

# **AI-Driven Stock Market Trading Strategy Generation**

A Paisa Vasool Team Presentation

- Divya Vemula
- Mani Krishna Tippani
- Rahul Chauhan
- Shaheryar Nadeem

# OVERVIEW

- PROJECT BACKGROUND & PURPOSE
- BUSINESS PROBLEM
- PROPOSED SOLUTION
- TOOLS, FRAMEWORKS & PLATFORMS
- DATA SOURCE
- TECHNICAL ARCHITECTURE
- DATA ENGINEERING & ETL
- DATA QUALITY
- STOCK INDICATORS BRIEF
- VISUALIZATIONS
- ML MODELLING
- GEN-AI
- RESPONSE GENERATION



# PROJECT BACKGROUND

- Financial markets generate **massive volumes of time-series data daily**, offering valuable opportunities for predictive modeling across investors, analysts, and businesses.
- **Raw stock market data - lacks the structure, enrichment, and quality required for meaningful insights and reliable predictions.**

## Purpose:

To unlock its full potential, market data must undergo systematic extraction, cleaning, feature engineering, and efficient storage before it becomes usable for machine learning applications.



# BUSINESS PROBLEM



## Traditional Models

Inaccurate predictions due to limited ability  
Complex financial time-series data.



## High volatility and non-linearity:

Difficulty in predicting market downturns and unexpected spikes.  
Incomplete or noisy data.



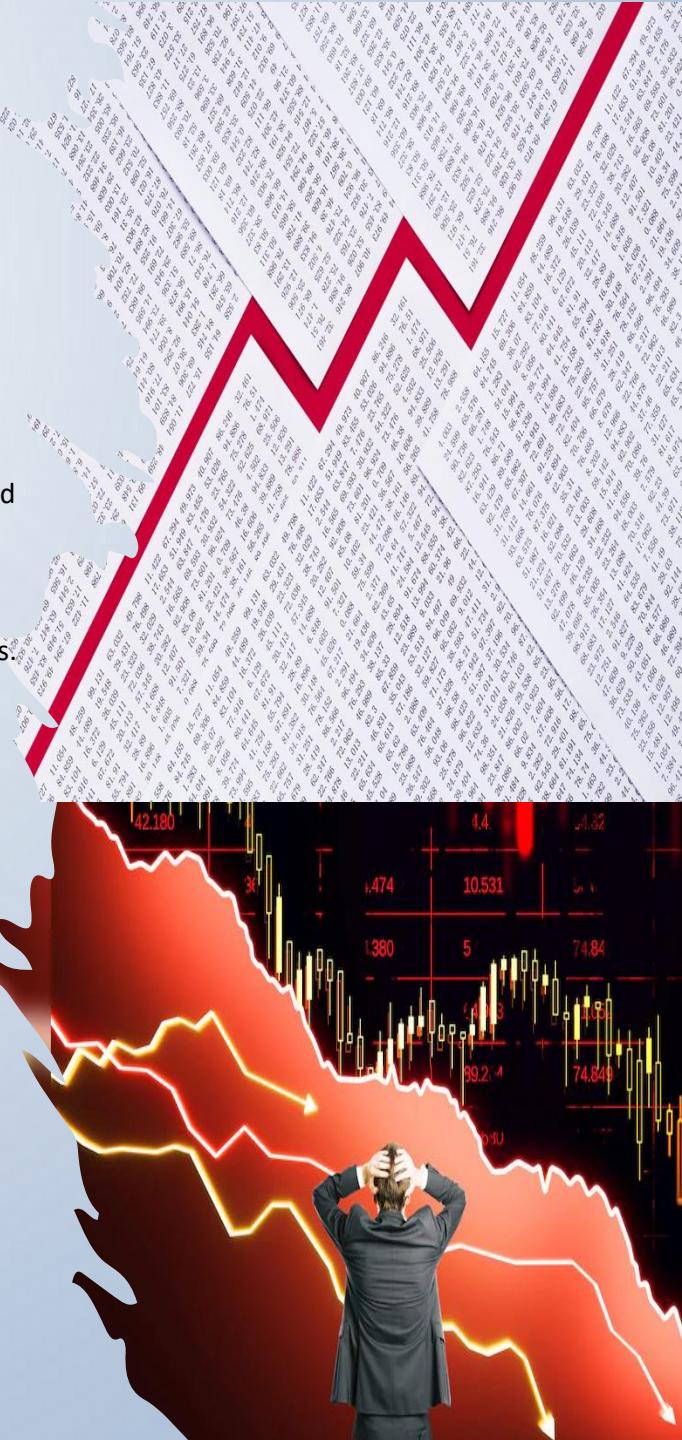
## Data Overload:

Inability to process and interpret large, unstructured datasets quickly and effectively.

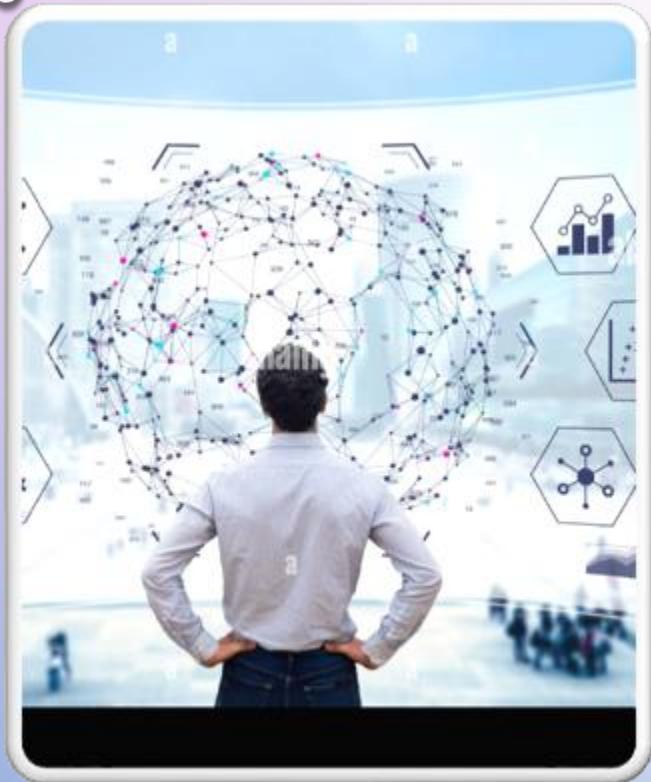
Difficulty in generating actionable trading strategies.

# Goal

- Predict stock price movements and generate explainable strategies using AI.



# PROPOSED SOLUTION



- TO DEVELOP AN ADVANCED PREDICTIVE MODEL THAT LEVERAGES LSTM WITH DRNN ENHANCED BY GANS, TO IMPROVE THE ACCURACY OF STOCK PRICE FORECASTING.
- BUILD A GEN-AI AGENT ("STOCK BUDDY") FOR SMART INVESTMENT DECISION-MAKING.
- LONG SHORT-TERM MEMORY (LSTM) NETWORKS WITH DEEP RECURRENT NEURAL NETWORKS
  - CAPTURES BOTH SHORT-TERM FLUCTUATIONS AND LONG-TERM TRENDS
  - IDEAL FOR TIME-SERIES DATA LIKE STOCK PRICES
- GENERATIVE ADVERSARIAL NETWORKS (GANS) INTEGRATION:
  - GENERATOR CREATES SYNTHETIC STOCK DATA
  - DISCRIMINATOR DISTINGUISHES BETWEEN REAL AND FAKE DATA
  - ENHANCES TRAINING BY PROVIDING MORE REALISTIC DATA SCENARIOS

# TOOLS, FRAMEWORKS, AND PLATFORMS OVERVIEW

Category	Tool/Framework	Purpose
Data Acquisition	Yfinance, pandas, numpy	Fetch around 5,000000 records + over 30 indicators
Data Lake	AWS S3	Store raw market data (Parquet/CSV)
Data Warehouse	Snowflake	Processed data analytics
ML Development	Amazon SageMaker AI Studio	LSTM price forecasting
ML Finetuning	Amazon SageMaker AI Studio	GAN synthetic data generation
ETL	Python, Pandas, SQLAlchemy	<b>Data Processing</b>
ML Platform	AWS SageMaker AI	Model training/deployment
Vector Database	Pinecone	for embedding/vector search
Agent Framework	LangChain	Orchestrate tools (SQL, RAG, ML)
LLM	OpenAI GPT-4	Natural language responses
Visualization	Power BI Dashboard	Interactive charts

# DATA SOURCING

# DATA SOURCE

## Yahoo Finance via Python library:

Use the **yfinance** library to fetch historical stock prices, volume, and other key financial metrics.

```
▶ import yfinance as yf  
import pandas as pd  
import matplotlib.pyplot as plt  
  
# Download stock data  
stock = yf.download('AAPL', start='2018-01-01', end='2023-01-01')
```

- Provides **in-depth financial analysis**, including **earnings reports, market sentiment, and stock news**.
- Data can be scraped or accessed via APIs to provide a richer context for stock predictions.

Date	Open	High	Low	Close	Volume	Dividends	Stock Splits	Ticker
2021-04-14 00:00:00	132.0037882	132.0624803	128.7951602	129.1571045	87222800	0	0	AAPL
2021-04-15 00:00:00	130.9081982	132.0625152	130.732107	131.5733948	89347100	0	0	AAPL
2021-04-16 00:00:00	131.3777063	131.7396505	130.3798967	131.2407532	84922400	0	0	AAPL
2021-04-19 00:00:00	130.6049134	132.5222718	130.4386143	131.9059753	94264200	0	0	AAPL
2021-04-20 00:00:00	132.0820509	132.5809483	128.941892	130.2136078	94812300	0	0	AAPL
2021-04-21 00:00:00	129.1700206	130.8306834	128.4420060	130.5951233	68817100	0	0	AAPL

# DATA SOURCE

```
[ ] df.columns
```

```
Index(['date', 'open', 'high', 'low', 'close', 'volume', 'dividends',
       'stocksplits', 'ticker', 'money_flow_idx', 'on_balance_volume',
       'accumu_dist_line', 'chaikin_money_flow', 'vol_wtd_avg_price',
       'simple_moving_avg', 'relative_strength_idx', 'price_vol_trend',
       'vol_rate_of_change', 'mov_avg_conver_divergence', 'bb_mid', 'bb_upper',
       'bb_lower', 'avg_true_range', 'rsi_volume', 'avg_direc_idx',
       'ich_tenkan_sen', 'ich_kijun_sen', 'comm_channel_idx',
       'stochastic_osc', 'fib_23_6_retrace_lvl', 'fib_38_2_retrace_lvl',
       'fib_61_8_retrace_lvl', 'std_dev', 'exp_mov_avg_12', 'exp_mov_avg_26'],
      dtype='object')
```

```
[ ] df.shape
```

```
→ (498989, 35)
```

rsi_volume	float64
avg_direc_idx	float64
ich_tenkan_sen	float64
ich_kijun_sen	float64
comm_channel_idx	float64
stochastic_osc	float64
fib_23_6_retrace_lvl	float64

```
df.dtypes
```

		0
date	object	
open	float64	
high	float64	
low	float64	
close	float64	
volume	int64	
dividends	float64	
stocksplits	float64	
ticker	object	
money_flow_idx	float64	
on_balance_volume	int64	
accumu_dist_line	float64	
chaikin_money_flow	float64	
vol_wtd_avg_price	float64	
simple_moving_avg	float64	
relative_strength_idx	float64	
price_vol_trend	float64	
vol_rate_of_change	float64	
mov_avg_conver_divergence	float64	
bb_mid	float64	
bb_upper	float64	
bb_lower	float64	
avg_true_range	float64	
rsi_volume	float64	

# DATA STORAGE

## Raw Data Source

- Yahoo Finance API (yfinance library) – Historical stock prices, volumes, technical metrics.

