

# 시계열 데이터 분석

# 시계열 데이터 분석

## 데이터 분류(Data Classification)

Data Classification		
데이터 타입	int, float 등의 구분	Data Type
데이터 특성	categorical, numerical 구분	Data Attribute
데이터 유형	정형/비정형/반정형 데이터 구분	Data Category


# 시계열 데이터 분석

## 데이터 분류 (Data Classification)

### 1) 데이터 타입 (Type)

- int (정수형)
- float (실수형)
- str (문자열)
- bool (불리언)

### 3) 데이터 유형 - (Category)

- 정형
  - 비정형(시계열)
  - 반정형
- 

### 2) 데이터 특성 (Data Attribute)

- 범주형 데이터 (Categorical Data)
  - 명목형(Nominal)
  - 서열형(Ordinal)
- 수치형 데이터 (Numerical Data)
  - 이산형(Discrete)
  - 연속형(Continuous)

# 데이터.변수\_분류.pdf

# 시계열 데이터 분석

## 시계열 데이터 - Time Series

시계열 : 일정 시간 간격으로 배치된 데이터들의 수열

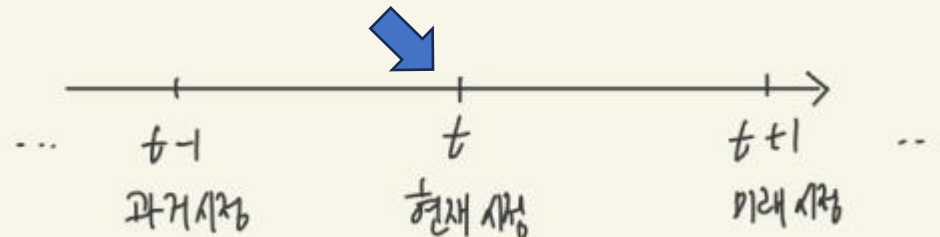
시계열 해석 : 시계열을 해석하고 이해하는 데 쓰이는 여러 가지 방법을 연구하는 분야. - wiki

Time-series data : 시간 순서에 따라 관측된 데이터

주식 가격, 기상 정보, 웹사이트의 사용자 트래픽 등 다양한 분야에서 사용

시계열 데이터의 주요 특징은 시간적 순서를 따르는 점과 연속성이 있음

time series data 표기법



# 시계열 데이터 분석

## 시계열 데이터 - Time Series

Time-series data : 시간 순서에 따라 관측된 데이터

**Air Passengers**

Month	#Passengers
Jan-49	112
Feb-49	118
Mar-49	132
Apr-49	129
May-49	121
Jun-49	135
Jul-49	148
Aug-49	148
Sep-49	136
Oct-49	119

**Temperature Data**

1960_3_m	1960_4_m	1960_5_m	1960_6_m	1960_7_m	1960_8_m
7.9	3.6	13.5	17.3	24	27.1
8.5	4.2	17	19	24.1	27.9
6.4	7.4	21.5	19.2	24.5	28.6
8.4	8.2	17.6	18	25.4	28.8
8.7	6	15.6	19.9	25.2	26.2
12.7	9.2	13.4	20.2	24.3	28
13.7	6.1	14.8	21.4	23.3	28.4
12.1	9.3	14.5	23.5	24.8	28.1

# 시계열 데이터 분석

## 시계열 데이터 - Time Series

### Q. 시계열 데이터?



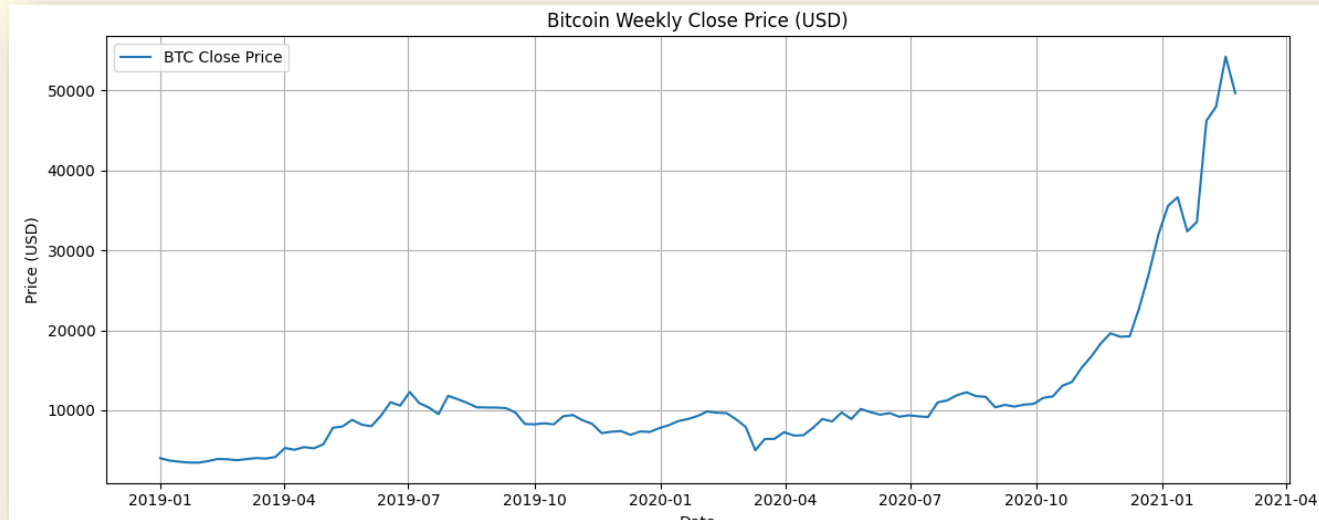
# 시계열 데이터 분석

## 시계열 데이터 - Time Series

- 실습

### 5.02.Timeseries.YF.Data.ipynb

Yahoo Finance 데이터 가져오기



# 시계열 데이터 분석

## 시계열 데이터 - Sequence

### ■ Sequence Data                      cf, Time-series data

- 시퀀스 데이터란 순서대로 정렬된 데이터의 연속이다.
- 이 데이터 유형은 특정 순서에 따라 배열된 항목들로 구성되며, 각 항목의 순서는 데이터 해석에 있어 중요한 역할을 한다.
- 시퀀스 데이터는 다양한 분야에서 나타나며, 예를 들어 생물학에서는 DNA 서열, 컴퓨터 과학에서는 문자열이나 시간에 따른 이벤트 로그, 금융에서는 시간에 따른 주식 가격 변동 같은 형태로 나타난다.

- wiki



# 시계열 데이터 분석

## 시계열 데이터 - Sequence

### ■ Sequence Data

- 시퀀스 데이터란 순서대로 정렬된 데이터의 연속이다.

“The movie is not fun”

```
9 embedding_layer_weights[4]
```

```
Embedding layer weights shape: (10000, 64)
```

```
Initial vector for the first word:
```

```
[-0.00259084  0.01449068 -0.0360497  0.04829049  0.03736104  0.0439479
 0.04184875  0.03720633 -0.00947944 -0.02207879  0.01567849  0.01323043
-0.00623485 -0.02969201  0.02015174  0.0419557  0.03964961  0.03541067
-0.04929231 -0.02532177 -0.01945633  0.0294967  -0.02734786 -0.00235792
 0.04007456  0.01345647 -0.02578268  0.02072446 -0.01814718 -0.00438038
-0.00013449  0.04992953  0.00172102 -0.02740901 -0.04531569  0.01248584
 0.03048811 -0.01235138 -0.03553003 -0.0029698  -0.00998558 -0.0273586
-0.01493003 -0.00096365 -0.0060817  0.01747804 -0.00963595  0.04506803
-0.02521282 -0.04632554 -0.0016393  -0.02918473  0.02854191 -0.00418777
-0.01391535 -0.01575425 -0.00093541 -0.00559113 -0.02639507  0.0006377
 0.01263886 -0.00140343  0.00334054  0.03750087]
```

# 시계열 데이터 분석

---

## Data Category

- 정형 데이터 (Structured Data)
- 시계열 데이터 (비정형, 반정형 데이터)
  - Timeseries Data
  - Sequence Data
- 비정형 데이터 (Unstructured Data)

