빅데이터 저장 및 관리

인터넷버즈 DB 및 방송프로그램별 (VOD, 실시간) 시청시간, 시청률 데이터 분석을 통한 최신 방송 컨텐츠의 트랜드 분석

2021.12.14

* 산업정보시스템전공 박사과정 공용택
* 데이터사이언스학과 석사과정 김재호

주제선정 배경 (1/3)

※ 주제명 : 인터넷버즈 DB 및 방송프로그램별 (VOD, 실시간) 시청시간, 시청률 데이터 분석을 통한 최신 방송 컨텐츠의 트랜드 분석

[IT돋보기] "케이블 적고 IPTV 많다"…시청률 조사 표본·기업 쏠림 '심각'

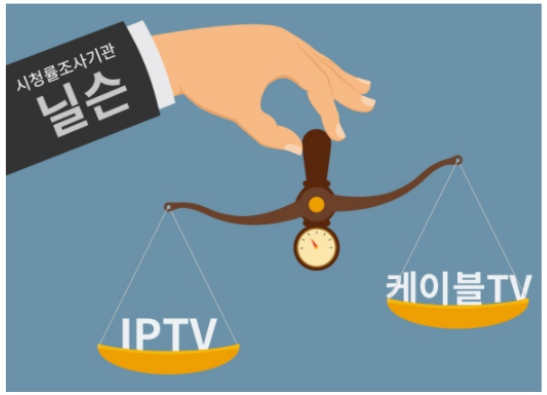
심지혜 기자 입력 2021.10.18 10:04 | 수정 2021.10.18 10:20 댓글 쓰기

조사 업체 독점상황 …표본 패널 구성 문제로 결과 왜곡

[아이뉴스24 심지혜 기자] 현행 방송 시청률 조사 방식을 두고 업계 불만이 상당하다.

시청률 조사를 위한 패널이 각 유료방송 매체를 대표할 수 있도록 구성돼야 함에도 치우치게 선정됐다는 것. 현실을 반영하지 않아 왜곡이 발생하면서 그에 따른 정당성에 흠결이 날 수 있다는 주장이다.

시청률 조사는 표본조사의 형식을 취하고 있고 시청률 조사에 참여하는 표본 가구를 '패널'이라고 하는데, 이 패널이 각 플랫폼, 지역, 연령, 성별 등을 대표하지 못하면 결과가 왜곡될 수 있어 검증 작업이 필요하다는 분석이다.



현행 방송 시청률 조사 방식을 두고 업계 불만이 상당하다. [사진=그래픽=조은수 기자]

정보통신신문 '시청률 데이터 신뢰성 문제' 심화...케이블TV방송협회, 공론화

HOME > 뉴스 > 통신방송

'시청률 데이터 신뢰성 문제' 심화...케이블TV방송협회, 공론화

박남수 기자 | 승인 2021.10.10 20:17 | 댓글 0

미국도 시청환경 변화에 따른
기존 방식에 제동 대안 찾기 몰두

[정보통신신문=박남수기자]

디지털 온라인 매체가 늘어나고 시청형태도 다양해지는 등 시청환경이 급격히 변화하고 있다.

하지만 시청률조사 방법은 급변하는 미디어 환경에 뒤쳐질 뿐 아니라 이로 인해 산업의 피해와 시장의 왜곡 현상이 심화 되어 방송업계의 불만이 가중되고 있다.

시청률 조사 업체 한 곳이 사실상 독점하고 있는 시장에서 이 업체의 패널구성의 왜곡에 따른 문제점을 더 이상 방치해서는 안 된다는 문제를 직시하고 케이블TV방송협회와 학계는 10월 중 세미나를 통해 문제점을 공론화할 예정이다.

현재 국내 시청률 조사기관은 민간사업자인 AGB닐슨 코리아가 시장의 95%이상을 점유해 사실상 독점 형태를 취하고 있는데 산출 데이터에 대한 규제기관 및 검증기구의 부재가 문제로 꼽히고 있다.

또한 무엇보다 케이블업계가 꾸준히 제기하고 있는 패널 구성이 왜곡된 채 시장에서 자료가 통용되고 있다는 것도 심각한 상황이다.

시청률 조사 결과에 대한 문제 제기는 비단 우리나라만의 갈등도 아니다.

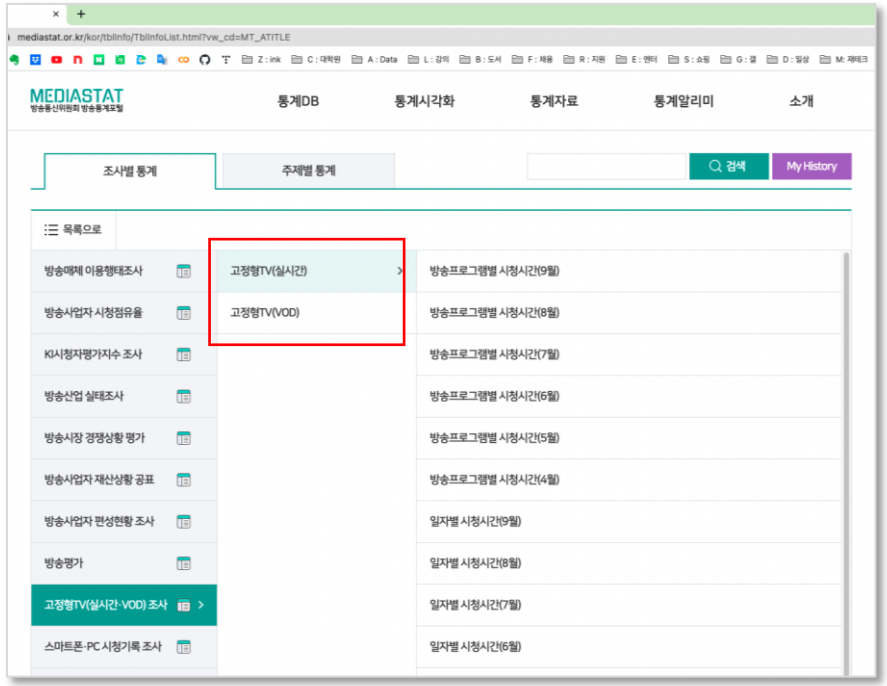
최근 미국에서도 콘텐츠 제작사와 방송사로 구성된 광고관련 기구인 VAB(Video Advertising Bureau)가 시청률위원회에 닐슨의 시청률 조사 방법에 대한 문제를 제기한 바 있다.

이에 미국 시청률검증위원회는 닐슨에 대한 인증을 중지하라는 요청을 정식 수용해 미디어시청률위원회(MRC)가 지난 9월 1일자로 인증을 중지하는 충격적인 사건이 발생했다. 이것은 유례없는 사태로 방송 업계가 시청률 데이터에 대한 신뢰를 거두었다는 의미이다.

- ✓ 디지털 온라인 매체 증가에 따른 시청률 조사 데이터의 신뢰성이 의심 받고 있음
- ✓ 또한 시청률 조사방법은 급변하는 미디어 환경에 뒤쳐질 뿐만 아니라 시장의 왜곡 현상도 초래하여 방송 업계의 불만이 가중되고 있음

주제선정 배경 (2/3)

✓ 방송통신위원회 방송통계포털 고정형TV(실시간,VOD) 조사 Data



방송프로그램별	2021. 09								
	장르	채널	개인 전체_시청시간 (분)	개인 전체_시청률 (%)	4-9세 남자_시청시간 (분)	4-9세 남자_시청률 (%)	4-9세 여자_시청시간 (분)	4-9세 여자_시청률 (%)	10대 남 간 (분
KBS뉴스9	보도	KBS1	76.299	5.023	36.066	2.374	28.629	1.885	34.898
일일드라마<빨강구두>	드라마&영화	KBS2	60.292	8.409	23.412	3.265	25.268	3.524	21.526
주말드라마<오케이광자매>	드라마&영화	KBS2	53.445	14.484	22.629	6.133	34.032	9.223	14.896
6시내고향	정보	KBS1	48.507	3.799	18.844	1.476	10.065	0.788	12.327
아침마당	정보	KBS1	45.961	3.316	17.236	1.244	6.232	0.450	10.020
일일드라마<속아도꿈결>	드라마&영화	KBS1	43.886	7.124	15.044	2.442	17.001	2.760	21.942
KBS뉴스광장 2부	보도	KBS1	39.370	3.463	24.166	2.125	4.482	0.394	6.093
MBC뉴스데스크	보도	MBC	38.912	2.574	23.903	1.581	54.090	3.577	21.672
2TV생생정보	정보	KBS2	38.874	2.884	22.845	1.695	20.600	1.528	15.289
갯마을차차차(본)	드라마&영화	tvN	34.032	5.368	16.719	2.637	14.934	2.356	21.404
인간극장(KBS1)	정보	KBS1	32.425	4.580	16.526	2.334	4.001	0.565	6.266
미운우리새끼	오락	SBS	27.184	6.488	12.385	2.956	13.706	3.271	15.173
KBS뉴스광장 1부	보도	KBS1	26.910	2.016	6.590	0.494	2.353	0.176	3.008
TV조선뉴스9	보도	TV조선	25.513	2.194	13.049	1.122	1.934	0.166	3.678
SBS8시뉴스	보도	SBS	25.107	1.693	16.647	1.122	20.435	1.378	9.043
MBN뉴스와이드	보도	MBN	23.357	1.067	3.046	0.139	1.617	0.074	1.562
일일아침연속극<아모르파티>	드라마&영화	SBS	23.263	3.441	13.748	2.034	7.742	1.145	9.138
갯마을차차차(재)	드라마&영화	tvN	21.368	0.698	18.513	0.605	13.899	0.454	7.360
MBN종합뉴스	보도	MBN	21.313	1.331	3.400	0.212	3.078	0.192	5.624
KBS뉴스12	보도	KBS1	20.523	1.763	7.835	0.673	3.896	0.335	5.664
슬기로운의사생활시즌2(본)	드라마&영화	tvN	20.264	8.373	15.300	6.322	18.956	7.833	13.922
뉴스파이터	보도	MBN	19.812	1.469	2.090	0.155	1.977	0.147	2.010
모닝와이드2부	정보	SBS	19.675	1.293	3.047	0.200	9.480	0.623	1.612

주제선정 배경 (3/3)

✓ 따라서 기존의 시청률 데이터를 보완하기 위한 미디어 뉴스(수), 동영상(수), 게시글(수), 댓글(수) 등을 조사한 인터넷 반응 (시청자버즈 + 미디어버즈) 데이터 활용이 필요

인터넷반응 | 조사대상 및 지표정의

■ 조사대상

29개 채널(지상파 5+종편PP 4+일반PP 20)에서 방송되는 드라마 및 예능

※해외콘텐츠/뉴스/스포츠중계/어린이프로그램은 제외, 다른채널에서 재방송된 경우에는 최초 방송채널을 기준으로 집계.

