

A Report on

Vendor Performance Analysis with a huge dataset

Project Overview

This project analyzes vendor and brand performance using six large datasets, including one with more than twelve million sales records. It explores how sales performance, purchasing behavior, and inventory movement influence overall profitability, helping identify where value is created or lost and where operational improvements can drive better margins.

The analysis was performed in Jupyter Notebook using a MySQL server to handle the high-volume data. After building a consolidated summary dataset, the key insights were visualized in Power BI to present a clear, business-ready view of the findings.

Dataset Summary

Feature Datasets with file size:

- Begin Inventory:
 - inventory ID, store, city, brand, description, size, onHand, price, start date
 - 9*columns - 206529*rows
 - End Inventory:
 - inventory ID, store, city, brand, description, size, onHand, price, end date
 - 9*columns - 224489*rows
 - Purchase Prices:
 - Brand, description, price, size, volume, classification, purchase price, vendor number, vendor name
 - 9*columns - 12261*rows
 - Purchases:
-

- Inventory ID, store, brand, description, size, vendor number, vendor name, PO number, PO date, receiving date, invoice date, pay date, purchase price, quantity, dollars, classification
 - 16*columns - 2372474*rows
- Sales:
 - Inventory ID, store, brand, description, size, sales quantity, sales dollars, sales price, sales date, volume, classification, excise tax, vendor number, vendor name
 - 14*columns - 12825363*rows
- Vendor Invoice:
 - Vendor number, vendor name, invoice number, PO number, PO date, pay date, quantity, dollars, freight, approval
 - 10*columns - 5543*rows

Prepared a final dataset as a summary data across all 6 files, “vendor_sales_summary” with 18*columns & 10692*rows

With values: vendor number, vendor name, brand, description, purchase price, actual price, volume, total purchase quantity, total purchase dollars, total sales quantity, total sales dollars, total sales price, total excise tax, freight cost, gross profit, profit margin, stock turnover, sales to purchase ratio

Missing Data: 178 null values in final summary dataset

Tools Used: Python (Numpy, Pandas, Matplotlib, Seaborn), MySQL, Power BI

Complete EDA Analysis using Python

The exploratory data analysis was carried out by connecting MySQL to Jupyter Notebook through pandas. Initial samples from each dataset were examined to understand the structure, assess data quality, and identify fields relevant to the business problem.

After reviewing the first and last few records and validating key attributes, a consolidated summary table was created in Python using pandas, integrated directly with the MySQL server.

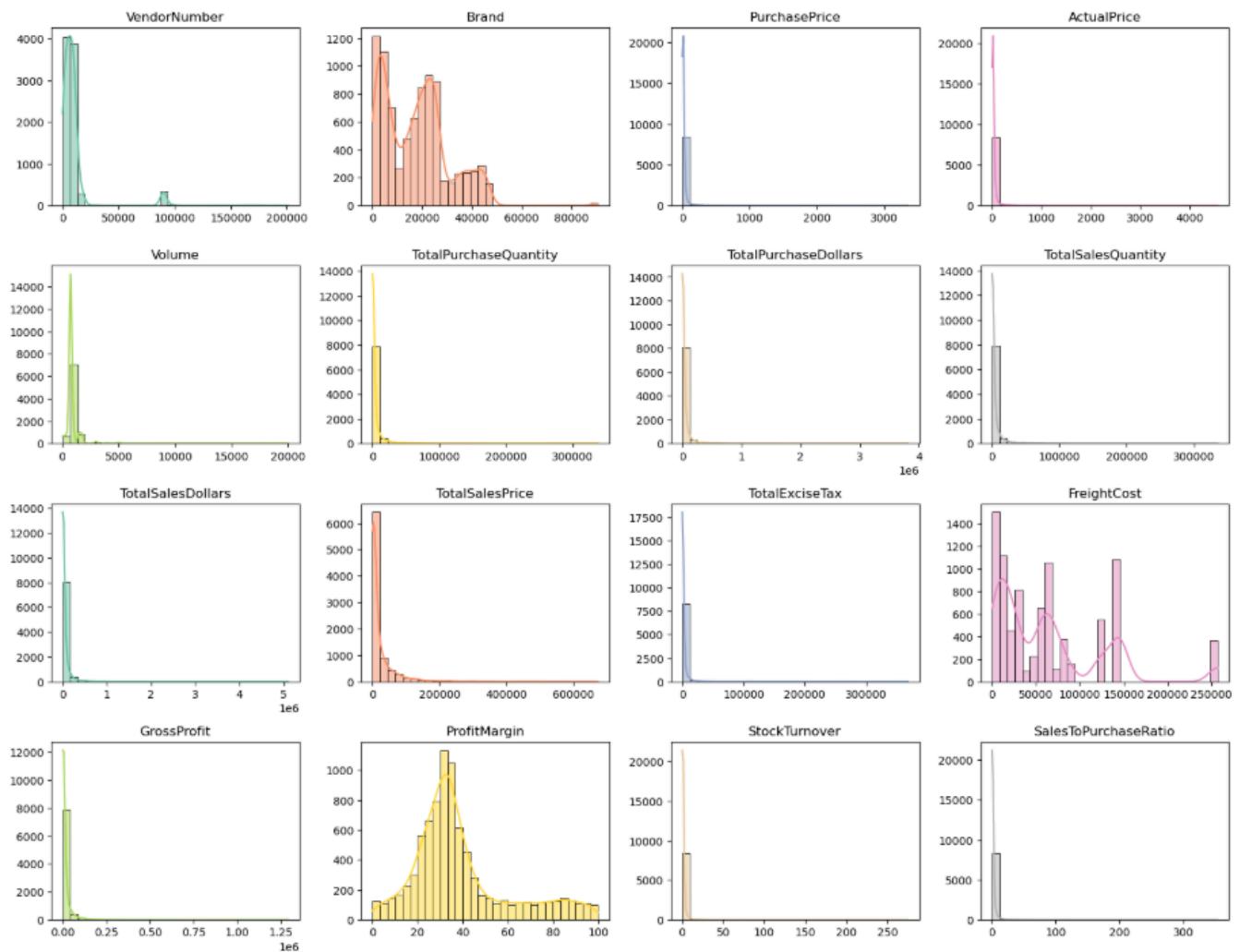
All further analysis was performed on this refined dataset to ensure accuracy, consistency, and efficient processing of large volumes of data.

