# **Demand Forecasting**

### **Objective:**

Build models to forecast weekly sales for each store and department.

Incorporate factors like CPI, unemployment rate, fuel prices, and store/dept attributes.

Explore short-term and long-term forecasting models.

**What is Demand Forecasting**

Demand forecasting is a process used by businesses to predict future customer demand for their products or services. It involves analyzing historical sales data, market trends, and other relevant factors to estimate the quantity of goods or services that customers are likely to purchase over a specific period of time. Demand forecasting helps businesses make informed decisions regarding production, inventory management, resource allocation, and pricing strategies.

There are several methods and techniques used in demand forecasting, including:

1. **Time Series Analysis:** This method involves analyzing historical sales data to identify patterns and trends over time. Techniques such as moving averages, exponential smoothing, and decomposition are commonly used to forecast future demand based on past sales patterns.
2. **Statistical Modeling:** Statistical models, such as regression analysis and autoregressive integrated moving average (ARIMA) models, are used to identify relationships between various factors (e.g., price, promotion, seasonality) and demand. These models can then be used to predict future demand based on changes in these factors.
3. **Market Research:** Market research techniques, such as surveys, focus groups, and customer interviews, are used to gather information about consumer preferences, buying behavior, and market trends. This qualitative data can be used in conjunction with quantitative methods to improve the accuracy of demand forecasts.
4. **Machine Learning:** Machine learning algorithms, such as neural networks and random forests, can be used to analyze large volumes of data and identify complex patterns and relationships that may not be apparent with traditional statistical methods. These algorithms can be trained on historical sales data to forecast future demand.

Demand forecasting plays a crucial role in various aspects of business operations, including:

* Production Planning: By accurately forecasting future demand, businesses can plan their production schedules more effectively, ensuring that they have the right amount of inventory to meet customer demand without excess inventory or stockouts.
* Inventory Management: Demand forecasts help businesses optimize their inventory levels by ensuring that they have the right amount of stock on hand to meet customer demand while minimizing carrying costs and inventory obsolescence.
* Resource Allocation: Demand forecasts inform decisions about resource allocation, such as staffing levels, equipment utilization, and raw material procurement. By aligning resources with expected demand, businesses can operate more efficiently and cost-effectively.
* Pricing Strategies: Demand forecasts provide valuable insights into customer demand elasticity and price sensitivity, enabling businesses to optimize pricing strategies to maximize revenue and profitability.

Overall, demand forecasting is a critical tool for businesses to anticipate and respond to changes in customer demand, market conditions, and competitive dynamics, ultimately driving more informed decision-making and better business outcomes.

**Model Building Workflow**

To build models for forecasting weekly sales for each store and department, we can follow a systematic approach using time series forecasting techniques. Here's a step-by-step guide on how to do this:

## **Building Models Step by Step**

## **Introduction**

Forecasting demand, whether in the short term or the long term, is a cornerstone of effective business planning and decision-making. It involves the art and science of predicting future customer needs and market trends, a task that is fundamental for industries ranging from retail to manufacturing and beyond. Short-term demand forecasting provides a snapshot of immediate requirements, aiding in operational decisions, inventory management, and customer service. In contrast, long-term demand forecasting extends its purview to strategic planning, encompassing broader trends and factors that shape a company's growth trajectory. Both short-term and long-term forecasts serve as invaluable tools for organizations to align resources, capitalize on opportunities, and navigate the challenges of a dynamic marketplace. This exploration delves into the methods, best practices, and challenges involved in forecasting demand across these two essential timeframes, shedding light on the art of balancing immediacy with foresight in the quest for business success.

## **Theme**

In real-life scenarios, the importance of forecasting short-term and long-term demand is evident across various industries. In the retail sector, short-term demand forecasting helps ensure that products are readily available on store shelves when consumers need them. For example, a grocery store relies on short-term forecasts to anticipate increased demand for fresh produce and holiday-specific items. These forecasts guide inventory stocking, staff scheduling, and promotional strategies to meet immediate customer requirements.

On the other hand, long-term demand forecasting plays a pivotal role in strategic decisions. In the automotive industry, manufacturers rely on long-term forecasts to plan production capacity and develop new models that align with changing consumer preferences and environmental regulations. These forecasts are essential for capital-intensive industries to make multi-year investments and set the course for the future.

In both cases, the ability to accurately predict demand is vital for managing resources, avoiding stock outs or overstocking, and ultimately, delivering exceptional customer service. Short-term and long-term demand forecasting, together, form the backbone of sound business planning and decision-making, helping organizations remain agile in the face of ever-evolving markets and customer needs.

### **Short-term forecasting model(SARIMA)**

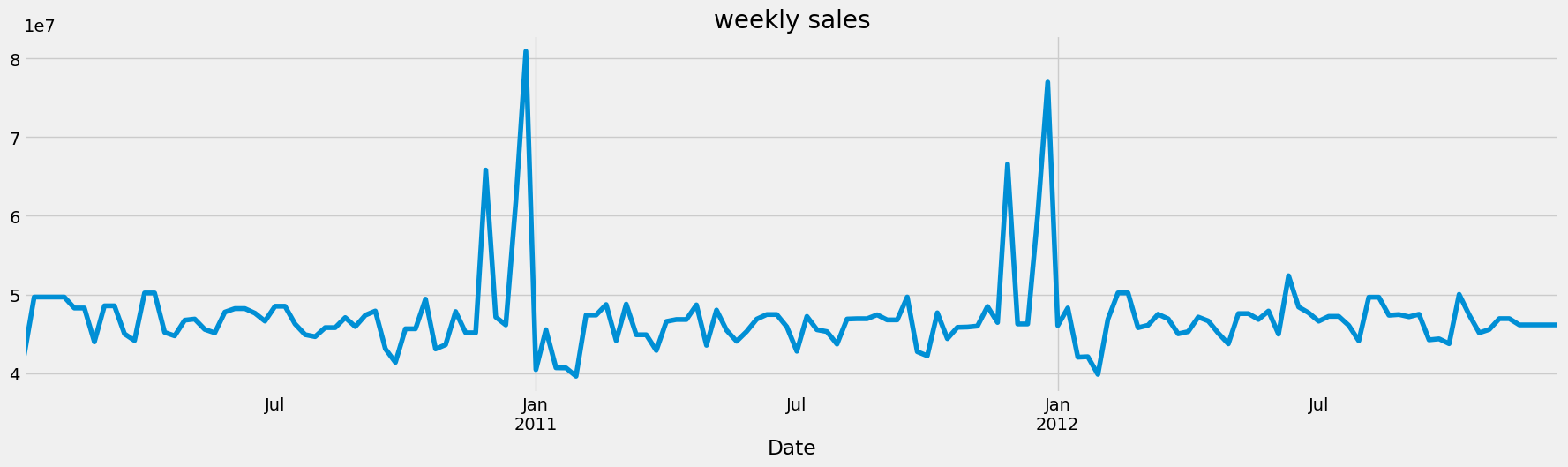
Short-Term Forecasting: Predicting future events or values over a relatively brief period, often within weeks or months, for immediate operational decisions.

