📝 Apple Stock Price Forecasting using Time Series Analysis

A Deep Dive into AAPL Stock Trends from 2012–2019

A Deep Dive into the World Of Time Series Analysis

👨‍🏫👩‍💻👨‍💼Project Team & Mentorship

Mentor

- Mr. Dilawar Basha (Sir)

Team Members

- Karedla Ganesh Reddy
- Amrutha A S
- Hritik Vivek Patil
- Nikhil Talapaneni
- O LOKESH
- Lokesh Laxman Sohanda
- Nisha Ashish Wandile

🧠Project Objective

Analyze historical Apple (AAPL) stock data (2012–2019) using time series models.

Forecast stock prices and derive business insights for strategic decision-making.

Data Analysis

Utilize historical AAPL data.

Model Selection

Linear Regression, SARIMAX, XGBoost, SVR, Random Forest, etc.

Strategic Insights

Informed decision-making, Improve Stock Management, Improved Business Functioning, etc.

Dataset Details

AAPL stock prices from Jan 3, 2012, to Dec 30, 2019.

No missing values or duplicates.

Data Source

Apple (AAPL) stock prices

Time Span : Jan 3, 2012 – Dec 30, 2019

Key Columns

- Date, Open, High, Low
- Close, Adj Close, Volume
- Target : Close Column

Data Highlights

- No missing values/duplicates
- Close Price: 55.79 to 291.52
- Volume: 11.36M to 376.53M
- Date converted to datetime

	Date	Open	High	Low	Close	Adj Close	Volume
0	03-01-2012	58.485714	58.928570	58.428570	58.747143	50.765709	75555200
1	04-01-2012	58.571430	59.240002	58.468571	59.062859	51.038536	65005500
2	05-01-2012	59.278572	59.792858	58.952858	59.718571	51.605175	67817400
3	06-01-2012	59.967144	60.392857	59.888573	60.342857	52.144630	79573200
4	09-01-2012	60.785713	61.107143	60.192856	60.247143	52.061932	98506100
...	...	...	...	...	...	...	...
2006	23-12-2019	280.529999	284.250000	280.369995	284.000000	282.054138	24643000
2007	24-12-2019	284.690002	284.890015	282.920013	284.269989	282.322266	12119700
2008	26-12-2019	284.820007	289.980011	284.700012	289.910004	287.923645	23280300
2009	27-12-2019	291.119995	293.970001	288.119995	289.799988	287.814392	36566500
2010	30-12-2019	289.459991	292.690002	285.220001	291.519989	289.522614	36028600

2011 rows × 7 columns

🔧 Exploratory Data Analysis (EDA)

EDA revealed key trends and characteristics of the AAPL stock data, guiding our model selection.

Mostly Upward Trend

2012–2019 stock price.

Volume Peaks

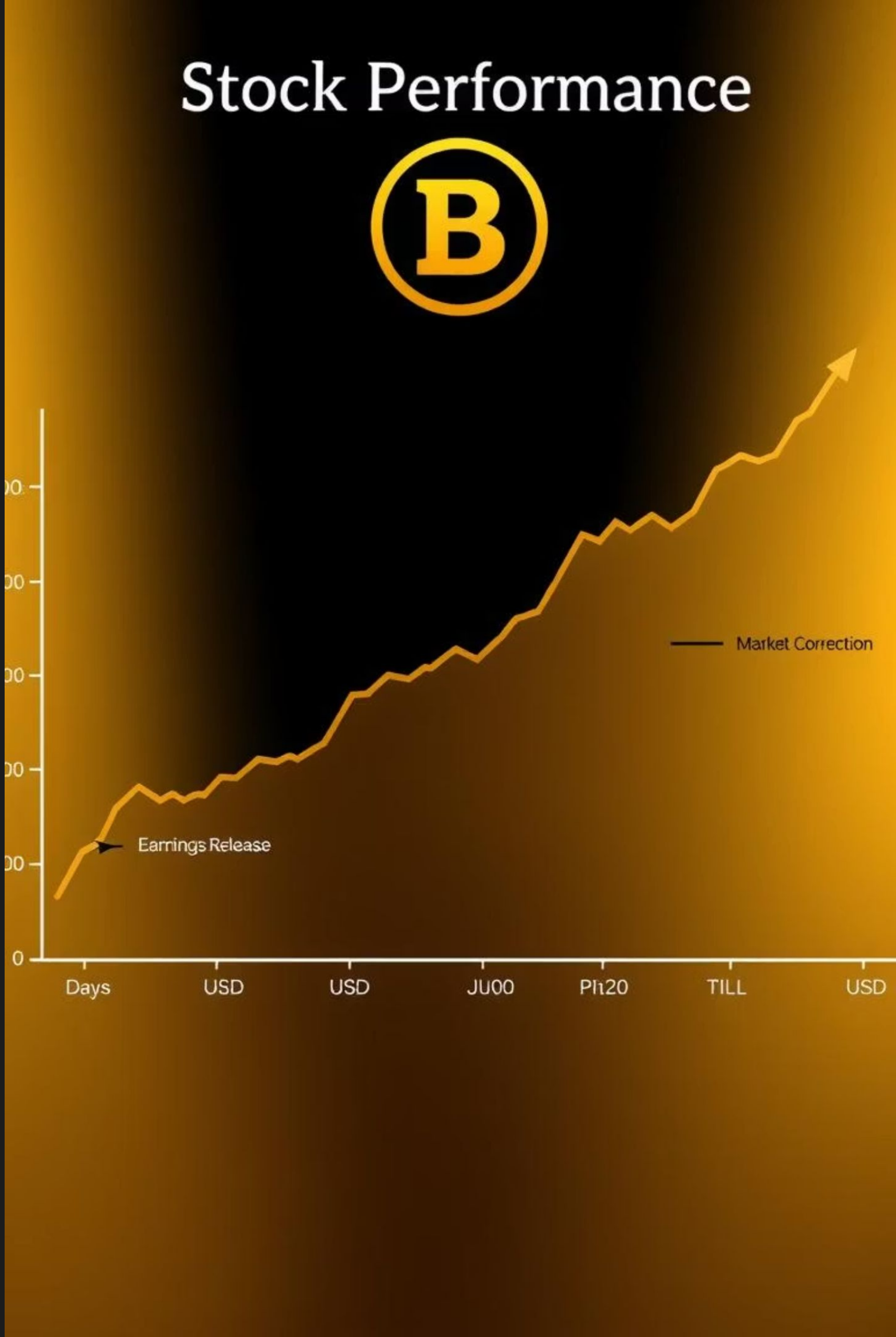
Indicating investor activity.

Non-Stationarity

Inconstant Rolling mean & standard deviation.

ADF Test

Constant Rolling mean & standard deviation, p-value improved to 0.0.



