

유통 환경 변화에 따른 구매 감소 고객 예측 솔루션

1팀 [임희진, 최윤진, 한아름]

21.10.18 presentation



Team Leader

- 프로젝트 기획 총괄
- 독립변수 체계화 총괄
- 전략 및 솔루션 체계화

- [Lim Heejin](#)
- dg961108@naver.com
- <https://github.com/heejvely>



Team Member

- Business issue finding 총괄
- 구매 상품 변동 독립변수 체계화
- Domain, 외부 자료 총괄

- [Choi Yunjin](#)
- cjunjin@gmail.com
- <https://github.com/ete-llorona>

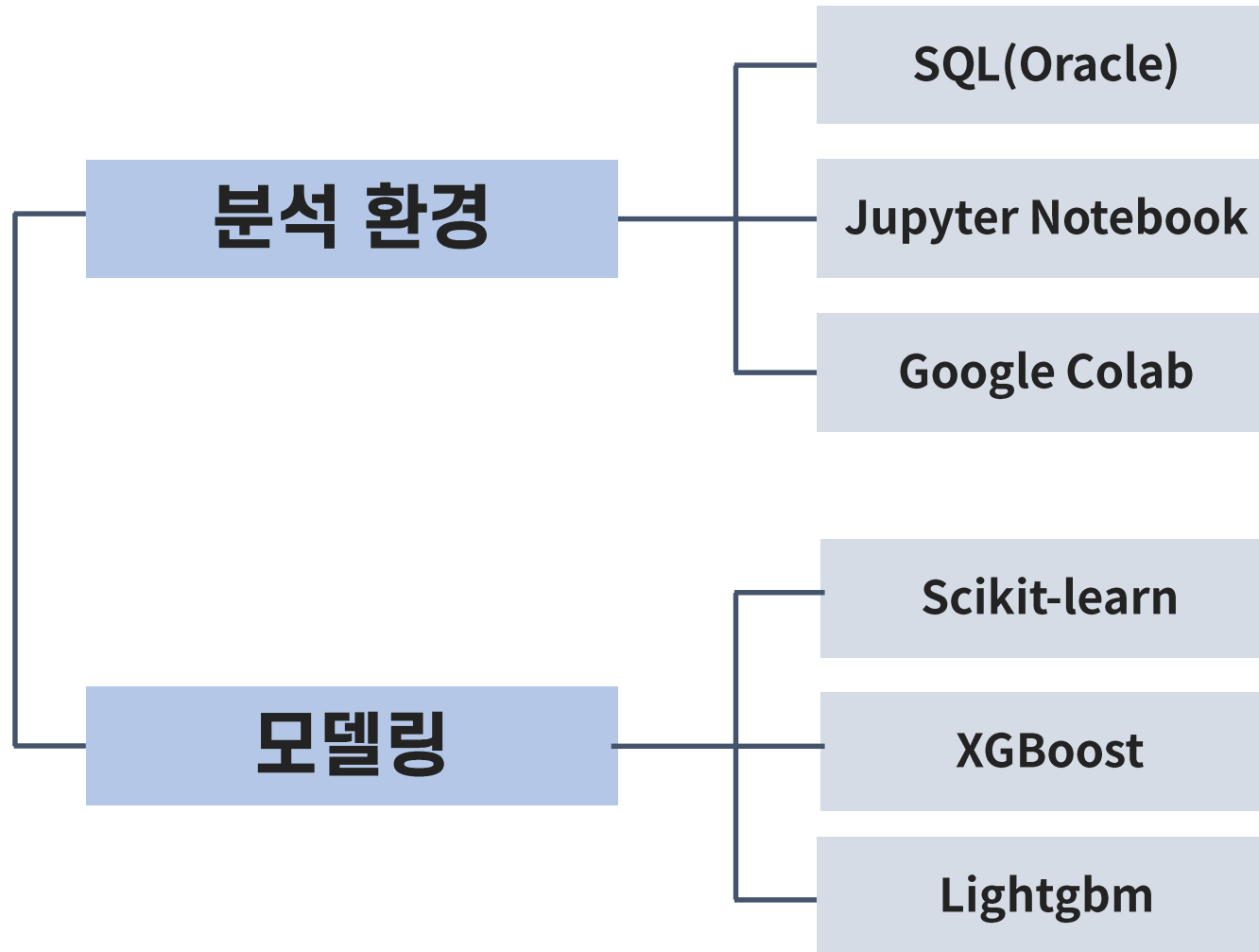


Team Member

- 기술 및 코드 총괄
- 구매 변동 패턴 독립변수 체계화
- 예측 모델링 체계화

- [Han A-Leum](#)
- 20172012@sungshin.ac.kr
- <https://github.com/zena-H>

프로젝트 진행 환경



ORACLE®
DATABASE

 **Jupyter**

 **colab**


 **scikit learn**

XGBoost

 **LightGBM**

Index

-	Issue	유통 산업 환경 변화
-	Problems	L사 감소 고객 분석
-	Problems	Business Issue 정의
-	Modeling	감소 고객 예측 모델링
	Analysis	예측 모델링 평가 및 분석
	Solution	감소 고객 유형에 따른 마케팅 제언



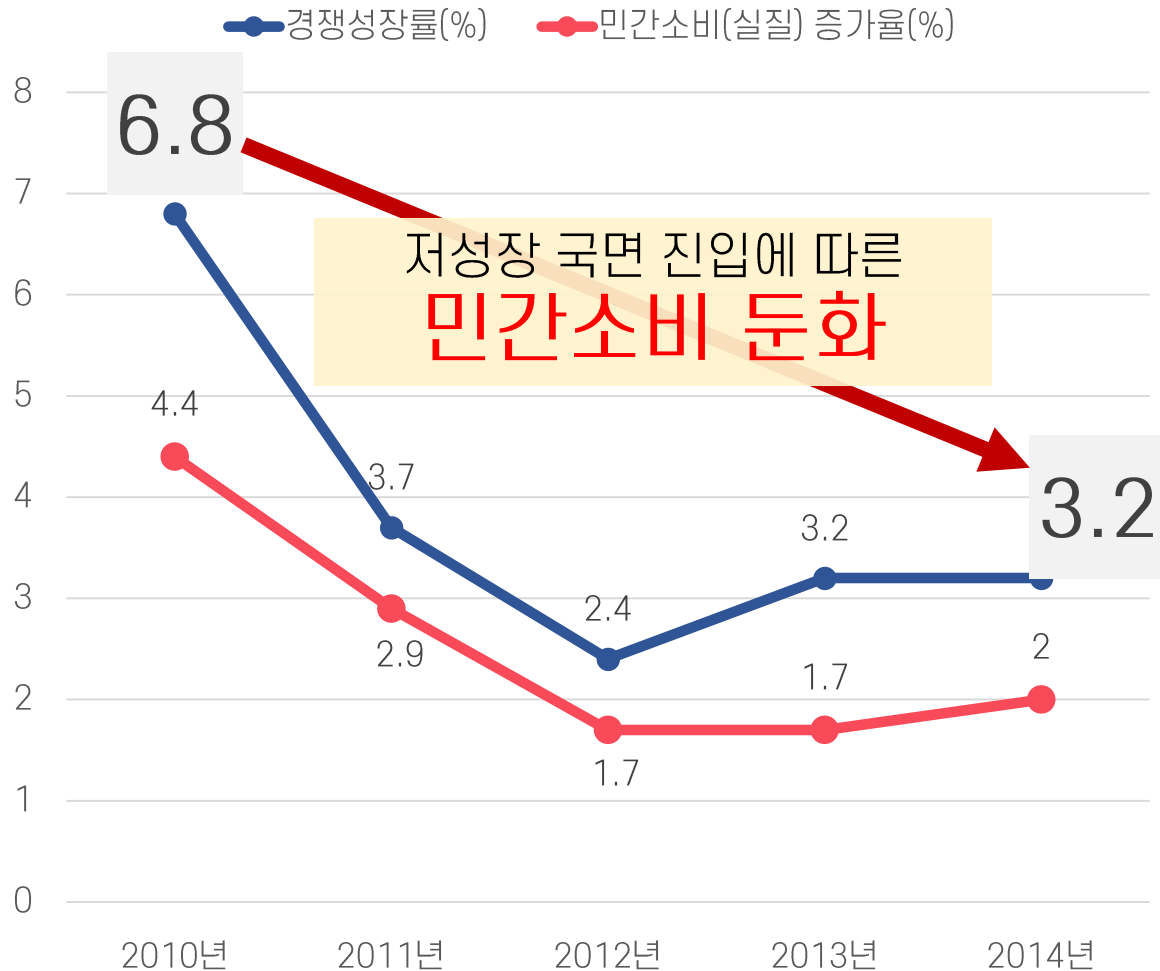
Issue.

