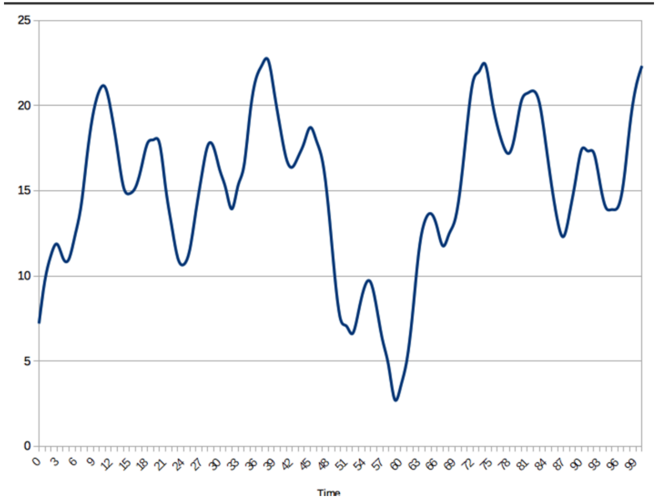
Time Series

- Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.
- Time series forecasting is the use of a model to predict future values based on previously observed values.
- Time series are widely used for non-stationary data, like economic, weather, stock price, and retail sales etc.

Understanding Problem Statement

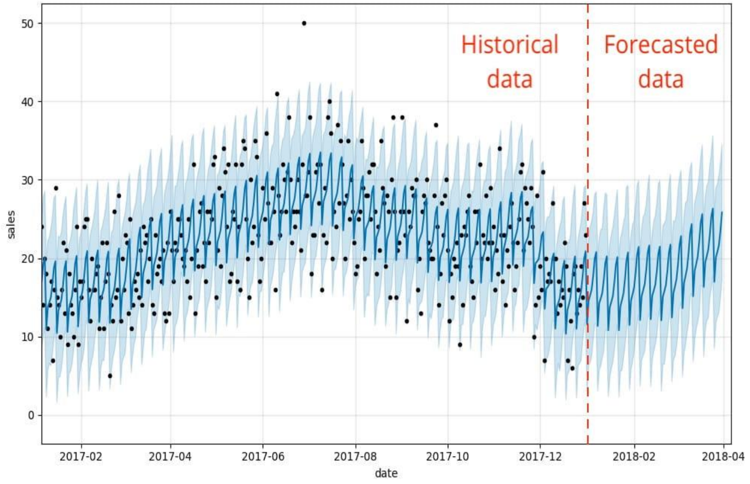
What is Time Series

- Time Series is a sequence taken at successive equally placed numerical points in time.
- It is a sequence of discrete time data.



Time Series Forecasting

- Time Series forecasting involves taking models fit on historical data and using them to predict future observations.
- Making predictions about the future is called **extrapolation** in the classical statistical handling of time series data.



Understanding Problem

- Predict the future customers in a shop.
    - Data contains the number of customers visiting the shop monthly from the year 1949 to 1960.
- We will make use of time series models to forecast the future customers visiting the shop.

Treating Missing Values

Missing Values

- Missing data or missing values are defined as the data value that is not stored for a variable in the observation of interest.
- The values which are missing from the data are called missing values.

Missing Values Example

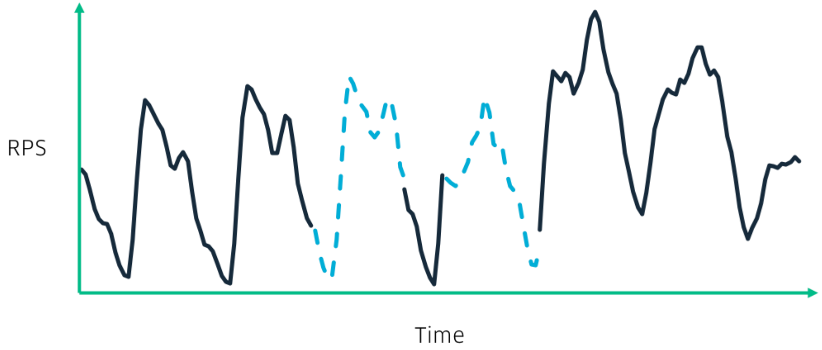
- There are some NaN values in the table.
- NaN value corresponds to the value which is missing from the data.

```
df
```

	column_a	column_b	column_c	column_d
0	1.0	1.2	a	True
1	2.0	1.4	?	True
2	4.0	NaN	c	NaN
3	4.0	6.2	d	None
4	NaN	NaN	--	False
5	NaN	1.1	NaN	True
6	6.0	4.3	d	False

