



# L사 유통 데이터 분석

구매 감소 고객 예측 모델을 통한  
마케팅 솔루션

...

김예지, 김형중, 김희진, 양정민, 이정호

# 프로젝트 목차



01

## 분석 배경

데이터 소개, 연관 자료, 개발 환경



02

## 데이터 분석 및 주제 선정

데이터 분석, 데이터 처리 및 정제



03

## 분류모델 개발 및 평가

라벨 설정, 변수 설정, 분류모델 학습 및 평가



04

## 분류 데이터 군집화

군집 변수 설정, 군집 후 분석



05

## 마케팅 솔루션 제안

고객 군집별 솔루션 제안



01

분석배경

...

# 분석 배경

## 분석 대상

L사 4개 제휴사의  
2014년~2015년 유통  
데이터

## 사용 데이터

**멤버십여부**  
고객번호 / 멤버십명 / 가입년월

**채널이용**  
고객번호 / 제휴사 / 이용횟수

**경쟁사이용**  
고객번호 / 제휴사 / 경쟁사/  
이용년월

## 사용 데이터

**고객DEMO**  
고객번호 / 성별 / 연령대 / 거주지역

**상품분류**  
제휴사 / 대중소분류코드 / 중,소분류명

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

## 외부 데이터

**유통상품 표준DB :**  
**대한상공회의소**  
**유통물류진흥원**

출처 :  
[http://www.allproductkorea.or.kr/  
products/database](http://www.allproductkorea.or.kr/products/database)

# 분석 배경

## 옴니채널

소비자가 모든 플랫폼이나 디바이스에 관계없이  
일관된 브랜드 구매 경험을 창출하는 유통방식

2015년 유통업계 키워드는 “옴니채널·모바일·글로벌”

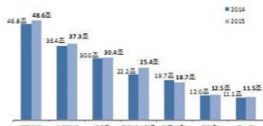
박지영 2014.12.01 11:39:54

가+ 가-

주요 연매매 예상 매출 성장률

구분	'14년(E)	'15년(F)
소매업 전체	2.2%	2.4%
인터넷쇼핑몰	13.0%	14.3%
편의점	3.1%	4.0%
대형마트	3.8%	3.8%
홈쇼핑	3.0%	3.2%
슈퍼마켓	1.9%	2.3%
백화점	1.0%	1.3%
전통시장	-4.8%	-5.0%

주요 연매매 예상 매출 규모



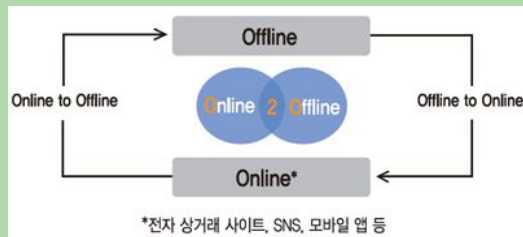
• 2013년 통계청 소매업(소매업) 기준으로 이상 상향세를 반영해 추정치  
• 인터넷쇼핑몰과 홈쇼핑은 통계청 소매업(소매업) 부문으로 별도 추정  
• 전통시장의 소매업(소매업)은 2013년도 전통시장 매출 추산치 참조

• 전년대비 전통시장 성장률: 추정치, F: 예측치

## O2O

### Online to Offline

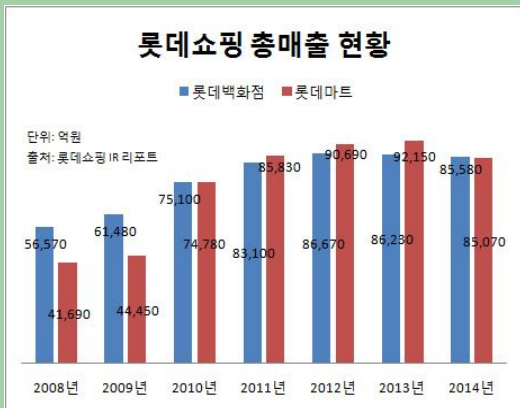
온라인에서 구매를 하고, 오프라인으로 물건을 받는  
것



# 분석 배경

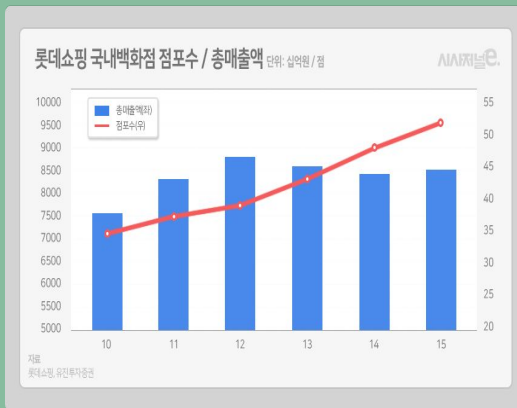
## L 마트

### 마트 매출 부진



## L 백화점

### 점포수 대비 총매출액 부진



## 분석 배경

데이터 탐색 및  
전처리



python



SQL



예측 모델 개발

PYCARAT



LightGBM



추천 시스템 개발

surprise

A Python scikit for  
recommender systems.



