Machine Leaning

Time Series

Cyrus Lentin

Time Series - Definition

- A Time Series Is A Set Of Observation Taken At Specified Times, Usually At ‘Equal Intervals’
- A Set Of Data Depending On The Time
- A Series Of Values Over A Period Of Time
- Collection Of Data Belonging To Different Time Periods Of Some Variable Or Composite Of Variables
- Example

Production Of Steel	Per Capita Income
Gross National Income	Price Of Tobacco
Index Of Industrial Production	Price Of Share

- Mathematically A Time Series Is Defined By The Values Y1, Y2…of A Variable Y At Times t1, t2
- Thus,
  Y= F(t)
- In Time Series, Time Act As An Independent Variable To Estimate Dependent Variables

Variations In Time Series

- Social Customs, Festivals Etc.
- Seasons
- The Four Phase Of Business : Prosperity, Decline, Depression, Recovery
- Natural Calamities: Earthquake, Epidemic, Flood, Drought Etc.
- Political Movements/Changes, War Etc.

Importance Of Time Series Analysis

- Popular Tool For Business Forecasting
- Basis For Understanding Past Behavior
- Can Forecast Future Activities; Planning For Future Operations
- Evaluate Current Accomplishments; Evaluation Of Performance
- Facilitates Comparison
- Estimation Of Trade Cycle

Components Of Time Series

What Is Component

- Characteristic Movements Or Fluctuations Of Time Series

Components Of Time Series

What Is Component

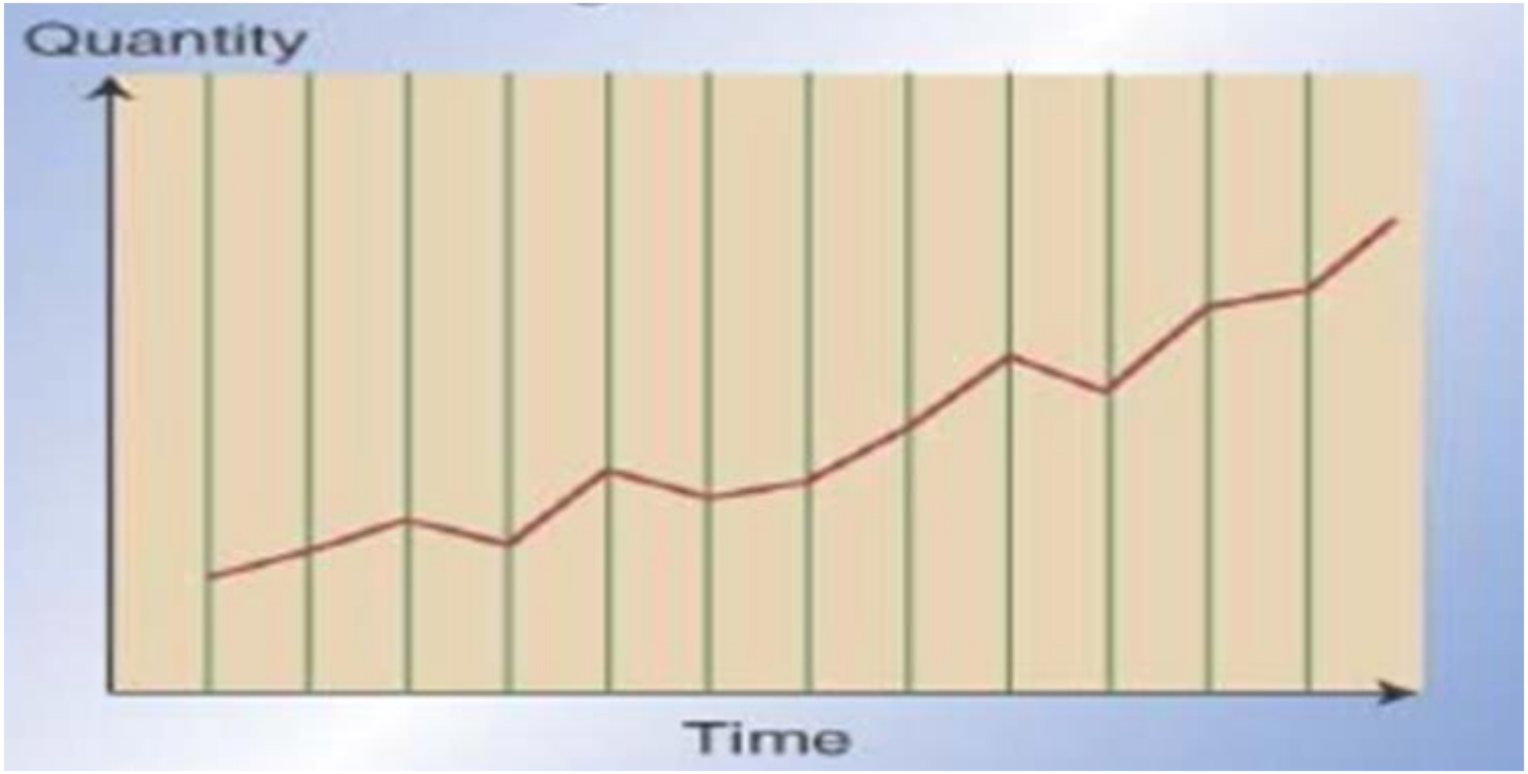
- Characteristic Movements Or Fluctuations Of Time Series

