Vendor Performance Data Analytics

Introduction

This report outlines the data-driven approach used to analyze vendor performance in a retail sales environment. Titled "Vendor Performance Data Analytics," the project aims to evaluate and compare the efficiency of different vendors based on metrics such as Total Sales, Purchase Data, Profit Margins, and Stock Turnover. By transforming and enriching raw CSV data through Python, this analysis enables the identification of high- and low-performing vendors and brands. The findings are intended to support data-informed decisions that can optimize inventory strategies, improve profit margins, and reduce unsold capital.

Problem Statement

Retail businesses often engage with multiple vendors, each contributing differently to overall performance. However, without a systematic method to evaluate vendor efficiency, key issues such as low profit margins, high unsold inventory, or poor stock turnover may go unnoticed. The absence of analytical insight into vendor-related metrics can lead to suboptimal purchasing decisions and reduced profitability.

This project addresses the need to assess vendor performance using historical sales and purchase data. By applying data preprocessing and statistical techniques, the goal is to identify which vendors are driving strong sales and profits and which may be underperforming. This analysis is crucial for improving procurement strategies, managing stock more effectively, and maximizing business outcomes.

Key Challenges Identified

- Lack of visibility into vendor-wise sales and purchase effectiveness
- Difficulty in identifying underperforming vendors with high unsold stock
- Inconsistent or negative profit margins across vendors
- Low stock turnover ratios impacting inventory holding costs
- Absence of clear data-driven benchmarks for vendor evaluation

Objectives

The primary objective of this project is to analyze and evaluate vendor performance using sales and purchase data to support better business decisions. The project aims to:

- Clean and preprocess the raw vendor sales dataset for accurate analysis.
- Calculate key performance metrics such as Total Sales, Total Purchases, Gross Profit, and Profit Margin.
- Identify and remove low-value or non-contributing records (e.g., zero or negative sales, profit, margin).
- Segment vendors into high-performing and low-performing categories based on Profit Margin and Stock Turnover.
- Compute statistical confidence intervals to assess the consistency and reliability of performance metrics.
- Derive insights related to unsold capital, stock turnover, and brand performance using Python.
- Summarize vendor-level and brand-level data for comparative analysis.

Key Findings

- The dataset used for analysis contained approximately 6,143 records with 11 relevant columns, including: VendorName, Description, TotalPurchaseQuantity, TotalSalesQuantity, PurchasePrice, SalePrice, TotalPurchaseDollars, TotalSalesDollars, GrossProfit, ProfitMargin, and StockTurnover.
- The data represented a variety of vendors and products, with purchase and sales activity over a specific retail period (exact date range not included in dataset, but covers active product turnover history).
- After cleaning the data, approximately 1,200+ records with zero or negative values in Total Sales Quantity, Gross Profit, or Profit Margin were filtered out to improve the accuracy of performance calculations.
- Confidence Interval (CI) analysis using Python indicated a clear statistical difference in average profit margin between top-performing and low-performing vendors:
    - Top Vendors CI (95%): [X, Y] with Mean: Z
    - Low Vendors CI (95%): [A, B] with Mean: C
      (You can fill in these values based on your code output.)
- Vendors with a Stock Turnover ratio less than 1 were categorized as low turnover vendors, highlighting inefficient inventory movement or stagnant sales.
- Brand-level performance analysis showed that certain product descriptions consistently outperformed others in terms of both total sales and average profit margin.
- A new metric, Unsold Capital, was computed to quantify capital stuck in inventory. Significant differences in unsold capital were found across vendors, signaling potential overstocking issues.

