

# Build a Model for Anomaly Detection in Time Series Data

---

Introduction to Time Series Data and Anomaly Detection



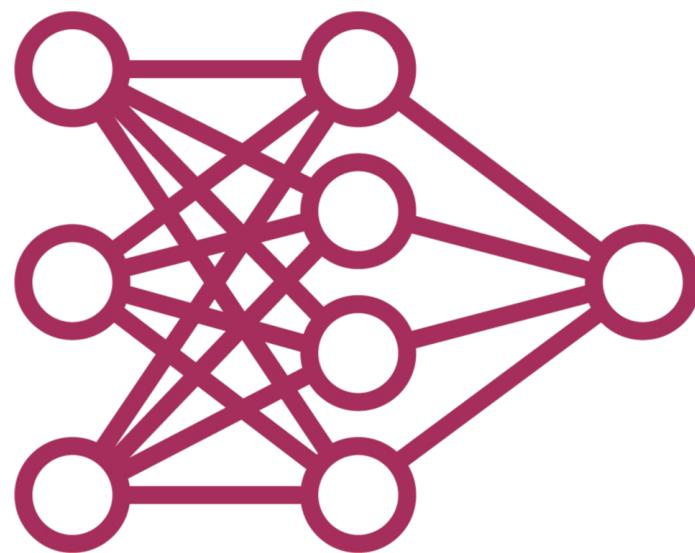
**Pratheerth Padman**  
Freelance Data Scientist



