



백개의 리뷰

: 전자제품 리뷰 총정리 서비스

Team 09

김정국 (데이터사이언스학과)
신성구 (데이터사이언스학과)
이준규 (데이터사이언스학과)
조유진 (데이터사이언스학과)
허상우 (데이터사이언스학과)

Motivation

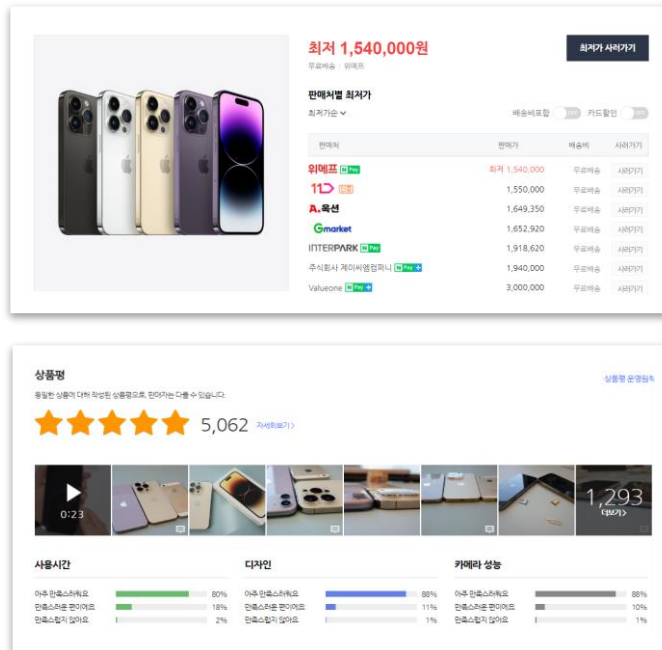


01. Motivation

When purchasing an electronic product,
we take various factors into consideration



01. Motivation



- The reviews provided by other buyers are very helpful for our consumption decision
- However, review data exists separately for each distribution platform, and it takes a lot of time to look it up.
- It will be very helpful for consumers to make decisions if they can see reviews from different platforms in one setting.

Let's gather all the reviews of each product in one page and show them!



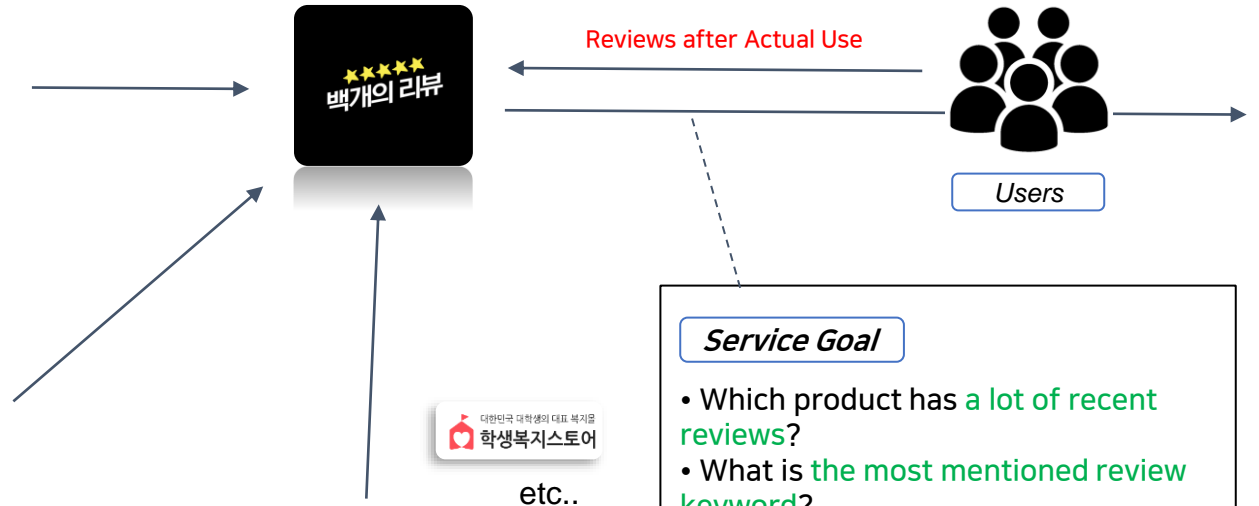
The background illustration features a large, light-blue rectangular screen on a dark blue stand. On the screen, there are four yellow stars arranged in a horizontal row. A man in a red long-sleeved shirt and dark blue pants stands to the left of the screen, gesturing with his right hand towards the stars. A woman in a dark blue long-sleeved shirt and a red skirt stands to the right, reaching up to place a fourth yellow star on the screen. The screen itself has some faint, abstract shapes: a horizontal bar at the top, a horizontal bar below the stars, and a white horizontal bar at the bottom. Surrounding the screen are various decorative elements: a heart icon, a speech bubble with a checkmark, and a speech bubble with three dots. Dashed lines connect these elements to the screen and the people.

Project Scenario

02. Project Scenario - Overview

The screenshot shows the Coupang homepage with various product listings. At the top, there's a navigation bar with '상품상세', '상품별 (1,204)', '상품문의', and '배송/교환/반품 안내'. Below this, there's a section for '온라인' with a star rating and a list of products. The main content area features several product cards, including an Apple MacBook Air (2022) and a Samsung Galaxy S23 Ultra. Each card displays the product name, price, and a '구매하기' button. Below the product cards, there's a section for '인텔 맥북 프로 16인치' with a detailed description and a '구매하기' button. The bottom of the page shows a '배송' section with a map and a '배송지' dropdown menu.

"Review everywhere everything all at once"



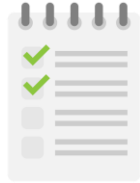
Service Goal

- Which product has a lot of recent reviews?
- What is the most mentioned review keyword?
- How to decide on each review to be a positive vs. negative?
- How to minimize the impact of advertising reviews?

