HELLO.

#### 4조

# 매출감소 유지고객 예측모델 개발 및 솔루션

2020.10.18

#### -- 조원 소개



오수문 (조장) ansvltm501@gmail.com github.com/sumunoh 기획 총괄 데이터 과제화, 이슈 파인딩, 전처리, 기술 및 코드, SQL



신인철 snc4656@naver.com github.com/InChil2 데이터 과제화, 이슈 파인딩, 자동화, 기술 및 코드 총괄 기획, SQL

## -- 조원 소개



우동주 w.dj7601@gmail.com github.com/hamstack

기획, 이슈 파인딩, 시각화, 기술 및 전처리, 검증



김동일 dongil0324@gmail.com

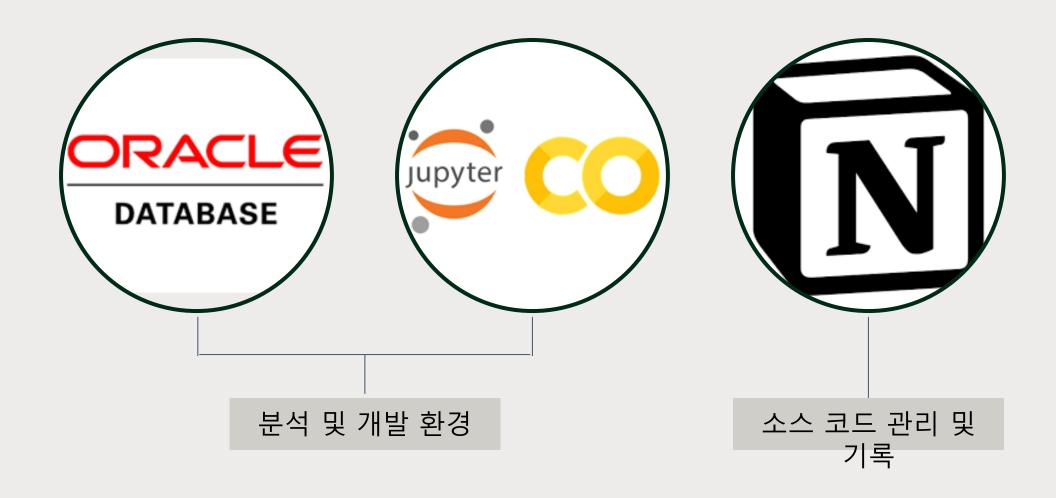
기획, 이슈 파인딩, 자료수집, 전처리

## -- 조원 소개



박기범	기획, 자료수집
-----	----------

