

IFITV- 맞춤형 IPTV 콘텐츠 추천 서비스

당신의 TV, 손끝으로 FIT된다.

백지연, 이경준, 이예림, 최동연

CONTENTS

01

프로젝트 개요

02

데이터 정의 및 수집 가능성

03

데이터 수집 아키텍처 설계

04

데이터 수집 및 전처리

05

데이터 검증 및 분석

06

시스템 아키텍처 설계

07

데이터 시각화 및 Dashboard 생성

08

앞으로 구현할 기능 & 기대효과

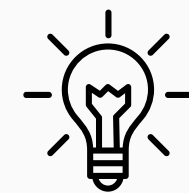
IFITV

프로젝트 개요



IFITV란 무엇인가?

01



맞춤형
추천

02



손동작
인식

03



사용자
경험

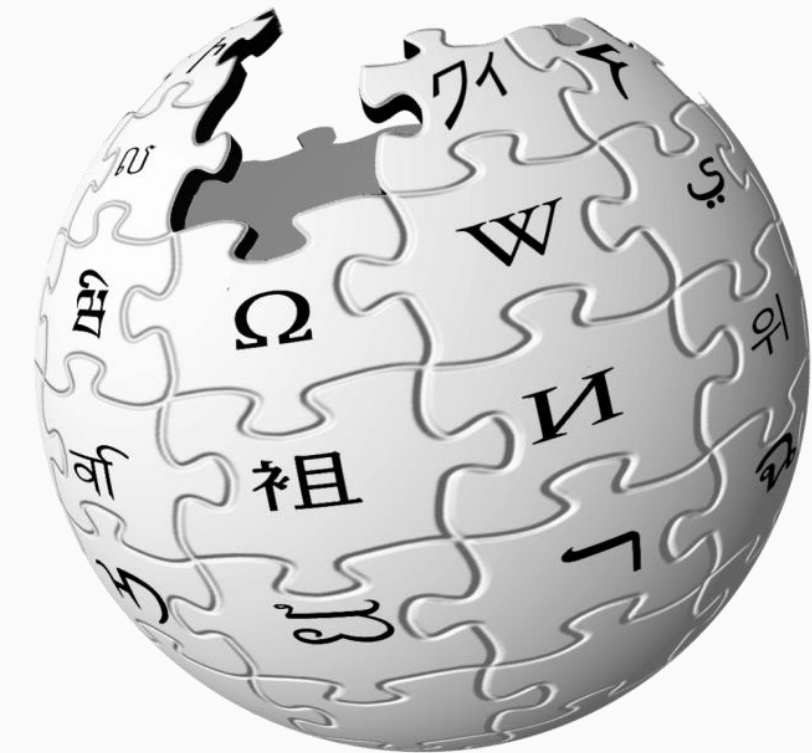
IFITV

데이터 정의 및 수집 가능성

✓ 데이터 소스



THE
MOVIE
DB



WIKIPEDIA



TVMAZE

NAVER
Web Search

데이터 정의 및 수집 가능성

✓ 콘텐츠 메타 데이터

- 콘텐츠 메타 데이터
제목, 설명, 장르, 서브장르

TMDB 영화 TV 프로그램 인물 More

개요 미디어 팬덤 공유

범죄도시 4 (2024)
15 2024/04/24 (KR) • 액션, 범죄, 드라마, 스릴러, 코미디 • 1h 49m

68% 회원 점수
당신의 바이브는 어떨까요? 0

트레일러 재생

거침없이 싹쓸어버린다

개요
신종 마약 사건 3년 뒤, 괴물형사 마석도와 서울 광수대는 배달앱을 이용한 마약 판매 사건을 수사하던 중 수배 중인 앱 개발자가 필리핀에서 사망한 사건이 대규모 온라인 불법 도박 조직과 연관되어 있음을 알아낸다. 필리핀에 거점을 두고 납치, 감금, 폭행, 살인 등으로 대한민국 온라인 불법 도박 시장을 장악한 특수부대 용병 출신의 빌런 백창기와 한국에서 더 큰 판을 짜고 있는 IT업계 천재 CEO 장동철. 마석도는 더 커진 판을 잡기 위해 장이수에게 뜻밖의 협력을 제안하고 광역수사대는 물론, 사이버수사대까지 합류해 범죄를 소탕하기 시작하는데...