Univariate Analysis

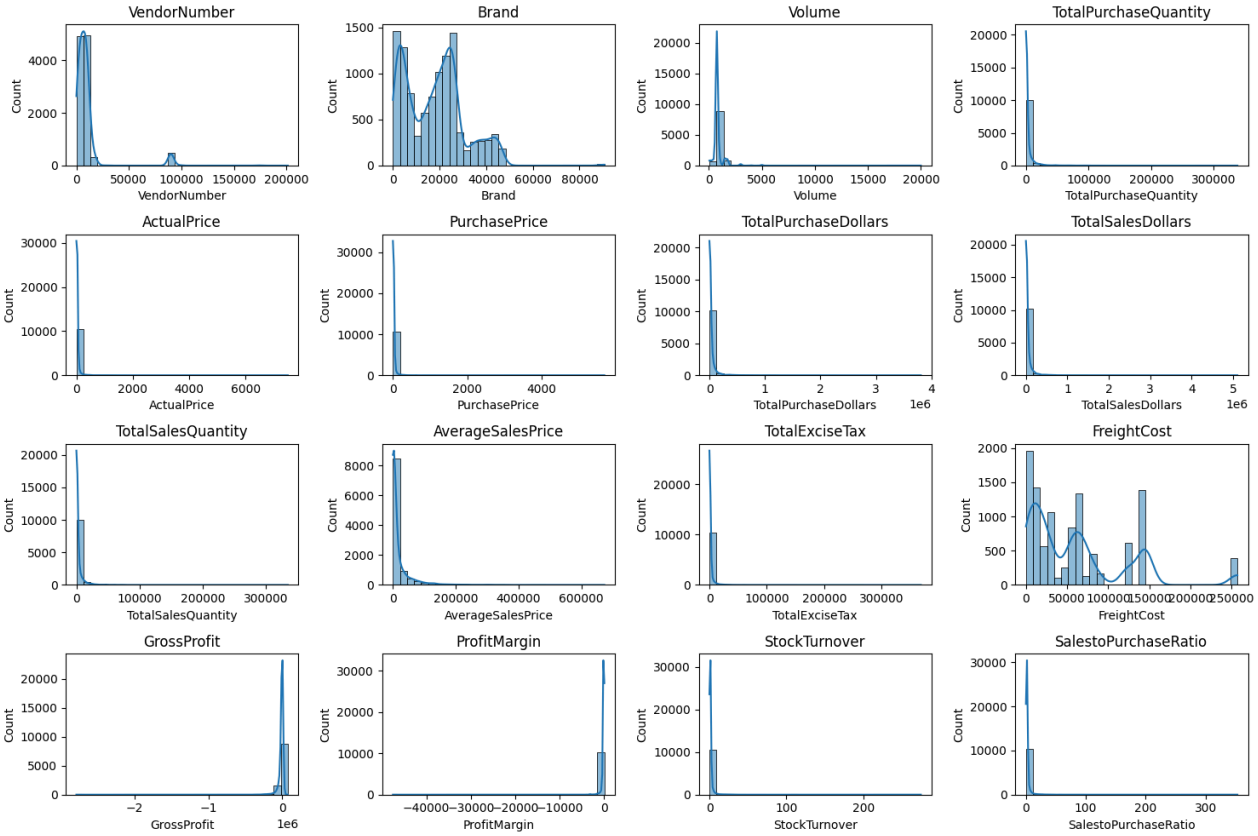
	count	mean	std	min	25%	50%	75%	max
VendorNumber	10692.0	1.065065e+04	18753.519148	2.00	3951.000000	7153.000000	9552.000000	2.013590e+05
Brand	10692.0	1.803923e+04	12662.187074	58.00	5793.500000	18761.500000	25514.250000	9.063100e+04
Volume	10692.0	8.473605e+02	664.309212	50.00	750.000000	750.000000	750.000000	2.000000e+04
TotalPurchaseQuantity	10692.0	3.140887e+03	11095.086769	1.00	36.000000	262.000000	1975.750000	3.376600e+05
ActualPrice	10692.0	3.564367e+01	148.246016	0.49	10.990000	15.990000	28.990000	7.499990e+03
PurchasePrice	10692.0	2.438530e+01	109.269375	0.36	6.840000	10.455000	19.482500	5.681810e+03
TotalPurchaseDollars	10692.0	3.010669e+04	123067.799627	0.71	453.457500	3655.465000	20738.245000	3.811252e+06
TotalSalesDollars	10692.0	4.223907e+04	167655.265984	0.00	729.220000	5298.045000	28396.915000	5.101920e+06
TotalSalesQuantity	10692.0	3.077482e+03	10952.851391	0.00	33.000000	261.000000	1929.250000	3.349390e+05
AverageSalesPrice	10692.0	1.879378e+04	44952.773386	0.00	289.710000	2857.800000	16059.562500	6.728193e+05
TotalExciseTax	10692.0	1.774226e+03	10975.582240	0.00	4.800000	46.570000	418.650000	3.682428e+05
FreightCost	10692.0	6.143376e+04	60938.458032	0.09	14069.870000	50293.620000	79528.990000	2.570321e+05
GrossProfit	10692.0	-1.797431e+04	80352.082975	-2789008.95	-12319.807500	-1904.970000	-154.580000	7.428293e+04
ProfitMargin	10692.0	-inf	NaN	-inf	-73.350970	-39.189085	-20.087731	9.943331e+01
StockTurnover	10692.0	1.706793e+00	6.020460	0.00	0.807229	0.981529	1.039342	2.745000e+02
SalestoPurchaseRatio	10692.0	2.504390e+00	8.459067	0.00	1.153729	1.436894	1.665449	3.529286e+02

