

Rec=VoD



www.hellorvdworld.com

LG헬로비전 DX DATA SCHOOL 1기

목차



프로젝트 개요

모델링

상세 기능

팀 회고

개 요

팀원 소개



김지민(PM)
Backend
CI/CD



김한경
Frontend



박효심
Frontend
Backend



송준호
Data Analysis

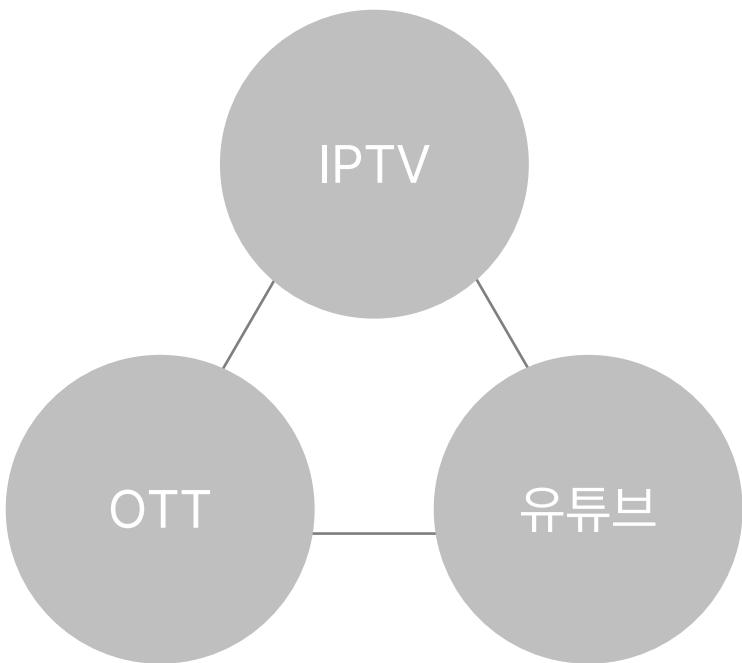


김도현
Data Analysis



김수정
Data Analysis

프로젝트 배경



VS



프로젝트 배경

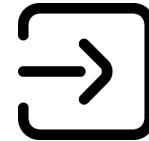
신규 고객 유입 및 기존 고객 이탈 방지를 위한
새로운 서비스 제공 필요



프로젝트 소개



웹 사이트 개발



회원가입 및 로그인 기능



1대의 셋톱박스에 多 아이디



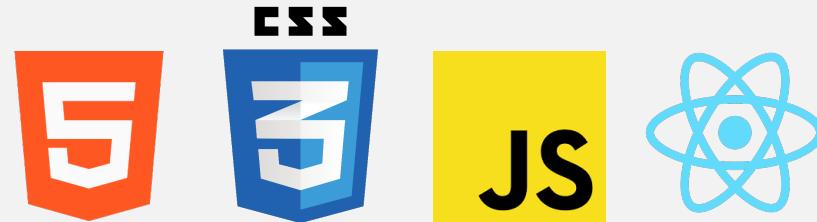
프로필 사진 설정으로 구분



더욱 개인화된 추천 가능

프로젝트 소개

Front-end



Data-analysis



Infra



Back-end



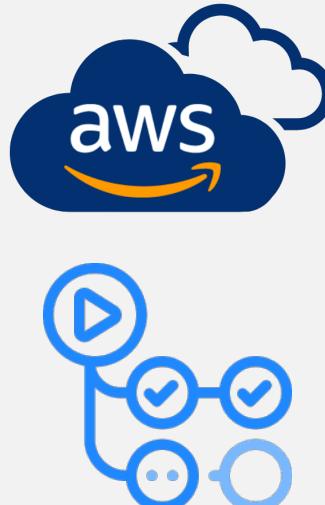
Database



Collaborate

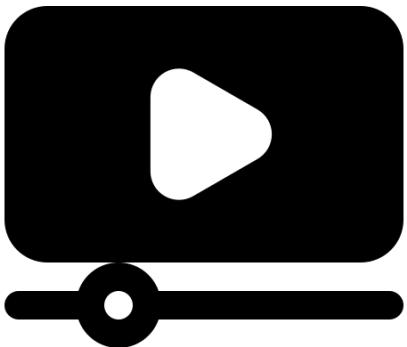


OS, Language



프로젝트 소개

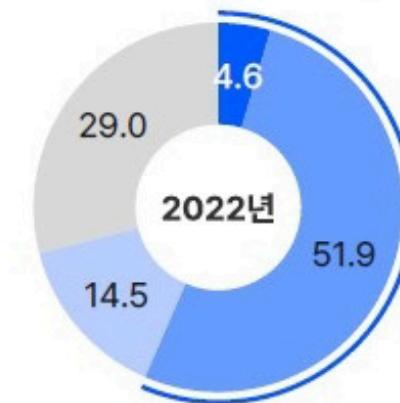
숏폼 콘텐츠 제공



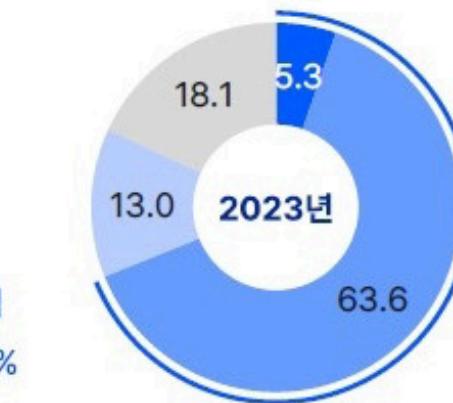
숏폼 콘텐츠 인지 및 이용 행태

[Base: 전체 응답자, N=5000, 단일 응답, %]

- 콘텐츠 게시 경험 있음
- 시청 경험, 콘텐츠 게시 경험 없음
- 인지, 시청 경험 없음
- 비인지



시청 경험
56.5%

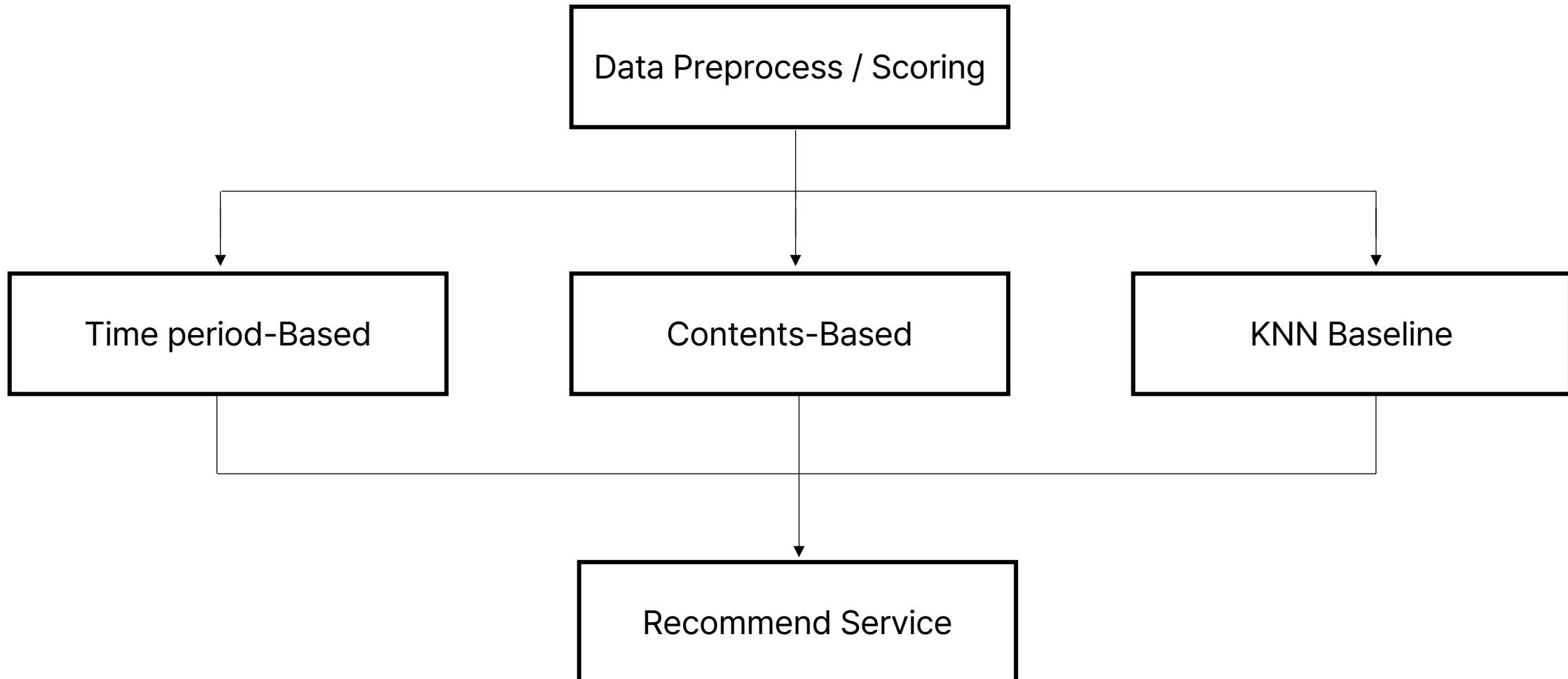


시청 경험
68.9%



모델링

모델링 흐름



데이터 전처리 / Scoring

- 데이터 소개

VOD 시청 로그 데이터 - VOD (8, 9, 10월)

