

롯데멤버스 빅데이터 경진대회

고객구매 데이터 기반 추천 시스템 개발 및 개인화 마케팅 전략 제안

프로젝트 요약

롯데멤버스 제공 내부 데이터를 기본 데이터 셋으로 사용하여 개인화 마케팅에 활용 가능한 상품 추천 시스템 개발
→ 사용자 기반 협업 필터링을 활용해 사용자 간 유사도를 분석하여 선호도가 비슷해 구매할 것으로 예상되는 상품 추천

진행 타임라인

- 2022.07.04 EDA 및 주제 선정
- 2022.07.11 데이터 전처리 및 외부데이터 선정
- 2022.07.18 모델 개발 및 성능 개선
- 2022.07.27 개인화 마케팅 제안
- 2022.08.05 보고서 작성
- 2022.08.08 프로젝트 종료

팀원 소개

Team – LP Player

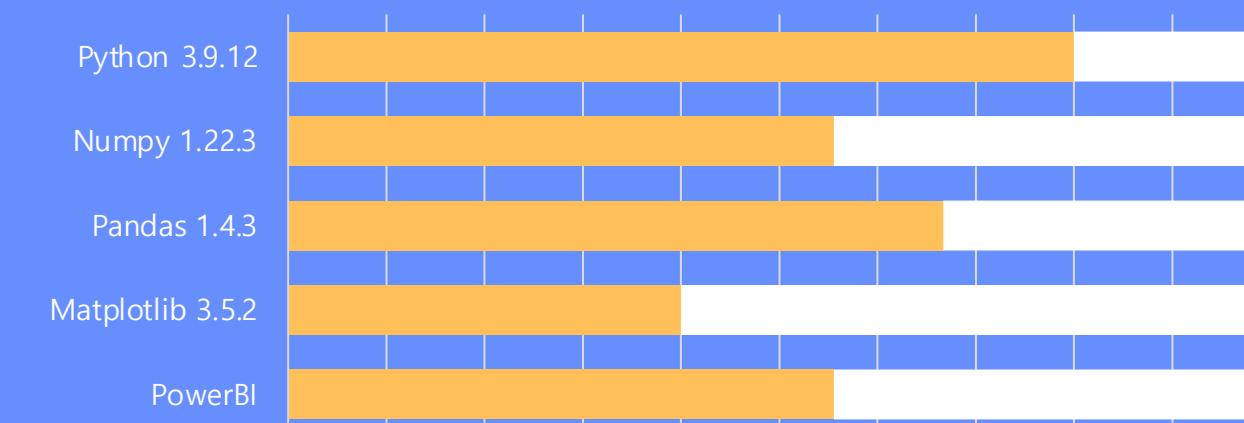


팀장 김민균 / 팀원 조수현, 한유림

WORK RULE

DATA SET: 롯데멤버스 제공 고객 데이터
Language: Python
사용 모델: 사용자 기반 협업 필터링

모델 개발 환경 및 버전



목차

01

서비스 개요

02

EDA 및
인사이트 분석

03

모델 개발

04

마케팅 제안

05

결론

팀원 소개



김민균 (팀장)

- 프로세스 및 모델 기획 실무 총괄
- 분석 기획 및 기획서 작성(부)
- 외부 데이터 수집 및 전처리
- 모델 개발 및 성능 평가(부)
- 분석 결과 기반 마케팅 전략(정)
- 결과 시각화 및 보고서 작성



조수현 (팀원)

- 분석 기획 및 기획서 작성(정)
- 데이터 수집 및 전처리
- 모델 개발 및 성능 평가(부)
- 분석 결과 기반 마케팅 전략(부)
- 결과 시각화 및 보고서 작성
- 보고서 작성 총괄



한유림 (팀원)

- 프로젝트 전체 일정관리
- 분석 기획 및 기획서 작성(부)
- 데이터 수집 및 전처리
- 모델 개발 및 성능 평가(정)
- 분석 결과 기반 마케팅 전략(부)
- 결과 시각화 및 보고서 작성

I . 개요

- 프로젝트 추진 방향
- 현상 분석
- 개인화 마케팅
- 개인화 마케팅의 추천 시스템
- 개인화 마케팅의 협업 필터링

01

롯데멤버스
빅데이터 경진대회

1.1 프로젝트 추진 방향

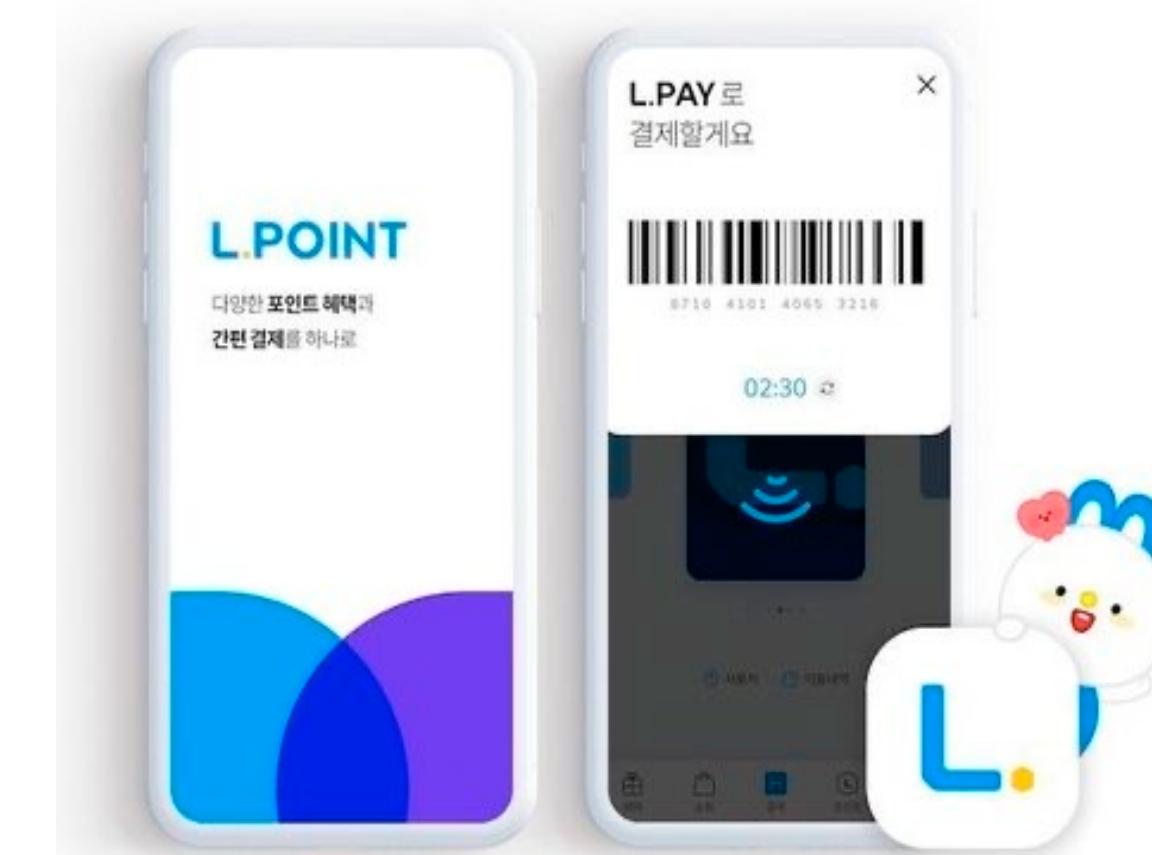
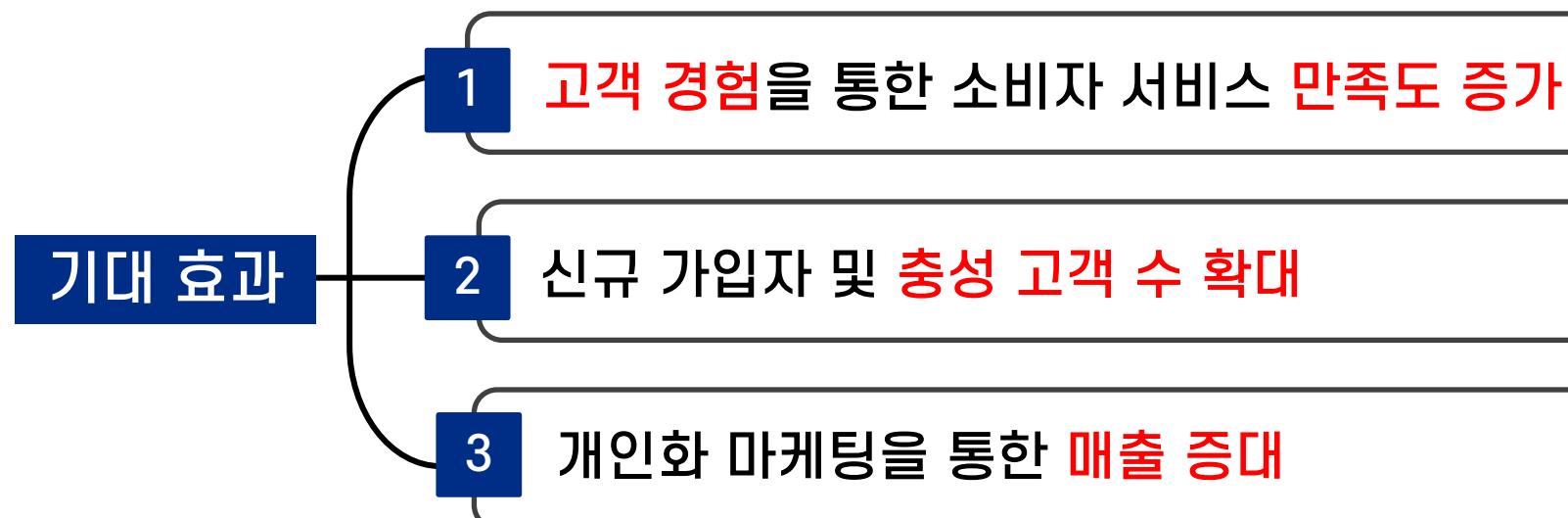
I. 개요

• 프로젝트 개요 및 목표

주제: 고객구매 데이터에 기반한 예측 모델 개발 및 개인화 마케팅 전략 제안

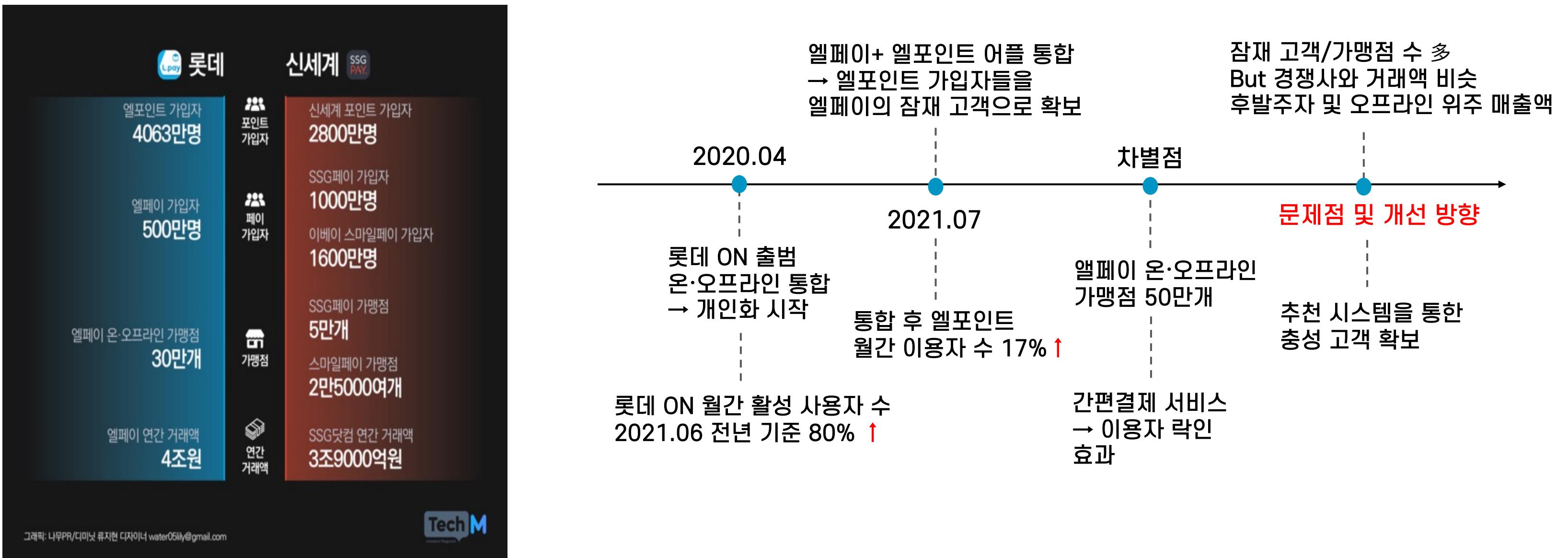
롯데멤버스에서 제공해주는 고객 구매 데이터를 가공, 상품 구매 예측(추천) 모델을 개발하여 개인화 마케팅에 활용

개발 방향 및 목표: 소비자가 온라인을 통해서 쉽게 접근할 수 있다는 점을 고려하여 원하는 상품을 찾는데 오랜 시간이 소요되는 수고로움 없이 비슷한 사용자가 구매한 상품을 추천해줌으로서 효율적이고 개인 맞춤형의 추천 시스템을 사용할 수 있다.



