Time Series Forecasting

Agenda

- Forecasting
  - Why forecast?
  - How to forecast?
  - What is time series?

Agenda

- Regression
- Time series
  - Examples of time series
  - Features of time series
  - Components of time series
  - Decomposition of time series

What is forecasting?

Forecasting

Forecasting is an important problem for many fields like economics, businesses, environmental sciences, medicine, social science and politics.

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

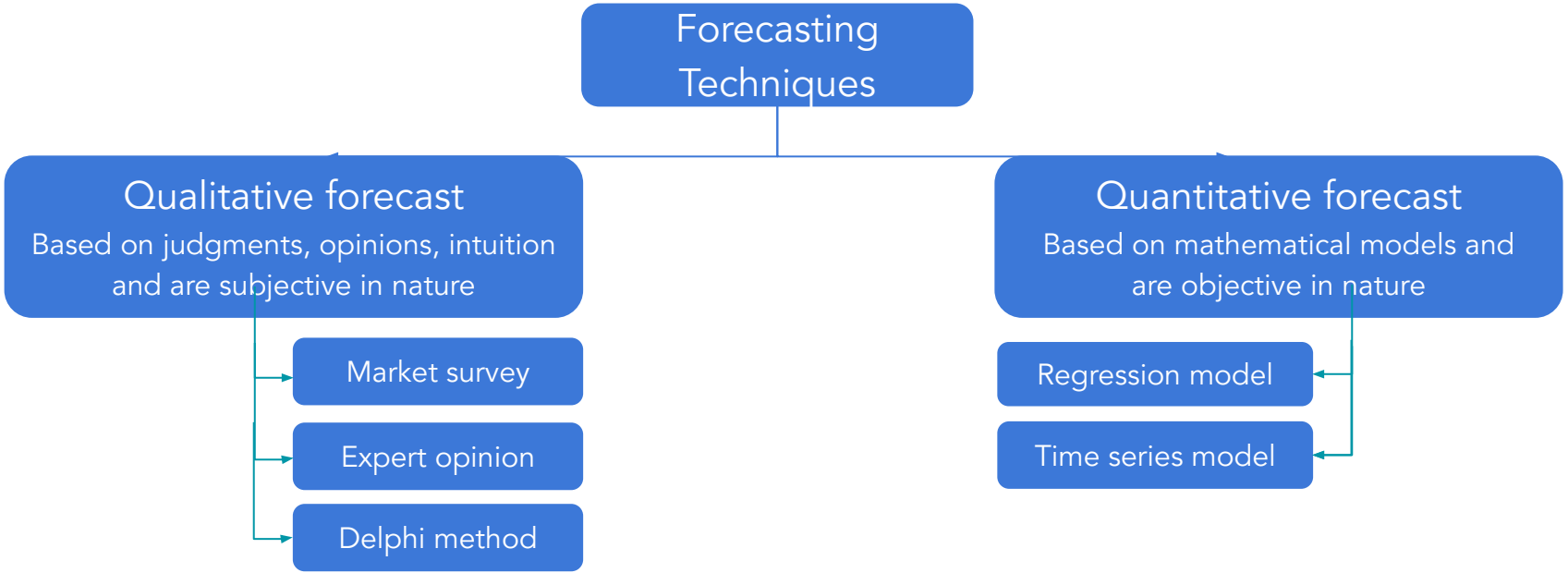
Why forecast?

Forecasting of future events is an important input for planning and decision making processes in the business. Forecasting helps businesses to solve the problems encountered in:

- Marketing
- Production planning and inventory control
- New product launching
- Finance and risk management

How to forecast?

- Forecasting techniques can be broadly classified into two types



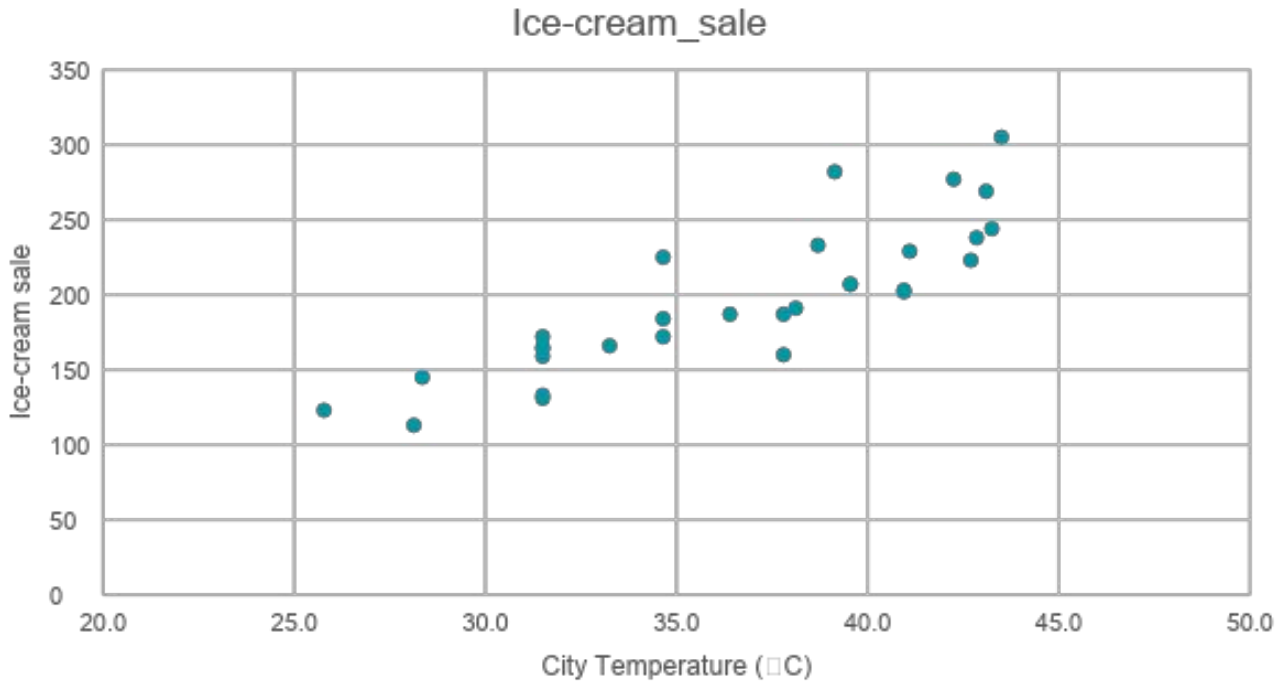
- Consider a one month sample data for Ice-cream store.

Date	Temperature(°C)	Sale
01 July 2016	31.5	164
02 July 2016	31.5	165
03 July 2016	36.4	187
04 July 2016	34.3	233
05 July 2016	42.2	277
06 July 2016	31.5	172
07 July 2016	43.2	244
09 July 2016	41.1	229
10 July 2016	42.8	238
…….	…….	…….
30 July 2016	28.3	145
31 July 2016	23.8	123

- **Correlation:-** Correlation is used to analyze the strength of a relationship between two variables. It not only shows the kind of relation in terms of direction but also how strong the relationship is.

$$\text{correlation } (x, y) = \frac{\text{Cov}(x, y)}{S_x * S_y}$$

- The correlation coefficient is a dimensionless metric and its value ranges from -1 to +1. The closer it is to +1 or -1, the more closely the two variables are related.


Ice-cream_sale
Ice-cream sale
350
300
250
200
150
100
50
0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
City Temperature (□C)

Regression

Regression

- Regression analysis is a statistical technique for modeling the relationships between a response variable and one or more predictor variables.

- The simple linear regression model involves a single predictor variable and is written as

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where y is the response, x is the predictor variable, β_0 and β_1 are unknown parameters, and ε is an error term.

