TIME SERIES FORECASTING

Why Forecast?

Every organization faces internal and external risks, such as high competition, failure of technology, labor unrest, inflation, recession, and change in government laws.

Every business operates under risk and uncertainty

Forecast is necessary to lessen the adverse effects of risks

There are varied methods of forecast – some of which you have already covered. Such as

- Regression
- Data mining methods

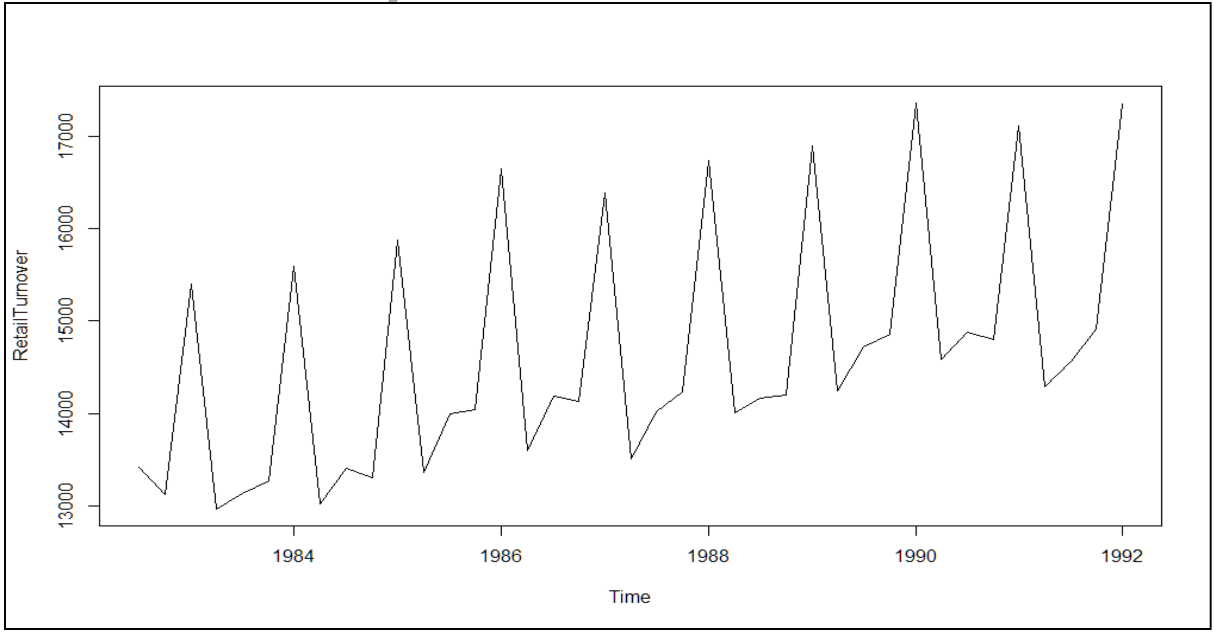
Why Forecast?

Time Series is another technique for forecasting

Why do we need so many techniques for forecasting? Because not all data are the same, or similar. Because different types of data possess different features, different methods of forecast becomes applicable

For example: In regression or CART you have one response and a number of predictors

Quarterly Retail Turnover



Based on the above how to predict turnover for the next two years?

How to Forecast?

This is an example of Time Series Data

The objective of this lesson is to learn about

- What is time series?
- What are the special features of time series data?
- What are the typical situations where time series methods are applied?

What is Time Series?

- A time series is a sequence of measurements on the same variable collected over time.

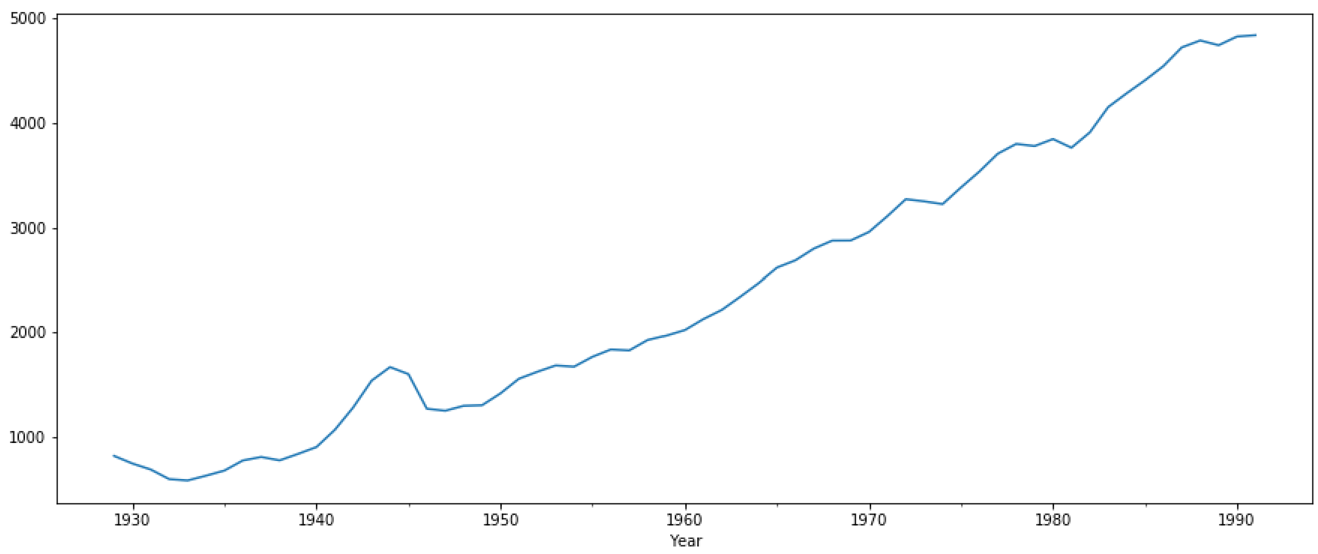
- The measurements are made at regular time intervals.

Intervals of Time Series

- **Yearly** — *GDP, Macro-economic series*
- **Quarterly** — *Revenue*
- **Monthly** — *Sales, Expenditure*
- **Weekly** — *Demand*
- **Daily** — *Closing price of stock*
- **Hourly** — *AAQI*

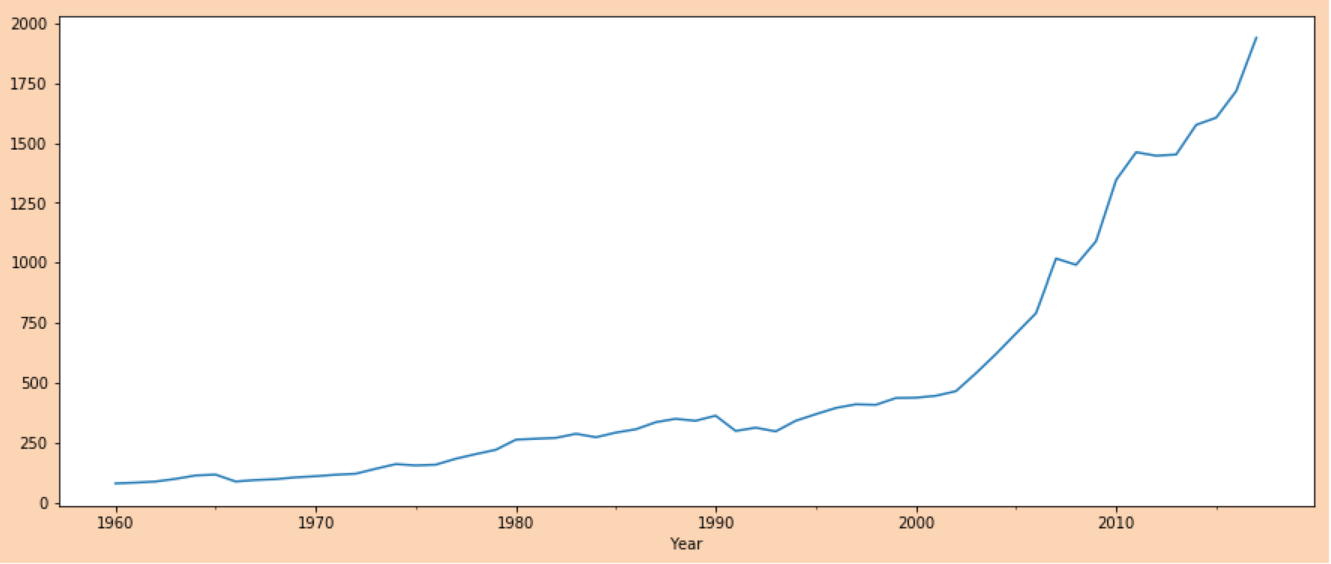
EXAMPLES OF TIME SERIES

Year	1929	1930	1931	1932	……	1989	1990	1991
US GDP (b. USD)	821.8	748.9	691.3	599.7	……	4739.2	4822.3	4835



