

데이터 분석 및 시각화

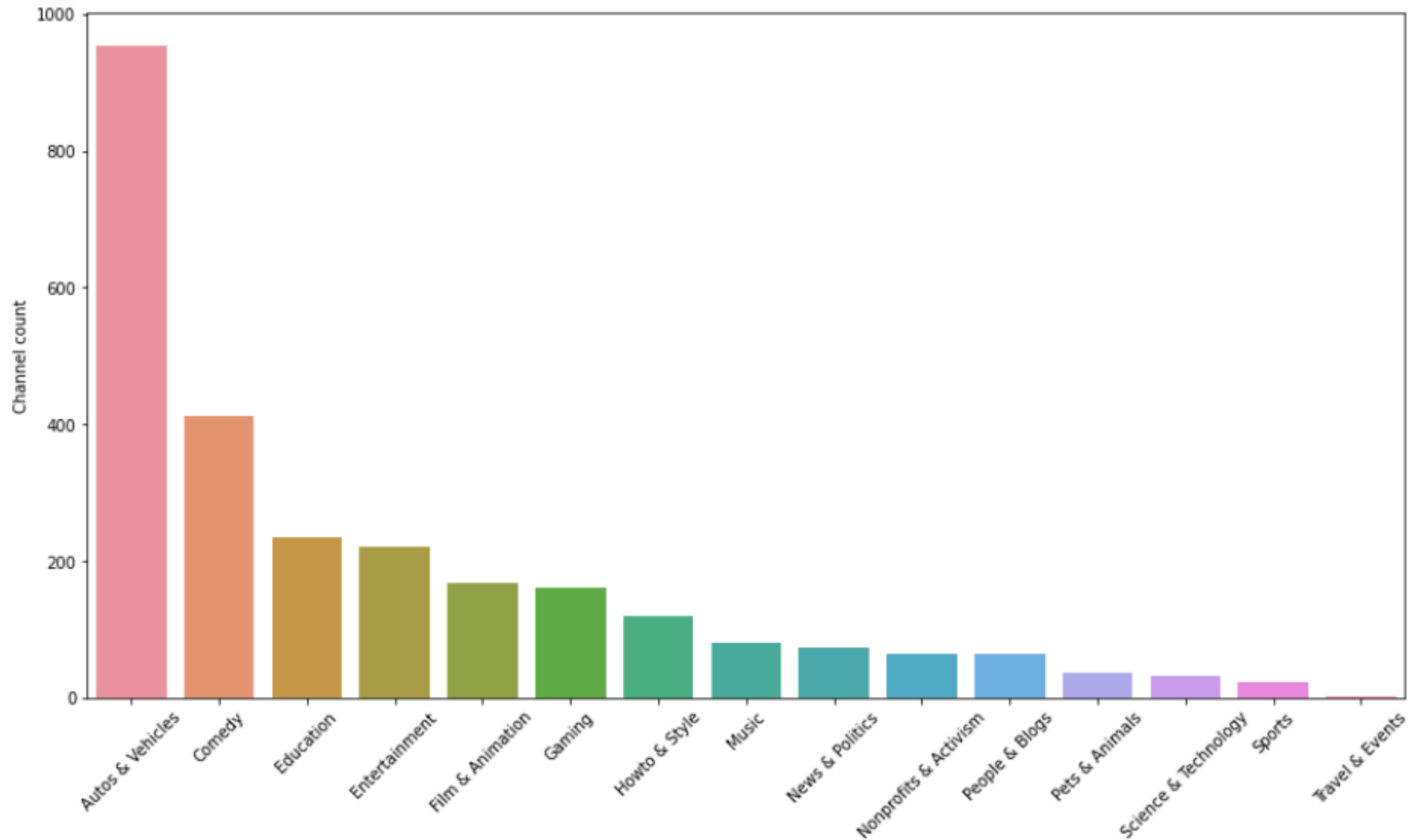
7팀 조하진

Q1. 데이터 타입별 시각화

- 전체기간 카테고리 > 채널 > 비디오 개수
- 월별 카테고리 > 채널 > 비디오 개수
- 월별 TOP10 채널 (분류기준은 비디오 개수)
- 주별 TOP5 채널 (분류기준은 비디오 개수)
- 월별 카테고리별 태그 키워드 순위

