

# 유아용품 온라인 고객 맞춤형 상품추천 시스템을 통한 매출 증대



A1조 류지환 박미령 박선우 신정윤 유태욱 한수빈

## 온라인 고객 맞춤형 상품추천 시스템을 통한 유아용품 매출 증대



- **회사 개요**

- 한국의 대표적인 기업 중 하나로, 유아용품을 중심으로 다양한 생활용품을 제조 및 판매하는 회사
- 1963년에 설립되어서 지금까지 지속적으로 성장해온 기업이며, 특히 기저귀와 같은 유아용품 분야에서 강한 신뢰를 받고 있음



- **대표 브랜드 : 하기스**

- 대한민국 No.1 기저귀 브랜드
- 대한민국 부모들이 믿고 쓰는 기저귀
- 국내 시장 점유율 연속 1위
- 40년간 아기에게 가장 좋은 기저귀만을 연구



- **고품질 제조와 연구 개발**

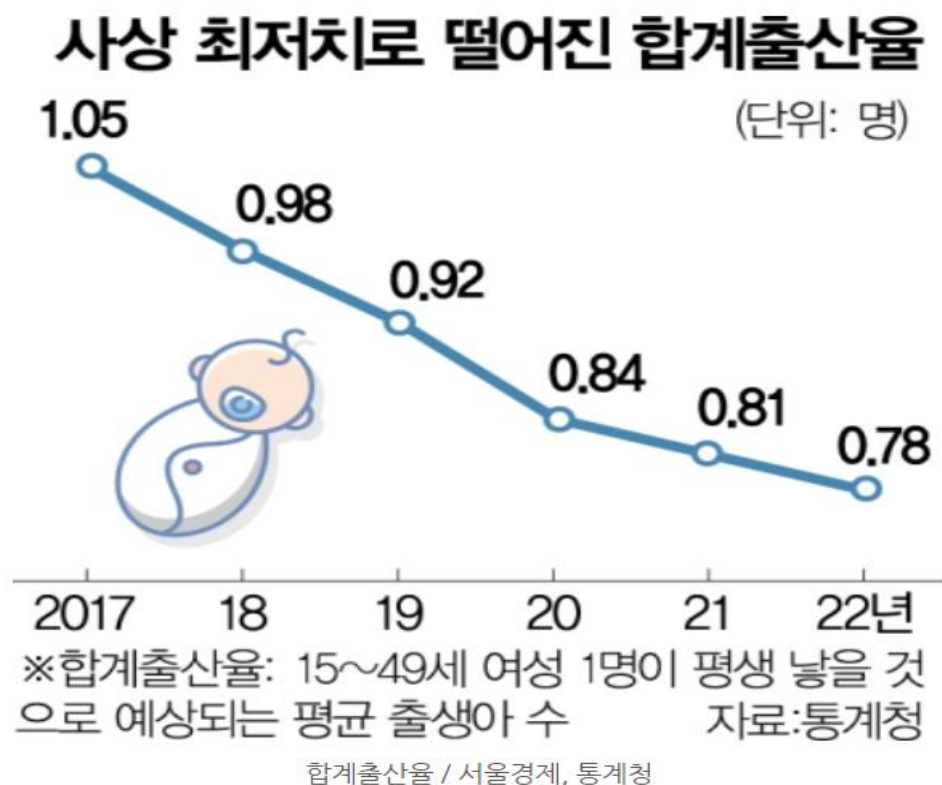
- 유아용품 제조에 있어 피부 친화성과 최고 품질을 고려한 연구와 개발을 진행
- 지속적인 연구와 개발로 기존 제품의 품질을 향상시켜, 부모들이 더 나은 제품을 선택할 수 있도록 지원

