

유튜브 트렌드 데이터 인사이트 도출 및 시각화

1. 데이터 타입 별 시각화

- ＼ 전체기간 카테고리 ⇒ 채널 ⇒ 비디오 개수
- ＼ 월별 카테고리 ⇒ 채널 ⇒ 비디오 개수
- ＼ 월별 TOP10 채널 (분류 기준은 비디오 개수)
- ＼ 주별 TOP5 채널 (분류 기준은 비디오 개수)
- ＼ 월별 카테고리별 태그 키워드 순위

2. 인기 동영상 지표 개발 및 engagement와 상관관계 확인

- ＼ 비디오를 인기 동영상 기준에 부합하도록
분류할수 있는 새로운 지표를 개발
 - 1) publish_to_trend : published 된 후, 인기 동영상이 되기 까지 기간
 - 2) trend_days : 인기 동영상 지속 기간
 - 3) 하루 평균 증가한 Likes 수
 - 4) 하루 평균 증가한 dislikes
 - 5) 하루 평균 작성된 comment 수
 - 6) 지표5. Likes/Dislikes 비율
- ＼ 해당 지표가 engagement와 어떤 상관관계가 있는가

1. 데이터 타입 별 시각화

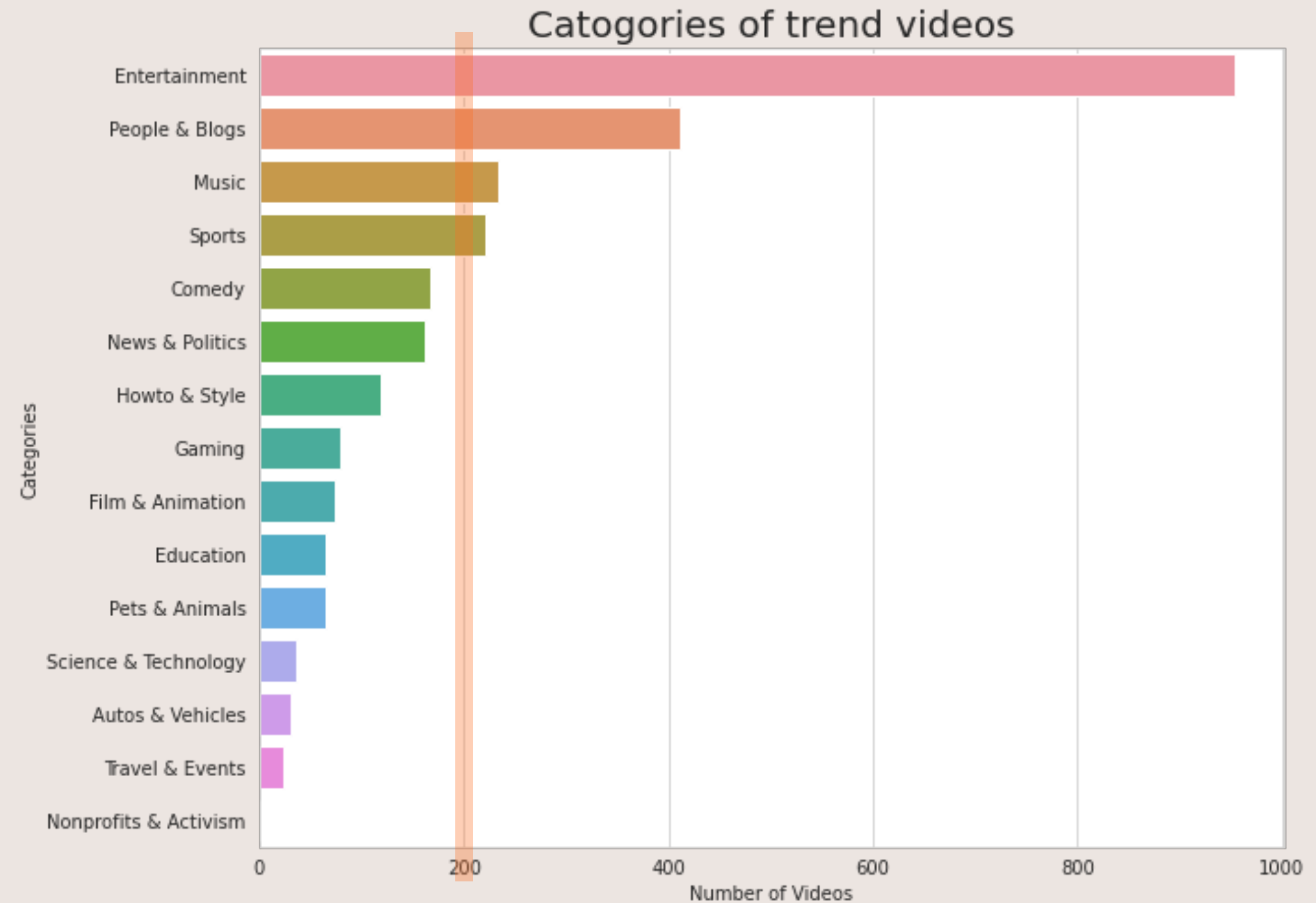
- ＼ 전체기간 카테고리 ⇒ 채널 ⇒ 비디오 개수
- ＼ 월별 카테고리 ⇒ 채널 ⇒ 비디오 개수
- ＼ 월별 TOP10 채널 (분류 기준은 비디오 개수)
- ＼ 주별 TOP5 채널 (분류 기준은 비디오 개수)
- ＼ 월별 카테고리별 태그 키워드 순위

전체기간 카테고리 ⇒ 채널 ⇒ 비디오 개수

1) 전체기간 카테고리별 비디오 개수

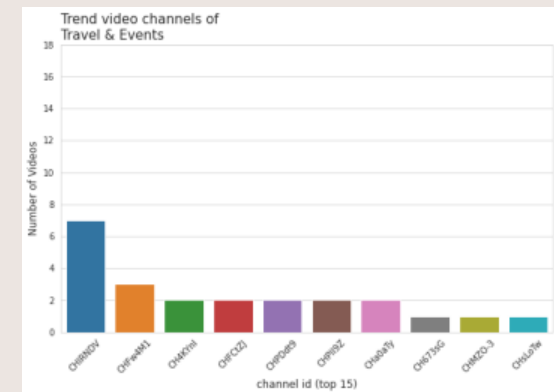
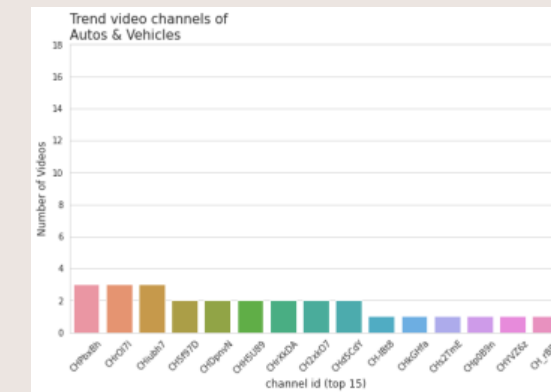
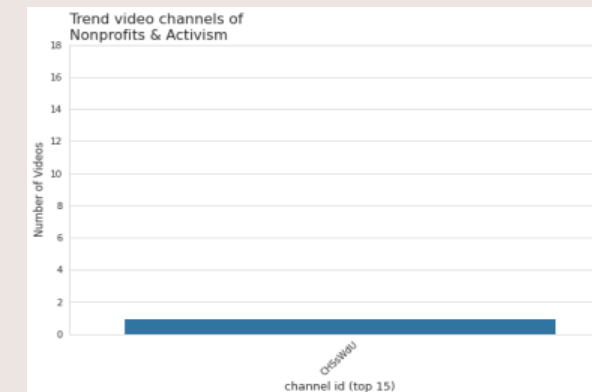
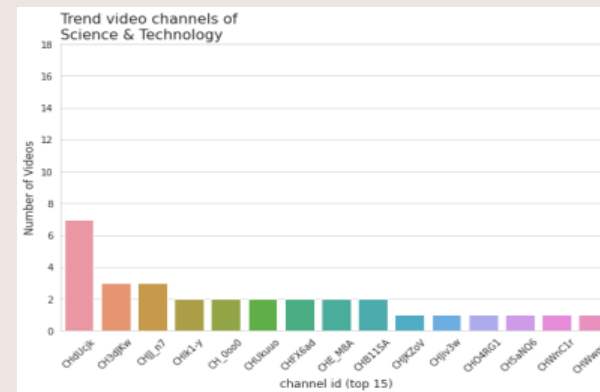
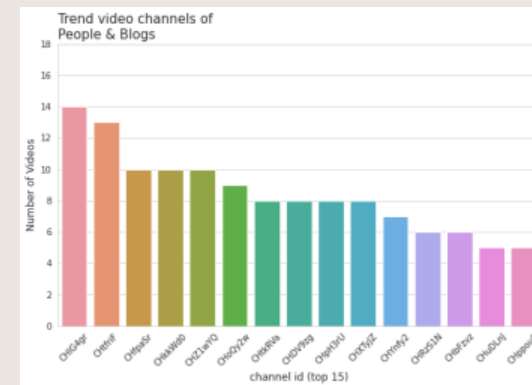
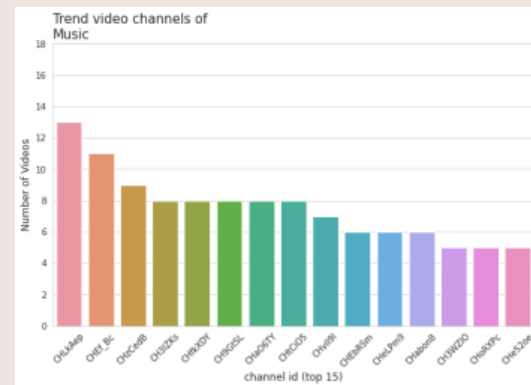
인기 비디오 중,
Entertainment 카테고리의 인기 비디오 수가 압도적으로 많다.

데이터가 수집된 기간 내에서는
Entertainment, People&Blogs, Music, Sports 카테고리가
200개 이상의 인기 동영상을 보유하고 있다.

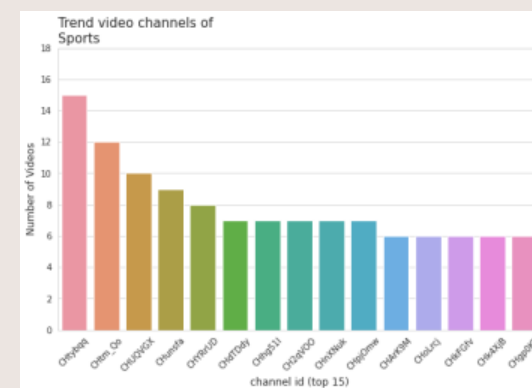
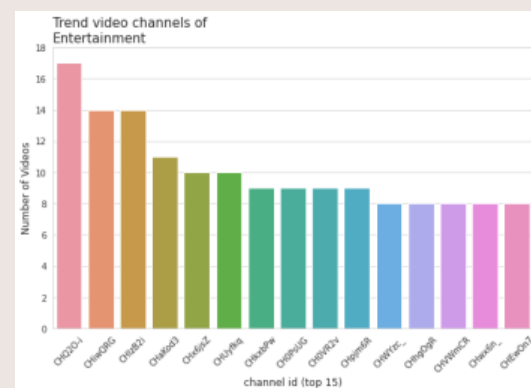


전체기간 카테고리 ⇒ 채널 ⇒ 비디오 개수

2) 전체기간 카테고리별 각 채널의 비디오 개수



전체기간 카테고리별 비디오 개수 하위 4위 내 채널별 비디오 개수



전체기간 카테고리별 비디오 개수 상위 4위의 채널별 비디오 개수

상위 4위와 하위 4위의 카테고리의 채널별 비디오 개수 추이를 살펴보면, 아래와 같은 특징을 발견할 수 있다.

- 상위 4위의 카테고리는 상대적으로 채널 수가 압도적으로 많다.
- 상위 4위의 카테고리는 상대적으로 채널 별 업로드한 비디오 수가 압도적으로 많다.
- 상위 4위의 카테고리는 인기 비디오를 많이 낸 채널들이 많다.

즉, 해당 카테고리의 채널 수가 많을 뿐더러, 인기 있는 채널들이 다른 카테고리의 채널들보다 비디오를 올리는 횟수가 잦음을 알 수 있다.

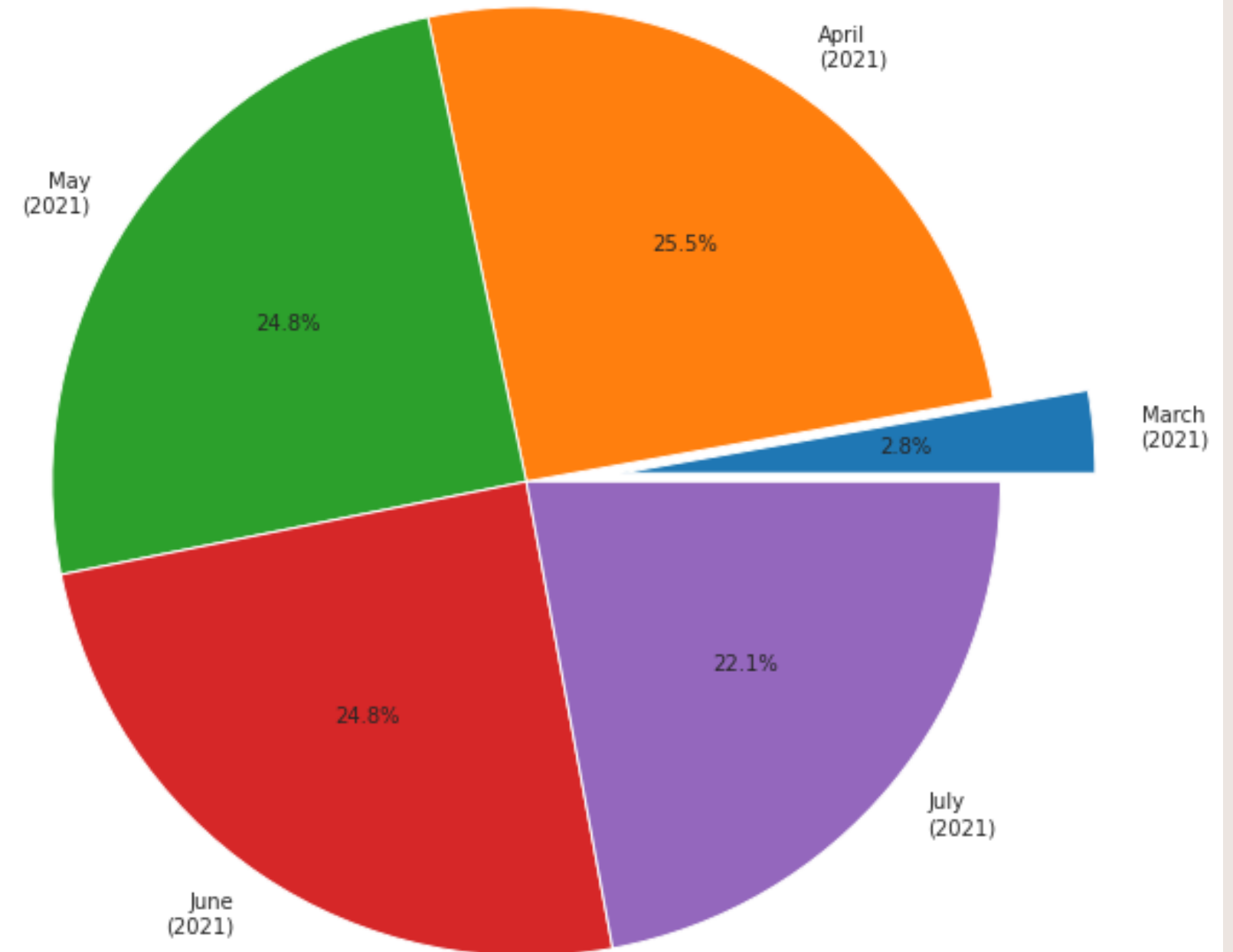
월별 카테고리 ⇒ 채널 ⇒ 비디오 개수

1) 월별 인기 비디오 개수 시각화

- * '월 별'의 기준은 데이터셋의 published_date 컬럼이다.
- * 해당 데이터는 2021.03~2021.07 데이터로 이루어져 있다.
- * 데이터가 3월 며칠부터 수집되었는가, 7월에 언제까지 수집되었는가를 확인해볼 필요가 있다.

