

개인화 상품 추천, 효율적 마케팅의 시작

자체 선호도 점수 기반 상품 추천 모델 & 마케팅 활용 전략

Team 비타플

1. 프로젝트 개요
2. EDA & 전처리
3. 모델링
4. 활용방안

[부록]



1. 프로젝트 개요

■ 문제 정의

무분별한 광고, 피로한 소비자

직장인 10명 중 7명이 소셜네트워킹서비스(SNS)에 피로감을 느낀다는 조사결과가 나왔다.

평생교육 전문기업 휴넷은 직장인 821명을 대상으로 설문조사를 한 결과 응답자의 69.4%가 SNS에 대해 피로감을 느끼고 있다고 답했다고 30일 밝혔다.

피로감을 느끼는 이유는 '원하지 않는 내용이 많이 보여서'(27.7%)와 '광고, 마케팅이 너무 잦아서'(26.1%)라는 이유가 많았다.

이어 '업무와 사생활의 경계가 모호해져서'(15.8%), '시간을 많이 소비하게 돼서'(14.6%), '사생활이 노출돼서'(8.6%), '상사 및 동료와 연결되는 것이 싫어서'(7.2%) 등이 뒤따랐다.

출처 : 김은경, " "원치 않는 내용·광고 너무 많다"... 직장인 70%, SNS에 피로감 ", 서울연합뉴스, 2017년 08월 30일자

<기준>

- 일괄적인 광고로 낮은 효율성
- 무분별한 광고로 소비자의 피로도 증가



<개선 방향>

- 데이터를 활용한 개인화 마케팅을 통해 효율성 증가 및 기업과 상품에 대한 선호도 증가

■ 데이터 활용 범위 & 목표 구체화

“고객 구매 데이터에 기반한 예측 모델 개발 및 개인화 마케팅 전략 제안”

예상 개인화 마케팅 방법

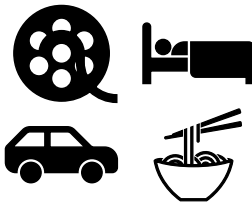
목표 달성 방식



제공 데이터



유통사



제휴사

- ✓ 고객별 상품 구매 성향 파악 및 미래 구매 예측 모델 개발
- ✓ 고객별 차별화된 상품 추천을 통한 마케팅

- ✓ 각 고객의 추가, 지속적 소비 가능성 활용
- ✓ 고객별 혜택 차별화를 통한 마케팅

예측 모델 활용한 추천 시스템 개발



데이터를 활용해
① 유의미한 모델 개발
② 개인화 마케팅을 위하여
유통사로 한정하여 진행

기존의 간단한 VIP 시스템 존재
(예측 모델 필요성X)

■ 데이터 활용 범위 & 목표 구체화

프로젝트 최종 주제 : 사용자-상품 간 선호도(구매가능성) 점수예측 모델 개발 및 추천시스템을 통한 개인화 마케팅

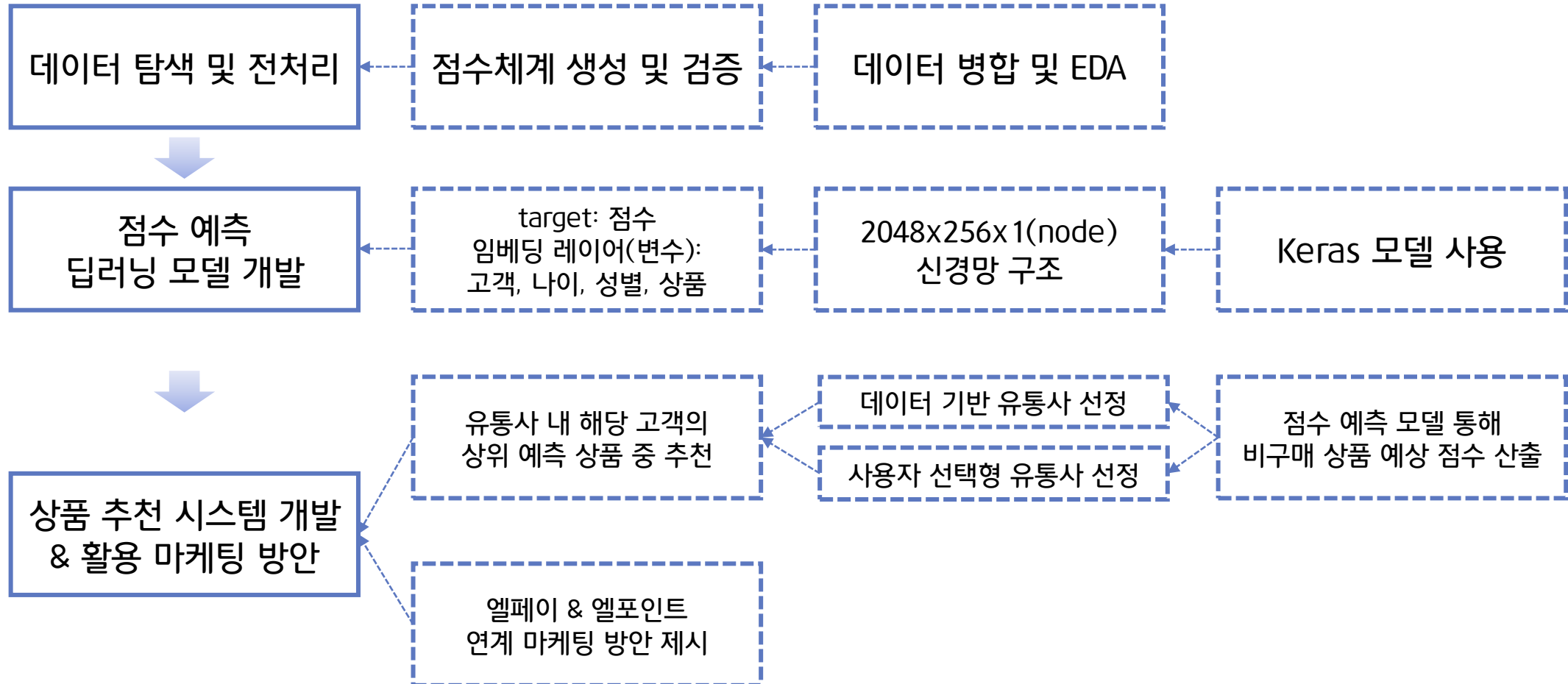
<분석 목표>

1. 데이터 활용해 자체적으로 고안한 선호도(구매 가능성) 점수 산출
2. 점수 예측 딥러닝 모델 개발
3. 예측 모델을 이용한 상품 추천 시스템 개발

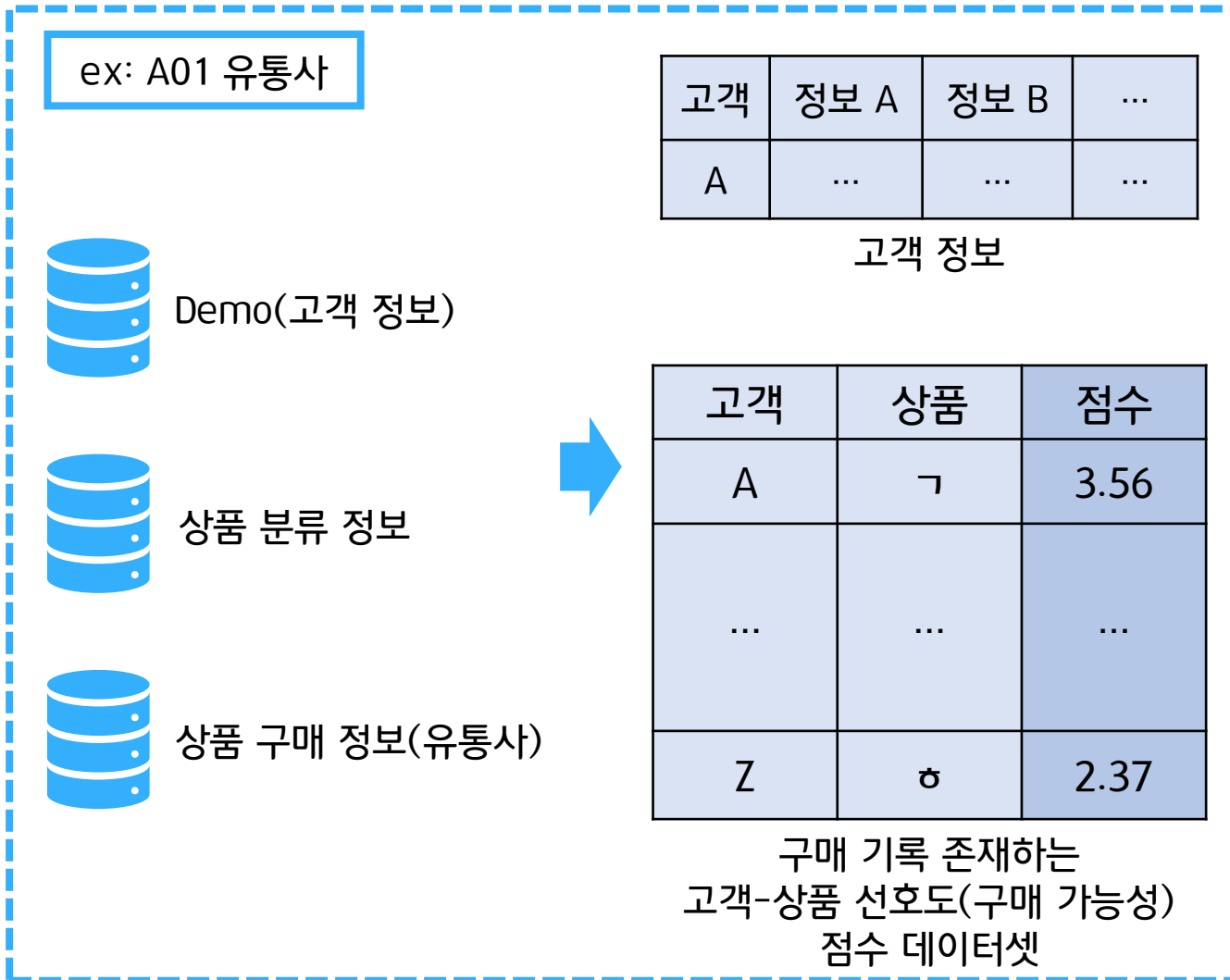
<마케팅 목표>

1. 추천시스템 기반, 고객 구매 기록 없는 상품 추천. 구매 범위 확장 유도
2. 개인화 상품 추천 with Lpay & Lpoint 마케팅전략

데이터 분석 프레임워크



데이터 활용 계획 (예측 모델 개발)



<구매 기록 X>

고객	상품	고객정보A	고객정보B	점수
A	ㄴ	NA



점수 예측
딥러닝 모델 적합



<예측 점수 산출>

고객	상품	고객정보A	고객정보B	점수
A	ㄴ	4.12

■ 점수 예측 모델 활용한 추천시스템

ex: A01 유통사



고객	상품	점수
A	0042	4.32
A	0139	4.11
A	0003	3.94
A	0837	3.94
...

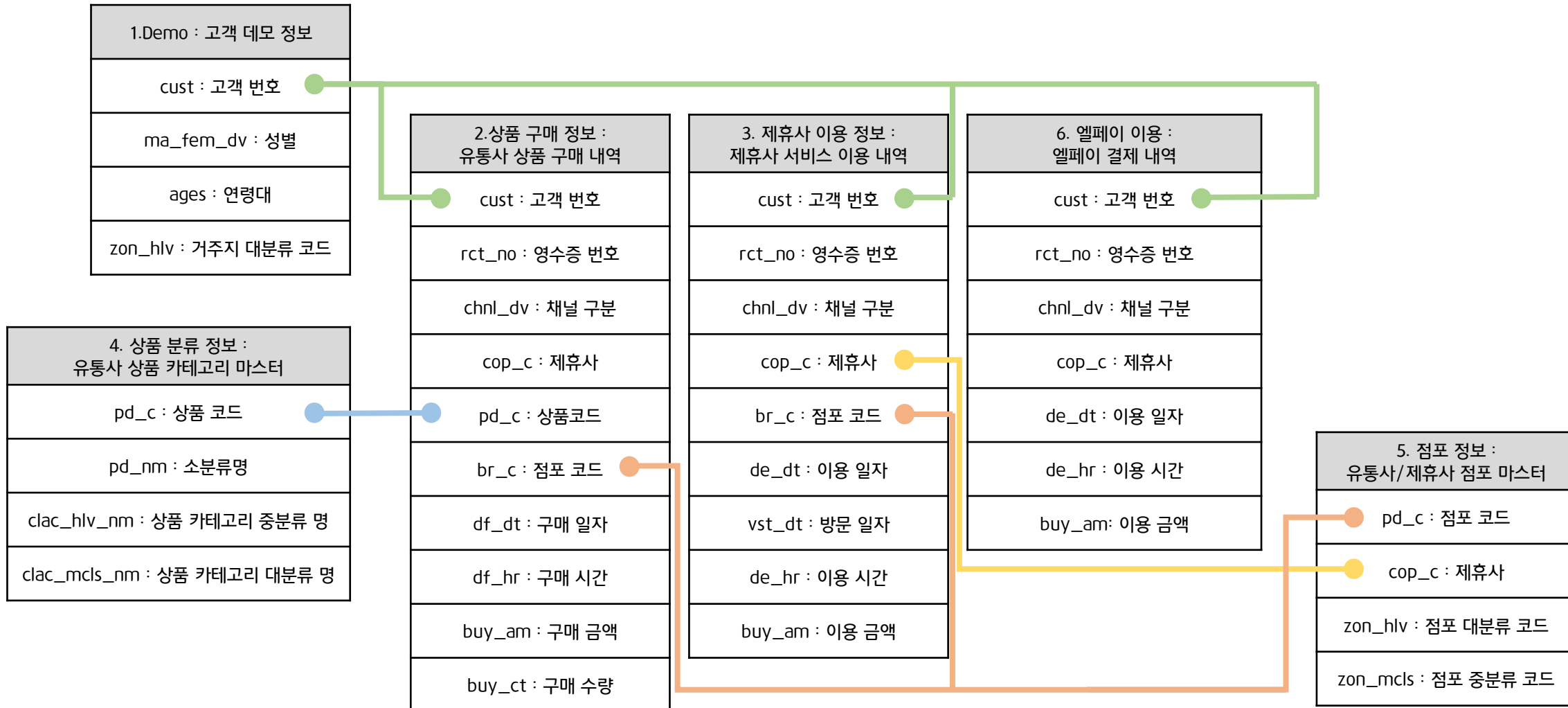


상품 추천

점수 예측 모델 이용해 상위 점수 상품 추천

2. EDA & 전처리

데이터 정보 및 관계도



유통사 별 데이터 분리

고객번호	...	유통사	...	Lpay 사용 여부
B	...	A01	...	0
...
...	...	A02
...
...	...	A03
...
...	...	A04
...
...	...	A05
...
C	...	A06	...	0