Regression

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- Parameters β_0 and β_1 are typically unknown and must be estimated from a sample of data using the method of least squares.
- The error term ε, also known as residuals, accounts for deviations of the actual data from the straight line specified by the model equation.
- The method of least squares chooses the model parameters (β) such that the sum of the squares of the errors, ε, is minimized.

Statistical inference in linear regression

- In linear regression problems, certain tests of hypotheses about the model parameters and confidence interval estimates of these parameters are helpful in measuring the usefulness of the model.

- R^2 :- It is a measure of the amount of variation in target variable y explained by the predictor variables in the model.

- In almost all cases, when a variable is added to the regression model R^2 increases or stays the same. Adding a variable to the model will never cause a decrease in R^2 , even in situations where the additional variable is not statistically significant.

Statistical inference in linear regression

- Adjusted R^2 :- The adjusted R^2 value increases when the variables added are of statistical significance. In fact, if unnecessary regressors are added, the value of the adjusted R^2 may decrease.

- If R^2 and adjusted R^2 are very similar, it is a good sign that the regression model does not contain unnecessary predictor variables.

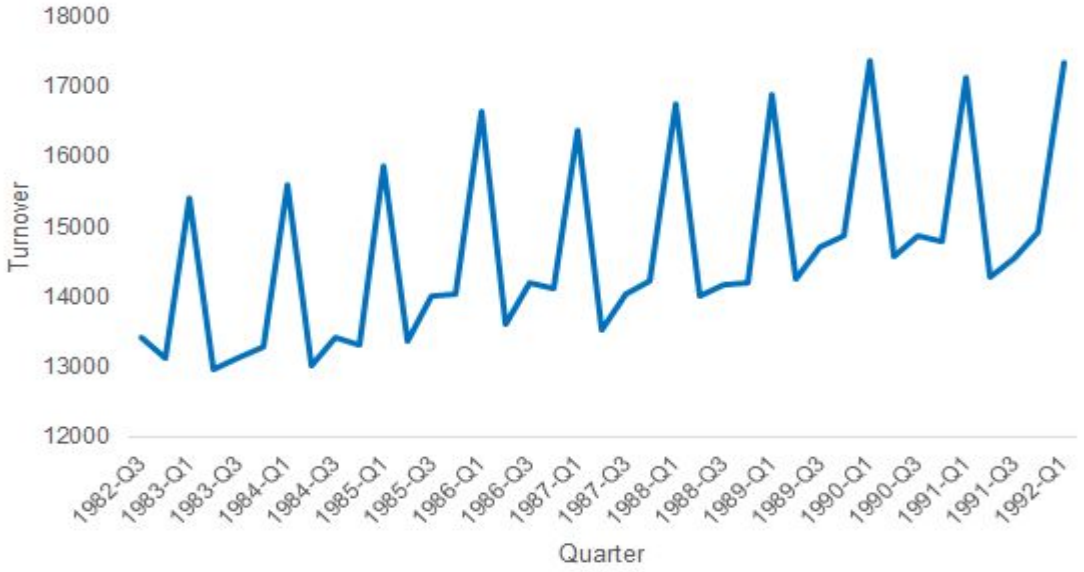
Time Series

What is time series?

- A time series is a sequence of measurements on the same variable collected over time.
- It is a set of observations, each one being recorded at equally spaced time interval.
- Following table shows quarterly data of a retail store turnover, where data is recorded at equal time intervals;

Quarter	1982-Q3	1982-Q4	1983-Q1	1983-Q2	.....	1991-Q4	1992-Q1
Turnover	13432.2	13128.8	15398.8	12964.2	.....	14914.3	17342.3

- Below is the time series plot for turnover of the retail store recorded on quarterly basis from Q3 of 1982 through Q1 of 1992.



Univariate and Multivariate time series

- Univariate time series:- Data is collected for only one variable over a period of time.
  For example, data collected from a sensor measuring the temperature of a room every minute.
- Multivariate time series:- Data is collected for multiple variables over the same period of time.
  For example, US data for Unemployment, GDP and Inflation recorder every year.

Year	1951	1952	1953	.....	2019
Unemployment	0.03	0.03	0.05	.....	0.04
GDP	0.08	0.04	0.05	.....	0.02
Inflation	0.06	0.01	0.01	.....	0.02

Time series intervals

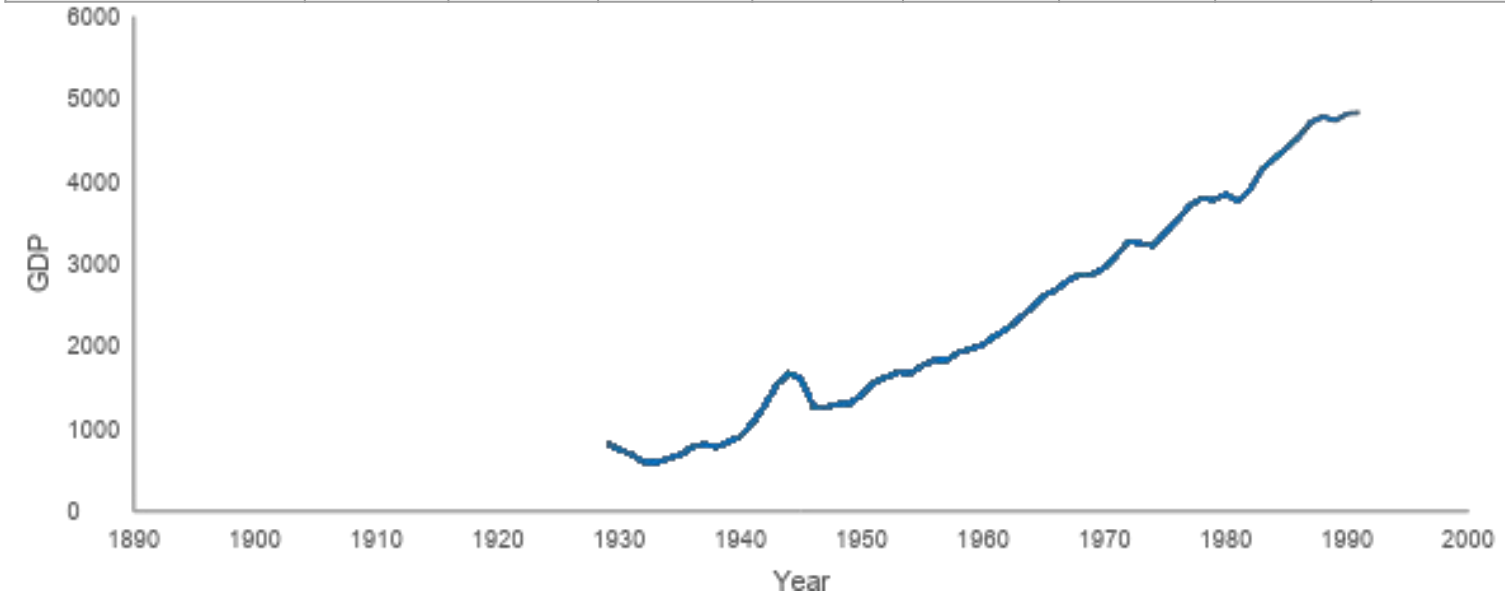
- Business applications generally utilize daily, weekly, monthly, quarterly, or annual time series data.