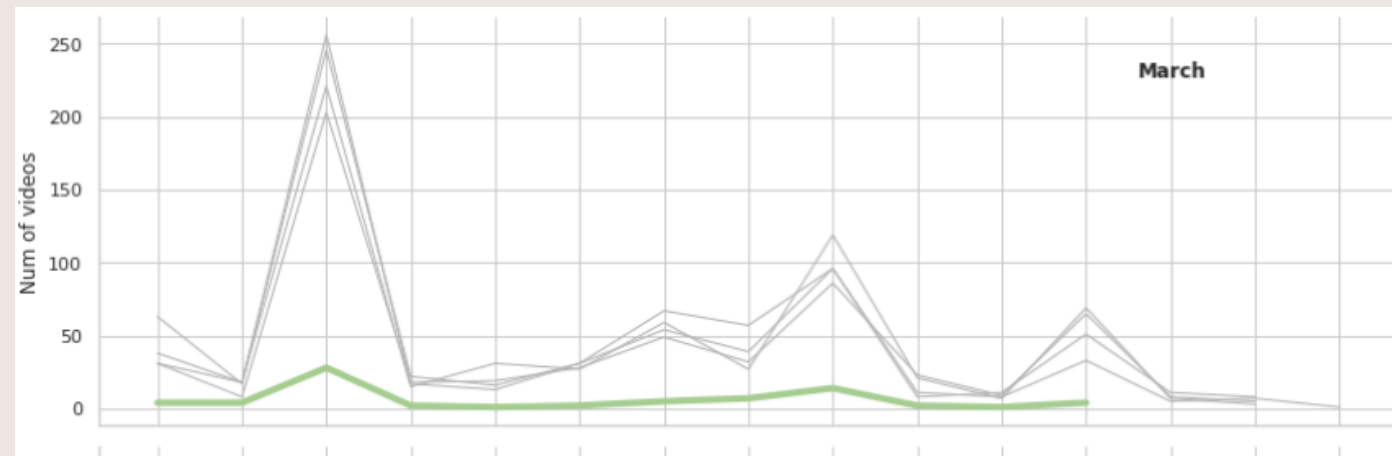
3월을 제외하고,
각 달마다 비슷한 비율로 인기 동영상 데이터가 수집되었음을 알 수 있다.

Ratio of monthly trend videos



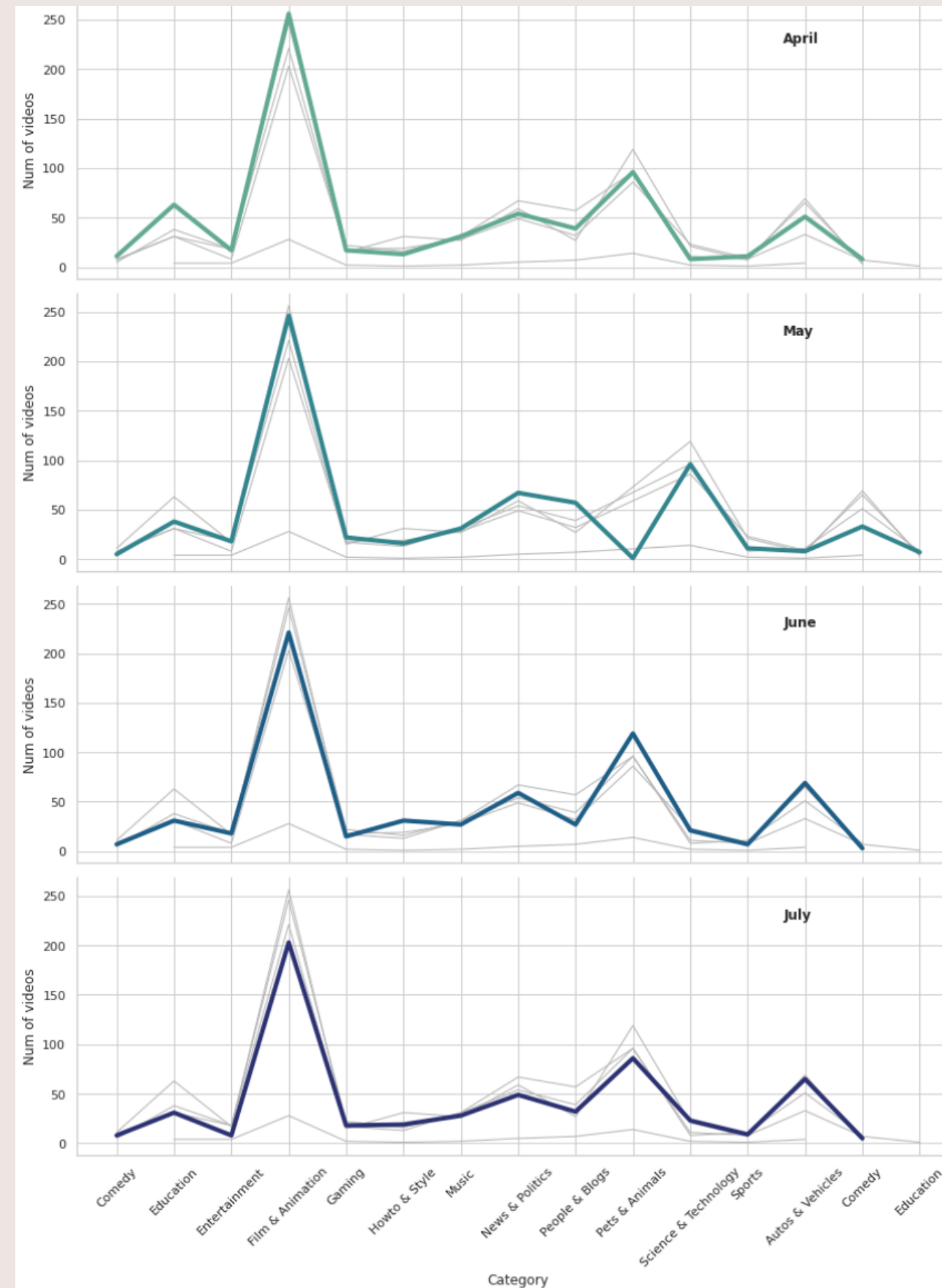
월별 카테고리 ⇒ 채널 ⇒ 비디오 개수

2) 월별 카테고리별 비디오 개수 시각화



수집된 인기 비디오 데이터 수가 상대적으로 적었던 3월을 제외하고,

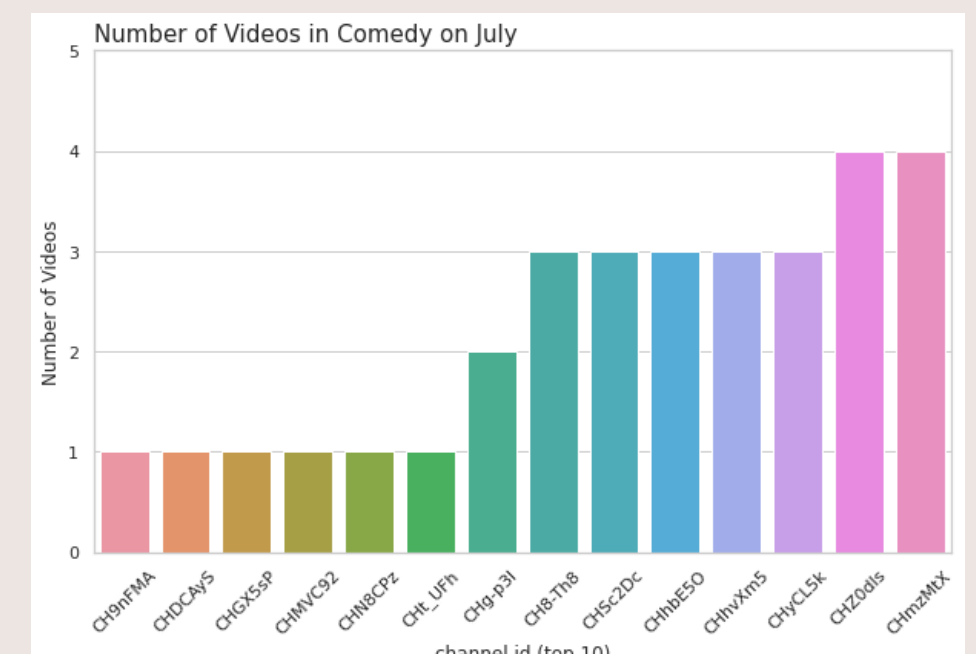
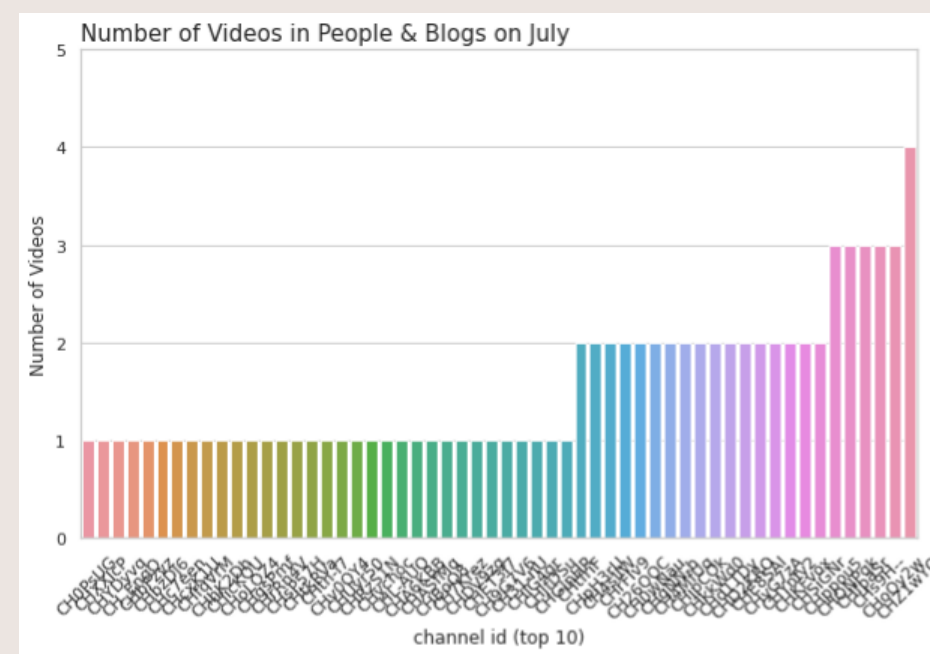
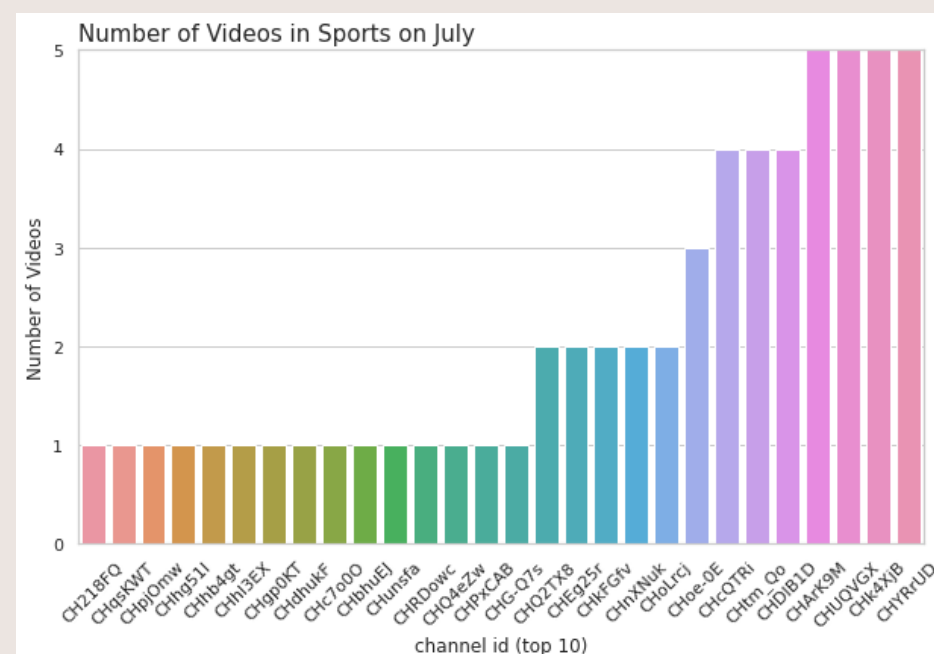
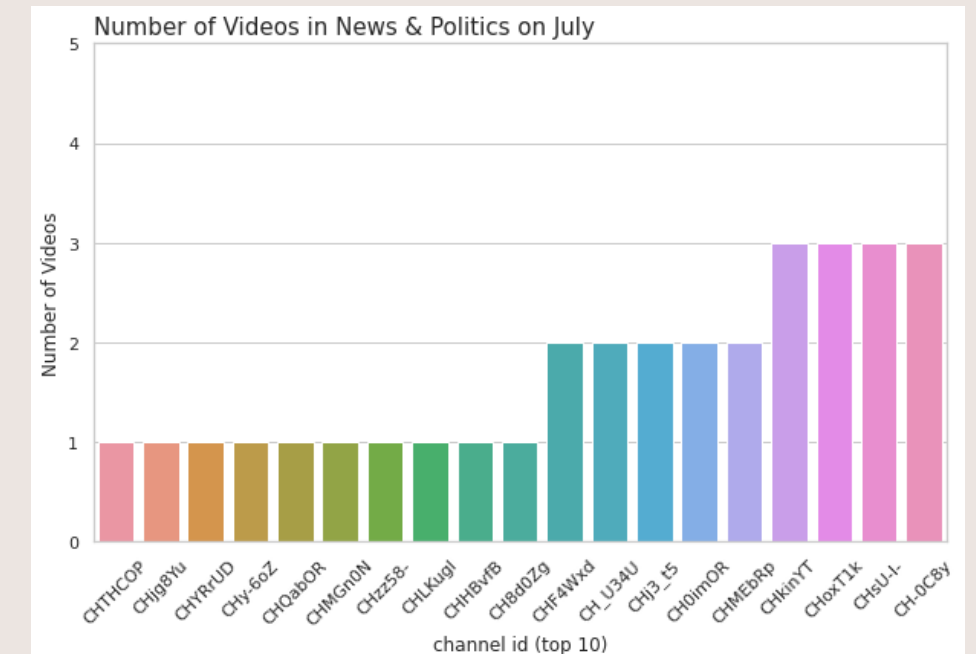
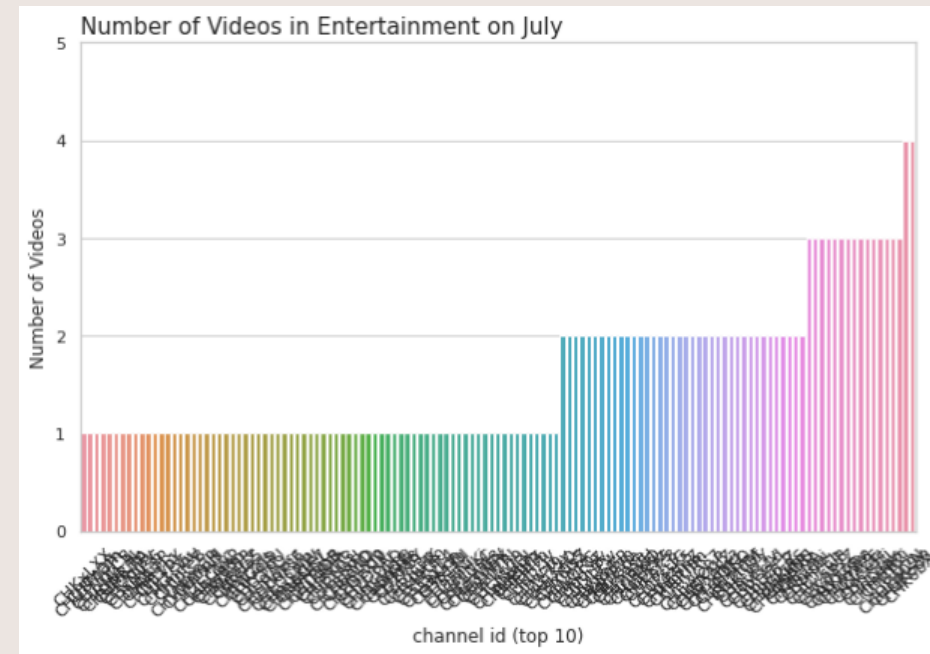
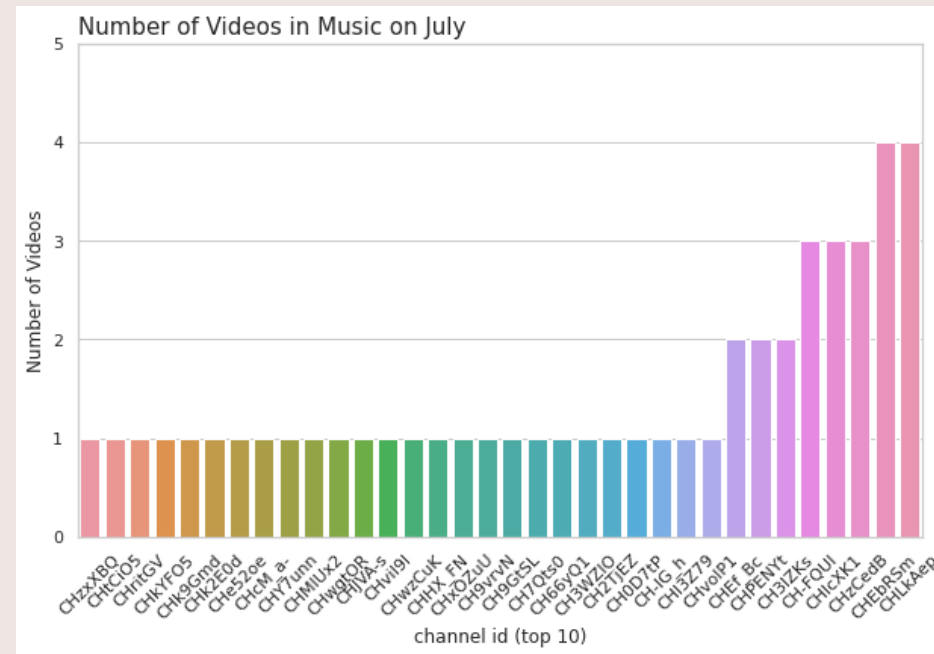
4월, 5월, 6월, 7월은 모두
카테고리 별로 인기 비디오 개수가 비슷한 추이를 보이는 것을 확인할 수 있다.



