



오늘의 광고

날씨에 따른 집 단별 맞춤 광고 제작부터 기사까지

Weather Big Data Contest

Wake-Up

김도영 김도현 이다솔
이세령 장예서 조신형

01. 공모 배경

- I. 사람과 날씨
- II. 소비와 광고

02. 데이터 정의와 전처리

- I. 데이터 정의와 활용 근거
- II. 날씨 데이터 전처리
- III. 판매량 데이터 전처리
- IV. SNS 언급량 데이터 전처리

03. 분석과 모델링

- I. 모델 선정
- II. 데이터셋 전처리
- III. 모델 결과
- IV. 앙상블 및 최종 선정
- V. 모델 시연

04. 활용 방안

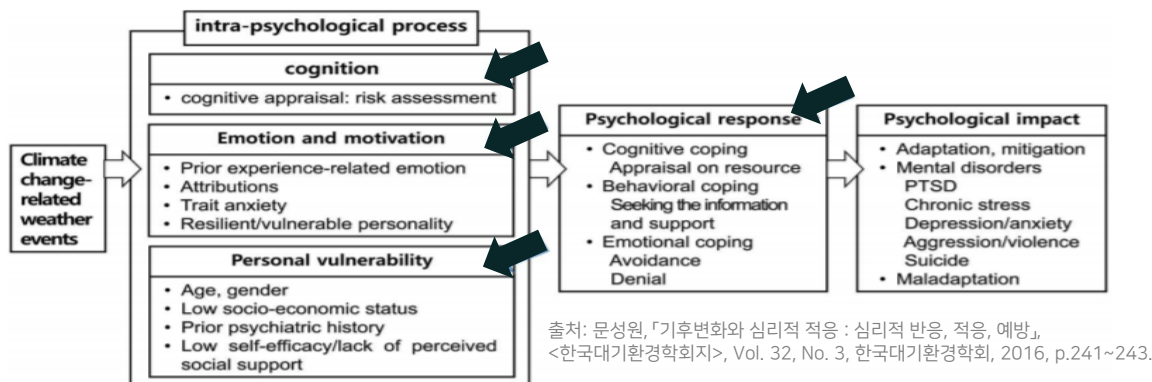
- I. 서비스 개요
- II. 서비스 프로세스 Stage 1
- III. 서비스 프로세스 Stage 2
- IV. 서비스 프로세스 Stage 2 예시
- V. 서비스 프로세스 Stage 3

05. 기대 효과

- I. 기존 광고와의 차별점
- II. 기대 효과 모식도

01. 공모 배경 | 사람과 날씨

기후 변화에 따른 심리적 변화

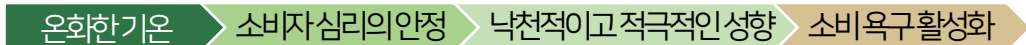


인간의 행동은 정서, 인지와 함께 작용한다. 태도는 행동에 그 행동은 다시 태도에 영향을 주며, 밀접하고 구체적인 연속적 변동성을 가지게 된다. (Myers, 2003)

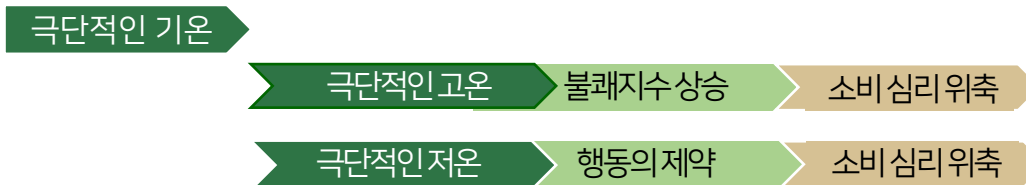
소비 심리에 영향을 주는 온도

A. 온화한 기온이 소비자의 소비 심리에 미치는 영향

출처: 산업 기상 연구소



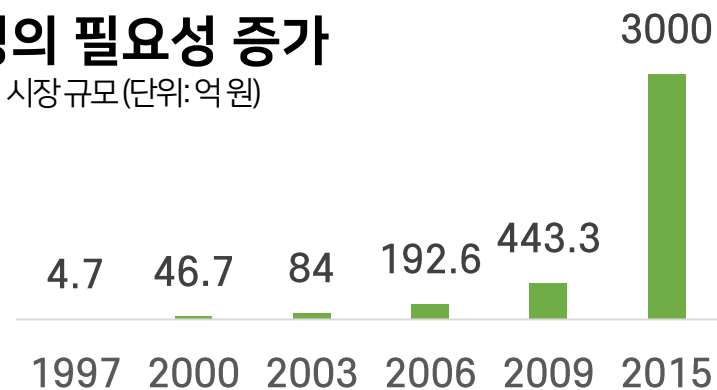
B. 극단적인 기온이 소비자의 소비 심리에 미치는 영향



날씨경영의 필요성 증가

한국 기상산업 시장규모 (단위: 억 원)

자료: 기상청



국내 GDP의 52%와 산업의 70~80%가 날씨에 직·간접적으로 영향 받음
'날씨를 어떻게 관리하느냐'가 기업의 주요 관심사항
날씨정보를 활용해 리스크를 최소화하고 이익을 창출 중

출처: KDI 경제정보센터

날씨경영 사례



파리바게트 | 날씨 판매지수 도입

기온 27도 이상의 맑은 날씨에는 샌드위치,
비가 오는 20도 안팎의 날씨에는 기름기가 많은
조리빵(피자빵, 소시지빵 등)이 잘 팔린다는 데이터 제공



GS25 | 가맹주에게 발주 시 참고용 날씨 정보 제공

기온이 0~5도 일 때는 양주,
6~10도일 때는 맥주가 더 잘 팔린다.
점주들은 본사에서 제공하는 날씨 정보를 발주 시 참고

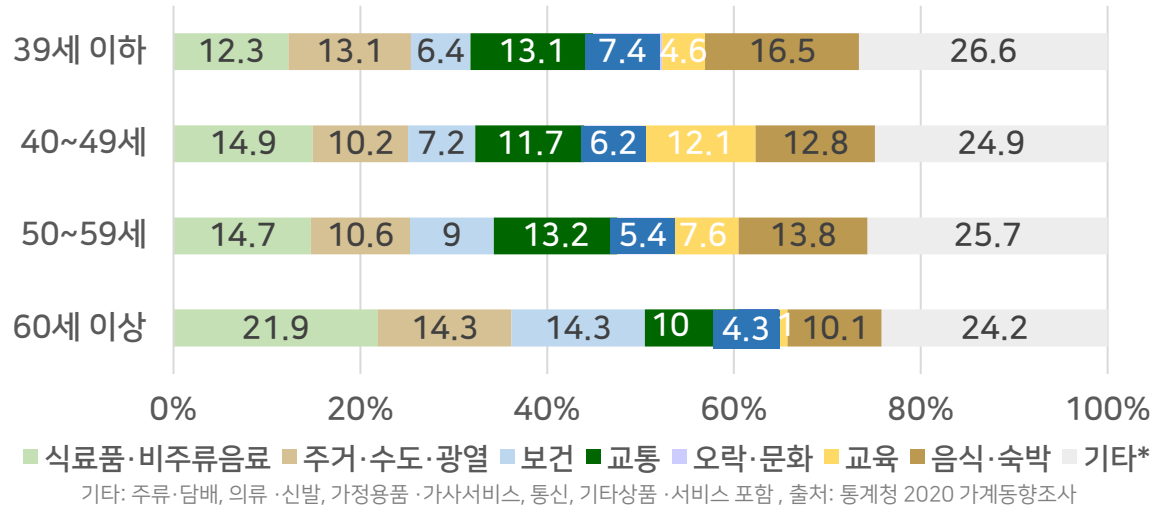


CJ제일제당 | 곡물 가격 변동을 대비한 원가 절감에 활용

곡물가의 급변동을 사전에 대비, 구매 시기 조정
날씨를 전문적으로 분석하는 곡물구매조직을 운영

01. 공모 배경 | 소비와 광고

가구주 연령별 소비지출 구성비



연령대별 소비 특징

| 남성 | | 여성 | |
|--------|----------------------------|--------|----------------------------|
| 20대 초반 | 주점 | 20대 초반 | 인터넷 쇼핑 |
| 30대 초반 | 차량 유지비 인터넷 쇼핑 비중 증가 | 30대 초반 | 영유아 자녀를 위한 인터넷쇼핑과 유아 교육 |
| 30대 후반 | 재테크 시작 | 30대 후반 | 자녀교육, 인터넷쇼핑 |
| 40대 초반 | 건강, 재테크 | 40대 초반 | 자녀교육, 문화·여가생활 |
| 40대 후반 | 재테크 비중 축소 | 40대 후반 | 의료비 지출비중 증가 |
| 50대 이상 | 다시 나를 찾는 소비, 차량지출 비중 증가 | 50대 이상 | 다시 나를 찾는 소비, 의류 소비비중 증가 |