- Time interval depends on the nature of the forecasting;
  Yearly:- GDP, Macro-economic series
  Quarterly:- Revenue
  Monthly:- Sales, Expenditure
  Weekly:- Price of petrol/diesel
  Daily:- Closing price of stock
  Hourly:- Air Quality Index

Examples of time series

Yearly time series: US GDP

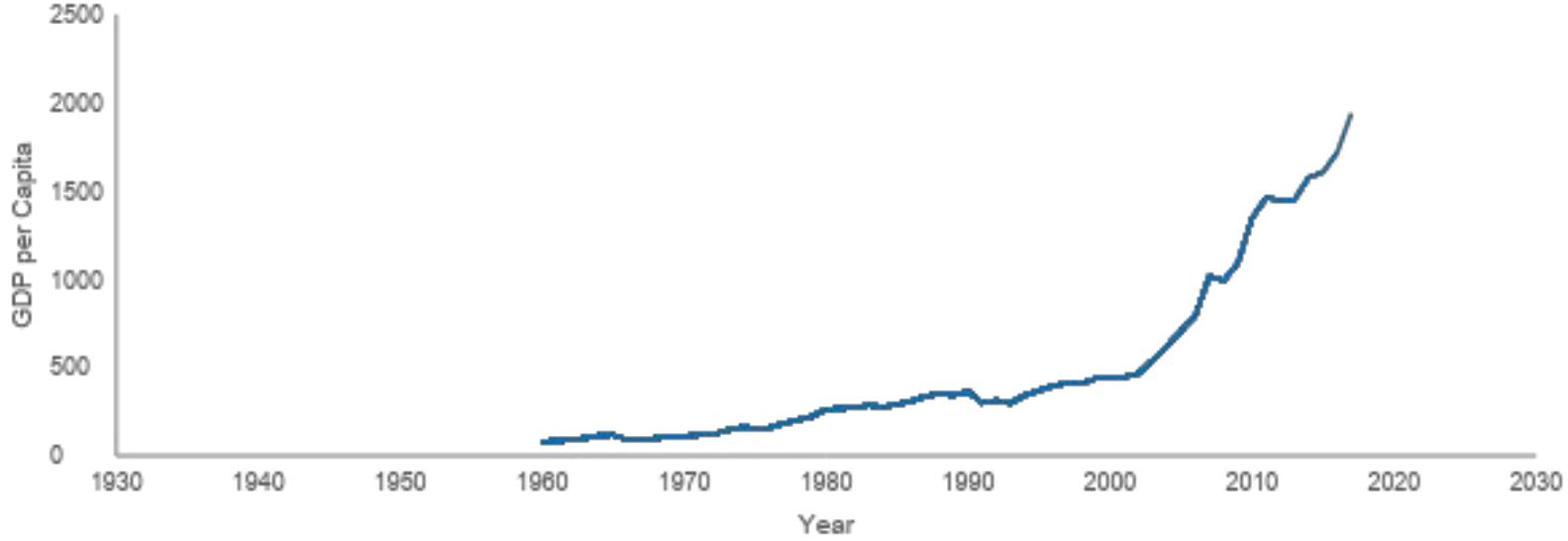
Year	1929	1930	1931	1932	……	1989	1990	1991
US GDP (b. USD)	821.8	748.9	691.3	599.7	……	4739.2	4822.3	4835



Examples of time series

Yearly time series: Per capita GDP of India

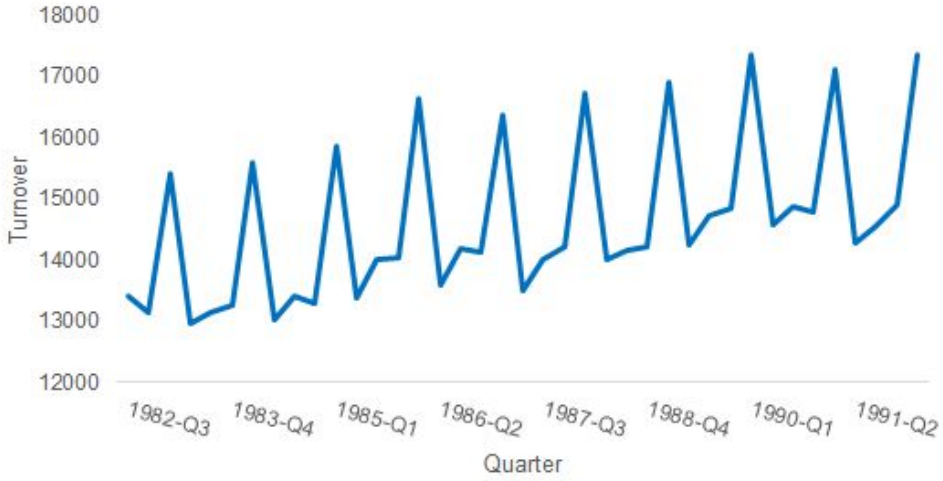
Year	1960	1961	1962	1963	…..	2015	2016	2017
GDP per Capita	81.2848	84.4264	88.9149	100.049	……	1606.04	1717.47	1939.61



Examples of time series

Quarterly Time Series: Retail Turnover

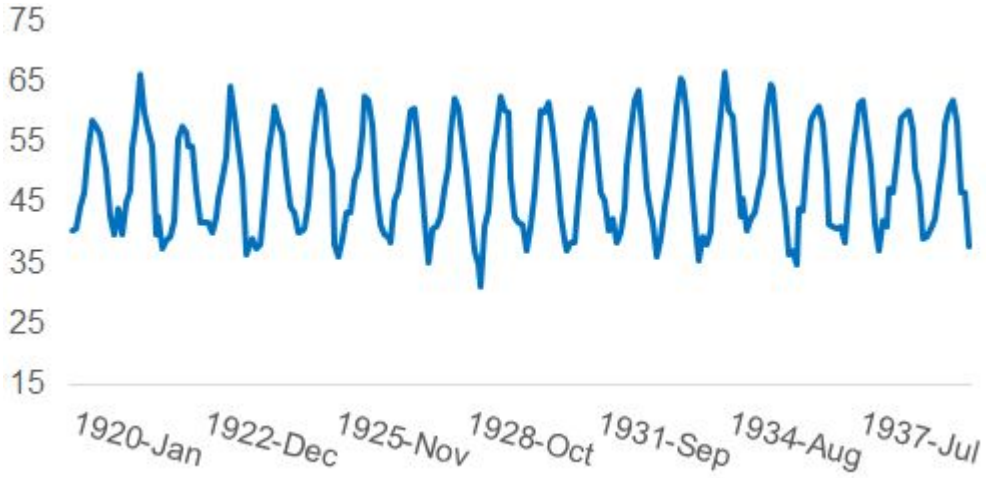
Year	1982	1982	1983	…	1991	1992
Quarter	Q3	Q4	Q1	…	Q4	Q1
Turnover	13423.2	13128.8	15398.8	…	14914.3	17342.3



Examples of time series

Monthly Time Series: Average Temperature

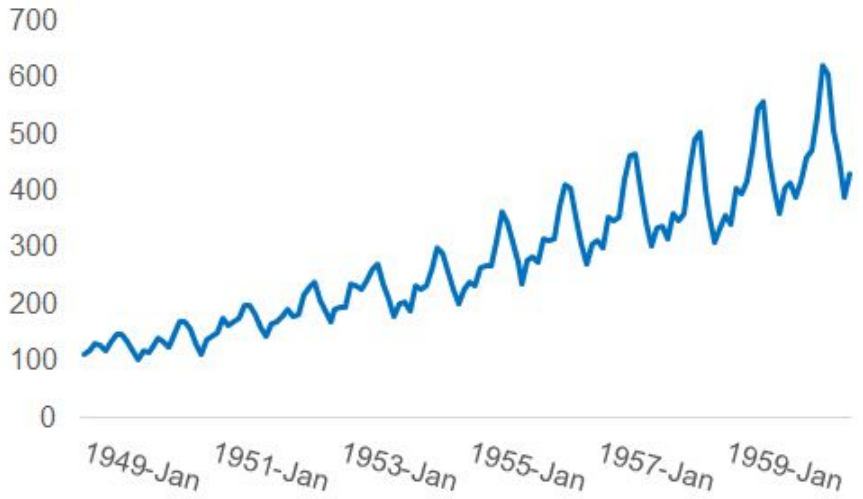
Year	1920	1920	1920	…	1939	1939
Month	Jan	Feb	Mar	…	Nov	Dec
Avg. Temp.	40.6	40.8	44.6	…	46.6	37.8



Examples of time series

Monthly Time Series: Airlines passengers

Year	1949	1949	1949	…	1960	1960
Month	Jan	Feb	Mar	…	Nov	Dec
Pax	112	118	132	…	432	390



What is not a time series?