VOD 진입 전 클릭 로그 데이터 – Content (8, 9월)

subsr	series_nm	super_asset_nm	ct_ct	genre_of_ct_ct	SMRY	ACTR_DISP	disp_rtm	log_dt
0	(영어더빙)핑크퐁 동물 동요	핑크퐁 동물 동요 20회	키즈	기타	겨울잠. 숲, 바다, 극지방 등 세계 여러 곳 에 사는 동물 친구들을 만나러 가자! ...	핑크퐁	0:01	20230813071155



name	Summary	actors	director	BigCategory	SmallCategory	Category	rename	counts	vod_id	vod_name	running_time	genres
그것 이 알고 싶은 다 인 시사 교양 프로그램	사회 전반의 다양한 문제점을 들여다보는 SBS의 대표적인 시사교양 프로그램	김상중	한재신, 장경주, 문치영, 신진주, 정문명	TV 시사/교양	기타	TV프로그램	그것이 알고 싶다	298 297	그것이 알고 싶다	67분	TV시사 교양 기타	

데이터 전처리 / Scoring

방법 1	시간 가중치를 도입한 평점
데이터	VOD 및 Content 데이터
추가 전처리	회차 제거
공식	$a_{ij} = \sum_k w_k \left(\frac{1}{2} \right)^{\frac{t_0 - t_k}{T}}$

Hong Ooi. 'Package 'SAR' ', Smart Adaptive Recommendations , MIT + file LICENSE, 2020-10-07

방법 2	클릭률	회차를 이용한 평점 생성
데이터	Content 데이터	VOD 데이터
추가 전처리	회차 제거	
공식	유저-프로그램 당 클릭 횟수 / 유저의 총 클릭 횟수	$r_p(u,i) = \frac{n_p(u,i)}{N_p(i)} = \frac{\sum_{j \in p} r(u,j)}{N_p(i)}$

데이터 전처리 / Scoring

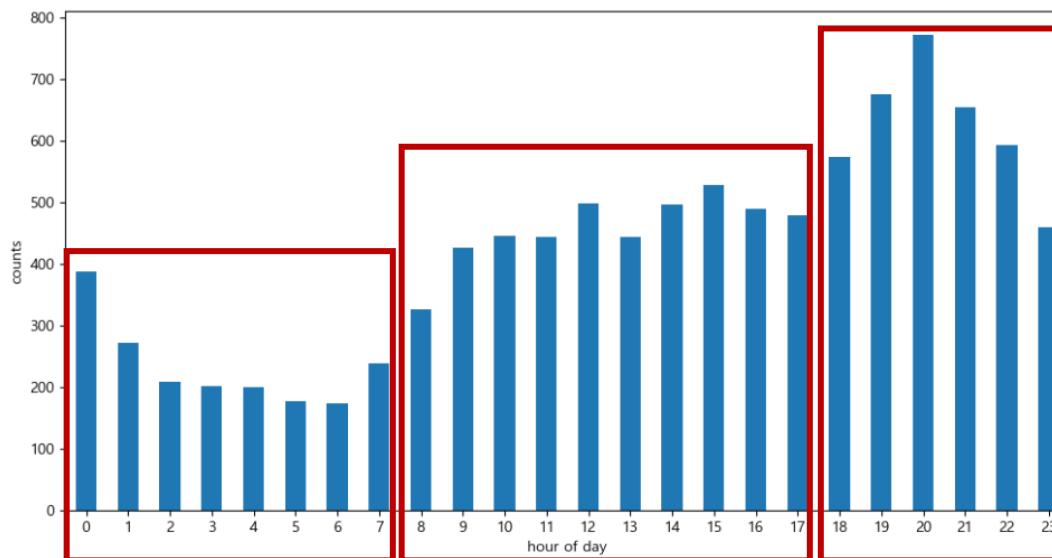
Train set	VOD 8~9월 시청 데이터 Content 8~9월 클릭 데이터
Test set	VOD 10월 시청 데이터
Metric	precision@10

precision@k 란 ?

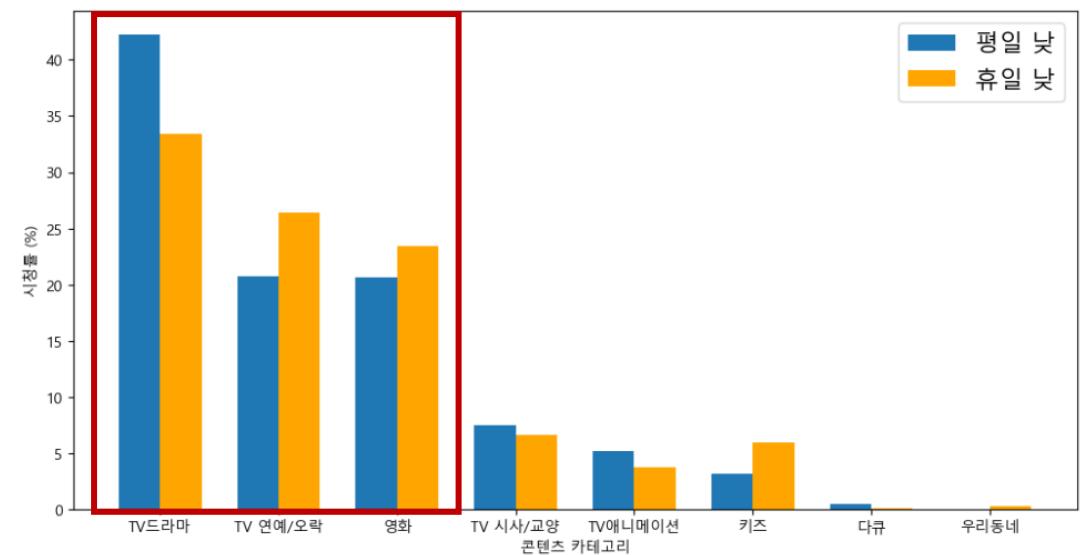
정보 검색에서 모델의 정밀도를 측정하는 지표로, 상위 k개의 결과 중 정답이 포함된 비율을 나타냄

