구매 감소 고객 예측을 통한 마케팅 전략 제시

K-digital AI solution developer based on Bigdata

2021.10.18

권준기 김광훈 진유훈 채승혜

0 팀원소개

팀장 **채승혜**

- 총괄
- EDA 및 데이터 분석
- Presentation

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

팀원 **권준기**

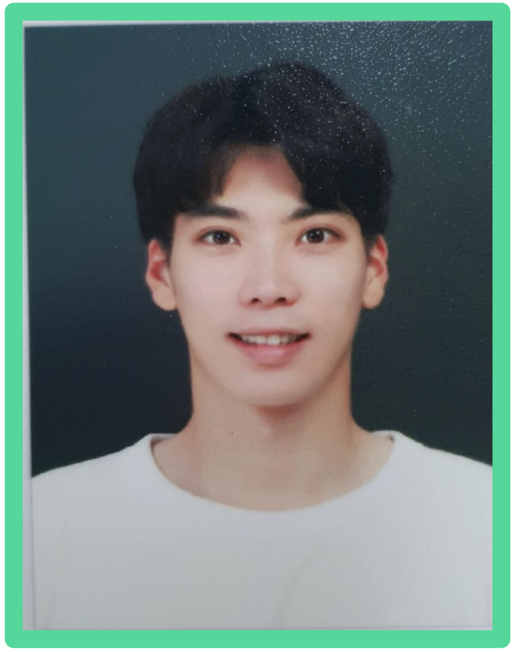
- EDA 및 데이터 분석
- 데이터 전처리
- 데이터 정제 및 생성

- 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

1 개요

1 개요

목적

- L사의 데이터를 활용하여 고객의 구매 패턴을 파악
- 구매감소고객 예측 모델을 통해 특정 패턴을 보이는 각 고객 유형별로 니즈를 해결할 수 있는 마케팅을 제언

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

1 개요

사용 데이터 (외부)

OpenAPI : 한국천문연구원_천문우주정보_특일_정보제공_서비스 : 공휴일 날짜 추출

작업환경

Google Colaboratory

Oracle SQL Developer

사용 Module

matplotlib Version 3.2.1

scikit learn

Pandas

seaborn

사용 Model

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

2 EDA/주제선정

2 EDA 및 주제선정

2 EDA 및 주제선정

기존고객 정의

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

이탈고객 – 구매이력 간격이 180일 이상이 된 고객

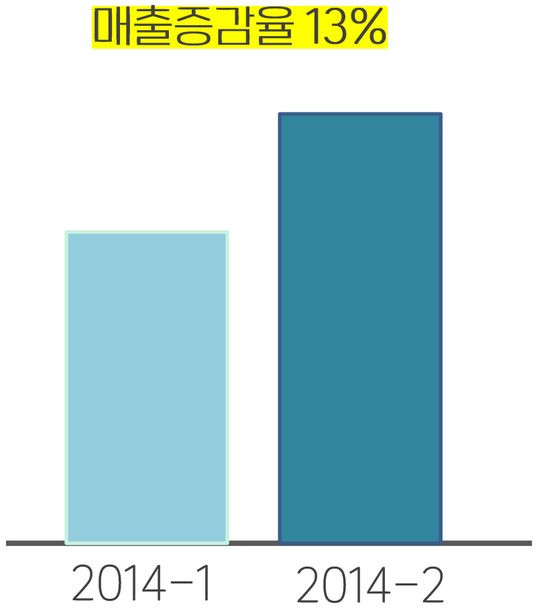
신규고객 – 다시 구매 이력이 생긴 이탈고객

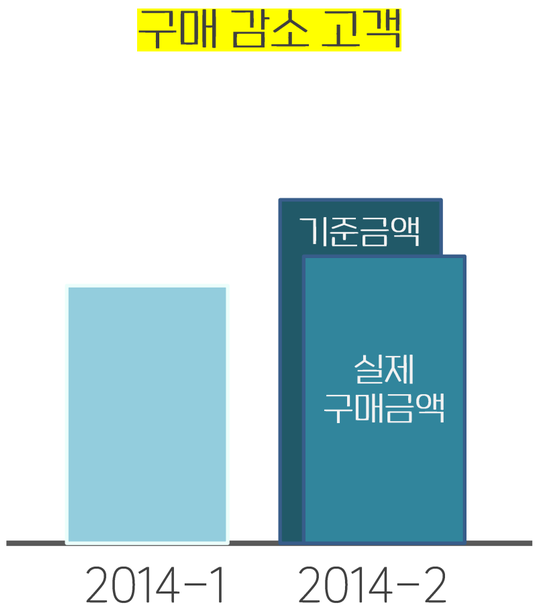
기존고객 Total 고객 수(19383명) – 신규/이탈고객 (298명) = 19085명

