
구매 감소 고객 예측을 통한 마케팅 전략 제시

K-digital AI solution developer based on Bigdata

2021.10.18

권준기 김광훈 진유훈 채승혜



팀장 채승혜

- 총괄
- EDA 및 데이터 분석
- Presentation

- <https://github.com/SeunghyeChae>
- lloves2743@gmail.com



팀원 권준기

- EDA 및 데이터 분석
- 데이터 전처리
- 데이터 정제 및 생성

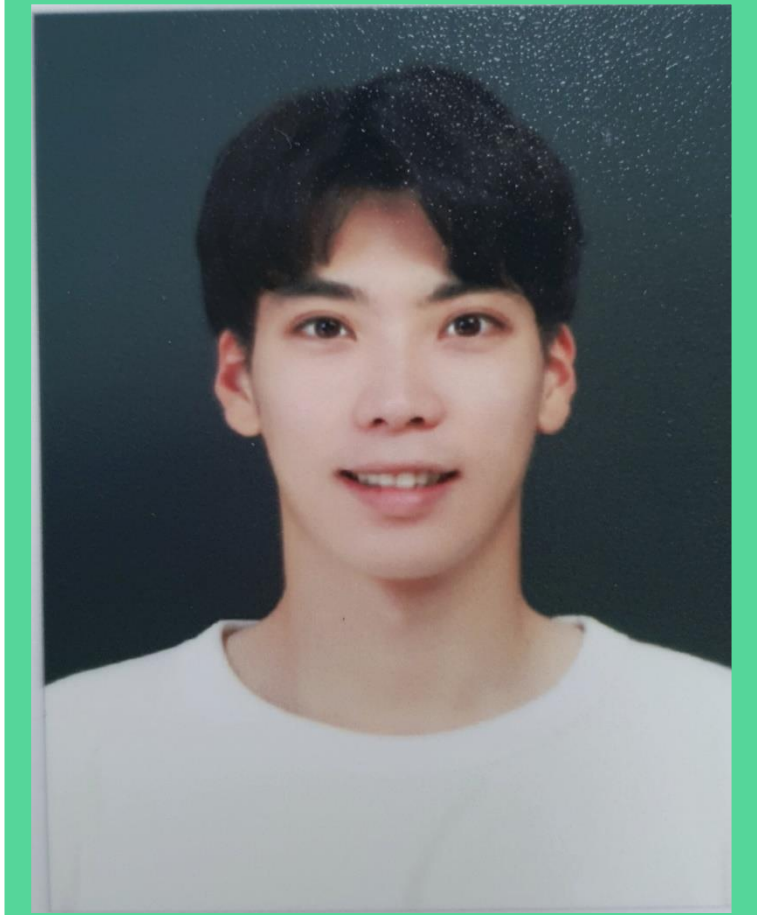
- <https://github.com/pkwon35>
- pkwon0307@naver.com



팀원 김광훈

- EDA 및 데이터 분석
- Research
- 평가지표 개발 및 성능개선

- <https://github.com/Kikiru328>
- kikiru328@gmail.com



팀원 진유훈

- EDA 및 데이터 분석
- Machine Learning
- 시각화

- <https://github.com/JINYUHOON>
- jyhoon77@nate.com

Contents





개요



목적

- L사의 데이터를 활용하여 **고객의 구매 패턴**을 파악
- 구매감소고객 예측 모델을 통해 특정 패턴을 보이는 **각 고객 유형별로 니즈를 해결**할 수 있는 **마케팅을 제안**

사용
데이터
(내부)

고객DEMO

고객번호 / 성별 / 연령대 / 거주지역

구매상품TR

고객번호 / 영수증번호/ 대, 중, 소분류코드 / 구매일자 / 구매시간/ 구매금액

멤버십여부

고객번호 / 멤버십명 / 가입년월

상품분류

제휴사 / 대, 중, 소분류코드 / 중, 소분류명

사용
데이터
(외부)

OpenAPI

한국천문연구원_천문우주정보_특일_정보제공_서비스 : 공휴일 날짜 추출

사용
Module

Google Colaboratory



Oracle SQL Developer

사용
Model

- Logistic-Regression
- SVC
- DecisionTree-Classififer
- RandomForest-Classififer
- XGB-Classififer
- LGBM-Classififer

2

EDA/주제 선정

기존고객
정의

분석의 배경
(Issue)

분석의 과제

기존고객
정의

기존고객 - 모든 구매이력 간격이 180일 이하인 고객
(각 반기에 구매이력이 1개 이상 있는 고객)

이탈고객 - 구매이력 간격이 180일 이상인 고객

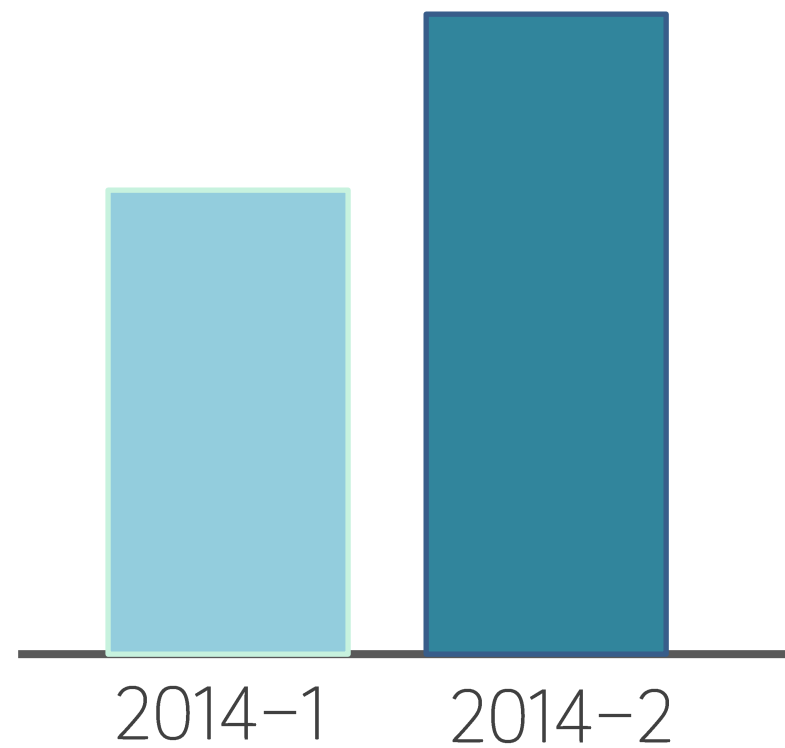
신규고객 - 다시 구매 이력이 생긴 이탈고객

기존고객 $\text{Total 고객 수}(19383\text{명}) - \text{신규/이탈고객}(298\text{명}) = 19085\text{명}$

