

**L.POINT**

# 미방문 고객 예측 및 모델 성능에 따른 개인화 마케팅

---

L.Point Big Data Competition

Team ROYAI

Members 나신영 박지수 최명진

# 목차

- 01** 데이터 탐색
- 02** 데이터 전처리
- 03** 모델링
- 04** 개인화 마케팅

# 데이터 탐색

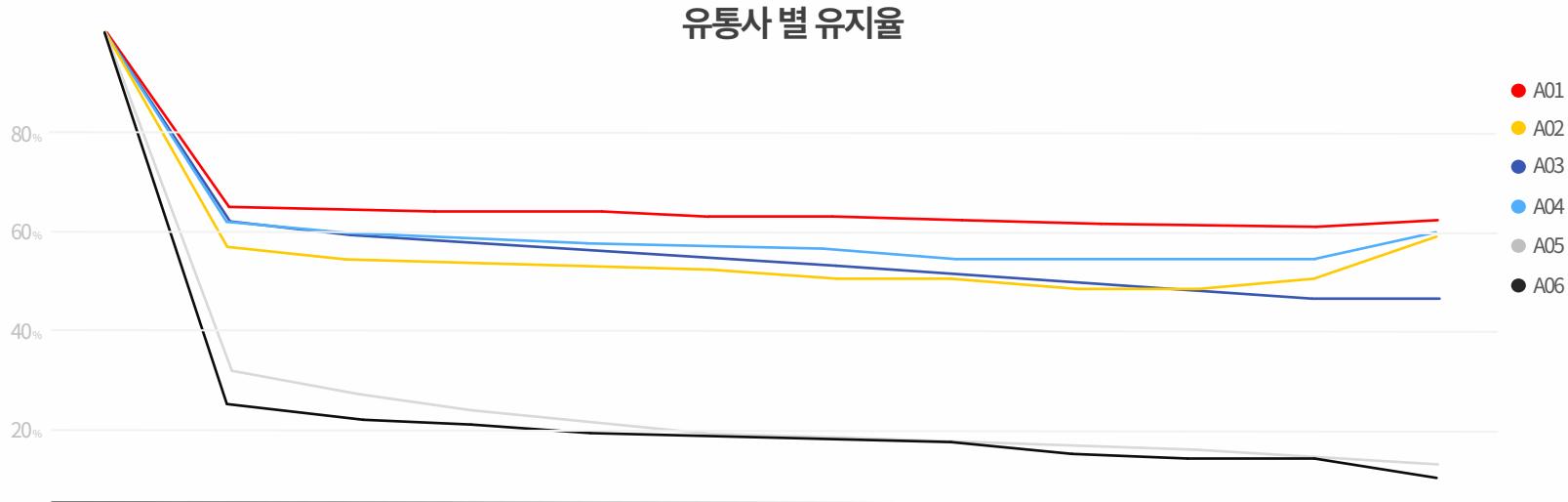
# 01 데이터 탐색

유통사 전체 유지율

1월	100	99	98	98	97	96	95	94	92	90	85	76
2월	100	94	93	91	89	87	85	83	78	71	57	
3월	100	89	87	84	83	79	75	71	62	48		
4월	100	83	80	76	73	69	64	57	43			
5월	100	78	75	71	67	63	56	42				
6월	100	72	68	63	58	51	41					
7월	100	66	62	56	47	35						
8월	100	59	54	47	38							
9월	100	51	41	30								
10월	100	41	31									
11월	100	30										
12월	100											

해당 차트는 특정 일자 이후부터 상품을 구매한 모든 고객의 유지율을 누적으로 나타낸 것이다(unbound retention). 1월 달에 첫 구매를 한 고객은 12월까지 잔류 비율이 높지만 첫 구매일이 12월에 가까워 질수록 잔류 비율이 낮아지는 것을 확인할 수 있다.

# 01 데이터 탐색

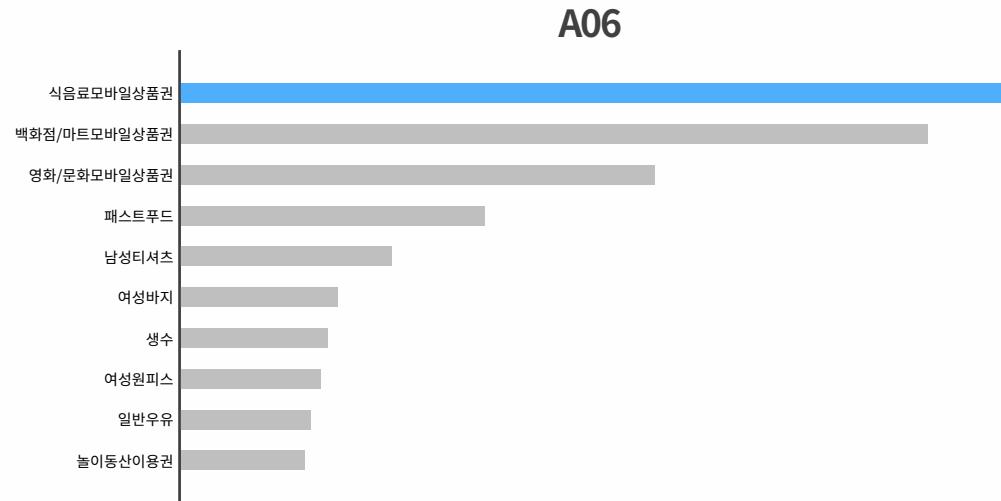
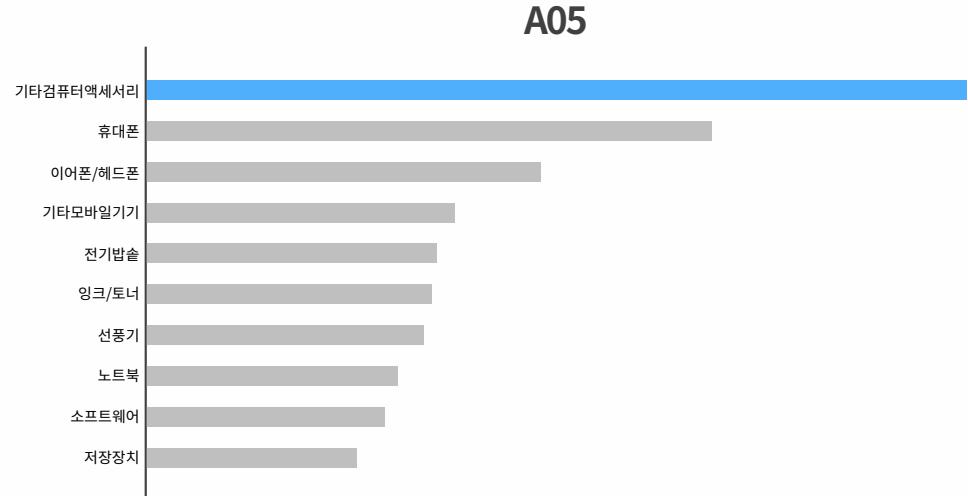


다음은 유통사 별로 유지율을 나타낸 것이다. A01 ~ A04 와 A05 ~ A06 이 서로 다른 양상을 보여주는 것을 알 수 있다.

