Leveraging Weather Data for Robusta and Arabica Coffee Trading Strategies

QF621 Sample Report - Group 3

Table of Contents

Executive Summary

Our project investigates how weather data can be used to develop trading strategies in the coffee market. Arabica and Robusta coffee are highly sensitive to weather, especially temperature and rainfall, which are often the reason behind major price disruptions. We wanted to find out whether certain weather patterns could help us spot trading signals early to react to real-world conditions instead of relying on past price movements.

Our goal was to build a quantitative trading model that meaningfully combines financial and environmental data. We created a dataset that merges weekly coffee futures prices with historical weather data from ten important producing regions. We selected these locations carefully to represent various climate regions and their importance in production.

We trained an XGBoost machine learning model to predict 8-week-ahead returns for Arabica and Robusta contracts. The model uses features derived from weather data, such as rolling averages, lagged temperature and rainfall values, and drought indicators, to detect patterns that might signal upcoming changes in supply. To make our backtests more realistic, we added a one-week delay between signal generation and trade execution and followed strict time-series validation to avoid data leakage.

Our weather-based strategy slightly outperformed a simple buy-and-hold strategy and a momentum baseline, especially in the Arabica market. The model also showed strong accuracy during upward-trending markets, correctly forecasting the price direction over 80% of the time in some periods. These results support our idea that weather signals, especially in climate-sensitive regions, can act as early indicators of supply risk and be useful for market timing.

However, we also encountered limitations. The strategy was less effective for Robusta, which may be due to Robusta's lower sensitivity to weather or more unstable price behavior overall. Another challenge was the structure of our trading windows. Since many trades overlapped, isolating which specific signals were responsible for performance became harder. We would aim to refine this in future iterations by filtering signals more carefully and using non-overlapping trade setups.

Looking ahead, there is much potential to improve the approach further. One of the most promising directions is incorporating weather forecasts, not just historical observations, to make the strategy even more forward-looking. We also see opportunities to adapt the model to different market regimes, such as high-volatility periods or strong directional trends, and tailor strategies more closely to the characteristics of Arabica and Robusta individually.

Overall, this project shows that alternative data, such as weather information, can bring real value to systematic trading strategies. By combining data science with knowledge of agriculture and commodity markets, we built a model that links climate signals to financial decisions. The strategy could become even more effective with more accurate inputs and improved structure.

Project Foundation

Overview of Coffee Market

Coffee is one of the most widely consumed beverages worldwide and one of the most actively traded agricultural commodities. As of 2023, the global coffee market was valued at approximately USD 223.78 billion and is expected to grow at a compound annual growth rate of 5.4 percent until 2030 (Grand View Research). This expected growth is driven by several factors, including increasing demand in developing countries, the rise of specialty coffee, and evolving consumer preferences in both mature and emerging markets.

Two main types of coffee beans dominate international trade: Arabica and Robusta. Arabica accounts for about 60 percent of global production and is known for its smooth, mild flavor, making it popular in premium blends. Robusta makes up around 38 percent of the market and is more resistant to pests and climatic stress (Bilen et al., 2022). It is commonly used in instant coffee and espresso blends, contributing to its growing demand, particularly in Asia and parts of Africa (International Coffee Organization).

The global coffee supply chain is concentrated in a few countries, with Brazil, Vietnam, Colombia, Indonesia, and Ethiopia accounting for over 70 percent of total exports (Shahbandeh, 2025). These countries are often affected by extreme weather conditions such as droughts, floods, and frosts, making coffee production highly sensitive to environmental change (Food and Agriculture Organization of the United Nations, 2025). Because of this, coffee prices tend to react quickly to climate-related events, creating opportunities for systematic and data-driven trading strategies.

Figure 1: Global Leading Countries in Coffee Market

Coordinates Used

To ensure the weather data reflected real growing conditions, we selected specific coordinates from coffee-producing regions that are both relevant to the market and climatically vulnerable. For Arabica, we focused on areas known for their exposure to cold snaps and variable rainfall. For Robusta, we selected regions with high production volumes that also experience unpredictable monsoon patterns. These location-specific coordinates allowed the model to link climate variability directly to supply risk.

Figure 2: Coordinates Used for Arabica and Robusta

Coffee Futures: How the Market Works

Coffee futures are standardized contracts traded on regulated exchanges. Arabica contracts are traded on ICE US in lots of 37,500 pounds, while Robusta contracts are traded on ICE Europe in 10 metric ton units. These futures allow producers, exporters, and traders to hedge against price fluctuations without physically delivering coffee.

One of the defining features of coffee futures is that they are highly reactive to anticipated weather events. For example, an early forecast of frost in Brazil can trigger significant price changes even before harvest reports are released. This behavior highlights the importance of incorporating forward-looking variables such as climate data into trading models (Zwart, 2021).

When a coffee futures trade is initiated, the contract's value is marked to the market daily. This means any change in market price is settled through a margin account, even before the contract is closed. In practice, most market participants settle their trades before expiration to avoid physical delivery. For this reason, understanding how futures respond to new information, especially weather data, was essential for this project.

Temperature and Rainfall: Key Indicators of Coffee Growth

Temperature and rainfall are the two most important climate variables in determining coffee yield and quality. Arabica grows best at cooler, high-altitude locations with temperatures between 18 and 24 degrees Celsius, while Robusta thrives in lowland areas where the ideal temperature is between 24 and 30 degrees Celsius (Kath et al., 2020). Even minor deviations from these ranges can disrupt flowering and reduce crop viability.

Rainfall is equally critical as it controls the timing of the flowering cycle. Optimal coffee production requires between 1,500 and 2,000 millimeters of rainfall annually, distributed evenly throughout the year. A lack of rain during key growth periods can delay or reduce flowering, while excessive rainfall increases the risk of disease and harvest delays. Events such as El Niño and La Niña further impact rainfall distribution and often affect coffee-producing regions differently (Peter S. Sephton, 2019).

Other weather variables like humidity and sunlight were considered in the early stages of the project. However, these were excluded due to either data availability issues or strong correlation with existing variables such as rainfall. By focusing on temperature and rainfall, we ensured that the model used biologically meaningful features and was widely available across geographies.

Coffee Futures Prices over TIme

The history of coffee futures prices shows a pattern of sharp movements triggered by unexpected events. During 2010 and 2011, frost and drought in Sul de Minas caused widespread crop damage and increased prices (Stauffer & Rochas, 2014). A similar pattern occurred in Brazil's 2013–2014 drought, which caused a nearly 90 percent increase in futures prices (Reuters, 2014).

More recently, the COVID-19 pandemic in 2020 and 2021 disrupted logistics and labor in producing countries while global demand recovered. This resulted in tighter supply and a price surge (Teixeira et al., 2020). In April 2025, new U.S. tariffs on coffee imports from major exporters like Vietnam and the EU created another wave of volatility, as market participants anticipated increased costs and restricted access (Teixeira, 2025).

These examples confirm that coffee prices are highly reactive to climate-related disruptions and global events. They also highlight the need for trading models to anticipate supply shocks rather than simply reacting after the fact.

Figure 3: Coffee Prices Over Time

Business Problem and Opportunity

Most traditional trading models for coffee rely on technical indicators and historical price patterns. However, these methods do not consider the real-time impact of weather on supply. This presents a clear opportunity to improve prediction by integrating climate variables into trading decisions.

We aimed to develop a more responsive strategy by using weather data as a leading indicator of supply risk. This approach supports better-informed decision-making and aligns with the broader trend in quantitative trading that seeks to incorporate alternative data sources. In the case of coffee, where weather events often drive sudden price changes, climate data can provide a valuable signal for generating timely and targeted trades.

Data Pipeline

Data Sources

To develop a meaningful model that could link weather patterns to coffee price movements, we began by sourcing data from two main platforms: Investing.com and OpenMeteo. Both were selected not only for their accessibility but also for the consistency and granularity of the data they offered.

Investing.com provided us with historical price data for Arabica and Robusta coffee futures. For Arabica, the data extended from 2000 to 2025, while for Robusta, the available records began in 2008. These prices were recorded daily, which gave us a rich and detailed view of price movements over time.

Alongside this, we obtained meteorological data from OpenMeteo. This platform allowed us to collect historical weather data using geographic coordinates. Since our trading strategy was built on the assumption that climate affects crop yields and, in turn, market prices, it was essential to use data that came directly from coffee-producing regions. We specifically focused on rainfall and temperature, which are widely recognized as the two most influential climate variables for coffee production. Although OpenMeteo also provided data on other variables, such as wind or humidity, we deliberately excluded them due to poor global coverage or because they did not provide any additional predictive value.

Granularity and Frequency

Once the data was collected, we aligned all variables weekly. This decision was not arbitrary but grounded in how climate events and agricultural responses typically unfold.

Firstly, using weekly data helped reduce the level of noise present in daily data. For weather and financial variables, daily changes can often be erratic and driven by short-term anomalies not reflective of broader trends. For example, one unusually hot day does not cause a failed coffee harvest. However, a trend of rising temperatures over several weeks might lead to drought stress or disrupted flowering. We could focus the model's attention on more stable and meaningful signals by averaging or summing data weekly.

Secondly, weekly intervals reflect how decisions are made in practice. Farmers, exporters, and traders rarely respond to isolated daily weather events. Instead, they are more likely to act based on sustained patterns, such as a multi-week dry spell or ongoing temperature anomalies. This is especially true in agricultural markets where logistical and operational responses take time. Aligning our data with these decision-making rhythms helped increase the model's real-world relevance.

Lastly, a weekly frequency allowed us to reduce the number of total observations without sacrificing information. This helped the model avoid overfitting and improved its ability to generalize across different periods.

Feature Engineering

After collecting the raw data, we moved on to feature engineering. This was a crucial step that involved preparing the data for modeling by cleaning it, transforming it, and creating new features that captured important relationships.

1. Standardizing Dates

Since the data came from different sources, date formats were inconsistent. Using Python's pandas.to_datetime() function, we converted all timestamps into a consistent calendar week format. This ensured that financial and climate data could be correctly aligned across the same time intervals.