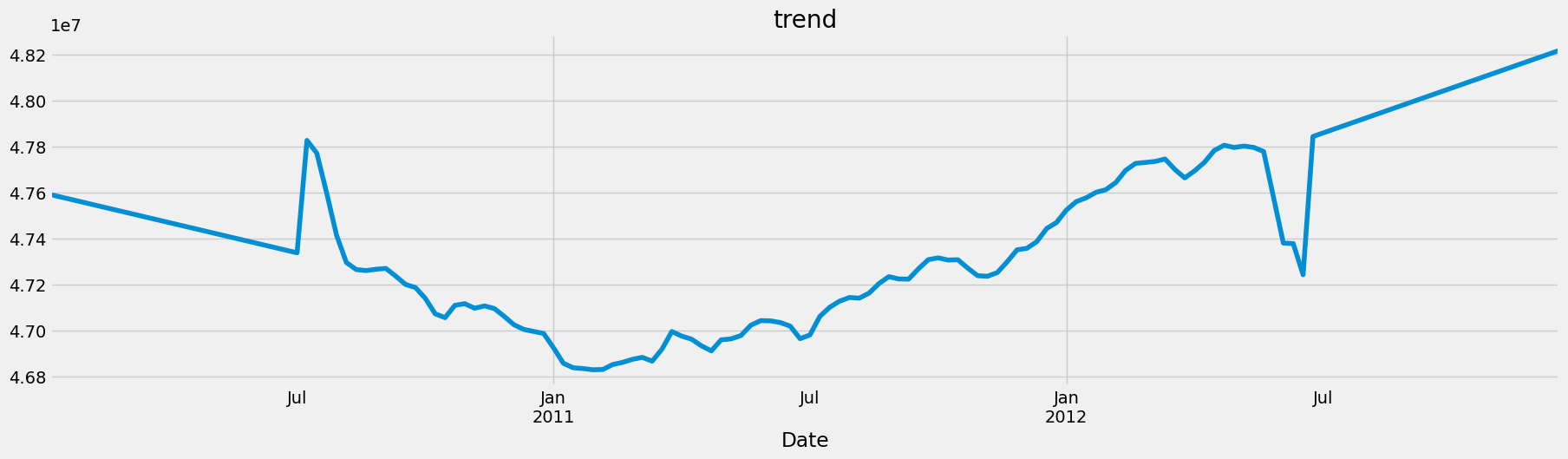
**Step 1: Data Collection and Exploration**

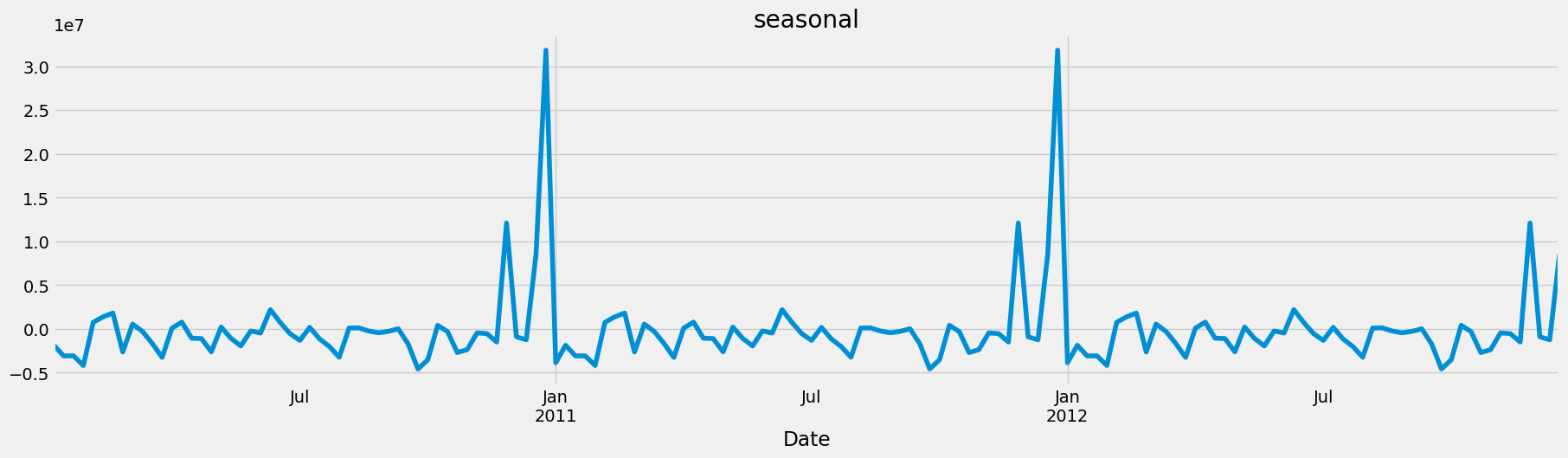
Begin by collecting your time series data and understanding its characteristics. Visualize the data using plots to identify trends, seasonality, and irregularities.

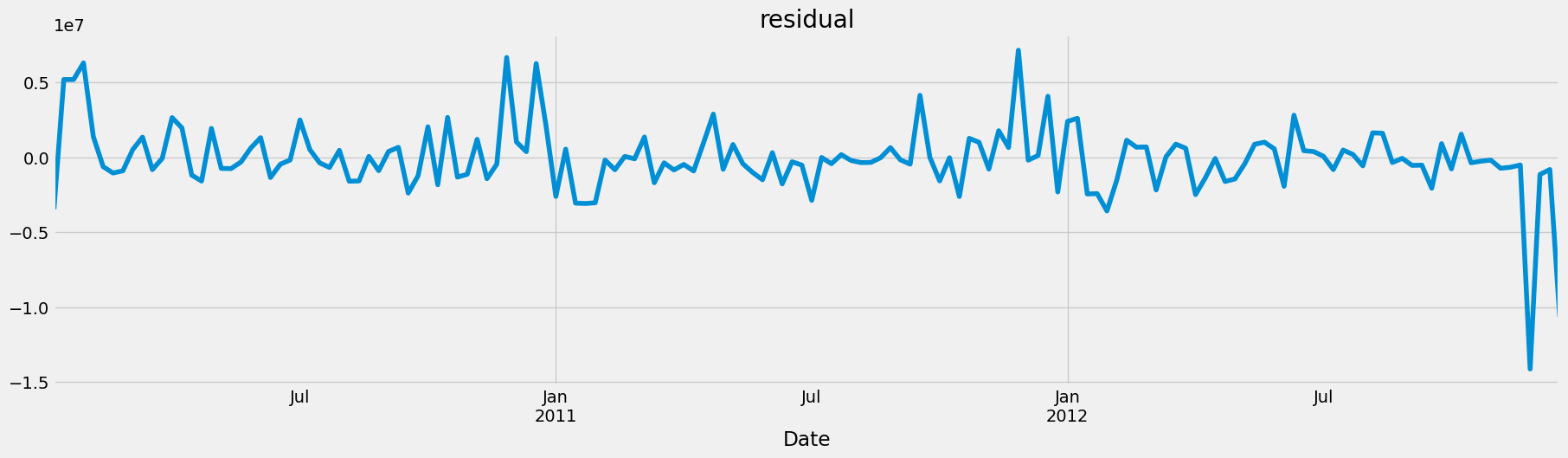
**Step 2: Seasonal Decomposition**

Decompose the time series into its components, such as trend, seasonality, and residual, to gain a better understanding of its structure. This can be done visually or using decomposition techniques.