고객번호	...	유통사	...	Lpay 사용 여부
C	...	A01	...	0
...
...	...	A01

고객번호	...	유통사	...	Lpay 사용 여부
B	...	A02	...	0
...
...	...	A02

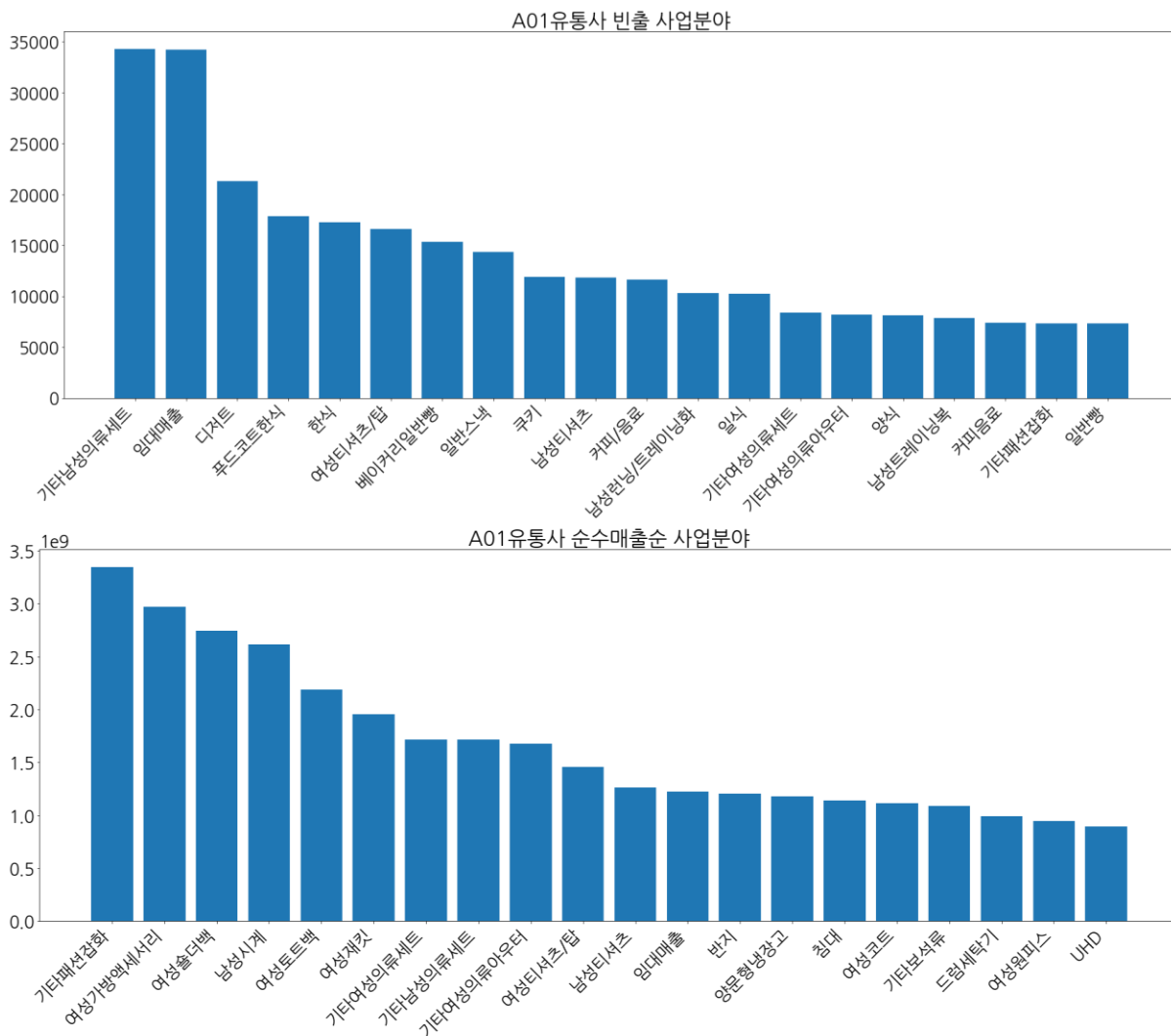
고객번호	...	유통사	...	Lpay 사용 여부
C	...	A03	...	0
...
...	...	A03

고객번호	...	유통사	...	Lpay 사용 여부
D	...	A04	...	0
...
...	...	A04

고객번호	...	유통사	...	Lpay 사용 여부
E	...	A05	...	0
...
...	...	A05

고객번호	...	유통사	...	Lpay 사용 여부
F	...	A06	...	0
...
...	...	A06

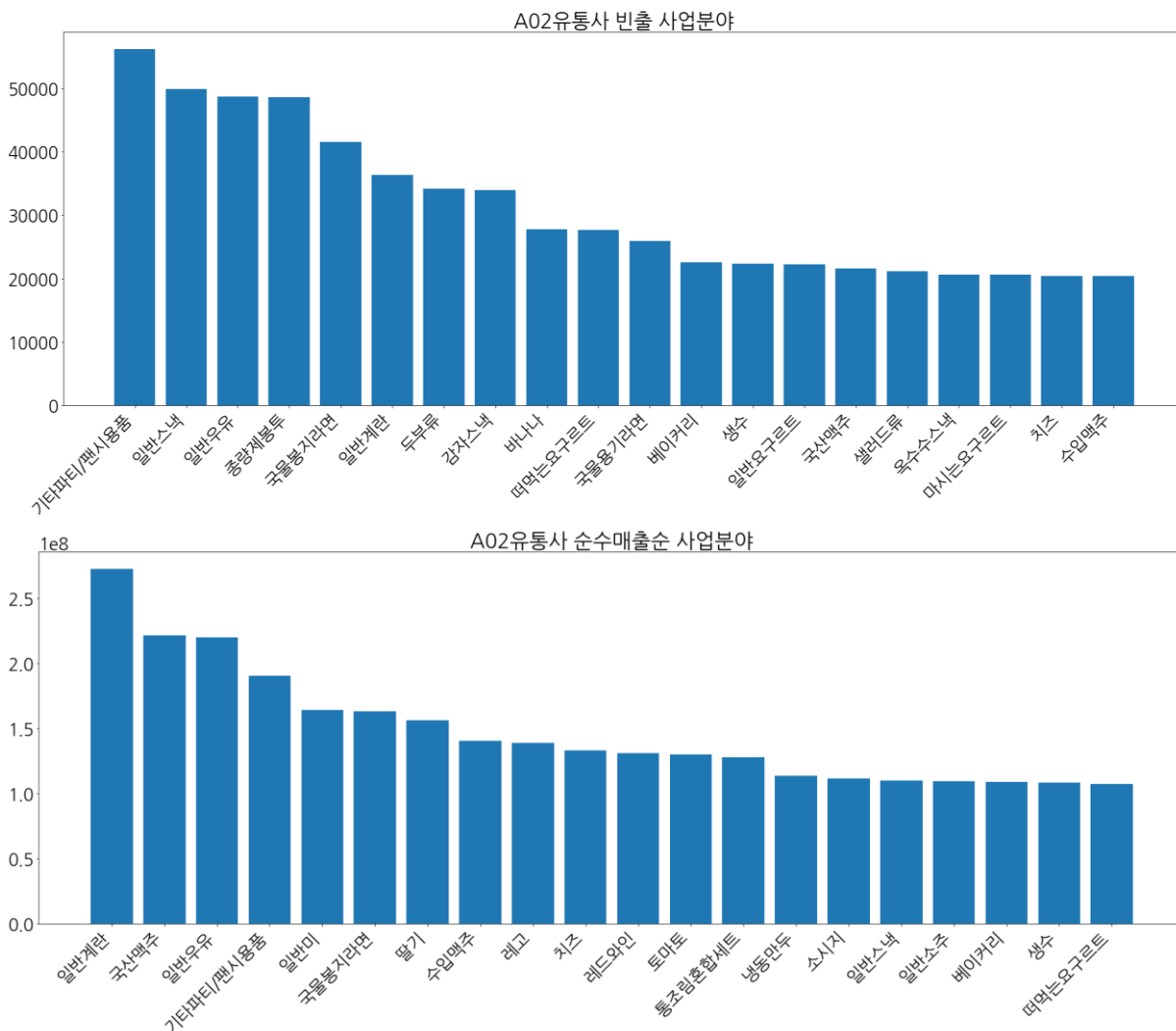
유통사 매출 구조를 통한 유통사 테마 추측 – A01



A01 : 오프라인 종합쇼핑몰(ex: 롯데백화점)

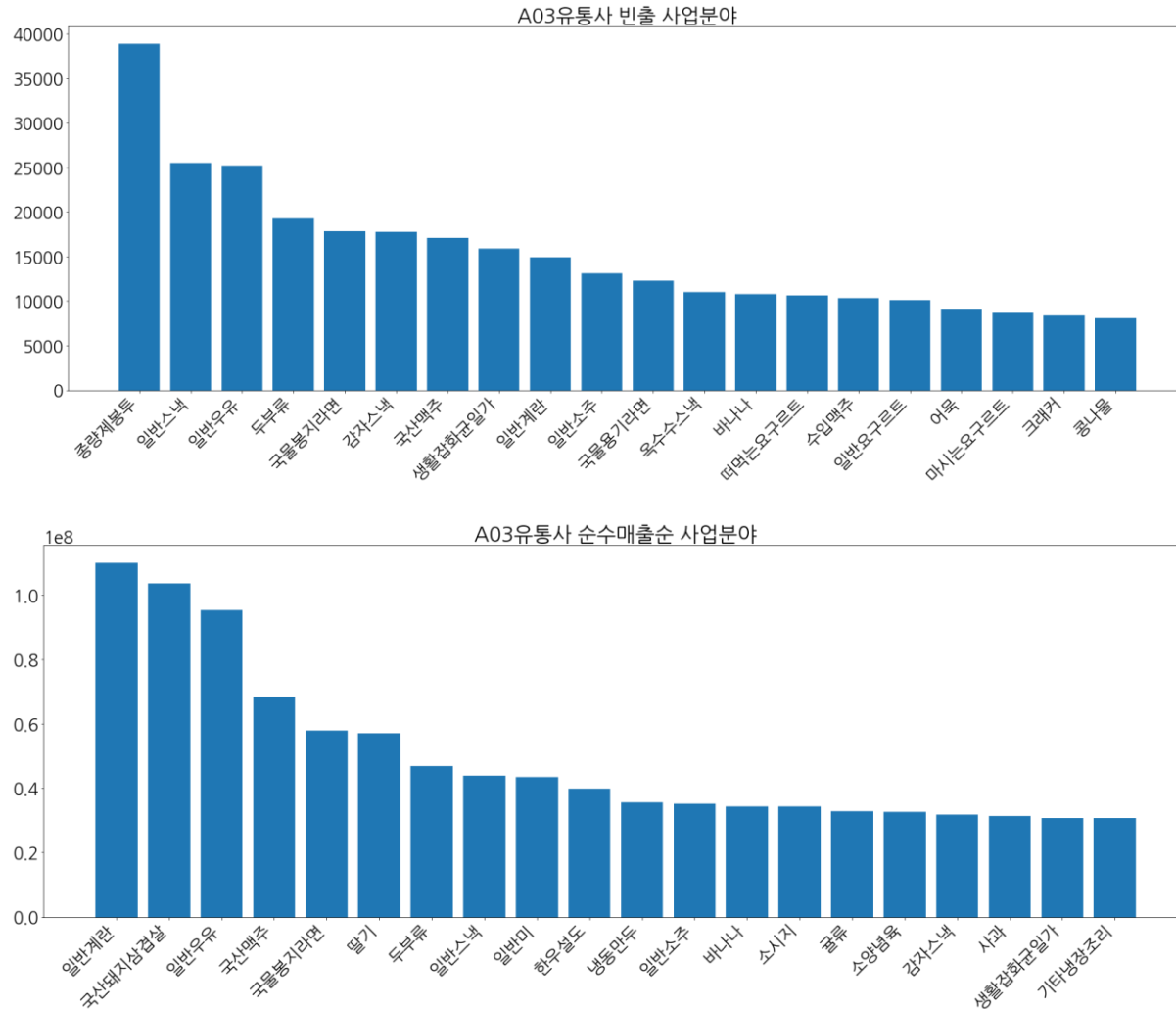
- 임대매출, 푸드코트, 의류가 상위 사업 분야임
- 점포코드는 종합쇼핑몰에 내점한 세부 점포들로 추측

유통사 매출 구조를 통한 유통사 테마 추측 – A02



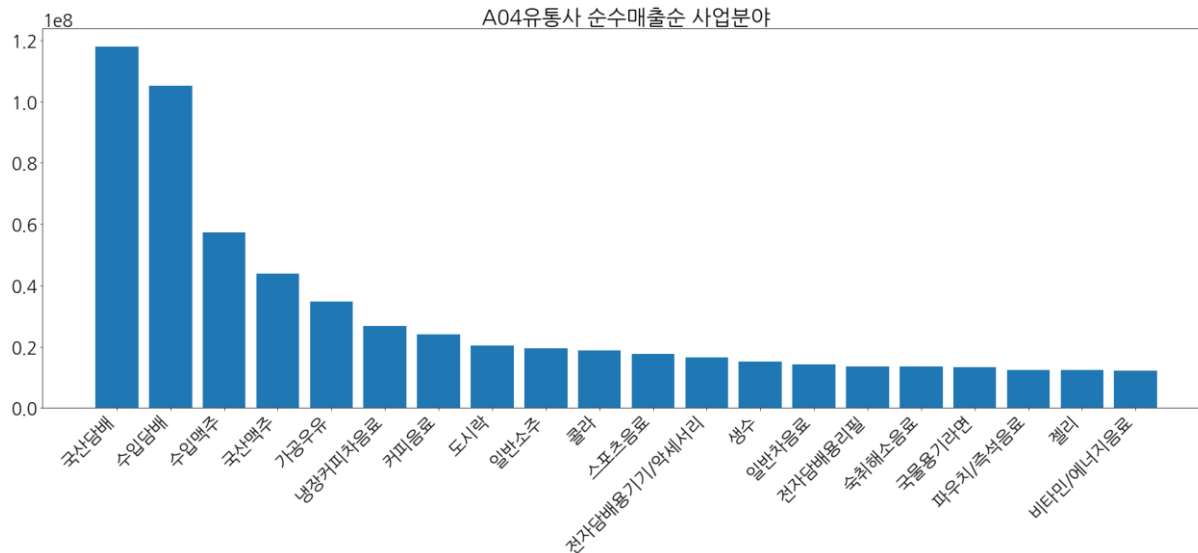
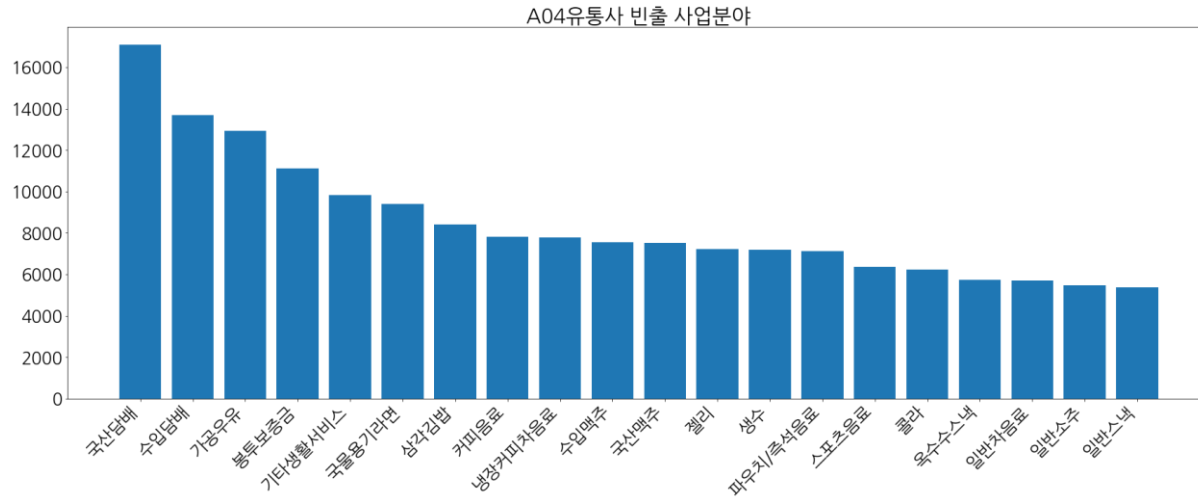
A02 : 대형마트 계열의 유통사 (ex: 롯데마트)
- 생활용품 및 팬시 용품이 상위 사업 분야임

유통사 매출 구조를 통한 유통사 테마 추측 – A03



A03 : 소규모 슈퍼마켓 계열 유통사 (ex: 롯데슈퍼)
 - A02와 흡사한 경향을 띄지만 일반미, 팬시&파티용품이 목록에 없고, 종량제 봉투가 상위권에 있음