# 1. 추진 배경

3

출산율 감소에도 불구하고 유아용품 온라인 시장 수요가 증가 추세이나, **당사의 22년 매출액은 지속적으로 하락**하여 경영 환경이 악화

## 지속적인 출산율 감소



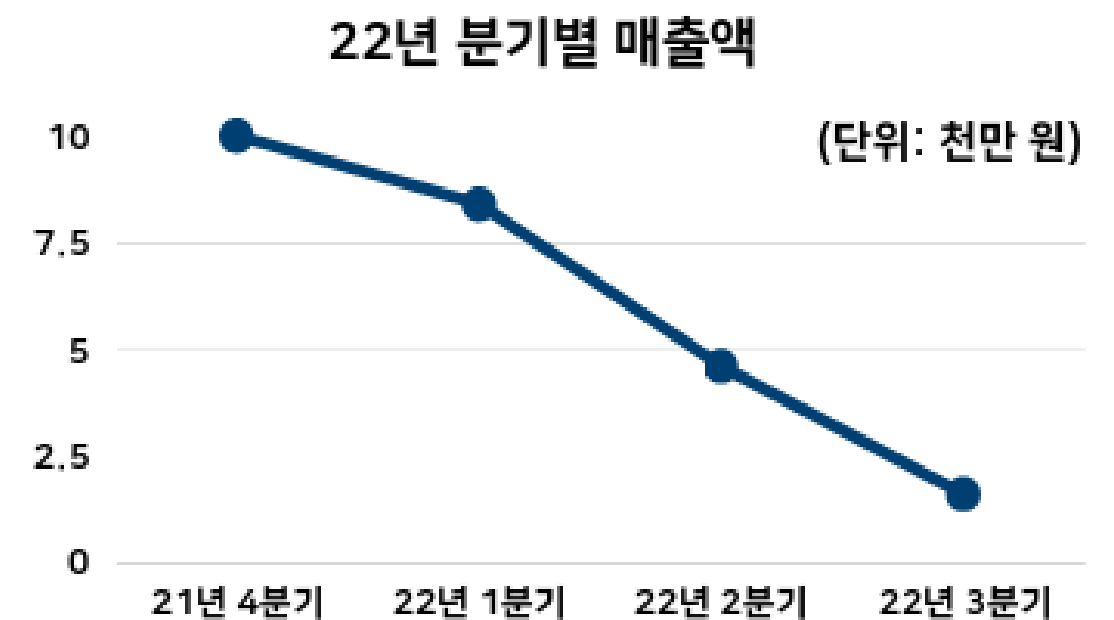
지속적인 출산율 감소로  
유아용품 시장에 변화가 있을 것이라는 전망

## 유아용품 온라인 구매 증가



출생률 감소에도 불구하고, **골드키즈 현상**  
으로 유아용품 온라인 구매 시장이 성장

## 당사 매출액 지속 감소



대형 유통사들의 온라인 시장 입점으로  
경쟁이 점점 심화되며, **매출액 감소**로 이어짐

매출 감소를 막기 위해 **차별화된 마케팅 및 고객 맞춤형 프로모션 전략**이 필요함

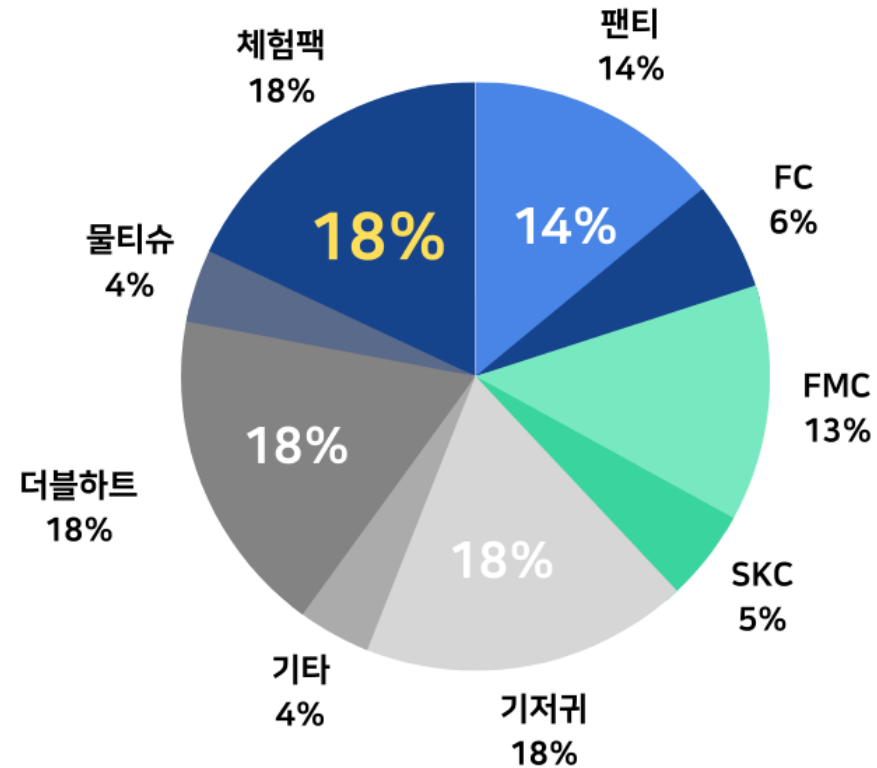
## 2. 현상 및 개선기회

신규 고객의 유입을 위한 콘텐츠 부족, 낮은 재구매율, 배송 관련 불만 증가로 인하여 당사의 매출액 감소에 영향

### 고객 유입 프로모션 콘텐츠 부족

현재 체험팩을 통한 마케팅이 존재하나,  
이를 발전시켜 고객을 유인할 콘텐츠 부족

#### 물품 대분류별 판매량 비율

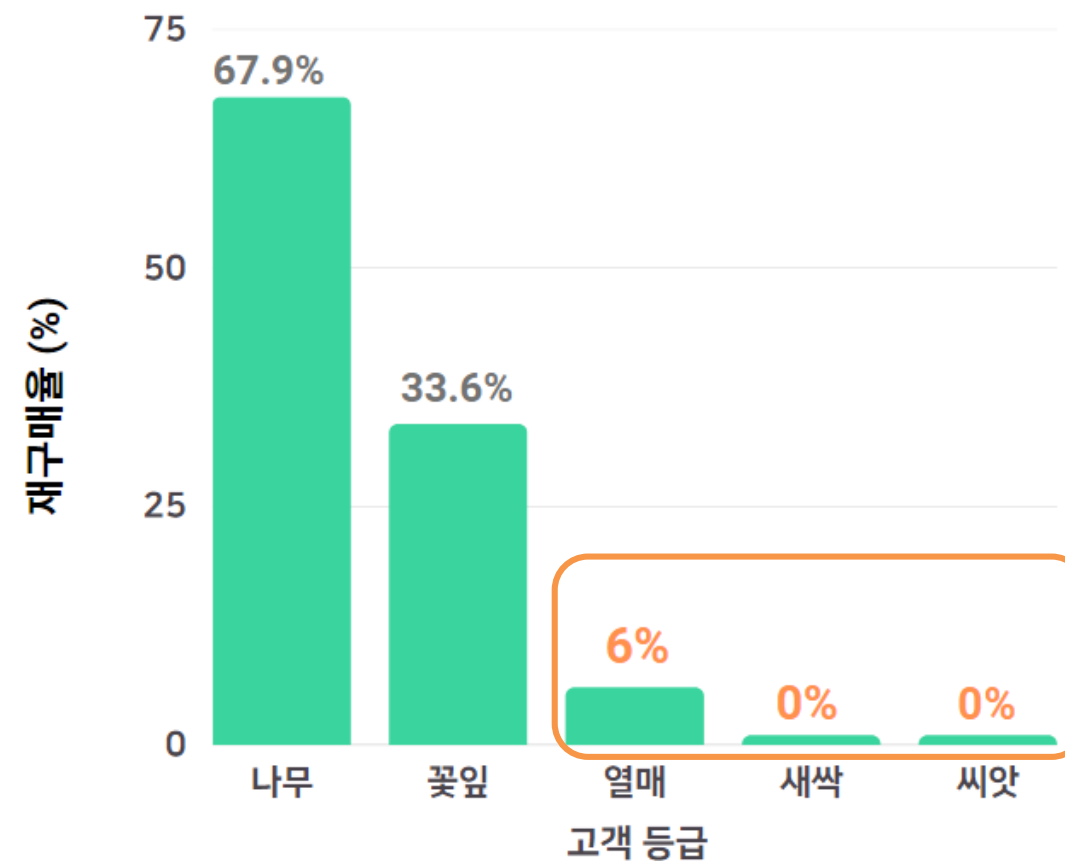


▶ 신규고객 유입을 위한  
프로모션 콘텐츠 필요

### 낮은 등급 회원 재구매율 부족

회원 재구매율 분석이 부족하여 고객 충성도를 높이지 못함

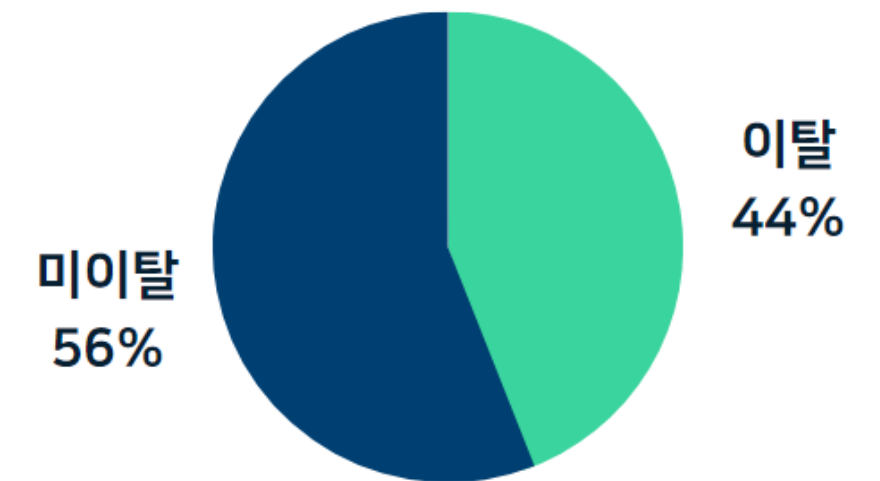
#### 고객 등급 별 재구매율



▶ 고객의 특성 분석을 통하여 하위등급의  
재구매를 유도하는 프로모션 필요

### 배송 관련 고객 불만 증가

온라인 시장 수요 관리 실패에 따라  
온라인 배송지연에 따른 고객의 불만 증가



▶ 물품 적기 조달 방안을 마련하여  
고객불만 감소를 통해 고객이탈방지

개선목표 : 향후 1년 내 현재 매출 (21.7억원) 대비 10% 이상(약 2.1억원) 향상 목표

# 3. 데이터 분석 계획

## 1) 데이터셋

- Member\_data (81360 \* 11 ) : 고객 회원정보 데이터셋    - Sales\_data (199999 \* 12 ) : 온라인 쇼핑몰 주문 내역 데이터셋    - Product\_data ( 3084 \* 5 ) : 판매 상품 정보 데이터셋

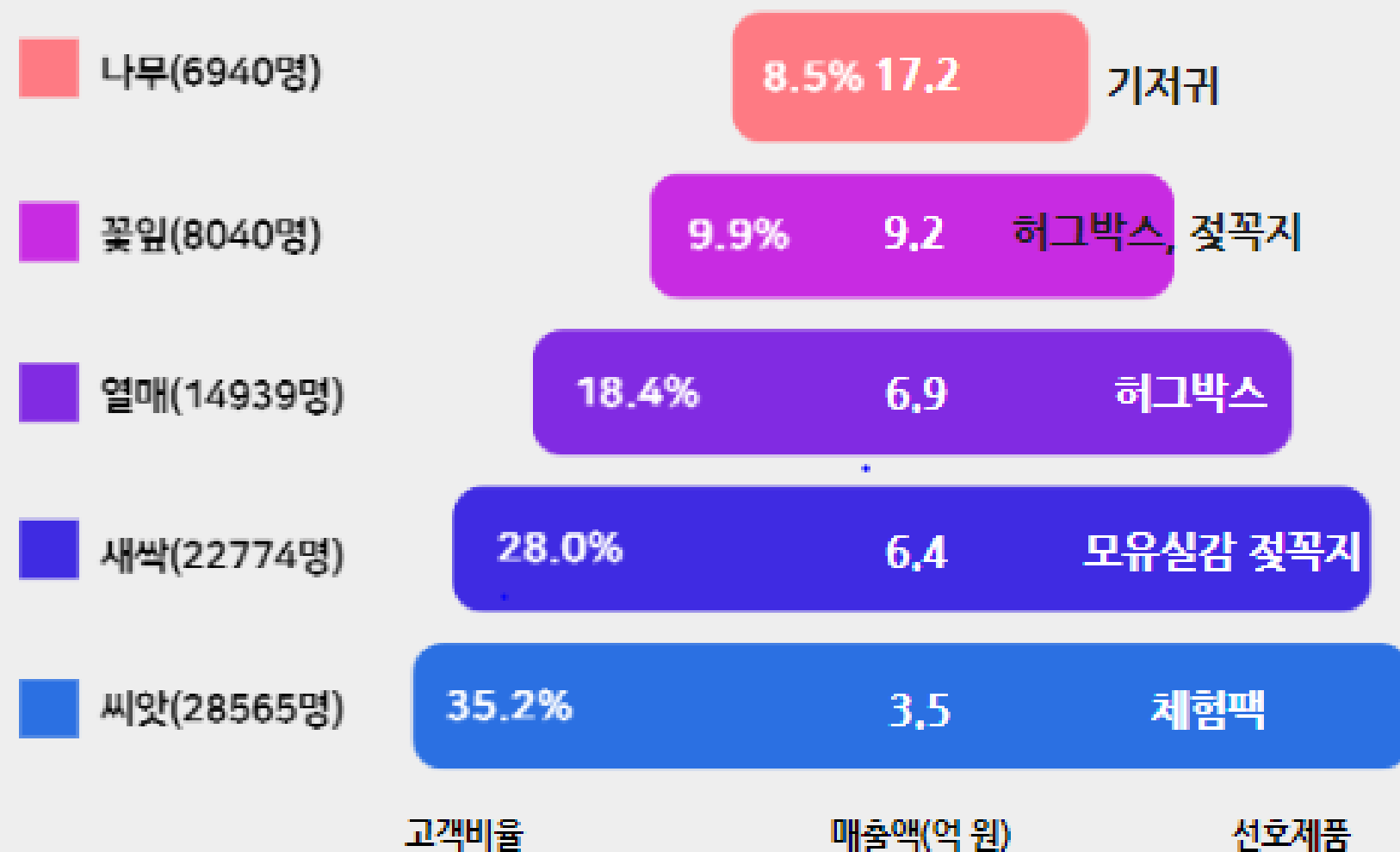
## 2) 데이터 분석계획

목적	분석방법	주요내용	담당자	일정
신규고객 유입을 위한 프로모션 콘텐츠 필요	bar chart pie plot	체험팩의 월령별 상품 판매량 및 매출액을 확인	한수빈	8/15
	Shapiro-Wilk paired t-test	체험팩 사용 이전과 이후의 판매량 및 매출액 확인	한수빈	
	bar chart Pie plot	고객 유입경로(SNS), 회원의 거주지역 및 결제카드 확인	류지환	
상품추천 시스템을 통한 기존회원 유지 및 재구매율 증대	연관규칙	월령별 상품 판매기록을 확인하여 시기에 맞는 상품 추천 모델링	신정윤	8/15
	Decision Tree, RandomForest, XGB	고객의 재구매여부 예측 모델링	유태욱	
	bar chart 꺾은선 그래프	고객의 등급별 연령대, 자녀월령 분포 확인	유태욱	
	bar chart	연령별, 월령별 물품대분류별 매출액 및 판매량 확인	신정윤	
고객 불만사항을 예방하여 고객이탈감소	Logistic Regression, XGB, RandomForest, LGBM	고객 이탈가능성과 주요 요인 확인을 위한 예측모델	박미령	8/15
	Histogram, Pie chart	배송 지연에 따른 고객 이탈 여부 파악	박선우	
	ARIMA	시계열 분석을 통한 제품 수요 예측	유태욱	8/15



