Table of Contents

Abhishek Varshney

Abhishek Varshney

CHAPTER 1: INTRODUCTION

1.1 Background

Vendor management is a critical component of organizational procurement strategy. Effective monitoring of vendor behavior, purchase patterns, quality consistency, pricing, and delivery performance directly influences operational efficiency and cost management. Traditionally, vendor evaluation is carried out through manual record-keeping or basic ERP-generated reports, which do not provide detailed analytics required for strategic decisions.

VENTrack – Vendor Tracking and Management System – is designed to automate vendor performance evaluation through a structured data pipeline using Python, SQLite, and Power BI. By integrating purchase, sales, freight, and pricing data, the system generates meaningful performance indicators to support informed decision-making.

1.2 Purpose of the Project

The objective of this project is to build a data-driven analytical system that consolidates vendor-related data and presents performance insights through automated processing and interactive dashboards. The system minimizes manual workload, enhances accuracy, and supports procurement departments in identifying top-performing and low-performing vendors.

1.3 Scope of the Project

This project focuses on data integration, preprocessing, KPI calculation, and visualization. It covers all steps from raw data extraction to dashboard creation. The system identifies vendor profitability, purchase contribution, stock turnover, unsold inventory, and sales-to-purchase efficiency.

The scope of the project includes:
• Data extraction and integration
• Data cleaning and transformation
• Creation of Vendor Sales Summary model
• Exploratory data analysis
• Analytical computations
• Visualization using Power BI
The scope excludes vendor contract management, real-time systems, and ERP integration.

1.4 Problem Statement

Organizations often store procurement and sales details in separate files or systems, leading to difficulties in assessing vendor performance. Manually identifying vendor contribution, profit margins, stock movement, or unsold capital becomes time-consuming and error-prone.

There is a need for a unified vendor performance evaluation system that integrates multi-source data, generates automated KPIs, and presents insights through an interactive dashboard.

Abhishek Varshney

CHAPTER 2: LITERATURE REVIEW / EXISTING SYSTEM

2.1 Existing Systems

Existing systems focus on maintaining vendor master data and transaction logs but do not offer integrated analytics. Many organizations rely on spreadsheets and static reports, which limit the ability to evaluate vendor efficiency based on profit, turnover, or contribution metrics.

2.2 Limitations of Existing Systems

• Lack of consolidated data

• No automated KPI computation

• Limited visualization capabilities

• Dependence on manual effort

• Difficulty identifying performance trends or anomalies

VENTrack aims to overcome these limitations by combining data engineering, analytics, and BI reporting.

Abhishek Varshney

CHAPTER 3: PROPOSED SYSTEM

3.1 System Overview
The proposed system integrates raw data files, including purchases, sales, vendor invoices, purchase prices, begin_inventory, and end_inventory, into a unified SQLite database. The system uses Python to preprocess and clean the data, compute KPIs, and produce analytical datasets. Power BI is used to visualize vendor performance using dashboards.

3.2 System Objectives
The system is designed with the following objectives:

Extract and consolidate vendor-related data using Python and SQL.

Clean, preprocess, and validate the data.

Generate analytical metrics such as profit margin, stock turnover, purchase contribution, and estimated profit.

Perform exploratory analysis to understand vendor behavior.

Build interactive dashboards using Power BI to support managerial decisions.

3.3 Features of the System
• Automated ingestion of multiple files

• Chunk-based data loading for large datasets

• Database table cleanup to prevent duplicate ingestion

• KPI calculation including gross profit, profit margin, stock turnover, sales-to-purchase ratio, estimated profit

• Vendor dependency analysis

• Pareto-based purchase contribution analysis

• Analysis of low-performing vendors

• Calculation of unsold inventory values

3.4 Hardware and Software Requirements
Hardware Requirements

• Intel Core i5 or higher

• Minimum 8 GB RAM (recommended 16 GB)