Time period-Based

가정 : VOD 시청자들의 하루 스케줄 유사할 것 / 평일과 휴일 간 시청 프로그램 차이 존재

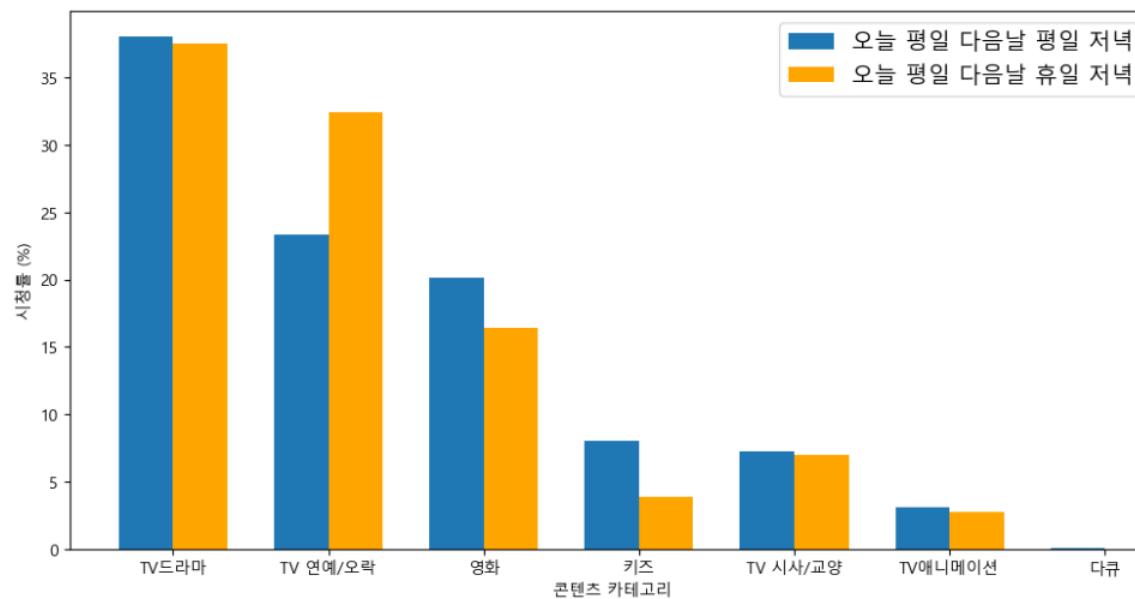


0-7 / 8-17 / 18-23



평일 / 휴일 간 시청비율 차이

Time period-Based



다음날 평/휴일 여부에 따른 시청비율 차이

새벽 : 전날, 당일의 평/휴일 여부에 영향

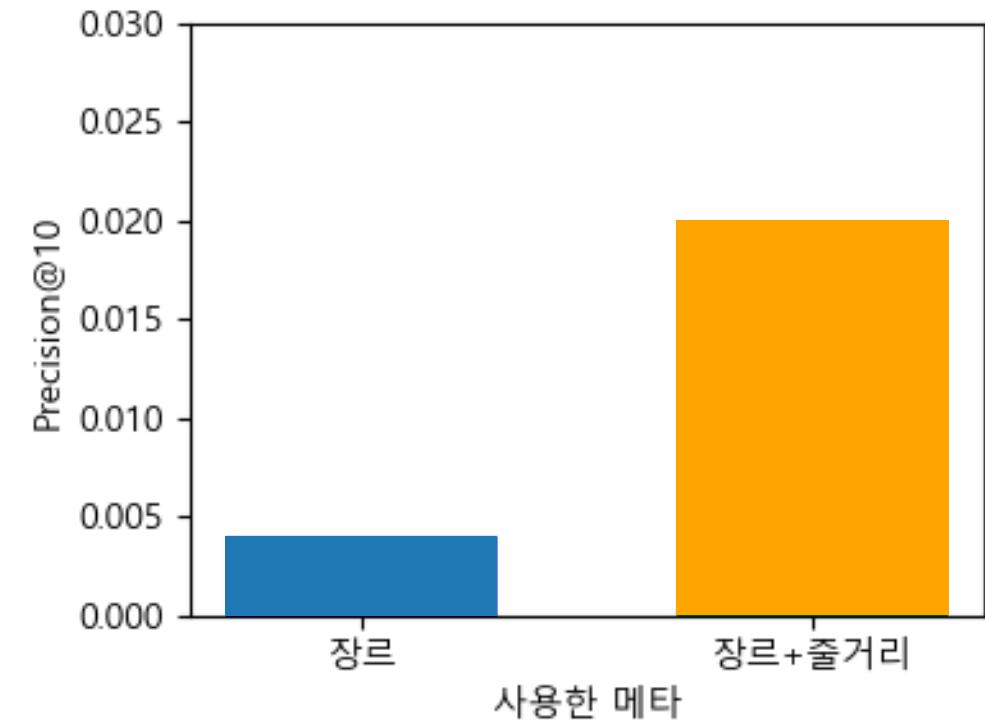
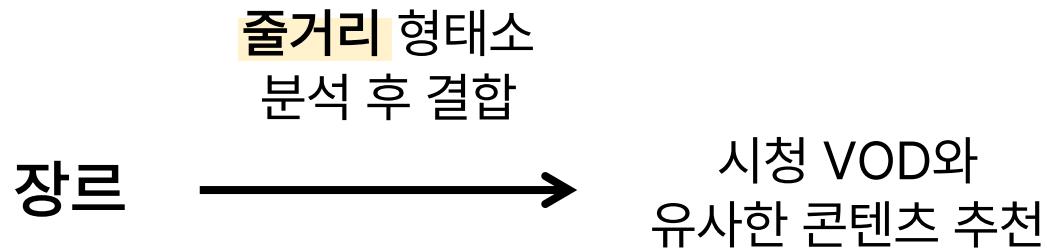
저녁 : 당일, 다음날의 평/휴일 여부에 영향



시간대를 **10가지** 경우로 나누어 추천

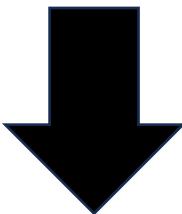
precision@10 : 0.34

Contents-Based



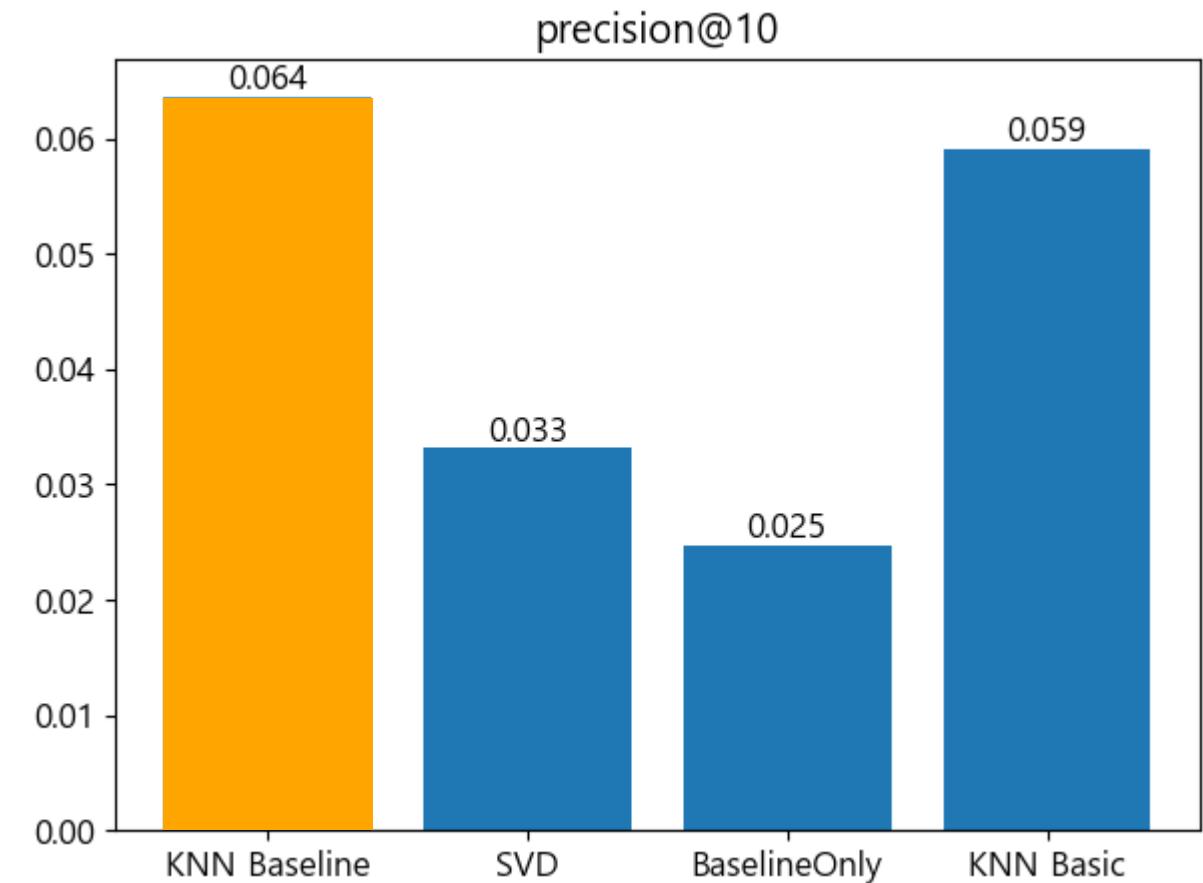
KNN-Baseline - 알고리즘 선정

Surprise 라이브러리 내장 알고리즘 중 4가지 비교

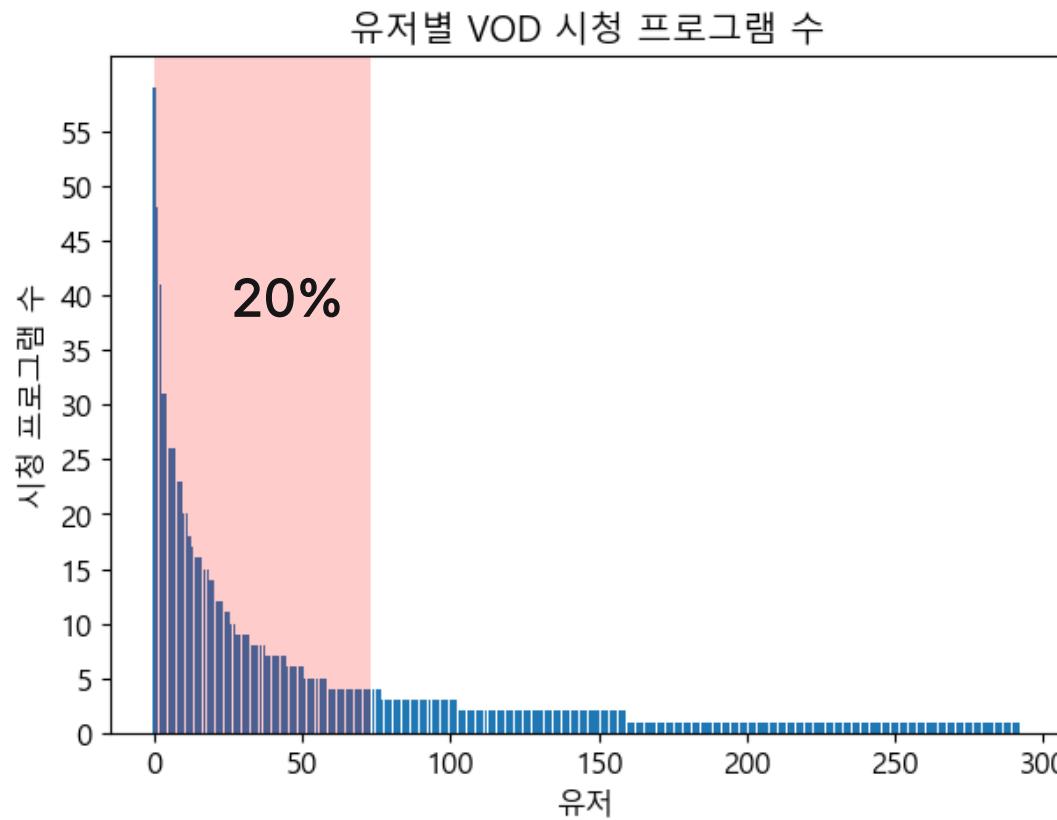


KNN 기반 모델의 성능이 가장 우수

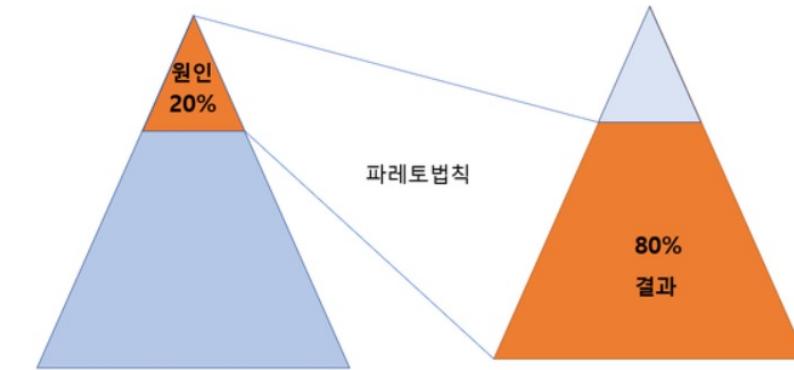
Model	Train set
Item-base	VOD 8~9월 시청 데이터
User-base	Content 8~9월 클릭 데이터



