TCS Stock Data

Live and Latest (With Given Dataset)

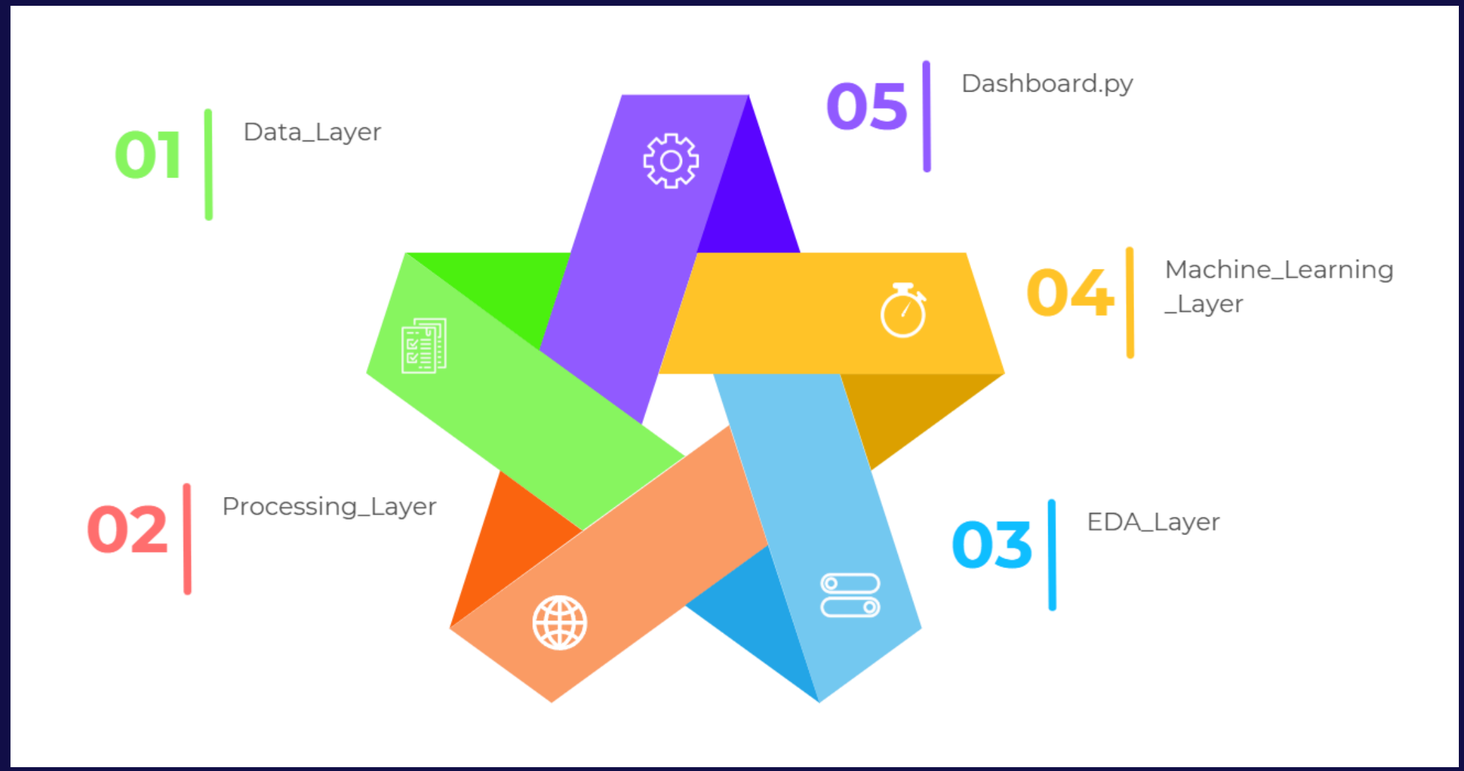
OIKANTIK BASU
Data Science Intern at Unified Mentor

MAY, 2025 @ 09:00 A.M.

OBJECTIVE

The primary objective of this project is to perform a comprehensive analysis of Tata Consultancy Services (TCS) stock by leveraging over two decades of historical financial data. This includes dissecting patterns in price movements, identifying seasonality, detecting technical indicator signals, and uncovering underlying trends across various timeframes. By engineering meaningful features like moving averages, MACD, RSI, and lagged returns, the goal is to understand the behavior of TCS stock during corporate actions such as dividends and stock splits, as well as during high-volume trading periods. Beyond analysis, the project aims to build and compare predictive models including classical machine learning algorithms (Linear Regression, Random Forest, XGBoost), deep learning models like LSTM, and advanced time-series models like Facebook Prophet to forecast future stock prices. These insights can support traders in short-term strategy planning and assist long-term investors in making data-backed decisions by understanding price dynamics, trend momentum, and future projections under uncertainty.

System Architecture Star Diagram



Dataset Overview

TCS_STOCK_HISTORY.CSV

This primary dataset contains historical OHLCV (Open, High, Low, Close, Volume) records of TCS stock, providing the foundation for time-series modeling and technical analysis. It spans multiple years, capturing daily trading activity and enabling trend, volatility, and return-based computations.

TCS_STOCK_ACTION.CSV

This dataset records corporate actions like stock splits, bonus issues, and dividend events. These events often correlate with significant market responses and are crucial for understanding anomalies, spikes in volume, or abrupt price shifts.

TCS_STOCK_INFO.CSV

A metadata file that includes essential reference details about the TCS stock. It contains identifiers, company profile details, and static information which can be useful for labeling, display, and context in the visual dashboard.

Feature Engineering

DAILY RETURN

Calculated as the percentage change in closing price from one day to the next, this feature reveals day-to-day volatility and helps assess market momentum.

MACD & SIGNAL LINE

The Moving Average Convergence Divergence (MACD) and its signal line are momentum indicators used to identify bullish or bearish crossovers, helping detect trend shifts early.

52-WEEK HIGH

Represents the maximum price reached in the past 252 trading days. It's a widely used indicator of investor sentiment and breakout opportunities.

LAG FEATURES

Lagged versions of the closing price (Lag_1, Lag_2) help the models capture short-term memory and temporal dependencies.

CUMULATIVE RETURN

A compound measure of returns over time, showcasing the long-term investment growth trajectory of the stock.

RSI (RELATIVE STRENGTH INDEX)

This oscillator indicates whether a stock is overbought or oversold, providing critical entry/exit signals for traders.

