Data-Driven Stock Analysis: Organizing, Cleaning, and Visualizing Market Tre

1. Project Overview

This project aims to analyze and visualize the performance of Nifty 50 stocks over the past year using a data-driven approach. The dataset, structured in YAML format, was parsed and transformed into clean, usable CSV files for further analysis. Tools such as Python, SQLite, Streamlit, and Power BI were used to derive insights.

2. Tools & Technologies

- Python (Pandas, YAML, Matplotlib, Seaborn)
- SQLite Database
- Streamlit (Interactive Web App)
- Power BI (Dashboard Visualizations)

3. Data Processing Workflow

1. Extract YAML files containing daily stock data
2. Organize data by Ticker and save individual CSVs
3. Clean data, handle missing values, convert types
4. Calculate: daily returns, volatility, cumulative returns
5. Store key metrics in SQLite and CSVs

4. Key Analysis & Metrics

- Top 10 Green & Red Stocks by Yearly Return
- Volatility (standard deviation of returns)
- Sector-wise performance
- Cumulative return chart for top stocks
- Monthly gainers/losers table
- Stock price correlation heatmap

5. Results & Insights

- ADANIENT, TATAMOTORS were highly volatile
- IT sector had highest average return
- INFY and TCS were highly correlated
- Consistent gainers identified using cumulative and monthly trends

6. Conclusion

The project successfully transformed raw YAML stock data into actionable insights using data science techniques. With dashboards in Power BI and Streamlit, it empowers investors and analysts to make better

decisions based on trend analysis, volatility, and sector performance.