

VOD 데이터 분석

• 데이터 분석 목록

1. 시간에 따른 VOD 시청 분석
2. 가장 많이 본 VOD
3. 예고편 시청이 실제 시청으로 이어지는 전환율
- 4-1. 상세페이지 머무는 시간 – 실제 시청 전환율
- 4-2. 상세페이지 머무는 시간 – 시청 시간 사이의 관계

데이터 전처리

1. 시청 시간이 1분 미만인 데이터 삭제

시청지표 (시청률)

닐슨코리아, TNMS 멀티미디어 등 시청률 조사기관에 따르면 시청률 조사 방법

TV 시청률 지표는 1분 단위로 산정된 값이기 때문에 1분 미만의 시청 시간은 전처리 삭제

22557 rows 중에서 17424 rows가 남고 1분 미만인 데이터는 5133 rows

2. 제목 끝 .(온점) 제거

제목 끝 .이 붙어 있는 경우 제목 정규화가 어려워서 삭제

0	66875000	장군재상 10회 (A)	TV드라마	외화 시리즈	2472	조옥근은 업소에게 호청의 짝사랑을 전하지만 업소는 오히려 호청을 찾아가 호되게 나무...	마사순,성일륜,정천,왕초연,장준명	00:43	20240123085102
1	66705000	수호지의 무송 04회.	TV드라마	외화 시리즈	1891	무대는 무송을 집으로 데려와 회포를 푼다. 반금련은 흰칠하고 잘생긴 무송을 보고 못...	유대경,변장강,손요기,장한	00:45	20240116203741
2	66705000	수호지의 무송 04회.	TV드라마	외화 시리즈	632	무대는 무송을 집으로 데려와 회포를 푼다. 반금련은 흰칠하고 잘생긴 무송을 보고 못...	유대경,변장강,손요기,장한	00:45	20240117193153

데이터 전처리

※ 수정 사항

- 분석 코드 중 disp_rtm(러닝 타임) -> use_tms(시청 시간)으로 수정하여 분석
- 분석 코드 중 strt_dt 문제로 전처리 파일 재생성
- 시청 시간이 1분 미만인 데이터

: 예고 시청 분석할 때를 제외한 경우 삭제하기 위해 별도의 파일로 생성

인코딩+ '제거+1분 미만 제거) vod_전처리_19, vod_movie_전처리_19, vod_series_전처리_19

인코딩+ '제거(예고편 포함)) vod_preview, vod_movie_preview, vod_series_preview

데이터 분석

• 데이터 전처리 – 1, 2, 4 분석에 사용하는 dataframe에 적용

프로그램 이름 정규화 함수

```
def normalize_program_name(name):  
    # 회차 제거 (모범택시 2(10회) -> 모범택시 2())  
    name = re.sub(r'\d+회', '', name)  
    # ()를 포함한 문자열 제거 (인어공주(2023)(예고) -> 인어공주)  
    name = re.sub(r'\(.*?\)', '', name)  
    # 대괄호 제거 ([설 특집 영화]비공식작전 -> 비공식작전)  
    name = re.sub(r'\[.*?\]', '', name).strip()  
    return name
```

00:00 형식의 datetiem -> n분으로 변경 1:43 = 103분

```
vod['disp_m'] = vod['disp_rtm'].dropna().apply(lambda x: int(str(x).split(':')[0]) * 60 + int(str(x).split(':')[1])  
if ':' in str(x) else float('nan'))  
# 실수형 -> 정수형 형변환  
vod['disp_m'] = vod['disp_m'].fillna(0).astype(int)
```

데이터 분석

• 데이터 전처리 - 3 분석에 사용하는 dataframe에 적용

00:00 형식의 datetime -> n분으로 변경 1:43 = 103분

```
vod['disp_m'] = vod['disp_rtm'].dropna().apply(lambda x: int(str(x).split(':')[0]) * 60 + int(str(x).split(':')[1])  
if ':' in str(x) else float('nan'))
```

실수형 -> 정수형 형변환

```
vod['disp_m'] = vod['disp_m'].fillna(0).astype(int)
```

프로그램 이름 정규화 함수 - [preview]가 True/False인 열 추가 후 정규화

```
def normalize_program_name(name):
```

```
    # 회차 제거 (모범택시 2(10회) -> 모범택시 2())
```

```
    name = re.sub(r'\d+회', '', name)
```

```
    # ()를 포함한 문자열 제거 (인어공주(2023)(예고) -> 인어공주)
```

```
    name = re.sub(r'\(..*?\)', '', name)
```

```
    # 대괄호 제거 ([설 특집 영화]비공식작전 -> 비공식작전)
```

