

Project Title: Retail Sales Performance Analysis System

1. Objective

The goal of this project is to build a small data analysis system that helps a retail company understand its sales performance — such as which products sell best, which regions generate the most revenue, and how sales trends change over time. This project combines Python, Pandas, NumPy, SQL, and Power BI to perform data cleaning, analysis, and visualization.

2. Tools & Technologies

Python (for data cleaning and automation)
Pandas (for data manipulation and analysis)
NumPy (for mathematical operations)
SQL (for storing and querying structured data)
Power BI (for creating dashboards and visual reports)
GitHub (for project storage and version control).

3. Problem Statement

Retail stores often deal with large amounts of sales data but struggle to extract meaningful insights like which products perform best or which regions generate the highest revenue. This project aims to solve these issues by creating a simple data-driven reporting system.

4. Proposed Solution

The system will:

1. Import sales data from a CSV file using Python.
2. Clean and transform data using Pandas and NumPy.
3. Store and query data using SQL (SQLite).
4. Visualize results using Power BI (without DAX measures).
5. Summarize insights in a simple report.

5. Key Features

- ✓ Load and clean raw sales data
- ✓ Calculate total revenue and profit per product, region, and month
- ✓ Generate charts (Python & Power BI)
- ✓ SQL-based data querying
- ✓ Export cleaned data and summaries
- ✓ Dashboard for visualization

6. Workflow / Methodology

1. Data Collection → Gather sales CSV file
2. Data Cleaning → Use Pandas & NumPy
3. Data Storage → Use SQL (SQLite)
4. Analysis → Query insights
5. Visualization → Power BI dashboards
6. Upload → GitHub repository

7. Expected Outcome

A clean dataset, Python scripts, SQL queries, and Power BI reports — all uploaded to GitHub as a complete project.

8. Learning Outcomes

Learn how to handle data end-to-end, strengthen Python, Pandas, NumPy, SQL, and Power BI concepts, and manage projects on GitHub.

9. Deliverables

Python scripts, SQL scripts, Power BI report, Documentation (README.md), GitHub Repository link.

10. Timeline

Week 1: Data collection & setup
Week 2: Python + Pandas cleaning
Week 3: SQL database & queries
Week 4: Power BI dashboard
Week 5: GitHub upload & documentation

11. Conclusion

The Retail Sales Performance Analysis System provides a complete and logical demonstration of all the tools and skills learned, showing a real-world data analysis workflow from start to finish.