

# L-STARGRAM

영수증도 스웨이드예요?

SSUDA

김성연 이동현 조경덕

# PART I

고객과 고객을 연결하는  
개인화 마케팅 C2C플랫폼 :

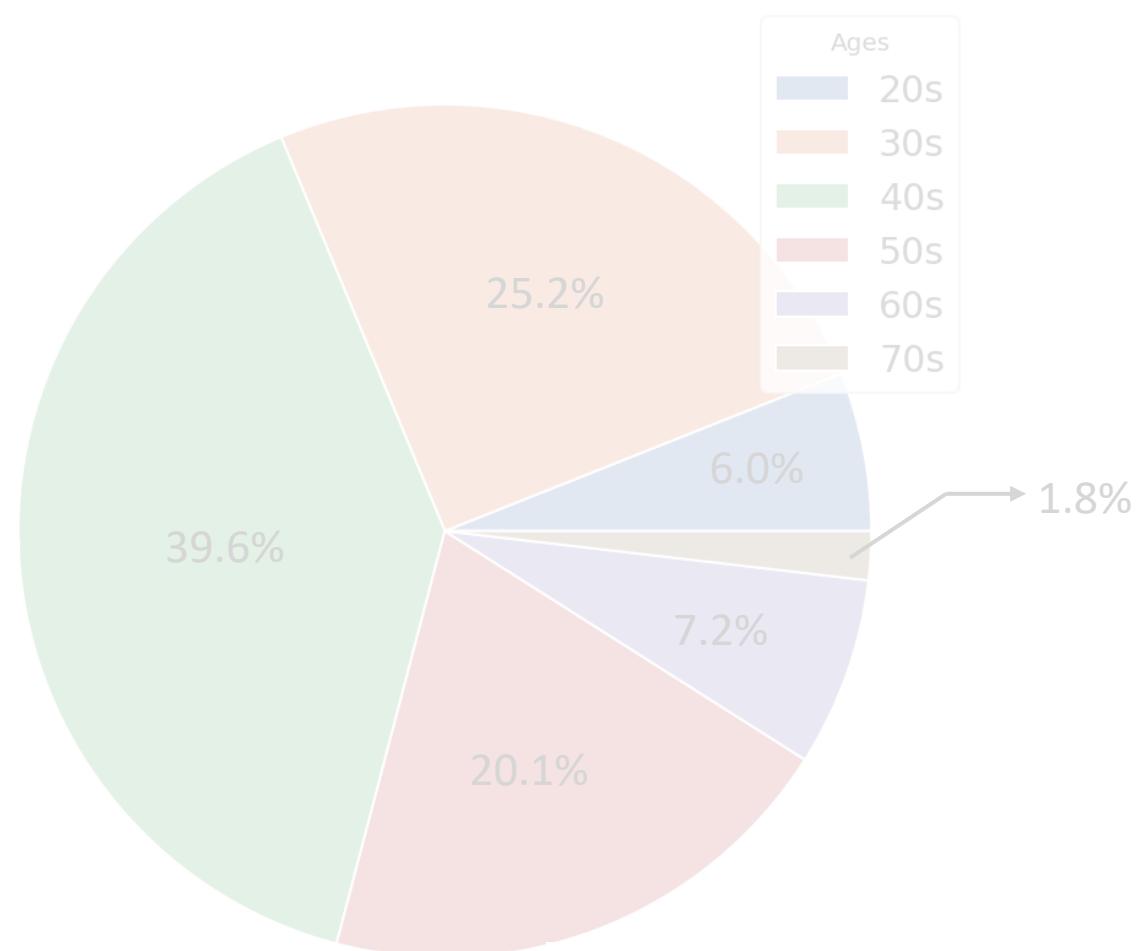
## L-Stargram

- EDA
- Recommend System

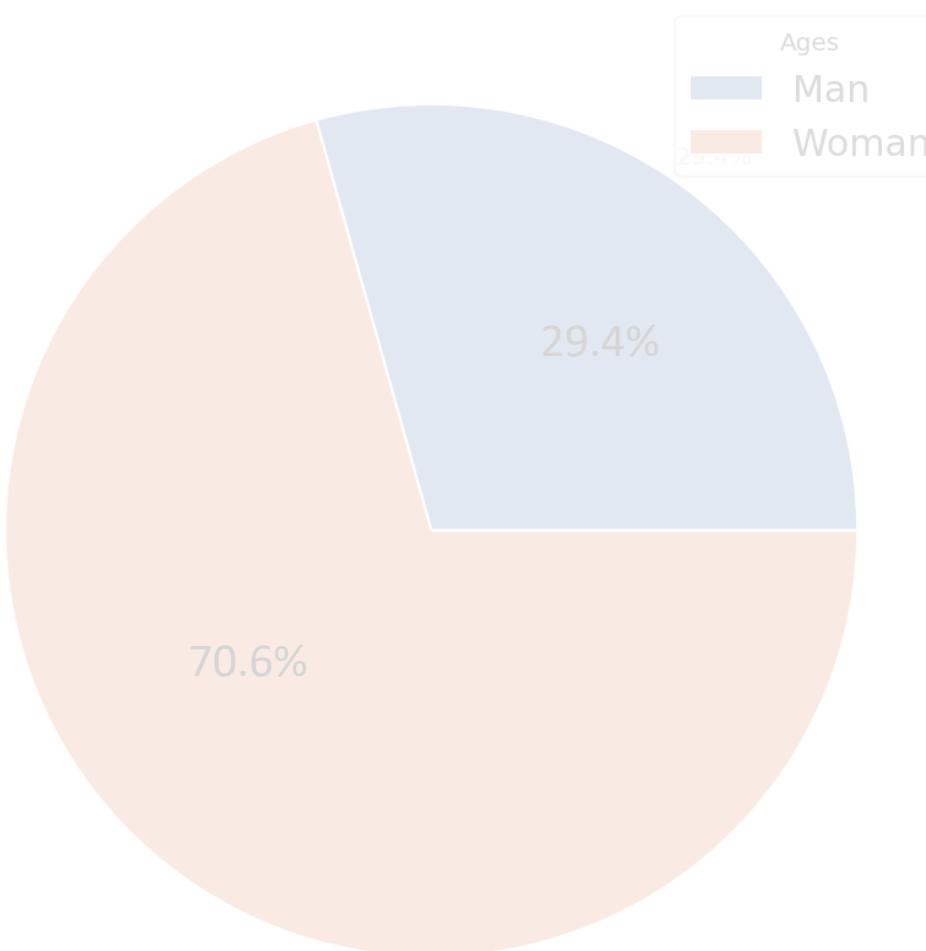


# PART I

purchasing power

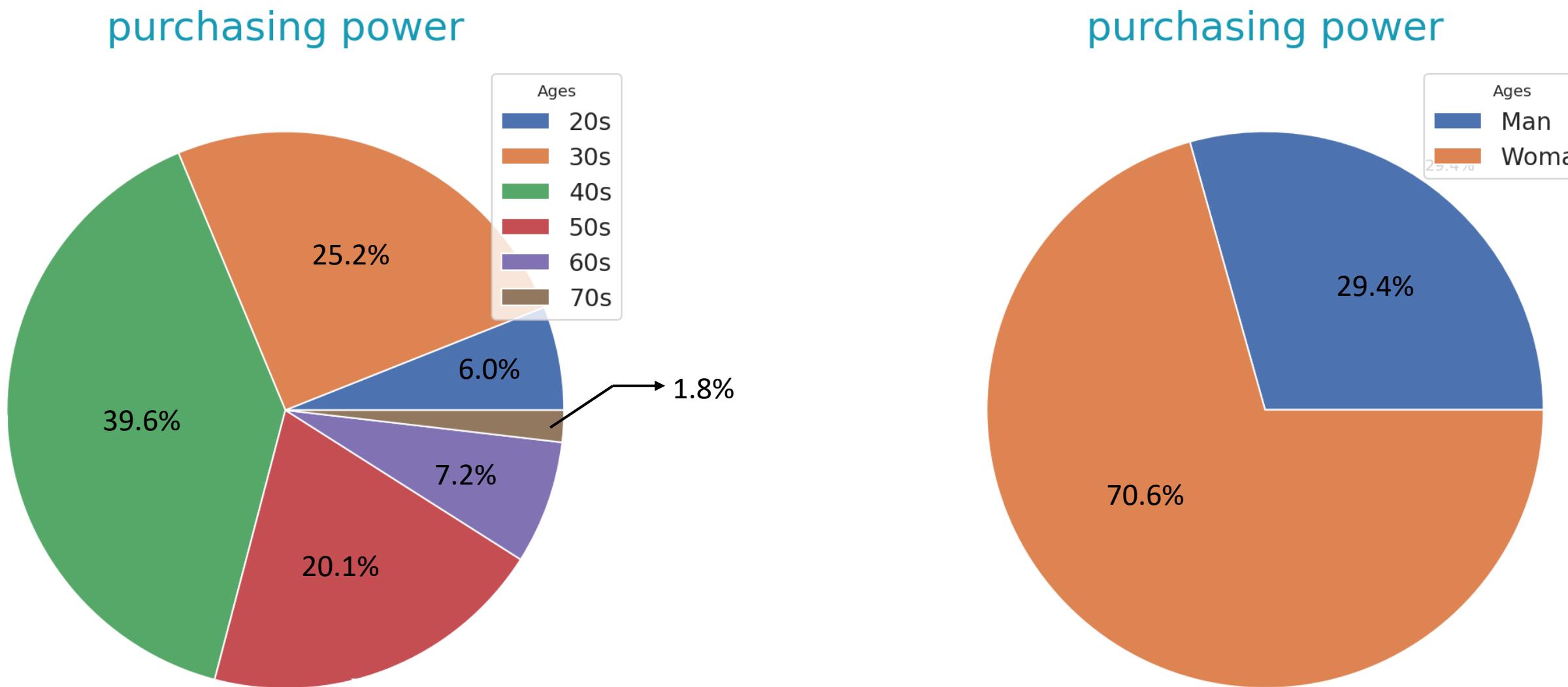


purchasing power



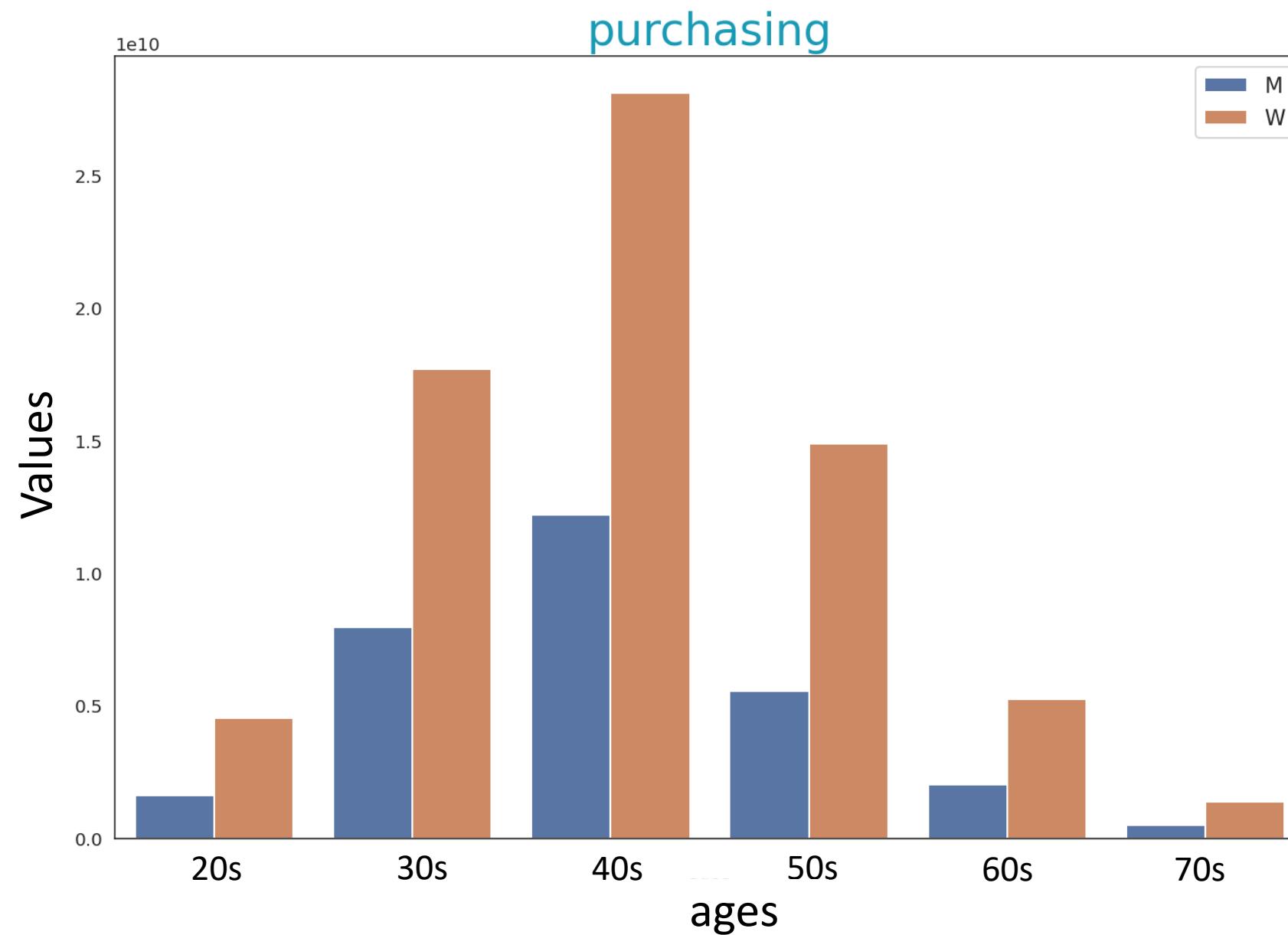
# EDA

## 연령대 및 성별 그룹별 구매력(결제금액합) 백분율

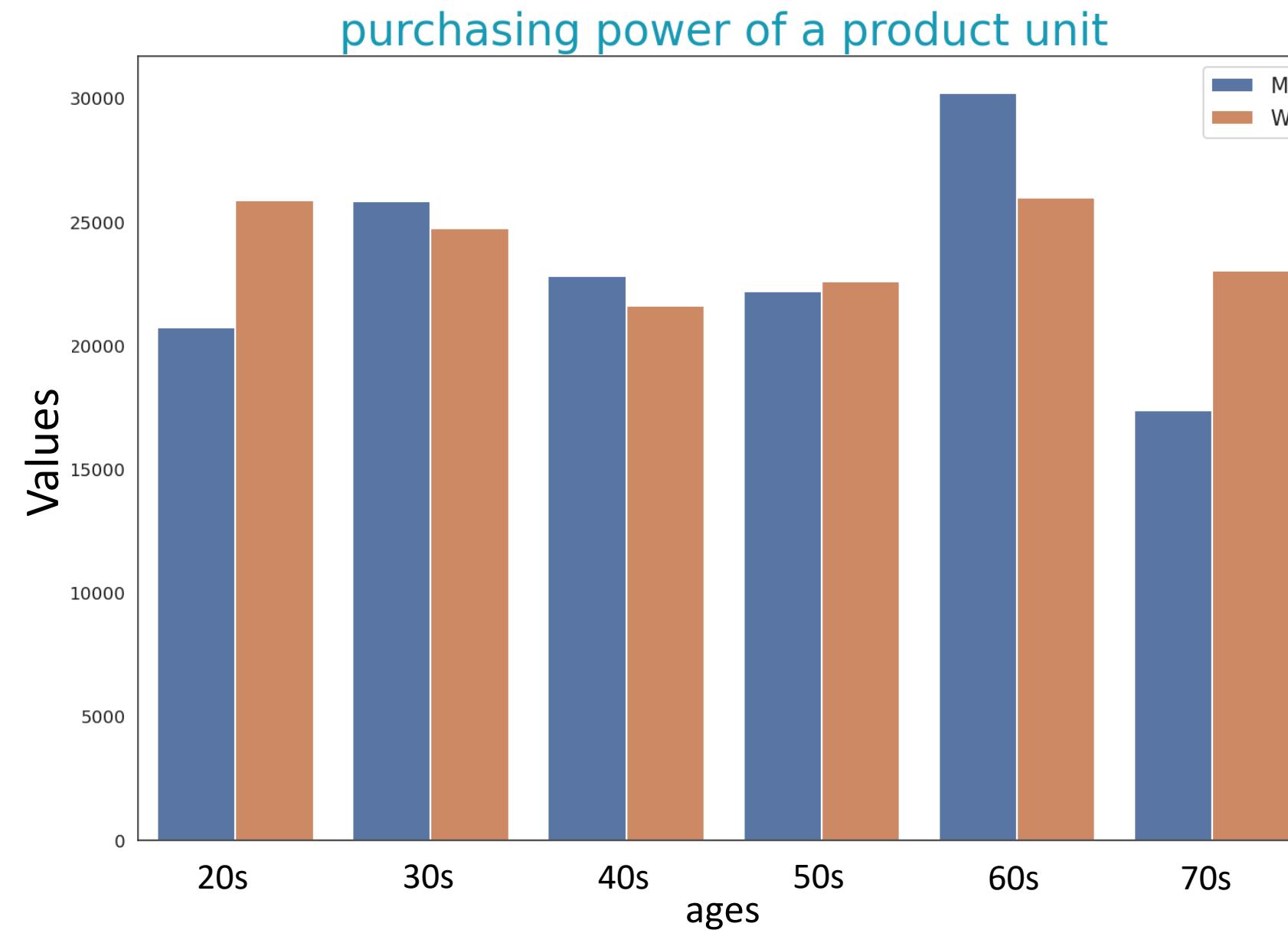


전체 결제 금액의 대부분을 30, 40, 50대 그룹이 차지하며  
여성 그룹의 구매력(결제금액합)은 남성 그룹의 2배 이상이다.