월별 카테고리 ⇒ 채널 ⇒ 비디오 개수

** 3월부터 7월까지 총 5개월, 카테고리 수 총 15개 → 총 45개의 시각화가 나오는데, 자료에 담기에 너무 많기 때문에 마지막 달인 7월을 예시로 시각화 자료를 구성하였습니다.

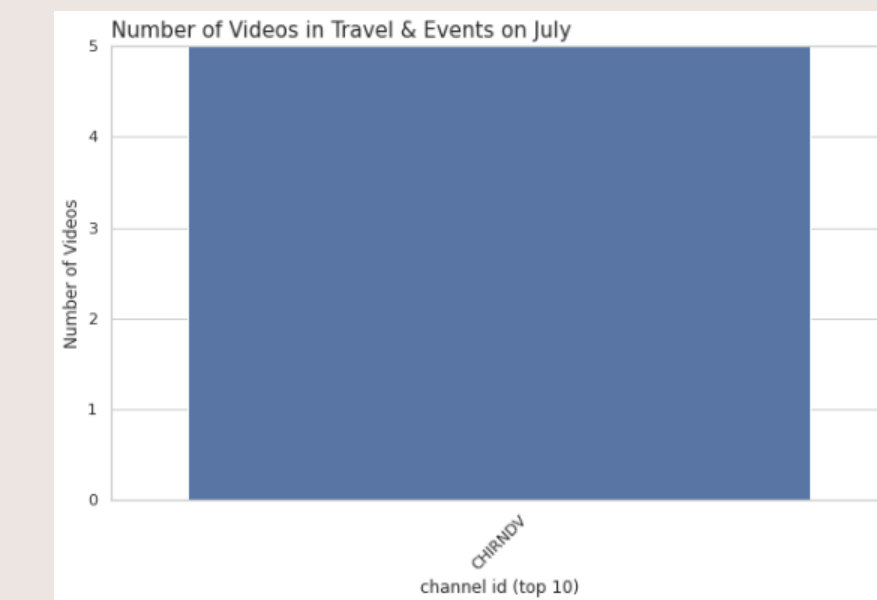
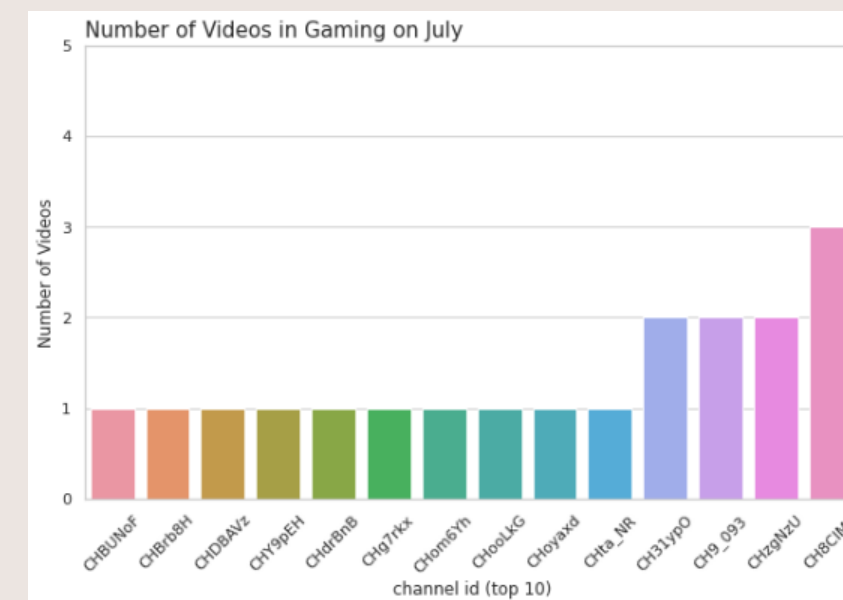
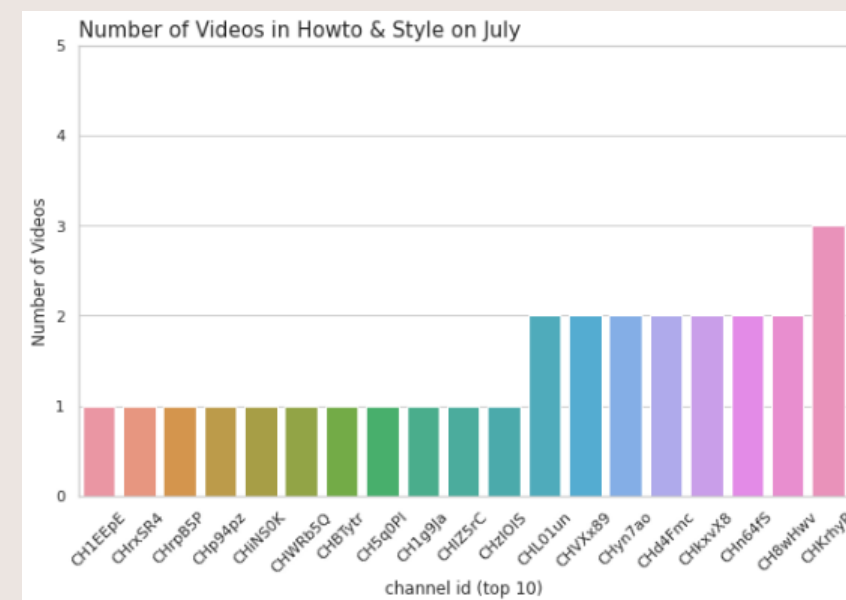
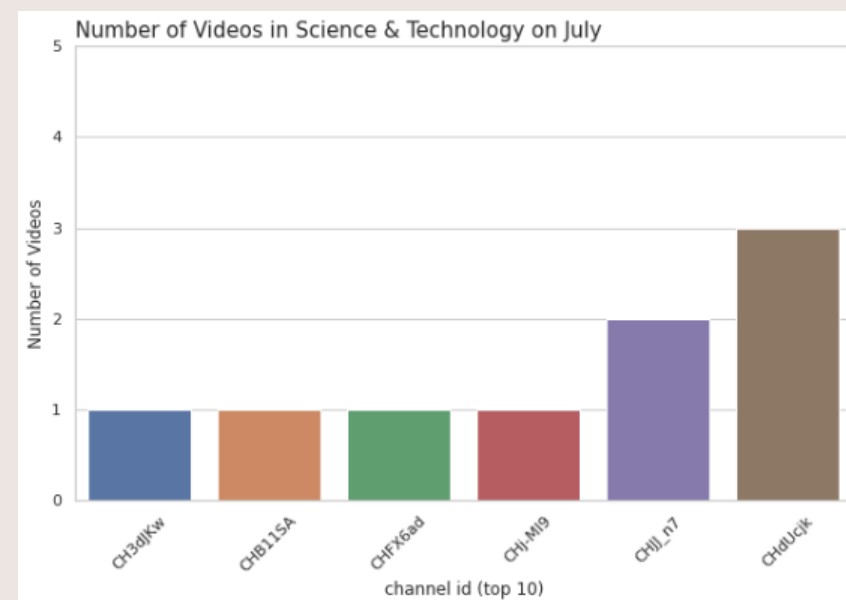
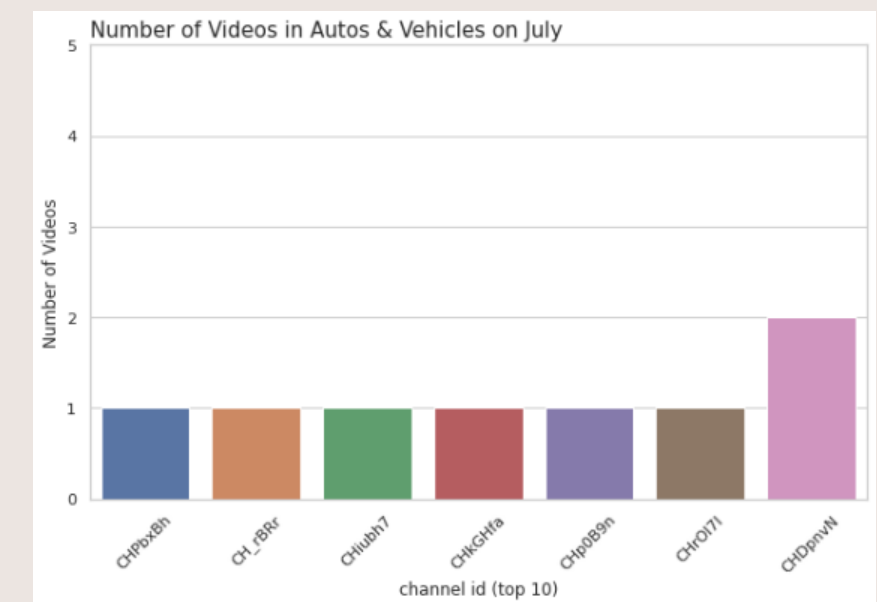
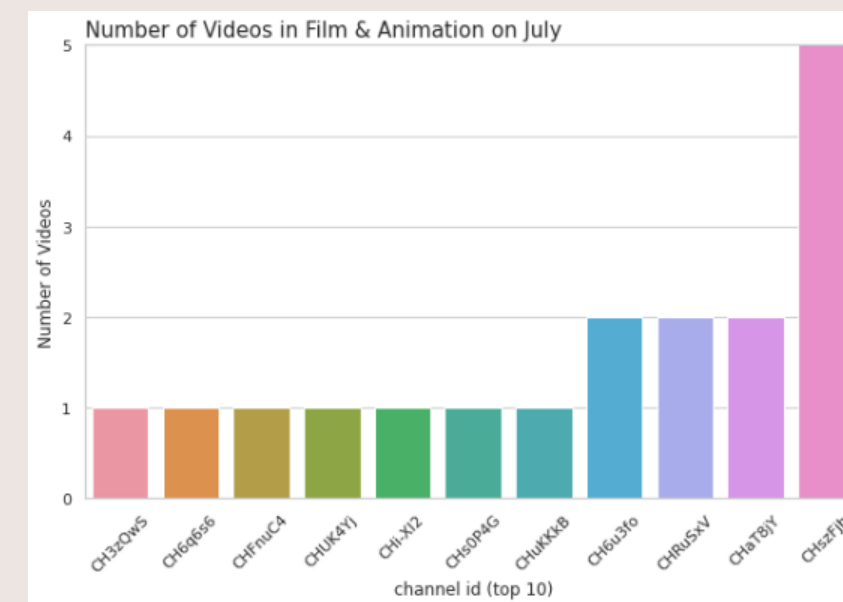
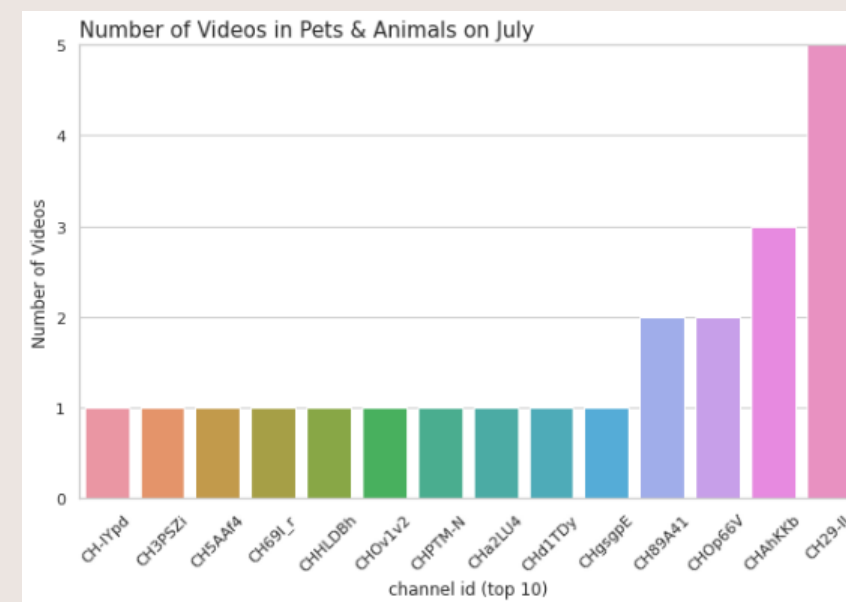
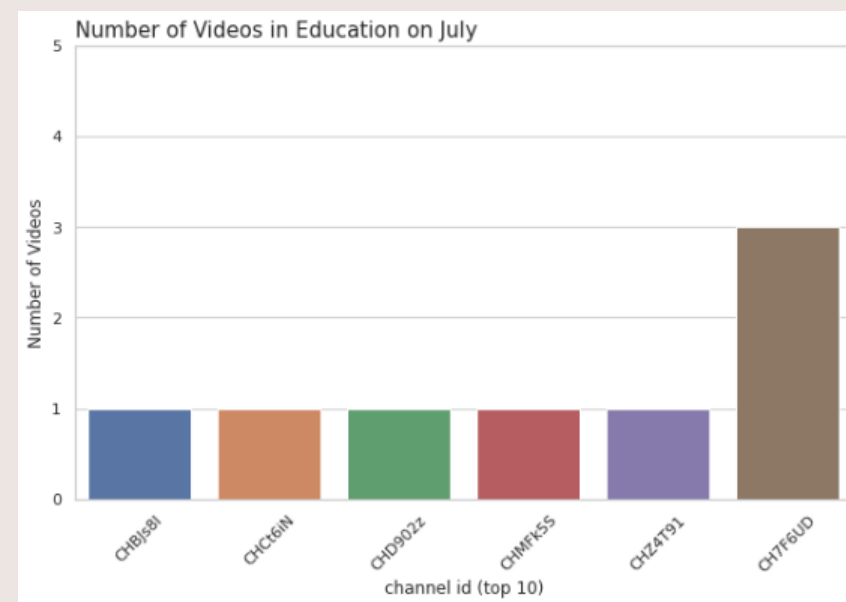
3) 월별 카테고리별 채널별 비디오 개수 (7월)



