Project Report

# State your forecasting question and its importance to you

What will be the total weekly sales across all given Walmart stores in dataset for the next 10 weeks, and how can this forecast inform better business decisions?

Forecasting total weekly sales is crucial for:

* Inventory Management: Ensuring sufficient stock levels to meet predicted demand.
* Revenue Planning: Setting realistic revenue targets for the coming weeks and aligning marketing strategies accordingly.
* Strategic Decision-Making: Preparing for seasonal patterns or special promotions, such as holiday sales, which often drive higher revenue.

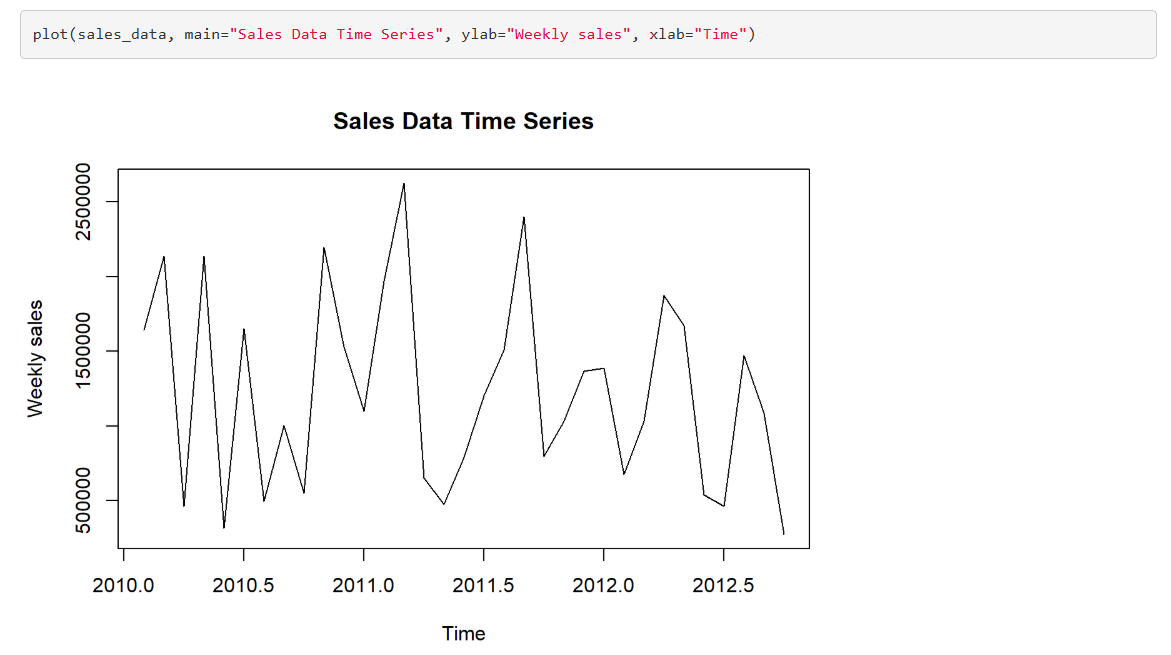
# Describe the data

The dataset contains weekly sales data for Walmart stores across the United States.

Structure of the Dataset

* Store: Unique identifier for each Walmart store.
* Date: The week-ending date for the sales data.
* Weekly\_Sales: Total sales for the store in a specific week.
* Holiday\_Flag: Binary indicator (1 if the week includes a major holiday; 0 otherwise).
* Temperature: Average temperature (in Fahrenheit) for the week.
* Fuel\_Price: Weekly average fuel price (in dollars per gallon).
* CPI: Consumer Price Index, indicating changes in the cost of goods and services.
* Unemployment: Unemployment rate for the region.

