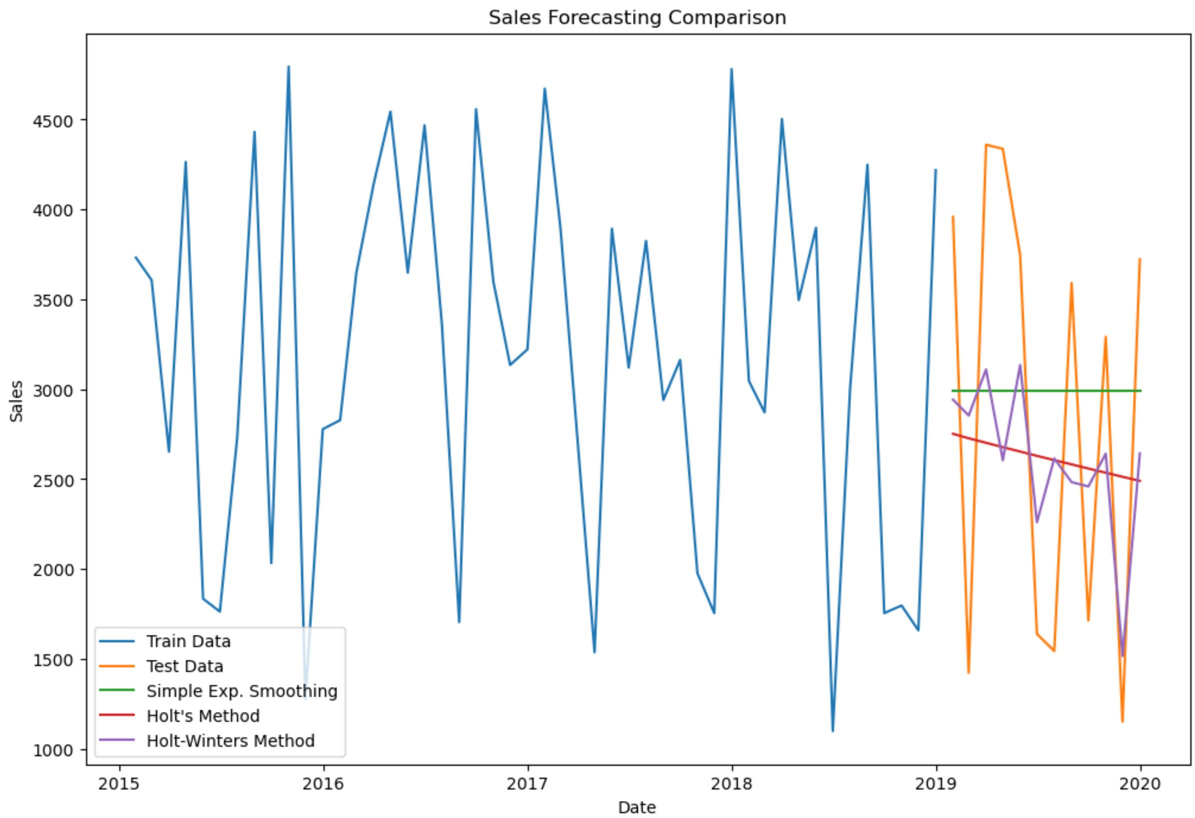
# Exploring Time Series Forecasting: A Comprehensive Study with Practical Datasets

Benjamin Nicholson
Seton Hill University



Introduction:
Forecasting is the process of using existing data to make predictions about future values. The process of forecasting is essential for any data scientist as the techniques that are going to be discussed can be applicable to any field. The techniques are as follows:

- Simple Exponential Smoothing (SES)
- Holt's Linear Method
- Holt Winter's Seasonal Method

Each of these techniques are used to smooth out the changes in the data so that you can extrapolate into the future. There is an incremental process in these methods, which is why Holt's Linear Method is something called $2^{nd}$ Exponential Smoothing and Holt Winter's Seasonal Method is called $3^{rd}$ Exponential Smoothing. The complexity of these methods increase.

Simple Exponential Smoothing
Simple Exponential Smooth (SES) is a very basic technique that can be used for some basic forecasting. SES assumes that data is stationary, meaning that there is no association between the time and the values in the data. This assumes that there is no change in trend levels nor change in standard deviation. It weights the most recent results with a greater importance. It creates a horizontal linear that looks to generalise the data that has taken place. If you are going to use SES it is important to check for stationarity which can be done by using the Augmented Dickey Fuller Test. The limitation of SES is that it does not incorporate any trend or seasonality in its forecasting.

The following is the equation for Simple Exponential Smoothing in time series

$$Y_t = B_0 + E_t$$

Holt's Linear Method
Holt's Linear Method's main component is that it includes the linear trend that takes place in the data. Holt's Linear Method can capture the trend by decomposing the data. The model consists of two components, the level, and the trend. One of the great advantages of this model is that it has a dampening variable that can be utilised to mitigate the extent of the growth that takes place in the forecast. It is unlikely that growth with continue indefinitely which is where the dampening variable can help increase the accuracy of the model. With the linear trend included this is a forecast that is similar to a simple linear regression which takes a line through the existing data to predict future outcomes.

The following are the equations for the components of Holt's Linear Method

$$Overall\ Equation : \hat{y}_{t+h} = l_t + hb_t$$

$$Level\ Equation : l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1})$$

$$Trend\ Equation : b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$$

https://towardsdatascience.com/forecasting-with-holts-linear-trend-exponential-smoothing-af2aa4590c18

Holt Winter’s Seasonal Method

Holt Winter’s Seasonal Method expands on SES and Holt’s Linear Method by including the seasonality component of the data. This final progression allows for the level, linear trend and seasonality to be a part of the forecast’s predictions. This is going to result in much more specific predictions. There are two components to this method, the additive and multiplicative method. The additive method is used for when the seasonal variation remains constant, meaning that with time there is no change in the magnitude of oscillations in seasonality. Whereas the multiplicative method observes that the seasonality has changes in its magnitude. These models are usually much more specific and will include changes compared to the overall growth or decay trend as seen in Holt’s Linear Method.

The following are the equations for the additive and multiplicative component of Holt Winter’s Seasonal Method.

**Additive Seasonality**

$$\boxed{F_{t+k} = L_t + (k * T_t) + S_{t+k-M}}$$

**Multiplicative Seasonality**

$$\boxed{F_{t+k} = [L_t + (k * T_t)] * S_{t+k-M}}$$

https://medium.com/analytics-vidhya/a-thorough-introduction-to-holt-winters-forecasting-c21810b8c0e6

**Problem 0** Decompose the four Models

To decompose models it is first very important to understand certain properties of the graph. These include whether they contain the following properties

- Linearity
- Seasonality
- Cyclical

Linearity represents an average change in values over time

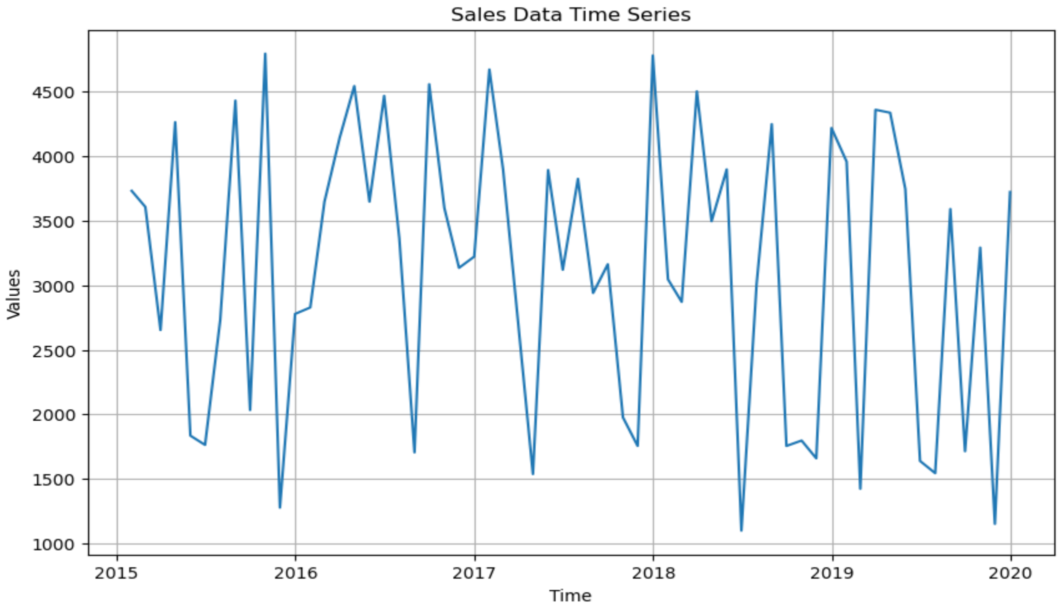
Seasonality displays the repetitive patterns that take place within the data over time

Cyclical is the change in patterns in data that often take place on an irregular basis

Another important measure to consider is the Augmented Dickey Fuller Test (ADF test) which is a means of measuring the stationarity of the data. Understanding if data is stationary or not is important in understand the type of forecasting that can performed on the data.

