

# 유튜브 광고 영상 분석

정민서 (소프트웨어융합전공) 박유희 (통계학과) 김지혜 (컴퓨터과학전공)

December 13, 2020

## 1 Introduction

유튜브는 전 세계적으로 가장 주요한 미디어 플랫폼이다. 최근 유튜브에서 시청자들에게 유료 광고임을 고지하지 않고 제품을 추천하는 등 이른바 '뒷광고' 논란이 일어났다. 이와 관련해 시청자들이 광고 영상을 어떻게 느끼길래 유튜버들이 광고를 표시하고 싶어 하지 않는지 궁금증이 생겼고, 광고 여부가 시청자에게 미치는 영향을 통계적으로 분석해보고자 했다.

본 프로젝트에서는 유튜브 사이트 크롤링을 통해 데이터를 추출했다. 분야 범주로는 뷰티, 먹방, 게임을 선정했고 각 분야에서는 15명의 유튜버를 미리 선정했다. 세 가지 범주를 나눈 이유는 크게 유튜브 콘텐츠 카테고리가 위의 세 개로 나뉘어있고, 각 분야별로 광고 비율과 영상 제작 방식, 그에 따른 시청자들의 반응이 다를 거라고 예상했기 때문이다. 각 분야별로는 15명의 유튜버를 선정하여 총 45명의 유튜버들의 동영상 정보를 크롤링하고자 했다. 2019년 1월 1일 이후에 유튜브에서 광고 고지에 대한 원칙을 확고히 했고, 유튜버들이 그 이후에 광고 고지에 대한 인식이 생성되었다는 점을 고려하여 2019년 1월 1일 이후의 영상들을 추출했다. 광고 및 협찬 여부, 좋아요 수, 싫어요 수, 조회수, 댓글 수, 댓글 등의 데이터를 추출해 데이터셋을 완성해 분석을 진행했다.

광고 여부에 따른 시청 데이터의 변화를 분석하기 위해 반응변수를 시청 데이터, 설명변수를 광고 여부로 설정했고 유튜버 별 반응변수 값이 이질적일 수 있다고 판단하여 설명변수에 유튜버 이름을 추가하여 다중선행회귀 분석을 진행했다. 광고 여부가 시청자에게 미치는 영향은 좋아요 수, 싫어요 수, 조회수, 댓글 수, 댓글 내용으로 추정 가능할 것이다. 이러한 착안점에 기인하여 우리는 광고 및 비광고 영상과 비교하여 시청자의 반응에 차이가 있는지, 차이가 있다면 어느 부분에서 크게 두각을 나타내는지 예측해볼 수 있다. 나아가 유튜브 콘텐츠별로 어떤 차이가 있는지 비교해볼 수 있다. 유튜브 분석을 다룬 타 프로젝트나 논문은 단순하게 워드 클라우드를 통한 댓글 분석에만 그치고 있다. 이와 다르게 해당 프로젝트는 댓글뿐만 아니라 싫어요 수, 좋아요 수, 조회수, 댓글 수의 영향도 고려했다는 점, 뷰티/게임/먹방 콘텐츠별로 차이를 알아보았다는 점, 광고 영상의 댓글, 비광고 영상의 댓글을 나누어 분석하여 광고 여부에 따른 댓글 반응을 알아보았다는 점, word2vec을 이용해 광고 영상의 시청자 감정 반응을 알아보았다는 점에서 차별점이 있다. 추가적으로 분야별, 유튜버별 광고 영상의 비율을 시각화를 통해 보여줌으로써 시청자에게는 합리적인 영상 소비를 돕고 영상 제작자에게는 분야별 적절한 광고 비율을 예측할 수 있도록 돕는다. 광고 여부에 따른 시청 데이터 분석으로 광고와 시청 데이터와의 연관성을 파악하고 정도를 수치로 나타낸다.

## 2 Background

대다수 유튜버가 영상의 더보기란을 통해 유료 광고 및 협찬 여부 등을 포함한 영상 설명 정보를 노출한다. 시청자들은 유튜버가 올린 영상에 대해 좋아요/싫어요를 클릭하거나 댓글을 남김으로써 영상을 평가하고 소통한다. 영상의 좋아요 수와 댓글 수가 많으면 유튜브 내에서는 좋은 품질의 영상이라고 자체적으로 판단을 하여 상위 검색에 노출시킨다고

한다. 싫어요 배너의 역할은 지속적으로 싫어요를 눌렀을 때, 시청자에게 부적절한 영상이라고 자체적으로 판단이 되어 해당 시청자에게 적게 노출된다고 알려져 있다.

### 3 Data

유튜브 사이트의 html문서 스크래핑을 통해 미리 정해두었던 데이터 추출을 진행했고 분석 목적에 따라 총 3가지의 데이터셋을 완성했다. 데이터 수집을 완료한 후 영상 업로드 날짜 고려, 결측치 제거, 결측치 대체 등의 데이터 전처리 과정을 진행했다.

#### 1. 유튜버 분야별로 15명씩 선정

유튜버 이름 name 변수

##### 데이터 수집 과정

분석 목표는 광고에 대한 시청자들의 반응이기 때문에 "뒷광고" 논란이 터지지 않고, 지금까지 광고 표시를 꾸준히 해 왔던 유튜버들을 표본으로 삼아야 한다. 우선 각 카테고리별로 유튜버들을 랜덤으로 50명 추출했다. 유튜버를 추출하는 과정은 전 세계 유튜버 랭킹을 알려주는 "블링" 사이트를 참고했다. 추출한 50명의 유튜버 중에서 1. 뒷광고 논란이 없는 유튜버들을 추출했고 2. [10만 이상 50만 이하], [50만 이상 100만 이하], [100만 이상]인 유튜버들을 적당한 비율로 정하여 최종 15명을 선정했다.

#### 2. 각 유튜버별 동영상 url 수집하기

유튜버 이름 name 변수, 영상의 url 변수

##### 데이터 수집 과정

선정한 유튜버들의 채널에 접속하여 모든 업로드한 동영상을 볼 수 있는 [동영상] 페이지에 접속한다. 동영상들은 최신 날짜 순으로 정렬되어 있다. selenium을 이용하여 가장 오래전에 업로드한 동영상이 로딩되도록 웹페이지의 끝까지 스크롤을 내린 후 BeautifulSoup을 이용하여 올린 모든 동영상들의 url을 리스트로 저장했다.

#### 3. 동영상 정보 수집하기

id	cate	name	sub	title	like	hate	view	ad	spn	url
key	분야	유튜버 이름	구독자 수	제목	좋아요 수	싫어요 수	조회수	광고 여부	협찬 여부	영상의 url

##### 데이터 수집 과정

2에서 url리스트를 얻었다면 반복문을 통해 해당 url에 순차적으로 접속해 동영상 정보를 얻어온다. 페이지 html 문서에 저장되어 있는 더보기란 내용은 string 형태로 저장했다. 이때, 더보기란 내용에서 유튜버별 광고, 협찬 설명 작성 방식에 따라 if 문을 통해 ad변수에 광고면 1 광고가 아니면 0, spn협찬 변수에 협찬이면 1, 협찬이 아니면 0을 저장했다. 좋아요 수, 조회 수, 싫어요 수, 구독자수는 int로 저장했다. 모든 원소의 저장이 완료되면 반복문을 통해 배열의 원소들을 데이터셋에 저장했다.

#### 4. 영상 별 댓글 수, 댓글 내용 수집하기

id	name	title	url	comment	comm_count
key	유튜버 이름	제목	영상의 url	댓글 내용	댓글 수

## 데이터 수집 과정

selenium에서 x-path를 통하여 댓글 수를 크롤링했다. 댓글은 동영상 정보의 하단에 있으므로 스크롤을 한번 내림으로써 댓글 수를 가지고 올 수 있었다. 간혹 댓글 창이 막혀있는 경우가 있었는데, 이러한 경우에는 정상적인 값을 얻을 수 없으므로 "중지"가 들어오도록 했다. 댓글 내용과 같은 경우에는 모든 댓글을 크롤링하기에는 시간, 하드웨어적으로 한계가 있기 때문에 스크롤을 20번 밑으로 내려서 얻어지는 댓글 내용들만 가지고 오기로 했다. 스크롤 내린 댓글들은 하나의 텍스트로 추가하여 데이터셋의 각 동영상 정보에 "comment" 변수로 저장했다.

동영상 정보와 댓글 정보는 이후 진행할 분석의 목적과 방법이 서로 다르기 때문에 한 데이터셋으로 함께 저장할 필요성이 없었고 댓글 정보가 상당한 양의 데이터를 담고 있어 원활한 분석을 진행하기 위해 두 가지의 데이터셋으로 나누어 생성했다 (영상 정보 데이터셋, 댓글 정보 데이터셋. 이후 두 데이터셋의 원활한 병합을 위해 key변수로 url 변수를 지정해주었다.