- When the series of observations are recorded at unequal interval of time.
  Example:
  - In a single time series both yearly and quarterly data cannot be mixed.
  - Natural calamities such as floods or earthquakes typically occur at irregular time intervals. Therefore not a time series.
- When data is collected on multiple items at the same point of time.
  Example:
  - DowJones closing price recorded for different stocks on a single day is not a time series

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

Where do we encounter TS Data?

Power Sector	Movement of the demand of electricity consumption pattern provides policy makers impetus on when the demand of electricity consumption is at peak.
Finance	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.
Aviation	Based on the demands of airline tickets between cities, airlines create their dynamic ticket pricing.
Hospitality	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.
Production	For production planning and inventory control

Special features of TS Data

- The most important feature that make Time Series analysis challenging and ~~none~~ many of the other machine learning techniques non-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 the data is cross- sectional (not a time series)

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 the data is time series

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

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

Scope of this module

- To understand Time Series data and important features of time series.
- To learn different smoothing techniques for time series forecasting.
- To understand ARMA Processes for times series forecasting.
- To understand ARIMA model for nonstationary and seasonal time series.
- Forecasting for Multivariate time series.

Time series EDA

Missing values

- 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.
- Appropriate graph captures the inherent features of 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.
- 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 time series.
- A time series can be split into several components, each representing one of the underlying categories of patterns,

Time series components

- Trend
- Seasonality

Systematic component

- Irregular component (error or random component)

Systematic Component of Time Series

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

Irregular Components

- 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.

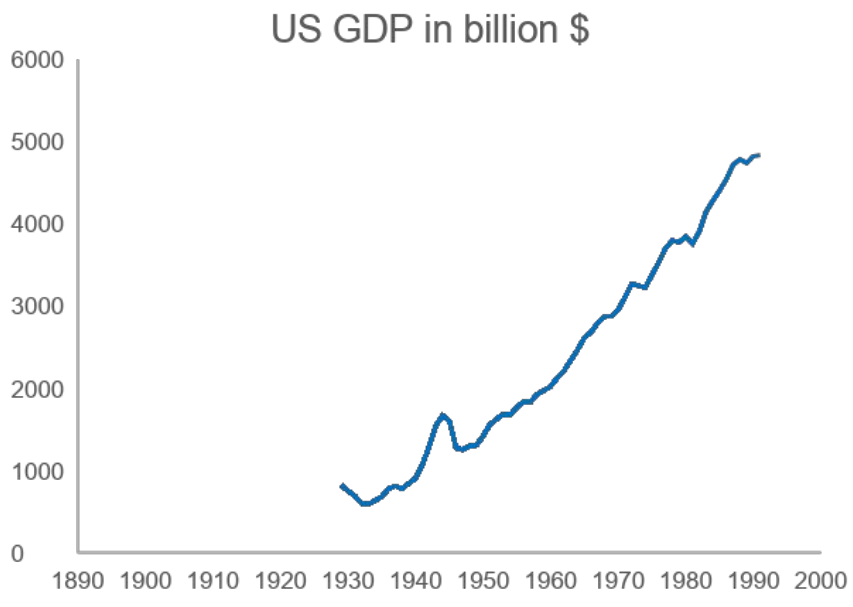
- Other names of this component is Error or White Noise.

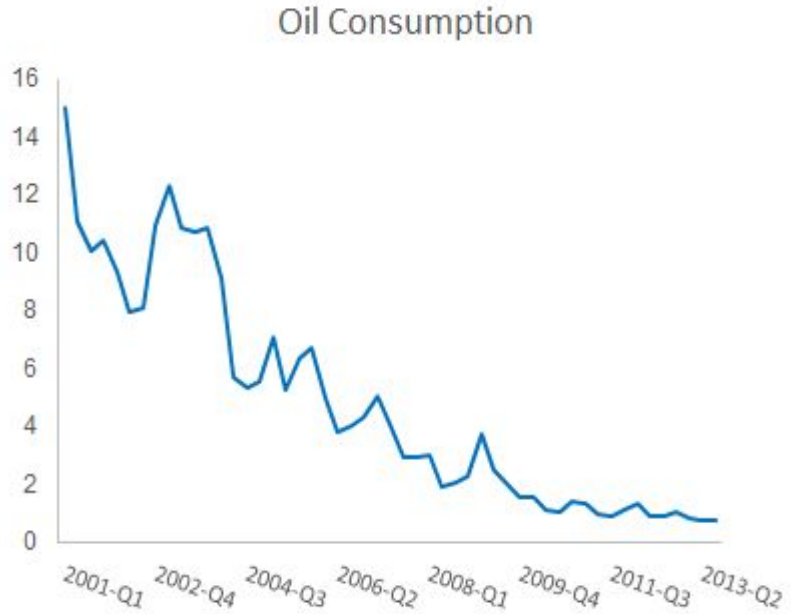
- This component is assumed to have a normal distribution with 0 mean and

  constant variance (σ^2) .