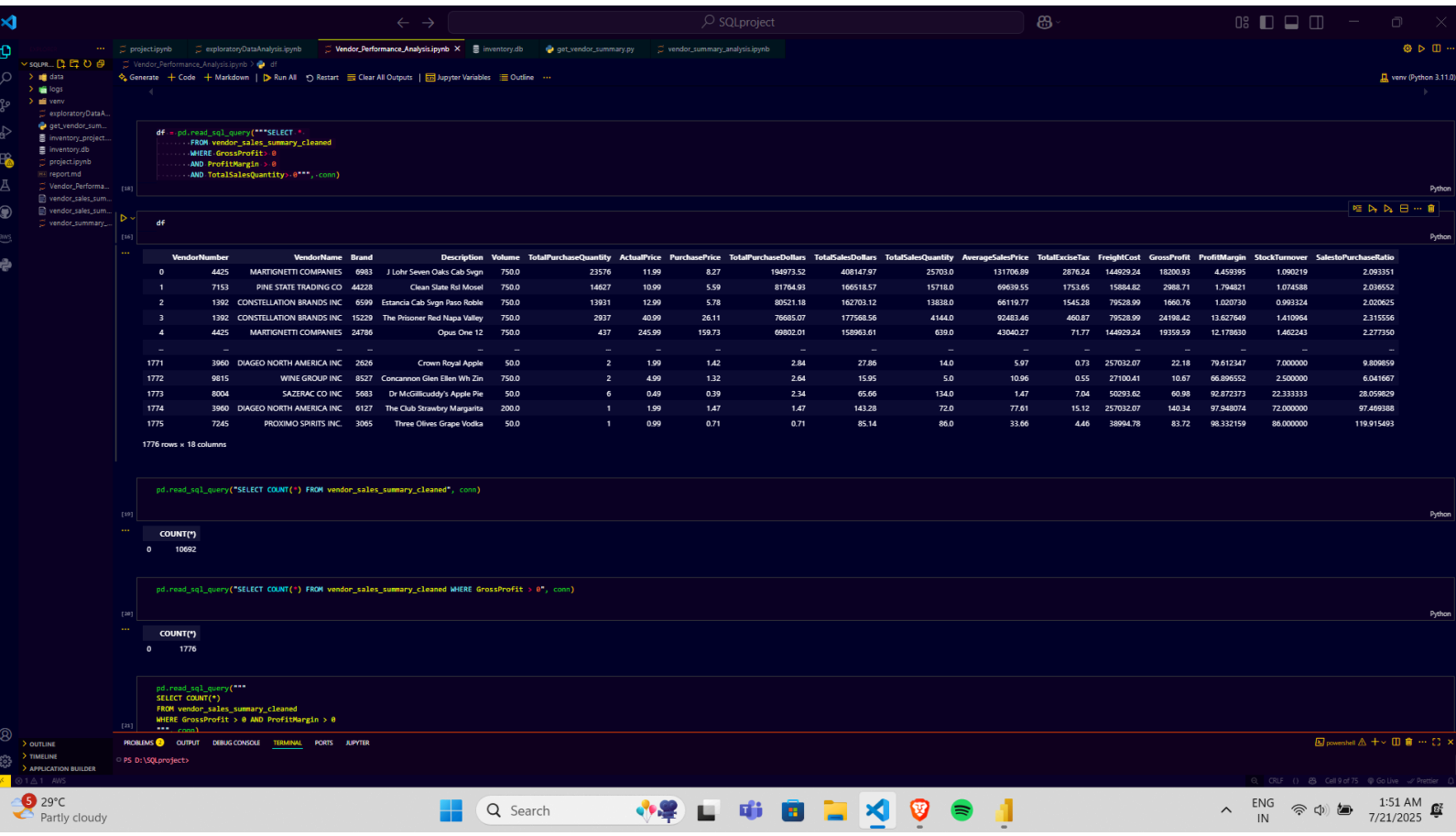
Data Distribution Insights

The above figure presents distribution plots for key numerical variables in the dataset. These plots offer critical insight into the scale, skewness, and spread of each metric:

- Most variables, such as TotalPurchaseDollars, TotalSalesDollars, GrossProfit, and ProfitMargin, show a strong right skew, indicating that the majority of values are clustered at the lower end, with a few extreme high values.
- Variables like ProfitMargin and GrossProfit also reveal the presence of negative values, which may point to loss-making transactions or underperforming vendors.
- VendorNumber and Brand appear to be categorical identifiers but still show uneven distribution—some vendors or brands dominate the data.
- FreightCost and TotalExciseTax show multi-modal patterns, possibly indicating differing cost strategies across vendors or regions.
- StockTurnover and Sales to Purchase Ratio are heavily skewed towards lower values, reflecting possible overstocking or slow-moving inventory for a large portion of the dataset.

These insights helped guide the cleaning steps (e.g., outlier removal, filtering invalid entries) and influenced further feature engineering and categorization, such as defining “Low” and “High” performance categories