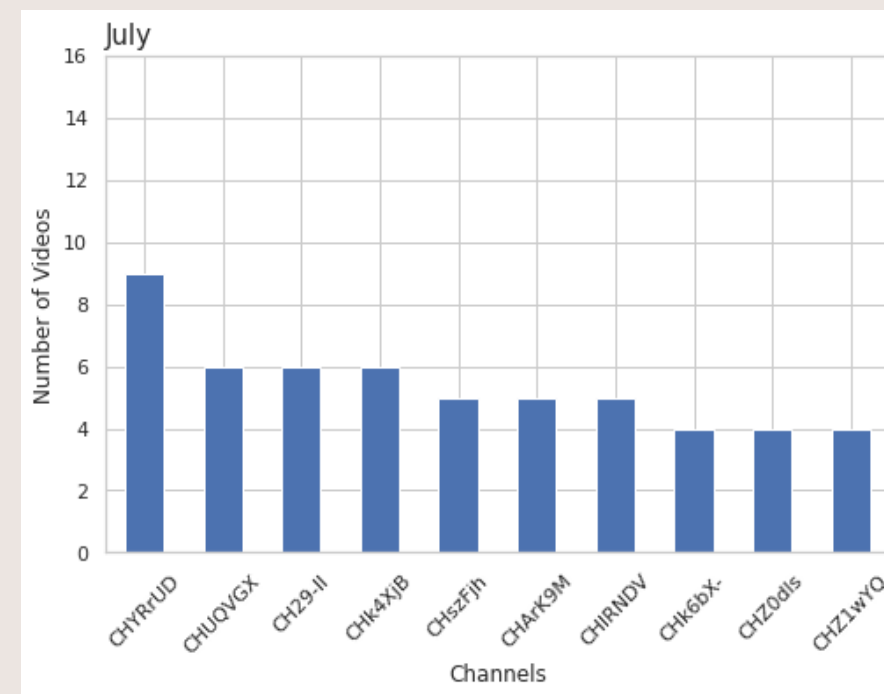
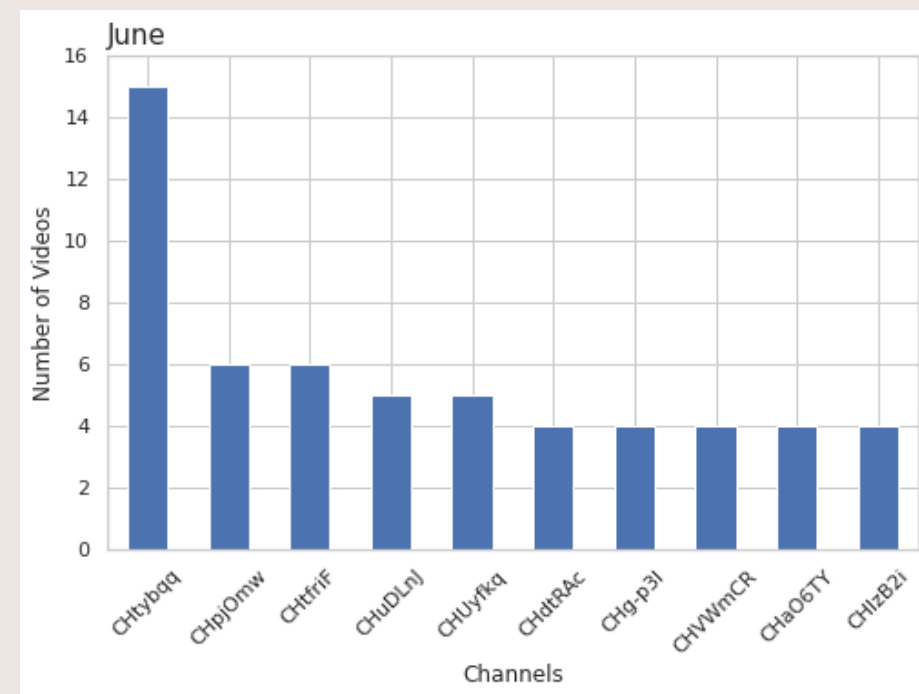
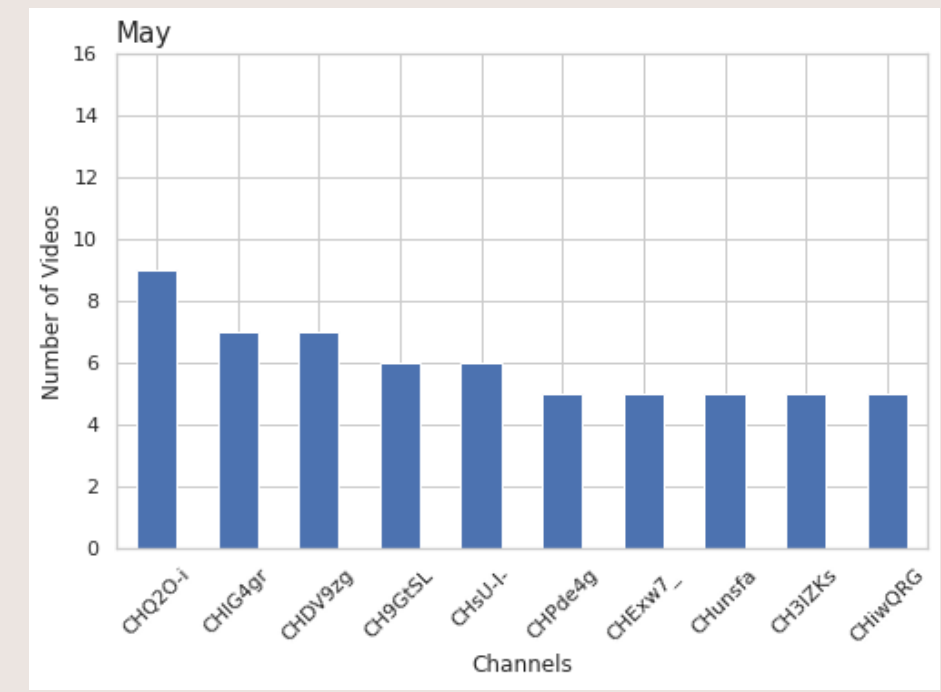
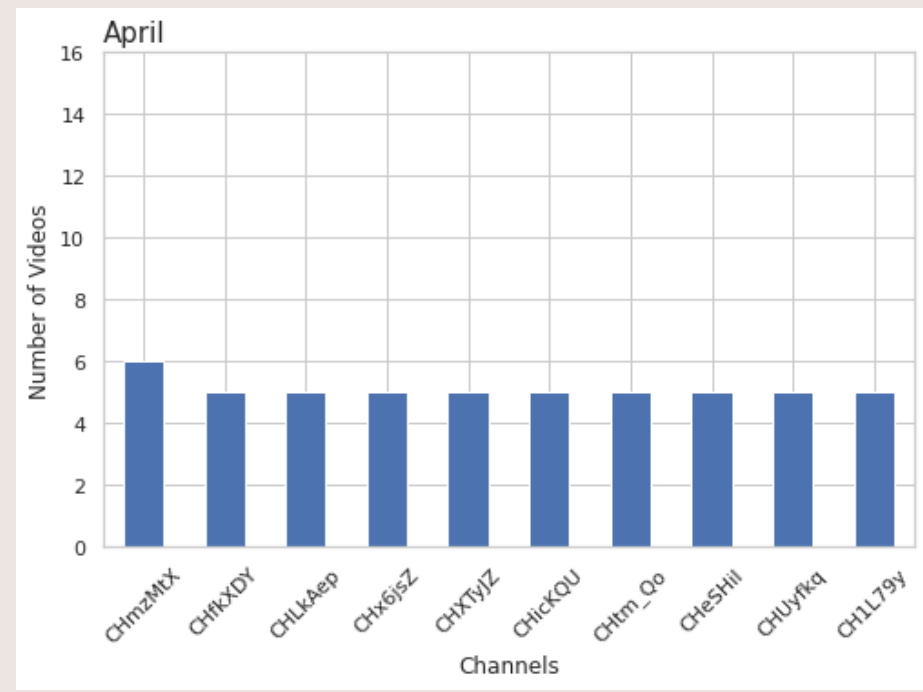
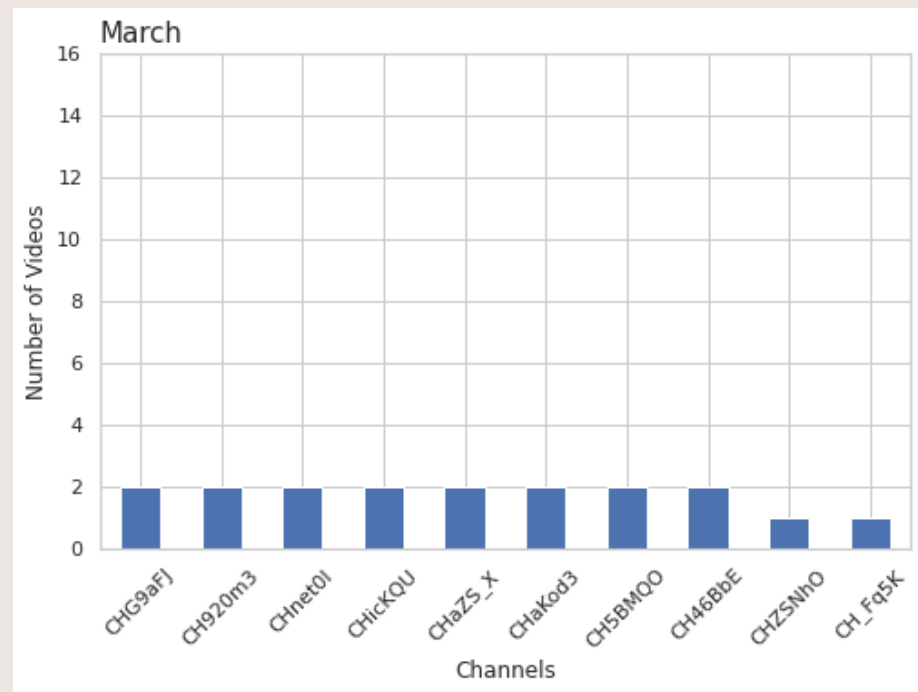
월별 카테고리 ⇒ 채널 ⇒ 비디오 개수

3) 월별 카테고리별 채널별 비디오 개수 (7월)

