

The background is a dark blue-grey color. It is decorated with various geometric shapes in orange and white. These include circles of different sizes, some with dotted patterns inside; hexagons, some solid orange and some outlined; triangles; and lines. Some shapes are partially cut off by the edges of the frame. There are also horizontal and vertical dotted lines. The overall aesthetic is modern and minimalist.

Time Series & Stock Analysis

By Aurellia Christie
Group Leader at Supertype

01.

What is Time Series Data?

02.

Time Series Data Visualization

03.

Forecasting Time Series Data

04.

How to Get Stock Data

05.

Stock Analysis

06.

Forecasting Stock Price



What is Time Series Data?

Time Interval

...
Daily, monthly,
yearly, etc.

Pattern

...
Constant, trend,
seasonal, etc.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432

`ts(data, start, end,
frequency) or
xts package`
...

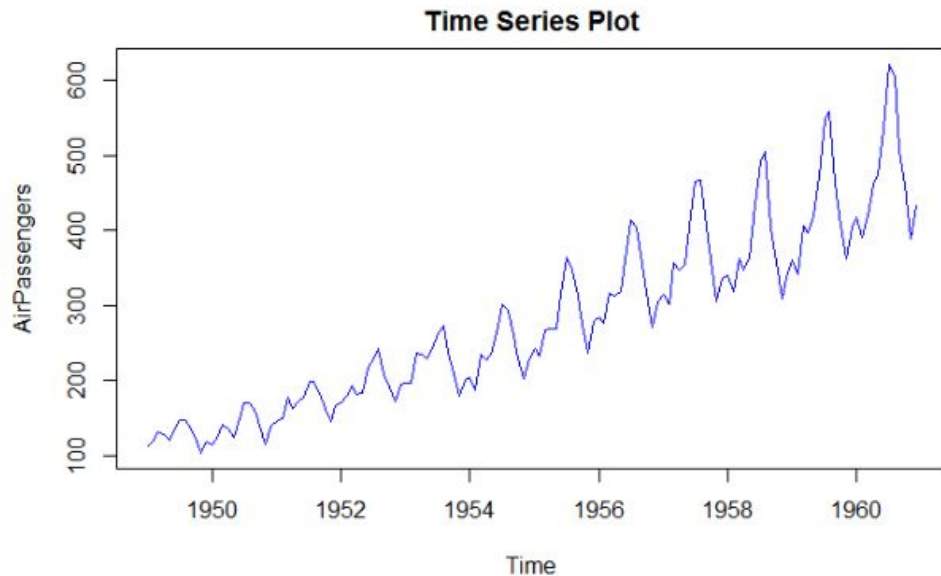
Create time series
object in R

`window(ts-data,
start, end)`
...

Subset the data



Time Series Data Visualization



- `plot(ts-data, col, main):`
 - x-axis : time
 - y-axis : observation
- Identify the patterns

Forecasting Time Series Data

Training Data

- Data used to build up the model
- Should be more than 50 % of data (70-80%)

Testing Data

- Data used to validate the model built
- Must not include the training data

Forecasting Time Series Data

Naive Forecasting

- Forecast = data from the previous period
- `naive(data, h=...)`

MA(q) Process

- Forecasting model with the past residual as the predictor
- $y_t = a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + c$
- `arima(data, order=c(0,0,q))`

AR(p) Process

- Forecasting model with the past observation as the predictor
- $y_t = b_0 + b_1 y_{t-1} + \dots + b_p y_{t-p} + e_t$
- `arima(data, order=c(p,0,0))`

ARIMA Model

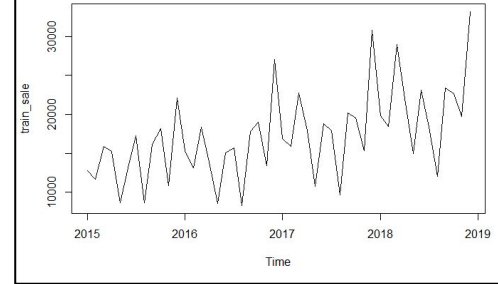
- Combination of AR, differencing, and MA
- `auto.arima(data)`

