

효율측정 모델을 통한



TV광고 성과 예측

- 21년 광고최적화를 위한 전략 제시-

ADA팀 박상욱, 최민형, 허원강

CONTENTS

1. 개요 및 프로젝트 목표

- 팀소개 및 타임라인
- 배경
- 문제인식 및 과제도출
- 정리
- 데이터 선정 및 이유
- 프로젝트 목표
- 데이터 셋 및 전처리

3. TV광고 효율측정 모델

- feature engineering
- 인코딩
- 상관계수파악
- 모델링

2. 데이터 분석 및 인사이트

- 월별 광고집행현황 분석
- 광고 삽입위치 및 광고길이 분석
- 캠페인 기간 및 요일별 분석
- 채널 및 프로그램 시간 분석
- 프로그램 분석
- 정리

4. 결론

- 모델 예측 결과
- 총평 및 향후 발전방향

개요 및 프로젝트 목표

팀소개 및 타임라인

ADA(ATL Data analyst)



허원강

프로젝트 목표설정 및 모델 기획
인사이트 도출 및 전략 수립
머신러닝 모델 성과 테스트



박상묵

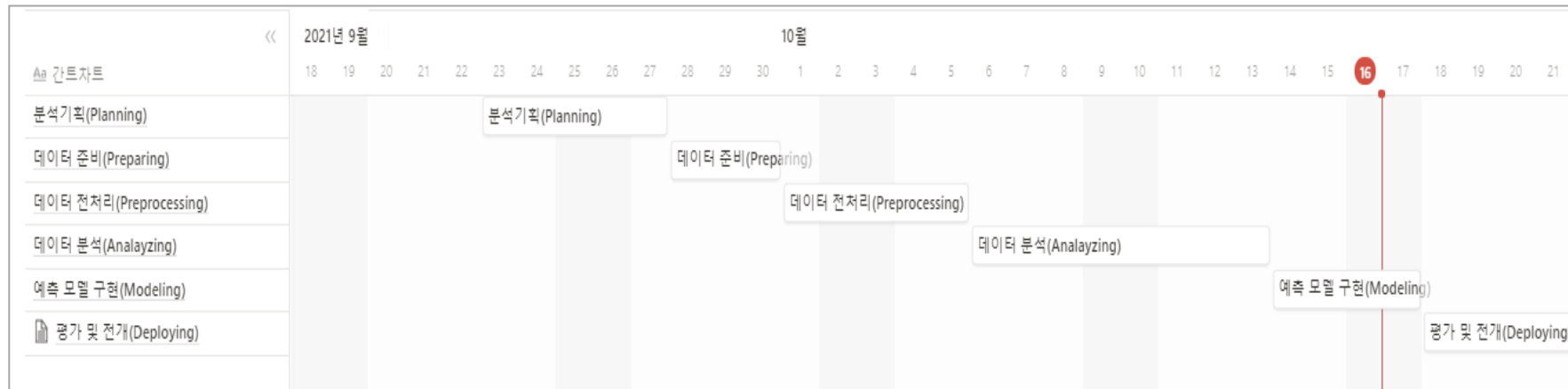
성과측정 모델 구축 및 검증
데이터 분석 및 시각화
보완점 개선 및 발전방향 수립



최민형

시장 리서치 및 과제 수립
프로젝트 구성 및 스토리텔링
인사이트 도출 및 적용

프로젝트 타임라인 : 9월 3째주 ~ 10월 4째주



배경

마케터들의 공감...

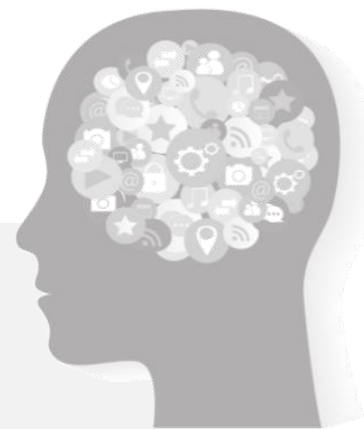
광고로 우리 제품/서비스/행사를 알려야 하는데...

요즘 많이 하는 '디지털미디어'만으로
광고를 집행 해야 하는 걸까?



경쟁이 치열한 디지털 광고만으로 제대로 알릴 수 있을까?

그렇다면... 레가시 미디어를 대표하는 TV는 어떨까...?
(사실 잘 된 광고는 대부분 TV광고도 집행하니까...)



문제인식 및 과제도출

문제인식

TV광고는 성과를 측정하기도 어렵고
결과를 컨트롤하기도 어렵잖아...



해결방안 모색

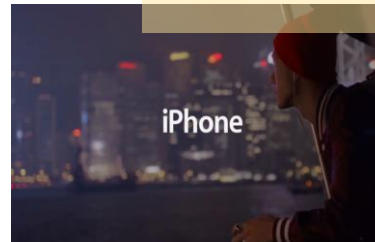
**어떻게 하면
TV 광고효율을 높일 수 있을까?**

심플하게?

병맛으로?

재미있게?

근데 너무 주관적이다...결과를 보장할 수 없음...



명확한 무언가 없을까..? 이를 테면, 광고 길이 라든지, 언제 보여줘야 한다든지,,,?