KNN-Baseline - User Grouping



파레토 법칙

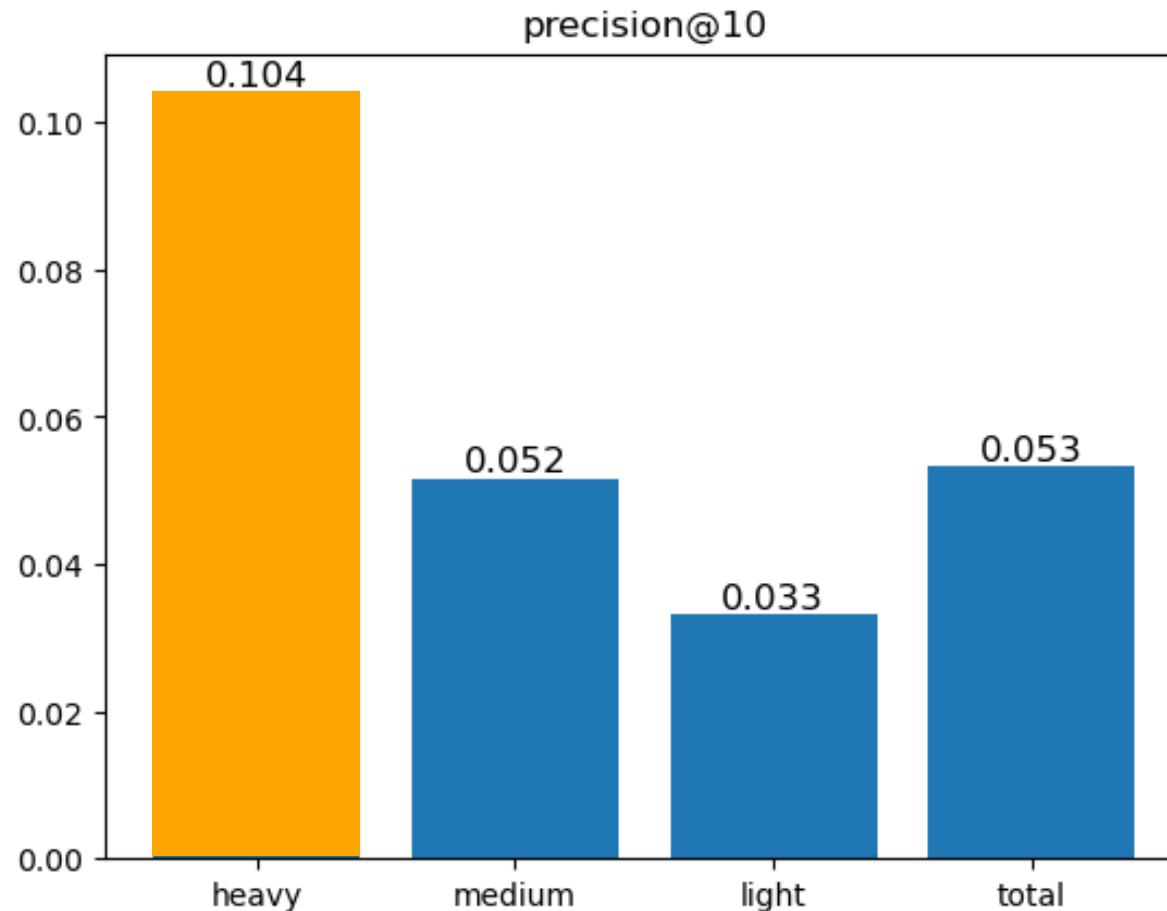


8,9월 VOD 데이터에서 5개 프로그램 이상 시청한 유저
293명 중 59명 (약 20%)
전체 유저의 20%가 전체 시청 수의 70% 를 차지

KNN-Baseline - User Grouping

유저 구분	유저 수	특징
Heavy user	59	프로그램 5개 이상 시청한 유저
Medium user	44	프로그램 3~4개 시청한 유저
Light user	813	프로그램 1~2개 시청한 유저 + 시청하지 않고 클릭만 한 유저

KNN-Baseline - User Grouping 의 효과



Heavy 유저
추천 정확도



상세 기능



유저 별 추천

추천 다양화

숏필름

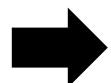
사용자 경험
고객 만족도
극대화

트렌드에 맞는
UI 구성

1. 유저 별 추천 – 기존 USER

평소 시청한 목록

	program_nm	count
0	최강 배달꾼	242
1	경이로운 소문	79
2	사랑의 불시착	74
3	응답하라 1988	72
4	응답하라1994	46
5	경이로운 소문2: 카운터 편치	44
6	야인시대	37
7	슬기로운 감빵생활	27
8	명불허전	25
9	골 때리는 그녀들	24
10	38 사기동대	20



KNN-Baseline

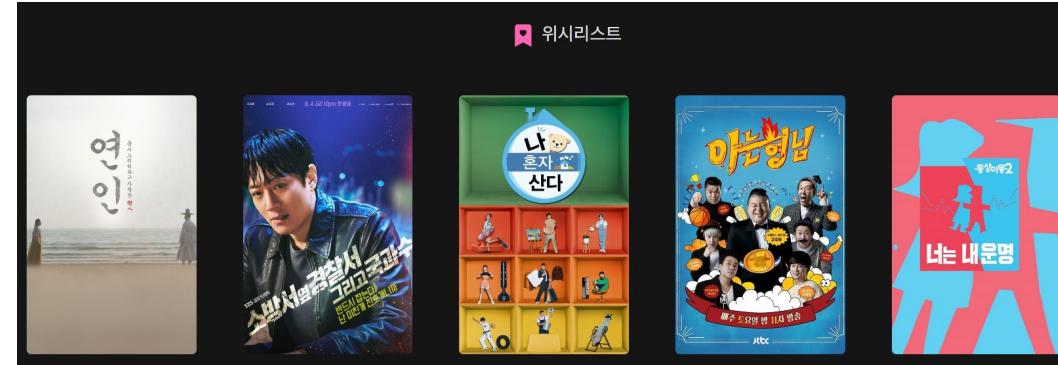


Content Based



1. 유저 별 추천 – 기존 USER

위시리스트



위시리스트



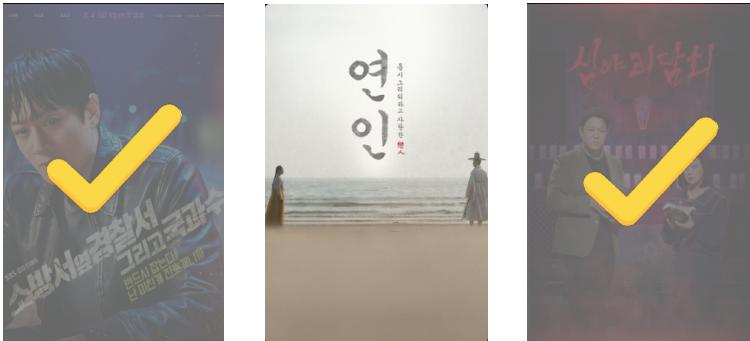
KNN-Baseline



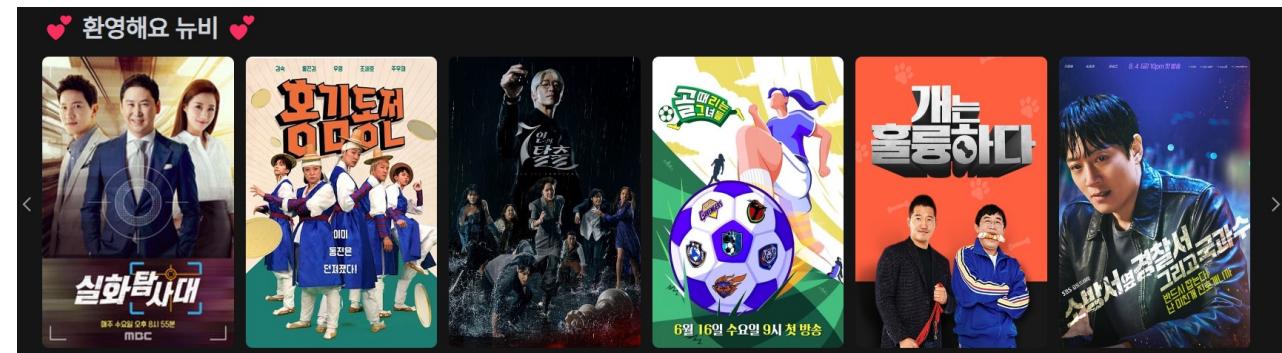
당신의 위시리스트를 토대로 추천해드렸어요

1. 유저 별 추천 – 새로운 USER

선호 프로그램 선택



KNN-Baseline



Content Based



2. 추천 다양화

⌚ 시간대별 추천

특정 시간대에 선호도 높은 프로그램을 사용자에게 추천하는 서비스

ȩ 다른 사용자의 Pick

비슷한 성향을 가진 사용자의 추천 항목을 다른 사용자에게 추천하는 서비스

▲ 시즌 별 추천

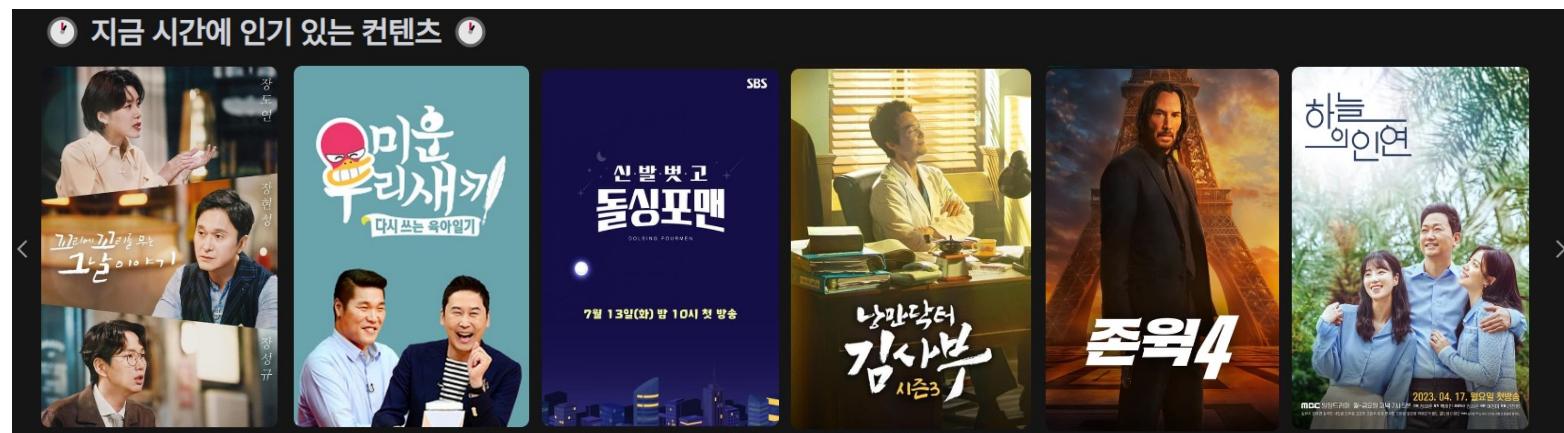
특정 시즌마다 인기 있는 VOD를 추천하는 서비스

2. 추천 다양화 – 시간대 별 추천

낮 (8시 ~ 17시)