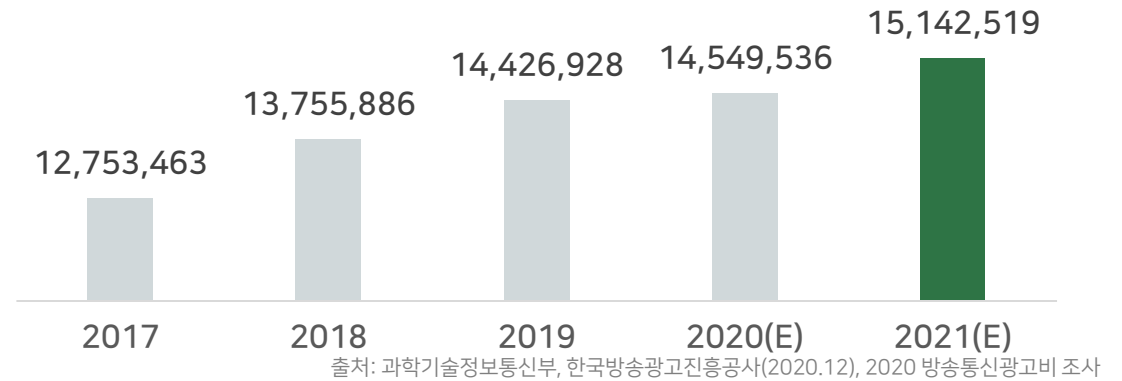
출처: 매일경제신문

⇒ 연령별·성별 소비 특징이 상이하다.

광고 산업의 성장

국내 전체 광고비 시장 규모

(단위: 억 원)



- ✓ 지난 10년간 한국의 광고시장은 꾸준히 성장
- ✓ 2020년 한국 광고시장의 전년대비 총 광고비 성장률은 0.8%로 추정
- ✓ 커지는 광고 시장에서 효율적인 광고 전략 필요성 증대

날씨는 기분에 영향을 주고, 기분은 의사결정에 영향을 준다.

날씨는 물품 구매에 영향을 주며, 날씨 경영의 효과는 입증되었다.

연령별·성별로 소비성향은 다르다.

광고 산업은 성장세를 그리며, 더 효율적인 광고 방안이 필요하다.



“연령·성·날씨를 고려해 광고 서비스를 제공하면 어떨까?”

02. 데이터 정의와 전처리 | 데이터 정의와 활용 근거

본 분석에서는 크게 1) 날씨 데이터, 2) 구매이력 데이터, 3) 소셜 데이터를 사용하였다.

| 항목 | 종관기상관측 일별 데이터 | 대기환경 정보 데이터 | 상품별 온라인 구매이력 데이터 | SNS에서의 상품명 언급량 데이터 |
|----------|---|---|---|---|
| 상세 내용 | 평균기온, 일교차, 일강수량, 평균 풍속, 평균 상대습도, 평균 전운량 | 오존, 미세먼지 | 뷰티, 식품, 냉난방가전으로 분류되는 383개의 품목의 연령 성별별 판매량 | 블로그, 커뮤니티, 인스타그램의 각 채널별 게시글 10만건 당 물품 키워드가 포함된 게시글의 상대적 문서수 |
| 기간 | 2018.01.01 ~ 2019.12.31 (2년) | | | |
| 출처 |  기상청 |  |  |  |

INPUT

종관기상관측 데이터

- 날씨에 따른 매출 분석을 위해
온라인 구매이력 데이터와 같은 기간의 날씨 데이터 사용
- 인구수 상위 8개 도시가 전체 인구수의 46%, 대표값으로 사용
 - 상위 8개 도시 = 서울, 부산, 인천, 대구, 대전, 광주, 울산, 수원
 - 백령도, 강화, 관악산 : 그 지역을 대표하기 어렵다고 판단, 제외
 - 대구(기) : 지점 상세보기 정보가 나와있지 않아 제외

대기환경 정보 데이터

- 날씨와 더불어 대기환경 또한 매출에 영향을 끼칠 것이라 판단하여
날씨 데이터에 병합하여 사용
 - 종관기상관측 데이터와 동일하게 상위 8개 도시의 데이터 사용

OUTPUT

상품별 온라인 구매이력 데이터

- 날씨에 따른 판매량 예측을 위해 사용

SNS 데이터

- 온라인 구매라는 특성상, 상품이 SNS에 많이 언급될 수록
매출에 도움을 줄 것이라 가정하여 함께 분석하기 위해 사용

02. 데이터 정의와 전처리 | 날씨 데이터 전처리

날씨 데이터의 전처리는 1) 대기환경 데이터 전처리, 2) 종관관측 데이터 전처리, 3) 데이터 병합, 4) 지역별 가중치 적용 순으로 진행되었다.

대기환경 정보 데이터

종관기상관측 데이터

결측치 처리 후 병합



지역별 인구수 가중치



날씨 데이터

1. 대기환경 데이터

1. 전체 대기 요소 중, 에어코리아의 예보에 사용되는 O3, PM10 데이터만 추출
2. 인구수 상위 8개 도시 [서울, 부산, 인천, 대구, 대전, 광주, 울산, 수원]의 데이터 추출, 결측치 없음 -> 완료

| 지역 | 망 | 측정소코드 | 측정소명 | 측정일시 | S02 | CO | O3 | NO2 | PM10 | PM25 | 주소 | |
|----|-------|-------|--------|------|------------|-------|-----|-------|-------|------|------|---------------|
| 0 | 서울 중구 | 도시대기 | 111121 | 중구 | 2019010101 | 0.003 | 0.7 | 0.003 | 0.054 | 39.0 | 24.0 | 서울 중구 덕수궁길 15 |
| 1 | 서울 중구 | 도시대기 | 111121 | 중구 | 2019010102 | 0.003 | 0.8 | 0.002 | 0.056 | 38.0 | 27.0 | 서울 중구 덕수궁길 15 |
| 2 | 서울 중구 | 도시대기 | 111121 | 중구 | 2019010103 | 0.003 | 0.9 | 0.002 | 0.057 | 42.0 | 28.0 | 서울 중구 덕수궁길 15 |
| 3 | 서울 중구 | 도시대기 | 111121 | 중구 | 2019010104 | 0.003 | 0.8 | 0.002 | 0.054 | 42.0 | 31.0 | 서울 중구 덕수궁길 15 |
| 4 | 서울 중구 | 도시대기 | 111121 | 중구 | 2019010105 | 0.003 | 0.8 | 0.002 | 0.048 | 49.0 | 33.0 | 서울 중구 덕수궁길 15 |



| 지역 | 날짜 | 오존(O3) | 미세먼지(PM10) |
|-----|---------------|----------|------------|
| 0 | 수원 2018-01-01 | 0.008958 | 42.766234 |
| 1 | 수원 2018-01-02 | 0.007590 | 41.896970 |
| 2 | 수원 2018-01-03 | 0.010512 | 34.670886 |
| 3 | 수원 2018-01-04 | 0.005339 | 45.394558 |
| 4 | 수원 2018-01-05 | 0.008345 | 54.585987 |
| ... | ... | ... | ... |

2. 종관 관측 데이터

1. 8개 도시의 데이터 추출
2. 일 강수량과 풍속의 결측치 처리
일 강수량 : 0.1 mm 미만 강수량이 NaN으로 기입되어 있어, 0으로 대체
풍속 : 결측치는 인근 지역의 데이터로 대체 (인천 -> 파주, 대전 -> 금산, 대구 -> 영천, 울산 -> 부산)
3. 기상청의 기상예보와 종관관측 데이터의 [공통 요소]와 최고기온에서 최저기온을 뺀 값인 [일교차]를 최종 피처로 선정 -> 완료

| 지역 | 날짜 | 평균기온(* C) | 일교차(* C) | 일강수량(mm) | 평균 풍속(m/s) | 평균 상대습도(%) | 평균 전운량(1/10) |
|----|---------------|-----------|----------|----------|------------|------------|--------------|
| 0 | 서울 2018-01-01 | -1.3 | 8.9 | 0.0 | 1.4 | 39.1 | 1.0 |
| 1 | 대전 2018-01-01 | -0.7 | 10.2 | 0.0 | 1.0 | 60.0 | 0.5 |

02. 데이터 정의와 전처리 | 날씨 데이터 전처리

날씨 데이터의 전처리는 1) 대기환경 데이터 전처리, 2) 종관관측 데이터 전처리, 3) 데이터 병합, 4) 지역별 가중치 적용 순으로 진행되었다.