※드라마 전수조사

예능:프라임타임(평일20~23시 및 주말 19~23시)방송작

정보:프라임 시간대 방송되는 프로그램 (평일 20시~23시 / 주말 19시~23시) / 1회성 프로그램은 등록 제외 / 다큐프로그램은 3회 이상 방송되는 경우 등록 / 방송국 홈페이지에서 등록되지 않은 프로그램은 등록 제외

*표시:유의미한 버즈량을 보이는 프로그램을 추가조사함

■ 발표주기 주간

■ 지표정의

유형	정의	지표	수집사이트
미디어 버즈	방송프로그램에 대한 언론사 · 방송사 반응	뉴스(수)	네이버 뉴스, 다음 뉴스, 네이트 뉴스
		동영상(수)	네이버TV, 카카오TV
시청자 버즈	방송프로그램에 대한 시청자 반응	게시글(수)	네이버 블로그, 네이버카페, 다음카페, 오늘의유머, 인스티즈, 웃긴대학, 뽐뿌, MBL Park,DC Inside
		댓글(수)	수집된 뉴스/게시글/동영상의 댓글
		동영상(조회수)	네이버TV, 유튜브, 카카오TV

RACOI 방송통신위원회 방송콘텐츠 가치정보분석시스템 | 인터넷반응 | 종합반응 | 해외반응 | RACOI | 검색

종합반응 > 종합반응DB (이용가이드)

기간: 2021년 09월

드라마: 전체 | 월·화 | 수·목 | 금·토 | 토·일 | 일일아침 | 일일저녁

예능/정보: 전체 | 월 | 화 | 수 | 목 | 금 | 토 | 일

채널: 전체 | 지상파 | 종편PP | 일반PP | 채널선택

《 2021년 09월 (2021.09.01~2021.09.30) 》 데이터다운로드 | EXCEL다운

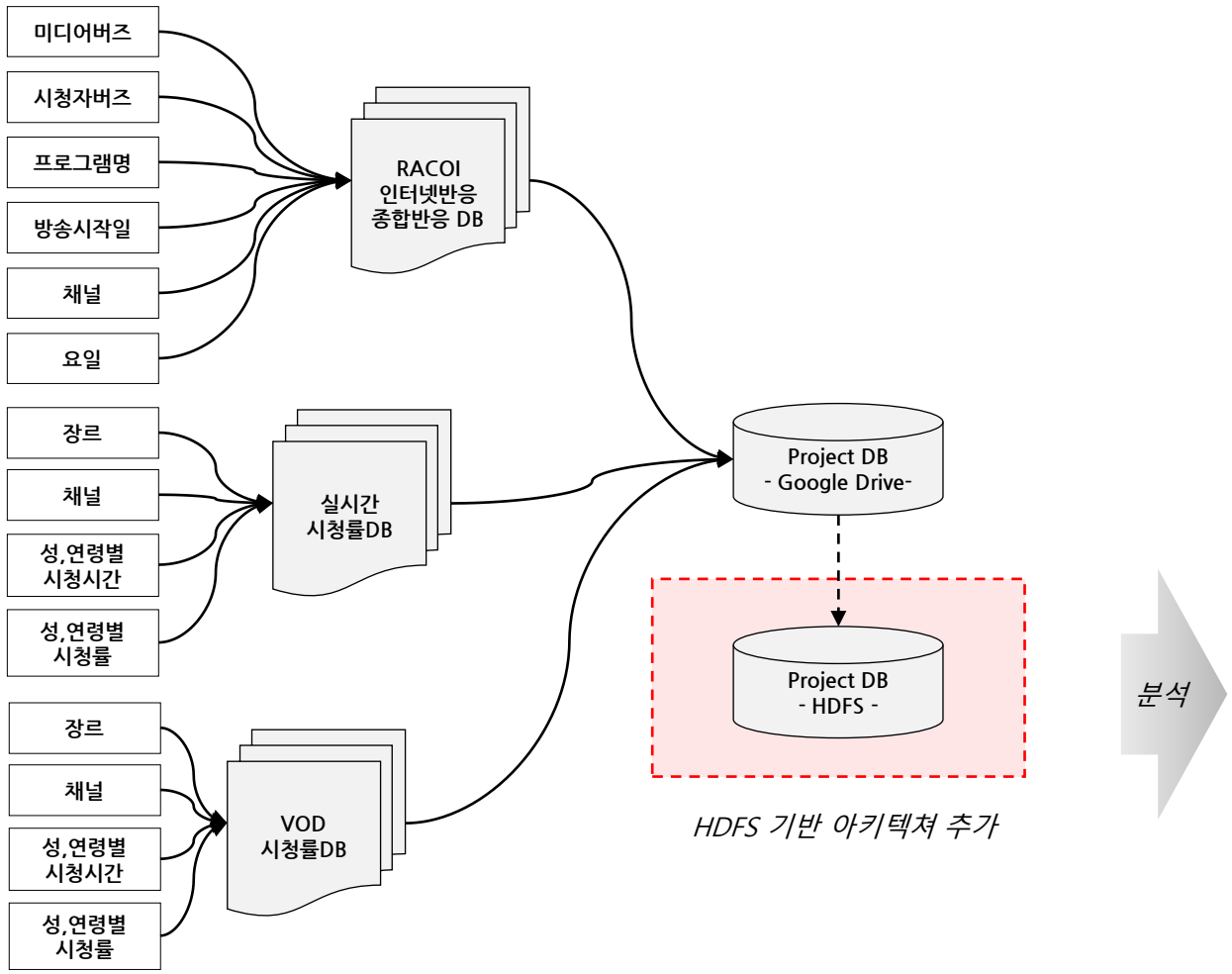
프로그램명	방송시작일	채널	요일	인터넷반응					TV									
				시청자버즈			미디어버즈		실시간 시청자수보기 (k) | 시청률보기 (%)									
				게시글	댓글	동영상조회(k)	뉴스	동영상	가구	개인	남성	여성	10대	20대	30대	40대	50대	60대+
스트릿 우먼 파이터	2021.08.24	Mnet	화	31,057	355,162	49,392	4,015	468	397	614	244	370	42	109	139	92	113	70
갯마을 차차차	2021.08.28	tvN	토일	17,801	127,581	17,935	3,506	548	1,625	2,830	1,211	1,619	197	240	337	692	794	488
슬기로운 의사생활2(종영)	2021.06.17	tvN	수목	9,525	62,833	11,083	2,353	116	2,482	4,371	1,818	2,553	295	457	622	1,172	1,103	497
걸스플래닛999 : 소녀...	2021.08.06	Mnet	금	7,757	89,362	11,166	817	297	138	318	122	196	87	29	55	73	50	15
펜트하우스3(종영)	2021.06.04	SBS	금	7,453	54,141	11,278	2,242	80	3,298	5,404	2,252	3,152	343	473	630	947	1,593	1,188
검은 태양	2021.09.17	MBC	금	6,760	30,311	5,780	3,469	154	1,464	2,238	1,051	1,186	111	146	245	559	586	522
홍천기	2021.08.30	SBS	월화	6,556	30,419	7,632	3,255	215	1,600	2,419	1,072	1,347	126	145	193	621	676	600

RACOI 방송통신위원회 방송컨텐츠 가치정보분석시스템 종합반응DB

데이터 수집모델 및 분석목표

✓ 인터넷반응 DB. 실시간 시청률 DB, VOD 시청률 DB, 프로그램 출연진 정보 등을 통한 분석용 DB 구축을 선행

✓ 각 수집된 DB를 활용한 최신 방송 트랜드 및 출연진 선호도 등의 분석을 진행할 계획



PROPOSAL 평가의견 (점수 : 8)

- 연구동기와 배경을 명확하고 구체적으로 제시함
- 수집 데이터에 대한 샘플을 구체적으로 제시하였으며, 다양한 소스로부터 수집한 데이터로부터 의미를 도출함
- 기존 유사 서비스와의 비교 및 차별화가 필요함. 빅데이터 기반 분석을 수행하기 위한 Hadoop 기반 아키텍쳐 설계 및 필요 시스템의 구성에 대한 제시가 필요함
- 최종 발표 시에는 중간 발표 및 최종 결과물을 포괄한 보고서 제출이 필요

프로젝트 목표

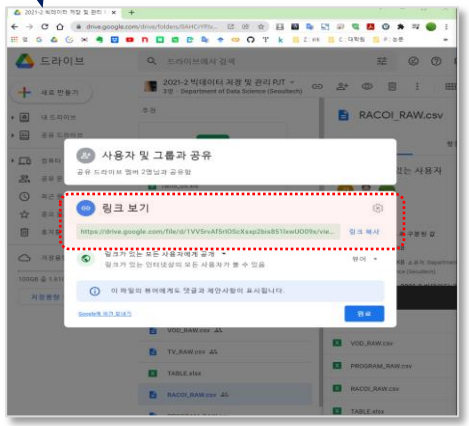
- <u>HDFS 기반 아키텍쳐 설계</u>
- 성, 연령별 미디어버즈, 시청자버즈 분석 및 시각화
- 성, 연령, 장르별 실시간 시청률, 시청시간 분석 및 시각화
- 성, 연령, 장르별 VOD 시청률, 시청시간 분석 및 시각화
- 방송 프로그램별 시청자 선호도 및 동향 분석
- 방송 프로그램별 출연진 선호도에 따른 시청률 및 버즈데이터 분석 등

