Introduction to Time Series

👋🏽 *Goodbye* (for now): infrastructure, AWS, databases, docker
Hello (again): machine learning, jupyter, plots!

Time Series Analysis

Q: What examples of time series data can you think of?

A:

- sales data
- weather data
- climate change data
- COVID data
- stock market
- usage data (e.g. website, app)
- sensor / machinery components data (for predictive maintenance)
- inflation rates, economic growth
- population dynamics

Q: What is special about time series data?

A:

- Ordered (arrow of time, non-repeatable, unique)
- Periodic/seasonal
- Trend
- Not independent

- We predict outside of the domain of our training data (i.e. future)

Two step approach:

- *Decomposition (Tuesday am)*
  - Time series elements:
    - Trend
    - Seasonality
    - **Remainder**
      - This is the part that we are interested in this week; this is the part that we are trying to model
      - The models we use have an assumption of stationarity — statistical characteristics like mean, standard deviation, correlation don't change over time — which is why we first decompose time series into its elements
  - How do we decompose time series:
    - Additive:
      - `y = trend + seasonal + remainder`
      - when the seasonal variation remains ~constant over time
    - Multiplicative:
      - `y = trend * seasonal * remainder`
      - when the seasonal variation changes over time
- *Autoregression* (AR; later also ARIMA) (*Tuesday & Thursday*)
  - Look at the remainder and model (e.g. linear regression) correlation between the value in point t and the value in points before t

Your Project for the Week

Goal:

> Build a model that is able to make short term temperature forecasts 🧙

Data:

- For the project:
  - *temperature data,* daily mean temperatures for the weather station of your choice!
  - follow instructions in the course materials
- For the encounters:
  - *flights dataset*, # of monthly airline passengers between 1949 and 1960.
  - in `week_07/data`

Your Tasks

- Download the data
- EDA
  - Clean the data
  - Plot the data
  - Decide on the final dataset/timeframe to use