

# LOTTE MEMBERS

## BIGDATA COMPETITION

TEAM        흥흥흥  
MEMBERS    김창현, 김예린, 정 업



## Exploratory Data Analysis

EDA의 전반적인 시각화 및 EDA를 통해 얻은 Insight에 대한 문제 제기



## Recommendation Algorithm

추천시스템을 구현하기 위해 시도한 몇 가지 방법과 알고리즘 종류,  
구현 코드 Layer 설명



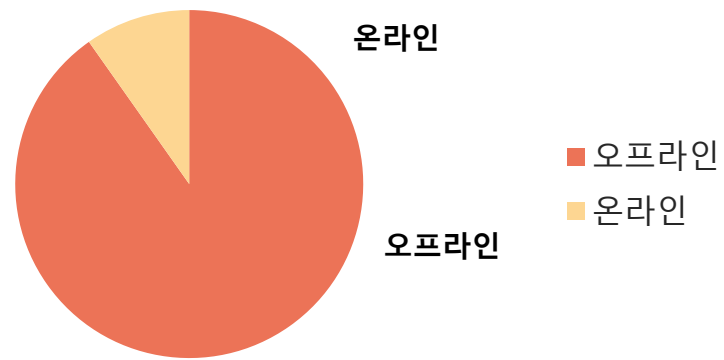
## IDEA

EDA에서 얻은 Keyword로 IDEA를 Development에 대한 과정  
Concept, Target, Case Study, Benchmarking, IDEA

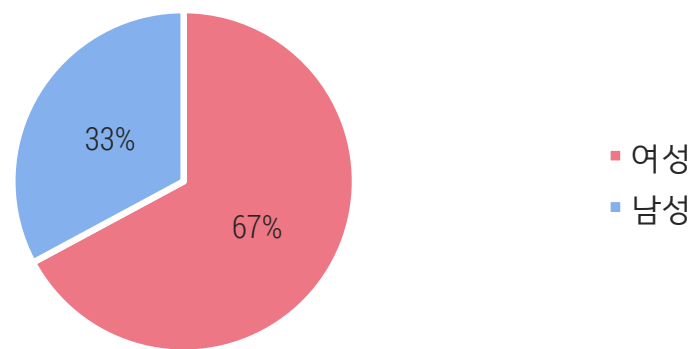
# Exploratory Data Analysis

# EDA Analysis

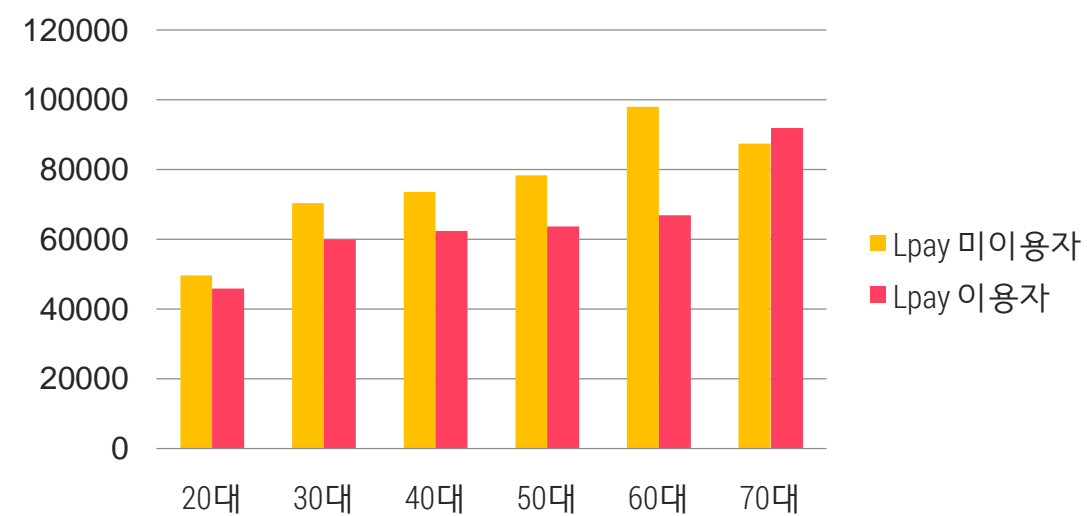
온·오프라인 결제 비율



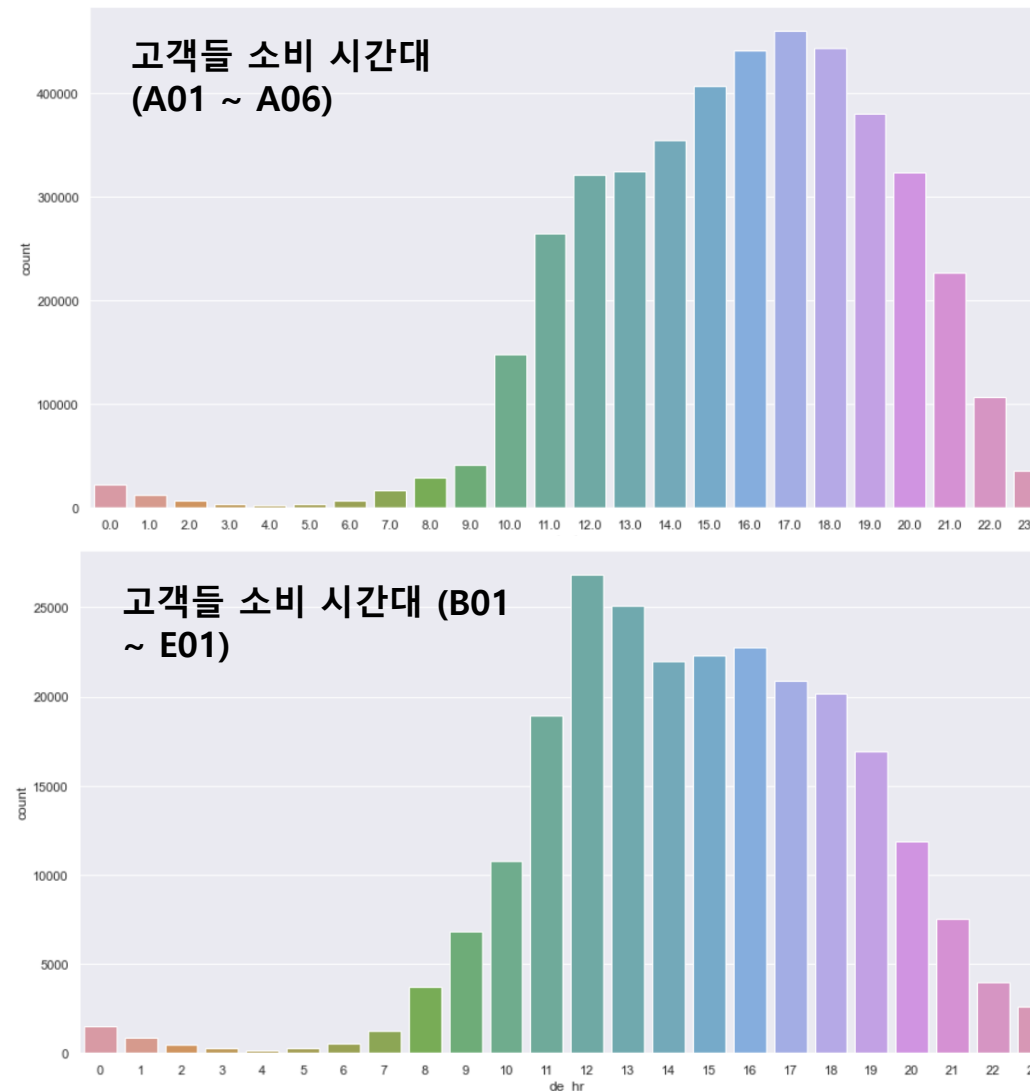
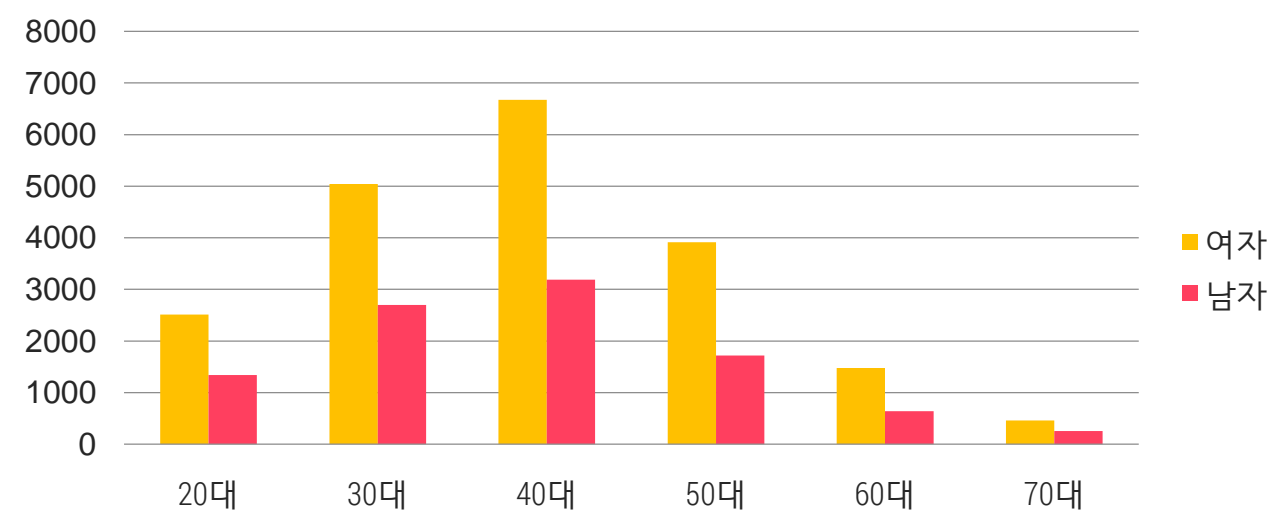
고객 남/여비율



