Introduction to Time Series

Saarthak Mehra

Agenda

Intro to Time Series

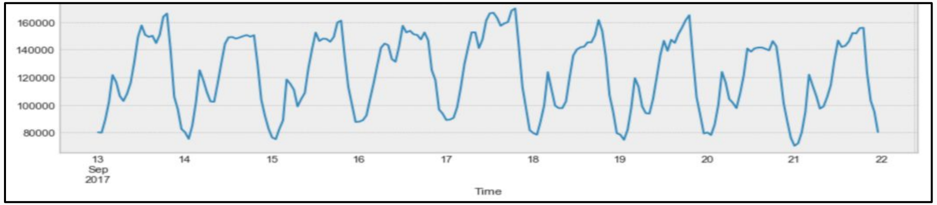
Time-Series Behavior

Forecasting Time-Series

What is time-series?

A sequence of observations taken over equal intervals over time.

What do they look like?



Univariate

datetime	A
2018-01-01 00:00:00	35
2018-01-01 01:00:00	2
2018-01-01 02:00:00	67
2018-01-01 03:00:00	7
2018-01-01 04:00:00	61
2018-01-01 05:00:00	94
2018-01-01 06:00:00	63
2018-01-01 07:00:00	90
2018-01-01 08:00:00	56
2018-01-01 09:00:00	27

Multivariate

datetime	A	B	
2018-01-01 00:00:00	35	NaN	104.0
2018-01-01 01:00:00	2	NaN	104.0
2018-01-01 02:00:00	67	104.0	104.0
2018-01-01 03:00:00	7	76.0	76.0
2018-01-01 04:00:00	61	135.0	135.0
2018-01-01 05:00:00	94	162.0	162.0
2018-01-01 06:00:00	63	218.0	218.0
2018-01-01 07:00:00	90	247.0	247.0
2018-01-01 08:00:00	56	209.0	209.0
2018-01-01 09:00:00	27	173.0	173.0

Why do we care?

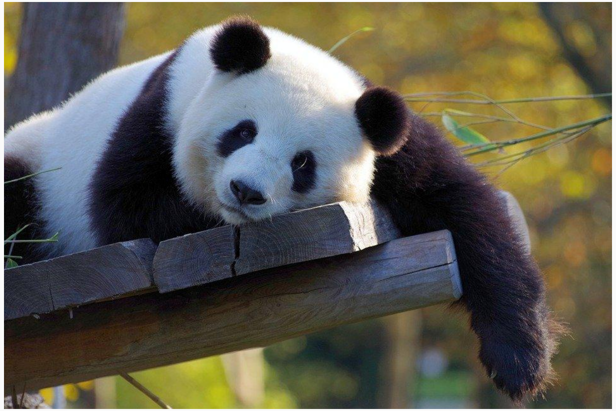
Time series can be analyzed in **any** business that operates an online platform (website, blog, etc.)

Trivial Time Series Analysis

Around 10 AM, there are **100,000** online users	→	At 10AM, we need all of our servers on.
Around 8 PM, there are **15,000** online users	→	At 8PM, we only need some of our servers on.

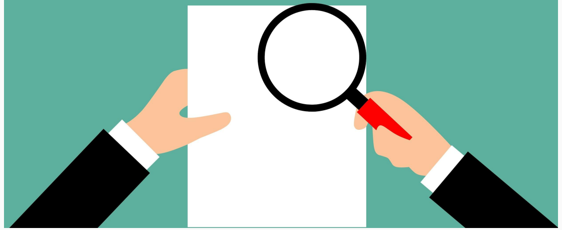
From the above observations, we are able to **save money** by **turning off an X amount of servers** at 8 PM, thereby **increasing efficiency**.

Important libraries for time-series modeling you need to know



Statsmodels

Provides tools for statistical modeling



Scikit Learn

Anything machine learning (from regression to clustering to parameter tuning)

Types of Time-Series Data

Deterministic vs Non-Deterministic

Additive vs Multiplicative

Stationary vs Non-Stationary

Deterministic vs Non-Deterministic

Deterministic time series can be expressed with a single equation.

