

K-Digital Training 371

# 유통기업 데이터 분석

부제 : 구매 감소고객 예측 및 활용방안

2021.10.18 | 2조 | 송유빈, 정하림, 박시우, 조경림, 최혜정

# CONTENTS

---

**01** 추진배경

**03** 타겟 고객 INSIGHT

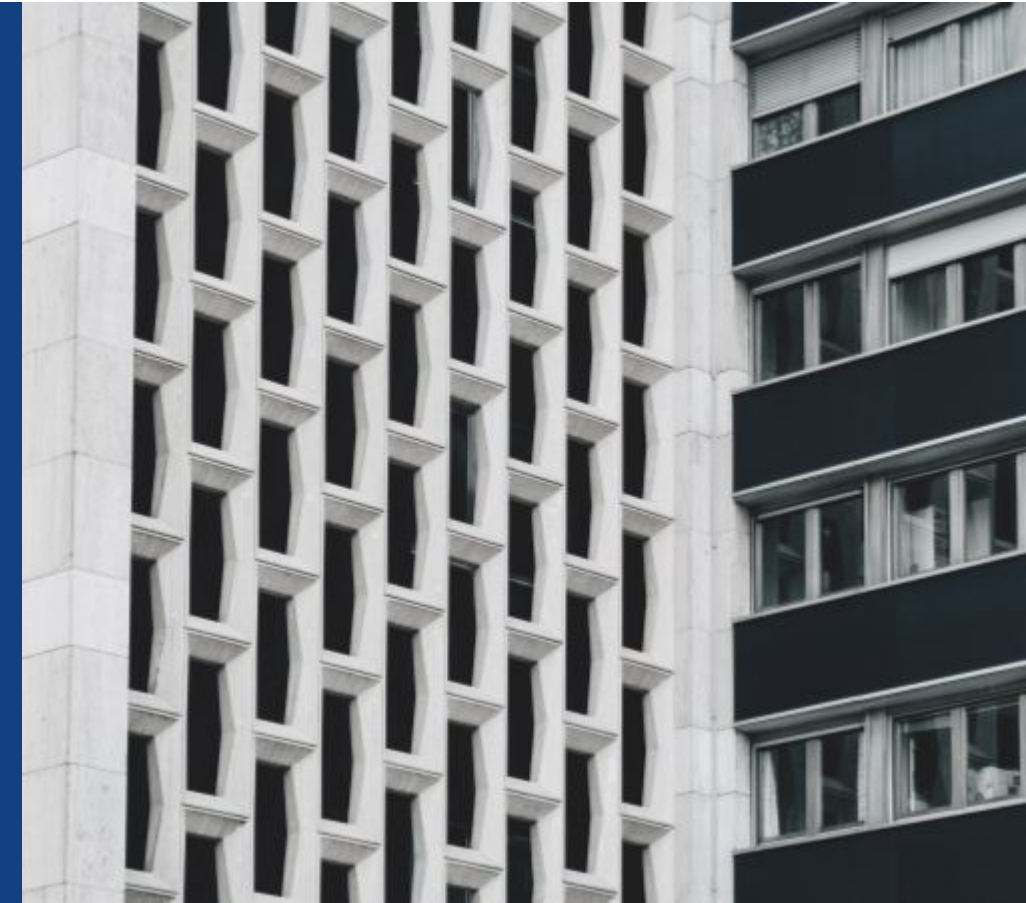
**02** 분석 과제화

**04** 제안 및 정리





# 이 추진배경



# 01 추진배경 유통업계 현안

전년 동분기 대비 매출 증감률 추이 (단위: %)



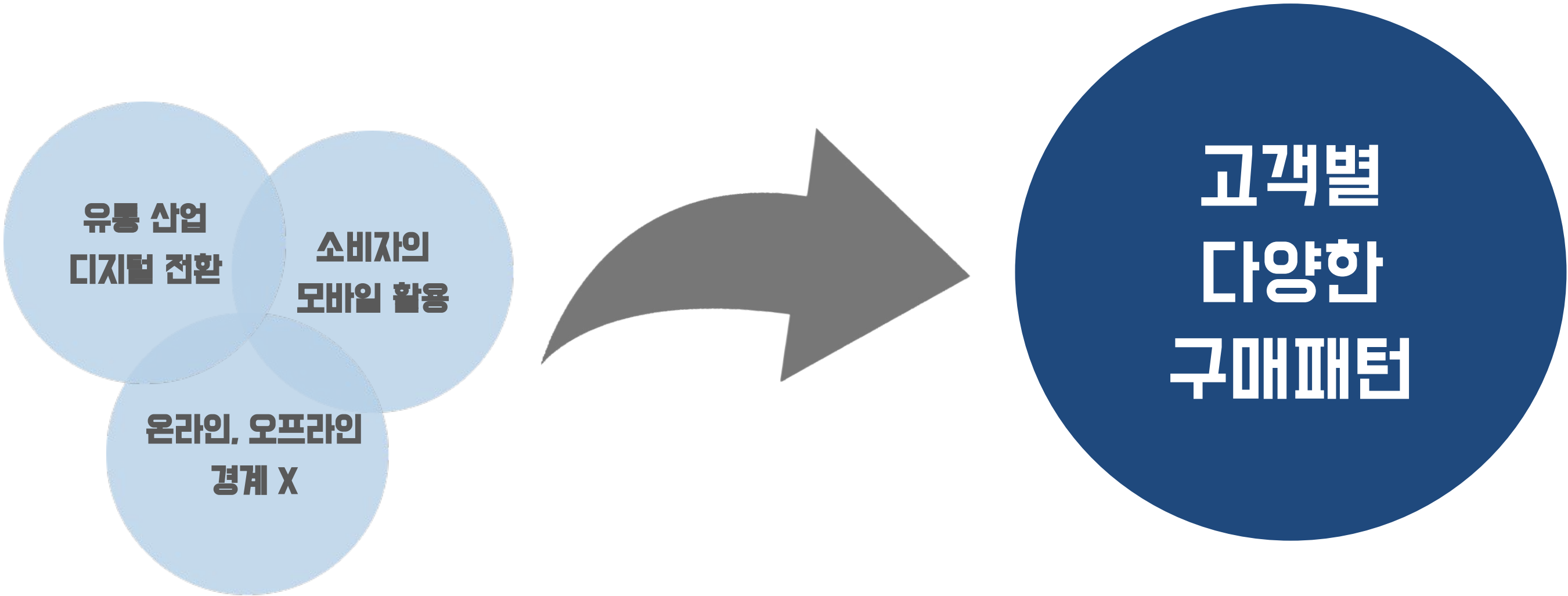
구매의 **편의성, 접근성**이 강점인

온라인 유통, 편의점, 홈쇼핑 등에서 보이는 높은 고객 구매 행동 패턴

01

추진배경

유통업계 현안



01

추진배경

보유 데이터

( 데이터수, 항목수 )



고객 Demo

( 19383, 4 )