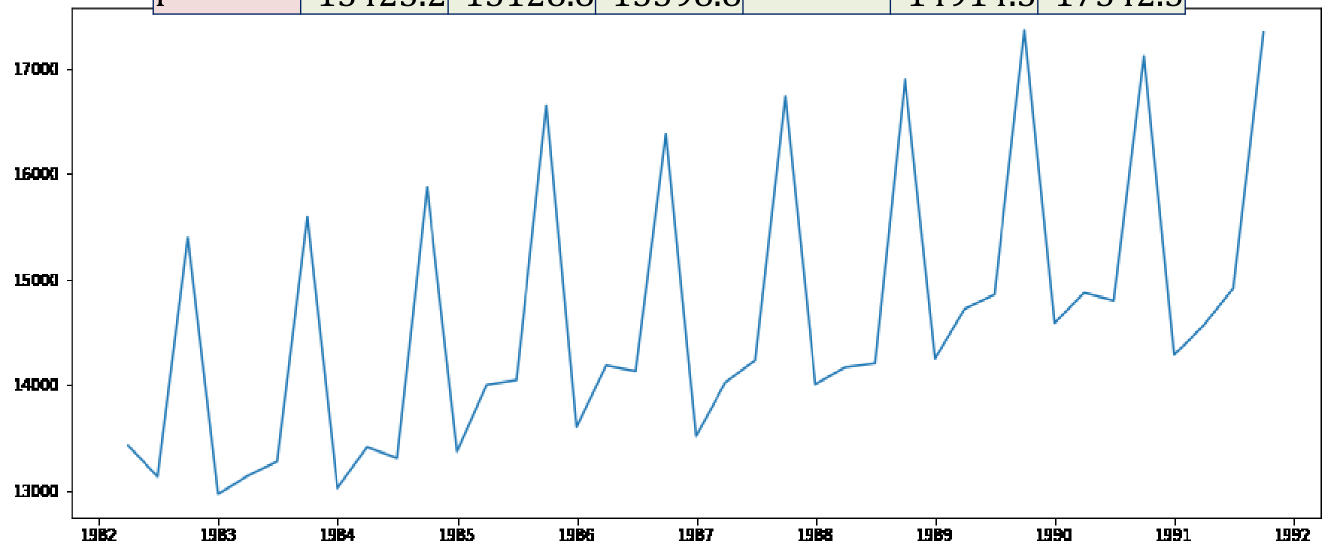
Yearly Time Series: US GDP

Year	1960	1961	1962	1963	.....	2015	2016	2017
GDP per capita	81.2848	84.4264	88.9149	100.049	……	1606.04	1717.47	1939.61



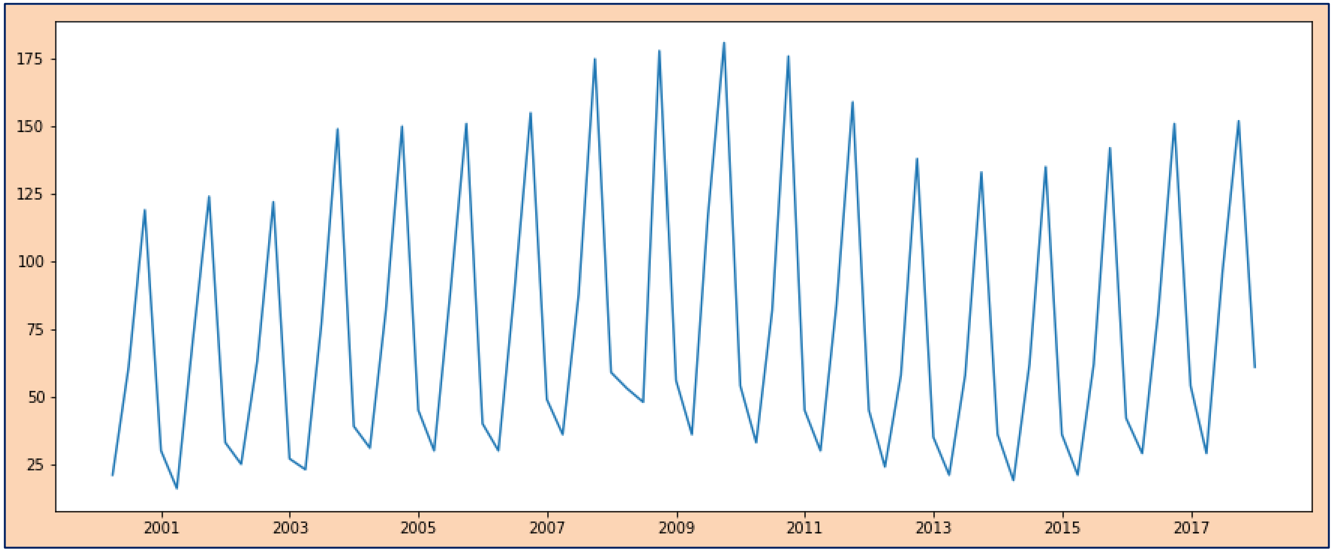
Yearly Time Series: Per Capita GDP India

Year	1982	1982	1983	...	1991	1992
Quarter	Q3	Q4	Q1	...	Q4	Q1
Turnover	13423.2	13128.8	15398.8	...	14914.3	17342.3

