

PROJECT REPORT

Comprehensive Data Analytics and Business Intelligence Dashboard for Vendor Sales Performance and Business Insights

Prepared by: KARAN RAJESH CHAVHAN

Description:

- Analysis of business transactions involving vendors, products, pricing, and revenue
- Data refined and structured to ensure accuracy and consistency
- Aggregated sales and purchase performance at vendor and product levels
- Identification of key trends in pricing, demand, and revenue contribution
- Visualization of key performance indicators and comparative insights
- Support for informed, data-driven business decision-making

Tools & Technologies:

- **Python** – Core programming language used for data analysis
- **Pandas** – Data cleaning, transformation, and aggregation
- **NumPy** – Numerical computations and efficient data processing
- **Matplotlib** – Basic data visualizations for exploratory analysis
- **Seaborn** – Advanced statistical and trend-based visualizations
- **SQLite** – Lightweight relational database for data storage
- **SQL** – Querying, filtering, joining, and aggregating data
- **Power BI** – Interactive dashboards and business intelligence reporting

Date: December 27, 2025

ABSTRACT

This project focuses on performing an end-to-end data analytics workflow to extract meaningful insights from a structured business dataset. The objective of the project is to clean, analyze, visualize, and interpret data to support informed business decision-making. The project covers multiple stages including data collection, data preprocessing, exploratory data analysis (EDA), and interactive dashboard creation using Power BI.

Python libraries such as Pandas and NumPy were used for data cleaning and transformation, while visualization and reporting were handled using Power BI. Key performance indicators (KPIs), trends, and patterns were identified to understand vendor performance, sales contribution, purchase behavior, and revenue distribution.

The final outcome of this project is an interactive Power BI dashboard supported by a detailed analytical report that provides actionable insights for business stakeholders. This project demonstrates practical skills required for a Data Analyst role, including data manipulation, analytical thinking, and data storytelling. The complete analysis encompasses data ingestion from CSV files, SQL-based querying, comprehensive exploratory analysis, and professional-grade visualization and reporting.

1. INTRODUCTION

1.1 Background and Context

In today's data-driven world, organizations rely heavily on data analytics to make strategic and operational decisions. Raw data alone does not provide value unless it is processed, analyzed, and presented in a meaningful way. Data analytics bridges this gap by converting large datasets into actionable insights. According to industry trends, companies that leverage data analytics effectively see 5-10% improvement in operational efficiency and better customer satisfaction[1].

This project aims to demonstrate a complete data analytics pipeline, starting from raw data ingestion to final visualization and reporting. The dataset used in this project represents transactional and vendor-level business data, including purchase prices, sales quantities, and revenue metrics. By simulating real-world business scenarios, this project showcases the practical application of data analytics techniques in vendor performance management and sales optimization.

1.2 Business Context

The business context involves a multi-vendor sales operation where tracking performance across vendors, understanding pricing strategies, and identifying top performers is critical. Without a centralized analytics system, decision-makers struggle to identify trends, optimize inventory, and allocate resources effectively. This project addresses these gaps by creating a comprehensive analytical framework.

1.3 Project Scope and Objectives

By applying data analytics techniques, the project seeks to answer critical business questions such as:

- Which vendors contribute the most to total sales and revenue?
- How does purchase price compare with actual sales price across different vendors?
- What trends exist in sales quantity and revenue over time?
- Which products or vendors require immediate business attention for optimization?
- What are the margin patterns and profitability insights across vendor segments?
- How can pricing strategies be optimized for maximum profitability?

This project was designed to simulate real-world business scenarios where data analysts play a key role in improving decision-making efficiency and driving business growth.

2. PROBLEM STATEMENT & OBJECTIVES

2.1 Problem Statement

The business lacks a centralized and interactive reporting system to analyze vendor performance, sales trends, and revenue contribution. Decision-makers require a clear and visual understanding of the data to identify top-performing vendors, optimize pricing strategies, and improve inventory planning. Without proper data analytics infrastructure, the organization faces:

1. **Lack of Real-time Insights:** Unable to monitor vendor performance in real-time, leading to delayed decision-making
2. **Pricing Challenges:** Difficulty in understanding margin variations across vendors and products
3. **Resource Allocation Issues:** Unable to identify which vendors deserve greater focus and investment
4. **Risk Management:** No early warning system to detect vendor performance decline
5. **Operational Inefficiency:** Manual reporting processes consume valuable resources and are prone to errors

