

# 중간발표 자료

## 주제 선정



우리가 즐기면서 할 수 있는 주제를 잡자!

### 1. 한국 커뮤니티 대시보드

커뮤니티 게시글 / 댓글 크롤링을 통해 커뮤니티별 특징 분석

1. 자주 나오는 단어들 워드클라우드 or 토픽모델링
2. 게시글 / 댓글이 달리는 시간대 시각화

등등... 가능하지 않을까 생각했습니다.

### 커뮤니티 사이트 순위 2023년 05월

1위	<b>디시인사이드</b> dcinside.com 종합커뮤니티	206M
2위	<b>에펨코리아</b> fmkorea.com 종합커뮤니티 스포츠	103.8M
3위	<b>뽐뿌</b> ppomppu.co.kr 쇼핑	54.4M
4위	<b>루리웹</b> ruliweb.com 게임	48.7M
5위	<b>더쿠</b> theqoo.net 여성	44.6M

인기 커뮤니티 사이트 목록도 찾았습니다. 각자 두 개씩 맡아서... 한 달치 크롤링 하는 거죠...

일부 커뮤니티에 대해서 폭력적인 게시글이 많을까 걱정되긴 합니다.

### 2. BOJ 크롤링

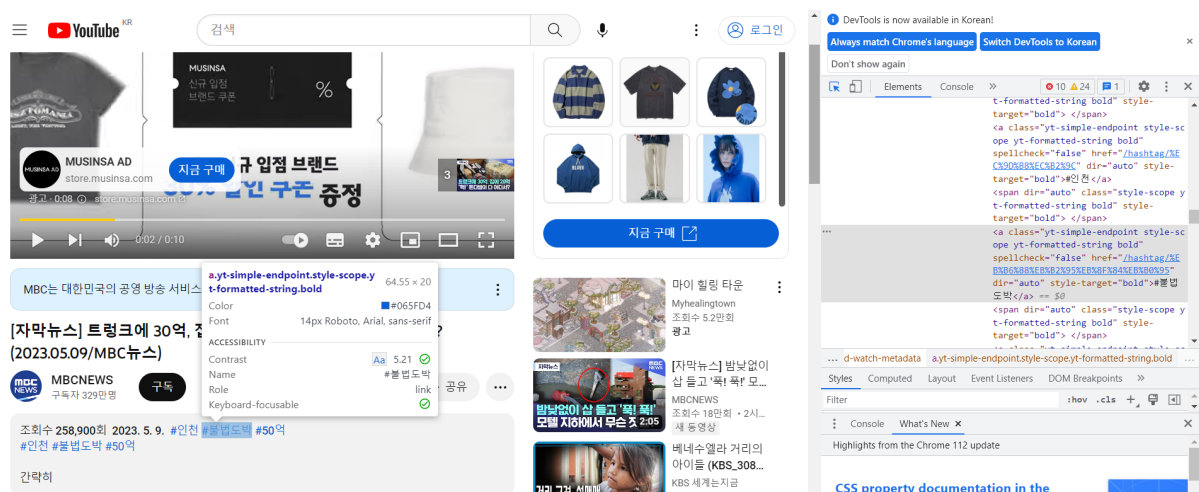
크롤링할 만한 사이트가 뭐가 있을까 하다가 알고리즘 풀이 사이트가 생각났습니다.

<https://www.acmicpc.net/status?top=23000>

채점 현황에서 13년 전 채점데이터까지 제공하고 있는 것을 확인했습니다, 엄청난요

시간의 흐름에 따른 언어 트렌드 변화나, 년도별로 인기있는 문제.. 등을 찾을 수도 있다고 생각했습니다.

### 3. 유튜브(동적 크롤링) 🏰



- 썸네일 or 영상 링크
- 제목
- 조회수
- 태그
- 댓글

⇒ 사진, 영상 댓글 등 크롤링 해올 데이터가 많아서 이것저것 많이 해볼 수 있을 듯!