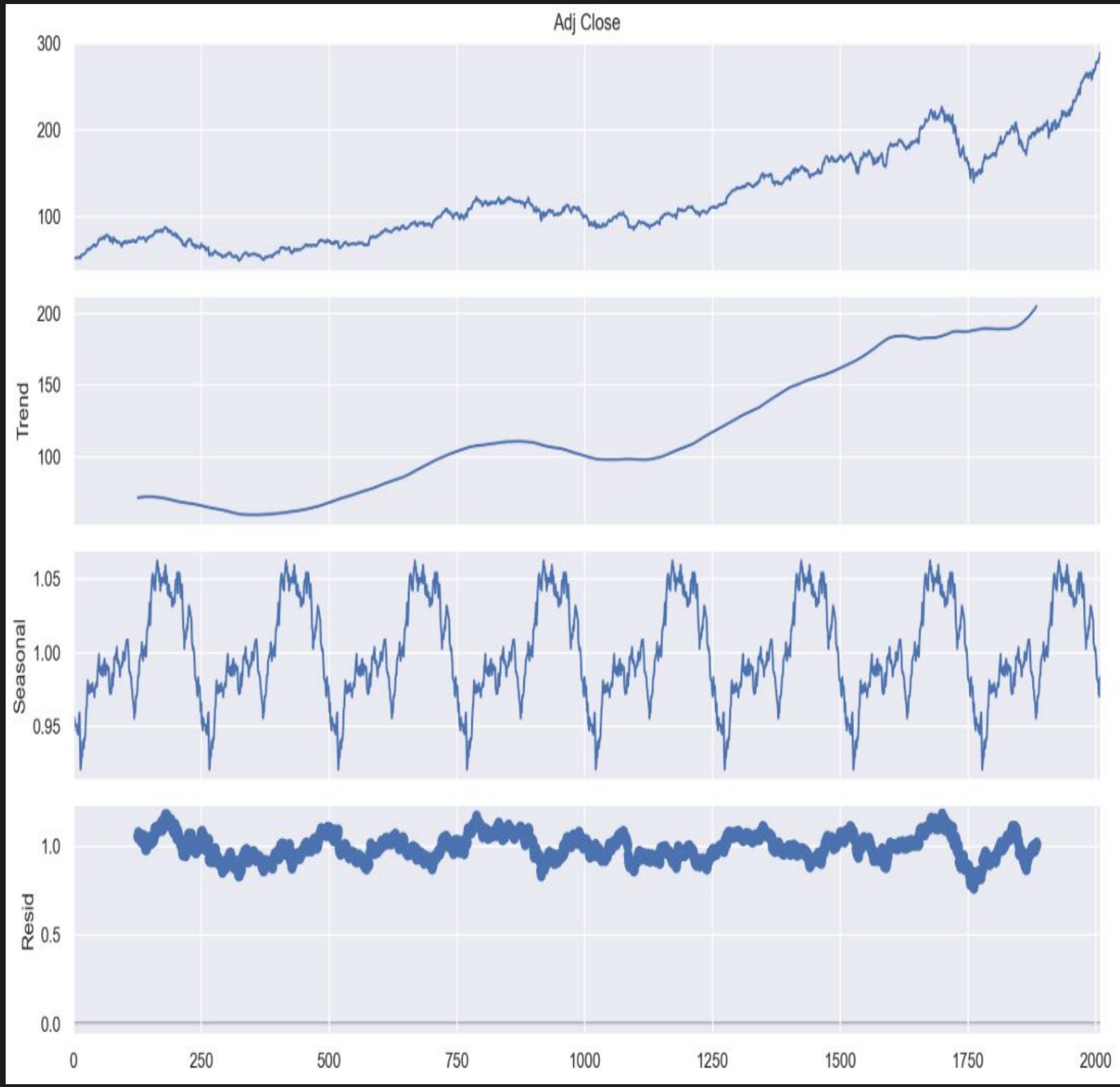
🔧 EDA Process

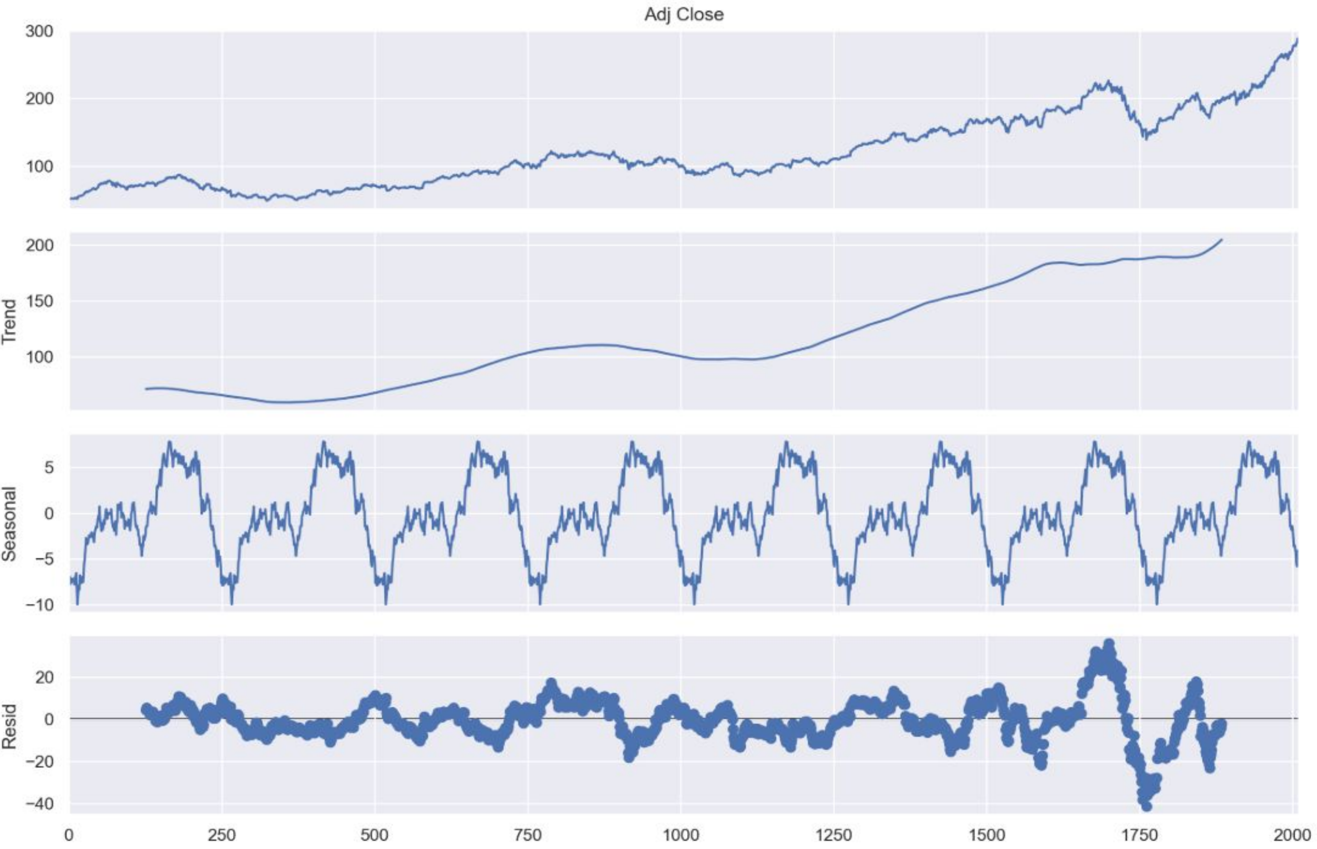
Descriptive Statistics

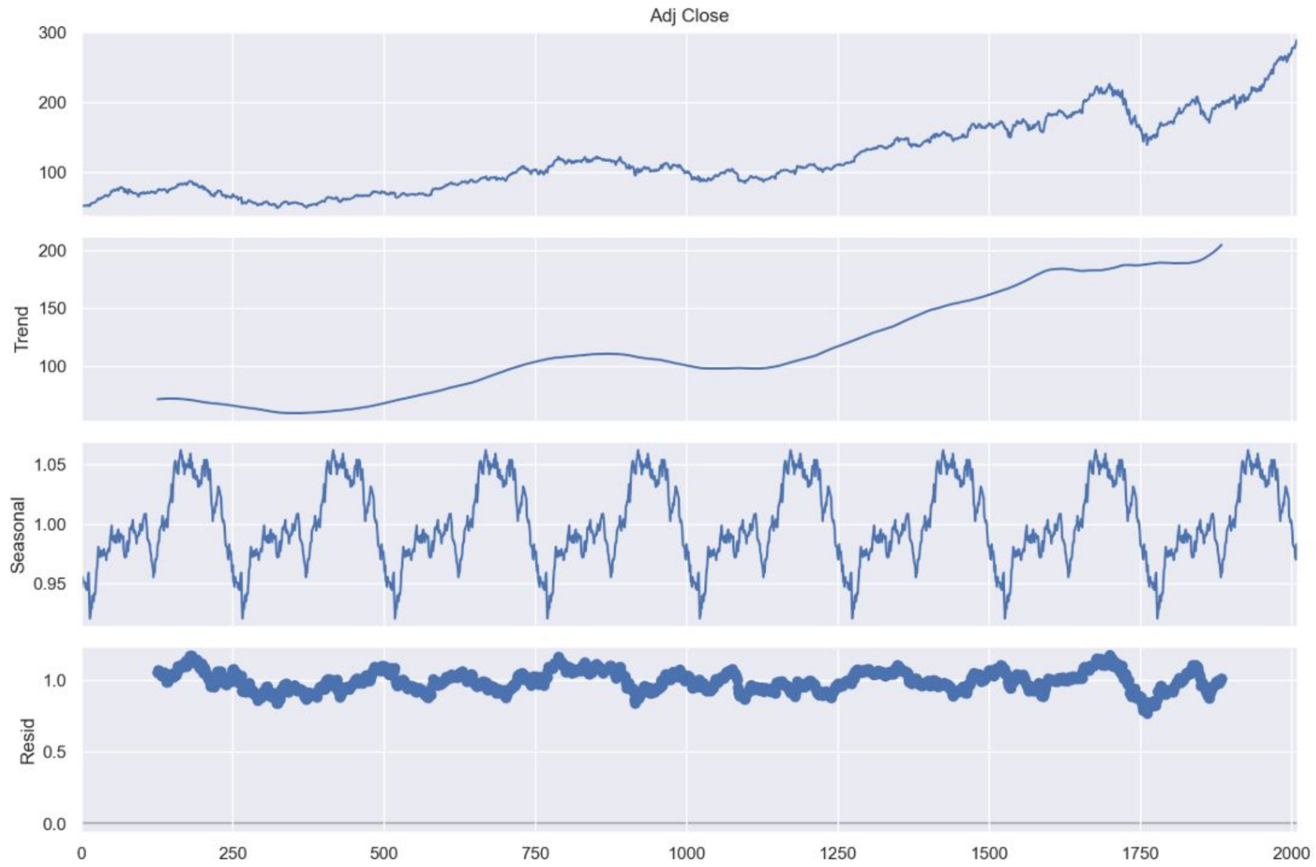
- Checked the Quartile values & checked the info of the data.
- Checked for missing values.
- Checked for duplicates.
- Converted Date column's format to DateTime.

Data Visualizations

- Histogram (for skewness)
- Heatmap (for co-relations)
- Scatter Plots (for trends between 2 columns)
- Pairplots (for correlations & outliers)
- Line Plots (for trends)
- Seasonal Decompose Chart (for decomposition & knowing core components)
