Modul 3

Time Series Forecasting

Data Science Program

Purwadhika
Startup and Coding School

Outline

What is Time Series Data ?

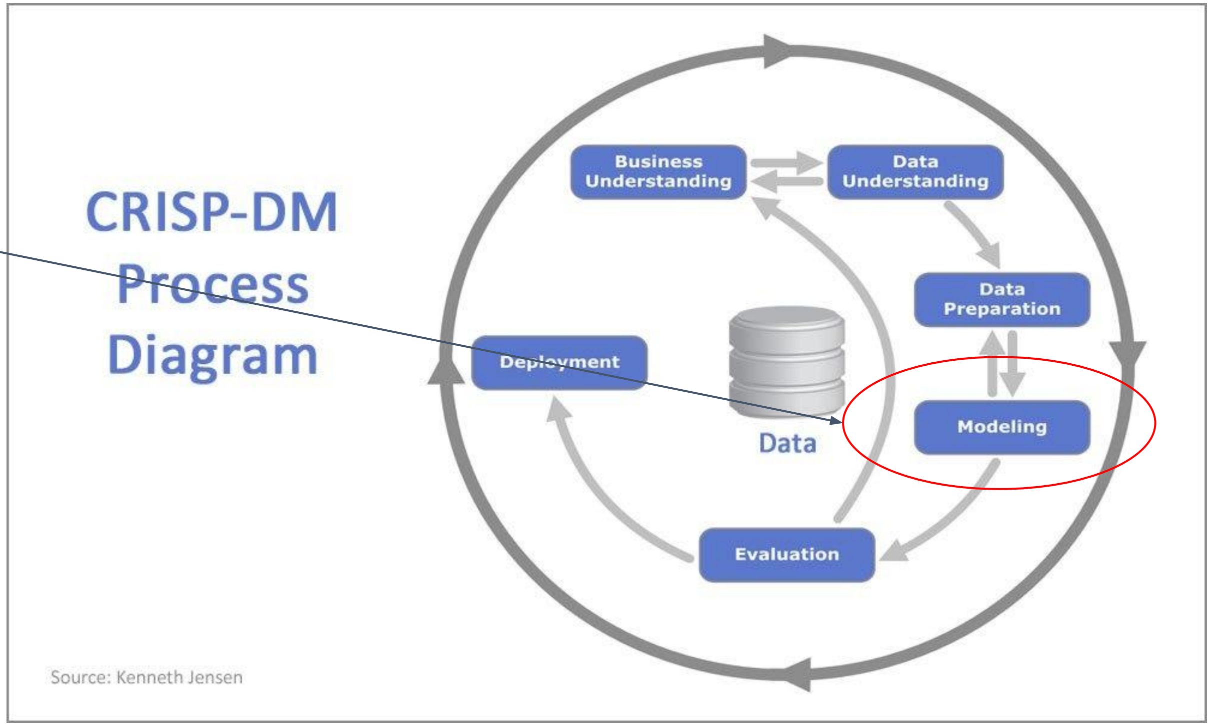
Time Series Forecasting

- Univariate
- Multivariate

Time Series Model with Exogenous Variable

Time Series Model Feature Engineering

Time Series Model Evaluation



What Is Time Series Data ?

Purwadhika
Startup and Coding School

Time Series Data

- Time series is a sequence of observations recorded at a regular time
- The frequency could be Yearly, Monthly, Daily or even milliseconds
- Not necessarily within the same interval
- The data analysis for time series is inherently different compared to the other data because:
  - It is time dependent
  - Time series could contain trend, cycle and seasonality

Time Series Data Example : Univariate

	Month	Sales
0	1-01	266.0
1	1-02	145.9
2	1-03	183.1
3	1-04	119.3
4	1-05	180.3

....

31	3-08	407.6
32	3-09	682.0
33	3-10	475.3
34	3-11	581.3
35	3-12	646.9

This dataset describes the **monthly number of sales** of shampoo over a 3 year period.

The units are a sales count and there are 36 observations. The original dataset is credited to Makridakis, Wheelwright and Hyndman (1998)

Only **one variable**

Purwadhika
Startup and Coding School

Time Series Data Example : Multivariate

Day	Average Temperature	Ice Cream Sales
1	25	2600
2	20	2100
3	44	8000
4	35	5100
...	...	...

This dataset describes the **daily number of sales** of ice cream.

besides number of sales, the dataset also provide **daily average temperature**.

More than one variables (**two variables**)

Why Is Time Series Data ?

Can provide massive business advantages:

- imagine you already know the number of sales of shampoo for several month ahead
- you can prepare the stock accordingly (not too much nor too little)
- same goes with the ice cream sales

Time Series Forecasting

Time Series Forecasting

Forecasting : Predicting the value (e.g shampo sales, ice cream sales) for several period ahead

Univariate Time Series Forecasting : predicting using their own value

- shampo sales

Time Series Forecasting with Exogenous Variable : predicting using their own value and another variable

- ice cream sales (with the help of average temperature)

Purwadhika
Startup and Coding School

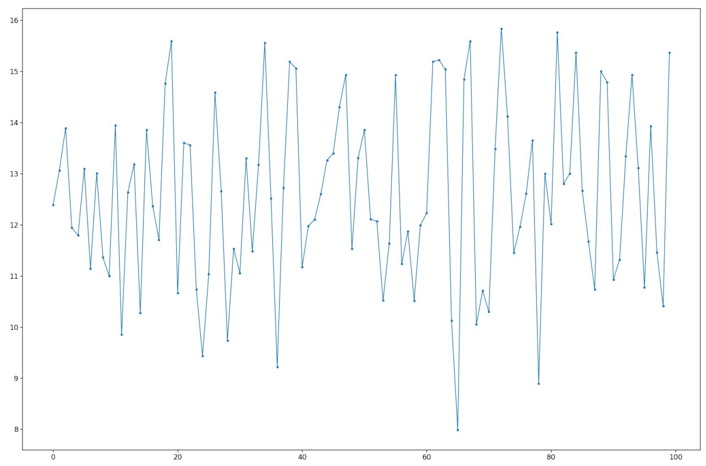
Must Know Term

These term will be very helpful to understand forecasting method in time series :

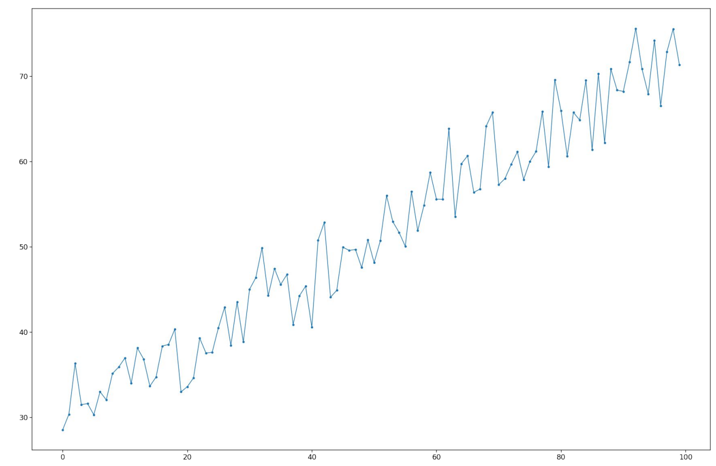
- Time series data pattern
- Stationarity