HDFS 기반 아키텍쳐 설계

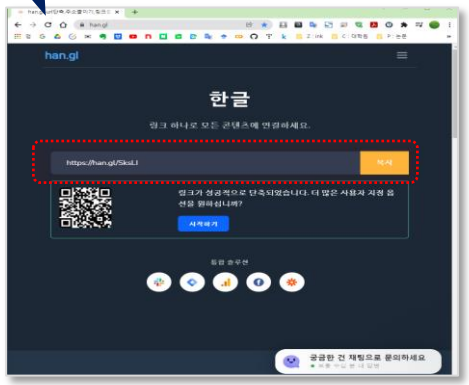
✓ HDFS 기반 아키텍쳐 설계를 위한 분석용 RAW DATA 준비

✓ HOST 로컬 드라이브 공유시 오류가 발생하여 구글 드라이브 링크를 통하여 가상머신 서버로 파일 공유를 진행함

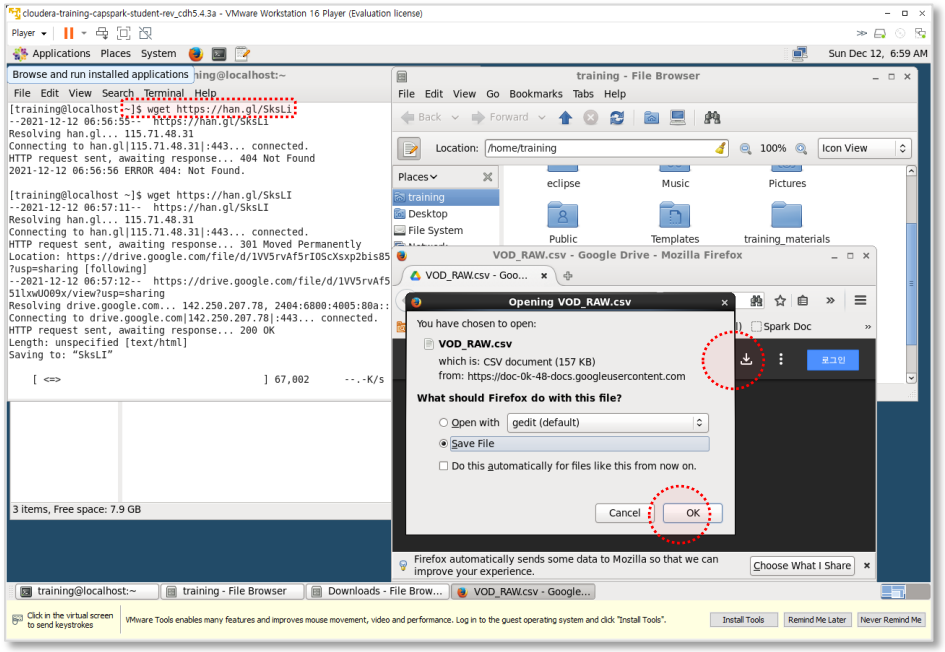
① 구글 드라이브 공유 링크 생성



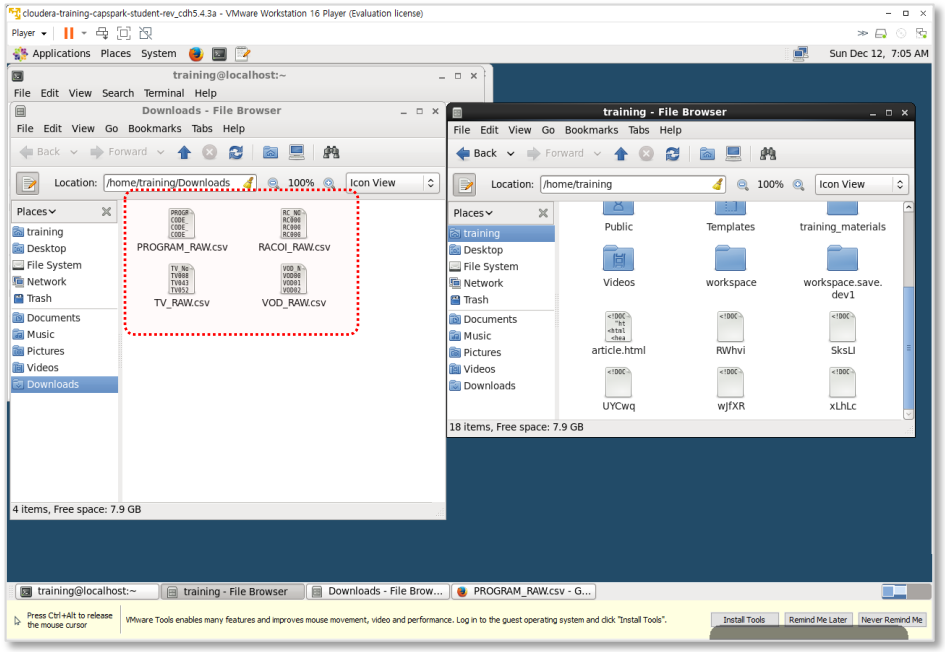
② URL 링크 단축



③ wget URL주소로 파일 다운



④ 분석용 RAW DATA 준비



Preprocessing ① 변수명 정의 # 컬럼정의서 (1/3)

✓ 데이터 분석을 위한 한글 변수명을 영문 변수로 재정의

▶ RACOI Data Set (2021.05 ~ 2021.10 x6회분)

조사 년,월		
프로그램코드명		
프로그램명		
방송시작일		
채널		
요일		
인터넷 버즈	시청자 버즈	게시글
		댓글
		동영상조회(K)
	미디어 버즈	뉴스
		동영상
실시간 시청자수 (K)		가구
		개인
		남성
		여성
		10대
		20대
		30대
		40대
		50대

RC_MONTH	object
PROGRAM_CODE	object
RC_PROGRAM_NAME	object
RC_PROGRAM_STR_DATE	object
RC_CHNNEL_NM	object
RC_PROGRAM_DAYS	object
RC_BUZZ_VIWER_POST	int64
RC_BUZZ_VIWER_COMMENT	int64
RC_BUZZ_VIWER_VIDEO	int64
RC_BUZZ_MEDIA_NEWS	int64
RC_BUZZ_MEDIA_VIDEO	int64
RC_LIVE_FAMILY_VIEWRS	float64
RC_LIVE_INDIVIDUAL_VIEWRS	float64
RC_LIVE_MALE_VIEWRS	float64
RC_LIVE_FEMALE_VIEWRS	float64
RC_LIVE_10S_VIEWRS	float64
RC_LIVE_20S_VIEWRS	float64
RC_LIVE_30S_VIEWRS	float64
RC_LIVE_40S_VIEWRS	float64
RC_LIVE_50S_VIEWRS	Float64

```
racoi_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1292 entries, 0 to 1291
Data columns (total 20 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   RC_MONTH                   1292 non-null   object
 1   RC_PROGRAM_NAME            1292 non-null   object
 2   PROGRAM_CODE               1292 non-null   object
 3   RC_PROGRAM_STR_DATE        1262 non-null   object
 4   RC_CHNNEL_NM               1292 non-null   object
 5   RC_PROGRAM_DAYS            1292 non-null   object
 6   RC_BUZZ_VIWER_POST         1292 non-null   int64
 7   RC_BUZZ_VIWER_COMMENT      1292 non-null   int64
 8   RC_BUZZ_VIWER_VIDEO        1292 non-null   int64
 9   RC_BUZZ_MEDIA_NEWS         1292 non-null   int64
 10  RC_BUZZ_MEDIA_VIDEO        1292 non-null   int64
 11  RC_LIVE_FAMILY_VIEWRS      1182 non-null   float64
 12  RC_LIVE_INDIVIDUAL_VIEWRS  1182 non-null   float64
 13  RC_LIVE_MALE_VIEWRS        1182 non-null   float64
 14  RC_LIVE_FEMALE_VIEWRS      1182 non-null   float64
 15  RC_LIVE_10S_VIEWRS         1162 non-null   float64
 16  RC_LIVE_20S_VIEWRS         1174 non-null   float64
 17  RC_LIVE_30S_VIEWRS         1180 non-null   float64
 18  RC_LIVE_40S_VIEWRS         1180 non-null   float64
 19  RC_LIVE_50S_VIEWRS         1182 non-null   float64
dtypes: float64(9), int64(5), object(6)
memory usage: 202.0+ KB
```

Preprocessing ① 변수명 정의 # 컬럼정의서 (2/3)

▶ **고정형TV 실시간 시청률, 시청시간 Data Set (2021.05 ~ 2021.10 x6회분)**

프로그램명	TV_PROGRAM_NAME
장르	TV_PROGRAM_GENRE
채널	TV_CHNNEL_NM
전체_시청시간 (분)	TV_ALL_WTCHNG_MIN
전체_시청률 (%)	TV_ALL_WTCHNG_RT
4-9세 남자_시청시간 (분)	TV_MALE_4_9YO_WTCHNG_MIN
4-9세 남자_시청률 (%)	TV_MALE_4_9YO_WTCHNG_RT
4-9세 여자_시청시간 (분)	TV_FEMALE_4_9YO_WTCHNG_MIN
4-9세 여자_시청률 (%)	TV_FEMALE_4_9YO_WTCHNG_RT
10대 남자_시청시간 (분)	TV_MALE_N10S_WTCHNG_MIN
10대 남자_시청률 (%)	TV_MALE_N10S_WTCHNG_RT
10대 여자_시청시간 (분)	TV_FEMALE_N10S_WTCHNG_MIN
10대 여자_시청률 (%)	TV_FEMALE_N10S_WTCHNG_RT
20대 남자_시청시간 (분)	TV_MALE_N20S_WTCHNG_MIN
20대 남자_시청률 (%)	TV_MALE_N20S_WTCHNG_RT
20대 여자_시청시간 (분)	TV_FEMALE_N20S_WTCHNG_MIN
20대 여자_시청률 (%)	TV_FEMALE_N20S_WTCHNG_RT
30대 남자_시청시간 (분)	TV_MALE_N30S_WTCHNG_MIN
30대 남자_시청률 (%)	TV_MALE_N30S_WTCHNG_RT
30대 여자_시청시간 (분)	TV_FEMALE_N30S_WTCHNG_MIN
30대 여자_시청률 (%)	TV_FEMALE_N30S_WTCHNG_RT
40대 남자_시청시간 (분)	TV_MALE_N40S_WTCHNG_MIN
40대 남자_시청률 (%)	TV_MALE_N40S_WTCHNG_RT
40대 여자_시청시간 (분)	TV_FEMALE_N40S_WTCHNG_MIN
40대 여자_시청률 (%)	TV_FEMALE_N40S_WTCHNG_RT
50대 남자_시청시간 (분)	TV_MALE_N50S_WTCHNG_MIN
50대 남자_시청률 (%)	TV_MALE_N50S_WTCHNG_RT
50대 여자_시청시간 (분)	TV_FEMALE_N50S_WTCHNG_MIN
50대 여자_시청률 (%)	TV_FEMALE_N50S_WTCHNG_RT
60세 이상 남자_시청시간 (분)	TV_MALE_N60S_WTCHNG_MIN
60세 이상 남자_시청률 (%)	TV_MALE_N60S_WTCHNG_RT
60세 이상 여자_시청시간 (분)	TV_FEMALE_N60S_WTCHNG_MIN
60세 이상 여자_시청률 (%)	TV_FEMALE_N60S_WTCHNG_RT

```
tv_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 251 entries, 0 to 250
Data columns (total 36 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   TV_No                       251 non-null    object
 1   TV_MONTH                    251 non-null    object
 2   PROGRAM_CODE                251 non-null    object
 3   TV_PROGRAM_NAME             251 non-null    object
 4   TV_PROGRAM_GENRE            251 non-null    object
 5   TV_CHNNEL_NM                251 non-null    object
 6   TV_ALL_WTCHNG_MIN           251 non-null    float64
 7   TV_ALL_WTCHNG_RT            251 non-null    float64
 8   TV_MALE_4_9YO_WTCHNG_MIN    251 non-null    float64
 9   TV_MALE_4_9YO_WTCHNG_RT     251 non-null    float64
 10  TV_FEMALE_4_9YO_WTCHNG_MIN  251 non-null    float64
 11  TV_FEMALE_4_9YO_WTCHNG_RT   251 non-null    float64
 12  TV_MALE_N10S_WTCHNG_MIN     251 non-null    float64
 13  TV_MALE_N10S_WTCHNG_RT      251 non-null    float64
 14  TV_FEMALE_N10S_WTCHNG_MIN   251 non-null    float64
 15  TV_FEMALE_N10S_WTCHNG_RT    251 non-null    float64
 16  TV_MALE_N20S_WTCHNG_MIN     251 non-null    float64
 17  TV_MALE_N20S_WTCHNG_RT      251 non-null    float64
 18  TV_FEMALE_N20S_WTCHNG_MIN   251 non-null    float64
 19  TV_FEMALE_N20S_WTCHNG_RT    251 non-null    float64
 20  TV_MALE_N30S_WTCHNG_MIN     251 non-null    float64
 21  TV_MALE_N30S_WTCHNG_RT      251 non-null    float64
 22  TV_FEMALE_N30S_WTCHNG_MIN   251 non-null    float64
 23  TV_FEMALE_N30S_WTCHNG_RT    251 non-null    float64
 24  TV_MALE_N40S_WTCHNG_MIN     251 non-null    float64
 25  TV_MALE_N40S_WTCHNG_RT      251 non-null    float64
 26  TV_FEMALE_N40S_WTCHNG_MIN   251 non-null    float64
 27  TV_FEMALE_N40S_WTCHNG_RT    251 non-null    float64
 28  TV_MALE_N50S_WTCHNG_MIN     251 non-null    float64
 29  TV_MALE_N50S_WTCHNG_RT      251 non-null    float64
 30  TV_FEMALE_N50S_WTCHNG_MIN   251 non-null    float64
 31  TV_FEMALE_N50S_WTCHNG_RT    251 non-null    float64
 32  TV_MALE_N60S_WTCHNG_MIN     251 non-null    float64
 33  TV_MALE_N60S_WTCHNG_RT      251 non-null    float64
 34  TV_FEMALE_N60S_WTCHNG_MIN   251 non-null    float64
 35  TV_FEMALE_N60S_WTCHNG_RT    251 non-null    float64
dtypes: float64(30), object(6)
memory usage: 70.7+ KB
```