# 一 개발 환경

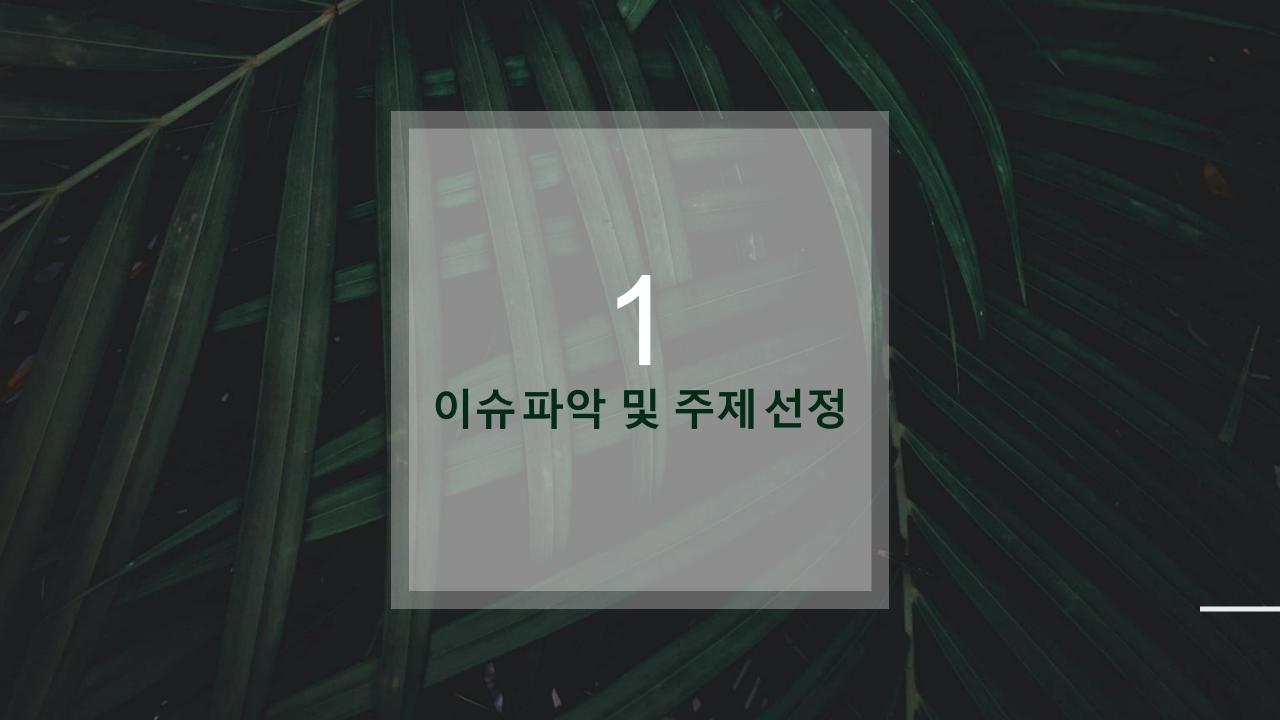


# INDEX.

**1** . 이슈 파악 및 주제 선정

 2.
 데이터 분석 및 모델링

제안 및 정리



1차 데이터 파악

#### 분석 대상

2년간 (2014년~2015년) L사 4개의 계열사에서 구매한 고객

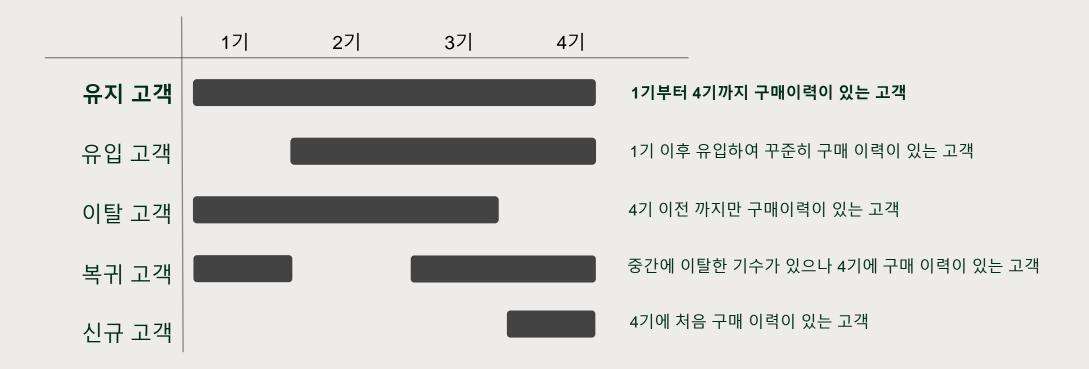
#### 제공 범위

- L사의 4개 계열사의 구매이력
- 고객의 성향을 파악 할 수 있는 데이터 제공

#### 데이터 목록 및 컬럼정보

- 고객세부정보: 성별, 연령대, 거주지
- 고객구매내역: 제휴사, 영수증, 구매상품목록, 구매일자, 구매시간, 구매금액
- 상품분류: 대분류/중분류/소분류 코드 및 분류명 정보
- 고객경쟁사이용: 고객의 경쟁사 이용 정보
- 고객멤버십여부: 멤버십 가입 정보 및 가입 년월
- 고객의 채널이용: 최근 3 개월간의 모바일/APP 로그인 횟수 / 온라인쇼핑몰 구매 횟수