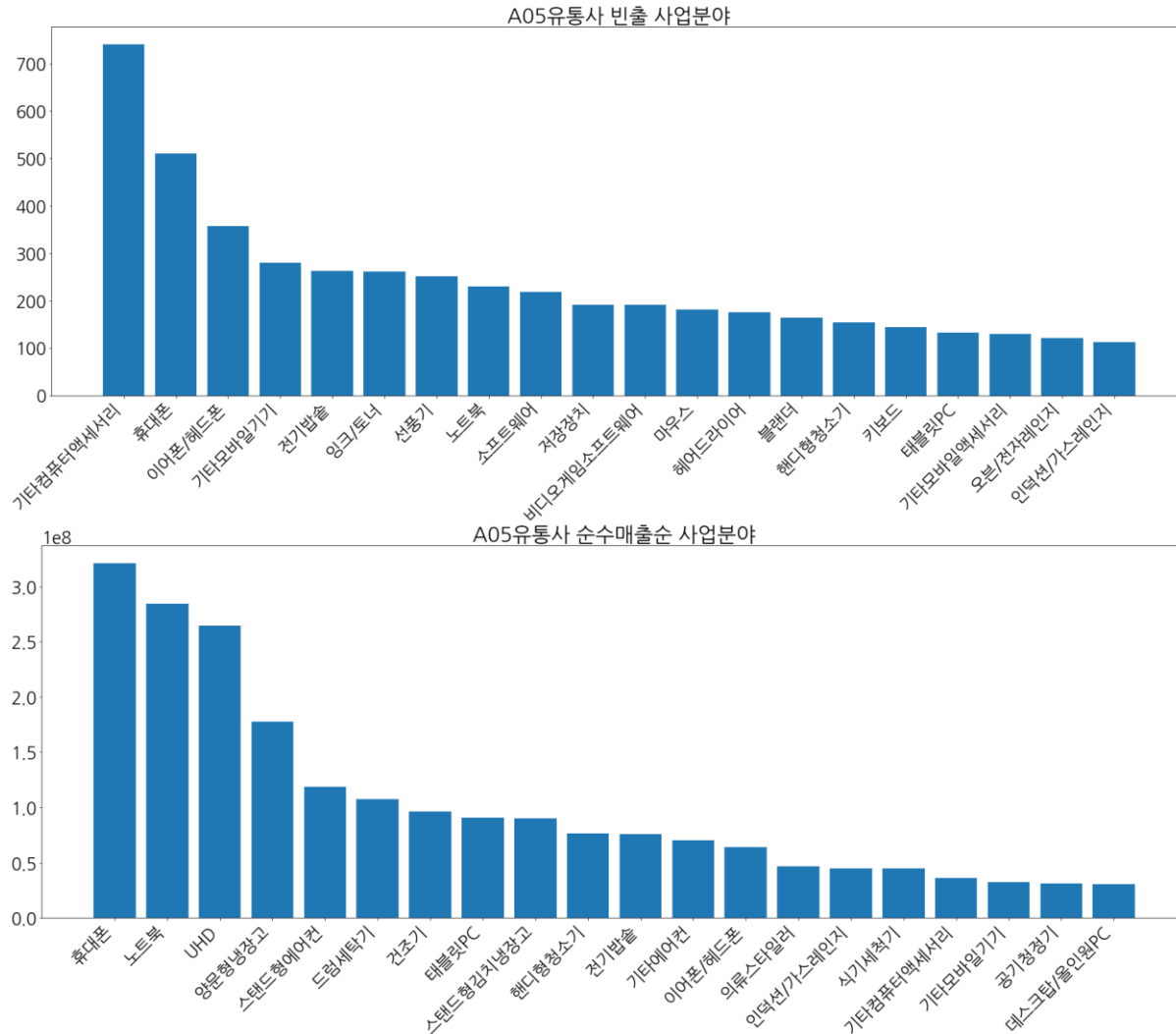
유통사 매출 구조를 통한 유통사 테마 추측 – A04



A04 : 편의점 (ex: 세븐일레븐)

- 현장결제만 존재
- 담배, 맥주, 삼각김밥 등이 상위 상품임

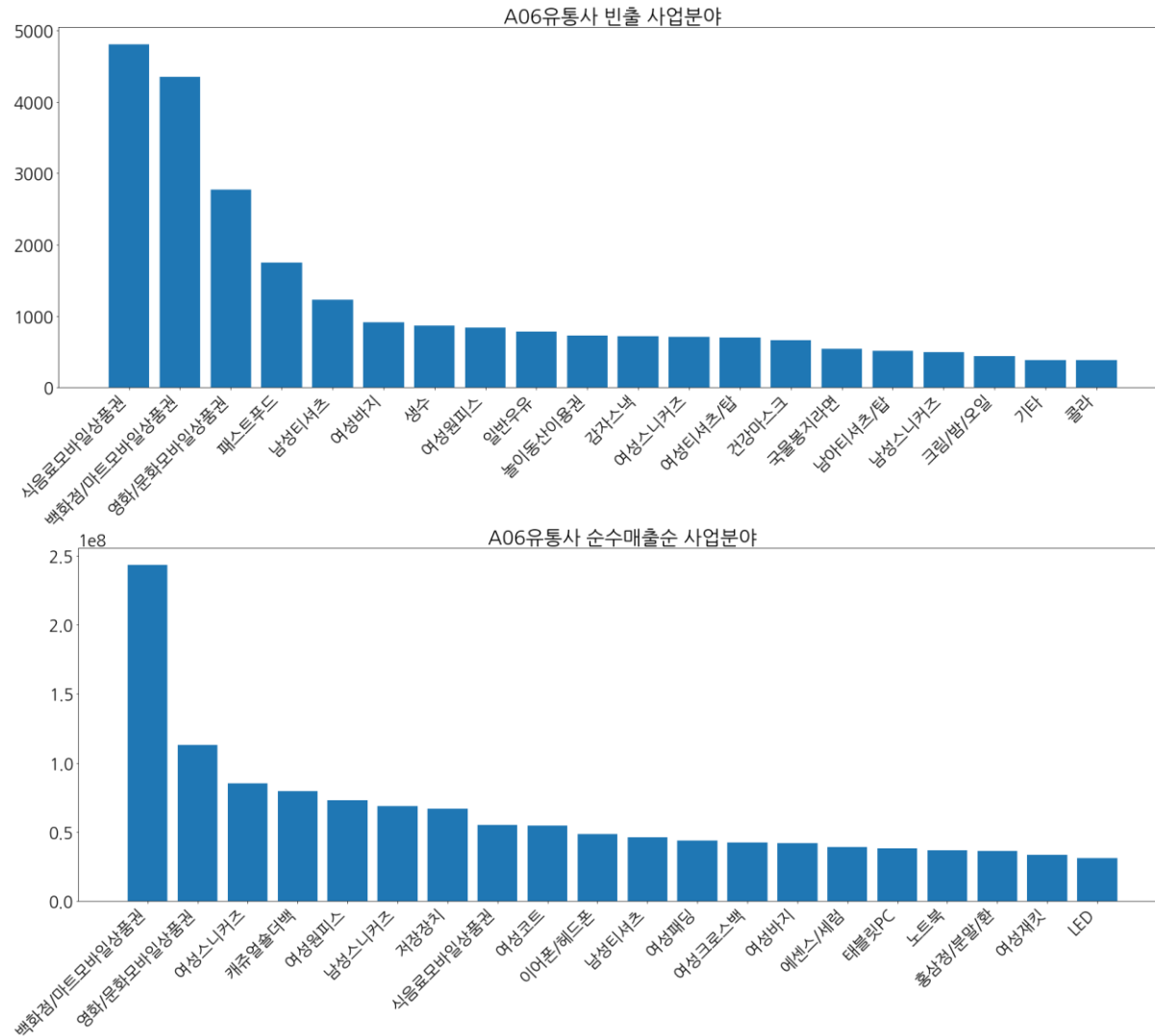
유통사 매출 구조를 통한 유통사 테마 추측 – A05



A05 : 가전제품 종합 쇼핑몰(ex: 하이마트)

- 품목이 모두 전자제품임

유통사 매출 구조를 통한 유통사 테마 추측 – A06



A06 : 온라인 쇼핑몰 (ex: 롯데ON)
- 모두 온라인 결제



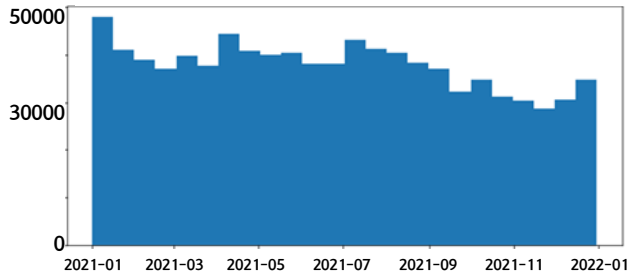
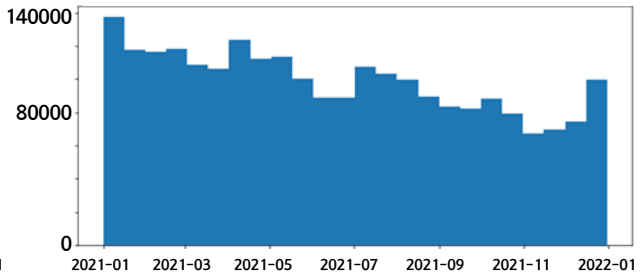
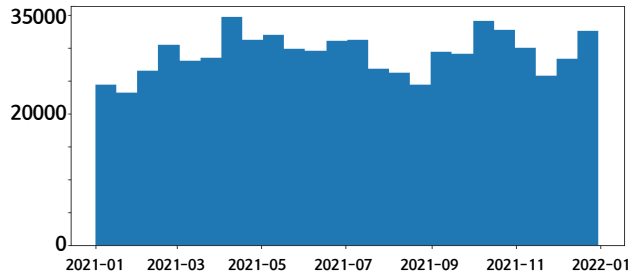
결제 경향 |

A01(백화점)

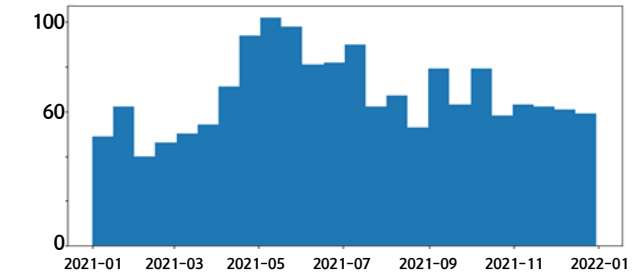
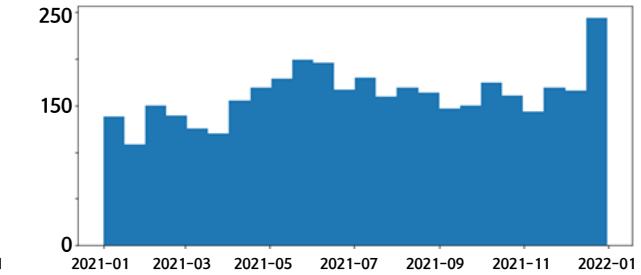
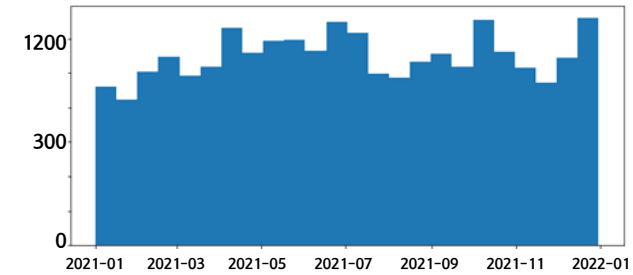
A02(대형마트)

A03(슈퍼마켓)

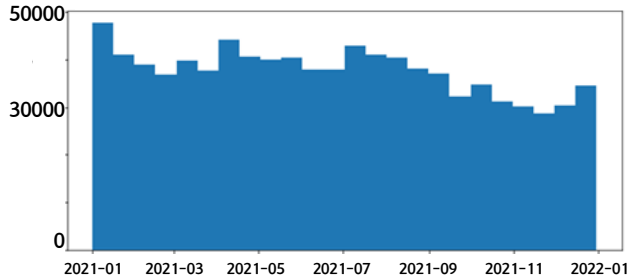
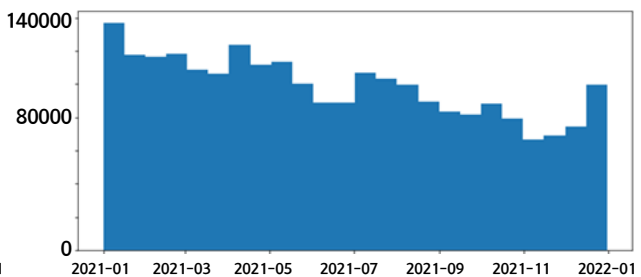
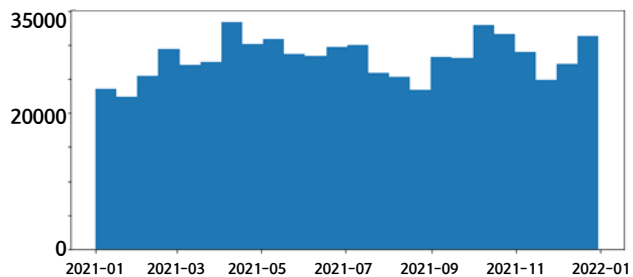
전체
결제



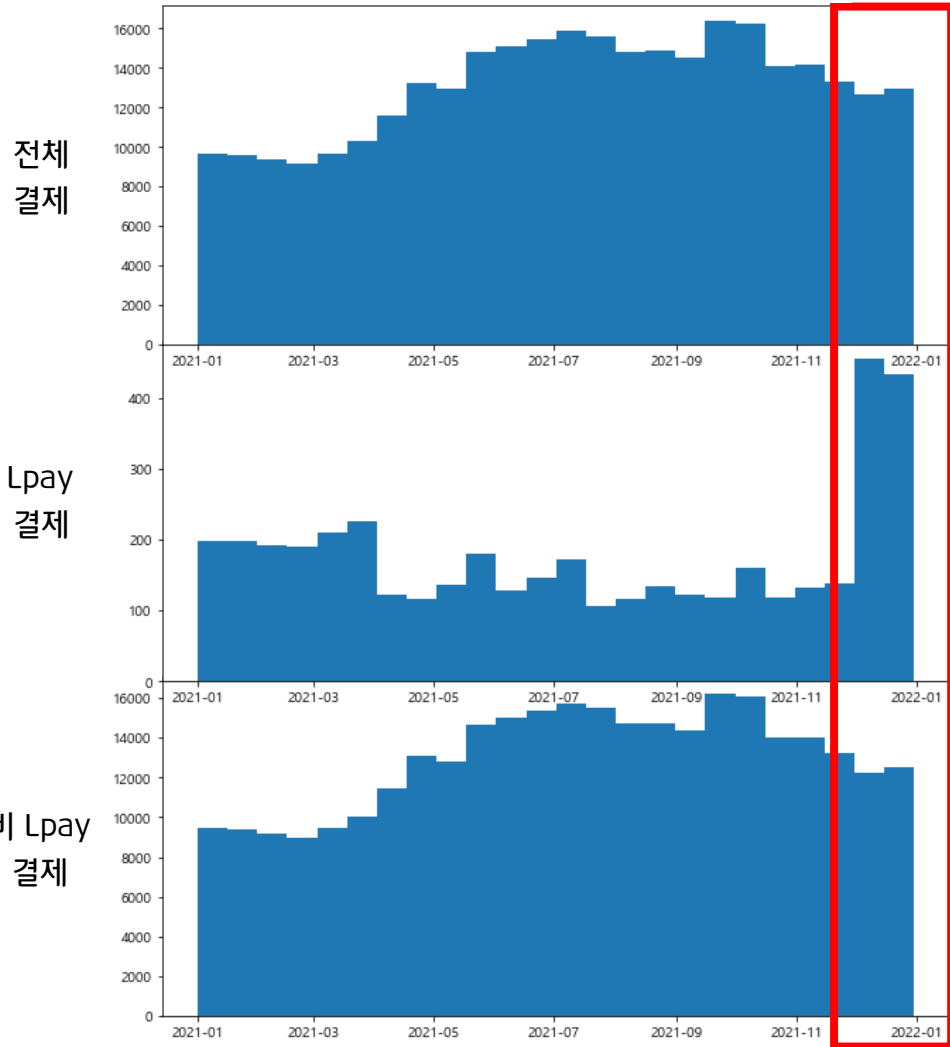
Lpay
결제



비 Lpay
결제



결제 경향 | A04(편의점)



세븐일레븐 12월 LPAY프로모션 존재

12월 lpay로 세븐일레븐 10000원 이상 결제 시, 5000p 또는 10000p 지급

*롯데멤버스 계열사이고, 12월에 위와 같은 프로모션이 있었기 때문에 A04 유통사는 세븐일레븐으로 추측

100% 당첨 룰렛 이벤트

L.PAY결제 시

L.POINT 100%증정

L.PAY로 1만원 이상 결제하면 룰렛 게임을 통해 L.POINT를 증정해드립니다.

12.1 ~ 12.31

L.POINT 받는 법

L.PAY로 1만원 이상 결제하면 룰렛 게임을 통해 L.POINT를 증정해드립니다.