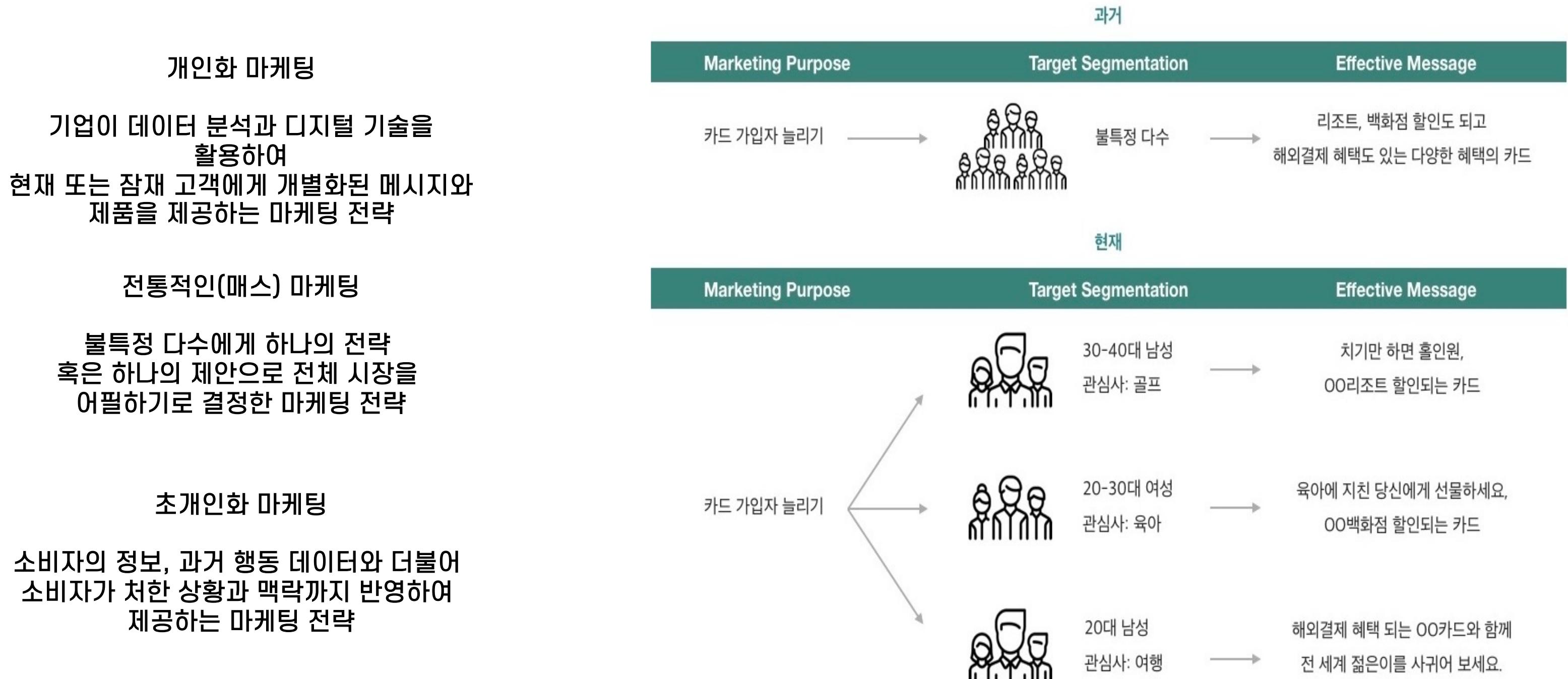
1.2 현상 분석

· 현상 분석 및 제작 목적



1.3 개인화 마케팅

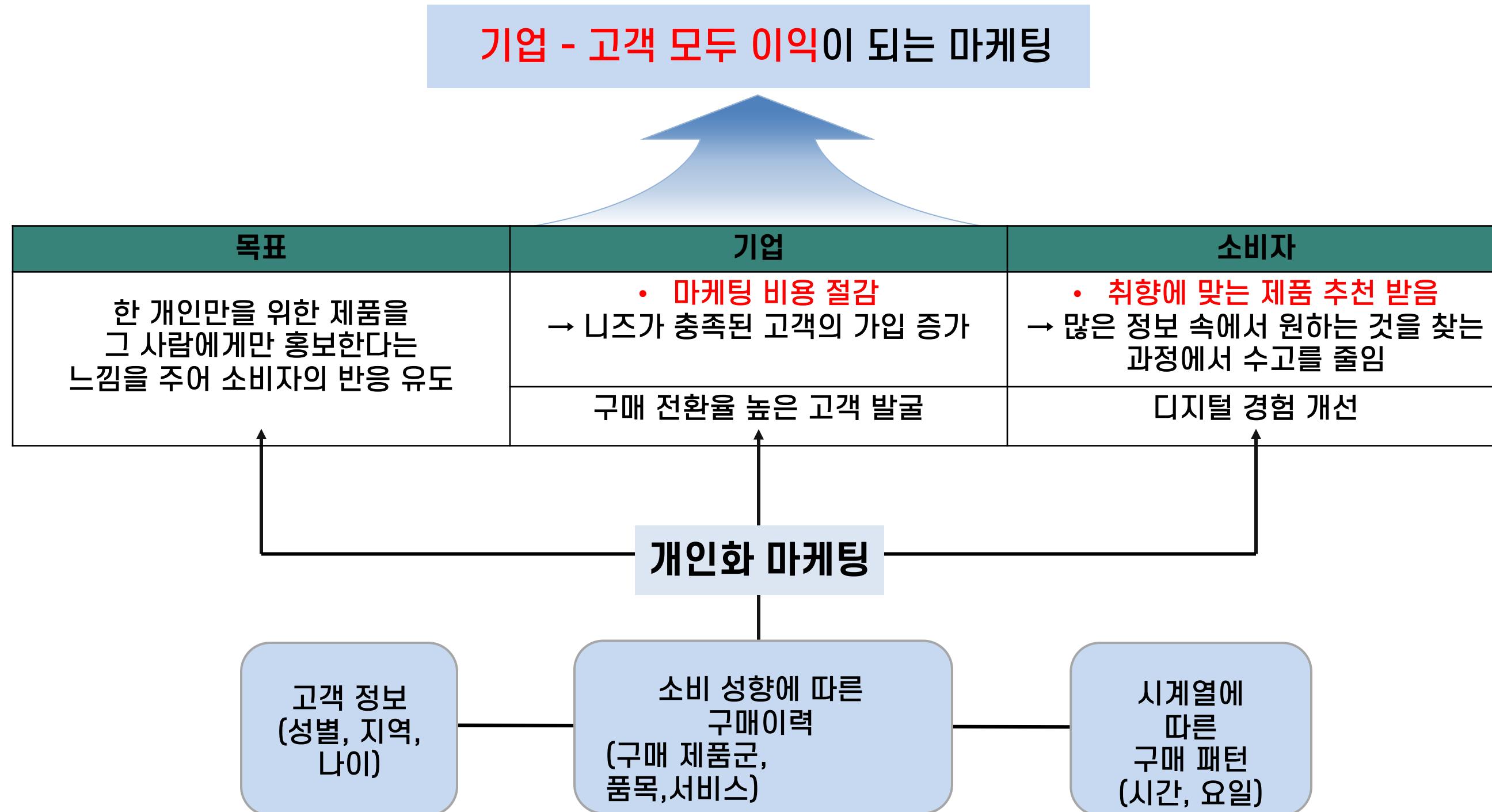
· 개인화 마케팅과 마케팅의 변천사



1.3 개인화 마케팅

I. 개요

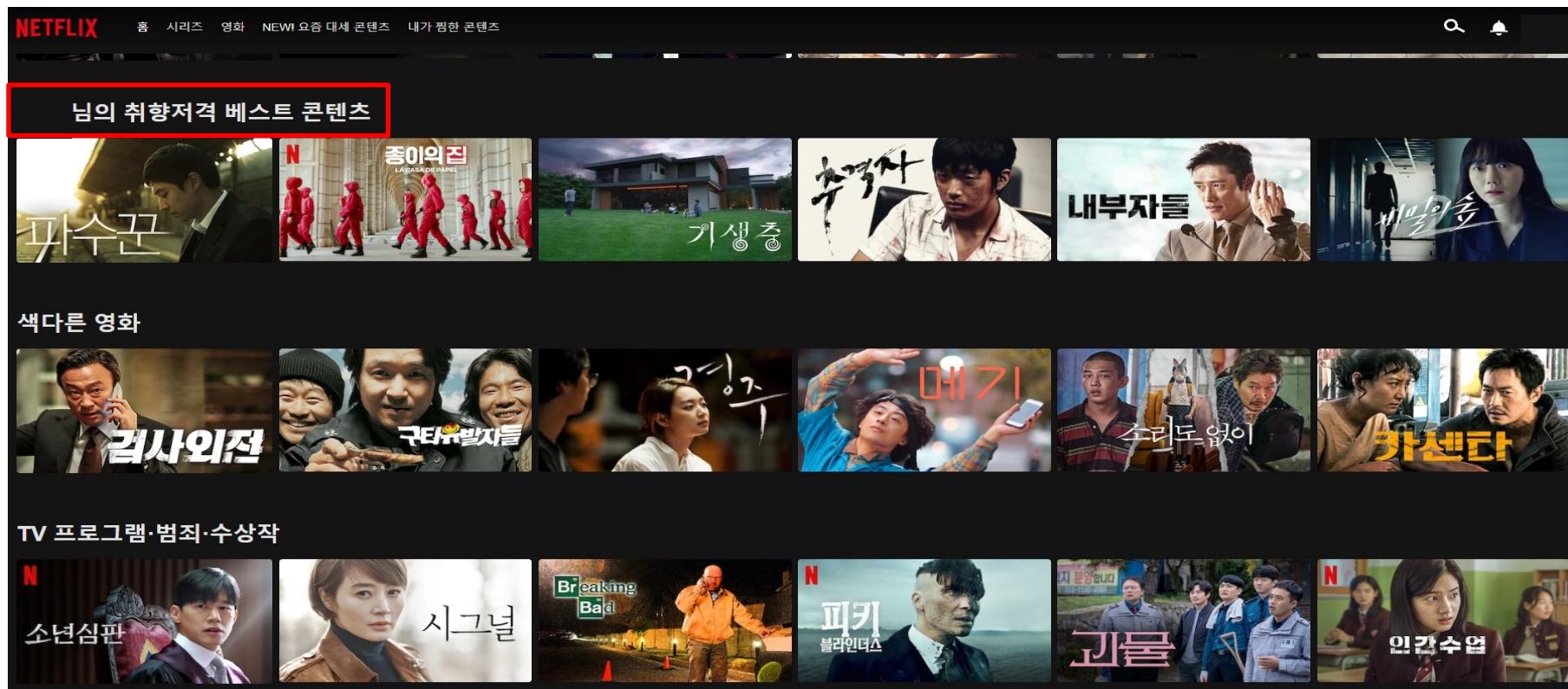
· 개인화 마케팅의 목표와 필요성



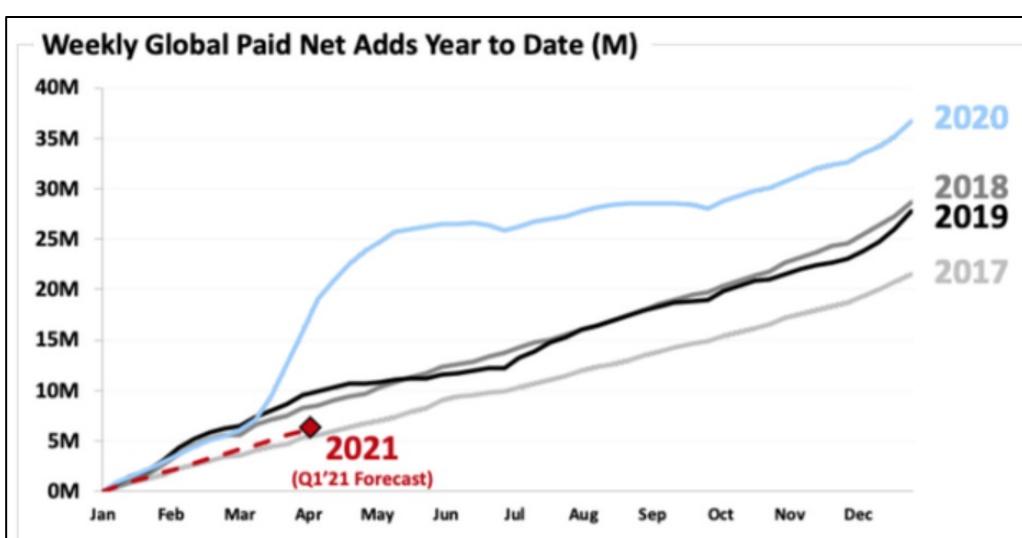
1.4 개인화 마케팅의 추천 시스템

I. 개요

• 추천 시스템의 개념과 종류



<예시> - 넷플릭스 개인화 추천 시스템



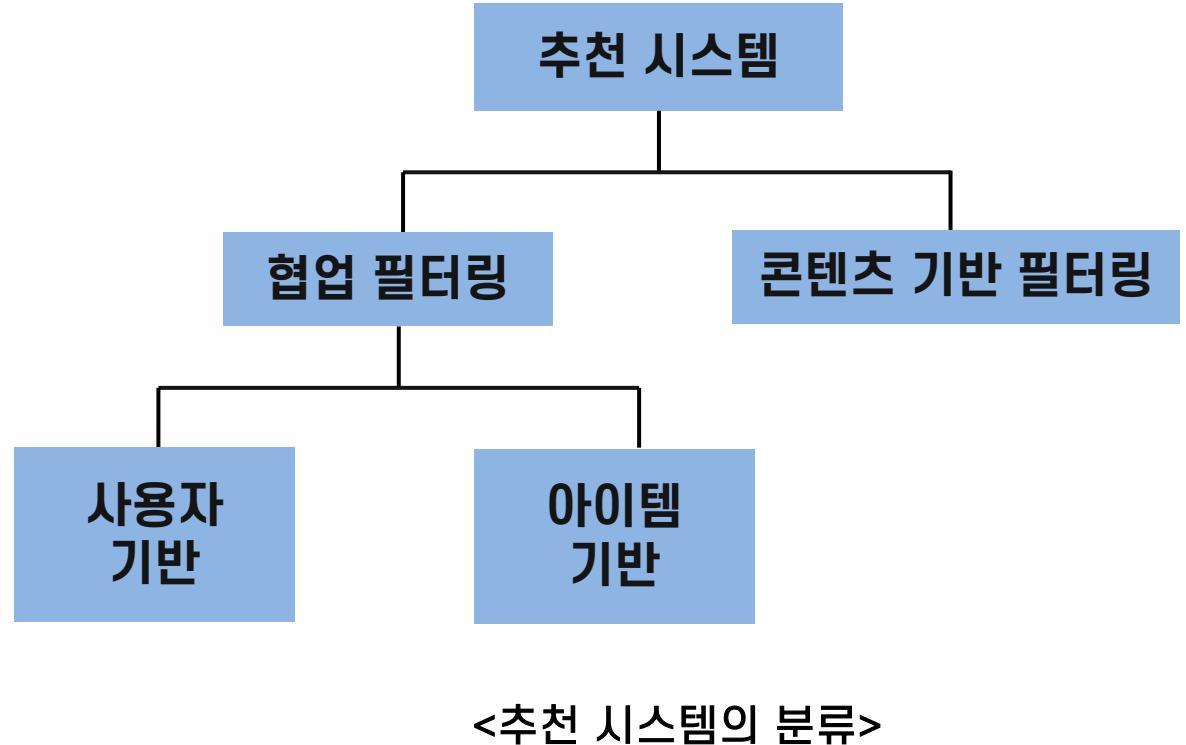
<개인화 추천>

각 개인의 관심사나 선호도를 분석해서 그에 맞는 정보나 제품을 추천해 주는 것

<추천 시스템>

사용자의 과거 행동 데이터나 다른 데이터를 바탕으로 사용자에게 필요한 정보나 제품을 골라서 제시해주는 시스템

◀ 넷플릭스 유료 구독자 가입 수 추이



<추천 시스템의 분류>

협업 필터링 (Collaborative Filtering : CF)

구매, 소비한 제품에 대한 각 소비자의 평가를 받아 평가 패턴이 비슷한 소비자를 한 집단으로 보고 그 집단에 속한 소비자들의 취향을 활용하여 추천하는 기술

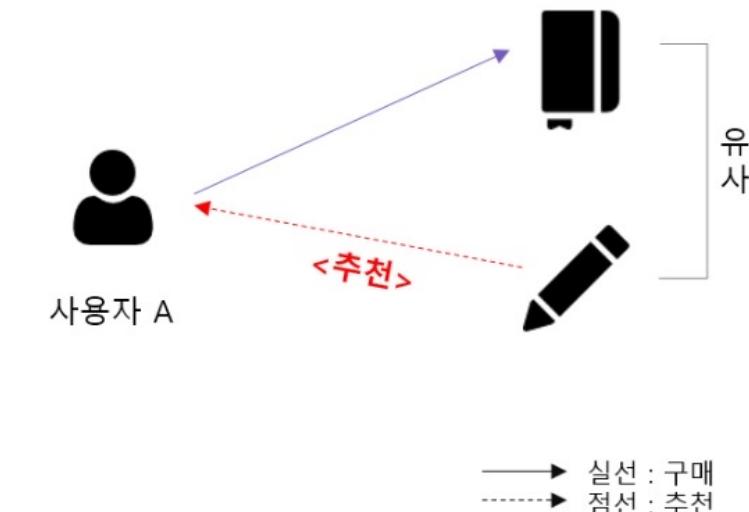
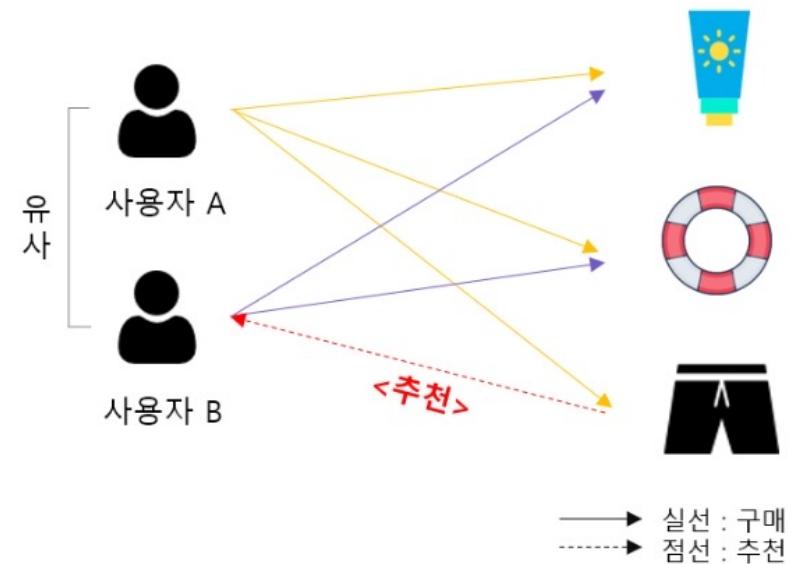
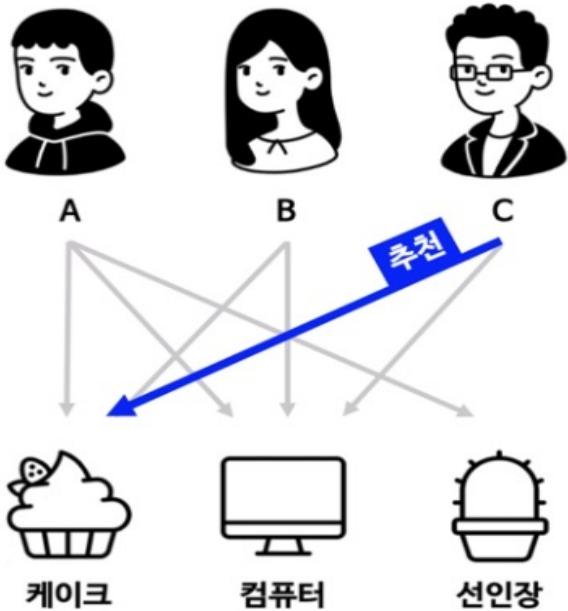
콘텐츠 기반 필터링 (Content-based filtering : CB)

제품의 내용을 분석해서 추천하는 기술

1.5 개인화 마케팅의 협업 필터링

I. 개요

- 협업 필터링 개념 설명 및 한계점을 고려한 선정 이유



협업 필터링 – 논문 탐색

탐색 과정

- 추천 시스템의 가장 대표적인 예시인 영화 추천 시스템을 바탕으로 적절한 모델 탐색
- 개인화 마케팅에 사용되는 대표적인 알고리즘이 협업 필터링과 컨텐츠 기반 필터링인 것 확인
- 장단점 비교 결과 협업 필터링이 적합하다 판단하여 선정 (구체적)

사용자 기반 협업 필터링

- 고객 간의 선호도를 분석하여 나와 유사한 성향의 고객이 좋아하거나 구매했던 상품/콘텐츠를 추천하는 기법
- 유사도 측정 효율성 (비용 및 시간)
: 아이템 기반 < 사용자 기반
→ 사용자 수에 비해 아이템 수가 많음

아이템 기반 협업 필터링

- 아이템 간의 유사도를 측정하여 사용자가 아이템을 조회했을 때 유사 상품을 추천하는 기법
- 협업 필터링 유형 중 가장 많이 쓰이는 방법 (넷플릭스, 아마존 등)
- 사용자 기반 협업 필터링의 단점 보완

02

롯데멤버스
빅데이터 경진대회

II. EDA 및 인사이트 분석

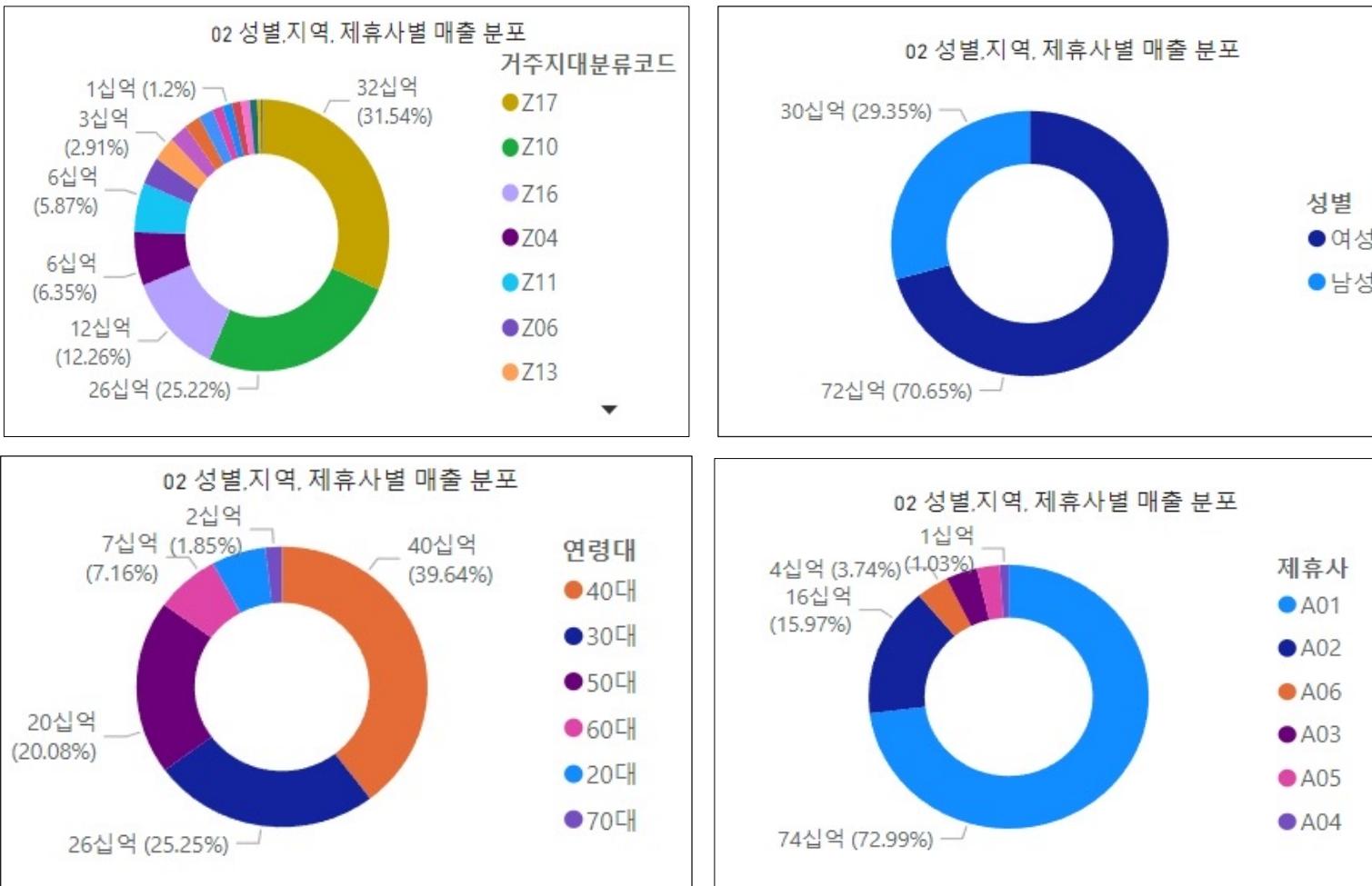
- 데이터 탐색
- 데이터 전처리와 가공
- 외부 데이터 선정

2.1 데이터 탐색

EDA 분석을 통한 데이터 탐색

서울/경기 거주 30-40대 여성

< 01 고객 데모 > & < 02 상품 구매정보 >



1

성별, 연령, 지역, 제휴사 별로 매출 분포 탐색
→ 핵심 고객층(충성 고객)을 파악함으로써
데이터의 신뢰성/정확도를 높임
→ 이후 진행한 고객 데이터 분석은 지정한
타겟층을 대상으로 새로운 파일을 만들어 진행함

2

상품 구매정보 파일의 전체 데이터 양은 400만
건으로 타겟층을 선정하고 **데이터 양을
줄임**으로서(100만 건) 데이터 분석 소요 시간과
속도를 절감
→ 동질적인 하나의 집단을 가지고 데이터 분석을 할
때 **오류율 최소화**
→ 충성고객으로 모델링해서 전체로 확대

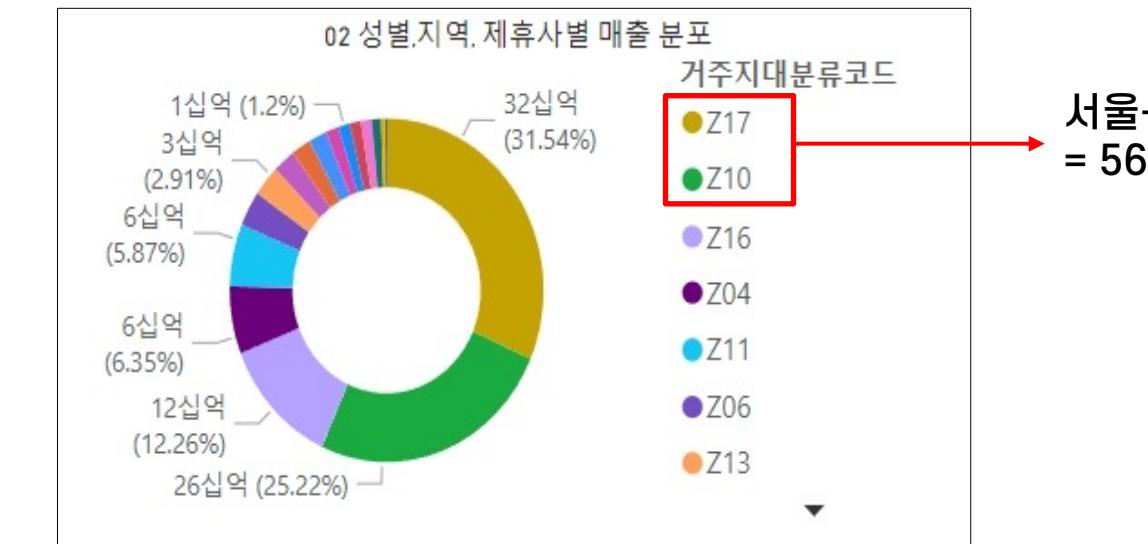
2.1 데이터 탐색

· EDA 분석을 통한 데이터 탐색

- 코드화 된 지역 정보 파악

< 01 고객 데모 > & 2021 통계청 인구 동향조사

행정구역(시군구)별	총인구수 (명)	거주지대분류코드	고객번호 (개)
경기도	13,565,450	Z10	8157
서울특별시	9,509,458	Z17	7881
부산광역시	3,350,380	Z16	2503
경상남도	3,314,183	Z11	1840
인천광역시	2,948,375	Z04	1750
경상북도	2,626,609	Z06	1080
대구광역시	2,385,412	Z14	988
충청남도	2,119,257	Z13	896
전라남도	1,832,803	Z12	803
전라북도	1,786,855	Z08	739
충청북도	1,597,427	Z15	677
강원도	1,538,492	Z09	642
대전광역시	1,452,251	Z03	600
광주광역시	1,441,611	Z05	576
울산광역시	1,121,592	Z01	474
제주특별자치도	676,759	Z07	204
세종특별자치시	371,895	Z02	103