TV광고의 성과를 결정짓는 핵심 지표를 찾아보고 이 지표에 영향을 끼치는 것을 찾아보자

정리

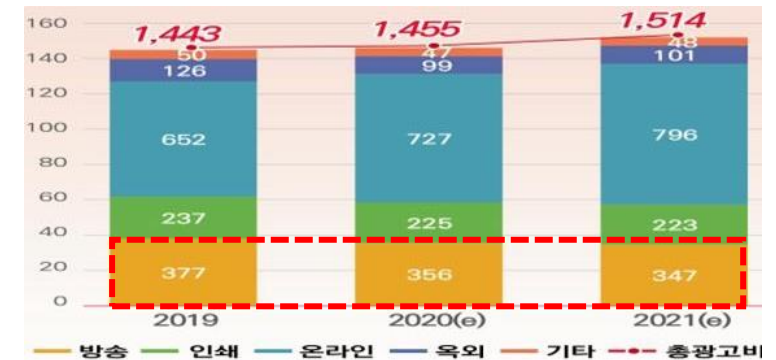
- ✓ **현업에서 광고 데이터 분석은 주로 디지털 매체에 대한 분석으로 한정 ※ 현업 마케터 10명에게 설문조사**
- ✓ **TV광고는 전통적 마케팅 수단인 ATL(TV,신문,라디오,잡지) 중에서 가장 강력한 매체**
※ 단 한번의 광고로 임팩트있게 노출 할 수 있는 유일무이한 매체 (하나뿐인 내편 30초 1회 광고 → 약 1천만명 노출)
- ✓ **온라인 광고 확대로 방송광고가 다소 하락했지만 여전히 높은 비중**
※ 2019년 26%, 2020년(E) 24%, 2021년(E) 23%
- ✓ **TV광고에서 도달(Reach)이라는 지표가 TV광고의 성과를 대표**
- ✓ **도달을 높이는 핵심변수를 찾는게 핵심**

▶ B브랜드 19년 2월 캠페인 스폷별 광고 시청률 TOP10 (개인 전체 기준)

	채널	집행일	프로그램명	시청률 (%)	시청자수 (명)
1	KBS2	2/24	주말드라마 (하나뿐인내편)	21.40%	10,337,221
2	JTBC	2/7	금토드라마 (SKY캐슬)	13.90%	6,714,363
3	SBS	2/10	마운우리새끼다사쓰는육아일기	9.40%	4,540,648
4	KBS2	2/24	해파선데이	7.10%	3,429,639
5	JTBC	2/9	아는형님	6.70%	3,236,420
6	SBS	2/7	드라마스페셜 (황후의품격)	6.60%	3,188,115
7	SBS	2/23	금토드라마 (열혈사제)	6.50%	3,139,810
8	MBC	2/22	나혼자산다	5.90%	2,849,981
9	SBS	2/23	장군의법칙IN북마리아나	5.70%	2,753,372
10	MBC	2/23	전자적참견시험	5.20%	2,511,848

출처 : AGBNMR (4세 이상 전체 인구 4,830만명 기준으로 시청자수 산출)

[그림3]



이에 ADA팀은 ‘마케터의 숙원과제’인 TV광고 데이터를 분석하여 인사이트를 도출하는 프로젝트를 진행하고자 합니다.

데이터 선정 및 이유



관광

무형유산

축제

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보

티엔엠에스 >

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보(202108)

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보(202107)

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보(202106)

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보(202104)

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보(202012)

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보(202011)

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보(202010)

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보(202009)

월별 지역축제 tvN등 PP 채널 광고 집행 DB 정보(202008)

✓ TV 시청 도달률에 영향을 끼치는 다양한 피처 보유
(채널, 광고유형, 지역, 시기, 러닝타임 등)

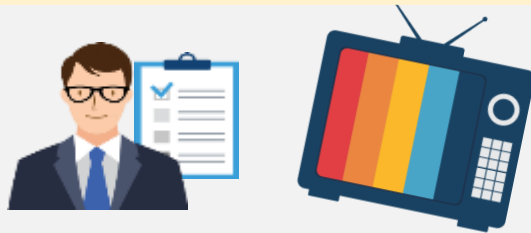
✓ 지역축제는 매년 꾸준히 TV광고 집행

✓ 광고 최적화를 통한 지역경제 활성화 견인



프로젝트 목표

**20년 TV광고 집행결과
의 효율 측정 및 핵심 지표 추출**



**21년 TV광고 편성
전략 제시**

**TV광고 성과측정 모델을
통한 도달(Reach) 예측**

목표 세부내용

- ✓ 지역 축제에 TV광고가 효율적으로 집행되었는가 판단
- ✓ 2021년 10월에 예정된 지역 축제에 대한 TV광고 편성 전략 제시
- ✓ 머신러닝 모델을 활용하여 ADA팀의 편성전략의 정확도 체크
(2020년과 동일한 양[광고 시간]의 광고를 집행했을 때 Reach를 얼마나 개선할 수 있을지 예측)

데이터셋 및 전처리 과정

✓ 기간

- 2020년 7~12월

✓ 컬럼수

- 35 → 32개

✓ 주요변수(y값)

- 도달율1+~5+는 프리퀀시 의미 (1명에게 몇회 노출되었는지)

✓ 데이터 전처리

- 2020년 하반기 데이터 Merge

- 컬럼명 한글화

- 포맷 통일 (날짜, "제거")

- 광고 시간 전처리(10, 15, 20, 30, 40 으로 정리)

- 중복데이터 제거 (시청측정단위코드 : 유료매체, 전국)

- 결측치 제거 (TV광고 미집행 데이터 - 도달율 1+ : 0인 행)