* 1인 1회 한정

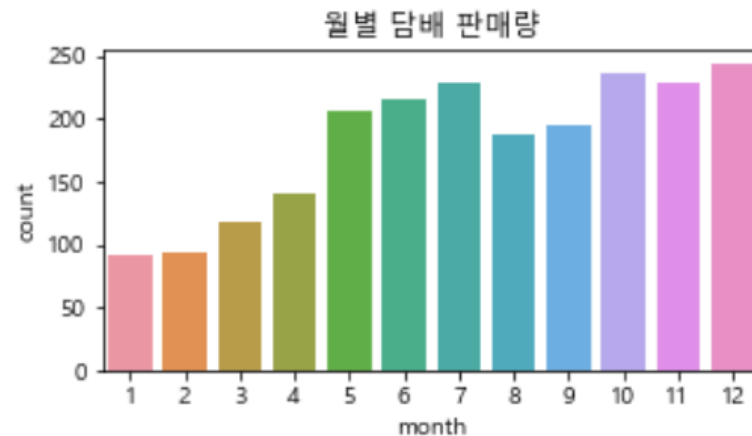
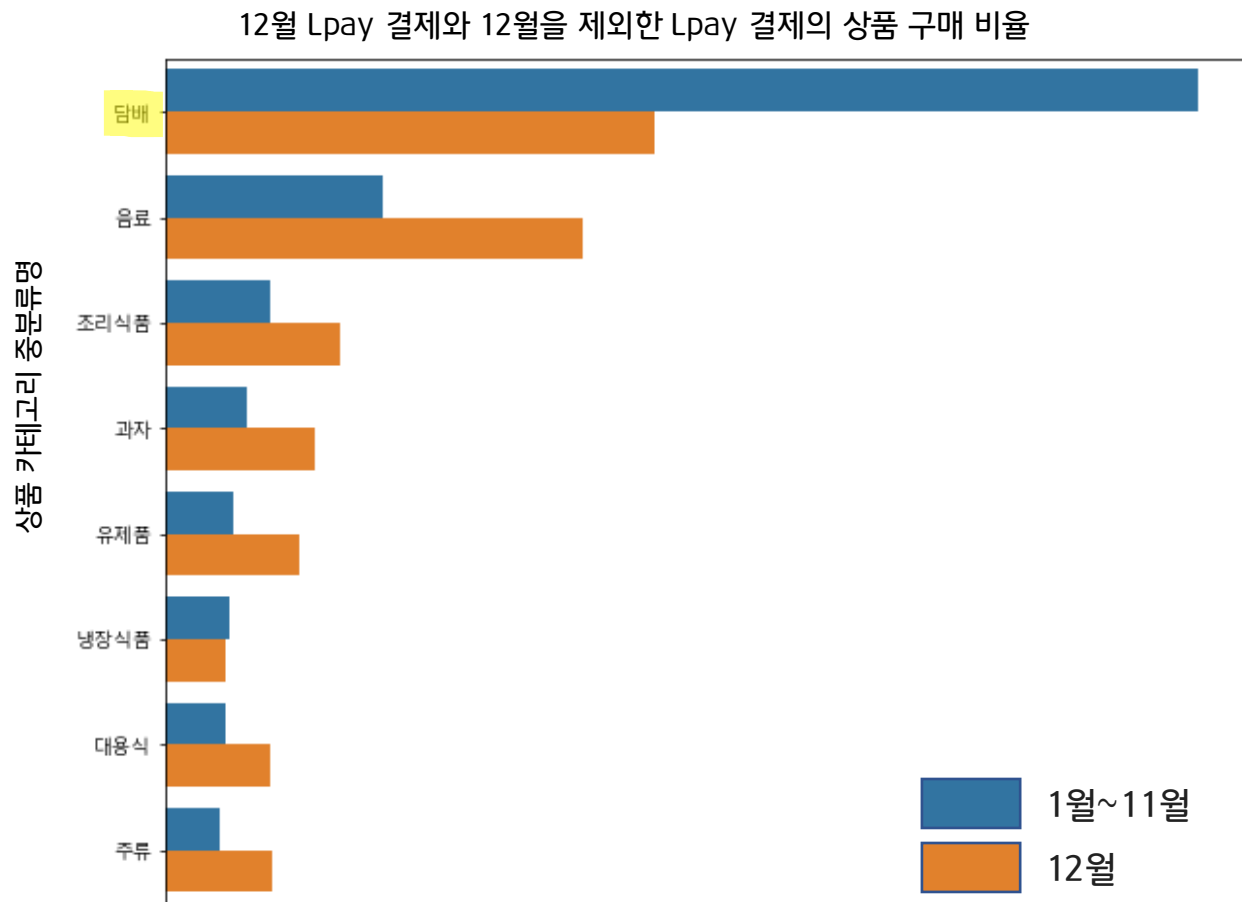
100% 당첨 룰렛 게임을 통해 L.POINT를 드려요

<p>5000P</p> <p>10,000명</p>	<p>1000P</p> <p>10,000명</p>
-----------------------------	-----------------------------

* 행사 참여 후 원거래에 대한 반품 또는 취소 시 룰렛을 통해 지급된 포인트는 강제 회수 될 수 있습니다.

<2021.12 진행한 세븐일레븐 LPAY 프로모션>

결제 경향 | A04(편의점)

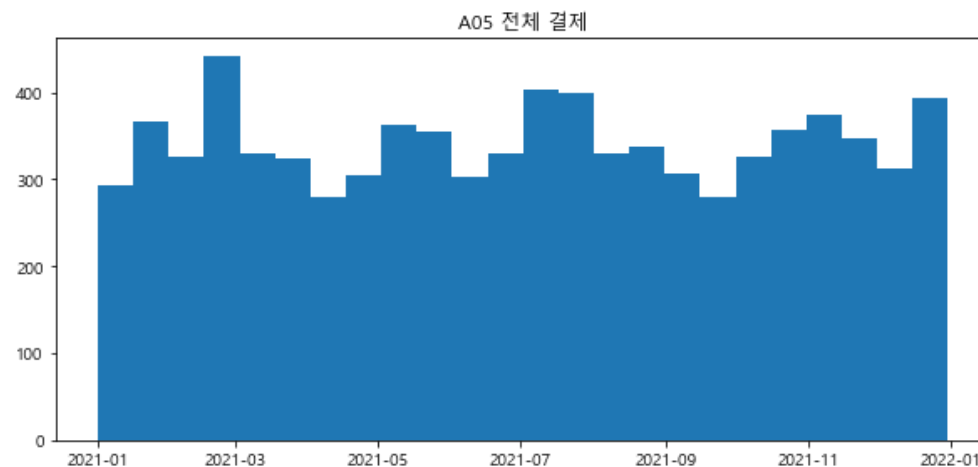
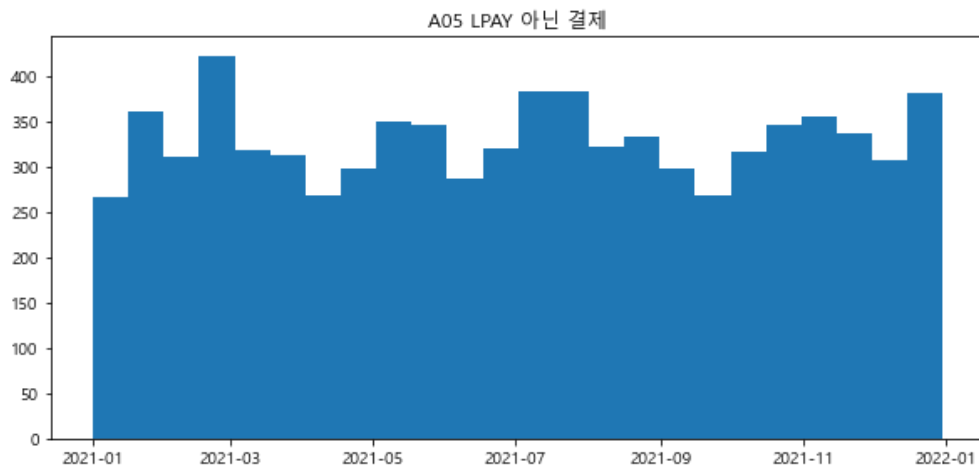
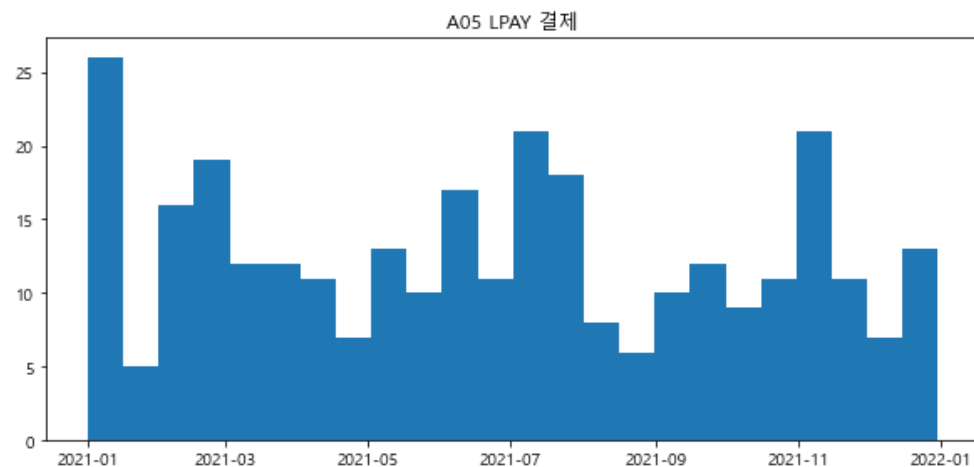


위와 같이 담배 구매량은 늘었으나,
대부분 상품의 구매 비율이 늘고
담배 구매 비율은 급격히 줄어 들음
(프로모션에 담배가 포함되지 않았기 때문)

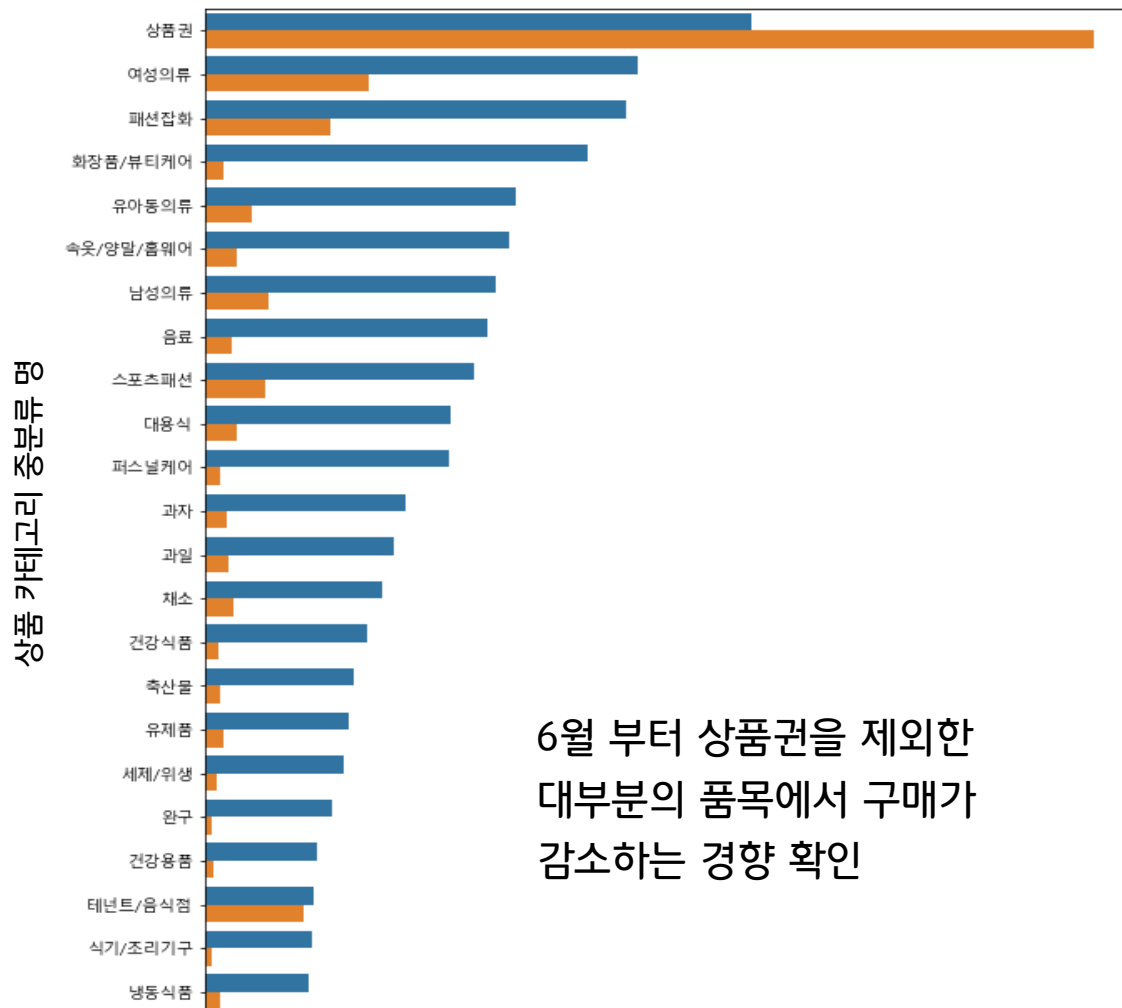
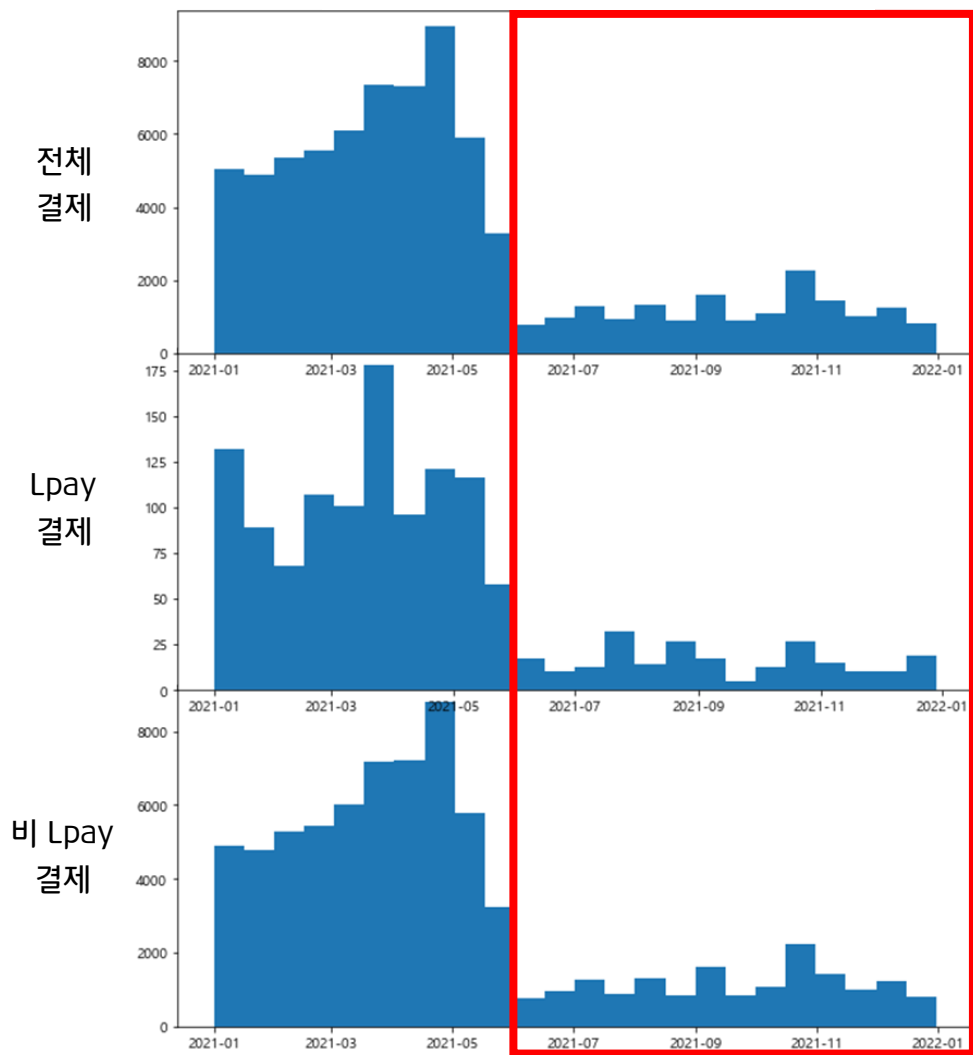


즉, 해당 프로모션으로 인해 Lpay 사용량이 증가
하고, 적용 상품 매출 증가에 긍정적 영향 주는 것
으로 확인

결제 경향 | A05(가전제품 종합 쇼핑몰)



결제 경향 | A06(온라인 쇼핑몰)



6월 부터 상품권을 제외한 대부분의 품목에서 구매가 감소하는 경향 확인

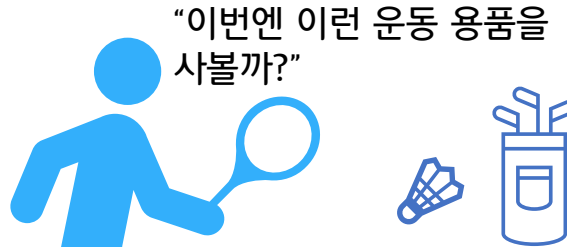
점수 구조

고객 별 상품 구매 횟수
0~5점

고객 별 대분류 구매 비율
0~5점

고객 별 상품 구매 간격 비율
0~5점

ex)



점수 가중치 평균

고객 별 상품 선호도(구매 가능성) 점수
0~5점

■ 생성 변수 데이터프레임 (by 고객, 상품)

cust	pd_c	① 고객 별 상품 구매 횟수	hlv_buy_am	buy_sum	② 고객 별 대분류 구매 비율	last_dt	range_mean	③ 고객 별 상품 구매 간격 비율
고객	상품	상품 구매 횟수	대분류 구매 횟수	고객 총 구매 횟수	대분류 구매 횟수 / 고객 총 구매 횟수	마지막 구매 날짜	상품의 평균 구매 간격	(오늘 - 마지막 구매 날짜) / 상품의 평균 구매 간격
A	c
A	e	간격이 없는 경우는 같은 대분류의 최대 구매 간 격으로 대체	...
B	a
B	c
...

