

데이터 조회 프로그램

데이터 조회 프로그램이란?

- 프로젝트에 필요한 데이터들을 저장, 조회하고 필요에 따라서 수정도 할 수 있는 프로그램

수집한 데이터의 정보

- 데이터 크기

All_category(meta+review) : 376GB

Meta_All_Beauty : 213MB

Review_All_Beauty : 327MB

Category	#User	#Item	#Rating	#R-Token	#M-Token
All_Beauty	632.0K	112.6K	701.5K	31.6M	74.1M

- 데이터 타입

- json

프로그램의 예시

- 파이썬
- R
- Mysql
- Excel

각 프로그램의 이점

	핵심 기능	상세 기능	장점	단점
파이썬	데이터 분석 및 처리	유연하고 강력한 프로그래밍 언어. 데이터 분석, 시각화, 머신러닝, 인공지능 등 다양한 분야에서 널리 사용.	직관적인 문법 / 풍부한 라이브러리와 프레임워크 / 다양한 분야에서 활용 / 높은 플랫폼 호환성	실행 속도가 느린 편 / 많은 메모리 소비
R	데이터 분석 및 처리	통계 분석 및 데이터 시각화에 특화된 프로그래밍 언어. 통계 분석 기능, 통계적 추론, 시뮬	통계적 분석에 강점 / 10,000개 이상의 다양한 패키지 / 뛰어난 시각화 기능	프로그래밍, 통계, 수학에 대한 기본적인 해가 필요

		레이션, 머신러닝 등 다양한 작업을 수행.		
MySQL	데이터 분석 및 처리	SQL을 사용하여 데이터를 관리하는 데이터베이스 관리 시스템. SQL을 활용하여 데이터를 추출, 필터링, 집계하는 등 다양한 작업을 수행.	무료 및 오픈 소스 / 빠른 처리 속도와 높은 성능 / 다양한 운영 체제 호환성	복잡한 설정 및 관리
Microsoft Excel	데이터 분석 및 처리	사용자 친화적인 인터페이스와 다양한 기능을 제공하여 데이터 분석 및 시각화를 손쉽게 할 수 있는 사무용 프로그램. 대부분의 회사원들이 익히고 사용하는 프로그램.	사용자 친화적인 인터페이스 / 다양한 기능 제공 / 포매팅 및 조건부 서식	대용량 데이터 처리 어려움 / 복잡한 분석 작업 제한

mysql을 선정합니다.

MySQL을 사용하는 이유

- **현재 시스템과의 호환성** - Django와 호환성이 뛰어나 바로 사용 가능함.
- **JSON 형식의 데이터 처리 지원** - JSON 데이터를 네이티브로 지원해 변환 과정 없이 저장·조회 가능함.
- **확장 가능성** - 데이터 증가 시에도 성능 유지하며 확장 가능함.
- **성능과 안정성** - 성능과 안정성이 입증된 관계형 데이터베이스임.
- **미래 확장성** - 추가 기능이나 복잡한 처리에도 유연하게 대처 가능함.

MySQL에 들어가는 형식

meta							
item 데이터							
Num	PK	AI	FK/Null	Name	Type	Default	Comment
1	✓	✓		meta_id	INT	0	sql에 저장시에 사용되는 pk
2				main_category	VARCHAR(50)		item의 메인 카테고리
3				title	VARCHAR(50)		상품의 이름
4				average_rating	DOUBLE		상품의 평균 리뷰점수
5				rating_number	INT		상품이 리뷰된 횟수
6			✓	features	VARCHAR(50)		상품의 특징
7			✓	description	VARCHAR(50)		상품의 설명
8			✓	price	VARCHAR(50)		상품의 가격
9			✓	images	VARCHAR(50)		상품의 이미지
10			✓	videos	VARCHAR(50)		상품의 영상
11				store	VARCHAR(50)		상품을 파는 상점 이름
12			✓	categories	VARCHAR(50)		상품의 세부 카테고리
13			✓	details	VARCHAR(50)		상품의 세부사항 (상품의 크기 같은 것) (parent_asin도 여기 안에
14				status	INT	0	0: 살아있음 1: 삭제됨
15			✓	deleted_time	TIMESTAMP		상품이 삭제된 시각

review							
리뷰 데이터							
Num	PK	AI	FK/Null	Name	Type	Default	Comment
1	✓	✓		review_id	INT	0	시트 자동 증가되는 key 값
2				rating	DOUBLE		0.0~ 5.0까지의 유저가 아이템에 남기는 리뷰 점수
3				title	VARCHAR(100)		리뷰의 제목
4				text	VARCHAR(500)		리뷰 내용
5			✓	images	VARCHAR(500)		리뷰에 사용된 image
6			✓	asin	VARCHAR(15)		리뷰하는 아이템의 id
7				parent_asin	VARCHAR(15)		리뷰하는 아이템의 부모 id
8				user_id	VARCHAR(30)		리뷰하는 유저의 id
9				timestamp	VARCHAR(20)		리뷰가 작성된 timestamp(unixtime)
10			✓	helpful_vote	INT		다른 유저가 도움이 되었다고 누른 횟수 0~
11			✓	verified_purchase	VARCHAR(10)		구매 확인
12				status	INT	0	0: 살아있음 1: 삭제됨
13			✓	deleted_time	TIMESTAMP		삭제 처리 된 시각

- 원래 json값을 최대한 살리고 삭제를 요청하더라도 Delete가 아닌 Update로 status만 변환 해줘서 나중에 다시 사용할 수 있게 처리함, 삭제된 시각도 같이 표현해줌