

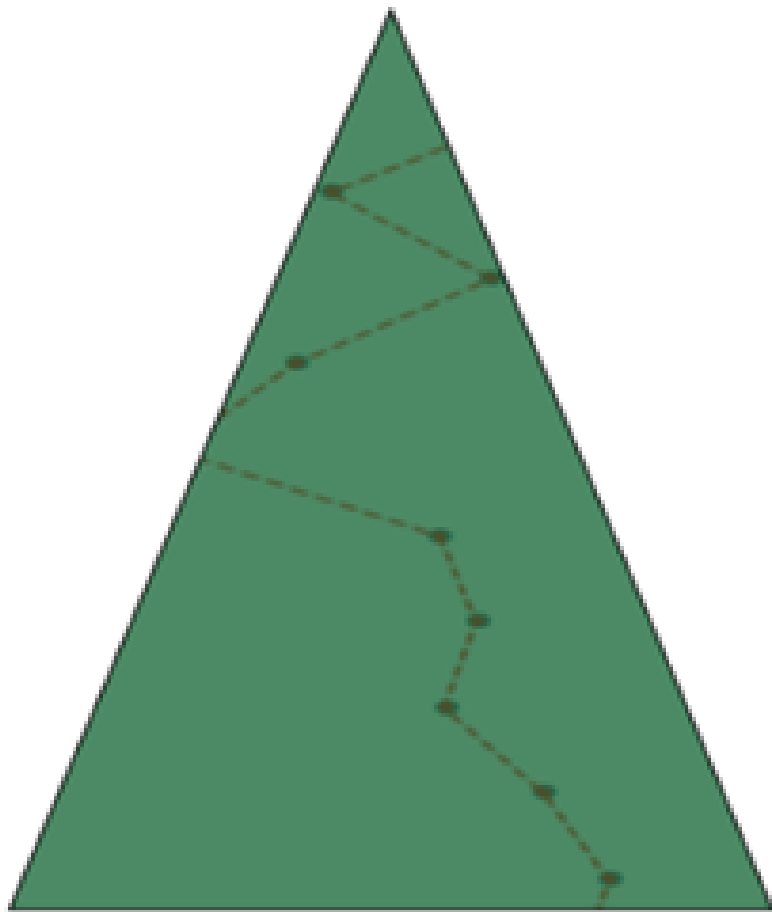
유튜브로 바라본

V-Choice



9팀 김준수 고통성 고주연 김방현 김성규 최청현

목차



01 서비스 소개

02 주요 기능

03 데모 시연

04 개별발표

05 Q&A

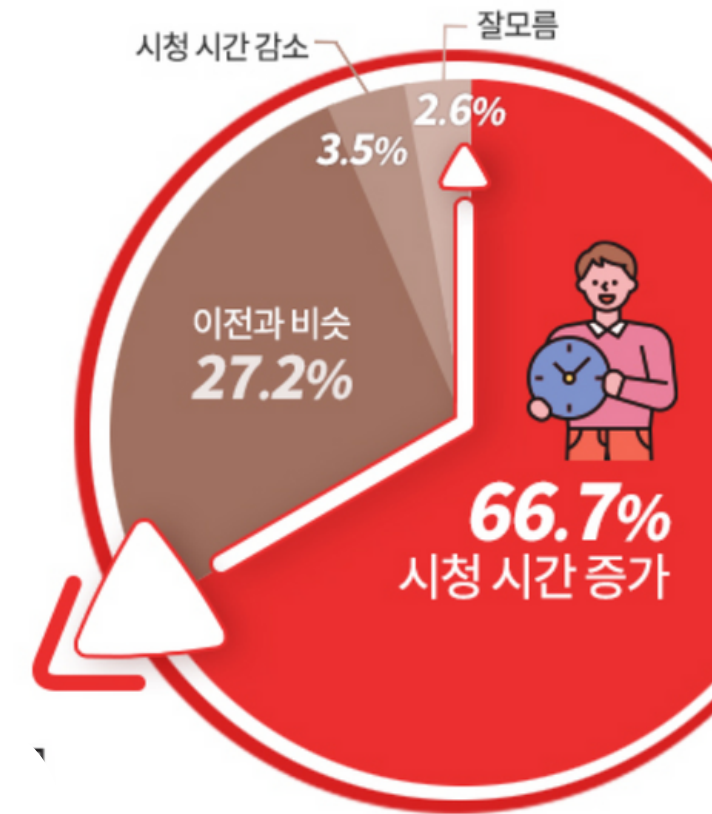
코로나와 유튜브?