Treating Missing Values

- Mean Imputation
- Last Observation Carried Forward
- Linear Interpolation
- Seasonal Interpolation

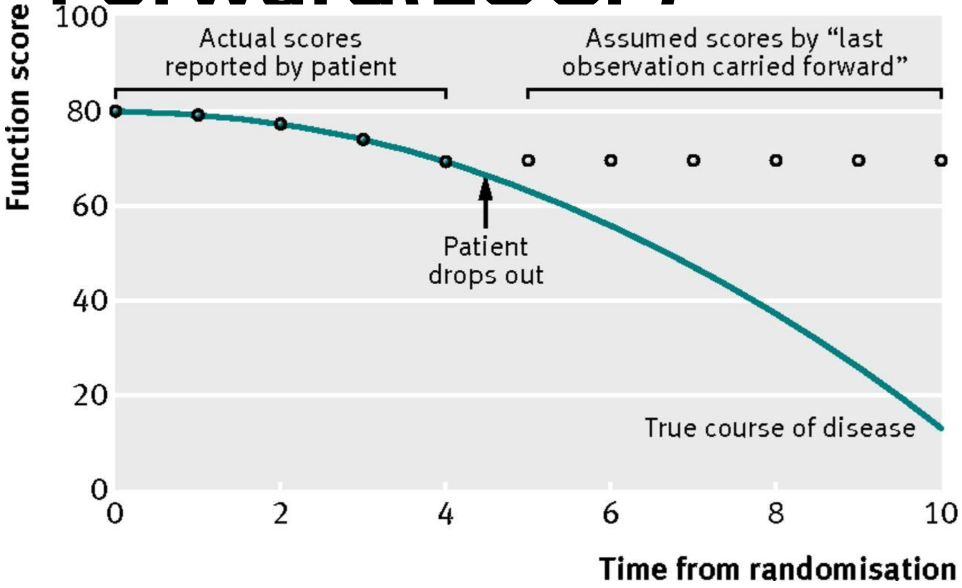


Mean Imputation

- We fill the missing values by the mean of the dataset.
- It can distort the seasonality of the dataset and won't take into consideration the nature of the data.

Item	Y	X
1	9	7
2	?	10
3	11	19
4	?	10
5	15	14
6	19	18
7	21	5
8	8	4
9	19	21
10	21	17

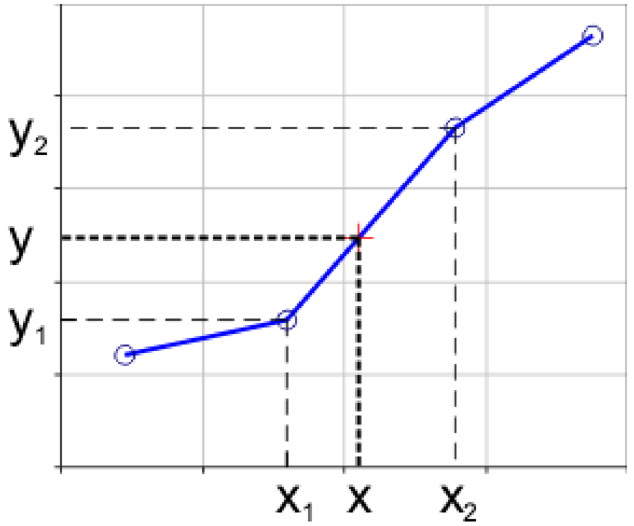
Last Observation Carried Forward(LOCF)



- We impute the missing values with previous value in the data.