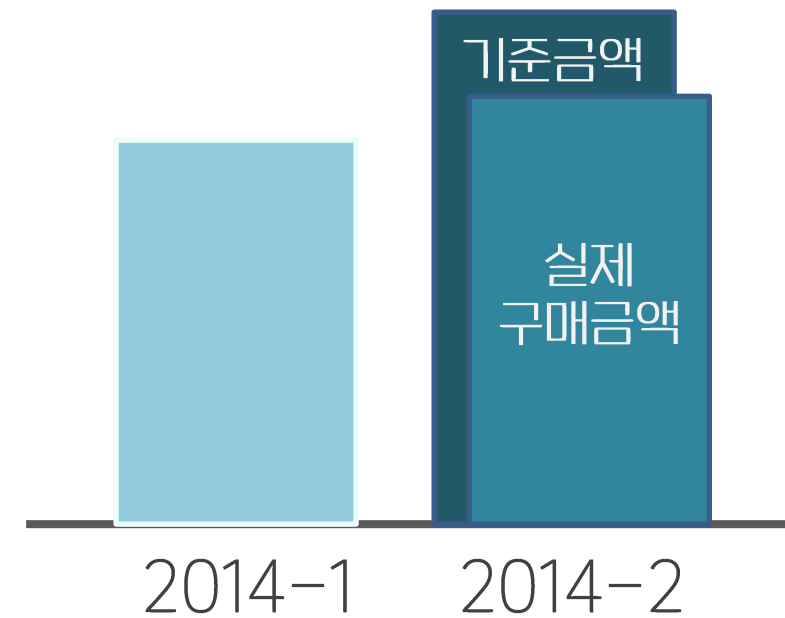
구매 감소고객
정의

각 반기의 실제 구매금액이 매출증감율을 고려한 기준금액보다 낮은 고객

매출증감율 13%



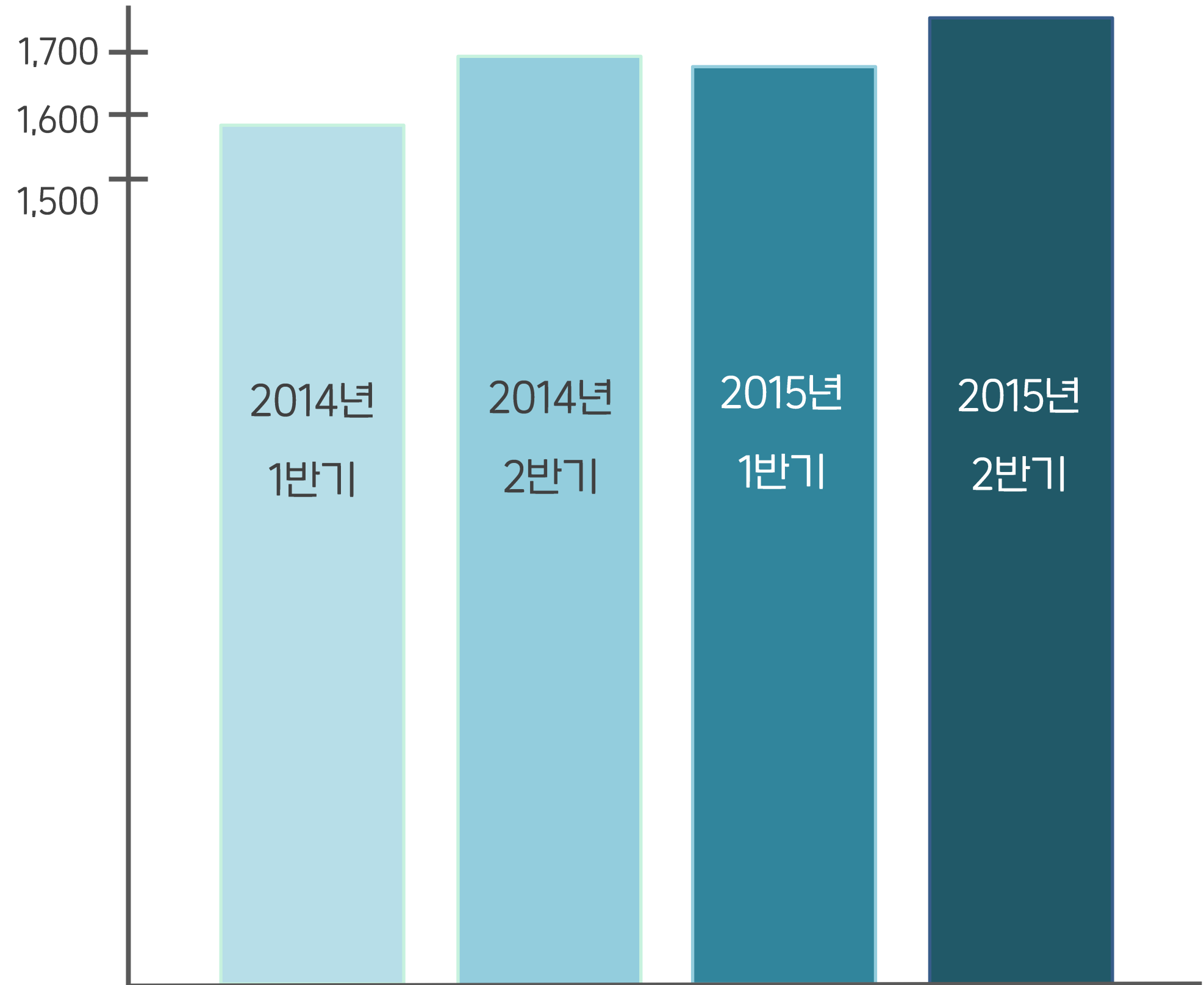
구매 감소 고객



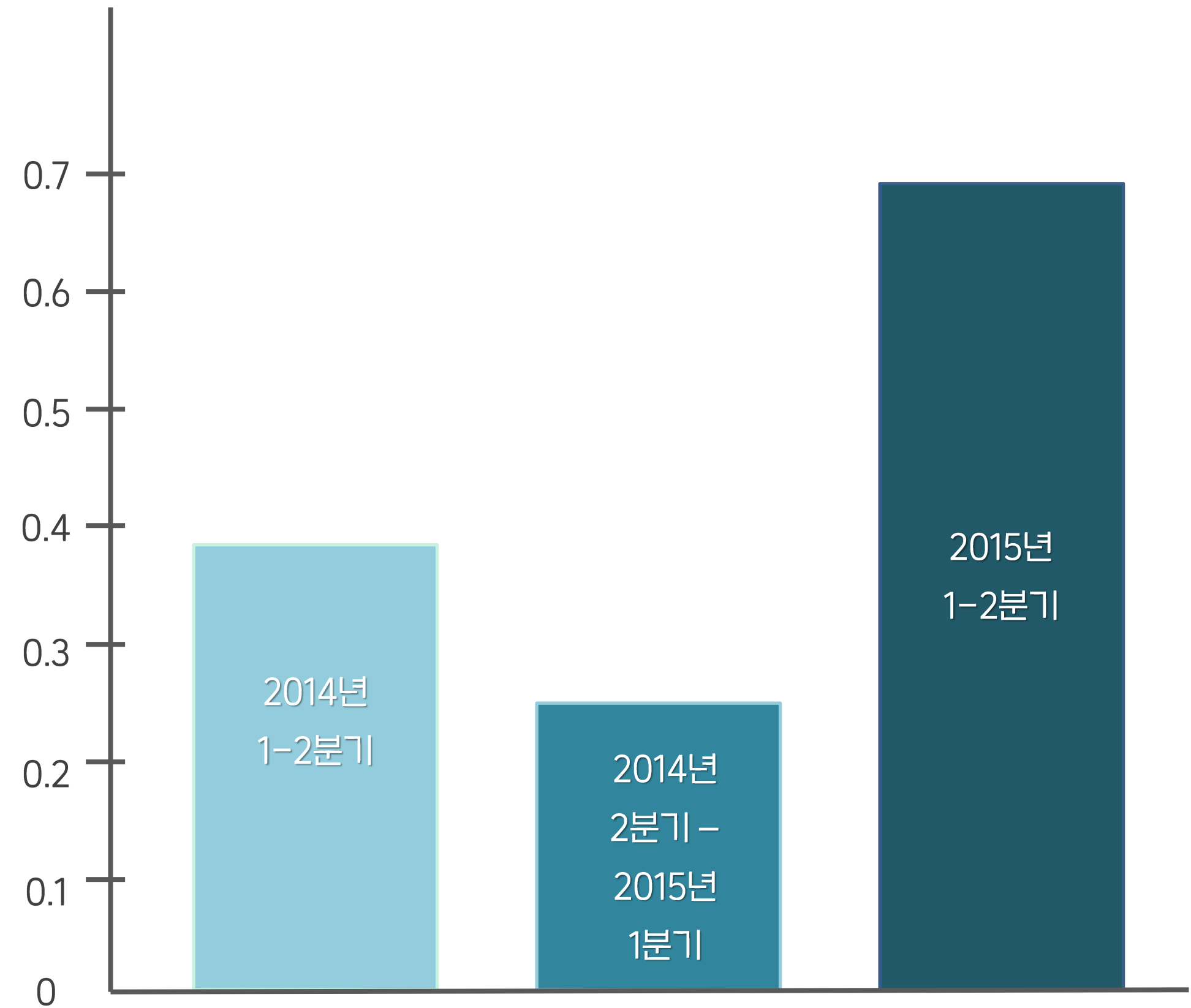
기준금액:
2014-1 구매금액 * 1.13

L사 반기별 총 매출액

매출이 증가하면 고객들의
구매감소는 줄어드는가?