2. Handling Missing Values

Gaps in weather data are common due to sensor issues or API limitations. To deal with this, we applied linear interpolation to fill small gaps. In cases where missing values exceeded a set threshold, the affected entries were dropped. This approach helped maintain the quality of the dataset while avoiding distortions.

3. Accounting for Seasonality

Coffee production is seasonal, but using numeric values for months (from 1 to 12) treats December and January as far apart. To solve this, we used sine and cosine transformations to represent the cyclical nature of time. This preserved the relationship between adjacent months and allowed the model to capture seasonal effects more accurately.

Figure 4: Cyclical Nature of Seasonality

<u>Rolling Statistics</u>

We computed rolling statistics over 4, 8, and 12-week windows to capture sustained weather conditions. These included rolling averages, standard deviations, minimums, and maximums for temperature and rainfall.

The purpose of rolling windows was to reflect that extreme events like droughts or heavy rainfall usually build up over time. A hot day might not be significant, but several weeks of elevated temperatures could damage crops. Likewise, a consistent drop in rainfall over multiple weeks can serve as an early warning for reduced harvests.

Rolling statistics helped the model see through short-term noise and identify more gradual patterns that may predict price movements. They also allowed us to capture changes in volatility, which can be a sign of unstable weather conditions and, by extension, unstable supply.

Figure 5: Rolling Statistics

<u>Lag Features</u>

There is often a time lag between cause and effect in agricultural markets. Weather events may take several weeks to impact crop yields, and any supply disruption will only be reflected in market prices after that. To account for this, we created lagged versions of key features by shifting them backward by 1, 2, 4, 8, and 12 weeks.

These lagged features allowed the model to learn how past conditions influence future prices. For example, a dry spell four weeks ago might now impact flowering, which could soon affect expected yields. Including lags allowed the model to pick up on delayed relationships between weather and returns.

Interaction Features

Finally, we constructed interaction features to capture the effects of combined stressors. Some environmental threats only become critical when multiple conditions occur simultaneously. For instance, high temperatures are much more damaging when they coincide with low rainfall. To reflect this, we created variables that measured the interaction between temperature and rainfall at each location.

We also added binary indicators to flag extreme weather. The drought indicator was set to one if there was both unusually high temperature and low rainfall. The flood indicator was triggered by excessive rainfall regardless of temperature. These simple but meaningful flags helped the model identify periods of potential disruption that may not be apparent from individual variables alone.

Interaction Feature	Description
Temp_rain_interact_loc{i}	This measures the interaction between temperature and rainfall at a given location over a specific week, capturing the compounding effects of heat and moisture. For example, high temperatures during a dry week can significantly intensify crop drought stress.
Drought_indicator_loc{i}	This binary variable flags weeks with unusually high temperatures and low rainfall relative to historical norms. This indicator is intended to detect periods of potential drought stress, where crops are likely to experience strain due to insufficient water and elevated heat. Since drought conditions often evolve gradually, this flag acts as an alert for emerging threats to yield and thus helps the model anticipate supply disruptions before they fully materialize in market prices.
Flood_indicator_loc{i}	This binary variable flags weeks with abnormally high rainfall, regardless of temperature levels. This feature mainly captures flood-like conditions, which may damage crops, delay harvests, or disrupt transportation infrastructures in coffee-producing regions. By flagging these weeks, the model becomes more responsive to the short-term risk that heavy precipitation poses to production and logistics.

By incorporating these interaction features, we allowed the model to detect more realistic scenarios where the combination of factors reflects real agricultural stress rather than relying on isolated variables.

Figure 6: Interaction Features

Strategy and Results

Engineered Data Frame

Once all features were generated, we compiled them into a single data frame that served as the foundation of our predictive model. This step organized the data so that the model could capture the relationship between weather conditions and future price returns.

	Feature-engineered columns	Explanation
Lag Features	return_lag_t	Return from *t* weeks ago to capture recent momentum or reversal patterns.
	price_lag_t	Raw price from *t* weeks ago to identify price anchoring effects
Moving Averages and Volatility	price_ma_t	Average price over the past *t* weeks to smooth short-term fluctuations
	price_std_t	Price volatility over *t* weeks to quantify market risk
	return_ma_t	Average return over *t* weeks to track sustained performance trends
Moving Averages and Volatility	rsi_t	Relative Strength Index over *t* periods to measure overbought/oversold conditions
	ema_t	Exponential moving average over *t* weeks to give more weight to recent prices
	macd	Difference between short- and long-term EMAs to identify trend direction
	macd_signal	Signal line (EMA of MACD) used to generate trading signals
Weather Lag & MA Features	temp_locX_lag_t	Weekly average temperature at location *X*, lagged by *t* weeks.
	rain_locX_lag_t	Weekly rainfall at location *X*, lagged by *t* weeks
	temp_loc_ma_t	Temperature moving average at location *X* over the past *t* weeks
Extreme Weather Indicators	extreme_high_temp_loc	Binary flag indicating extreme high temperature events at a location
	extreme_low_temp_loc	Binary flag indicating extreme low temperature events at a location
	extreme_rain_loc	Binary flag for excessive rainfall events at a given location

By transforming raw weather and price data into interpretable features, we created a dataset that captured past market behavior and represented real environmental conditions that influence supply and price movements.

<u>Three Key Challenges That Shaped Our Strategy Design</u>

Throughout the strategy's development, we encountered several challenges that significantly influenced our approach. Addressing these issues was necessary not only to improve model performance but also to ensure that our results were valid and could be realistically applied in a trading context.

1. Temporal Leakage in Validation

Temporal leakage occurs when the model is given access to future data during training, even if only unintentionally. In time series modeling, this creates unrealistic results, as the model may learn patterns that would not be available in real-world trading.

To prevent this, we avoided using standard cross-validation techniques that randomly split the data. Instead, we applied time series cross-validation, which always trains the model on data from the past and tests it on data from the future. This ensured that our model was not relying on information it could not have known at the time of prediction. More importantly, it forced us to treat the problem the same way a real trader would: making decisions without the benefit of hindsight.

2. Misalignment of Features and Targets

Another issue was the potential mismatch between weather features and the returns they were intended to predict. Because the effects of rainfall or temperature are not immediate, there is often a time lag between when a weather event occurs and when it impacts market prices.

To solve this, we created a time-aware merge between the weather and return data. This meant that, at any given point in the dataset, only past or current weather conditions were used to predict future returns. This adjustment was critical in ensuring the model's predictions were causally meaningful rather than statistically convenient.

3. Inconsistent Prediction Horizon

The final challenge was related to the prediction horizon. If the model is asked to predict "future returns," it must be clear what time frame that refers to. If the target shifts from one week to the next, the model becomes confused and loses the ability to generate consistent signals.

We resolved this by fixing the prediction target to a specific horizon. In our case, we defined the target as the 8-week forward return. This meant the model always tried to answer the same question: "Given the weather patterns up until today, what is the expected return of coffee prices eight weeks from now?" By keeping the objective fixed, we improved the reliability of the model's predictions and avoided conflicting signals.

<u>Implementation of Time Series Cross Validation (TSCV)</u>

As mentioned earlier, TSCV was used to assess model performance. Traditional validation methods, such as random sampling, are unsuitable for time-dependent data, as they allow information from the future to be included in the training set. This creates overly optimistic results that do not reflect how the model would perform.

TSCV avoids this problem by preserving the data in chronological order. The model is trained on earlier periods and validated on later ones, mimicking how a live strategy operates. Although this approach requires more careful implementation and longer training times, it results in a much more realistic estimate of model accuracy.

<u>Choice of Model: XGBoost</u>

We selected XGBoost as the main model for our strategy. XGBoost is a gradient-boosting algorithm that builds a sequence ensemble of decision trees. Each tree is trained to correct the errors made by the previous ones, which allows the model to learn complex, non-linear relationships (NVIDIA, 2025).

The weather-price relationship was not likely to be simple or linear as a moderate temperature increase might improve yields in one region but hurt them in another, depending on historical averages, rainfall, or soil conditions. XGBoost's flexibility made it better suited for handling these nuanced relationships.

We also tuned several hyperparameters to control the model's complexity and avoid overfitting. For example, we limited tree depth and adjusted the learning rate to stabilize training. While more complex than a basic linear model, XGBoost allowed us to capture the interactions and delays often present in real-world agricultural systems.

<u>XGBoost Variables</u>

To fine-tune the model, we tested various configurations and adjusted the following key hyperparameters:

Hyperparameter	**Value**	**Description**
n_estimators	[100,200]	Number of boosting rounds (trees). Allows the model to learn complex relationships without becoming too deep.

max_depth	[3,5]	Controls the complexity of each tree. A smaller depth prevents overfitting to short-term or noisy fluctuations.
learning_rate	[0.01, 0.1]	Reduces the size of each update. A lower rate stabilizes training and avoids volatile learning behavior.
subsample	[0.8, 1.0]	Uses at least 80% of the data per tree. Introduces randomness to help reduce variance and improve generalization.
colsample_bytree	[0.8]	Uses 80% of features per tree. Prevents the model from over-relying on any single variable.

The result was a model that could identify valuable patterns in the data without becoming too dependent on any single feature or noise in the training set.

Time Aware Merge

A crucial part of the strategy was ensuring that the data fed into the model reflected the flow of real-world information. When merging price and weather data, we ensured that each price observation was only matched with weather features from the same week or earlier. We also built a seven-day tolerance to handle slight timing differences, such as rainfall being recorded on a Sunday and prices being captured on a Monday.

This step ensured that the data respected the order in which information became available. Without it, the model might use future weather data that would not have been known at the time, undermining the purpose of a predictive model.

Figure 7: Time Aware Merge

Setting Future Returns Targets

The model's target was defined as the 8-week forward return. This meant we shifted the price data eight weeks into the future and calculated the percentage change from the current week. For example, if prices rose by 5 percent over the next eight weeks, that became the value the model tried to predict based on today's weather conditions.

This structure ensured that the model was predictive rather than reactive. It also aligned with the practical needs of a trading strategy, which relies on anticipating market movements rather than chasing them after the fact.