Linear Interpolation

- We draw a straight line joining the previous and next points of the missing values.



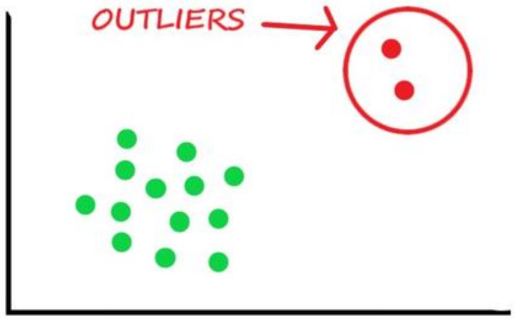
Handling Outliers

Outliers

- Outliers are extreme values that fall a long way outside of the other observations.
- Outliers are those values which are significantly very much different than other values present in the data.

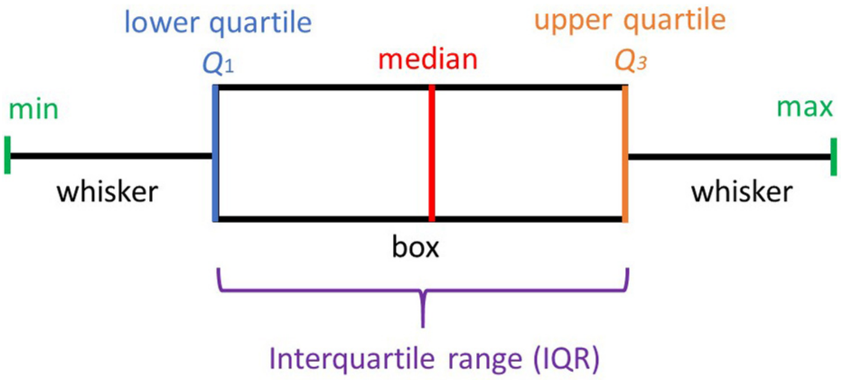
Reasons for Outliers

- In the image shown, the red colored observations are the outliers of the data.
- **Reasons for Outliers**
  - Entry Errors
  - Measurement Errors
  - Natural Errors



Detecting Outliers

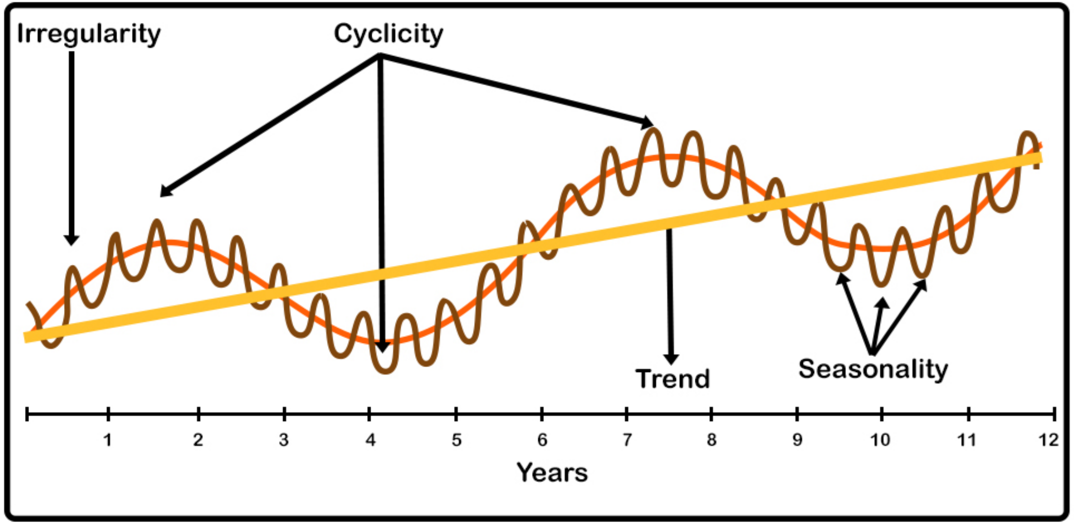
- A **Boxplot** is a standardized way of displaying the distribution of data.
- Any data which are less than **Q1-1.5*IQR** or greater than **Q3+1.5*IQR** is considered to be outliers.
  - IQR**-** interquartile range (Q3-Q1)