## 완성 데이터셋

id	cate	name	sub	title	like	hate	view	ad	spon	url
4612	beauty	회사원	1180000	지워지지 않게 립스틱 바르는 법/How to keep your lip color all day long	13000	437	1295686	0	0	https://www.youtube.com/watch?v=lwygTyL72Ho
4613	beauty	회사원	1180000	Blue eyeliner makeup/깊고 푸른 겨울밤 파티 메이크업	1800	60	265250	0	0	https://www.youtube.com/watch?v=N56p_w4VgrE
4614	beauty	회사원	1180000	1분만에 통머리 만들기/1分で作る簡?お?子ヘア/1Minute bun hair updo style	4800	137	656673	0	0	https://www.youtube.com/watch?v=DiN16R-0gpQ
4615	beauty	회사원	1180000	Products I've used up/다 쓴 제품 공병 리뷰/使い切った化粧品紹介	2100	58	302932	0	0	https://www.youtube.com/watch?v=Rrug1PU5uMI
4616	beauty	회사원	1180000	Fashion film : Tropical Girl By Calary Girl	1000	51	137781	0	0	https://www.youtube.com/watch?v=LBnBtEYvx1Y

Table 1: 영상 정보\_뷰티

id	name	title	url	comment	comm_count
4612	회사원	지워지지 않게 립스틱 바르는 법/How to keep your lip color all day long	https://www.youtube.com/watch?v=lwygTyL72Ho	"아...회사원님 예전꺼보니까 적응안되는 차분"	646
4613	회사원	Blue eyeliner makeup/깊고 푸른 겨울밤 파티 메이크업	https://www.youtube.com/watch?v=N56p_w4VgrE	"이젠 몰랐지 내가 회사원님"	178
4614	회사원	1분만에 통머리 만들기/1分で作る簡?お?子ヘア/1Minute bun hair updo style	https://www.youtube.com/watch?v=DiN16R-0gpQ	"지금보니 인사가 또"	449
4615	회사원	Products I've used up/다 쓴 제품 공병 리뷰/使い切った化粧品紹介	https://www.youtube.com/watch?v=Rrug1PU5uMI	"달라ㅋㅋㅋ 처음:여러분"	118
4616	회사원	Fashion film : Tropical Girl By Calary Girl	https://www.youtube.com/watch?v=LBnBtEYvx1Y	"2020년에도 윗"	86

Table 2: 댓글 정보\_뷰티

id	cate	name	sub	title	like	hate	view	ad	spon	url
0	food	아미아미	167000	제가 할머니라노...추어탕 주어튀김 총각김치 먹방 :)	401	16	3400	1	0	https://www.youtube.com/watch?v=0Mnbn3jTa8U
1	food	아미아미	167000	직접만든 진짜 수제버거10개만 먹을게요 :) (feat 밀리터리버거) 10 homemade hamburgers mukbang	1300	44	31553	0	0	https://www.youtube.com/watch?v=czyWVUcblw
2	food	아미아미	167000	이차들 간단하게 차돌박이 15인분 풀면 김치말이국수 스파게티 돌초밥 먹고왔어요!! mukbang	1600	47	45937	0	0	https://www.youtube.com/watch?v=r-MpEzaUzHc
3	food	아미아미	167000	신전떡볶이 위에 고구마무스 올리고 춘권고구마치즈스틱 먹방!! (feat **비어안주) Spicy Teokbokki Cheese stick mukban	1400	65	50235	0	0	https://www.youtube.com/watch?v=C9F_FJtPVIA
4	food	아미아미	167000	열라면이랑 봉구스밥버거 먹방 :) Spicy ramen Rice burger mukbang eating show	2500	122	105344	0	0	https://www.youtube.com/watch?v=jg5KMw7EIEI
5	food	아미아미	167000	원 팔방에 고등어구이 올려서 된장찌개랑 쌀밥 먹방 :) Grilled Korean Mackerel mukbang	1400	91	49612	1	0	https://www.youtube.com/watch?v=HvOmPbM9aZk

Table 3: 영상 정보\_먹방

id	name	title	url	comment	comm_count
0	아미아미	제가 할머니라노...추어탕 주어튀김 총각김치 먹방 :)	https://www.youtube.com/watch?v=0Mnbn3jTa8U	편집없는 일본영상(생방송)Full-Vide( Live)	178
1	아미아미	직접만든 진짜 수제버거10개만 먹을게요 :) (feat 밀리터리버거)	https://www.youtube.com/watch?v=czyWVUcblw	이집 맛집이네요 ㅋㅋㅋ 여기 아미 수제버거 5개 포장주문이요 !!	248
2	아미아미	이차들 간단하게 차돌박이 15인분 풀면 김치말이국수 스파게티 돌초밥 먹고왔어요!!	https://www.youtube.com/watch?v=r-MpEzaUzHc	저는 아쉬운대로 집에 우삼겹이랑 이차들 차돌면 포장해와서 먹었어요	225
3	아미아미	신전떡볶이 위에 고구마무스 올리고 춘권고구마치즈스틱	https://www.youtube.com/watch?v=C9F_FJtPVIA	떡볶이는 그냥 맛있는 음식	144
4	아미아미	열라면이랑 봉구스밥버거 먹방	https://www.youtube.com/watch?v=jg5KMw7EIEI	아미님 진짜 맛있게드실ㅎㅎ 천천히 드세요	200

Table 4: 댓글 정보\_먹방

id	cate	name	sub	title	like	hate	view	ad	spon	url
7621	game	테드TV	968000	1:1:1 이런경우를 많이있죠?? 같이 극복해봅시다	3500	131	402779	0	0	https://www.youtube.com/watch?v=oDsNyWWf.C4
7622	game	테드TV	968000	불, 웰라 다 있죠? 이맵에서 풀빠세요	2400	90	276996	1	0	https://www.youtube.com/watch?v=1InjswORDzQ
7623	game	테드TV	968000	매우 아픈 스파이크 500+에서 직접 플레이 해봤습니다	5000	210	724722	1	0	https://www.youtube.com/watch?v=BVUjyHnEmZc
7624	game	테드TV	968000	물주면 3복 주는 나무가 있네.. 직접 확인해봐야겠지?	1800	97	204517	0	0	https://www.youtube.com/watch?v=oRpxYraOm48
7625	game	테드TV	968000	막 만든 1렐 계정에 100만원 상자 오픈 해봤습니다 흠 전설뜨네?	51000	4300	8010131	1	0	https://www.youtube.com/watch?v=gsGfCIdWfw

Table 5: 영상 정보\_게임

id	name	title	url	comment	comm_count
7621	테드TV	1:1:1 이런경우들 많이있죠?? 같이 극복해봅시다	https://www.youtube.com/watch?v=oDsNyWWfC4	상황판단 지켰다리 10:11 본인이 포코 리뷰할때 말함...	662
7622	테드TV	불, 켈리 다 있죠? 이땅에서 꿈꿔주세요	https://www.youtube.com/watch?v=1lnjswoRDzQ	요즘 30대분들도 브롤 많이하네 홈 0:21 우당탕 건홍탕	786
7623	테드TV	매우 아픈 스파이크 500+에서 직접 플레이 해봤습니다	https://www.youtube.com/watch?v=BVUjyHnEmZc	1 쓰는놈들 형이 다 두카팬다 방금 브롤 상자에서 스파이크	1467
7624	테드TV	물주면 3쪽 주는 나무가 있네.. 직접 확인해봐야겠지?	https://www.youtube.com/watch?v=oRpxYraOm48	한배도 참신한거 많이 업뎃좀 해주셈.. 2:41	527
7625	테드TV	막 만든 1렐 계정에 100만원 상자 오른 해였습니다 홈 전설쓰네?	https://www.youtube.com/watch?v=gsGfCldWvfw	클래시로알 초장기 : 전설 3대장 시절 (프린세스, 얼음 마법사	7676

Table 6: 댓글 정보\_게임

## 5. 데이터 전처리

유튜버의 모든 영상에 대해 분석을 진행하면 옛날에는 광고를 거의 표시하지 않았거나 뒷광고 논란이 있는 후 급하게 광고 설명을 추가한 영상들에 대해서 분석 결과의 오류가 나타날 것이라고 판단했다. 데이터 수집을 완성한 데이터셋을 보아도 비교적 최근 영상에 광고, 협찬 표시가 활성화되어 있다는 점을 확인할 수 있었다. 따라서 좀 더 정확한 분석 결과를 얻기 위해 2년 이내에 업로드된 영상을 기준으로 분석을 진행하기로 결정했고 2019.01.01 이전의 영상은 분석 대상에서 제외했다. 좋아요 수, 싫어 요수, 구독자 수, 조회수는 크롤링한 결과 int형으로 저장되어있지 않고, 1.5만/ 2.5천 과 같은 string 형태로 저장되어있었다. 이후 분석의 편의성을 위해 만, 천, 백 단위로 문자열을 구분 후 각각 float 형으로 변환하고 10000, 1000, 100을 곱하여 int형으로 바꿔줬다. 또한, 좋아요 수, 싫어요 수를 비공개 설정해둔 경우 크롤링 과정에서 오류를 발생시켰다. 우선 예외 처리를 진행하며 정상적으로 가져온 데이터와 구분할 수 있도록 null 또는 음수 값으로 대체하여 크롤링을 진행했다. 이후 결측치는 광고 여부, 유튜버 이름에 따라 그룹화를 진행해 얻은 평균값으로 대체했다. 댓글 기능을 막아둔 동영상에서 댓글 수는 "중지"라는 대체 값이 들어오도록, 댓글 내용에는 아무런 데이터도 들어오지 않도록 크롤링을 우선 진행했다. 댓글 수와 같은 경우에는 좋아요 / 싫어요 수와는 달리 데이터 누락 레코드 수가 5개 미만으로 매우 적었기 때문에 전체 분석 결과에 미치는 영향이 미비할 것으로 판단하고 대체하지 않고 해당 행을 삭제했다. ad와 spon에 1이 든 동영상의 index를 리스트로 뽑아 비교해 본 결과, 겹치는 경우가 굉장히 많았다. 대부분 유료 광고와 협찬이 같이 진행되는 경우가 많았으며, 유튜버들도 이 개념을 굉장히 혼동하는 경우가 많았음을 추가적으로 알게 되었다. 나아가 시청자들도 광고나 협찬 여부에 관심을 가지는 것이지 둘의 차이에 따라 반응이 달라지지 않음을 고려하여 spon과 ad의 변수를 통합하여 모두 ad에 저장하기로 했다.

