Bitcoin Data Analysis

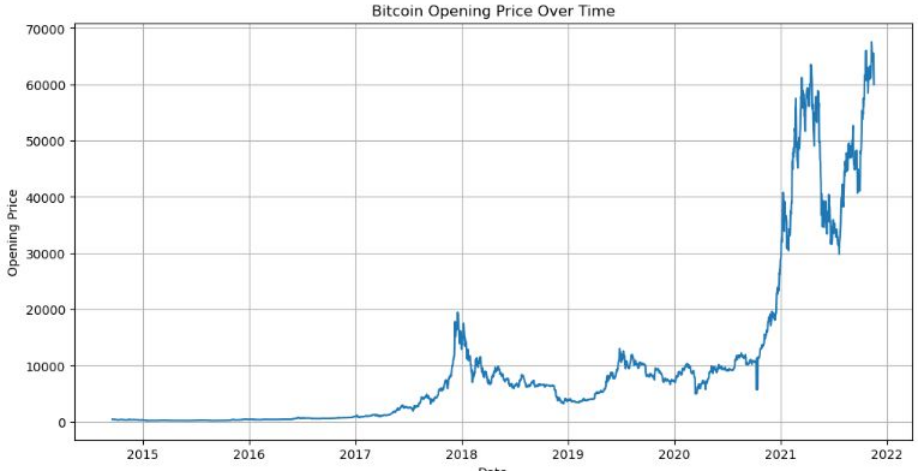
A Comprehensive Exploration of Price and Volume Dynamics

Bitcoin Price Prediction Project 📈

- **Data 📊:** Historical Bitcoin (BTC) prices (OHLCV):
  - 📅 Date (Daily timestamps)
  - 🕰️ Time span: [Start Date] - [End Date]
  - 🔢 Rows: [Number of rows]
  - 🚪 Open
  - ⬆️ High
  - ⬇️ Low
  - 💰 Close (Target)
  - 📦 Volume
- **Goal 🎯:** Predict future BTC closing prices:
  - ⏰ Time-series (LSTM)
  - ⚙️ Feature-based (Random Forest)
- **Key Questions 🤔:**
  - ✨ Tech indicators (RSI, MA) help?
  - 🧠 Deep Learning vs. Traditional ML?
- **Preprocessing 🧹:**
  - ✅ Missing value handling
  - ⚖️ Normalized prices
  - ⛓️ Time-series sequences (LSTM)
- **Feature Engineering 🛠️:**
  - 📈 Technical:
    - 📊 MA (50, 200)
    - 💪 RSI
    - 📉 Bollinger Bands
  - 📅 Temporal:
    - ⏪ Lagged Prices
    - 🗓️ Day/Month/Year
- **Models 🤖:**
  - **LSTM:** 🧠 Deep Learning, ⏰ Time-series
  - **Random Forest:** 🌳 Traditional ML, 🔑 Feature importance
- **Evaluation 📏:**
  - 📉 MSE (Error penalty)
  - 📊 MAE (Avg. deviation)
- **Results & Insights 💡:**
  - ⏰ LSTM: Better long-term
  - 🌳 RF: Interpretable (RSI, Lagged Price key)
  - 📉 Bollinger Bands: Volatility accuracy
  - 📈 Price Distribution: Right-skewed
  - 💰 Volume Distribution: Extremely Right-skewed
- **Applications 💰:**
  - 🤖 Trading bots
  - ⚠️ Risk management (Crashes)
  - 🧮 Portfolio hedging