01 고객 데모

가정: 고객 수는 인구 수와 비례할 것

과정: 2021년 통계청 인구 동향조사 자료를 참고하여

코드화 된 지역 추측

→ 가장 고객이 많이 포진되어 있는 서울/경기 지방의
고객들을 대상으로 데이터 활용 결정

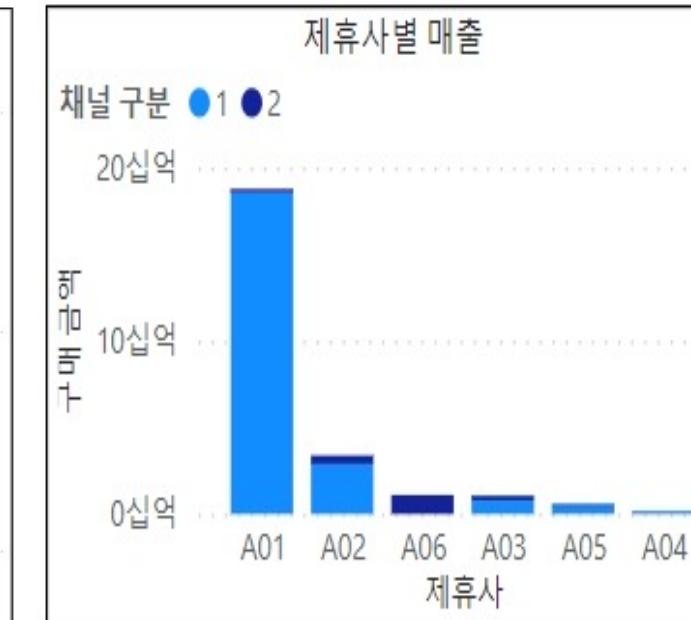
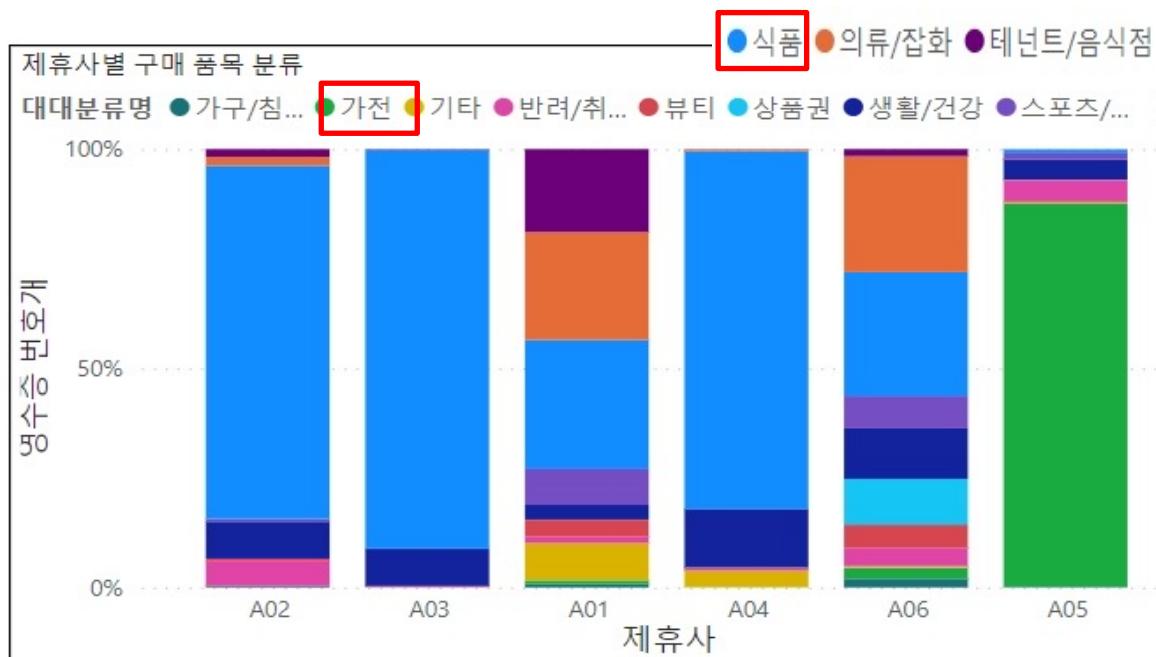
결과: 전체의 56% 정도를 차지하여 일부의 데이터지만
전체로 이용하기에 문제 없다고 판단

2.1 데이터 탐색

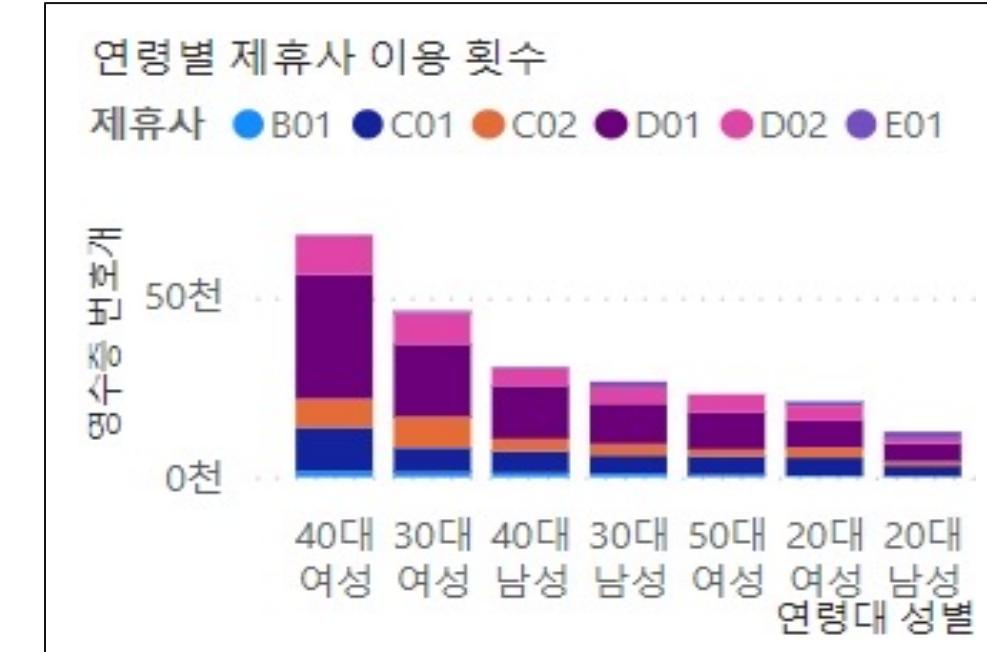
II. EDA 및 인사이트 분석

EDA 분석을 통한 데이터 탐색 – 제휴사 및 유통사

지정된 타겟층을 대상으로 가공한 F3040 파일을 기반으로 분석



< 01 고객 데모 > & < 03 제휴사 이용 정보 >



제휴사	근거
A01(롯데백화점)	다양한 구매 품목+압도적 매출과 오프라인 지출
A02(롯데마트)/A03/A04	식품 위주의 구매 품목, 롯데마트 점포 수: 경기=서울x2
A05(하이마트)	87%의 가전 판매율
A06(롯데 ON)	다양한 구매 품목+100% 온라인 지출

기존 : 백화점 중심의 오프라인 매출이 강점
목표 : 개인화 마케팅은 온라인 위주, 온·오프라인 통합이 핵심
→ 온라인 매출을 더욱 활성화 할 방안 모색
해결 방안 : 추천 시스템 개발 및 향후 모델 고도화가 도움이 될 수 있도록 목표 설정

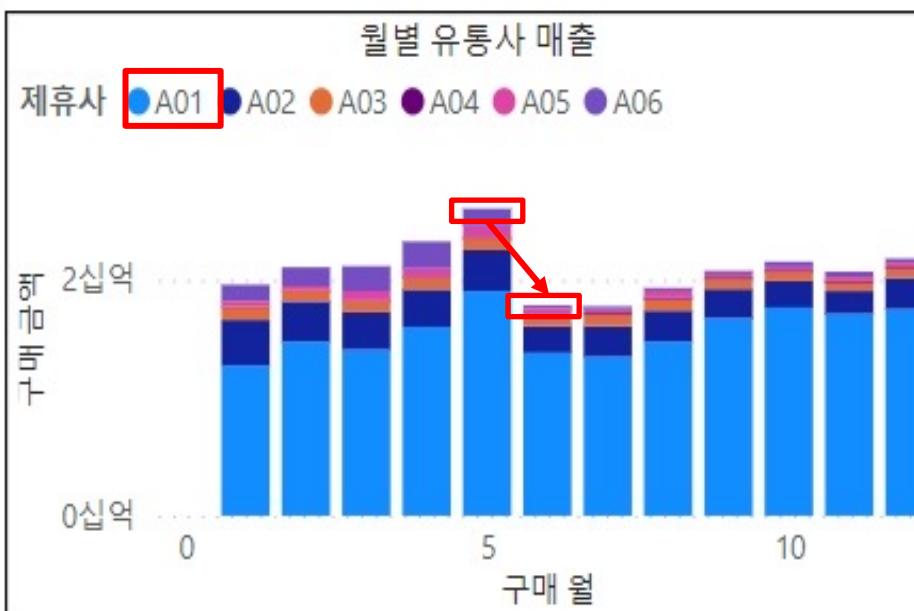
제휴사 사용 비율

- 30-40대 여성이 눈에 띠게 1,2위를 차지
- 30-40대 여성 포함 전 연령층에서 F&B(D01+D02)의 비율이 가장 많음

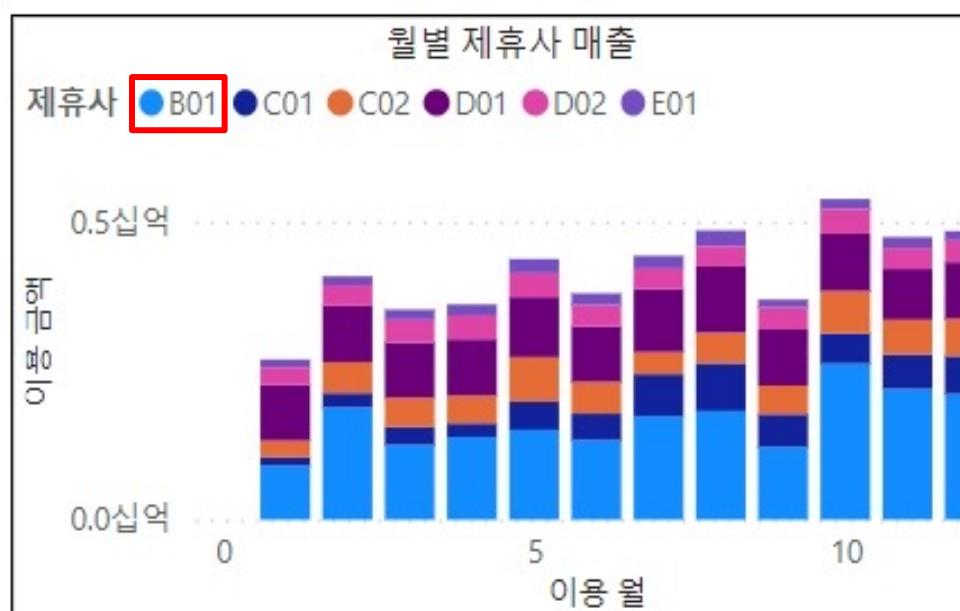
2.1 데이터 탐색

· EDA 분석을 통한 데이터 탐색 – 시계열&매출

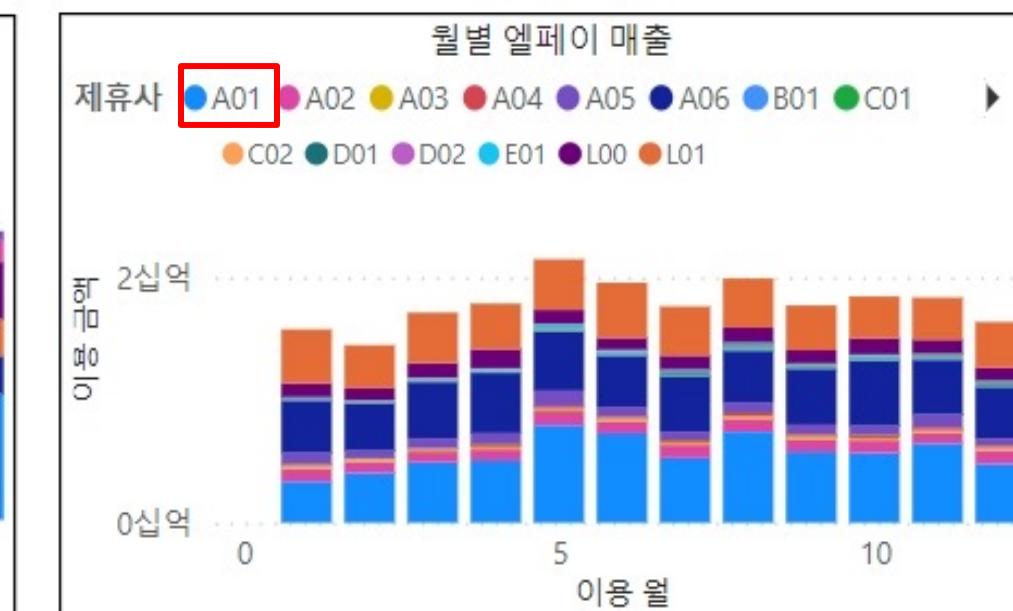
< F 3040 >



< 03 제휴사 이용 정보 >



< 06 엘폐이 이용 내역 >



<결과>

1. 세 파일의 데이터가 공통된 추이를 보이지는 않음

2.1 유통사: 6월부터 큰 폭으로 매출이 감소함.

오프라인에서는 A01, 온라인에서는 A06의 매출이 동시에 떨어진 것이 영향을 끼친 것으로 추측

2.2 제휴사: B01의 매출 영향을 가장 많이 받음, 가장 결제 금액이 크기 때문으로 예측

대부분 오프라인에서 결제가 이루어지기 때문에 외부데이터인 코로나 확진자 수와 연관하여
매출이 영향을 받았을 것이라는 가설을 세웠으나 관련 없음 확인

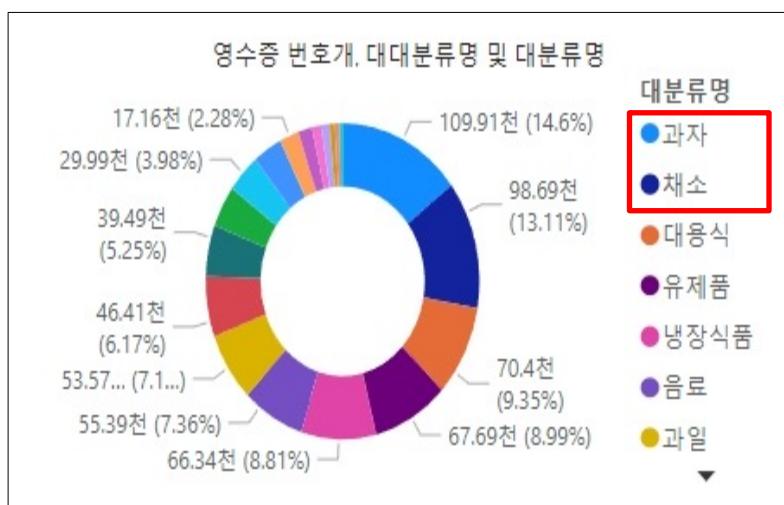
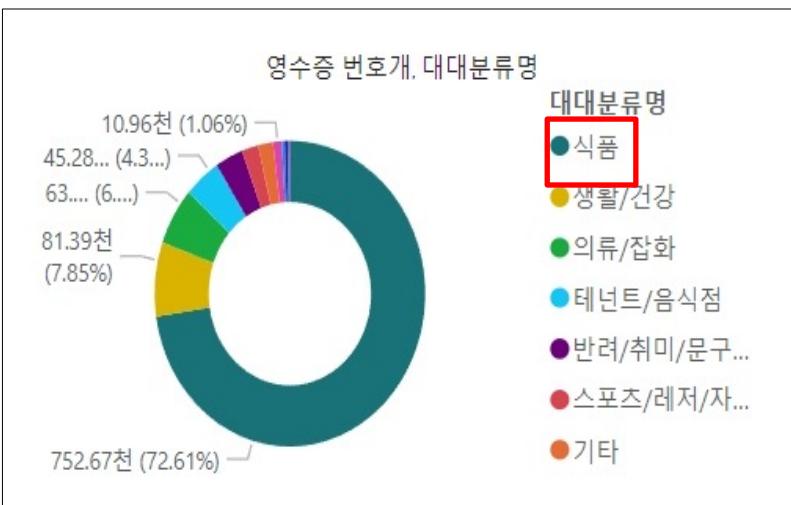
2.3 엘폐이: 유통사와 마찬가지로 A01의 매출에 따라 전체 매출이 큰 영향을 받음

2.1 데이터 탐색

II. EDA 및 인사이트 분석

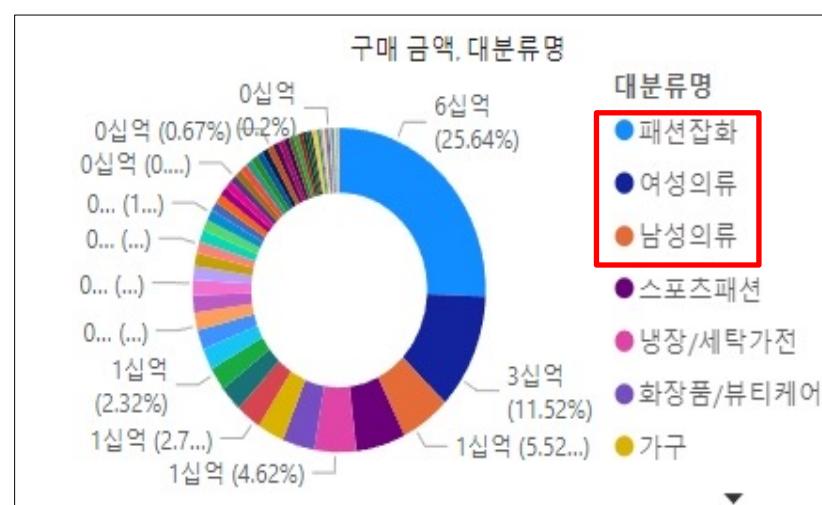
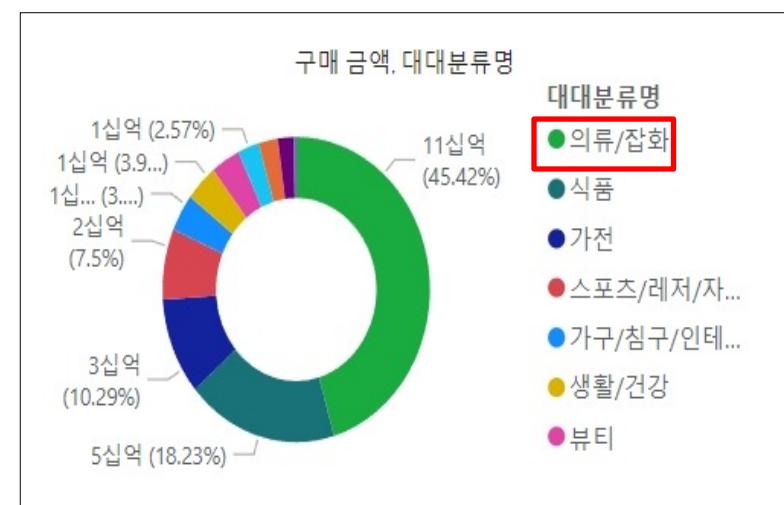
- EDA 분석을 통한 데이터 탐색
 - 타겟층의 상품 소비 패턴 분석

< 상품 분석 1 : 가장 많이 구매하는 품목 >



가설	채소, 유제품 등의 식자재가 가장 많이 구매된 품목으로 예상
결과	과자 14.6% 채소 13.11% (+1.5%↑) 가량 더 높은 비율로 → 판매율 1위

< 상품 분석 2 : 매출이 가장 큰 품목 >



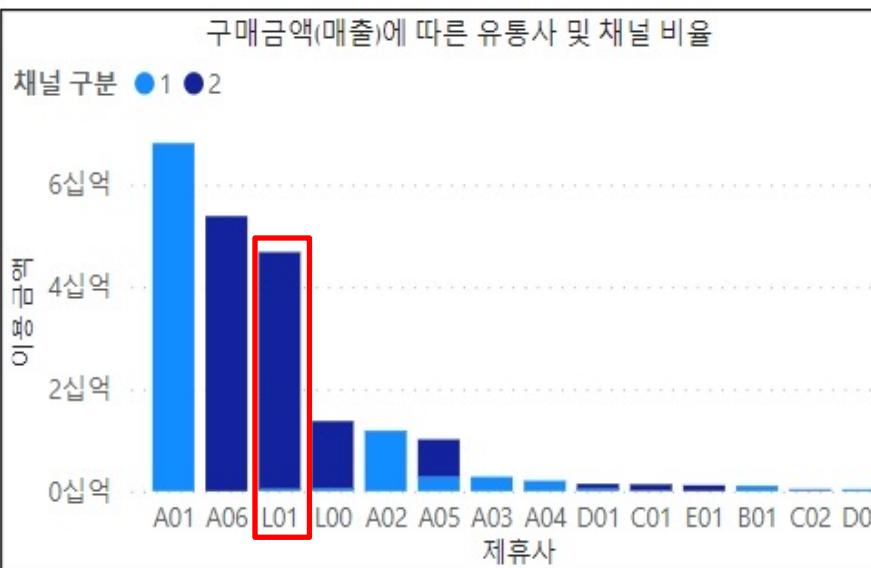
가설	30-40대 여성의 최다 소비 상품군은 육아/식품군으로 예상
결과	육아 분야가 아닌 의류/잡화 중 가방, 야외, 악세서리 등의 패션잡화, 여성의류 즉, 본인이 직접 착용할 상품을 구입하는데 지출이 많음
비고	- 유아동의류는 대분류 기준 매출 1.96%, 횟수 12.9% - 여성이 구매하는 남성 의류도 상당 부분 차지함 (전체 매출 3위)