<ratings> DataFrame

① 상품 구매 횟수가 많다면, 그 상품을 선호한다고 판단할 수 있음. 예) 키위를 1년간 20회 넘게 구매한 고객

② 고객마다 선호하는 대분류가 존재하며, 이에 해당하는 상품은 구매 가능성이 높음. 예) 스포츠 관련 상품들 주로 구매하는 고객 등...

③ 마지막 구매 날짜와 오늘 날짜 (2022.01.01)의 간격이 길수록 상품을 구매할 가능성은 높음.

상품의 평균 구매 간격이 짧을수록 구매 가능성이 높음.

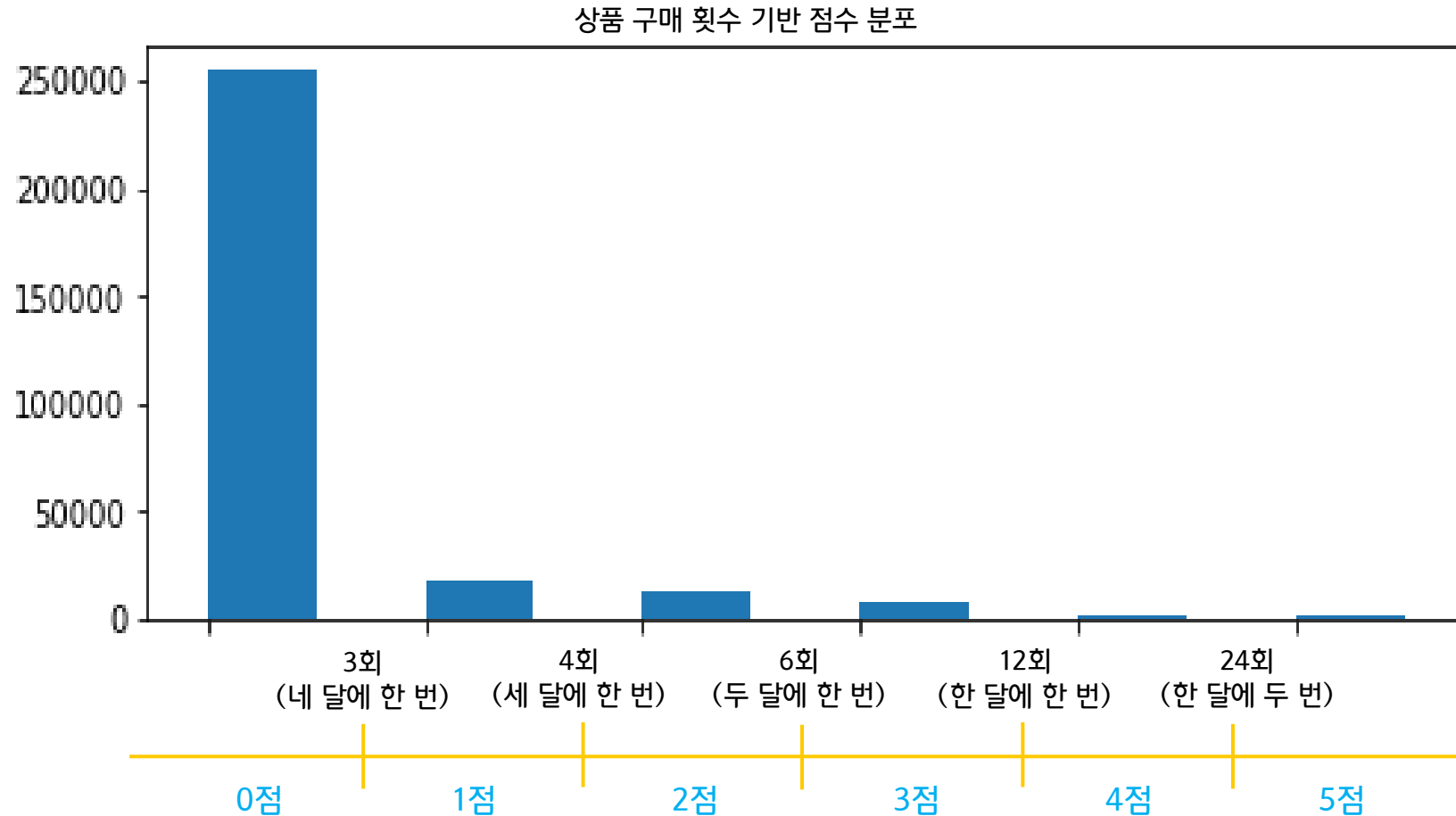
예) 평균적으로 100일만에 재구매한 상품보다 20일만에 재구매한 상품이 더 구매 가능성이 높음.

3가지 파생 변수를 바탕으로 0~5 사이의 3개 점수 생성

ex: A01 유통사

기준에 맞춰 0점 ~ 5점으로 점수화

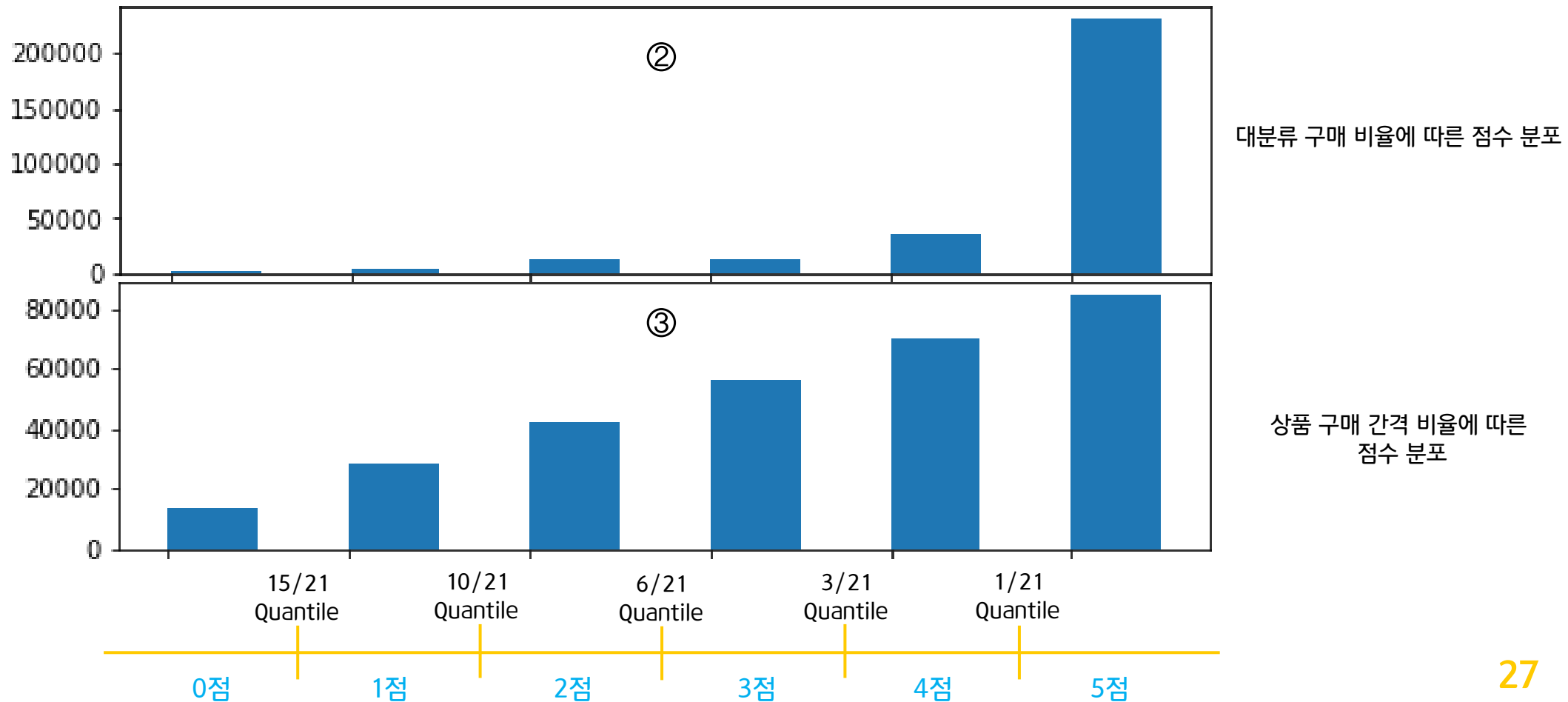
① 고객 별 상품 구매 횟수 기반 점수 : **buy_level**



ex: A01 유통사

② 고객 별 대분류 구매 비율 기반 점수 : low_clac

& ③ 고객 별 상품 구매 간격 비율 기반 점수 : range_score



ex: A01 유통사

구매 횟수의 중요도가 더 높다고 판단

$$\text{score} = \frac{\text{buy_level} \times 2 + \text{low_clac} + \text{range_score}}{4}$$

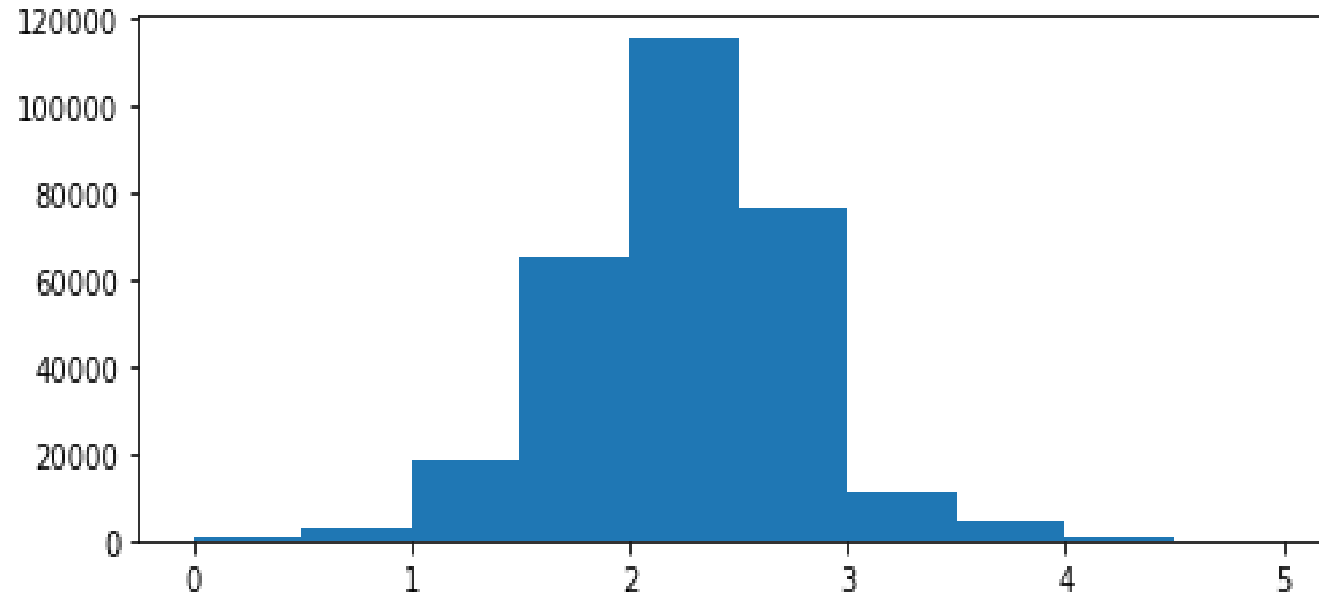
score
선호도 점수
(구매 가능성)

buy_level
고객 별 상품
구매 횟수 점수

low_clac
고객 별 대분류
구매 비율 점수

range_score
고객 별 상품
구매 간격 비율 점수

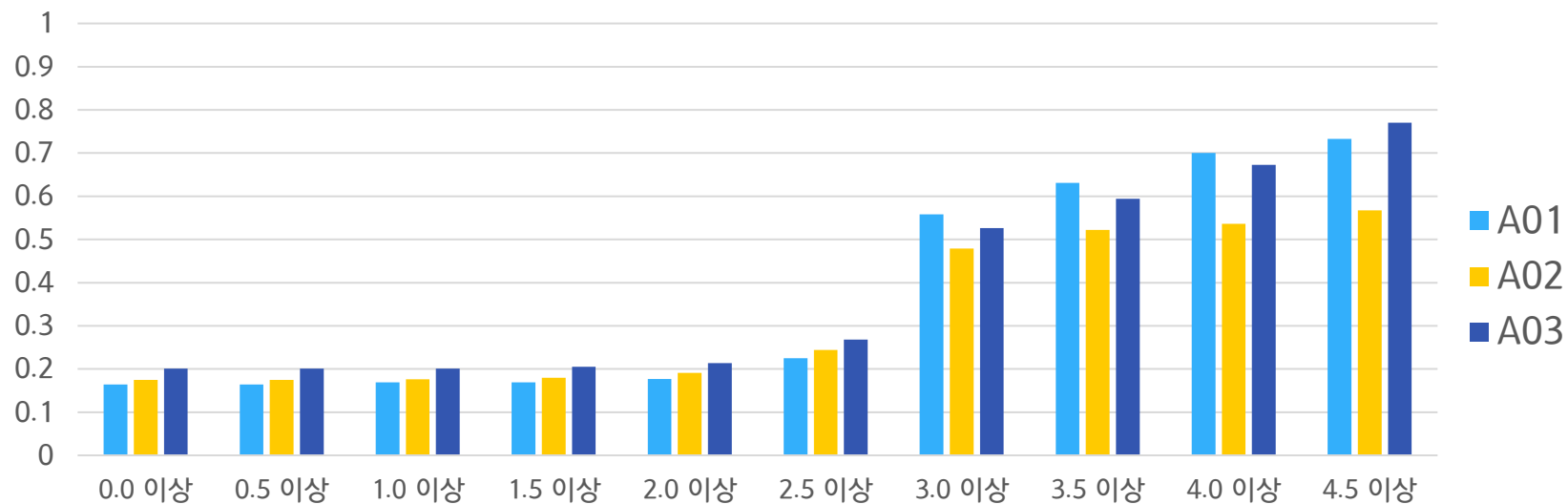
선호도 점수 분포



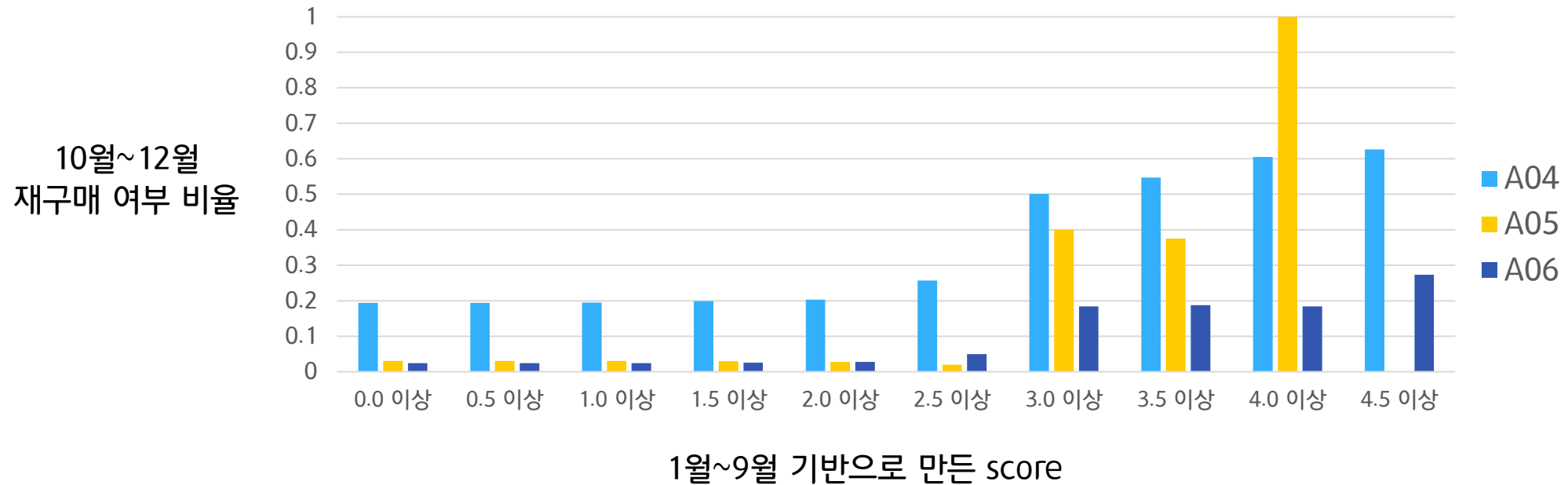


점수 체계의 신뢰성을 검증하기 위해, 앞선 9개월의 구매 데이터로 점수를 생성한 후
특정 선호도 점수(임계치) 이상의 상품들이 뒤 3개월에 재구매 된 비율을 확인

10월~12월
재구매 여부 비율



1월~9월 기반으로 만든 score



유독 재구매 비율이 낮은 A06

6월부터 급격하게 전체 구매 수가 줄었기 때문에, 재구매 여부 비율도 낮은 것으로 예상됨.