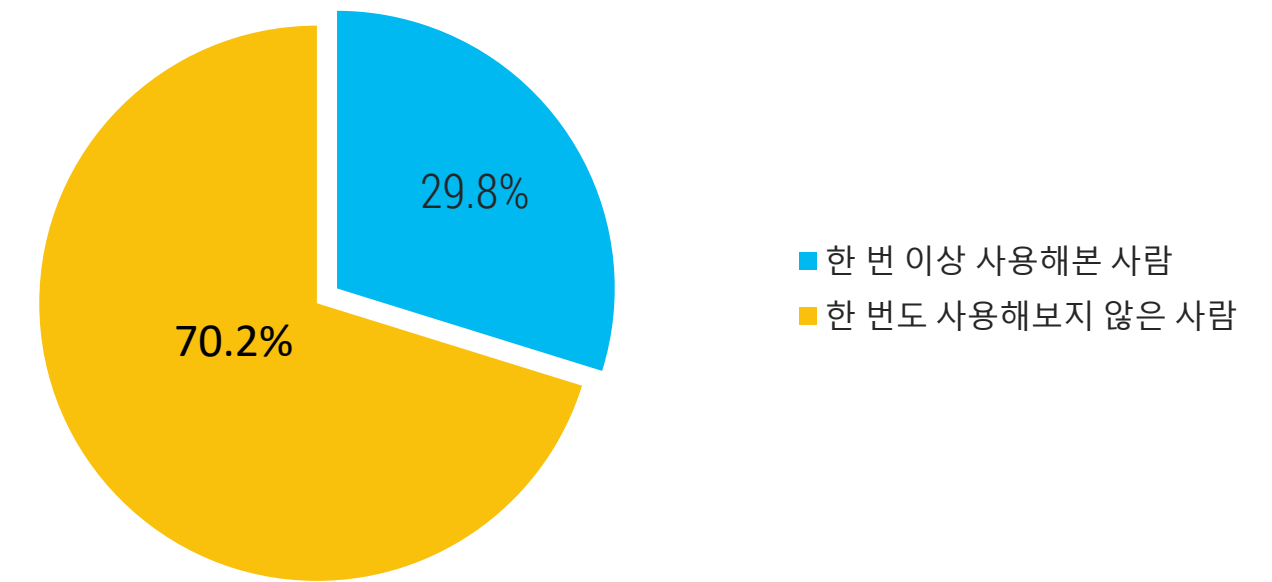
나이대별 평균 구매 액수



나이대별 고객수



Lpay 이용률



L-pay 이용률이 30% 밖에 되지 않는다

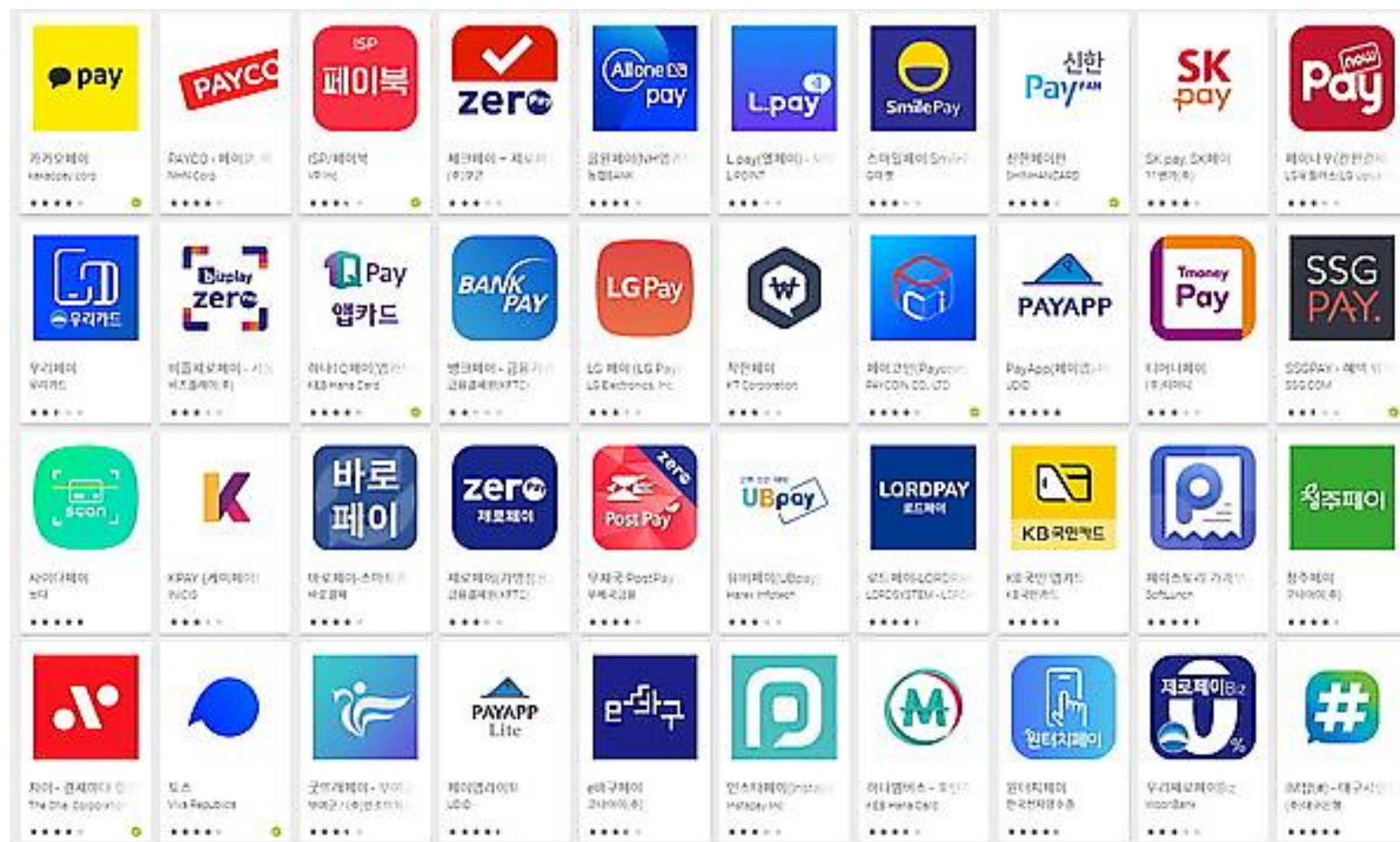
# Pay Issue

## ‘페이 전쟁’

### 유통업계, 자체페이 경쟁 치열...후발주자 가세

강소슬 기자 | 승인 2022.06.19 11:44 | 댓글 0

## 다시 뜨거워진 유통街 '페이 전쟁'...이유는



라다 기자 | 입력 2022-06-14 18:01

요즘 국내 외 크고 작은 기업들이 ‘간편 결제 시스템’ 전쟁을 벌이고 있음

# Pay Benefit

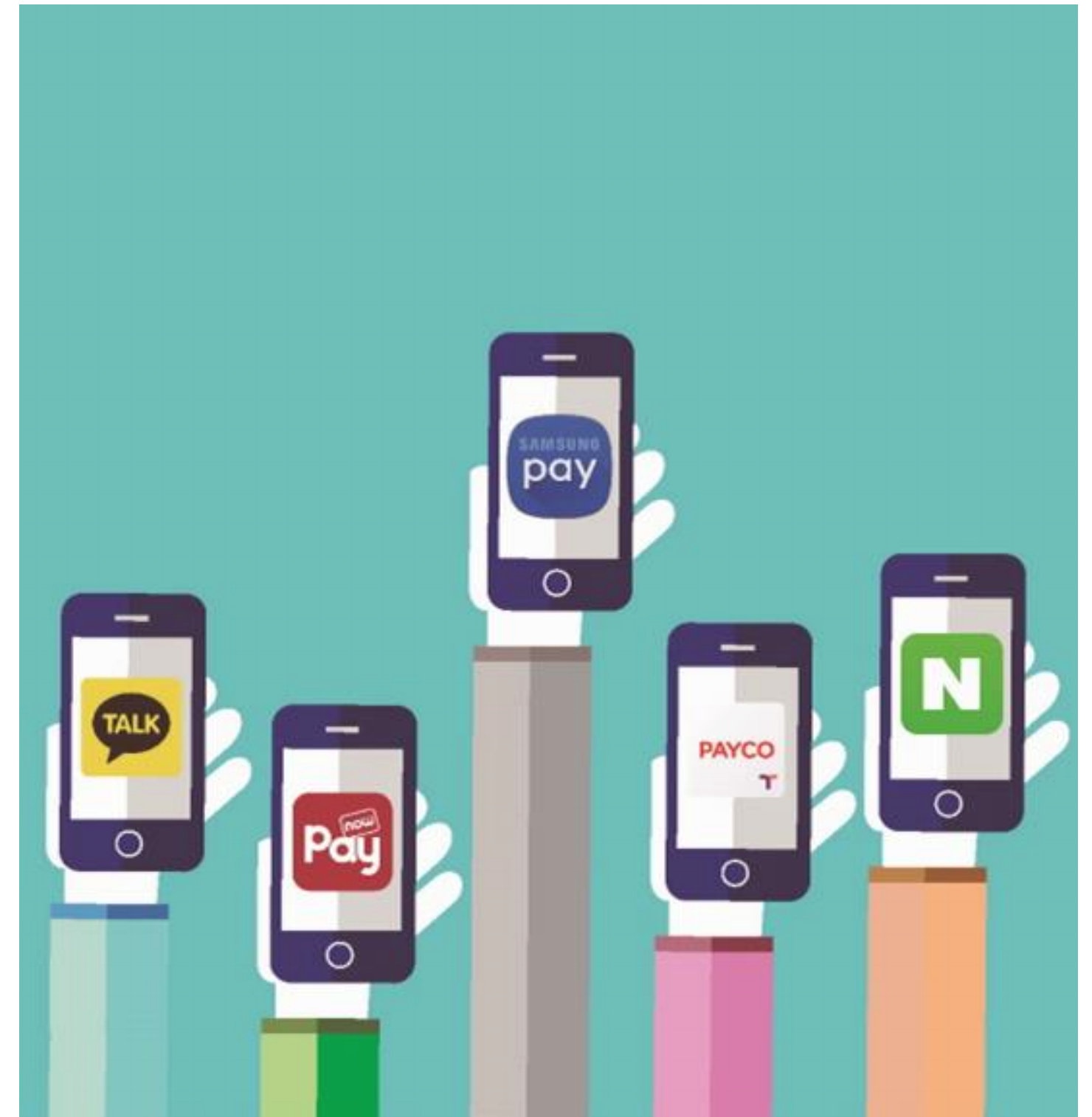
## 왜 모두 OO페이를 출시하려고 할까?

