Massachusetts		5			5
Nevada				1	1
New Hampshire		1			1
New Jersey		2			2
New Mexico				1	1
New York		47			47
North Carolina			68		68
Ohio		209			209
Oregon				43	43
Pennsylvania		278			278
Rhode Island		4			4
Tennessee			65		65
Texas	486				486
Washington				17	17
West Virginia		1			1
Grand Total	**741**	**553**	**259**	**318**	**1871**

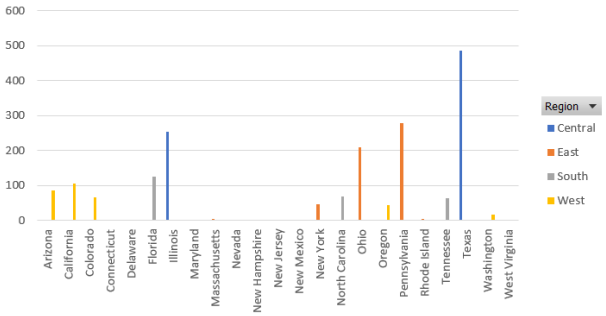


Figure 5. Left: Pivot table of regions, states, and cities. Right: Clustered Column Chart.

Relying on the data, **Texas state shows the highest amount of non-profitable results**, also three more states can be selected: Illinois, Ohio, and Pennsylvania. All the states are located in Central and East regions that show the most unpleasant performance.

Sales Analysis

To find out how to increase sales, it's important to check whether or not there is a correlation between Sales and Profit values. One of the approaches to estimate the correlation is linear regression analysis (Figure 6).

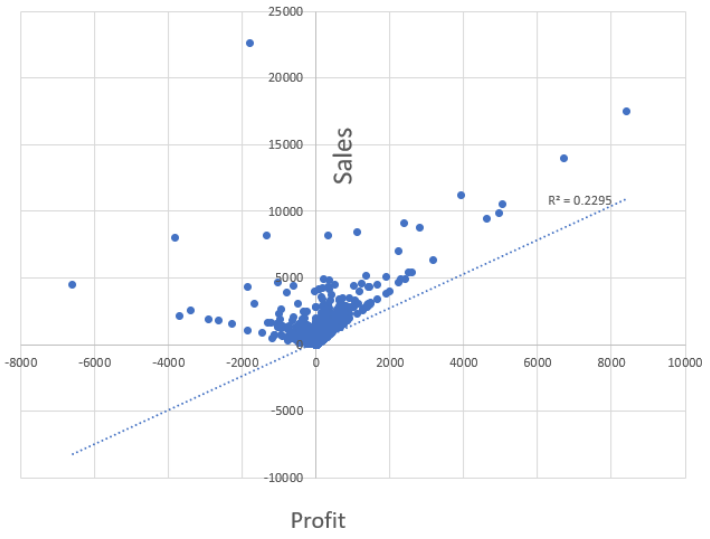


Figure 6. Graph of Sales Relative to Profit. R^2 = 0.23, correlation = 0.48.

According to the R squared value from linear regression, the correlation is positive and equal to 0.48 which means that one variable affects another. Consequently, **if profit increases by implementing the above insights, sales are also likely to increase.**

The next stage of understanding the sales performance and factors that help to improve it is looking at categories and region (states) performance (Figures 7, 8, 9).

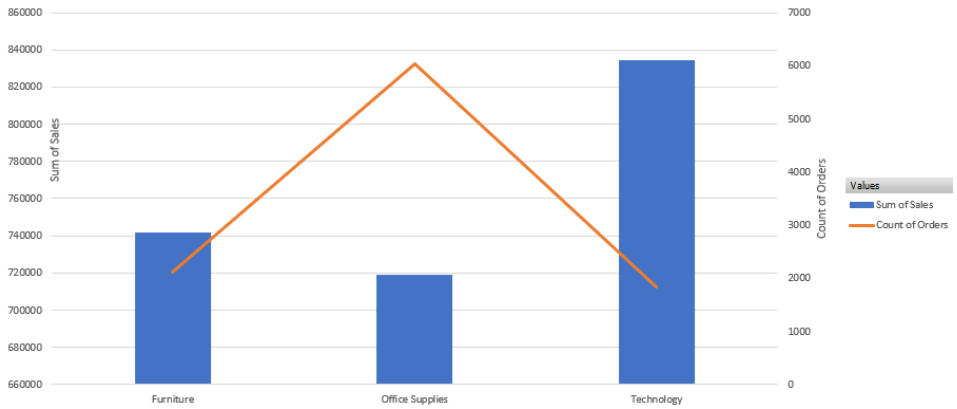


Figure 7. Pivot Char- Combo (Column Clustered and Line) of categories.

For categories, there is a high demand for Office Supplies which has the lowest value of the sum of sales, whereas the opposite situation for Technology (Figure 7).

