구매감소 고객 
분석 및 맞춤형 솔루션

2023.06

Data Universe

팀원소개

▣ Data Universe

박주경 (총괄, 데이터분석, 피처엔지니어링, 모델링, 클러스터링)

김예슬 (데이터 분석, 모델링, 시각화, ppt)

남동연 (데이터 분석, 피처엔지니어링, 모델링, 클러스터링)

오윤택 (데이터 분석, 데이터 전처리, 피처엔지니어링, 모델링)

조차선 (데이터 분석, 피처엔지니어링, 모델링, 클러스터링)

진행일정

1주차 : 분석 주제 선정 및 계획 수립

2주차 : 데이터 분석 및 기획안 발표

3주차 : Feature 개발 및 마스터 데이터 세팅

4주차 : 최적 모델 탐구 및 Hyper parmeter 조정

5주차 : 적정 군집화 수집 및 고객 세그멘테이션

6주차 : 고객 상품 추천 및 최종발표

※ 세그멘테이션 : 시장을 공통적인 수요와 구매행동을 가진 층으로 나누어서 그 층의 욕구와 필요에 맞추어서 제품을 디자인하여 제공하는 것

Contents

개요 I

목적

목적

1. L사의 제휴사 백화점, 마트, 슈퍼, 드럭스토어 고객의 거래 데이터 분석
2. 구매 감소가 예측되는 고객 분류 후 군집화
3. 각 그룹에 알맞는 마케팅 제언 → 매출 감소 예방, 매출 증대

외부요인

옴니 채널 (Omni Channel)

온.오프라인 모든 쇼핑 채널이 유기적으로 연결돼
고객이 하나의 매장을 이용하는 것처럼 느끼도록 모든 채널을 융합한 것
→ 즉, 소비자들이 시간 · 장소 구애없이 제품을 구매할 수 있게 하는 쇼핑 시스템

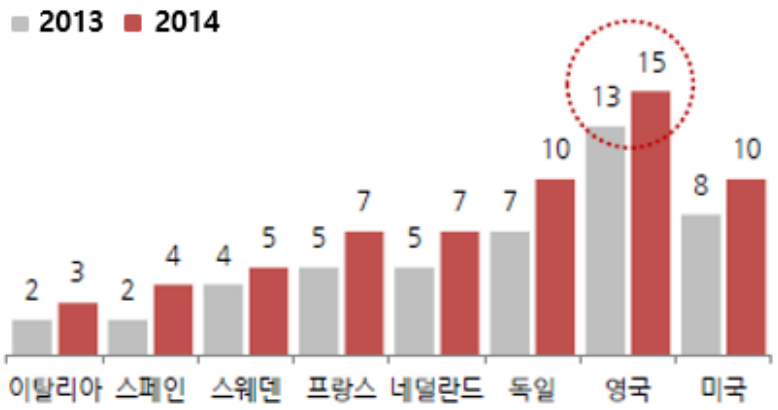
❖ 옴니 채널의 필요성

- ✓ 모바일 기기의 보편화
- ✓ 빅데이터, 모바일 결제 등의 기술 발전
- ✓ 밀레니얼 세대 부상
- ✓ 소비 패턴의 비정형화
- ✓ 스마트 컨슈머의 출현

※ 밀레니얼 : 1980 ~ 2000년대 초반까지의 출생자
※ 스마트 컨슈머 : 여러가지 종합적인 정보를 바탕으로 합리적인 선택을 하는 소비자

외부요인

<전체 소매 매출 중 온라인 매출(서유럽 &미국)>



주요 국가 오프라인
유통 업체들의 파산

대응 필요
(옴니채널)

<영국 &미국 오프라인 유통 업체 파산 사례>

업체명	국가	설명	파산일
Circuit City	미국	미국 2위 전자제품 판매 전문점	09.01
Borders	미국	미국 대형 2위 서점 체인	11.03
Comet	영국	영국 최대 전자제품 판매 전문점	12.12
Jessops	영국	광학제품(카메라, 망원경 등) 판매 전문점	13.01
HMV	영국	음반, 영화, 게임 소매 유통체인	13.01
Blockbuster	영국	비디오 대여 체인	13.01

고객 중심의
구매패턴 파악 및 클러스터링

→ 맞춤형 마케팅 제시 전략

활용 데이터

사용 데이터

고객 DEMO	고객번호, 성별, 연령대, 거주지역
구매상품 TR	고객번호, 영수증번호, 대.중.소분류코드, 구매일자, 구매시간, 구매금액
멤버십 여부	고객번호, 멤버십명, 가입년월
상품 분류	제휴사, 대.중.소분류코드, 중.소분류명

데이터 전처리 – 연령대, 지역, 시간

연령대

- 19세 이하
- 20~24세
- 25~29세
- 30~34세
- 35~39세
- 40~44세
- 45~49세
- 50~54세
- 55~59세
- 60세 이상

지역

서울	1-99
경기	100
인천	210
강원	240
충북	270
세종	300
충남	310
대전	340
경북	360
대구	410
울산	440
부산	460
경남	500
전북	540
전남	570
광주	610
제주	630

※ 신우편번호 앞 3자리

시간

- 6~10시 아침(MOR)
- 11~13시 점심(LUN)
- 14~17시 오후(AFT)
- 18~21시 저녁(EVE)
- 22~5시 심야(MID)