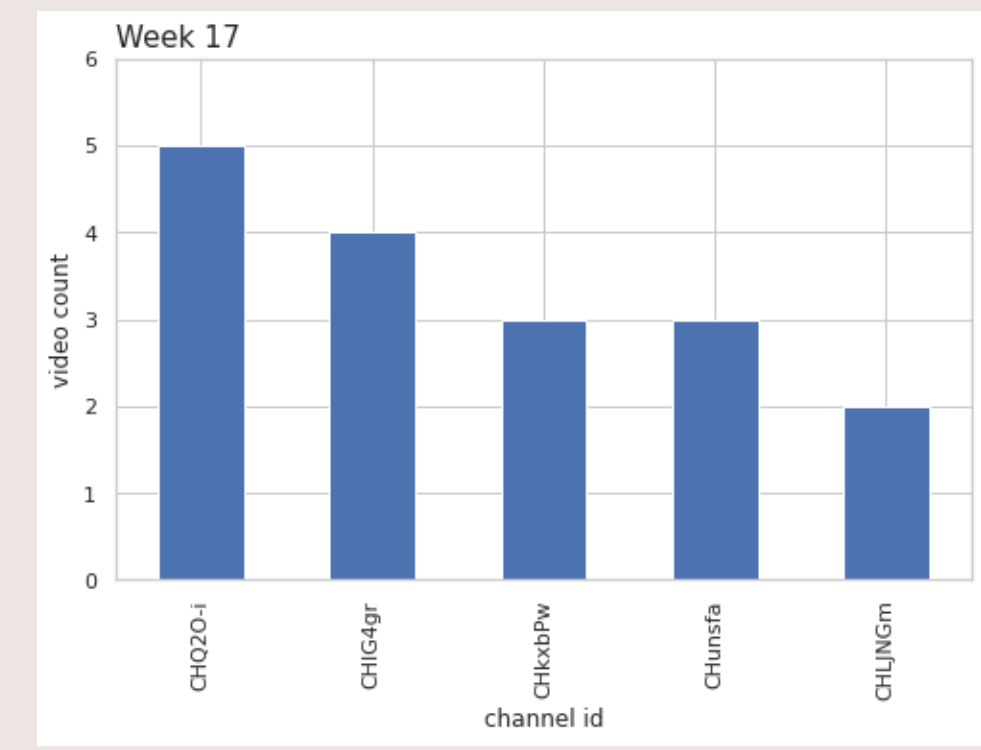
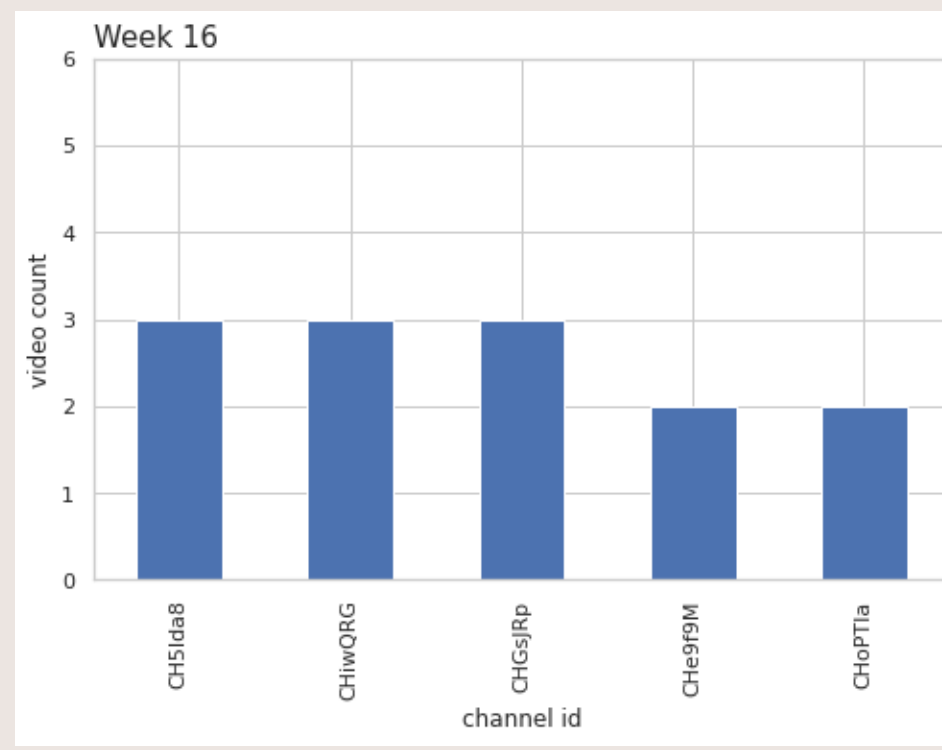
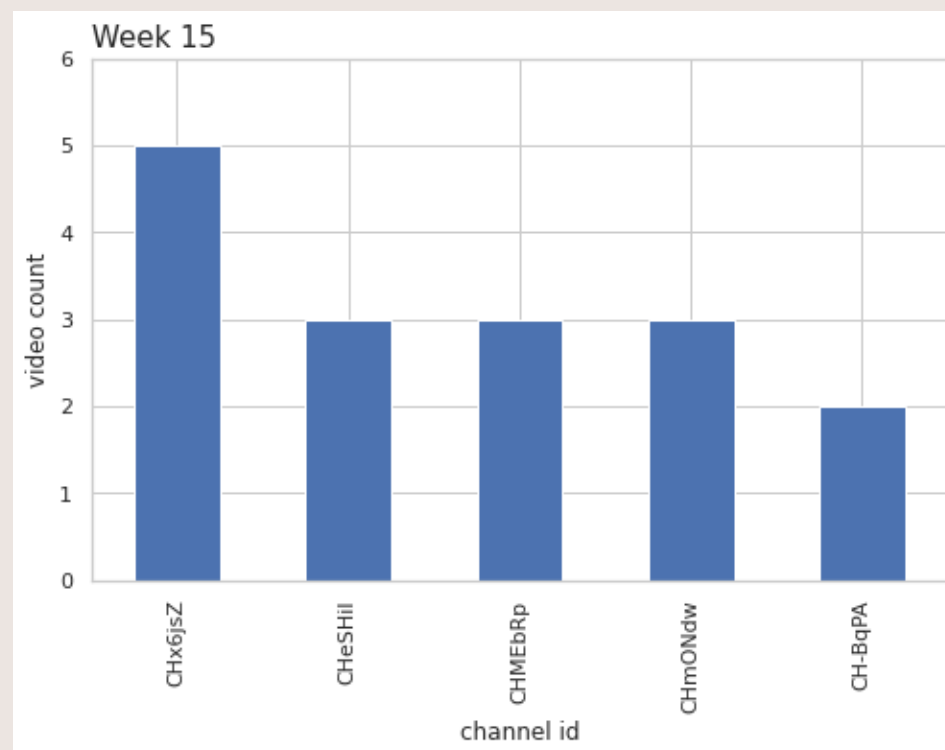
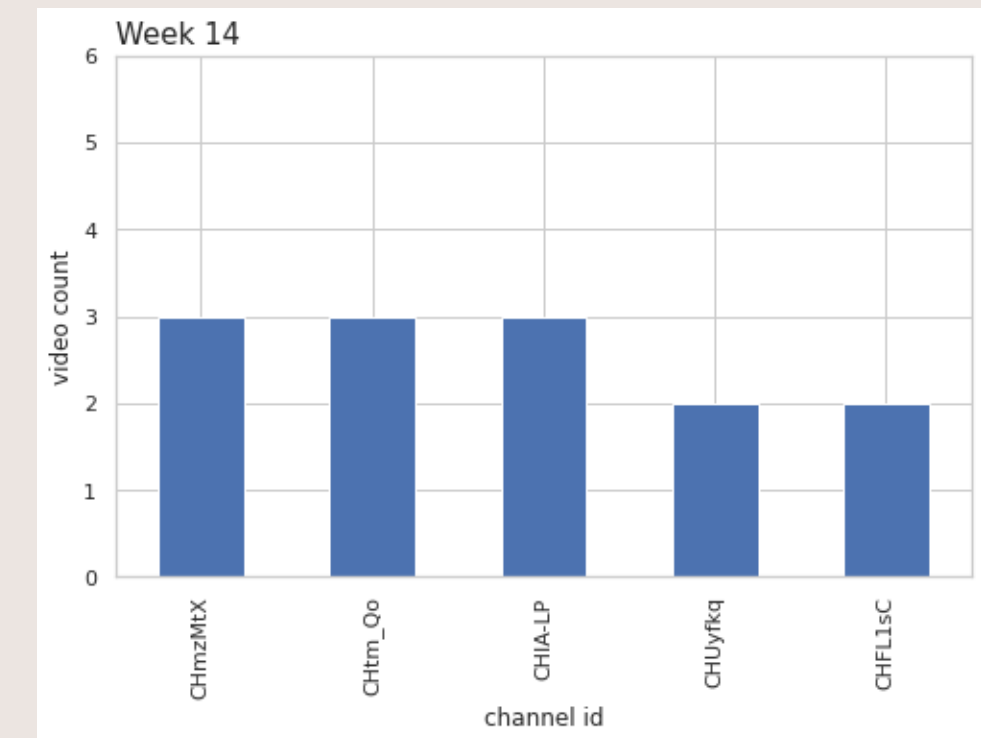
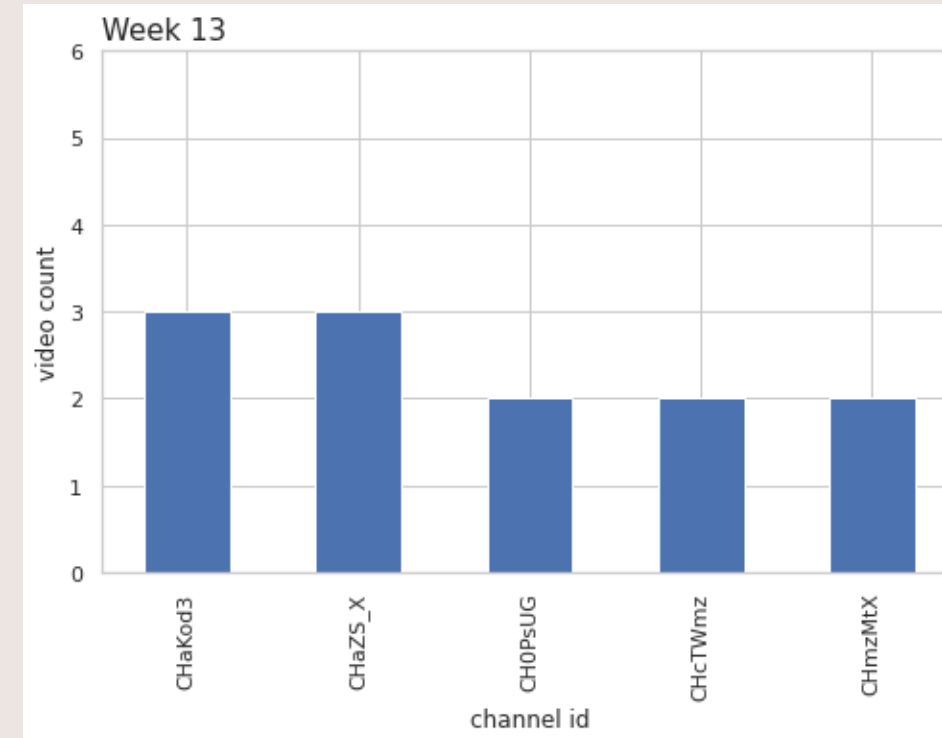
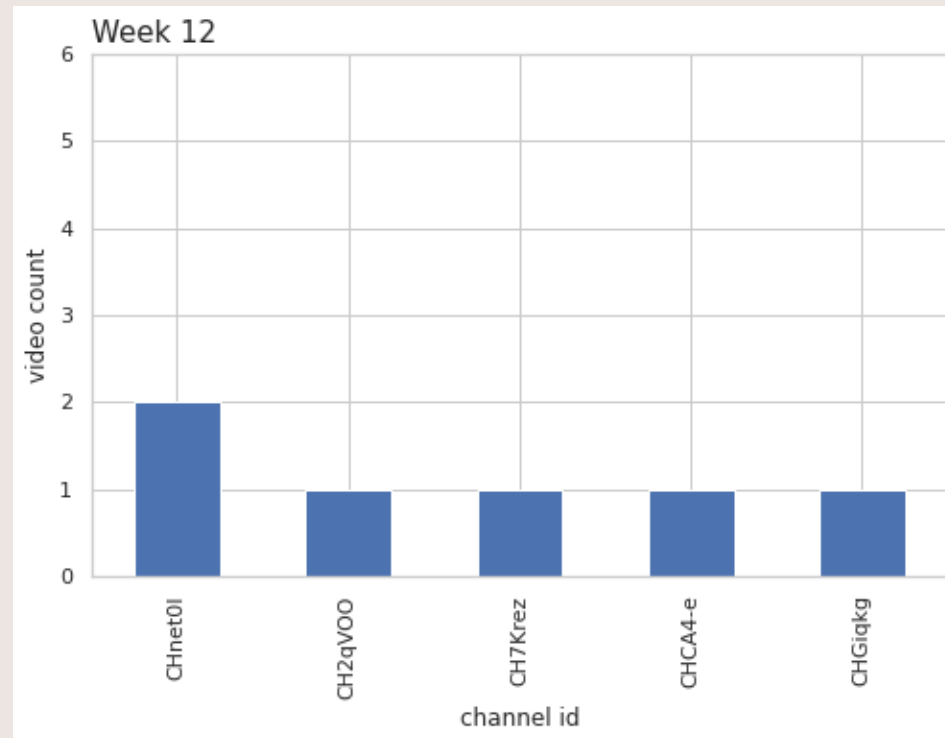
** 3월부터 7월까지 총 5개월, 카테고리 수 총 15개 → 총 45개의 시각화가 나오는데, 자료에 담기에 너무 많기 때문에 마지막 달인 7월을 예시로 시각화 자료를 구성하였습니다.



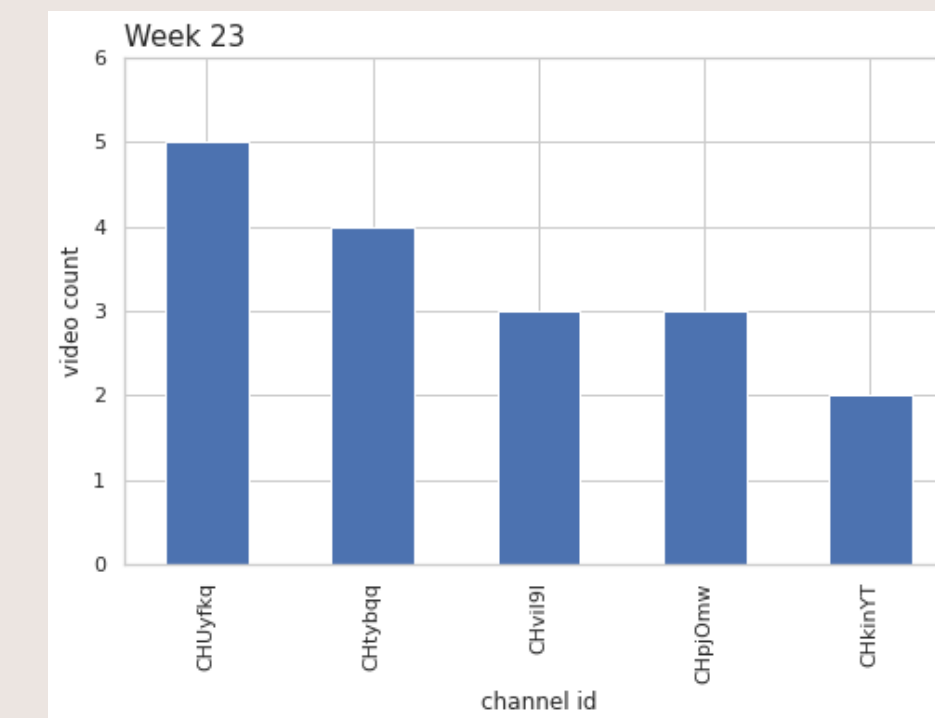
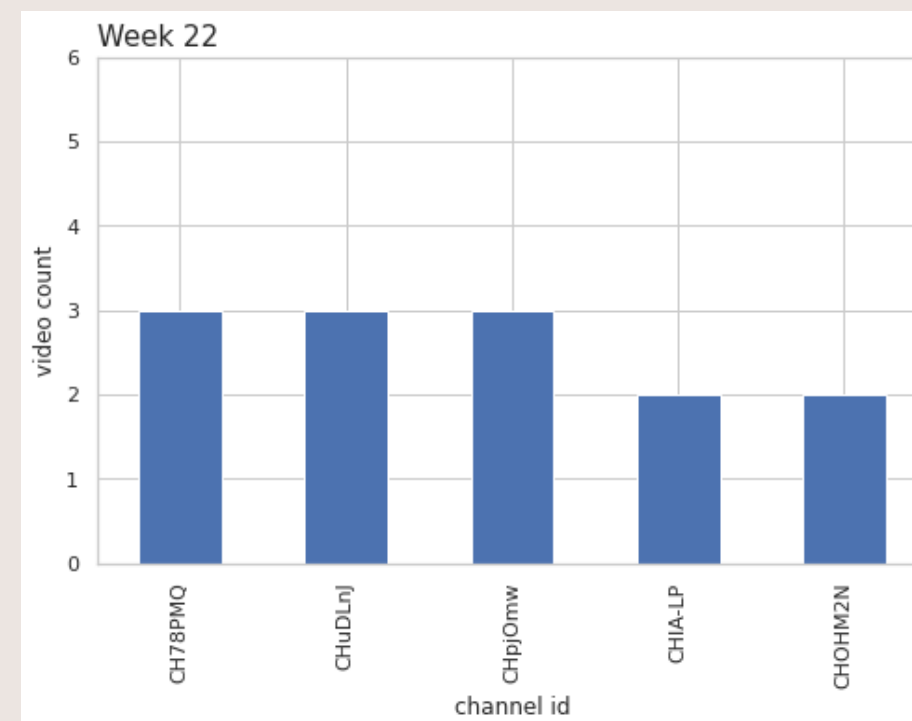
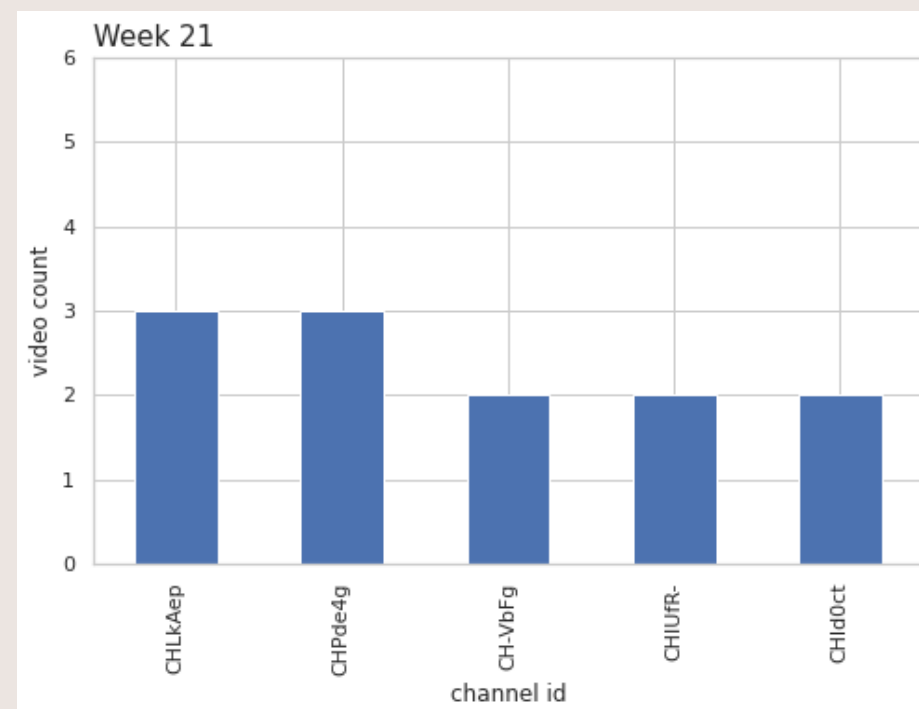
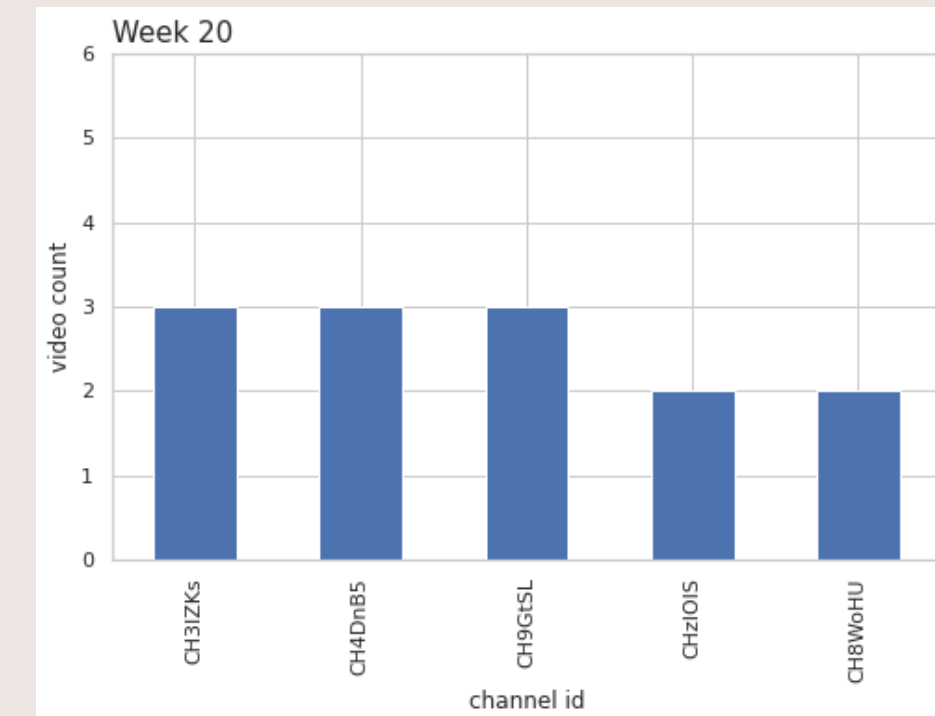
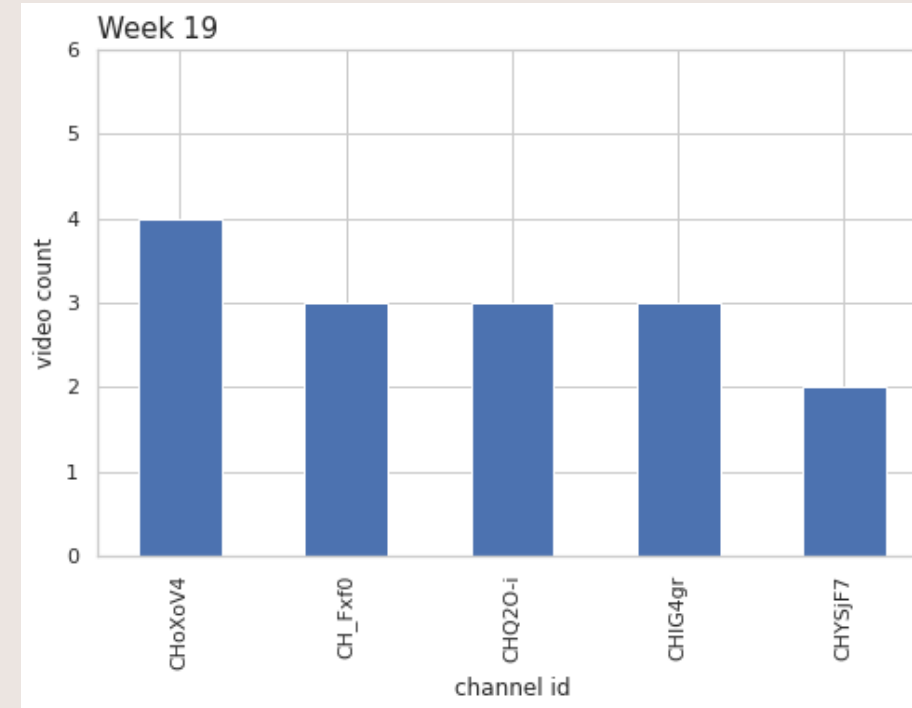
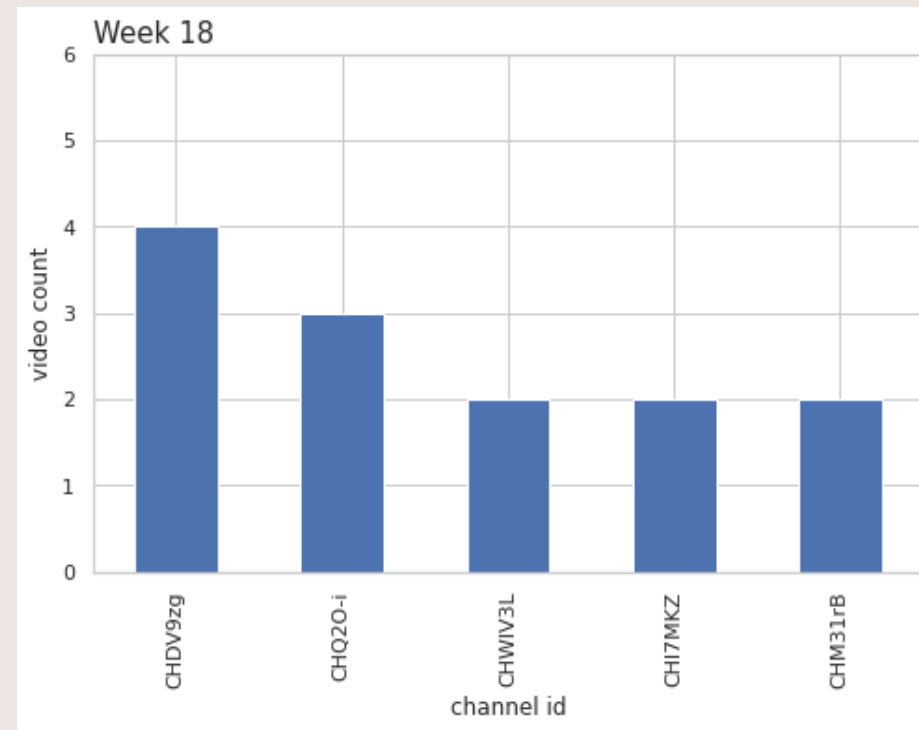
월별 TOP10 채널 (분류 기준은 비디오 개수)



