

# 파이썬 기반 데이터 사이언스 입문

2024.06.

디지털신기술팀 이창훈 과장

# 데이터 분석 워크플로우



















데이터 입수, 데이터 전처리, 분석, 시각화, 결과 공유, 보고서 작성

# 파이썬(Python)

- 파이썬이란?

- 네덜란드 프로그래머 귀도 반 로섬(Guido van Rossum)이 1989년 크리스마스 연휴에 취미로 시작한 프로젝트에서 탄생
- 디자인 철학: 더 높은 생산성 제공, 쉬운 학습(명확하고 간결한 문법), 오픈소스, 유연성, 확장성
- 1991년 첫 번째 버전(0.9.0) 공개 이후, 현재 3.12.4 까지 공개
- 버전별로 문법이 달라 호환성 문제 있었지만, 3 이후 크게 완화
- 현재 전 세계적으로 가장 인기 있는 프로그래밍 언어(접근성, 범용성, 오픈소스 라이브러리)

# TIOBE index

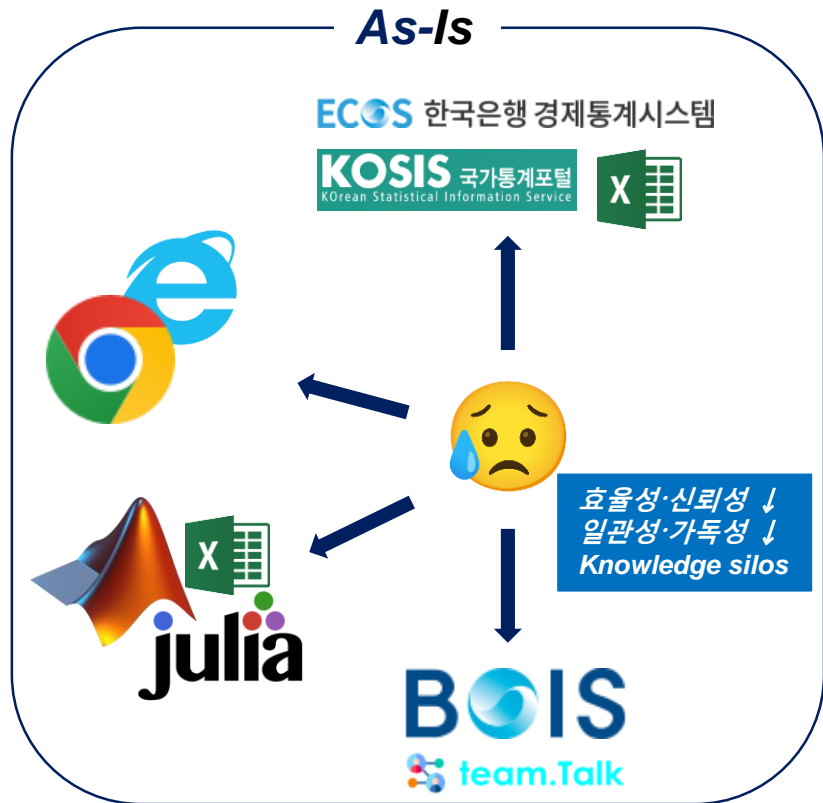
Jun 2024	Jun 2023	Change	Programming Language	Ratings	Change
1	1		 Python	15.39%	+2.93%
2	3	▲	 C++	10.03%	-1.33%
3	2	▼	 C	9.23%	-3.14%
4	4		 Java	8.40%	-2.88%
5	5		 C#	6.65%	-0.06%
6	7	▲	 JavaScript	3.32%	+0.51%
7	14	▲	 Go	1.93%	+0.93%
8	9	▲	 SQL	1.75%	+0.28%
9	6	▼	 Visual Basic	1.66%	-1.67%
10	15	▲	 Fortran	1.53%	+0.53%
11	11		 Delphi/Object Pascal	1.52%	+0.27%
12	19	▲	 Swift	1.27%	+0.33%
13	10	▼	 Assembly language	1.26%	-0.03%
14	12	▼	 MATLAB	1.26%	+0.14%
15	8	▼	 PHP	1.22%	-0.52%
16	13	▼	 Scratch	1.17%	+0.15%
17	20	▲	 Rust	1.17%	+0.26%
18	18		 Ruby	1.11%	+0.17%
19	29	▲	 Kotlin	1.01%	+0.50%
20	22	▲	 COBOL	0.96%	+0.22%

# 파이썬 데이터 분석 라이브러리와 인터페이스

- 라이브러리
  - Pandas(직관적인 데이터 처리), Numpy(수치 계산, matlab 대체), Matplotlib(시각화), scikit-learn, TensorFlow, Pytorch(AI/ML 분야) 등
- 인터페이스(IDE)
  - Jupyter notebook, Jupyterlab, VS code, Google colab

# BIDAS 인프라 기반 데이터 분석 업무 효율화

As-Is



To-Be