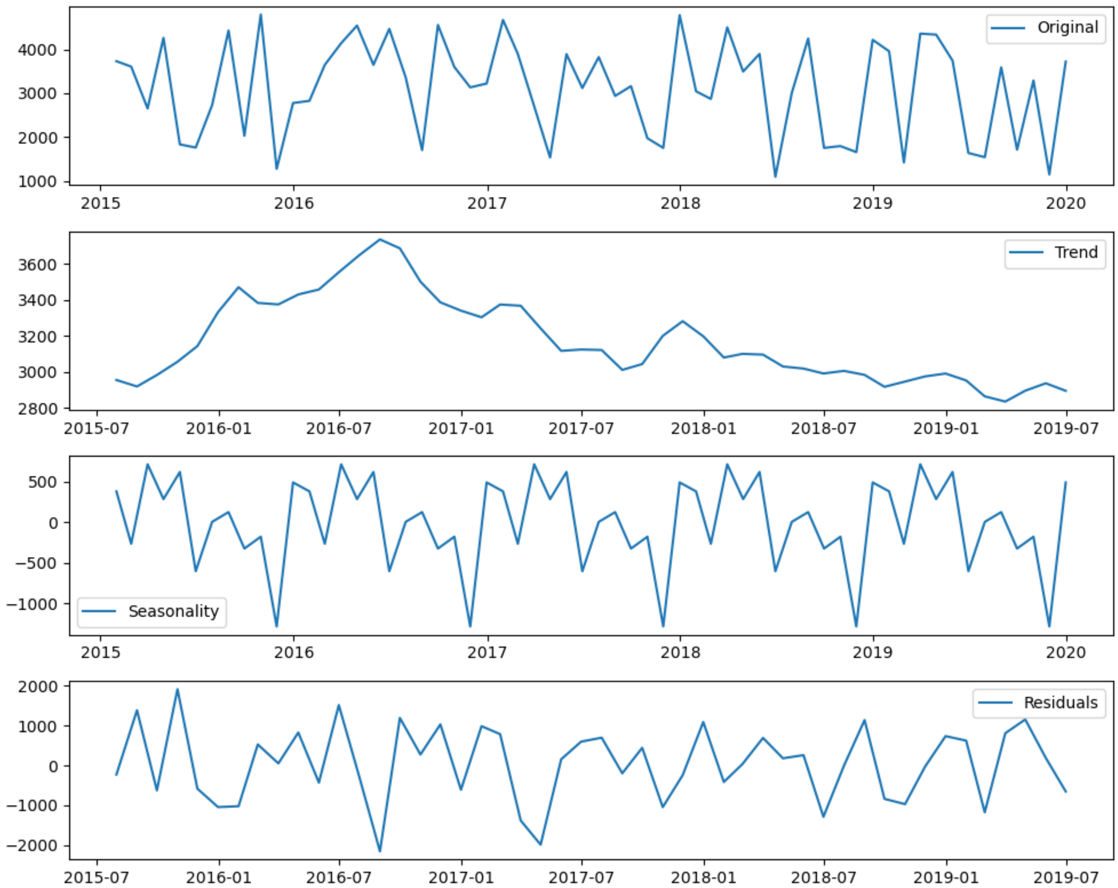
The ADF test will return a value, this is the ADF statistic. Based on the particular value you can give a statement on whether the data is stationary or not with a certain confidence interval. These confidence intervals are at 90%, 95% and 99%.

*Graph 1: Sales Data Time Series*



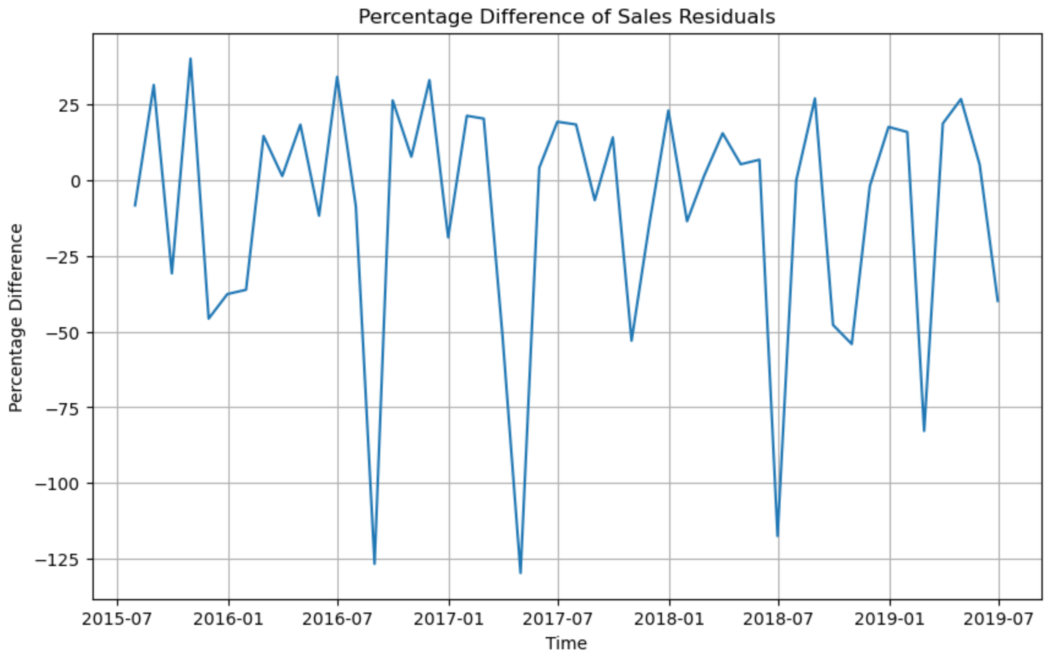

The data does not show any immediate signs of linear trend, seasonality or cyclical changes. Further seasonal decomposition, ACF plots and ADF Tests can be used to further prove this.

*Graph 2: Decomposition of Sales Data*

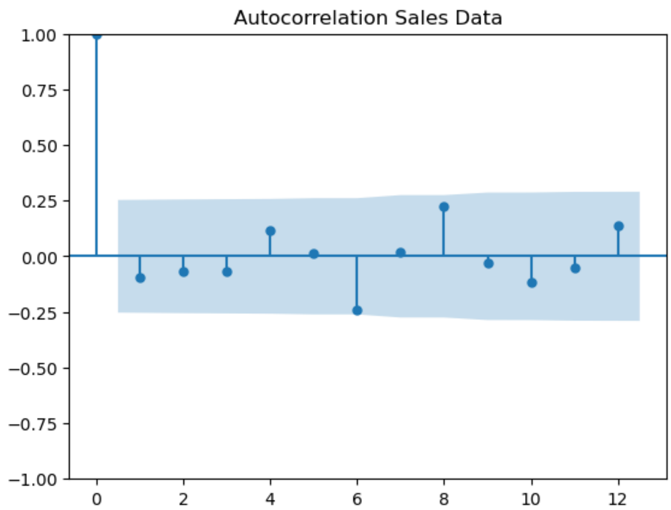


When you first observe the data, you might assume that there is some seasonality that is described in the 3rd graph as the residuals seem to oscillate above and below a mean of 0. However, if you look further you can see that these residual oscillations are taking place between 500 and 2000 sales above and belove the mean. Which I am now going to represent the residuals as a percentage of the original data to get a better understanding of how far off the decomposition model is.

*Graph 3: Percentage Difference of Sales Data Residuals*




A part from the 3 large negative oscillations the residuals fell mainly between -10% and 25% which is not that accurate of a model. Next it is time to look at the ACF plot to see if there is any correlation and finally the ADF test. If all of these come back representing what I assume to be results that represent no seasonality then this data is likely to be stationary. The trend that is exhibited is only 9% increase from the mean of 3000 while the residuals have much greater changes than that which is likely to represent a graph that is stationary and does not exhibit any of the linear trend, cyclical changes or seasonality that is being tested for.

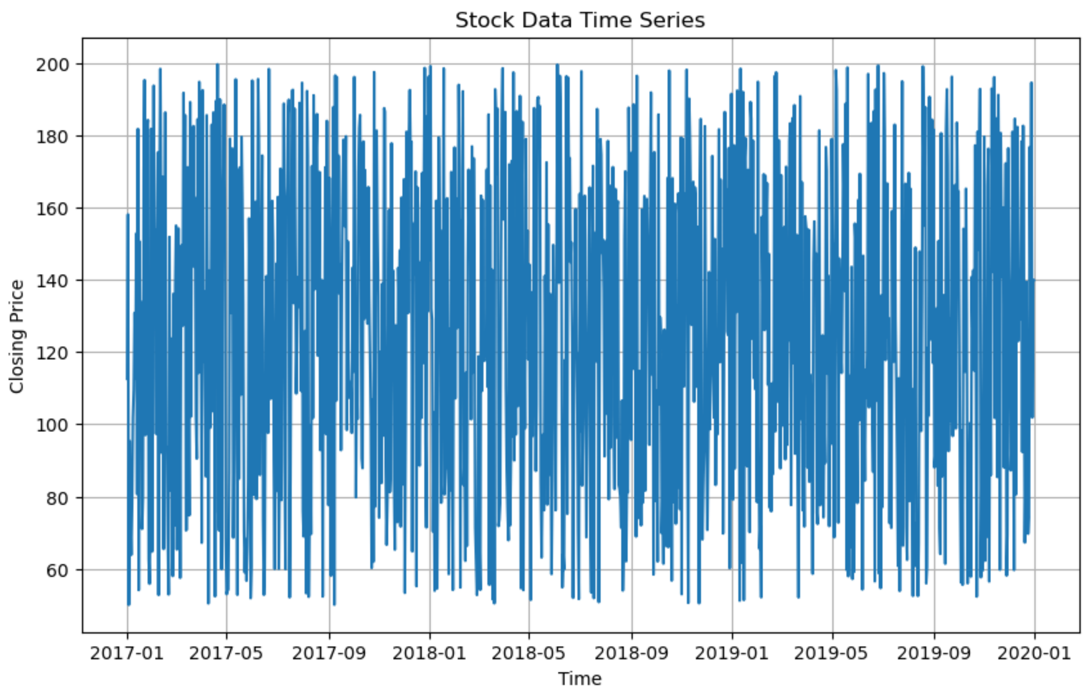



*Graph 4: Autocorrelation of Sales Data*

```
ADF Statistic:  -8.325005277374798
P-Value:  3.489721856712595e-13
Critical Value:
        1%: -3.55
        5%: -2.91
        10%: -2.59
```

