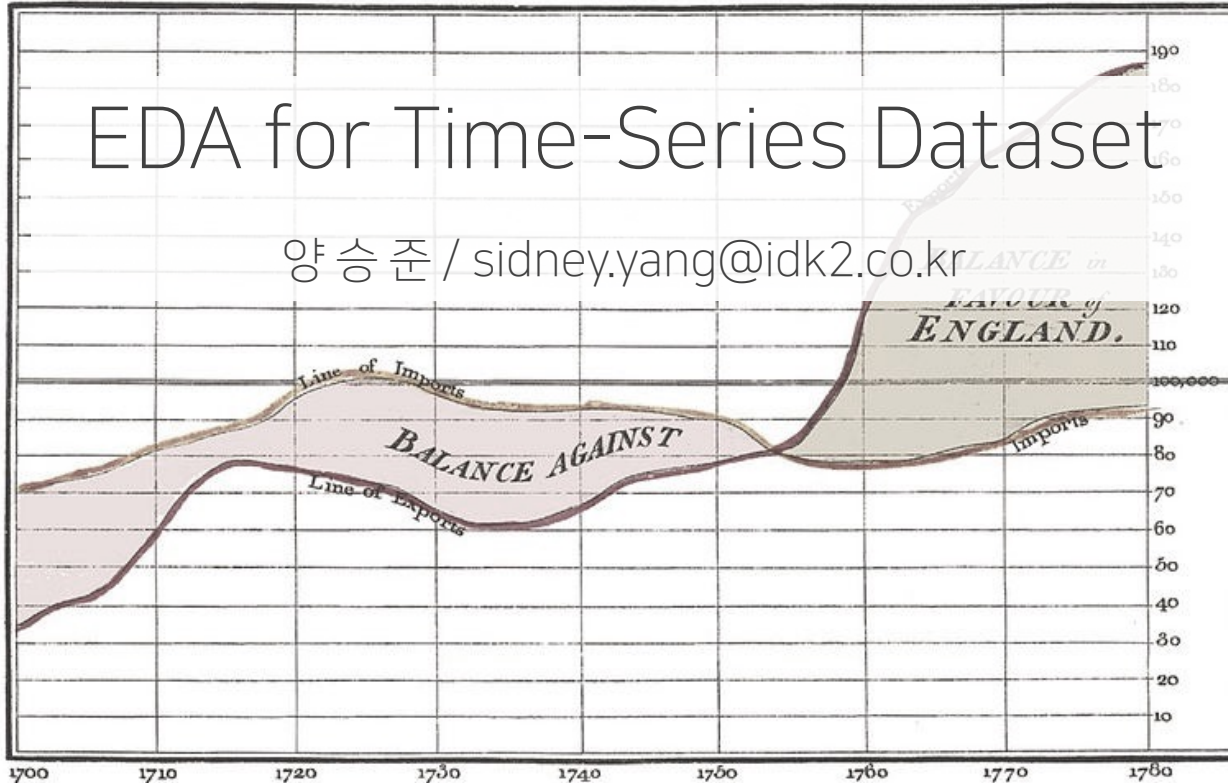


Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.

EDA for Time-Series Dataset

양승준 / sidney.yang@idk2.co.kr



The Bottom line is divided into Years, the Right hand line into L10,000 each.

Published as the Act directs, 1st May 1786, by W^m Playfair

Noble sculpt 352 Strand, London.

Definition: Time-Series Analysis

O'REILLY

Practical Time Series Analysis

Prediction with Statistics & Machine Learning



Aileen Nielsen

시계열 분석은 일정 간격의 시간순으로 저장된 데이터에서 유의미한 정보를 뽑아 내는 작업으로 과거의 행동을 이해하고 미래의 행동을 예측하는데 사용됩니다.