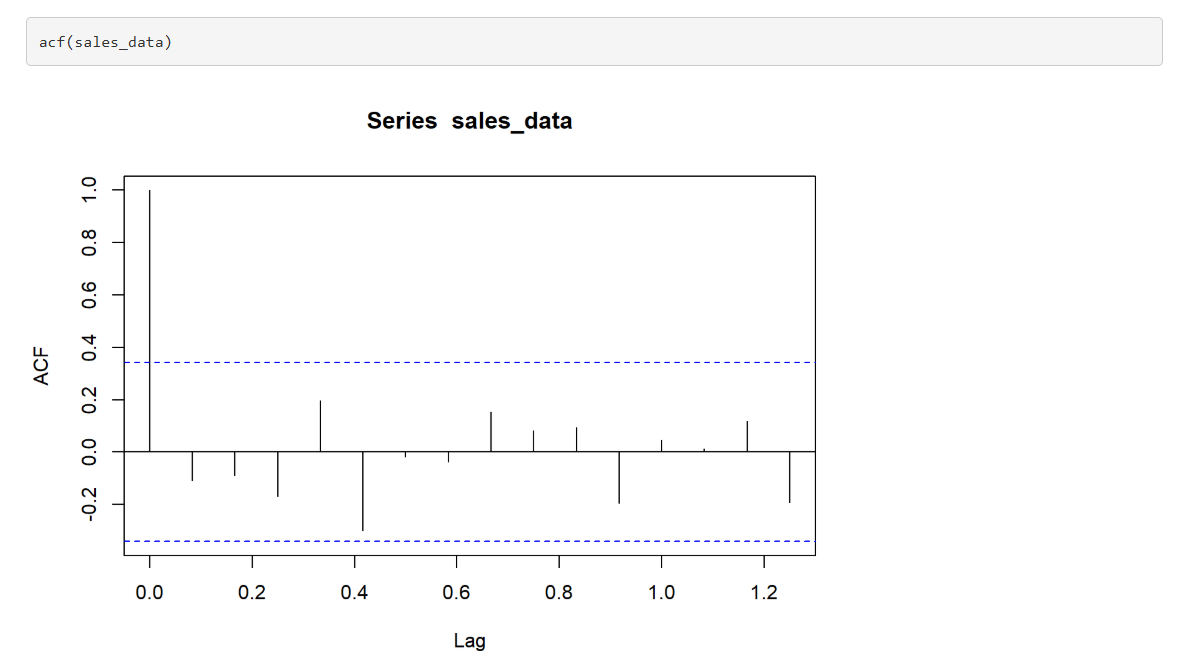
# Insights from Exploratory Data Analysis



**Overall Trend**: The graph shows a general upward trend in sales over time. This indicates that, despite the weekly fluctuations, there has been an overall increase in sales from 2010 to mid-2012.

**Seasonal Patterns**: There appear to be recurring peaks and troughs, suggesting seasonal patterns. These could correspond to specific times of the year when sales typically increase or decrease.

**Periodic Fluctuations**: Regular fluctuations within each year can be seen, likely representing cyclical variations in sales. These might be linked to monthly or quarterly business cycles.



From the Autocorrelation Function (ACF) plot for the time series dataset you provided, here are some key insights:

1. **Significant First Lag**:
   * The first lag has a significant positive autocorrelation, indicating that sales in one period are positively correlated with sales in the next period. This suggests a strong persistence in sales from one period to the next.
2. **Subsequent Lags**:
   * The subsequent lags show smaller or negative autocorrelations, and most of these values fall within the confidence intervals (dashed blue lines). This means that the correlation between sales data and its past values quickly diminishes after the first lag, indicating less dependence on further past values.

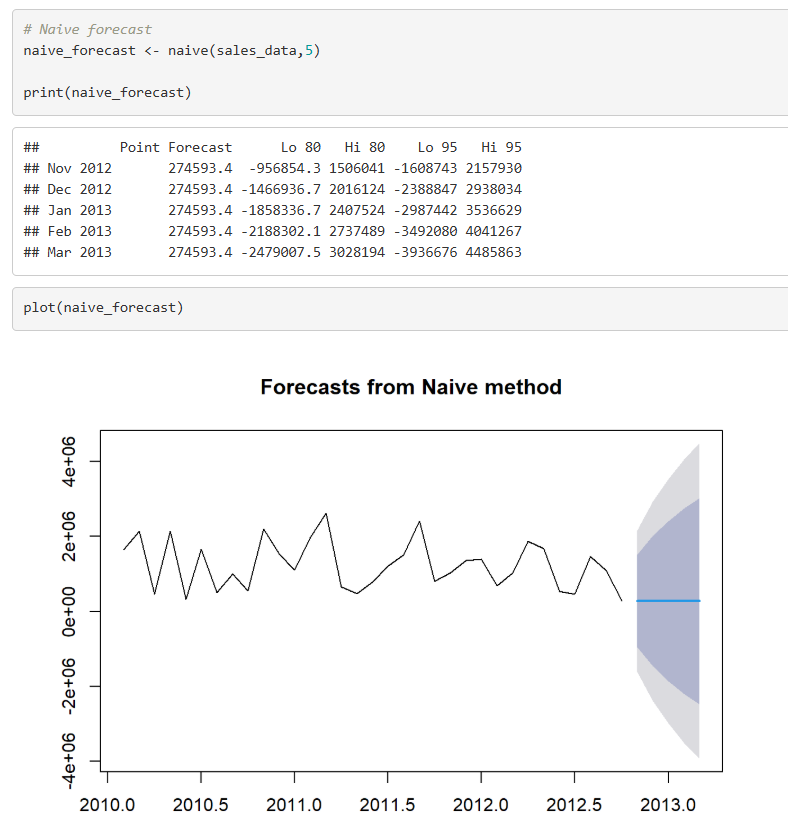
# State your Accuracy measure and its importance to you

