Building an end to end Data Analytics using Snowflake, Airflow, dbt, and a BI tool

Ameya Satish Khond
Department of Applied Data Intelligence
San Jose State University
ameyasatish.khond@sjsu.edu

Prathamesh Mankar
Department of Applied Data Intelligence
San Jose State University
prathmesh.mankar@sjsu.edu

Abstract

The stock analytics pipeline established in Lab 1 is further developed in this lab through the adoption of dbt data transformations, their orchestration by Apache Airflow, and the leveraging of the Preset BI platform for the visualization of analytical insights. Among the calculated technical indicators are Moving Averages (SMA20, SMA50), Relative Strength Index (RSI-14), daily percentage returns, and trading volume; these are all performed in dbt models which are applied to two stocks (AAPL and TSCO.L). With Airflow in place, the execution of the dbt model is automated thus making the transformations reproducible. A comprehensive BI dashboard is set up to deliver user-friendly insights into four aspects of stocks: trend, momentum, volatility, and market activity. The ELT process, system architecture, transformation logic, Airflow orchestration, and BI dashboard outcomes are all discussed in this report.

Keywords

Snowflake, dbt, Airflow, Preset, RSI, Moving Averages, Data Engineering

I. Introduction

Modern financial analytics demand the automation of data ingestion, transformation, and visualization.

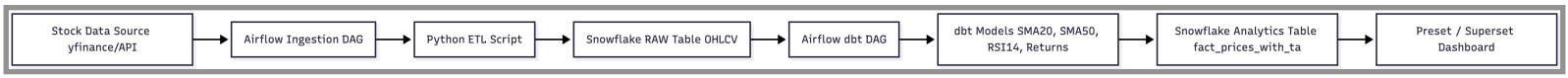
Lab 1 subjected to pipeline ingestion and forecasting, whereas Lab 2 is doing the following:

- The transformations are done with dbt, such as SMA20, SMA50, RSI-14, and daily returns
- Airflow is being used for the scheduling of dbt runs
- A BI dashboard development in Preset for the purpose of visualizing technical indicators.

It similarly to Lab 1, the pipeline with the same engineering style is built complete ELT pipeline (the pipeline diagrams on the pages 1-2 will help you).

The findings lead to the determination of stock momentum, volatility, and market structure that are ready for implementation.

II. System Architecture



- **Components:** The architecture consists of **eight major components**, forming an end-to-end ELT pipeline. Each component's purpose is described clearly and concisely.
    - **Stock data source** (yfinance API) used to extract daily OHLCV values for selected tickers.
    - An **Airflow ingestion DAG** that schedules and triggers a Python ETL script to download and load raw price data into Snowflake.
    - A **Snowflake RAW table** that stores unmodified OHLCV data for both AAPL and TSCO.L
    - An **Airflow dbt DAG** that runs dbt models to compute technical indicators including SMA20, SMA50, RSI-14, and daily returns.
    - **dbt transformation models** that apply SQL window functions to generate technical analysis features.
    - A **Snowflake analytics table (**fact_prices_with_ta**)** that consolidates all technical indicators into a single, curated dataset.
    - A **Preset/Superset dashboard** that visualizes key metrics such as stock prices, moving averages, RSI, daily returns, and volume.

III. Data Model

We define Two tables within database PRATHMESH_1:

- **ANALYTICS.FACT_PRICES_WITH_TA**:
    - **Columns:** DT, OPEN, HIGH, LOW, CLOSE, SMA_20, VOLUME, SYMBOL, SMA_50, RSI_14, DAILY_RETURN

- **SNAPSHOT.FACT_PRICES_WITH_TA**:
    - **Columns:** DT, OPEN, HIGH, LOW, CLOSE, SMA_20, VOLUME, SYMBOL, SMA_50, RSI_14, DAILY_RETURN, DBT_UPDATED_AT, DBT_VALID_FROM, DBT_VALID_TO

This design enforces uniqueness, supports idempotent loads, and clearly separates observed versus forecasted values while allowing unified consumption.

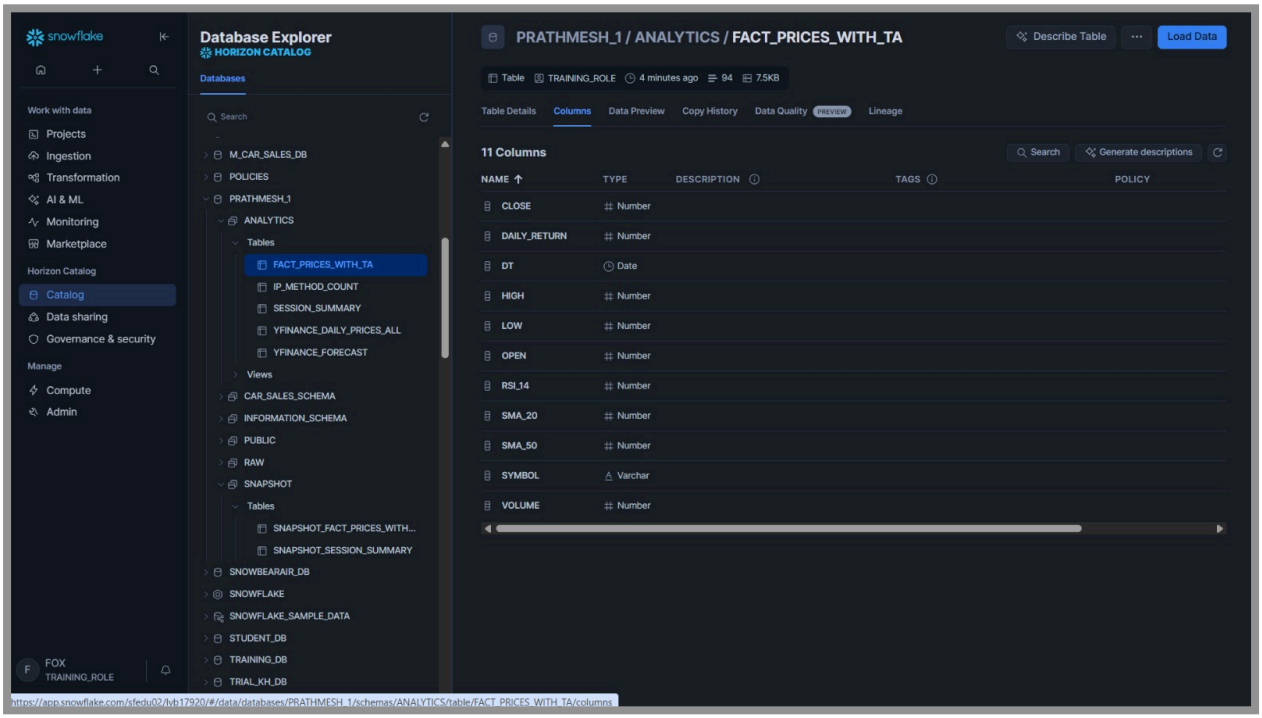


Fig. FACT_PRICES_WITH_TA Table (Fields, Attributes)

