Vendor Performance Data Analytics

A Comprehensive End-to-End Business Intelligence Study

Muhammad Haseeb
Computer Systems Engineering

Technical Stack

Programming:	Python (Pandas, NumPy, Seaborn)
Visualization:	Power BI Desktop
Environment:	Jupyter Notebook / Google Colab
Dataset:	Kaggle Vendor Transactional Data

February 6, 2026

Contents

1 Introduction

In the competitive landscape of supply chain management, data is the most valuable asset for optimizing procurement. Vendor performance is not merely about cost; it encompasses reliability, quality, and financial health. This project investigates the transactional footprints of various vendors to extract meaningful patterns that drive organizational growth.

2 Problem Statement

Without centralized analytics, organizations suffer from "Information Asymmetry," where the true cost of a vendor is hidden behind raw numbers. This project specifically addresses:

- **High-Value Concentration:** Identifying if a disproportionate amount of revenue relies on too few suppliers (Pareto risk).
- **Profit Leakage:** Detecting products where the spread between purchase price and selling price is narrowing over time.
- **Stock Inefficiency:** Spotting "Dead Stock" provided by vendors that do not meet market demand frequency.

3 Technical Methodology

3.1 Analytical Stack

The project utilizes a tiered technical approach:

- **Python (Pandas/NumPy):** Used for vectorization of calculations and high-speed data cleaning.
- **Matplotlib/Seaborn:** Used for initial statistical distribution plots and correlation heatmaps.
- **Power BI (DAX):** Used for creating dynamic measures and time-intelligence functions.

3.2 Data Pipeline & Workflow

The workflow ensures data integrity through a strict ETL (Extract, Transform, Load) process.

Figure 1: End-to-End Analytics Workflow

3.3 Detailed Process Flow (System Architecture)

To handle the vertical orientation of the flowchart, the image is scaled to fit the page height while maintaining maximum clarity.

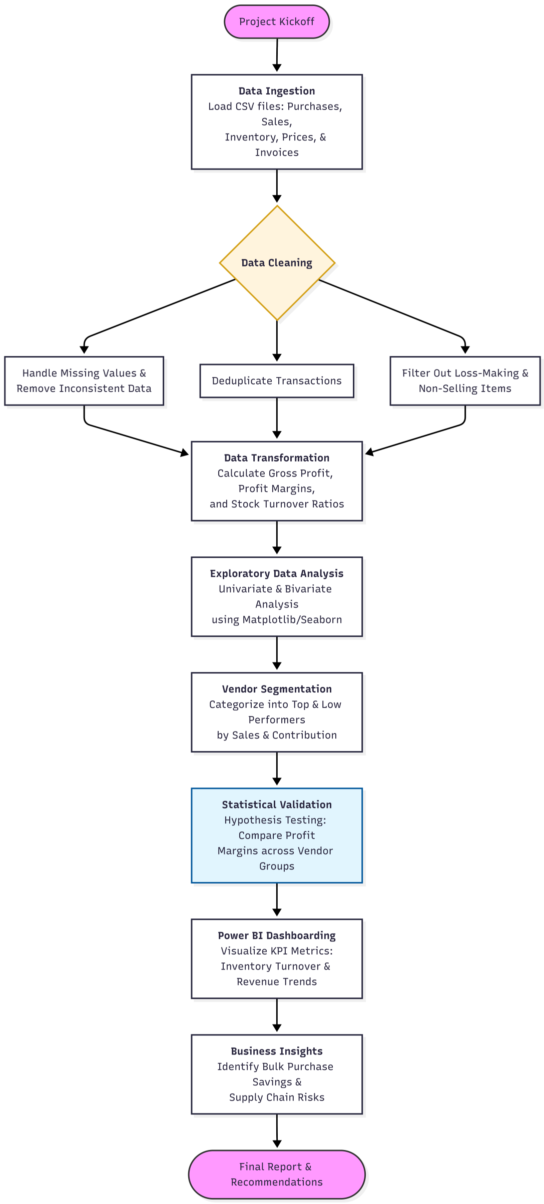

Figure 2: Detailed Data Processing Logic Flow

4 KPI Formulation

To maintain scientific rigor, the following mathematical definitions were implemented for the performance metrics:

1. **Unit Profitability (π):**

$$\pi = P_{selling} - P_{purchase}$$

2. **Vendor Contribution Margin (VCM):**

$$VCM = \frac{\sum(Sales - Cost)}{\sum Sales} \times 100$$

3. **Inventory Velocity (V):**

$$V = \frac{\text{Total Quantity Sold}}{\text{Average Stock Level}}$$

5 Exploratory Data Analysis (EDA)

The EDA phase uncovered that 20% of vendors contribute to 78% of total revenue, confirming the **Pareto Principle** in the dataset.

