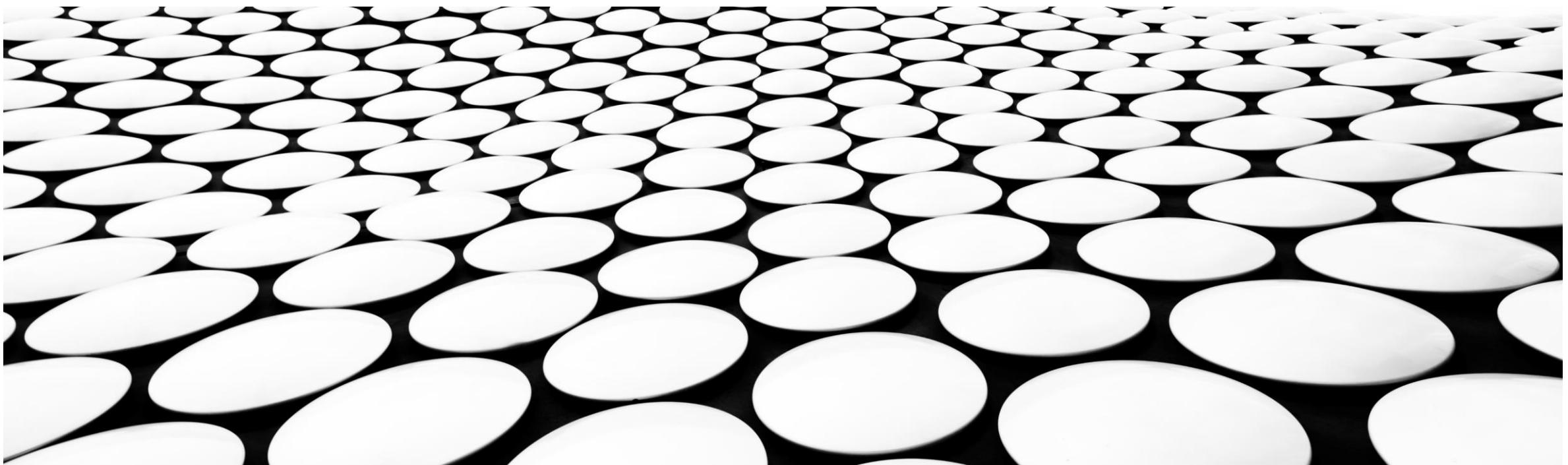


# 빅데이터 분석

## 1. 빅데이터 분석 기초





데이터로 추론하기



# 데이터 추론 1

## ■ 편의점 매출 손실 사건

- ☆☆ 편의점에서 최근 몇 주 동안 매출 손실이 발생했습니다. 점주는 이를 조사하기 위해 데이터 분석을 결정했고,
- 알바생 중 한 명이 매출 손실 사건에 연루된 것으로 추정됩니다. 여러분은 주어진 데이터를 분석하여 범인을 찾아내야 합니다.

근무 시작 시간	근무 종료 시간	종업원 이름
00:00	04:00	Alice
04:00	08:00	Bob
08:00	12:00	Charlie
12:00	16:00	David
16:00	20:00	Eve
20:00	24:00	Frank



# 데이터 추론 1

- 편의점 매출 손실 사건
  - 제공 데이터
    - 실제 판매된 수량과 매출 금액을 비교하여 매출 손실이 발생한 시점을 파악해야 합니다.

Sales Data	
Date	날짜
Time_Slot	시간대
Sales_Amount	매출액

Inventory Record Data	
Date	날짜
Time_Slot	시간대
Product_ID	제품 ID
Product_Price	제품 가격
Stock_Amount	재고 수량
Remaining_Stock	잔여 재고량

# 데이터 추론 2

## ■ 지명수배자 밀항 배 추적

- 한 지명수배자가 밀항을 시도할 계획이라는 정보를 입수했습니다.  
이 수배자의 출국을 막기 위해 그의 탑승 배를 찾아내야 합니다.  
여러가지 힌트를 바탕으로 지명수배자가 타는 배의 이름을 추리하세요.

### 힌트

1. 인천항에서 탑승한다.
2. 지명수배자의 이름에 '김'이 포함되어 있다.
3. 아시아 국가로 가는 배에 탑승한다.
4. 출발 시간이 08:00~12:00 사이다.
5. 도착 시간이 18:00~23:00 사이다.