유통 산업 환경 변화

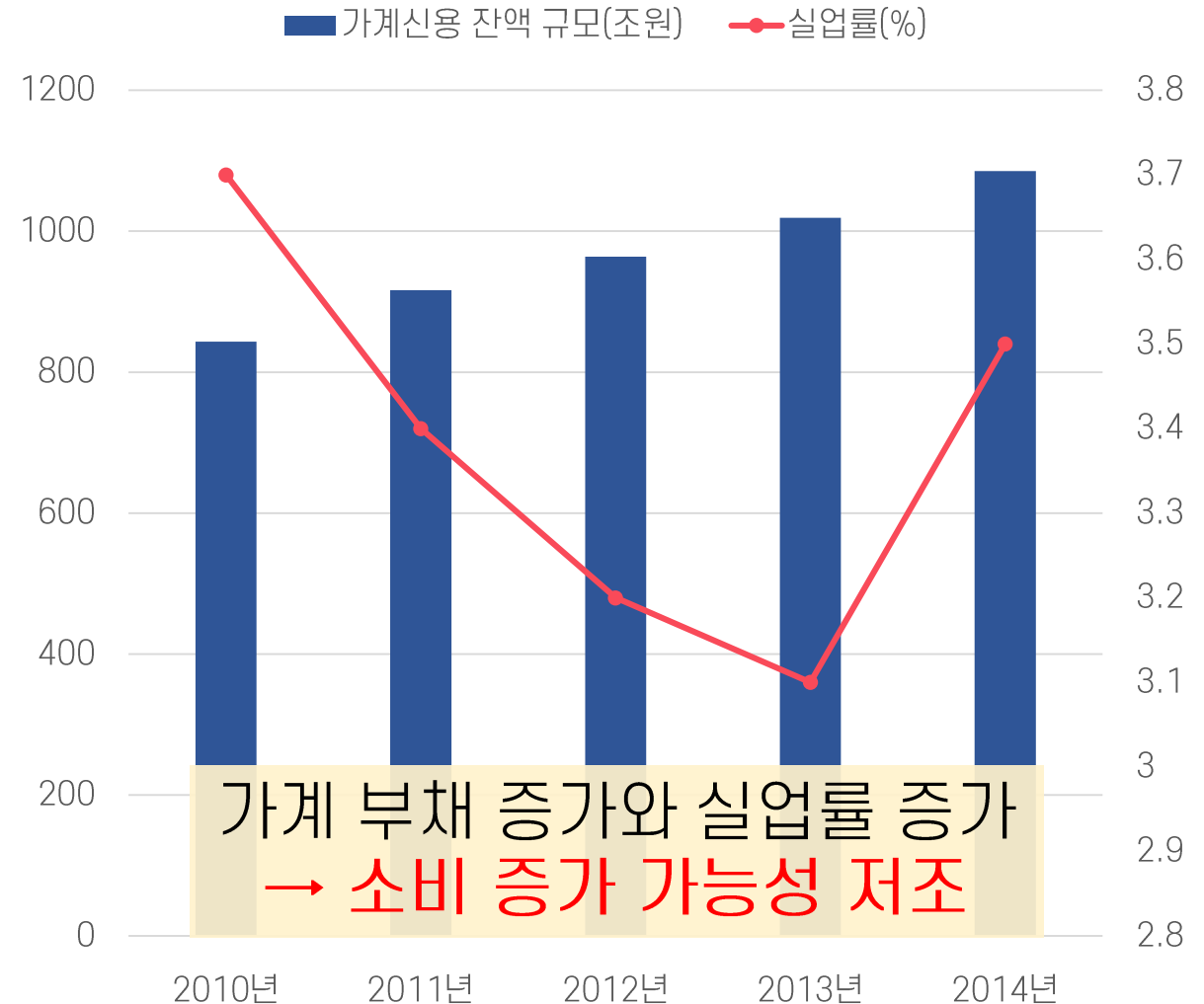
- 민간 소비 둔화 기조
- 정부 내수 활성화 노력

유통 업계 현실 - [민간소비 둔화 기조]

* 경제 성장률 및 소비 증가율

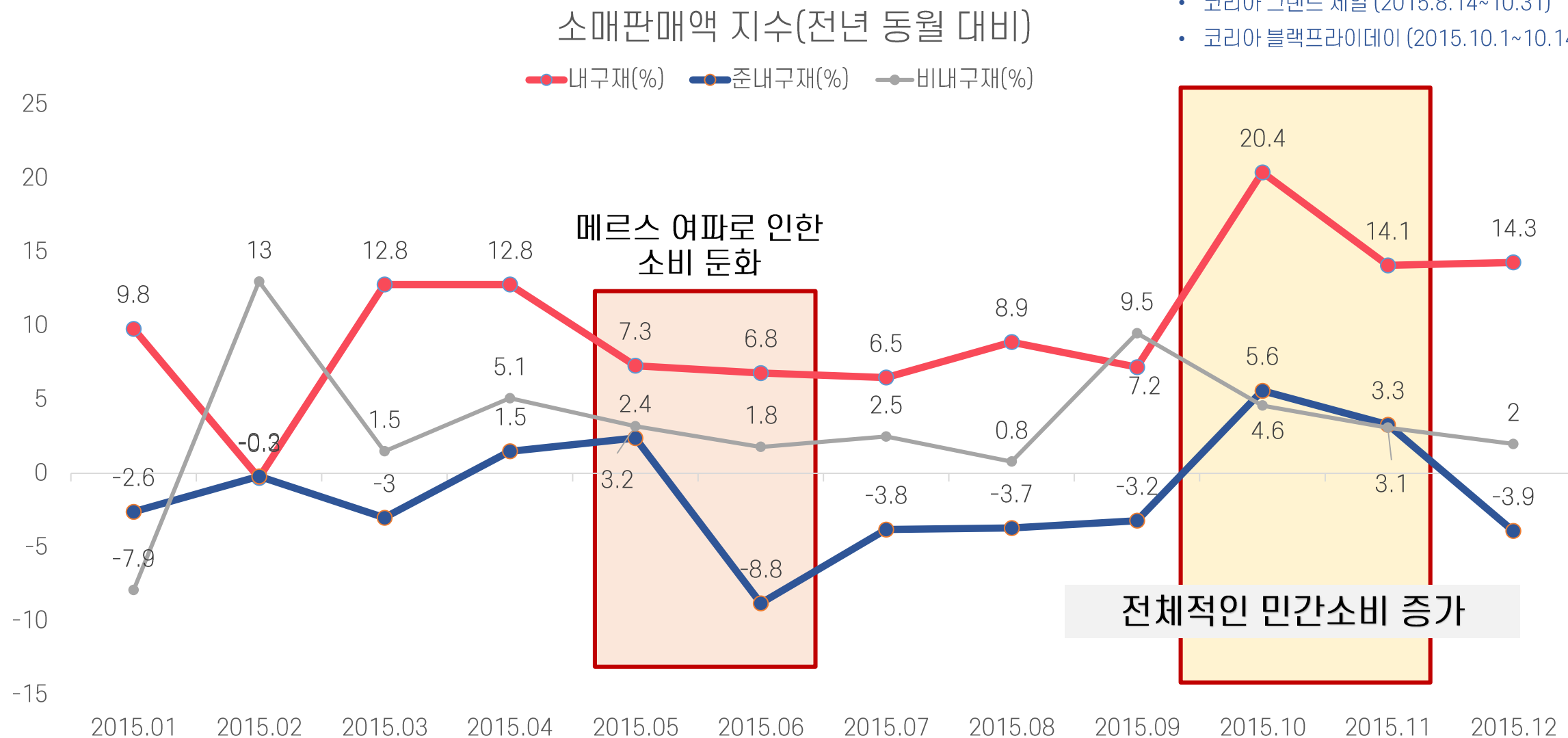


* 가계신용 및 실업률

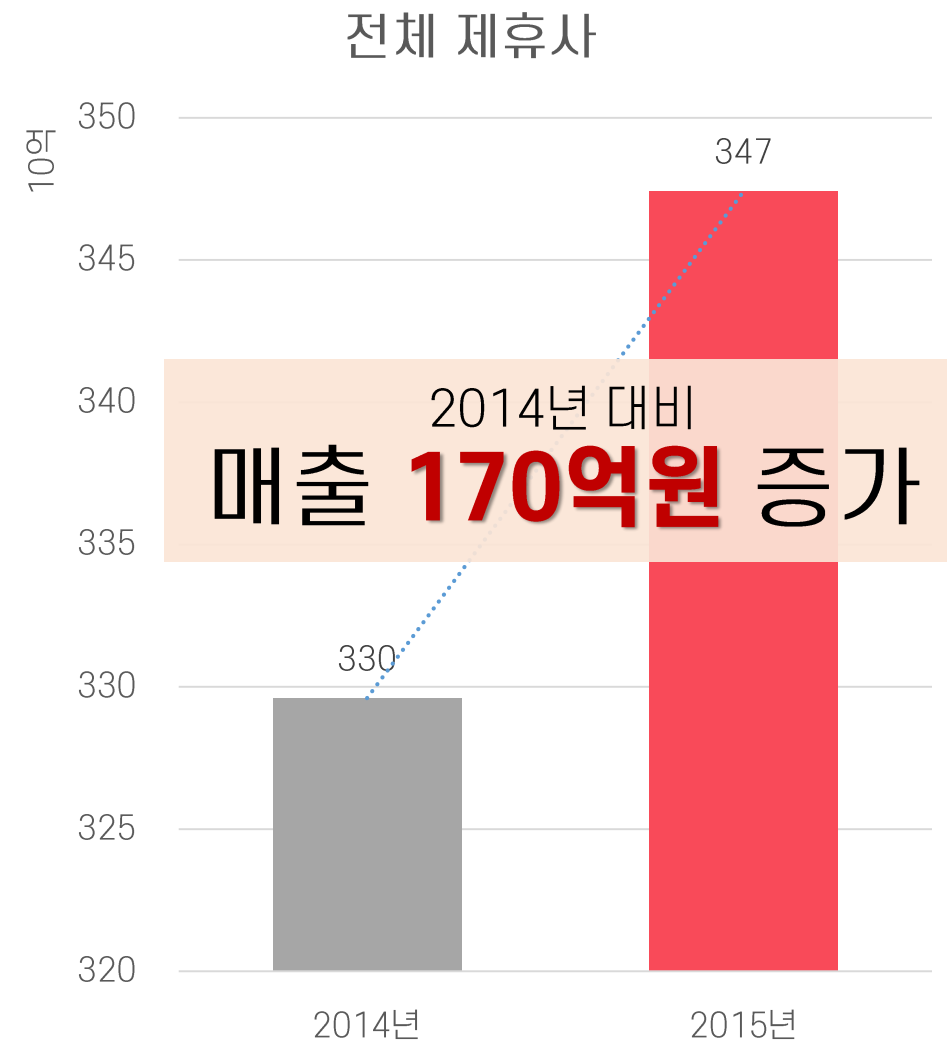
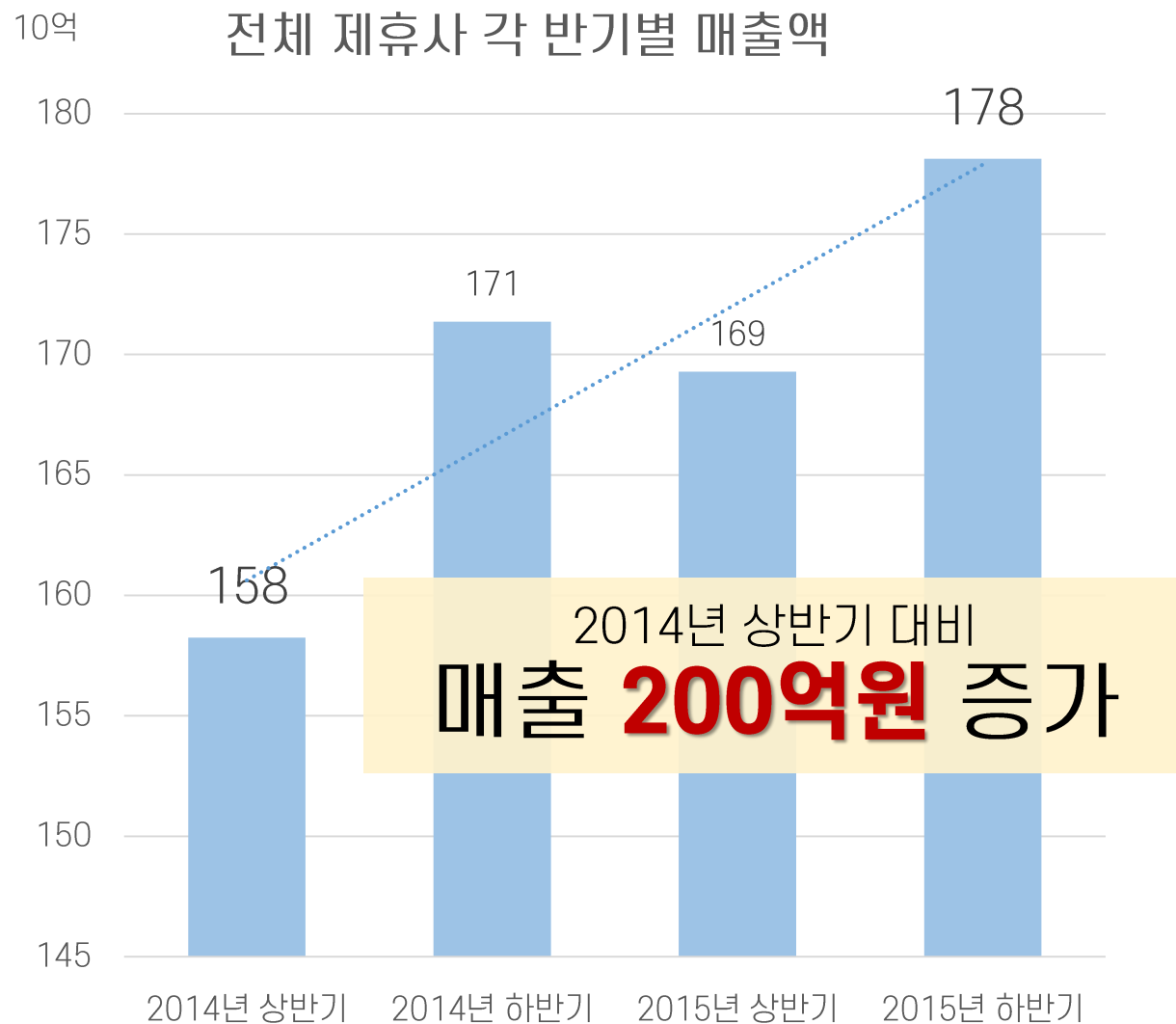