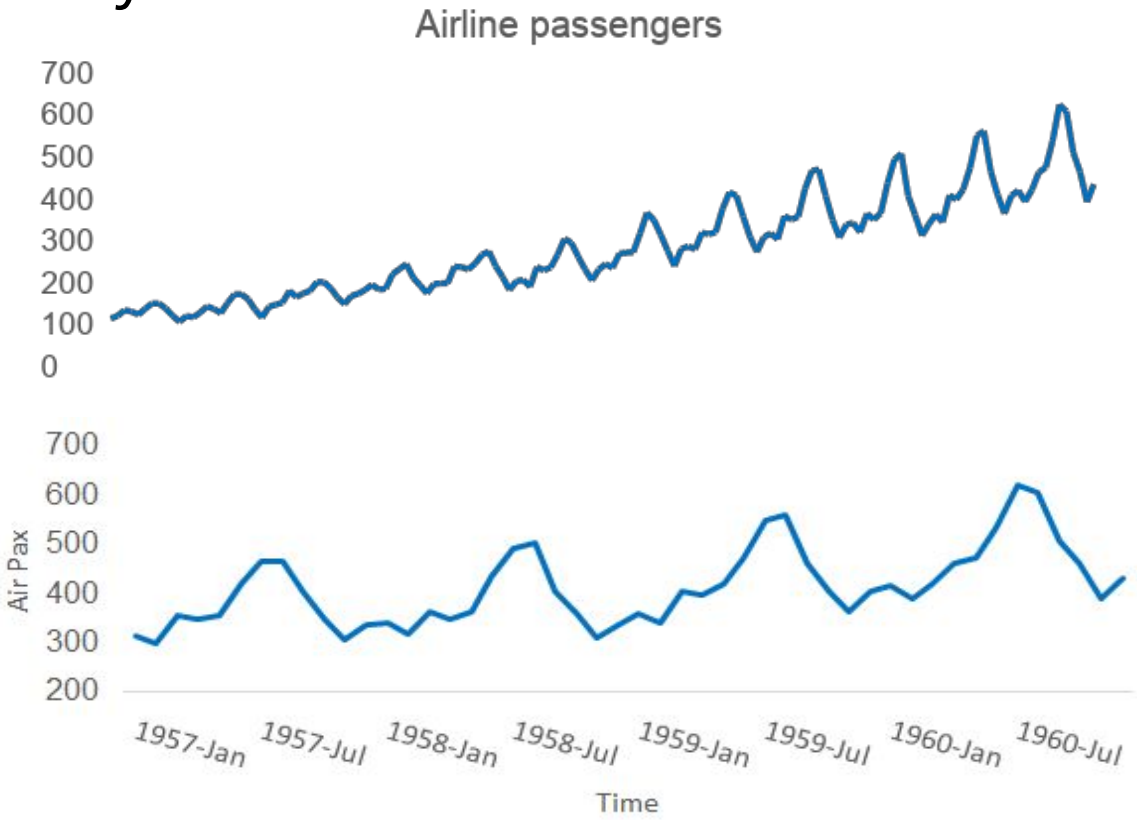
Trend

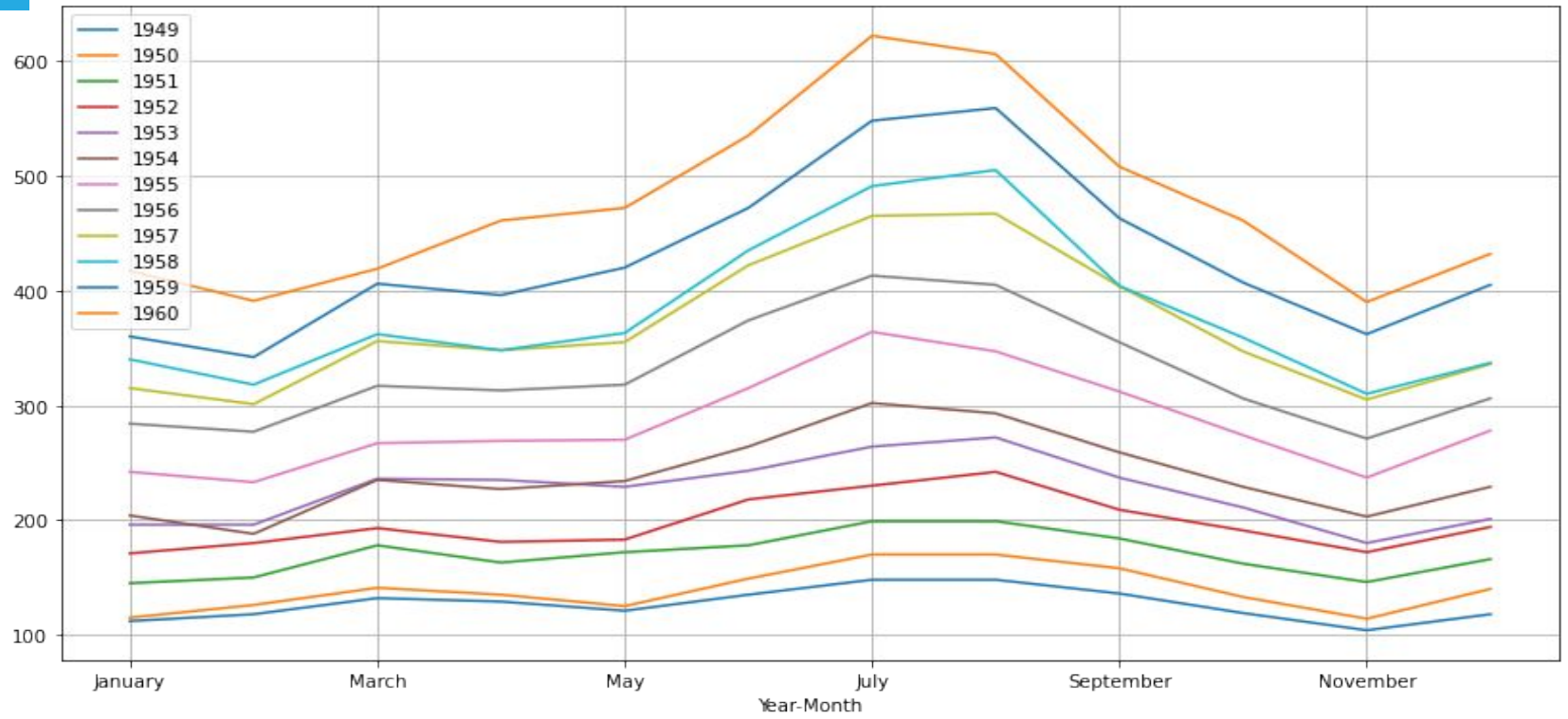
- Trend is a long term movement of a series; either increasing or decreasing.



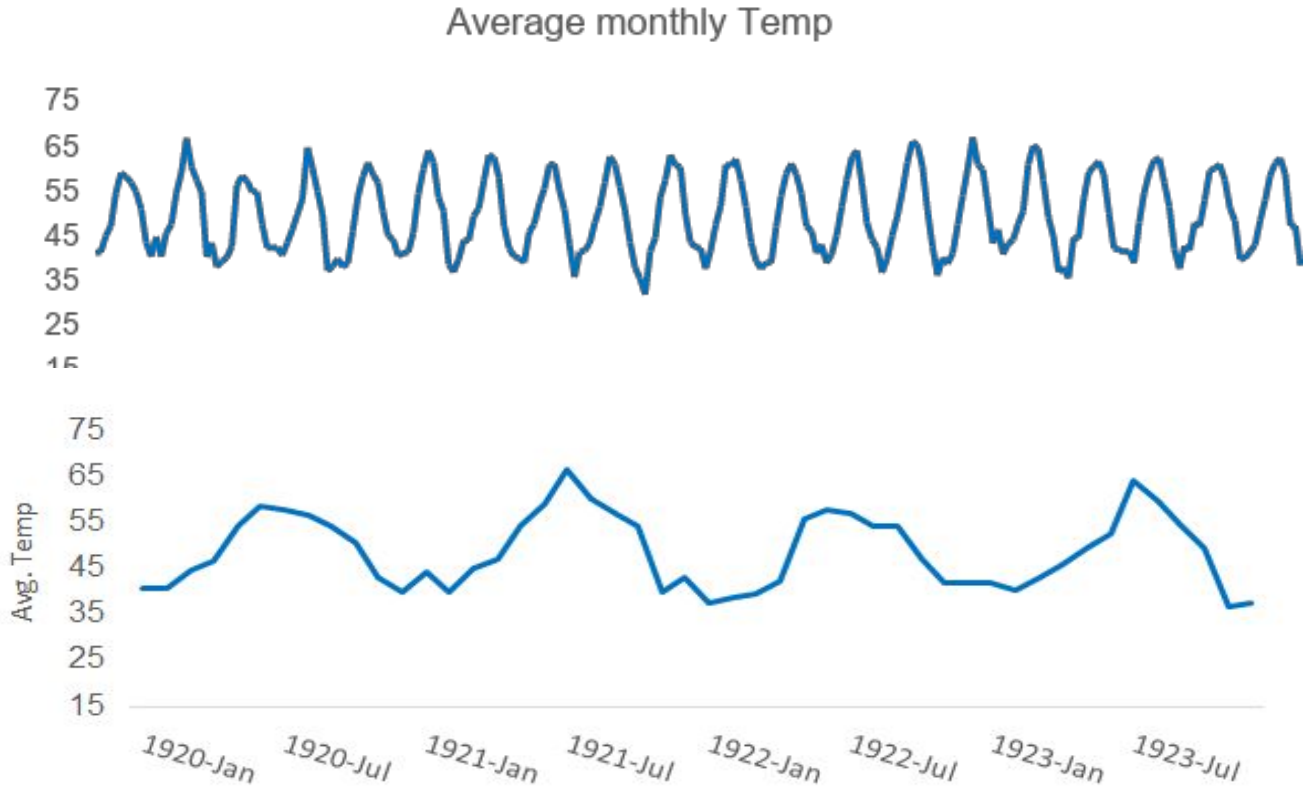


Seasonality




1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
600
500
400
300
200
100
January
March
May
July
September
November
Year-Month