• 고객번호

• 연령대

• 성별

• 거주지역

멤버십 가입

( 7456, 3 )

• 고객번호

• 멤버십명

• 가입년월

구매데이터

( 28593030, 10 )

• 제휴사

• 영수증번호

• 대분류코드

• 중분류코드

• 소분류코드

• 고객번호

• 점포코드

• 구매일자

• 구매시간

• 구매금액

상품분류

(, 4926, 6 )

• 제휴사

• 대분류코드

• 중분류코드

• 소분류코드

• 중분류명

• 소분류명

채널 이용

( 8824, 3 )

• 고객번호

• 제휴사

• 이용횟수

# 01 추진배경 데이터 탐색 기준

## 데이터 분석 기준

- 2014년 상반기 : 1분기
- 2014년 하반기 : 2분기
- 2015년 상반기 : 3분기
- 2015년 하반기 : 4분기
  
- train data = 1~3분기
- test data = 4분기



## 구매감소고객 기준

전체 매출의 증감에서  
고객별로 차지하는 구매 총금액의  
증감의 비교가 1보다 작은 고객

고객별 구매 총금액의 증감률

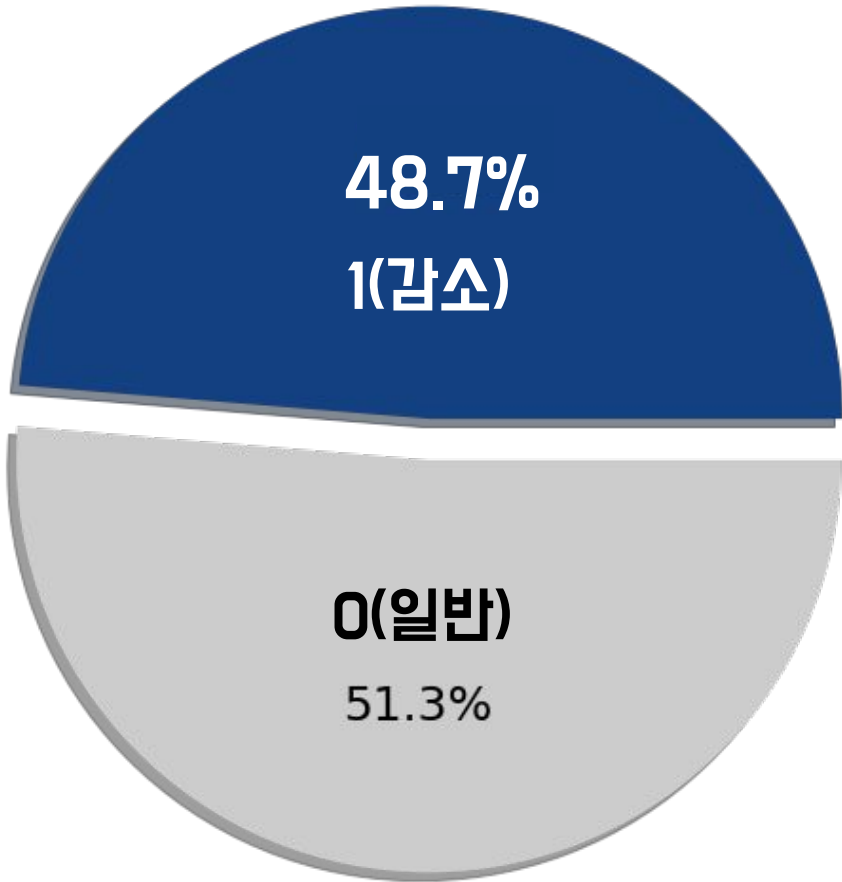
전체 매출의 증감률

< 1

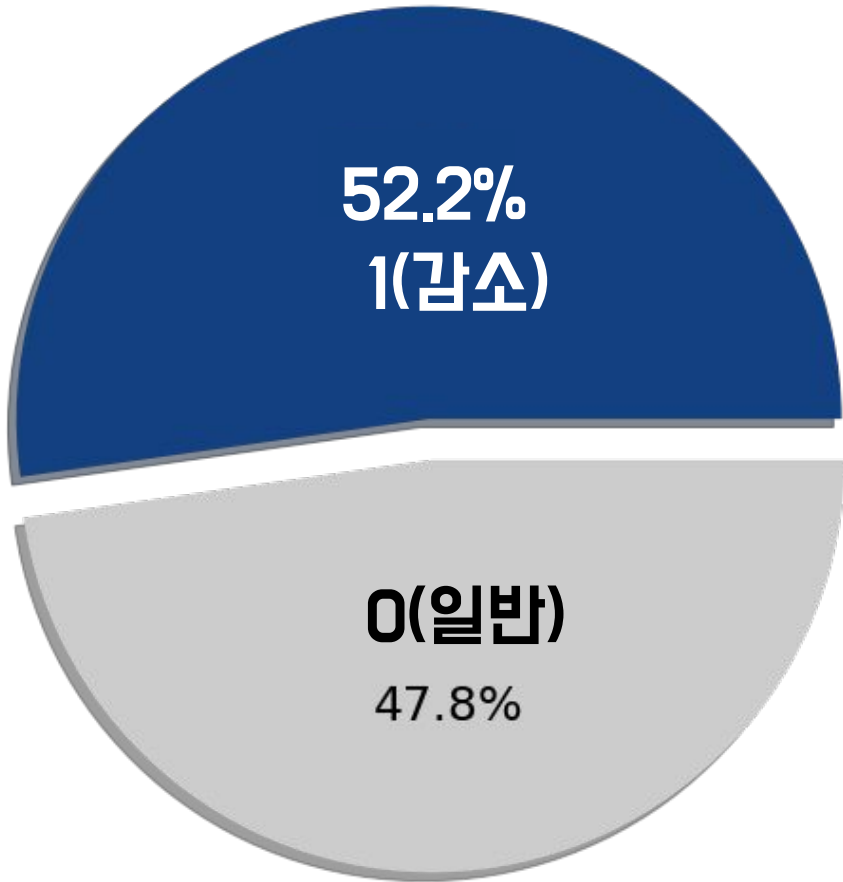
1분기, 3분기 데이터의 증감과 전체매출의 증감을 비교하여  
구매감소고객 패턴 분석

# 01 추진배경

1.3반기 구매건수 증감 고객 비교



1.3반기 구매금액 증감 고객 비교



구매금액/건수의  
구매감소고객 분석 결과,  
각각 **48.7%, 52.2%**



# 01 추진배경

통합데이터 활용을 위해 분류체계 통일 후 다시 탐색

- ❑ **새분류** : 카테고리 분류. \*명품 또한 중분류명에 카테고리로 분류되어 있어서 이용함.  
[가전, 가구/인테리어, 의류/잡화, 보석, 명품, 레저/스포츠, 식품, 일상용품, 뷰티, 기타, 유아/아동용품, 교육/문화/사무용품]