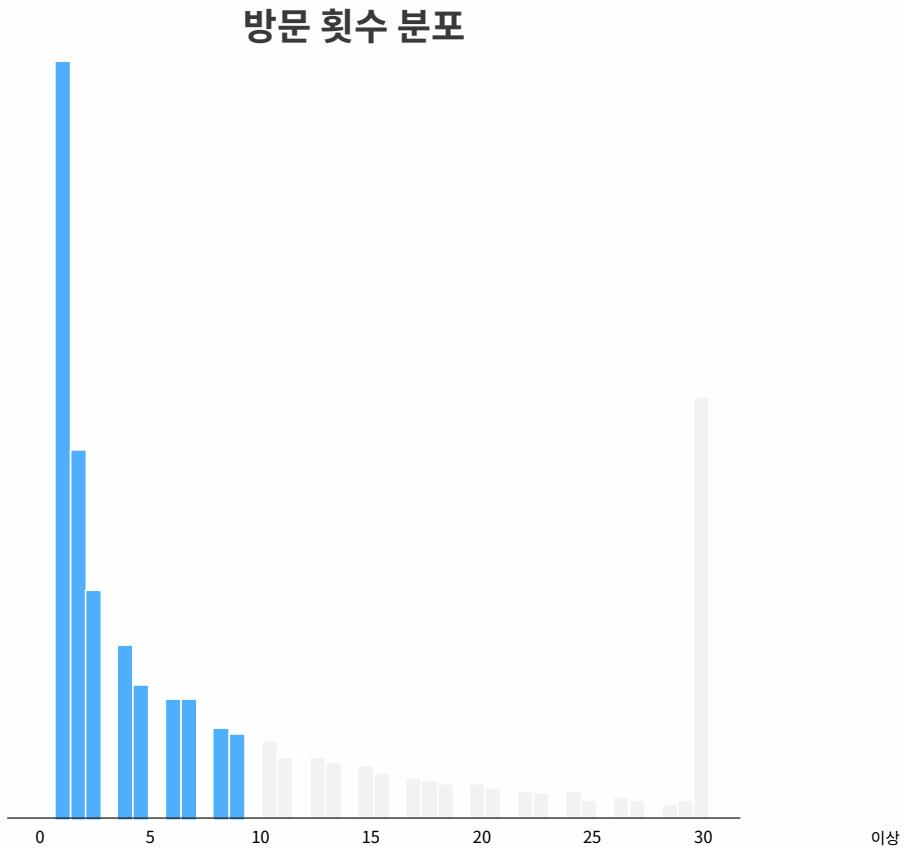
# 01 데이터 탐색

다음은 A01 ~ A06 유통사와 다른 양상을 보여준  
A05, A06 유통사의 주요 판매 품목들을 나열한 그래프이다.

A05의 경우 전자기기 / 가전제품 종류가 많은 비율을 차지하고,  
A06의 경우 상품권 류가 많은 비율을 차지하는 것을 알 수 있다.