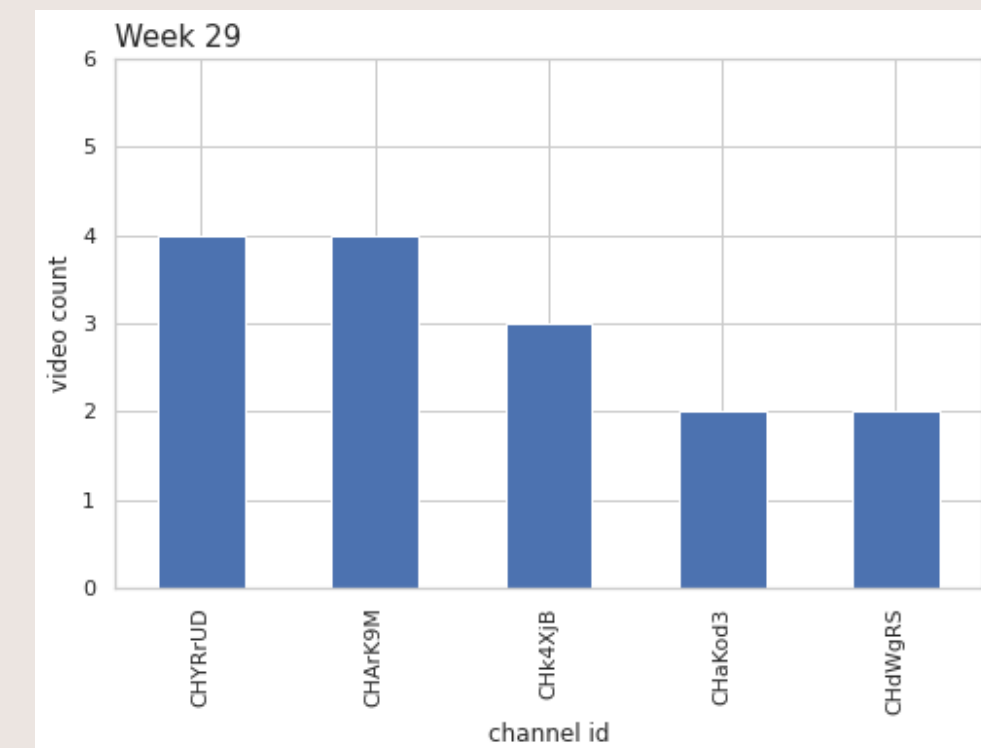
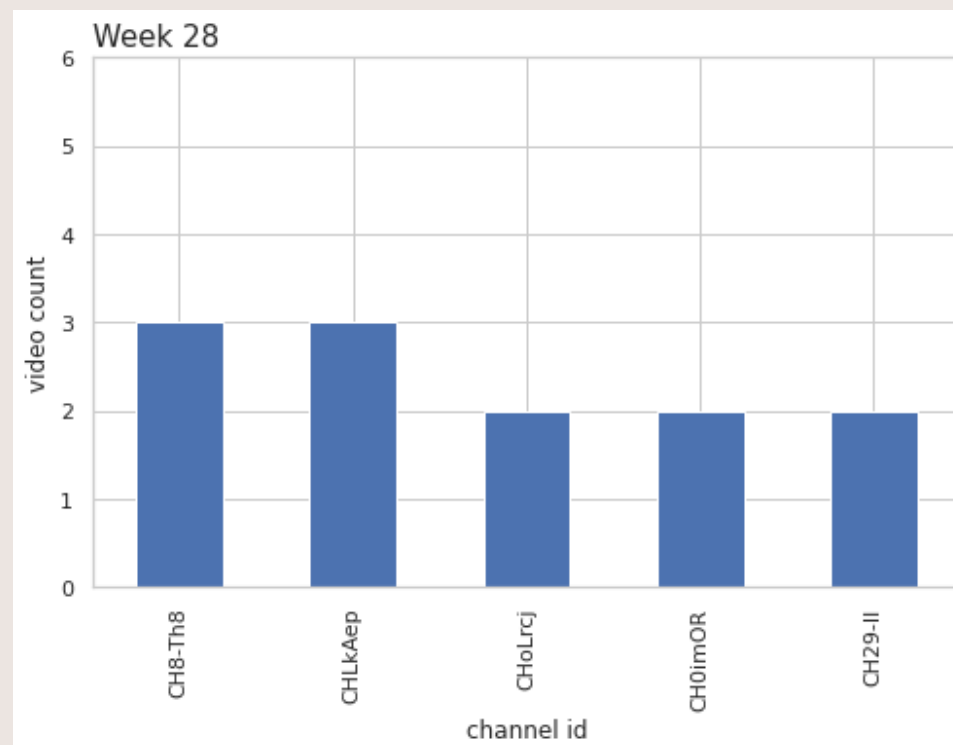
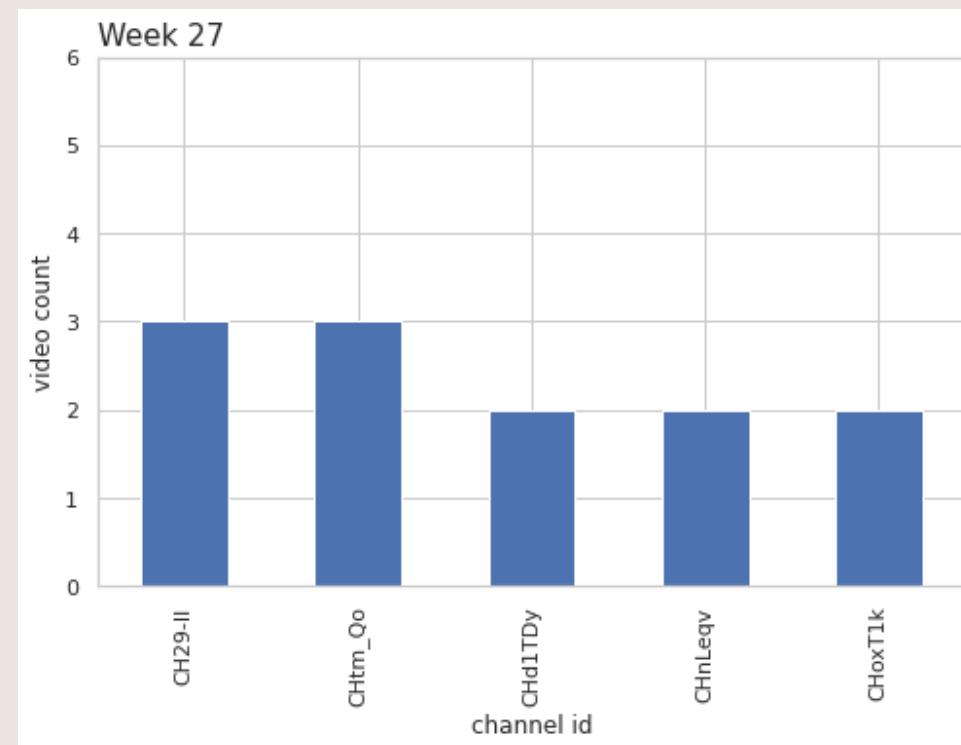
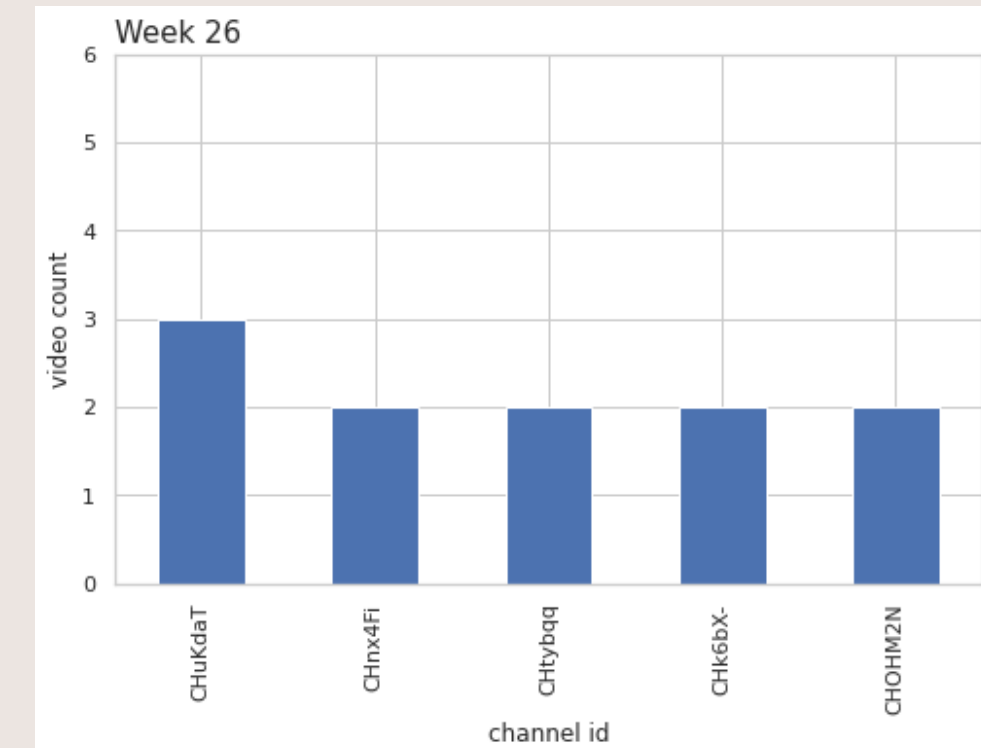
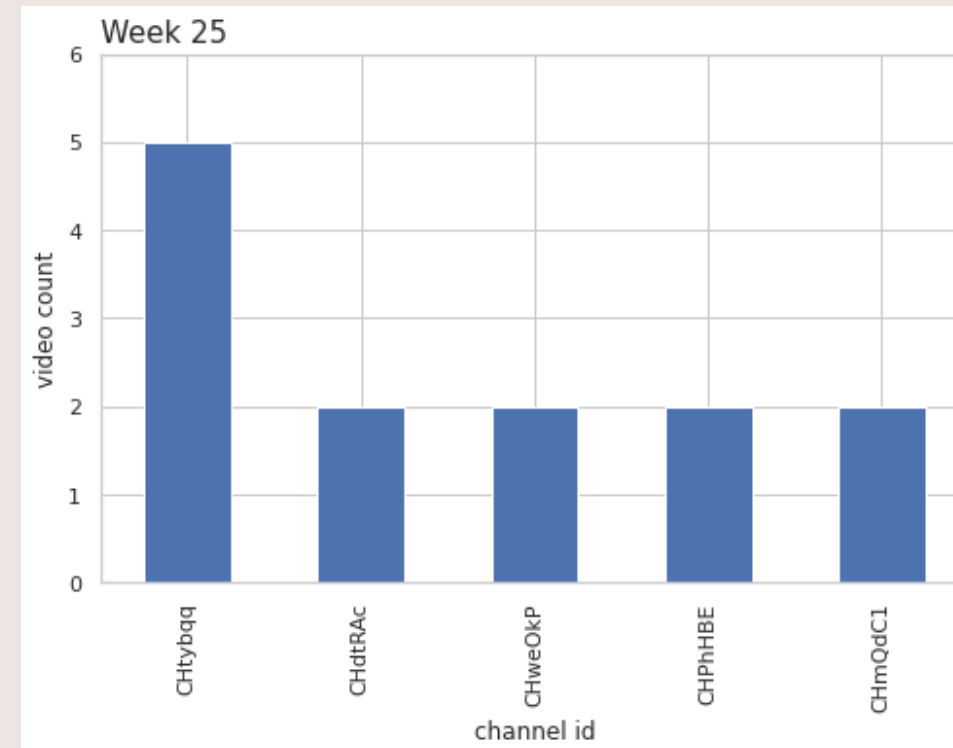
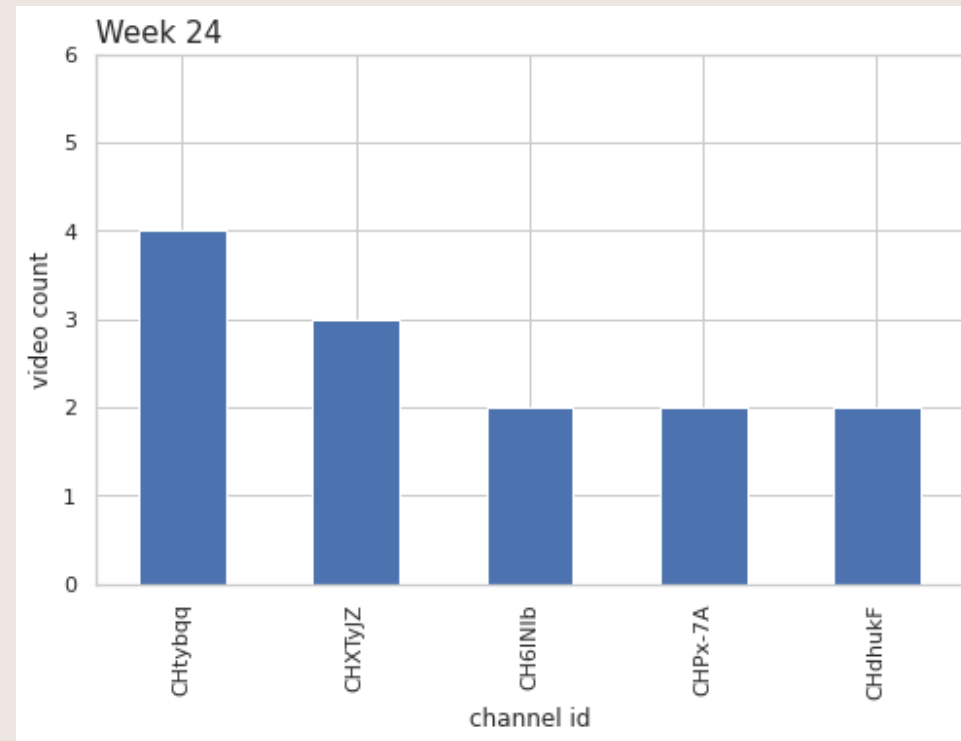
주별 TOP5 채널 (분류 기준은 비디오 개수)