- ❑ **삼분류** : 편의품/선매품/전문품

- 편의품 : 식품, 일상용품, 교육/문화/ 사무용품, 기타
- 선매품 : 가전, 뷰티, 스포츠/레저, 의류/잡화, 유아/아동용품
- 전문품 : 명품, 보석, 가구/인테리어, 악기류 (소분류명)

- 삼분류 기준 : 소비재의 종류와 마케팅

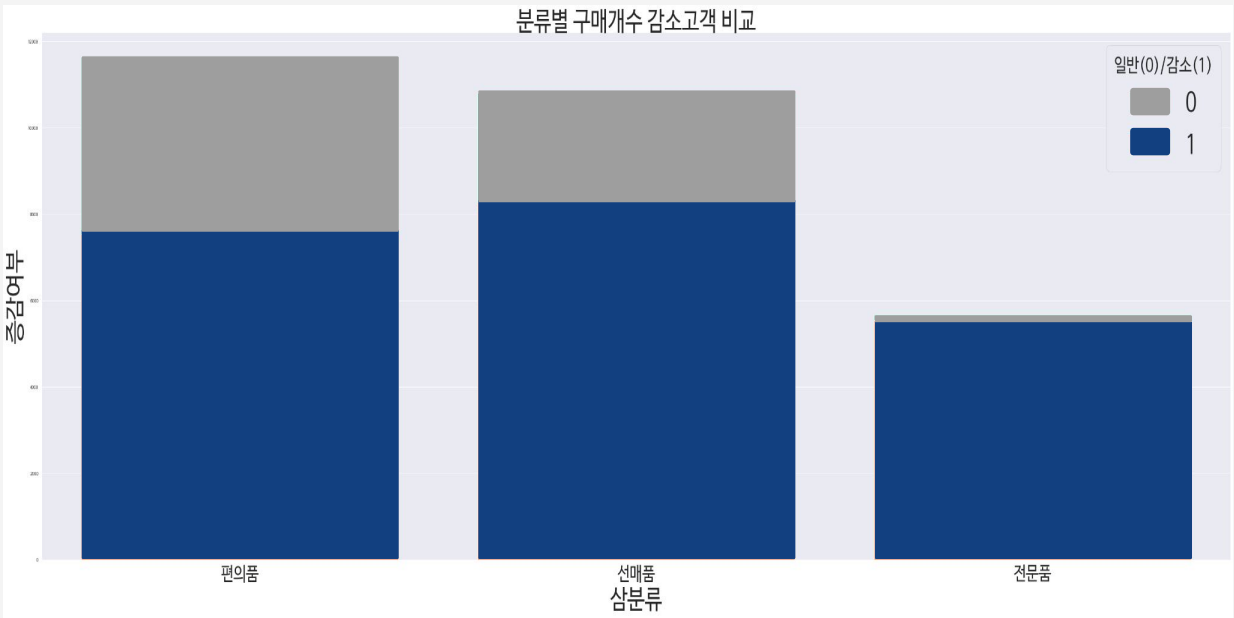
	편의품	선매품	전문품
구매빈도	높다	중간	낮다
관여도 수준	낮다	비교적 높다	매우 높다
제품 예시	치약, 세제, 비누	패션, 승용차, 가구	고급 제품, 보석류
가격	저가	고가	매우 높은 가격

새로운  
분류체계

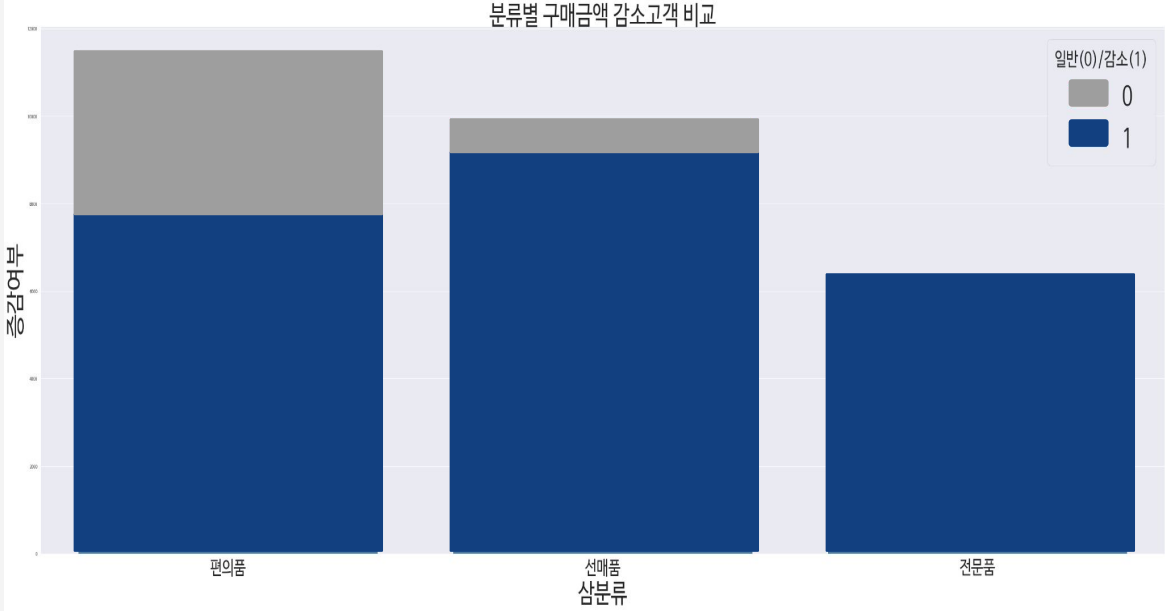
# 01 추진배경

## 새로운 분류로 고객 탐색

삼분류별 상품구매건수 고객 누적 그래프



삼분류별 상품구매금액 고객 누적 그래프



( x축 : 품목명, y축 : 고객수 )

**편의품/선매품/전문품의 감소고객 탐색**

: 주로 다나가를 고려함 (저/중/고)