작성한 실사용 리뷰 13

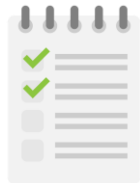
The screenshot shows a user review for a Samsung PM2.5S sensor. The review is titled '사용의 비헤이' and is dated 2023.12.14. The user has a rating of 4 stars. The review text mentions that the sensor is easy to use and provides accurate readings. There is a photo of the sensor and a link to the product page.

02. Project Scenario - Requirement & Assumption



Assumption

- 🤔 We need a lot of quality reviews, but it is very difficult to collect review data from all different online shopping malls and reviews sites in reality.....
- 😬 Naver Shopping and Coupang data mainly used for review data collection 😬



Requirement

- ✓ Using Python to crawl review data
- ✓ Integrating reviews from multiple data sources into a single database and linking them to MySQL
- ✓ Extract the necessary review data from such an integrated DB and use it for Wordcloud, positive/negative prediction, and keyword inquiry.

02. Project Scenario - Technical Challenges



It was difficult to organize a review database from a web page source code with only a simple code, so we borrowed webscraping models released to the online community such as GitHub.



As reviews were written freely regardless of grammar, it was difficult to apply grammatical rules to standard preprocessing for NLP..



It was difficult to extract sentence-unit data because there were many cases where the proper grammar was not used.

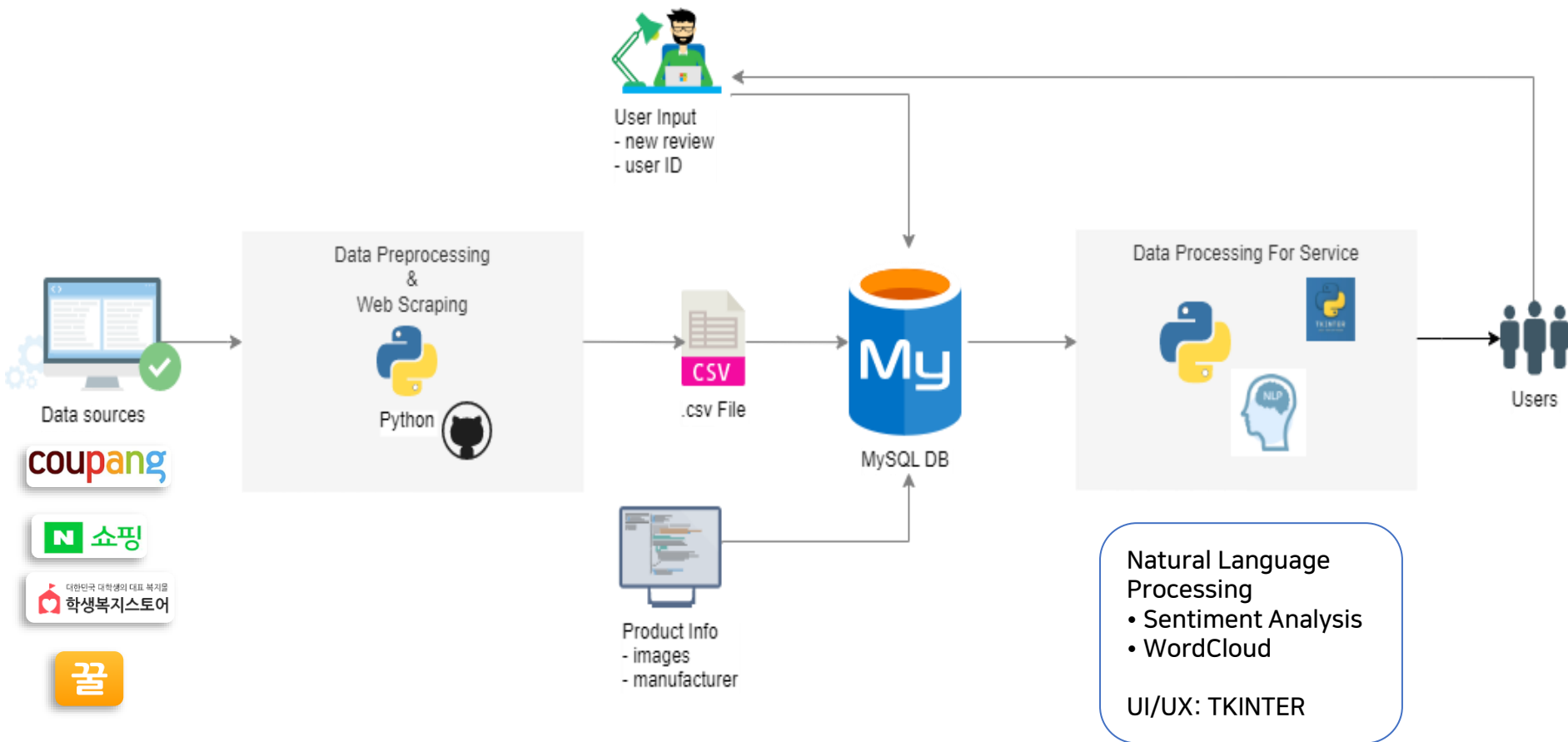


An illustration featuring two stylized human figures interacting with a large, light-blue rectangular screen. The screen displays four yellow five-pointed stars arranged in a horizontal row. Above the stars is a solid blue horizontal bar, and below them is a thin blue horizontal line. The screen is mounted on a dark blue pedestal. To the left of the screen, a man in a red long-sleeved shirt and dark blue trousers stands with his right hand on his hip and his left hand pointing towards the screen. To the right, a woman in a dark blue long-sleeved shirt and a red skirt is reaching up with both hands to place or adjust the fourth star. The background is a solid grey. Various decorative elements are scattered around: a blue heart icon in the upper left, a speech bubble with a checkmark in the middle left, a speech bubble with three dots in the upper right, and another speech bubble with a checkmark in the middle right. Dashed lines connect some of these icons to the screen. At the bottom of the screen, there are three small blue dots and a white horizontal brushstroke.