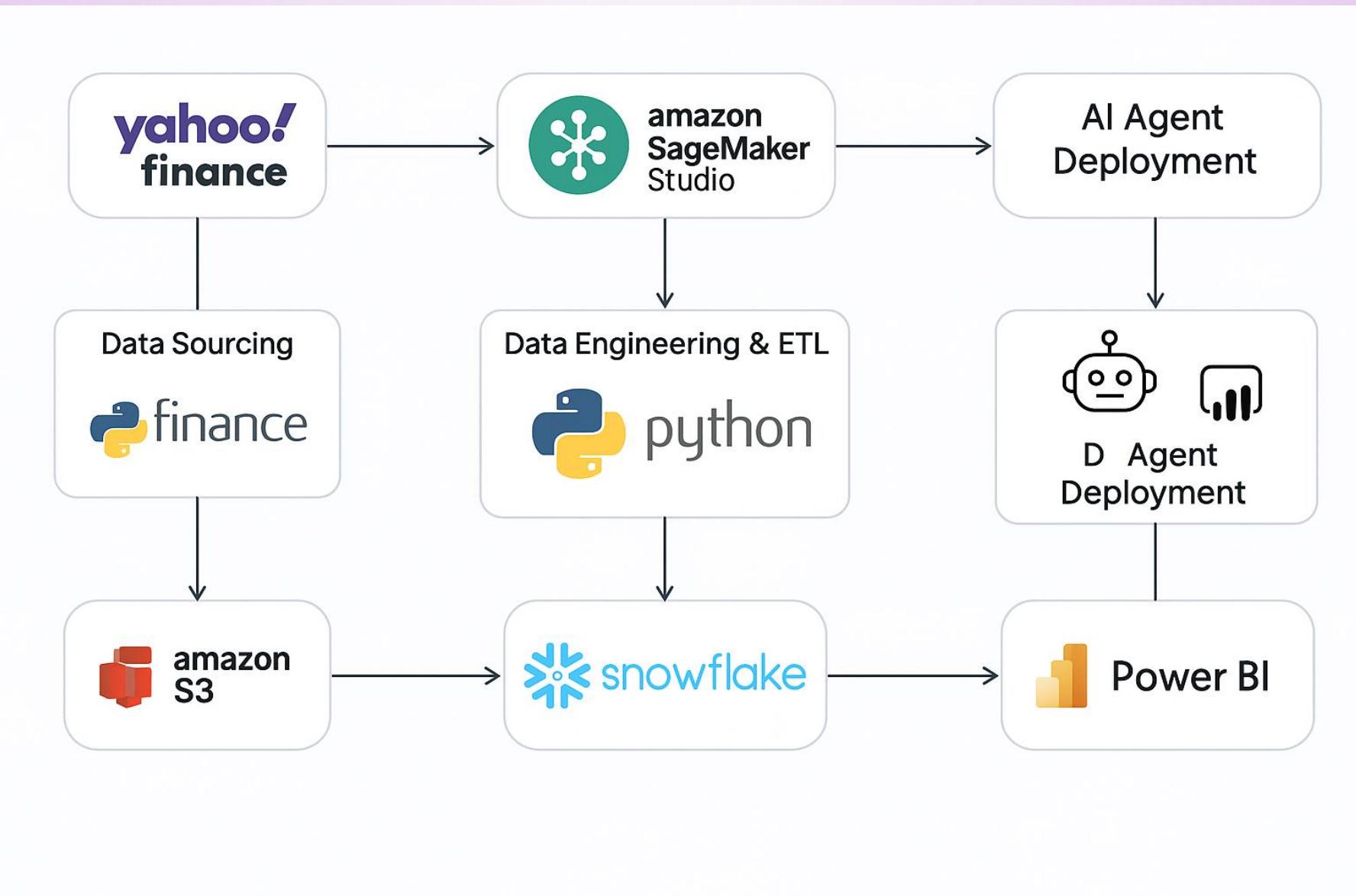
## Storage Formats

- Initial Raw Data: CSV files (larger size ~870 MB).
- Optimized Storage: Parquet files (compressed ~34 MB).
- Faster read/write.
- Supports complex types (arrays, maps).
- Predicate pushdown for faster querying.

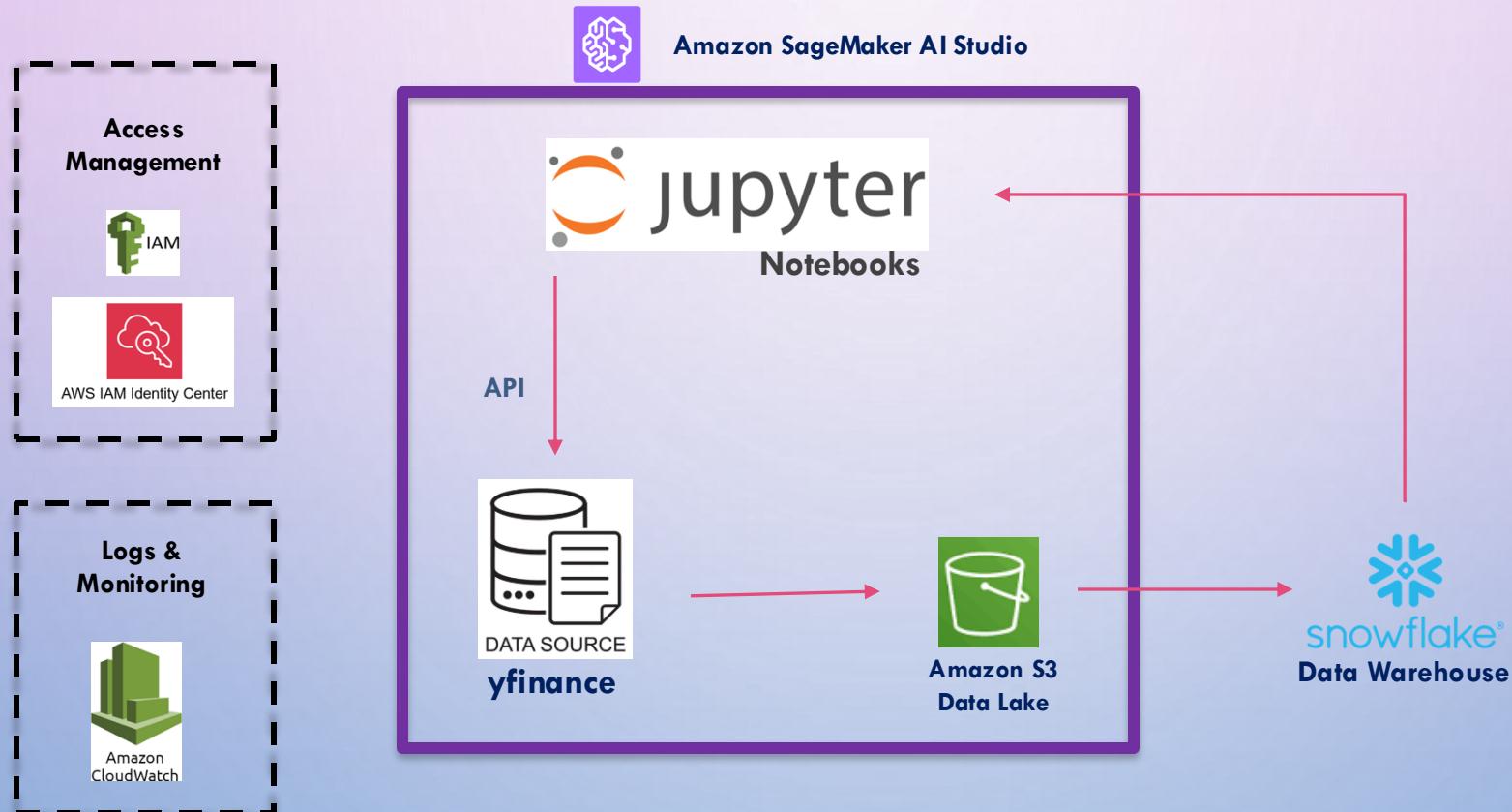
## Cloud Storage and Platforms

- Amazon S3 Buckets:
- Stored raw and cleaned datasets securely.
- Integrated with Amazon SageMaker for direct ML model access.
- Snowflake Data Warehouse:
- Hosted cleaned, transformed data (STOCK\_FINAL\_CLEANED table).
- Ready for downstream ML modeling and Power BI visualizations.

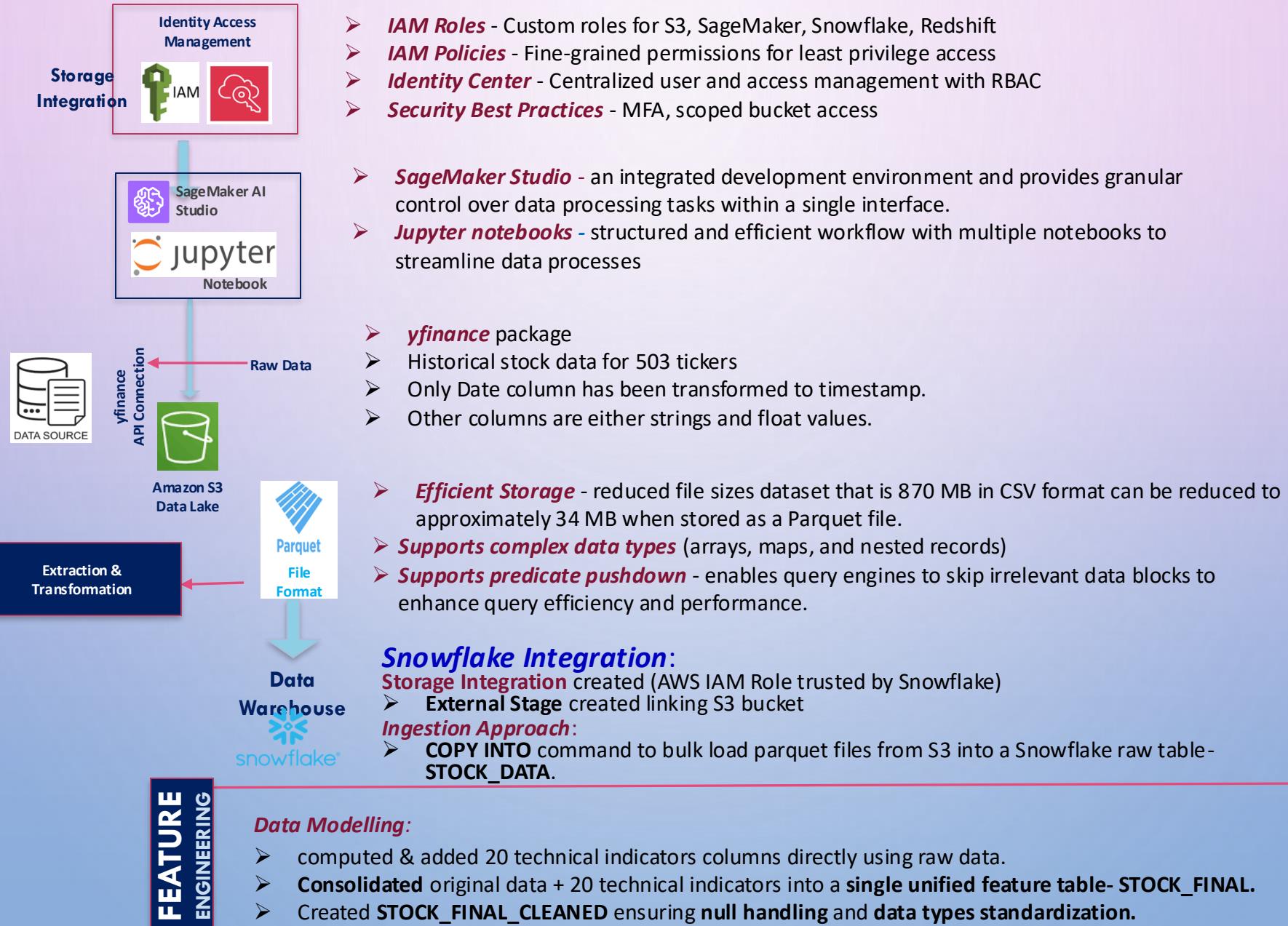
# AI-DRIVEN STOCK TRADING STRATEGY ANALYSIS- FULL ARCHITECTURE FLOW



# END-TO-END DATA ENGINEERING FLOW



# DATA ENGINEERING & ETL



## Data Ready-to-use

- Downstream ML modeling
- Reporting (Dashboards)