During testing, we also created multiple versions of the target (1-week, 2-week, and 4-week returns) to explore how the model responded to different time horizons. However, we found that the 8-week target offered the most consistent patterns and clearest performance, and we adopted it as the final version.

Standardisation and Correlations

Before training the model, we standardized the input features. This step placed all variables on a similar scale, making it easier for the model to compare them and avoid bias toward features with larger numeric values. For example, temperature and rainfall are measured in very different units, but both carry valuable information. Standardization allowed the model to treat them fairly.

We also examined the correlation between input features and future returns. While many individual features had weak correlations, we observed that certain rainfall sums and lagged temperature features had moderate relationships with forward returns. These correlations were not strong enough to serve as trading signals, but they did validate the idea that weather affects prices measurably.

Figure 8: Correlation Matrix

Model Features and Importance

To understand which variables influenced the model's predictions, we looked at feature importance scores generated by XGBoost. These scores showed how frequently each variable was used in tree splits across the model.

We found that features with longer look-back windows and interaction effects tended to be more influential. For example, rainfall and temperature from four to eight weeks ago appeared more frequently in the model than short-term weather changes. This supported our earlier belief that climate-related effects on coffee supply unfold gradually rather than instantly.

This insight also helped guide the design of our final strategy. We aimed to create a more stable and reliable trading approach by focusing on longer-term patterns rather than immediate fluctuations.

Figure 9: Top Weather Features Correlation with Future Returns

XGBoost vs. Linear Regression

To evaluate the effectiveness of our chosen model, we compared XGBoost with a traditional linear regression model. The purpose of this comparison was to measure accuracy and understand whether XGBoost's added complexity was justified in this context.

Linear regression is often used as a baseline model because of its simplicity and interpretability. It assumes a linear relationship between input features and the target variable. This helps to understand whether there are any apparent trends in the data (Beers, 2025). However, this simplicity can also be a limitation. Regarding coffee prices, the relationship between weather patterns and future returns is unlikely to be strictly linear. A slight temperature change might be beneficial under certain conditions but harmful under others. Linear regression is not equipped to capture such interactions.

The results supported this intuition when we ran both models on the same dataset. The XGBoost model consistently outperformed linear regression across multiple evaluation metrics, including mean absolute error and R-squared. While the linear model was more straightforward to interpret, it failed to account for the complexity of the underlying data. In contrast, XGBoost was able to pick up on the non-linear and delayed effects of weather on price returns, as well as the interactions between variables such as rainfall and temperature.

This comparison demonstrated that while linear models can help establish a baseline, they may oversimplify the dynamics of commodity markets. Given climate's unpredictable and often layered effects on agriculture, a more flexible model like XGBoost is better suited for capturing these nuances.

Strategies: Buy-and-Hold vs. Weather-Based Strategy

To test the practical value of our model, we implemented two distinct trading strategies: a passive Buy-and-Hold approach and an active Weather-Based strategy driven by our return predictions.

The Buy-and-Hold strategy served as a benchmark. It involved purchasing Arabica and Robusta coffee futures and holding the position without any adjustments. This approach reflects what a long-term investor might do without access to predictive insights. The simplicity of this strategy makes it helpful in evaluating whether a more complex model adds value.

Meanwhile, the Weather-Based strategy made decisions based on the model's predicted 8-week forward returns. Specifically, we only entered a long position when the predicted return was within the top 25 percent of historical predictions. The idea was to take advantage of the model's strongest signals while avoiding periods of uncertainty. This rule-based filtering allowed us to reduce exposure during low-confidence weeks and concentrate on high-probability opportunities.

At first glance, the Weather-Based strategy achieved slightly higher returns and a better Sharpe Ratio than the Buy-and-Hold. It also experienced marginally lower volatility. These results suggested that integrating climate data into trading decisions could improve performance and risk management.

Figure 10: Annualised Performance between Buy & Hold and Weather-Based Strategy

However, further inspection revealed important limitations. Since the model was evaluated weekly but predicted returns over an 8-week horizon, the same high return forecast could influence trading decisions for multiple weeks. This created overlapping trades that closely resembled the Buy-and-Hold strategy. As a result, both strategies often shared the same exposure, making it difficult to differentiate between them in practice.

Figure 11: Performance Comparison: Weather-based vs. Buy-and-Hold

Moreover, the correlation between predicted and actual returns was relatively weak. While the model was sometimes directionally correct, its signals lacked the precision to generate timed entries and exits. This reduced the strategic advantage we had hoped to gain from using weather data as a trading signal.

While the Weather-Based strategy showed early promise, it also revealed the need for more careful trade execution and better-aligned timing rules.

Train and Test Split

To assess the model's real-world performance, we applied a train-test split that preserved the temporal order of the data. This means the model was trained only on past data and tested on future data, mimicking how a trading system would operate in practice.

We reserved the final year of available data as the test set and used all preceding data for training. This approach ensured that the model's predictions were based only on information that would have been available then. Any attempt to mix training and testing periods would have led to artificial results and overestimated performance.

Before finalising the split, we removed features and targets with more than 25 percent missing values to ensure reliability. We also replaced infinite values with NaNs and used median imputation based on the training set to fill in any remaining gaps. This was done carefully to avoid leaking information from the test set into the training process.

Separating and cleaning the data was not just a technical requirement. It was a necessary step to make sure the model was evaluated fairly. In financial forecasting, even a small amount of lookahead bias can completely distort the results. By treating the test set as truly unseen data, we were able to get a clearer picture of how the model would perform in a real trading environment.

Figure 12: Train and Test Split

Adapted Strategies

The initial version of the Weather-Based strategy involved simply making weekly predictions of 8-week returns and entering a position whenever the forecast was positive. However, this led to several problems.

Since the model made a new prediction each week, and each prediction referred to an 8-week time frame, many trades overlapped. This caused the strategy to remain in a long position for extended periods, often without meaningful changes in exposure. As a result, the strategy behaved very similarly to Buy-and-Hold. While this confirmed that the model was generally optimistic, it also indicated that the trading signals were not strong or distinct enough to justify frequent decisions.

Another issue was that the model's signals were not consistently reliable. Many predictions hovered near zero, making it unclear whether a position should be entered or avoided. The lack of sharp contrast between high-confidence and low-confidence weeks made the strategy vulnerable to false positives.

To address these weaknesses, we adapted the strategy to trade only when the model's prediction fell into the top quartile of historical returns. This added threshold helped filter out weak or ambiguous signals. We also reduced the frequency of trading decisions, allowing trades to run their full 8-week course before evaluating a new entry. This helped avoid overlapping positions and made each trade more independent.

These adjustments improved the strategy's structure and logic. Rather than trying to act on every forecast, we focused only on the most confident signals. This change reflected a more realistic approach to trading, where not every model output should be acted upon immediately.

Performance and Backtest

XGBoost Strategy Performance Evaluation

After training and validating the model, we conducted a performance evaluation to determine how well the strategy would have worked if it had been deployed in a real trading environment. This step was not simply about measuring accuracy. Instead, it was about testing whether the model could generate returns that were not only positive but also stable and repeatable under realistic market conditions.

The performance was assessed using the final test set, representing the most recent data year. The model was applied to forecast 8-week forward returns, and trades were executed based on whether those predictions fell within the top 25 percent of all historical forecasts.

Over this evaluation period, the weather-based strategy generated a slightly higher annualized return than the passive Buy-and-Hold benchmark. It also demonstrated a higher Sharpe Ratio, which suggests a better balance between return and risk. Although the differences in raw performance were modest, these results showed that the model could identify periods of relatively favorable returns with some level of consistency.

However, it is important to note that these results should be interpreted cautiously. Many of the long positions the model issued were overlapping, meaning that multiple forecasts pointed in the same direction. This reduced the variation in exposure and made the strategy's performance appear smoother than it might have been in a live trading setting. The model showed potential in this sense, but its signals were not yet sharp or distinct enough to consistently outperform a passive approach.

In evaluating performance, we also considered downside risk. The model avoided certain drawdowns by staying out of the market during lower-confidence weeks. This was a promising sign, as it suggested that climate data could be useful not only for identifying opportunities but also for managing risk. Nevertheless, the overall advantage was still marginal, and further improvements were clearly needed.

Figure 13: XGBoost Strategy Performance Evaluation

Figure 17: Backtest Performance (Linear Scale): XGBoost (8w) vs. Buy-and-Hold

Arabica vs. Robusta Performance Divergence

One unexpected insight from our analysis was the difference in model performance between Arabica and Robusta futures. While our strategy was initially developed with Arabica as the primary focus, we also applied the same model to Robusta data to assess its adaptability.

However, we found that the strategy performed noticeably better with Arabica futures. The signals were stronger, the feature correlations were more stable, and the return forecasts were slightly more accurate. This divergence likely reflects structural differences between the two markets.

Arabica is generally more sensitive to weather disruptions, frigid temperatures, and drought conditions. It is often grown at higher altitudes in regions like Brazil and Colombia, where climate anomalies directly and immediately impact crop quality. Conversely, Robusta is grown in more resilient environments and tends to be less affected by short-term weather fluctuations. It is also traded in different volumes and exchanges, which may influence how prices respond to new information.

As a result, the climate-based strategy captured more meaningful signals in the Arabica market. While this does not mean that weather data is irrelevant for Robusta, it suggests that different variables or model configurations may be required to achieve similar performance levels.

This observation reinforced the importance of understanding the physical realities behind the traded asset. A strategy that works in one market cannot be assumed to work equally well in another, even if the two are closely related. Successful modeling in commodity markets requires technical accuracy and a firm grasp of the underlying fundamentals.

Figure 18: Good Arabica Results and Poor Robusta Results

XGBoost vs. Trend Following vs. Linear Regression

To fully understand our model's relative performance, we compared XGBoost against two additional strategies: a simple trend-following model and linear regression. This comparison was intended not just to measure which model produced the highest returns but also to reflect on each model's assumptions and limitations.