# 데이터 추론 2

- 지명수배자 밀항 배 추적
  - 제공 데이터
    - 주어진 힌트에 맞는 데이터를 필터링하여 분석해야 합니다.

Passenger List	
Ship_ID	배 이름
Passenger_ID	승객 ID
Passenger_Name	승객 이름
Boarding_Port	탑승 항구

Ship Schedules	
Ship_ID	배 이름
Destination	목적지
Departure_Time	출발 시간
Arrival_Time	도착 시간

# 빅데이터의 개념 및 특성

# 빅데이터의 개념 및 특성

- 데이터의 유래

\*Dare : 주다

**Data : 주어진 것 이라는 뜻의 라틴어 Datum의 복수형**

**“Present / gift that which is given / debit”**

**우리 눈앞에 주어진 실재하는 이 세상의 모든 것**

# 빅데이터의 개념 및 특성

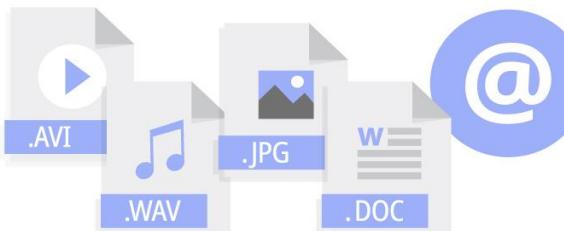
## ■ 데이터의 형식

### ■ 정형 데이터 (Structured Data)

- 미리 정해놓은 형식과 구조에 따라 저장된 데이터
- 관계형 DB, 스프레드시트, csv 데이터

ID	Name	AGE	SEX
01	KIM	32	M
02	LEE	26	F
03	PARK	72	F
04	CHOI	15	M

structured  
data



unstructured  
data

### ■ 비정형 데이터 (Unstructured Data)

- 정해진 구조가 없이 저장된 데이터
- text, image, sound, multimedia data

# 빅데이터의 개념 및 특성

## ■ 데이터 구분



성별 : [남자, 여자]

온도

지역 : [서울, 울산, 부산, 대구]

키

계절 : [봄, 여름, 가을, 겨울]

몸무게

월 : [1월, 2월, … 12월]

