University of Bremen

DataNord

DATA SCIENCE CENTER

Time Series Analysis and Prediction

Using Python and Jupyter

CONTACT

DSC Data Scientists

movahedm@uni-bremen.de

anolte@uni-bremen.de

NEWSLETTER

Sign Up!

The Trainers

Data Scientists for Consultation & Training

M.Sc. ANNIKA NOLTE

- 2019 – 2023 Scientific Researcher & PhD Candidate, Climate Service Center Germany (GERICS) & Universität Hamburg
  - Hydrogeology; Geochemistry; Climatic and non-climatic drivers
  - Database development; Time series & Machine Learning using Python
- 2019 Environmental Sciences (M.Sc.) Technical University Braunschweig

Dr. MARYAM MOVAHEDIVAR

- 2021 – 2024 Postdoctoral scholar, Institute for Statistics, University of Bremen, Germany
- 2016 – 2020 PhD in Statistics, Dep.of Statistics, University of Payame noor, Tehran, Iran
- 2007 – 2009 Master's degree in Mathematical Statistics Dep. of Statistics, University of Payame noor, Shiraz, Iran
- 2001 – 2005 Bachelor's degree in Mathematical Statistics Dep. of Statistics, University of Payame noor, Neyshabur, Iran

Training Goals

1. Know fundamental concepts in time series analysis and develop an understanding of different characteristics of time series data

2. Understand how to use Python to work with time series data - focus: <u>Earth Sciences</u>

3. Acquiring skills to analyze, visualize, and forecast time series using key Python packages

4. Applying shareable and reproducible data science techniques with Jupyter notebooks

Source: Sumanley, Pixabay

Training Outline

University of Bremen

Day 1 **(~ 9 am - 5 pm)**	**Time Series Concepts & Exploratory Time Series Data Analysis**
Day 2 **(~ 9 am - 5 pm)**	**Stationary Time Series Models and Forecasting**
Day 3 **(~ 9 am - 12 pm)**	***[continues]* AND/OR “Bring your own data”**

Visual Examples of Time Series

[Climate Time Machine (nasa.gov)](https://climate.nasa.gov/interactives/climate-time-machine)

What is a Time Series?

- A time series is a sequential set of data points, measured typically over successive times.
- Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

Categories and Terminologies

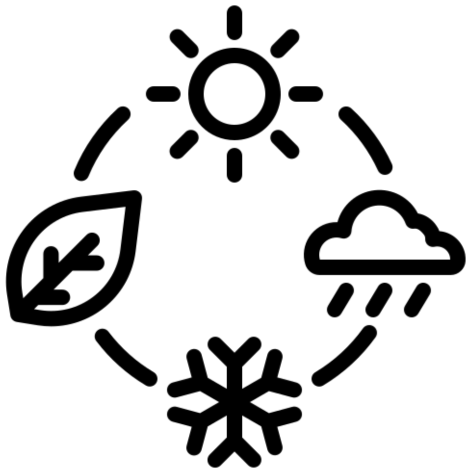
- **Time-domain** vs. **Frequency-domain**
  - Time-domain approach: how does what happened today affect what will happen tomorrow?
    These approaches view the investigation of lagged relationships as most important, e.g. autocorrelation analysis.
  - Frequency-domain approach: what is the economic cycle through periods of expansion and recession?
    These approaches view the investigation of cycles as most important, e.g. spectral analysis and wavelet analysis.
- This lecture will focus on time-domain approaches.

Categories and Terminologies

- **univariate** vs. **multivariate**
  A time series containing records of a single variable is termed as univariate, but if records of more than one variable are considered then it is termed as multivariate.
- **linear** vs. **non-linear**
  A time series model is said to be linear or non-linear depending on whether the current value of the series is a linear or non-linear function of past observations.
- **discrete** vs. **continuous**
  In a continuous time series observations are measured at every instance of time, whereas a discrete time series contains observations measured at discrete points in time.