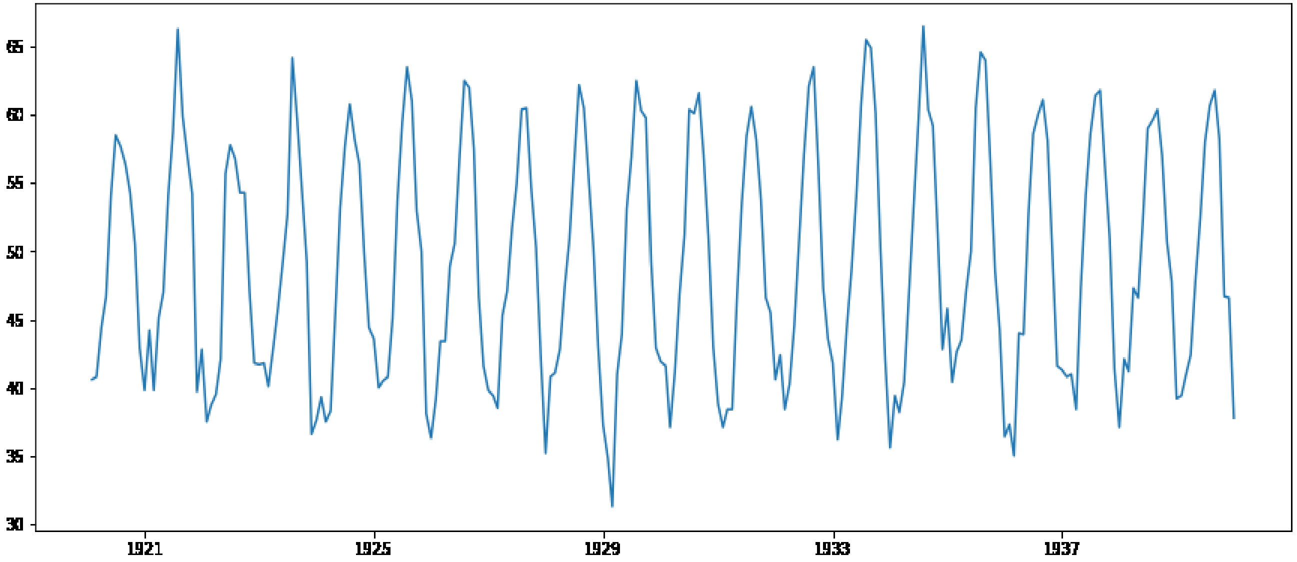


Quarterly Time Series: Retail Turnover

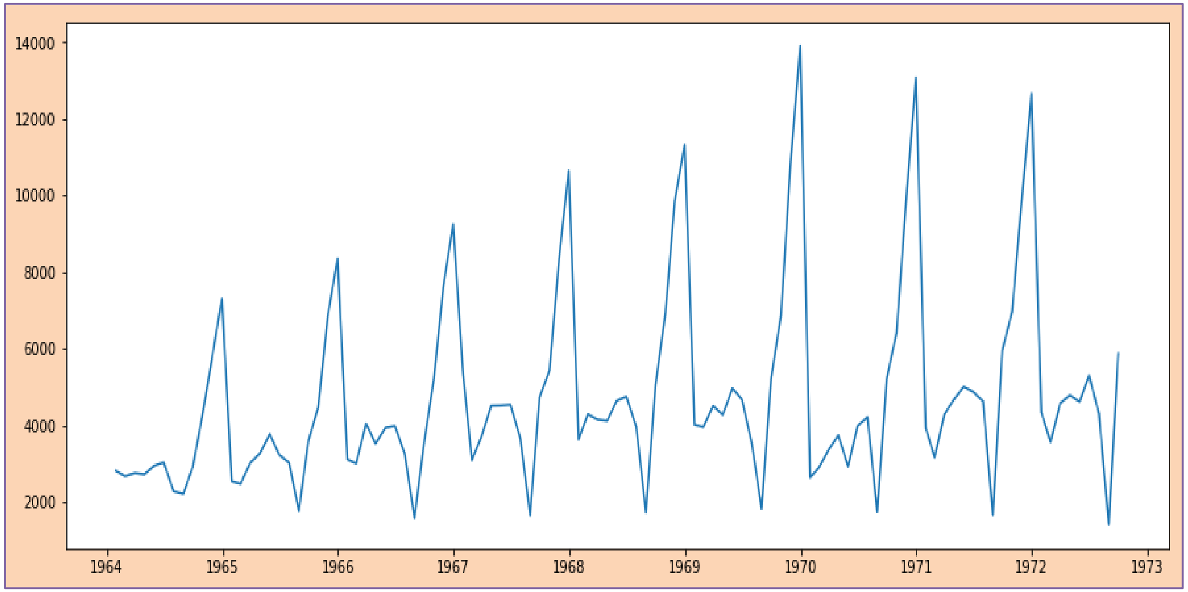
Year	2000	2000	2000	.....	2017	2017	2017
Quarter	Q1	Q2	Q3	.....	Q2	Q3	Q4
Pax	21	61	119	.....	96	152	61



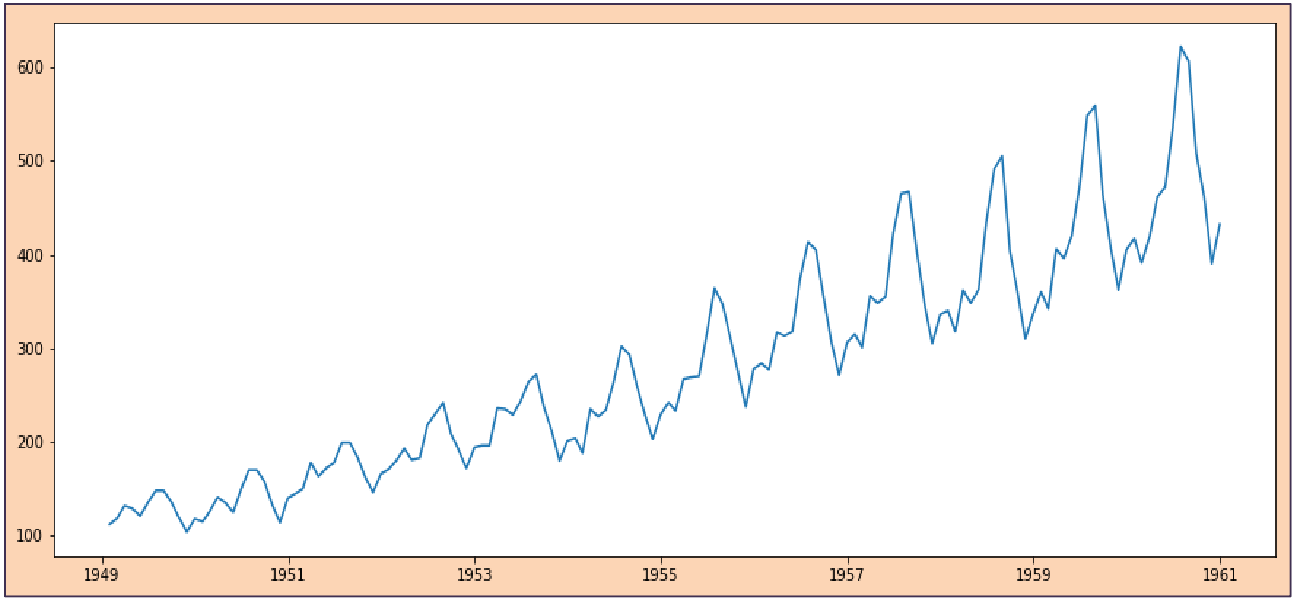
Quarterly Time Series: Portugal Port



Monthly Time Series: Average Temperature



Monthly Time Series: Champagne Sales



Monthly Time Series: Airlines Passenger

What is Time Series?

Weekly TS: Price of petrol or diesel

Daily time series –

- daily closing price or daily closing volume of a certain stock
- Sensex value
- Total daily transaction volume of your nearest ATM machine

All are examples of time series

What is **not** a Time Series?

Data collected on multiple items at the same point of time is not a time series!

Example: DowJones average on a single day

Date	MMM	T	AXP	BA	CAT	CVX	CSCO	KO	DD	XOM
1/9/2012	0.0046	-0.0094	0.0037	0.0042	0.0332	-0.008	0.0183	-0.0173	-0.0017	-0.0058
	GE	GS	HD	INTC	IBM	JPM	JNJ	MCD	MRK	MSFT
1/9/2012	0.0272	-0.007	0.0259	0.0379	-0.0253	0.0164	-0.0144	0.0081	0.0023	0.0362
	NKE	PFE	PG	TRV	UTX	UNH	VZ	V	WMT	DIS
1/9/2012	0.0159	-0.0068	-0.0028	0.0119	-0.0075	0.0239	-0.0217	-0.03	-0.0191	0.0376

What is **not** a Time Series?

When time periods are not the same: For example in a single time series both yearly and quarterly data cannot be mixed

Where do we encounter TS Data?

Companies need to evaluate their manpower requirements from historic data and take a decision regarding hiring

Portfolio managers try to understand stock movements based on past data so that they can be more effective in advising their clients how best to invest

Movement of the demand of electricity consumption pattern provides policy makers impetus on where to build the new electricity production plant and when

Where do we encounter TS Data?

Based on the demands of airline tickets between cities, airlines create their dynamic ticket pricing

Based on past data on booking pattern hotels decide on whether any discounts are to be offered in room pricing at certain times of the year

In short, time series data is being collected and utilized in all data driven decision mechanisms

What are the special features of TS Data?

The most important feature that make TS analysis challenging and none of the other machine learning techniques applicable is because

- Data are not independent
- One defining characteristic of time series is that this is a list of observations where the ordering matters.
- Ordering is very important because there is dependency and changing the order will change the data structure

If data is cross-sectional

Make	Year of Make	MPG	Horsepower	Weight
AMC	80	24.3	90	3003
Audi	80	34.3	78	2188
Buick	81	22.4	110	3415
Chevy	82	27	90	2950
Chrysler	82	26	92	2585
Datsun	81	32.9	100	2615

Order of the Observations Does not matter

Make	Year of Make	MPG	Horsepower	Weight
Datsun	81	32.9	100	2615
Audi	80	34.3	78	2188
Chevy	82	27	90	2950
Chrysler	82	26	92	2585
AMC	80	24.3	90	3003
Buick	81	22.4	110	3415

If data is time series

greatlearning

Year	Quarter	Turnover
1982	Q3	13423.2
1982	Q4	13128.8
1983	Q1	15398.8
1983	Q2	12964.2
1983	Q3	13133.5
1983	Q4	13271.7
1984	Q1	15596.3

Order of the Observations is All Important

WRONG

Year	Quarter	Turnover
1982	Q3	13423.2
1983	Q1	15398.8
1984	Q1	15596.3
1983	Q2	12964.2
1983	Q3	13133.5
1983	Q4	13271.7
1982	Q4	13128.8

Objective of this Module

To learn to forecast!

Not all series is equally easy or difficult to forecast. It depends on

- How well the contributing factors are understood
- How much data is available

Year	Quarter	Turnover
1982	Q3	13423.2
1982	Q4	13128.8
1983	Q1	15398.8
1983	Q2	12964.2
1983	Q3	13133.5
1983	Q4	13271.7
1984	Q1	15596.3
1984	Q2	13018
1984	Q3	13409.3
1984	Q4	13304.2

We should not try to forecast if the historical data available is for a short duration

Will not provide reliable forecast for
Next 8 quarters

Might get a working forecast for the
Next 3 – 4 quarters

Missing Data

- Time Series does not admit missing data
- All data observations must be contiguous
- Impute missing data *to the best of your knowledge*

Exploratory analysis

First step: Plot the time series.