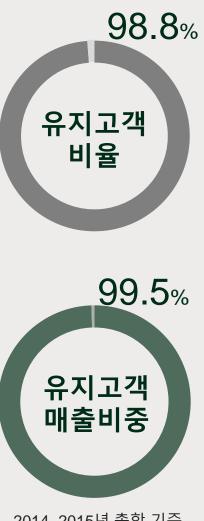
방문 유형별 고객 구분



방문 유형별 고객 구분

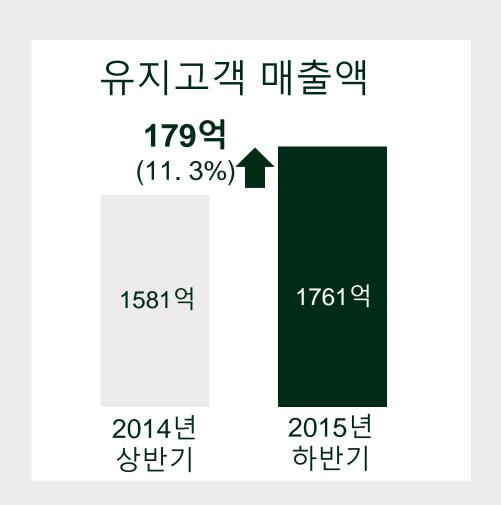
	고객번호	1기_이용금액	2기_이용금액	3기_이용금액	4기_이용금액	14-1	14-2	15-1	15-2	고객구분
0		19925355	34116312	17898365	13102518					유지
1	2	24323230	21922100	27791610	25843820					유지
2		851215	1042554	592280	679118					유지
3	4	5042106	4257510	3656937	3737257					유지
4	5	5460150	2225730	2934930	2616260					유지
19378	19379	0	1190620	2395920	2192740	0				유입
19379	19380	0	522700	837950	2552327	0	1			유입
19380	19381	0	0	2797171	3243498	0	0			유입
19381	19382	0	0	1826739	2575034	0	0			유입
19382	19383	0	0	27000	2357400	0	0			유입
19383 ro	ws : 10 col	umns								

