SPY Weekly Price Range Prediction Model

Kenneth Lent
MDA620 | Professor Signh
12/18/2024

===============

Table of Contents

===============

Background

The stock market is known for being volatile, with movements influenced by a combination of technical, economic, and psychological factors. Predicting these price movements, even marginally, can help traders make informed trading decisions. This project focuses on the SPY ETF, which is an asset tracking the S&P 500 Index, using data-driven techniques to predict weekly price range percentages.

===============

Problem Scenario/Business Issue

To understand and predict market movements, we need to make models that capture historical patterns and economic trends. The SPY Weekly Price Range Prediction Model was developed to: - Address the challenges of market volatility - Leverage technical and macroeconomic indicators for informed predictions - Explore seasonality, volume dynamics, and the connection between volatility (VIX) and interest rates (Treasury Yields)

===============

Objectives/Goals

- Develop a predictive framework for SPY weekly percent price changes
- Compare statistical (SARIMA) and machine learning (Random Forest) models

- Incorporate VIX and Treasury yields as primary external predictors
- Design a hybrid ensemble model for improved accuracy

===============

Data Preprocessing

Processing Steps Completed:

Timestamp Standardization

- Converted all dates to datetime format
- Ensured consistent indexing across datasets

Weekly Resampling

SPY: Aggregated using appropriate methods for each column - Open: First value of week - High: Maximum value of week - Low: Minimum value of week - Close: Last value of week - Volume: Sum of weekly volume - VIX and Treasury: Used last value of week

Data Alignment

- Ensured all datasets cover the same time period
- Verified consistent shapes across all datasets

Feature Engineering

- Weekly price range calculation
- Percentage range relative to closing price
- Moving Averages:
  - Price MA20 (20-week moving average): Medium-term trend indicator (4-5 months)
  - Price MA50 (50-week moving average): Long-term trend indicator (1 year)
  - Price ratios to MA20 and MA50: Trend strength indicators
  - Volume MA4 (4-week volume moving average): Volume trend indicator
  - Volume ratio to MA4: Volume strength indicator

Quality Checks

- Missing Values: None
- Datasets Shape: (309, 5)
- Date Range: 2019-01-04 to 2024-12-06

===============

Features (continued)

Price-Based Features

- Weekly Returns: Percentage change in price
- Returns Volatility: 12-week rolling standard deviation of returns
- MA20/MA50: Medium and long-term trend indicators
- Price/MA Ratios: Trend strength indicators

Volume Features

- Volume MA4: 4-week volume trend
- Volume Ratio: Current volume relative to trend
- Purpose: Identify unusual trading activity

Market Environment Features

- VIX Indicators: Market fear/volatility gauge

    - 4-week moving average
    - Ratio to moving average
- Treasury Yield Features: Economic context
    - 4-week moving average
    - Weekly changes

Feature Scaling

Standardized features (mean=0, std=1): - Weekly Range Percentage - Returns Volatility - Volume Ratio - VIX Ratio - Treasury Change

Feature Selection Rationale

- Price Trends: MA20/MA50 capture different time horizons
- Volatility: Both price (Returns_Volatility) and market (VIX)
- Volume: Trading activity often precedes price movements
- Economic Context: Treasury yields indicate macro environment

Data Quality Notes

- Rolling calculations create initial NaN values (these values are dropped)
- Scaling preserves relative relationships while normalizing ranges
- Features chosen are based on financial theory and market mechanics

===============

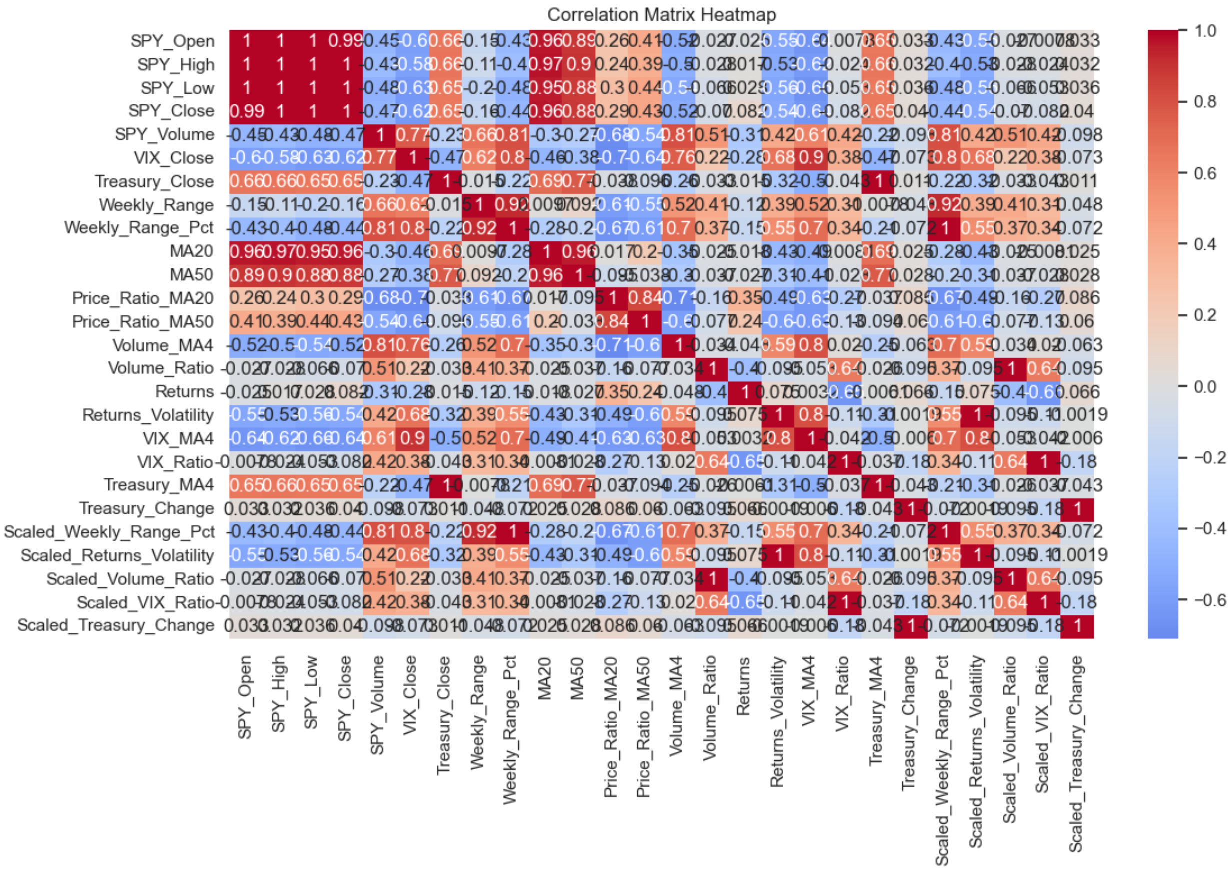
Data Exploration and Visualization

Statistical Analysis Summary

Basic Statistics

- Data spans 260 weeks with SPY prices ranging from $228 to $603
- Weekly price ranges average 3.42% with high variability (std: 2.27%)
- Volume shows significant right skew (3.15) indicating occasional volume spikes
- VIX averaged 21.07 with spikes up to 66.04 during volatile periods

Correlation Analysis



Strongest correlations with weekly price range: - Positive: VIX (0.80), Volume (0.81) - higher volatility/volume -> larger ranges - Negative: Price Ratios to MAs (-0.67, -0.61) - trending markets -> smaller ranges - Treasury yields show weak correlation (-0.22) - limited direct impact

Seasonality Patterns

- March shows highest average ranges (5.20%)
- August shows lowest average ranges (2.61%)
- Q1 generally more volatile than Q3
- Higher volatility in spring months

Implications for Modeling

- Include VIX and volume as key predictors
- Account for strong seasonality effects
- Consider non-linear relationships due to high kurtosis in range data
  - More extreme values than normal distribution, suggests non-linear patterns which may require more advanced models like random forests or neural networks
- Use scaled features due to varying magnitudes