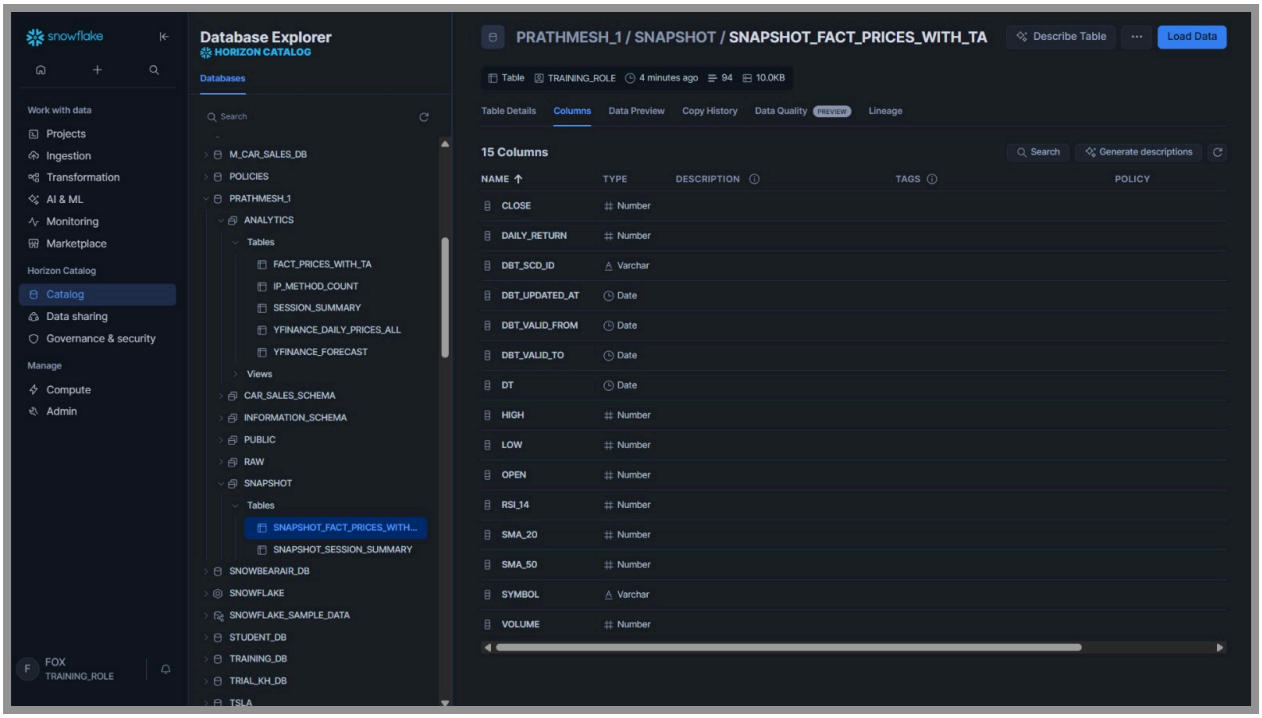


Fig. SNAPSHOT.FACT_PRICES_WITH_TA Table (Fields, Attributes)

IV. Methods

A. Data Preparation and Loading:

- Raw stock price data for AAPL and TSCO.L is obtained from the yfinance API through a Python ETL script scheduled in Airflow.
- The script standardizes field names, enforces datatypes, and ensures that each record contains a consistent OHLCV structure.
- The processed dataset is loaded into the RAW schema of Snowflake, which acts as the initial storage layer for all downstream transformations.

B. Analytical Transformations Using dbt:

The primary objective of Lab 2 is to compute technical indicators required for financial analysis. dbt (data build tool) is used to implement SQL-based transformations within Snowflake.

The dbt models create a refined analytical table that contains:

- Simple Moving Averages (SMA20, SMA50)
- Relative Strength Index (RSI-14)
- Daily percentage returns
- Normalized OHLCV features

C. Visualization with Preset (Superset):

The final analytical dataset is visualized using Preset, a cloud-hosted version of Apache Superset. A dashboard is built to display key financial metrics through interactive charts, enabling comparison between AAPL and TSCO.L.

The dashboard includes:

- RSI chart to observe momentum
- Daily returns chart to assess volatility
- Price + SMA charts for trend analysis
- Volume chart for trading activity
- Cross-stock price comparison chart

V. Orchestration with Airflow and DBT

- In this part, the usage of Apache Airflow for the automation of the execution of dbt models that produce analytical indicators like SMA20, SMA50, RSI-14, and daily returns is illustrated. The orchestration takes care of the transformations to run in a manner that is reliable, consistent, and no human intervention is required.

A. Airflow DAG for dbt Execution:

- A dedicated Airflow DAG is created to automate dbt transformations on a scheduled basis. The DAG runs the following tasks sequentially:

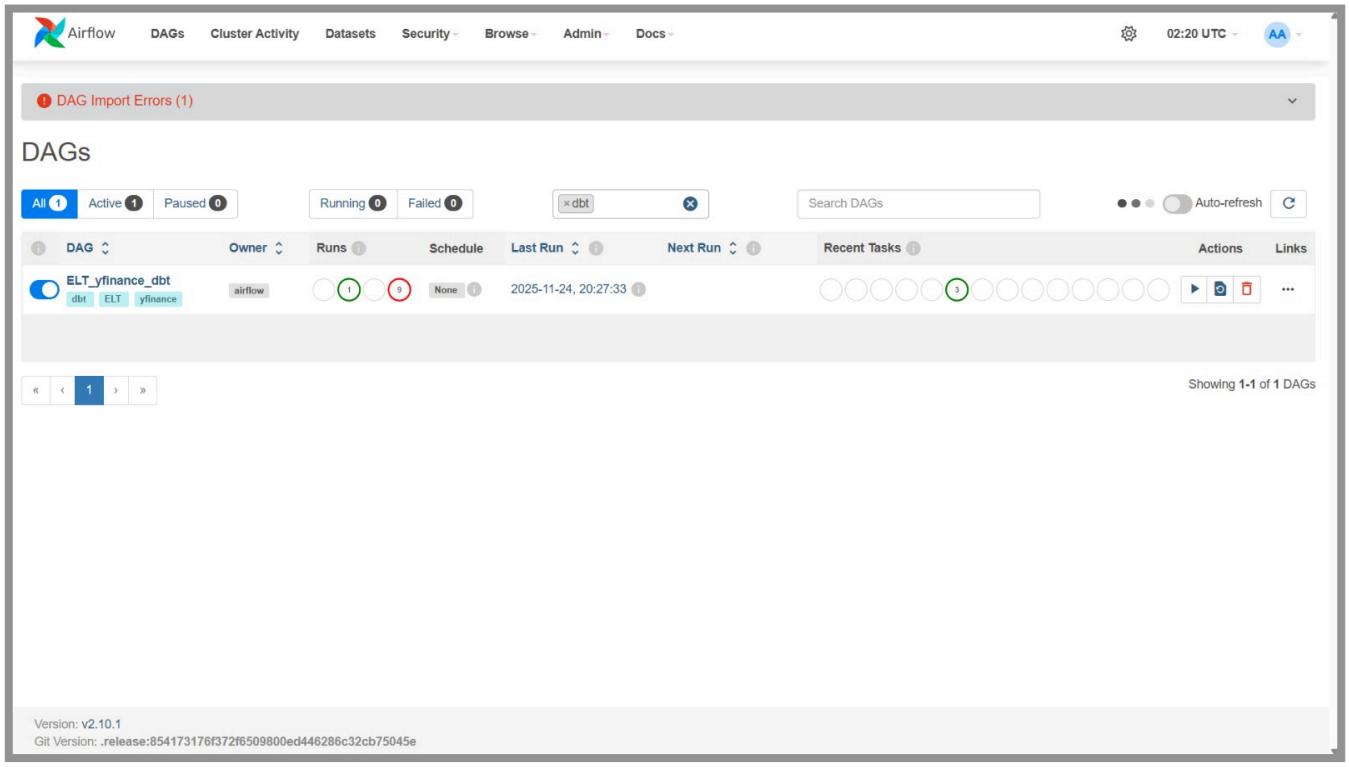


Fig. ELT DAG

1. **dbt run** to build models

```
09:23:47  All checks passed!
PS D:\SJSU\DATA-226\Airflow\yfinance_elt> dbt run
09:23:59  Running with dbt=1.8.5
09:24:00  Registered adapter: snowflake=1.8.2
09:24:00  Unable to do partial parsing because of a version mismatch
09:24:03  [WARNING]: Deprecated functionality
The `tests` config has been renamed to `data_tests`. Please see
https://docs.getdbt.com/docs/build/data-tests#new-data_tests-syntax for more
information.
09:24:04  Found 2 models, 1 snapshot, 2 data tests, 1 source, 459 macros
09:24:04
09:24:08  Concurrency: 1 threads (target='dev')
09:24:08
09:24:08  1 of 1 START sql table model analytics.fact_prices_with_ta ..................... [RUN]
09:24:10  1 of 1 OK created sql table model analytics.fact_prices_with_ta ................ [SUCCESS 1 in 2.41s]
09:24:10
09:24:10  Finished running 1 table model in 0 hours 0 minutes and 6.15 seconds (6.15s).
09:24:10
09:24:10  Completed successfully
09:24:10
09:24:10  Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
```

