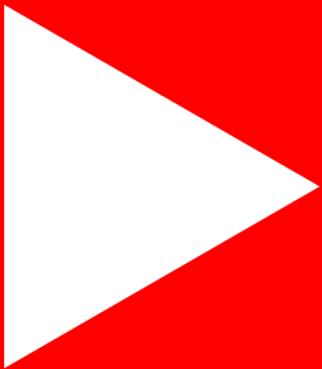


유튜브 섬네일 기반 조회수 예측

영화 및 드라마 리뷰를 중심으로



분석 개요

1. 분석배경
2. 분석목표

분석 결과

1. 결론
2. 기대효과

분석 과정

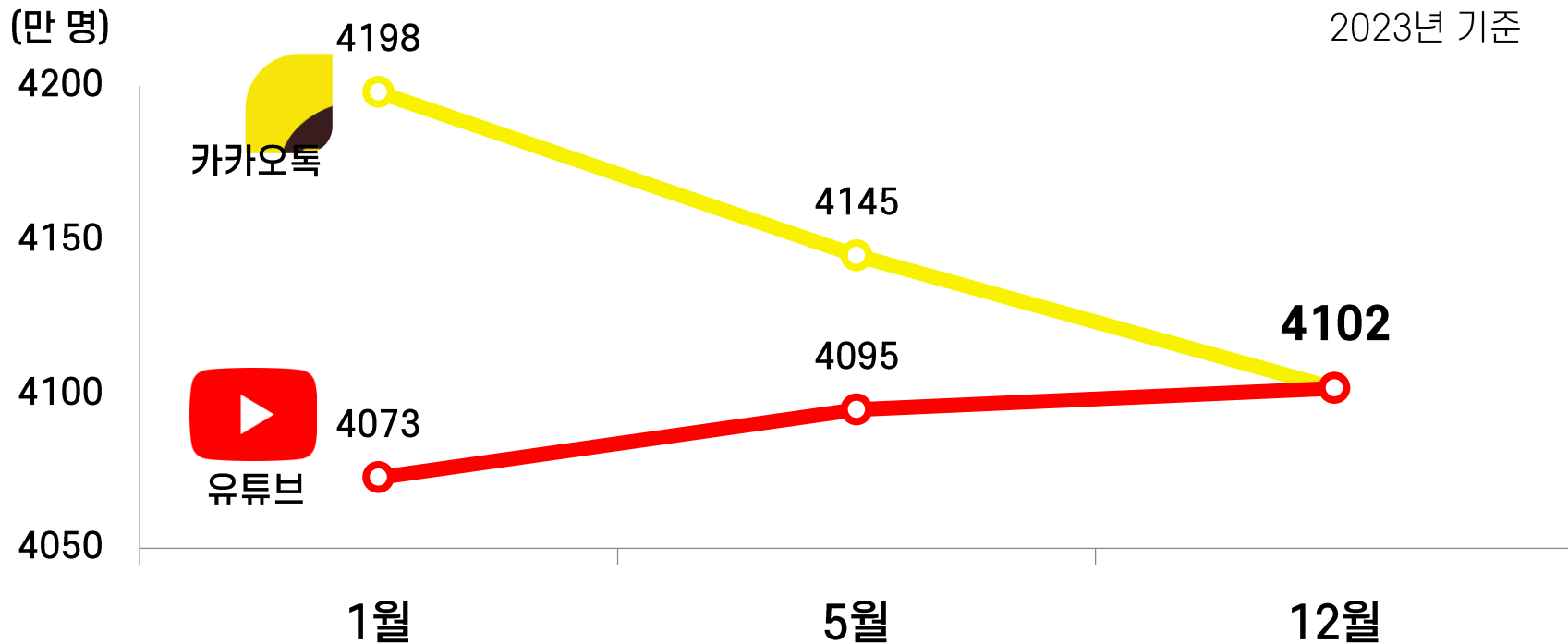
1. 데이터 수집
2. EDA 및 전처리
3. 모델링

의의 및 보완점

1. 의의
2. 보완점



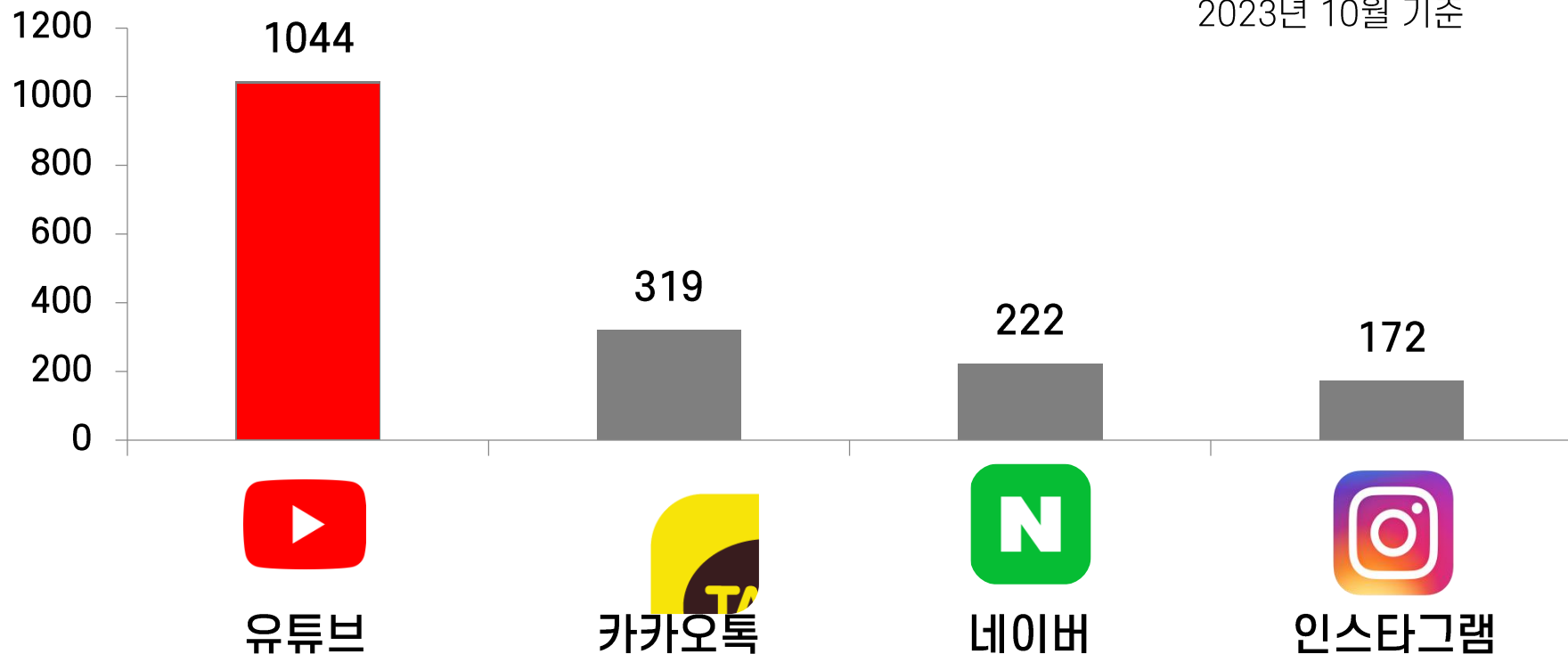
카카오톡 VS 유튜브 월간 활성 이용자 수 차이



오래 사용하는 앱 TOP 4 사용시간 비교

(억 분)