- & more...







📈 Graph Interpretations: Seasonal Decompose Chart

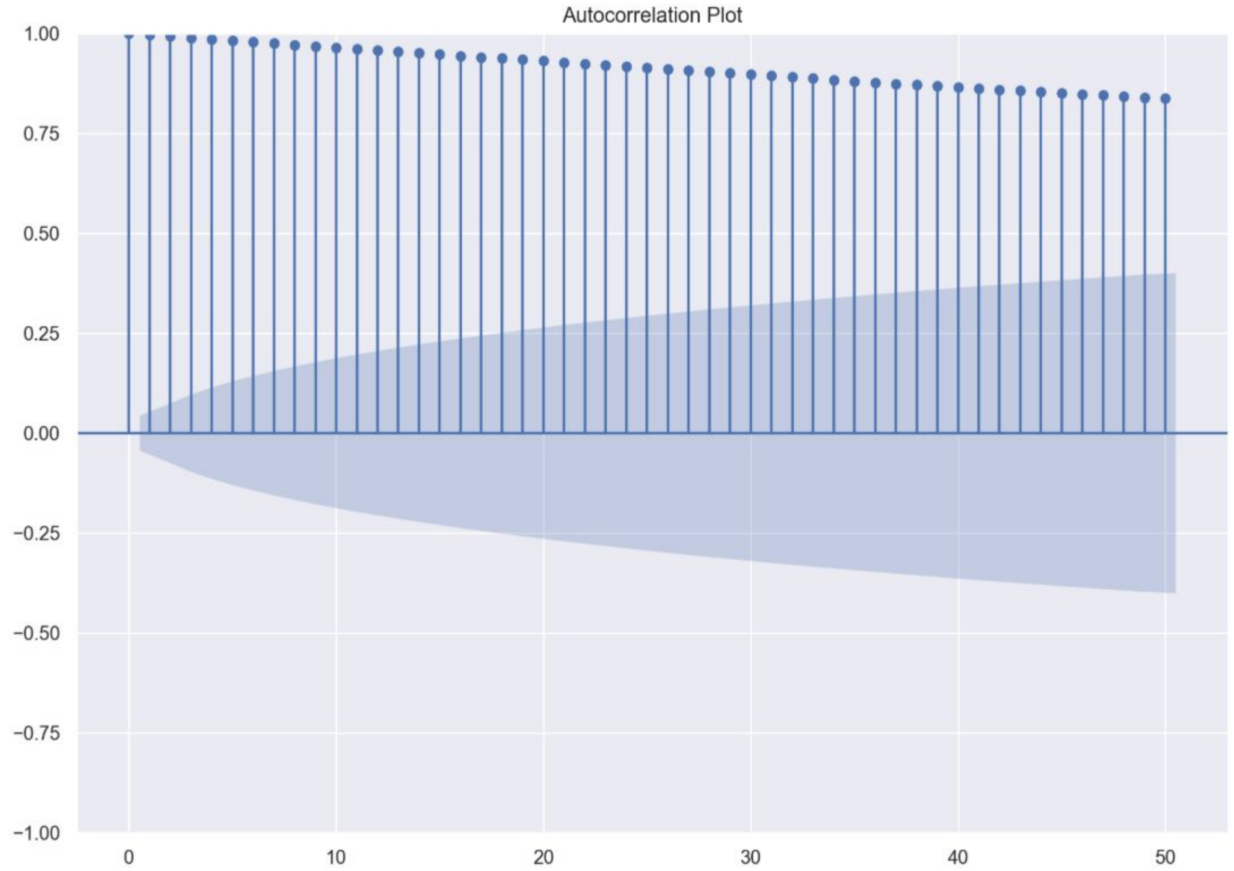
1. Observed Data (Adj Close)
   - The top subplot represents the adjusted closing price of Apple stock over time.
   - The trend suggests **fluctuations with possible growth**, alongside recurring patterns.

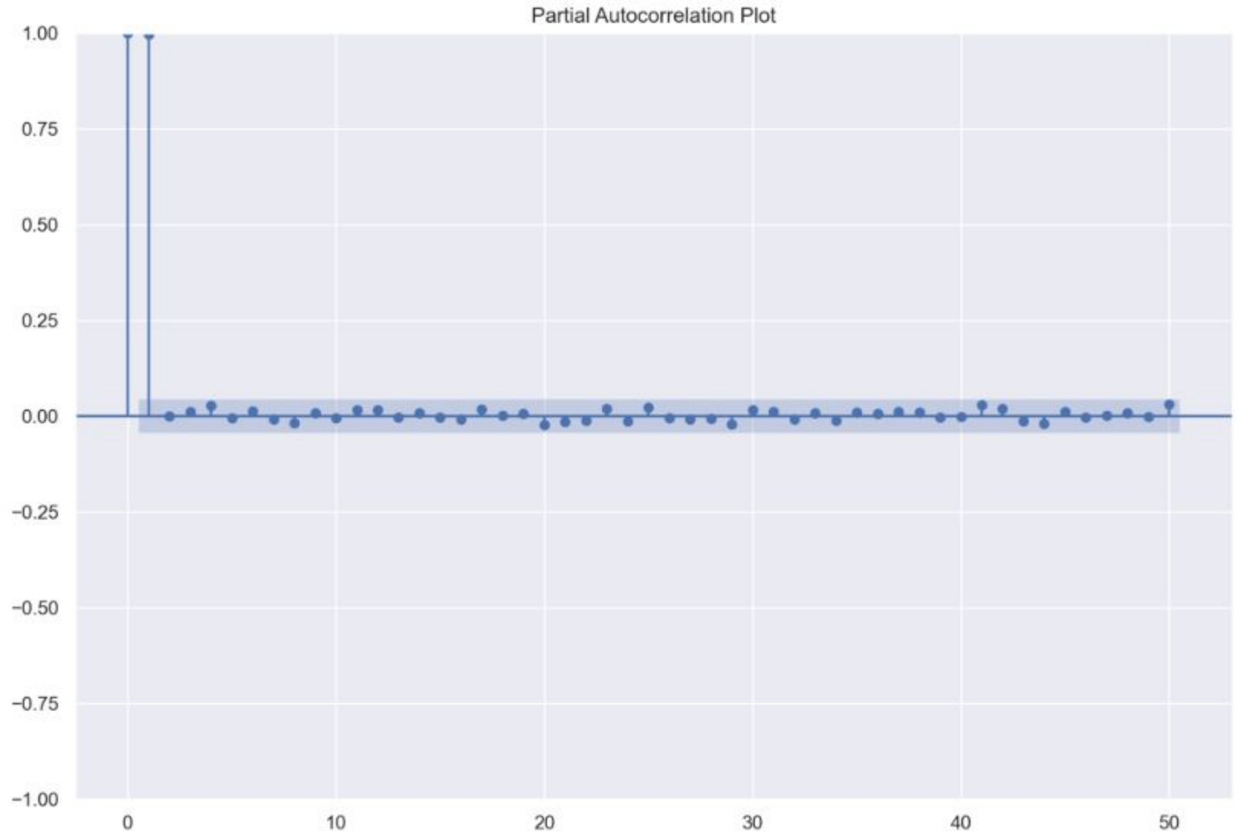
2. Trend Component
   - The second subplot showcases the **long-term movement** in stock prices.
   - It smooths out short-term fluctuations, revealing an overall **upward trend towards later years.