3. 전처리가 완료된 대기환경 데이터와 종관관측 데이터를 병합

4. 지역별 인구수의 가중치를 구하여 날씨 데이터의 각 요소에 적용 -> 최종 완료

대기환경 정보 데이터

종관기상관측 데이터

결측치 처리 후 병합



지역별 인구수 가중치



날씨 데이터

지역별 인구수 가중치

$$\begin{aligned} &\text{전체 인구수 중 각 도시가 차지하는 인구수의 비율} \\ &= \frac{\text{각 도시의 인구수}}{\text{8개 도시의 인구수의 합}} \end{aligned}$$

※ 인구 데이터 기준연도
: 2018~19년

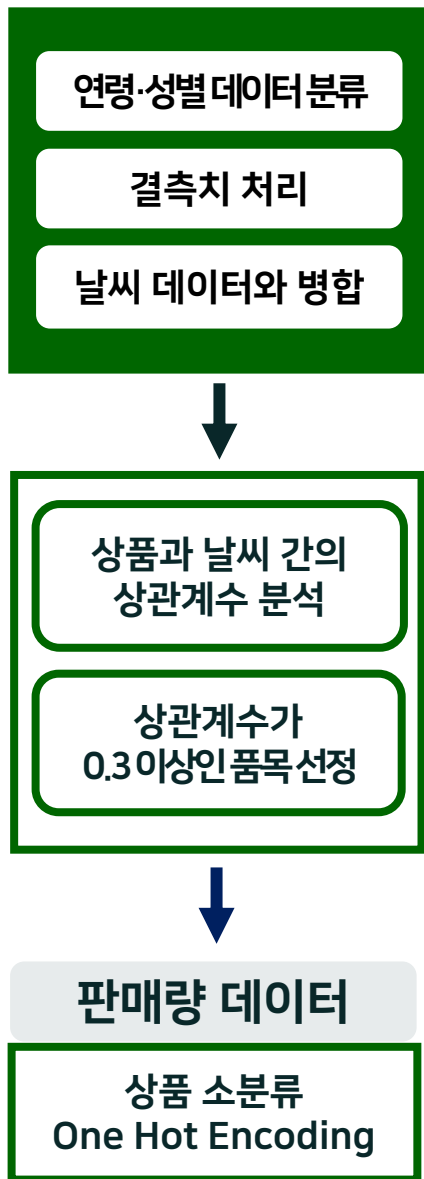
(예) 평균 기온의 경우,
(서울인구가중치*서울평균기온) + (부산인구가중치*부산평균기온) + ... + (수원인구가중치*수원평균기온) = 전국 평균 기온

날짜 평균기온(° C) 일교차(° C) 일강수량(mm) 평균 풍속(m/s) 평균 상대습도(%) 평균 전운량(1/10) 오존(O3) 미세먼지(PM10)

| | | | | | | | | |
|------------|-----------|----------|----------|----------|-----------|----------|----------|-----------|
| 2018-01-01 | 0.066157 | 8.198700 | 0.000000 | 1.844893 | 41.267581 | 0.731871 | 0.015544 | 44.032069 |
| 2018-01-02 | -0.000201 | 7.910404 | 0.000000 | 2.003189 | 47.229263 | 2.787500 | 0.013179 | 53.742579 |

02. 데이터 정의와 전처리 | 판매량 데이터 전처리

판매량 데이터의 전처리는 1) 연령·성별 데이터 분류, 2) 결측치 처리, 3) 날씨 데이터와 병합, 4) 날씨와 판매량의 상관관계 분석 및 최종 품목 선정, 5) 물품별 원-핫 인코딩 순으로 진행되었다.



1. 연령·성별로, 각 날짜마다 10개의 동일 상품 판매량 존재 -> 10개의 컬럼으로 바꿔주었다.

예) 10개의 행으로 존재하는 2018.01.01의 공기청정기 판매량

| | 날짜 | 성별 | 연령 | 대분류 | 소분류 | 판매량 |
|---|----------|----|----|-------|-------|-----|
| 0 | 20180101 | F | 20 | 냉난방가전 | 공기청정기 | 1 |
| 1 | 20180101 | F | 30 | 냉난방가전 | 공기청정기 | 28 |
| 2 | 20180101 | F | 40 | 냉난방가전 | 공기청정기 | 16 |
| | | | | ... | | |
| 7 | 20180101 | M | 40 | 냉난방가전 | 공기청정기 | 18 |
| 8 | 20180101 | M | 50 | 냉난방가전 | 공기청정기 | 4 |
| 9 | 20180101 | M | 60 | 냉난방가전 | 공기청정기 | 3 |

| 날짜 | 대분류 | 소분류 | 20대 여성 판매량(개) | 20대 남성 판매량(개) | ... | 60대 여성 판매량(개) | 60대 남성 판매량(개) |
|------------|-------|-------|---------------|---------------|-----|---------------|---------------|
| 2018-01-01 | 냉난방가전 | 공기청정기 | 1.0 | 2.0 | ... | 2.0 | 3.0 |

품목 1개의 일별 판매량을 연령 성별별로 1행에 담기도록 정렬

2. 결측치 (NaN) 은 0으로 대체

3. 앞서 전처리한 날씨 데이터와 판매량 데이터를 날짜 기준으로 병합

| 날짜 | 대분류 | 소분류 | 평균기온 (°C) | 일교차 (°C) | 일강수량 (mm) | 평균 풍속 (m/s) | 평균 상대습도 (%) | 평균 전운량 (1/10) | 오존 (03) | 미세먼지 (PM10) | 20대 여성 판매량 (개) | 20대 남성 판매량 (개) | 30대 여성 판매량 (개) | 30대 남성 판매량 (개) | ... | 60대 여성 판매량 (개) | 60대 남성 판매량 (개) |
|------------|-------|--------|-----------|----------|-----------|-------------|-------------|---------------|----------|-------------|----------------|----------------|----------------|----------------|-----|----------------|----------------|
| 2018-01-01 | 냉난방가전 | 가열식가습기 | 0.066157 | 8.1987 | 0.0 | 1.844893 | 41.267581 | 0.731871 | 0.015544 | 44.032069 | 1.0 | 1.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 |