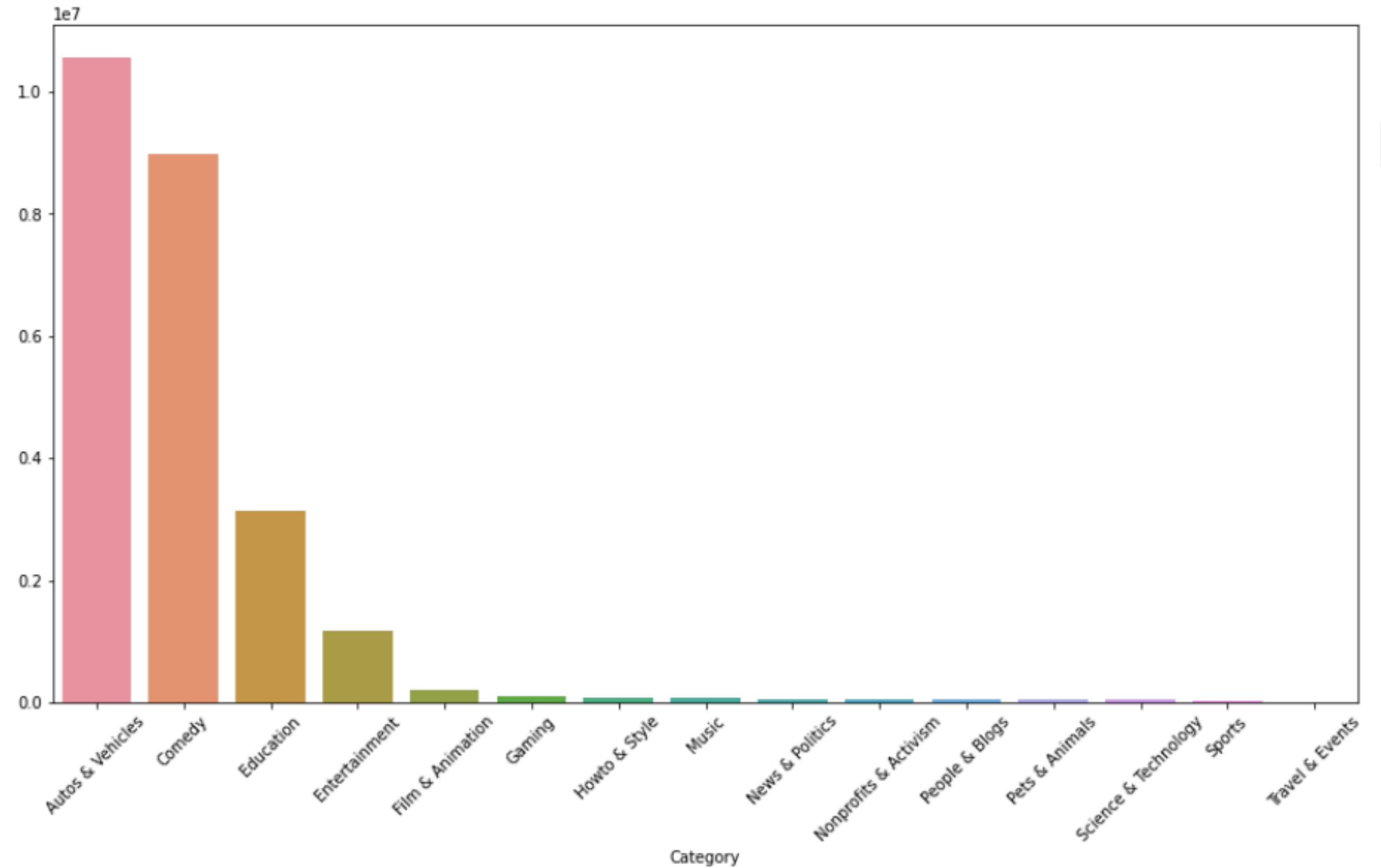
Q1. 데이터 타입별 시각화

전체기간 카테고리별 채널개수



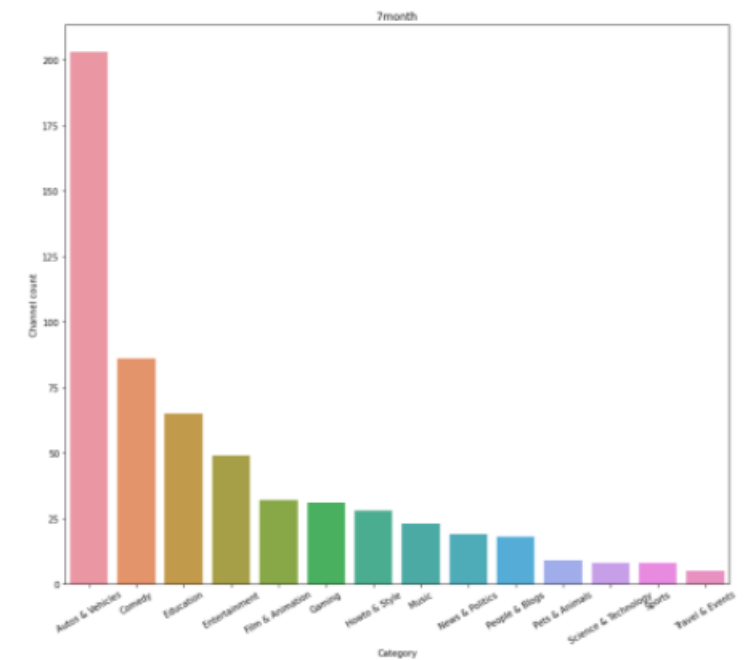
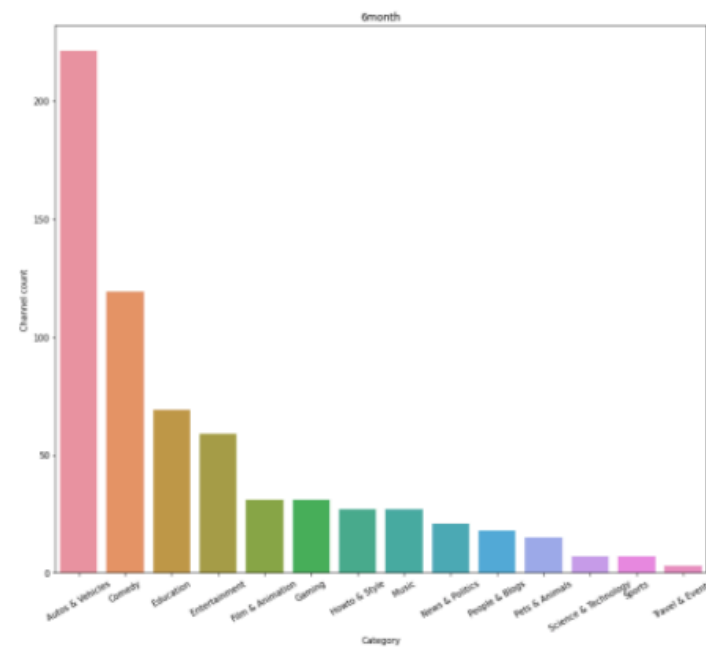
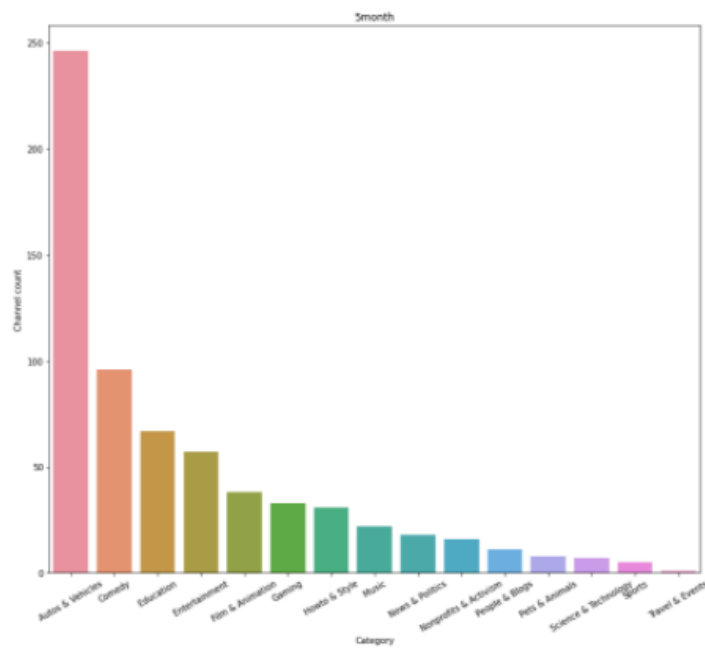
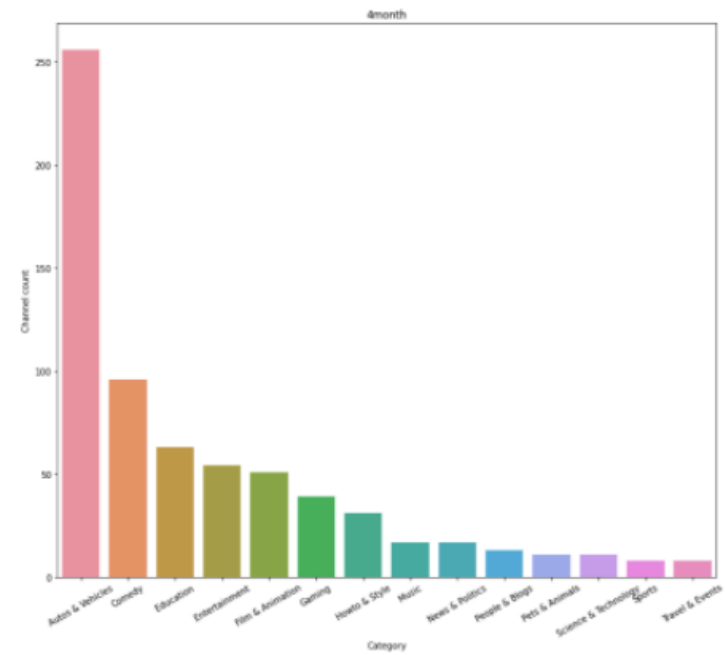
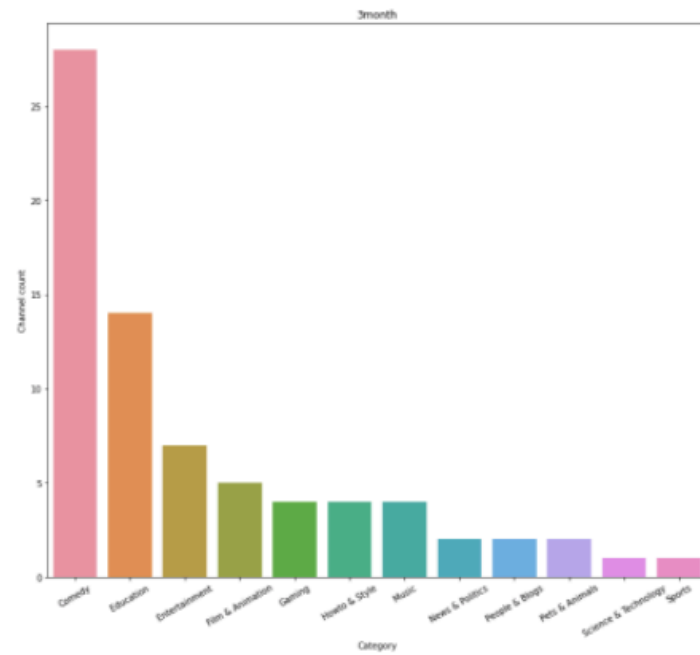
Q1. 데이터 타입별 시각화