Types of Components

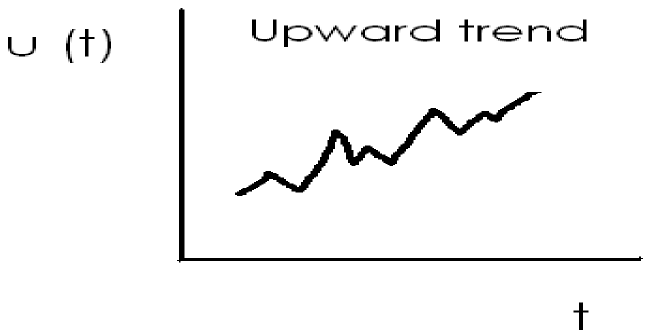
- Trend - Secular Trend
- Seasonality - Seasonal Variations / Fluctuations
- Cyclical - Cyclical Variations / Fluctuations
- Irregular Variations / Movements

Trend

- The General Tendency Of The Data To Grow Or Decline Over A Long Period Of Time.
- The Forces Which Are Constant Over A Long Period (Or Even If They Vary They Do So Very Gradually) Produce The Trend.



Trend – Characteristics



- Downward Trend-declining Death Rate
- Upward Trend-population Growth

Mathematically Trend May Be Linear Or Non-linear

Secular Trend - Examples

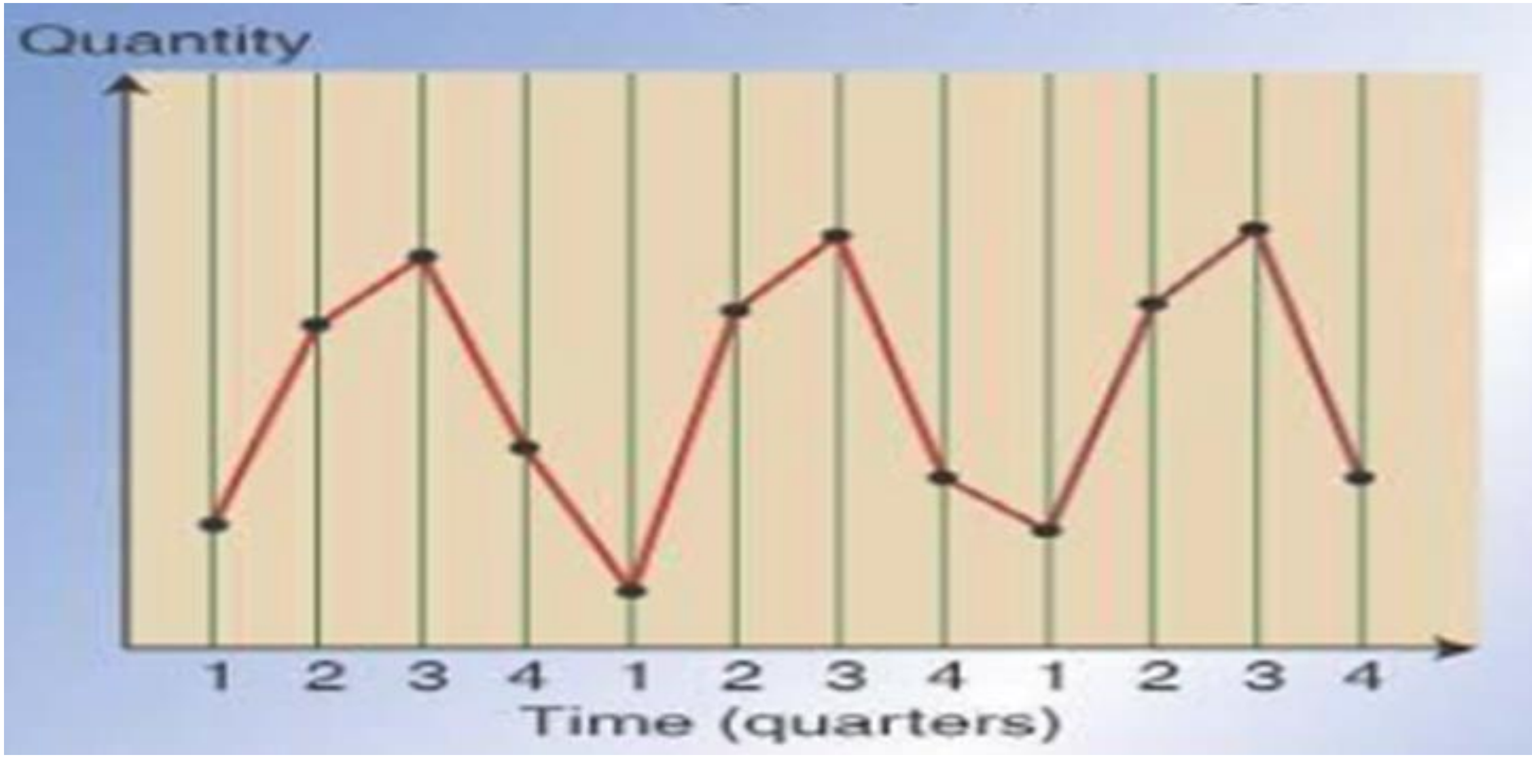
- Population Change
- Technological Progress
- Global Temperature
- Improvement In Business
- Better Medical Facility
- Formation Of Rocks

Trends – Why Measure?