Summary Statistics for the final Dataset

	count	mean	std	min	25%	50%	75%	max
VendorNumber	10692.0	10650.649458	18753.519148	2.000000	3951.000000	7153.000000	9552.000000	2.013590e+05
Brand	10692.0	18039.228769	12662.187074	58.000000	5793.500000	18761.500000	25514.250000	9.063100e+04
PurchasePrice	10692.0	24.385303	109.269375	0.360000	6.840000	10.455000	19.482500	5.681810e+03
ActualPrice	10692.0	35.643671	148.246016	0.490000	10.990000	15.990000	28.990000	7.499990e+03
Volume	10692.0	847.360550	664.309212	50.000000	750.000000	750.000000	750.000000	2.000000e+04
TotalPurchaseQuantity	10692.0	3140.886831	11095.086769	1.000000	36.000000	262.000000	1975.750000	3.376600e+05
TotalPurchaseDollars	10692.0	30106.693372	123067.799627	0.710000	453.457500	3655.465000	20738.245000	3.811252e+06
TotalSalesQuantity	10692.0	3077.482136	10952.851391	0.000000	33.000000	261.000000	1929.250000	3.349390e+05
TotalSalesDollars	10692.0	42239.074419	167655.265984	0.000000	729.220000	5298.045000	28396.915000	5.101920e+06
TotalSalesPrice	10692.0	18793.783627	44952.773386	0.000000	289.710000	2857.800000	16059.562500	6.728193e+05
TotalExciseTax	10692.0	1774.226259	10975.582240	0.000000	4.800000	46.570000	418.650000	3.682428e+05
FreightCost	10692.0	61433.763214	60938.458032	0.090000	14069.870000	50293.620000	79528.990000	2.570321e+05
GrossProfit	10692.0	12132.381048	46224.337964	-52002.780000	52.920000	1399.640000	8660.200000	1.290668e+06
ProfitMargin	10514.0	-15.885227	447.289882	-23730.638953	15.353839	30.778375	40.210967	9.971666e+01
StockTurnover	10692.0	1.706793	6.020460	0.000000	0.807229	0.981529	1.039342	2.745000e+02
SalesToPurchaseRatio	10692.0	2.504390	8.459067	0.000000	1.153729	1.436894	1.665449	3.529286e+02

The summary statistics for most numerical fields of the final dataset show most fields show a wide range of values. This points to vendors of very different sizes, pricing structures, and sales activity. The numbers also suggest noticeable differences in how products move through the pipeline, from purchasing to sales.

Visual Representation of the dataset



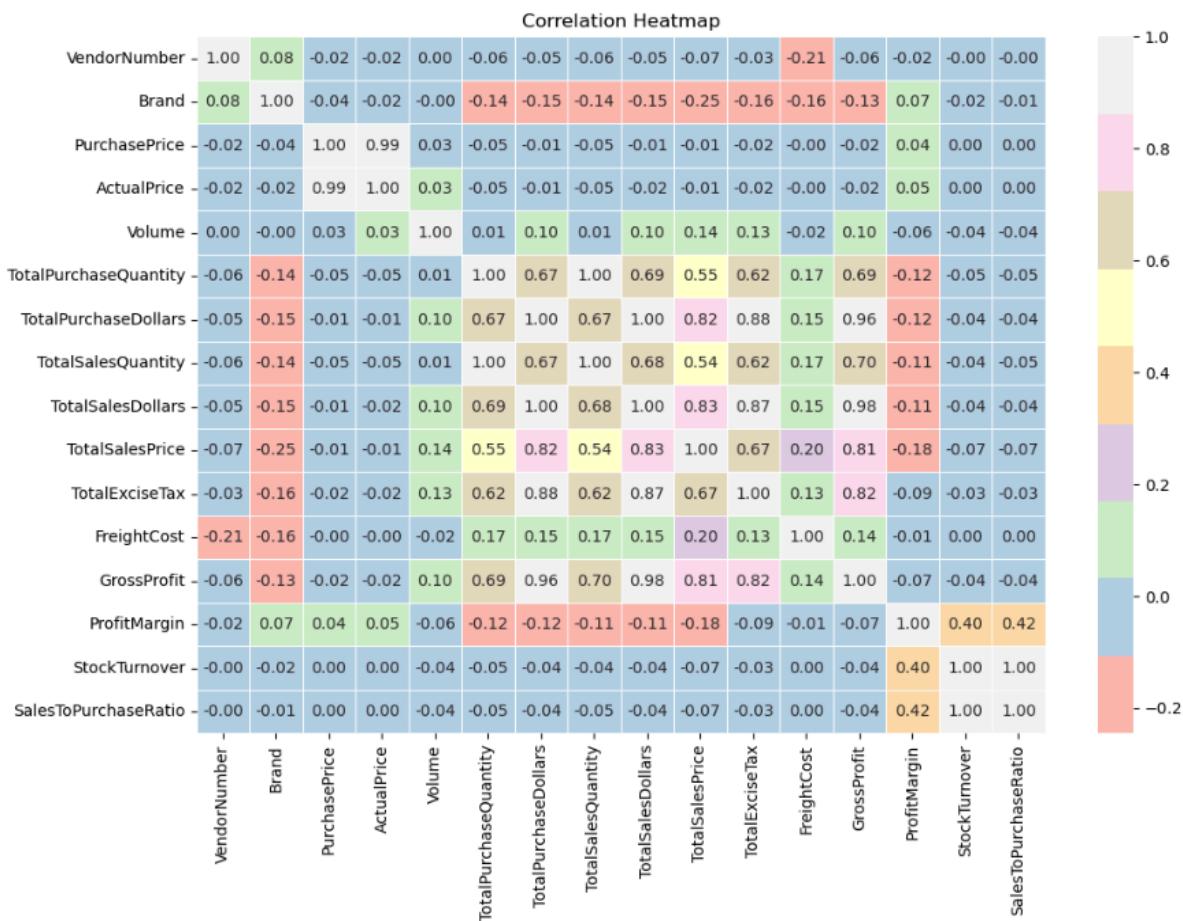
These summary statistics highlight a few areas that need attention:

- Negative gross profit values are detected, which means **certain products** are being **sold below cost**.
- Profit margin also drops to negative infinity in cases where revenue is zero, pointing to unprofitable or **mispiced items**.
- Sales quantity and revenue hitting zero suggest inventory that never moved, which may signal obsolete or **mispiced stock**.