Architectural Design

03. Architectural Design

Project Architecture



An illustration of two people, a man and a woman, interacting with a large digital screen. The man on the left is wearing a red sweater and blue pants, giving a thumbs up. The woman on the right is wearing a blue top and red skirt, placing a yellow star on the screen. The screen displays four yellow stars in a row. There are also speech bubbles with checkmarks and a heart icon floating around the screen. The text "CREATE Team 09.DB" is overlaid in the center.

CREATE Team 09.DB

04. Database - Which Products?



Seven products from three manufacturers:
Samsung, LG, and Apple



Five products from two manufacturers:
Samsung and Apple



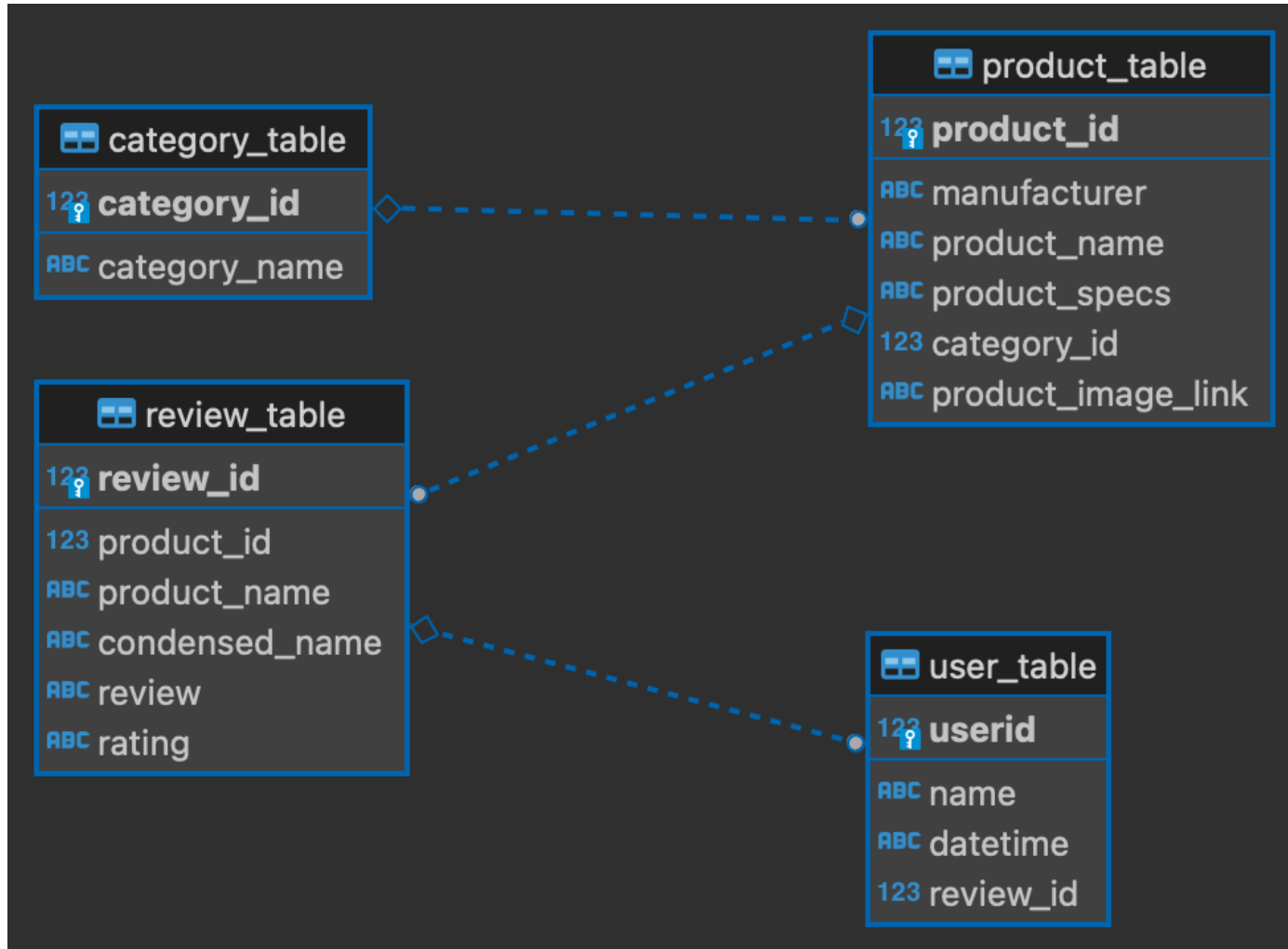
Four products from two manufacturers:
Samsung and Apple

Data primarily consolidated into a single spreadsheet and saved as csv files, which are then configured to work with MySQL