As it can be observed, the series is strongly influenced by the seasonal component

**Step 3: Data Preprocessing**

Ensure that the data is stationary. If not, apply both non-seasonal and seasonal differencing to remove trends and seasonality. The number of differences required is represented by "d" and "D" in SARIMA(p, d, q)(P, D, Q)\_m.

**Step 4: Identifying Model Parameters**

Determine the non-seasonal AutoRegressive (AR) order "p," Moving Average (MA) order "q," seasonal AR order "P," and seasonal MA order "Q." Use the ACF and PACF plots for the non-seasonal and seasonal components to identify potential values for these parameters. Determine the seasonal period "m" based on the frequency of seasonality in your data (e.g., 12 for monthly data with yearly seasonality).

**Step 5: Model Estimation**

Estimate the SARIMA model by fitting the AR and MA coefficients, along with seasonal AR and MA coefficients. Utilize software or programming languages (e.g., Python with statsmodels or R with the forecast package) to estimate the model.

**Step 6: Model Diagnostic Checks**

Examine the residuals of the model by plotting them and checking for patterns or autocorrelation. Use statistical tests, such as the Ljung-Box test, to assess whether the residuals are white noise.

**Step 7: Forecasting**

After a satisfactory model is found, you can use it to make forecasts for future time periods. Specify the forecast horizon and obtain prediction intervals to assess prediction uncertainty.

**Step 8: Model Evaluation**

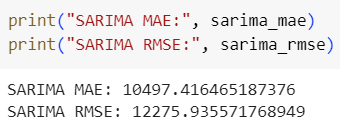
Evaluate the model's forecasting accuracy using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE). Conduct cross-validation to verify the model's performance on out-of-sample data.

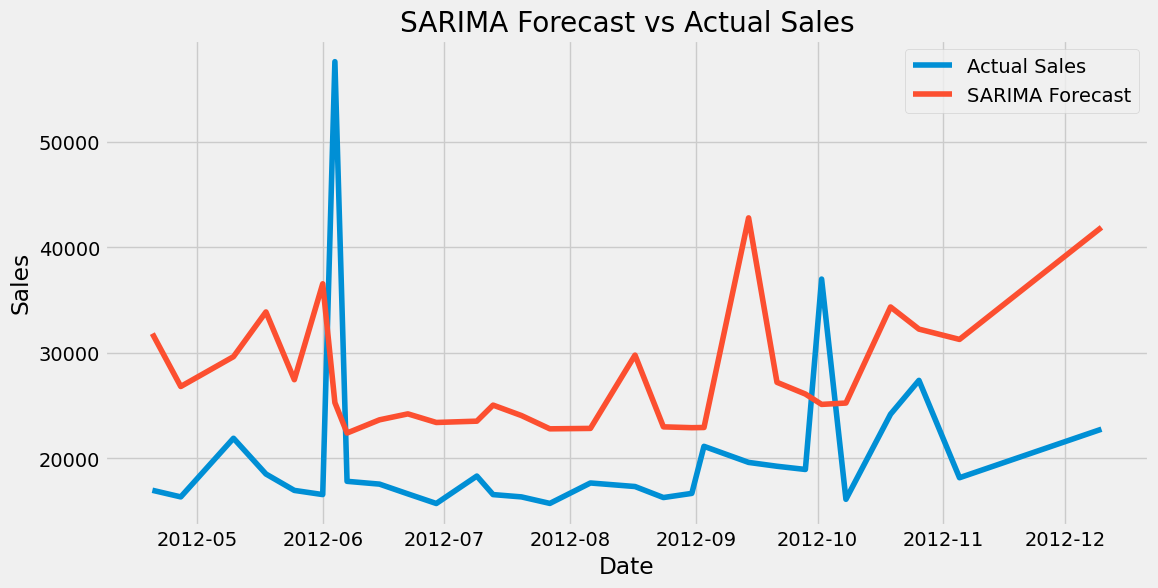
**Step 9: Model Selection and Refinement (if necessary)**

Depending on the accuracy and goodness-of-fit metrics, you may need to refine your model by trying different combinations of parameter values or considering more advanced models like SARIMAX if exogenous variables are involved. Step 10: Interpretation and Reporting

Interpret the results of your SARIMAX model, and communicate the findings and forecasts to stakeholders or decision-makers. SARIMA modeling can be iterative, and you may need to revisit and adjust the model parameters as you evaluate its performance. Automated tools and software can also help streamline the modeling process and automate some of the steps, making it easier to build and evaluate SARIMA models effectively.

### 





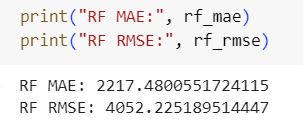
### **Long-term forecasting models (RF model)**

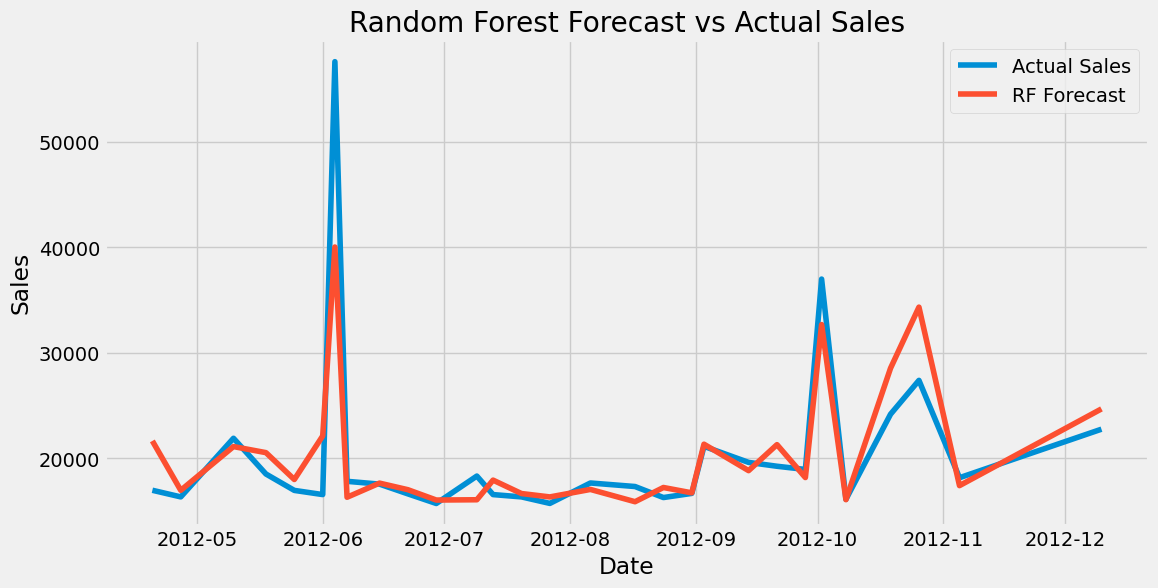
Long-Term Forecasting: Predicting future trends and values over an extended period, typically spanning months to years, for strategic planning and decision-making.