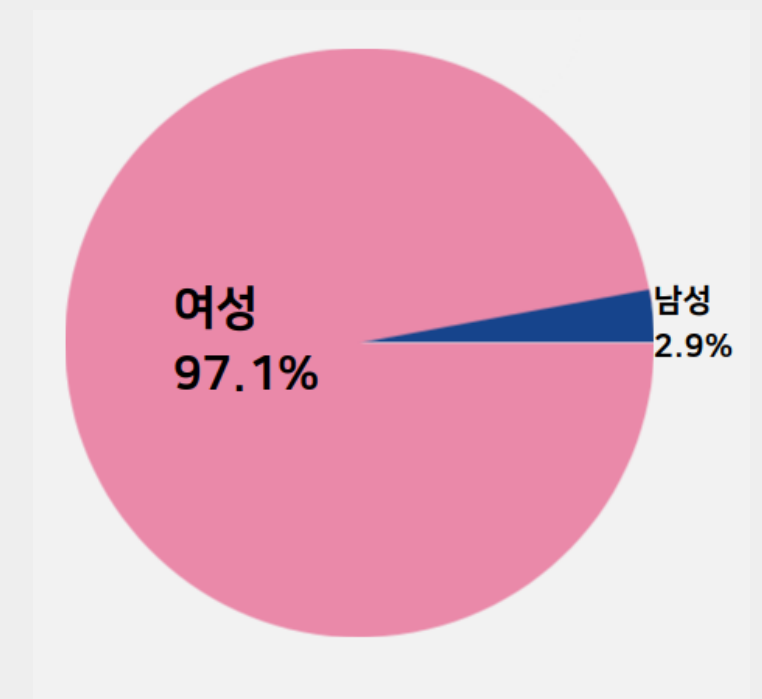
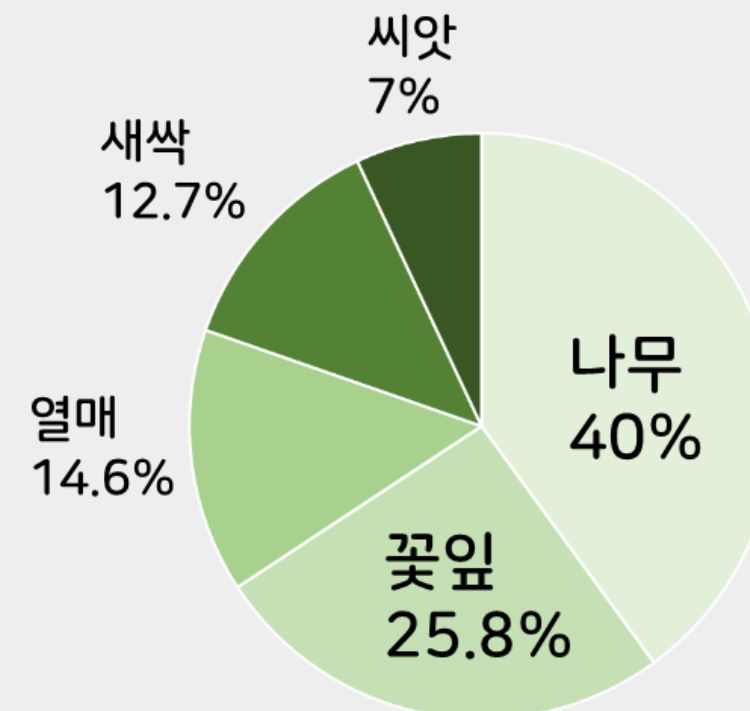
## 4. 분석결과 - RFM

RFM이란 고객이 얼마나 최근에 상품을 구입했고(R), 고객이 얼마나 자주 상품을 구입하고(F), 고객이 지출한 총 금액은(M) 얼마인가에 관한 정보를 활용하여 구입 가능성이 높은 고객들을 선별하는 모델링의 구성요소



- ▶ 당사 RFM 점수 =  $0.4 * R + 0.3 * F + 0.3 * M$
- ▶ 고객 선호도와 구매 이력에 기반 개인화된 추천 시스템을 도입하여 고객에게 맞춤형 제안을 제공

- 상위 2개 등급(나무, 꽃잎)의 고객이 전체 매출의 상당수(65.8%)를 차지
- 상위 2개 등급(나무, 꽃잎)의 고객의 97.1%가 자녀를 가진 30대 여성



- ▶ 전체 매출의 대부분을 차지하는 나무, 꽃잎 등급의 자녀를 가진 30대 여성을 타겟으로 프로모션 적용

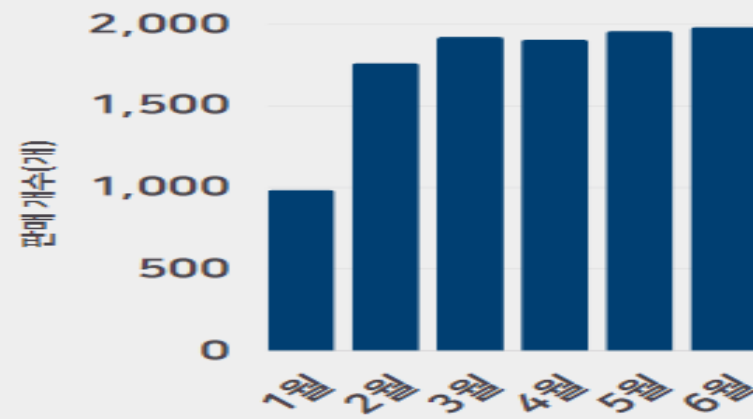
고객 층별화를 통한 등급 별 맞춤형 추천 시스템 도입 및 주요 고객 파악

## 4-1. 분석결과 - 고객 유입을 위한 프로모션 콘텐츠 도입

7

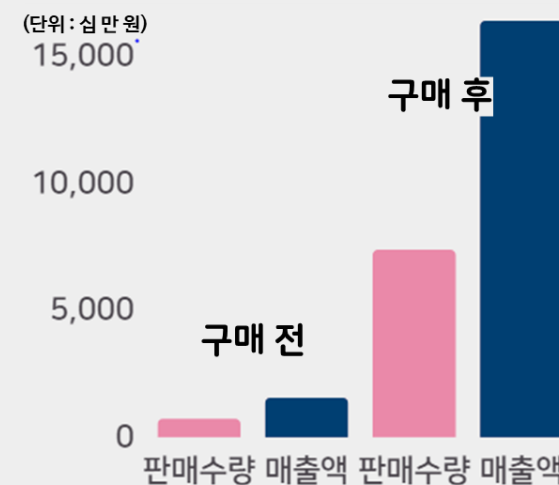
### 체험팩을 이용한 프로모션

22년 체험팩 판매수량

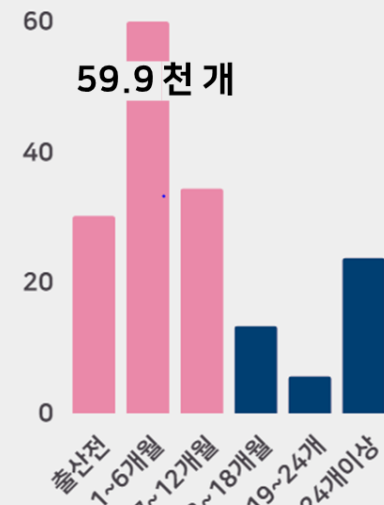


- ▶ 자사의 판매수량이 줄어들어도 불구하고 **체험팩 판매수량은 증가** 및 유지하고 있음  
체험팩은 고객에게 꾸준히 판매되는 상품이며 이는 중요한 요소임을 확인

[ 체험팩 구매 전/후 매출액 및 판매수량 변화량 ]



[ 자녀 월령별 판매량 ]



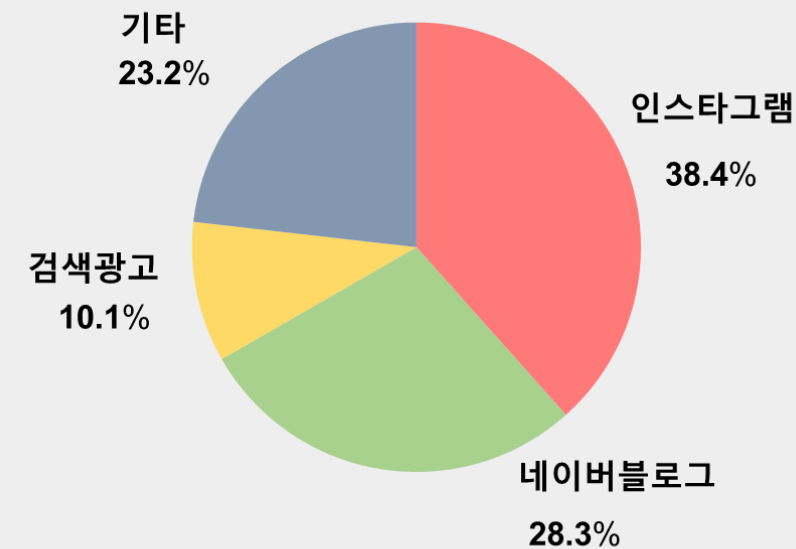
- ▶ 체험팩 구매 유무에 따라 매출액과 판매수량이 **약 10배 증가**  
즉, 1년간 매출액이 약 1.48억 증가 (연간 매출액 6%상승 효과)
- ▶ **출산 전 ~ 12개월 자녀를 가진 고객에게** 체험팩 마케팅 가능성을 확인