- 전체고객 19383명 중 19147명이 유지고객에 해당



2014~2015년 총합 기준

고객 특성 파악 및 매출 추이

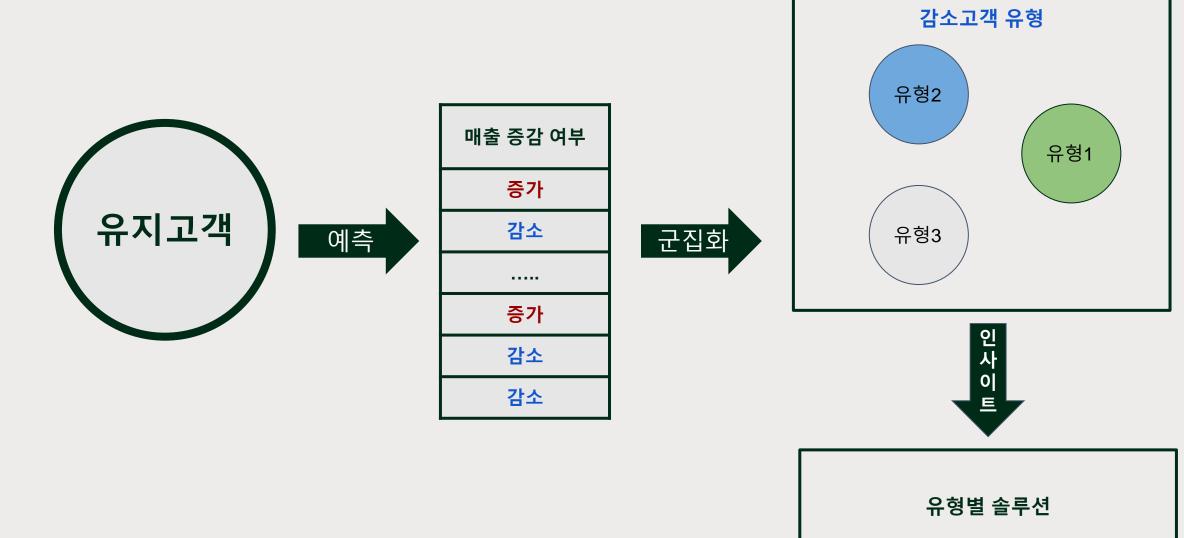


2014년 상반기 대비 2015년 하반기 매출 감소 유지고객 비율

유지고객 19147명 중 **8243명** 

43%

주제선정





FIRST 데이터 전처리

데이터 가공 및 재범주화

1차 데이터인 상품분류, 고객정보 등을 재범주화 및 병합 SECOND 매출감소 예측

매출감소 예측모델 개발

전처리 과정을 거친 데이터를 머신러닝, 차원축소 등을 고려하여 예측모델 개발 및 검증 THIRD 매출감소 솔루션

매출고객군집별 솔루션

예측 모델 기반 감소 고객 군집화, 유형별 구매패턴을 파악 하고, 솔루션 제안

# 2. 데이터 분석 및 모델링 종합 데이터셋

데이터셋 종속변수		데이터셋 독립변수
구매변동_1-4(1기 대비 4기 매출 증감 여부)	고객번호	고객번호
1기 대비 4기 매출 증감 여부	성별	성별
	연령대	연령대
	구매금액_1_3P	1기 대비 3기 구매금액 등급 변동
	구매금액_1_3P_A	1기 대비 3기 A제휴사 구매금액 등급 변동
	구매금액_1_3P_B	1기 대비 3기 B제휴사 구매금액 등급 변동
	구매금액_1_3P_C	1기 대비 3기 C제휴사 구매금액 등급 변동
	구매금액_1_3P_D	1기 대비 3기 D제휴사 구매금액 등급 변동
	가구/가전/전자기기_1_3P	1기 대비 3기 가구/가전/전자기기 구매금액 등급 변동
	기타_1_3P	1기 대비 3기 기타 구매금액 등급 변동
	생활/건강_1_3P	1기 대비 3기 생활/건강 구매금액 등급 변동
	식품_1_3P	1기 대비 3기 식품 구매금액 등급 변동
	의류/잡화_1_3P	1기 대비 3기 의류/잡화 구매금액 등급 변동
	반려/취미/문구/도서_1_3P	1기 대비 3기 반려/취미/문구/도서 구매금액 등급 변동
	스포츠/레저/자동차_1_3P	1기 대비 3기 스포츠/레저/자동차 구매금액 등급 변동
	출산유아동_1_3P	1기 대비 3기 출산유아동 구매금액 등급 변동
	방문횟수_1_3P	1기 대비 3기 방문횟수 구매금액 등급 변동
	중분류개수_1_3P	1기 대비 3기 중분류개수 구매금액 등급 변동
	상위10%물품_1_3P	1기 대비 3기 물건 가격 상위 10%에 해당하는 물품 구매개수 등급 변동

# 주요독립변수

PART1. 제휴사별 매출변화

제휴사별 그룹화 매출 증감 등급화

# 종속변수

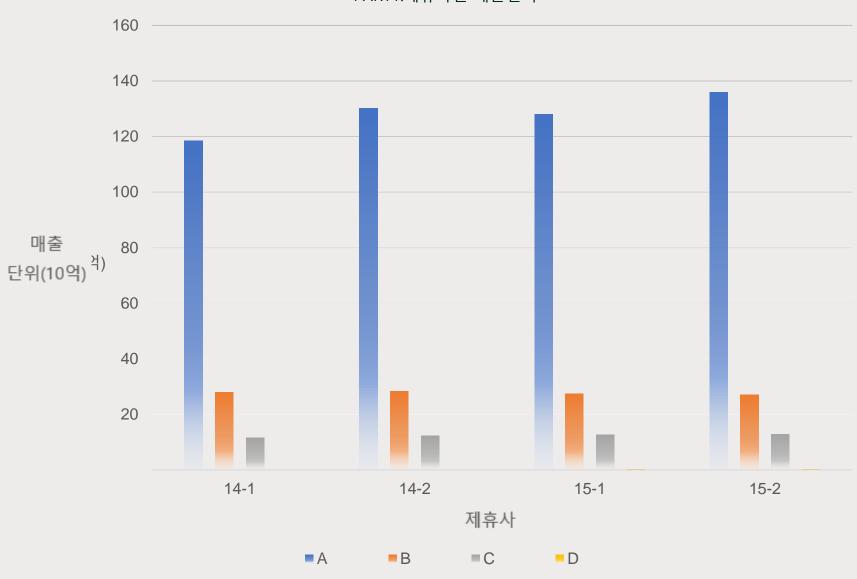
유지고객의 1기 대비 4기 매출증감 여부 PART2. 대분류별 매출변화/구매제품군 변화