데이터 전처리 – 상품

- **필요성**
  - 4개의 제휴사의 상품분류 코드가 상이
  - 2014 → 2015년 상품 카테고리의 이동, 통합 등의 이슈

카테고리

식품관
잡화, 화장품
아동
의류
과일
과자,음료
케어 용품
스포츠
주방용품
유제품
⋮
⋮

A,B,C,D
제휴사
14개의
분류로 통합

카테고리

01 가공식품
02 신선식품
03 외식/편의시설
04 일상용품
05 의약품/의료기기
06 교육/문화용품
07 뷰티(화장품,미용)
08 디지털/가전
09 가구/인테리어
10 의류
12 전문스포츠/레저
12 패션잡화
13 유아/아동
14 명품

※ 유통상품지식뱅크 기준

데이터 전처리 – 계절성 제거

*단위 : 십만원(14,15년 고객별 구매금액)

고객 번호	14년도 1분기	14년도 2분기	… …	15년도 3분기	15년도 4분기	14,15년 총 구매금액
15999	4,608	2,923	… …	3,868	8,465	35,718
6207	6,293	3,952	… …	1,116	1,588	17,210
⋮	⋮	⋮	… …	⋮	⋮	⋮
총 구매액의 비율	24%	24%	… …	23%	29%	
모두 25%의 비율로 맞춰 줌 (EX> 1분기 – (0.25-0.24) * (14,15년 총 구매금액/4)						