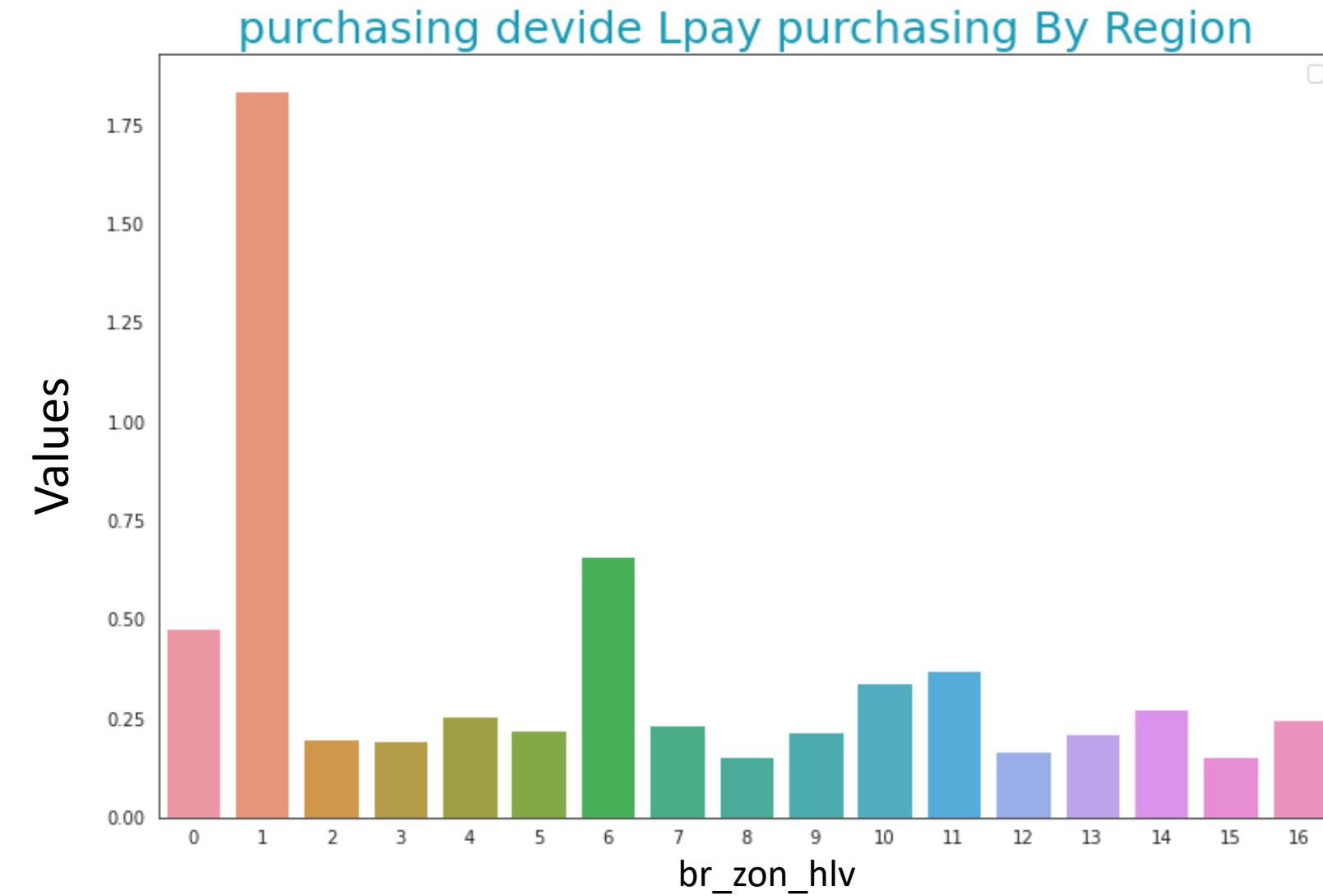
## 성별&amp;연령대 그룹별 구매력(결제금액합) 막대 그래프



전 연령대에서 여성 그룹이 남성 그룹보다 구매력이 높으며  
특히, 40대 여성 그룹이 압도적으로 높다.

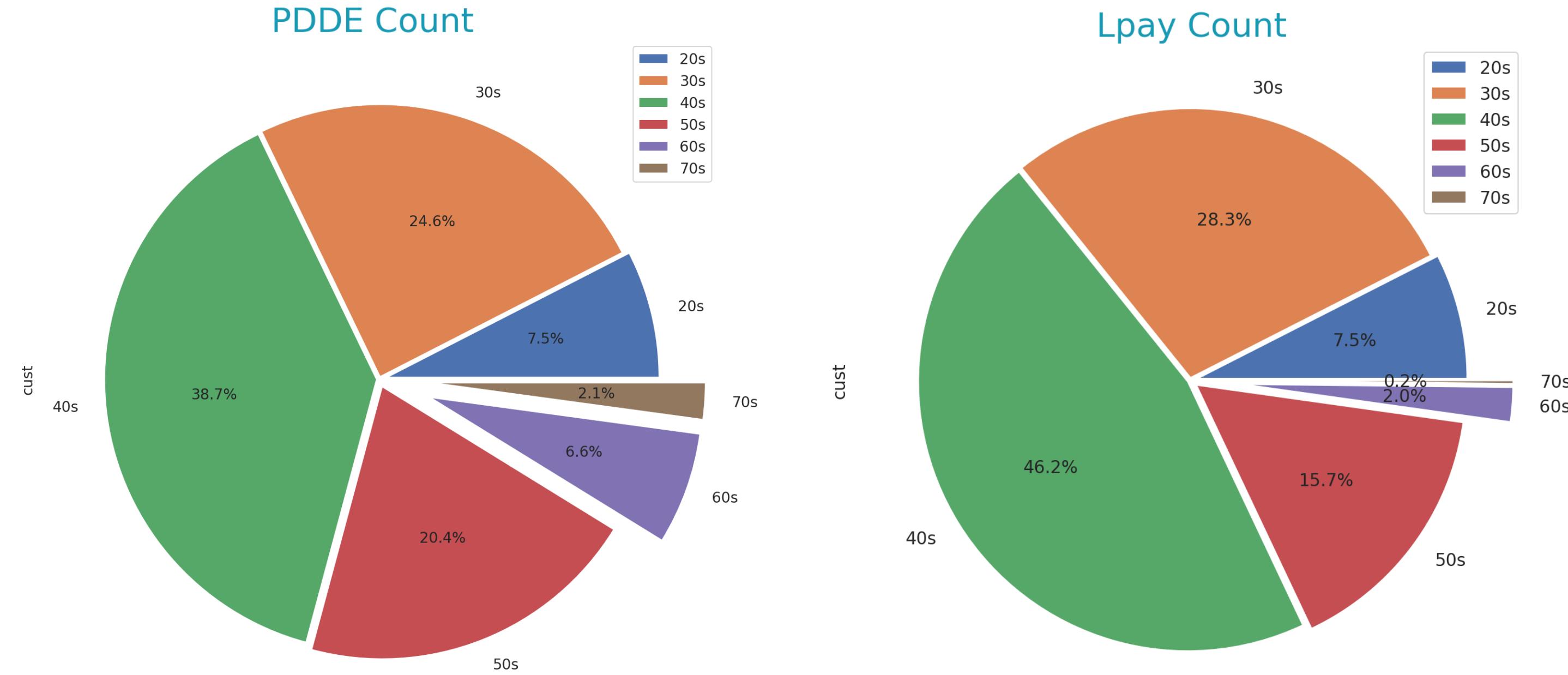


개인으로 본다면 20대도 다른 세대에 비해 크게 밀리진 않는다.  
60대 남성들은 평균적으로 비싼 제품을 많이 산다.  
(그룹의 평균값은 개인으로 볼 수 있다.)



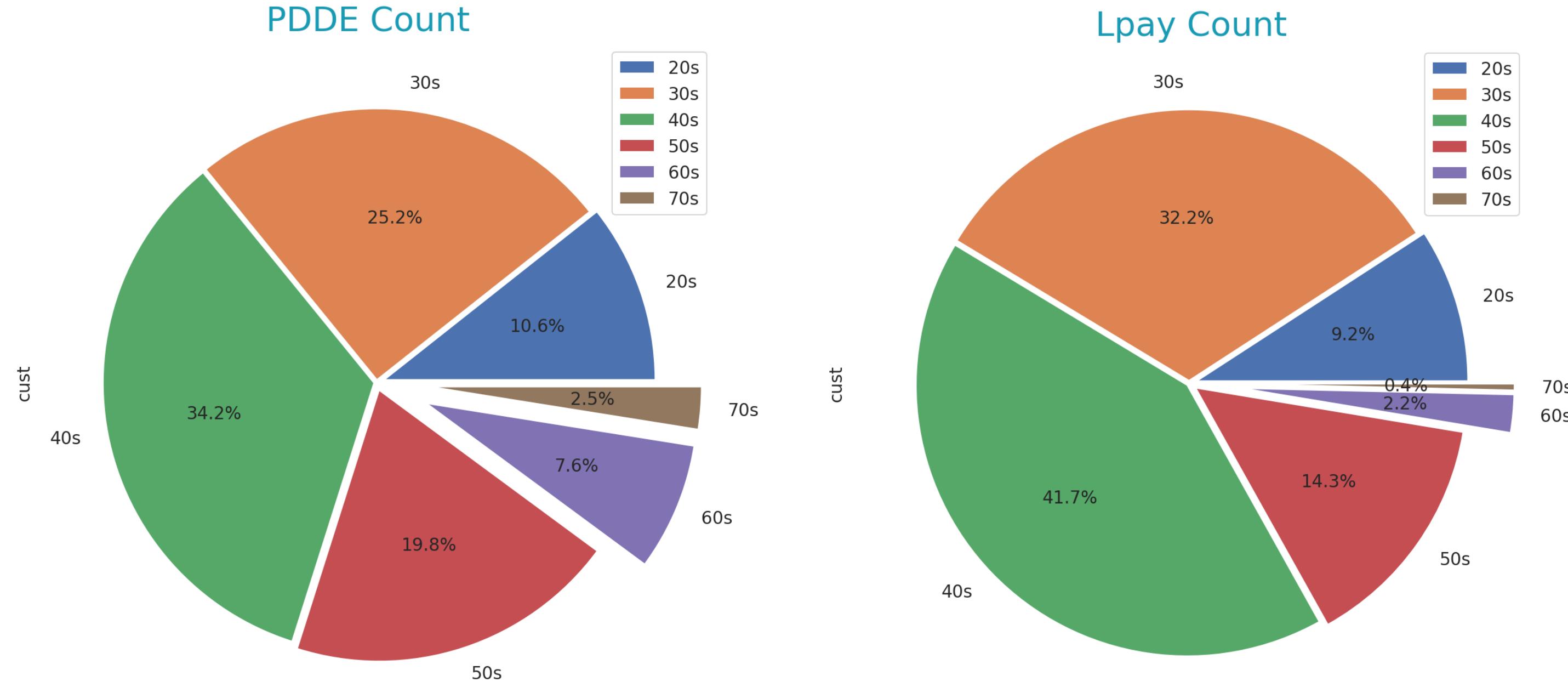
지역별로 LPay 이용률이 조금씩 차이가 있다.

## PDDE와 LPay 데이터 내 연령대별 이용량(결제건수) 분포 비교



60대 이상은 전체 상품결제 비율에 비해 LPay 이용량이 적었다.  
20대는 거의 동일한 비율로 사용하는 것을 볼 수 있다.

## PDDE와 LPay 데이터 내 연령대별 사용자 수 분포 비교



60대 이상은 전체 상품결제 사용자 수 비율보다 LPay를 사용하는 비율이 매우 적다.  
앞선 이용량 비교에서 20대의 이용량이 동일했으나 LPay의 사용자 비율은 적다.(10.6% → 9.2%)  
젊은층이 LPay 접근성이 높아 사용이 잦을 것으로 기대되는 것과 달리 사용자 수가 뒤쳐진다.

# EDA

## 연령대별 이용량 비율 차, 사용자 비율 차 비교

$$\begin{aligned}\text{이용량 비율 차} &= \text{LPay 이용량 비율} - \text{PDDE 이용량 비율} \\ \text{사용자 비율 차} &= \text{Lpay 사용자 비율} - \text{PDDE 사용자 비율}\end{aligned}$$

위의 지표를 통해 다음과 같이 해석할 수 있다.