Seasonality



Seasonality

- Seasonality represents intra-year stable fluctuations repeatable year after year with respect to timing, direction and magnitude.
- It is a normal variation 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

Examples of seasonality

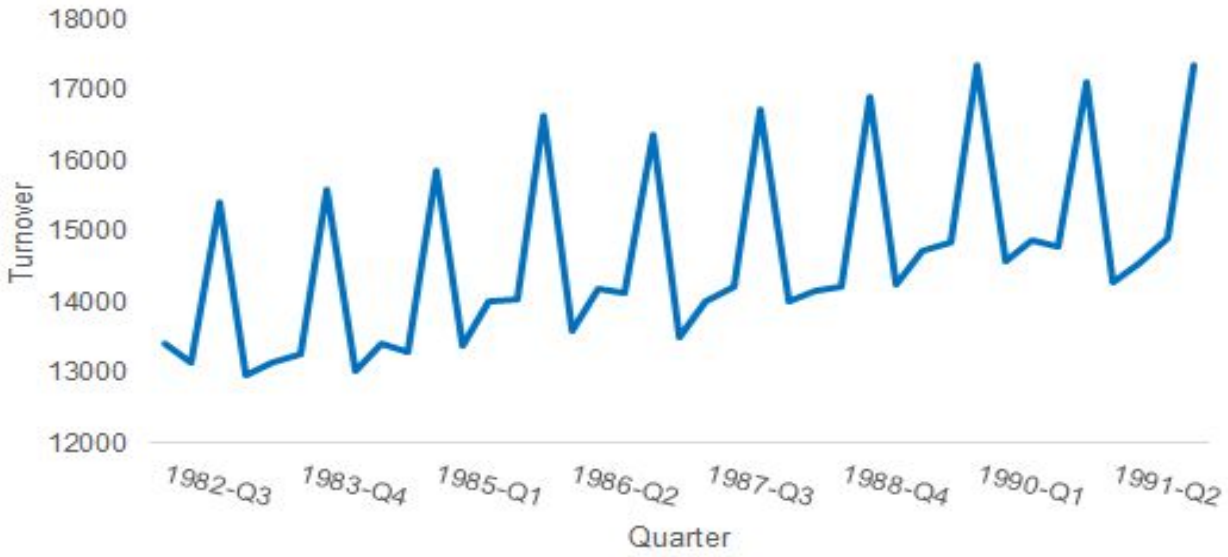
- Demand for winter clothes .
- Airlines and train ticket demands.
- Incidence of influenza or other vector-borne diseases.
- Will stock prices show any seasonal pattern?

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 or seasonality, 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!

Time series decomposition

Quarterly Time Series: Retail Turnover



Why decomposition?

- 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

$$Y_t = T_t * S_t * I_t$$

- Useful when the seasonal variation depends upon trend.

Additive model

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

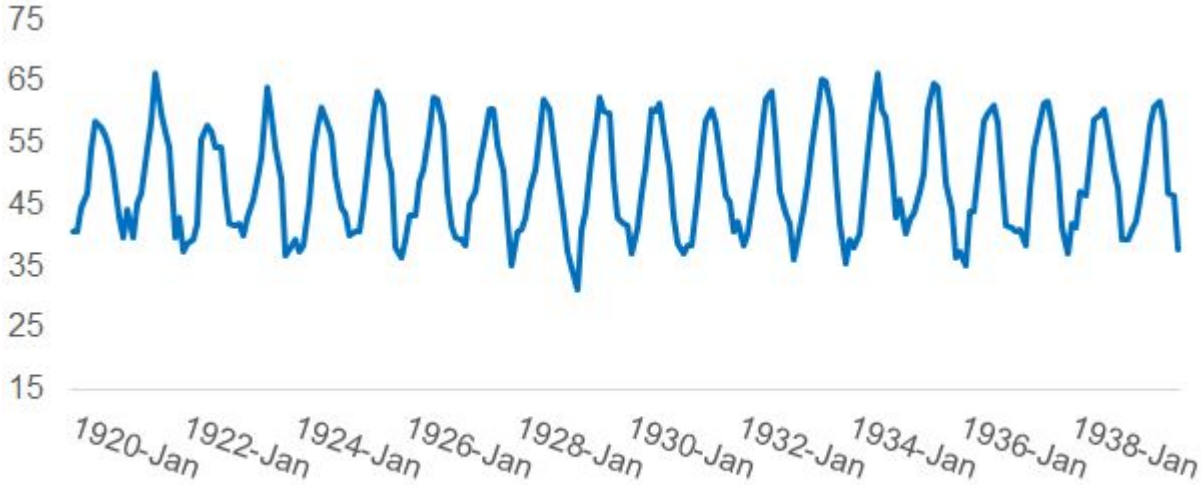
Example: Additive Seasonality Model of Sales

$$Y_t = T_t + S_t + I_t$$

Sales = Trend + Seasonality + Error

Additive model

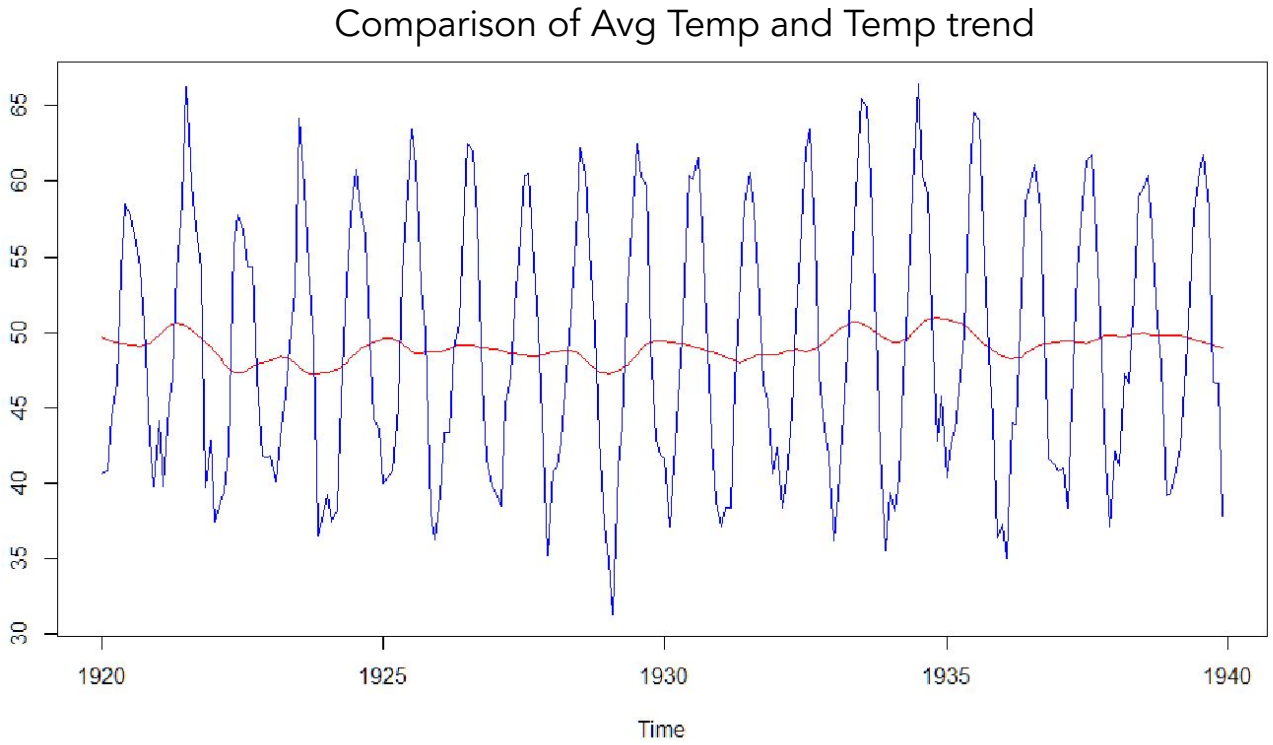
Example: Average Monthly Temp



Additive seasonality: If seasonal fluctuations do not change with trend.