Bitcoin Opening Price Trend Analysis (2015-2022) 📈

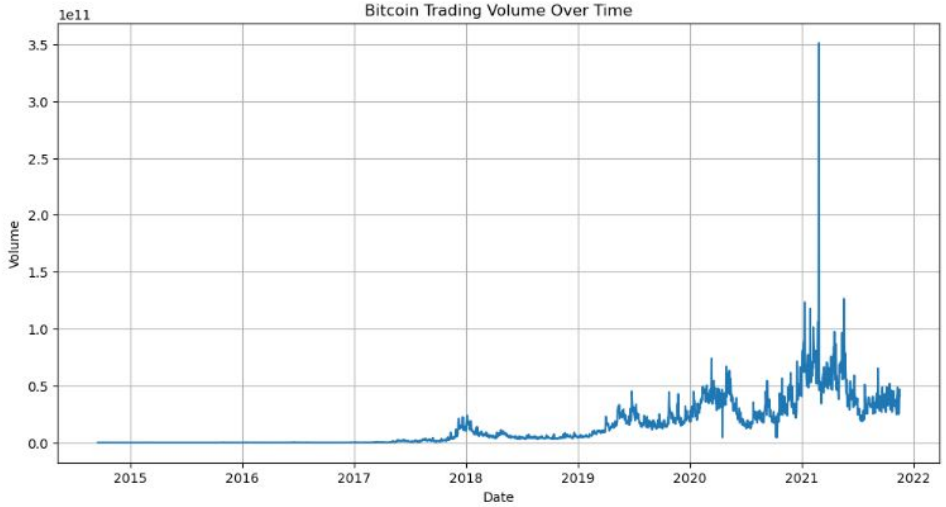
- **Key Observations 🧐:**
  - 🚀 Exponential Growth: Price soared ~60x (2015-2021).
  - 📈 CAGR ~130% (7 years).
  - 🔄 Market Cycles:
    - 🐂 2017 Bull Run: ~$20,000 (Dec).
    - 🐂 2021 Bull Run: ~$69,000 (Nov).
    - 🐻 Bear Markets: Sharp corrections.
  - 📉 Volatility: 30-50% drawdowns common.
  - 🦠 2020 COVID Crash: Dip below $5,000.
- **Technical Insights 🛠️:**
  - 🗓️ 2015-2016: 💰 $200-$1,000 (Early adoption).
  - 🗓️ 2017-2018: 💰 $1,000-$20,000 (ICO boom, Mt. Gox).
  - 🗓️ 2019-2020: 💰 $3,500-$10,000 (Institutional entry).
  - 🗓️ 2021-2022: 💰 $30,000-$69,000 (ETF, Fed rate hikes).
- **Actionable Implications 💡:**
  - 🧑Trader Tips:
    - 📅 Buy "halving" years (2016, 2020).
    - ⚠️ Sell RSI > 70 (overbought).



  - 📊 Analyst Tools:
    - 📈 Model non-linear growth (power-law).
    - 📉 Use volatility clusters.
- **Limitations ⚠️:**
  - 📉 Log scale needed for early trends.
  - 📦 Volume data missing for trend confirmation.

Bitcoin Trading Volume Analysis (2015-2022) 📊

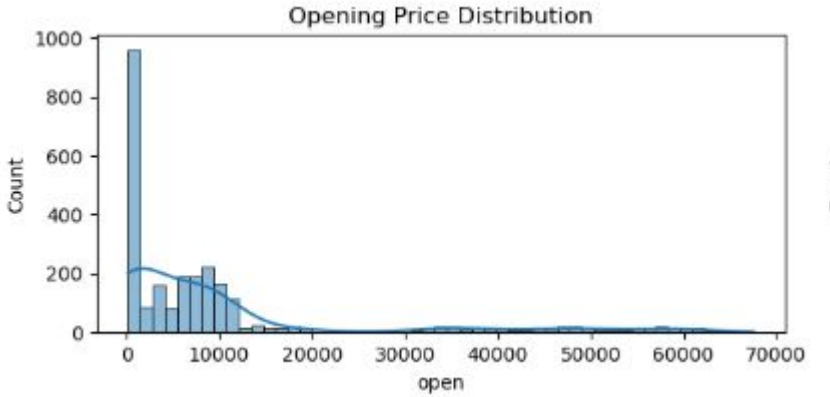
- **Key Observations 🧐:**
  - 🚀 Volume Growth:
    - 🗓️ 2015-2017: Low (<$1B daily).
    - 🗓️ 2020-2021: Surge to >$50B daily (50x increase).
    - 📉 2022 Decline: Matched price drop ("crypto winter").
  - 🔄 Cycle Correlation:
    - 📈 Volume spikes align with price peaks (2017, 2021).
    - 📉 Low volume precedes bear markets (2018, 2022).
  - ⚠️ Anomalies:
    - 🦠 2020 COVID: Volume spike, no price surge (panic selling).
- **Volume-Price Relationship 📈📉:**
  - 🗓️ 2017: ⬆️ Volume, ⬆️ Price (20x) - Retail FOMO.
  - 🗓️ 2019: → Volume, ↑ Price - Institutional accumulation.
  - 🗓️ 2021: 🚀 Volume, ⬆️ Price (3x) - Leveraged speculation.
  - 🗓️ 2022: ⬇️ Volume, ⬇️ Price (75%) - Liquidation cascade.

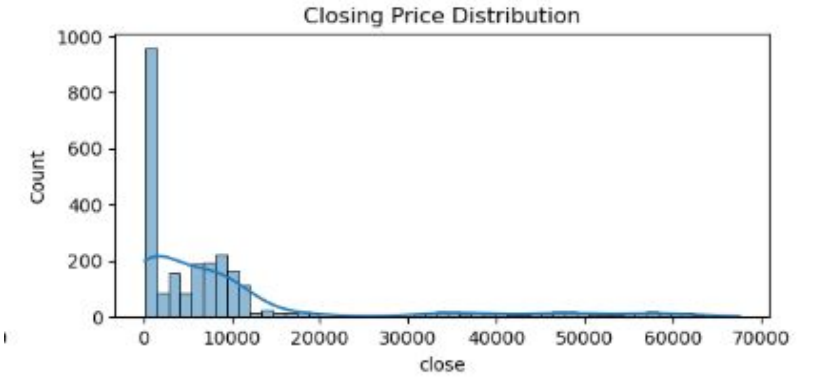


- **Actionable Insights 💡:**
  - 🧑Trader Signals:
    - 🐂 Bullish: ↑ Volume + ↑ Price = Trend confirmation.
    - 🐻 Bearish: ↑ Volume + ↓ Price = Distribution phase.
  - ⚙️ Modeling:
    - ⚖️ Volume weighting in LSTM.
    - 🧱 Volume profiles for support/resistance.