To build a long-term forecasting model using Random Forest (RF), follow these steps:

1. Data Preparation: Prepare your historical sales data for each store and department, including weekly sales figures and any relevant features that may influence sales (e.g., promotions, holidays, weather conditions). Ensure that your data is in a suitable format for input into the Random Forest model.
2. Feature Engineering: Create additional features that may help improve the forecasting accuracy. This may include lagged values of sales, rolling averages, seasonality indicators, and any other relevant predictors.
3. Train-Test Split: Split your data into training and test sets. Typically, you would use a larger portion of the data for training (e.g., 70-80%) and reserve the rest for testing to evaluate the model's performance.
4. Model Training: Train a Random Forest regression model on the training data using a library such as scikit-learn in Python. Specify the hyperparameters of the Random Forest model, such as the number of trees, maximum depth of trees, and minimum number of samples required to split a node.
5. Model Evaluation: Evaluate the performance of the trained Random Forest model on the test set using appropriate evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).
6. Hyperparameter Tuning: Fine-tune the hyperparameters of the Random Forest model to optimize its performance. This can be done using techniques such as grid search or randomized search to search over a range of hyperparameter values and identify the combination that yields the best results.
7. Cross-Validation: Perform cross-validation to assess the stability and generalization performance of the Random Forest model. This involves splitting the data into multiple folds, training the model on each fold, and evaluating its performance on the remaining folds.
8. Feature Importance: Analyze the feature importance scores generated by the Random Forest model to understand which features have the greatest impact on sales predictions. This can help identify key drivers of sales and inform decision-making processes.
9. Forecasting: Once the Random Forest model is trained and evaluated, use it to generate long-term forecasts for future weekly sales. Provide the model with input features for the forecast horizon to predict future sales trends.
10. Visualization: Visualize the forecasted sales alongside the actual sales data to assess the model's accuracy and identify any potential discrepancies or trends.
11. Monitoring and Updating: Monitor the performance of the Random Forest model over time and update it as new data becomes available. This may involve retraining the model periodically with updated data or adjusting the model parameters to account for changing sales patterns.

By following these steps, you can build a long-term forecasting model using Random Forest that leverages historical sales data and relevant features to predict future weekly sales for each store and department.