전체기간 카테고리별 비디오 개수



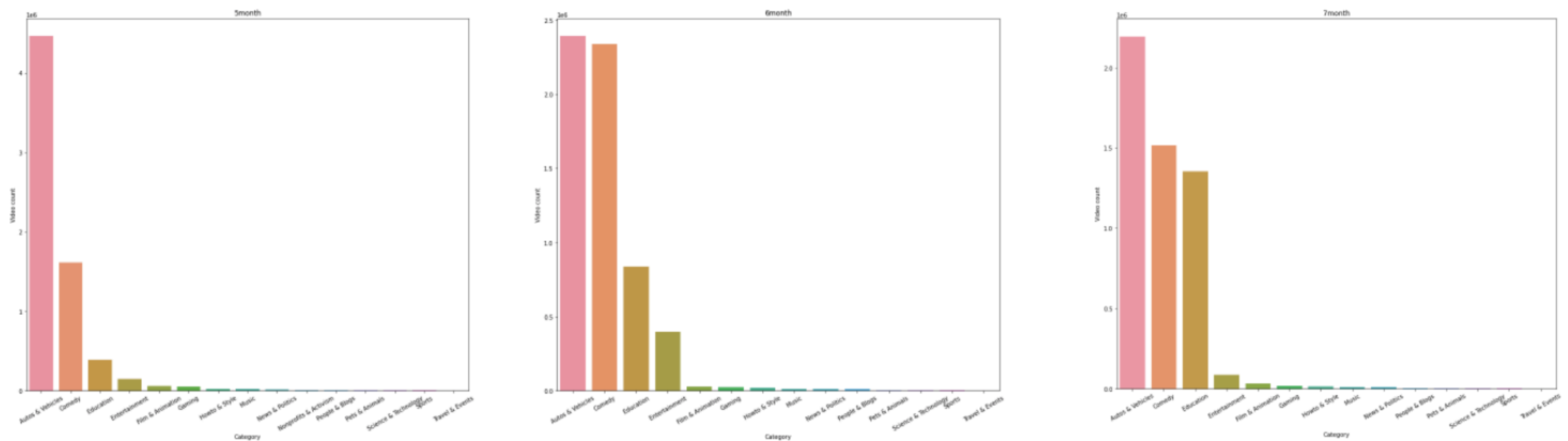
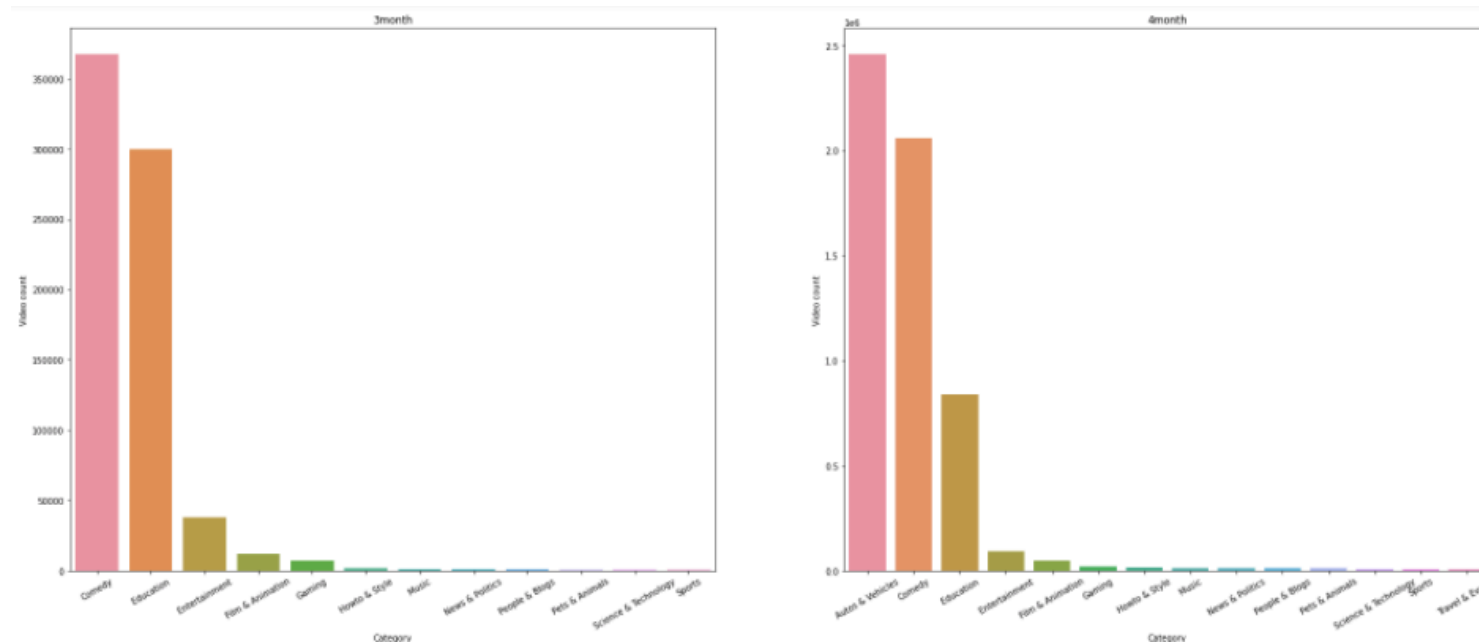
Q1. 데이터 타입별 시각화