- 주제
  - 태그별로 조회수 높은 영상 추천
  - 댓글 워드 클라우드
  - 추천 영상 제목 워드 클라우드

### 4. 크림

KREAM | 한정판 거래의 FLEX

한정판을 거래하는 가장 확실한 방법

 <https://kream.co.kr/>

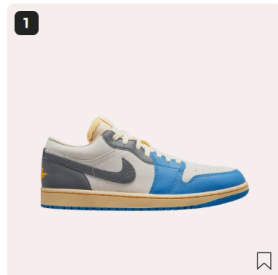
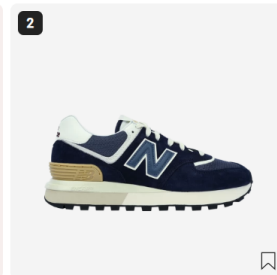
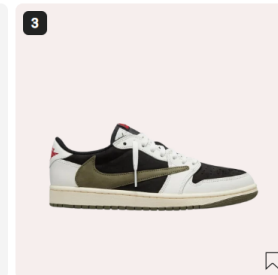
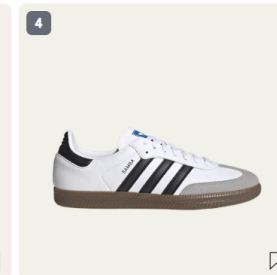


KREAM

HOME STYLE SHOP MY Q

힙크림 2차티켓 추천 **랭킹** 남성 여성 브랜드

지금 주목받는 신발

 Jordan Jordan 1 Low SE Tokyo 96 빠른배송 175,000원 즉시 구매가	 New Balance New Balance 574 Legacy Navy 빠른배송 180,000원 즉시 구매가	 Jordan (W) Jordan 1 x Travis Scott Retro Low OG SP Medium Olive 빠른배송 569,000원 즉시 구매가	 Adidas Adidas Samba Vegan White Black 빠른배송 164,000원 즉시 구매가
---	---	--	---

더보기

- 가격대 설정하면 내가 살 수 있는 물품 이미지로 띄워주기?
  - 이미지 클릭하면 상세 정보
- 트렌드(남성/여성/랭킹)
- 트렌드 키워드 → 워드 클라우드해도 될 듯
  - 상품 카테고리
  - 브랜드
  - 등등

## 5. 영화 리뷰 대시보드

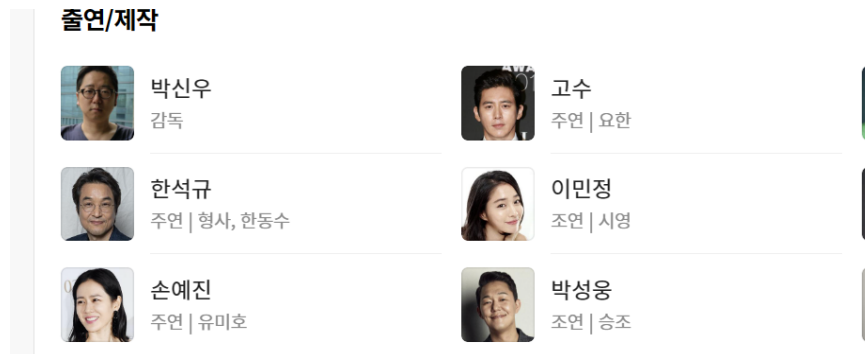
site : 왓차 피디아 [왓차피디아](http://watcha.com) - 영화, 책, TV 프로그램 추천 및 평가 서비스 ([watcha.com](http://watcha.com)).

다음영화 홈 | [다음영화](http://다음영화.daum.net) ([daum.net](http://다음영화.daum.net)).