2.1 데이터 탐색

II. EDA 및 인사이트 분석

EDA 분석을 통한 데이터 탐색 – 엘폐이 이용 현황

< 06 엘폐이 이용 내역 >



1 오프라인 2 온라인



<엘폐이 사용 현황>

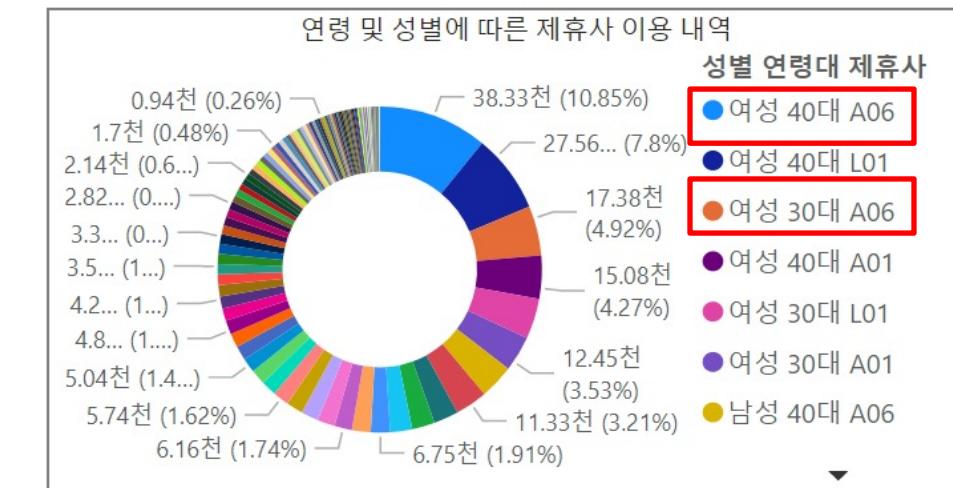
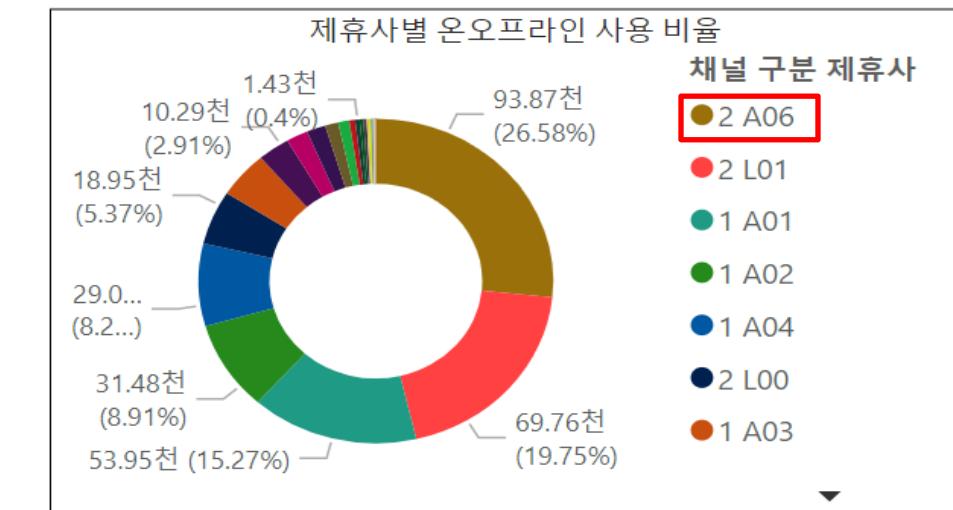
- 대부분 온라인에서 사용, **사용률이 가장 높은 유통사는 A06**
- ⇒ 온라인 매출의 활성화를 목표로 하고 있기 때문에 타겟으로 설정하기 적절
- ↔ 엘폐이(간편결제 시스템)가 온라인에 특화 되어있음에도 불구하고 **A01(오프라인)의 매출이 가장 높음**

< 제휴사 및 비제휴사 >

- (+) L01(비제휴사)에서의 엘폐이 이용 금액이 3위, 구매 횟수가 2위를 차지
→ 30만개의 엘폐이 온·오프라인 가맹점에서의 이용이 가능한 메리트를 직접적으로
- 본여줌

엘폐이 사용률 제휴사(B01~E01) < 유통사 & 비제휴사

<온·오프라인 별 제휴사 이용 비율>



온라인 마케팅 – 유통사 위주의 개인화 마케팅 고안
→ **A06**을 대상으로 선정하기에 적절하다고 결론

2.2 데이터 전처리 및 가공

- 결측치, 이상치, 중복값 탐색을 통한 데이터 전처리

< 결측치 >

chnl_dv	cop_c	br_c
2	A01	

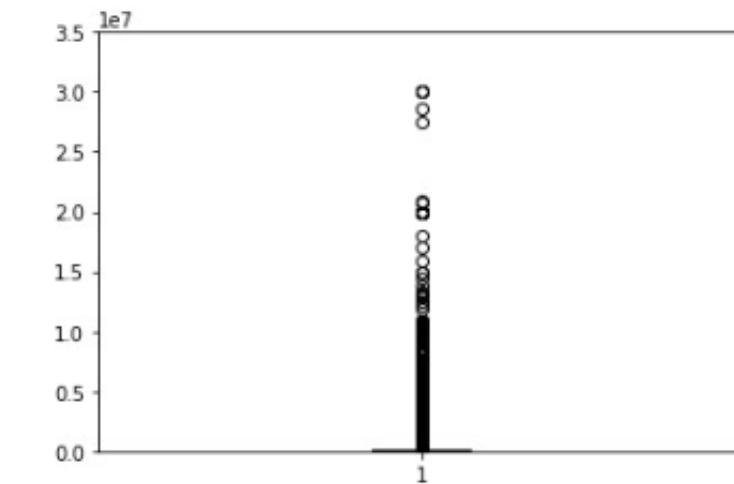
2번 파일 온라인의 점포코드는 결측치
→ 0으로 채움

< 중복값 >

cust	rct_no	de_dt	de_hr	buy_am	buy_ct	pd_nm
M000225114	A01073830386	2021-03-02	10	8500	1	일반스낵
M000225114	A01073830386	2021-03-02	10	12900	1	일반스낵
M000225114	A01073830386	2021-03-02	10	7900	1	일반스낵
M000225114	A01082337204	2021-03-07	20	9000	1	일반스낵
M000225114	A01115043060	2021-04-01	18	1000	1	일반스낵
M000225114	A01133564660	2021-04-13	18	2000	1	일반스낵
M000225114	A01133564660	2021-04-13	18	1000	1	일반스낵
M000225114	A01133564660	2021-04-13	18	3000	2	일반스낵
M000225114	A01142874144	2021-04-20	18	9000	1	일반스낵

→ 중복값은 각자 다른 데이터로서 삭제하지 않음

< 이상치 >



구매금액 이상치는 삭제하지 않고
로그화로 처리
이상치에 해당하는 금액을 소비한
고객은 충성 고객일 것이라고 판단
→ 매출의 중요한 축을 담당하므로
삭제하기 적절치 않다고 판단
→ 로그화를 통해 데이터 사용

de_dt	vst_dt
20201231	20201231
20201231	20201231
20201230	20201231
20201231	20201231

3번 파일 내 포함되어 있는
전년(2020년) 데이터 삭제,
이용일자 기준으로 데이터 활용

< 데이터 타입 변환 >

```
# 데이터 타입 변환
PDDE['de_dt'] = PDDE['de_dt'].astype('str')
PDDE['de_dt'] = pd.to_datetime(PDDE['de_dt'])
PDDE['buy_am'] = PDDE['buy_am'].astype('int')
```

de_dt(구매 일자) 칼럼 데이터: 실수 → 문자 → datetime 변경
buy_am(구매 금액) 칼럼 데이터: 실수 → 정수 변경

2.2 데이터 전처리 및 가공

II. EDA 및 인사이트
분석

· 데이터 가공 시 발생 문제점과 파생변수 생성

< 문제점 >

파일 명	02 상품 구매 정보	03 제휴사 이용 정보	06 엘페이 이용
영수증 번호	A01000011560	21102612B015763935	210803190302973
	A02208132879	21122311C015770490	210803190302973
비고	유통사에 따라 영수증 앞자리가 결정	제휴사 이용 정보에 따라 가운데 영수증 번호 결정	사용일자에 따라 앞 6개 자리 영수증 번호 결정

파일 병합 시 문제점



1. 영수증 번호가 모두 다를 때
2. 파일 별 제휴사, 유통사 정보를 활용하는 칼럼 명과 데이터가 모두 다를 때
3. 06번 파일 고객 < 02번 파일 고객

<파생 변수>

파생변수	설명	활용
GROUP	상품 분류 정보 파일 중 clac_hlv_nm 의 상위 그룹	타겟층의 소비 패턴을 분석할 때 사용
DE_DAY	각 파일 de_dt 에 해당하는 요일 변수	각종 시계열 분석 시 사용
DE_DT_MONTH	각 파일 de_dt 에 월만 추출한 월별 변수	각종 시계열 분석 시 사용
뉴스 심리 지수	인터넷 경제뉴스 데이터를 문장 단위의 긍정, 부정, 중립 감성으로 분류하고, 긍/부정 문장의 차이를 지수화	상관관계 분석 및 관계성이 높은 변수를 기준으로 모델 고도화에 사용
트렌드 지수	카카오 데이터 트렌드에서 일별 '롯데'를 검색한 횟수를 서울/경기 지역을 기준으로 나타냄	상관관계 분석 및 관계성이 높은 변수를 기준으로 모델 고도화에 사용

2.2 데이터 전처리 및 가공

- 정확한 결과값 도출을 위한 **로그화/정규화** 진행

로그화

로그화를 통해 데이터 간 편차를 줄여주어 정규성을 높여 정확한 값 도출

☞ 문제점 : 이용금액(buy_am)은 이상치를 제거하지 않았고 데이터 간의 편차가 커서 그대로 사용 시 데이터의 정확도가 떨어짐

→ 결과 : 데이터 간 편차를 줄여 결과값이 더 정확해짐

정규화

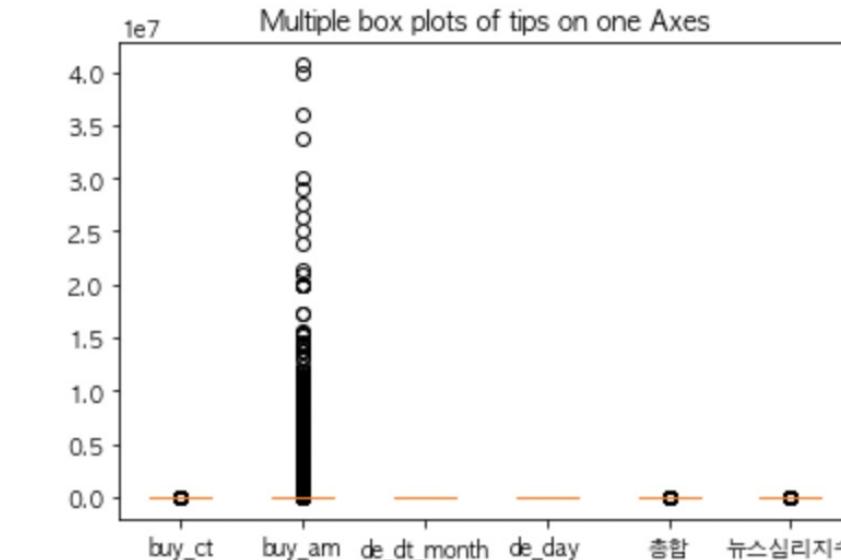
최소 최대 정규화 (Min max normalization) 이용
데이터의 범위가 [0 ~ 1]가 되도록 비례적으로 맞추기

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

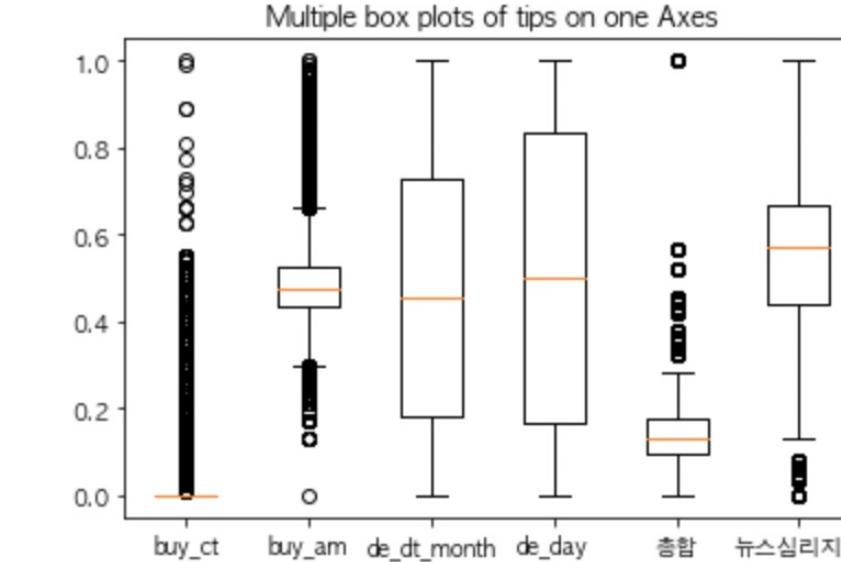
☞ 문제점 : 데이터의 단위가 모두 다르기 때문에 RMSE 값이 일정하지 않고 다르게 나옴
→ 상대적인 크기 차이를 제거해야 함

→ 결과 : 데이터의 범위를 모두 [0 ~ 1]로 바꾸어 줌

< 원본 데이터 >



< 로그화/정규화한 데이터 >



2.2 데이터 전처리 및 가공

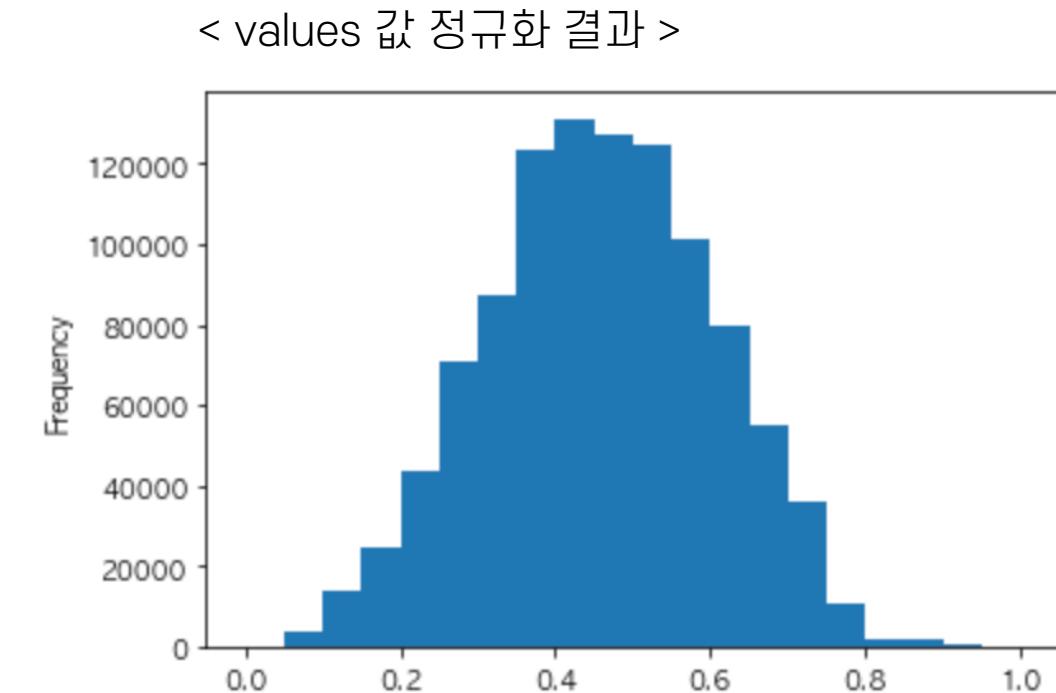
- 모델에 사용할 데이터 전처리 과정 (파생변수, 중복 데이터 묶기, 데이터 저장)

- ✓ 정규화한 데이터들을 더해 평가에 사용될 values 값 만들기

→ values 값 생성 후 다시 한 번 정규화를
함

buy_ct	buy_am	de_dt_mont h	de_day	총합	뉴스심리지수
0.0	0.516377	0.181818	0.166667	0.221591	0.435341
0.0	0.540185	0.181818	0.166667	0.221591	0.435341
0.0	0.512199	0.181818	0.166667	0.221591	0.435341
.

values
0.251112
0.258547
0.249808
.



- ✓ 'cust'와 'pd_nm' 값을 기준으로 중복 데이터 묶기 (groupby 사용)

- ✓ 데이터를 빠르게 로드할 수 있도록 피클(pickle) 형태로 데이터 저장

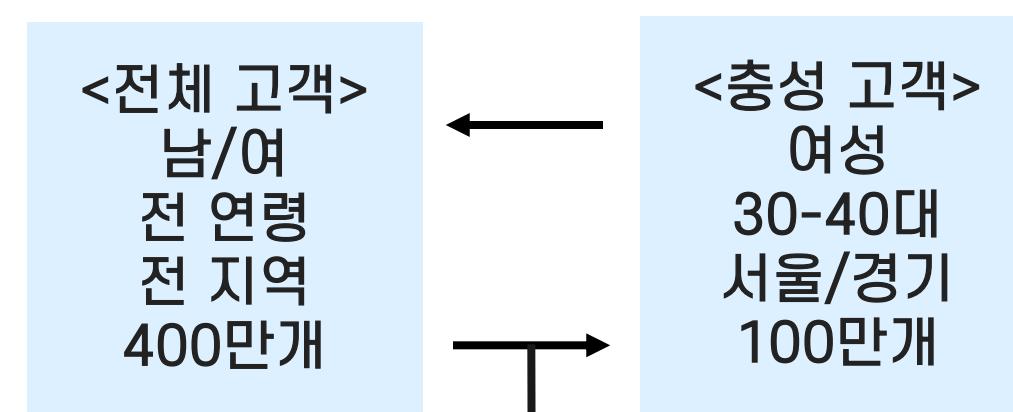
2.2 데이터 전처리 및 가공

- 파생변수 추가 및 EDA 분석 결과에 따른 타겟층 파일 생성

파생변수 추가

de_dt_month	de_day	group
3	1	식품
3	1	식품
3	1	식품
3	6	식품
4	3	식품
4	1	식품
4	1	식품
.	.	.