- Knowledge Of Past Behavior
- Estimation
- Study Of Other Components

Seasonality

- The Component Responsible For The Regular Rise Or Fall (Fluctuations) In The Time Series During A Period Not More Than 1 Year.
- Fluctuations Occur In Regular Sequence The Period Being A Year, A Month, A Week, A Day, Or Even A Fraction Of The Day, An Hour Etc.



Seasonality

- The Component Responsible For The Regular Rise Or Fall (Fluctuations) In The Time Series During A Period Not More Than 1 Year.
- Fluctuations Occur In Regular Sequence (Periodical)
- The Period Being A Year, A Month, A Week, A Day, Or Even A Fraction Of The Day, An Hour Etc.
- Term “Seasonal” Is Meant To Include Any Kind Of Variation Which Is Of Periodic Nature And Whose Repeating Cycles Are Of Relatively Short Duration.
- The Factors That Cause Seasonal Variations Are:
  - Climate & Weather Condition
  - Customs Traditions & Habits

Seasonality - Examples

- The Factors That Cause Seasonal Variations Are:
  - Climate & Weather Condition
  - Customs Traditions & Habits

- Examples
  - Demands for woolen clothes goes up in winter
  - Price increases during festivals
  - Withdraws from banks are heavy during first week of the month.
  - The number of letter posted on Saturday is larger

Seasonality – Characteristics

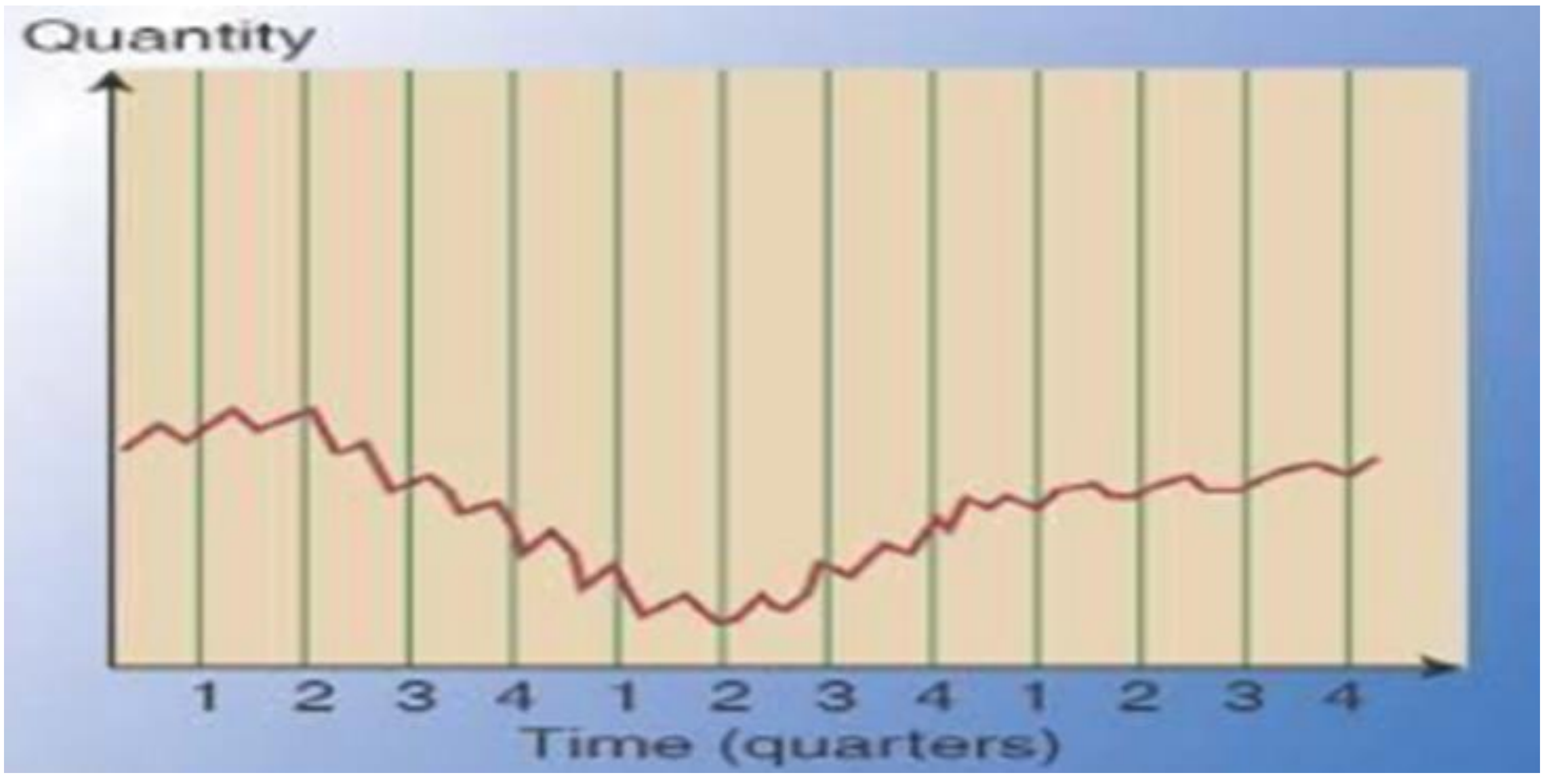
- Regularity
- Fixed Proportion
- Increase Or Decrease
- Easy To Forecast