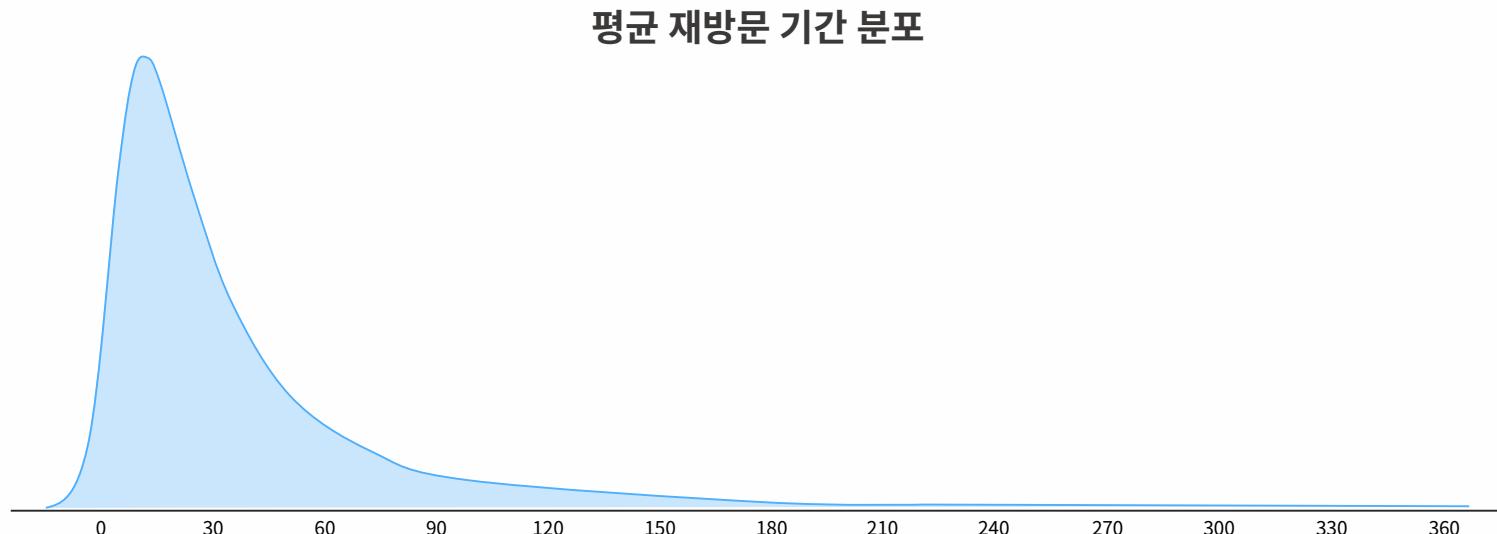


# 01 데이터 탐색



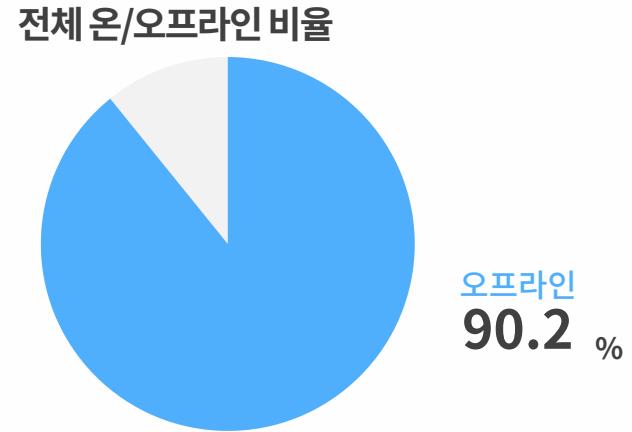
왼쪽의 그래프는 사용자의 방문 횟수 분포를 나타낸 그래프이다.  
대부분의 사용자의 방문 횟수가 10회 이하임을 확인할 수 있다.

# 01 데이터 탐색



위는 고객들의 평균 재방문에 걸리는 시간을 나타낸 그래프이다. 주로 30일 근처에 집중되어있는 것을 확인할 수 있다.

## 01 데이터 탐색



유통사 거래 데이터를 보면 오프라인 판매 비율이 90.2%로 지배적이었다.

# 01 데이터 탐색

## 문제 정의

위의 내용들을 요약해보면 다음과 같다.

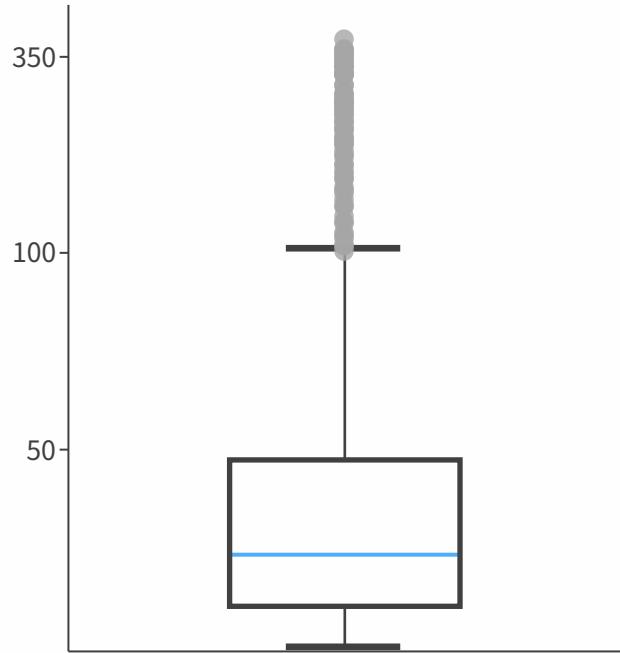
1. 첫 구매일자가 12월에 가까울 수록 잔류 비율이 급격하게 줄어든다.
2. A01 ~ A04 유통사와 A05 ~ A06 유통사의 잔류 비율 양상이 서로 다르다.
3. 구매 고객들의 재방문 빈도는 30일 근처가 가장 많았다.
4. 고객들의 방문 횟수가 대부분 10회 이하였다.
5. 오프라인 구매 비율: 온라인 구매 비율 = 9 : 1

구매고객의 빈도는 높지 않고 재방문까지 긴 시간이 걸린다. 따라서, 유지율과 더불어 고객의 재구매 기간을 단축하기 위해 본 프로젝트 주제를 선정하였다.

# 데이터 전처리

## 02 데이터 전처리

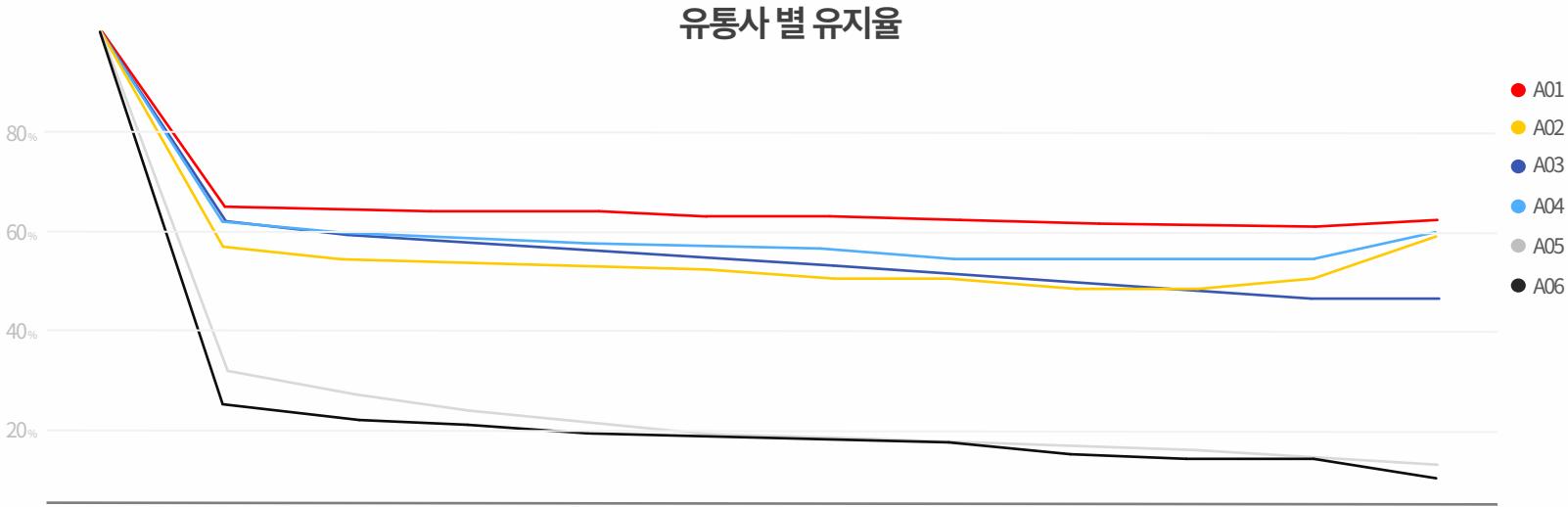
### 관찰기간 설정



관찰 기간(observation window)이 길면 활동을 더 이상 하지 않는 고객들이 포함될 확률이 높아지고, 관찰 기간을 짧게 잡으면 예측을 위한 데이터가 충분하지 않을 가능성이 있다.