Time Series Pattern

Plotting time at x-axis and the data or variable of interest at y-axis



Random Pattern

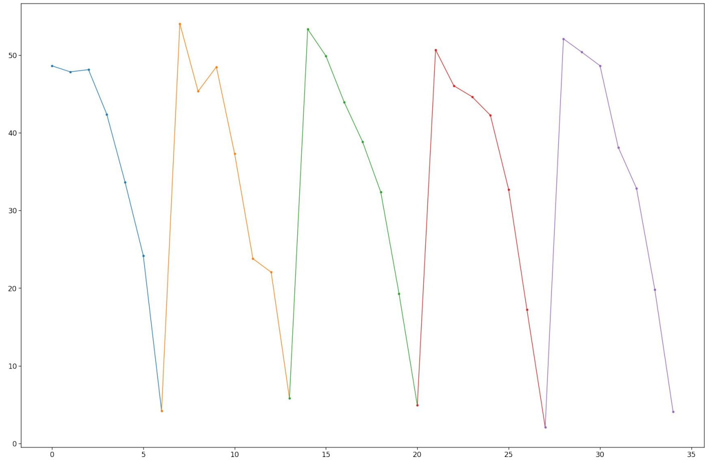


TRENDS : increasing or decreasing slope observed in the time series

Purwadhika
Startup and Coding School

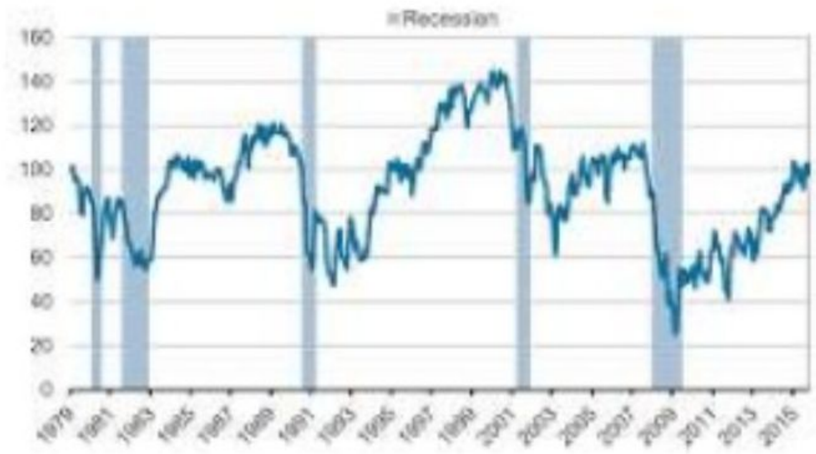
Time Series Pattern

Plotting time at x-axis and the data or variable of interest at y-axis



SEASONAL : a distinct repeated pattern at fixed period of time. Can be affected by seasonal factors such as

- weekly
- daily
- etc



CYCLICAL : a distinct repeated pattern at unpredicted period of time and extend beyond a year

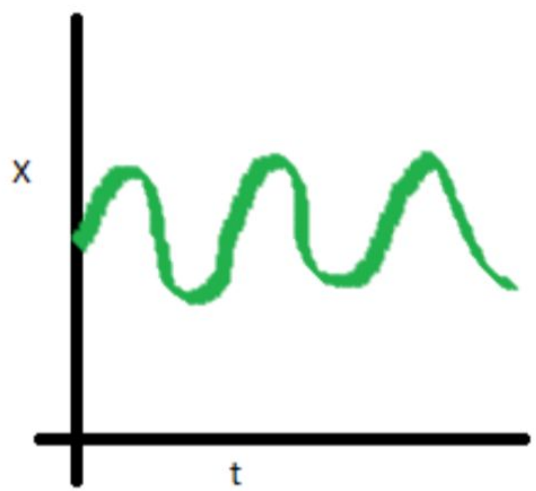
Purwadhika
Startup and Coding School

Stationarity

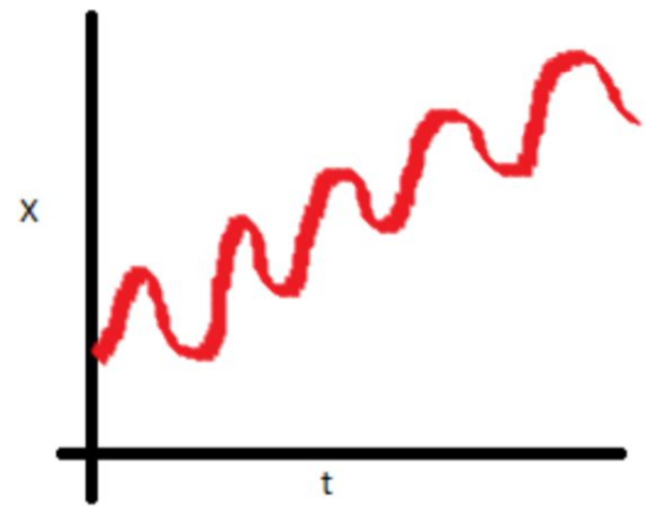
- Stationarity may has an important role in forecasting.
- Stationarity mirrors the behaviour of the process that happen in the data.
- There is some forecasting method that require stationarity for good performance
- There is also some method that able to achieve good performance regardless stationarity
- Stationarity :
  - Mean
  - Variance