월별 카테고리별 채널개수



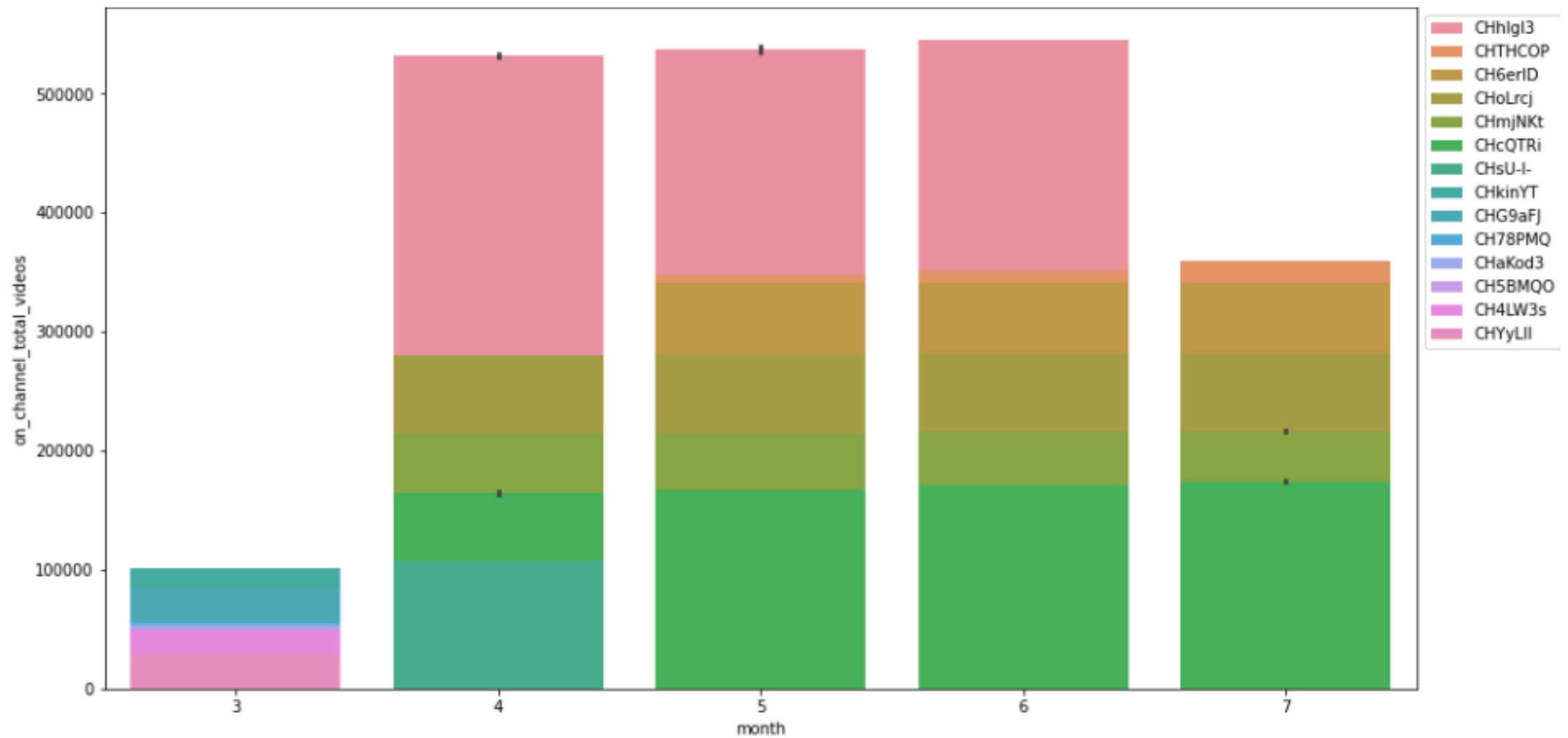
Q1. 데이터 타입별 시각화

월별 카테고리별 비디오개수



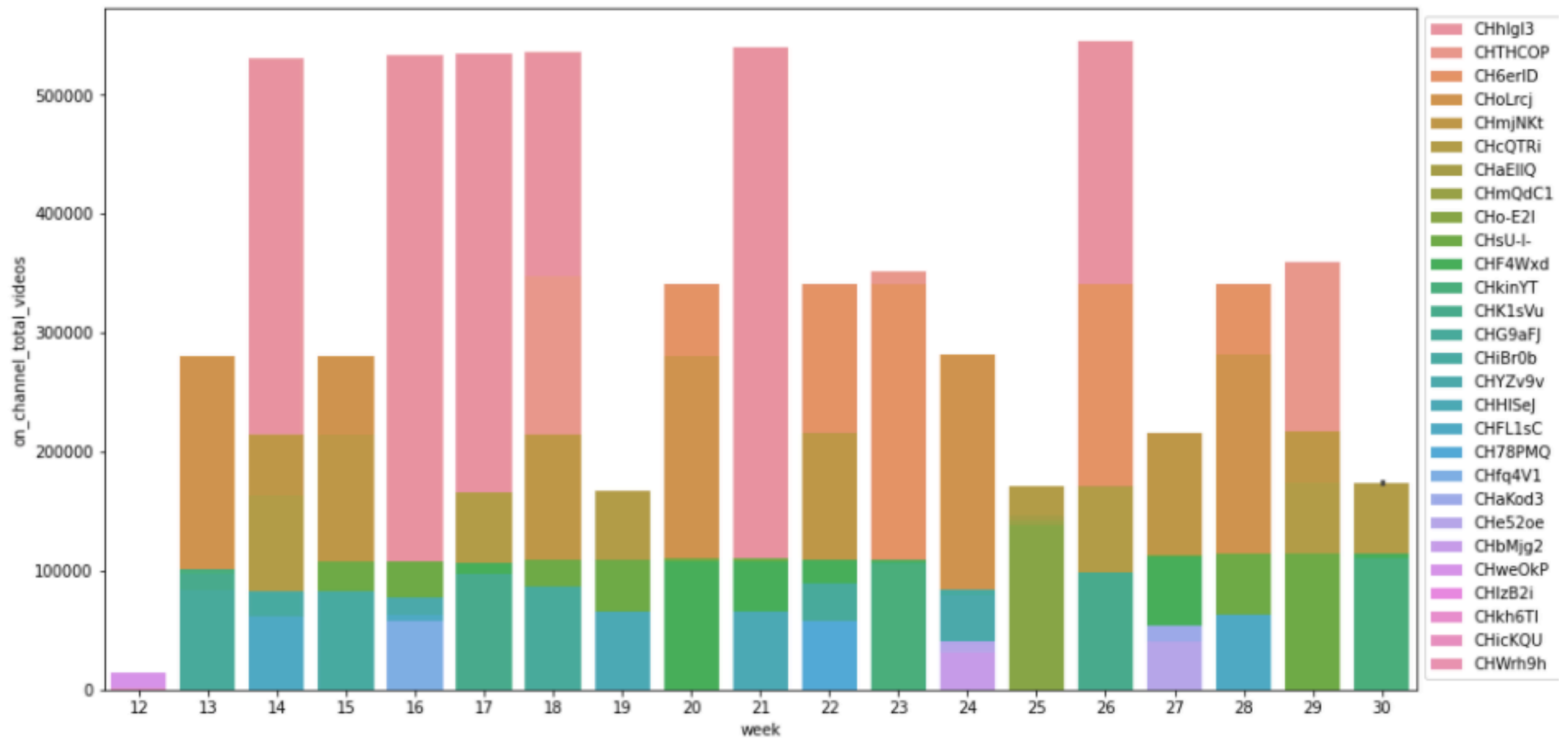
Q1. 데이터 타입별 시각화

월별 TOP10 채널 (분류기준 : 비디오개수)



Q1. 데이터 타입별 시각화

주별 TOP5채널 (분류기준 : 비디오개수)



Q1. 데이터 타입별 시각화

월별 태그 키워드 순위



Q1. 데이터 타입별 시각화

카테고리별 태그 키워드 순위



Q2. 각각의 비디오는 시청자의 호응도를 판단 할 수 있는 객관적인 지표들이 있음

- 비디오를 인기 동영상 기준에 부합하도록 분류할 수 있는 새로운 지표를 개발하고
- 이 지표를 사용하여 engagement와 어떤 상관관계가 있는지 설명하시오.

Q2. 인기동영상 지표

2644 개의 데이터

	on_rank	off_rank	on_views	off_views	on_likes	off_likes	on_dislikes	off_dislikes	on_comments	off_comments	c
count	2644.000	2644.000	2644.000	2644.000	2644.000	2644.000	2644.000	2644.000	2644.000	2644.000	
mean	20.234	35.795	953481.912	1432732.441	54658.126	68453.837	534.280	860.402	9032.085	11112.025	
std	12.833	10.377	3200374.580	6836354.131	323933.712	498321.676	2070.643	4754.238	115532.131	157460.092	
min	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
25%	9.000	28.000	235352.000	306854.000	4208.500	4932.750	86.000	108.000	511.000	597.750	
50%	19.000	37.000	427924.500	571508.000	7851.000	9507.000	159.000	211.000	1157.000	1311.500	
75%	30.000	45.000	826880.000	1119498.750	16883.500	20859.500	315.000	435.000	2720.000	3150.750	
max	50.000	50.000	97276666.000	248876265.000	8097173.000	15759925.000	37349.000	133721.000	4625133.000	6659938.000	

본격적 탐색에 앞서
describe로 훑어보기

```
print(df_copy["on_likes"].value_counts().head(1))
print(df_copy["on_comments"].value_counts().head(1))
print(df_copy["on_channel_subscribers"].value_counts().head(1))
```

```
0    22
Name: on_likes, dtype: int64
0    17
Name: on_comments, dtype: int64
0    64
Name: on_channel_subscribers, dtype: int64
```

좋아요, 댓글, 구독자가
0인 경우도 있다.

```
df_copy[df_copy["on_views"]==0]
```

video_id	channel_id	published_date	category_name	duration	tags	description	on_trending_date
1855	VhAxq	CHdq61m	2021-04-21	Film & Animation	PT40S	NaN	This year's annual Earth Day Doodle highlights... 2021-04-23

1 rows × 27 columns

심지어 조회수 0인
데이터도 존재

Q2. 인기동영상 지표

조회수 TOP10 리스트

전부 music 카테고리이며 BTS, NCT, 에스파, 엑소
아이돌 그룹이 조회수가 압도적으로 많다.