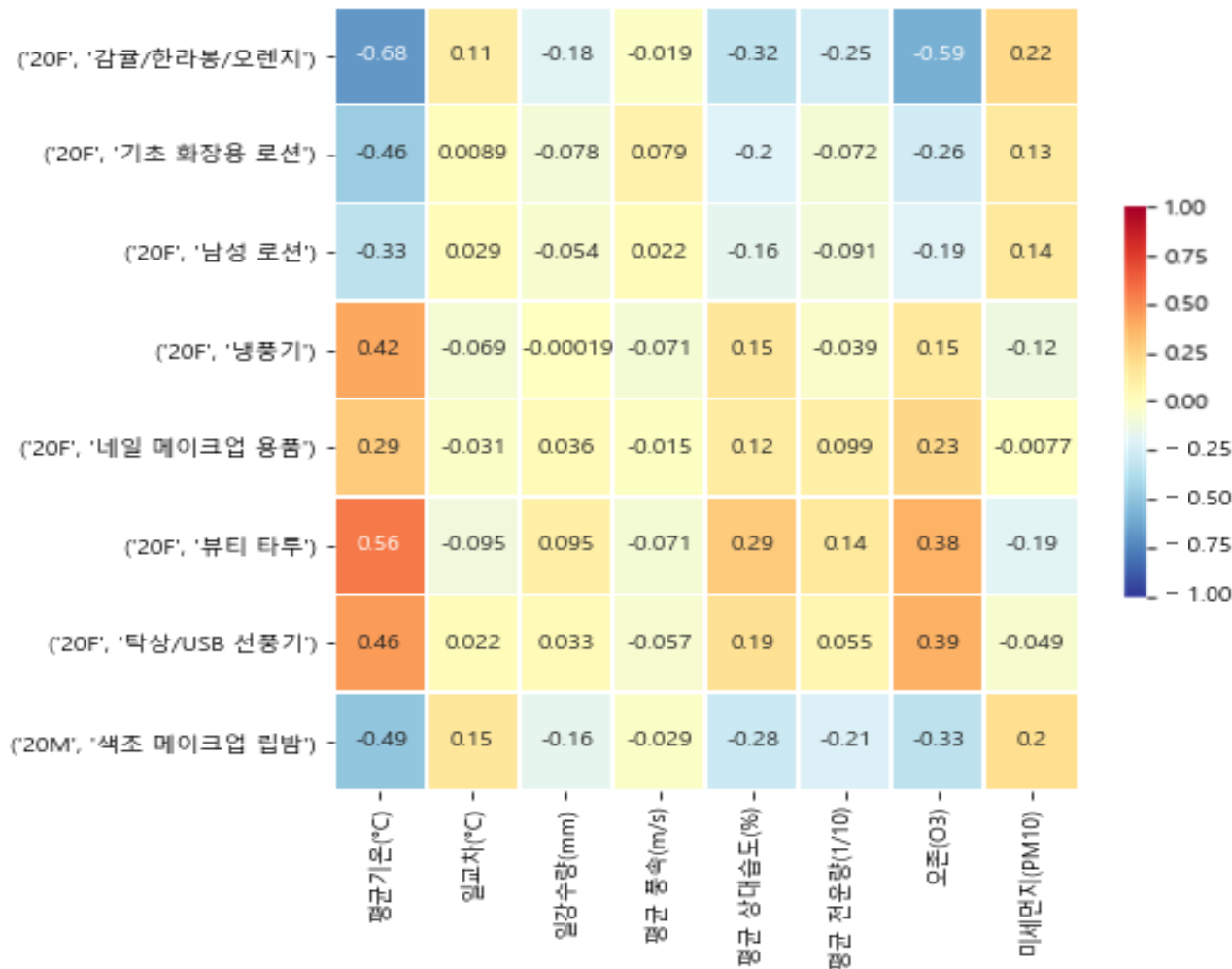
02. 데이터 정의와 전처리 | 판매량 데이터 전처리

판매량 데이터의 전처리는 1) 연령·성별 데이터 분류, 2) 결측치 처리, 3) 날씨 데이터와 병합, 4) 날씨와 판매량의 상관관계 분석 및 최종 품목 선정, 5) 물품별 원-핫 인코딩 순으로 진행되었다.



4. 날씨와 연령·성별 판매량 간의 상관관계 계산

어떤 소비자 집단에서도 0.3 이상의 수치를 가지지 않는 품목 제외,
383개 품목 중 126개 선정



5. 소분류(물품명)을 기준으로 원-핫 인코딩 -> 완료

Diagram showing the mapping from weather data to product sub-category one-hot encoding. The '날씨' (Weather) data is mapped to '소분류' (Sub-category) via '대분류' (Major category).

| 날씨 | 대분류 | 소분류_가열식 가습기 | 소분류_감귤/한라봉/오렌지 |
|------------|-------|-------------|----------------|
| 2018-01-01 | 냉난방가전 | 1 | 0 |

02. 데이터 정의와 전처리 | SNS 언급량 데이터 전처리

SNS 언급량 데이터의 전처리는 1) 판매량 데이터와 동일한 품목 추출과 병합, 2) 물품별 원-핫 인코딩 순으로 진행되었다.

1. 앞서 상관계수를 기준으로 선별한 126개의 품목의 데이터를 추출하고 날씨 데이터와 병합

| 날짜 | 대분류 | 소분류 | 평균기온 (°C) | 일교차 (°C) | 일강수량 (mm) | 평균 풍속 (m/s) | 평균 상대습도 (%) | 평균 전운량 (1/10) | 오존 (O3) | 미세먼지 (PM10) | SNS언급량 |
|------------|-----|--------------|--------------|-------------|--------------|----------------|----------------|------------------|------------|----------------|-----------|
| 2018-01-01 | 뷰티 | 기능성 링클케어 화장품 | 0.066157 | 8.1987 | 0.0 | 1.844893 | 41.267581 | 0.731871 | 0.015544 | 44.032069 | 12.154295 |
| 2018-01-01 | 뷰티 | 기능성 모공관리 화장품 | 0.066157 | 8.1987 | 0.0 | 1.844893 | 41.267581 | 0.731871 | 0.015544 | 44.032069 | 36.000828 |
| 2018-01-01 | 뷰티 | 기능성 아이케어 화장품 | 0.066157 | 8.1987 | 0.0 | 1.844893 | 41.267581 | 0.731871 | 0.015544 | 44.032069 | 0.895782 |

동일 품목추출

날씨 데이터와 병합

2. 물품별 원-핫 인코딩 -> 완료

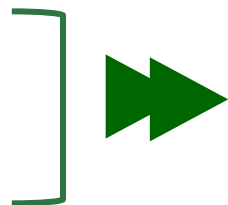
| 날짜 | 대분류 | 소분류_가공란 | 소분류_가스온수기 | 소분류_가열식가습기 | 소분류_가자미 | 소분류_갈비/찜/바비큐용돈육 | ... | 소분류_히터 | 평균기온 (°C) | 일교차 (°C) | 일강수량 (mm) | 평균 풍속 (m/s) | 평균 상대습도 (%) | 평균 전운량 (1/10) | 오존(O3) | 미세먼지 (PM10) | SNS언급량 |
|------------|-----|---------|-----------|------------|---------|-----------------|-----|--------|--------------|-------------|--------------|----------------|----------------|------------------|----------|----------------|-----------|
| 2018-01-01 | 뷰티 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0.066159 | 8.198689 | 0.0 | 1.844892 | 41.267527 | 0.73187 | 0.015544 | 44.032018 | 12.154295 |
| 2018-01-01 | 뷰티 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0.066159 | 8.198689 | 0.0 | 1.844892 | 41.267527 | 0.73187 | 0.015544 | 44.032018 | 36.000828 |

SNS 언급량 데이터

상품 소분류
One Hot Encoding

목표 모델

- ☑ 성별·연령별 판매량을 예측하는 회귀모델
- ☑ SNS 언급량을 예측하는 회귀모델



정형 데이터이므로

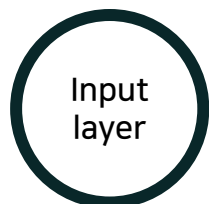
머신러닝, 딥러닝 모두 시도

머신러닝

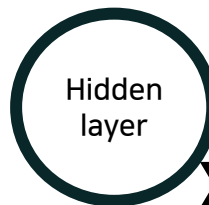
- Linear Regression 계열
 - Linear Regression
 - Ridge Regression
 - Lasso Regression
- Boosting 계열
 - Light GBM
 - XGBoost
 - CatBoost