✓ 데이터 수

- 70,528 → 33,060개

✓ 작업환경

- Jupyter notebook3.8.5

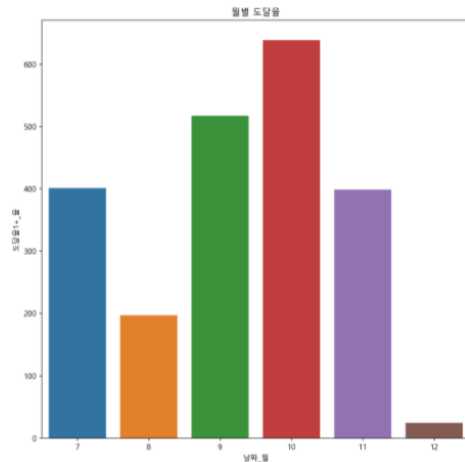
DATE_DE	날짜_날짜
CHNNEL_CD	채널_코드
ADVRTS_NM_NM	광고명_이름
CMPGN_CODE_CD	캠페인코드_코드
CMPGN_BEGIN_DATE_DE	캠페인시작날짜_날짜
ADVRTS_BEGIN_TM_TM	광고시작시간_시각
ADVRTS_LT_TMLT	광고길이_시간길이
ADVRTS_BRDCST_AREA_CD	광고방송지역_코드
ADVRTS_TY_TY	광고유형_유형
ADVRTS_CMPGN_AT	광고/캠페인_여부
BLCK_INNER_ADVRTS_LC_LC	블록내광고위치_위치
BLCK_INNER_ADVRTS_CO_CO	블록내광고갯수_수
ADVRTS_OWNR_NM	광고주_이름
ADVRTS_OWNR_CODE_CD	광고주_코드_코드
BRAND_NM	브랜드_이름
BRAND_CODE_CD	브랜드_코드_코드
ADVRTS_INDUTY_LCLAS_NM	광고업종대분류_이름
ADVRTS_INDUTY_LCLAS_CODE_CD	광고업종대분류_코드_코드
ADVRTS_INDUTY_MLSFC_NM	광고업종중분류_이름
ADVRTS_INDUTY_MLSFC_CODE_CD	광고업종중분류_코드_코드
ADVRTS_INDUTY_SCLAS_NM	광고업종소분류_이름
ADVRTS_INDUTY_SCLAS_CODE_CD	광고업종소분류_코드_코드
ADVRTS_TY_DC	광고유형별_상세설명_설명
PROGRM_BRDCST_AREA_CD	프로그램_방송지역_코드
ADVRTS_PROGRM_NM	광고_프로그램_이름
PROGRM_BEGIN_TM_TM	프로그램_시작시간_시각
PROGRM_END_TM_TM	프로그램_종료시간_시각
PROGRM_BRDCST_AREA_RCIV_CD	프로그램_광고_지역(수신)_코드
AUDE_MESUREUNIT_CD	시청_측정단위_코드
TVGRP_BRDCST_AREA_RT	TVGRP_방송지역_율
REACH_RATE_1_RT	도달율1+_율
REACH_RATE_2_RT	도달율2+_율
REACH_RATE_3_RT	도달율3+_율
REACH_RATE_4_RT	도달율4+_율
REACH_RATE_5_RT	도달율5+_율

< 컬럼 소개 >

데이터 분석 및 인사이트

월별 광고집행현황 분석

현황



날짜_월	도달율1+_율	광고길이_시간길이
7	1478439	52360
8	825607	88220
9	2788691	147750
10	7011832	276060
11	1913034	138220
12	8819	14320

***광고효율성**
(광고 시간 대비 도달)



10월 분석

날짜_월	도달율1+_율	광고길이_시간길이	시간대비도달	
0	7	1478439	52360	28.23604
1	8	825607	88220	9.35850
2	9	2788691	147750	18.87439
3	10	7011832	276060	25.39967
4	11	1913034	138220	13.84050
5	12	8819	14320	0.61585

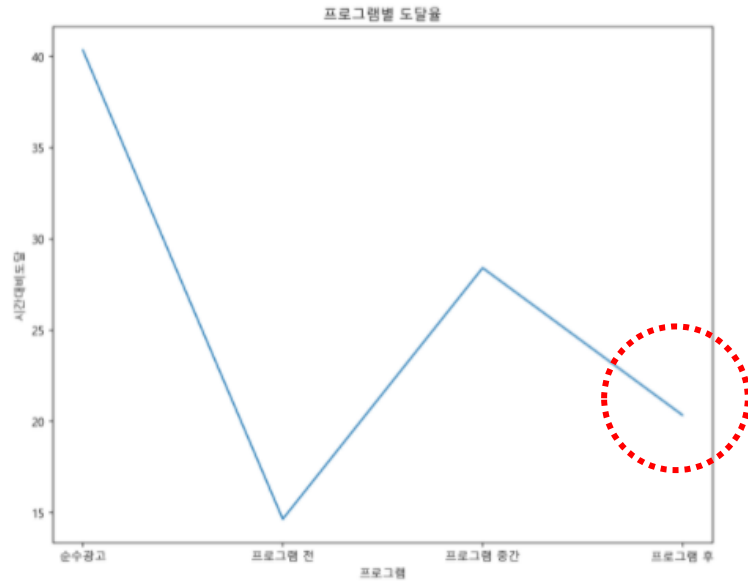
- ✓ 총 276,060초의 광고 집행, 총 7,011,832명 도달
- ✓ 광고효율성은 25.3으로 동기간 2위로
비교적 높은 효율로 집행