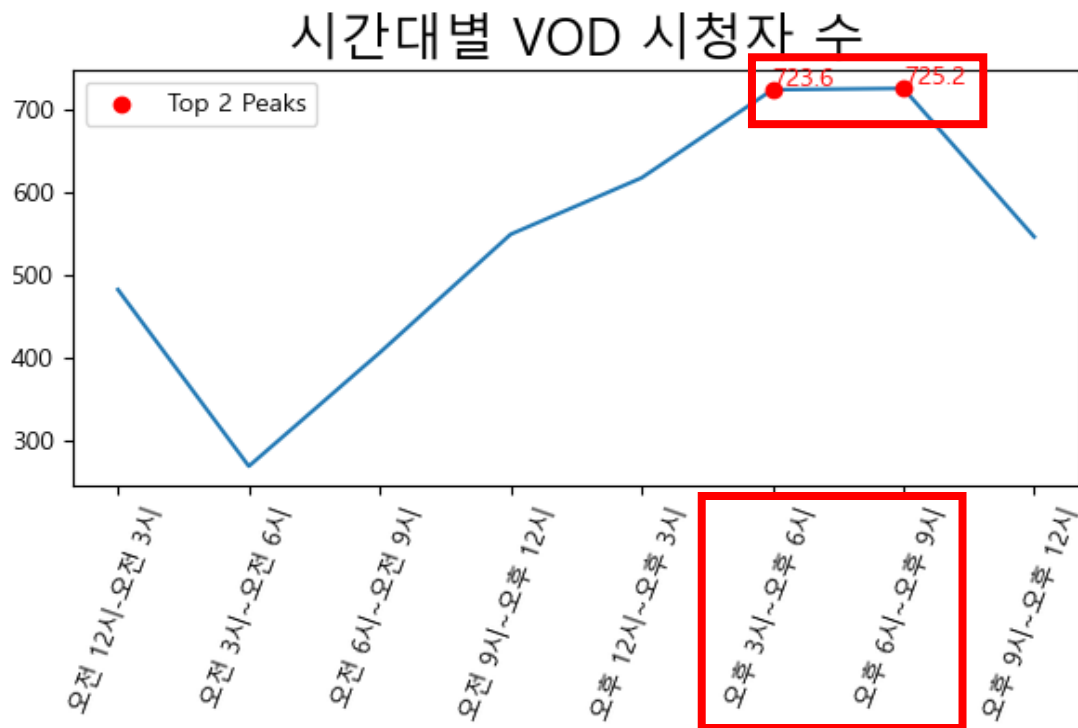
```
    name = re.sub(r'\[.*?\]', '', name).strip()
```

```
    return name
```

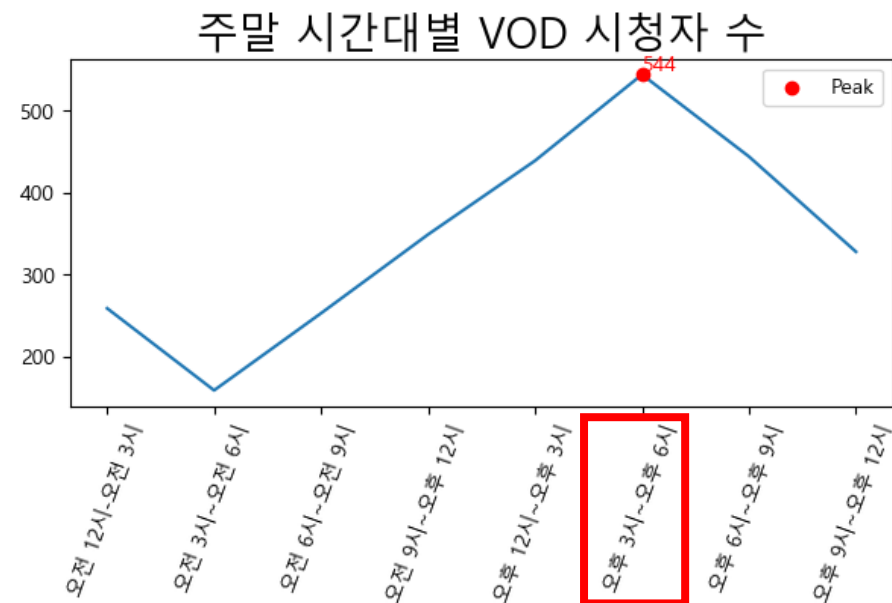
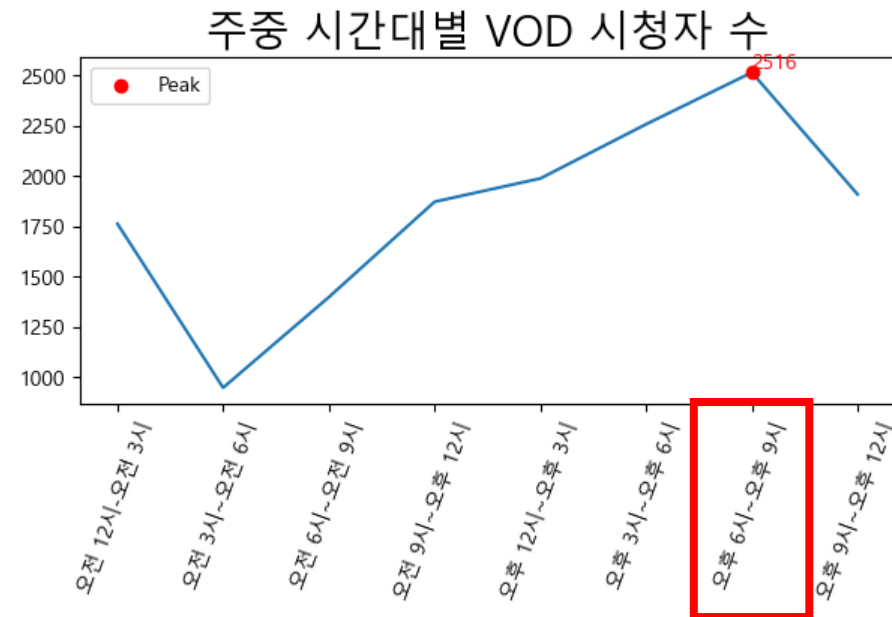
데이터 분석