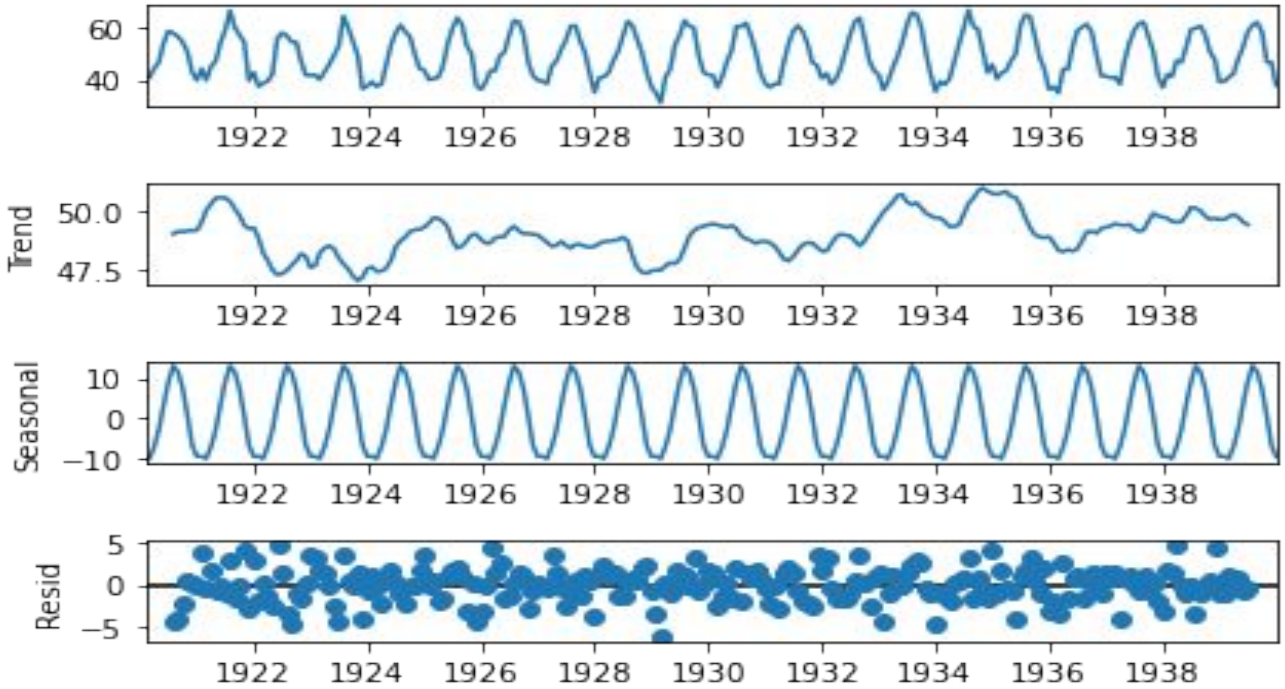
Additive model

Example: Average monthly temp TS



Additive model

Example: Decomposition of average monthly temp TS

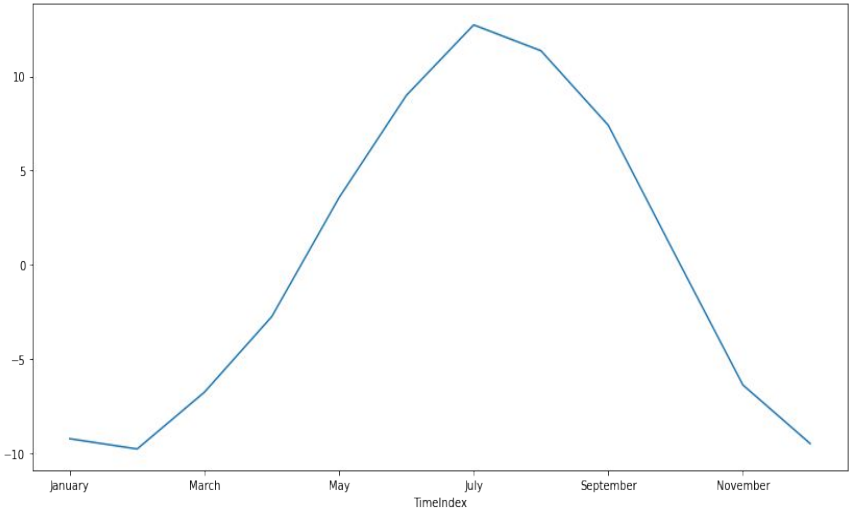


Additive model

Example: Average monthly temp TS

- Constant seasonality
- Trend contribution

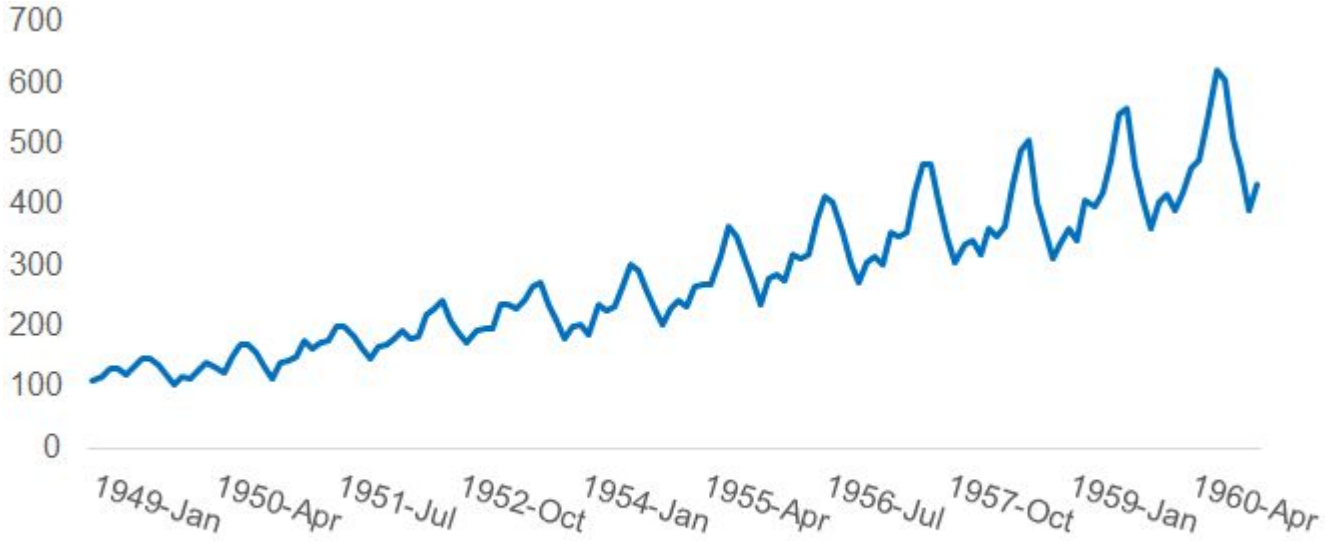
Month	Seasonal Index	Month	Seasonal Index
Jan	-9.24	Jul	12.73
Feb	-9.79	Aug	11.35
Mar	-6.77	Sep	7.40
Apr	-2.76	Oct	0.47
May	3.55	Nov	-6.39
Jun	8.99	Dec	-9.50