• Minimum 20 GB free storage

• 64-bit Windows 10 or higher

Software Requirements

• Python 3.x

• SQLite

Abhishek Varshney

- Pandas, NumPy, Seaborn, Matplotlib libraries

- Jupyter Notebook or VS Code

- Power BI Desktop

Abhishek Varshney

- Power BI Desktop

CHAPTER 4: SYSTEM DESIGN
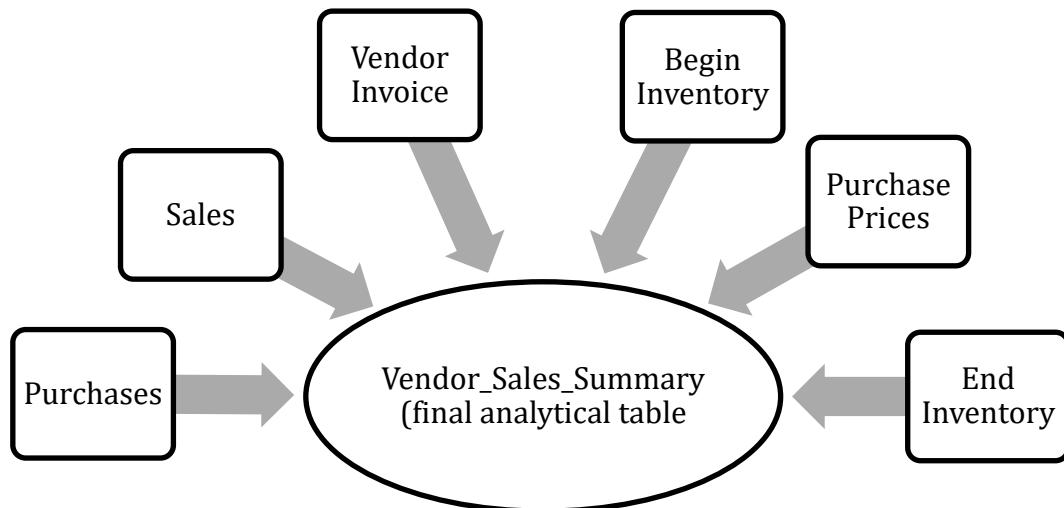
4.1 Modular Architecture

1. The system consists of six core modules:
2. Data Collection Module
3. Data Preprocessing Module
4. Analytical Computation Module
5. Visualization Module
6. Reporting Module
7. User Interface Module (Power BI)
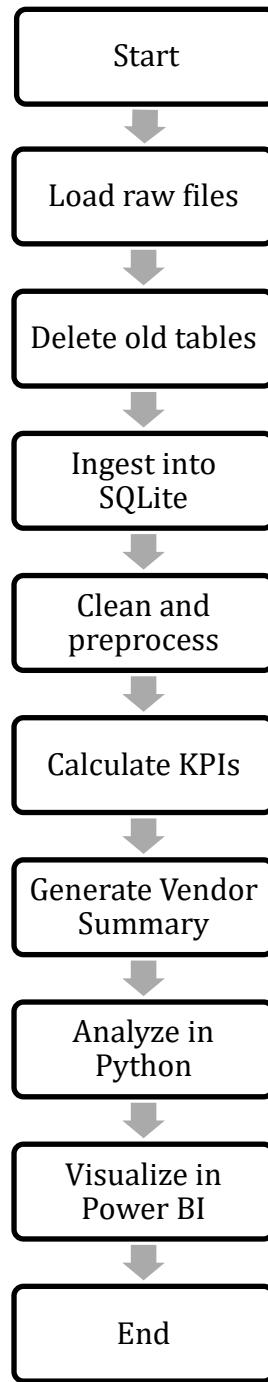
4.2 Data Flow Diagram (DFD Level 0)

user	→	python ingested script	→	SQLite database	→	JUpiter notebooks	→	Vendor summary table	→	Power bi dashboard

4.3 Entity Relationship Diagram

```
   Sales          Vendor          Begin
                  Invoice        Inventory        Purchase
                                                  Prices
Purchases  →   Vendor_Sales_Summary   ←   End
                (final analytical table        Inventory
```

Abhishek Varshney

4.4 Flowchart

```
        ┌─────────────┐
        │    Start    │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │ Load raw files │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │ Delete old tables │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │  Ingest into  │
        │    SQLite     │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │  Clean and   │
        │  preprocess  │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │ Calculate KPIs │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │ Generate Vendor │
        │    Summary    │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │  Analyze in  │
        │    Python    │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │ Visualize in │
        │  Power BI    │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │     End     │
        └─────────────┘
```

Abhishek Varshney

CHAPTER 5: IMPLEMENTATION

5.1 Data Ingestion

The ingestion_db.py script imports CSV files into the SQLite database using smart chunking based on column count. Old tables are dropped to avoid duplication. Logs are generated for monitoring.