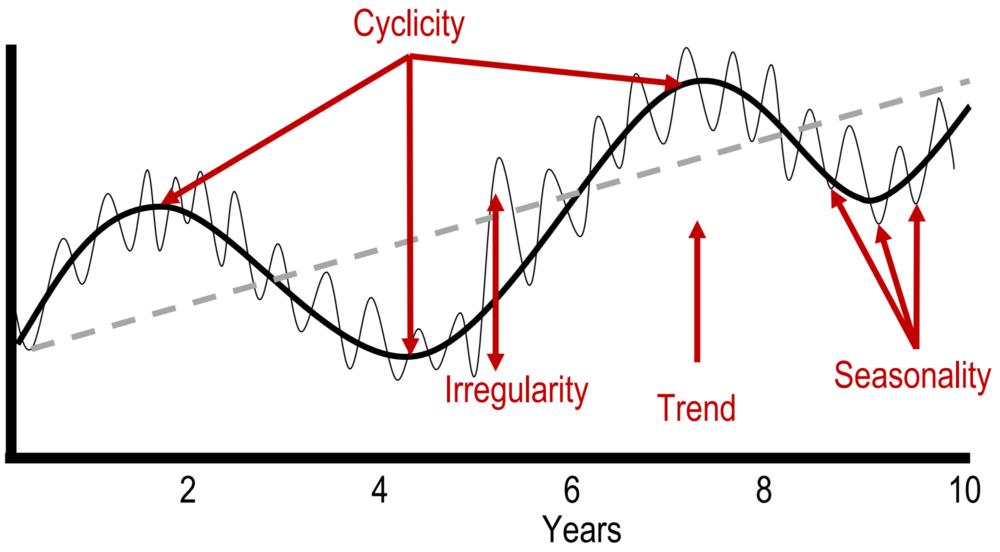
Components of a Time Series

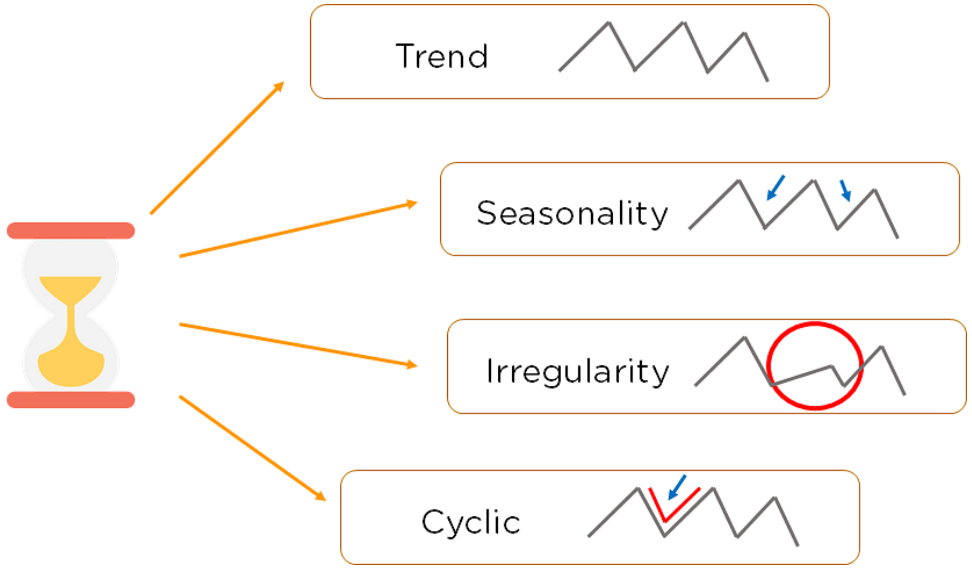
- Seasonality
- Trend
- Cyclinicality
- Irregularities