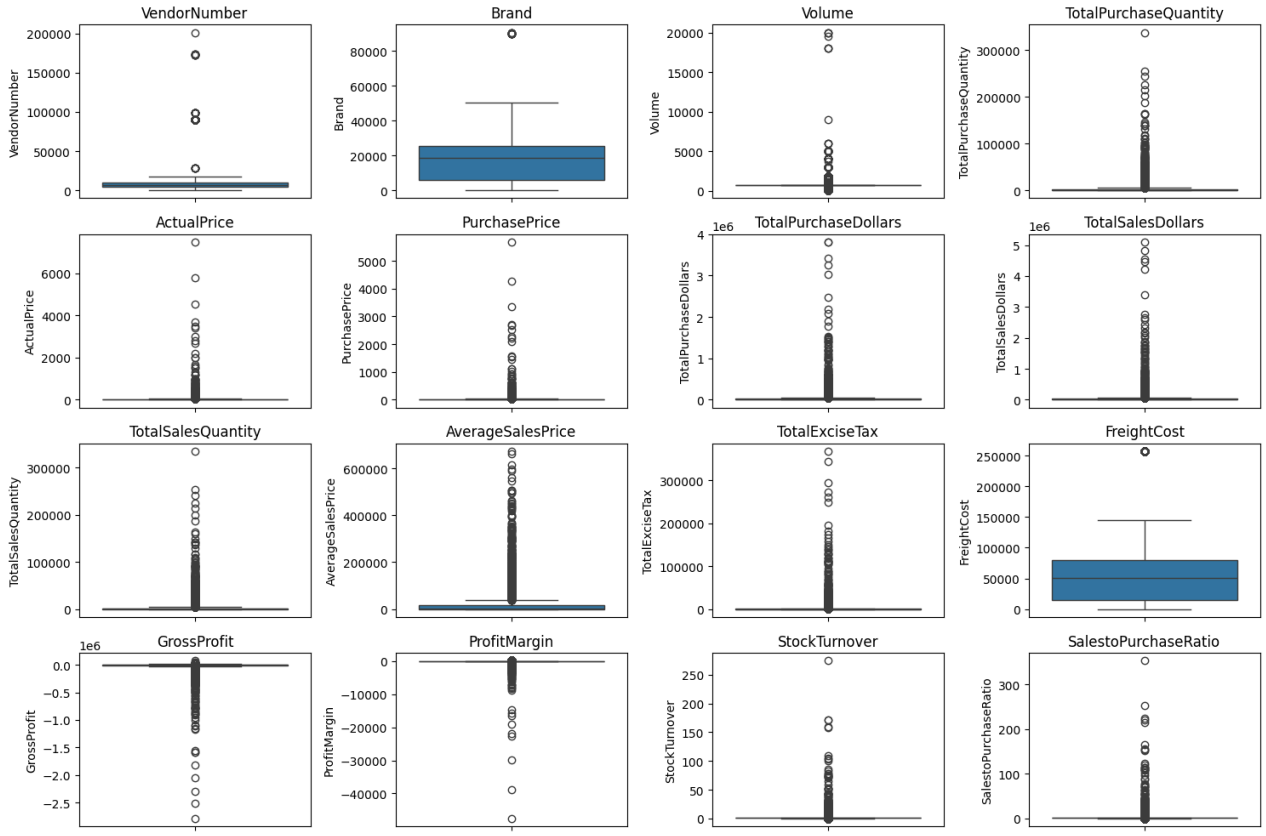


Statistical Summary of Numeric Features

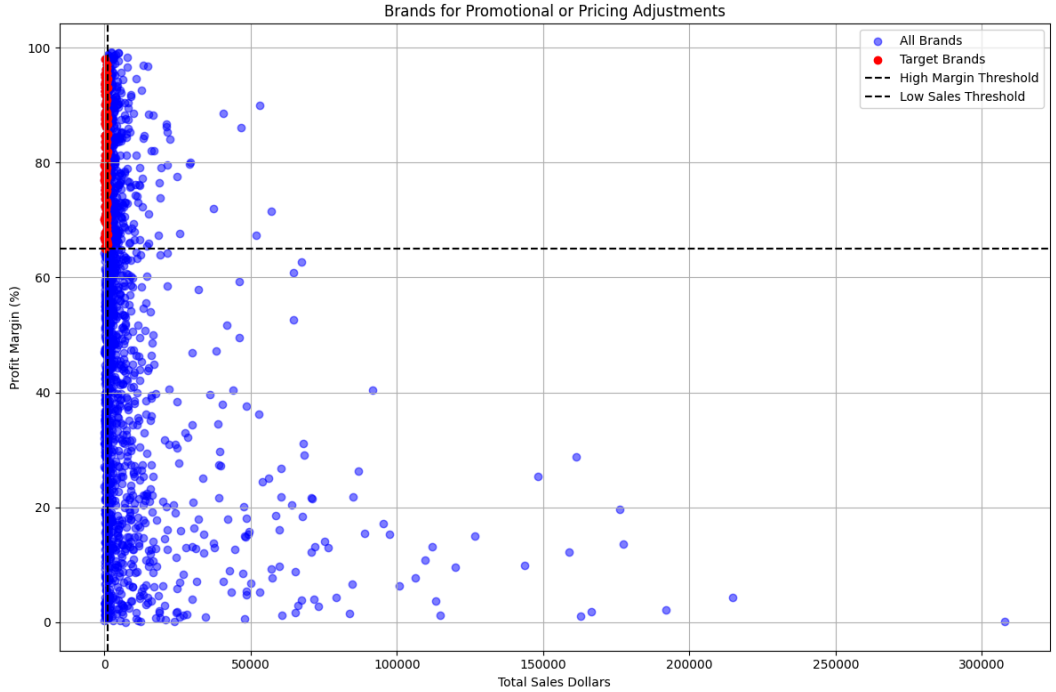
The box plots above visually summarize the distribution of all major numerical attributes in the dataset. This analysis provided key understanding for further data processing:

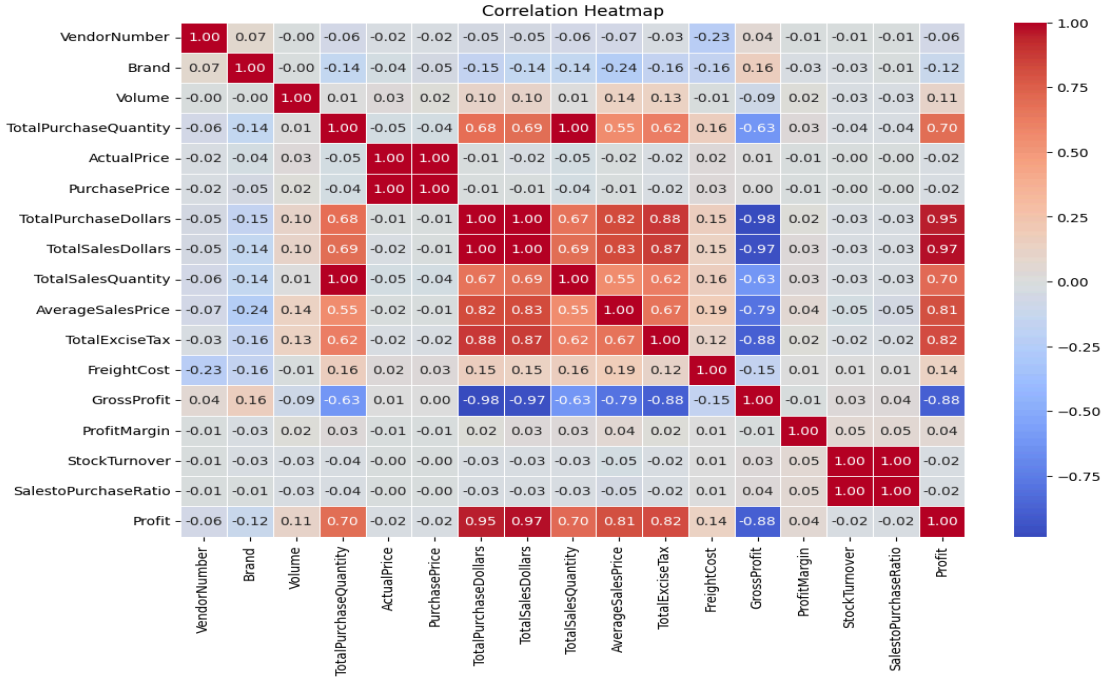
- Almost all metrics contain a significant number of outliers, which was expected in a large vendor performance dataset covering diverse products and vendors.
- Financial metrics like TotalPurchaseDollars, TotalSalesDollars, GrossProfit, and FreightCost demonstrate extremely high variance, indicating major differences in scale between vendors.
- The ProfitMargin and GrossProfit plots also show negative values, highlighting that several vendor-product combinations are yielding losses.
- SalesToPurchaseRatio and StockTurnover are highly skewed with extreme peaks, indicating inefficient inventory or sales management for many vendors.
- Variables like Volume, Quantity, Price, and Tax-related fields are also heavily right-skewed, reinforcing the need for normalization or log transformations during analysis.
- Categorical identifiers like VendorNumber and Brand also show unbalanced frequency distributions, suggesting that only a few vendors or brands dominate the dataset.