5.2 Data Validation

The check.py script verifies whether CSV row counts match database table row counts, ensuring data accuracy.

5.3 Vendor Summary Model

The get_vendor_summary.ipynb notebook uses SQL Common Table Expressions (CTEs) to combine purchase, sales, and freight data. KPIs such as gross profit, estimated profit, profit margin, stock turnover, and sales-to-purchase ratio are computed.

5.4 Exploratory Data Analysis

The eda.ipynb notebook includes histograms, boxplots, correlation matrix, brand-level performance insights, and identification of high-margin low-sales brands.

5.5 Vendor Performance Analysis

The VendorPerformanceAnalysis.ipynb notebook includes purchase contribution analysis, Pareto identification of top 10 vendors, unsold inventory calculations, confidence interval analysis, and hypothesis testing.

5.6 Power BI Dashboard

The final dashboard displays Total Sales, Total Purchase, Profit Margin, Unsold Inventory, Purchase Contribution Donut Chart, Top Vendors and Top Brands by Sales, Low-Performing Vendors, and a scatter plot showing vendor performance patterns.

Abhishek Varshney

CHAPTER 6: TESTING AND RESULTS

6.1 Test Cases

The VENTrack system was tested across key functional and analytical components to ensure accuracy, reliability, and data integrity. The following major tests were conducted:

6.1.1 CSV Ingestion and Data Integrity Testing

• All raw CSV files were successfully ingested into SQLite without loss or mismatch.

• Row counts of each CSV file were validated against their respective database tables.

• Chunk-based ingestion was tested to ensure consistent loading of large files.

6.1.2 Duplicate Prevention Testing

• Old tables were automatically dropped before ingestion.

• No duplication, multi-loading, or redundant row creation occurred during repeated executions.

6.1.3 KPI Computation Validation

• All analytical metrics (Profit Margin, Gross Profit, Stock Turnover, Sales-to-Purchase Ratio, Estimated Profit) were manually cross-verified.

• No computational discrepancies or division errors (e.g., zero sales) were observed.

6.1.4 SQL Join Correctness

• SQL joins across Purchases, Sales, Vendor Invoice, and Purchase Prices were verified.

• All CTEs produced clean merged datasets without missing or duplicated vendor records.

• NULL values were handled properly using IFNULL statements.

6.1.5 Dashboard Visualization Accuracy

• KPI cards, donut chart, bar charts, scatter plots, and contribution visuals correctly reflected backend data.

• Visuals dynamically updated with no inconsistencies between data and display.

6.1.6 Hypothesis Testing Validation

• A two-sample t-test was used to compare profit margins of top-performing and low performing vendors.

• The test returned a statistically significant p-value (<0.05), confirming meaningful performance differences.

• This validated that vendor segmentation based on sales performance is analytically justified.

6.2 Results and Observations

This chapter presents a comprehensive analysis of vendor performance, profitability patterns, sales behavior, pricing dynamics, inventory efficiency, and correlation-based insights derived from the consolidated dataset of 10,692 records. The results are structured into multiple analytical themes as per advanced data analytics reporting standards.

6.2.1 Statistical Summary of Vendor Performance Dataset

A descriptive statistical analysis was conducted on key numerical variables such as purchase price, sales price, total purchase dollars, gross profit, profit margin, and stock turnover. Several important patterns and irregularities were identified.