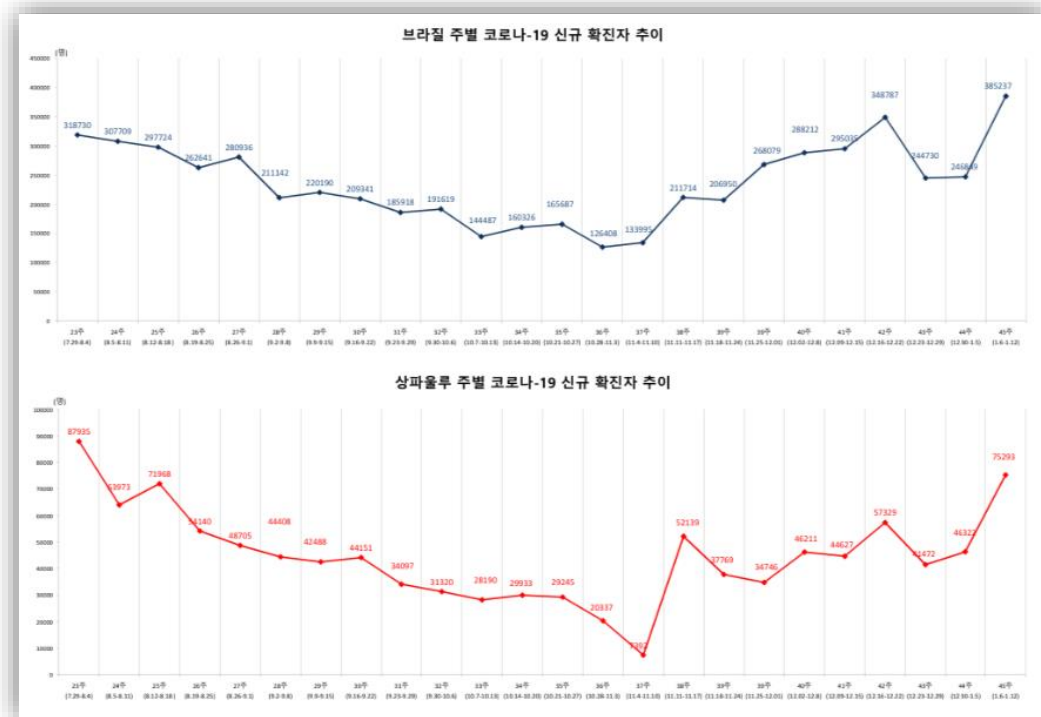
# 시계열 데이터 분석

## • 시계열 예측

: 시간 순서(Time series)로 수집 or 정렬한 순차적인(Sequential) 데이터 활용, 미래 시점의 상태 예측

ex) 일별 코로나19 확진자 수 데이터 -> 확진자 수 예측 -> 방역 정책 수립 활용

: 주 활용 분야 : 경제, 금융, 기상학



# 시계열 데이터 분석

- 시계열 예측

- ✓ 완벽한 예측X, 변동 가능성 높음
- ✓ 데이터셋 구성 변형 어려움
- ✓ 좋은 예측 성능을 위해서 데이터의 성질 고려필요
- ✓ 데이터가 정상성을 띠지 않을 때 시계열 예측 정확도가 높음  
정상성 갖추지 못했다면 변환 필요 (차분, 로그변환 등)

# 시계열 데이터 분석

## 시계열 데이터의 특성 - 1

- **시간 종속성 (Time Dependence)**
  - 시간 순서대로 배열 = 데이터 포인트가 시간적으로 서로 종속적인 관계를 지님
  - 과거의 데이터가 미래를 예측
- **불규칙성 (Irregularity)**
  - 데이터에 노이즈로 작용하는 예측할 수 없는 변동성
  - 외부 이벤트나 예측 불가능한 변화에 의해 발생 <- 예측 불가능성 UP
- **자기상관(Autocorrelation)**
  - 특정 시점의 값이 **이전 시점의 값들과 통계적으로 유의미한 관계**를 가짐
  - 예측 모델은 이 관계를 학습하여 미래 값을 예측
  - ex, 오늘의 기온은 어제와 비슷할 가능성이 높음

# 시계열 데이터 분석

## 시계열 데이터의 특성 - 2

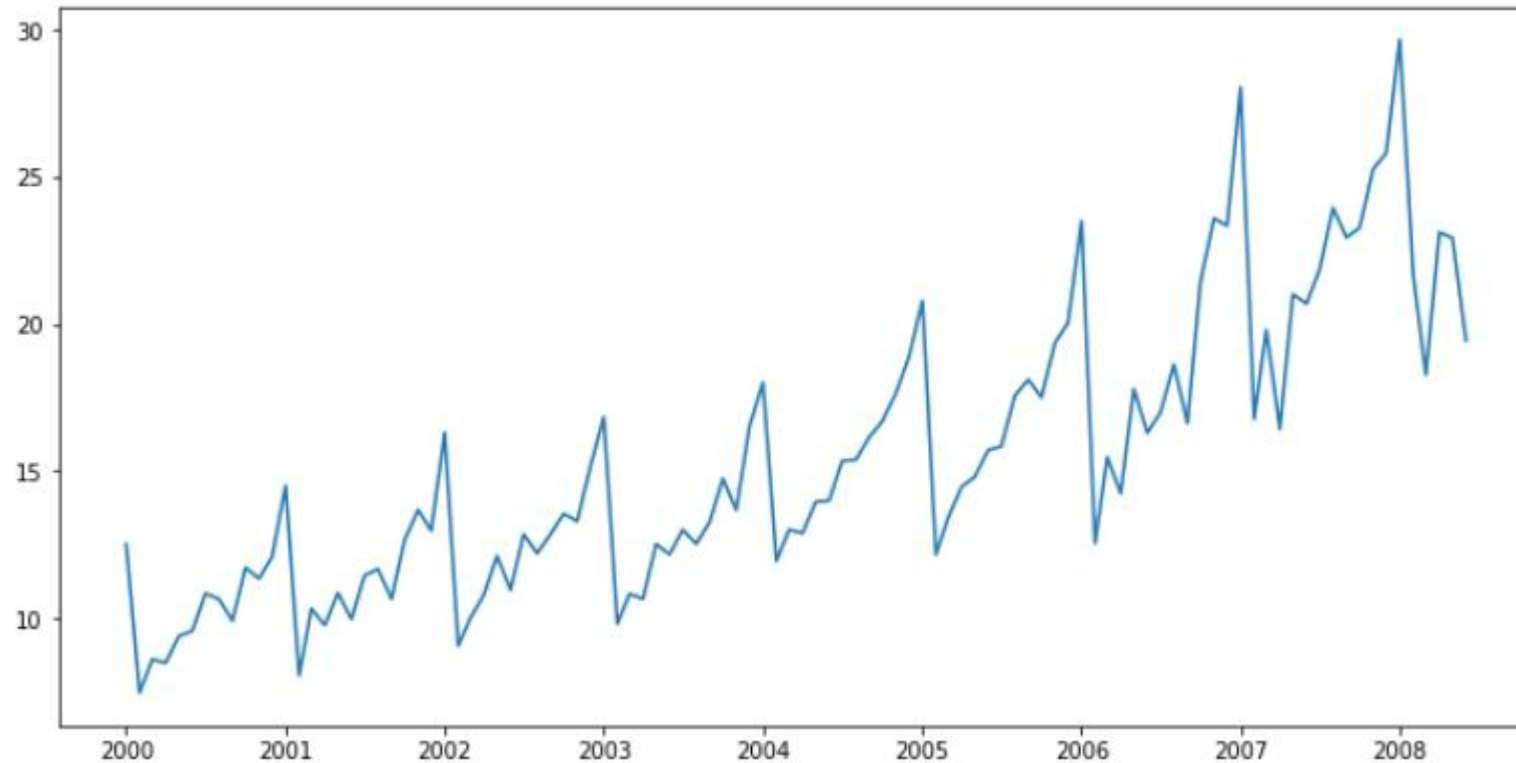
- **계절성** (Seasonality)
  - 일정 기간을 주기로 반복되는 패턴  
ex, 매년 겨울 - 난방 수요가 증가, 여름 - 냉방 수요가 증가 <= 계절적 요인.
- **추세** (Trend)
  - 데이터에서 장기간에 걸쳐 나타나는 상승 또는 하락 경향  
ex, 인구 증가
- **주기성** (Cyclic)
  - 고정된 기간 동안 발생하지 않고(cf 계절성), 불규칙적인 간격으로 발생  
ex, 경제 불황기 or 호황기 경제 주기
- **정상성** (stationarity)
  - 통계적 속성(평균, 분산, 공분산 등)이 시간에 따라 일정하게 유지 되는 성질
  - 시계열 분석의 대부분이 데이터의 정상성을 가정하여 수행

# 시계열 데이터 분석

## 시계열 데이터의 특성 - 계절성 (Seasonality)

주기적으로 반복되는 변화

ex, 매년 겨울 - 난방 수요가 증가, 여름 - 냉방 수요가 증가 <= 계절적 요인



호주 당뇨병 치료약 월별 sales데이터

# 시계열 데이터 분석

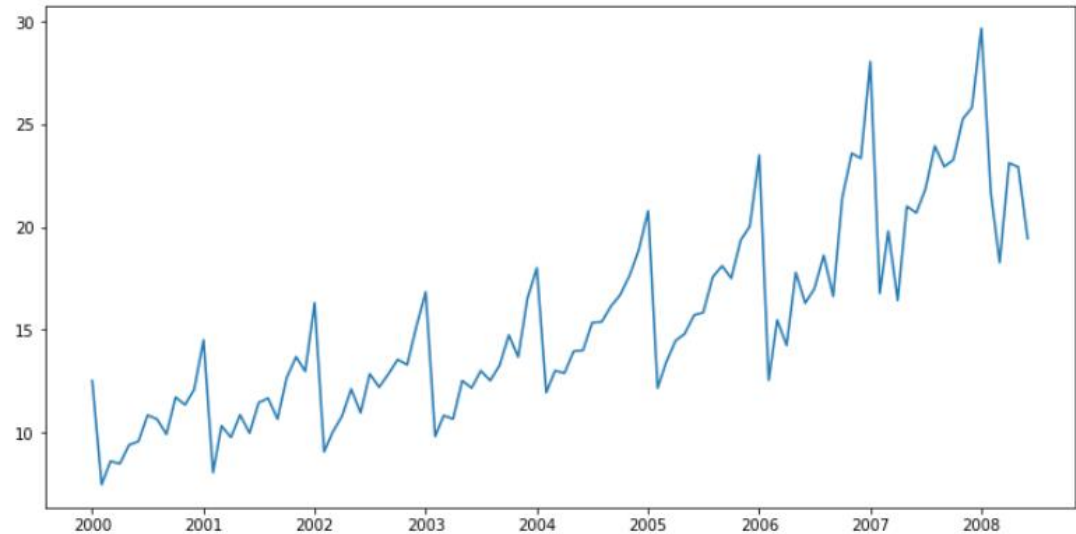
## 시계열 데이터의 특성 – 추세 (Trend)