Source: Mutualism, Flaticon

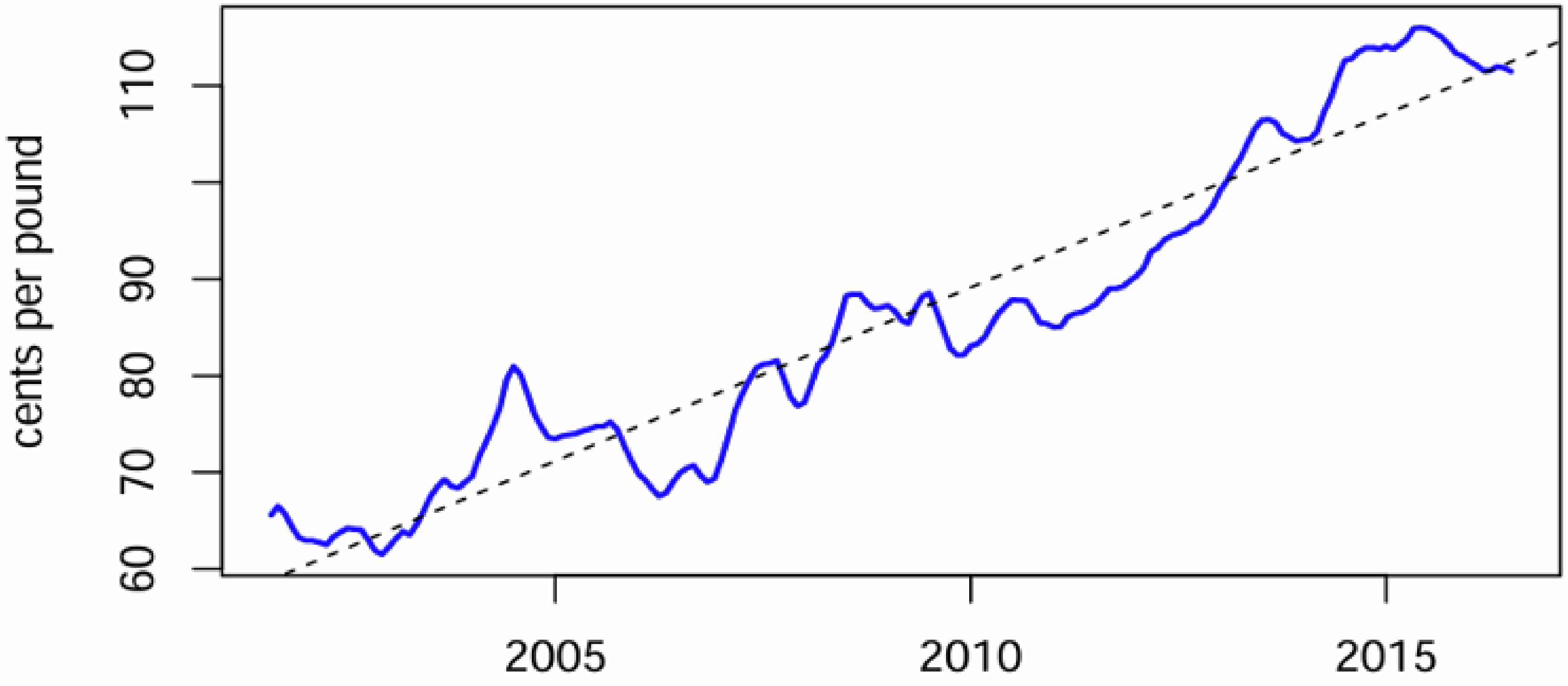
Source: Fajar Studio, Noun Project





Components of a Time Series

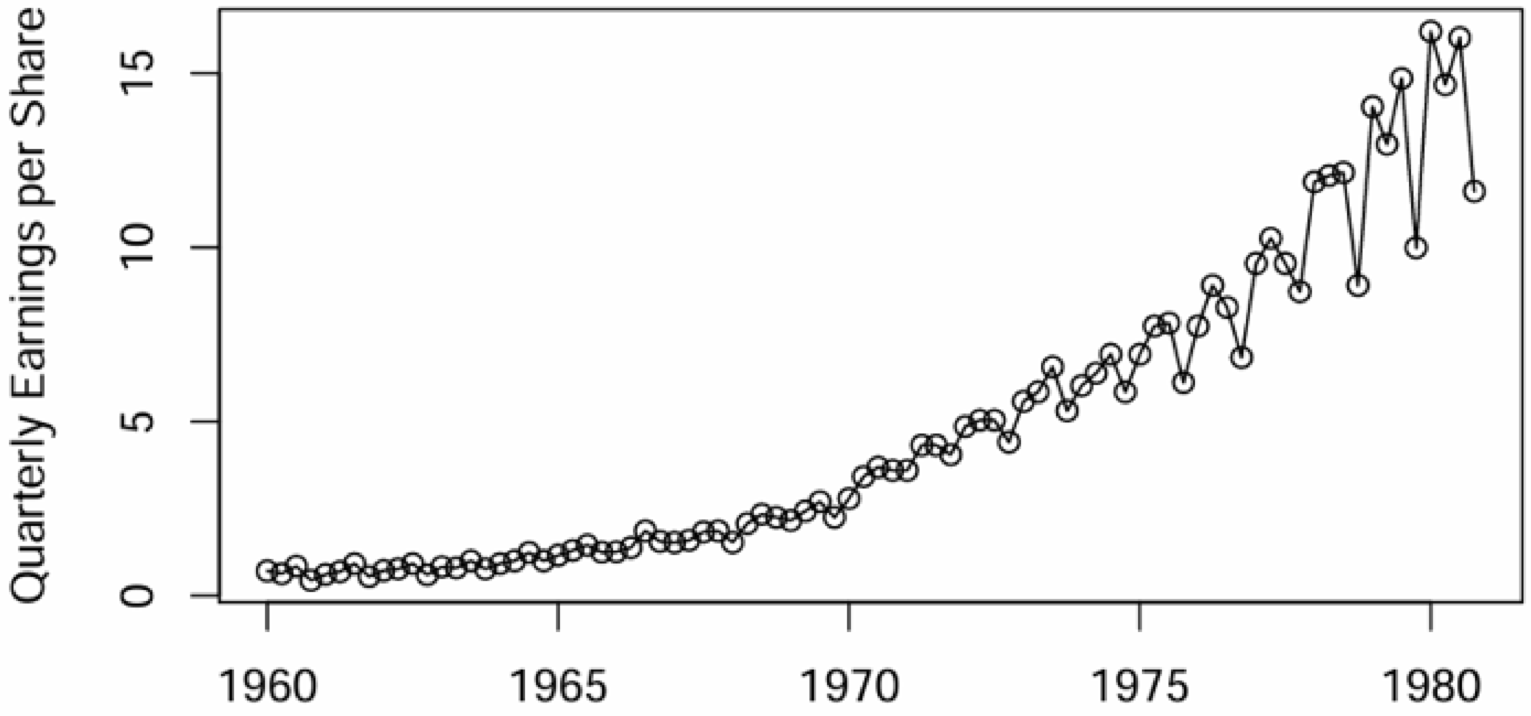
- **Trend**
  The general tendency of a time series to increase, decrease or stagnate over a long period of time.



The price of chicken: monthly whole bird spot price, Georgia docks, US cents per pound, August 2001 to July 2016, with fitted linear trend line.