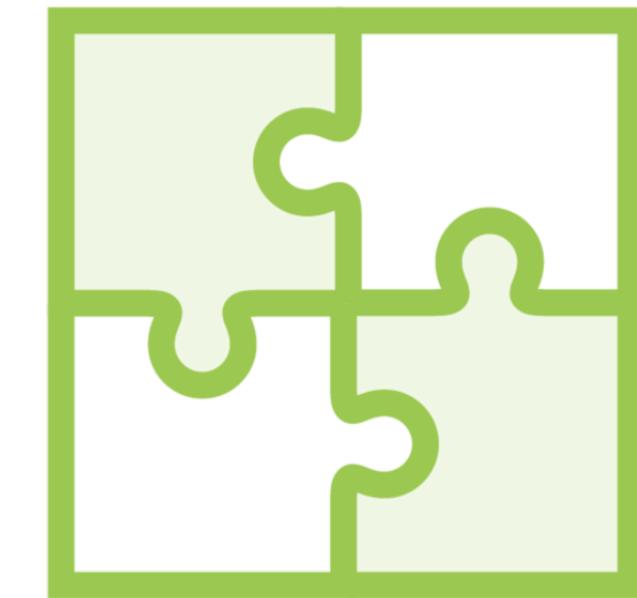
# Course Overview



**Intro to time series  
and anomaly  
detection**



**Building a model to  
automate anomaly  
detection**



**Dealing with the  
detected anomalies**



# Module Overview



**What is time series data?**

**Analysing time series data**

**Stationarity and autocorrelation**

**Introduction to anomaly detection**

**Demo: Setting up your environment**



# What Is Time Series Data?

---





