



부동산 금융에 데이터 과학 적용하기

3주차. 데이터 분석 프로그래밍 언어 학습하기, 파이썬 패키지



1. 파이썬 이론, 실습
2. 국내 주식 데이터 수집하기
 - 3-1. 네이버 증권 - 국내증시 [최근 영업일 가져오기]
 - 3-2. 한국거래소(KRX 정보데이터시스템) 업종분류 현황 및 개별지표 크롤링
 - 3-3. 개별종목 지표 크롤링, 데이터 전처리하기
 - 3-4. FnGuide Index - 산업 섹터 정보 크롤링
 - 3-5. 네이버 증권 - 개별종목 수정주가 크롤링
 - 3-6. FnGuide - CompanyGuide - 재무제표 크롤링
 - 3-7. 가치지표(PER, PBR, PCR, PSR, DY) 계산

What to Cover

3

The screenshot shows the KRX Information System interface. On the left, a sidebar navigation includes '증권 시세' (Stock Prices), '증권 정보' (Stock Information), '거래 실적' (Trading Performance), '기타 증권' (Other Securities), '세부 안내' (Detailed Information), '증권 상품' (Securities Products), '채권' (Bonds), '파생 상품' (Derivatives), '일반 상품' (General Products), '해외 연계 시장' (Overseas Linked Markets), and '파생 결합 증권' (Derivative Composite Securities). A search bar at the top has the placeholder '검색어를 입력해주세요.' (Please enter your search term). The main content area displays a table titled '[12021] PER/PBR/배당수익률(개별종목)' (PER/PBR/Dividend Yield by Individual Stock) with columns for 종목코드 (Stock Code), 종목명 (Stock Name), 종가 (Closing Price), 대비 (Change), 등락률 (Return Rate), EPS (EPS), PER (P/E Ratio), 선행 EPS (Prior EPS), 선행 PER (Prior P/E Ratio), BPS (Beta), PBR (P/B Ratio), and 주당 배당금 (Dividend per Share). The table lists various stocks like GS, AJ 네트웍스, AK 투링스, etc. To the right, a developer tools Network tab is open, showing a list of requests with details like name, headers, payload, and response.

<http://data.krx.co.kr/contents/MDC/MDI mdiLoader/index.cmd?menuId=MDC0201020502>

수동적인 학습이 아닌 자신만의 OUTPUT을 생산하기 위한 적극적인 학습



배움보다 중요한 것은 야망이다 (최진석 교수)



STUDIAN
1.74M subscribers

Subscribed

1.4K



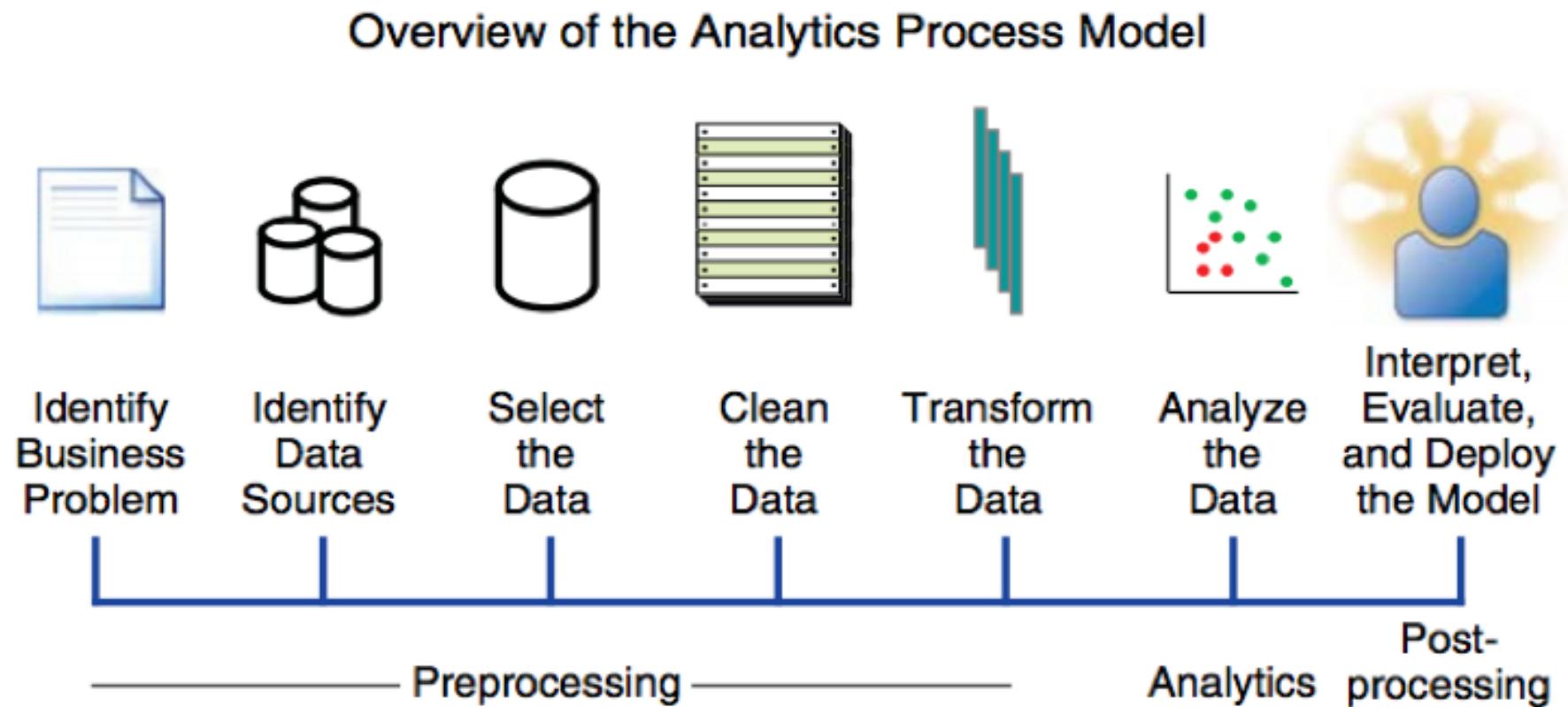
Share

Download

Clip

...

<https://www.youtube.com/watch?v=SSOs-FEOHRU>



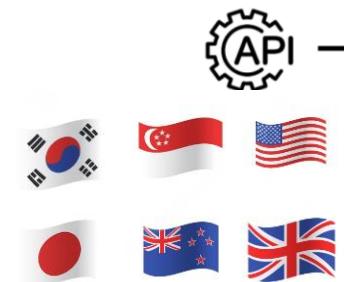
<https://blogs.sas.com/content/sgf/2019/05/14/big-data-in-business-analytics-talking-about-the-analytics-process-model/>

Real Estate Securities Data Analytics

6

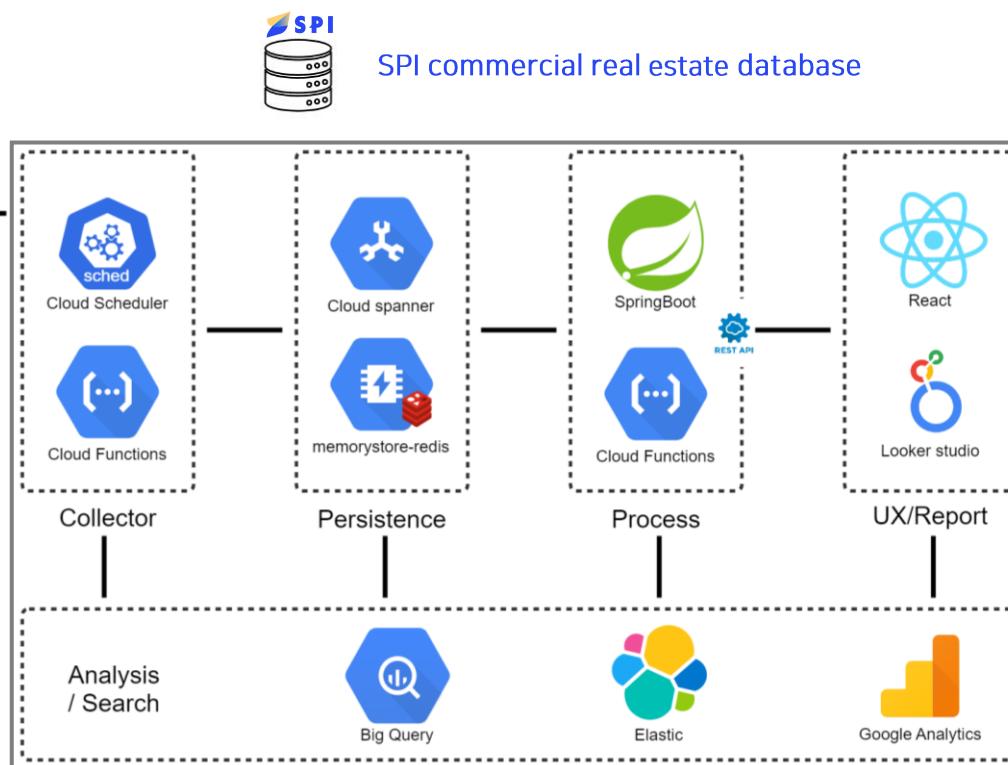
Data architecture specially designed for commercial real estate beneficiary certificates, providing data visualization and quantitative analysis

stocks, macro-economic, spatial,
or other data from key real
estate financial markets
are connected through API



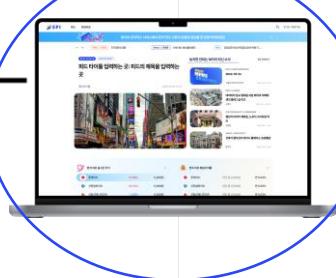
DART
SElbro
SGIS PLUS
Map SGX
ACR
Geospatial.SG

Nasdaq
EDGAR
Refinitiv
ICE
Google Map
Cherre



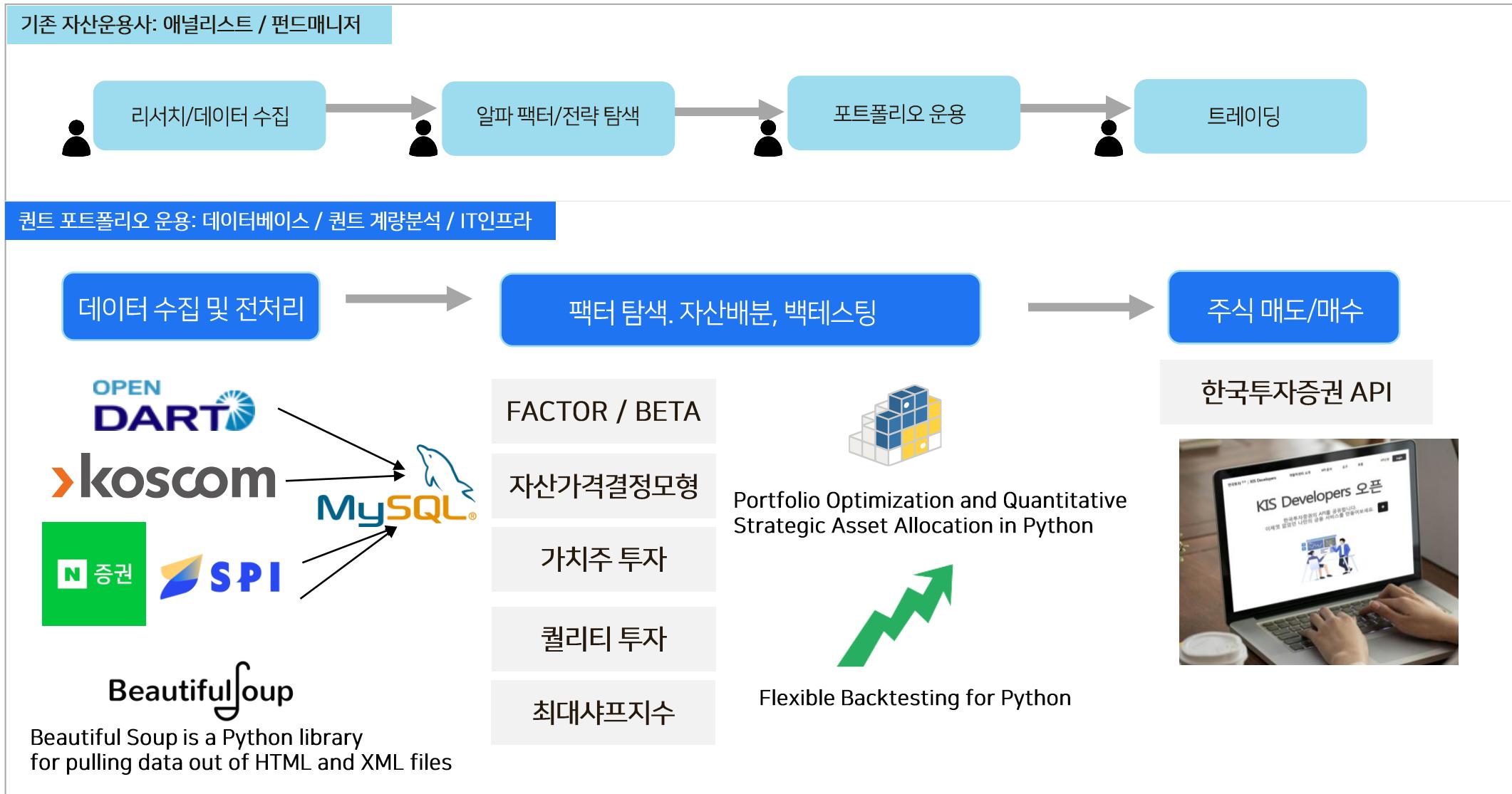
SPI commercial real estate database

Data visualization and
quantitative analysis



- ✓ Data analytics for publicly traded REITs
- ✓ Data visualization of private equity funds
- ✓ Information on properties on sale from institutional investors

- 기존 자산운용[데이터 수집 – 알파 팩터/전략 설정/포트폴리오 운용- 트레이딩]과정을 데이터베이스 및 IT인프라를 기초로 고도화 및 효율화



Robo Advisor

TEST BED CENTER

참여신청 | 운용정보 | 심사결과 | 정보안내 | 로그인 | 회원가입

ROBO ADVISOR TESTBED

맞춤형 자산관리서비스의 대중화를 위한 로보어드바이저의 신뢰성/안정성 확인

참여신청하기 → | 심사진행현황 → | 운용RA 정보 →

NOTICE

28 2023.04.	개인정보 처리방침 변경 안내
07 2023.03.	RA테스트베드 심사 업무 수행을 ...
06 2023.03.	주문 수량 계산 공개 로직 - 파이...

/main/main.do

... 자주 이용하시는 질문에 대한 답변을 확인

Daily Portfolio Report

투자유형별 지수

투자유형	지수	전일대비
안정주구형	1,113.25	▲ 1.11
위험증립형	1,197.86	▲ 0.36
적극투자형	1,361.04	▼ 1.11
종 합	1,222.17	▲ 0.17

Daily Hot Algorithm Provider

HOT 알고리즘

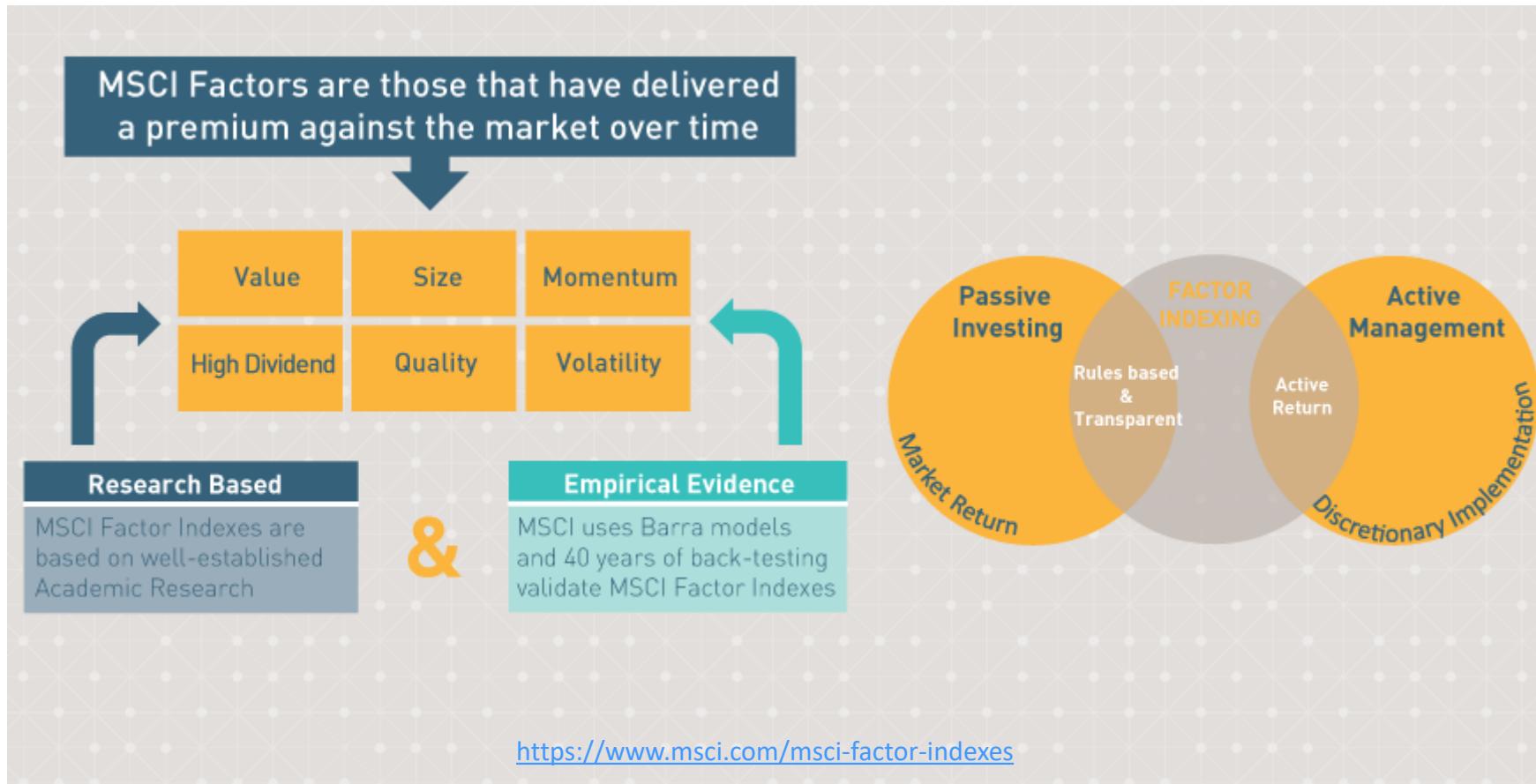
메타로고스... KB증권... shkim... 뉴지스탁...
 큐티백자산... 미래에셋대우... KB증권... 알파로직...
 이베스트투... 신한예아이... 대신증권...
 키움증권... NH투자증권... NH투자증권...
 에프앤가이... 고부저널...
 ...