분석을 진행하면서 좋아요, 싫어요, 조회수 이외의 또 다른 반응 변수를 고민하던 중 앞서서 댓글 정보 데이터셋에 따로 저장해 두었던 댓글 수 변수도 영상 시청자의 시청 반응을 나타내는 척도 중 하나로 쓰일 수 있을 거라 판단하고 변수에 추가하게 되었다. key변수로 url변수를 미리 지정하고 데이터셋을 나누어 저장했기 때문에 이 변수를 기준으로 삼아 데이터셋 병합을 진행하였고 변수 이름은 동일하게 comm\_count로 추가했다.

## 최종 데이터셋

id	cate	name	sub	title	like	hate	view	ad	url	comm_count
1540	beauty	회사원	1180000	Reducing Time For Washing &	2900	130	194371	0	https://www.youtube.com/watch	359
1541	beauty	회사원	1180000	YouTube 1M Subscribers Gold Button Unboxing!!!	7300	83	169156	0	https://www.youtube.com/watch	905
1542	beauty	회사원	1180000	Fake K-Beauty Cosmetic Review	31000	825	1799680	0	https://www.youtube.com/watch	2291
1543	beauty	회사원	1180000	Improving skin elasticity just by change	11000	460	861171	0	https://www.youtube.com/watch	1009
1544	beauty	회사원	1180000	[Edited Ver.] Boyfriend Is Doing	2600	134	204816	0	https://www.youtube.com/watch	353

Table 7: 최종 영상 정보\_뷰티

id	cate	name	sub	title	like	hate	view	ad	url	comm_count
0	food	아이마미	167000	제가 알마나라노... 주어탕 주어진것 총각김치 먹방 :) Korean food loach soup Fried loach mukbang	401	16	3400	1	https://www.youtube.com/watch?v=0Mbn3jTa8U	178
1	food	아이마미	167000	직접만든 진짜 수제버거10개만 먹을까요 :) (feat 밀리터리버거) 10 homemade hamburgers mukbang	1300	44	31553	0	https://www.youtube.com/watch?v=czyWVlUclfw	248
2	food	아이마미	167000	이차돌 간단하게 작동박이 15인분 졸면 김치말이국수 스파게티 돌조밥 먹고왔어요!! mukbang	1600	47	45937	0	https://www.youtube.com/watch?v=rMpEzaUzHc	225
3	food	아이마미	167000	신진백복이 위에 고구마무스 올리고 춘권고구마치즈스틱 먹방!! (feat **비어안주) Spicy Tteokbokki Cheese stick mukban	1400	65	50235	0	https://www.youtube.com/watch?v=C9F_FhPViA	144
4	food	아이마미	167000	열려면이랑 봉구스밥버거 먹방 :) Spicy ramen Rice burger mukbang eating show	2500	122	105344	0	https://www.youtube.com/watch?v=jg5KMWTElE	200
5	food	아이마미	167000	흰 쌀밥에 고등어구이 올려서 된장찌개랑 찰밥 먹방 :) Grilled Korean Mackerel mukbang	1400	91	49612	1	https://www.youtube.com/watch?v=HwOmPbM9aZk	173

Table 8: 최종 영상 정보\_먹방

id	cate	name	sub	title	like	hate	view	ad	url	comm_count
7621	game	테드TV	968000	1:1:1 이런경우들 많이있죠?? 같이 극복해봅시다	3500	131	402779	0	<a href="https://www.youtube.com/watch?v=oDsNyWWf_C4">https://www.youtube.com/watch?v=oDsNyWWf_C4</a>	662
7622	game	테드TV	968000	불, 웰리 다 있죠? 이렐에서 끝빠세요	2400	90	276996	1	<a href="https://www.youtube.com/watch?v=1InjswoRDzQ">https://www.youtube.com/watch?v=1InjswoRDzQ</a>	786
7623	game	테드TV	968000	매우 아픈 스파이크 500+에서 직접 플레이 해왔습니다	5000	210	724722	1	<a href="https://www.youtube.com/watch?v=BVUjyHnEmZc">https://www.youtube.com/watch?v=BVUjyHnEmZc</a>	1467
7624	game	테드TV	968000	물주면 3쪽 주는 나무가 있네.. 직접 확인해봐야겠지?	1800	97	204517	0	<a href="https://www.youtube.com/watch?v=oRpxYraOm48">https://www.youtube.com/watch?v=oRpxYraOm48</a>	527
7625	game	테드TV	968000	막 만든 1렐 계정에 100만원 장자 오른 해왔습니다 흠 전설뜨네?	510000	4300	8010131	1	<a href="https://www.youtube.com/watch?v=gsGfCldWfwf">https://www.youtube.com/watch?v=gsGfCldWfwf</a>	7676

Table 9: 최종 영상 정보-게임

전처리까지 진행한 뷰티, 먹방, 게임 분야 최종 데이터 셋의 변수 개수는 11개, 레코드 개수는 각각 1545, 7879, 7626 개로 데이터 수집과 정제를 마쳤다.

## 4 Methods and hypotheses

광고 여부에 따른 좋아요 수, 싫어요 수, 조회수, 댓글 수의 변화를 분석하기 위해 총 3가지 회귀 모형을 고려했다.

- (1) 모델:  $\text{outcome} \sim 1 + \text{ad}$
- (2) 모델:  $\text{outcome} \sim 1 + \text{ad} + \text{name}$
- (3) 모델:  $\text{outcome} \sim 1 + \text{ad} + \text{name} + \text{ad}:\text{name}$

반응변수는 각각 좋아요 수, 싫어요 수, 조회수, 댓글 수이고, 설명변수 중 ad는 광고 여부, name은 유튜버 이름이다. (1) 모델은 광고 여부에 따른 평균 비교가 가능하지만 유튜버별로 반응변수 값이 이질적인 점을 설명하지 않는 모델이다. (2) 모델은 (1) 모델에 유튜버 이름 변수를 추가하여 유튜버별로 반응변수 값이 이질적인 점을 설명하지만 광고 여부의 효과가 유튜버 간에 동일하다고 가정한 모델이다. (3) 모델은 (2) 모델에서 광고 여부의 효과가 유튜버별로 이질적이라고 가정하여 교호 작용을 추가한 모델이다. 3가지 모델 중 광고 여부, 유튜버 이름 변수가 범주형이고 뷰티, 먹방, 게임 3가지 카테고리별로 비교를 하여 차이를 파악하고 싶다는 점을 고려하여 (2) 모델을 최종적으로 결정했다.

다중회귀 분석 진행을 위해 광고 여부 변수인 ad, 유튜버 이름 변수인 name은 범주형 데이터이므로 각각 수치형, 문자형 데이터에서 범주형으로 형 변환했다. 설명변수인 광고 여부는 광고/비광고로 2개, 유튜버 이름은 분야별 15명 유튜버를 의미하는 15개 범주를 가진다. 반응변수인 좋아요 수, 싫어요 수, 조회수, 댓글 수는 모두 수치형 변수이다. 또한, 다중회귀 분석에 앞서 수집한 데이터를 바탕으로 뷰티, 먹방, 게임 분야별 광고 비율 시각화를 진행했다. 뷰티, 먹방, 게임 분야의 광고/비광고 비율을 파이 차트로 나타냈다. 파이 차트는 17p 부록 1. 분야별 광고 비율에서 확인 가능하다. 뷰티 분야는 10.9%, 먹방 분야는 5.4 %, 게임 분야는 4.5% 로 광고 비율을 대략적으로 파악했다. 뷰티, 먹방, 게임 분야 순으로 광고 비율이 높았으며 뷰티 분야의 경우 타 분야보다 광고 비율이 두배 이상이고, 게임과 먹방은 광고 비율이 비슷하다는 사실을 발견했다. 이는 뷰티 분야에서 하나의 영상을 찍을 때 다양한 브랜드의 제품에 대한 광고/협찬 제의를 많이 받을 수 있는 반면, 먹방 및 게임의 경우에는 하나의 영상을 찍을 때 1개 광고 상품만 진행할 수 있어서 광고/협찬 제의를 뷰티 분야에 비해 적게 받는다고 판단했다.

## 5 Experimental results

뷰티, 먹방, 게임 3가지 분야로 나뉘어 다중선형회귀 분석을 진행했다. 각각 분야별 광고 여부에 따른 좋아요 수, 싫어요 수, 조회수, 댓글 수의 변화를 총 4가지 다중선형회귀 적합 결과로 광고 여부에 따른 변화를 살펴보고, 분야별 분석 결과를 바탕으로 세 분야의 광고 여부에 따른 좋아요 수, 싫어요 수, 조회수, 댓글 수의 변화를 비교했다. 또한, 뷰티, 먹방, 게임 3가지 분야로 나뉘어 반응변수가 좋아요 수, 싫어요 수, 조회수, 댓글 수일 때 회귀모형의 잔차산점도, 정규확률그림을 살펴봤을 때 오차항의 등분산성, 정규성 가정에 문제점을 발견하여 Box-Cox 변환 방법을 시도했다.

그 결과  $\lambda$  값이 모두 0에 수렴하는 값이 나왔고, 반응변수 값의 범위가 매우 크기 때문에 해석의 편의상 반응변수 로그변환을 시도했다. 로그변환 후에도 모형 진단을 했을 때 여전히 오차항의 등분산성, 정규성에 문제가 있었으나 변환 전보다 훨씬 안정되었고,  $R^2$  값도 증가했기 때문에 로그변환된 형태를 최종 모델로 결정했다. 가독성을 위해 각 모델의 모형진단, 변환의 선택 문제를 해결할 수 있는 체계적인 방법인 Box-Cox 과정은 생략하고 서술했다.