Seasonality – Why Measure?

- Analysis Of Past Behavior Of The Series
- Forecasting The Short Time Fluctuations
- Elimination Of The Seasonal Variations For Measuring Cyclic Variations

Cyclicity

- Cycle Refers To Recurrent Variations In Time Series; Variations That Usually Last Longer Than A Year
- Cyclic Fluctuations / Variations Are Long Term Movements That Represent Consistently Recurring Rises And Declines In Activity



Cyclicity - Characteristics

Business Cycle Consists Of 4 Phases

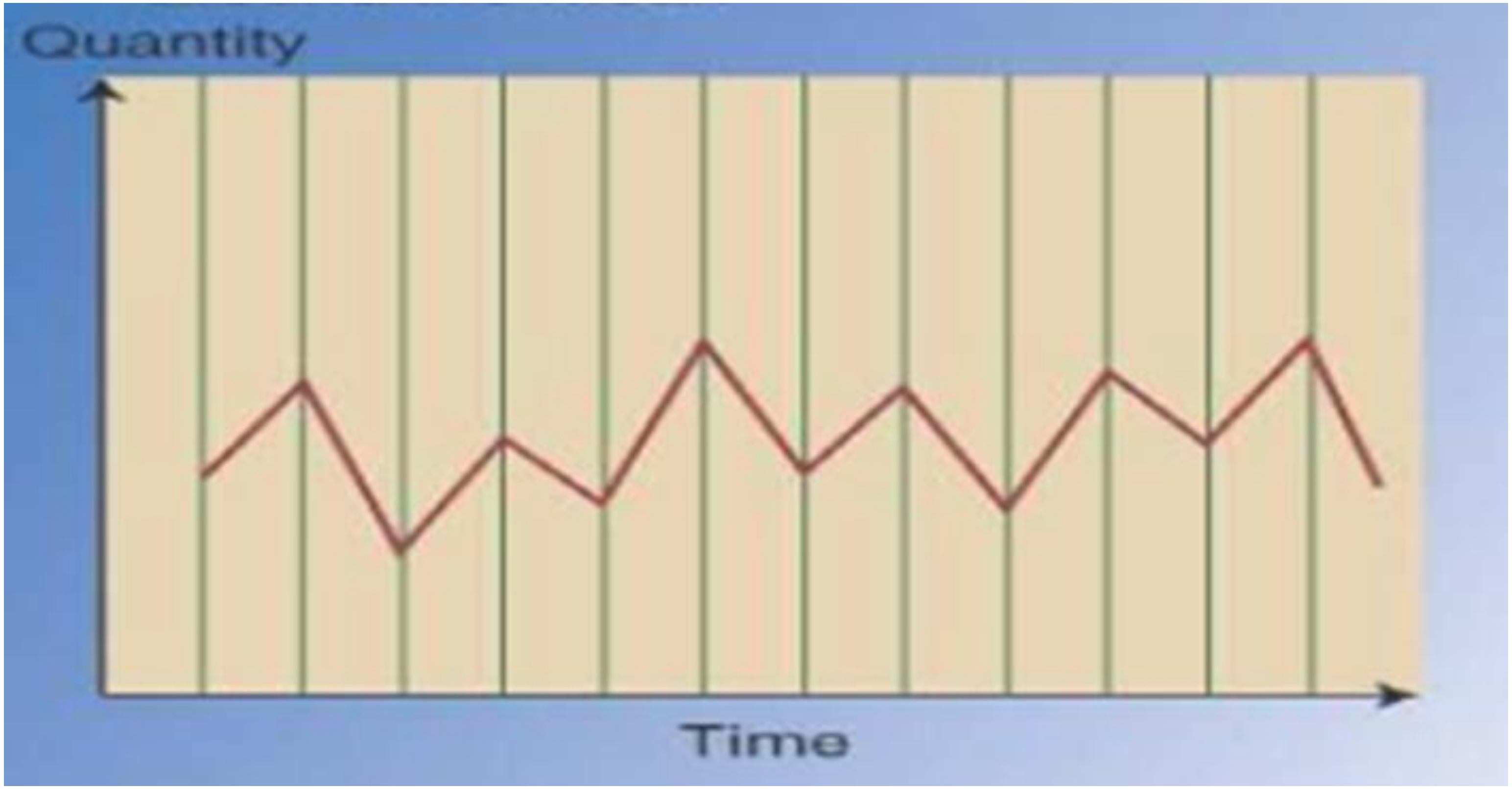
- Prosperity
- Decline
- Depressions
- Recovery

Cyclicity – Why Measure?