2 EDA 및 주제선정

구매 감소고객 정의

각 반기의 실제 구매금액이 매출증감율을 고려한 기준금액보다 낮은 고객



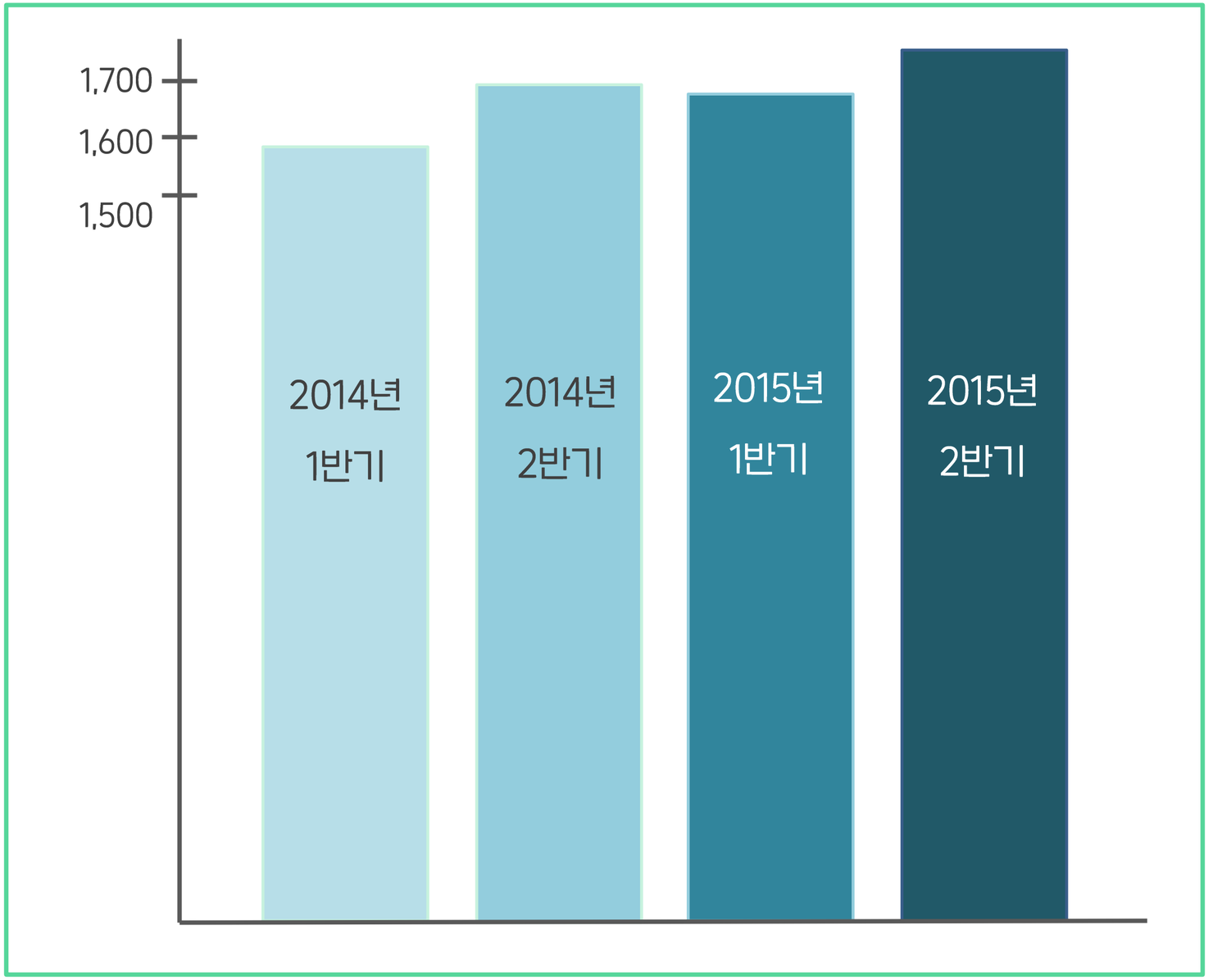


기준금액:
2014-1 구매금액 * 1.13

2 EDA 및 주제선정

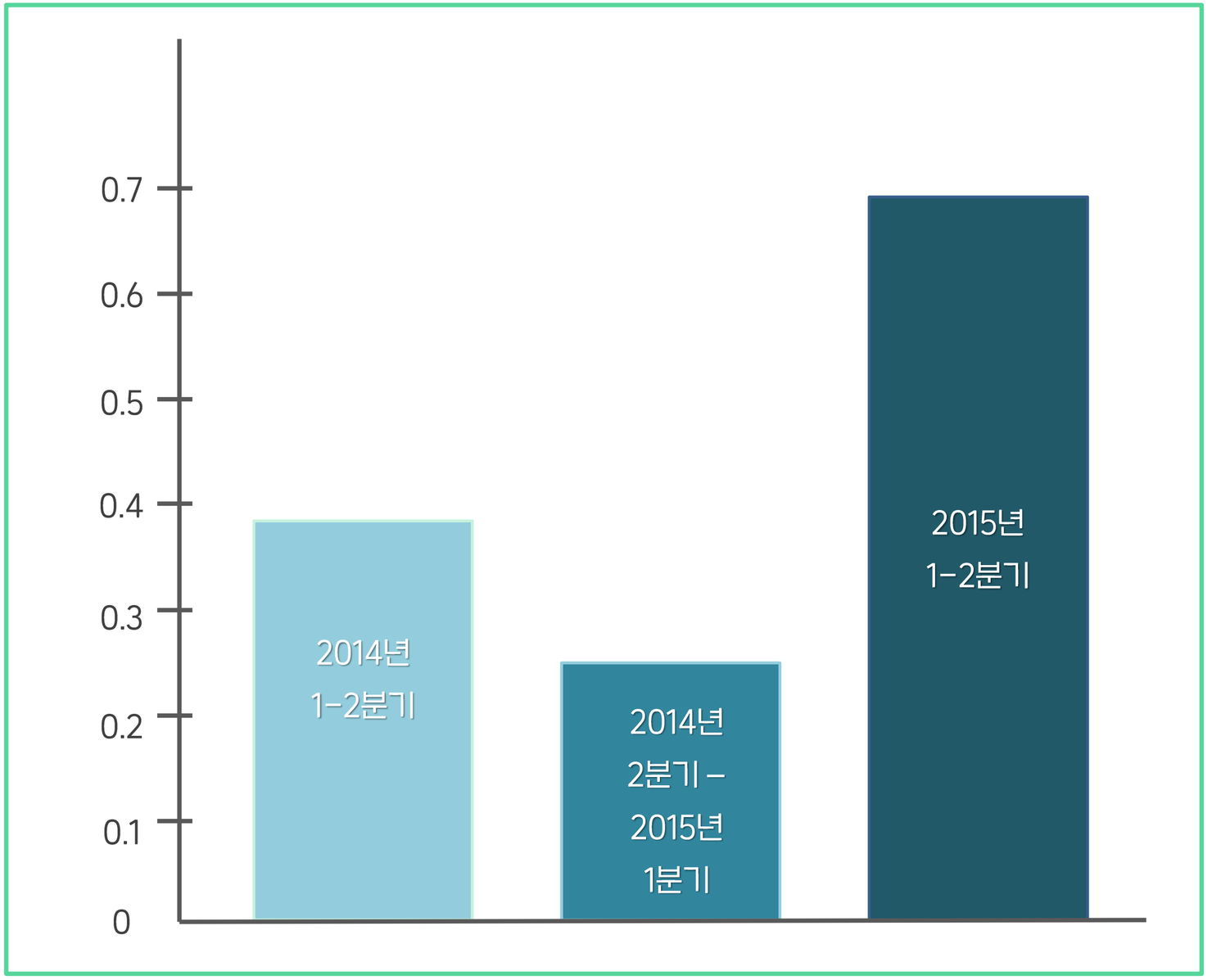
L사 반기별 총 매출액

매출이 증가하면 고객들의
구매감소는 줄어드는가?