대분류별 매출 변화 등급화 구매 제품군 변화량 측정

PART3. 상위 10% 물품

물품 가격의 10%에 해당하는 물품 횟급 변화 중국

**주요독립변수** PART1.제휴사별 매출변화



#### 주요독립변수

PART2.대분류별 매출변화/구매제품군 변화

#### 중분류명

담배

화장품

미용소품

두발용품

기기

헤어케어

과일류

채소류

반찬류

헤어스타일링

L사 온라인 쇼핑몰 상품 카테고리 참고15개의 대분류 87개의 중분류 재범주화



#### 가구/가전/전자기기

['인테리어', '침구', '전자기기', '가전', '디지털', '가구', '게임']

#### 기타

['바디용품', '다이소', '화장', '일반식당', '편의시설', '헤어', '케어', '화장소품', '담배', '풋/핸드', '남성 케어', '남성향수', '일반의약외품', '여성향수', '패스트푸드', '카페', '레스토랑', '해외직구']

#### 반려/취미/문구/도서

['문구', '미술', '파티/팬시', '반려', '교육', '장난감', '도서', '핫팩', '식물', '음악', '스포츠', '여성의류']

#### 생활/건강

['여성용품', '세제/방향/탈취', '주방용품', '생활집화', '청소용품', '정리용품', '욕실용품', '구강용품', 위생용품', '세탁용품', '안전용품', '헬스용품', '공구/안전용품']

#### 스포츠/레저/자동차

['아웃도어', '스포츠', '인테리어', '레저', '자동차']

#### 식품

['음료', '주류', '동산품', '유제품', '가공식품', '수산품', '라면/통조림/즉석밥', '빵류', '육류', '과일', '반찬', '소소/양념/조미료', '간식', '동산물', '쌀/잡곡', '건강식품', '건해산물', '견과류', '회/초밥', '수산물']

#### 의류/잡화

['남성의류', 'ACC', '속옷', '패션잡화', '여성의류', '신발', '명품', '주얼리', '가방', '시계']

#### 출산유아동

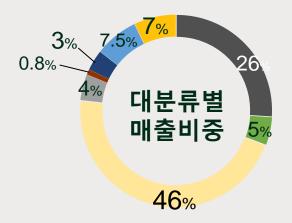
['출산유아동']

매출 분포가 비슷한 분류끼리 대분류 재통합 8개의 대분류

기존 데이터 중분류명 개수: 717개

주요독립변수

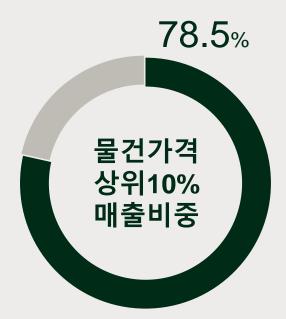
PART2.대분류별 매출변화/구매제품군 변화



■의류/잡화 ■식품 ■스포츠/레저/자동차 ■기타・■가구/가전/전자기기 ■생활/건강 ■출산유아동 ■반려/취미/문구/도서

#### 주요독립변수

PART3. 상위 10% 물품 구매횟수



물건가격 상위 10% 매출 증감 영향에 절대적 1. 구<mark>매내역 테이블 구매금액 내림차순 정렬</mark>

time\_all = all[all['구매일자']<=20141231] bbb = time\_all.sort\_values(by=['구매금액'],ascending=False) bbb