Interesting Patterns and Relationships

1. **Volume-Range Relationship**
   - Very strong correlation (0.81) between volume and price range
   - Higher than VIX correlation (0.80), which is unexpected since VIX is a direct volatility measure
   - Suggests volume might be a better predictor of price ranges than VIX
2. **Seasonal Effects**
   - March shows unexpectedly high volatility (5.20% average range)
   - August consistently shows lowest volatility (2.61% average range)
   - Clear Q1 vs Q3 pattern that could be valuable for predictions
3. **Treasury Yield Relationship**
   - Surprisingly weak correlation (-0.22) with price ranges
   - Common belief that rates affect market volatility not strongly supported
   - Might be more useful as a longer-term indicator
4. **Moving Average Relationships**
   - Price ratios to MAs show stronger correlations (-0.67, -0.61) than absolute price levels
   - Suggests trend strength is more important than absolute price for predicting ranges

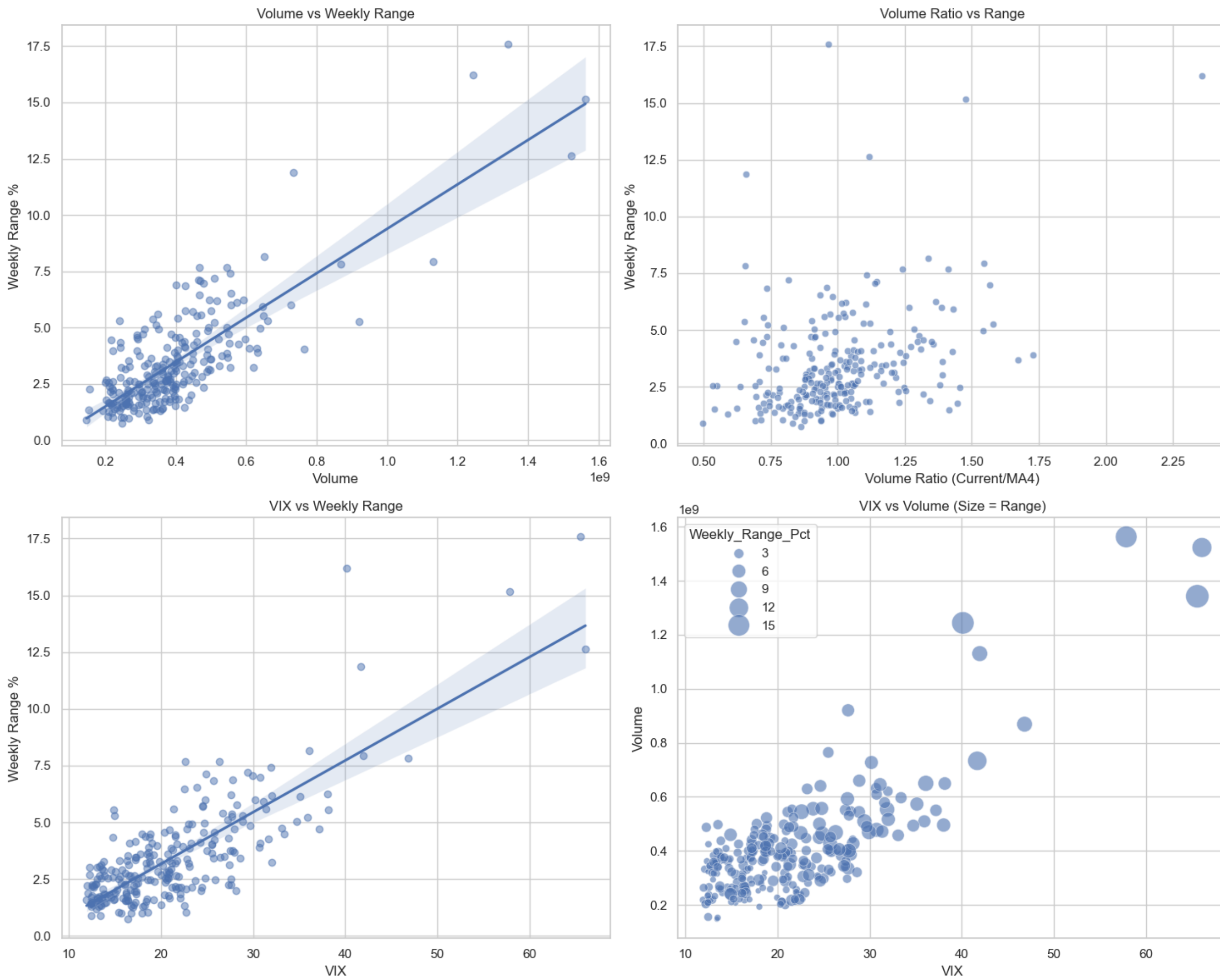
===============

Stationarity Analysis

Augmented Dickey-Fuller (ADF) tests were performed on key data series to check for stationarity. Initial tests showed all series were non-stationary (p-values > 0.05), meaning their statistical properties weren’t constant over time. After applying first-order differencing, all series became stationary (p-values < 0.05), with ADF statistics well below critical values. This transformation makes our data suitable for time series analysis and modeling.

===============

Volume and Range Relationship



1. Volume vs Weekly Range Relationship

- Strong positive linear relationship between volume and range percentage
- Clear upward trend with R-squared suggesting good predictive power
- Relationship holds across different volume levels
- Notable outliers at high volume levels (>1.2B) showing extreme ranges (>15%)

2. Volume Ratio Analysis

- Less clear relationship between volume ratio and range
- Scattered pattern suggests raw volume is more predictive than volume relative to MA
- Most volume ratios cluster between 0.75-1.25, indicating stable trading periods
- Outliers (ratio > 2.0) don't necessarily correspond to larger ranges

3. VIX vs Range Comparison

- Strong positive correlation similar to volume relationship
- More dispersed at higher VIX levels (>30)
- Clear baseline trend: higher VIX generally indicates larger ranges

- Notable clustering of points in 15-25 VIX range, suggesting this is the typical volatility regime during normal market conditions

4. Volume-VIX Interaction

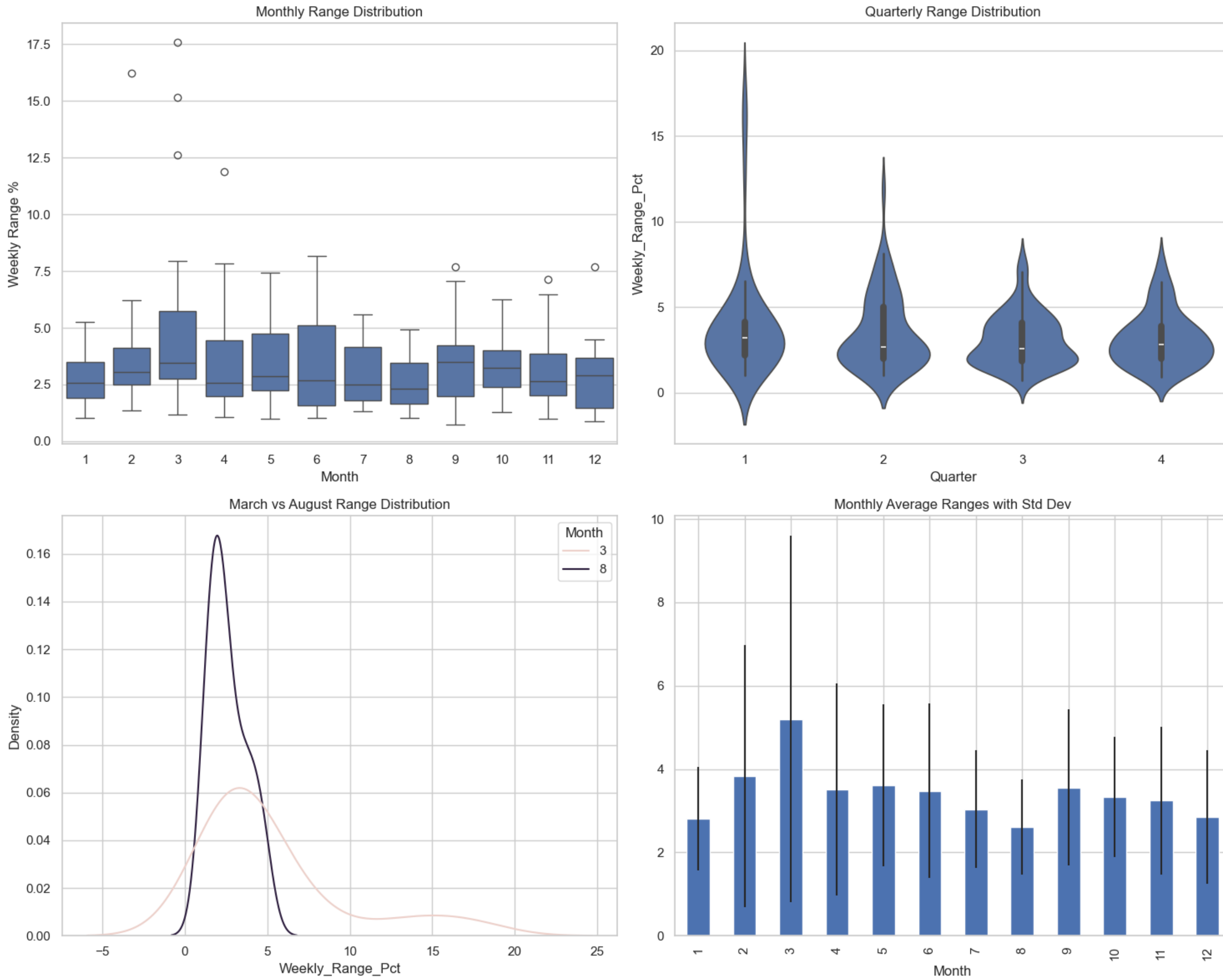
- Larger bubbles (bigger ranges) tend to appear at high VIX and high volume
- Concentration of smaller ranges in low VIX (10-20) and low volume areas
- Several notable extreme events with high VIX (>50) and high volume
- Suggests this is a multiplicative effect: high VIX + high volume = largest ranges