EDA 분석 후 타겟층 선정



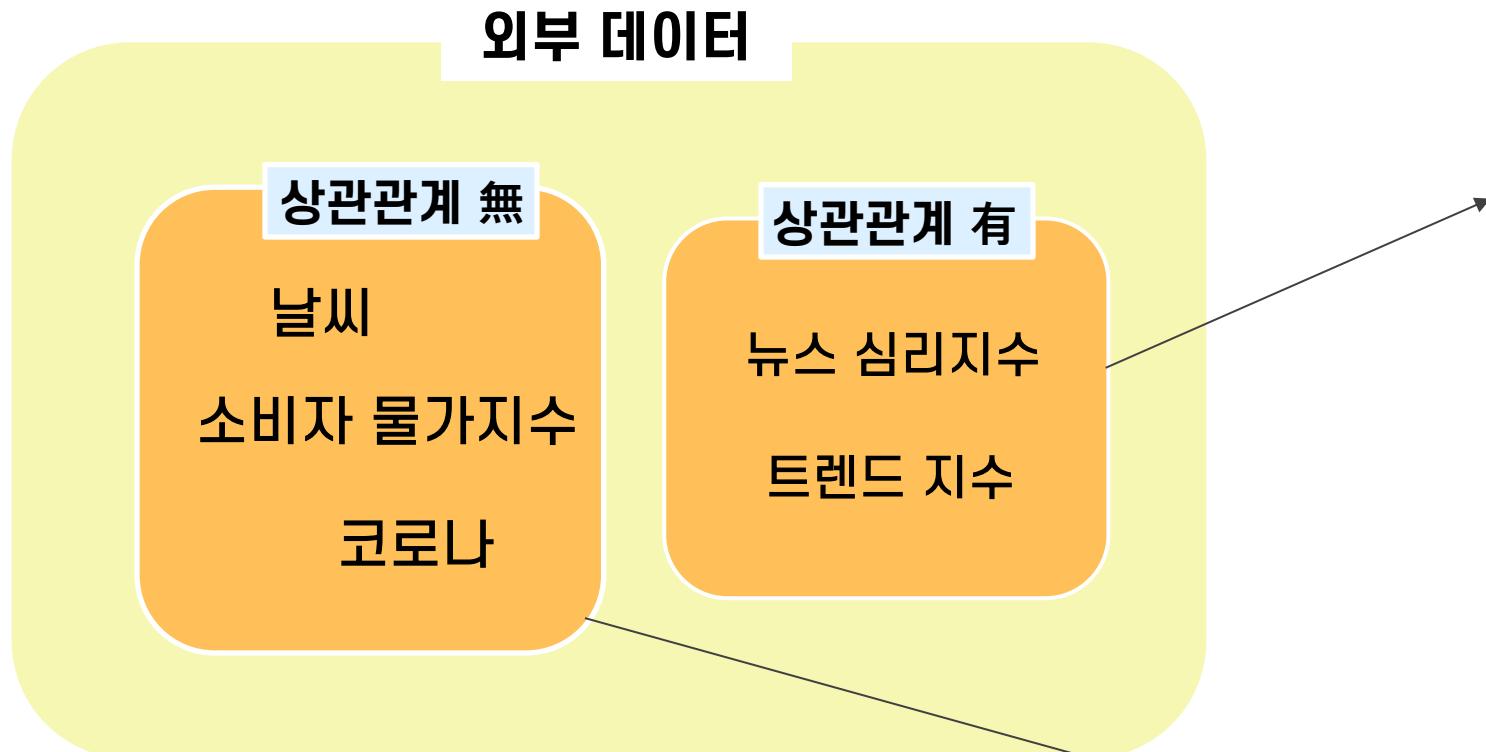
타겟층 한정 새 파일 생성

cust	ma_fem_dv	ages	zon_hlv
M000419293	여성	40대	Z17
M000494848	여성	30대	Z10
M000705571	여성	40대	Z17
M001714196	여성	40대	Z17
M003468477	여성	30대	Z17
M003531710	여성	30대	Z17
M003765725	여성	40대	Z10
M003834772	여성	40대	Z17

추후 전체 고객
대상으로 확대 가능

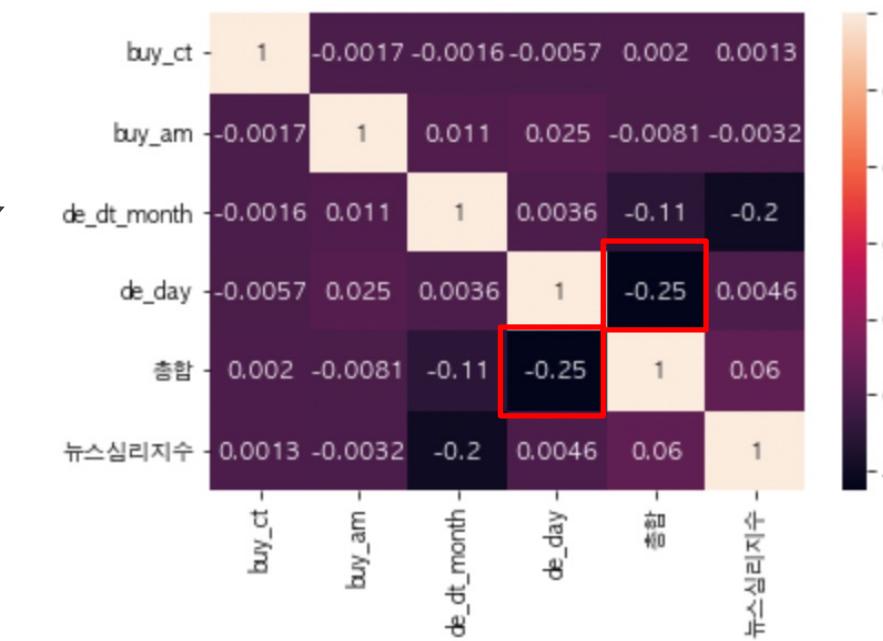
2.3 외부 데이터 선정

- EDA 분석과 상관분석을 통한 타겟 선정 – 외부 데이터 선별 과정



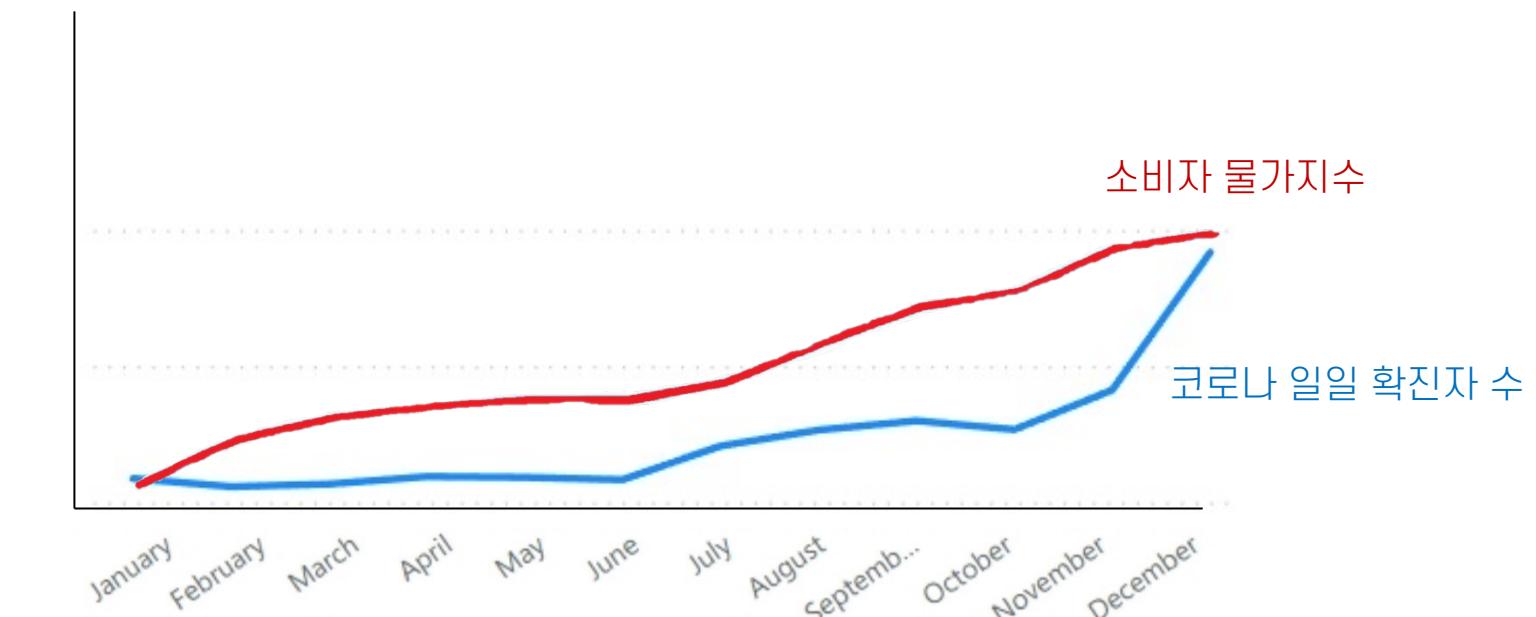
소비자의 소비 패턴에 영향을 끼칠 수 있는 요소 탐색
→ 히트맵 및 시각화 상관계수 확인 후 데이터 선별

- 뉴스 심리지수: 경제심리지표 및 실물경제지표에 1-2개월 선행하여 높은 상관관계를 나타냄



<히트맵 상관계수>

뉴스심리지수-트렌드지수 총합-요일(de_day)
뉴스심리지수-월(de_dt_month)
→ 음의 상관관계를 나타냄



2.3 외부 데이터 선정

- 모델에 활용할 타겟층 고객 파일과 외부 데이터 셋 병합

data_F_3040을 기준으로 뉴스 심리 지수, 트렌드 데이터 병합

cust	buy_am	buy_ct	...	de_dt	de_hr
M000225114	8500	1	...	2021-03-02	10
M000419293	21800	2	...	2021-04-24	12
M000494848	15000	1	...	2021-04-07	12

【 data_F_3040 】

【 뉴스 심리 지수 】

날짜(de_dt)	뉴스 심리 지수
2021-01-01	107.97
2021-01-02	110.03

【 트렌드 지수 】

일(de_dt)	서울	경기	총합
2021-01-01	16	15	31
2021-01-02	24	25	49

2.3 외부 데이터 선정

II. EDA 및 인사이트
분석

- 모델에 사용할 데이터 전처리 후 필요한 열 추출

【필요한 열 추출】

cust	pd_nm	buy_ct	buy_am	de_dt_month	de_day	총합	뉴스심리지수
M000225114	일반스낵	1	8500	3	1	63	107.35
M003531710	일반스낵	1	1500	3	1	63	107.35
M818104759	이유식	1	2700	4	3	59	117.47
M281654571	디저트	2	8600	9	6	51	111.03
M260717185	종량제봉투	1	540	8	2	53	100.48
.
.
.

* 총 1036659개의 행, 8개의 열 데이터 생성

III. 모델 개발

- 데이터 아키텍쳐
- 모델 선정
- 모델 개발
- 모델 평가지표
- 성능 개선
- 결과 검증

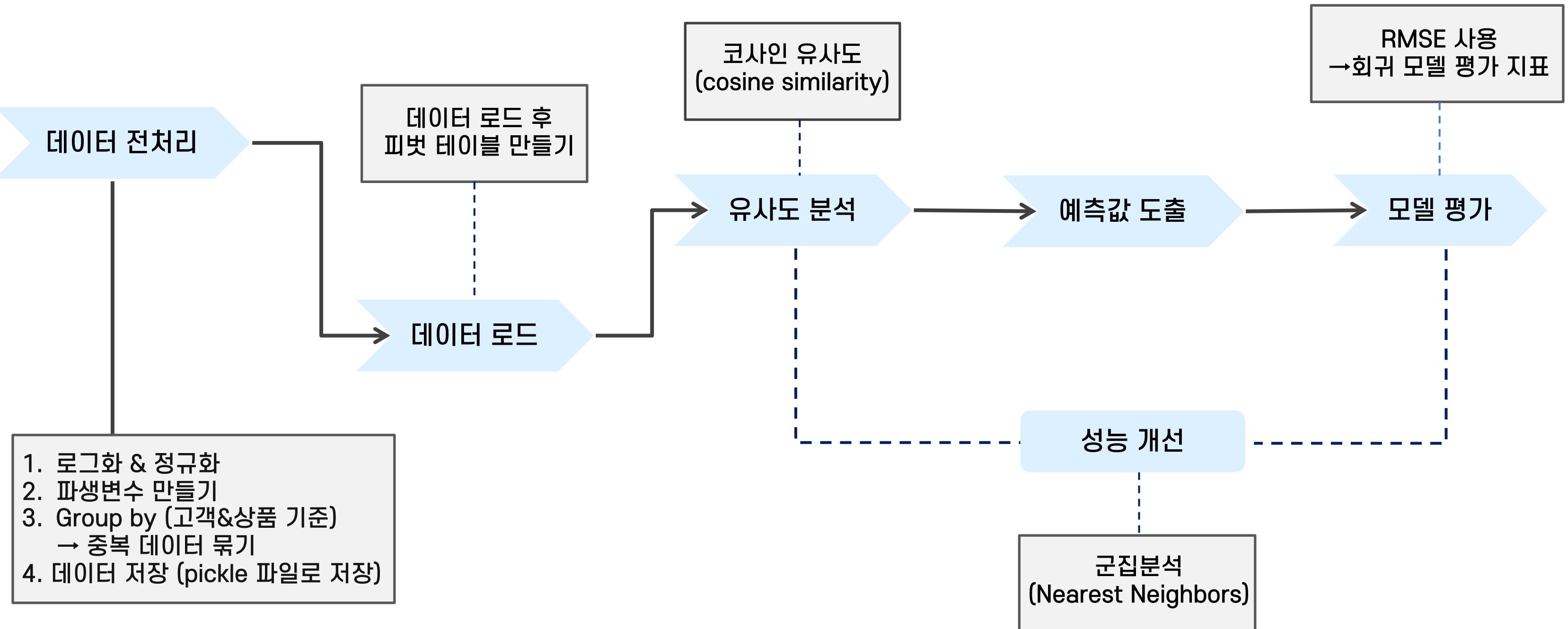
03

롯데멤버스
빅데이터 경진대회

3.1 데이터 아키텍쳐

III. 모델 개발

- 협업 필터링 모델 내 사용되는 데이터 셋의 전체적인 프로세스 시각화

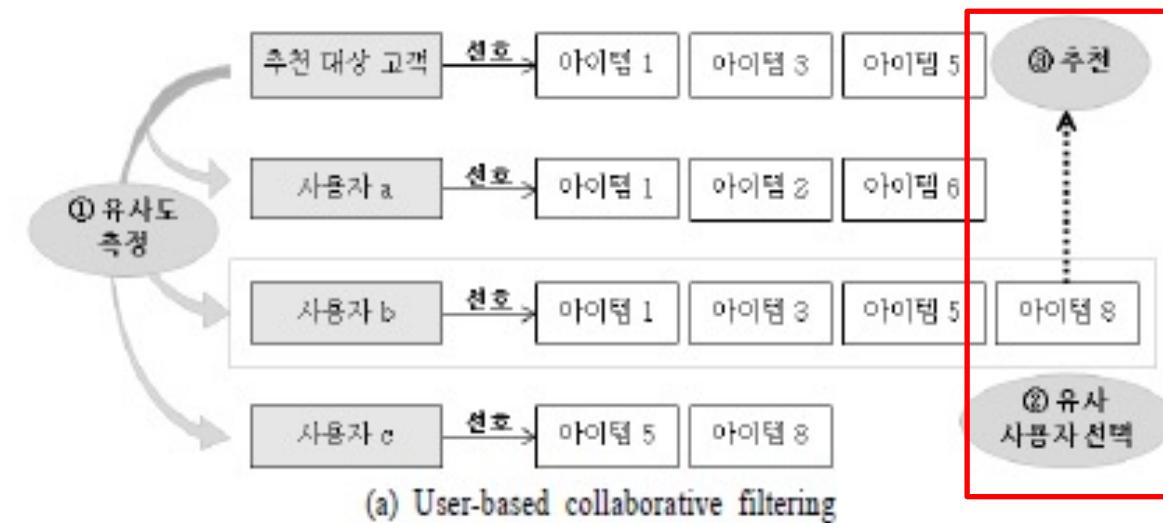


3.2 모델 선정

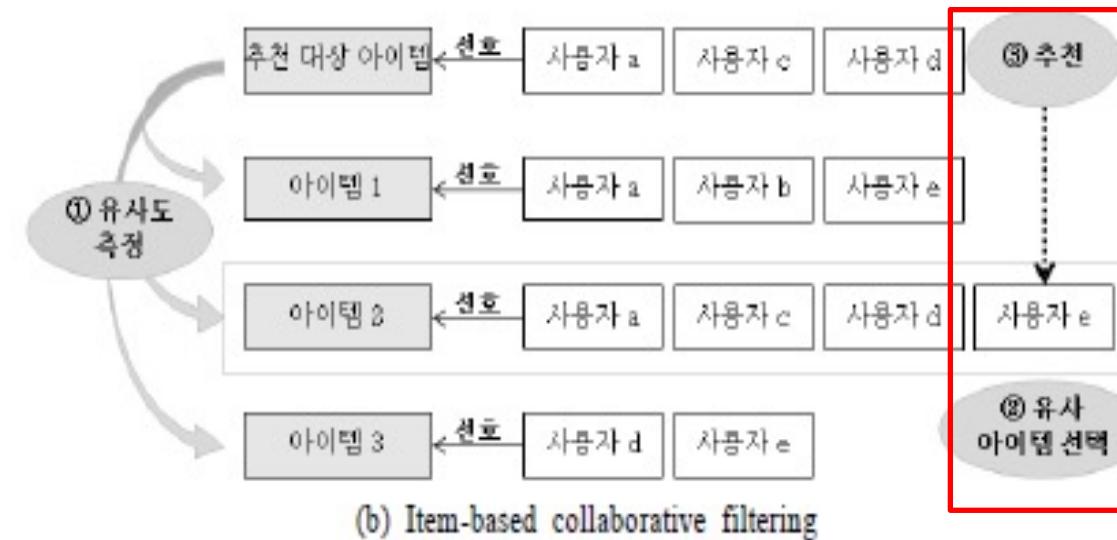
III. 모델 개발

• 협업 필터링의 장단점을 고려한 모델 설정

< 사용자 기반 협업 필터링 >



< 아이템 기반 협업 필터링 >



✓ 장점

- 아이템 자체의 정보 없이 추천 가능
- 알고리즘이 간단하여 구축하기 쉬움
- 결과의 설명력이 좋음

✓ 단점

- 초기 데이터 양이 충분하지 않을 때 추천 정확도가 떨어짐 (Cold Start 문제)
- 데이터가 너무 많아지면 속도가 저하됨 (확장성) → 군집 분석 적용
- 신규 가입자의 경우 추천 정확도가 떨어짐 → 아이템 기반 협업 필터링 적용

✓ 사용자 기반 협업 필터링의 단점을 보완하기 위해 추가적으로 사용

- 기존 고객의 구매 이력을 기반으로 추천해주는 사용자 기반 협업 필터링은 신규 가입자에 대한 추천 정확도가 떨어진다는 단점을 보완하기 위해 사용

✓ 한계점

- 여전히 Cold Start 문제가 발생
→ 신규 가입자의 기본 정보(나이, 성별, 거주지 등)를 활용해서 추천
→ 기본 정보도 부족하다면 전체적으로 인기 많은 제품을 추천

3.3 모델 개발

III. 모델 개발

- 데이터 로드 후 피벗 테이블 만들기

```
ratings = data.pivot_table(index='cust', columns='pd_nm', values='values')
ratings.fillna(0, inplace=True) # na 값 0으로 채우기
```

【ratings 피벗 테이블】

cust \ pd_nm	2단 우산	3단우산	BB/파운데이션/컴팩트류	DIY 가구	...	휴지통	흑미	히터	힙색/사이드백
cust	2단 우산	3단우산	BB/파운데이션/컴팩트류	DIY 가구	...	휴지통	흑미	히터	힙색/사이드백
M000225114	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
M000261625	0.0	0.0	1.275642	0.0	...	0.0	0.0	0.0	0.0
M000419293	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
M000494848	0.0	0.0	0.527808	0.0	...	0.0	0.0	0.0	0.0
..
M997831586	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
M998680257	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
M999227380	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
M999268192	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0

* 5705 rows x 1836 columns

3.3 모델 개발

III. 모델 개발

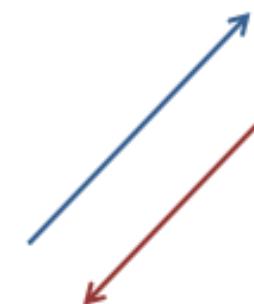
• 코사인 유사도를 이용해 고객 간 유사도 도출

✓ 코사인 유사도의 사용

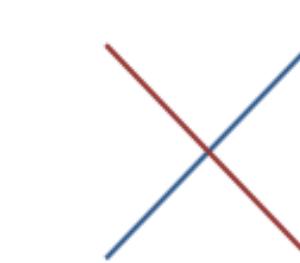
- 사용자 기반의 협업 필터링의 핵심은 사용자 간 유사도를 구하는 것
- 대표적인 유사도 측정 척도에는 코사인 유사도, 피어슨 상관계수가 있음
- 논문에 따라서 코사인 유사도를 사용해 유사성을 도출하기로 함

✓ 코사인 유사도란?