More about it:
Future values can easily be predicted.

Variance remains constant
Mean changes with time

More about it:
Harder to predict future values.

Variance changes time
Mean changes time

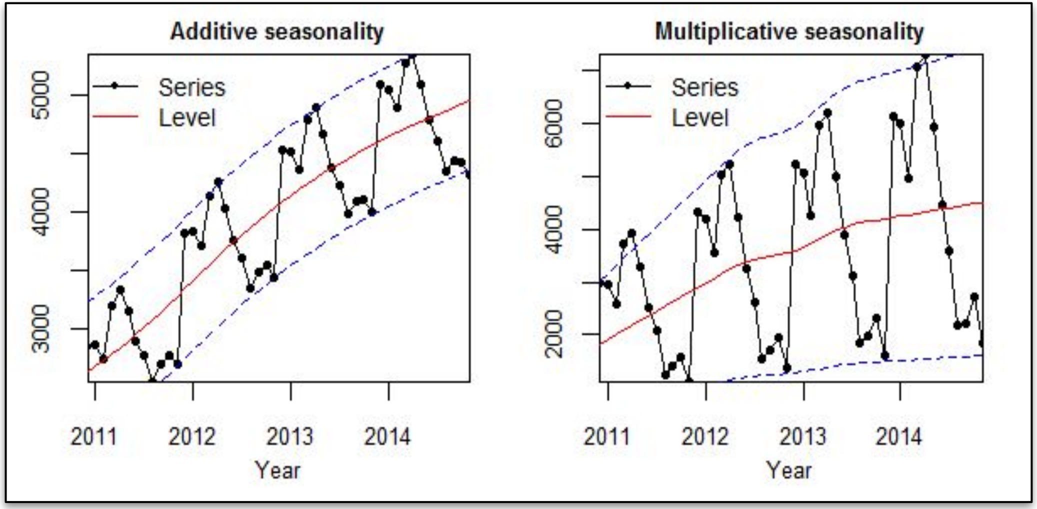
Non-Deterministic time series can **not** be expressed with a single equation

Additive vs Multiplicative

Additive:
<u>sum</u> of the components;
tends to show a linear trend;
data is measured in absolute quantity

Multiplicative:
<u>product</u> of the components;
tends to show an exponential trend;
Data is measured in percent change

Data can be modeled as either additive or multiplicative



Stationary vs Non-Stationary

Stationary

Statistical properties are constant

- Mean
- Variance
- Autocorrelation ➜ correlation of the series with its previous values

Independent of time

Easier to predict future values

No seasonality

Non-Stationary

Trend ➜ A general pattern for a long time period

- Example: A general increase in the number of UBER rides called in San Francisco.

Seasonality ➜ Repeating pattern in the data during a fixed time period

- Example: An increase in the number of UBER rides on Friday nights in downtown San Francisco.

Irregularity ➜ Non-repeating behavior within a short time period

- Example: A sudden decrease in the number of UBER rides due to a sudden snowstorm.