The trend-following strategy operated with a basic rule: if the return over the past 8 weeks was positive, we would go long the following week. If the return was negative, we would stay out of the market. This common rule-based approach is used in financial markets because of its simplicity and historical success in specific commodity cycles. It does not rely on external data like weather but instead assumes that price momentum carries forward.

As mentioned earlier, linear regression served as a baseline model. It used the same weather features as XGBoost but assumed a linear relationship between these variables and future returns.

Among the three, XGBoost consistently produced the most favorable balance between return and risk, especially regarding the Sharpe Ratio. The trend-following model generated competitive returns during strong uptrends but performed poorly during flat or volatile periods. Linear regression, although easy to implement and interpret, showed the weakest predictive performance. It often fails to detect inflection points or adapt to complex interactions in the data.

While XGBoost is not flawless, it provides a better framework for handling weather's non-linear and delayed effects on commodity prices. While the trend-following model responded well during persistent price movements, it could not anticipate shifts caused by upcoming supply disruptions. XGBoost, by integrating weather data, offered a more forward-looking perspective that proved valuable in certain market conditions.

Figure 16: Strategy Comparison

Directional Accuracy Analysis

Aside from return metrics, we also evaluated the model's directional accuracy, focusing on how often the model correctly predicted whether returns would be positive or negative over the 8-week horizon. This metric is critical in trading contexts because it reflects the model's ability to anticipate market direction, which can guide risk positioning even when return magnitudes are uncertain.

1. Up-Trend Markets: The model demonstrated an exceptional 83 percent win rate during upward market trends, consistently maintaining long positions during favorable conditions.
2. Down-Trend Markets: During bearish phases, the strategy achieved a 58 percent win rate, successfully shorting the market approximately half the time during downturns.
3. Overall Directional Accuracy: Across all market conditions, the model correctly predicted price direction 67 percent of the time, providing a substantial edge over random chance.

This insight highlighted that the model may not always be accurate. What matters more is knowing when to trust it. By trading only during high-confidence periods, we could filter out weaker signals and avoid unnecessary exposure. This kind of conditional accuracy plays a vital role in real-world strategy design, where not every prediction needs to be acted upon.

Figure 19: Distribution of Realized Trade Returns by Market Regime

Performance Limitations

Although the model demonstrated early signs of promise, several limitations prevented it from reaching its full potential.

1. Market Regime Dependence

The strategy performed best under specific market conditions, particularly when the market structure aligned well with the model’s signal-generation capabilities. This raises concerns about its resilience during sudden regime shifts or atypical market behavior, suggesting that the model may lack the flexibility to adapt effectively across all market environments.

2. Robusta Model Challenges

While the strategy performed strongly in the Arabica market, its application to Robusta futures was less successful. The underwhelming results in this segment point to potential gaps in the model’s feature set or calibration, indicating a need for variety-specific adjustments. Tailoring the model more closely to the characteristics of the Robusta market, whether through feature engineering, additional data inputs, or model tuning, may improve performance.

3. Temporal Consistency

Although the strategy performed well overall, its effectiveness varied over time. There were distinct periods of underperformance that reduced cumulative returns and introduced volatility into the backtest results. This inconsistency suggests the need for further investigation into the model's stability and performance sustainability, particularly under different temporal conditions or external factors influencing market behavior.

Our weather-based trading strategy demonstrated statistically significant and economically meaningful outperformance for Arabica coffee futures. The model's strong directional accuracy and ability to navigate different market conditions validate our hypothesis that weather data can be effectively leveraged to enhance coffee futures trading strategies. Further refinements are needed to address identified limitations and extend the approach to Robusta markets.

Future Improvements

Based on the performance outcomes and the limitations we identified, there are three main areas where our weather-based coffee trading strategy could be meaningfully improved. These refinements aim to improve accuracy and ensure the strategy remains relevant, resilient, and responsive across changing market environments.

Understanding When Weather Matters More and When It Doesn’t

Currently, the model assumes that weather has a relatively stable impact on coffee prices. However, in reality, the influence of weather is likely to fluctuate depending on broader market conditions, seasonal timing, or even geopolitical context. There may be periods when coffee prices respond sharply to climate events and other times when the same events are already priced in or are overshadowed by other drivers, such as policy changes or shipping disruptions.

Recognizing when the weather has a stronger or weaker effect would allow the model to adapt more effectively. For instance, when weather appears to have minimal predictive power, the strategy could reduce exposure, delay entry, or avoid trades altogether. In doing so, capital would be preserved for moments when the relationship between weather and prices becomes clearer and stronger. Instead of assuming that weather always matters equally, the strategy should learn to distinguish high-impact periods from low-impact ones and adjust accordingly.

Adapting the Strategy to Changing Market Conditions

Another limitation we observed was that the model was primarily tested in a range-bound and somewhat volatile Arabica market. While this environment was useful for testing how the strategy performs under pressure, it may not represent the full spectrum of market conditions. Many promising signals in sideways markets became less reliable when prices began trending strongly in one direction.

This raises the question of how the strategy would behave in different market regimes, such as sustained bull runs, deep corrections, or high-volatility events linked to macroeconomic shifts. In our case, the model struggled particularly during prolonged upward trends, failing to capture momentum. This suggests it was more reactive than proactive and possibly too anchored in mean-reversion logic.

To address this, the strategy needs to become more regime-aware. This does not necessarily mean overhauling the entire model but equipping it with the ability to detect broader market structures and adapt its behavior. For example, during trending markets, the model may need to prioritize different features or use longer time horizons to remain relevant. Enhancing robustness across varied regimes will be key if the strategy is to succeed beyond its current testing window.

Moving from Historical to Predictive: Using Weather Forecasts

Currently, the model relies entirely on historical weather data. While this gives us a solid understanding of how past conditions may have influenced price, it does little to anticipate what will happen next. This creates a lag in decision-making and limits the model’s ability to respond quickly to upcoming risks.

Incorporating weather forecasts could help close this gap. Unlike historical averages, forecasts provide a forward-looking view that reflects expected temperature shifts, rainfall anomalies, or drought risk. These developments are often the triggers that move commodity prices, particularly in sensitive crops like coffee. By integrating forecasts into the model, we could increase its ability to act preemptively rather than reactively.

This improvement would make the strategy more agile, particularly when environmental changes are imminent but not yet visible in historical records. For example, early warnings about El Niño conditions or unusual cold fronts could inform better positioning before market prices react.

Addressing these three key areas could move the strategy closer to real-world viability. These enhancements would not only reduce unnecessary trades and improve timing but also help the model withstand diverse and unexpected conditions.

References

Beers, B. (2025). Regression: Definition, analysis, calculation, and example. Investopedia. https://www.investopedia.com/terms/r/regression.asp

Bilen, C., El Chami, D., Mereu, V., Trabucco, A., Marras, S., & Spano, D. (2022, December 25). A systematic review on the impacts of climate change on Coffee Agrosystems. MDPI. https://www.mdpi.com/2223-7747/12/1/102

Do, T. N. (2024). Commentary: Climate change leaves a bitter taste in world's Coffee Supply. CNA. http://www.channelnewsasia.com/commentary/climate-change-making-coffee-more-expensive-supply-quality-4642536#:~:text=COFFEE%20WILL%20GET%20MORE%20EXPENSIVE&text=The%20combined%20effect%20could%20significantly,in%20infrastructure%20like%20irrigation%20systems

Food and Agriculture Organization of the United Nations. (2025, March 17). Adverse climatic conditions drive coffee prices to highest level in years. PreventionWeb. https://www.preventionweb.net/news/adverse-climatic-conditions-drive-coffee-prices-highest-level-years

Grand View Research. (n.d.). Coffee market to reach $323.72bn by 2030: Cagr 5.4%. https://www.grandviewresearch.com/press-release/global-coffee-market

ICE. (2025). Coffee C ® Futures. Coffee C Futures. https://www.ice.com/products/15/Coffee-C-Futures

International Coffee Organization. (n.d.). December 2023 Coffee Report and Outlook. https://icocoffee.org/documents/cy2023-24/Coffee_Report_and_Outlook_December_2023_ICO.pdf

Kath, J., Byrareddy, V. M., Craparo, A., Nguyen‐Huy, T., Mushtaq, S., Cao, L., & Bossolasco, L. (2020). Not so robust: Robusta coffee production is highly sensitive to temperature. Global Change Biology, 26(6), 3677–3688. https://doi.org/10.1111/gcb.15097

NVIDIA. (2025). What is XGBoost?. NVIDIA Data Science Glossary. https://www.nvidia.com/en-us/glossary/xgboost/

Peter S. Sephton, P. (2019). El Niño, La Niña, and a cup of Joe. Science Direct. https://doi.org/10.1016/j.eneco.2019.104503

Reuters. (2014). Coffee prices at 47-year high on Brazil crop concerns | Reuters. https://www.reuters.com/markets/commodities/coffee-prices-47-year-high-brazil-crop-concerns-2024-11-29/

Shahbandeh, M. (2025). Leading countries worldwide based on coffee area harvested 2022 | statista. Statista. https://www.statista.com/statistics/279162/top-countries-worldwide-based-on-coffee-area-harvested/

Stauffer, C., & Rochas, A. (2014). Record Brazil heat pressures crops, energy prices, government | Reuters. Reuters. https://www.reuters.com/article/us-brazil-heat/record-brazil-heat-pressures-crops-energy-prices-government-idUSBREA0U1PM20140201

Teixeira, M. (2025). Import tax on coffee pressures us roasters already facing high prices | reuters. Reuters. https://www.reuters.com/markets/commodities/import-tax-coffee-pressures-us-roasters-already-facing-high-prices-2025-04-03/

Teixeira, M., Angel, M., & Cobb, J. (2020, March 30). Coffee importers stockpiling on fears over Coronavirus Lockdowns. The Irish Times. https://www.irishtimes.com/business/agribusiness-and-food/coffee-importers-stockpiling-on-fears-over-coronavirus-lockdowns-1.4215594

UNUniversity. (2014). Climate change impacts to drive up coffee prices. Our World. https://ourworld.unu.edu/en/climate-change-impacts-to-drive-up-coffee-prices

Zwart, R.-J. (2021, March 17). Coffee futures: The Basics for Coffee Roasters. Trabocca. https://www.trabocca.com/coffee-knowledge/coffee-supply-chains/coffee-futures-the-basics-for-coffee-roasters/

Appendix

Code Files

The adjunct files contain the code and data used to perform the trading strategy:

- *QF206FinalProject.ipynb* - This is the main code used for the project.
- *ArabicaWeekly2000_2024.csv* - Weekly prices of Arabica coffee futures.
- *Robusta_resampled_weekly.csv* - Weekly prices of Robusta coffee futures
- *meteoArab_5places_00_24_weekly.csv* - The temperature and precipitations of the 5 places we used for Arabica coffee.
- *weekly_df.csv* - The temperature and precipitations of the 5 places we used for Robusta coffee.