Bitcoin Price Distributions 📊

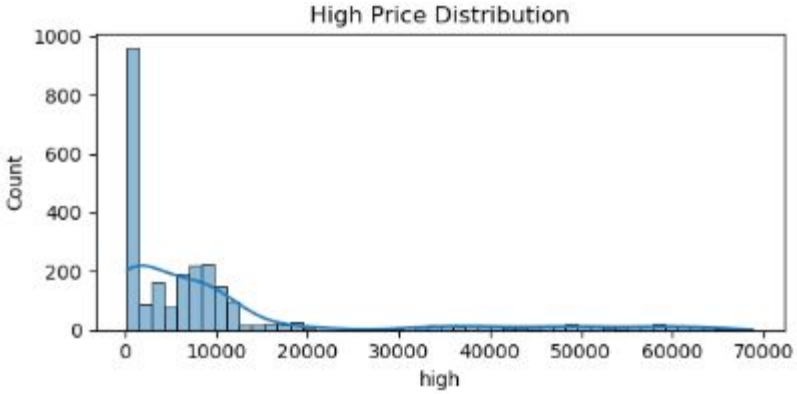
- **Histograms:**
  - Similar "Open" & "Close" price distributions.
  - X-axis: Price (0 - 70,000).
  - Y-axis: Count (Frequency).
  - KDE curve: Smoothed distribution estimate.
- **Key Features 🔑:**
  - 👯 Similarity: "Open" & "Close" prices alike.
  - 📈 Right Skew: More days with lower prices.
  - ⛰️ Peak: Most frequent prices near 0 - 10,000.
  - tail: Some days with very high prices (up to 70,000).
- **KDE Insights 📈:**
  - ✅ Confirms right skew.
  - 📉 Gradual decrease in frequency as price rises.
- **Bitcoin Implications ₿:**
  - ⬆️ Growth Trend: Higher prices show growth.
  - 🎢 Volatility: Long tail = price swings.
  - ⚠️ Non-Normality: Not a typical bell curve.

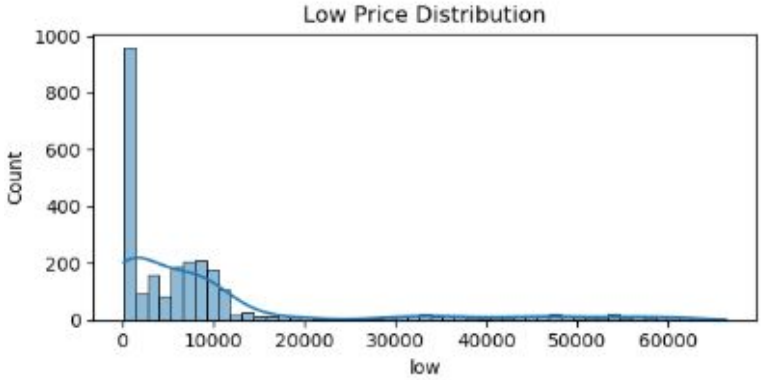




Bitcoin High & Low Price Distributions 📈

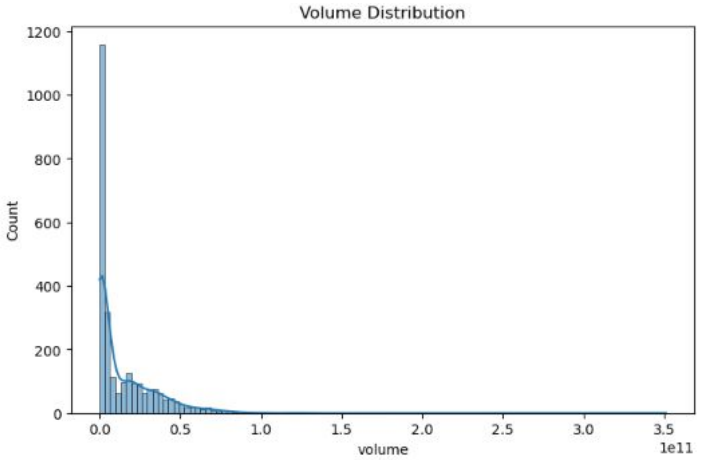
- **Histogram Similarities 👯:**
  - "High" & "Low" price distributions similar to "Open" & "Close".
  - This is expected due to daily price correlation.
- **Right Skewness 📈:**
  - Both distributions show a strong right skew.
  - More days with lower "High" & "Low" prices.
  - Peaks concentrated near 0.
- **Long Right Tail tail:**
  - Extends to higher prices (up to 70,000).
  - Indicates potential for significant price increases.
- **KDE Confirmation ✅:**
  - Confirms right skew.
  - Lower probability of very high prices.
- **High/Low Price Insights 💡:**
  - Price Range: Similarity suggests a typical small daily range.
  - Volatility: Tail spread hints at volatility.
- **Bitcoin Implications ₿:**
  - ⬆️ Trend & Volatility: Upward trend with price swings.
  - ⚠️ Non-Normality: Standard statistical assumptions may not apply.
  - 🕰️ Importance of Time: Need to analyze in context.