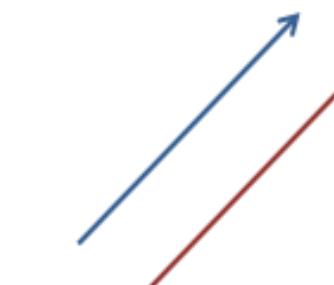
- 내적 공간의 두 벡터 간 각도의 코사인 값을 이용해 두 벡터가 얼마나 유사한지 측정하는 방법
- 두 벡터의 방향이 180도 반대이면 -1, 90도 각을 이루면 0, 완전히 동일한 경우엔 1의 값을 가짐



코사인 유사도 -1



코사인 유사도 0



코사인 유사도 1

→ 코사인 유사도는 [-1 ~ 1] 사이의 값을 가지며 1에 가까울 수록 유사도가 높다고 판단

【 사용자 유사도 】

	사용자 A	사용자 B	사용자 C	사용자 D	사용자 E
사용자 A	1.000000	0.000000	0.035050	0.061841	0.000000
사용자 B	0.000000	1.000000	0.000000	0.167633	0.069774
사용자 C	0.035050	0.000000	1.000000	0.028871	0.000000
사용자 D	0.061841	0.167633	0.028871	1.000000	0.135265
사용자 E	0.000000	0.069774	0.000000	0.135265	1.000000

< 유사도가 높은 사용자 찾기 예시 >



recommend_user('M999227380')	
cust	
M999227380	1.000000
M801858738	0.865736
M267665211	0.843617
M360187347	0.821696
M967393794	0.803842
M369057001	0.774366

3.3 모델 개발

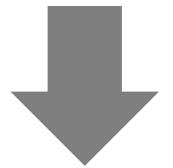
III. 모델 개발

· 유사도를 기반으로 예측값을 도출하고 새로운 상품 추천

- ✓ 사용자 유사도를 기반으로 예측값을 도출함
- ✓ 전체 상품 리스트와 고객이 구매한 리스트를 비교하여 구매하지 않은 상품 리스트를 생성
- ✓ 고객이 구매하지 않은 상품 리스트 중 예측값이 높은 상위 5개 상품 추천

【 예측값 도출 】

	상품 1	상품 2	상품 3	상품 4	상품 5	상품 6	상품 7
사용자 A	0.600323	0.459231	0.035612	0.003765	0.000000	0.497322	0.003941
사용자 B	0.000279	0.000000	0.232847	0.717943	0.000871	0.519183	0.493714
사용자 C	0.023759	0.492198	0.000000	0.794421	0.000658	0.746021	0.582017
사용자 D	0.015171	0.167633	0.014809	0.983641	0.000000	0.002396	0.003362
사용자 E	0.000109	0.010202	0.641457	0.641555	0.012987	0.034674	0.495713
사용자 F	0.012021	0.003823	0.610339	0.000000	0.600342	0.006581	0.000000
사용자 G	0.000000	0.23939	0.000543	0.487175	0.514822	0.392105	0.374962



【 새로운 상품 추천 예시 】

pd_nm	pred_score
일반스낵	0.997749
종량제봉투	0.717935
임대매출	0.641555
일반계란	0.601339
국물봉지라면	0.600323

3.4 모델 평가 지표

III. 모델 개발

- 회귀 모델 평가 지표인 RMSE를 사용해 모델 평가

✓ 평가지표 - RMSE



논문 탐색

『추천 시스템 기법 연구동향 분석』

추천 정확도는 아이템의 점수를 예측하는 알고리즘이 사용되고
실제 선호도와 예측 값의 차이로 계산하여 널리 사용되는 방법으로 RMSE 가 있다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Divide by the total number of data points

Predicted output value

Actual output value



MSE와 RMSE 모델 비교 후 선정

MSE(Mean Squared Error)

- 실제 값과 예측 값의 차를 제곱한 평균
- 값이 낮을수록 좋음
- 장점: 지표 자체가 직관적이고 단순함
- 단점: 에러를 제곱하기 때문에 값의 왜곡이 생김

RMSE(Rooted Mean Squared Error)

- MSE에 루트를 씌운 값
- 값이 낮을수록 좋음
- MSE의 장단점을 거의 그대로 따름
- 에러를 제곱한 뒤 다시 루트를 씌움
→ 왜곡을 줄여줌 (MSE의 단점 보완)
- MSE보다 에러 그 자체의 평균을 직관적으로 보여줌

RMSE 선정 이유

MSE의 단점을 보완한 평가 지표로 에러를 제곱한 뒤 다시 루트를 씌워 값의 왜곡을 줄여 주기 때문에 선정
→ 이상치의 영향을 적게 받음 (덜 민감함)

RMSE 결과 값

train: 2.037621899809216

test: 2.1644356924512076

3.5 성능 개선

III. 모델 개발

• 군집 분석을 통한 평가 지표 개선



논문 탐색

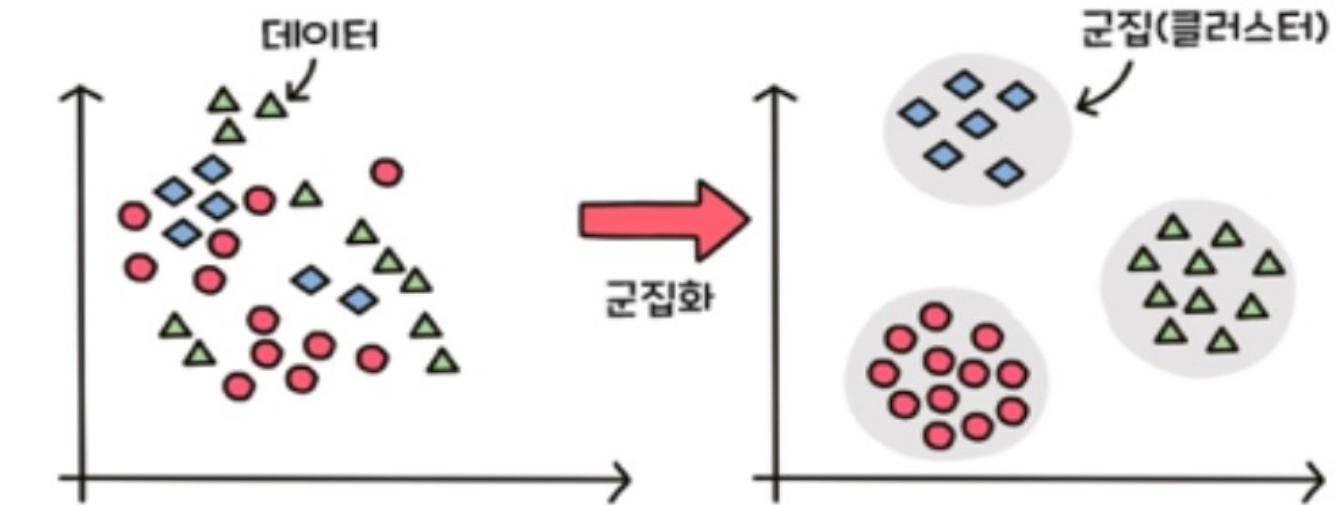
『협동적 필터링을 이용한 K-최근접 이웃 수강 과목 추천 시스템』
추천 시스템의 예측 능력을 향상시키기 위하여 사용자와 유사한 선호도를
가지는 이웃을 선정하는 기법으로 K-최근접 이웃 방법을 사용한다.

군집화(Clustering)

- 데이터를 유사한 정도에 따라 군집으로 분석하는 것
- 다양한 데이터들이 서로 섞여 있지만, 군집화 과정을 진행하면 비슷한
데이터끼리 군집으로 묶을 수 있음
- 사용 이유: 협업 필터링의 문제점인 확장성(scalability)의 문제점 해결
(데이터 셋이 커질수록 컴퓨터 계산량이 많아지며 예측력과 정확도 저하와 직결)
- 평가지표(RMSE) 값을 개선을 위해 사용

최근접 이웃(Nearest Neighbors)

- 군집화의 대표 알고리즘
- 비지도 학습용으로 이웃을 검색할 때 사용
- 가장 비슷한 몇 명의 사용자끼리만 비교하도록 개선



【군집 분석 후 RMSE 값 개선】

train: 2.037621899809216
test: 2.1644356924512076

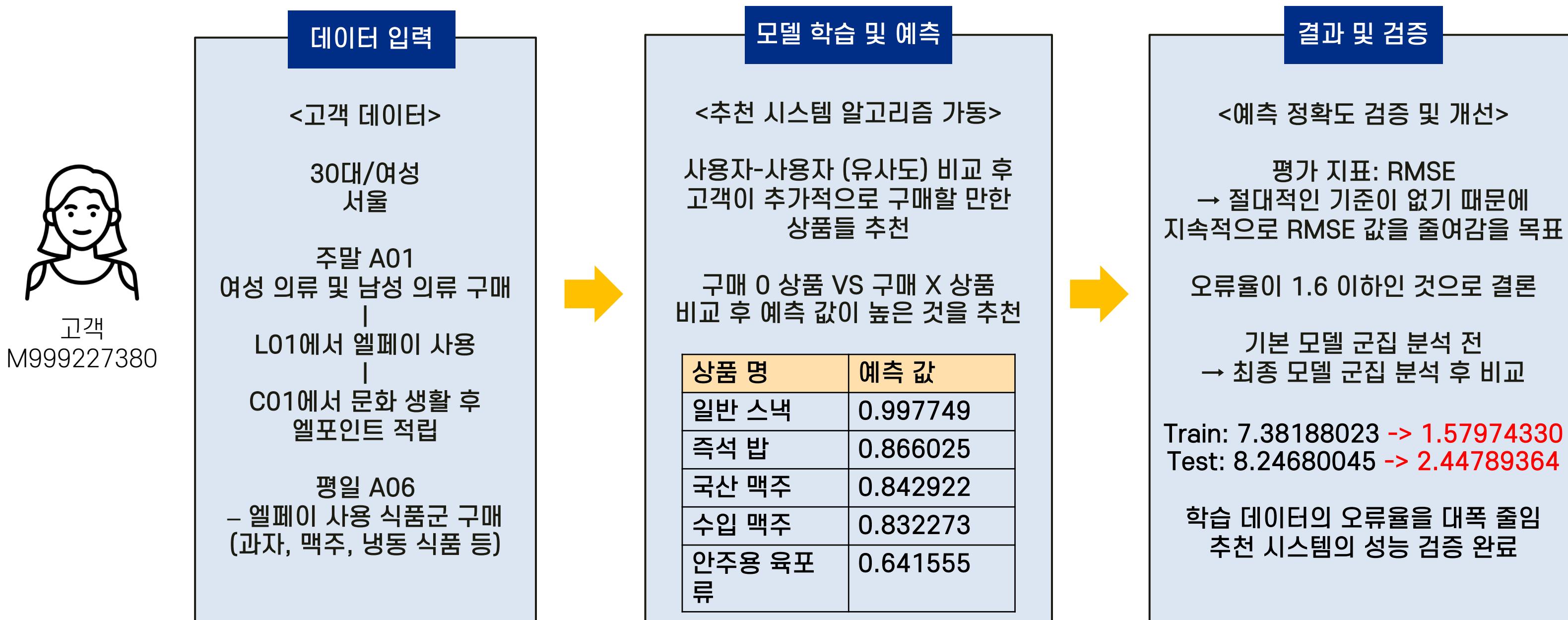
train: 1.5797433059454178
Test: 2.4478936412833026

train 값만 하락, test 값 상승한 이유
→ 데이터 양이 불충분하기 때문이거나
데이터 자체의 샘플 문제로 추측

3.6 결과 검증

III. 모델 개발

- 추천 시스템 실행 시 검색 결과에 대한 검증



04

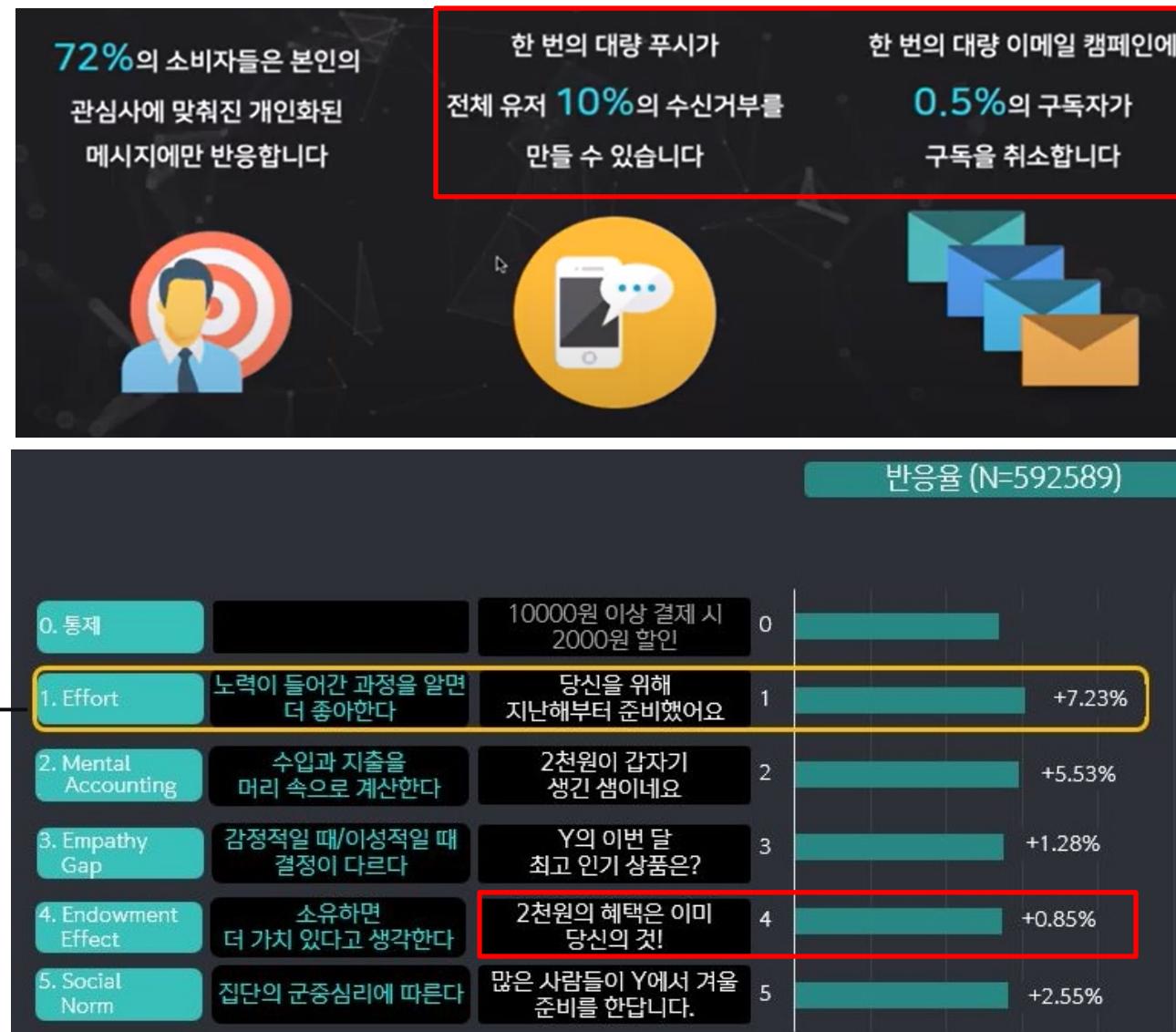
롯데멤버스
빅데이터 경진대회

IV. 마케팅 제안

- 마케팅 제안 근거
- 마케팅 제안 요약
- 온라인 마케팅 제안
- 제휴사 활용 방안
- 엘폐이 활용 방안
- 서비스 기대 효과

4.1 마케팅 제안 근거

· 개인화 마케팅의 필요성 대두



- 모든 고객에게 일률적으로 마케팅을 진행하는 것은 역효과 발생
- 개인화 마케팅의 긍정적 효과와 동일하게 해석

논문 - 『개인화 상품 추천이 구매의도에 미치는 영향에 관한 연구 : Matrix Factorization 알고리즘을 중심으로』

표 2. 실험결과

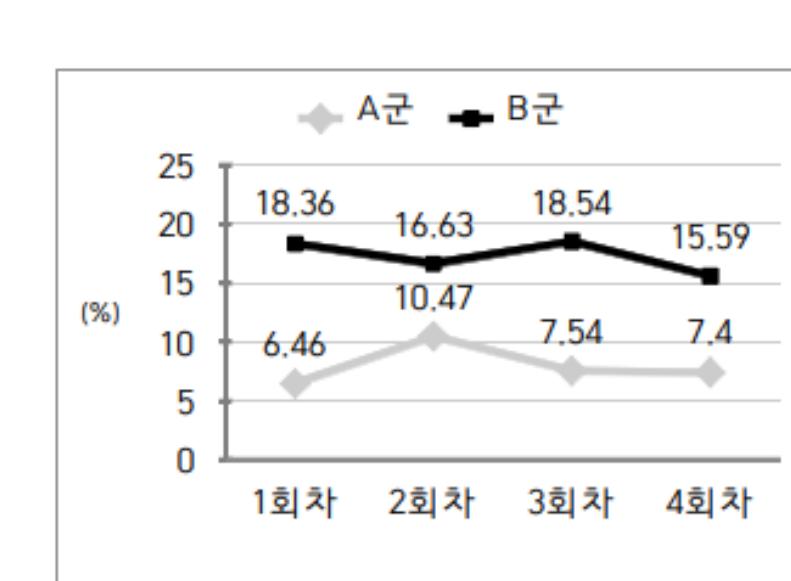


그림 4. 실험차수별 상품클릭 전환율(CTR)

구분	A (MD추천)	B (개인화추천)	비고
발송 메일수	408,716	372,597	-
메일 오픈수	159,376	147,914	-
메일 오픈비율	39%	39.7%	-
상품 클릭수	12,688	25,578	+101.6%
상품클릭 전환율	7.97%	17.3%	+117.2% P값 0.000
연관 거래액	17,598,510	38,198,040	+117.1%
메일당 거래액	110	258	+133.9%

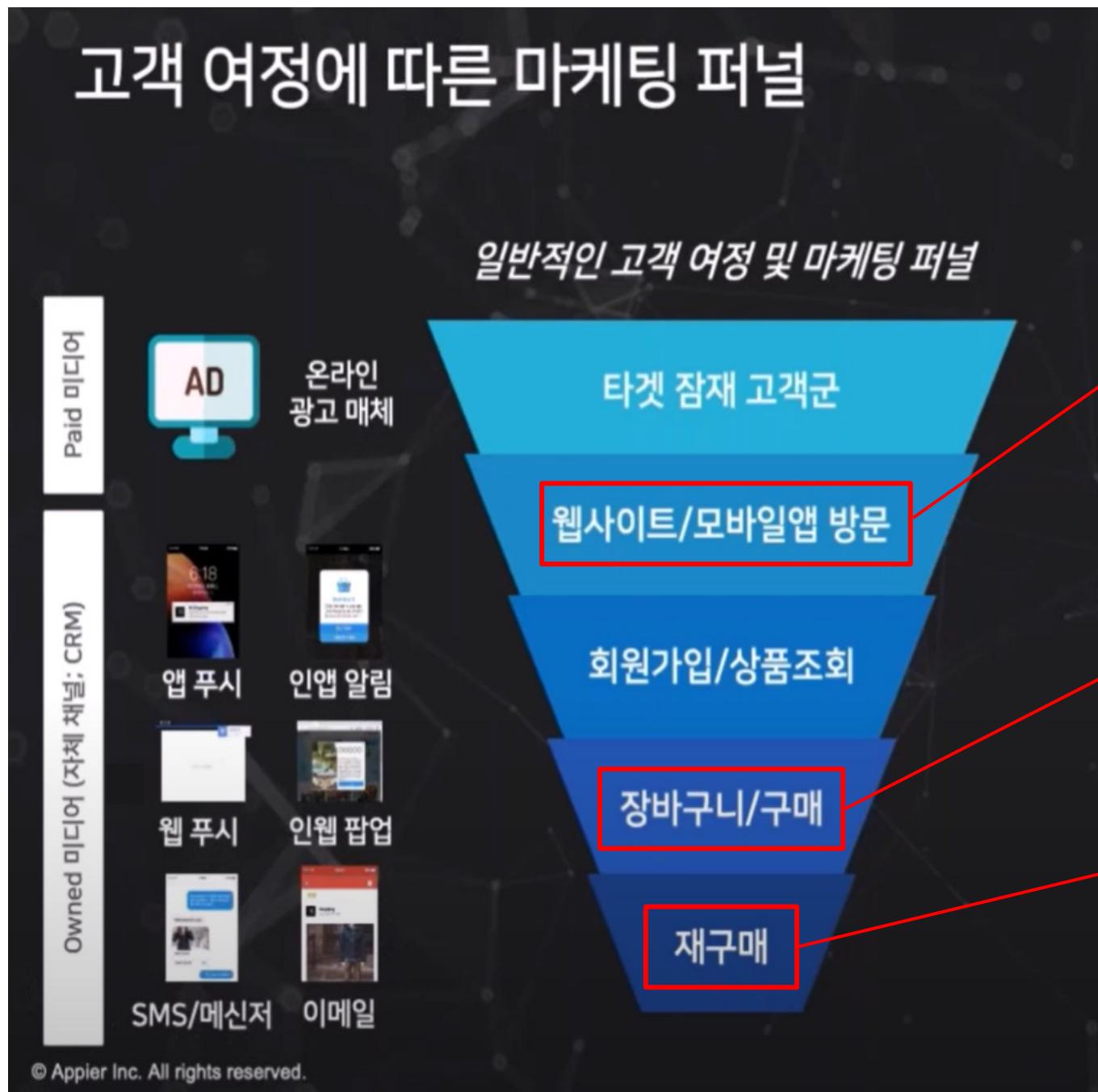
<실증 분석>

사람(MD)이 추천하는 상품 vs 추천 알고리즘이 추천하는 상품

<결과> 알고리즘에 의한 개인화 추천이 기존 방식 대비 상품 클릭 전환율 117.1% 증가하여 실제 효과가 있음을 검증

4.2 마케팅 제안 요약

· 고객 여정에 따른 마케팅 퍼널 기반 단계별 마케팅 제안



1. 웹 사이트 및 모바일 앱 방문 시

- 웹 사이트 / 블로그 / 포털 사이트에 알고리즘으로 도출한 개인화 맞춤 추천 상품 위주로 광고 및 홍보를 진행
- 모바일 어플도 같은 방식
- 이메일, 문자 발송
- 이 때, 사용자별로 많이 사용하는 요일/시간을 고려해 적용

2. 상품 조회/장바구니 및 구매 시

- 상품 탐색 시 추천 상품을 상단에 노출, 장바구니에 물건을 담으면 그와 관련된 상품 추천 (장바구니 추천)
- 롯데ON 홈페이지에 적용

3. 재구매

- 일정 기간이 지난 후 업데이트 된 정보를 바탕으로 재구매 유도를 위해 상품 추천 광고 및 홍보 어플 알림

※ 웹으로 접속했을 때 보다 모바일로 접속했을 때
개인화 마케팅 성공률이 더 높음

4.3 온라인 마케팅 제안

IV. 마케팅
제안

1. 웹 사이트 방문 유도 – 포털 사이트 광고 노출을 통한 고객 유입 유도



<추천 시스템> 비슷한 사용자가 산 상품 TOP5	
pd_nm	pred_score
콘아이스크림	0.997749
기타냉동간편식	0.717935
냉동간편식세트	0.641555
냉동디저트	0.601339
아이스크림케이크	0.600323

The screenshot shows the Naver Shopping homepage. A recommendation banner for Lotte On is displayed, targeting a 30-year-old female customer for instant food and dessert needs. The banner includes a deal for 2 items at 50% off. A red arrow points from the 'pd_nm' column of the recommendation system table to the banner, indicating the connection between the predicted scores and the recommended products.