Kim Min-sung
Characters

강윤성
Characters

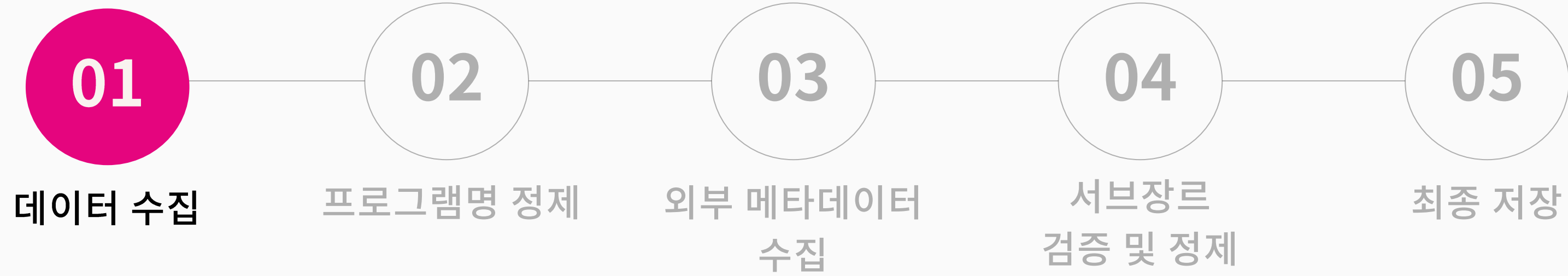
차우진
Characters

허명행
Director

오상호
Writer

지금 스트리밍 중
시청하기

데이터 수집 아키텍처 설계



데이터 수집

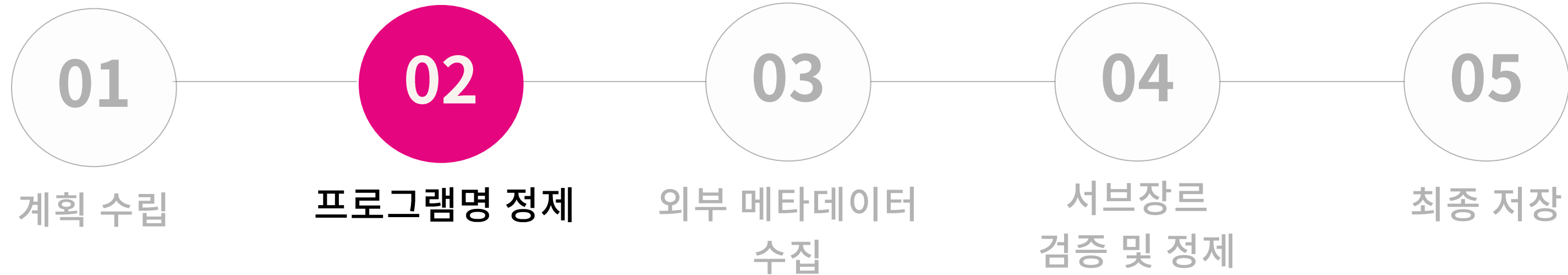


LGU+

Beautifulsoup



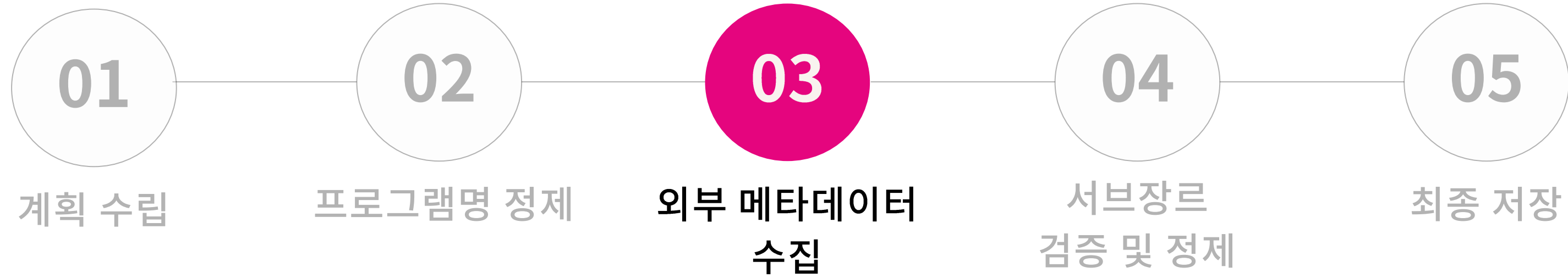
데이터 수집 아키텍처 설계



프로그램명 정제

- 함수 : `clean_name()`
- 목적 : 괄호, 특수문자, 회차 정보, 중복 공백 제거
- 결과 : TMDb, NAVER 등 API 호출용 클린 타이틀 확보