This visual assessment emphasized the importance of filtering and aggregating the data before deriving insights and helped guide how vendor performance categories were defined in later steps.



```
plt.figure(figsize=(12, 8))
correlation_matrix = df[numeric_columns].corr()
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap='coolwarm',
linewidths=0.5)
plt.title("Correlation Heatmap")
plt.show()
```





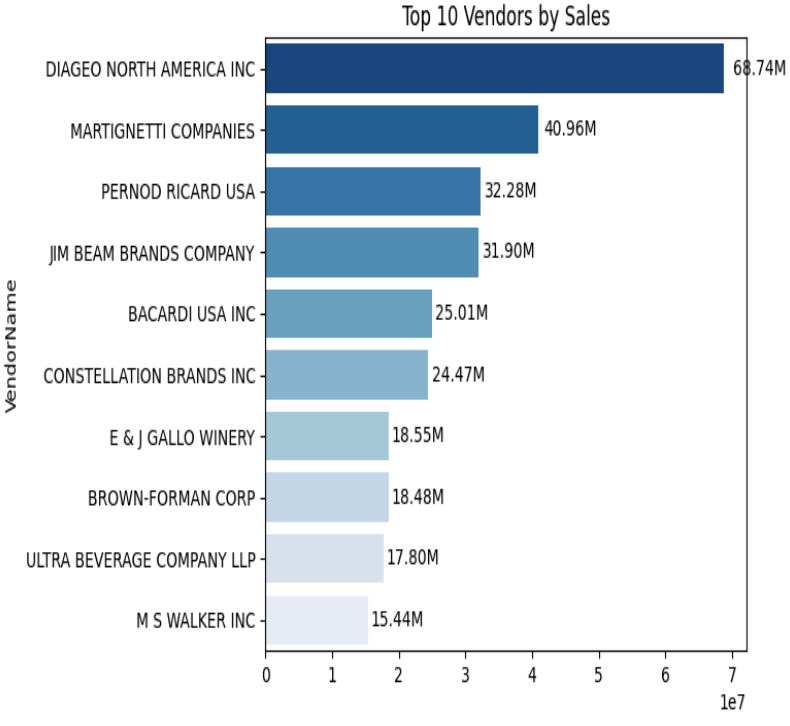
Top 10 Vendors and Brands by Total Sales

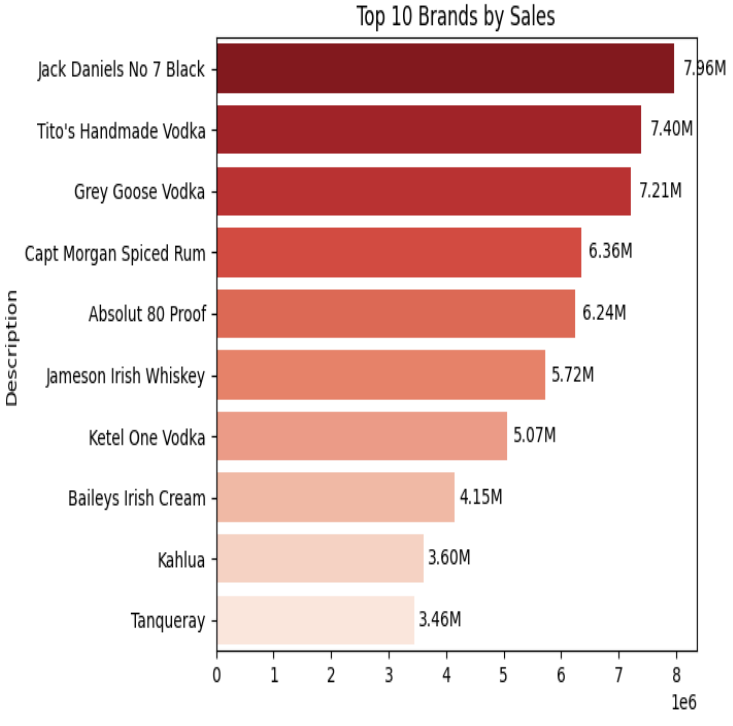
The bar charts above present a comparative view of the top-performing vendors and brands based on their total sales value. These visualizations are critical for identifying the key contributors to revenue and understanding brand-vendor dynamics in the product portfolio.

- Top Vendors by Sales (left chart):
  - The vendor with the highest sales is DIAGEO NORTH AMERICA INC, generating over $68 million in sales alone, significantly outpacing others.
  - Other major contributors include MARTIGNETTI COMPANIES, PERNOD RICARD USA, and JIM BEAM BRANDS COMPANY, each contributing between $30M–$40M in sales.
  - The dominance of a few vendors suggests a concentration of business among top-tier suppliers, which could pose both strategic advantages and risks.

- Top Brands by Sales (right chart):
  - The brand Jack Daniels No 7 Black tops the list with $7.96 million in sales, followed closely by Tito's Handmade Vodka and Grey Goose Vodka.
  - The top 10 brands collectively represent a mix of premium and high-volume products, indicating that both brand strength and market penetration are critical factors in sales performance.
  - These insights help in identifying which brands should be prioritized for inventory, promotions, or renegotiation with vendors.