- On the higher end, large gaps between the mean and maximum values for **purchase price**, **actual sale price**, and **freight cost indicate outliers**. These could be premium products or signs of inconsistent pricing and logistics spending.
- Freight costs in particular show wide variation, which may reflect **inefficiencies in shipment planning**.
- Stock turnover ranges from zero to extremely high levels. **Some items sell quickly, while others sit untouched**.
- Values above one** show that sales are being fulfilled from older inventory, which is fine operationally but should be **monitored to avoid stockouts** or inaccurate forecasting.

Overall, the data points to a mix of pricing gaps, slow-moving inventory, and operational inconsistencies that merit deeper review.

Correlation Overview



The correlation checks show that product pricing has little influence on revenue or profit.

The strongest relationship is between **purchase quantity** and **sales quantity**, which confirms that inventory is moving as expected.

This visual representation of correlation also confirms that **higher selling prices** tend to **reduce margins**, likely due to competitive pressures.

Finally, **turnover speed** shows almost **no link to profitability**, suggesting that *fast-moving products are not always the most profitable*.

Business questions with deliverables

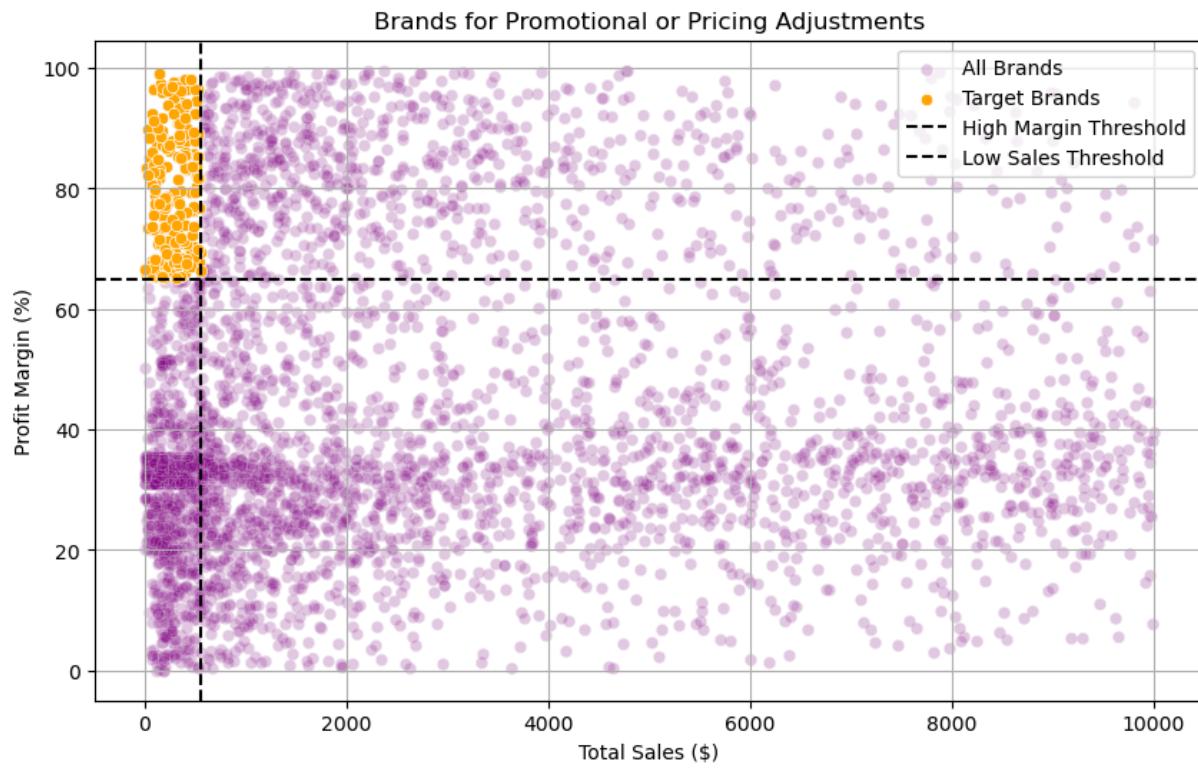
Effective inventory and sales management are critical for optimizing profitability in the retail and wholesale industry. Companies need to ensure that they are not incurring losses due to inefficient pricing, poor inventory turnover, or vendor dependency. The goal of this analysis is to:

01 Brands that need Promotional or Pricing Adjustments which exhibit lower sales performance but higher profit margins.

Brands with Low Sales but High Profit Margins:

	Description	TotalSalesDollars	ProfitMargin
6199	Santa Rita Organic Svn Bl	9.99	66.466466
2369	Debauchery Pnt Nr	11.58	65.975820
2070	Concannon Glen Ellen Wh Zin	15.95	83.448276
2188	Crown Royal Apple	27.86	89.806174
6237	Sauza Sprklg Wild Berry Marg	27.96	82.153076
...
5074	Nanbu Bijin Southern Beauty	535.68	76.747312
2271	Dad's Hat Rye Whiskey	538.89	81.851584
57	A Bichot Clos Marechaudes	539.94	67.740860
6245	Sbragia Home Ranch Merlot	549.75	66.444748
3326	Goulee Cos d'Estournel 10	558.87	69.434752

198 rows × 3 columns

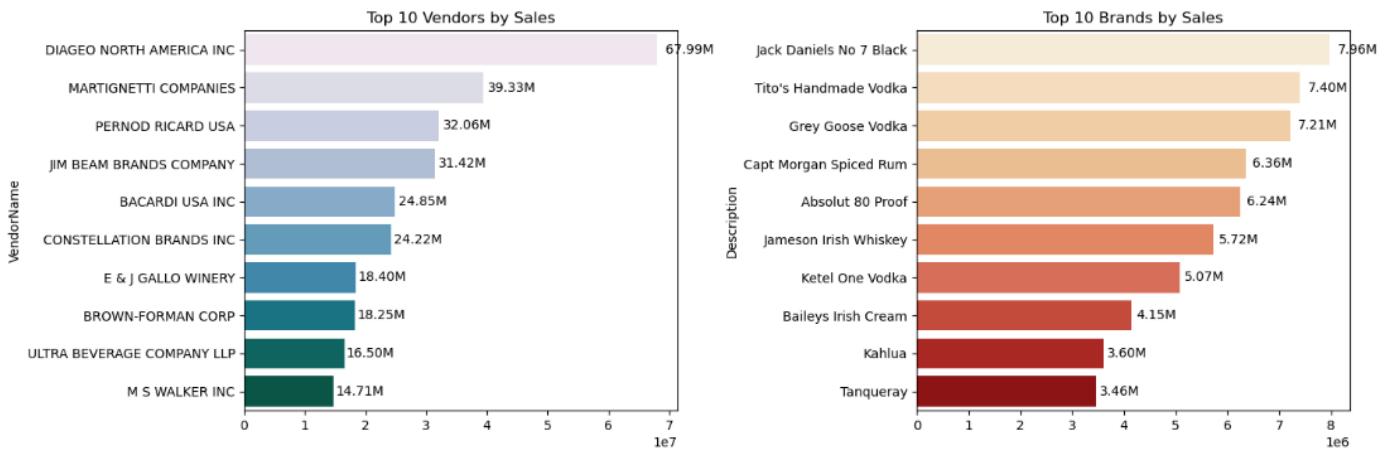


The insights clearly shows that **198 brands aren't selling in high volumes but are still delivering strong margins**. This group stands out as a growth opportunity.