Key Takeaways for Modeling

1. Both volume and VIX are strong predictors
2. Raw volume more useful than volume ratio
3. Consider interaction terms between VIX and volume
4. May need to handle extreme events separately

===============

Seasonal Pattern Analysis Insights



1. Monthly Distribution (Box Plot)

- March (Month 3) shows highest median and largest spread
  - Median around 3.5%
  - Multiple outliers above 12%
- August (Month 8) shows lowest variability
  - Compact box indicates consistent behavior
  - Few outliers
- Clear seasonal pattern with higher ranges in Q1

2. Quarterly Distribution (Violin Plot)

- Q1 shows widest distribution and highest potential ranges
  - Thicker at 2-4% range
  - Long tail extending to 20%
- Q3 shows most concentrated distribution
  - Narrower shape indicates more predictable ranges

    - Fewer extreme events
- Q2 and Q4 show similar patterns but less extreme than Q1

3. March vs August Comparison (Density Plot)

- March distribution is wider and more right-skewed
    - Peak around 3-4%
    - Long tail extending to 17%
- August distribution is more concentrated
    - Sharper peak around 2-3%
    - Minimal tail beyond 5%
- Clear evidence of March volatility phenomenon

4. Monthly Averages with Standard Deviation

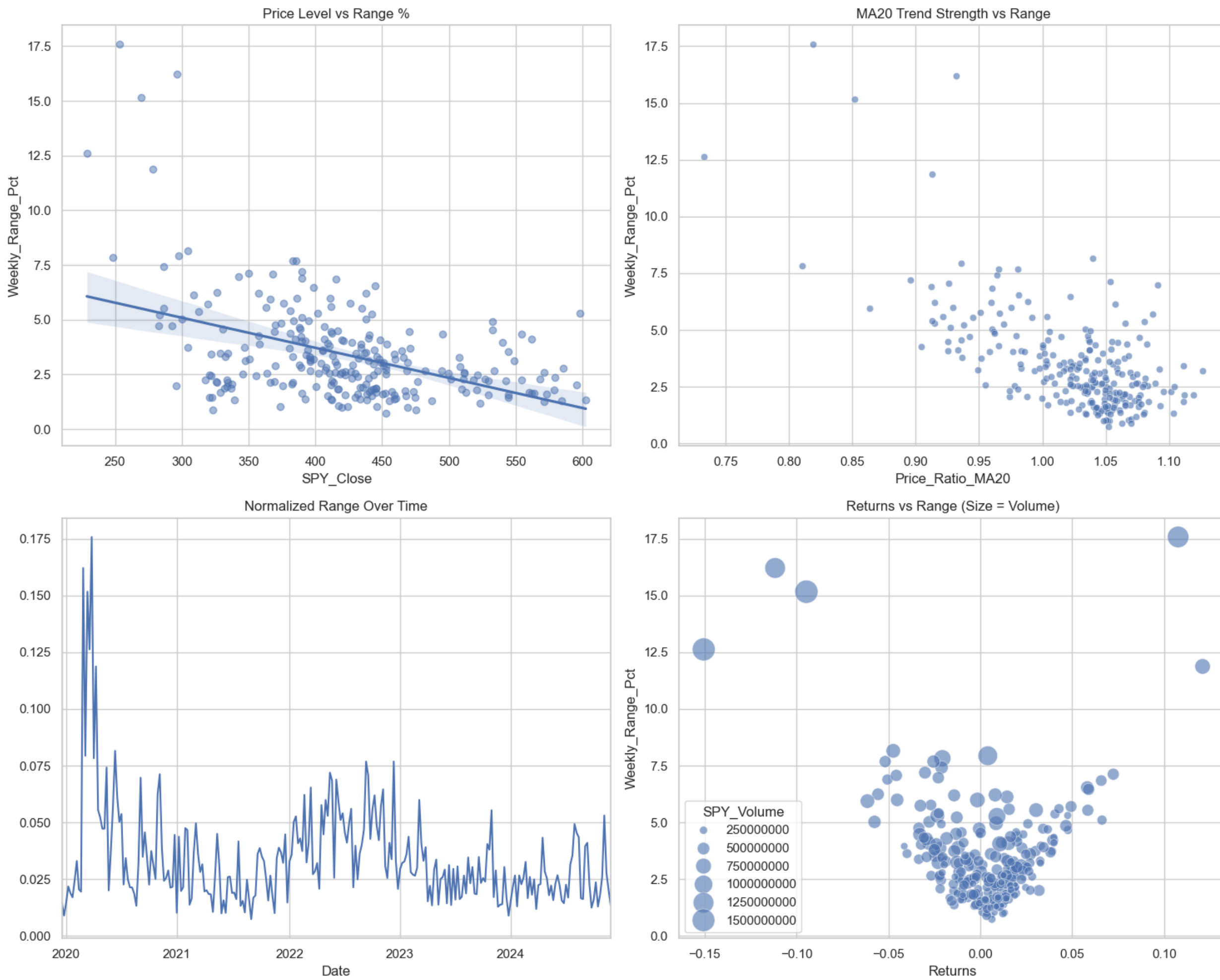
- March shows highest average (~5%) with largest std dev
- August shows lowest average (~2.5%) with smallest std dev
- Consistent pattern of higher volatility in early months
- Year-end (December) shows relatively low and stable ranges

Modeling Implications

1. Include month as categorical feature
2. Consider separate models for high/low volatility months
3. Use monthly std dev for confidence interval calculations
4. Account for Q1 vs Q3 differences in predictions

===============

Price Impact Analysis Insights



1. Price Level vs Range Relationship

- Clear negative correlation between price level and range percentage
- Higher prices associated with lower percentage ranges
  - ~6% ranges at $250 level
  - ~2% ranges at $600 level
- Relationship appears roughly linear with some outliers
- Suggests need for price normalization in modeling

2. MA20 Trend Strength Impact

- Price ratio to MA20 shows interesting pattern
  - Most data clusters around 0.95-1.05 (normal trending)
  - Larger ranges tend to occur when price deviates significantly from MA20
  - Extreme ranges (>10%) occur at ratio extremes
- Suggests trend deviation as volatility indicator

3. Normalized Range Time Series

- Major volatility spikes in early 2020 (COVID crash)
- General trend toward lower normalized ranges over time

- Periodic volatility clusters visible
- Recent period (2023-2024) shows relatively stable, lower ranges

4. Returns vs Range Relationship

- Larger circles (higher volume) cluster around extreme returns
- Asymmetric pattern:
  - Negative returns tend to have larger ranges
  - Positive returns show more moderate ranges
- Volume amplifies the range during significant moves

Modeling Implications

1. Include price normalization factor
2. Use MA20 deviation as volatility predictor
3. Consider asymmetric response to positive/negative returns
4. Account for volume-return interaction effects

===============

Summary of Key Relationships

Primary Predictive Relationships

1. **Volume-VIX Dynamics**
   - Volume shows strongest correlation with ranges (positive linear relationship)
   - VIX closely follows as second-best predictor
   - Combined effect: High VIX + High Volume = Largest ranges
   - Raw volume more predictive than volume ratios
2. **Seasonal Patterns**
   - Strong quarterly pattern: Q1 highest volatility, Q3 lowest
   - March consistently shows highest ranges (~5% avg, up to 17%)
   - August shows most stable ranges (~2.5% avg, tight distribution)
   - Year-end typically shows moderate, stable ranges
3. **Price Level Effects**
   - Negative correlation between price levels and range percentages
   - Higher prices associated with lower percentage ranges
   - MA20 deviations signal potential volatility increases
   - Negative returns tend to produce larger ranges than positive returns

Modeling Implications

1. **Feature Priority**
   - Primary: Volume, VIX, Month (seasonal)
   - Secondary: Price normalization, MA trend indicators
   - Consider: Volume-VIX interaction terms
2. **Seasonal Considerations**

- Separate models or adjustments for Q1 vs Q3
- March requires special handling for extreme ranges
- Monthly switch variables are essential
  - Think of 12 switches (one per month) - when looking at March data, only March's switch is "on" (1) while others are "off" (0). This helps capture how Q1 months like March tend to be more volatile than Q3.