데이터를 가지고 놀 수 있는 **놀이터**

인사이트를 **공유**



(단위: %)
코로나19 이후 (2020년 1월~)
유튜브 시청 시간 변화 여부



팀 개발 원칙 : 배운 것을 최대한 활용하자

대상

- 유튜브 크리에이터, 유튜브 시청자 등 유튜브 트렌드가 궁금한 사람들
- 코로나와 유튜브의 상관관계가 궁금한 사람들

차별성

- 단순히 코로나로 인해 유튜브 사용자가 늘었다가 아니라, 구체적으로 코로나가 유튜브 영상에 어떤 영향을 끼쳤는지 확인 가능
- 총 **622개**의 방대한 시각화 자료 제공
 - 예시) 카테고리 선택 + 코로나 정보 선택 + 유튜브 정보 선택(조회수..) 와 같이 선택
- 유튜브에 특화된 감성사전을 구축하여 유튜브 제목에서 드러나는 **긍부정 분석** 제공
- 다수의 이용자와 함께 **의논하며 직접 인사이트 도출** 가능

방법

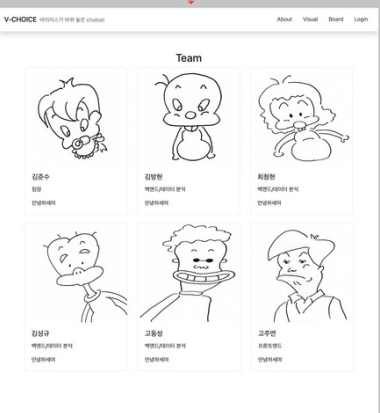
- **회원가입** 및 **로그인** 후 코로나와 유튜브의 상관관계에 대한 기존 의견 수집
- 코로나와 유튜브 트렌딩 비디오의 관계를 시각화한 **그래프**들 자유롭게 확인
- 확인 후 코로나와 유튜브의 상관관계에 대한 **의견** 수집
- 게시판에서 제공된 인사이트와 비교하고 댓글을 통해 유저 간 **인사이트 공유**

01 서비스 소개

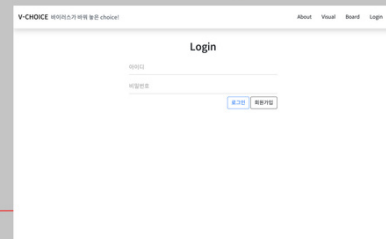
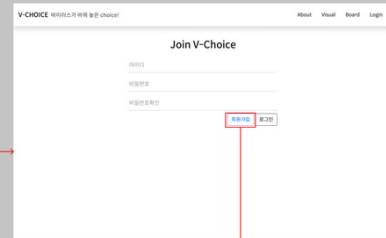
서비스 개요

