

# 6년 연속 베테랑 공급기업 크라우드웍스 데이터바우처 성공 사례집

고품질 데이터를 정부지원으로 저렴하게 구매할 수 있습니다.

# 전문 사업팀의 경험과 기술력을 바탕으로 한 1:1 맞춤 컨설팅을 제공합니다.

## 사업지원

### STEP 1



#### 1:1 컨설팅

전문 컨설턴트의  
전략 수립 및 사전 준비

### STEP 2



#### 서류·발표 심사

지원 서류부터  
심사까지 완벽 지원

### STEP 3



#### 프로젝트 수행

최적화된 작업자 매칭과  
전문 PM의 데이터 관리

### STEP 4



#### 감리 진행

데이터 품질 검증 및  
원활한 감리를 위한 지원

### STEP 5



#### 바우처 성공

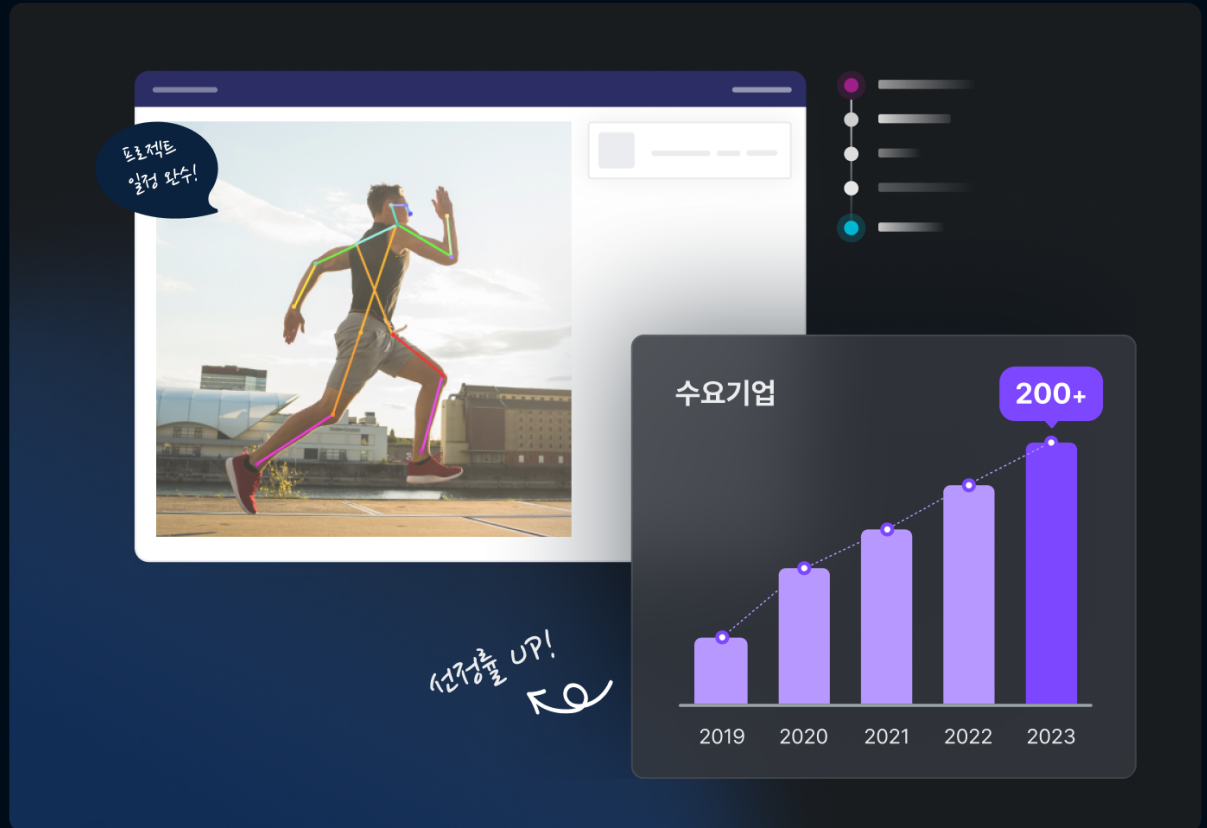
사업화 컨설팅부터  
우수사례 선정까지

## 사업수행

6년 연속 한국데이터바우처산업진흥원 공식 파트너

# 데이터바우처 누적 수행 경험 최다 보유

제조, 유통, 의료, 스마트 산업 등 50개 이상의 산업 분야와 이미지, 영상, 음성, 텍스트 등 전 데이터 분야를 아우르는 데이터바우처 프로젝트 경험을 보유하고 있습니다.