# Time Series Data

**Time series data is a sequence or collection of data points obtained through repeated measurements over time.**



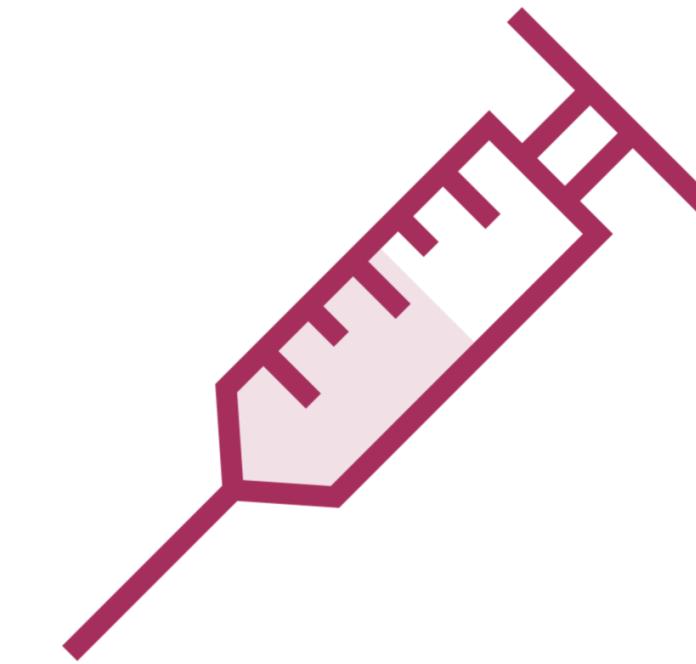
# Examples of Time Series Data



Daily stock prices



Daily  
temperature/weather  
records



People getting  
vaccinated each day



# Components of Time Series Data

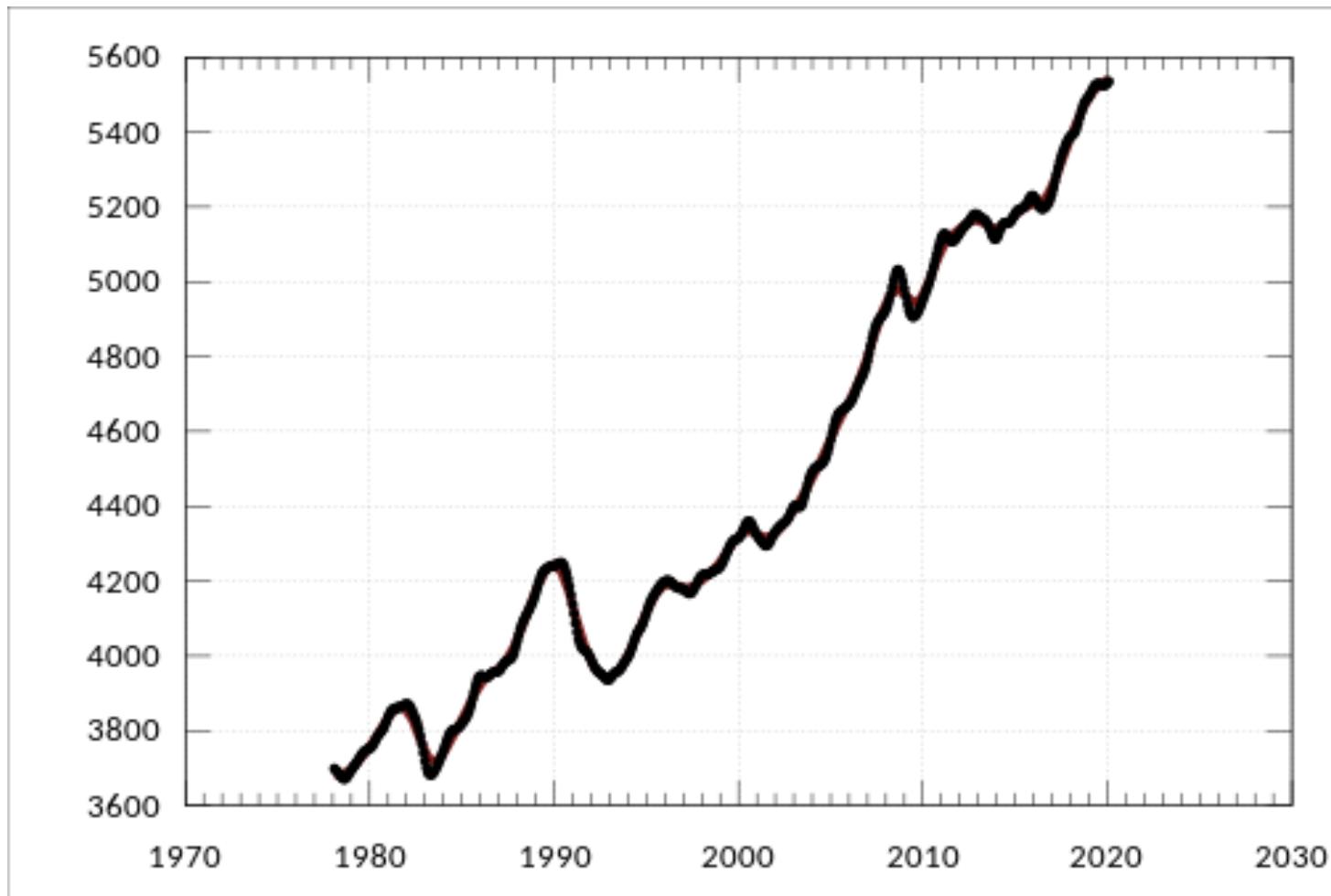
Trend

Seasonality

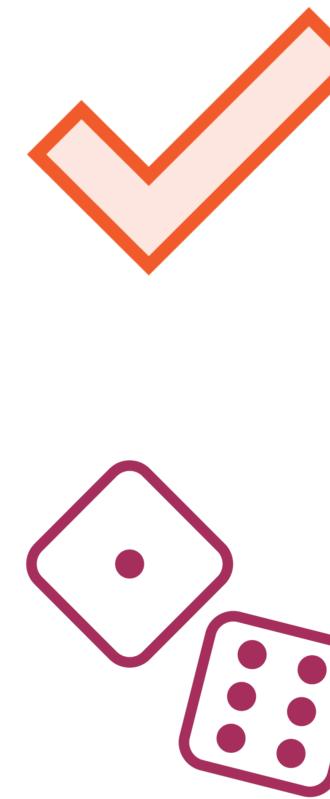
Noise



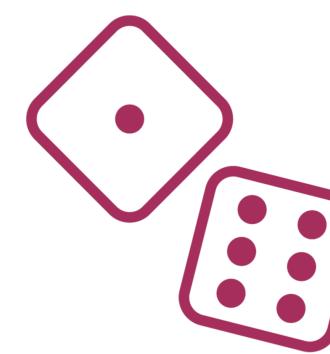
