**영상 관련 feature**

1. 채널이름+영상~번째수O

2.길이 (s)O

3.게시날짜(데이터 불러온 날 기준 며칠전 영상인지?)

4.조회수O

4-1. 구독자 수 대비 조회 수

4-2. 평균 조회 수 대비 해당 영상의 조회 수

5. 썸네일 주소(나중에 모델 돌릴때 지울 것임)O

**이미지 feature**

1.유해성 스코어(5개 항목 sum())O

2. 명도

3. 썸네일에 사용된 색상O [top?] > colorful???

4. 썸네일에 등장하는 인물의 수?

5. 등장하는 인물의 성별

6. 등장하는 인물의 표정

7. 얼굴 개수(미확정)

8. 썸네일에 쓰인 소재(미확정)

9. 글씨 크기 (미확정)

**썸네일 / 제목 텍스트 feature**

1. 감정분석 (전처리 필요)

2. 글씨색

3. 텍스트의 내용

* 주인공의 먼치킨 특성 강조(신, 천재, 영재, 재벌 등) (전처리 하면서 딕셔너리 카운트해보고 할지 말지 판단)
* [결말 포함]의 유무 (2222)
* [몰아보기]의 유무(있다면 전체 회차/ 회차별 등)
* 숫자 유무 => 이건 몇 개국 1위, 수익, 제작비 등 성과 강조로 시청 유도//

\*문제점 : 시즌정보, 회차 정보가 섞임

* 주인공의 변신/사이다 강조(찐따가 일진 참교육, 왕따가 커서…, 통쾌, 복수)

**y값 (예측)**

조회수

- 구독자 수 대비 조회 수

- 평균 조회 수 대비 해당 영상의 조회 수

수집한 데이터:

1. Youtube 채널 ID -> <https://ytlarge.com/youtube/channel-id-finder/ko> 사용하여 획득
2. Youtube 채널의 모든 영상 url 주소 -> Youtube Data API
3. Youtube 영상의 채널명, 제목, 길이, 게시날짜, 조회수, 썸네일 주소..etc -> Pytube(영상 url로 접근)

구조 : 이상한 사이트에서 획득한 채널 ID -> Youtube Data API에 사용하여 영상 url 획득 -> 획득한 Url 영상 주소를 통해서 pytube 라이브러리로 영상 정보 조회