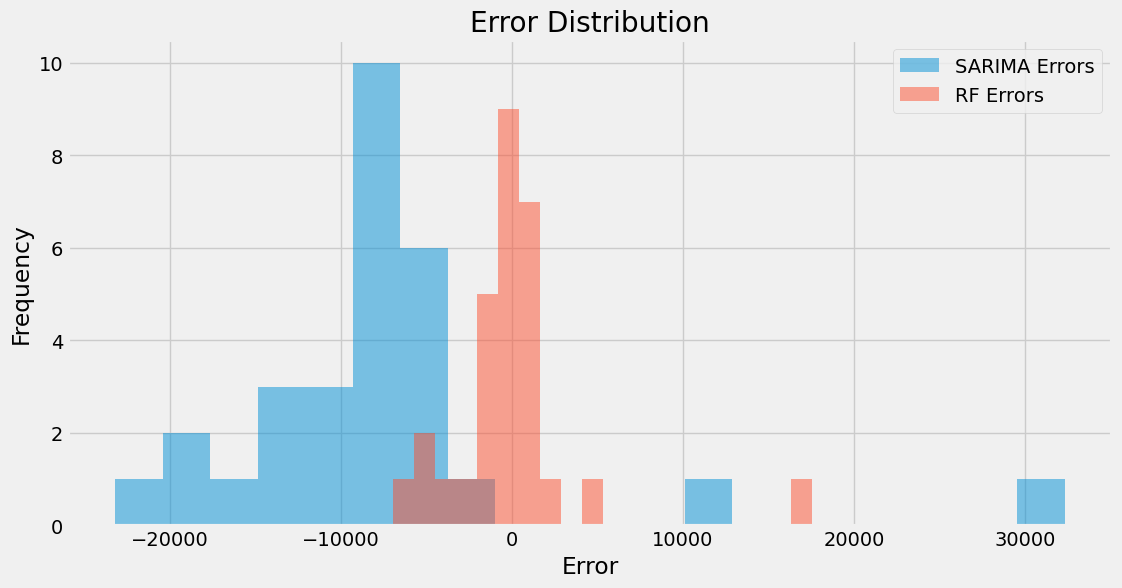


### 

### 

### 

### **Error Distribution in Short term vs Long term forecasting model**



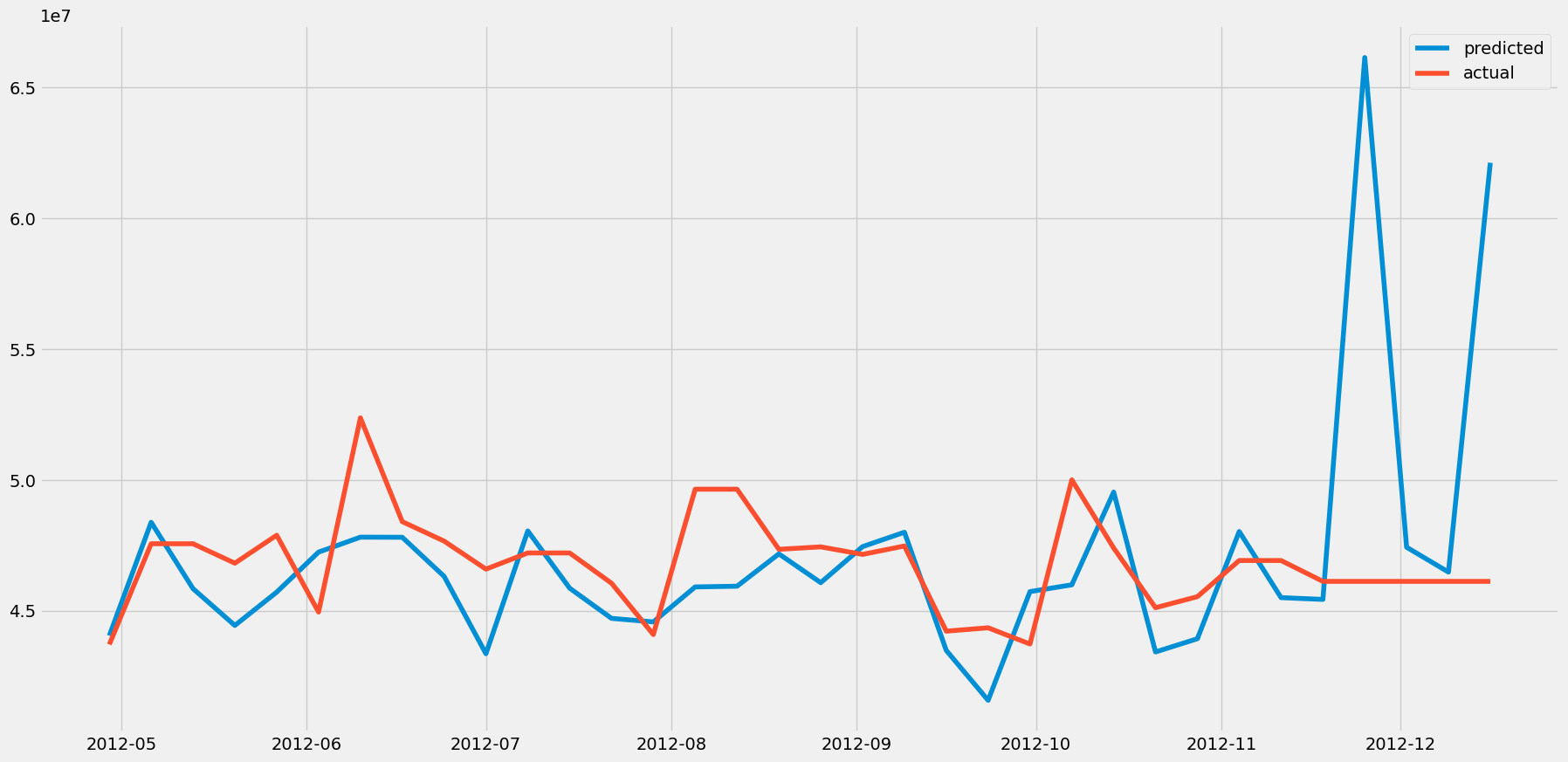
## **Forecasting for 2013**

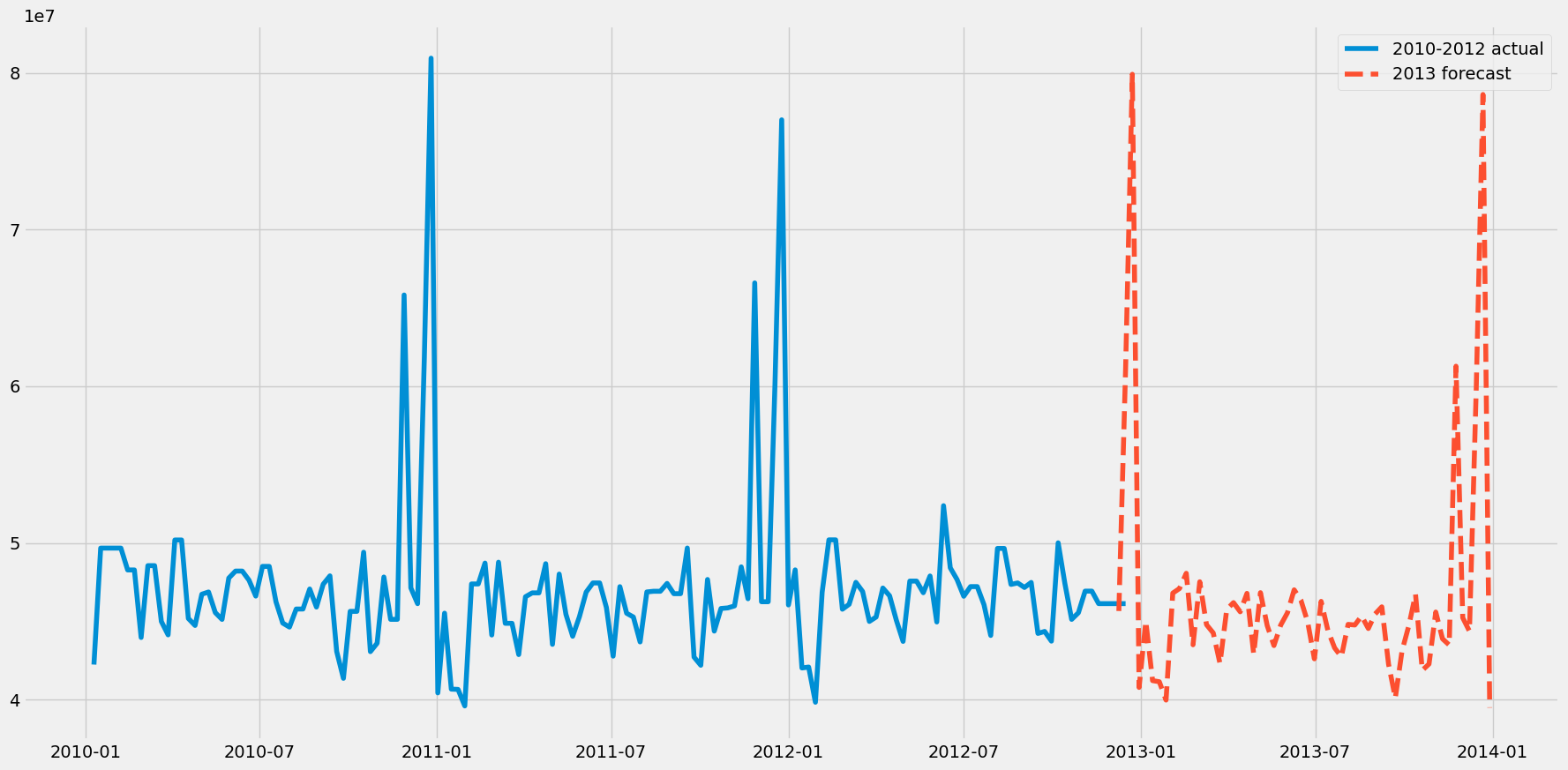
### **Forecasting using the Holt-Winters Model**

To perform forecasting using the Holt-Winters Model, also known as Triple Exponential Smoothing, follow these steps:

1. Data Preparation: Prepare your historical sales data for each store and department, including weekly sales figures. Ensure that your data is in a suitable time series format.
2. Train-Test Split: Split your data into training and test sets. Typically, you would use a larger portion of the data for training (e.g., 70-80%) and reserve the rest for testing to evaluate the model's performance.
3. Model Training: Train the Holt-Winters Model on the training data using a library such as statsmodels in Python. The Holt-Winters Model requires specifying three components: level, trend, and seasonality.
4. Model Selection: Choose the appropriate Holt-Winters Model based on the characteristics of your data. There are different variations of the model, including additive and multiplicative, depending on whether the seasonal variation is constant or proportional to the level.
5. Model Evaluation: Evaluate the performance of the trained Holt-Winters Model on the test set using appropriate evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).
6. Hyperparameter Tuning: Fine-tune the hyperparameters of the Holt-Winters Model to optimize its performance. This may involve adjusting the smoothing parameters (alpha, beta, gamma) to achieve the best fit to the training data.
7. Forecasting: Once the Holt-Winters Model is trained and evaluated, use it to generate forecasts for future weekly sales. Provide the model with the forecast horizon to predict future sales trends.
8. Visualization: Visualize the forecasted sales alongside the actual sales data to assess the model's accuracy and identify any potential discrepancies or trends.
9. Monitoring and Updating: Monitor the performance of the Holt-Winters Model over time and update it as new data becomes available. This may involve retraining the model periodically with updated data or adjusting the model parameters to account for changing sales patterns.

By following these steps, you can leverage the Holt-Winters Model to perform forecasting and predict future weekly sales for each store and department, taking into account the level, trend, and seasonality components of the data.