Components of a Time Series

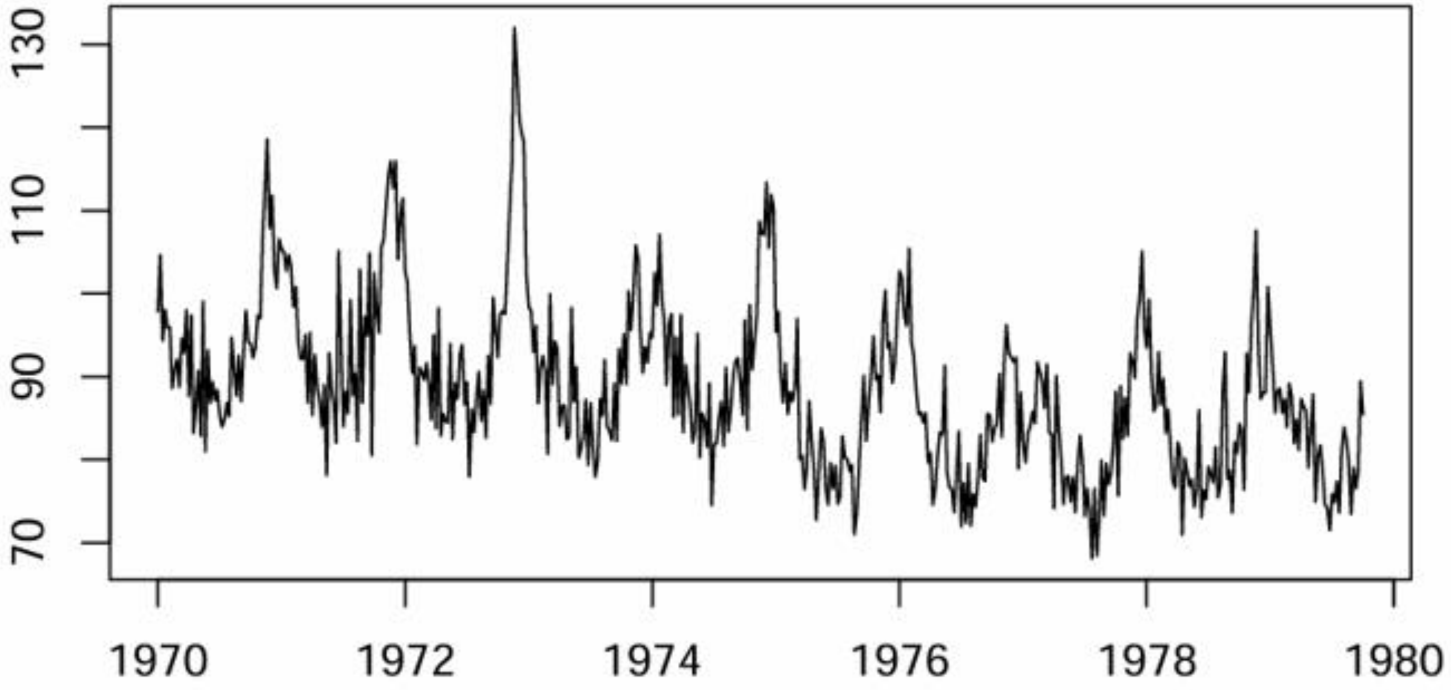
- **Seasonal variation**
  This component explains fluctuations within a year during the season, usually caused by climate and weather conditions, customs, traditional habits, etc.



Johnson & Johnson quarterly earnings per share, 84 quarters, 1960-I to 1980-IV.

Components of a Time Series

- **Cyclical variation**
  This component describes the medium-term changes caused by circumstances, which repeat in cycles. The duration of a cycle extends over longer period of time.



Average weekly cardiovascular mortality in Los Angeles County. There are 508 six-day smoothed averages obtained by filtering daily values over the 10 year period 1970-1979.

Components of a Time Series

- **Irregular variation**
  Irregular or random variations in a time series are caused by unpredictable influences, which are not regular and also do not repeat in a particular pattern.
  These variations are caused by incidences such as war, strike, earthquake, flood, revolution, etc.
  There is no defined statistical technique for measuring random fluctuations in a time series.

Combination of Four Components

- Considering the effects of these four components, two different types of models are generally used for a time series.

  - Additive Model

    $$Y(t) = T(t) + S(t) + C(t) + I(t)$$

    Assumption: These four components are independent of each other.

  - Multiplicative Model

    $$Y(t) = T(t) \times S(t) \times C(t) \times I(t)$$

    Assumption: These four components of a time series are not necessarily independent and they can affect one another.