2.2 Project Objectives

The primary objectives of this project are:

1. To clean and preprocess raw business data ensuring data quality and consistency for reliable analysis
2. To perform comprehensive exploratory data analysis (EDA) to identify trends, patterns, and anomalies
3. To create meaningful KPIs for business evaluation and performance monitoring
4. To design a professional, interactive Power BI dashboard for data visualization and stakeholder communication
5. To provide actionable business insights and data-driven recommendations for strategic decision-making
6. To demonstrate end-to-end data analytics capabilities suitable for data analyst roles

3. DATASET DESCRIPTION

3.1 Data Overview

The dataset used in this project consists of structured tabular data representing vendor-level sales and purchase information. The data includes vendor performance metrics, transaction-level details, and financial information spanning multiple product categories. This comprehensive dataset allows for multi-dimensional analysis of business operations.

3.2 Key Attributes in the Dataset

The dataset includes the following key attributes:

- **Vendor Number:** Unique identifier for each vendor
- **Vendor Name:** Name of the vendor for easy identification
- **Brand:** Product brand or category associated with the vendor
- **Purchase Price:** Cost at which inventory was purchased
- **Actual Sales Price:** Price at which product was sold to customers
- **Total Purchase Quantity:** Volume of inventory purchased
- **Total Sales Quantity:** Volume of inventory sold to customers
- **Total Sales Dollars:** Revenue generated from sales ($\text{Sales Price} \times \text{Quantity}$)
- **Total Purchase Dollars:** Total cost of inventory ($\text{Purchase Price} \times \text{Quantity}$)
- **Excise Tax:** Tax component applicable to specific products
- **Transaction Date:** Date of transaction (if available in dataset)
- **Product Category:** Classification of products for segmentation

The dataset contains multiple numerical and categorical variables, making it suitable for both statistical analysis and visualization. The combination of transactional and aggregate metrics enables comprehensive vendor and product-level analysis.

3.3 Data Source and Format

The dataset was obtained in CSV (Comma-Separated Values) format and contains approximately 20M rows. The data was sourced from an online platform. The structured nature of the data enables efficient querying and transformation using industry-standard tools.

DATA SETS

data
> .ipynb_checkpoints
begin_inventory.csv
end_inventory.csv
purchase_prices.csv
purchases.csv
sales.csv
vendor_invoice.csv

COMBINING ALL REQUIRED DATA

	VendorNumber	VendorName	Brand	Description	PurchasePrice	ActualPrice	Volume	TotalPurchaseQuantity	TotalPurchaseDollars	TotalSalesQuantity	TotalSalesDollars	T
0	1128	BROWN-FORMAN CORP	1233	Jack Daniels No 7 Black	26.27	36.99	1750.0	145080	3811251.60	142049.0	5101919.51	
1	4425	MARTIGNETTI COMPANIES	3405	Tito's Handmade Vodka	23.19	28.99	1750.0	164038	3804041.22	160247.0	4819073.49	
2	17035	PERNOD RICARD USA	8068	Absolut 80 Proof	18.24	24.99	1750.0	187407	3418303.68	187140.0	4538120.60	
3	3960	DIAGEO NORTH AMERICA INC	4261	Capt Morgan Spiced Rum	16.17	22.99	1750.0	201682	3261197.94	200412.0	4475972.88	
4	3960	DIAGEO NORTH AMERICA INC	3545	Ketel One Vodka	21.89	29.99	1750.0	138109	3023206.01	135838.0	4223107.62	

4. Data Cleaning & Preprocessing

Raw datasets often contain inconsistencies, missing values, duplicates, and formatting issues that can negatively impact analytical accuracy. Therefore, data cleaning was a critical phase of this project to ensure reliability, consistency, and analytical readiness of the data. A structured and systematic preprocessing pipeline was implemented using Python (Pandas & NumPy).