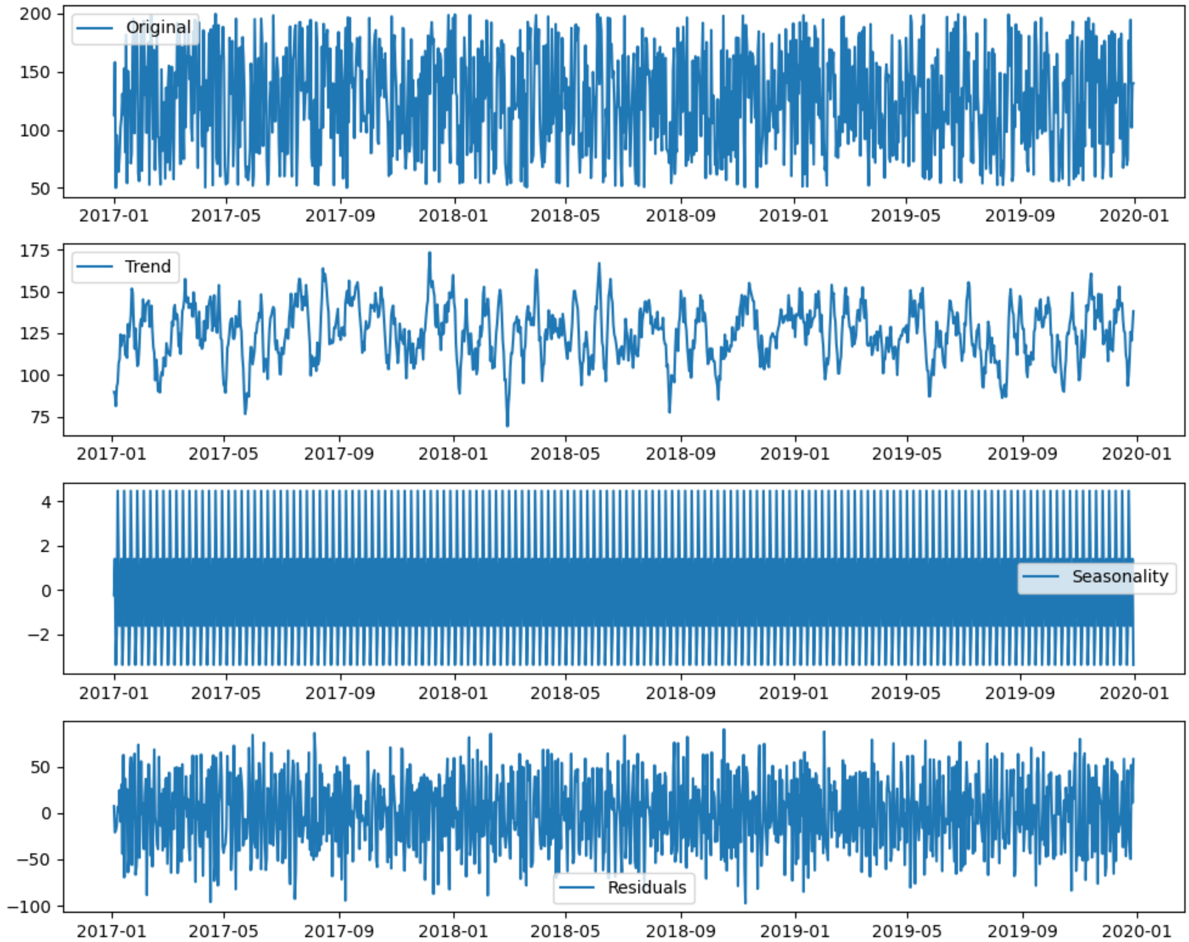
With all of the tests that have been done. There seems to be no seasonality, which is proven by the auto correlation and the large residual values that were given in the decomposition of the data. The linear trend has a small peak of about 9% but as the residuals extend far greater than this it is fair to assume that there is no linear movement in the data, further proven by the ADF test returning a statistic far higher than what is required to have 99% confidence that the graph is stationary. There is no cyclical change either depicted by the visualisation of the graph and the linear trend.

*Graph 5: Stock Data Time Series*




Similar to the above graph, the data does not show any immediate signs of linear trend, cyclical changes or seasonality. However the main difference is the noise that is displayed, you can see how this seems to be similar to white noise graphs. However, a seasonal decomposition is going to be used to gain a better insight into underlying patterns.

*Graph 6: Seasonal Decomposition of Stocks Data*

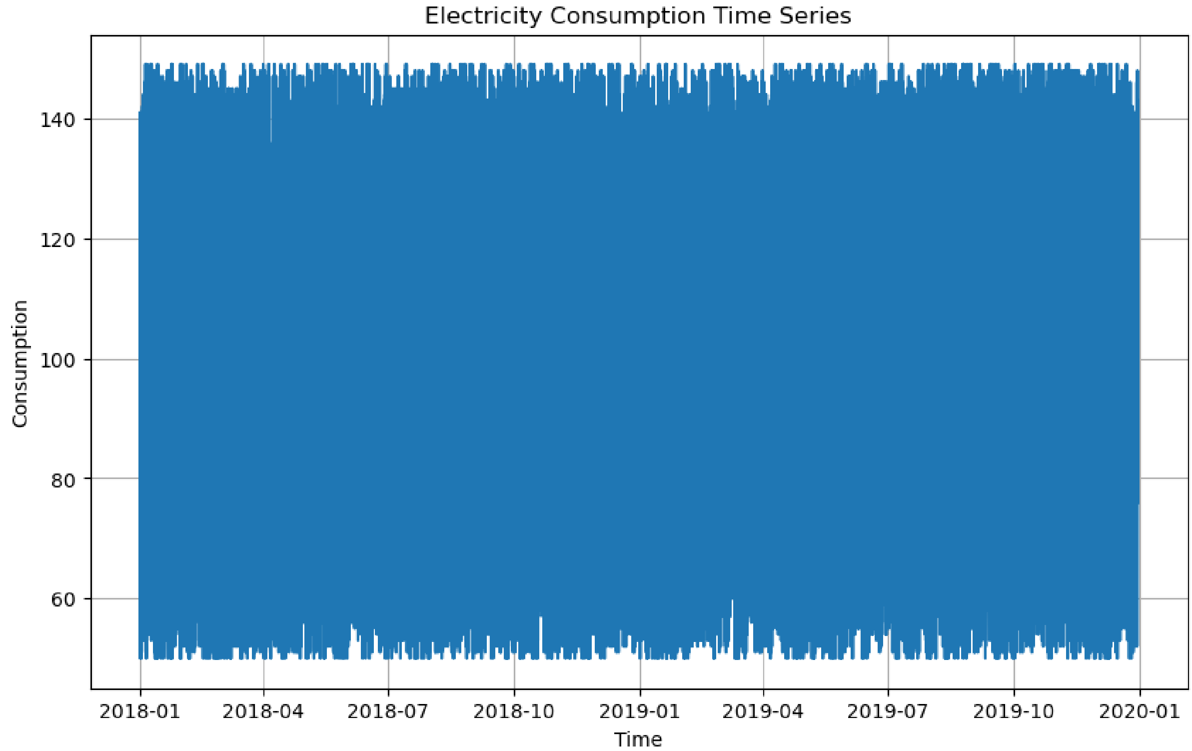



The Stock Data is even more comparable to white noise. You can see that the mean and standard deviation essentially remains the same as time continues. This graph indicates no seasonality, cyclical nature or linearity. Hence why the ADF statistic is so high and there is a 99% confidence in the assumption that the data is stationary.

```
ADF Statistic:  -22.340369622532915
P-Value:  0.0
Critical Value:
        1%: -3.44
        5%: -2.86
        10%: -2.57
```

It can also be assumed that there is no trend, cyclical changes or seasonality based off of the seasonal decomposition.

*Graph 7: Electricity Data Time Series*

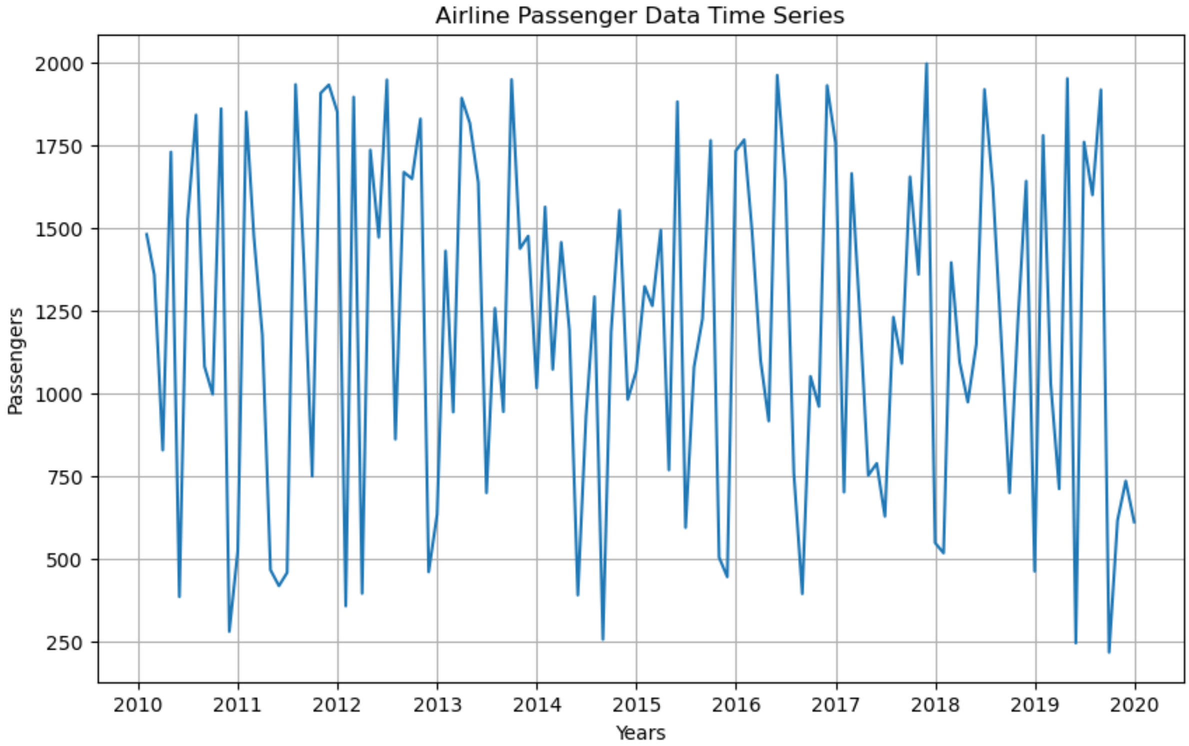



There is so much noise and error in this data you can not see the movement. This is not going to display any cyclical changes, seasonality or linear trends. This can be proved by the extent of the ADF test below.