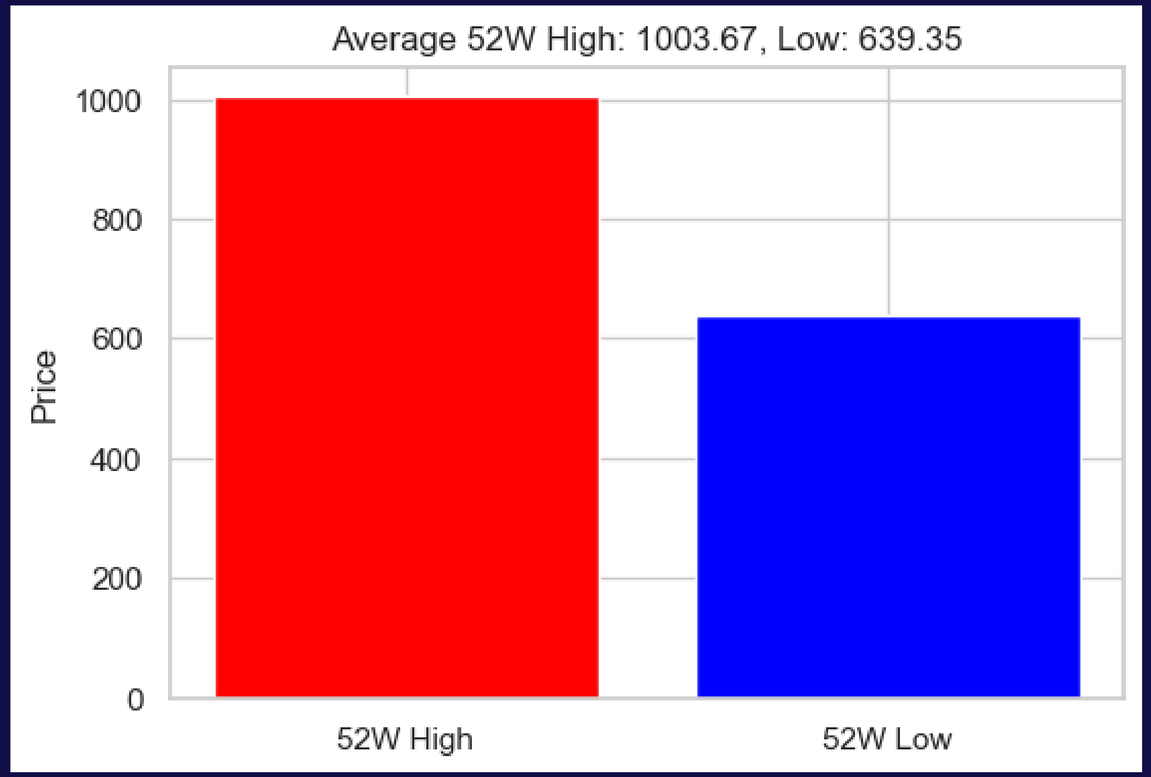
ROLLING STATISTICS

Includes 7-day rolling mean and standard deviation, which smooth out fluctuations and detect short-term trends and volatility.

Exploratory Data Analysis

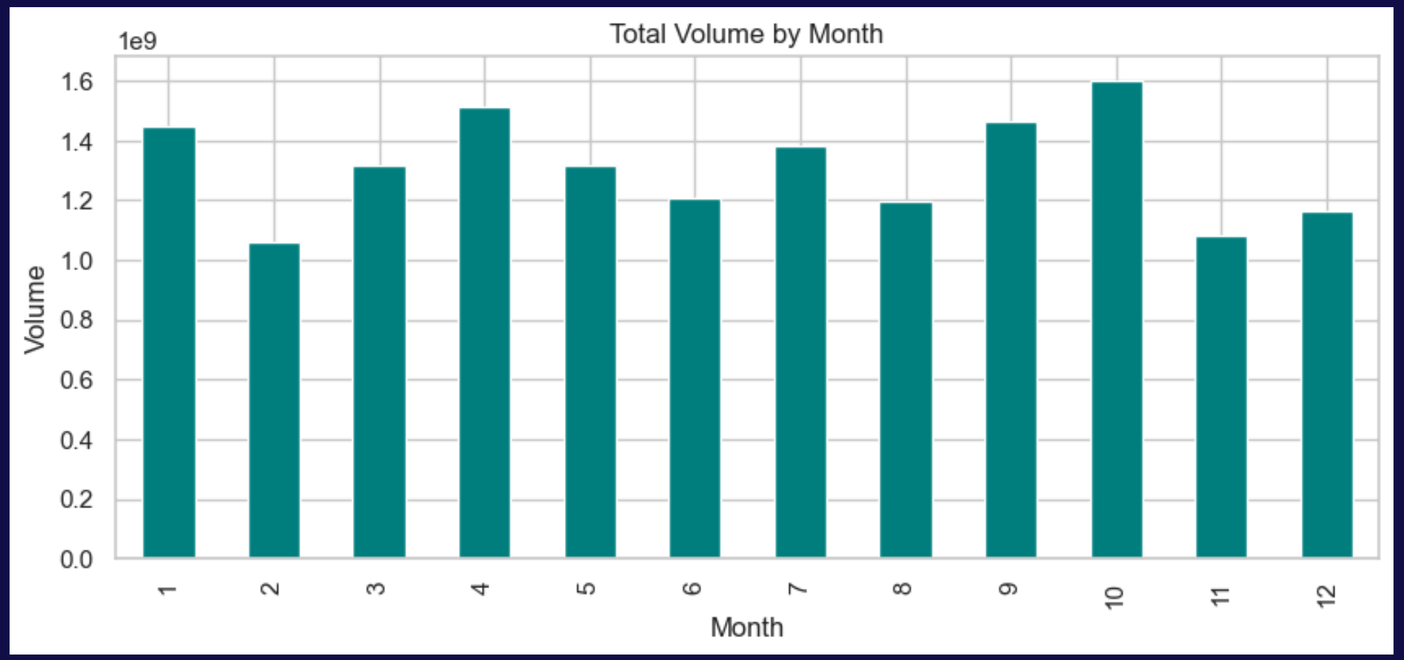
The Exploratory Data Analysis (EDA) provided critical insights into TCS’s stock behavior by visualizing patterns in price, volume, technical indicators, and event-driven movements. Through comparisons like 52-week highs vs lows, monthly volume trends, moving averages, and model-friendly features like MACD and RSI, the analysis revealed long-term growth consistency, short-term volatility triggers, and the stock’s outperformance against sector benchmarks laying a solid foundation for forecasting and strategy development.

Average 52Weeks High vs Low



Average 52W High vs Low

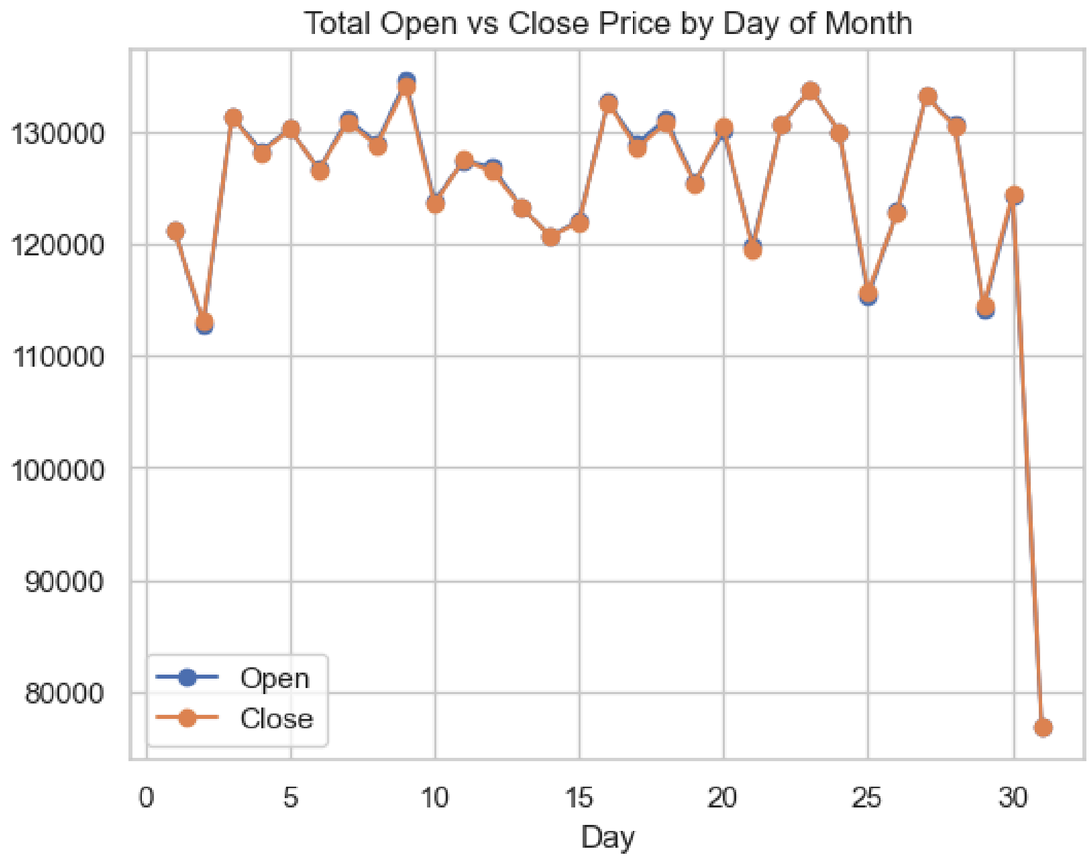
The average 52-week high (1003.67) far exceeds the 52-week low (639.35), highlighting significant price range and volatility. This spread reflects strong investor momentum and potential for long-term bullish growth.