With the right push through marketing or calibrated pricing, these brands can drive higher sales while keeping their profitability intact.

02_Vendors and brands demonstrating highest sales performance

VendorName	TotalSalesDollars
DIAGEO NORTH AMERICA INC	67.99M
MARTIGNETTI COMPANIES	39.33M
PERNOD RICARD USA	32.06M
JIM BEAM BRANDS COMPANY	31.42M
BACARDI USA INC	24.85M
CONSTELLATION BRANDS INC	24.22M
E & J GALLO WINERY	18.40M
BROWN-FORMAN CORP	18.25M
ULTRA BEVERAGE COMPANY LLP	16.50M
M S WALKER INC	14.71M
Name: TotalSalesDollars, dtype: object	

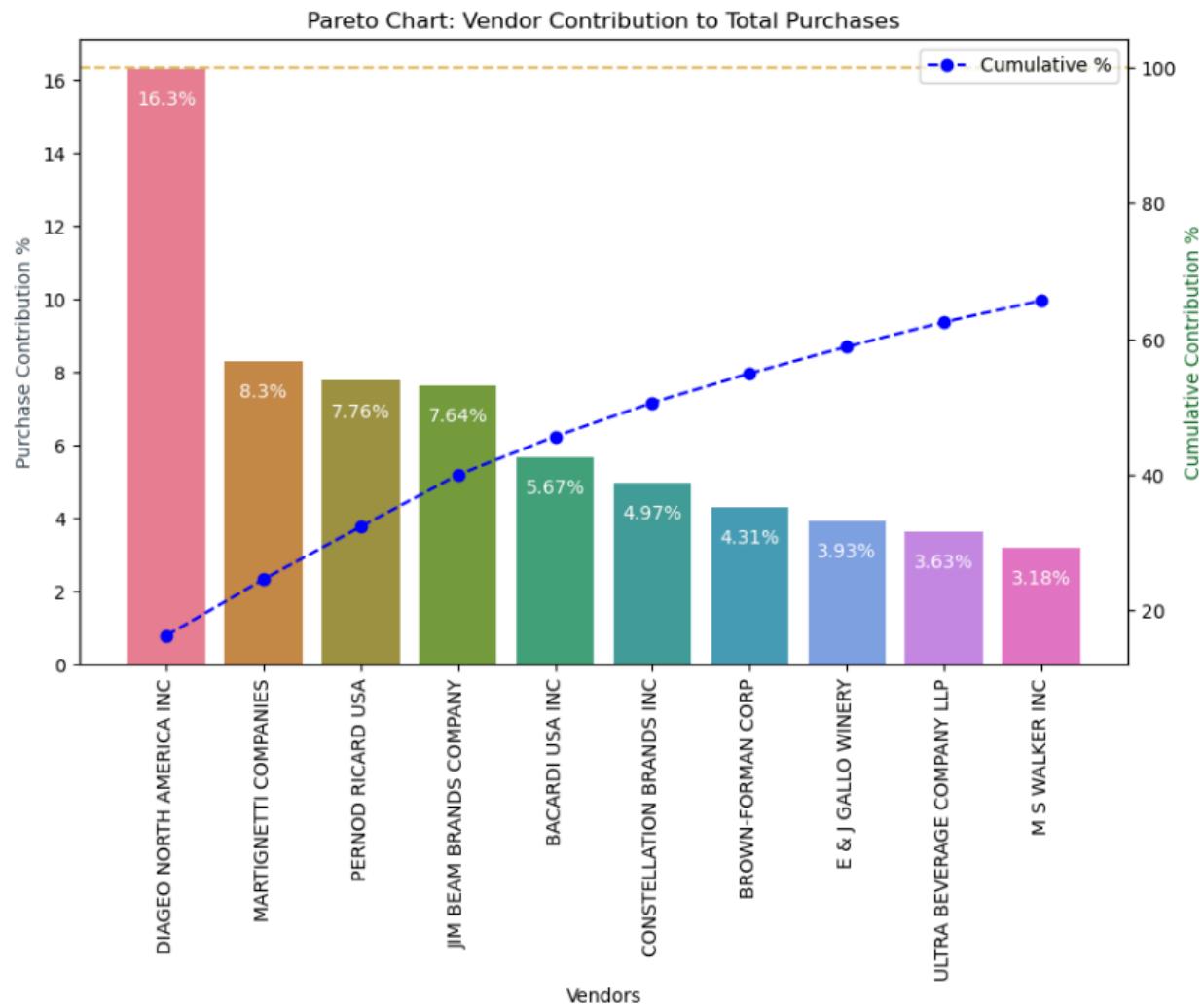


- Figure shows top brands & vendors by sales performance

The sales distribution shows that a small group of vendors and brands drive a significant share of revenue. DIAGEO North America leads vendor sales by a large margin, while Jack Daniel's No. 7 tops brand-level performance. This concentration highlights where customer demand is strongest and where strategic partnerships are most valuable.