가장 많은 축제와 많은 광고 예산이 투입된
기간인 만큼 심층 분석을 통해 광고성과에 영향을 미치는
핵심 지표를 찾아내어 도달 개선 필요

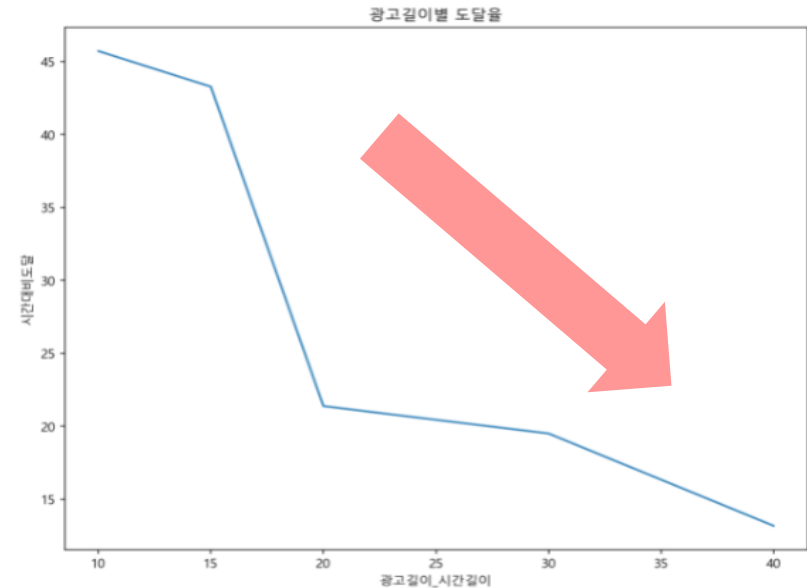
- ✓ 2020년(하) 10월에 지역 축제가 가장 많았음
- ✓ TV 광고 역시 동월에 가장 많은 집행과 가장 높은 도달 발생

광고삽입위치 및 광고길이 분석

광고가 삽입되는 위치를 '프로그램 전', '중간', '후'
그리고 프로그램과 관계없는 '순수광고' 시간으로 구분하여 효율 비교

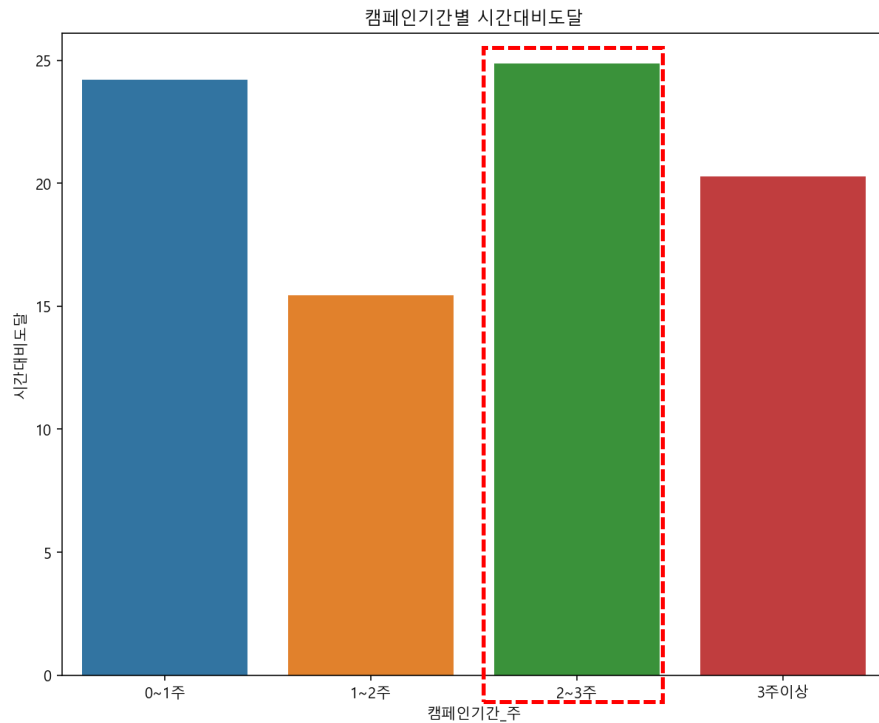


- ✓ 시간대비 도달 효율이 순수광고가 우수하게 나타나지만, 독립변수간 데이터 크기가 상이해 비교가 어려움
- ✓ 데이터 크기가 적은 순수광고와 프로그램 중간 데이터를 배제하고 프로그램 전/후에 광고 효율을 비교하면 프로그램 후에 광고를 집행하는 것이 높은 효율을 나타냄