Graphs enable many features of the data to be visualized, including patterns, unusual observations, changes over time, and relationships between variables, if more than one series is considered

Appropriate graph captures the inherent features of time series

What do graphs reveal?

These observations give clue to inherent features of the time series, known as components of time series

Components of Time Series

Graphs highlight variety of patterns inherent to TS

A TS can be split into several components, each representing one of the underlying categories of patterns,

Time Series Components

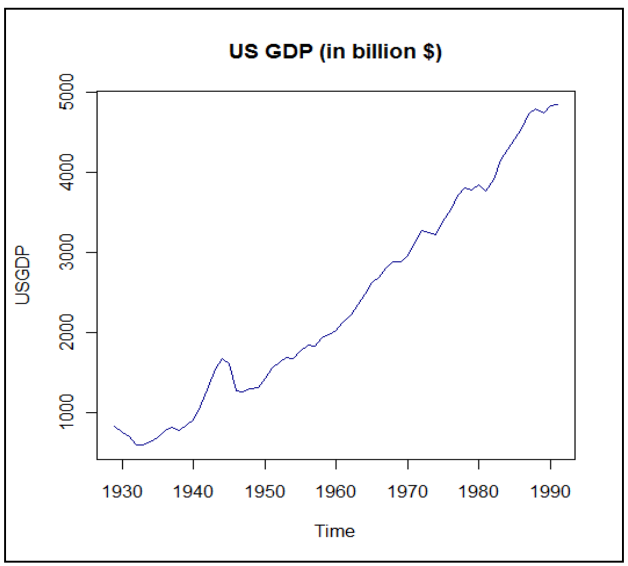
- Trend
- Seasonal component

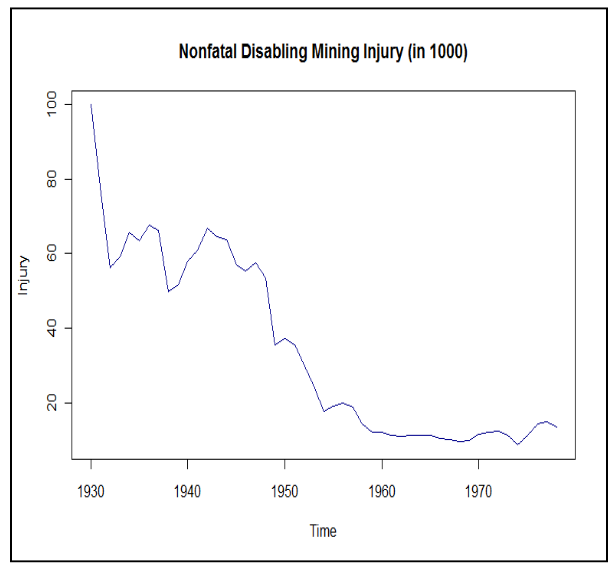
Systematic Component

- Irregular component (Error or Random Component)