정부의 국내 경기 활성화 노력

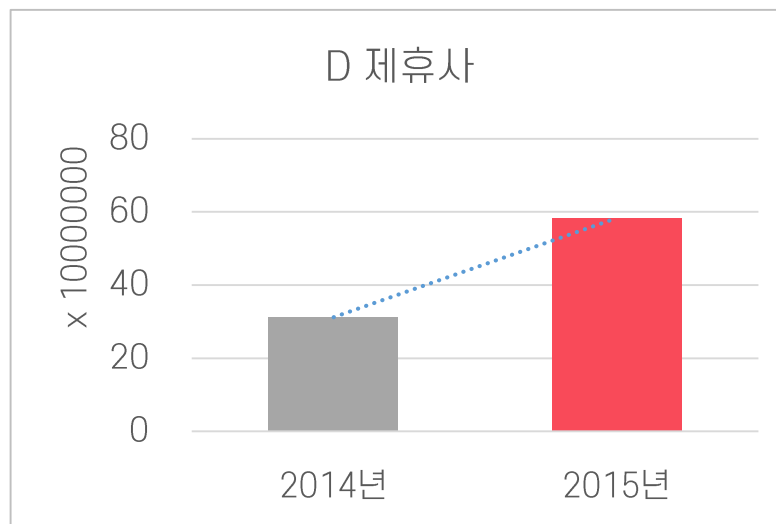
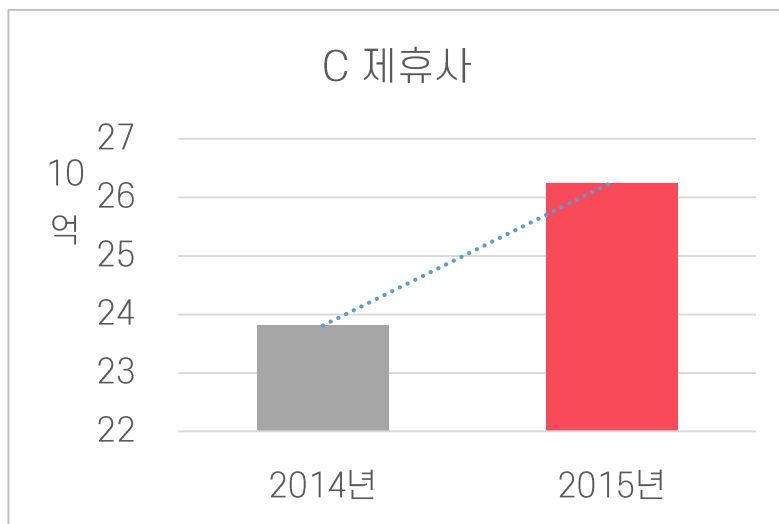
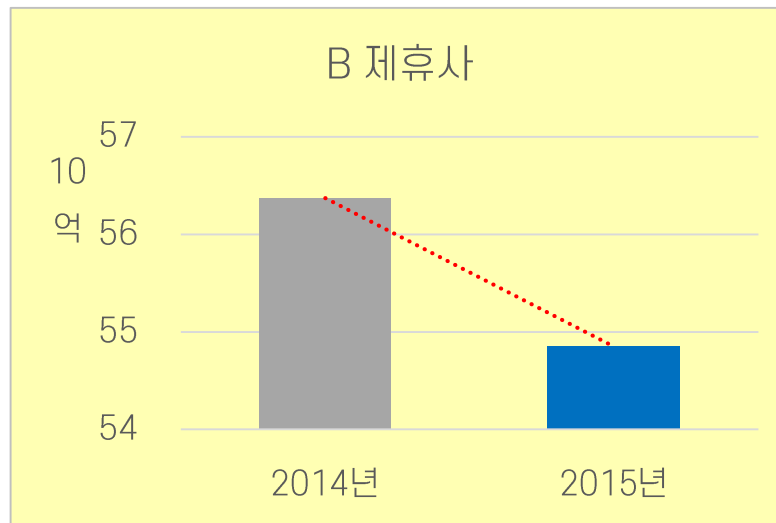
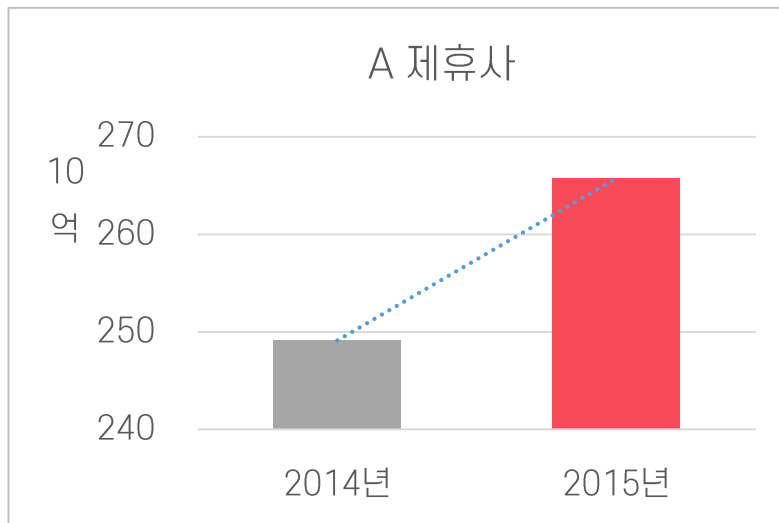
- 개별소비세 인하 (2015.8.27~2015.12.31)
- 코리아 그랜드 세일 (2015.8.14~10.31)
- 코리아 블랙프라이데이 (2015.10.1~10.14)



L사 전체 매출 동향



[L사 14년도 ~ 15년도 제휴사 별 매출 현황표]