Forecasting Time Series Data

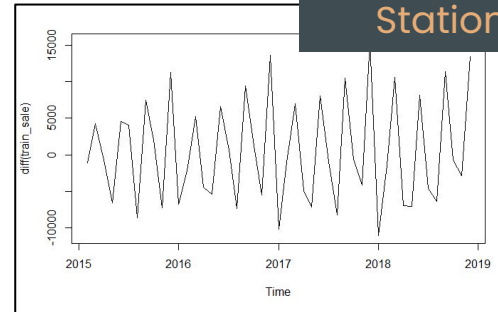
Stationarity

- Constant mean
(data fluctuates around a mean value over time)
- Differencing

Non Stationer



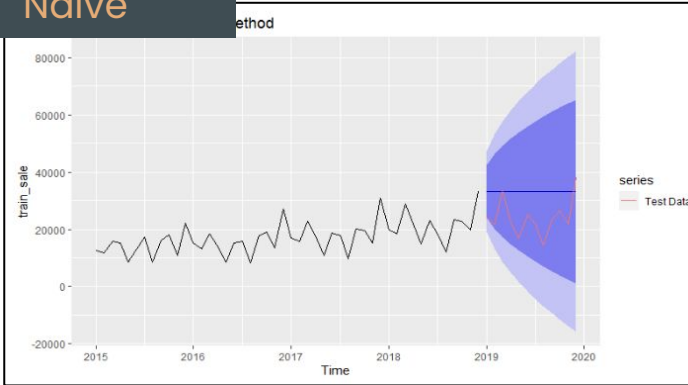
Stationer



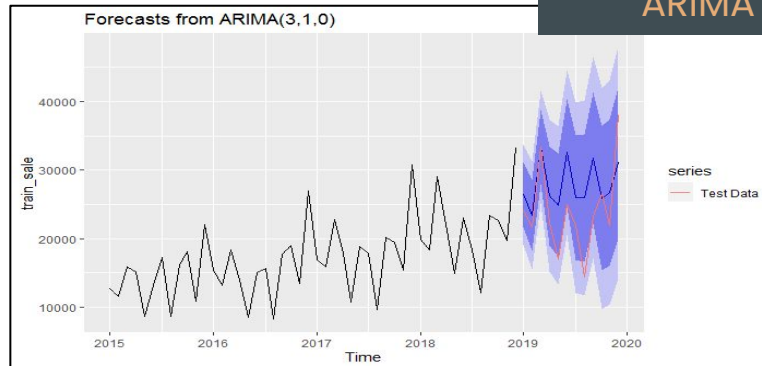
Forecasting Time Series Data

Monthly Sale

Naive



ARIMA



Forecast Evaluation

checkresiduals(model)

Residual

- The difference between the forecast data and the actual data
- Have to be randomly distributed without an obvious pattern

Assumptions

- Normally distributed (mean=0) :
 - Normal curve
- Have a constant variance :
 - Residual plot
- Have no autocorrelation :
 - ACF plot
 - Ljung-Box Test :
 - H_0 : no autocorrelation
 - H_1 : there is autocorrelation
 - P-value > alpha 0.05