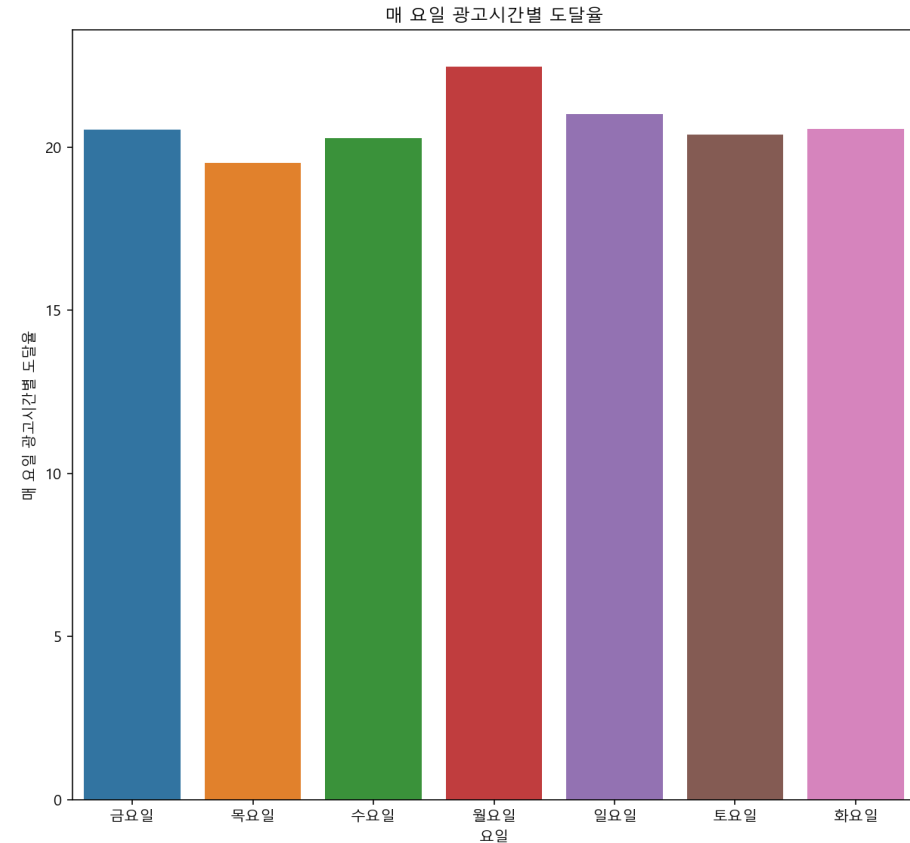


- ✓ 광고길이가 길어질수록 효율이 떨어짐

캠페인 기간 및 요일별 분석



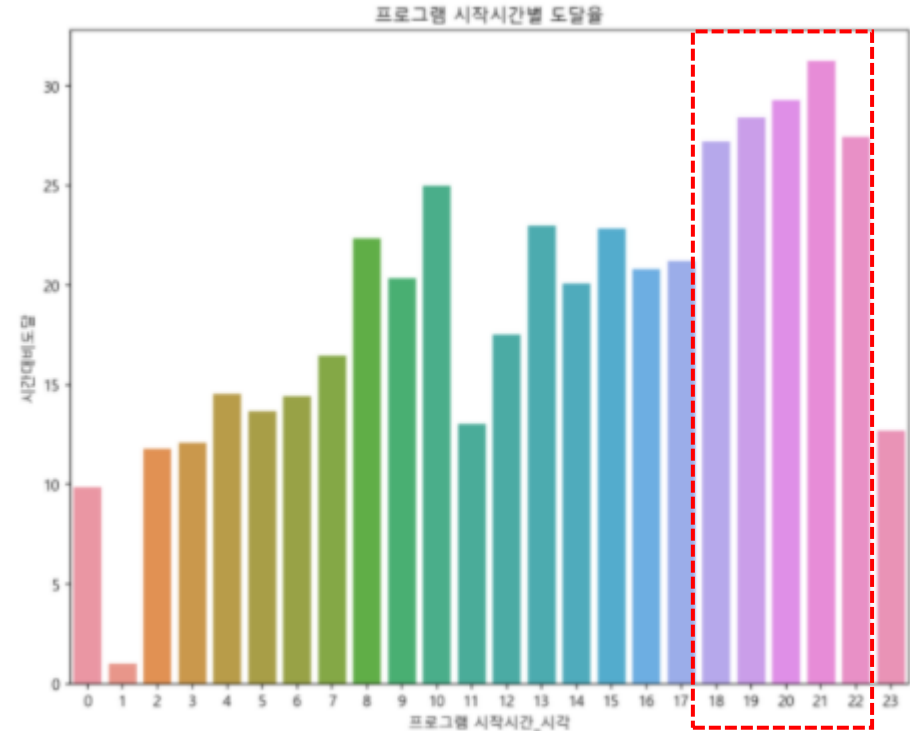
- ✓ 행사 당일 2~3주 전에 집행한 광고 효율이 가장 높음
- ✓ 후순위로는 1주 전 광고



- ✓ 요일별로는 월요일이 가장 효율이 좋고 목요일이 가장 낮았으며, 그 외는 큰 차이 없음

채널 및 프로그램 시간 분석

	채널_코드	광고길이_시간길이	도달율1+_율	시간대비도달
4	KBS2	7800	228553	29.30167
7	YTN	148300	3798555	25.61399
8	연합뉴스TV	235420	5734303	24.35776
2	KBS JOY	20550	463407	22.55022
6	SBS	3220	69018	21.43416
5	MBC	1560	24858	15.93462
0	JTBC Golf	126000	1823878	14.47522
1	JTBC2	126080	1513113	12.00121
3	KBS N SPORTS	48000	370737	7.72369



- ✓ 9개의 채널에 광고 집행
- ✓ 가장 효율이 높은 채널은 **KBS2, YTN**으로 확인
- ✓ 가장 효율이 낮은 채널은 **JTBC2, KBS N SPORTS**로 확인
- ✓ 광고 집행 금액을 데이터상 확인할 수 없지만, 만약 동일하다는 가정을 할 경우 해당 채널에는 광고를 집행하지 않는 것 추천

- ✓ 광고 주목도는 저녁시간대 가장 높은(18~22시)

프로그램 분석

총 147개의 프로그램

	광고 프로그램_이름	채널_코드	광고길이_시간길이	도달율1+_율	시간대비도달
32	SBB왕금차트(재)	KBS JOY	2250	103038	45.79467
52	개는 훌륭하다	KBS JOY	1800	77069	42.81611
33	YTN24(0900)	YTN	3000	123434	41.14467
129	이브닝뉴스	YTN	4200	165378	39.37571
35	YTN뉴스와이드(1000)	YTN	2400	93749	39.06208
97	더뉴스	YTN	8400	312979	37.25940
134	차트를 달리는 남자(재)	KBS JOY	2400	86724	36.13500
131	일요일아이드	연합뉴스TV	2400	73156	30.48167
14	6면시선뉴스큐브	연합뉴스TV	14820	442972	30.29904
113	슈퍼맨이 돌아왔다(재)	KBS JOY	3000	90626	30.20867