<https://www.ratetestbed.kr:7443/portal/main/contents.do?menuNo=200070>

- 퀀트 투자는 크게 '[포트폴리오 운용](#)' 전략과 '[트레이딩 전략](#)'으로 나뉨
- [포트폴리오 운용](#) 전략은 과거 주식 시장을 분석해 좋은 주식의 기준을 찾아낸 후 해당 기준에 만족하는 종목을 매수한다. 반대에 있는 주식은 공매도하기도 한다. 다수의 종목을 하나의 포트폴리오로 구성해 운용한다. [재무제표 분석, 증권 분석, 거시경제 분석]
- [트레이딩](#) 전략은 주식이 오르거나 내리는 움직임을 연구한 후 각종 지표를 이용해 매수 또는 매도한다. 투자 속도가 빠르다. Buy and Sell Quickly [기술적 지표(모멘텀 등), 차트 기반, 데이 트레이딩]

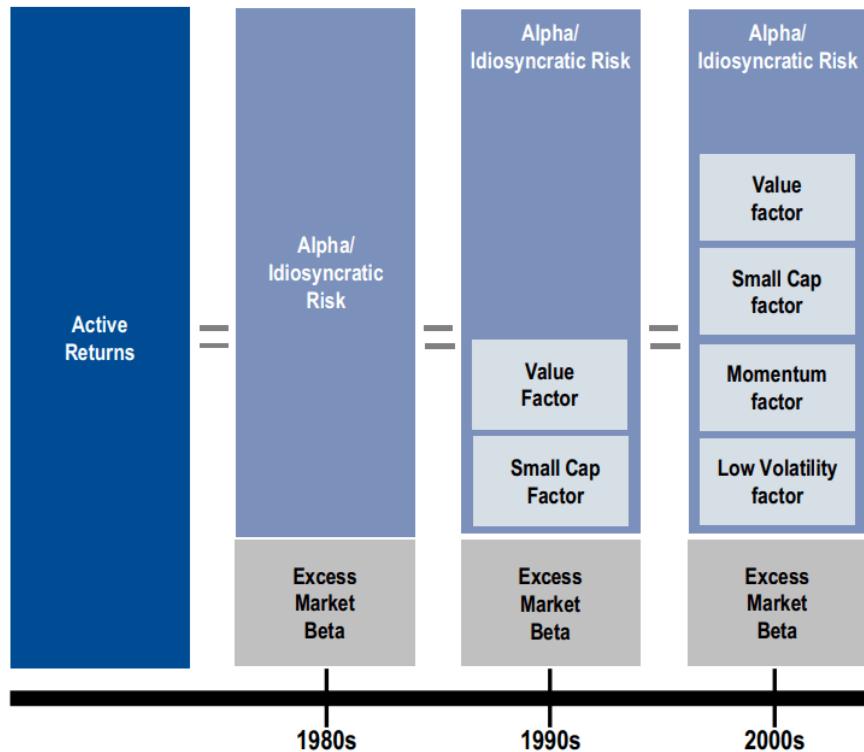
팩터(Factor)

주식 수익률에 영향을 미치는 특성들을 팩터(factor)라고 한다.



The MSCI factor indexes are **rules-based indexes** that capture the returns of systematic factors that have historically earned a persistent premium over long periods of time—such as Value, Low Size, Low Volatility, High Yield, Quality and Momentum and Growth.

그림3. 90년대 이후 알파(Alpha)를 다양한 팩터(Factor) 프리미엄으로 설명하는 추세



주식 투자 스타일을 크게 나누자면 Active 와 Passive 투자로 나눌 수 있습니다. Active 투자 전략은 본인이 적극적으로 종목을 발굴하여 주가 지수보다 초과 수익을 낼 수 있다고 믿는 사람들이 사용하고, Passive 투자 전략은 ETF 등을 활용하여 주가 지수를 추종하는 방식

스마트베타 전략은 단순히 지수를 추종하는 것에서 나아가 밸류, 모멘텀, 퀄리티로우볼 등의 팩터를 정량적으로 구성하고 지수화하여 투자하는 전략. Passive를 추구하면서 동시에 Active 적인 요소를 가미하여 초과 수익을 추구

자료: Allianz Global Investors, MSCI

팩터	설명
Value	이익(Earnings), 장부가치(Book Value), 현금흐름(Cash Flow), 매출(Sales) 등 펀더멘탈 대비 주가가 낮은 종목에 투자
Low Volatility	주가 변동성이 낮은 종목에 투자
Momentum	최근 주가 모멘텀이 큰 종목에 투자
Quality	ROA, ROE, 부채비율 등의 지표를 이용해 이익률, 안정성 등이 높은 종목에 투자
Dividend	배당주(배당성장, 고배당)에 투자
Size	시가총액 가중지수 대비 중, 소형주 비중을 높이는 전략

자료: NH 투자증권 리서치본부

시장을 이기는 법 : 팩터를 활용한 스마트베타 투자 전략 | 러닝스푼즈 (learningspoons.com)

데이터 분석	목적	대응
데이터 시각화	데이터가 어떻게 생겼는지 알고 싶다	<ul style="list-style-type: none"> (1) 전체 데이터를 한눈에 확인 <ul style="list-style-type: none"> - 점그림, 선그림, 영역그림 - 막대그림, 등고선그림, 분포그림(히스토그램) → 통계 사용 (2) 데이터를 뿌리고 통계를 계산하기 위해 데이터 값 하나하나를 표현 <ul style="list-style-type: none"> → 확률/컴퓨터 사용
기술적 분석	데이터가 어떻게 생겼는지 알고 싶다	<ul style="list-style-type: none"> (1) 전체 데이터 특성을 몇 개의 숫자들로 확인 <ul style="list-style-type: none"> → 통계 사용 (2) 통계를 계산하기 위해 데이터 값 하나하나를 표현 <ul style="list-style-type: none"> → 확률/컴퓨터 사용
상관관계/인과관계	여러종류 데이터끼리의 관계를 알고 싶다	<ul style="list-style-type: none"> (1) 각 데이터를 몇 개의 숫자들로 표현 <ul style="list-style-type: none"> → 통계 사용 (2) 표현된 숫자들을 비교 <ul style="list-style-type: none"> → 확률/통계 사용
통계추론	일부 데이터로 전체를 알고 싶다	<ul style="list-style-type: none"> (1) 일부 데이터의 특성을 확인 <ul style="list-style-type: none"> → 통계 사용 (2) 반복적으로 실험 진행 및 통계치 재확인 <ul style="list-style-type: none"> → 컴퓨터 사용 (3) 전체 특성을 추론 <ul style="list-style-type: none"> → 통계 사용
알고리즘학습	전체 데이터로 미래를 알고 싶다	<ul style="list-style-type: none"> (1) 데이터의 관계를 수학적으로 표현 <ul style="list-style-type: none"> → 확률/통계/함수/컴퓨터 사용 (2) 미래를 예측한 후 정확성 확인 <ul style="list-style-type: none"> → 확률/통계/컴퓨터 사용
가설검정(A/B Test)	뭔가 진실과 가까운 의사결정을 하고 싶다	<ul style="list-style-type: none"> (1) 기존 데이터와 새로운 데이터 비교를 위해 숫자들로 표현 <ul style="list-style-type: none"> → 통계 사용 (2) 표현된 숫자들을 비교 <ul style="list-style-type: none"> → 확률/통계/컴퓨터 사용

글로벌 상장 리츠 투자 서비스

상장 리츠 데이터 분석력을 기초로 글로벌 리츠 투자 서비스 [Active Index REIT ETF] 출시, 시가총액가중지수(capitalization-weighted index)를 단순 추종하는 Passive REIT 시장과 정성적 리서치 기반의 Active REIT 시장을 혁신

Active Index REIT ETF

[통계/머신러닝 분석방법론 기반 factor Index 설계]

Passive Index REIT ETF

[지수 단순 추종]



Vanguard®

Vanguard REAL ESTATE ETF

TIGER ETF

미래에셋 TIGER 미국MSCI리츠 ETF

미래에셋 TIGER 리츠부동산인프라 ETF

ACE ETF

한국투자신탁 ACE 싱가포르리츠 ETF

SPI

QRAFT
AI ETFs

Active REIT Fund

[펀드매니저 리서치 기반]



FRANKLIN
TEMPLETON



Fidelity
INVESTMENTS

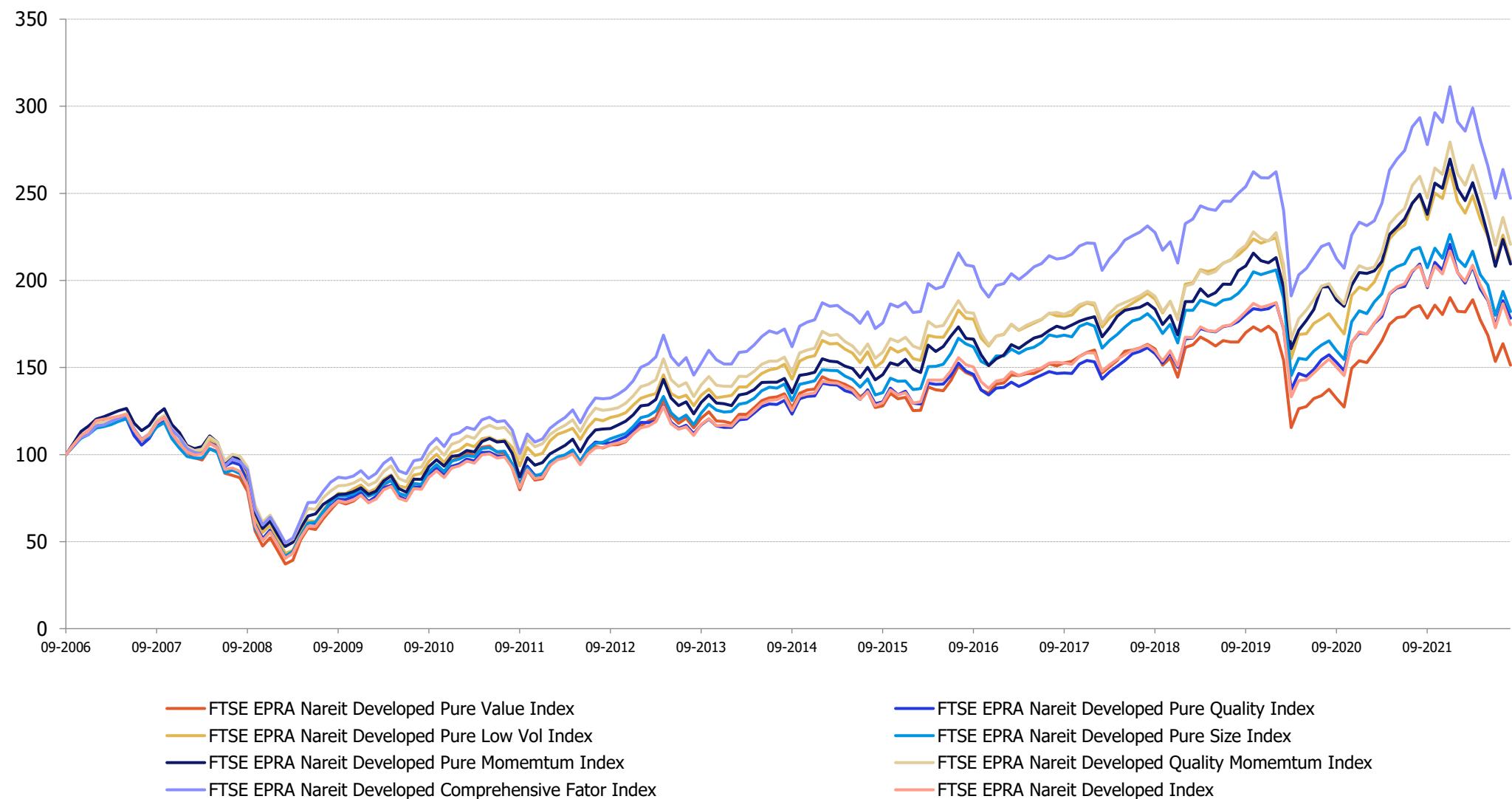
Mastern

INVESTMENT MANAGEMENT

이지스자산운용
IGIS Asset Management

글로벌 상장 리츠 투자 서비스

Active REIT Index - 통계/머신러닝 분석방법론 기반 factor Index 설계



출처: LSE FTSE INDEX

Value(가치주) Factor [PER, PBR] / Quality(우량주) Factor [지속적인 실적 성장, 자기자본 이익률]

- Warren Buffett, Benjamin Graham

정의 12 : FFO

$FFO = \text{지배순이익} - \text{우선주현금배당금총액} + \text{투자부동산조정} + FVPL \text{ 적용투자증권조정}$
+ 관계기업공동기업투자조정 + 지분법적용투자증권조정 + 종속기업투자증권조정 + FX조정

account_nm	si_div	account_id
지배순이익	CIS	ifrs-full_ProfitLossAttributableToOwnersOfParent
우선주현금배당금총액	-	-
투자부동산조정		
FVPL 적용투자증권조정		
관계기업공동기업투자조정		
지분법적용투자증권조정	하단참조	
종속기업투자증권조정		
FX 조정		

FFO, Fiscal Quarters	
구분	
지배순이익	+
우선주현금배당금총액	-
투자부동산감가상각비	+
투자부동산평가손실	+
투자부동산평가이익	-
투자부동산처분손실	+
투자부동산처분이익	-
투자부동산손상차손	+
투자부동산손상차손환입	-
투자부동산 기타조정	+
당기손익인식금융자산평가손실	+
당기손익인식금융자산평가이익	-
당기손익인식금융자산처분손실	+
당기손익인식금융자산처분이익	-
FVPL 적용투자증권 기타조정	+
관계기업/공동기업투자처분손실	+
관계기업/공동기업투자처분이익	-
관계기업/공동기업투자손상차손	+
관계기업/공동기업투자손상차손환입	-
관계기업공동기업투자 기타조정	+
지분법손실	+
지분법이익	-
지분법적용투자주식배당금	+
지분법투자주식처분손실	+
지분법투자주식처분이익	-
지분법적용주식 기타조정	+
종속기업투자주식처분손실	+
종속기업투자주식처분이익	-
종속기업투자주식 기타조정	+
외화환산손실	+
외화환산이익	-
외환차손	+
외환차익	-
FX 기타조정	+

NAV, Fiscal Quarters	
구분	
지배기업의 소유주에게 귀속되는 자본	+
우선주자본금	-
우선주식발행초과금	-
투자부동산	-
투자부동산공정가치	+
기타조정	+

정의 8: NAV

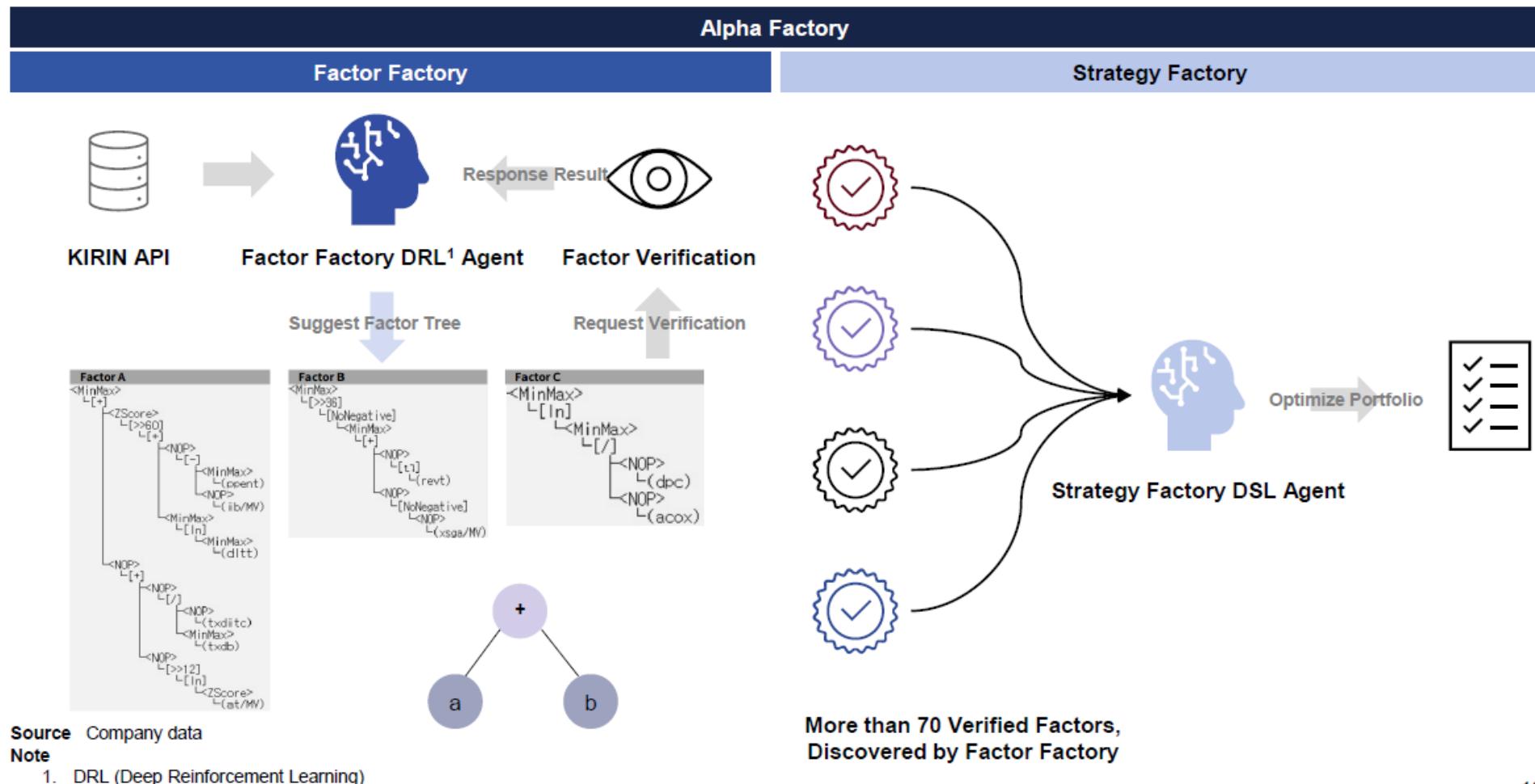
$NAV = \text{지배기업의 소유주에게 귀속되는 자본} - \text{우선주자본금} - \text{우선주식발행초과금} - \text{투자부동산}$
+ 투자부동산공정가치 ± 기타조정

account_nm	si_div	account_id
지배기업의 소유주에게 귀속되는 자본	BS	ifrs-full_EquityAttributableToOwnersOfParent
우선주자본금	BS	dart_IssuedCapitalOfPreferredStock
우선주식발행초과금	-	-
투자부동산	BS	ifrs-full_InvestmentProperty
투자부동산공정가치	-	-
기타조정	-	-

Value(가치주) Factor [NAV, FFO] / Quality(우량주) Factor [FFO growth]

크래프트 태크놀로지스(Qraft Technology)

Alpha Factory = [Factor Factory + Strategy Factory] Factor Factory는 랜덤 서치, 강화학습 알고리즘을 통해서 알파가 될 팩터를 자동으로 찾는다. Strategy Factory는 Factor Factory로 발견한 알파 팩터들을 기초로 최적의 포트폴리오를 구성한다.