총 구매금액에서 분기별 비율을 모두 맞춰
계절성 제거

※ 계절성 : 특정한 주기나 계절에 따라 데이터나 현상에 반복적인 패턴이 나타나는 것

작업 환경

개발 환경

Jupyter


Colab


Github

분석 언어


Python

SQL

라이브러리 & 프레임워크

EDA를 통한 고객 정의

EDA – 기준 고객 정의

2014 ~ 2015년 8개 분기 기준

구매가 연속적인 고객

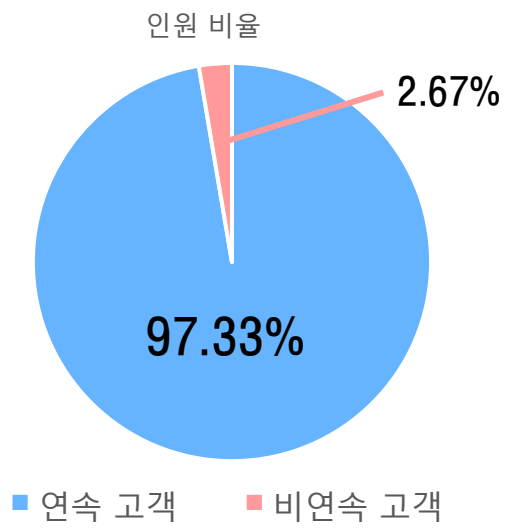
구매가 비연속적인 고객

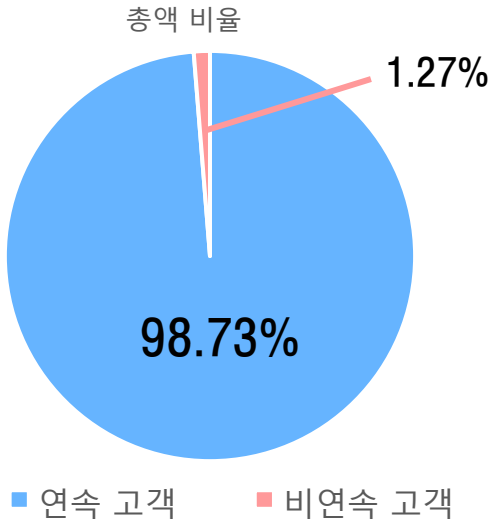
고객 리텐션 전략에 따라 연속 고객을 기존 고객으로 정의

※ 리텐션 마케팅 : 기존 고객의 이탈률을 최소화하는 동시에, 리 텐션율(고객 유지율)을 높이는 마케팅 방식

EDA - 기준 고객 정의

연속 고객과 비연속 고객의 인원,총액 비율





*단위: 명

고객	인원 수
연속	18,866
비연속	517

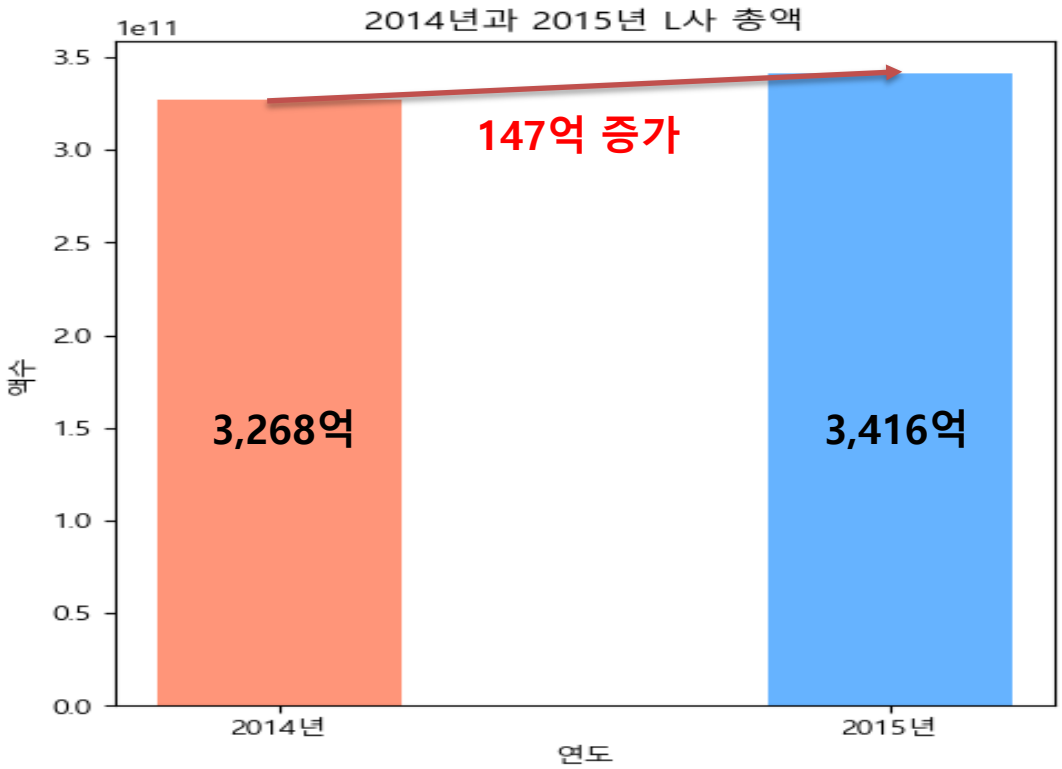
*단위: 천만원

고객	총액
연속	66,843
비연속	858

비연속 고객 비중이 2.67%(인원), 1.27%(총액)로 전체 데이터에 대한 영향도가 적음

∴ 기존고객 데이터를 활용한 분석 진행

EDA - 감소 고객 정의



2014년 총액 3,268억, 2015년 3,416억으로

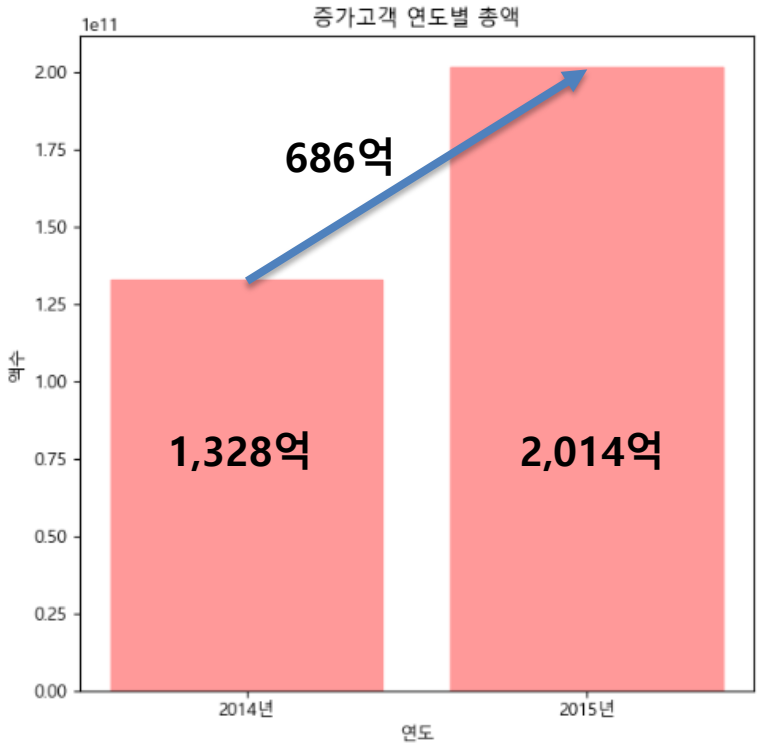
2014년 대비 2015년의 총액 증감률은 4.52%(147억)