### 장기적으로 증가하거나 감소하는 패턴

ex, 인구 증가, 농업 생산, 제조 품목, 출생 및 사망자 수, 산업 또는 공장 수, 학교 또는 대학 수 등



주가 시계열



Q. 각 그래프의 트렌드는?



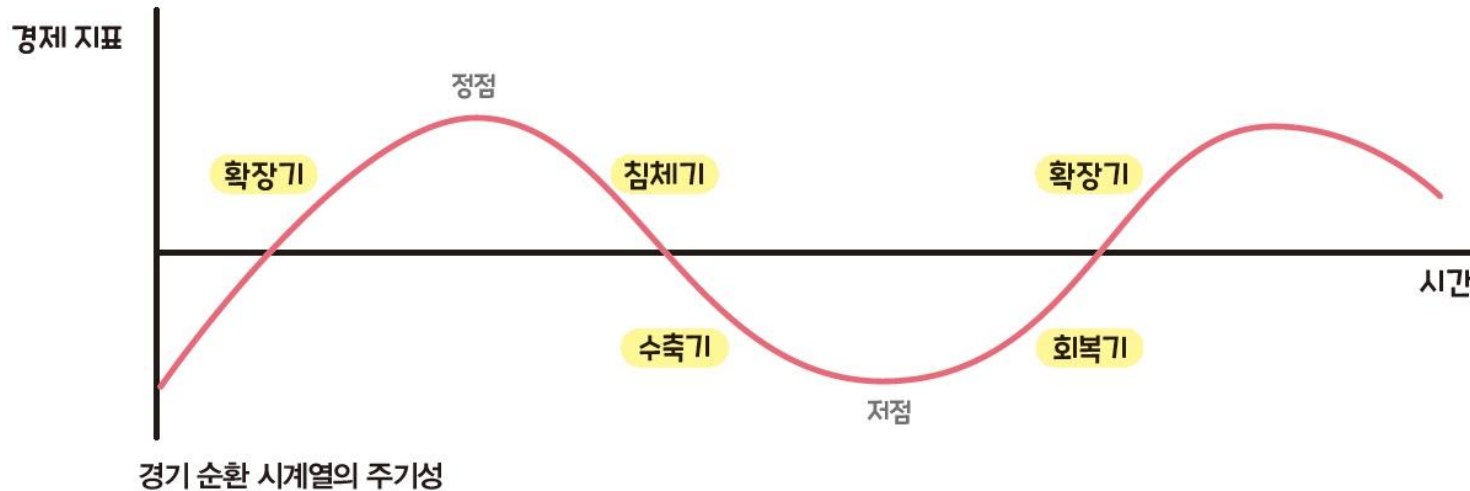
# 시계열 데이터 분석

## 시계열 데이터의 특성 – 주기성 (Cyclic)

정해진 빈도 없이 값의 증가나 감소가 반복되는 패턴

규칙적이지 않은 변동. 보통 불규칙한 기간을 가지며, 사이클의 길이가 고정되어 있지 않음.

cf, 계절성. ex, 경제 불황기 or 호황기 경제 주기



# 시계열 데이터 분석

## 시계열 데이터의 특성 – 계절성 (Seasonality) vs 주기성 (Cyclic)

- 계절성 : 주기적으로 반복되는 변화
- 주기성 : 정해진 빈도 없이 값의 증가나 감소가 반복되는 패턴

구분	계절성 (Seasonality)	주기성 (Cyclic)
반복	규칙적인 간격으로 반복됨	불규칙한 간격으로 반복됨
주기 길이	고정 (ex) 12개월, 7일 등)	가변적 (예측 불가, ex) 몇 년)
원인	달력 기반 (계절, 요일, 월 등)	경제적/사회적 흐름, 장기적 영향
예시	여름 아이스크림 매출 증가, 월말 카드 사용량 급증	경기 순환(확장-수축), 부동산 호황/불황 주기

# 시계열 데이터 분석

시계열 데이터의 특성 - 정상성 (stationarity) - 시간에 따른 변화가 없는 상태

- ✓ 그래프에 지속적인 상승 또는 하락 추세가 없음  
= 특정한 트렌드 (추세)가 존재하지 않는 성질
- ✓ 시계열의 확률적인 성질들이 시간의 흐름에 따라 불변  
= 과거의 변동폭과 현재의 변동폭이 같음
- ✓ 평균, 분산, 공분산 등이 시간의 경과와 무관하게 일정  
= 계절성이 없음

## 정상성(Stationarity)

"시계열의 안정적 수준, 정상이란 변하지 않고 일정한 상태를 의미"

"과거 관찰값을 바탕으로 미래를 예측하기 위해서는 수집된 관측값이 안정적으로 유지되고 있는 상태인지, 또는 계속해서 변동하는 상태인지를 확인"

# 시계열 데이터 분석

---

시계열 데이터의 특성 – 정상성 (stationarity)    - 시간에 따른 변화가 없는 상태

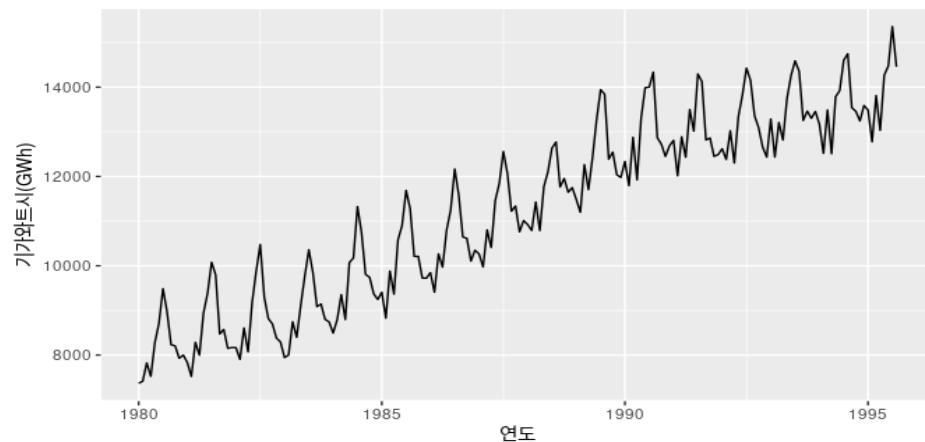
- Stationary : 변화 없는
- Stationarity : 정상성  
: 시계열 데이터가 시간에 따라 통계적 성질(평균, 분산, 공분산 등)이 변하지 않는 상태

# 시계열 데이터 분석

## 시계열 데이터의 특성 - 정상성 (stationarity)

- ✓ 그래프에 지속적인 상승 또는 하락 추세가 없음  
= 특정한 트렌드 (추세)가 존재하지 않는 성질
- ✓ 시계열의 확률적인 성질들이 시간의 흐름에 따라 불변  
= 과거의 변동폭과 현재의 변동폭이 같음
- ✓ 평균, 분산, 공분산 등이 시간의 경과와 무관하게 일정  
= 계절성이 없음

시계열의 패턴은  
추세, 계절성, 주기성이 복합적으로 발생



Figure

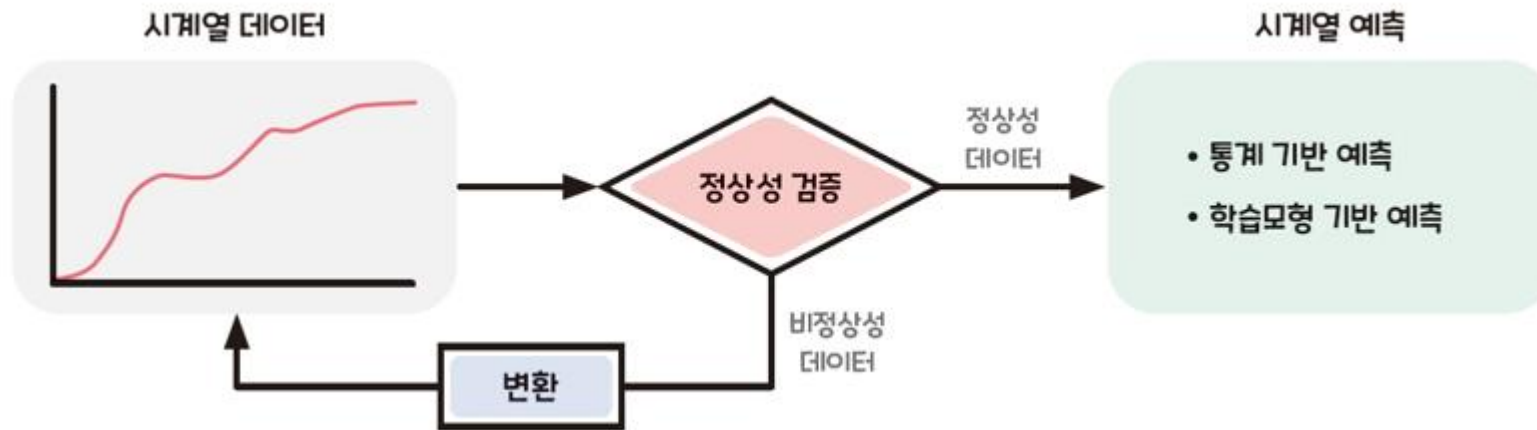
1부터 1995년까지 호주 월별 전력 수요.