Figures

Figure 1: Global Leading Countries in Coffee Market

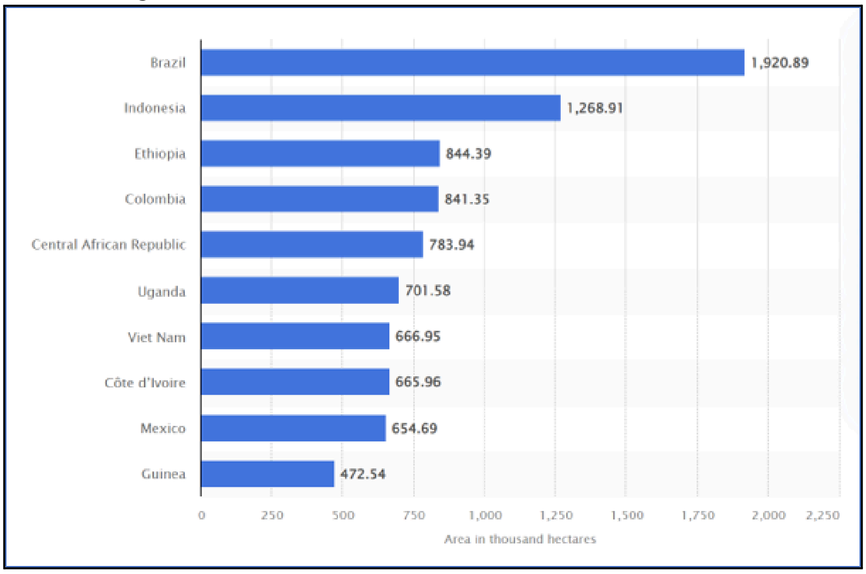


Figure 2: Coordinates Used for Arabica and Robusta

Arabica	
Brazil – Sul de Minas	Largest Arabica-producing region; highly sensitive to frost and drought, driving global price spikes.
Brazil – Cerrado Mineiro	Technologically advanced and consistent output; stable weather makes it a key benchmark region.
Colombia – Cauca	High-elevation zone with top-quality beans; strong link between climate shifts and yield variability.
Colombia – Eje Cafetero	Historic core of Colombia's coffee industry; two harvests per year enable richer data analysis.
Ethiopia – Sidamo	Birthplace of Arabica; extreme altitude and biodiversity make it highly responsive to climate changes.

Robusta	
India – Chikkamagaluru, Karnataka	Heart of India's coffee belt; strong monsoon cycles offer clear seasonal signals.
Uganda – Hoima District	Emerging Robusta hub; erratic rainfall adds Sub-Saharan context to climate modeling.
Indonesia – Lampung, Southern Sumatra	Top Robusta exporter; exposure to volcanoes and El Niño adds weather-type diversity.
Brazil – Vila Valério, Espírito Santo	Main Robusta-producing state; allows Arabica vs. Robusta climate comparison within Brazil.
Vietnam – Đắk Lắk, Central Highlands	World's top Robusta producer; highly sensitive to shifting monsoon and drought patterns.

Figure 3: Coffee Prices Over Time

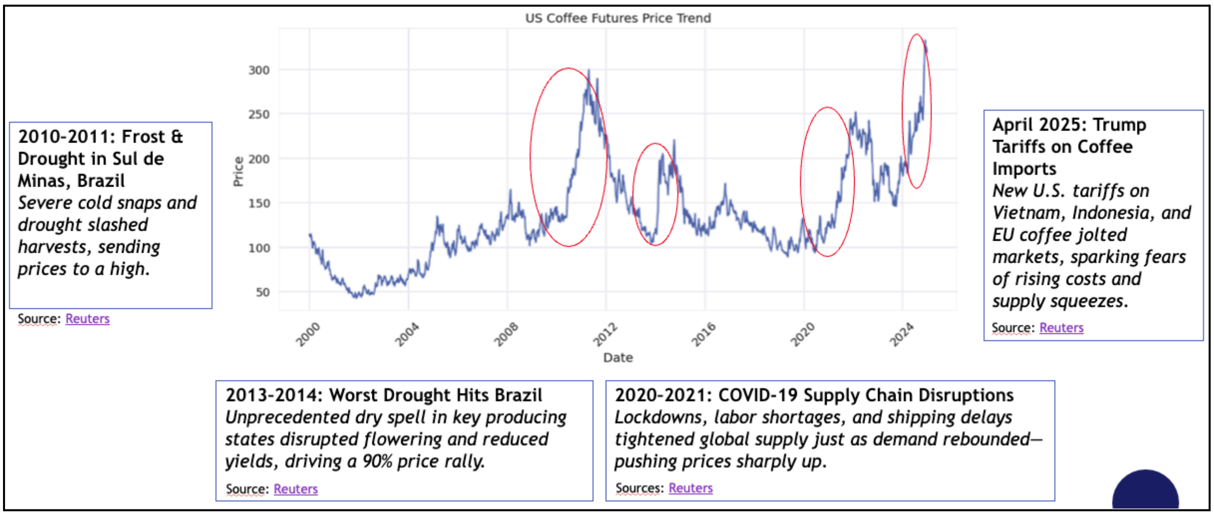


Figure 4: Cyclical Nature of Seasonality

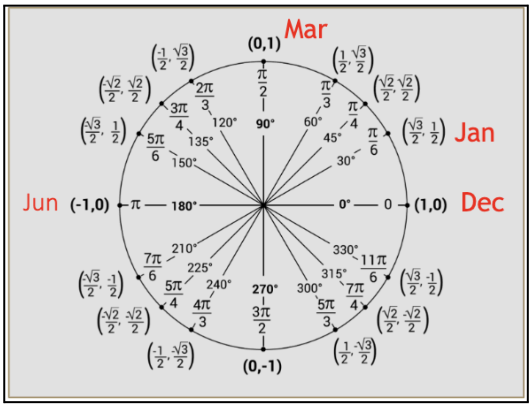


Figure 5: Rolling Statistics

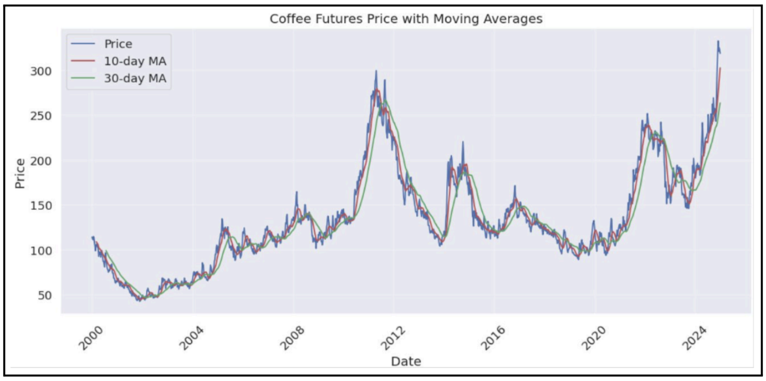


Figure 6: Interaction Features

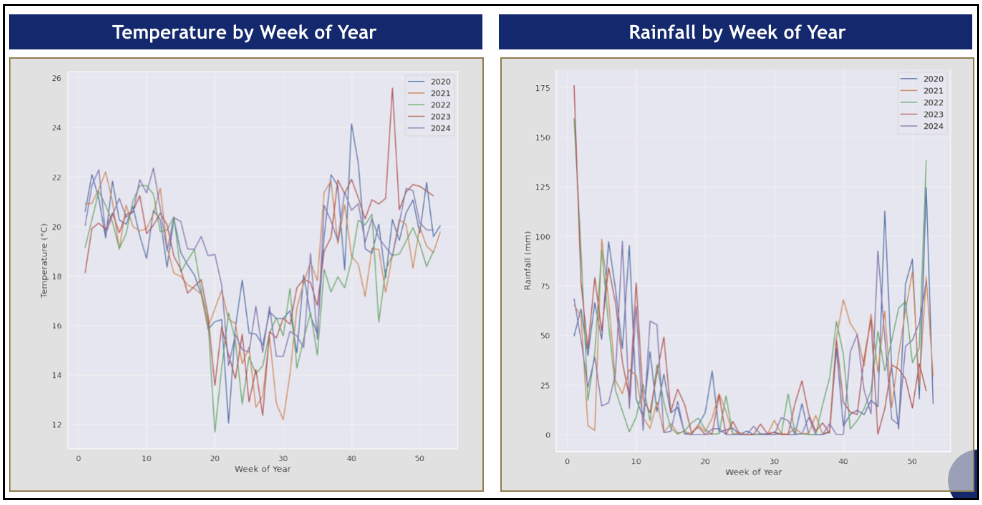
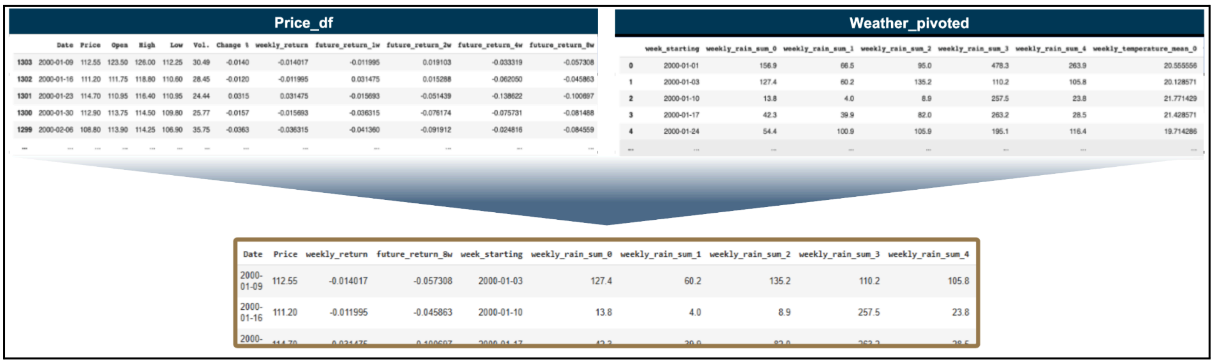