Stationarity - Mean

The mean is constant, not be a function of time



Stationary series

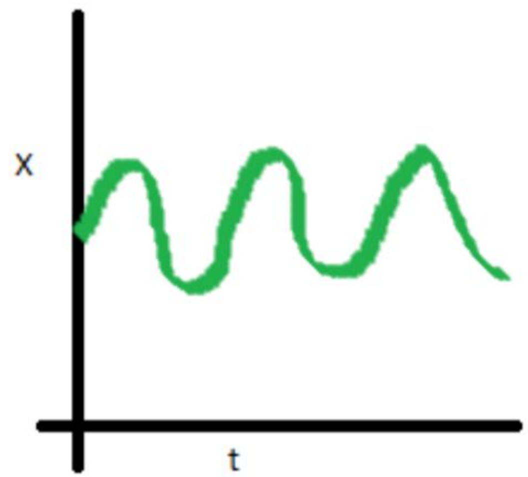


Non-Stationary series

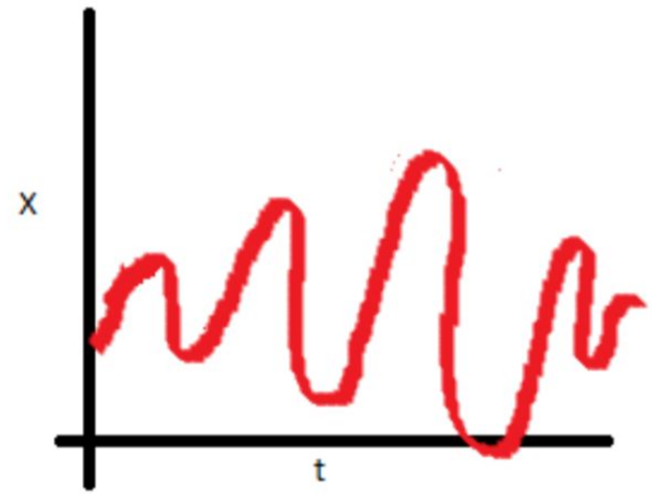
Purwadhika
Startup and Coding School

Stationarity - Variance

The variance is constant, not be a function of time



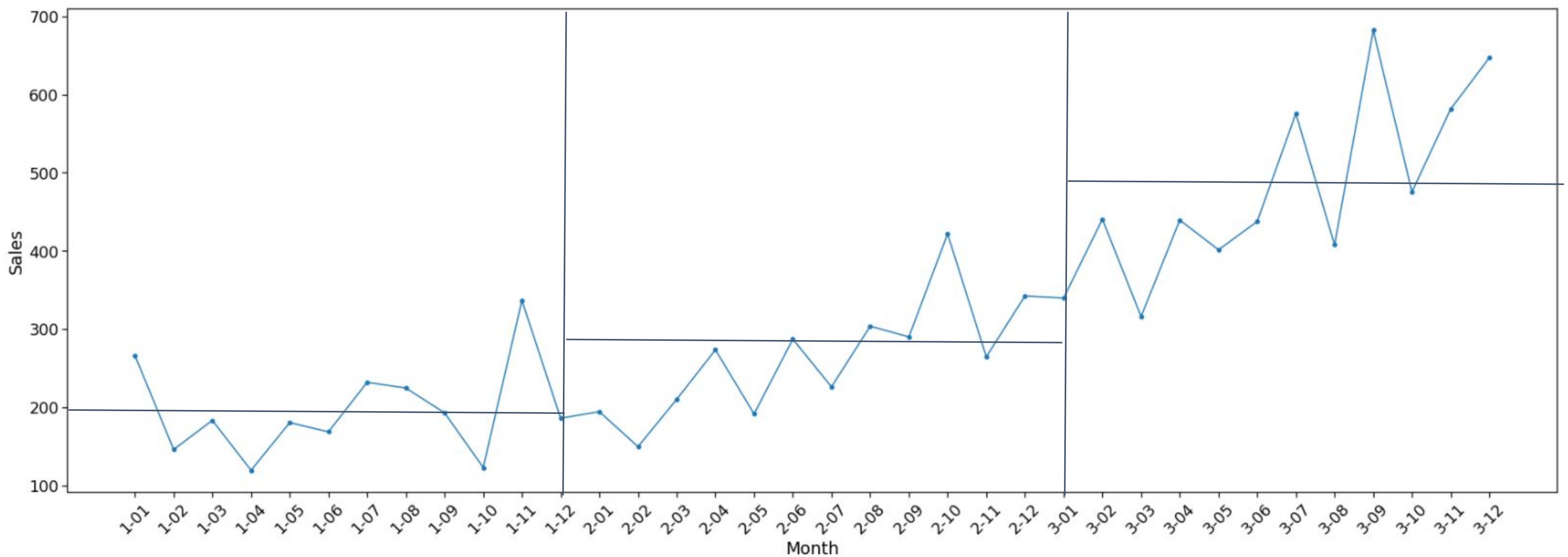
Stationary series



Non-Stationary series

Purwadhika
Startup and Coding School

Time Series Plot for Shampoo Sales



Violate Mean Stationarity

Purwadhika
Startup and Coding School

Differencing Level 1

	Month	Sales
1	1-01	266.0
2	1-02	145.9
3	1-03	183.1
4	1-04	119.3
5	1-05	180.3
6	1-06	168.5
7	1-07	231.8
8	1-08	224.5
9	1-09	192.8
10	1-10	122.9

	Month	Sales	Sales Stationary
1	1-01	266.0	NaN
2	1-02	145.9	-120.1
3	1-03	183.1	37.2
4	1-04	119.3	-63.8
5	1-05	180.3	61.0
6	1-06	168.5	-11.8
7	1-07	231.8	63.3
8	1-08	224.5	-7.3
9	1-09	192.8	-31.7
10	1-10	122.9	-69.9

Due to the needs of stationarity, we often can't directly analyze the data. As a solution we can transform the data using differencing method to achieve stationarity.

transform the data : $Y_t \rightarrow Z_t = Y_t - Y_{t-1}$ (first differencing)

: e.g $Z_2 = Y_2 - Y_1 = 145.9 - 266 = -120.1$

: e.g $Z_3 = Y_3 - Y_2 = 183.1 - 145.9 = 37.2$, and so on

Purwadhika
Startup and Coding School

Differencing Level 2

transform the data : $Z_t \rightarrow W_t = Z_t - Z_{t-1}$ (second differencing)

: $W_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

: e.g $W_2 = Z_2 - Z_1 = 37.2 - (-120.1) = -82.9$, and so on

	Month	Sales
1	1-01	266.0
2	1-02	145.9
3	1-03	183.1
4	1-04	119.3
5	1-05	180.3
6	1-06	168.5
7	1-07	231.8
8	1-08	224.5
9	1-09	192.8
10	1-10	122.9

→

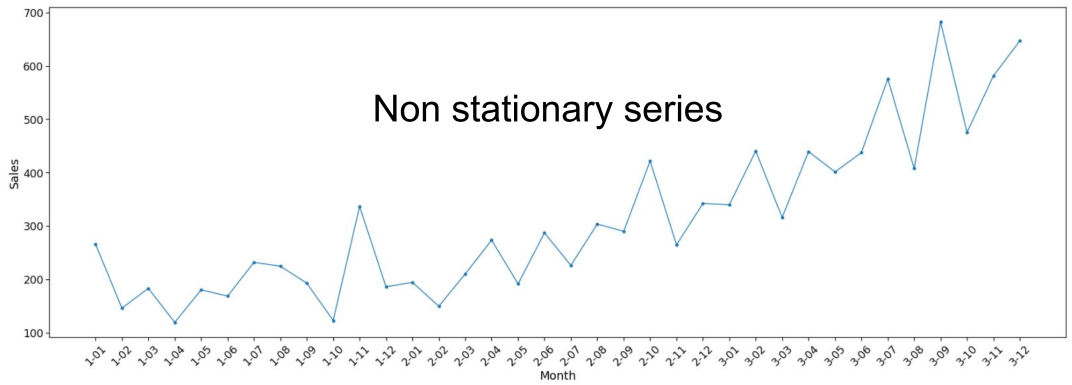
	Month	Sales	Sales Stationary
1	1-01	266.0	NaN
2	1-02	145.9	-120.1
3	1-03	183.1	37.2
4	1-04	119.3	-63.8
5	1-05	180.3	61.0
6	1-06	168.5	-11.8
7	1-07	231.8	63.3
8	1-08	224.5	-7.3
9	1-09	192.8	-31.7
10	1-10	122.9	-69.9

→

	Month	Sales	Sales Stationary	Sales Stationary 2
1	1-01	266.0	NaN	NaN
2	1-02	145.9	-120.1	NaN
3	1-03	183.1	37.2	-82.9
4	1-04	119.3	-63.8	-26.6
5	1-05	180.3	61.0	-2.8
6	1-06	168.5	-11.8	49.2
7	1-07	231.8	63.3	51.5
8	1-08	224.5	-7.3	56.0
9	1-09	192.8	-31.7	-39.0
10	1-10	122.9	-69.9	-101.6