락인 효과 (Lock-in) : ‘페이’로 고객을 유입해 충성고객을 만들기 위한 의도  
ex) 쿠팡의 쿠팡플레이, 네이버 멤버십

자체 결제 시스템을 통해 수수료 절감 효과

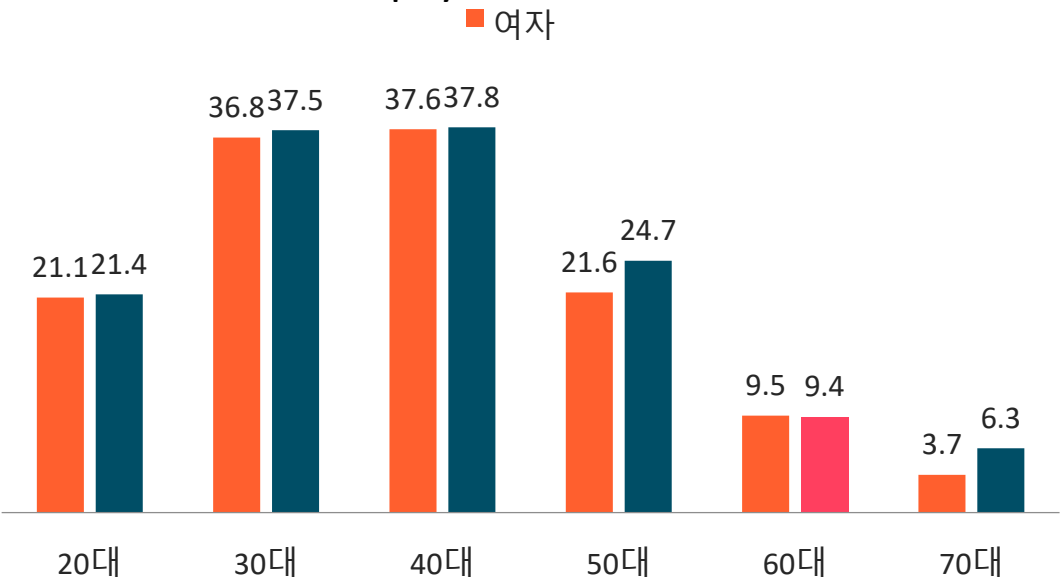
넓은 고객 기반을 토대로 금융업을 연계한다면 추가 수익원 확보 가능

간편결제를 통해 쌓인 소비자 데이터를 활용할 수 있음

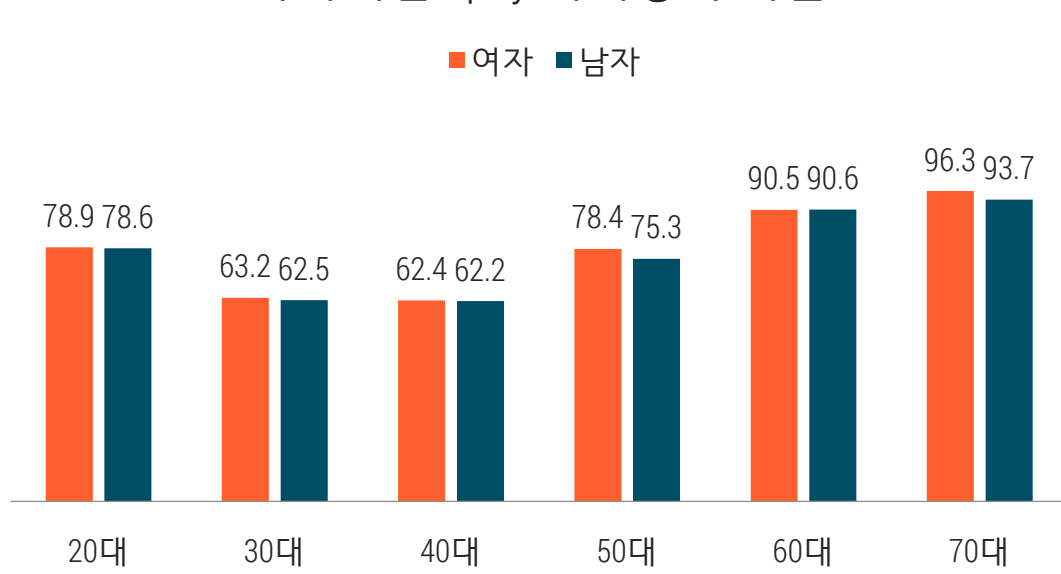


# EDA Analysis of L-pay

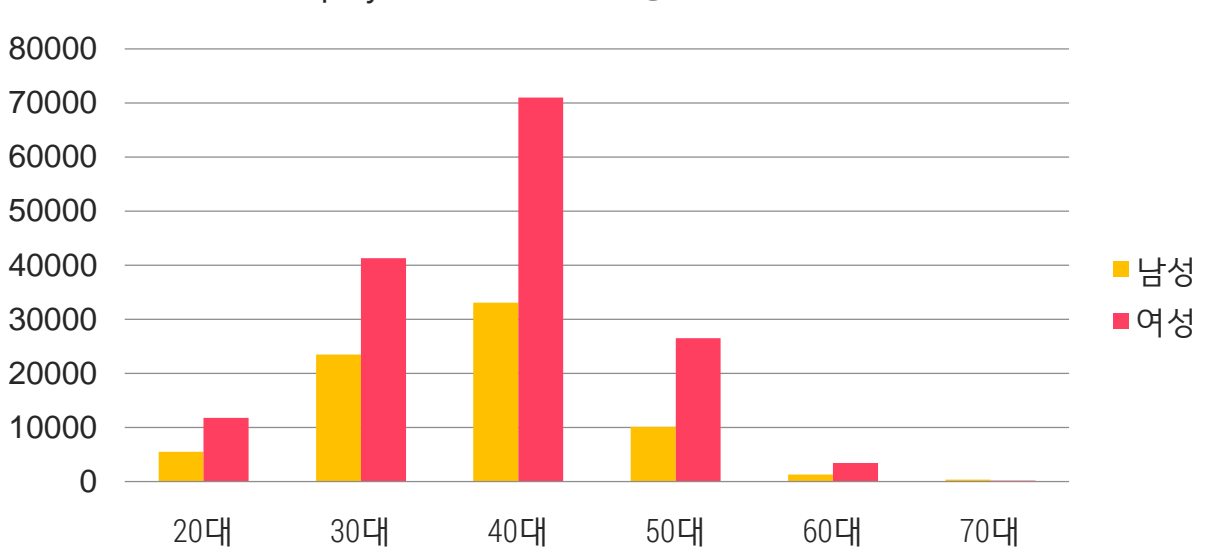
Lpay 이용자 비율



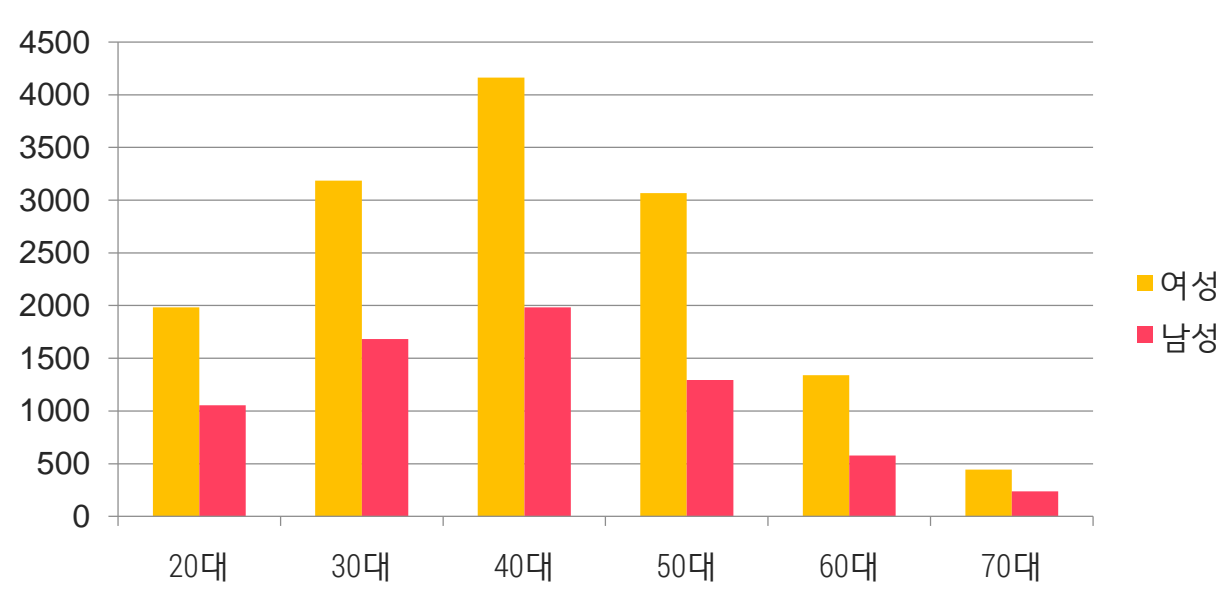
나이대별 Lpay 미이용자 비율



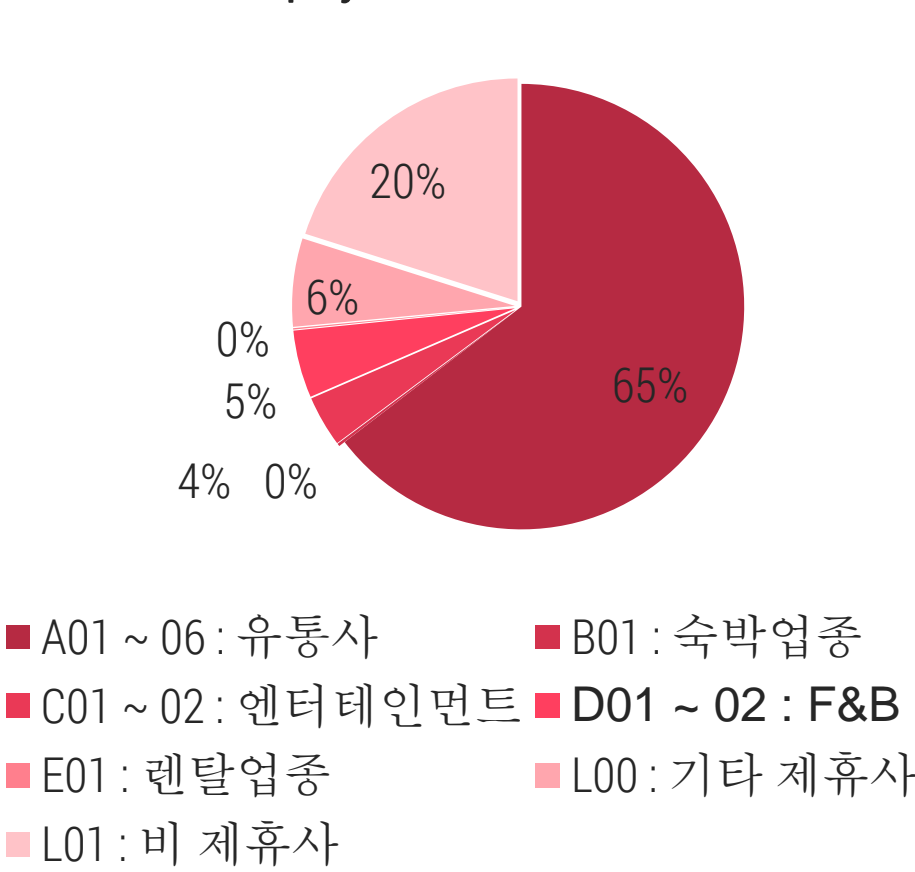
Lpay 제휴사 A 연령별 구매횟수



Lpay 나이대별 미이용자수



Lpay 제휴사 이용비율



엘페이의 사용률을 증진시키기 위해서는 어플을 원활하게 사용 할 수 있는 연령과 엘페이 사용 연령층을 비교하여 최적인 30 ~ 40대로 타겟을 선정

L-pay 결제 중 대부분은 유통사 A가 65%를 차지하고 있었고 상대적으로 데이터가 많아 A 유통사 상품을 추천하는 것으로 정하였다.