Fig. dbt run

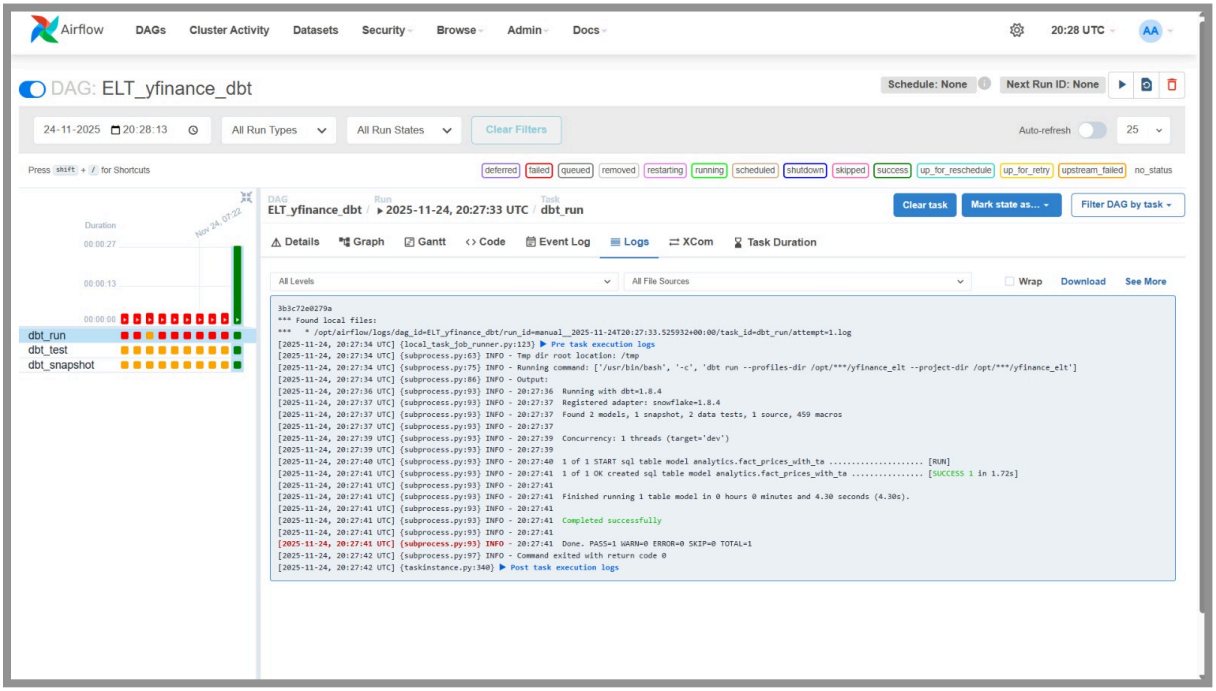


Fig. Dbt run logs

2. **dbt test** to validate data quality

```
09:24:10
09:24:10  Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
PS D:\SJSU\DATA-226\Airflow\yfinance_elt> dbt test
09:24:49  Running with dbt=1.8.5
09:24:50  Registered adapter: snowflake=1.8.2
09:24:51  Found 2 models, 1 snapshot, 2 data tests, 1 source, 459 macros
09:24:51
09:24:53  Concurrency: 1 threads (target='dev')
09:24:53
09:24:53  1 of 2 START test not_null_fact_prices_with_ta_dt ..... [RUN]
09:24:54  1 of 2 PASS not_null_fact_prices_with_ta_dt ..... [PASS in 1.34s]
09:24:54  2 of 2 START test not_null_fact_prices_with_ta_symbol ..... [RUN]
09:24:55  2 of 2 PASS not_null_fact_prices_with_ta_symbol ..... [PASS in 1.07s]
09:24:55
09:24:55  Finished running 2 data tests in 0 hours 0 minutes and 4.78 seconds (4.78s).
09:24:55
09:24:55  Completed successfully
09:24:55
09:24:55  Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
PS D:\SJSU\DATA-226\Airflow\yfinance_elt>
```

Fig. dbt test

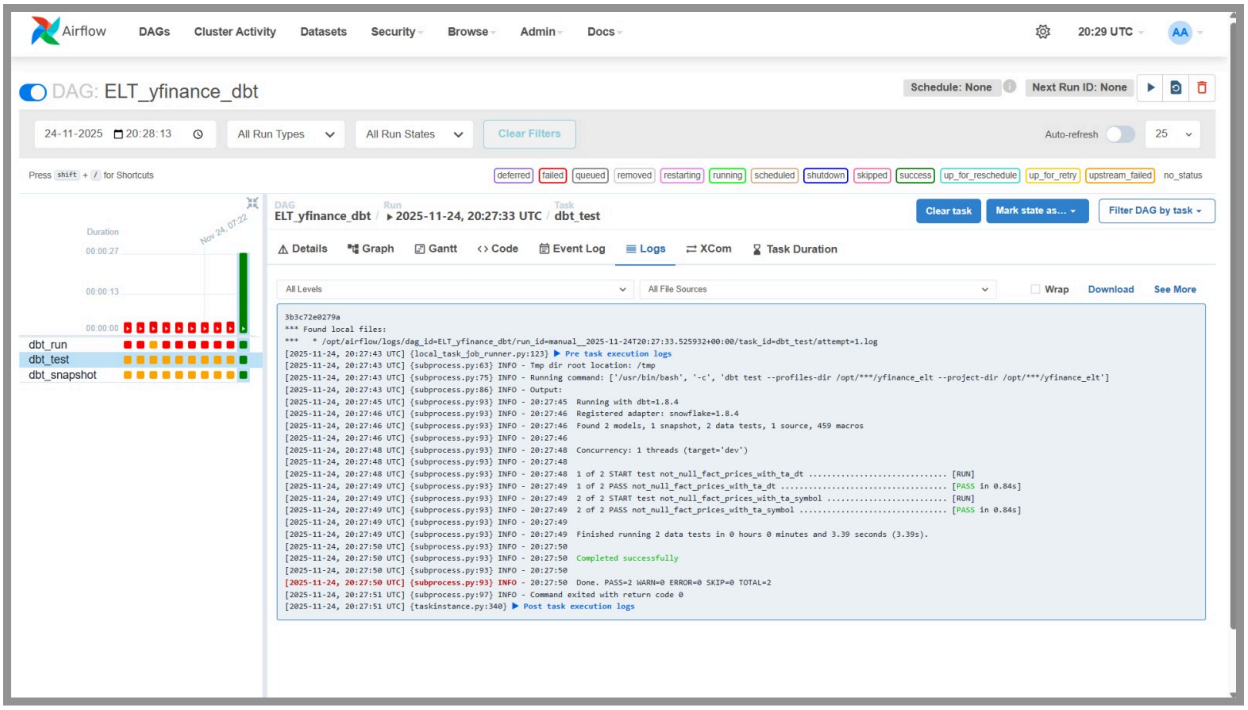


Fig. dbt test logs

3. dbt snapshot

The DAG ensures that new market data in Snowflake RAW tables is processed into analytical models daily.

```
09:24:55
09:24:55  Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
PS D:\SJSU\DATA-226\Airflow\yfinance_elt> dbt snapshot
09:25:27  Running with dbt=1.8.5
09:25:29  Registered adapter: snowflake=1.8.2
09:25:29  Found 2 models, 1 snapshot, 2 data tests, 1 source, 459 macros
09:25:29
09:25:33  Concurrency: 1 threads (target='dev')
09:25:33
09:25:33  1 of 1 START snapshot SNAPSHOT.snapshot_fact_prices_with_ta ..................... [RUN]
09:25:38  1 of 1 OK snapshotted SNAPSHOT.snapshot_fact_prices_with_ta ..................... [SUCCESS 0 in 4.95s]
09:25:38
09:25:38  Finished running 1 snapshot in 0 hours 0 minutes and 8.80 seconds (8.80s).
09:25:38
09:25:38  Completed successfully
09:25:38
09:25:38  Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
PS D:\SJSU\DATA-226\Airflow\yfinance_elt>
```