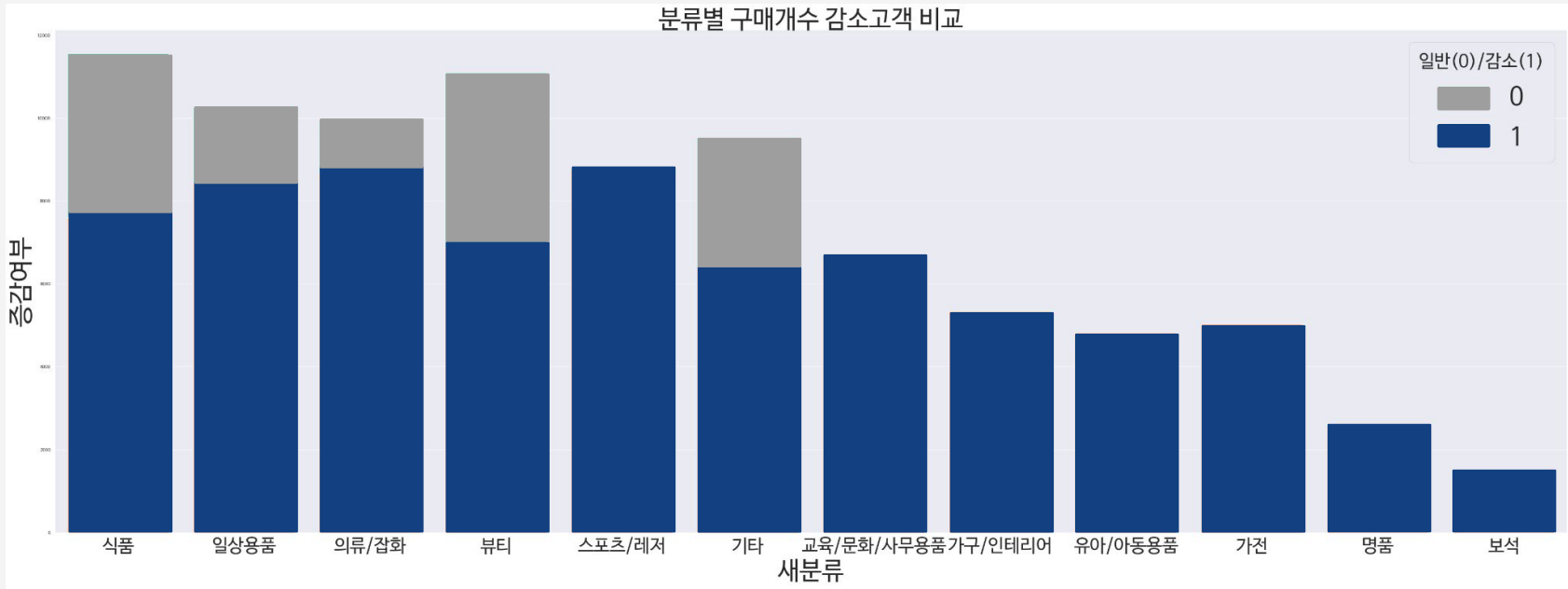
**구매건수, 구매금액 감소고객이 더 많거나 비슷한 품목**

→ **전문품**

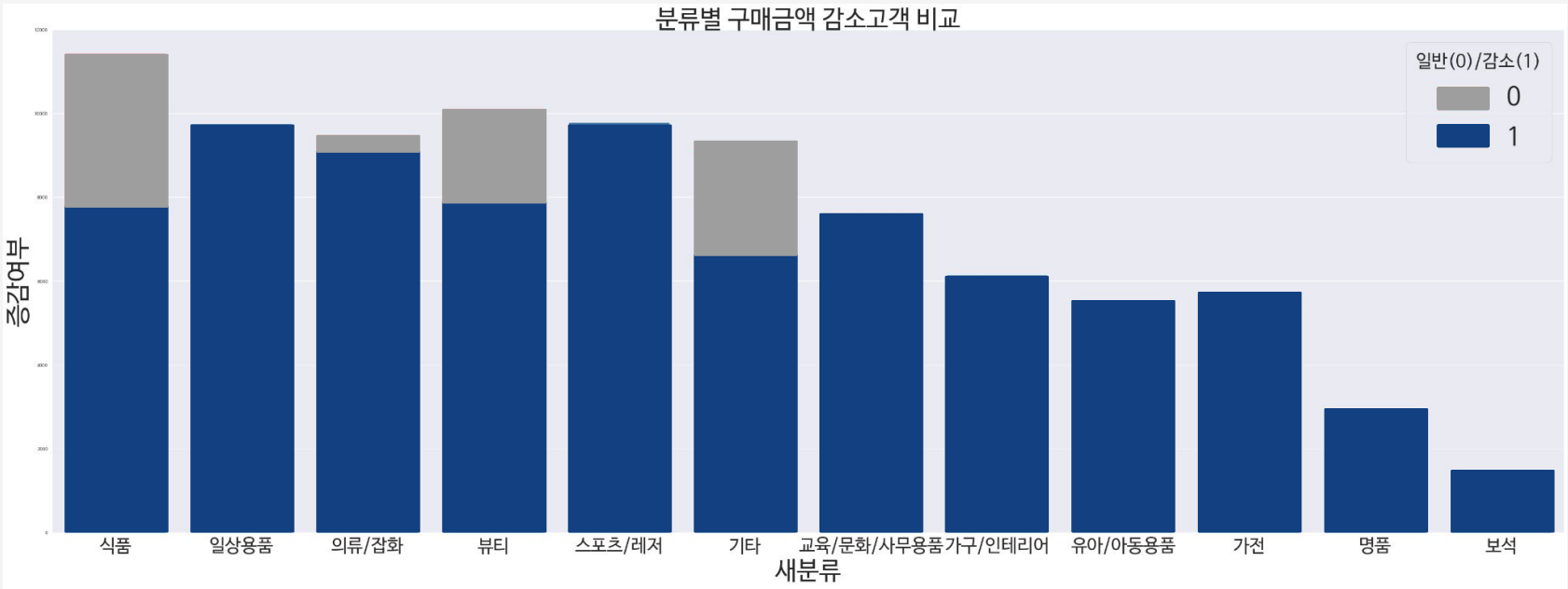
# 01 추진배경

## 새로운 분류로 고객 탐색

분류별 상품구매건수 감소고객 누적 그래프



분류별 상품구매금액 감소고객 누적 그래프



( x축 : 품목명, y축 : 고객수 )

(구매건수감소 품목) : 스포츠/레저, 교육/문화/사무용품, 가구/인테리어, 유아/아동용품, 가전, 명품, 보석

→ 12개 항목 중 7개 항목에서 구매건수 감소

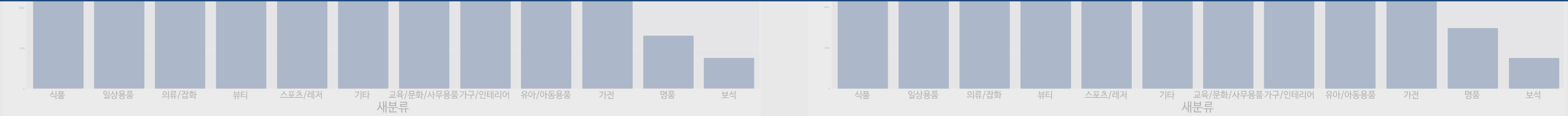
(구매금액감소 품목) : 일상용품, 스포츠/레저, 교육/문화/사무용품, 가구/인테리어, 유아/아동용품, 가전, 명품, 보석

→ 12개 항목 중 8개 항목에서 구매비용 감소

# 01 추진배경

새로운 분류로 고객 탐색

구매 감소 고객을 예측해보고자 함



( x축 : 품목명, y축 : 고객수 )

