

한국은행 금요강좌  
제888회(2022. 6. 10(금))

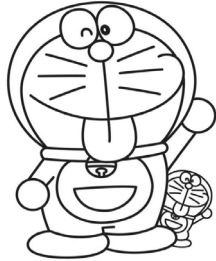
# 효과적인 금융경제 데이터 분석방법

BIS 통화경제부 권병천 박사

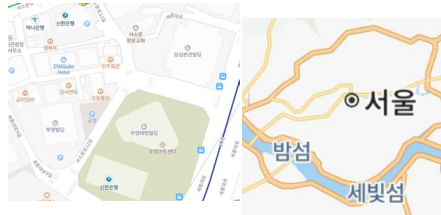


다양한 금융경제 데이터 분석 업무 경험 공유

강의목표



Big data? Small data??



Micro data? Macro data??

금융경제 데이터 구분

### 시계열 데이터

(예: GDP, CPI)



### 패널 데이터

(예: 서베이)

	순자산 (A-B)	자산 (A)
2019년 <sup>1)</sup>	35,281	43,191
2020년 <sup>1)</sup>	36,287	44,543
증감률	2.9	3.1

### 텍스트 데이터

(예: 뉴스, SNS)

The phenomenal growth of cryptocurrenc financial system. What role are traditional f what drives their engagement? Are new no a novel global supervisory database of bar complementary data sources for other type

### 이미지 데이터

(예: 사진, 그림)



### 스트리밍 데이터

(예: 음성, 영상)



```
[[[0. 0. 0. 0.]
[0. 1. 1. 0.]
[0. 1. 1. 0.]
[0. 0. 0. 0.]]
```

금융경제 데이터 구분

### 시계열 (Time Series)

Date	Observation	
2022-06-10	D	100 Index
2022-06	M	100 level
2022-06	Q	300
2022	A	100
2022	A	1200

주기  
(frequency)

단위  
(unit)

### 패널 (Panel)

Date	Column-1	Column-2	...	Column-n
2020		410		100
2021	BIS	756		99
2022		840		90
2020	Bank of Korea	410		90
2021		756		99
2022		840		100

Country Name	Country Codes	
	ISO3	Code
Afghanistan	AFG	004
Albania	ALB	008

시계열, 패널 데이터



<https://fred.stlouisfed.org/series/GDP>

Your trusted data source since 1991.

Categories > National Accounts > National Income & Product Accounts > GDP/GNP

☆ Gross Domestic Product (GDP)

Observation: Q1 2022: **24,382.683** (+ more)  
Updated: Apr 28, 2022

Units: Billions of Dollars, Seasonally Adjusted Annual Rate

Frequency: Quarterly





<https://ecos.bok.or.kr>

18.1.4.2 국내총생산(GDP)

년 1950 - 2021 10 개(최근시점 기준)

항목선택 원자료 가로로보기 백만달러

통계표	항목명1	단위	변환
18.1.4.2 국내총생산(GDP)	한국	백만달러	원자료



시계열 데이터 예시

<https://kosis.kr>

1) 가구특성별 소득원천별 가구소득 「가계금융복지조사」, 통계청, 한국은행, 금융감독원 (자료)

수록기간: 년 2017 ~ 2021 / 자료경신일: 2021-12-16 / 주석정보

시점 중감(중감률) 행렬전환 열고정

새 탭 열기 화면복사

(단위: 만원)

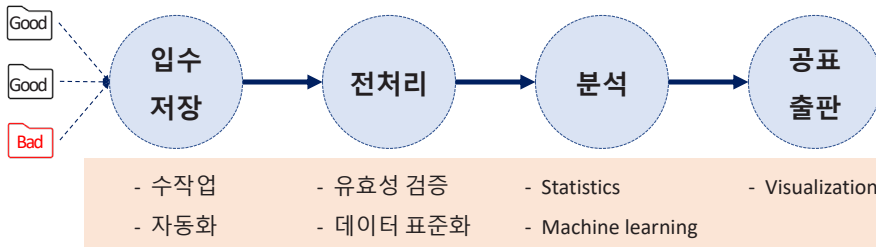
시점	가구특성별	원천별	2) 가구소득(전년도) 평균 (만원)	3) 가구소득(전년도) 중앙값 (만원)
2017	(가구주 성) 남자	경상소득	6,219	5,032
	(가구주 성) 여자	경상소득	2,923	1,891
2018	(가구주 성) 남자	경상소득	6,479	5,184
	(가구주 성) 여자	경상소득	2,986	1,965

