README

Project Overview

This project demonstrates the creation of a real-time data streaming and analysis pipeline using Amazon Web Services. The primary objective was to collect, process, store, query, and visualize financial market data in near real-time.

Using the yfinance Python library, we collected historical five-minute interval stock data (high and low prices) from 10 major companies over a 9-day trading period. An AWS Lambda function transformed the data into JSON format and streamed it to Amazon Kinesis, which delivered it to Amazon S3 for storage.

The data was then cataloged with AWS Glue and made queryable using Amazon Athena, enabling interactive SQL-based analysis. We calculated daily average, maximum, and minimum volatility per company and exported the results to a CSV file.

For the final step, a Jupyter Notebook was used to generate two visualizations:

1. A line chart of maximum volatility trends per company.
2. A grouped bar chart of daily maximum volatility per company.

Together, these tools demonstrate the end-to-end process of real-time data ingestion, transformation, querying, and visualization on the cloud. This setup can be adapted for any similar scenario involving high-frequency, continuously updating datasets.