# **Recommendation** Algorithm



# Recommendation Column

## Recommendation System

기존 협업 필터링 User-based Filtering , Item-based Filtering, Matrix Factorization 추천시스템, Deep Learning 기반 추천시스템 등 롯데에서 제공 받은 데이터로 많은 시도를 해보았다.

위의 알고리즘은 평점데이터(Rating)의 데이터가 주어지고 그것을 토대로 예측을 한다. 하지만 제공 받은 데이터에는 Rating이라는 정보가 없어서 직접 빈도수를 통해서 Rating column을 만들었지만, 이것은 구매 빈도일 뿐 정말로 사용자가 만족해서 점수를 이렇게 주었을 지는 알 수 없었다.

또한 위의 시스템은 제공받은 데이터에 적용하기에는 무언가 무리가 있어 보였다. 적절한 알고리즘을 찾아보도록 시도하였고 Xgboost, Faiss, Gru4rec의 알고리즘을 이용하여 추천시스템을 구축하였다.

### XGboost

Xgboost는 Tree기반 알고리즘으로 현재 가장 인기 있는 분류 알고리즘이다.

결론적으로는 Xgboost를 사용하지는 않는다. 하지만 이 알고리즘을 사용한 경험을 토대로 새로운 아이디어를 얻을 수 있었다.

### Faiss

처음에는 이 알고리즘으로 전체적인 추천시스템의 틀을 만들어 보려고 하였었다.

Xgboost를 사용해보면서 단순히 하나만 추천한다는 것은 모델 성능, 정확도 양쪽을 잡을 수 없다는 사실을 알았고 추천 범위를 축소할 필요성을 느끼게 되어 유사도를 채택하여 사용하게 해 보았다.