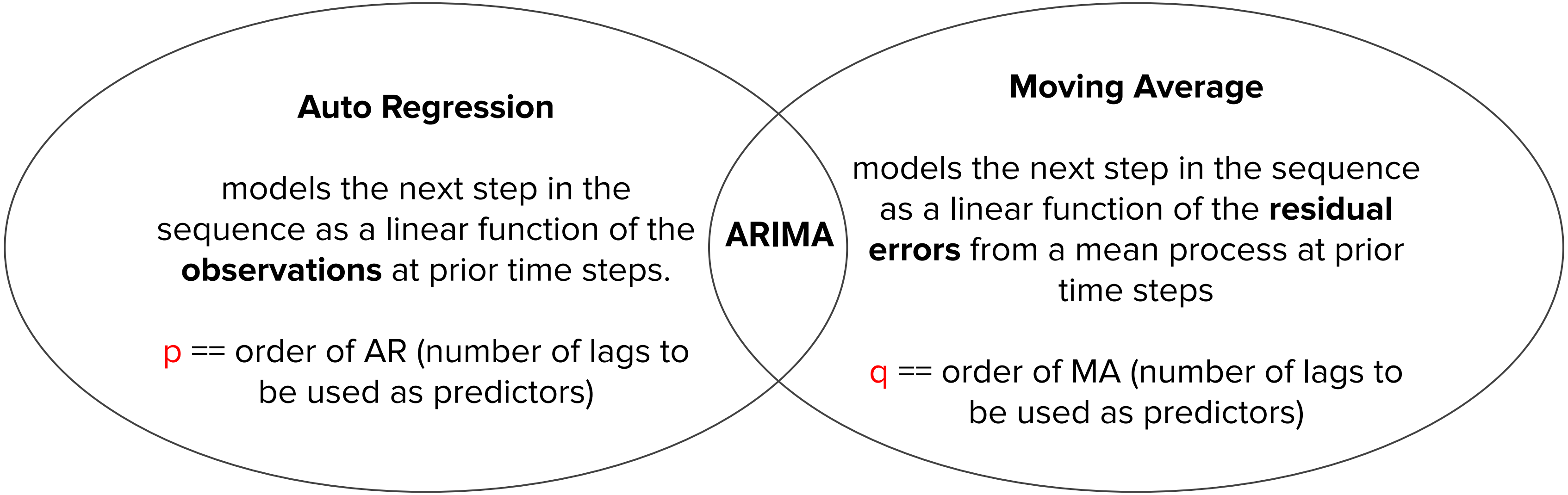
Cyclic ➜ Repeating behavior occurring at irregular time intervals