Preprocessing ① 변수명 정의 # 컬럼정의서 (3/3)

▶ **VOD 시청률, 시청시간 Data Set (2021.05 ~ 2021.10 x6회분)**

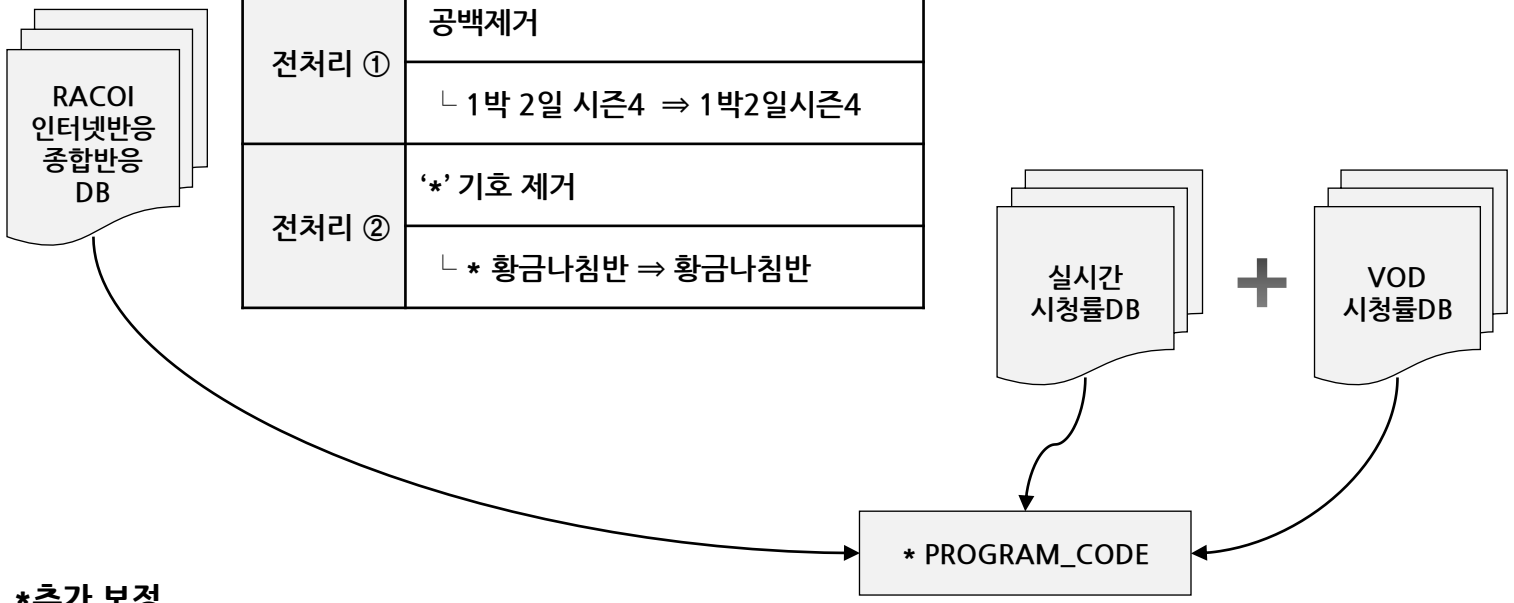
프로그램명	VOD_PROGRAM_NAME
장르	VOD_PROGRAM_GENRE
채널	VOD_CHNNEL_NM
가시청_가구_시청시간 (분)	VOD_FAMILY_WTCHNG_MIN
가시청_가구_시청률 (%)	VOD_FAMILY_WTCHNG_RT
가시청_개인_시청시간 (분)	VOD_INDIVIDUAL_WTCHNG_MIN
가시청_개인_시청률 (%)	VOD_INDIVIDUAL_WTCHNG_RT
남자_시청시간 (분)	VOD_MALE_WTCHNG_MIN
남자_시청률 (%)	VOD_MALE_WTCHNG_RT
여자_시청시간 (분)	VOD_FEMALE_WTCHNG_MIN
여자_시청률 (%)	VOD_FEMALE_WTCHNG_RT
4-9세_시청시간 (분)	VOD_4_9YO_WTCHNG_MIN
4-9세_시청률 (%)	VOD_4_9YO_WTCHNG_RT
10대_시청시간 (분)	VOD_N10S_WTCHNG_MIN
10대_시청률 (%)	VOD_N10S_WTCHNG_RT
20대_시청시간 (분)	VOD_N20S_WTCHNG_MIN
20대_시청률 (%)	VOD_N20S_WTCHNG_RT
30대_시청시간 (분)	VOD_N30S_WTCHNG_MIN
30대_시청률 (%)	VOD_N30S_WTCHNG_RT
40대_시청시간 (분)	VOD_N40S_WTCHNG_MIN
40대_시청률 (%)	VOD_N40S_WTCHNG_RT
50대_시청시간 (분)	VOD_N50S_WTCHNG_MIN
50대_시청률 (%)	VOD_N50S_WTCHNG_RT
60대 이상_시청시간 (분)	VOD_N60S_WTCHNG_MIN
60대 이상_시청률 (%)	VOD_N60S_WTCHNG_RT

4-9남_시청시간 (분)	VOD_MALE_4_9YO_WTCHNG_MIN
4-9남_시청률 (%)	VOD_MALE_4_9YO_WTCHNG_RT
4-9여_시청시간 (분)	VOD_FEMALE_4_9YO_WTCHNG_MIN
4-9여_시청률 (%)	VOD_FEMALE_4_9YO_WTCHNG_RT
10대남_시청시간 (분)	VOD_MALE_N10S_WTCHNG_MIN
10대남_시청률 (%)	VOD_MALE_N10S_WTCHNG_RT
10대여_시청시간 (분)	VOD_FEMALE_N10S_WTCHNG_MIN
10대여_시청률 (%)	VOD_FEMALE_N10S_WTCHNG_RT
20대남_시청시간 (분)	VOD_MALE_N20S_WTCHNG_MIN
20대남_시청률 (%)	VOD_MALE_N20S_WTCHNG_RT
20대여_시청시간 (분)	VOD_FEMALE_N20S_WTCHNG_MIN
20대여_시청률 (%)	VOD_FEMALE_N20S_WTCHNG_RT
30대남_시청시간 (분)	VOD_MALE_N30S_WTCHNG_MIN
30대남_시청률 (%)	VOD_MALE_N30S_WTCHNG_RT
30대여_시청시간 (분)	VOD_FEMALE_N30S_WTCHNG_MIN
30대여_시청률 (%)	VOD_FEMALE_N30S_WTCHNG_RT
40대남_시청시간 (분)	VOD_MALE_N40S_WTCHNG_MIN
40대남_시청률 (%)	VOD_MALE_N40S_WTCHNG_RT
40대여_시청시간 (분)	VOD_FEMALE_N40S_WTCHNG_MIN
40대여_시청률 (%)	VOD_FEMALE_N40S_WTCHNG_RT
50대남_시청시간 (분)	VOD_MALE_N50S_WTCHNG_MIN
50대남_시청률 (%)	VOD_MALE_N50S_WTCHNG_RT
50대여_시청시간 (분)	VOD_FEMALE_N50S_WTCHNG_MIN
50대여_시청률 (%)	VOD_FEMALE_N50S_WTCHNG_RT
60대이상남_시청시간 (분)	VOD_MALE_N60S_WTCHNG_MIN
60대이상남_시청률 (%)	VOD_MALE_N60S_WTCHNG_RT
60대이상여_시청시간 (분)	VOD_FEMALE_N60S_WTCHNG_MIN
60대이상여_시청률 (%)	VOD_FEMALE_N60S_WTCHNG_RT

Preprocessing ② 프로그램명 Key 코드 변수 생성

✓ 조시 기관별 동일 프로그램의 제목이 상이하게 표기되어 있어서, Data Set을 연결해줄 수 있는 PROGRAM_CODE를 KEY코드로 생성하였음

✓ 전처리 후에도 매칭되지 않는 프로그램명은 별도의 수작업 검증을 통하여 처리하였음



전처리 ①	공백제거
	└ 1박 2일 시즌4 ⇒ 1박2일시즌4
전처리 ②	'*' 기호 제거
	└ * 황금나침반 ⇒ 황금나침반

전처리 ①	프로그램 장르 ('보도', '스포츠') 제거
	└ KBS뉴스9, 2020도쿄올림픽(생) ⇒ X
전처리 ②	재방송 data 제거 right(text,3) = (재)
	└ 당신이바로보이스킹(재) ⇒ X
전처리 ③	본방송 '(본) ' 표시 제거
	└ 뽕숭아학당(본) ⇒ 뽕숭아학당
전처리 ④	'드라마요일', '<>' 표시 제거
	└ 주말드라마<오케이광자매> ⇒ 오케이광자매

***추가 보정**

RACOI DB		시청률 DB
LOUD	↔	라우드
보이스킹	↔	당신이바로보이스킹
라디오스타	↔	황금어장라디오스타
아모르파티-사랑하라,지금	↔	아모르파티

RACOI BUZZ 데이터셋 분석 (1/2)

✓ 각 변수들을 프로그램 월별 평균값으로 Groupby() 함수를 이용하여 mean값으로 연산하였으며, 각 변수들 통계량 분석 및 시각화 진행

```
In [23]: # racoi_df 데이터 확인
         racoi.head(10)
```

Out[23]:

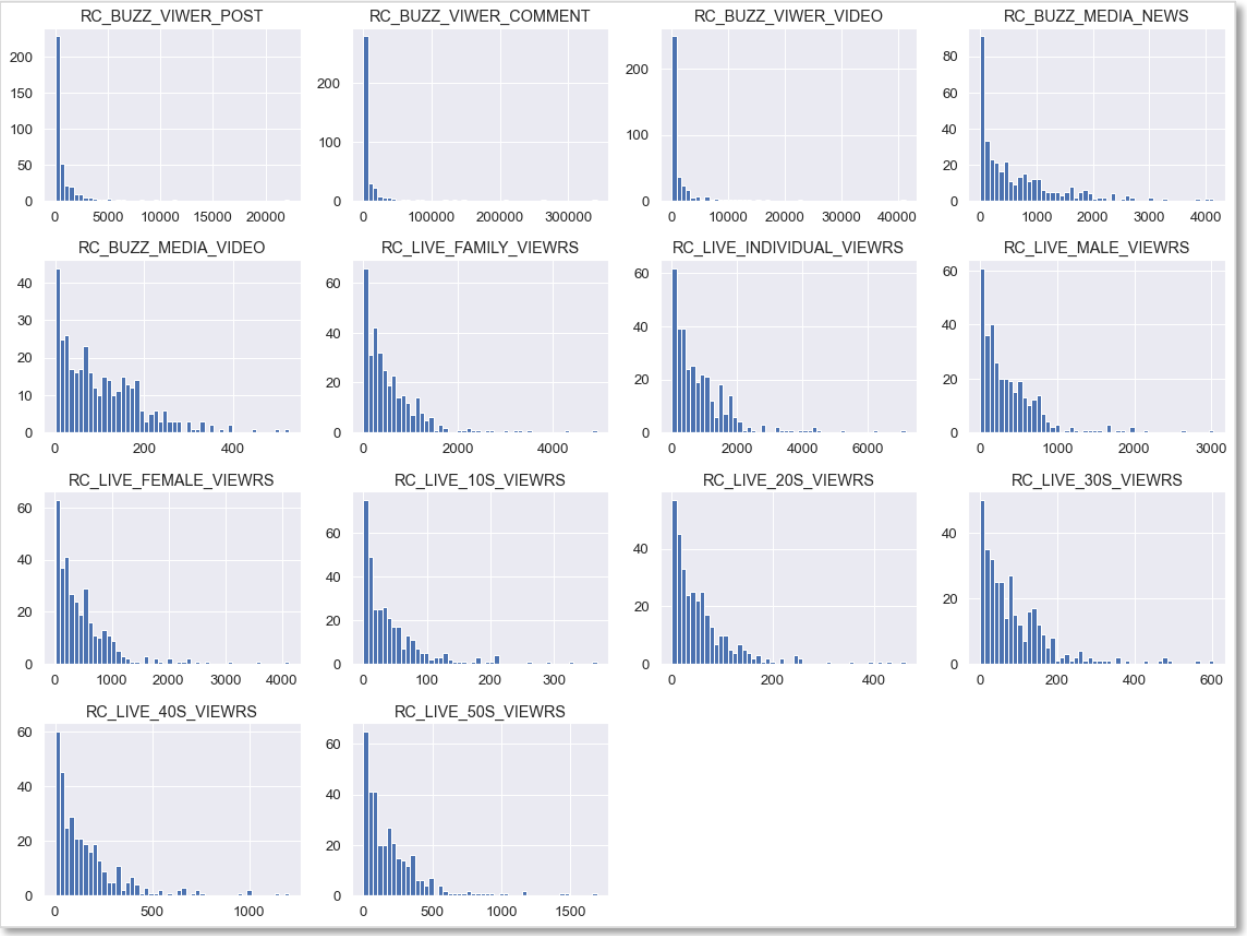
	RC_MONTH	RC_PROGRAM_NAME	PROGRAM_CODE	RC_PROGRAM_STR_DATE	RC_CHNNEL_NM	RC_PROGRAM_DAYS	RC_BUZZ_VIWER_POST	RC_BUZ
0	2021_05	휴먼다큐사노라면	CODE_0001	2012.03.07	MBN	금	23	
1	2021_06	휴먼다큐사노라면	CODE_0001	2012.03.07	MBN	금	20	
2	2021_07	휴먼다큐사노라면	CODE_0001	2012.03.07	MBN	금	31	
3	2021_08	휴먼다큐사노라면	CODE_0001	2012.03.07	MBN	금	10	
4	2021_09	휴먼다큐사노라면	CODE_0001	2012.03.07	MBN	금	11	
5	2021_10	휴먼다큐사노라면	CODE_0001	2012.03.07	MBN	금	19	
6	2021_05	환경스페셜	CODE_0002	2021.03.04	KBS2	목	56	
7	2021_06	환경스페셜	CODE_0002	2021.03.04	KBS2	목	77	
8	2021_07	환경스페셜	CODE_0002	2021.03.04	KBS2	목	220	
9	2021_08	환경스페셜	CODE_0002	2021.03.04	KBS2	목	156	

Groupby()

```
In [270]: racoi_groupby_name = racoi.groupby(
              ['PROGRAM_CODE','RC_PROGRAM_NAME','RC_CHNNEL_NM','RC_PROGRAM_DAYS'],
              as_index=False).mean().sort_values(by=["RC_BUZZ_VIWER_POST"], ascending = [False])
          display(racoi_groupby_name[0:10])
```

	PROGRAM_CODE	RC_PROGRAM_NAME	RC_CHNNEL_NM	RC_PROGRAM_DAYS	RC_BUZZ_VIWER_POST	RC_BUZZ_VIWER_COMMENT	RC_BUZZ_VIWER
137	CODE_0138	스트릿우먼파이터	Mnet	화	22175.333333	265441.333333	41326
296	CODE_0297	갯마을차차차	tvN	토일	11486.333333	80168.000000	13308
143	CODE_0144	쇼미더머니10	Mnet	금	9453.500000	78347.000000	13948
27	CODE_0028	펜트하우스3	SBS	금	9340.800000	84294.800000	22316
133	CODE_0134	슬기로운의사생활2	tvN	수목	8414.800000	57241.200000	11166
294	CODE_0295	걸스플래닛999:소녀대전	Mnet	금	6368.000000	145622.000000	16606
300	CODE_0301	강철부대	채널A/SKY	화	6319.000000	64218.000000	8212
163	CODE_0164	빈센조	tvN	토일	5942.000000	37782.000000	4216

	VIWER_POST	VIWER_COMMENT	VIWER_VIDEO	MEDIA_NEWS	MEDIA_VIDEO
count	364	364	364	364	364
mean	844.7	9,060.5	1,485.1	652.5	104.9
std	1,819.3	29,424.9	3,561.4	742.7	91.8
min	0	0	0	0	0
max	22,175.3	342,113.5	41,326.0	4,140.0	523.5



RACOI BUZZ 데이터셋 분석

✓ RAOCI BUZZ 데이터셋 각 변수별 상위 TOP 10 프로그램 분석

▶ 시청자 게시글 Top10

순	프로그램명	채널명	요일	시청자 게시글
1	스트릿우먼파이터	Mnet	화	22,175
2	갯마을차차차	tvN	토일	11,486
3	쇼미더머니10	Mnet	금	9,454
4	펜트하우스3	SBS	금	9,341
5	슬기로운의사생활2	tvN	수목	8,415
6	걸스플래닛999:소녀대전	Mnet	금	6,368
7	강철부대	채널A/SKY	화	6,319
8	빈센조	tvN	토일	5,942
9	모범택시	SBS	금	5,879
10	홍천기	SBS	월화	5,172

▶ 시청자 댓글 Top10

순	프로그램명	채널명	요일	시청자댓글
1	킹덤:레전더리워	Mnet	목	342,114
2	스트릿우먼파이터	Mnet	화	265,441
3	사랑의콜센타	TV CHOS	목	211,636
4	걸스플래닛999:소녀대전	Mnet	금	145,622
5	문명특급컴눈명스페셜	SBS	금	132,841
6	뽕숭아학당	TV CHOS	수	117,925
7	펜트하우스3	SBS	금	84,295
8	갯마을차차차	tvN	토일	80,168
9	쇼미더머니10	Mnet	금	78,347
10	강철부대	채널A/SKY	화	64,218

▶ 시청자 동영상 조회

순	프로그램명	채널명	요일	시청자 동영상조회
1	스트릿우먼파이터	Mnet	화	41,326
2	펜트하우스3	SBS	금	22,317
3	킹덤:레전더리워	Mnet	목	17,099
4	걸스플래닛999:소녀대전	Mnet	금	16,607
5	마인	tvN	토일	15,216
6	쇼미더머니10	Mnet	금	13,949
7	갯마을차차차	tvN	토일	13,308
8	모범택시	SBS	금	12,677
9	원더우먼	SBS	금	11,725
10	간떨어지는동거	tvN	수목	11,665

▶ 미디어 뉴스 Top10

순	프로그램명	채널명	요일	미디어 뉴스
1	모범택시	SBS	금	4,140
2	스트릿우먼파이터	Mnet	화	3,981
3	마인	tvN	토일	3,864
4	홍천기	SBS	월화	3,244
5	나혼자산다	MBC	금	3,018
6	펜트하우스3	SBS	금	3,015
7	슬기로운의사생활2	tvN	수목	2,677
8	놀면뭐하니?	MBC	토	2,661
9	오월의청춘	KBS2	월화	2,615
10	라디오스타	MBC	수	2,605

▶ 시청자 댓글

순	프로그램명	채널명	요일	미디어 동영상
1	마인	tvN	토일	524
2	놀라운토요일	tvN	토	500
3	스트릿우먼파이터	Mnet	화	441
4	유퀴즈온더블럭	tvN	수	392
5	걸스플래닛999:소녀대전	Mnet	금	391
6	어느날우리집현관으로멸망이	tvN	월화	370
7	모범택시	SBS	금	352
8	갯마을차차차	tvN	토일	347
9	미스몬테크리스토	KBS2	월화수목금	330
10	아는형님	JTBC	토	330

▶ 실시간 시청자수(K)

순	프로그램명	채널명	요일	실시간 시청자수(K)
1	오케이광자매	KBS2	토일	4,925
2	신사와아가씨	KBS2	토일	4,305
3	미스몬테크리스토	KBS2	월화수목금	3,491
4	빨강구두	KBS2	월화수목금	3,289
5	펜트하우스3	SBS	금	3,188
6	속아도꿈결	KBS1	월화수목금	3,038
7	국가대표와이프	KBS1	월화수목금	2,691
8	모범택시	SBS	금	2,569
9	빈센조	tvN	토일	2,460
10	미운우리새끼	SBS	일	2,334

TV 실시간, 시청률 데이터셋 분석

✓ TV 시청률 데이터 셋의 각 변수들을 프로그램 월별 평균값으로 Groupby() 함수를 이용하여 mean값으로 연산하였으며, 각 변수들 통계량 분석 및 시각화 진행