# Trend



**Long-term consistent up or down movement in a series**



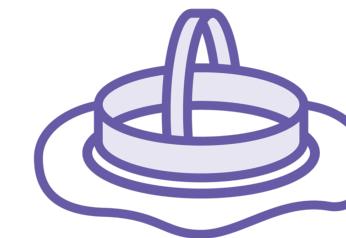
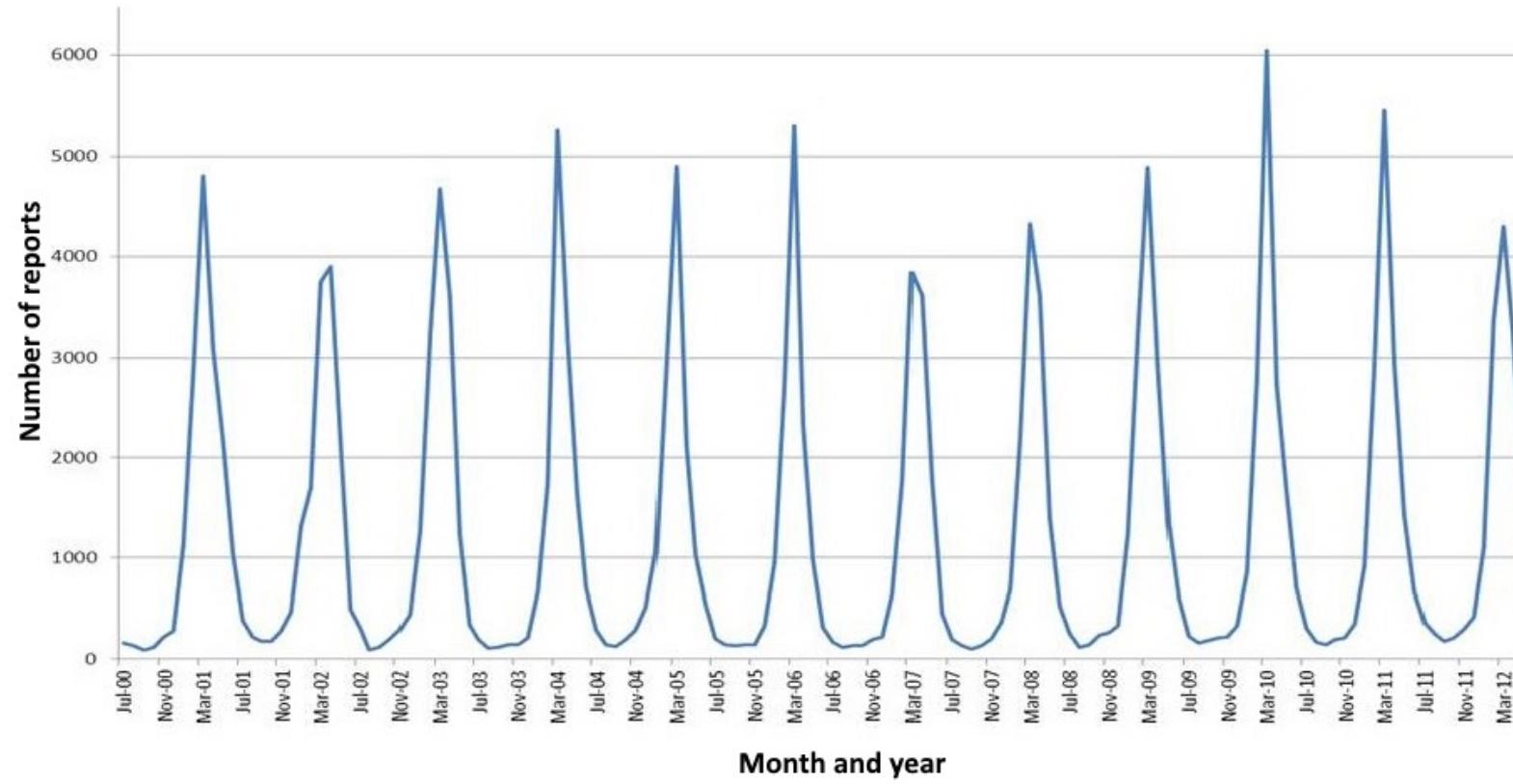
**Deterministic – can find underlying cause**



**Stochastic – random and unexplainable**



# Seasonality



**Seasonality – variations that occur at a predictable and fixed rate**



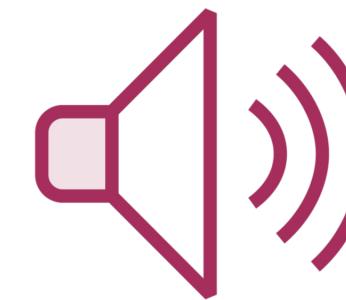
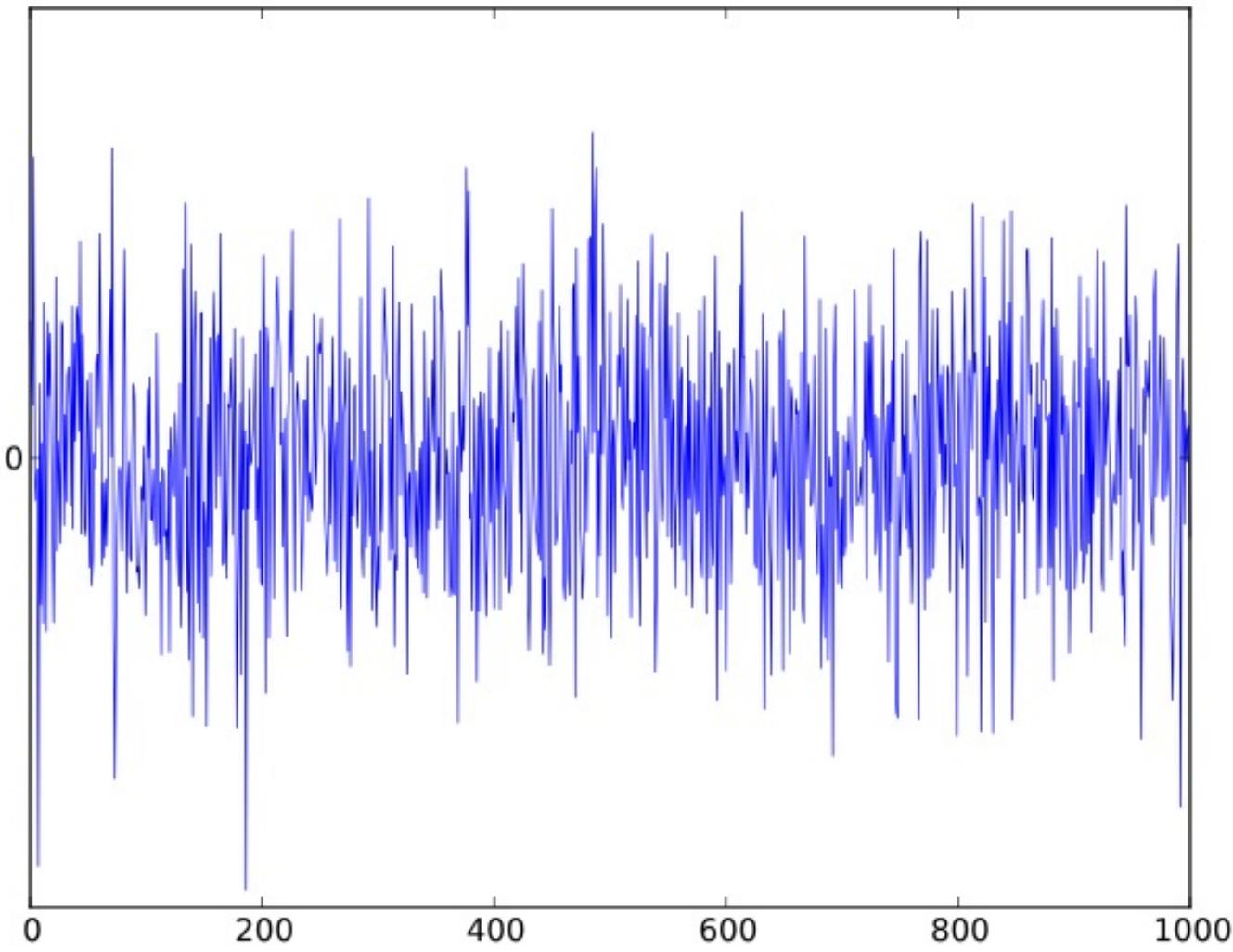
**Rise in sales of ice cream in the summer**