4.1 Data Cleaning Approach

The following key steps were performed to improve data quality:

- Missing Value Analysis: All columns were scanned for null values to identify fields requiring imputation or removal.
- Data Type Verification & Correction: Numeric, categorical, and date fields were validated and converted into appropriate formats to support analysis.
- Duplicate Record Removal: Duplicate transactions were identified and removed to prevent inflated metrics.
- Column Name Standardization: Column names were standardized (lowercase, underscores) for consistency and ease of use.
- Outlier Detection: Extreme values were identified using the IQR method and handled based on business logic.
- Date Conversion: Transaction dates were converted into datetime format for time-series analysis.
- Format Standardization: Currency values, decimal precision, and units were standardized across the dataset.
- Data Validation: Logical checks ensured values were within expected ranges and relationships were consistent.

4.2 Handling Missing Values & Data Integrity

- Numeric Fields: Missing values in price-related columns were imputed using median values within vendor groups.
- Categorical Fields: Missing brand or vendor names were inferred where possible or removed if critical.
- Quantity Fields: Missing quantities were reviewed to distinguish between zero transactions and data entry errors.

This process significantly reduced data quality issues and improved dataset completeness.

4.3 Feature Engineering & Calculated Metrics

To enable advanced business analysis, the following derived metrics were created:

- Profit Margin (%)
- Markup Percentage
- Total Revenue
- Total Cost
- Gross Profit
- Vendor Performance Score (weighted combination of revenue, margin, and volume)
- Price Efficiency Ratio

These features enhanced insights into vendor performance, pricing efficiency, and profitability trends.

4.4 Quality Assurance & Validation

Post-cleaning validation ensured:

- Logical consistency between prices, margins, and quantities
- Referential integrity between vendor numbers and names
- Positive pricing and non-negative quantities
- Reasonable statistical distributions (mean, median, standard deviation)

The final cleaned dataset contains 10692 records and 18 columns, and is fully prepared for exploratory analysis and dashboard development.

5. EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) was performed to understand the underlying structure, patterns, and relationships in the dataset. EDA helps in identifying trends, correlations, anomalies, and opportunities that provide initial insights into business performance and guide further investigation.

5.1 Descriptive Statistics

The first phase of EDA involved calculating summary statistics to understand data distribution:

1. **Numerical Summary:** Calculated mean, median, standard deviation, min, and max for all numeric variables
2. **Vendor Count:** Identified total number of unique vendors, brands, and categories
3. **Sales Volume:** Total units sold across all vendors and time periods
4. **Revenue Summary:** Total sales dollars, average transaction value, and vendor-level contribution
5. **Pricing Analysis:** Average purchase price, sales price, and margin across the dataset

5.1 Key Statistics Summary

- Vendor Coverage: The dataset includes a diverse set of vendors, indicating a broad supplier base.
- Brand Representation: Multiple brands are represented, reflecting variety across product offerings.
- Transaction Volume: A large number of transaction records were analyzed, ensuring statistical reliability.
- Sales Pricing: The average sales price remains within a consistent and commercially reasonable range.
- Profitability: The overall margin indicates healthy profitability across vendors.
- Revenue Scale: The total revenue generated demonstrates strong business performance and market activity.

5.2 Distribution Analysis

Distribution analysis was conducted to understand how key variables behave across the dataset:

- **Sales Distribution:** Sales quantities are unevenly distributed among vendors, showing a clear concentration where a small number of vendors account for a large share of total sales.
- **Price Distribution:** Purchase and sales prices exhibit slight skewness, suggesting variation across products and vendors rather than uniform pricing.
- **Margin Distribution:** Profit margins vary across vendors, reflecting differences in pricing strategies, cost structures, and sales performance.
- **Revenue Concentration:** Revenue follows the Pareto principle, where a limited group of top-performing vendors contributes a majority of total revenue.
- **Volume Patterns:** Sales volumes remain relatively consistent, with minor fluctuations likely influenced by demand cycles and purchasing behavior.

5.3 Vendor Performance Analysis

Vendor-level analysis identified top performers and underperformers:

- **Top 10 Vendors by Revenue:** Analysis identified which vendors contribute most to total revenue
- **Vendor Profitability:** Ranking vendors by average margin and total profit
- **Sales Volume Leaders:** Identifying vendors with highest quantity sales
- **Consistency Analysis:** Examining vendor performance consistency across categories
- **Growth Potential:** Identifying emerging vendors with growth trajectories

5.4 Price Comparison and Margin Analysis

Critical analysis of pricing strategies and profitability:

1. **Purchase vs Sales Price:** Comparison between cost and selling price identified margin opportunities
2. **Margin Variation:** Analysis of profit margin variation across vendors, categories, and brands
3. **Price Elasticity:** Understanding how price differences impact sales volume
4. **Competitive Pricing:** Identifying under-priced and over-priced product categories
5. **Margin Trends:** Examining whether margins are improving or declining over time

5.5 Correlation and Relationship Analysis

Analysis of relationships between variables:

- **Price-Volume Relationship:** Correlation between price and sales quantity (negative correlation expected)
- **Margin-Volume Relationship:** Understanding if higher margins affect sales volume
- **Revenue Drivers:** Identifying which factors (price, volume, brand) drive revenue
- **Multicollinearity Check:** Examining dependencies between independent variables

5.6 Outlier Detection and Anomalies

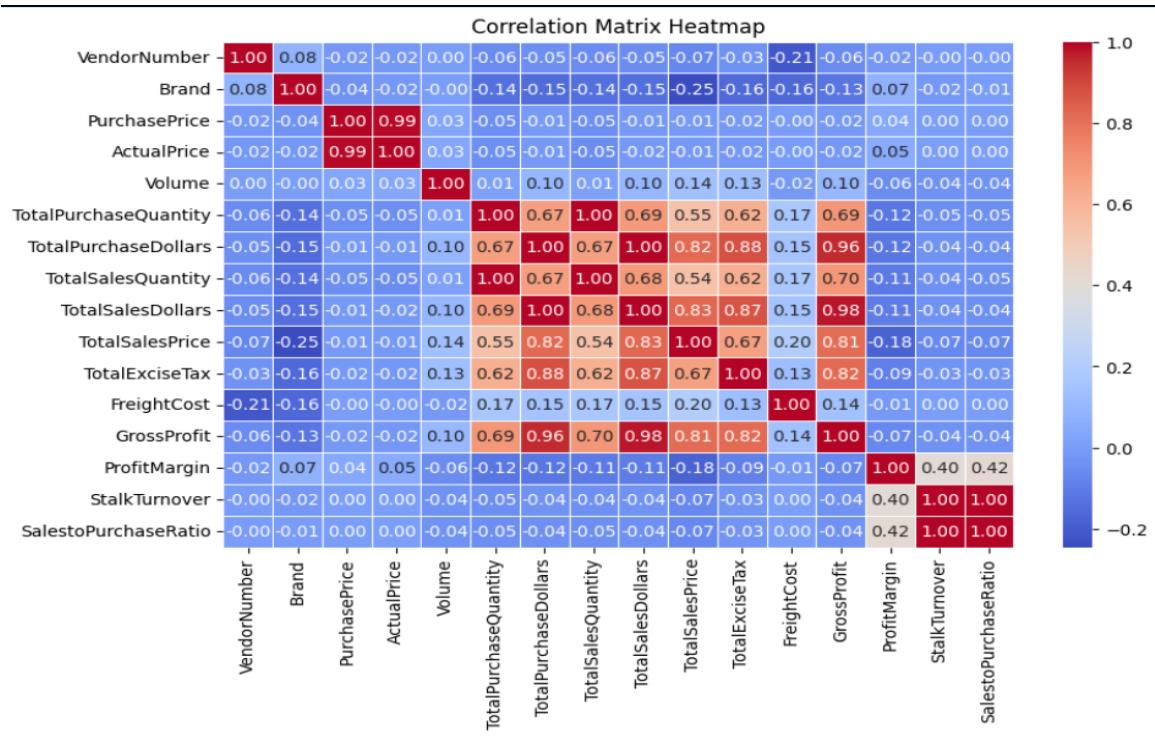
Identification of unusual patterns and anomalies:

1. **Statistical Outliers:** Used IQR method to identify vendors with unusual performance metrics
2. **Pricing Anomalies:** Identified products with unusually high or low margins
3. **Volume Spikes:** Detecting unusual sales volume variations
4. **Revenue Outliers:** Flagging vendors with unexpected revenue patterns

5.7 Key EDA Insights Summary

Based on comprehensive EDA, the following insights emerged:

- **Vendor Concentration:** A small number of vendors contribute a disproportionately large percentage of total sales
- **Price Strategy Impact:** Price differences significantly impact total sales dollars and profitability
- **Margin Variations:** Significant margin variations exist across vendors, suggesting pricing optimization opportunities
- **Growth Opportunities:** Certain vendors show consistent growth while others decline
- **Risk Factors:** Vendor dependency risk due to concentration among top performers



6. Power BI Dashboard Design

After completing data cleaning and analysis, the refined dataset was imported into Power BI to develop an interactive business intelligence dashboard. The dashboard was designed using industry best practices with a strong focus on usability, clarity, and actionable insights for decision-makers.