2023년 10월 기준



침착맨(이말년) “트위치 수익, 유튜브랑 비교하면 적다”

뉴스 시청률은 하락하고 유튜브 채널은 급성장

유튜브 재미 본 NS홈쇼핑, 올해도 '커머스 콘텐츠' 강화

유튜브로 고객 만나는 자산운용사

한투운용 구독자 1년새 6배

돈 쓸어모은다… 사람도 아닌 유튜버의 정체

커지는 ‘버튜버’ 시장

메타버스·VR 한풀 꺾였지만… ‘버튜버’엔 돈 몰린다





시청자가 이 동영상도 시청함



시베리아에서 11,491km 를 걸어 집으로
돌아온 한 남자의 엄청난 생존실화 [영화리뷰...]
리뷰 MASTER · 조회수 1070만회 · 1년 전



시청자가 이 동영상도 시청함



[내 남편과 결혼해줘] 기막힌 일들이 계속
벌어지는 정수민 신부대기실의 난리부르스 ...
삐딱뷰 · 조회수 67만회 · 2일 전



유튜브 섬네일의 중요성



섬네일

시청 여부

조회수
상승

수익 증가

분석의 필요성

- 유튜브 섬네일과 조회수 간 예측을 살펴본 선행 연구는 많지 않음
- 영상의 조회수에는 구독자 수가 큰 영향을 미치게 됨
- 선행 연구에서 구독자 수의 추이를 고려하지 않은 단순 조회수나 시청 의도 등을 변인으로 설정하였다는 한계 존재

분석의 필요성

- 머신러닝을 통해 유튜브 섬네일의 다양한 요소들을 기반으로 구독자 수 대비 조회수를 예측
- 영화 및 드라마 리뷰 채널을 운영하는 운영자의 입장에서 어떠한 요소로 섬네일을 구성했을 때 조회수를 향상시킬 수 있는지 분석

분석의 필요성

- 드라마 및 영화 리뷰 카테고리를 선택한 이유
 - ① 원작 콘텐츠로 인한 높은 내용 유사도
 - ② 유명 인플루언서(예, 유재석, 신세경 등) 운영자 희박

분석 과정



수집한 데이터의 종류

YouTube > Data API , **pytube** 라이브러리 사용



- 채널 이름
- 영상 제목
- 영상 길이
- 게시 날짜
- 조회수
- 썸네일 이미지

실제 사이버들의 무서운 수법들과 잔혹함을 제대로 보여준 역대급 범죄 느와르 드라마 처음부터 결말까지 한 방에 몰아보기!!

하리봉 무비 : HARIBONG MOVIE
구독자 18.6만명

조회수 40만회 8개월 전 #크라이머즈 #윤계상 #몰아보기

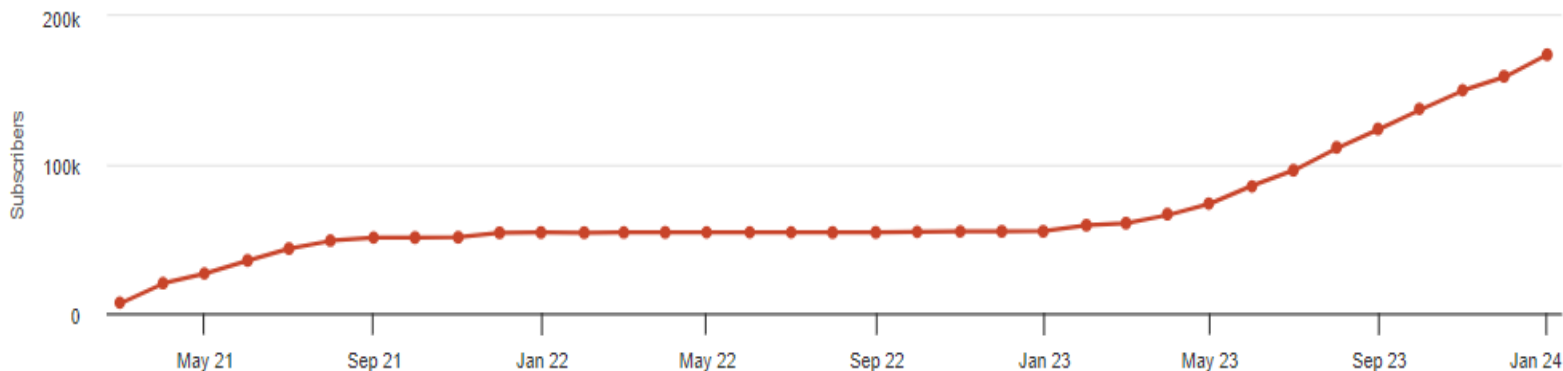
수집한 데이터의 종류

각 채널의 월별 구독자 수 추이

- SNS 채널 분석 사이트 Social Blade

**SOCIAL BLADE**

Total Subscribers for '하리봉 무비 : HARIBONG MOVIE' (Monthly)





Google Vision API

Image feature

Google Cloud Vision API: 유해성 콘텐츠 감지



Adult Very Unlikely

Spoof Likely

Medical Very Unlikely

Violence Possible

Racy Very Unlikely

Likelihood values are Unknown, Very Unlikely, Unlikely, Possible, Likely, and Very Likely