- Measures of past cyclical behavior
- Forecasting
- Useful in formulating policies in business

Irregularity

- Also Called Erratic, Random, Or “Accidental” Variations
- Do Not Repeat In A Definite Pattern
- Strikes, Fire, Wars, Famines, Floods, Earthquakes
- Unpredictable



Irregularity - Characteristics

- Irregular & Unpredictable
- No Definite Pattern
- Short Period Of Time
- No Statistical Technique

Handling Missing Values - ffill

- When ffill() is applied across the index then any missing value is filled based on the corresponding value in the previous row.

	A	B	C	D
0	5.0	NaN	4	5.0
1	3.0	2.0	3	4.0
2	NaN	4.0	8	2.0
3	4.0	3.0	5	NaN

	A	B	C	D
0	5.0	NaN	4	5.0
1	3.0	2.0	3	4.0
2	3.0	4.0	8	2.0
3	4.0	3.0	5	2.0

Handling Missing Values - bfill

- When bfill() is applied across the index then any missing value is filled based on the corresponding value in the next row.

	A	B	C
0	NaN	11.0	NaN
1	1.0	5.0	5.0
2	2.0	NaN	10.0
3	3.0	NaN	11.0
4	NaN	NaN	NaN
5	NaN	8.0	8.0

	A	B	C
0	1.0	11.0	5.0
1	1.0	5.0	5.0
2	2.0	8.0	10.0
3	3.0	8.0	11.0
4	NaN	8.0	8.0
5	NaN	8.0	8.0

Handling Missing Values - interpolate

- When interpolate() is applied across the index then any missing value is filled with mean of the values in the previous and next rows.

	A	B	C	D
0	12.0	NaN	20.0	14.0
1	4.0	2.0	16.0	3.0
2	5.0	54.0	NaN	NaN
3	NaN	3.0	3.0	NaN
4	1.0	NaN	8.0	6.0

	A	B	C	D
0	12.0	NaN	20.0	14.0
1	4.0	2.0	16.0	3.0
2	5.0	54.0	9.5	4.0
3	3.0	3.0	3.0	5.0
4	1.0	3.0	8.0	6.0

Handling Missing Values – KNN Mean

- Compute the mean of K nearest rows up & down

Handling Missing Values – Seasonal Values

- Compute the mean of corresponding seasonal periods

Smoothing

- There Are No Proven "Automatic" Techniques To Identify Trend Components In The Time Series Data
- However, As Long As The Trend Is Monotonous That Part Of Data Analysis Is Typically Not Very Difficult
- If The Time Series Data Contains Considerable Error, Then The First Step In The Process Of Trend Identification Is Smoothing
- Smoothing Always Involves Some Form Of Local Averaging Of Data Such That The Nonsystematic Components Of Individual Observations Cancel Each Other Out
- The Most Common Technique Is Moving Average Smoothing Which Replaces Each Element Of The Series By Either The Simple Or Weighted Average Of N Surrounding Observations

Moving Average

- Time Series Data Such As Stock Prices, Weather Observations Or Experimental Results Often Contain Random Fluctuations Which Are Referred To As "Noise"
- Noise Can Make It Difficult To Infer Useful Things From The Data Stream Such As Trend And Cycles
- The Simplest Way Of Reducing The Effect Of Noise Is To Use An Historical Average Of N Observations
- Moving Averages Are Described By The Number Of Observations Used In The Calculations
- For Example, MA20 Is An Average Based On The 20 Most Recent Observations

Simple Moving Average

- Simple Moving Average Is Mean / Average Of Last N Readings
- Simple Moving Average Is A Lagging Indicator Since It Uses Historical Data

Rolling average with a window size of 3

df_rainfall

year	accumulated_rainfall
1786	722.2
1787	625.1
1788	728.8
1789	342.5
1790	855.7

(722.2+625.1+728.8)/3 = 692.0333

(625.1+728.8+342.5)/3 = 565.4666

(728.8+342.5+855.7)/3 = 642.3333

df_rainfall.rolling(3).mean()

year	accumulated_rainfall
1786	NaN
1787	NaN
1788	692.033333
1789	565.466667
1790	642.333333

Cumulative Moving Average

- The Cumulative Moving Average is unweighted mean of the previous values up to the current time t.
- The simple moving average has a sliding window of constant size M. On the contrary, the window size becomes larger as the time passes when computing the cumulative moving average.