SageMaker AI  
Studio



**GEN AI Agent**



Power BI

# FINAL DATA VIEW SNOWFLAKE

	DATE	OPEN	HIGH	LOW	CLOSE	VOLUME	DIVIDENDS	STOCKSPLITs	TICKER	MONEY_FLOW_IDX			
1	2024-03-06	175.54	173.51	176.46	173.26	32090900	0	0	AMZN	55.409			
2	2024-03-06	68.19	68.17	68.814	67.682	3953300	0	0	FIS	86.941			
3	2024-03-06	574.91	575.97	582.52	567.01	787800	0	0	SNPS	49.908			
4	2024-03-06	29.811	29.541	29.869	29.329	4716100	0	0	APA	47.727			
5	2024-03-06	282.61	285.24	289.41	281.26	561400	0	0	CPAY	55.361			
6	2024-03-06	233.719	232.075	234.379	230.095	276700	0	0	ESS	60.106			
7	2	ON_BALANCE_VOLUME		ACCUMU_DIST_LINE		CHAIKIN MONEY_FLOW		VOL_WTD_AVG_PRICE		SIMPLE_MOVING_AVG		RELATIVE_STRENGTH_I	
8	2	1	217225200	714042265650.181		7.682		172.388		170.972		61.0	
	2	2	-44274900	41524584358.609		7.876		64.097		62.691		80.9	
	3	3	30850800	10417694803.356		9.656		570.025		561.38		53.9	
	4	4	417547000	62707757505.888		3.053		29.376		28.97		59.1	
	5	5	10840500	4987319511.295		43.127		276.305		273.313		67.4	
	6	6	-15329500	4853262465.768		27.528		221.926		219.503		69.8	
	7	7	66422100	24932426645.382		13.454		55.405		55.007		62.8	
8	8	RELATIVE_STRENGTH_IDX		PRICE_VOL_TREND		VOL_RATE_OF_CHANGE		MOV_AVG_CONVER_DIVERSION		BB_MID	BB_UPPER	BB_LOW	
9	9	1	61.067	-38426466.674		-25.048		4.02	170.972	177.75	164.1		
	2	2	80.914	-13980802.507		-19.315		2.074	62.691	68.738	56.6		
	3	3	53.919	701079.518		-15.236		10.313	561.38	592.047	530.7		
	4	4	59.169	6942130.895		-23.209		-0.325	28.97	29.887	28.0		
	5	5	67.412	-167141.485		19.269		0.45	273.313	286.506	260.1		
	6	6	69.882	-112648.908		-39.559		0.705	219.503	228.792	210.2		
	7	7	62.847	-2721739.127		44.033		0.048	55.007	56.294	53.7		
	8	8	69.933	-1571524.07		-33.13		10.364	189.795	214.24	165.		
	9	9	53.454	290485.307		-13.625		2.796	244.879	251.72	238.0		

# DATA QUALITY

289.33

# DATA QUALITY

## Accuracy

Stock prices (open, close, high, low) and volume data accurately reflected real trading activity from Yahoo Finance.

## Validity

All fields adhered to expected formats and rules, such as valid ticker symbols, non-negative volume values, and appropriate date ranges to maintain dataset integrity.

## Uniqueness

Duplicate records were removed, ensuring that each stock's daily trading data point (ticker + date) was unique without unnecessary repetitions.

## Integrity

- Enforced schema consistency during Snowflake ingestion to ensure overall data integrity
- Verified no corruption during S3 upload and retrieval process.

## Relevance

- Only relevant data necessary for stock trend prediction and technical indicator computation was retained
- Eliminated unrelated or unnecessary fields.

## Completeness

All columns like date, open, high, low, close, volume, and ticker are fully populated.

## Timeliness

Stock data was extracted covering the last 5 years and updated at the time of retrieval, ensuring relevance and freshness for both training and future prediction tasks.

## Consistency

- Standardized formats (e.g., datetime format)
- Ensured that each ticker had uniform schema and no contradictory records across files.

## Traceability

Preserved the ability to trace back every data file to its original extraction timestamp by proper file naming, S3 bucket structuring, and Snowflake tables.

# **Major Types of Stock Indicators**

## **Trend-Following Indicators**

Identify market direction & strength

## **Momentum/Oscillator Indicators**

Measure overbought/oversold conditions & reversals.

.....

## **Volume-Based Indicators**

Analyze trading volume to confirm trends or divergences

## **Volatility Indicators**

Often grouped separately but can fit under Trend or Momentum

# Trend-Following Indicators

## Moving Averages (SMA, EMA)

- Moving average that gives more weight to recent prices.
- Reacts faster than SMA; used for entry/exit signals.
- Crossovers (price vs EMA) indicate trend changes.
- Average of closing prices over a period.
- Smooths out price data to identify trends.
- Commonly used in crossover strategies.

## MACD (Moving Average Convergence Divergence)

- Tracks the difference between short and long EMAs.
- Signals momentum and potential reversals.
- MACD line crosses signal buy/sell points.

## Ichimoku Cloud

- All-in-one trend, support/resistance, and momentum tool.
- Price above cloud = bullish; below = bearish.
- Cloud thickness shows support/resistance strength.

## Average Directional Index

- Measures trend strength (not direction).
- $ADX > 25$  indicates a strong trend;  $< 20$  = weak/no trend.
- Often used with +DI and -DI lines.

# Momentum/Oscillator Indicators

## RSI (Relative Strength Index)

- Measures the speed and change of price movements.
- RSI  $>70$  = overbought;  $<30$  = oversold.
- Useful for spotting reversals and divergences.

## Stochastic Oscillator

- Compares a stock's closing price to its price range over a period.
- Identifies overbought ( $>80$ ) and oversold ( $<20$ ) levels.
- Crossovers provide trading signals.

## Money Flow Index

- Combines price and volume to measure buying and selling pressure.
- Helps identify overbought or oversold conditions.
- Graph fluctuates between 0–100; values  $>80$  indicate overbought,  $<20$  oversold.

## VOLUME RSI

- Combines volume and RSI concepts to gauge strength.
- Indicates bullish/bearish volume momentum.
- Above 70 = strong volume buying; below 30 = strong selling.

# Volume-Based Indicators

## OBV (On-Balance Volume)

- Cumulative volume-based indicator that reflects buying/selling pressure.
- Rising OBV suggests accumulation; falling indicates distribution.
- Graph direction supports or contradicts price trends.

## Chaikin Money Flow (CMF)

- Measures volume-weighted accumulation over a period.
- Positive CMF = accumulation; negative = distribution.
- Crosses zero line for trend signals.

## VWAP (Volume-Weighted Average Price)

- Average price weighted by volume throughout the day.
- Traders use it as a benchmark for fair value.
- Price above VWAP = bullish; below = bearish.

## Accumulation/ Distribution Line

- Uses price and volume to determine whether stock is accumulated or distributed.
- Rising line = buying pressure; falling = selling pressure.
- Divergence from price can predict reversals.

# Volatility Indicators

## Bollinger Bands

- A band of standard deviation lines around an SMA.
- Price near upper band = overbought; lower band = oversold.
- Band squeeze suggests upcoming volatility.

## Fibonacci Retracements

- Horizontal lines indicate potential support/resistance levels.
- Based on % retracement levels (23.6%, 38.2%, etc.).
- Used to predict reversals within trends.

## Standard Deviation

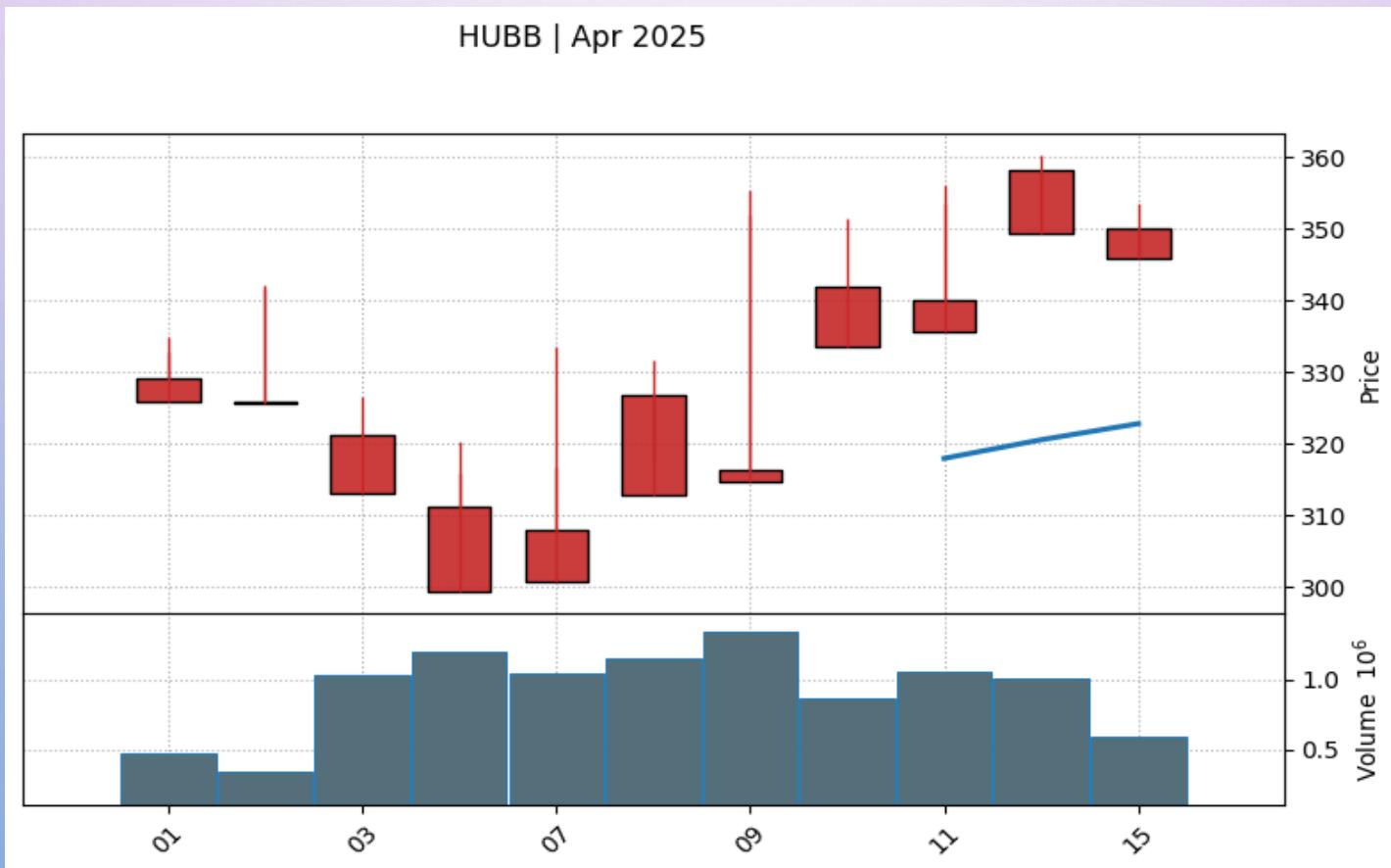
- Measures price volatility from the mean.
- High value = volatile market; low = stable.
- Used to set risk levels and detect breakouts