자녀 연령대가 낮은 신규고객의 판매량 증대를 위해,  
체험팩 제품을 이용한 차별화 콘텐츠 필요

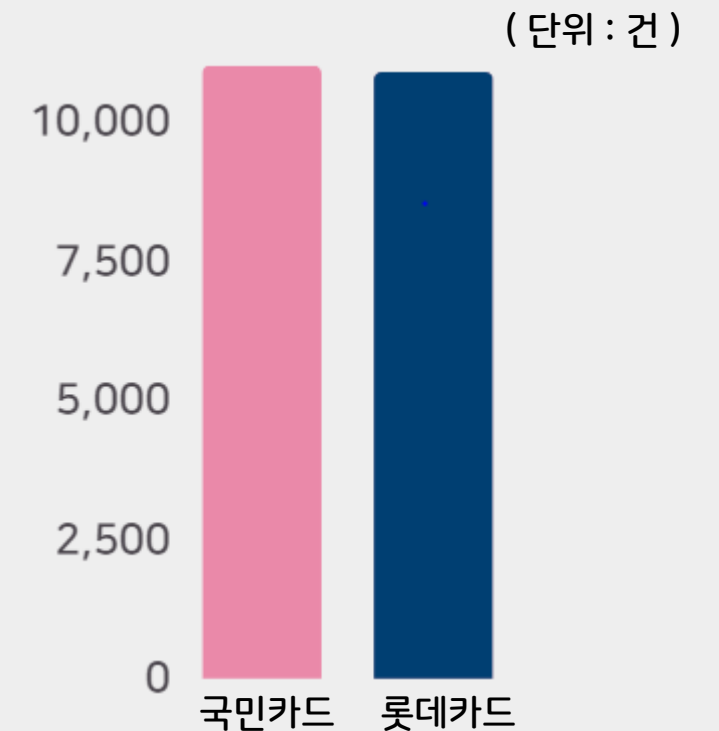
### 주요 고객 마케팅 전략

- 주요고객인 30대 여성이 선호하는 유입경로 및 카드사 분석

[ 30대 고객 유입경로 ]



[ 결제 카드사 별 주문 수 ]



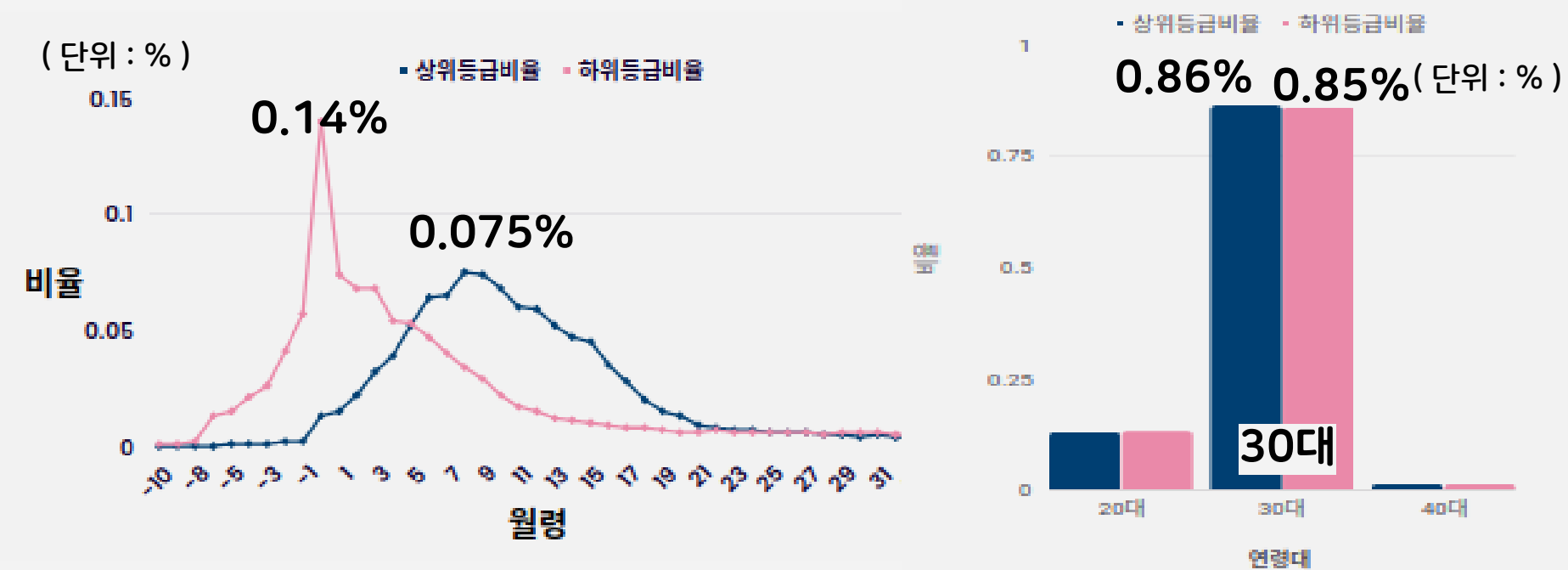
- ▶ 주 고객의 유입경로는 인스타그램과 네이버 블로그의 유입이 많은 것을 확인
- ▶ 주 고객이 이용하는 카드사는 국민카드와 롯데카드가 많은 것을 확인

주요 고객층의 선호를 통해 결제경로를 파악해  
맞춤 마케팅과 혜택 방안 필요



## 4-2. 분석 결과 - 상품 추천 시스템을 통한 기존회원 유지 및 재구매율 증대

### 등급별 자녀 월령, 연령 분포



#### 하위 등급 고객은 자녀 월령이 0개월에 치중

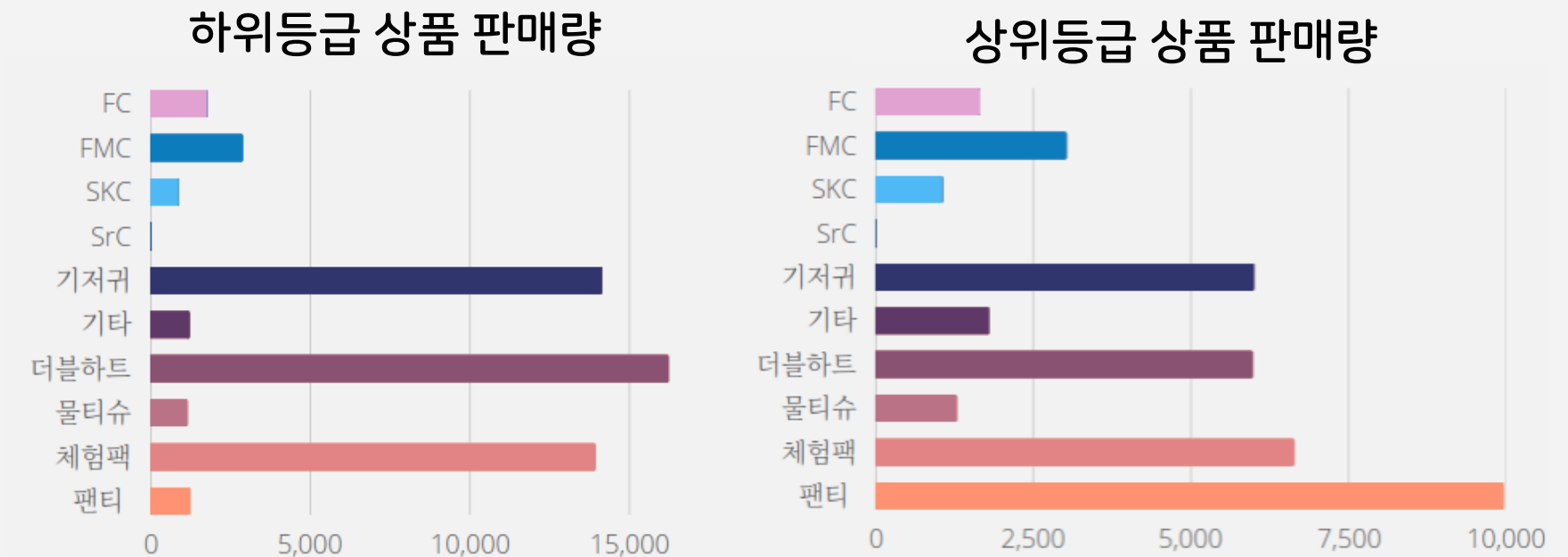
하위 등급 고객의 자녀는 주로 0개월 근처에 분포  
상위 등급 고객의 자녀는 조금 더 넓은 범위로 9개월 근처에 분포



#### 등급별 연령 차이는 없음

소비 연령대에서는 낮은 등급, 높은 등급에서 모두 동일하게 나타남,  
30대가 대다수를 차지

### 물품대분류 별 판매 수량 비교



#### 하위등급 주 고객층 상품 분포

하위 등급 고객층은 체험팩과 대표 상품인 기저귀, 더블하트 제품을 주로 구매  
팬티 종류의 제품 구매는 저조



#### 상위등급 주 고객층 상품 분포:

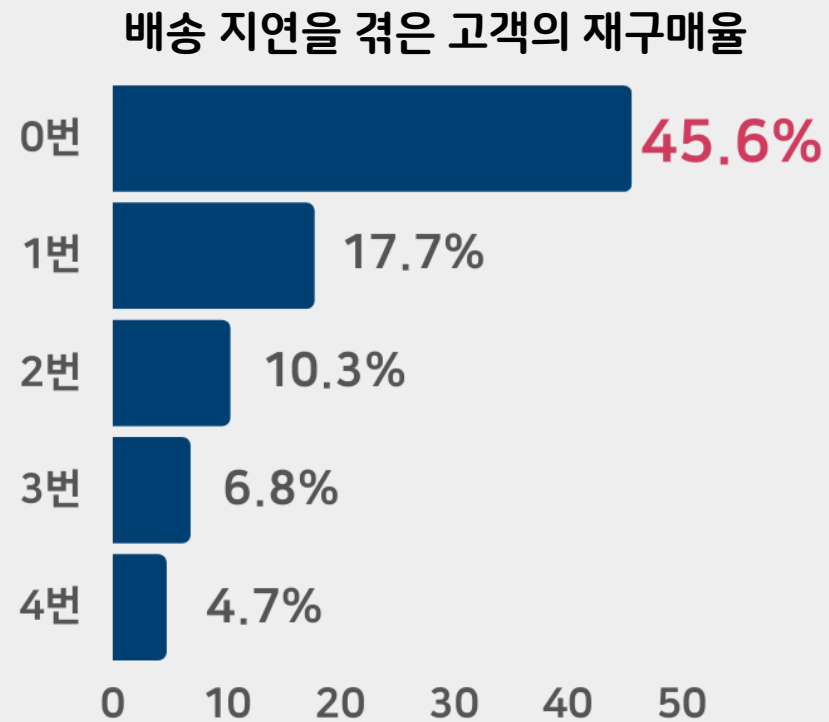
상위 등급 고객층은 하위 등급에 비해 팬티 제품군의 구매가 압도적으로 높음

