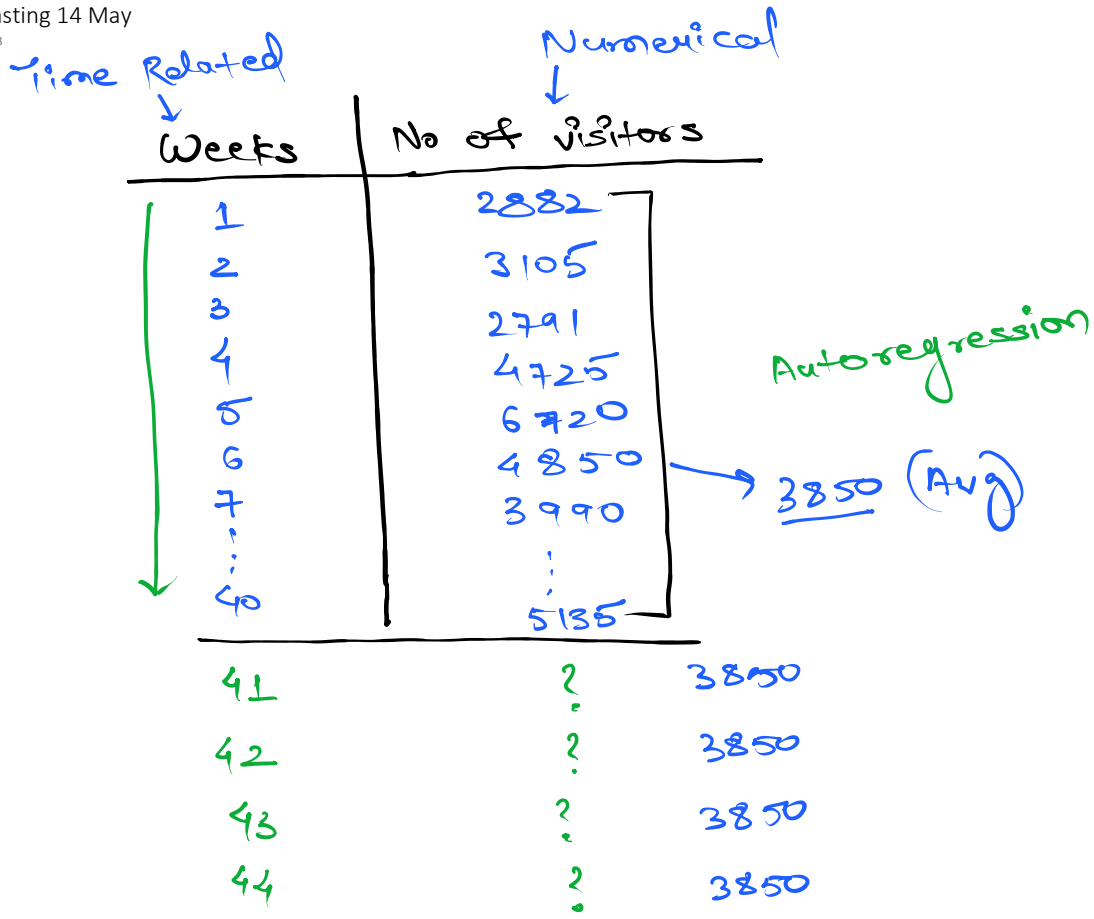
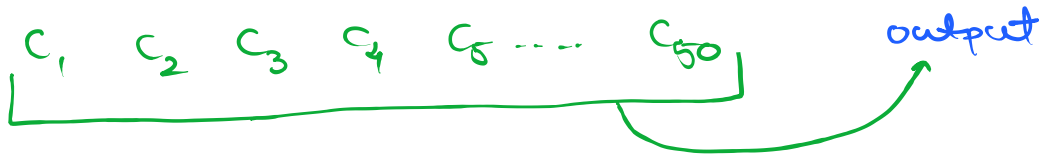
Time Series Forecasting 14 May

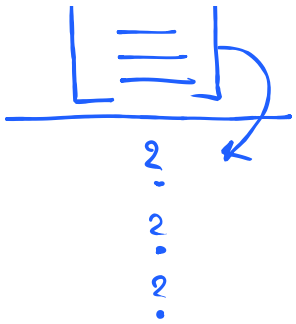
Wednesday, 14 May 2025 07:03



Regression



Autoregression



Ways of Time Series Forecasting →

Quantitative

→ Data available

→ Historical patterns repeat mostly.