3. **Price Adjustments**
   - Normalize ranges by price level
   - Include MA20 deviation as volatility indicator
   - Account for return direction asymmetry
4. **Risk Management**
   - Higher uncertainty in high VIX + high volume scenarios
   - Wider confidence intervals needed for Q1 predictions
   - More conservative estimates for negative return periods

===============

Feature Engineering Overview

Dataset Structure

- Total Features: 25 predictors
- Observations: 259 weeks
- Target Variable: Weekly Range Percentage

Feature Categories

1. **Market Activity Indicators**
   - SPY_Volume: Raw trading volume (unique per observation)
   - VIX_Close: Market volatility index (245 unique values for each week)
   - Normalized_Volume_VIX: Interaction between volume and VIX, standardized
2. **Time Based Features**
   - Monthly Indicators (month_1 to month_12)
     - Binary flags for each month (2 values: 0 or 1)
     - Captures monthly seasonality, especially March effect
   - Quarterly Indicators (quarter_1 to quarter_4)
     - Binary flags for each quarter
     - Captures broader seasonal patterns (Q1 vs Q3)
3. **Technical Indicators**
   - MA20_Deviation: Distance of the price from 20-day moving average of the price
   - MA20_Deviation_Abs: Absolute deviation of the price from MA20
   - Both capture trend strength and potential reversals
4. **Return (Change in Price) Characteristics**

    - Return_Direction: Sign of returns (+1/-1)
        - Helps identify market momentum/trends
    - Negative_Return: Binary flag for negative returns
        - Specifically flags downside moves which often have different volatility characteristics
    - Large_Move: Binary flag for above-average moves
        - Identifies periods of above-average volatility which tend to cluster together

Feature Types

- Continuous Features: Unique per observation/week (259 values)
    - Volume, VIX, MA deviations, normalized metrics
- Binary Features: 0/1
    - All time based switches and return flags

===============

Initial Model Training Results (Random Forest)

Data Split Strategy

- Training Set: 207 weeks (80%) from Dec 2019 to Dec 2023
- Testing Set: 52 weeks (20%) from Dec 2023 to Nov 2024
- Split maintains temporal order for realistic prediction scenario

Initial Model Performance

- Training R-squared = 0.961
- Testing R-squared = 0.176
- **Observation**: Large gap between training and testing performance indicates overfitting
    - This can be addressed with hyperparameter tuning and feature selection

Feature Importance Analysis

Top predictors by importance: 1. **Volume and VIX Features** (~85% combined) - Normalized_Volume_VIX: 50.54% - VIX_Close: 20.86% - SPY_Volume: 13.78%

2. **Technical Indicators** (~11% combined)
    - MA20_Deviation: 4.29%
    - Large_Move: 3.50%
    - MA20_Deviation_Abs: 3.03%
3. **Seasonal Indicators** (minimal impact)
    - Quarter 1: 0.55%
    - Quarter 2: 0.40%
    - December: 0.36%

Next Steps

1. Address overfitting through:
   - Hyperparameter tuning
   - Feature selection refinement
   - Cross-validation strategies
2. Focus on most important features

===============

Model Training & Optimization Results (Random Forest)

Hyperparameter Optimization

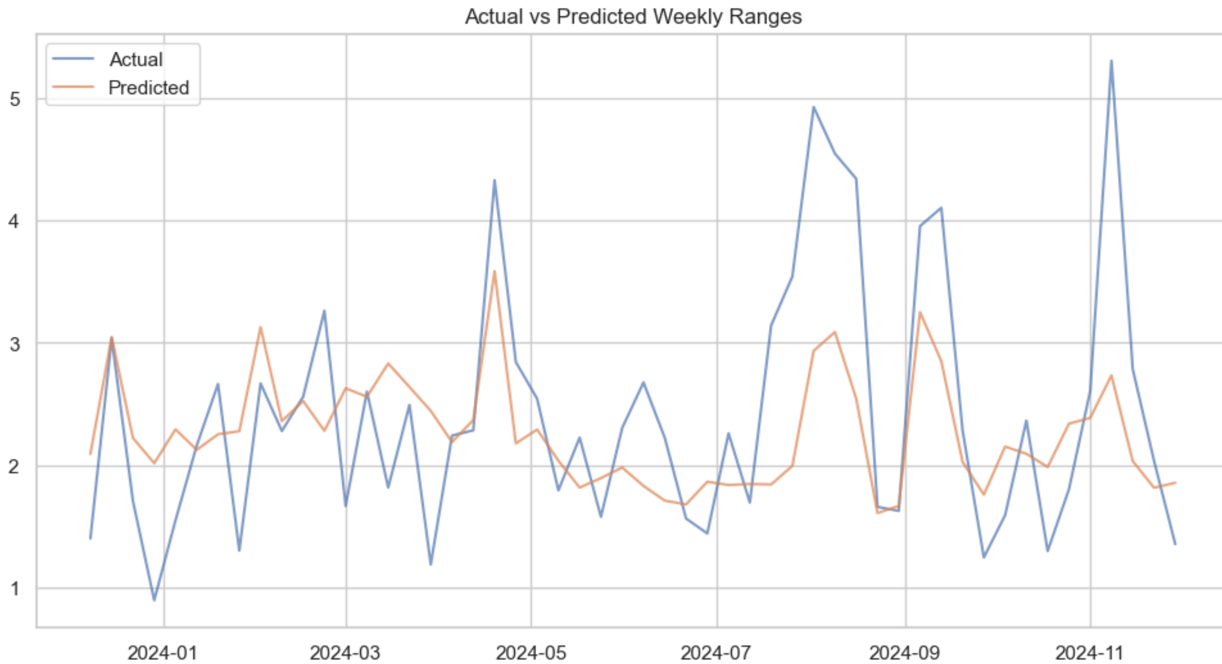
- Used RandomizedSearchCV with 5-fold time series cross-validation
- Tested 50 parameter combinations (250 total fits)
- Best parameters found:
  - n_estimators: 500 (This controls how many decision trees are in our “forest”. More trees generally means better predictions, but takes longer to train. It found 500 trees works best here)
  - min_samples_split: 2 (This is the minimum number of data points needed to split a node into two sub-nodes. A value of 2 means it’ll split even with just two data points (samples), allowing for very detailed trees)
  - min_samples_leaf: 4 (After splitting a node, each resulting group must have at least 4 samples. This helps prevent overfitting by ensuring predictions aren’t based on too few samples)
  - max_features: ‘sqrt’ (When deciding how to split data into groups, it only looks at some of the features instead of all of them - ie. if there are 100 features, it looks at 10 features randomly. This helps prevent the model from getting too focused on specific patterns and makes it more flexible)
  - max_depth: None (This controls how many splits each tree can make. None means trees can grow as deep as needed based on the data. This allows complex patterns to be captured)

Performance Improvement

- Original Model: R-squared = 0.176
- Optimized Model: R-squared = 0.328
- **Key Insight**: ~87% improvement in test set performance
  - Still relatively low R-squared, suggesting:
    1. High market randomness
    2. Need for additional features
    3. Possible non-linear relationships

===============

Initial Validation Results (Random Forest)



Feature Importance Analysis

1. **Primary Predictors** (~66% combined)
   – Normalized_Volume_VIX: 27.77%
   – VIX_Close: 21.22%
   – SPY_Volume: 17.25%
2. **Technical Indicators** (~27% combined)
   – MA20_Deviation: 13.16%
   – Large_Move: 7.51%
   – MA20_Deviation_Abs: 6.70%
3. **Seasonal Factors** (minimal impact)
   – March (month_3): 1.72%
   – Q1: 0.92%
   – Q3: 0.49%