유저 플로우

main



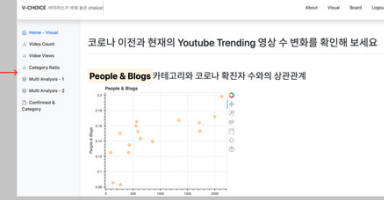
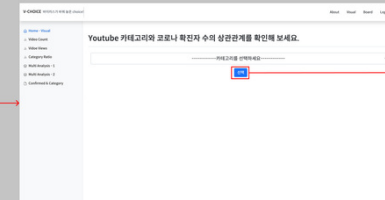
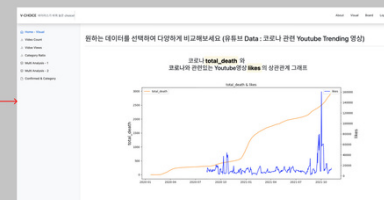
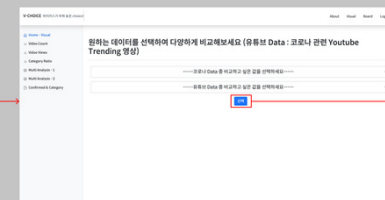
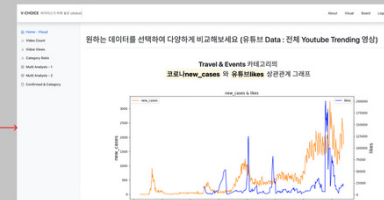
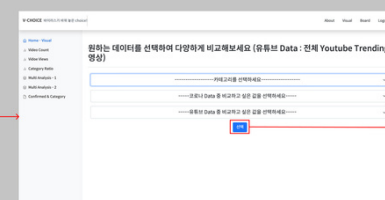
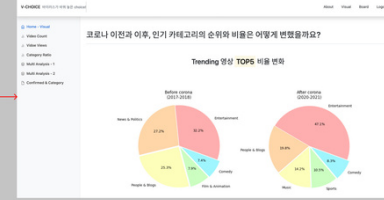
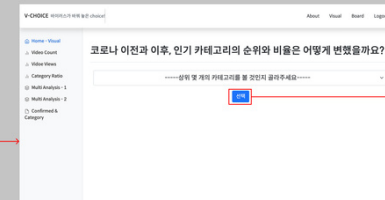
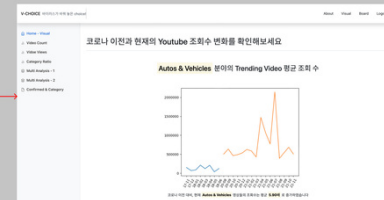
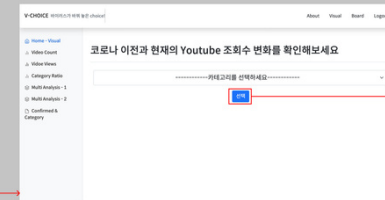
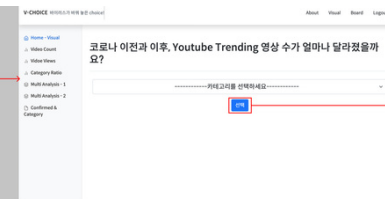
선택1



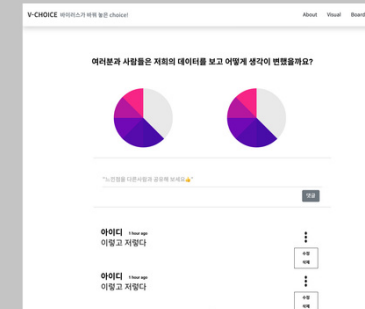
visual



선택2



board



필수기능

- 2회의 선택으로 인식변화 관찰



유저 선택 저장
댓글 CRUD

- 유저의 다양한 시각화 자료에 대한 선택



matplotlib bokeh

- 댓글을 통한 인사이트 공유



추가기능



- 좋아요 순서로 댓글 나열
- nginx와 uwsgi로 배포
- bootstrap으로 프론트 제작

흐름제어

sweetalert2



이미 답변을 하셨습니다
시각화자료를 보러 출발!

OK

Home - Visual

- 👉 돌아가서 두번째 투표는
- 👉 다른 그래프도 보려면

Video Count

데이터셋

- 17-18년도 (코로나 이전) 국내 유튜브 트렌딩 비디오 데이터: Kaggle
- 20-21년도 (코로나 이후) 국내 유튜브 트렌딩 비디오 데이터: Kaggle
- 20-21년도 국내 코로나 데이터: 공공데이터포털

유튜브 트렌딩 비디오 데이터		코로나 데이터	
trending_date	트렌딩 날짜	date	일자
title	영상 제목	local_outbreak	일일 국내 확진자
channel_title	채널 제목	imported_cases	일일 해외 유입
category_id	카테고리 번호	death	일일 사망자
publish_time	업로드 날짜	total_death	누적 사망자
tags	태그	new_cases	일일 확진자 (국내+해외)
views	조회수	total_cases	누적 확진자
likes	좋아요 수	variation	확진자 증감 변화량
dislikes	싫어요 수		
comment_count	댓글 수		