## 5.1 뷰티 분야

### 5.1.1 반응변수가 좋아요 수인 경우

- (1) 로그변환 전 모델:  $\text{like} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{like}) \sim 1 + \text{ad} + \text{name}$

like는 좋아요 수, ad는 광고 여부(1: 광고, 0: 비광고), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중회귀 적합 결과는 다음과 같다.

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	21165.8670	0.0000*	9.3003	0.0000*
ad1	-1587.9264	0.0864	-0.1495	0.0111*
데이지	-15919.1264	0.0000*	-1.2625	0.0000*
디렉터파이	-15909.8482	0.0000*	-1.0096	0.0000*
민스코	-16737.8030	0.0000*	-1.1699	0.0000*
수뷰티	-19875.4544	0.0000*	-2.5480	0.0000*
씨니	-16065.0057	0.0000*	-1.0459	0.0000*
썬님	-14970.7542	0.0000*	-0.8238	0.0000*
연두콩	-18042.1577	0.0000*	-1.5956	0.0000*
유애아인	-12924.7692	0.0000*	-1.0788	0.0000*
유티루	-18364.2233	0.0000*	-1.6087	0.0000*
이사배	-11733.6064	0.0000*	-0.5004	0.0001*
쥬씨	-14838.0252	0.0000*	-1.8421	0.0000*
포니	12918.7746	0.0000*	0.9033	0.0000*
푸어디	-14587.2482	0.0000*	-1.0944	0.0000*
회사원	-13615.1363	0.0000*	-0.7426	0.0000*
	(1) F-statistic	(1) p-value	(1) R-squared	
	29.37	2.2e-16 미만	0.2169	
	(2) F-statistic	(2) p-value	(2) R-squared	
	86.86	2.2e-16 미만	0.4561	

Table 10: 좋아요적합결과-뷰티

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.2169이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 크기 때문에  $\beta_1$ 가 0이라는 귀무가설을 채택한다. 따라서 광고 여부 변수는 유튜버 변수가 동일하다고 할 때 반응변수와 선형의 상관관계가 있다고 할 수 없다.

(2) 로그변환 후 다중회귀 적합시, 로그변환 후  $R^2$ 가 0.4561로 로그변환 전  $R^2$ 인 0.2169 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.1495으로, 동일 유튜버의 영상들에 대하여 광고 영상의  $\log(\text{like})$ 가 비광고 영상의 동일 반응 값보다 0.1495 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 좋아요 수이고,  $y_0$ 가 비광고 영상의 평균적 좋아요 수라고 한다면,  $\log(y_1) - \log(y_0) = -0.1495$ 이고  $\log(\frac{y_1}{y_0}) = -0.1495$ 이다. 좌우변에 자연상수  $e$ 를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.1495} = \text{약 } 0.861$ 임을 알 수 있다. 따라서 광고인 경우 평균적 좋아요 수가 광고가 아닌 경우의 평균적 좋아요 수보다 약 14% 감소하는 것을 알 수 있다. 즉, 광고인 경우 좋아요 수가 감소하는 경향이 보인다.

### 5.1.2 반응변수가 싫어요 수인 경우

- (1) 로그변환 전 모델:  $\text{hate} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{hate}) \sim 1 + \text{ad} + \text{name}$

hate는 싫어요 수, ad는 광고 여부(1: 광고, 0: 비광고), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중회귀 적합 결과는 다음과 같다.

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	335.9364	0.0000*	4.9369	0.0000*
ad1	-32.9902	0.2000	-0.1175	0.0887
데이지	-238.2295	0.0001*	-1.1767	0.0000*
디렉터파이	-201.6841	0.0004*	-0.5475	0.0004*
민스코	-287.8440	0.0000*	-1.4150	0.0000*
주부티	-314.3979	0.0000*	-2.5490	0.0000*
썬니	-277.4070	0.0003*	-1.1666	0.0000*
썬님	-223.0262	0.0001*	-0.5376	0.0003*
연두콩	-276.6120	0.0000*	-1.2341	0.0000*
유애아인	-165.1078	0.0055	-0.8532	0.0000*
유트루	-267.1384	0.0000*	-0.9173	0.0000*
이사배	-189.4039	0.0005*	-0.3301	0.0239*
주씨	-264.9676	0.0004*	-1.8358	0.0000*
포니	377.7293	0.0000*	0.9874	0.0000*
퓨어디	-167.6239	0.0014*	-0.8288	0.0000*
회사원	-71.3813	0.1766	0.1886	0.1829
	(1) F-statistic	(1) p-value	(1) R-squared	
	16.75	2.2e-16 미만	0.1333	
	(2) F-statistic	(2) p-value	(2) R-squared	
	81.5	2.2e-16 미만	0.4392	

Table 11: 싫어요적합결과-뷰티

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.1333이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 크기 때문에  $\beta_1$ 가 0이라는 귀무가설을 채택한다. 따라서 광고 여부 변수는 유튜버 변수가 동일하다고 할 때 반응변수와 선형의 상관관계가 있다고 할 수 없다.

(2) 로그변환 후 다중회귀 적합시, 로그변환 후  $R^2$ 가 0.4392로 로그변환 전  $R^2$ 인 0.1333 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 크기 때문에  $\beta_1$ 가 0이라는 귀무가설을 채택한다. 따라서 **광고 여부 변수는 유튜버 변수가 동일하다고 할 때  $\log(\text{hate})$ 와 선형의 상관관계가 있다고 할 수 없다.**

### 5.1.3 반응변수가 조회수인 경우

- (1) 로그변환 전 모델:  $\text{view} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{view}) \sim 1 + \text{ad} + \text{name}$

view는 조회수, ad는 광고 여부(1: 광고, 0: 비광고), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중회귀 적합 결과는 다음과 같다.

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	928908.8544	0.0000*	13.1478	0.0000*
ad1	-48952.3833	0.2342	-0.1600	0.0157*
페이지	-653454.7814	0.0000*	-1.2117	0.0000*
디렉터	-572112.3870	0.0000*	-0.7797	0.0000*
미스코	-741127.5714	0.0000*	-1.4079	0.0000*
수뷰티	-883343.1332	0.0000*	-2.9595	0.0000*
썸니	-731091.7826	0.0000*	-1.2015	0.0000*
썸님	-564631.3016	0.0000*	-0.6356	0.0000*
연두콩	-779268.2775	0.0000*	-1.6013	0.0000*
유엔아이	-559541.1704	0.0000*	-1.1450	0.0000*
유투트루	-772876.8528	0.0000*	-1.3815	0.0000*
이사배	-454464.2229	0.0000*	-0.3472	0.0133*
쥬썸	-772074.4559	0.0000*	-2.2177	0.0000*
포니	18026.4789	0.8589	0.3802	0.0198*
퓨어디	-502559.3061	0.0000*	-1.0645	0.0000*
회사원	-460165.4071	0.0000*	-0.4718	0.0005*
	(1) F-statistic	(1) p-value	(1) R-squared	
	16.95	2.2e-16 미만	0.06521	
	(2) F-statistic	(2) p-value	(2) R-squared	
	87.24	2.2e-16 미만	0.4572	

Table 12: 조회수적합결과-뷰티

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.06521이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 크기 때문에  $\beta_1$ 가 0이라는 귀무가설을 채택한다. 따라서 광고 여부 변수는 유튜버 변수가 동일하다고 할 때 반응변수와 선형의 상관관계가 있다고 할 수 없다.

(2) 로그변환 후 다중회귀 적합시, 로그변환 후  $R^2$ 가 0.4572 로그변환 전  $R^2$ 인 0.06521 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.1600으로, 동일 유튜버의 영상들에 대하여 광고 영상의  $\log(\text{like})$ 가 비광고 영상의 동일 반응 값보다 0.1600 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 좋아요 수이고,  $y_0$ 가 비광고 영상의 평균적 좋아요 수라고 한다면,  $\log(y_1) - \log(y_0) = -0.1600$ 이고  $\log(\frac{y_1}{y_0}) = -0.1600$ 이다. 좌우변에 자연상수  $e$ 를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.1600} = \text{약 } 0.8521$  임을 알 수 있다. 따라서 광고인 경우 평균적 좋아요 수가 광고가 아닌 경우의 평균적 좋아요 수보다 약 15% 감소하는 것을 알 수 있다. 즉, 광고인 경우 조회수가 감소하는 경향이 보인다.

#### 5.1.4 반응변수가 댓글 수인 경우

유튜버의 수익 창출 및 대중의 평가를 위해 반응변수로 채택한 좋아요 수, 싫어요 수, 조회수는 중요한 척도지만 댓글 수는 별개이기 때문에 가독성을 위해 부록 17p에 기록했다.



## 5.2 먹방 분야

### 5.2.1 반응변수가 좋아요 수인 경우

- (1) 로그변환 전 모델:  $\text{like} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{like}) \sim 1 + \text{ad} + \text{name}$

like는 좋아요 수, ad는 광고 여부(1: 광고, 0: 비광고), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중회귀 적합 결과는 다음과 같다.