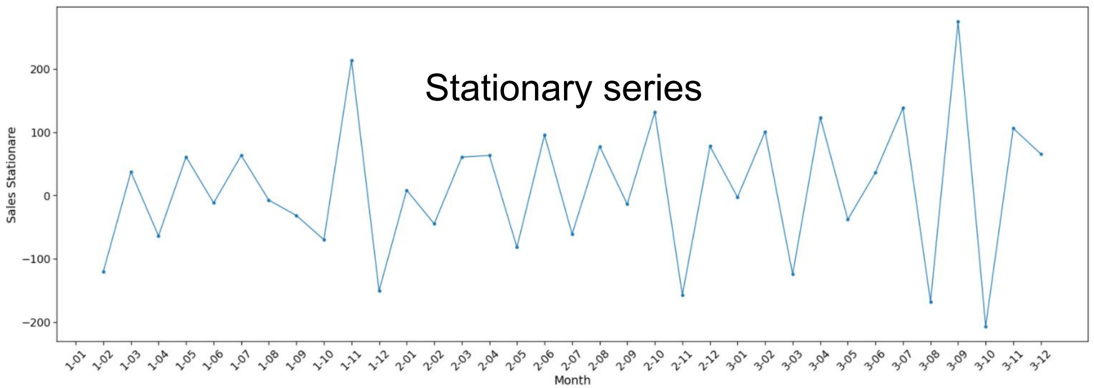
Purwadhika
Startup and Coding School

Time Series Plot for Shampoo Sales After Stationarity



Yt

transform the data : Yt → Zt = Yt - Yt-1



ARIMA

Univariate Time Series Forecasting : ARIMA

ARIMA : use older observation as features to predict future value

ARIMA = Autoregressive Integrated Moving Average

hyperparameter ARIMA(p,d,q): p, d, q

p(Autoregressive) : how many previous observation used

d(Integrated) : adjust stationarity

q(Moving Average) : avoid correlated error (autocorrelation)

ARIMA Models

ARIMA = Autoregressive Integrated Moving Average

e.g.:

ARIMA(1,0,0) or AR(1) : Yt = a + b*Yt-1 + et

ARIMA(2,0,0) or AR(1) : Yt = a + b*Yt-1 + c*Yt-2 + et

ARIMA(1,1,0) or ARI(1,1) : Zt = a + b*Zt-1 + et

: (Yt - Yt-1) = a + b*(Yt-1 - Yt-2) + et

: Yt = a + (1+b) Yt-1 - b*Yt-2 + et

ARIMA(0,0,1) or MA(1) : Yt = a + et + r*et-1

ARIMA(1,0,1) or ARMA(1,1) : Yt = a + b*Yt-1 + et + r*et-1

Autoregressive (AR) : p

ARIMA(1,0,0) : Yt = a + b*Yt-1 + et

ARIMA(2,0,0) : Yt = a + b*Yt-1+ c*Yt-2 + et

Uses 1 previous period (Yt-1) as feature

	Month	Sales
0	1-01	266.0
1	1-02	145.9
2	1-03	183.1
3	1-04	119.3
4	1-05	180.3
5	1-06	168.5
6	1-07	231.8
7	1-08	224.5
8	1-09	192.8
9	1-10	122.9

Yt

	Month	Sales	lag1 Sales
0	1-01	266.0	NaN
1	1-02	145.9	266.0
2	1-03	183.1	145.9
3	1-04	119.3	183.1
4	1-05	180.3	119.3
5	1-06	168.5	180.3
6	1-07	231.8	168.5
7	1-08	224.5	231.8
8	1-09	192.8	224.5
9	1-10	122.9	192.8

Yt-1

Uses 2 previous period (Yt-1 and Yt-2) as feature

	Month	Sales	lag1 Sales	lag2 Sales
0	1-01	266.0	NaN	NaN
1	1-02	145.9	266.0	NaN
2	1-03	183.1	145.9	266.0
3	1-04	119.3	183.1	145.9
4	1-05	180.3	119.3	183.1
5	1-06	168.5	180.3	119.3
6	1-07	231.8	168.5	180.3
7	1-08	224.5	231.8	168.5
8	1-09	192.8	224.5	231.8
9	1-10	122.9	192.8	224.5

Yt-1 Yt-2

Purwadhika
Startup and Coding School

Integrated (I) : d

ARIMA(1,1,0) : $Zt = a + b*Zt\text{-}1 + et$

: $(Yt - Yt\text{-}1) = a + b*(Yt\text{-}1 - Yt\text{-}2) + et$

: $Yt = a + (1+b)\, Yt\text{-}1 - b*Yt\text{-}2 + et$

	Month	Sales	lag1 Sales	Sales Stationary	lag1 Sales Stationary
0	1-01	266.0	NaN	NaN	NaN
1	1-02	145.9	266.0	-120.1	NaN
2	1-03	183.1	145.9	37.2	-120.1
3	1-04	119.3	183.1	-63.8	37.2
4	1-05	180.3	119.3	61.0	-63.8
5	1-06	168.5	180.3	-11.8	61.0
6	1-07	231.8	168.5	63.3	-11.8
7	1-08	224.5	231.8	-7.3	63.3
8	1-09	192.8	224.5	-31.7	-7.3
9	1-10	122.9	192.8	-69.9	-31.7
		Yt	Yt-1	Zt	Zt-1

Uses Zt as target variable and Zt-1 as feature

Purwadhika
Startup and Coding School

Moving Average (MA) : q

ARIMA(0,0,1) : Yt = a + et + r*et-1

	Month	Sales	lag1 et = sales - mean(sales)
0	1-01	266.0	NaN
1	1-02	145.9	-46.6
2	1-03	183.1	-166.7
3	1-04	119.3	-129.5
4	1-05	180.3	-193.3
5	1-06	168.5	-132.3
6	1-07	231.8	-144.1
7	1-08	224.5	-80.8
8	1-09	192.8	-88.1
9	1-10	122.9	-119.8