그럼에도 선호도 점수가 상승할수록 재구매 비율은 증가하는 경향을 보여줌 .

유통사 별 검증 (10월~12월 재구매 여부 비율)

A01	1월~9월 기반으로 만든 score	0.0 이상	0.5 이상	1.0 이상	1.5 이상	2.0 이상	2.5 이상	3.0 이상	3.5 이상	4.0 이상	4.5 이상
	10월~12월 재구매 여부 비율	0.164	0.164	0.169	0.169	0.177	0.225	0.558	0.631	0.7	0.733
A02	1월~9월 기반으로 만든 score	0.0 이상	0.5 이상	1.0 이상	1.5 이상	2.0 이상	2.5 이상	3.0 이상	3.5 이상	4.0 이상	4.5 이상
	10월~12월 재구매 여부 비율	0.175	0.175	0.176	0.18	0.191	0.244	0.479	0.522	0.536	0.567
A03	1월~9월 기반으로 만든 score	0.0 이상	0.5 이상	1.0 이상	1.5 이상	2.0 이상	2.5 이상	3.0 이상	3.5 이상	4.0 이상	4.5 이상
	10월~12월 재구매 여부 비율	0.201	0.201	0.201	0.205	0.214	0.268	0.526	0.594	0.673	0.77
A04	1월~9월 기반으로 만든 score	0.0 이상	0.5 이상	1.0 이상	1.5 이상	2.0 이상	2.5 이상	3.0 이상	3.5 이상	4.0 이상	4.5 이상
	10월~12월 재구매 여부 비율	0.194	0.194	0.195	0.199	0.203	0.257	0.501	0.547	0.605	0.626
A05	1월~9월 기반으로 만든 score	0.0 이상	0.5 이상	1.0 이상	1.5 이상	2.0 이상	2.5 이상	3.0 이상	3.5 이상	4.0 이상	4.5 이상
	10월~12월 재구매 여부 비율	0.031	0.031	0.031	0.03	0.028	0.02	0.4	0.375	1	-
A06	1월~9월 기반으로 만든 score	0.0 이상	0.5 이상	1.0 이상	1.5 이상	2.0 이상	2.5 이상	3.0 이상	3.5 이상	4.0 이상	4.5 이상
	10월~12월 재구매 여부 비율	0.024	0.024	0.024	0.026	0.028	0.05	0.184	0.188	0.184	0.273

재구매 비율 0.5이상

점수 체계 신뢰성 검증 결론

점수가 높아질수록 실제 재구매로 이어진 경우가 많음.

즉, 자체적으로 만든 점수가 고객의 실제 구매 여부를 일정 수준 예측한다고 볼 수 있음

또한 구매 기록이 있고 점수가 높을 경우 재구매 비율이 이미 높다고 할 수 있기에 예측이 필요하지 않음.

따라서 상품 추천은 고객마다 구매 이력이 없는 상품에 집중하여 구매 범위를 확장시켜야 한다는 근거가 됨

3. 모델링

■ 점수 예측 모델 선정 | 딥러닝 예측 모델

왜 많은 추천시스템 모델 중 딥러닝 모델을 택하였는가?



본 프로젝트는 고객-상품 간의 모든 경우의 수를 다뤄 추천시스템을 만들어야 함



데이터가 매우 희박할 수밖에 없음



흔히 쓰이는 MF 모델을 사용하면 성능 저하가 발생할 가능성이 높기 때문에 딥러닝 모델 채택

Label Encoding

고객	label
A	0
B	1
C	2
D	3
E	4
...	...

<customer_info>

상품	label
a	0
b	1
c	2
d	3
e	4
...	...

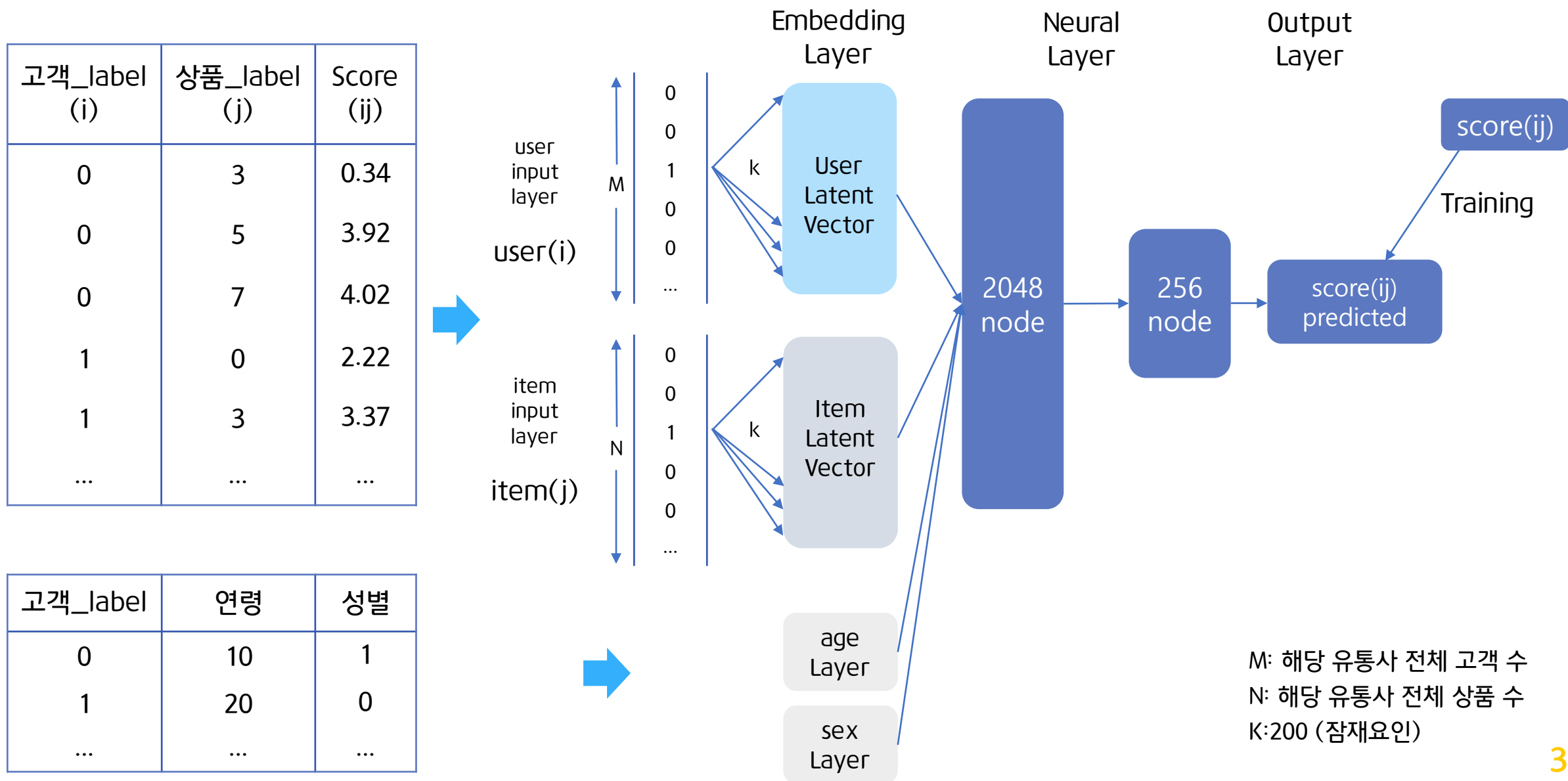
<pd_info>

성별	label
남성	0
여성	1

나이대	ages
20대	20.0
30대	30.0
40대	40.0
50대	50.0
60대	60.0
70대	70.0

모델에 적용하기 위해 고객, 상품, 성별에 label을 할당

나이대는 값 그대로 수치형으로 변환



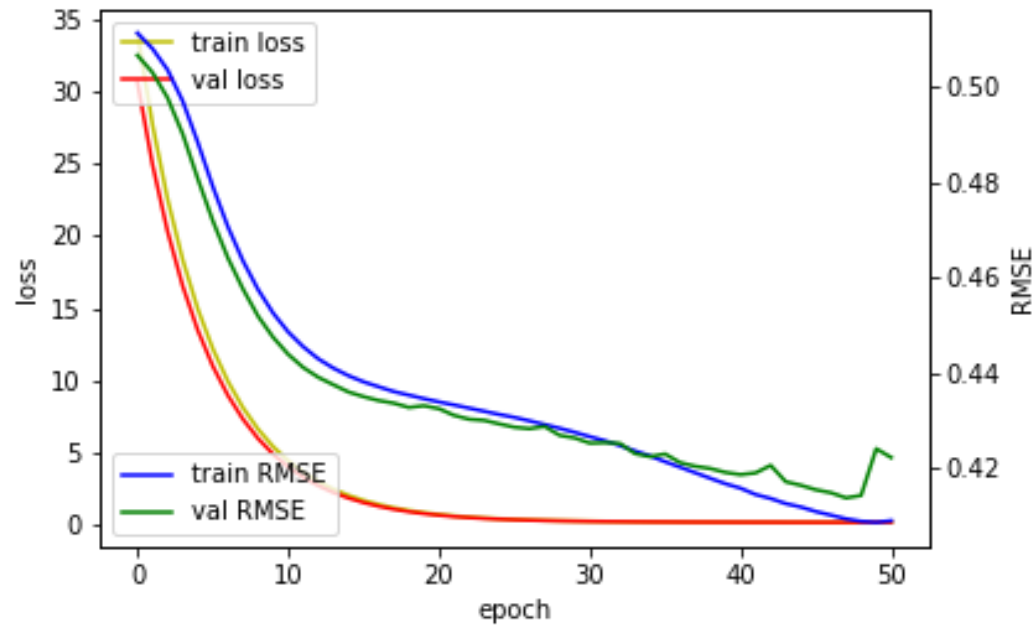
■ 학습 & 검증 데이터

상품 고객	0	1	2	3	4	5	...
0	3.xx					4.xx	
1			3.xx		1.xx		
2		3.xx			0.xx	2.xx	
3		0.xx	2.xx		1.xx		
4				1.xx			
5		4.xx				1.xx	
6	3.xx				5.xx		
...							

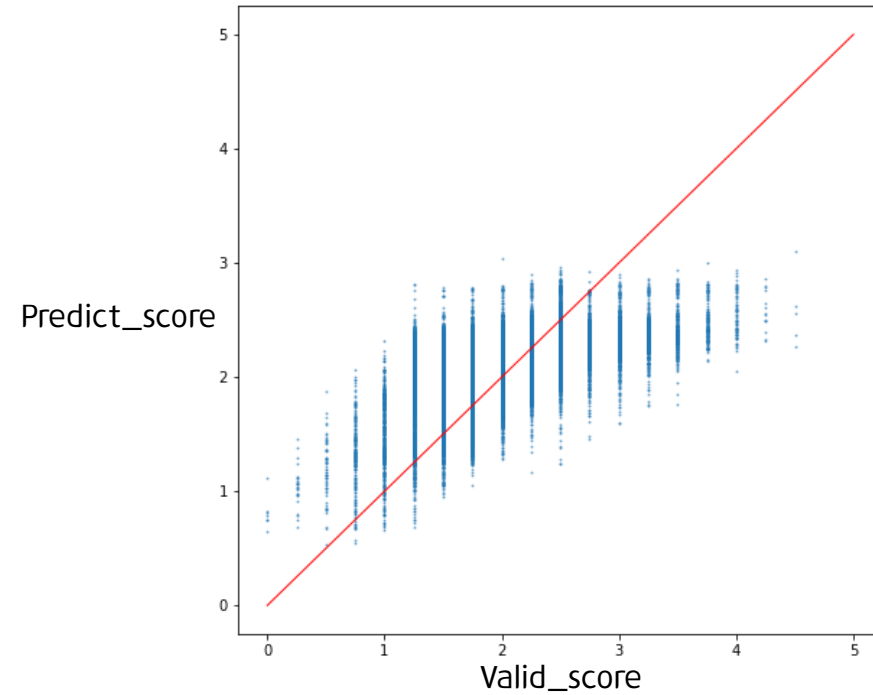
train : valid = 9 : 1

점수 존재하는 고객-상품 데이터를
9:1 비율로 나누어 학습 및 검증 진행

ex: A01 유통사



valid set의 RMSE는 약 0.42



실제 선호도 점수가 높을 때
예측 점수도 높은 경향을 보임

Output 데이터 생성

상품 \ 고객	0	1	2	3	4	5	...
0	3.xx	x.xx	x.xx	x.xx	x.xx	4.xx	x.xx
1	x.xx	x.xx	3.xx	x.xx	1.xx	x.xx	x.xx
2	x.xx	3.xx	x.xx	x.xx	0.xx	2.xx	x.xx
3	x.xx	0.xx	2.xx	x.xx	1.xx	x.xx	x.xx
4	x.xx	x.xx	x.xx	1.xx	x.xx	x.xx	x.xx
5	x.xx	4.xx	x.xx	x.xx	x.xx	1.xx	x.xx
6	3.xx	x.xx	x.xx	x.xx	5.xx	x.xx	x.xx
...	x.xx	x.xx	x.xx	x.xx	x.xx	x.xx	x.xx

고객	상품	예측 점수
0	1	x.xx
0	2	x.xx
0	3	x.xx
0	4	x.xx
1	0	x.xx
1	1	x.xx
1	3	x.xx
1	5	x.xx
...

<Output>

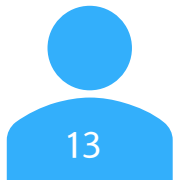
예측 점수를 포함한 Output
데이터프레임 생성

나머지 유통사(A02~a06)에 대해서도 동일한
모델링 및 OUTPUT 생성 과정 진행 [부록 참조]

4. 활용 방안

■ 마케팅 예시

모델_활용_방안.ipynb | 구현 완료



고객	A01	A02	A03	A04	A05	A06
13	0.34	0.001	0.47	0.12	0.01	0.058

고객	상품	예측 점수
13	1	2.87
13	2	3.11
13	3	3.97
13	4	0.02
...

<A03 Output>

■ 개발 시스템 실제 실행 결과 예시

고객
M430112881 (label:4038)

A03	A02	A01	A04
0.64	0.26	0.09	0.007



<A03 Output>

고객_label	상품_label	나이	성별	예측점수
4038	666	50	1	2.899
4038	200	50	1	2.894
...
4038	551	50	1	2.477
4038	623	50	1	2.472
...
...
...