- Exploratory(Descriptive) Analysis: 주로 시각화 작업을 통해 trends, cycles, seasonality 등의 패턴을 이해하는 작업
- Forecasting: 과거 관측값으로 모델을 만들어 미래에 발생할 값을 예측하는 작업

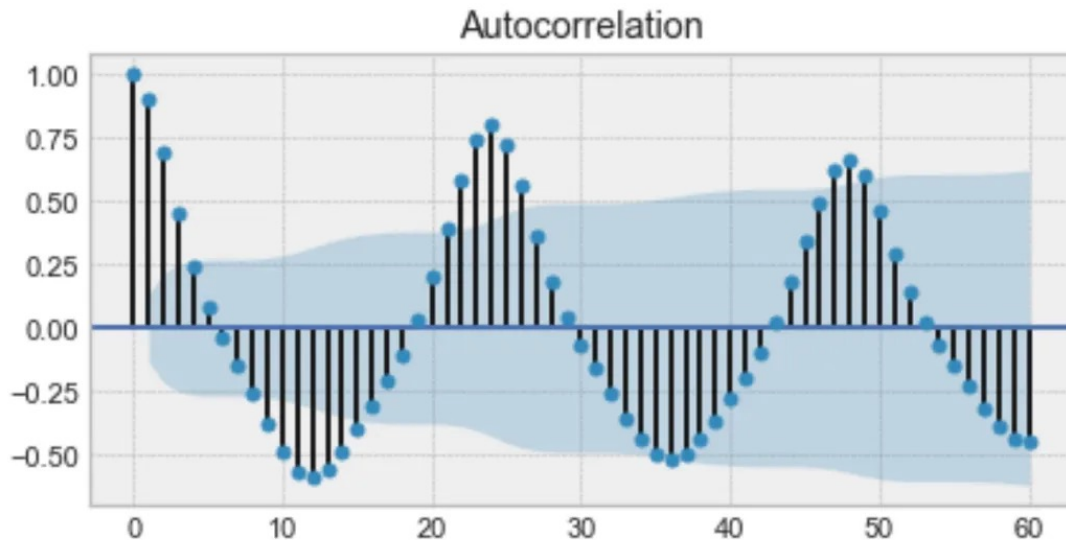
참고

- 샘플 데이터셋: UCI Machine Learning Repository (<https://archive-beta.ics.uci.edu/>)
- 책: Forecasting (<https://otexts.com/fpp2/index.html>)

Time-Series Analysis – Key Concept

Temporal Behavior of Data (시간의 흐름에 따른 데이터의 특성):

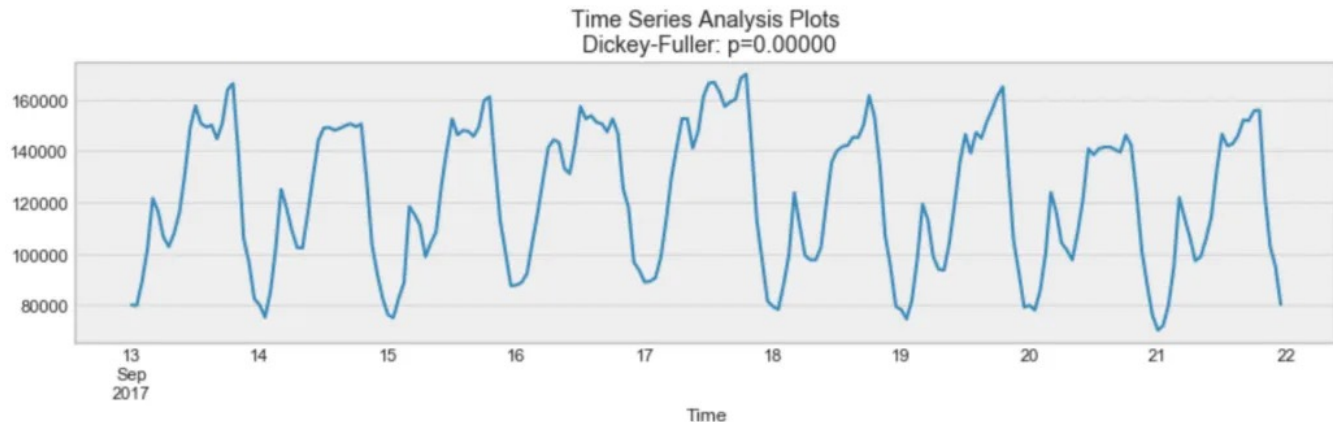
- Autocorrelation
 - the similarity between observations as a function of the time lag between them.
 - “자기”와 “특정 시차에 위치한 과거의 자기”와 유사한 정도
- 아래 그림: 24 time unit마다 유사한 관측값이 목격됨. (hint for seasonality)