Yt et-1

ARIMA(1,0,1) : Yt = a + b*Yt-1 + et + r*et-1

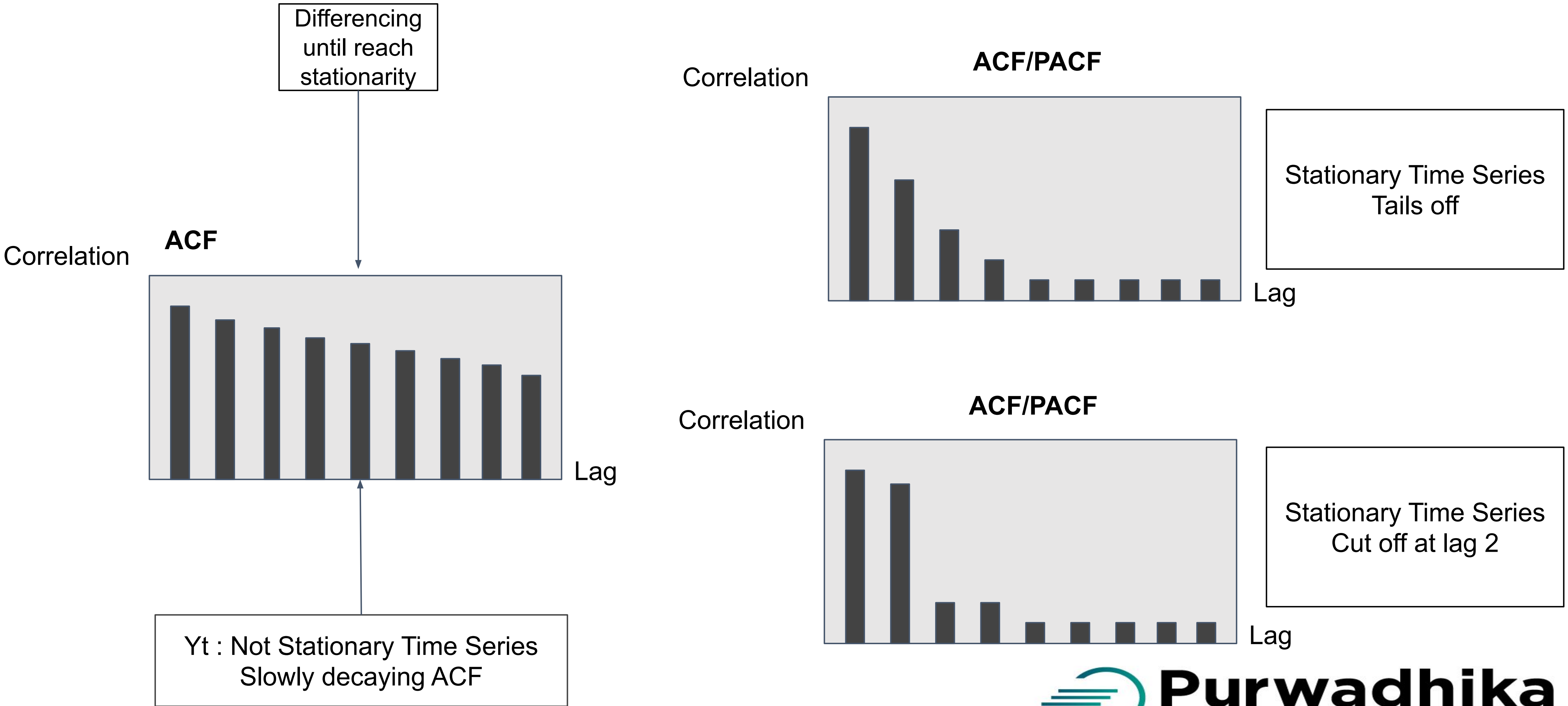
	Month	Sales	lag1 Sales	et-1
0	1-01	266.0	NaN	NaN
1	1-02	145.9	266.0	-139.193703
2	1-03	183.1	145.9	-8.510723
3	1-04	119.3	183.1	-101.266317
4	1-05	180.3	119.3	9.394083
5	1-06	168.5	180.3	-49.886864
6	1-07	231.8	168.5	22.597975
7	1-08	224.5	231.8	-33.973237
8	1-09	192.8	224.5	-59.991091
9	1-10	122.9	192.8	-105.216566

Yt Yt-1 et-1

ACF-PACF

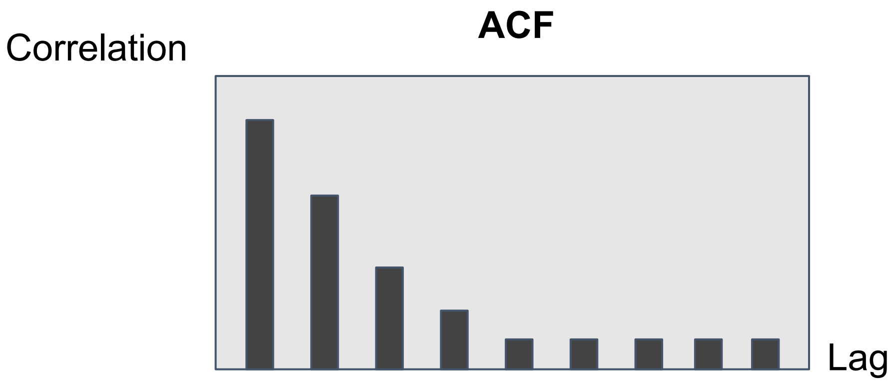
- We can use ACF-PACF to determine the best combination of p d and q
- ACF is a measured of the correlation between the time series and their own lags
- PACF measures the correlation between the time series with their own lags but after eliminating the variations such as Trend and Seasonality

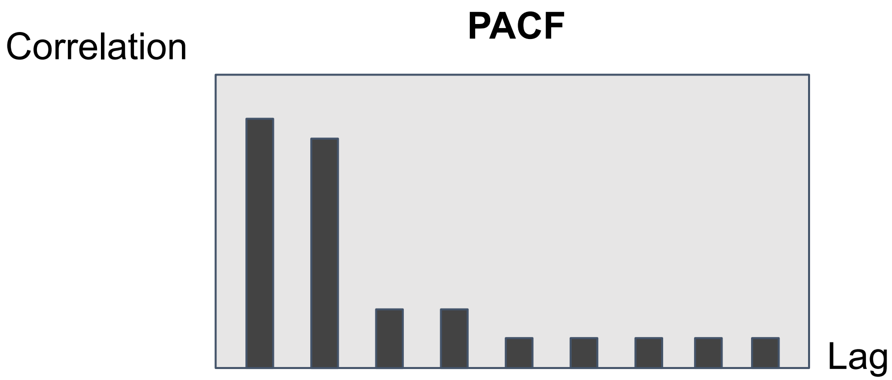
Determining p d q Based On ACF PACF



Purwadhika
Startup and Coding School

Determining p d q Based On ACF PACF





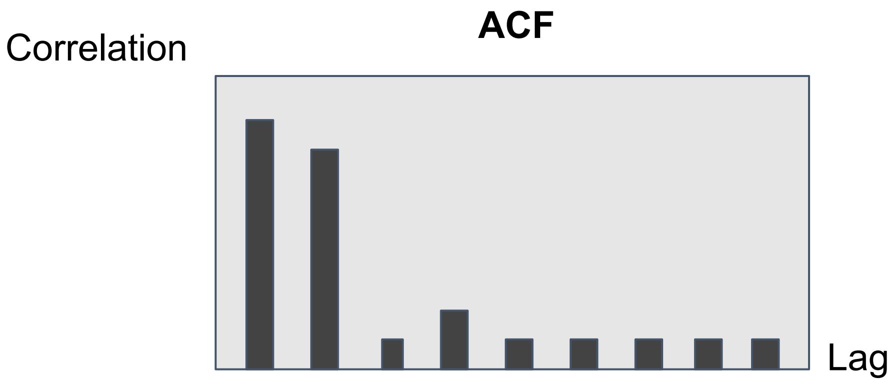
rules

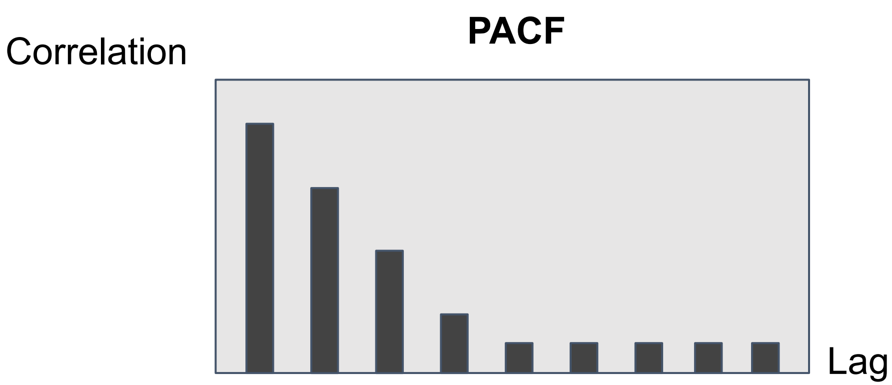
- ACF Tails off
- PACF Cut Off at lag ?
- Model AR

example :