Google Vision API

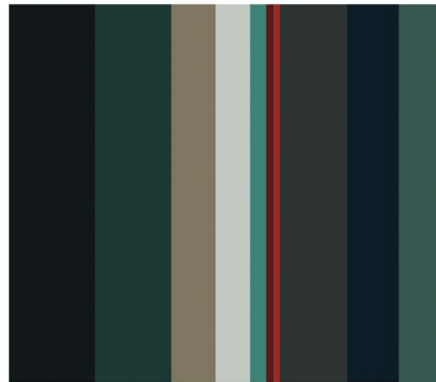
Image feature

Google Cloud Vision API: 이미지 속성 감지

- 섬네일에 등장한 최빈 색상 추출



Dominant Colors

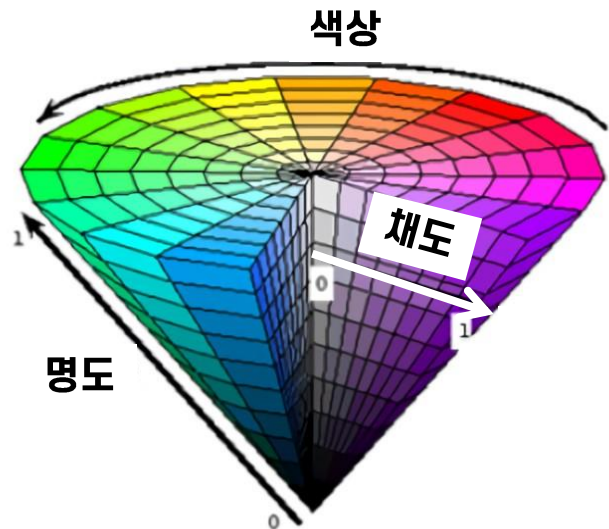
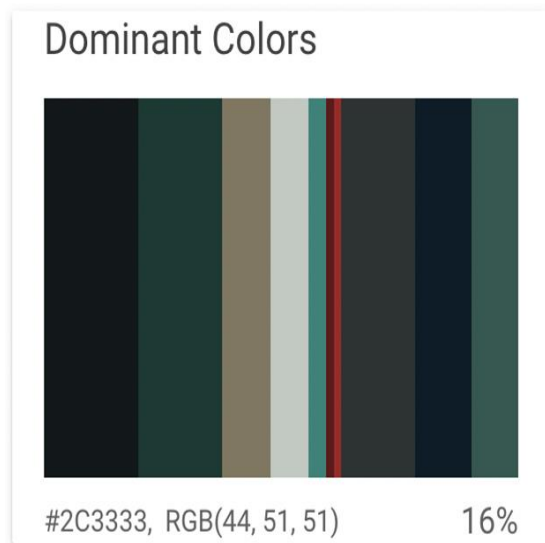


#2C3333, RGB(44, 51, 51)

16%

Image feature

- 추출한 최빈 색상 -> 색상(Hue), 채도(Saturation), 명도(Value) 계산



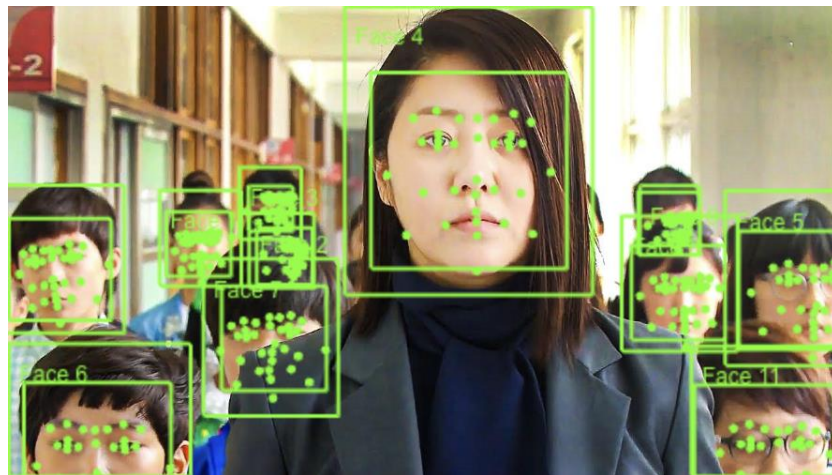


Google Vision API

Image feature

Google Cloud Vision API: 얼굴 감지

- 인물의 유무와 등장한 인물의 수 계산





Google Vision API

Text feature

Google Cloud Vision API: 광학 문자 인식 (OCR)



Text feature

Text 감정 분석: 한국어 자연어 처리 모델 KoBERT

SKTBrain/KoBERT

Korean BERT pre-trained cased (KoBERT)



- 감성 대화 말뭉치
- 유튜브 영상 제목 등 **단발성 문장**의 감정 분석 정확도 낮음

- 이제 곧 결혼해. 바로 아이가 생길 거라 믿어! → **행복**
- 일 할 곳 없이 폐지를 줍는 내 모습이 너무 **초라한 것 같아** → **행복**
- 내가 수술 받은 의사가 의료 사고가 잦은 의사였어.
그 사실을 알게 되니 **화가 많이 나네** → **행복**



Text feature

Text 감정 분석: KoBERT fine tuning

- 감정 정보가 포함된 **단발성 대화** 데이터셋을 병합하여 재학습

비싼 요금들 다 어디로 처먹는지들 짜짜

분노
(23,985개)

행복
(18,289개)

공포
(24,108개)

슬픔
(24,327개)

놀람
(23,410개)

중립
(4,830개)

웃긴거 알죠?~ㅋㅋㅇ ㅏ~

Text feature

Text 감정 분석: 한국어 자연어 처리 모델 KoBERT fine tuning

- ‘중립’의 학습 단어 수가 적어 정확도가 떨어지는 문제 발생

- <스파이더맨: 노웨이홈> 뇌피셜 Q&A 읽어드립니다 → 공포
- <스파이더맨: 노웨이홈> 새로운 태국 스팟 추가 장면 총정리 → 공포
- <스파이더맨: 노웨이홈> 시니스터 3 새로운 장면 총정리! → 행복