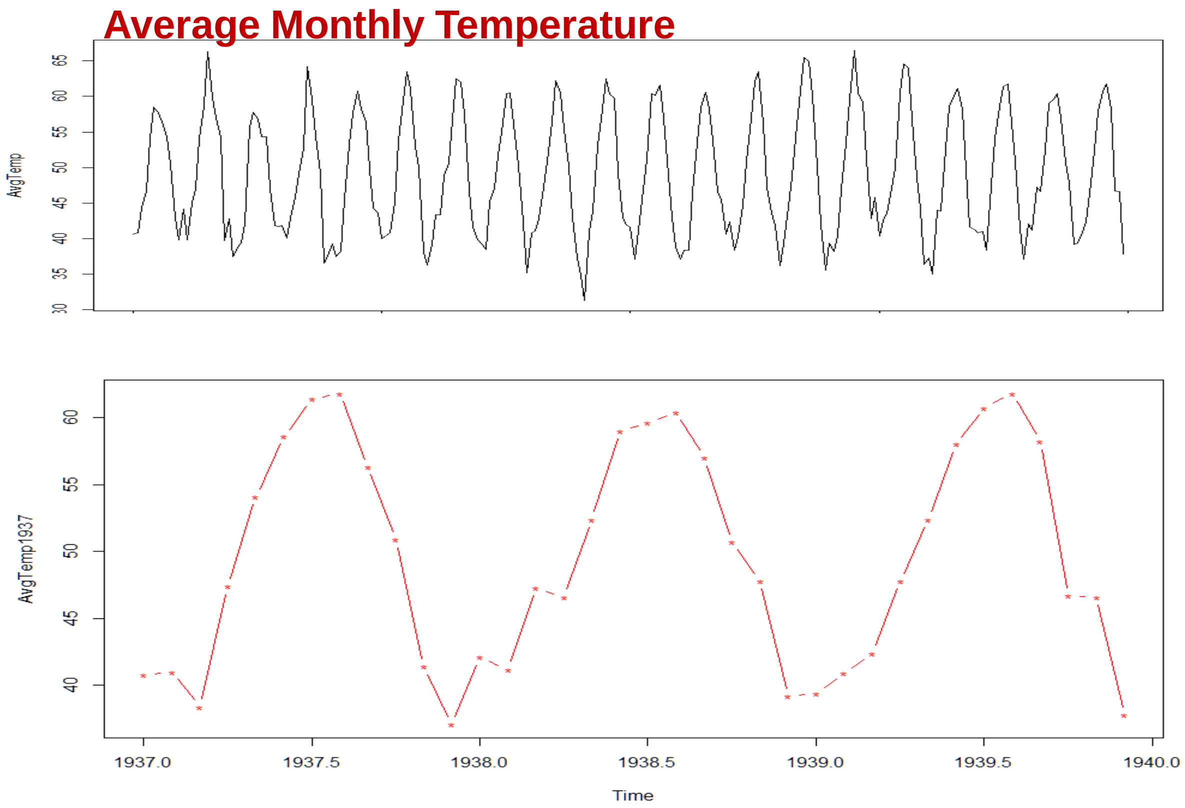
Trend

- Long term movement of a series: either increasing or decreasing

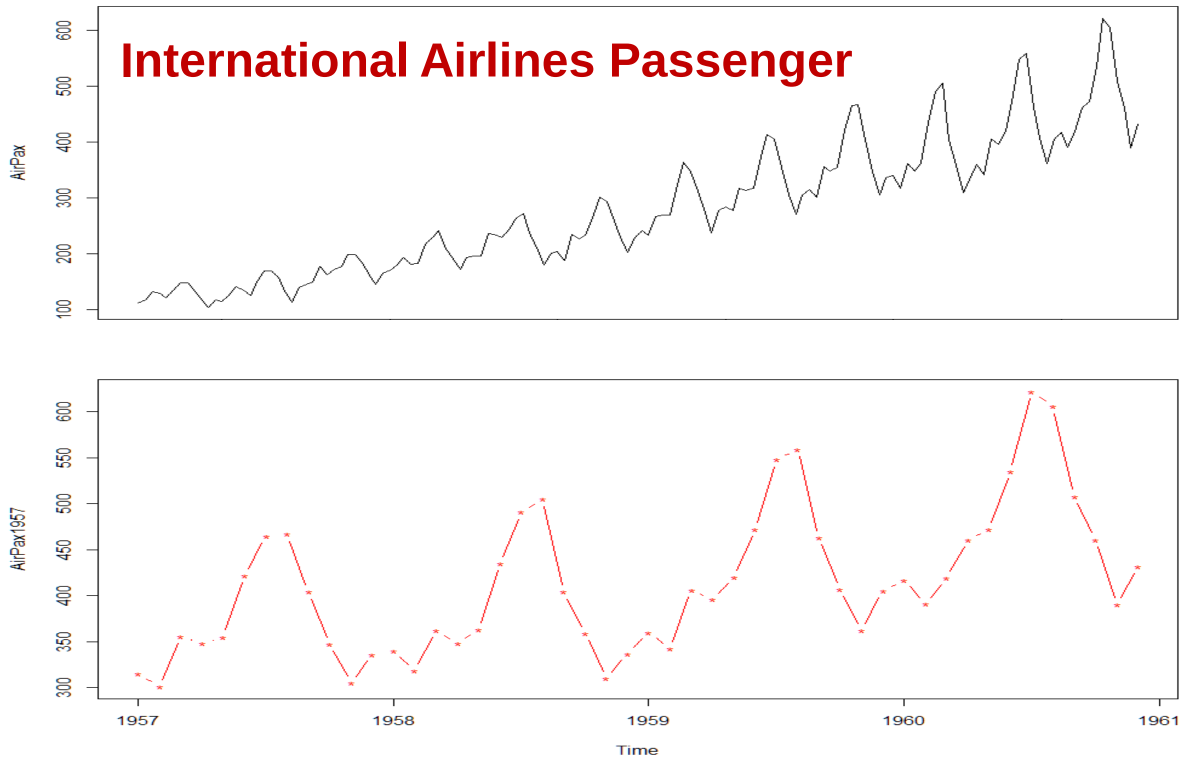




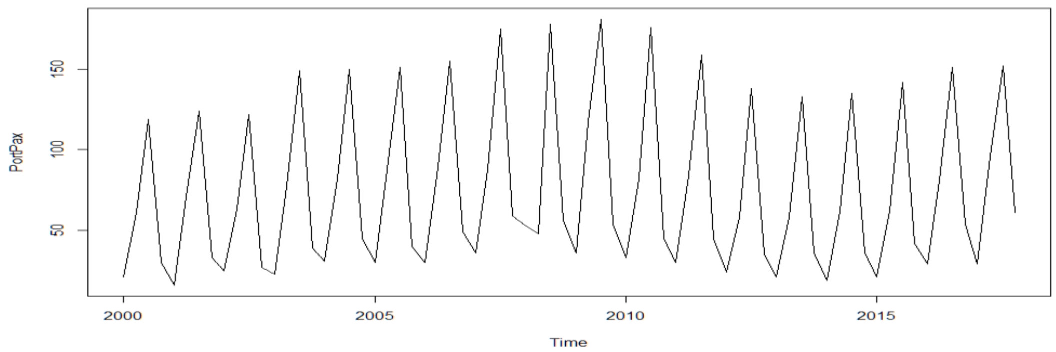
Seasonality

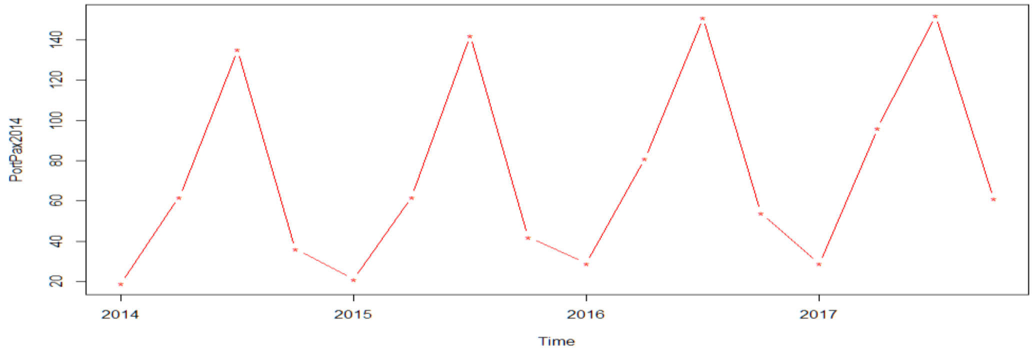


Seasonality



Seasonality





Seasonality

- Representing **intra-year stable fluctuations repeatable** year after year with respect to timing, direction and magnitude
  - Normal variations that recur every year to the same extent
  - A Yearly series does not have seasonality

What is Seasonality?

Seasonality is the relative increase or decrease of sales (demand or consumption) every period (quarter or month) compared to the yearly average

Heuristic example with 4 quarters

Yearly sale = 400 units

Quarterly average = 100 units

Actual sales

Q1 = 80 units Q2 = 70 units Q3 = 200 units Q4 = 50 units

Seasonality estimate

Q1 = -20 Q2= -30 Q3 = +100 Q4 = - 50

Seasonality

- Demand for winter clothes
- Airlines and train ticket demands
- Incidence of influenza or other vector-borne diseases

Stock prices typically will not show any seasonal pattern

Graphically identify Important Characteristics

- Time Series plots are the first step in understanding the pattern of the data

- Not only it identifies whether there are trend, seasonality or cyclicality, it also identifies
  - Which historical horizon to include for forecasting
    - Is there any abrupt change in the level of the series?
  - Whether there are any unusual observations in the series
    - sudden spikes or sudden drops!

Systematic Components

- Trend and Seasonality are part of systematic component
- These patterns are interpretable
- These can be estimated
- Forecast of time series involves estimation and extrapolation of these components

Irregular Component

The error or variability associated with the series is the Irregular component

This component is a random component

The part of the series that cannot be explained through Systematic component forms the Irregular Component

Other names of this component is Error or White Noise

This component is assumed to have a normal distribution with 0 mean and constant variance σ^2

Decomposition

Why Decompose

- To understand revenue generation without the quarterly effects
  - De-seasonalize the series
  - Estimate and adjust by seasonality

- Compare the long-term movement of the series (Trend) vis-a-vis short-term movement (seasonality) to understand which has the higher influence

- If revenue for multiple sector are to be compared and if the sectors show non-uniform seasonality, de-seasonalized series needs to be compared

Decomposition Model

Y_t : time series value (actual data) at period t.

S_t : seasonal component (index) at period t.

T_t : trend cycle component at period t.

I_t : irregular (remainder) component at period t

Additive model: Observation = Trend + Seasonality + Error

$$Y_t = T_t + S_t + I_t$$

Useful when the seasonal variation is relatively constant over time

Multiplicative model: Observation = Trend * Seasonality * Error

$$Y_t = T_t * S_t * I_t$$

Example: Average Monthly Temp

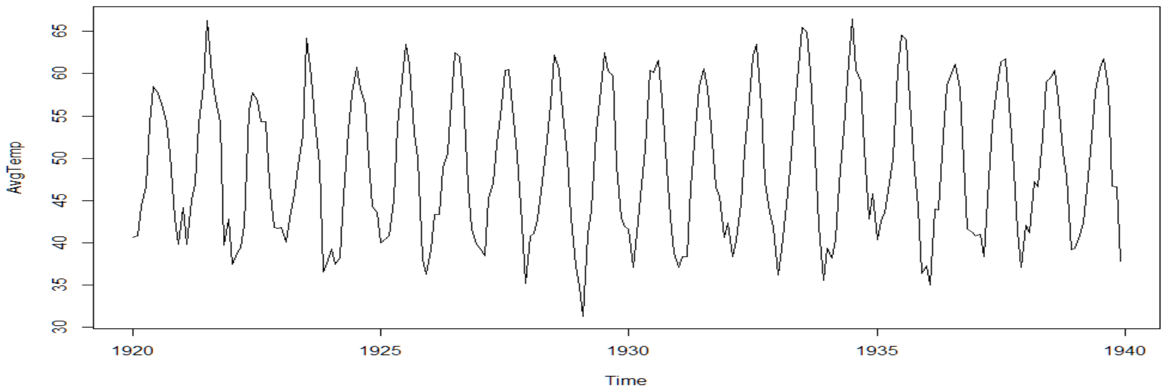
- Monthly temp is a sum of Trend, Seasonality and Irregular component
- Would like to understand relative effects of the 12 months
- Would like to understand whether there is at all any movement of the temp series after the seasonal fluctuations are eliminated
- Example of an Additive Seasonality Model