Time Series Decomposition

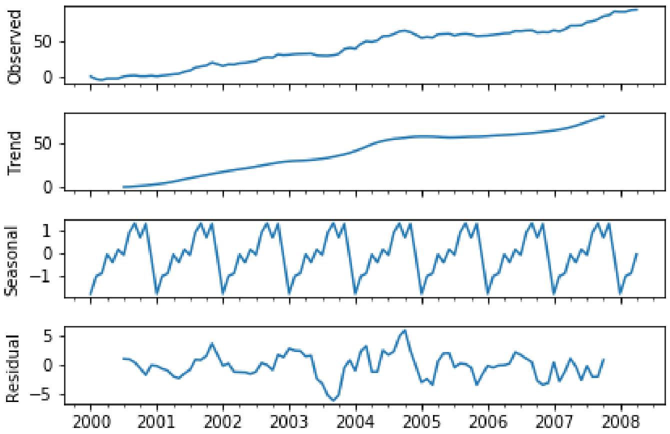
Components of Time Series

- Three main components of Time Series
  - Level
  - Trend
  - Seasonality



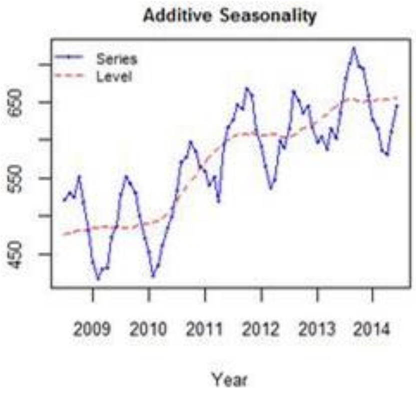
Time Series Decomposition

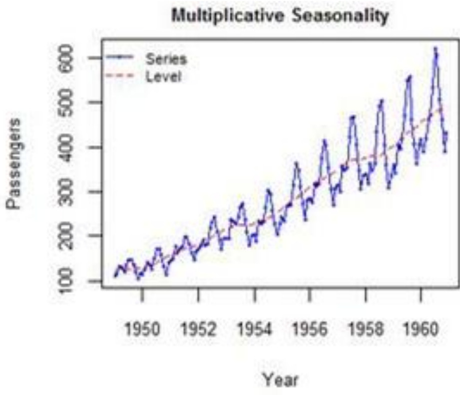
- Time series decomposition involves thinking of a series as a combination of level, trend, seasonality, and noise components.
- It provides a useful abstract model for thinking about time series and for better understanding problems during time series analysis and forecasting.



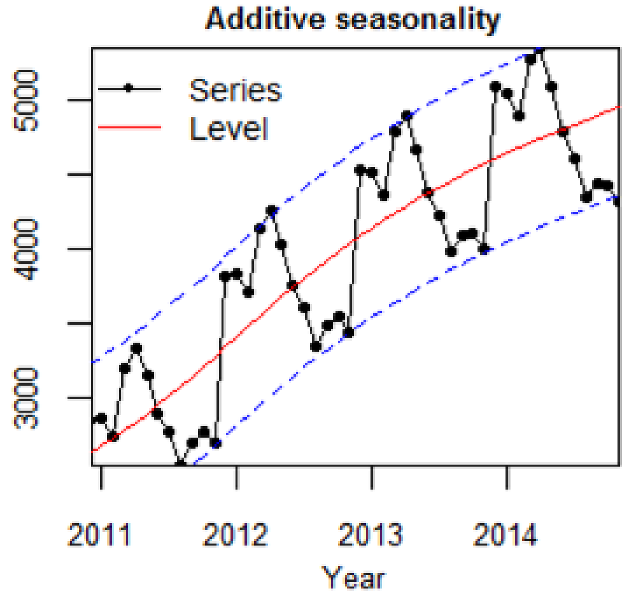
Time Series Decomposition

- There are two ways of decomposing time series data:-
  - Additive Seasonal decomposition
  - Multiplicative Seasonal decomposition