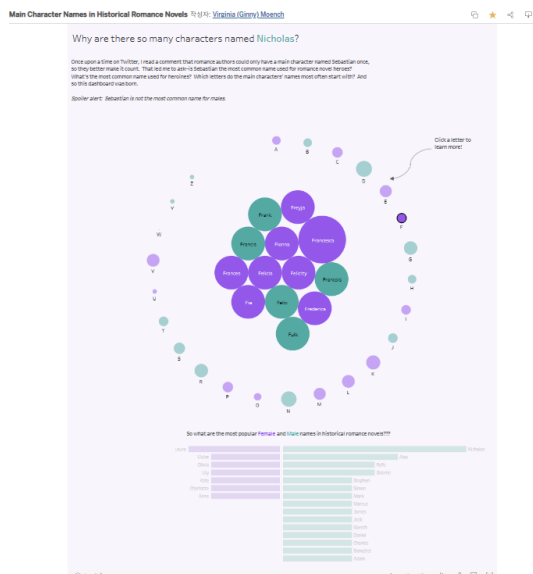
cf ) [Crawling](http://Crawling.velog.io) 크롤링 연습 2-1 다음 영화 리뷰 크롤링 ([velog.io](http://velog.io)).

- 영화 리뷰 워드 클라우드
- 장르별...

- 평점별...
- 국가별...
- 년도별...
- 배우별...
- 영화 캐릭터 이름 차트



왓차 사이트에 들어가면 이렇게 영화 캐릭터의 이름이 나오는데 이걸 이용해서



<https://public.tableau.com/app/profile/virginia.moench/viz/MainCharacterNamesinHistoricalRomanceNovels/MostPopularNames>

요런 느낌으로 영화 캐릭터 이름 버블 차트? 만들어도 재밌을 것 같습니다

## 6.전시회 정보 제공 대시보드

site : [ARTMAP | Art is Everywhere\(art-map.co.kr\)](http://ARTMAP.Art is Everywhere(art-map.co.kr))

- 전시회 키워드 관련 워드 클라우드
- 나이, 성별대별 랭킹

- 지역별로 전시회 지도

## 7. 무신사 크롤링 (지그재그나 브랜드 등 쇼핑몰 크롤링)

### 1. 특정 브랜드 정보 및 리뷰 크롤링

- 특정 브랜드(ex. 무신사 스탠다드) 선택
- 해당 브랜드 정보 크롤링
  - 상품 상세 정보
  - 좋아요 수 / 해시태그 / 구매 만족도
  - 상품 후기
    - 스타일 후기 / 상품 후기 / 일반 후기 中 최신순 / 인기순
    - 후기가 없는 경우도 有 ⇒ 후기순 정렬 후, n번째 페이지까지만 크롤링
- 상품 별 리뷰 내용
- 가격대에 따른 상품

### 2. 실시간 랭킹 상품

- 실시간 순위, 의류 브랜드명, 의류명, 링크
- 상품 상세 정보
- 성별, 조회수(1개월), 누적판매(1년), 가격 정보

## 8. TV 시청률 크롤링 (닐슨 코리아)

- 지상파 일일 시청률 (가구 / 시청자 수 top20)
- 실시간 시청률 정보를 찾아서 시간대 별 방송 편성해보는 대시보드를 만들고 싶었는데 ...
- 실시간 시청률 정보를 제공하는 사이트가 없어서, n일치 시청률 비교

### [코드잇] TV 시청률 최종 프로젝트

티비랭킹닷컴의 데이터를 DataFrame으로 만들어서 분석해보려 합니다. 주어진 결과 예시와 같은 DataFrame을 만들어 보세요. import time import pandas as pd import requests from bs4 import

 <https://dlsalfkd11.tistory.com/156?category=920775>

period	rank	channel	progra
.28 ~ 2010.01.03	1	KBS2	주말연속극(수상한삼형제)
.28 ~ 2010.01.03	2	KBS1	일일연속극(다함께차차차)
.28 ~ 2010.01.03	3	KBS2	해피선데이
.28 ~ 2010.01.03	4	MBC	MBC연기대상2
.28 ~ 2010.01.03	5	SBS	주말극장(천만번사랑해)

## 9. 멜론

시대별 가사, 장르별 가사 비교

트렌드가 어떻게 변했는

플레이리스트

- 하트, 곡수, 태그, 시기 크롤링해서
- 시기별 플레이리스트 평균 하트 수 변화(이 시기에는 이러한 몇명의 사람들이 하트를 눌렀고~, 줄었다가~)
- 시기별 인기 태그 변화
- 하트 많이 받은 플레이리스트는 어떤 특징있는지(태그, 노래수 등)