# DATA VISUALIZATIONS

**Power BI Dashboard Visualizations Link**

# Price Action Analysis - HUBB (April 2025)

- Shows daily price movement between ~300-360 with a slight downward trend
- Volume spikes correlate with key price movements (potential institutional activity)
- The 12-day EMA (blue) crossing below 26-day EMA (orange) suggests bearish momentum
- Key levels to watch: 320 (support), 340 (resistance)

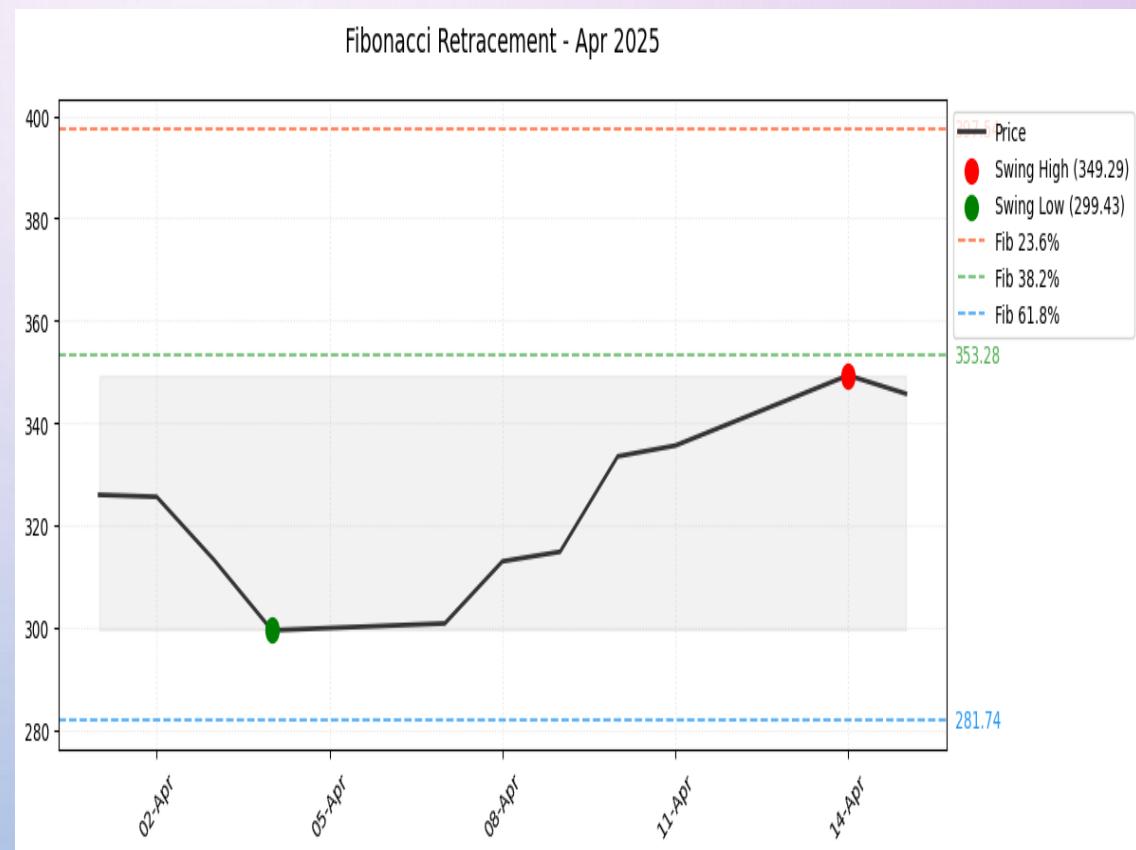


# Support/Resistance Levels

## Fibonacci Retracement

- Swing High: 349.29 | Swing Low: 299.43 (19.8% price range)
- Key retracement levels:
  - 23.6% at 353.28 (immediate resistance)
  - 38.2% at 335.50 (mid-level)
  - 61.8% at 310.75 (strong support)
- Current price hovering near 38.2% level - decisive break could indicate next trend direction

• • • •



# Technical Analysis Of Stock Performance

## Stock Performance Overview

Year	Quarter	Month	Day	TICKER								
All	All	All	4	AAPL								
TICKER	Year	Quarter	Month	Day	RELATIVE_STRENGTH_IDX	CLOSE	MOV_AVG	CONVER_DIVERGENCE	LOW	OPEN	HIGH	% Price Change
AAPL	2021	Qtr 2	June	4	40.72	121.36		-1.36	123.62	121.58	123.36	-0.1%
AAPL	2023	Qtr 1	January	4	18.51	123.64		-4.58	127.18	125.43	124.91	-1.4%
AAPL	2021	Qtr 2	May	4	24.91	123.94		-0.42	128.63	128.34	125.07	-3.4%
AAPL	2022	Qtr 4	November	4	42.81	132.84		-1.06	141.03	140.46	136.79	-5.4%
AAPL	2021	Qtr 4	October	4	28.14	135.69		-2.02	139.56	139.12	136.55	-2.4%
AAPL	2022	Qtr 4	October	4	36.92	142.37		-4.38	144.30	143.13	144.18	-0.9%
AAPL	2021	Qtr 3	August	4	48.11	143.34		2.48	144.82	144.31	144.00	-0.3%
AAPL	2021	Qtr 4	November	4	69.77	147.83		1.48	149.59	148.76	148.15	-0.6%
AAPL	2022	Qtr 2	May	4	38.14	156.72		-2.68	163.83	157.12	163.37	-0.2%
AAPL	2022	Qtr 1	March	4	42.82	159.52		-1.61	162.91	161.87	160.57	-1.4%
AAPL	2022	Qtr 3	August	4	80.09	162.05		5.21	164.77	163.60	163.41	-0.9%
AAPL	2023	Qtr 2	May	4	51.65	162.67		2.36	165.37	163.24	164.14	-0.6%
AAPL	2023	Qtr 2	April	4	88.97	163.46		3.97	165.17	164.94	163.98	-0.3%
AAPL	2022	Qtr 1	February	4	49.74	167.96		0.16	171.32	168.94	169.64	-0.1%
AAPL	2024	Qtr 2	April	4	40.56	168.03		-2.37	171.11	169.49	168.03	-0.3%
AAPL	2023	Qtr 4	October	4	42.42	169.73		-2.47	172.94	169.85	172.40	-0.3%
AAPL	2022	Qtr 2	April	4	84.29	171.66		3.37	175.64	171.79	175.60	-0.3%

- Displays AAPL stock metrics on the 4th day across years and quarters.
- Key indicators: RSI, Closing Price, MACD, Low, Open, High, % Price Change.
- Highlights stock volatility, momentum, and price trends over time.
- Interactive filters: Year, Quarter, Month, and Day for custom analysis.
- Supports data-driven insights for investment strategy proposals.

- Oversold conditions like January 2023 presented **strong buying opportunities** — ideal for value investors.
- Overbought conditions like April 2023 and April 2022 warned of **possible pullbacks**, suggesting profit booking.
- Negative MACD and Low RSI combo (e.g., October 2022, January 2023) reinforced **bearish biases**.
- Positive MACD with High RSI (e.g., April 2023) confirmed **strong bullish momentum**, good for momentum traders but risky for late entries.

The market behaviour in 2022-2023 demonstrates how momentum indicators can predict shifts well before price changes are visible.