딥러닝 _ DNN

- DNN : 딥러닝의 기본 구조

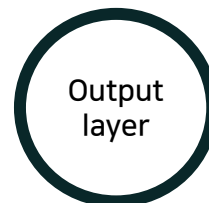


- 134개의 node



- 64개의 node
- Batch normalization
- Relu 함수 적용

X2



- 판매량 예측 모델 : 10개의 node
- SNS 언급량 예측 모델 : 1개의 node

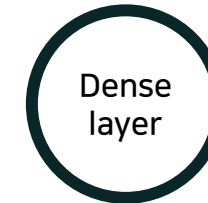
딥러닝 _ LSTM

- 긴 의존 기간을 필요로 하는 시계열 데이터 학습에 적합
- 따라서 시계열 데이터인 기상 데이터 학습에 적합



X5

- 5층의 LSTM layer
- 100개의 node
- 40%의 Dropout 적용



- 판매량 예측 모델 : 10개의 node
- SNS 언급량 예측 모델 : 1개의 node

03. 분석과 모델링 | 모델 선정

분석에 사용한 모든 기법으로 성별·연령별 판매량을 예측하는 모델과, SNS 언급량을 예측하는 모델을 각각 구현

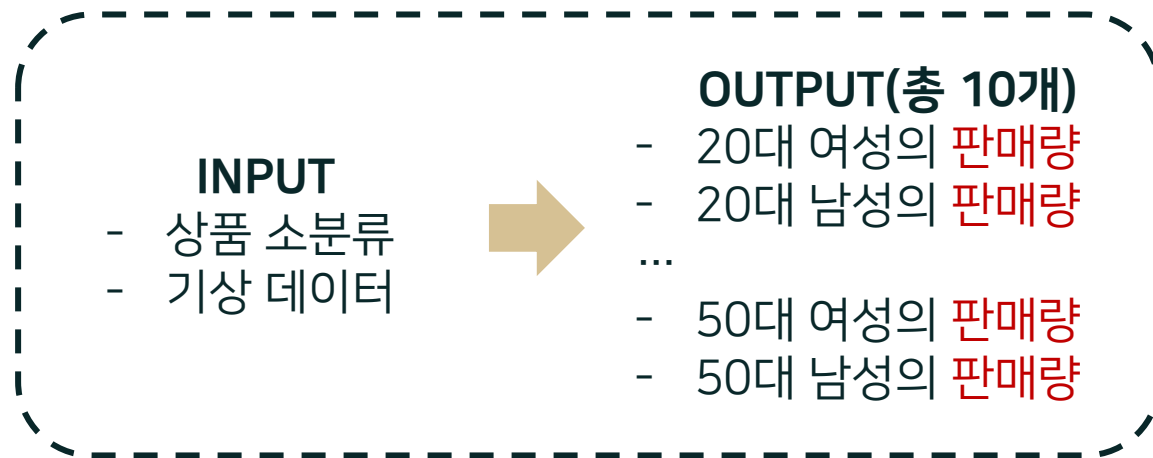
Linear
Ridge
Lasso
Light GBM
XGBoost
CatBoost
DNN
LSTM



성별·연령별 판매량
예측 모델



SNS 언급량
예측 모델



머신러닝·딥러닝(DNN) 전처리

- ☑ DNN은 시계열 특성을 고려하지 않아도 된다.
-> 머신러닝과 동일한 데이터셋 사용
- ☑ 기상 데이터_Min-Max scaler로 정규화
- ☑ 상품 소분류_One Hot Encoding

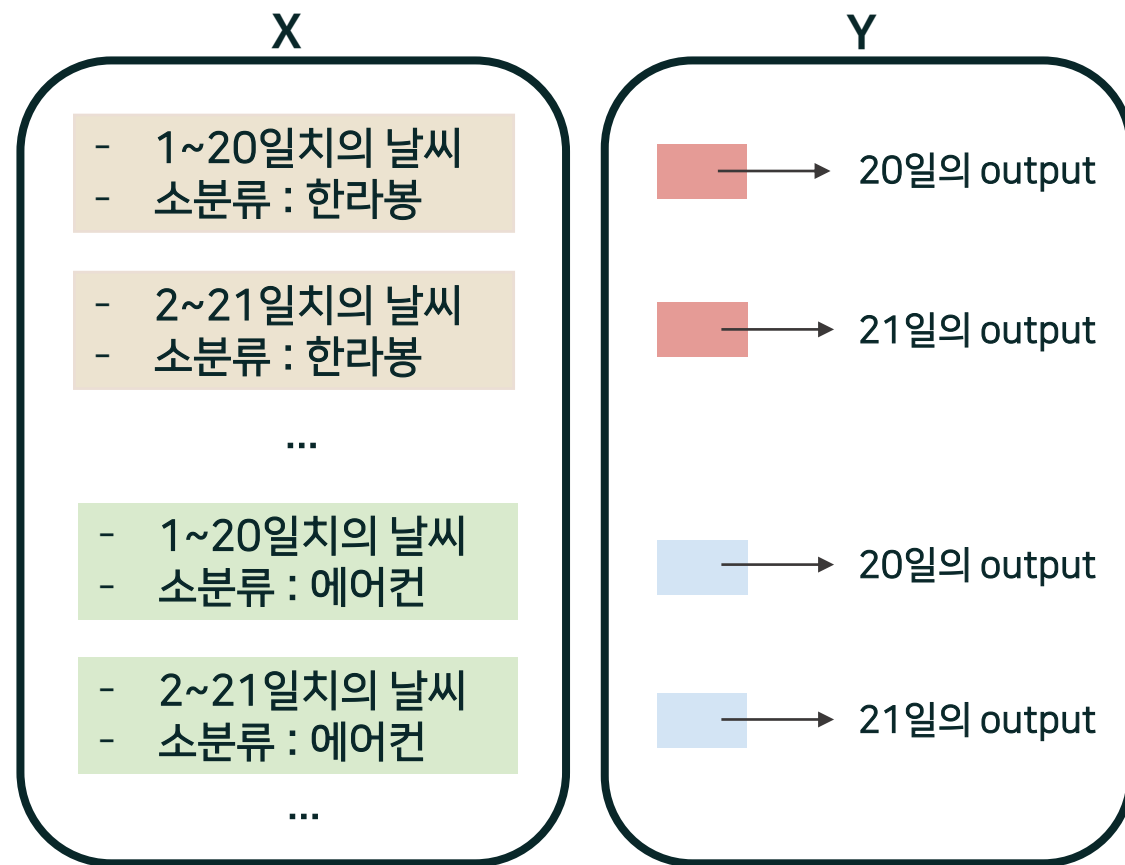
| 소분류_감 | 소분류_굴 | ... | 평균기온 | 강수량 | ... | output |
|-------|-------|-----|------|-----|-----|--------|
| 1 | 0 | ... | 0.4 | 0.7 | ... | 257 |

One Hot
Encoding

Min_Max
Scaling

LSTM 전처리

- ☑ 정규화, One Hot Encoding 동일 적용
- ☑ 예측 시점으로부터 과거 20일치의 기상 데이터로 예측



03. 분석과 모델링 | 모델 결과

성별·연령별 판매량 예측

소수점 넷째자리에서 반올림함

| | Linear Regression | Ridge | Lasso | LGBM | XGBoost | CatBoost | DNN | LSTM |
|------|-------------------|---------|---------|---------|---------|----------|---------|---------|
| MSE | 468.124 | 476.776 | 643.621 | 439.514 | 459.429 | 437.208 | 467.179 | 414.381 |
| RMSE | 21.636 | 21.835 | 25.370 | 20.965 | 21.434 | 20.910 | 21.614 | 20.356 |
| MAE | 8.319 | 8.323 | 13.079 | 7.958 | 9.159 | 7.981 | 7.736 | 7.180 |

SNS 언급량 예측

| | Linear Regression | Ridge | Lasso | LGBM | XGBoost | CatBoost | DNN | LSTM |
|------|-------------------|---------|---------|---------|---------|----------|---------|---------|
| MSE | 817.357 | 744.664 | 835.954 | 674.180 | 626.738 | 650.987 | 549.207 | 472.266 |
| RMSE | 28.589 | 27.289 | 28.913 | 25.965 | 25.035 | 25.514 | 23.435 | 21.732 |
| MAE | 12.095 | 12.282 | 18.480 | 10.480 | 13.338 | 10.524 | 10.249 | 8.849 |

LGBM, Catboost, DNN, LSTM 선정 ➡ 4개 모델로 Ensemble

03. 분석과 모델링 | 앙상블 및 최종 선정

앙상블이란?