03_Vendors contributing the most to total purchase dollars

	VendorName	TotalPurchaseDollars	GrossProfit	TotalSalesDollars	Purchase_Contribution%
25	DIAGEO NORTH AMERICA INC	50.10M	17.89M	67.99M	16.30
57	MARTINETTI COMPANIES	25.50M	13.83M	39.33M	8.30
68	PERNOD RICARD USA	23.85M	8.21M	32.06M	7.76
46	JIM BEAM BRANDS COMPANY	23.49M	7.93M	31.42M	7.64
6	BACARDI USA INC	17.43M	7.42M	24.85M	5.67
20	CONSTELLATION BRANDS INC	15.27M	8.95M	24.22M	4.97
11	BROWN-FORMAN CORP	13.24M	5.01M	18.25M	4.31
30	E & J GALLO WINERY	12.07M	6.33M	18.40M	3.93
106	ULTRA BEVERAGE COMPANY LLP	11.17M	5.34M	16.50M	3.63
53	M S WALKER INC	9.76M	4.94M	14.71M	3.18



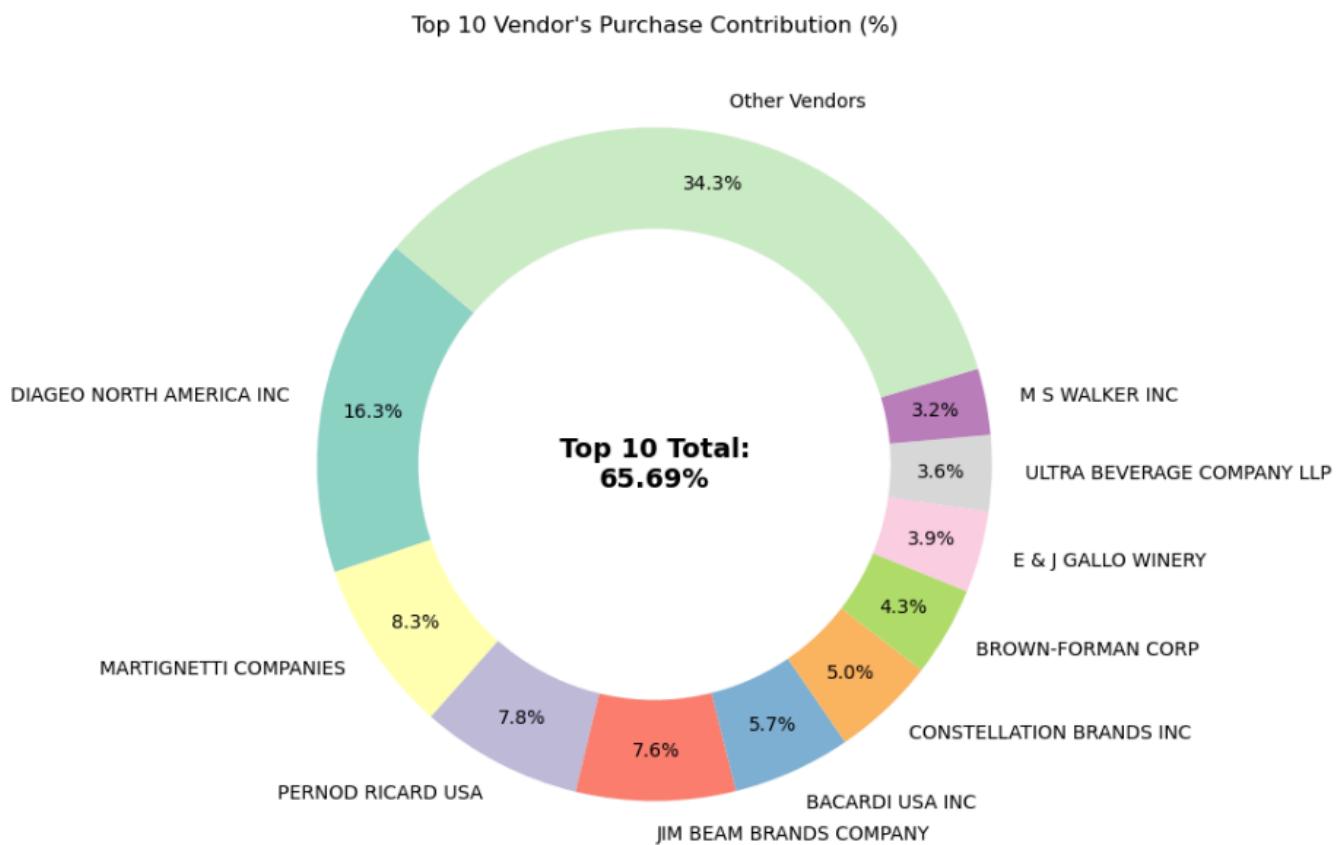
The analysis shows that a small group of vendors accounts for a major share of total purchase spending.

For example, Diageo North America alone contributes over 16 percent, followed by a few others with notable shares.

This concentrated **spending pattern highlights** the vendors that drive most of the procurement budget and signals **where negotiation, partnership management, or risk-mitigation efforts should be prioritized**.

