# Business Forecasting Assignment 4

Name: Parth Thaker

RUID: 227008583

**Understand and explain your model output.**  
Sunspots are temporary phenomena in the Sun's photosphere that appear as spots darker than the surrounding areas. Sunspots usually appear in pairs of opposite magnetic polarity. Their number varies according to the approximately 11-year solar cycle. This dataset comprises of the monthly mean total sunspots from 1/31/2000 to 1/31/2021.  
  
Explanation of various outputs:

1. Average Method: This method takes the average of the past observations. This forecast can decide the weights and past observations that go into the forecast. It is also known as the Mean Forecast Method. From the mean method, we can determine that with years on the X axis and Monthly Mean Total Sunspot number on the Y axis, we see a certain observation. From the years 2000-2002, we see a rise in the sunspot numbers observed. From 2003, we can observe a downward trend till 2010. From 2011, there is a significant rise till 2016 and then a gradual fall till 2020. The plot shows that this method is working for this dataset.
2. Naive Method: This is a forecasting method where the most recent data point in a dataset is referred to make predictions for the future data points. Also known as the “no change forecast” method. It assumes that there will be no change or growth in the variable and that historical data is a reliable predictor of the future. The Naive Method is incredibly simple to use and understand. It requires minimal computation and can be applied quickly. The Naive Method assumes that the future will be the same as the most recent observation. This assumption is often unrealistic, and in many cases, it leads to inaccurate forecasts. Does not look very accurate in the Dataset used in the assignment as the output relies solely on the value of the sunspot measured in 2021.
3. Random Walk Method: Random Walk Method is commonly used for short-term forecasting in situations where the underlying data is relatively stable, and where quick and simple forecasts are sufficient. It shows the results same as the above methods used hence it’s not very accurate for the current dataset.
4. Seasonal Naive Method: The Seasonal Naive Method is a variation of the Naive Method in business forecasting that considers seasonality in data. It's particularly useful when dealing with time series data that exhibit regular and predictable seasonal patterns. Like other forecasting methods, the Seasonal Naive Method relies on historical data. This data does not show a lot of seasonal shifts due to which the plot would not show any different result than the previous methods used which leads to an inaccurate output in the current dataset used.
5. Moving Averages Method: This method is used particularly when dealing with time series data. It is designed to smooth out fluctuations and noise in historical data to reveal underlying trends and patterns. The method involves calculating an average of past data points over a specified time, and this average is then used as a forecast for future periods. In the current dataset, this method shows an accurate output as it shows a downright trend of the mean monthly sunspot data from 2003 to 2010, then a gradual rise till 2015 and then a fall from 2015 to 2021. As there is no seasonality in the current dataset, it handles the data well.
6. Decomposition Method: It is a commonly used statistical technique in business forecasting, especially for time series data, where the objective is to break down a time series into its component parts to better understand and forecast future values. The method decomposes a time series into three main components: trend, seasonality, and random variation or error.   
   Trend Component: The trend component represents the long-term movement or direction in the time series data.  
   Seasonal Component: The seasonal component accounts for recurring patterns that occur at fixed intervals within the time series. These patterns are often related to calendar or seasonal factors and can include monthly, quarterly, or annual variations.  
   Error Component: It represents the residual variation in the time series data that cannot be explained by the trend or seasonality.  
     
   This is showing the most accurate results in this time series forecast of the dataset of Sunspots as it is able to capture complex patterns and giving a better understanding of the factors driving the variation in the data.
7. Holt Winters Filtering Method: It is a double exponential smoothing method since we must estimate the current level and the slope for the trend. It accounts for three parameters: level, trend and seasonality. For the current dataset, it is effective at capturing and forecasting time series data with trends and seasonality.   
     
   Level Component(α): The level component represents the estimated baseline or average value of the time series. It accounts for the overall level of the data.  
     
   Trend Component (β): The trend component accounts for the rate of change or trend in the data. It captures whether the data is increasing or decreasing over time.  
     
     
     
   Seasonality Component (γ): The seasonality component accounts for the regular, recurring patterns in the data, often associated with specific time periods such as days of the week, months, or seasons. Seasonality is updated using a weighted average of the most recent seasonally adjusted observation and the previous seasonality estimate. The weight is determined by the smoothing parameter γ.  
     
   This is the most accurate model used for forecasting for the current dataset.

**Pick an accuracy model: based on the accuracy comparison:**

Accuracy Measures Calculated:  
  
Mean Error (ME)  
Root Mean Squared Error (RMSE)  
Mean Absolute Error (MAE)  
Mean Percentage Error (MPE)  
Mean Absolute Percentage Error (MAPE)  
Mean Absolute Scaled Error (MASE)  
Autocorrelation of errors at lag 1 (ACF1)

Accuracy Measure Selected for Model Comparison:

Root Mean Squared Error (RMSE): RMSE is the square root of the MSE and is often used to express error in the same units as the data. Generally, a lower RMSE indicates a higher level of accuracy in forecasting. RMSE is the lowest for the Exponential Smoothing Forecast which is 18.71306 and it shows the lowest error rate. However, it is important to note that the Exponential Smoothing Forecast can lead to less accurate forecasts when dealing with a longer time frame data of 20 years as this of Mean Sunspot Data. Therefore, the most suitable model for this time series dataset is the Holt-Winters forecast.