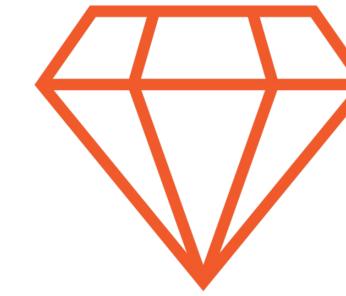
**Increase in revenue for e-commerce websites during holiday season**



# Noise



**Noise is short-term fluctuation which is not predictable**



**Harder to forecast**



**Daily stock prices is an example**



# Time Series Decomposition

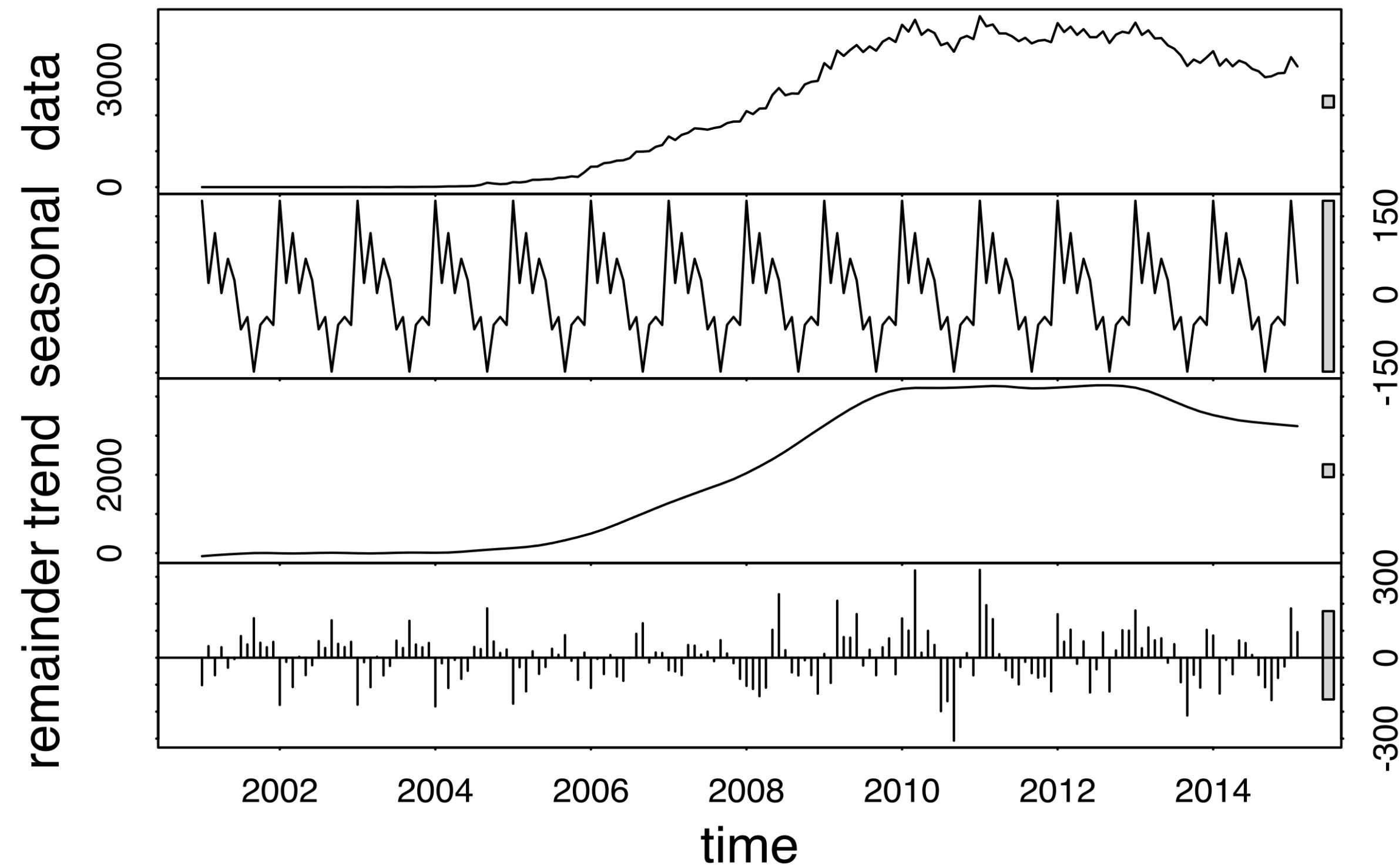


Image citation: EpochFail, CC BY-SA 4.0 <<https://creativecommons.org/licenses/by-sa/4.0>>, via Wikimedia Commons