가격

# 빅데이터의 개념 및 특성

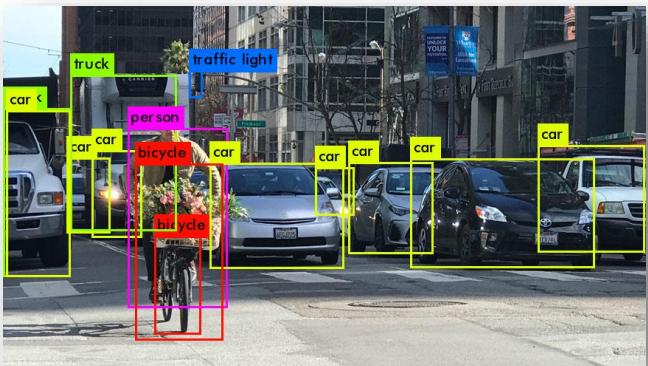
## ■ 데이터 구분

비정형데이터

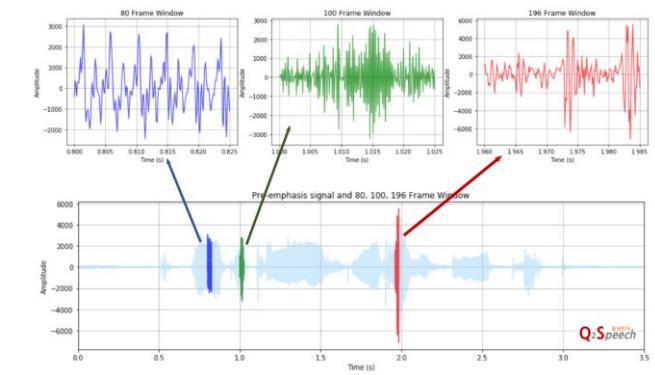
이미지 데이터

자연어 데이터

소리 데이터



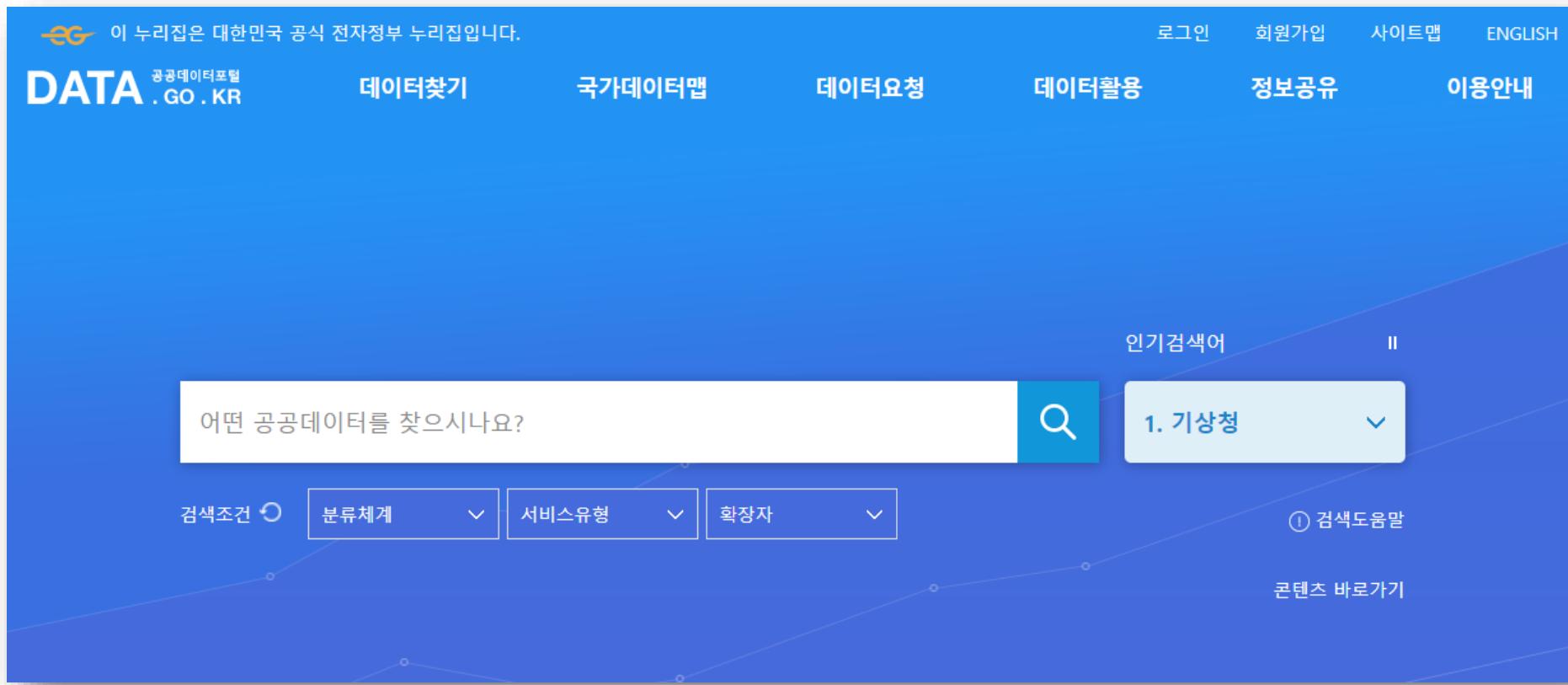
ATTRACTIVE PRETTY  
WINNING BONNY PICTURE  
LOVELY FETCHING BEGUILING  
APPEALING ENGAGING BEDDABLE  
NUBILE GORGEOUS HANDSOME RAVISHING  
SULTRY AROUSING GOOD-LOOKING SEXY INFORMAL FANCIALE  
SEXUAL BEWITCHING CHARISMATIC BEAUTIFUL CAPTIVATING  
LUSH AS DESIRABLE DELIGHTFUL ENCHANTING  
PRECIOUS CHARMING ARRESTING SEXUALLY  
LUSCIOUS VALUABLE STUNNING  
STRIKING ALLURING EROTIC  
SENSUAL