→ we can easily capture complex patterns through available data.

Qualitative

- No data available
- we have no idea about the historical pattern.
- We can't capture/identify the complex patterns because of unavailability of data.

3 Components of Time Series Forecasting →

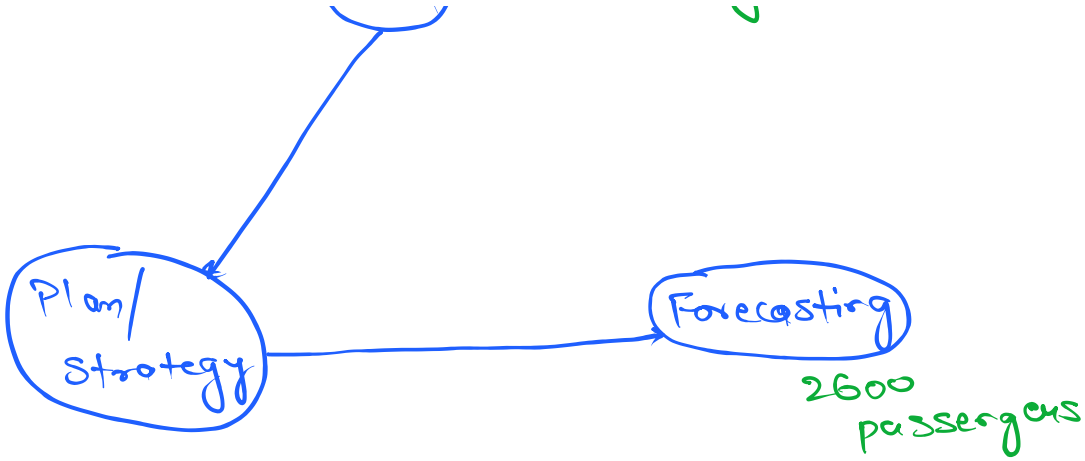
① Time Series Data.

② Time Series Analysis.

③ Time Series Forecasting.

Basic Terminologies

Goal occupancy → 90%.



A1 : ~1000

A2 : ~2000

A3 : ~3000

Steps in Forecasting

→ Define the problem statement.

→ Collect the data

→ Analyze the data.

→ Build and evaluate the forecasting models.

Some caveats associated with Time-Series forecasting →

(1) The granularity rule: The more aggregate your forecasts are, the more accurate they will be.

(2) The frequency rule: Keep updating your

forecasts regularly to capture any new trend/information that comes.

(3) The Horizon Rule: When you have forecasted for many future weeks/months, your forecasts are more likely to be very accurate in the earlier weeks/months as compared to the later ones.

Three important characteristics of a time-series data:

(1) Relevant: Data should be relevant to our goal/objective.

(2) Accurate: Data should be accurate in terms of capturing the timestamps and the related observations.

(3) Long Enough: Data should be long enough to forecast accurately. This is important to identify all the patterns in the past.

Basic Approaches for Time Series Forecastings →

(1) Naive Approach →

Forecasted value = Last observed value

(2) Simple Average Approach

(3) Moving Average Approach

(4) weighted moving average approach.

x_1	x_2	x_3	x_4	x_5
2000	2100	1980	3150	2950

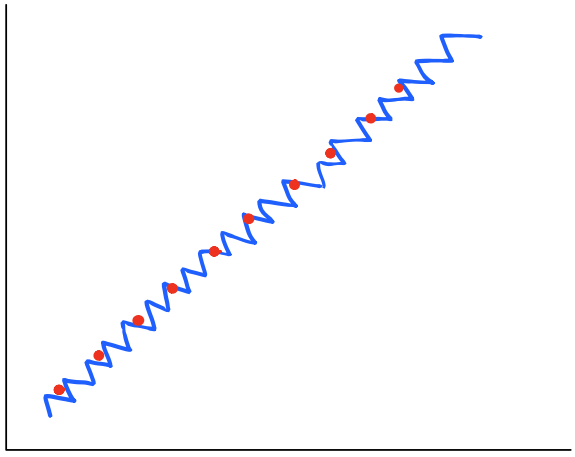
$$\text{Avg} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

$$\text{weighted avg} = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5}{5}$$