$$Y_t = T_t + S_t + I_t$$

Sales = Trend + Seasonality + Error

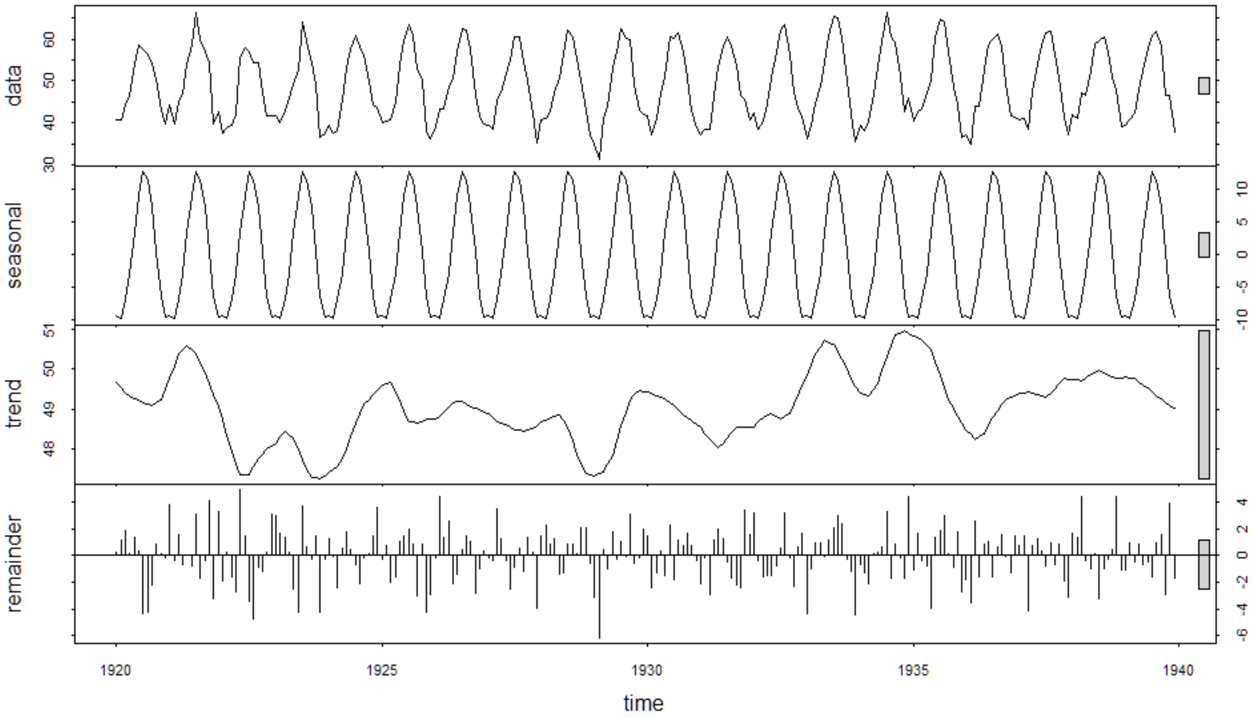
Example: Average Monthly Temp

Average Monthly Temperature



Additive Seasonality : If seasonal fluctuations do not change with trend

Example: Average Monthly Temp

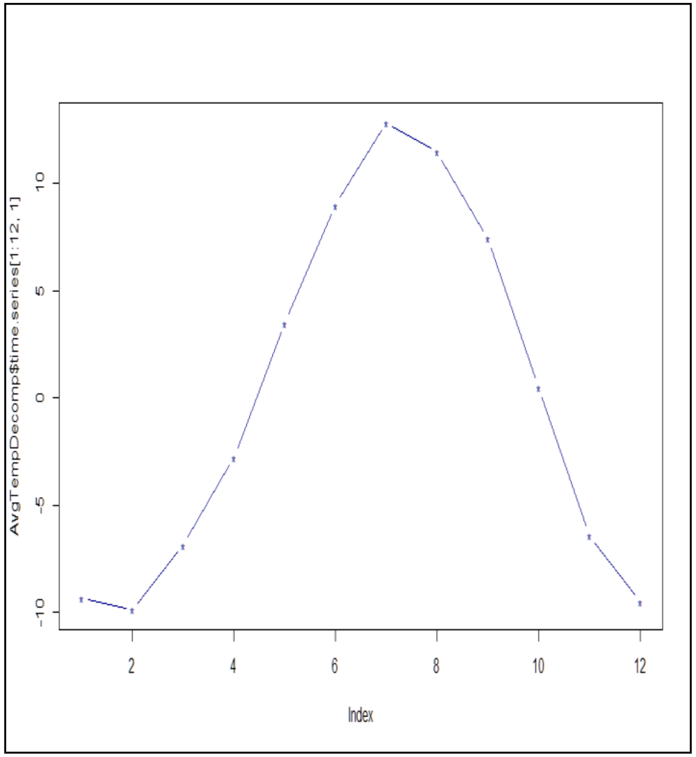


Example: Average Monthly Temp

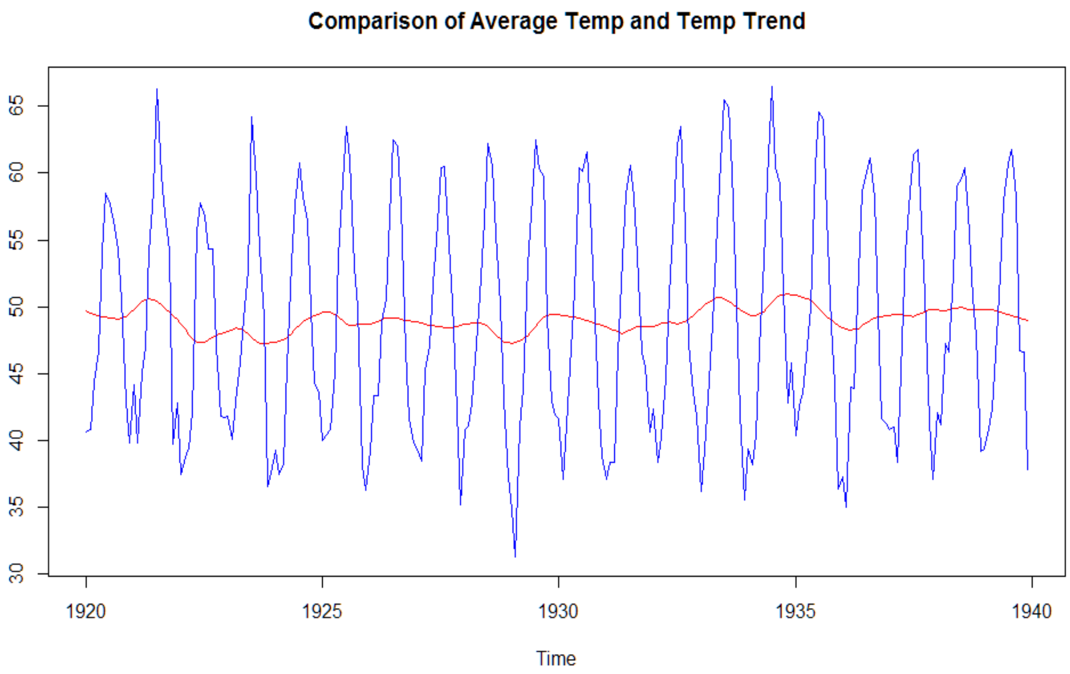
- Constant seasonality
- Trend contribution minimal

Month	Seasonal Index	Month	Seasonal Index
Jan	-9.33	Jul	12.84
Feb	-9.86	Aug	11.47
Mar	-6.85	Sep	7.44
Apr	-2.77	Oct	0.47
May	3.50	Nov	-6.42
Jun	8.98	Dec	-9.49