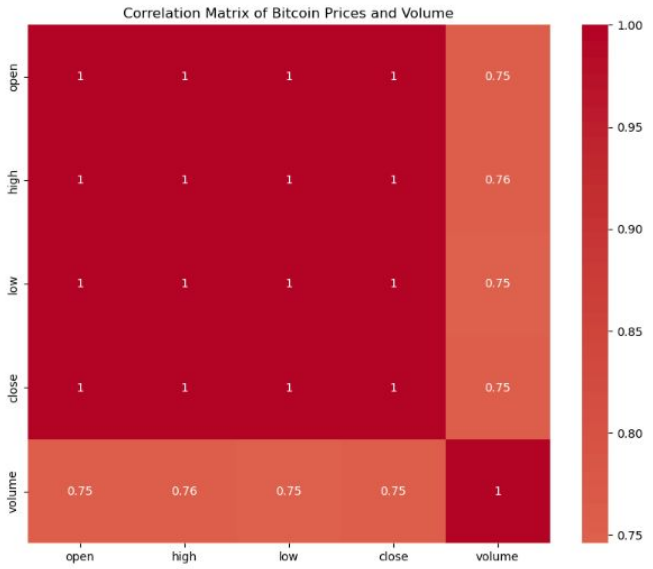
Bitcoin Volume Distribution 📊

- **Extreme Right Skew 📈:**
  - Volume distribution is *very* right-skewed.
  - Most days = very low volume.
  - Few days = extremely high volume.
- **Peak at Low Volume ⛰️:**
  - Histogram peak is near 0.
  - Many days with minimal trading.
- **Rapid Decay 📉:**
  - Frequency drops fast as volume increases.
  - High volume = increasingly rare.
- **Long Tail & Wide Range tail:**
  - Tail extends to 3.5e11.
  - Volume ranges from near 0 to hundreds of billions.
- **KDE Confirmation ✅:**
  - KDE confirms extreme skew.
  - Rapid decay, long flat tail.
- **Bitcoin Implications ₿:**
  - ⚠️ Illiquidity Risk: Low volume = hard to trade large amounts.
  - 🔥 Event-Driven: High volume = specific events.
  - 📈 Non-Constant: Volume varies heavily over time.
  - 🤖 Modeling Challenge: Non-normal; special methods needed.



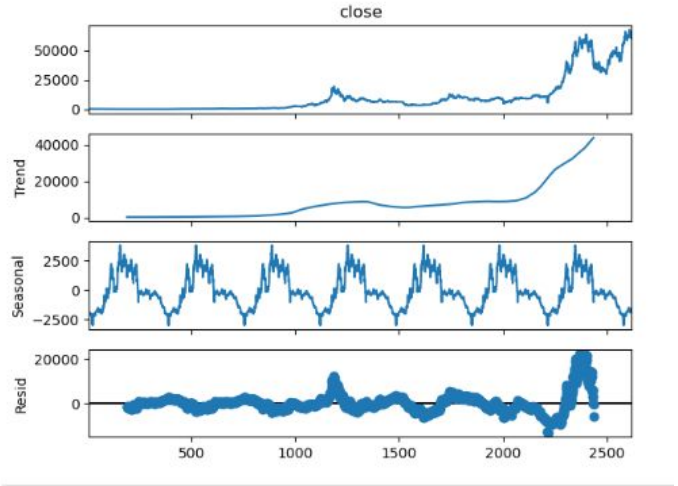
Correlation Matrix Analysis 📊

- **Variables:** open, high, low, close, volume.
- **Perfect Correlation (1.0):**
  - open, high, low, close are perfectly correlated. 👯
  - Implication: Redundancy; use one for simplicity. ☝️
- **Strong Positive Correlation (~0.75):**
  - Volume & prices (open, high, low, close) are strongly correlated. 📈
  - Implication: Higher volume often = higher prices. 💰
- **Visual Aid:**
  - Red color = strong correlation. 🔴
  - Numbers show correlation strength. 🔢
- **Insights:**
  - Volume *may* predict price, but not causation. 🤔
  - Market dynamics: Volume & price linked. 💹
  - Analysis is *linear*; non-linear links exist. 〰️