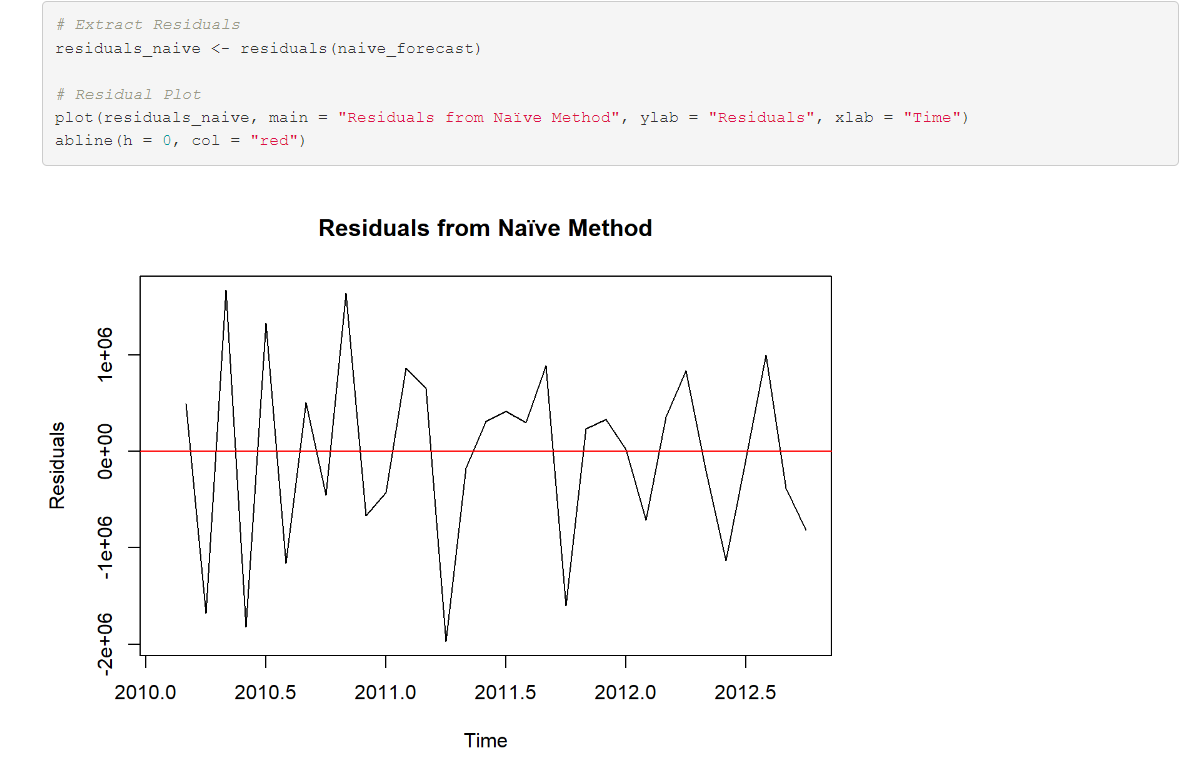
**Mean Absolute Percentage Error (MAPE)** is a common metric used to measure the accuracy of forecasting models. It calculates the average percentage error between the forecasted values and the actual values or average deviation of the predictions from the actual values. A lower MAPE value indicates higher accuracy, meaning the model's predictions are closer to the actual values. For example, a MAPE of 5% means that, on average, the predictions are 5% off from the actual values

# Insights from different forecasting methods and their residual analysis

When applying various forecasting methods to Walmart's sales data, each method provides unique strengths and insights. Residual analysis further evaluates their performance by examining the difference between observed and predicted values.