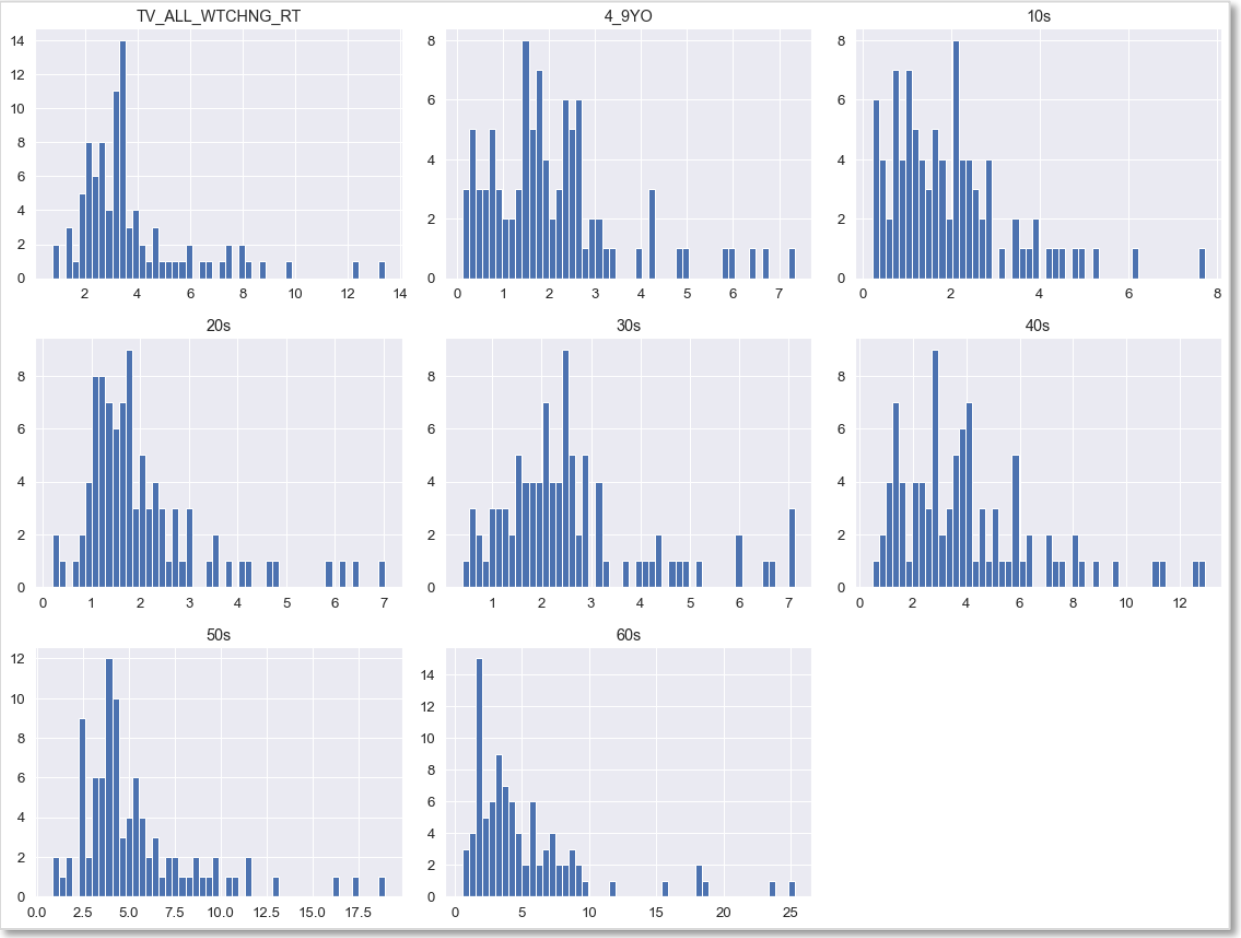
```
tv.head(10)
```

	TV_No	TV_MONTH	PROGRAM_CODE	TV_PROGRAM_NAME	TV_PROGRAM_GENRE	TV_CHNNEL_NM	TV_ALL_WTCHNG_MIN	TV_ALL_WTCHNG_RT	TV_MALE_4_9YO_WTCHNG_MIN
0	TV0086	2021_05	CODE_0003	화요청백전(본)	오락	TV조선	11.325	2.545	0.625
1	TV0434	2021_09	CODE_0004	월화드라마<홍천기>	드라마&영화	SBS	16.789	4.550	4.751
2	TV0523	2021_10	CODE_0004	월화드라마<홍천기>	드라마&영화	SBS	22.966	4.640	3.232
3	TV0600	2021_10	CODE_0012	해치지않아(본)	오락	tvN	8.427	2.071	3.618
4	TV0495	2021_09	CODE_0020	하이클래스(본)	드라마&영화	tvN	8.740	1.625	7.774
5	TV0584	2021_10	CODE_0020	하이클래스(본)	드라마&영화	tvN	9.451	2.068	4.722
6	TV0599	2021_10	CODE_0024	풍류대장(본)	오락	JTBC	8.520	1.487	3.263
7	TV0112	2021_06	CODE_0028	금요드라마<펜트하우스3>	드라마&영화	SBS	29.205	8.904	16.239
8	TV0214	2021_07	CODE_0028	금요드라마<펜트하우스3>	드라마&영화	SBS	30.339	9.601	14.440
9	TV0316	2021_08	CODE_0028	금요드라마<펜트하우스3>	드라마&영화	SBS	32.305	10.288	16.387

Groupby()

```
# tv groupby
tv_groupby_name = tv.groupby(['TV_PROGRAM_NAME', 'PROGRAM_CODE','TV_PROGRAM_GENRE'],as_index=False).mean().sort_values(by=["TV_ALL_WTCHNG_RT"], ascending = [False] )
display(tv_groupby_name[0:10])
```

	TV_PROGRAM_NAME	PROGRAM_CODE	TV_PROGRAM_GENRE	TV_ALL_WTCHNG_MIN	TV_ALL_WTCHNG_RT	TV_MALE_4_9YO_WTCHNG_MIN	TV_MALE_4_9YO_WTCHNG_RT	TV_FEMALE_4_9YO_WTCHNG_MIN
85	주말드라마<오케이광자매>	CODE_0085	드라마&영화	72.290600	13.43440	24.605400	4.66460	43.92000
84	주말드라마<신사와아가씨>	CODE_0125	드라마&영화	53.891000	12.17650	21.896500	4.98050	28.58950
11	금요드라마<펜트하우스3>	CODE_0028	드라마&영화	27.041250	9.78000	14.853250	5.63875	23.93025
77	일일드라마<미스몬테크리스토>	CODE_0189	드라마&영화	66.769500	8.73250	20.364000	2.66350	43.67400
79	일일드라마<빨강구두>	CODE_0162	드라마&영화	52.754250	8.14225	20.228250	3.07200	23.31175
42	빈센조(본)	CODE_0164	드라마&영화	14.193000	7.97300	10.096000	5.67200	4.95400
13	금토드라마<모범택시>	CODE_0198	드라마&영화	46.325000	7.96000	25.931000	4.45500	30.58900
80	일일드라마<속아도꿈결>	CODE_0148	드라마&영화	39.208600	7.38740	11.944000	2.21060	16.02340



TV 실시간, 시청률 데이터셋 분석

✓ TV 실시간, 시청률 데이터셋 각 변수별 상위 TOP 10 프로그램 분석

▶ 전체 시청률 TOP10

순	프로그램명	장르	전체 시청률
1	오케이광자매	드라마&영화	13.434
2	신사와아가씨	드라마&영화	12.177
3	펜트하우스3	드라마&영화	9.780
4	미스몬테크리스토	드라마&영화	8.733
5	빨강구두	드라마&영화	8.142
6	빈센조	드라마&영화	7.973
7	모범택시	드라마&영화	7.960
8	속아도꿈결	드라마&영화	7.387
9	슬기로운의사생활시즌2	드라마&영화	7.374
10	미운우리새끼	오락	7.173
-	**평균**	-	**9.013**

▶4-9 시청률 TOP10

순	프로그램명	장르	4-9s
1	펜트하우스3	드라마&영화	7.035
2	놀면뭐하니	오락	4.254
3	오케이광자매	드라마&영화	6.105
4	슬기로운의사생활시즌2	드라마&영화	5.870
5	신사와아가씨	드라마&영화	4.641
6	원더우먼	드라마&영화	3.841
7	모범택시	드라마&영화	4.720
8	빈센조	드라마&영화	6.437
9	나혼자산다	오락	2.767
10	미스몬테크리스토	드라마&영화	2.979
-	**평균**	-	**4.865**

▶10대 시청률 TOP10

순	프로그램명	장르	10s
1	펜트하우스3	드라마&영화	7.728
2	슬기로운의사생활시즌2	드라마&영화	6.078
3	빈센조	드라마&영화	5.278
4	오케이광자매	드라마&영화	5.004
5	갯마을차차차	드라마&영화	4.739
6	모범택시	드라마&영화	4.553
7	신사와아가씨	드라마&영화	4.405
8	미운우리새끼	오락	4.191
9	놀면뭐하니	오락	3.926
10	나혼자산다	오락	3.842
-	**평균**	-	**4.974**

▶ 20대 시청률 TOP10

순	프로그램명	장르	20s
1	펜트하우스3	드라마&영화	7.035
2	빈센조(본)	드라마&영화	6.437
3	오케이광자매	드라마&영화	6.105
4	슬기로운의사생활시즌2	드라마&영화	5.870
5	모범택시	드라마&영화	4.720
6	신사와아가씨	드라마&영화	4.641
7	놀면뭐하니	오락	4.254
8	갯마을차차차	드라마&영화	4.079
9	원더우먼	드라마&영화	3.841
10	미운우리새끼	오락	3.537
-	**평균**	-	**5.052**

▶ 30대 시청률 TOP10

순	프로그램명	장르	30s
1	슬기로운의사생활시즌2(본)	드라마&영화	7.133
2	빈센조(본)	드라마&영화	7.106
3	펜트하우스3	드라마&영화	7.027
4	놀면뭐하니	오락	6.694
5	모범택시	드라마&영화	6.530
6	신사와아가씨	드라마&영화	6.004
7	오케이광자매	드라마&영화	5.933
8	나혼자산다	오락	5.189
9	미운우리새끼	오락	4.872
10	미스몬테크리스토	드라마&영화	4.769
-	**평균**	-	**6.126**

▶ 40대 시청률 TOP10

순	프로그램명	장르	40s
1	슬기로운의사생활시즌2	드라마&영화	12.946
2	빈센조	드라마&영화	12.577
3	모범택시	드라마&영화	11.449
4	펜트하우스3	드라마&영화	11.047
5	미운우리새끼	오락	9.615
6	오케이광자매	드라마&영화	8.849
7	지리산	드라마&영화	8.360
8	갯마을차차차	드라마&영화	8.113
9	놀면뭐하니	오락	8.062
10	나혼자산다	오락	7.543
-	**평균**	-	**9.856**