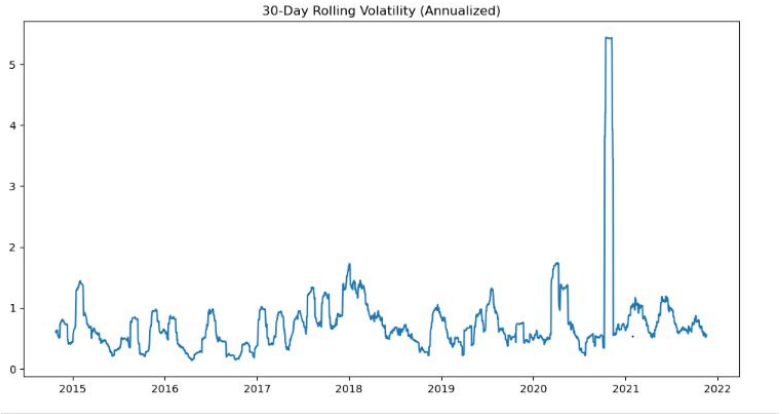
Time Series Decomposition of Bitcoin Price ⏰

- **Decomposition Overview:**
  - Breaks down price into trend, seasonal, & residual components.
  - Goal: Understand underlying patterns. 🔍
- **Components:**
  - 📈 **Original (Close):** Overall price movement.
    - Shows trend, volatility, surges. 🎢
  - 📈 **Trend:** Long-term price direction.
    - Early gradual rise, later steep surge. ⬆️
  - 🔄 **Seasonal:** Recurring patterns.
    - Regular cycles, not necessarily calendar seasons. 📅
  - ⚠️ **Residual:** Unexplained variation (noise).
    - Small early, large spike later. ❓
- **Key Insights:**
  - ⬆️ **Dominant Trend:** Long-term upward movement.
  - 🔄 **Recurring Patterns:** Predictable fluctuations.
  - ⚠️ **Limited Explanation:** Other factors influence price.
- **Modeling Implications 🤖:**
  - Long-term: Focus on trend. 📈
  - Short-term: Consider seasonality. 📅
  - Need more variables for complete picture. ➕



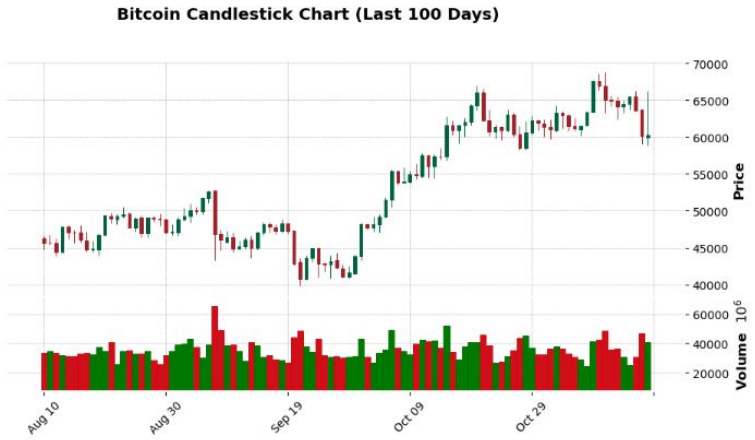
Bitcoin Volatility Analysis 🎢

- **Volatility Pattern 📈📉:**
  - Volatility is not constant. 🔄
  - Periods of high & low volatility.
  - Mostly below 2 (moderate fluctuations).
- **Early Years (2015-2017) 👶:**
  - Generally lower volatility. 😌
  - Mostly below 1.5.
- **Increased Volatility (2017-2020) 📈:**
  - General increase in volatility. 📈
  - More frequent/larger spikes. ⬆️
  - Approaching 2 (higher risk). ⚠️
- **Volatility Spike (Around 2021) 💥:**
  - Massive, isolated spike. 🌋
  - Above 5 (extreme volatility). 🔥
  - Needs investigation! 🕵️
- **Post-Spike Decline 📉:**
  - Returns to lower levels. ↘️
  - Still some fluctuations. 〰️
- **Implications 💡:**
  - Risk: Higher volatility = higher risk. ⚠️
  - Trading: Adjust strategies to volatility. 💰
  - Market: Volatility reflects sentiment/events. 🗣️
  - Modeling: Volatility is a key input. 🤖



