Day 8

Agenda:

1. Time Series Analysis
2. Checks:
    1. Seasonality
    2. Trend
    3. Cyclical
    4. Irregularity
3. Stationarity
4. How to convert Non Stationary to stationary
5. Code
6. KNN
7. Churn Prediction
8. MCQs

Time Series Analysis

-- A Time-Series represents a series of time-based orders. It would be Years, Months, Weeks, Days, Horus, Minutes, and Seconds.

-- The time variable/feature is the independent variable and supports the target variable to predict the results.

-- Time Series Analysis (TSA) is used in different fields for time-based predictions – like Weather Forecasting, Financial, Signal processing, Engineering domain – Control Systems, Communications Systems.

-- Assumptions: There is one and the only assumption that is "stationary", which means that the origin of time, does not affect the properties of the process under the statistical factor.

Components of Time Series Analysis :

Trend:

In which there is no fixed interval and any divergence within the given dataset is a continuous timeline. The trend would be Negative or Positive or Null Trend

Seasonality :

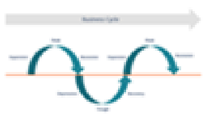
In which regular or fixed interval shifts within the dataset in a continuous timeline. Would be bell curve or saw tooth

Cyclical :

In which there is no fixed interval, uncertainty in movement and its pattern

Irregularity :

Unexpected situations/events/scenarios and spikes in a short time span.

	Trend	Seasonality	Cyclical	Irregularity
Time	Fixed Time Interval	Fixed Time Interval	Not Fixed Time Interval	Not Fixed Time Interval
Duration	Long and Short Term	Short Term	Long and Short Term	Regular/Irregular
Visualization				
Nature - I	Gradual	Swings between Up or Down	Repeating Up and Down	Errored or High Fluctuation
Nature – II	Upward/Down Trend	Pattern repeatable	No fixed period	Short and Not repeatable
Prediction Capability	Predictable	Predictable	Challenging	Challenging
Market Model				Highly random/Unforeseen Events – along with white noise.

With help of "Time Series" we can prepare numerous time-based analyses and results.

Forecasting
Segmentation
Classification
Descriptive analysis`
Intervention analysis

Data Types of Time Series :

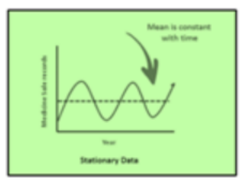
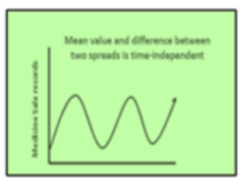
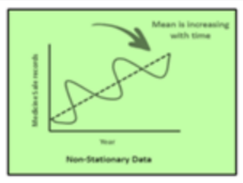
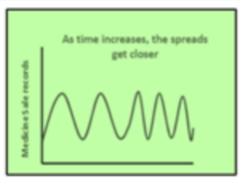
1. Stationary:

A dataset should follow the below thumb rules, without having Trend, Seasonality, Cyclical, and Irregularity component of time series

-- The MEAN value of them should be completely constant in the data during the analysis
-- The VARIANCE should be constant with respect to the time-frame
-- The COVARIANCE measures the relationship between two variables.

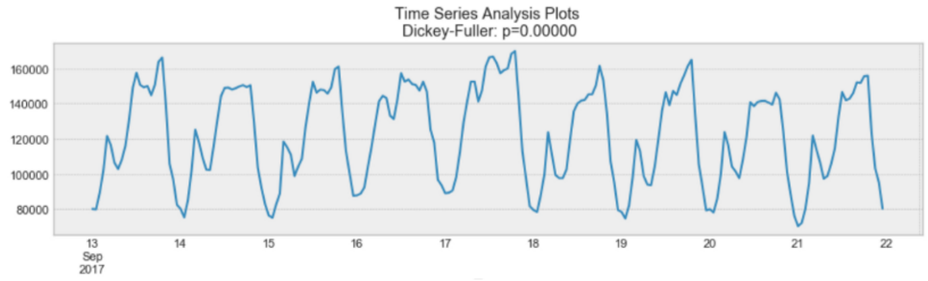
2. Non- Stationary:

This is just the opposite of Stationary.

	MEAN	Variance	Covariance
Stationary			
Non-Stationary			

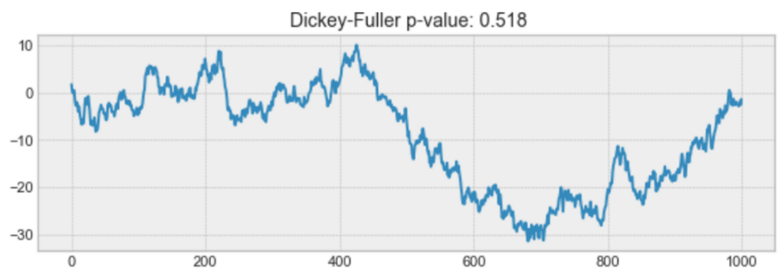
How to analyze Time Series?

1. Collecting the data and cleaning it
2. Preparing Visualization with respect to time vs key feature
3. Observing the stationarity of the series
4. Developing charts to understand its nature.
5. Model building – AR, MA, ARMA and ARIMA
6. Extracting insights from prediction



Looking again at the same plot, we see that the process above is stationary. The mean and variance do not vary over time.

Ideally, we want to have a stationary time series for modelling. Of course, not all of them are stationary, but we can make different transformations to make them stationary.