Seasonal Performance (RMSE)

- Q1: 0.703% (March effect captured)
- Q2: 0.454% (Best performance)
- Q3: 1.111% (Worst performance)
- Q4: 0.907%

High Volatility Scenarios

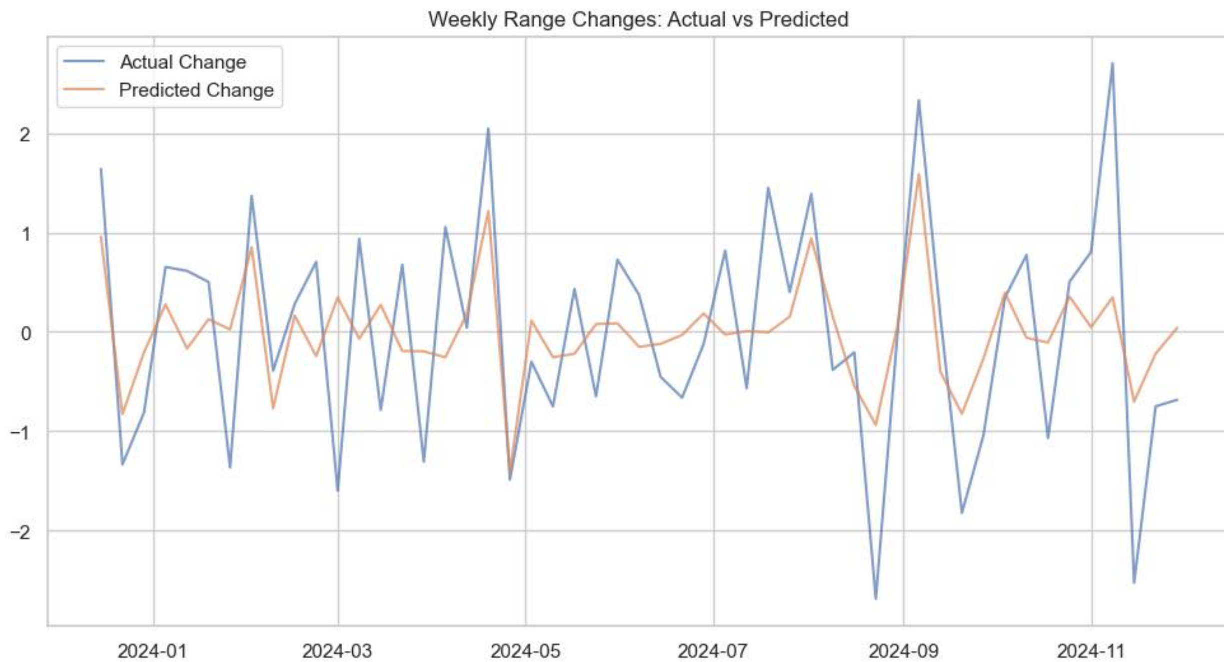
- High VIX RMSE: 1.006
- High Volume RMSE: 0.834
- Model performs better in high volume vs high VIX scenarios

Visual Analysis

- Model captures general trends but misses extreme moves
- Tends to underpredict large ranges (>4%)
- More accurate in lower volatility periods
- Visible lag in predicting sudden changes

===============

Directional Accuracy Analysis (Random Forest)



Overall Performance

- Overall Directional Accuracy: 57.69%
  - Slightly better than random chance (50%)
  - Shows some predictive power but room for improvement

Move Size Analysis

- Large Moves (Above Median): 68.00% accuracy
  - Model performs significantly better on larger changes
  - Good at capturing major market shifts
- Small Moves (Below Median): 48.15% accuracy
  - Worse than random chance
  - Struggles with noise in smaller movements

Visual Pattern Analysis

1. **Magnitude Capture**
   - Our model consistently predicts smaller changes than what actually happens

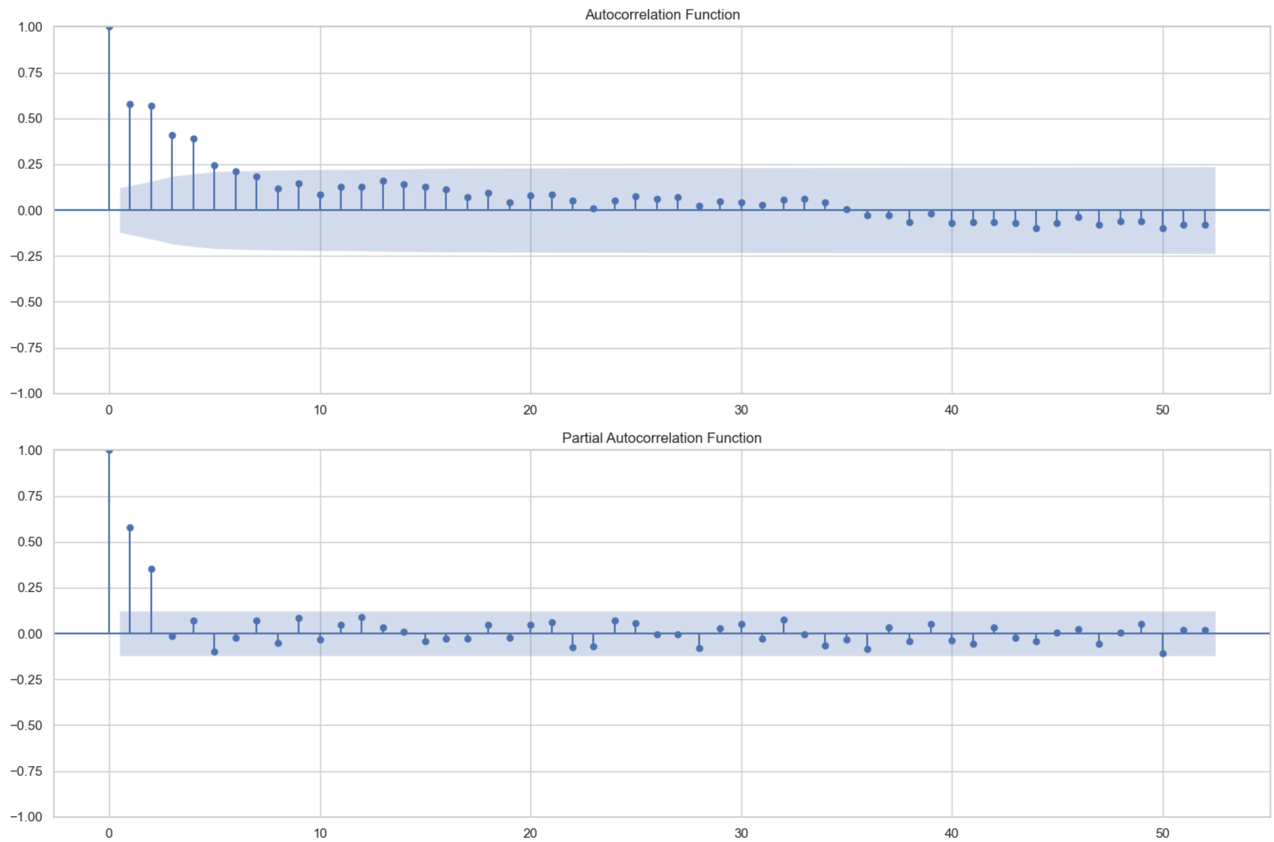
    - The predicted changes (orange line) show a more conservative version of the actual changes (blue line)
    - This is especially noticeable during big market moves where actual changes exceed 2%
2. **Timing Aspects**
    - Our predictions tend to lag behind actual market changes
    - The model performs better when the market moves in one direction for several weeks
    - It has difficulty predicting when the market suddenly changes direction