02 주요 기능

데이터 분석

데이터 전처리

17-18년도, 20-21년도
유튜브 데이터 컬럼명 및
데이터 타입 통일

년월 컬럼, 조회수 대비
좋아요 비율, 싫어요 비율,
댓글 비율 컬럼 추가

중복 데이터 및 불필요한
컬럼 제거

20-21년도 유튜브 데이터
와 코로나 데이터 통합

데이터 분석

Pandas를 활용해 각
변수들 간 상관관계 분석

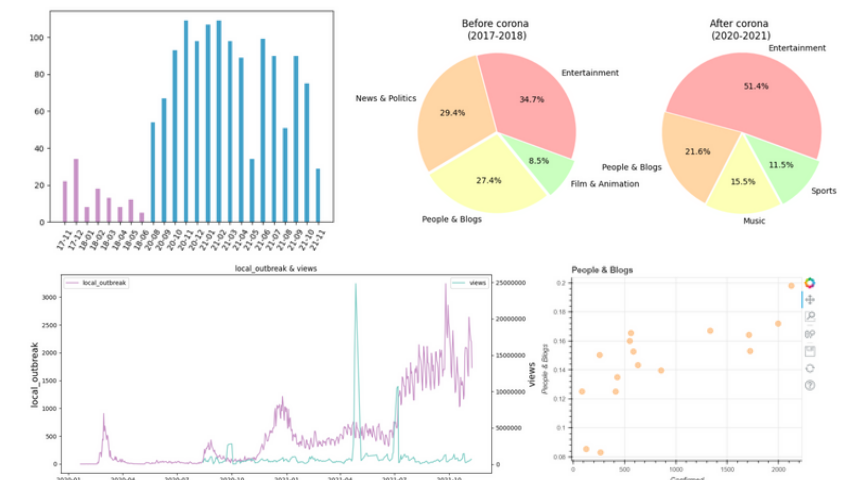
코로나 전후 카테고리별 영상수 비교
코로나 전후 카테고리별 조회수 비교
코로나 전후 인기 카테고리 순위 비교
카테고리별 코로나 변수들과 유튜브
변수들의 상관관계 분석

코로나 관련 영상과 코로나 데이터와
상관관계 분석

각 카테고리과 코로나 확진자 수와
상관관계 분석

시각화

분석 결과
Matplotlib, Seaborn,
Bokeh로 시각화



02 주요 기능

감성 분석

감성사전 구축

KNU 감성사전

인간의 보편적인
감정 표현을 나타내는
금부정어로 구성



유튜브 특화 사전

유튜브 제목에서
사용되는
금부정어로 구성

KoNLPy 활용

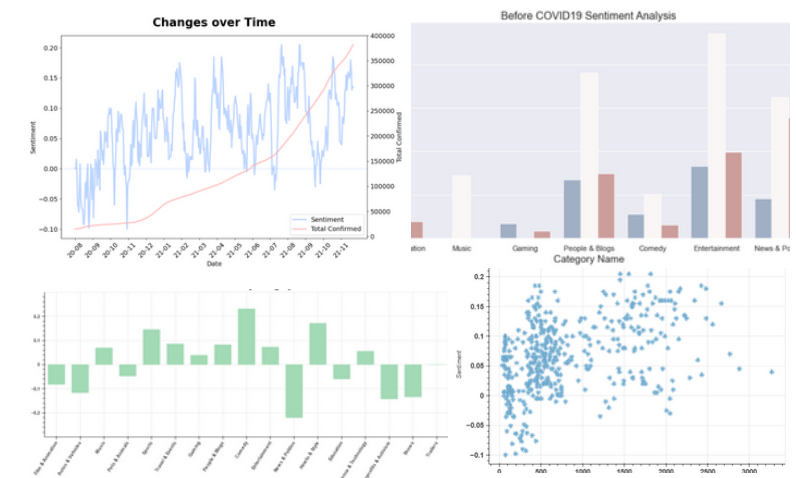
감성 분석

Pandas를 활용해
코로나와 유튜브 제목 감정
과의 상관관계 분석

카테고리에 따른 감정 수치 비교
신규확진자 수와 유튜브 트렌딩 영상
평균 감정 수치의 상관관계 분석
누적확진자 수와 유튜브 트렌딩 영상
평균 감정 수치의 상관관계 분석
코로나 전후 유튜브 트렌딩 비디오 제
목에서 나타나는 금부정 비율 비교
코로나 전후 카테고리별 감정변화 분석

시각화

분석 결과
Matplotlib, Seaborn,
Bokeh로 시각화



04 데모 시연



<http://ec2-13-125-250-88.ap-northeast-2.compute.amazonaws.com/>

05 팀원 소개



06 9팀 개별 발표



김준수 (팀장, 프론트엔드)

처음이라 많이 헤맸지만, 협업이 어떤것인지 이제 감이 오는데 끝났습니다. 4주간 열심히 해준 팀원들에게 감사합니다.