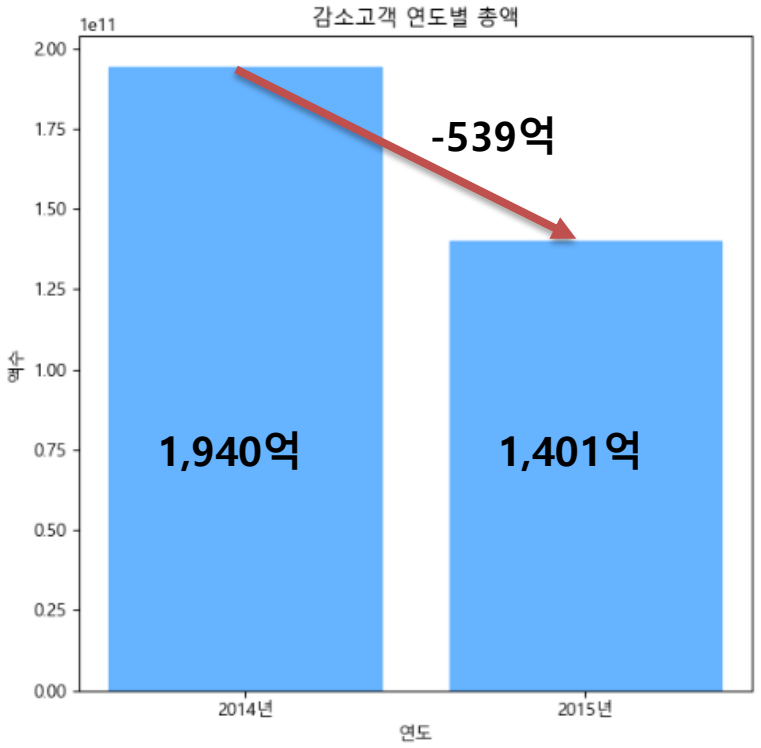
EDA – 감소 고객 정의

인플레이션을 고려한 실질적 성장률 계산식으로
L사 성장률, 고객당 성장률 계산

※인플레이션 : 일정 기간 동안 물가가 지속적이고 비례적으로 오르는 현상
※ 경제성장률 = {(금년도 실질 GDP - 전년도 실질 GDP) ÷ 전년도 실질 GDP} × 100

EDA – 감소 고객 정의





증가고객 총액 21.01% 증가

감소고객 총액 16.49% 감소

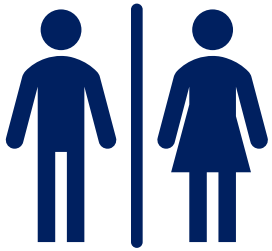
→ 전체적인 매출은 증가했지만 감소고객 그룹 -539억

Feature Engineering

범주화

Label Encoding

회귀와는 달리 숫자의 크기나 순서에 덜 민감한
분류 알고리즘의 특성을 감안



성별

남,여 0과 1로
Encoding

연령대

10개의 연령대
Encoding

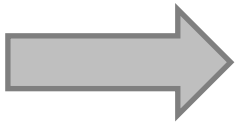
거주지역

17개의 거주지역
Encoding

파생변수

Feature

- 성별
- 연령대
- 거주지역
- 카테고리 14개
- 제휴사별 총액
- 총액
- 연평균 증가율
- 방문횟수
- 시간대별
- RECENCY(최근 방문 경과일수)



개발한 Feature 지수화 후 머신러닝

지수화

지수화

1. 분기별 총 구매금액을 10분위로 등급화 (구매금액 0원일 경우 1등급)
2. 전 분기 대비 등급 변동값 계산 (2분기 – 1분기, 3분기 – 2분기, …)
3. 변동값을 모두 합한 순증감지수와 절대값들을 모두 합한 변동지수 계산
4. 제휴사별, 분기별 매출, 구매횟수, 방문횟수, 방문 시간대, 상품 카테고리
   → **지수화**

분기별 등급화

1분기	2분기	3분기
4	5	4
2	5	7
10	9	9
7	5	3

→

전 분기 대비 변동값

1_2변동값	2_3변동값
1	-1
3	2
-1	0
-2	-2

→

순증감지수, 변동지수

순증감지수	변동지수
0	2
5	5
-1	1
-4	4

모델링 IV

데이터셋 분류

데이터를 8:2 비율로 나누어 학습, 검증, 테스트를 진행

독립변수

- 고객 속성 변수 : 고객이 고유하게 가지고 있는 속성을 의미하는 변수
- 구매 패턴 변수 : 분기 별 변화하는 구매 패턴을 의미하는 변수

성능 평가

	Random Forest*	Decision Tree	Logistic Regression	Light GBM	XGBoost
Accuracy	0.7491	0.6445	0.7427	0.7452	0.7315
Precision	0.7114	0.6014	0.6955	0.7083	0.6946
Recall	0.7503	0.6383	0.7682	0.7438	0.7267
F1 score	0.7304	0.6193	0.7301	0.7256	0.7103
ROC AUC	0.8263	0.6440	0.8235	0.8253	0.8077