- Usually caused by the business itself or other socioeconomic factors

Time-Series Forecasting

1. Attempting to make the data stationary
2. Selecting the appropriate model
3. Evaluating model accuracy

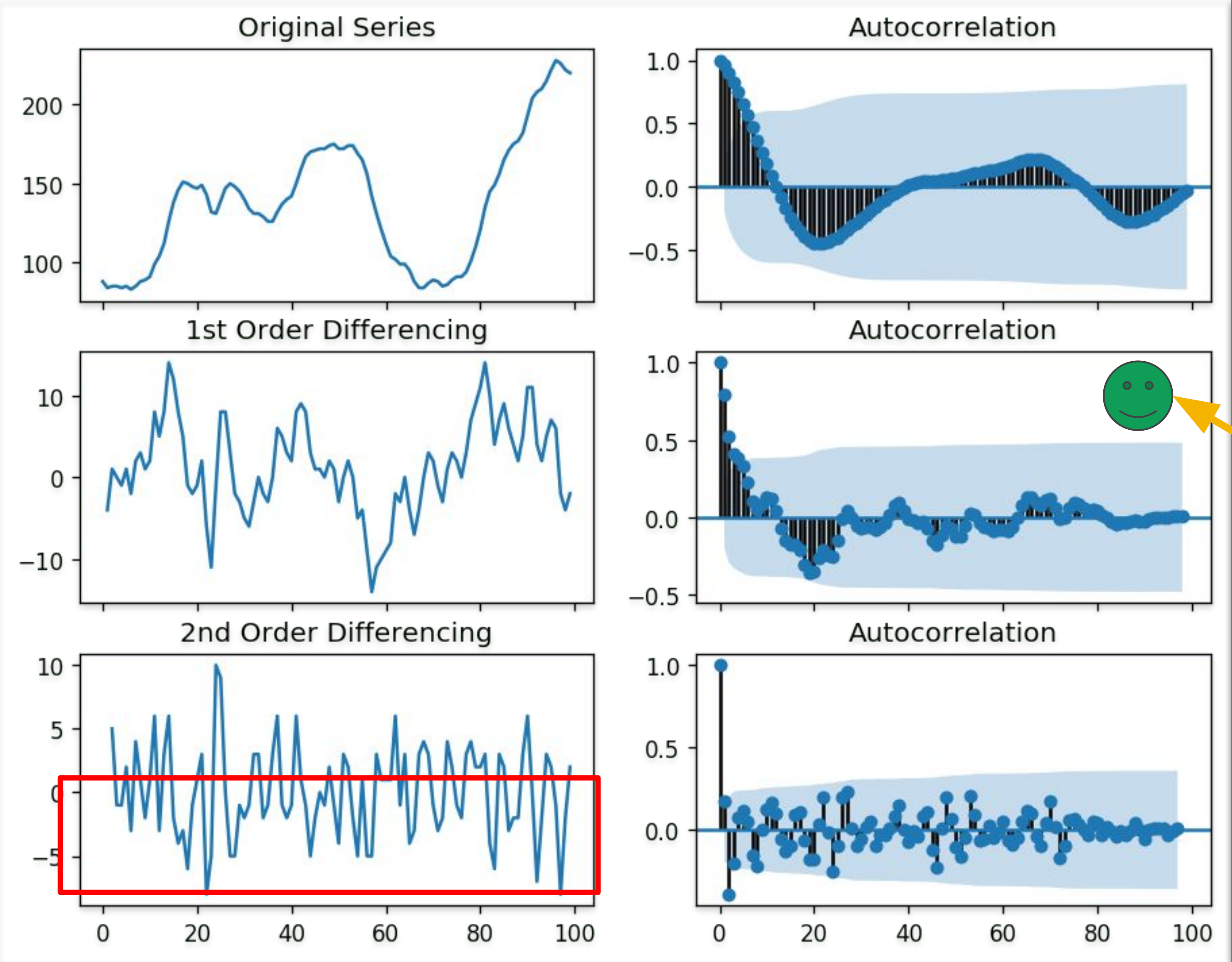
Time-Series Models: ARIMA



ARIMA → works well with time series with trend but without seasonality

The “**I**” in ARIMA represents the number of differences (d) made in order to make time-series stationary

Differencing?



Purpose of differencing?
To minimize autocorrelation while the time-series remains greater than 0