(구매건수감소 품목) : 스포츠/레저, 교육/문화/사무용품, 가구/인테리어, 유아/아동용품, 가전, 명품, 보석

→ 12개 항목 중 7개 항목에서 구매건수 감소

(구매금액감소 품목) : 일상용품, 스포츠/레저, 교육/문화/사무용품, 가구/인테리어, 유아/아동용품, 가전, 명품, 보석

→ 12개 항목 중 8개 항목에서 구매비용 감소





## 02 분석 과제화

## 02 분석 과제화 변수 및 데이터셋 설정

### - 종속변수 설정 -

증감률 =

(고객별 3분기 총구매금액 - 고객별 1분기 총구매금액)

고객별 1분기총구매금액

( 3분기 총매출액 - 1분기 총매출액)

1분기 총매출액

( 1분기를 기준으로 )

Target

구매 감소고객 : 증감비 < 1 → 1

\* 감소고객이 아닌 경우 0

고객별 구매 총금액의 증감률

< 1

전체 매출의 증감률

Y(타겟) 구매감소고객  
(감소 : 1, 증가 : 0)

02

분석 과제화

변수 및 데이터셋 설정

- 변동변수 산출 -

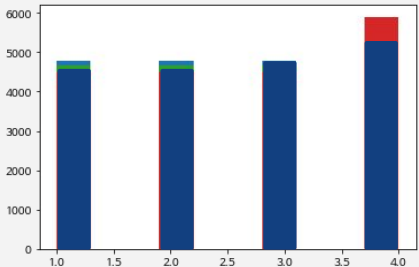
1) 고객별 집계

1,2,3,4분기별	새분류별	삼분류별	A,B,C,D 제휴사별
총구매금액 및 총구매 갯수 ABS, NET			

2) 고객 집계항목 등급 나누기

- 사분위수에 따라 4등급으로 부여

1등급 25%	2등급 50%	3등급 75%	4등급 100%
------------	------------	------------	-------------



3) 등급 간의 변동 계산하여 지수 산출

- 반기별 변동지수 -> abs (absolute value sum, 변동 폭)
- 총 변동(증감) 지수 -> Net 계산

## 02 분석 과제화 변수 및 데이터셋 설정

### - 독립변수 선정 -

#### 1 고객 특성 변수

- 성별  
(여성 : 1, 남성 : 0)
- 연령대
- 구매시간대
- 멤버십 여부

#### 2 변동폭(ABS) 변수

- 반기별 총 구매금액 변동폭
- 제휴사별 구매금액 변동폭
- 상품분류별 구매금액 변동폭
- 상품금액대분류별  
구매금액 변동폭

#### 3 변동 Net 변수

- 반기별 총 구매금액 증감
- 제휴사별 구매금액 증감
- 상품분류별 구매금액 증감
- 상품금액대분류별  
구매금액 증감

다중공선성, 변수중요도를 고려하여  
전체 96개의 변수중에 40개를 분석변수로 선정



02

분석 과제화

테이터 분석 모델링

구매감소 고객 예측모델 탐색

(혼동행렬 평가 지표(Accuracy, Precision, Recall, F1 score))

DecisionTree	LightGBM	RandomForest	XGBoost	catBoost
<div><div>❑ Accuracy : 0.6853</div><div>❑ Precision : 0.6976</div><div>❑ Recall : 0.7164</div><div>❑ F1 score : 0.7069</div></div> <div><div>- 구매증감지수 중요도 : 1순위 - 편의품 비중 2순위 - 선매품 비중</div><div>- 분기 sum_net이 중요도 1순위</div><div>- 상품 관점에서 뷰티, 스포츠/레저, 가구/인테리어, 가전 구매건수 Net으로 중요도 상위권</div></div>	<div><div>❑ Accuracy : 0.7052</div><div>❑ Precision: 0.7312</div><div>❑ Recall : 0.7013</div><div>❑ F1 score : 0.7159</div></div> <div><div>- 편의품 구매금액 Net이 중요도 1순위</div><div>- 전체 구매건수, 구매금액 Net 중요도 상위권에 존재.</div><div>- 뷰티, 가전, 가구/인테리어, 스포츠/레저 구매건수 Net 으로 중요도 상위권.</div><div>- 명품, 편의품, 선매품 구매금액 변동폭이 중요도 하위권에 존재함.</div></div>	<div><div>❑ Accuracy : 0.7137</div><div>❑ Precision: 0.7353</div><div>❑ Recall : 0.7180</div><div>❑ F1 score : 0.7265</div></div> <div><div>- 구매증감지수 중요도 : 편의품 &gt; 선매품 &gt; 전문품</div><div>- 분기 sum_net 의 중요도가 1순위이며, 제휴사별 구매금액 변수로는 A &gt; B &gt; C &gt; D 순서임.</div><div>- 구매횟수보다 구매비용에 관련된 변수의 중요도가 전체적으로 모두 높음.</div><div>- 상품 관점에서 식품, 인류/잡화 관련변수의 중요도가 높음.</div></div>	<div><div>❑ Accuracy : 0.6932</div><div>❑ Precision: 0.7118</div><div>❑ Recall : 0.7071</div><div>❑ F1 score : 0.7094</div></div> <div><div>- 구매증감지수 중요도 : 1순위 - 편의품 비중 2순위 - 선매품 비중 (둘의 차이는 크지 않고 다른 피처와는 큰 차이가 나타남)</div><div>- 상품관점에서 봤을때, 일상용품 sum_net 스포츠/레저 sum_net 가전 count_net 등이 중요도가 높음</div></div>	<div><div>❑ Accuracy : 0.7084</div><div>❑ Precision: 0.7289</div><div>❑ Recall: 0.7158</div><div>❑ F1 score : 0.7223</div></div> <div><div>- 편의품 비중이 중요도 1순위</div><div>- 분기sum_net이 중요도 2순위</div><div>- 전반적으로 net에대한 항목의 중요도가 높음</div><div>- B-&gt;A사 순으로 변수 중요도가 높음</div><div>- 12가지로 나뉜 새분류보다 삼분류에 대한 변수의 중요도가 높음</div></div>