Total Volume by Month

Total Volume by Month

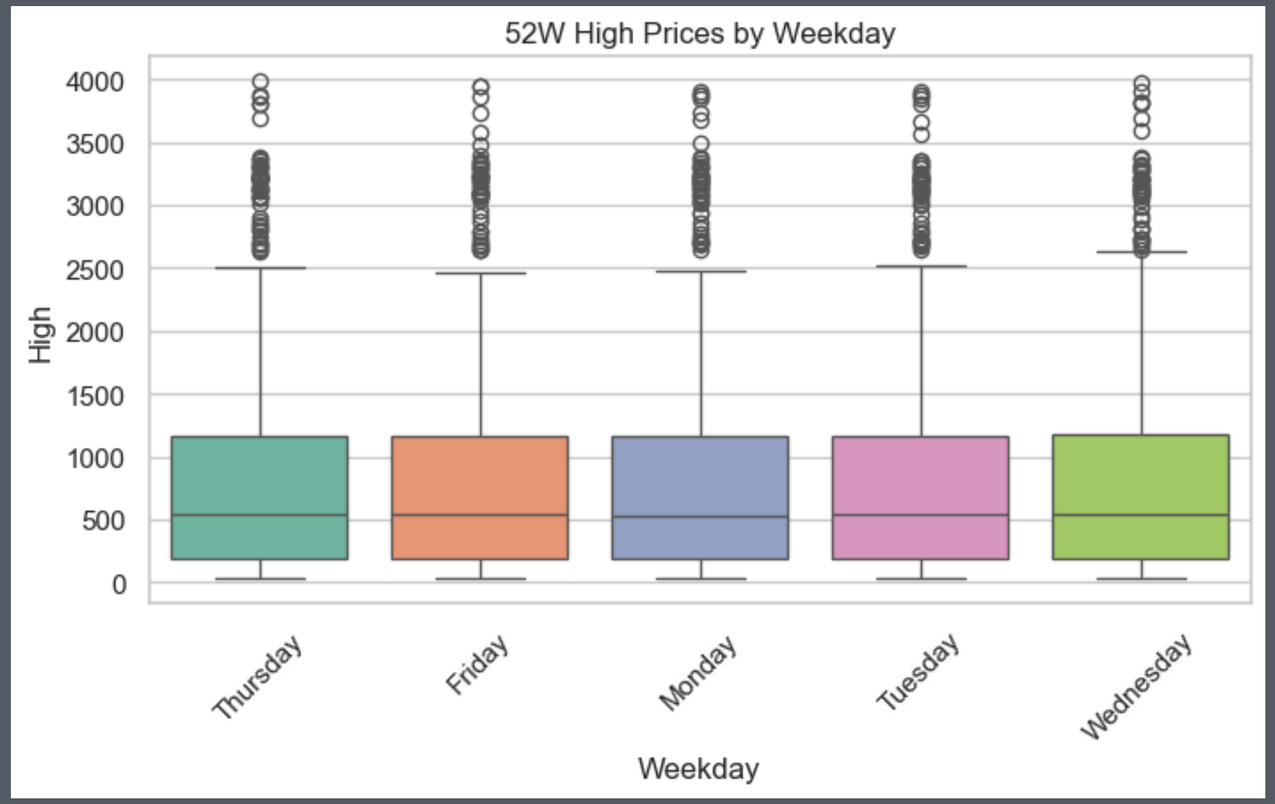
October and April consistently record the highest trading volumes, likely due to earnings announcements, dividends, or fiscal events. In contrast, February and November show lower engagement.



Open vs Close Price by Day of Month

Open vs Close Price by Day of Month

Daily stock movement remains relatively steady across the month. However, a notable dip on the 30th suggests possible month-end profit booking or reduced trading activity.



52Weeks High by Weekday

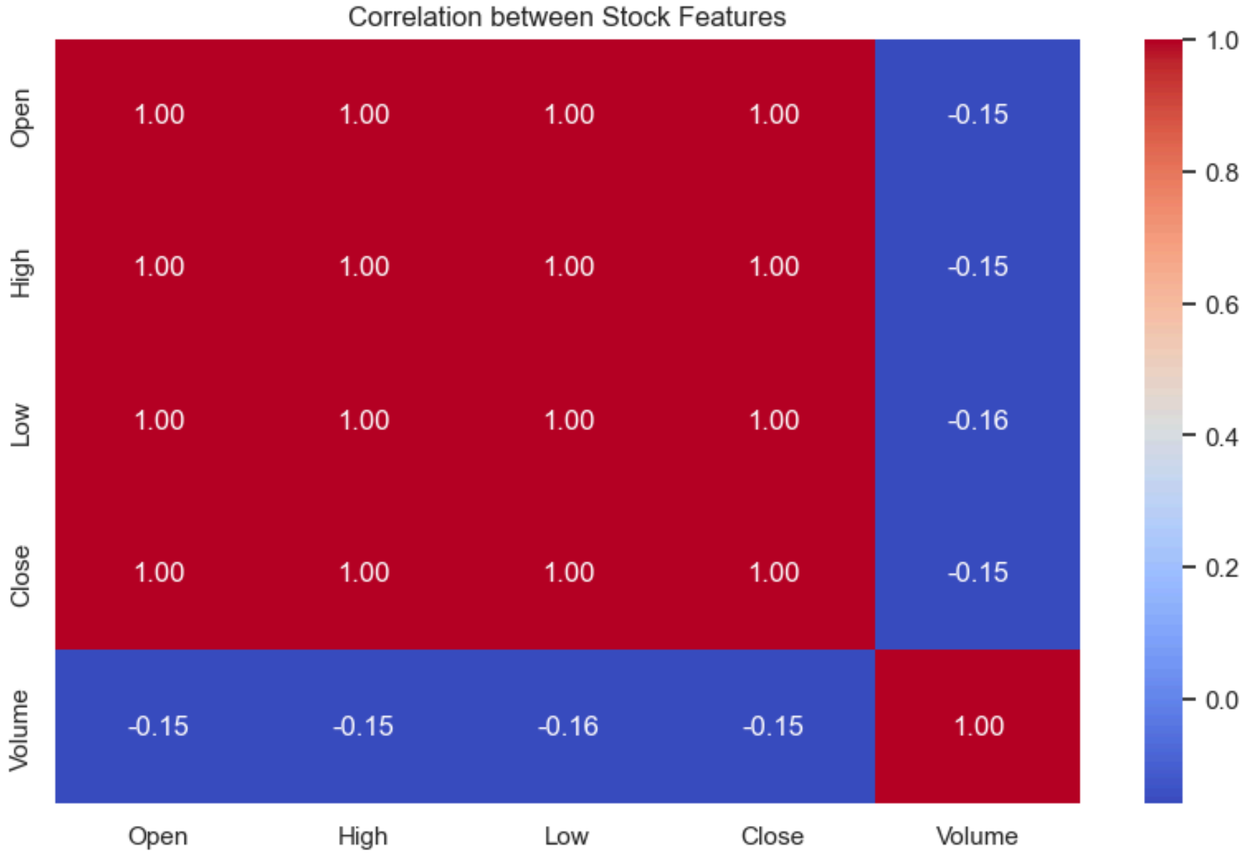
52Weeks High by Weekday

While all weekdays show similar medians, Tuesday and Thursday occasionally produce higher outliers. This may indicate strategic mid-week activity or price breakouts. (Boxplot)