# Price and Volume Analysis

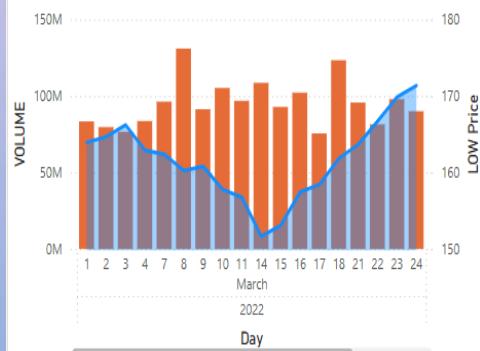
## Open vs Close Price

● Sum of CLOSE ● Sum of OPEN



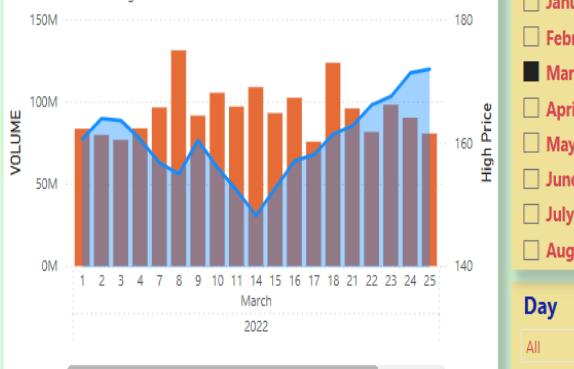
## Volume vs Price Change

● VOLUME ● LOW Price



## VOLUME Vs High Price

TICKER ● AAPL ● High Price



## TICKER

- A
- AAPL
- ABBV
- ABNB

## Year

- 2021
- 2022
- 2023
- 2024
- 2025

## Month

- January
- February
- March
- April
- May
- June
- July
- August

## Day

- All

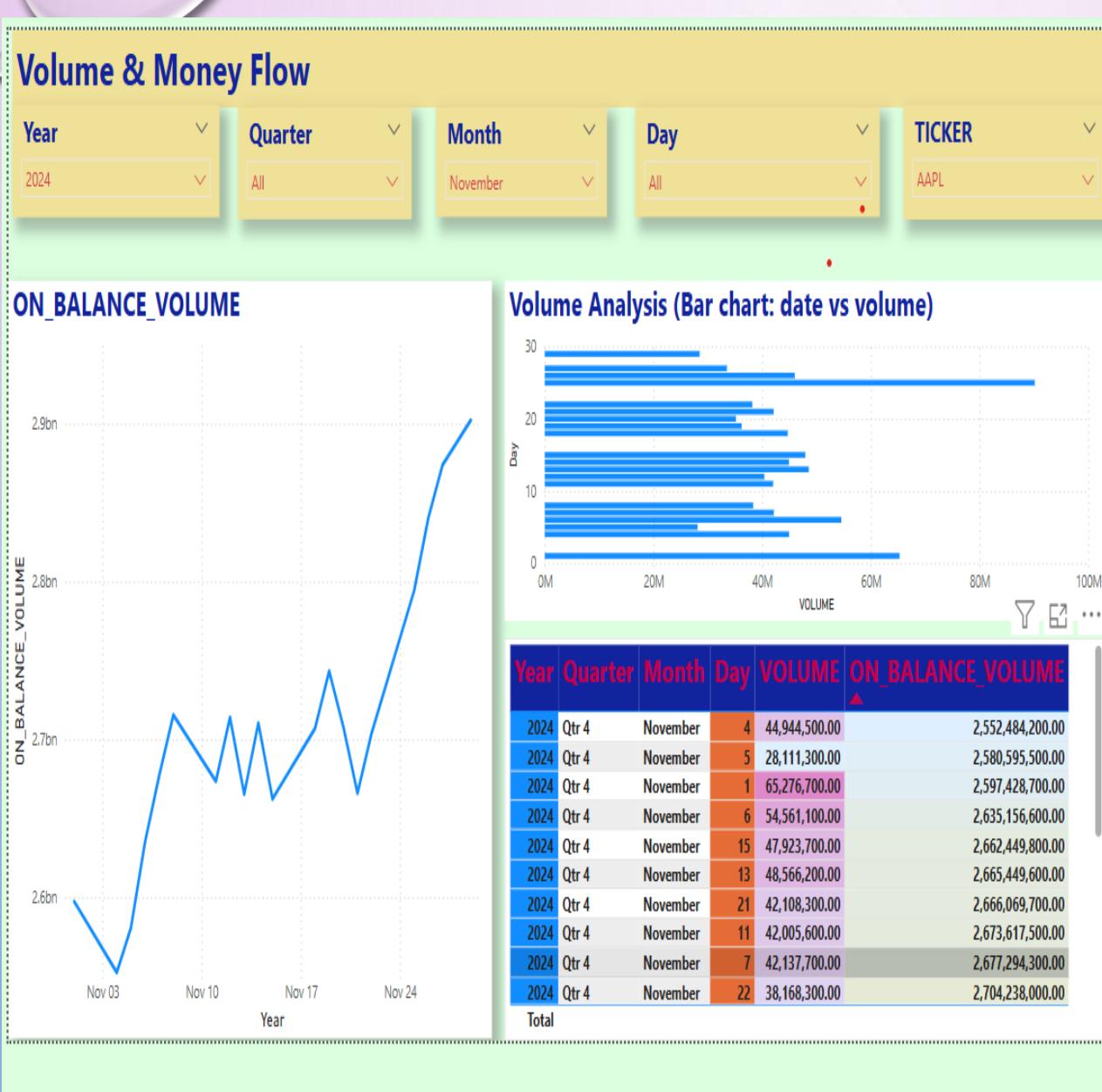
## Open vs Close Price Analysis

- **Early March:** Open price > Close price → bearish daily closes.
- **Mid-March:** Sharp dip in both open and close prices.
- **Post Mid-March:** Strong recovery with open and close prices rising together.
- **End of March:** Bullish sentiment returns, market strength improving.

## Volume and Price-Volume Trend

- **Early** Volume remained stable despite early March price drops.
- No panic selling observed during market dip.
- Increased volume supported price recovery after mid-March.
- Strengthening price-volume trend suggests rising investor confidence.

# Volume and Money Flow Analysis



## Volume Trends:

- Significant volume spikes on November 1st (65.3M shares) and 6th (54.6M shares)
- Generally high trading activity throughout the month (>38M shares daily)
- Typical volume range: 38M-65M shares/day

## Money Flow (OBV):

- Consistent upward trajectory in On-Balance Volume (2.55B → 2.70B)
- Every trading day shows positive OBV accumulation
- Strongest accumulation on November 1st (+37.8M OBV)

## Notable Patterns:

- Higher volume days (1st, 6th) correlate with larger OBV increases
- Even lower volume days (>38M) show continued accumulation
- No distribution days (negative OBV) in this period

# Momentum & Strength - AAPL

## Momentum & Strength

Year

2024

Quarter

All

Month

November

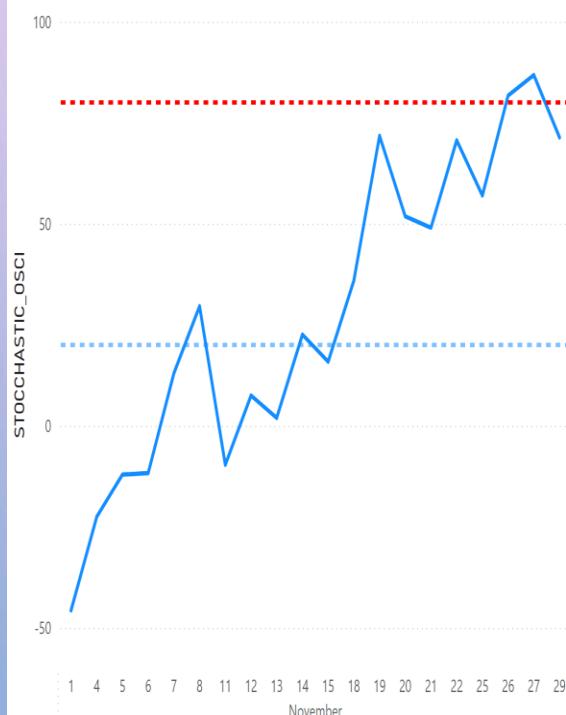
Day

All

TICKER

AAPL

### Line chart with thresholds (20/80)



### MACD over time Table

MOV_AVG_CONVER_DIVERGENCE	Year	Quarter	Month	Day
-0.98	2024	Qtr 4	November	13
-0.92	2024	Qtr 4	November	12
-0.90	2024	Qtr 4	November	11
-0.87	2024	Qtr 4	November	6
-0.85	2024	Qtr 4	November	7
-0.84	2024	Qtr 4	November	14
-0.77	2024	Qtr 4	November	15

### MOV\_AVG\_CONVER\_DIVERGENCE by Day



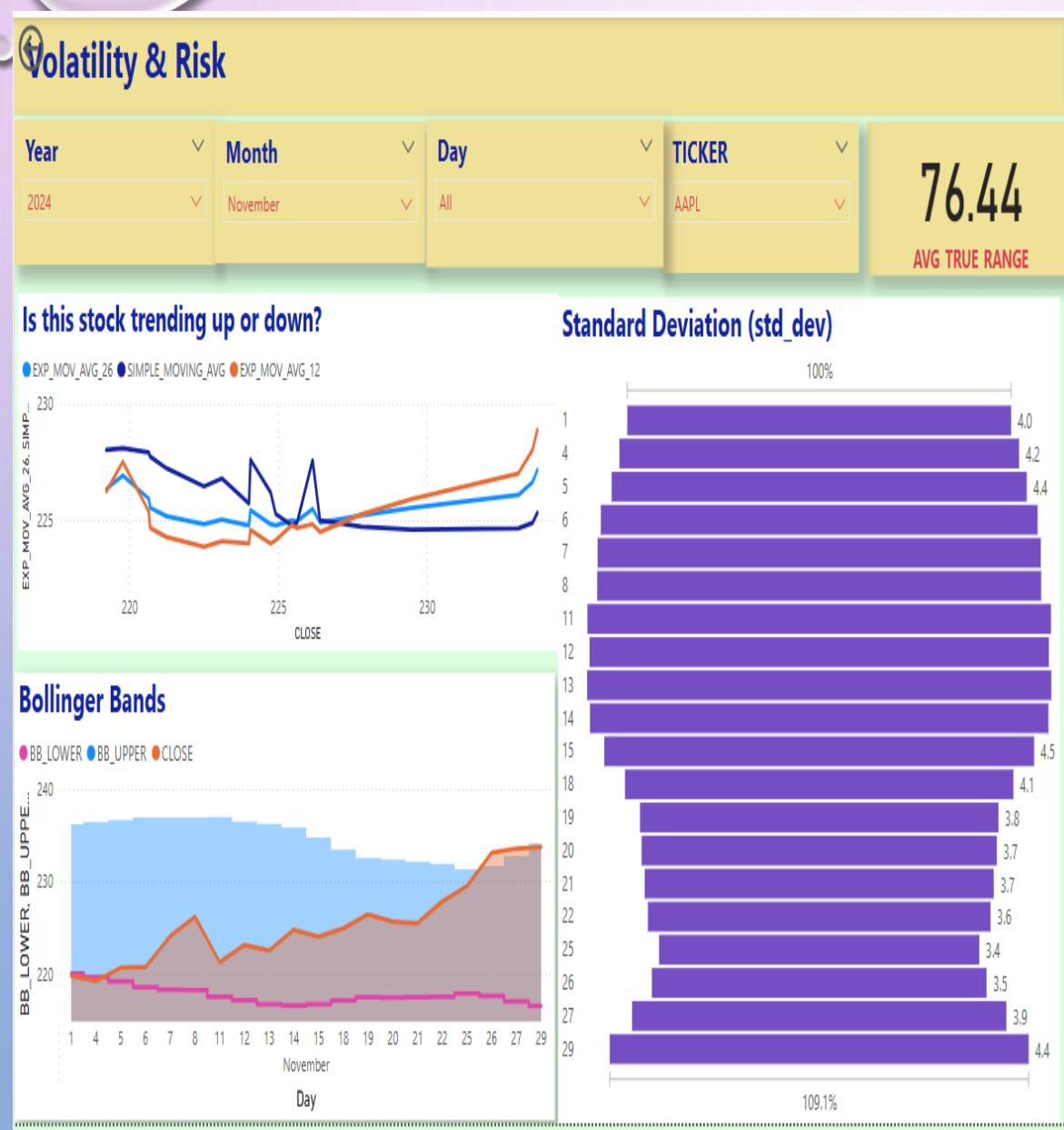
## Key Findings:

- Early month: Bearish momentum (MACD negative, Oscillator negative).
- Mid to late month: Bullish reversal (MACD turns positive, Oscillator high).
- Late month: Overbought risk.

## Impact & Strategy:

- Early November: Risk-off / cautious trading.
- Mid November: Opportunity for entry (momentum shift).
- Late November: Monitor for pullback signs due to overbought levels.
- Emphasize the need for adaptive strategies based on momentum indicators.

# Volatility Analysis



## Standard Deviation (Std\_Dev) Plot

- **Volatility Baseline:** Values hover around 1.0–1.5 → Stable/low volatility.
- **Spike Detection:** Higher values (e.g., Nov 29) → Possible news/event-driven moves.
- **Action:** Monitor spikes for breakout confirmation.

## Bollinger Bands Plot

- **Range-Bound Trading:** Price consistently oscillates between upper and lower bands → No strong trend.
- **Squeeze Alert:** Bands are relatively narrow → Low volatility, potential breakout imminent.
- **Critical Levels:**
  - *Support:* Lower band (buy zone)
  - *Resistance:* Upper band (sell zone)

## Moving Averages (EXP\_MOV\_AVG\_12/26)

### Trend Direction:

- *Positive Angle:* Uptrend likely
- *Negative Angle:* Downtrend likely

### Crossover Watch:

- 12EMA > 26EMA → Bullish signal
- 12EMA < 26EMA → Bearish

# Volume and Money Flow Analysis



## Oversold Signal (RSI)

- RSI at 34.31 (Near oversold territory)
- Below neutral 50 level → Potential buying opportunity
- Not yet at extreme oversold (<30) → Caution still advised

## Money Flow Weakness

- MFI at 38.97 (Below critical 40 level)
- Shows distribution pressure
- Divergence from price action would be concerning

## Chaikin Money Flow

- Value: +6 (Positive but weak)
- Above zero = Net buying pressure
- Below 10 = Not strong accumulation

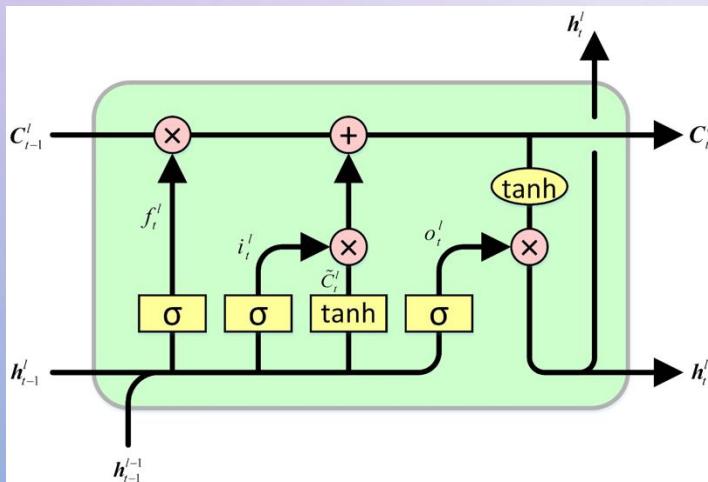
## Volatility (ATR)

- ATR: 3.40 (Absolute dollar volatility)
- Context needed:
  - If stock price = \$180 → ~1.9% daily range
  - If stock price = \$100 → ~3.4% daily range

# MACHINE LEARNING MODELLING

289.33

# LSTM AND MODEL ARCHITECTURE



## What we built ! - Introduction to LSTM

- LSTM (Long Short-term Memory) is a type of deep learning model (RNN) built to understand sequential data over time — like stock prices, weather patterns, or speech.