	channel_id	description	published_date	category_name	on_views	on_likes	on_dislikes	on_comments	on_channel_subscribers
1384	CH3IZKs	BTS (방탄소년단) 'Butter' Official MV Credits: Dire...	2021-05-21	Music	97276666	8097173	13899	4625133	55800000
603	CH3IZKs	BTS (방탄소년단) 'Permission to Dance' Official MVC...	2021-07-09	Music	62496315	6409443	17720	2125783	58900000
1138	CHEf_Bc	NCT DREAM's 1st album Hot Sauce is out!Listen ...	2021-05-10	Music	46629333	1780217	16466	474730	27200000
2617	CH3IZKs	BTS (방탄소년단) 'Film out' Official MVSong Credits...	2021-04-01	Music	35829903	4762803	23017	1632895	53100000
784	CHLkAep	Join the 🕺 Permission to Dance 🕺 ChallengeHow to...	2021-07-23	Music	34744351	3471855	37327	47234	54700000
1087	CH3IZKs	BTS (방탄소년단) 'Butter' Official TeaserBTS (방탄 소년단...	2021-05-18	Music	34439082	3750872	24065	1050306	54700000
220	CHEf_Bc	aespa's new single Next Level is out!Listen an...	2021-05-17	Music	33029974	1447202	37349	218154	27300000
135	CHEf_Bc	EXO's Special Album DON'T FIGHT THE FEELING is...	2021-06-07	Music	30069456	3356841	19971	2091078	27800000

Q2. 인기동영상 지표

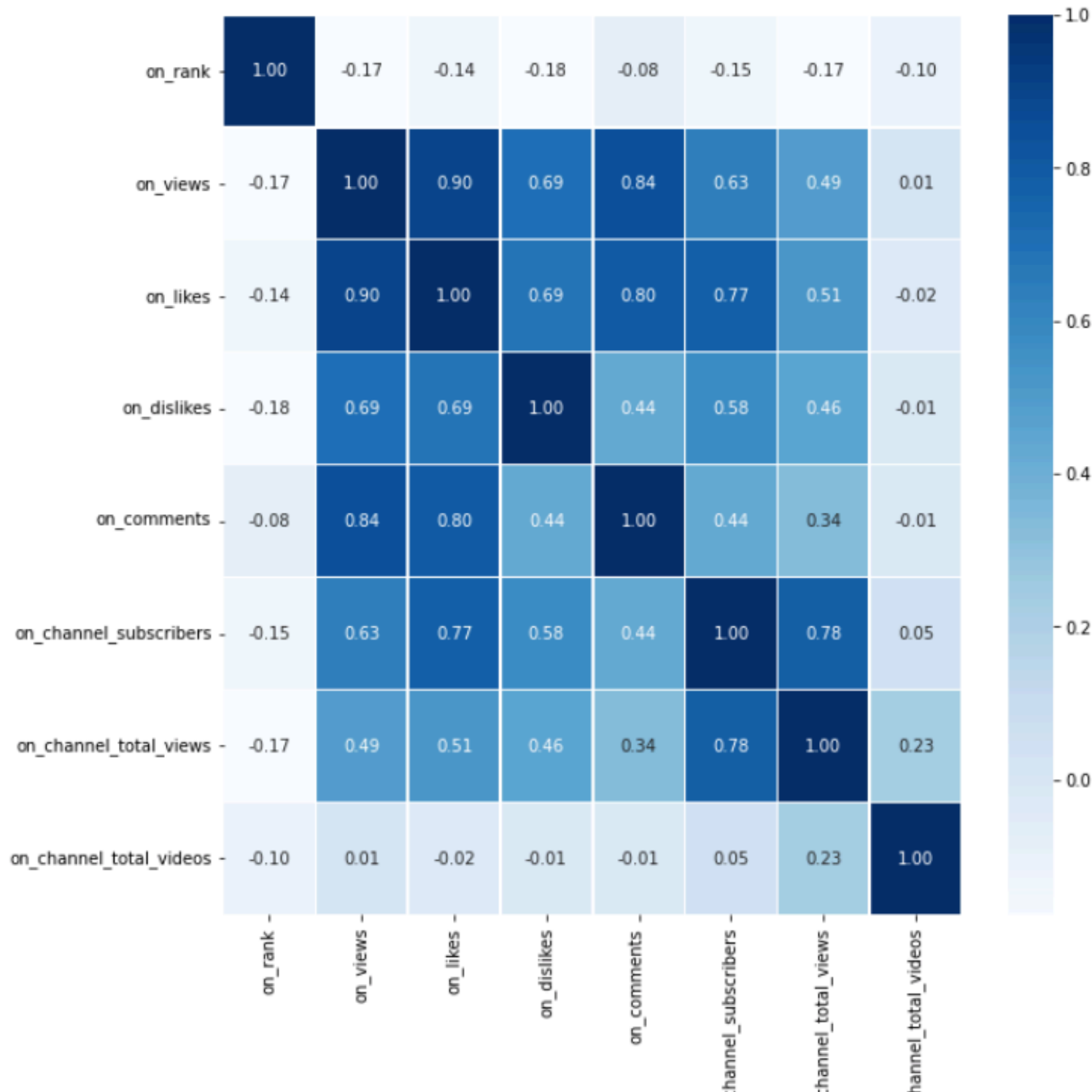
조회수 하위 5 리스트

최소 조회수가 5만 이상은 되어
인기동영상에 오를 수 있는 것으로 보인다.

	channel_id	description	published_date	category_name	on_views	on_likes	on_dislikes	on_comments	on_channel_subscribers
2641	CHO4RG1	애플 봄 스페셜 이벤트를 통해서 공개된 아이패드 프로 5세대, 신형 아이맥, 에어태...	2021-04-20	Science & Technology	59192	729	46	438	73100
207	CHhna8l	여덟번째 브회로그 생일 축하해주신 모든 분들 너무 감사합니다앞으로도 겸손한 마음으로 ...	2021-06-12	People & Blogs	55563	5983	68	1048	139000
2570	CHRQ7Rw	브레이브걸스 인터뷰 기사에도 많은 관심 부탁드립니다. 감사합니다!https://ww...	2021-04-12	Music	53551	2118	9	477	1470
367	CH9gxOp	아프리카 최대시장 탐방기복잡하지만 짬있는고	2021-04-26	People & Blogs	53297	1877	30	345	85700
1855	CHdq61m	This year's annual Earth Day Doodle highlights...	2021-04-21	Film & Animation	0	0	0	0	1890000

Q2. 인기동영상 지표

상관관계 히트맵 시각화



조회수와 상관관계가 가장 높은 컬럼은
순서대로 좋아요, 댓글수, 싫어요, 구독자수이며,
좋아요와 댓글수는 조회수와 연관이 크다고 나타난다.

채널의 총 조회수는 0.49정도의 수치를 보여줬다.

채널의 총 비디오 수는 조회수와 큰 연관이 없었다.

Q2. 인기동영상 지표

- 인기동영상에 랭크되었을 때 평균 조회수는 953,481회 이고, 평균 구독자 수는 약 170만명 이다.
- 좋아요, 싫어요, 댓글 같은 경우 각 채널에서 표시 되지 않게 설정 한 경우 0으로 수집되었다.
- 조회수의 최소값이 0인 데이터가 있어서 찾아보니 조회수, 좋아요, 댓글 등 다 0이지만 구독자 수가 백만이 넘고 채널 토탈뷰는 3억이 넘었다. 위에 찾아본 인기동영상 기준처럼 외부요인이 영향을 준건지, 채널안에서 비교가 영향을 준건지 조금 애매한 데이터이다.
- 조회수 상위 채널은 대부분 아이돌 그룹이다.
- 좋아요, 댓글수, 싫어요, 구독자수 순으로 조회수와 연관이 깊다.
- 그럼 지금까지 분석한 내용을 바탕으로 몇 가지 컬럼을 더 추가해 분석해보자.

Q2. 인기동영상 지표

- likes/views, dislikes/views, comment/views : 조회수 대비 좋아요/싫어요/댓글수 비율
- on_period : 동영상 올린 후 인기동영상에 올라가게 된 기간
- off_period : 인기동영상 게재 후 내려간 기간
- views_increase_day : 게재된 기간 하루 동안 조회수 증가폭
- tags_len : tag 개수

총 일곱개의 새로운 지표를 추가 했다.