30대 여성
오전
간편식
디저트

000 님, 더운 날씨엔 간단 조리하세요!
무더위 타파 간편식&아이스크림까지
★ 1+1/2개 50%★ 인기 먹거리 모음

고객 개개인에 맞춘 상품을
추천하여 매번 다른 화면

CATEGORY

everbody	SUNNYLOVE	MISS CANDY	닥터피엘	LOTTE ON	Gmarket	FASHIONPLUS
----------	-----------	------------	------	----------	---------	-------------

4.3 온라인 마케팅 제안

IV. 마케팅
제안

2. 웹 사이트 방문 유도 – 개인화 맞춤형 이메일 전송 활용

- 롯데ON [프로모션] (광고) 8월의 시작은 퍼스트먼데이 05:34
- 롯데ON [프로모션] (광고) 7월의 마지막 72시간! 10%+10% 더블할인 놓7 수 없지! (~7/31) 05:34
- 롯데ON [프로모션] (광고) 출국없이 즐기는 면세쇼핑 최대 84% 할인! 05:34
- 롯데ON [프로모션] (광고) [이로운X풀무원] 비건 맛있게 먹는 비법 알려주세요! 05:34
- 롯데ON [프로모션] (광고) 온세일 2차전 시작! [최대 20% 쿠폰] 받으셨나요? 05:34
- 롯데ON [프로모션] (광고) 7월 온세일로 가격사냥! 최대 20% 쿠폰 + 오늘의 특가 + BIG BRAND 할인 05:34
- 롯데ON [프로모션] (광고) 최대 25% 쿠폰! 시원스레 찐 특가! 최대 7% 카드 할인 더! 05:34
- 롯데ON [프로모션] (광고) 오늘 초복! 삼계탕부터 장어/전복까지 여름보양식 쟁기세요 05:34
- 롯데ON [프로모션] (광고) 온앤더뷰티 WEEK 백화점 뷰티 2종쿠폰 외 05:34
- 롯데ON [프로모션] (광고) 혜택 안고 찾아온 7월의 새친구 소개해요! 05:34
- 롯데ON [프로모션] (광고) [바이브온] ~70% 브랜드 할인은 내일까지! 05:34
- 롯데ON [프로모션] (광고) [롯데슈퍼 Fresh] 무더운 여름 쿨~한 장보기 주말엔 1만원 쿠폰 찬스! 05:34

05:34
07-30 05:35
07-28 18:12
07-23 05:34
07-21 17:43
07-18 17:44
07-17 17:26
07-16 06:22
07-14 17:47
07-11 17:41
07-09 17:23
07-08 07:37



더위 사냥! 가격 사냥↓
당신을 위한 온세일 쿠폰

1차 쿠폰
18 19 20
월 화 수
2차 쿠폰
21 22 23 24
목 금 토 일



쿠폰 한번에 받기 ↗

온세일 쿠폰 쇼핑 TIP

- 상품을 검색하고, 원피스 ↗
- 필터를 ON 해보세요. 온세일 ↗
- 스티커 확인 후, 쿠폰 적용 온세일 ↗

온세일 상품 보러가기 >

개인화 적용
↓

- 롯데ON [프로모션] (광고) 당신을 위한 맞춤형 장바구니템을 확인해보세요!
- 롯데ON [프로모션] (광고) 000고객님을 위한 온 세일 핫딜이 도착했어요!

08-04 09:05
08-03 08:45

기존 모든 고객에게 일괄적으로 보내는 할인 정보가 아닌
모델을 통한 개인화된 맞춤 상품에 대한 할인 정보를 맞춤 시간에
제공
→ 웹사이트 방문을 유도

4.3 온라인 마케팅 제안

IV. 마케팅
제안

1. 장바구니/구매 – 고객에게 개인 맞춤형 상품 추천 후 구매 유도

The screenshot shows the LOTTE Mart mobile application interface. At the top, there is a search bar with the text "대한민국 농활" and a magnifying glass icon. To the right of the search bar are icons for messaging (1 notification), "나의 찜" (My Favorites), "마이롯데", and a shopping cart icon labeled "장바구니" (Cart) with a red notification badge showing the number 3. Below the header, there is a navigation bar with categories: 홈, 전단행사, 베스트 (Best), HOT딜, 신상품, and 추석선물. A red arrow points from the "장바구니" icon to a callout box titled "<제안 사항>".

000 고객님을 위한 맞춤 상품 추천

Below the navigation bar, there are category icons for 전체 (All), 과일 (Fruit), 채소 (Vegetables), 쌀/잡곡 (Rice/Breakfast Cereals), 견과/건과 (Nuts/Confectionery), 정육/계란 (Meat/Eggs), 수산 (Seafood), 건해산물 (Processed Seafood), 밀키트/델리 (Meal Kit/Deli), and 냉장/냉동/간편식 (Refrigerated/Frozen/Convenience Food). A green arrow points from the "장바구니" icon to the "맞춤 상품 추천" section.

맞춤 상품 추천

The recommended products are displayed in a grid:

- 1** 온리프라이스 Only Price 1등급 우유 (930ML*2입)
3,500 원
★ 4.9 (288,596)
- 2** 제주 삼다수 그린 (2L*6입)
5,880 원
★ 4.9 (12,408)
- 3** 행복생생란 (특란, 30입) (1.8KG)
6,790 원
★ 4.8 (116,141)
- 4** 농심 농심 올리브 짜파게티 (5개입)
4,280 원
★ 4.9 (78,188)
- 5** 양반 양반 참기름 식탁김 (4G*16봉)
8,480 원
★ 4.7 (319)

1. 장바구니에 상품을 담으면
관련 상품 추천

2. 홈페이지 메인 화면에
맞춤 상품 추천을 띄워 노출 확률 ↑

3. 상품 간 가격 편차가 적은
식품군으로 제한하여 선택하여
상품 추천 정확도 ↑

⇒ 식품군에 한정되는 경향이
있지만
추후 다른 품목 및 다른 유통사
사이트에도 확대 적용 가능할 것

4.3 온라인 마케팅 제안

IV. 마케팅
제안

2. 장바구니/구매 – 온·오프라인 통합 기반 개인화 상품 구매 방안

오늘 장보기 고민중이세요?

우리동네 당일배송 점포의 인기상품이에요

우리동네 당일배송 점포에서
당신을 위한 상품을 추천해드려요

LOTTE Mart 오늘 20:50 도착

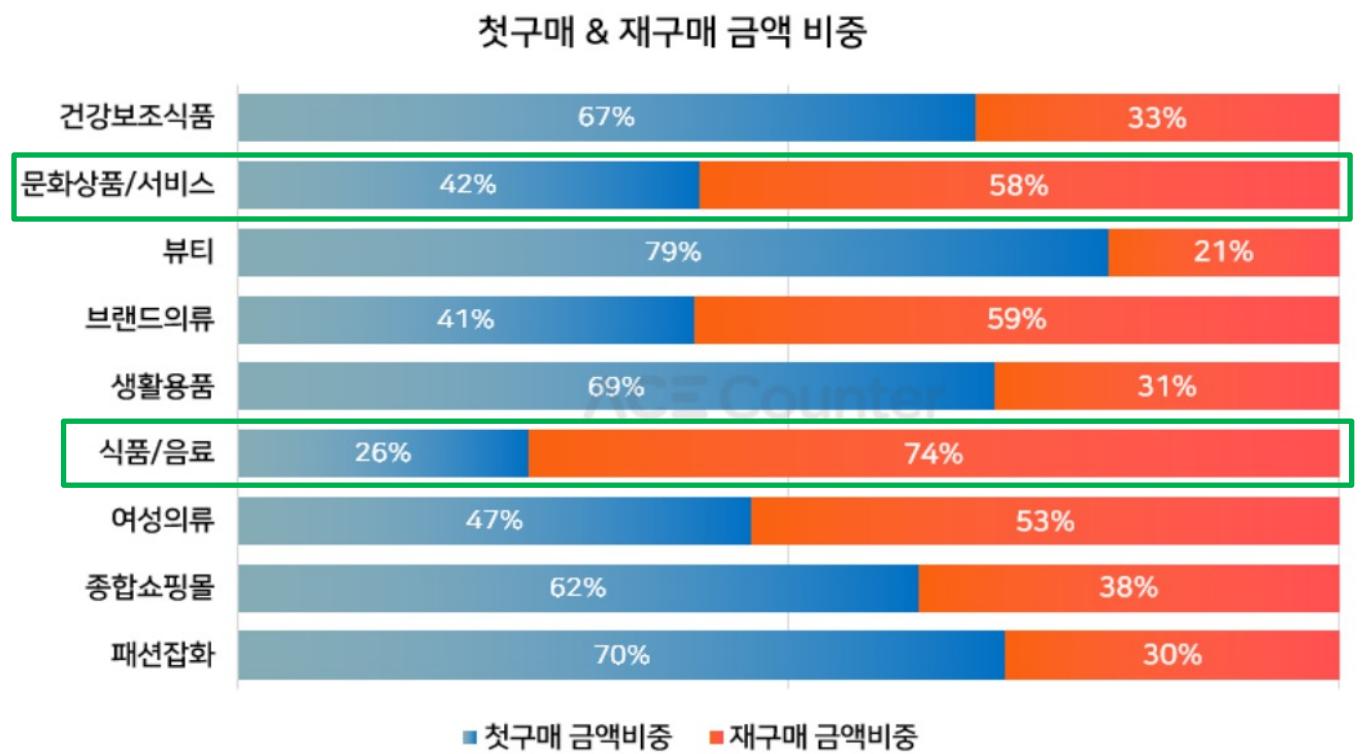
켈로그 허쉬 초코크런치 (500G) 8,580원 ★ 4.8 (130)	캐모마일 릴렉스 제로 (1.5L) 3,000원	종가집 대상 종가집 오래맛있는 열 무김치 (900G) 12,900원 ★ 4.7 (983)	초이스엘 한우 잡뼈 (1.5KG) 50% 11,000원 5,500원 ★ 4.8 (3,374)	서울우유 서울 저지방우유기획 (1,000ML*2) 4,280원 ★ 4.8 (14,837)
---	------------------------------	--	--	--

웹사이트 및 어플(온라인)로 상품 주문
→ 사용자 근처 점포(오프라인)에서
당일 받아볼 수 있는 상품 추천

인기상품이 아닌 고객이 구매하고자
하는 상품을 추천 (개인화 추천)

4.3 온라인 마케팅 제안

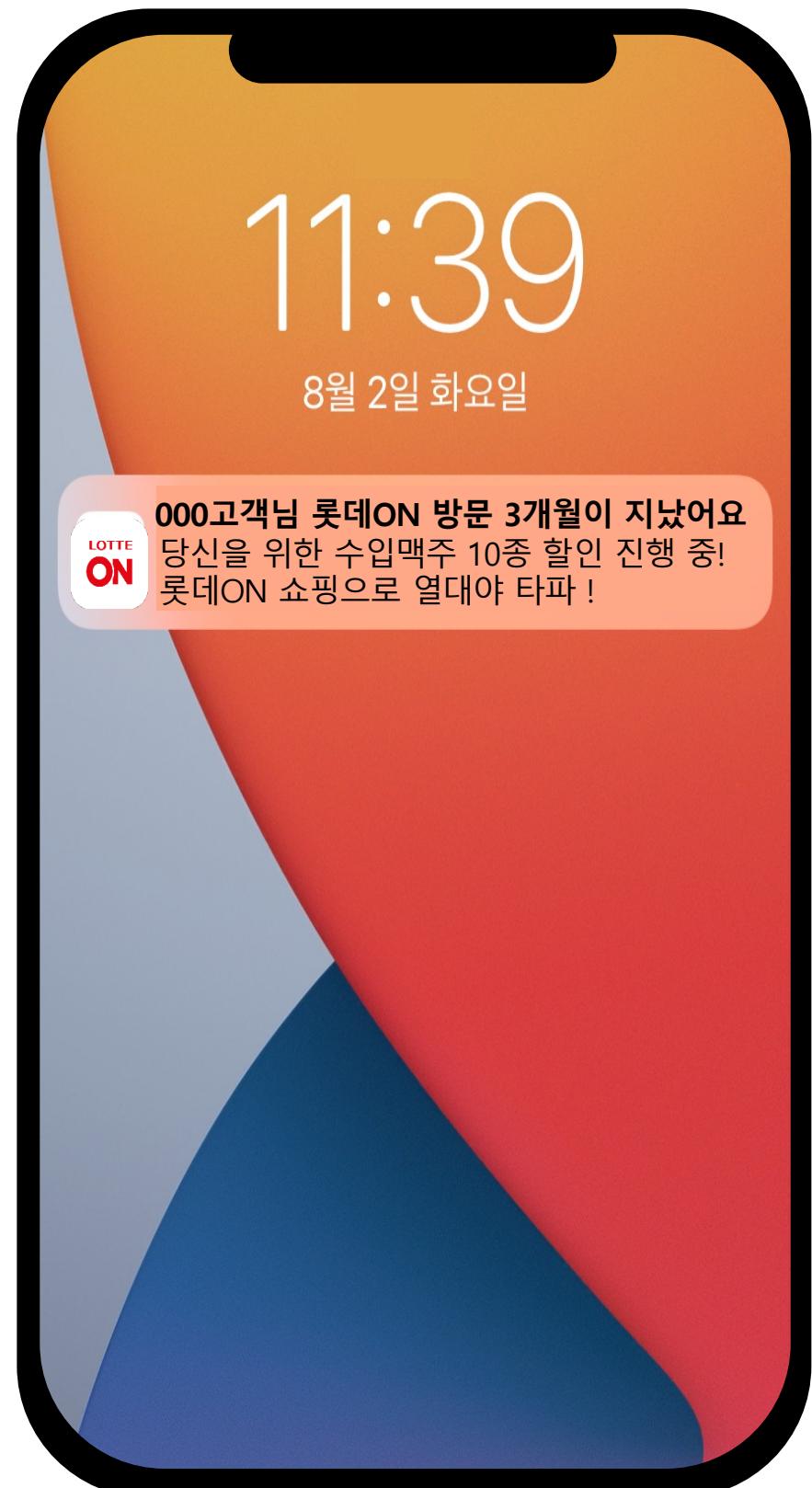
재구매 - 재구매 유도를 위한 마케팅 제안



신규 고객이 아닌 구매 이력이 있는 고객들의 재구매를 유도하는 것은 매우 중요 → 충성 고객으로 유도 가능

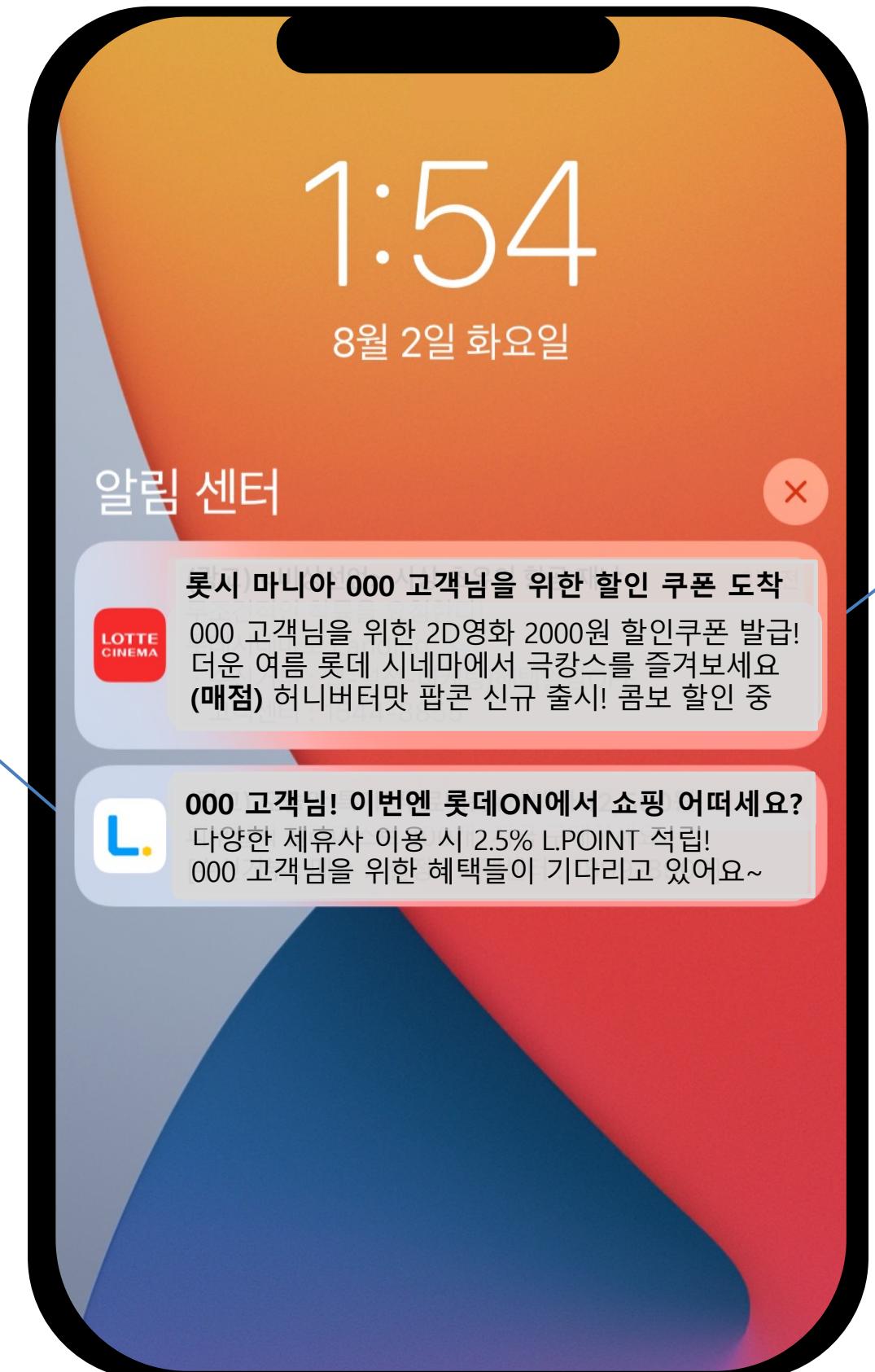
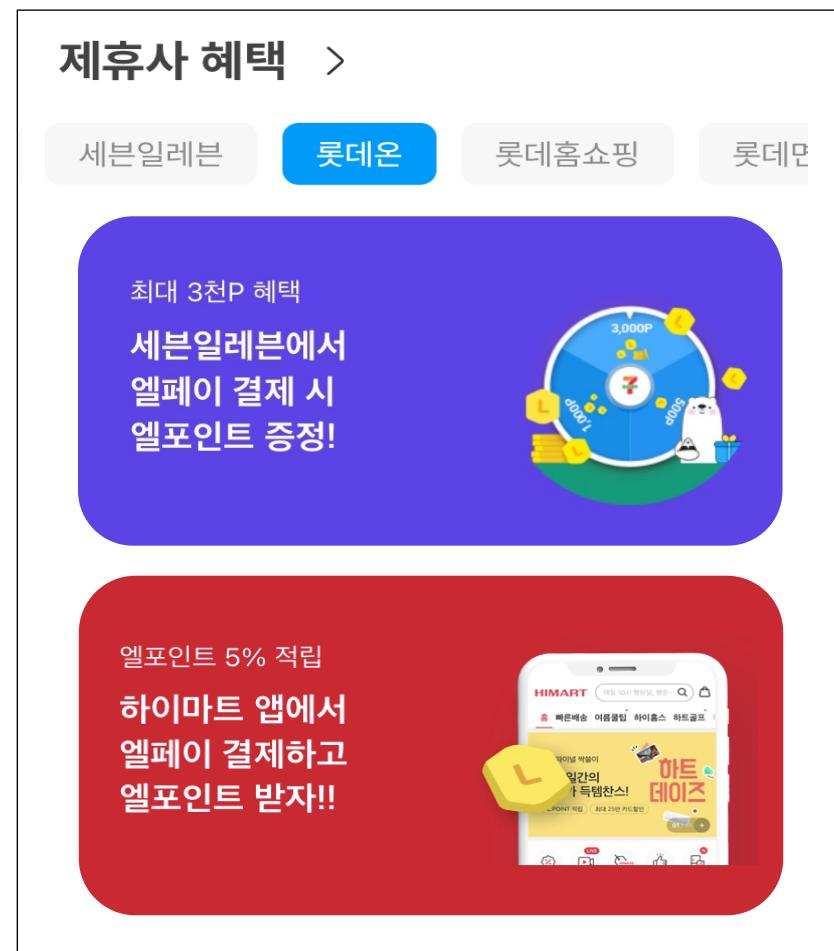
일정 기간이 지난 뒤 재구매를 하지 않을 시 새로운 상품 품목 및 고객 정보를 기반으로 고객에게 맞춤형 알림톡 발송 (발송 시 참고사항: 식품군의 경향은 5~14일 이내에 나타남)