Rolling Averages (MA20 vs MA50)

Rolling Averages (MA20 vs MA50)

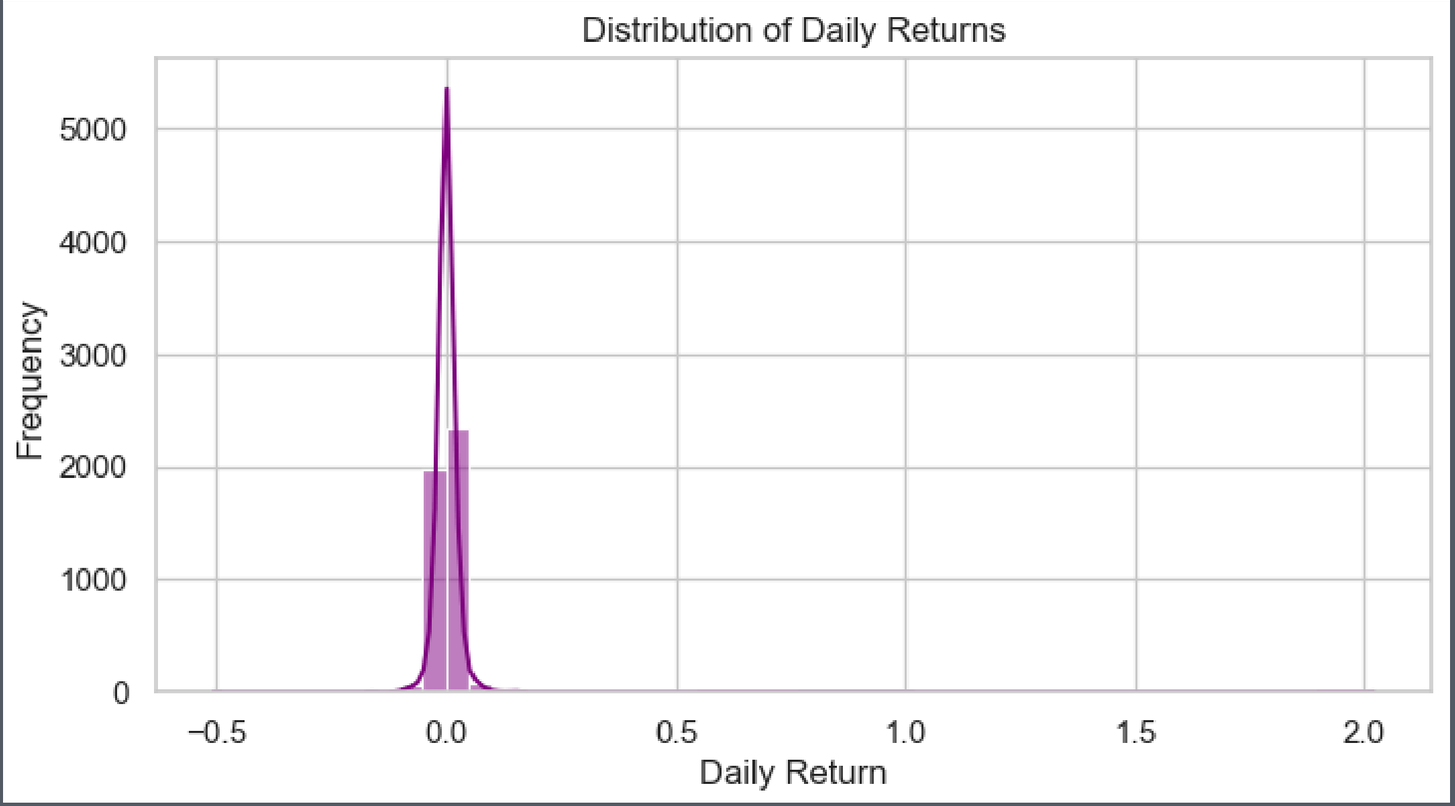
Short-term (MA20) and medium-term (MA50) moving averages track price momentum closely. MA20 reacts quicker to shifts, helping identify bullish or bearish crossovers and trend reversals.



Correlation Heatmap

Correlation Heatmap

A strong positive correlation exists between Open, High, Low, and Close prices. Interestingly, Volume behaves more independently, showing little correlation with price movements.



Distribution of Daily Returns

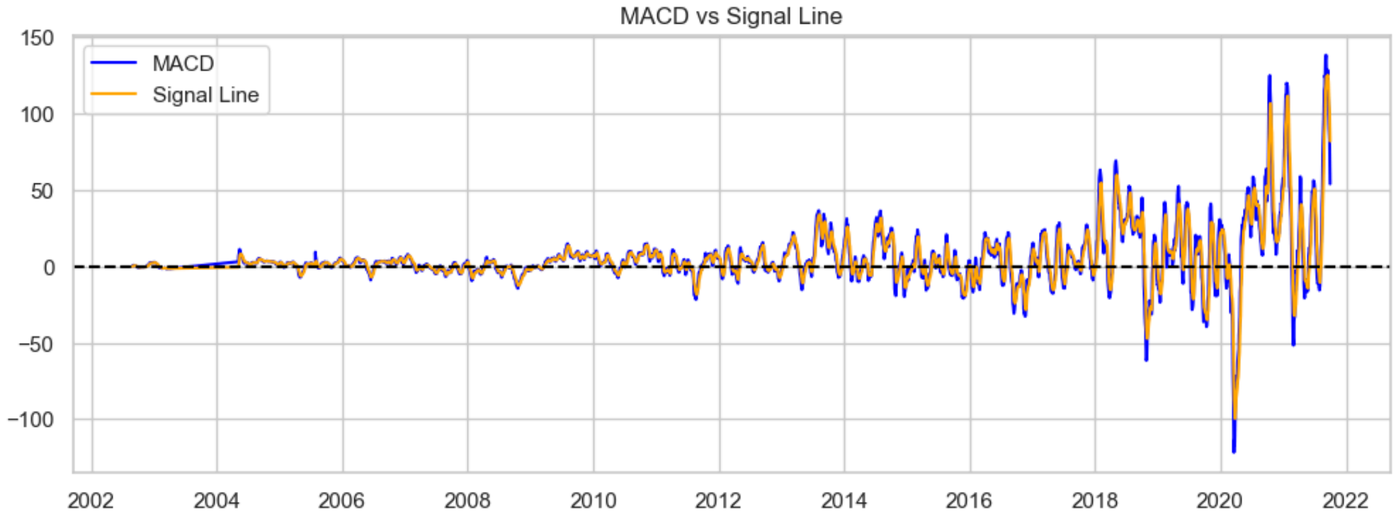
Distribution of Daily Returns

The return distribution forms a near-normal bell curve, centered around zero. This indicates mostly stable day-to-day movement, with some sharp gain/loss outliers driven by market news or earnings.