Homework (30%) - Score: 85
- Quizzes (20%) - Score: 90
- Final Exam (50%) - Score: 80
- Multiply each score by its respective weight:
- Homework: (85 \times 0.3 = 25.5)
- Quizzes: (90 \times 0.2 = 18)
- Final Exam: (80 \times 0.5 = 40)
- Multiply each score by its respective weight:
- Homework: (85 \times 0.3 = 25.5)
- Quizzes: (90 \times 0.2 = 18)
- Final Exam: (80 \times 0.5 = 40)

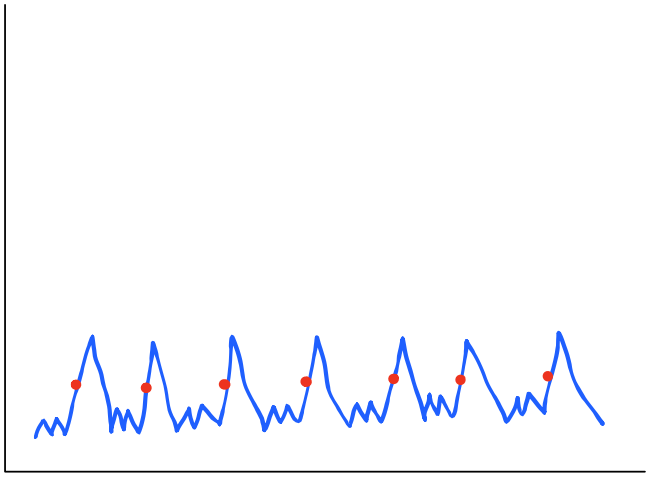
Some examples of a time series data plot →

Upward Trend + Seasonality



varying Mean & Constant variance → Non-stationary Time Series Data.

No Trend + Seasonality

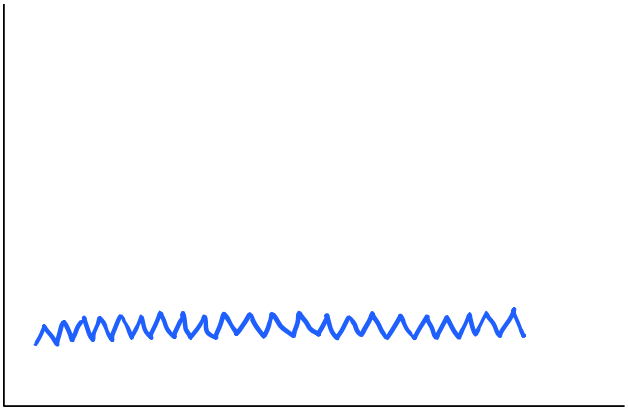


Constant Mean & varying variance → Non-Stationary Data.

No Trend

+

Seasonality



Constant Mean
&
Constant variance
→ Stationary Data.

Every Time Series Algorithm; requires the data given to it to be 'Stationary'.

If the data is "Non-Stationary", then we perform necessary transformation on the data to make it 'Stationary'.

Time Series Transformations →

① Differencing → To be used when your data has varying mean but constant variance.

Varying mean
+
Constant variance
→ Diff →
Constant Mean
+
Constant Variance

(2) Box-Cox Transformation → To be used when your data has constant mean but varying variance.