# 시계열 데이터 분석

## 시계열 데이터의 특성 – 정상성 (stationarity)

✓ 비정상성 데이터는 정상성을 가지도록 변환 후 분석

-> 통계 기반 시계열 예측의 예측 정확도 향상



# 시계열 데이터 분석

## 시계열 데이터의 특성 – 정상성 (stationarity)

### 정상성 판단

시각적 판단	통계적 판단
그래프에 지속적인 상승 또는 하락 추세가 없음	평균이 일정
과거의 변동폭과 현재의 변동폭이 같음	분산이 시점에 독립적
계절성이 없음	공분산이 시차에 의존적이나 시점에 독립적

# 시계열 데이터 분석

## 시계열 데이터의 특성 – 정상성 (stationarity)

### 정상성 판단

```
ADF test with TSLA time-series  
ADF Stats: -1.301076  
p-value: 0.628680  
Critical values:  
  1%: -3.4356  
  5%: -2.8639  
 10%: -2.5680
```

귀무 가설 : 데이터가 비정상성

대립 가설 : 데이터가 정상성

- p-값이 0.05 이하이거나
- ADF 검정 통계량(ADF Stats)이 임계값(Critical values) 보다 작으면  
=> 정상성 판단

```
ADF_result = adfuller(df_tsla.values)
```



# 시계열 데이터 분석

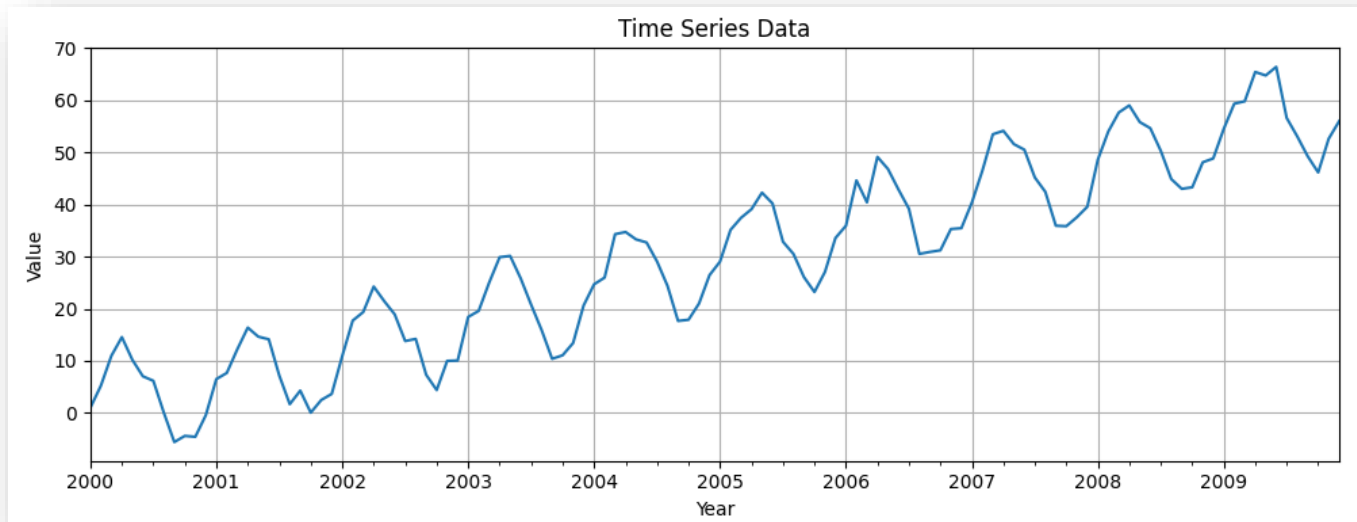
시계열 데이터의 특성 – 정상성 (stationarity)

정상성 판단 – 통계적 방법

- 실습

## 5.02.TimeSeries.Staionarity

Q. 그래프를 통한 판단은?



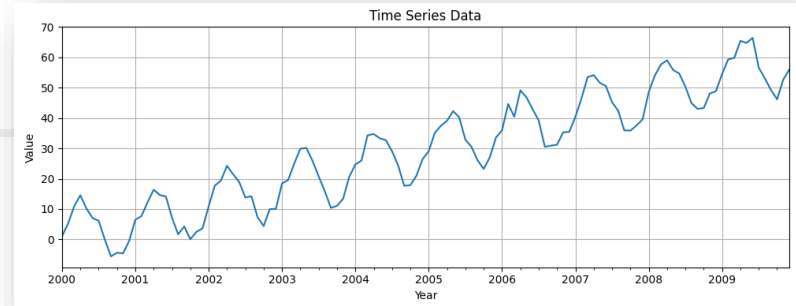
# 시계열 데이터 분석

## 시계열 데이터의 특성 – 정상성 (stationarity)

### 정상성 판단 – 통계적 방법

- 실습

## 5.02.TimeSeries.Staionarity



```
ADF 통계량: 0.906757559310063  
p-value: 0.9931778836425875  
임계값: {'1%': np.float64(-3.4936021509366793), '5%': np.float64(-2.8892174239808703), '10%': np.float64(-2.58153320754717)}
```

ADF 테스트의 귀무가설( $H_0$ ) :

$H_0$ : 이 시계열은 비정상(non-stationary)이다

$H_1$ : 대립가설, 정상 시계열(stationary)이다

p-value: 0.9931

→ 매우 크므로 귀무가설을 기각할 수 없음

→ 이 시계열은 정상성이 없다, 즉 비정상(non-stationary) 으로 판단

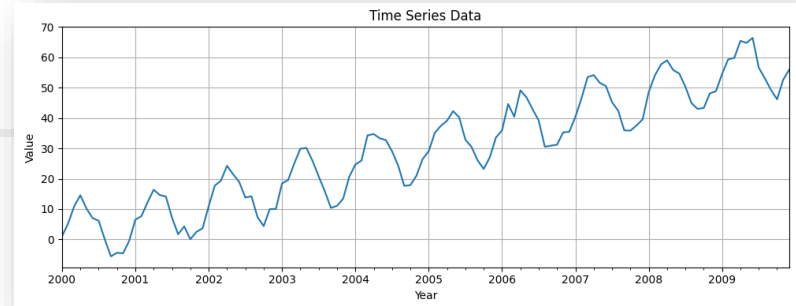
# 시계열 데이터 분석

## 시계열 데이터의 특성 – 정상성 (stationarity)

### 정상성 판단 – 통계적 방법

- 실습

## 5.02.TimeSeries.Staionarity



```
ADF 통계량: 0.906757559310063
p-value: 0.9931778836425875
임계값: {'1%': np.float64(-3.4936021509366793), '5%': np.float64(-2.8892174239808703), '10%': np.float64(-2.58153320754717)}
```

ADF 테스트의 귀무가설( $H_0$ ) :

$H_0$ : 이 시계열은 비정상(non-stationary)이다

$H_1$ : 대립가설, 정상 시계열(stationary)이다

p-value: 0.9931

→ 매우 크므로 귀무가설을 기각할 수 없음

→ 이 시계열은 정상성이 없다, 즉 비정상(non-stationary) 으로 판단

# 시계열 데이터 분석

시계열 데이터의 특성 - 정상성 (stationarity)

정상성 판단 - 통계적 방법

- 실습

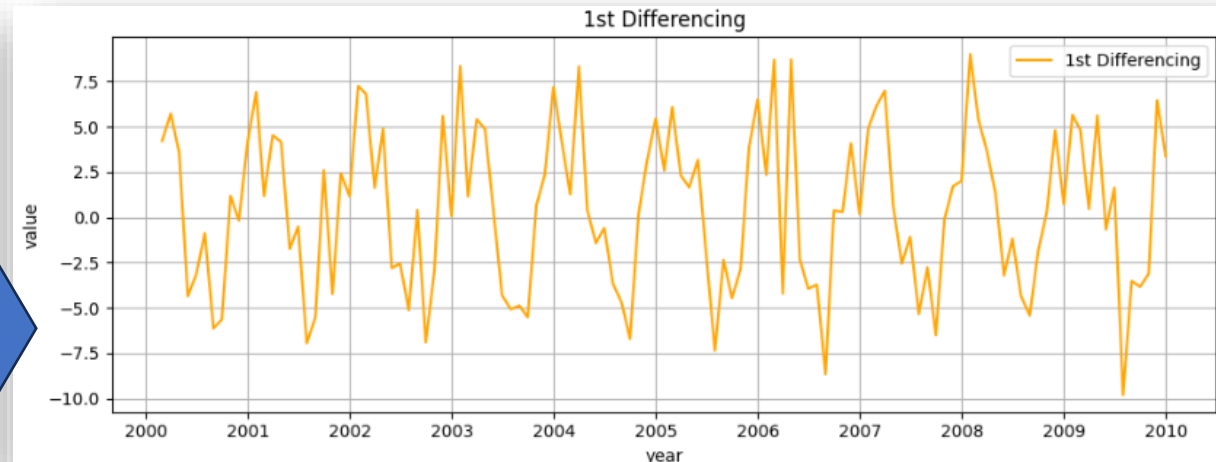
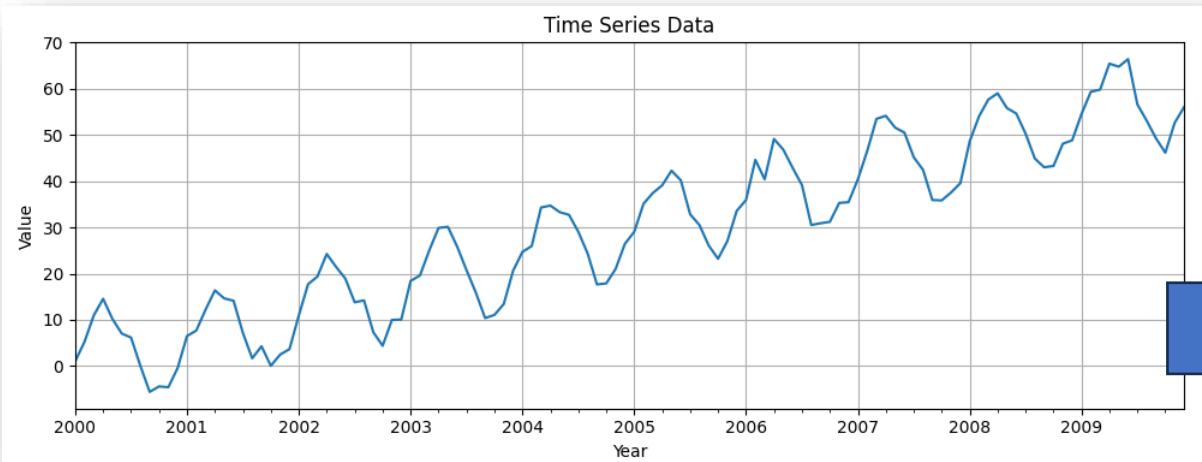
## 5.02.TimeSeries.Staionarity