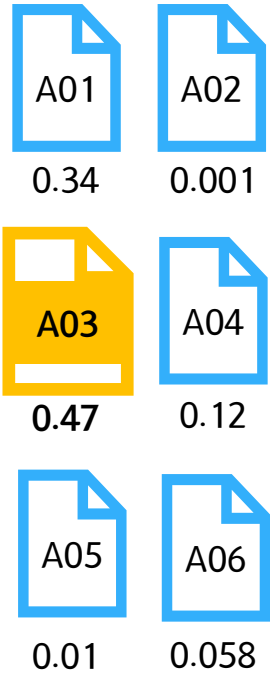
상위 30개 상품 예측점수 기반 가중 샘플링



상품추천	상품명	대분류	중분류
PD1595	기저귀크림/파우더	출산/육아용품	유아스킨/바디케어

모델_활용_방안.ipynb 내 실행 결과

방안 1. 비율 기반 유통사 선택 추천 일반 고객 대상



방안 2. 고객 선택형 구독 서비스 충성 고객 대상



■ 고객이 선택한 유통사의 구독 추천 서비스

- 구독 유통사 내 추천 상품 Lpay 구매 시 할인 혜택
- 구독 유통사 내 추천 상품 구매 시 Lpoint 추가 적립

등의 구독 고객 맞춤 혜택 제공 및 모델 기반한 상품 추천

충성 고객에 대한 맞춤 혜택 제공 및 상품 추천 가능



최근 시작한 롯데멤버스 엘페이 프로모션 확인

온·오프라인 전 지점에서 Lpoint 추가 적립 혜택을 받을 수 있는 유료 멤버십 서비스 “Lpay 프리미엄” 이 존재

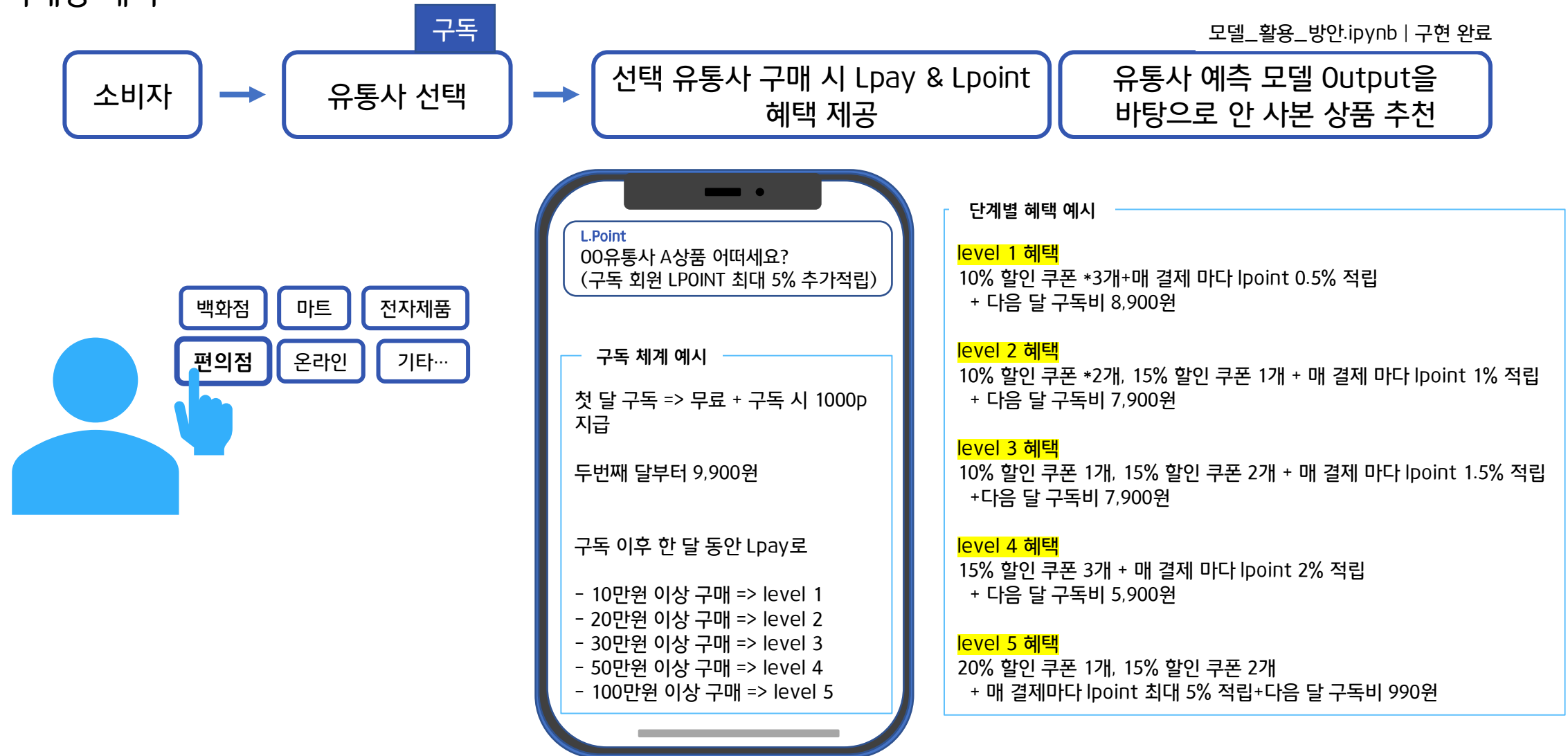


본 조가 만든 모델을 바탕으로

고객이 **유통사를 선택하여 맞춤형 혜택을 제공**받을 수 있는 기회를 제공

예측 점수 기반 상품 추천으로 인한 **개인화 마케팅** 가능

■ 마케팅 예시



■ 개발 시스템 실제 실행 결과 예시

고객
M430112881 (label:4038)

A02 구독 선택!

<A02 Output>

고객_label	상품_label	나이	성별	예측점수
4038	666	50	1	2.899
4038	200	50	1	2.894
...
4038	551	50	1	2.477
4038	623	50	1	2.472
...
...
...

상위 30개 상품 예측점수 기반 가중 샘플링

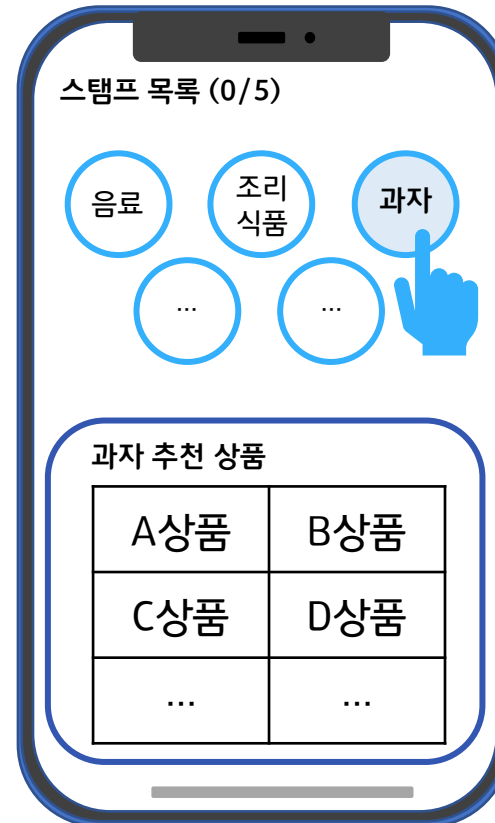


상품추천	상품명	대분류	중분류
PD1153	일반요구르트	유제품	요구르트

모델_활용_방안.ipynb 내 실행 결과

마케팅 예시

모델_활용_방안.ipynb | 구현 완료



예측 점수 기반 상품 추천

■ 개발 시스템 실제 실행 결과 예시

이번달 스탬프는
A04에서 진행!

고객
M430112881 (label:4038)

<A04 Output>

고객_label	상품_label	나이	성별	상품 대분류	예측점수
4038	666	50	1	생활/렌탈 서비스	2.899
...
4038	200	50	1	음료	2.894
...
4038	551	50	1	조리식품	2.477
...
4038	623	50	1	음료	2.472
...

평균 예측 점수 TOP5 대분류 선정

생활/렌탈서비스	음료	조리식품	...	스탬프 대분류
A1상품	B1상품	C1상품	...	추천상품1
A2상품	B2상품	C2상품	...	추천상품2
A3상품	B3상품	C3상품	...	추천상품3
A4상품	B4상품	B4상품	...	추천상품4
...

× 5개

모델_활용_방안.ipynb 내 실행 결과

활용 방안 1. 비율 기반 유통사
선택 추천
일반 고객 대상

활용 방안 2. 고객 선택형 구독
서비스
충성 고객 대상

활용 방안 3. 스탬프 미션
챌린지
전체 고객 대상

<기대 효과>

✓ 무분별하고 일괄적인 광고를 효율적으로 변경, 광고비용 절감 및 소비자 피로도 감소

✓ Lpay 결제를 조건으로 둔다면 Lpay 홍보 or 촉진 효과

21P | A04 Lpay 사용량
증가 프로모션을 근거로

✓ 상품 추천을 통한 구매 범위 확장 및 직접적 이윤 창출

✓ 고객 맞춤 추천을 통해 기업 추천
서비스 신뢰 상승, 일반 고객을 충
성 고객으로 유도 가능

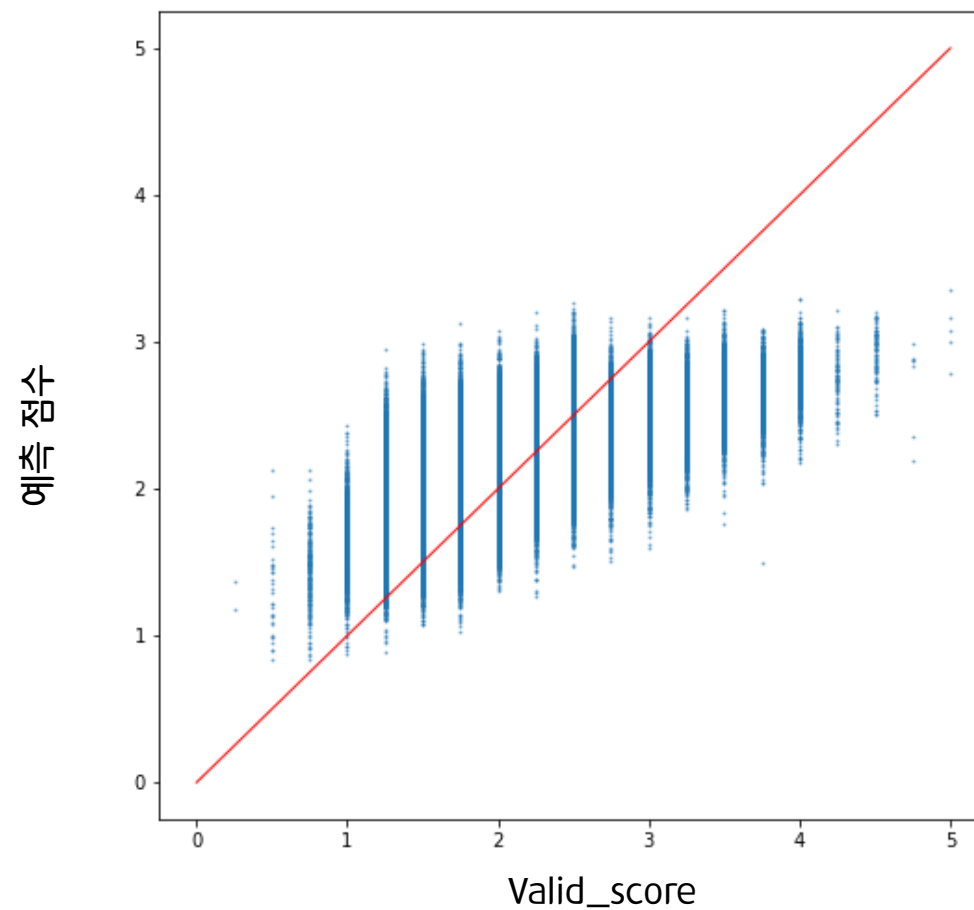
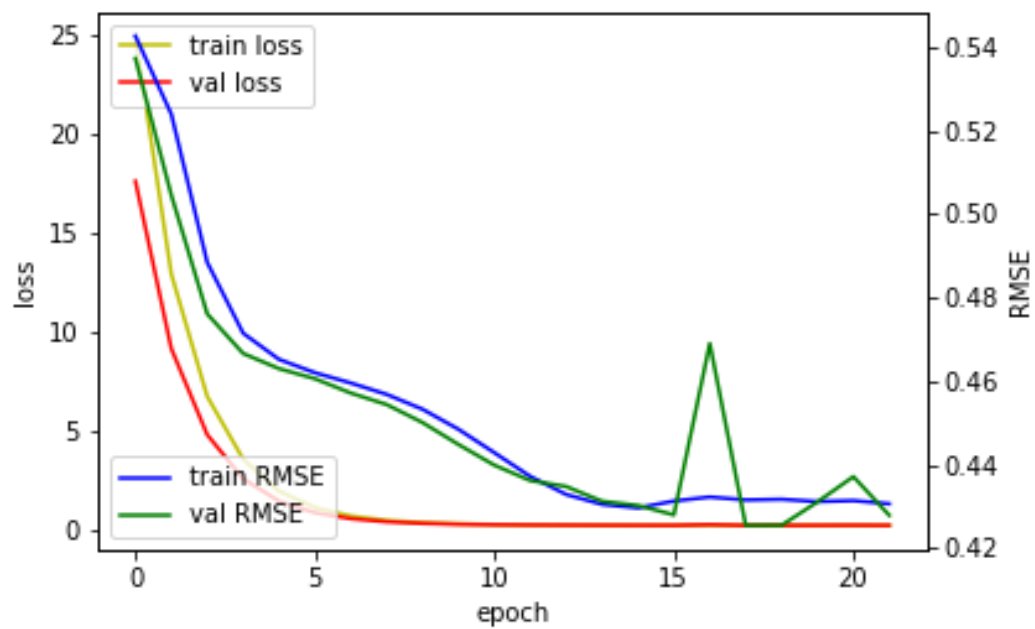
✓ 구독 비용을 통한 이윤 창출
✓ 실제 소비자가 원하는 유통사에
맞춰서 예측 가능
✓ 혜택 및 추천에 대한 만족, 충성
고객 이탈 방지

✓ 챌린지 진행을 통한 소비자의
흥미 유도
✓ 스탬프 대분류 해당 상품
구매 유도 가능
✓ 신규 고객 유입 증가

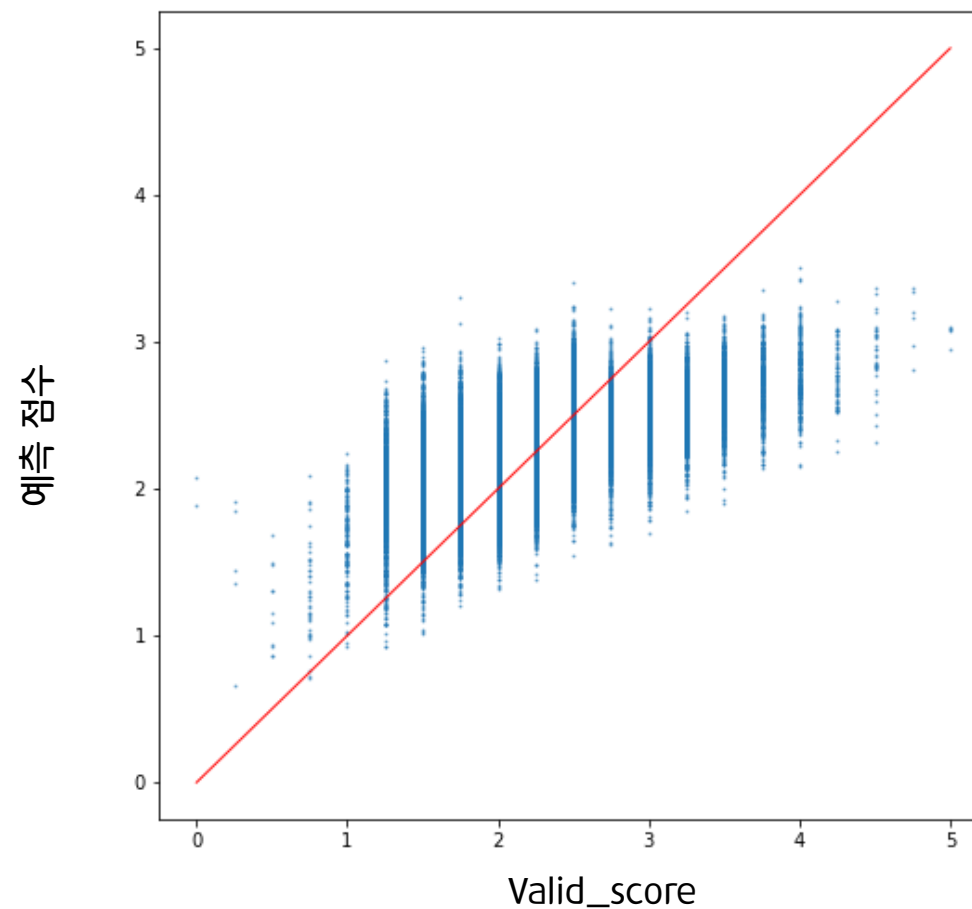
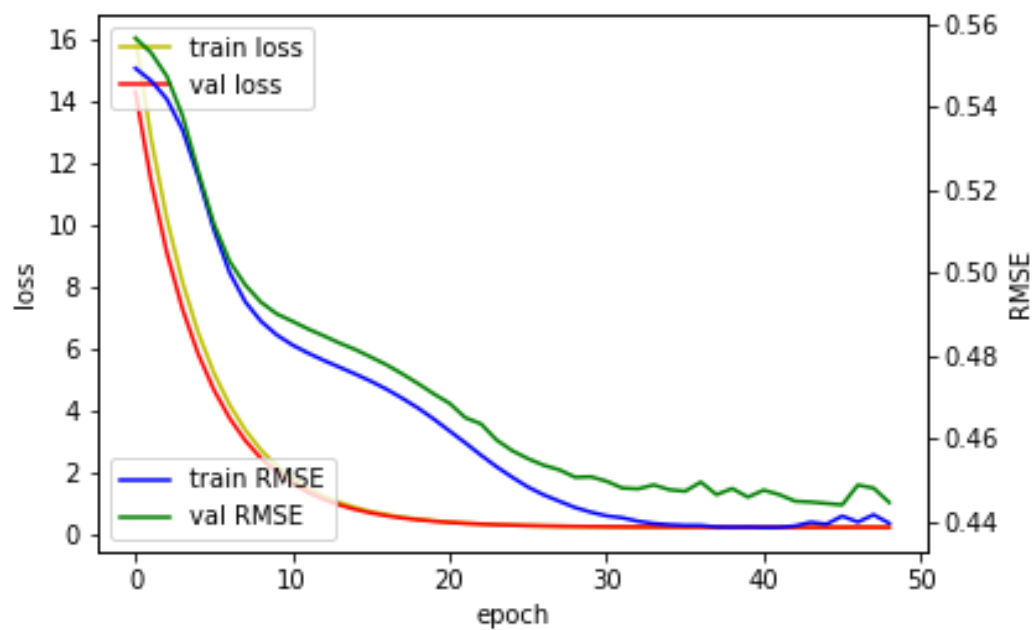