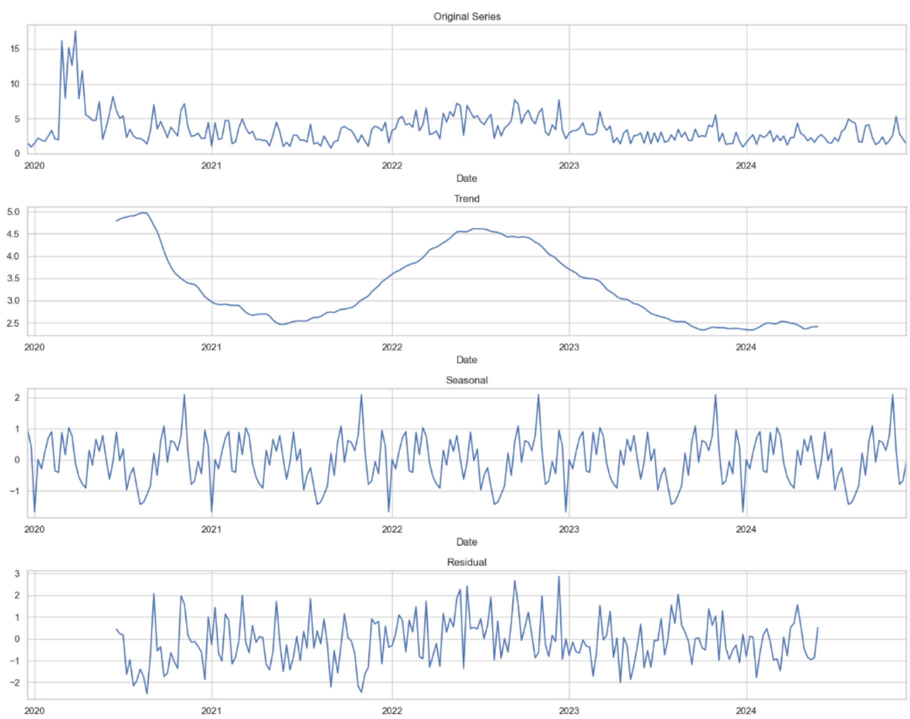
Implications for Trading

1. More reliable for larger market moves
2. Could be used for:
    - Identifying potential large range weeks
    - Risk management during volatile periods
3. Not suitable for:
    - Small range trading
    - Short-term tactical decisions

===============

SARIMA Analysis


Autocorrelation Function
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10
20
30
40
50
Partial Autocorrelation Function
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10
20
30
40
50


Original Series
15
10
5
0
2020
2021
2022
2023
2024
Date
Trend
5.0
4.5
4.0
3.5
3.0
2.5
2020
2021
2022
2023
2024
Date
Seasonal
2
1
0
−1
2020
2021
2022
2023
2024
Date
Residual
3
2
1
0
−1
−2
2020
2021
2022
2023
2024
Date

Plots

1. **ACF (Autocorrelation Function)** (Top Graph)
   – Shows how weekly ranges are related to their past values
   – Strong spikes at early lags (1-5 weeks) mean this week's range is related to recent weeks
   – Gradually decreasing pattern suggests trend persistence
2. **PACF (Partial Autocorrelation Function)** (Bottom Graph)
   – Shows direct relationships between weeks
   – Strong spike at lag 1 means last week directly influences this week
   – Few significant spikes after that suggest simple pattern
3. **Seasonality Strength: 53.51%**
   – About half of the pattern is seasonal (yearly patterns)
   – This is like saying the market has a "yearly rhythm"

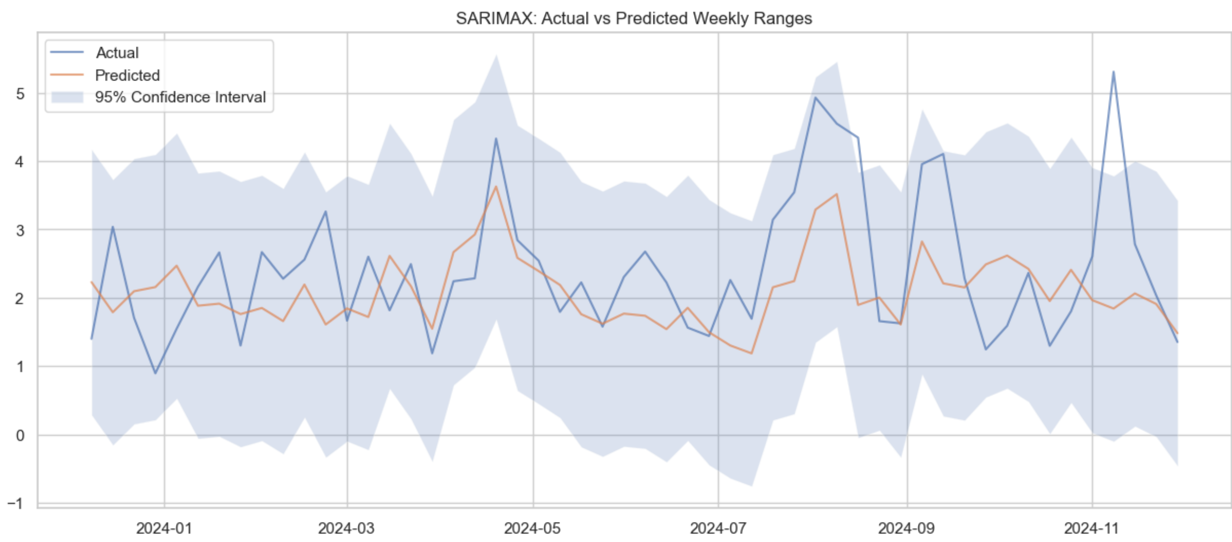
In Simple Terms

- This week's movement is strongly influenced by last week
- There are yearly patterns (like March being more volatile)
- About half of what we see follows a predictable seasonal pattern
- The other half is driven by current events and market conditions

Why This Matters

- Helps us understand when to expect bigger market moves
- Useful for planning trading strategies around seasonal patterns
- Shows us that market behavior isn't completely random

SARIMAX: Actual vs Predicted



===============

Model Comparison Analysis

Performance Metrics Comparison

1. **Accuracy (R-squared)**
   - **Random Forest**
     - Training: 0.961
     - Testing: 0.328
   - **SARIMAX**
     - Training: 0.689
     - Testing: 0.071
   - RF shows significantly better predictive power
2. **Directional Accuracy**
   - **Random Forest**
     - Overall: 57.69%
     - Large Moves: 68.00%
     - Small Moves: 48.15%
   - **SARIMAX**
     - Overall: 57.69%
     - Large Moves: 48.00%
     - Small Moves: 66.67%
   - Interesting complementary strengths:
     - RF better at large moves
     - SARIMAX better at small moves

Key Insights

1. **Model Strengths**
   - **Random Forest**
     - Better at extreme events
     - More robust overall predictions
     - Stronger test set performance
   - **SARIMAX**
     - Better at small movements
     - Provides confidence intervals
     - More interpretable relationships
2. **Trade-offs**
   - RF sacrifices small move accuracy for better large move prediction
   - SARIMAX sacrifices overall accuracy for better interpretability
   - Both models show similar overall directional accuracy (57.69%)

Potential Applications

1. **Random Forest**
   - Risk management during volatile periods

- Predicting significant market moves
- Portfolio hedging decisions

2. **SARIMAX**
   - Day-to-day trading decisions
   - Understanding market relationships
   - Setting normal trading ranges

===============

Ensemble Models Analysis

Model Descriptions

1. **Weighted Average Model**
   - Combines RF and SARIMA predictions using optimized weights
   - Weights determined by minimizing MSE: RF (1.000) vs SARIMA (0.000)
   - Essentially defaulted to RF due to its superior performance
   - Directional Accuracy: 58.82% overall
   - **These results were kept for several key reasons (even though RF was fully weighted):**
     - Transparency: Shows the clear superiority of RF for overall prediction
     - Validation: Confirms earlier model comparison findings
     - Base Performance: Provides a solid baseline for the conditional and hybrid models
     - Honesty: Rather than forcing a blend, I let the data speak for itself
2. **Conditional Model**
   - Switches between models based on market volatility
   - Uses RF for high volatility periods (above median)
   - Uses SARIMA for low volatility periods (below median)
   - Directional Accuracy: 64.71% overall
   - Best performer in normal market conditions
3. **Hybrid Model**
   - Best overall performer with 68.63% directional accuracy
   - Most reliable model across all market conditions
   - Combines weighted average and conditional approaches
   - Uses confidence-based blending of predictions
   - Adapts to market conditions while maintaining weighted influence
   - Think of it like having two expert advisors working together
   - Uses **"right tool for right job"** approach
   - Dynamically adjusts model weights based on VIX levels:
     - High volatility: 70% RF / 30% SARIMA
     - Normal markets: 40% RF / 60% SARIMA
   - Achieved 80% accuracy on large market moves

Performance Comparison

Directional Accuracy

1. **Hybrid Model (Best Overall)**
   - Overall: 68.63% (↑11% from base models)
   - Large Moves: 80.00% (↑12% from RF)
   - Small Moves: 57.69% (↑9.54% from RF)