- 정상성 변환 - 차분(Differencing):

현재 값에서 이전 값을 빼는 방법으로, 시계열에서 추세를 제거하고 정상성을 만드는 방법

```
result2 = adfuller(ts_diff)
```

ADF 통계량: -7.853306135842218  
p-value: 5.526927424569128e-12  
임계값: {'1%': np.float64(-3.4936021509366793),



# 시계열 데이터 분석

시계열 데이터의 특성 - 정상성 (stationarity)

- 정상성 변환 - 차분(Differencing):

현재 값에서 이전 값을 빼는 방법으로, 시계열에서 추세를 제거하고 정상성을 만드는 방법

```
result2 = adfuller(ts_diff)
```



**Q. 변형(차분 등)된 데이터로 분석,  
그 결과를 원래 데이터 관점에서 어떻게 해석?"**

# 시계열 데이터 분석

---

## 시계열 데이터의 특성 – 실습 : 테슬라 주가 추세

### 5.02.TimeSeries.yfinance.tsla.ipynb

```
# 야후 파이낸스 주가 데이터
import yfinance as yf
import pandas as pd
import matplotlib.pyplot as plt

# 테슬라
tsla = yf.download('TSLA', start='2018-01-01', end='2022-12-31')
tsla

# 테슬라 주식 종가
df_tsla = pd.DataFrame(tsla['Close'])

# 시각화
df_tsla.plot(figsize=(12.2, 6.4))
```

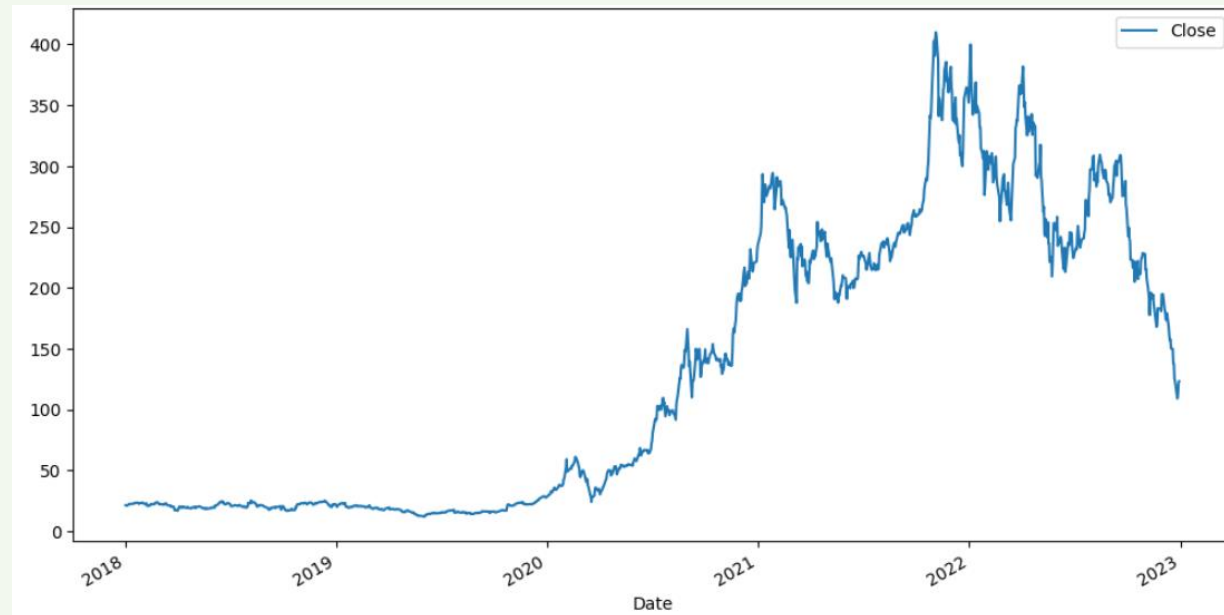
# 시계열 데이터 분석

시계열 데이터의 특성 – 실습 : 테슬라 주가 추세

Q. 추세 ?

Q. 계절성 ?

Q. 주기성 ?



# 시계열 데이터 분석

## 시계열 데이터의 특성 - 정상성 (stationarity) 검증

```
from statsmodels.tsa.stattools import adfuller # Import the adfuller function

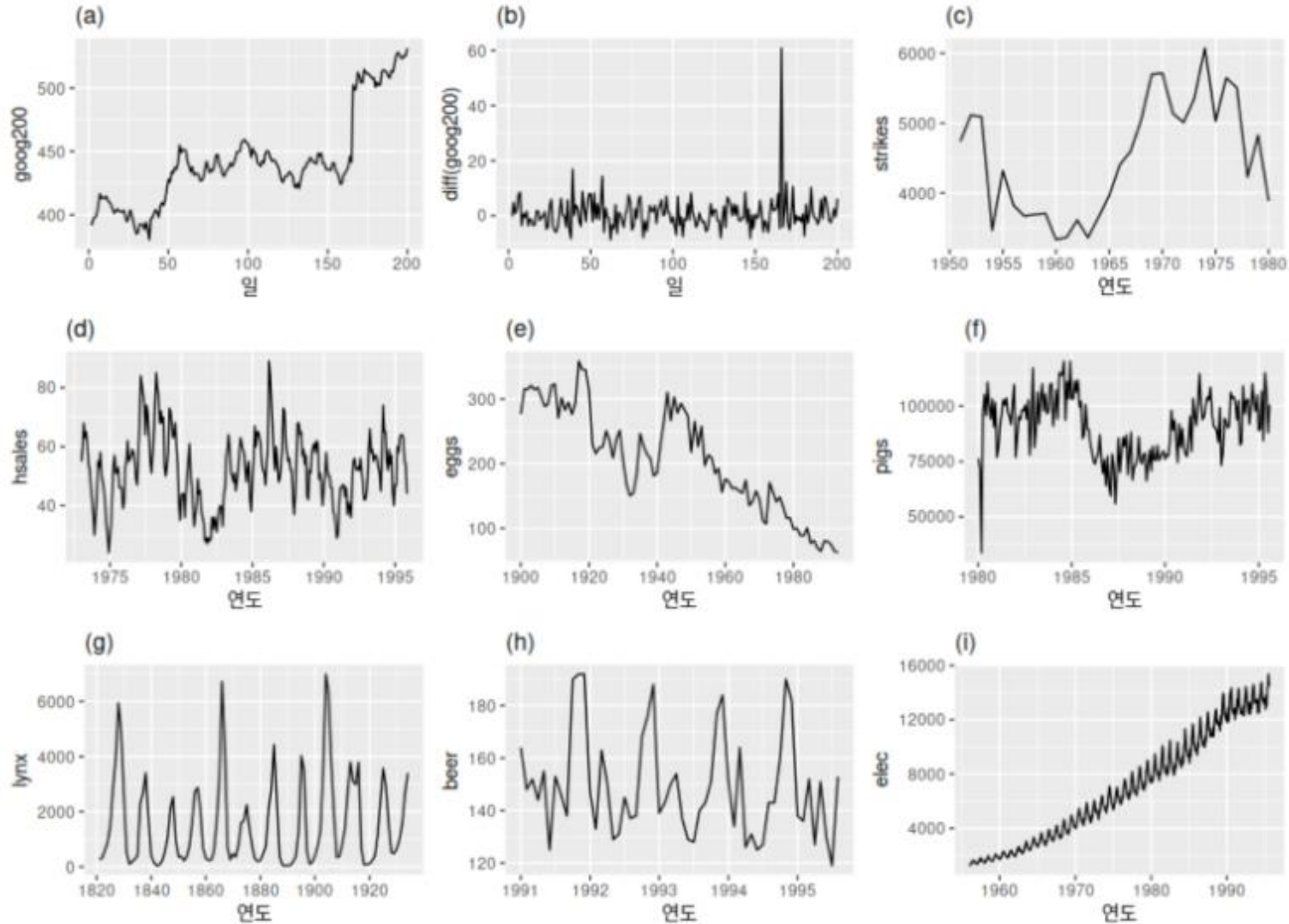
print('ADF test with TSLA time-series')
ADF_result = adfuller(df_tsla.values)
#ADF 통계량
print('ADF Stats: %f' % ADF_result[0])
#p-값
print('p-value: %f' % ADF_result[1])
#임계값
print('Critical values: ' )
for key, value in ADF_result[4].items():
    print('\t%s: %.4f' % (key, value))
```

```
ADF test with TSLA time-series
ADF Stats: -1.301076
p-value: 0.628680
Critical values:
          1%: -3.4356
          5%: -2.8639
         10%: -2.5680
```