# 빅데이터의 개념 및 특성

## ■ 데이터 사이트

- 공공데이터 포털(<https://www.data.go.kr/>)



# 빅데이터의 개념 및 특성

## ■ 데이터 사이트

- AI-Hub(<https://aihub.or.kr/>)

The screenshot shows the homepage of the AI-Hub website. At the top, there is a navigation bar with links for AI 데이터찾기, AI 허브소개, 참여하기, 커뮤니티, AI 개발지원, 고객지원, 로그인, and 회원가입. Below the navigation bar, there is a banner featuring icons for various data categories: 금융, 한국어, 영상이미지, 헬스케어, 교통물류, 재난안전환경, 농축수산, and 문화관광. A search bar with placeholder text "찾으시는 데이터를 입력해주세요" and a magnifying glass icon is located below the banner. On the left side, there is a login form with fields for email and password, and buttons for "로그인", "내이버 로그인", and "카카오 로그인". Below the login form, there is a red banner with the text "AI 데이터 분석부터 활용까지 데이터도전전설". In the center, there is a section titled "인기 데이터 TOP3" with three items: 1. [한국어] 감성 대화 말뭉치 (78,086), 2. [영상이미지] 음식 이미지 및 영양정보 텍스트 (51,188), and 3. [한국어] 한국어 음성 (48,451). To the right, there is a large blue rectangular area with the text "AI 데이터" and a descriptive paragraph: "지능정보산업 인프라 조성사업으로 추진한 AI 학습용 데이터(14개 분야)와 국내외 기관/기업에서 보유한 AI 학습용 데이터를 공개". Below this text, there are two rows of icons representing different data domains: 한국어, 영상이미지, 헬스케어, 교통물류, 재난안전환경, 농축수산, 문화관광; and 스포츠, 교육, 로보틱스, 제조, 지식재산, 법률, and 금융.

# 빅데이터의 개념 및 특성

## ■ 데이터 사이트

- Kaggle(<https://www.kaggle.com/datasets>)

The screenshot shows the Kaggle Datasets homepage. At the top, there's a search bar labeled "Search datasets" and a "Filters" button. Below the search bar are several category filters: "All datasets", "Computer Science", "Education", "Classification", "Computer Vision", "NLP", "Data Visualization", and "Pre-Trained Model". A "Trending Datasets" section is displayed, showing three datasets:

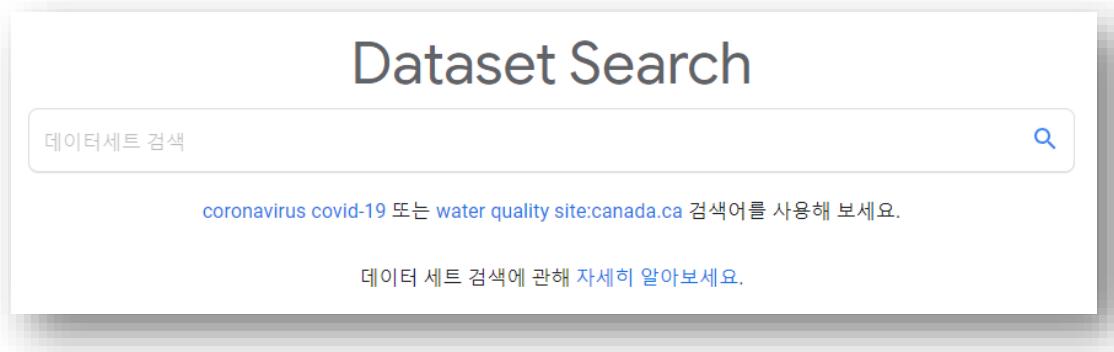
- Greenhouse gas giants** by Konrad Banachewicz (Updated 10 hours ago, Usability 10.0, 4 files, 2 MB) - Ranked 11th, Bronze level.
- Cataract Classification Dataset** by AkshayRamakrishnan (Updated 3 hours ago, Usability 8.8, 410 files, 20 MB) - Ranked 3rd.
- Food Demand Forecasting** by Shivani Rana (Updated 21 days ago, Usability 9.4, 5 files, 6 MB) - Ranked 14th, Bronze level.

On the right side of the page, there are buttons for "New Dataset" and "Your Work", and a "Relevance" dropdown menu.