주별 TOP5 채널 (분류 기준은 비디오 개수)

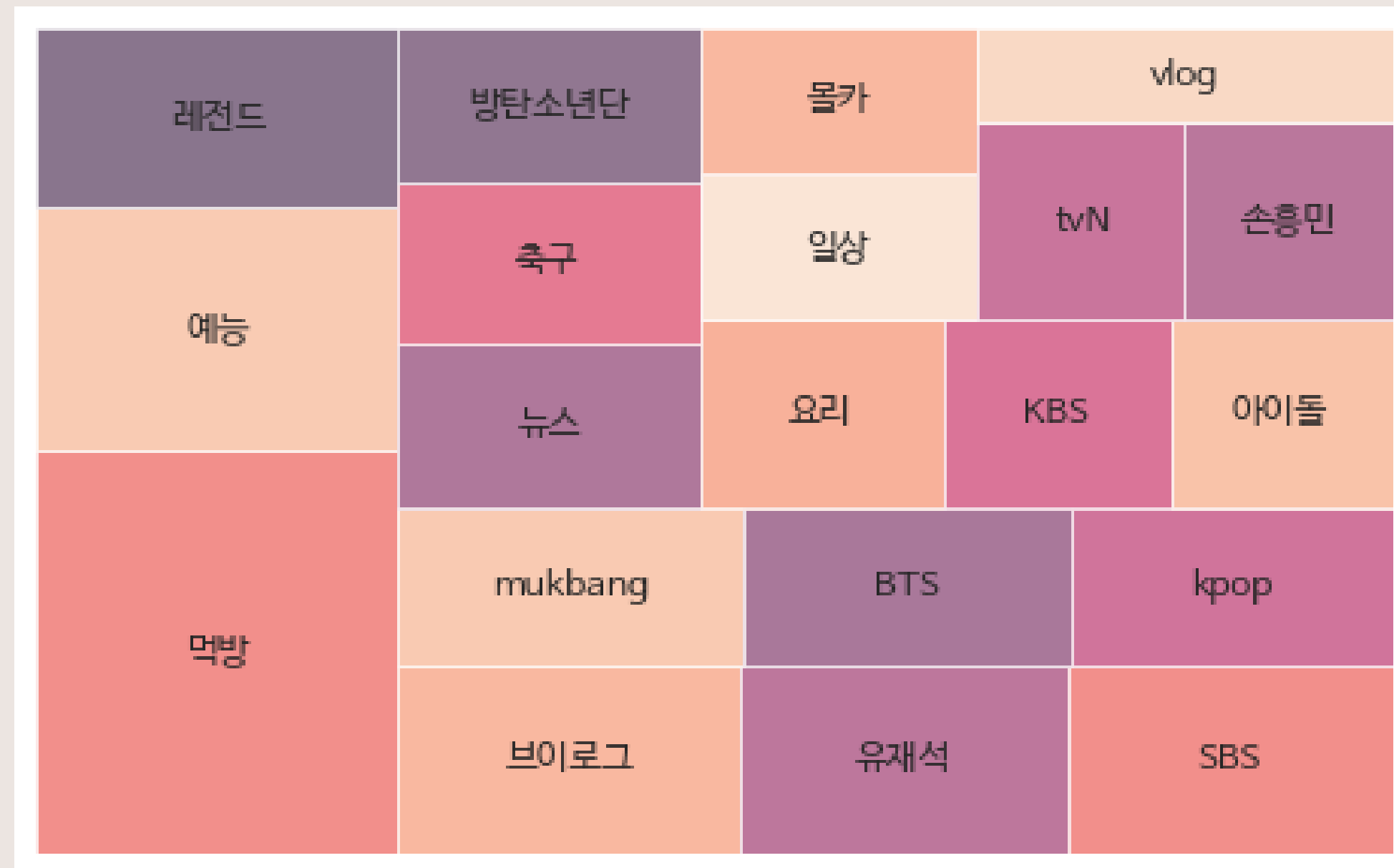


주별 TOP5 채널 (분류 기준은 비디오 개수)



월별 카테고리별 태그 키워드 순위

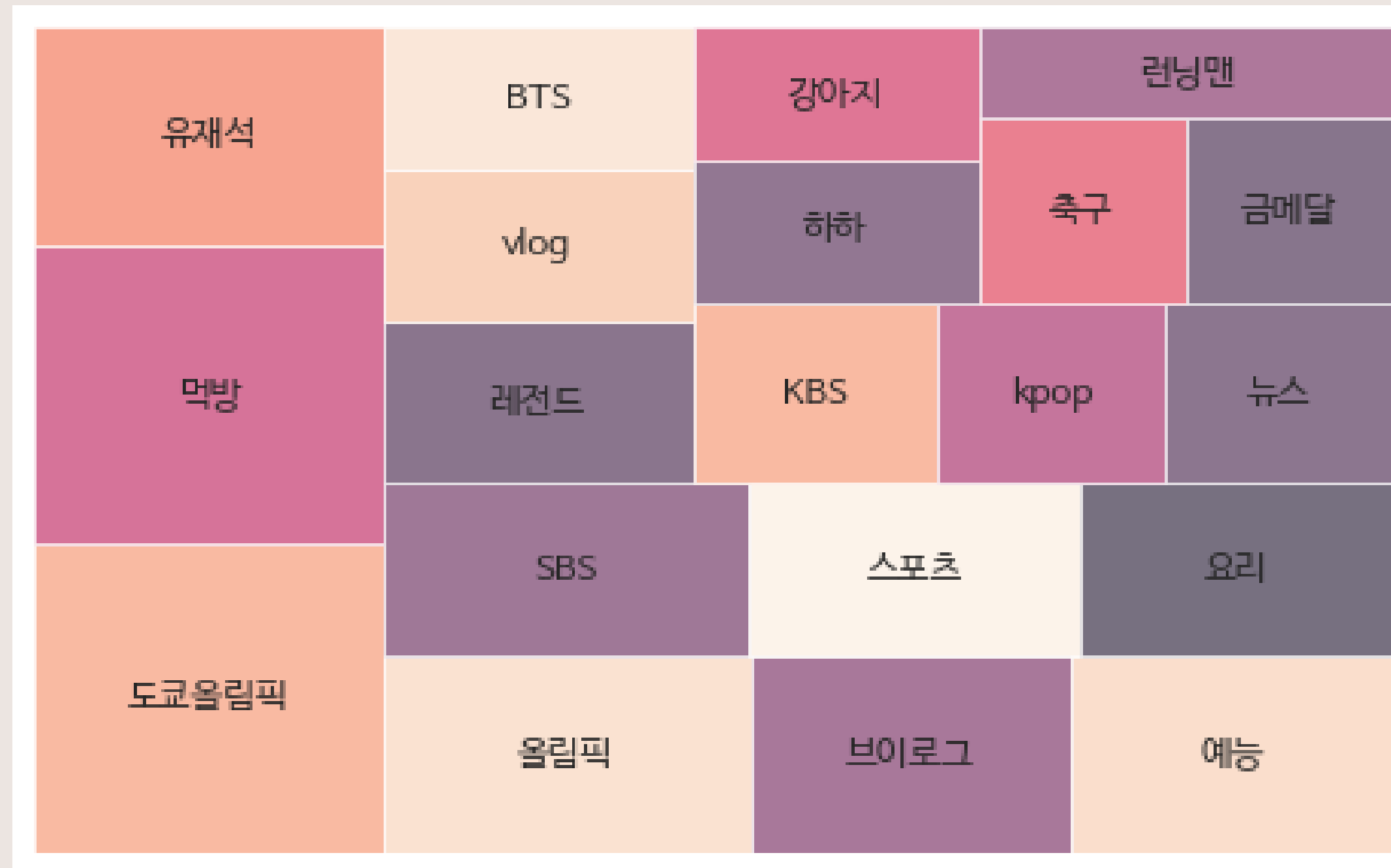
1) 3월~7월 전체 등장 비율 상위 20위 안의 태그 시각화



월별 카테고리별 태그 키워드 순위

2) 7월 전체 등장 비율 상위 20위 안의 태그 시각화

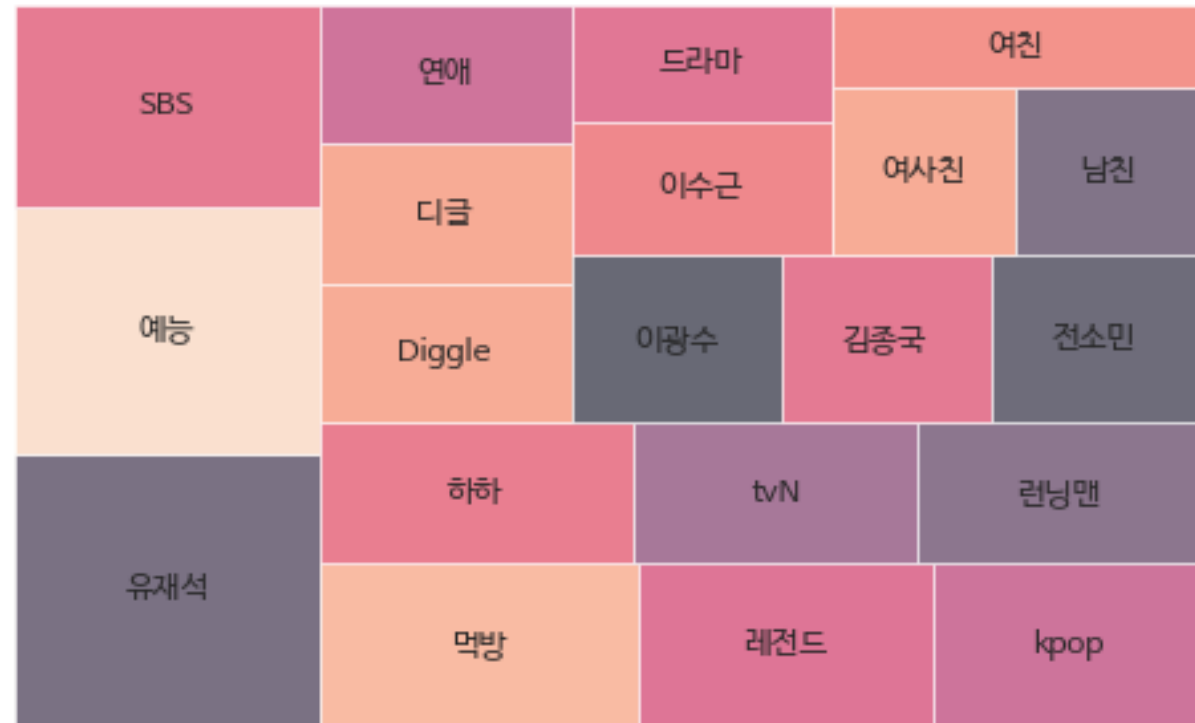
** 3월부터 7월까지 총 5개월, 카테고리 수 총 15개 → 총 45개의 시각화가 나오는데, 자료에 담기에 너무 많기 때문에 마지막 달인 7월을 예시로 시각화 자료를 구성하였습니다.