재구매율 증대를 위해 등급별 주 고객층의 상품 수요 예측을 통한 상품 추천 시스템 필요

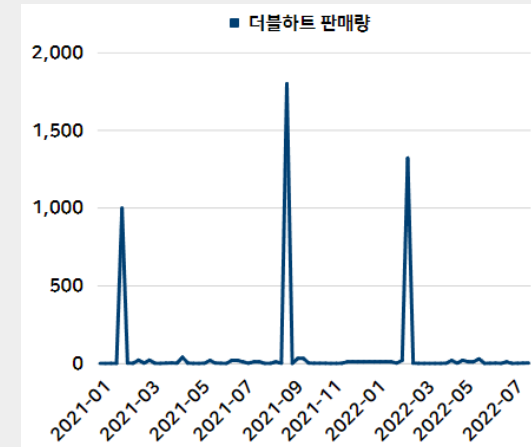
## 4-3. 분석 결과 - 택배 배송 지연에 따른 고객 불만족

배송 지연을 겪은 고객은 재구매율이 매우 낮음

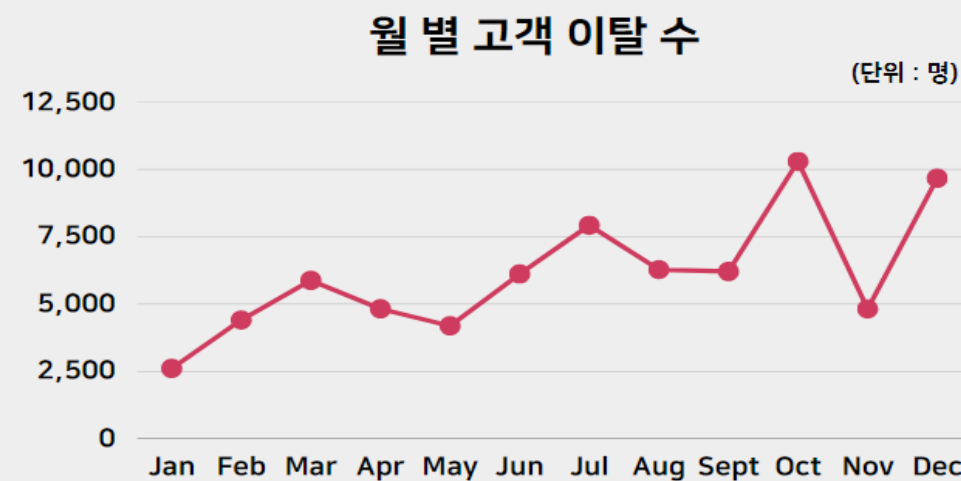
→ 배송 지연 상품 파악



배송지연이 심한 품목을 찾고  
ARIMA Model로 수요 예측 실시



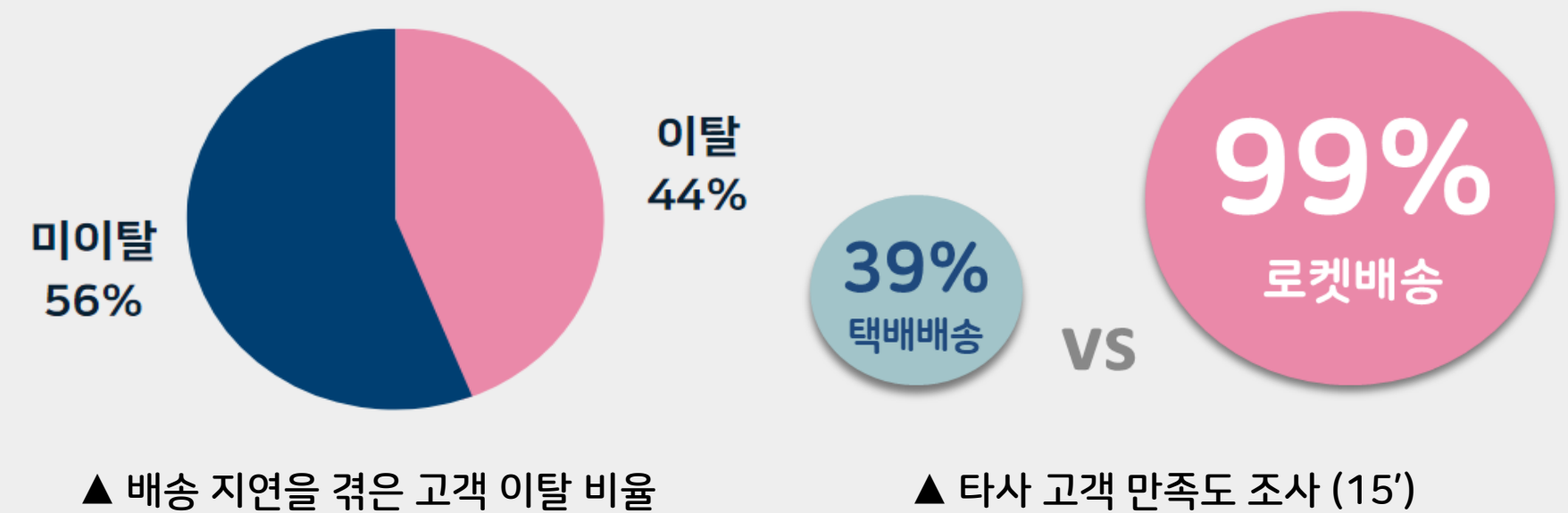
p / d / q	BIC	AIC
(3, 2, 2)	6480	6455



월 별 고객 이탈이  
지속적으로 발생함,

고객 이탈의 원인은  
배송지연임을 확인

배송 지연을 겪은 고객 44% 이탈, 고객 이탈과 배송기간은 연관이 있음

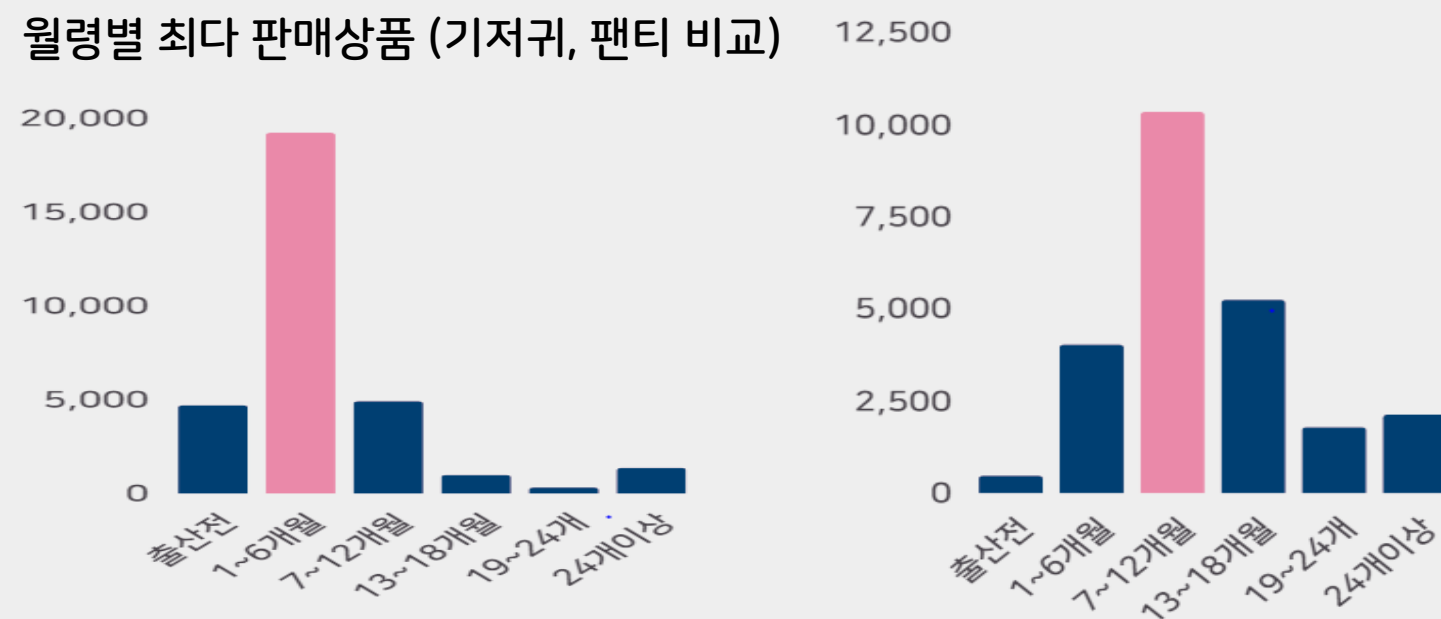


- 배송 지연은 고객 이탈률과 재구매 여부에 영향을 끼침
- '더블하트 Soft Touch 모유실감 젖꼭지'의 경우, 배송 지연률(48%)이 높음  
-> 해당 제품에 대한 ARIMA Model을 구성하여, 수요 예측 실시  
이를 이용해 사전 재고 확보를 통한 배송 지연 대비
- 배송 지연에 따른 고객의 이탈을 방지한다면,  
재구매율을 높이고 매출을 올릴 수 있을 것으로 기대

배송지연에 따른 고객 이탈을 방지하고 재구매율을 높이기 위해,  
배송 지연 상품의 재고를 대비하고 이탈 가능성 고객을 예측하여 맞춤 서비스 필요

## 5-1. 개선안 및 적용방안 - 팬티형 체험팩 추가

### ● 월령별 최다 판매상품 (기저귀, 팬티 비교)



▶ 1~6개월은 기저귀를 주로 구매, 7~12개월은 팬티를 주로 구매

### ● 다양한 팬티 체험용 체험팩 구성

### 팬티 체험팩 구성 예시

- 
 하기스 네이처메이드 오가닉 팬티 3단계 2P
- 
 하기스 네이처메이드 팬티 3단계 2P
- 
 하기스 네이처메이드 드라이 팬티 3단계 2P

▶ 현재 1~6개월 자녀를 위한 기저귀 위주에 체험팩 구성에서  
7~12개월 자녀를 위한 팬티 체험팩을 추가로 구성한다.

