Project Overview

This document provides detailed documentation for the Vendor Performance Data Analytics End To End Project. The objective of this project is to analyze vendor level transactional data and convert it into meaningful business insights using SQL, Python, and data visualization techniques.

Problem Statement

Organizations often struggle to identify high performing vendors, optimize pricing strategies, and manage inventory efficiently due to fragmented data sources. This project addresses these challenges by building a unified analytical pipeline that evaluates vendor performance holistically.

Data Sources

Multiple datasets were used including purchase transactions, sales records, vendor invoices, pricing data, and freight cost information. These datasets were integrated using SQL joins and common table expressions.

Data Ingestion and Storage

Raw CSV files were ingested into an SQLite database using Python and SQLAlchemy. Each dataset was stored as a separate table to maintain data integrity and support scalable querying.

SQL Aggregation and Modeling

Advanced SQL queries were written using Common Table Expressions (CTEs) to compute vendor level summaries. Metrics such as total sales, total purchases, freight cost, excise tax, and profit margins were calculated.

Exploratory Data Analysis

Python libraries such as Pandas, Matplotlib, and Seaborn were used to explore data distributions, detect outliers, and analyze relationships between numerical variables.

Visualization and Insights

Various charts including bar charts, scatter plots, distribution plots, pie charts, and correlation heatmaps were created to visually communicate trends and insights effectively.

Statistical Analysis

Confidence interval analysis was performed to compare profit margins between top and low performing vendors, providing statistically supported conclusions.

Key Business Findings

The analysis revealed that a small group of vendors contributes a majority of sales revenue, profit margins vary significantly across brands, and freight cost plays a critical role in overall profitability.

Conclusion

This project demonstrates a complete data analytics workflow from raw data ingestion to business insight generation. The findings can directly support vendor optimization, pricing decisions, and inventory management.

Author

Satyam Gajjar
Portfolio: https://www.satyamgajjar.com
GitHub: https://github.com/satyamgajjar21