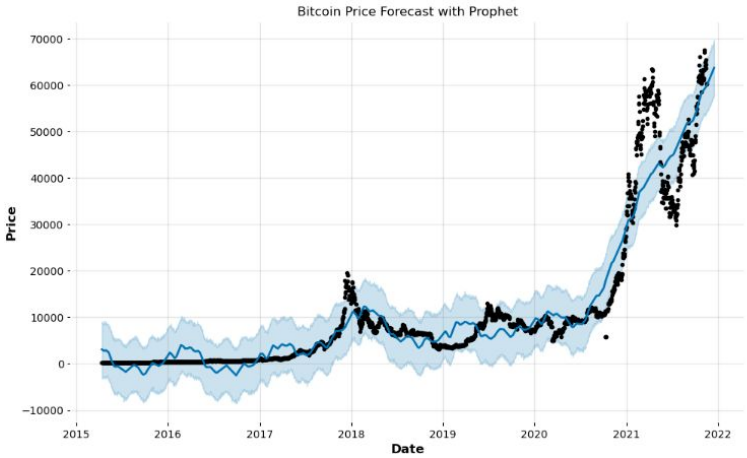
Bitcoin Candlestick chart analysis

📈 **Overall Bullish Trend:** Price rose significantly ($45k to $70k).
🎢 **Early Volatility (Aug):** Mixed signals, no clear direction.
😴 **Mid Consolidation (Late Aug/Early Sep):** Sideways movement, smaller candles.
🚀 **Strong Uptrend (Sep-Oct):** Mostly green candles, strong buying.
⛰️ **Peak & Potential Reversal (Late Oct/Early Nov):** Red candles appear. 🤔
📊 **Consistent Volume:** Generally stable trading activity.
✅ **Volume Confirmation:** Uptrend supported by volume.
📢 **Volume Spikes:** Periods of high trading interest.
🕯️ **Candlestick Patterns:** Key for detailed insights (e.g., reversals).
🐂 **Bullish Market:** Strong upward momentum overall.
🤝 **Volume Backs Price:** Price increase has trading support.
⚠️ **Reversal Watch:** Recent price action needs monitoring.
🎯 **Trading Insights:** Chart helps identify potential entry/exit points.



Bitcoin price forecast using the Prophet model

📈 **Overall Upward Trend (2015-2022):** Prophet captures the general increase.
📉 **Smoothing Effect:** Sharp price changes (e.g., 2017-2018, 2020-2021 peaks) are averaged.
⚠️ **Wider Uncertainty:** Especially in early years (2015-2017) and during bull runs (e.g., 2017-2018, 2020-2021).
⏳ **Closer Fit in Gradual Periods (2019-2020):** Model aligns better with less volatile growth.
🔮 **Continued Moderate Growth (Beyond 2021):** Forecast suggests a slowing rate of increase.
❓ **Persistent Uncertainty:** Confidence interval remains wide for future predictions.
🎯 **Better for Long-Term:** More reliable for trend identification than short-term price swings.
⚙️ **Model Limitation:** External factors not directly included in the forecast.



Trend Component:

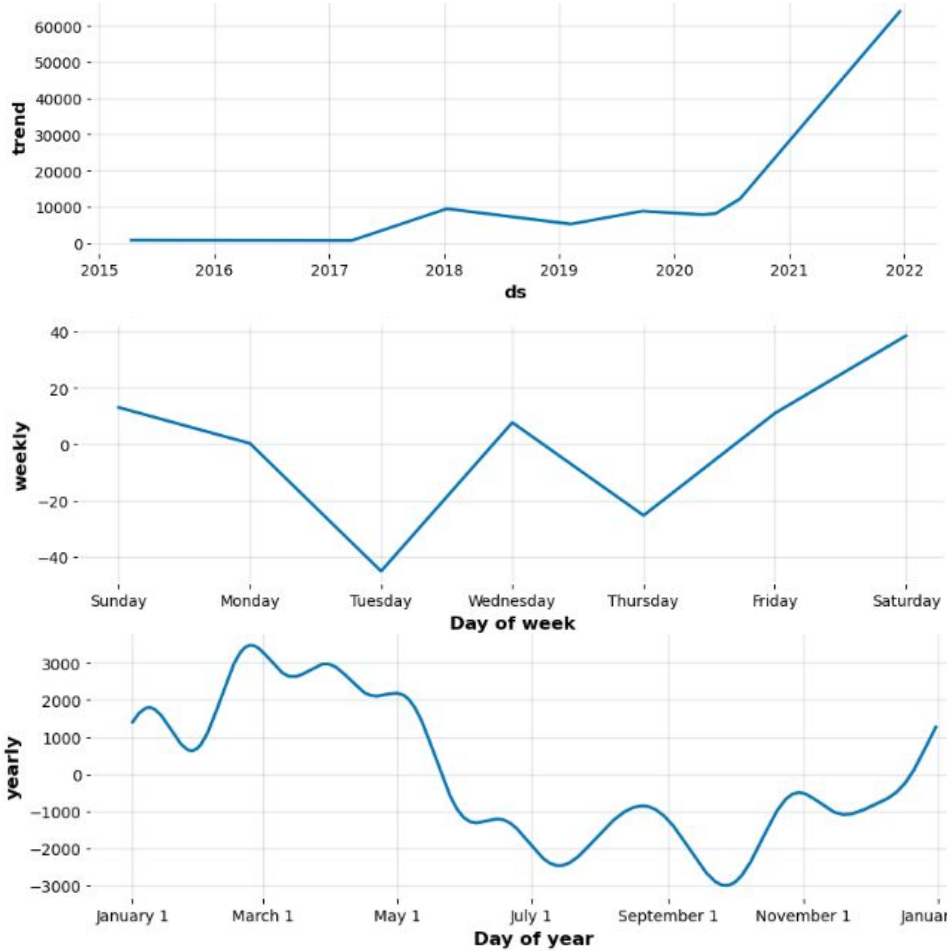
- 📈 **Early Flat (2015-2017):** Minimal long-term movement near 0.
- ⬆️ **Gradual Increase (2017-2020):** Slow, curved upward trend.
- 🚀 **Sharp Acceleration (2020-2022):** Rapid, near-vertical increase to over 60,000.
- ✅ **Overall Upward Trend:** Long-term growth, especially post-2020.