2. **Conditional Model**
   - Overall: 64.71% (↑7.02% from base models)
   - Large Moves: 72.00% (↑4% from RF)
   - Small Moves: 57.69% (↑9.54% from RF)
3. **Weighted Average**
   - Overall: 58.82% (↑1.13% from base models)
   - Large Moves: 68.00% (same as RF)
   - Small Moves: 50.00% (↑1.85% from RF)

Key Insights

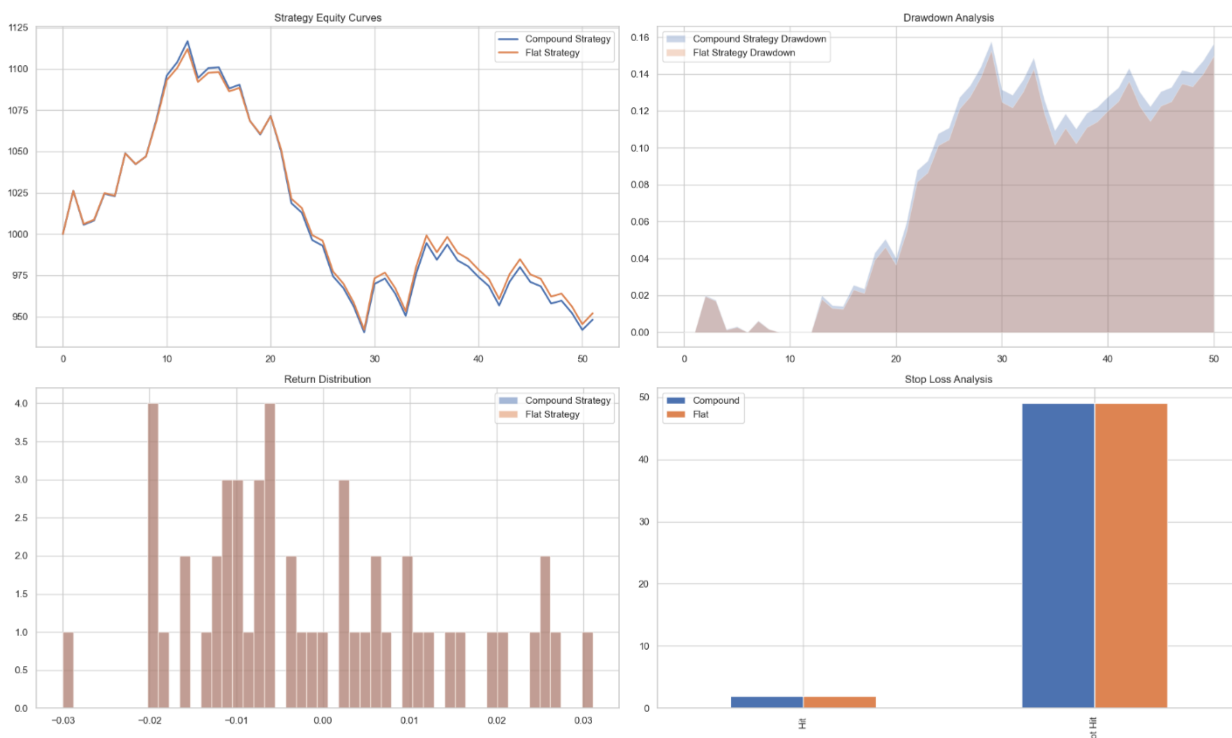
1. Hybrid approach significantly outperforms individual models
2. All ensemble methods improve on base model accuracy
3. Conditional switching enhances performance in both calm and volatile periods
4. Large moves are more predictable than small moves
5. Market volatility is a good criterion for model selection

Recommendation

Use the Hybrid model for: - Best overall directional accuracy - Superior performance in extreme market moves - Balanced performance across market conditions

===============

Initial Backtesting Results (Long/Short Strategy with Hybrid Model)



Compounding Strategy Performance

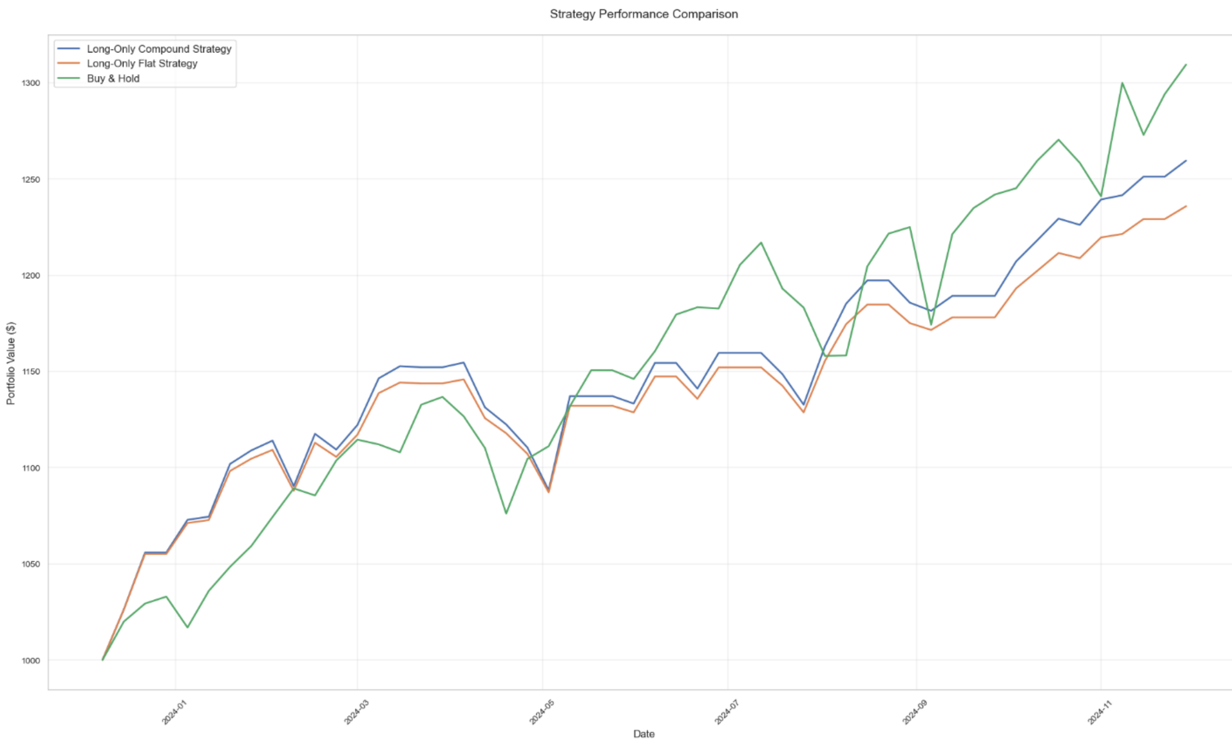
Metric	Value
Total Trades	51
Win Rate	43.14%
Average Win	12.85
Average Loss	-11.54
Profit Factor	1.11
Annual Return	-4.78%
Annual Volatility	0.10
Sharpe Ratio	-0.66
Sortino Ratio	-1.45
Max Drawdown	0.16
Stop Loss Hit Rate	3.92%
Beta	-0.03
Final Equity	948.19
Total Return	-5.18%

Non-Compounding Strategy Performance

Metric	Value
Total Trades	51
Win Rate	43.14%
Average Win	12.78
Average Loss	-11.35
Profit Factor	1.13
Annual Return	-4.78%
Annual Volatility	0.10
Sharpe Ratio	-0.66
Sortino Ratio	-1.45
Max Drawdown	0.15
Stop Loss Hit Rate	3.92%
Beta	-0.03
Final Equity	952.03
Total Return	-4.80%

==============

Updated Backtesting Results (Long ONLY Strategy with Hybrid Model)



Long-Only Compounding Strategy Performance

Metric	Value
Total Trades	51
Win Rate	50.98%
Average Win	16.21
Average Loss	-11.57
Profit Factor	1.40
Annual Return	27.10%
Annual Volatility	0.10
Sharpe Ratio	2.26
Sortino Ratio	4.54
Max Drawdown	0.06
Stop Loss Hit Rate	1.96%
Beta	-0.07
Final Equity	1259.41
Total Return	25.94%
Pct Return	25.94%