✓ 광고 많은 프로그램

- 아는형님 > 1호가될순없어 > 라이브레슨70 > YTN뉴스특보 > 뉴스포커스

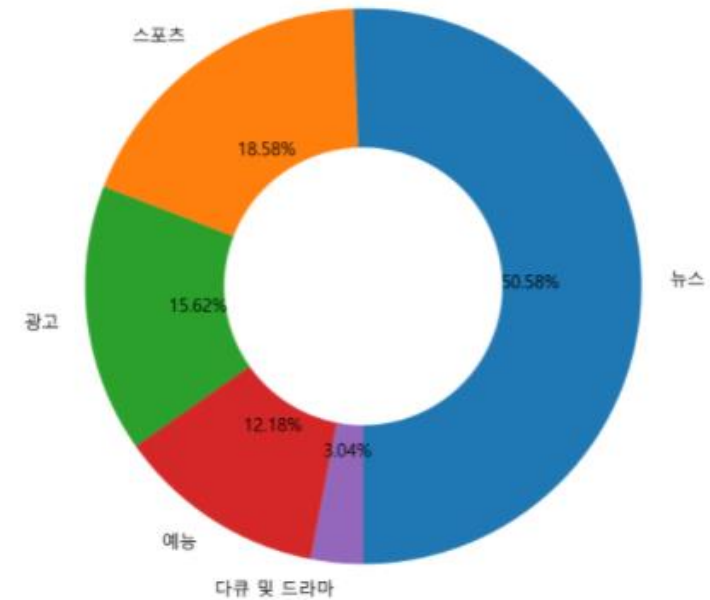
✓ 도달율

- 뉴스워치 > 1호가될순없어 > 6면시선뉴스큐브 > 뉴스포커스

✓ 광고시간대비 도달 효율

- SBB왕금차트(재) > 개는 훌륭하다 > YTN24(0900) > 이브닝뉴스

전체 TV프로그램에 대한 데이터가 아니므로 경향성 파악을 통한 인사이트를 얻기 위해 프로그램 카테고리화



✓ 뉴스 > 스포츠 > 순수광고 > 예능 > 다큐&드라마

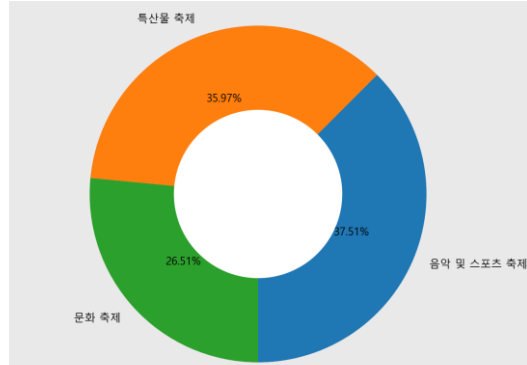
정리

- ✓ 하반기 중 10월에 가장 많은 TV광고 집행 → 21년 10월의 광고 성과 예측
- ✓ '순수광고' 효율이 가장 높았으며, 프로그램 전/후 비교에서는 '프로그램 후'가 가장 높게 나타남
- ✓ 광고길이가 길어질수록 효율이 떨어짐 → 짧고 임팩트있는 '숏폼 콘텐츠'로 제작할 필요성 있음
- ✓ 행사 당일 '2~3주 전'에 집행한 광고 효율이 가장 높음
- ✓ 요일별로는 '월요일'이 가장 효율이 좋고 목요일이 가장 낮음
- ✓ 가장 효율이 높은 채널은 'KBS2, YTN' 낮은 채널은 JTBC2, KBS N SPORTS
- ✓ 광고 주목도는 '저녁시간대' 가장 높음(18~22시)
- ✓ 뉴스 > 스포츠 > 순수광고 > 예능 > 다큐&드라마 순으로 광고효율이 높음

TV광고 효율측정 모델

Feature engineering

① 축제 브랜드명을 'Food', 'Culture', 'Music_sports' 로 카테고리화

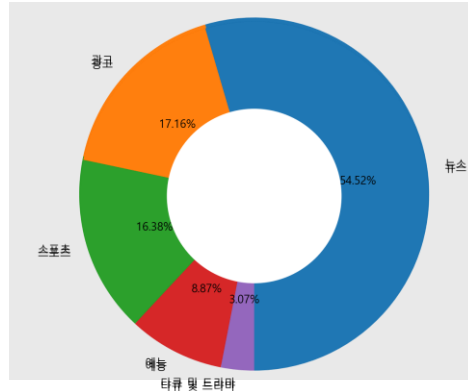


food = ['보은대추축제', '온라인금산인삼축제', '강경젓갈축제', '회인대추장터',
'횡성한우축제', '홍천인삼한우온라인할인판매행사', '명품전주복숭아큰잔치',
'영주풍기인삼축제', '전주한지패션대전', '괴산고추축제', '괴산김장축제',
'향수옥천포도복숭아온라인축제', '금산인삼축제', '고추구기자청양농특산물온라인기획전']

culture = ['옥천지용제', '전주한복오감', '위아자나눔장터', '세계개천문화대축제',
'한산모시', '산청한방약초축제', '울산옹기축제', '온라인보령머드축제',
'진안고원촌스런마을여행']

music_sports = ['서울드림페스티벌', '전주세계소리축제', '여수해양레저스포츠무료체험',
'웰컴대학로', '온한류축제', '집콕방콕콘서트', '포춘밸리페스티벌']