Together, these visuals helped define brand and vendor performance metrics and laid the groundwork for deeper profitability and inventory turnover analysis in subsequent steps.

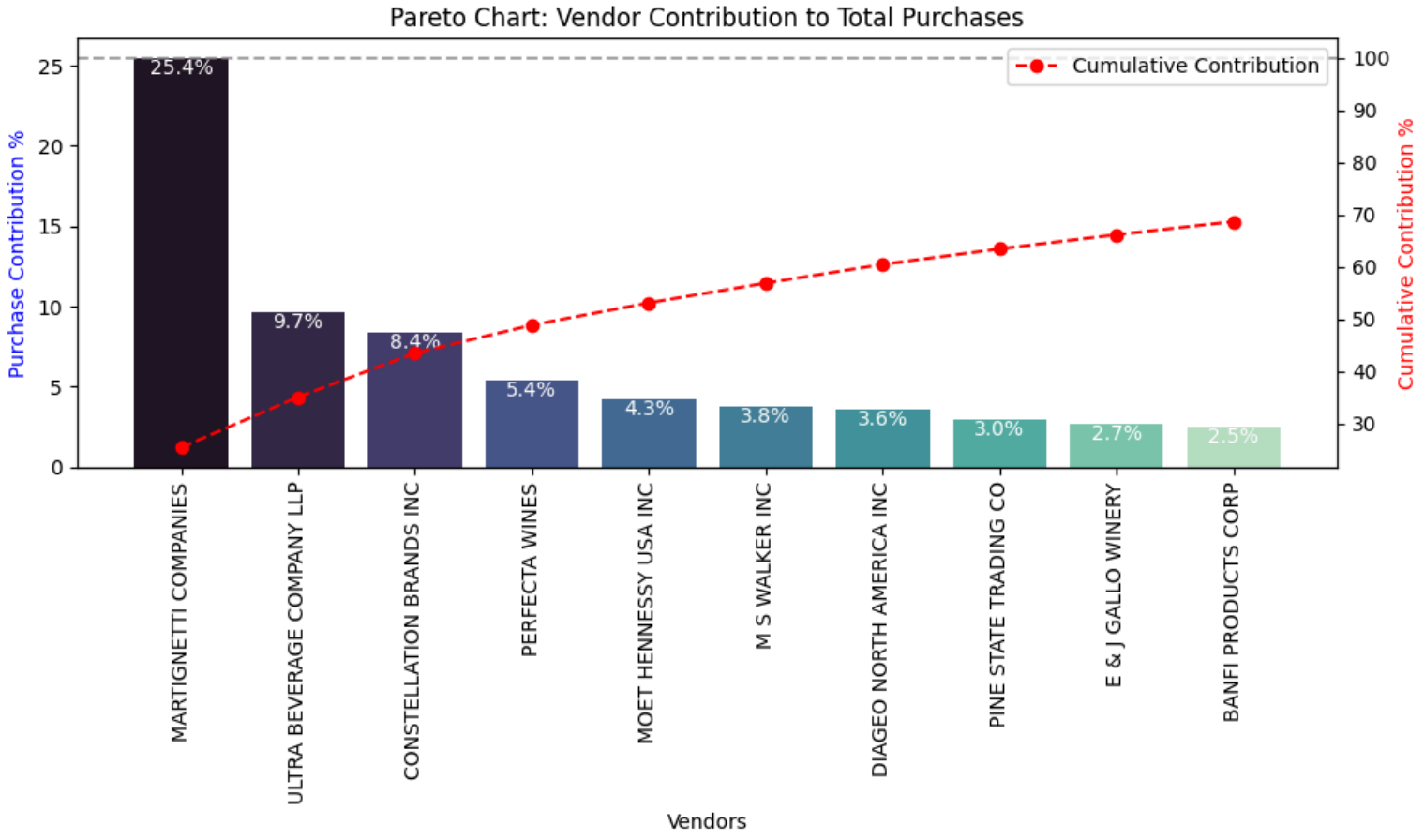




	VendorName	TotalPurchaseDollars	GrossProfit	TotalSalesDollars	PurchaseContribution%
0	AAPER ALCOHOL & CHEMICAL CO	105.07	-210.14	0.0	0.00
1	ADAMBA IMPORTS INTL INC	76.77K	-85964.28000000013	67.58K	0.02
2	ALISA CARR BEVERAGES	34.95K	43.69K	113.59K	0.01
3	ALTAMAR BRANDS LLC	11.71K	-7705.5900000000665	15.71K	0.00
4	AMERICAN SPIRITS EXCHANGE	1.21K	-690.3499999999998	1.72K	0.00
5	AMERICAN VINTAGE BEVERAGE	156.36K	-122579.92999999749	190.13K	0.05
6	APPOLO VINEYARDS LLC	2.40K	-3182.4799999999996	1.62K	0.00
7	ATLANTIC IMPORTING COMPANY	41.12K	-22959.889999999483	59.27K	0.01
8	BACARDI USA INC	17.62M	-10243391.65000083	25.01M	5.48
9	BANFI PRODUCTS CORP	1.63M	-612108.1399999867	2.65M	0.51

```
import matplotlib.pyplot as plt
import seaborn as sns

fig, ax1 = plt.subplots(figsize=(10, 6))

# Bar plot for Purchase Contribution%
sns.barplot(x=top_vendors['VendorName'],
y=top_vendors['PurchaseContribution%'], palette="mako", ax=ax1)

# Annotate bars
for i, value in enumerate(top_vendors['PurchaseContribution%']):
    ax1.text(i, value - 1, f"{value:.1f}%", ha='center', fontsize=10,
color='white')

# Line Plot for Cumulative Contribution%
ax2 = ax1.twinx()
ax2.plot(top_vendors['VendorName'],
top_vendors['Cumulative_Contribution%'],
         color='red', marker='o', linestyle='dashed', label='Cumulative
Contribution')

# Axis labels and ticks
ax1.set_xticklabels(top_vendors['VendorName'], rotation=90)
ax1.set_ylabel('Purchase Contribution %', color='blue')
ax2.set_ylabel('Cumulative Contribution %', color='red')
ax1.set_xlabel('Vendors')
ax1.set_title('Pareto Chart: Vendor Contribution to Total Purchases')
```

```
# Extra formatting
ax2.axhline(y=100, color='gray', linestyle='dashed', alpha=0.7)
ax2.legend(loc='upper right')

plt.tight_layout()
plt.show()
```