# Analysing Time Series Data

---



# Time Series Analysis

**Time-series analysis is a technique for analysing time series data and extracting meaningful statistical information and characteristics from it.**



# Examples of Time Series Analysis



Retail sales



Stock market analysis



Weather analysis



Healthcare



Anomaly detection



Astronomy



# What Kind of Data Do We Need for Time Series Analysis?

Date	Run Distance (KM)
12/02/2022	5.31
13/02/2022	3.45
14/02/2022	5.51
15/02/2022	4.98
16/02/2022	5.01
17/02/2022	2.79
18/02/2022	6.01
19/02/2022	5.89
20/02/2022	4.88
21/02/2022	5.23

Time Series Data

Date	User	Run Distance (KM)
12/02/2022	Jay	5.99
12/02/2022	Ariana	3.45
12/02/2022	Kapil	6.02
12/02/2022	Chris	7.45
12/02/2022	Dwayne	4.51
12/02/2022	Pratheerth	7.55
12/02/2022	Aaron	3.31
12/02/2022	Anthony	4.01
12/02/2022	Bo	6.99
12/02/2022	Adam	7.01

Cross-sectional Data



# Stationarity and Autocorrelation

---



# What Is Stationarity?

**A time series is stationary if its statistical properties do not change over time**

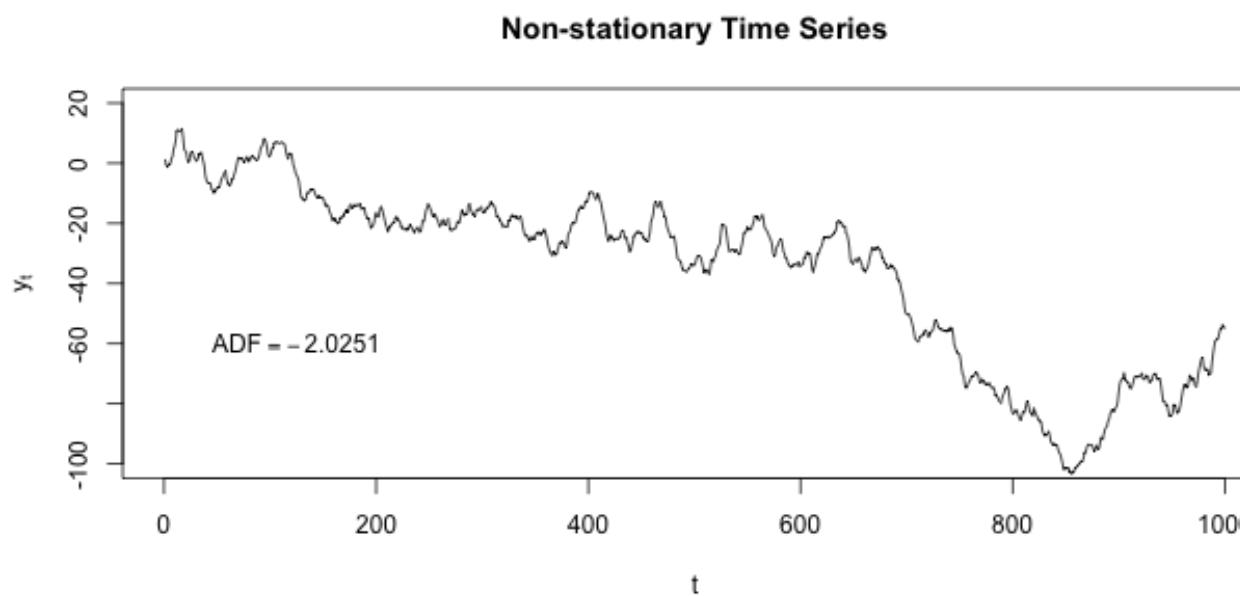
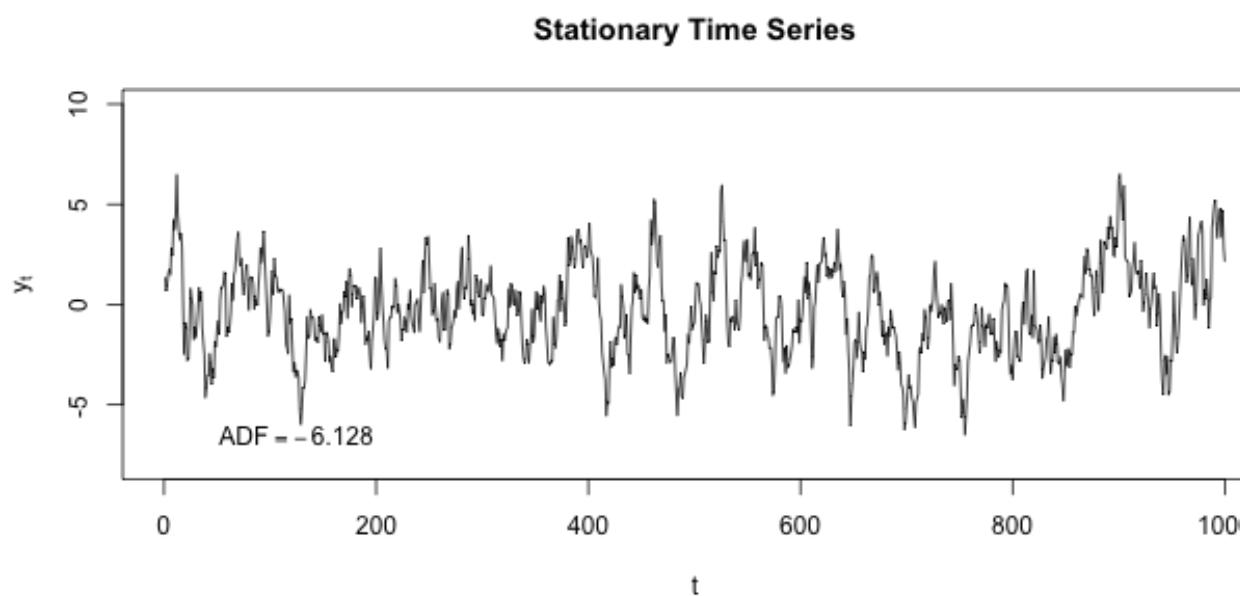
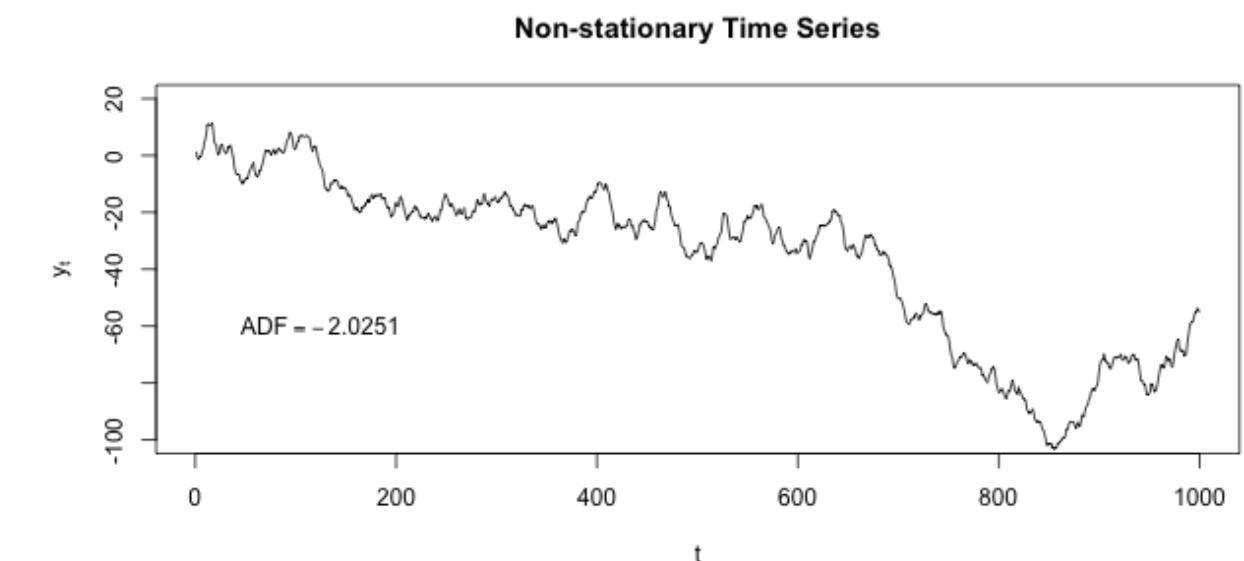
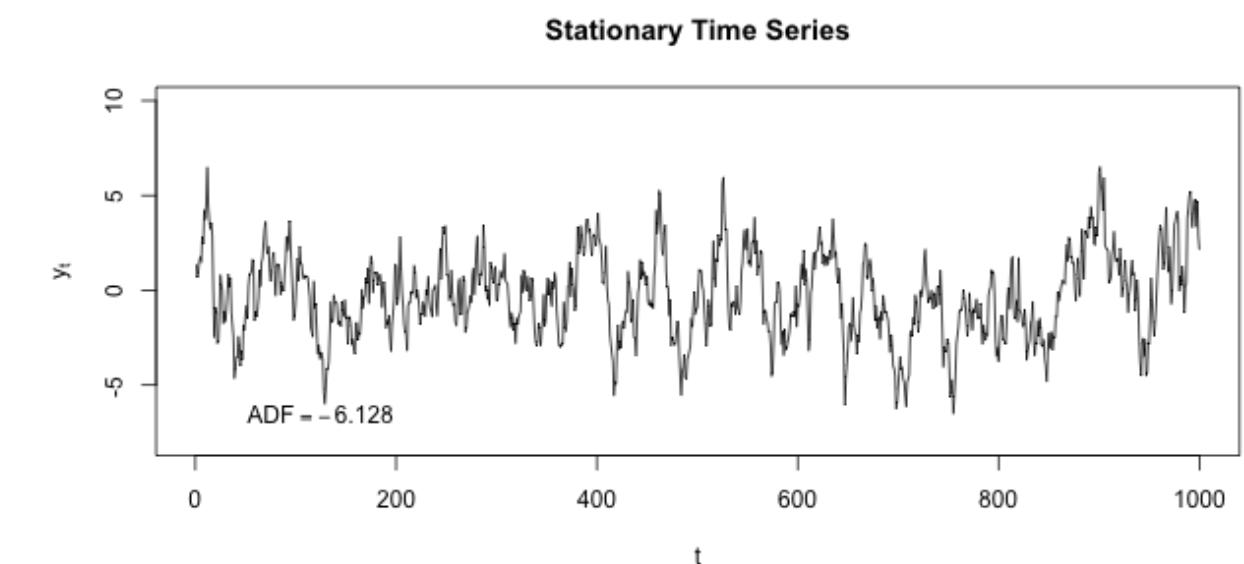


Image citation: Protonk at English Wikipedia, CC BY-SA 3.0 <<https://creativecommons.org/licenses/by-sa/3.0>>, via Wikimedia Commons