Rolling average with a window size of 3

df_rainfall

year	accumulated_rainfall
1786	722.2
1787	625.1
1788	728.8
1789	342.5
1790	855.7

(722.2+625.1+728.8)/3 = 692.0333

(625.1+728.8+342.5)/3 = 565.4666

(728.8+342.5+855.7)/3 = 642.3333

df_rainfall.rolling(3).mean()

year	accumulated_rainfall
1786	NaN
1787	NaN
1788	692.033333
1789	565.466667
1790	642.333333

Exponential Smoothing

- It Bases Its Forecasts On A Weighted Average Of Past Observations, With More Weight Put On The More Recent Observations
- Simple Exponential Smoothing
  For A Series With No Pronounced Trend Or Seasonality

Exponential moving average

`df_rainfall`

year	accumulated_rainfall
1786	722.2
1787	625.1
1788	728.8
1789	342.5
1790	855.7

722.2 = 722.20

0.1*625.1+0.9*722.20 = 712.49

0.1*728.8+0.9*712.49 = 714.12

0.1*342.5+0.9*714.12 = 676.958

0.1*855.7+0.9*676.958 = 694.833

`df_rainfall.ewm(alpha=0.1, adjust=False).mean()`

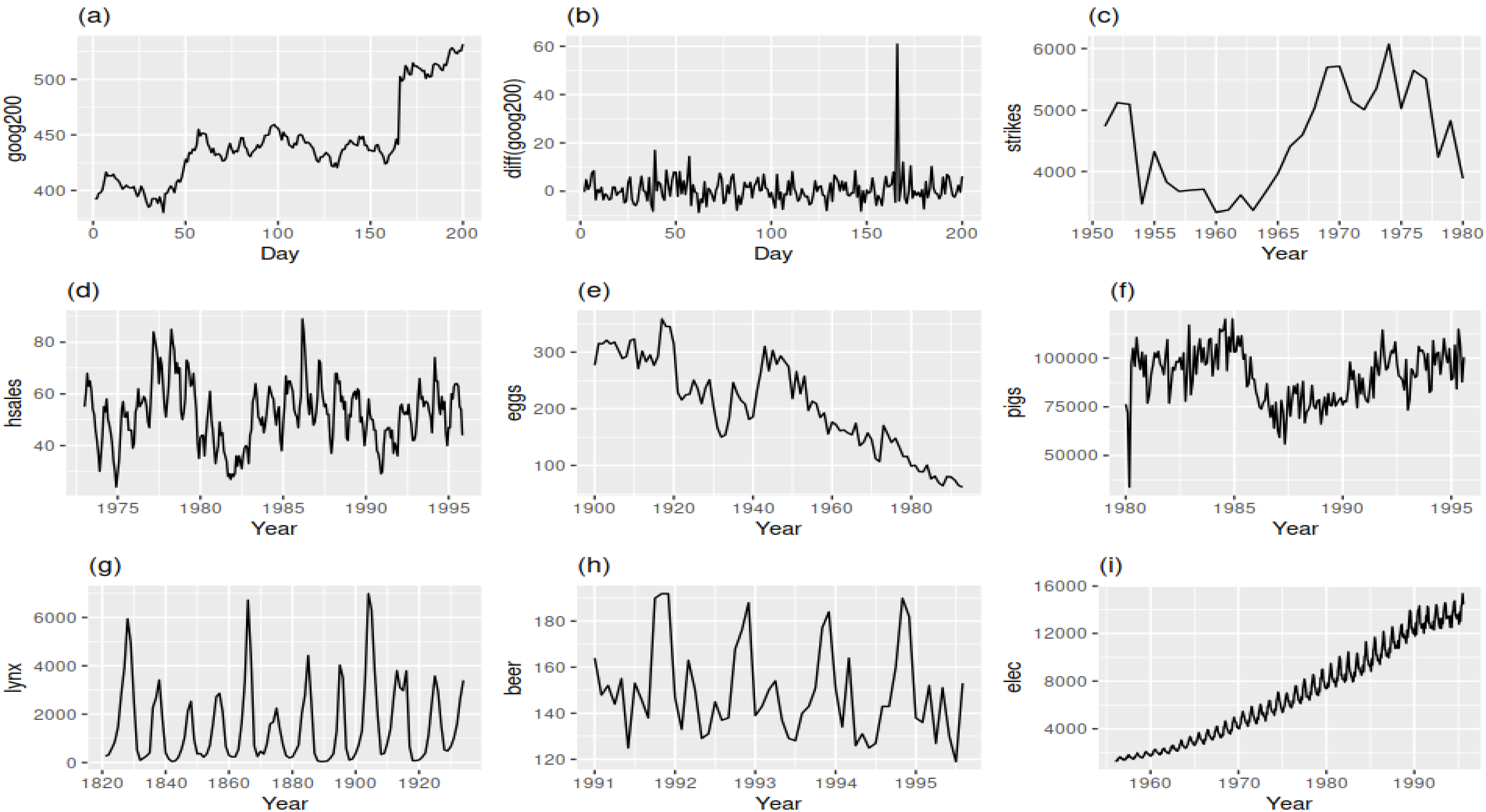
year	accumulated_rainfall
1786	722.20000
1787	712.49000
1788	714.12100
1789	676.95890
1790	694.83301

Stationarity

- A stationary time series is one whose properties do not depend on the time at which the series is observed.
- Thus, time series with trends, or with seasonality, are not stationary — the trend and seasonality will affect the value of the time series at different times
- You can check if your time series is stationary by looking at a line plot of the series over time.
- Sign of obvious trends, seasonality, or other systematic structures in the series are indicators of a non-stationary series.
- A more accurate method would be to use a statistical test, such as the Dickey-Fuller test.
- Should you make your time series stationary? Generally, yes.
- If you have clear trend and seasonality in your time series, then model these components, remove them from observations, then train models on the residuals.