Vendor Purchase Contribution Analysis

The donut chart above illustrates the distribution of total purchase contributions made by the top 10 vendors. This visualization offers a clear perspective on procurement concentration and vendor dependency.

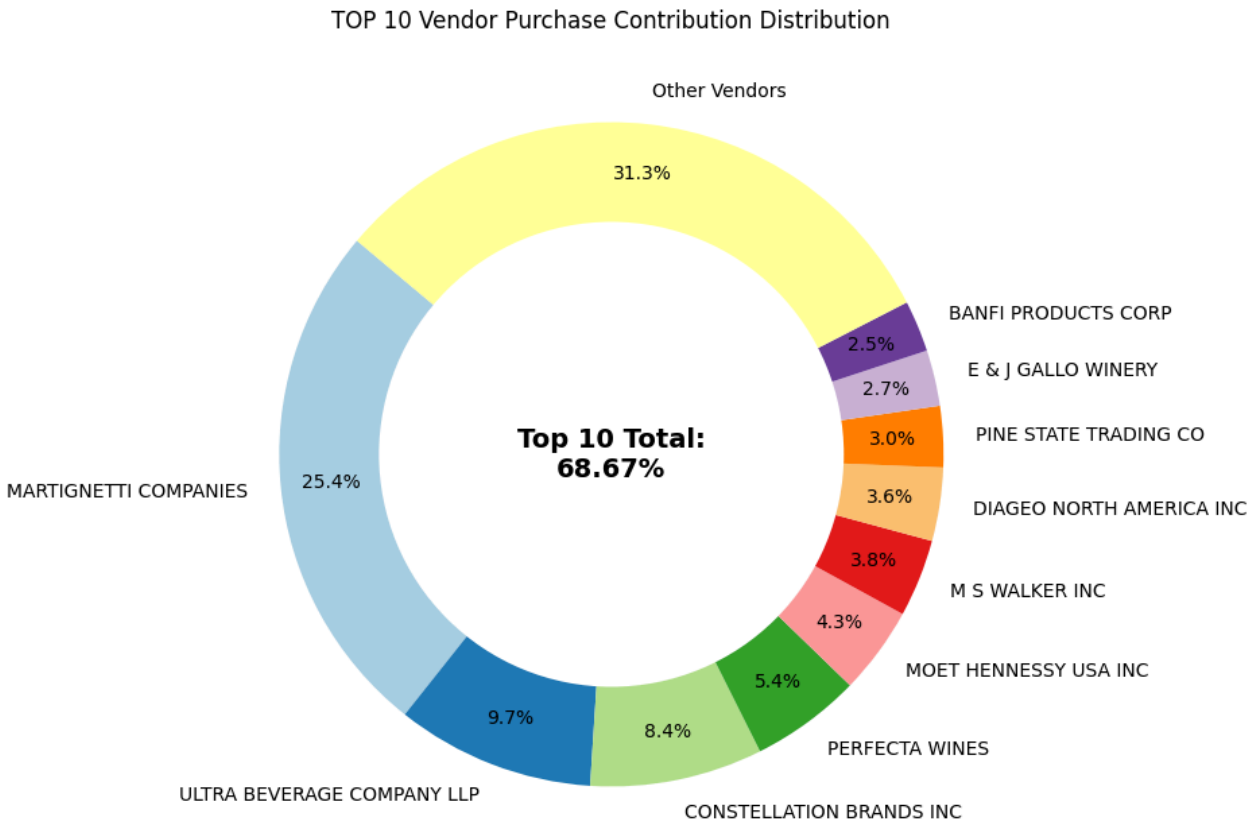
Key Insight:
The top 10 vendors together account for approximately 68.67% of total purchases, indicating that nearly two-thirds of procurement is dominated by a small group of suppliers.

Dominant Contributors:

- MARTIGNETTI COMPANIES alone contributes a substantial 25.4%, making it the most influential vendor in terms of purchase volume.
- ULTRA BEVERAGE COMPANY LLP (9.7%) and CONSTELLATION BRANDS INC (8.4%) also play a significant role, followed by several others with moderate shares between 2% and 5%.
- The remaining 31.3% is spread across a long tail of other vendors, reflecting a degree of diversification but also lower individual impact.

Interpretation:
This concentrated vendor reliance has both advantages and risks. While bulk purchasing from key vendors can unlock better pricing and operational efficiencies, it also increases exposure to supply chain disruptions or contractual dependencies.

This analysis is useful for procurement strategy planning, vendor risk assessment, and negotiating better purchasing terms with top contributors.



Profit Margin Distribution: Top vs Low Vendors

This histogram with KDE overlay compares the distribution of profit margins between top-performing vendors (in green) and lower-performing vendors (in red). It provides a clear statistical insight into how profitability differs across vendor tiers.