② 광고 프로그램 이름을 'Sports', 'Funny', 'News', 'Drama', 'AD'로 카테고리화



sport = ['에브리샷', '경기', '배구', '축구', 'PGA', '씨름', '야구', '대회', '클럽', 'KBO', 'TOUR', '골프', '농구', '레슨',
'GTOURWOMENS', '하이라이트', '올림픽', '프로', '맘스터치레전드빅매치', '리그', '스포츠', '베이스볼']

funny = ['신상출시편스토랑', '옥탑방의문제아들', '히든싱어', '싱어게인', '방구석1열', '개는 훌륭하다', '갬성캠핑',
'몽쳐야찬다', '아는형님', '슈퍼맨', '사장님', '코미디', '무엇이든 물어보살', '차트를달리는남자', '이십세기Hits',
'SBB황금차트', '서울엔우리집이없다', '1호가될순없어', '몽쳐야찬다']

news = ['뉴스', '특보', 'YTN', 'LIVE', '와이드', '출발640', '증권', '라이브']

drama = ['18어게인', '연애', '허쉬', '우리사랑했을까', '다큐', '우아한친구들', '경우의수', '사생활']

ad = ['광고']

③ 채널코드를 같은 방송국끼리 묶어줌

연한뉴스TV	11625
VTN	6942
JTBC Golf	2250
KBS N SPORTS	2190
JTBC2	1850
KBS JOV	1170
KBS2	332
SBS	144
MBC	52

문자 Data 인코딩

One-hot encoding

Label encoding

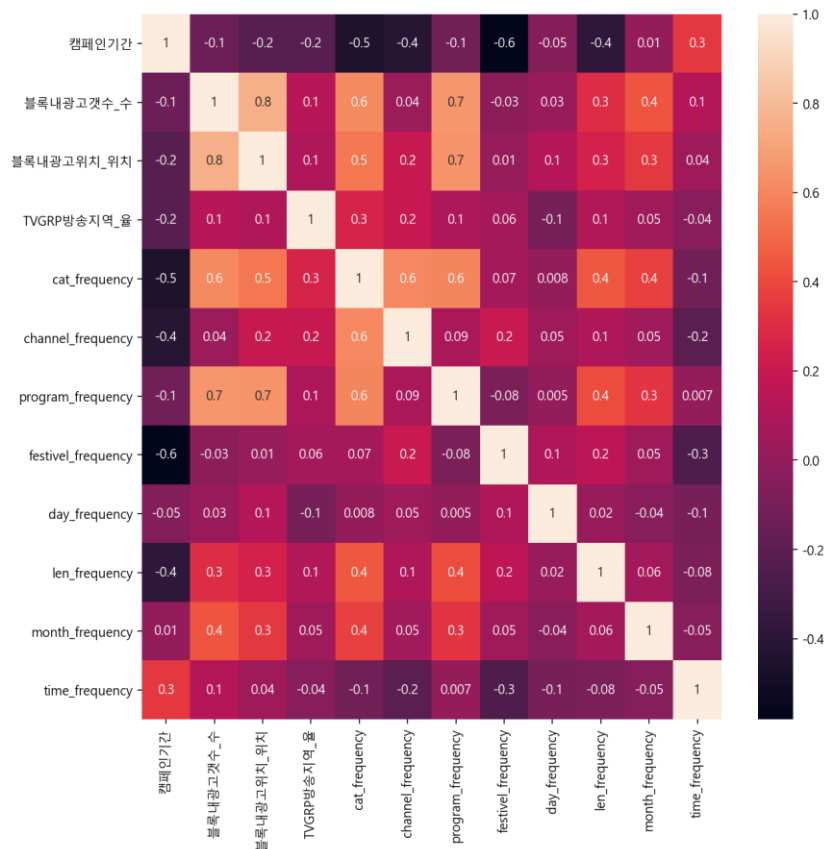
머신러닝 성능 低

Frequency encoding

	캠페인 기간	블록내 광고개 수_수	블록내 광고위 치_위치	TVGRP 방송지 역_플	cat_frequency2	channel_frequency2	program_frequency2	festivel_frequency2	day_frequency2	len_frequency2	r
4669	-0.64573	-0.28445	-0.83064	-0.60159	0.16559	0.15167	0.53174	0.38986	0.13807	0.52263	
1786	-0.57020	0.19282	1.10416	-0.60159	0.53473	0.43809	0.53174	0.26455	0.15283	0.52263	
5465	-0.64573	-0.68218	-0.95963	-0.60159	0.16559	0.15167	0.53174	0.38986	0.15017	0.52263	
2621	0.14730	-1.15946	-0.95963	0.75233	0.17388	0.43809	0.17437	0.38986	0.15017	0.15316	
9640	-0.23033	1.46556	-0.05672	-0.00726	0.53473	0.25079	0.53174	0.38986	0.13807	0.52263	
...
1598	-0.26810	-1.00037	-0.95963	1.59923	0.17388	0.15167	0.17437	0.26455	0.15017	0.14537	
6	-0.57020	-0.20491	0.45923	0.08288	0.53473	0.43809	0.53174	0.26455	0.15283	0.52263	
2858	0.14730	-1.15946	-0.95963	-0.32555	0.17388	0.43809	0.17437	0.38986	0.17288	0.15316	
18463	-0.38139	2.26102	1.74910	-0.23729	0.53473	0.25079	0.53174	0.34560	0.17288	0.52263	
9371	-0.23033	-0.28445	-0.70166	1.57106	0.53473	0.43809	0.29289	0.38986	0.15482	0.52263	