데이터 수집 아키텍처 설계



외부 메타데이터 수집

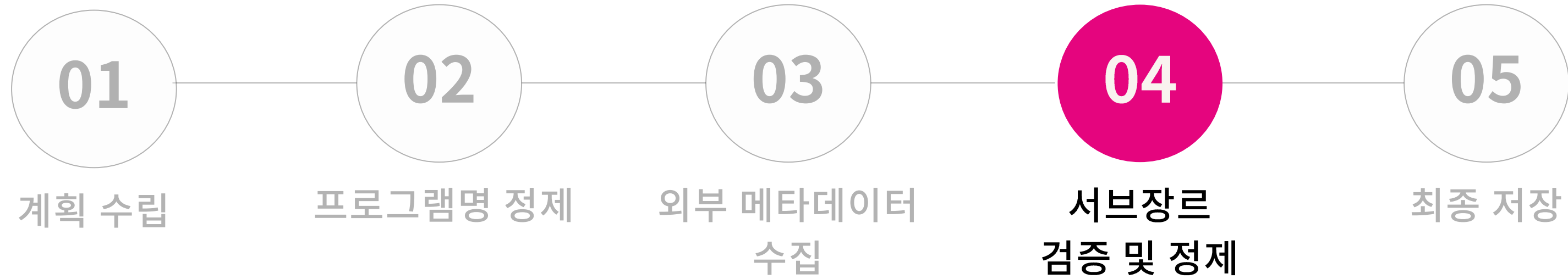
THE
MOVIE
DB

TVMAZE



NAVER

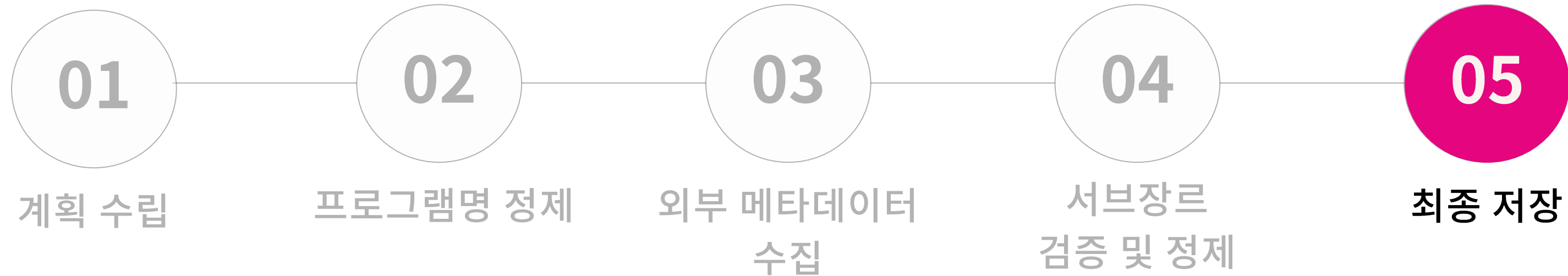
데이터 수집 아키텍처 설계



서브장르 검증 및 정제

- 허용된 서브장르 목록(allow_subgenres_by_genre)과 대조
- TMDb/ Tvmaze / Wikipedia / NAVER의 서브장르 추론 결과가 유효하지 않으면 키워드 기반 fallback(desc_keywords.json) 사용
- 예능에 드라마 서브장르가 들어가지 않도록 필터링