	A	B	C	D	E	F	G	
1	상품명	구매자	구매처	필리프	제품리뷰	내용		
2	Apple 202	보라돌이	5	애플이 야만병처럼요, 최근 해외 장기	그래픽/전문가용			
3	Apple 202	나정원	5	개발자용 가성비 좋은 아수스 노트북	그래픽/전문가용			
4	Apple 202	김*철	5	등록된 하자는 핸드폰이 아이폰을 사용	그래픽/전문가용			
5	Apple 202	다사즈아	5	그냥 사서 20여 년 동안 외길만 걸어온	사무용			
6	Apple 202	고끼랑	5	서프라이즈에 계시는 대장님이 맥북을	학습용			
7	Apple 202	김*주	5	m2맥북프로인텔맥북에어 2020 사용하다	기타 용도			
8	Apple 202	김성수	5	512gb 24" 가전제품은 이제 쿠팡이 답인	그래픽/전문가용			
9	Apple 202	테이스티	5	할인할때 * 구매 일자 : 2022. 6. 29.선물	사무용			
10	Apple 202	오팔삼칠	5	기대 이상 실리곤 프로세서용 제품은	사무용			
11	Apple 202	고*열	5	노트북은 쿠팡 사랑합니다.10코어, 24g	그래픽/전문가용			
12	Apple 202	빛나염마	5	원도우에 갓, 맥북 상태 아무이상없	학습용			
13	Apple 202	김*복	5	개발자이고 업무용으로 구입	사무용			
14	Apple 202	연상호	5	맥린이 여 대학교 입학할때 13년도	사무용			
15	Apple 202	김*철	5	가성비 좋구형 디자인의 맥북에 M2	그래픽/전문가용			
16	Apple 202	김*학	5	등록된 허맥북에어살가 고민하다가	이 학습용			
17	Apple 202	최*승	5	앱개발 용클린텐 24메모리 m2 앱개발	그래픽/전문가용			
18	Apple 202	김*훈	5	기존 m1 2022년 10월구매 m2 pro 512g	그래픽/전문가용			
19	Apple 202	이*도	5	등록된 허M1 맥북프로 13인치 8기가	사무용			
20	Apple 202	홍*모	5	등록된 허SNS에서 새로나온 맥북에어	사무용			
21	Apple 202	서*해	5	M1 pro랑 그날 그날 할인용, 가격이	다르 그래픽/전문가용			
22	Apple 202	nana02	5	등록된 허아이 학원다니면서 쓰려고	구 사무용			
23	Apple 202	황*나	5	가법계 쓸사무용 학습용이라 용량이나	학습용			
24	Apple 202	서*수	5	이래서 맥맥북 처음써보는 사람이고	코 그래픽/전문가용			

[illegible]

04. MySQL Entity Relationship Diagram



Entity Relationship Diagram of four tables with unique primary keys and foreign keys

Database Workloads

An illustration featuring a large, light-blue rectangular screen on a dark blue stand. The screen displays four yellow stars in a horizontal row. Overlaid on the screen is the text 'Database Workloads' in a large, white, sans-serif font. To the left of the screen, a man in a red long-sleeved shirt and dark blue pants stands with his hands on his hips, gesturing towards the screen. To the right, a woman in a dark blue long-sleeved shirt and a red skirt is reaching up to place a yellow star on the screen. The background is a solid dark grey. Various decorative elements are scattered around: a heart icon, a speech bubble with a checkmark, and a three-dot menu icon are on the left; a speech bubble with three dots and another with a checkmark are on the right. Dashed lines connect some of these icons to the screen area.

05. Database Workloads

OLTP(Update DB when our user added a new review)

: INSERT INTO user_table, review_table



The screenshot shows a web browser window with a form titled '리뷰를 작성해주세요' (Please write a review). The form contains the following text: '제품명: 맥북 에어' (Product name: MacBook Air), '사용자 ID: 32442' (User ID: 32442), and '리뷰 작성일시: 2022-12-08 12:49' (Review creation time: 2022-12-08 12:49). Below this is a large text input area with a placeholder '리뷰를 작성해주세요.' (Please write a review.) and a small blue button labeled '작성' (Write). The text '리뷰 작성하기' (Write review) is visible at the bottom of the form.

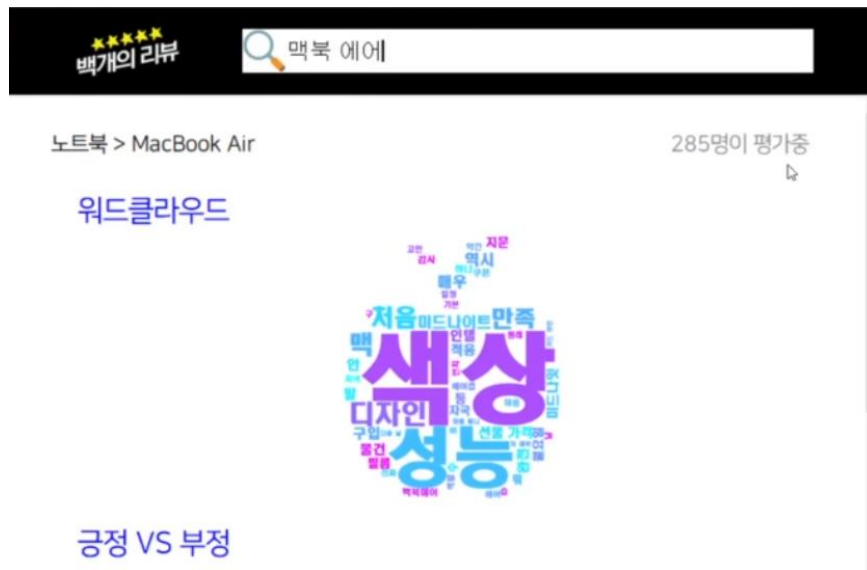
• *INSERT INTO teamdb9.user_table VALUES (35, 'Sangwoo Heo', '2022-12-12', 11074, '맥북 에어');*

• *INSERT INTO teamdb9.review_table VALUES (7, 'Apple 2022 맥북 에어, 실버, M2 8코어, GPU 8코어, 256GB, 8GB, 30W, 한글', '맥북 에어', '쿠팡답게 칼배송입이다 미드나잇과 스그중 무엇을 살까 고민하다 미드나잇을 케이스 씌우고 쓰기로했습니다 지문때문에 케이스는 필수이며 실물은 무지무지 이쁩니다 미드나잇 쓰실분은 꼭 케이스도 쓰시길..', 5);*

05. Database Workloads

OLAP

: f'SELECT r.review FROM review_table r WHERE r.condensed_name = {search}'



- When a product name is entered, the product review data is processed to enable word cloud, positive/negative classification, keyword classification, and YouTube link search

05. Database Workloads

OLAP

: SELECT r.condensed_name, r.review FROM review_table r JOIN user_table u ON u.review_id = r.review_id ORDER BY u.datetime DESC LIMIT 7

- View the 7 most recently written reviews

The screenshot shows a web interface for a Korean e-commerce site. At the top, there's a search bar with a magnifying glass icon and the text '백개의 리뷰' (100 reviews). Below the search bar, there's a section titled '지금 HOT한 제품' (Currently HOT products). This section displays three products: LG 울트라 PC (LG Ultra PC) with 364 reviews, MacBook Air with 285 reviews, and Galaxy Buds 2 with 144 reviews. Each product has a small image and a gold medal icon. Below this section, there's a red-bordered box titled '최근 등록된 리뷰' (Recently registered reviews). This box contains a list of reviews for various products, including '갤럭시북 플렉스' (Galaxy Book Flex), '갤럭시북 이온' (Galaxy Book Ion), '갤럭시북2 프로' (Galaxy Book2 Pro), '갤럭시북2 프로 그램' (Galaxy Book2 Pro Gram), and '갤럭시북 이온' (Galaxy Book Ion). Each review entry includes the product name and a short snippet of the review text.