# 시계열 데이터 분석

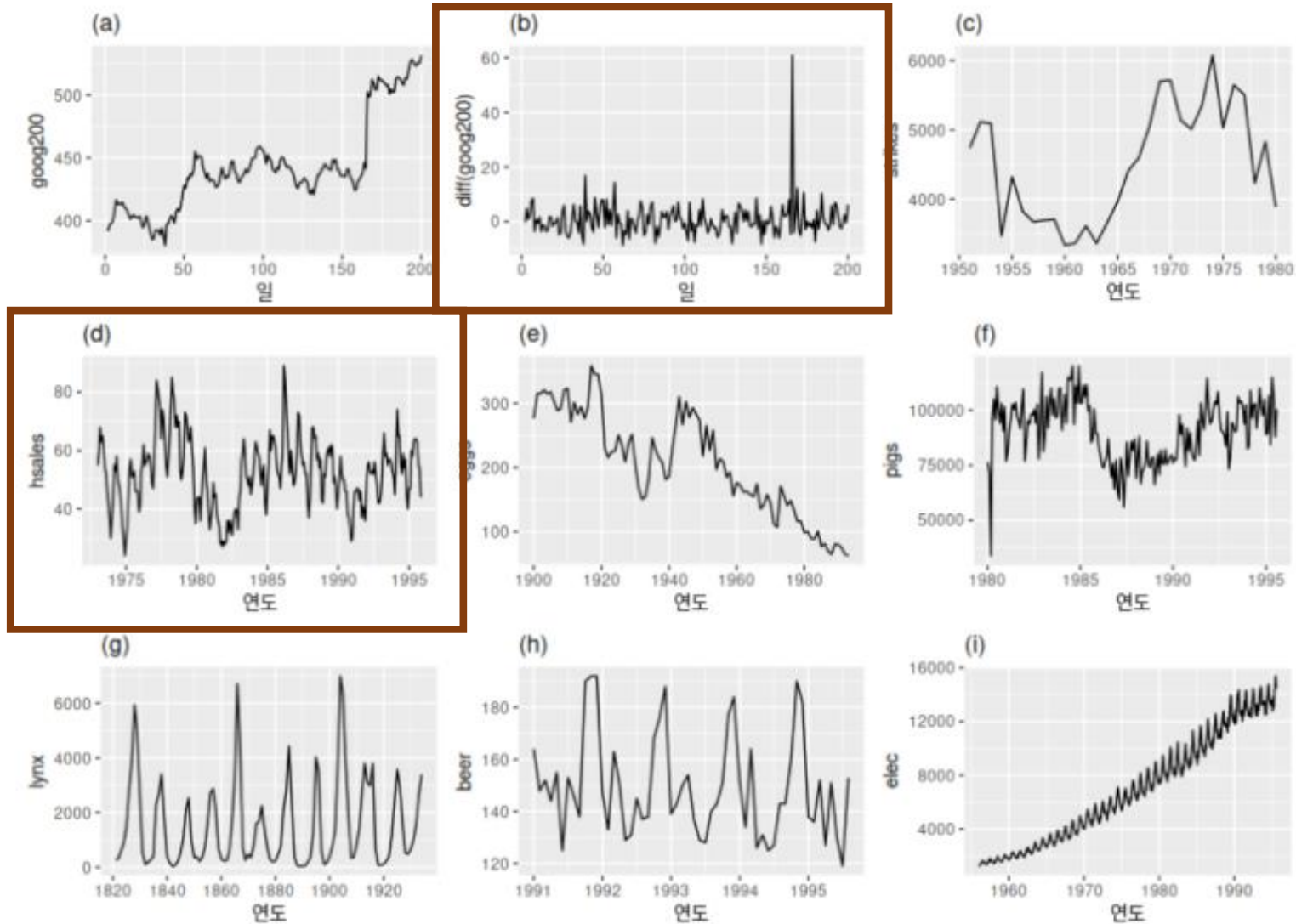
## 시계열 데이터의 특성 - 정상성 (stationarity)



Q. '정상성'을 지닌 데이터는?

# 시계열 데이터 분석

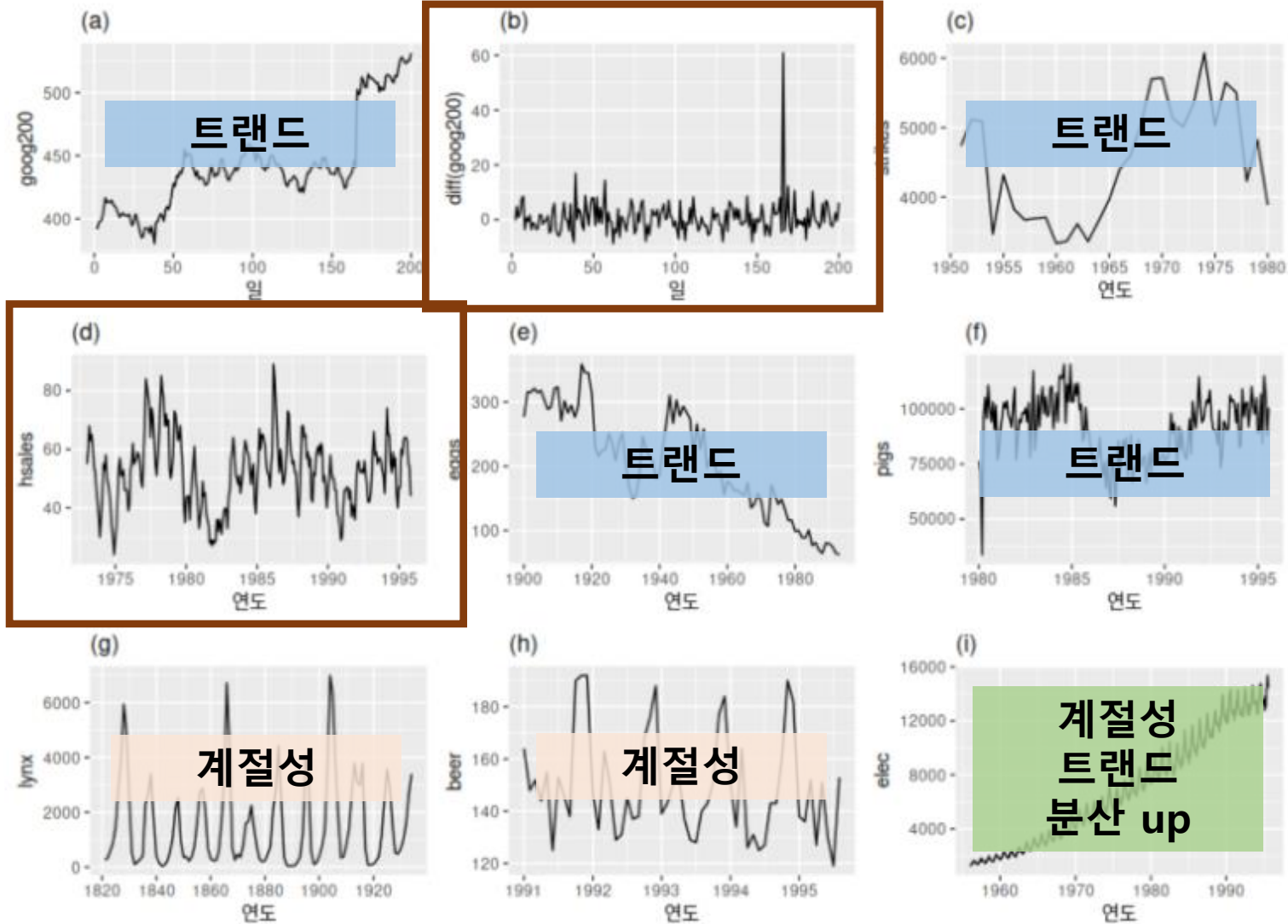
## 시계열 데이터의 특성 - 정상성 (stationarity)



Q. '정상성'을 지닌 데이터는?

# 시계열 데이터 분석

## ■ 정상성(stationarity)



Q. '정상성'을 지닌 데이터는?

# 시계열 데이터 분석

## 시계열 데이터 분석 기법

- **ARIMA (자동회귀누적이동평균)**

시계열 데이터의 비계절성 부분을 모델링할 때 사용  
데이터의 추세와 불규칙성을 설명하는 데 유용

- **계절적 ARIMA (SARIMA)**

ARIMA의 확장 형태로 계절성 요소를 추가하여 계절적 변동을 모델링

- **지수평활법**

데이터의 추세와 계절성을 예측하는 데 적합  
다양한 평활 파라미터를 조정하여 미래 값을 예측합니다.

- **LSTM (Long Short-Term Memory):**

Deep Learning 모델  
시계열 데이터에서 장기 의존성을 모델링하는 데 효과적

# 시계열 데이터 분석

## 시계열 데이터 분석 기법 - ARIMA

- **Auto**Regulation **I**ntegrated **M**oving **A**verage
- 자기회귀 모형과 이동평균 모형을 결합한 모형.
- 자기회귀(AR) 모형: 데이터의 과거 값들을 선형으로 조합하여 미래 값을 예측.
- 이동평균(MA) 모형: 데이터가 어떤 방향성을 가지고 증가하거나 감소할 때, 즉 데이터가 이동할 때마다 평균을 구하여 예측에 활용.
- ARMA 모형: AR 모형과 MA모형을 결합한 모형.