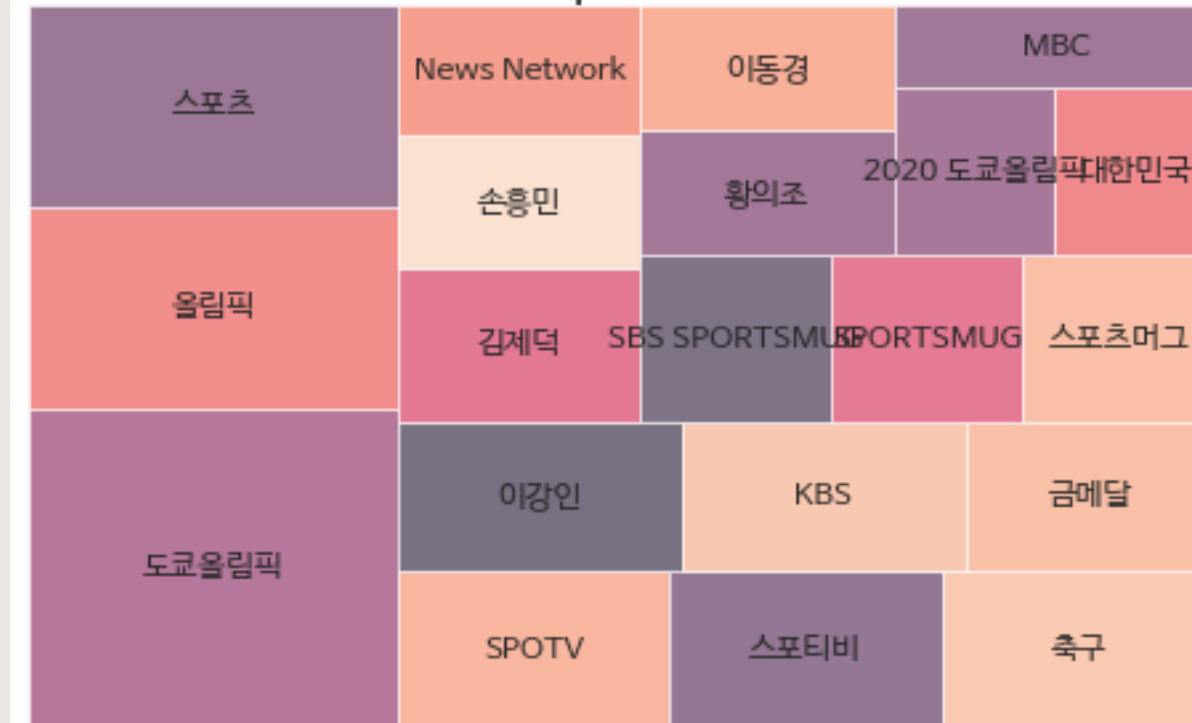
월별 카테고리별 태그 키워드 순위

3) 7월 내 각 카테고리 별 상위 20개 키워드 (상위 6개 카테고리의 키워드)

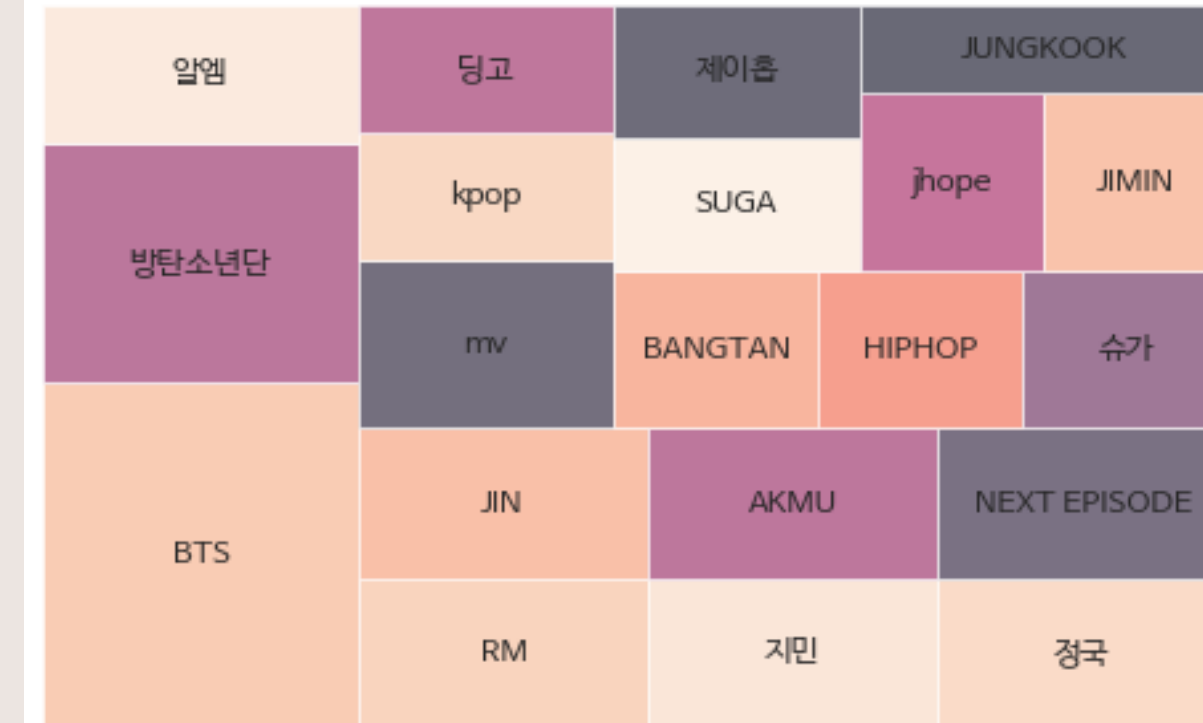
Entertainment



Sports

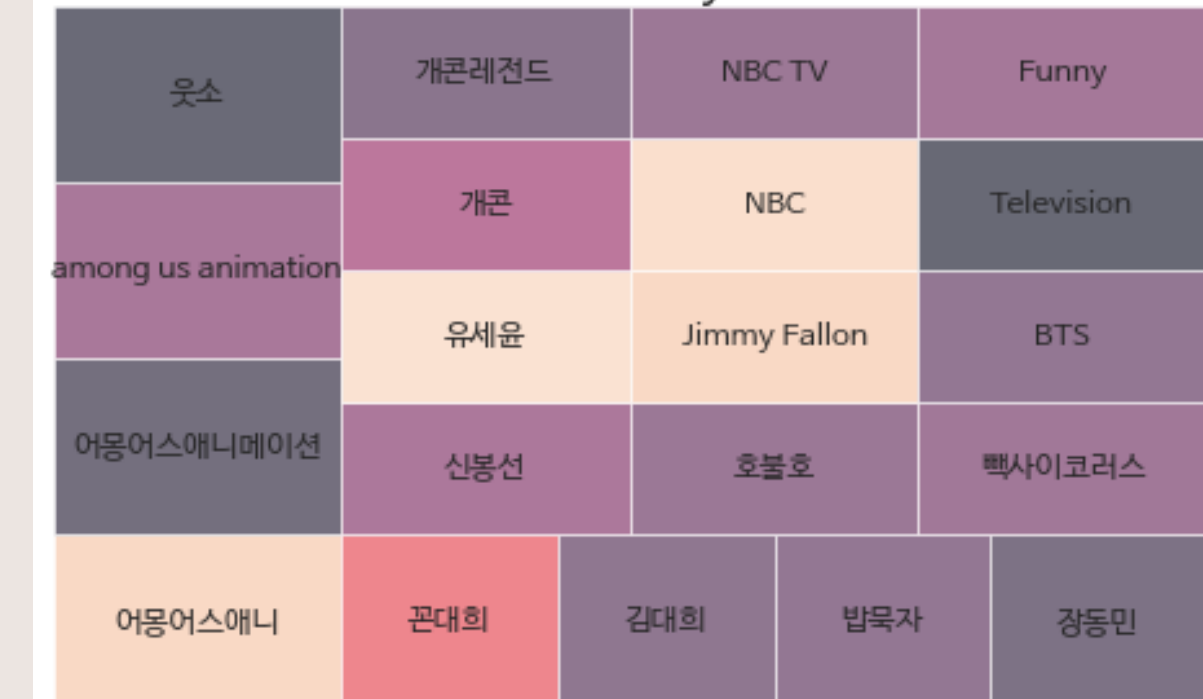
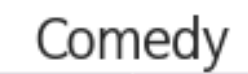
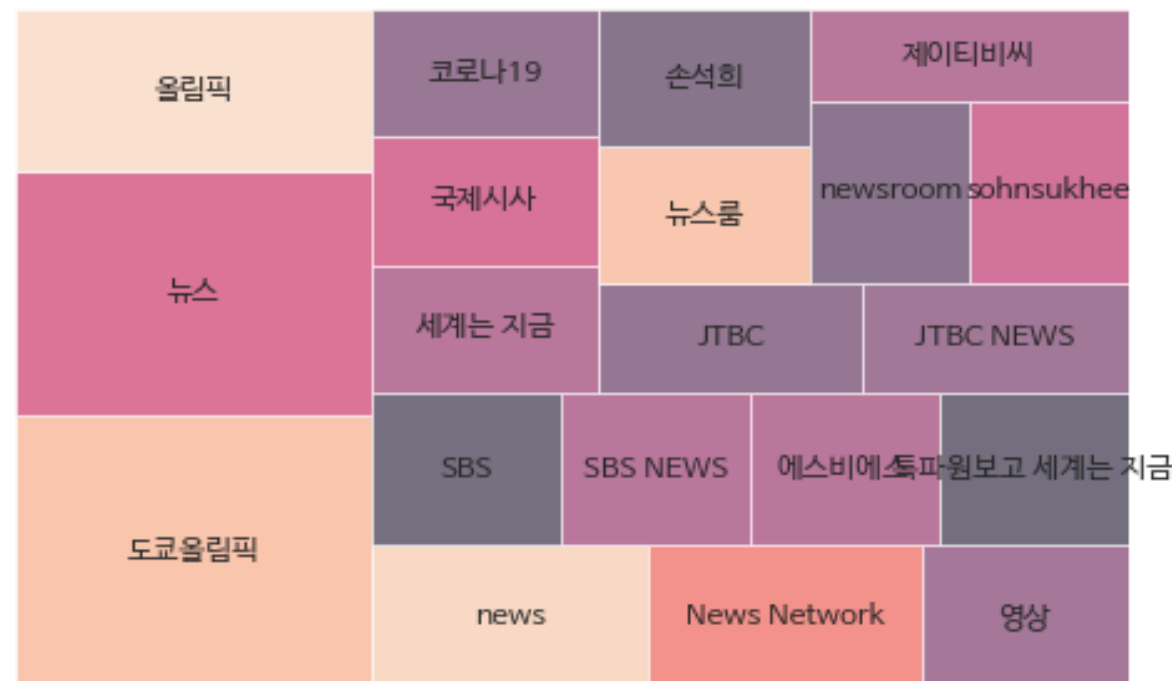
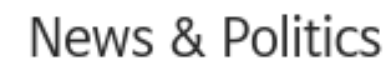


Music



월별 카테고리별 태그 키워드 순위

3) 7월 내 각 카테고리 별 상위 20개 키워드 (상위 6개 카테고리의 키워드)



2. 인기 동영상 지표 개발 및 engagement와 상관관계 확인

＼ 비디오를 인기 동영상 기준에 부합하도록 분류할수 있는 새로운 지표를 개발하기

1) publish_to_trend : published 된 후, 인기 동영상이 되기 까지 기간

2) trend_days : 인기 동영상 지속 기간

3) 하루 평균 증가한 Likes 수

4) 하루 평균 증가한 dislikes

5) 하루 평균 작성된 comment 수

6) 지표5. Likes/Dislikes 비율

＼ 해당 지표가 engagement와 어떤 상관관계가 있는지 설명하기

지표 1-1. publish_to_trend :published 된 후, 인기 동영상이 되기 까지 기간

▶ $\text{publish_to_trend} = \text{published_date} - \text{on_trending_date}$

지표 1-2. trend_days : 인기 동영상 지속 기간

▶ $\text{trend_days} = \text{off_trending_date} - \text{on_trending_date}$

지표 2-1. likes_growth/dislikes_growth : 인기 동영상 기간 동안 증가한 Likes 수/Dislikes 수

▶ $\text{likes_growth} = \text{off_likes} - \text{on_likes}$

▶ $\text{dislikes_growth} = \text{off_dislikes} - \text{on_dislikes}$

지표 2-2. likes_growth_per_day/Dislikes_growth_per_day : 인기 동영상 기간 내 하루 평균 증가한 Likes 수/ Dislikes 수

▶ $\text{likes_growth_per_day} = \text{likes_growth} / \text{trend_days}$

▶ $\text{Dislikes_growth_per_day} = \text{dislikes_growth} / \text{trend_days}$

지표 3-1. comments_growth : 인기 동영상 기간 동안 증가한 comments 수

▶ $\text{comments_growth} = \text{off_comments} - \text{on_comments}$

지표 3-2. comments_growth_per_day : 인기 동영상 기간 내 하루 평균 증가한 comments 수

▶ $\text{comments_growth_per_day} = \text{comments_growth} / \text{trend_days}$

지표 4. Likes Dislikes Ratio (기준: category_name)

▶ $\text{like_dislike_ratio} = \text{total_df.groupby('category_name')['likes_growth'].agg('sum')} / \text{total_df.groupby('category_name')['dislikes_growth'].agg('sum')}$

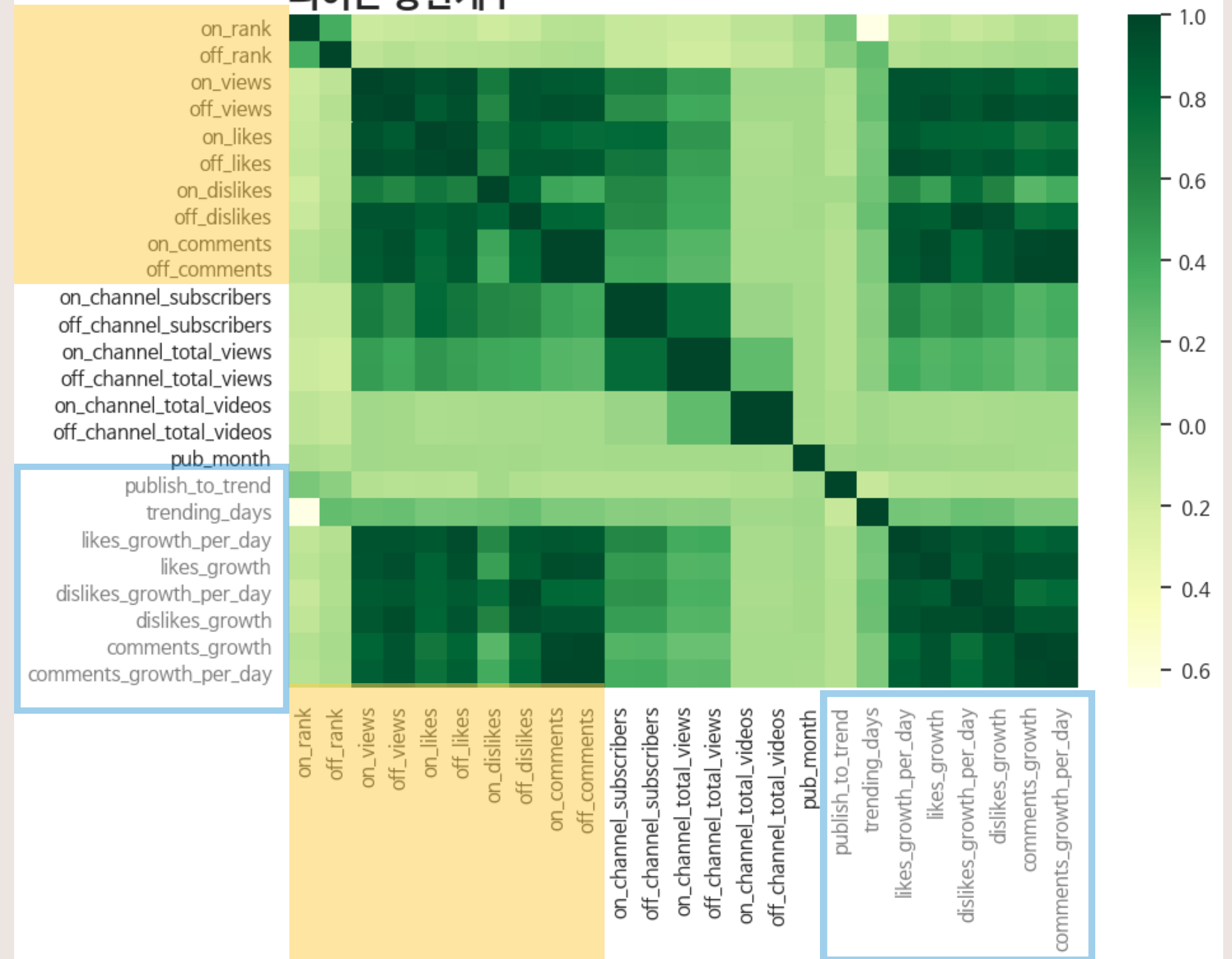
* 히트맵으로 데이터셋의 컬럼 간 상관관계를 시각화한 자료

- 판단지표와 새롭게 만든 지표가 겹치는 부분을 보면 두 지표가 얼마나 연관성이 있는지 확인할 수 있다.
- 새롭게 지표를 만들 때, 사용한 기존 호응도 관련 컬럼은 상관관계가 높게 나올 수 밖에 없기 때문에
→ 새로운 지표와 호응도 관련 상관관계를 확인할 때, 이 점을 유의하고 봐야한다.

engagement
판단 지표

새롭게 만든 지표

피어슨 상관계수

engagement
판단 지표

새롭게 만든 지표

지표 4. Likes Dislikes Ratio (기준: category_name)

* like_dislike_ratio 컬럼은 데이터셋에 추가해주지 않고 따로 시각화로 확인했다.

- * Dislikes 수 대비 Likes의 수가 많은지 확인할 수 있다.
: 해당 지표를 통해 호불호가 많이 갈리는 카테고리인지 추측할 수 있다.
- * Likes, Dislikes는 모두 (off_ - on_) 기준
e.g. Likes : 'off_Likes' - 'on_likes' (= 'likes_growth' 컬럼)