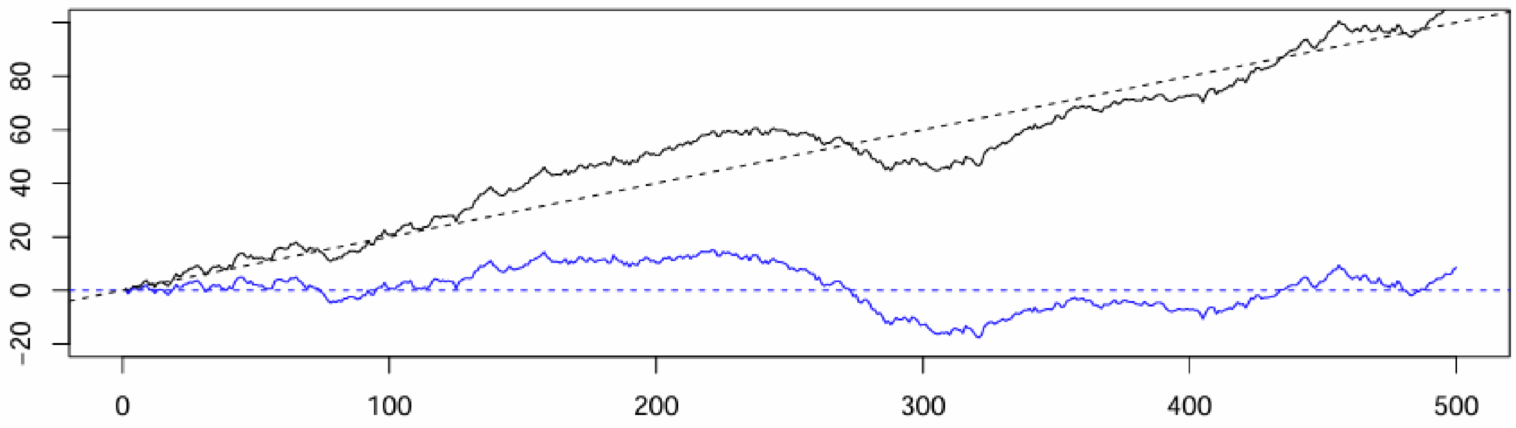
Time Series Examples: White Noise

- **White Noise**
  - A simple time series could be a collection of uncorrelated random variables, $\{w_t\}$, with zero mean $\mu = 0$ and finite variance σ_w^2, denoted as $w_t \sim wn(0, \sigma_w^2)$.
- **Gaussian White Noise**
  - A particular useful white noise is Gaussian white noise, wherein the w_t are independent normal random variables (with mean 0 and variance σ_w^2), denoted as $w_t \sim \text{iid } \mathcal{N}(0, \sigma_w^2)$.
- White noise time series is of great interest because if the stochastic behavior of all time series could be explained in terms of the white noise model, then classical statistical methods would suffice.

Time Series Examples: Random Walks

- A random walk is the process by which randomly-moving objects wander away from where they started.
- Consider a simple 1-D process:
  - The value of the time series at time t is the value of the series at time $t-1$ plus a completely random movement determined by w_t. More generally, a constant drift factor δ is introduced.