Forecast Evaluation

Monthly Sale

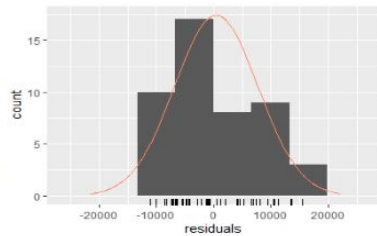
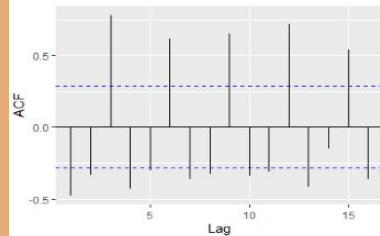
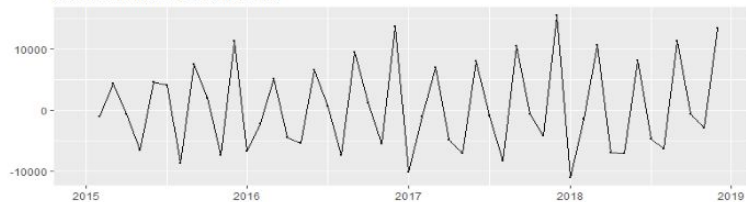
Naive

Ljung-Box test

data: Residuals from Naive method
 $Q^* = 132.17$, $df = 10$, $p\text{-value} < 0.00000000000000022$

Model df: 0. Total lags used: 10

Residuals from Naive method



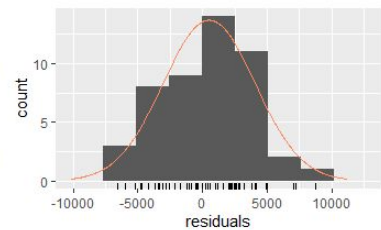
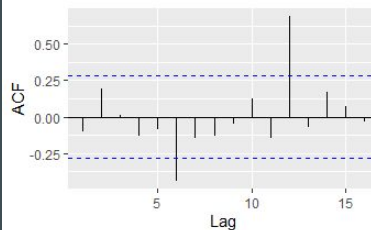
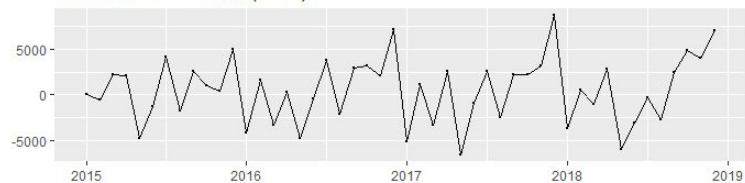
Ljung-Box test

data: Residuals from ARIMA(3,1,0)
 $Q^* = 18.276$, $df = 7$, $p\text{-value} = 0.01079$

Model df: 3. Total lags used: 10

ARIMA

Residuals from ARIMA(3,1,0)



Forecast Evaluation

accuracy(model, testing data)

RMSE

Root Mean Squared Error

MAPE

Mean Absolute Percentage Error

MAE

Mean Absolute Error

MASE

Mean Absolute Scaled Error

Forecast Evaluation

Monthly Sale

Naive

	ME	RMSE	MAE
Training set	435.7021	7222.766	6144.468
Test set	-9077.9167	10958.031	9916.917
	MPE	MAPE	MASE
Training set	-7.376794	37.79182	2.382706
Test set	-46.357012	48.57149	3.845588
	ACF1	Theil's U	
Training set	-0.47882874	NA	
Test set	-0.09319801	1.25836	

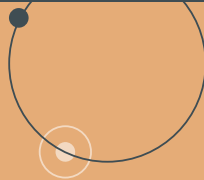
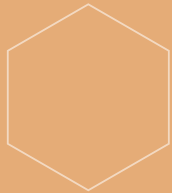
ARIMA

	ME	RMSE	MAE
Training set	510.3676	3554.848	2950.114
Test set	-3726.0933	6047.061	5023.016
	MPE	MAPE	MASE
Training set	-0.9610355	18.07454	1.143997
Test set	-20.6906129	24.23912	1.947828
	ACF1	Theil's U	
Training set	-0.1000926	NA	
Test set	0.1138575	0.7967281	