- 서로 다른 모델을 결합하는 것
- 각 모델의 예측 결과를 산술평균

성별·연령별 판매량 예측

| | LGBM | CatBoost | DNN | LSTM | Ensemble |
|------|---------|----------|---------|---------|----------|
| MSE | 439.514 | 437.208 | 441.137 | 414.380 | 405.364 |
| RMSE | 20.965 | 20.910 | 21.003 | 20.356 | 20.134 |
| MAE | 7.958 | 7.981 | 7.805 | 7.180 | 7.322 |

최종 모델 선정

판매량 → Ensemble

SNS 언급량 예측

| | LGBM | CatBoost | DNN | LSTM | Ensemble |
|------|---------|----------|---------|---------|----------|
| MSE | 674.180 | 650.987 | 650.930 | 472.266 | 552.024 |
| RMSE | 25.965 | 25.514 | 25.513 | 21.732 | 23.495 |
| MAE | 10.480 | 10.524 | 11.937 | 8.849 | 9.719 |

SNS
언급량 → LSTM

03. 분석과 모델링 | 모델 시연

21년 7월 29일의 탄산음료 판매량과 SNS 언급량 예측

품목 One-Hot Encoding

```
array([[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]])
```

날씨 (21.07.09~29)

| | tm | avgTa | temp_diff | sumRn | avgWs | avgRhm | avgTca | O3 | PM10 |
|---|------------|-----------|-----------|----------|----------|-----------|----------|----------|-----------|
| 0 | 2021-07-09 | 25.035173 | 5.770388 | 2.963365 | 1.758153 | 81.953459 | 7.053429 | 0.029647 | 15.430012 |
| 1 | 2021-07-10 | 25.675155 | 6.256063 | 2.532003 | 2.388350 | 81.726342 | 6.932619 | 0.042718 | 25.255703 |
| 2 | 2021-07-11 | 25.883850 | 5.511758 | 0.534969 | 2.314651 | 84.890819 | 7.065721 | 0.041383 | 25.969001 |

Ensemble

```
def predict_ss(base_input, lstm_input):
    sale_pred = ensemble_sale(base_input, lstm_input)
    sns_pred = lstm_sns(lstm_input).numpy()
    return sale_pred, sns_pred
```

LSTM

판매량 예측

20대 여성 판
매량(개)

60대 여성 판 매량(개)

| | | |
|------|-------------|-------------|
| 탄산음료 | 3342.951199 | 1778.154984 |
|------|-------------|-------------|

384.585808 249.978117

SNS 언급량 예측

SNS 이용 강도

탄산음료 341.766113

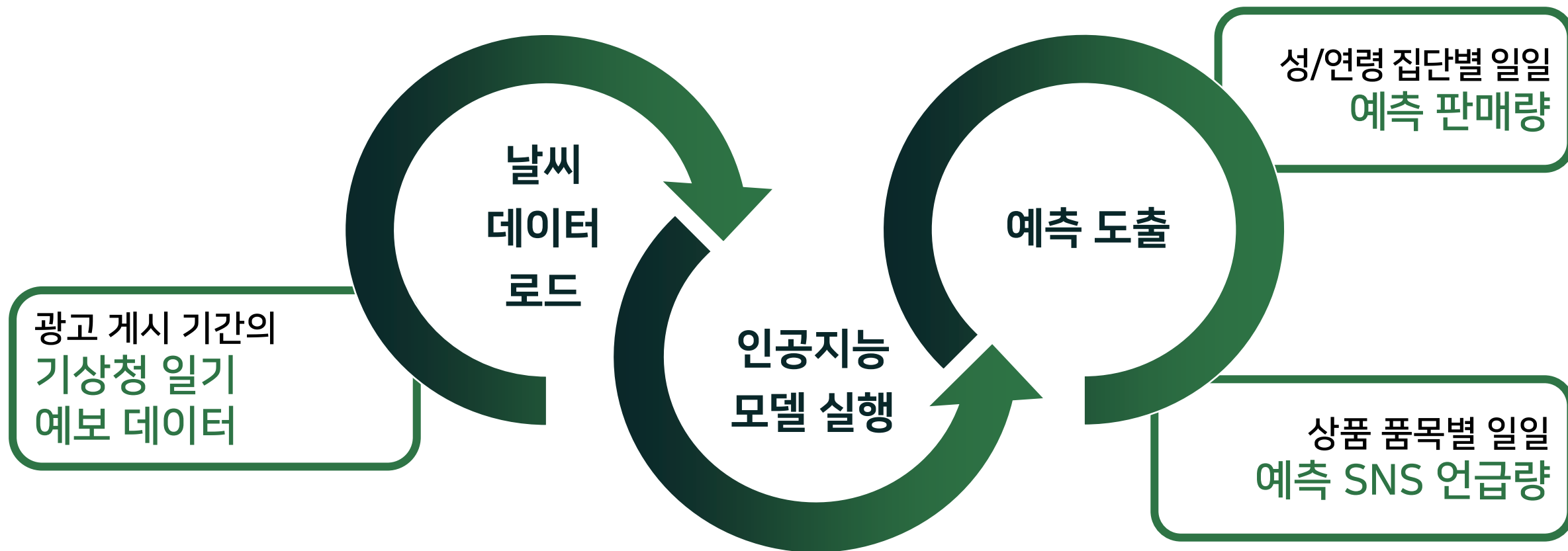


오늘의 광고

“AI를 활용한 **날씨별/연령별/성별 맞춤형** 광고 제작 및 맞춤형 게시 서비스”