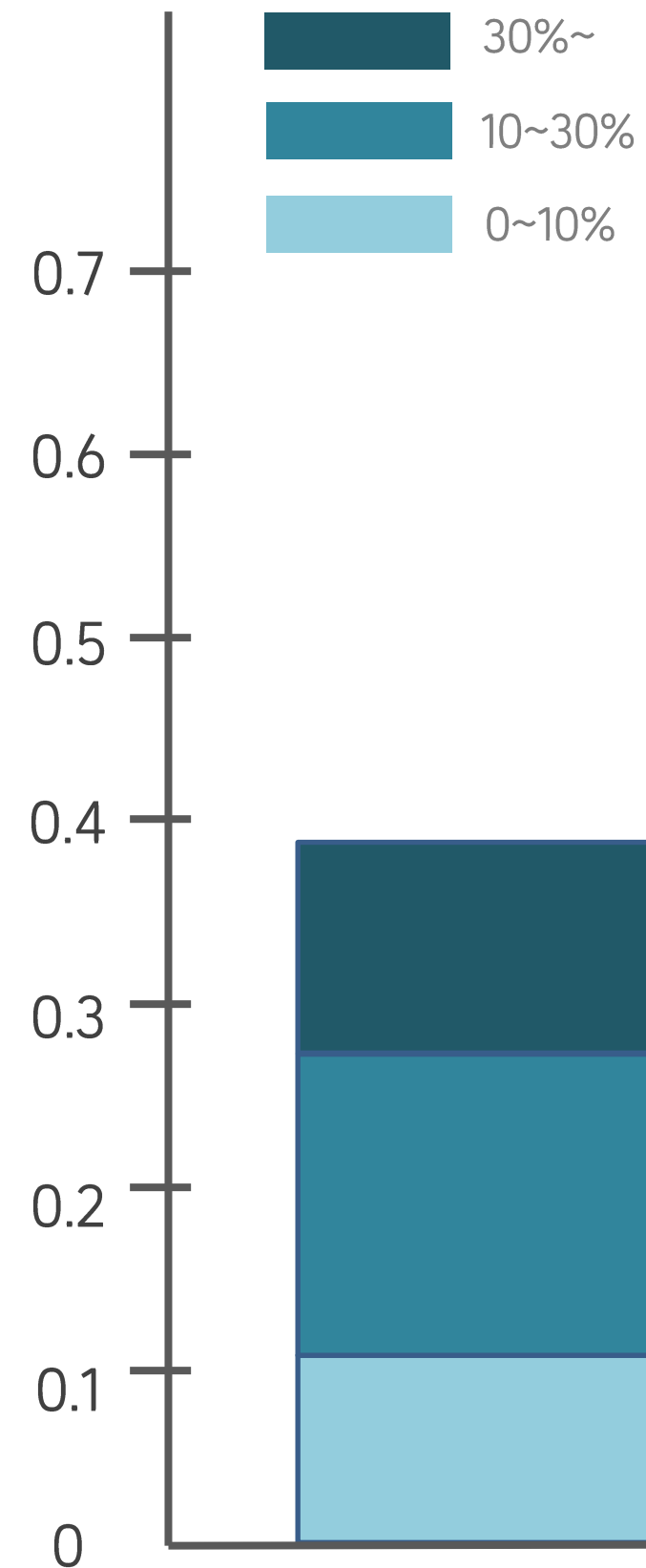
구매감소고객의 비율



구매감소고객의
구매감소액 비율

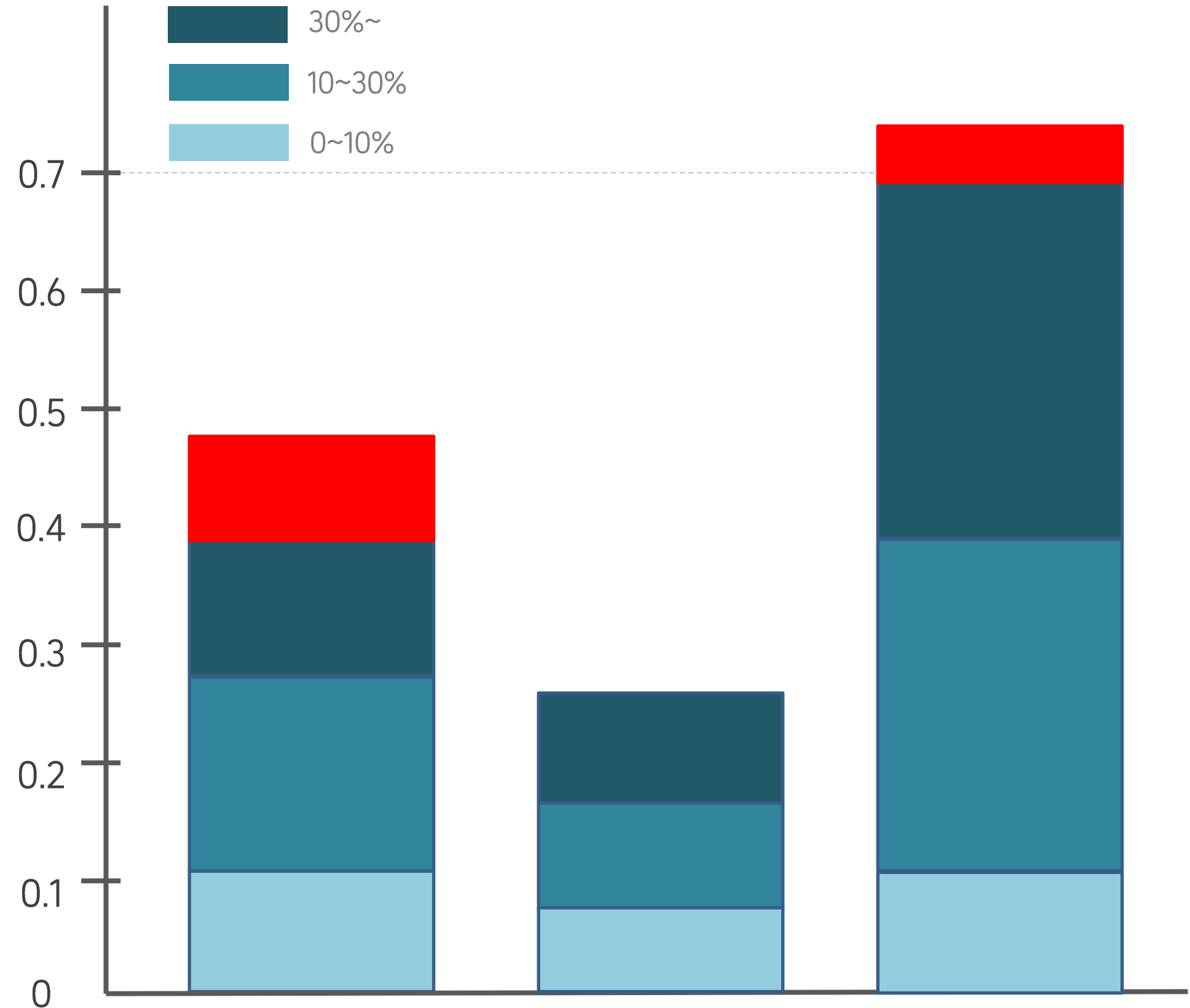
10~30% 감소한 사람의 비율
약 2배 증가

30% 이상 감소한 사람의 비율
약 3배 증가



매출증감율 고려한
구매감소고객 비율

매출증감율을 고려한
2015년 1-2분기의
구매감소 고객은
전체 고객의 **70% 이상**



분석의 과제

- 구매 감소 패턴을 보이는 **고객의 니즈**를 해결
- 다른 특성을 갖는 각 고객군들의 **패턴 파악** 및 **마케팅 제언**

주제 선정

구매 감소 고객 예측 모델 생성 및 **마케팅 제언**



데이터 분석

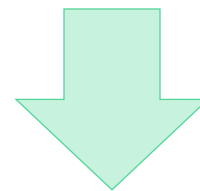


고객DEMO 고객번호 / 성별 / 연령대 / 거주지역

구매상품TR 고객번호 / 영수증번호 / 대, 중, 소분류코드 / 구매일자 / 구매시간 / 구매금액

멤버십여부 고객번호 / 멤버십명 / 가입년월

상품분류 제휴사 / 대, 중, 소분류코드 / 중, 소분류명



고객 패턴 파악을 위해
고객중심데이터로 가공 및 통합

종속변수
(target)

학습-검증 Dataset

– (2014년 1분기 대비) 2015년 1분기 구매감소유무

Test Dataset

– (2014년 1분기 대비) 2015년 2분기 구매감소유무

독립변수

학습-검증 Dataset

– 2014년 1분기 ~ 2014년 2분기 데이터 사용

Test Dataset

– 2014년 1분기 ~ 2015년 1분기 데이터 사용

성별

총 구매금액(누적)

연령대

앵겔지수 (Engel Coefficient)

거주지역 (랭크)

R / F / M 의 추세선 기울기

독립변수

방문별 거래횟수 증감/변동

객단가 증감/변동

평일 구매 비중 증감/변동

엔겔지수 증감/변동

A/B/C/D 제휴사별
구매금액 증감/변동

Recency 증감/변동

Monetary 증감/변동

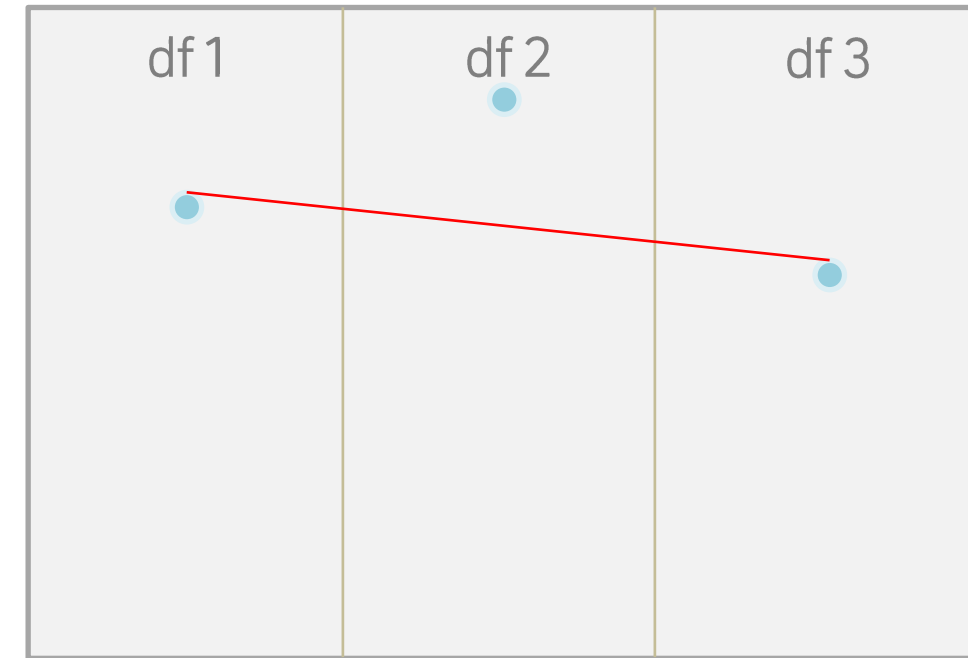
오전 구매 비중 증감/변동

패션 구매금액 증감/변동

고관여 제품 / 저관여 제품 / 사치품
구매금액 증감/변동

증감
(NET)

- 첫 반기와 마지막 반기의 증감

변동
(Coefficient of
Variation)

- 상대 표준 편차
- 표준편차 / 산술평균
- 상대적인 차이를 동일한 평균기준으로 값의 비교가 가능

고객속성

성별

- One hot encoding 을 통해 두 개의 컬럼 생성

연령대

- 연령대 범위를 넓혀서 재 범주화 (10대, 20대 ... 60대이상)

거주지역 (랭크)