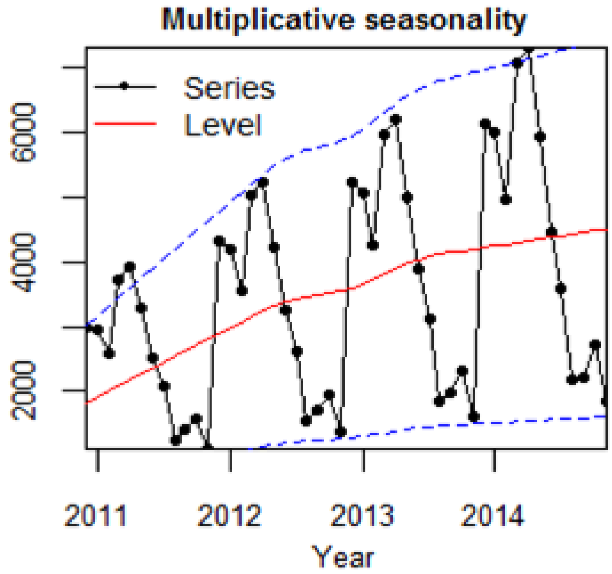
Additive Seasonal Decomposition



Additive Seasonal decomposition is when we add the individual components to get the time series data.

y(t) = Level + Trend + Seasonality + Noise

Multiplicative Seasonal Decomposition



Multiplicative Seasonal decomposition is when we multiply the individual components to get the time series data.

y(t) = Level * Trend * Seasonality * Noise

Additive vs Multiplicative Decomposition

- We use additive model, the magnitude of seasonality does not vary with the level of the time series.
- On the other hand, we use multiplicative models when the variation in the seasonal pattern appears to be directly proportional to the level of the time series data.

Holt’s Winter Exponential Smoothing

Holt’s Winter Exponential Smoothing

- Holt’s Winter Exponential Smoothing technique is a method which includes all the three components of the time series:- level, trend and seasonality.
- Holt winters is also called Triple Exponential Smoothing.
- The model predicts future value by computing the combined effects of three influences.

Holt’s Winter Exponential Equation

- The Holt’s Winter Exponential Smoothing technique equation is:-
  - y(t+1) = l(t) + b(t) + s(t+1-m)
    - m is the number of times a season repeats in a time period.

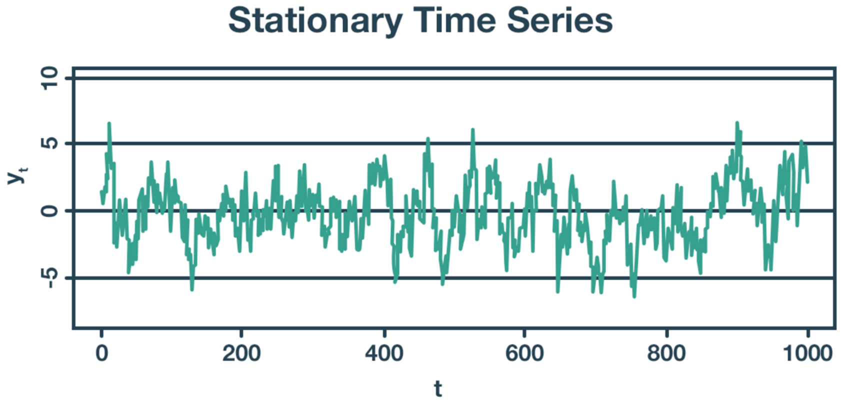
Assumptions of the SARIMA Model

Assumptions of building SARIMA model

- There are two assumptions of applying the SARIMA model.
    - Stationarity
    - Autocorrelation

Stationarity

- Stationarity means that the statistical properties of a process generating a time series do not change over time.
- The statistical properties are
  - **Mean**
  - **Variance**
  - **Covariance**.



How to Check Stationarity?

- There are two popular statistical tests using which we can test the stationarity of a time series.
  - **Augmented Dickey-Fuller(ADF) test.**
  - **Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.**

Ad Fuller Test vs KPSS Test

Ad Fuller Test

- **Null Hypothesis:** The time series is not stationary.
- **Alternate Hypothesis:** The time series is stationary.