Abhishek Varshney

	count	mean	std	min	25%	50%	75%	max
VendorNumber	10692.0	10650.649458	1.875352e+04	2.000000	3951.000000	7153.000000	9552.000000	2.013590e+05
Brand	10692.0	18039.228769	1.266219e+04	58.000000	5793.500000	18761.500000	25514.250000	9.063100e+04
AvgPurchasePrice	10692.0	24.385303	1.092694e+02	0.360000	6.840000	10.455000	19.482500	5.681810e+03
TotalPurchaseQuantity	10692.0	3140.886831	1.109509e+04	1.000000	36.000000	262.000000	1975.750000	3.376600e+05
TotalPurchaseDollars	10692.0	30106.693372	1.230678e+05	0.710000	453.457500	3655.465000	20738.245000	3.811252e+06
AvgActualPrice	10692.0	35.643671	1.482460e+02	0.490000	10.990000	15.990000	28.990000	7.499990e+03
TotalVolume	10692.0	265550.515853	1.069263e+06	50.000000	2250.000000	13500.000000	114750.000000	2.484000e+07
TotalSalesDollars	10692.0	42239.074419	1.676553e+05	0.000000	729.220000	5298.045000	28396.915000	5.101920e+06
TotalSalesPrice	10692.0	18793.783627	4.495277e+04	0.000000	289.710000	2857.800000	16059.562500	6.728193e+06
TotalSalesQuantity	10692.0	3077.482136	1.095285e+04	0.000000	33.000000	261.000000	1929.250000	3.349390e+05
TotalExciseTax	10692.0	1774.226259	1.097558e+04	0.000000	4.800000	46.570000	418.650000	3.682428e+05
FreightCost	10692.0	61433.763214	6.093846e+04	0.090000	14069.870000	50293.620000	79528.990000	2.570321e+05
EstimatedProfit	10692.0	-49301.382166	7.121940e+04	-271965.100000	-74160.377500	-30361.360000	-7724.617500	1.222066e+06
GrossProfit	10692.0	12132.381048	4.622434e+04	-52002.780000	52.920000	1399.640000	8660.200000	1.290668e+06
ProfitMargin	10692.0	-15.620770	4.435553e+02	-23730.638953	13.324515	30.405457	39.956135	9.971666e+01
SalesToPurchaseRatio	10692.0	2.504390	8.459067e+00	0.000000	1.153729	1.436894	1.665449	3.529286e+02
StockTurnover	10692.0	1.706793	6.020460e+00	0.000000	0.807229	0.981529	1.039342	2.745000e+02

6.2.1.1 Skewness and Heavy-Tailed Distributions

From the histograms and summary statistics:

- TotalPurchaseDollars, TotalSalesDollars, and TotalVolume displayed extreme right-skewness.

- Very high maximum values indicate the presence of bulk product lines or wholesale-level purchases.

- A small number of vendors contribute disproportionately to sales and purchases.
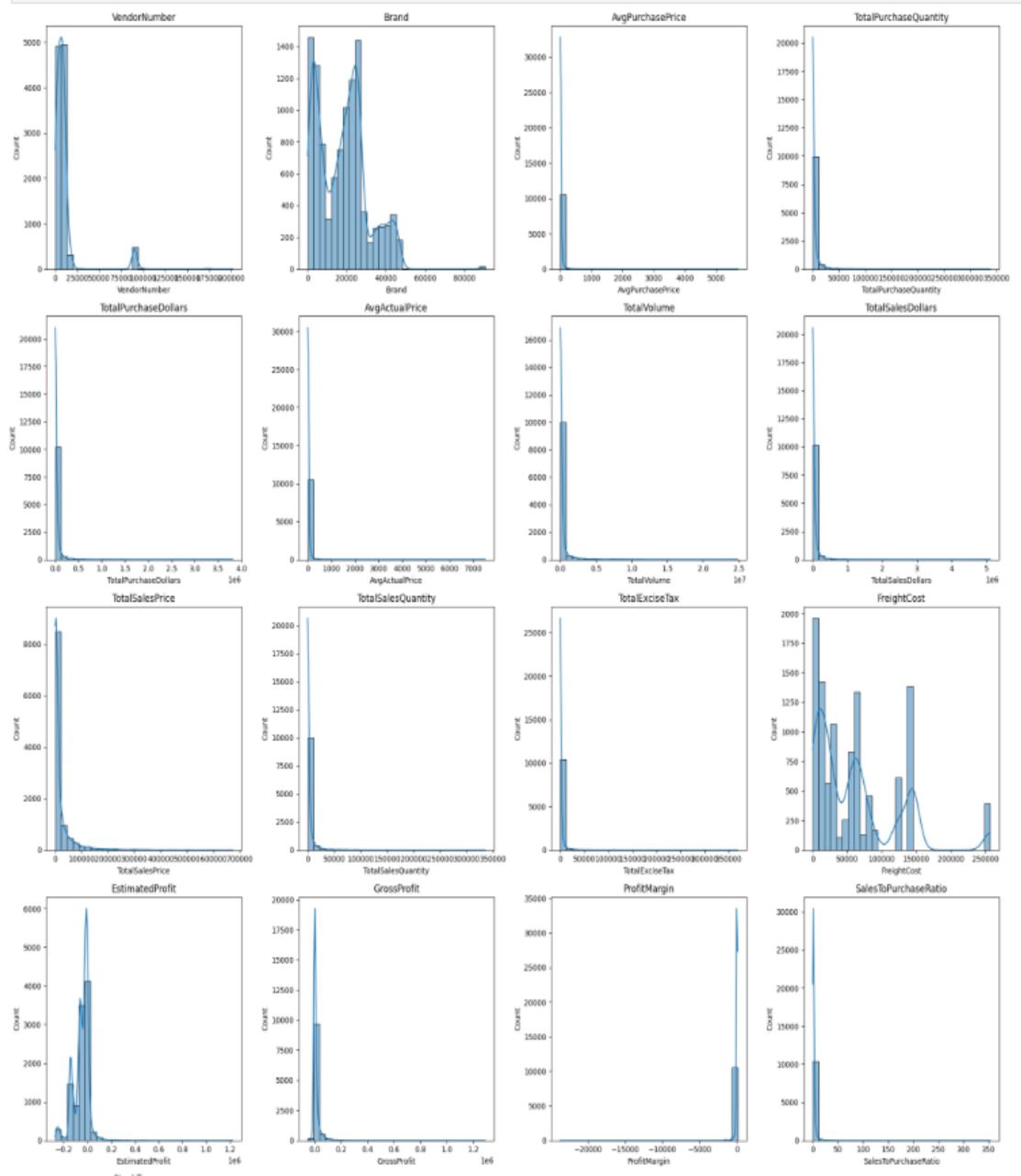
This pattern confirms a Pareto effect in vendor performance.