<https://stats.bis.org>

**Dataset:** Locational Banking Statistics - disseminated data  
**Retrieved on:** Wed, Apr 27, 2022 11:58 AM GMT  
**Subject:** BIS locational banking  
**Frequency:** Quarterly  
**Type of instruments:** All instruments  
**Currency denomination:** All currencies  
**Currency type of reporting country:** All currencies (=D+F+U)  
**Parent country:** Germany  
**Type:** Repo  
**Coun:** Coun  
**Posit:** Posit

Balance sheet position		Total claims			
Counterparty sector		All sectors	Banks, related offices	Banks, total	Non-banks, total
TIME_PERIOD	Measure				
2020-Q1	Amounts outstandin...	€ 2,535,426.7	€ 833,591.9	€ 1,431,779.7	€ 1,099,527.4
	FX and break adjuste...	157,348.4	79,746.0	103,345.5	53,034.9
2020-Q2	Amounts outstandin...	€ 2,572,261.4	€ 844,397.5	€ 1,448,217.3	€ 1,120,252.3
	FX and break adjuste...	-33,472.7	-564.8	-13,441.4	-19,665.7

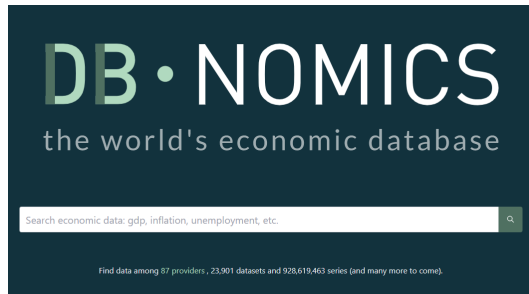
패널 데이터 예시



금융경제 데이터분석 프로세스

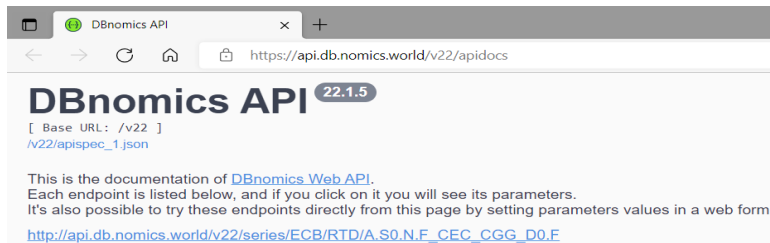
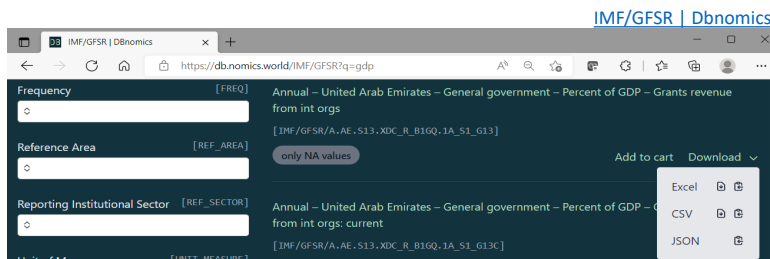
## DB·NOMICS\*

- ▶ 프랑스 중앙은행에서 무료로 제공중인 금융경제 데이터베이스
- ▶ IMF, BIS 등 전 세계 87개(2022년 5월) 데이터베이스 취합 및 제공



\* <https://db.nomics.world>

실습 - 데이터 입수

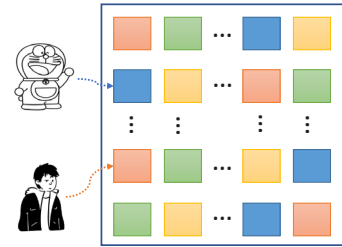
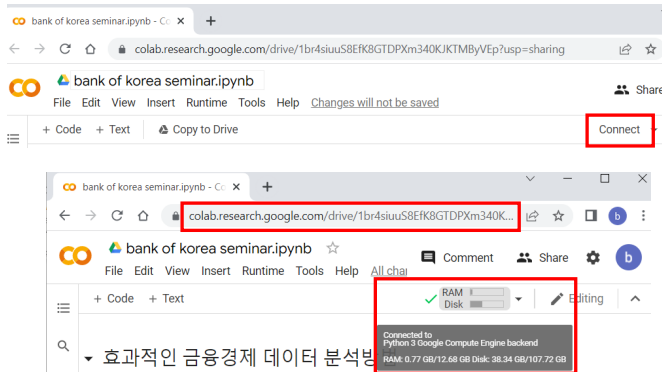


입수 프로세스

## Google Colaboratory 을 이용한 실습환경

- Google 에서 제공하는 무료 가상 컴퓨팅 환경

[Google Colab 링크 접속](#)

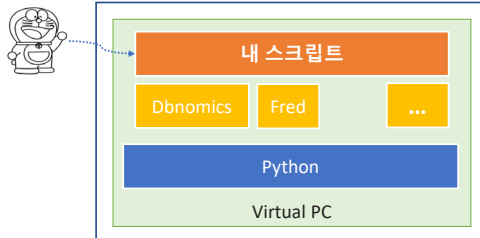


Google 클라우드

실습환경

## 실습 언어 - Python

- 1989년 범용 목적으로 개발된 프로그래밍 언어
- 무료로 배포되고 Python 개발 재단을 통해 운영관리



Google Colab

실습환경

실습 영상 - 엑셀 수작업

수작업

실습 영상 - 자동화

자동화작업

## 전처리 - 유효성 검증



데이터는 메타데이터 (주기 frequency, 단위 unit) 를 따르는가?