변동성(위험지표)를 낮추면서 수익률을 높이는 알고리즘을 개발



그룹 나누기 그룹 모으기 전체그룹 R* 그룹 R+ 그룹 RI 그룹 참고지표

운용기준일 : 2023년 05월 18일

- R* 그룹 (테스트베드를 통과하여 상용서비스가 가능한 알고리즘으로 95개의 포트폴리오가 공시되고 있습니다.)

용어 설명

알고리즘명 [계좌별명]	업체명	수익률(%)									위험지표				공시시작일	그래프 보기
		1주	1개월	3개월	6개월	1년	2년	3년	연환산	누적	표준 편차	샤프 지수	최대 손실률	보상 비율		
R* 관텍 국내주식형 대형3호 [적극투자형]	관텍투자일임 주식	2.81 2,695,923	-10.59 -3.94	-2.39 -15.07	-19.93 20.78	2.83 13.28	0.25 0.05	-36.92 0.29	0.29 3.5년	2019-11-04						
R* 관텍 해외주식형 대형5호 [적극투자형]	관텍투자일임 주식	-1.49 2,298,145	-5.36 -6.03	-4.40 -5.43	0.91 67.99	6.92 25.58	0.29 0.18	-48.16 0.48	-48.16 3.5년	2019-11-04						
R* 관텍 스테디 국내 EMP [스테디 국내 EMP 적극투자형3]	관텍투자일임	-0.15 6,800,000	0.39 2.41	3.21 0.12			-4.58 -6.35	0.09 -0.01	-12.87 -0.49	2022-02-07 1.3년						
R* 퀀터백 글로벌자산배분 국 [퀀터백자산운용주식회사]	퀀터백자산운용 특이	-0.04 5,000,000	-0.25 2.35	5.20 0.19	-6.47 -6.47	16.24 2.57	18.11 0.11	0.09 0.09	-23.35 0.63	2016-10-17 6.6년						
R* 파운트 블루웨일(ETF)1.0 [적극투자형]	파운트투자자문	0.46 5,000,000	1.64 3.04	-0.69 -1.85	0.33 0.33	7.62 0.08	-0.28 -0.28	0.14 0.14	-28.06 -0.01	2018-05-08 5.0년						
R* NH로보 ETF형 모멘텀 [NH로보 ETF형 모멘텀(공격형5)]	NH투자증권 특이	0.37 5,000,000	-0.69 3.23	2.04 -1.37	0.32 0.32	36.82 4.40	22.26 0.12	0.22 0.22	-31.83 0.66	2018-11-05 4.5년						

규칙 기반 프로그래밍

Rule-based Programming

vs

머신러닝

Machine Learning

금융 분야에서 머신러닝/딥러닝 적용은 아직 초기 단계

부동산 금융 분야 분석방법론은
학계에서는 AIRMA 시계열 분석 수준,
민간에서는 계산금융 수준에 머무름

가설: 부동산 금융 시계열의 특징이 어느 정도 유지되는 패턴을 보이고,
그 패턴으로부터 미래의 가격 움직임을 예측할 수 있다고 가정

계량경제학(econometrics)과 계산 금융(Computational Finance)의 발전:
회귀분석과 같은 **통계 분석방법론** 및 **프로그래밍 가능한 API**로 비약적 발전

규칙 기반 프로그래밍

Rule-based Programming

계량경제학(econometrics)

Econometrics is the use of statistical and mathematical models to develop theories or test existing hypotheses in economics and to forecast future trends from historical data. It subjects real-world data to statistical trials and then compares the results against the theory being tested.

from investopedia

계산 금융 (Computational Finance)

계산 금융 (Computational Finance) 계산금융이란, 금융시장과 거래되는 금융상품과 관련된 문제들을 수리적으로 모델화 하고 이를 효율적 계산 방법들을 이용하여 풀어내는 연구 분야이다. 이를 위하여, 다양한 최신 과학계산 기법들과 수학적, 통계적 모델 기법들을 연구하고 개발한다.

From 서울대학교 산업공학과 통계학습 및 계산금융 연구실

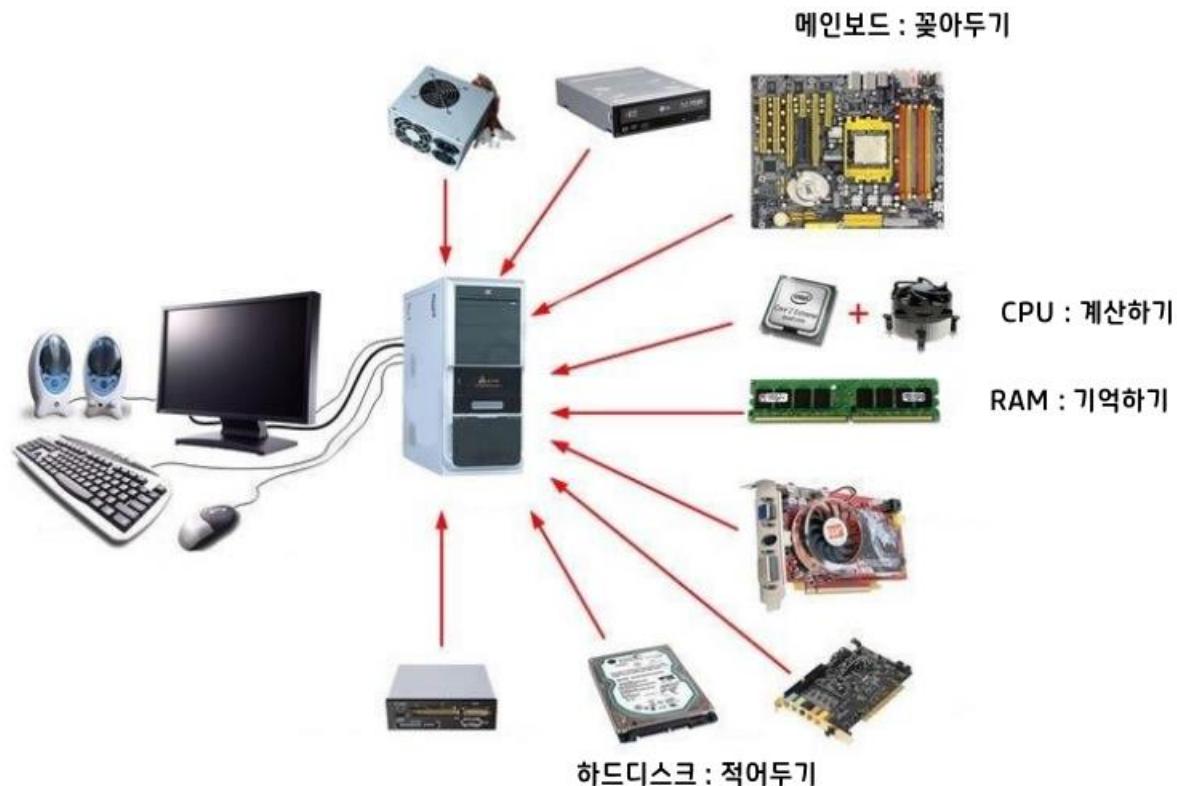
데이터 기반 금융(data-driven finance)

프로그래밍 가능한 API (Programmable API)

계량경제학(econometrics)

계산 금융(Computational Finance)

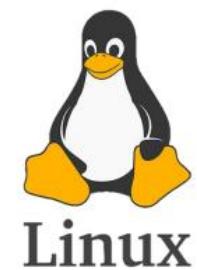
컴퓨터란 무엇인가?



컴퓨터는 기본적으로 기계를 모아둔 것

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

그걸 가능하게 하는 게 OS (운영체제)



즉, 인간이 기계와 소통할 수 있게 만드는 기반 프로그램
(휴대폰이라는 기계를, 컴퓨터라는 기계를, TV라는 기계를..)

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

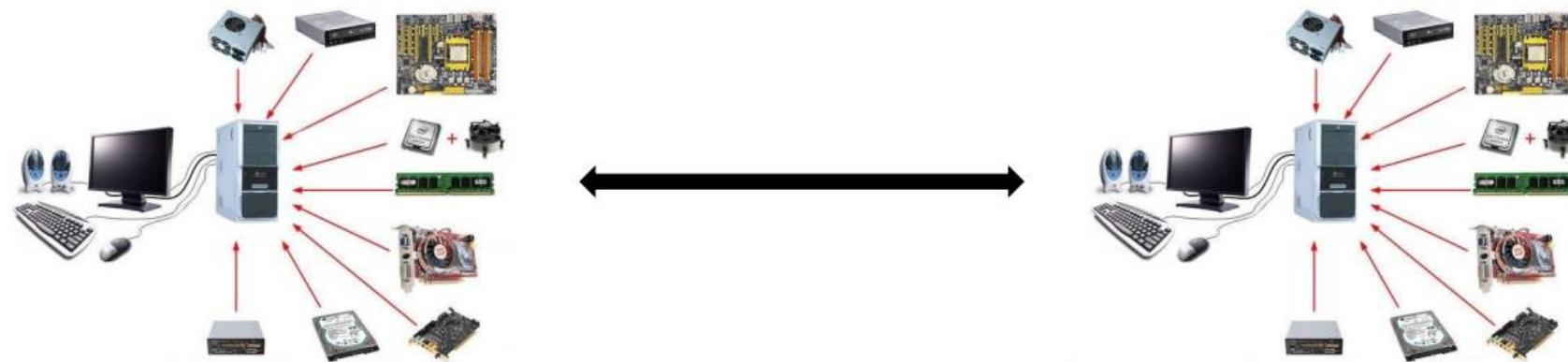
이걸 어떤 말로 하냐가



“프로그래밍 언어”

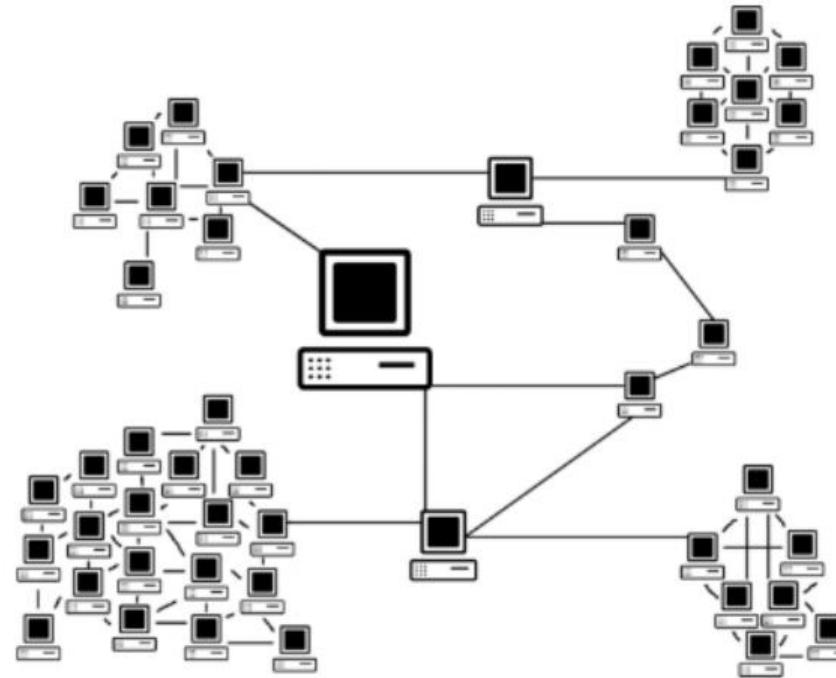
영어로 할거나, 이탈리아어로 할거나 차이
즉, 어떤 말로든 명령만 내리면 됨

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설



컴퓨터끼리 대화를 주고 받고 싶으면?

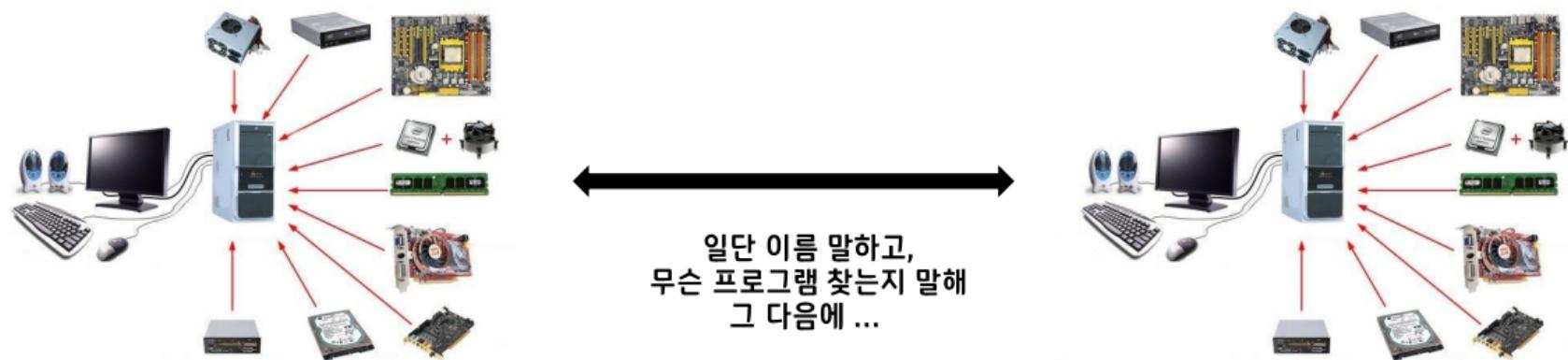
출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설



그게 바로 인터넷

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

근데 대화를 주고 받으려니 규칙이 필요하다



그러한 규칙 중 하나가 웹 = http

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

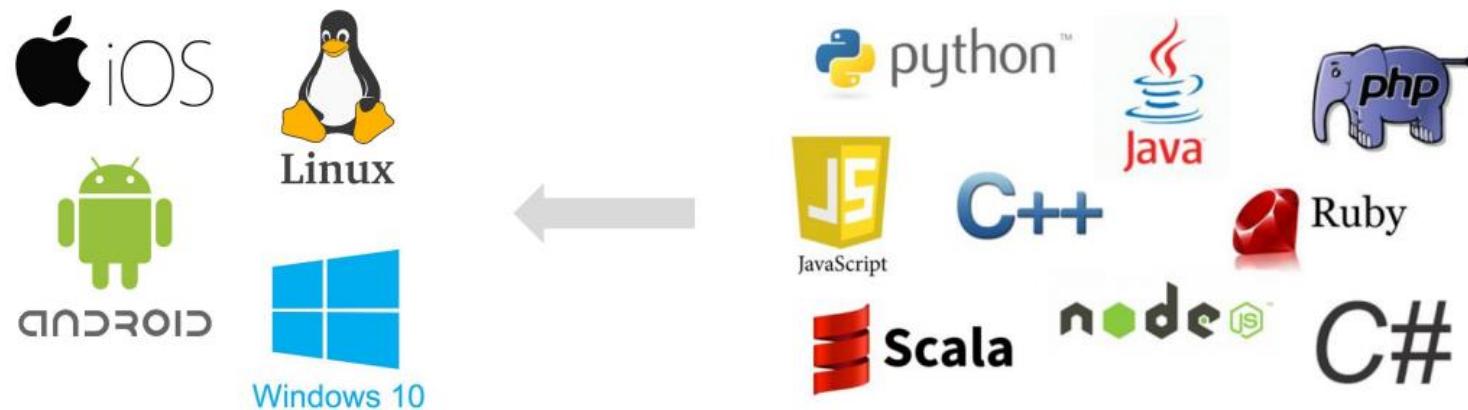
근데 대화를 주고 받으려니 규칙이 필요하다

<http://spartacodingclub.kr/>

웹 통신을 하겠다는 의미

“내가 지금부터 웹 규칙대로 얘기할 거니까 잘 들어~”

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설



어떤 명령이냐면.