6.2.1.2 Negative and Zero Values

Several KPIs showed negative or zero values, suggesting inefficiencies:

- Gross Profit ranged from -52,002.78 to over 1.29 million.

- Profit Margin recorded values as low as negative infinity when sales revenue was zero.

- TotalSalesQuantity = 0 for some entries indicates unsold or obsolete stock.

These cases required filtering to ensure meaningful analysis.

Abhishek Varshney

6.2.2. Data Filtering Observations

To enhance analytical reliability, records were filtered using:

- GrossProfit > 0

- ProfitMargin > 0

- TotalSalesQuantity > 0

Abhishek Varshney

6.2.4 Correlation-Based Insights

6.2.4.1 Purchase Price vs Sales Dollars and Gross Profit

Correlation values were near zero. Price variations do not significantly impact revenue or profit.

#### 6.2.4.2 Total Purchase Quantity vs Total Sales Quantity
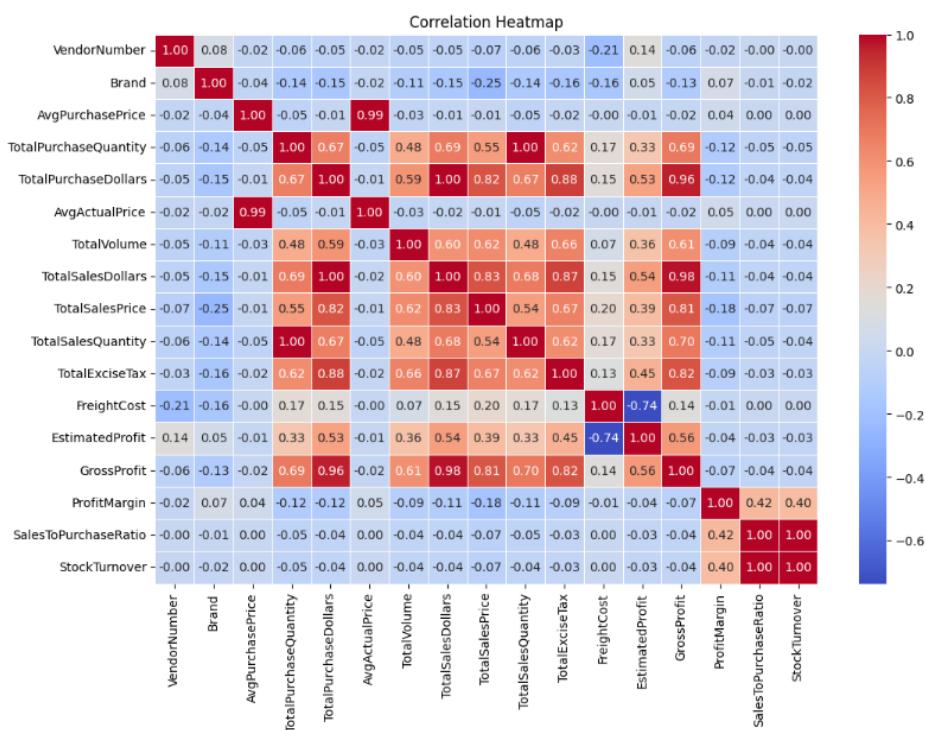
A very strong correlation (0.999) was observed, confirming efficient inventory turnover.