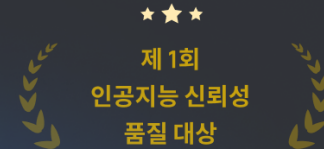


美 가트너, CB인사이트 선정 데이터 가공 글로벌 대표 기업

# 최고 품질의 데이터 제공

국내 최고의 품질 관리 프로세스를 보유하고 있으며,  
국내외 기관으로부터 데이터 품질과 기술 경쟁력을  
인정 받았습니다.

정부 기관으로부터 데이터 품질 대상을 수상하는 등  
데이터 품질을 보장합니다.



우수상



Gartner®



가트너, CB인사이트  
데이터 가공 대표 기업 등재



국내최초! 학습 데이터 플랫폼  
GS인증 1등급 획득



수요기업 평가 상위 20% 선정 기업 다수 배출

# 데이터바우처 다년 지원에 최적

2022년부터 데이터바우처 수행 결과 상위 20%에  
해당할 경우 여러 해 지원이 가능합니다.  
클라우드웍스와 함께라면 좋은 결과로 내년, 내후년  
지원사업의 혜택을 받을 수 있습니다.



데이터 정확도 99% 보장

# 180+개의 특허로 작업을 스마트하게

다양한 기술 특허를 통해 데이터 생산 프로세스를 고도화했습니다. 스마트 라벨링, 데이터 비식별화 등 클라우드웍스만의 AI 솔루션을 이용해 객체를 인식하거나, 데이터 전처리를 쉽고 빠르게 수행할 수 있습니다.



AI 선두기업과 파트너십

# 500+개 고객사의 선택

클라우드웍스는 IT, 금융, 에듀테크, 자율주행 등 다양한 산업군의 고객사를 보유하고 있습니다. 데이터bauer로 국내 유수의 고객사가 신뢰하는 고품질 데이터를 직접 경험해 보세요.

NAVER

kakaoenterprise

SAMSUNG

LG전자

kt

SK telecom

kakaobank

HYUNDAI

KAIST

# 200+개 수요기업의 성공에는 크라우드웍스가 함께하고 있습니다

크라우드웍스는 2019년부터 데이터bauer 지원사업의 시작을 함께한 베테랑 공급기업입니다.



데이터bauer 수요기업 성공 사례집에서는 수요기업이 지원 당시 겪고 있었던 상황에 대한 이야기와 더불어 크라우드웍스의 프로젝트 진행 과정, 상황별 데이터 솔루션, 성공 사례 분석 등이 담겨있습니다.

제조, 교육, IT 등 다양한 분야에서 크라우드웍스와 성공적인 프로젝트를 마친 수요기업들의 사례를 통해 2024년 데이터bauer 지원사업의 파트너사를 선택하는 데 많은 도움이 되길 바랍니다.

# [제조] 화장품 구독 서비스 기업 ‘스킨랩코리아’, ‘피부 분석 AI 모델’ 개발을 위한 330명 여성 피부 데이터 수집



SKIN & LAB

스킨랩코리아는 피부 데이터를 기반으로 28일 마다 새로운 화장품을 제공하는 구독 서비스 전문 기업으로 피부 분석 AI 모델을 개발하고 있었습니다.

## 공급 기업으로 클라우드웍스를 선택한 이유는?

특정 상황, 특정 연령, 성별의 데이터 등 대량 수집의 어려움을  
해결해 줬어요!

(국내 최대 58만 명 이상의 작업자 풀에서 20·30 여성 모집 원활)



다년간 유사 프로젝트 경험에서 나오는 노하우가 달라요!

(현장 촬영 전문 인력 모집 및 노쇼 인원 발생 시 이탈을 대비한 순차 모집 계획 등 제공)

민감한 데이터를 안전하게 보호해줘요! (철저한 개인 정보 보호 관리)

문의 및 요청에 대한 피드백이 빠르고 효율적이에요!

(프로젝트별 담당 PM 배정)

“남다른 노하우와 빠른 피드백으로 까다로운 데이터 수집도 완벽했어요!”