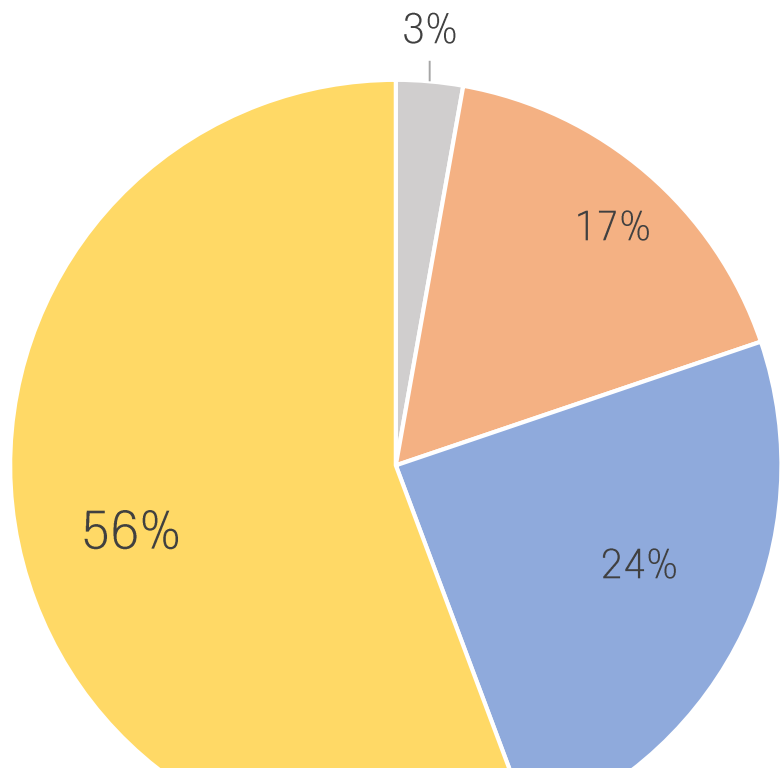
전체적으로 매출 증가한 반면

**제휴사 B
약 10억원 감소**

L사 전체 매출 동향 - 준내구재, 비내구재 매출 감소

L사 매출 구성

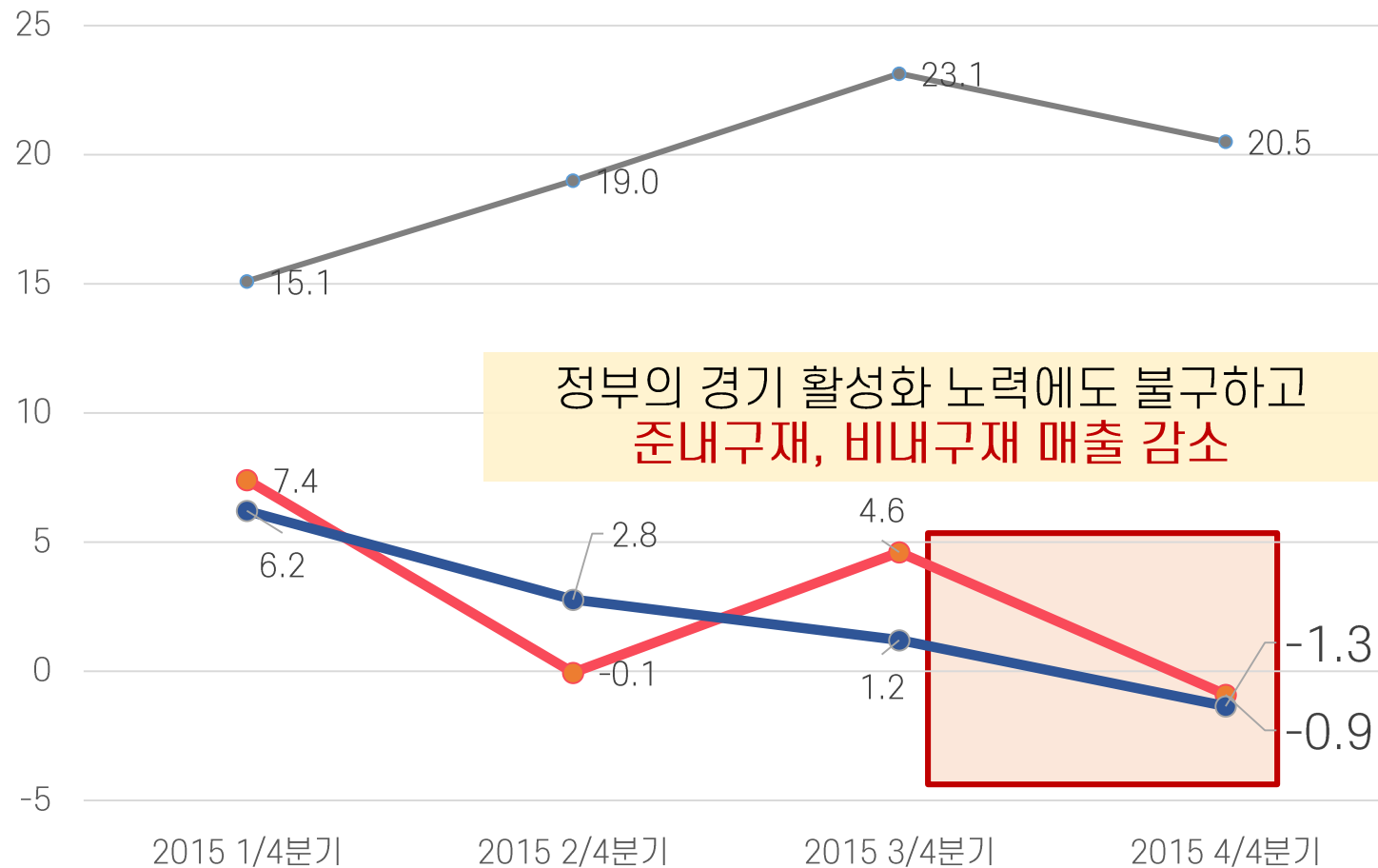
■ 서비스 및 기타 ■ 내구재 ■ 비내구재 ■ 준내구재



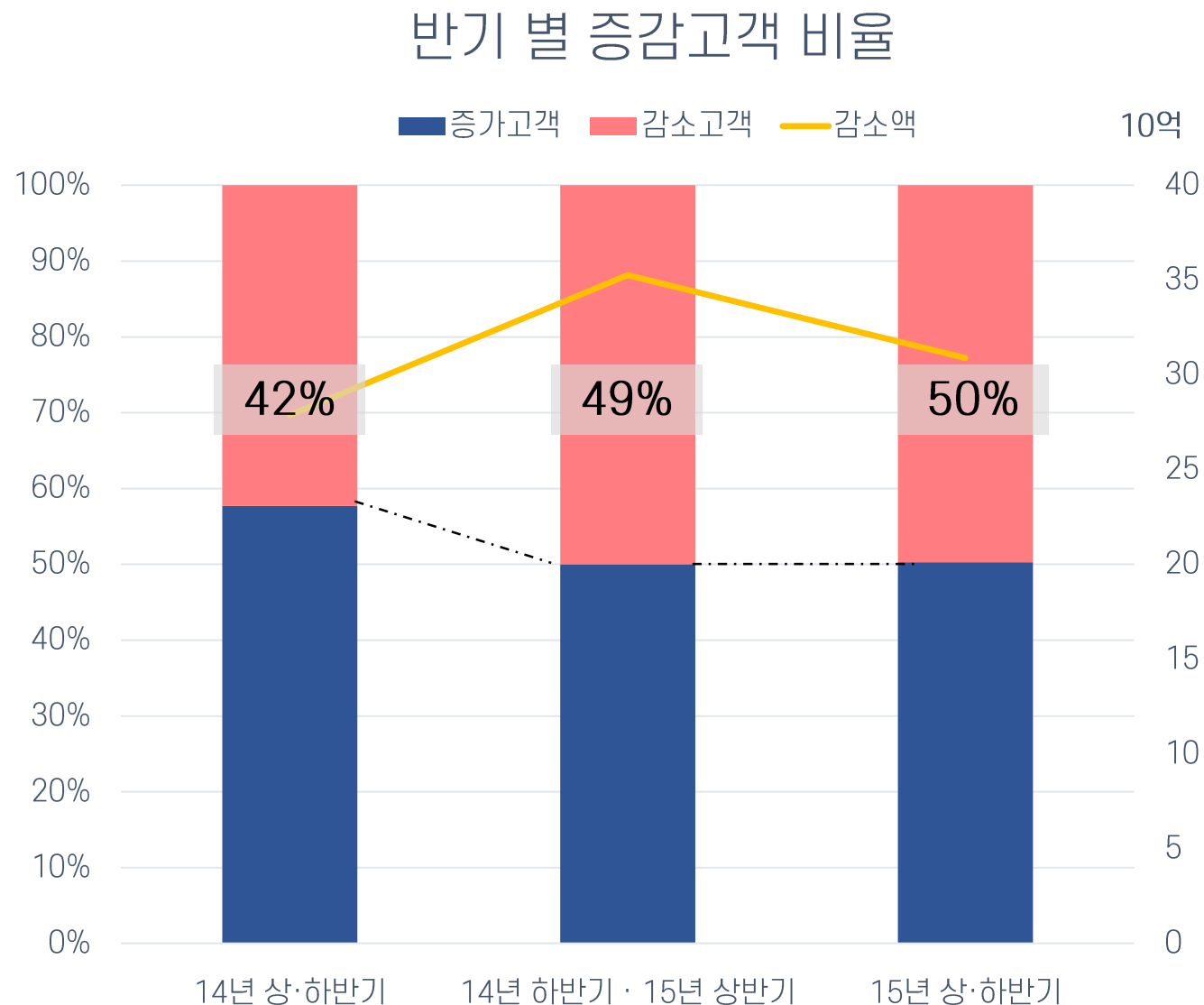
주요 매출 구성 항목
“준내구재, 비내구재”

L사 전년 동기 대비 매출 동향

● 내구재(%) ● 준내구재(%) ● 비내구재(%)



L사 전체 매출 동향 - 구매 감소 고객 비중 증가



구매 금액이 감소하는 고객 비중 증가

[42% → 50%]