▶ 50대 시청률 TOP10

순	프로그램명	장르	50s
1	오케이광자매	드라마&영화	18.950
2	신사와아가씨	드라마&영화	17.190
3	펜트하우스3	드라마&영화	16.121
4	모범택시	드라마&영화	12.804
5	빈센조	드라마&영화	11.530
6	원더우먼	드라마&영화	11.437
7	슬기로운의사생활시즌2	드라마&영화	10.635
8	미운우리새끼	오락	10.292
9	내일은국민가수	오락	9.856
10	미스몬테크리스토	드라마&영화	9.578
-	**평균**	-	**12.839**

▶ 60대 시청률 TOP10

순	프로그램명	장르	60s
1	오케이광자매	드라마&영화	25.315
2	신사와아가씨	드라마&영화	23.482
3	미스몬테크리스토	드라마&영화	18.599
4	빨강구두	드라마&영화	18.162
5	속아도꿈결	드라마&영화	17.968
6	국가대표와이프	드라마&영화	15.423
7	내일은국민가수	오락	11.614
8	사랑의콜센타(본)	오락	9.811
9	아침마당	정보	9.423
10	펜트하우스3	드라마&영화	9.055
-	**평균**	-	**15.885**

VOD 시청률 데이터셋 분석

✓ VOD 시청률 데이터 셋의 각 변수들을 프로그램 월별 평균값으로 Groupby() 함수를 이용하여 mean값으로 연산하였으며, 각 변수들 통계량 분석 및 시각화 진행

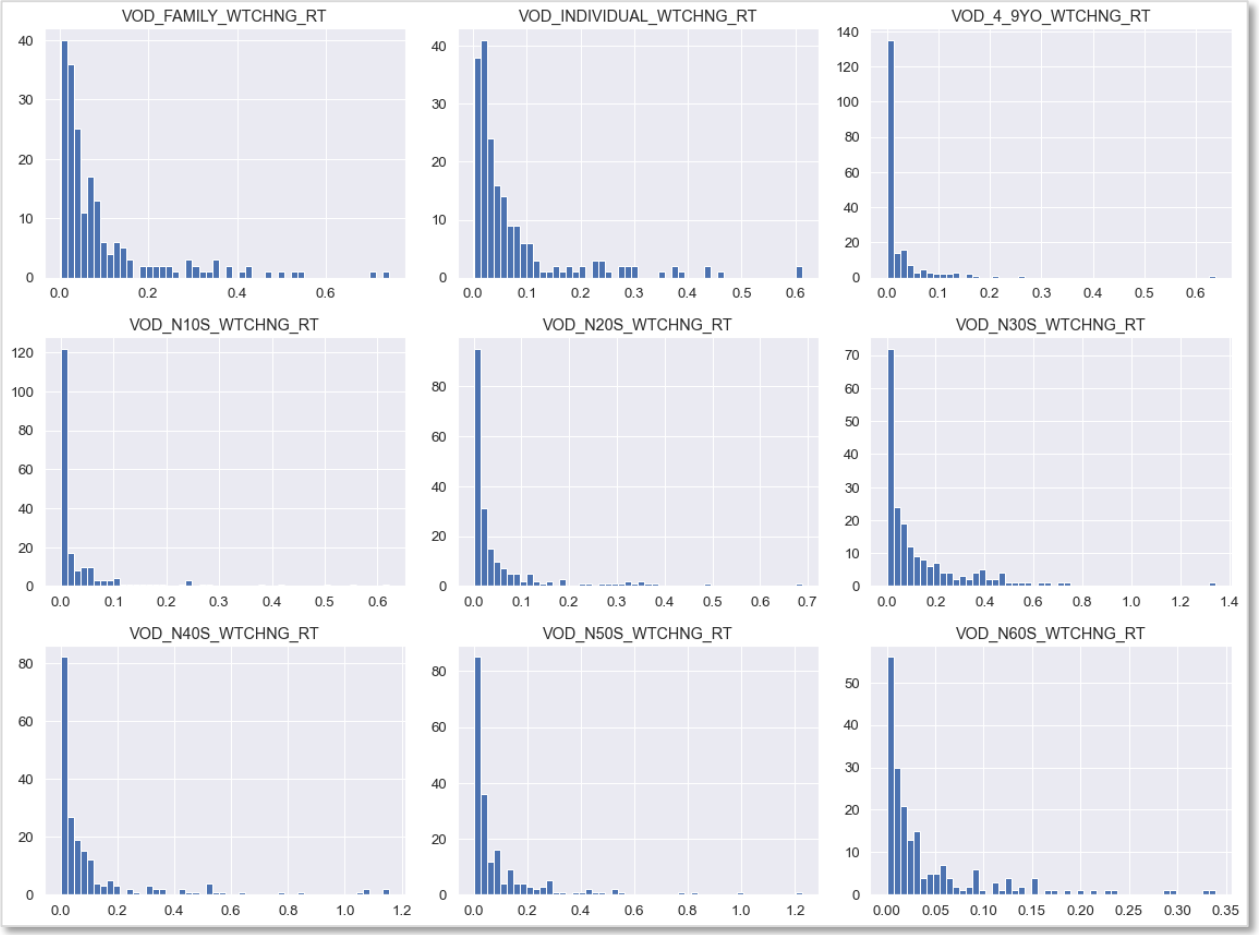
```
vod.head(10)
```

	VOD_NO	VOD_MONTH	VOD_PROGRAM_NAME	PROGRAM_CODE	VOD_PROGRAM_GENRE	VOD_CHNNEL_NM	VOD_FAMILY_WTCHNG_MIN	VOD_FAMILY_WTCHNG_RT	VOD_INDIVI
0	VOD0069	2021_05	휴먼다큐사노라면(본)	CODE_0001	정보	MBN	0.095	0.042	
1	VOD0147	2021_06	휴먼다큐사노라면(본)	CODE_0001	정보	MBN	0.174	0.076	
2	VOD0286	2021_07	휴먼다큐사노라면(본)	CODE_0001	정보	MBN	0.031	0.011	
3	VOD0358	2021_08	휴먼다큐사노라면(본)	CODE_0001	정보	MBN	0.096	0.042	
4	VOD0471	2021_09	휴먼다큐사노라면(본)	CODE_0001	정보	MBN	0.039	0.017	
5	VOD0580	2021_10	휴먼다큐사노라면(본)	CODE_0001	정보	MBN	0.040	0.014	
6	VOD0490	2021_09	환경스페셜	CODE_0002	정보	KBS2	0.023	0.009	
7	VOD0314	2021_08	월화드라마<홍천기>	CODE_0004	드라마&영화	SBS	0.548	0.382	
8	VOD0403	2021_09	월화드라마<홍천기>	CODE_0004	드라마&영화	SBS	1.359	0.368	
9	VOD0504	2021_10	월화드라마<홍천기>	CODE_0004	드라마&영화	SBS	2.341	0.474	

Groupby()

```
# vod groupby
vod_groupby_name = vod.groupby(['VOD_PROGRAM_NAME', 'PROGRAM_CODE', 'VOD_PROGRAM_GENRE'],as_index=False).mean().sort_values(by=["VOD_FAMILY_WTCHNG_RT"],
display(vod_groupby_name[0:10])
```

	VOD_PROGRAM_NAME	PROGRAM_CODE	VOD_PROGRAM_GENRE	VOD_FAMILY_WTCHNG_MIN	VOD_FAMILY_WTCHNG_RT	VOD_INDIVIDUAL_WTCHNG_MIN	VOD_INDIVIDUAL_WT
112	슬기로운의사생활시즌2(본)	CODE_0134	드라마&영화	2.267500	0.743000	1.848750	
10	갯마을차차차(본)	CODE_0297	드라마&영화	3.128667	0.704333	2.778333	
87	빈센조(본)	CODE_0164	드라마&영화	0.950000	0.536000	0.772000	
152	월화드라마<연모>	CODE_0094	드라마&영화	2.178000	0.522000	1.911000	
62	로스쿨(본)	CODE_0212	드라마&영화	1.620000	0.500500	1.203500	
27	금토드라마<검은태양>	CODE_0294	드라마&영화	2.112000	0.470000	1.923000	
28	금토드라마<모범택시>	CODE_0198	드라마&영화	2.506000	0.432000	2.138000	
64	마인(본)	CODE_0206	드라마&영화	2.413500	0.425000	2.005000	
154	월화드라마<홍천기>	CODE_0004	드라마&영화	1.416000	0.408000	1.000333	
31	꼬리에꼬리를무는그날이야기	CODE_0267	정보	0.658000	0.382000	0.660000	



VOD 실시간 시청률 데이터셋 분석

✓ VOD 실시간 시청률 데이터셋 각 변수별 상위 TOP 10 프로그램 분석

▶ 전체 시청률 TOP10

순	프로그램명	장르	전체시청률
1	갯마을차차차	드라마&영화	0.613
2	슬기로운의사생활시즌2	드라마&영화	0.603
3	연모	드라마&영화	0.458
4	검은태양	드라마&영화	0.435
5	빈센조(본)	드라마&영화	0.435
6	꼬리에꼬리를무는그날이야기	정보	0.384
7	로스쿨	드라마&영화	0.373
8	모범택시	드라마&영화	0.369
9	마인	드라마&영화	0.354
10	홍천기	드라마&영화	0.296
-	**평균**	-	**0.432**

▶4-9 시청률 TOP10

순	프로그램명	장르	4-9s
1	꼬리에꼬리를무는그날이야기	정보	0.638
2	모범택시	드라마&영화	0.258
3	검은태양	드라마&영화	0.207
4	슬기로운의사생활시즌2	드라마&영화	0.171
5	알고있지만	드라마&영화	0.164
6	강철부대	오락	0.163
7	마인	드라마&영화	0.138
8	놀면뭐하니	오락	0.135
9	다시뜨거워지고싶은애로부부	오락	0.129
10	미치지않고서야	드라마&영화	0.127
-	**평균**	-	**0.213**

▶10대 시청률 TOP10

순	프로그램명	장르	10s
1	너는나의봄(본)	드라마&영화	0.623
2	슬기로운의사생활시즌2	드라마&영화	0.558
3	갯마을차차차	드라마&영화	0.507
4	로스쿨	드라마&영화	0.418
5	라켓소년단	드라마&영화	0.376
6	빈센조	드라마&영화	0.279
7	검은태양	드라마&영화	0.262
8	강철부대	오락	0.246
9	꼬리에꼬리를무는그날이야기	정보	0.239
10	연모	드라마&영화	0.239
-	**평균**	-	**0.375**

▶ 20대 시청률 TOP10

순	프로그램명	장르	20s
1	꼬리에꼬리를무는그날이야기	정보	0.689
2	마인	드라마&영화	0.488
3	지리산	드라마&영화	0.381
4	강철부대	오락	0.367
5	너를닮은사람	드라마&영화	0.355
6	갯마을차차차	드라마&영화	0.349
7	로스쿨	드라마&영화	0.342
8	놀면뭐하니	오락	0.323
9	출장십오야	오락	0.323
10	빈센조	드라마&영화	0.307
-	**평균**	-	**0.392**