- ACF Tails off
- PACF Cut Off at lag 2
- Model ARIMA(2,d,0)
- with d the order of differencing needed until reach stationarity

Determining p d q Based On ACF PACF

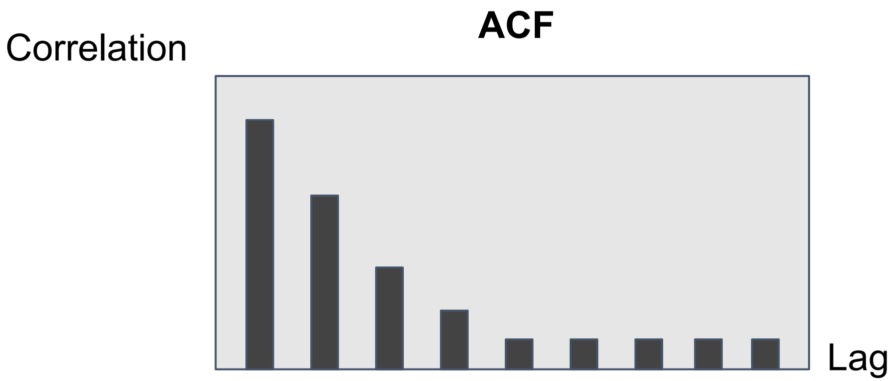


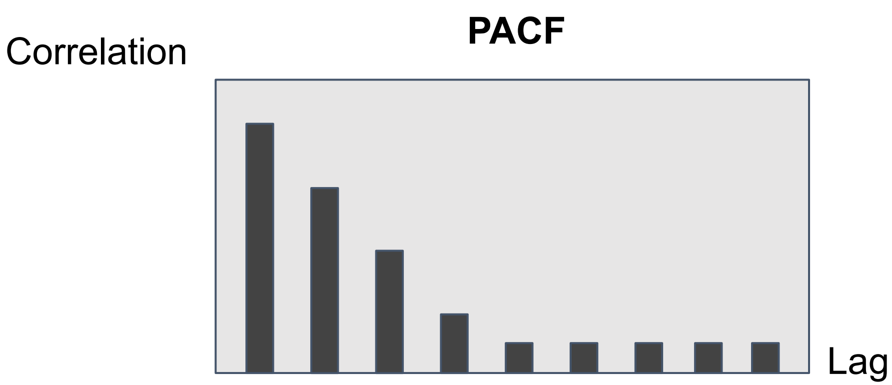


rules
- ACF Cut Off at lag ?
- PACF tails off
- Model MA

example :
- ACF Cut Off at lag 2
- PACF tails off
- Model ARIMA(0,d,2)
- with d the order of differencing needed until reach stationarity

Determining p d q Based On ACF PACF





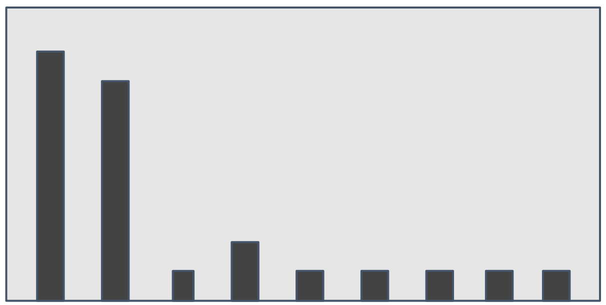
rules

- Both ACF and PACF tails off
- Model ARIMA
- choose all possible combination
    - ARIMA(1,d,0), ARIMA(1,d,1), ARIMA(2,d,0), ARIMA(0,d,1), ARIMA(1,d,1). ….
    - with d the order of differencing needed until reach stationarity

Determining p d q Based On ACF PACF

Correlation

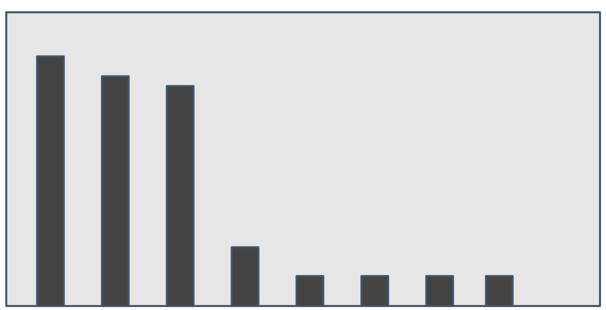
ACF



Lag

Correlation

PACF



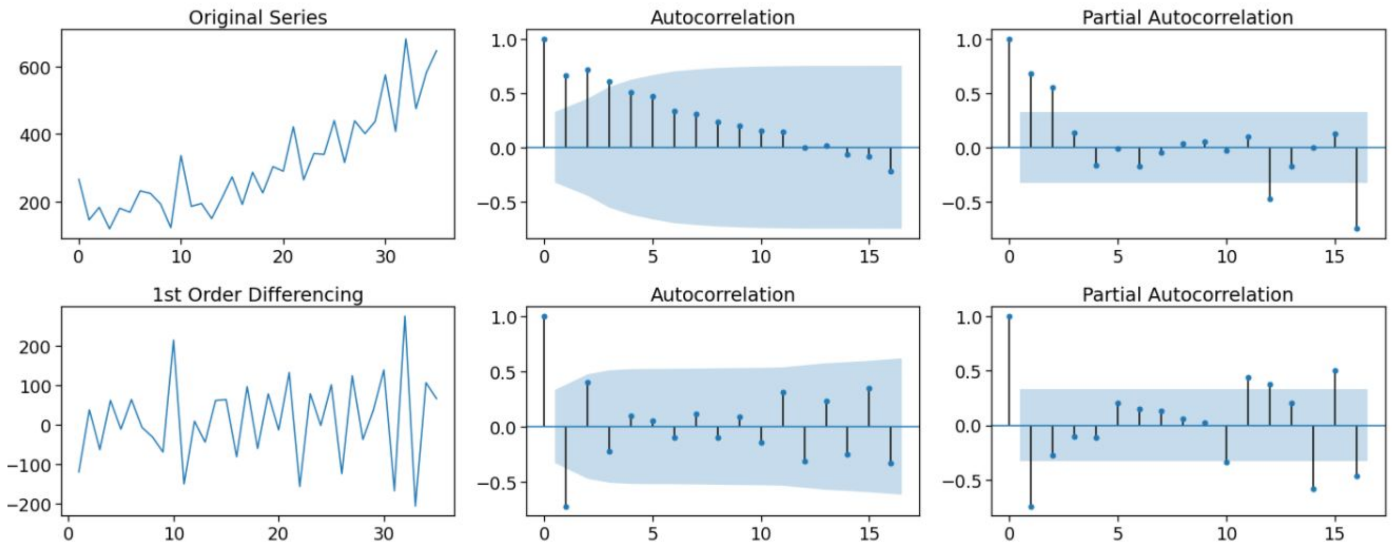
Lag

rules

- Both ACF and PACF cut off at certain leg
- Model AR or MA
- choose between
    - AR(3) or MA(2)
    - with d the order of differencing needed until reach stationarity