2 EDA 및 주제선정

구매감소고객의 비율

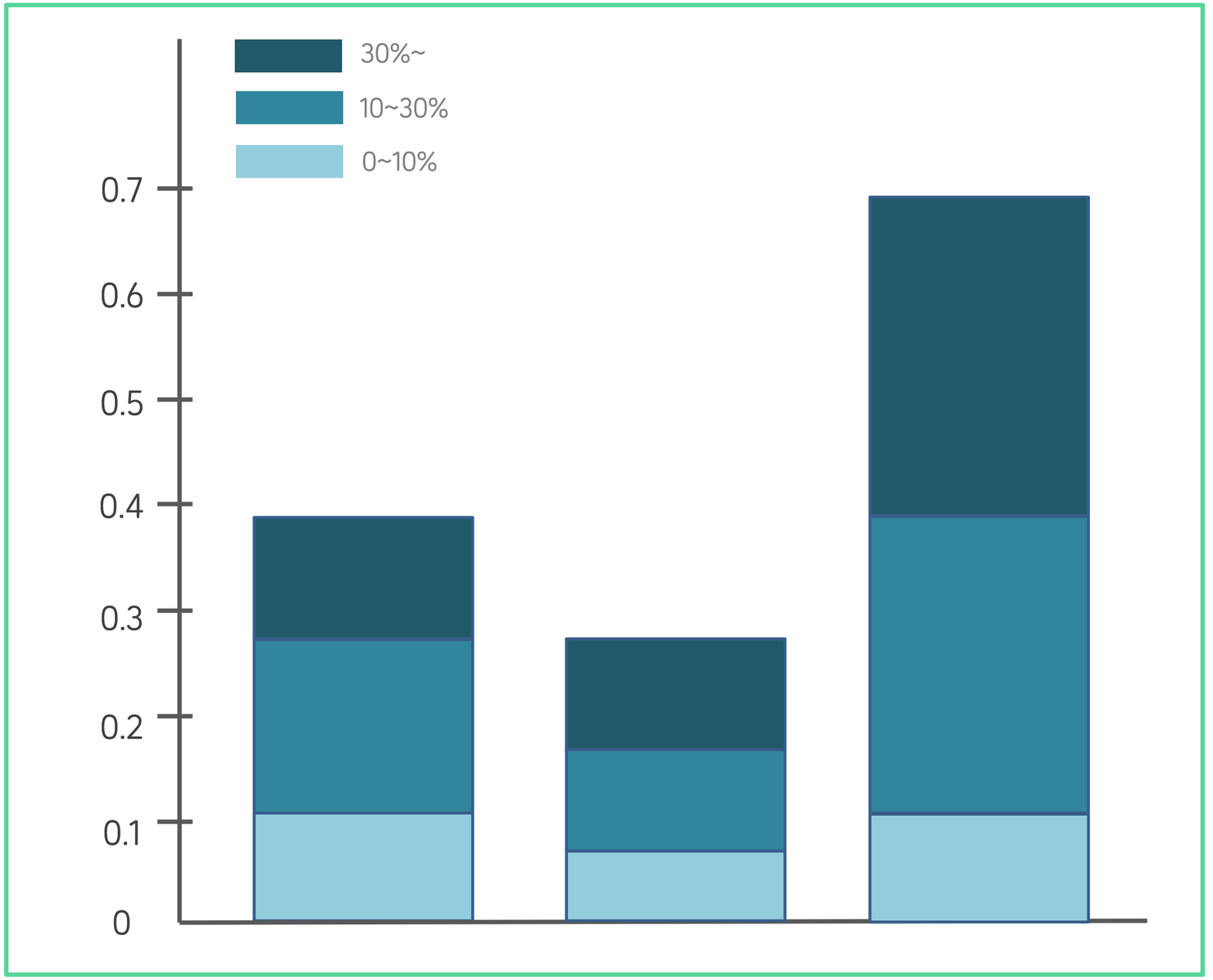


2 EDA 및 주제선정

구매감소고객의
구매감소액 비율

10~30% 감소한 사람의 비율
약 2배 증가

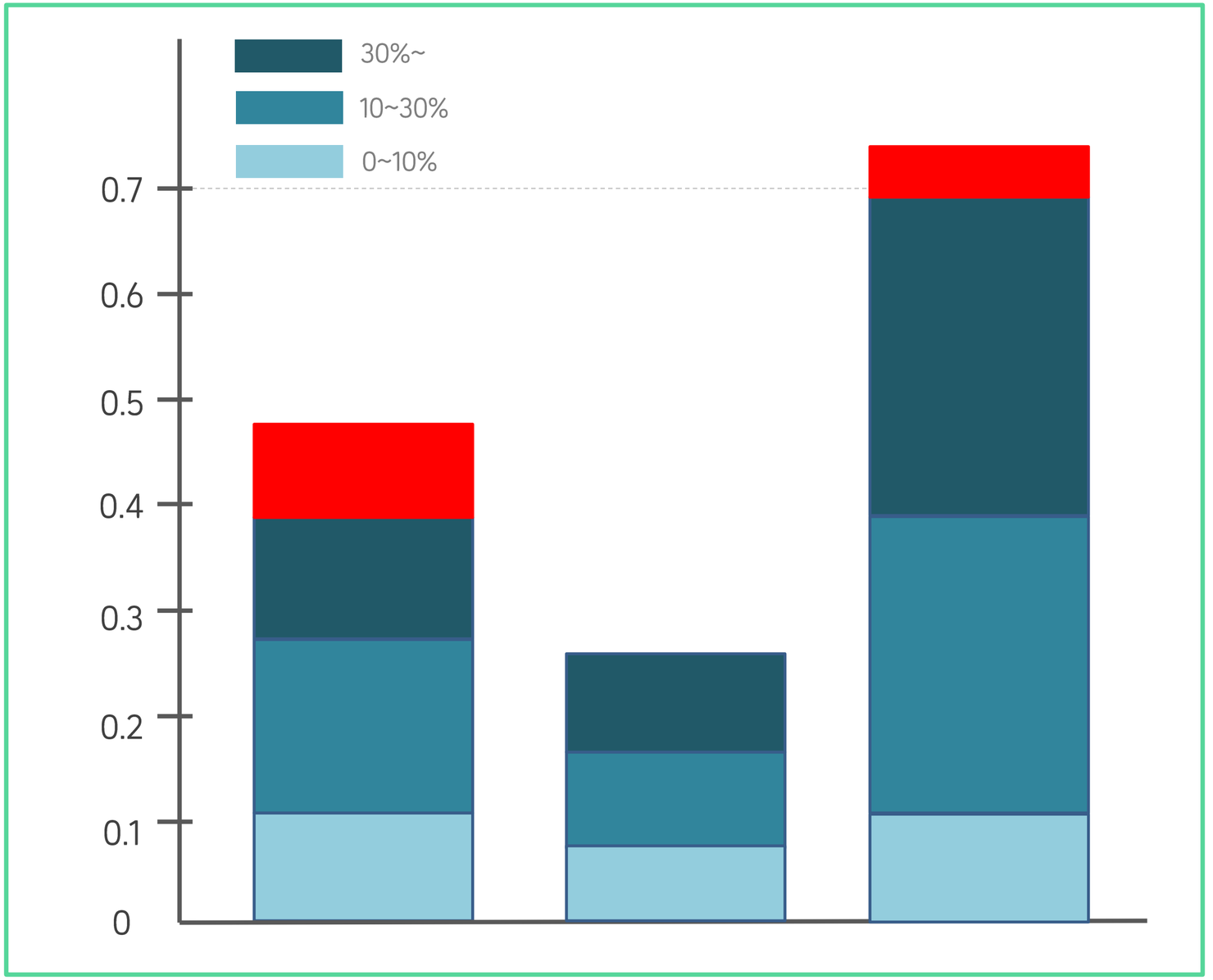
30% 이상 감소한 사람의 비율
약 3배 증가



2 EDA 및 주제선정

매출증감율 고려한
구매감소고객 비율

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



2 EDA 및 주제선정

분석의 과제

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

주제 선정

구매 감소 고객 예측 모델 생성 및 마케팅 제언

3 데이터 분석

3 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

고객중심 데이터로 가공 및 통합

3 데이터 분석

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

- 성별
- 연령대
- 거주지역 (랭크)

Test Dataset
– 2014년 1분기 ~ 2015년 1분기 데이터 사용

- 총 구매금액 (누적)
- 앵겔지수 (Engel Coefficient)
- R / F / M 의 추세선 기울기

3 데이터 분석

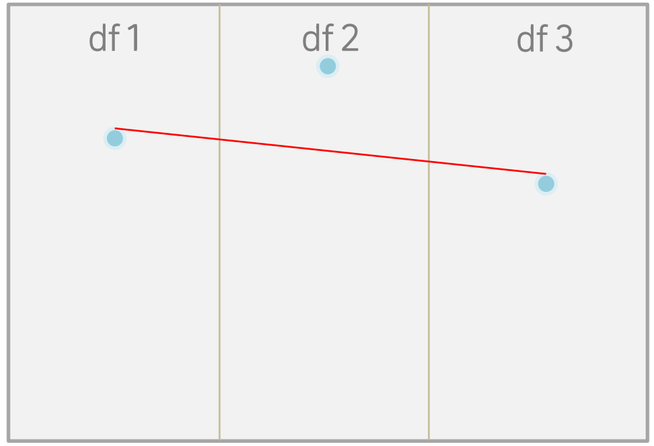
독립변수

- 방문별 거래횟수 증감/변동
- Recency 증감/변동
- 객단가 증감/변동