이용량 비율 차 > 사용자 비율 차

=> LPay가 해당 연령층에게 저평가 되고 있거나 충성고객이 존재함

이용량 비율 차, 사용자 비율 차 > 0

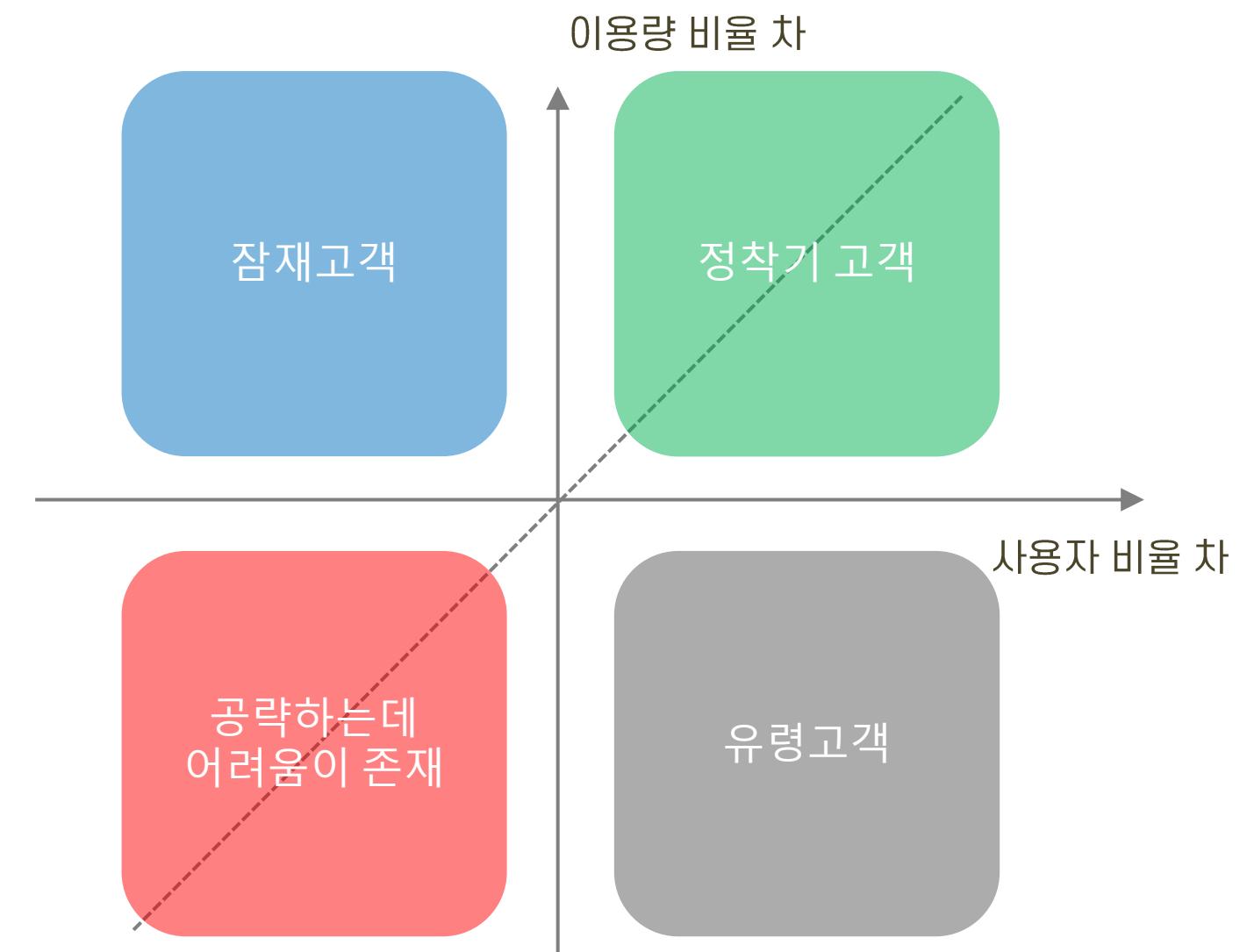
=> 해당 연령층의 Lpay 이용이 정착기에 있음

이용량 비율 차, 사용자 비율 차 < 0

=> 해당 연령층을 공략하는데 어려움이 존재

이용량 비율 차 > 0, 사용자 비율 차 < 0

=> 해당 연령층내 LPay의 잠재고객이 많이 존재함.

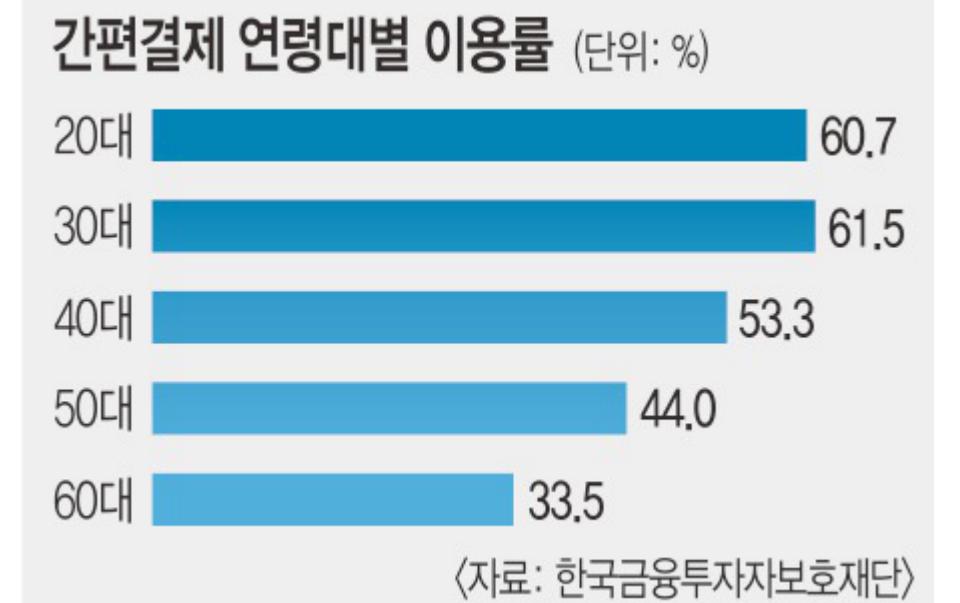
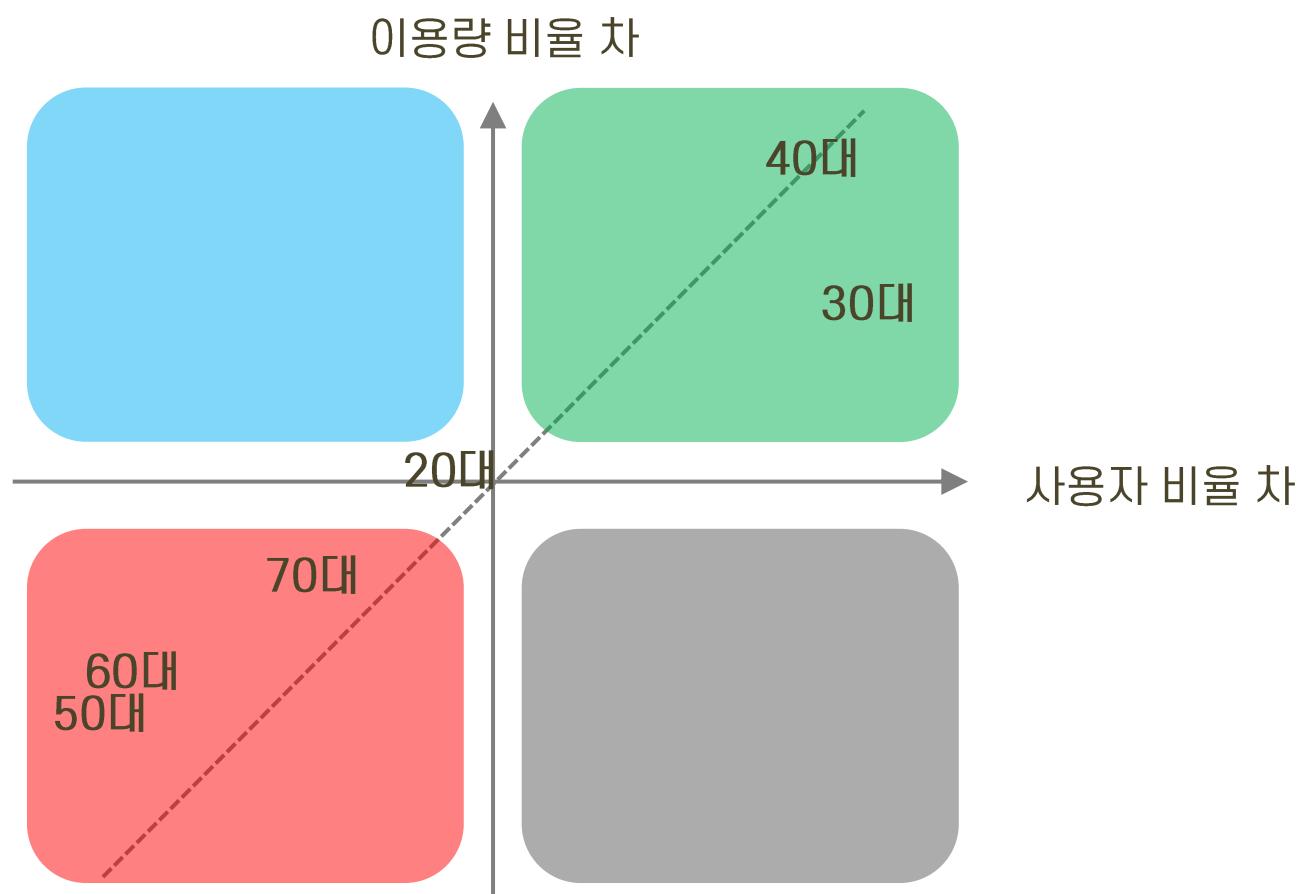


# EDA

## 연령대별 이용량 비율 차, 사용자 비율 차 비교

### 이용량 비율 차 vs 사용자 비율 차

- 20대 : 0 % > -1.4%
- 30대 : 3.7% < 7%
- 40대 : 7.5% = 7.5%
- 50대 : -4.7% > -5.5%
- 60대 : -4.6% > -5.4%
- 70대 : -1.9% > -2.1%



출처 : 2018년 한국금융투자자보호재단

## 이용량 비율 차 vs 사용자 비율 차

- 20대 : 0 % > -14% 주 고객층임과 동시에 충성고객이 존재할 것으로 판단된다.
- 30대 : 3.7% < 7%
- 40대 : 7.5% = 7.5% 잠재고객이 많을 것으로 생각되는 고객층은 20대로
- 50대 : -4.7% > -5.5% 전체 고객에서 큰 비중을 가지진 않았지만
- 60대 : -4.6% 사용자에 비해 이용량이 많은 일부 인원은 적극적으로 Lpay를 사용하며
- 70대 : -1.9% > -5.4% 타 간편결제 서비스를 많이 이용하는 고객층이므로 공략의 대상이 된다.

40대

30대

20대

사용자 비율 차

70대

60대

50대

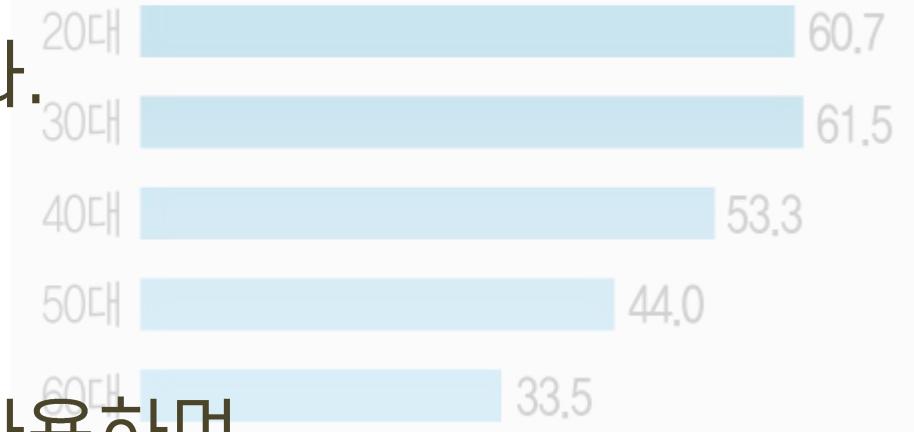
40대

30대

20대

10대

## 간편결제 연령대별 이용률 (단위: %)



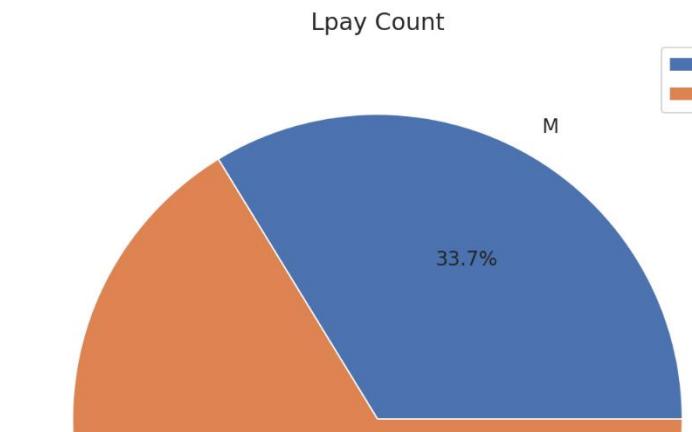
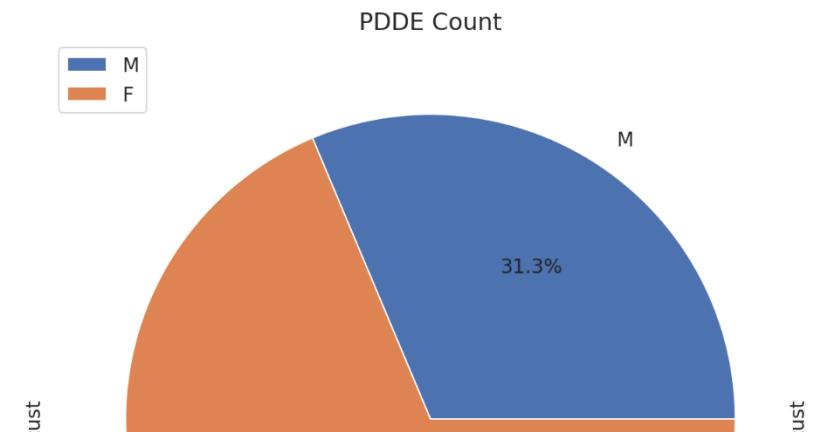
&lt;자료: 한국금융투자자보호재단&gt;

출처: 2018년 한국금융투자자보호재단

50대부터는 공략하기 어려운 연령층으로 보인다.  
다만, 50대는 큰 비중을 차지하는 고객층이므로  
이탈을 방지하는 방향으로 생각해볼만 하다.

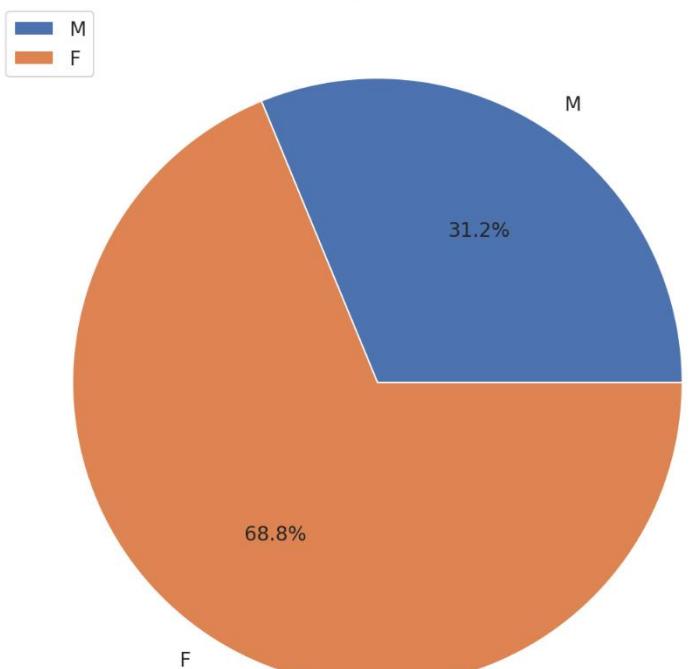
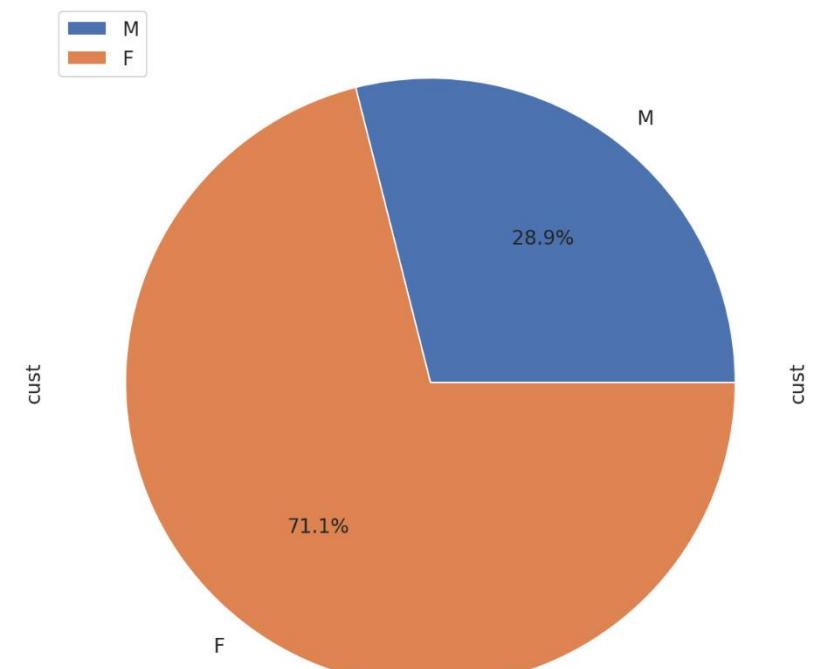
## PDDE와 LPay 데이터 내 성별에 따른 이용량(결제건수) 및 사용자 비교

이용량



이용량, 사용자 모두 Lpay에서 남성의 비율이 높으며 이용량 비율차이는 2.3%, 사용자 비율차이는 2.4%로 크게 다르지 않은 모습을 보였다.