Stationarity

- Which of these time series is stationary?



Stationarity – Why?

- If we fit a stationary model to data, we assume our data are a realization of a stationary process.
- So our first step in an analysis should be to check whether there is any evidence of a trend or seasonal effects and, if there is, remove them.
- Unless Time Series Is Stationary, An ARIMA Model Can Not Be Built
- In Cases Where The Stationary Criterion Are Violated, The First Requisite Becomes To "Stationarize" The Time Series
- Only After This, We Can Use ARIMA Time Series Models For Forecasting
- There Are Multiple Ways Of Bringing This Stationarity Like Detrending, Differencing Etc.

Differencing

- One of the most common methods of dealing with both trend and seasonality is differencing.
- In this technique, we take the difference of the observation at a particular instant with that at the previous instant.
- This mostly works well in improving stationarity.

De-Trending With Differencing

- First Order Differencing
  When differencing is done once it is called First Order Differencing
- Second Order Differencing
  When differencing is done on already difference data it is called Second Order Differencing

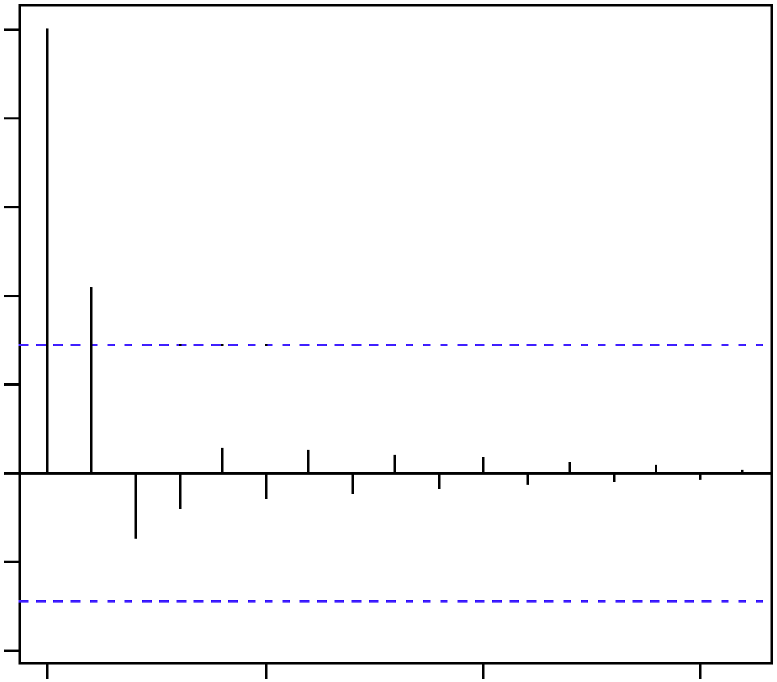
De-Seasonality With Differencing

- Find out the seasonality period

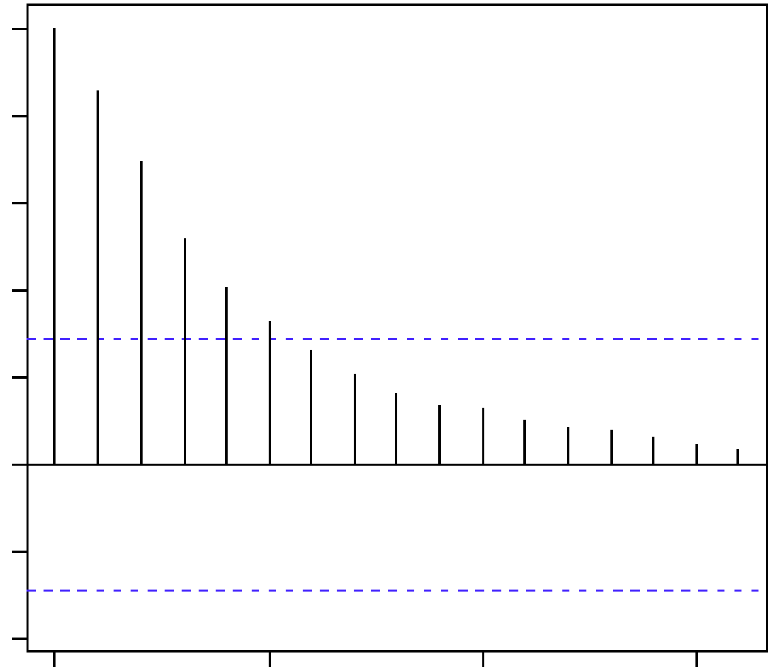
Augmented Dickey-Fuller Test Of Stationarity