저녁 (18시 ~ 23시)



2. 추천 다양화 – 다른 사용자의 Pick

유사 사용자 상위 3명



- | | | |
|---|---|-------------|
| 1 | ○ | [1번 list] |
| 2 | ○ | [2번 list] |
| 3 | ○ | [3번 list] |

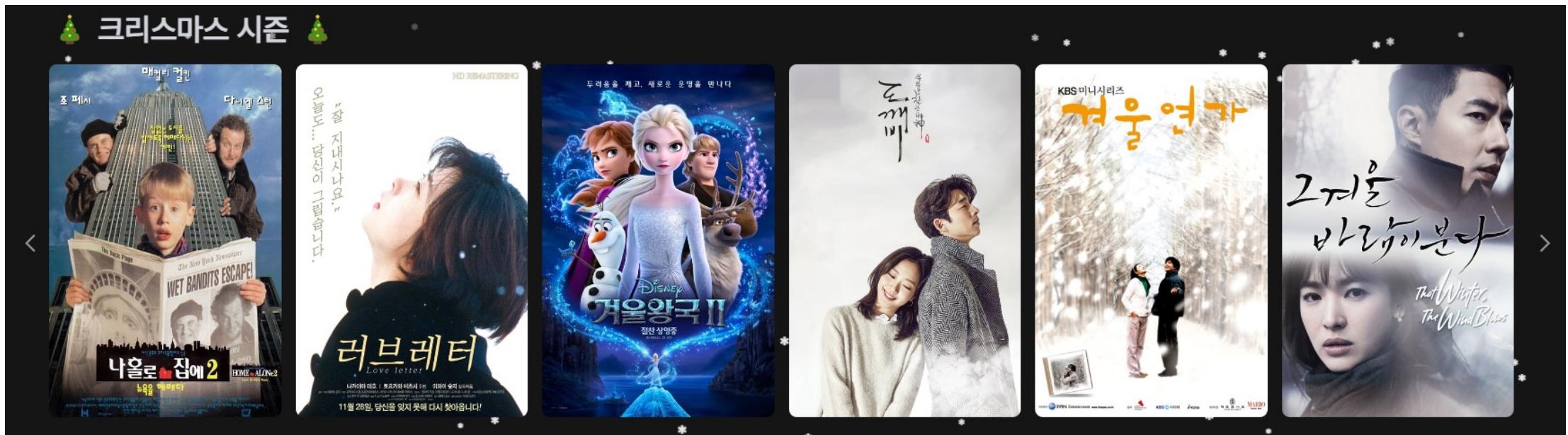
추천 리스트 결합

[[1번 list] + [2번 list] + [3번 list]]

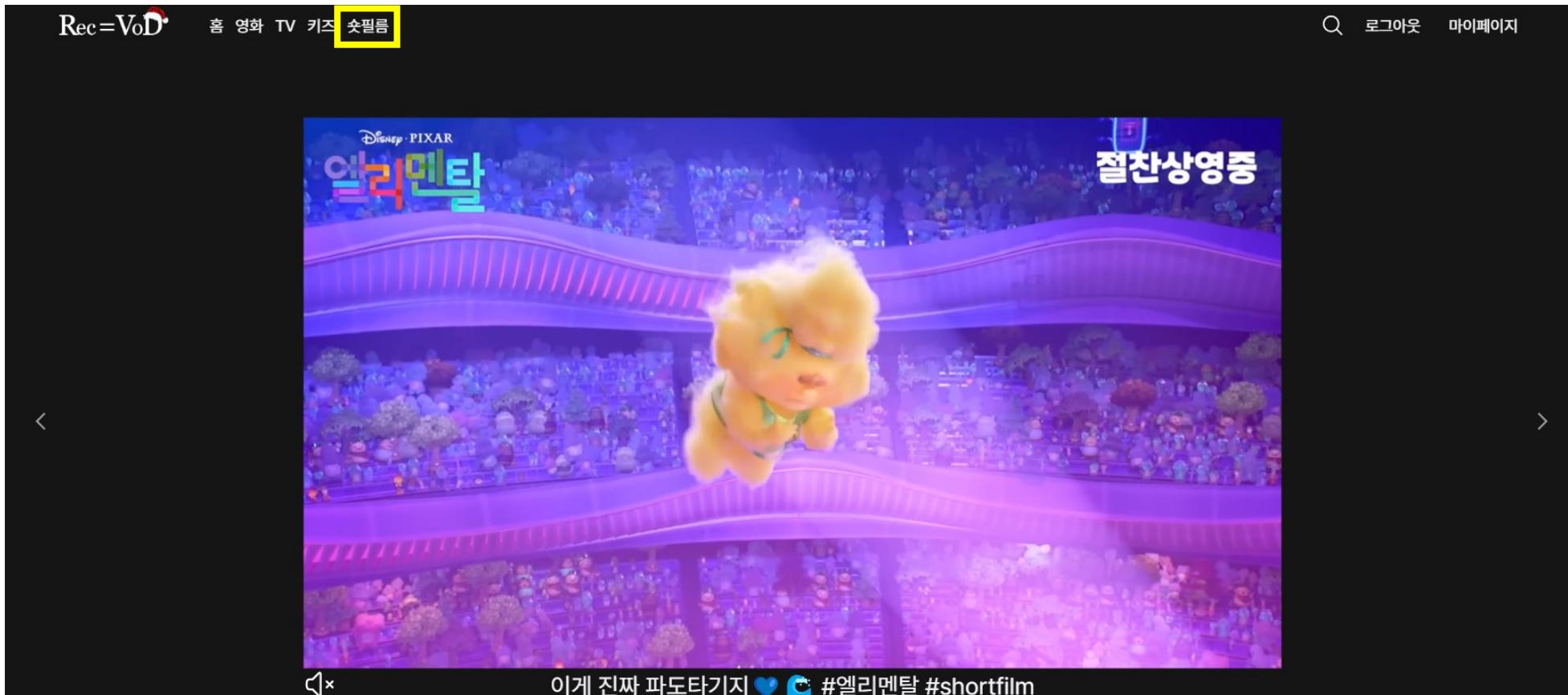
이미 추천된 항목 제외

[□ □ □ □ □ □ ↗ □ ...]

2. 추천 다양화 – 시즌 별 추천



3. 솟필름



→ 마케팅 트렌드로 급부상한 솟폼 콘텐츠를 차용하여 차별화된 고객 경험 제공

4. 사용자 경험 및 고객 만족도 극대화 – 다양한 검색 기능

검색어 : 아

Search Vods

GET /contents/%EC%95%84

HTTP 200 OK
Allow: GET, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

```
[  
  {  
    "id": 3,  
    "name": "마는 형님",  
    "smallcategory": "기타",  
    "imgpath": "/zak7mqMxJGnnkplSdbjDtV7bg2Y.jpg",  
    "count": 50  
  },  
  {  
    "id": 18,  
    "name": "감자마저씨의 크리스마스 공연 외",  
    "smallcategory": "애니메이션",  
    "imgpath": "/aTaGTqkbMlwoy0ef0M48qJkMltxE.jpg",  
    "count": 1  
  },  
  {  
    "id": 22,  
    "name": "할아버지와 배 타기 외",  
    "smallcategory": "애니메이션",  
    "imgpath": "/aTaGTqkbMlwoy0ef0M48qJkMltxE.jpg",  
    "count": 2  
  },  
]
```

기본 검색 API

검색어 : 으ㄴㅎㄴ

Search Vods By Choseong

GET /contents/search/%E3%85%87%E3%84%B4%E3%85%8E%E3%84%B4

HTTP 200 OK
Allow: GET, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

```
[  
  {  
    "id": 3,  
    "name": "마는 형님",  
    "smallcategory": "기타",  
    "imgpath": "/zak7mqMxJGnnkplSdbjDtV7bg2Y.jpg",  
    "count": 50  
  }  
]
```