DJ 플레이리스트 정보



Morning Coffee

모닝 커피: 여유롭게, 분위기 있는 카페테라스에서  
DJ 멜균 

2023.05.08 수정

[#커피](#) [#카페](#) [#분위기](#) [#휴식](#) [#테라스](#) [#편안한](#) [#어쿠스틱](#) [#여유](#) [#아침](#)






 7,379   신고 >



소개글

하늘을 보며 여유를 느끼고 싶은 날, 잠시 머릿속 복잡한 생각들은 내려놓고  
살랑이는 바람 같은 선율과 기분 좋은 멜로디로 지금, 이 플레이스트와 함께 커피 한 잔 어떠세요?  
예전 노래들과 신곡들을 적절히 선곡하였습니다.

#분위기 #휴식 #커피 #카페 #테라스 #편안한 #어쿠스틱 #여유 #아침

□ 번호	곡정보	앨범	좋아요	듣기	담기	다운	유비
□ 1	 <b>Call U After Rehab</b> Devon Cole	Call U After Rehab	♡ 152	▶	+	↓	□ <sup>q</sup>
□ 2	 <b>California</b> Paige	California	♡ 901	▶	+	↓	□ <sup>q</sup>
□ 3	 <b>Honeymoon</b> Johnny Stimson	Honeymoon	♡ 20,782	▶	+	↓	□ <sup>q</sup>
□ 4	 <b>Morning Coffee</b> Chevy, Nalba	Morning Coffee	♡ 1,567	▶	+	↓	□ <sup>q</sup>
□ 5	 <b>Too Much To H8</b> Paige	Always Growing EP	♡ 16,515	▶	+	↓	□ <sup>q</sup>







치열한 접전 끝에 유튜브 대시보드를 만들기로 되었습니다!

## 주제 구체화 회의



채널 선정 + 대시보드 흐름 정하기 + 유튜브 크롤링

### 1. 채널선정

유튜브의 어떤 부분을 크롤링 하는 것이 좋을까?

→ 채널을 정해서 그 채널에 대한 정보를 제공하는 대시보드를 만들자!

유튜브 운영을 잘하고 있는 사례를 분석해서

→ 그럼 어떤 채널을 선정하는 게 좋을까?

그래도 근거를 가지고 선정하는 게 좋지 않을까?

#### SNS 랭킹

유튜브 랭킹, 인스타 랭킹, 아프리카TV 랭킹, 페이스북 랭킹, Vlive 랭킹, 트위터 랭킹, 순위, 구독자, 크리에이터, 유튜버, 영상, 조회 수, 인기동영상, 순위 실시간/일간/주간 차트

 <https://youtube-rank.com/#>











TOP 300	게임	BJ/인물/연예인	음악/댄스/가수	TV/방송	음식/요리/레시피	패션/미용	뉴스/정치/사회	취미/라이프	IT/기술/컴퓨터
회사/오피셜	교육/강의	영화/만화/애니	키즈/어린이	애완/반려동물	스포츠/운동	국내/해외/여행	자동차	주식/경제/부동산	미분류
전체순위	구독자순	동영상순	재생순	조회순	이름	카테고리순	Random		

여기 있는 구분 방식을 활용해도 좋을 것 같기도 해서 우선 가져왔습니다.



그런데 차트의 상위에 있는 채널들이 우리가 자주 보는 채널이 아니었습니다.

주제를 더 좁히는 것도 괜찮지 않을까? 주제를 (예를 들어) 코미디로 잡아볼까?