02

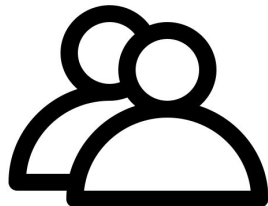
## 데이터 분석 및 주제 설정

...

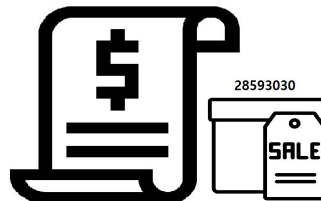


# DataBase 분석

19383



11086939



28593030

경쟁사이용				
Untitled5				
고객번호	cust_num2	Domain	NUMBER	
제휴사	affiliates	Domain	VARCHAR2	
경쟁사	competitor	Domain	VARCHAR2	
이용년월	using_date	Domain	NUMBER	

고객DEMO				
Untitled				
고객번호	cust_num	Domain	NUMBER	
성별	gender	Domain	VARCHAR2	
연령대	age	Domain	VARCHAR2	
거주지역	Field4	Domain	VARCHAR2	

구매상품TR				
Untitled2				
고객번호	cust_num2	Domain	NUMBER	
제휴사	affiliates	Domain	VARCHAR2	
영수증번호	receipt	Domain	VARCHAR2	
대분류코드	l_category_code	Domain	NUMBER	
중분류코드	m_category_code	Domain	VARCHAR2	
소분류코드	s_category_code	Domain	VARCHAR2	
점포코드	store_num	Domain	VARCHAR2	
구매일자	pur_date	Domain	NUMBER	
구매시간	pur_time	Domain	NUMBER	
구매금액	pur_price	Domain	NUMBER	

상품분류			
Untitled3			
제휴사	affiliates	Domain	VARCHAR2
대분류코드	l_category_code	Domain	NUMBER
중분류코드	m_category_code	Domain	VARCHAR2
소분류코드	s_category_code	Domain	VARCHAR2
중분류명	m_category_name	Domain	VARCHAR2
소분류명	s_category_name	Domain	VARCHAR2

채널이용				
c				
고객번호	cust_num2	Domain	NUMBER	
고객번호	cust_num	Domain	NUMBER	
제휴사	affiliates	Domain	VARCHAR2	
이용횟수	Field3	Domain	NUMBER	

멤버십여부				
Untitled4				
고객번호	cust_num	Domain	NUMBER	
고객번호	Field	Domain	NUMBER	
멤버십명	Field2	Domain	VARCHAR2	
가입년월	Field3	Domain	NUMBER	

제휴사	판매수	판매액
A	5770318	514841144030
B	13338074	111221686478
C	9379236	50060970889
D	105402	895355544

분류	
대분류	92
중분류	656
소분류	4386

## 01 유통 데이터 대분류 추출

'전체 분류' 칼럼 안에 '대분류>중분류>소분류>세분류' 형태로 입력되어 있음

-> 대분류만 추출하여 칼럼 추가

	KAN_CODE	전체 분류	현재 분류명	상품예시	분류1
0	1010101.0	가공식품>조미료>종합조미료>천연/발효조미료	천연/발효조미료	다시다, 미원, 멸치가루, 버섯가루	가공식품
1	1010102.0	가공식품>조미료>종합조미료>식초	식초	과일초, 발효초, 감식초, 사과식초, 현미식초, 홍초	가공식품
2	1010103.0	가공식품>조미료>종합조미료>천일염	천일염	구운천일염, 꽃소금, 굵은소금, 가는소금, 볶은소금	가공식품
3	1010104.0	가공식품>조미료>종합조미료>가공염	가공염	정제염, 맛소금, 허브솔트, 깨소금	가공식품
4	1010105.0	가공식품>조미료>종합조미료>설탕	설탕	슈가파우더, 과당, 각설탕, 황설탕, 백설탕, 흑설탕	가공식품
...	...	...	...	...	...
3107	11150301.0	패션잡화>양말>유아동주니어양말/기타양말>유아동주니어양말	유아동주니어양말	유아동양말, 주니어양말, 유아동캐릭터양말	패션잡화

## 02 상품명 분리 후 대분류 지정

	중분류명	대분류명
0	가공	가공식품
1	가공식품	가공식품
2	식품	가공식품
3	조미료	가공식품
...	...	...
26388	실내화	패션잡화
26398	기타상품	기타상품