Time-Series Analysis – Key Concept

Temporal Behavior of Data (시간의 흐름에 따른 데이터의 특성):

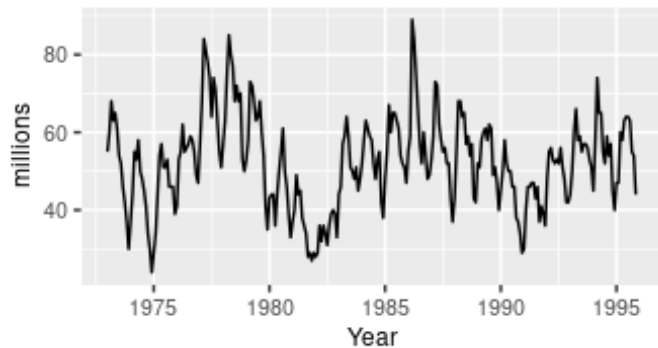
- **Seasonality:** 분기, 달, 요일, 시간대(오전/오후) 등 계절적 요인에 따라 주기적 패턴을 가지고 변동하는 경우
 - 참고) Cyclic Pattern: 주기적이진 않지만(fixed period X) 증가, 하락하는 패턴을 보이는 경우 (경기의 부침)
- **Stationarity:** 평균, 분산과 같은 통계적 특성이 변하지 않는 경우
 - Dickey-Fuller test: 시계열 데이터가 stationary한지 판별하는 통계적 검정방식