인플루언서 랭킹		한국	Nox카테고리	코미디
기본 정보	구독자	평균 조회수	NoxScore	
1 -  영국남자 Korean Englishman	560만 ↑ 0.2%	228.5만 ↓ 1.8%	★★★★★	
2 -  CuRe 구래	519만 -	810.17만 -	★★★★★	
3 ↓ 1  장빠꾸	355만 ↑ 0.3%	109.82만 ↑ 1.9%	★★★★★	
4 ↓ 1  슈카월드	268만 ↑ 0.4%	70.46만 ↓ 2.1%	★★★★★	
5 ↓ 1  숏박스	252만 ↑ 0.4%	367.65만 ↓ 0.9%	★★★★★	
6 ↓ 1  피식대학Psick Univ	208만 ↑ 1.5%	71.03만 ↑ 13.2%	★★★★★	
7 ↓ 1  대생 Daesaeng	204만 ↑ 0.5%	9.08만 ↑ 0.9%	★★★★★	
8 ↓ 1  퇴경아 약먹자	198만 -	40.04만 ↓ 1.5%	★★★★★	

어차피 저런 차트를 그대로 사용한다고 해도, 슈카월드 같은 채널들은 코미디로 넣기 어려워서 저희가 직접 고르는 작업을 하긴 해야할 것 같아요. 슈카월드는 코미디로 분류되는데 미미미누는 교육으로 들어가 있습니다.

침착맨 / 오징 / 숏박스 / 피식대학 등으로 좀 유명하다 싶은 걸 **우리끼리 골라야 하긴 할 것 같습니다.**



근데 어차피 우리가 선택해야 한다면, 굳이 차트를 참고해야 하나?



우리에게 익숙한 주제(웹드라마, 코미디, 플레이리스트 등) 에서 우리가 자주 보는 채널로 고르자!

- 음악(플리) : 에센셜 (120만)
- 웹드라마 : 짧은대본 (114만)
- 코미디 / 예능 : 피식대학 (209만)
- 요리 / 먹방 :
- 게임 : 우왁굳(153만)

- 뷰티 : 이사배(226만) /
- 패션 : 깡스타일리스트(122만) / 주우재
- 여행 : 빠니보틀 / 곽튜브 /
- 브이로그 :
- IT : 잇섭
- 운동 : 땅끄부부(최근 영상이 없어서 △) . 흑자헬스 (64만), 권혁 (68만), 피지컬갤러리(기획영상 말고 운동 영상들만 크롤링해도... 될지도...)



세 개 정도 하면 되지 않을까?

에센셜, 웹드라마, 코미디의 세 개로 결정되었습니다.

## 2. 유튜브 크롤링



모두가 힘을 합쳐 크롤링을 진행했습니다.

### 우리가 가진 데이터

- 채널 내
  - 모든 영상 각각 제목
  - 모든 영상 조회수
  - 모든 영상 링크
  - 모든 영상 길이
  - 모든 영상 썸네일 이미지 링크
- 동영상 화면
  - 모든 영상 각각 좋아요 수
  - 모든 영상 댓글 수
  - 모든 영상 업로드 날짜

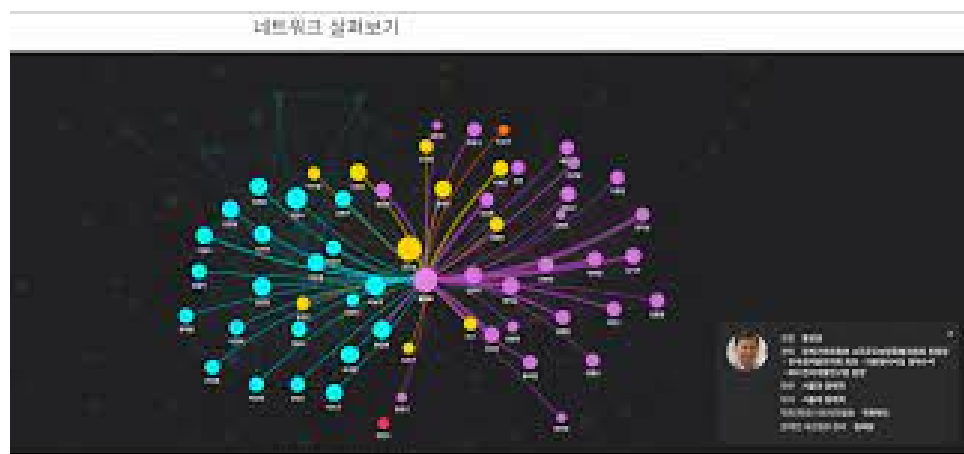
- 모든 영상 태그
- 선택된 영상 댓글 내용
- 선택된 영상 댓글 좋아요
- 선택된 영상 댓글 작성자(ID)
- 재생목록
  - 재생목록 제목
  - 재생목록 영상 수
  - 재생목록 링크

