



데이터 수집부터 분석까지

데이터는 현대 사회의 핵심 자원입니다. 이 프레젠테이션에서는 데이터의 기본 개념부터 실제 수집과 분석까지 다룹니다.

공공 데이터 포털을 활용한 실습을 통해 실질적인 경험을 쌓아보겠습니다.



by Lim

데이터의 정의와 중요성

데이터란?

데이터는 관찰이나 측정을 통해 수집된 사실이나 통계입니다.
. 정보의 원천이 되는 기초 자료입니다.

정보와의 차이

데이터는 원재료, 정보는 가공된 결과물입니다. 데이터 분석을 통해 의미 있는 정보를 얻습니다.

중요성

데이터는 의사결정의 기반이 됩니다. 객관적이고 과학적인 접근을 가능하게 합니다.



데이터의 유형

정형 데이터

구조화된 형태의 데이터입니다.
관계형 데이터베이스에 저장되며,
쉽게 검색하고 분석할 수 있습니다.

예: 엑셀 스프레드시트, SQL 데이터베이스

반정형 데이터

부분적으로 구조화된 데이터입니다.
태그나 마크업 언어를 사용하여 구조를 나타냅니다.

예: XML, JSON 파일

비정형 데이터

정해진 구조가 없는 데이터입니다.
. 텍스트, 이미지, 동영상 등 다양한 형태가 있습니다.

예: 소셜 미디어 포스트, 이메일 내용



데이터 수집의 중요성

1

데이터 기반 의사결정

객관적인 데이터를 바탕으로 더 나은 결정을 내릴 수 있습니다. 직관이나 경험에만 의존하지 않습니다.

2

인사이트 도출

데이터 분석을 통해 숨겨진 패턴이나 트렌드를 발견할 수 있습니다. 새로운 비즈니스 기회를 창출합니다.

3

예측 및 최적화

과거 데이터를 기반으로 미래를 예측하고 프로세스를 최적화할 수 있습니다. 효율성이 크게 향상됩니다.

데이터 수집 방법: 1차 vs 2차 데이터

1차 데이터

직접 수집한 원본 데이터입니다. 특정 목적에 맞춰 수집되어 정확성이 높습니다.

예: 설문조사, 실험, 관찰

2차 데이터

이미 수집된 데이터를 재사용하는 것입니다. 시간과 비용을 절약할 수 있습니다.

예: 정부 통계, 학술 논문, 기업 보고서

데이터 수집 방법: 설문조사와 웹 스크래핑

설문조사

직접적인 방법으로 원하는 정보를 얻을 수 있습니다.
온라인 도구를 활용하면 효율적입니다.

주의점: 샘플의 대표성, 질문의 객관성

웹 스크래핑

웹사이트에서 자동으로 데이터를 추출하는 기술입니다.
. 대량의 데이터를 빠르게 수집할 수 있습니다.

주의점: 법적, 윤리적 문제, 데이터 품질

데이터 수집 방법: API와 센서 데이터

API 활용

애플리케이션 간 데이터 교환을 위한 인터페이스입니다. 실시간 데이터 수집이 가능합니다.

예: 소셜 미디어 API, 기상 데이터 API

센서 데이터

IoT 기기나 센서를 통해 물리적 환경 데이터를 수집합니다. 지속적인 모니터링이 가능합니다.

예: 스마트홈 센서, 산업용 센서

센서 데이터의 종류



온도 센서

환경이나 기기의 온도를 측정합니다. 에너지 관리, 기후 연구에 활용됩니다.



GPS 센서

실시간 위치 정보를 추적합니다. 내비게이션, 물류 관리에 사용됩니다.





공공 데이터의 이해

정의

정부 및 공공기관이 생성하거나 취득한 데이터입니다. 국민에게 개방되어 활용될 수 있습니다.