- Monetary 증감/변동
- 평일 구매 비중 증감/변동
- 오전 구매 비중 증감/변동
- 엥갤지수 증감/변동
- 패션 구매금액 증감/변동
- A/B/C/D 제휴사별 구매금액 증감/변동
- 고관여 제품 / 저관여 제품 / 사치품 구매금액 증감/변동

3 데이터 분석

증감 (NET)

- 첫 반기와 마지막 반기의 증감



변동 (Coefficient of Variation)

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

3 데이터 분석

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

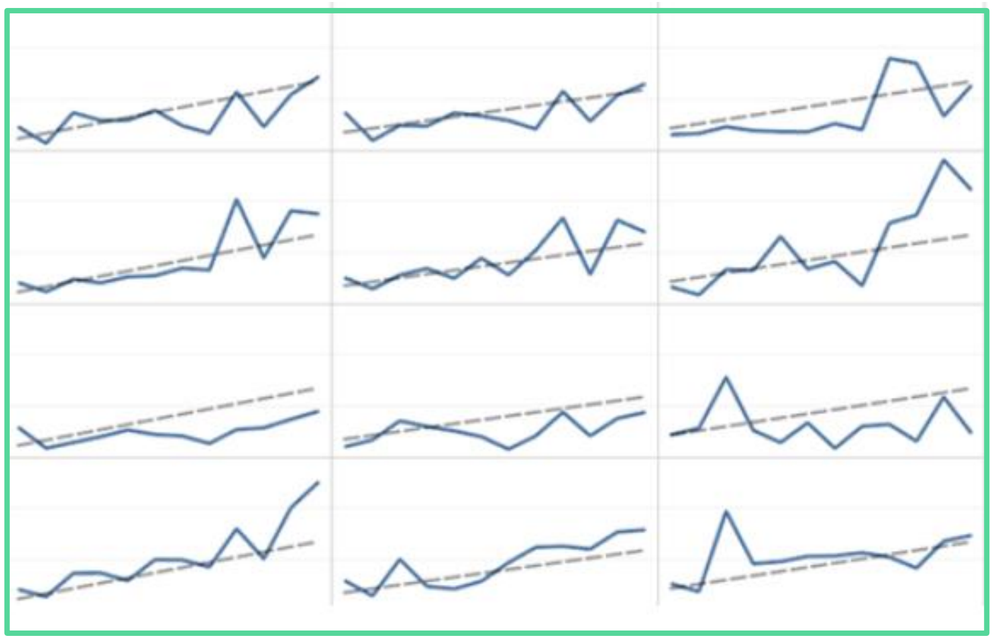
3 데이터 분석

고객속성

R/F/M의
추세선 기울기

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

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



3 데이터 분석

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

3 데이터 분석

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)

3 데이터 분석

고객 패턴

Engel index (엥겔지수)
변동 / 증감

– 식품 구매금액 / 총 구매금액 (앵겔지수) 패턴

엥겔법칙이란,

소득의 증가에 따라 지출중 음식비 지출의 비중이

점차 감소한다는 법칙이다.