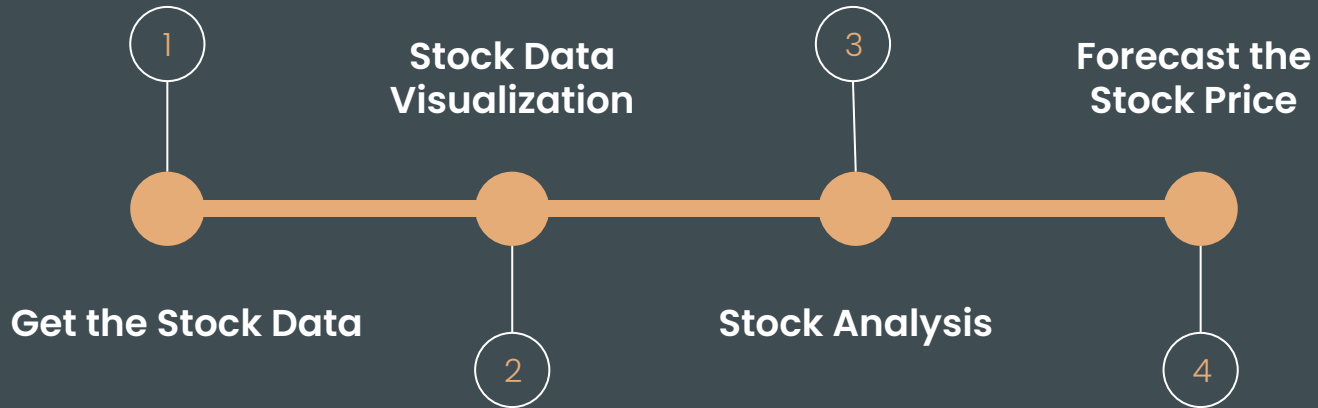


ANOTHER REAL WORLD APPLICATION

Stock Analysis



Workflow



How to Get Stock Data



**quantmod
package**



**getSymbols
(ticker, auto.assign =
FALSE, from=..., to=...)**



xts object



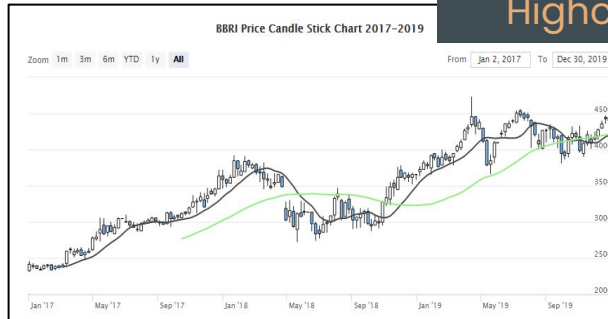
Stock Data Visualization

Plot



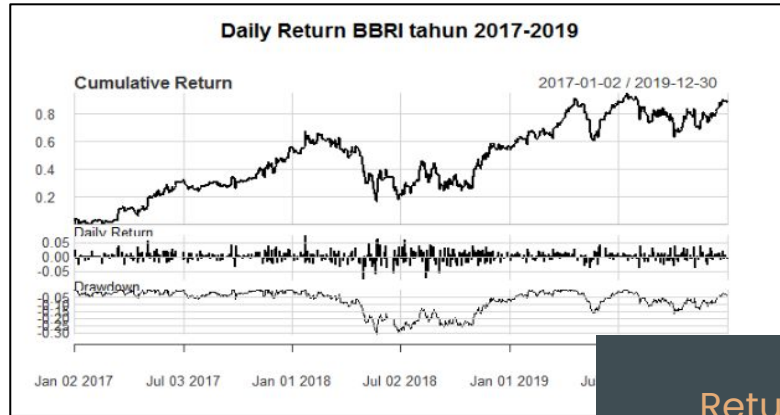
- Use the closing price data : `CI(data)`
- Alternatives:
 - `plot()`
 - `highchart()`