초성 검색 API

검색어 : 김혜수

Searchactors

GET /contents/search/actor/%EA%B9%80%ED%98%9C%EC%88%98

HTTP 200 OK
Allow: GET, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

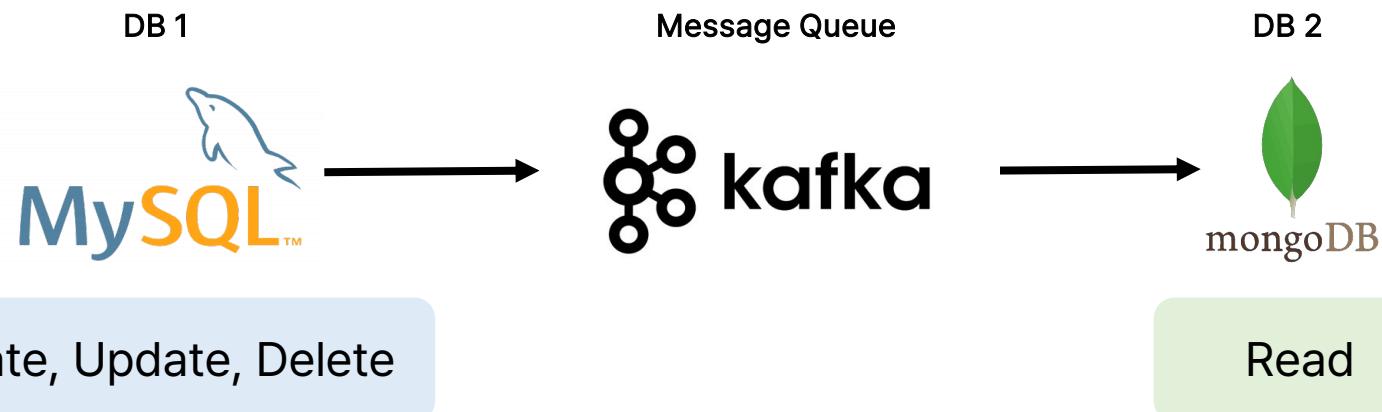
```
[  
  {  
    "id": 179,  
    "name": "모던 보이",  
    "smallcategory": "드라마",  
    "imgpath": "/2X0hg9w0tZtiIPCHzqErAyS48DM.jpg",  
    "count": 4  
  },  
  {  
    "id": 351,  
    "name": "영원한 제국",  
    "smallcategory": "다큐멘터리",  
    "imgpath": "/cetowKINhYsyBz3YTRn024I7H4Z.jpg",  
    "count": 2  
  },  
  {  
    "id": 593,  
    "name": "UHD로 만나는 TV문학관",  
    "smallcategory": "기타",  
    "imgpath": "/1PucPXOm6pGgJEYFKmEUinX9hQP.jpg",  
    "count": 4  
  },  
  {  
    "id": 709,  
    "name": "타짜",  
    "smallcategory": "드라마",  
    "imgpath": "/qjCl5nggBaOfy2lhVraE4iSBzP3.jpg",  
    "count": 8  
  },  
]
```

배우 검색 API

4. 사용자 경험 및 고객 만족도 극대화 – CQRS를 통한 성능개선 및 예외처리

CQRS란 ?

- 데이터 저장소의 읽기와 업데이트 작업을 분리하는 패턴
- 어플리케이션의 퍼포먼스, 확장성, 보안성을 극대화
- 복수의 업데이트 명령들에 대한 충돌도 방지 가능



4. 사용자 경험 및 고객 만족도 극대화 – DB화를 통한 속도 개선



stbnum	method	rec1_id	rec10_id	rec2_id	rec3_id
59,900,000	1	111	296	14	3,028
59,930,000	1	728	398	456	1,156
60,050,000	1	213	398	360	1,156
60,067,000	1	954	884	4,782	992
60,224,000	1	410	747	4,787	1,886
60,326,000	1	1,639	4,833	4,763	887
60,463,000	1	699	398	690	1,156
60,506,000	1	1,156	1,395	1,347	2,979
60,510,000	1	1,156	1,395	1,347	2,979
60,593,000	1	1,506	30	14	2,092
60,939,000	1	1,156	1,395	1,347	2,979
60,955,000	1	958	1,347	738	3,570
61,065,000	1	835	3,028	861	3

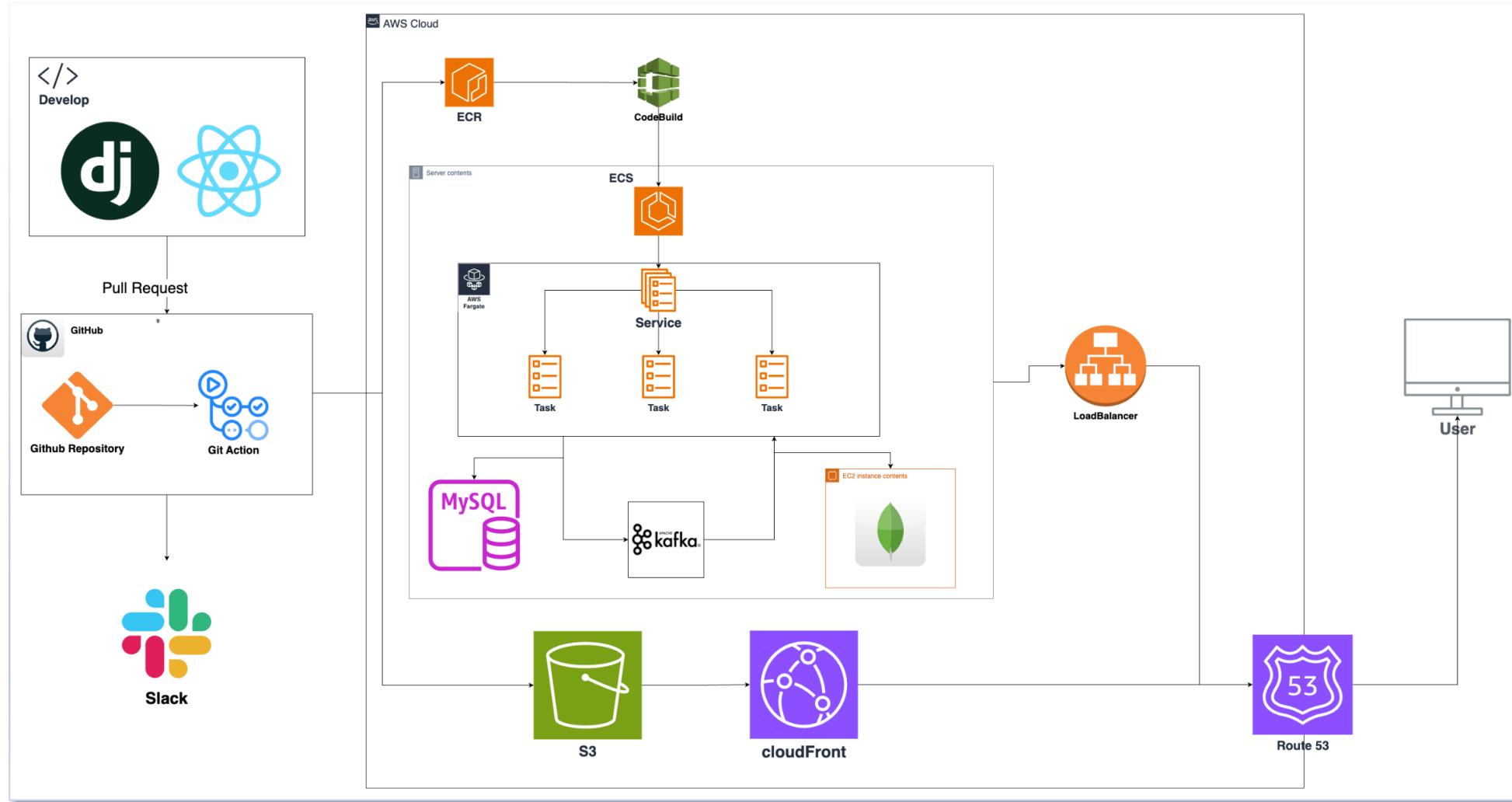


id	stbnum	watched	rec1
1	5941	59879000	3405
2	5942	59879000	1860
3	5943	59879000	1168
4	5944	59882000	4023
5	5945	59882000	4302
6	5946	59882000	1677
7	5947	59886000	1399
8	5948	59886000	3893
9	5949	59886000	818

