Forecasting Models

ARIMA, SARIMA, and Facebook Prophet

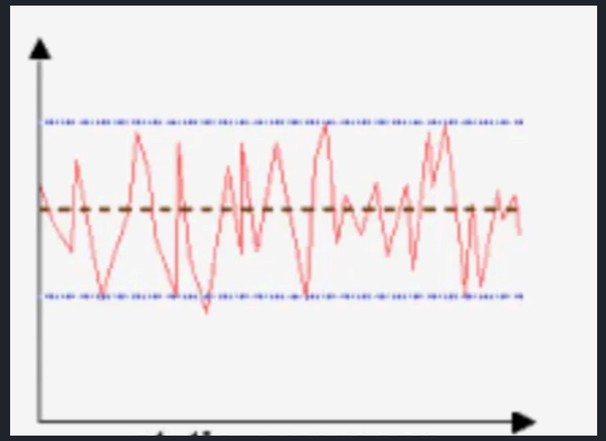
Done by: Safia Abalkhail.

Supervised By: Mohammad Modallal

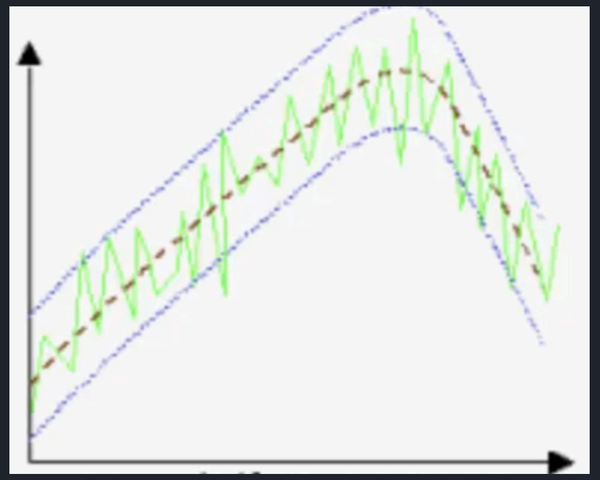
Outline

What Do We Mean by ‘Stationary’ & ‘Seasonal’ data?