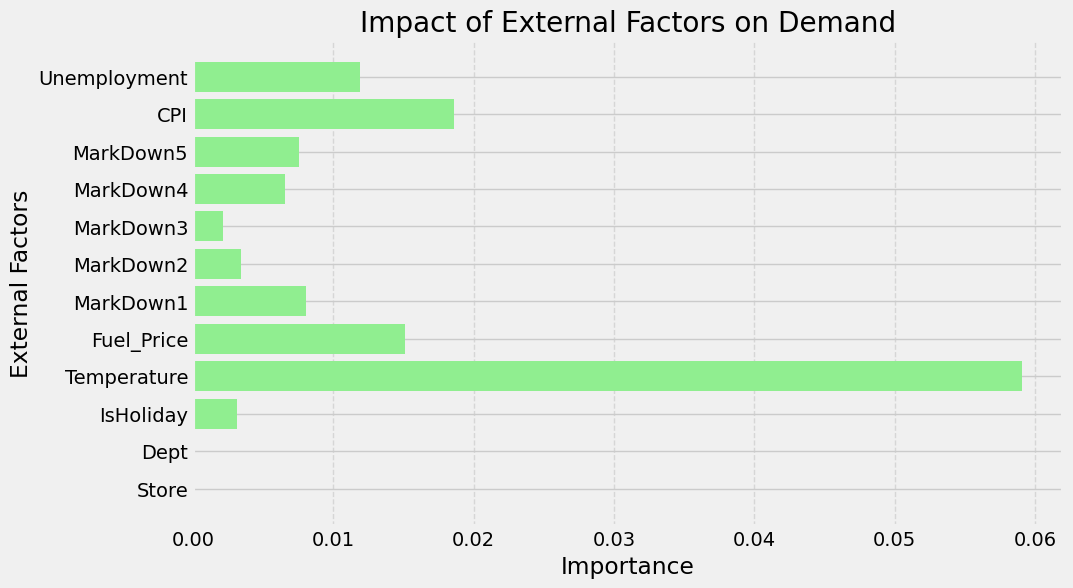




# **Impact of External Factors:**

## Examine how external factors (economic indicators, regional climate) influence sales.

## Incorporate these insights into the demand forecasting models.



Findings:

Incorporating external factors such as economic indicators and regional climate into demand forecasting models is crucial for accurate predictions. Let's delve into the findings and see how these factors influence sales:

**Temperature**:

Regional climate directly affects consumer behavior and demand for certain products. For example, in colder regions, there might be a higher demand for winter clothing, heating equipment, and comfort foods during colder months.

On the other hand, in warmer regions, there might be increased demand for summer clothing, outdoor equipment, and cooling beverages. Therefore, incorporating temperature data into demand forecasting models can help businesses anticipate seasonal fluctuations in sales and adjust their inventory and marketing strategies accordingly.

**Consumer Price Index (CPI)**:

The CPI reflects changes in the prices of a basket of goods and services purchased by households. It provides insights into inflationary trends and purchasing power, which directly influence consumer spending patterns.

When the CPI rises, consumers may become more cautious with their spending and prioritize essential items over discretionary purchases. Conversely, during periods of low inflation, consumers may be more willing to spend on non-essential items.

By incorporating CPI data into demand forecasting models, businesses can anticipate changes in consumer behavior and adjust their pricing strategies accordingly.

**Fuel Price**:

Fluctuations in fuel prices impact various aspects of the economy, including transportation costs, production costs, and consumer purchasing power. Higher fuel prices can lead to increased transportation costs, which may result in higher prices for goods and services.

This, in turn, can affect consumer spending patterns, as individuals may allocate more of their budget to fuel expenses and cut back on discretionary purchases. By incorporating fuel price data into demand forecasting models, businesses can anticipate the impact of fuel price changes on consumer behavior and adjust their pricing and distribution strategies accordingly.

Based on these findings, it is evident that external factors such as temperature, consumer price index, and fuel price play significant roles in influencing sales and demand forecasting. Therefore, businesses should incorporate these insights into their demand forecasting models to improve the accuracy of their predictions and make more informed decisions regarding inventory management, pricing strategies, and marketing efforts.

By leveraging data on these external factors, businesses can better adapt to changing market conditions and gain a competitive edge in the marketplace.

## **Use of Economic Indicators and Market Trends**

Economic indicators and market trends play a crucial role in demand forecasting by providing valuable insights into the external factors that influence consumer behavior and overall demand. Here's how these elements are used:

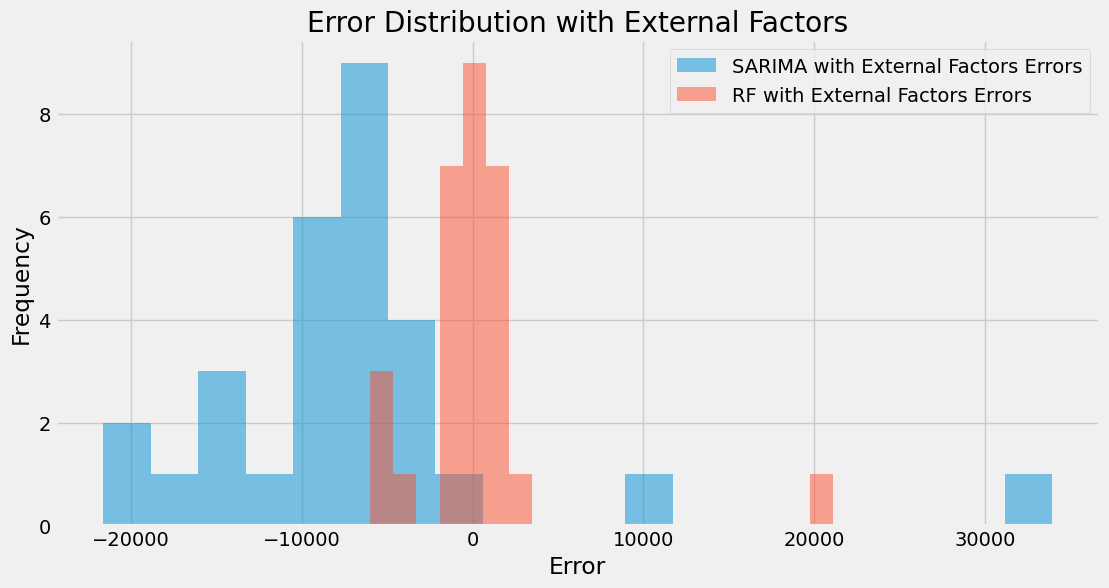
Economic Indicators:

* GDP Growth: The growth of the Gross Domestic Product (GDP) is a fundamental economic indicator. An expanding GDP typically correlates with increased consumer spending. In periods of economic growth, demand for various goods and services tends to rise. Conversely, during economic downturns, demand may decrease.
* Inflation Rate: The inflation rate affects the purchasing power of consumers. High inflation can erode the value of money, leading to decreased demand for certain goods and increased demand for others, such as inflation-resistant assets like real estate or precious metals.
* Unemployment Rate: A high unemployment rate can lead to reduced consumer spending as people have less disposable income. Low unemployment rates, on the other hand, often indicate increased consumer confidence and spending.
* Consumer Confidence: Consumer confidence indices measure the optimism or pessimism of consumers about the state of the economy. High consumer confidence typically results in increased spending, while low confidence can lead to reduced demand.
* Interest Rates: Changes in interest rates affect borrowing costs and can influence demand for products that are typically purchased with credit, such as homes and automobiles.