# 빅데이터의 개념 및 특성

## ■ 데이터 사이트

- 구글 데이터셋 서치(<https://datasetsearch.research.google.com/>)



The screenshot shows a Google search results page for the query 'helmet'. The search bar at the top has 'helmet' in it. Below the search bar are several filter buttons: '최종 업데이트 날짜', '다운로드 형식', 'Croissant' (sorted by relevance), '사용 권한', '주제', '제공업체', and '무'. The main search results area displays three datasets:

- kaggle Helmet Detection**  
kaggle.com  
zip  
Jun 1, 2020에 업데이트됨
- osf.io HELMET dataset**  
osf.io  
Mar 2, 2024에 업데이트됨  
+ 추가 버전
- datasetninja.com Safety Helmet Detection Dataset**  
datasetninja.com  
kaggle.com  
Jan 23, 2024에 업데이트됨

On the right side of the search results, there's a sidebar for the first result, showing a large 'kaggle' logo, the dataset name 'Helmet Detection', '764 images belonging to 2 classes', a 'Kaggle | kaggle.com' button, and a download link 'zip(410125849 bytes)'.

# 빅데이터의 개념 및 특성

- 빅데이터 활용 사례 1
  - 주차장 공유 서비스 스타트업 “뉴틴”



비대면 월주차 공유 플랫폼

공공(6종)	금융 빅데이터 플랫폼(5종)
국토부 표준지 공시지가	유동인구(KT)
전국 주차장 정보	아파트 주차 정보(기종)
행정구역 분류	상권 지역별 업종별 데이터(부산광역시)
부산광역시 주차장 정보	읍면동·집계구별 가구별 인구수(부산광역시)
부산시설관리공단 실시간 주차 정보	읍면동·집계구별 연령 성별 인구수(부산광역시)
부산시설관리공단 주차시설 정보	

민간, 공공 데이터 11종 활용

# 빅데이터의 개념 및 특성

- 빅데이터 활용 사례 2
  - 의료 헬스케어 플랫폼 “닥터바이스”