엥겔지수가 낮을수록 고소득층으로 취급하는게 보통이다.

3 데이터 분석

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

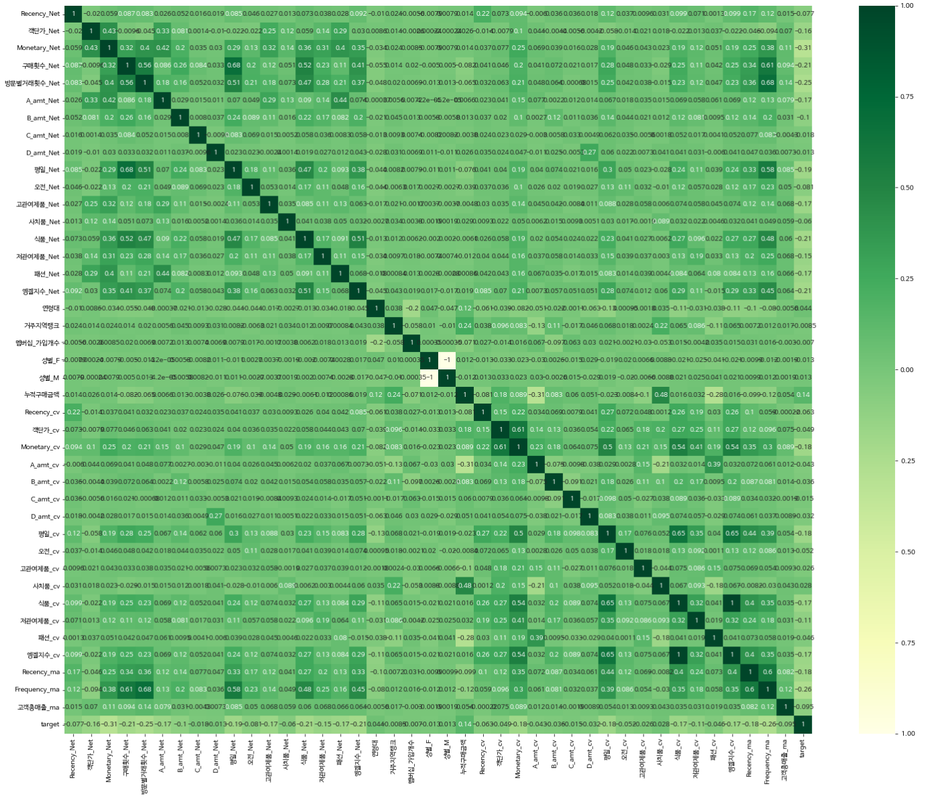
4 모델링

4 모델링

Model Selection

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

HeatMap



4 모델링

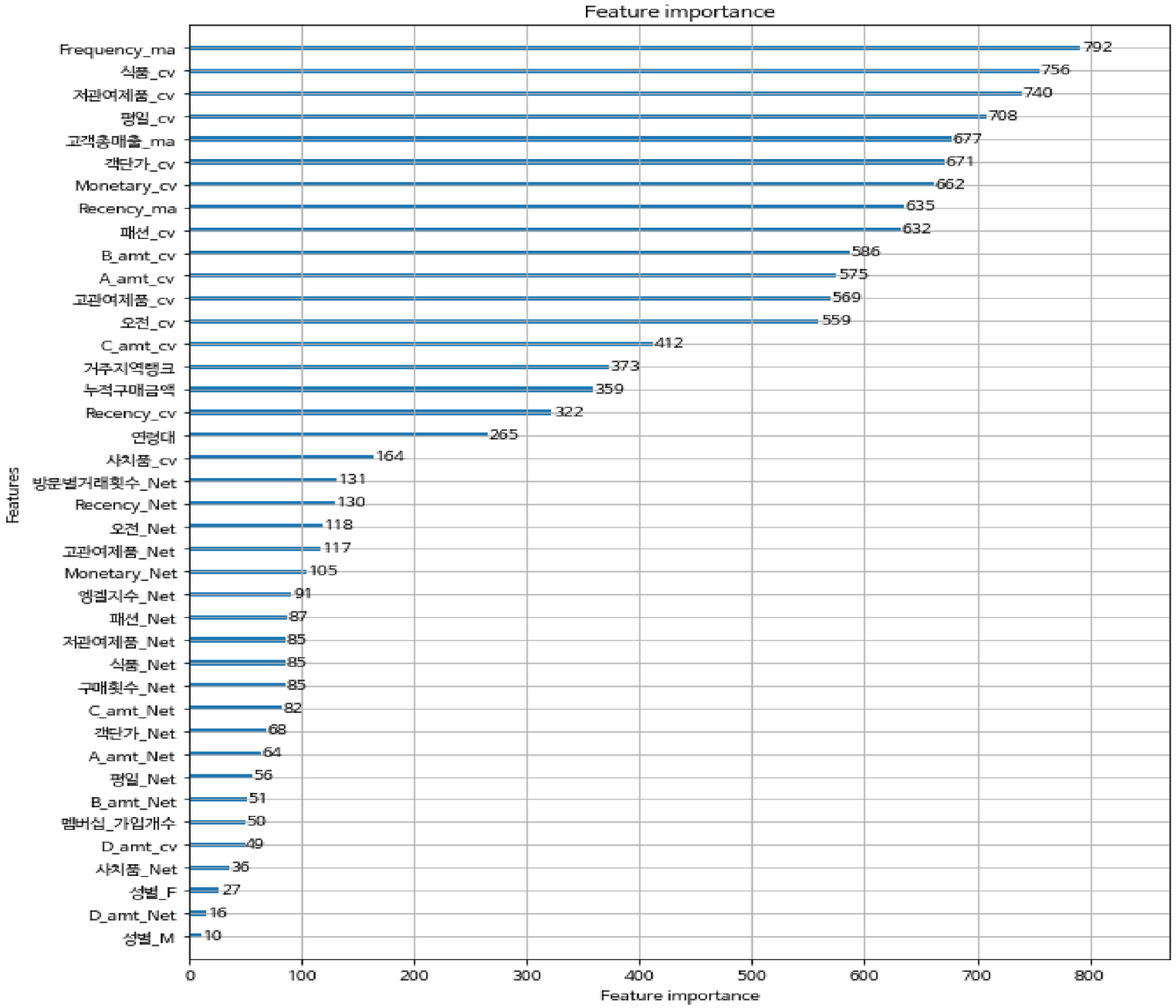
Modeling Accuracy ([accuracy] [precision] [recall] [f1_score] [roc_auc])