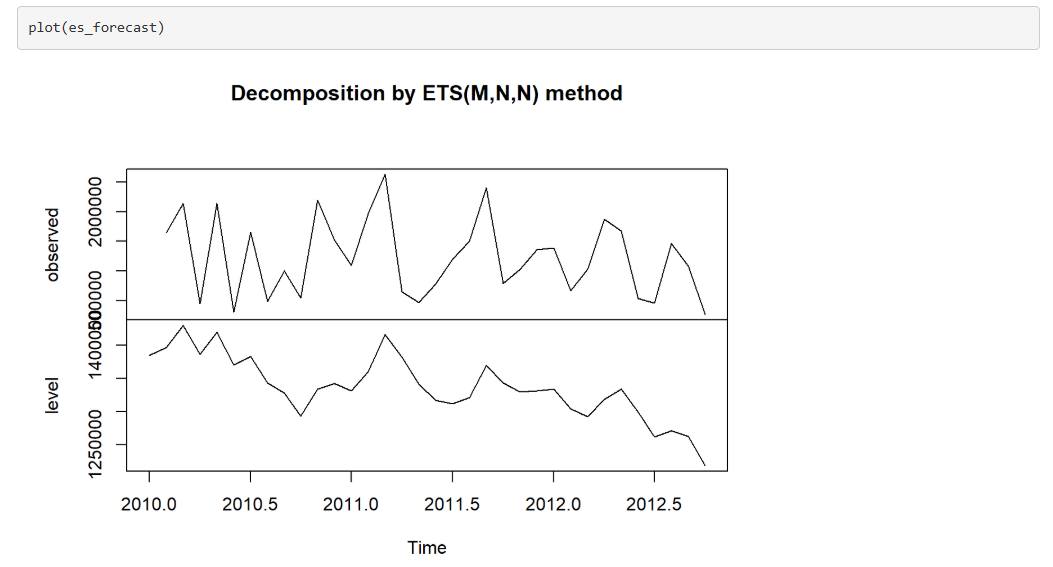
1. **Naive method**

 Sales data shows strong and predictable seasonal patterns.

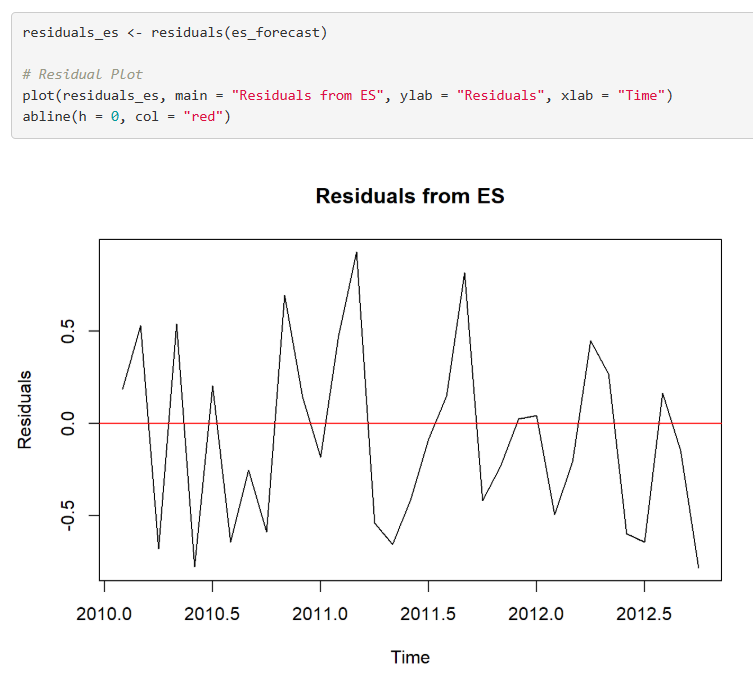
Sales data shows strong and predictable seasonal patterns.

The residuals fluctuate around the zero line, suggesting that the Naive method produces unbiased forecasts on average. This means there isn't a consistent overestimation or underestimation of the sales data.