재방문하는 평균 기간이 39일이기 때문에, 관찰 기간을 2개월로 정하였다.

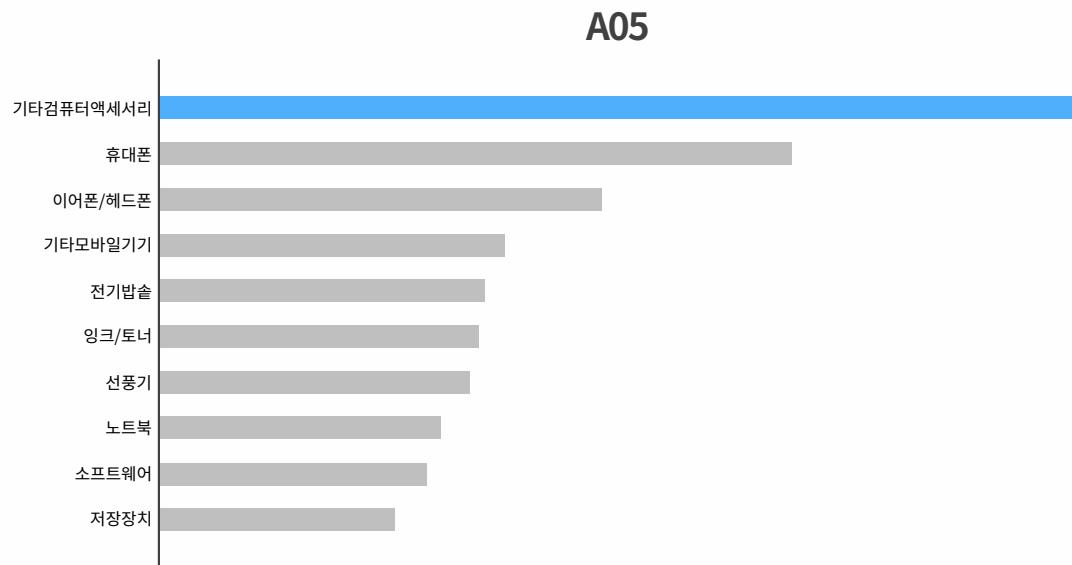
## 02 데이터 전처리 A05, A06 데이터 제거



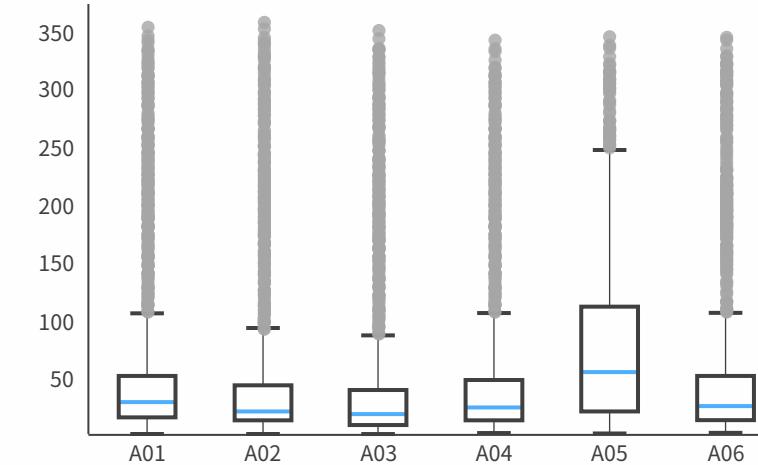
데이터 셋으로부터 학습에 필요한 feature들을 생성하기 전에 모델의 학습 성능을 높이기 위해 성능 저하의 원인이 되는 몇 가지 데이터들(결측치 등)을 제거 해야할 필요성이 있다.

위의 그래프는 앞에서 설명했었던 유통사별 고객 잔류 비율을 나타낸 것이다. 여기서 A05와 A06 유통사의 경우 다른 양상을 보여주고 있으므로 왜 이런 양상을 보이는지를 알아보고 제거 여부를 결정하기로 했다.

## 02 데이터 전처리 A05, A06 데이터 제거



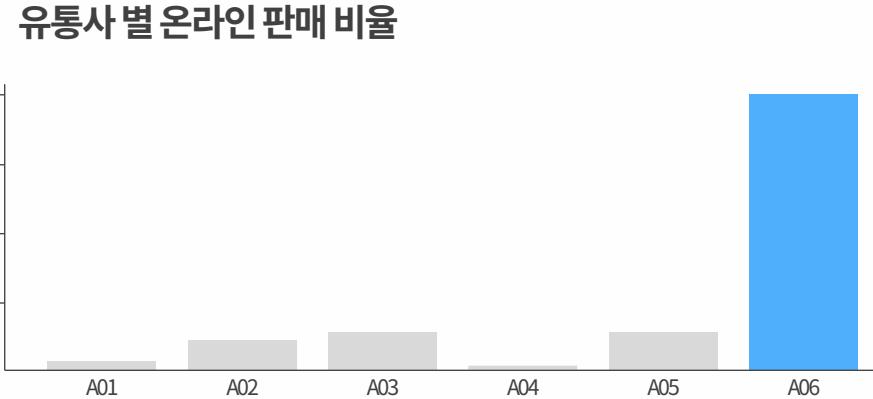
유통사별 방문 빈도



먼저, 유통사 A05의 주요 품목들을 나열해 보았다. 품목명들을 살펴보면 주로 전자제품과 가전기기 들로 구성되어 있다는 것을 알 수 있다.

전자제품과 가전기기의 경우는 한 번 구매하면 오래 쓰는 경우가 많다. 실제로 오른쪽 유통사별 방문 빈도를 그래프로 보면 다른 유통사들에 비해 재방문 텀이 긴 것을 확인할 수 있다. 따라서, A05 유통사의 데이터는 앞에서 설정한 관찰 기간 2개월에 맞지 않는다고 판단하여 데이터 셋에서 제거하기로 하였다.