04_Procurement dependence on top vendors

Total Purchase Contribution of top 10 vendors is 65.69 %

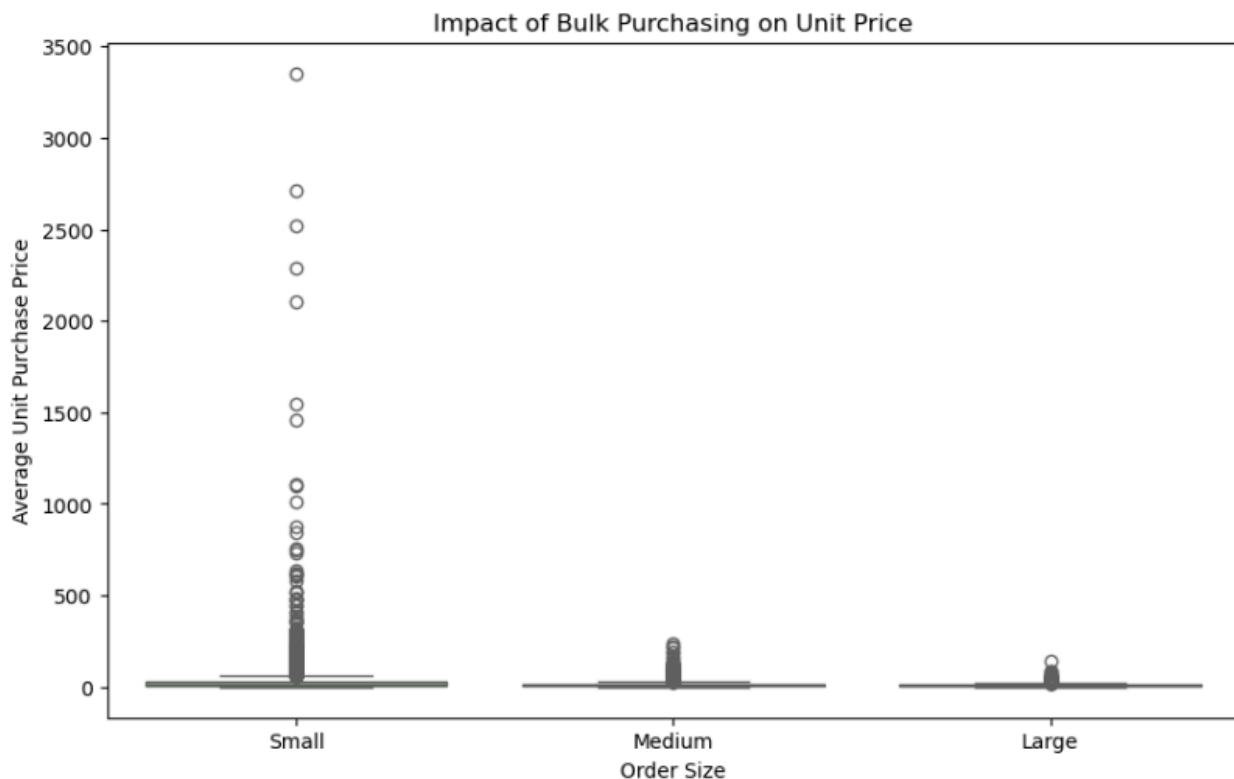


This analysis shows that 65.69 percent of total purchases come from a small group of vendors, while the rest account for only 34.31 percent. Relying so heavily on a few suppliers can create problems if any of them face delays or issues.

It's a signal that the **vendor base should be broadened** for the business **to reduce operational risk and stabilize supply**.

05_Bulk purchasing impact on unit cost

OrderSize	UnitPurchasePrice	
0	Small	39.057543
1	Medium	15.486414
2	Large	10.777625



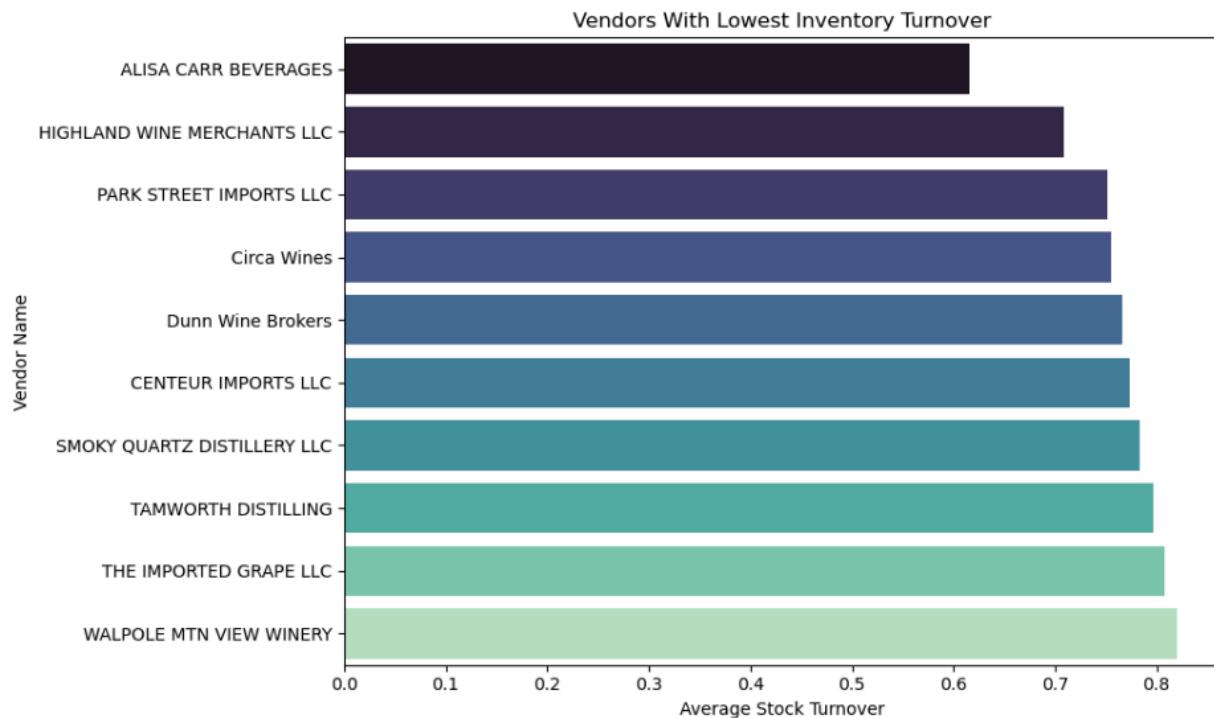
Based on the analysis, bulk buying clearly drives down costs.

Large orders secure the lowest unit price at around 10.78, which is roughly a 72 percent drop compared to small-volume purchases.

This pattern shows that **vendors benefit the most when they move into the large-order bracket**, making it the most cost-efficient purchase size for the business.

06 Vendors with Low Inventory Turnover

	VendorName	StockTurnover
0	ALISA CARR BEVERAGES	0.615385
36	HIGHLAND WINE MERCHANTS LLC	0.708333
60	PARK STREET IMPORTS LLC	0.751306
19	Circa Wines	0.755676
26	Dunn Wine Brokers	0.766022
15	CENTEUR IMPORTS LLC	0.773953
78	SMOKY QUARTZ DISTILLERY LLC	0.783835
90	TAMWORTH DISTILLING	0.797078
91	THE IMPORTED GRAPE LLC	0.807569
101	WALPOLE MTN VIEW WINERY	0.820548