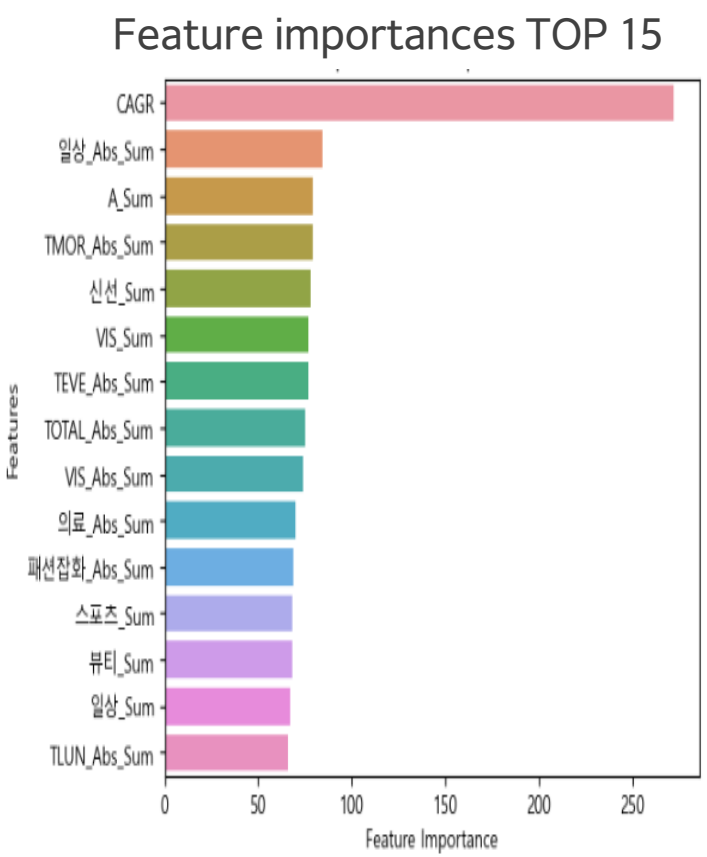
GridsearchCV를 통한 최적화 Hyper Parameter

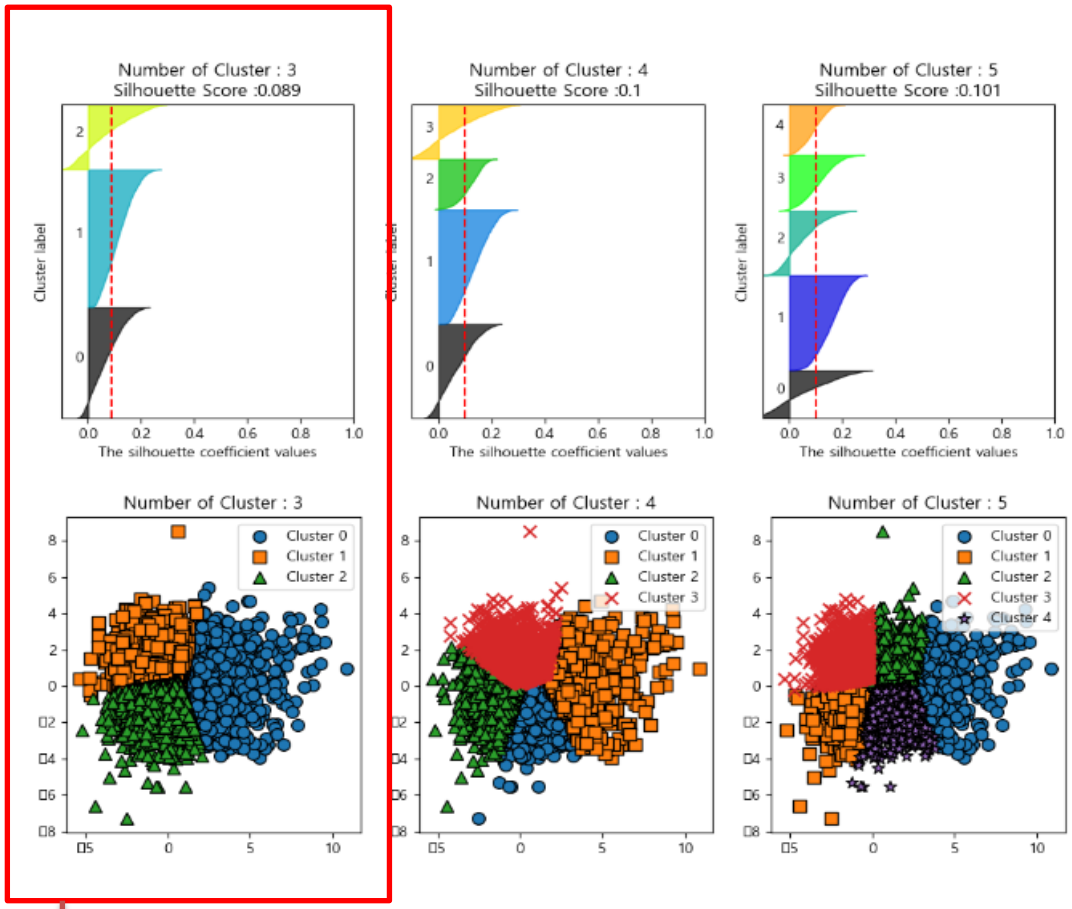
Accuracy	0.7536
Precision	0.7130
Recall	0.7634
F1 score	0.7373
ROC AUC	0.8320