Q2. 인기동영상 지표

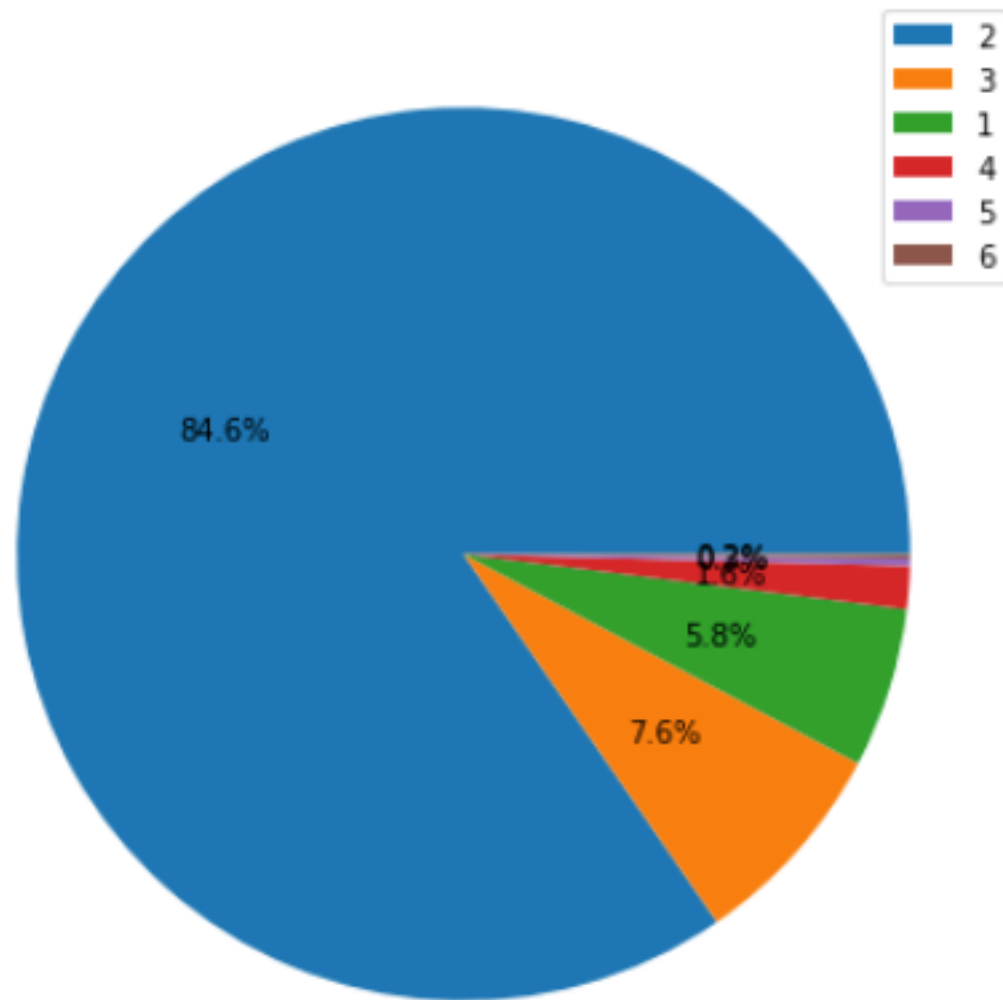
새로 추가한 컬럼 describe

	likes/views	dislikes/views	comment/views	on_period	off_period	views_increase_day	tags_len
count	2643.000	2643.000	2643.000	2644.000	2644.000	2644.000	2274.000
mean	0.032	0.001	0.004	2.064	1.330	123965.719	21.118
std	0.040	0.001	0.006	0.490	1.170	631111.815	18.998
min	0.000	0.000	0.000	1.000	0.000	0.000	1.000
25%	0.012	0.000	0.002	2.000	0.000	0.000	8.000
50%	0.019	0.000	0.003	2.000	1.000	40919.000	15.000
75%	0.032	0.001	0.005	2.000	2.000	96402.833	28.000
max	0.328	0.021	0.097	6.000	7.000	18949949.875	132.000

- 조회수 대비 좋아요는 평균 3.2프로 최대 32프로.
- 조회수 대비 싫어요는 대부분 0프로 최대 2프로.
- 조회수 대비 댓글수는 대부분 낮은편이지만 최대 9.7프로 인걸로 봐서 특이값이 있을 것으로 예상 된다.
- 보통 게시된지 이틀 이내에 인기동영상에 올랐고, 최고 6일까지도 존재하지만 대부분 이틀 이내에 올라가는 것으로 확인된다.
- 보통 인기동영상에 하루 이내로 랭크 되어있고, 최장 7일 까지도 있을 수 있다.
- 인기동영상에 랭크되면 약 12만 정도 조회수가 올라갔으며, 하위권에 존재하는 0은 랭크되자마자 내려온 경우로 보인다.

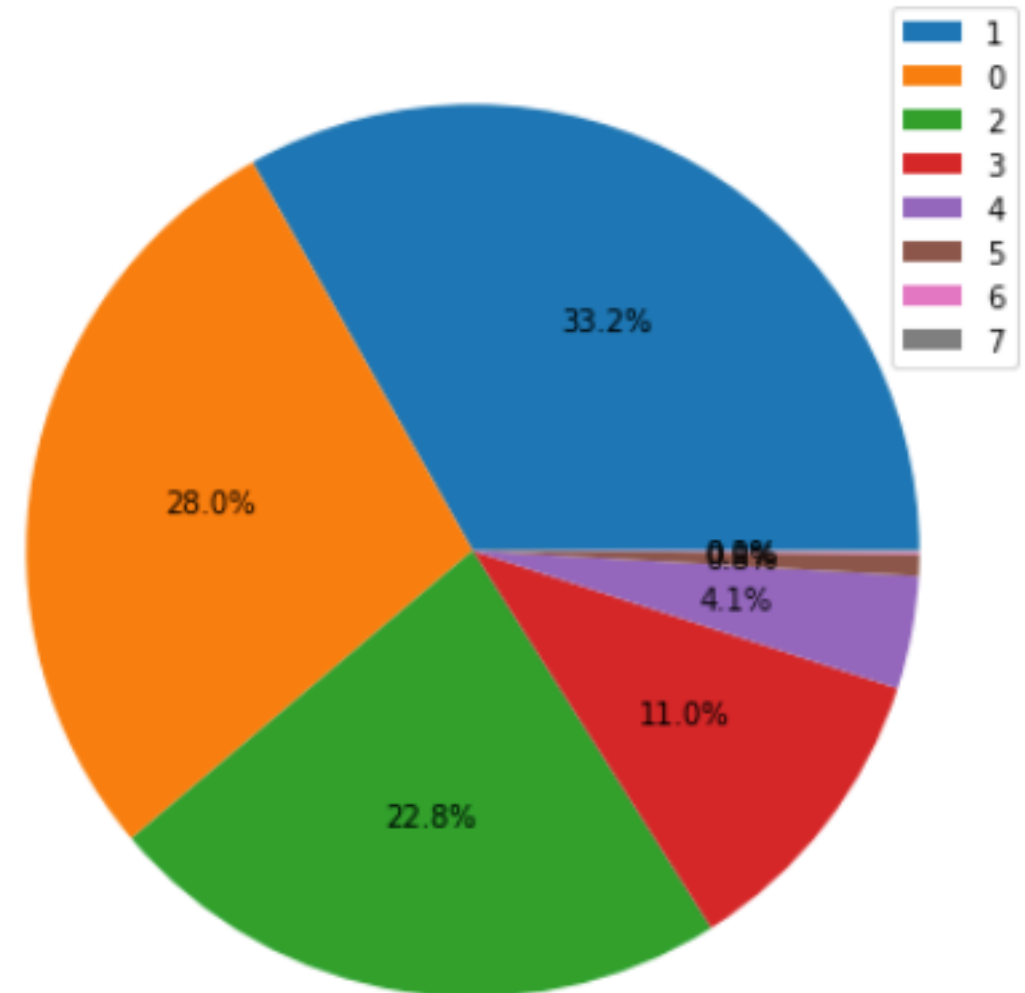
Q2. 인기동영상 지표

동영상 올린 후 인기동영상에
올라가는 기간 비율 차트



이틀째에 랭크된 경우가 84.6%로 거의 대부분이고
3일이 지나면 인기동영상에 오를 확률이 희박해진다.

인기동영상에 게재된 비율 차트



하루도 못가는 경우가 28%로 두번째로 많고
보통 이틀 이내로 내려간다.