방식	속도
실시간 추천 함수	1.10272 초
DB를 통한 추천	0.06280 초

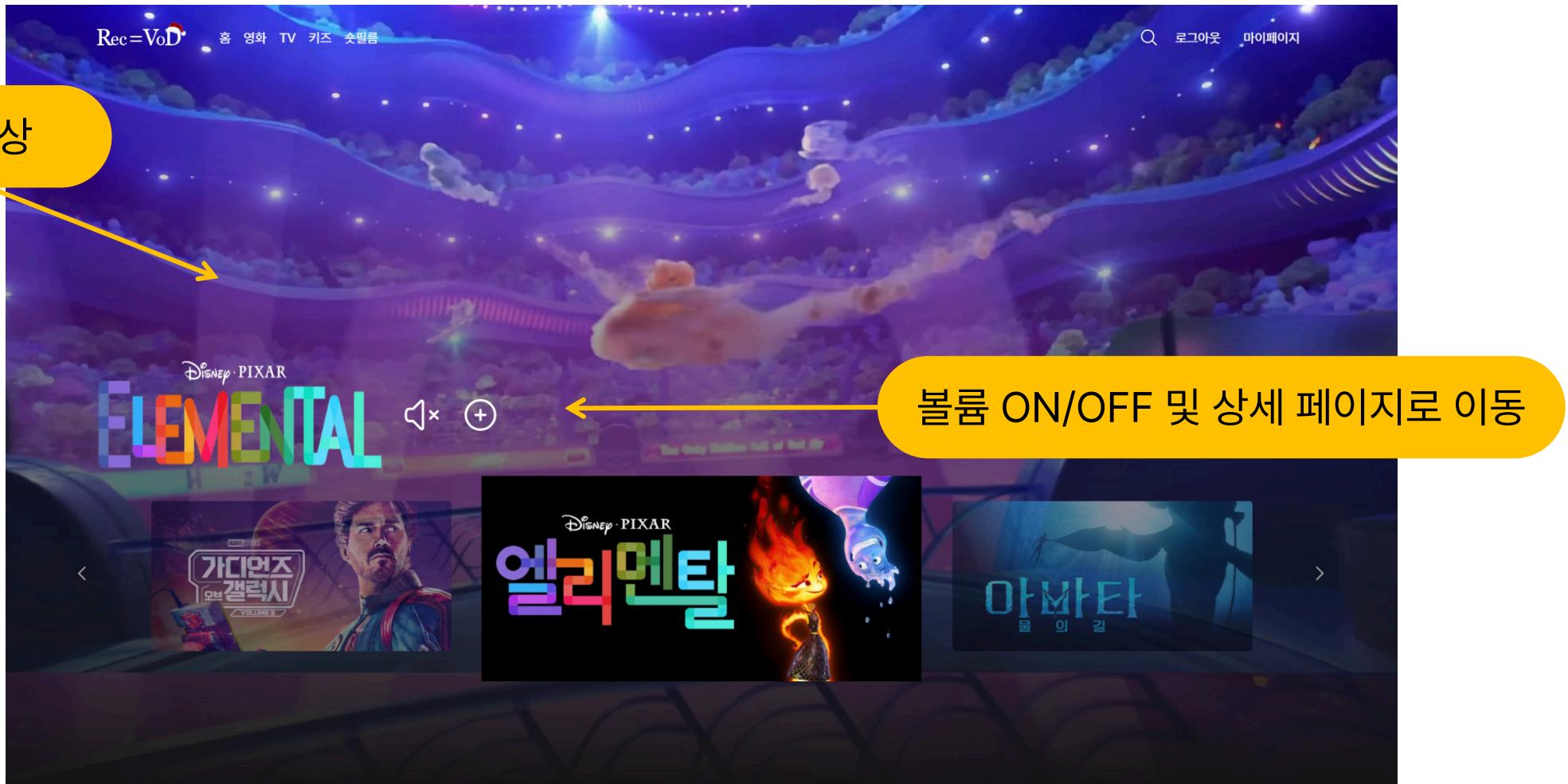
18배
빠른 조회

4. 사용자 경험 및 고객 만족도 극대화 – 무중단 배포 및 HTTPS보안



5. 트렌드에 맞는 UI 구성

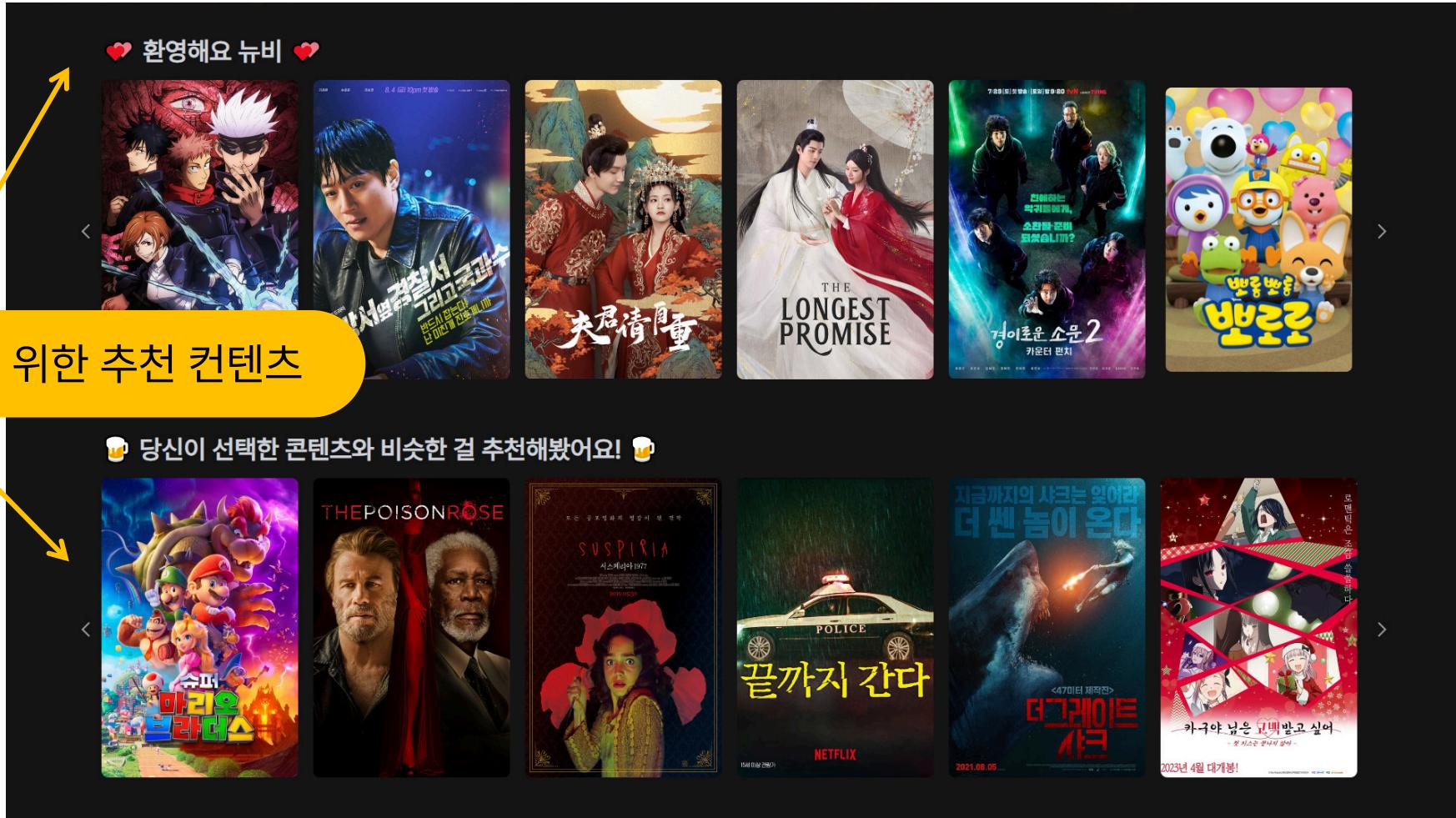
메인 페이지 (상)



5. 트렌드에 맞는 UI 구성

메인 페이지 (하-1)

신규 유저를 위한 추천 컨텐츠

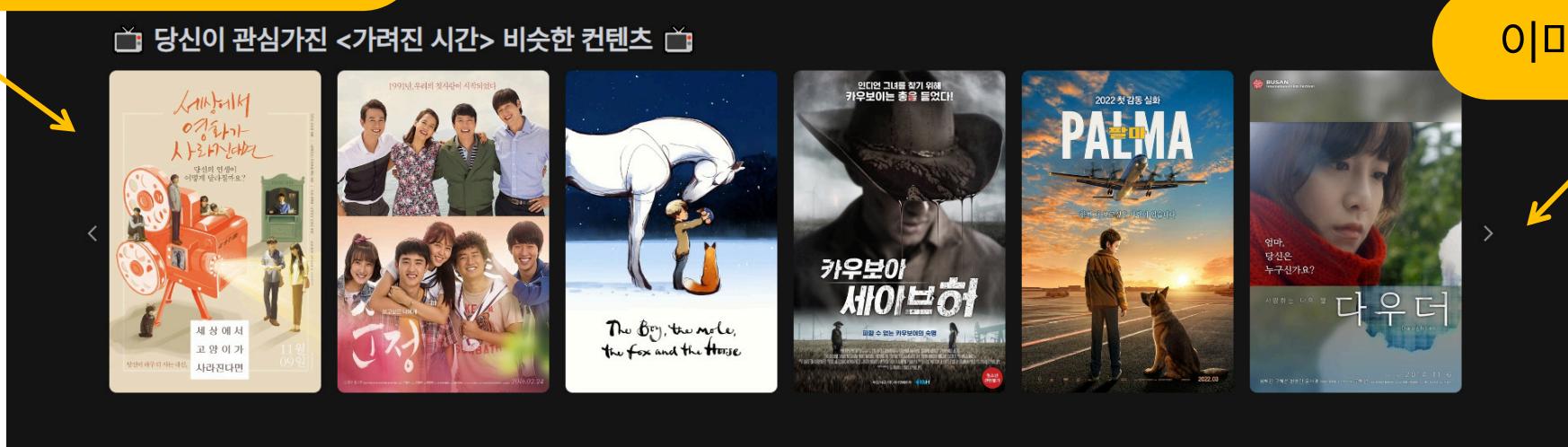


5. 트렌드에 맞는 UI 구성

메인 페이지 (하-2)