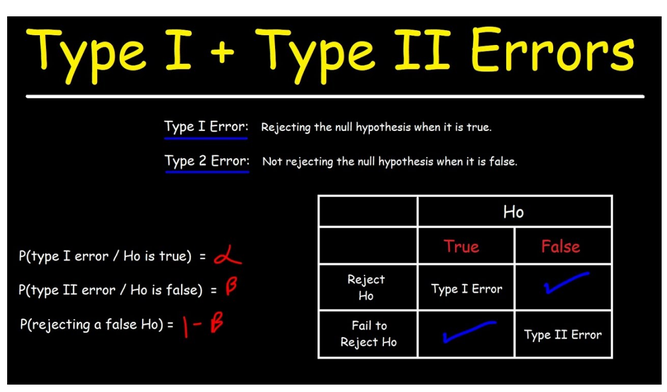
How to check stationarity :

1. AD fuller test

The ADF test is the most popular statistical test and with the following assumptions.

Null Hypothesis (H0): -- Series is non-stationary
Alternate Hypothesis (H1): -- Series is stationary

if p-value >0.05 Fail to reject (H0)/do not reject the null hypothesis/Accept H0
p-value <= 0.05 reject (H0)/reject the null hypothesis/Accept (H1)

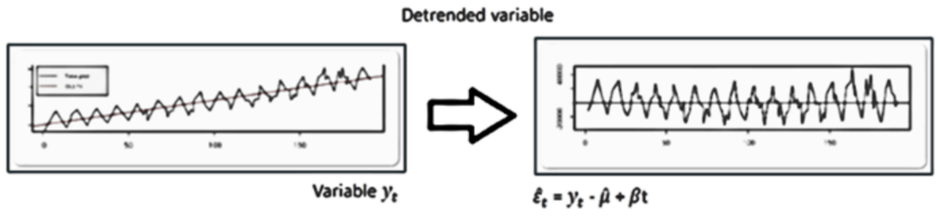


Null Hypothesis: it treats everything same or equal.
P-Value: It is minimum significant level at which you can reject null hypothesis.
Lower the P value better it is to reject the null hypothesis.

Converting Non- stationary into stationary

1. Detrending:

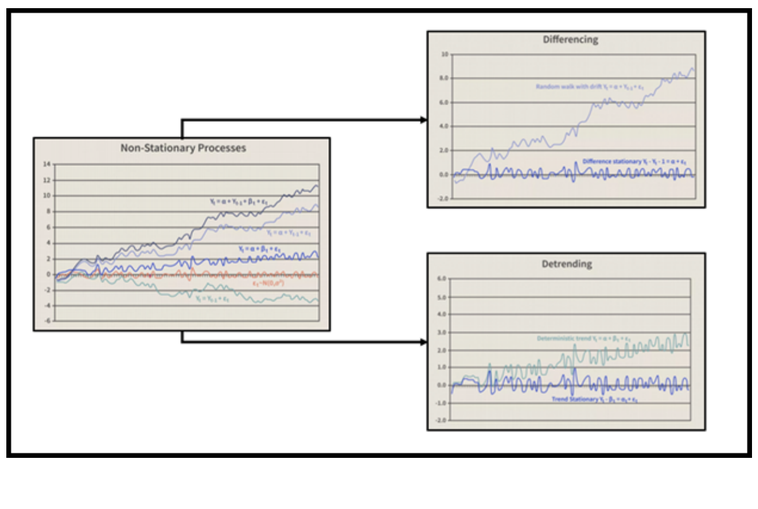
It involves removing the trend effects from the given dataset and showing only the differences in values from the trend.
it always allows the cyclical patterns to be identified.



2. Differencing:

This is a simple transformation of the series into a new time series, which we use to remove the series dependence on time and stabilize the mean of the time series, so trend and seasonality are reduced during this transformation.

Yt= Yt – Yt-1
Yt=Value with time



3. Transformation:

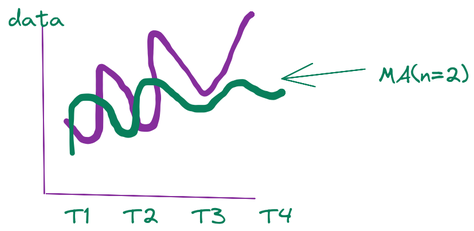
This includes three different methods they are Power Transform, Square Root, and Log Transfer., most commonly used one is Log Transfer.

Rolling Stats

T1	T2	T3	T4	T5
300	310	420	530	640

MA(Moving Average): lets n= 2 (Window)

	MA1	MA2	MA3	MA4	MA5
MA with window 2 =	-	310+300/2	310+420/2		



-- Used for smoothening of Time series
--- Used for feature engineering as well

Time Series Analysis in Data Science and Machine Learning

Lagged Series:

Lag is essentially delay. Just as correlation shows how much two timeseries are similar, autocorrelation describes how similar the time series is with itself.

temp = { Jan, Feb, March, April, May}
Lagged temp series by 1 = { Feb, March, April, May}

Otherwise take a example of a fish price today and last month and before last month.

Auto Regression (AR) :

An autoregressive model predicts future behavior using data from the past. When there is some correlation between values in a time series and the values that precede and succeed them, it is utilized for forecasting.

Autocorrelation Function (ACF) :

- ACF is an (complete) auto-correlation function which gives us values of auto-correlation of any series with its lagged values
- In simple terms, it describes how well the present value of the series is related with its past values.

- We care about direct and all indirect correlation here that price of fish was on july and august will effect price at september.

Partial Autocorrelation Function (PACF) :

- it is also a correlation function but removes the relationship explained by previous lags.
- We see the direct effect that t-2 is having on t.

- we consider the direct effects which tells us how much price of fish in september correlated with july.