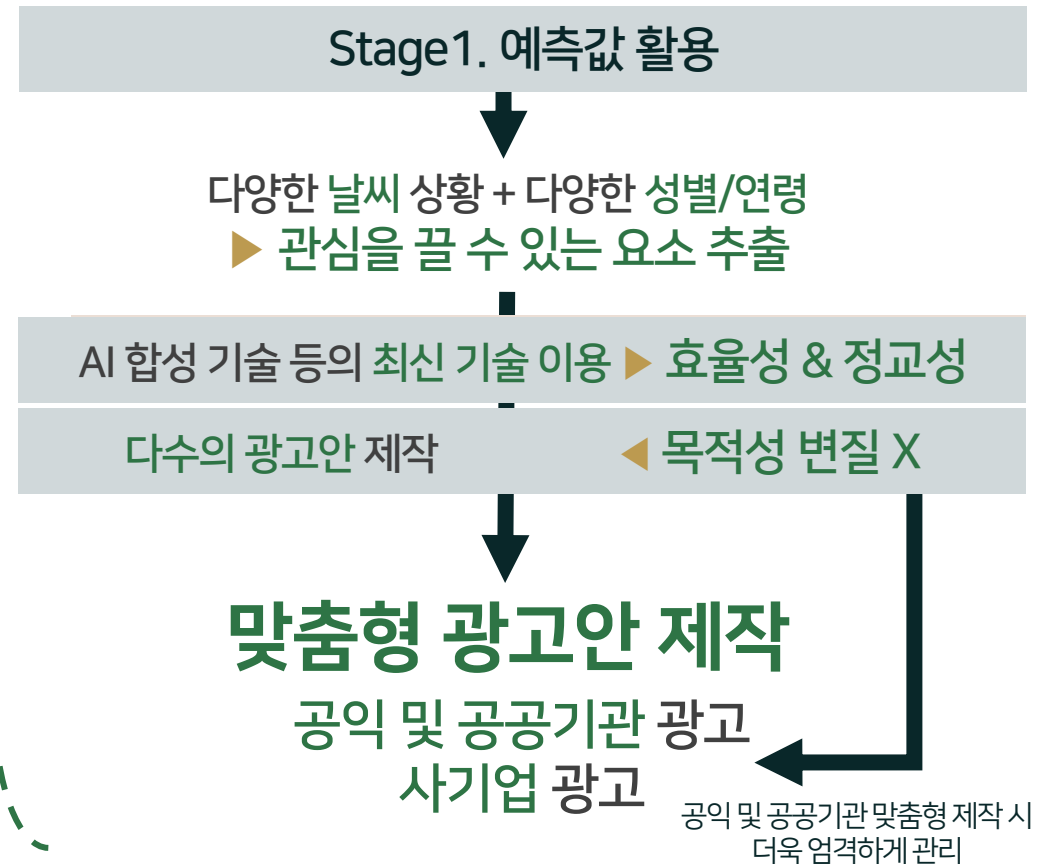
Stage1. 판매량 및 SNS 언급량 예측



Stage2. 성별+연령별+날씨별 맞춤형 광고 제작



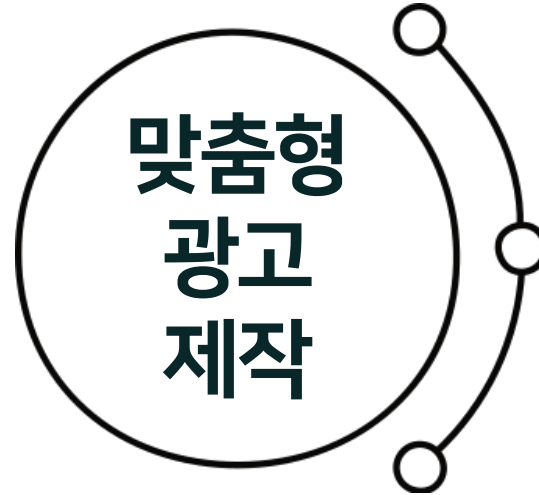
성별/연령별/날씨별 맞춤형 광고 제작 프로세스



맞춤형 광고 제작 예시: 기업 광고

맑은 날 | 맑은 날씨 배경 사용 

 여성 모델 사용 | 남성 고객
트로트 가수 모델 사용 | 60대 고객



눈 오는 날 | 눈 오는 배경 사용 

 남성 모델 사용 | 여성 고객
아이돌 가수 모델 사용 | 20대 고객



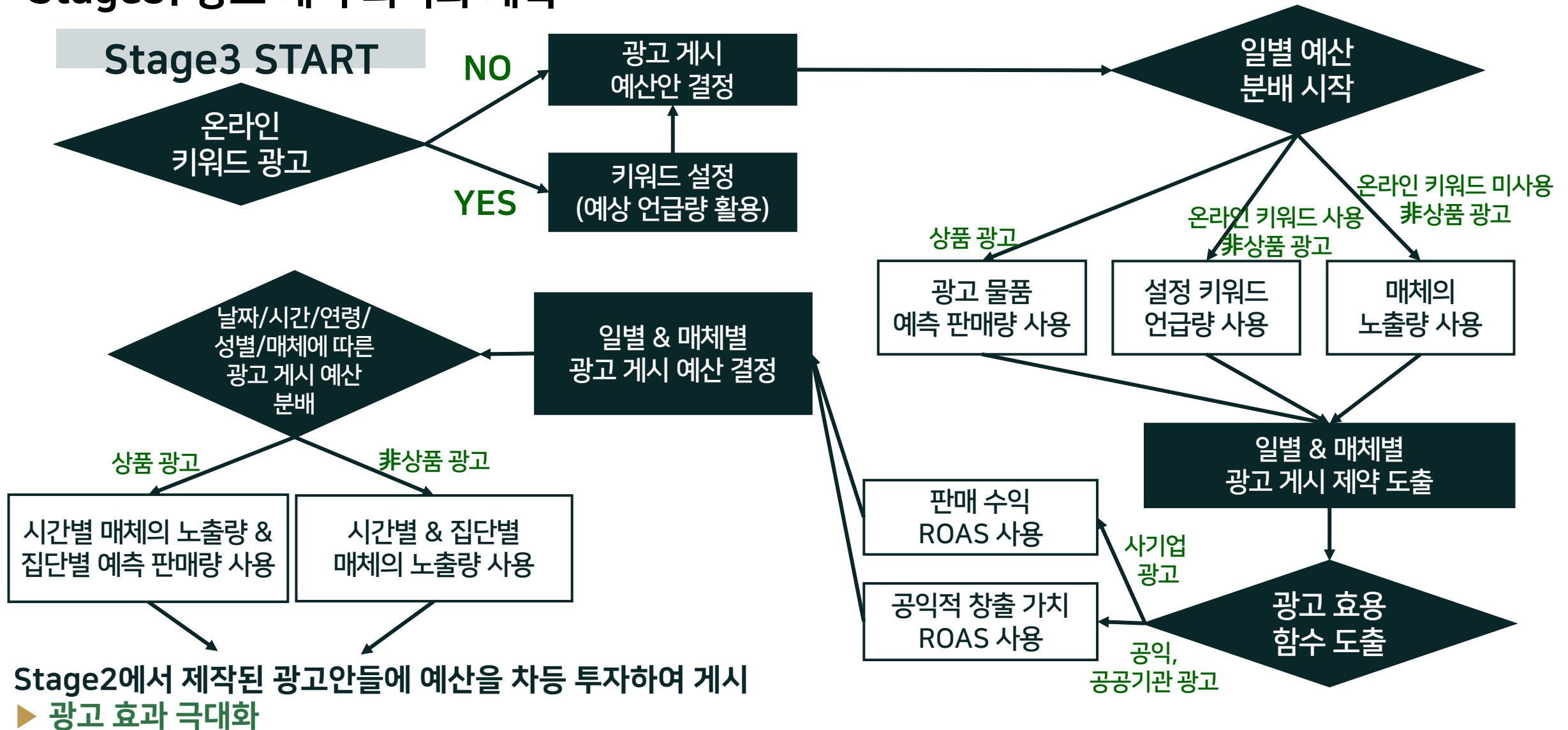
Target Category

| 맑음 | 눈 |
|-----|-----|
| 60대 | 20대 |
| 남성 | 여성 |

맞춤형 광고 제작 예시: 공익 광고

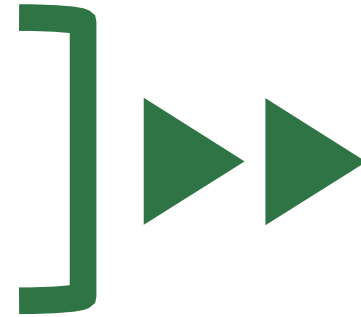


Stage3. 광고 게시 최적화 계획



온라인 키워드 광고 설정

- ☑ 일기 예보를 이용하여 도출한
일별 물품들의 예상 언급량의 변화
- ☑ 물품들과 광고의 연관성



광고의 키워드로
설정할 물품명 결정

예시) A물품 vs B물품

광고 날짜 | A물품 광고 게시한 날
예 측 | B물품 예상 언급량 급증
관 계 | A물품과 B물품 보완재 관계

A물품 광고 키워드 ► B물품 설정

일별 품목의 예상 언급량은 Stage 1에서 도출한 것을 이용

맞춤형 광고 게시 예산안 결정 과정

01. 전체 광고 게시 예산 결정 및
일별 광고 게시 예산 분배

I. 상품 광고

광고 게시 기간 동안의 일별 일기예보를 통해 도출한
광고 품목별 일별 예측 판매량 사용

II. 온라인 키워드를 사용한 비(非)상품 광고 (ex: 검색 시 나오는 광고)
광고 게시 기간 동안의 일별 일기예보를 통해 도출한
설정 키워드의 일별 예측 언급량 사용

III. 온라인 키워드를 사용하지 않는 비(非)상품 광고 (ex: 배너 광고, 랜딩 페이지)
광고 게시 기간 동안의 일별 일기예보를 통해 도출한
광고 게시 시 사용하는 전 매체의 일별 노출량 사용
일별 노출량은 과거 데이터를 통해 예측

▶ 광고일 광고 게시 예산

$$= \text{전체 광고 게시 예산} \times \frac{\text{광고일 예측 판매량(or 언급량 or 노출량)}}{\text{광고기간 예측 총판매량(or 총언급량 or 총노출량)}}$$

예시) A물품

| | |
|--------------|------|
| 광고 대상 | A물품 |
| 2일간 전체 광고 예산 | 3억 |
| 1일차 예측 판매량 | 100개 |
| 2일차 예측 판매량 | 200개 |

1일차 광고 게시 예산 ▶ 3억 원 X (1/3) = 1억 원

02. 일별 & 매체별 광고 게시 예산 제약 도출

$$W = \sum_{i=1}^n p_i x_i$$

(p: 광고 게시 횟수, x:광고 게시 비용, W: 전체 예산)

예시) A물품 + X매체 & Y매체

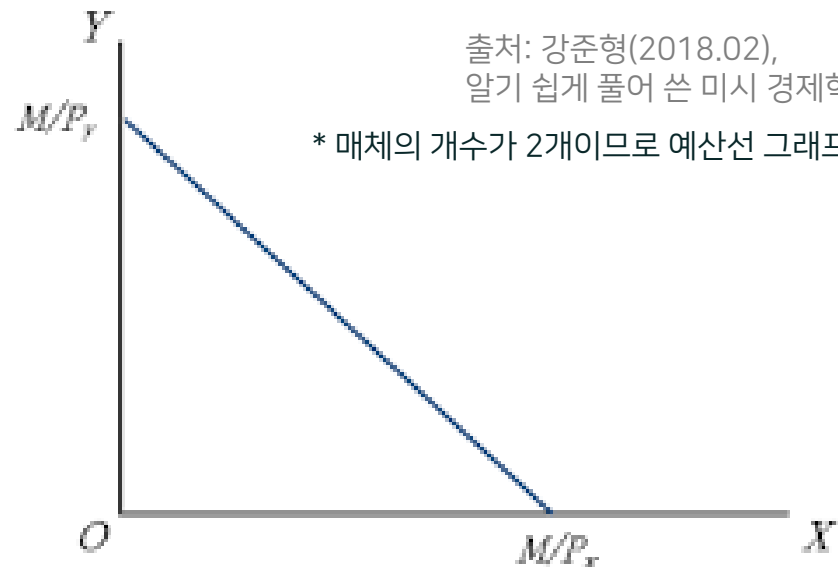
광고 날짜 | 광고 1일차

게시 예산 | 1억

광고 매체 X의 광고 게시 횟수 | X

광고 매체 Y의 광고 게시 횟수 | Y

1일차 광고 게시 예산 제약 ► 예산선 도출



03. 광고 효용 함수 도출

일별 매체별 예상 광고비용 대비 수익률 (효용, ROAS) 계산

$$\text{매체별 평균 ROAS(Return On Ad Spend)} = \left(\frac{\text{수익}}{\text{광고비}} \right) \times 100$$