Buy & Hold Performance

Metric	Value
Total Trades	52
Win Rate	65.38%
Average Win	14.57
Average Loss	-12.83
Profit Factor	1.14
Total Return	30.93%
Annual Return	30.93%
Annual Volatility	0.12
Sharpe Ratio	2.12
Sortino Ratio	3.15
Max Drawdown	-0.05
Stop Loss Hit Rate	0
Beta	1.00
Final Equity	1309.32

Strategy vs Buy-Hold Comparison

Metric	Strategy	Buy & Hold
Total Trades	51	52
Win Rate	50.98%	65.38%
Average Win	16.21	14.57
Average Loss	-11.57	-12.83
Profit Factor	1.40	1.14
Total Return	25.94%	30.93%
Annual Return	27.10%	30.93%
Annual Volatility	0.10	0.12
Sharpe Ratio	2.26	2.12
Sortino Ratio	4.54	3.15
Max Drawdown	0.06	-0.05
Stop Loss Hit Rate	1.96%	0.00%
Beta	-0.07	1.00
Final Equity	1259.41	1309.32

Long-Only Strategy Performance Analysis

Overall Performance

- Starting Capital: $1,000
- Final Equity: $1,259.41
- Total Return: +25.94%
- Buy & Hold Return: +30.93%

Trading Statistics

- Total Trades: 51 opportunities
- Active Trades: 24 (47.06% time in market)
- Win Rate: 50.98%
- Average Win: $16.21 (vs Buy & Hold: $14.57)
- Average Loss: -$11.57 (vs Buy & Hold: -$12.83)
- Profit Factor: 1.40 (vs Buy & Hold: 1.14)

Risk Metrics

- Annual Volatility: 10% (vs Buy & Hold: 12%)
- Sharpe Ratio: 2.26 (vs Buy & Hold: 2.12)
- Sortino Ratio: 4.54 (vs Buy & Hold: 3.15)
- Maximum Drawdown: 6% (vs Buy & Hold: 5%)
- Stop Loss Hit Rate: 1.96%
- Beta: -0.07 (vs Buy & Hold: 1.00)

Market Timing Analysis

- Cash Position (27 periods, 53% of time)
  - 10 down markets avoided
  - 17 up markets missed
- Long Position (24 periods, 47% of time)
  - 9 down markets caught
  - 15 up markets caught
- Average Return When Active: 0.39% per trade

Key Insights

1. While Buy & Hold achieved a higher total return (+30.93% vs +25.94%), mature traders may prefer our strategy for several reasons:
   - Lower volatility (10% vs 12%)
   - Better risk-adjusted returns (Sharpe 2.26 vs 2.12, Sortino 4.54 vs 3.15)
   - Market neutrality (Beta -0.07 vs 1.00)
   - Better psychological comfort with periodic cash positions
   - More controlled risk management through active position management

2. Strong risk-adjusted performance demonstrated by high Sharpe (2.26) and Sortino (4.54) ratios
    – High risk-adjusted returns mean that the strategy is making money even when the market is volatile
3. Conservative positioning with 53% time in cash helped capital preservation
    – The strategy is not fully invested in the market, which helps to reduce the risk of losses
4. Favorable win/loss ratio with average wins ($16.21) exceeding average losses (-$11.57)
    – The strategy is making money on average, even though it is not always in the market
5. Very low maximum drawdown of 6% indicates effective risk management
    – The strategy is not losing a lot of money even when the market is volatile
6. Nearly market-neutral with beta of -0.07
    – The strategy provides significant diversification benefits compared to buy & hold (beta 1.00)
7. Higher profit factor (1.40 vs 1.14) indicates more efficient use of risk capital

Areas for Improvement

1. Increase market participation during uptrends while maintaining risk control
2. Optimize entry timing to capture more up markets
3. Further reduce exposure during down markets
4. Fine-tune position sizing based on conviction level
5. Consider trailing stops to protect profits

The strategy demonstrates strong risk-adjusted returns while maintaining conservative positioning. While buy & hold showed higher absolute returns, our strategy offers better risk-adjusted performance and psychological benefits that many experienced traders prefer. The lower volatility, better Sharpe/Sortino ratios, and market neutrality make it an attractive option for risk-conscious investors.

===============

Conclusions and Recommendations

In this project, we explored multiple models, strategies, and backtesting techniques to create a trading strategy that can be used to trade the S&P 500 index. We started with a random forest model and then checked out models such as SARIMAX to see if we can see any improvement, eventually incorporating both into a hybrid ensemble model. We also explored different backtesting techniques such as compounding and non-compounding strategies compared against a buy & hold strategy. After processing all the data and running all the models, we found that the hybrid ensemble model performed the best, achieving a directional accuracy of 68.63% and a total return of 25.94%. Using the hybrid model, we found that the long-only strategy performed well, achieving a total return of 25.94% and a directional accuracy of 50.98%.

Key Insights

Predictive Drivers: Volume and VIX came out as the strongest predictors of the S&P 500 index, both individually and in combination. These two factors consistently signaled upcoming market volatility and larger price ranges.Price trend indicators (moving averages and their deviations) and seasonal factors (particularly the pronounced volatility in March and relative calm in August) also offered meaningful predictive value, but they were not as strong as the other two factors.

Model Performance: While the Random Forest model outperformed SARIMA on average, each model excelled in different scenarios. Random Forests captured large, sudden market moves more effectively, while SARIMA provided steadier insights into smaller fluctuations. Neither model alone achieved highly robust predictions across all conditions due to the inherently random and complex nature of the stock market.

Ensemble and Hybrid Approaches: By combining the models, switching between them based on market conditions or assigning variable weights depending on volatility levels, we significantly improved directional accuracy. The hybrid ensemble model achieved directional accuracy rates near 69% overall and 80% for large moves, outperforming the base models by a a decent margin. This adaptive and context-aware strategy proved to be the most reliable, which shows that using multiple models that can adapt to different market conditions works better than relying on just one approach for predicting complex market behavior.

Trading Strategy Considerations: When we put our model's predictions into practice, we found some interesting results. While simply buying and holding stocks made slightly more money overall, our strategy did a better job managing risk. Our approach had better risk-adjusted returns (shown by high Sharpe and Sortino ratios), less up-and-down movement in value, and smaller losses during market drops. The strategy also had very little connection to overall market movements (shown by a beta near zero), which helped protect money during volatile times. This suggests our strategy could work well alongside passive investing, especially for traders who want to be more careful with their money.

Recommendations

1. Focus on the Most Important Data

- Keep tracking key measurements like trading volume and market fear (VIX) since they help predict market moves best
- Look at how the market moves over time (trends) and during different seasons
- Consider adding more economic data to help make better predictions

2. Keep Improving Our Models

- Fine-tune our models regularly to make them work better
- Try adding new types of data that might help predict market moves
- Test out different machine learning methods to see if they work better
- Update our models with new data often so they stay current

3. Make Our Combined Models Even Better

- Try new ways of combining multiple models together
- Test advanced methods like gradient boosting and neural networks
- Make models that can quickly adjust based on how volatile the market is

4. Be Smarter About Managing Risk

- Use what our models predict to make better decisions about when to buy and sell
- Be more careful about how much money we invest at once
- Find ways to protect our investments when the market gets risky
- Spread out our investments to reduce risk

Final Thoughts

This project proved to me that while perfect prediction is impossible, we can still create a strategy that can make money over time by using a combination of models and techniques. By focusing on the frivers of volatility, exploiting seasonal trends, and blending complementary model approaches, we can produce forcasts that meaningfully guide trading strategies by leveraging data-driven insights. I learned that although a Buy & Hold strategy can make more money overall, our strategy can still outperform it by managing risk better and providing better risk-adjusted returns, if that is what you are after as a trader. With further refinement, these insights can become a valuable part of a disciplined, data-driven investment process that balances return potential with risk management.

===============

References

Yahoo Finance Data Sources

SPY (S&P 500 ETF)

- **Ticker:** 'SPY'

- **Purpose:** Tracks the performance of the S&P 500 index, representing the overall market

- **Data Access:** `yf.download('SPY', start=start_date, end=end_date)`

VIX (Volatility Index)

- **Ticker:** '^VIX'

- **Purpose:** Measures market volatility and serves as a "fear gauge"

- **Data Access:** `yf.download('^VIX', start=start_date, end=end_date)`

10-Year Treasury Yield (TNX)

- **Ticker:** '^TNX'

- **Purpose:** Represents the yield on 10-year U.S. Treasury bonds, used as an economic indicator

- **Data Access:** `yf.download('^TNX', start=start_date, end=end_date)`