1. 시간에 따른 VOD 시청 분석

- 전체/주중/ 주말 시간대별 VOD 시청자수



주중 데이터는 5일, 주말 데이터는 2일로 나누어
일별 vod 시청자 수로 확인



데이터 분석

1. 시간에 따른 VOD 시청 분석

- 주중/주말과 viewing_time과의 관계

	viewing_time
viewing_weekday_label	
주말	2024-04-24 00:36:04.532326400
주중	2024-04-24 00:36:12.372162560

데이터 분석

1. 시간대별 VOD 콘텐츠 분석

- 시간대별 사용자 구분
- 시간대별 시청 장르 분석 – 추천에 사용

데이터 분석

2. 가장 많이 본 VOD

- 가장 많이 본 VOD 대분류/소분류

```
ct_cl
TV드라마      6893
영화          4159
TV 연예/오락  2688
기타          2118
TV 시사/교양   661
TV애니메이션  447
키즈          398
다큐          36
스포츠        18
우리동네      3
라이프        2
미분류        1
Name: count, dtype: int64
```

```
genre_of_ct_cl
기타      8811
외화 시리즈  3926
성인      1307
액션/어드벤처  973
드라마     754
애니메이션  325
코미디     237
공포/스릴러  205
멜로       194
인물/다큐   123
명랑/코믹   108
액션/모험    86
무협        80
SF/환타지   80
학습        49
다큐멘터리    34
오락        30
인물        19
추리/미스터리  18
레슬링       18
동물        9
로맨틱코미디  7
무협/환타지   6
역사        5
자연        5
호러/공포     4
연예/오락     3
학원/순정/연애  3
단편        2
운동/건강     2
미분류       1
```

데이터 분석

2. 가장 많이 본 VOD

- 가장 많이 본 VOD 프로그램 상위 30개

asset nm		
성인물		409
성인물		308
세 번째 결혼	266	
러니매	251	
성인물	234	
꼬리에 꼬리를 무는 그날 이야기		221
효심이네 각자도생	187	
우당탕탕 패밀리	178	
밤에 피는 꽃	175	
연화루	155	
내 딸 서영이	146	
안락전	137	
놀면 뭐하니?	133	
장풍도 : 운명의 바람		121
성인물	115	
상군새장	111	
미쓰우리새끼	109	
성인물	108	
성한찬란	105	

낭만닥터 김사부3	102	
육남매	99	
서울의 봄	98	
한문철의 블랙박스 리뷰		97
재벌X형사	95	
천고결진	93	
신발 벗고 돌싱포맨		90
미스터 로또	89	
랑군불여의 : 내 낭군은 늑대		87
나 혼자 산다	84	
도둑놈 도둑님	83	

데이터 분석

2. 가장 많이 본 VOD

- 가장 오래 시청한 VOD 장르(대분류) & VOD 시청 시간 평균

	ct_cl	sum	mean
2	TV드라마	13422562	1947.274336
1	TV 연예/오락	7829345	2912.702753
9	영화	6156824	1480.361625
0	TV 시사/교양	1777283	2688.779123
4	기타	1663911	785.604816
3	TV애니메이션	349827	782.610738
11	키즈	140030	351.834171
5	다큐	55404	1539.000000
8	스포츠	14531	807.277778
10	우리동네	360	120.000000

데이터 분석

2. 가장 많이 본 VOD

- 가장 오래 시청한 VOD 장르(소분류) & VOD 시청 시간 평균

	genre_of_ct_cl	sum	mean
2	기타	17749160	2014.431960
21	외화 시리즈	6734674	1715.403464
17	액션/어드벤처	2393224	2459.634121
6	드라마	1557171	2065.213528
27	코미디	485385	2048.037975
24	인물/다큐	445009	3617.959350
15	애니메이션	421715	1297.584615
1	공포/스릴러	345479	1685.263415
14	성인	295741	226.274675
9	멜로	224623	1157.850515

데이터 분석

2. 가장 많이 본 VOD

- 가장 오래 시청한 VOD 프로그램 & 시청 시간 평균

	asset_nm	ct_cl	sum	mean
738	런닝맨	TV 연예/오락	718154	2861.171315
339	꼬리에꼬리를무는그날이야기	TV 시사/교양	647077	2927.950226
993	미운우리새끼	TV 연예/오락	436900	4008.256881
1321	세 번째 결혼	TV드라마	413331	1553.875940
482	내딸서영이	TV드라마	403395	2762.979452
2659	한문철의 블랙박스 리뷰	TV 시사/교양	377212	3888.783505
987	미스트롯 3	TV 연예/오락	370066	5212.197183
1039	밤에 피는 꽃	TV드라마	363731	2078.462857
1548	아내의 속사정:음란한 거래	기타	356491	871.616137
1295	서울의 봄	영화	356120	3633.877551
2771	효심이네 각자도생	TV드라마	344150	1840.374332
519	놀면 뭐하니?	TV 연예/오락	324136	2437.112782
1486	신발 벗고 돌싱포맨	TV 연예/오락	308147	3423.855556
1794	연화루	TV드라마	300702	1940.012903
2121	재벌X형사	TV드라마	268857	2830.073684