사용자



비율차이가 모두 양수인 것으로 보아 남성은 여성에 비해 평균적으로 LPay에 잘 정착하는 것으로 해석할 수 있다.

20대 : MZ세대의 절대적인 구매총액은 적으나 구매액의 평균 값은 크게 뒤쳐지지 않는다.  
즉, 20대의 구매량이 늘어난다면 구매 총액이 크게 늘어날 것이다.

30대 : 이제 갓 사회에 진출한 사회초년생인 경우가 많다.  
구매총액이 두 번째로 큰 주 고객층으로 30대 사용자의 이탈 방지 및 소비 장려 전략이 필요하다.

40대 : 구매 총액이 가장 큰 주 고객층이다.  
특히 40대 여성층은 소비가 가장 많은 핵심 타겟층으로 이탈 방지 및 소비 장려 전략이 필요하다.

50대 : 공력이 어려울 것으로 생각되는 고객층이다.  
개인화 마케팅 전략을 시행하면서 배제되어 이탈하지 않도록 해야 한다.

# EDA

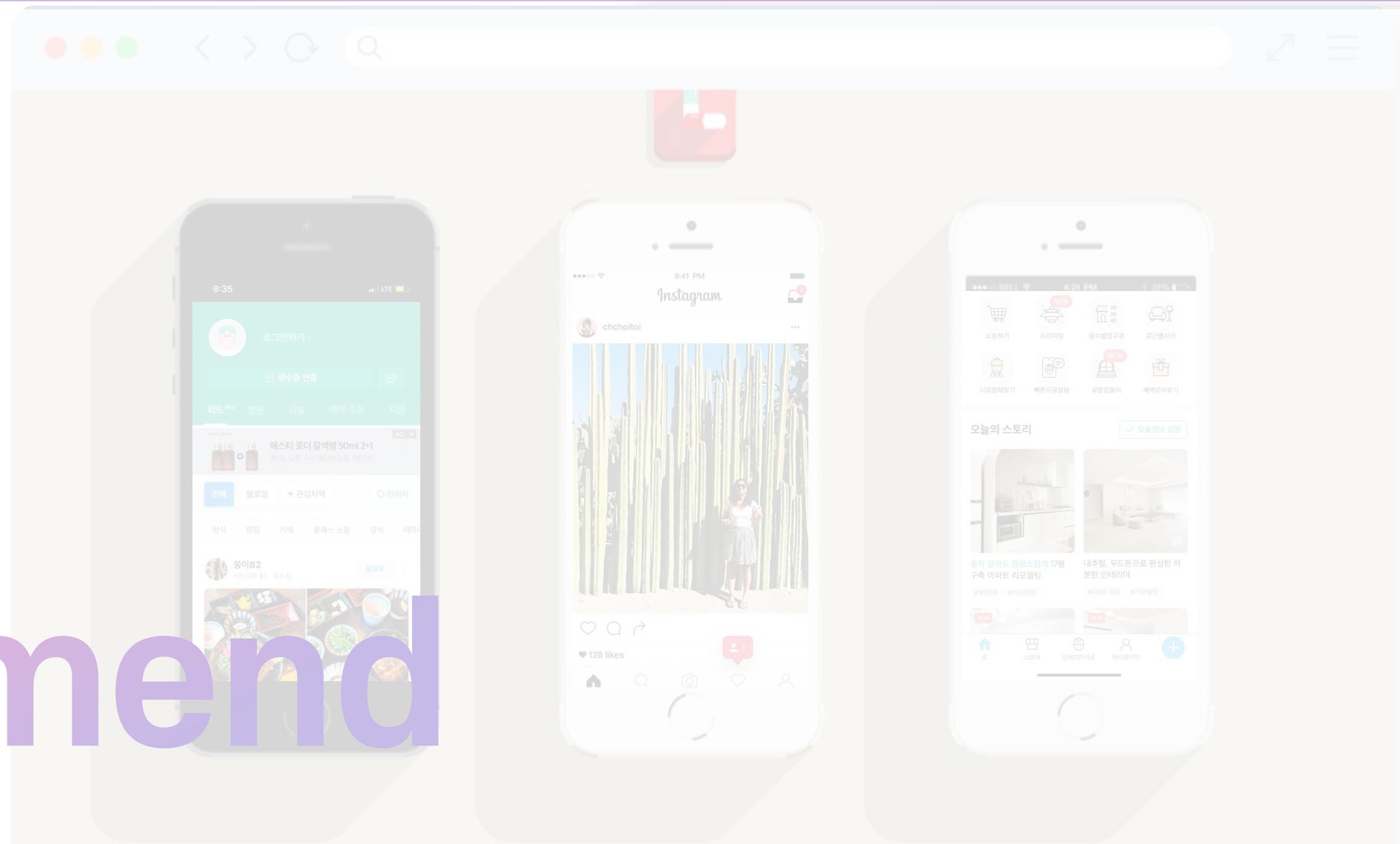
20대, 30대 : 인스타그램, 페이스북, 틱톡 등의 SNS 상에서 개인의 개성을 표현하는 세대  
40대, 50대 : 네이버, 다음 등의 포털사이트 카페 내 알뜰, 유용한 구매정보 공유하는 세대



두 세대 모두 정보를 경제적 재화로 취급하는 것을 볼 수 있으며  
영향력을 가진 이들의 정보를 신뢰함과 동시에 스스로 그렇게 되길 바라기도 한다.  
고객들은 정보를 통해 공유, 개성 표현, 재화의 재생산이 가능한 플랫폼에서 자주 활동하고 있다.

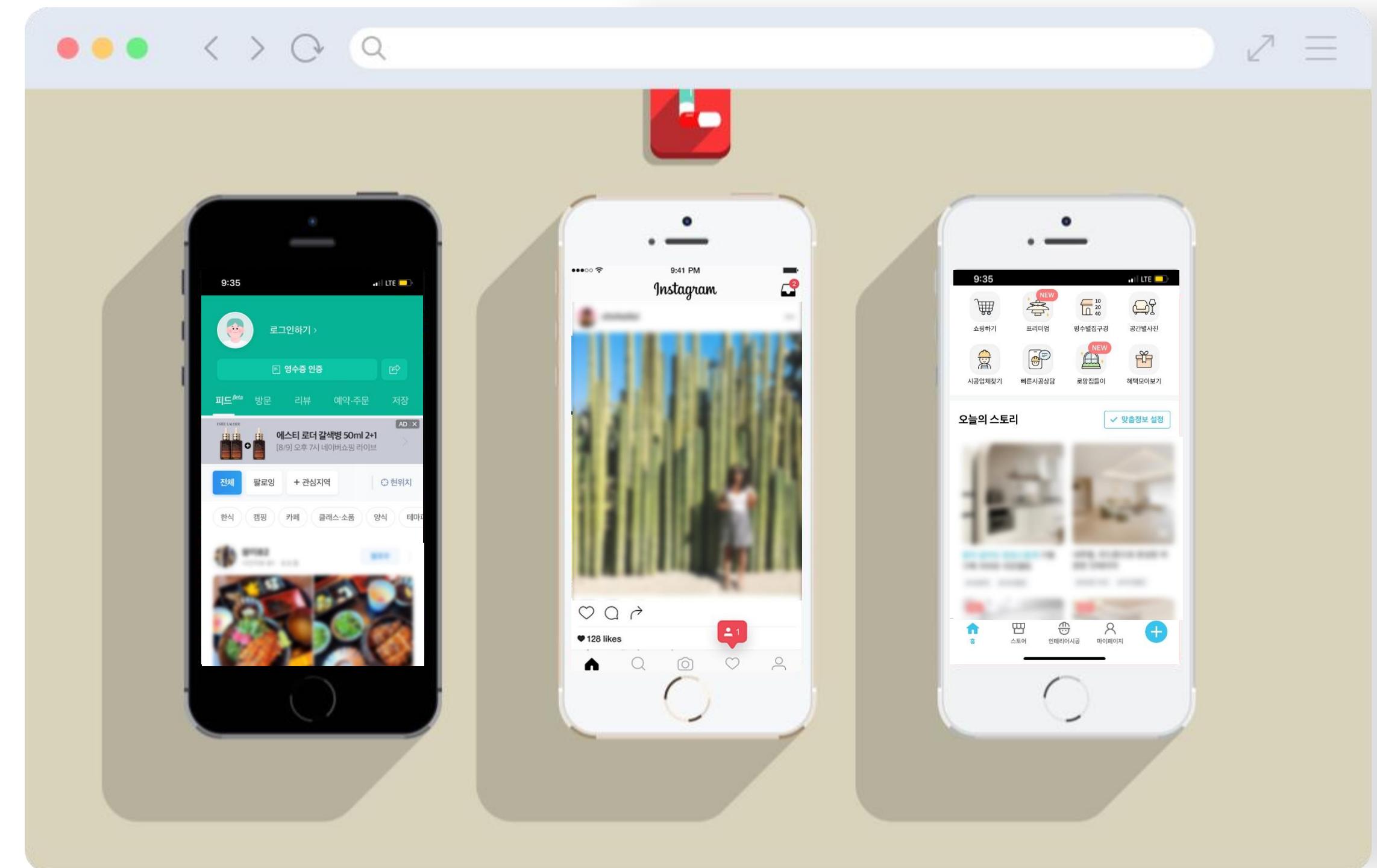
고객과 고객을 이어주는 마케팅 전략이 필요하다.

# Recommend System

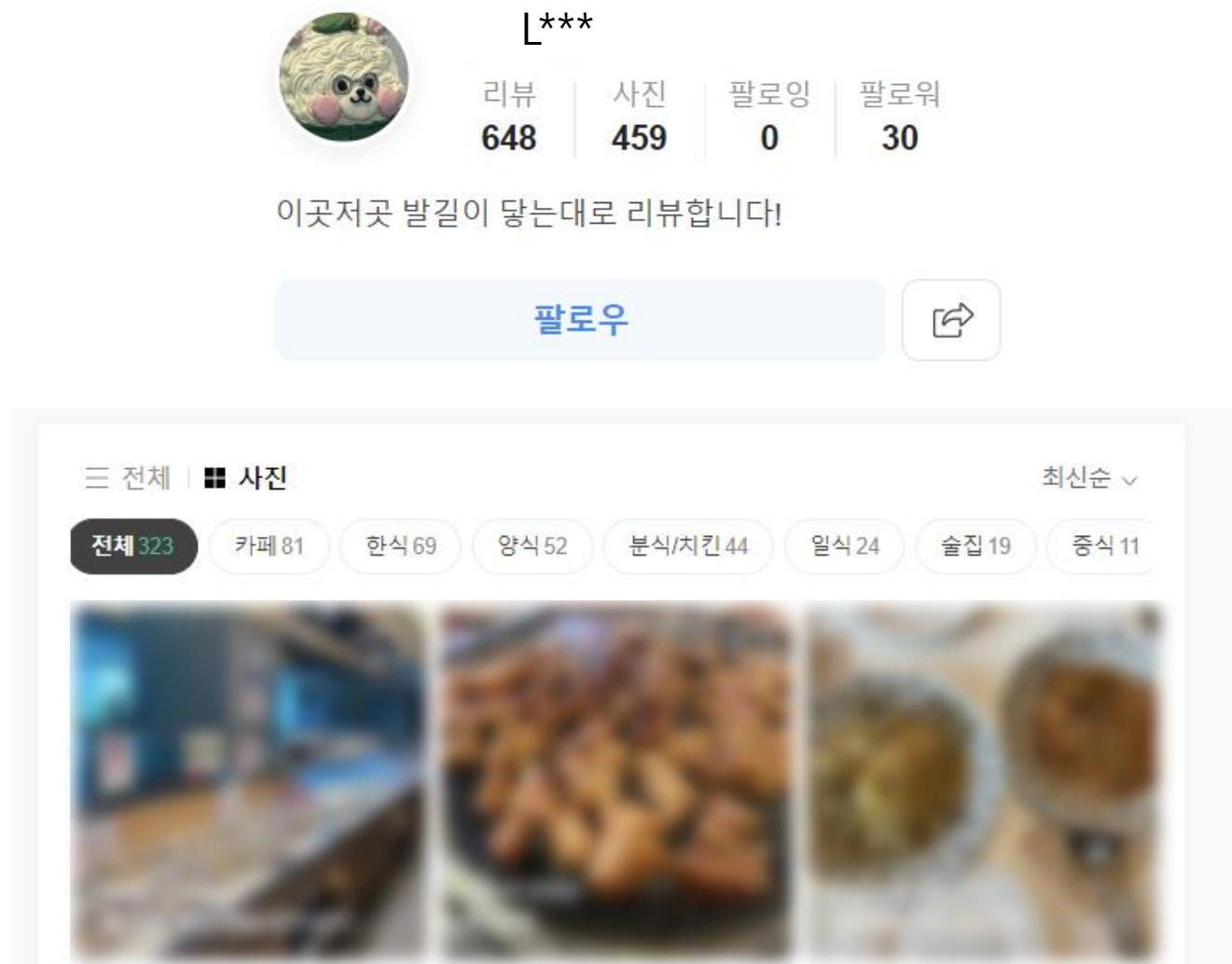


# Recommend System

- 네이버 리뷰시스템
- 인스타그램
- 오늘의 집



# Recommend System



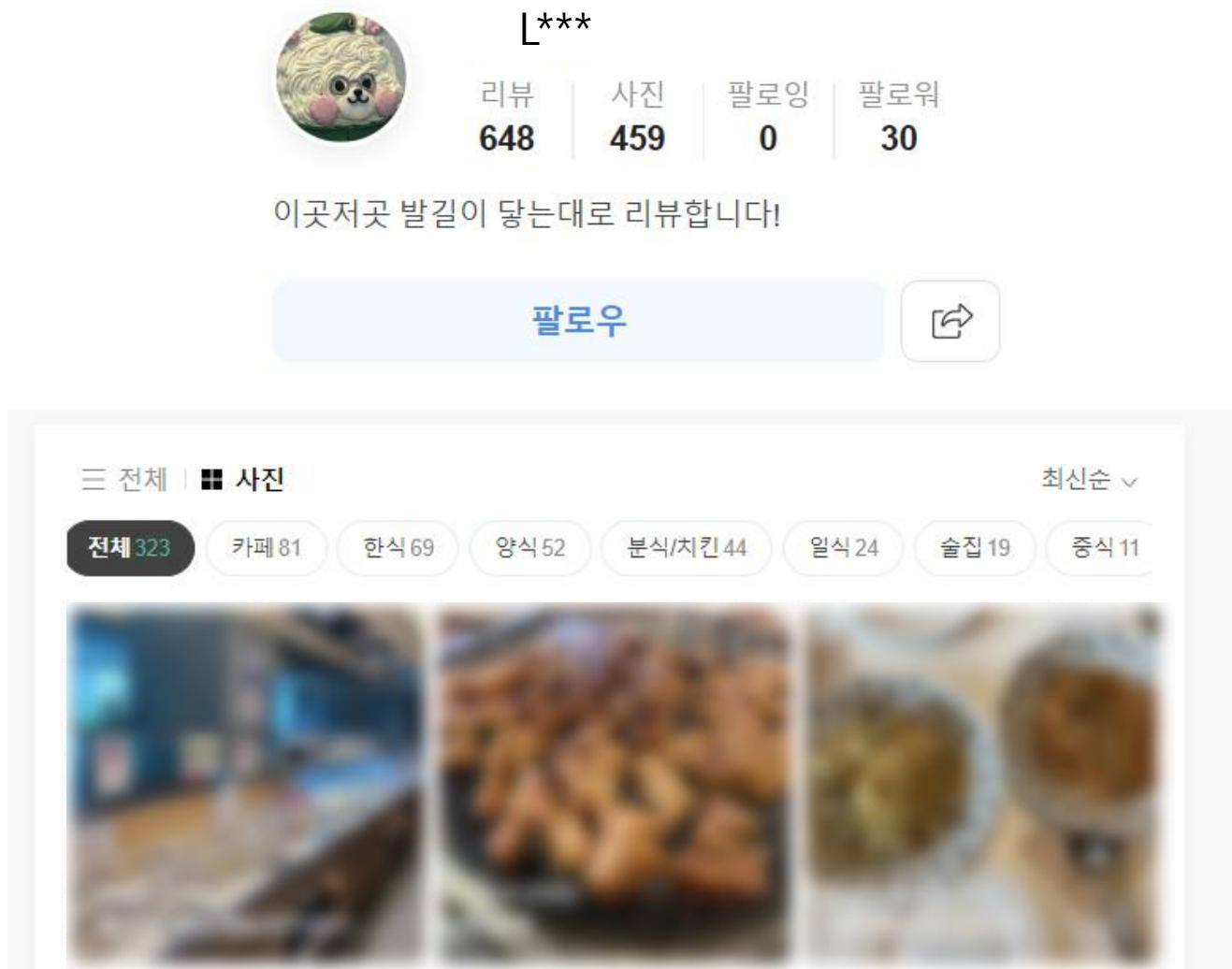
## N MY플레이스

### 네이버 MY플레이스 리뷰시스템

음식점과 장소에 대한 리뷰 시스템으로  
팔로워와 추천수 기능이 존재하며  
고객이 사진으로 결제 내역을 인증하고  
리뷰하면 리워드를 지급한다.

