1. PROJECT DESCRIPTION

This project documents the execution and validation of an automated ETL and visualization pipeline designed to process IBM stock data sourced from Yahoo Finance.

The pipeline consists of two independent automated scripts. The first script performs data extraction from Yahoo Finance, cleans and standardizes the dataset, transforms the date index into a structured date column, and loads the transformed data into a PostgreSQL Database which serves as a central point of truth. The second script retrieves the updated stock data from the database and generates a scatter plot, which is saved to a dedicated directory that retains all previously generated visualizations.

A controlled 10-second delay is implemented between the execution of the two scripts to ensure data availability before visualization.

---

2. AUTOMATION CONFIGURATION

- Execution frequency: Hourly

- Total execution duration: 12 hours

- Start time: 14:00 (2 PM)

- End time: 02:00 (2 AM)

- Execution method: Automated task scheduling

- Script dependency: Sequential execution with buffer delay

---

3. TEST OBJECTIVE

The objective of this test was to validate the reliability, stability, and consistency of the automated pipeline over a continuous 12-hour execution window, while applying controlled changes to the data extraction start date to confirm correct handling of varying historical data ranges.

---

4. TEST EXECUTION RECORD

5. 14:00 – Execution 1
Result: Successful execution
Output: Scatter plot generated with correct date timestamps

6. 15:00 – Execution 2
   Result: Successful execution
   Output: Scatter plot generated with correct date timestamps

7. 16:00 – Execution 3
   Result: Successful execution
   Output: Scatter plot generated with correct date timestamps
   Change applied: Extraction start date changed from 2025-12-20 to 2024-12-20

8. 17:00 – Execution 4
   Result: Successful execution
   Output: Scatter plot generated with correct date timestamps
   Change applied: Extraction start date changed from 2024-12-20 to 2025-01-01

9. 18:00 – Execution 5
   Result: Successful execution
   Output: Scatter plot generated with correct date timestamps
   Change applied: Extraction start date changed from 2025-01-01 to 2023-01-01

10. 19:00 – Execution 6
    Result: Successful execution
    Output: Scatter plot generated with correct date timestamps
    Change applied: Extraction start date changed from 2023-01-01 to 2025-06-01

11. 20:00 – Execution 7
    Result: Successful execution
    Output: Scatter plot generated with correct date timestamps
    Change applied: Extraction start date changed from 2025-06-01 to 2025-08-01

12. 21:00 – Execution 8
    Result: Successful execution
    Output: Scatter plot generated with correct date timestamps
    Change applied: Extraction start date changed from 2025-08-01 to 2025-09-01

13. 22:00 – Execution 9
    Result: Successful execution
    Output: Scatter plot generated with correct date timestamps

14. 23:00 – Execution 10
    Result: Successful execution
    Output: Scatter plot generated with correct date timestamps

15. 00:00 – Execution 11
    Result: Successful execution
    Output: Scatter plot generated with correct date timestamps
    Change applied: Extraction start date changed from 2025-09-01 to 2025-10-01

16. 01:00 – Execution 12
    Result: Successful execution
    Output: Scatter plot generated with correct date timestamps
    Change applied: Extraction start date changed from 2025-10-01 to 2025-11-01

Final execution completed at 02:00.

    Result: Successful execution
    Output: Scatter plot generated with correct date timestamps

---

5. TEST RESULTS SUMMARY

- Total executions: 13

- Successful executions: 13

- Failed executions: 0

- Visual output generation: 100% successful

- Data consistency: Maintained across all date range modifications

- Automation stability: Confirmed over continuous operation

---

6. CONCLUSION

The automated ETL and visualization pipeline operated successfully for the full 12-hour test window without interruption or failure.

The system consistently handled dynamic changes in data extraction parameters while maintaining reliable data loading and visualization generation. The results confirm the pipeline's operational stability and suitability for continuous, scheduled execution in a production-style environment.