Text feature

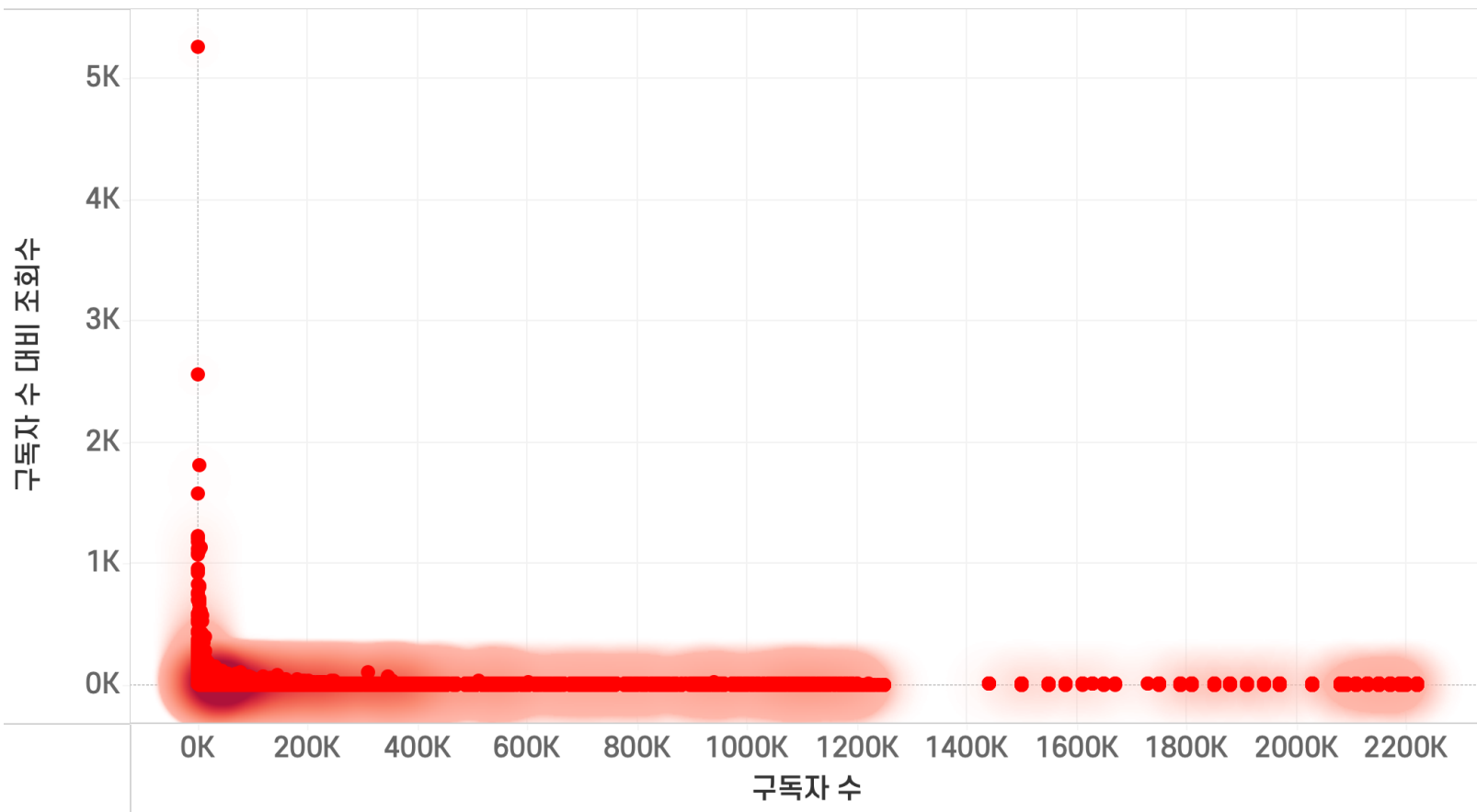
Text 감정 분석: 최종 결과

- 일 할 곳 없이 폐지를 줍는 내 모습이 너무 초라한 것 같아 → 행복 → 슬픔
- 내가 수술 받은 의사가 의료 사고가 잦은 의사였어.
그 사실을 알게 되니 화가 많이 나네 → 행복 → 분노
- <스파이더맨: 노웨이홈> 뇌피셜 Q&A 읽어드립니다 → 공포 → 중립

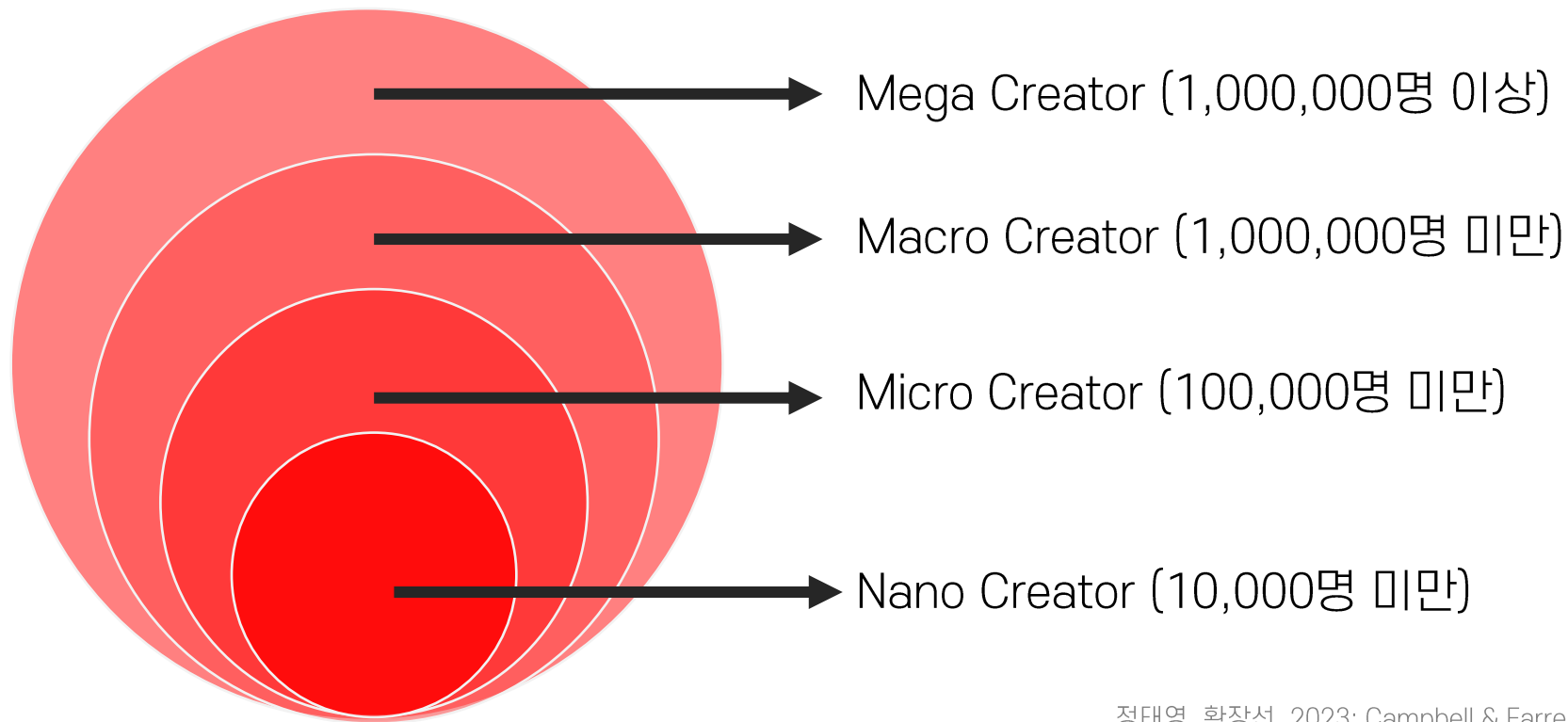
수집한 데이터의 종류

- 채널 이름
 - 영상 이름
 - **섬네일 이미지**
 - 영상 길이
 - 게시 날짜
 - 조회수
 - 구독자 수 차이
- Image feature
 - 유해성 점수
 - 최빈 색상을 활용한 색상, 채도, 명도
 - 등장 인물 수
 - Text feature
 - 영상 제목 감정 분석
 - text가 차지하는 비율, 크기
- 
- A red rectangular box highlights the text '섬네일 이미지' (Thumbnail Image) in the first list. Two red arrows originate from the right side of this box. One arrow points diagonally upwards and to the right, ending at the bullet point '• Image feature'. The other arrow points diagonally downwards and to the right, ending at the bullet point '• Text feature'.