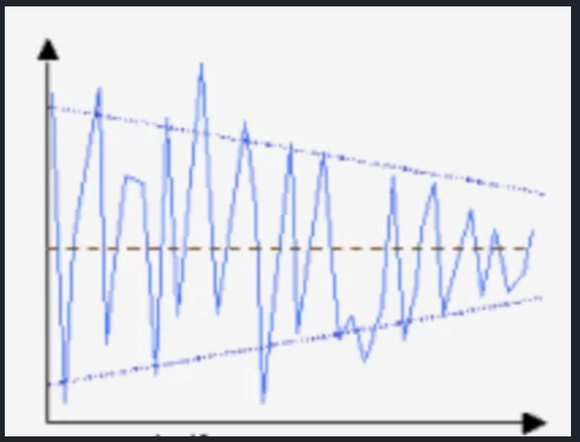
Stationary Mean
Stationary Variance



Non-Stationary Mean
Stationary Variance

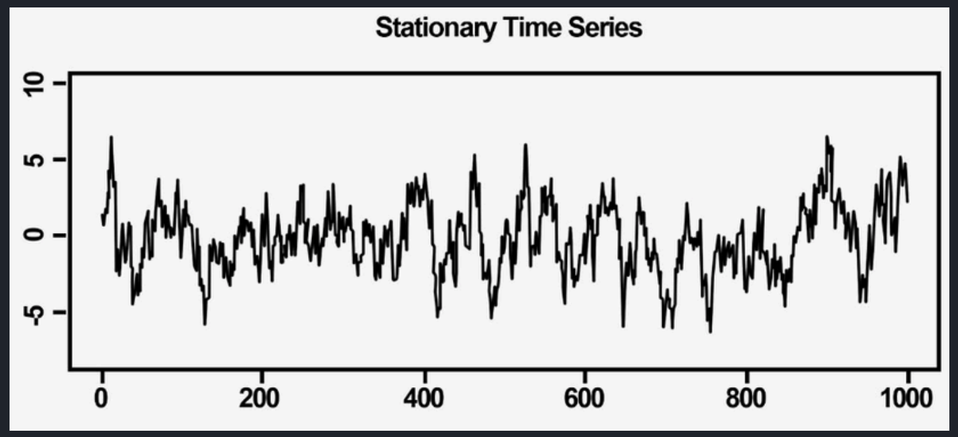


Stationary Mean
Non-Stationary Variance

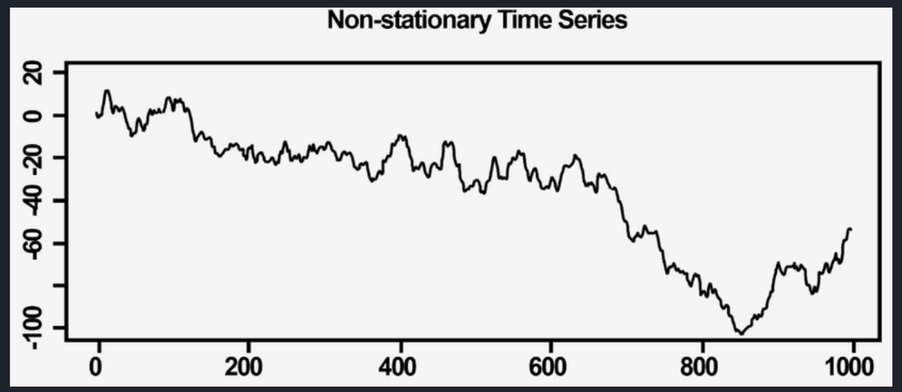


In Conclusion...

Stationary Data



Non-stationary/Seasonal Data



ADF (Augmented Dickey-Fuller) Test

Purpose:
Is a statistical test used to determine whether a time series is stationary or non-stationarity.

When Do De Use It?
Ensure stationarity of time series data for accurate modeling (e.g., ARIMA models).

Interpretation:
Reject (Null Hypothesis H0): The time series is stationary.
Fail to Reject (H0): The time series is non-stationary.

Example:

- Test Statistic: -3.50
- Critical Value (1%): -3.43
- P-Value: 0.05

Conclusion:
Since the test statistic is less than the critical value at the 1% significance level, we reject the null hypothesis and conclude that the time series is stationary. ☑️

ARIMA Algorithm

- Useful for data that shows patterns or trends over time.
- ARIMA captures standard temporal structures (patterned organizations of time) in the input database

ARIMA Algorithm Parameters:

p: The number of lag observations (previous time points) to include in the model.

d: The number of times the data needs to be differenced to make it stationary.

q: The number of lagged forecast errors to include in the model.

Suppose we have monthly sales data and want to build an ARIMA Model:

p=1: Use last month’s sales to predict this month’s sales

d=1: Data Is differenced ONCE to make it stationary

q=1: Use last month’s forecast error to adjust this month’s forecast.

Differencing..?

- **Differencing** is a method used in time series analysis to MAKE time series stationary.
- **Transforms** a time series by subtracting the previous observation from the current; stabilizing the mean

Original Time Series:

[3,5,8,12]

After d=1:

[2,3,4]

ARIMA Algorithm

Advantages	Disadvantages
Simple, easy to understand	Assumes linearity, and can struggle with nonlinear data
Many resources available	Requires stationary data

Use Case

Can be used with shorter to medium timezones.
AKA Several months to a few years.

Example:
Sales Forecasting: Short term sales prediction for inventory management

SARIMA Algorithm

- Stands for ‘Seasonal ARIMA’
- It extends ARIMA to capture the seasonality in the data

SARIMA Algorithm Parameters

Nonseasonal Components (= ARIMA)

p: The number of lag observations (previous time points) to include in the model.

d: The number of times the data needs to be differenced to make it stationary.

q: The number of lagged forecast errors to include in the model.

+

Seasonal Components (= SARIMA)

P: Seasonal AR order

D: Seasonal Differencing Order

Q: Seasonal MA Order

m: Seasonal Period Length

Suppose we have monthly sales data and want to build an SARIMA Model:

p=1: Use last month’s sales to predict this month’s sales

d=1: Data Is differenced ONCE to make it stationary

q=1: Use last month’s forecast error to adjust this month’s forecast.

p=1: Use the sales from the SAME MONTH LAST YEAR

D=1: Differencing with the SAME MONTH LAST YEAR

Q=1: Use the forecast error from the SAME MONTH LAST YEAR.

m=12: There are 12 months in a year.