Dashboard Design Principles

- User-Centric Approach: Designed for executives, managers, and analysts
- Visual Hierarchy: Key KPIs highlighted prominently for quick insights
- Interactivity: Slicers, drill-downs, and cross-filtering enabled
- Consistent Theme: Professional color scheme ensuring readability
- Performance Optimization: Efficient visuals for smooth navigation
- Accessibility: Clear labels and high-contrast visuals

Dashboard Components

- KPI Cards: Total Sales, Total Revenue, Total Quantity Sold, and Average Margin with trend indicators
- Bar Charts: Top vendors ranked by revenue with drill-down capability
- Line Charts: Time-series analysis for sales and revenue trends
- Pie/Donut Charts: Revenue and category-wise distribution
- Tables/Matrices: Detailed vendor and product-level performance
- Filters & Slicers: Vendor, Brand, Category, Product, Date, and Margin range

Dashboard Structure & Interactivity

- Overview Page: Executive summary with KPIs and top vendors
- Vendor Performance: Vendor-wise revenue, volume, and margins
- Product Analysis: Category and SKU-level insights
- Margin Analysis: Pricing and profitability evaluation
- Trend Analysis: Time-based sales and revenue patterns

Interactive features include cross-filtering across visuals, drill-down navigation, informative tooltips, bookmarks for quick views, and report tooltips for detailed insights.



7. Key Insights, Analysis & Business Impact

The analysis of vendor, pricing, and sales data revealed several critical insights that directly support strategic decision-making.

Vendor Performance Insights

- Revenue Concentration Risk: A small group of top vendors contributes a large share of total revenue, indicating dependency risk.
- Single-Vendor Exposure: The leading vendor alone accounts for a significant portion of revenue, creating potential supply and negotiation risk.
- Long-Tail Vendors: A large number of vendors contribute minimal revenue, increasing operational overhead.
Insight: Revenue diversification is required while maintaining strong partnerships with high-performing vendors.

Profitability & Margin Analysis

- Wide Margin Variation: Significant differences in profit margins across vendors and categories indicate pricing inefficiencies.
- High-Margin Vendors: Certain vendors consistently maintain strong margins, reflecting pricing power and brand strength.
- Low-Margin Segments: Some vendors and categories operate with thin margins due to cost pressure or high competition.
Insight: Targeted pricing and cost optimization can improve overall profitability.

Sales Volume & Pricing Behavior

- Volume Concentration: A small set of vendors drives most unit sales.
- Price–Volume Relationship: Higher prices generally correlate with lower volumes, though strong brands defy this trend.
- Seasonality: Sales volumes show predictable seasonal patterns.
Insight: Pricing strategies should balance volume and margin while leveraging seasonal demand.

8. Business Recommendations

Strategic Recommendations

- Vendor Portfolio Optimization: Classify vendors into core, growth, and maintenance tiers to reduce risk and improve focus.
- Pricing & Margin Optimization: Adjust prices for low-margin products using elasticity insights to improve profitability.
- Inventory Alignment: Prioritize inventory for high-margin and high-velocity products to reduce carrying costs.

Operational Improvements

- Performance Monitoring: Use dashboards for regular vendor and category performance reviews.
- Automation: Enable automated reporting and refresh cycles to improve data timeliness.
- Governance: Establish monthly review cycles with defined KPIs and accountability.

Quick Wins

- Immediate price adjustments on select products
- Strengthening partnerships with top and emerging vendors
- Broader dashboard adoption among stakeholders

9. Conclusion

This project demonstrates an end-to-end data analytics lifecycle, transforming raw analytical exploration, and interactive visualization, the project identifies revenue concentration risks, margin optimization opportunities, and growth drivers.

Key outcomes include:

- Identification of vendor dependency and diversification opportunities
- Discovery of margin improvement potential across vendors and categories
- Clear growth and risk signals supported by data
- Actionable recommendations aligned with business objectives

The project successfully bridges technical analysis and business decision-making, making it suitable for professional portfolio presentation.