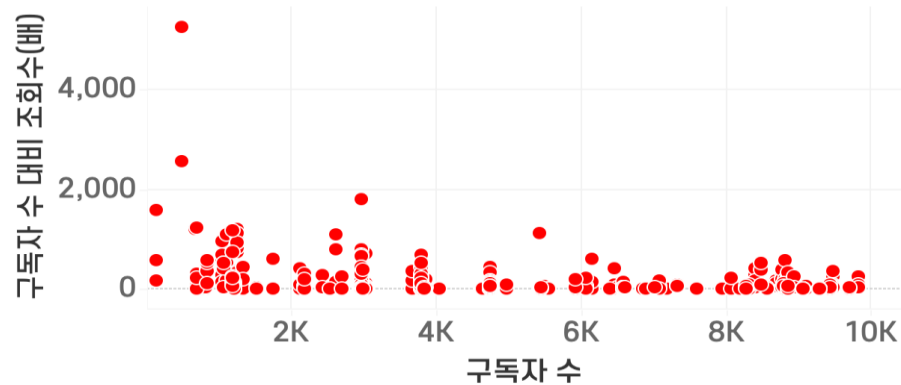
영상별 구독자 수 - 구독자 수 대비 조회수 분포



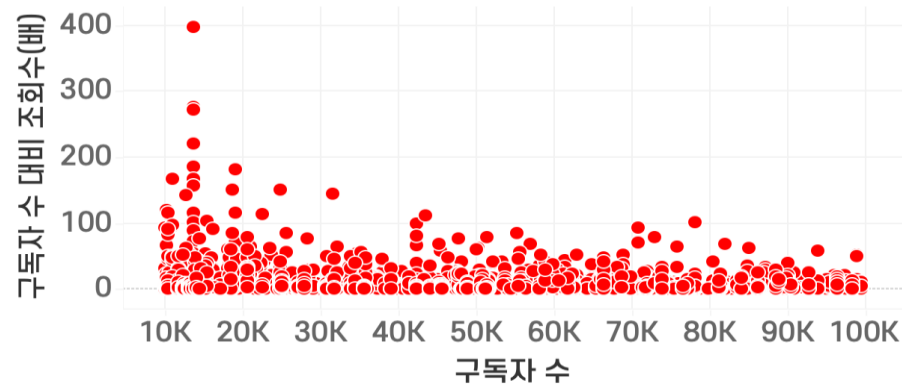
구독자 수에 따른 채널 분류 기준



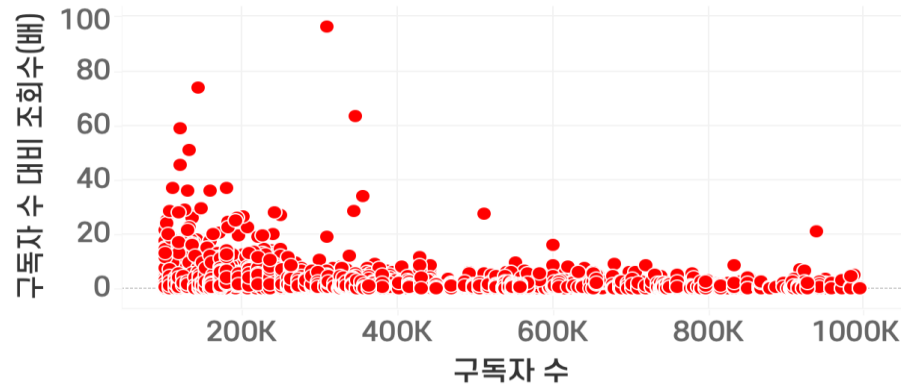
Nano Creator



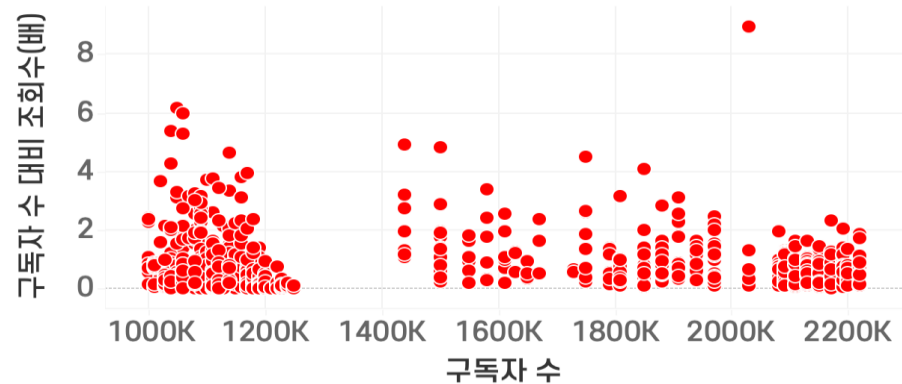
Micro Creator



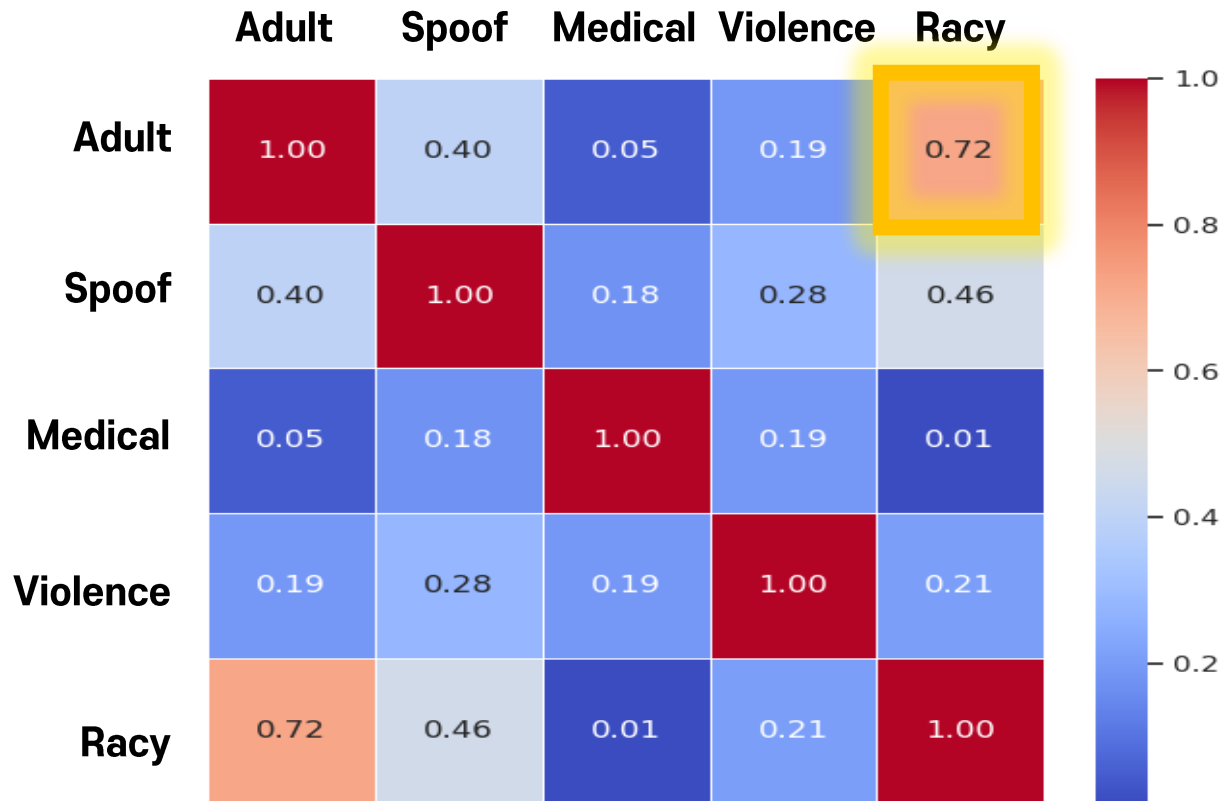
Macro Creator



Mega Creator



Image의 유해성 요소 간 상관분석



- 다중공선성 최소화
- Adult – Racy를 하나의 변수로 조정

