Project 3 is focuses on building a real-time data pipeline for finance data records using AWS Lambda. The pipeline involves data collection, analysis, and visualization to enable interactive querying and provide insights into the streaming data.

This project leverages AWS services, including Lambda, Kinesis, Glue, and Athena, to create a real-time data pipeline for finance data records. The Lambda function collects stock price data using the yfinance module and pushes it to the Kinesis stream. The data is then stored in S3 and made available for interactive querying and analysis using AWS Athena. Finally, visualizations are generated based on the query results to provide insights into the average volatility trend and daily highest volatility per company.

**screenshot\_of\_s3\_bucket.png**

A screenshot of a computer

Description automatically generated

Collects the data from yfinance and puts it into the firehose stream on S3 bucket

**kinesis\_monitor.jpeg**

A screenshot of a computer

Description automatically generated

AWS Kinesis configuration page

**exec\_results.jpeg**

A screenshot of a computer

Description automatically generated with medium confidence Data is collected via Lambda Function

**results.jpeg**

**A screenshot of a computer

Description automatically generated with medium confidence**

## The query is run through AWS Athena

## Data Visualization

1. Graph the average volatility trend per company (A single Line Chart: Each line refers to a company)

Which company is the most volatile?

A picture containing text, diagram, plot, line

Description automatically generated

As you can see from Average volatility trend per company, company COST is the most volatile company, and the average volatility number is 0.683444.

1. Graph the daily highest volatility per company (A Grouped Bar Chart: Each group refers to a company and the bars refer to the daily highest volatility)

Do the findings from this graph support your conclusion from the first graph?

A picture containing screenshot, text, plot, diagram

Description automatically generated

Company COST exhibited the highest daily volatility, reaching approximately 5 on April 6, 2023. The second-highest daily volatility was observed on April 11, 2023, with a value of around 4.

TGT had a daily highest volatility of approximately 2.5, while WMT recorded a daily highest volatility of approximately 2. Both TGT and WMT experienced their highest volatilizes on the same day as COST.

Therefore, these findings in the second chart are consistent with the conclusion reached in the first chart, which identifies COST companies as the most volatile companies based on the highest average volatility.