▶ 30대 시청률 TOP10

순	프로그램명	장르	30s
1	슬기로운의사생활시즌2	드라마&영화	1.344
2	모범택시	드라마&영화	0.728
3	갯마을차차차	드라마&영화	0.714
4	놀면뭐하니	오락	0.665
5	슬기로운산촌생활	오락	0.628
6	로스쿨	드라마&영화	0.576
7	런닝맨	오락	0.559
8	출장십오야	오락	0.527
9	빈센조	드라마&영화	0.490
10	마인	드라마&영화	0.480
-	**평균**	-	**0.671**

▶ 40대 시청률 TOP10

순	프로그램명	장르	40s
1	빈센조	드라마&영화	1.157
2	갯마을차차차	드라마&영화	1.146
3	슬기로운의사생활시즌2	드라마&영화	1.074
4	마우스	드라마&영화	1.065
5	어쩌다사장	오락	1.042
6	모범택시	드라마&영화	0.837
7	검은태양	드라마&영화	0.787
8	로스쿨	드라마&영화	0.642
9	너는나의봄	드라마&영화	0.561
10	원더우먼	드라마&영화	0.549
-	**평균**	-	**0.886**

▶ 50대 시청률 TOP10

순	프로그램명	장르	50s
1	연모	드라마&영화	1.228
2	갯마을차차차	드라마&영화	0.990
3	검은태양	드라마&영화	0.833
4	홍천기	드라마&영화	0.776
5	로스쿨	드라마&영화	0.543
6	슬기로운의사생활시즌2	드라마&영화	0.533
7	구경이	드라마&영화	0.516
8	라켓소년단	드라마&영화	0.468
9	빈센조	드라마&영화	0.466
10	꼬리에꼬리를무는그날이야기	정보	0.435
-	**평균**	-	**0.679**

▶ 60대 시청률 TOP10

순	프로그램명	장르	60s
1	너를닮은사람	드라마&영화	0.339
2	연모	드라마&영화	0.330
3	꼬리에꼬리를무는그날이야기	정보	0.292
4	구경이	드라마&영화	0.285
5	펜트하우스3	드라마&영화	0.234
6	갯마을차차차	드라마&영화	0.230
7	원더우먼	드라마&영화	0.216
8	마인	드라마&영화	0.202
9	슬기로운의사생활시즌2	드라마&영화	0.184
10	홍천기	드라마&영화	0.171
-	**평균**	-	**0.248**

RACOI, TV, VOD 데이터셋 관계 분석

✓ RACOI, TV. VOD 데이터셋의 상관관계 분석 및 시각화 진행

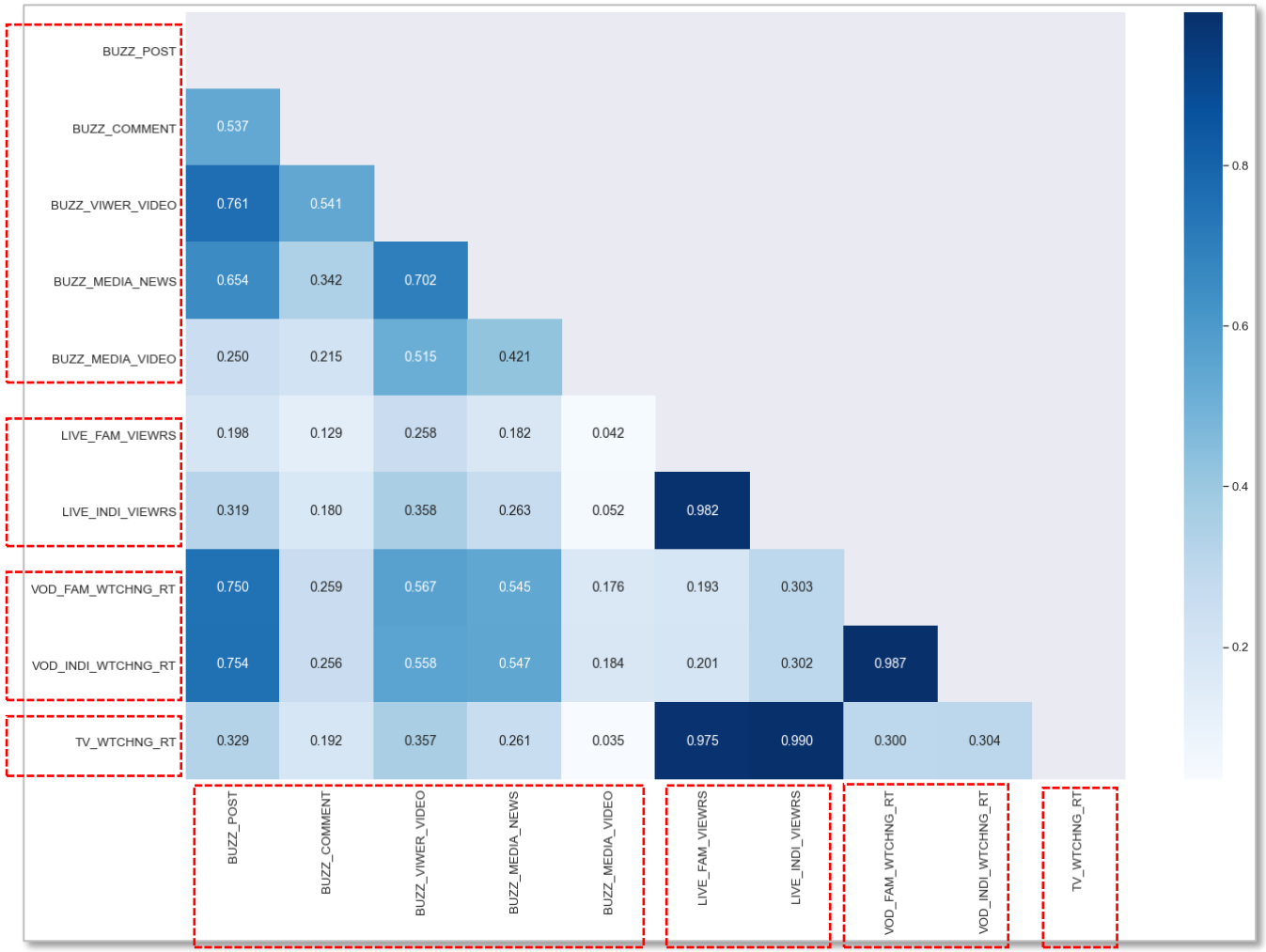
Merge ("inner Join")

```
racoi_vod_mean = pd.merge(left = racoi_groupby_name, right = vod_mean, how = "inner", on = "PROGRAM_CODE")
racoi_vod_tv_mean = pd.merge(left = racoi_vod_mean, right = tv_mean, how = "inner", on = "PROGRAM_CODE")
racoi_vod_tv_mean[:10]
```

Corr분석 & 시각화

```
cols = ['BUZZ_POST', 'BUZZ_COMMENT', 'BUZZ_VIWER_VIDEO',
        'BUZZ_MEDIA_NEWS', 'BUZZ_MEDIA_VIDEO', 'LIVE_FAM_VIEWRS',
        'LIVE_INDI_VIEWRS', 'VOD_FAM_WTCHNG_RT', 'VOD_INDI_WTCHNG_RT', 'TV_WTCHNG_RT']

corr = df[cols].corr(method = 'pearson')
mask = np.zeros_like(corr)
mask[np.triu_indices_from(mask)] = True

display(corr)

cols_view = ['BUZZ_POST', 'BUZZ_COMMENT', 'BUZZ_VIWER_VIDEO',
        'BUZZ_MEDIA_NEWS', 'BUZZ_MEDIA_VIDEO', 'LIVE_FAM_VIEWRS',
        'LIVE_INDI_VIEWRS', 'VOD_FAM_WTCHNG_RT', 'VOD_INDI_WTCHNG_RT', 'TV_WTCHNG_RT']

sns.set(rc = {'figure.figsize':(20,14)}, font_scale = 1.2)
df_hm = sns.heatmap(corr.values, cmap = "Blues", cbar = True, annot = True, xticklabels = cols_view,
                    yticklabels = cols_view, fmt = '.3f', annot_kws = {'size':14}, mask = mask)

plt.tight_layout()
plt.show()
```

	BUZZ_POST	BUZZ_COMMENT	BUZZ_VIWER_VIDEO	BUZZ_MEDIA_NEWS	BUZZ_MEDIA_VIDEO	LIVE_FAM_VIEWRS	LIVE_INDI_VIEWRS	VOD_FAM_WTCHNG_RT
BUZZ_POST	1.000000	0.536678	0.760964	0.654063	0.249632	0.197835	0.318776	0.750229
BUZZ_COMMENT	0.536678	1.000000	0.541280	0.341520	0.214543	0.129164	0.179647	0.258509
BUZZ_VIWER_VIDEO	0.760964	0.541280	1.000000	0.701919	0.514956	0.257770	0.357734	0.567256
BUZZ_MEDIA_NEWS	0.654063	0.341520	0.701919	1.000000	0.420610	0.181670	0.262811	0.545036
BUZZ_MEDIA_VIDEO	0.249632	0.214543	0.514956	0.420610	1.000000	0.042418	0.051870	0.176333
LIVE_FAM_VIEWRS	0.197835	0.129164	0.257770	0.181670	0.042418	1.000000	0.982446	0.192669
LIVE_INDI_VIEWRS	0.318776	0.179647	0.357734	0.262811	0.051870	0.982446	1.000000	0.302694
VOD_FAM_WTCHNG_RT	0.750229	0.258509	0.567256	0.545036	0.176333	0.192669	0.302694	1.000000
VOD_INDI_WTCHNG_RT	0.753789	0.255776	0.557962	0.546778	0.183689	0.201352	0.301665	0.987057
TV_WTCHNG_RT	0.329019	0.191692	0.357336	0.260655	0.034791	0.975301	0.990222	0.300072



✓ TV시청률이 높다고 BUZZ 데이터 지표가 높게 나타나는 것은 아님

✓ TV시청률보다 VOD 시청률이 높을수록 BUZZ 데이터 (게시글, 댓글, 동영상 조회수, 뉴스) 지표가 높게 나타남

Conclusion

- ✓ 컨텐츠를 접할 수 있는 경로가 다양화 됨에 따라서 실시간 시청률이 높지 않더라도 온라인에서 화재가 되고 동영상 조회 및 VOD로 다시 찾아보는 프로그램들이 많아짐
- ✓ 예능, 오락 프로그램들이 회차별 주제와 출연진이 다양함에 따라 버즈 데이터가 시청률에 비하여 더 높고 화제가 될 수 있음을 확인
- ✓ 전처리 과정 및 HDFS 시스템 구축에 시행착오 발생하고 많은 시간이 소요되었으며, 실제 수업에서 진행했던 실습과 실제 데이터 분석은 더 어려웠음을 체감하였음
- ✓ 회차별 주제 정보, 출연진 정보, 방송시간 등 조금더 다채로운 정보들이 수집되어 분석이 되면 더 의미 있는 결과를 도출할 수 있을 것이라고 생각하며, 향후 프로그램별 광고나 마케팅에 활용도 검토 가능하겠다고 판단함

- 끝 -