## Why LSTM for stock data?

- Stock prices are sequential — today's price depends on the past.
- LSTM models capture historical patterns better than traditional feedforward networks.

## Architecture:

LSTM introduces a **memory cell** and **gates** to control information flow

- **Input gate**: decides which information from the input should be kept.
- **Forget gate**: decides what information should be discarded from memory.
- **Output gate**: determines what the output should be based on the memory and input.

# STOCK PRICE PREDICTION LSTM

Let's now walk through how to build a stock prediction using machine learning by leveraging an LSTM network to forecast stock price movements.





# DATA CLEANING

- **SOURCE:** STOCK DATA RETRIEVED SECURELY FROM SNOWFLAKE DATABASE USING **SQLALCHEMY**.
- **FILTERING:** SELECTED DATA FOR A SPECIFIC TICKER (E.G., AAPL).
- **MISSING VALUES:** ASSUMED CLEAN DATASET (NO EXPLICIT HANDLING SHOWN).
- **SELECTED COLUMNS:** FOCUSED ONLY ON RELEVANT NUMERICAL COLUMNS FOR MODELING.

# FEATURE ENGINEERING

## Feature Selection:

- Chose **14 technical indicators and stock attributes**:
- open, high, low, volume, simple\_moving\_avg,
- mov\_avg\_conver\_divergence, bb\_mid, bb\_upper, bb\_lower,
- avg\_true\_range, rsi\_volume, std\_dev, exp\_mov\_avg\_12, exp\_mov\_avg\_26.
- **Target variable**: close price (next day).

## Normalization:

- Applied **MinMaxScaler** to scale all features and target to the range **[-1, 1]**.
- Improves training speed and convergence for LSTM and GAN models.

## Sequence Creation:

- Used a **sliding window** of size **10** (sequence\_length = 10).
- Each input sample = 10 past days of features.
- Target = close price on the **11th day**.

## Tensor Conversion:

- Converted processed sequences into **PyTorch tensors** to feed into LSTM-based models.

# MODEL HYPERPARAMETERS

## INPUT SEQUENCE CREATION:

Used a rolling window of 60 days

- For each prediction, the model sees the previous **60 days** of closing prices as input.
- Predicts the **61st day's** closing price.

Formed sequences as

- **Samples** = number of training examples.
- **Timesteps** = 60.
- **Features** = 1 (closing price only).

## TRAINING SETUP

- Train-test split: 80% for training, 20% for testing.

## BATCH SIZE: 32

- The model processes 32 sequences at a time before updating weights.
- This balances training speed and model generalization.

# MODEL PERFORMANCE AND PREDICTIONS

## Stock Price Prediction Report

### Prediction Accuracy (First 10 Samples)

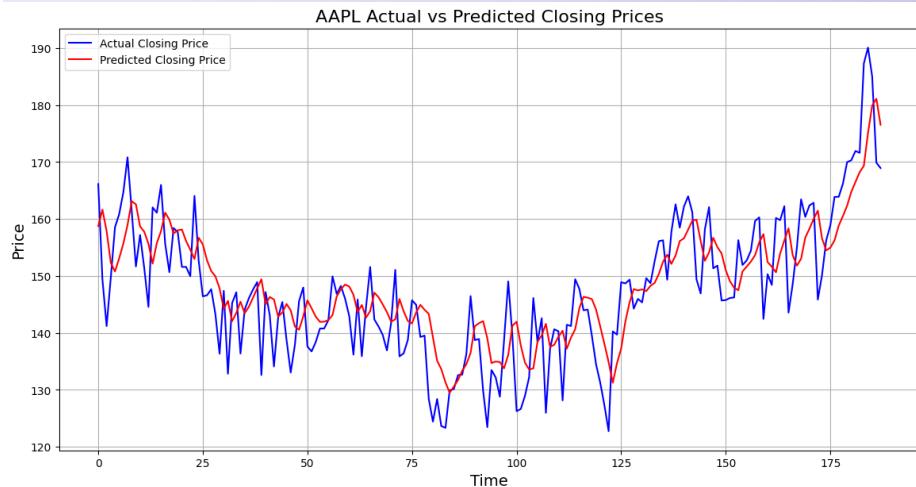
Index	Actual	Predicted	Difference	Accuracy
0	166.14	158.76	7.38	95.56%
1	149.39	161.66	-12.26	91.79%
2	141.20	157.88	-16.68	88.19%
3	149.74	152.31	-2.57	98.28%
4	158.56	150.79	7.77	95.10%
5	160.86	153.11	7.75	95.18%
6	164.71	155.70	9.01	94.53%
7	170.85	158.96	11.88	93.04%
8	161.41	163.12	-1.71	98.94%
9	151.67	162.59	-10.91	92.80%

### Prediction Performance Summary

Most Recent Prediction: \$176.55  
Average Prediction: \$148.49  
Prediction Range: \$129.49 – \$181.12  
Mean Absolute Error (MAE): \$5.99  
Mean Squared Error (MSE): \$53.68

## Base Model Performance Metrics:

- **Absolute Error (MAE):** 5.98
- **Mean Squared Error (MSE):** 53.67
- **Root Mean Squared Error (RMSE):** 7.32
- **R<sup>2</sup> Score:** 0.64



# **STOCK PRICE PREDICTION (GAN + LSTM) - FINE TUNED MODEL**

## **CHALLENGES WITH THE CURRENT MODEL**

- Stock data is limited and noisy.
- Traditional models overfit on small datasets.
- Need for smarter, richer datasets for more reliable forecasting.

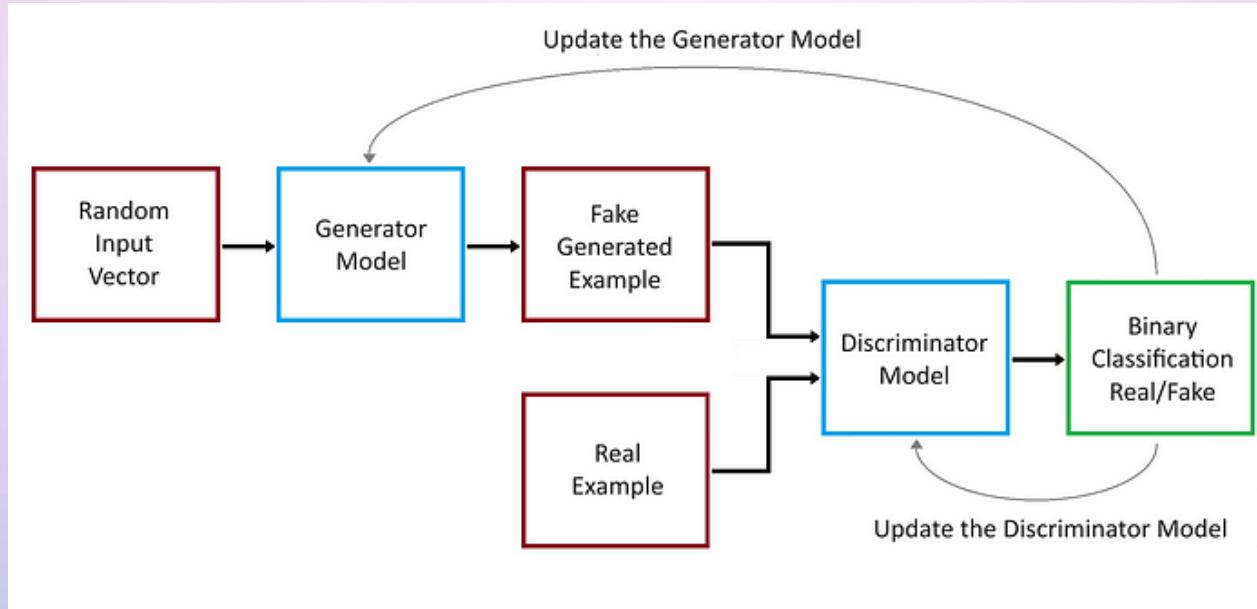
## **OUR SOLUTION**

Build a robust, ai-driven synthetic data pipeline. Combine long short-term memory (LSTM) and generative adversarial networks (gans) to enhance forecasting accuracy.

- GAN + LSTM create highly realistic stock patterns.
- Expand real-world data with generated sequences.
- Enable better model generalization and higher prediction accuracy.

# GENERATIVE ADVERSARIAL NETWORK

# A GENTLE INTRODUCTION TO GAN



A GAN consists of two neural networks — the generator and the discriminator — that are trained together in a competitive process. The generator creates fake data, while the discriminator evaluates how realistic the data is, leading the generator to improve.

## ➤ Architecture:

- **Generator:** A deep neural network that tries to generate fake data (e.g., Stock data).
- **Discriminator:** A deep neural network that attempts to classify if the data is real (from the dataset) or fake (generated by the generator).
- **Training process:** both networks compete. The generator tries to "fool" the discriminator, while the discriminator tries to distinguish between real and fake data.

## ➤ GAN's goal:

To generate realistic synthetic data that is indistinguishable from real data. In stock predictions, gans can simulate future stock prices or trends.

# LSTM-GAN ARCHITECTURE

## LSTM Generator

**Purpose:** Create realistic synthetic stock sequences from random noise.

- Input: Random vector (latent space).
- Output: Multifeatured stock sequence (Open, High, Low, Volume, Indicators).

## LSTM Discriminator

**Purpose:** Distinguish between real and generated sequences.

- Input: Real or Fake stock data.
- Output: Probability (Real or Fake).

## Training Process:

Alternate between:

- **Discriminator:** Improve real/fake classification.
- **Generator:** Improve deception skills/fooling Discriminator.