상관계수/다중공선성

① 피어슨 상관계수를 통해 피처간의 상관계수 파악



② 다중공선성 평가를 위해,

VIF(Variance Inflation Factors, 분산팽창요인) 파악

→ VIF값이 10이상이면 머신러닝 효율 저해

	VIF Factor	features
0	2.24492	캠페인기간
1	3.97307	블록내광고갯수_수
2	2.89774	블록내광고위치_위치
3	1.10557	TVGRP방송지역_율
4	24.97464	cat_frequency
5	19.21032	channel_frequency
6	18.50553	program_frequency
7	60.59443	festivel_frequency
8	46.11736	day_frequency
9	6.69381	len_frequency
10	9.41495	month_frequency
11	12.18452	time_frequency

③ 해결을 위해 PCA와 정규화 진행

→ PCA는 성능이 떨어져서 정규화만 진행

모델링

구분

R2_score

선형회귀

```
1 lr_r2 = r2_score(y_test, lr.predict(X_test))  
2 print('r2_score:', lr_r2)
```

r2_score: 0.1273031969180406

Ridge

계수를 제공한 값을 기준으로 규제 적용

```
1 ridge_r2 = r2_score(y_test, ridge.predict(X_test))  
2 print('r2_score:', ridge_r2)
```

r2_score: 0.12724378065521058

Lasso

계수의 절댓값을 기준으로 규제 적용

```
1 lasso_r2 = r2_score(y_test, lasso.predict(X_test))  
2 print('r2_score:', lasso_r2)
```

r2_score: 0.12233444941037319

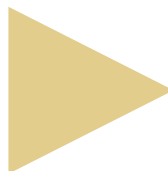
✓ R2_score가 3가지 다 비슷하게 나옴

모델링

Lasso기준 그리드 서치를 이용해서
하이퍼 파라미터를 조정하니
0.127로 높아졌으나 여전히 부족

```
1 lasso3 = Ridge(alpha=0.001)
2 lasso3.fit(X_train, y_train)
3 lasso3_r2 = r2_score(y_test, lasso3.predict(X_test))
4 print('r2_score:', lasso3_r2)
```

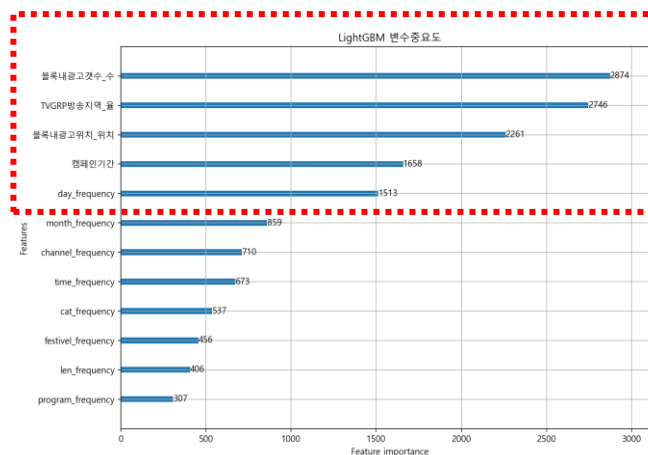
r2_score: 0.1273031397319213



light GBM 활용 0.313까지 상승
→ light GBM 모델 선정

```
1 lgbm_r2 = r2_score(y_test, lgbm_model.predict(X_test))
2 print(lgbm_r2)
```

0.31319453808220477



✓ 중요도 높은 4개 변수를 통해
10월 데이터셋 신규생성을 및 도달 파악

→ 블록 내 광고갯수, 광고삽입위치

→ 광고기간, 광고요일 조정

결론

예측 결과

```
1 y_forecast = lgbm_model.predict(X_predict)
2
3 y_forecast

array([216.97809163, 216.97809163, 441.81483639, ..., 743.81585583,
       421.98388102, 421.98388102])
```

```
1 y_forecast.mean() # 예측값

402.6570886216889
```

```
new_df[new_df["날짜_월"] == 10]["도달율1+_율"].sum() - y_forecast.sum()

-507054.39963483624
```

2020년 10월 광고를 통한 도달량

2021년 예측값

총 500,000회 도달량 개선



상품구성	1 배너 (1029 x 204px)	2 랜딩페이지
	광고주 페이지	
	애드뷰 - 플뷰, 콤팩트뷰	
	채팅방 - 챗봇	
	채팅방 - 특채널 메시지	
	특비즈니스폼	
선택가능한 랜딩페이지	카카오 커머스 플랫폼 - 토크투어, 선물하기, 메이커스	
	#태그 검색결과	
	채널별뷰	
	특채널리더	
	포스트	
최소 입찰 가이드	CPC 10원 / CPM 4,000원	
집행 및 운영 플랫폼	카카오 모먼트	
광고 리포트	모먼트 리포트 제공	

카카오 비즈보드 광고기준

CPM 4,000원(1,000회 노출당 비용)

2,000,000원 광고비용 절감 효과

결론 및 향후 발전방향

✓ 결론

- TV 광고에 영향을 미칠 수 있는 변수는 생각보다 다양(32개 컬럼)
- 현업에서 TV 매체광고 효과와 퍼포먼스를 결정짓는 주요 변수에 대한 의문이 많았는데
본 프로젝트를 통해 영향을 끼치는 요소들을 알 수 있었고 이를 통한 효율 향상 및 비용절감이 가능한 것 확인

✓ 개선방안 및 향후 발전방향

- numeric 데이터를 수집/보완하여 성능 개선 필요(예 : 지역별 총가구수, 시청인구수, 연령)
- 지역축제 뿐만 아니라, 다른 분야의 TV 광고 효율 예측
- 업계에서 성과 측정 지수를 도달(Reach)에만 의존하고 있는데
도달과 실제 전환의 연관성에 대한 파악은 프로젝트를 더 의미있게 만들어줄 것이라고 생각
(계속적인 리서치와 고민 필요)