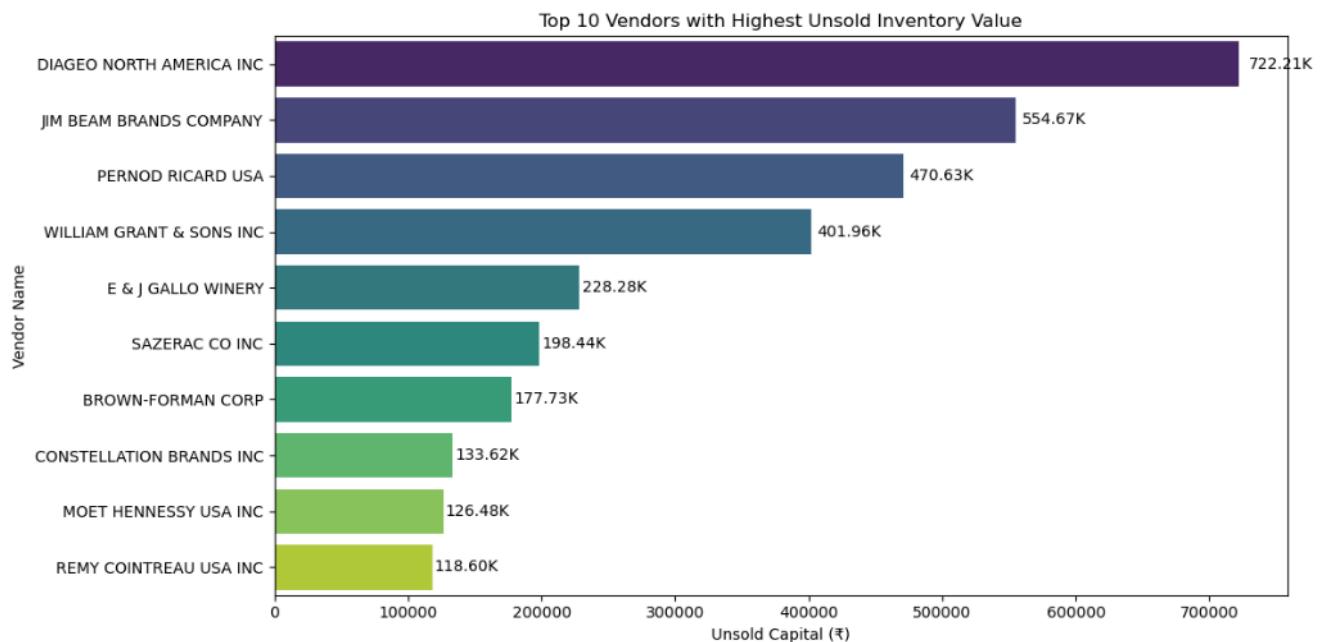


This analysis shows that nearly **\$2.71M is tied up in unsold inventory**, which puts pressure on storage capacity and limits available working capital.

Highlighting the vendors with low turnover helps to pinpoint where stock isn't moving as expected, allowing to **improve inventory planning and reduce unnecessary financial load**.

07_Uncollected Inventory Capital by Vendor

	VendorName	UnsoldInventoryValue
25	DIAGEO NORTH AMERICA INC	722.21K
46	JIM BEAM BRANDS COMPANY	554.67K
68	PERNOD RICARD USA	470.63K
116	WILLIAM GRANT & SONS INC	401.96K
30	E & J GALLO WINERY	228.28K
79	SAZERAC CO INC	198.44K
11	BROWN-FORMAN CORP	177.73K
20	CONSTELLATION BRANDS INC	133.62K
61	MOET HENNESSY USA INC	126.48K
77	REMY COINTREAU USA INC	118.60K

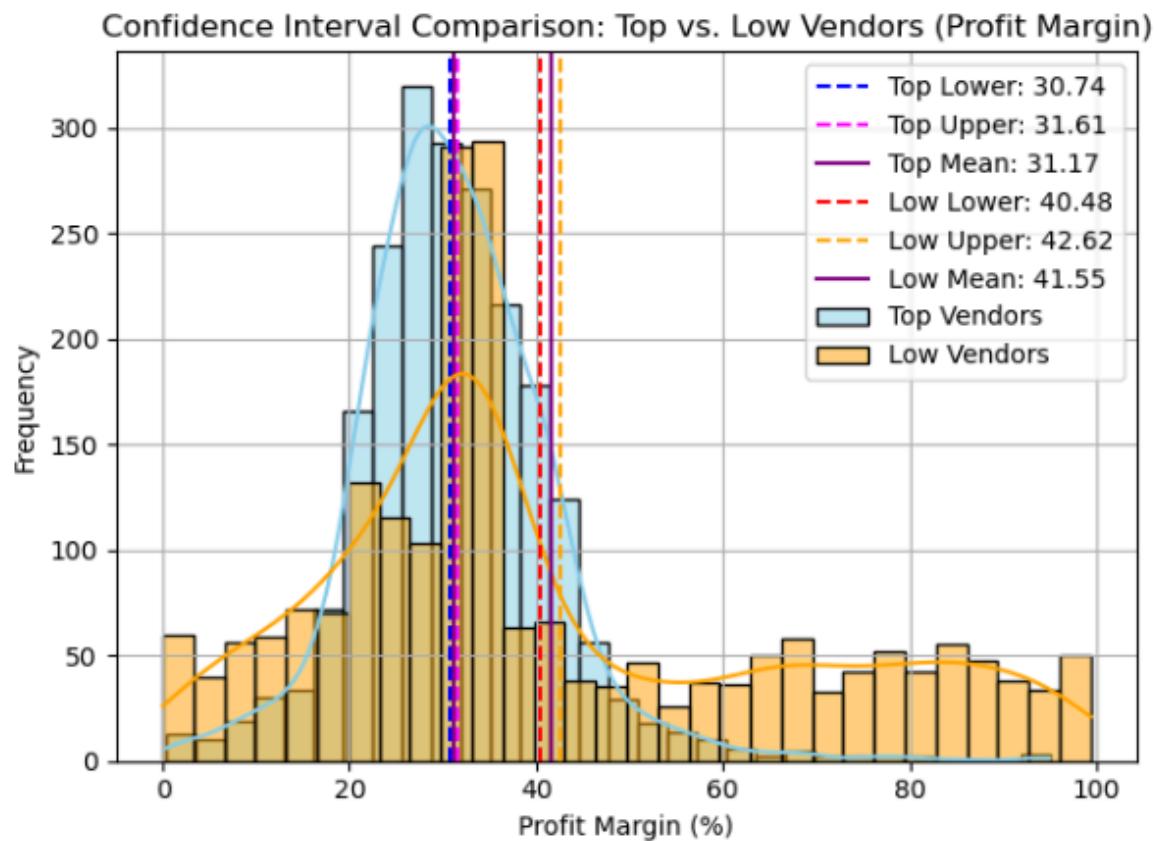


A large share of unsold inventory value sits with a few vendors, led by Diageo, Jim Beam, and Pernod Ricard. This concentration signals **a need to review demand planning and ordering decisions** for these suppliers to avoid tying up more capital than necessary.

08_Vendor Profit Margin Comparison for High Vs Low performing Vendors

Top Vendors 95% CI: (30.74, 31.61), Mean: 31.17

Low Vendors 95% CI: (40.48, 42.62), Mean: 41.55



The analysis shows a clear split in profitability.