- P-value should be less than 0.05
- Small p-values suggest the data is stationary

ACF / PACF For MA

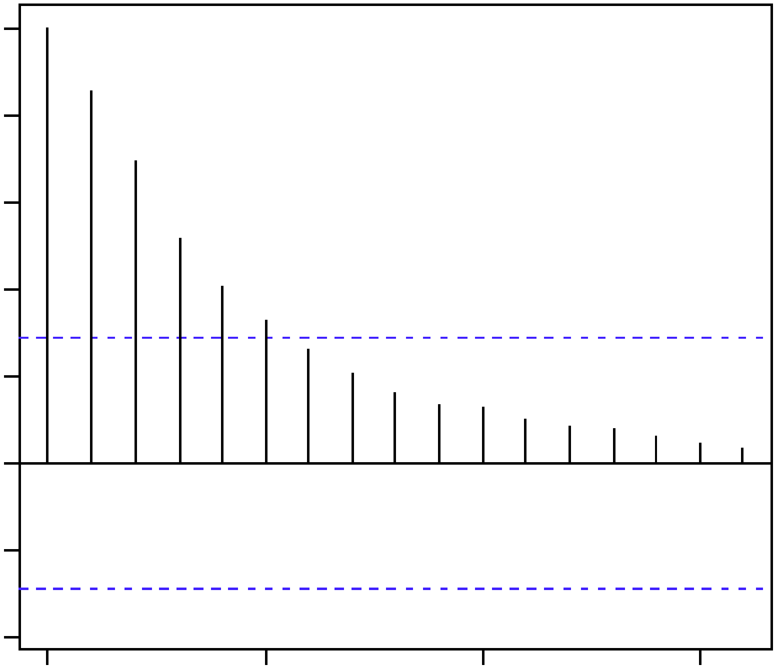


Auto Correlation Function

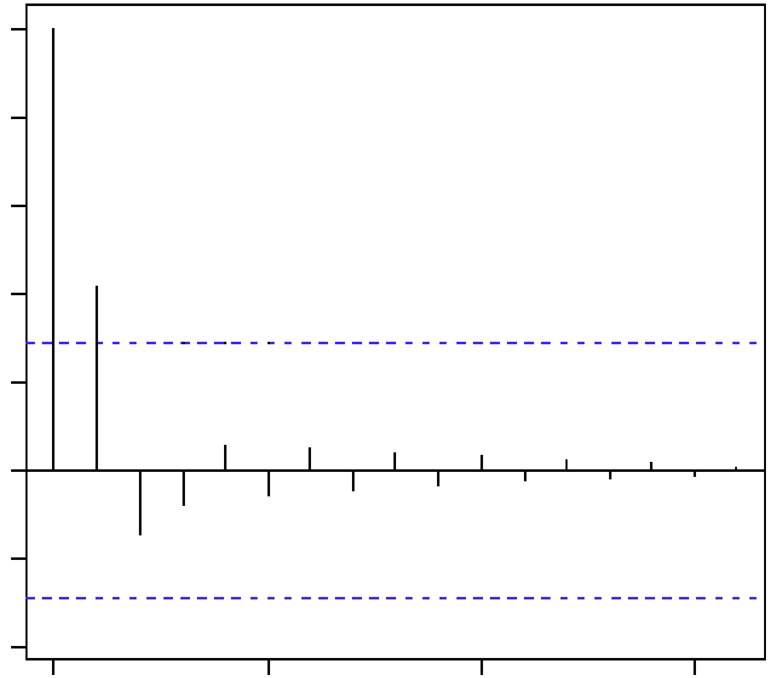


Partial Auto Correlation Function

ACF / PACF For ARIMA



Auto Correlation Function



Partial Auto Correlation Function

ARIMA (Auto Regressive Integrated Moving Averages)

- ARIMA Models Provide Another Approach To Time Series Forecasting
- ARIMA Models Are, In Theory, The Most General Class Of Models For Forecasting A Time Series.
- Arima Models Aim To Describe The Autocorrelations In The Data
- Pre Requisite Of ARIMA Is That The Time Series Needs To Be Stationary.
- A Stationary Series Has No Trend, Its Variations Around Its Mean Have A Constant Amplitude, And It Wiggles In A Consistent Fashion,
  I.E., Its Short-term Random Time Patterns Always Look The Same In A Statistical Sense
- This is a modeling approach that can be used to calculate the probability of future value lying between two specified limits

Steps For Time Series Models

Simple Moving Average

- Visualize Time Series
- Check Stationarity
- Plot ACF / PACF Charts
- Build SMA Model
- Make Forecast

Auto Regressive Integrated Moving Average

- Visualize Time Series
- Check Stationarity
- Plot ACF / PACF Charts
- Build ARIMA Model
- Make Forecast

Thank you!

Contact:

Cyrus Lentin
cyrus@lentins.co.in
+91-98200-94236