The Electricity Consumption Time Series has so much noise that it is difficult to even understand what the data points are. This is also because of the density of the data points in the time interval. The ADF statistic is 100 times greater than the threshold for 99% confidence that the graph is stationary. This graph has no linearity, seasonality or cyclical nature.

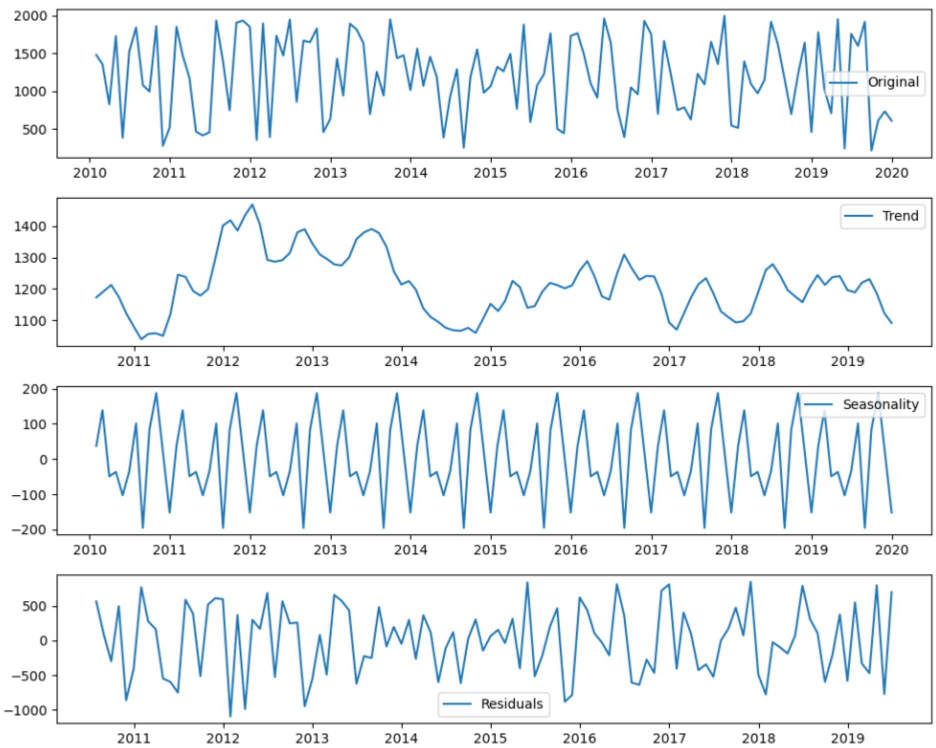
```
ADF Statistic:  -131.64004315876778
P-Value:  0.0
Critical Value:
        1%: -3.43
        5%: -2.86
        10%: -2.57
```

*Graph 8: Airline Passenger Data Time Series*




This data is similar to the first graph where the data does not show any immediate signs of linear trend, seasonality or cyclical changes. Further seasonal decomposition, ACF plots and ADF Tests can be used to further prove this.

*Graph 9: Airline Data Decomposition*



Similar to the first data set, the residuals show such a large change compared to the change in trend, it can be assumed that the seasonality has been artificially created and does not accurately depict any seasonality in the data.

The Airline Passenger data is similar to the Sales Data time series. The data around 2014 and 2015 has a smaller variation than data in the beginning and end intervals. However there is no overall change in data which means that there is no linearity, the graph also does not represent any seasonality or cyclical nature. The ADF statistic is above the threshold for a 99% confidence interval that the data is stationary.

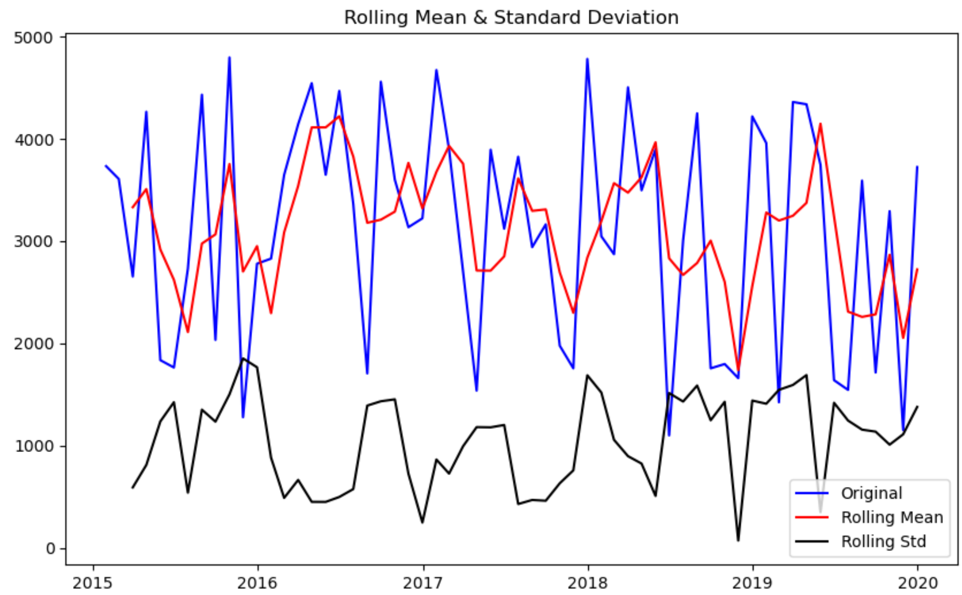
```
ADF Statistic:  -7.600002845561867
P-Value:  2.403322262296728e-11
Critical Value:
        1%: -3.49
        5%: -2.89
        10%: -2.58
```

**Problem 1: Basics of Moving Averages**

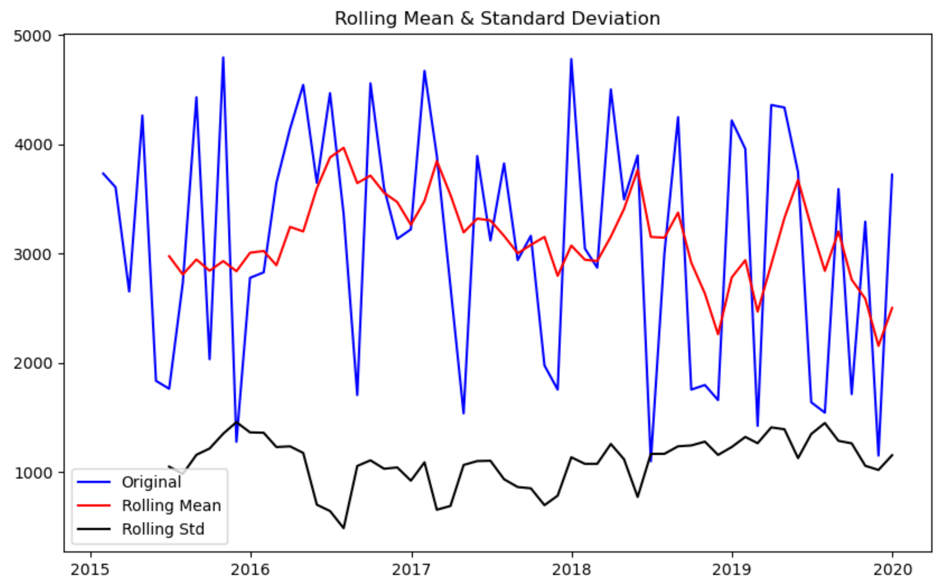
1. Use Dataset A to calculate and plot simple moving averages with different window sizes.
2. Discuss the impact of window size on the smoothness of the trend.

Moving averages is the technique that is used to smooth out data and the variation that takes place. By mitigating the magnitude of oscillations, you are able to pick up on overall trends better rather than data that may be skewed by outliers. The following problem is going to incorporate different windows for data. As the window number increases the smoothness of the graph also increases. This is because there is a much higher magnitude that is required to overcome the trends that occur in the data.

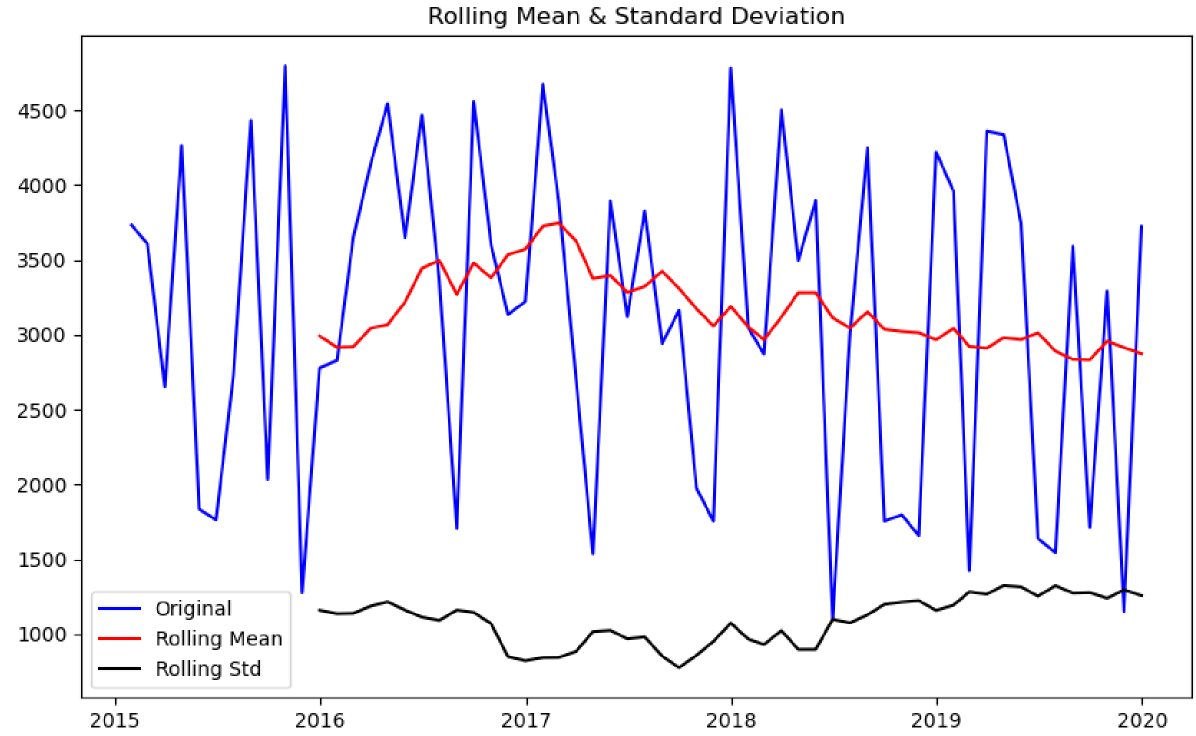
*Graph 10: Quarterly Rolling Mean & Standard Deviation (window=3)*