V

군집화와 마케팅 제언

군집화





3개의 군집으로 구매 감소 고객 분류

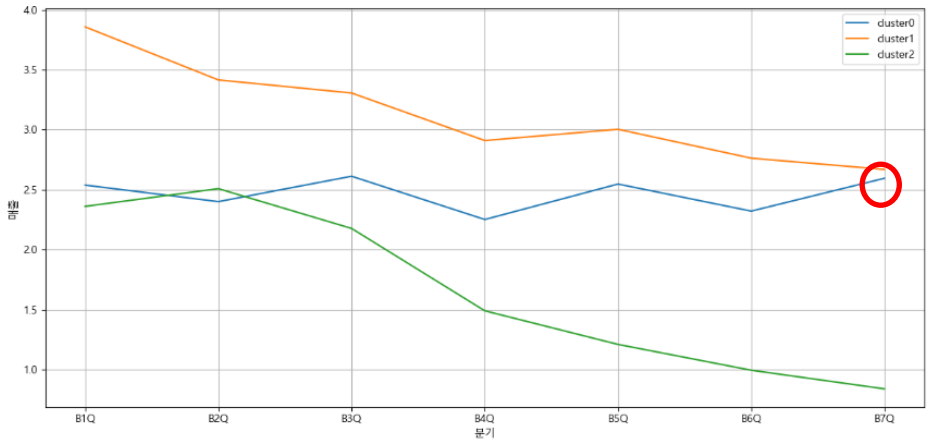
군집별 분석 - 0번 군집

군집 특징

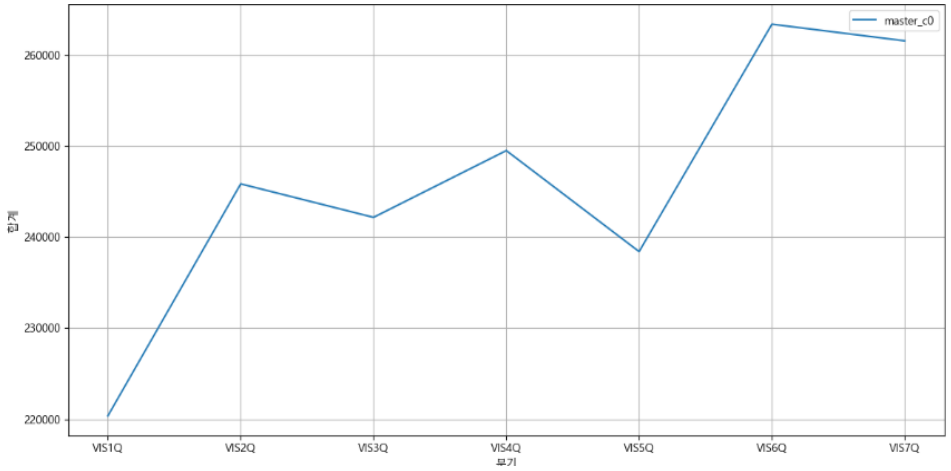
1분기 대비 8분기
증가를 보인 항목

- 전체 방문 횟수
- 저녁 시간대 방문 횟수
- 마트, 슈퍼의 분기별 매출
- 가공,신선 상품 매출

분기별 마트 매출

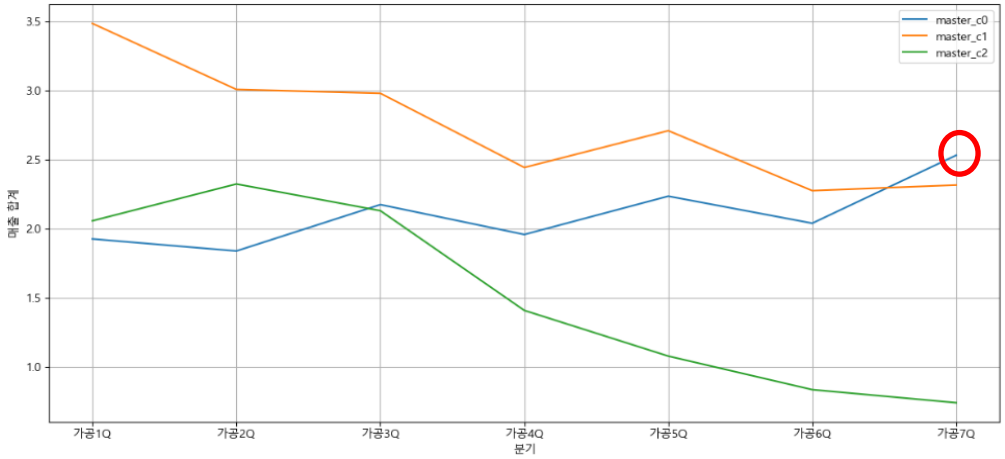


분기별 방문횟수