Highchart



Stock Data Analysis

Compare with other
stock prices

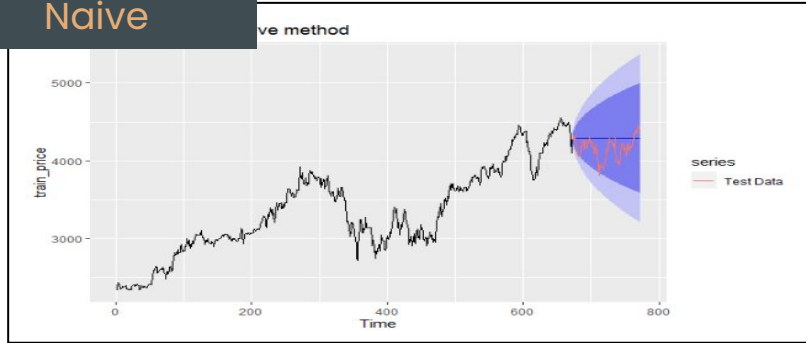


Return Analysis

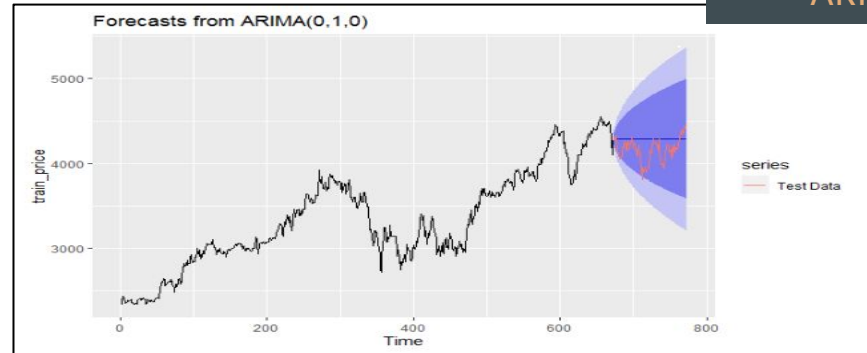


Forecasting Stock Price

Naive



ARIMA



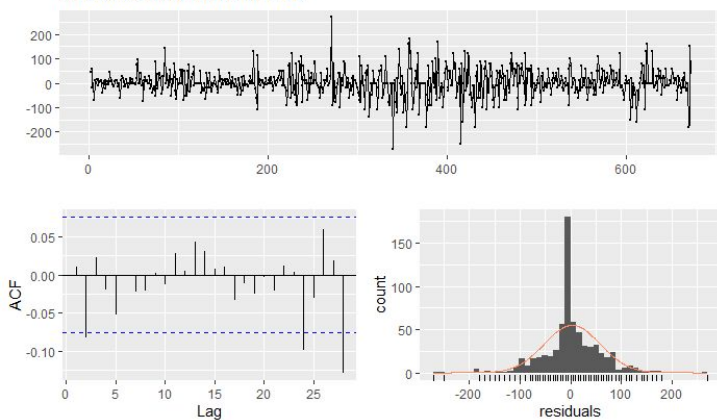
Forecast Evaluation

Naive

Ljung-Box test

data: Residuals from Naive method
 $Q^* = 7.7907$, $df = 10$, $p\text{-value} = 0.6493$
Model df: 0. Total lags used: 10

Residuals from Naive method

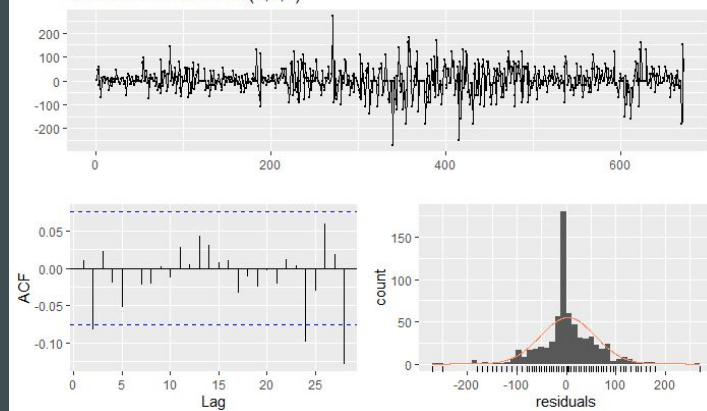


ARIMA

Ljung-Box test

data: Residuals from ARIMA(0,1,0)
 $Q^* = 7.8026$, $df = 10$, $p\text{-value} = 0.6481$
Model df: 0. Total lags used: 10