Cumulative Returns Over Time

Cumulative Returns Over Time

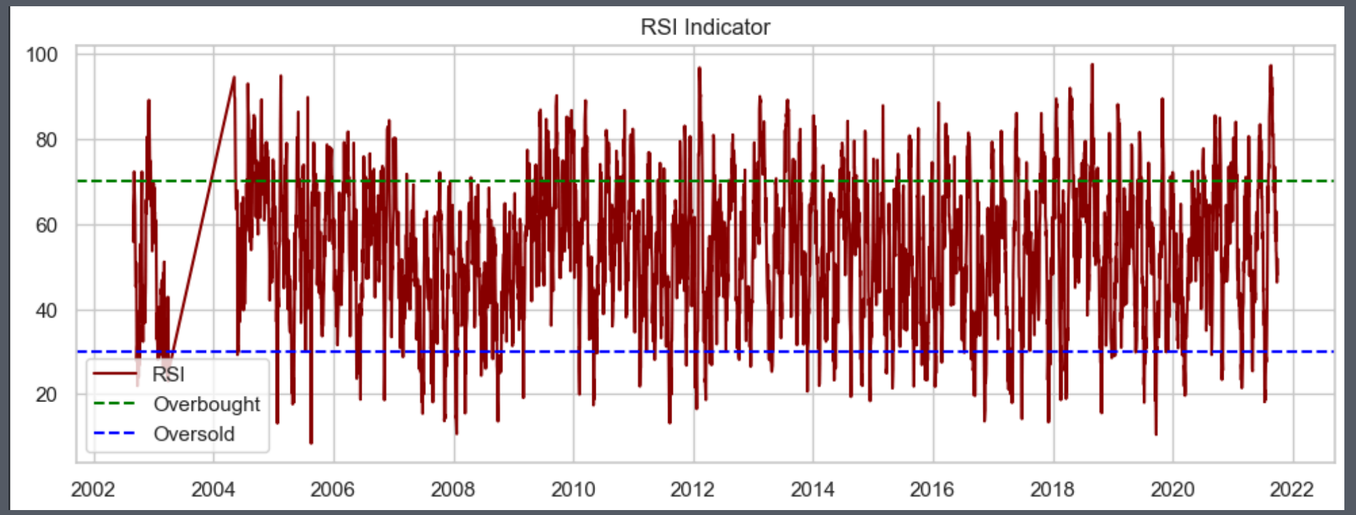
TCS shows a smooth and consistent growth trajectory over the years. The steady climb in cumulative returns makes it a reliable stock for long-term investors.



MACD vs Signal Line

MACD crossovers above the signal line represent bullish triggers, while drops below suggest bearish reversals. These technical signals often align with key price swings.

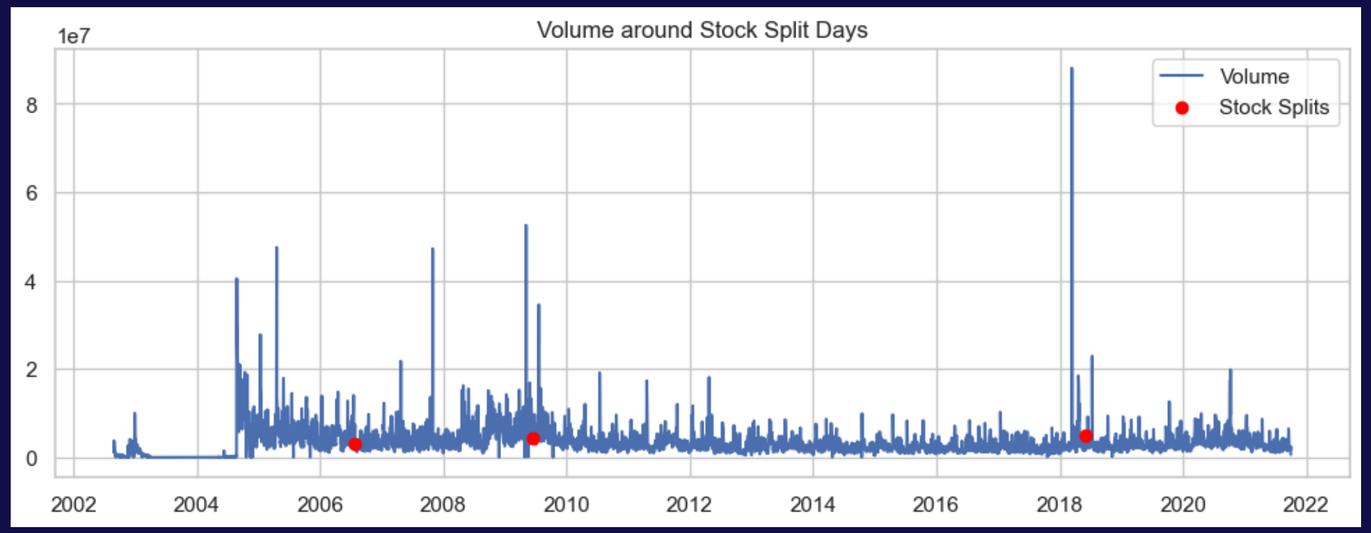
Forecasting future stock prices with Prophet model



RSI Indicator

RSI Indicator

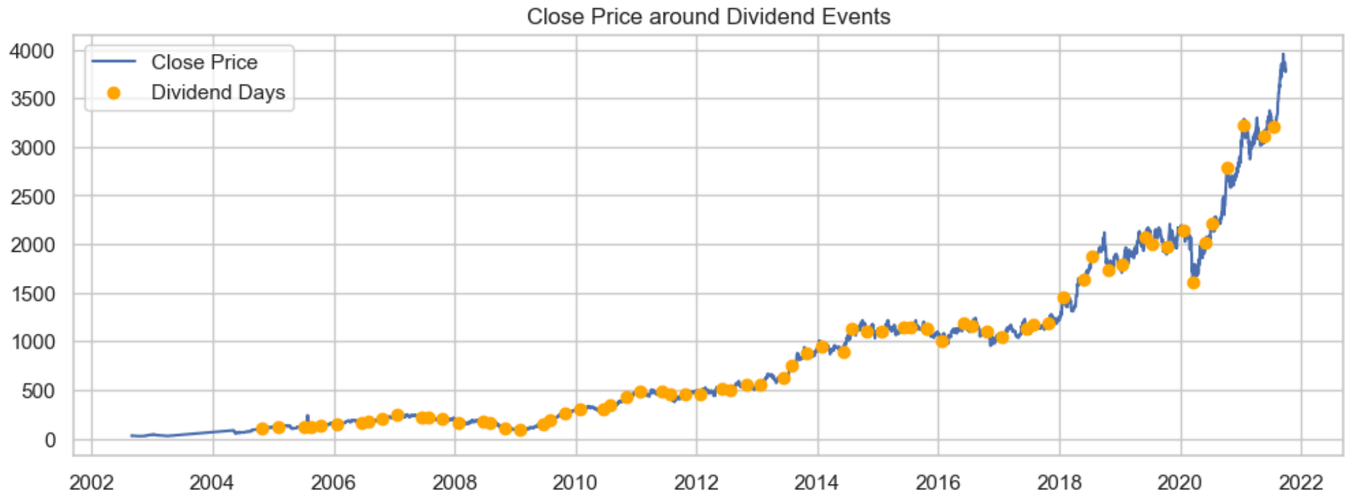
The RSI graph frequently enters overbought (>70) and oversold (<30) zones. These oscillations provide effective signals for short-term trade entries and exits based on market sentiment.