3. Seasonal Component
   - The third subplot captures the **recurring short-term cycles** in stock prices.
   - Peaks and troughs indicate periodic fluctuations, possibly due to **market trends, earnings reports, or macroeconomic factors**.

4. Residual Component
   - The bottom subplot visualizes the **remaining variations** after removing the trend and seasonal influences.
   - The residuals appear centered around zero, signifying **random noise or external events affecting price movements**.

Insights
- **Trend analysis** can help predict future stock performance.
- **Seasonality detection** may inform better entry or exit points for trades.
- **Residual component analysis** can highlight anomalies or external shocks.



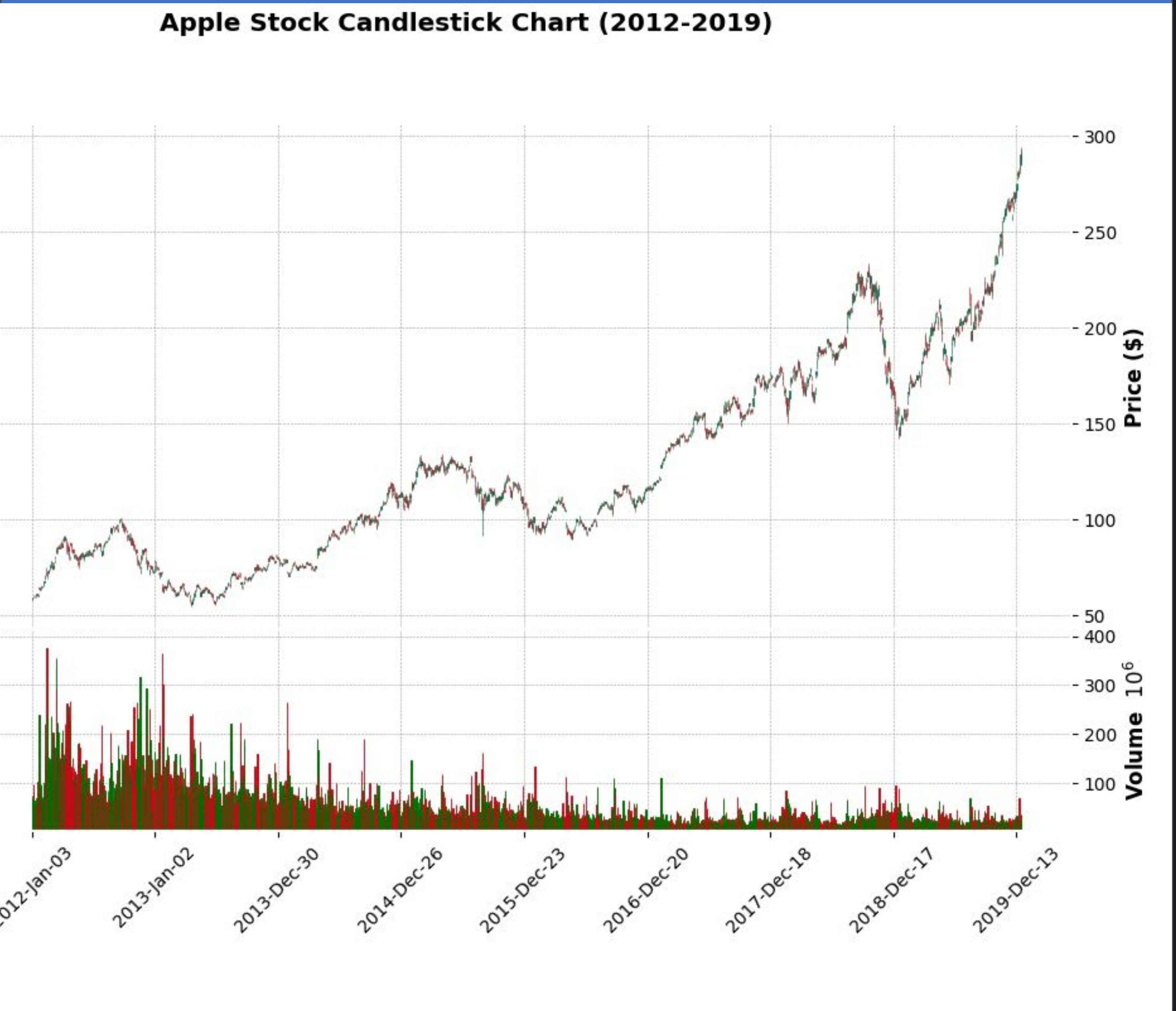


📈 Graph Interpretations:

ACF Plot - Interpretation

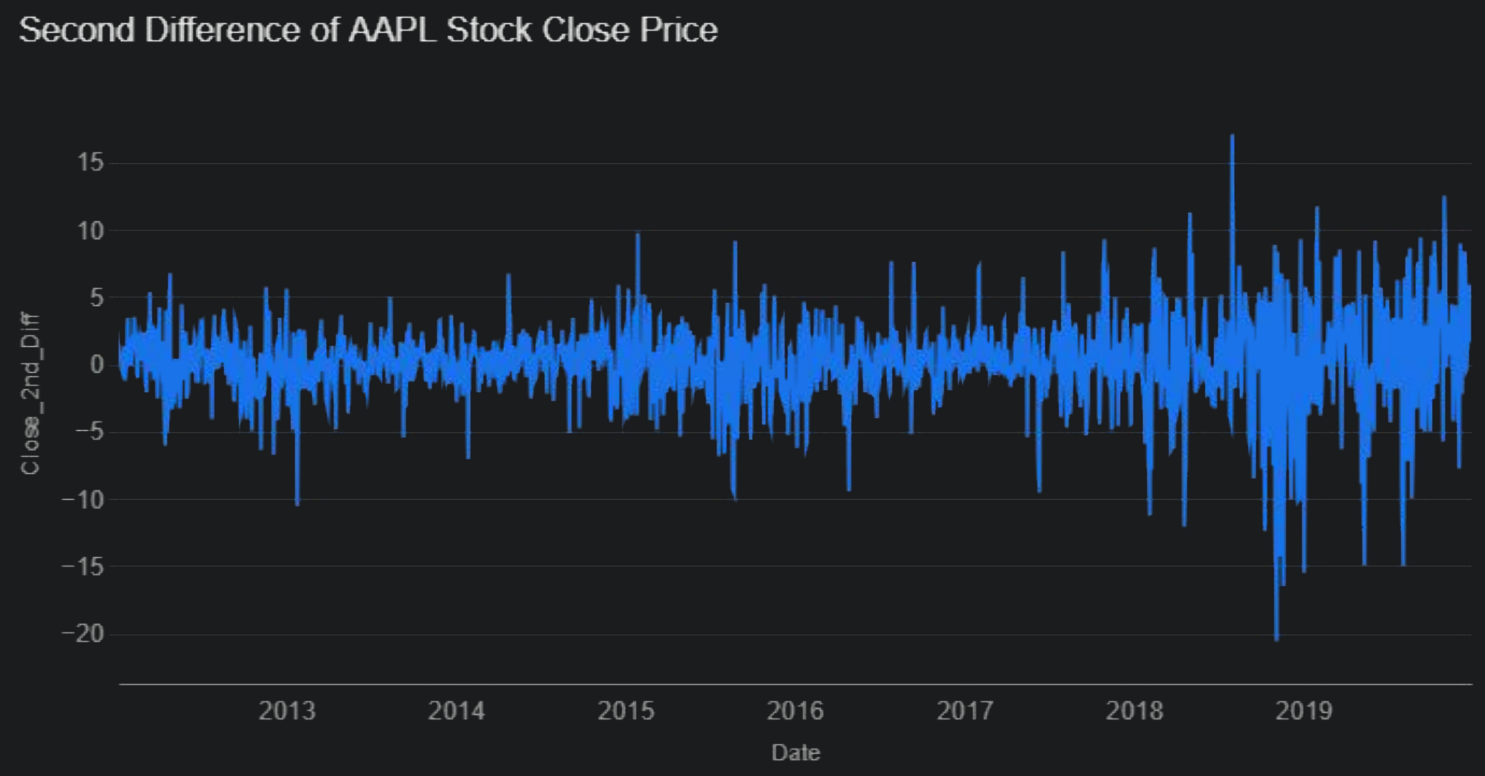
- Autocorrelation Function (ACF): The plot represents how past values of the dataset influence future values.
- Y-Axis (Autocorrelation Values): Ranges from -1 to 1, showing the correlation strength.
- X-Axis (Lag Values): Indicates different time lags.
- Blue Vertical Lines: Represent autocorrelation values at different lags.
- Shaded Confidence Interval: Helps determine statistical significance of autocorrelations.
- Observation: Autocorrelation values remain high, suggesting strong correlation over multiple time lags.
- Potential Insight: The dataset may exhibit a trend or seasonality, requiring differencing or transformation for modeling.

PACF Plot - Interpretation
- Partial Autocorrelation Function (PACF): Measures the direct relationship between a time series and its past values while removing the effect of intermediate lags.
- Y-Axis (Partial Autocorrelation Values): Ranges from -1 to 1, showing the strength of correlation.
- X-Axis (Lag Values): Indicates the different time lags.
- Significant Spikes: High values at lag 1 and lag 2 suggest strong partial autocorrelation at these lags.
- Near-Zero Values for Other Lags: Implies that further lags do not significantly contribute to the model once the primary lags are accounted for.
- Application: Helps in identifying relevant lag values for AR models in time series forecasting.
- Modeling Insight: This pattern suggests an autoregressive (AR) structure—typically useful for ARIMA modeling.



📈 Graph Interpretation

This candlestick chart highlights Apple's stock performance from 2012 to 2019, showing a consistent long-term upward trend with periodic fluctuations.
Each candlestick captures the stock's open, close, high, and low prices over time, while the volume bars below—green for gains, red for losses—reveal peaks in investor activity, often coinciding with major market events.
The rising price trajectory reflects increasing investor confidence and sustained growth, making it a strong visual cue for strategic forecasting and investment decisions in your presentation.



Stationarity Tests

The Dataset was not Stationary at first, which was interpreted by performing ADF Test & KPSS Test. Used Differencing Techniques to make the data stationary.

The Augmented Dickey-Fuller (ADF) test results indicate that the time series of Apple (AAPL) stock's closing prices is **non-stationary**.

Here's an interpretation of the results:

- **ADF Statistic (1.219):** This value is positive and much higher than the critical values.
- **p-value (0.996):** The p-value is significantly greater than the typical significance level of 0.05. This means we **fail to reject the null hypothesis**.
- **Critical Values: 1, 5, 10**

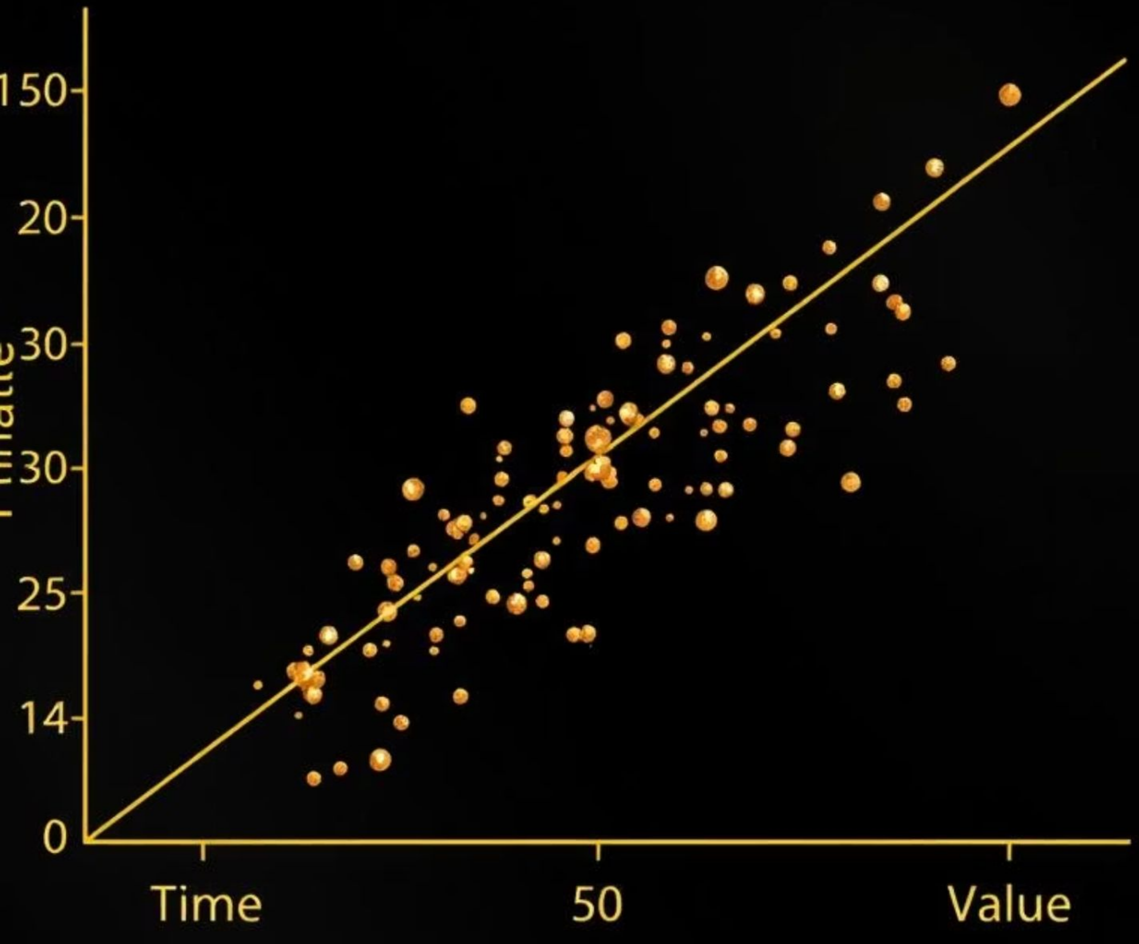
Insight:

Since the p-value (0.996) is greater than 0.05 (or any of the critical values), we conclude that the time series has a unit root and is **non-stationary**. This implies that the statistical properties of the series, such as mean and variance, change over time. For example, stock prices often exhibit trends and are not mean-reverting, which is characteristic of non-stationary processes.

What does this mean for analysis?