# **FINETUNE LSTM WITH GAN SYNTHETIC DATA**

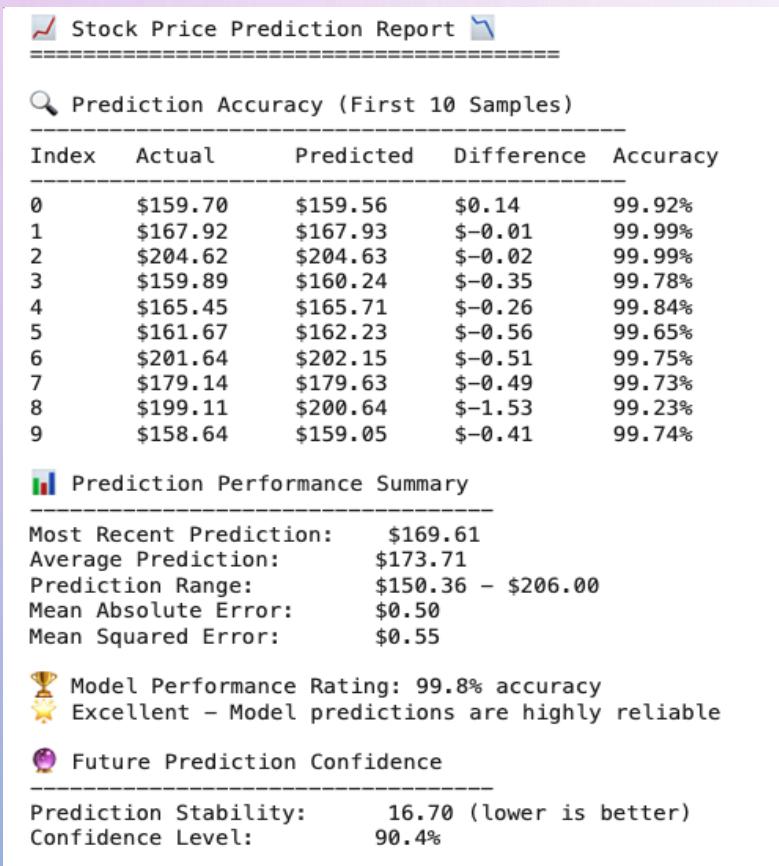
Once we generate a rich, expanded dataset (real + synthetic), we train a final LSTM model. This time focused only on making stock price predictions.

- Build synthetic stock data (GAN + LSTM).
- Expand dataset.
- Train final LSTM on richer dataset.
- Achieve better stock price predictions.

## **Synthetic data creation**

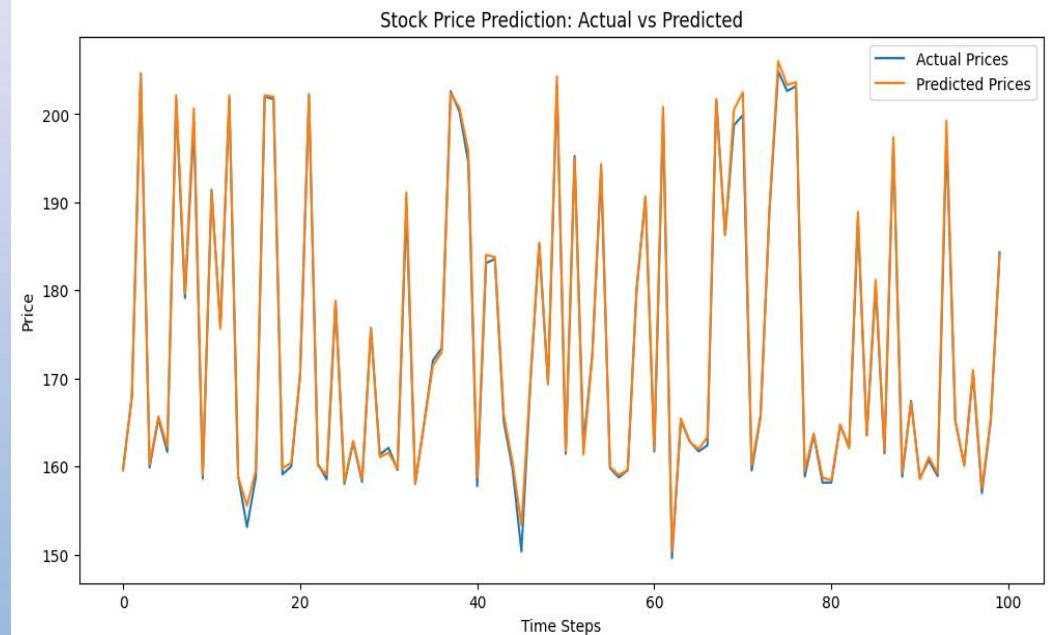
- Generate 100,000+ new synthetic stock samples.
- Blend with real-world stock data to create an enriched, powerful training dataset.

# FINE TUNED MODEL (LSTM+GAN) PERFORMANCE AND PREDICTION METRICS



## Fine Tuned Model Performance:

- Mean Squared Error (MSE) : 0.5497
- Mean Absolute Error (MAE) : 0.4961
- R<sup>2</sup>: 0.9980



# GENERATIVE AI AGENT

289.33

# Stock Buddy Gen-AI Agent

Empowering Smart Investment Decisions

## Why This Agent?

In the ever-evolving world of stock trading, making informed decisions is critical to achieving success. However, the process of analyzing market trends, understanding company performance, and predicting stock movements can be overwhelming for both novice and experienced traders. The sheer volume of data, the complexity of technical indicators, and the rapid pace at which markets move can make it difficult to keep up and make well-timed investment choices.

## What It Does?

The Stock Buddy Gen-AI Agent is designed to help both new and experienced investors make smarter, more informed decisions by providing them with personalized, AI-driven stock market analysis. By removing the complexity of traditional trading methods and offering real-time, actionable insights, Stock Buddy takes the guesswork out of investing and gives users the tools they need to succeed in today's fast-paced financial markets.



# Pain Points & Solutions

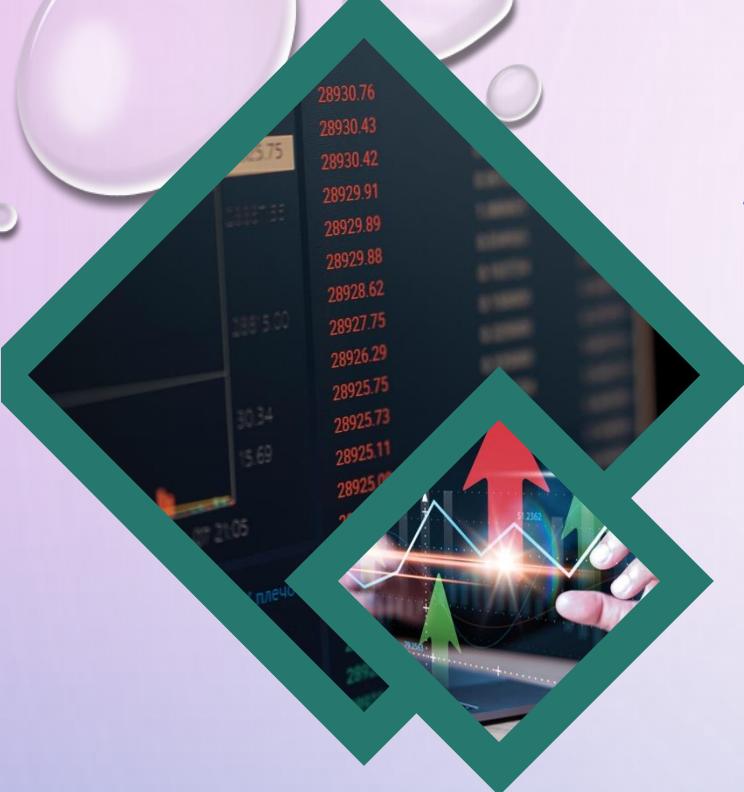


## Problems

-  **Overwhelming Data:** Investors face difficulty navigating through thousands of stocks, financial reports, and massive market data to extract actionable insights.
-  **Lack of Expertise:** Many users, especially beginners, struggle with understanding technical analysis, trading strategies, and market trends.
-  **Emotion-Driven Decisions:** Emotional biases like fear and greed often lead to irrational and impulsive trading behaviour.
-  **Real-Time Market Movements:** Rapid market changes make it hard to get timely, data-backed insights without advanced tools.
-  **Lack of Personalization:** Generic stock advice doesn't cater to individual goals, risk appetite, or investment strategies.

## Solutions

- **Empowers users with data-driven insights** to replace guesswork in investment decisions.
- **Automates technical analysis** to simplify complex trading concepts for investors.
- **Reduces emotional trading** with AI-backed advice and real-time market updates.
- **Provides personalized recommendations** based on user goals and risk profiles.
- **Saves time** by handling market research and analysis automatically.



# *Proposed Use Cases*

## Novice Investors

Stock Buddy simplifies the stock market for new investors by providing easy-to-understand recommendations, guiding them through the complexities of market data, and offering educational resources.



## Portfolio Management

Active traders can use Stock Buddy's real-time analysis and predictive capabilities to spot trends and make fast, informed decisions in a fast-paced environment.

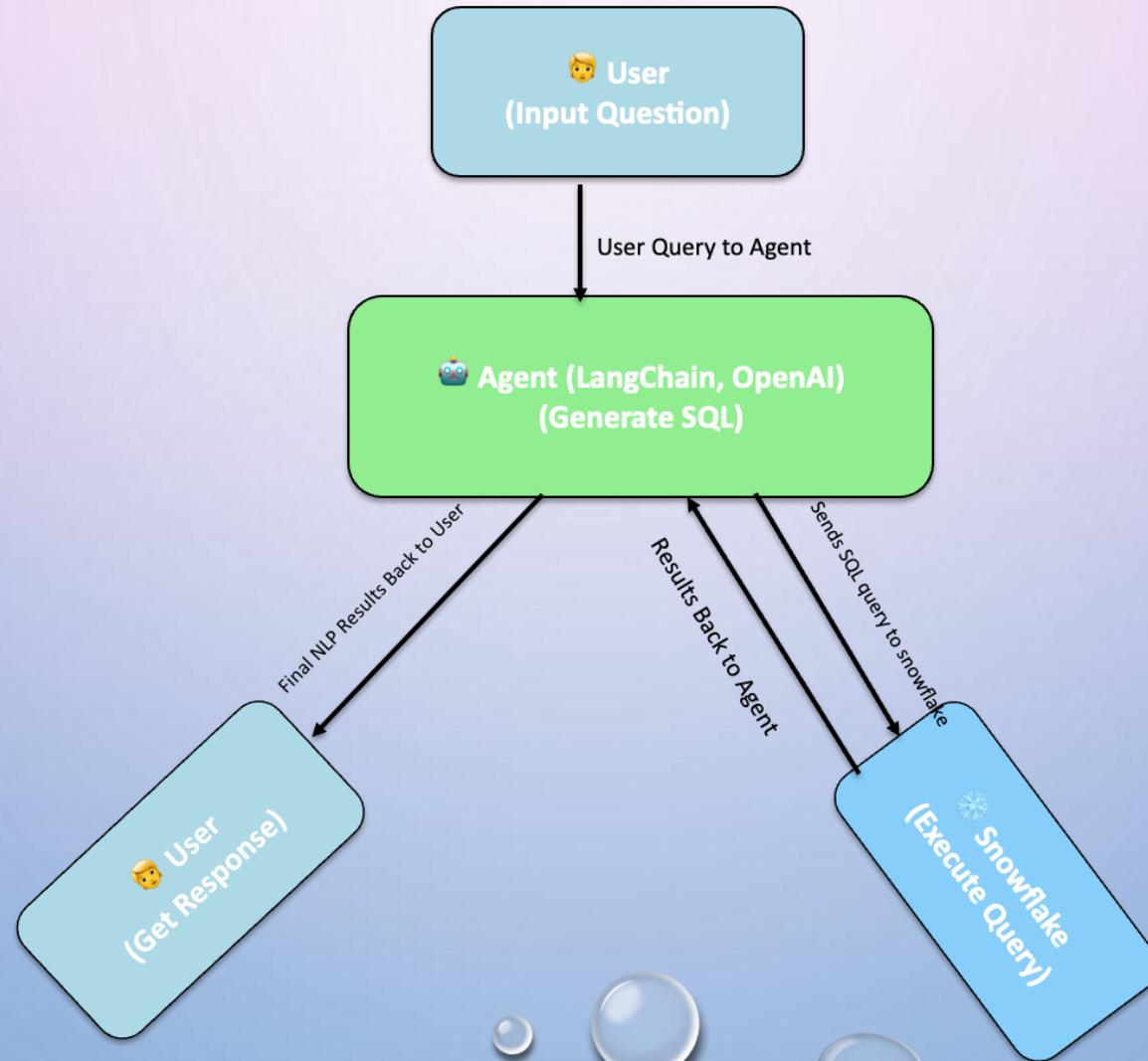
## Active Traders

Active traders can use Stock Buddy's real-time analysis and predictive capabilities to spot trends and make fast, informed decisions in a fast-paced environment.