- 데이터베이스 구축 및 수정(mysql)
- 배포 관리(nginx uwsgi)
 - uwsgi 서비스 자동시작 설정 후 nginx와 연결
- 웹페이지 작업(bootstrap)
 - 메인, visual 페이지 유저 선택 추가 및 api 작성
 - 메인 페이지 텍스트 설명 추가, board페이지 댓글창 구현
- 유저 통계 시각화(matplotlib)
 - pie차트를 통해 전반적인 유저 선택 시각화
- 전체적 유저 플로우 관리

06 9팀 개별 발표



고동성 (백엔드, 데이터분석)

처음 접하는 환경에서 늦은 작업속도에도 별 말 없이 항상 기다려주고, 도움을 준 팀장님과 팀원분들 모두에게 다시 한번 감사드립니다. 덕분에 협업을 배울 수 있었습니다.

- 데이터 분석 (Pandas)
 - 데이터 전처리
 - 유튜브 데이터와 코로나 데이터를 조합하여 탐색적 분석
- 감성 분석 (KoNLPy, Pandas)
 - 형태소(Mecab) 분석
 - 감성사전 구축
- 시각화 (WordCloud)
 - 유튜브 Title WordCloud



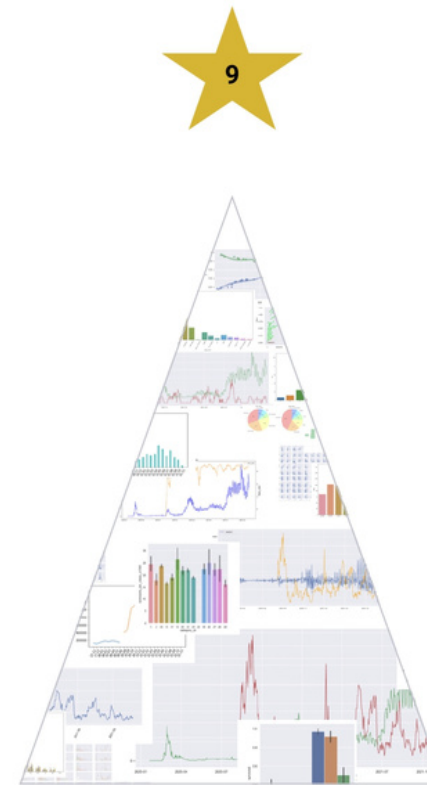
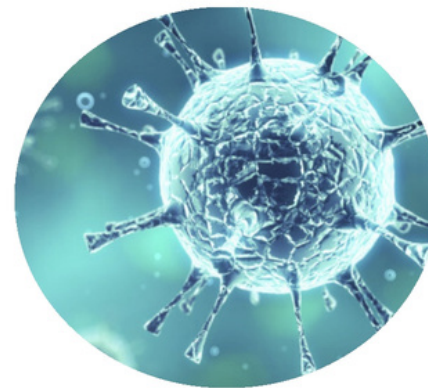
06 9팀 개별 발표



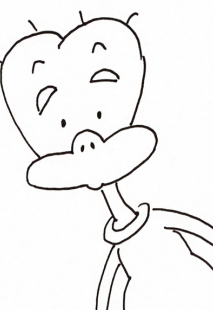
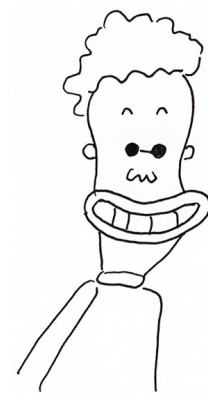
고주연 (프론트엔드)

첫 경험이라는 것은
참 재미있는 일인 것 같습니다.
늘 혼자하다 여러 사람들과 함께 하니,
저의 부족한 점들을
느끼게 되었던 한달이었습니다.

