**Time Series Data Analysis**

1. Explain your model.

The dataset set shows the Retail sales of Beer, Wine, and Liquor Stores from January 2018 to July 2023. The sales are in millions of dollars over the period mentioned above.

The time series graph of this dataset shows the seasonality in sales every year. The sales values grow over the year, especially during the holiday season. The graph shows the spike highest during the year's end and second highest during the month mid-year i.e., May and July. The graph shows a decline in sales at the beginning of the year.

Looking at the time series data it can also be inferred that there has been a visible increasing trend.

Sales growth has been shown over the last 5 years. It has increased from 2018 to 2023.

**Models used in Forecasting:**

**1. Naive Methods**

○ **Naive Forecast:** The Naive forecasting method relies solely on the most recent data point to make predictions for future data points. It is a straightforward approach suitable for small datasets with minimal or no discernible trends or seasonality. However, in the context of the time series data provided in this assignment, the Naive method tends to produce an inaccurate result, as it may not capture complex patterns or underlying trends over a longer time.

**○ Random walk without a drift**: The Random Walk method is a simple forecasting technique that assumes that future values will be the same as the most recent observed value. In other words, it forecasts by repeating the last observed value for the specified number of periods. In the dataset, it is showing a similar result as the Naïve forecast which does not work for this dataset.

**○ Random walk with a drift:** This method considers the most recent data point

with a drift term and hence can provide good results for a short time

period. In this case, this is not capturing the seasonality but showing a trend over time.

**○ Seasonal Naive:** Seasonal data exhibits a repeating pattern or seasonality, typically occurring at regular intervals annually. For example, sales data increases during holidays. This captures the accurate result for the provided dataset with high seasonality.

**2. Average Methods**

○ **Mean Forecast:** Mean Forecasting predicts future values by taking the average (mean) of historical observations. In this case, the forecasted values for the next 6 periods will be calculated as the average of the historical values in the **Tsales** data. It calculates the mean of the observed values in the time series which is not relevant to predicting future values as this data set has seasonality and trend.

○ **Moving Averages:** This method adds weights to the data points with a

higher weight to the recent data points and can provide a better

prediction. This forecasts future values by averaging a specified number of the most recent observations (the order) to reduce noise and highlight underlying trends. In my case, it is showing an increase over time.

**3 Exponential smoothing**

**● ETS**: This method applies exponential smoothing to the time series data

and predicts the data points. It does not work for this dataset as it does not account for

trend or seasonality.

**● Holt-Winters:** Holt-Winters is a triple exponential smoothing method that considers three main components: level (the current value of the series), trend (the direction in which the series is moving), and seasonality (repeating patterns within the data). This is particularly useful in this dataset which deals with time series data that exhibit trend and seasonality.

**●SSE\_Simple:** Applies the Holt-Winters exponential smoothing method to this time series data with specific settings for the smoothing parameters.

Below are the parameters that have been used in SSE\_Simple:

**beta=TRUE, gamma=FALSE:**

beta = FALSE indicates that the model will not include a trend component. It implies that the data is assumed to have no linear trend.

gamma = TRUE indicates that the model will include a seasonality component, capturing seasonal patterns in the data.

**beta=TRUE, gamma=TRUE:**

beta=TRUE, it means that the model will consider a trend component in addition to the level and seasonality components. This is suitable when data exhibits a linear trend.

gamma = TRUE it means that the model will consider a seasonality component in addition to the level and trend components. This is suitable when data exhibits repeating patterns or seasonality.

**beta=FALSE, gamma=TRUE:**

beta=FALSE, This parameter setting means that the model will not include a trend component. In other words, the Holt-Winters model will assume that the data does not exhibit a linear trend over time. It will focus on capturing seasonality and the level of the data.

gamma=TRUE, this parameter setting means that the model will include a seasonality component. It indicates that the Holt-Winters model will consider the presence of repeating patterns or seasonality in the data.

**2. Pick an accuracy measure, compare your models, and state the best model**

**based on the accuracy comparison:**

1. Accuracy Measures Calculated:
2. Mean Error (ME)
3. Root Mean Squared Error (RMSE)
4. Mean Absolute Error (MAE)
5. Mean Percentage Error (MPE)
6. Mean Absolute Percentage Error (MAPE)
7. Mean Absolute Scaled Error (MASE)
8. Autocorrelation of errors at lag 1 (ACF1)

**Accuracy Measure Selected for Model Comparison:**

* Mean Absolute Percentage Error (MAPE):

The Mean Absolute Percentage Error (MAPE) quantifies the average forecast errors in terms of percentages. It is a valuable accuracy metric because it provides a percentage-based understanding of forecast accuracy that is easily comprehensible to a wide audience. For instance, a MAPE of 3% signifies a 3% disparity between the actual and forecasted data. Generally, a lower MAPE indicates a higher level of accuracy in forecasting.

**MAPE for different Models:**

● Mean Forecast: 13.37699

● Naive Forecast: 9.680379

● Random Walk Forecast: 9.680379

● Random Walk Forecast with Drift: 9.606246

● Simple Naive Forecast: 5.764731

● Exponential smoothing Forecast: 1.826156

● Holt-Winter Forecast: 2.181053

● SSE Simple Forecast: 11.0221

● SSE Simple Forecast1: 3.023347

● SSE Simple Forecast2: 2.363931

Looking at the MAPE for all the models the best models to use for forecasting are:

● Exponential Smoothing Forecast

● Holt-Winters Forecast

However, it's important to note that the Exponential Smoothing Forecast does not factor in trends or seasonality, which can lead to less accurate forecasts when dealing with a longer time horizon. Therefore, the most suitable model for this particular time series dataset is the Holt-Winters Forecast.