데이터 수집 아키텍처 설계



최종 저장

- Pandas DataFrame 정리
- CSV 저장 : ./data__live/{채널명}_crawlingprogram_list.csv

데이터 수집 및 전처리

LG U+ 실시간 편성표 수집

실시간 편성표 내 프로그램명 기반 콘텐츠
메타데이터 수집



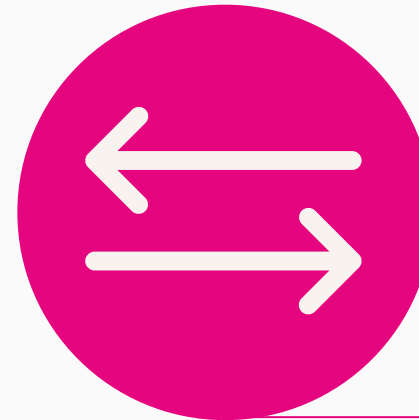
중복 콘텐츠 제거

동일 콘텐츠의 중복을 식별하고 제거



Tving 내 콘텐츠 수집

Tving 내 콘텐츠 장르, 서브장르 크롤링



결측값 처리

결측값을 '정보없음'으로 처리

데이터 검증 및 분석

✔ 라이브러리 불러오기

```
import os
import re
import time
import pandas as pd
import requests
from bs4 import BeautifulSoup
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.chrome.options import Options
import html
from urllib.parse import quote
from datetime import datetime, timedelta
import json
from dotenv import load_dotenv
```

데이터 검증 및 분석

✓ 크롤링 하고자 하는 채널 리스트 선언

```
# 채널 리스트
channel_list = [
    # 전국 지상파
    'KBS1[9]', 'KBS2[7]', 'MBC[11]', 'SBS[5]',

    # 종편 + 공영 + 교양
    'JTBC[15]', 'MBN[16]', '채널A[18]', 'TV조선[19]',
    'EBS1[14]', 'EBS2[95]', 'OBS[26]',

    # 드라마/예능/영화 전문 채널
    'tvN[3]', 'OCN[44]', '스크린[46]', '씨네프[47]', 'OCN Movies2[51]',
    '캐치온1[52]', '캐치온2[53]', '채널액션[54]',
    '드라마큐브[71]', 'ENA[72]', 'ENA DRAMA[73]',
    'KBS Story[74]', 'SBS플러스[33]', 'MBC드라마넷[35]',

    # 애니메이션/키즈 채널
    '투니버스[324]', '카툰네트워크[316]'
]
```

데이터 검증 및 분석

✓ LG U+ 사이트 내 실시간 편성표 접속 및 사이트 내 버튼 변수 선언

```
# 크롬 드라이버 설정
options = Options()
options.add_argument('--headless')
options.add_argument('--no-sandbox')
options.add_argument('--disable-dev-shm-usage')

driver = webdriver.Chrome(options=options)
wait = WebDriverWait(driver, 10)

url = 'https://www.lguplus.com/iptv/channel-guide'
table_btn_xpath = '//a[contains(text(), "채널 편성표 안내")]'
all_channel_btn_xpath = '//a[contains(text(), "전체채널")]'
```

데이터 검증 및 분석

✓ 반복문을 사용해 채널 리스트의 티비 프로그램 정보 수집

```
# 채널별 반복 크롤링
for channel in channel_list:
    try:
        driver.get(url)
        driver.execute_script("document.body.style.zoom='50%'")
        time.sleep(1)

        wait.until(EC.element_to_be_clickable((By.XPATH, table_btn_xpath))).click()
        time.sleep(1)

        wait.until(EC.element_to_be_clickable((By.XPATH, all_channel_btn_xpath))).click()
        time.sleep(2)

        # 채널 팝업 다시 열기
        wait.until(EC.element_to_be_clickable((By.CSS_SELECTOR, "a.c-btn-outline-2-s.open"))).click()
        time.sleep(1)

        # 채널 버튼 클릭
        channel_xpath = f'//a[contains(text(), "{channel}")]'
        wait.until(EC.element_to_be_clickable((By.XPATH, channel_xpath))).click()
        time.sleep(2)

        page_source = driver.page_source
        soup = BeautifulSoup(page_source, 'html.parser')
        program_soup_list = soup.select('tr.point')

        program_list = []
```


데이터 검증 및 분석

✓ 반복문을 사용해 채널 리스트의 티비 프로그램 정보 수집

```
for item in program_soup_list:
    try:
        tds = item.select('td')
        time_text = tds[0].text.strip()
        name_parts = tds[1].text.split('\n')
        raw_name = name_parts[1].strip() if len(name_parts) > 1 else tds[1].text.strip()
        name = clean_name(raw_name)

        if name in ["방송 시간이 아닙니다", "방송시간이 아닙니다."]:
            continue

        genre = genre_map.get(tds[2].text.strip(), tds[2].text.strip())

        result = get_program_metadata(name, driver, genre)
        if not result:
            print(f"[메타데이터 없음] {name}")
            continue
        original_genre, sub_genre, desc, thumbnail = result
        program_list.append([time_text, name, original_genre, sub_genre, desc, thumbnail])

        time.sleep(0.2)
    except Exception as e:
        print(f"[프로그램 처리 오류] {e}")
        continue
```