지금 HOT한 제품

- LG 울트라 PC
등록된 리뷰 수: 364
- MacBook Air
등록된 리뷰 수: 285
- Galaxy Buds 2
등록된 리뷰 수: 144

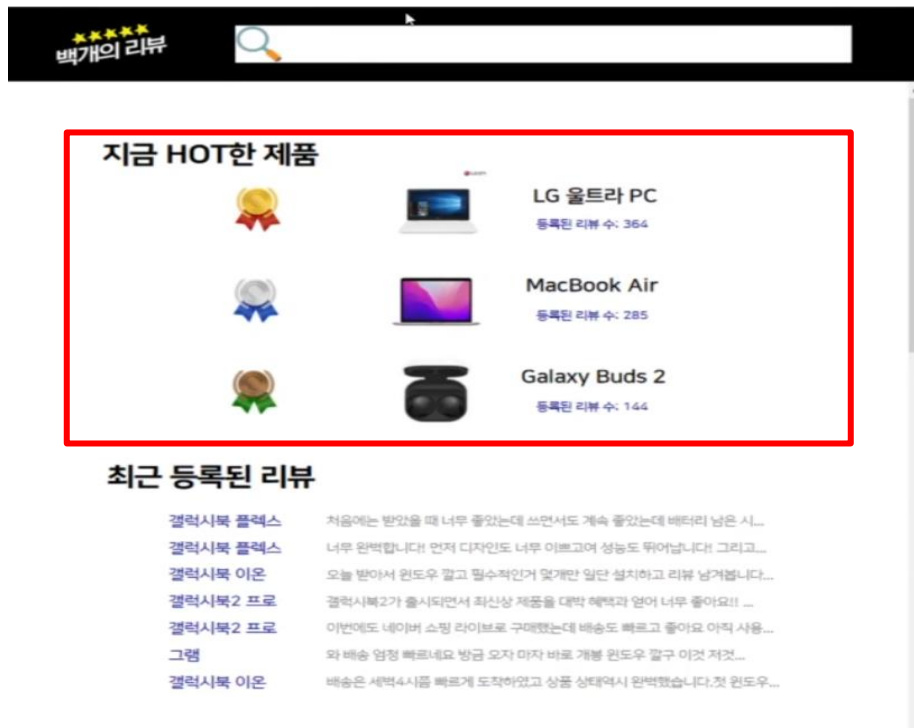
최근 등록된 리뷰

- 갤럭시북 플렉스: 처음에는 받았을 때 너무 좋았는데 쓰면서도 계속 좋았는데 배터리 남은 시...
- 갤럭시북 플렉스: 너무 완벽합니다! 먼저 디자인도 너무 이쁘고요 성능도 뛰어납니다! 그리고...
- 갤럭시북 이온: 오늘 받아서 윈도우 깔고 필수적인거 몇개만 일단 설치하고 리뷰 남겨봅니다...
- 갤럭시북2 프로: 갤럭시북2가 출시되면서 최신상 제품을 대박 혜택과 얻어 너무 좋아요!! ...
- 갤럭시북2 프로: 이번에도 네이버 쇼핑 라이브로 구매했는데 배송도 빠르고 좋아요 아직 사용...
- 그램: 와 배송 엄청 빠르네요 방금 오자마자 바로 개봉 윈도우 깔구 이것 저것...
- 갤럭시북 이온: 배송은 새벽4시쯤 빠르게 도착하였고 상품 상태역시 완벽했습니다. 첫 윈도우...

05. Database Workloads

OLAP

: SELECT r.condensed_name, p.product_image_link, COUNT(r.review_id) AS rcnt FROM review_table r JOIN user_table u ON u.review_id = r.review_id JOIN product_table p ON p.product_id = r.product_id GROUP BY r.condensed_name ORDER BY rcnt DESC LIMIT 3 WHERE u.datetime > (Last day of Last month)



- View the product names, images, and the number of registered reviews for the top three products with the highest number of reviews in the last month

An illustration featuring a large, light-blue rectangular screen on a dark blue stand. The screen displays a rating system with four yellow stars in a row. A man in a red long-sleeved shirt and dark blue pants stands to the left of the screen, gesturing towards it with his right hand. A woman in a dark blue long-sleeved shirt and a red skirt stands to the right, reaching up to place a fourth yellow star on the screen. The background is a solid grey. Various decorative elements are scattered around: a heart icon, a speech bubble with a checkmark, and a three-dot menu icon are on the left; a speech bubble with three dots and another with a checkmark are on the right. Dashed lines connect some of these icons to the screen. The text 'Data Modeling' is centered over the screen in a large, white, sans-serif font.

Data Modeling

06. Data Modeling



☺ Keyword Classification

- Explore the data to select only those reviews that mention color, size, and price within the sentence after a sentence-by-sentence tokenization of the Natural Language Toolkit (NLTK) package

```
sent_lst = sent_tokenize(row)
for sent in sent_lst:
    words = sent.split(' ')
    for word in words:
        if word in size_words:
            result.append(sent)
return result[:5]

print(size_extract(review_db1))
print(size_extract(review_db2))
```



['충전기 자체가 너무 크기 때문에 완충해서 나가면 오랜시간 사용할 수 있어서 노트북만 들고 나가면 될 것 같아요!']