## 02 데이터 전처리 A05, A06 데이터 제거



다음은 A06 유통사이다. 전체 데이터 셋에서 온라인 구매의 비율이 10%밖에 되지 않는다는 것을 앞에서 설명하였다. 따라서 **미방문 고객 예측 시에는 오프라인 고객을 기준으로 해야 하는데, 이때 모든 데이터가 온라인 구매 데이터로 구성된 A06 유통사의 경우는 이에 적절하지 않다고 판단하여 제거하였다.**

## 02 데이터 전처리

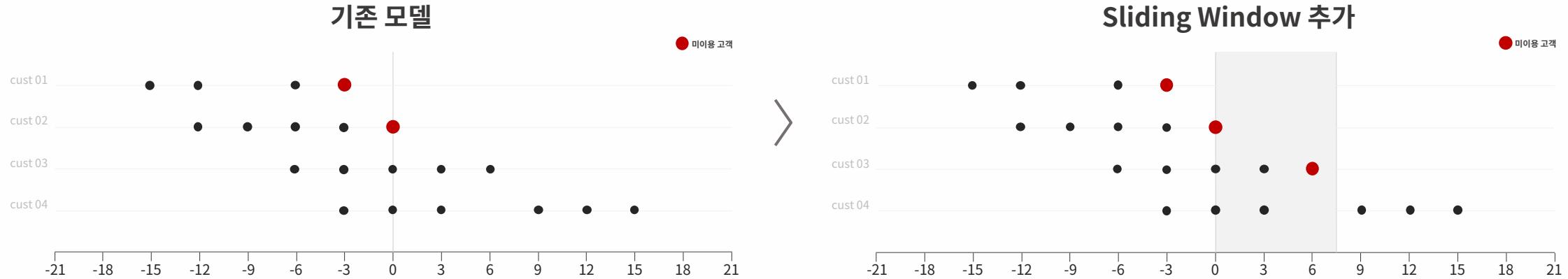
### A05, A06 데이터 제거

“LPOINT\_BIG\_COMP\_02\_PDDE.csv” 파일에서 데이터를 불러와 아래와 같이 2개월 단위로 데이터 셋을 분류하였다. 훈련 데이터 셋을 하나의 시간 기준으로 나누지 않고, 여러 시간 기준을 가져왔다. 이렇게 하게 되면, 데이터의 양이 더 방대해지고, 특정 시점에서의 bias가 줄어들고, overfitting을 방지할 수 있다(Gattermann-Itscherl and Thonemann, 2021).



여기서 교차검증(cross validation)을 사용하지 않고 validation 셋을 따로 두었는데, 학습과 검증 데이터 셋의 시점이 역전되는 것을 막기 위해 교차검증 대신 Validation 셋을 따로 생성하게 되었다.

## 02 데이터 전처리 Sliding Window

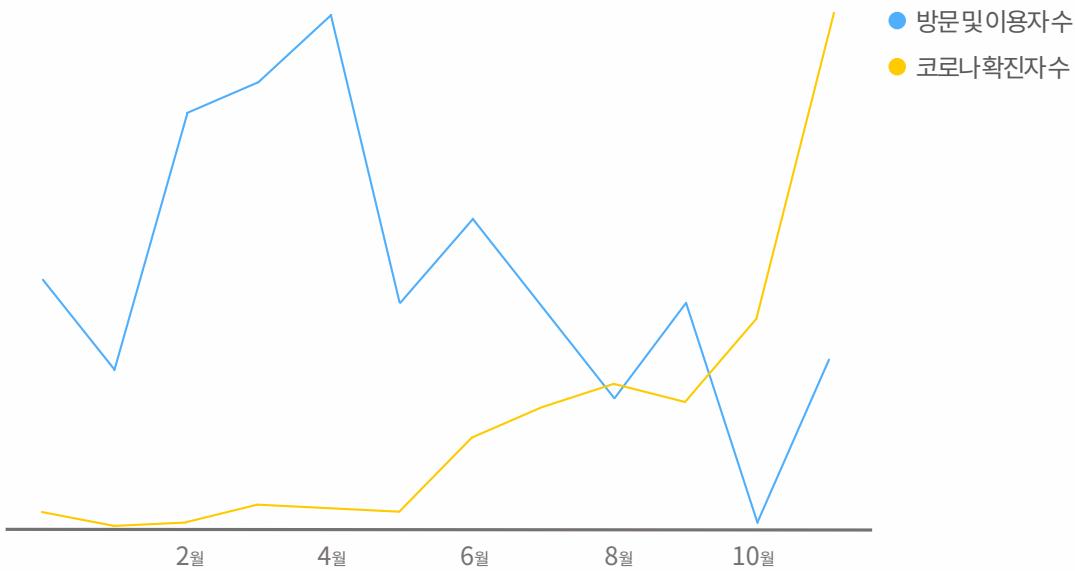


기준 모델은 기준 날짜 이후 거래가 존재하지 않았을 때 미방문 고객으로 분류가 된다. 해당 방법은 매우 엄격한 기준을 가지고 있으며 고객들의 다양한 소비 패턴이 고려되지 않는다.

사람들은 일반적으로 주 단위 생활 사이클을 가지고 있으므로 적어도 7일간의 sliding window(Hadjiyi et al., 2014)를 설정하여 기준 모델의 엄격한 기준을 완화했다. 예외적으로, Sliding window 안에 거래 내용이 존재해도 30일간 연속으로 거래가 없다면 미이용 고객으로 분류된다.

## 02 데이터 전처리 가중치 변수 생성

월 별 방문자 및 코로나 확진자 수 비교



회귀분석 결과

계수	P-값	신뢰구간
-0.04	0.000	-0.059 ~ -0.021

현재 데이터는 2021년도를 기준으로 하고 있으므로 때문에, 코로나 확진자 수와 오프라인 방문자 수 간에 상관관계가 있을지 궁금하여 회귀 분석을 진행하였다. 분석 결과 오른쪽의 표를 보면 회귀 분석 결과 상관관계가 있다고 할 수 있다. 따라서 데이터에 코로나 관련 데이터를 추가하기로 하였다.

## 02 데이터 전처리 가중치 변수 생성

$$Rt = \frac{\text{(감염률)}}{\text{(회복률)}}$$

$Rt > 1$  (확진세)  
 $Rt < 1$  (감소세)

코로나 상황을 고려하여 감염 재생산 지수를 가중치 변수로 사용하였다. 감염재생산지수(Rt)는 감염자가 발생하였을 때 감염자가 평균적으로 감염시킬 수 있는 2차 감염자의 수를 나타낸 것이다. 만약 1 보다 크면( $Rt > 1$ ) 확진세가 1 보다 작으면( $Rt < 1$ ) 감소세인 것으로 해석된다.

감염재생산지수 데이터를 질병관리청에서 제공 받아서, 각 거래 주차 별 감염재생산지수들의 평균을 내어 모델 학습에 사용하였다.