Volume Around Stock Split Days

Volume Around Stock Split Days

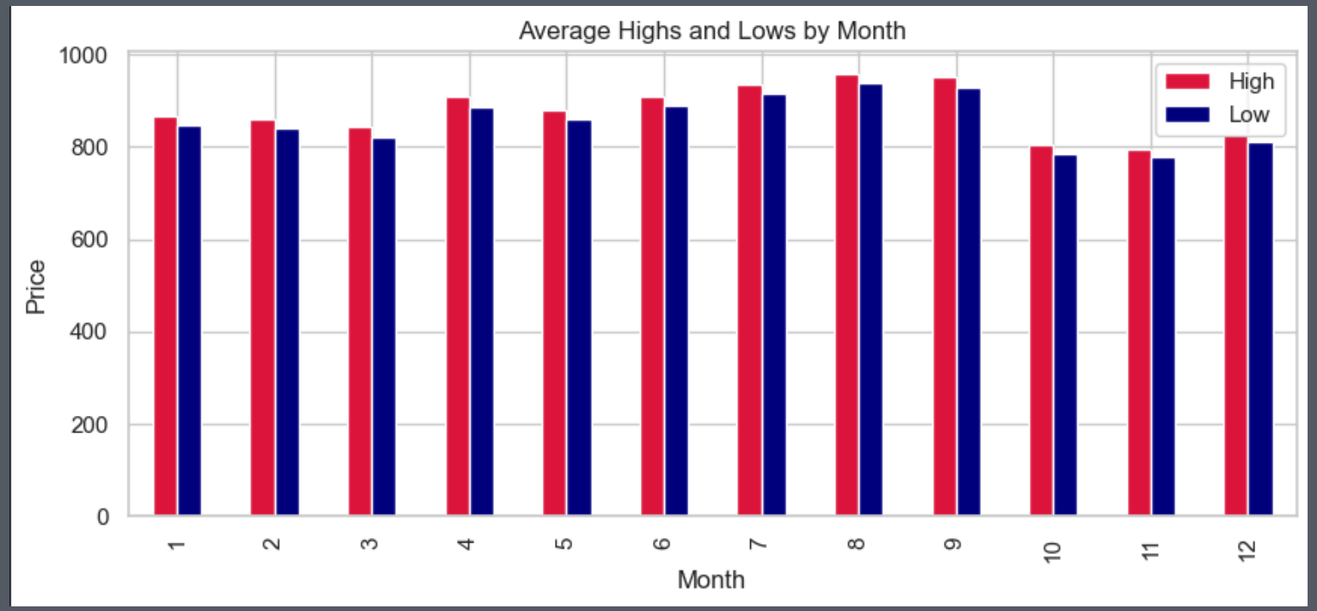
Volume spikes significantly around stock split dates, especially during 2018. This reflects increased speculative interest and liquidity driven by share restructuring events.



Close Price Around Dividend Events

Close Price Around Dividend Events

Dividend events lead to slight fluctuations in the close price. However, TCS maintains its upward momentum, indicating investor confidence and long-term value creation through consistent dividends.



Average Highs and Lows by Month

August and September stand out with the highest average highs and lows, suggesting bullish seasonal trends. October and November dips may relate to post-earnings corrections or market adjustments.

Cumulative returns and correlation analysis of TCS stock performance

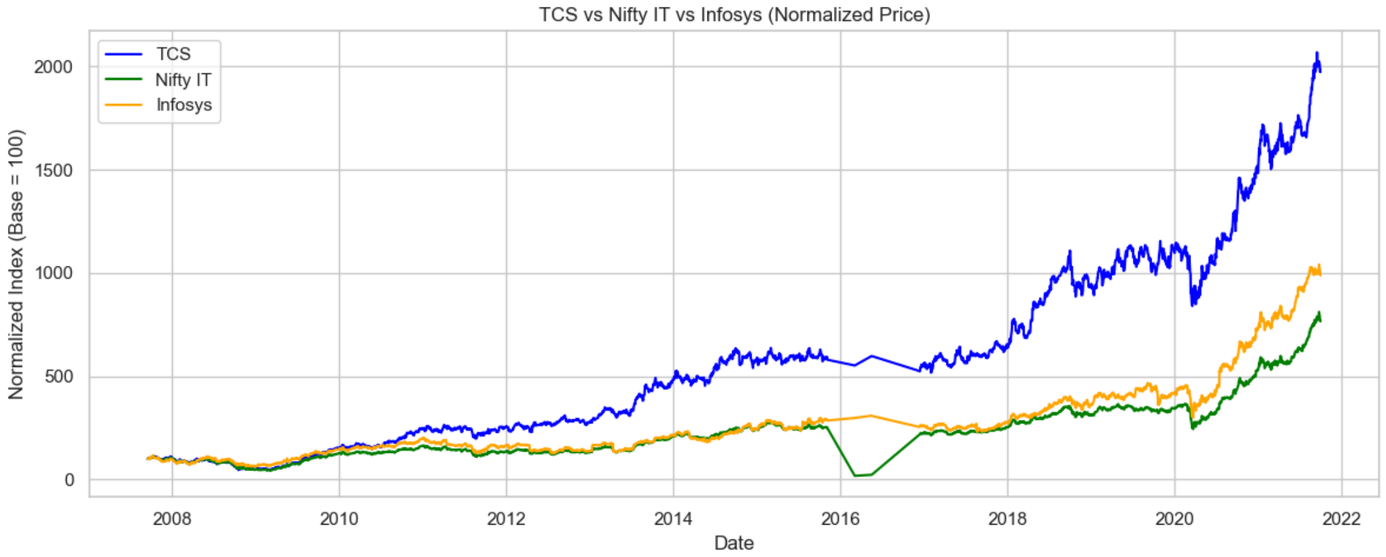
Average Close Price by Year

There’s a clear upward trend in average yearly close prices, with a notable spike from 2018 to 2021. This confirms TCS’s position as a steadily growing large-cap stock.

Model performance metrics for stock price forecasting

TCS vs Nifty IT vs Infosys

When normalized to a base of 100, TCS consistently outperforms both Infosys and the Nifty IT index over time. The smoother and stronger trajectory demonstrates its market leadership and investor trust.



TCS vs Nifty IT vs Infosys (Normalized Index)

Model Comparison

CLASSICAL MODELS (LR, RF, XGBOOST)

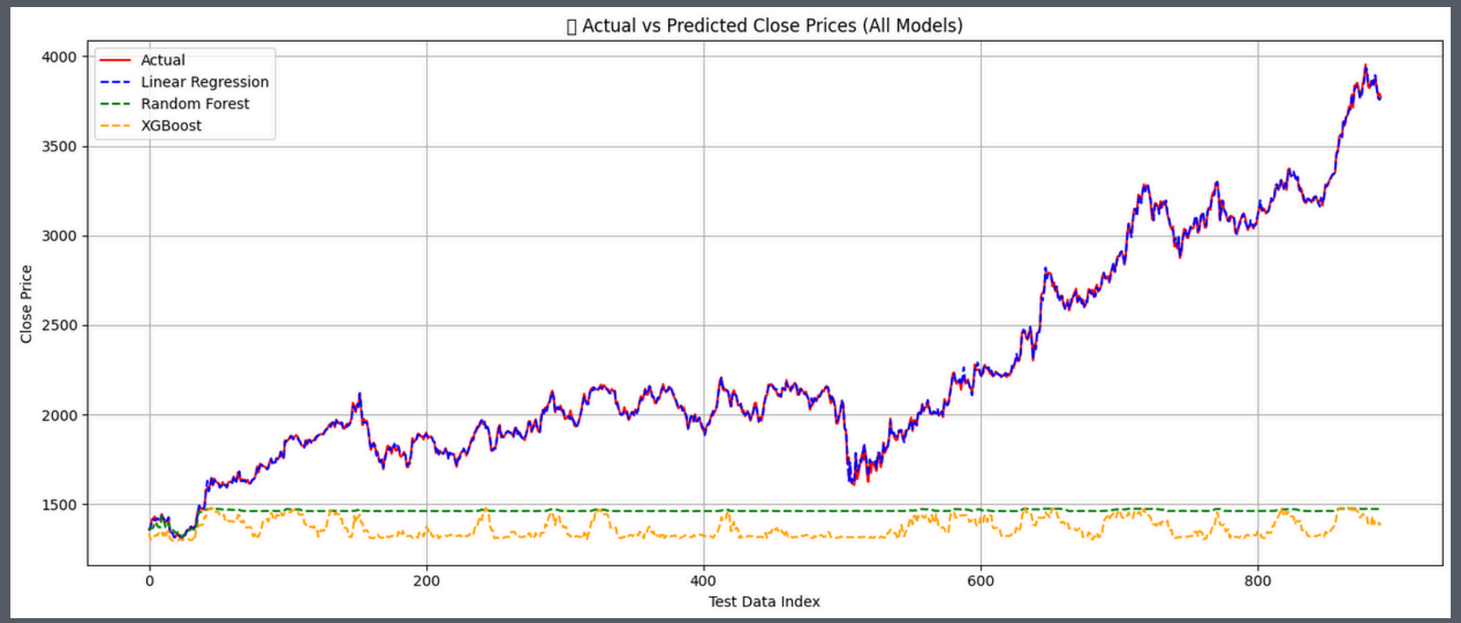
Linear Regression achieved high accuracy (RMSE: 14.71, R^2: ~0.999), while **Random Forest** and **XGBoost** struggled due to overfitting and lack of temporal understanding, with XGBoost showing the weakest performance (R^2: -2.22).