제휴사	명수증번호	소분류코드	고객번호	구매일자	구매시간	구매금액	소분류명
Α	1416069	A080106	14323	20140221	14	101330000	명품잡화
Α	1500785	A080106	14323	20140307	14	101110000	명품잡화
Α	1677262	A080106	14323	20140405	16	99960000	명품잡화
Α	2220608	A020303	8079	20140627	10	82800000	색조 화장품
Α	1862939	A080106	14323	20140503	15	72915000	명품잡화

2. 상위 10% 물건가격 확인 
 제휴사
 영수증번호
 소분류코드
 고객번호
 구매일자
 구매시간
 구매금액
 소분류명

 B
 6450546
 B060114
 679
 20140522
 11
 29000
 게

3. 상위 10%해당 기수별 /고객별 횟수 count

상위10%물품_3	상위10%물품_2	상위10%물품_1	고객번호
87	122	114	1
118	104	124	2
2	0	1	3
24	31	40	4
12	17	31	5

# 2. 데이터 분석 및 모델링 등급화 및 등급 변화량 측정

각 독립변수별 **MEDIAN값** 확인 (2014년 상반기, 2014년 하반기 데이터 활용)



MEDIAN값을 중앙 등급(3등급) 으로하여 **등급간 간격 FIX** 



1~5등급 범주화



기수간 등급변화량 측정

가구/가전/전자기기_1	기타_1_3P	생활/건강_1_3P	식품_1_3P	의류/잡화_1_3P	반려/취미/문구/도서_	스포츠/레저/자동차_1	출산유아동_1_3P
2	-1	0	-2	0	0	0	0
0	0	0	-2	0	2	0	0
0	0	0	0	0	0	0	0
0	1	-1	0	-2	0	-2	1
-1	-2	-2	-1	-1	0	-1	0
3	0	-1	0	0	0	1	0
1	0	4	0	0	0	-2	0
0	0	0	0	-1	0	-1	0
0	4	0	0	0	0	0	0



27

1기 대비 3기 매출 증감 예측

하이퍼파라미터 서치 및 모델 검증



TRAIN / TEST 17

3기

1기 대비 4기 매출 예측

1기 = 2014년 상반기 2기 = 2014년 하반기 3기 = 2015년 상반기

4기 = 2015년 하반기

# ㅁ데벼 저하다 미 ᄌᆼ 버ᄉ

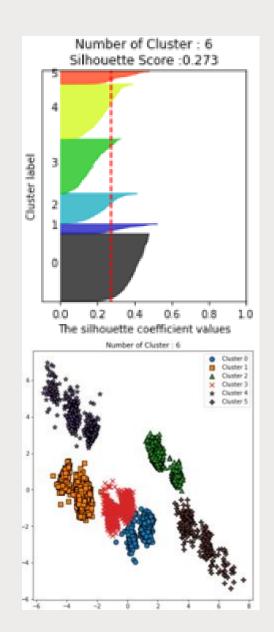
모델멀 경왁노 및 중요 먼수					
Random Forest  ['max_depth': 13, 'min_samples_leaf': 40, 'min_samples_split': 2, 'n_estimators': 215]	0.7337	<ul> <li>→ 성별</li> <li>→ 연령대</li> <li>→ 구매금액_1-3P</li> <li>→ 구매금액_1-3P_B</li> </ul>			
<pre>Decision Tree {'criterion': 'gini', 'max_depth': 20, 'max_features': 'log2', 'min_samples_leaf': 80, 'min_samples_split': 2}</pre>	0.6298	<ul> <li>→ 성별</li> <li>→ 연령대</li> <li>→ 구매금액_1-3P</li> <li>→ 구매금액_1-3P_B</li> </ul>			
XGBoost  ['max_depth': 2, 'min_samples_leaf': 20, 'min_samples_split': 1, 'n_estimators': 200]	0.7407	<ul> <li>→ 방문횟수_1-3P</li> <li>→ 구매금액_1-3P</li> <li>→ 식품_1-3P</li> <li>→ 의류/잡화_1-3P</li> <li>→ 구매금액_1-3P_A</li> </ul>			
LightGBM  ['max_depth': 12, 'min_samples_leaf': 20, 'min_samples_split': 1, 'n_estimators': 210}	0.7439	→ 방문횟수_1-3P         → 식품_1-3P         → 구매금액_1-3P_A         → 의류/잡화_1-3P         → 구매금액_1-3P			