	학습-검증 Dataset Accuracy	Test Dataset Accuracy
Logistic-Regression	0.7129 / 0.7129 / 0.7219 / 0.7093 / 0.7857	0.6791 / 0.6791 / 0.6826 / 0.6887 / 0.7464
Decision-Tree	0.6306 / 0.6306 / 0.6306 / 0.6234 / 0.6306	0.7120 / 0.7120 / 0.6925 / 0.7143 / 0.7128
Random Forest	0.7123 / 07123 / 0.7034 / 0.7034 / 0.7875	0.7255 / 0.7255 / 0.7033 / 0.7272 / 0.7886
XGB	0.7079 / 0.7079 / 0.7111 / 0.7026 / 0.7913	0.6997 / 0.6997 / 0.6930 / 0.7059 / 0.7752
LGBM	0.7084 / 0.7084 / 0.7025 / 0.7093 / 0.7820	0.7227 / 0.7227 / 0.7035 / 0.7251 / 0.7925

4

모델링

Feature Importance



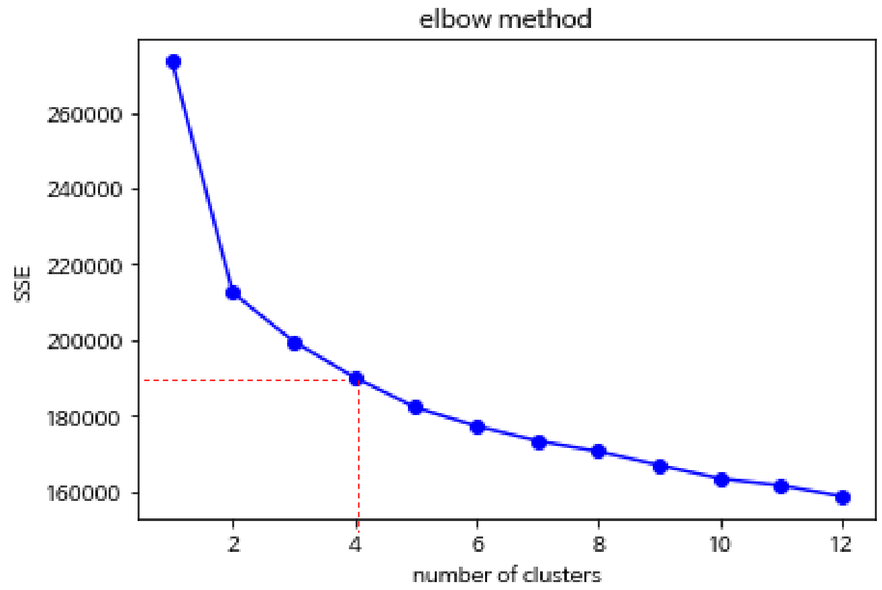
5 INSIGHT

5 INSIGHT

Clustering

- 중요변수 22개를 추출하여 군집화

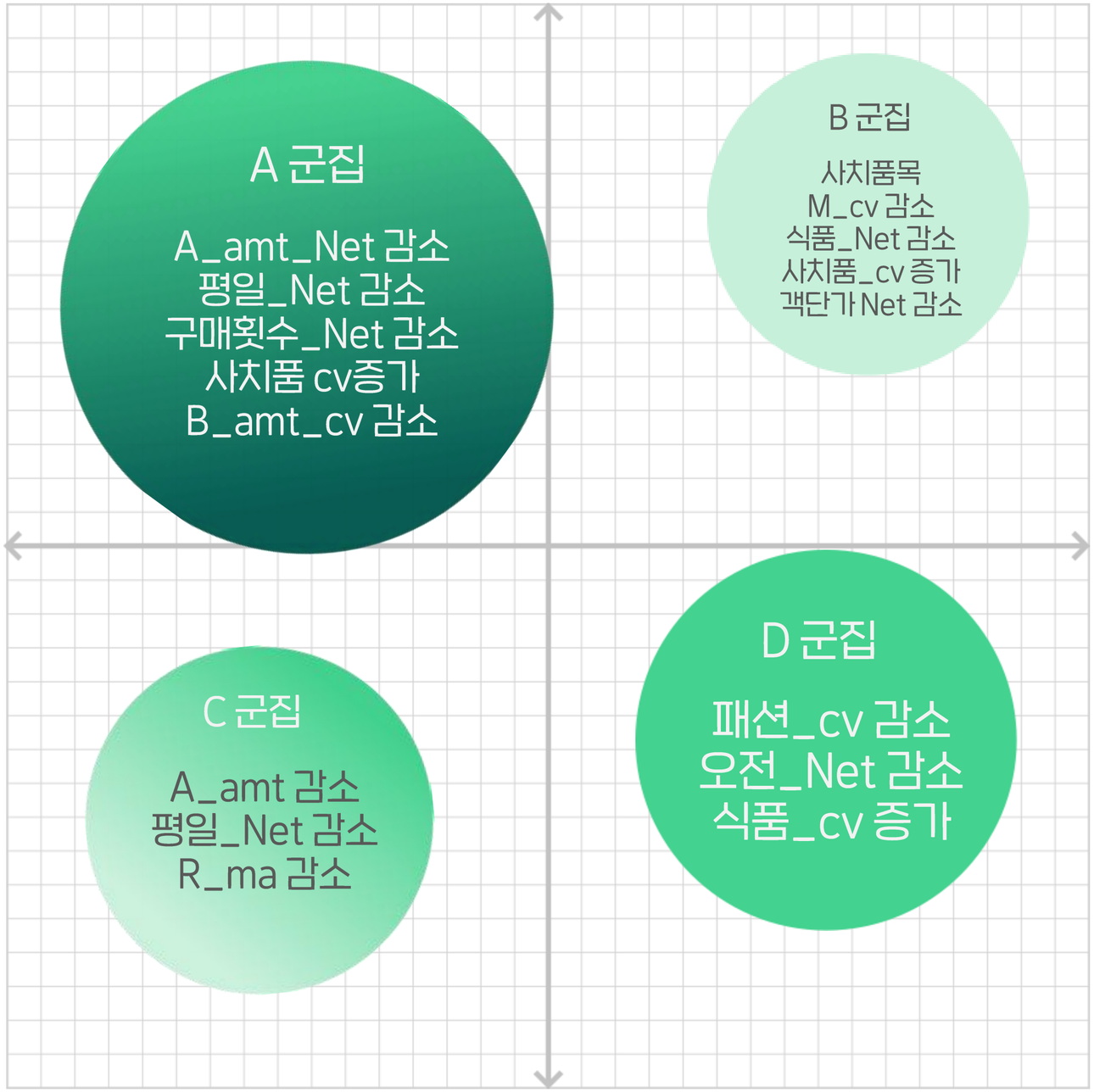
Elbow-Method