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	4261.8025	0.0000*	8.0196	0.0000*
ad1	-2073.8691	0.0367*	-0.1423	0.0010*
도남이먹방	11037.9839	0.0000*	1.0224	0.0000*
맛상무	-1885.4609	0.1212	-0.5798	0.0000*
밥굽남	666.9325	0.6009	0.0564	0.3094
비연	425.9912	0.8362	0.3481	0.0001*
삼대장	1035.2813	0.4171	0.2398	0.0000*
설기양	29110.6171	0.0000*	1.3970	0.0000*
소프	-3229.7968	0.0316*	-1.3187	0.0000*
순이엄마	1040.6452	0.5150	0.0922	0.1851
아미아미	-2203.0993	0.1325	-0.7482	0.0000*
에드마	-1930.4004	0.1278	-0.4480	0.0000*
입짧은햇님	-721.1130	0.5103	-0.1604	0.0008*
정육왕	-1901.0564	0.1714	-1.0375	0.0000*
햄지	64784.4863	0.0000*	3.0353	0.0000*
히밥	379.4188	0.8199	0.1377	0.0577
	(1) F-statistic	(1) p-value	(1) R-squared	
	223.4	2.2e-16 미만	0.3899	
	(2) F-statistic	(2) p-value	(2) R-squared	
	432.3	2.2e-16 미만	0.5535	

Table 13: 좋아요적합결과-먹방

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.3899이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다. 따라서  $\beta_1$ 는 유튜버 변수가 동일하다고 할 때,  $\hat{\beta}_1$ 는 -2073.8691로 광고 영상인 레코드와 비광고 영상인 레코드는 평균적으로 좋아요 수가 약 2074의 차이가 있었다.

(2) 로그변환 후 다중회귀 적합시, 로그변환 후  $R^2$ 가 0.5535로 로그변환 전  $R^2$ 인 0.3899 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.1423으로, 동일 유튜버의 영상들에 대하여 광고 영상의  $\log(\text{like})$ 가 비광고 영상의 동일 반응 값보다 0.1423 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 좋아요 수이고,  $y_0$ 가 비광고 영상의 평균적 좋아요 수라고 한다면,  $\log(y_1) - \log(y_0) = -0.1423$ 이고  $\log(\frac{y_1}{y_0}) = -0.1423$ 이다. 좌우변에 자연상수  $e$ 를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.1423} = \text{약 } 0.867$  임을 알 수 있다. 따라서 광고인 경우 평균적 좋아요 수가 광고가 아닌 경우의 평균적 좋아요 수보다 약 14% 감소하는 것을 알 수 있다. 즉, 광고인 경우 좋아요 수가 감소하는 경향이 보인다.

### 5.2.2 반응변수가 싫어요 수인 경우

- (1) 로그변환 전 모델:  $\text{hate} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{hate}) \sim 1 + \text{ad} + \text{name}$

hate는 싫어요 수, ad는 광고 여부(1: 광고, 0: 비광고), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중회귀 적합 결과는 다음과 같다.

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	280.6591	0.0019*	5.1852	0.0000*
ad1	-165.4039	0.1036	-0.1957	0.0001*
도남이	806.2306	0.0000*	0.6519	0.0000*
맛상무	-120.7899	0.3321	-0.5458	0.0000*
밥굽남	-70.8919	0.5871	-0.3338	0.0000*
비여	-169.3725	0.4221	-0.6642	0.0000*
삼대장	-52.7103	0.6866	-0.2507	0.0001*
설기양	1938.4846	0.0000*	1.1834	0.0000*
소프	-221.8975	0.1493	-1.4885	0.0000*
순이엄마	33.8878	0.8360	-0.2405	0.0026*
아미아미	-165.0340	0.2711	-0.9013	0.0000*
에드머	-224.7410	0.0833	-1.6143	0.0000*
입짧은햇님	-146.1529	0.1925	-0.5914	0.0000*
정육왕	-141.3803	0.3205	-1.2930	0.0000*
햄치	3009.2981	0.0000*	2.7727	0.0000*
히밥	-107.5367	0.5285	-0.3761	0.0000*
	(1) F-statistic	(1) p-value	(1) R-squared	
	60.26	2.2e-16 미만	0.1455	
	(2) F-statistic	(2) p-value	(2) R-squared	
	367.4	2.2e-16 미만	0.5129	

Table 14: 싫어요적합결과-먹방

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.1455이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 크기 때문에  $\beta_1$ 가 0이라는 귀무가설을 채택한다. 따라서 광고 여부 변수는 유튜버 변수가 동일하다고 할 때 반응변수와 선형의 상관관계가 있다고 할 수 없다.

(2) 로그변환 후 다중회귀 적합시, 로그변환 후  $R^2$ 가 0.5129로 로그변환 전  $R^2$ 인 0.3899 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.1957으로, 동일 유튜버의 영상들에 대하여 광고 영상의  $\log(\text{hate})$ 가 비광고 영상의 동일 반응 값보다 0.1957 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 좋아요 수이고,  $y_0$ 가 비광고 영상의 평균적 좋아요 수라고 한다면,  $\log(y_1) - \log(y_0) = -0.1957$ 이고  $\log(\frac{y_1}{y_0}) = -0.1957$ 이다. 좌우변에 자연상수 e를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.1957} = \text{약 } 0.822$ 임을 알 수 있다. 따라서 **광고인 경우 평균적 싫어요 수가 광고가 아닌 경우의 평균적 싫어요 수보다 약 18% 감소하는 것**을 알 수 있다. 우리가 일반적으로 생각하는 것과 달리 시청자들은 광고 영상이라고 해서 영상에 대한 반감을 크게 가지지 않는다. 따라서 싫어요 수 증가는 유튜버의 이미지나 영상 자체의 콘텐츠에 영향을 받는다고 추론해볼 수 있다.

### 5.2.3 반응변수가 조회수인 경우

- (1) 로그변환 전 모델:  $\text{view} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{view}) \sim 1 + \text{ad} + \text{name}$

view는 조회수, ad는 광고 여부(1: 광고, 0: 비광고), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중회귀 적합 결과는 다음과 같다.

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	363591.9696	0.0001*	12.2217	0.0000*
ad1	-198693.5993	0.0618	-0.2445	0.0000*
도남이	927882.8229	0.0000*	0.8475	0.0000*
맛상무	-127426.8614	0.3284	-0.3054	0.0000*
밥	161536.6087	0.2372	0.4310	0.0000*
비연	62822.1068	0.7761	0.6039	0.0000*
삼대장	113584.3390	0.4062	0.4566	0.0000*
설기	1890517.7733	0.0000*	1.1895	0.0000*
소프	-264568.5638	0.1005	-1.1542	0.0000*
순이엄마	-60051.9486	0.7259	-0.3478	0.0000*
아미아미	-233860.9728	0.1362	-1.0957	0.0000*
에드머	-234696.7589	0.0841	-0.7336	0.0000*
입짧은	-66797.9625	0.5693	0.0358	0.5234
정육왕	-33077.0187	0.8243	-0.3487	0.0000*
햄지	4751448.0523	0.0000*	3.0910	0.0000*
히밥	-31734.6341	0.8590	0.0949	0.2662
	(1) F-statistic	(1) p-value	(1) R-squared	
	102.1	2.2e-16 미만	0.2251	
	(2) F-statistic	(2) p-value	(2) R-squared	
	291.2	2.2e-16 미만	0.4548	

Table 15: 조회수적합결과-먹방

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.2251이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 크기 때문에  $\beta_1$ 가 0이라는 귀무가설을 채택한다. 따라서 광고 여부 변수는 유튜버 변수가 동일하다고 할 때 반응변수와 선형의 상관관계가 있다고 할 수 없다.

(2) 로그변환 후 다중회귀 적합시, 로그변환 후  $R^2$ 가 0.4548로 로그변환 전  $R^2$ 인 0.2251 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.2445으로, 동일 유튜버의 영상들에 대하여 광고 영상의  $\log(\text{view})$ 가 비광고 영상의 동일 반응 값보다 0.2445 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 조회수이고,  $y_0$ 가 비광고 영상의 평균적 조회수라고 한다면,  $\log(y_1) - \log(y_0) = -0.2445$ 이고  $\log(\frac{y_1}{y_0}) = -0.2445$ 이다. 좌우변에 자연상수  $e$ 를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.2445} = \text{약 } 0.783$ 임을 알 수 있다. 따라서 **광고인 경우 평균적 조회수가 광고가 아닌 경우의 평균적 조회수보다 약 22% 감소하는 것**을 알 수 있다. 즉, 광고인 경우 조회수가 감소하는 경향이 보인다. 대부분의 먹방 유튜버의 경우에 광고 영상은 제목을 [광고 영상]이라고 붙이는 경향이 있었는데, 이는 시청자들로 하여금 동영상을 클릭하지 않고도 광고임을 알 수 있다. 따라서 조회수 감소에 영향을 끼친 것으로 해석할 수 있다.

#### 5.2.4 반응변수가 댓글 수인 경우

유튜버의 수익 창출 및 대중의 평가를 위해 반응변수로 채택한 좋아요 수, 싫어요 수, 조회수는 중요한 척도지만 댓글 수는 별개이기 때문에 가독성을 위해 부록 18p에 기록했다.

## 5.3 게임 분야

### 5.3.1 반응변수가 좋아요 수인 경우

- (1) 로그변환 전 모델:  $\text{like} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{like}) \sim 1 + \text{ad} + \text{name}$

like는 좋아요 수, ad는 광고 여부(1: 광고, 0: 비광고), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중 회귀 적합 결과는 다음과 같다.