## 02 분석 과제화 모델 선정

### Decision Tree

- 활용 변수 : 40개
- Accuracy : 0.6853

### XGBoost

- 활용 변수 : 40개
- Accuracy : 0.6932

### Random Forest

활용 변수 : 40개  
Accuracy : 0.7137

### LightGBM

- 활용 변수 : 40개
- Accuracy : 0.7057

### CatBoost

- 활용 변수 : 40개
- Accuracy : 0.7084

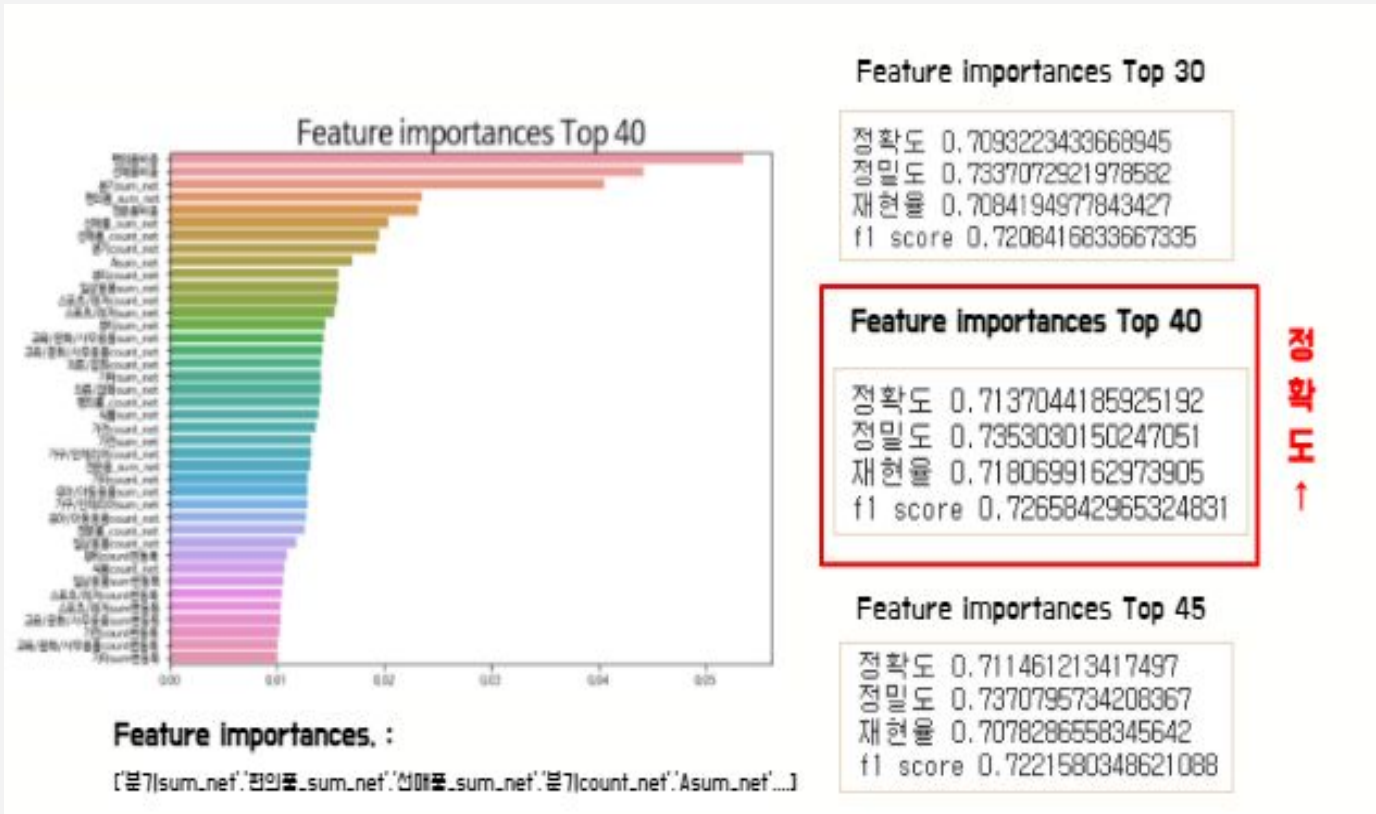
02

분석 과제화 모델 선정

- RandomForest -

x_train : 1-2반기 등급변동폭(절대값) , Net(+,-)	y_train : 1-3기 타겟(감소고객)
x_test : 1-3반기 등급변동폭(절대값) , Net(+,-)	y_test : 1-4기 타겟(감소고객)

- 총 고객 정보(n=19169)를 이용한 모델 정확도 : **0.7137**  
(Feature Importance 상위 40개 변수 채택)



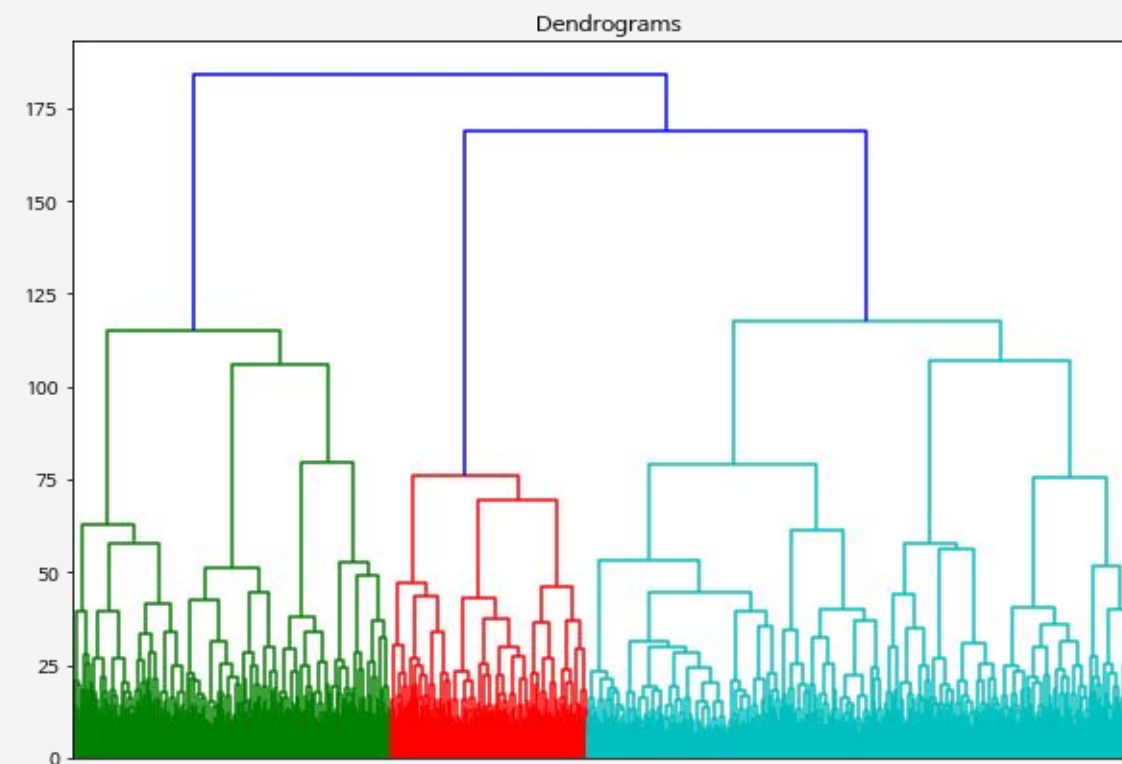
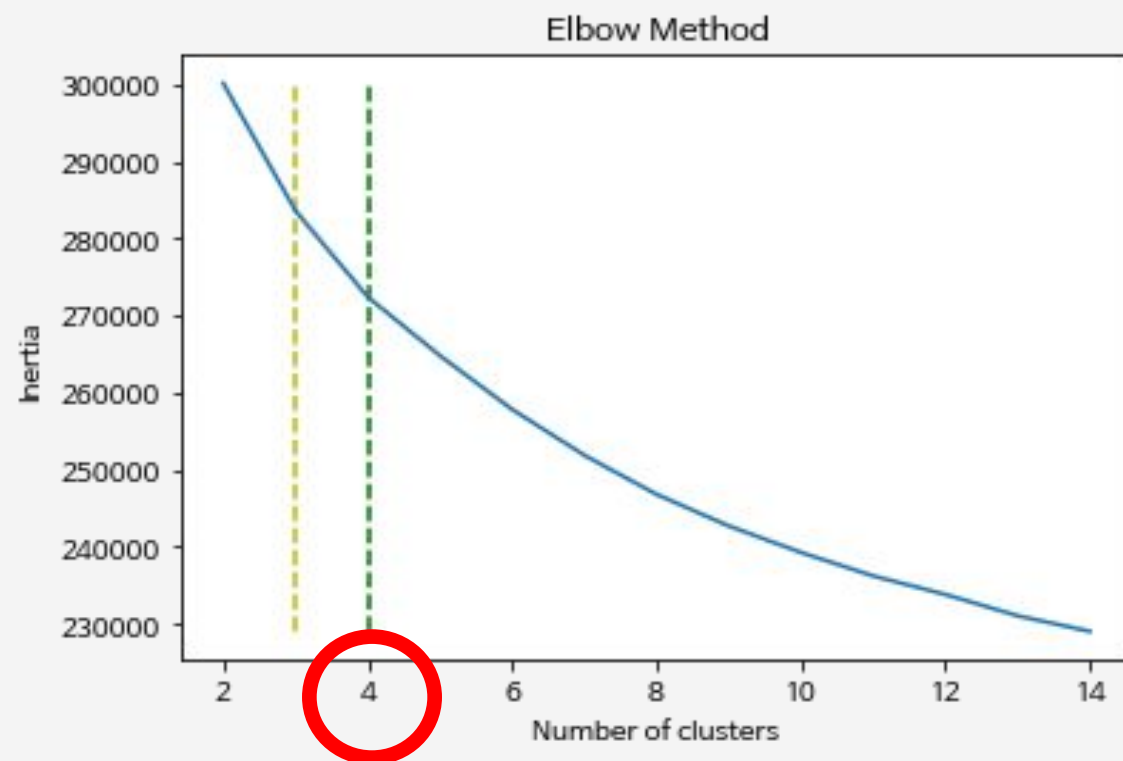
## 모델 기반 변수 선택