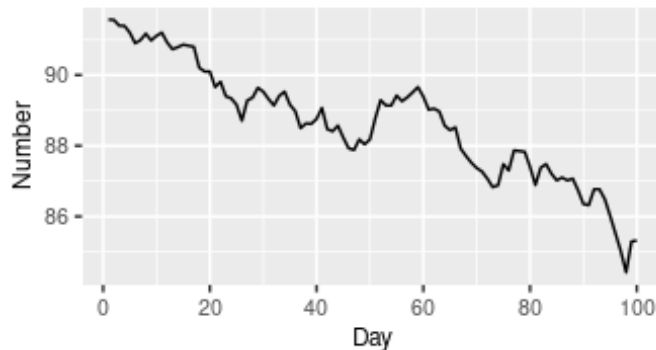
Trends, Seasonal, Cyclic

Trends: X
Seasonal: O
Cyclic: O

① Sales of new one-family houses, USA

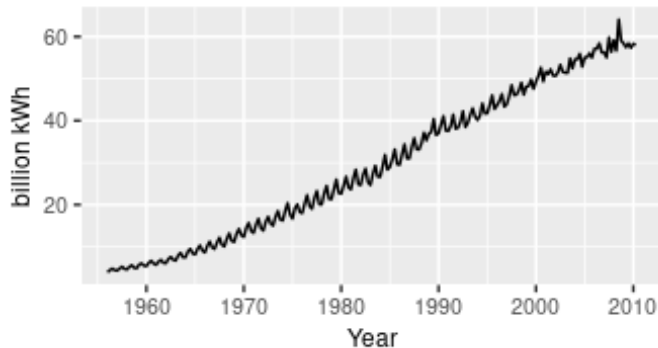


② US treasury bill contracts

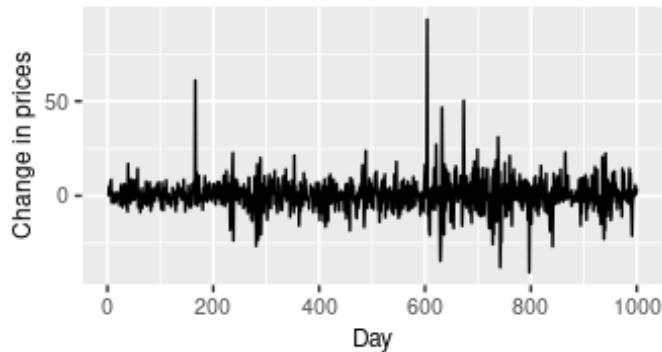


Trends: O
Seasonal: ?
Cyclic: ?

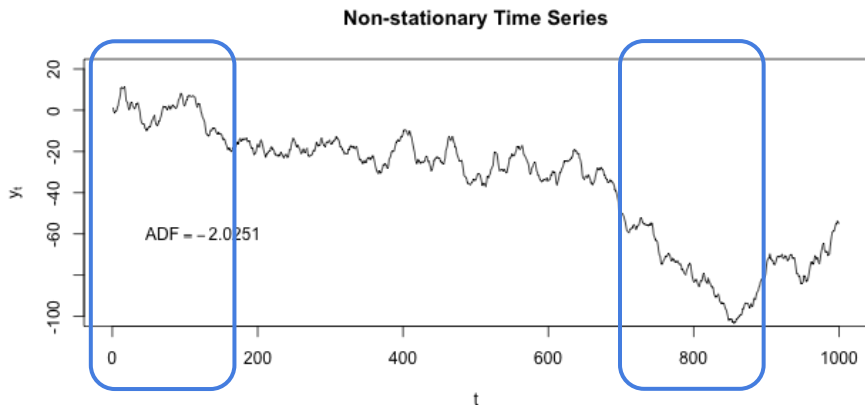
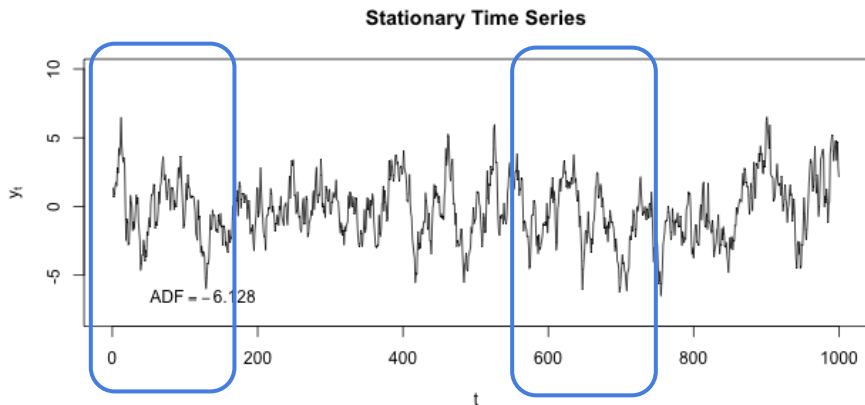
③ Australian quarterly electricity production