06. Data Modeling

☹ Sentiment analysis

- Data is preprocessed through the process of removing all but Korean alphabet using regular expressions.
- Use the morpheme analyzer Mecab to perform tokenization.
- The frequency of each word is calculated under two cases: which words appear mainly in positive reviews and which words appear mainly in negative reviews.
- The data then get integer encoded and classified using LSTM.

	상 품 명	구 매 자 이 름	구 매 자 평 점	리 뷰 제 목	리 뷰 내 용	긍 정 부 정	긍 정 부 정 확 률
0	Apple 2022 맥북 에어, 실버, M2 8코어, GPU 8코어, 256GB, ...	이*현	5.0	아이패드 짝궁	당근에서 미개봉 구하다가 실버 색상은 없어서 쿠팡에서 구매했습니다와우 회원 쿠폰 쓰...	1	0.894794
1	Apple 2022 맥북 에어, 미드나이트, M2 8코어, GPU 8코어, 256G...	한*기	4.0	맥북병은 맥북을 사야 낫죠. 완치했습니다.	- 구매 계기20대부터 맥북을 써보야지 했는데 매 번 윈도우 노트북을 사용하다 보니...	1	0.944754
2	Apple 2022 맥북 에어, 미드나이트, M2 8코어, GPU 8코어, 256G...	강*중	5.0	와씨... 완전 개 이쁨...	8월 5일 배송예정이었는데, 8월1일 배송받았어요! 압도적감사!!!미드나이트 컬러로...	1	0.957828
3	Apple 2022 맥북 에어, 미드나이트, M2 8코어, GPU 8코어, 256G...	김*현	5.0	등록된 헤드라인이 없습니다	M1 제품이 너무 저렴하게 잘 나왔다보니 상대적으로 평가가 애매한 M2 제품입니다....	1	0.815718
4	Apple 2022 맥북 에어, 스페이스그레이, M2 8코어, GPU 8코어, 25...	김*린	5.0	고민말고 GO!! 완전만족~~ 역시 애플	얼마전 애플 스토어에 방문했다가 넘 이뻐서 고민고민 끝에 구입했어요ㅎㅎ 윈도우만 써...	1	0.984701

06. Data Modeling

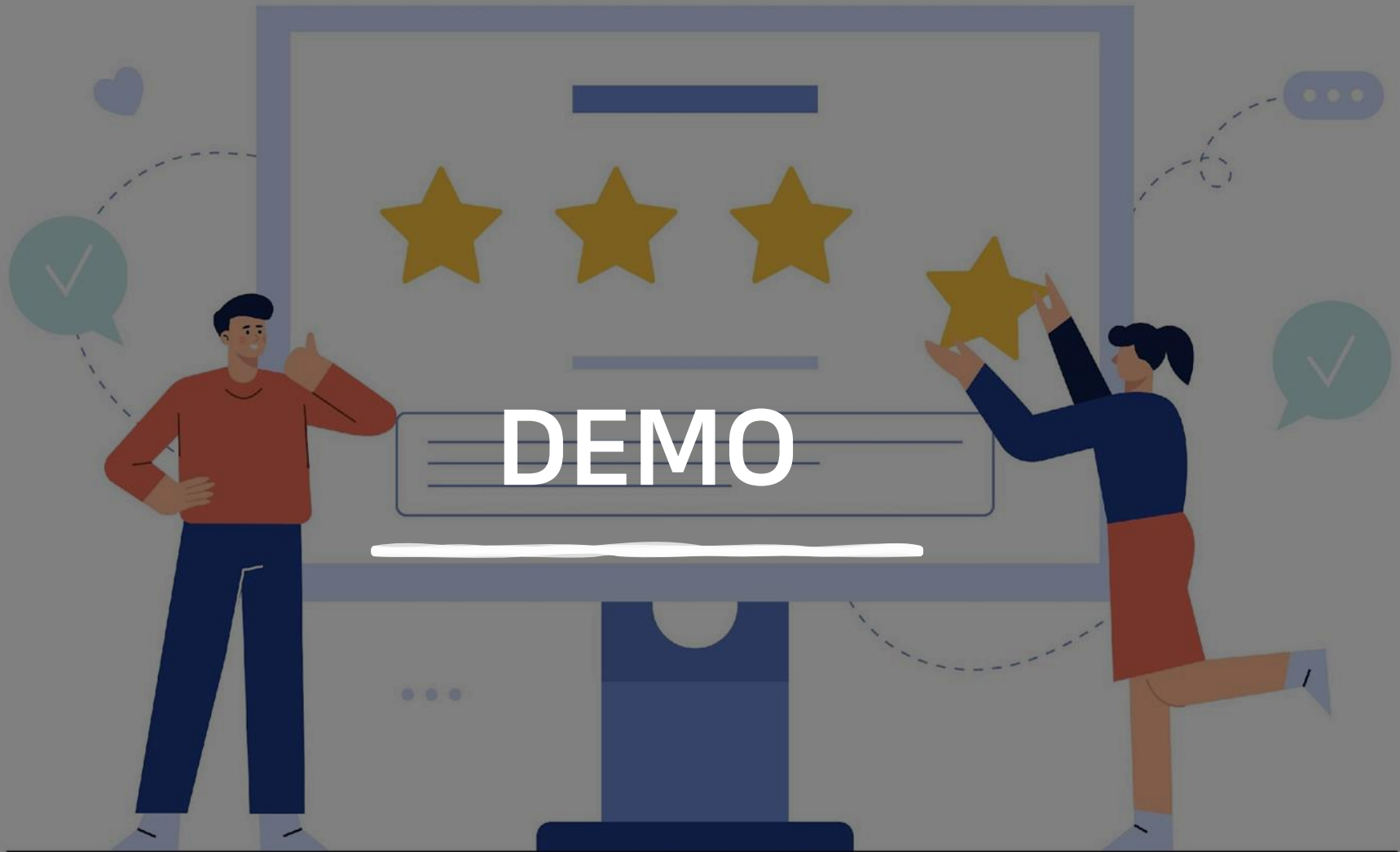


WordCloud

- WordCloud visualizes and shows the size of most frequently appearing words in relative sizes, providing an intuitive glimpse of the most commonly used words.
- KoNLPy (package for natural Korean language processing) was used for Korean text analysis.
- Use repetitive sentences to classify shapes and match parts



```
mask_image = np.array(Image.open('/content/drive/MyDrive/사과.jpg'))
wc = WordCloud(font_path='/content/drive/MyDrive/HMKMRHD.TTF',
               background_color='white',
               max_words=50,
               mask=mask_image,
               max_font_size=200,
               colormap='cool').generate(text4wc)
```





