

빅데이터 분석과 관련한 가장 보편적인 데이터 유형

- **웹 데이터** : 방문 수, 페이지 조회, 검색, 구매 등과 같은 고객 수준의 웹 행동 데이터 등
- **텍스트 데이터** : 이메일, 뉴스 기, Facebook 피드, Word 문서 등 텍스트 소스에서 생성된 데이터로서 비구조적 데이터 중 가장 크고 가장 널리 사용되는 유형
- **시간 및 위치 또는 특정 지역 관련(geospatial) 데이터** : GPS와 휴대전화는 물론 Wi-Fi 연결 덕분에 시간과 위치 정보가 흥미로운 데이터 출처로서 두각. 여기에는 도로, 건물, 호수, 주소, 사람, 직장, 이동 경로 등 지리적 정보 시스템에서 생성된 정보와 관련된 지리적 데이터도 포함.
- **실시간 미디어** : 실시간 데이터 소스에는 실시간 스트리밍이나 이벤트 기반 데이터.
- **스마트 그리드와 센서 데이터** : 자동차, 송유관, 풍차 터빈 및 여타 센서에서 생성한 센서 데이터를 극히 높은 빈도로 수집
- **소셜 네트워크 데이터** : Facebook, LinkedIn, Instagram 등과 같은 소셜 네트워크 사이트에서 생성된 비구조적 텍스트(댓글, 좋아요 등)
- **링크드 데이터(Linked data)** : HTTP, RDF, SPARQL 및 URL과 같은 표준 웹 기술을 사용해 수집되어온 데이터입니다.
- **네트워크 데이터** : Facebook과 Twitter 같은 초대형 소셜 네트워크나 인터넷, 전화 및 교통수단 네트워크 등의 기술 네트워크와 관련한 데이터.

[참고] 빅데이터 분석 : <https://www.databricks.com/kr/glossary/big-data-analytics>

빅데이터 분석을 위한 공개 데이터 제공처 (국외)

- **Kaggle** : 데이터 과학 및 머신 러닝 경진대회를 주최하며, 이를 위한 데이터셋도 제공.
<https://www.kaggle.com/datasets>
- **UCI Machine Learning Repository** : 머신 러닝 분야에서 가장 유명한 데이터 제공처 중 하나로, 다양한 분야와 문제에 대한 데이터셋을 제공. <https://archive.ics.uci.edu/ml/index.php>
- **Google Dataset Search** : 구글에서 제공하는 데이터 검색 엔진으로, 다양한 분야와 주제에 대한 무료 데이터셋을 검색.
<https://datasetsearch.research.google.com/>
- **Data.gov** : 미국 정부에서 제공하는 공공 데이터 포털로, 정부 기관에서 수집한 다양한 데이터를 제공.
<https://www.data.gov/>
- **Amazon Web Services (AWS) Public Dataset** : 아마존 웹 서비스에서 제공하는 무료 공개 데이터셋으로, 텍스트, 음성, 이미지, 비디오 등 다양한 형태의 데이터를 제공. <https://aws.amazon.com/ko/datasets/>
- **World Bank Open Data** : 세계은행에서 제공하는 다양한 분야와 주제에 대한 데이터셋을 제공.
<https://data.worldbank.org/>
- **CDC Open Data** : 미국 질병통제예방센터에서 제공하는 건강 및 의료와 관련된 데이터를 제공.
<https://www.cdc.gov/datastatistics/index.html>
- **European Union Open Data Portal** : 유럽 연합에서 제공하는 공개 데이터 포털로, 다양한 분야와 주제에 대한 데이터셋을 제공. <https://data.europa.eu/euodp/ko/home>

빅데이터 분석을 위한 공개 데이터 제공처 (국내)

- **공공데이터포털**(www.data.go.kr): 국내 공공데이터를 한 곳에서 검색하고 다운로드할 수 있는 대표적인 포털입니다. 행정, 교육, 경제, 환경, 보건 등 다양한 분야의 데이터를 제공.
- **한국국가통계포털**(www.kosis.kr): 통계청에서 운영하는 국가통계포털로, 인구, 경제, 사회, 문화 등 다양한 분야의 통계 데이터를 제공.
- **서울시 열린데이터광장**(data.seoul.go.kr): 서울시에서 제공하는 열린데이터로, 서울시의 교통, 환경, 보건, 문화, 관광 등 다양한 분야의 데이터를 제공.
- **국립중앙도서관 디지털자료실**(dibrary.or.kr): 국립중앙도서관에서 운영하는 디지털자료실로, 문화, 예술, 역사, 지리, 자연 등 다양한 분야의 디지털 자료를 제공.
- **한국환경공단 환경데이터포털**(data.airkorea.or.kr): 한국환경공단에서 제공하는 환경 데이터 포털로, 대기오염, 수질오염, 소음 등 다양한 환경 데이터를 제공.
- **한국농수산식품유통공사**(www.at.or.kr): 농수산물식품 유통 정보를 제공하는 포털로, 농산물 도매시장 가격 정보, 실시간 상품 정보 등을 제공.
- **한국문화정보원**(www.culturecontent.or.kr): 한국문화정보원에서 운영하는 문화콘텐츠 정보 포털로, 한국의 문화유산, 문화재, 관광지 정보 등을 제공.
- **한국관광공사**(www.visitkorea.or.kr): 한국관광공사에서 제공하는 국내 관광정보 포털로, 지역별, 테마별, 여행코스별 정보 등을 제공.
- **데이터허브** : <https://dataapi.co.kr/>
- **데이터 안심구역** : <https://dsz.kdata.or.kr/svc/main/main.do>