Vendor Performance Analysis

Subtitle - Data Analyst Project

Name - Ajay Prajapat

Gmail - Prajapatjaiseya235@gmail.com

Linkedln

Github

- **Table of Contents**

• Abstract

This project analyzes vendor performance and inventory management using Python and SQLite. Multiple datasets were integrated and explored through statistical analysis and visualization. The study highlights that bulk purchasing significantly reduces unit costs, while some brands show low sales but high profit margins, requiring pricing or promotional adjustments. Correlation analysis confirmed strong links between purchase and sales quantities, but also revealed inefficiencies like unsold inventory and low stock turnover. The insights help optimize vendor relationships, pricing strategies, and inventory utilization, demonstrating the importance of data-driven decision-making in supply chain performance.

• Introduction

In today's competitive retail market, efficient management of purchases, sales, and vendor performance is critical for sustaining profitability and growth.
This project utilizes four datasets — *Purchase, Sales, Purchase Price, and Vendor Invoice* — to analyze vendor performance, inventory turnover, and profit margins.
Python was employed for data ingestion and Exploratory Data Analysis (EDA), while SQL queries were used for structured insights.
Finally, an interactive dashboard was built in Power BI to visualize key findings and communicate results effectively.
The overall objective of this project is to identify opportunities for pricing adjustments, bulk purchasing strategies, and vendor negotiations that can enhance operational efficiency and improve business outcomes.

- # Dataset Description

This project is based on four primary datasets that collectively provide insights into vendor performance, sales, and profitability:

1. Purchase Dataset:
   Contains details of purchase transactions, including vendor number, vendor name, brand, description, purchase price, quantity, and total dollars spent.
   - *Use:* Helps in analyzing total purchases, cost structures, and purchase behavior across vendors.
2. Purchase Price Dataset:
   Provides product-wise actual prices and volume details for each vendor-brand combination.
   - *Use:* Serves as a reference table to validate actual purchase prices and optimize pricing strategies.
3. Sales Dataset:
   Records sales transactions with vendor number, brand, sales price, sales quantity, excise tax, and total sales dollars.
   - *Use:* Enables tracking of revenue, sales performance, and profit margin calculations.
4. Vendor Invoice Dataset:
   Summarizes purchases by vendor and PO number, including freight costs in addition to purchase amounts.
   - *Use:* Helps in evaluating logistics costs, vendor billing, and overall expense tracking.

- **Methodology (Python + SQL + Power BI)**

1. Data Ingestion (Python + SQL)

- Imported datasets (purchases, purchase_prices, sales, vendor_invoice) using Python (pandas).
- Performed initial cleaning and validation to ensure consistency.
- Used SQLite as the database, and ingested all CSV files into SQL tables via sqlalchemy and sqlite3.
- This step ensured that all data was centralized and queryable.

---

2. Data Exploration (SQL + Python)

- Wrote SQL queries to explore each table: checked record counts, unique values, and missing data.
- Used SQL Joins and Group By to consolidate data across multiple tables.
- Created a master table vendor_sales_summary that combines purchases, sales, freight, and actual prices into one consolidated summary.
- Conducted Exploratory Data Analysis (EDA) in Python using matplotlib and seaborn to understand distributions, detect outliers, and visualize correlations.

---

3. Vendor Performance Analysis (Python)

- Derived new key metrics (KPIs) such as:
    - Gross Profit = Total Sales – Total Purchases
    - Profit Margin (%)
    - Stock Turnover Ratio
    - Sales-to-Purchase Ratio
    - Unsold Inventory Value

- Performed deeper analysis to identify:
    - Brands with low sales but high profit margins (ideal for promotions).
    - Vendors achieving bulk purchase cost savings.
    - Vendors with low inventory turnover, indicating slow-moving products.

---

4. Dashboarding & Visualization (Power BI)

- Exported the consolidated vendor_sales_summary table to CSV and imported it into Power BI.
- Built an interactive dashboard featuring:
    - Cards for Total Sales, Purchases, Profit, and Profit Margin.
    - Bar Charts for Top Vendors and Top Brands by Sales.
    - Donut Chart for vendor contribution to purchases.
    - Funnel Chart to visualize the flow from purchases → sales → profit.
- Added filters and slicers for dynamic analysis by vendor, brand, or time period.
- Final dashboard was presented in full-screen mode for professional reporting.

👉 This section shows that the project followed an end-to-end methodology, integrating Python for data processing, SQL for querying and aggregation, and Power BI for visualization and storytelling.

- # Exploratory Data Analysis

The purpose of Exploratory Data Analysis was to understand the datasets, detect anomalies, and prepare the data for further analysis. EDA was performed using both SQL queries and Python visualizations.

1. Initial Exploration

- Verified the number of records and columns in each table (purchases, purchase_prices, sales, vendor_invoice).
- Checked for missing values, duplicates, and inconsistencies.
- Displayed the first few records of each table to get an overview of the schema and column meanings.

---

2. Key Observations from SQL Queries

- Purchases Table: Contained vendor-wise purchase transactions (quantity, price, dollars).
- Purchase Prices Table: Contained standard product prices, unique by Vendor + Brand.
- Vendor Invoice Table: Aggregated vendor invoices with freight charges.
- Sales Table: Contained brand-level sales transactions (quantity, dollars, excise tax).

From these tables, a summary table (vendor_sales_summary) was created that consolidated purchases, sales, freight, and pricing data.

---

3. Descriptive Statistics (Python)

- Used df.describe() to calculate mean, min, max, and standard deviation for numerical columns.