\* AR, MA, ARMA 모형은 모두 정상성 시계열 데이터를 전제

# 시계열 데이터 분석

---

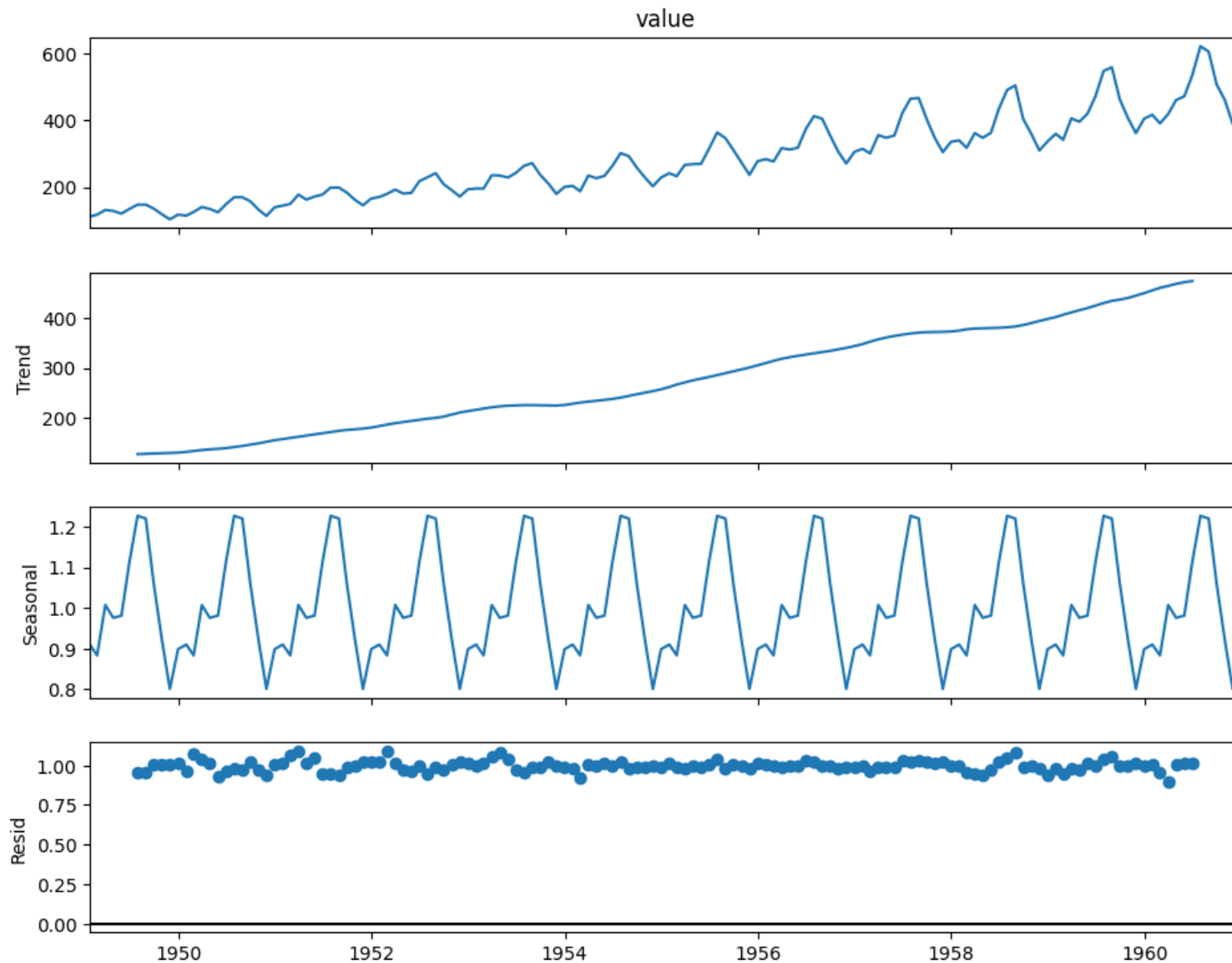
- 실습

5.02.Time series forecasting.exam.ipynb

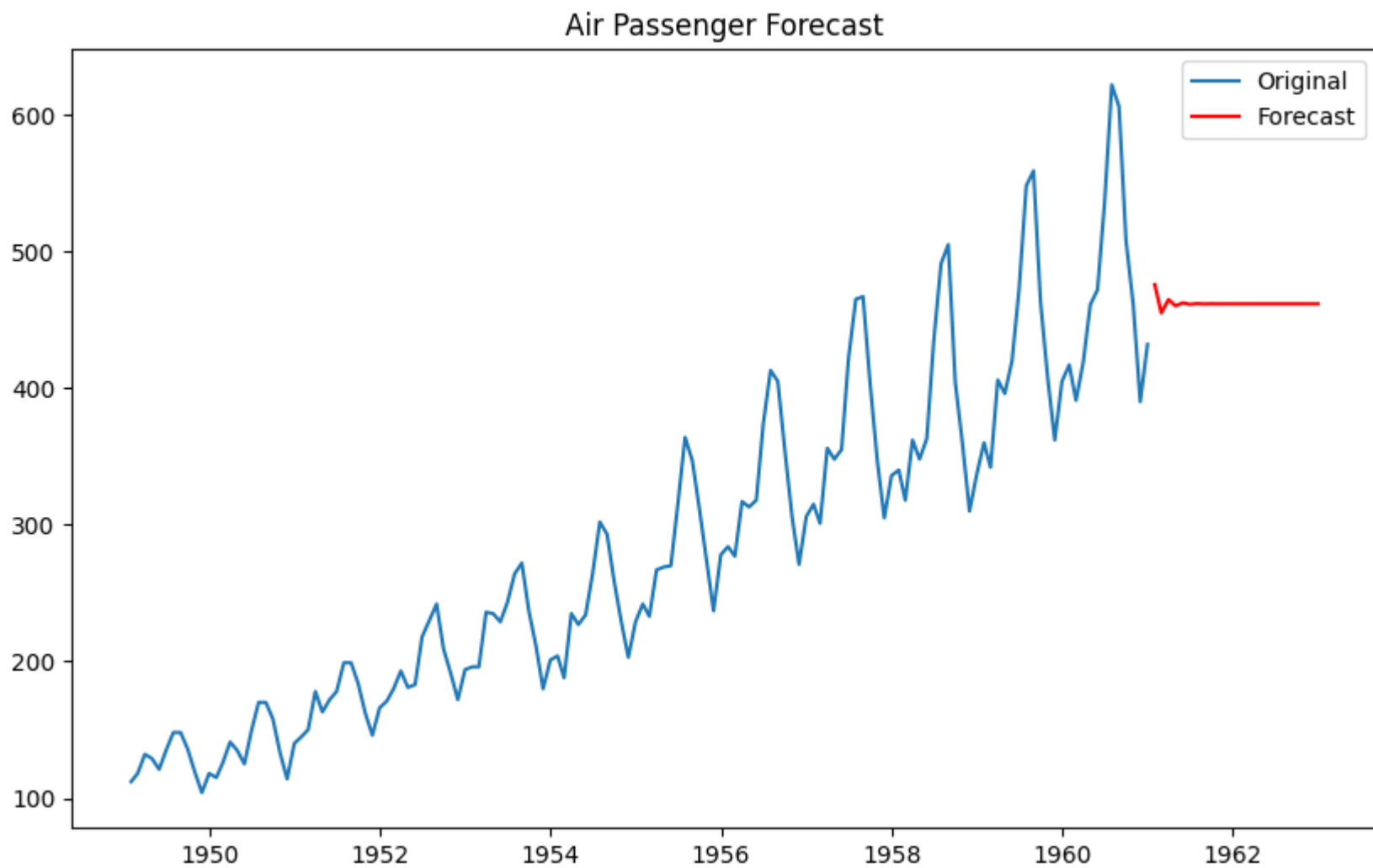
5.02.Time series forecasting.AirPassengers.ipynb

# 시계열 데이터 분석

## AirPassengers



# 시계열 데이터 분석



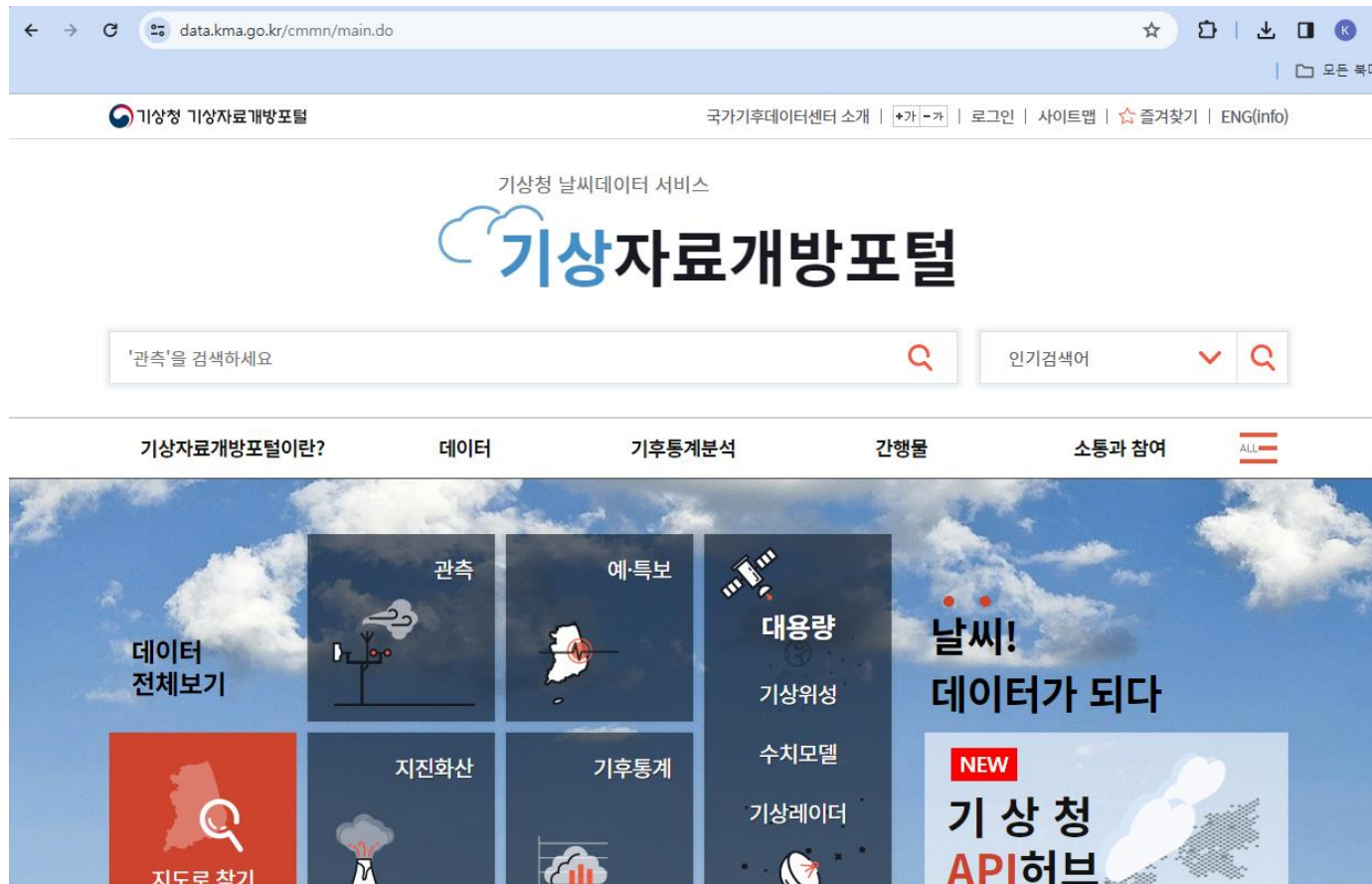


# 시계열 데이터 분석

## 광주 기온 데이터 - 실습

- 1960년 1월 부터 2021년 8월 까지의 광주지역 일일 기온 데이터 수집하기

<https://data.kma.go.kr/cmmn/main.do>



# 시계열 데이터 분석

## 광주 기온 데이터 - 실습

<https://data.kma.go.kr/climate/RankState/selectRankStatisticsDivisionList.do?pgmNo=179>

