NVIDIA (NVDA) Stock Price and Volatility Analysis

Project Objective: To analyze and model the behavior of NVIDIA's (NVDA) Adjusted Closing price, focusing specifically on price forecasting and volatility estimation.

---

1. Data Loading and Preprocessing

The process begins by loading the nvidia-2025-09-04T19-54_export.csv file into a pandas DataFrame. Key initial preprocessing steps include:

- **Indexing:** The Unnamed: 1 column is renamed to Date and converted to datetime format to serve as the temporal index.
- **Renaming:** Column NVDA.4 is renamed to Adj Close for clarity, representing the adjusted closing price.
- **Exploratory Visualization:** Plotting the full Adj Close time series (2000–2025) reveals a clear long-term bullish trend and increasing volatility, particularly in recent years.
- **Seasonal Decomposition:** A decomposition analysis shows a strong trend, low-magnitude seasonality (with a 252-day period, equivalent to a trading year), and residuals that grow over time, indicating **heteroskedasticity**.
- **Stationarity Testing:** Autocorrelation (ACF) and Partial Autocorrelation (PACF) functions confirm the original series is non-stationary. This is verified by the **Augmented Dickey-Fuller (ADF) Test** (p-value = 1.0000).

---

2. Price Modeling: ARMA and ARIMA

To achieve stationarity, differencing is applied. The models are evaluated across two segments: the last trading year (252 observations) and the last trading month (21 observations).

Annual Analysis (Last 252 Observations)

- **Stationarity:** The annual series exhibits a trend. Applying a first difference (d=1) makes the series stationary, confirmed by an ADF test (p-value = 0.0000) and ACF/PACF plots.
- **Model Selection:** I tested **ARMA(1,0,1)** on the differenced series and **ARIMA(1,1,1)** on the original series.

- **Diagnostics:** Residual analysis (distribution plots, QQ-plots, and correlograms) and the **Ljung-Box test** indicate that the residuals are white noise, confirming the models capture the underlying structure.
- **Evaluation:** Using **TimeSeriesSplit** (rolling window), the ARIMA(1,1,1) achieved a **MAPE of 3.09%**. The ARMA(1,0,1) showed a misleadingly high MAPE (101.98%), a common issue when metrics deal with values close to zero.
- **Forecasting:** Forecasts were generated for the last 5 historical values and 10 future steps. **SARIMA** was discarded due to the lack of significant seasonal patterns in daily stock data.

Monthly Analysis

- The same methodology was applied to the monthly segment. The **ARIMA(1,1,1)** model outperformed with a **MAPE of 1.64%**.

---

3. Volatility Modeling: GARCH

To address the inherent risk in stock prices, a **GARCH (Generalized Autoregressive Conditional Heteroskedasticity)** model was implemented.

- **Log Returns:** Closing prices were transformed into log returns and scaled by 100. This stabilizes variance and prevents numerical issues during optimization.
- **Structural Break:** Analysis of the 1-year rolling volatility revealed two distinct "eras":
    1. **Old Era (Pre-2018):** Characterized by high instability.
    2. **Modern Era (Post-2018):** A different risk profile influenced by NVIDIA's growth in AI and Data Centers.
- **Model Fit:** A **GARCH(1,1)** model with a **Student’s t-distribution** (to account for "fat tails" in finance) was fitted to the "Modern Era" data.
- **Walk-Forward Validation:** To simulate a real-world scenario, a walk-forward validation was performed over the last 252 days.
    - **MAE:** 1.8456
    - **RMSE:** 2.4689
- **The MAPE Paradox:** The extremely high MAPE (595.99%) in volatility prediction is noted as a metric limitation. When actual returns are near zero, even small

prediction deviations result in massive percentage errors; however, visual inspection confirms the model tracks volatility trends effectively.