### 3. 대시보드 흐름 정하기

## 만들 차트

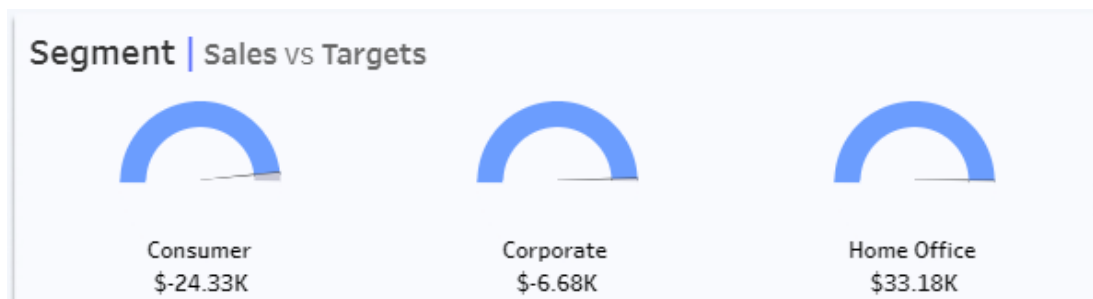
### 지금 할 수 있는 차트

- 댓글 워드클라우드 (선택된 영상)
- 댓글 토픽모델링 (네트워크 차트, 위에 지후가 그려둠)



- 제목 워드클라우드
- 좋아요 대비 댓글 비율 (유의미할지는 잘 모르겠지만 함, 에센셜이랑 짧대랑은 좀 차이가 보일듯)
- 구독자 수 / 동영상 수 / 예상 수익 (애네는 그냥 .. 손으로 크롤링)
- 구독자 수 증가 추이 (이 데이터가 아직 없긴 한데 유튜브 외부에서 찾아서 할 수 있을지도 몰라)
- 영상 업로드 달력 만들기

- 시간 흐름에 따른 제목 조회 수
- 조회수 대비 영상 길이 산점도
- 재생목록별 영상 수 막대차트
- 조회 수 막대
- 버터플라이 차트( 이전 년(2022)도와 해당(2023) 년도의 월별 조회수 비교)
- 지표
  - \* 조회수 대비 좋아요수 : 영상의 긍정도
  - \* 조회수 대비 댓글 수 : 영상의
  - \* 구독자수 대비 좋아요수 : 채널의 긍정도?

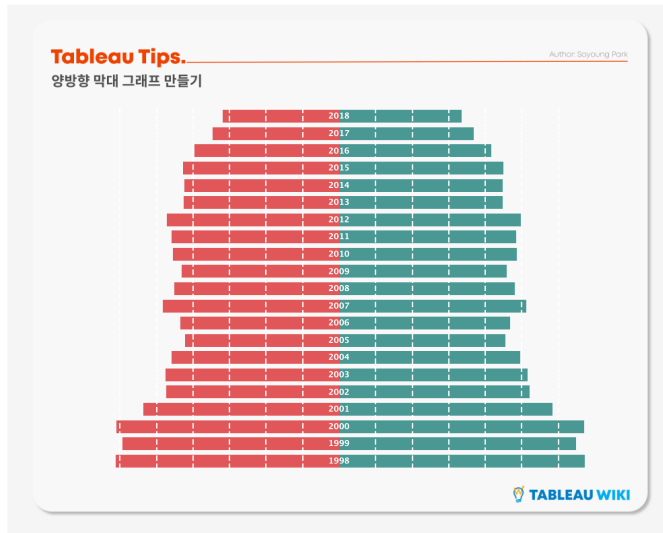


게이지 차트로 만들어봐도 좋을 듯 (가능하다면 ....)