Q2. 인기동영상 지표

5일 이상 걸려 인기동영상에 올라간 영상 리스트

category_name	on_likes	on_comments	on_channel_subscribers	on_views
Entertainment	43037	1077	203000	2873289
Entertainment	241640	13572	63100	2706027
Sports	41903	1482	84600	2542033
Education	26412	1596	175000	2199900
People & Blogs	29952	1941	41000	1729099
Entertainment	26166	1954	3420000	873061
Music	12908	1270	9380000	585612
Entertainment	5631	1929	45200	299433
Entertainment	3470	232	121000	269834
People & Blogs	2809	70	2980	248716
Music	4340	1432	395000	110504

최소 10만에서 최대 287만 정도 조회수가 나와야 시간이 오래 걸려도 인기동영상에 오를 수 있었다.
카테고리는 상위 조회수의 대부분 이었던 music이 아닌 entertainment의 비율이 높다.
조회수가 비교적 낮은 10만 짜리 영상의 경우에는 좋아요, 댓글, 구독자 수가 확실히 많았다.

Q2. 인기동영상 지표

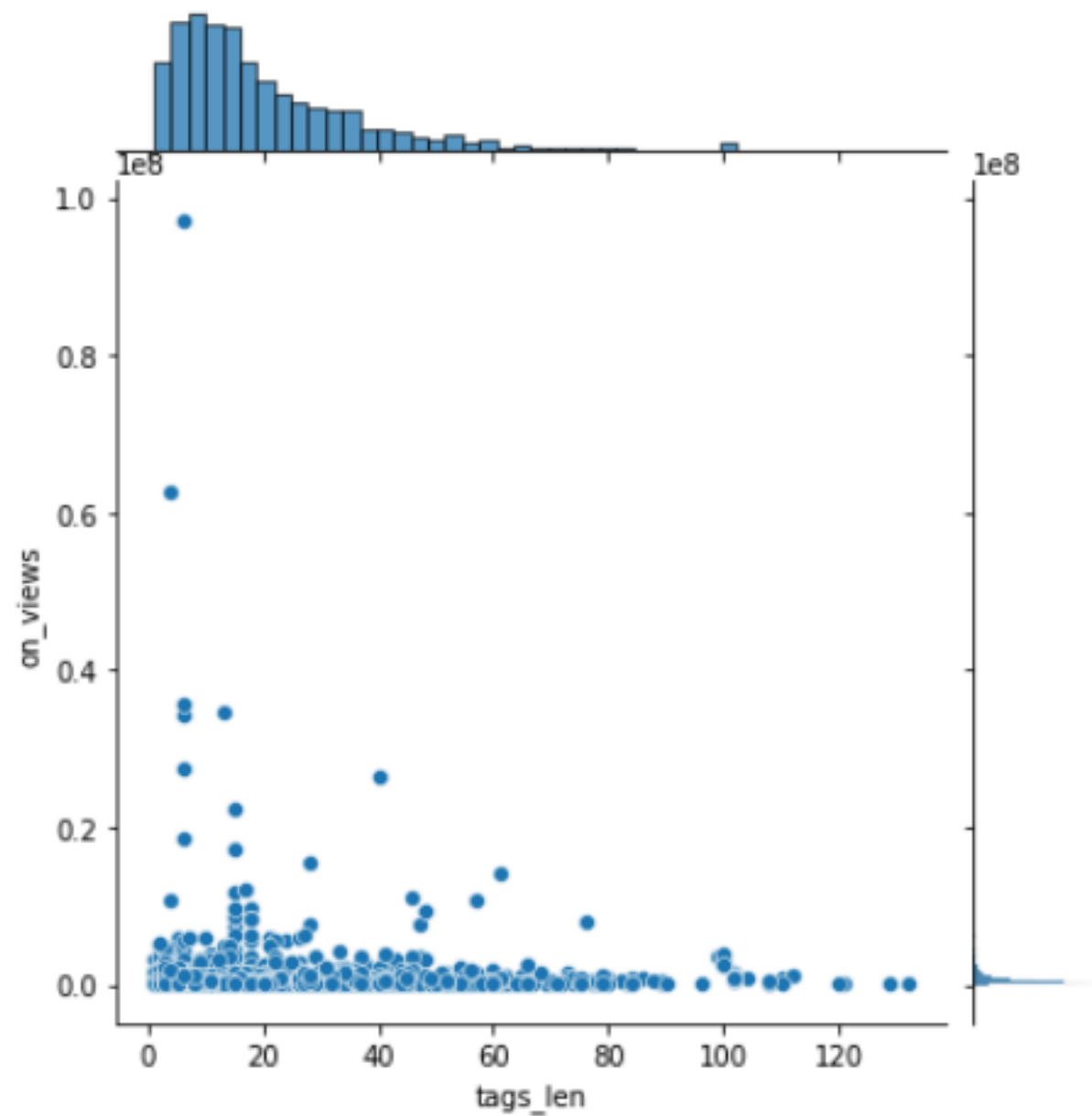
5일 이상 인기동영상에 올라가있는 영상 조회수 TOP10

on_rank	off_rank	category_name	on_likes	on_comments	on_channel_subscribers	on_views
1	35	Music	8097173	4625133	55800000	97276666
1	28	Music	6409443	2125783	58900000	62496315
1	29	Music	2414549	121023	50500000	9661320
1	34	Entertainment	825421	47712	22200000	8553414
1	47	News & Politics	34896	14422	359000	5134545
1	29	People & Blogs	67117	10171	2110000	4600715
1	30	Entertainment	155973	17532	1400000	3619163
1	50	Music	269601	22681	2850000	3540343
1	50	People & Blogs	91976	28741	817000	3414658
1	39	Entertainment	224962	16599	2280000	3284254

5일 이상인 비교적 긴 시간 인기동영상에 올라가있는 리스트로 총 24개중 조회수 순으로 10개를 뽑았다.
대부분 당시 화제성이 엄청 뛰어난 영상들 이었다.(BTS, 아이유, 당시 핫했던 뉴스나 유튜버들)
시작 순위가 전부 1위로 높은 순위로 올라가면 오래 머무를 수 있다는 것을 알 수 있다.

Q2. 인기동영상 지표

태그개수와 조회수의 산점도



큰 상관관계는 없는 것으로 보인다.
태그는 대부분 20개 이하로 달기 때문에
그 안에서 조회수가 높은 몇몇 영상이 발견됐다.

Q2. 인기동영상 지표

조회수 대비 좋아요 비율 상위 3

description	category_name	on_rank	off_rank	on_views	on_likes	on_dislikes	on_comments	on_period	off_period	views_increase_day
안녕하세요 백현 입니다! 사랑하는 구독자 여러분 5.6일 훈련소에 들어가게 되었습 니다...	People & Blogs	18	18	518808	170108	170	23363	2	0	0.000
D.O.'s 1st Mini Album 공감 will be released on J...	Music	20	34	771408	245800	293	31448	2	1	65552.500
D.O.'s 1st Mini Album 공감 will be released on i	Music	21	21	605001	186484	151	14719	2	0	0.000

그룹 엑소 멤버들이 상위권을 차지하고 있다.

계속해서 보여지는 현상으로, 아이돌 팬의 화력 때문에 조회수도 좋아요도 상당히 높게 나온다.

Q2. 인기동영상 지표

조회수 대비 싫어요 비율 상위 3

description	category_name	on_rank	off_rank	on_views	on_likes	on_dislikes	on_comments	dislikes/views	on_period	off_period
안녕하세요 오메킴입니다제가 이번 영상을 올리는 이유는 5월 9일 경 있었던 큰 잘못 ...	Gaming	6	49	384865	1024	8160	5175	0.021	2	3
전기야 나와서보자 ^^리뷰 영상 출처 https://youtu.be/53nbU26EgB8	People & Blogs	3	49	1287367	3415	26335	13992	0.020	3	4
BJ봉준입니다 저는 2021년 5월 10일 새벽 1시경 동료BJ들과 방송을 진행하던...	Comedy	2	28	1137805	3333	22830	19228	0.020	2	3

영상들을 직접 찾아보니 싫어요 상위권 답게 대부분 어그로성(?) 영상이 많았다.
하지만 조회수대비 좋아요 영상들 보다 인기동영상에 더 오래 머물러있음을 확인 할 수 있었는데,
싫어요 비율도 어쨌든 화제의 척도가 되는 것을 다시 한번 알 수 있었다.