Figure 7: Time Aware Merge



Price_df

	Date	Price	Open	High	Low	Vol.	Change %	weekly_return	future_return_1w	future_return_2w	future_return_4w	future_return_8w
1303	2000-01-09	112.55	123.50	126.00	112.25	30.49	-0.0140	-0.014017	-0.011995	0.019103	-0.033319	-0.057308
1302	2000-01-16	111.20	111.75	118.80	110.60	28.45	-0.0120	-0.011995	0.031475	0.015288	-0.062050	-0.045863
1301	2000-01-23	114.70	110.95	116.40	110.95	24.44	0.0315	0.031475	-0.015693	-0.051439	-0.138622	-0.100697
1300	2000-01-30	112.90	113.75	114.50	109.80	25.77	-0.0157	-0.015693	-0.036315	-0.076174	-0.075731	-0.081488
1299	2000-02-06	108.80	113.90	114.25	106.90	35.75	-0.0363	-0.036315	-0.041360	-0.091912	-0.024816	-0.084559
...	...	...	...	...	...	...	...	...	...	...	...	...

Weather_pivoted

	week_starting	weekly_rain_sum_0	weekly_rain_sum_1	weekly_rain_sum_2	weekly_rain_sum_3	weekly_rain_sum_4	weekly_temperature_mean_0
0	2000-01-01	156.9	66.5	95.0	478.3	263.9	20.555556
1	2000-01-03	127.4	60.2	135.2	110.2	105.8	20.128571
2	2000-01-10	13.8	4.0	8.9	257.5	23.8	21.771429
3	2000-01-17	42.3	39.9	82.0	263.2	28.5	21.428571
4	2000-01-24	54.4	100.9	105.9	195.1	116.4	19.714286
...	...	...	...	...	...	...	...

Date	Price	weekly_return	future_return_8w	week_starting	weekly_rain_sum_0	weekly_rain_sum_1	weekly_rain_sum_2	weekly_rain_sum_3	weekly_rain_sum_4
2000-01-09	112.55	-0.014017	-0.057308	2000-01-03	127.4	60.2	135.2	110.2	105.8
2000-01-16	111.20	-0.011995	-0.045863	2000-01-10	13.8	4.0	8.9	257.5	23.8
2000-	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Figure 8: Correlation Matrix

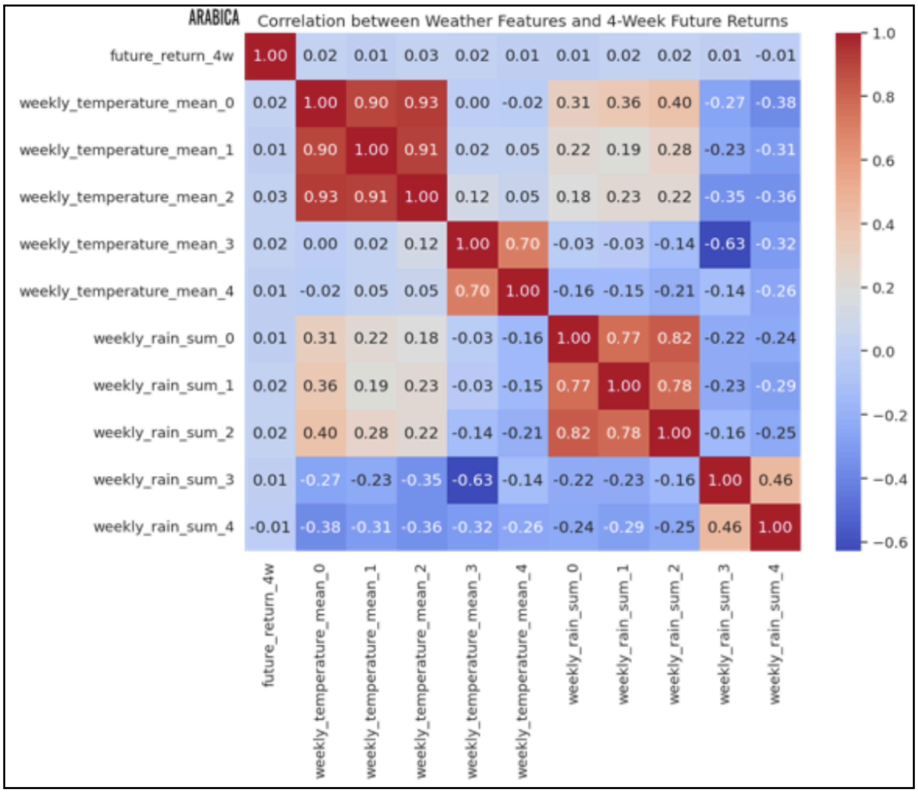


Figure 9: Top Weather Features Correlation with Future Returns

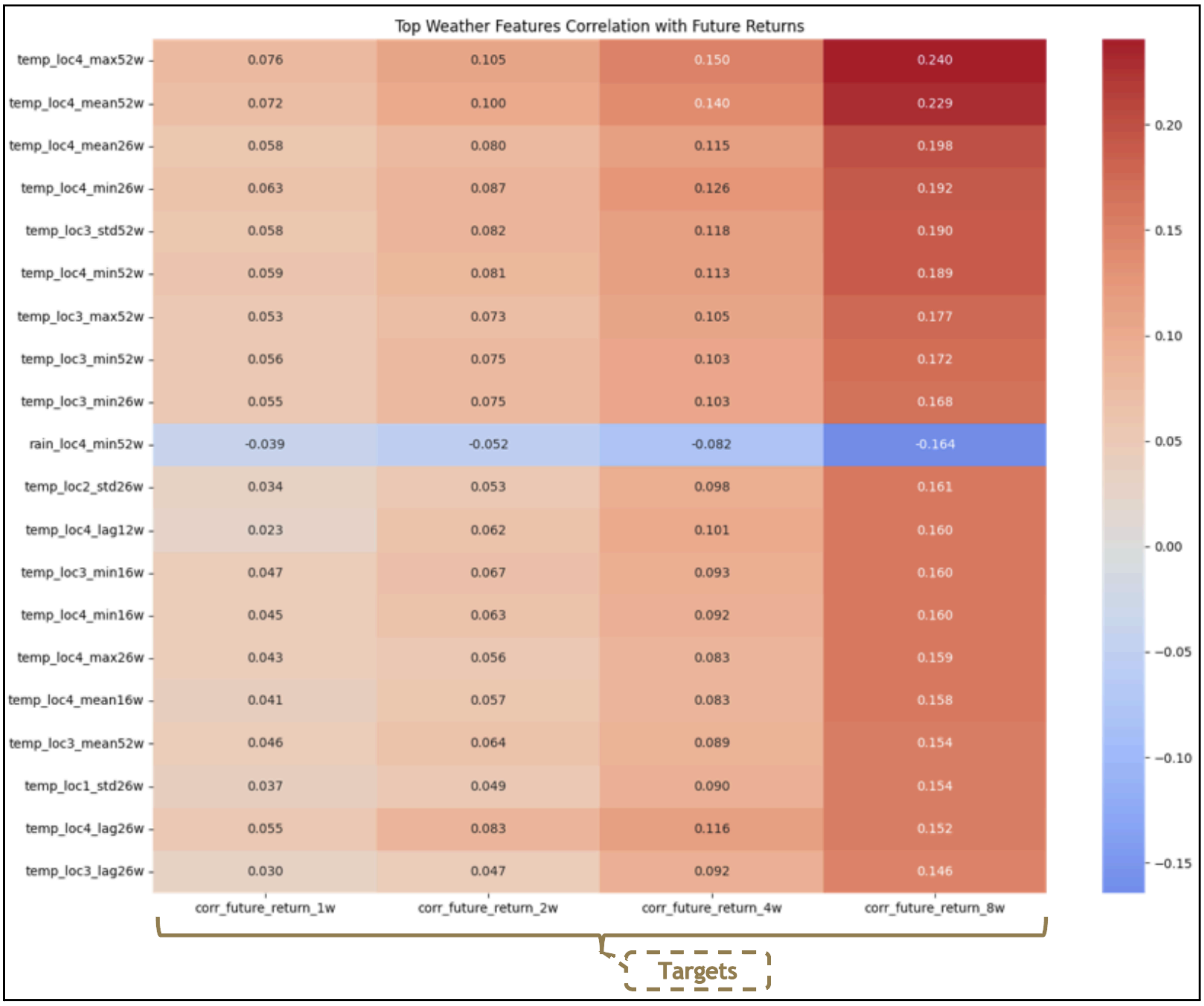


Figure 10: Annualised Performance between Buy-and-Hold and Weather-based

Annualised Performance		
	Buy and Hold	Weather Based
Sharpe Ratio	6.6353	6.9146
Return	4.1513	4.2424
Volatility	0.6256	0.6135