Fig. dbt snapshots

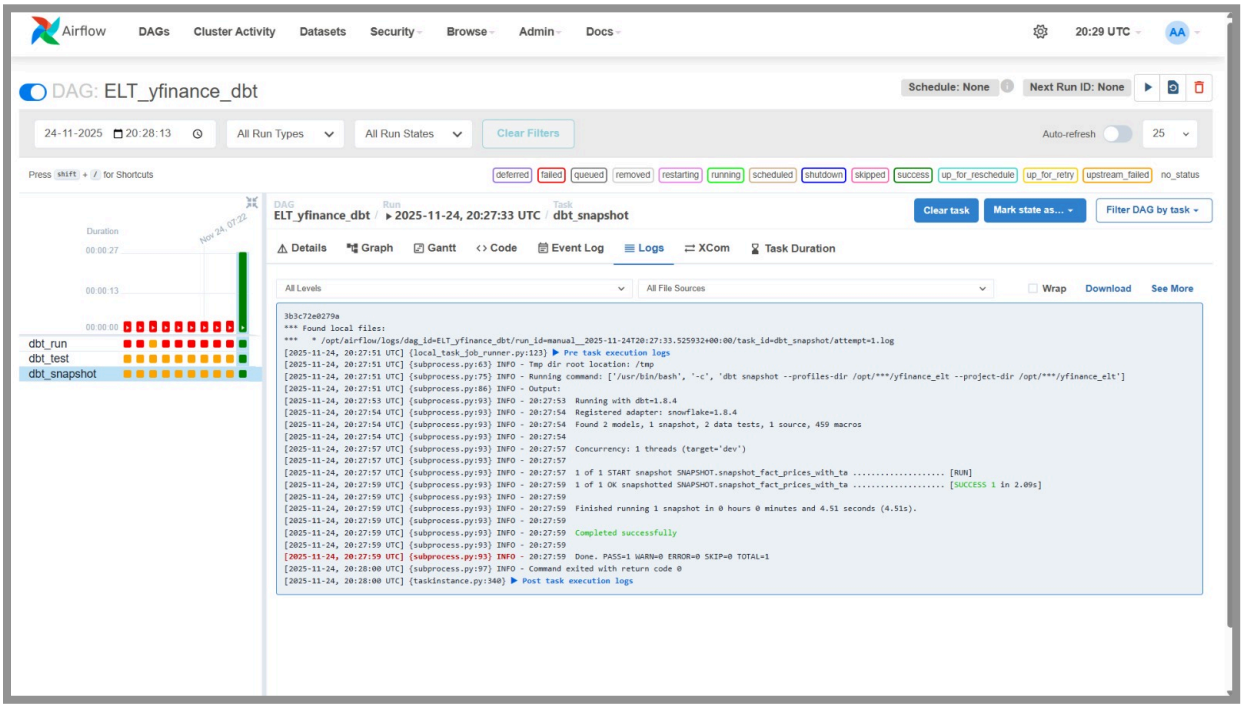


Fig. dbt snapshots logs

B. Airflow Connections and Variables:

Airflow uses:

- **Snowflake Connection** – credentials and warehouse settings
- **Variables** – symbols, database, schema.
- **Schedule** – CRON expression for dbt DAG

These parameters allow the pipeline to be modified without changing code.

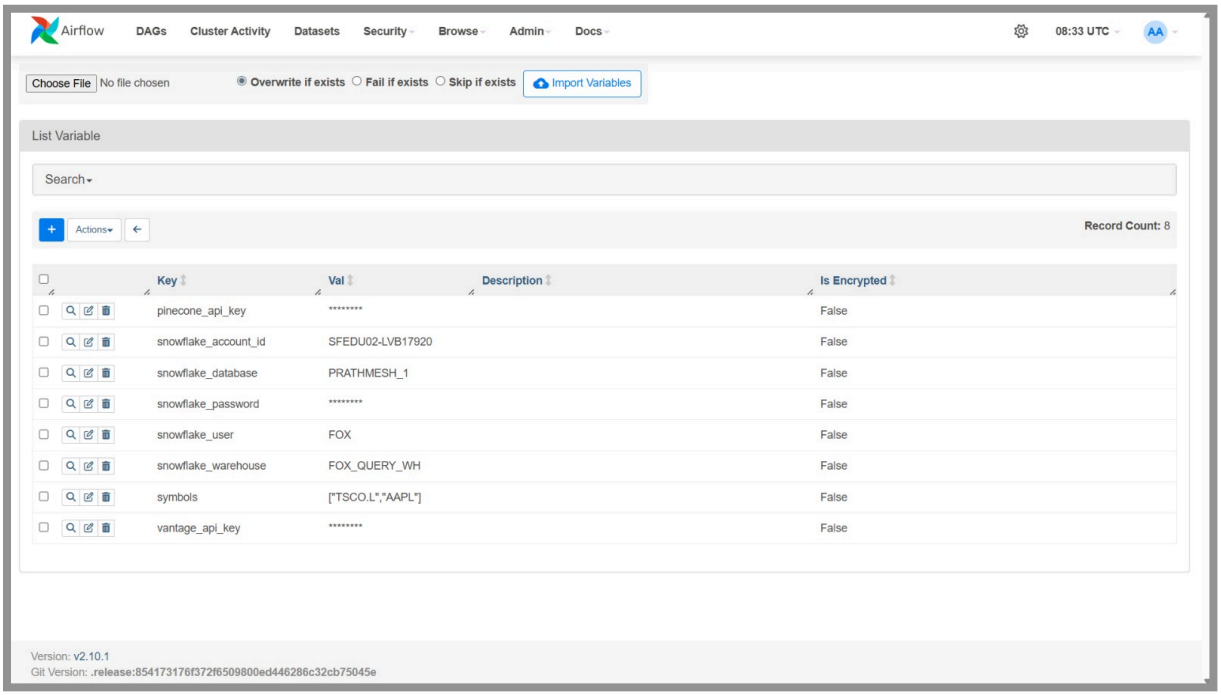


Fig. Airflows Variables

C. Integration Between Airflow and dbt:

Airflow triggers dbt commands using either:

- **dbt CLI execution** inside a BashOperator

The dbt execution produces the final transformed table:

ANALYTICS.FACT_PRICES_WITH_TA

This table is then consumed by Preset for visualization.

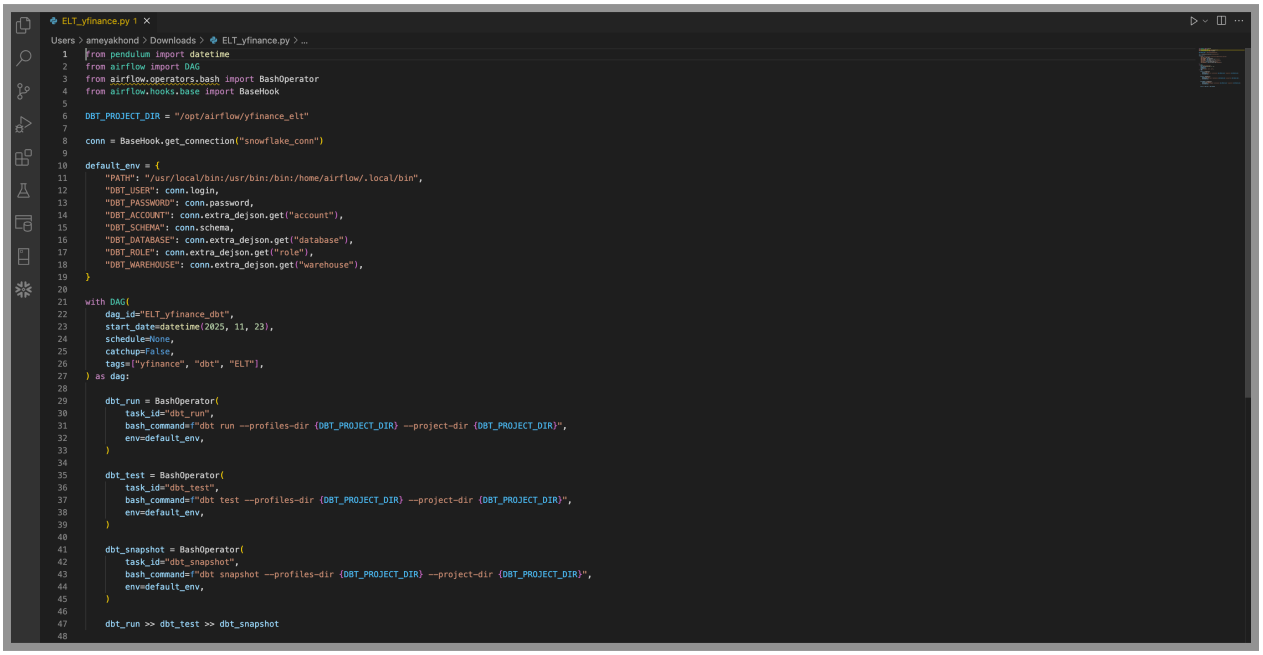


Fig. Dbt Bash Operator code

VI. BI VISUALIZATION RESULTS (PRESET)

Preset was used to create a dashboard titled **Stock Analysis Dashboard**, containing the following charts:

A. Relative Strength Index (RSI):

- Compare momentum for AAPL and TSCO.L

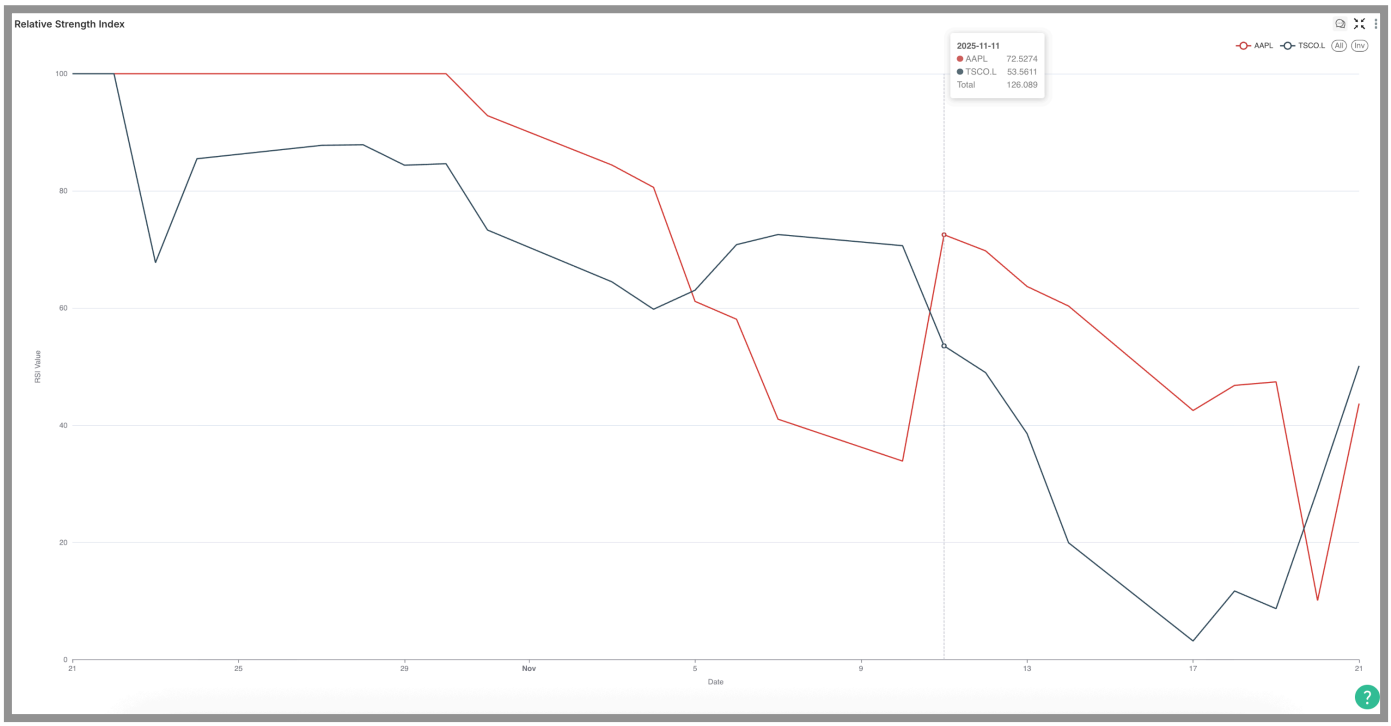


Fig. RSI Visualization

B. Daily Percentage Returns:

- Bar chart showing volatility and day-to-day fluctuation.

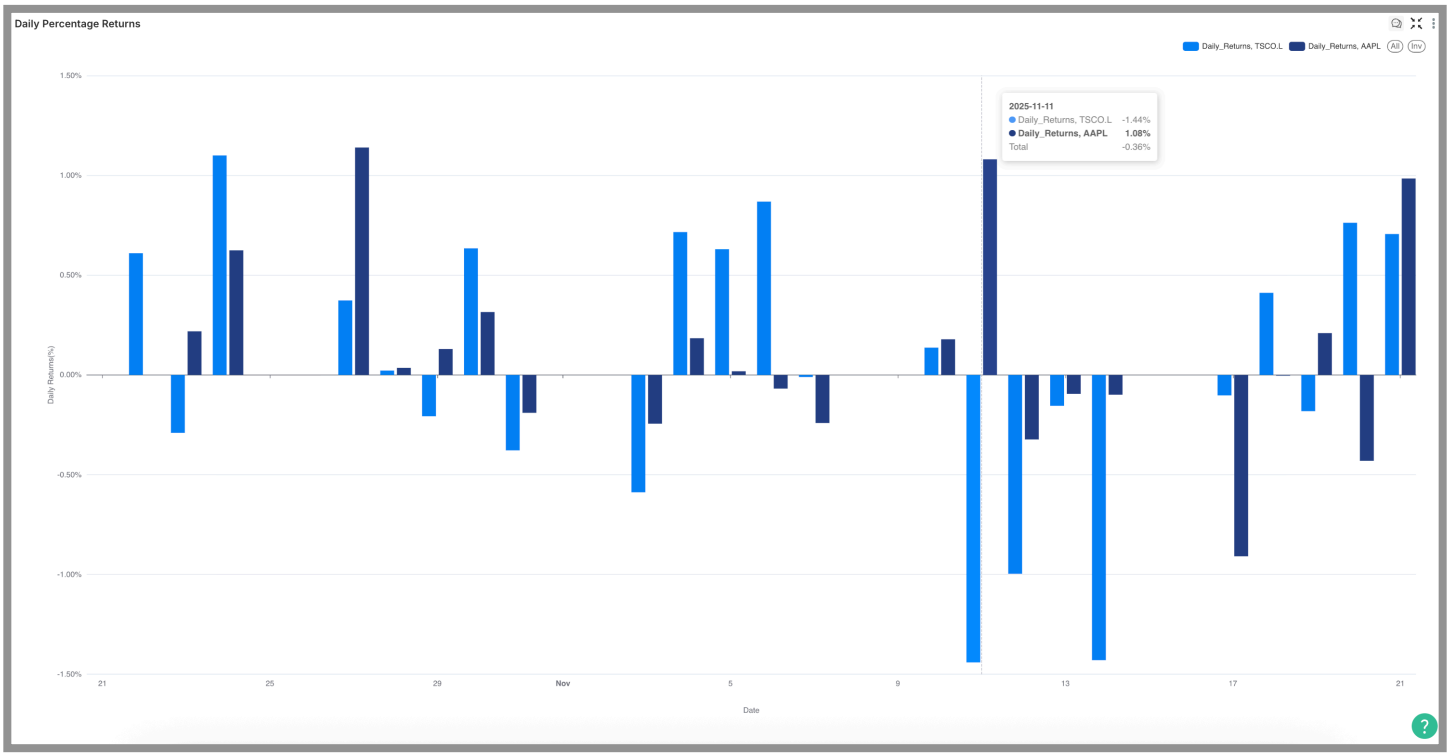


Fig. Daily Percentage Return Visualization

C. AAPL Price with SMA20 & SMA50:

- Shows trend-following indicators and potential crossover signals.