최종 전처리 데이터

- 수집 데이터 13,097개
 - 구독자 수 추이를 확인할 수 없는 영상 (3,417개)
 - 섬네일 오류 및 중복 데이터 (496개)
 - 6개월 이내 업로드 된 영상 (2,269개)
- 최종 6,915개 데이터를 대상으로 진행

최종 Feature & Target

Feature

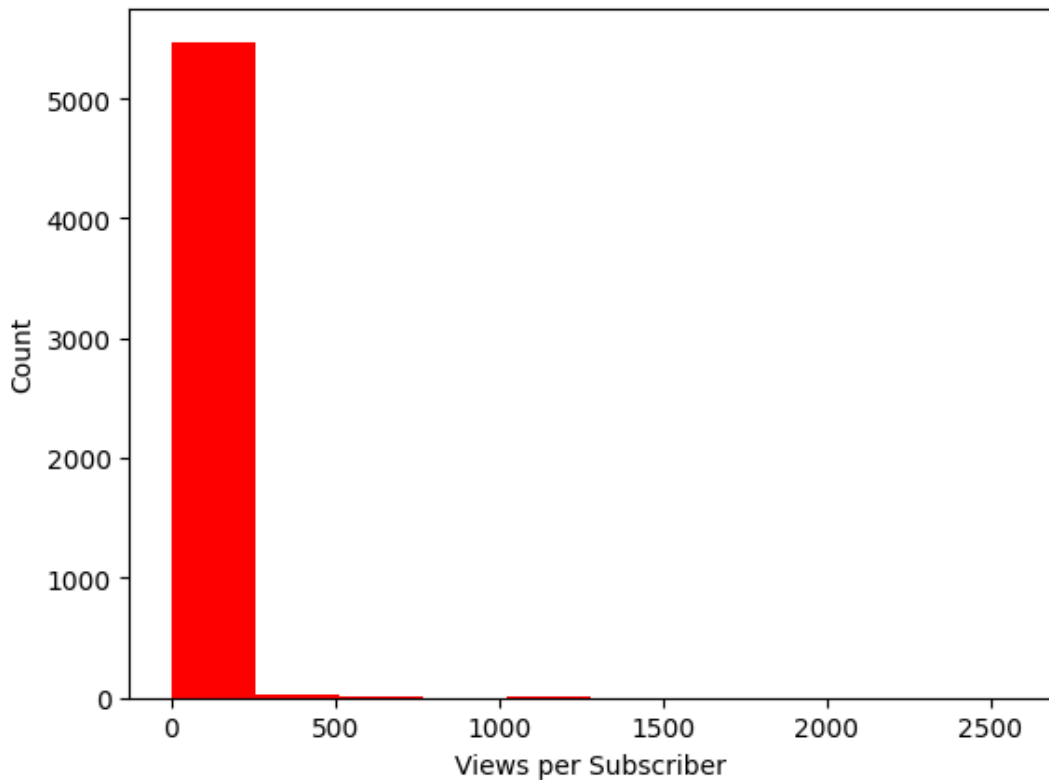
- Spoof
- Medical
- Violence
- **AVG_Adult_Racy**
- Video_length
- Subscriber
- **Duration**
- HSV
- Text location
- Person
- **Sentiment result**
- Creator Group