리뷰와 팔로잉 시스템을 통해 고객의 개성을  
표현하고 정보를 공유할 수 있는 공간이 된다.

# Recommend System



## N MY플레이스

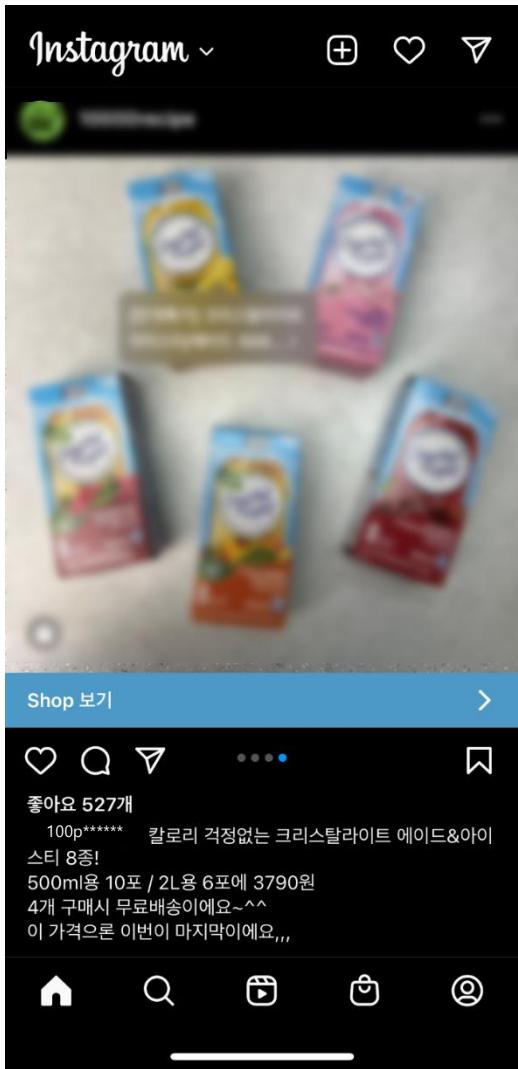
### 네이버 MY플레이스 리뷰시스템

LPay에 적용된다면 이용 범위가 확장될 수 있다. 상품 결제 데이터를 활용하여 특정 상품에 대한 정보 공유, 리뷰, 개성 표현 등이 가능한 시스템을 구축할 수 있다.

# Recommend System



## 인스타그램



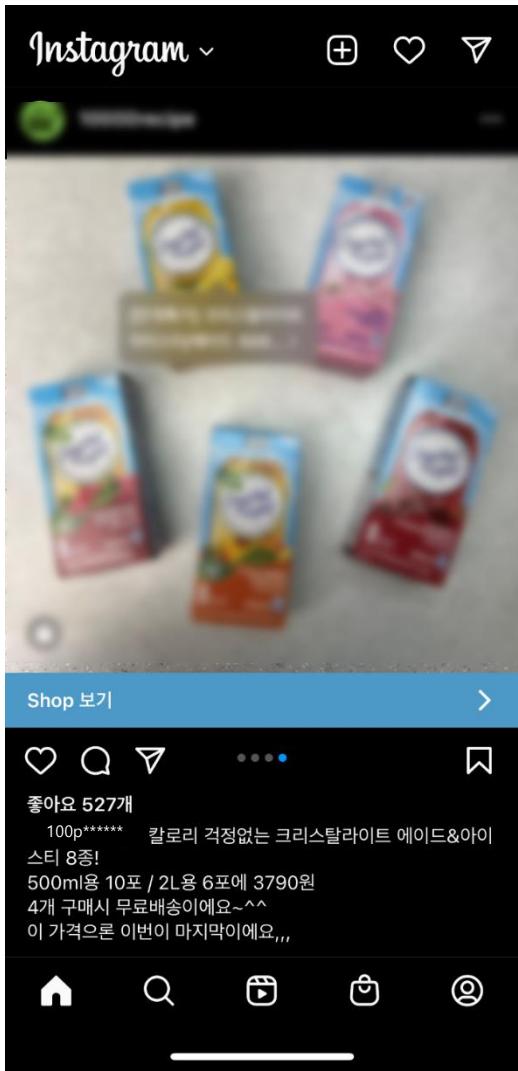
검색 기록을 바탕으로 고객에게 적합한 인플루언서를 추천하고 상품, 광고 피드를 노출한다.

고객은 컨텐츠의 소비자이자 제공자가 되어 활발히 컨텐츠를 생산한다. 좋은 컨텐츠를 제공하는 고객은 인플루언서로써 광고를 통해 수익을 창출할 수 있다.

# Recommend System



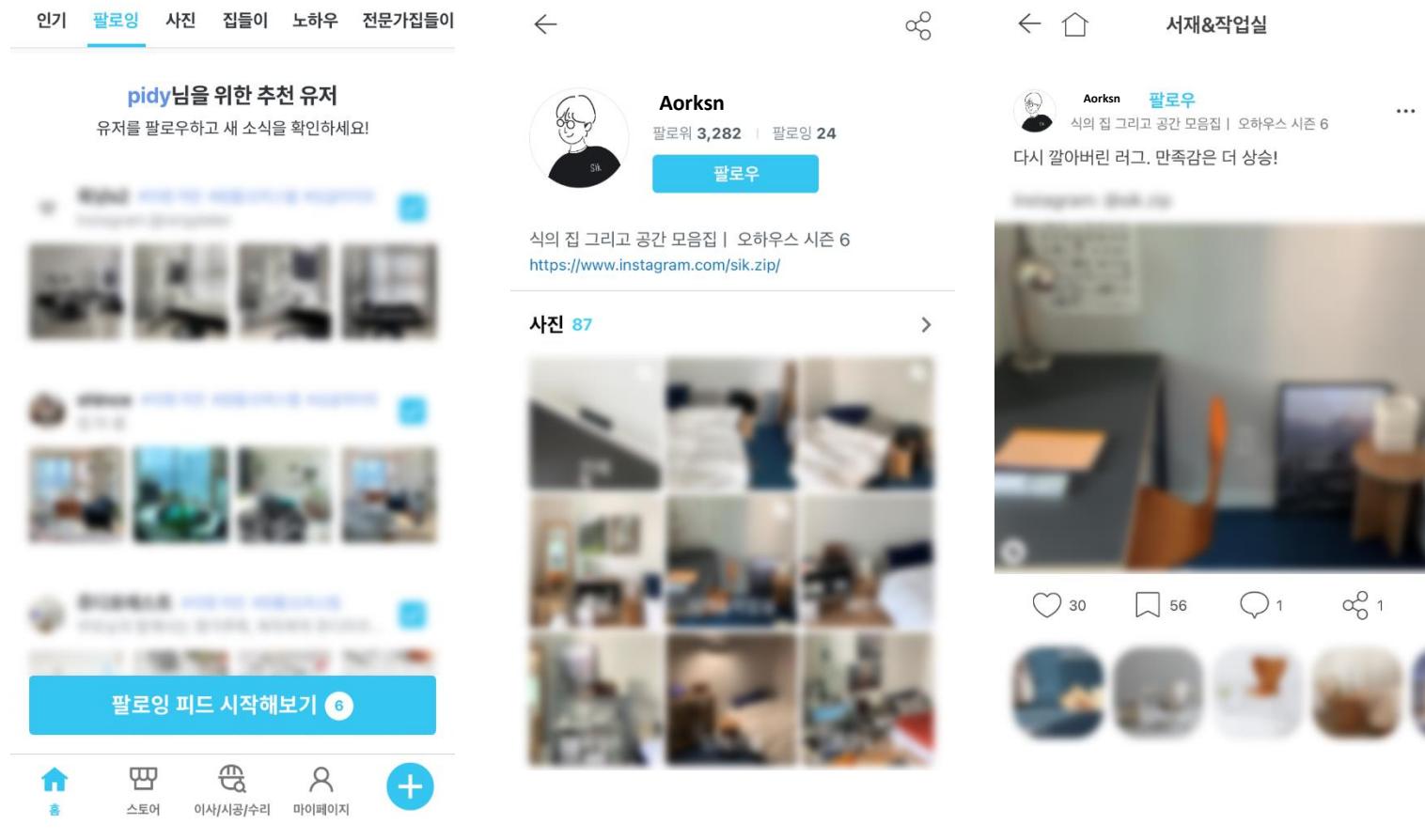
## 인스타그램



LPay의 영수증을 하나의 컨텐츠로 다룰 수 있고, 많은 사람들이 구매할 만한 영수증을 만든 고객에게 리워드가 돌아갈 수 있다면 LPay로 결제할 매력적인 이유가 될 것이다.

또한, 인스타그램에 비해 리워드를 얻기까지의 진입장벽이 낮아 많은 사람들이 이용할 만한 충분한 동기가 제공될 것이다.

# Recommend System

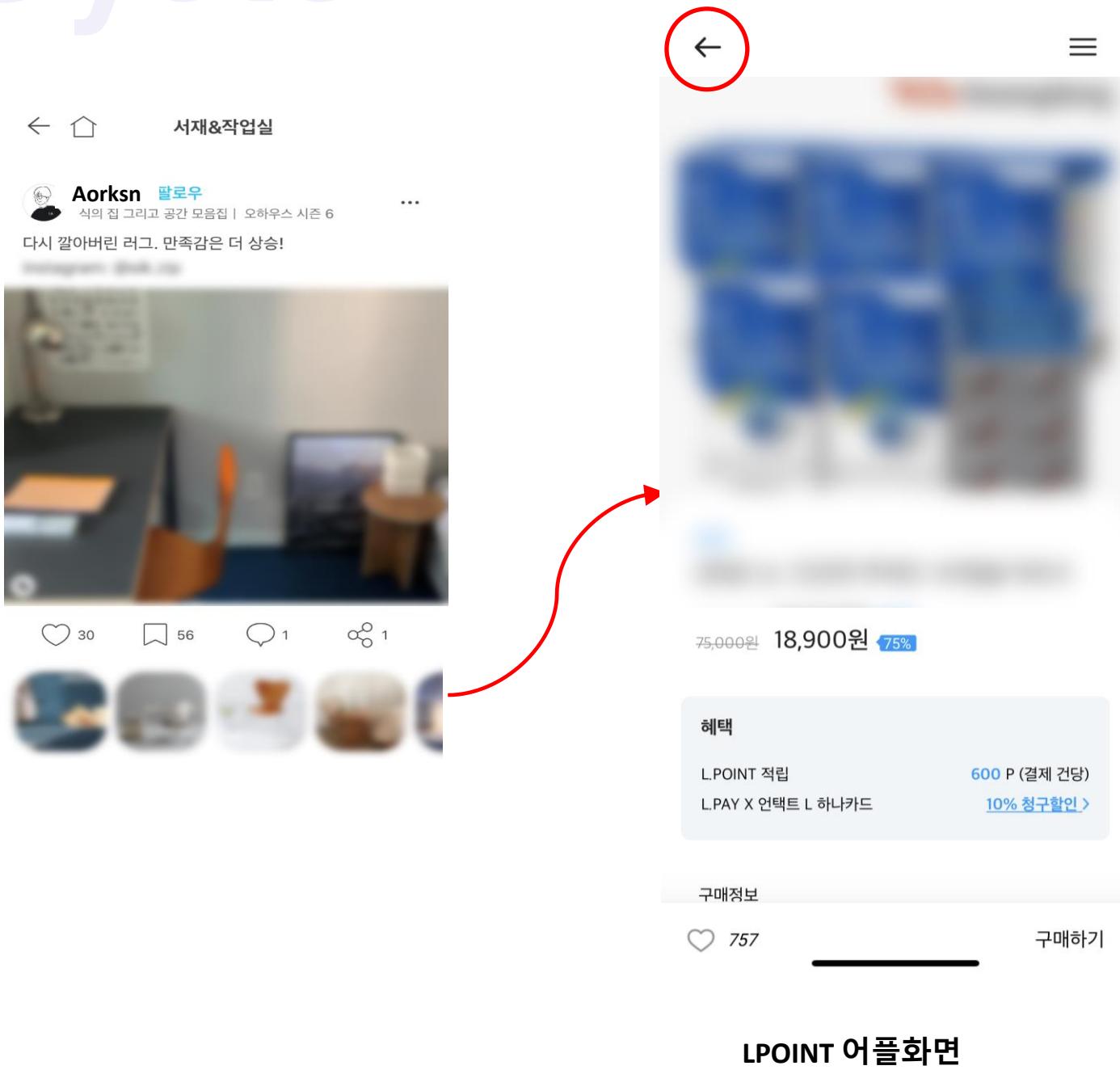


## 오늘의 집

구매한 가구나 생활용품을 통해 개성을 표현할 수 있는 구매 정보 공유 플랫폼이 존재한다.

게시물은 고객에게 상품을 인지시키고 흥미를 유발할 수 있다. 특히, 하단의 구매 목록은 해당 상품들의 링크로 연결되어 구매로 이어질 확률이 높아진다.

# Recommend System



PART I



## 오늘의 집

LPay 어플에도 동일한 시스템을 적용시킬 수 있다면 다른 고객의 게시물이 고객들의 구매 심리를 자극하여 상품 결제율을 높일 것이다.

LPay 어플 내 사용자의 페이지 방문 기록이 존재하므로 게시물의 링크로 접속하여 구매한 경우 게시자와 구매자에게 모두 보상이 지급되는 시스템을 구현할 수 있다.

고객과 고객을 연결하는  
개인화 마케팅 C2C플랫폼 :

# L-Stagram

영수증도 컨텐츠가 될 수 있다

고객은 영수증을 게시함으로써 구매정보를 공유함과 동시에 개인의 개성을 표현할 수 있다. 게시된 영수증은 다른 고객들에게 하나의 컨텐츠로써 소비되며 흥미를 유발하고 구매심리를 자극할 수 있다.

단, 영수증 게시는 고객의 선택이다.



고객과 고객을 연결하는  
개인화 마케팅 C2C플랫폼 :

# L-Stagram

## 양방향 리워드 제도

게시된 영수증 페이지에서 상품 링크로 구매 시  
영수증 게시자와 구매한 고객 모두에게 소정의 리워드를 지급한다.  
LPay로 결제하고 영수증을 게시하면 누구나 리워드를 얻을 수 있어  
수익 창출까지의 진입장벽이 낮기 때문에 적은 리워드로도 고객들은 적극적으로  
영수증을 게시할 것이다.