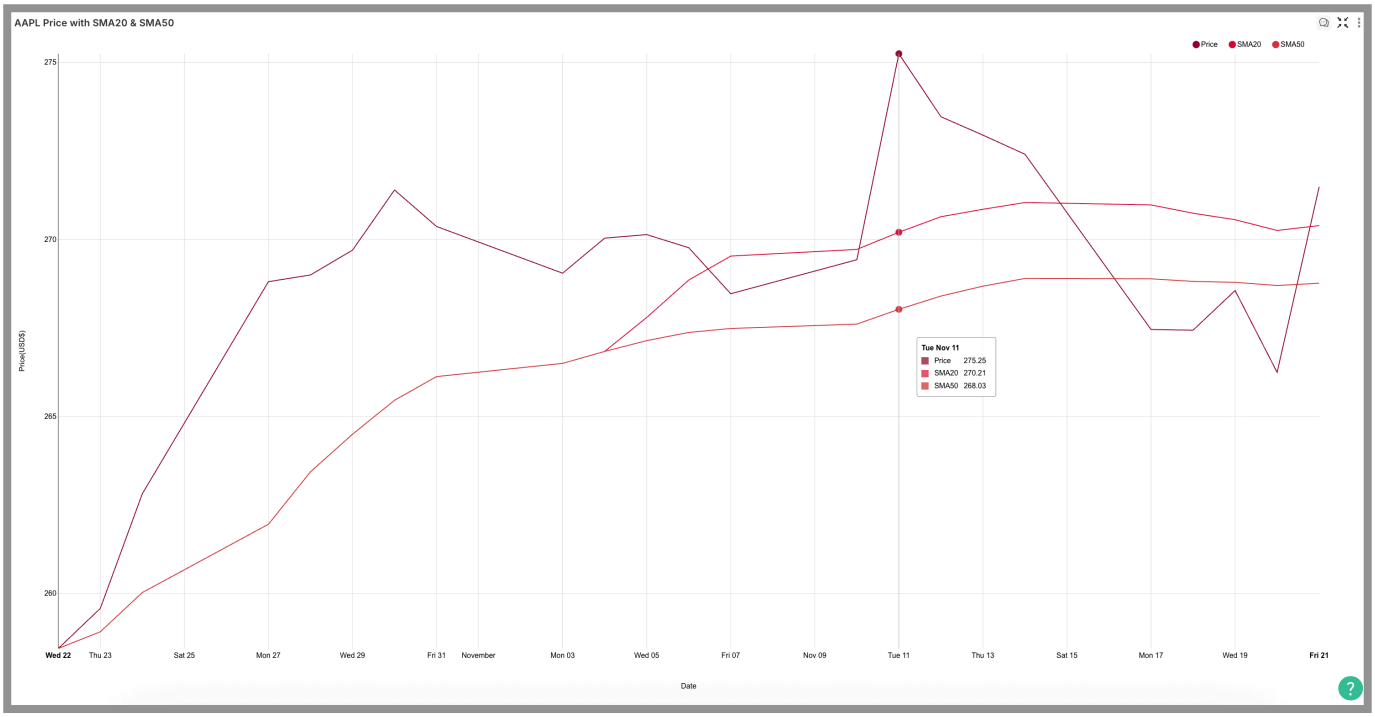


Fig. Apple Stock SMA20 and SMA50 Visualization

D. TSCO.L Price with SMA20 & SMA50:

- Same analysis for second stock.

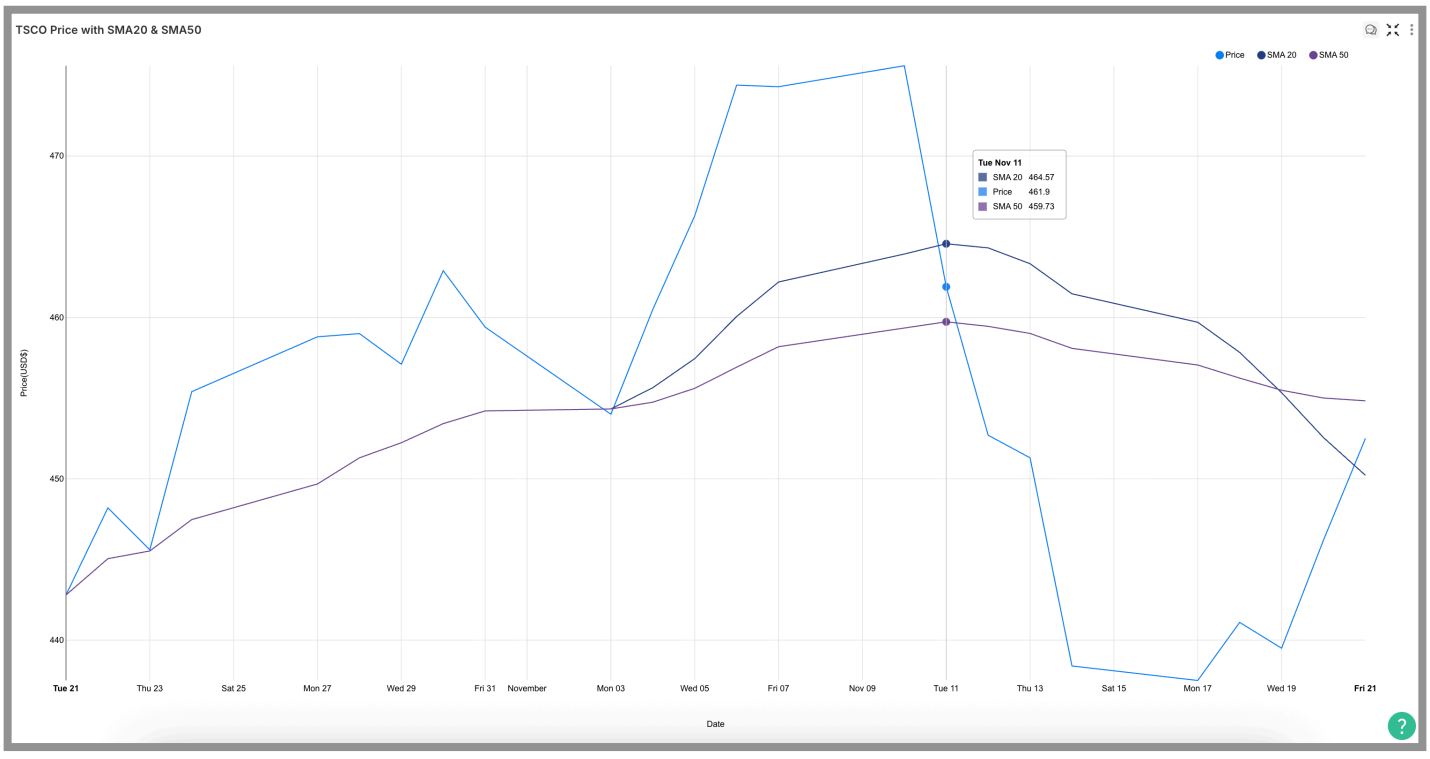


Fig. Tesco Stock SMA20 and SMA50 Visualization

E. Stock Price Comparison:

- Line chart comparing raw prices of AAPL vs TSCO.L.