## 02 데이터 전처리

### 사용 변수

변수 이름	설명
A01~A04	유통사 더미 변수
first	2021년도 첫 거래 일
recency	최근 거래 일
frequency	방문 빈도
monetary	총 사용금액
time	평균 방문 시간대
chnl	온라인과 오프라인 구매 비율
Rt	감염재생산지수

id	A01	A02	A03	A04	first	recency	frequency	monetary	time	chnl	Rt
3	1	0	0	0	297	5	2	1,570,000.0	15.25	1.0	0.975
6	1	0	0	0	244	2	8	395,230.0	17.9706	1.0	0.9274
9	1	0	0	0	265	5	7	2,148,600.0	14.1429	1.0	0.99
10	1	0	0	0	276	16	4	325,400.0	17.375	1.0	0.9825
12	1	0	0	0	303	2	7	764,480.0	14.8235	1.0	0.9671
...	...	...	...	...	...	...	...	...	...	...	...
56884	0	0	0	1	15	9	2	18,000.0	7.0	1.0	0.87
56889	0	0	1	0	283	1	1	16,710.0	4.0	2.0	1.06
56890	0	0	0	1	226	4	13	77,700.0	10.4848	1.0	0.9967
56894	1	0	0	0	297	10	4	1,792,500.0	16.0	1.0	0.8933
56895	0	1	0	0	297	2	3	123,320.0	18.125	1.0	1.0438

## 02 데이터 전처리

### RFM 변수를 사용한 고객 세분화

수만 명에 이르는 고객 개개인에 대하여 대응하는 것은 효율적이지 못하다. 따라서 예측 분석을 진행하기 전에 고객 세분화를 하여 기업에 이득이 되는 고객과 그렇지 않은 고객을 분류하였다.

클러스터	최근성(평균)	빈도(평균)	금액(평균)	RFM
0	3.84	3.17	399,091.19	R↓F↓M↓
1	55.07	1.17	138,196.10	R↑F↓M↓
2	30.59	2.04	231,337.59	R↑F↓M↓
3	2.92	22.63	1,437,950.43	R↓F↑M↑
4	4.14	10.34	1,152,868.73	R↓F↑M↑
5	42.306	1.60	191,673.90	R↑F↓M↓
6	11.855	3.31	378,942.71	R↓F↓M↓
7	20.785	2.70	321,332.15	R↑F↓M↓
전체 평균	18.97	4.36	462,030.12	

고객 세분화를 위해 최근성(R: recency), 방문빈도(F: frequency), 그리고 사용금액(M: monetary value) 변수들을 사용하여 K-Means 클러스터링을 실시하였다. 위 표에서  $\uparrow$ ,  $\downarrow$ 는 각 RFM이 평균보다 크거나 작은지를 나타낸다. RFM과  $\uparrow$ ,  $\downarrow$ 를 조합했을 때 나올 수 있는 총 경우를 고려하여 k=8로 클러스터링을 하였다(Liu and Shih, 2005).

## 02 데이터 전처리

### RFM 변수를 사용한 고객 세분화

Cluster	R	F	M
0, 6	↓	↓	↓
3, 4	↓	↑	↑
1, 2, 5, 7	↑	↓	↓

클러스터링한 결과 위 표처럼 총 3가지 특성으로 분류되었다. 여기서 3번째 행에 해당하는 클러스터를 해석해 보자면, 최근에 방문한 기록이 없고(R↑) 방문 빈도(F↓)와 사용 금액(M↓)이 평균보다 낮다. 해당 클러스터는 기업의 입장에서 이득이 되지 않는 집단이라 생각할 수 있으므로 관찰 대상에서 제외하였다.

모델링

# 03 모델링

## 모델 선택

### 사용한 모델: XGBoost

#### 사용한 이유

1. 과적합(Overfitting) 방지를 위한 여러 파라미터를 제공한다.
2. 병렬처리로 학습 및 분류 속도가 빠르다.
3. 불균형 데이터(Imbalanced data) 세트를 처리할 수 있는 파라미터(scale\_pos\_weight)를 제공한다.

종속 변수의 클래스 비중이 약 2.5:1정도로 scale\_pos\_weight 파라미터 값을 2.5로 설정하였다.

# 03 모델링

## 모델 성능 평가

### Validation

	TRUE	Pred	FALSE
TRUE	19,528	4,858	
Act			
FALSE	3,928	6,718	

### Test

	TRUE	Pred	FALSE
TRUE	9,014	3,412	
Act			
FALSE	1,893	4,666	

Precision	Recall	F1 점수	정확도
0.58	0.63	0.60	0.75

Precision	Recall	F1 점수	정확도
0.58	0.71	0.64	0.72

## 03 모델링 적정 마케팅 비용 측정

모델의 정확도뿐만 아니라 실제 모델을 비즈니스에 적용하였을 때 이익을 최대화하는 적정 마케팅 비용을 검증하기 위해 Profit 지수(Lee et al., 2018)를 사용하여 모델을 검증하였다.

$$\text{Profit} = \text{CLV}(\gamma \times \text{TP}) - \text{C}(\text{TP} + \text{FP})$$

**CLV** = 고객생애가치

**C** = 개인별 마케팅 비용

**$\gamma$**  = 전환율

우리가 미방문으로 예측한 고객 중 실제로 방문한 고객(FP)은 마케팅과 상관없이 방문이 예정된 고객이기 때문에 추가적인 이득 없이 비용만 낭비된다. 그리고 고객마다 생애가치가 다름으로 모델 성능에 따라 마케팅 비용을 다르게 책정할 필요가 있다.

## 03 모델링 적정 마케팅 비용 측정

$$\text{Profit} = \text{APV} \times \text{TP} - \text{C}(\text{TP} + \text{FP})$$

$$\text{APV} = \frac{\text{총 구매금액}}{\text{총 방문 횟수}}$$

기존 Profit 지수는 고객 이탈 방지를 위한 이탈예측모델을 평가할 때 사용된 지수이다. 하지만 우리의 목적은 고객이탈 방지가 아닌 다음 월 방문 여부를 예측하여 기존 고객들의 적극적인 참여를 유도하는 것이다. 마케팅이 성공적으로 진행되어 방문 유도에 성공하였을 때 예상되는 평균구매가치(APV)로 대신하는 것이 맞다고 판단하였고, 개인별 적정 마케팅 비용(C)을 산정하였다. 해당 지수로 구해진 적정 마케팅 비용을 통해, 개인화 마케팅에 최종적으로 적용하였다.

전환율은 현재 데이터상에서는 확인이 불가하여, 지수 식에서 제외하였고 유통사에서 발생하는 재고유지비용과 판매 수수료 등을 고려하여 순매출액을 매출액의 25%라 가정하였다.