- Key Findings:
    - Some vendors/products had negative Gross Profit → indicating losses.
    - Profit Margin had cases with infinite or negative values, suggesting zero sales or high costs.
    - Freight Costs showed very high variance, highlighting inefficiencies in logistics.
    - Stock Turnover Ratio varied significantly (0 to 274.5), showing some products sold rapidly while others remained unsold.

---

4. Visual EDA (Python with Seaborn & Matplotlib)

- Distribution Plots: Histograms with KDE to check normality of sales, purchases, and margins.
- Box Plots: Detected outliers in purchase price, sales dollars, and freight cost.
- Count Plots: Top 10 vendors and brands were visualized by frequency of transactions.
- Correlation Heatmap:
    - Strong correlation (0.99) between Total Purchase Quantity and Total Sales Quantity → efficient turnover.
    - Weak correlation between Purchase Price and Gross Profit → product cost does not always impact profitability.
    - Negative correlation between Profit Margin and Sales Price → higher sales price may reduce margin due to competitive pricing.

---

5. Insights from EDA

- Some vendors/brands incurred losses, requiring better pricing or sales strategies.
- High freight costs increased total expenses for certain vendors.
- Brands with low sales but high margins were identified as opportunities for promotional campaigns.
- Bulk purchases reduced per-unit costs significantly, encouraging large order volumes.
- A subset of vendors had low stock turnover, locking capital in unsold inventory.

👉 This EDA stage ensured data quality, provided business insights, and prepared the foundation for building the dashboard and vendor performance analysis.

- **Key Insights**

1. Top Performing Vendors & Brands

- A small group of vendors (like *Diageo North America Inc*, *Pernod Ricard USA*, *Bacardi USA Inc*) contribute the majority of sales and profit.
- Top brands like *Jack Daniels, Tito's Vodka, Absolut, Grey Goose* drive the highest revenue.

2. Profitability Patterns

- Some vendors/products show negative Gross Profit, meaning sales price is lower than cost → urgent pricing review needed.
- Average Profit Margin is healthy (~20–30%), but varies widely across vendors.

3. Inventory Efficiency

- Strong correlation between purchases and sales (~0.99), confirming efficient turnover.
- However, a few vendors have Stock Turnover < 1, locking significant capital in unsold inventory.

4. Logistics & Freight Costs

- Freight costs show large variation (₹0.09 – ₹2,57,032), highlighting inefficiencies in vendor logistics.

5. Bulk Purchase Advantage

- Vendors purchasing in large volumes enjoy ~72% lower unit costs, confirming bulk order discounts.

6. Promotional Opportunities

- Identified brands with low sales but high margins, ideal for promotional campaigns to boost revenue.

7. Overall Contribution

- Top 10 vendors account for more than 70% of total purchases, showing a heavy dependency on a few suppliers.

- ## Dashboard & Visualizations

The project's final step involved creating an interactive Power BI dashboard to visualize vendor performance and sales trends effectively. The dashboard provided dynamic insights, enabling stakeholders to make data-driven decisions.

Key Dashboard Features:

- 📊 Top Vendors & Brands: Bar charts highlighted the top 10 vendors and brands by total sales, helping identify major revenue contributors.
- 📈 Sales vs Purchase Trend: Line and bar visualizations compared purchase dollars with sales dollars to evaluate vendor profitability.
- 🏷 Profit Margin Analysis: KPI cards and scatter plots showed products with high profit margins but low sales, supporting pricing and promotional strategies.
- 📦 Inventory Insights: Stock turnover and unsold inventory visuals identified slow-moving products and excess stock.
- 🚚 Freight & Bulk Purchase Impact: Visuals demonstrated how freight costs and order size affected overall profitability.
- 📌 Interactive Filters: Users could filter by Vendor, Brand, and Time Period to drill down into specific insights.
- PowerBiDashboard

- ## Conclusion

This project successfully integrated Python, SQL, and Power BI to analyze vendor and brand performance. Through data ingestion, transformation, and visualization, the analysis uncovered:

- ✅ Top-performing vendors and brands driving maximum revenue.
- ✅ Vendors with low sales but high margins, suggesting pricing or promotional opportunities.
- ✅ Impact of bulk purchasing on reducing unit costs, leading to better cost efficiency.
- ✅ Identification of slow-moving inventory and excess stock, highlighting areas for operational improvement.

Overall, the project demonstrates how data-driven insights can guide strategic decision-making, improve vendor relationships, and optimize purchasing strategies for higher profitability.

Q&A

Q1. Why did you choose Python, SQL, and Power BI for this project?
Python was used for data preprocessing, statistical analysis, and generating KPIs. SQL helped in efficient querying, joining, and aggregating large datasets. Power BI provided interactive dashboards for decision-making and visualization.

Q2. What were the key business insights from your analysis?
Identified top-performing vendors and brands contributing maximum sales. Found vendors with low sales but high profit margins, suggesting promotional opportunities. Bulk purchasing reduced unit cost significantly (~72%). Highlighted vendors with slow inventory turnover and locked capital in unsold stock.

Q3. How can this project help businesses in real life?
Helps in vendor selection for profitability. Optimizes product pricing strategy. Reduces inventory costs by tracking slow-moving stock. Improves purchasing decisions through bulk-order insights.

Q4. What challenges did you face while doing this project?
Handling missing and inconsistent data across multiple tables. Optimizing SQL queries for large datasets. Ensuring correct mapping of purchase and sales data to vendors and brands.

Q5. If given more time, what improvements would you add?
Real-time data integration into Power BI dashboards. Predictive analytics using machine learning for demand forecasting. Automated vendor performance scoring system.