ㅁ데벼	저화 도	미즈이	出人
エヨヨ	O 의 각 도	大石山	민구

보델멸 성왁노 및 중요 면수				
Random Forest  ['max_depth': 13, 'min_samples_leaf': 40, 'min_samples_split': 2, 'n_estimators': 215]	0.7337	<ul> <li>→ 성별</li> <li>→ 연령대</li> <li>→ 구매금액_1-3P</li> <li>→ 구매금액_1-3P_B</li> </ul>		
<pre>Decision Tree  {'criterion': 'gini', 'max_depth': 20,</pre>	0.6298	<ul> <li>→ 성별</li> <li>→ 연령대</li> <li>→ 구매금액_1-3P</li> <li>→ 구매금액_1-3P_B</li> </ul>		
XGBoost  ['max_depth': 2, 'min_samples_leaf': 20,   'min_samples_split': 1, 'n_estimators': 200}	0.7407	<ul> <li>→ 방문횟수_1-3P</li> <li>→ 구매금액_1-3P</li> <li>→ 식품_1-3P</li> <li>→ 의류/잡화_1-3P</li> <li>→ 구매금액 1-3P A</li> </ul>		
LightGBM ['max_depth': 12, 'min_samples_leaf': 20, 'min_samples_split': 1, 'n_estimators': 210]	0.7439	→ 방문횟수_1-3P   → 식품_1-3P   → 구매금액_1-3P_A   → 의류/잡화_1-3P   → 구매금액 1-3P		

#### 2. 데이터 분석 및 모델링 <sub>군집화</sub>

변수명	설명
구매금액_1P_A	A제휴사의 구매금액
구매금액_1P_B	B제휴사의 구매금액
구매금액_1P_C	C제휴사의 구매금액
식품_1P	식품군 구매금액
의류/잡화_1P	의류/잡화 구매금액
방문횟수_1P	방문횟수
중분류개수_1	중분류 개수 변화
상위10%물품_1P	물품가 상위10% 상품의 개수
성별	성별
연령대	연령대
구매변동_1_2	1기 2기 구매변동(감소=1,증가=0)



## 매출 감소고객 대상 군집화

최종 Cluster 수: 6개





# 번 군 집

구분	대상
주 연령대	50대~60대
주 성별	여성
총 누적 구매금액	84,345,644,686원
A사 누적 구매금액	71,000,668,450원
B사 누적 구매금액	8,982,227,402원
C사 누적 구매금액	4,341,178,754원
총 방문횟수	2,605,988회
총 중분류개수	220,735개
방문당 소비금액(2년)	32,366원
인원당 소비금액(2년)	43,634,580원

구분	평균 증감
구매금액	-3,039,278원
A사 구매금액	-2,809,117원
의류/잡화 구매금액	-1,919,783원
B사 구매금액	-165,230원
식품 구매금액	-212,724원

## 총 1,933명(23.4%) (중요 1순위)

방문당 매출 비중이 높은 고객층 방문당 소비금액, 인원당 소비금액이 높아 꾸준한 관리가 필요하며 구매 품목의 폭이 넓어 전체적인 감소 고객층 매출의 대부분을 차지 이탈 방지를 위해 이용 기간이 길어지면 길어질 수록 혜택이 늘어나야 함



# 번 굽 집

구분	대상
주 연령대	30대~40대
주 성별	여성
총 누적 구매금액	74,442,474,142원
A사 누적 구매금액	57,953,718,220원
B사 누적 구매금액	11,435,068,197원
C사 누적 구매금액	4,990,461,556원
총 방문횟수	2,892,579회
총 중분류개수	238,327개
방문당 소비금액(2년)	25,736원
인원당 소비금액(2년)	38,731,780원

구분	평균 증감
구매금액	-2,250,308원
A사 구매금액	-1,963,486원
의류/잡화 구매금액	-1,270,790원
B사 구매금액	-221,223원
식품 구매금액	-211,223원