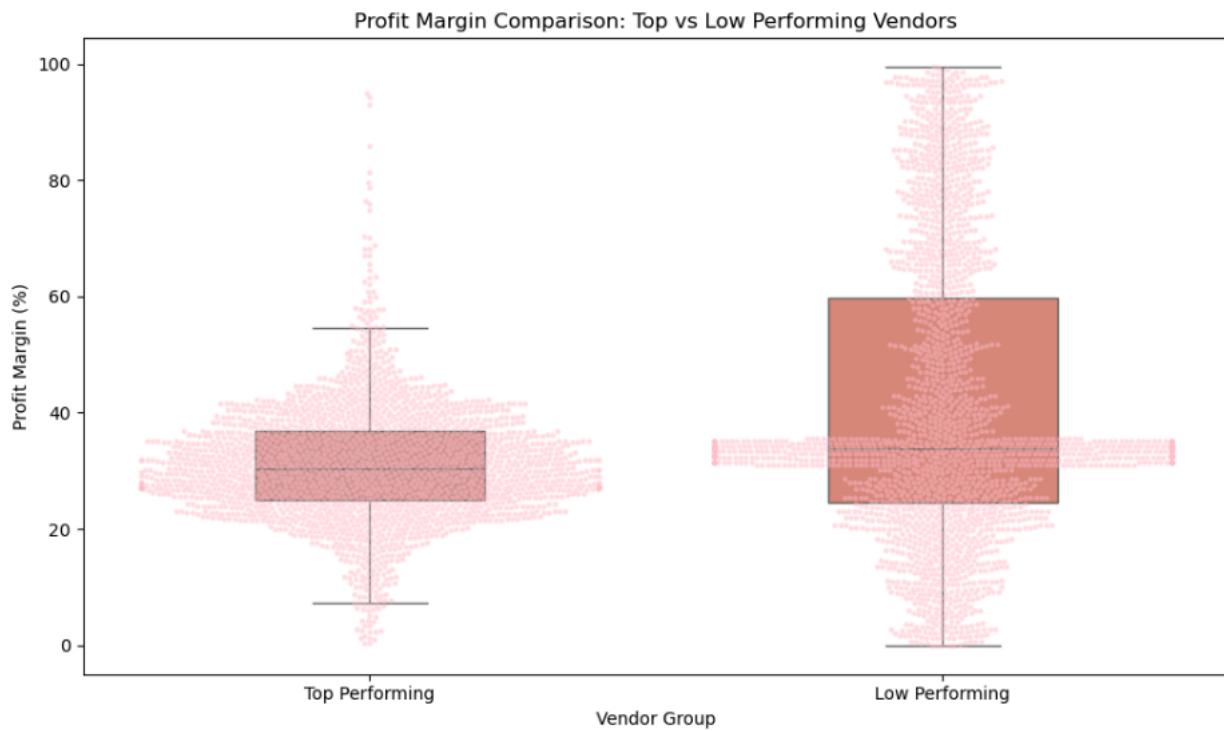
Top-performing vendors operate within a 95% confidence range of roughly 30.7% to 31.6%, while low-performing vendors show higher margins at about 40.5% to 42.6%. This suggests that **vendors with stronger sales may be pricing more aggressively**, while **those with lower sales are relying on higher margins but not converting volume**.

From a business standpoint, high-sales vendors could explore pricing or cost adjustments to lift margins further. **Vendors with lower sales but high margins may benefit from stronger marketing and distribution efforts** to improve reach and increase volume.

09_Profit Margin Gap Between Vendor Groups

T-Statistic: -17.6440, P-Value: 0.0000

Reject H_0 : There is a significant difference in profit margins between top and low-performing vendors.



The statistical test shows a clear difference in profit margins between top-selling and low-selling vendors.

This means both groups follow different profitability patterns.

High-margin vendors appear to **rely more on pricing strength**, while **high-volume vendors may gain more by improving cost efficiency**.

Business Recommendations and key insights

The analysis highlights several areas where the business can strengthen profitability, improve operational control, and reduce risk.

1. Unlock growth from high-margin, low-volume brands

A group of 198 brands is generating healthy margins but not reaching meaningful sales volumes. These brands present a strong growth opportunity. With focused marketing or more competitive pricing, they can contribute significantly more revenue without compromising profitability.

2. Revenue is concentrated among a few high-performing vendors and brands

Diageo North America leads vendor-level sales by a large margin, while Jack Daniel's No. 7 stands out at the brand level. This concentration shows where customer demand is strongest and where strategic relationships are most valuable. These vendors and brands should remain core partners as they consistently drive performance.

3. Procurement spending is heavily concentrated, raising risk exposure

A small vendor group accounts for most of the procurement budget. For example, Diageo North America alone contributes more than 16 percent. In total, 65.69 percent of purchases come from a limited set of suppliers. Such dependency increases the risk of supply disruptions. Broadening the vendor base would help stabilize supply and reduce operational vulnerabilities.

4. Bulk purchases provide the strongest cost advantage

The cost analysis shows a clear drop in unit price for larger order sizes. Large-volume buying reduces the unit cost to roughly 10.78, nearly 72 percent lower than small orders. This confirms that bulk purchasing is the most cost-efficient option and should be used strategically to strengthen margins.

5. High inventory holding needs corrective action

About 2.71 million dollars in capital is locked in unsold inventory. This ties up resources and adds storage pressure. Much of this stock sits with a few vendors, particularly Diageo, Jim Beam, and Pernod Ricard. Reviewing order quantities and demand forecasts for these suppliers will help prevent unnecessary build-ups and free up working capital.

6. Different profitability patterns between vendor groups

Top-selling vendors operate with profit margins in the 30.7 to 31.6 percent range, while low-selling vendors show higher margins between 40.5 and 42.6 percent. This gap signals two distinct operating models:

- High-volume vendors depend on lower margins and faster turnover.
- Low-volume vendors rely on higher margins but lack reach.

Improving pricing for high-volume vendors and expanding marketing support for low-volume vendors can balance performance on both sides.

Recommended Actions

- Revisit pricing for low-volume, high-margin brands to improve sales momentum.
- Expand vendor partnerships to reduce reliance on a small group of suppliers.
- Use bulk purchasing strategically to control costs without overstocking.
- Address slow-moving inventory through adjusted purchase quantities, clearance strategies, or improved storage planning.
- Strengthen marketing and distribution efforts for vendors struggling with volume.

By acting on these areas, the business can boost profitability, reduce risk, and improve overall operational efficiency.