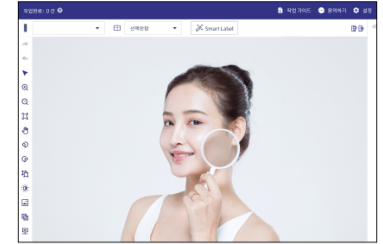
“클라우드웍스는 한번도 수집해 본 적 없는 형태의 사진 및 피부 데이터였지만 다양한 레퍼런스 수집을 통한 초기 기준 설정 및 수년간 쌓아온 노하우를 바탕으로 한 데이터 수집 등을 통해 상당히 까다로운 데이터를 수집 및 검수하였습니다. 무엇보다 빠른 피드백으로 사업 진행이 상당히 수월했으며 단순히 사업 진행을 위한 계약 관계가 아닌 수집 데이터에 진심으로 대하는 모습에 감사했습니다.”

## 수집 데이터

1) 20·30 여성 330명의 세 가지 타입 데이터

- 📷 피부 자가 진단 데이터
- 📷 피부 현미경 데이터
- 📷 피부 측정기를 통한 피부 데이터

2) 랜덤으로 부여된 개인 식별번호를 통해 세 가지 데이터 매핑

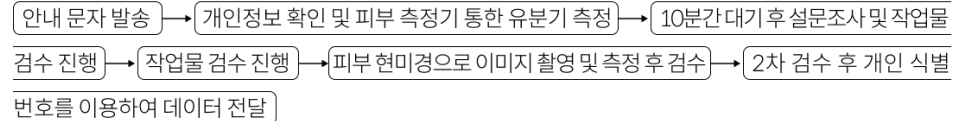


## 클라우드웍스의 솔루션

### 작업자 선발 과정

- 촬영 인원 모집 : 20·30 여성 300명 이상, 평일 저녁/주말 병행 촬영, 노쇼 비율 고려하여 스케줄링
- 프로젝트 현장 안내 인력 모집 : 유사 프로젝트 참여 경험이 있는 평일/주말 현장 안내 인력 채용, 사전 교육

### 프로젝트 진행과정



## 프로젝트 성과

### “샘플 데이터와 빠른 진행 상황 공유로 기간 내 데이터 수집 완료”

프로젝트 초기에 고객사에 샘플 데이터와 진척 상황을 수시로 공유했습니다. 샘플 데이터에 대한 고객사의 피드백을 빠르게 수령하여 니즈에 부합하는 데이터를 구축할 수 있었고, 원하는 일정 내 총 330명의 인원에 대한 피부 데이터 수집을 완료했습니다.

# [IT] AI 학습 어시스턴트 교육 플랫폼 '테크트리', AI 챗봇 학습 고도화를 위한 QA 데이터셋 생성



테크트리는 AI 챗봇 어시스턴트의 자연어 처리와 머신러닝 기술을 활용한 온라인 교육 플랫폼 "격차"를 운영 중입니다.

## 공급 기업으로 클라우드웍스를 선택한 이유는?

국내 최대 58만 명 이상의 작업자 풀로 자연어/챗봇 학습 데이터 관련 작업 및 검수가 가능한 보장된 인력을 제공받을 수 있어요!  
(다양한 프로젝트 경험과 검증된 인력 제공)

상세한 작업 기준 가이드 구축과 실시간 검수 모니터링으로 최적의 데이터를 제공해요!  
(신속한 오류 개선 및 높은 퀄리티 데이터 제공)

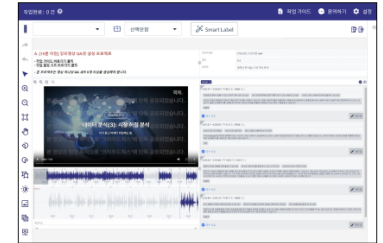


## “상세한 초기 가이드 및 담당자와의 원활한 소통으로 양질의 QA셋 데이터를 구축할 수 있었어요!”

“프로젝트 특성상 작업 기준이 다소 모호할 수 있는데 상세하게 정리된 초기 가이드를 제공해주셔서 이를 참고해 정해진 일정 내 프로젝트를 끝낼 수 있었습니다. 또한 온라인 강의 영상이라는 비정형 데이터의 특수성이 있어, 샘플 및 중간 작업물 검토 과정에서의 가이드 보완 작업이 중요했는데 담당자 간 원활한 소통으로 양질의 QA셋 데이터를 만들어낼 수 있었습니다”