Constant Mean + varying variance → (BCT) → Constant Mean + Constant Variance

(3) Box-Cox Transformation + Differencing

varying Mean + varying variance → (BCT + diff) → Constant mean + Constant vari-ance

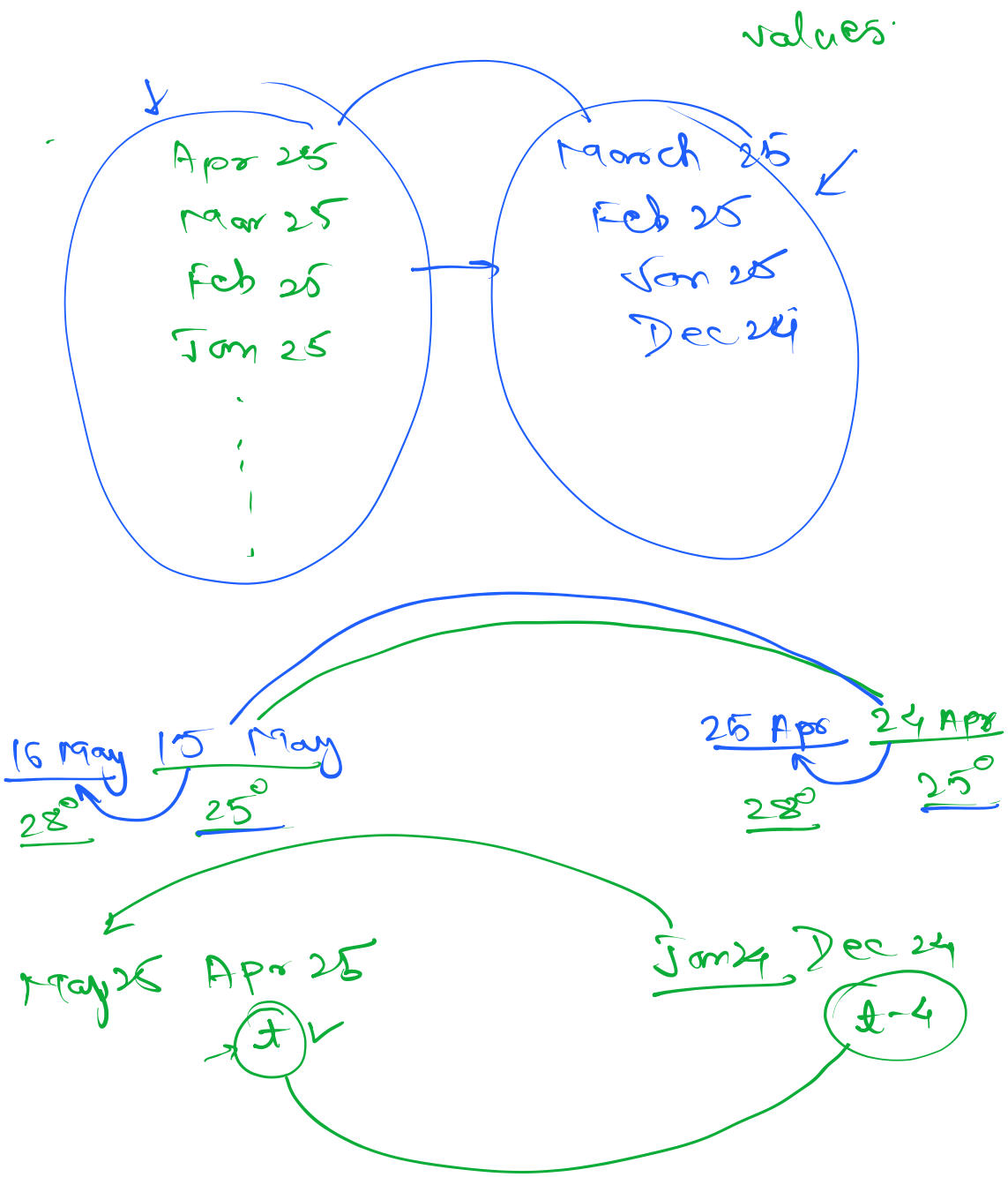
Autocorrelation Plots →

Sales — Profit

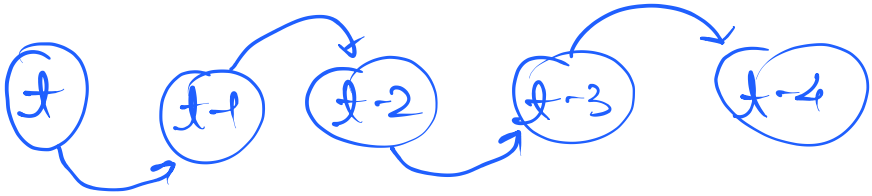
0.9

Sales

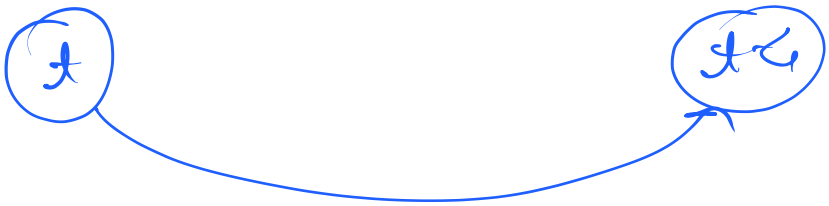
correlation of the sales column values with its own previous



ACF Plot



PACF plot



Time Series Transformation & Reversal Process

train test NS

t (months)	1	2	3	4	...	16	17	18	19	20
y (sales)(k)	25	7	19	31	...	51	38	15	46	50
y boxcox	3.22	1.95	2.94	3.43	...	3.93	3.64			