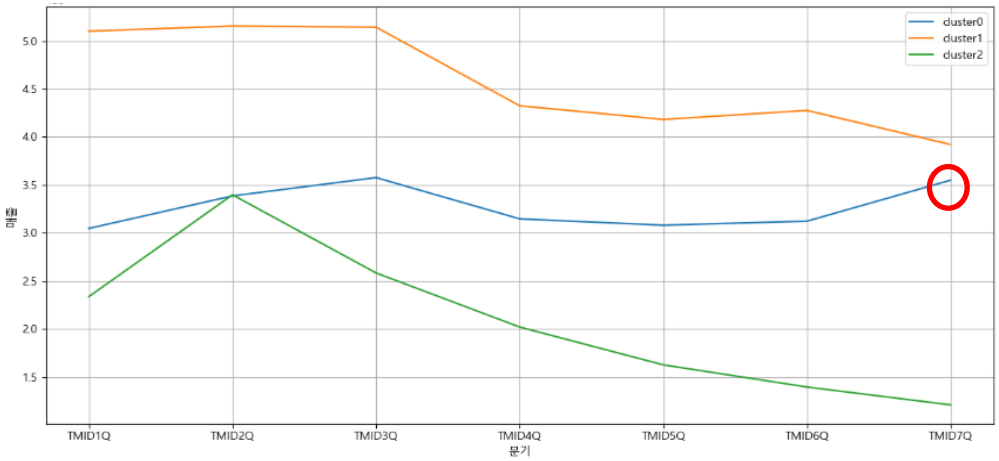


군집별 분석 – 0번 군집

분기별 가공 식품 매출



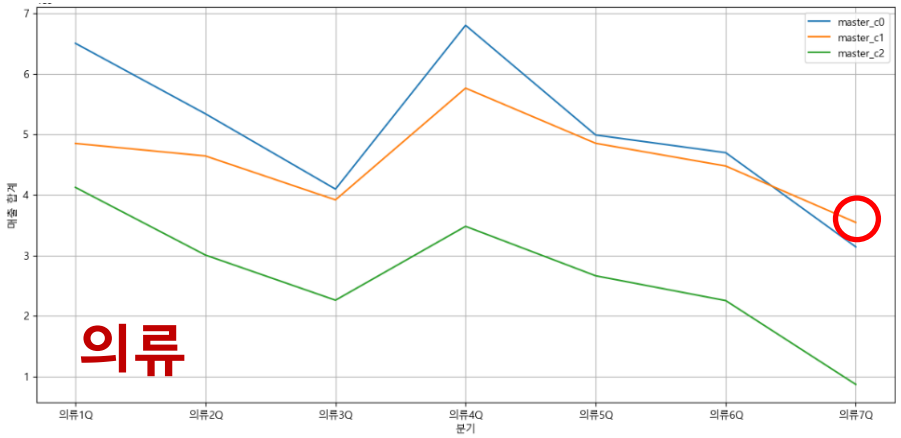
분기별 심야 매출

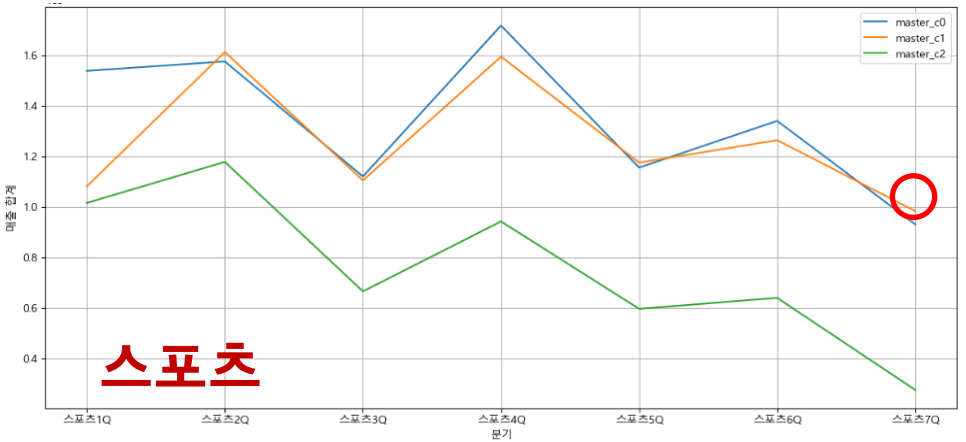


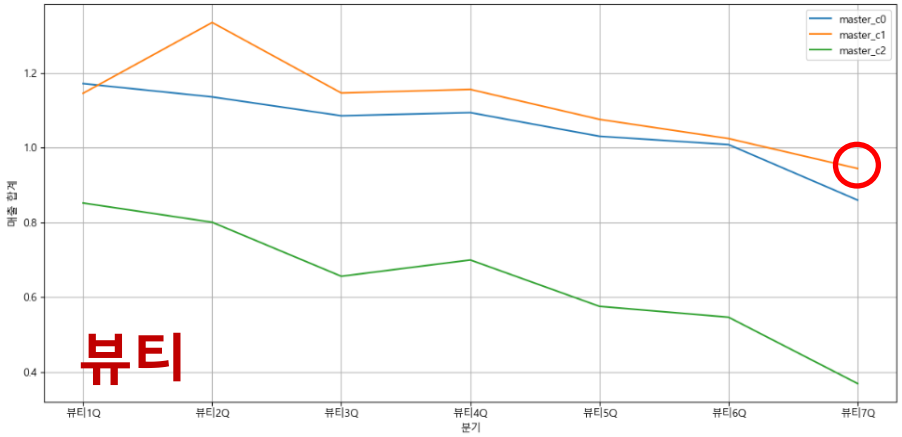
마케팅 제언

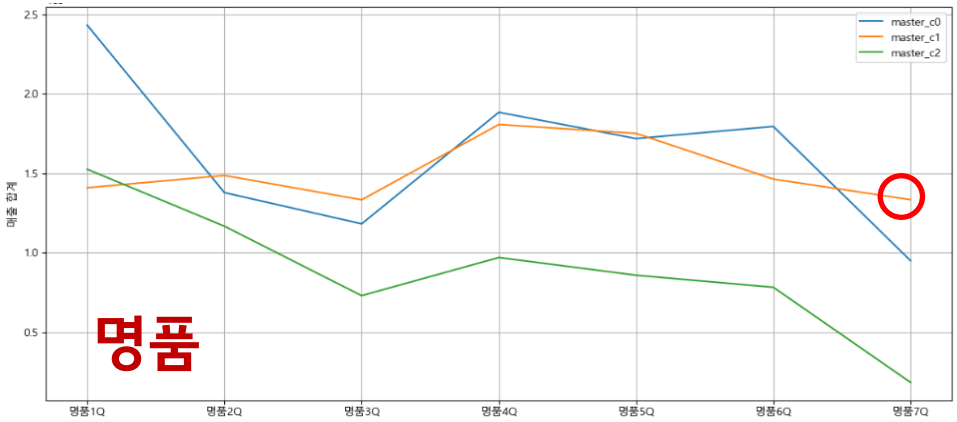
- 가공 상품 카테고리 위주 추천시스템 제안
- 심야 시간대 고객과 상품 특성을 고려한 접근