Residuals from ARIMA(0,1,0)



Forecast Evaluation

Naive

	ME	RMSE	MAE
Training set	2.909226	55.04361	37.35863
Test set	-126.100000	182.06867	146.10000
	MPE	MAPE	MASE
Training set	0.07657466	1.131002	1.000000
Test set	-3.13232520	3.587806	3.910743
	ACF1		
Training set	0.01062694		
Test set	NA		

ARIMA

	ME	RMSE	MAE
Training set	2.908373	55.00278	37.30659
Test set	-126.100000	182.06867	146.10000
	MPE	MAPE	MASE
Training set	0.07660946	1.129470	0.998607
Test set	-3.13232520	3.587806	3.910743
	ACF1		
Training set	0.01061502		
Test set	NA		



LET'S TRY!



References

- Supertype & Algoritma Learning Materials
- <https://otexts.com/fpp2/seasonal-arma.html>
- <https://www.aptech.com/blog/introduction-to-the-fundamentals-of-time-series-data-and-analysis/>
- <https://www.datacamp.com/community/blog/r-xts-cheat-sheet>
- <https://www.econometrics-with-r.org/14-ittsraf.html>
- <https://www.investopedia.com/trading/candlestick-charting-what-is-it/>
- <https://www.investopedia.com/ask/answers/121114/what-difference-between-golden-cross-and-death-cross-pattern.asp>
- <https://www.kaggle.com/podsyp/time-series-starter-dataset> (Monthly Sale.csv) with some adjustment

A vertical orange sidebar on the left side of the slide. It contains several geometric elements: a large circle with a small dark dot inside, a square with a diagonal line and a dotted pattern, a series of dots arranged in a grid, and a dark circle. The background of the sidebar is a solid orange color.

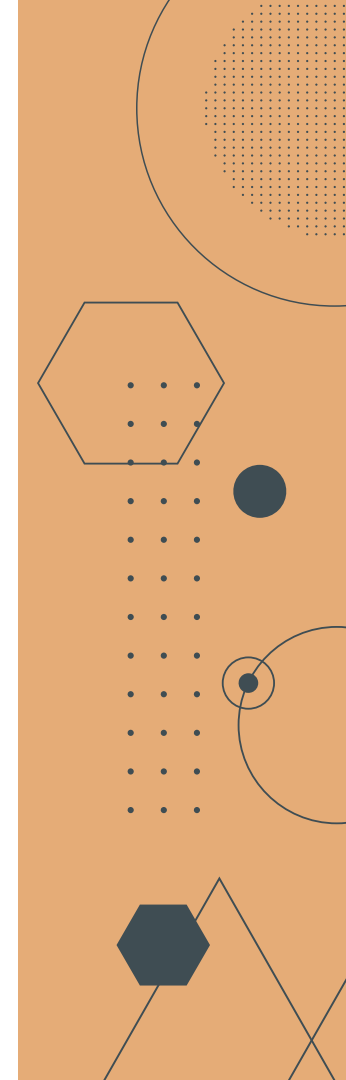
Thanks!

Do you have any questions?

aurelliachristie77@gmail.com

LinkedIn : Aurellia Christie

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

A vertical orange sidebar on the right side of the slide. It contains several geometric elements: a large circle with a dotted pattern inside, a hexagon with a dotted pattern inside, a dark circle, a series of dots arranged in a grid, and a dark hexagon. The background of the sidebar is a solid orange color.