- **Spurious Regressions:** If you were to perform regression analysis on non-stationary time series, you might encounter spurious regressions, where variables appear to be related when they are not.
- **Forecasting Models:** Many time series forecasting models (e.g., ARIMA) assume stationarity. To use these models effectively, you would typically need to transform the series to make it stationary (e.g., by differencing).

Post-differencing, the Apple stock price series exhibits stationarity—evidenced by a stabilized mean and variance, and confirmed by the ADF test p-value dropping from 0.996 to 0.0—making the data suitable for reliable time series modeling.



🏗️ Model Building

Model building isn't just about fitting algorithms to data—it's about translating patterns into power, and turning questions into quantifiable insight.

Models Trained

Linear Regression, XGBoost, Random Forest, ARIMA, SARIMAX, etc.

Most Important Predictor Variable

Time as the sole predictor.

Purpose

Generate business insights to support strategic investment and operational planning.

🏗️ Model Building: Linear Regression

Linear Regression is treated as the baseline forecasting model. Simple yet powerful model leveraging linear patterns in time series data. Overall A Good Model amongst all others.

Model

Linear Regression

Preprocessing

Ensured data was clean with no missing values; datetime formatting applied.

Hyperparameters

- None used in the Model Building.

Tuning Method

Performance evaluated without complex hyperparameter tuning due to model simplicity.

🏗️ Model Building: SARIMAX

SARIMAX addresses seasonality and trends, essential for accurate time series forecasting.

Model

SARIMAX (Seasonal ARIMA with Exogenous features).

Preprocessing

Ensured stationarity via differencing.

Hyperparameters

- (p, d, q) = (1, 1, 1)
- (P, D, Q, s) = (1, 1, 0, 12)

Tuning Method

GridSearchCV, TimeSeriesSplit.

🏗️Model Building: XGBoost

XGBoost leverages supervised learning with lag features for robust forecasting, capturing non-linear relationships.

Model Type

XGBoost (Extreme Gradient Boosting).

Approach

Supervised learning with lag features.

Best Parameters

- n_estimators = 1000
- max_depth = 5
- learning_rate = 0.01

Tuning Method

GridSearchCV, TimeSeriesSplit.

Model Evaluation

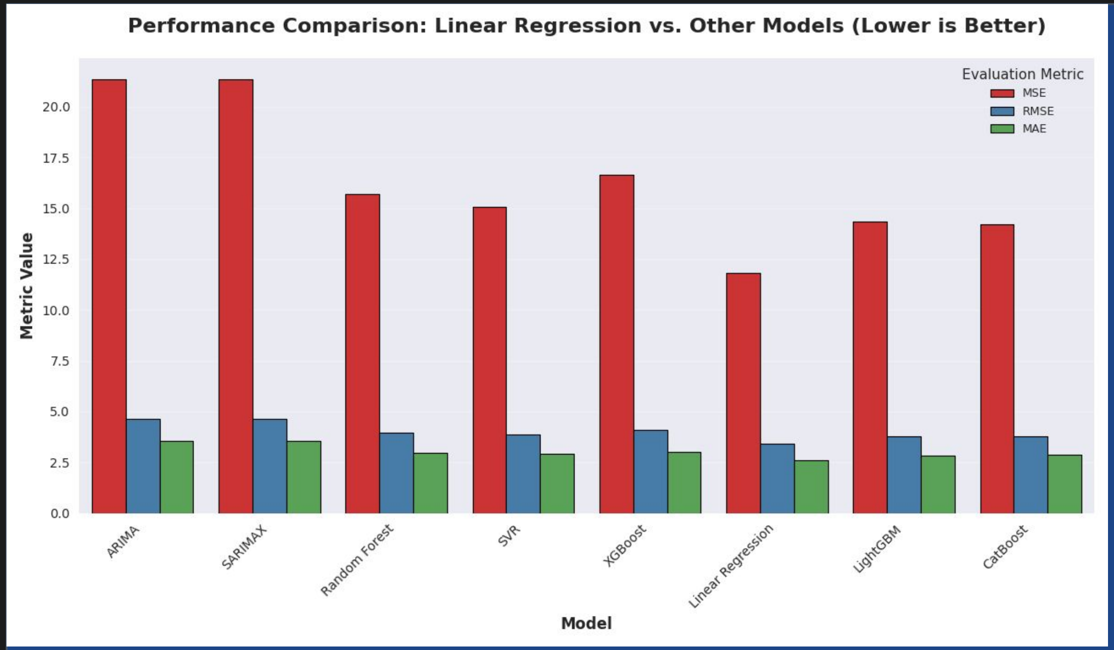
MSE (Mean Squared Error): Measures average squared error—best used when large errors need to be penalized more heavily.

RMSE (Root Mean Squared Error): Square root of MSE—great for interpreting errors in original units of the target variable.

MAE (Mean Absolute Error): Averages absolute errors—ideal when all deviations, regardless of direction, matter equally.

MAPE (Mean Absolute Percentage Error): Expresses error as a percentage—useful for comparing model accuracy across different scales.

Model	MSE	RMSE	MAE	MAPE
—-----------------------------------	—-----------------------------------	—-----------------------------------	—-----------------------------------	—-----------------------------------
ARIMA	21.33	4.61	3.56	103.92
SARIMAX	21.33	4.61	3.56	99.61
Random Forest	15.69	3.96	2.97	140.82
SVR	15.05	3.88	2.91	133.59
XGBoost	16.65	4.08	2.58	140.41
Linear Regression	11.81	3.43	2.58	136.57
LightGBM	14.35	3.78	2.84	138.81
CatBoost	14.23	3.77	2.85	134.74



🎯Final Model: Linear Regression

Why Linear Regression Outperforms Other Models ?

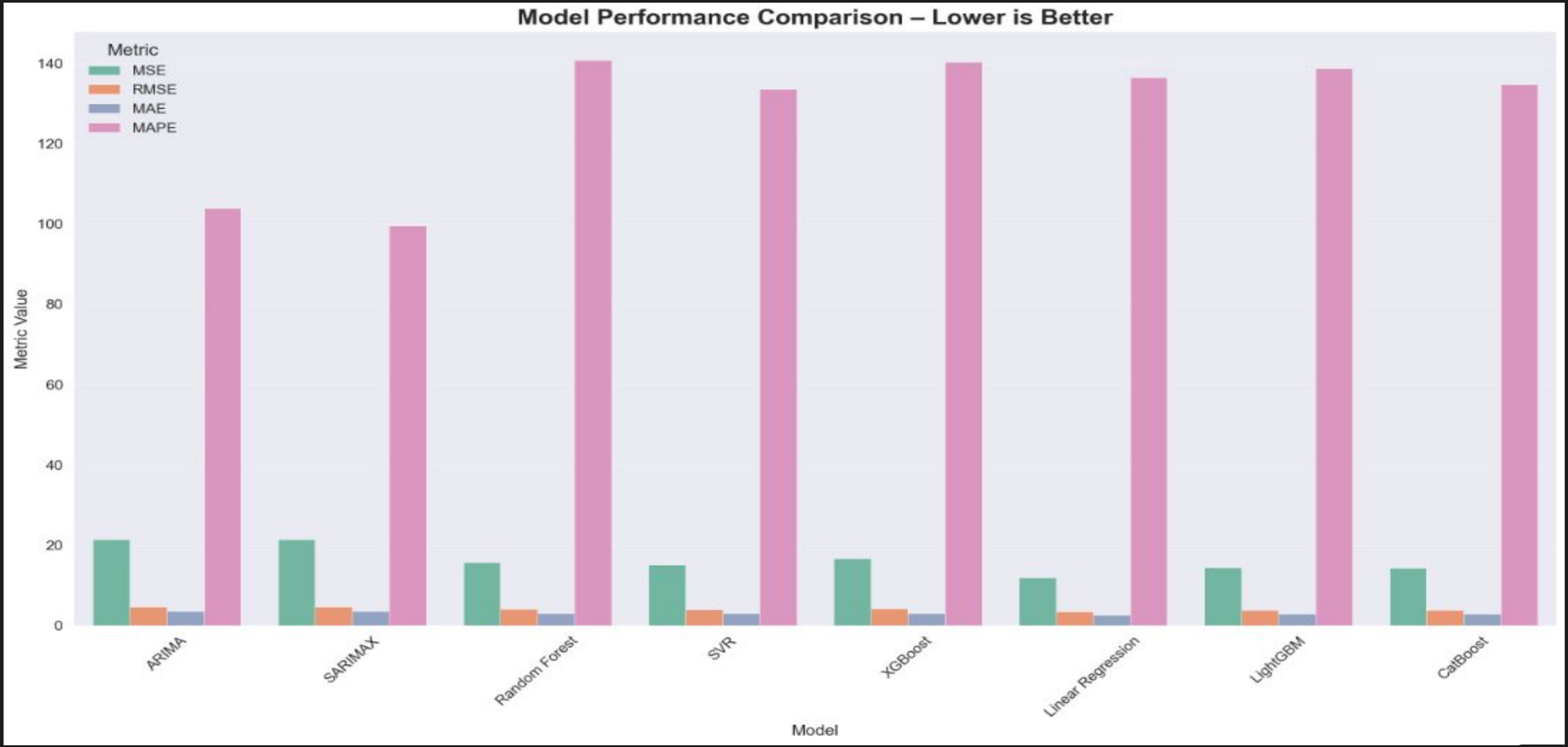
Linear Regression excelled in forecasting Apple stock prices (2012–2019), achieving the lowest MSE (11.8157), RMSE (3.4374), and MAE (2.5888) compared to ARIMA, SARIMAX, Random Forest, SVR, XGBoost, LightGBM, and CatBoost.