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설



“잘 듣고 있다가 누가 http 요청 하면 이 파일을 돌려줘”

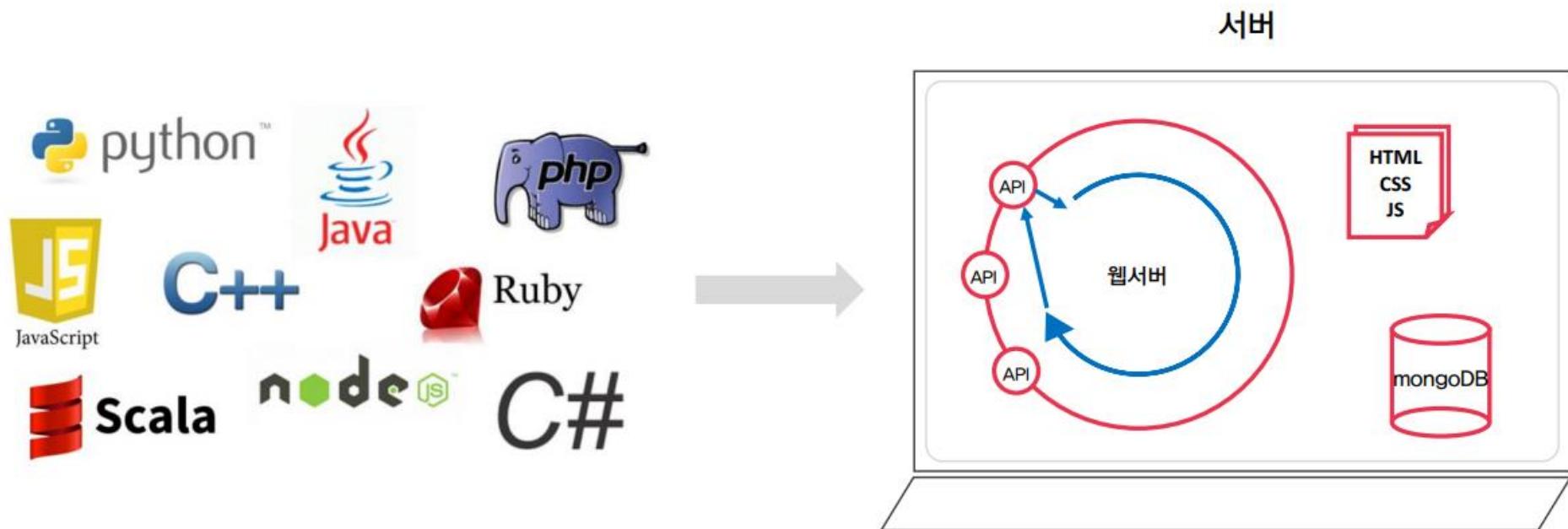
라고 명령을 한다면

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설



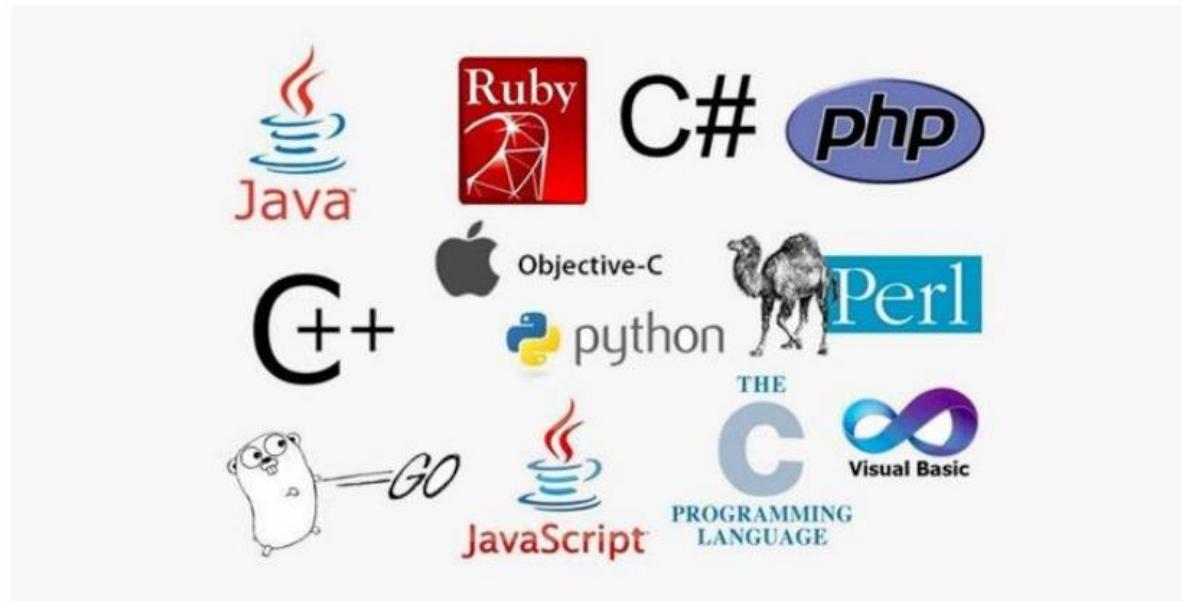
그게 바로 “**웹 서버**”를 만들었다고 하는 것

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설



출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

자, 그럼 언어 별로 용처와 특징을 알아보겠습니다.



대부분의 언어로 서버를 만들 수 있다고 봐도 무방
(HTML, CSS, JS를 제외하고)

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

자, 그럼 언어 별로 용처와 특징을 알아보겠습니다.



HTML 뼈대, CSS 꾸미기, Javascript 움직이게
이것만 갖고 만들면 프레임워크 없이 만드는 것(이건 별로 안 어려움)

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

자, 그럼 언어 별로 용처와 특징을 알아보겠습니다.

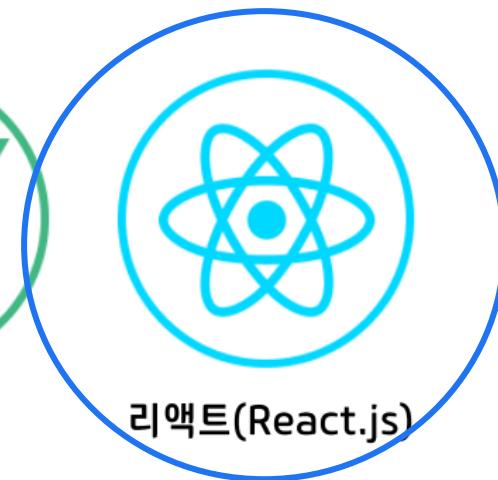
모두 다 Javascript 로 만든 프레임워크



앵귤러(Angular)



뷰(Vue.js)



리액트(React.js)

HTML, CSS, Javascript를 더 잘 쓸 수 있게 하는 프레임워크

체감상 리액트가 대세. 그런데 셋 중 하나 할 줄 알면 금방 배운다는 이야기가!

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

자, 그럼 언어 별로 용처와 특징을 알아보겠습니다.



어렵다. 형식이 많음
큰 회사에서 많이 씀



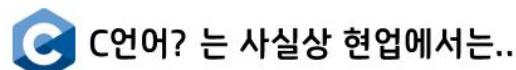
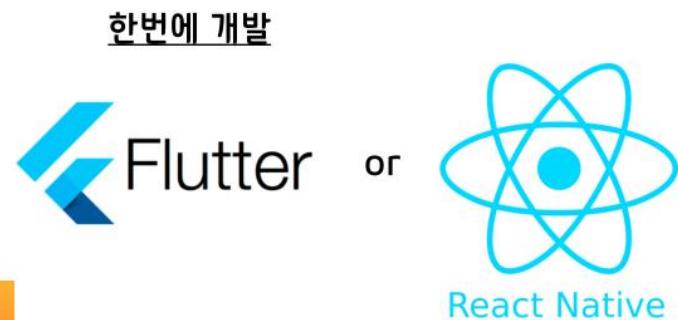
쉽게 입문. 자유로움
데이터분석에도 씀



Javascript와 거의 같음
스타트업에서 많이 쓴다
(백/프론트언어가 같으면 편하니까)

출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

그 외 - 앱을 만들때는



출처: 스파르타코딩클럽 - 1시간만에 끝내는 직장인 필수 코딩 용어 해설

WEB

HTML, CSS, Javascript

JS Framework

- React, Vue, Angular

SERVER

Java(Spring)

Python
(Django, Flask)

Node.js (Express)

C++

PHP

...

DATA ANALYTICS

Python

SQL

APP

안드로이드(Kotlin)

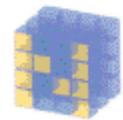
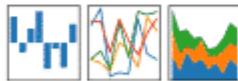
아이폰(Swift)

Flutter, ReactNative



Scientific Computing

pandas
 $y_t = \beta' x_{it} + \mu_i + \epsilon_{it}$



NumPy



SciPy

Data Access

Quandl

SQLAlchemy



AI & Machine Learning



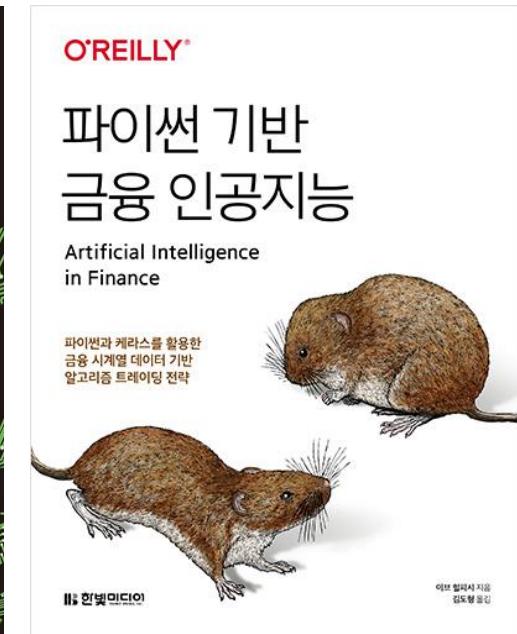
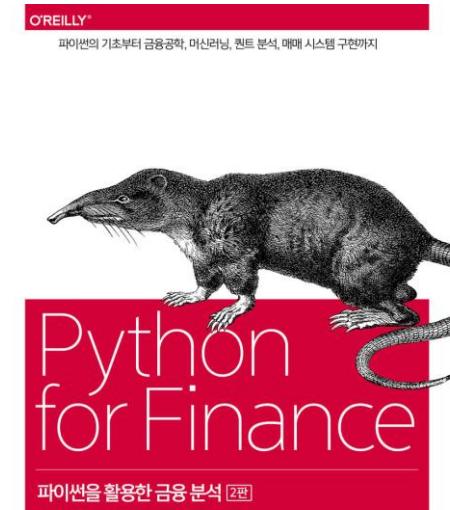
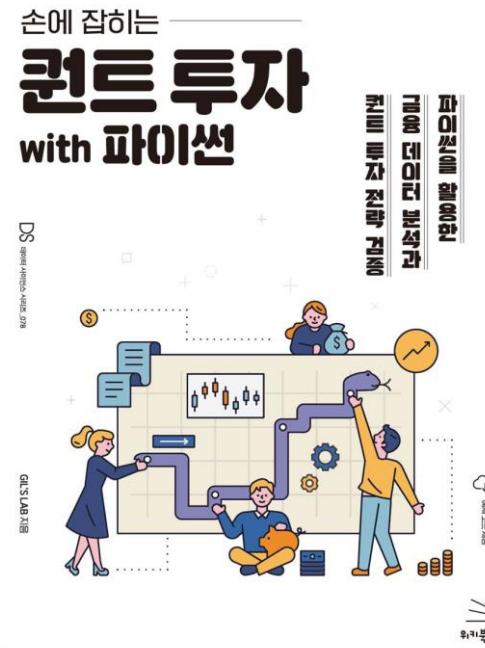
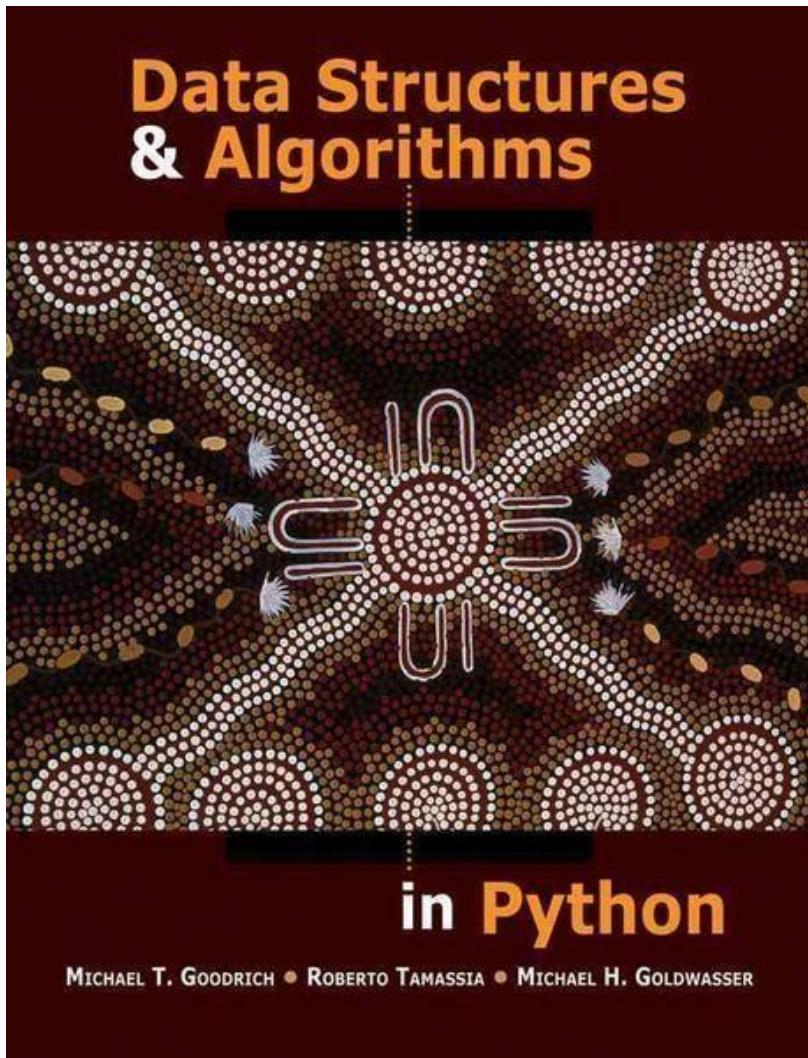
theano

TensorFlow

Keras

파이썬 추천 도서

점프투파이썬 <https://wikidocs.net/12>



Introduction to Data Analysis and Visualization



pandas

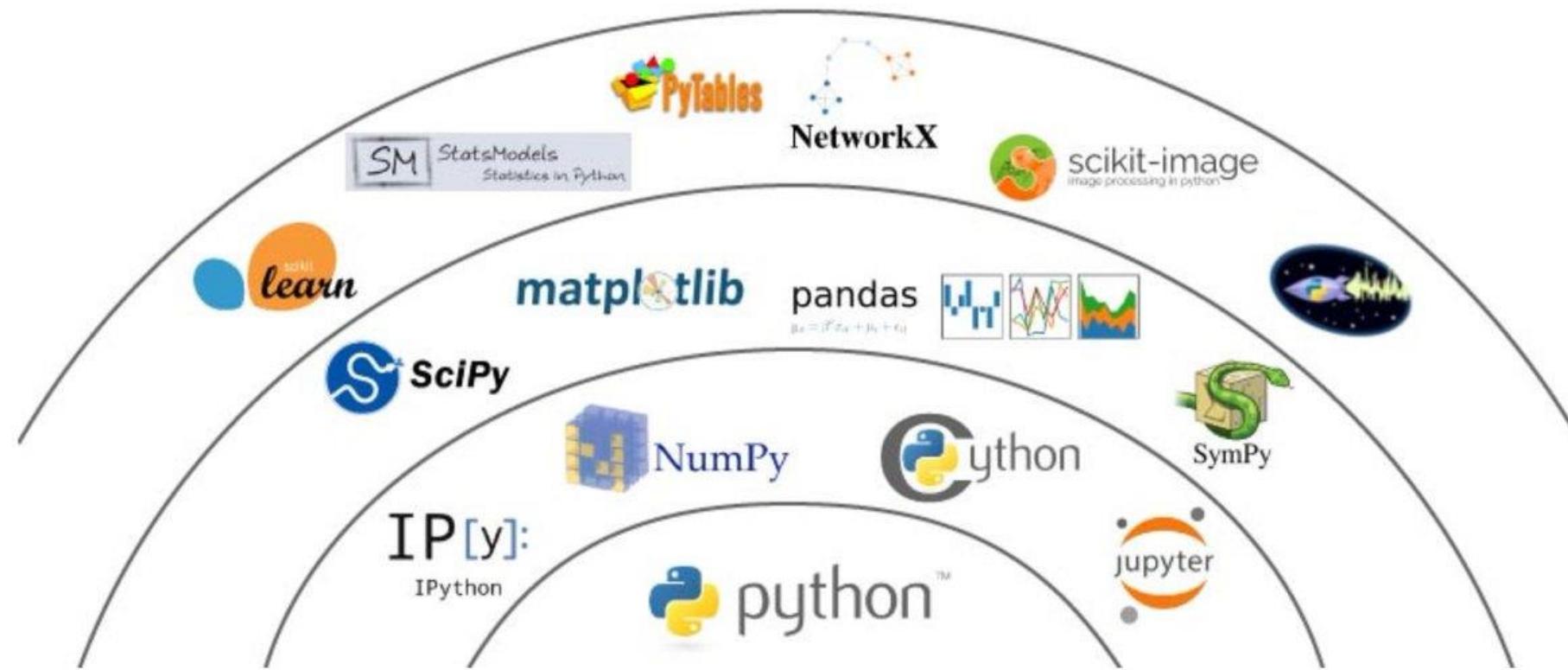
matplotlib



seaborn

<https://www.freecodecamp.org/news/exploratory-data-analysis-with-numpy-pandas-matplotlib-seaborn/>

Python Ecosystem





파이썬은 1991년 네덜란드계 소프트웨어 엔지니어인 귀도 반 로섬이 발표한 고급 프로그래밍 언어; 인터프리터식, 객체지향적, 동적 타이핑 대화형 언어이다.