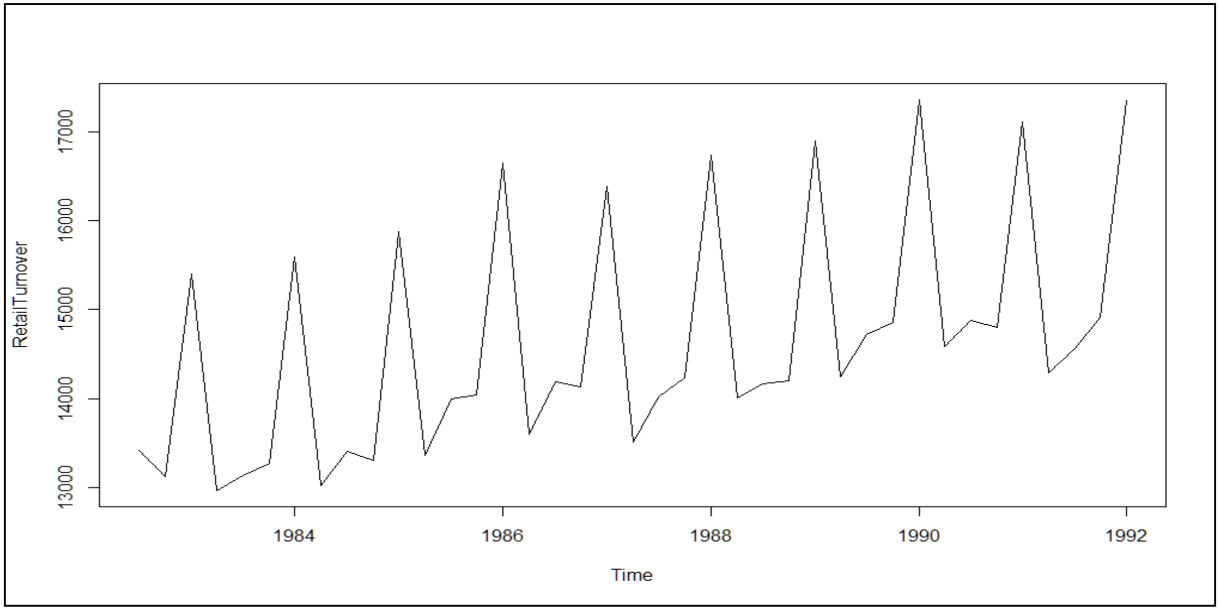
Sum of additive seasonal indices = 0



Example: Average Monthly Temp

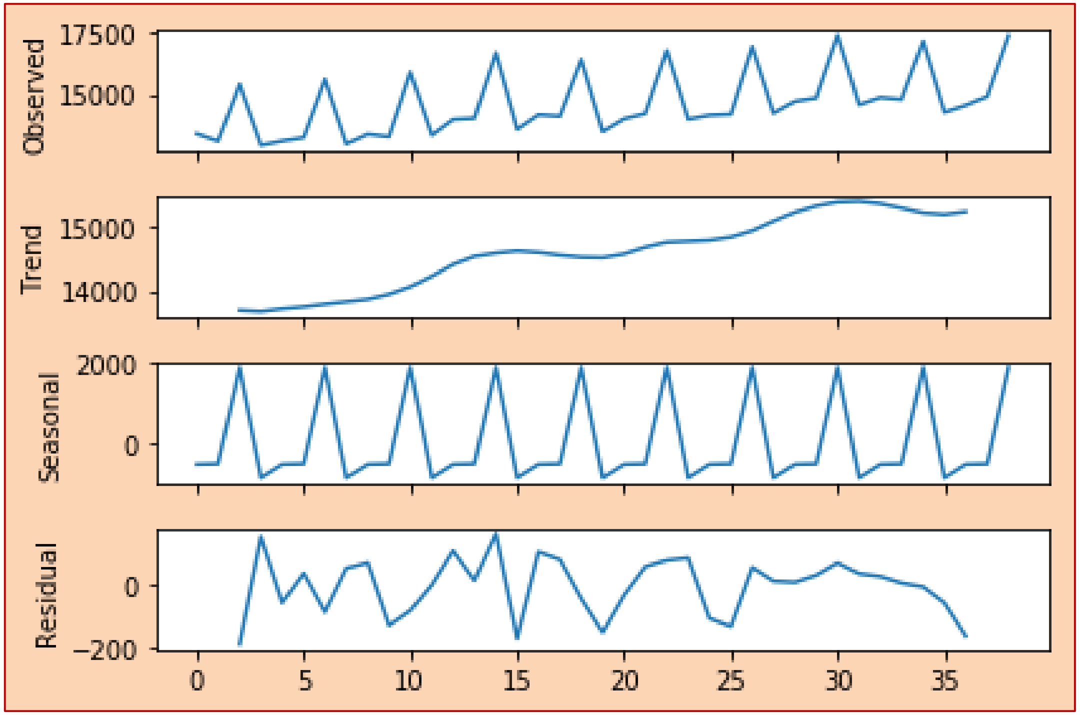


Example: Quarterly Turnover



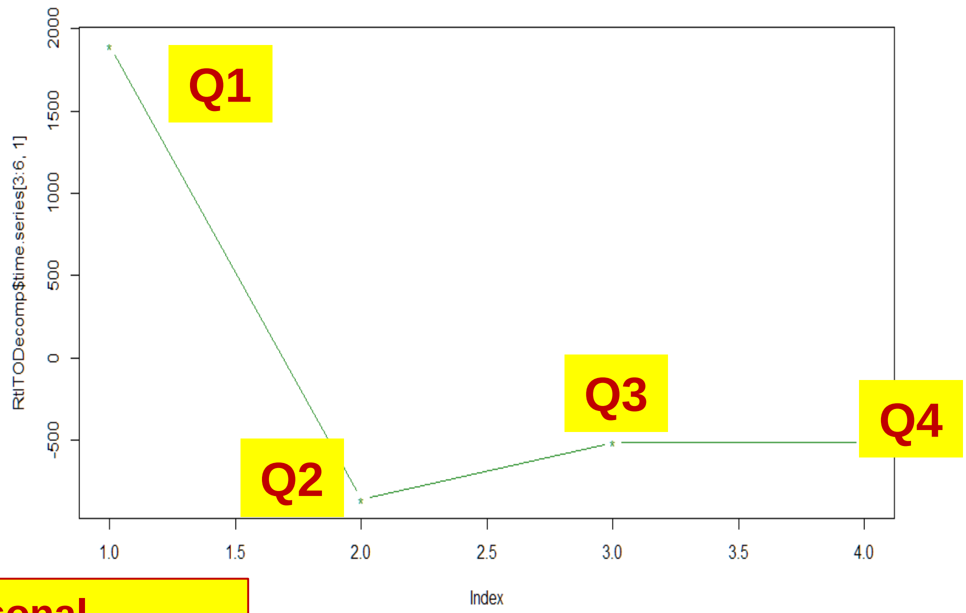
Additive Seasonality : Seasonal fluctuations do not change with trend

Example: Quarterly Turnover



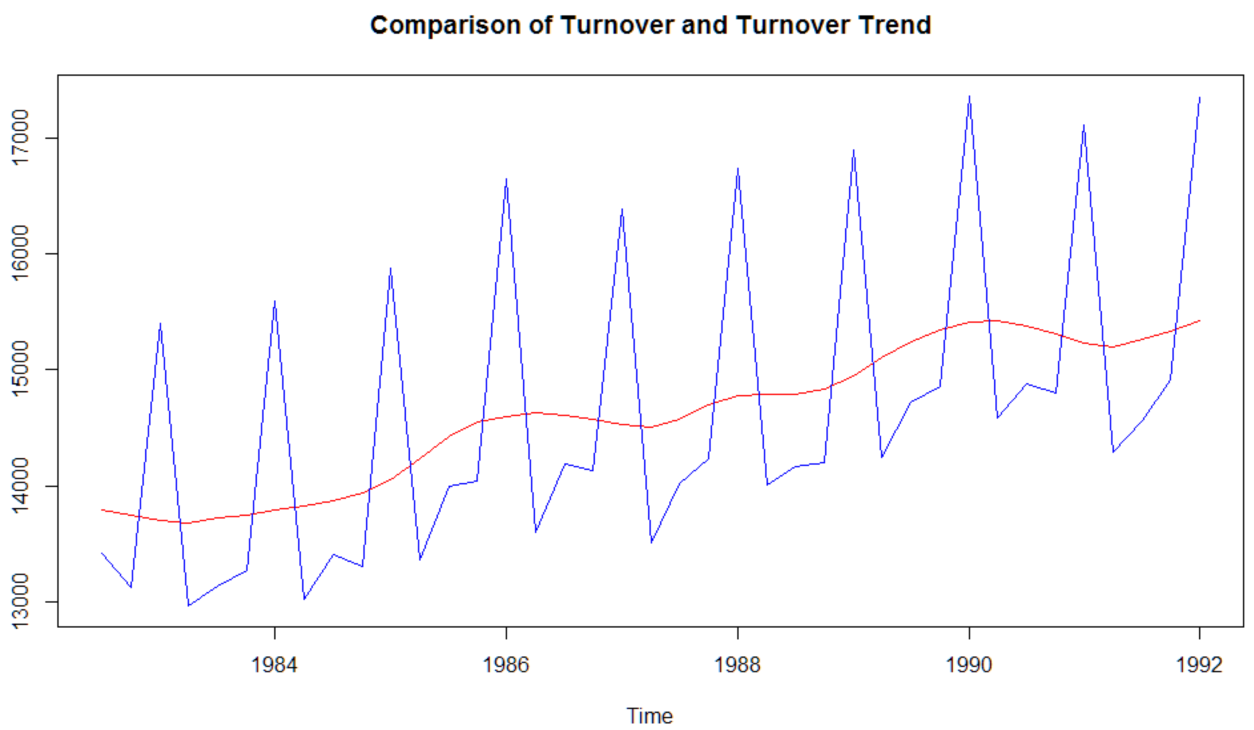
Example: Quarterly Turnover

Quarter	Seasonal Index
Q3	**-518.03**
Q4	**-515.26**
Q1	**1896.48**
Q2	**-863.19**