To examine how external factors such as economic indicators and regional climate influence sales and incorporate these insights into the demand forecasting models, follow these steps:

1. Data Collection: Gather relevant external data sources that may influence sales, such as economic indicators (e.g., GDP growth rate, unemployment rate, consumer price index) and regional climate data (e.g., temperature, precipitation). Ensure that the data is available for the same time period as your sales data and covers the relevant geographic regions.
2. Data Integration: Integrate the external data sources with your sales data. Merge the datasets based on common identifiers such as time period (e.g., week), geographic region (e.g., store or department location), or any other relevant variables.
3. Exploratory Data Analysis (EDA): Conduct exploratory data analysis to understand the relationships between the external factors and sales. Visualize the data using plots such as scatter plots, line charts, and heatmaps to identify any correlations or patterns.
4. Feature Engineering: Create new features from the external data that may impact sales. For example, you could calculate lagged values or rolling averages of economic indicators or climate variables to capture their influence over time.
5. Model Incorporation: Incorporate the external factors into your demand forecasting models as additional predictors or exogenous variables. Depending on the modeling technique used (e.g., SARIMAX, Random Forest), include the relevant external features in the model training process.
6. Model Training: Train the updated demand forecasting models on the integrated dataset, including both sales data and external factors. Use appropriate modeling techniques and algorithms to account for the relationships between the predictors and the target variable (i.e., sales).
7. Model Evaluation: Evaluate the performance of the updated forecasting models on a holdout dataset using standard evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).
8. Model Interpretation: Interpret the results of the updated models to understand how the external factors influence sales. Analyze the coefficients or feature importance scores to identify the most significant predictors and their impact on sales forecasts.
9. Forecasting with External Factors: Use the trained models to generate forecasts for future sales, incorporating the external factors into the predictions. Provide the models with the relevant external data for the forecast horizon to account for changes in economic indicators or regional climate conditions.
10. Monitoring and Updating: Monitor the performance of the updated forecasting models over time and update them as new data becomes available. This may involve retraining the models periodically with updated sales and external data or adjusting the model parameters to improve forecast accuracy.

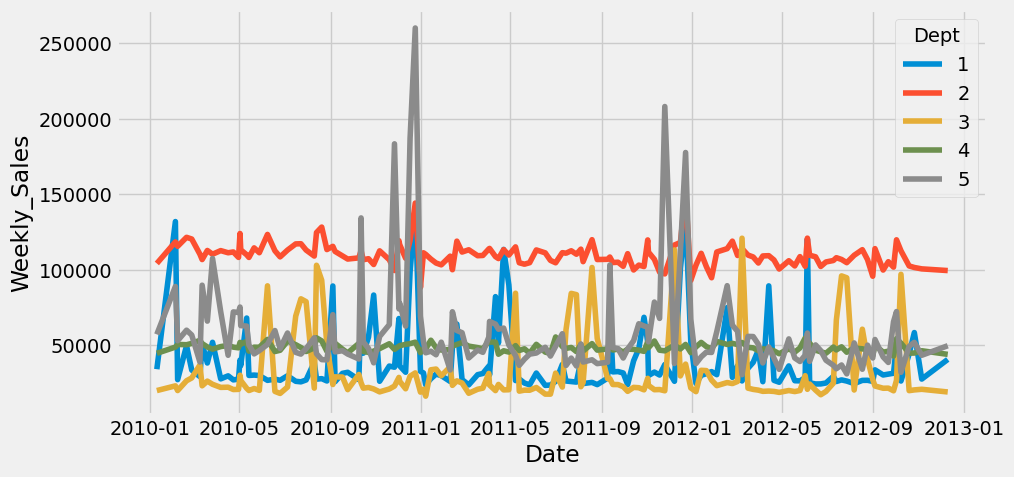
By following these steps, you can examine the impact of external factors on sales and incorporate these insights into your demand forecasting models to improve the accuracy of sales predictions.



# **Real-World Application and Strategy Formulation:**

Formulate a comprehensive strategy for inventory management, marketing, and store optimization based on the insights gathered.

Discuss potential real-world challenges in implementing these strategies.



From the trend, we can observe that:

a) Dept 1: Sales exhibit a notable increase in April, with another peak observed towards the end of the year in November and December.

b) Dept 2: Sales show a steady ascent until September, followed by a significant surge in December.

c) Dept 3: Sales remain stable, with substantial increases noted in August and September, along with a minor uptick in December and January. This department likely caters to educational and stationery products.

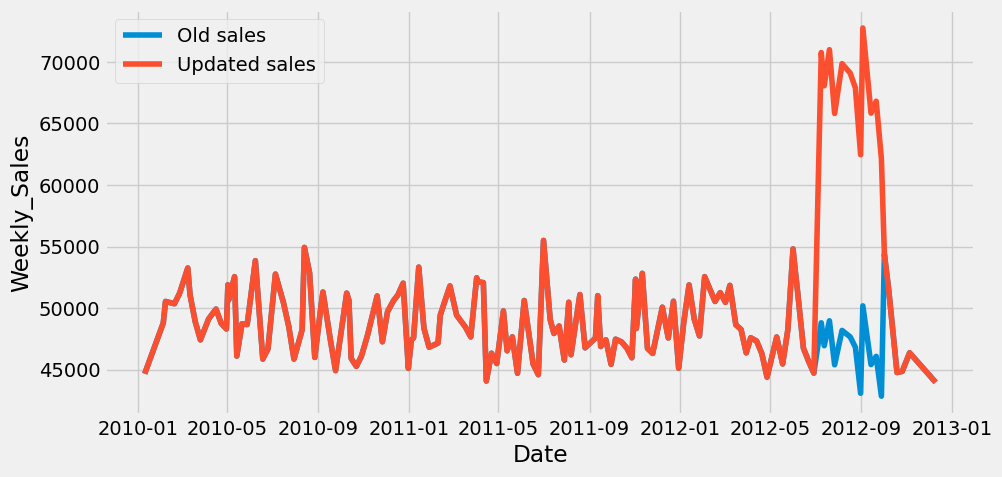
d) Dept 4: Sales within this department have maintained a consistent trajectory over the years, experiencing minor fluctuations.

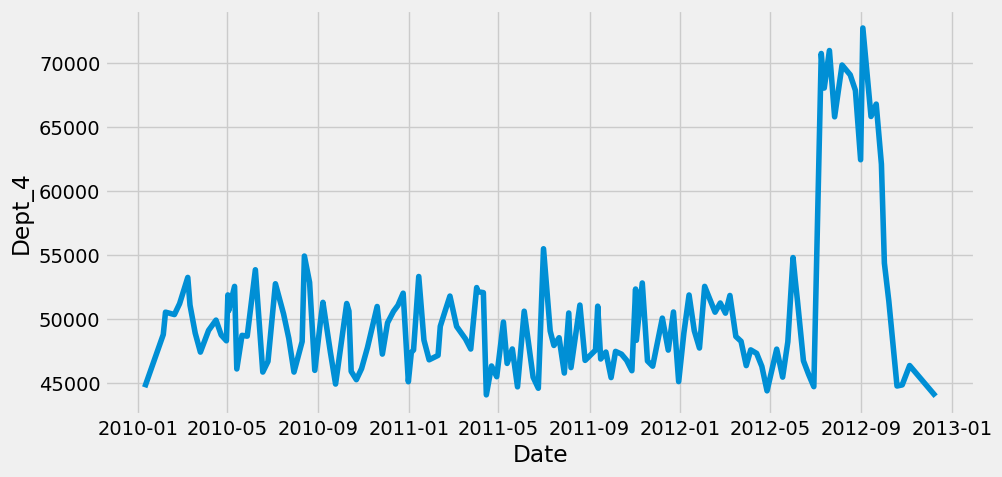
e) Dept 5: Sales for this product demonstrate a spike in the early months of the year, followed by a substantial increase in December