## Financial Advisors

Financial advisors can leverage Stock Buddy's capabilities to provide their clients with data-backed recommendations, market analysis, and personalized investment strategies.

# Generative AI Agent Workflow



# Response Generation

The image shows a Jupyter Notebook interface with two main panes. The top pane displays a code cell in Python 3 (ipykernel) containing:

```
return response

[8]: # Example query
response = agent.run("Show me stocks with RSI > 70 and volume spike last week")
print(response)
```

The bottom pane shows the output of the previous cell, which is a list of stock tickers and their details:

```
> Finished chain.
Here are some stocks with an RSI greater than 70 and a volume spike last week:

1. **Ticker:** ROST
   - **Date:** April 10, 2025
   - **RSI:** 70.481
   - **Volume:** 4,736,500 (Previous Volume: 1,698,000)

2. **Ticker:** UNH
   - **Date:** April 4, 2025
   - **RSI:** 84.668
   - **Volume:** 9,019,000 (Previous Volume: 6,457,900)

3. **Ticker:** ORLY
   - **Date:** April 4, 2025
   - **RSI:** 74.07
   - **Volume:** 1,113,600 (Previous Volume: 536,700)

4. **Ticker:** AMT
   - **Date:** April 4, 2025
   - **RSI:** 70.356
   - **Volume:** 9,877,200 (Previous Volume: 5,852,900)

5. **Ticker:** TAP
   - **Date:** April 3, 2025
   - **RSI:** 71.591
   - **Volume:** 4,413,200 (Previous Volume: 2,449,900)

6. **Ticker:** HCA
   - **Date:** April 3, 2025
   - **RSI:** 83.271
   - **Volume:** 1,946,900 (Previous Volume: 1,249,400)

7. **Ticker:** KR
   - **Date:** April 3, 2025
   - **RSI:** 70.531
   - **Volume:** 14,958,000 (Previous Volume: 5,909,600)
```

The right side of the interface lists the generated responses:

8. \*\*Ticker:\*\* MSI
  - \*\*Date:\*\* April 3, 2025
  - \*\*RSI:\*\* 73.127
  - \*\*Volume:\*\* 1,497,700 (Previous Volume: 822,700)
9. \*\*Ticker:\*\* WM
  - \*\*Date:\*\* April 3, 2025
  - \*\*RSI:\*\* 89.474
  - \*\*Volume:\*\* 2,216,100 (Previous Volume: 1,411,300)
10. \*\*Ticker:\*\* CMS
  - \*\*Date:\*\* April 3, 2025
  - \*\*RSI:\*\* 73.167
  - \*\*Volume:\*\* 3,543,300 (Previous Volume: 2,172,000)

A note at the bottom right states: "These stocks have shown significant trading volume increases along with high RSI values, indicating strong buying interest and potentially overbought conditions."

At the very bottom, a status bar shows: "Python 3 (ipykernel) | Idle" and "Fully initialized".

# Response Generation

```
# Example query
response = agent.run("tell me which sector do i invest")
print(response)
```

## > Finished chain.

Based on the database schema and available tables, here are a few options for investing your money:

1. **Stock Market**: You can invest in stocks listed on the NYSE. The database contains detailed stock data including historical prices, volume, and technical indicators which can help in making informed decisions. Tables like `nyse\_stock\_data`, `stock\_data\_cleaned\_divya`, `stock\_final`, and `stock\_final\_cleaned` provide extensive data on stock performance.
2. **S&P 500 Companies**: Investing in companies that are part of the S&P 500 index can be a safer bet as these are large and stable companies. The `companies\_sp500` table lists companies in the S&P 500, including their sectors and sub-industries, which can help you choose a diversified portfolio.
3. **Technical Analysis**: If you are interested in technical analysis, tables like `stock\_data\_cleaned\_divya` and `stock\_final\_cleaned` offer various technical indicators such as Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Bollinger Bands, and more. These indicators can help predict future market movements based on historical data.
4. **Diversified Sectors**: The `companies\_sp500` table also categorizes companies by sectors such as Industrials, Health Care, etc. Investing across different sectors can help in risk management.

If you need more specific recommendations or data analysis from these options, I can run queries to provide detailed insights or trends based on historical data.

# Response Generation

```
# Example query
response = agent.run("can you tell, based on previous day performance, which company is likely to fall in stock market?")
print(response)
```

Invoking: `sql\_db\_query` with `{'query': 'SELECT ticker, date, close, relative\_strength\_idx FROM stock\_final\_cleaned WHERE date = (SELECT MAX(date) FROM stock\_final\_cleaned) AND relative\_strength\_idx > 70 ORDER BY close - LAG(close, 1) OVER (PARTITION BY ticker ORDER BY date) DESC LIMIT 10'}`

responded: To predict which company's stock is likely to fall based on the previous day's performance, I can use indicators such as the Relative Strength Index (RSI) and the closing price movement. A high RSI (typically above 70) might indicate that a stock is overbought and could potentially fall, while a significant drop in the closing price from the previous day could also be a sign of potential decline.

I will query the `stock\_final\_cleaned` table to find stocks with a high RSI and a decrease in the closing price from the previous day. Let's execute this query.

[('UNH', datetime.date(2025, 4, 15), 582.01, 74.124), ('ROST', datetime.date(2025, 4, 15), 140.67, 70.369)]Based on the previous day's performance, the following companies are likely to see a fall in their stock prices:

1. \*\*UnitedHealth Group Incorporated (UNH)\*\* - With a closing price of \$582.01 and a Relative Strength Index (RSI) of 74.124, indicating it may be overbought.
2. \*\*Ross Stores, Inc. (ROST)\*\* - With a closing price of \$140.67 and an RSI of 70.369, also suggesting it might be overbought.

These indicators suggest that these stocks could potentially decline, as they are currently considered overbought.

> Finished chain.

Based on the previous day's performance, the following companies are likely to see a fall in their stock prices:

1. \*\*UnitedHealth Group Incorporated (UNH)\*\* - With a closing price of \$582.01 and a Relative Strength Index (RSI) of 74.124, indicating it may be overbought.
2. \*\*Ross Stores, Inc. (ROST)\*\* - With a closing price of \$140.67 and an RSI of 70.369, also suggesting it might be overbought.

These indicators suggest that these stocks could potentially decline, as they are currently considered overbought.

# Response Generation

```
# Example query
```

```
response = agent.run("based on few technical indicators analyze apple and microsoft stocks, and suggest if i should buy or sell"
print(response)|
```



> Finished chain.

Here's the analysis of Apple (AAPL) and Microsoft (MSFT) stocks based on the latest data:

### Apple (AAPL)

- \*\*Date\*\*: 2025-04-15
- \*\*Closing Price\*\*: \$199.80
- \*\*Relative Strength Index (RSI)\*\*: 39.044
  - An RSI below 40 suggests the stock is potentially oversold, indicating a buying opportunity.
- \*\*Moving Average Convergence Divergence (MACD)\*\*: -9.24
  - A negative MACD indicates a bearish trend, suggesting caution.
- \*\*Bollinger Bands\*\*:
  - Lower Band: \$166.988
  - Upper Band: \$239.522
  - The closing price is closer to the lower band, which might indicate a potential upward movement.

### Microsoft (MSFT)

- \*\*Date\*\*: 2025-04-15
- \*\*Closing Price\*\*: \$384.16
- \*\*Relative Strength Index (RSI)\*\*: 47.838
  - An RSI near 50 suggests the stock is neither overbought nor oversold, indicating a neutral position.
- \*\*Moving Average Convergence Divergence (MACD)\*\*: -4.26
  - A negative MACD indicates a bearish trend, suggesting caution.
- \*\*Bollinger Bands\*\*:
  - Lower Band: \$346.968
  - Upper Band: \$402.487
  - The closing price is within the bands, suggesting stability but leaning towards the lower band, which might indicate a potential upward movement.

### Recommendations:

- \*\*Apple (AAPL)\*\*: Consider buying as the indicators suggest the stock might be undervalued and could potentially rebound.
- \*\*Microsoft (MSFT)\*\*: Hold or monitor closely. The indicators suggest a neutral to slightly bearish position, so it might be prudent to wait for more bullish signals before buying.

Please note that these recommendations are based on technical analysis and it's important to consider other factors such as market conditions, news, and fundamental analysis before making investment decisions.

# DETAILED SUMMARY OF WORK & FINDINGS

- Collected historical stock data using the **yfinance** Python library (503 tickers).
- Built an **ETL pipeline** to extract, clean, and store stock data efficiently (Parquet format).
- Used **Amazon SageMaker Studio** and **Snowflake** for data processing and storage.
- Engineered features using **technical indicators** (SMA, EMA, MACD, RSI, etc.).
- Ensured **data quality** through missing value imputation and volume checks.
- Analyzed **price action, volume trends, and momentum indicators**.
- Created **interactive stock analysis dashboards** using **Power BI** (price, volume, MACD, RSI trends).
- Visualized **support/resistance levels** and **Fibonacci retracements** using Power BI.
- Built a **base LSTM model** to predict stock closing prices.
- Achieved **base LSTM model accuracy**:  $R^2$  score of **0.64** with RMSE of **7.32**.
- Identified limitations with real-world stock data (noise, overfitting risks).
- Developed a **Generative Adversarial Network (GAN)** to synthesize stock sequences.
- Fine-tuned the LSTM using **GAN-generated synthetic data**.
- Achieved **fine-tuned model accuracy**:  $R^2$  score of **0.9980** and MAE of **0.4961**.
- Proved that **GAN + LSTM** combination boosts prediction reliability and robustness.
- Built a **Gen-AI agent "Stock Buddy"** for smart investment insights.
- Stock Buddy **automates technical analysis** and gives real-time stock recommendations.
- Addressed pain points like **information overload and emotional trading**.
- Designed use cases for **novice investors, active traders, and financial advisors**.
- Delivered a complete **AI-driven framework** from data ingestion, analysis, to prediction.

# QUICK SUMMARY OF OUR WORK AND KEY FINDINGS

- Extracted and cleaned stock data from **Yahoo Finance** using **Python** and **Snowflake**.
- Engineered stock indicators like **SMA, EMA, MACD, RSI** for better modeling.
- Built a **baseline LSTM model** achieving **R<sup>2</sup> score of 0.64**.
- Identified stock market data challenges like **noise and overfitting**.
- Created **synthetic stock data** using **GAN (Generative Adversarial Network)**.
- Fine-tuned LSTM using GAN data, achieving a **R<sup>2</sup> score of 0.9980**.
- Developed the "**Stock Buddy**" **Gen-AI agent** for real-time investment support.
- Simplified complex trading concepts for **novice and experienced investors**.
- Automated technical analysis and reduced emotional bias in trading.
- Delivered a full **AI-driven trading strategy framework** from data to prediction.
- Created interactive stock analysis dashboards using Power BI (price, volume, MACD, RSI trends).

**Power BI Dashboard Visualizations Link**



**THANK YOU**