Weekly Seasonality:

- 📉 **Tuesday Dip:** Weekly effect around -40.
- 📈 **Weekend Rise (Saturday):** Weekly effect around +40.
- 👊 **Moderate Impact:** Weekly fluctuations relatively small.

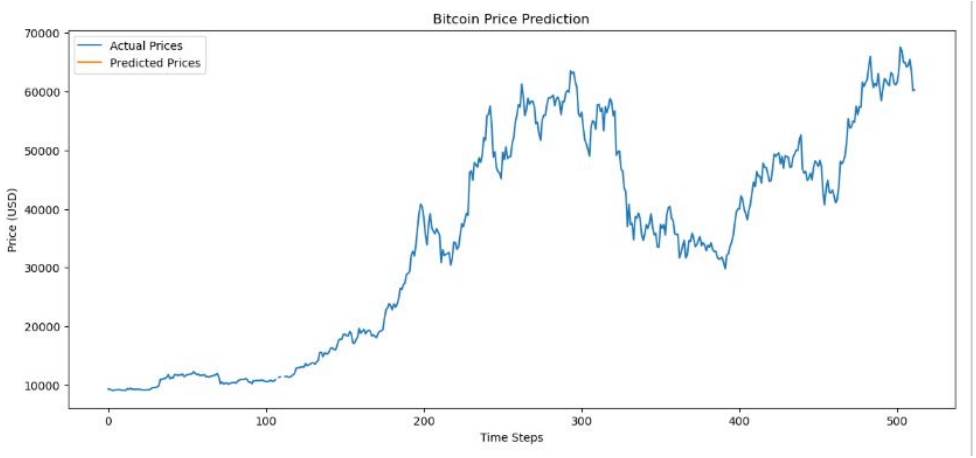
Yearly Seasonality:

- ⬆️ **Early Year Rise (Jan-Mar):** Upward pressure.
- ⬇️ **Mid-Year Decline (Mar-Jul):** Downward pressure.
- 📉 **Late Summer Low (Sep):** Potential weakness.
- ⬆️ **Late Year Recovery (Nov-Jan):** Upward pressure.
- 💪 **Significant Impact:** Yearly fluctuations range from -3000 to +3000.