## 03 모델링 적정 마케팅 비용 측정

$$ROI = \frac{\text{이익금액}}{\text{투자금액}}$$

기업은 현재 ROI(투자자본수익률) 10% 달성을 목표로 하고 있음을 가정하였고 다음과 같은 결과를 도출해냈다.

### Validation

개인별 마케팅 비용	총 마케팅 비용	순이익
15,832원	168,545,235원	16,854,523원

### Test

개인별 마케팅 비용	총 마케팅 비용	순이익
18,362원	120,434,170원	12,043,417원

# 개인화 마케팅

## 04 개인화 마케팅

### 마케팅 전략 제안

#### 개인화 마케팅 전략

개인화 마케팅에서는 추천시스템을 통해 기존 사용자와 비슷한 구매 성향을 보인 사용자들의 데이터를 이용하여 사용자에게 기존에 구매하지 않은 새로운 상품을 추천한다. 이러한 추천은 쿠폰 제공이 동반하며, 쿠폰 제공을 통해 사용자의 구매 욕구를 촉진하고, 고객의 구매 물품 종류가 다양해져 방문 빈도를 높이게 될 것이다.

## 04 개인화 마케팅

### 추천 시스템 모델링

추천 시스템의 기본적인 동작 방식은 아래의 **고객-아이템 행렬**에서 “?(기존에 구매하지 않았던 제품)”로 표시된 부분들의 값을 유사한 사용자와의 비교를 통해 예측 후, 예측된 점수가 높은 순서대로 해당 사용자에게 추천해주는 방식이다.

cust	Item A	Item B	Item C	...
1	2	?	?	...
2	?	2	3	...
3	?	3	5	...
4	4	?	1	...
...	...	...	...	...



cust	Item A	Item B	Item C	...
1	2	2	4	...
2	4	2	3	...
3	1	3	5	...
4	4	5	1	...
...	...	...	...	...

## 04 개인화 마케팅

### 추천 시스템 모델링

고객 - 아이템 행렬에서 각 열에 들어가는 값들은 고객이 해당 열에 있는 아이템에 대해서 평가를 한 정보가 들어간다. (구매 후기 평점 등) 하지만, 현재 데이터 셋에는 사용자가 구매한 물건에 대해 평가를 한 정보가 들어있지 않다. 따라서 우리는 구매 빈도를 대신하여 사용하였다. 고객이 같은 물건을 여러 번 구매하면 그 물건에 대해 좋은 평가를 한 것이라 생각하여 아래와 같이 구매 횟수에 따라 평가 점수를 부여했다.

The diagram illustrates the transformation of a wide user-item matrix into a sparse rating matrix. On the left, a wide matrix has columns for 'cust' (customer ID), 'item' (item ID), and '...' (ellipsis). It contains five rows of data: (1, A, ...), (1, A, ...), (2, B, ...), (3, B, ...), and (4, A, ...). An arrow points from this matrix to a more compact, sparse matrix on the right. The right matrix has columns for 'cust', 'Item A', 'Item B', and '...'. It contains five rows of data: (1, 2, 0, ...), (2, 0, 1, ...), (3, 0, 1, ...), (4, 1, 0, ...), and (... , ..., ..., ...).

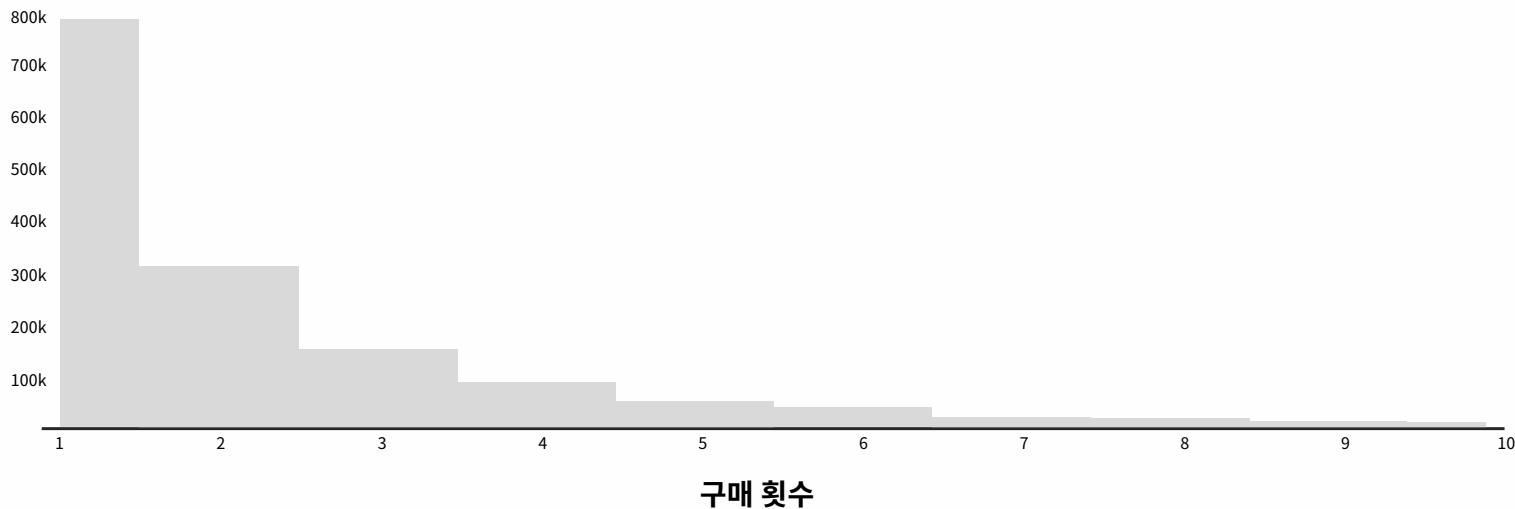
cust	item	...
1	A	...
1	A	...
2	B	...
3	B	...
4	A	...

>

cust	Item A	Item B	...
1	2	0	...
2	0	1	...
3	0	1	...
4	1	0	...
...	...	...	...

## 04 개인화 마케팅 추천 시스템 모델링

데이터에서 아이템 별 구매 횟수가 10이 넘어가는 경우가 전체의 5% 정도 되기 때문에, 대부분의 동일한 물건 재구매 횟수가 10회 이하라 생각하여 10 이상인 경우 10으로 고정한다.



## 04 개인화 마케팅 추천 시스템 모델링

데이터를 정리해본 결과 데이터 셋 내의 고객과 아이템 수는 다음과 같다.