### Gru4rec

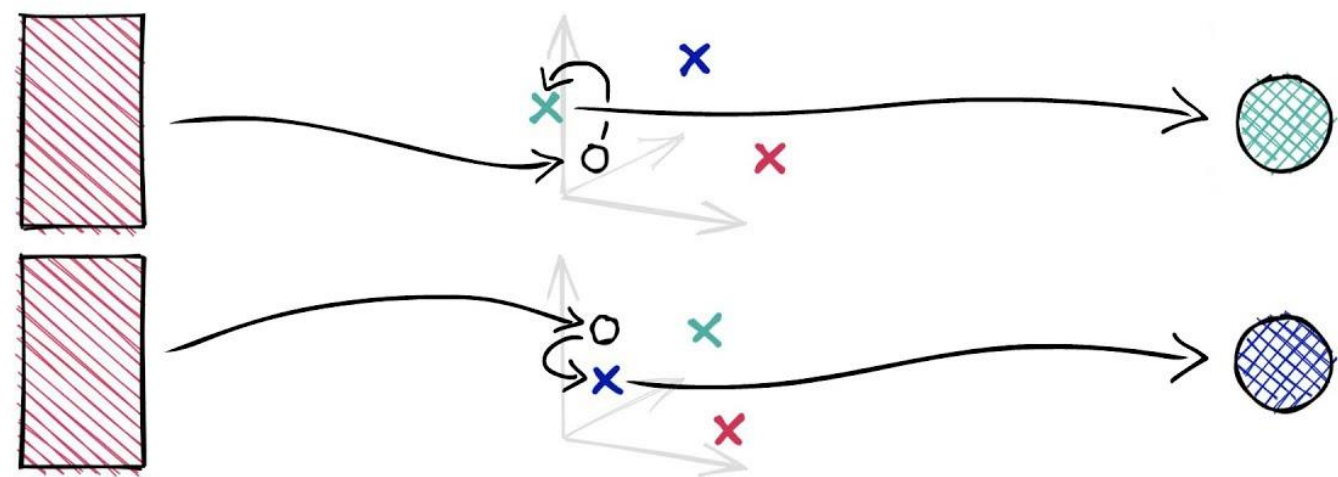
“고객의 선호도는 이전 선호도와 다를 것이다”라는 의문이 들어서 가장 최근 선호하는 아이템 위주로 추천시스템을 구현해 보려고 하였고 Deeplearning 기반의 추천시스템 (RNN, LSTM)을 주로 찾아보다 GRU4REC를 발견하게 되었고.

LSTM의 변형이며 Session-based 추천시스템으로 최근 Sequential한 정보를 이용해서 다음 상품 예측이 가능하다는 점에서 적극 활용하게 되었다.

# Recommendation Faiss

## Similarity Search

**IndexPQ and IVFPQ in Faiss**



### FAISS

Input = [고객번호, 나이, 성별, 날짜, 시간, 구매개수, 온/오프라인, 지역 대분류, 기온]

vector : (4616646, 9)

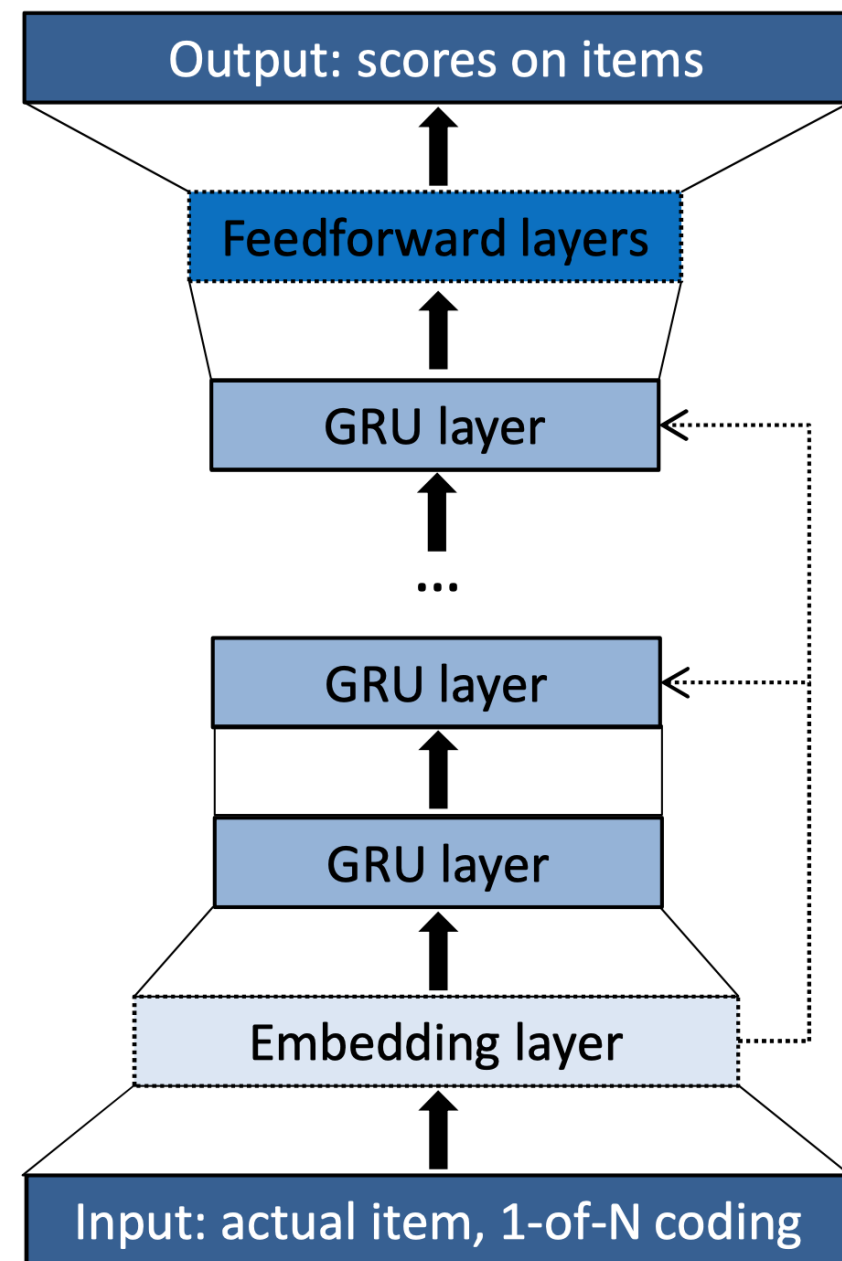
Label : 상품 대분류(60)

Faiss은 성분 분해 알고리즘인 Product Quantization 알고리즘이며 Optimized Product Quantization for Approximate Nearest Neighbor Search를 기반으로 작성되었다.

Cold start에 해당하는 고객의 데이터를 다른 고객 데이터와 Distance L2 유사도 비교를 통해서 비슷한 벡터를 알 수 있으며 비교적 적은 데이터로 cold start에 해당하는 고객들에게 아이템을 추천하기에 적절한 알고리즘이라고 판단되었다.

Faiss 유사도는 Top 10을 기준으로 데이터를 추출하지만 유사한 벡터가 몇 가지 없을 경우  $1 < x < 11$  사이의 데이터가 추출 된다.

# Recommendation Gru4rec



## GRU4REC

Input = [고객번호, 구매날짜 시간(timestamp)]

Train : (3154854, 9)      Val: (350540, 9)      Test : (876349, 9)

Label : 상품 대분류(60)

Accuracy : 0.2790      TOP5\_Accuracy : 0.4629  
TOP10\_Accuracy : 0.6287      TOP15\_Accuracy : 0.7016