<https://remarble.org/2023/03/18/python-package-for-data-science/>



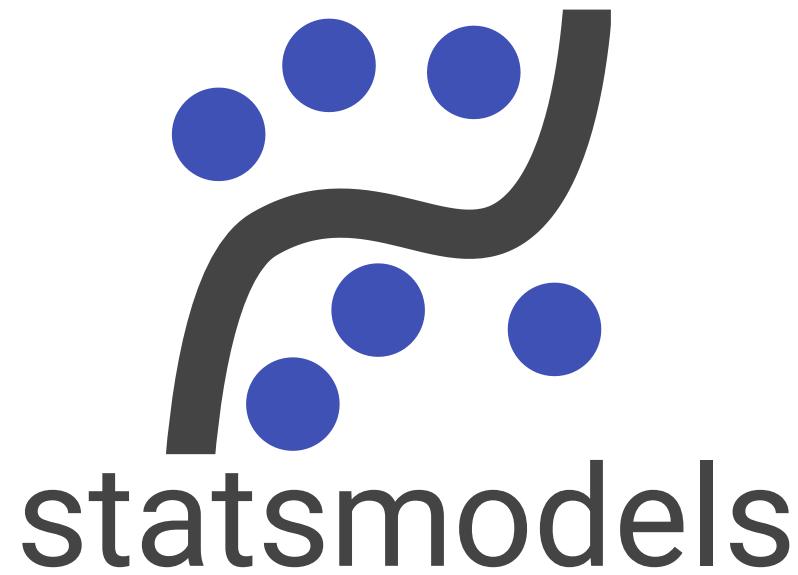
Numpy는 Numerical Python의 약자로 수치 해석 라이브러리. 특히 선형대수 연산에 필요한 다차원 배열과 배열 연산을 수행하는 다양한 함수를 수행한다.



판다스는 2008년 금융데이터 분석용으로 개발. 이후 오픈소스 커뮤니티를 통해 발전하고 있으며 통계, 데이터과학 분야에서 가장 중요한 소프트웨어로 성장. 데이터 수집 및 전처리에 최적화

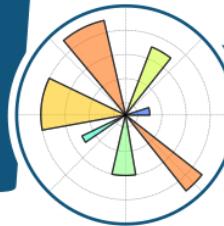


사이파이(SciPy)는 수학 연산에 필요한 다양한 패키지를 모아 놓은 라이브러리이다. 미적분, 선형대수, 행렬 연산, 방정식 계산 등에 필요한 함수를 지원한다.



다양한 통계, 회귀분석, 시계열 분석 모델을 지원한다.

matplotlib



그래프타 차트 등 그래픽으로 표현하는데 사용하는 파이썬 기반 2D 시각화 도구. 판다스와 연계하여 다양한 방식으로 데이터를 시각화할 수 있다. 데이터를 시각화하면 데이터의 구조와 패턴을 파악하기에 용이하며 이를 통해 인사이트를 발견할 수 있다.



scikit-learn은 머신러닝 학습을 위한 파이썬 라이브러리이다.

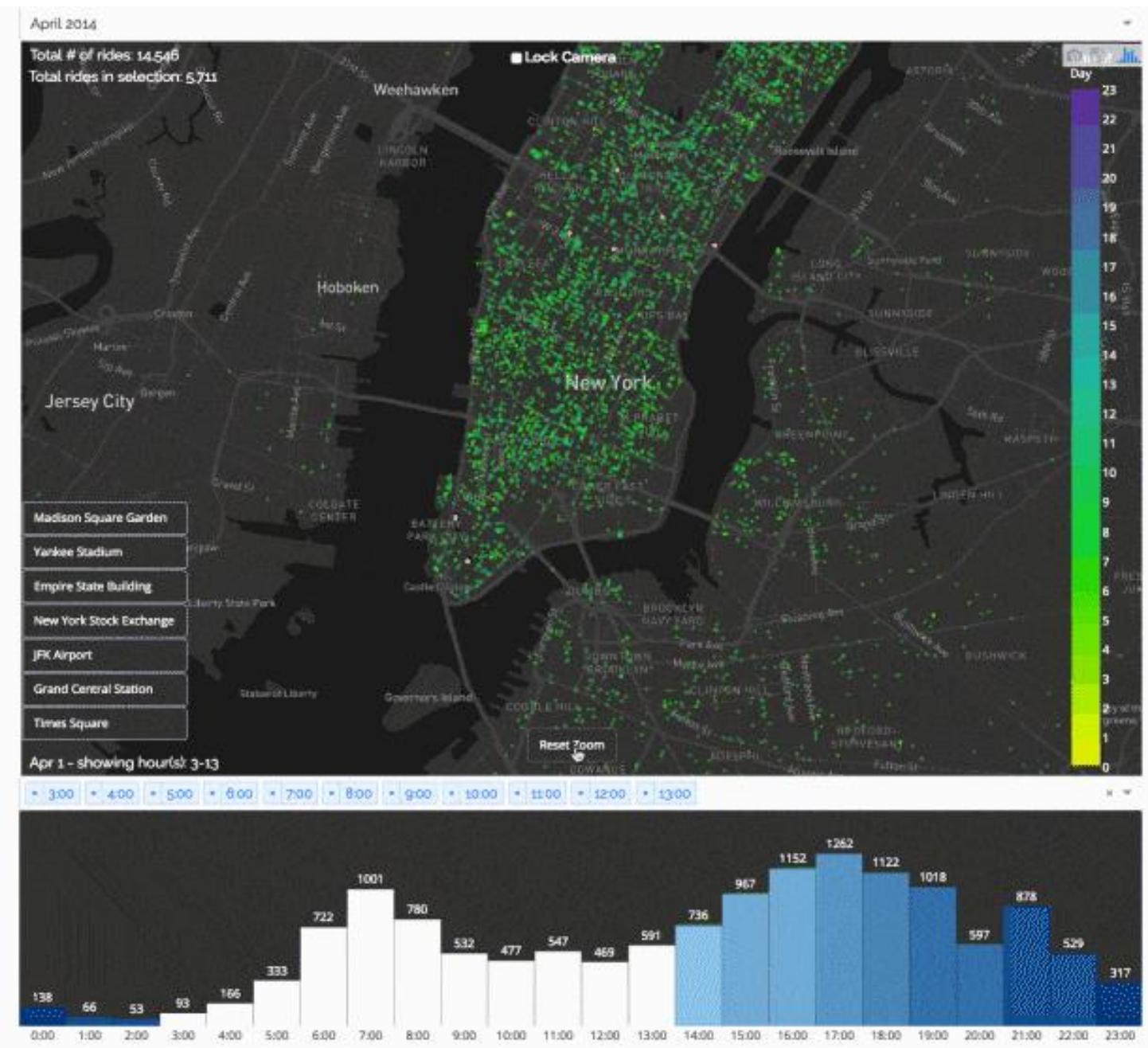
GeoPython 2021

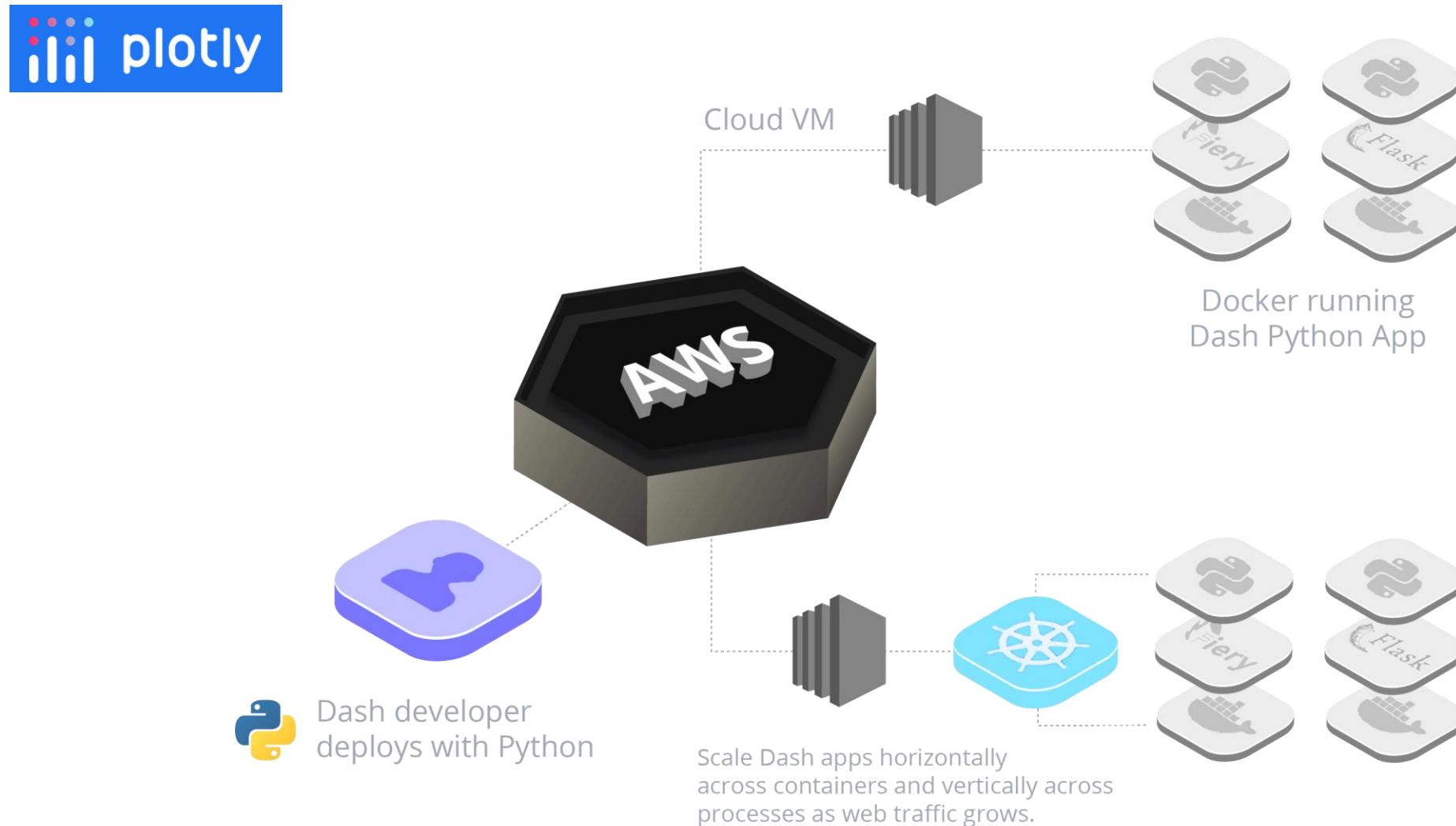
Worldwide Online

April 22-23



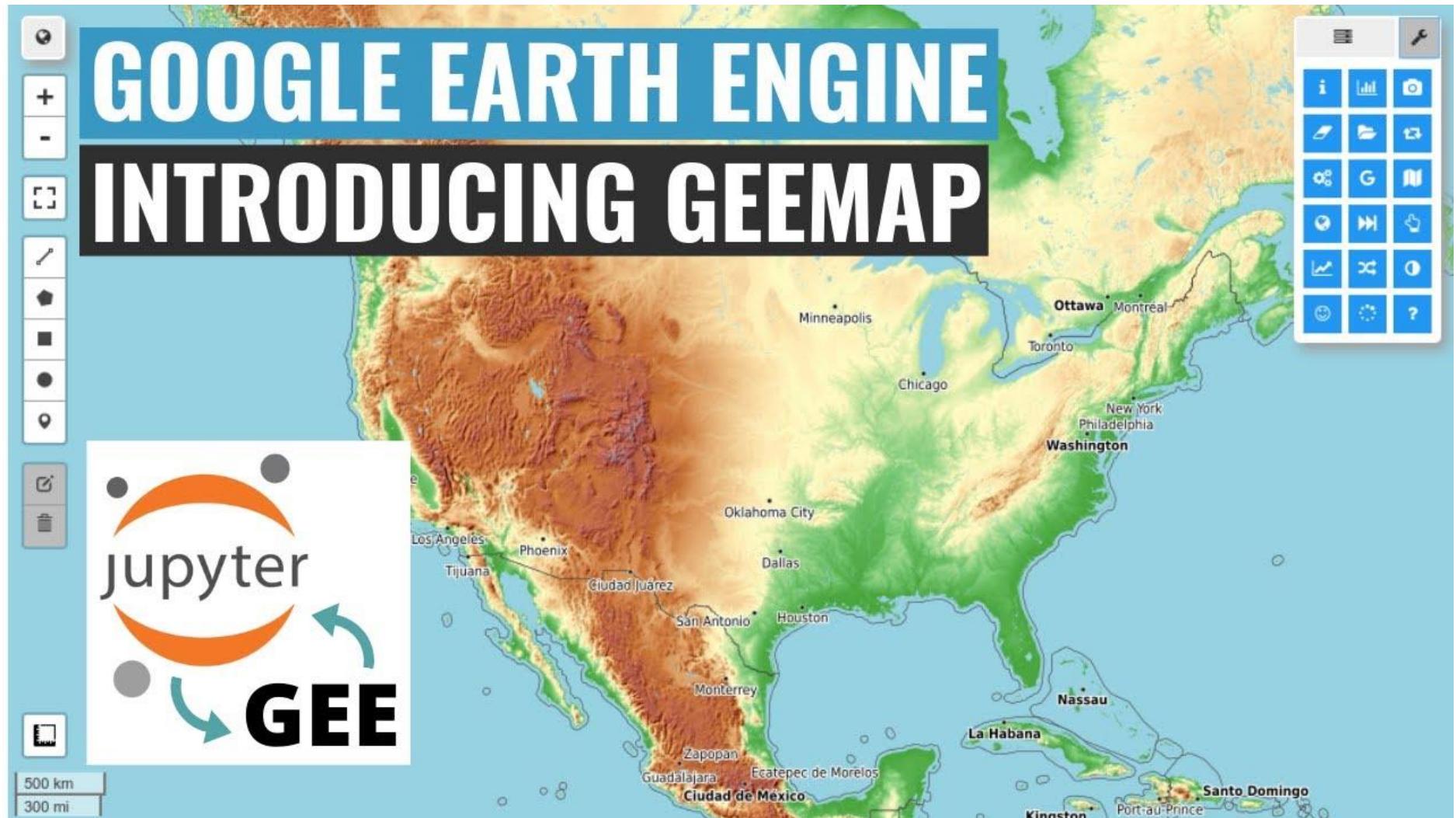
Main Topics:GIS/Mapping
Geography / Geophysics / Geodesy / Geomatics
Earth Sciences / Environmental Sciences
Geovisualization
Smart Cities
Spatial Data / Geodata
Geospatial Webservices
Big Data
Data Processing
(Spatial) Databases
Computer Vision
Remote Sensing
Image Processing
Machine Learning / Deep Learning
Mobile Mapping
Indoor Mapping and Modelling





Plotly is a production grade visualization platform with wrappers in not just python but other languages like R, Julia, MATLAB, etc. Plotly provides visualizations, online plotting, statistical tools along with a suite of solutions like Dash and Chart Studio to cater to different needs.

Geospatial Data Science



A Python package for interactive mapping with Google Earth Engine, ipyleaflet, and ipywidgets.



GeoPandas is an open source project to add support for geographic data to pandas objects. It currently implements GeoSeries and GeoDataFrame types which are subclasses of pandas.Series and pandas.DataFrame respectively. GeoPandas objects can act on shapely geometry objects and perform geometric operations.

PYTHON IDE



Home — Spyder IDE

The screenshot shows the Spyder IDE interface with the following components:

- Left Panel:** Shows the file tree with several Python files under the `spyder/plugins/plots` directory, including `plugin.py`, `chart_plot_example.py`, and `ipythonconsole.py`.
- Code Editor:** Displays the `plugin.py` file content, which includes imports for `qtpy.QtCore`, `Plugins`, `SpyderDockablePlugin`, and `PlotsWidget`. It defines a `Plots` class that implements the `SpyderDockablePlugin` API.
- Variable Explorer:** A table showing variables and their values:

Name	Type	Size	Value
bool	bool	1	True
data	Array of str128	(3, 3)	ndarray object of numpy module
datetime_object	datetime	1	2021-04-14 17:35:14.687085
df	DataFrame	(2, 2)	Column names: Col1, Col2
filename	str	53	/Users/Documents/spyder/spyder/tests/test_dont_use.py
li	list	5	['abcd', 745, 2.23, 'efgh', 70.2]
myset	set	3	{'2', '1', '3'}
r	float	1	6.46567886443
t	tuple	5	('abcd', 745, 2.23, 'efgh', 70.2)
tinylist	list	2	[123, 'efgh']
x	float64	1	1.1235123099439
- 3D Plot:** A 3D surface plot showing a terrain model with axes labeled from -84.41 to 90° and 36.70 to 700.
- Bottom Status Bar:** Shows the environment as `conda: spyder-dev Python 3.8.5`, the LSP status as `LSP Python: ready`, and system information like master, Line 10, Col 1, UTF-8, LF, RW, Mem 64%.

Spyder is a free and open source scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It features a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

<https://www.youtube.com/watch?v=WV9bm4ey7Cg>

자료형

- 기본 자료형: 숫자형(number) / 문자형 (string)
- 구조 자료형: 배열형 – 리스트(list) / 튜플(tuple)
 - 집합(Set)
 - 사전 (Dictionary)
 - 부울 (Boolean)

자료형은 프로그래밍의 기본이자 핵심 단위.
프로그래밍을 할 때 쓰이는 숫자, 문자열 등 자료 형태로 사용되는 모든 것을 의미.

제어문

- 조건문(if문)
- 반복문(while문)
- 반복문(for문)

제어문: 자료형을 제어한다.

함수, 클래스

함수: 프로그램의 반복적인 입력(Input)과 출력(Output)을 실행

클래스: 인스턴스를 생성하는 틀. 속성(attribute)과 속성의 동작을 수행하는 메소드(method)로 구성.

메소드는 클래스 내에 정의 된 함수이며, 속성은 함수를 통해 할당됨

클래스로부터 객체를 생성하는 것을 인스턴스화(instantiation)라고 함.