6.2.4.3 Profit Margin vs Total Sales Price

A moderate negative correlation suggests competitive pressures reduce margins on high-priced items.

6.2.4.4 Stock Turnover vs Gross Profit and Profit Margin

Weak negative correlation indicates faster turnover does not guarantee higher profitability.

Correlation Heatmap

6.2.5 Research Questions and Key Findings

6.2.5.1 High-Margin Low-Sales Brands

198 brands were identified as high-profit but low-sales, suitable for marketing and pricing adjustments.

Abhishek Varshney

Target Brands Found: 198
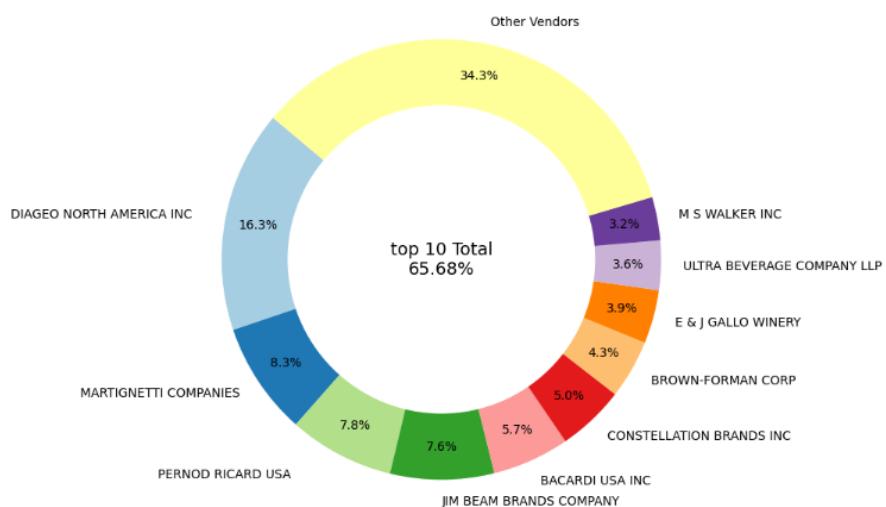
	Description	TotalSalesDollars	ProfitMargin
6199	Santa Rita Organic Svgn Bl	9.99	66.466466
2369	Debauchery Pnt Nr	11.58	65.975820
2070	Concannon Glen Ellen Wh Zin	15.95	83.448276
2188	Crown Royal Apple	27.86	89.806174
6237	Sauza Sprklg Wild Berry Marg	27.96	82.153076
...	...	...	...
5074	Nanbu Bijin Southern 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

6.2.6 Vendor Contribution and Dependency Analysis

Top 10 vendors contribute approximately 65.69 percent of total purchases, indicating dependence on a small vendor group and the need for diversification.

Purchase Contribution Distribution: Top 10 Vendors vs Others

6.2.7 Inventory Efficiency and Capital Utilization

Unsold inventory analysis highlighted slow-moving or dead stock, requiring procurement optimization.

6.2.8 Profitability Segmentation and Statistical Testing

Confidence interval and hypothesis testing confirmed significant differences between top- and low-performing vendors.

6.2.9 Overall Summary

Abhishek Varshney

- Vendor contribution is unevenly distributed.

- Profit performance varies significantly across vendors.

- High-margin low-sales brands present strategic opportunities.

- Unsold inventory indicates inefficiencies.

- Strong purchase-sales correlation confirms operational accuracy.

- Statistical evidence supports performance-based vendor segmentation.

Abhishek Varshney

CHAPTER 7: CONCLUSION AND FUTURE SCOPE

7.1 Conclusion

VENTrack provides a reliable, structured, and analytical approach to vendor performance evaluation by automating data ingestion, integrating datasets, computing KPIs, and offering dashboard-driven decision support.

7.2 Future Scope

Integration with ERP systems, real-time performance alerts, predictive vendor ranking using machine learning, automated reporting, and additional KPI integration.

Abhishek Varshney