Q2. 인기동영상 지표

조회수 대비 댓글 비율 상위 3

description	category_name	on_rank	off_rank	on_views	on_likes	on_dislikes	on_comments	comment/views	on_period	off_period
무 정체성 무 컨 텐츠 채널인데도 구독해주셔서 감 사합니다선물 관 련하여 추후 회사 컨펌을...	Music	7	19	98823	7401	11	9608	0.097	1	1
EXO's Special Album DON'T FIGHT THE FEELING is...	Music	11	34	30069456	3356841	19971	2091078	0.070	2	2
죄송합니다 앞으 로 안 까볼겠습니 다그래도 즐거웠 습니다영상 편집 : 바빠서 근 한달 만에...	Music	41	41	71496	5672	27	4115	0.058	2	0

댓글은 아까 describe에서 특이값이 좀 크게 있는것으로 봤었는데, 찾아보니 댓글 추첨을 통해 선물을 주는 이벤트가 있는경우 유독 댓글이 더 많이 달린 것을 알 수 있었다.

Q2. 인기동영상 지표

하루당 조회수 증가 상위 10

description	category_name	on_rank	off_rank	on_views	on_likes	on_dislikes	on_comments	on_period	off_period	views_increase_day
BTS (방탄소년단) 'Butter' Official MV Credits: Dire...	Music	1	35	97276666	8097173	13899	4625133	1	7	18949949.875
BTS (방탄소년단) 'Permission to Dance' Official MVC...	Music	1	28	62496315	6409443	17720	2125783	1	6	12937656.714
ITZY 마.피.아. In the morning M/Vhttps://orcd.co/...	Music	2	19	11185695	1476092	17512	166847	1	3	11335215.500
TWICE Alcohol-Free M/V TWICE The 10th Mini Alb...	Music	1	43	26416805	2190938	32200	649113	2	4	9377917.000
aespa's new single Next Level is out!Listen an...	Music	1	28	33029974	1447202	37349	218154	2	3	6536013.750
NCT DREAM's 1st Album Repackage Hello Future i...	Music	5	24	13319546	1262135	5226	248198	2	2	5966474.000
BTS (방탄소년단) 'Butter' Official TeaserBTS (방탄소년단...	Music	3	15	34439082	3750872	24065	1050306	2	3	5731408.500
BTS (방탄소년단) 'Film out' Official MVSong Credits...	Music	3	33	35829903	4762803	23017	1632895	2	3	5343625.500
NCT DREAM's 1st album Hot Sauce is out!Listen ...	Music	4	13	46629333	1780217	16466	474730	2	2	5196969.333
Join the 🦋 Permission to Dance 🦋 ChallengeHow to...	Music	3	15	34744351	3471855	37327	47234	2	3	5022095.750

조회수 상위권답게 대부분 아이돌 그룹이며 BTS당시 신곡 Butter는 게재 기간도 7일로 최고치 였다.

Q2. 인기동영상 지표

하루당 조회수 증가 하위

```
print(len(df_copy[df_copy["views_increase_day"]==0]))  
print(len(df_copy[(df_copy["views_increase_day"]==0) & (df_copy["off_period"]==0)]))
```

740

739

조회수 증가량이 0인 영상이 많아서 찾아보니
740개의 영상중 739개가 당일 올라갔다 바로 내려온 영상으로 나타난다.

	on_rank	off_rank	on_views	on_likes	on_dislikes	on_comments	views_increase_day
5	14	14	10638842	442766	1993	38182	0.000
267	41	41	9706443	52432	263	6659	0.000
1534	31	31	9205614	321317	3200	68849	0.000
2509	36	36	8036827	78727	878	8521	0.000
1961	24	24	6474905	206201	2457	13313	0.000
1561	34	34	6256882	1216334	4389	83966	0.000
1254	12	12	6195556	1293467	1953	37584	0.000
2329	25	25	6132077	622720	3007	55162	0.000
1069	29	29	5947821	116728	1901	17760	0.000
55	26	26	5521217	235946	5219	0	0.000

증가량 0인 영상을 좀더 살펴봤다.

조회수나 좋아요가 부족하진 않은데 현재 나온 지표로는
어떤점 때문에 금방 랭크 아웃됐는지 파악하기 어려운것 같다.

Q2. 인기동영상 지표

인기동영상 개수별 채널수

1	407
2	159
3	128
4	72
5	48
6	37
7	33
8	21
10	10
9	9
11	6
14	4
13	3
17	1
12	1
15	1

Name: channel_id, dtype: int64

대부분의 채널이 인기동영상을 한개 보유하고 있다.

인기동영상이 1개인 채널 카테고리 현황

Entertainment	103
People & Blogs	74
Music	56
Gaming	29
Sports	27
News & Politics	21
Comedy	18
Howto & Style	15
Film & Animation	14
Education	14
Science & Technology	11
Pets & Animals	11
Autos & Vehicles	10
Travel & Events	3
Nonprofits & Activism	1

Name: category_name, dtype: int64

Entertainment, People&blog, music 순으로
인기동영상을 보유 하고 있다.

Q2. 인기동영상 지표

인기동영상 개수가 2~3개인 채널 카테고리 현황

```
Entertainment      252
People & Blogs      116
Music               61
Name: category_name, dtype: int64
```

2~3개의 동영상 보유 경우에는 한개 보유 일때와 1~3순위 동일

인기동영상 개수가 5개 이상인 채널 카테고리 현황

```
Entertainment      492
Sports             154
People & Blogs      138
Name: category_name, dtype: int64
```

- 대부분의 채널이 3개이내의 인기동영상 보유하고 있었다.
- Entertainment, People & Blogs 는 꾸준히 많은 보유량을 보여주었고, Sports는 비교적 다수의 인기동영상을 보유한 카테고리임을 알 수 있었다.
- 조회수가 높은 카테고리는 거의 대부분 music임에 비해 인기동영상 보유 비중이 높은것은 Entertainment 이다.

Q2. 인기동영상 지표

결론

일단 인기동영상이 되는데 가장 큰 영향을 미치는 것은 조회수이며 평균 95만 정도, 최소 5만 이상의 조회수가 나와야 인기동영상에 랭크 될 수 있다.

조회수가 가장 상관관계가 높은 것은 좋아요 수치이며, 좋아요 > 댓글수 > 싫어요 > 구독자 순으로 조회수와 연관이 높고, 채널의 총 비디오수는 큰 연관이 없다.

3일 이내에 인기동영상이 되지 않으면 가능성이 점점 떨어진다.

랭크된 후 보통 이틀이내로 내려가며 당일 바로 내려가는 경우도 28%정도 있다.

조회수 대비 좋아요 비율은 아이돌 그룹이 대부분이다.

조회수 대비 싫어요 비율은 비호감적인 영상이 많았다.

조회수 대비 댓글 비율의 경우에는 좋아요의 비율보단 낮으나, 댓글추첨을 통한 경품추첨이 걸려있을 경우 확실하게 높은 수치를 보여줬다.

전반적으로 Entertainment, people & Blogs, music 카테고리 순으로 인기동영상 비중이 높았다. 그러나 차이가 많이 나는 높은 조회수 부분에서는 단연 아이돌 그룹이 앞섰다.

높은 Engagement를 얻기 위해선...

- 최소 5만 이상의 조회수를 얻을 만한 콘텐츠 제작 아이디어 필요.
- 빠른 시일 내에 인기동영상에 오르지 못하면 기회가 없어지기 때문에 목적에 따라 경품이나 노이즈 마케팅으로 랭크 시킬 기회가 있을 것으로 예상됨.
- 태그 개수는 조회수와 크게 상관이 없는 것으로 보이나, 대부분 20개 내외로 사용함.
- Entertainment, people & Blogs 카테고리가 유리함. music은 아이돌 위주여서 제외.