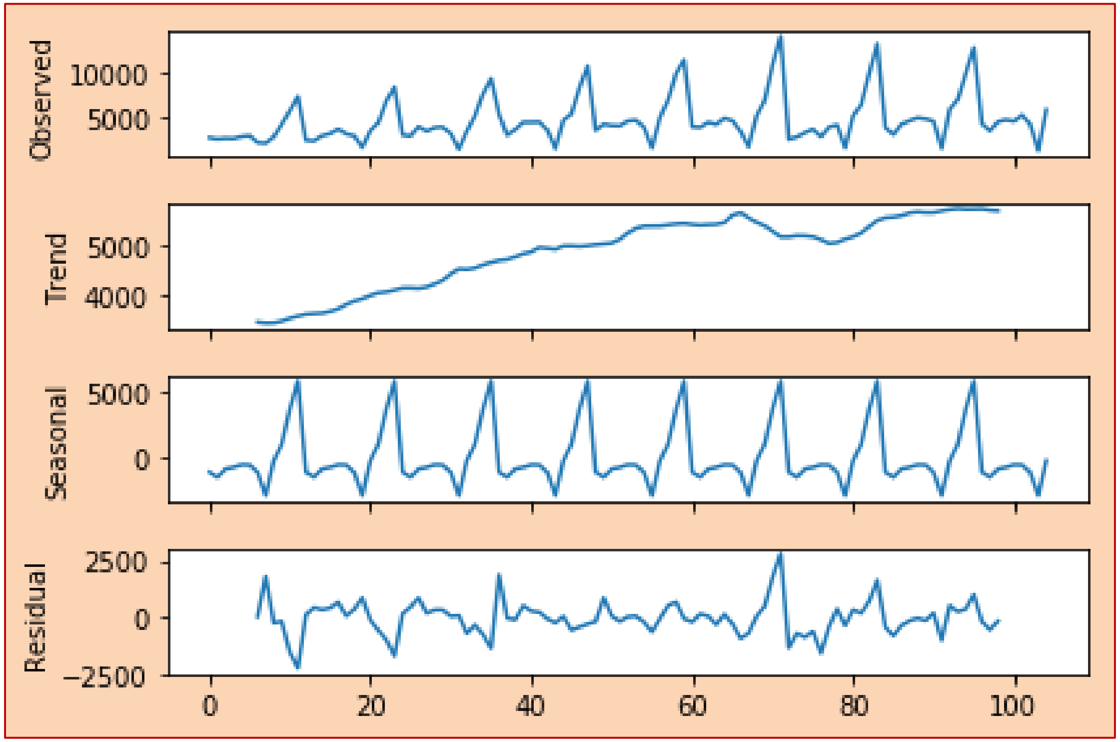
Sum of additive seasonal indices = 0

Example: Quarterly Turnover

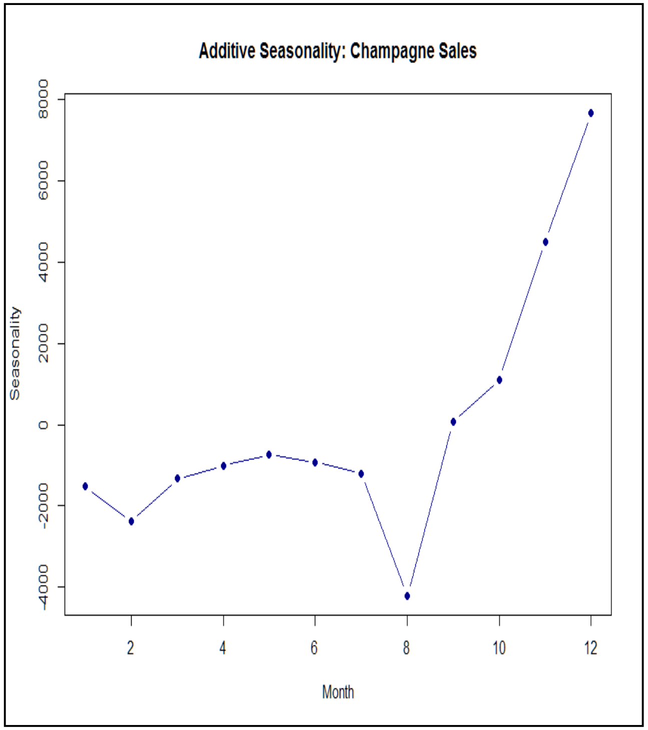


greatlearning

Example: Champagne Sales



Example: Champagne Sales



Critical look at seasonality

- During first part of the year almost no change
- Sharp drop in sales in August
- Last 4 months show steep increase in sales

Example: Passenger Volume

- There is a definite upward movement YOY
- Seasonal fluctuations increasing as total volume increases
- Example of an **Multiplicative** Seasonality Model

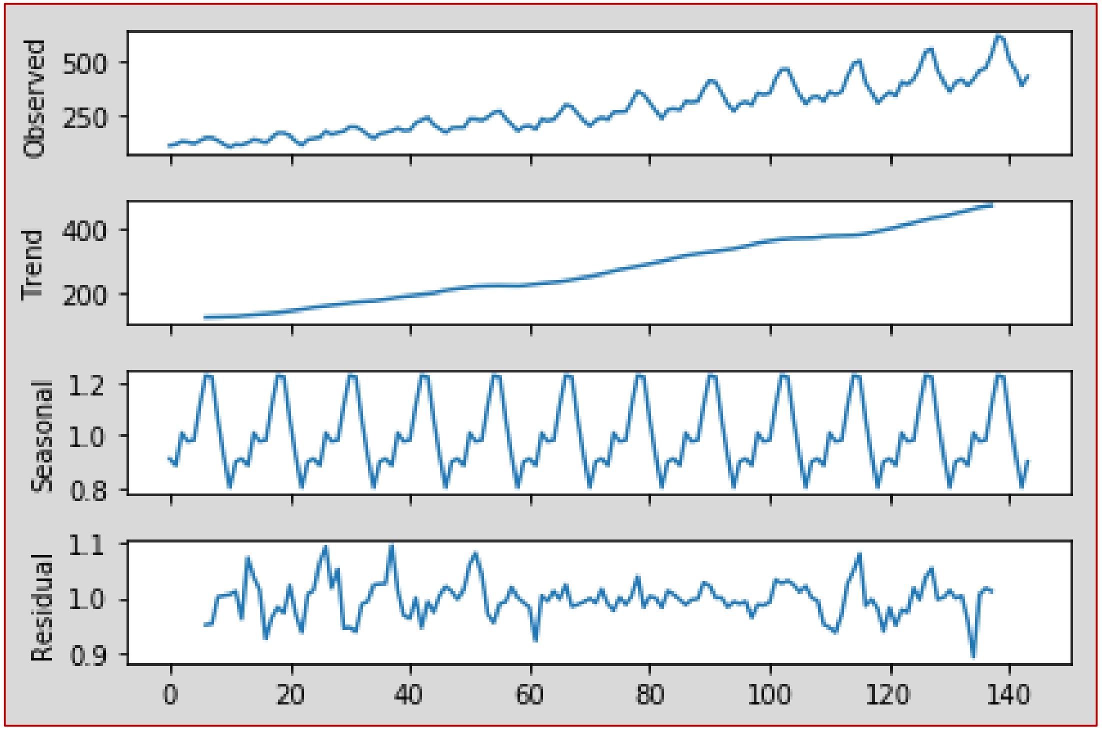
$$Y_t = T_t * S_t * I_t$$

Volume= Trend *Seasonality * Error

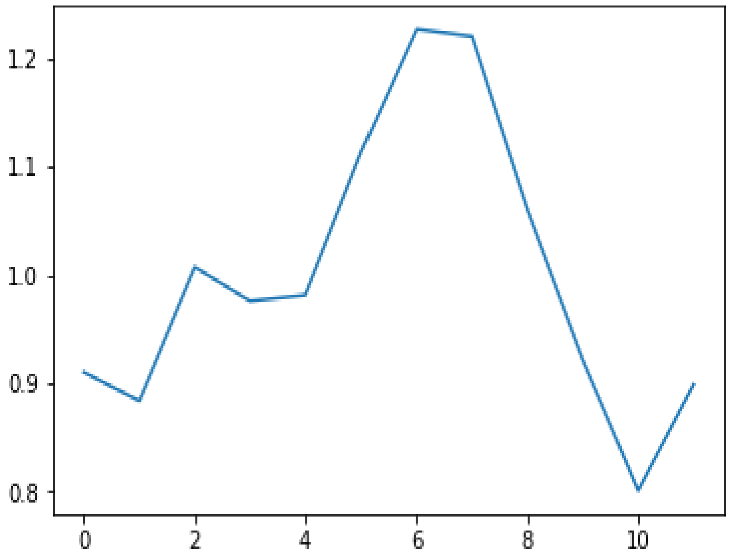
- Need logarithmic transformation to convert into an additive series

$$\text{Log(Vol)} = \log(\text{Trend}) + \log(\text{Seasonality}) + \log(\text{Error})$$

Example: Passenger Volume



Passenger Volume



Critical look at seasonality

- From Feb passenger volume starts increasing
- Jun – Sep shows high volume
- Jul – Aug has highest volume
- Dec shows slight increase

HAPPY LEARNING