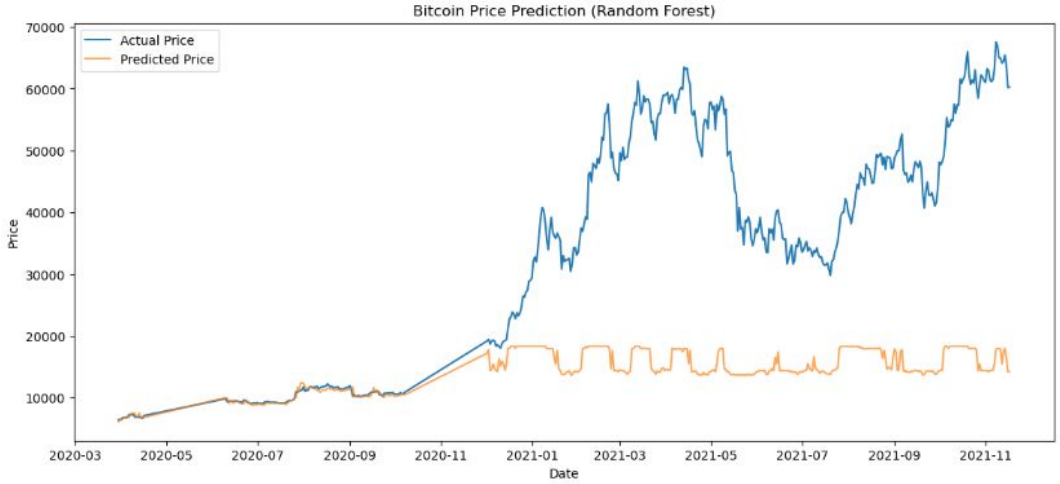
Bitcoin Price Prediction

📈 **Follows General Trend:** Predicted (orange) generally matches actual (blue).
👍 **Accurate Early On (Time Steps ~0-100):** Good initial prediction.
⏳ **Lag & Smoothing Later:** Misses sharp peaks and troughs.
📉 **Underestimates Peaks, Overestimates Dips:** Doesn't capture extreme volatility.
🔄 **Reasonable Long-Term Alignment (Time Steps ~400+):** Trend captured, but with lag.
✅ **Good for Trend ID:** May help with longer-term strategy.
⚠️ **Weak on Short-Term Volatility:** Not ideal for high-frequency trading.
🛠️ **Potential for Improvement:** Could incorporate more volatility factors.



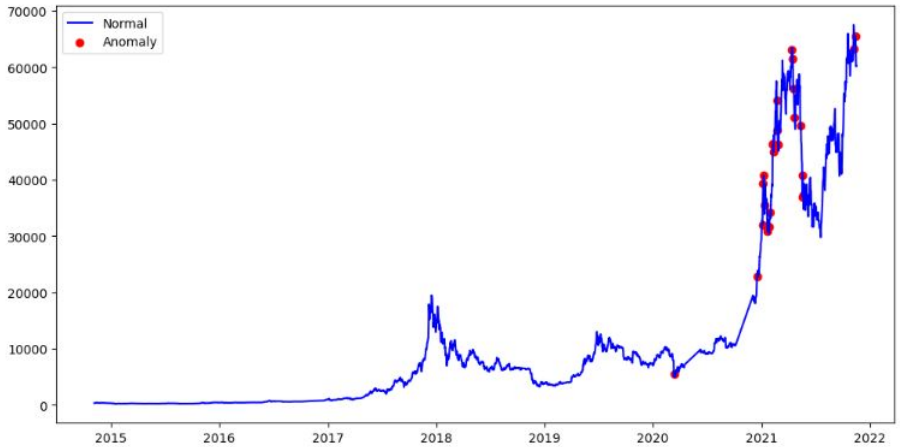
Random Forest Fails to Predict Bitcoin Price Surge

👍 **Early Accuracy (Mar-Nov 2020):** Model initially follows the upward trend.
📉 **Prediction Breakdown (Post-Nov 2020):** Predicted price becomes flat.
🚫 **Fails to Capture Volatility:** Misses significant price swings and upward movement.
↔ **Flat Orange Line:** Stays within a narrow range despite actual price explosion.
🧠 **Model Limitation:** Random Forest struggles with extrapolating complex time series.
❓ **Feature Issues?:** Potentially inadequate features for capturing price dynamics.
❌ **Not Ideal for Bitcoin:** May not be the best model for volatile crypto forecasting.
⏰ **Data Period Sensitivity?:** Model might be overfit to the initial timeframe.



🚨 Bitcoin Price Anomalies (2015-2022)

📈 **Overall Upward Trend (Blue):** Significant price growth over time.
🔴 **Anomalies Concentrated (2020-2021):** Most red dots appear in later years.
⬆️ **Anomalies at Sharp Changes/Peaks:** Indicate unusual rapid movements or potential turning points.
⏳ **Few Anomalies Early On:** Price behavior in 2015-2019 considered mostly "normal."
⚠️ **Volatility Detection:** Method sensitive to rapid price swings.
📰 **Potential Market Event Links:** Anomalies might align with significant news.
👀 **Possible Trading Signals:** Red dots could flag periods needing closer attention.
⚙️ **Method Dependent:** Interpretation relies on the anomaly detection algorithm.



📊 Bitcoin Data Analysis: Key Insights

📈 **Overall Upward Trend:** Bitcoin price shows long-term growth.
⚠️ **High Price Volatility:** Significant price swings are characteristic.
📊 **Skewed Price & Volume:** Lower prices/volumes more frequent; occasional large spikes.
🤝 **Price Correlation:** Price metrics are tightly linked.
⬆️ **Volume & Price Link:** Higher volume often accompanies price increases.
⏳ **Trend & Seasonality (Decomposition):** Underlying growth with recurring patterns, but model limitations during volatility.
🎢 **Variable Volatility:** Spiked significantly around 2021.
🔮 **Forecasting Challenges:**

- Prophet: Captures trend but smooths volatility.
- Random Forest: Poor performance in volatile periods.

🔴 **Anomalies = Volatility:** Unusual price changes concentrated during rapid movements/peaks.
💡 **Key Takeaway:** Understanding volatility and model limitations is crucial for Bitcoin analysis.


THANK YOU!