▲ 대분류

	asset_nm	genre_of_ct_cl	sum	mean
738	런닝맨	기타	718154	2861.171315
339	꼬리에꼬리를무는그날이야기	기타	647077	2927.950226
993	미운우리새끼	기타	436900	4008.256881
1321	세 번째 결혼	기타	413331	1553.875940
482	내딸서영이	기타	403395	2762.979452
2662	한문철의 블랙박스 리뷰	인물/다큐	377212	3888.783505
987	미스트롯 3	기타	370066	5212.197183
1039	밤에 피는 꽃	기타	363731	2078.462857
1548	아내의 속사정:음란한 거래	기타	356491	871.616137
1295	서울의 봄	드라마	356120	3633.877551
2774	효심이네 각자도생	기타	344150	1840.374332
519	놀면 뭐하니?	기타	324136	2437.112782
1486	신발 벗고 돌싱포맨	기타	308147	3423.855556
1795	연화루	외화 시리즈	300702	1940.012903
2123	재벌X형사	기타	268857	2830.073684

▲ 소분류

가장 오래 시청한 VOD의 소분류가 제대로 되어있지 않고 기타로 되어 있음

데이터 분석

2. 가장 많이 본 VOD

- 가장 오래 시청한 VOD 프로그램 & 시청 시간 평균

- 실제로 csv파일 확인해보면 TV드라마의 경우는 기타/외화 시리즈로만 분류되어 있고 TV연예/오락의 경우 기타로 분류되어 있음

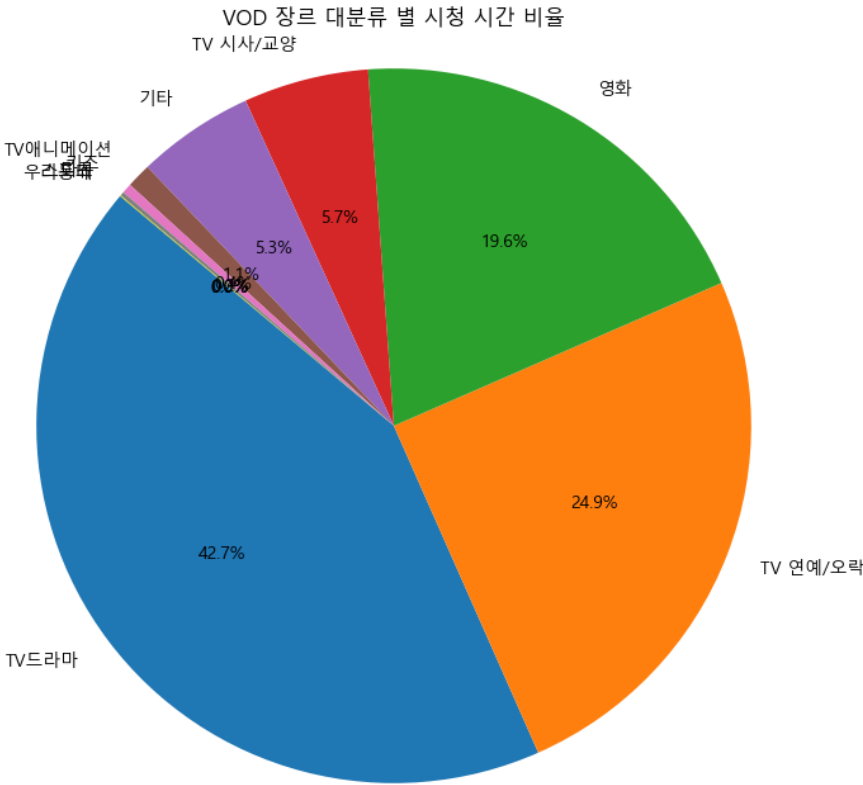
65015000	화요일은 밤이 좋아 68회(23/05/16)	TV 연예/오락	기타	62474000	국민사형투표 06회(23/09/14)	TV드라마	기타
62701000	최강야구 31회(23/01/23)	TV 연예/오락	기타	67303000	철인왕후 19회	TV드라마	외화 시리즈
62701000	최강야구 30회(23/01/16)	TV 연예/오락	기타	64521000	웰컴투 삼달리 07회(23/12/23)	TV드라마	기타
66873000	최강야구 75회(24/01/08)	TV 연예/오락	기타	64521000	웰컴투 삼달리 06회(23/12/17)	TV드라마	기타
67053000	골 때리는 그녀들 135회(24/03/20)	TV 연예/오락	기타	66056000	소방서 옆 경찰서 그리고 국과수 03회(23/08/11)	TV드라마	기타
67053000	골 때리는 그녀들 135회(24/03/20)	TV 연예/오락	기타	66056000	소방서 옆 경찰서 그리고 국과수 11회(23/09/08)	TV드라마	기타
60224000	몽처야 찬다 3 15회(24/01/21)	TV 연예/오락	기타	64984000	조선변호사 02회(23/04/01)	TV드라마	기타
65300000	미스트롯 3 05회(24/01/18)	TV 연예/오락	기타	64680000	웰컴투 삼달리 15회(24/01/20)	TV드라마	기타
62701000	최강야구 39회(23/05/01)	TV 연예/오락	기타	66056000	소방서 옆 경찰서 그리고 국과수 01회(23/08/04)	TV드라마	기타
65300000	미스트롯 3 01회(23/12/21)	TV 연예/오락	기타	64984000	조선변호사 01회(23/03/31)	TV드라마	기타
61141000	불타는 장미단 27회(23/10/17)	TV 연예/오락	기타	64984000	조선변호사 01회(23/03/31)	TV드라마	기타
62701000	최강야구 37회(23/04/17)	TV 연예/오락	기타	64984000	조선변호사 01회(23/03/31)	TV드라마	기타
66873000	최강야구 76회(24/01/15)	TV 연예/오락	기타	64984000	조선변호사 01회(23/03/31)	TV드라마	기타
65265000	(HD)미운우리새끼 374회(23/12/31)	TV 연예/오락	기타	64984000	조선변호사 14회(23/05/13)	TV드라마	기타
60593000	(HD)미운우리새끼 374회(23/12/31)	TV 연예/오락	기타	66056000	소방서 옆 경찰서 그리고 국과수 09회(23/09/01)	TV드라마	기타
62701000	최강야구 32회(23/01/30)	TV 연예/오락	기타	66056000	소방서 옆 경찰서 그리고 국과수 09회(23/09/01)	TV드라마	기타
65015000	화요일은 밤이 좋아 98회(24/01/30)	TV 연예/오락	기타	64984000	조선변호사 13회(23/05/12)	TV드라마	기타
65015000	화요일은 밤이 좋아 98회(24/01/30)	TV 연예/오락	기타	64984000	조선변호사 13회(23/05/12)	TV드라마	기타
60593000	미스트롯 3 03회(24/01/04)	TV 연예/오락	기타	66056000	소방서 옆 경찰서 그리고 국과수 05회(23/08/18)	TV드라마	기타
66792000	싱어게인 3 04회(23/11/16)	TV 연예/오락	기타	66056000	소방서 옆 경찰서 그리고 국과수 05회(23/08/18)	TV드라마	기타