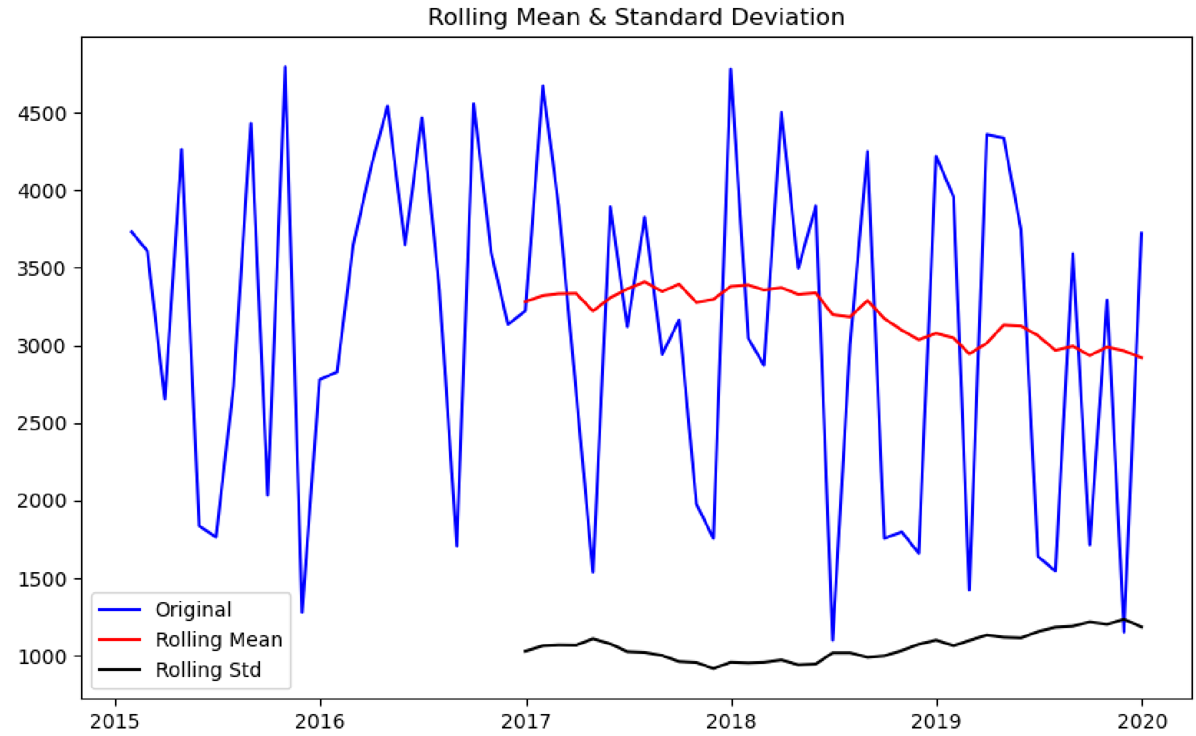

*Graph 11: Half-Yearly Rolling Mean & Standard Deviation (window = 6)*



*Graph 12: Yearly Rolling Mean & Standard Deviation (window = 12)*




*Graph 13: Biyearly Rolling Mean & Standard Deviation (window = 24)*



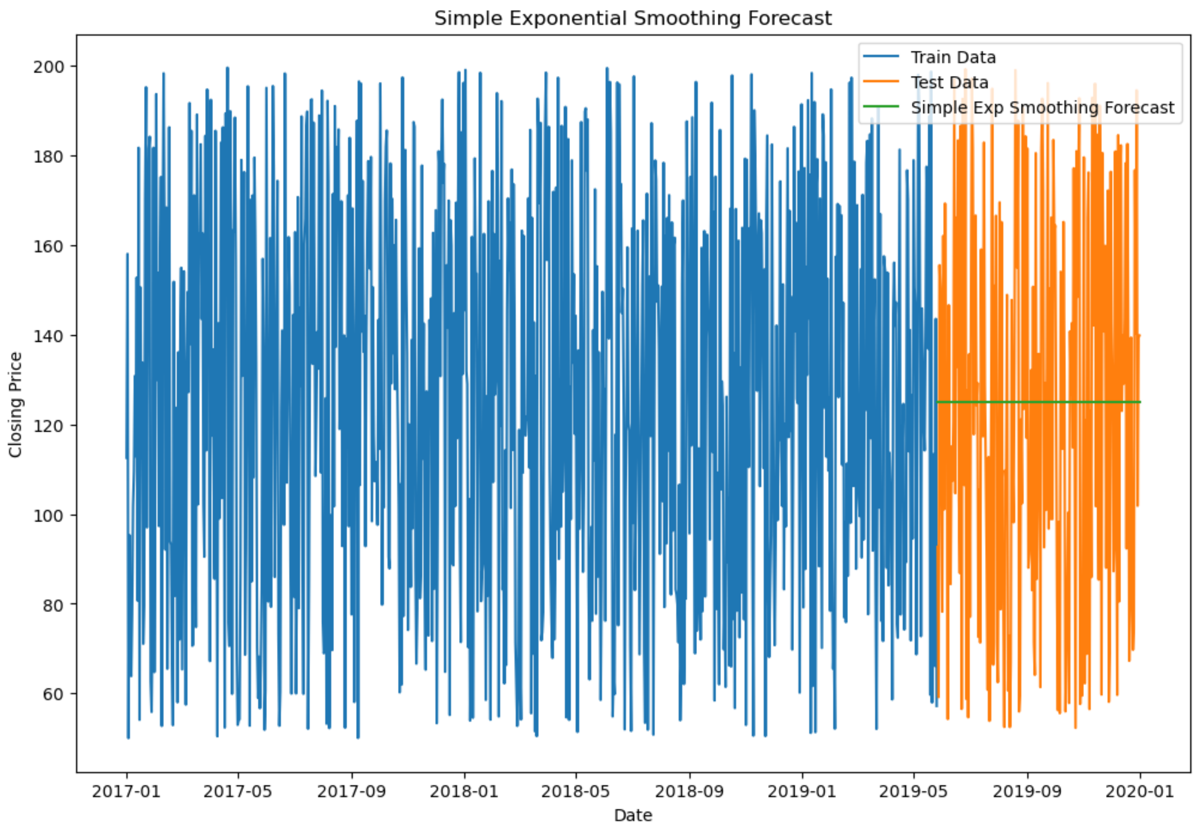

The first point of observation is that when you increase the window, the start of the rolling mean is pushed back. This is because it takes more data to get that first point. You can clearly see that as you increase the window value there is less oscillations and an overall trend of the data takes form. Graph 5 and Graph 6 with a window of 3 and 6 respectively have a much greater increase in variability of the mean. If you want the data to show the overall trend then increase the window. However if you want to just smooth out some of the outliers using a smaller window such as a quarterly or half-yearly window number is going to be more effective.

**Problem 2: Introduction to Exponential Smoothing**

1. Apply single exponential smoothing on Dataset B and compare it with moving averages.

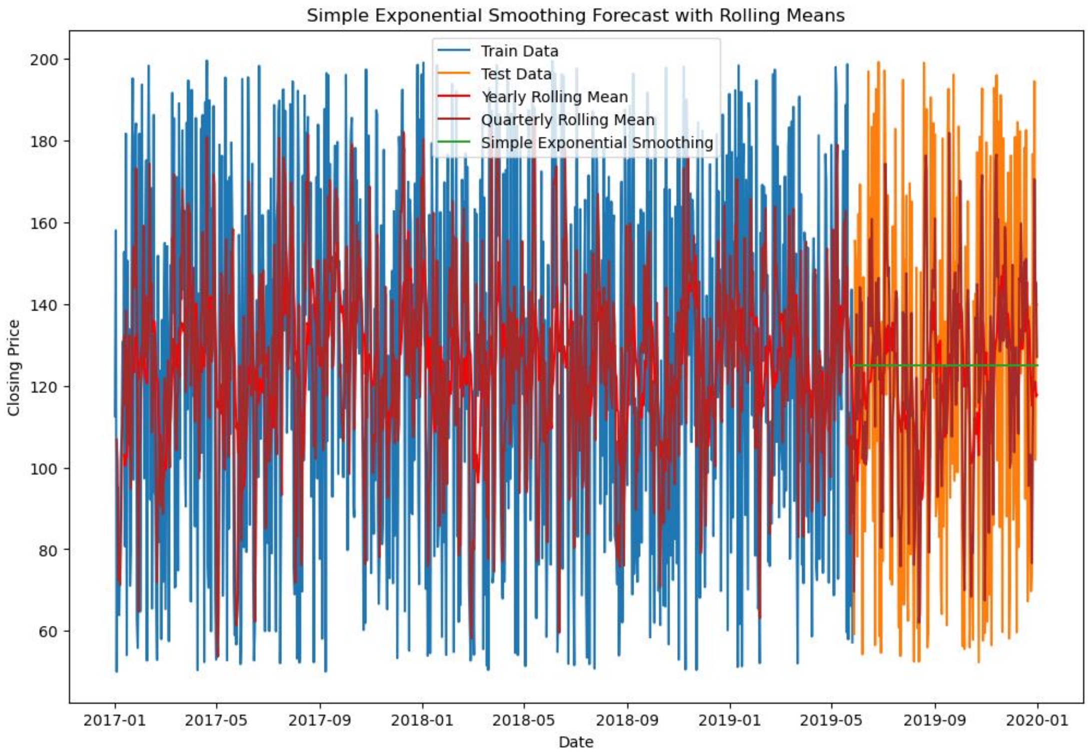
For SES I am going to be splitting the data into 80% training and 20% testing data. This split can allow to compare how the SES did with the original data. Dataset B is the stock data time series which was stationary. So it would be assumed that the SES would take the average that took place and put a horizontal line through that, with the most recent data having more of an impact.