KPSS Test

- **Null Hypothesis:** The time series is stationary.
- **Alternate Hypothesis:** The time series is not stationary.

Converting Non-stationary into Stationary

- There are two tools for converting a non-stationary series into a stationary series:
  - **Box Cox Transformation**
  - **Differencing**

Autocorrelation

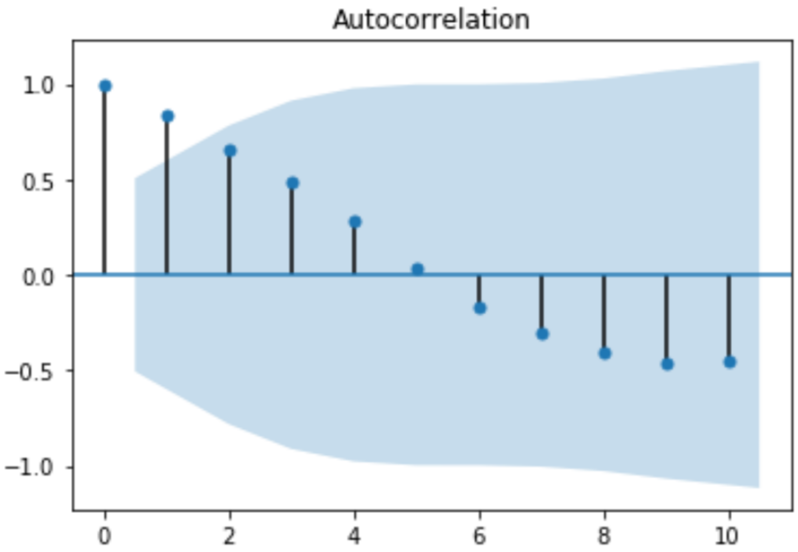
- Autocorrelation is the representation of the degree of similarity between a given time series and the lagged version of itself over successive time intervals.
- It helps us to know a variable is influenced by its own lagged values.

Types of Autocorrelation

- There are two types of Autocorrelation measures:
  - **Autocorrelation function**
  - **Partial Correlation function**

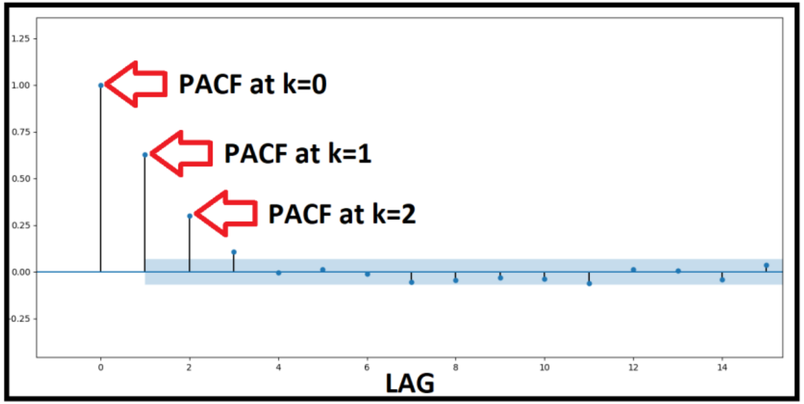
Autocorrelation function

- ACF is an (complete) autocorrelation function which gives us values of auto-correlation of any series with its lagged values.



Partial Autocorrelation function

- Partial autocorrelation function(PACF) gives the partial correlation of a stationary time series with its own lagged values, regressed the values of time series at all shorter lags.



SARIMA Model

Key Elements of SARIMA Model

- The Time series is differenced to make it stationary.
- The SARIMA equation is a linear combination of past observations and past errors.
- Seasonal differencing is performed on the time series.
- SARIMA models future seasonality as a linear combination of past seasonality observations and past seasonality errors.