1. **Exponential Smoothing (ETS) method**

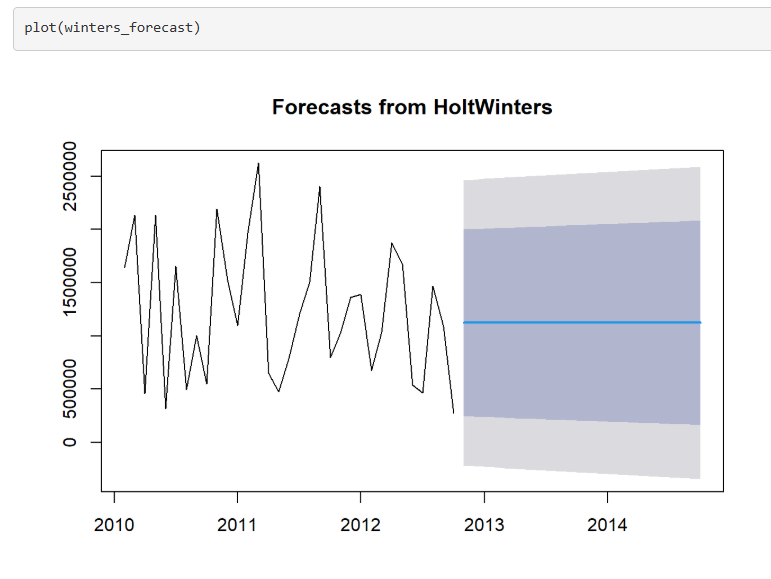


The level (or trend) component reveals an overall downward trend. This indicates that the underlying trend in sales data is decreasing over time, even though there are fluctuations.

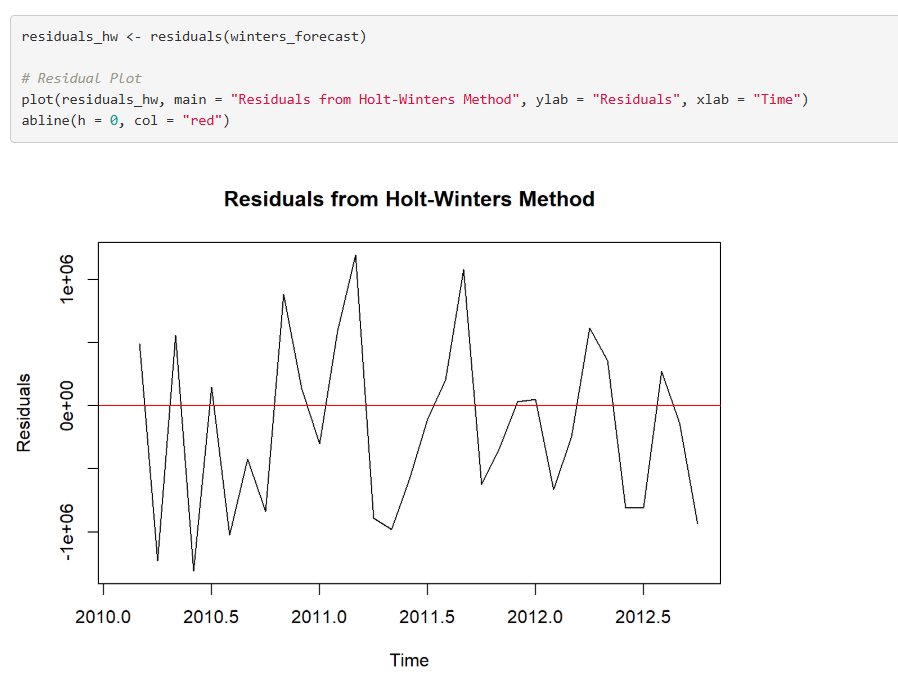


The residuals fluctuate around the zero line, indicating that the ES model produces forecasts that are unbiased on average. This means there's no systematic overestimation or underestimation in the forecasts.

1. **Holt Winters method**

****

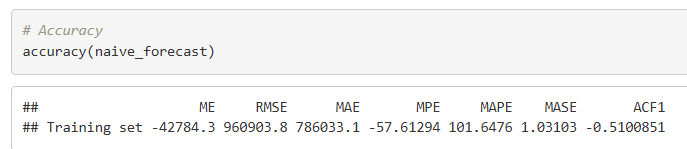
The plot displays a noticeable upward trend in the sales data over time, indicating that sales are generally increasing. This suggests a positive growth trend in your business.

****

The residuals fluctuate around the zero line, suggesting that the Holt-Winters method produces unbiased forecasts on average. This means there isn't a consistent overestimation or underestimation of the sales data.