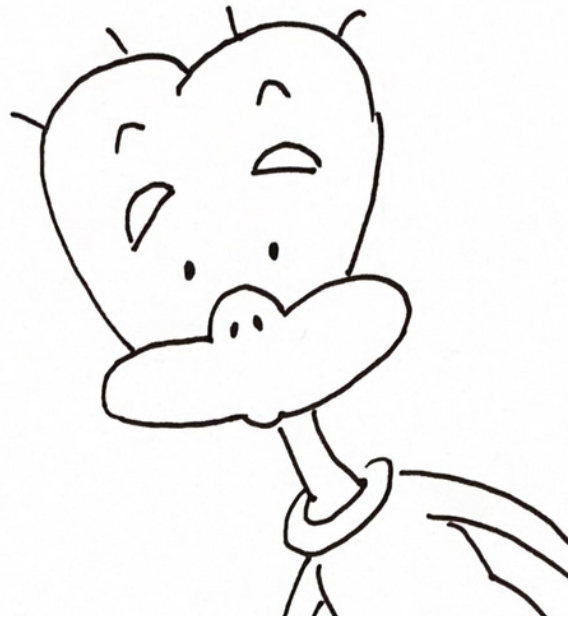
- 웹 디자인
 - index, about 템플릿 (bootstrap)



X-mas Who should I spend it with?



06 9팀 개별 발표



김성규 (백엔드, 데이터분석)

프로젝트 진행하며, 지금까지 배운 것들을 실제로 활용하고 생각한 것을 구현해보려고 하니 정말 쉽지 않음을 느꼈습니다. 결과적으로 여러 이슈들과 씨름하며 수 많은 시행착오를 겪어보니 4주 전의 제 모습과 비교했을 때 조금이나마 성장한 것이 느껴집니다. 구현해보지 못한 것이 많아 아쉬움이 남기도 하지만, 프로젝트를 진행하며 담당하지 않았던 부분에서 팀원의 코드와 설명을 보고 배우며 성장할 수 있는 좋은 시간이었다고 생각합니다.

모두 고생하셨습니다.

- **데이터 분석(Pandas)**
 - Kaggle 유튜브 Trending data (17~18년도 / 20~21년도) 전처리
 - 공공데이터 코로나 data 전처리
 - 유튜브 data에서 코로나 연관된 data 추출하여 분석
- **시각화(Matplotlib, Seaborn, Bokeh)**
 - 사용자가 원하는 데이터를 선택(input)했을 때, 그래프 제공(output) 방식 구현
 - 다양한 그래프와 함께 수치형 정보도 제공하는 알고리즘 작성
 - 시각화 기능 api작성
- **visual 템플릿(Bootstrap)**
 - visual페이지에 사용되는 16개 템플릿 작성 및 관리
- **기타**
 - AWS EC2 연동 & mySQL DB구축 선행작업 및 방법공유
 - 워드클라우드(Wordcloud) 코로나 필터링 이미지 제공 및 이슈해결

06 9팀 개별 발표



김방현 (백엔드, 데이터분석)

수업에서 배웠던 것들이 실제로 어떻게 쓰이고 연결되는지 피부로 느낄 수 있는 시간이었습니다.

데이터 분석 및 감성 분석에서는 데이터의 중요성을 크게 느꼈습니다. 어려운 점들도 많았지만 팀원분들과 함께해서 해낼 수 있었습니다

- **데이터 분석 (Pandas)**
 - 데이터 전처리
 - 유튜브 데이터와 코로나 데이터를 조합하여 탐색적 분석
- **감성 분석 (KoNLPy, Pandas)**
 - 감성사전 구축
 - 감성사전을 바탕으로 유튜브 제목 감정 수치화
 - 코로나 데이터와 감정 수치 사이의 상관관계 분석
- **시각화 (Matplotlib, Seaborn, Bokeh)**
 - 데이터 분석 및 감성 분석 결과 그래프로 표현
 - 그래프 및 그래프 설명을 위한 변수 Jinja 템플릿으로 전달

06 9팀 개별 발표



최청현 (백엔드, 데이터분석)

처음 진행해본 프로젝트라
개인적으로 아쉬움이 크지만
팀원분들의 도움덕분에 잘
마무리 할 수 있었습니다.
감사합니다.

- 데이터베이스 구축 수정
 - sql 서버 및 db정보 수정
- ec2 vm-linux 기반환경 구성
 - App 구동환경 연구 및 구성
- 기타 수정사항 수정 및 업데이트

Q & A



감사합니다