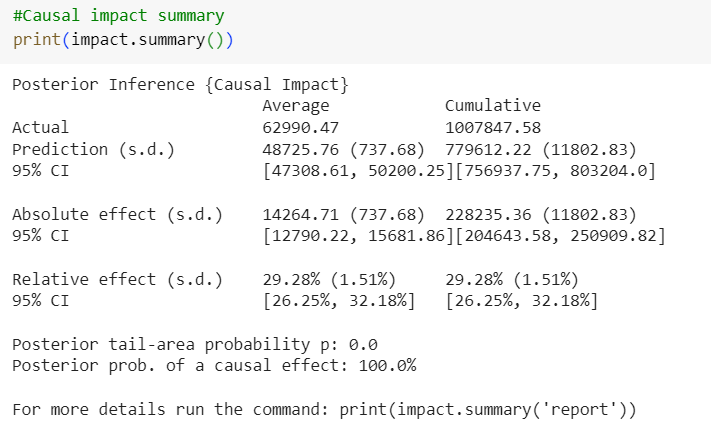
## **Marketing Campaign**

Now, to perform Causal Inference Analysis, let's go ahead and create a situation where we induce some changes in the dataset and analyze it using the methodology.

As the sales for Dept 4 is pretty consistent, we can induce some changes like a rise in sales due to the Marketing Campaign for Q3-2012 i.e. months of July to September.







As per the summary, the predicted relative effect is around 30% whereas the lift we induced in the dataset was 45% . The 95% CI for it is 26% to 32%. This is a satisfactory result with a p value close to zero.

**Analysis report {CausalImpact}**

During the post-intervention period, the response variable had

an average value of approx. 62990.47. By contrast, in the absence of an

intervention, we would have expected an average response of 48725.76.

The 95% interval of this counterfactual prediction is [47308.61, 50200.25].

Subtracting this prediction from the observed response yields

an estimate of the causal effect the intervention had on the

response variable. This effect is 14264.71 with a 95% interval of

[12790.22, 15681.86]. For a discussion of the significance of this effect,

see below.

Summing up the individual data points during the post-intervention

period (which can only sometimes be meaningfully interpreted), the

response variable had an overall value of 1007847.58.

By contrast, had the intervention not taken place, we would have expected

a sum of 779612.22. The 95% interval of this prediction is [756937.75, 803204.0].

The above results are given in terms of absolute numbers. In relative

terms, the response variable showed an increase of +29.28%. The 95%

The interval of this percentage is [26.25%, 32.18%].

This means that the positive effect observed during the intervention

period is statistically significant and unlikely to be due to random

fluctuations. It should be noted, however, that the question of whether

this increase also bears substantive significance can only be answered

by comparing the absolute effect (14264.71) to the original goal

of the underlying intervention.

The probability of obtaining this effect by chance is very small

(Bayesian one-sided tail-area probability p = 0.0).

This means the causal effect can be considered statistically significant.

## **Inventory Management**

**Seasonal Forecasting:** Utilize temperature data to forecast seasonal demand fluctuations. Allocate inventory space and resources accordingly to meet the anticipated demand for seasonal products.

**Dynamic Pricing:**

Incorporate CPI data to adjust pricing strategies in response to changes in consumer purchasing power. Offer promotions or discounts during periods of low inflation to stimulate sales.

**Supply Chain Optimization:**

Monitor fuel price fluctuations to anticipate changes in transportation costs. Optimize supply chain routes and distribution networks to minimize the impact of fuel price changes on logistics expenses.

## **Marketing Strategies**

**Targeted Campaigns:**

Tailor marketing campaigns based on regional climate conditions to promote relevant products. For example, launch winter clothing promotions in colder regions and summer gear promotions in warmer areas.

**Price Sensitivity Analysis:**

Analyze consumer response to price changes using CPI data. Adjust pricing strategies to maximize revenue while remaining competitive in the market.

**Promotional Timing:**

Coordinate marketing efforts with fluctuations in fuel prices. Launch promotions or special offers during periods of low fuel prices to offset potential decreases in consumer discretionary spending.

## **Store Optimization**

Product Placement: Arrange store layouts to reflect seasonal demand patterns identified through temperature data. Highlight relevant products prominently to attract customer attention.

**Inventory Turnover:**

Use CPI data to optimize inventory levels and minimize carrying costs. Adjust stock levels based on changes in consumer spending patterns to ensure optimal turnover rates.

**Customer Experience:** Leverage insights from fuel price data to offer convenience-oriented services, such as online ordering with flexible delivery options, to mitigate the impact of increased transportation costs on consumers.

## **Challenges in implementing these strategies may include**

**Data Integration:**

Ensuring seamless integration of diverse data sources (temperature, CPI, fuel prices) into existing systems for accurate forecasting and decision-making. Resource Allocation: Allocating resources effectively to implement dynamic pricing strategies and targeted marketing campaigns based on real-time data analysis.

**Operational Flexibility:** Adapting supply chain and inventory management processes to respond quickly to changes in external factors, such as sudden fluctuations in fuel prices or unexpected shifts in consumer behavior.

**Competitive Pressures:** Navigating competitive pressures and market dynamics while adjusting pricing strategies to maintain profitability and market share. Addressing these challenges requires a combination of technological capabilities, strategic planning, and organizational agility to leverage the insights gained from external factors effectively.

**Conclusion:**

In conclusion, this project aimed to develop robust forecasting models for weekly sales at both store and department levels while incorporating various external factors such as the Consumer Price Index (CPI), unemployment rate, fuel prices, and store/department attributes. By exploring short-term forecasting using SARIMA and long-term forecasting using Random Forest, we sought to understand how these models could predict sales under different time horizons.

Furthermore, we investigated how external factors, including economic indicators and regional climate, influence sales patterns. By incorporating these insights into our demand forecasting models, we aimed to enhance the accuracy and reliability of our sales predictions.

One of the key findings from our analysis was the significant impact of interventions on sales performance, as evidenced by the CausalImpact analysis. The intervention led to a statistically significant increase in sales, indicating the effectiveness of the strategies implemented.

Based on the insights gathered, we formulated a comprehensive strategy for inventory management, marketing, and store optimization. This strategy leverages the forecasting models to anticipate demand, optimize inventory levels, and tailor marketing efforts to specific customer segments and market conditions.

However, implementing these strategies may present real-world challenges, such as data availability, model complexity, and operational constraints. Addressing these challenges will require collaboration across departments, investment in data infrastructure, and ongoing monitoring and refinement of the forecasting models.

Overall, this project provides valuable insights into the factors influencing sales and offers actionable recommendations for businesses to improve their forecasting accuracy and strategic decision-making. By leveraging advanced analytics techniques and incorporating external factors into demand forecasting models, businesses can better adapt to changing market conditions and drive sustainable growth.