마지막으로 구매한 상품정보와 업데이트 된 소비자 정보 기반으로 추천 시스템의 상품목록을 확대하여 알림으로 전송

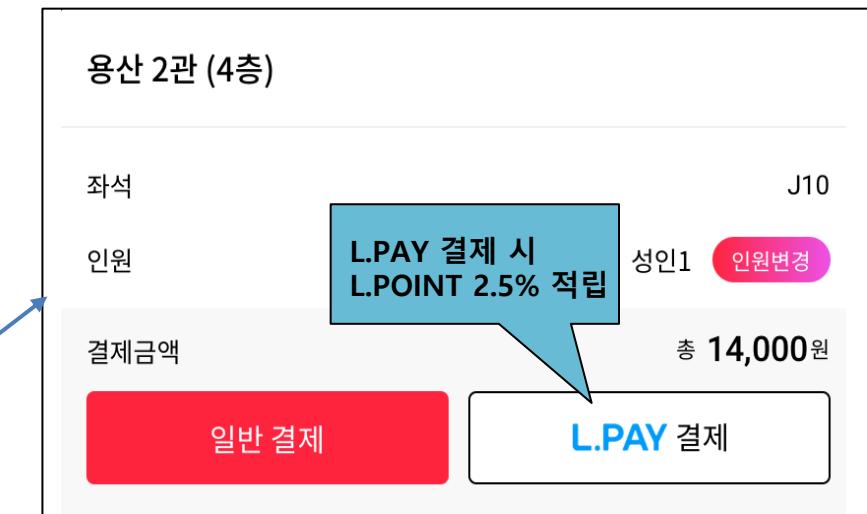


4.4 제휴사 활용 방안

과거 제휴사 이용 정보 기반 제휴사 추천



<예시- 롯데시네마를 주로 이용하는 고객 >



2. 특정 제휴사를 누적 이용 이력이 있는 고객
→ 자주 이용하는 제휴사 할인 쿠폰 제공하여
재이용 권유

이들이 **충성고객이 될 수 있게 유도**하는 것이 핵심

→ 문화 상품 및 서비스에 대한 재구매 비율이
높은 것을 감안
특히 오프라인 제휴사 고객들에게 재방문을
유도할 수 있는 **고객 경험을 선사**하는 것이 중요

1. 특정 제휴사에 대한 누적 이용 이력이 없는 신규 고객
→ 추천 시스템 기반 이용 경험이 없는
제휴사를 추천하여 **다양한 온·오프라인 경험
유도**

4.5 LPAY 활용 방안

IV. 마케팅
제안

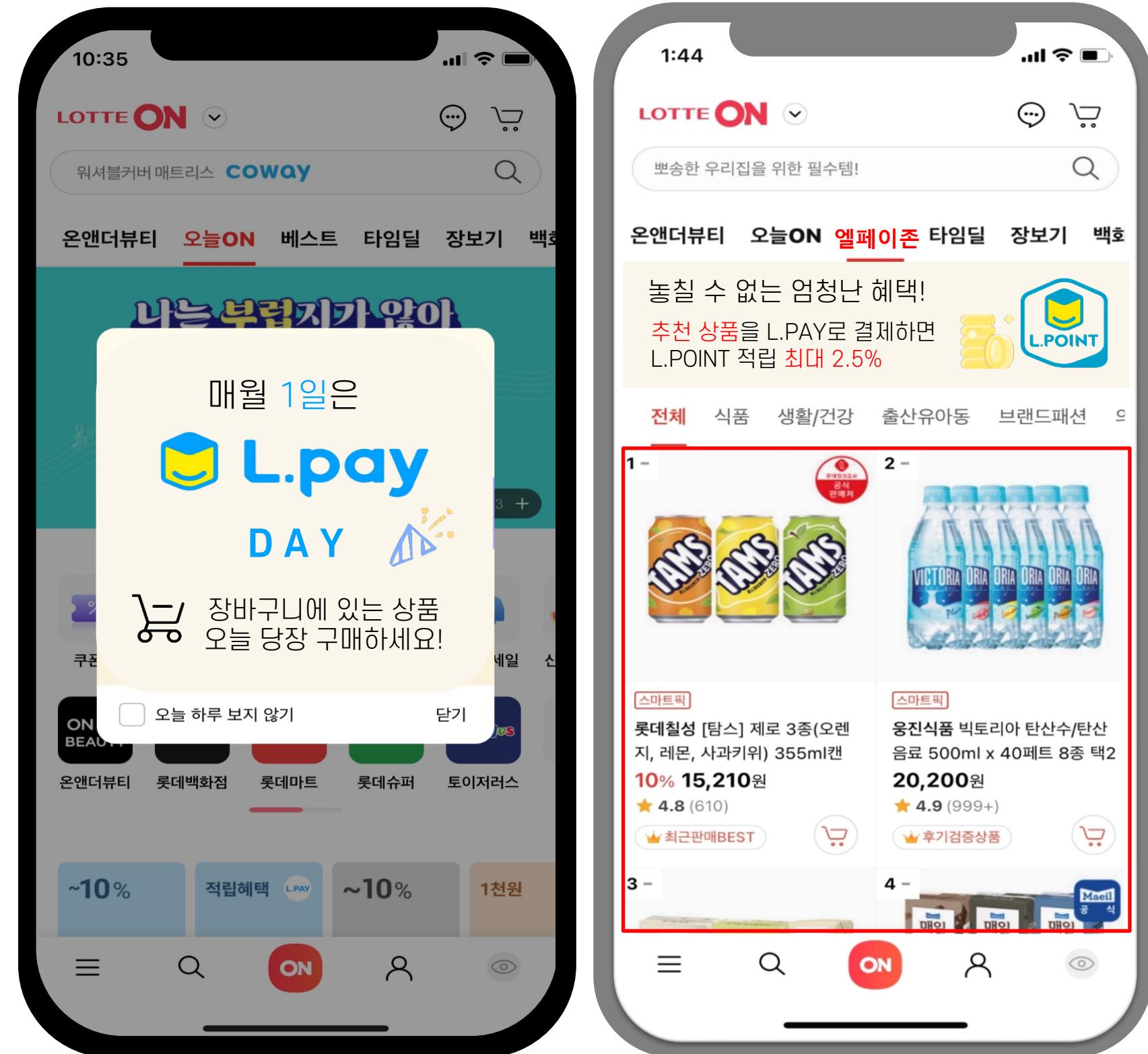
엘페이 사용 유도를 위한 마케팅 제안

1. 엘페이존

- 엘페이로 결제하면 혜택이 더 좋은 상품들을 모아 판매
- 개인화 추천 시스템 모델링 결과 활용해서 추천 상품을 띠움
- 추천 상품을 엘페이로 결제하면 일반 결제보다 혜택이 더 다양하다는 것을 홍보하기 위해 광고 배너 활용

2. 엘페이 데이

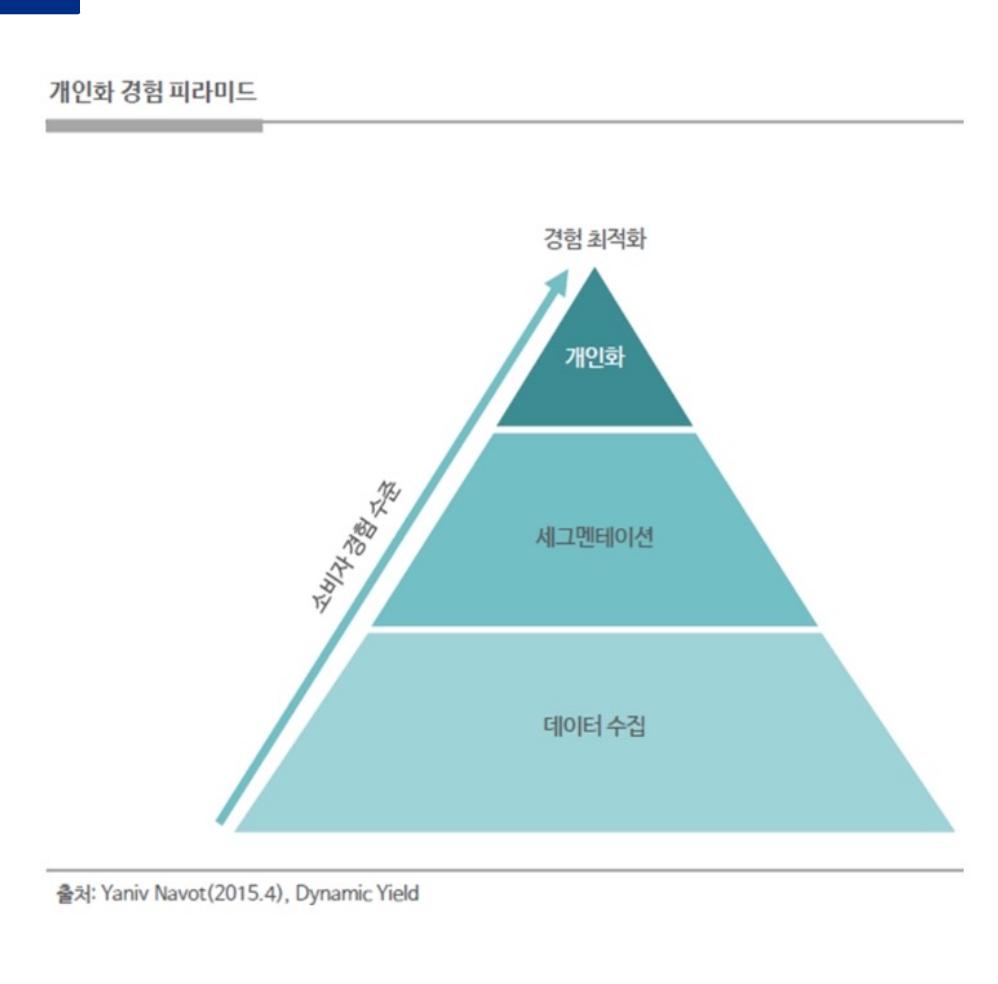
- 한달에 한번 엘페이 데이에 결제하면 다양한 혜택 제공
- 멤버십 등급별 당일에만 사용할 수 있는 특별 할인 쿠폰 제공
- 하루만 구매 금액에 관계 없이 무조건 무료배송



4.6 서비스 기대효과

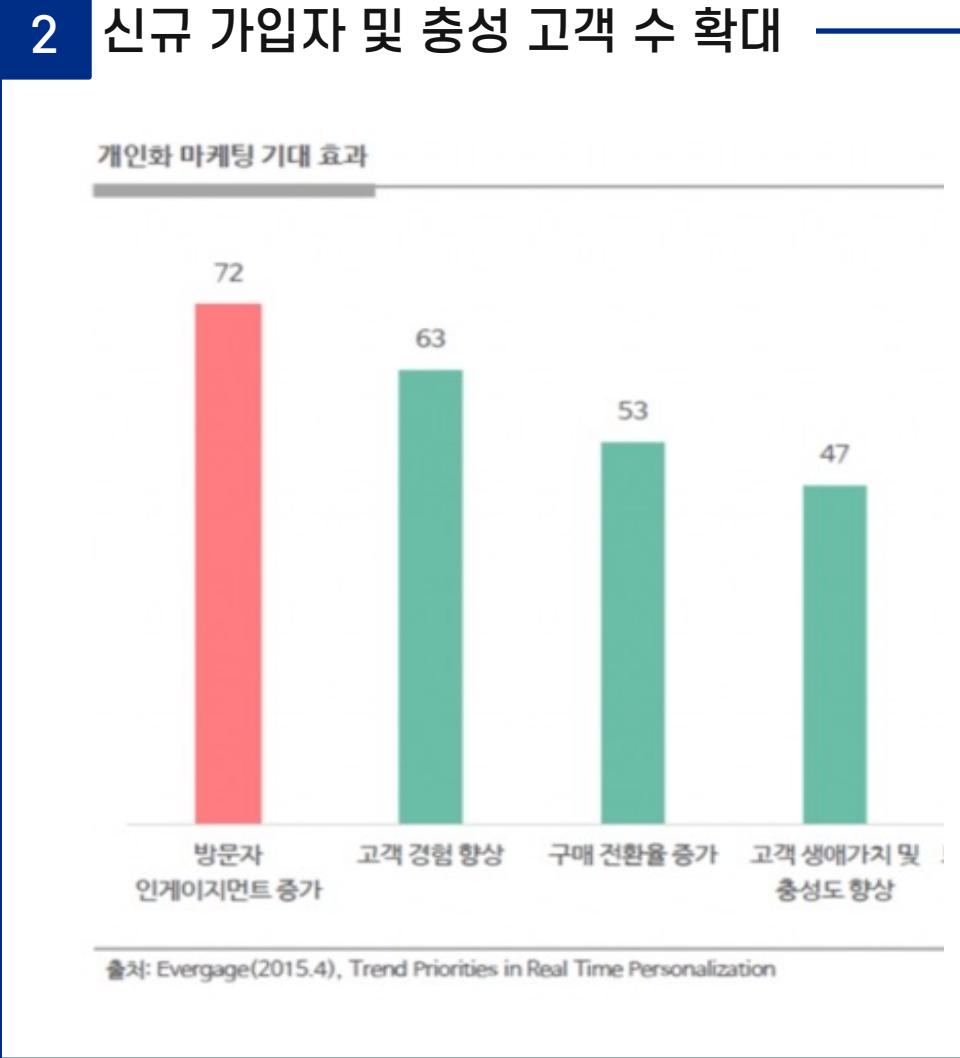
추천 시스템 기반 개인화 마케팅 기대 효과

1 고객 경험 및 서비스 이용 만족도 증가



개인화 마케팅을 통해 소비자의 경험 수준을 개인에 최적화되어 만족도를 증가시킴

2 신규 가입자 및 충성 고객 수 확대



소비자의 경험 수준을 끌어올려 충성고객 확보 및 신규 가입자에 대한 방문 유도

3 개인화 마케팅을 통한 매출 증대

17년 하반기 ~ 18년 상반기 기준



*Marketing Cloud: 통합 개인화 마케팅 플랫폼

Marketing Cloud 통한 매출:
17Q1/18Q1 대비 136% 상승

고객 경험 및 만족도 증대와 충성 고객 확보
→ 결과적으로 매출 증대 실현

V. 결론

- 결론
- 한계점 및 보완점
- 참고 자료

05

롯데멤버스
빅데이터 경진대회

5.1 결론

V. 결론

프로젝트 과정 요약 및 결론

EDA 요약

< 타겟층 및 주요 유통사 탐색 >

30-40 대 서울 경기 여성이 구매 큰 축을 담당 (충성고객) → 타겟층 설정

⇒ 타겟층의 주 이용 온라인 유통사: A06

A06 (롯데온) - 엘페이지 구매 횟수 1위, 온라인 매출 상승을 중점 → 온라인 위주 마케팅 방안 제안

모델

< 코사인 유사도와 예측 값을 통한 모델 개발 >

- 사용자 기반 협업 필터링을 활용한 상품 추천 모델

과정: 코사인 유사도 + 예측값 + RMSE 평가지표

외부데이터 : 상관 분석(히트맵) → 뉴스 심리지수& 트렌드 지수 활용

군집분석 : NearestNeighbors(최근접 이웃) → RMSE 값 개선

서비스 제안

< 고객 여정에 따른 개인화 마케팅 제안 >

1-1. 웹사이트 방문 유도 → 포털 사이트 광고 배너와 이메일로 개인 맞춤 상품 광고

1-2. 장바구니/ 구매 → 온·오프라인 통합 고객을 위한 맞춤 상품 추천하여 바로 배송으로 수령 가능

1-3. 재구매 유도 → 일정 기간 이후 구매이력이 없는 고객에게 개인 맞춤형 알림 톡 발송

2. 엘페이지 → 엘페이지 존과 엘페이지 데이: 엘페이지 결제 시 다양한 혜택 제공 (포인트 적립 및 쿠폰)

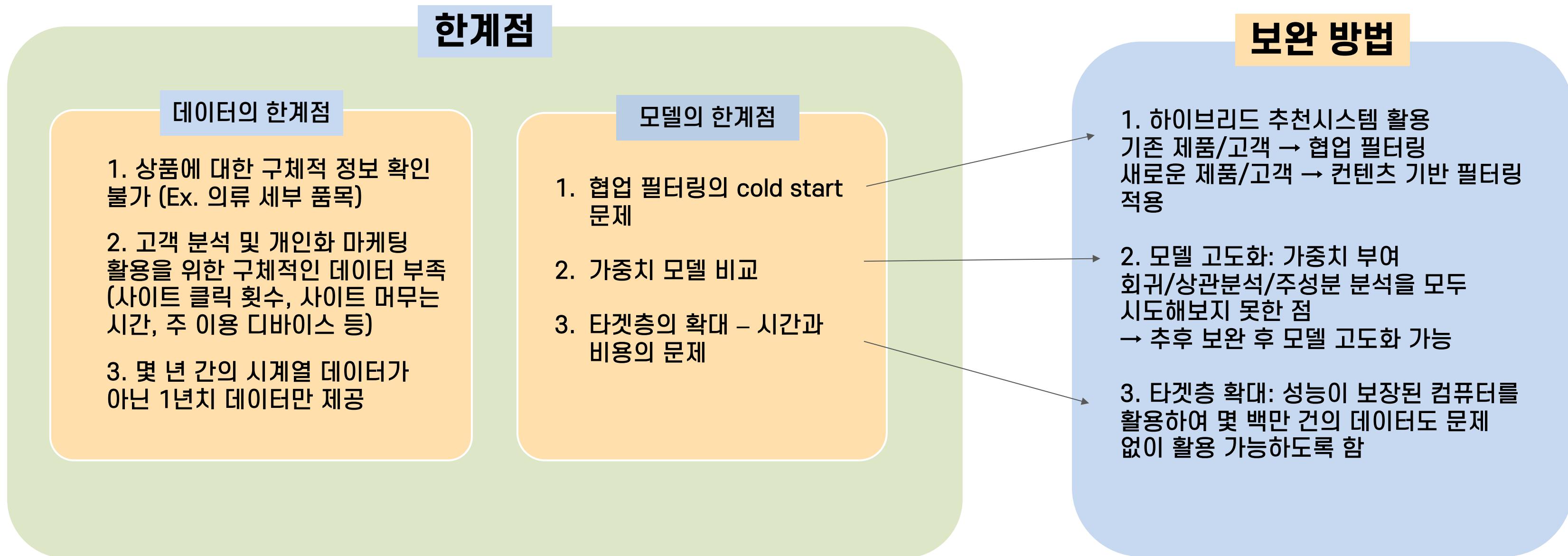
3. 제휴사 → 과거 제휴사 사용 이력을 바탕으로 새로운 제휴사 추천 또는 기존 제휴사 할인쿠폰 제공

“타겟층 확대”
제한적 데이터
(서울/경기 30-40대 여성)를
적용한 모델에서
→ 전체 고객을 대상으로
확대하여 모델의 확장성 실현

5.2 한계점 및 보완점

V. 결론

프로젝트 진행 과정 전반의 한계점 및 보완점



5.3 참고자료

V. 결론

<https://newsroom.koscom.co.kr/30518>

<https://platum.kr/archives/184517>

<https://www.codingworldnews.com/news/articleView.html?idxno=2477>

<https://brunch.co.kr/@biginsight/10>

<https://data-science-hi.tistory.com/150>

<https://www.youtube.com/watch?v=euflBOKwWAc>

<https://dailyheumsi.tistory.com/167>

<https://datascienceschool.net/02%20mathematics/07.05%20%EA%B3%B5%EB%B6%84%EC%82%B0%EA%B3%BC%20%EC%83%81%EA%B4%80%EA%B3%84%EC%88%98.html>

<https://brunch.co.kr/@dightly/20>

<http://boost.bizspring.co.kr/index.php/2017/08/07/case-29/>

<https://leebaro.tistory.com/entry/%EB%8D%B0%EC%9D%B4%ED%84%B0-%EB%B6%84%EC%84%9D-%EC%8B%9C-%EC%8B%9D%EC%97%90-%EB%A1%9C%EA%B7%B8%EB%A5%BC-%EC%B7%A8%ED%95%98%EB%8A%94-%EC%9D%B4%EC%9C%A0>

<https://www.clvs.co.kr/post/case-study-%EA%B0%9C%EC%9D%B8%ED%99%94-%EB%A7%88%EC%BC%80%ED%8C%85%EC%9D%84-%ED%86%B5%ED%95%9C-%EB%A7%A4%EC%B6%9C-136-%EC%83%81%EC%8A%B9%EC%9D%98-%EB%B9%84%EA%B2%BO>

<https://www.dailyimpact.co.kr/news/articleView.html?idxno=65408>

<논문>

개인화 상품 추천이 구매의도에 미치는 영향에 관한 연구 : Matrix Factorization 알고리즘을 중심으로

모바일 개인화 서비스에 대한 소비자 지각이 모바일 거래 만족도와 충성도에 미치는 영향 연구 -지각된 통제력의 조절효과를 중심으로

Comparison of Customer Satisfaction Indices Using Different Methods of Weight Calculation

Review and Analysis of Recommender Systems

<통계청 자료>

https://kostat.go.kr/portal/korea/kor_nw/1/2/3/index.board 2021년

<외부 데이터>

- 뉴스 심리지수

<https://ecos.bok.or.kr/#/SearchStat>

- 트렌드 지수

<https://datatrend.kakao.com/>

감사합니다!

잘 부탁드립니다!

LP PLAYER

김민균, 조수현, 한유림

롯데멤버스 빅데이터 경진대회