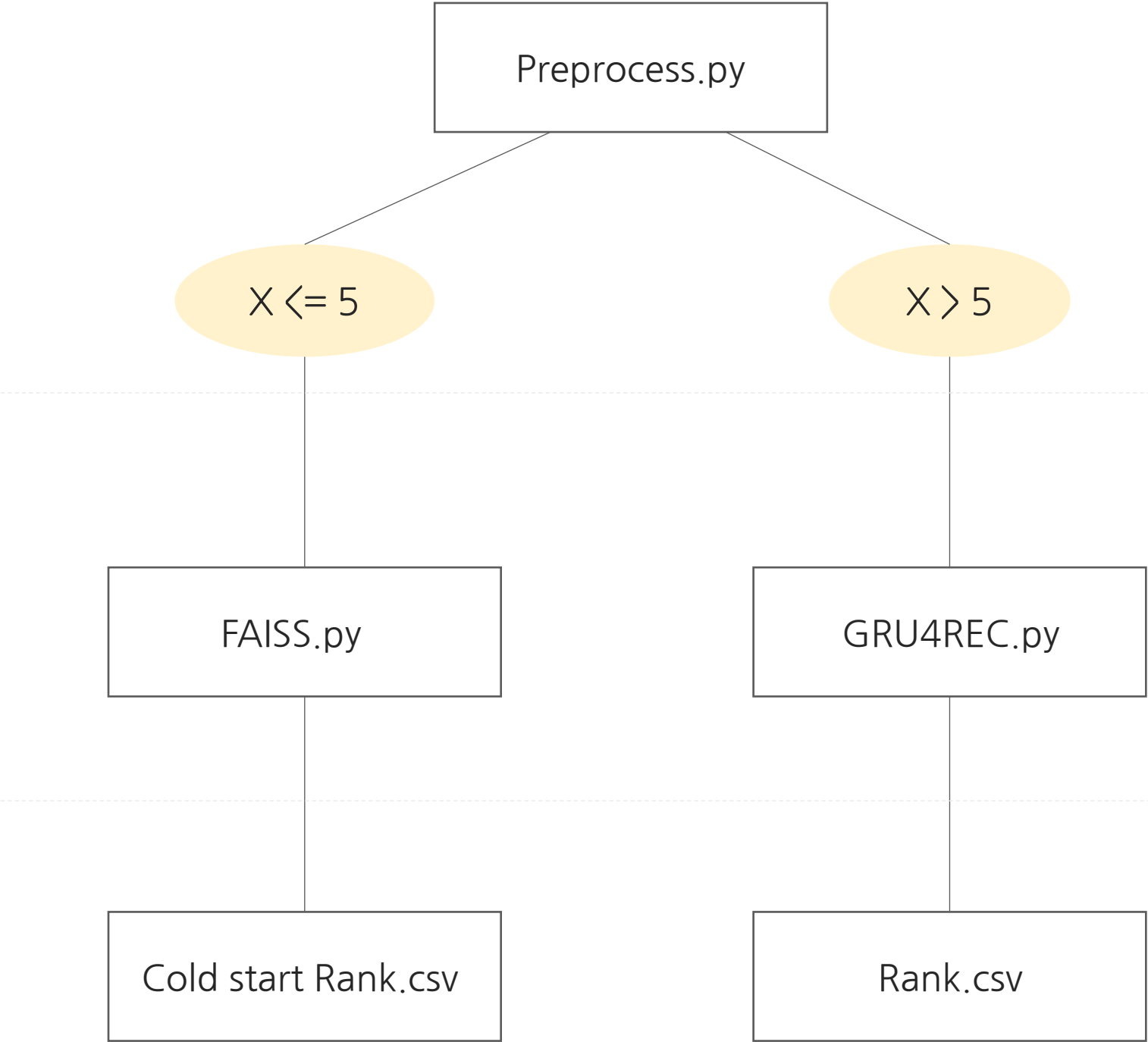
### Session based recommendation

GRU4REC는 Session이라는 Sequence 데이터를 GRU 레이어에 입력하여 다음에 올 확률이 가장 높은 모델을 추천하는 모델 (LSTM 변형 모델)

GRU Layer에서는 Sequence 상 모든 아이템들에 맥락적 관계를 학습한다.

고객의 선호도는 고정된 것이 아니기 때문에 가장 최근 고객이 선호하는 상품이 무엇인지 알 필요가 있고 또한 그것을 추천해 주어야 한다고 생각 되었다.  
따라서 최근의 정보를 고려 할 수 있는 GRU4REC 알고리즘이 적합할 것이라고 생각된다.

# Recommendation Layer



## Layer I - preprocess

고객데이터 5개 이하, 5개 이상으로 데이터를 나눔  
csv format으로 저장

## Layer II - Algorithm

고객데이터 5개 이하.csv → Faiss.py (유사도 알고리즘)  
고객데이터 5개 초과.csv → GRU4REC.py (Session-based)

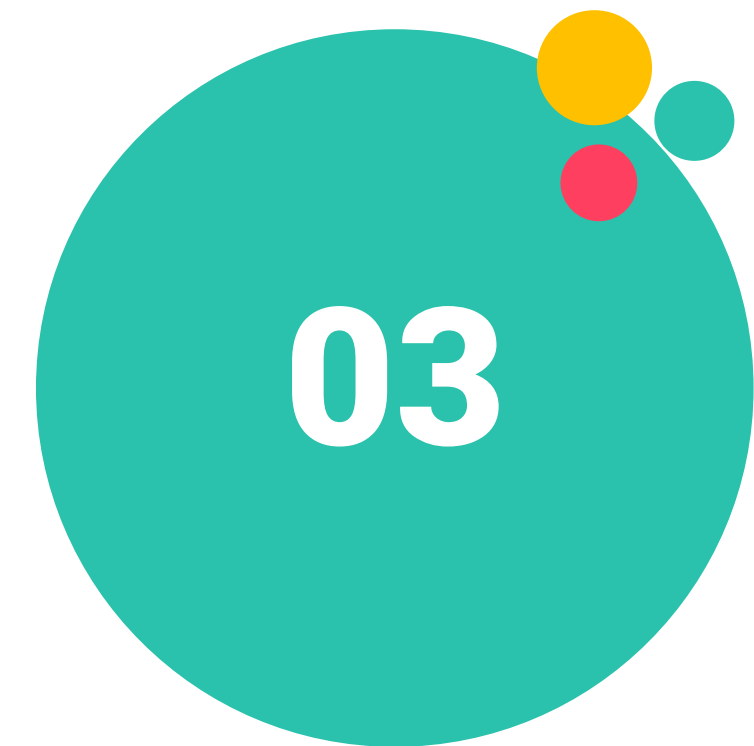
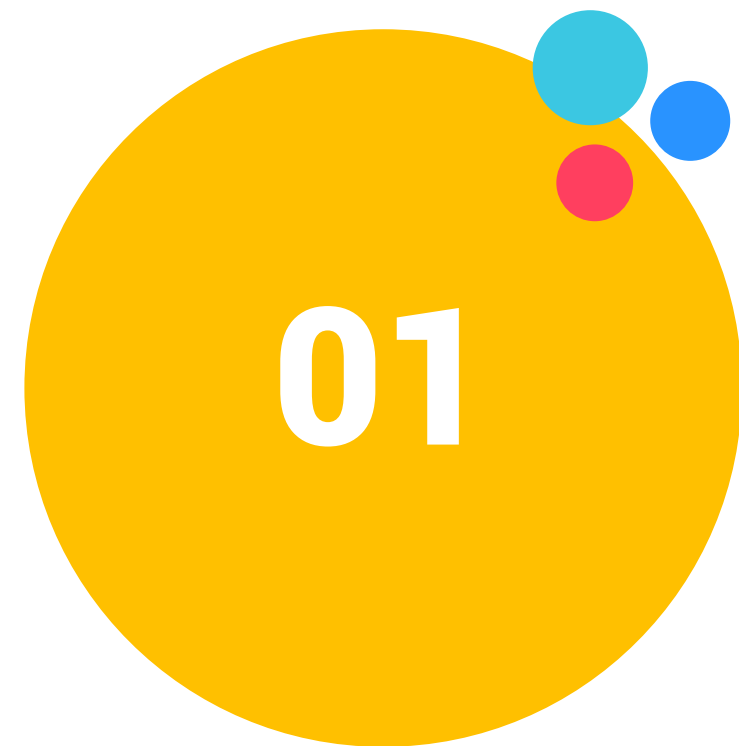
각각 상위 TOP 10 데이터 추출