# Stationarity

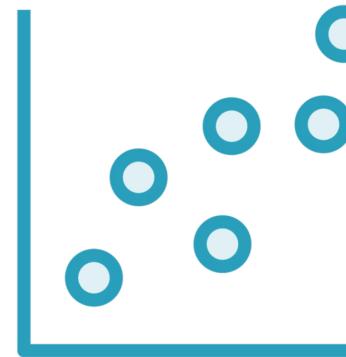
**Constant mean**  
**Constant variance**  
**Constant autocorrelation structure**  
**No periodic component**



Non-stationary data is  
unpredictable and cannot be  
modeled or forecasted



# How to Test for Stationarity?



**Plotting data and visually checking for trend and seasonal components**



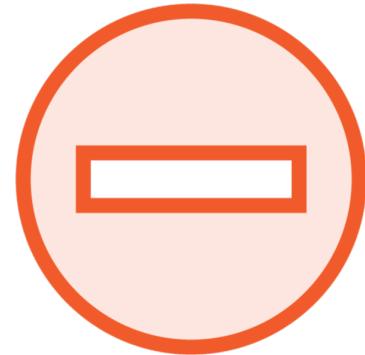
**Statistical tests**

$$\{\gamma, \alpha\}$$

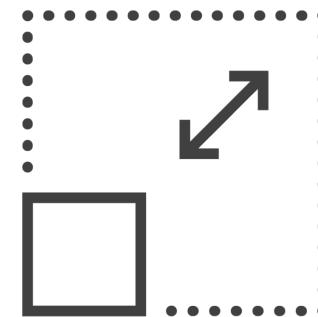
**Augmented Dickey-Fuller (ADF) Test and Kwiatkowski-Phillips-Schmidt-Shin(KPSS) Test**



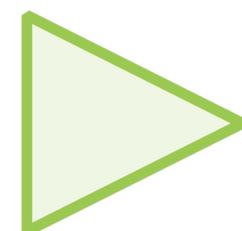
# Making Time Series Stationary



**Stationarity through differencing**



**Transformations – taking log or square root of time series**



**General recommendation is to only use transformations when differencing isn't working**



# Autocorrelation

Numerical values that indicate how a data series is related to itself over time

The time interval between correlated values is called lag

Ranges from -1 to +1



# Introduction to Anomaly Detection

---

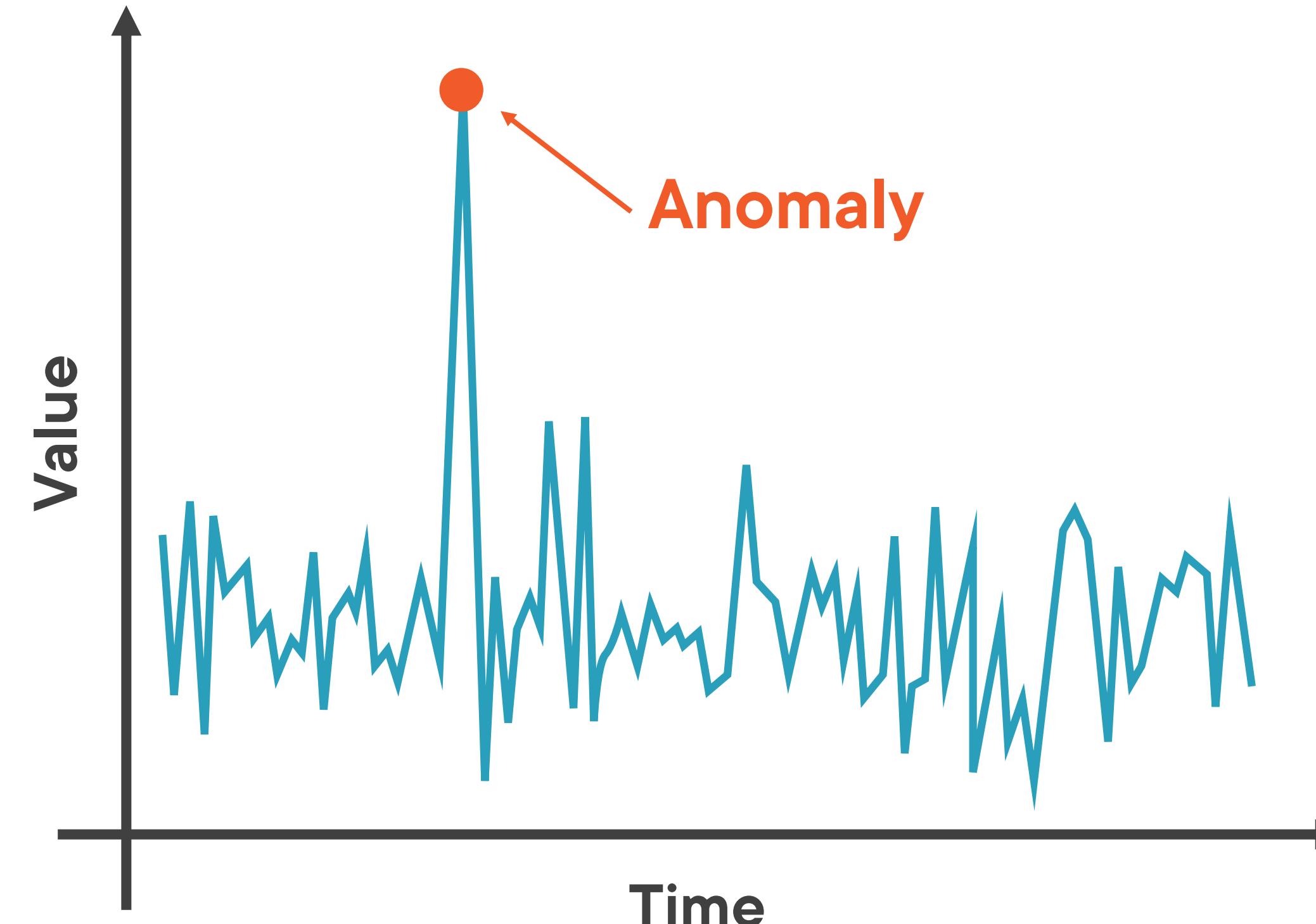


# Anomaly

**An observation, data point, or collection of data points that doesn't seem to follow the same pattern as the rest of the dataset**