Purwadhika
Startup and Coding School

ACF-PACF Shampoo Dataset



Identification

- needs first difference
- ACF cut off at 2
- PACF cut off at 2

We need to choose between ARIMA(2,1,0) or ARIMA(0,1,2)

Python Exercise : Time Series ARIMA

Analyze data shampoo sales.csv

- identified data pattern
- build ACF PACF plot until second differencing
- identified the most suitable model based on ACF PACF plot
- build ARIMA(2,1,0) model
- forecast for 6 periods ahead

Time Series Model with Exogenous Variable

Purwadhika
Startup and Coding School

Time Series Model with Exogenous Variable

- In ARIMA, we utilize its own data as feature to predict future value/forecast
- You can develop machine learning method to predict future value with the help of exogenous variable
- The only requirement to use an exogenous variable is we need to know the value of the variable during the forecast period as well, such as Date.

	Date	Consumption	
0	2006-01-01	1069.18400	
1	2006-01-02	1380.52100	
2	2006-01-03	1442.53300	
3	2006-01-04	1457.21700	
4	2006-01-05	1477.13100	
...	...	...	
4378	2017-12-27	1263.94091	3
4379	2017-12-28	1299.86398	5
4380	2017-12-29	1295.08753	5
4381	2017-12-30	1215.44897	7
4382	2017-12-31	1107.11488	7

Model

Recommended Model:

- Linear Regression
- Support Vector Regression
- Any model with ability to extrapolate

Tree based model such as Decision Tree and Random Forest are not recommended because they can't extrapolate:

https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why

While time series forecasting method is an extrapolation

Purwadhika
Startup and Coding School

Time Series Feature Engineering

- date
- lag variable
- differencing

Date

Features that may be created from Date:

- Day of month
- Day of week
- Day of year
- Weekend or weekday
- Payday
- Holiday
- Quarter
- Start of Quarter
- End of Quarter
- Days to month-end
- Days to month start
- Days to holiday
- Season of year
- Certain Event

Example:

	Date	year	month	day	weekday
0	2006-01-01	2006	1	1	6
1	2006-01-02	2006	1	2	0
2	2006-01-03	2006	1	3	1
3	2006-01-04	2006	1	4	2
4	2006-01-05	2006	1	5	3
...	...	...	...	...	...
4378	2017-12-27	2017	12	27	2
4379	2017-12-28	2017	12	28	3
4380	2017-12-29	2017	12	29	4
4381	2017-12-30	2017	12	30	5
4382	2017-12-31	2017	12	31	6

Purwadhika
Startup and Coding School

Lag Variable

Uses 1 previous period (Yt-1) as feature

	Month	Sales
0	1-01	266.0
1	1-02	145.9
2	1-03	183.1
3	1-04	119.3
4	1-05	180.3
5	1-06	168.5
6	1-07	231.8
7	1-08	224.5
8	1-09	192.8
9	1-10	122.9

Yt

	Month	Sales	lag1 Sales
0	1-01	266.0	NaN
1	1-02	145.9	266.0
2	1-03	183.1	145.9
3	1-04	119.3	183.1
4	1-05	180.3	119.3
5	1-06	168.5	180.3
6	1-07	231.8	168.5
7	1-08	224.5	231.8
8	1-09	192.8	224.5
9	1-10	122.9	192.8

Yt-1

Uses 2 previous period (Yt-1 and Yt-2) as feature

	Month	Sales	lag1 Sales	lag2 Sales
0	1-01	266.0	NaN	NaN
1	1-02	145.9	266.0	NaN
2	1-03	183.1	145.9	266.0
3	1-04	119.3	183.1	145.9
4	1-05	180.3	119.3	183.1
5	1-06	168.5	180.3	119.3
6	1-07	231.8	168.5	180.3
7	1-08	224.5	231.8	168.5
8	1-09	192.8	224.5	231.8
9	1-10	122.9	192.8	224.5

Yt-1 Yt-2

Purwadhika
Startup and Coding School

Differencing

transform the data : $Z_t \rightarrow W_t = Z_t - Z_{t-1}$ (second differencing)

: $W_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

: e.g $W_2 = Z_2 - Z_1 = 37.2 - (-120.1) = -82.9$, and so on

	Month	Sales
0	1-01	266.0
1	1-02	145.9
2	1-03	183.1
3	1-04	119.3
4	1-05	180.3
5	1-06	168.5
6	1-07	231.8
7	1-08	224.5
8	1-09	192.8
9	1-10	122.9

	Month	Sales	Sales Stationary
0	1-01	266.0	NaN
1	1-02	145.9	-120.1
2	1-03	183.1	37.2
3	1-04	119.3	-63.8
4	1-05	180.3	61.0
5	1-06	168.5	-11.8
6	1-07	231.8	63.3
7	1-08	224.5	-7.3
8	1-09	192.8	-31.7
9	1-10	122.9	-69.9

	Month	Sales	Sales Stationary	Sales Stationary 2
0	1-01	266.0	NaN	NaN
1	1-02	145.9	-120.1	NaN
2	1-03	183.1	37.2	-82.9
3	1-04	119.3	-63.8	-26.6
4	1-05	180.3	61.0	-2.8
5	1-06	168.5	-11.8	49.2
6	1-07	231.8	63.3	51.5
7	1-08	224.5	-7.3	56.0
8	1-09	192.8	-31.7	-39.0
9	1-10	122.9	-69.9	-101.6