### 수집 데이터

- 1) 강의 내용 요약한 수업노트 데이터셋 생성
- 2) QA 데이터셋 구축
- 3) 강의 음성을 전사한 스크립트 추출
- 4) 데이터 생성



### 클라우드웍스의 솔루션

#### 작업자 선발 과정

- 자연어/교육 프로젝트 경력자 선발 : 온라인 강의 콘텐츠의 구성과 강의 핵심 내용을 잘 파악할 수 있는 인력 모집

#### 프로젝트 진행과정

- 1) 강의 스크립트 중심의 수업노트 데이터셋 생성
  - STT(Speech-to-Text, 음성 인식) 엔진 활용 및 전사 정확도 검수 작업 병행으로 정확성과 신속성 확보
- 2) 작업자/검수자/담당 PM의 상시적인 소통을 통해 최적의 QA셋 태깅 생성
  - 작업 가이드와 예시를 활용하여 작업 방향성 강조

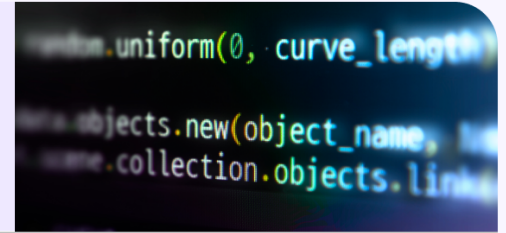
### 프로젝트 성과

#### “STT 엔진 활용과 검수 작업 병행으로 정확성과 신속성 확보”

강의 데이터별 수업노트 데이터셋은 스크립트 중심으로 정확한 전사 여부가 가장 중요해 STT 엔진 활용과 검수 병행으로 결과를 산출하며 프로젝트의 정확도와 신속성을 모두 확보했습니다. 또한 작업자, 검수자, 담당 PM 간의 상시 소통을 통해 작업 방향성을 강조하고, 이를 모든 분들이 숙지할 수 있도록 업데이트 사항을 수시로 공유했습니다. 이 결과, 소수의 인원으로도 데이터 품질의 일관성과 정확도를 확보할 수 있었고 최적의 QA셋 태깅을 생성할 수 있었습니다.



# [교육 · IT] 교육 AI 솔루션 기업 '엘리스', 코딩 학습 질의응답 대화 데이터셋 5만 건 라벨링



/\*elice\*/

엘리스는 교육 인공지능 솔루션 전문 기업으로 1:1 맞춤형 학습 문의 응대 서비스를 고도화 하기 위해 AI 튜터 서비스 'AI 헬피'를 개발 및 운영 중입니다.

## 공급 기업으로 클라우드웍스를 선택한 이유는?

교육 분야를 비롯한 다양한 산업군의 텍스트 데이터셋 가공 경험  
풍부해 걱정이 없어요!  
(코딩 학습 관련 질의응답 대화 데이터셋 라벨링 가능)

작업-검수 병행 및 실시간 검토로 피드백 개선이 빨라 오류율을  
최소화할 수 있어요!  
(효율적인 프로세스로 신속한 피드백 가능)

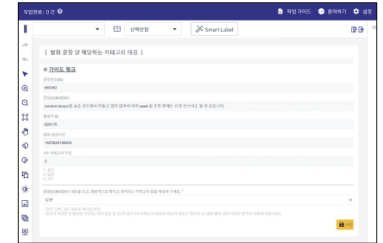


**“프로젝트 운영 경험이 풍부하고, 능숙하게 작업할 수 있는 인력이 보장되어 있어  
작업 진척도 & 태깅 품질을 만족할 수 있었어요!”**

“코딩 학습 플랫폼 내 튜터-수강생 간 대화 데이터를 가공하는 작업이므로, 학습자의 의도 파악 및 대화 맥락을 확인하여 질문-답변-단순 코드 정보로 잘 분류할 수 있는 스킬이 필요했습니다. 클라우드웍스는 다수의 유사 프로젝트 운영 경험이 풍부할 뿐 아니라, 이를 능숙하게 작업 및 검수할 수 있는 인력 풀이 보장되어 있었기 때문에 작업 진척도나 태깅 품질에 만족할 수 있었습니다.”