- 거주지역별 고객의 평균 이용금액을 랭크화함으로써
label encoding의 단점을 보완

누적 총 구매금액

- 각 고객의 누적 총 구매금액

3

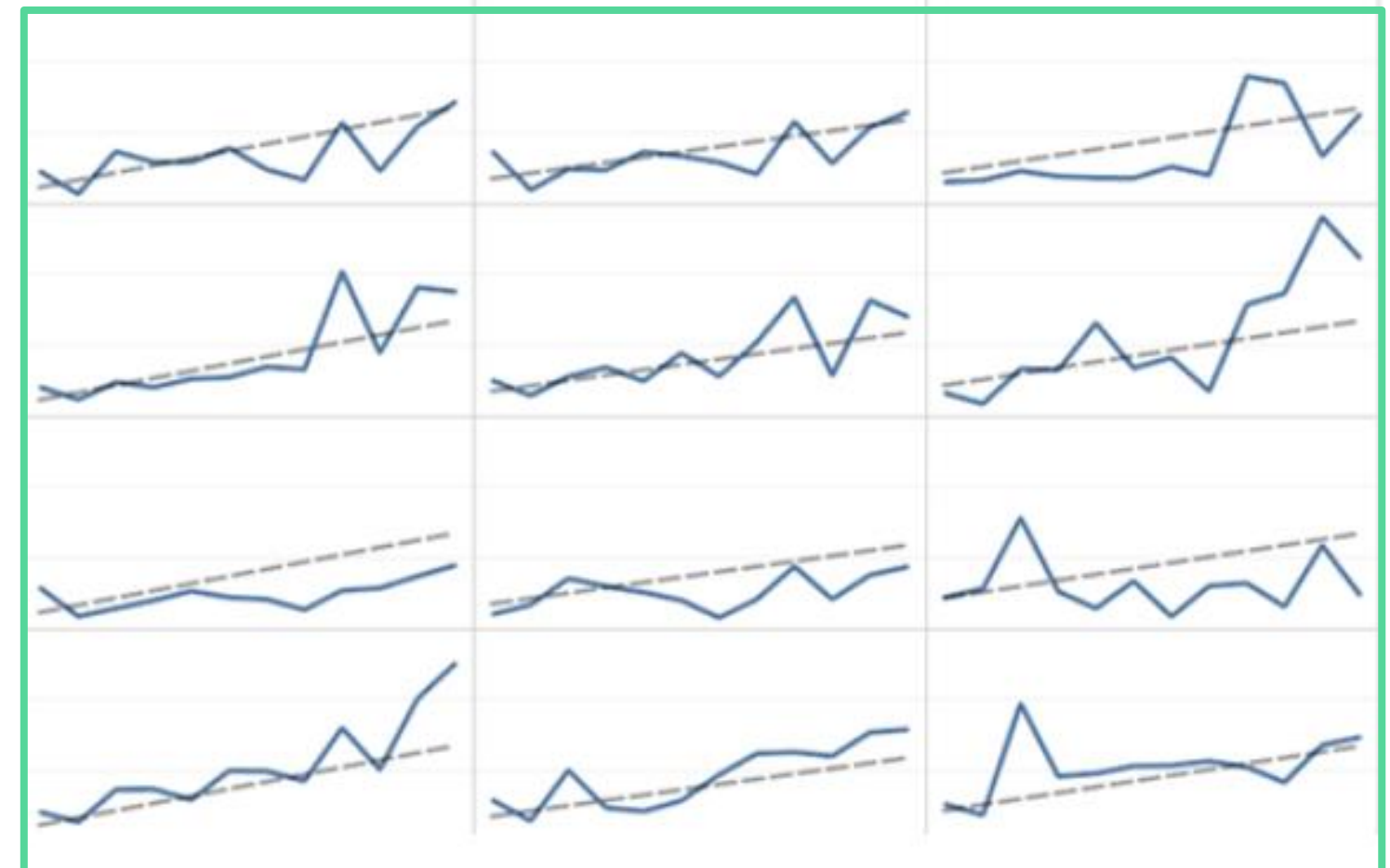
데이터 분석

고객속성

R/F/M의
추세선 기울기

-고객의 데이터(월별)를 산점도로 나타냈을 경우
데이터 산점도의 선형회귀선 기울기

$$m = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$



고객 패턴

방문별 거래횟수
변동 / 증감

- 방문할 때마다 머무는 시간을 고려하기 위해
방문별 거래횟수의 변동과 증감을 통해 패턴 측정

Recency (고객의 최근성)
변동 / 증감

- 값이 클수록 더 최근에 구매했음을 의미하는 Recency의 패턴

Monetary
변동 / 증감

- 고객이 돈을 얼마나 썼는가를 쓰는지를 의미하는 구매금액의 패턴

고객 패턴

객단가
변동 / 증감

- 한 번 구매할 때마다 얼마를 쓰는지를 의미하는 객단가의 패턴

평일 구매 횟수
변동 / 증감

- 공휴일을 제외한 평일에 구매한 횟수의 패턴
(외부데이터를 활용하여 공휴일 날짜 제외)

오전 구매 횟수
변동 / 증감

- 오전에 구매한 횟수의 패턴
(오전/오후 시간 0~12시 / 12~24시)

고객 패턴

Engel index (엔겔지수)
변동 / 증감

– 식품 구매금액 / 총 구매금액 (엔겔지수) 패턴

엔겔법칙이란,

소득의 증가에 따라 지출중 음식비 지출의 비중이
점차 감소한다는 법칙이다.

엔겔지수가 낮을수록 고소득층으로 취급하는게 보통이다.

고객 패턴

패션 구매금액
변동 / 증감

- 11개의 대분류중 패션 관련 구매금액 패턴

대분류 재범주화

- 대분류 코드를 국가기준에 맞추어 11개의 대분류명으로 재범주화
- 대분류 속성에 맞추어 각 대분류별로 중분류 재조합

가공식품/ 신선식품/ 가구,인테리어/ 교육 문화용품/ 디지털,가전
/ 의류/ 패션잡화/ 일상용품/ 전문스포츠,레저/ 명품/ 기타

3

데이터 분석

고객 패턴

제휴사별
구매금액
변동 / 증감

- A,B,C,D 제휴사별 구매금액 패턴

고관여,저관여 제품
,사치품 변동 / 증감

- 구매품목을 제품의 특성별로 분류한 구매금액 패턴 (중분류 기준)