Y boxcox diff	-	-1.27	1.00	0.49	...	0.19	-0.29			
Y pred								0.03	0.25	0.18
Appended Series	-	-1.27	1.00	0.49	...	0.19	-0.29	0.03	0.25	0.18
Cumulative Series	-	-1.27	-0.27	0.22	...	0.71	0.42	0.45	0.70	0.88
Initial Value Adjustment	-	1.95	2.94	3.43	...	3.93	3.64	3.67	3.92	4.10
Exponential	-	7	19	31	...	51	38	38.98	49.83	59.73

AR
ARIMA
SARIMA

y-pred

y-pred in original form

$\log(25) = 3.22$

$e^{3.22} = 25$

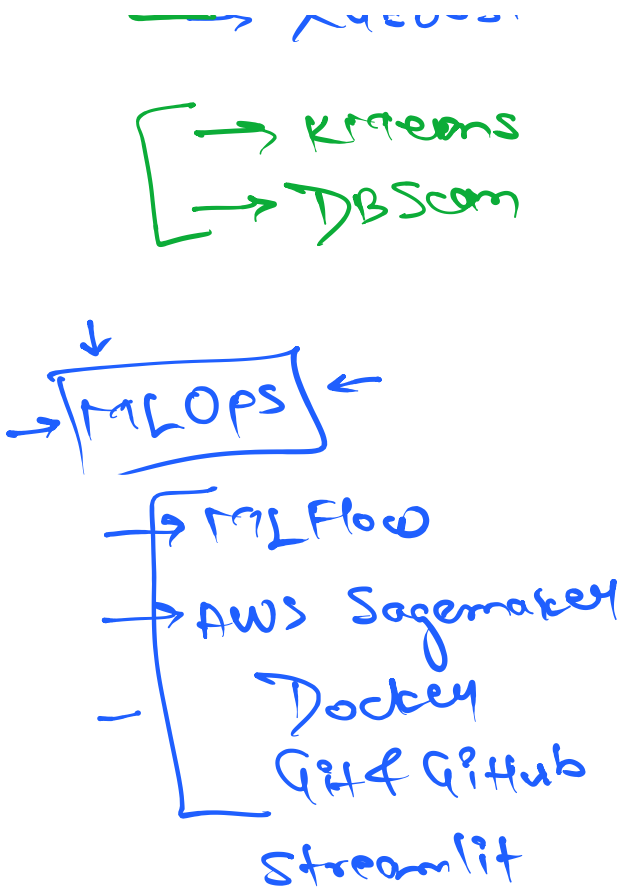
df_boxcox[0]

$e^{1.95} = 7$

$e^{2.94} = 19$

SQL
↓
Python
↓
statistics
↓
ML

- → Linear Reg
- → Log Reg
- → DT
- → RF
- → Gradient Boost
- → XGBoost



→ Personal Brand

Share knowledge

→ Substack