LSTM (DEEP LEARNING)

LSTM captured time-dependent trends well (RMSE: 246.29, R^2: 0.827), outperforming tree-based models in momentum prediction. It's effective for sequential forecasting but can lag during sharp spikes.

PROPHET

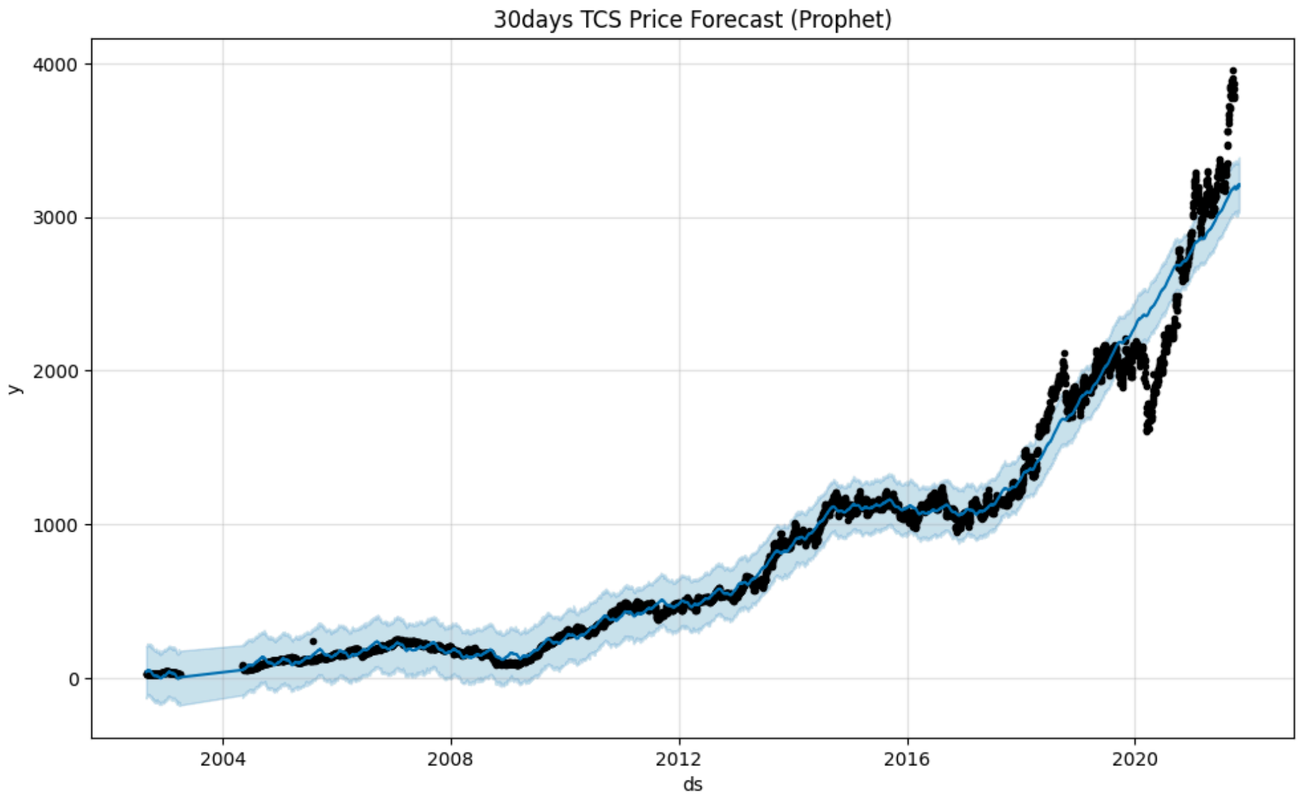
Prophet generated smooth, interpretable long-term forecasts using only the 'Close' price and time. Ideal for strategic planning with seasonal trends and uncertainty bands, especially in corporate forecasting contexts.



Actual vs Linear Regression, Random Forest and Xgboost

LR, RF and Xgboost

The classical models applied to TCS stock forecasting revealed diverse outcomes. Linear Regression performed exceptionally well with an RMSE of 14.71 and R^2 of ~0.999, benefiting from the smooth upward trend in TCS stock. In contrast, Random Forest captured non-linear patterns moderately well but showed signs of overfitting with an RMSE of 990.85 due to its lack of temporal awareness. XGBoost struggled the most, with an RMSE of 1067.58 and a negative R^2 (-2.22), likely due to poor hyperparameter tuning and insufficient handling of sequential patterns.



30days TCS Price Forecast (Prophet)

Prophet

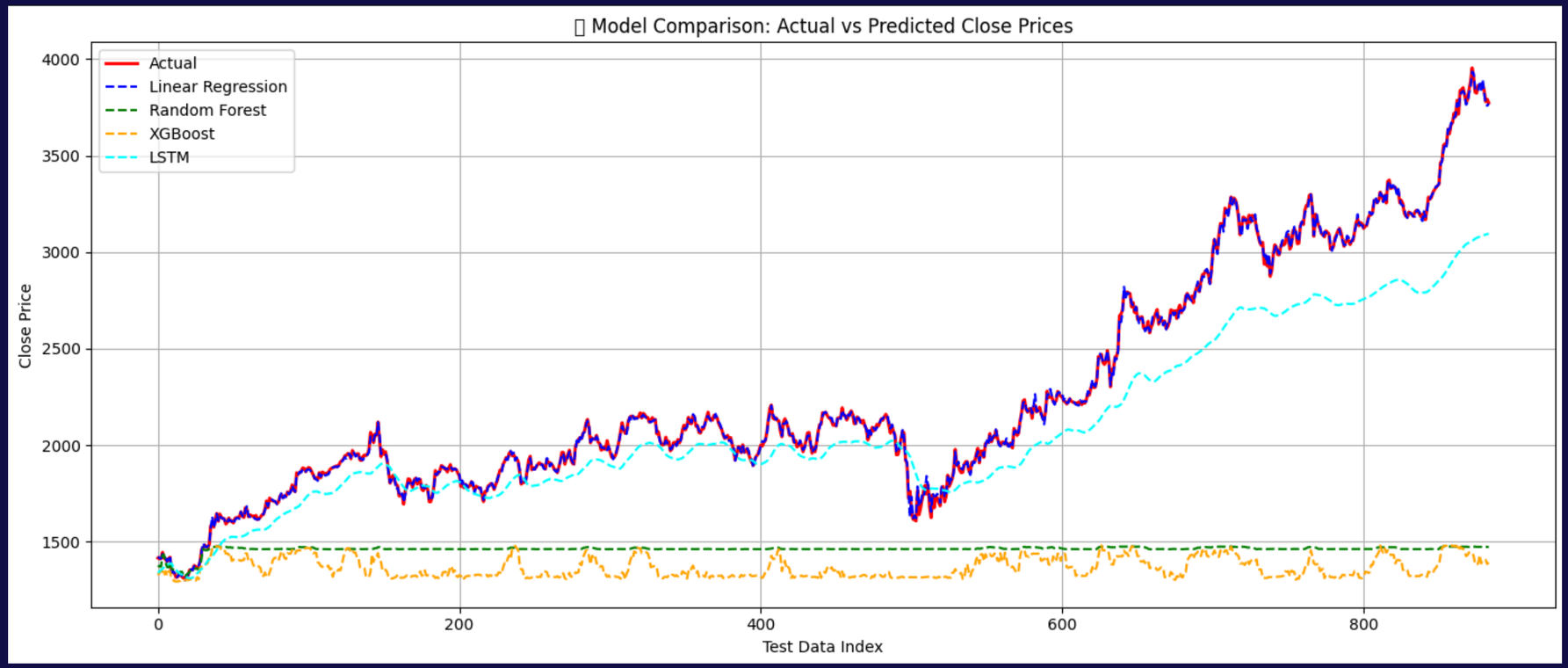
Prophet is a time-series forecasting model built by Meta, optimized for long-term trend and seasonality detection. It only used the ‘Close’ price and date as inputs. Prophet provided visually smooth and interpretable forecasts with 95% confidence intervals, making it suitable for investor-level outlooks and reporting. Though not measured in RMSE here, its clear structure and robust handling of missing data, holidays, and trends make it ideal for long-term corporate forecasting.