Key reasons include:

1. Linear Trend Capture: The dataset's stable, upward price trend aligns with Linear Regression's strength in modeling linear relationships, unlike complex models that may overfit.
2. Model Simplicity: With fewer parameters, Linear Regression avoids the overfitting risks of ensemble methods like XGBoost, ensuring robust generalization.
3. Feature Suitability: Engineered features (e.g., lagged prices, moving averages) likely exhibit linear correlations with the target (Close price), favoring Linear Regression's assumptions.
4. Effective Preprocessing: Stationarity adjustments and scaling optimized the data for linear modeling, reducing the need for complex techniques like SARIMAX.
5. Balanced Error Metrics: Linear Regression consistently minimizes both small and large errors, outperforming SARIMAX, which had a lower MAPE (99.61 vs. 136.57) but higher MSE and RMSE.
6. Recommendation: Linear Regression is ideal for this dataset due to its accuracy, simplicity, and interpretability, making it a reliable choice for stock price forecasting.

✅ Final Model Evaluation

Comparing all models, Linear Regression offered the best overall accuracy, while SARIMAX achieved the lowest MAPE.



🚀 Business Insights & Suggestions

Strategic & Operational Insights

1. Upward Trend for Strategic Planning: Apple's consistent price growth from 2012–2019 reflects strong investor confidence—useful for projecting long-term growth and supporting product expansion strategies.
2. Seasonality Supports Timing: Clear yearly seasonal trends suggest businesses can time product launches, marketing campaigns, and inventory planning to align with favorable market cycles.
3. Volatility Signals Risk Awareness: Analysis of 'High' and 'Low' price ranges allows risk assessment and hedging strategy development to navigate stock price fluctuations confidently.

🚀 Business Insights & Suggestions

Market Activity & Communication Strategy

1. Volume Reflects Market Sentiment: Trading volume spikes indicate major news or earnings events—a cue for businesses to sync announcements with high investor engagement periods.
2. Liquidity & Capital Planning: Understanding periods of high trading activity helps evaluate stock liquidity for institutional investors and plan capital-raising initiatives accordingly.
3. Seasonal Volume Cycles: Monitor cyclical volume trends to schedule marketing rollouts, investor calls, and financial reports effectively.

🚀 Business Insights & Suggestions

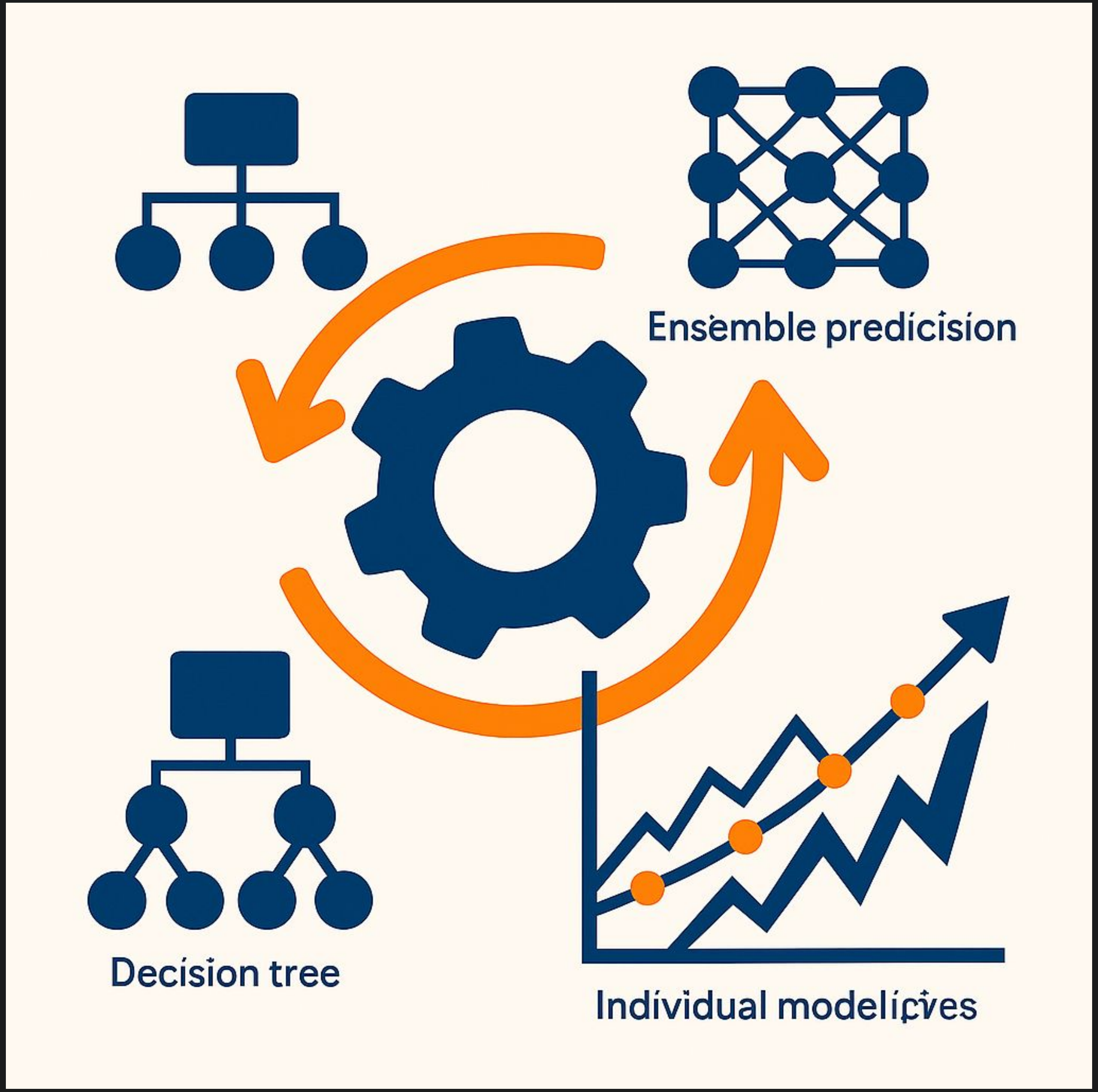
Model-Driven Stock Management

1. Forecasting Supports Decision-Making: Models like Linear Regression and SARIMAX provide predictive insights that support informed buy, sell, or hold decisions.
2. Model Monitoring for Adaptability: Regularly reviewing forecast performance ensures agility—switch models as markets evolve.
3. Pursue Continuous Optimization: Combine models or adopt ensemble approaches to refine forecast accuracy and enable more resilient, data-driven planning.