07. Limitation

Difficult to identify advertised reviews

- Inauthentic reviews such as advertised reviews are extracted using keywords such as events and sponsorships.
- The more review data directly input by service users accumulate, the lower the proportion of advertising and event reviews
- However, in case of reviews that are not as obvious as advertising articles, it is necessary to build a sophisticated model using ML/DL because the intention and tone of the writing must be comprehensively considered.

이온 후기

C	D	E	F	G
갤럭시북 이온	삼성갤럭시 북리온 후기 저는 1월5일날 구매하고, 1월6일날 방문수령을 통해서 바로 노트북 이온과 그램 중에 어떤 노트북을 살지 고민이 많았습니다. 그래서 유 이온과 그램 중에 어떤 노트북을 살지 고민이 많았습니다. 그래서 유 그리고 최종적으로 이온을 선택하게 되었는데요 결론부터 말씀드리 우선, 걱정했던 소음과 키감 부분은 생각보다 불편하지 않았습니 그리고 디스플레이는 예상했던대로 아주 훌륭하구요. 발열감 또한 거 무튼, 전체적으로 잘 만든 제품인거 같습니다. 매우만족!!			
갤럭시북 이온	웹서핑과 문서작업용으로 사용할 노트북 고민하다가 최종선택했고 웹서핑과 문서작업용으로 사용할 노트북 고민하다가 최종선택했고 제가 고려한 부분은 - 완전 고사양은 아니지만 인터넷 검색, 유튜브, 오피스 문서 작성할때 - 주로 집에서 쓸것 같지만 가끔은 들고다닐 수 있으니 가볍적 가벼우 - 인터넷, 문서가 주 용도지만 그래픽 성능이 너무 뛰어나고 하면 편집 - PD 충전 제공되었으면 좋겠음 D1159 잘되었으면 좋겠음 100만원대 초반까지			
갤럭시북 이온	삼성갤럭시 북리온 후기 저는 1월5일날 결제를하고, 1월6일날 방문수령을 통해서 바로 노트북을 받았어요. 처음에는 리눅스버전이라서 윈도우10설치까지 직접해야되는 번거로움이 있지만, 온라인에 많은 영상들을 참고 하며했더니, 10분도 안걸리고 바로 설치를 할수있었어요. 윈도우 10 설치후 이제 이런저런 기능도 테스트해보고 광고속에서 나온 그 무선충전을 해보니깐 정말 신기하고, 따로 휴대폰 배터리 걱정 을 할필요가 없다는 생각에 너무 행복했습니다. 그리고 그전 노트북이랑 색깔차이를 비교해보니, 정말 화면속 같은 옷인데 다른 색달로 보이더라고요. 이제 인터넷에서 옷구매를 할때 걱정할 필요 가없을거 같아요. 이제 3월달에 북학준비도할겸 어떤노트북을 살까 고민하다가, 무게도 괜찮고 살짝 게임도 하고싶어서 적당한스펙을 찾다가 이렇게 저한테 딱맞는 노트북을 찾을수있어서 정말 만족합니다. 저처럼 가볍고 살짝 성능을 원하시는 분이있다면 갤럭시북 이온 적극추천드려요!! 아 그리고 ssd 256gb 추가를 했는데, 따로 아무것도 안해도 설치를 해서 주시더라고요, 그대신 제가 만져 오픈을 못한다는 점ㅠㅠㅠ 그래도 만족합니다!!! #삼성전자#갤럭시북#리온#노트북#대학생노트북 #입학선물#갤럭시북#갤럭시북리온#갤럭시북s #후기이벤트#후기응모			

후기이벤트 61/64

1.가볍기도 소문난 그램과 무게 별차이 느끼지 못할 정도로 가볍고
2.자판소음이 없어서 스터디카페에서 사용하기 좋다고 아이가 좋아함
3.배터리완충후 장시간 사용할수있어 충전 걱정없이
4.백라이트기능 있어서 밤에도 불켜지않고 자판이 잘보여요 좋아요^
5.보호필름 안붙이고 사용중으로~~
단점 중 하나인 발열문제.