SARIMA Algorithm

Advantages	Disadvantages
Handles Seasonality Well	More complex, due to additional parameters.
Retains ARIMA's strengths	Requires more computational resources, due to additional parameters

What makes it special?

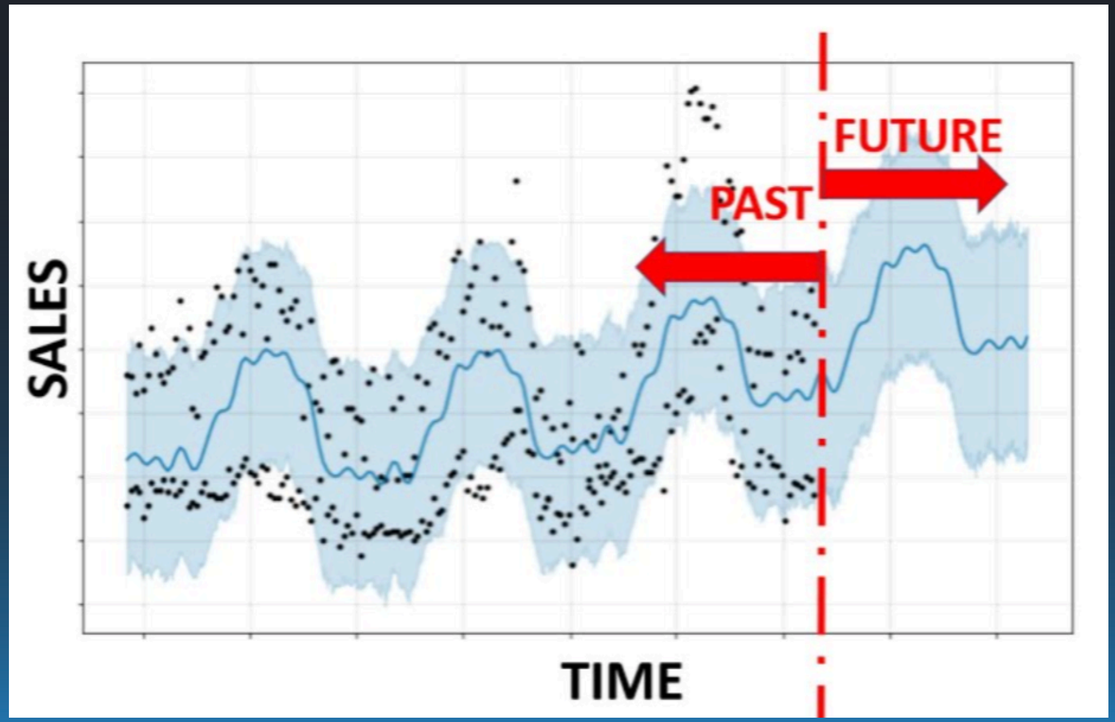
Flexibility:
SARIMA extends the ARIMA model by including both non-seasonal and seasonal components.
This flexibility allows it to model a wider range of time series patterns.

Use cases:

- **Retail Sales:** Predicting future sales for retail stores, especially when there are strong seasonal effects such as increased sales during holidays or specific seasons.
- **Patient Admissions:** Forecasting the number of patient admissions in hospitals which may have seasonal variations (e.g., flu season).
- **Energy Consumption:** Predicting electricity or gas demand where usage patterns often vary seasonally (e.g., higher consumption during winter or summer).

Facebook Prophet Algorithm

- Open-source released by **facebook** for forecasting time series data.
- Designed to handle various types of time series data, including data with **strong seasonal effects** & missing data.
- Especially suited for data with **clear seasonal patterns** and potential trend changes.



Facebook Prophet Parameters

1. growth= ‘linear’ or ‘Logistic’

2. yearly_seasonality = True or False:

- **Usage**: Do you want the model account for patterns that repeat every year?

3. weekly_seasonality = False or True:

- **Usage**: Do you want the model account for patterns that repeat every week?

4. daily_seasonality = False or True:

- **Usage**: Do you want the model account for patterns that repeat every day?

5. seasonality_mode = ‘additive’ or ‘Multiplicative’:

- **Description**: Specifies the type of seasonal component.

6. interval_width=0.95:

- **Description**: Sets the width of the uncertainty intervals.

Facebook Prophet Algorithm

Advantages	Disadvantages
Allows customizations for different seasonal patterns and holidays.	Slow on very large datasets, as the computational power increases.
Provides confidence Intervals to handle real-life uncertainty.	This makes it less efficient for real-time forecasting needs.