🚀 Business Insights & Suggestions

Continuous Model Improvement

1. Model Comparison for Growth: Benchmark performance across SARIMAX, XGBoost, and Linear Regression to identify improvement opportunities.
2. Explore Ensemble Techniques: Combine strengths of multiple models for more stable and accurate predictions, especially in volatile market conditions.
3. Commit to Iterative Tuning: Integrate feedback loops for retraining and fine-tuning models based on real-world deviations and performance drift.



🚀 Deployment Overview

Apple Stock Deployment

The Streamlit application leverages Linear Regression to deliver accurate and user-friendly 30-day forecasts for Apple stock prices (2011–2020).
Achieving the lowest MSE (11.8157), RMSE (3.4374), and MAE (2.5888) among competing models, Linear Regression effectively captures the dataset's linear trends, benefiting from robust feature engineering (lagged prices) and preprocessing (standard scaling).
The app's intuitive interface, featuring a candlestick chart, predicted price plot, and strategic insights, empowers users to select custom date ranges and make informed investment decisions based on forecasted trends, volatility, and price changes.
This solution combines simplicity, accuracy, and interactivity, making it a reliable tool for stock price forecasting.
The interface snapshot on the next slide showcases a robust forecasting dashboard for Apple stock, combining interactive features like date selection, candlestick visualizations, predictive analytics, and strategic insights—all in one intuitive layout designed to support informed investment decisions.

🚀 Deployment Page (Streamlit)

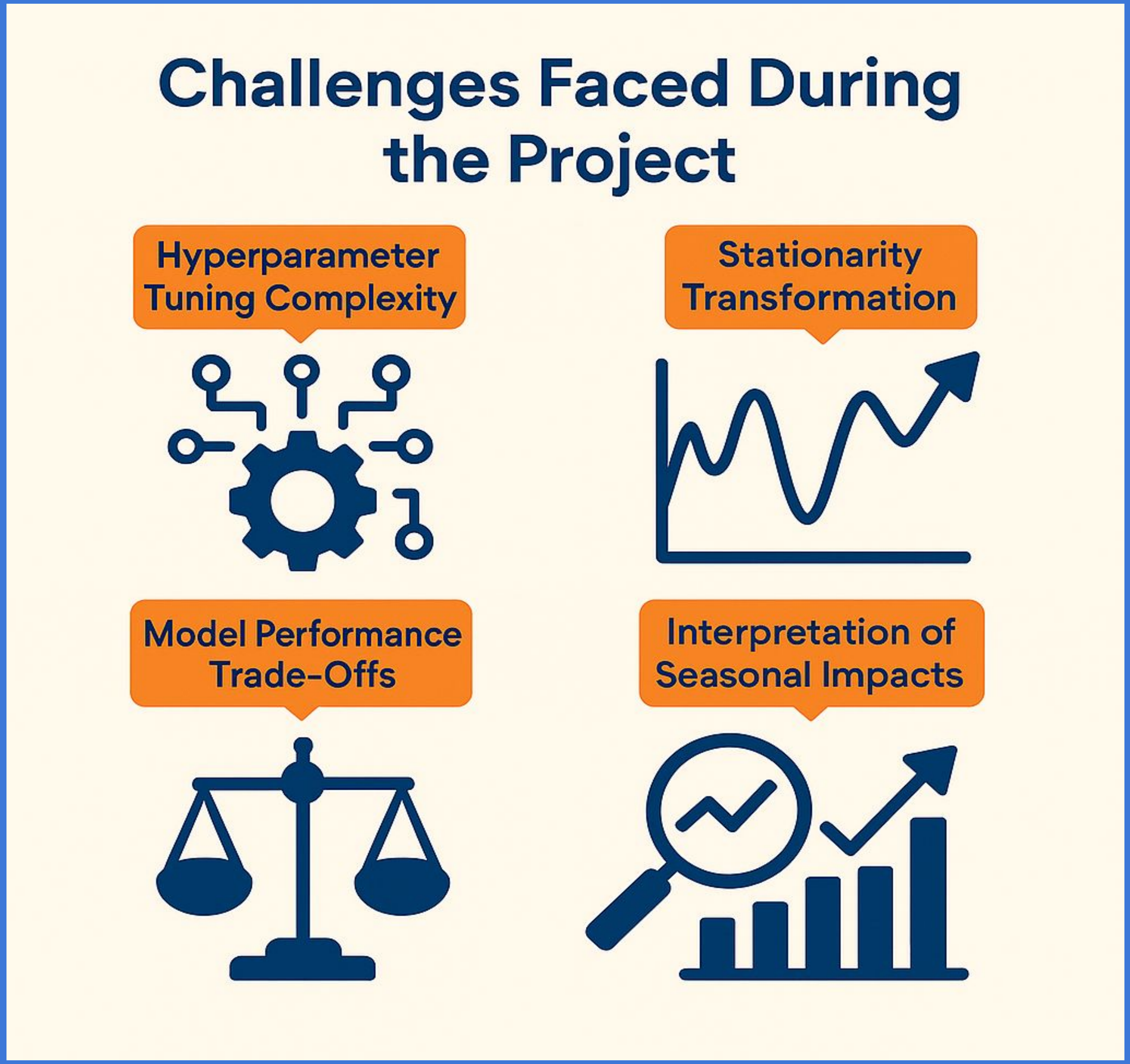
🧗 Challenges Faced

- **Hyperparameter Tuning Complexity:** Fine-tuning parameters for models like SARIMAX, XGBoost, and SVR required extensive experimentation using GridSearchCV and cross-validation, significantly increasing computational time and model iteration cycles.

- **Stationarity Transformation:** The non-stationary nature of the original stock price series demanded multiple transformations (e.g., differencing) and diagnostic tests (ADF Test) to ensure accurate modeling.

- **Model Performance Trade-offs:** While Linear Regression provided the best accuracy metrics, more advanced models like SARIMAX and XGBoost offered advantages in seasonal adaptation or handling nonlinear patterns, highlighting the need to balance simplicity vs. flexibility.

- **Interpretation of Seasonal Impacts:** Isolating and interpreting seasonal trends from noise required decomposition analysis and domain understanding—challenging when multiple factors influence price movement across quarters.



🎯Conclusion & Future Work

Our analysis provides valuable insights for strategic decision-making in stock market investments.

Key Takeaways

- ✅ Linear Regression emerged as the best-performing model, offering the lowest error metrics (MSE, RMSE, MAE) for forecasting Apple stock prices.
- 📈 The dataset exhibited strong linear trends and seasonality, making it well-suited for time series modeling after applying stationarity transformations.
- 🎯 Business insights derived from trend, volume, and seasonality analysis enable informed investment, communication, and risk mitigation strategies.
- 🛠️ Model evaluation balanced accuracy with interpretability, comparing models like SARIMAX, XGBoost, and Random Forest for performance trade-offs.
- 🔁 Continuous improvement through model tuning and ensemble techniques ensures adaptability in real-world market conditions.
- 🚀 Streamlit-based deployment demonstrated practical value, empowering users with an intuitive forecasting tool for strategic decision-making.

Next Steps

- 🔄 Enhance Feature Engineering: Integrate macroeconomic indicators (e.g., interest rates, inflation) and news sentiment scores to enrich the model inputs.
- 🧠 Explore Advanced Models: Investigate deep learning techniques like LSTM or GRU for capturing long-term dependencies and nonlinear dynamics.
- 📡 Deploy Real-Time Forecasting: Set up APIs and automate data pipelines for live prediction and real-time dashboard updates.
- 🧪 Run Scenario Simulations: Develop "what-if" analyses to evaluate potential shocks (e.g., earnings misses, policy changes) on stock performance.
- 🎯 Expand to Multi-Asset Forecasting: Apply the same framework to other stocks or sectors to build a diversified investment insight platform.
- 📊 Incorporate Alternative Data Sources: Leverage sentiment analysis from social media, financial news, or earnings transcripts to enhance model sensitivity to real-world events.


Thank You
FOR
LISTENING