고관여 제품

- 고객이 상품을 구매할 때 많은 고민을 거치는 성격의 상품

저관여 제품

- 습관적으로 구매하거나 크게 비교가 필요하지 않은 상품

사치품 - 명품, 보석 등

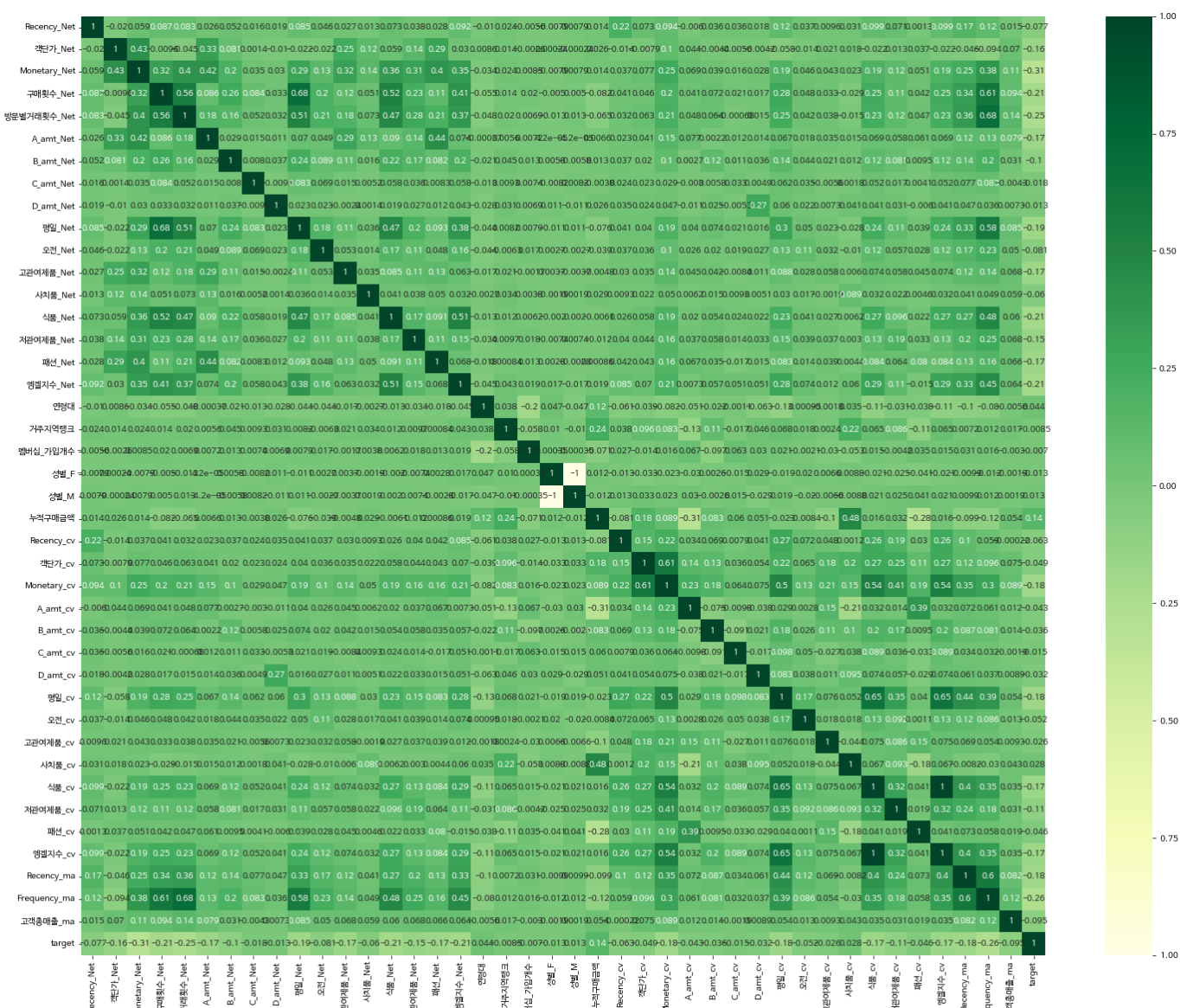
4

모델링

Model Selection

- 121개의 feature 에서 Model selection을 통해 42개의 feature로 축소

HeatMap



Modeling Accuracy

([accuracy] [precision] [recall] [f1_score] [roc_auc])