5 INSIGHT

Clustering 특성

구매감소고객 총 9924명



6

마케팅 제언

6 마케팅 제언

A 군집

특성

- A 제휴사의 이용횟수 감소
- 평일 방문 , 구매 횟수 감소
- 사치품 구매 변동률 증가

마케팅 제언

- A 제휴사에서 평일에도 실시하는 문화행사
- A 제휴사에서 할인율 폭이 큰 상품을 대거 활용해 평일에만 판매하도록 함

6 마케팅 제언

B 군집

특성

- 식품의 매출액 감소
- 사치품 변동률 증가
- A 제휴사 매출액 감소
- 객단가 감소

마케팅 제언

- PB상품 개발
- 외국 식재료 or 외국 식품 판매를 목적으로 마케팅

6 마케팅 제언

C 군집

특성

- A 제휴사 매출 변동 감소
- 평일 방문 횟수 감소
- 방문 추세 감소

마케팅 제언

- 주말에만 열리는 행사를 평일까지 연장
- 평일 방문객들이 많이 구매하는 상품 이벤트
- 최근 트렌드를 반영한 광고 및 상품 판매

6 마케팅 제언

D 군집

특성

- 오전 구매 횟수 감소
- 식품 매출 변동 증가

마케팅 제언

- 신선식품 / 계절식품 오전 선착순 마케팅
- 지속적인 쿠폰 제공
- 고객관리를 통해 해당 계절, 기온, 트렌드 등을 고려한 맞춤 식품 서비스 진행

THANK YOU

감사합니다

K-digital AI solution developer based on Bigdata

2021.10.18

권준기 김광훈 진유훈 채승혜