기저귀에서 팬티로 넘어가는 시기를 활용하여  
7~12개월 자녀를 위한 팬티용 체험팩을 통해 매출상승을 기대할 수 있음



< 02-0000-0000

### [Web발신]

000 고객님의, 안녕하세요.

A1의 공식 온라인 물인 A1234에서 복귀회원  
고객님께 다양한 혜택을 드립니다.

어디서도 만나볼 수 없는 온라인 전용 사은품 목  
록을 지금 바로 확인해 보세요

### ■ 복귀회원 혜택

- 1) 복귀회원 위한 15%할인 쿠폰 발급
- 2) 3만원이상 구매 시 유아용품 상품권 또는  
체험팩 사은품 제공
- 3) 육아 노하우 best5 도서 증정

감사합니다.

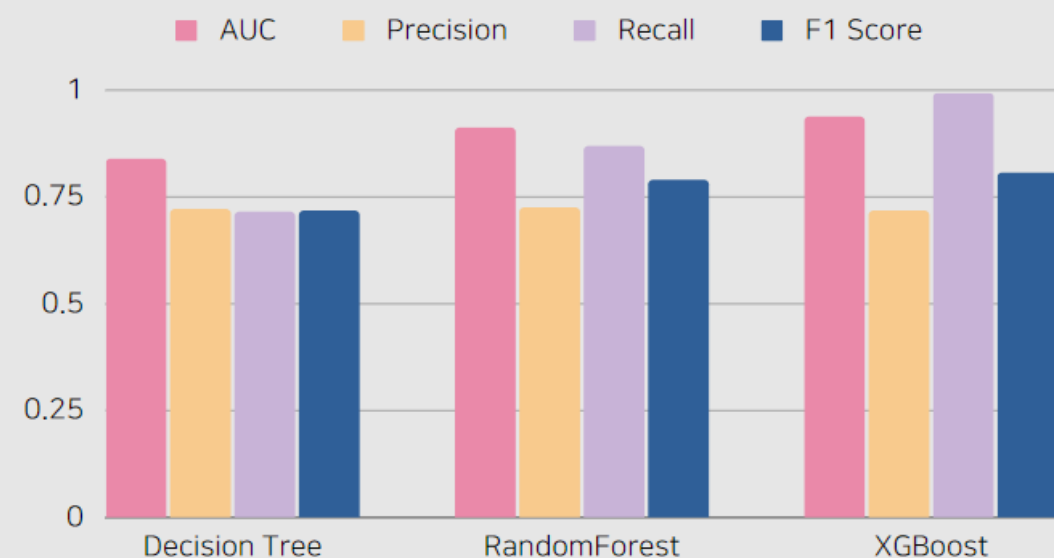
자사 인스타 계정 운영 및 문자 알리를 통한 홍보 전략

## 5-2. 개선안 및 적용방안 - 재구매율 증가를 위한 상품추천 시스템



### 고객 정보 기반 재구매 여부 예측 모델

예측 모델 평가 지표



f1 score 기준 XGBOOST 모델 선정 (f1 score= 0.81)



### 자녀 월령/ 구매자 연령층별 상품추천 모델

연령대	자녀연령(개월)	최다판매상품	추천상품	Lift
20대	1~6	기저귀	FC	4.44
	7~12	팬티	SKC	2.77
	13~18	팬티	물티슈	2.43
30대	1~6	기저귀	체험팩, 더블하트	2.56
	7~12	팬티	FC	2.57
	13~18	팬티	물티슈	2.43



### 고객의 재구매를 유도한 매출 증대

재구매 여부 예측 모델

+

상품 추천 모델



목표: 고객의 재구매 여부를 예측해 추천 상품 소비를 촉진



X = 고객 정보 (최근 주문일, 고객 등급, 구매금액 등)  
Y = 고객 재구매 여부 (재구매 기준: 90일 이내 구매 2회 이상)  
train data : test data = 7 : 3



재구매 의사가 없는 것으로 예측되는 고객의  
연령대, 자녀 월령별 최다 판매 물품대분류를 구해  
연관규칙 분석 진행 후 최대 향상도(lift)를 가지는 상품 추천

재구매 여부를 예측해 고객의 특성에 맞는 상품을 추천해주는 서비스를 제공하면  
고객의 재구매율 증가 및 기존 회원 유지 효과를 기대할수있음

## 5-3. 개선안 및 적용 방안 - 재고 사전 대비 및 고객 이탈가능성 예측

### 배송 지연이 잦은 상품 재고 대비



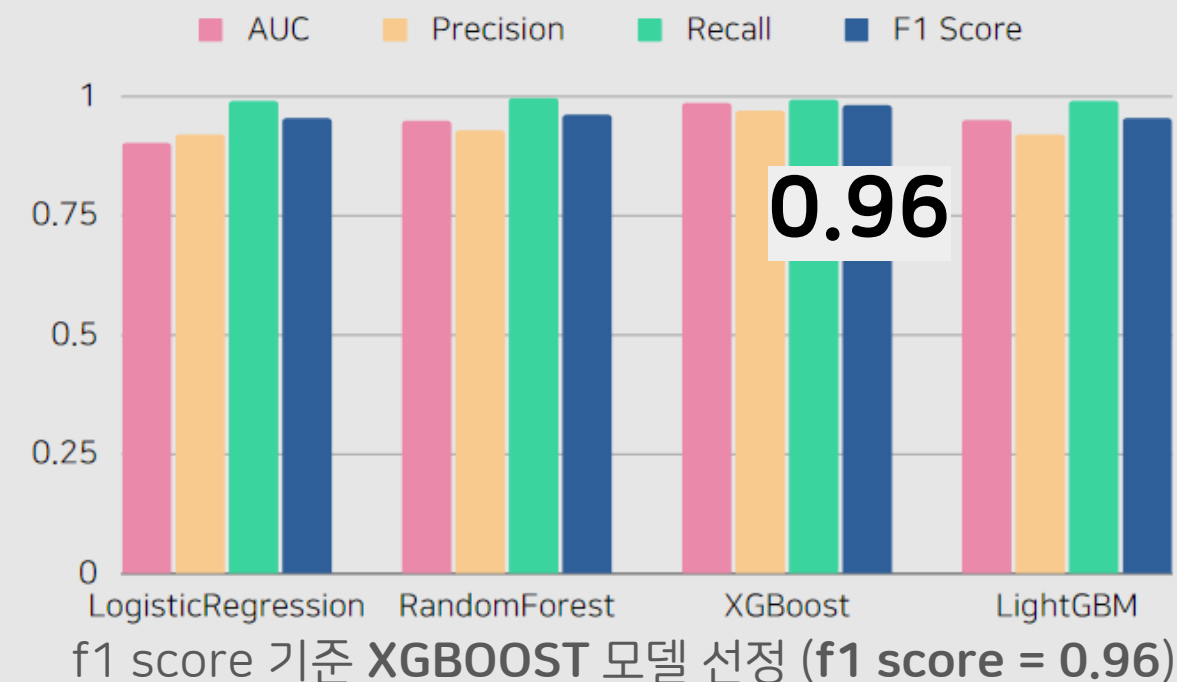
배송 지연률이 높은  
상품 탐색

ARIMA를 이용해  
상품의 수요 예측

상품 재고 집중 관리

- ✓ 목표:  
배송 지연이 잦은 상품을 파악하여 사전에 재고 대비
- ✓ 상품의 수요량을 미리 예측
- ✓ 해당 상품의 재고를 집중적으로 관리