Time Series Model Evaluation

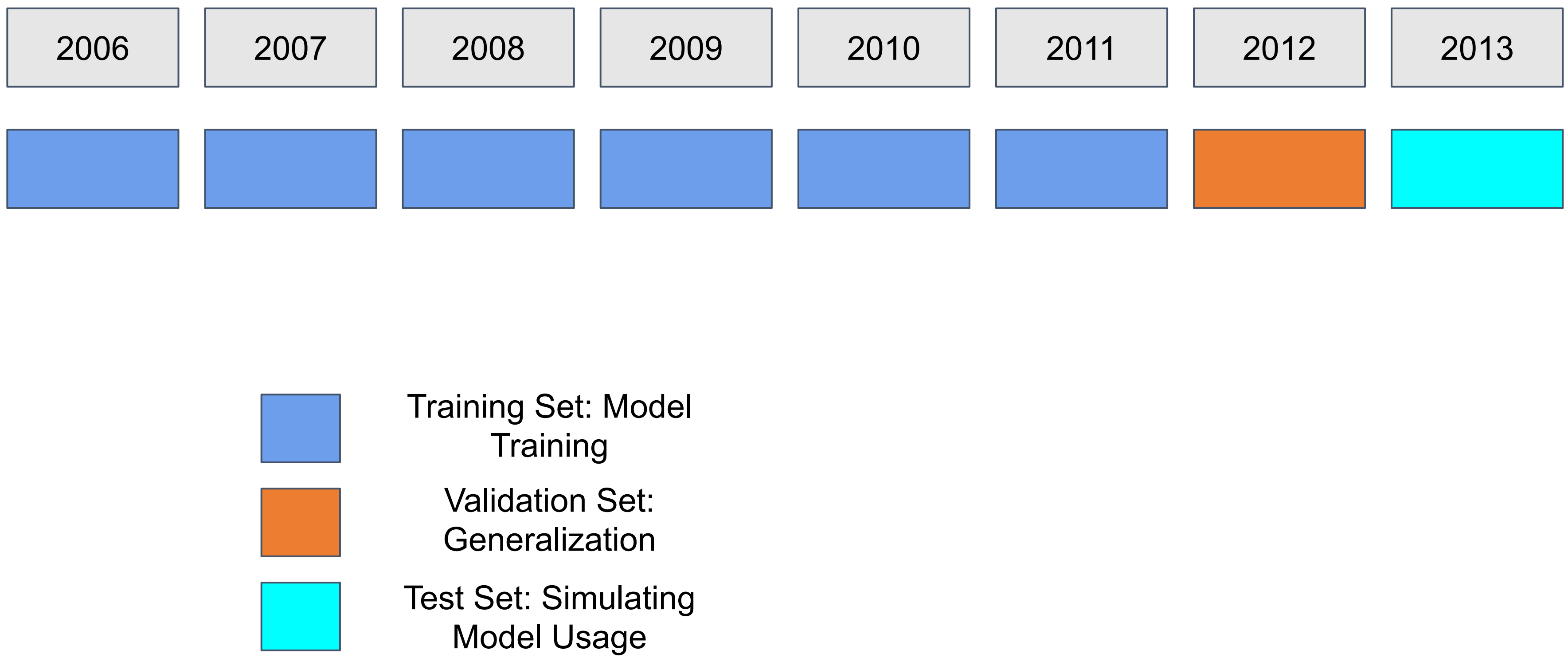
Purwadhika
Startup and Coding School

Metrics

Same as regression metrics :

- R-square
- MSE
- MAE
- MSPE
- MAPE
- MSLE
- etc

Data Splitting



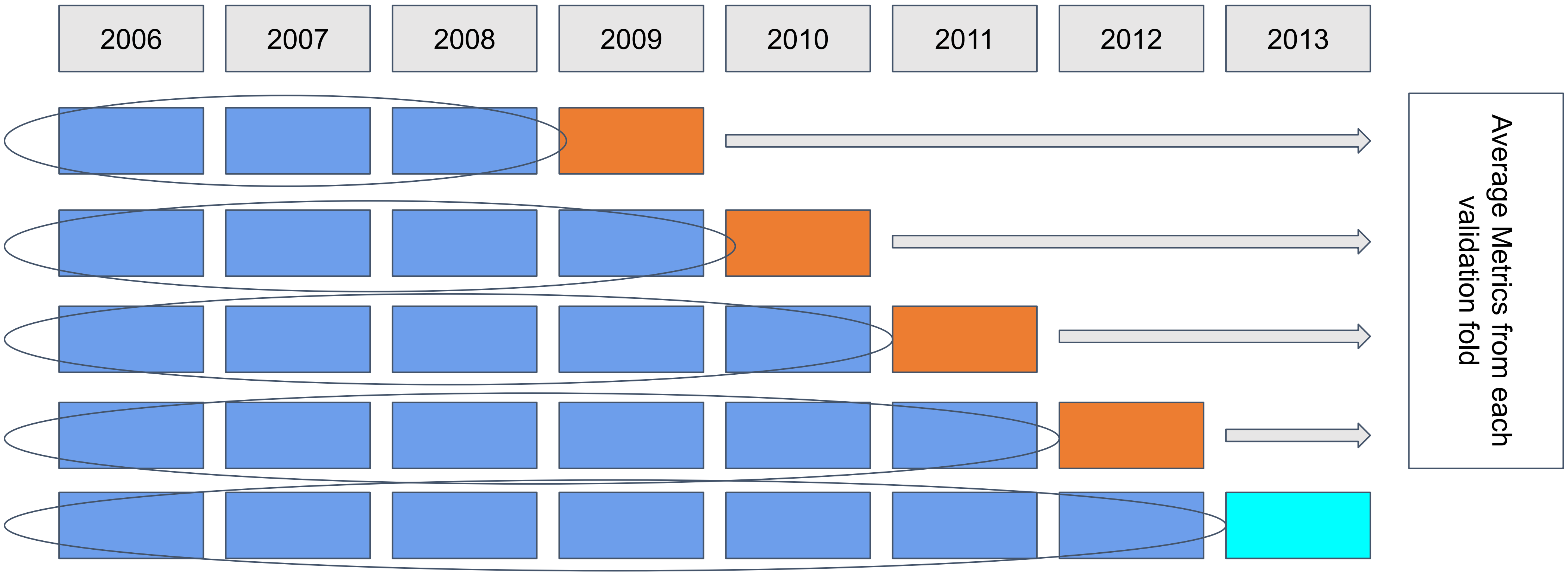
Forward Chaining Strategy

Training Fold: Model Training

Validation Fold: Generalization

Simulating Model Usage

Some kind of cross validation in time series



Purwadhika
Startup and Coding School

Python Exercise : Time Series with Feature Engineering

Analyze data opsd_germany_daily.csv

- build a time series model using linear regression
  - target : Consumption
  - feature : Date
  - FE Date 1 : year, month, day, weekday
  - FE Date 2 : year, month, day, weekday, year 2009, year > 2014, christmas, winter
- Split data
  - training : 2006 - 2015
  - testing : 2016 - end
- Compare the result (FE1,FE2) using following evaluation metrics :
  - explained variance
  - mean square log error
  - r2
  - MAE
  - MSE
  - RMSE

- plot test data, FE Date 1 forecasting result, FE Date 2 forecasting result

Purwadhika
Startup and Coding School

Python Exercise : Time Series Evaluation Method

Continue Analyze data opsd_germany_daily.csv

- With FE Date 2, try several models and find the best model based on R-square in forward chaining strategy (5 splits)
- those models are : ridge, lasso, elastic net, SVR
- optimize the best model based on R-square using hyperparameter tuning
- check the final performance : explained variance, mean square log error, r2, MAE, MSE, RMSE
- plot test data, FE Date 1 forecasting result, FE Date 2 forecasting result, FE Date 2 (Tuned Model) forecasting result

Purwadhika
Startup and Coding School

References

https://machinelearningmastery.com/time-series-datasets-for-machine-learning/

https://medium.com/@ODSC/machine-learning-for-time-series-data-e3971d38005b

https://www.stat.ipb.ac.id/en/uploads/STK352/STK352_10.pdf

https://towardsdatascience.com/time-series-modeling-using-scikit-pandas-and-numpy-682e3b8db8d1

https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why

https://www.ethanrosenthal.com/2018/03/22/time-series-for-scikit-learn-people-part2/

https://otexts.com/fpp2/case-studies.html

https://www.slideshare.net/ElegantJ-BusinessIntelligence/what-are-data-trends-and-patterns-and-how-do-they-impact-business-decisions

Purwadhika
Startup and Coding School