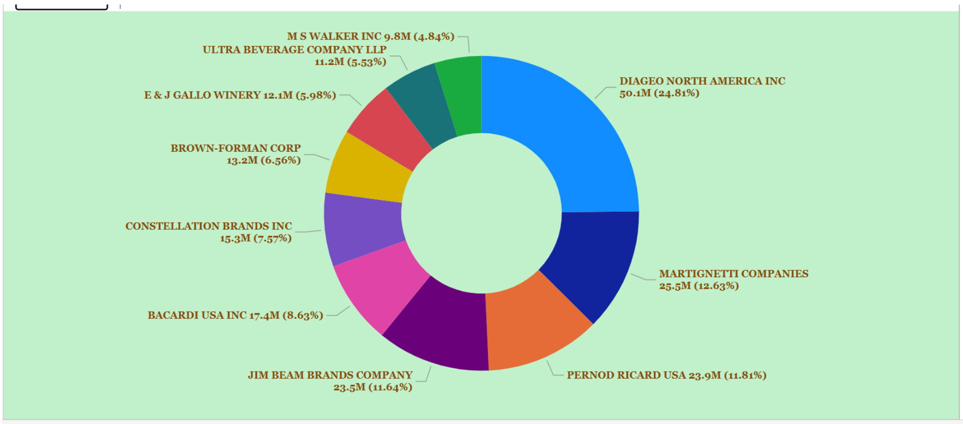

Figure 3: Distribution of Sales across Vendor Tiers

6 Business Insights & Recommendations

Executive Insights

- **Price Volatility:** Five key vendors showed a price variance of $> 15\%$, suggesting a need for fixed-price contracts.
- **Inventory Gaps:** Identified 124 SKUs with high inventory gaps, representing roughly \$45,000 in tied-up capital.
- **Top Performers:** Vendors in the 'Electronics' category showed the highest turnover but the lowest individual margins.

7 Visualization Dashboard

The Power BI dashboard serves as the final delivery vehicle for these insights, allowing stakeholders to filter data by date, category, and vendor region.

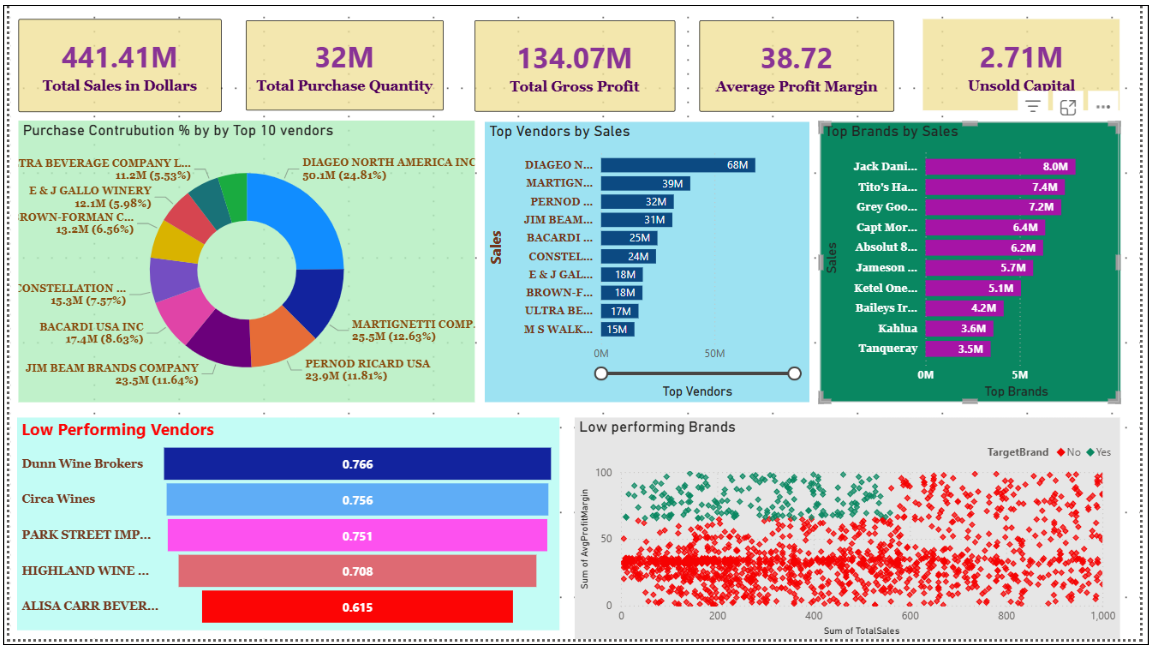

Figure 4: Final Interactive Business Intelligence Dashboard

8 Conclusion

By bridging the gap between raw Python-based data science and interactive Business Intelligence, this project provides a scalable template for vendor auditing. The results facilitate proactive procurement rather than reactive spending.

9 Future Improvements

- **Machine Learning:** Implement K-Means clustering to automatically group vendors into 'Strategic', 'Tactical', and 'High-Risk' categories.
- **Real-time ETL:** Transition from flat files to a cloud-based SQL database with automated refresh cycles.

References

- Dataset Source: Haseeb-Zai30 / Vendors Data
- Project Source Code: GitHub Repository