④ Google daily changes in closing stock price



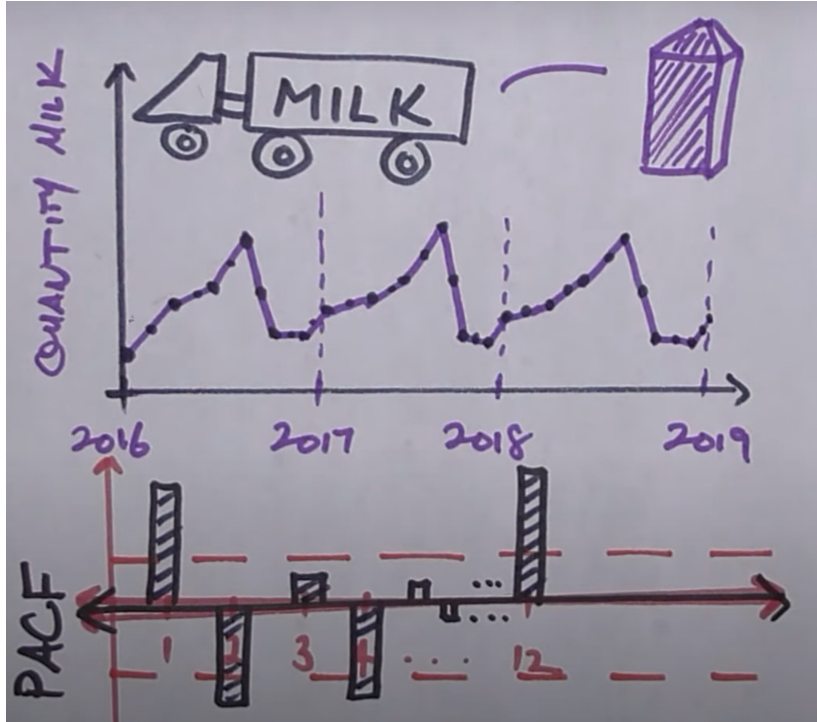
Stationarity (정상성)



Stationarity

- the statistical properties of a process generating a time series do not change over time.
- stationarity is important because, in its absence, a model describing the data will vary in accuracy at different time points.

AR Model (자기회귀 모델)



AR Model

- AR: AutoRegressive
- 자기를 사용해서 자기를 예측

월별 우유 생산량이 궁금함

- M_t : 이번달 우유 생산량
- M_{t-1} : 지난달 우유 생산량 (Lag:1)
- M_{t-2} : 두달 전 우유 생산량 (Lag:2)
- 모든 이전 달 생산량을 사용할 수도 있지만 M_{t-n} 을 최소화하는 게 좋음 → PACF 활용

PACF(편자기상관함수) 차트

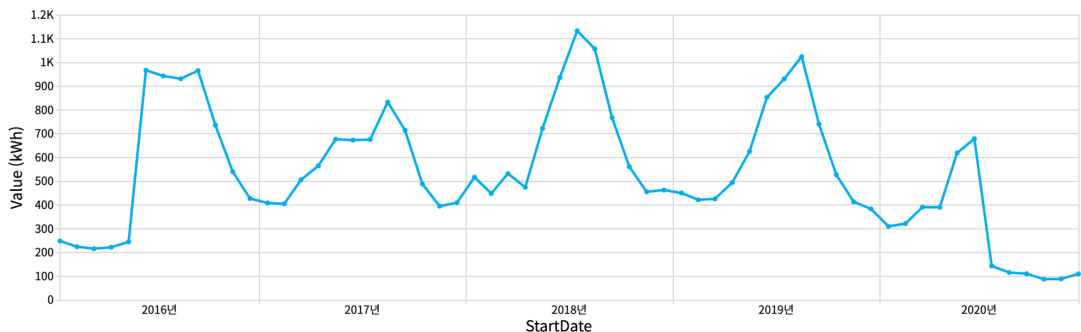
- M_{t-1}, M_{t-12} 가 이번달 생산량과 높은 + 상관관계
- M_{t-2}, M_{t-4} 가 이번달 생산량과 높은 - 상관관계

$$m_t = \beta_0 + \beta_1 m_{t-1} + \beta_2 m_{t-2} + \beta_4 m_{t-4} + \beta_{12} m_{t-12} + \varepsilon_t$$