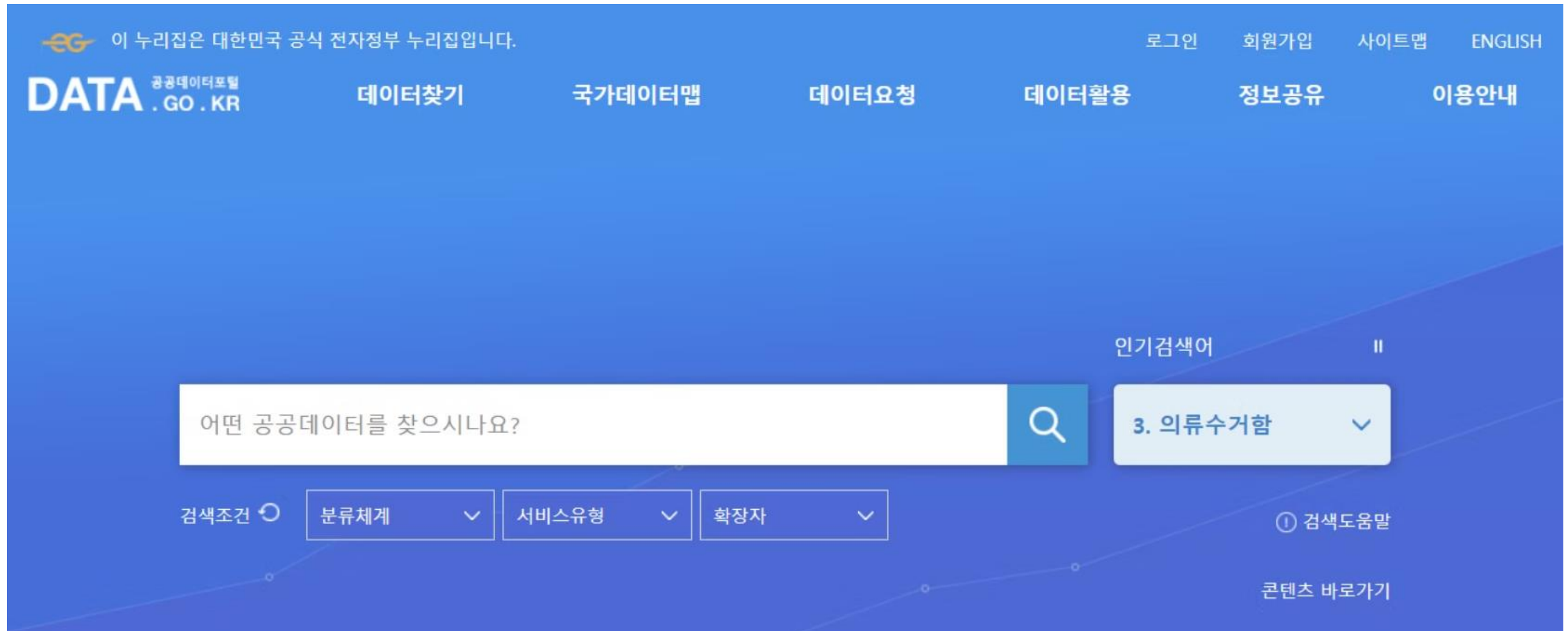
특징

공신력 있고 방대한 양의 데이터입니다. 대부분 무료로 제공되며, 정기적으로 업데이트됩니다.

활용 장점

신뢰할 수 있는 데이터로 연구나 비즈니스에 활용할 수 있습니다. 사회 문제 해결에 기여합니다.

공공 데이터 포털 소개



[공공데이터포털 \(data.go.kr\)](http://data.go.kr)



공공 데이터 포털 접속 및 이용

포털 접속

한국의 공공데이터포털(data.go.kr)에 접속합니다. 영어 버전도 제공되어 외국인도 이용 가능합니다.