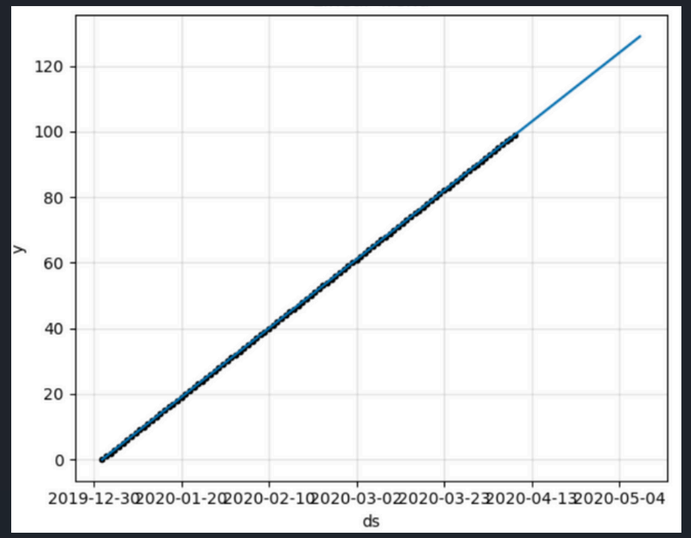
What makes it special?

Automatic Seasonality detection:

- Can automatically detect yearly, weekly, and daily seasonality patterns.
- Seasonality like SARIMA + holds account holidays and special events that might affect the data.

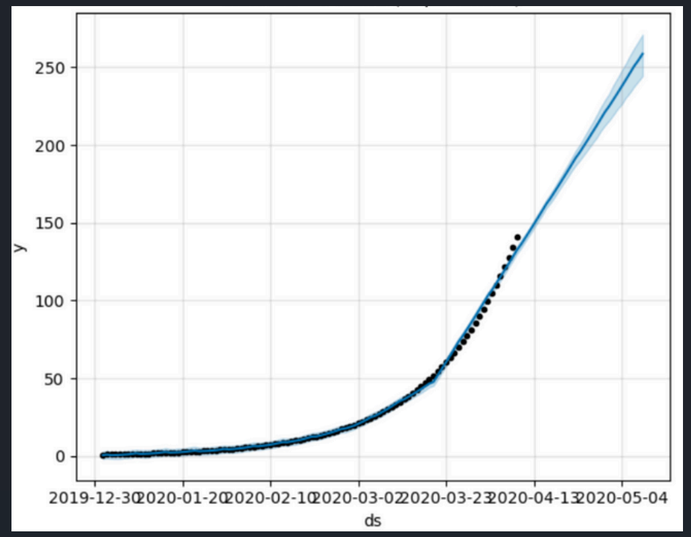
How Do I know when to use Facebook Prophet?

Linear Trend

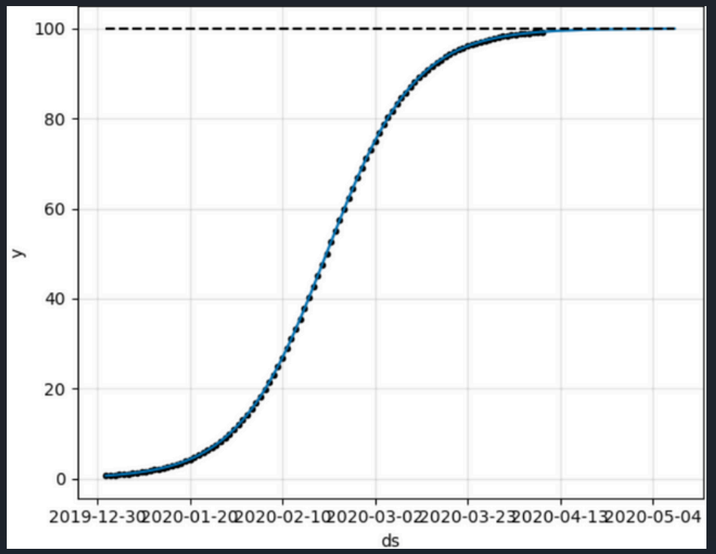


If you expect your sales to grow linearly.

Nonlinear Trend



Logistic Trend



If you expect your sales to grow BUT level off at a certain point.

Questions?