: 트리 기반 모델(또는 선형모델)들이  
특성 중요도(Feature Importance)를  
제공하는 것에 기반한 방법

## 02 분석 과제화 군집 및 유형화

### - 감소고객 분석 -

1. 군집분석에 사용할 중요 변수를 선정
2. 엘보우 방법과 Dendrogram 통해 최적의 **군집수 4개** 도출
3. 군집유형 파악





02

분석 과제화

군집 및 유형화

기준 :

등급증가

← -3 -2 -1 0 1 2 3 →

등급감소

구분	고객 특성				비중
구매 감소 고객	예측되는 구매감소고객 (9740/19169)  50.22%	충분기 구매건수 증감 > 0.5	가전 총구매액 증감 > 0.5	가전 총구매액 변동폭 > 1  A사 총구매액 증감 > 0.5	군집1 (43.8%)
			충분기 총구매금액 증감 > 0.5	A사 총구매액 증감 > 0.5	군집2 (18.4%)
				인류잡화 총구매액 증감 > 0.5  A사 총구매액 증감 > 0.5	군집3 (19.3%)
				선매품 총구매액 증감 > 0.5	군집4 (18.5%)



# 03 타겟 고객 INSIGHT

# 03 타겟 고객 INSIGHT : 군집1

4269명(감소고객 전체의 43.8%)

평균 구매 건수 :

1448 건

1801 건

평균 구매 금액 :

27,629,772 원

47,522,023 원

[ 구매 변화 상위 3개, 하위 2개 ]

구분	평균 구매건수 등급증감
스포츠/레저	0.29
식품	0.28
인류/잡화	0.27
가구/인테리어	-0.21
가전	0.020

구분	평균 구매금액 등급증감
식품	0.39
인류/잡화	0.26
일상용품	0.22
가구/인테리어	-0.23
가전	-0.018

[ 구매 등급증가 변화 ]

구분	평균 구매건수 등급증감	평균 구매금액 등급증감
A	0.11	0.13
B	0.10	0.14
C	0.06	0.06
D	-0.06	-0.04
편의품	0.27	0.37
선매품	0.24	0.20
전문품	0.02	-0.11

## 〈인사이트〉

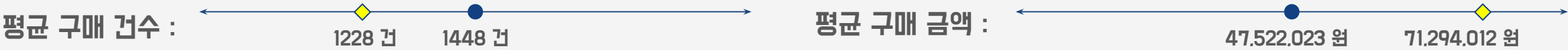
1. 군집A,B,C,D 평균 대비 평균 구매금액은 낮고 평균 구매건수는 높은 고객이다.
2. 편의품과, 선매품의 구매건수 및 구매금액 감소가 눈에 띄게 높은 고객이다.
3. 편의품에서는 식품이, 선매품에서는 스포츠/레저와 인류/잡화의 구매건수가 감소하였다.
4. A,B,C사 평균 구매건수 및 구매금액이 전반적으로 감소한 고객이다.

## 〈마케팅 방안〉

- 평균 구매금액이 낮고 평균 건수가 높은 성질을 띠고 있으므로 중저가의 상품 위주로 마케팅 필요함.
- 감소한 선매품과 편의품에 대한 할인 프로모션과 품목의 다양성 확보, 신제품 발굴 및 빠른 확보 등이 필요할 것으로 생각됨.
- 편의품의 구매 건수와 금액이 전반적으로 감소하였으므로 편의품에 대한 베네핏(ex.카테고리 한정 쿠폰, 카테고리한정 적립 2배)을 제공하여 구매 욕구 상승을 유도.

# 03 타겟 고객 INSIGHT : 군집2

1791명(감소고객 전체의 18.4%)



[ 구매 변화 상위 3개, 하위 2개 ]

구분	평균 구매건수 등급감소
명품	0.42
인류/잡화	0.20
뷰티	0.19
가전	-0.87
가구/인테리어	-0.25

구분	평균 구매금액 등급감소
명품	0.21
인류/잡화	0.20
식품	0.15
가전	-0.93
가구/인테리어	-0.27

[ 구매 등급증가 변화 ]

구분	평균 구매건수 등급증감	평균 구매금액 등급증감
A	0.10	0.12
B	-0.03	-0.02
C	0.01	-0.01
D	-0.10	-0.08
편의품	0.13	0.14
선매품	0.14	0.09
전문품	-0.09	-0.11

## <인사이트>

1. 군집 평균 대비 구매건수는 낮고 구매금액은 높았다.
2. 명품과 의류/잡화의 구매가 감소했고, 편의품의 구매가 감소한 고객이다.
3. 가전, 가구/인테리어와 A제휴사에서 구매가 증가한 고객이다.
4. A제휴사에서 구매가 감소, C제휴사에서 변화가 미약, B,D제휴사에서는 증가한 고객임을 확인할 수 있다.