Figure 11: Performance Comparison: Weather-based vs. Buy-and-Hold

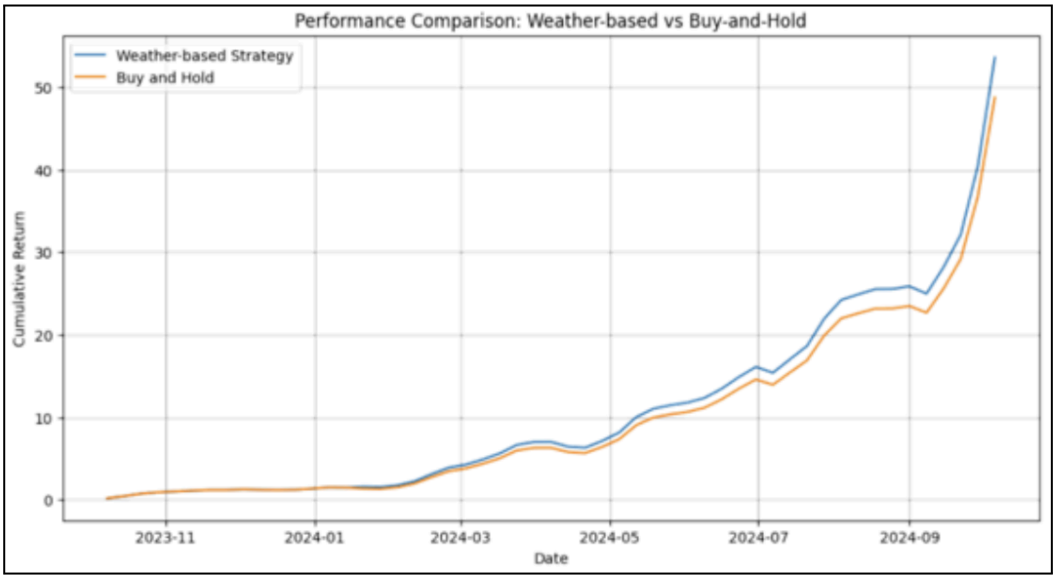


Figure 12: Train and Test Split

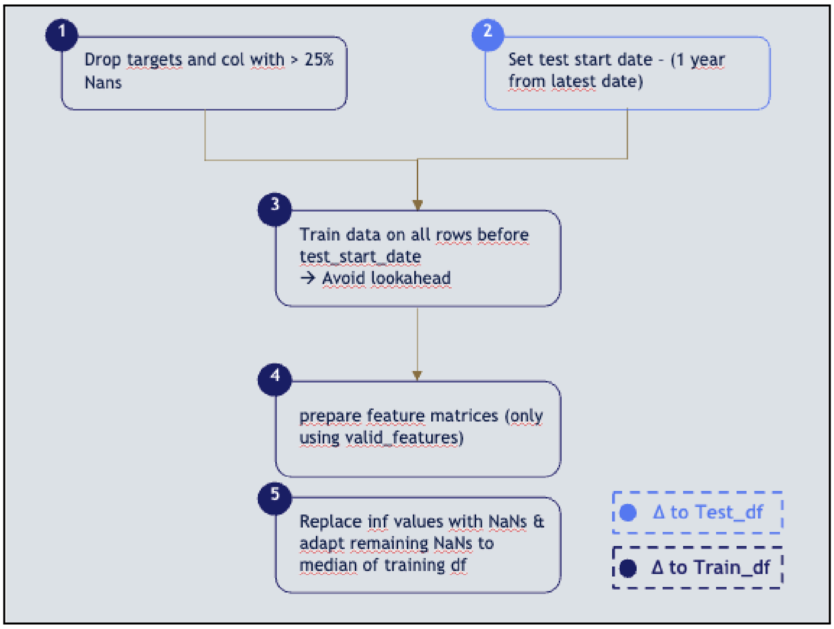
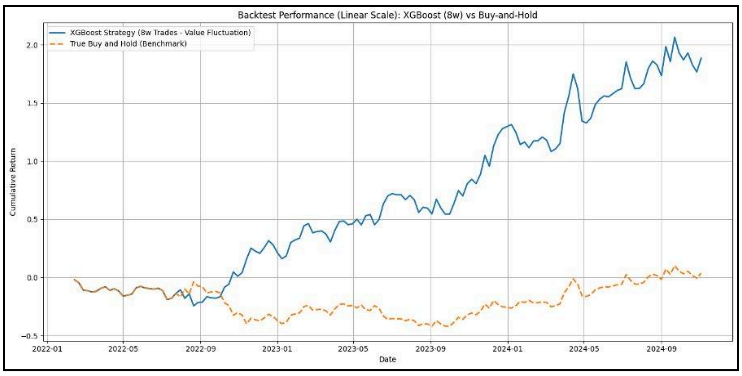
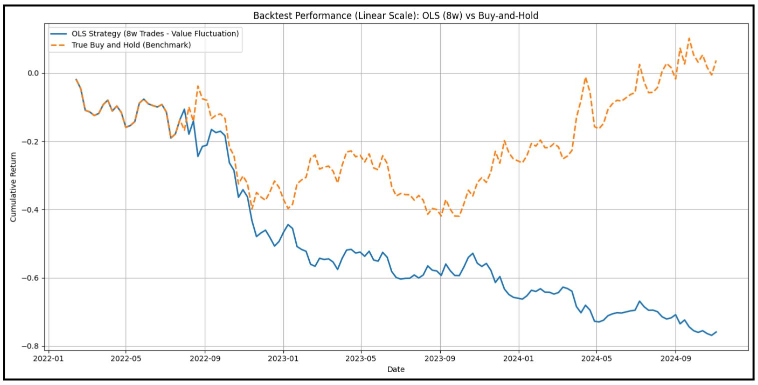


Figure 13: XGBoost Strategy Performance Evaluation



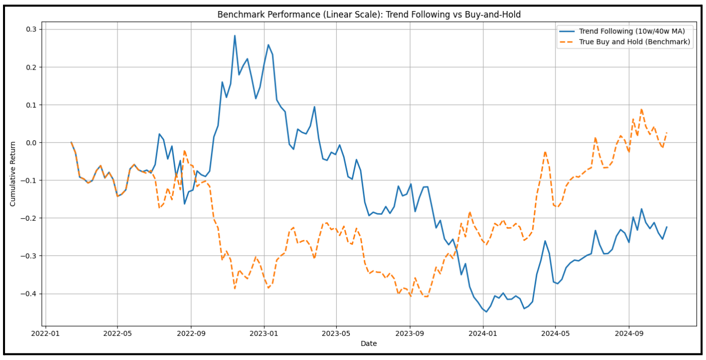

```
--- Strategy Performance (Annualized)
XGBoost Strategy (8w Trades):
  ARITHMETIC:
    Sharpe Ratio:          1.2418
    Annualized Return:     0.3993
    Annualized Volatility: 0.3216
  GEOMETRIC (Compounded):
    CAGR:                  0.4761

True Buy-and-Hold:
  ARITHMETIC:
    Sharpe Ratio:          0.2046
    Annualized Return:     0.0681
    Annualized Volatility: 0.3331
  GEOMETRIC (Compounded):
    CAGR:                  0.0135
```

Figure 14: OLS Strategy Performance Evaluation




```
--- Strategy Performance (Annualized)
OLS Strategy (8w Trades):
  ARITHMETIC:
    Sharpe Ratio:            -0.9977
    Annualized Return:       -0.3329
    Annualized Volatility: 0.3337
  GEOMETRIC (Compounded):
    CAGR:                    -0.4073

True Buy-and-Hold:
  ARITHMETIC:
    Sharpe Ratio:            0.2046
    Annualized Return:       0.0681
    Annualized Volatility: 0.3331
  GEOMETRIC (Compounded):
    CAGR:                    0.0135
```

Figure 15: Trend Following Strategy Performance Evaluation




```
--- Benchmark Strategy Performance (Annualized) --
Trend Following Strategy (10w / 40w MA Crossover):
  ARITHMETIC:
    Sharpe Ratio:            -0.1124
    Annualized Return:       -0.0373
    Annualized Volatility: 0.3323
  GEOMETRIC (Compounded):
    CAGR:                    -0.0891

True Buy-and-Hold (Test Period):
  ARITHMETIC:
    Sharpe Ratio:            0.1926
    Annualized Return:       0.0640
    Annualized Volatility: 0.3322
  GEOMETRIC (Compounded):
    CAGR:                    0.0096
```