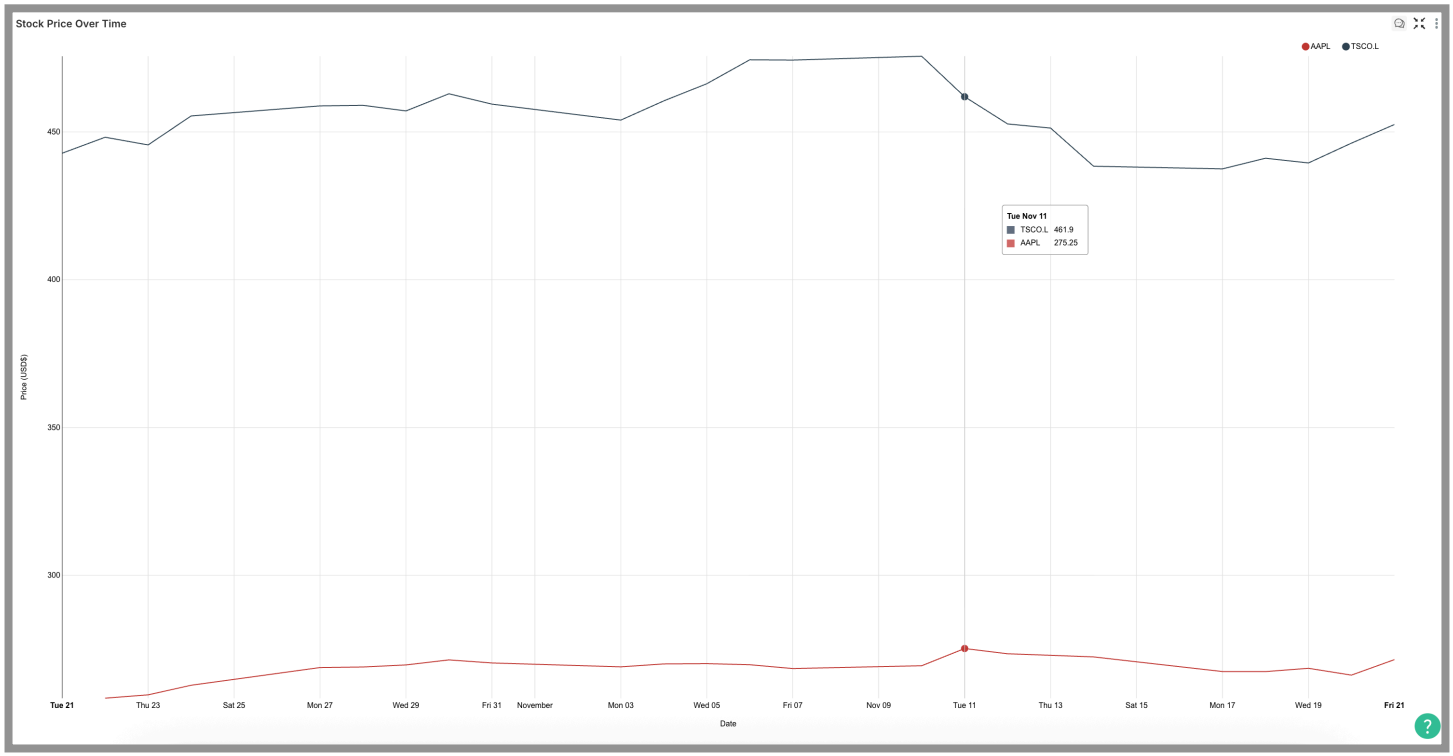


Fig. Stock Comparison Visualization

F. Trading Volume Over Time:

- Bar chart showing comparative trading activity for both stocks.

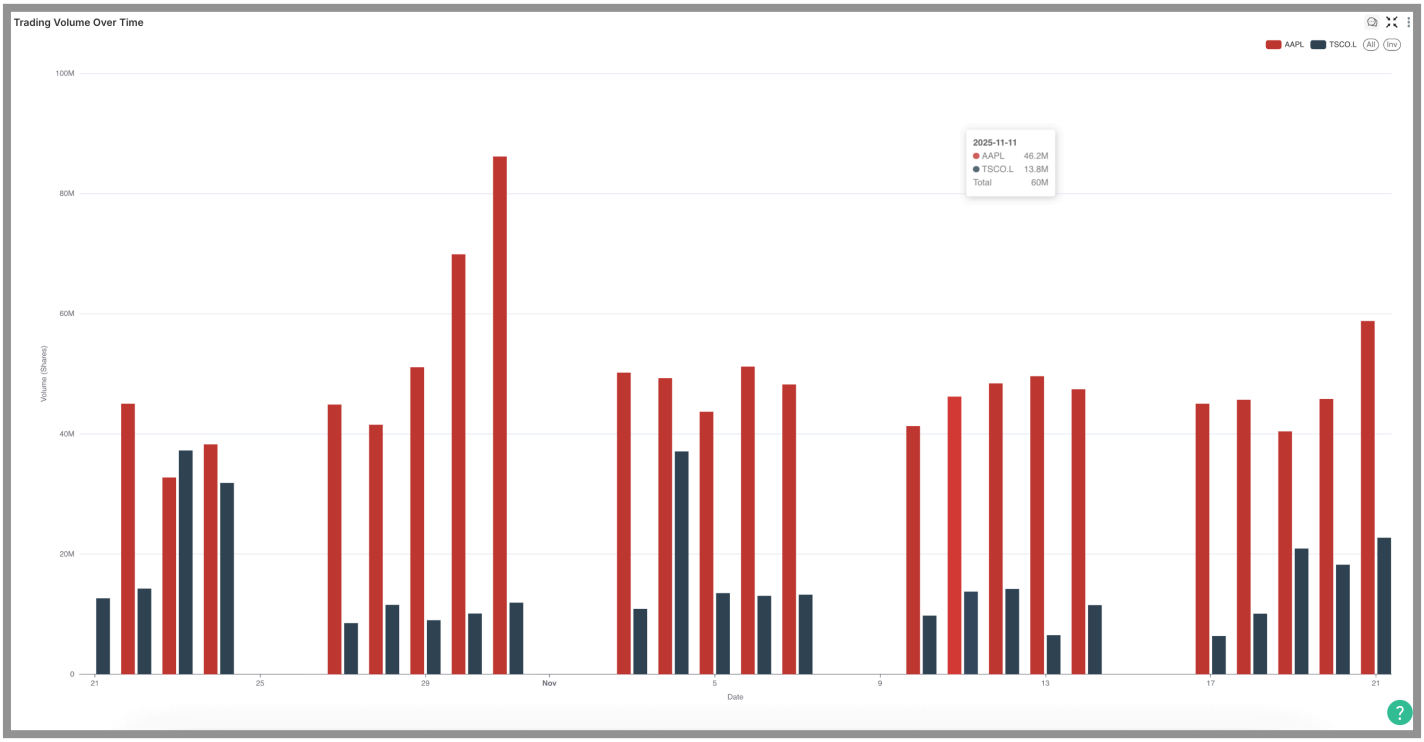


Fig. Trading Volume Visualization

G. Dashboard Filters

Filters allow users to adjust:

- Stock symbol(s)