Sum of additive seasonal indices = 0

Multiplicative model

Example: Passenger Volume



Multiplicative model

Example: Passenger Volume

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

$$Y_t = T_t * S_t * I_t$$

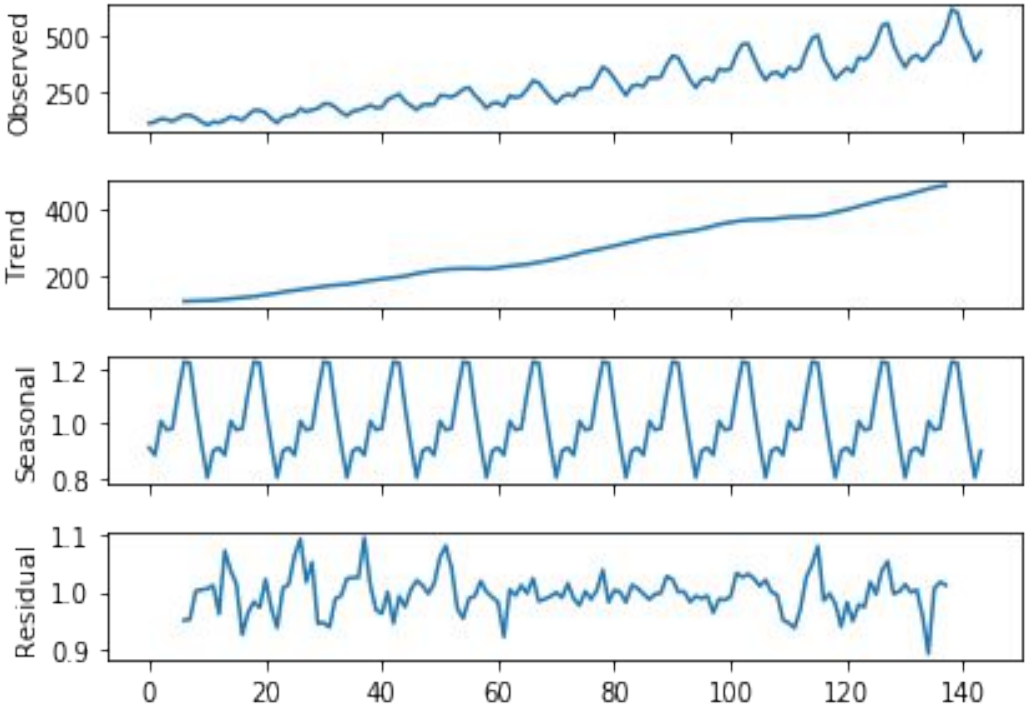
Volume = Trend *Seasonality * Error

- Need logarithmic transformation to convert into an additive series

Log(Vol) = log(Trend) + log(Seasonality) + log(Error)

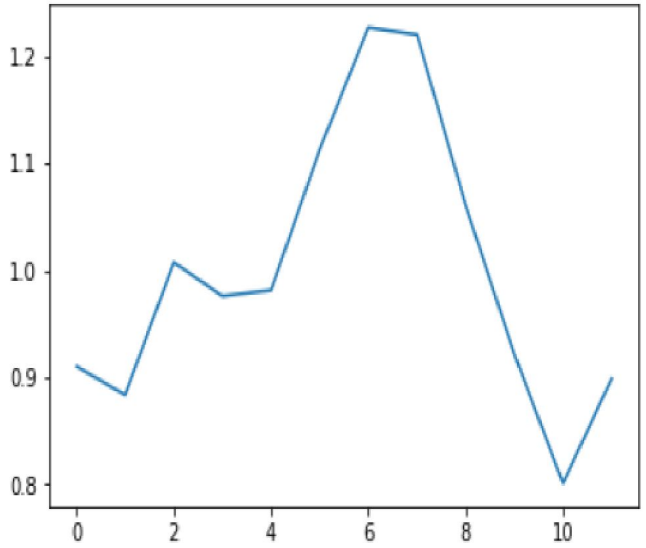
Multiplicative model

Example: Decomposition of Passenger Volume TS



Multiplicative model

Example: Passenger Volume TS



Critical look at seasonality

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

Moving average

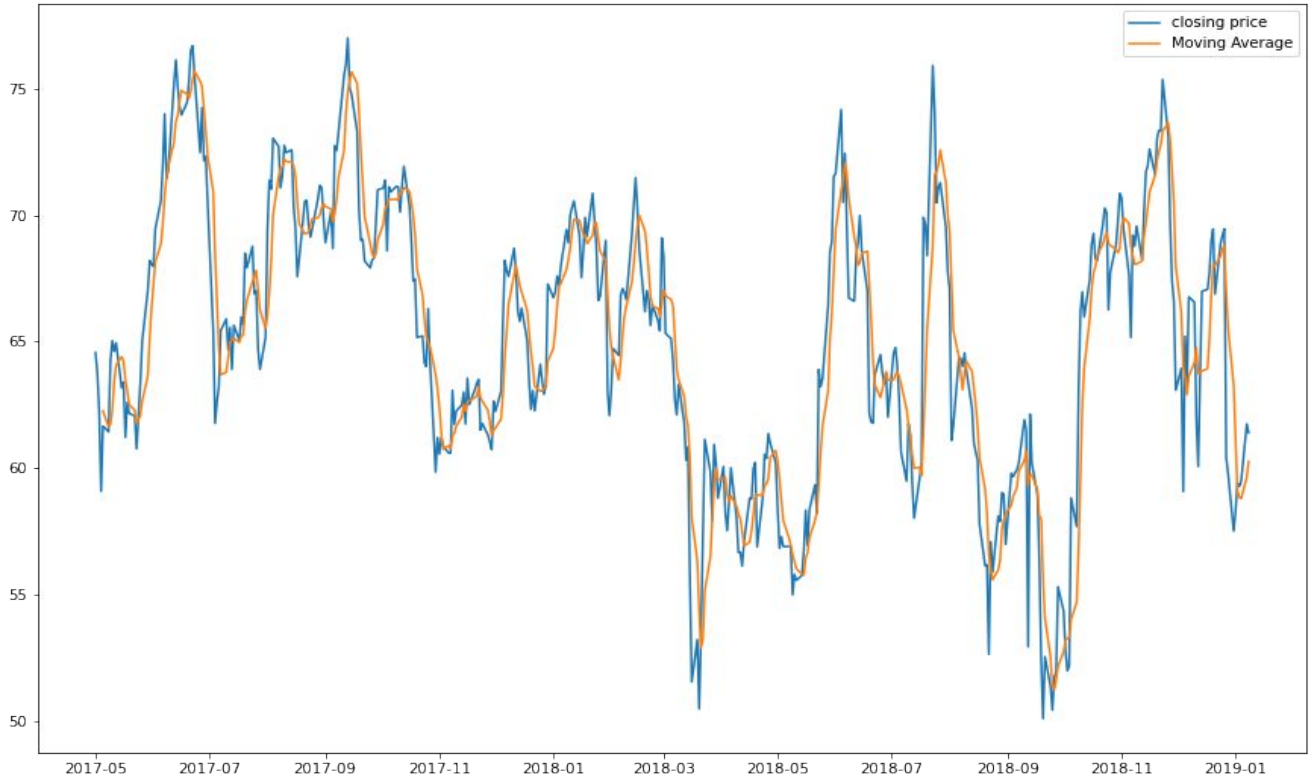
- Moving average is one of the most popular and often-used technical indicators to understand stock movement

- The moving average is easy to calculate and, once plotted on a chart, is a powerful visual trend-spotting tool.

- A simple moving average has two components:

Average: Take average of values over a window of certain width of the time series Moving: Move the window to update the average value.

Tesla stock price (MA 5)



Tesla stock price (MA 30)

Summary

- Forecasting techniques
- What is time series
- Examples of time series
- Features of time series
- Decomposition of time series

Thank You