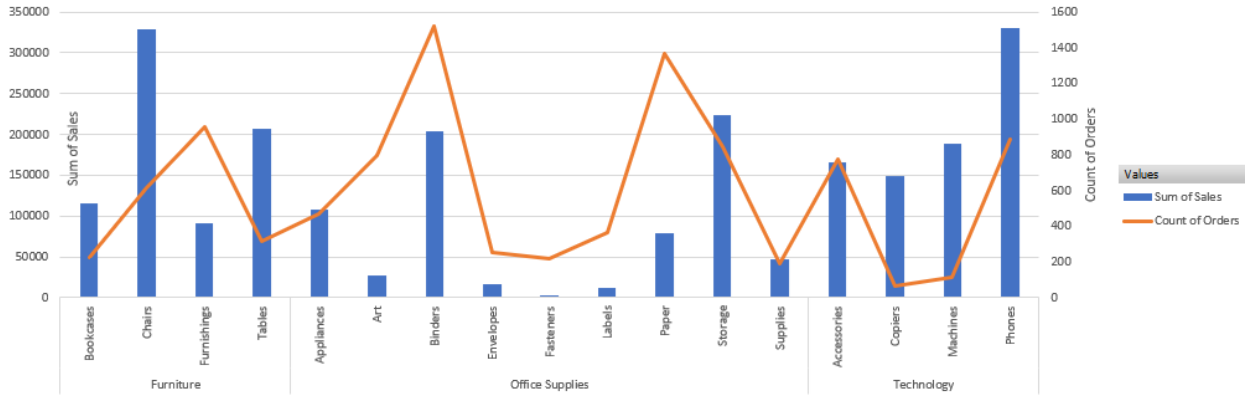


Figure 8. Pivot Char- Combo (Column Clustered and Line) of sub-categories.

Further analysis of the data presented in Figure 8 reveals that **enhancing the sales of Art and Paper products is likely to boost the overall sales of Office Supplies**. Additionally, since **Copiers and Machines are the least in demand, focusing on increasing their sales could positively impact total profits**.

To identify successful regions (states), a filter was applied to select only those with values greater than 600, as determined by the sales histogram in the Appendix. After filtering, the total number of rows amounted to 944. This refined dataset was then utilized to determine which states demonstrate the most favorable outcomes, as shown in Figure 9.

Count of City	Column Labels				
Row Labels	Central	East	South	West	Grand Total
Alabama			11		11
Arizona				13	13
Arkansas			7		7
California				191	191
Colorado				13	13
Connecticut		6			6
Delaware		6			6
District of Columbia		2			2
Florida			32		32
Georgia			18		18
Idaho				2	2
Illinois	37				37
Indiana	18				18
Iowa	2				2
Kentucky			20		20
Louisiana			4		4
Maryland		10			10
Massachusetts		14			14
Michigan	24				24
Minnesota	12				12
Mississippi			3		3
Missouri	9				9
Montana				2	2
Nebraska	3				3
Nevada				8	8
New Hampshire		4			4
New Jersey		10			10
New Mexico				1	1
New York		127			127
North Carolina			22		22
North Dakota	1				1
Ohio		34			34
Oklahoma	10				10
Oregon				6	6
Pennsylvania		43			43
Rhode Island		11			11
South Carolina			5		5
Tennessee			15		15
Texas	76				76
Utah				6	6
Vermont		4			4
Virginia			25		25
Washington				57	57
West Virginia		1			1
Wisconsin	18				18
Wyoming				1	1
Grand Total	210	272	162	300	944

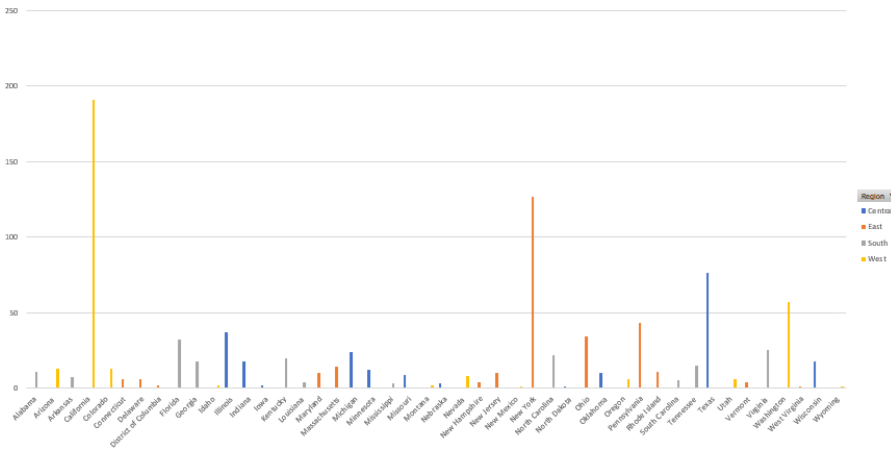


Figure 9. Left: Pivot table considering regions, states, and cities. Right: Clustered Column Chart.

Based on Figure 9, **California and New York can be identified as the primary states for sales focus**.

Returns Analysis

Returns influence the overall profit (sales) performance which also should be considered. The total amount of returns equals 800 which is approximately 8% of overall sales.