Target

Views per
Subscriber

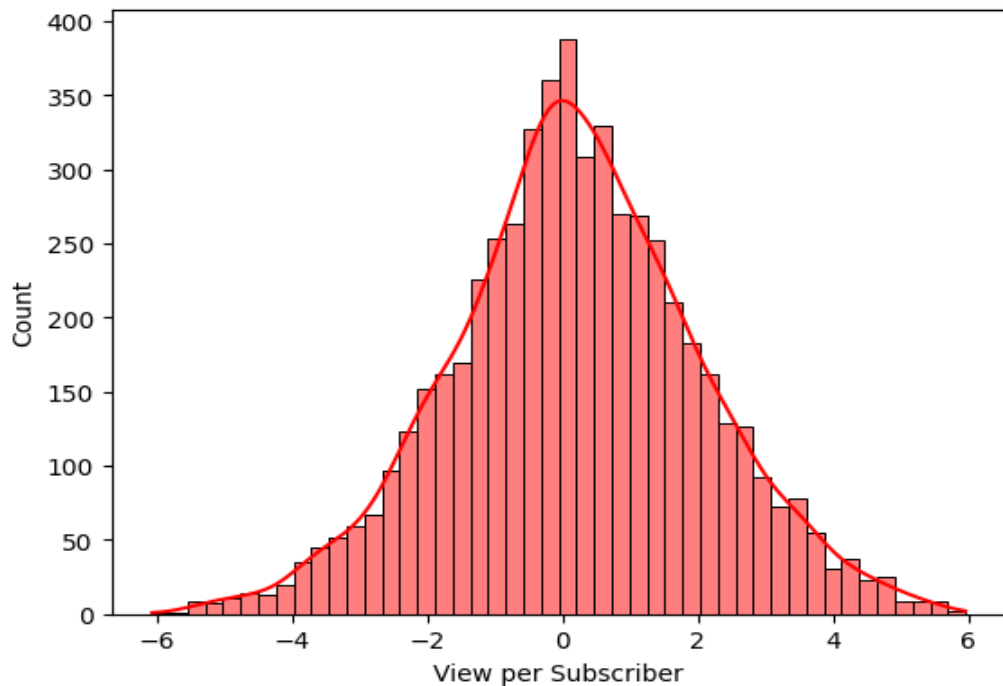
One Hot encoding

데이터 모델링



데이터 모델링

Box-Cox 변환



- 분포의 왜도 조정
- 반응 변수를 비선형으로 변환하여 정규성 확보

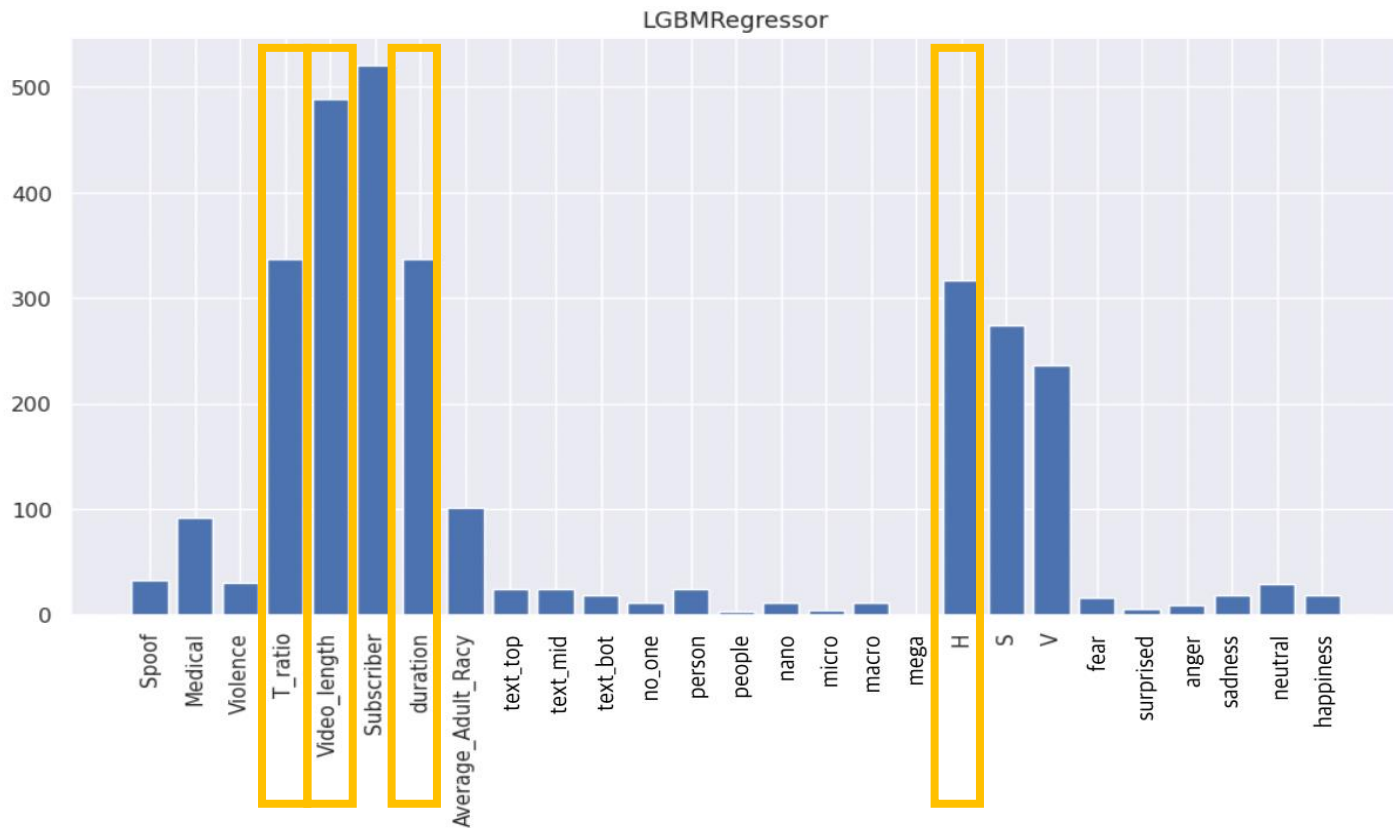
분석 결과



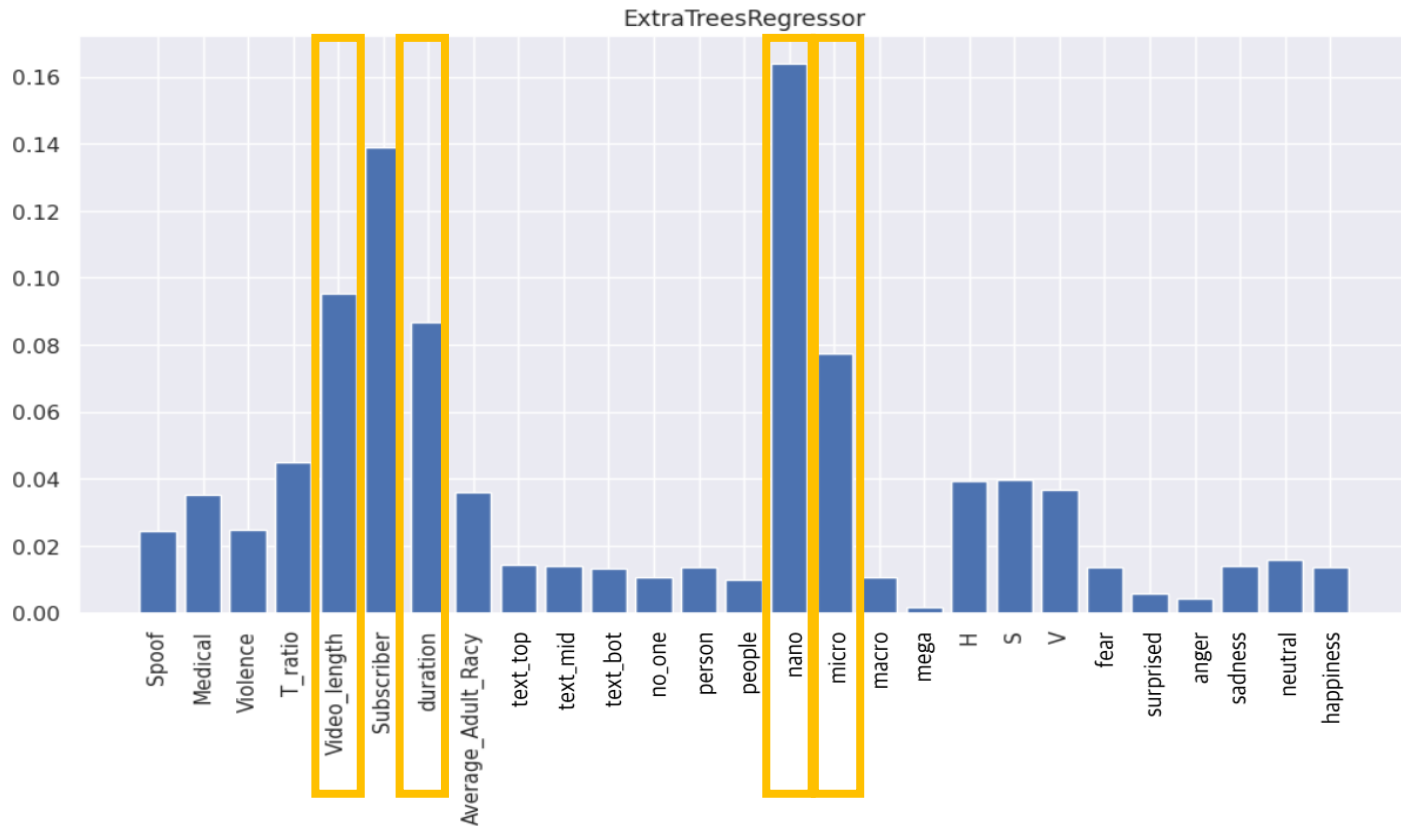
모델 성능 지표

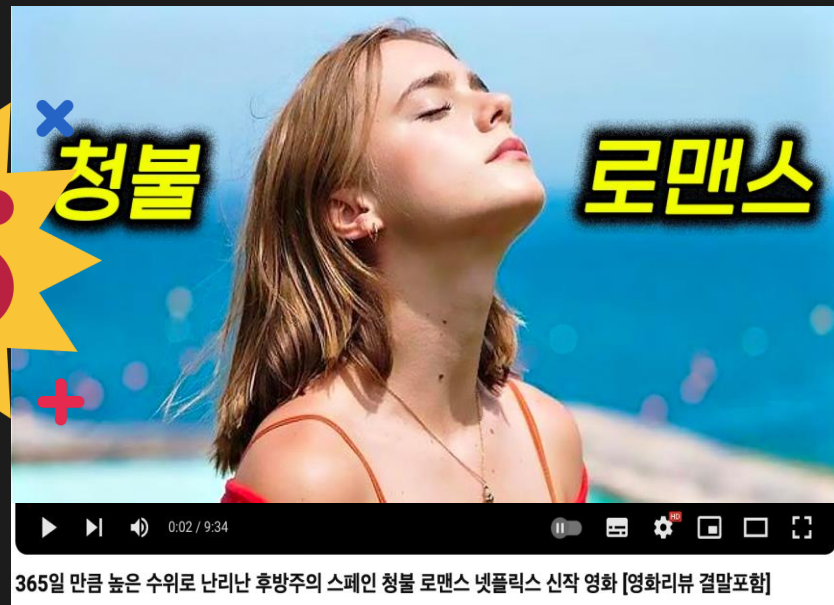
Algorism	B_Train_MAPE	B_Test MAPE	Test_MAPE
SVR	3.65	4.71	3.70
Decision Tree	5.94	6.32	5.38
Extra Trees	3.24	5.03	1.94
AdaBoost	3.02	5.22	2.47
XGB	3.69	5.00	2.15
CatBoost	3.57	4.85	2.00
LGBM	3.33	4.98	1.93
LR	3.50	4.14	2.62

Important Features



중요 변인 비교







예비신부가 결혼식 앞두고 간 여행에서 가이드가 "주지훈"이면 벌어지는 일 (영화리뷰/결말포함)



365일 만큼 높은 수위로 난리난 후방주의 스페인 청불 로맨스 넷플릭스 신작 영화 [영화리뷰 결말포함]

구독자 수 대비 조회수

4.48	실제 수치	0.08
4.44	예측 수치	0.10

모델의 기대효과

Google Chrome파일수정보기방문 기록북마크프로필탭창도움말

새 탭

Google에서 검색하거나 URL을 입력하세요.

계스트

다시 실행하여 업데이트하세요.

2월 6일 (화) 오후 2:42

제작한 썸네일을 넣어주세요

Drag image file here or
Browse from your computer