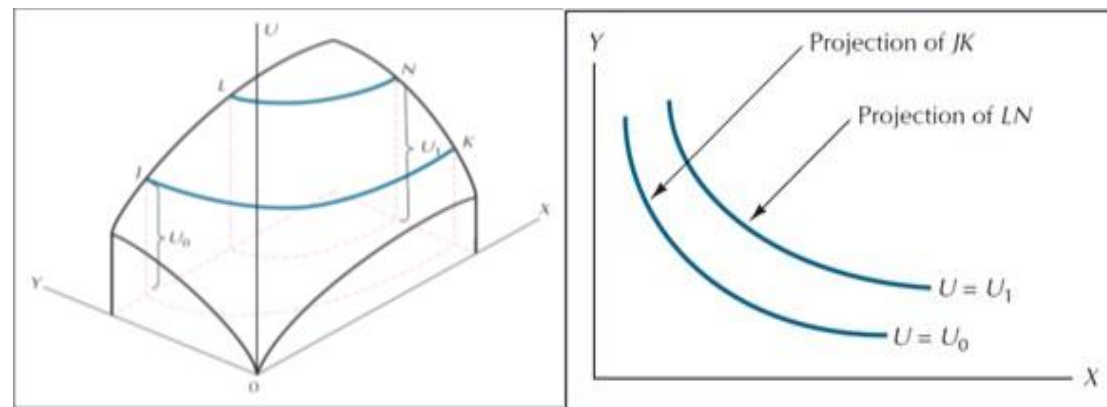
공익, 공공기관 광고는 위의 식 분자에 수익 대신 공익적 창출가치(비용) 사용

예시) A물품 + X매체 & Y매체

A물품의 광고 효용 함수 $U(X, Y) =$

$(X\text{매체 예상 ROAS} * X) + (Y\text{매체 예상 ROAS} * Y)$

X는 X매체 광고 게시 횟수, Y는 Y매체 광고 게시 횟수



출처: Robert H. Frank(2009.09), Microeconomics and Behavior

* 매체의 개수가 2개이므로
3차원의 효용곡면과 2차원의 무차별지도 도출

효용 함수 ► 효용곡면 & 무차별지도 도출

04. 일별 & 매체별 광고 게시 예산 설정

예산 제약 하에서 가장 효용함수 값이 높은 조합을 선정

▶ 일별 & 매체별 광고 게시 예산을 분배

예시) A물품 + X매체 & Y매체

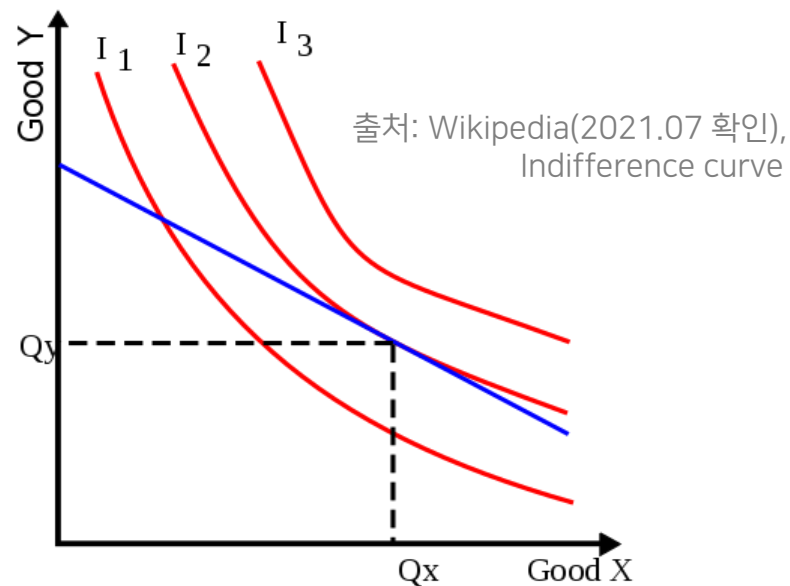
무차별곡선과 예산선의 교점 도출
무차별지도(I3, I2, I1순으로 효용이 높은 곡선)에서
효용이 높은 무차별곡선과 예산선의 교점이 최적의 조합

최적의 X매체 광고 게시 횟수 | Q_x X매체 광고 게시 횟수 | X축
최적의 Y매체 광고 게시 횟수 | Q_y Y매체 광고 게시 횟수 | Y축

광고 게시 예산

▶ X매체 | $Q_x * X매체 광고 단가$
Y매체 | $Q_y * Y매체 광고 단가$

1일차 X매체 광고 게시 예산은 3,000만 원, Y매체 광고 게시 예산은 7,000만원이라 가정



* 매체의 개수가 2개이므로 그래프는 2차원

05. 일별 & 매체별 광고 게시 예산 내 시간/성/연령별 노출량에 따른 광고 게시 예산 분배

I. 상품 광고

맞춤형 광고 게시 예산

$$= \text{특정 일, 특정 매체 광고 게시 예산} \times \frac{\text{특정 시간대, 특정 매체 평균 노출량}}{\text{특정 매체의 일 평균 노출량}} \times \frac{\text{특정 연령 · 성별 집단의 광고 물품 일 예측 판매량}}{\text{광고 물품 일 예측 판매량}}$$

II. 비(非)상품 광고

맞춤형 광고 게시 예산

$$= \text{특정 일, 특정 매체 광고 게시 예산} \times \frac{\text{특정 연령 · 성별 집단의 특정 시간대, 특정 매체 평균 노출량}}{\text{특정 매체의 일 평균 노출량}}$$

예시) A물품 + X매체

| | |
|------------------------|----------|
| 1일차 X매체 광고 게시 예산 | 3,000만 원 |
| 12-18시 X매체 평균 노출량 | 600회 |
| X매체 일 평균 노출량 | 1000회 |
| 1일차 20대 여성의 A물품 예측 판매량 | 200개 |
| 1일차 A물품 예측 판매량 | 500개 |

1일차 12-18시 X매체에서, 20대 여성 맞춤형 광고 예산

▶ (3,000만 원) X (600/1000) X (200/500) = 720만 원

05. 기대 효과 | 기존 광고와의 차별점

기대 효과1. 맞춤형 광고 서비스에 **날씨 정보** 추가

기존 광고 | 날씨에 따라 변하는 사람들의 관심 고려 불가

오늘의 광고 ▶ 기존에 없던 **다양한 날씨 요소들의 변화**가 포괄적, 과학적으로 **고려**된 맞춤형 광고 서비스

기대 효과2. 다양한 고려 요소를 **조합하여 다수의 맞춤형 광고** 제공

기존 광고 | 맞춤형 광고를 진행함에도
맞지 않는 광고를 시청하게 되는 개인들 존재

오늘의 광고 ▶ 성별/연령/날씨 맞춤형 광고들을 합성기술을 이용해 **다수의 맞춤형 광고**를 제작한 뒤, 이를 제공

기대 효과3. **다양한 플랫폼에서의 활용**

기존 광고 | 플랫폼 선정에서의 한계 존재

오늘의 광고 ▶ **여러 온라인 플랫폼과 오프라인 플랫폼**에서 맞춤형 광고 서비스 제공 가능

예시) **온라인** | 개인정보 활용 광고 서비스, 랜딩 페이지 배너 광고 서비스

오프라인 | 지금까지 제공 불가능했던 정교한 맞춤형 전광판 광고 서비스

기대 효과4. **공공기관에도 맞춤형 광고 서비스** 제공 가능

기존 광고 | 대부분 사기업 광고

오늘의 광고 ▶ **공익 및 공공기관 광고**에서도 적절한 맞춤형 광고 서비스 제공 가능

05. 기대 효과 | 기대 효과 모식도