유저: 26,917명

아이템: 1,933개

이를 기본적인 협업 필터링 방식(Memory based approach)으로 하려면

$26,917 \times 1,933 = 52,030,561$  개의 파라미터가 필요하다.

파라미터의 수가 너무 많기 때문에 파라미터 수를 줄일 수 있는 방식을 선택해야 한다.

파라미터를 적게 사용하는 방법에서 아래와 같이 2 가지를 선택하였다.

1. 잠재 요인 기반 협업 필터링 (Truncated SVD)
2. 최대우도추정을 이용한 랜덤 샘플링

## 04 개인화 마케팅

### 추천 시스템 모델링

두 번째, 잠재 요인 기반 협업 필터링의 결과이다.

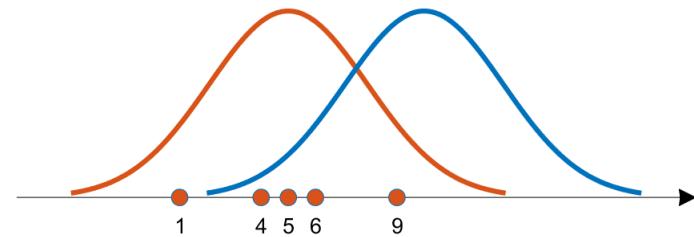
추천 항목 결과

id	중분류명	예측 평가 점수
1	유아의류하의	6.53
2	남아특수목적의류	4.29
3	어류	4.07
4	조개류	4.02
...	...	...

하지만 여기서 개인별 추천 항목이 거의 모든 유저에게서 비슷하게 나타나서, 개인화 마케팅이 불가능해 실제로 적용하기에는 어려웠다.

## 04 개인화 마케팅 추천 시스템 모델링

두 번째, 최대 우도 추정 기반 추천 시스템이다.



최대 우도 추정이란 간단하게 위의 그래프를 보면 Point 1~9의 원래 분포가 파란색 보다는 빨간색일 가능성이 더 높다는 것을 직관적으로 알 수 있다. 이처럼, 일 부분의 데이터를 분석하여 모평균과 모분산을 추정하는 것이다.

## 04 개인화 마케팅

### 추천 시스템 모델링

Surprise 모듈에는 최대 우도 추정을 이용한 추천 시스템 모델을 만들어주는 `NormalPredictor()` 클래스가 있다.

The prediction  $\hat{r}_{ui}$  is generated from a normal distribution  $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$  where  $\hat{\mu}$  and  $\hat{\sigma}$  are estimated from the training data using Maximum Likelihood Estimation:

공식 API의 설명에도 나와있듯이 최대 우도 추정을 통해 구한 모평균과 모분산으로 만든 정규 분포 내에서 랜덤으로 샘플링을 해준다.

## 04 개인화 마케팅

### 추천 시스템 모델링

현재 데이터셋에 적용한다면 아래와 같이 해석할 수 있다.

cust	Item A	Item B	...
1	2	0	...
2	0	1	...
3	0	1	...
4	1	0	...
...	...	...	...

사용자의 아이템 평가 데이터로부터 최대 우도 추정을 통해 만든 정규분포에서 랜덤하게 샘플하여 아직 평가가 없는 부분(0)의 값을 예측하는 방식이다.

## 04 개인화 마케팅 추천 시스템 모델링

앞서 언급한 surprise 모듈의 NormalPredictor()를 통해 만든 추천 시스템의 평가 결과는 아래와 같다.

	Fold1	Fold2	Fold3	Fold4	TEST
RMSE	2.49	2.49	2.49	2.49	2.5
MSE	6.21	6.22	6.22	6.22	6.25

이제 완성된 추천 시스템에 사용자 아이디를 넣으면, 사용자가 구매하지 않은 물품 중에서 평가 점수가 높게 나온 물품들을 추천해주게 된다. 아래는 앞에서 만든 이탈고객 모델에 9~11월 사용자 중 이탈고객으로 분류된 사용자들을 추천시스템에 넣어서 추천을 받은 결과이다.[사용자아이디, 추천항목]

```
[500003, ['여아패딩', '볶음반찬', '남성슬립온', '풍선', '도서/교육', '중식', '남성골프바지', '소프트캔디', '남성백팩', '텐트']]  
[500019, ['자전거기타액세서리', '할로윈완구', '남성캐쥬얼스포츠양말', '채색도구', '막걸리', '호주산소목심', '서류정리용품', '기타탁구용품', '유아동일반양말', '기타란']]  
[500021, ['조리도구세트', '스냅백', '여성스포츠티셔츠/탑', '냉동흉시', '랩스터', '미국산소토시살', '바닥재', '호빵찐빵', '기타테니스용품', '과채혼합음료']]  
[500028, ['전화기', '여성트렌치코트', '기타자동차가전기기', '족욕기', '한우정육', '음반', 'TV장식장/거치대', '바디케어세트', '오리고기', '유아동스포츠전신/원피스']]
```

## 결론 및 기대효과

본 프로젝트는 구매 데이터를 기반으로 해, 소비자의 내달 재방문 여부를 예측하고, 이탈로 예측된 고객들을 개인화 마케팅을 통해 회유하는 것을 목표로 하였다. 이 과정에서 profit지수를 통해서 개인별 마케팅 비용을 산정하였다. 이렇게 산정된 마케팅 비용과 상품 추천 시스템을 통해 개인화 마케팅에 사용할 수 있을 것이다.

## 참고문헌

Gattermann-Itscher, T., & Thonemann, U. W. (2021). How training on multiple time slices improves performance in churn prediction. *European Journal of Operational Research*, 295(2), 664-674.

Hadjii, F., Sifa, R., Drachen, A., Thurau, C., Kersting, K., & Bauckhage, C. (2014, August). Predicting player churn in the wild. In 2014 IEEE Conference on Computational Intelligence and Games (pp. 1-8). ieee.

Lee, E., Kim, B., Kang, S., Kang, B., Jang, Y., & Kim, H. K. (2018). Profit optimizing churn prediction for long-term loyal customers in online games. *IEEE Transactions on Games*, 12(1), 41-53.

Liu, D. R., & Shih, Y. Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, 42(3), 387-400.

## 데이터 출처

### 코로나 확진자 데이터

KDX한국데이터거래소. (2022) 코로나바이러스감염증-19(COVID-19) 현황 누적 데이터,  
<https://kdx.kr/data/view/25918>

### 감염재생산지수

질병관리청 (정보공개포털)