"Weak stationarity"

In the second order of differencing, autocorrelation went negative too quick, which suggests...

← overdifferencing

Time-Series Models: SARIMA

If our data is measured in days, and the data shows a repeating pattern every seven days (**weekly seasonality**), then $m = 7$. So at every future time step...

AR(p), I(d), and AM(q) will be dependent on every time step in the past
AR(P), I(D), and AM(Q) will be dependent on past time steps from every 7 days (week)

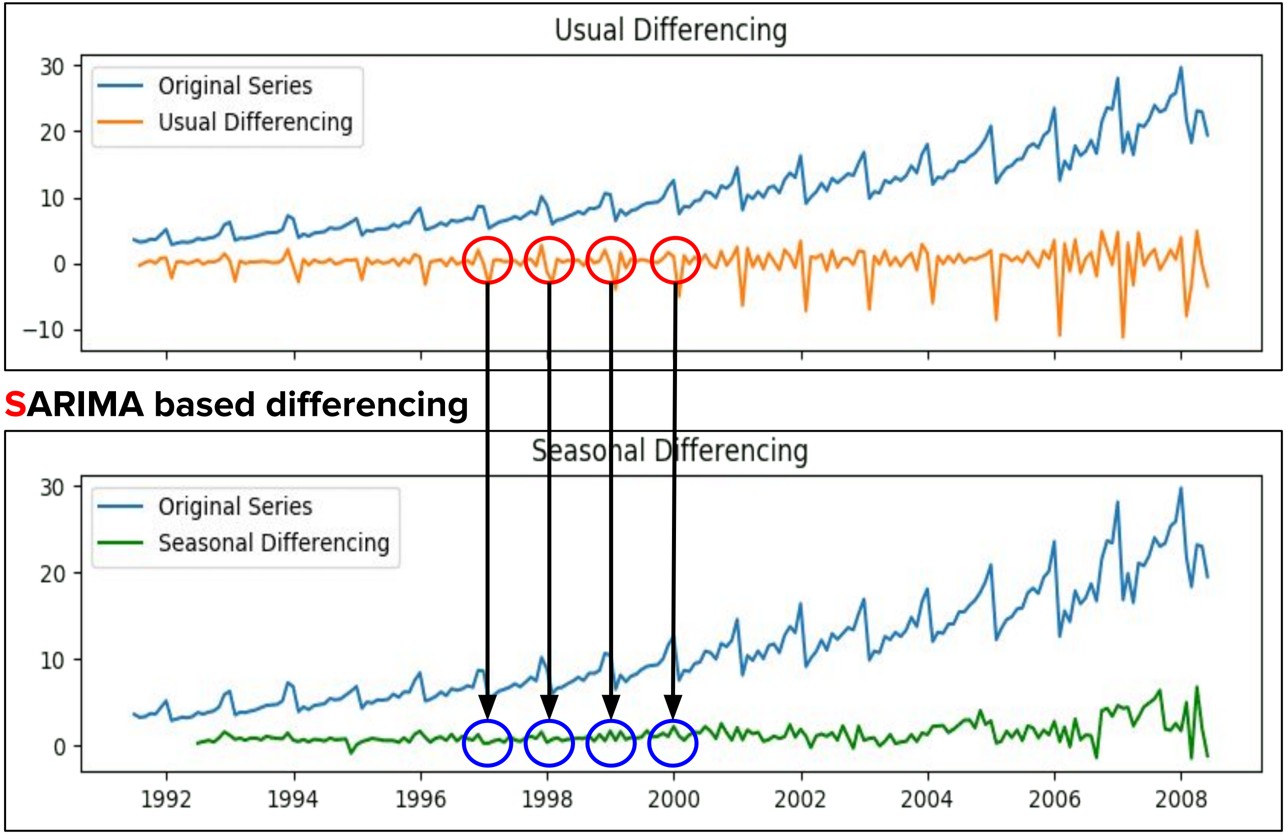
Auto Regressive	**Integrated**
Moving Average	**Seasonality**