### 수집 데이터

- 1) 수강생 - 튜터 간 1:1 대화 데이터 제공
- 2) 데이터 가공
  - 대화 데이터 내 질문과 답변 분류
  - 학습자가 제출한 코드 영역 구분

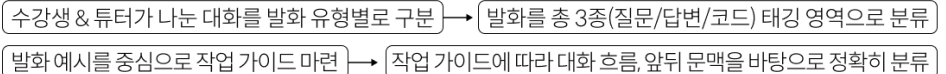


### 클라우드웍스의 솔루션

#### 작업자 선발 과정

- 상담 및 학습 대화 데이터 가공 프로젝트 경력자 선발: 유관 데이터 가공 경험이 있거나 실제 코딩 학습 플랫폼 사용 경험 등 가이드 및 대화 내용을 잘 이해할 수 있는 작업자로 선발

#### 프로젝트 진행과정

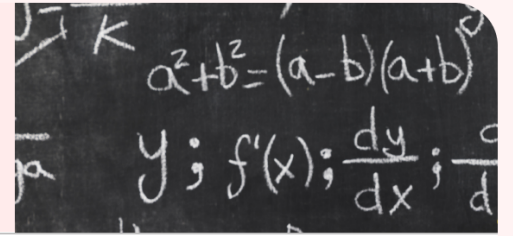


### 프로젝트 성과

#### “발화 유형별 적정 속성값 태깅 작업으로 고품질의 텍스트 데이터셋 제공”

코딩 학습 플랫폼 내 수강생-튜터 대화 데이터를 활용하여, 대화의 문맥과 흐름에 따라 질문과 답변을 구분하는 태깅 작업을 진행했습니다. 이를 위해 예시 중심의 작업 가이드를 작성하고, 모호한 사례는 PM의 판단과 피드백을 통해 작업을 완료했습니다. 그 결과, 플랫폼 고도화를 위한 학습 데이터가 적절하게 라벨링 되어 전달되었습니다.

## [교육 · IT] 수학자료 인식 OCR 솔루션 기업 '아이텍솔루션' 이미지 수집 및 수식/텍스트 구분 라벨링 데이터 수집



iTeXSolution

인공지능 딥러닝 기술을 활용한 교육 콘텐츠의 광학문자인식(OCR) 솔루션 개발 기업으로 수학교육에서 인식이 어려운 수식 및 그래프 등을 정확하게 인식 해주는 'Qn.AI' 서비스를 운영 중입니다.

### 공급 기업으로 클라우드웍스를 선택한 이유는?

수학 문제 솔루션 관련 작업자 확보가 용이해요!  
(자격 시험을 통해 검증된 작업자 선발 가능)

데이터 수집부터 가공까지 체계적인 운영으로 일정 지연  
걱정이 없어요!  
(수학 문제 필기 데이터 수집 & 가공 병행 가능)

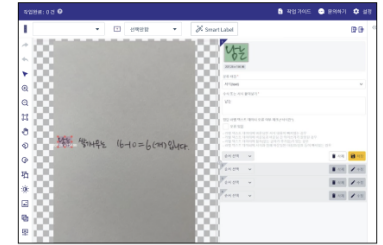


**“원활한 의사소통과 신속한 가이드로 기간 내 수학 문제 손글씨 데이터 수집/가공을  
마칠 수 있었어요!”**

“한정된 기간 내 수학 문제 손글씨 필기 데이터를 수집함과 동시에, 해당 데이터를 가공하는 작업을 병행하는 프로젝트였는데요. 원활한 의사소통과 신속한 작업 가이드 마련으로 일정상 큰 무리 없이 수행할 수 있었습니다. 작업물에 발생한 오류 사항에 대한 피드백도 적극적으로 반영 및 개선할 수 있어 긍정적인 협업이었습니다.”

### 수집 데이터

- 1) 수식 + 텍스트(한글 및 영문) 포함된 필기로 작성된 이미지
- 2) 수식/텍스트 구분(태깅) 후 라벨링한 데이터

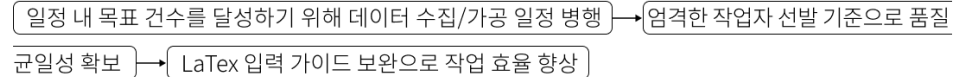


### 클라우드웍스의 솔루션

#### 작업자 선발 과정

- 1) LaTeX 문법 이해도 파악
  - 자격시험을 통해 1차적인 이해도 확인
  - 유사 프로젝트 경험이 있는 작업자 위주로 선발
- 2) 작업 품질 향상을 위한 작업자 선발
  - 정확도 판단 기준이 모호한 수학 문제 손글씨 수집 작업 특성을 고려
  - 고객사 기준에 맞는 작업자를 최소한으로 선발