데이터 분석

2. 가장 많이 본 VOD

- 가장 오래 시청한 VOD 프로그램 & 시청 시간 평균

	asset_nm	genre_of_ct_cl	sum	mean
738	런닝맨	기타	718154	2861.171315
339	꼬리에 꼬리를 무는 그날 이야기	기타	647077	2927.950226
993	미운우리새끼	기타	436900	4008.256881
1321	세 번째 결혼	기타	413331	1553.875940
482	내 딸 서영이	기타	403395	2762.979452
2662	한문철의 블랙박스 리뷰	인물/다큐	377212	3888.783505
987	미스트롯 3	기타	370066	5212.197183
1039	밤에 피는 꽃	기타	363731	2078.462857
1548	아내의 속사정: 음란한 거래	기타	356491	871.616137
1295	서울의 봄	드라마	356120	3633.877551
2774	효심이네 각자도생	기타	344150	1840.374332
519	놀면 뭐하니?	기타	324136	2437.112782
1486	신발 벗고 돌싱포맨	기타	308147	3423.855556
1795	연화루	외화 시리즈	300702	1940.012903
2123	재벌X형사	기타	268857	2830.073684



▲ 소분류

→ TV드라마/TV 연예/오락이 차지하는 시청 시간 비율이 전체 시청의 67.6%로 크게 차지
추가로 장르를 분류할 필요가 있어 보임

2. 가장 많이 본 VOD

- 가장 인기있었던 출연자 10명의 출연작

ACTR_DISP	
유재석	450
마츠시타 미오리	423
안내상	413
하하	394
장도연	366
전노민	359
사사키 아키	322
김광규	278
양세찬	277
김상중	273
Name: count, dtype: int64	

['해피투게더 시즌4', '런닝맨', '유 퀴즈 온 더 블럭', '놀면 뭐하니?', '해피투게더 시즌3', '설에 놀면 뭐하니? 인생라면', '일로 만난 사이', '무한도전 Classic', '2023 런닝맨 게스트 어워즈', '투유프로젝트-슈가맨', '식스센스', '식스센스2', '투유프로젝트-슈가맨3'], dtype=object), array(['유부녀들-섹스 스캔들', '폭 적셔주는 질편한 마사지', '무방비 엉덩이의 간호사', '아내의 속사정:음란한 거래'], dtype=object), array(['세 번째 결혼', '소용없어 거짓말', '강철비2: 정상회담', '송곳', '서울의 봄', '무 법변호사'], dtype=object), array(['런닝맨', '놀면 뭐하니?', '오은영 리포트 결혼 지옥', '무한도전 Classic', '2023 런닝맨 게스트 어워즈'], dtype=object), array(['태어난 김에 세계일주3', '꼬리에꼬리 를무는그날이야기', '태어난 김에 세계일주2', '개는 훌륭하다', '전 국민 힐링쇼-푸바오와 할부지', ' 꼬리에 꼬리를 무는 그날 이야기 2', '꼬리에 꼬리를 무는 그날 이야기 1', '황금어장-라디오스타', ' 태어난 김에 세계일주2-미방송분깊이보기', '태어난 김에 세계일주', '인생술집', '금요일 금요일 밤 에', '푸바오와 할부지 2'], dtype=object), array(['세 번째 결혼', '황금빛 내인생', '선덕여왕'], dtype=object), array(['총각을 찾는 욕구불만 유부녀', '흥건한 제수씨:불타는 밤', '아내의 엄마와 몰 래 절정', '대리 만족-아내를 드립니다', '남편은 이렇게 못해줘', '내가 가져버린 사위는 몸짱', '애육 의 처제', '관리인 녹인 용광로 부인', '후들후들한 친구엄마', '발기부전약으로 폭발정사'], dtype=object), array(['밤에 피는 꽃', '나 혼자산다', '그놈 목소리', '홍천기'], dtype=object), array(['톡 파원 25시', '런닝맨', '2023 런닝맨 게스트 어워즈'], dtype=object), array(['밤에 피는 꽃', '그것이알 고싶다', '어쩌다 어른', 'tvN STORY 어쩌다 어른', '그것이 알고싶다 레전드 50선', '마리아와 여인숙']