학습-검증 Dataset Accuracy

Logistic-
Regression

0.7129 / 0.7129 / 0.7219 / 0.7093 / 0.7857

Decision-
Tree

0.6306 / 0.6306 / 0.6306 / 0.6234 / 0.6306

Random
Forest

0.7123 / 0.7123 / 0.7034 / 0.7034 / 0.7875

XGB

0.7079 / 0.7079 / 0.7111 / 0.7026 / 0.7913

LGBM

0.7084 / 0.7084 / 0.7025 / 0.7093 / 0.7820

Test Dataset Accuracy

Logistic-
Regression

0.6791 / 0.6791 / 0.6826 / 0.6887 / 0.7464

Decision-
Tree

0.7120 / 0.7120 / 0.6925 / 0.7143 / 0.7128

Random
Forest

0.7255 / 0.7255 / 0.7033 / 0.7272 / 0.7886

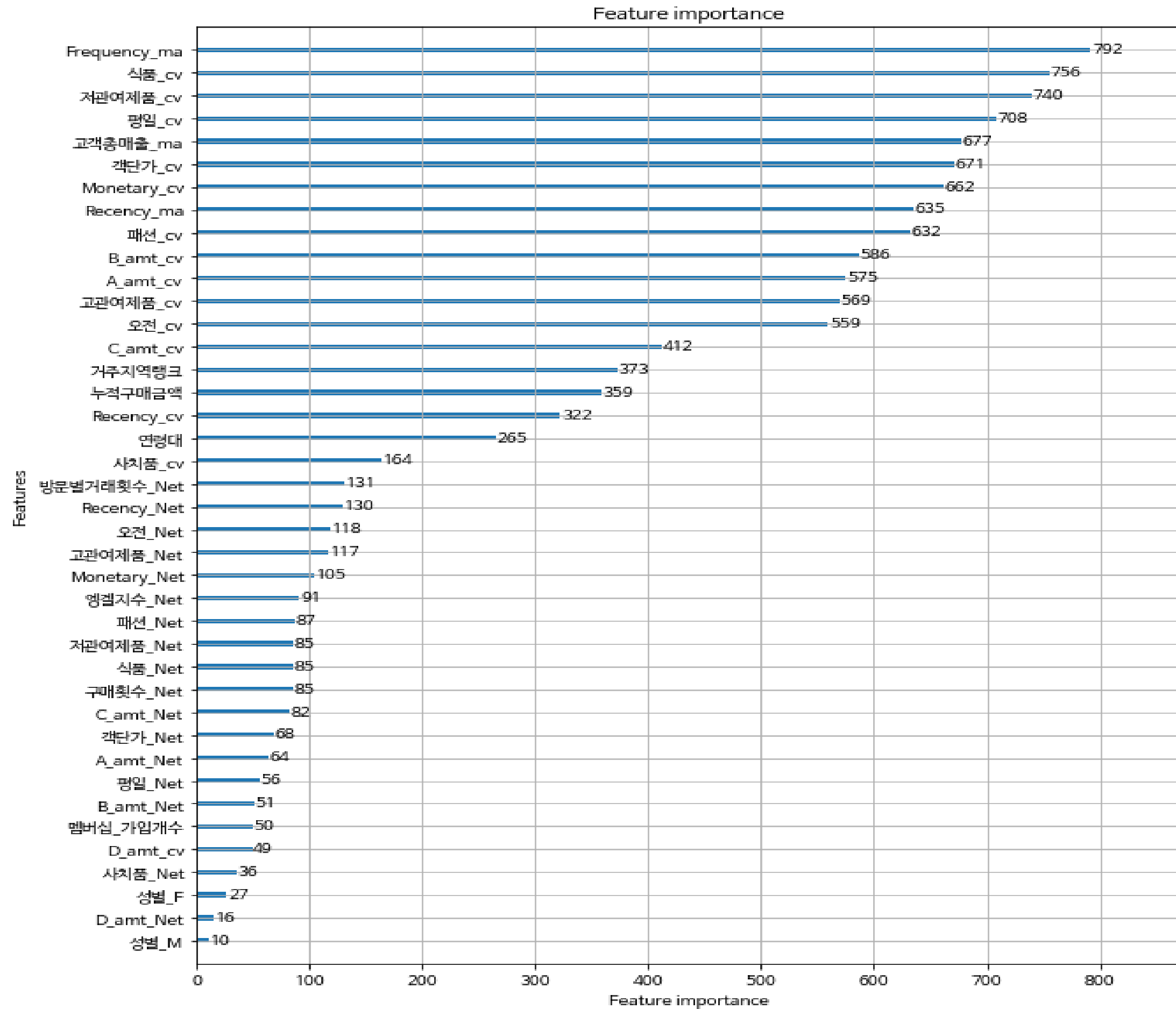
XGB

0.6997 / 0.6997 / 0.6930 / 0.7059 / 0.7752

LGBM

0.7227 / 0.7227 / 0.7035 / 0.7251 / 0.7925

Feature Importance





INSIGHT

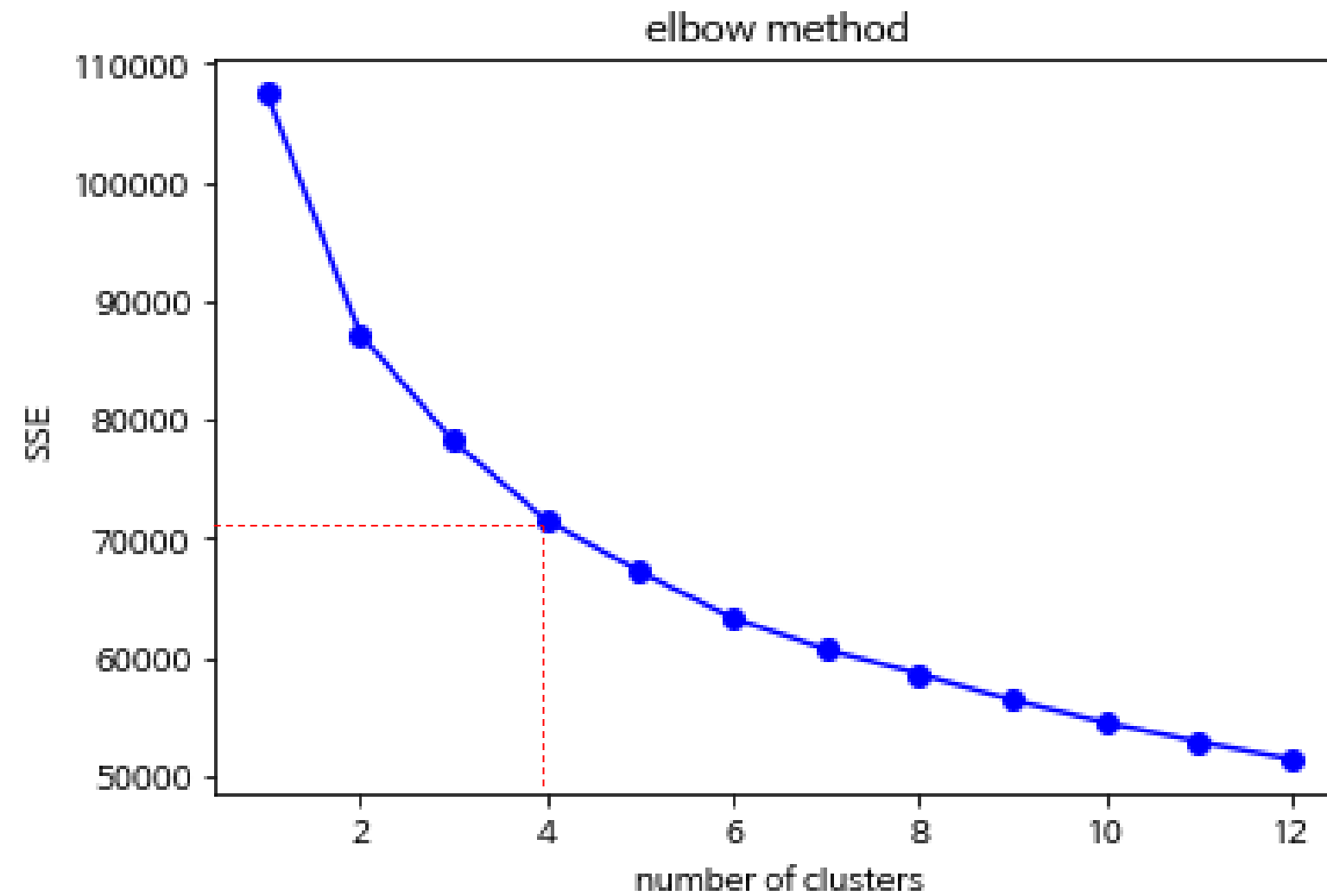
Clustering

- 중요변수 12개를 추출하여 군집화

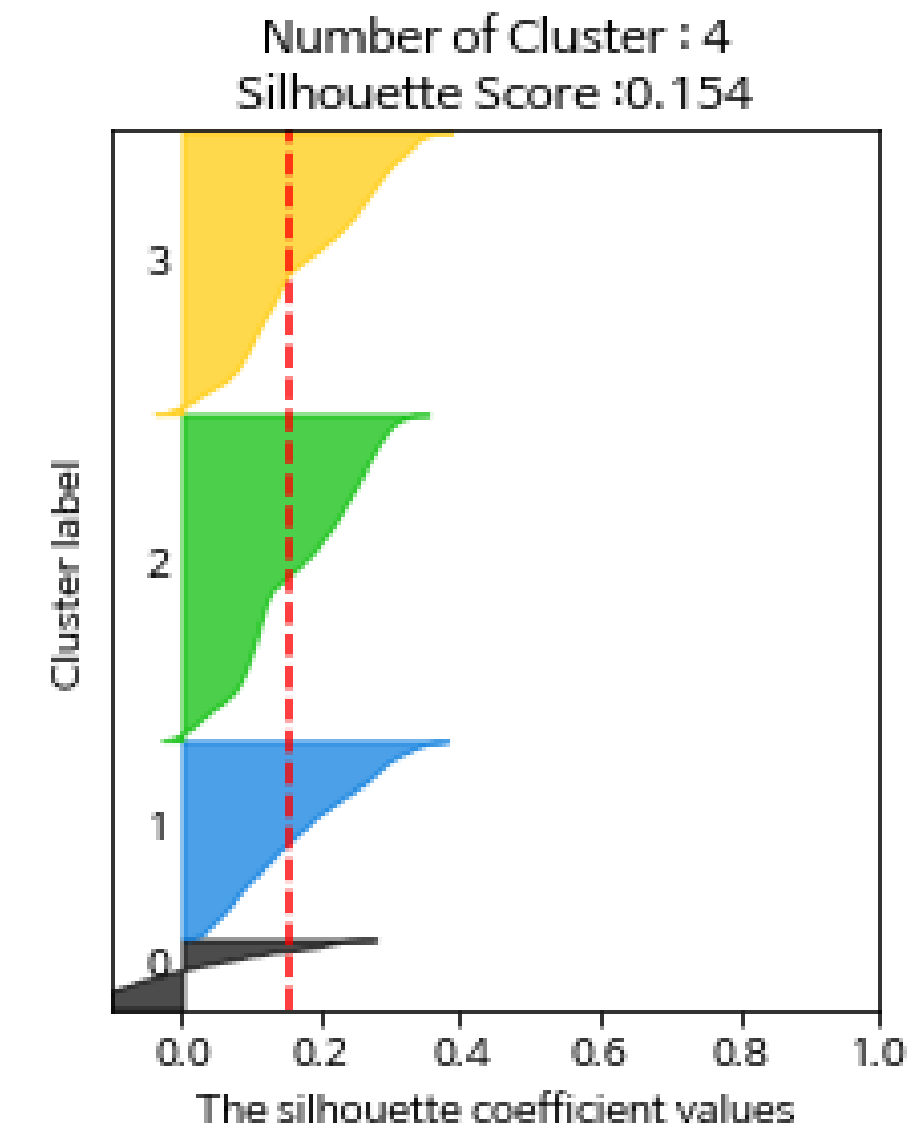
	방문별거래횟수_cv	객단가_cv	Frequency_cv	고관여제품_cv	저관여제품_cv	구매횟수_cv	연령대	패션_cv	식품_cv	사치품_cv	A_amt_cv	0	1	2	3
0	0.274695	-0.159943	-0.126403	0.874033	0.564284	0.537080	10	-0.951844	0.303052	-0.631570	-0.350468	0	1	0	0