기존 유저를 위한 추천 컨텐츠



이미지 슬라이더

5. 트렌드에 맞는 UI 구성

카테고리 페이지

★ 영화 Top10 ★

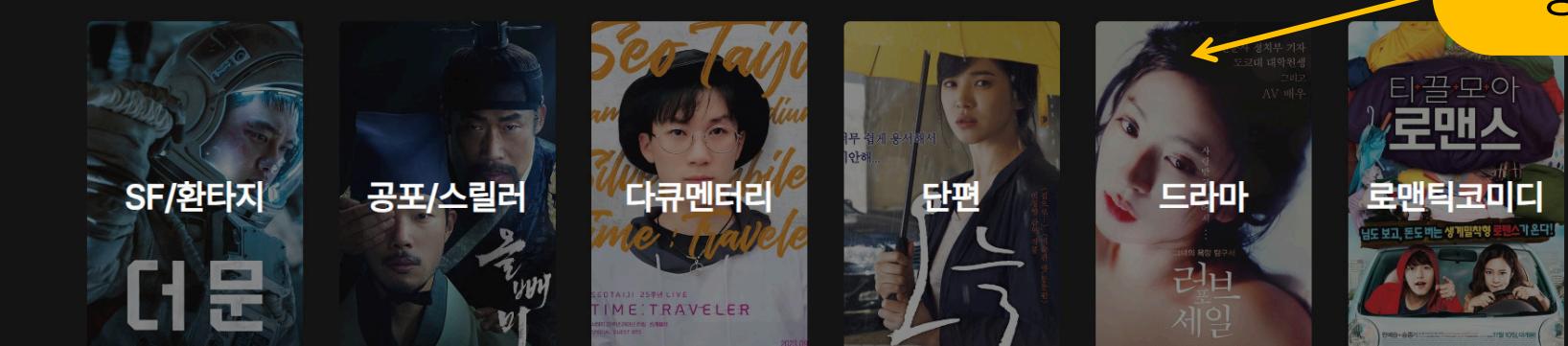


카테고리 별 TOP10

1 2 3 4 5 6

◀ ▶

拇指 카테고리를 골라보세요



SF/환타지
더문

공포/스릴러
을 빼면

다큐멘터리
Seo Taiji
Time Traveler

단편
나는 그녀

드라마
리브 세일

로맨틱코미디
티끌모아 로맨스

장르 이미지 버튼

5. 트렌드에 맞는 UI 구성

장르 별 페이지

The screenshot shows a dark-themed movie recommendation website. At the top, there's a navigation bar with the logo 'Rec=VoD', a search icon, and links for '로그아웃' (Logout) and 'マイ페이지' (My Page). Below the navigation, a category header '공포/스릴러' (Horror/Suspense) is displayed. A row of movie posters for '100 피트', '12명의 죽고 싶은 아이들', '21 브릿지: 테러 셧다운', '22 CHASER', '22년 후의 고백', and '47시간' is shown. To the right of this row, three sorting buttons are highlighted in red: '이름순' (Name Order), '인기순' (Popular Order), and '랜덤' (Random). An orange arrow points from a yellow callout bubble to the '랜덤' button. A yellow callout bubble at the bottom right contains the text '이름순, 인기순, 랜덤 정렬' (Name Order, Popular Order, Random Sort). Below the main row, there's another row of movie posters for '64', '기생충', '가면병동', '개인캐스', and '개인캐스' (repeated).

5. 트렌드에 맞는 UI 구성

VOD 상세 페이지 (상)

VOD 포스터 및 이미지

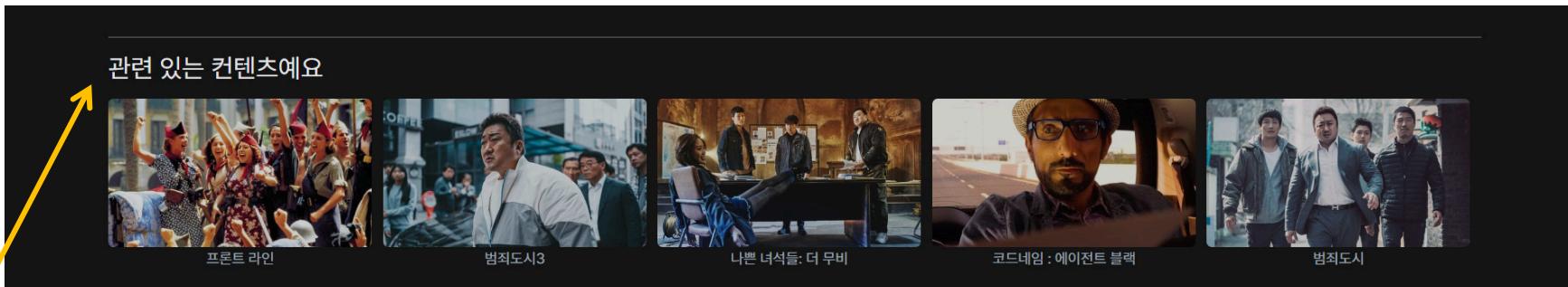
평점, 위치, 주소복사 등

VOD 상세 정보



5. 트렌드에 맞는 UI 구성

VOD 상세 페이지 (하)



줄거리 및 장르 기반 관련 컨텐츠

이 작품에 대한 생각을 자유롭게 작성해주세요...

리뷰 작성

선택하세요

총 1건

평점 ★★★★☆ 4점

lg*****com

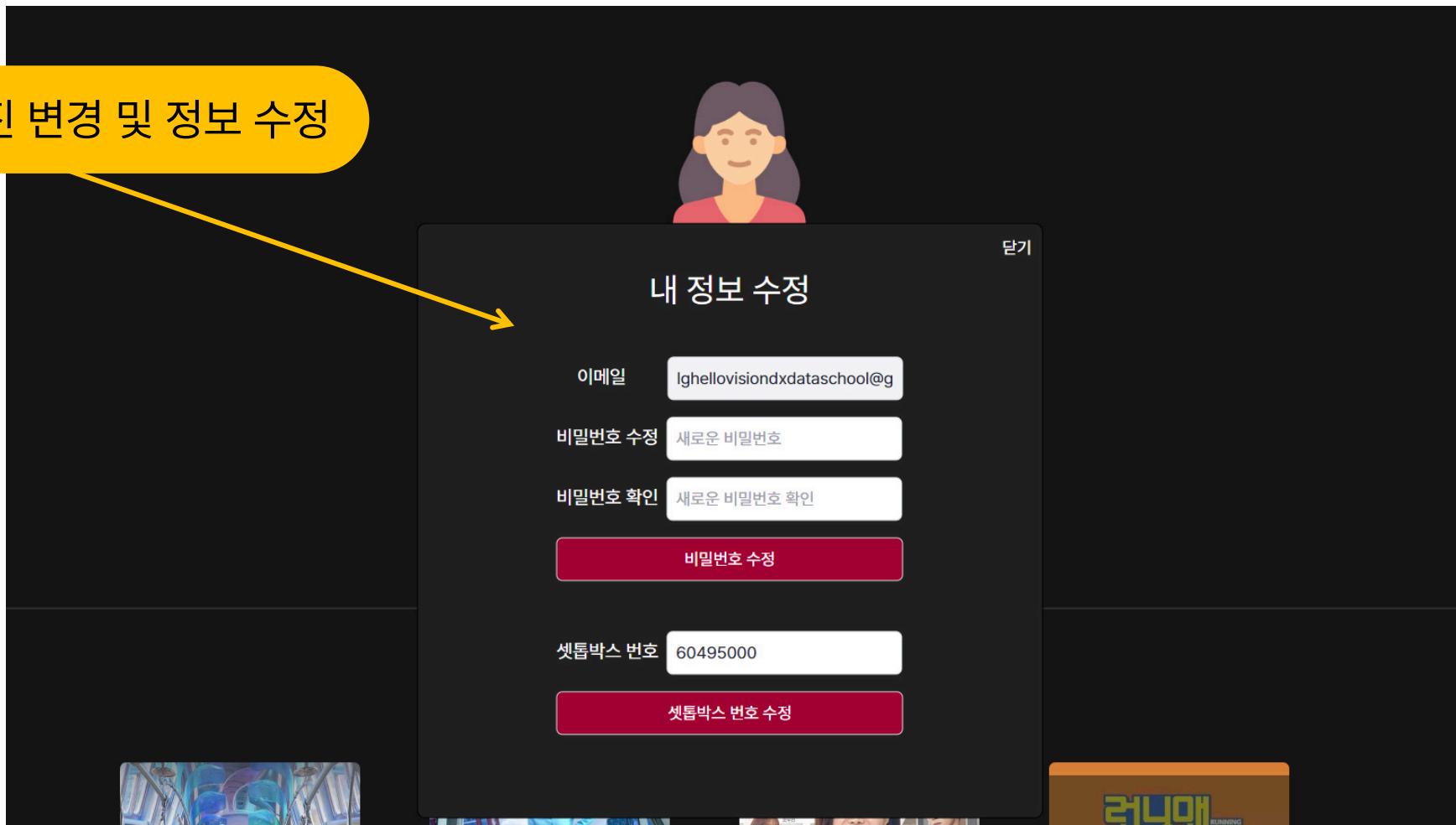
시즌 3도 너무 기대됩니다~

리뷰작성 및 확인

5. 트렌드에 맞는 UI 구성

마이페이지 (상)

프로필 사진 변경 및 정보 수정



5. 트렌드에 맞는 UI 구성

마이페이지 (하)

위시리스트 확인

위시리스트

총 4건



나의 리뷰

총 3건



범죄도시2

시즌 3도 너무 기대됩니다~

평점 ★★★★☆ 4점 수정 삭제

나의 리뷰 수정 및 삭제

팀 회고



커뮤니케이션 룰을 정하고 협업 툴 사용하여 원활한 소통을 한 것 같아 좋았습니다.



애자일 개발로 진행할 수 있어 좋았고, 처음으로 제대로 된 성과를 만들어 낸 것 같아 좋았습니다.
또한 협업에 대한 중요성을 많이 깨달을 수 있었던 프로젝트였습니다.



CQRS의 필요성에 대해 알게 되었고 이것을 통해 성능을 개선할 수 있는 방법을 학습할 수 있었습니다.



팀원과 함께 하나의 주제에 대해 깊게 토론하고 결과를 도출해볼 수 있는 시간이 주어져서 좋았고,
프로젝트에 집중하는 동안 성장할 수 있어서 뿌듯했습니다.



비정형 로그 데이터를 정형 데이터로 변환하는 과정이 흥미로웠습니다.



다양한 추천 알고리즘을 적용하는 과정이 흥미로웠습니다.
이를 통해 사용자 취향에 맞는 추천 시스템을 구축하는데 필요한 전문지식을 얻을 수 있었습니다.

Q & A

URL: www.hellorvdworld.com

ID : testing@hv.com

PW : hellovision

셋탑 번호 : 62248000

Thank You