SARIMA(p, d, q)(P, D, Q)m

- p, d, q → remember from ARIMA
- m → number of time steps in each season
- P, D, Q → same as (p, d, q), but only active based on m

ARIMA based differencing

SARIMA based differencing



With seasonality represented as a parameter, you can repeat the ability to perform:

i) autoregression
ii) differencing
iii) moving average

modeling like ARIMA did at the seasonal level.

More Time-Series Models

SARIMAX: Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors
VAR: Vector Autoregression **VARMA:** Vector Autoregression Moving-Average **VARMAX**: Vector Autoregression Moving-Average with Exogenous Regressors
SES: Simple Exponential Smoothing **HWES**: Holt Winter’s Exponential Smoothing
FbProhet: Open source model made by Facebook; takes holidays into account
LSTM: Long Short Term Memories
NNETAR: Neural Network Autoregression

Evaluating Model Accuracy

Accuracy Metrics

MAPE → Mean Absolute Percentage Error

ME → Mean Error

MAE → Mean Absolute Error

MPE → Mean Percentage Error

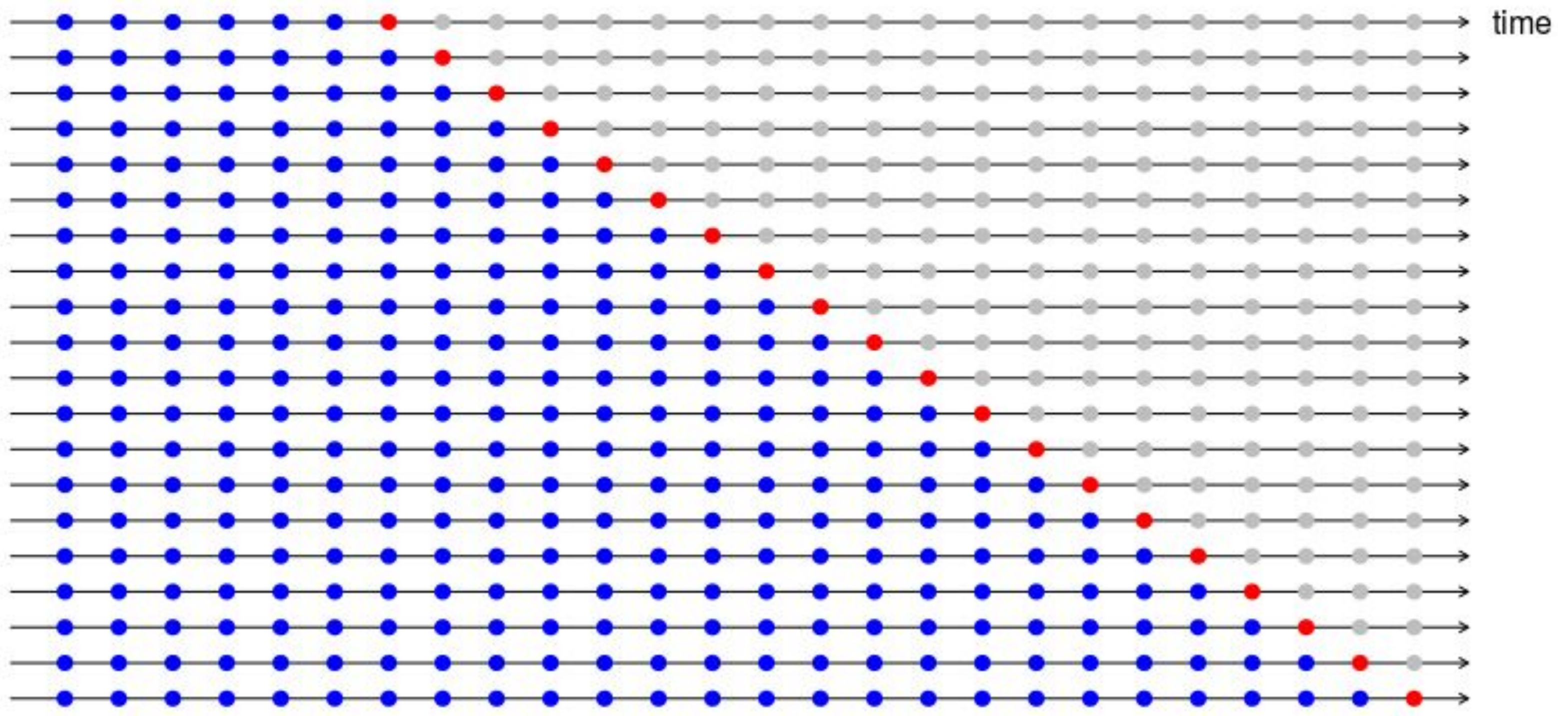
RMSE → Root Mean Squared Error

ACF1 → Lag 1 Autocorrelation of Error

corr → Correlation between the Actual and the Forecast

Minmax → Min-Max Error

Interesting/Useful Information



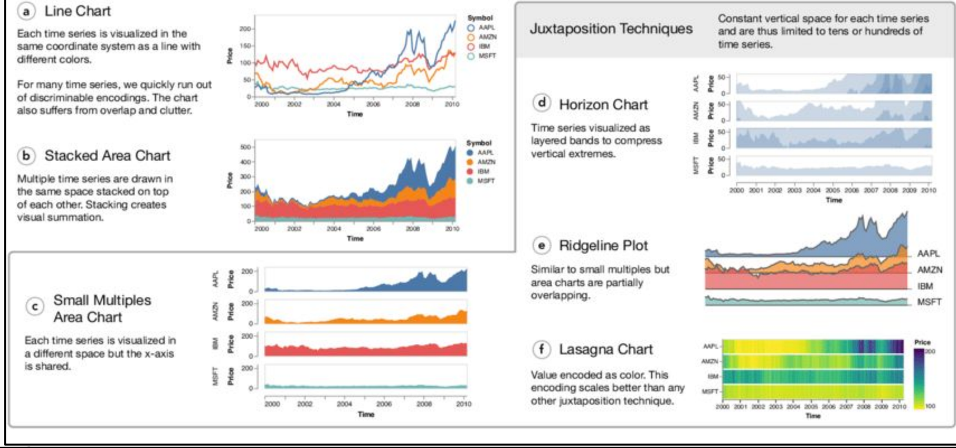
Cross-Validation technique used to validify model before testing it on the entire training set