## Layer III - Output

Cold start Rank.csv, Rank.csv라는 이름으로 특정 폴더에 저장

**IDEA**

# Three Steps to IDEA



## Concept

EDA 시각화 자료를 통해 얻은  
Insight로 컨셉 설정

## Case Study

다른 경쟁업체들 APP Event의  
종류와 의도 파악

## Benchmarking & Idea

경쟁사의 Event를 조사하고  
Benchmarking하여 IDEA Development

# Development Concept



## L Pay

EDA 데이터 상에서 L-PAY를 통해서 구매하는 사람은 29.8%로 전체의 약 1/3 수준

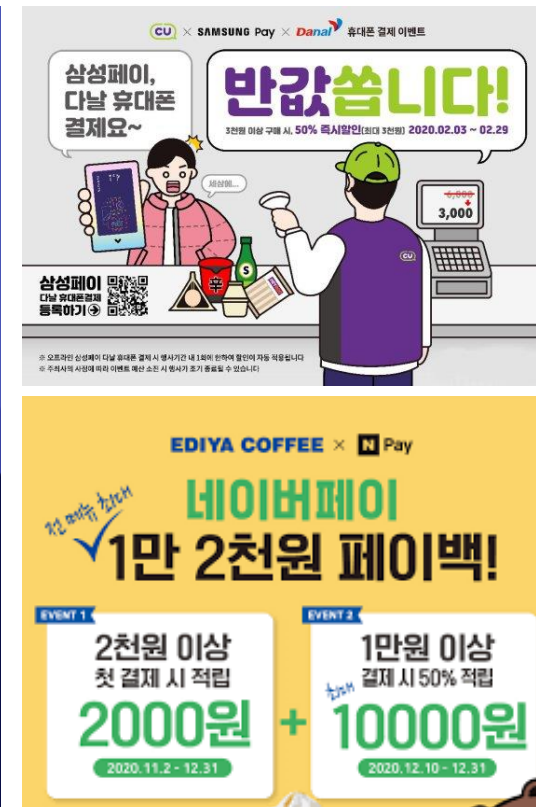
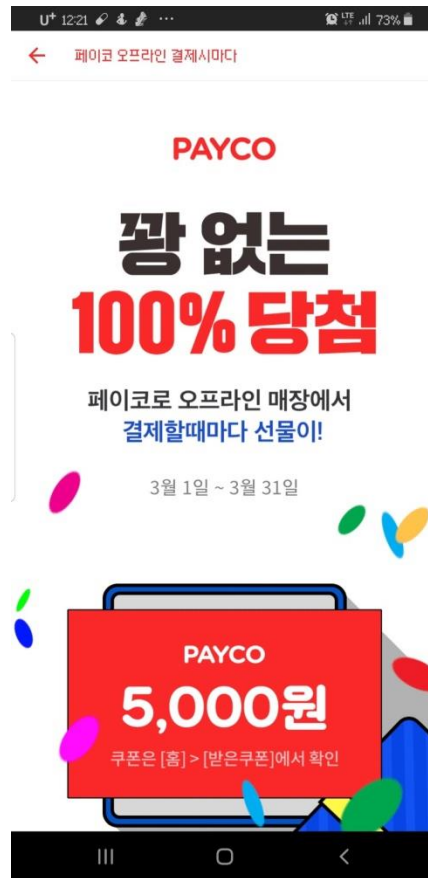
어떻게 L-PAY라는 수단이 고객들에게 매력적으로 느껴지게 만들 수 있을 까라는 부분에 Point를 두고 진행

## Main Target

APP 활용 확률이 높은 30 ~ 40 대의 연령 남 / 여 고객층



# Pay Event case



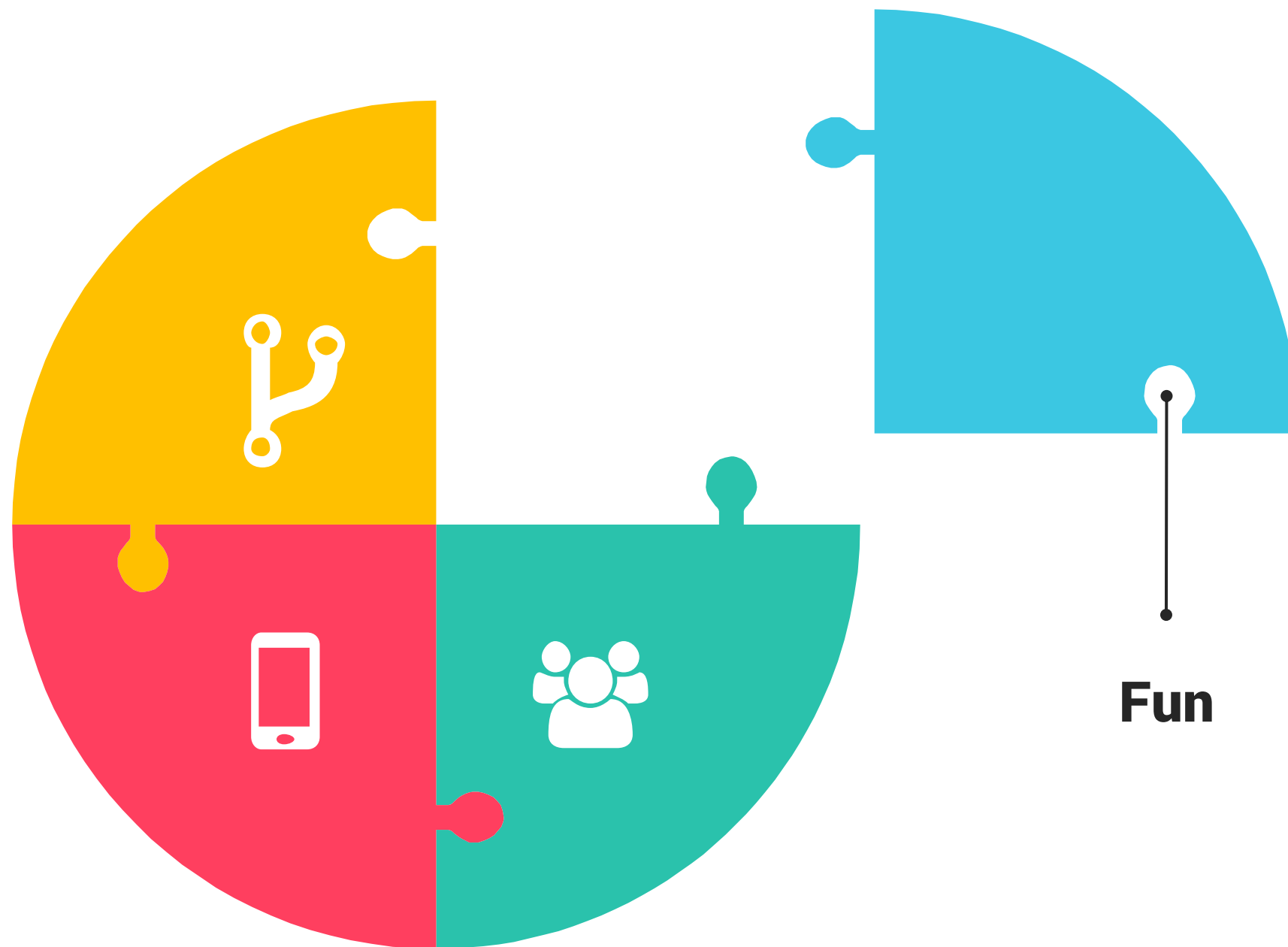
## Event

- 스탬프
- 반값
- 페이백
- 복권
- 당첨
- 추천
- 포인트
- 특가상품 할인

다른 pay의 이벤트도 위의 Keyword에서 크게 벗어나지 않는다.



# Development Concept



## Fun

지속적으로 L-pay App을 활용하기 위해서는 어떠한 요소가 필요하다고 판단되었고 그것이 “재미” 요소를 집어 넣는 것이다.

단순하게 쿠폰형식으로만 지급되는 것이 아닌 재미요소를 포함해서 기존 이용 고객들에게 유의미한 효과를 얻을 수 있다면 신규 고객 유치에도 큰 도움이 될 것이라 사료된다.

# Development Concept

## L Pay

엘페이 사용률 증진을 위한 방법



## Recommendation

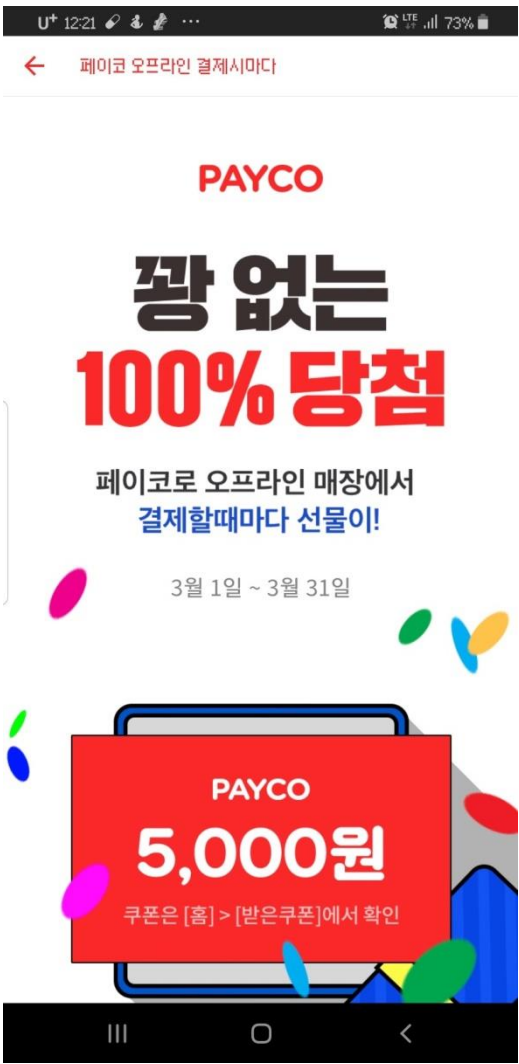
추천시스템 알고리즘으로 고객 맞춤 상품  
TOP 10 추천.