데이터 검증 및 분석

✓ 반복문을 사용해 채널 리스트의 티비 프로그램 정보 수집

```
# 런타임 계산 추가 적용
program_list = calculate_runtime(program_list)

# 결과 저장
safe_name = re.sub(r'\s*(\[^\]]*\)', '', channel).strip()
df = pd.DataFrame(program_list, columns = ['airtime', 'title', 'genre', 'subgenre',\
                                         'runtime', 'desc', 'thumbnail'])

# ✂ 쌍따옴표 제거
df['subgenre'] = df['subgenre'].apply(
    lambda x: x.replace('"', '') if isinstance(x, str) else x
)

# 저장
df.to_csv(f'./data_crawling_live/{safe_name}_program_list.csv', index=False, encoding='utf-8-sig')

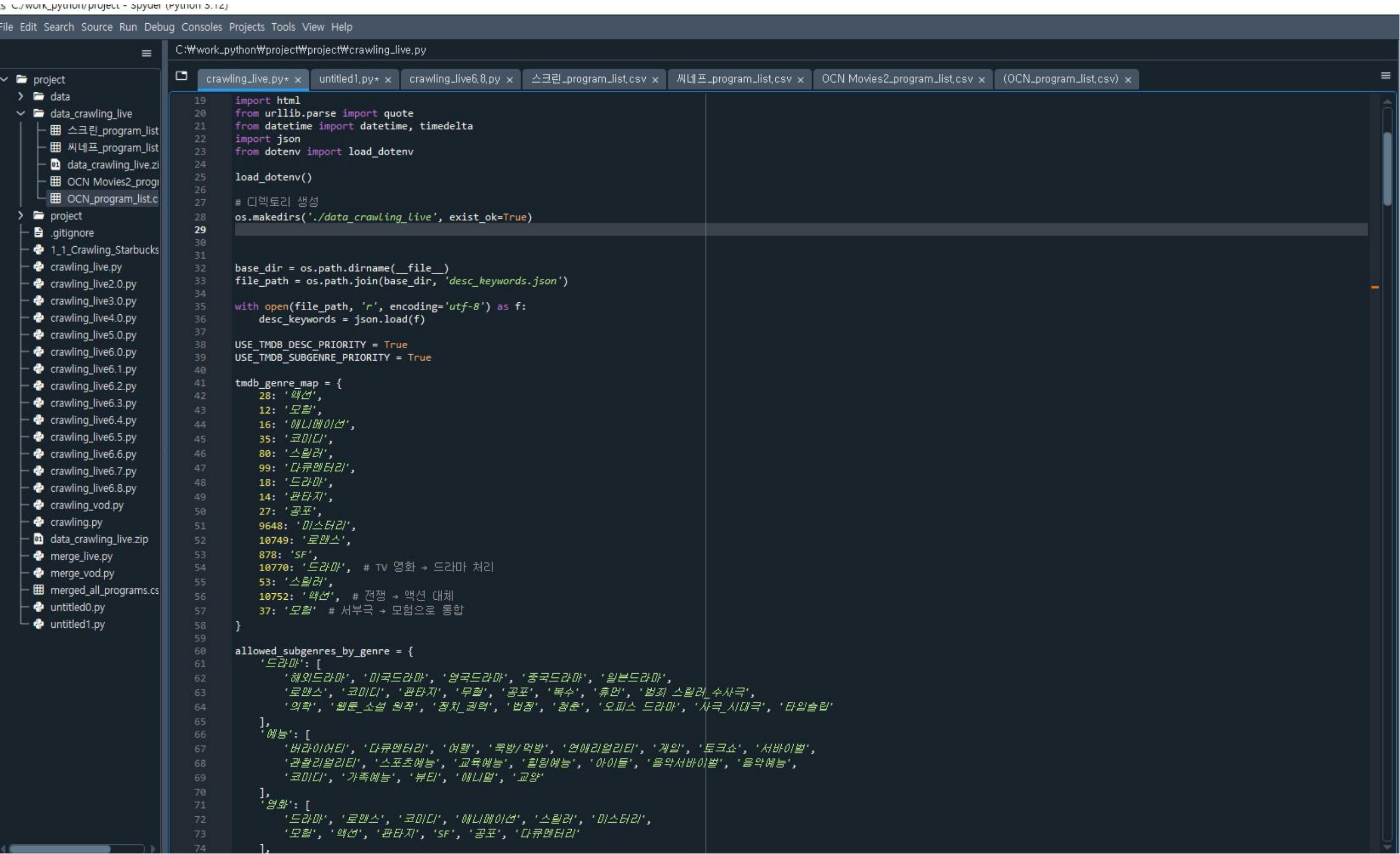
print(f"[완료] {channel} → 저장 완료")
time.sleep(1)

except Exception as e:
    print(f"[채널 오류] {channel} 처리 중 오류: {e}")
    continue

# 드라이버 종료
driver.quit()
print("[전체 완료] 모든 채널 크롤링 종료")
```