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	3789.0778	0.0000*	8.0431	0.0000*
ad1	-1543.5859	0.0000*	-0.3403	0.0000*
캠브링 GGAM BRING	1167.0039	0.1126	0.1267	0.2107
김재원의 즐거운 세상	8697.9499	0.0000*	1.3189	0.0000*
대도서관TV	-1866.5618	0.0100*	-0.7293	0.0000*
덕대-DD film	1708.4817	0.0271*	-0.9866	0.0000*
도로시Games	-2857.2257	0.0004*	-1.5849	0.0000*
런민기	126.7534	0.8741	-0.4633	0.0000*
배룡 TV	-3428.6296	0.0001*	-3.6907	0.0000*
악동 김블루	10926.7944	0.0000*	1.4554	0.0000*
악어 유튜브	-2128.6535	0.0034*	-0.7518	0.0000*
우왁굳의 게임방송	2863.3317	0.0001*	0.5069	0.0000*
코아 TV	1162.7078	0.1221	0.2824	0.0064*
테드TV	475.4968	0.5152	0.1136	0.2588
하창봉	-2261.0429	0.0022*	-0.9494	0.0000*
해물파전TV	-3265.2393	0.0000*	-1.9100	0.0000*
	(1) F-statistic	(1) p-value	(1) R-squared	
	304.1	2.2e-16 미만	0.3736	
	(2) F-statistic	(2) p-value	(2) R-squared	
	1056	2.2e-16 미만	0.6748	

Table 16: 좋아요적합결과-게임

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.3736이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다. 따라서  $\beta_1$ 는 유튜버 변수가 동일하다고 할 때,  $\hat{\beta}_1$ 는 -1543.5859로 광고 영상인 레코드와 비광고 영상인 레코드는 평균적으로 좋아요 수가 약 1544의 차이가 있었다.

(2) 로그변환 후 다중회귀 적합시, 로그변환 후  $R^2$ 가 0.6748로 로그변환 전  $R^2$ 인 0.3736 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.3403으로, 동일 유튜버의 영상들에 대하여 광고 영상의  $\log(\text{like})$ 가 비광고 영상의 동일 반응 값보다 0.3403 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 좋아요 수이고,  $y_0$ 가 비광고 영상의 평균적 좋아요 수라고 한다면,  $\log(y_1) - \log(y_0) = -0.3403$ 이고  $\log(\frac{y_1}{y_0}) = -0.3403$ 이다. 좌우변에 자연상수 e를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.3403} = \text{약 } 0.712$ 임을 알 수 있다. 따라서 광고인 경우 평균적 좋아요 수가 광고가 아닌 경우의 평균적 좋아요 수보다 약 29% 감소하는 것을 알 수 있다. 즉, 광고인 경우 좋아요 수가 감소하는 경향이 보인다.

### 5.3.2 반응변수가 싫어요 수인 경우

- (1) 로그변환 전 모델:  $\text{hate} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{hate}) \sim 1 + \text{ad} + \text{name}$

hate는 싫어요 수, ad는 광고 여부(광고: 1, 비광고: 0), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중회귀 적합 결과는 다음과 같다.

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	130.7928	0.5892	3.3692	0.0000*
ad1	-340.7721	0.0007*	-0.3273	0.0000*
캠브링 GGAM BRING	220.3243	0.3816	1.8093	0.0000*
김재원의 즐거운 세상	245.8462	0.3276	2.3656	0.0000*
대도서관TV	-66.6191	0.7884	0.1448	0.2256
덕대-DD film	1672.6344	0.0000*	1.4803	0.0000*
도로시Games	-83.8725	0.7599	-0.2841	0.0315*
런민기	56.7998	0.8358	0.8431	0.0000*
배룡 TV	-75.6207	0.7978	-1.8123	0.0000*
악동 김블루	195.0771	0.4509	2.1927	0.0000*
악어 유튜브	-89.6098	0.7184	0.0833	0.4863
우왁굳의 게임방송	54.3187	0.8281	1.4618	0.0000*
코아 TV	37.3961	0.8845	1.2583	0.0000*
테드TV	89.4677	0.7207	1.6272	0.0000*
하창봉	-62.0524	0.8061	0.3605	0.0031*
해물파전TV	-97.1641	0.6995	-0.2605	0.0316*
	(1) F-statistic	(1) p-value	(1) R-squared	
	16.26	2.2e-16 미만	0.02919	
	(2) F-statistic	(2) p-value	(2) R-squared	
	1048	2.2e-16 미만	0.6783	

Table 17: 싫어요적합결과-게임

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.02919이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다. 따라서  $\beta_1$ 는 유튜버 변수가 동일하다고 할 때,  $\hat{\beta}_1$ 는 -340.7721로 광고 영상인 레코드와 비광고 영상인 레코드는 평균적으로 싫어요 수가 약 341의 차이가 있었다.

(2) 로그변환 후 다중선형회귀 적합시, 로그변환 후  $R^2$ 가 0.6783로 로그변환 전  $R^2$ 인 0.02919 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.3273으로, 동일 유튜버의 영상들에 대하여 광고 영상의  $\log(\text{hate})$ 가 비광고 영상의 동일 반응 값보다 0.3273 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 싫어요 수이고,  $y_0$ 가 비광고 영상의 평균적 싫어요 수라고 한다면,  $\log(y_1) - \log(y_0) = -0.3273$ 이고  $\log(\frac{y_1}{y_0}) = -0.3273$ 이다. 좌우변에 자연상수  $e$ 를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.3273} = \text{약 } 0.721$  임을 알 수 있다. 따라서 **광고인 경우 평균적 싫어요 수가 비광고 영상의 평균적 싫어요 수보다 약 28% 감소하는 것**을 알 수 있다. 게임 분야의 경우 광고 여부에 따라 싫어요 수가 감소하는 경향이 있다고 판단하기 어렵고, 유튜버의 언행 및 영상의 품질에 따라 싫어요 수가 증가될 것으로 기대된다.

### 5.3.3 반응변수가 조회수인 경우

- (1) 로그변환 전 모델:  $\text{view} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{view}) \sim 1 + \text{ad} + \text{name}$

view는 조회수, ad는 광고 여부(광고: 1, 비광고: 0), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중회귀 적합 결과는 다음과 같다.

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	243113.9431	0.1788	11.7833	0.0000*
ad1	-322440.8521	0.0000*	-0.4687	0.0000*
캠브링 GGAM BRING	181415.9725	0.3346	0.8491	0.0000*
김재원의 즐거운 세상	647979.8966	0.0006*	1.7378	0.0000*
대도서관TV	-119374.3882	0.5194	-0.4188	0.0003*
덕대-DD film	1239093.3954	0.0000*	-0.0904	0.4628
도로시Games	-149983.4918	0.4641	-0.9778	0.0000*
런민기	139106.3371	0.4964	0.2434	0.0562
배룡 TV	-182205.3866	0.4005	-2.7244	0.0000*
악동 김블루	860549.7946	0.0000*	1.9740	0.0000*
악어 유튜브	-121996.5695	0.5108	-0.2485	0.0316*
우왁굳의 게임방송	243143.4013	0.1931	1.0484	0.0000*
코아 TV	183729.8446	0.3392	0.8307	0.0000*
테드TV	182950.5664	0.3274	0.8369	0.0000*
하창봉	-127529.7177	0.4992	-0.5418	0.0000*
해물파전TV	-159471.0795	0.3962	-0.6103	0.0000*
	(1) F-statistic	(1) p-value	(1) R-squared	
	36.37	2.2e-16 미만	0.06505	
	(2) F-statistic	(2) p-value	(2) R-squared	
	660.5	2.2e-16 미만	0.5647	

Table 18: 조회수적합결과-게임

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.06505이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다. 따라서  $\beta_1$ 는 유튜브 변수가 동일하다고 할 때,  $\hat{\beta}_1$ 는 -322440.8521로 광고 영상인 레코드와 비광고 영상인 레코드는 평균적으로 조회수가 약 322441의 차이가 있었다.

(2) 로그변환 후 다중회귀 적합시, 반응변수의 로그변환 후  $R^2$ 가 0.5647로 로그변환 전  $R^2$ 인 0.06505 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.4687으로, 동일 유튜브의 영상들에 대하여 광고 영상의  $\log(\text{view})$ 가 비광고 영상의 동일 반응 값보다 0.4687 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 조회수이고,  $y_0$ 가 비광고 영상의 평균적 조회수라고 한다면,  $\log(y_1) - \log(y_0) = -0.4687$ 이고  $\log(\frac{y_1}{y_0}) = -0.4687$ 이다. 좌우변에 자연상수  $e$ 를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.4687} = \text{약 } 0.626$ 임을 알 수 있다. 따라서 **광고인 경우 평균적 조회수가 광고가 아닌 경우의 평균적 조회수보다 약 37% 감소하는 것**을 알 수 있다. 즉, 광고인 경우 조회수가 감소하는 경향이 보인다.

#### 5.3.4 반응변수가 댓글 수인 경우

유튜버의 수익 창출 및 대중의 평가를 위해 반응변수로 채택한 좋아요 수, 싫어요 수, 조회수는 중요한 척도지만 댓글 수는 별개이기 때문에 가독성을 위해 부록 19p에 기록했다.