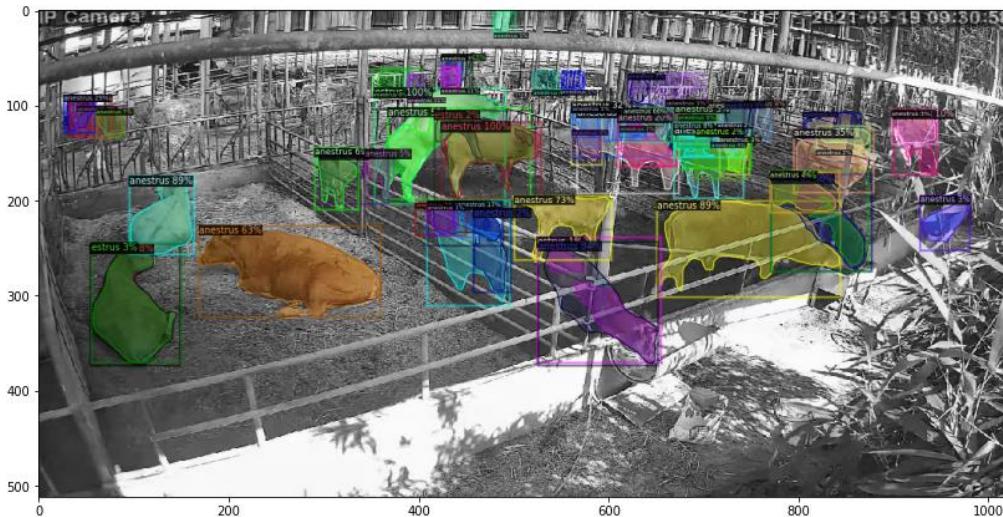
의료 데이터 통합 및 활용 플랫폼



외래·입원·개인·의사·병원·공공  
6개군 데이터 활용

# 빅데이터의 개념 및 특성

- 빅데이터 활용 사례 3
- 양돈 ICT 통합관리 솔루션



교배, 임신, 임신사고 분만 등 생산 정보 관리 시스템



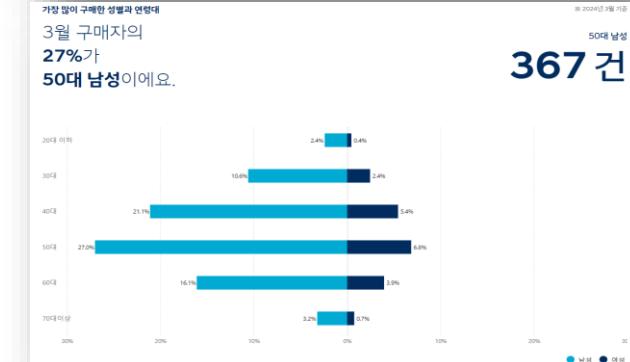
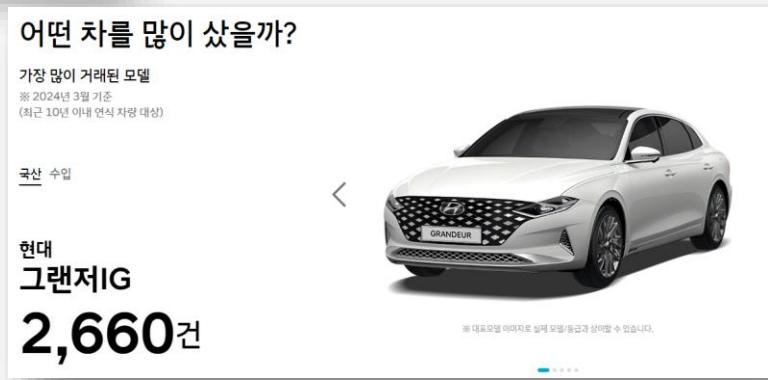
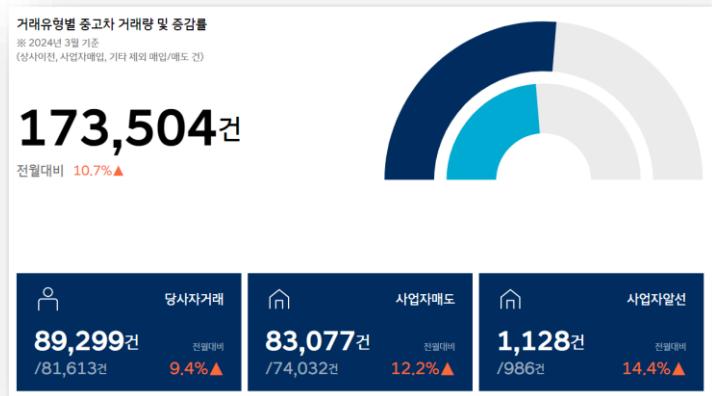
축산농가에 설치된 다양한 IoT 장비로 환경조건, 건강상태, 발정여부, 사료섭취량 등 데이터 활용

# 빅데이터의 개념 및 특성

## ■ 현대자동차 빅데이터 활용 사례

### ■ 중고차 통합정보 토클 하이랩(Hi-LAB)

- [시세 조회 결과 | Hi-LAB - 현대/제네시스 인증중고차 \(hyundai.com\)](#)



# 빅데이터의 개념 및 특성

- 현대자동차 빅데이터 활용 사례
  - 중고차 AI 가격산정 엔진

**차량에 대한 정보를  
불러왔어요.**

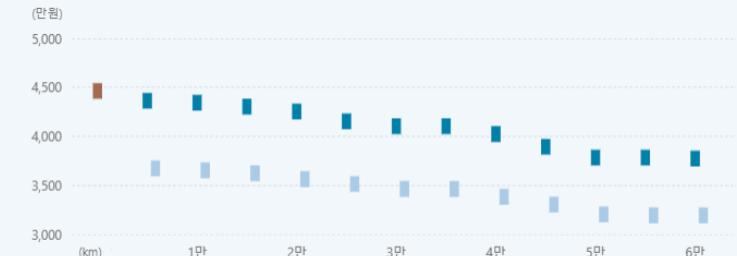


차량번호	206로5366	제조사/모델명	제네시스/G80
신차등급명	가솔린 2.5T 20인치 기본+ECS	최초등록일	2021.08.10
연식	2022년	주행거리	3,241 km <u>수정</u>
차종	대형차	연료	가솔린
배기량	2,497cc	변속기	오토
색상	태즈먼 블루	출고가격	69,300,000원

G80 5인승 가솔린 2.5T 2WD 20인치 기본+ECS의 시세는  
**4,390~4,540만원**입니다.