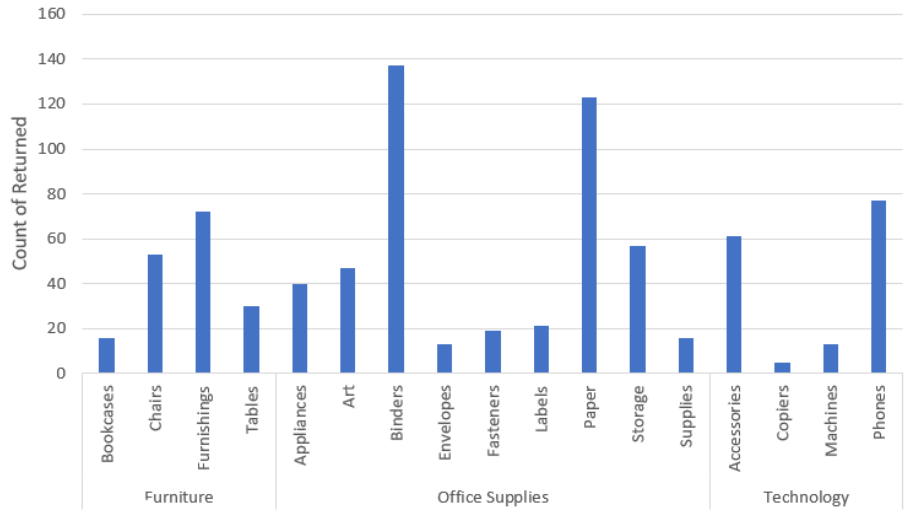


Figure 10. Clustered Column Chart of Pivot Table of sub-categories of the amount of returned products

Based on the data breakdown by sub-categories and states (Figures 10 and 11), the following observations can be highlighted:

- Binders and Paper have the highest return rates.
- Three states show the highest number of returns: New York, and Washington, with California recording the largest number at 338 returns.

These insights are valuable for identifying potential improvements in product quality or customer service.

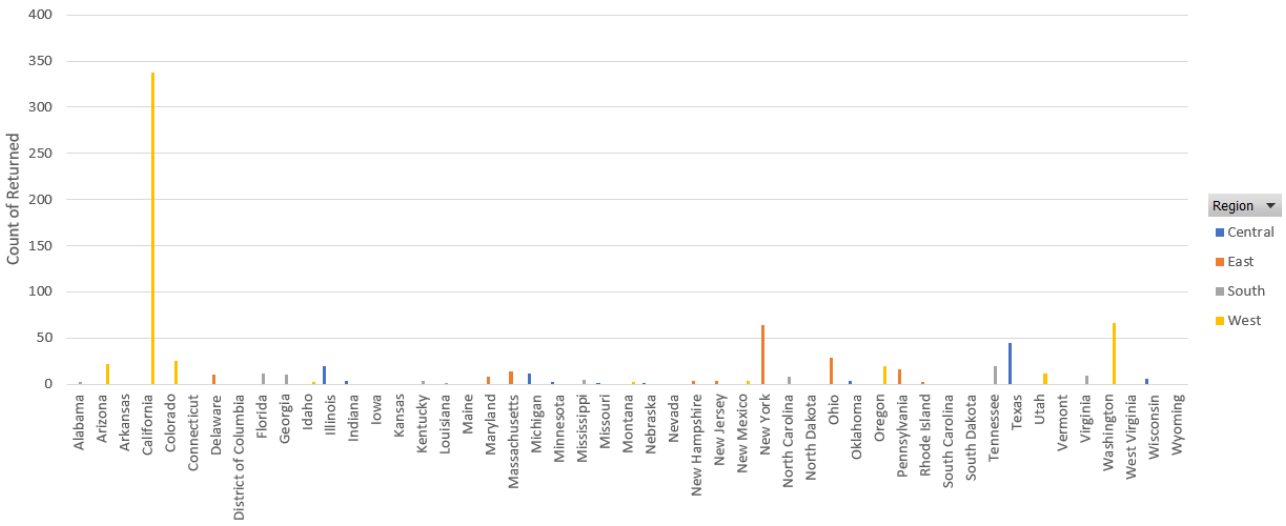


Figure 11. Clustered Column Chart of Pivot Table of states of the amount of returned products

Recommendations

Upon thorough analysis of the data, we propose the following strategic actions to enhance sales performance:

1) Discontinue the sale of Binders due to their high return rates and non-profitable results.
2) Re-evaluate Texas as a primary market, as current data suggests it may not be as profitable.
3) Prioritize California and New York as key markets, while ensuring the delivery of high-quality products and services.
4) Implement a price increase for Art and Paper products to leverage their strong market performance.
5) Consider reducing the prices of Copiers and Machines to stimulate demand and offset lower sales volumes.

For more nuanced and in-depth recommendations, a further detailed investigation is necessary, encompassing additional data and extending the analysis timeframe.

Appendix. Histograms of sales, quality, discount, and profit.

Distribution of sales, discounts, quantity, and profit is represented in the following histograms