생성된 인스턴스가 가지고 있는 속성과 메소드는 닷(.)을 사용해서 접근할 수 있음

패키지, 모듈

패키지: 라이브러리, 특정 기능과 관련된 여러 모듈을 한 그룹으로 묶은 것으로 패키지 안에 패키지가 있을 수도 있다. Import from 패키지

모듈: 클래스와 함수의 집합인 파이썬 파일. 변수, 함수, 클래스 등을 포함

<https://seoulpi.co.kr/65702/>

<https://hyunsitstory.tistory.com/entry/python-%ED%8C%A8%ED%82%A4%EC%A7%80-%EB%AA%A8%EB%93%88-%ED%95%A8%EC%88%98-%ED%81%B4%EB%9E%98%EC%8A%A4-%EA%B0%9C%EB%85%90-%EC%A0%95%EB%A6%AC>

자료형 객체

자료형 객체	의미	사용
숫자형 int	자연수 객체	자연수
숫자형 float	실수 객체	실수
문자 자료형 str	문자열 객체	글자
리스트형 list	가변 컨테이너	변화하는 레코드 집합
튜플형 tuple	불변 컨테이너	고정된 객체 집합, 레코드
집합형 set	가변 컨테이너	유일한 객체 집합
사전형 dict	가변 컨테이너	키와 값의 저장소
부울형 bool	비교, 논리연산	참 또는 거짓

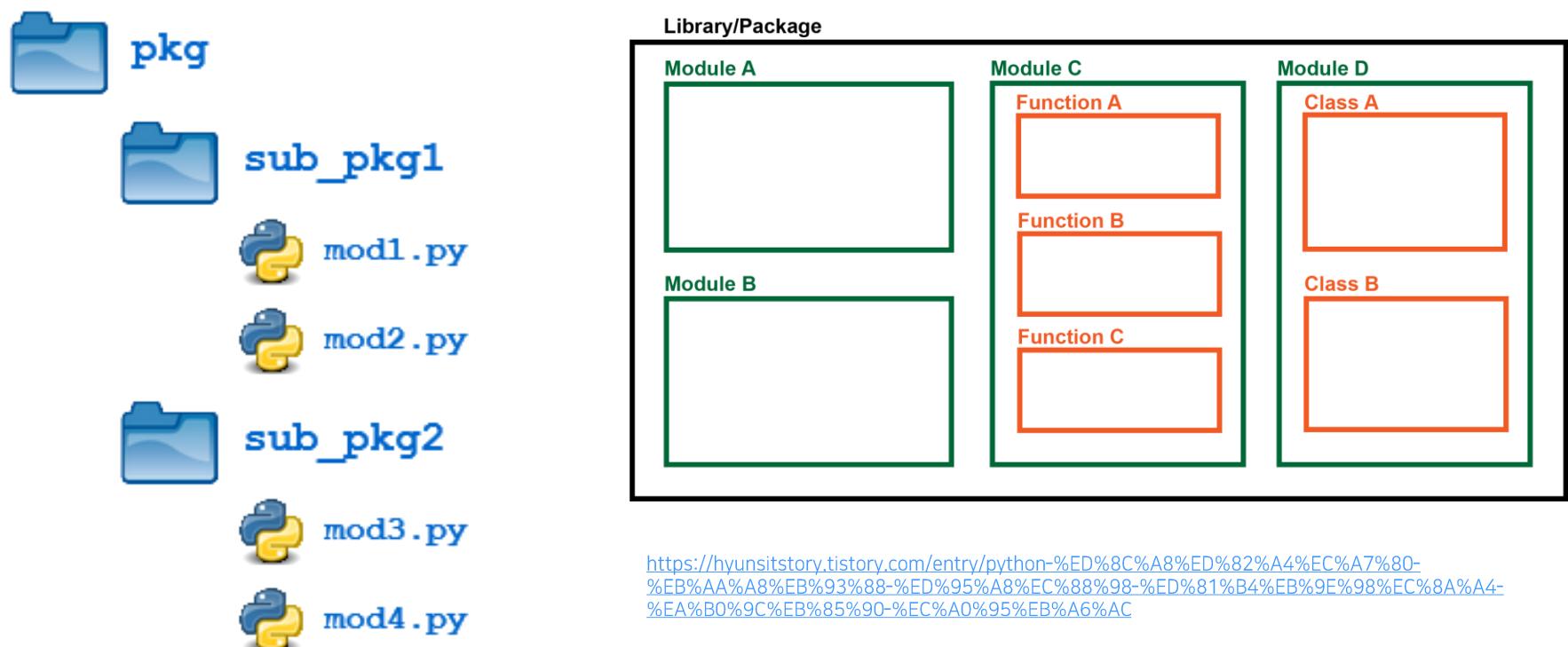
*객체지향적 프로그래밍: 소프트웨어 개발에 필요한 모든 요소를 객체화하여 프로그래밍하는 기법

라이브러리 패키지 > 모듈의 집합 > 변수, 함수, 클래스를 모아 놓은 파일

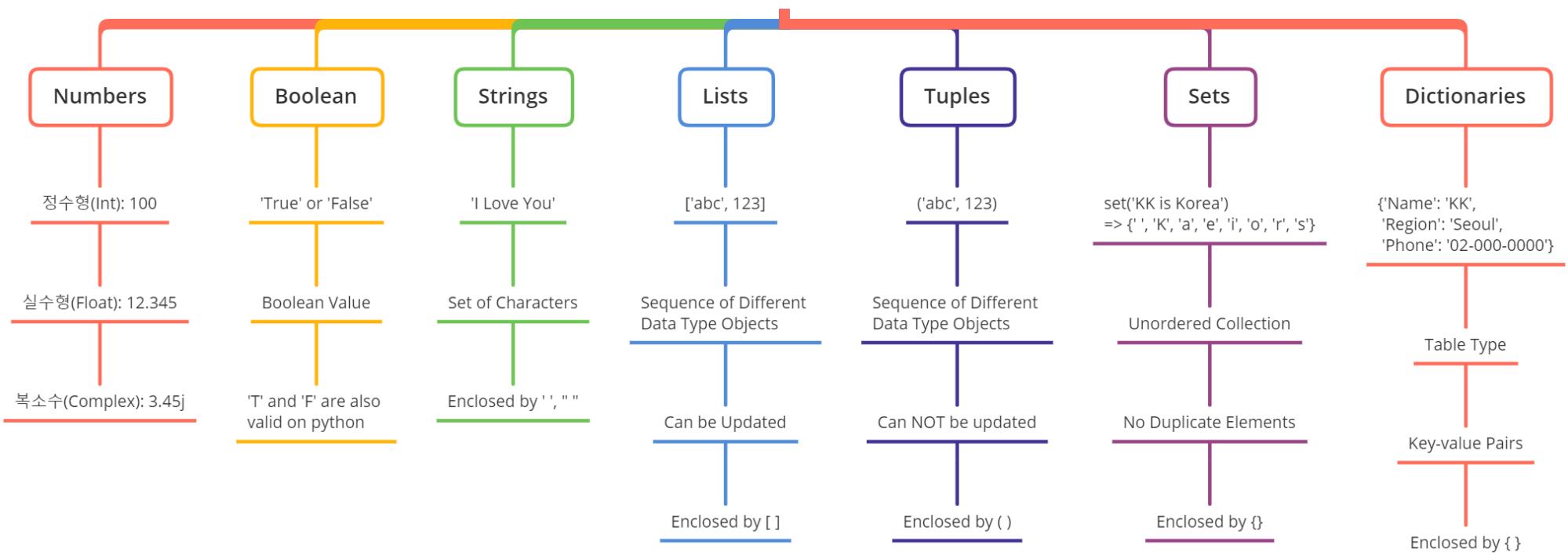
파이썬 최고의 강력한 장점 중 하나는 오픈소스 커뮤니티가 계속해서 발전시켜 나가는 **라이브러리 패키지**들을 사용할 수 있다는 것이다.

다른 컴퓨터 과학자가 모듈(module) 단위로 작성해 둔 코드를 가져다 쓸 수 있다. **모듈은 미리 만들어 놓은 파이썬(.py) 파일**이다.

패키지(package)은 여러 모듈을 특정 디렉토리에 모아 둔다. 단(.)을 사용하여 파이썬 모듈을 계층적으로 관리할 수 있게 한다. 예를 들어, `scipy.spatial`이 있으면 `scipy`는 패키지 이름이고 `spatial`은 패키지의 모듈이다. **모듈이란 변수, 함수, 클래스를 모아 놓은 파일**이다. 다른 파이썬 프로그램에서 불러와 사용할 수 있게 만들어졌다. 본인이 만들어 놓거나 다른 사람이 만들어 놓은 모듈을 가져와 활용할 수 있다.



Python Data Types

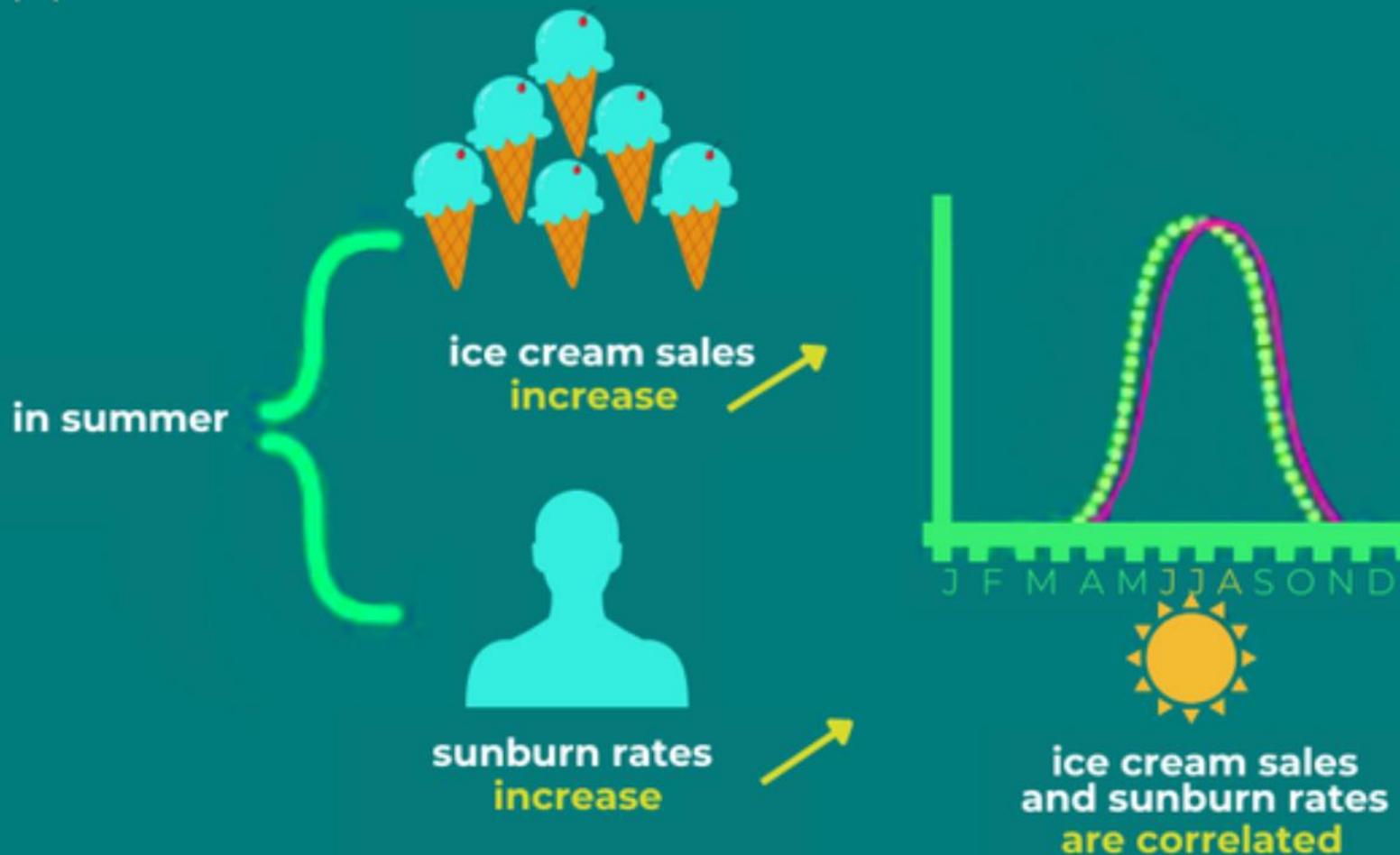


기준 금리와 부동산 매매지수의 상관관계 분석

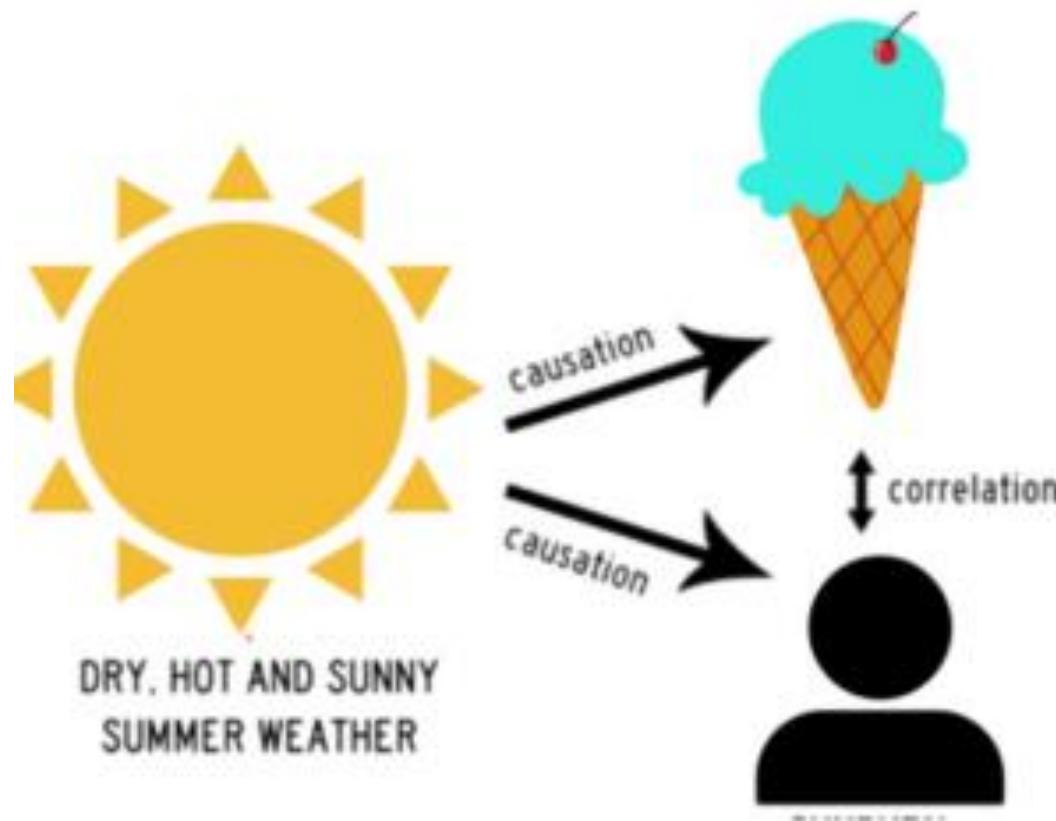
e- 나라지표 / KOSIS

Correlation

when two or more things appear to be related



does this mean eating ice cream increases your risk of sunburn?



피어슨 상관 관계 계수(Pearson Correalation coefficient)

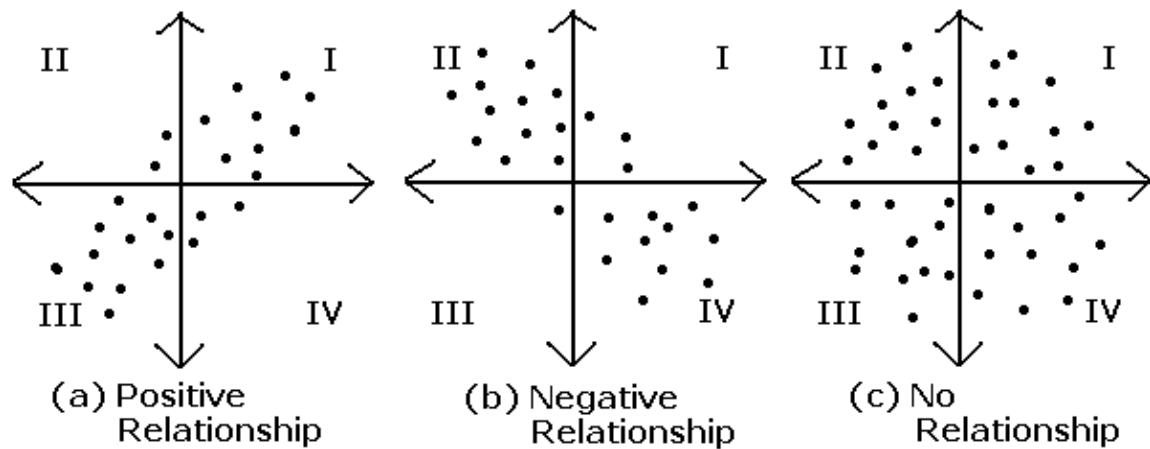
$$\text{피어슨상관계수} = \frac{\text{공분산}}{\text{표준편차} \cdot \text{표준편차}}$$

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X) Var(Y)}}, \quad -1 \leq \rho \leq 1$$

$$r_{XY} = \frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{\sum_i^n (X_i - \bar{X})^2}{n-1}} \sqrt{\frac{\sum_i^n (Y_i - \bar{Y})^2}{n-1}}}$$

$$r_{XY} = \frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2} \sqrt{\sum_i^n (Y_i - \bar{Y})^2}}$$

공분산 (Covariance, Cov)는 2개의 확률변수의 상관 정도를 나타내는 값이다.



$$E(X) = \mu, E(Y) = \nu$$

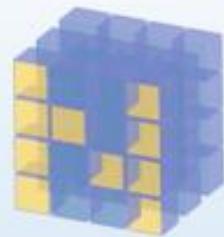
$$\text{Cov}(X, Y) = E((X - \mu)(Y - \nu))$$

$$\text{Cov}(X, Y) = E(X \cdot Y) - \mu\nu$$

$$E(X \cdot Y) = E(X) \cdot E(Y) = \mu\nu$$

<https://supermemi.tistory.com/entry/%ED%99%95%EB%A5%A0%EA%B3%BC-%ED%86%B5%EA%B3%84-%EB%B6%84%EC%82%B0variance-Var%EA%B3%BC-%EA%B3%B5%EB%B6%84%EC%82%B0Covariance-Cor-%EC%9D%B4%EB%9E%80-2>