*Graph 14: Simple Exponential Smoothing Forecast for Stock Sales Data*



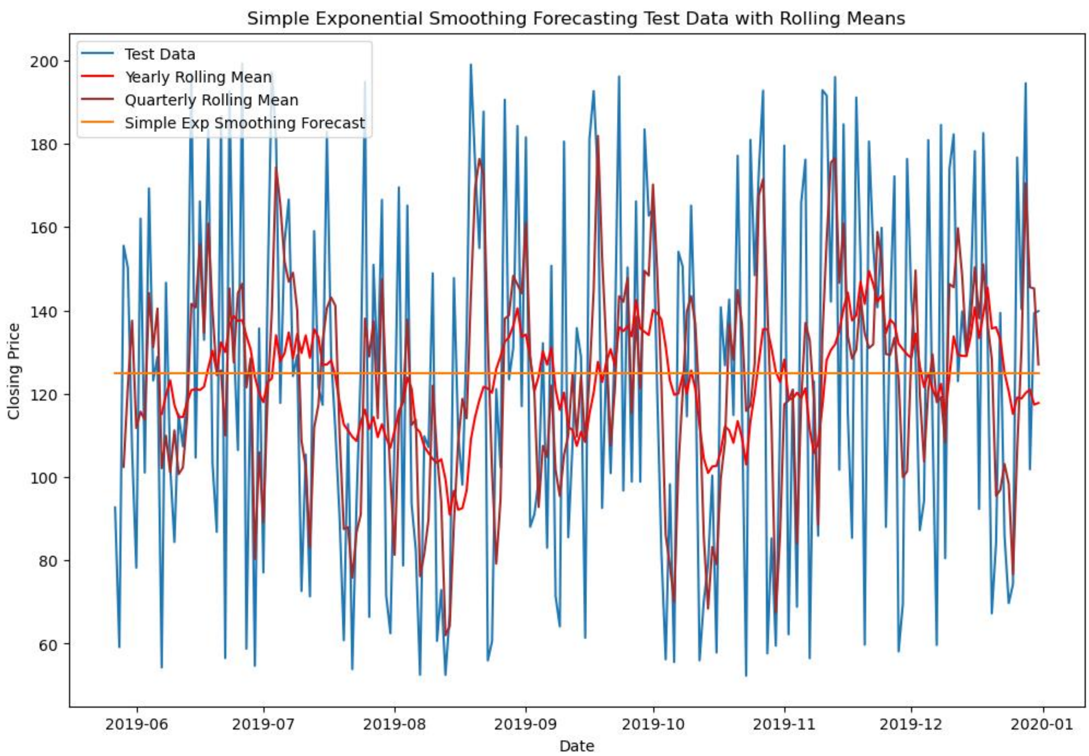

To better compared how the SES compared with differing rolling means I am going to include all of that on the one graph to compared how they differ.

*Graph 15: Simple Exponential Smoothing Forecast with Rolling Means*




It can be difficult to observe what is occurring so I am going to be looking at the test data with the same width so that the concentration of different graphs does not take away from the understanding of the behavior.

*Graph 16: Simple Exponential Smoothing Forecasting Test Data with Rolling Means*



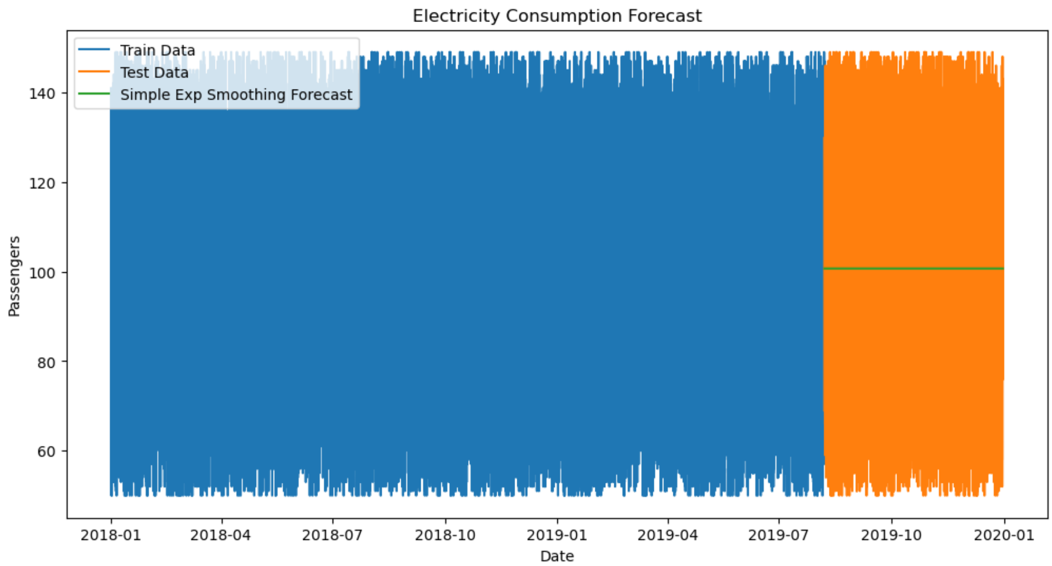
The SES data takes a horizontal line going through the data. Meanwhile the rolling means take on different behaviour. The quarterly rolling mean follows the similar oscillations that ake place in the original data however the magnitude is decreases. The yearly rolling mean has much better depiction of the overall trend in the data. That is that there are periods above and below the mean but ultimately this graph is stationary.

**Problem 3: Understanding Simple Exponential Smoothing (SES)**

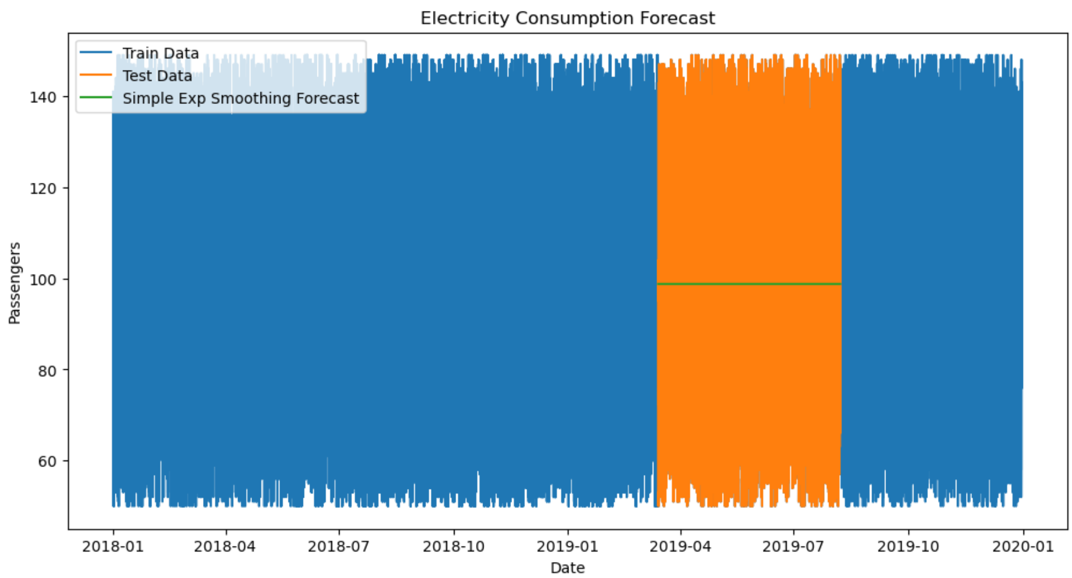
1. Implement SES on Dataset C and analyse its responsiveness to changes in data.

The SES value is going to change based on which data it is working on. I have included test data that is split 80:20 with test data. As well as the split data with intervals from 0:60, 80:20 and test intervals from 60:80. Comparing the two different SES outcomes would usually be beneficial to observe how different data will result in different smoothed trends. However, because of the extreme noise that takes place in the Electricity Consumption dataset it is going to be harder to observe.