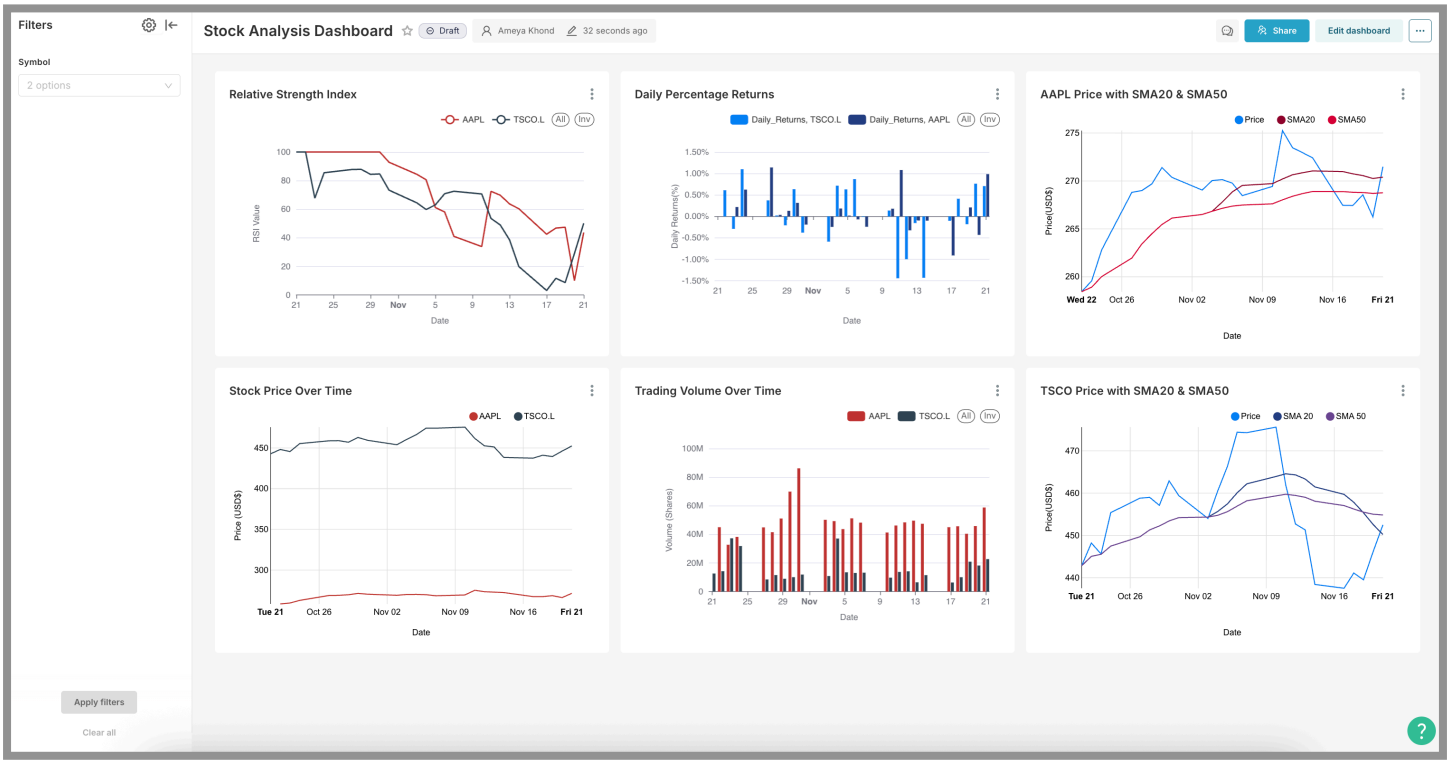


Fig. Stock Analysis Dashboard

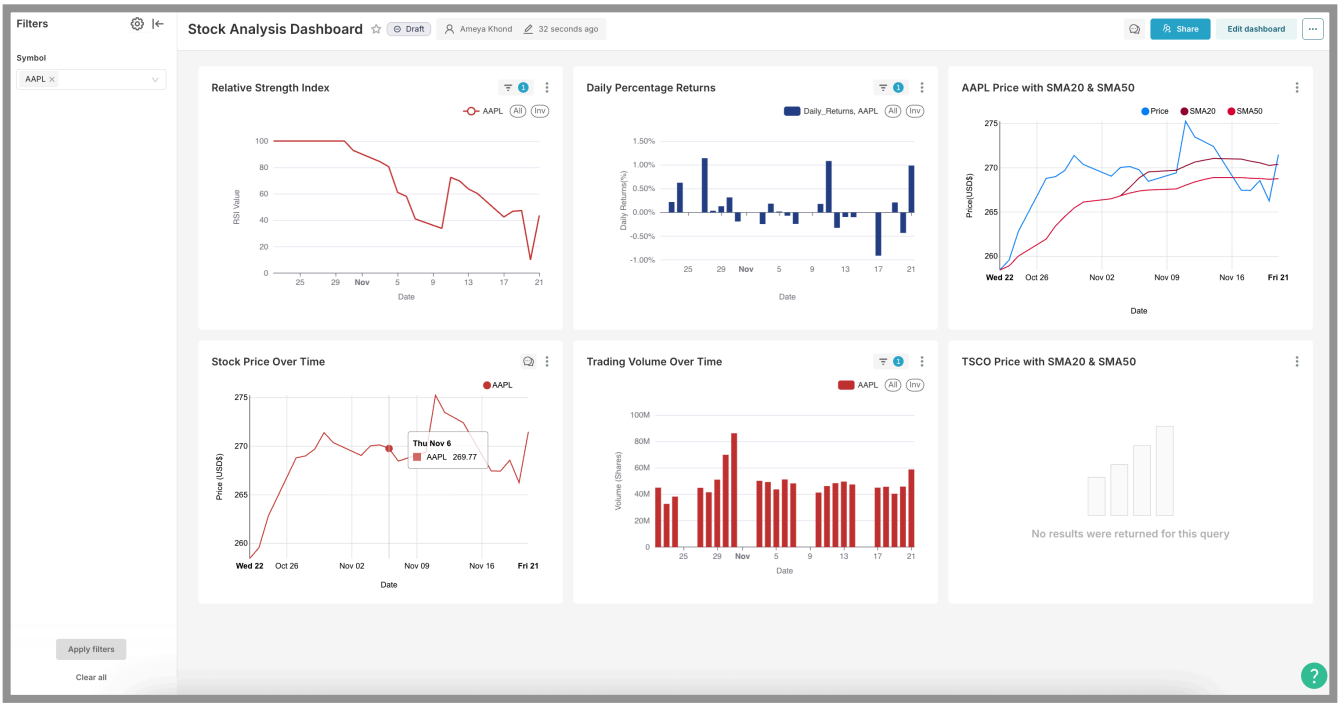


Fig. Dashboard using Apple Stock as Filter

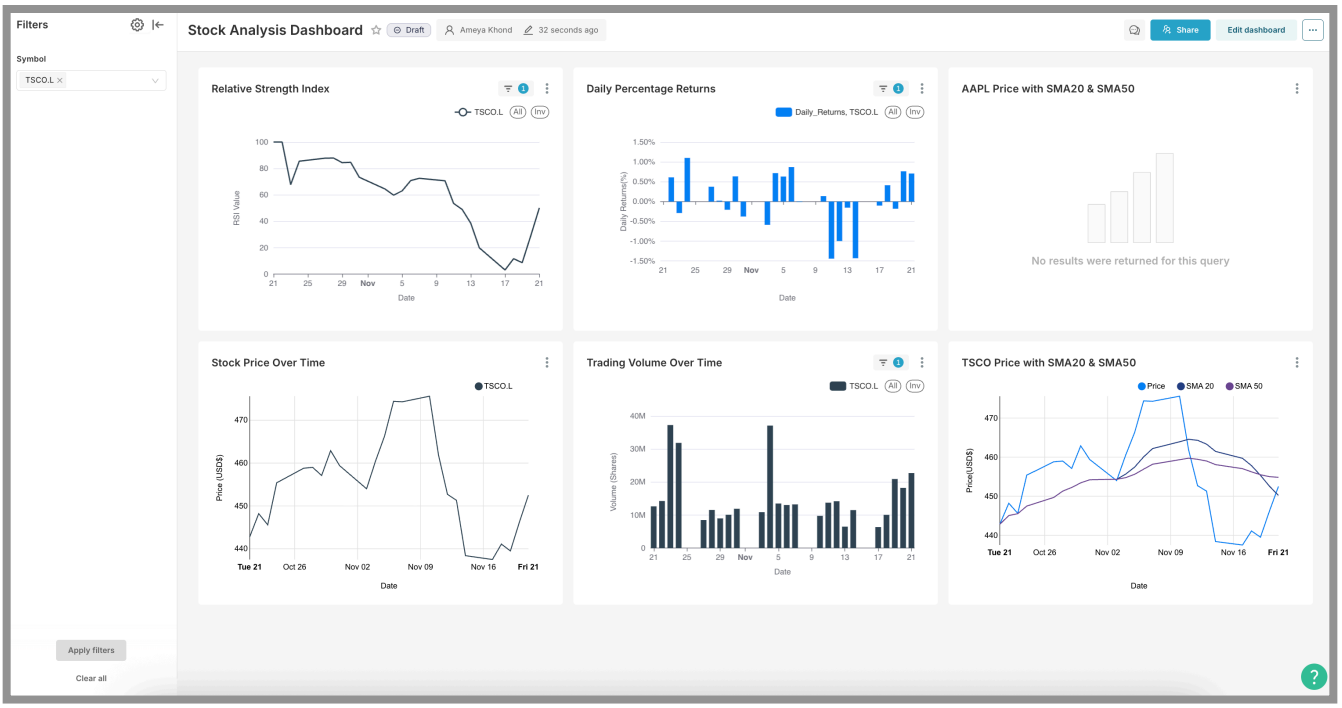


Fig. Dashboard using Tesco Stock as Filter

VIII. Discussion

The dashboard reveals clear insights:

- **AAPL** shows more stable price action compared to **TSCO.L**
- **SMA20** and **SMA50** crossovers reveal trend shifts
- **RSI** shows momentum differences and potential price reversals
- **Volume** patterns correlate with price volatility
- **Daily returns** highlight risk and fluctuation behavior

The ELT pipeline created is scalable and reusable for other financial datasets.

IX. Conclusion

This lab successfully demonstrated a complete ELT + BI workflow using modern data engineering tools. Raw data was transformed into actionable metrics using dbt, scheduled in Airflow, and visualized interactively in Preset.

The result is a functional analytics environment capable of technical stock analysis and can be extended with machine learning, forecasting, or automated alerts in future work.

Github Link:

https://github.com/prathmesh0550/LAB_2