1	0.098278	-0.764186	-0.273601	1.828496	2.223994	-0.497089	10	1.908615	-0.387719	-0.631570	1.747667	0	1	0	0
2	0.076567	-0.650952	0.584470	0.185151	1.626760	0.706902	10	-0.223993	0.906996	-0.631570	-0.766105	0	0	1	0
3	2.762970	-0.317480	2.867088	-0.656965	1.954265	2.661321	10	0.160968	1.478878	1.838729	0.108540	0	0	1	0
4	-0.541970	0.145126	-0.822011	-0.915137	-1.095517	-0.753061	10	1.586765	-0.883889	1.838729	-0.190156	0	0	1	0
...
9271	-0.070963	-0.610104	-0.388516	-0.773742	-0.616698	-0.250596	6	1.908615	-0.917900	-0.631570	1.747667	0	1	0	0
9272	0.899060	-0.602059	-0.637450	-1.296455	-0.525439	-0.241063	5	-1.223279	1.138404	-0.631570	-1.076555	1	0	0	0
9273	0.879869	-0.644054	-0.122789	1.828496	3.117756	0.368327	3	1.908615	-0.027766	-0.631570	-0.415071	1	0	0	0
9274	1.165992	-0.436630	1.197963	0.117756	-0.356103	0.174645	5	0.375669	0.267088	-0.631570	-1.076555	1	0	0	0
9275	-0.130945	-0.369737	0.212049	-1.296455	1.696743	0.486185	8	1.908615	0.430582	-0.631570	1.747667	0	1	0	0