2015년 하반기

구매 감소 고객 9,605명

구매 감소 고객 구성 - 일반 등급 고객 비중 증가

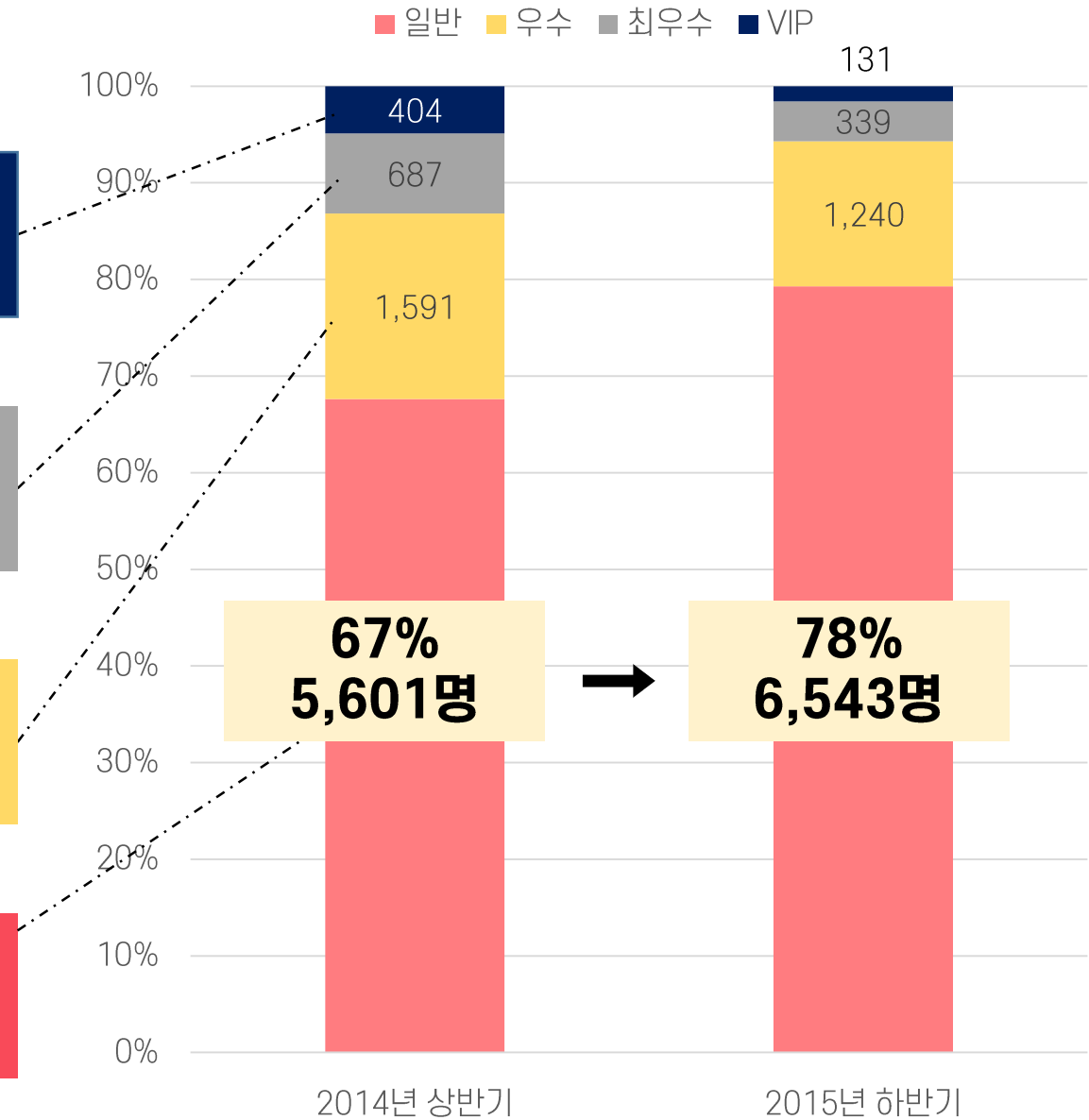
14년 상반기 대비 15년 하반기 감소 고객 등급 비율

VIP: 6개월 동안 3,435만원 이상 구매한 고객

최우수: 6개월 동안 2,015만원 이상 3,435만원 미만 구매한 고객

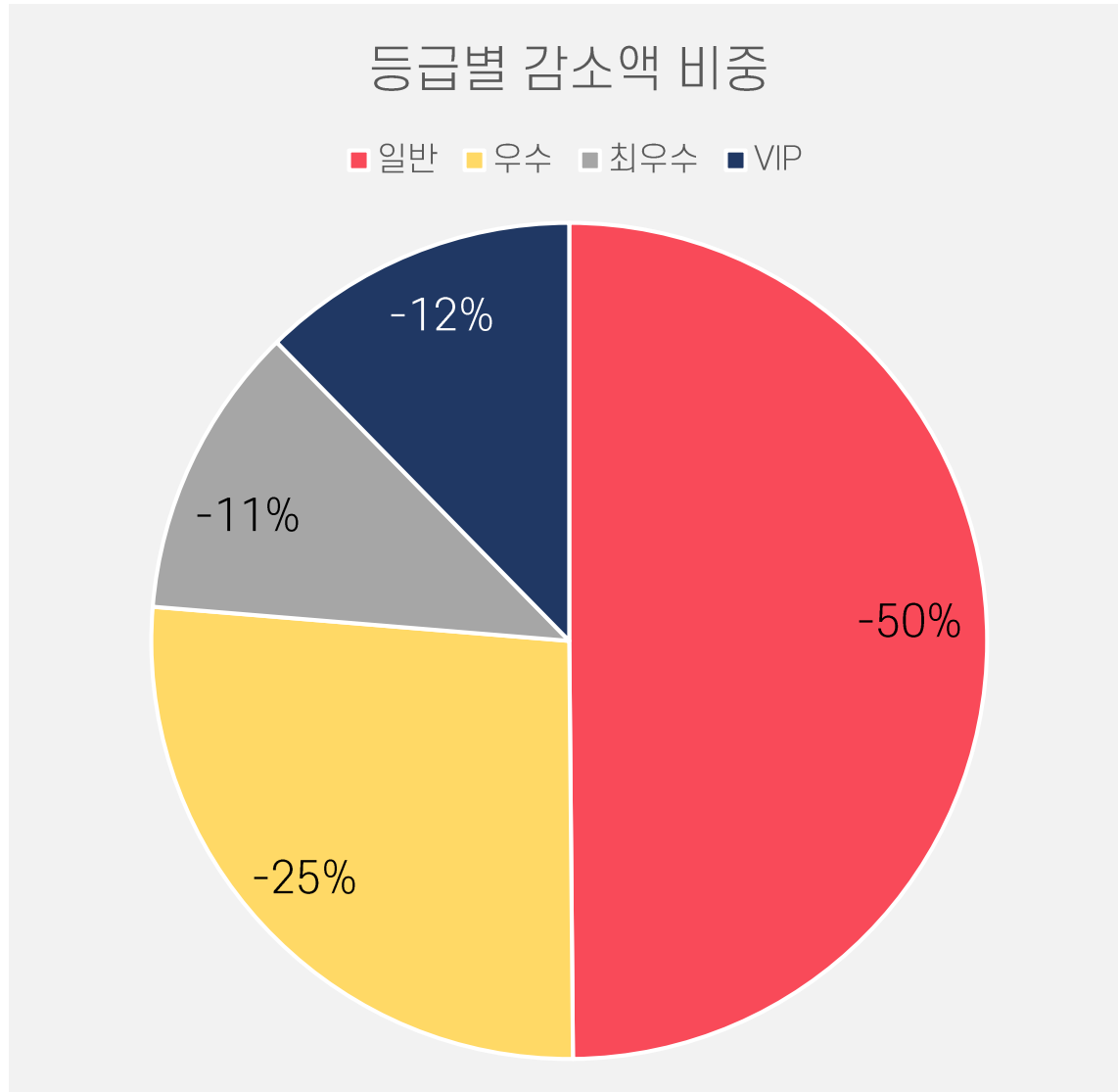
우수: 6개월 동안 870만원 이상 2,015만원 미만 구매한 고객

일반: 6개월 동안 870만원 미만 구매한 고객



*등급 기준: L사 각 제휴사 별 등급 기준 금액의 합

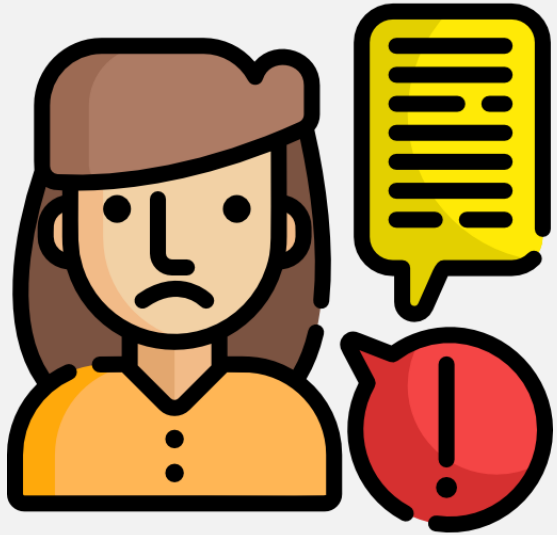
구매 감소 고객 구성 - 등급별 감소액 비중



2014년 상반기 대비
총 구매 감소 금액 **약 357억 원**

일반 등급으로 이동한 고객의
총 구매 감소 금액 **약 177억 원**


Business 과제 정의



- 구매액이 크지 않은 일반 등급의 구매 감소 고객
:L사를 구매처로 정착하지 않고
이탈 가능성이 높은 고객들로 정의



- 구매 감소 고객의 구매 패턴을 파악하여
구매 감소를 예방하고, 매출을 증대
하는 것을 Business 과제로 설정



Modeling.

감소 고객 예측 모델링

감소 고객 예측 모델링 Process

1.

구매 감소 고객 패턴 파악

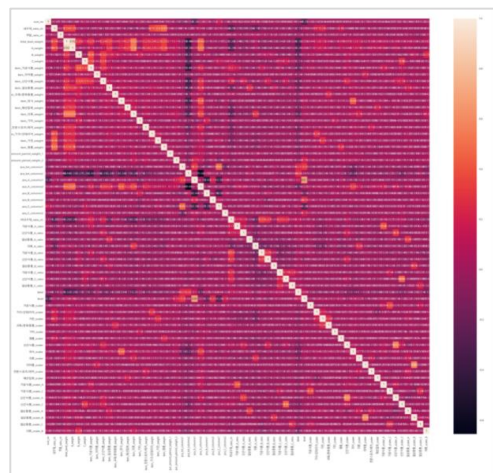
고객 속성 정보 및
구매 내역 분석을 통한
구매 감소 고객의 구매 패턴 파악



2.

독립변수 설정

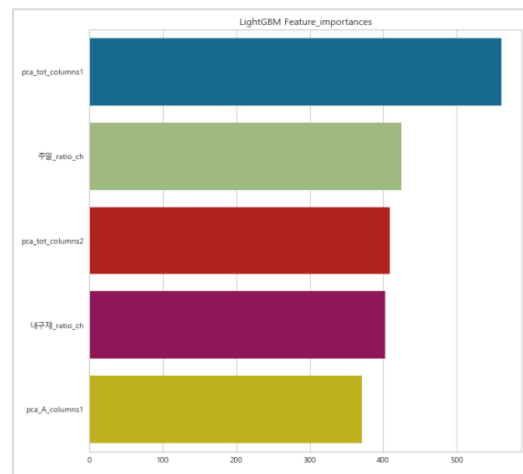
데이터에서 파악된 구매
패턴을 통해 독립변수로 설정



3.

학습 및 검증

학습용 데이터를 통한
학습 및 검증 진행



4.