# Time Series Anomaly



# Why Detect Anomalies?



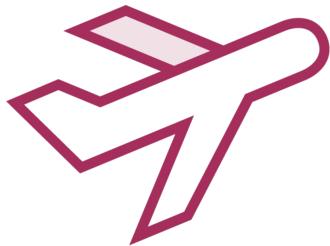
**Anomaly detection is very useful in a wide variety of fields**



**Credit card fraud detection**



**Stock market manipulation**



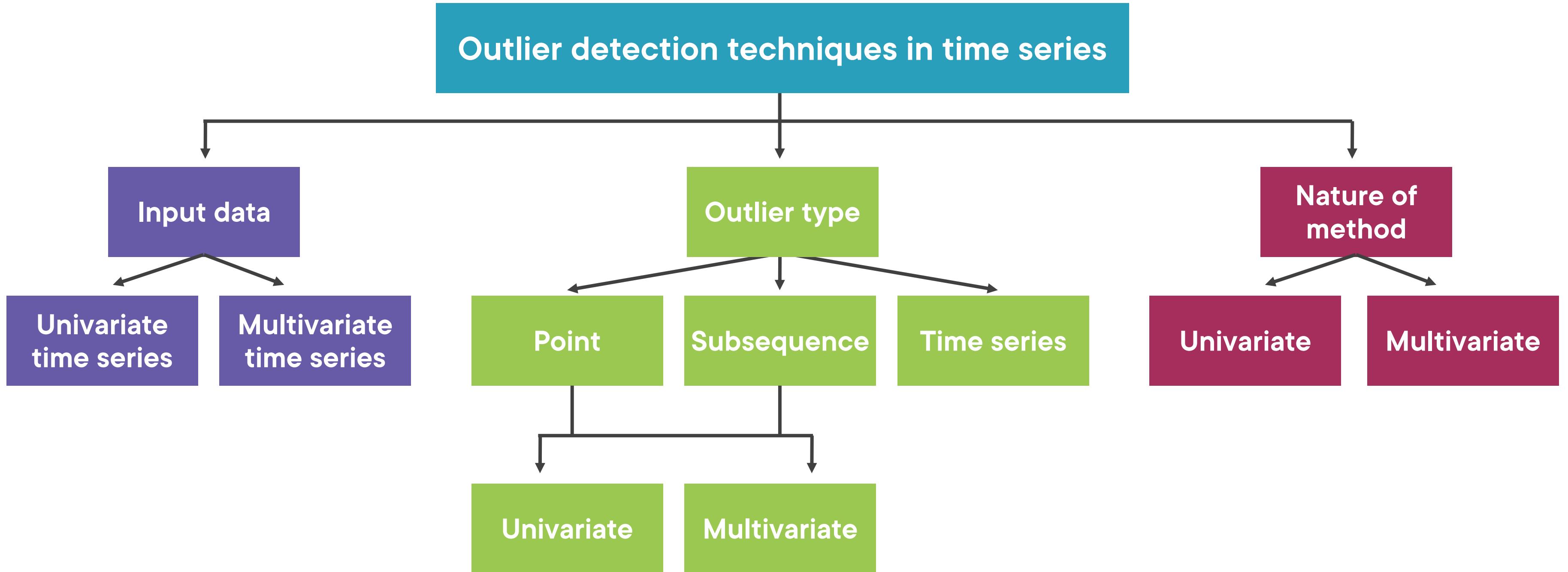
**In-flight safety**



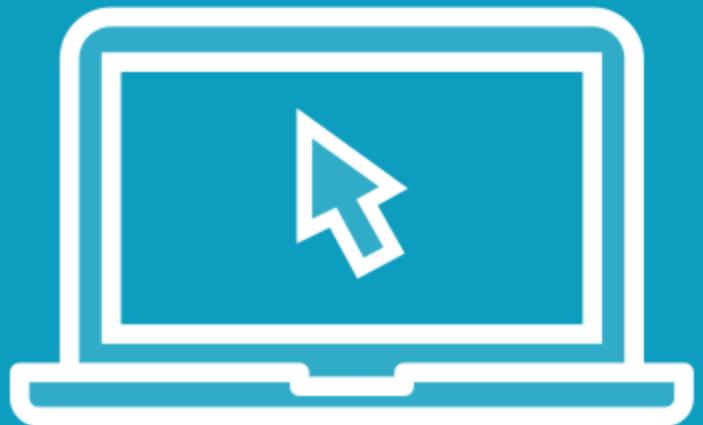
**Event detection systems in sensor data**



# Outlier Detection Techniques



Demo



**Setting up your environment**



## Summary



**Time series - sequence of data points obtained through repeated measurements**

**3 main components: Trend, Seasonality, Noise**

**Time series is stationary if its statistical properties do not change over time**



# Summary



**Statistical tests and transformations are used to test for and create stationary time series**

**Autocorrelation is a numerical value that indicates how a data series is related to itself over time**

**Anomaly is an observation that doesn't seem to follow the same pattern**



# Up Next: Building a Model to Automate Anomaly Detection

---