9276 rows × 15 columns

Elbow-Method

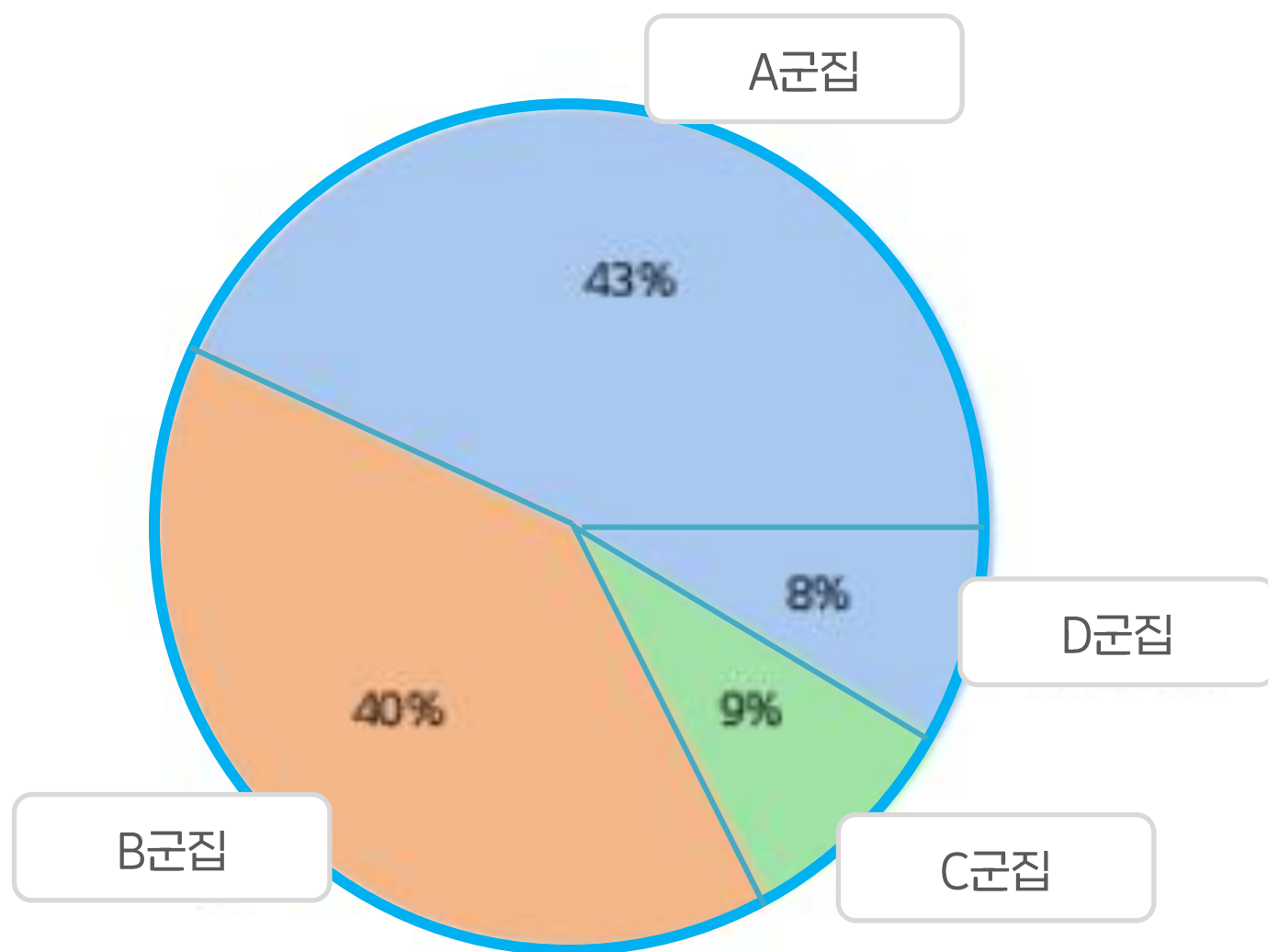


Silhouette Coefficient



Clustering 특성

구매감소고객 총 9924명





6

마케팅 제언

A 군집

특성

- 패션_NET 즉, 패션에 쓰이는 매출액이 감소된 고객들로 구성
- 그 중 사치품에 관한 구매가 감소된 고객의 비율 높음
- (패션잡화 및 의류에 대한 소비가 감소)

마케팅
제언

- 고객 맞춤 패션 의류 및 패션잡화 추천 서비스 시스템 도입

B 군집

특성

- 구매 횟수의 감소폭이 큼
- 구매 횟수가 크게 감소한 곳은 A 제휴사로 보임

마케팅
제안

- A 제휴사는 가장 높은 매출액을 보이는 제휴사로, A 제휴사에 맞는 마케팅을 진행.
- A 제휴사의 상품은 대부분 B,C,D 에는 없는 사치품과 고관여 상품이 주로 이룸.
따라서 A 제휴사의 이용과 소비를 늘리기 위해서는 A 사의 방문을 우선시하는
마케팅을 제공
- A 제휴사에서 다양한 문화행사와 이벤트 판매를 진행
(A 제휴사의 방문을 늘림과 더불어 구매횟수를 늘릴 수 있도록 함)

C 군집

특성

- 사치품에 대한 소비 감소폭이 큼

마케팅
제언

- 사치품은 고관여상품과는 달리 과시소비중 하나라고 볼 수 있음 (보석, 명품 등)
사치품에 대한 소비가 줄어들었다는 것은
개인을 과시하려는 상품이 부족 or 사치품 필요성 감소로 추론
- 사치품에 대한 마케팅을 진행.
ex. 사치품의 물량을 줄여 희소성을 높이거나, 가장 최신의 사치품을 보유하여
과시소비의 심리를 자극

D 군집

특성

- 식품의 소비 감소폭이 큼

마케팅 제언

- 식품: 가격의 변동이 크지 않고, 수요의 탄력성이 크지 않은 상품군
엔겔지수의 설명과 같이 식품은 소비가 항상 일어나는 상품
- 식품의 소비가 줄었다는 것은 해당 제휴사들 안에서 식품을 소비하지 않고,
경쟁사에서 소비하고 있다고 추론 가능
- 식품의 소비를 올리기 위해서 매번 다른 할인 상품 이벤트or
경쟁사와의 차별성 만들기

THANK YOU

감사합니다

2021.10.18

K-digital AI solution developer based on Bigdata

권준기 김광훈 진유훈 채승혜