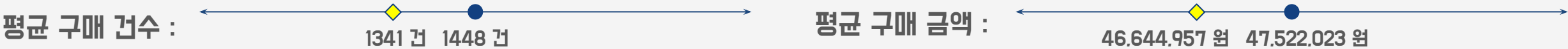
## < 마케팅 방안 >

- 평균 구매 금액이 높고 평균 건수가 낮은 성질을 띠고 있으며 전문품의 구매 증가가 눈에 띄므로 고가 상품 위주 마케팅이 필요함.
- 가구/인테리어, 가전에 대한 구매 건수 및 금액이 증가하였으므로 해당 항목에 대한 베스트 상품 등을 추천.
- 구매 건수 및 구매 금액이 감소한 의류/잡화/뷰티와 관련한 할인 행사를 진행한다.



# 03 타겟 고객 INSIGHT : 군집3

1875명(감소고객 전체의 19.3%)



[ 구매 변화 상위 3개, 하위 2개 ]

구분	평균 구매건수 등급감소
가전	2.14
명품	0.31
인류/잡화	0.25
가구/인테리어	0.12
뷰티	0.14

구분	평균 구매금액 등급감소
가전	2.02
식품	0.34
일상용품	0.28
가구/인테리어	0.10
명품	0.16

[ 구매 등급증가 변화 ]

구분	평균 구매건수 등급증감	평균 구매금액 등급증감
A	0.18	0.22
B	0.08	0.14
C	0.01	0.03
D	-0.09	-0.06
편의품	0.19	0.22
선매품	0.26	0.30
전문품	0.09	-0.01

## 〈인사이트〉

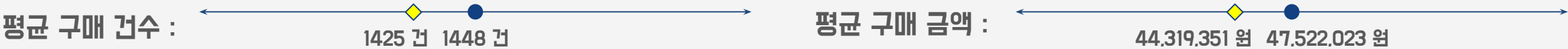
1. 군집 평균 대비 구매건수가 낮고 구매금액도 낮은 고객이다.
2. 가전에서의 구매가 크게 감소했고, 식품, 명품에서 감소가 있으며 가구/인테리어에서 구매감소가 낮다.
3. 편의품, 선매품의 구매건수 및 구매금액의 등급이 감소하였으며, 전문품의 평균 구매금액만 증가했다.
4. D 제휴사를 제외한 A.B.C 제휴사의 평균 구매건수, 금액 모두 감소했다.

## 〈마케팅 방안〉

- 평균 구매 금액이 낮고 평균 건수도 낮은 고객으로 이탈 고객이 될 가능성이 큰 군집.
- 전반적으로 모든 품목에서 구매 감소가 보이므로 고객 특성을 이용한 상품 추천 활용 필요
- 가전 관련된 전문 코너를 새로 오픈하는 등, 이탈 예방을 위한 대책 마련이 시급함.

# 03 타겟 고객 INSIGHT : 군집4

1805명(감소고객 전체의 19.3%)



[ 구매 변화 상위 3개, 하위 2개 ]

구분	평균 구매건수 등급감소
가구/인테리어	1.14
스포츠/레저	0.39
명품	0.30
일상용품	0.18
가전	0.06

구분	평균 구매금액 등급감소
가구/인테리어	1.27
스포츠/레저	0.32
식품	0.31
명품	0.11
가전	-0.01

[ 구매 등급증가 변화 ]

구분	평균 구매건수 등급증감	평균 구매금액 등급증감
A	0.24	0.22
B	0.10	0.15
C	0.01	0.01
D	-0.08	-0.07
편의품	0.22	0.29
선매품	0.24	0.15
전문품	0.48	-0.30

## 〈인사이트〉

1. 군집 평균 대비 구매금액이 낮고 구매건수도 낮은 고객이다.
2. 가구/인테리어, 스포츠/레저와 같은 선매품에서 구매건수 및 구매금액이 뚜렷하게 감소한 것을 확인할 수 있는 고객이다.
3. 삼분류에서 전문품의 평균 구매건수 감소가 가장 큰 고객이다.
4. A,B,C사에서 전반적으로 구매금액이 감소한 고객이다.

## 〈마케팅 방안〉

- 평균 구매 금액이 낮고 평균 건수도 낮은 고객으로 이탈 고객이 될 가능성이 큰 군집
- 가전에 대한 구매 건수 및 금액이 증가하였으므로 해당 항목에 대한 베스트 상품 등을 추천.
- 트렌디한 가구/인테리어 상품을 전시한 팝업스토어를 런칭하여 기존고객층의 이탈을 예방하고 새로운 고객층 확보를 시도.



## 04 제안 및 정리

## 04 제안 및 정리

### 추가적인 활용방안

< 군집별 고객별 맞춤 추천  
상품 >

- 
- 
- 
- 

< 뉴스레터 >

- 할인 쿠폰 발급
- 신제품 서포터즈 모집 정보
- 할인 행사 정보
- 기념일 축하 쿠폰 발급(연1회)

04

제안 및 정리

추가적인 활용방안

뉴스레터 구성 예시

Newsletter

클릭하고 선물받기 🎅

뉴스레터를 구독하면 선물이 와르르

예시 이미지 사이즈  
285 x 285

고객이 좋아할만한

고객맞춤 상품 추천1

상품 추천 🐱

추천

고객에게 딱 맞는

고객맞춤 상품 추천2

고객 맞춤 상품 추천

추천

GRAY

RED

PINK

GRAPE

VIOLET

INDIGO

BLUE

CYAN

TEAL

GREEN

LIME

YELLOW

ORANGE



what we used :



**조경림(팀장)**

- GitHub : <https://github.com/00ong>
- E-Mail : [philosophic.code@gmail.com](mailto:philosophic.code@gmail.com)
-

**정하림**

- GitHub : <https://github.com/hharimjung>
- E-Mail : [halim7401@naver.com](mailto:halim7401@naver.com)

**박시우**

- GitHub : [PerfectTruth \(github.com\)](#)
- E-Mail : [truestar0807@gmail.com](mailto:truestar0807@gmail.com)

**송유빈**

- GitHub : <https://github.com/yibnn>
- E-Mail : [hiyub.985@gmail.com](mailto:hiyub.985@gmail.com)

**최혜정**

- GitHub : <https://github.com/601chl>
- E-Mail : [gwe121292@gmail.com](mailto:gwe121292@gmail.com)



Direct us!

**CAKD3\_Team\_2 :** 팀워크와 분위기는 단연 원내 1등이라고 자부할 수 있었던 팀. 미숙한 부분은 서로 채워주며 **#협조 #긍정 #적극성** 어려움을 배움으로 승화시킬 수 있었다. 사랑을 담은 메시지를 Direct로 보냅니다.



# THANK YOU

K-Digital Training 3기  
2조 I 송유빈, 정하림, 박시우, 조경림, 최혜정

2021.10.18