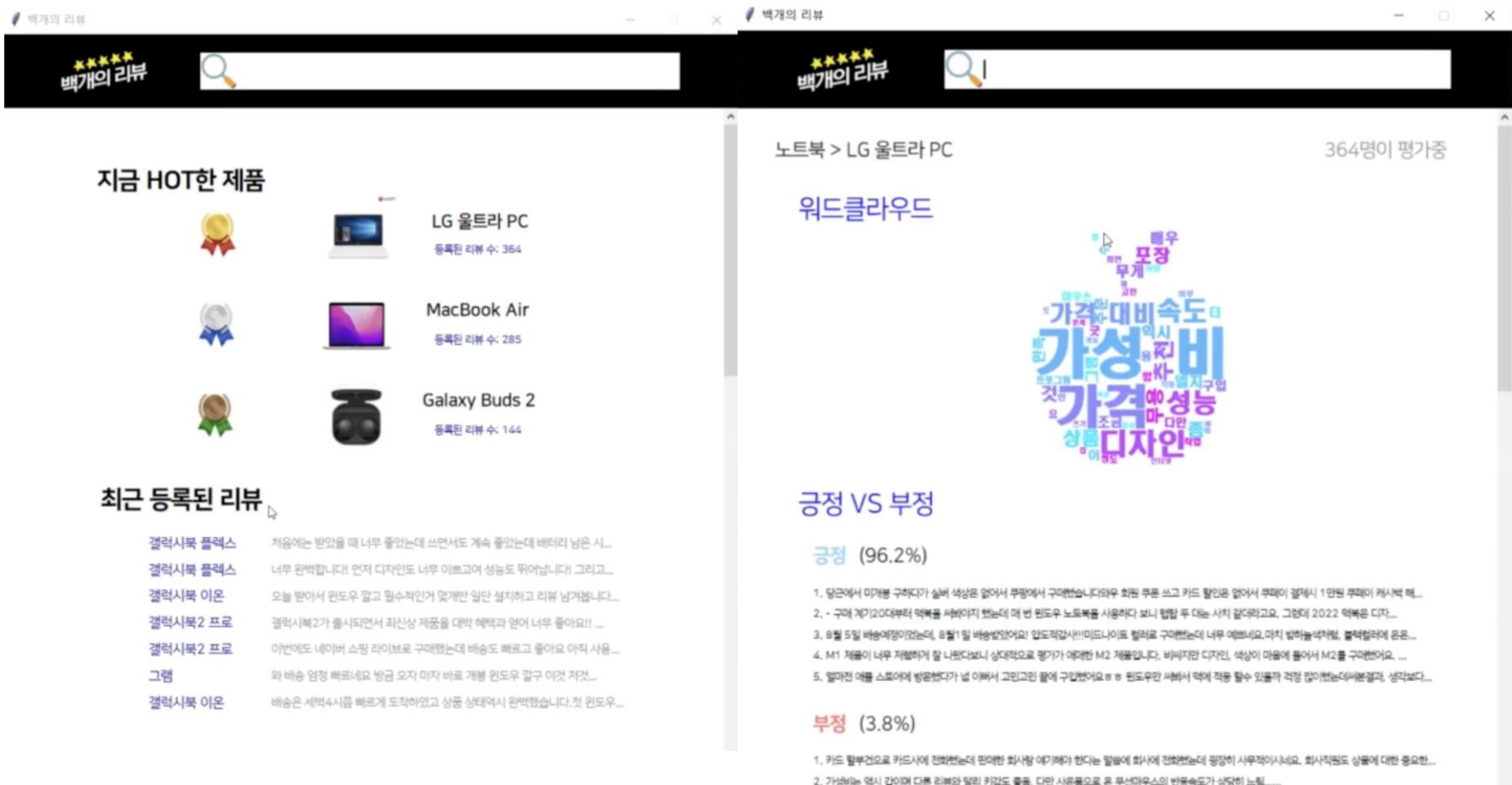
07. Limitation

Only Korean Reviews

- IT reviews are written in various languages around the world, and products sold first in other countries such as the United States may deliver more accurate results.

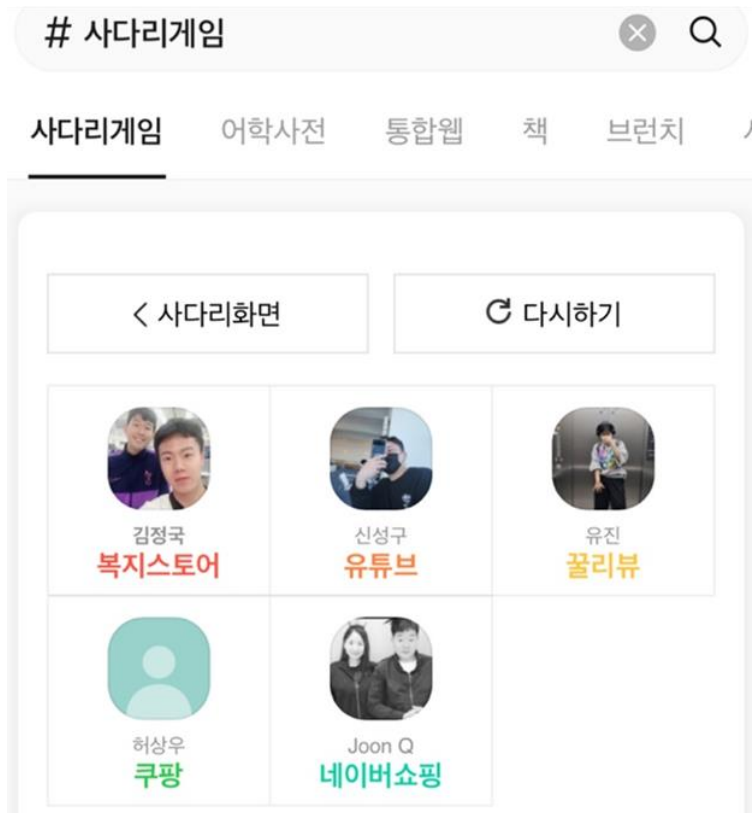
The collage consists of three screenshots. The top-left screenshot shows the 'Trusted Reviews' website with English reviews for fuboTV, Apple iPad Pro, Huawei MateView SE, and De'Longhi Capsule Fit Fan Heater. The top-right screenshot shows a YouTube search for 'macbook air m2 benchmark' with results in Korean, including videos from 'Max Tech', 'Dave2D', 'GregsGadgets', and 'ITSub잇섬'. The bottom-center screenshot shows the YouTube interface in Korean, with a sidebar menu and search results for 'macbook air m2 benchmark'.







09. Team 9 members



1. Seonggu Shin

- Leader of the team
- Project proposal, Final presentation
- Demo development
- MySQL Database

2. Joonkyu Lee

- Web Scraping(Naver shopping)
- MySQL Database & SQL Query

3. Jeongguk Kim

- Web Scraping(Coupang, Student Welfare Store)
- WordCloud Visualization

4. Yujin Jo

- Web Scraping(Ggul review)
- Sentiment analysis

5. Sangwoo Heo

- Web Scraping(Coupang)
- Keyword Extraction & Classification