데이터 분석

3. 예고편 시청이 실제 시청으로 이어지는 전환율

- 5분 이하의 disp_rtm의 경우)

1. 예고편) 00회, 000회, (예고), (사전구매), (커밍순), (예고편), (예약구매), 비하인드

2. 예외) 0분~6분

- 살림하는 남자들 2 287회
- 황금어장-라디오스타 762회
- 오늘 밤도 그놈을 기다리는 성인물
- (5CHOICE) -스파이더맨: 어크로스 더 유니버스/ 악마들

* 사이언스 인 모션 - 과학 원리를 모션그래픽으로 풀어낸 콘텐츠 (기타/기타)

- 제습의 원리
- 자물쇠의 원리
- 오르골의 원리
- 자외선과 피부보호
- 손난로의 발열원리
- 날개없는 선풍기의 원리
- 바코드의 원리
- 사람에게도 이런 기관이... 진화의 흔적
- 슈퍼맨, 아이언맨 마징가Z가 싸우면 누가 이길까
- 공룡이 많았던 한반도 왜 석유는 없을까?
- 투명 이빨을 드러내는 심해의 암살...

데이터 분석

3. 예고편 시청이 실제 시청으로 이어지는 전환율

- 예고편 비율

```
preview
False    21229
True      1328
Name: count, dtype: int64
```

- 예고편/예고편x VOD 그룹화

```
   subsr  asset_nm  use_tms  strt_dt  preview
0  59891000   서울의 봄      60 2024-02-21 11:15:11    True
1  59892000   같은 씨를 원하는 며느리    60 2024-02-19 21:45:08    True
2  59892000   불륜남에 미쳐버린 음탕한 부인    60 2024-02-21 13:45:36    True
3  59892000   유부녀 제자 욕정의 보충 수업    60 2024-02-21 13:43:25    True
4  59892000   입을 가져다 대는 시아버지    60 2024-02-19 21:42:59    True
...
1008 67344000   꼬리에 꼬리를 무는 그날 이야기    500 2024-03-14 14:43:51    True
1009 67344000   유산으로 남겨준 새엄마    0 2024-02-16 15:59:19    True
1010 67348000   노랑: 죽음의 바다    120 2024-02-21 18:40:54    True
1011 67348000   데드맨    60 2024-03-24 15:47:51    True
1012 67348000   도그데이즈    60 2024-03-24 15:45:07    True
[1013 rows x 5 columns]
```

'subsr', 'asset_nm'이 동일한 'use_tms'를 합친 df1

'strt_dt'를 가장 빠른 일자로 설정한 df2

'subsr'와 'asset_nm' 기준으로 df1, df2병합

#1013편의 예고편 시청 기록

#4797편의 본편 시청 기록

데이터 분석

3. 예고편 시청이 실제 시청으로 이어지는 전환율

- 예고편/예고편X VOD 병합

	subsr	asset_nm	use_tms_true	strt_dt_true	preview_true	use_tms_false	strt_dt_false	preview_false
0	60320000	1947 보스톤	60	2024-03-09 18:20:10	True	165	2024-03-15 22:16:31	False
1	60326000	마이데몬	60	2024-03-26 20:57:11	True	70879	2024-03-26 20:58:20	False
2	60593000	레트리뷰션	60	2024-02-08 06:37:05	True	5294	2024-02-10 04:17:09	False
3	60593000	메가로돈 2	120	2024-01-09 10:26:30	True	6387	2024-01-09 10:29:43	False
5	61173000	스토커	60	2024-03-13 14:27:42	True	8311	2024-03-13 15:09:03	False
...
95	66873000	재벌X형사	60	2024-01-30 21:20:34	True	49878	2024-01-27 21:20:55	False
96	66875000	외계+인 2부	120	2024-02-15 18:50:15	True	10521	2024-02-20 13:31:48	False
98	67262000	꼬리에 꼬리를 무는 그날이야기	5220	2024-01-17 01:18:48	True	9694	2024-01-16 22:31:45	False
99	67334000	재벌X형사	60	2024-02-26 10:10:00	True	3840	2024-02-26 10:13:00	False
100	67344000	꼬리에 꼬리를 무는 그날이야기	500	2024-03-14 14:43:51	True	52711	2024-02-16 13:31:57	False

- 예고편 -> 시청 : 1013rows 에서 78개의 rows (약 7.7% 전환율)

※ use_tms = 0 인 경우 삭제

4. 상세페이지 머무는 시간 – 실제 시청 전환율

- 상세 페이지의 strt_dt – vod 시청페이지의 strt_dt : 상세 페이지에 머문 시간

22448	60593000	한문철의 블랙박스 리뷰	70회(24/03/05)	TV 시사/교인물/다큐	6000	싱어게인3	한문철,이수근,이	1:40	20240311124944
22449	60593000	한문철의 블랙박스 리뷰	70회(24/03/05)	TV 시사/교인물/다큐	4650	싱어게인3	한문철,이수근,이	1:40	20240311190858