고객과 고객을 연결하는  
개인화 마케팅 C2C플랫폼 :

# L-Stagram

## 고객을 연결하는 추천 시스템

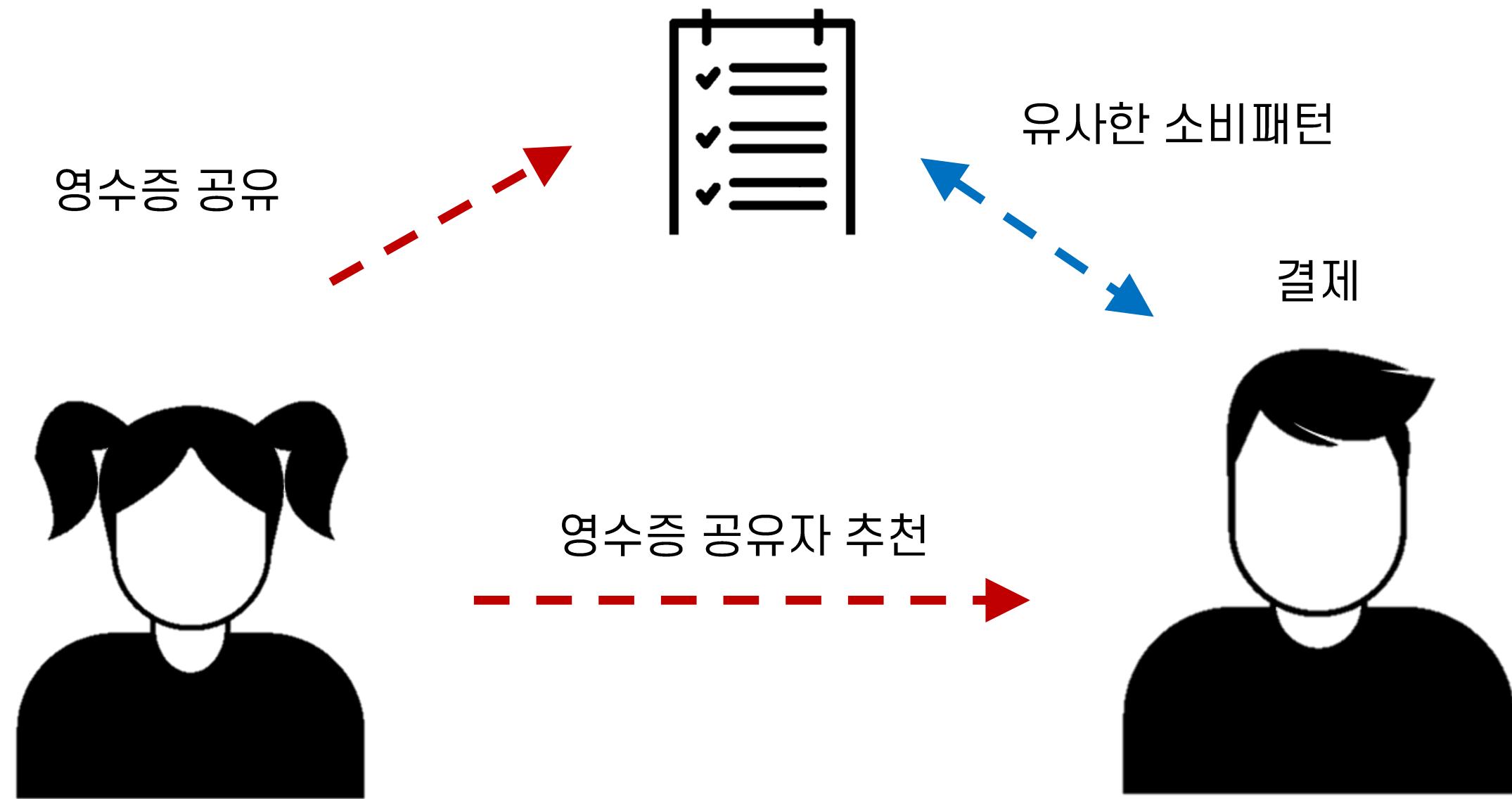
사용자가 실제로 소비할 영수증을 정확히 추천해준다면  
게시된 영수증 페이지에서의 소비가 늘어날 것이다. 이를 통해 게시자에게 리워드가  
돌아가면서 게시자의 수익구조가 개선된다면 영수증 게시활동이 활성화 되어  
전반적인 LPay 사용률이 높아질 것이다.



# PART II

## Recommend System

- Preprocess
- Modeling
- Performance



# PART II

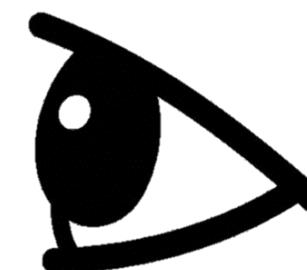
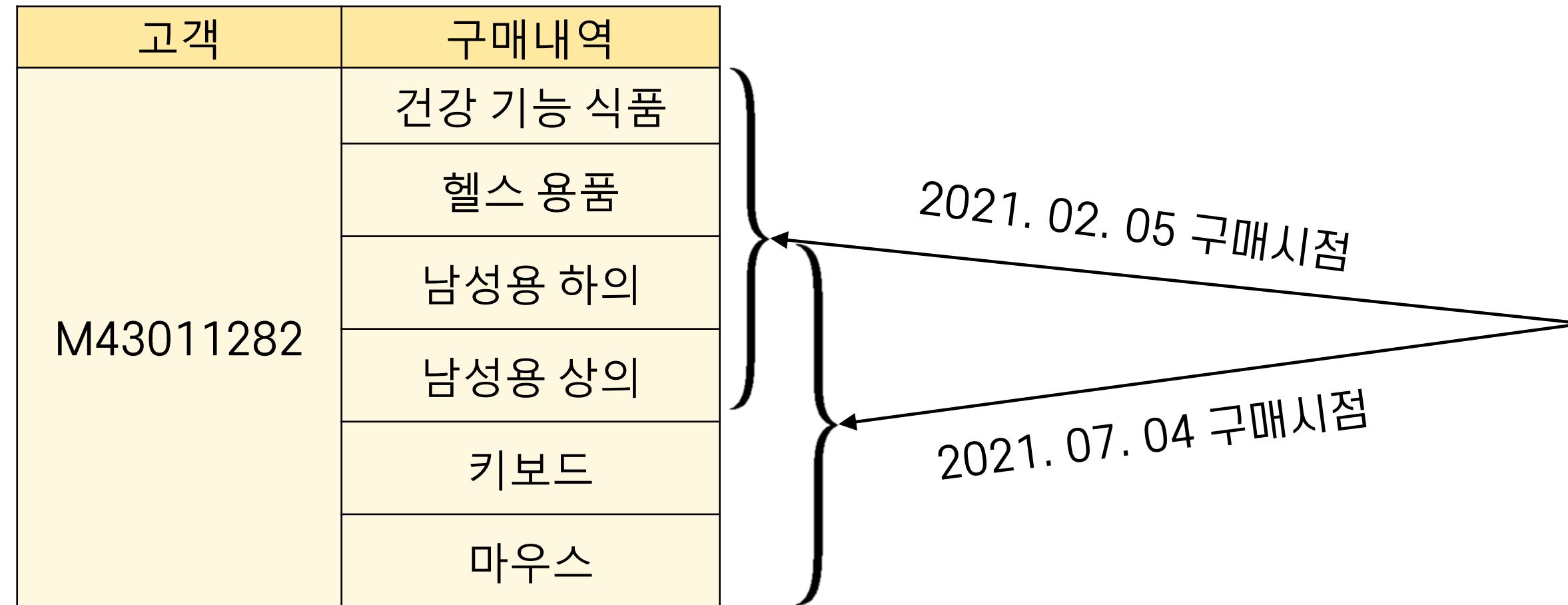


# Preprocess



고객 정보, 상품정보, 판매정보 데이터에 pivot table을 사용하여 오른쪽과 같이 변환한다.

## Preprocess



사람들의 관심사는 변화하기 때문에 2021년 2월에 운동에 관심있던 사람이  
2021년 7월에 전자기기에 관심을 가질 수 있다.

## 고객 “A01000001113”의 영수증 묶음

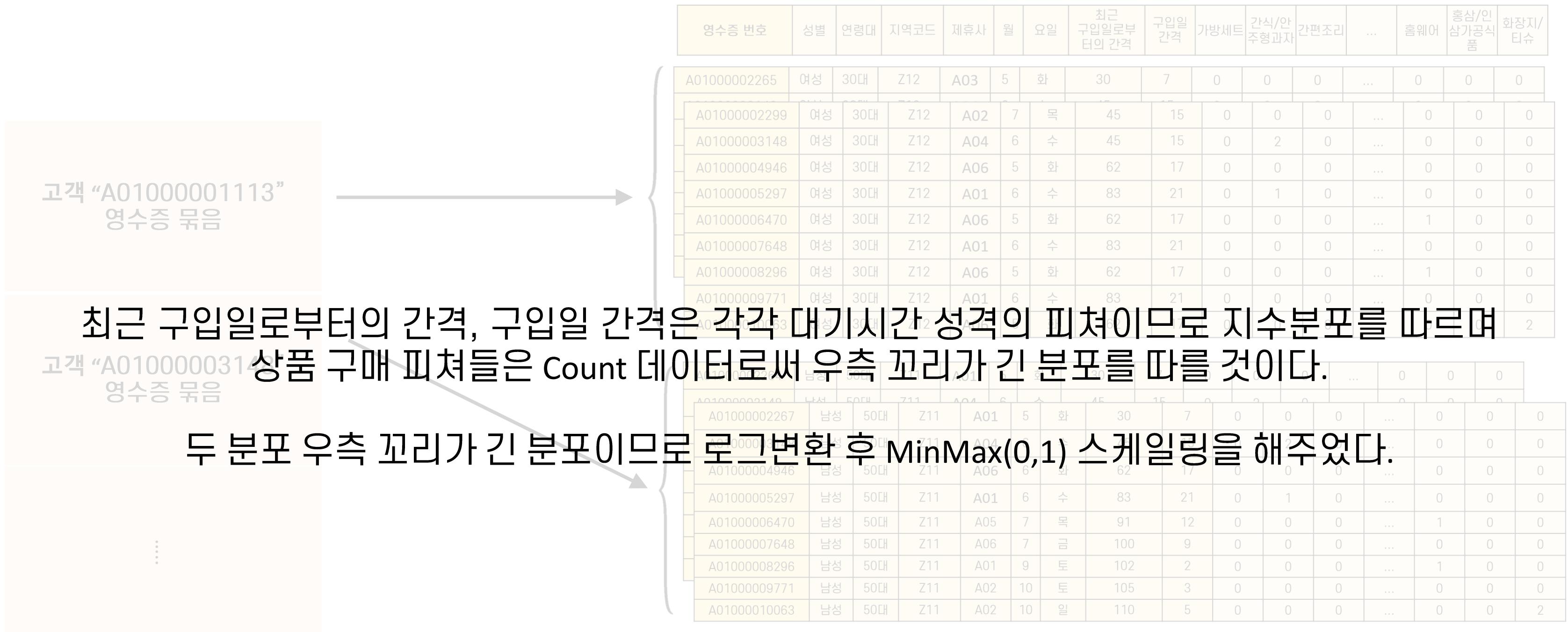
영수증 번호	성별	연령대	지역 코드	...	간식/안주 형과자	간편조리	...	홈웨어	홍삼/인삼 가공식품	화장지/EI 슈
A01000001113	남성	30대	Z11	...	0	0	...	0	0	0
A01000001113	남성	30대	Z11	...	0	0	...	0	0	0
A01000001113	남성	30대	Z11	...	2	0	...	0	0	0
A01000001113	남성	30대	Z11	...	0	0	...	0	0	0
A01000001113	남성	30대	Z11	...	1	0	...	0	0	0
A01000001113	남성	30대	Z11	...	0	0	...	1	0	0
A01000001113	남성	30대	Z11	...	0	0	...	0	0	0
A01000001113	남성	30대	Z11	...	0	0	...	1	0	0
A01000001113	남성	30대	Z11	...	0	0	...	0	0	0
A01000001113	남성	30대	Z11	...	0	0	...	0	0	2
A01000001113	남성	30대	Z11	...	0	0	...	0	0	0
A01000001113	남성	30대	Z11	...	0	0	...	0	0	0
A01000001113	남성	30대	Z11	...	2	0	...	0	0	0
A01000001113	남성	30대	Z11	...	0	0	...	0	0	0
A01000001113	남성	30대	Z11	...	1	0	...	0	0	0

또한 한 고객의 구매기록 내에서도 고객의 특성은 구매시점에 따라 다를 수 있기 때문에,  
정렬된 영수증을 5개씩 겹쳐가며 10개씩 묶는다.

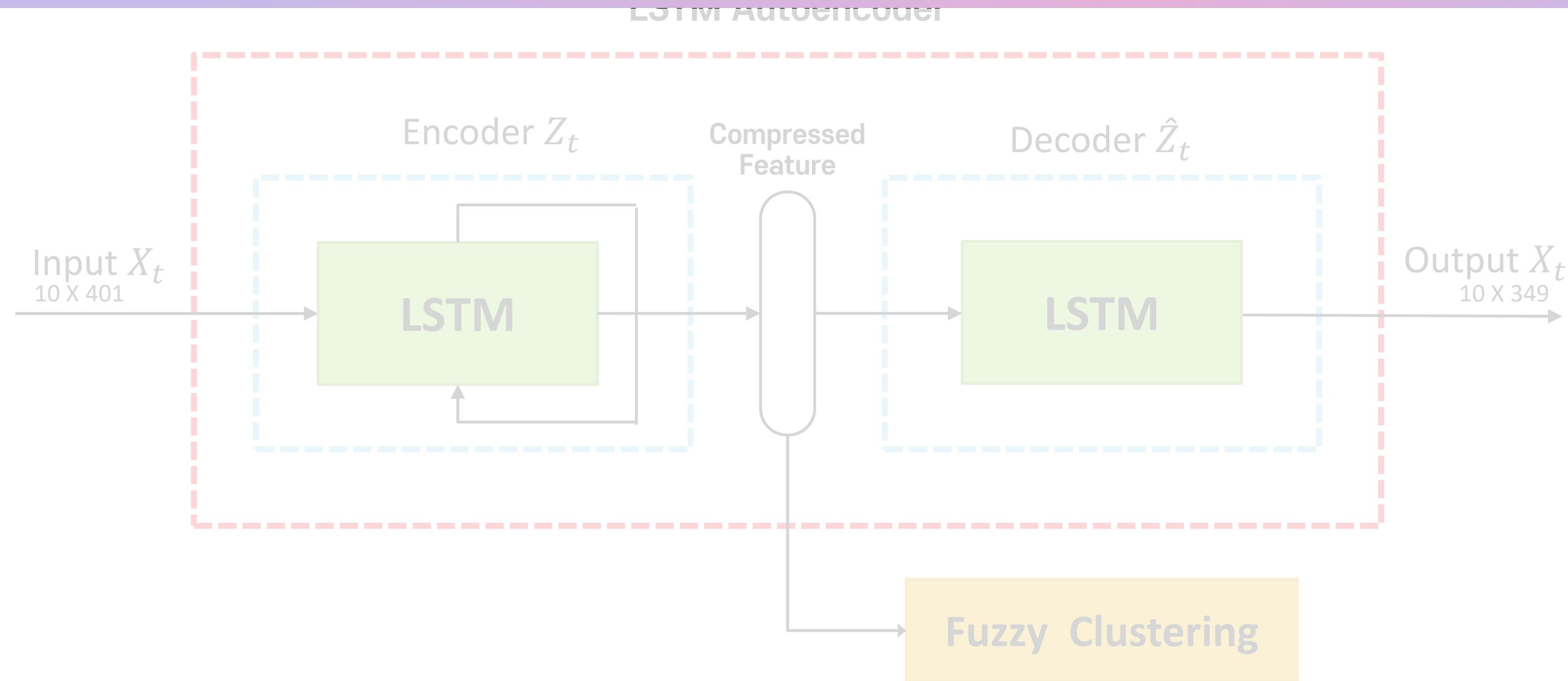


영수증 번호	성별	연령대	지역코드	제휴사	월	요일	최근 구입일로부터의 간격	구입일 간격	가방세트	간식/안주형과자	간편조리	...	홈웨어	홍삼/인삼가공식품	화장지/티슈
A01000002265	여성	30대	Z12	A03	5	화	30	7	0	0	0	...	0	0	0
A01000002299	여성	30대	Z12	A02	7	목	45	15	0	0	0	...	0	0	0
A01000003148	여성	30대	Z12	A04	6	수	45	15	0	2	0	...	0	0	0
A01000004946	여성	30대	Z12	A06	5	화	62	17	0	0	0	...	0	0	0
A01000005297	여성	30대	Z12	A01	6	수	83	21	0	1	0	...	0	0	0
A01000006470	여성	30대	Z12	A06	5	화	62	17	0	0	0	...	1	0	0
A01000007648	여성	30대	Z12	A01	6	수	83	21	0	0	0	...	0	0	0
A01000008296	여성	30대	Z12	A06	5	화	62	17	0	0	0	...	1	0	0
A01000009771	여성	30대	Z12	A01	6	수	83	21	0	0	0	...	0	0	0
A01000010063	여성	30대	Z12	A06	5	화	62	17	0	0	0	...	0	0	2
A01000002264	남성	50대	Z11	A01	5	화	30	7	0	0	0	...	0	0	0
A01000002267	남성	50대	Z11	A01	5	화	30	7	0	0	0	...	0	0	0
A01000003148	남성	50대	Z11	A04	6	수	45	15	0	2	0	...	0	0	0
A01000004946	남성	50대	Z11	A06	6	화	62	17	0	0	0	...	0	0	0
A01000005297	남성	50대	Z11	A01	6	수	83	21	0	1	0	...	0	0	0
A01000006470	남성	50대	Z11	A05	7	목	91	12	0	0	0	...	1	0	0
A01000007648	남성	50대	Z11	A06	7	금	100	9	0	0	0	...	0	0	0
A01000008296	남성	50대	Z11	A01	9	토	102	2	0	0	0	...	1	0	0
A01000009771	남성	50대	Z11	A02	10	토	105	3	0	0	0	...	0	0	0
A01000010063	남성	50대	Z11	A02	10	일	110	5	0	0	0	...	0	0	2