$$X_t = \delta + X_{t-1} + w_t = \delta t + \sum_{i=1}^{t} w_i$$

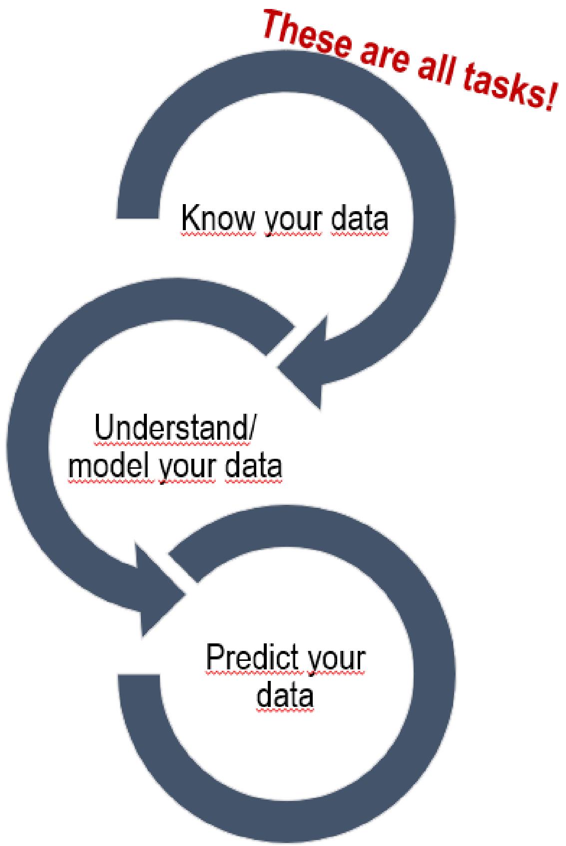


Source: jobhaug, Pixabay

Time Series Analysis and Forecasting

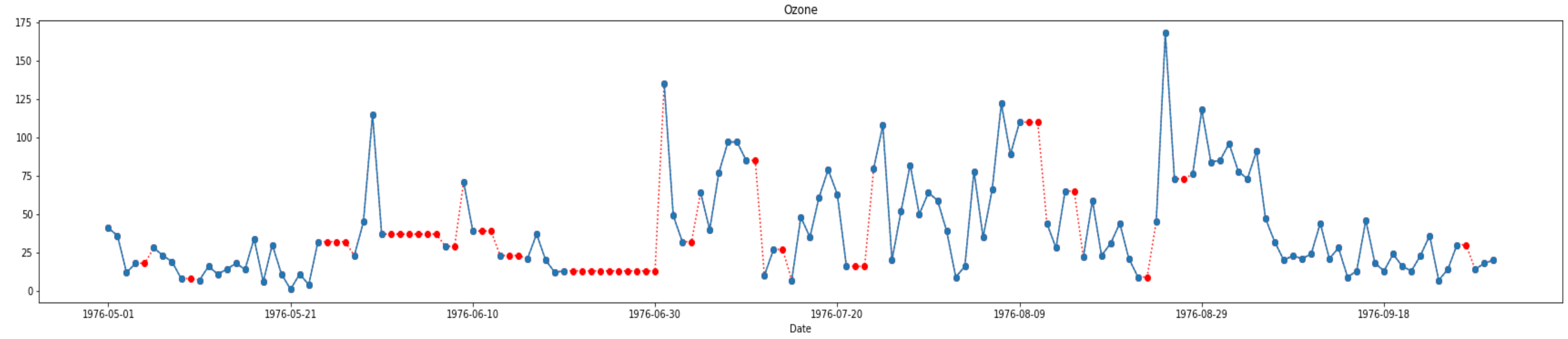
University of Bremen

- The procedure of using known data values to fit a time series with suitable model and estimating the corresponding parameters. It comprises methods that attempt to understand the nature of the time series and is often useful for future forecasting and simulation.
- There are several ways to build time series forecasting models, but this lecture will focus on **stochastic process**.
  - We assume a time series can be defined as a collection of random variables indexed according to the order they are obtained in time, $X_1, X_2, X_3, \ldots$ t will typically be discrete and vary over the integers $t = 0, \pm 1, \pm 2, \ldots$
  - Note that the collection of random variables $\{X_t\}$ is referred to as a stochastic process, while the observed values are referred to as a realization of the stochastic process.

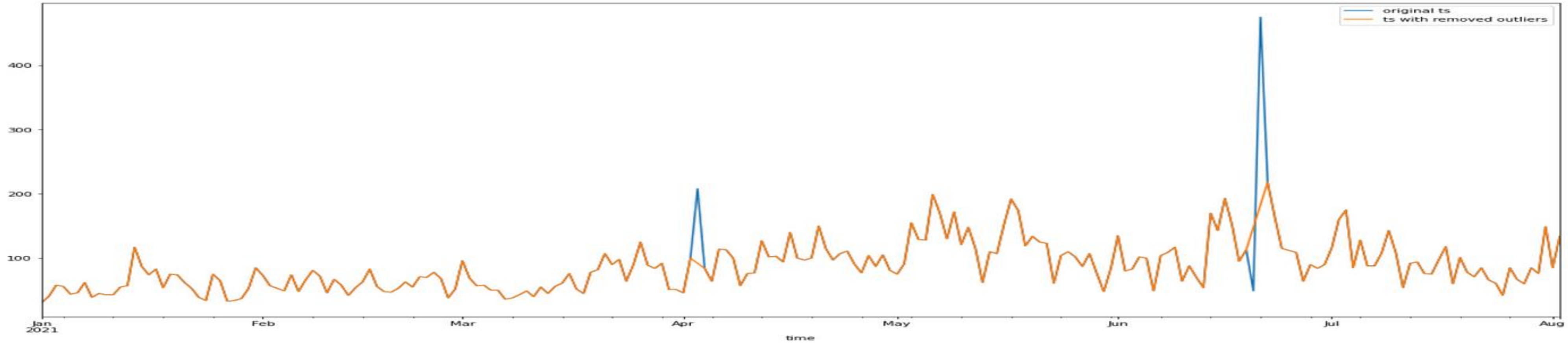


Missing Values and Outliers

- Missing Values



- Outliers



Time Series Analysis in Python