LSTM: Actual vs Predicted Close Price

LSTM

LSTM, a deep learning model designed for sequence prediction, was trained on rolling windows of historical data. It learned time-dependent patterns effectively, especially in short-term momentum shifts. It achieved an RMSE of 246.29 and an R^2 of 0.827, outperforming tree-based models in capturing sequential dependencies. While not perfect in handling sharp spikes, LSTM demonstrated strong potential for time series forecasting in financial domains.

All Model Comparison



This chart compares all models TCS stock prices with predictions from Linear Regression, Random Forest, XGBoost, and LSTM models. Linear Regression closely tracks the real trend, benefiting from the stock's smooth upward pattern. LSTM also performs well, capturing momentum with some lag. In contrast, Random Forest and XGBoost deviate significantly, showing poor temporal handling. Overall, Linear Regression and LSTM offer the most reliable predictions for TCS stock forecasting.

Key Takeaways

CONSISTENT LONG-TERM GROWTH

EDA revealed TCS’s strong upward trend in cumulative returns, supported by rising average yearly close prices—ideal for long-term investment models.

SEASONAL & VOLATILE PATTERNS

Volume and price shifts by month and weekday suggest seasonal investor behavior, valuable for time-aware models like LSTM and Prophet.

TECHNICAL INDICATOR STRENGTH

Features like MACD, RSI, and rolling averages captured momentum shifts well, enhancing model learning.

LINEAR REGRESSION SHINES

Surprisingly, simple Linear Regression outperformed complex models in RMSE due to the linear nature of TCS's long-term trend.

LSTM CAPTURES SEQUENCE

LSTM delivered the best balance between accuracy and trend learning by modeling sequential dependencies.

XGBOOST & RF LIMITATIONS

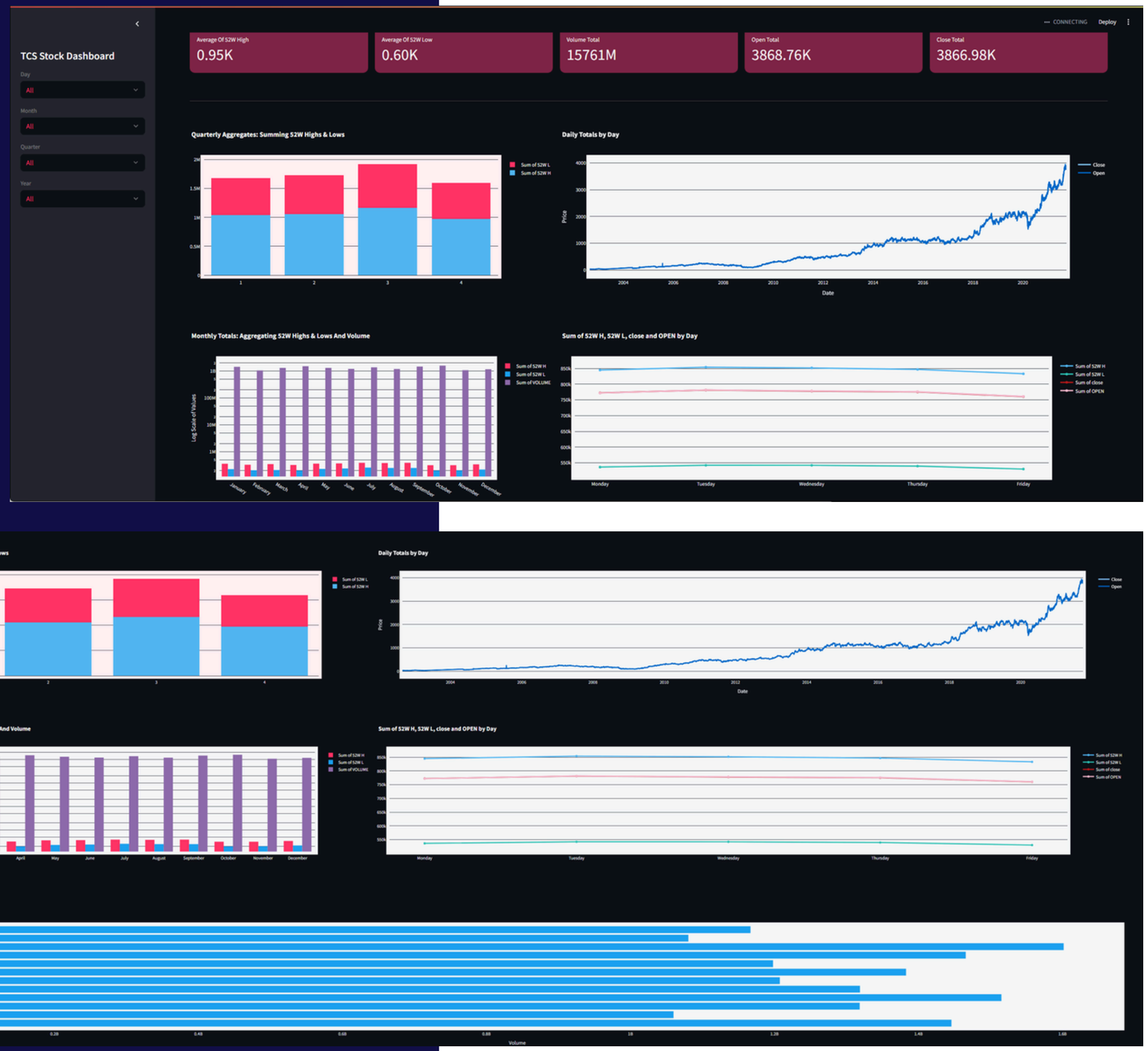
Tree-based models struggled due to lack of temporal awareness and required better tuning or additional lag/rolling features.

PROPHET FOR FORECASTING

Ideal for high-level, explainable long-term trend forecasting with built-in handling of seasonality and holidays.

TCS Stock Dashboard App

The **dashboard.py** script is a powerful Streamlit-based web application developed to visually explore and analyze historical stock performance of Tata Consultancy Services (TCS). This interactive dashboard serves as a live front-end for data exploration, combining real-time filters, key performance metrics, and rich visualizations.



Thank You