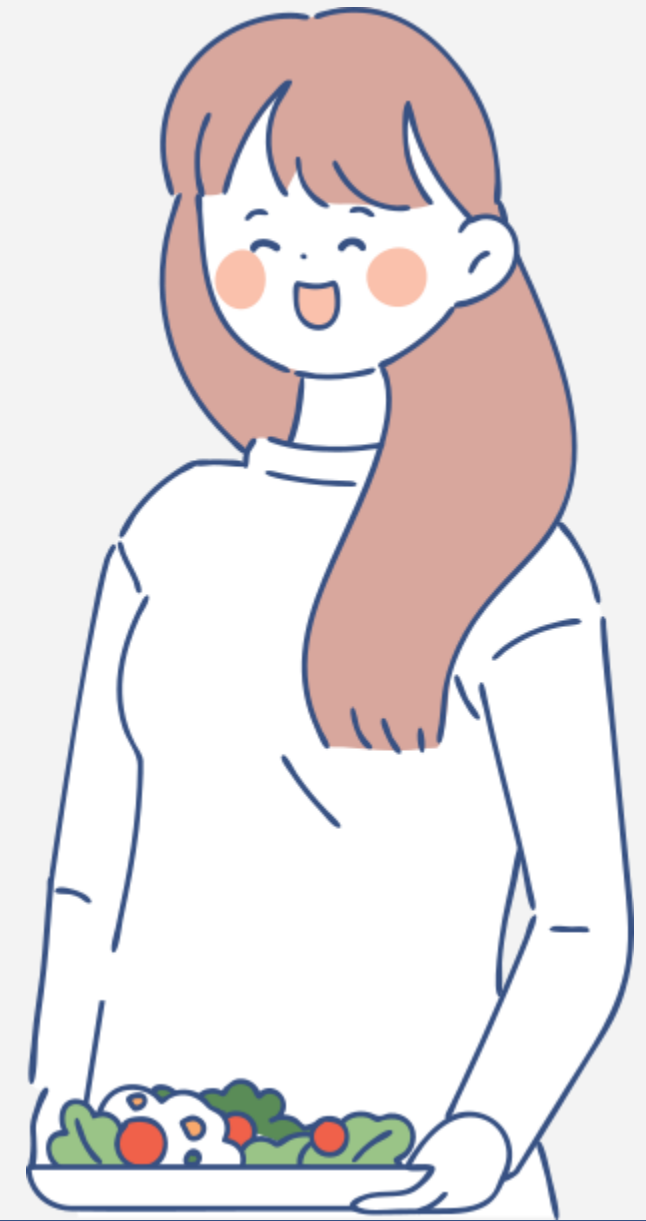
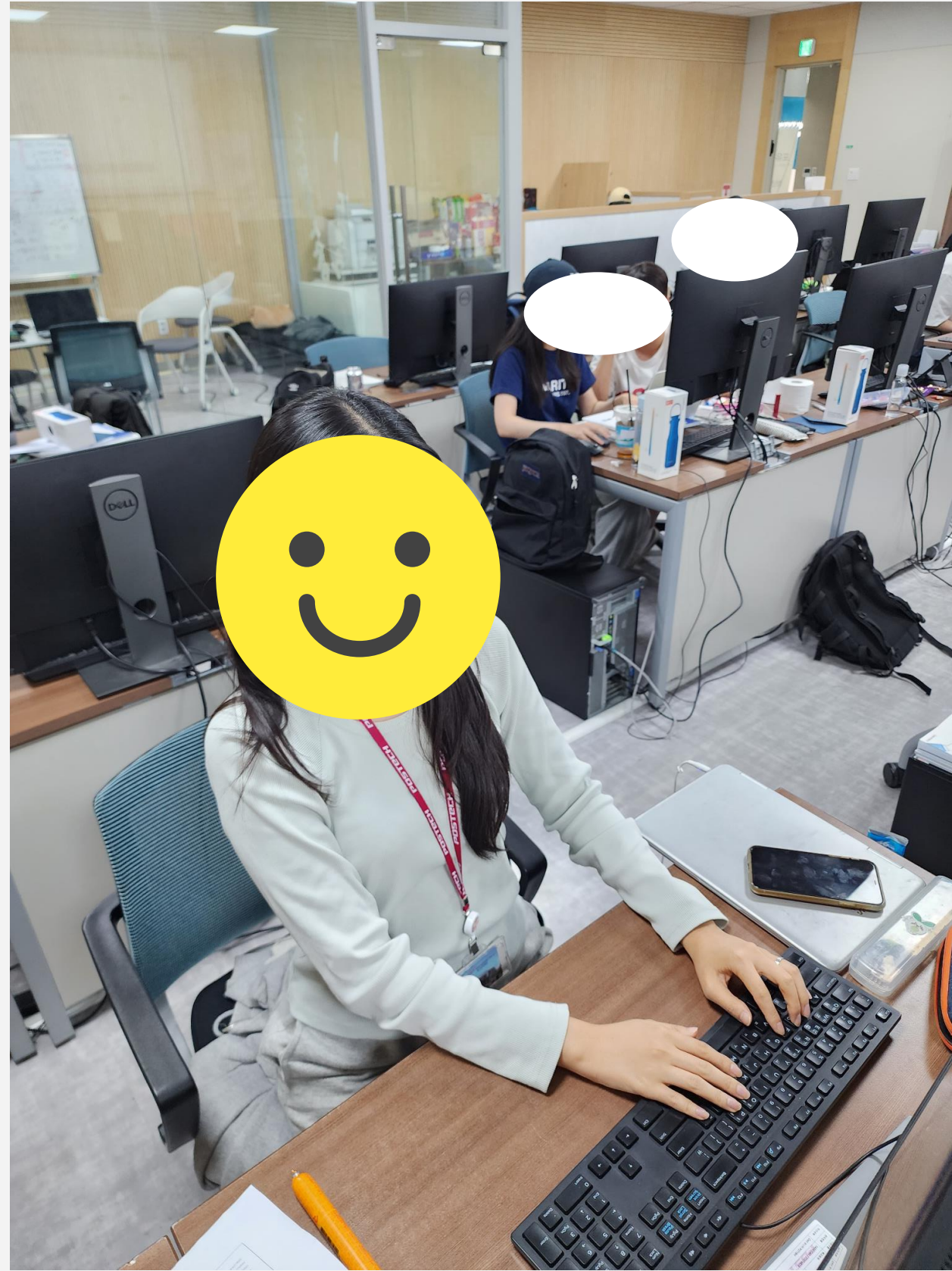
### 고객 이탈 가능성 예측 모델



- ✓ 목표: 고객의 이탈 가능성을 예측하여 이탈 방지
- ✓ X = 고객 구매 정보 (최근 주문일, 고객 등급, 구매금액 등)  
Y = 고객 이탈 여부 (이탈 기준 180일, True/False)  
Train data : test data = 7: 3
- ✓ 이탈 가능성에 따른 서비스 제공  
이탈 가능성이 있는 고객에게 3일 이내 사용 쿠폰 서비스를 제공

배송 지연이 잦은 상품을 사전에 재고 대비하고, 이탈 가능성 고객을 예측하여 맞춤 서비스를 제공한다면,  
고객 이탈을 방지하고 재구매율 증가 효과를 누릴 수 있음





오랜 작업으로 인해 시력이 안좋아진 정윤이...



## 7. Learned Lesson

14

### 류지환

주어진 데이터를 목적에 맞게 다양한 방식으로 조합하여 문제를 해결하는 과정에서 빅데이터 분석의 전반적인 과정을 다시 한번 적립할 수 있었습니다. 실제 기업의 데이터를 다뤄보며 얻은 결과뿐만 아니라 그 결과를 실제 업무에 적용하는 방안을 도출하는 과정이 좋았고 팀원들과 몰입을 통해 공동의 목표 달성한 경험이 뜻깊었습니다.

### 박미령

이번 프로젝트를 통해 여태 배웠던 데이터분석을 다시 복습할 수 있었고, 기업의 복잡한 상황에 문제해결 기법을 적용하는 방법을 배웠습니다. 또한 팀원내의 다양한 갈등이 불화로 번지지 않게 관리하며 협업을 강화하는 뜻깊은 시간이었습니다.

### 박선우

이번 프로젝트를 통해 협업을 통해 문제를 해결하는 방법을 배웠습니다. 혼자하는 일보다 더 논의하고 고려할 점이 많았지만 그만큼 더 좋은 결과가 나왔습니다. 또한 내가 할 수 없는 부분을 서로 보완해줄 수 있고 무엇보다 끈끈한 유대가 우리가 끝까지 달릴 수 있게 해주는 힘을 알게되었습니다.

### 신정윤

실제 기업의 문제 상황을 해결하기 위해 현상 파악과 원인 탐색부터 데이터 분석을 통한 개선안 도출까지 전체적인 과정을 배울 수 있어 좋았습니다. 또 단순히 데이터 분석 결과만 확인하는 데 그치지 않고 인사이트를 해석하고 활용할 방안까지 생각하는 주관을 키울 수 있는 좋은 기회였던 것 같습니다.

### 유태욱

실제 데이터를 갖고 빅데이터 분석 프로젝트 과정을 차례대로 경험하며, 데이터 분석보다 정제에 더 많은 시간이 든다는 점을 느끼게 되었습니다. 또한, 팀원들과 하나의 목표를 위해 열중한 경험이 되어 좋았습니다.

### 한수빈

하나의 raw data라도 각각의 필드를 사용하는 목적에 따라 다양한 데이터 프레임 만들어봄으로써 문제 해결을 여러 방향의 시각으로 고민해 보는 기회가 되었습니다. 또한, 데이터를 원하는 방향으로 나타내기 위해 알고리즘을 고민하며 computational thinking의 역량이 많이 늘어난 의미 있는 경험이었습니다.

## 고객 이탈 가능성 예측

최근 구매일 (일 전)

결제 금액

고객 등급

배송 기간 (일)

배송 지연 여부

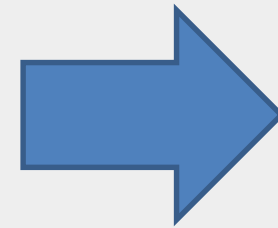
물품 대분류

구매 일자

년 - 월 - 일



이탈 가능성 예측



## 고객 이탈 가능성 예측

최근 구매일 (일 전)

45

결제 금액

1040000

고객 등급

나무



배송 기간 (일)