*Graph 17: Electricity Consumption SES Forecast 1st split*




*Graph 18: Electricity Consumption SES Forecast 2nd split*

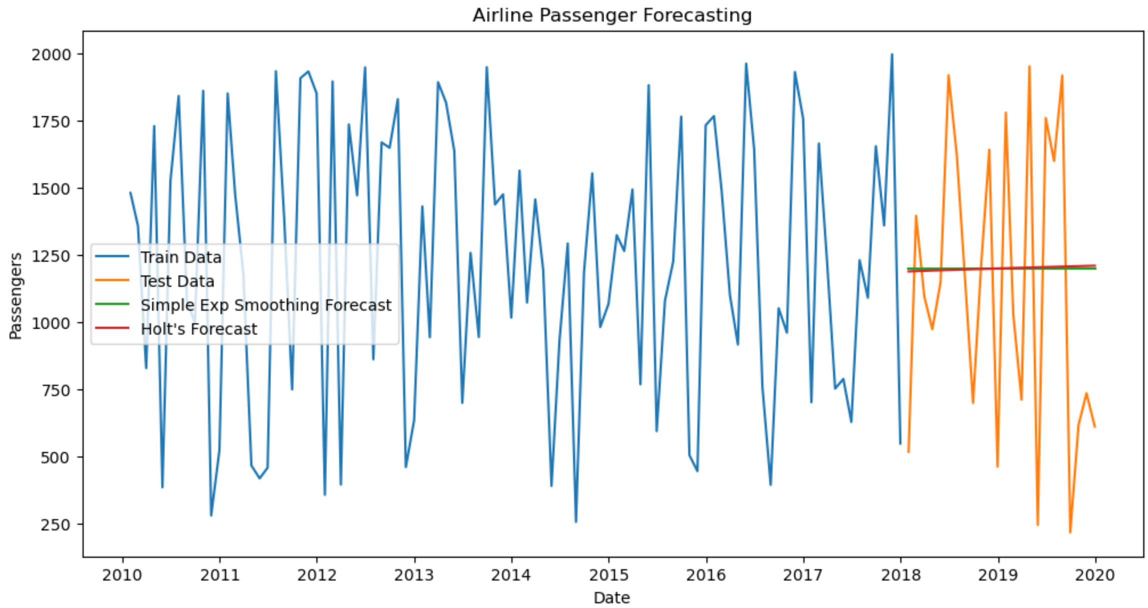


**Problem 4: Exploring Holt's Linear Trend Model**

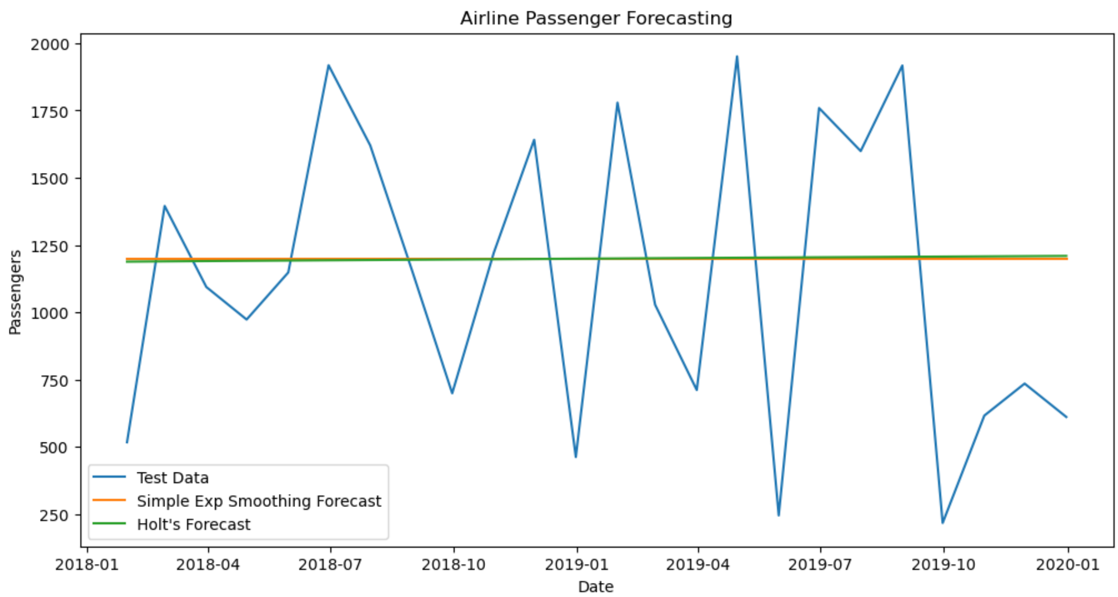
1. Apply Holt's Linear Trend Model to Dataset D and discuss its ability to capture both level and trend.
2. Explanation: Holt's model extends exponential smoothing by incorporating a trend component, allowing for more accurate forecasting when data exhibits a trend over time.

I am going to split the data 80:20 for the training and test dataset. This can allow for a comparison to what the real data represented and how accurate it was in forecasting.

*Graph 19: Airline Passenger Holt's Linear Trend Forecasting 1st Split*

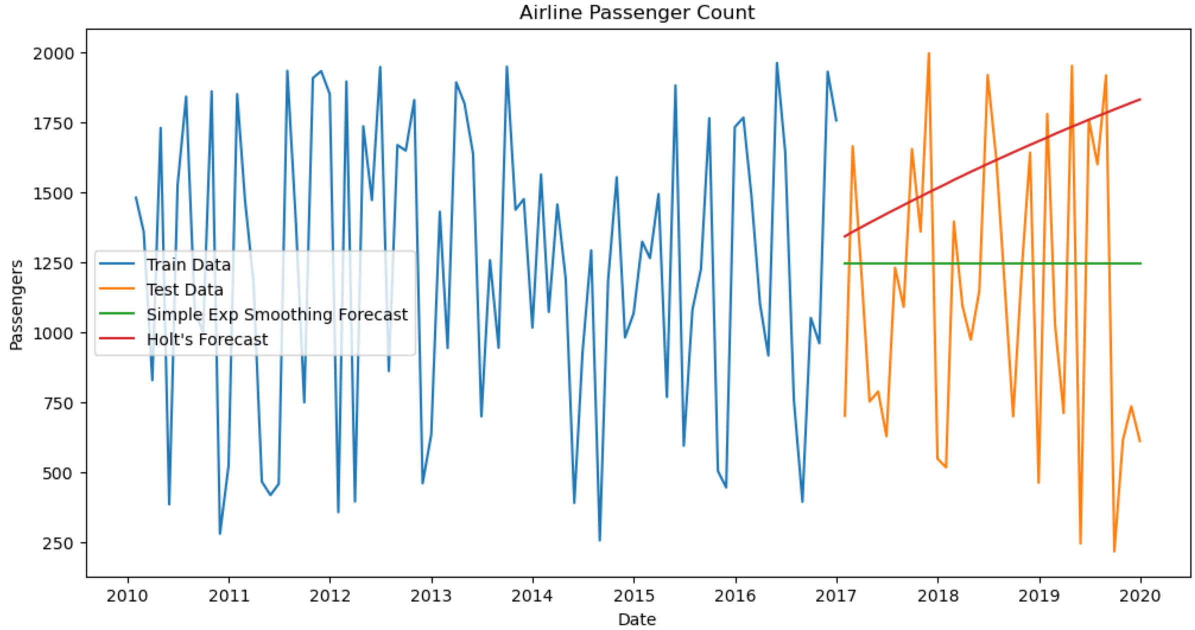



*Graph 20: Airline Passenger Holt's Linear Trend Forecasting 1st Split Zoomed in*

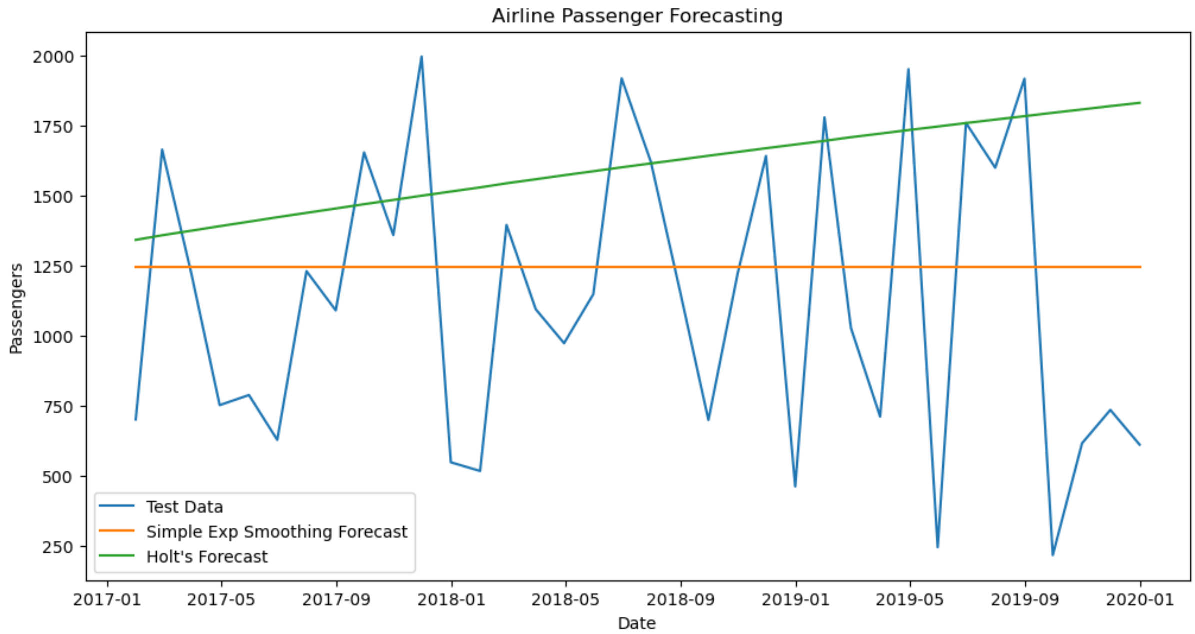



In this data split you can see how there is almost no different between SES forecast and Holt's forecast. This is a situation where you must understand the data to determine the best forecast. To see Holt's Method capture trend I am going to split the data 70:30. Also note that the model is additive and not multiplicative due to the variation in oscillations.

*Graph 21: Airline Passenger Holt's Linear Trend Forecasting 2nd Split*




*Graph 22: Airline Passenger Holt's Linear Trend Forecasting 2nd Split Zoomed in*



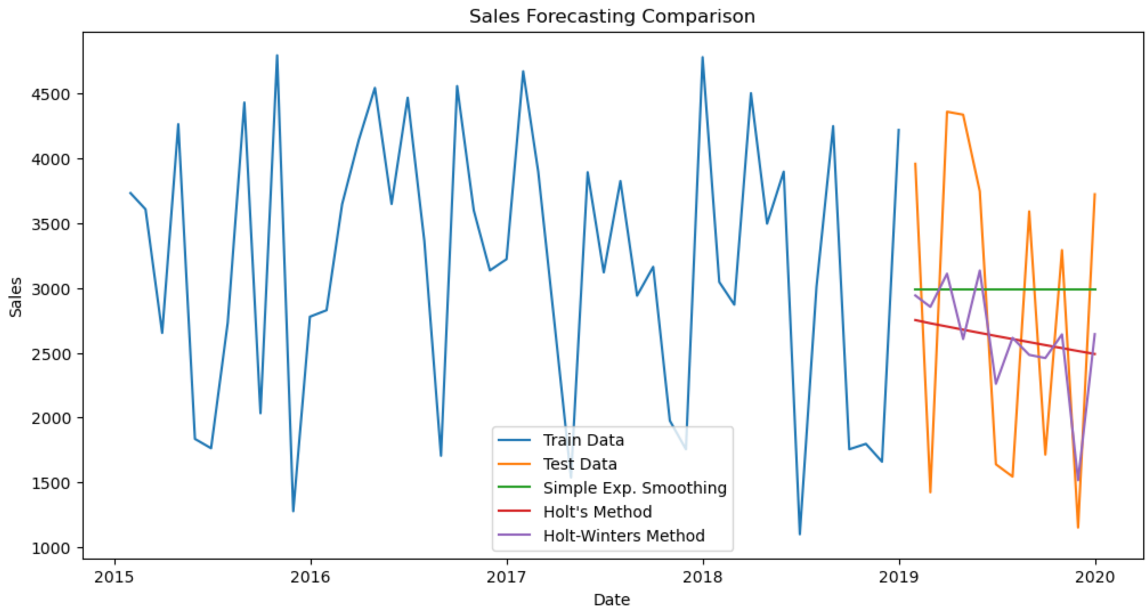

This data split represents a new story. While there is a much greater understanding of the linear trend it does not seem to accurately represent the trend in the data. Once again understanding that the data is stationary is an important factor into why the SES forecast seems to represent the movement of the data more effectively.