군집별 분석 – 1번 군집









군집 특징

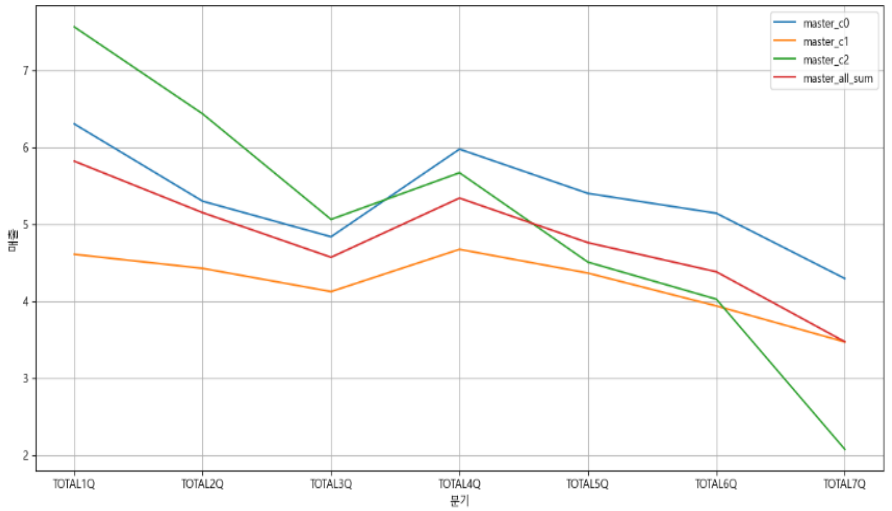
선방한 상품 카테고리 목록

마케팅 제언

추천 시스템 활용 제안

군집별 분석 – 2번 군집

분기별 전체 매출



군집 특징

- 세 군집 중 가장 적은 인원수, 매출액
- 모든 상품에서 감소추세
- 1인당 평균 금액 높음(구매력 높음)

마케팅 제언

할인쿠폰, 이벤트, 프로모션

** 단위 : 억*

Cluster	총 매출액
0	1,232
1	1,218
2	684
합계	3,134

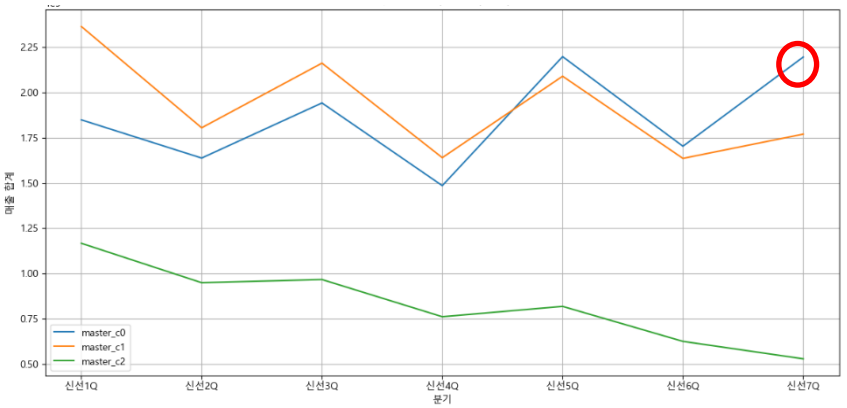
** 단위 : 명*

Cluster	인원 수
0	3,306
1	4,114
2	1,936
합계	9,356

** 단위 : 만원*
** (구매력 : 총 매출액 / 인원 수)*

Cluster	구매력
0	3,726
1	2,961
2	3,535
합계	10,223

고객 개인화 상품 추천




```python
# RMSE
reader = Reader(rating_scale=(0, df_0['score'].max()))
data = Dataset.load_from_df(df_0[['고객번호', '소분류명', 'score']], reader)

trainset, testset = train_test_split(data, test_size=.25, random_state=0)
algo = SVD(n_factors=50, random_state=0)
algo.fit(trainset)
predictions = algo.test(testset)
accuracy.rmse(predictions)
```

```
RMSE: 0.9411
```

RMSE : 0.9411

고객번호	소분류명
13148	치즈
13148	쿠키
13148	기능성우유
13148	바나나
13148	옥수수스낵

고객번호	소분류명
19304	청과
19304	유기농채소
19304	채소
19304	서적
19304	모피

고객번호	소분류명	카테고리	구매횟수
13148	유기농 채소	신선식품	1
	채소	신선식품	4
	청과	신선식품	4
	치즈	가공식품	72
	쿠키	가공식품	133

고객번호	소분류명	카테고리	구매횟수
19304	서적	교육/문화용품	18
	유기농 채소	신선제품	54
	채소	신선제품	31
	청과	가공제품	16

Q & A

감사합니다

박주경 - https://github.com/llikespike
김예슬 - https://github.com/KYeseul
남동연 - https://github.com/namdongyeon
오윤택 - https://github.com/ohyunteak
조차선 - https://github.com/chasuncho