▲ VOD 데이터의 asset_nm

49395	60593000	한문철의 블랙박스 리뷰	한문철의 블랙박스 리뷰	70회	TV 시사/교인물/다큐	싱어게인3	한문철,이수근,이	1:40	20240311124859
49396	60593000	한문철의 블랙박스 리뷰	한문철의 블랙박스 리뷰	70회	TV 시사/교인물/다큐	싱어게인3	한문철,이수근,이	1:40	20240311124908
49397	60593000	한문철의 블랙박스 리뷰	한문철의 블랙박스 리뷰	70회	TV 시사/교인물/다큐	싱어게인3	한문철,이수근,이	1:40	20240311111636

▲ contents 데이터의 super_asset_nm

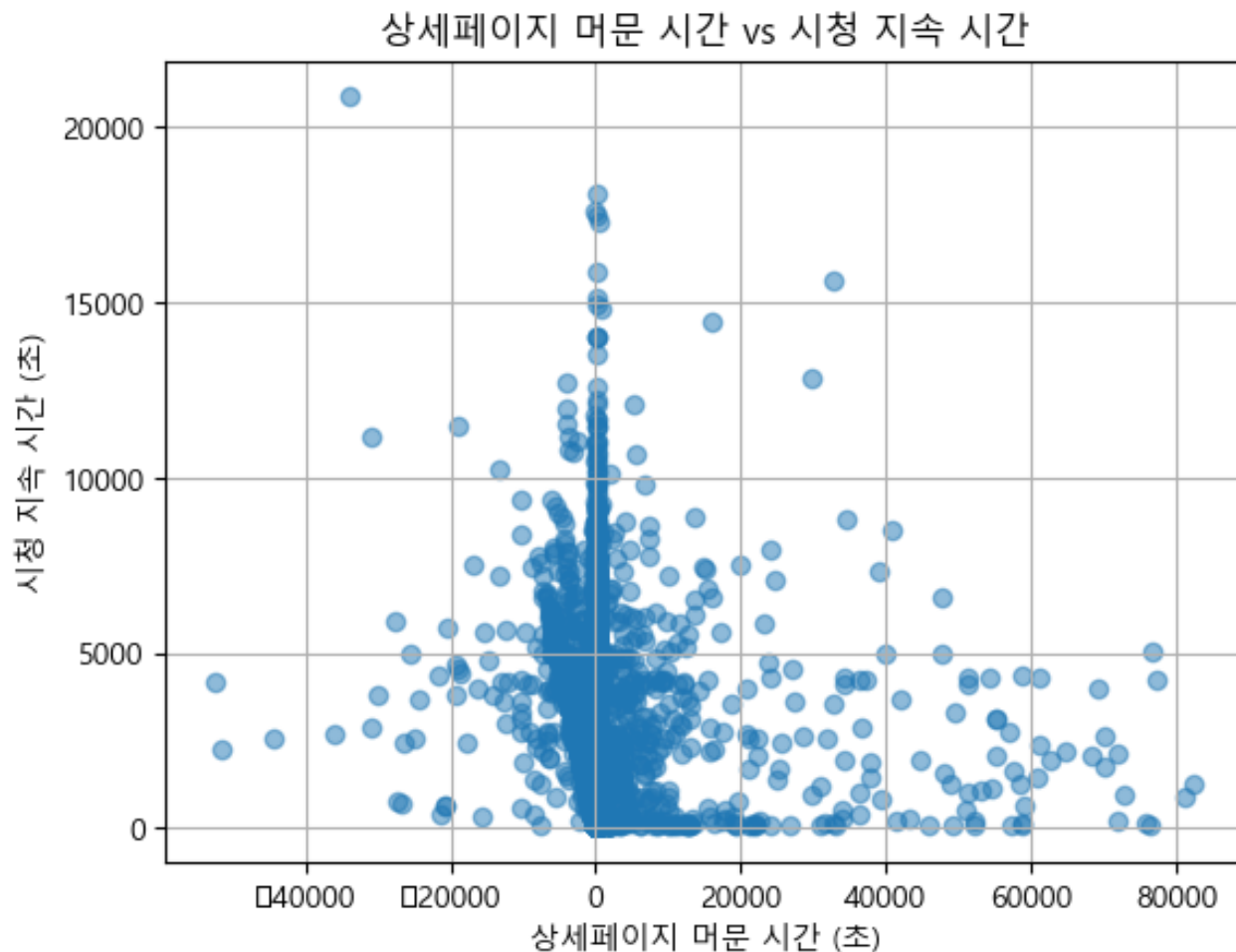
VOD 의 asset_nm 와 contents의 super_asset_nm의 정규화