# Prediction and Accuracy summary from different forecasting methods

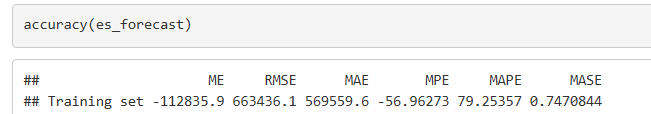
1. **Naive method**

The forecast assumes no change; the last observed value is repeated.

**Insights:**

1. **Underestimation Tendency**: Both ME and MPE are negative, indicating that the Naive model generally underestimates the actual sales.
2. **Error Magnitude**: High values of RMSE and MAE suggest that the forecast errors are significant. This indicates that while the Naive model might capture some trends, it may not be sufficiently accurate for precise forecasting.
3. **Percentage Errors**: A MAPE value over 100% indicates substantial errors in percentage terms, implying that the model's forecasts deviate significantly from actual values.
4. **Simple Exponential Smoothing (ETS) method**

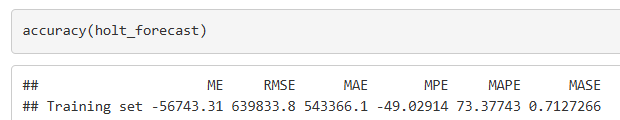
The level (or trend) component reveals an overall downward trend.



**Insights:**

1. **Underestimation Tendency**:
   * Both ME and MPE are negative, indicating that the model generally underestimates the actual sales. This could be due to systematic factors not captured by the model.
2. **Magnitude of Errors**:
   * High values of RMSE and MAE suggest that the forecast errors are significant. This indicates that while the es\_forecast model might capture some trends, it may not be sufficiently accurate for precise forecasting.
3. **Percentage Errors**:
   * A MAPE value of around 80% indicates substantial errors in percentage terms, implying that the model's forecasts deviate significantly from actual values.
4. **Holt Winters method**

The plot displays a noticeable upward trend in the sales data over time, indicating that sales are generally increasing. This suggests a positive growth trend in your business.



**Insights:**

1. **Underestimation Bias**:
   * Both ME and MPE are negative, indicating that the Holt-Winters model tends to underestimate actual sales. This bias needs addressing for more accurate forecasting.
2. **Magnitude of Errors**:
   * The high values of RMSE and MAE suggest significant errors in the model's predictions. Despite capturing some trends, the model may not be sufficiently accurate for precise forecasting.
3. **Relative Errors**:
   * With a MAPE of around 73.38%, the model's errors are substantial in percentage terms, indicating considerable deviation from actual values.

# State your decision based on the analysis

After evaluating the predictions and accuracy of the different forecasting methods, we can make an informed decision on the best model for forecasting Walmart sales. Here are the key takeaways from the analysis:

**1. Naïve Method**

* **Not suitable** for Walmart sales forecasting due to its lack of trend and seasonality consideration. It simply repeats the last value observed, which is highly inaccurate for data that exhibits trends and seasonality.
* **Decision**: Not recommended for this dataset.

**2.** **Simple Exponential Smoothing**

* **Only** accounts for trends and not seasonality. For data with consistent long-term growth or decline, this method can be useful but lacks the ability to adjust for seasonal peaks and valleys.
* **Decision**: A good short-term method, but not the best for capturing seasonality in Walmart sales.

**3. Holt-Winters Seasonal Method**

* **The best method** for this dataset. It accounts for both trends and seasonality, which are essential when forecasting Walmart sales. Retail sales tend to show strong seasonal patterns, especially around holidays or specific promotional periods. The Holt-Winters method’s ability to adjust for both trend and seasonality provides the most accurate and reliable forecasts.
* **Decision**: This method is the **recommended choice** for this dataset due to its flexibility and ability to capture both seasonal variations and trends.

# Provide some ideas to improve your forecasts

To improve the forecasts for Walmart sales, you can consider the following strategies :

**1. Decompose the Time Series Further**

Decompose the time series into its trend, seasonal, and residual components. By better understanding the individual components, you can apply different models to each component (e.g., applying Holt’s method for trend, Fourier series for seasonality, and traditional residual analysis).This helps isolate any noise or irregularities in the data, improving the forecast's overall accuracy.

* 1. **Use Higher Frequency Data (Daily or Hourly)**

If feasible, one can consider using **daily or even hourly** data for your forecasts, rather than weekly data. This can provide more granular insights into sales patterns, especially during high-traffic periods like weekends, holidays, or special sales events.