Test set 예측

Test 데이터를 통한
최종 모델 평가



종속 변수

반기 구매 금액을 기준으로 감소 고객 분류

* Train set 종속변수

: 2014년 상반기(1기) 구매 금액 > 2015년 상반기(3기) 구매 금액

* Test set 종속변수

: 2014년 상반기(1기) 구매 금액 > 2015년 하반기(4기) 구매 금액

고객 분류	Label
감소 예측 고객	1
유지 · 증가 예측 고객	0

2014년 상반기	2014년 하반기	2015년 상반기	2015년 하반기
1기	2기	3기	4기
Train set 독립 변수 시기		Train set 종속 변수 시기	
Test set 독립 변수 시기			Test set 종속 변수 시기

데이터 전처리 [상품 재분류]

* 상품 대분류

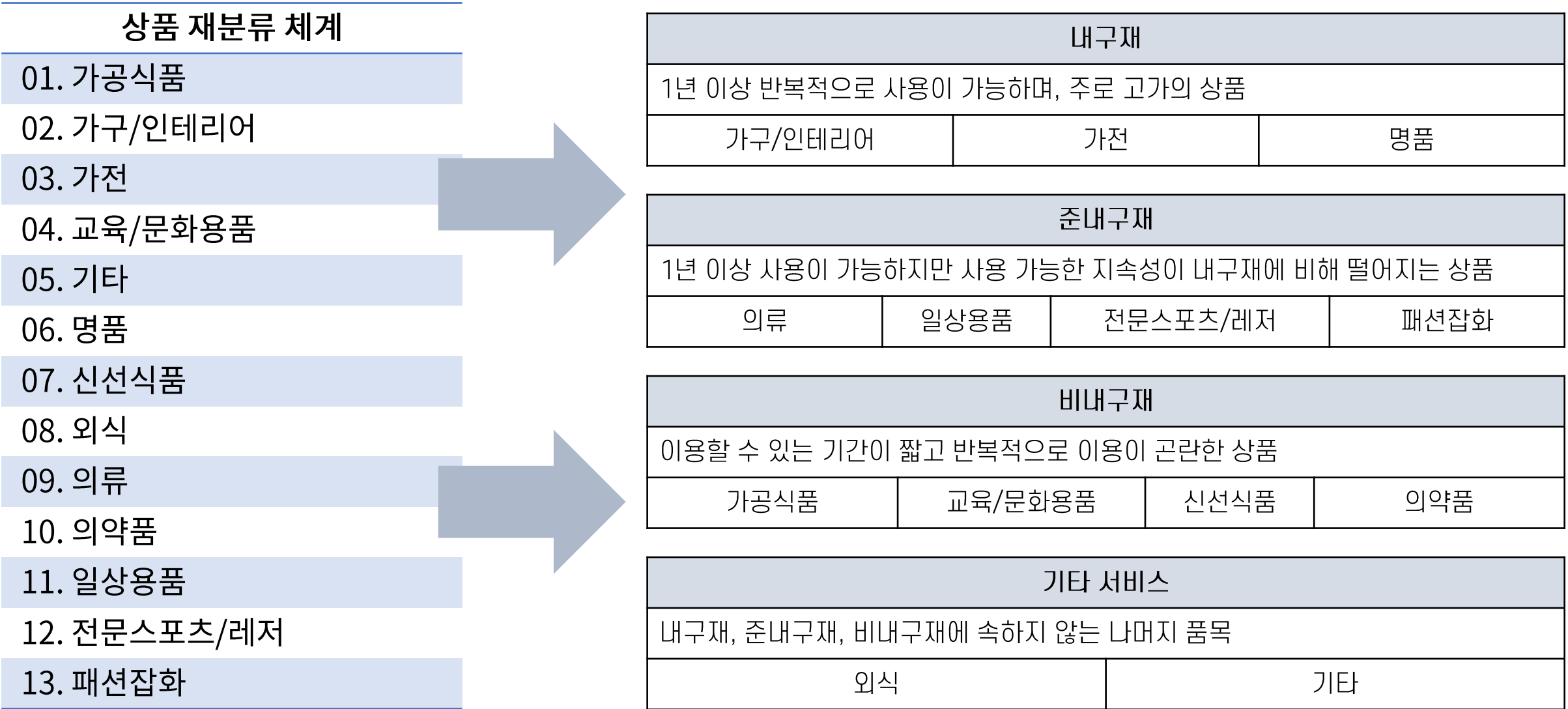
- 01. 가공식품
- 02. 신선식품
- 03. 일상용품
- 05. 의약품/의료기기
- 06. 교육/문화용품
- 07. 디지털/가전
- 08. 가구/인테리어
- 09. 의류
- 10. 전문스포츠/레저
- 11. 패션잡화
- 99. 기타 상품

제휴사 별 통합되지 않은
상품 분류 체계 통합

상품 재분류 체계

- 01. 가공식품
- 02. 가구/인테리어
- 03. 가전
- 04. 교육/문화용품
- 05. 기타
- 06. 명품
- 07. 신선식품
- 08. 외식
- 09. 의류
- 10. 의약품
- 11. 일상용품
- 12. 전문스포츠/레저
- 13. 패션잡화

데이터 전처리 [상품 재분류] - 소비 동향 파악을 위한 상품 분류



주요 독립 변수

독립변수 유형 1

고객 구매 행동 패턴 변수

- 1) 기간 간 구매 금액 변화
- 2) 기간 내 구매 금액
- 3) 기간 내 상품 구매 횟수
- 4) 기간 간 주말 구매 비중 차이

전체/ 제휴사별로 고객의 구매 행동 패턴을 파악

- * 기간 간: 반기/분기 간의 구매 금액/비중 변화
- * 기간 내: 14년 상반기 ~ 15년 상반기의 금액/횟수

독립변수 유형 2

고객 구매 상품 변수

- 1) 기간 내 상품 분류 별 구매액
- 2) 기간 내 제휴사별 특정 상품 구매액
- 3) 기간 내 특정 상품 구매 비중
- 4) 기간 간 내구재, 비내구재 비중 차이

전체/제휴사별 고객의 구매 상품 특성 파악

- * 기간 간: 반기/분기 간의 상품 구매 비중 변화
- * 기간 내: 14년 상반기 ~ 15년 상반기의 금액/비중

독립변수 유형1 - 고객 구매 행동 패턴 변수



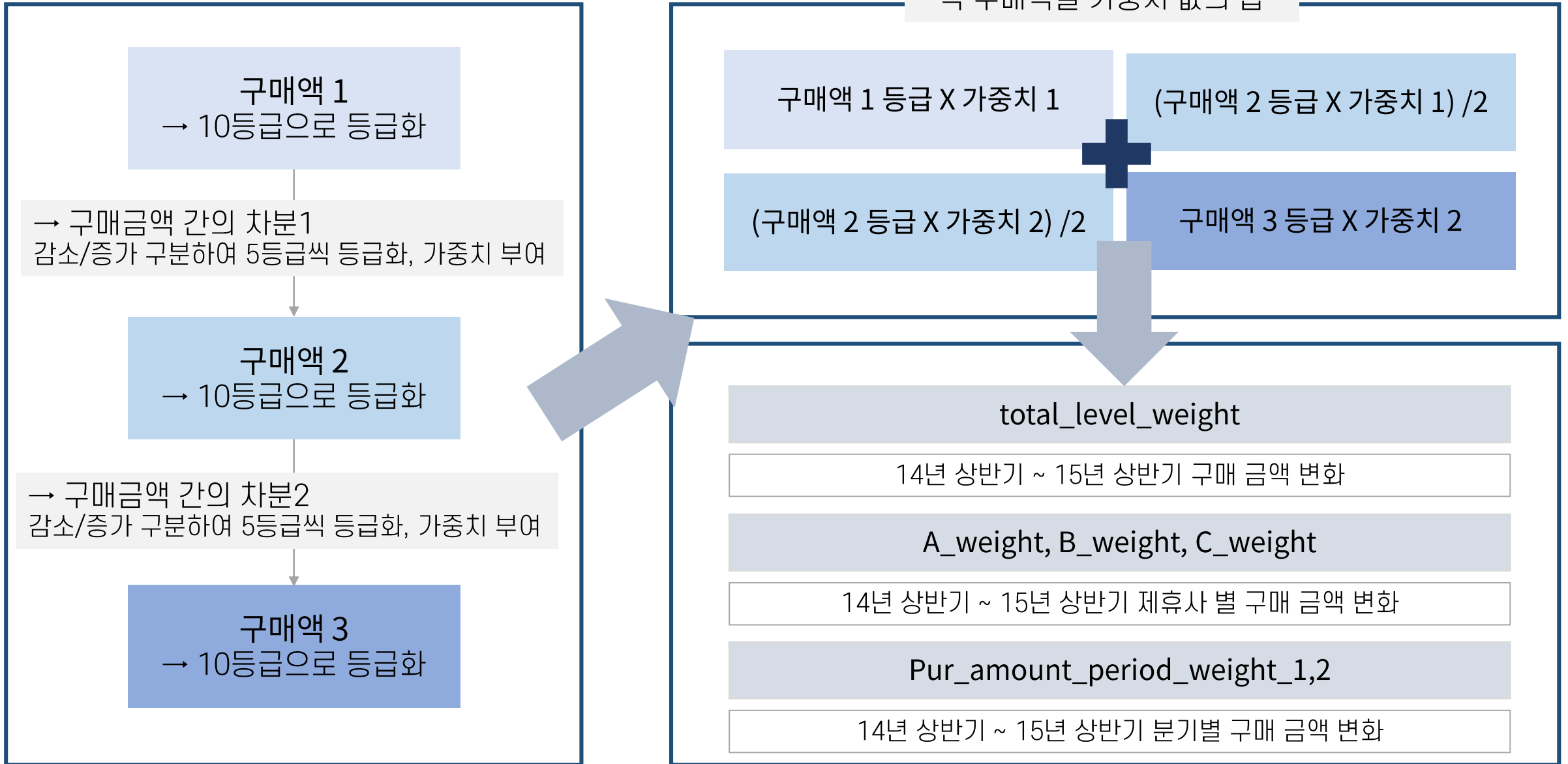
- 1) 반기 구매액 기준 상대성을 고려한 등급화 (구매액이 높을수록 10등급에 위치)
- 2) 반기 간 금액 차이에 따라 증가/ 감소/ 유지 구분
- 3) 증감액에 따라 상대성을 고려한 증가/감소 5등급화 후 가중치 부여
- 4) 등급이 유지인 경우에도, 구매액의 변동이 있을 경우, 가중치 부여
- 5) 유지일 경우, 가중치 0 (유지는 구매증감액이 0원일 경우 적용)
- 6) 구매액2의 경우, 중복 계산되므로 각각 계산 값의 1/2값 계산