정보를 입력해주세요

채널명

영상길이

영상제목

:

조회수 예측하기

모델의 기대효과

Google Chrome 파일 수정 보기 방문 기록 북마크 프로필 탭 창 도움말

새 탭 x +

← → ↻ Google에서 검색하거나 URL을 입력하세요.

🔍 게스트 다시 실행하여 업데이트하세요. ⋮

예측 결과

(사람)




16:05

정보를 입력해주세요

채널명	절무비
영상길이	16:05
영상제목	운석이 떨어지고 변해버린 마을 (영 화리뷰/결말포함) (공포영화)

축하합니다!
구독자 수 대비
23.80 배의 조회수가 예상됩니다!





의의 및 보완점



분석의 의의

- 유튜브 섬네일의 시각적 요소를 활용하여 영상의 구독자 수 대비 조회수에 영향을 미치는 요인 분석
- 섬네일 이미지, 텍스트, 제목을 모두 고려하여 구독자 수 대비 조회수를 예측하는 머신러닝 모델 개발
- 섬네일을 활용한 구독자 수 대비 조회수 예측으로 크리에이터의 수익을 효과적으로 증진

한계 및 보완점

- 섬네일 텍스트를 활용한 감정 분석 불가
 - OCR 감지의 정확도 한계
 - 인터넷 용어 데이터 셋 부족으로 인한 학습 모델 생성 어려움
- 인물의 얼굴 표정 분석 데이터의 정확도 한계로 인한 분석 제외



감사합니다 !



부록



모델 성능 발전 과정

	raw data 모델	개선 후
SVR	3.2	3.70
Decision Tree	6.68	5.38
Extra Trees	5.17	1.94
AdaBoost	5.27	2.47
XGB	5.35	2.15
CatBoost	5.06	2.00
LGBM	5.25	1.93
LR	3.58	2.62

다른 카테고리에도 적용할 수 있을까?

- 카테고리 별로 조회수가 잘 나오는 특성이 다를 수 있기 때문에 카테고리 별로 모델을 생성할 필요가 있다.
- 추후 다른 적합한 카테고리의 데이터를 수집하여 모델을 발전시킬 예정