Parameters of SARIMA Model

- The SARIMA model has the following parameters **p, q, d, m, P, Q, D**

  p- highest lag in the time series

  q- past errors included.

  d- degree of differencing to make time series stationary.

  m - the number of time steps for a single seasonal period

  P - the seasonal autoregressive order. **0<=P<=4**

  Q - the seasonal moving average order. **0<=Q<=4**

  D - the seasonal difference order. **D=0, 1, 2**

Major Takeaways

Major Takeaways

- You learnt about handling missing values in the time series data.
- You learnt about handling outliers present in the time series data.
- You learnt about two different decompositions of time series data.
- You learnt about Holt winter exponential smoothing model.
- You learnt about stationary data and check stationarity of the data.

Major Takeaways

- You learnt about the ways of converting non-stationary data to stationary data.
- You learnt about autocorrelation function and partial autocorrelation function.
- You learnt about implementation of the SARIMA model.
- You learnt evaluating time series model.

Prophet Model

Introduction to Prophet Model

- Prophet is an open source software released by Facebook's Core Data Science Team.
- It is available for download on CRAN and PyPI.
- Prophet is a forecasting procedure implemented in R and Python.
- It is fast and provides completely automated forecasts that can be tuned by hand by data scientists and analysts.

Prophet Model

- Prophet is a procedure used for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly and daily seasonality plus holiday effects.
- It provides us with the ability to make time series predictions with good accuracy using simple intuitive parameters.

Advantages of Prophet Model

- Prophet is used in many applications across Facebook for producing reliable forecasts for planning and goal setting.
- Prophet is robust to outliers, missing data and dramatic changes in our time series.
- It provides many possibilities for users to tweak and adjust forecasts.

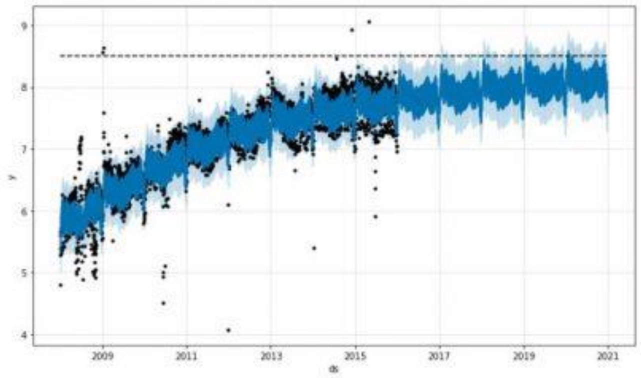
Working of Prophet Model

Working of Prophet Model

- The Prophet model equation is:-
  - $y(t) = g(t) + s(t) + h(t) + \epsilon(t)$
    - $g(t)$ is the piecewise linear growth curve for modelling non-periodic changes in time series.
    - $s(t)$ means periodic changes.
    - $h(t)$ means effects of holidays.
    - $\epsilon(t)$ is the error term

Components of Prophet Model

- **Trend**
  - Trend is modelled by fitting a piecewise linear curve over the trend or non-periodic part of the time series.
  - The linear fitting exercise ensures that it is least affected by spikes/missing data.



Components of Prophet Model

- **Seasonality**
  - The Seasonality s(t) provides an adaptability to the model by allowing periodic changes based on sub-daily, daily, weekly and yearly seasonality.
  - To fit and forecast the effects of seasonality, prophet relies on fourier series to provide a flexible model.

Components of Prophet Model

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right)$$

- P is the period
- The fourier order N that defines whether high frequency changes are allowed to be modelled is an important parameter to set here.

Fourier Series

- A fourier series is a periodic function composed of harmonically related sines and cosines combined by a weighted summation.

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right)\right)$$

- The fourier series has many applications such as:-
  - Electrical engineering
  - Vibration analysis

Applications of Fourier Series

- Optics
- Signal processing
- Image processing
- Econometrics

- The Fourier series of functions in the differential equation often gives some prediction about the behavior of the solution of the differential equation.

Components of Prophet Model

- **Holidays**
  - Impact of a particular holiday on the time series is often similar year after year, making it an important incorporation into the forecast.
  - Holidays and events incur predictable shocks to a time series.
  - For instance, Diwali in India occurs on a different day each year and a large portion of the population buy a lot of new items during this period.