총 1,922명(23.3%) (중요 2순위)

방문당 매출이 높은 고객층 식품, 의류/잡화의 비율이 비슷하며 구매품목의 폭이 좁다 구매품목의 폭을 넓히는 방식의 마케팅 전략 필요



# 번 군 집

구분	대상
주 연령대	30대~40대
주 성별	여성
총 누적 구매금액	87,527,866,911원
A사 누적 구매금액	64,232,505,680원
B사 누적 구매금액	15,684,530,332원
C사 누적 구매금액	7,512,687,979원
총 방문횟수	3,587,285원
총 중분류개수	292,242원
방문당 소비금액(2년)	24,399원
인원당 소비금액(2년)	36,607,220원

구분	평균 증감
구매금액	-3,617,834원
A사 구매금액	-2,406,646원
의류/잡화 구매금액	-1,312,844원
B사 구매금액	-967,072원
식품 구매금액	-719,181원

## 총 2,391명(29%) (중요 3순위)

인원이 가장 많아서 저가 제품의 마케팅 효과를 기대해 볼 수 있는 고객층 구매의 폭도 넓고 인원당 소비 금액도 높은 편이다 의류/잡화의 매출 비중이 높아 식품등 다른 저가 품목의 폭을 넓히는 방식의 마케팅 필요



# 번군집

구분	대상
주 연령대	30대~50대
주 성별	여성
총 누적 구매금액	12,638,272,228원
A사 누적 구매금액	9,333,255,230원
B사 누적 구매금액	2,456,101,250원
C사 누적 구매금액	833,085,448원
총 방문횟수	618,903회
총 중분류개수	52,613개
방문당 소비금액(2년)	20,420원
인원당 소비금액(2년)	28,336,930원

구분	평균 증감
구매금액	-3,399,527원
A사 구매금액	-3,138,293원
의류/잡화 구매금액	-2,189,371원
B사 구매금액	-270,253원
식품 구매금액	-258,905원

## 총 446명(5.4%) (중요 4순위)

구매의 폭도 좁고 방문당 매출도 적은 고객층 하지만 인원당 매출이 낮지 않음 적은 인원이지만 매출 상승의 잠재력을 가지고 있음 따라서 지속적인 주시가 필요하며 방문횟수와 구매폭을 넓히는 전략의 마케팅 필요



# 번 갑 집

구분	대상
주 연령대	30대~40대
주 성별	여성
총 누적 구매금액	8,623,777,760원
A사 누적 구매금액	6,042,843,820원
B사 누적 구매금액	2,021,391,472원
C사 누적 구매금액	554,129,018원
총 방문횟수	461,269회
총 중분류개수	40,865개
방문당 소비금액(2년)	18,696원
인원당 소비금액(2년)	26,053,710원

구분	평균 증감
구매금액	-1,450,231원
A사 구매금액	-1,123,241원
C사 구매금액	-93,784원
B사 구매금액	-235,052원
식품 구매금액	-308,303원

## 총 331명(4%) (중요 5순위)

인원도 가장 적고 구매 폭, 방문횟수도 가장 적은 고객층 매출의 감소가 가장 낮다 우선 순위는 낮지만 식품 등 저가의 상품 위주의 마케팅 전략 필요



# 번 군 진

구분	대상
주 연령대	30대~50대
주 성별	여성
총 누적 구매금액	21,876,190,016원
A사 누적 구매금액	12,059,614,560원
B사 누적 구매금액	7,346,424,562원
C사 누적 구매금액	2,447,017,144원
총 방문횟수	1,577,898호
총 중분류개수	135,447개
방문당 소비금액(2년)	13,864원
인원당 소비금액(2년)	20,599,050원

구분	평균 증감
구매금액	-1,318,175원
A사 구매금액	-987,975원
의류/잡화 구매금액	-590,928원
B사 구매금액	-309,117원
식품 구매금액	-233,693원

총 1,062명(12.9%) (중요 6순위)

방문당 매출 비중이 적은 고객층 의류/잡화 포인트 적립, 쿠폰 추천 등 부족한 매출을 보완

# 감사합니다