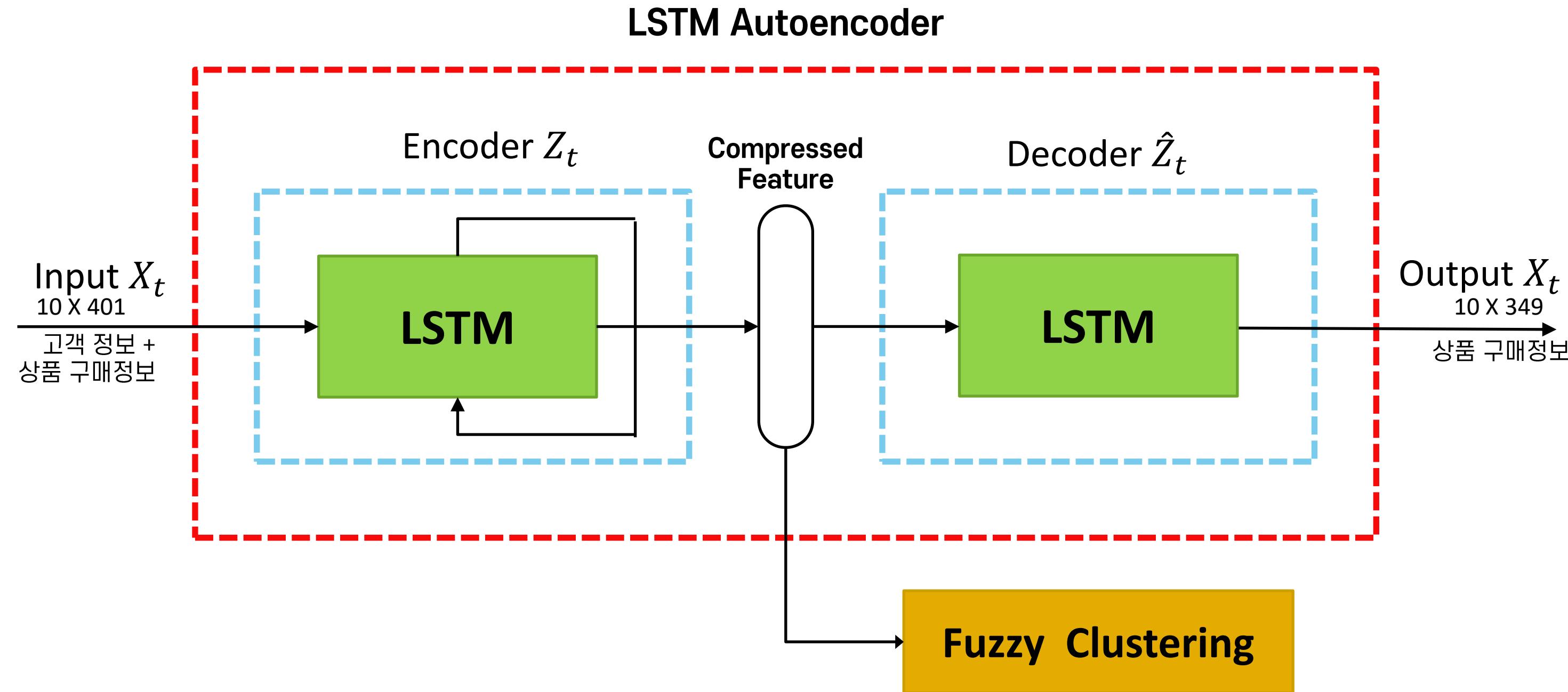
성별, 연령대, 지역코드는 One-Hot 인코딩을 실시하였다.  
최종 데이터 shape : (112953, 10, 401)



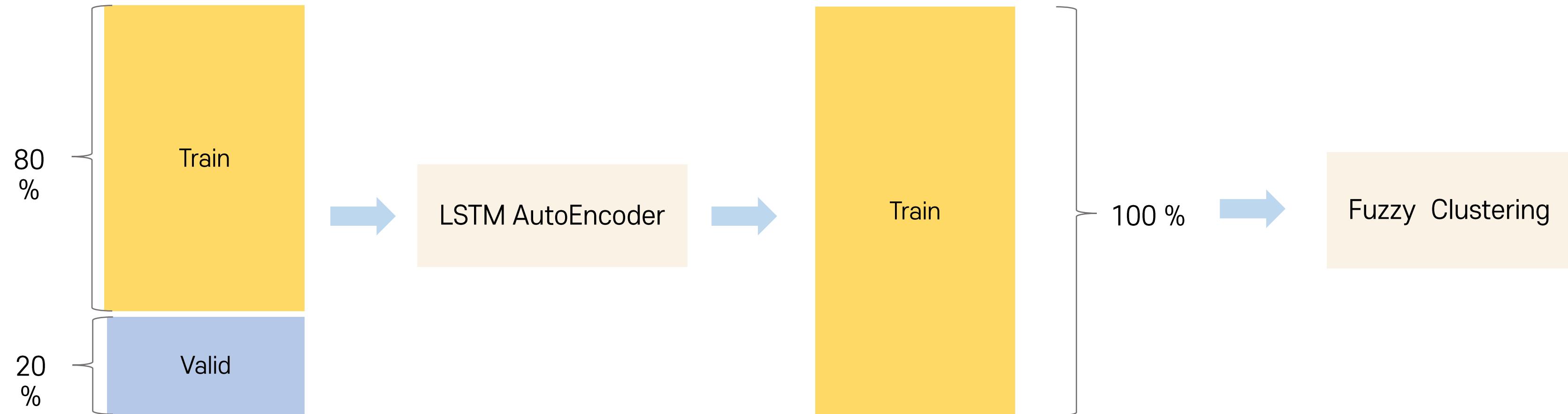
## Preprocess



# Modeling

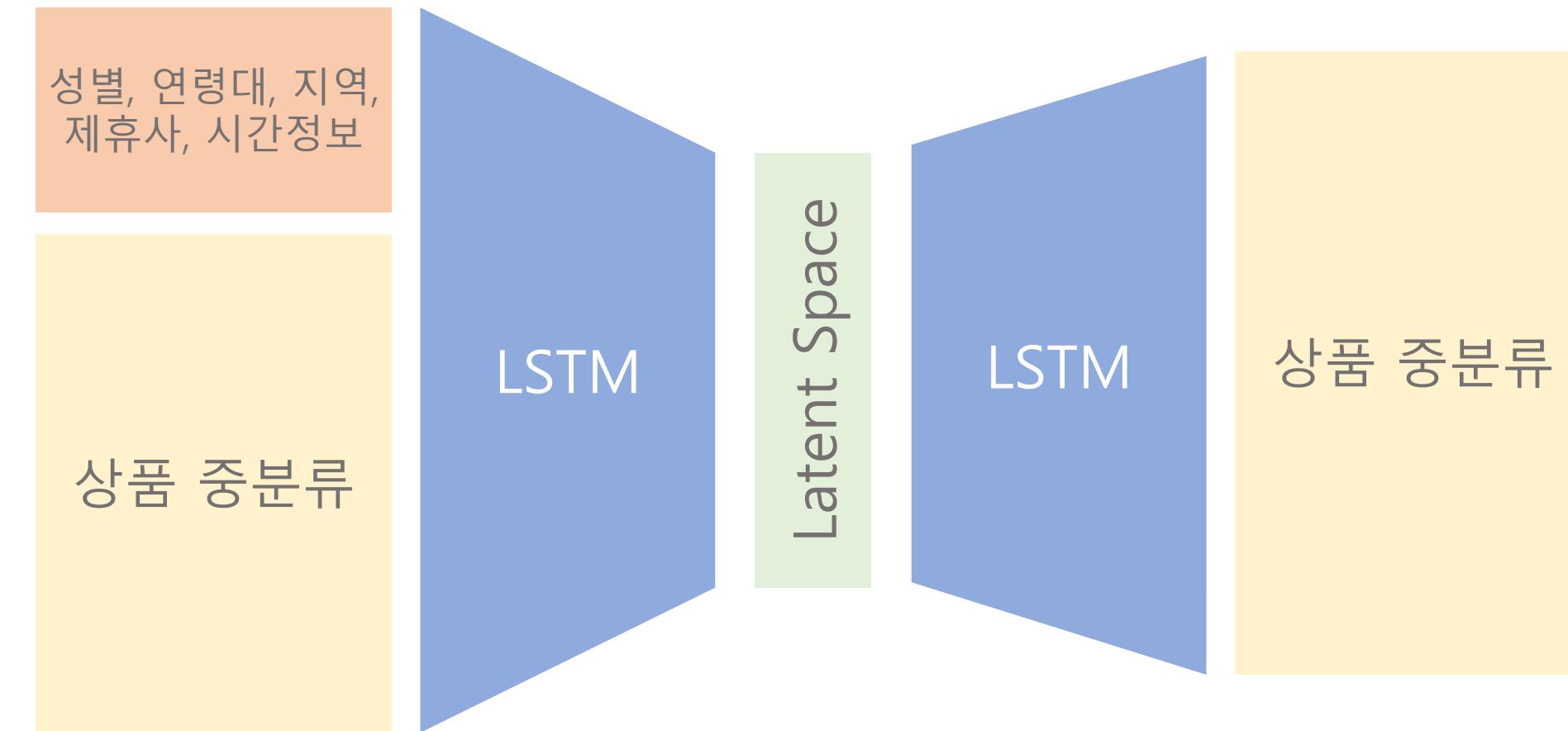


LSTM AutoEncoder를 통해 Latent Variable 추출 이후 Fuzzy Clustering 진행



전체 데이터의 80% TrainSet으로 사용하여 LSTM AutoEncoder를 학습시킨다.  
이후 전체 데이터를 Encoder에 입력하여 Latent Variable을 추출하여 Fuzzy Clustering을 실시하고  
Fuzzy Clustering으로부터 나온 확률로 확률분포 피쳐를 구한다.

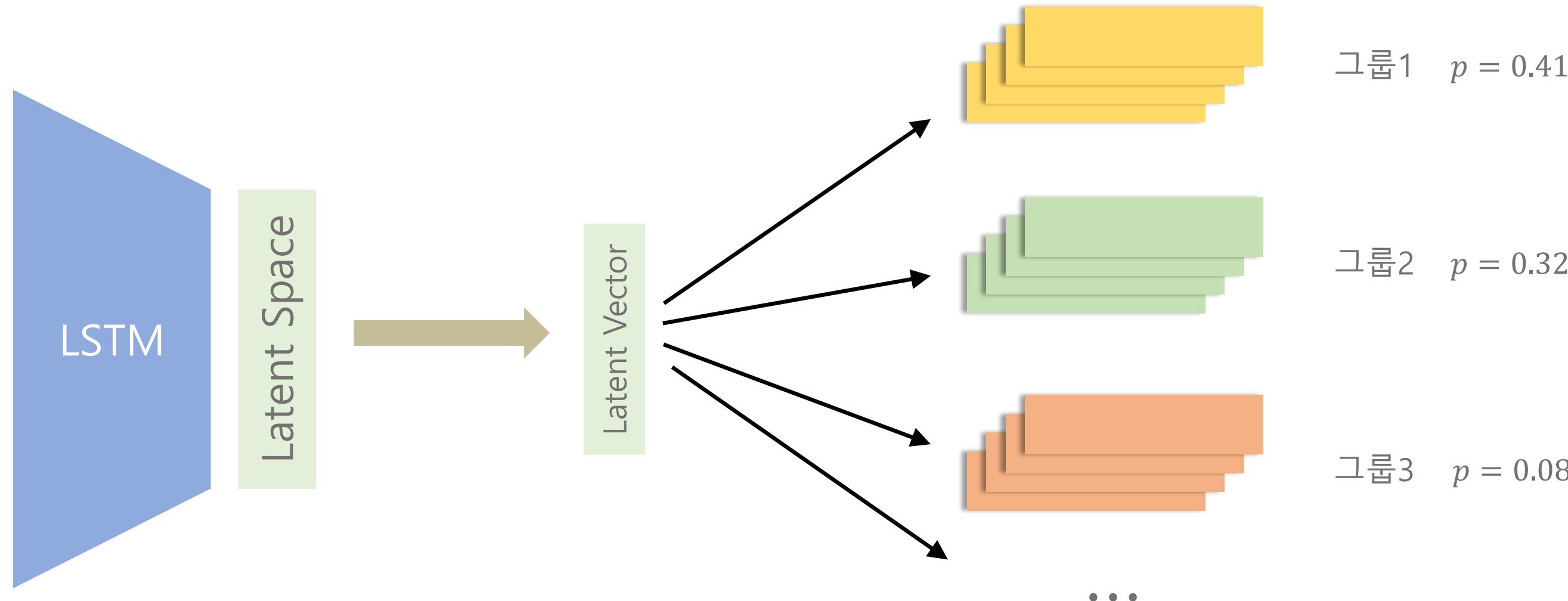
## LSTM AutoEncoder



고객의 정보와 영수증 묶음으로부터 유사한 영수증 묶음을 찾아내기 위해  
Latent Space는 고객정보, 영수증 묶음으로 영수증 묶음을 재구성하는 Feature의 벡터공간으로 학습될 것이다.  
인코더는 상품 중분류에 대한 정보와 성별, 연령대, 지역, 제휴사, 시간정보를 함께 포함하고  
디코더는 상품 중분류만 포함한다.