### 5.4 추가분석: 댓글 감정 분석

광고에 따른 시청자들의 댓글 반응을 보기 위하여 광고 영상들의 댓글 내용을 단어별로 분석해보았다. 우선 데이터 테이블에서 'ad' column을 광고와 비광고 테이블로 나누었고, 동영상 별 댓글 내용이 담겨져있는 'comment' 열을 반복적으로 돌면서 분석에 용이하게끔 전처리하였다. KoNLPy에서 제공하는 형태소 분석기 사전을 사용하여 명사인 것들만

골라내었다. 이를 통해 시청자들이 댓글 칸에 어떠한 단어를 공통적으로 많이 사용하는지 알 수 있고 나아가 광고의 효과를 알아볼 수 있다. 분야별 term frequency는 다음과 같다. 단어 빈도수 분석 결과, 뷰티, 먹방 게임 분야에서 자주

	광고						비광고					
	뷰티		먹방		게임		뷰티		먹방		게임	
	token	num	token	num	token	num	token	num	token	num	token	num
1	언니	3975	영상	1035	게임	578	영상	702	영상	537	영상	714
2	영상	2513	언니	422	영상	573	메이크업	356	언니	297	게임	379
3	피부	1041	고기	417	대도	343	제품	345	오늘	279	대도	278
4	메이크업	1030	오늘	395	코아	322	진짜	245	소리	227	사람	259
5	오늘	1004	광고	342	지원	277	화장	223	사람	197	오늘	248
6	제품	886	감사	295	광고	261	얼굴	189	햇님	188	코아	240
7	감사	817	정육	274	오늘	246	하루	177	음식	183	웃기	211
8	데이지	604	상무	236	감사	194	추천	145	진짜	151	진짜	199
9	사랑	592	사람	229	사랑	190	영상	134	감사	151	편집	177
10	진짜	567	진짜	220	사람	142	유튜버	120	고기	132	소리	175
11	광고	484	지원	218	방송	137	이사배	113	생각	125	사랑	165
12	추천	483	리뷰	212	진짜	135	사랑	111	방송	121	방송	155
13	화장	474	사랑	135	숙제	107	관리	104	모습	114	감사	144

Table 19: 상위키워드

뷰티	유료 0.80	포함 0.79	고지 0.76	개정 0.75	표기 0.75	적극적 0.74	계약 0.74	표시 0.73
<p>설마 이번에도 영혼없는 키엘 수분크림 광고인가 했는데 π π 솔찍히 너무 과대 광고예요 광고영상을 이렇게 재밌게 찍다니 사고싶어지잖아요!! 일상속에 아이소이광고가아니라 광고안에 브이로그느낌,,, 물론 제품좋은건 알겠지만 일상내용이 더담겼으면좋았을거같어용!그래두 잘보고가요!! 언니..유튜브 영상에 광고가 너무 많은거같아요..π π..전체의 거의 반.. 언니가 좋아하는 메이크업을 해주세요..</p>								
먹방	유료 0.91	표시 0.90	포함 0.89	시청자 0.88	광고주 0.87	난리 0.86	비난 0.86	간접 0.84
<p>광고는 이렇게 하면 서로 기분좋은듯 시청자들도 기분 좋고 광고주도 기분 좋고 다른 유튜버들보다 비난과 비판이 적은 이유. 다른 유튜버들은 애매모호하게 광고인지 험찬인지 정확히 시청자분들에게 알리지않았음 광고 숨기는 이유가 시청자들이 싫어하기 때문인데 그러면 숨기지 말고 이렇게 맛깔나게 살리는 곳에 광고를 주면 되잖아 와서 대박 우리 드머 개힘들게 올라오면서 광고올거같으면 시청자한테 먼저 말해주고 입꼬리 주체 못하고 좋아한다고 π π 옛날에 광고 받았을때도 유료 광고 적어놓고 참 정직해</p>								
게임	유료 0.96	포함 0.94	삼성 0.89	협찬 0.89	카메라 0.88	회사 0.88	표시 0.86	헌혈 0.56
<p>제목만 봤는데 광고라는게 느껴지네 ㅋㅋㅋㅋㅋㅋㅋㅋ 형 돈 많이벌으라고 광고 끝까지봤다ㅋㅋ 뒷광고를 의심할 여지없이 제목에 광고라고 써있네 뒷광고가 사라질때쯤 이제 앞광고가 유행이 된다. 아 제발 광고 좀 이제 그만봤으면 좋겠다..</p>								

Table 20: "광고" 단어에 대한 댓글반응

나오는 단어는 크게 세 카테고리로 나눌 수 있다. 감정 표현("사랑", "감사", "웃김", "수고")과 콘텐츠관련(2.1 유튜버 이름-"대도", "정육", "상무" 2.2 영상 등장 제품-"고기", "게임", "마스크", "쿠션") 과기술적 관련("소리", "방송", "편집", "영상")으로 분류된다

또한 주목할만한 점은 광고 영상들은 비광고 영상과 달리 자주 사용되는 단어 리스트에 광고와 관련된 단어가("광고",

”지원”, ”유료”, ”협찬”, ”숙제”)가 다수 랭킹 되었다는 점이다. 따라서 중심 단어로 주변 단어를 예측하는 Skip-gram 을 적용해 봄으로써 광고 영상 중 ”광고” 단어의 단어 간 유사도를 본다면 시청자들의 광고에 대한 반응을 보다 더 구체 적으로 알 수 있을 것이다. 이로써 자주 나오는 단어들을 기준으로 Word2vec을 이용하여 단어 간 유사도를 측정했다.

”광고” 단어 관련 말뭉치들을 보면, 시청자들이 유튜버들의 광고 영상을 달가워하고 응원하는 모습이 보였다. 최근 뒷광고 논란이 터지면서 열심히 광고 표기를 한 유튜버들에 대한 인식이 재조명되었기 때문이다. 이는 앞서 시청자들은 유튜버의 광고 영상을 싫어하지 않았다는 회귀분석 결과와 일맥상통하며, 시청자들은 광고 개수보다는 솔직하게 광고 여부를 밝히는 것을 긍정적으로 평가한다고 추론 가능하다.

## 6 Concluding remarks

뷰티 분야는 광고 영상의 평균적 좋아요 수는 비광고 영상의 평균적 좋아요 수 보다 약 14% 감소했고, 로그변환된 싫어요 수, 댓글 수에 유튜버 변수가 동일하다고 할 때 광고 여부 변수는 유의하지 않았다. 조회수는 광고 영상의 경우에 비광고 영상보다 약 15% 평균적으로 감소했다. 먹방 분야는 광고 영상의 평균적 좋아요 수는 비광고 영상의 평균적 좋아요 수 보다 약 14% 감소했다. 그에 비해 광고 영상의 평균적 싫어요 수는 비광고 영상일 때 보다 증가할 것이라 기대했지만 약 18% 감소했다. 조회수, 댓글 수는 광고 영상의 경우에 비광고 영상 보다 각각 약 22% , 12% 씩 평균적으로 감소했다. 게임 분야는 광고 영상의 평균적 좋아요 수는 비광고 영상의 평균적 좋아요 수 보다 약 29% 감소했다. 그에 비해 광고 영상의 평균적 싫어요 수는 비광고 영상일 때 보다 증가할 것이라 기대했지만 약 28% 감소했다. 조회수, 댓글 수는 광고 영상의 경우에 비광고 영상 보다 각각 약 37% , 10% 씩 평균적으로 감소했다.

	좋아요수	싫어요수	조회수
게임	29%감소	28%감소	39%감소
먹방	14%감소	18%감소	22%감소
뷰티	14%감소	유의하지않음	15%감소

Table 21: 카테고리 별 광고 영향

뷰티 분야에서는 광고 여부가 좋아요 수와 조회수, 먹방 분야와 게임 분야에서는 좋아요 수, 싫어요 수, 조회수, 댓글 수에 모두 영향을 미쳤으며, 세 카테고리 모두 광고에 대한 시청자들의 반응이 유사한 형태를 보였다. 광고 영상인 경우 좋아요, 싫어요 수, 조회수, 댓글 수가 모두 감소하는 양상을 보였다. 단, 뷰티 분야에서 싫어요 수, 댓글 수에 대한 설명은 제외한 경우이다. 다만 영향의 정도에 따라서 차이를 보였다. 즉, 게임, 먹방, 뷰티 분야 순으로 시청자들은 광고 영상의 영향을 크게 받는다고 볼 수 있다. 이러한 차이는 분야별로 콘텐츠 진행 양상과 시청자들의 영상에 대한 기대하는 바가 다르기 때문이다. 뷰티 분야의 경우, 시청자들은 유튜버가 어떠한 콘셉트로 얼마나 화장을 잘하느냐를 어떠한 제품을 썼느냐보다 우선시한다. 먹방의 경우도 마찬가지이다. 시청자들은 먹방 유튜버가 무슨 음식을 먹었냐 보다는 얼마나 잘, 많이, 맛있게 먹는지를 보는 경향이 있으므로 영상이 광고이냐는 덜 민감하게 받아들인다. 반면 게임과 같은 경우에는 게임별로 규칙과 내용이 다르기 때문에 플레이하는 방식이 달라진다. 즉, 광고 여부가 콘텐츠 진행 방식 자체를 결정지어버리기 때문에 시청자들도 광고의 영향을 가장 많이 받게 되는 것이다.