2020년 4월부터 수집된 누적 데이터 기반 동일 엔진/구동타입 중 주행거리 3,241km 차량의 거래가  
(2024.05.13 11시 07분 기준)

· 동일 트림 차량 중 ● 음선 포함 ■ 음선 비포함 ■ 조회 시세



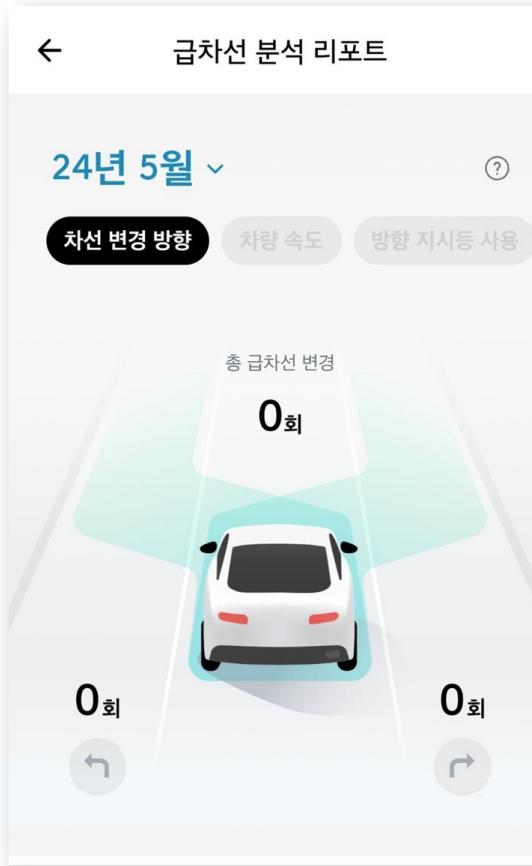
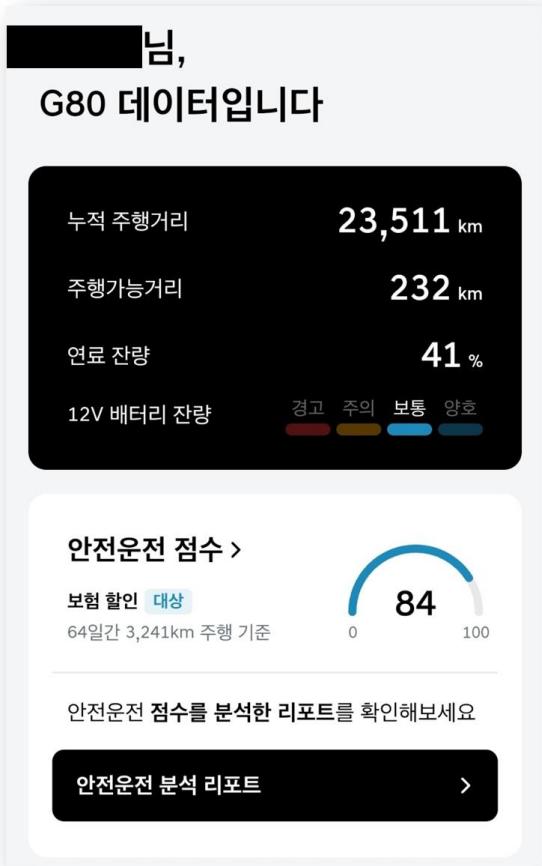
- 시세는 조회 시점으로부터 14일간 유효합니다.
- 방문평가 이후 사고유무 등 차량 상태에 따라 인내 받은 시세와 최종 견적이 다를 수 있습니다.

고객님의 차량에 적용된 선택 옵션 가격은  
총 920만원으로 현재 가치는 잔존 82%, 754만원으로 예상됩니다.

[옵션 상세정보 보기](#)

# 빅데이터의 개념 및 특성

- 현대자동차 빅데이터 활용 사례
- 모바일 차량관리 서비스



# 빅데이터의 개념 및 특성

- 현대자동차 빅데이터 활용 사례
  - 안전운전(UBI) 자동차 보험 할인할인 특약