 [https://public.tableau.com/app/profile/tableauwiki/viz/\\_16570383606250/sheet0](https://public.tableau.com/app/profile/tableauwiki/viz/_16570383606250/sheet0)

게이지 차트 참고용

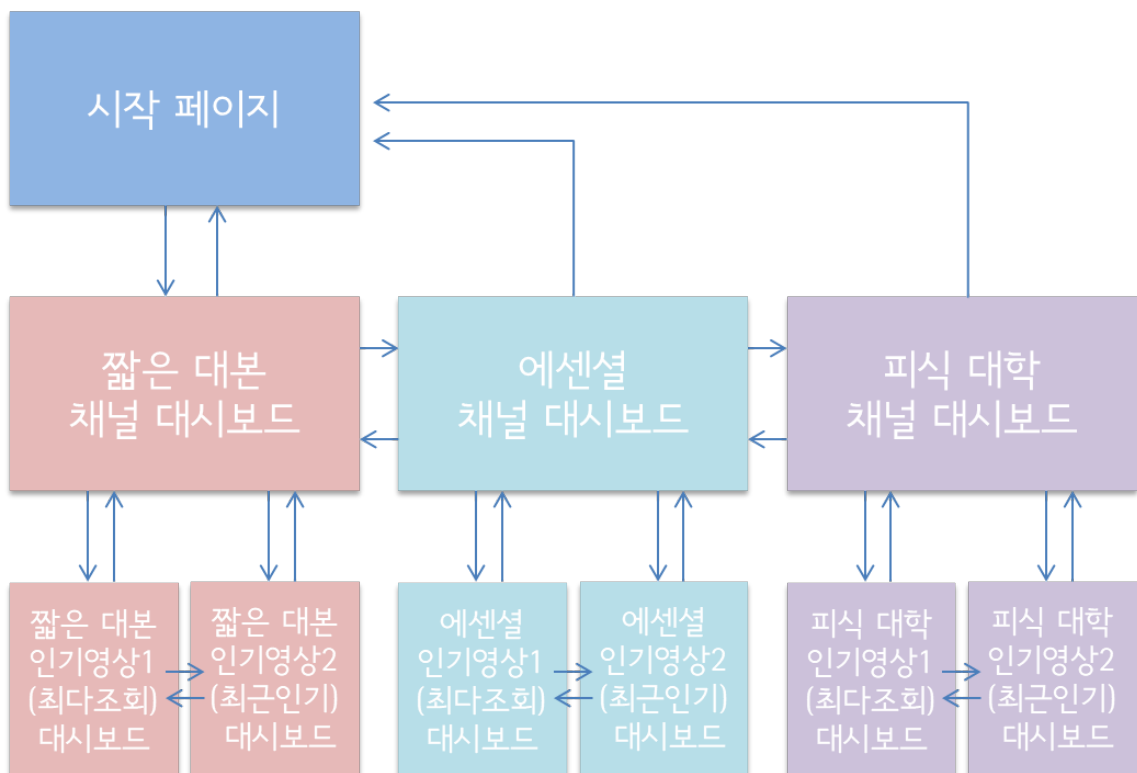
- 레이더차트 (댓글 수 / 조회 수 / 좋아요 수 / 동영상 수 / 영상 길이- 초 단위)



나비차트

### 지금 할 수 없는 차트

- 태그명에 따른 조회수 / 좋아요수 평균
- 태그별 순위
- 태그 워드클라우드



## 진행상황

크롤링 진행중 → 영상에 붙은 태그 크롤링이 뭔가 잘 안 되어서 진행중에 있습니다.

각 대시보드 페이지에 들어갈 차트 레이아웃 → 오늘 정할 예정입니다.

데이터 전처리 → 크롤링 진행된 데이터들부터 천천히 진행하고 있습니다.