**Problem 5: Diving into Holt Winter's Seasonal Model**

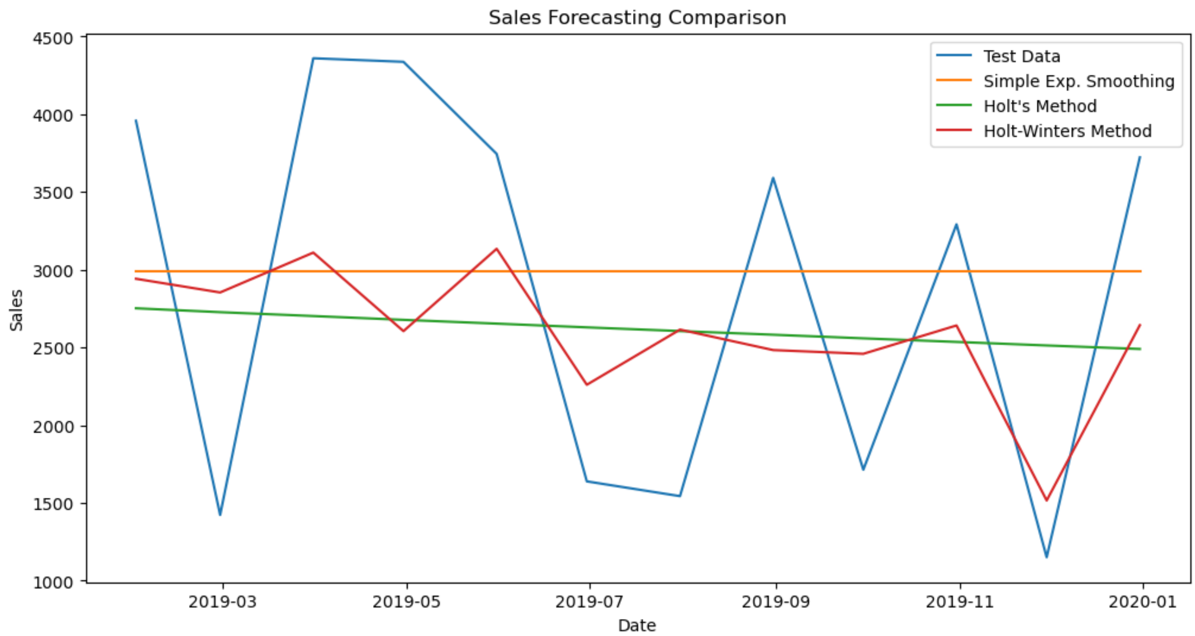
1. Use Dataset A to apply the Holt-Winters Seasonal Model, focusing on capturing seasonality in addition to level and trend.
2. Description: The Holt-Winters seasonal model includes three components: level, trend, and seasonality, offering a comprehensive approach to forecasting time series data that exhibits seasonal patterns.

Using dataset A which was the Sales Data is going to allow for a better understanding of how each forecasting model works. Continuing to split the data 80:20 between training and test, you are able to see how each forecast compares with the real data values.

*Graph 23: Sales Forecasting Comparison*




*Graph 24: Sales Forecasting Comparison Zoomed*



Zooming into the test data you can see how the Holt Winters Model does a good job of reflecting the future values. This would suggest seasonality, linearity and level components that need to be incorporated into the forecasting equation. However you can also see how the SES plays a role in generalising the data and finding a pretty accurate representation of the noise that goes above and below and viewing it as stationary. The linear trend model indicates a small decrement in average values.

**Problem 6: Application and Selection Criteria**

1. For each dataset, select the most appropriate forecasting model based on specific criteria such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Akaike Information Criterion (AIC).
2. Justify the selection of models for each dataset.

Mean Absolute Error (MAE) is the metric that is use to evaluate the accuracy of a forecasting model using the average absolute difference between the predicted and the actual values. This is in the same units as the original data so it should be easier to interpret. It is also less sensitive to outliers as it takes the mean of the difference.

$$mae = \frac{\sum_{i=1}^{n} abs\left(y_i - \lambda(x_i)\right)}{n}$$

https://medium.com/@20__80__/mean-absolute-error-mae-sample-calculation-6eed6743838a

Mean Squared Error (MSE) is another metric for evaluating the accuracy of a forecast. This is very commonly used and is differentiable everywhere which makes it easier to run in algorithms compared to MAE. However outliers are exuberated because of the square to those values.

$$\mathrm{MSE} = \overset{\text{Mean}}{\boxed{\frac{1}{n}\sum_{i=1}^{n}}}\Big(\overset{\text{Error}}{\boxed{Y_i - \hat{Y}_i}}\Big)^{\overset{\text{Squared}}{\boxed{2}}}$$

https://suboptimal.wiki/explanation/mse/

Akaike Information Criterion (AIC) is a statistical measure used for model selection. It can compare the goodness of fit of different models through penalising the complexity of them. This is likely going to see the SES results be more effective.

The data is going to be split 80:20 for training and test data.

The following suggestions for the forecasting based on the model evaluation techniques listed is based on the function that is defined in the .ipynb file. It finds the best forecasting model which minimises the error found in the graph.

Dataset A – Airline Data

| | Mean Absolute Error | Mean Squared Error | Akaike Information Criterion |
|---|---|---|---|
| Holt Winter's Seasonal Method | 506.662322 | 343509.634596 | 1226.064418 |
| Holt's Linear Method | 474.870138 | 300756.985495 | 1222.101347 |
| SES | 474.152955 | 299302.758008 | 1206.446991 |

Recommended Forecast Model: SES

Dataset B – Electricity Consumption Data

| | Mean Absolute Error | Mean Squared Error | Akaike Information Criterion |
|---|---|---|---|
| Holt Winter's Seasonal Method | 25.191066 | 849.543588 | 94352.093004 |
| Holt's Linear Method | 340.212538 | 154330.115639 | 95312.641336 |
| SES | 25.111212 | 839.305793 | 94071.497677 |

Recommended Forecast Model: SES

Dataset C – Sales Data

| | Mean Absolute Error | Mean Squared Error | Akaike Information Criterion |
|---|---|---|---|
| Holt Winter's Seasonal Method | 866.344961 | 9.977373e+05 | 690.610761 |
| Holt's Linear Method | 1187.301208 | 1.488388e+06 | 683.620635 |
| SES | 1129.763744 | 1.463894e+06 | 675.037348 |

Recommended Forecast Model: SES/Holt’s Winter Seasonal Method

Dataset D – Stocks Data

| | Mean Absolute Error | Mean Squared Error | Akaike Information Criterion |
|---|---|---|---|
| Holt Winter's Seasonal Method | 37.654897 | 1922.688153 | 6680.152921 |
| Holt's Linear Method | 430.894801 | 242911.616260 | 6764.283787 |
| SES | 37.035566 | 1876.550416 | 6619.021218 |

Recommended Forecast Model: SES

**Problem 7: Reflection on Models & Real-World Applications**

From the selection of models based on error evaluations techniques listed it is clear that SES was the best forecasting for these models. This is likely due to the stationarity of all of the graphs. As each dataset suggested a stationary model then the SES model would perform best as it would not try to fit a trend that was not there nor look for seasonality which was in fact just noise. These datasets are examples of where using a simple base forecast such as SES actually outperforms the more complex model predictions. This is a phenomenon called overfitting where you do not generalise your forecasts.

**Real World Applications of Each Model**

SES is utilised for forecasting data which has a constant trend, no seasonality and no major changes in fluctuations in data. This is quite limited in its real world applications but it can be used in website traffic and estimating demand for basic household goods. These are datasets that are likely to remain constant throughout long periods of time which is why SES can represent the continued constant data values.

Holt's linear Model has expanded real world capabilities because it is common for growth and decay to happen with data. Generalising this growth is a very important factor of forecasting. A real world example could be looking at the growth of a stock over long periods of times or looking at a successful start up company that has constant growth. Situations where data is likely to have a linear increase and not exponential is going to be the most useful situation for Holt's Linear Model.

Holt Winter's Seasonal Model has the most real world applications and that is because of its increased complexity of incorporating seasonality as well as trend in its forecast. The following are a few examples of different industries which utilise this model.

- Sales data for companies that have differing success based on certain parameters such as weather or economic cycles
- Agriculture: estimating crop yields based on the seasonal weather patterns
- Transportation: estimating the amount of people using public transportation based on certain holidays and celebrations

**Holt Winter's Seasonal Model over SES**

Although the data that was used in this investigation did not represent a situation where Holt Winter's Seasonal Model would be preferred to SES, there are many examples of this case. When the data exhibits a clear trend as well as noise as SES cannot capture the overall movement of data. Situations where the forecasting extends beyond the range of existing data because it is unlikely for data to continue completely stagnant. As well as looking at data that experiences repetitive patterns. Overall it is going to be more likely than not that Holt Winter's Seasonal Model is going to be more effective when looking at real world data. However SES does offer a great basis for understanding what stationary data would look like as time continues.