[계산 예시]

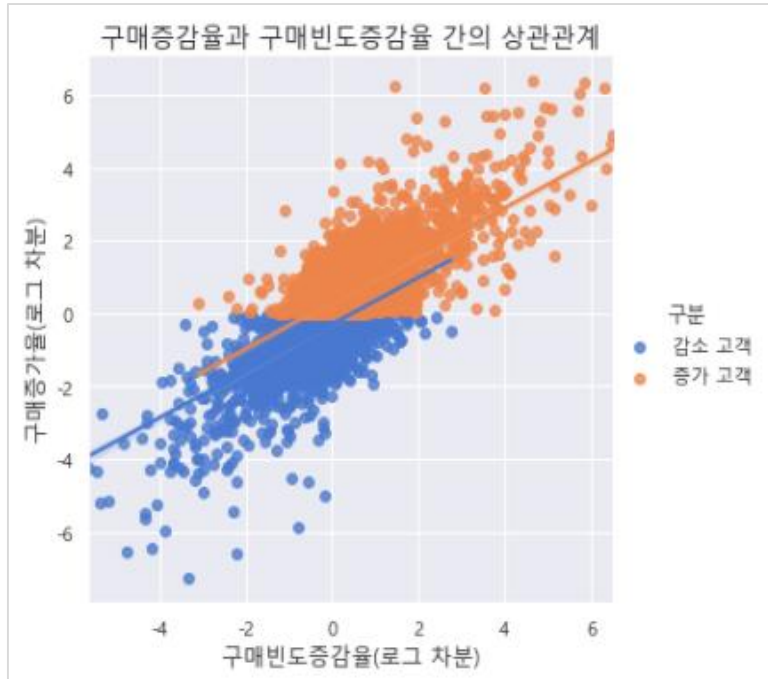
$$\{(5 \times 0.3) + (7 \times 0.3 / 2)\} + \{(7 \times -0.1 / 2) + (7 \times -0.1)\} = (2.55) + (-1.05)$$

최종 값 = 1.5

독립변수 유형1 - 고객 구매 행동 패턴 변수



독립변수 유형1 - 고객 구매 행동 패턴 변수



구매 빈도 감소 시 구매액 감소

total_freq_pca_1, total_freq_pca_2

14년 상반기 ~ 15년 상반기 전체 구매 횟수
주성분 변동성(설명력): [0.687 0.217]

A_freq_pca_1, A_freq_pca_2

14년 상반기 ~ 15년 상반기 A사 반기/분기구매 횟수
주성분 변동성(설명력): [0.867 0.057]

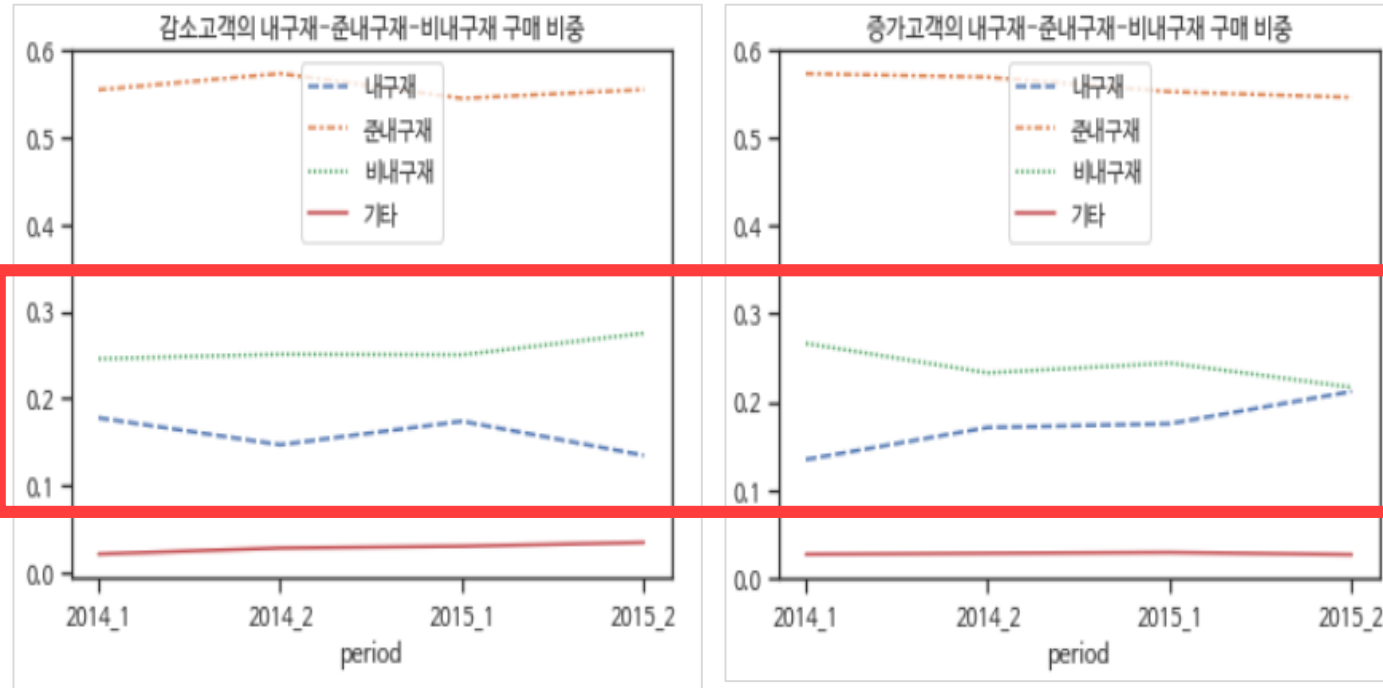
B_freq_pca_1, B_freq_pca_2

14년 상반기 ~ 15년 상반기 B사 반기/분기구매 횟수
주성분 변동성(설명력): [0.867 0.064]

C_freq_pca_1, C_freq_pca_2

14년 상반기 ~ 15년 상반기 C사 반기/분기구매 횟수
주성분 변동성(설명력): [0.914 0.043]

독립변수 유형2 - 고객 구매 상품 변수



내구재_ratio_ch

14년 상반기 ~ 15년 상반기 내구재 비율 변화

비내구재_ratio_ch

14년 상반기 ~ 15년 상반기 비내구재 비율 변화

내구재 구매 비중 감소 시 구매 금액 감소
비내구재 구매 비중 증가 시 구매 금액 감소

독립변수 유형2 - 고객 구매 상품 변수

상품 분류	비중
가공식품	14%
신선식품	10%
의류	31%
일상용품	11%

4개의 주요 상품이 **전체 매출의 66% 비중**
구매 비중이 높은 주요 상품은
고객의 구매 금액 변화에 영향이 큼.

주요 상품_제휴사별_ratio

14년 상반기 ~ 15년 상반기 제휴사별 주요상품 구매 비중

주요 상품_제휴사별_scaled

14년 상반기 ~ 15년 상반기 제휴사별 주요상품 구매액 상대성

독립변수 최종 목록

변수 종류	설명
Level_weight [Total, A/B/C]	구매 금액 변동 지수
Pur_amount_period_weight	반기별 구매 금액 변동 지수
주말_ratio_ch	주말 구매 비중 차분
level	14년 상반기~15년 상반기 기간 내 구매액 지수
Freq_pca	반기/분기 전체, 제휴사별 구매 횟수

변수 종류	설명
내구재/비내구재_ratio_ch	내구재, 비내구재 구매 비중 차분
상품별_weight	상품별 구매 금액 변동 지수
상품별 구매액_scaled	14년 상반기~15년 상반기 전체 구매액 상대성
주요 상품 제휴사 별_scaled	14년 상반기~15년 상반기 주요 상품 제휴사 별 구매액 상대성
주요 상품 제휴사별_ratio	14년 상반기~15년 상반기 주요 상품 제휴사 별 구매 비중

**총 61개
독립변수 설정**

예측 모델 평가

Hyperparameter
: n_estimators=1000

Logistic Regression

Accuracy	0.71
Precision	0.69
Recall	0.58
F1	0.63
ROC AUC	0.77

변수 중요도

1. 비내구재_ratio_ch
2. Total_freq_pca_2
3. 신선식품_C_ratio
4. 가공식품_C_ratio
5. 신선식품_A_scaled