영수증 끊음	Feature_0	Feature_1	Feature_2	Feature_3	Feature_4
1	-0.003499	-0.097034	0.132439	-0.097034	-0.153275
2	-0.161823	-0.124157	0.189549	-0.124157	-0.126616
3	-0.152364	-0.111995	0.171075	-0.111995	0.120717
4	-0.153275	-0.116798	0.176958	-0.116798	-0.238949
5	-0.126616	-0.038817	0.008592	-0.038817	-0.152364
6	0.120717	-0.219868	-0.379981	-0.219868	-0.153275
7	-0.238949	-0.348969	0.163122	-0.348969	-0.003499

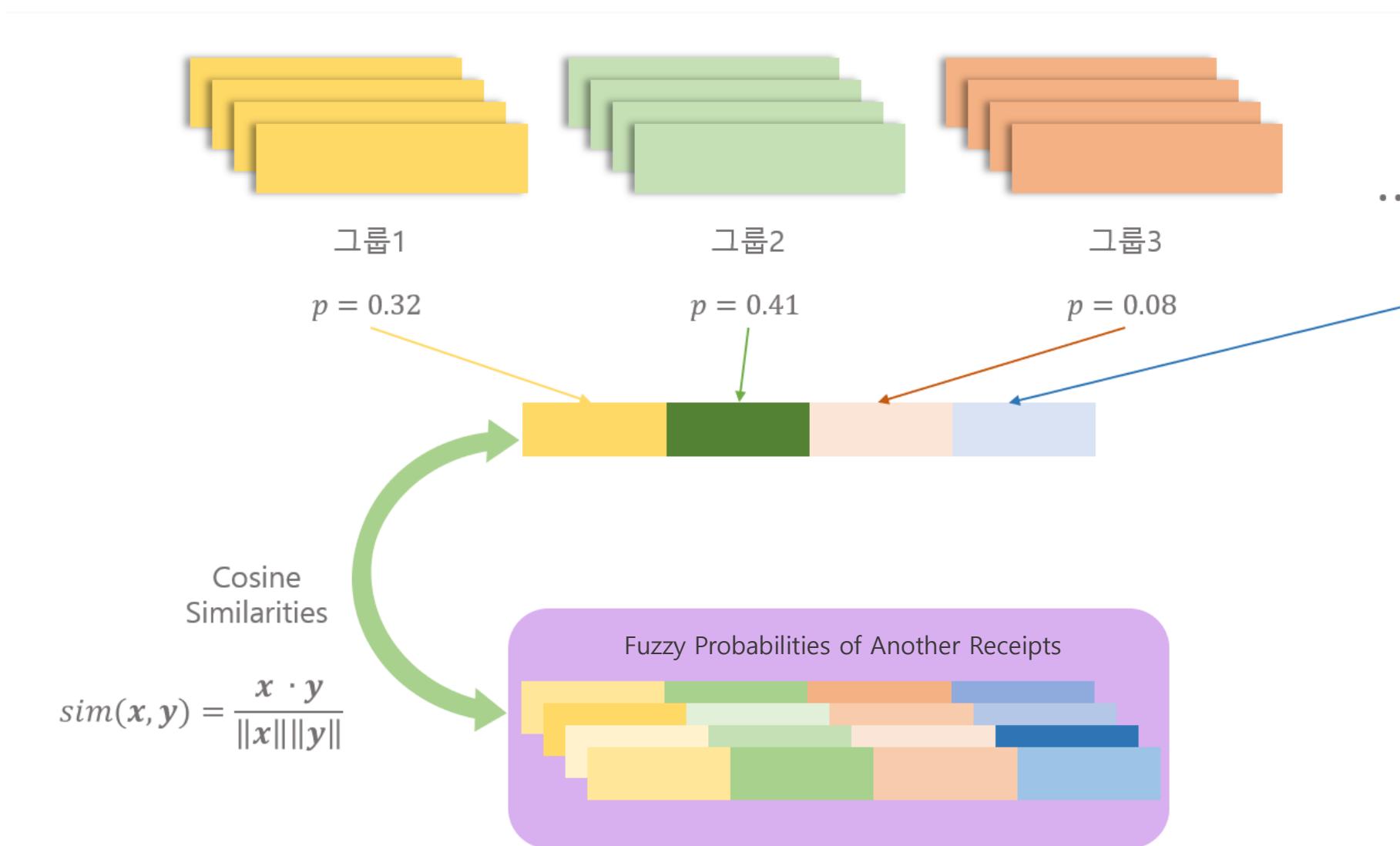
Encoder를 통해 얻어진 Latent Vector는 Clustering에 적합한 실수 값으로 변환된다.



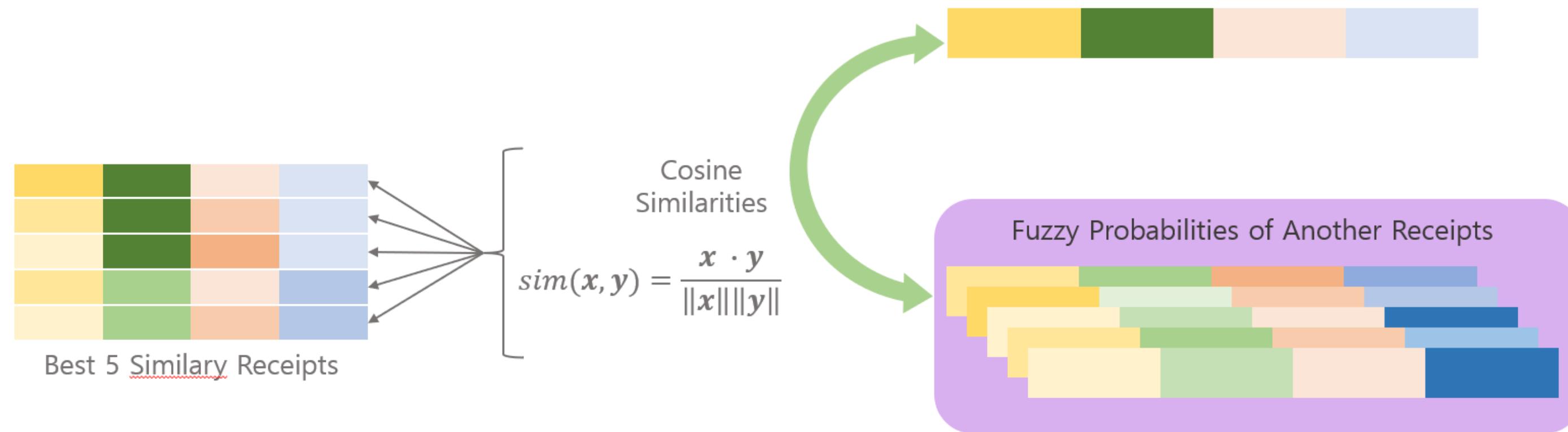
고객의 영수증 끓음의 특성은 특정 그룹에만 속하지 않을 수 있다.  
이러한 점을 군집화 기법에 반영하기 위해 Fuzzy 군집화를 하였다.  
Latent Variable들은 Fuzzy 군집화를 통해 특정 그룹에 속할 확률들을 얻는다.  
이 때 확률들은 영수증 끓음의 그룹에 대한 확률 분포이다.

영수증 묶음	그룹0_확률	그룹1_확률	그룹2_확률	그룹3_확률	...	그룹17_확률	그룹18_확률	그룹19_확률
1	0.04275	0.036111	0.003493	0.174078	...	0.059129	0.036329	0.029503
2	0.134232	0.85214	0.000414	0.004681	...	0.046461	0.014976	0.040607
3	0.073017	0.122784	0.315826	0.008248	...	0.601926	0.062129	0.037389
4	0.063566	0.96612	0.001637	0.011538	...	0.063872	0.069649	0.066932
5	0.039764	0.276133	0.000677	0.005389	...	0.035718	0.07497	0.183641

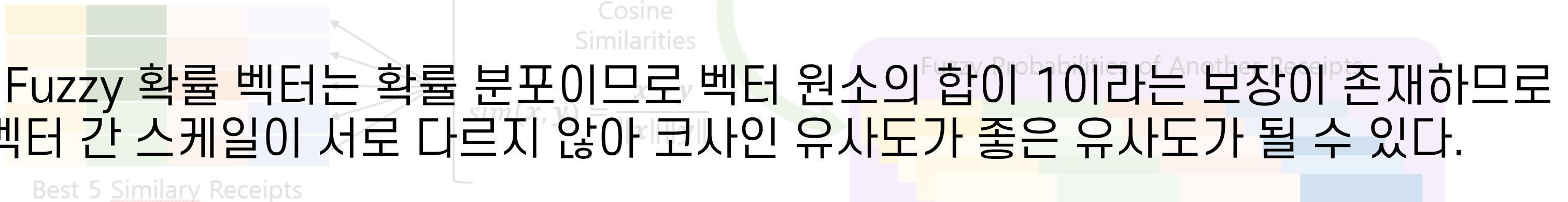
총 20개의 그룹으로 Fuzzy Clustering을 사용하였다.  
 특정 그룹에 확실하게 속하는 영수증 묶음이 있는 반면,  
 다양한 그룹에 속할 확률이 고른 영수증 묶음이 존재한다.



고객의 영수증 끓음이 갖는 확률분포와 비슷한 확률 분포를 갖는 영수증 끓음이라면,  
유사한 구매 패턴을 가진 영수증 끓음이라고 할 수 있다.



코사인 유사도는 두 벡터의 크기와 관계없이 방향만 같으면 큰 값을 갖는다.  
이 때 스케일이 서로 다른 벡터여도 코사인 유사도가 높을 수 있기 때문에  
코사인 유사도의 신뢰성이 떨어진다.



The diagram shows two vectors,  $x$  and  $y$ , originating from the same point. The angle between them is labeled  $\theta$ . A green shaded area represents the cosine similarity between the vectors. The text "Cosine Similarities" is written above the vectors. Below the vectors, the text "Best 5 Similary Receipts" is visible, indicating that the diagram is related to a recommendation system.

한편, Fuzzy 확률 벡터는 확률 분포이므로 벡터 원소의 합이 1이라는 보장이 존재하므로  
벡터 간 스케일이 서로 다르지 않아 코사인 유사도가 좋은 유사도가 될 수 있다.

따라서 코사인 유사도를 사용하여 유사한 구매패턴을 가진 영수증 묶음 상위 5개를 뽑는다.

Case 1 : 무작위 5개 상품 추천 (0.68208)

Case 2 : 상위 5개 영수증 묶음 내 상품 추천 (4.48946)

Case 3 : 상위 5개 영수증 묶음의 소유 고객들의 다른 구매목록 중 상품 추천 (4.77826)

각 수치는 아래의 평가 산식으로 300번 시뮬레이션하여 얻어진 결과이다.

# Performance

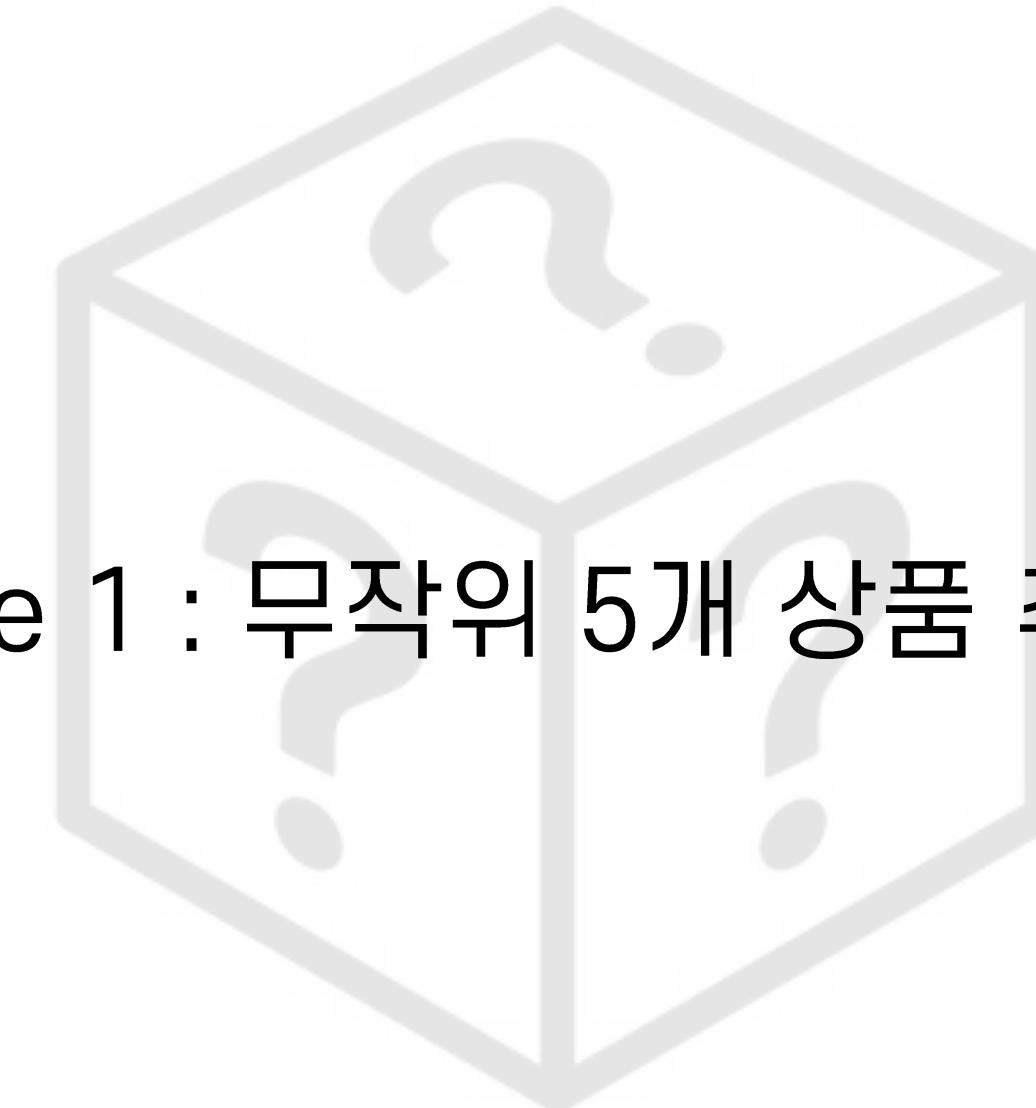
Case 1 : 무작위 5개 상품 추천 (0.68208)

Case 2 : 상위 5개 영수증 묶음 내 상품 추천 (4.48946)

Case 3 : 상위 5개 영수증 묶음의 소유 고객들의 다른 구매목록 중 상품 추천 (4.77826)

각 수치는 아래의 평가 산식으로 300번 시뮬레이션하여 얻어진 결과이다.

$$\text{평가산식} = \frac{\text{5개 추천된 상품 중 사용자가 이미 구매한 상품을 추천한 경우}}{\text{사용자의 총 구매내역}}$$



Case 1 : 무작위 5개 상품 추천

$$\text{평가산식} = \frac{\text{5개 추천된 상품 중 사용자가 이미 구매한 상품을 추천한 경우}}{\text{사용자의 총 구매내역}}$$

## Case 2 : 상위 5개 영수증 묶음 내 상품 추천



$$\text{평가산식} = \frac{\text{5개 추천된 상품 중 사용자가 이미 구매한 상품을 추천한 경우}}{\text{사용자의 총 구매내역}}$$

## Case 3 : 상위 5개 영수증 끓음의 소유 고객들의 다른 구매목록 중 상품 추천



$$\text{평가산식} = \frac{\text{5개 추천된 상품 중 사용자가 이미 구매한 상품을 추천한 경우}}{\text{사용자의 총 구매내역}}$$

무작위로 상품을 추천한 것 보다 (Case1)

비슷한 영수증 그룹 내 상품을 추천한 것(Case2) 이 성능평가지표가 약 6.6배 높다.

선발된 영수증 그룹 외 같은 사람이 구매한 상품을  
추천한 것 (Case3)이 무작위 상품을 추천한 것(Case1) 보다 지표가 약 7배 높다.

비슷한 영수증 그룹을 매칭 시켜서  
매칭된 그룹 외 같은 사람이 구매한 상품을 추천하는 것 (Case3) 이  
영수증 그룹 내 상품 추천 한 것 (Case2) 보다 6%정도 성능이 좋다.

고객과 고객을 연결하는  
개인화 마케팅 C2C플랫폼 :  
**L-Stargram**

고객과 고객을 연결하는 고도화된 영수증 추천시스템과  
리워드 시스템이 맞물린다면 개인의 개성을 보여주는 영수증 컨텐츠,  
L-STARGRAM0I LPAY의 성장을 견인할 것이다.



THANK YOU

SSUDA  
김성연 이동현 조경덕