Time Series

Status	Done
Files	
Link	
Type	All Around
Course	Data Analysis

❗ This makes quite a small portion of the exam, if any, maybe one or two multiple choice questions, so not really worth revising, but its very interesting for real world analysis of time series

- Temporal = related to time

Weak stationarity of a series

- **Weak stationarity of a (time) series** is the problem that arrises when we try to perform summary/inferential/descriptive statistics of a (time) series
    - We usually assume that there is no correlation between the variables, but if there is a correlation then the usual approaches are not correct
    - We need to remove this correlation before continuing so the time series “doesn’t change its probabilistic character” over time

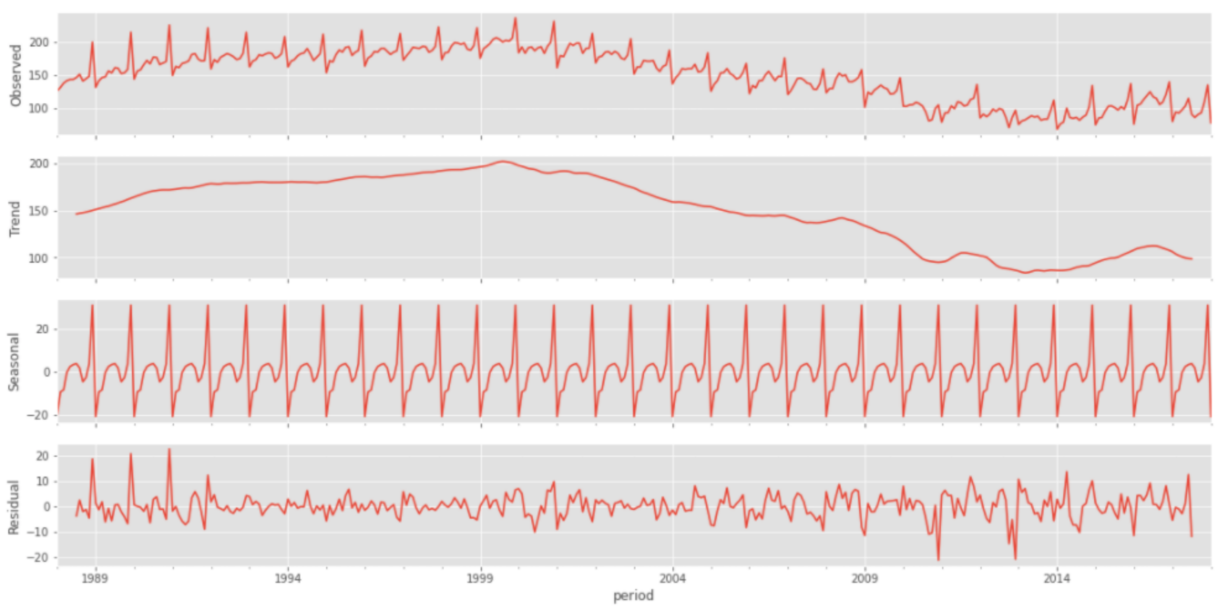
- Stationary → does not have these seasonal effects
  Non Stationary → has these seasonal effects
- Basically its like you wanna study the time series but you wanna look at overall trends and generalise the daily/seasonal/whatever fluctuations that repeat every x period because these are distractions and they are not important in the overall trend
- **Formally,** a time series Xt is said to be weakly stationary, if :

  - $E[X_t] = \mu$: all X have identical expected mean
  - $Var(X_t) = \sigma^2$: all X_t have identical variance
  - $Cov(X_t, X_{t+h}) = \gamma_h$: autocovariance depends only on lag h

- How can we deal with this? We can decompose a time series as follows

$$x_t = m_t + s_t + r_t$$
$$= \text{trend} + \text{seasonal effect} + \text{stationary remainder}$$

- Example, we can see some repeating patterns in this graph with the raw series



Stationarity statistical tests

- Augmented Dickey-Fuller (ADF) unit root test
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for stationarity

Correlogram

- Correlogram is a visual way to show serial correlation in data that changes over time
- The intuition behind it is simple, if its a purely random series, x_n can not help in predicting x_{n+1}
- We can put this in a value called **Sample Autocorrelation Coefficient**
  - The closer to 0 is it, the more random the source of the data is, especially if n is large

$$r_1 = \frac{\sum_{t=1}^{n-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$