

Damion Brown

5/13/25

This project simulates a real-time financial data pipeline using AWS services and Python. The objective was to build a cloud-native workflow capable of collecting, processing, querying, and visualizing financial data in near real-time. Historical stock data for ten publicly traded companies was gathered using the yfinance Python library. The data interval covered April 14 to 25, 2025, with records collected at five-minute intervals, each including high and low prices along with computed volatility. A custom AWS Lambda function transformed these records into JSON format. The data was streamed to Amazon Kinesis and delivered to an S3 bucket via Kinesis Firehose. I used AWS Glue to crawl the data and create a schema that enabled querying through Amazon Athena.

Once the data was accessible in Athena, a query was executed to compute daily average, maximum, and minimum volatility for each company. The results were exported as a CSV file and visualized using Jupyter Notebook. A line chart was created to show maximum volatility trends across all companies. A grouped bar chart was also developed to compare daily highest volatility by company. These visualizations revealed that Tesla was the most volatile stock during the period. The analysis used several Python libraries, including pandas for data manipulation, matplotlib.pyplot for plotting, and seaborn for statistical visuals. The boto3 library was used to interface with AWS services programmatically. Combining these technologies and

libraries, I was able to build a scalable and automated pipeline for streaming, analyzing, and visualizing financial market data in near real-time