## Fun

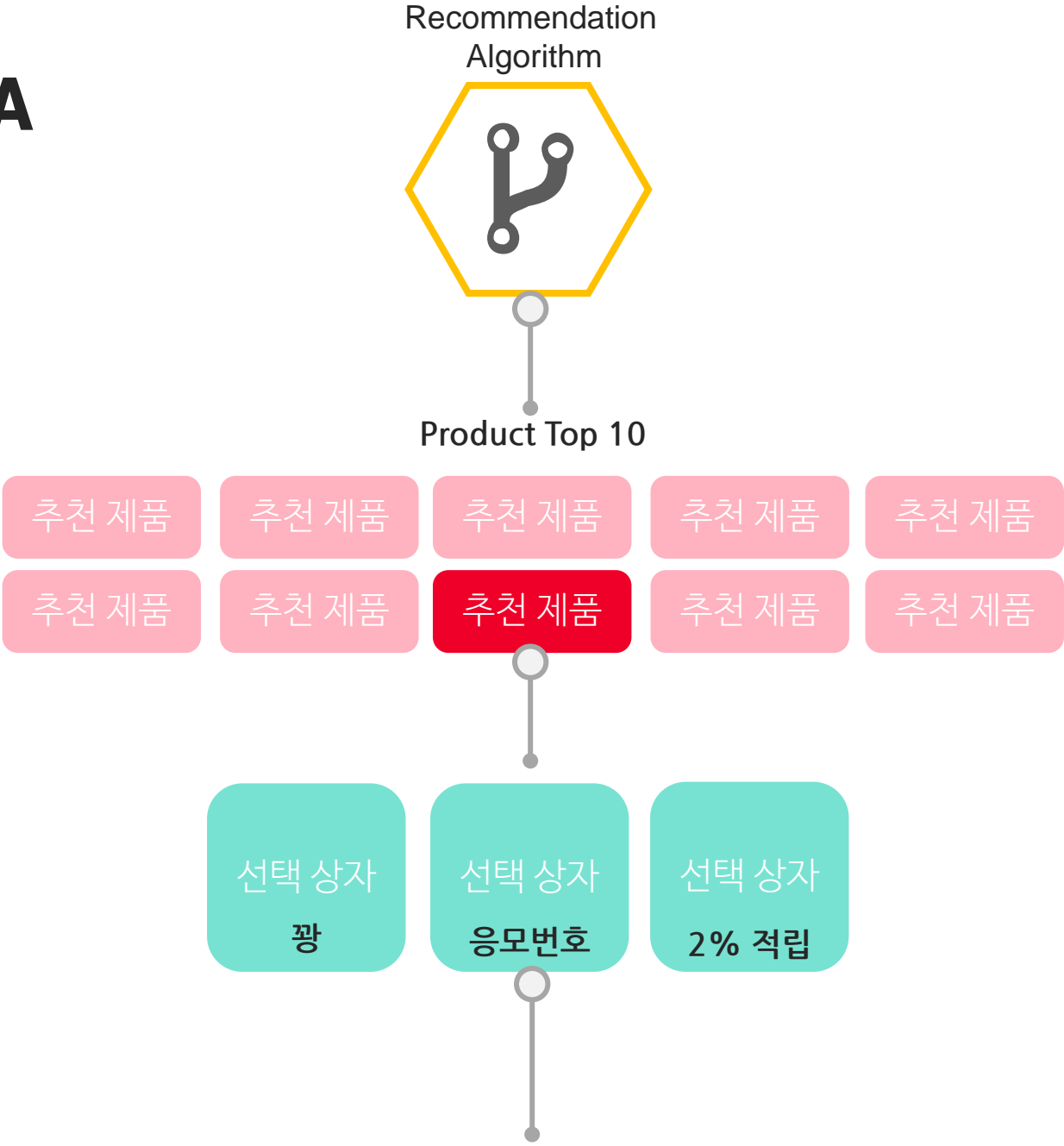
기존 고객과 신규 고객에게 재미라는  
새로운 경험으로 흥미유발

# Pay Event Benchmarking & IDEA

## Benchmarking

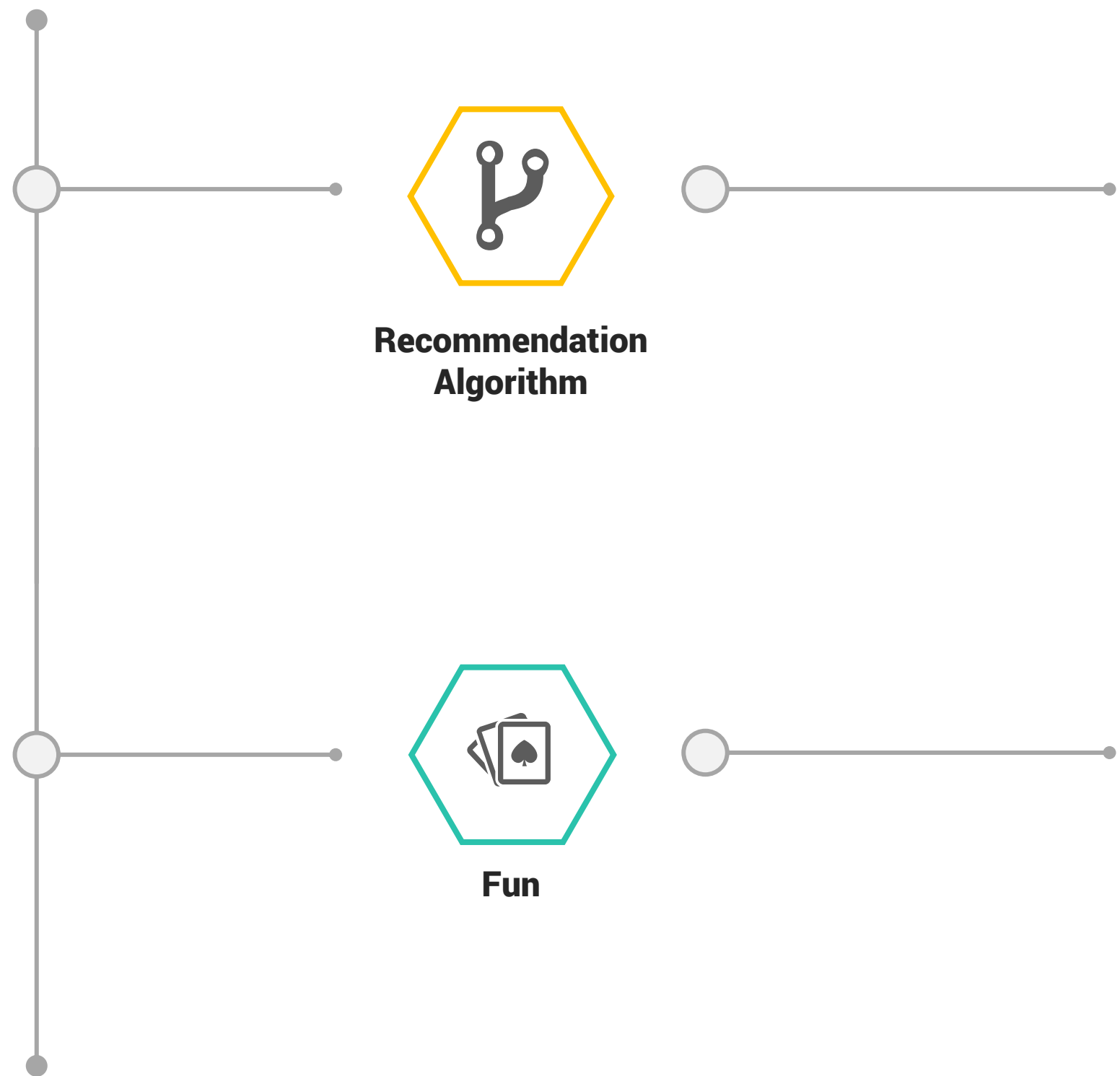


## IDEA



추천 된 제품을 구매하면 매번 선택상자를 선택할 기회를 줌

# Event IDEA Detail



- 추천된 제품이 고객이 원하는 제품일 확률은 62%이다.
- 이 중 32%는 원하는 제품이 아닐 수 있지만 해당 제품을 보여줌으로 해당 제품을 보고 구매율을 조금이라도 늘릴 수 있을 것이라 사료된다.



- 추천된 제품을 구매하면 상자를 선택할 기회가 주어진다
- 선택 상자에서 선택된 것이 응모번호라면 매주 ~ 매달 진행되는 경품 추첨에 참가 신청을 할 수 있다.
- 추천 제품을 많이 살 수록 응모 번호를 획득 할 확률이 증가하고 동시에 경품에 당첨될 확률이 상승한다라는 점을 강조한다