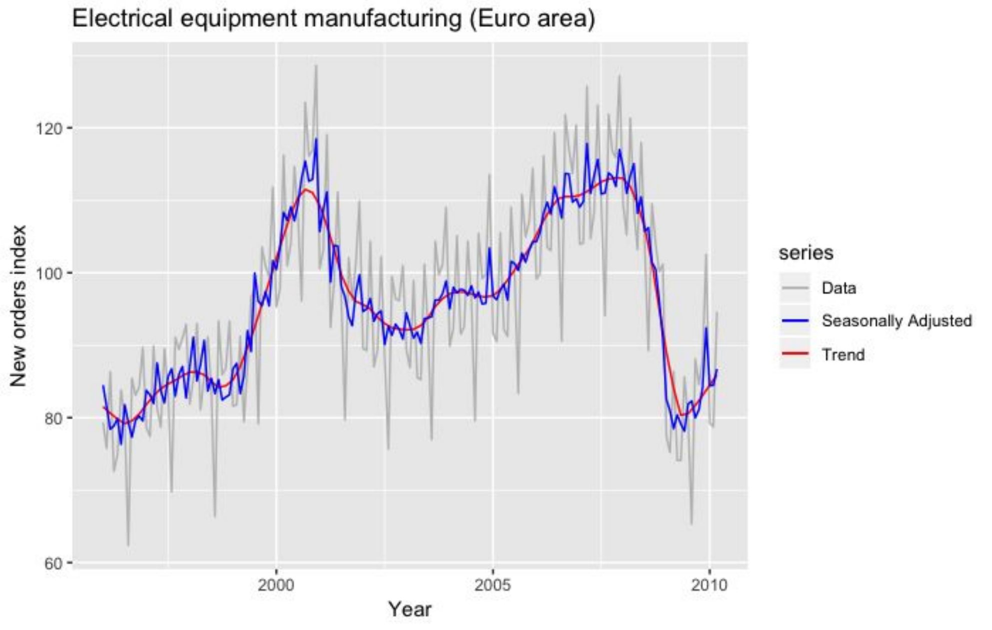
Blue → training set
Red → validation set

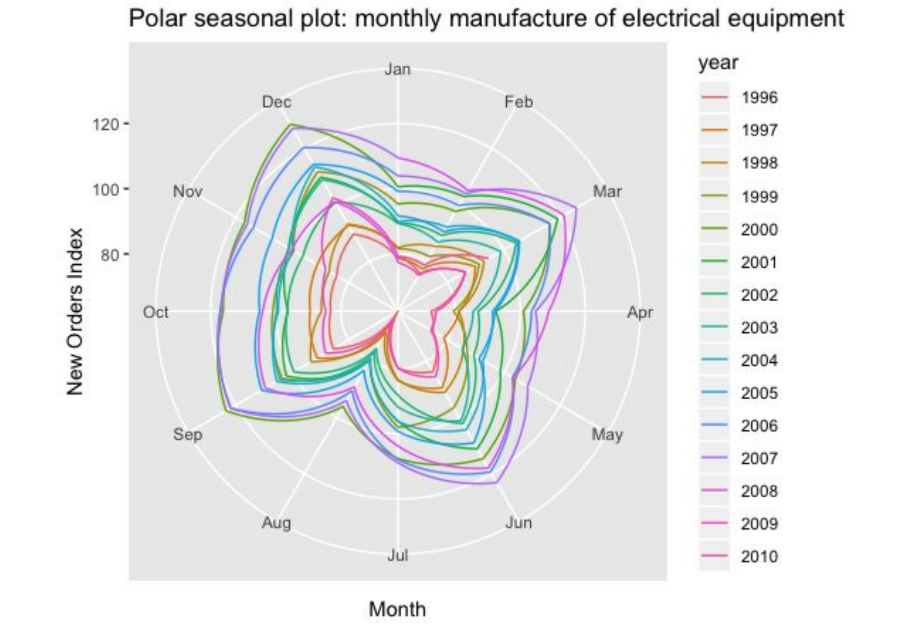
Visualizations

Avoid sticking to stock visualizations from matplotlib

Ensure that every visualization is practical, innovative, and creative







Thank you for attending!