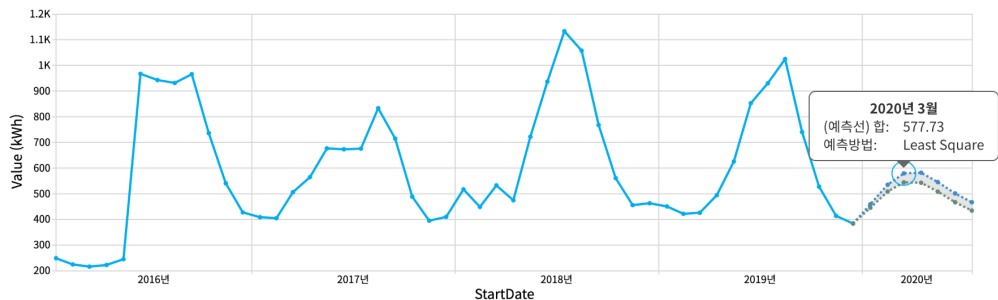
AR Model (자기회귀 모델)

AR model makes an assumption that the observations at previous time steps are useful to predict the value at the next time step.

Y: Value (kWh) (합) X: StartDate 하위그룹: 없음 화면분할: 없음 상관계수: -0.25



Y: Value (kWh) (합) X: StartDate 하위그룹: 없음 화면분할: 없음 상관계수: 0.18

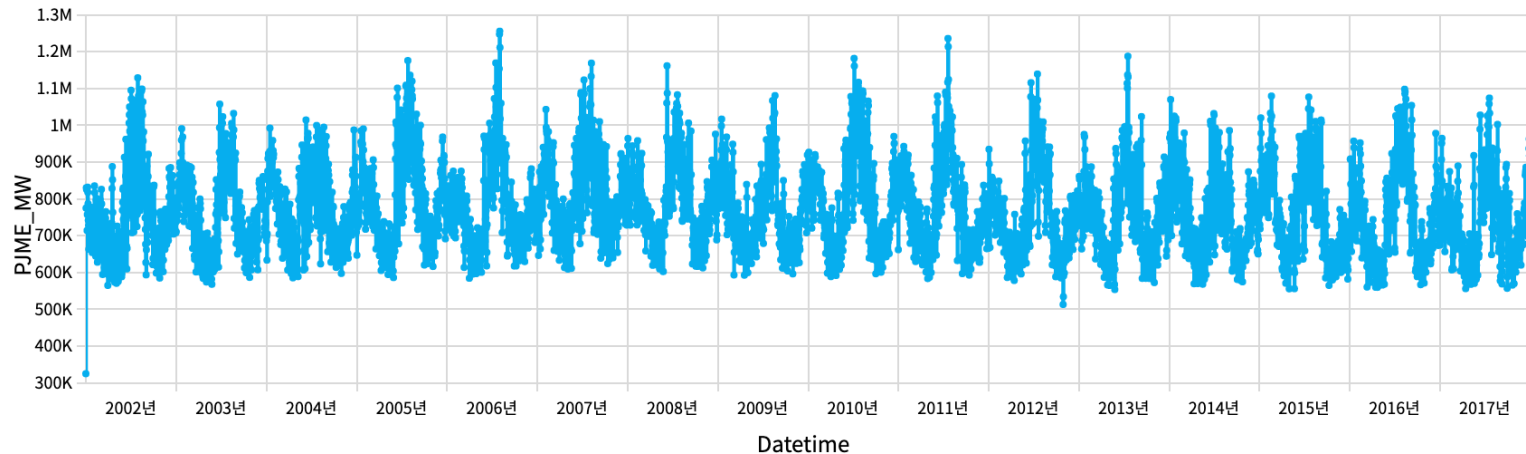


Time-Series ML Model(XGBoost)

PJM Hourly Energy Consumption Data

- PJM 전력회사의 시간별 에너지 소비 데이터
- XGBoost: AR 모델과 달리 시간에 대한 개념없이 feature를 사용해서 분류/예측하는 tree 모형
- <https://www.kaggle.com/code/soheiltehranipour/xgboost-timeseries-energy-consumption>

Y: PJME_MW (합) X: Datetime 🕒 하위그룹: 없음 화면분할: 없음 상관계수: -0.08 



실습 시간

<https://www.heartcount.io/login>