2

배송 지연 여부

있음



물품 대분류

기저귀

구매 일자

2023-08-18



이탈 가능성 예측

이탈 가능성 예측 결과

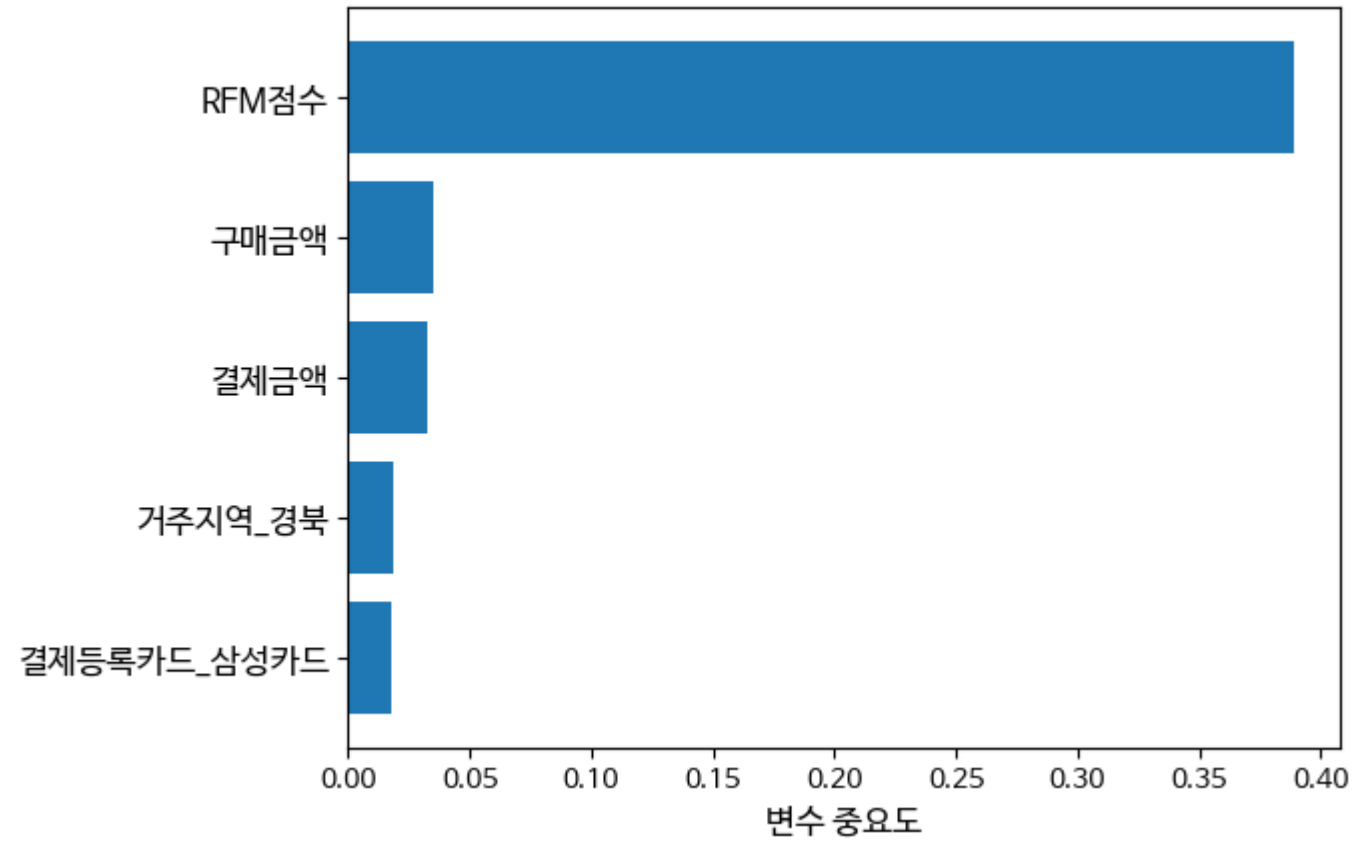
이탈 가능성: 2.76%



감사합니다  
THANK YOU!

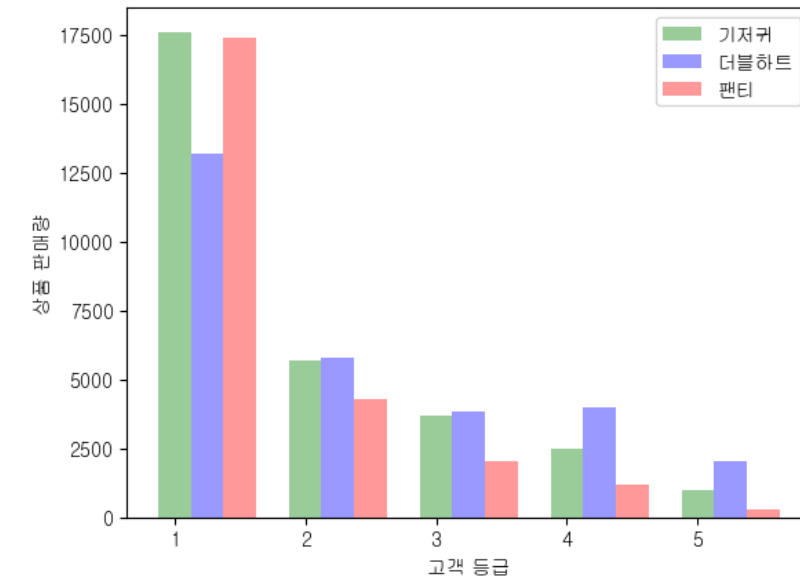
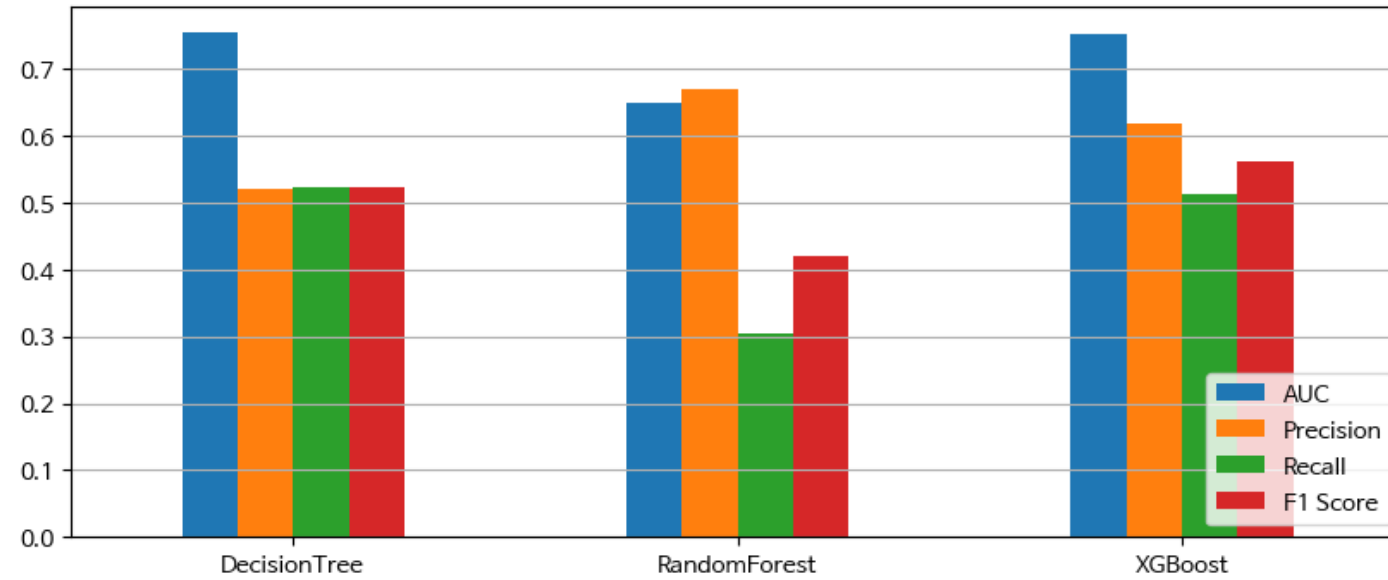
	AUC	Precision	Recall	F1 Score
LR	0.901	0.919	0.989	0.953
RF	0.947	0.927	0.996	0.960
XGB	0.985	0.968	0.992	0.980
LGBM	0.949	0.919	0.989	0.953

상위 5개 변수 중요도



```
: true_counts.index[:20].div()
```

```
Index(['맘큐 허그박스', '하گی스 네이처메이드 2단계 여아 156매(밴드형)',
      '더블하트 SofTouch 모유실감 젖꼭지 L (2입)', '하گی스 네이처메이드 2단계 남아 156매(밴드형)',
      '더블하트 SofTouch 모유실감 젖꼭지 M (2입)', '하گی스 네이처메이드 3단계 여아 132매(밴드형)',
      '하گی스 네이처메이드 3단계 남아 132매(밴드형)', '하گی스 맥스드라이 3단계 공용 20매(밴드형/체험팩)',
      '더블하트 SofTouch 모유실감 젖꼭지 S (2입)', '더블하트 SofTouch 모유실감 젖꼭지 LL (2입)',
      '하گی스 네이처메이드 1단계 공용 186매(밴드형)', '2020 하گی스 매직 컴포트 3단계 공용 20매(밴드형/체험팩)',
      '하گی스 맥스드라이 팬티 4단계 남아 20매(팬티형/체험팩)',
      '더블하트 SofTouch 모유실감 노꼭지 트윈팩 240ml(오렌지)', '하گی스 맥스드라이 3단계 공용 132매(밴드형)',
      '크리넥스 여행용티슈 카카오프렌즈 70매 X 3개입', '하گی스 맥스드라이 팬티 4단계 여아 20매(팬티형/체험팩)',
      '더블하트 젖꼭지 브러쉬 와이드형+젓병 스펀지 브러쉬 일반형',
      '[2019년형] 하گی스 에어솔솔 씹머기저귀 3단계 공용 144매(밴드형)',
      '2020 하گی스 매직팬티 컴포트 4단계 남아 20매(팬티형/체험팩)'],
      dtype='object')
```



항목명	의미	유형	이상치(건)	결측치(건)	확인결과	정제방안
고객번호	고객 구분 번호 ID	범주형	0	0		
성별	주문 고객의 성별	범주형	0	266	비회원 주문 고객	비회원 유무에 따른 데이터프레임 생성
결혼유무	주문 고객의 결혼유무	범주형	0	266		
거주지역	주문 고객의 거주지역	범주형	0	266		
연령	주문 고객의 나이	연속형	0	266		
결제등록카드	상품 주문 시 결제를 위한 카드	범주형	0	266		
유입경로	회원 가입 시 유입경로	범주형	<sup>19</sup> 0	266		
자녀 여부	회원의 자녀여부	범주형	0	266		
고객 등급	회원 RFM 점수에 따른 고객 등급	연속형	0	0		파생변수 생성
RFM 점수	회원 RFM 점수	연속형	0	0		파생변수 생성
고객 이탈 여부	180일 이상 미주문 시 이탈	범주형	0	0		파생변수 생성

Sales\_data.csv (199999 rows, 12 cols)

항목명	의미	유형	이상치(건)	결측치(건)	확인결과	정제방안
전체 매출액	당사의 1년 매출액	연속형				파생변수 추가
월/주/일간 매출	월/주/일 간 당사 매출액	연속형				테이블 추가
배송 지연 여부	배송 기간이 일정 기간 이상인 주문 건 수	범주형				파생변수 추가
배송기간	배송 완료일 - 배송 시작일	연속형				파생변수 추가
구매 시 월령	고객 자녀의 월령(00개월)	연속형	16655	31857	36개월 이상은 이상치로 판단	분석 시 데이터 삭제 -> NaN값으로 대체
결제금액	고객이 결제한 금액	연속형	1483	0	0원을 이상치로 판단	데이터 삭제

항목명	의미	유형	이상치(건)	결측치(건)	확인결과	정제방안
물품판매량	상품이 판매된 개수	연속형	1	0	1만 2천개 주문 데이터를 이상치로 판단	분석 시 데이터 삭제
물품대분류	상품이 어떤 분류인지	범주형	0	0		
제품단가	상품 별 가격	연속형	90	0	0원을 이상치로 판단	분석 시 데이터 삭제
상품명	상품의 이름	범주형	0	0		상품명, 주문 개수 분리
재구매율	상품대분류 별 재구매율	연속형				파생변수 추가