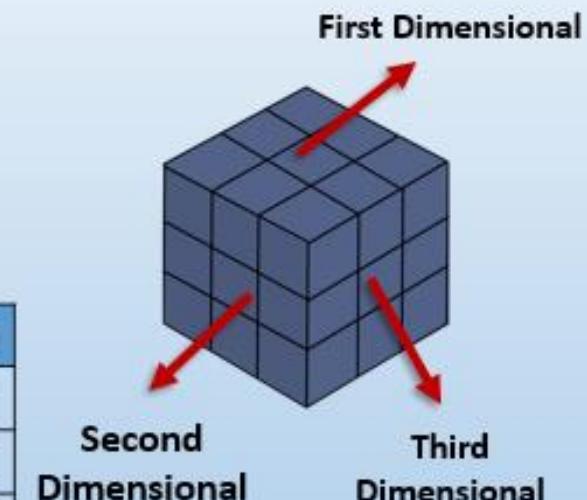
NumPy Ndarray

10	15	13	8	25
----	----	----	---	----

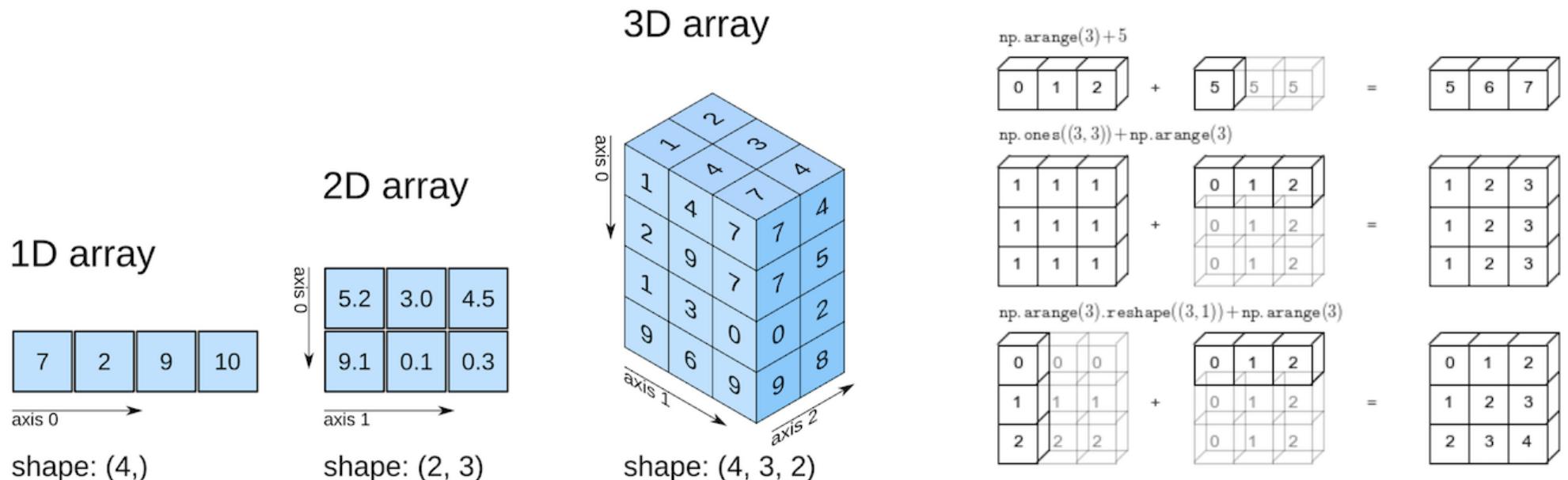
1D-Array

	Column 0	Column 1	Column 2
Row 0	X[0][0]	X[0][1]	X[0][2]
Row 1	X[1][0]	X[1][1]	X[1][2]
Row 2	X[2][0]	X[2][1]	X[2][2]

2D-Array



3D-Array



1D array

7	2	9	10
---	---	---	----

axis 0 →

shape: (4,)

2D array

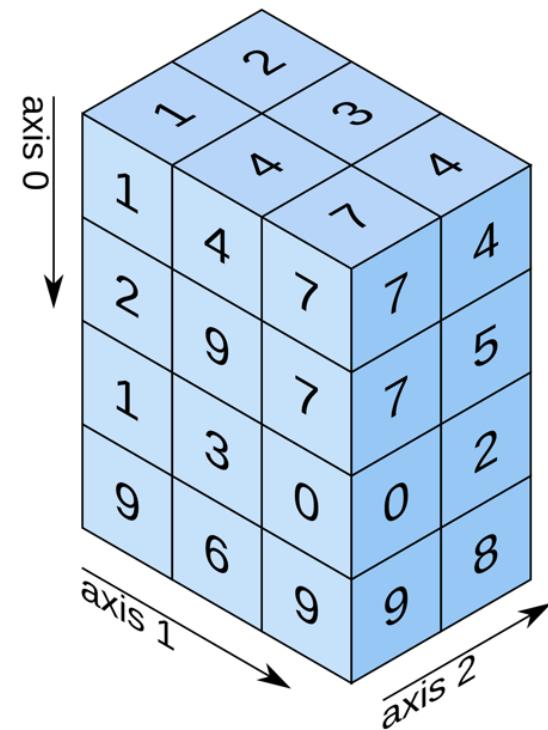
5.2	3.0	4.5
9.1	0.1	0.3

axis 0 ↓

axis 1 →

shape: (2, 3)

3D array



shape: (4, 3, 2)

`np.arange(3)+5`

A diagram illustrating element-wise addition of two 1D arrays. On the left, a 1D array `np.arange(3)` is shown as a horizontal box containing three cells with values 0, 1, and 2. To its right is a plus sign. Next is another 1D array `5` repeated three times, shown as a horizontal box containing three cells with values 5, 5, and 5. To the right of the plus sign is an equals sign. On the far right is the result, a 1D array `5+5` shown as a horizontal box containing three cells with values 5, 6, and 7.

`np.ones((3, 3))+np.arange(3)`

A diagram illustrating element-wise addition of a 3x3 matrix of ones and a 1D array. On the left, a 3x3 matrix of ones is shown as a 3x3 grid of boxes. To its right is a plus sign. Next is a 1D array `np.arange(3)` shown as a horizontal box containing three cells with values 0, 1, and 2. To the right of the plus sign is an equals sign. On the far right is the result, a 3x3 matrix where each element is the sum of the corresponding element from the first matrix and the 1D array, resulting in a 3x3 grid of boxes with values 1, 2, and 3.

`np.arange(3).reshape((3, 1))+np.arange(3)`

A diagram illustrating element-wise addition of a 3x3 matrix and a 3x1 column vector. On the left, a 3x3 matrix is shown as a 3x3 grid of boxes. To its right is a plus sign. Next is a 3x1 column vector `np.arange(3)` shown as a vertical stack of three boxes with values 0, 1, and 2. To the right of the plus sign is an equals sign. On the far right is the result, a 3x3 matrix where each element is the sum of the corresponding element from the first matrix and the column vector, resulting in a 3x3 grid of boxes with values 0, 1, 2, 1, 2, 3, 2, 3, 4.



판다스는 2008년 금융데이터 분석용으로 개발됨. 오픈소스 커뮤니티를 통해 계속해서 발전하고 있으며 통계, 데이터과학, 머신러닝 분야에서 가장 중요한 소프트웨어로 성장함

판다스 자료형

판다스는 데이터를 효율적으로 다루기 위해 시리즈(Series)와 데이터프레임(DataFrame)이라는 자료형을 사용한다. 데이터프레임은 엑셀에서 볼 수 있는 시트(Sheet)와 동일한 개념이며 시리즈는 시트의 열 1개를 의미한다. 파이썬으로 비유하여 설명하면 데이터프레임은 시리즈들이 각 요소가 되는 딕셔너리(Dictionary)와 유사하다.

판다스는 시리즈와 데이터프레임이라는 구조화된 데이터 형식을 제공한다. 서로 다른 종류의 데이터를 한곳에 담는 그릇(컨테이너)가 된다. 시리즈는 1차원 배열이고, 데이터프레임은 2차원 배열(행렬)이다.

판다스의 1차적인 목적은 서로 다른 유형의 데이터를 공통의 포맷으로 정리하는 것이다. 특히 행과 열로 이루어진 2차원 구조의 데이터프레임은 데이터 분석 실무에서 사용된다.

시리즈
(Series)

데이터프레임
(DataFrame)

정수형 위치 인덱스(integer position), 인덱스 이름(index name), 인덱스 라벨(index label)

인덱스 이름
(optional)

i_0

i_1

i_2

i_3

i_4

...

i_n

데이터 값
(원소)

v_0

v_1

v_2

v_3

v_4

...

v_n

위치 인덱스

0

1

2

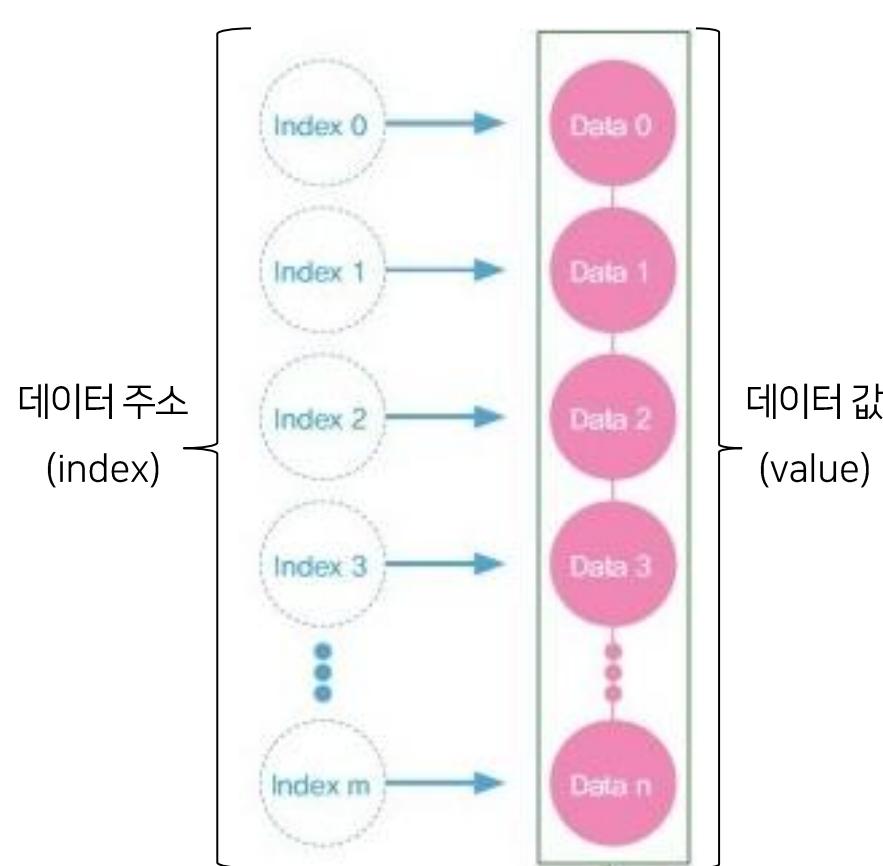
3

4

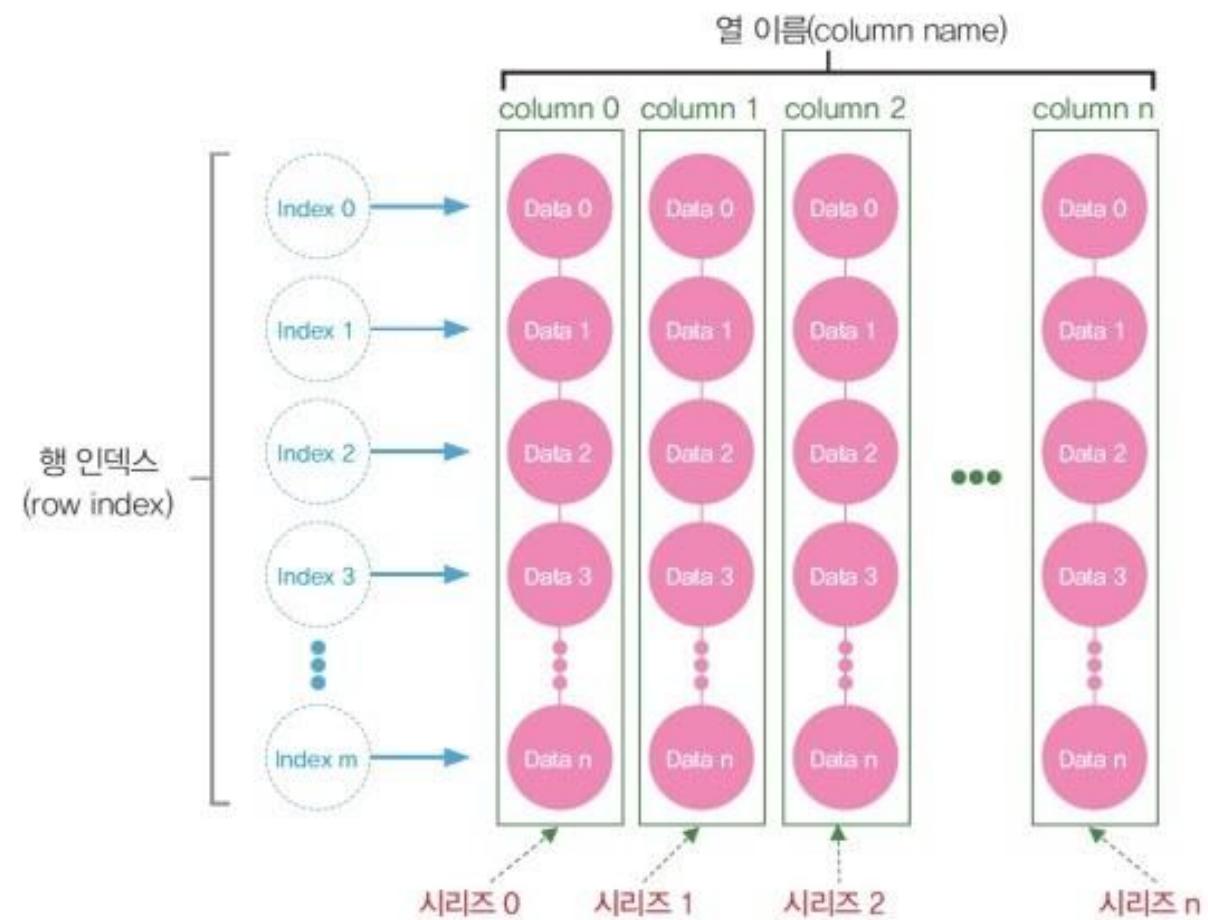
...

5

시리즈 (Series) 1차원 배열



데이터프레임 (DataFrame) 2차원 배열

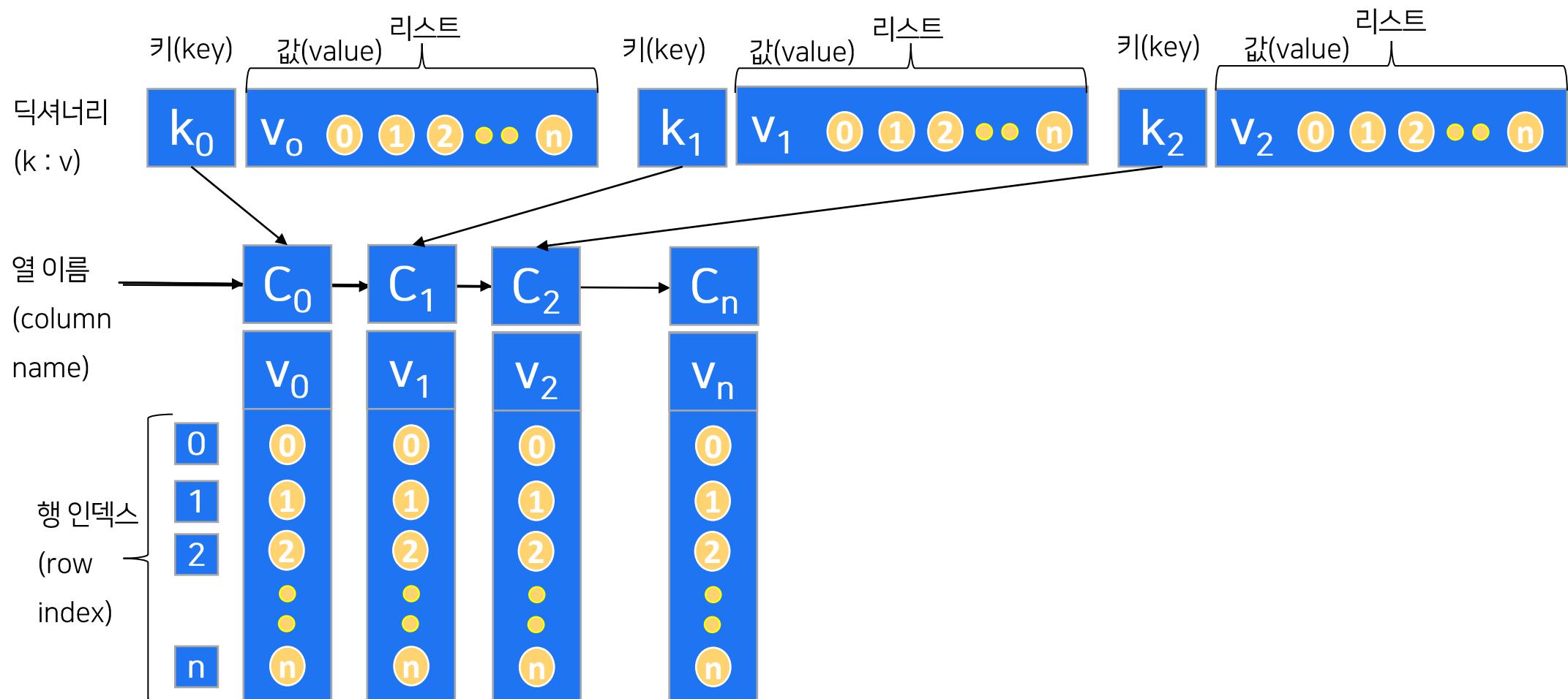


Python dictionary VS Pandas DataFrame

```
In [1]: import pandas as pd  
  
data = {'Key1': [1, 2, 3],  
        'Key2': ['Hank', 'Steve', 'Lisa'],  
        'Key3': [1.2, 3.1, 3.1]}  
  
df = pd.DataFrame(data)  
df.head()
```