#### 프로젝트 진행과정



### 프로젝트 성과

**“LaTeX 입력 가이드 보완으로 작업 효율 향상! 6,000여 건의 가공 데이터 추가 제공”**

고객사와의 긴밀한 소통으로 수학 문제 손글씨 이미지 수집 기반 데이터 전달 일정을 확인하며 목표 건수를 달성하기 위해 수집/가공을 병행했습니다. 작업자 선발은 LaTeX 문법에 대한 이해와 유사 경험을 고려하여 이루어졌으며, 작업 진행 중에는 빠른 피드백을 통해 LaTeX 입력 가이드를 보완하여 목표 건수 보다 6,000여 건의 가공 데이터를 추가해 전달할 수 있었습니다.



## [IT] 교육 콘텐츠 에듀테크 기업 '로보그램', 코딩교육 콘텐츠에 적용할 이미지 데이터셋 구축

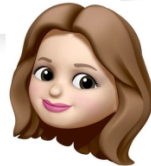


융합 알고리즘을 일반인이 쉽게 이해하고 접근할 수 있게 해주는 교육 콘텐츠를 개발하는 에듀테크 전문기업입니다.

### 공급 기업으로 클라우드웍스를 선택한 이유는?

짧은 기간 내 5만 장 이상의 다양한 각도의 사물 이미지 수집이  
가능해요!  
(각도가 중요한 까다로운 이미지 데이터를 빠르게 수집 가능)

작업자 이탈을 고려한 상품 리스트 5배수 모집으로 일정 지연  
걱정이 없어요!  
(작업자 이탈에 관한 프로세스 보유)

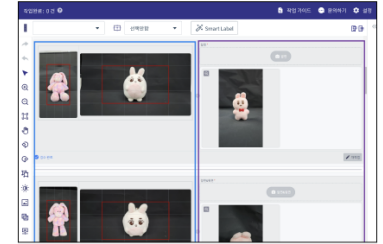


**“까다로운 수집에도 불구하고 고품질의 이미지 데이터를 전달받았고, 다음에도  
클라우드웍스의 데이터를 적극적으로 활용하고 싶어요!”**

“이미지 데이터 수집 과정에서 규칙이 엄격해 수집이 까다로웠지만, 질 높은 데이터를 제공받아 프로젝트를 원활하게 완료할 수 있었습니다. 다음에도 기회가 된다면, 다시 클라우드웍스의 데이터를 적극적으로 활용하고 싶습니다”

### 수집 데이터

- 1) 전자기기, 사무용품, 잡화 등의 다양한 카테고리의 이미지 수집
- 2) 카테고리별 6~18개 각도의 이미지 수집
- 3) 카테고리별 최소 1천 장의 이미지 수집



### 클라우드웍스의 솔루션

#### 프로젝트 진행과정

##### 1) 상품 리스트 선정

- 중복을 피해 수집 진행이 필요한 프로젝트
- 카테고리별로 작업자에게 1~5개 상품 정보 제출 요청
- 작업자 이탈 및 상품 파손 고려, 5배수 모집 후 중복 정제 진행
- 최종 선발된 작업자에게 참여 안내 및 1배수가 부족한 카테고리의 경우 수요기업과 대체 협의

##### 2) 작업자 선발

- 사전 조사 참여 작업자 중 촬영 및 제출 가능 작업자 선발 (중복 불가, 1상품 1작업자)

##### 3) 수집

- 데이터 품질 일관성을 위해 카테고리 & 각도별 올바른 촬영 예시 제공
- 고난이도 촬영으로 작업자 이탈 시 작업 난이도 조절 및 카테고리 변경 진행

### 프로젝트 성과

**“짧은 기간 내 다양한 각도의 사물 이미지를 중복 없이 6만 건 이상 수집”**

최대한 중복을 피해 5만 장 이상 구축이 필요했던 프로젝트로 샘플 데이터와 진척 상황을 고객사에 공유하고 피드백을 반영해 니즈에 부합하는 데이터 51종, 62,822장을 수집했습니다.