 기상청 기상자료개방포털

국가기후데이터센터 소개 | [\\*가|-가](#) | 로그인 | 사이트맵 | ☆ 즐겨찾기 | [ENG\(info\)](#)



[기상자료개방포털이란?](#) | [데이터](#) | [기후통계분석](#) | [간행물](#) | [소통과 참여](#) | 

  
기후통계분석

[평년값](#)   
[통계분석](#)   
[조건별통계](#)  
[기온분석](#)  
[강수량분석](#)  
[다중지점통계](#)  
[24절기](#)  
[순위값](#)  
[장마](#)  
[.....](#)

[Home](#) > [기후통계분석](#) > [통계분석](#) > [조건별통계](#)  

### 조건별통계

[자료설명](#) [> 사용법](#)

기온, 강수량, 바람 자료 대상으로 원하는 조건의 자료를 검색할 수 있습니다.

\* '지역/지점'의 '지역'은 전국 및 광역 단위의 평균 제공(1973년~)  
- 전국 및 광역별 평균 산출에 사용되는 지점은 62개 지점이며 제주도 4개 지점은 제주지역 산출에만 사용됩니다.  
[\[전국/지역별 통계 산출 지점 정보\(더보기\)\]](#)

#### 검색조건

분류 [지상](#) 

지역/지점 [광주](#) [선택](#)

요소 [기온](#) 

기간 [월](#)  2014  년 ~ 2024  년

조건

☐ 요소 [평균기온](#)  <  [선택](#) 

☒ 월 [01](#)  월 ~ [12](#)  월

# 시계열 데이터 분석

## 광주 기온 데이터 - 실습

temp\_mean\_gwangju.csv

	1960_1_m	1960_2_m	1960_3_m	1960_4_m	1960_5_m	1960_6_m	1960_7_m	1960_8_m	1960_9_m	1960_10_n	1960_11_n	1960_12_n	1961_1_m	1961_2_m	1961_3_m
0	2.7	1.7	7.9	3.6	13.5	17.3	24	27.1	24.9	18.8	14.2	1.8	-8.3	-6.8	9.4
1	2.6	0.9	8.5	4.2	17	19	24.1	27.9	24.8	17	10.1	2.1	-3.1	-2.4	13.1
2	6.4	2.5	6.4	7.4	21.5	19.2	24.5	28.6	25.3	15.2	9.4	4.6	0	-2.9	10.9
3	10.9	0.8	8.4	8.2	17.6	18	25.4	28.8	22.2	16.6	10.9	5.4	-3	-0.2	6.2
4	-0.2	1.2	8.7	6	15.6	19.9	25.2	26.2	21.6	16.3	10.2	3.5	-4.3	3.5	4.1
5	-2.2	6.5	12.7	9.2	13.4	20.2	24.3	28	19.5	17.9	9.8	0.6	-3	5	4.8
6	-0.3	9.7	13.7	6.1	14.8	21.4	23.3	28.4	21.3	18.4	10.7	1.6	-2.3	3.2	2.3
7	2.2	8.6	12.1	9.3	14.5	23.5	24.8	28.1	19.2	17	11.4	4.3	1.2	1.1	1.1
8	4.1	6.1	11.8	12.8	12.1	23.2	23.7	28.1	19.7	15.6	10.4	6.7	3.1	1.6	1
9	5.3	0.3	9.8	10.7	12.3	21.9	26.1	27.5	18.7	16.7	10.5	8.3	2.1	0.3	1.1
10	2	0	9.8	10.8	14.2	20.6	26.9	27.8	20.2	16.8	9.3	10	-6.2	1.4	2.9
11	0.4	-1.3	5	11.2	18.2	17.7	26.7	26.3	23.4	17.1	10	5.5	-5.4	-0.7	7.4
12	1.8	0.9	0.7	12.3	18.1	21.9	26.9	26.4	23.8	18.2	6.9	4.2	-2.3	1	9.6
13	3	1.9	1.8	9.1	14.6	22.8	26.6	24.9	19.7	18.1	7.2	6.7	-1.6	-2.7	5.6
14	1.5	0.3	3.6	10.6	17.1	21.5	26.6	26.4	19.3	19	7.5	4.9	-1.6	-2.3	6
15	-2.1	2.2	6.7	7.7	18.2	21.5	26.8	26.1	18.8	15.6	8.4	2.6	-3.7	1	8.3
16	-3.4	4	7.1	6.5	20	20.5	27.3	27.5	17.6	13.2	10.7	1.3	-4.7	1.3	8.5
17	-0.5	0.4	8.2	8.6	17.5	21.5	27.6	27.2		13	10	-3.6	-4	0.3	13.6
18	2.6	3.2	10	11.1	17.9	21.9	27.3	27.6		14.2	10.6	-1.3	-4.3	3	11.2
19	2.7	5	12	10.2	14.9	24.3	27.3	28.1		14.9	11.2	0.4	-1.3	5.4	9.5
20	-0.3	1.1	6.4	10.8	14.4	20.6	26.8	28.6		16.6	16.5	1.1	2.7	3.6	7.4

# 시계열 데이터 분석

---

광주 기온 데이터 - 실습

- 실습

5.02.TempAnal.ipynb

- TASK

성능향상

- 1) 파라미터 튜닝
- 2) NN모델 구조 변경

# 시계열 데이터 분석

## 광주 기온 데이터 - 실습

- 실습

### 5.02.TempAnal.ipynb

#### 기온 예측

30일을 입력값으로 사용 다음날 예측

설명변수 : 1일 ~30일까지의 날씨

목표변수 : 31일의 날씨

#### 학습 데이터

input	target
1960년 1월 1일 ~ 1월 30일	1960년 1월 31일
1960년 1월 2일 ~ 1월 31일	1960년 2월 1일
1960년 1월 3일 ~ 2월 1일	1960년 2월 2일
⋮	⋮
2021년 7월 18일 ~ 8월 15일	2021년 8월 16일

#### 테스트 데이터

input	predict
2021년 7월 19일 ~ 8월 16일	2021년 8월 17일
2021년 7월 20일 ~ 8월 17일	2021년 8월 18일
⋮	⋮
2021년 8월 1일 ~ 8월 30일	2021년 8월 31일

# 시계열 데이터 분석

광주 기온 데이터 - 실습

- 실습

5.02.TempAnal.ipynb

	0	1	2	3	4	5	6	7	8	9
1960_1_mean	2.7	2.6	6.4	10.9	-0.2	-2.2	-0.3	2.2	4.1	5.3
1960_2_mean	1.7	0.9	2.5	0.8	1.2	6.5	9.7	8.6	6.1	0.3
1960_3_mean	7.9	8.5	6.4	8.4	8.7	12.7	13.7	12.1	11.8	9.8
1960_4_mean	3.6	4.2	7.4	8.2	6.0	9.2	6.1	9.3	12.8	10.7
1960_5_mean	13.5	17.0	21.5	17.6	15.6	13.4	14.8	14.5	12.1	12.3

# 시계열 데이터 분석

## 광주 기온 데이터 - 실습

- 실습

### 5.02.TempAnal.ipynb

- bias

- $y = ax + b$

$$u_1 = \begin{bmatrix} w_1 & w_2 & w_3 & \dots & w_{30} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{30} \end{bmatrix} + b_1$$

# 시계열 데이터 분석

## 광주 기온 데이터 - 실습

- 실습

### 5.02.TempAnal.ipynb

$$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}^T = \begin{bmatrix} 0 & 3 & 6 \\ 1 & 4 & 7 \\ 2 & 5 & 8 \end{bmatrix}$$

	1960_1_mean	1960_2_mean	1960_3_mean	1960_4_mean	1960_5_mean	1960_6_mean	1960_7_mean	1960_8_mean
0	2.7	1.7	7.9	3.6	13.5	17.3	24.0	17.6
1	2.6	0.9	8.5	4.2	17.0	19.0	24.1	15.6
2	6.4	2.5	6.4	7.4	21.5	19.2	24.5	13.4
3	10.9	0.8	8.4	8.2	17.6	14.8	24.5	14.5
4	-0.2	1.2	8.7	6.0	15.6	13.4	24.5	12.1



df.T.head()

	0	1	2	3	4	5	6	7	8	9	...	21
1960_1_mean	2.7	2.6	6.4	10.9	-0.2	-2.2	-0.3	2.2	4.1	5.3	...	-3.4
1960_2_mean	1.7	0.9	2.5	0.8	1.2	6.5	9.7	8.6	6.1	0.3	...	1.8
1960_3_mean	7.9	8.5	6.4	8.4	8.7	12.7	13.7	12.1	11.8	9.8	...	7.3
1960_4_mean	3.6	4.2	7.4	8.2	6.0	9.2	6.1	9.3	12.8	10.7	...	13.6
1960_5_mean	13.5	17.0	21.5	17.6	15.6	13.4	14.8	14.5	12.1	12.3	...	15.1



**THANK YOU**