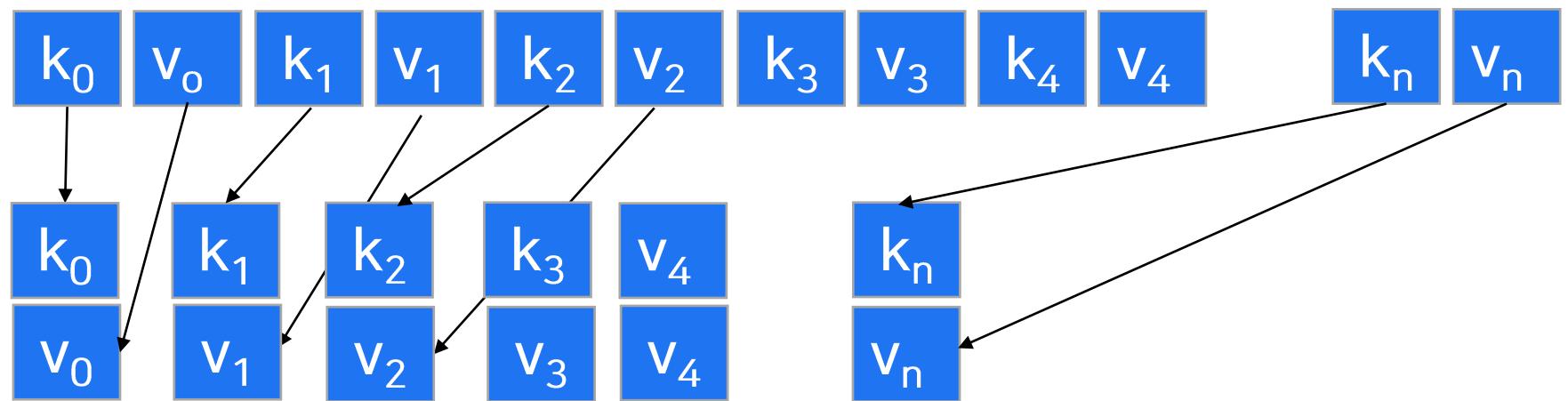
Out[1]:

	Key1	Key2	Key3
0	1	Hank	1.2
1	2	Steve	3.1
2	3	Lisa	3.1



키(key) 값(value)

딕셔너리
(k : v)



인덱스(index)

데이터 값(value)

파이썬 딕셔너리 자료형: {Key1 : Value1, Key2: Value 2, Key3: Value3, ...}

딕셔너리 -> 시리즈 변환 : pandas.Series(딕셔너리)

딕셔너리 -> 데이터프레임 변환 : pandas.DataFrame(딕셔너리 객체)

Series

apples	
0	3
1	2
2	0
3	1

Series

oranges	
0	0
1	3
2	7
3	2

+

DataFrame

apples		oranges
0	3	0
1	2	3
2	0	7
3	1	2

=

Series 1		Series 2		Series 3		DataFrame		
Mango	Apple	Banana		Mango	Apple	Banana		
0	4	0	5	0	2	0	4	2
1	5	1	4	1	3	1	5	3
2	6	2	3	2	5	2	6	5
3	3	3	0	3	2	3	3	2
4	1	4	2	4	7	4	1	7

판다스 라이브러리는 여러 종류의 클래스(class)와 다양한 내장 함수(built-in function)으로 구성된다. 시리즈와 데이터프레임은 구조를 표현하는 대표적인 클래스 객체이다. 시리즈와 데이터프레임 클래스의 속성과 메소드를 잘 이해하면 판다스를 자유자재로 다루는 데 어려움이 없다. 내장 함수로는 Series(), DataFrame(), read_csv(), read_excel() 등이 있다.

부동산 가격지수 시계열 예측 (ARIMA)

KOSIS

ARIMA(Autoregressive Integrated moving average)

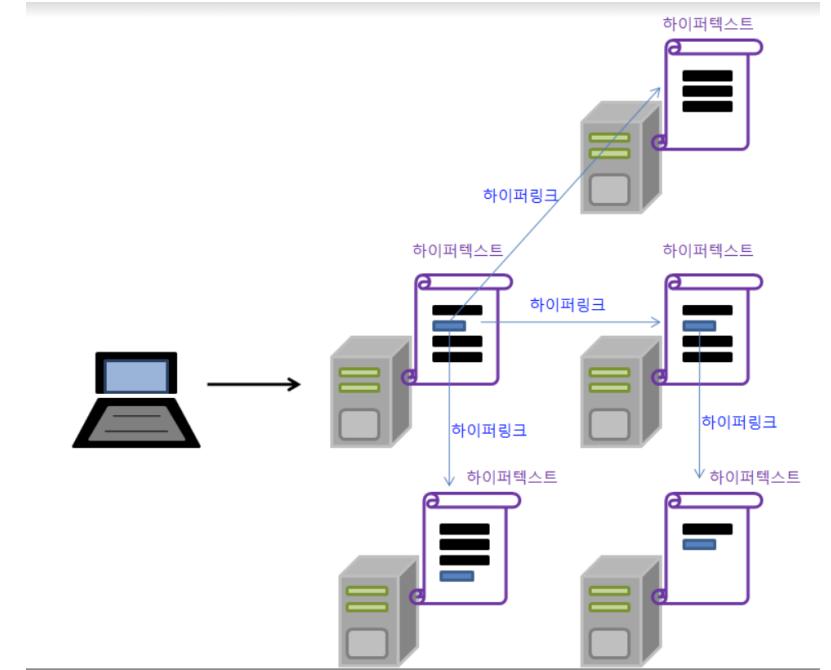
과거의 기록을 기반으로 주가 전망이나 수요 예측, AIRMA 모델은 과거의 관측치와 오차값을 기반으로 주가 전망이나 수요 예측에 사용된다. AIRMA 모델은 과거의 관측치와 오차값을 기반으로 현재의 시계열값을 설명하는 ARMA(autoregressive moving average) 모델을 일반화한 것이다. ARMA 모델이 안정적인 시계열(stationary time series)에만 적용 가능한 것에 비해 ARIMA 모델은 다소 비안정적인 시계열(non-stationary time series)에도 적용 가능

웹 크롤링하기



BeautifulSoup

BeautifulSoup, 해당 페이지의 HTML에서 원하는
데이터를 가져와서 파싱하는 라이브러리이다



끝없는 해체. 재구성이 '이어령표 이야기'를 만든다⑥ 이어령의 '하이パーテ스트'

<https://www.joongang.co.kr/article/6704227#home>

두 명의 지성인을 기리며

<http://www.latimes.kr/news/articleView.html?idxno=39650>

하이퍼텍스트를 기술하는 언어가 HTML(HyperText Markup Language)이다. HTML에서는 문장 표시 방법이나 하이퍼링크를 태그tag로 표현한다. 이런 언어를 일반적으로 마크업 언어라고 부른다.

HTML에서는 <태그의 종류> 태그의 의미를 적용할 대상이 되는 문장 </태그의 종류>라는 형태로 기술함으로써 문장에 의미를 줄 수 있다. 구체적으로는 해당 문장이 제목임을 나타내거나, 하이퍼링크임을 나타내거나, 테이블임 등을 나타낸다. 예를 들어, <title>웹 기술의 모든 것</title>은 해당 문자의 목이 '웹 기술의 모든 것' 이라는 의미이다.

HTML로 기술된 문서를 콘텐츠contents라고 부른다.

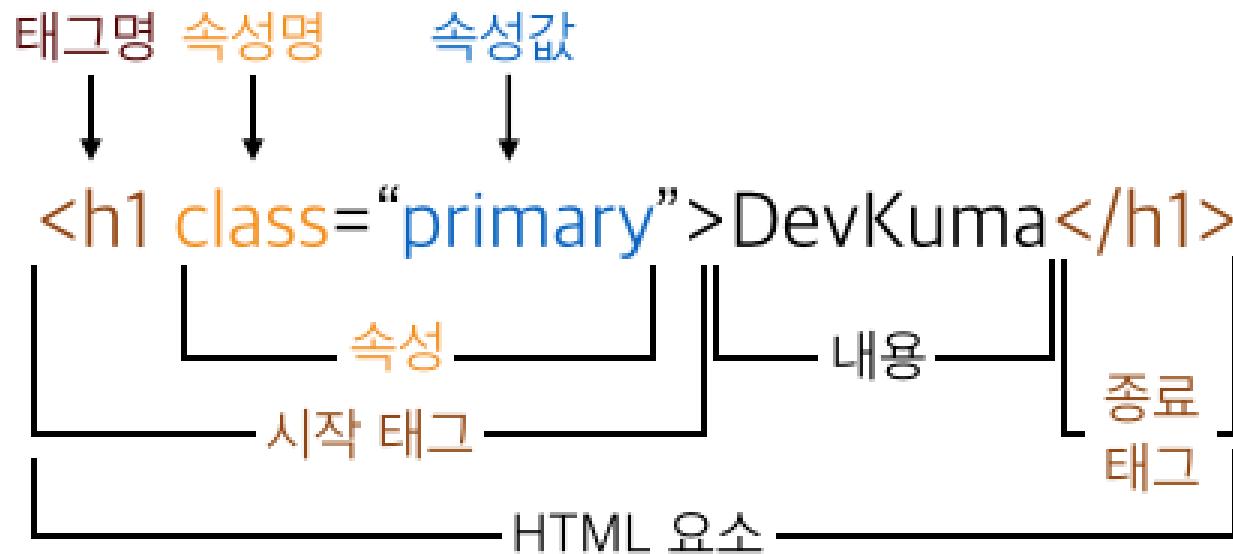
표시 프로그램 웹 브라우저

하이퍼텍스트는 문장에 태그로 의미를 붙인 것이며, 사람이 그 상태를 읽기에는 적절하지 않다. 그래서 하이퍼텍스를 해석해서 사람이 읽기 쉽도록 바꿔주는 것이 웹브라우저라는 프로그램이다. 일반적으로 사용하는 웹브라우저에는 크롬chrome, 에지edge 등이 있다.

웹브라우저의 종류에 따라 표시 방법이 다소 다르기는 하지만 HTML 규칙은 세계 공통이므로, 기본적으로 어떤 웹 브라우저에서도 동일하게 콘텐츠를 볼 수 있다.

출처: 웹의 기초 / 코바야시 쿄헤이, 사카모토 아키라 지음 / 위키북스

HTML 구성 요소 분석



한눈에 보는 HTML 요소(Elements & Attributes) 총정리

<https://heropy.blog/2019/05/26/html-elements/>

What's the Difference?



Create the structure

- Controls the layout of the content
- Provides structure for the web page design
- The fundamental building block of any web page



Stylize the website

- Applies style to the web page elements
- Targets various screen sizes to make web pages responsive
- Primarily handles the "look and feel" of a web page



Increase interactivity

- Adds interactivity to a web page
- Handles complex functions and features
- Programmatic code which enhances functionality

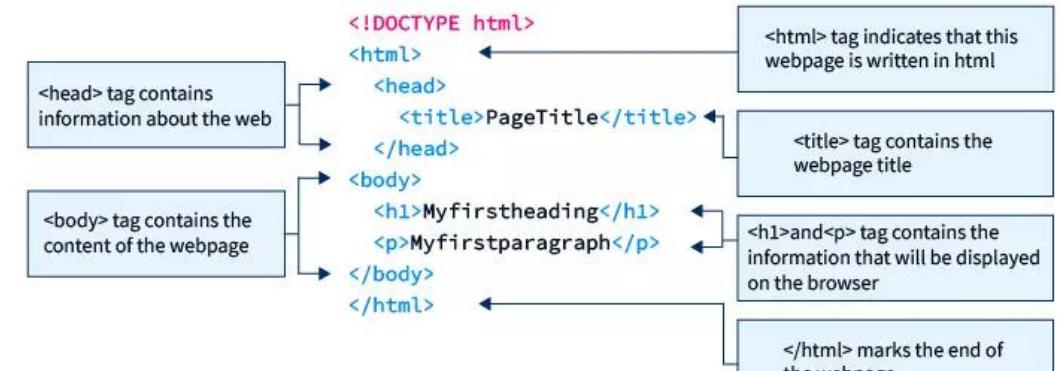
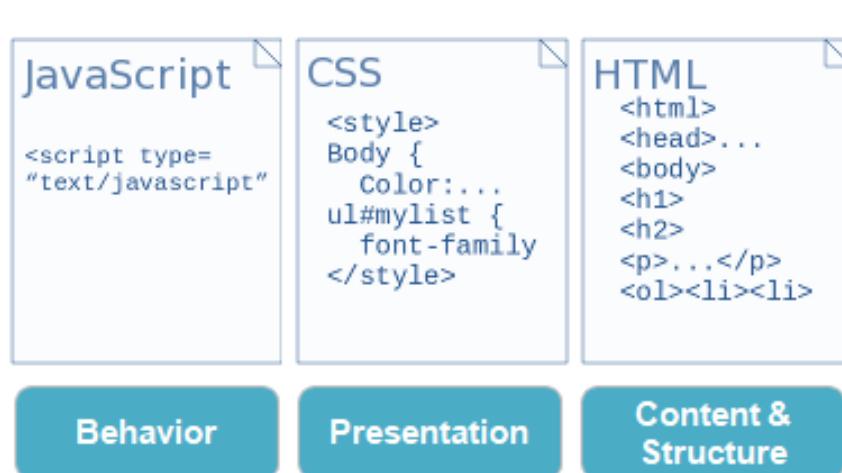
<https://seoulpi.io/article/32672>

https://charmingcity.github.io/spi_chatgpt/

HTML, CSS and JavaScript

These 3 web programming languages in conjunction take care of all the information webpage contains (from text to adding special effects).

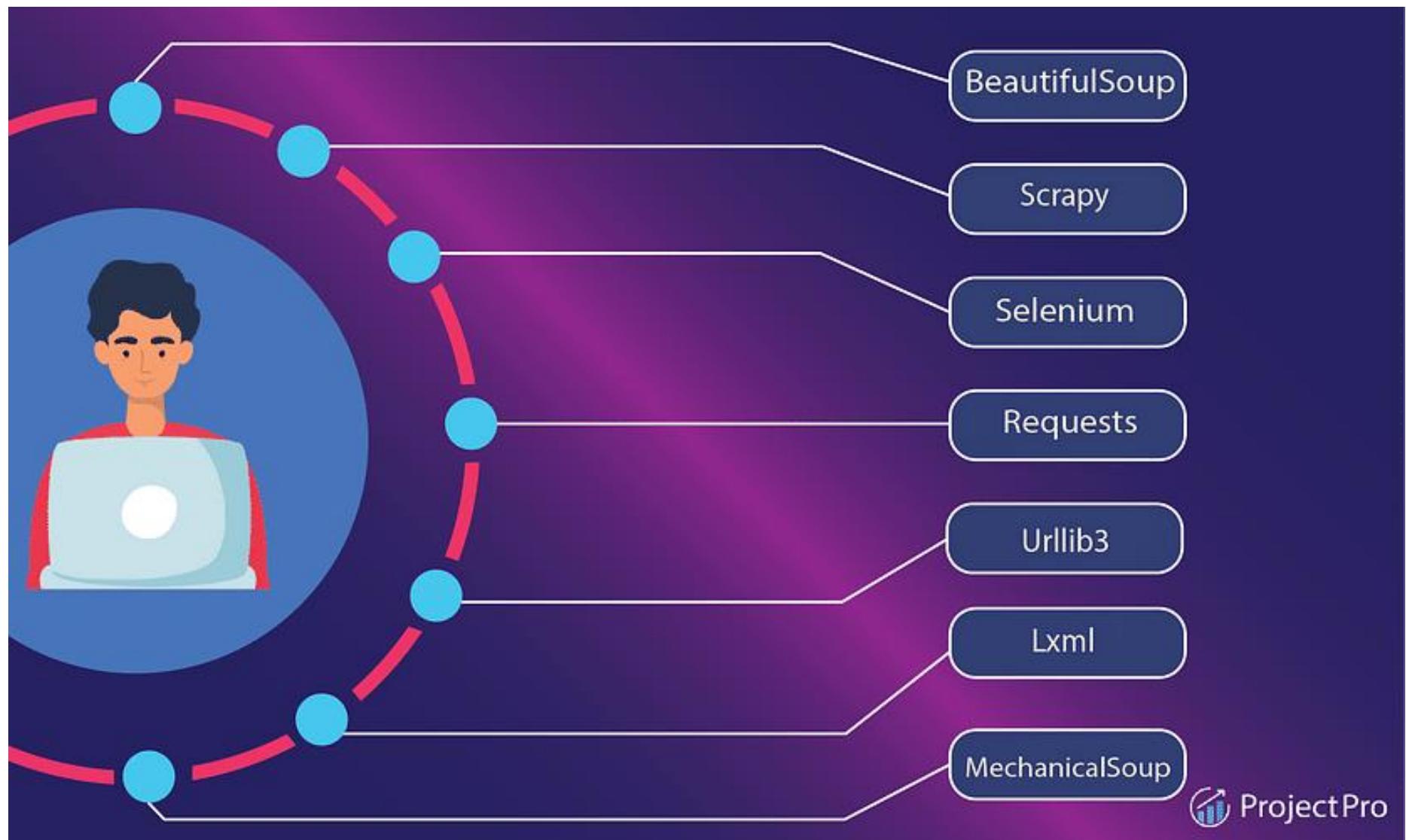
- HTML determines the content and structure of a page (header, paragraph, footer etc.)
- CSS controls how webpage would look like (color, font type, border etc.)
- JavaScript decides advanced behaviors such as pop-up, animation etc.



<https://sunjackson.github.io/2018/12/23/af11fc584cb237cd89ea175e9a713bbd/>

<https://www.scaler.com/topics/structure-of-html-document/>

SCALER
Topics



<https://www.projectpro.io/article/python-libraries-for-web-scraping/625>