1

2

회원가입

일반 회원가입 또는 간편 로그인을 통해 가입합니다. 개인정보 처리방침을 꼭 확인하세요



데이터 다운로드 및 활용

1

파일 다운로드

CSV, Excel 등의 형식으로 직접 다운로드합니다. 대용량 데이터

2

OpenAPI 활용

실시간 데이터 접근이 필요한 경우 API를 사용합니다. 개발자 문

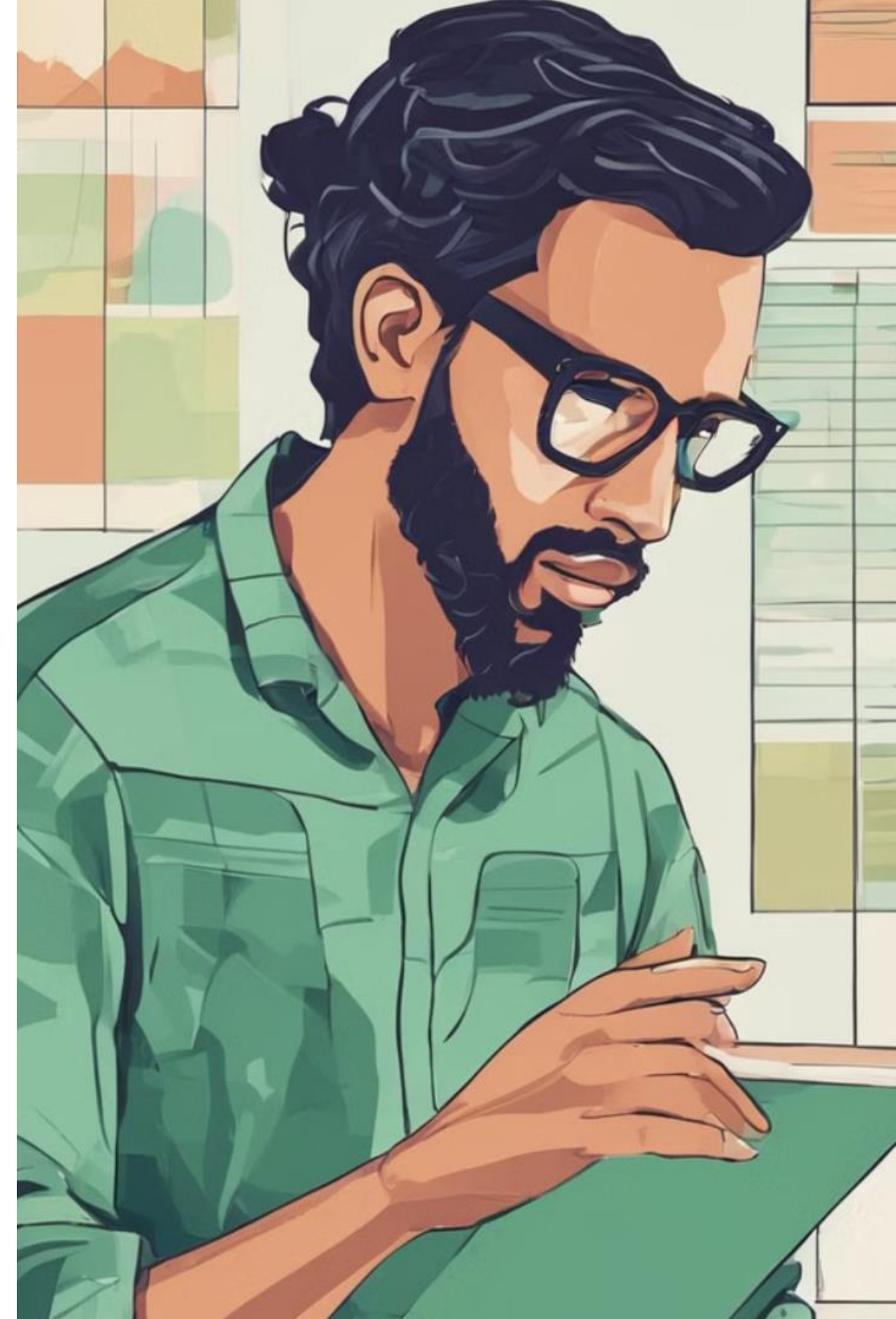
3

사용 조건 확인

라이선스와 사용 제한 사항을 반드시 확인합니다. 출처 표시 등의

데이터 기본 분석

단계	내용	도구
구조 파악	열(컬럼) 구조 이해	Excel, 텍스트 에디터, Power BI, Tableau, Google Sheets, R
데이터 유형 확인	숫자, 문자열, 날짜 등 구분	Excel, Python, Power BI, Tableau, Google Sheets, R
기본 통계	평균, 최대값, 최소값 계산	Excel, Python (pandas), Power BI, Tableau, Google Sheets, R



데이터 시각화 기초

1 matplotlib 소개

Python의 대표적인 데이터 시각화 라이브러리입니다. 다양한 그래프 생성이 가능합니다.

2 기본 그래프 유형

막대 그래프, 선 그래프, 산점도 등을 쉽게 만들 수 있습니다. 데이터 특성에 맞는 그래프를 선택하세요.

3 시각화의 중요성

복잡한 데이터를 직관적으로 이해할 수 있게 해줍니다. 패턴과 트렌드를 쉽게 파악할 수 있습니다.