데이터 검증 및 분석

✓ 시연 영상

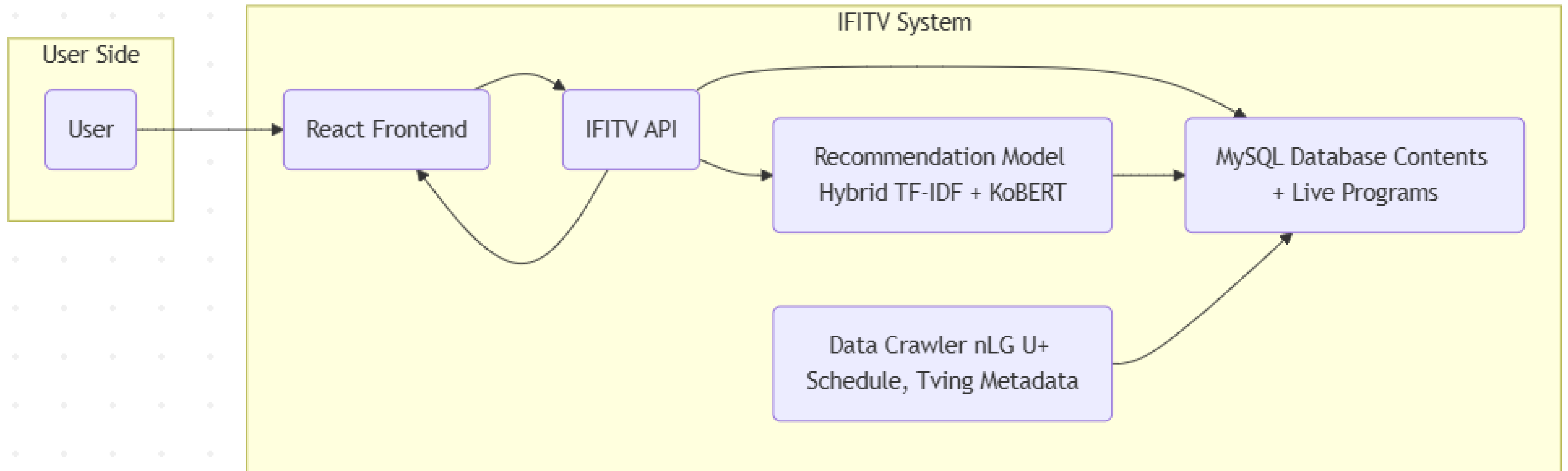


데이터 검증 및 분석

✔ Chnnel_list에서 방영시간, 방영 프로그램, 장르와
get_program_metadata로 프로그램 메타 데이터 추출

airtime	title	genre	subgenre	runtime	desc	thumbnail						
0:10:00	데스티네이션	영화	미스터리, 공포	110	끝없는 악몽	https://image.tmdb.org/t/p/w500/oS0b0iiwEW2T9vcGHFYpTlTPmul.jpg						
2:00:00	이브의 유혹 좋은 아내	영화	로맨스, 드라마	100	오늘도 영혼	https://image.tmdb.org/t/p/w500/vchR34tlwZRsf7XpatDGbKYava.jpg						
3:40:00	배틀 크랙	영화	액션, 공포, SF	100	2242년, 기	https://image.tmdb.org/t/p/w500/8ouNEhnjjX6NVnzhmHtdpTTbBs1.jpg						
5:20:00	주유소 습격사건 2	영화	코미디, 액션	120	노마크에게	https://image.tmdb.org/t/p/w500/28WZCtwJH9Is0QUVmOABWbrcZad.jpg						
7:20:00	스텔라	영화	SF, 모험, 드라마	100	세계 각국	https://image.tmdb.org/t/p/w500/evoEi8SBSvIIEveM3V6nCJ6vKj8.jpg						
9:00:00	퀵	영화	코미디, 스릴러, 액션	130	스피드 마	https://image.tmdb.org/t/p/w500/bnPd7UzSW8ZM0Aca5XI9woAl8IB.jpg						
11:10:00	쉬즈 더 맨	영화	코미디, 로맨스	120	말괄량이	https://image.tmdb.org/t/p/w500/bzJhzqDsRmoxdYLueZO50MUDEGk.jpg						
13:10:00	마이너리티 리포트	영화	액션, 스릴러, SF	150	2054년 워	https://image.tmdb.org/t/p/w500/oaVX63mgvSN4MRN98sg5molfKBd.jpg						
15:40:00	정도	영화	판타지, 모험, 액션	130	도성의 무	https://image.tmdb.org/t/p/w500/uvD7vPBlpzejysF9e6fQbT5n86O.jpg						
17:50:00	미션 임파서블	영화	스릴러, 모험, 액션	130	디지털상의	https://image.tmdb.org/t/p/w500/5Zxdorl5Tharll9S47YxoKzGCsi.jpg						
