Healthcare Claims Analysis - Executive Summary

This project demonstrates how data-driven decision-making can be powered by scalable healthcare claims analysis using Python, SQL, and data visualization tools. Using a synthetic dataset of 4,500+ healthcare claims, I built an end-to-end pipeline to clean, transform, and analyze claim records for trend discovery, cost optimization, and provider performance monitoring.

Tools & Technologies:

- Python (Pandas, Matplotlib, Seaborn), SQL
- Jupyter Notebook, GitHub, Visual Studio Code
- Modular scripts for ETL and EDA, SQL queries for KPIs

Objectives:

- Understand the distribution and cost impact of claim types
- Identify high-cost providers and seasonal claim spikes
- Automate report generation to replace manual Excel workflows

Key Findings:

- Outpatient claims represented ~43% of volume but only ~28% of cost
- Top 3 providers accounted for over 60% of total claim value
- Seasonal claim spikes observed in March and November
- Automated the pipeline, achieving 85% faster reporting turnaround

Impact:

This project simulates real-world analysis by a data analyst or engineer in healthcare. It highlights technical fluency, business insight extraction, and automation of critical reporting processes to

support operational efficiency.