**Time Series: Visualization**

**Time Plot:**

* Most basic and informative plot of visualizing a Time Series is a Time Plot.
* Time Plot is Line chart of series values over time.
* Time values are on horizontal axis and values are on vertical axis.

**Zooming In:**

* Looking at a shorter period of time in the series instead of the entire series.
* Especially, important when the time series is long as it can reveal patterns that are hidden when viewing the entire series.

**Adding Trendline:**

* By trying different trendline we can find what types of trend (e.g: linear, exponential, cubic) best approximates the data.

**Suppressing Seasonality:**

* It is easier to see the trend in the data when seasonality is suppressed.
* Suppressing seasonal patterns can be done by plotting the series at a cruder time scale.

**Lag Scatter Plot:**

* Scatter plots are usually used to explore the relationship between two variables.
* It is often the case in time series that the next values are based on the previous observations.
* Previous observations in a time series are called lags.
* Lag scatter plot helps us get a hint whether there is a relationship between current values and future values or not.

**Time Series: Feature Engineering**

A time series dataset must be transformed to be modeled as a supervised learning problem.

**Date Time Features -** Components of time step for each observation.

**Lag Features -** Values at prior time steps.

**Window Features -** Summary of values over a fixed window of prior time steps.

**Time Series: Resampling**

**Resampling:** Changing frequency of available data to match the frequency of the required forecast.

**Types:**

**Up sampling:** Increase the frequency of the samples such as from months to dates.

**Down sampling:** Decrease the frequency of the samples such as from months to years.

In both the cases, data must be invented.

In case of up sampling, interpolation technique is used to calculate fine-grained observation.

In case of down sampling, summary statistics is used to calculate new aggregate values.

Resampling is also done for the purpose of feature engineering.

**Time Series: Transformations**

**Power Transformation**

* A lot of the forecasting models do not work well when there is a non-linear trend.
* To better identify the shape of the trend, it is useful to change the scale of the series.

**Moving Average**

* MA Smoothing is creating a new series where values are averages of the raw observations.

**Simple Exponential Smoothing**

* Weighted Average -The importance of the latest values will be more than the older values.

*Moving Average Smoothing and Exponential Smoothing can only be used when there is no trend or seasonality in the serie.*

**Time Series: Important Concepts**

**White Noise:**

* It is a sequence of random numbers.
* If a series is white noise, it cannot be predicted.

We will try to see if our series is white noise or not.

* If it is white noise, it means it cannot be predicted and we will stop the process there only.
* If it is not white noise, we will do the predictions, find out the errors in the predictions (difference between actual and predicted values) and then we will check if the error values are white noise or not.
* If these error values are not white noise then this means that there is still information that could have been modeled, but the current model was not able to take out that information. In such case we will use more advanced forecasting models.
* Whereas, if the error values are white noise, it means all the information has been harnessed and all that is left is random fluctuations.

**In Summary:**

* Series should not be white noise.
* Error values should be white noise.

**Random Walk:**

* Next value of the series is the modification of the previous value in the sequence.
* In case of Random Walk, we use previous value as the forecast for the next value, Naïve Forecasting give the best result.
* Naïve Forecasting accuracy serves as the benchmark for other models as well.
* If the accuracy of the advanced model is better than the accuracy of the Naïve Forecast, then the model is finding out more information and we can say that the underlying series is not a random walk.
* But, if the accuracy of the advanced model is less than that of Naïve forecast, then probably the underlying series is the random walk.

**Differencing:**

* It is a simple method for removing trend and seasonality.