20:00:00	극한직업	영화	코미디, 스릴러, 액션	120	불철주야	https://image.tmdb.org/t/p/w500/jbHNkNydiZstlqhhBSvG19lm4NL.jpg						
22:00:00	소풍	영화	드라마	60	여든의 은	https://image.tmdb.org/t/p/w500/65rPts9EIWyd6qLBwX1CKGHNIPN.jpg						

시스템 아키텍처 설계



데이터 시각화 및 Dashboard 생성

✔ User가 선택한 콘텐츠 메타데이터와 유사한 콘텐츠 추천

```
>>> print(result_df)
```

	선택_프로그램	추천_VOD	장르	서브장르	추천_근거	최종_점수
0	데스티네이션	콰이어트 플레이스 2	영화	스릴러, 공포, SF	장르 겹침: ['공포'] 설명 키워드 겹침: ['공포', '죽음', '실체']	0.422610
1	데스티네이션	소풍	영화	드라마	설명 키워드 겹침: ['죽음', '자신', '영화']	0.396851
2	데스티네이션	걸프렌즈	영화	판타지, 코미디, 로맨스	설명 키워드 겹침: ['영화', '자신']	0.311240
3	데스티네이션	그녀가 죽었다	영화	스릴러, 미스터리	장르 겹침: ['미스터리'] 설명 키워드 겹침: ['영화', '미스터리']	0.307889
4	데스티네이션	사라진 밤	영화	스릴러, 미스터리	장르 겹침: ['미스터리'] 설명 키워드 겹침: ['영화', '미스터리']	0.290355
5	데스티네이션	정도	영화	판타지, 모험, 액션	설명 키워드 겹침: ['이제', '다음', '영화']	0.285861
6	데스티네이션	더 파일럿 드론 어택	영화		설명 키워드 겹침: ['가족', '영화']	0.279349
7	데스티네이션	황후화	영화	판타지, 액션, 드라마	설명 키워드 겹침: ['가족', '영화', '자신']	0.278630
8	데스티네이션	파주	영화	로맨스, 드라마	설명 키워드 겹침: ['영화', '자신']	0.274825
9	데스티네이션	언더월드 블러드 워	영화	판타지, 스릴러, 액션, SF	설명 키워드 겹침: ['영화', '자신']	0.274747

앞으로 구현할 기능 & 기대효과

✔ 앞으로 구현할 기능

데이터베이스 구축

사용자 정보,
콘텐츠 메타데이터 관리



실시간 추천

실시간 편성표 기반 맞춤형 콘텐츠 추천



멀티 프로필

사용자별 맞춤 프로필로 연령과
선호에 최적화된 추천을 제공



UI/UX 고도화

인터페이스 고도화로 사용자 경험 향상



앞으로 구현할 기능 & 기대효과

✓ 앞으로 구현할 기능

서비스 확장



OTT, 스마트TV 등으로
서비스 영역을 확대



사용자 경험 강화



빠른 로그인과 개인화로
편의성을 극대화



맞춤형 IPTV 시스템



완벽한 개인 맞춤형
시청 환경을 구현



2025.05.16

감사합니다

백지연, 이경준, 이에림, 최동연