Figure 16: Strategy Comparison

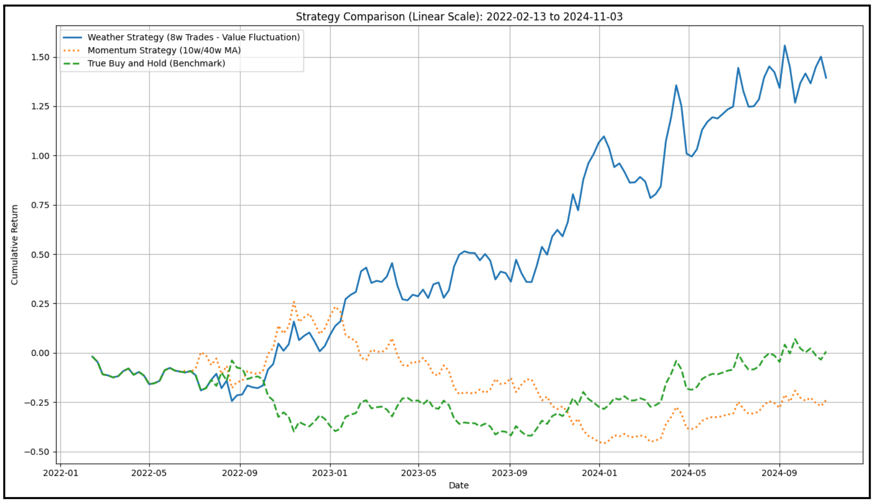


Figure 17: Backtest Performance (Linear Scale): XGBoost (8w) vs. Buy-and-Hold

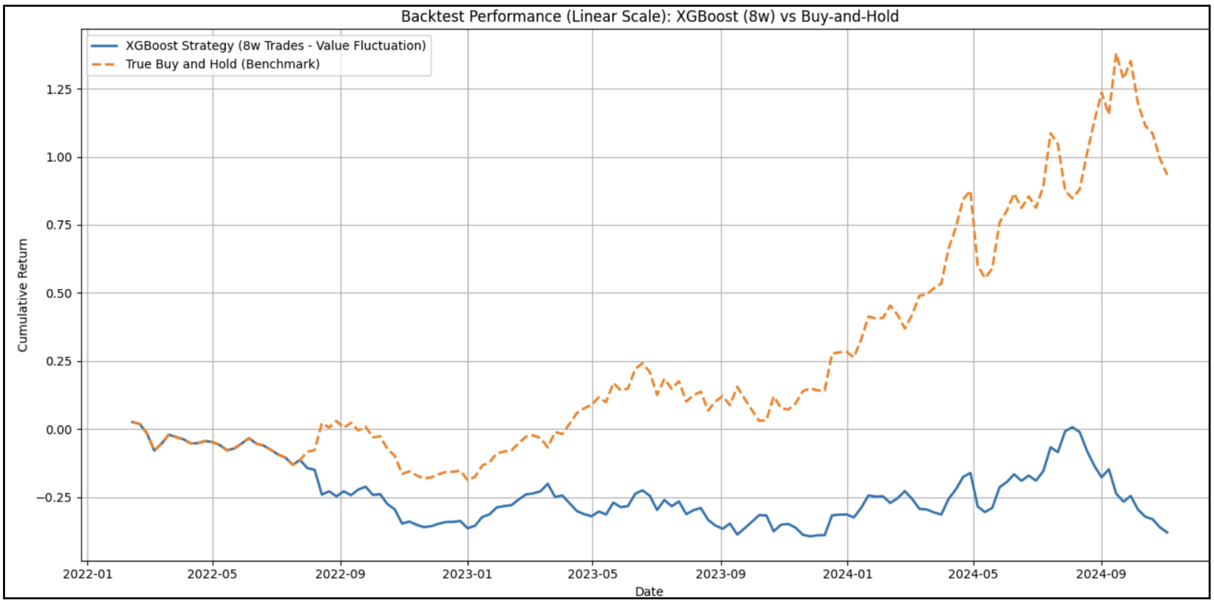
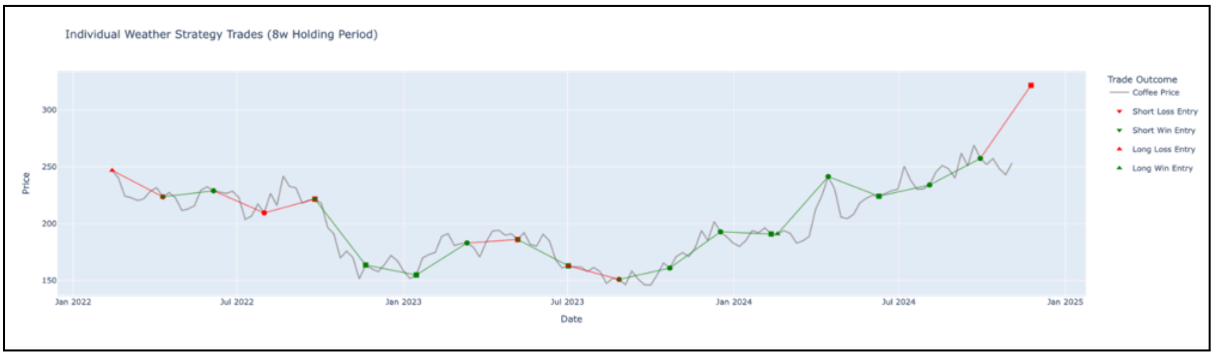


Figure 18: Good Arabica Results and Poor Robusta Results

Arabica Results



Robusta Results

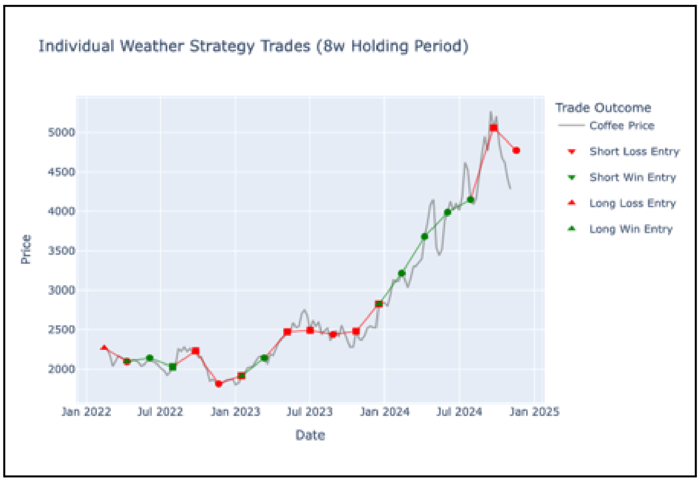
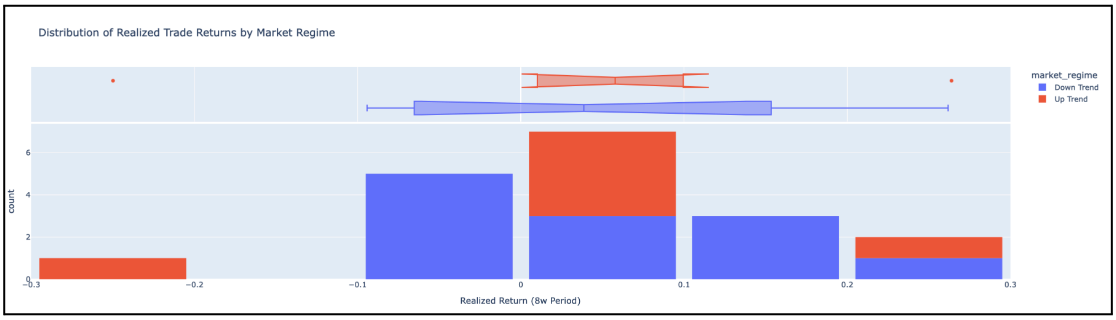


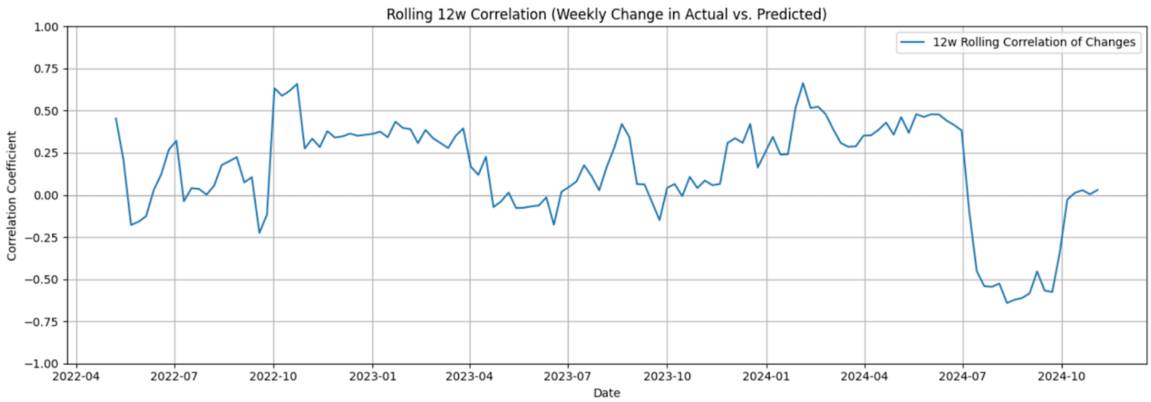
Figure 19: Distribution of Realized Trade Returns by Market Regime



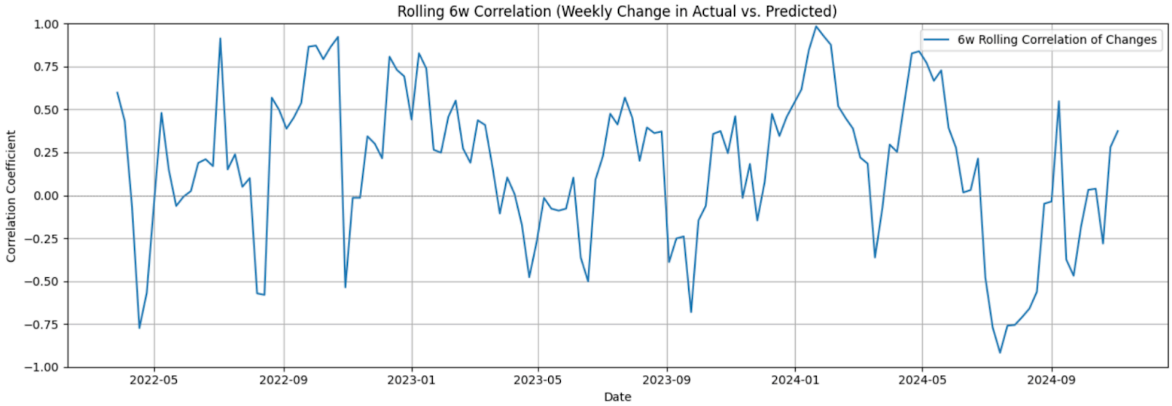

```
11. Defining Market Regimes...
Market regimes defined (based on price vs MA). Value Counts:
market_regime
Down Trend     90
Up Trend       52
Name: count, dtype: int64

12. Calculating Realized Returns & Win Rate...

Total Trades Initiated: 18
Winning Trades: 12
Overall Win Rate: 66.67%

13. Analyzing Win Rate by Market Regime...

Win Rate per Market Regime:
               num_trades  num_wins win_rate_pct
market_regime
Down Trend             12         7       58.33%
Up Trend                6         5       83.33%
```

Figure 20: Week-over-Week Change in Predicted Returns vs. Actual Returns

```
- Plotting Rolling Correlation of Changes...
  Saved rolling correlation plot to: /content/drive/MyDrive/QF206_Project/actual_vs_predicted_plot_multihoriz_robusta_changes_roll_corr.png
```



```
- Plotting Rolling Correlation of Changes...
  Saved rolling correlation plot to: /content/drive/MyDrive/QF206_Project/actual_vs_predicted_plot_multihoriz_robusta_changes_roll_corr.png
```




```
- Plotting Rolling Correlation of Changes...
```

```
- Plotting Rolling Correlation of Changes...
  Saved rolling correlation plot to: /content/drive/MyDrive/QF206_Project/actual_vs_predicted_plot_multihoriz_robusta_changes_roll_corr.png
```