- ()를 포함한 문자열 제거
- []를 포함한 문자열 제거

데이터 분석

4. 상세페이지 머무는 시간 - 실제 시청 전환율

- 상세 페이지의 strt_dt - vod 시청페이지의 strt_dt : 상세 페이지에 머문 시간



상세페이지에서 머문 시간과
시청 지속 시간 간의 상관 계수
: -0.05292964719142732

4. 상세페이지 머무는 시간 – 실제 시청 전환율

- 상세 페이지의 strt_dt – vod 시청페이지의 strt_dt : 상세 페이지에 머문 시간

```
OLS Regression Results
===== Dep. Variable:
use_tms R-squared: 0.003 Model: OLS Adj. R-squared: 0.003 Method: Least Squares F-statistic: 20.60 Date:
Mon, 29 Apr 2024 Prob (F-statistic): 5.75e-06 Time: 00:54:34 Log-Likelihood: -67093. No. Observations: 7335
AIC: 1.342e+05 Df Residuals: 7333 BIC: 1.342e+05 Df Model: 1 Covariance Type: nonrobust
===== coef
std err t P>|t| [0.025 0.975] ----- const 2048.7697
26.530 77.223 0.000 1996.762 2100.777 time_difference -0.0194 0.004 -4.539 0.000 -0.028 -0.011
===== Omnibus:
2360.675 Durbin-Watson: 1.134 Prob(Omnibus): 0.000 Jarque-Bera (JB): 7930.414 Skew: 1.627 Prob(JB): 0.00
Kurtosis: 6.919 Cond. No. 6.21e+03
===== Notes: [1]
Standard Errors assume that the covariance matrix of the errors is correctly specified. [2] The condition
number is large, 6.21e+03. This might indicate that there are strong multicollinearity or other numerical
problems.
```


데이터 분석

4. 상세페이지 머무는 시간 – 실제 시청 전환율

- 상세 페이지의 strt_dt – vod 시청페이지의 strt_dt : 상세 페이지에 머문 시간

1.R-squared (결정 계수): 0.003으로 매우 낮습니다. 이는 모델이 종속 변수의 변동을 매우 적게 설명한다는 것을 의미합니다. 모델이 상세페이지에서 머문 시간과 시청 지속 시간 간의 변동을 거의 설명하지 못한다는 것을 의미합니다.

2.회귀 계수 (Coefficients):

- 1. const (상수항):** 2048.7697으로, 이는 상세페이지에서 머문 시간이 0일 때의 시청 지속 시간을 나타냅니다. 회귀 모델에 따르면, 상세페이지에서 머문 시간이 0이더라도 시청 지속 시간이 2048.7697으로 나타납니다. 이는 통계적으로 유의합니다 (p-value < 0.05).
- 2. time_difference (상세페이지에서 머문 시간):** -0.0194으로, 이는 상세페이지에서 머문 시간이 증가할수록 시청 지속 시간이 0.0194초씩 감소한다는 것을 의미합니다. 이 계수도 통계적으로 유의합니다 (p-value < 0.05).

3.Prob (F-statistic): 5.75e-06으로 매우 낮습니다. 이는 회귀 모델이 통계적으로 유의하다는 것을 의미합니다. 회귀 분석이 종속 변수에 대한 설명력이 있는 것으로 나타납니다.

4.Durbin-Watson: 1.134로, 오차 항의 자기상관을 나타내는 지표입니다. 2에 가까우면 자기상관이 없다는 것을 나타냅니다. 여기서는 1.134로 자기상관이 거의 없는 것으로 보입니다.

5.Omnibus, Jarque-Bera, Skew, Kurtosis: 잔차의 정규성과 관련된 지표들입니다. Omnibus와 Jarque-Bera 통계량은 잔차의 정규성을 확인하기 위한 것이며, Skew와 Kurtosis는 잔차의 비대칭성과 첨도를 나타냅니다. 이러한 값들이 유의하다면 잔차가 정규분포를 따르지 않는다는 것을 의미합니다.

6.다중공선성 (Multicollinearity): Condition Number가 6.21e+03으로 매우 큼니다. 이는 다중공선성이나 다른 숫자적 문제가 있을 수 있다는 것을 나타냅니다.

데이터 분석

4. 상세페이지 머무는 시간 – 실제 시청 전환율

- 상세 페이지의 strt_dt – vod 시청페이지의 strt_dt : 상세 페이지에 머문 시간
- 상세 페이지 > 실제 시청 전환율로 재확인

5. 의견

- LG헬로비전 데이터 기준 인기 VOD 추천 - 신규 가입자 / 기존 가입자 추천 구분 필요

- 영화

- 1) 인기있는 배우/ 출연자(ACTR_DISP) 가 출연한 출연작 기준 추천
- 2) 인기있는 영화(LG헬로비전 데이터 중 시청 횟수가 가장 많은 작품)

-----신규 가입자

- 1) 예고편 시청한 경우 추천
- 2) Contents/channel에 머무른 시간이 긴 프로그램 추천

-----기존 가입자는 해당 추천 추가

- 시리즈물

- 1) TV드라마의 - 인기있는 배우/ 출연자(ACTR_DISP) 가 출연한 출연작 기준 추천
- 2) TV연예/오락 - 인기있는 제작진(tmdb), 게스트(네이버 데이터 활용)
- 3) 인기있는 프로그램(LG헬로비전 데이터 중 시청 횟수가 가장 많은 작품)

-----신규 가입자

- 1) 기본적으로 장르 중 대분류 시청 비율 기준 추천(영화/키즈/성인 제외)
- 2) 대분류 중 TV드라마/ TV 연예/오락의 경우 추가적인 장르 분류를 통한 추천(카테고리화/라벨링 필요)

ex) 음악/ 음식 / 유머 / 스포츠

- 3) 예고편 시청한 경우 추천
- 4) Contents/channel에 머무른 시간이 긴 프로그램 추천

-----기존 가입자는 해당 추천 추가