RandomForest Classifier

Accuracy	0.75
Precision	0.71
Recall	0.68
F1	0.7
ROC AUC	0.82

변수 중요도

1. Total_level_weight
2. Total_freq_pca_1
3. A_weight
4. Total_freq_pca_2
5. 의류_scaled

Hyperparameter
: n_estimators=500
Learning_rate=0.1
N_jobs=-1

XGB Classifier

Accuracy	0.74
Precision	0.7
Recall	0.67
F1	0.68
ROC AUC	0.81

변수 중요도

1. Total_freq_pca_1
2. 주말_ratio_ch
3. Total_freq_pca_2
4. 내구재_ratio_ch
5. A_freq_pca_1

Hyperparameter
: n_estimators=500
Learning_rate=0.1
N_jobs=-1

LGBM Classifier

Accuracy	0.74
Precision	0.71
Recall	0.69
F1	0.7
ROC AUC	0.81

변수 중요도

1. Total_level_weight
2. Total_freq_pca_1
3. Total_freq_pca_2
4. A_freq_pca_1
5. 가공식품_scaled

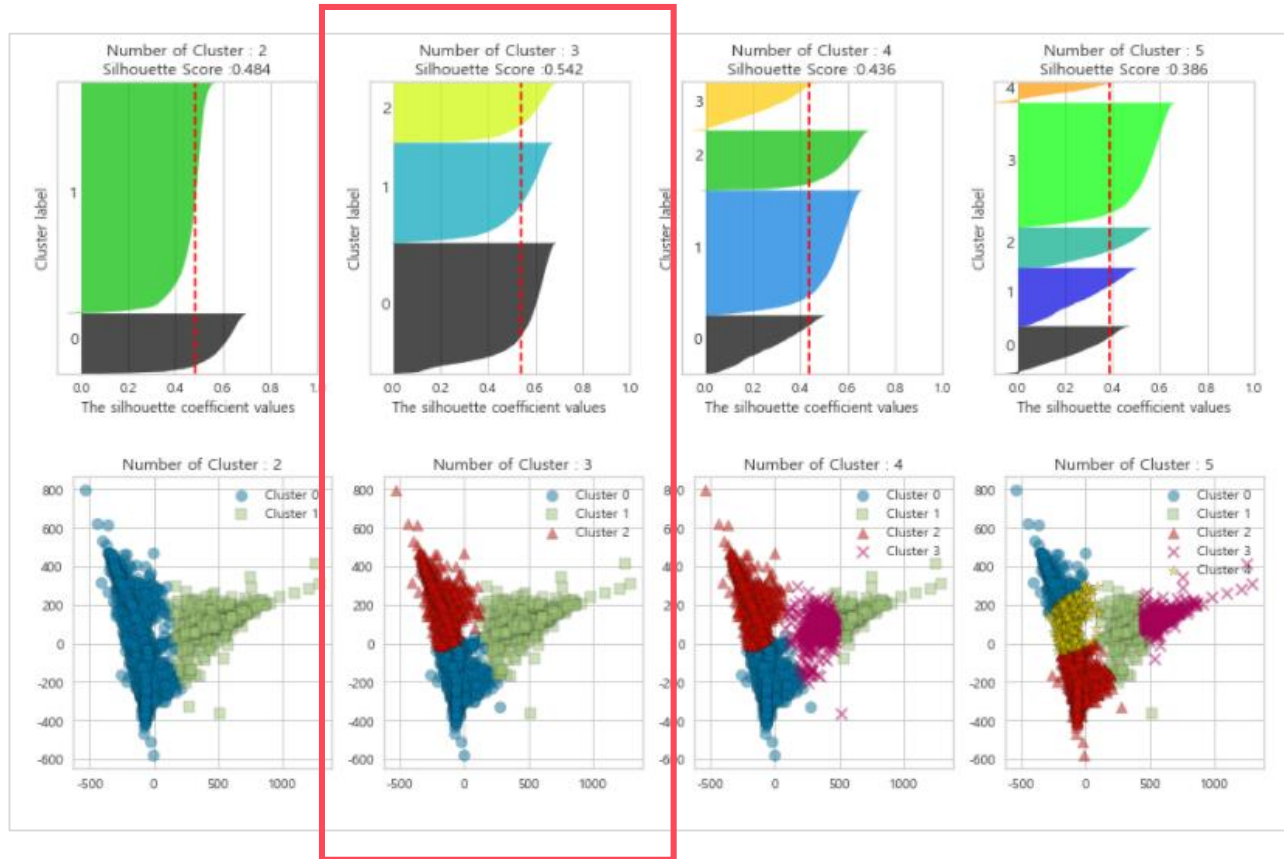


Solution.

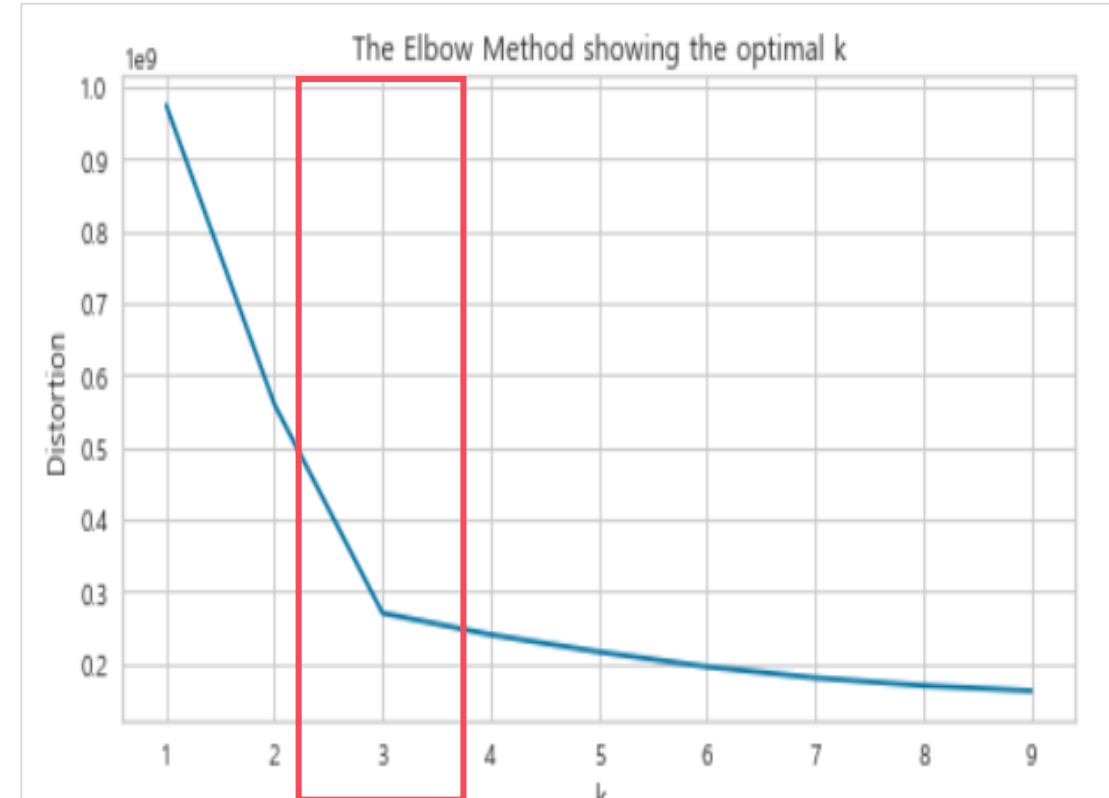
감소 고객 유형에 따른 마케팅 제언

군집 개수 설정

K-means 군집화 시행/ [최종 군집 수: 3개]



Silhouette Score: 0.542, 군집 일정하게 되어있는 3개 선택



경사가 완만해지는 시점인 3개 선택

cluster 0 [1번째 군집] - 경쟁사 이탈 가능성 높은 군집

상품 분류	평균 Weight 값
의류	-1.066416
전문스포츠/레저	-0.517517
일상용품	-0.502687
패션잡화	-0.462386
가전	-0.460210

구분	
감소 고객 수	3,722명(19%)
주요 이용 제휴사 (평균 이용 횟수)	A(126회)
내구재/준내구재 소비 감소	
주요 감소 품목	의류
1회 평균 구매 금액	98,567원
앱 로그인 횟수는 평균인 반면 온라인 구매 횟수는 적음	
중~고가 식품 구매 비중 65%	
A, B 경쟁사 이용률 최대 군집	

마케팅 제언

1. 경쟁사 이탈 방지를 위한 A, B 제휴사 오프라인 방문 쿠폰 발급
2. 의류 행사 진행 시 안내 문자 발송
3. 앱 로그인 유도 행사를 진행하여 로그인 유도 후 맞춤 상품 추천하여 구매 유도
4. 1회 평균 구매 금액이 큰 고객으로, 경쟁사와 차별성을 둘 수 있는 구매 금액 별 혜택 제공

cluster 1 [2번째 군집] - 온라인 상품 구매 및 외식 이용 최빈 군집

상품 분류	Weight 값
기타	0.842262
외식	0.291422
가공식품	-0.536940
전문스포츠/레저	-0.581232
신선식품	-0.474790
교육/문화용품	-0.465091
가전	-0.457283
의류	-0.406933

구분	
감소 고객 수	2,856명(15%)
주요 이용 제휴사 (평균 이용 횟수)	B(112회)
내구재/준내구재 소비 감소	
주요 감소 품목	가공식품, 전문스포츠/레저
1회 평균 구매 금액	37,189원
앱 로그인, 온라인 구매 최다 이용 군집	
저가 식품 구매 비중 62%	
전체 제휴사 평균 이용횟수가 증가 고객보다 많음	
기타 서비스 및 외식 이용률 증가	

마케팅 제언

1. 온라인 구매 금액에 따라 혜택 제공

2. 온라인으로 식품 구매 시 배송 혜택 제공

3. 구매 금액에 대한 혜택 차등 제공

4. 외식 이용률이 높아, 외식 구매 시 오프라인 상품 구매 혜택 제공

5. 저가형 상품 판매 마케팅 진행

cluster 2 [3번째 군집] – 신선식품, 가공식품 구매 이탈 고객

상품 분류	Weight 값
신선식품	-0.850616
가공식품	-0.738299
의류	-0.222581
가전	0.200059

구분	
감소 고객 수	1,705명(15%)
주요 이용 제휴사 (평균 이용 횟수)	C(165회)
주요 감소 품목	가공식품, 신선식품
1회 평균 구매 금액	20,941원
앱 로그인 횟수와 온라인 구매 횟수가 평균	
저가 식품 구매 비중 68%	
가전 구매 증가	

마케팅 제언

1. 온라인으로 식품 구매 시 배송 혜택 제공

2. 식품 타임세일 알림을 통해 구매 유도

3. 가전 프로모션 알람

4. 저가형 상품 판매 마케팅 진행

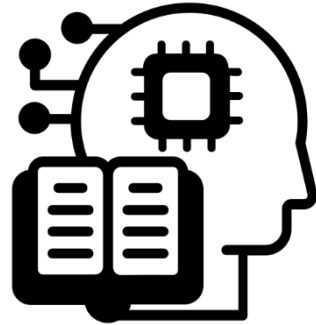
총정리

1. Business Issue finding



- 1) 일반 등급 이동 고객 = 이탈 가능성이 높은 고객
- 2) 구매 감소 고객 예측 중요

2. Modeling



- 1) 구매 감소 고객 패턴 파악에 따른 변수 설정
- 2) 구매 감소 고객 예측 모델 구축

3. Analysis



- 1) 감소 고객 분석 및 군집화
- 2) 군집에 따른 고객 유형 분석 및 구매 행동 패턴 파악

4. Solution



각 군집 별 고객 특성에 맞는 마케팅 솔루션 제언

- 1) 경쟁사 이탈 방지
- 2) 온라인 구매 증가 유도
- 3) 주요 품목 구매 증가 유도

감사합니다.

End of Document.

1팀 [임희진, 최윤진, 한아름]

21.10.18 presentation