## 부록

### 1. 분야별 광고 비율

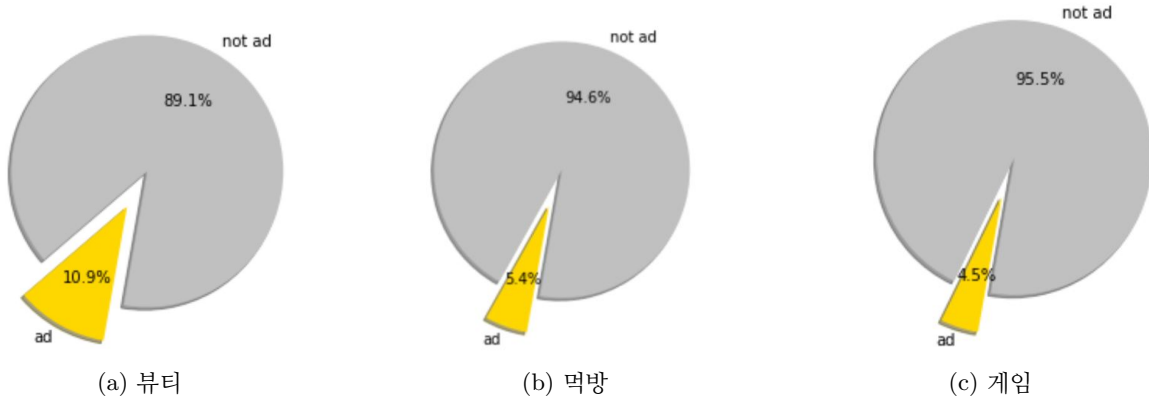


Figure 1: 분야별 광고 비율

### 2. 반응변수가 댓글수인 경우

- (1) 로그변환 전 모델:  $\text{comm\_count} \sim 1 + \text{ad} + \text{name}$ , (2) 로그변환 후 모델:  $\log(\text{comm\_count}) \sim 1 + \text{ad} + \text{name}$

comm.count는 댓글 수, ad는 광고 여부(광고: 1, 비광고: 0), name는 유튜버 이름을 의미하는 변수이다. (1), (2) 모델의 다중회귀 적합 결과는 다음과 같다.

(뷰티 분야)

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	815.5046	0.0000*	6.3543	0.0000*
ad1	15.5297	0.7969	0.0174	0.7635
데이지	-332.5756	0.0204*	-0.6586	0.0000*
디렉터파이	-306.1432	0.0226*	-0.3655	0.0044*
미스코	-508.0773	0.0000*	-0.8554	0.0000*
수뷰티	-702.5154	0.0000*	-1.8687	0.0000*
썸니	-508.8539	0.0043*	-0.7219	0.0000*
썸님	26.5789	0.8366	-0.2839	0.0213*
연두콩	-402.2507	0.0018*	-0.7805	0.0000*
유앤아인	-264.5772	0.0578	-0.5655	0.0000*
유투트루	-517.8772	0.0001*	-0.8445	0.0000*
이사배	-116.0216	0.3635	-0.1678	0.1693
쥬썸	-637.6436	0.0003*	-1.8968	0.0000*
포니	126.1956	0.3959	0.1974	0.1651
푸어디	-302.0701	0.0137*	-0.5414	0.0000*
회사원	-175.4483	0.1565	-0.2712	0.0221*
	(1) F-statistic	(1) p-value	(1) R-squared	
	8.143	2.2e-16 미만	0.06521	
	(2) F-statistic	(2) p-value	(2) R-squared	
	0.06521	2.2e-16 미만	0.3263	

Table 22: 적합결과-뷰티

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.06521이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은

유의수준  $\alpha=0.05$ 보다 크기 때문에  $\beta_1$  가 0이라는 귀무가설을 채택한다. 따라서 광고 여부 변수는 유튜버 변수가 동일하다고 할 때 반응변수와 선형의 상관관계가 있다고 할 수 없다.

(2) 로그변환 후 다중회귀 적합시, 로그변환 후  $R^2$ 가 0.3263로 로그변환 전  $R^2$ 인 0.06521 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 크기 때문에  $\beta_1$  가 0이라는 귀무가설을 채택한다. 따라서 광고 여부 변수는 유튜버 변수가 동일하다고 할 때  $\log(\text{comm\_count})$ 와 선형의 상관관계가 있다고 할 수 없다.

(먹방 분야)

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	208.6624	0.0000*	5.1461	0.0000*
ad1	-73.3362	0.1387	-0.1250	0.0025*
도남이	378.6876	0.0000*	0.7200	0.0000*
맛상무	324.8520	0.0000*	0.7570	0.0000*
밥굽남	378.0836	0.0000*	0.8804	0.0000*
비연	588.0404	0.0000*	1.3958	0.0000*
삼대창	539.9219	0.0000*	1.1706	0.0000*
설기양	1213.4385	0.0000*	1.2300	0.0000*
소프	-44.7014	0.5511	-0.2937	0.0000*
순이엄마	459.9675	0.0000*	0.9998	0.0000*
아미아미	-20.6107	0.7779	-0.0773	0.2044
에드마	280.6242	0.0000*	0.8783	0.0000*
입짧은햇님	4.0100	0.9415	-0.1903	0.0000*
정육왕	55.1220	0.4268	0.0796	0.1687
햄지	2532.1774	0.0000*	2.6304	0.0000*
히밥	288.3529	0.0005*	0.7622	0.0000*
	(1) F-statistic	(1) p-value	(1) R-squared	
	123.5	2.2e-16 미만	0.2605	
	(2) F-statistic	(2) p-value	(2) R-squared	
	273	2.2e-16 미만	0.4388	

Table 23: 댓글수적합결과-먹방

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.2605이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당한 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$  가 0이라는 귀무가설을 기각한다. 따라서  $\beta_1$ 는 유튜버 변수가 동일하다고 할 때,  $\hat{\beta}_1$ 는 -73.3362로 광고 영상인 레코드와 비광고 영상인 레코드는 평균적으로 댓글 수가 약 73의 차이가 있었다.

(2) 로그변환 후 다중회귀 적합시, 반응변수의 로그변환 후  $R^2$ 가 0.4388로 로그변환 전  $R^2$ 인 0.2605 보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$  가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.1250으로, 동일 유튜버의 영상들에 대하여 광고 영상의  $\log(\text{comm\_count})$ 가 비광고 영상의 동일 반응 값보다 0.1250 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 댓글 수이고,  $y_0$ 가 비광고 영상의 평균적 댓글 수라고 한다면,  $\log(y_1) - \log(y_0) = -0.1250$ 이고  $\log(\frac{y_1}{y_0}) = -0.1250$ 이다. 좌우변에 자연상수 e를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.1250} = \text{약 } 0.882$ 임을 알 수 있다. 따라서 광고인 경우 평균적 댓글 수가 광고가 아닌 경우의

평균적 댓글 수보다 약 12% 감소하는 것을 알 수 있다. 즉, 광고인 경우 댓글 수가 감소하는 경향이 보인다.

(게임 분야)

	(1) Estimate	(1) Pr(> t )	(2) Estimate	(2) Pr(> t )
(Intercept)	422.3710	0.0006*	5.8546	0.0000*
ad1	5.8547	0.9152	-0.1020	0.0300*
캠브링 GGAM BRING	639.9209	0.0000*	0.7238	0.0000*
김재원의 즐거운 세상	2193.1869	0.0000*	1.8777	0.0000*
대도서관TV	-277.5900	0.0272*	-1.2258	0.0000*
덕대-DD film	-272.8828	0.0739	-1.3653	0.0000*
도로시Games	-183.9936	0.1854	-0.7768	0.0000*
런민기	697.9677	0.0000*	0.4131	0.0005*
배룡 TV	-368.2362	0.0122*	-2.8022	0.0000*
악동 김블루	1410.4428	0.0000*	1.4885	0.0000*
악어 유튜브	-168.7232	0.1800	-0.5913	0.0000*
우왁굳의 게임방송	604.5210	0.0000*	0.8548	0.0000*
코아 TV	480.8360	0.0002*	0.6188	0.0000*
테드TV	684.2149	0.0000*	0.8348	0.0000*
하창봉	-75.0354	0.5577	-0.3103	0.0046*
해물파전TV	-326.7383	0.0104*	-1.3899	0.0000*
	(1) F-statistic	(1) p-value	(1) R-squared	
	348.5	2.2e-16 미만	0.4116	
	(2) F-statistic	(2) p-value	(2) R-squared	
	1048	2.2e-16 미만	0.6783	

Table 24: 댓글수적합결과-게임

(1) 로그변환 전 다중회귀 적합시,  $R^2$ 는 0.4116이다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관관계가 있다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 크기 때문에  $\beta_1$ 가 0이라는 귀무가설을 채택한다. 따라서 광고 여부 변수는 유튜버 변수가 동일 하다고 할 때 반응변수와 선형의 상관관계가 있다고 할 수 없다.

(2) 로그변환 후 다중회귀 적합시, 반응변수의 로그변환 후  $R^2$ 가 0.6783로 로그변환 전  $R^2$ 인 0.4116보다 증가했다. F-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에 설명변수의 회귀계수인  $\beta$ 가 모두 0이라는 귀무가설을 기각한다. 따라서 적어도 하나의  $\beta$ 는 0이 아니며 적어도 하나의 설명변수는 반응변수와 선형의 상관 관계가 있음을 알 수 있었다. 광고 여부 변수에 해당하는 t-검정의 p-값은 유의수준  $\alpha=0.05$ 보다 작고 0에 매우 가까운 작은 값이기 때문에  $\beta_1$ 가 0이라는 귀무가설을 기각한다.  $\hat{\beta}_1$ 는 -0.1020으로, 동일 유튜버의 영상들에 대하여 광고 영상의  $\log(\text{comm\_count})$ 가 비광고 영상의 동일 반응 값보다 0.1020 작았다. 즉,  $y_1$ 가 광고 영상의 평균적 댓글 수이고,  $y_0$ 가 비광고 영상의 평균적 댓글 수라고 한다면,  $\log(y_1) - \log(y_0) = -0.1020$ 이고  $\log(\frac{y_1}{y_0}) = -0.1020$ 이다. 좌우변에 자연상수 e를 각각 적용하면  $\frac{y_1}{y_0} = e^{-0.1020} = \text{약 } 0.903$ 임을 알 수 있다. 따라서 광고인 경우 평균적 댓글 수가 광고가 아닌 경우의 평균적 댓글 수보다 약 10% 감소하는 것을 알 수 있다. 즉, 광고인 경우 댓글 수가 감소하는 경향이 보인다.