→ 주기, 단위 e.g. unit: 2020년=100

결측치(Missing) 데이터가 있는가?

→ 어떻게 채워 넣을것인가? (e.g. forward filling)

→ 결측치 제거 전략

데이터는 유효한가?

→ 이상치 (outlier) 체크 (e.g. 표준편차\*3 < observation)

→ 컴포넌트 체크 - 시계열 A + 시계열 B = 시계열 C

전처리 프로세스

실습 영상 - 전처리

실습 - 전처리

## 분석



### Statistics

- 인과 추론 (Causal Inference)
- 수학적 모델링을 통한
  - 경제 시스템에 대한 이해
  - 가설 추론 및 검증

### Machine Learning

- 예측 (Prediction)
- 관측되지 않은 값에 대한
  - 가장 정확한 값을 계산
  - 인과적 이해는 충분조건

#### Statistics versus machine learning

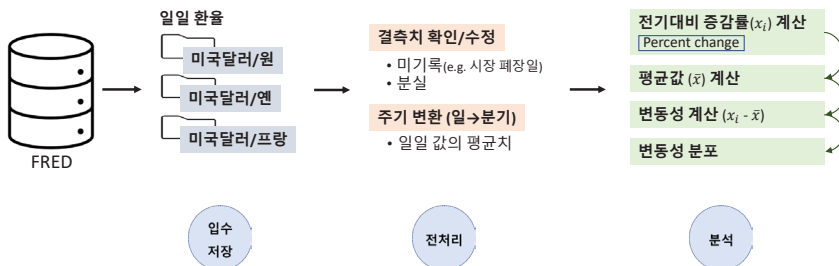
Daniilo Bzdok, Naomi Altman & Martin Krzywinski

*Nature Methods* 15, 233–234 (2018) | [Cite this article](#)

Bzdok, D., Altman, N. & Krzywinski, M. *Statistics versus machine learning*. *Nat Methods* 15, 233–234 (2018)

데이터 분석

## FRED\* 데이터베이스를 이용한 일일 환율데이터 분석



\*Federal Reserve Economic Data: 미 세인트루이스 연방준비은행이 제공하는 금융경제 데이터 서비스

종합 예제

실습 영상 - 종합예제

종합 예제

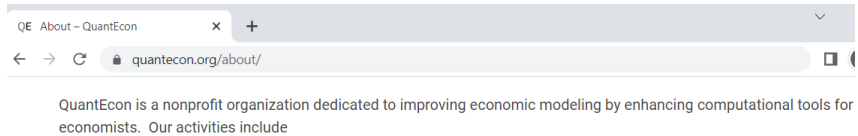
### 분석 전문가 (Data analyst/scientist) 가 되기 위한 필수 IT 지식



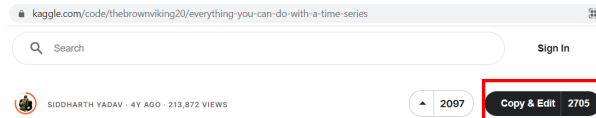
- 프로그래밍에 필요한 지식 습득은 최소화 → hands on example
- Interactive 프로그래밍에 익숙해져야 유리

## 참고할 만한 자료

- (한글 교재) Jump to Python – [링크](#) (기초 학습 4장까지)
- (영문) 금융경제를 위한 파이썬 프로그래밍 – [링크](#)



- (영문) 시계열 분석에 대한 모든 것 - [링크](#)



### Everything you can do with a time series

# 효과적인 금융경제 데이터 분석 방법

본 내용은 발표자 개인의 의견으로 국제결제은행의 공식 견해와는 무관합니다.

## 한국은행 소셜미디어 채널 오픈



[www.facebook.com/bankofkoreahub](http://www.facebook.com/bankofkoreahub)



[www.twitter.com/bok\\_hub](http://www.twitter.com/bok_hub)



[www.flickr.com/photos/bankofkorea](http://www.flickr.com/photos/bankofkorea)



[www.youtube.com/user/theBankofKoreakr](http://www.youtube.com/user/theBankofKoreakr)



[story.kakao.com/ch/bankofkorea](http://story.kakao.com/ch/bankofkorea)