부록

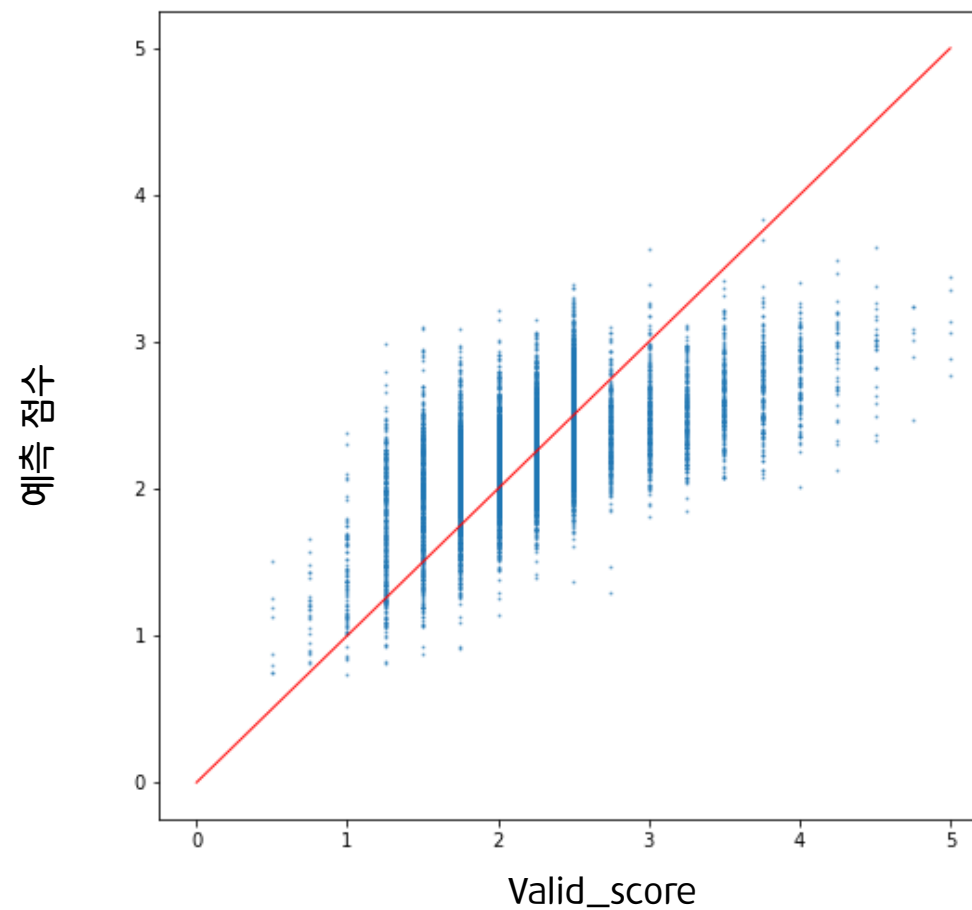
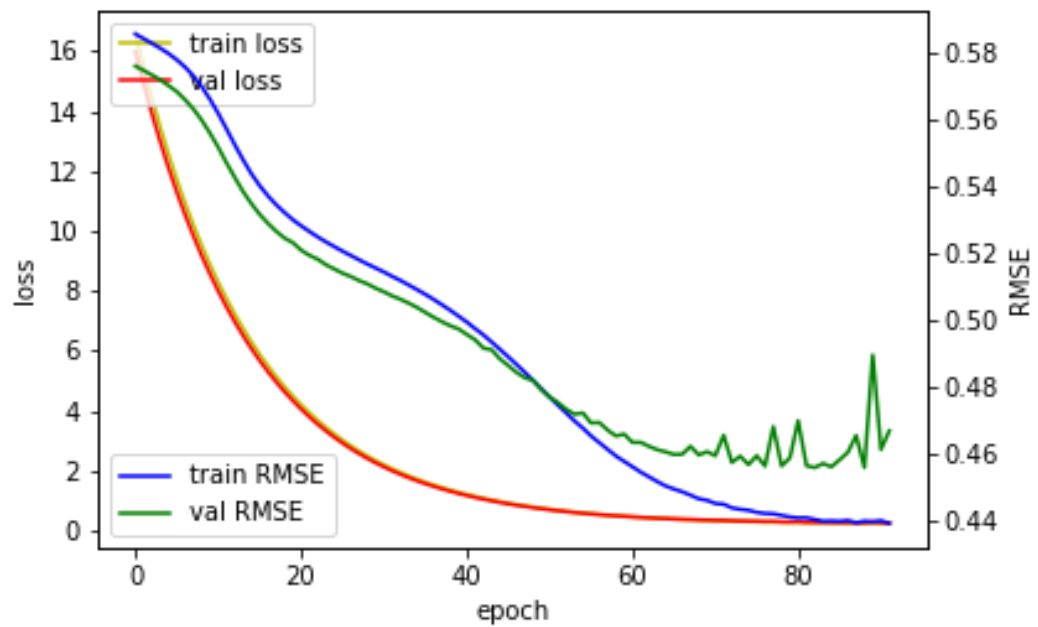
■ A02 점수 예측 모델 성능 검증



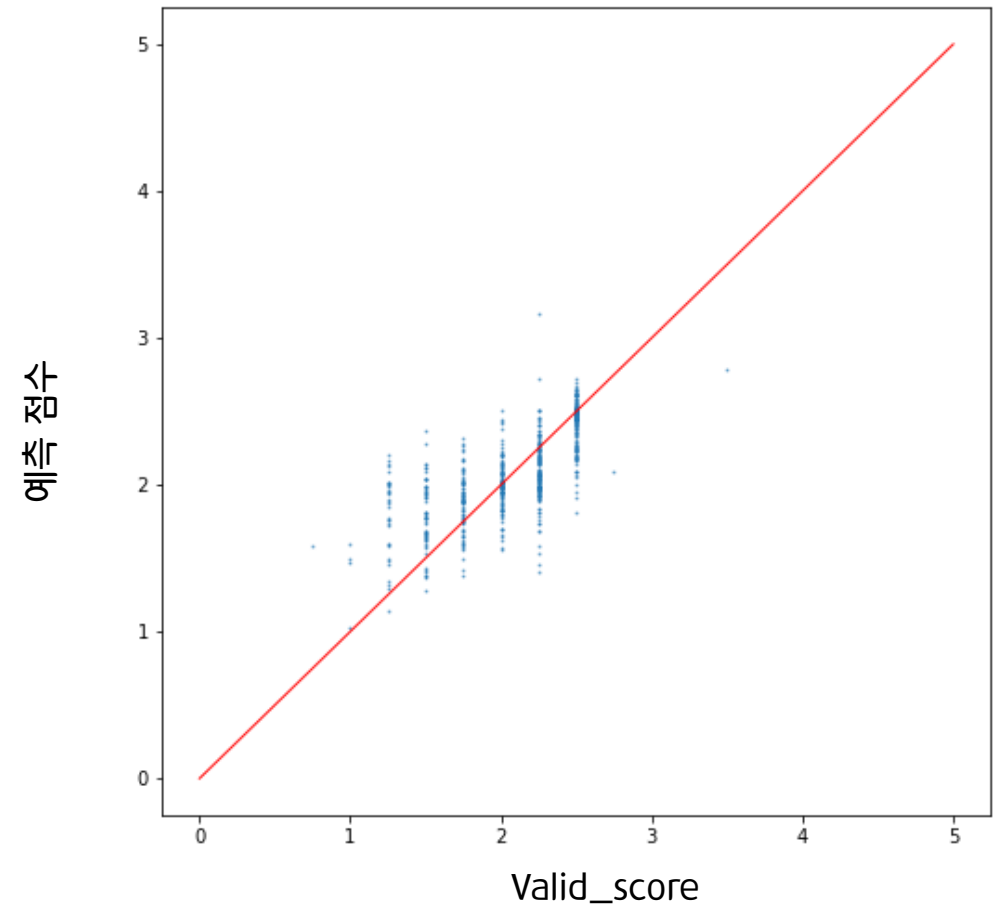
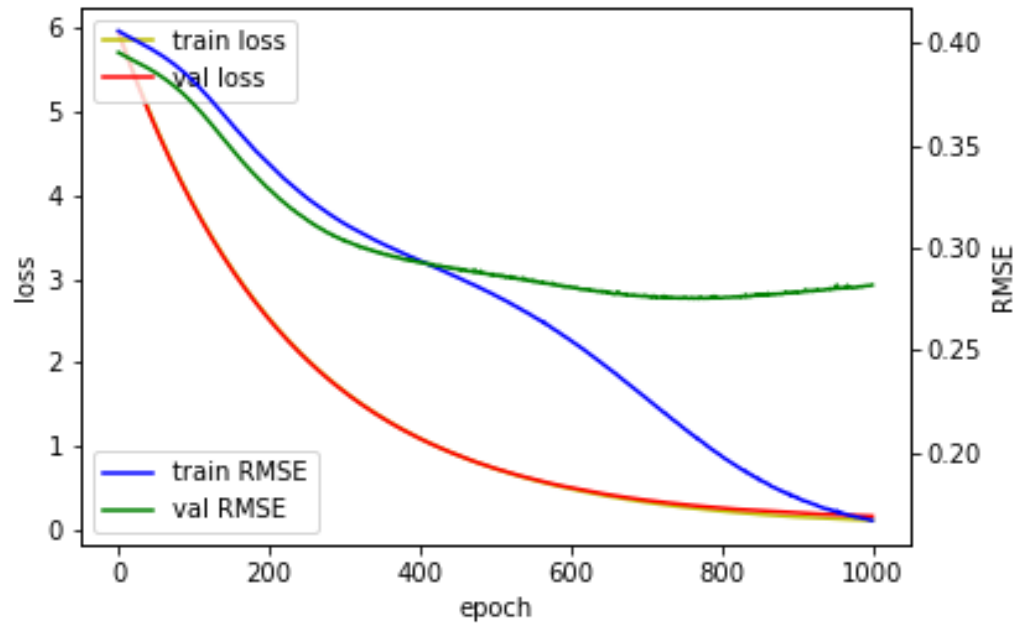
A03 점수 예측 모델 성능 검증



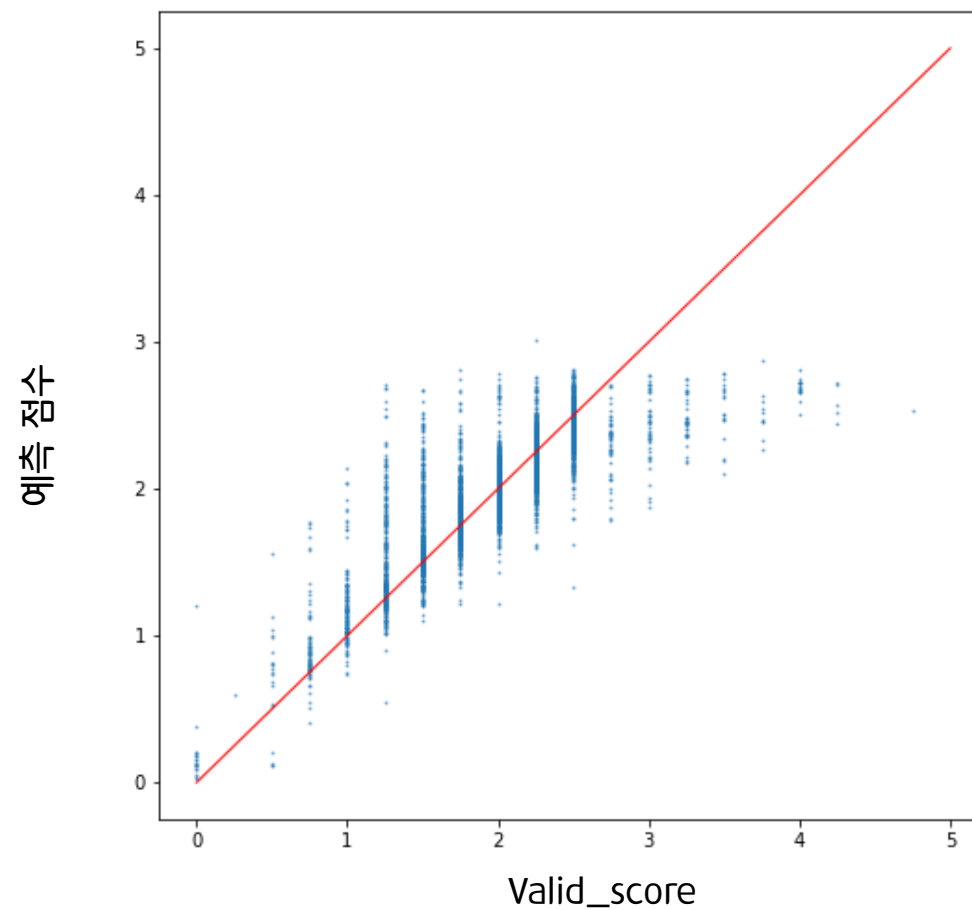
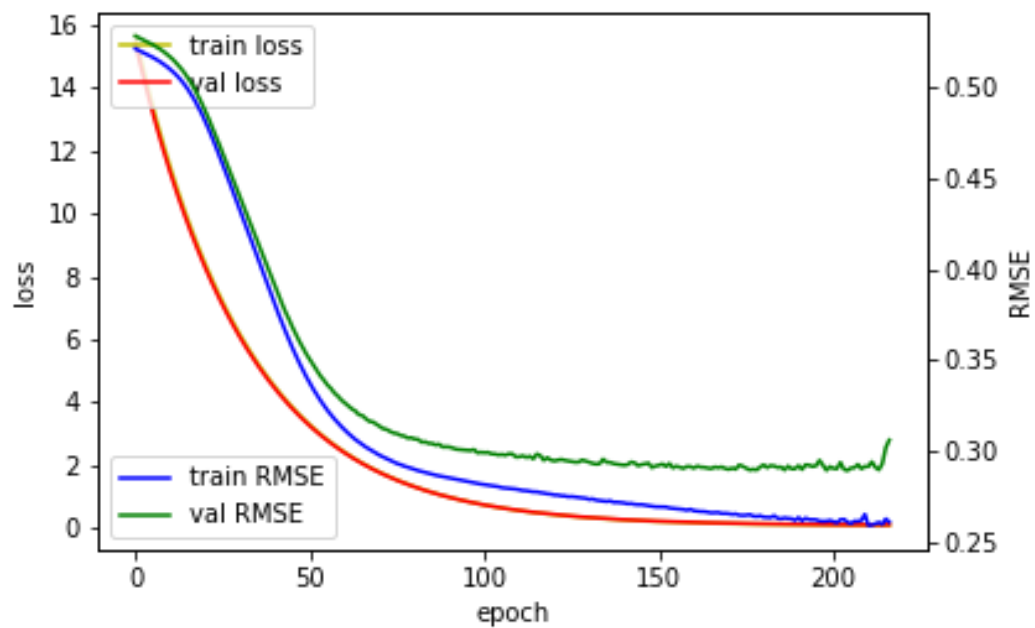
A04 점수 예측 모델 성능 검증



■ A05 점수 예측 모델 성능 검증

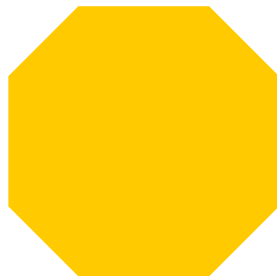


■ A06 점수 예측 모델 성능 검증





감사합니다



Team 비타플