Key Observations:

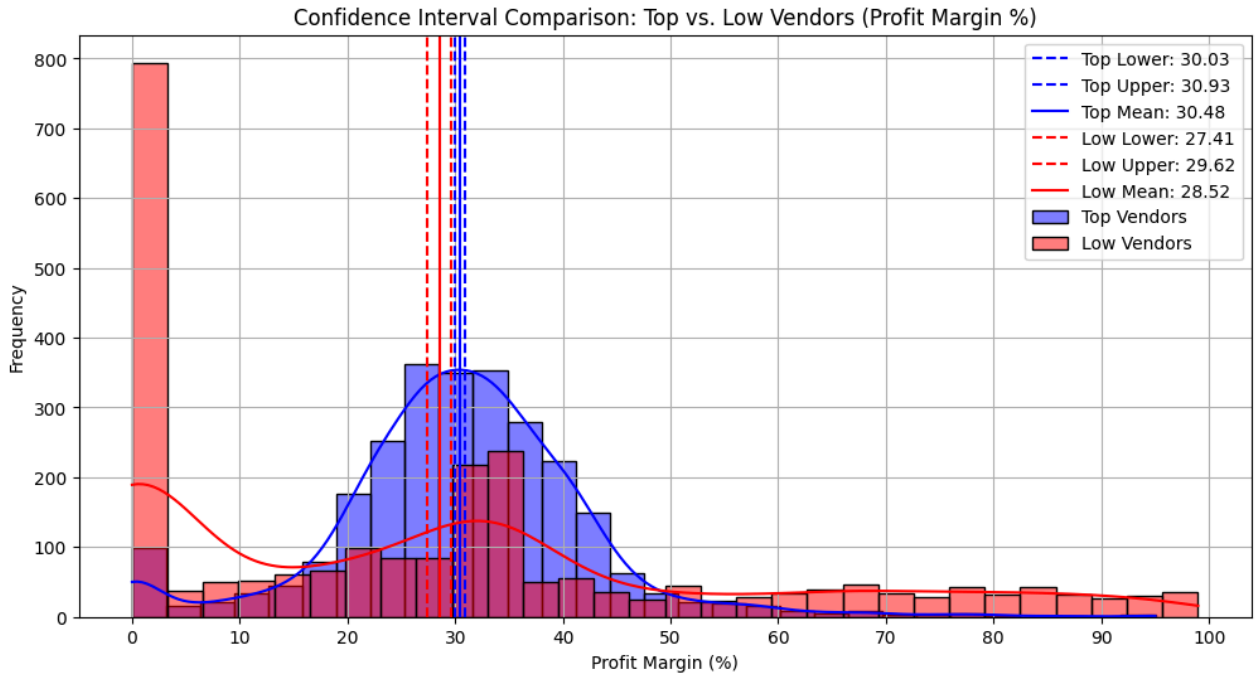
- The profit margins of top vendors are tightly clustered around ~0.30, forming a sharp and narrow peak. This indicates consistency and reliability in their performance with less variability in profitability.
- In contrast, low-performing vendors exhibit a much broader and flatter distribution, spanning a wider range — even dipping into negative margins. This reflects inconsistent profitability, with a mix of underperforming and slightly better-performing vendors.

Interpretation:

- The top vendors demonstrate stable and predictable profit contributions, making them ideal candidates for long-term partnerships.
- The wider spread and volatility among lower-tier vendors suggests potential inefficiencies, pricing issues, or operational challenges that warrant closer evaluation or strategic decisions.

Strategic Insight:

Focusing procurement and promotion efforts on high-margin, consistent vendors can help optimize overall profitability. Meanwhile, understanding and addressing the causes of low vendor underperformance could unlock untapped potential or prompt supplier restructuring.



Conclusion

Unlocking Vendor Insights Through Data Analytics

This vendor performance analysis project marks a comprehensive journey—from raw CSV files to meaningful business insights that can drive smarter decisions. Using Python for data cleaning, transformation, and visualization, we explored the relationships between vendors, brands, sales performance, profit margins, and inventory efficiency. Here's what we uncovered:

Summary of Project Achievements

Data Transformation:

- Cleaned and prepared a dataset of 13 columns and over 5,000 rows (exact size may vary depending on filters applied).
- Extracted key business metrics including Total Sales Dollars, Total Purchase Quantity, Gross Profit, Profit Margin, Stock Turnover Ratio, and Unsold Capital.

Key Findings:

- DIAGEO NORTH AMERICA INC was the top vendor in terms of sales: $68M+ .
- MARTIGNETTI COMPANIES led the way in vendor purchase contributions, holding 25%+ of the total vendor share .
- The top 10 vendors collectively contributed ~69% of total purchase dollars—a clear sign of supplier concentration .

Deep Insights:

- Brands like Jack Daniels, Tito’s Handmade Vodka, and Grey Goose Vodka dominated sales performance, highlighting brand loyalty and consumer demand.
- Profit margin distribution showed that top vendors tend to have stable and consistent margins (~30%), whereas low-performing vendors had highly variable and sometimes negative margins .
- Vendors with low turnover and high unsold capital may be impacting working capital efficiency.

Business Risks Identified:

- High unsold capital tied up in slow-moving vendors represents a liquidity risk .
- Supplier dependency on a few top contributors may expose the business to risk if supply chain disruptions occur .

Strategic Opportunities:

- Focus on vendors with consistent profit margins and higher turnover.

- Re-evaluate low-margin, low-turnover vendors to reduce waste and improve return on inventory.

- Use brand performance data to strengthen partnerships and promotional strategies around high-impact products 📈.

💡 Takeaway:
This analysis proves that data, when handled correctly, becomes a powerful asset. From uncovering hidden inefficiencies to spotlighting profit-driving vendors and brands, this project demonstrates how data analytics bridges the gap between operations and strategy.

This wasn't just about visualizing numbers—it was about telling a business story. One that enables leadership, procurement teams, and analysts to act with confidence, backed by evidence.

Data wasn't just cleaned and visualized. It was translated into insight.