## 03 상품 분류 데이터와 결합하여 대분류 추가

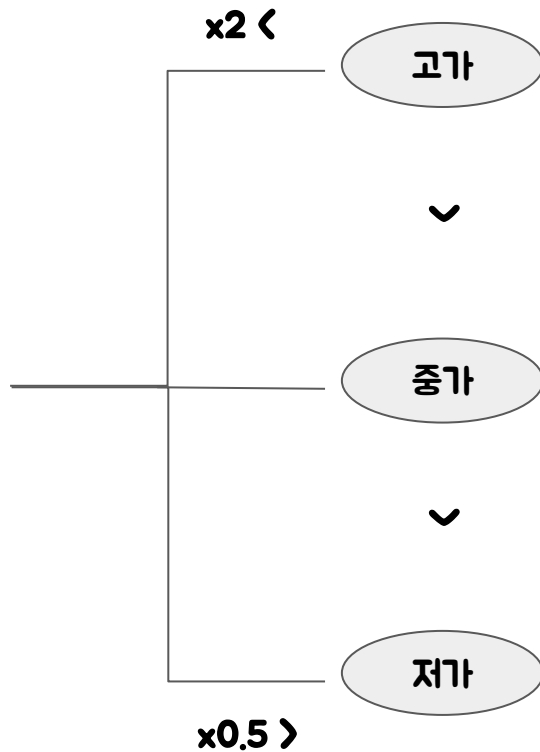
	제휴사	대분류코드	중분류코드	소분류코드	중분류명	소분류명	대분류
0	A	1	101	A010101	일용잡화	위생세제	일상용품
1	A	1	101	A010102	일용잡화	휴지류	일상용품
2	A	1	101	A010103	일용잡화	뷰티상품	일상용품
3	A	1	101	A010104	일용잡화	일용잡화	일상용품
4	A	1	101	A010105	일용잡화	세트상품	일상용품
...	...	...	...	...	...	...	...
4381	D	8	802	D080204	아이메이크업	아이섀도우	일상용품
4382	D	8	802	D080205	아이메이크업	메이크업세트	일상용품
4383	D	8	803	D080301	립메이크업	립스틱/립라이너	일상용품
4384	D	8	803	D080302	립메이크업	립글로즈/틴트	일상용품
4385	D	8	804	D080401	네일메이크업	일반네일/케어류	일상용품

4386 rows × 7 columns

# 칼럼 추가

## 가격분류

대분류	평균금액
가공식품	10636.40
가구/인테리어	392312.89
교육/문화용품	51179.62
기타	93786.60
디지털/가전	30811.79
신선식품	5811.59
음식점	6448.87
의류	337481.82
의약품/의료기기	15365.52
일상용품	36345.13
전문스포츠/레저	14538.35
패션잡화	147900.05

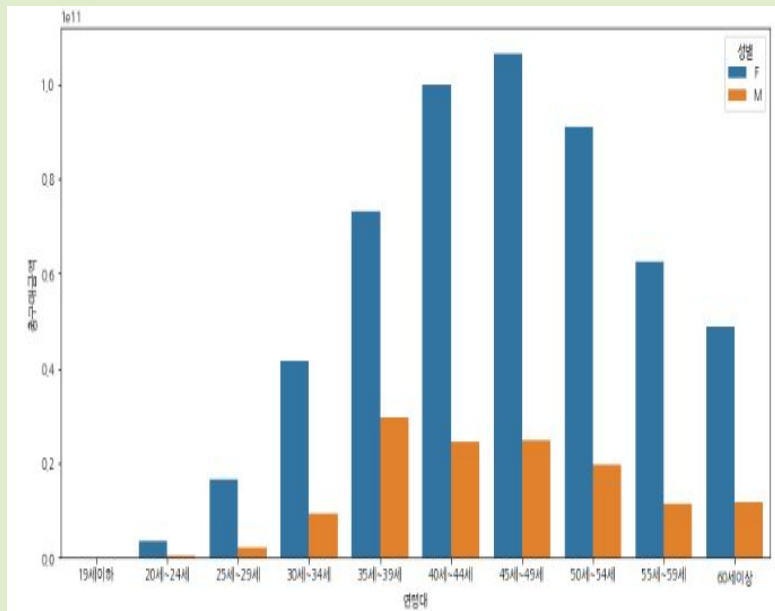


대분류	구매금액	분류
일상용품	13500	저가
가공식품	2500	저가
가공식품	6470	중가
신선식품	15800	고가
가공식품	8000	중가
일상용품	9000	저가
기타	647190	고가
디지털/가전	25000	중가
신선식품	3990	중가
패션잡화	392400	고가
교육/문화용품	22050	저가
일상용품	377330	고가
음식점	15800	고가
패션잡화	14950	저가

# 데이터 분석

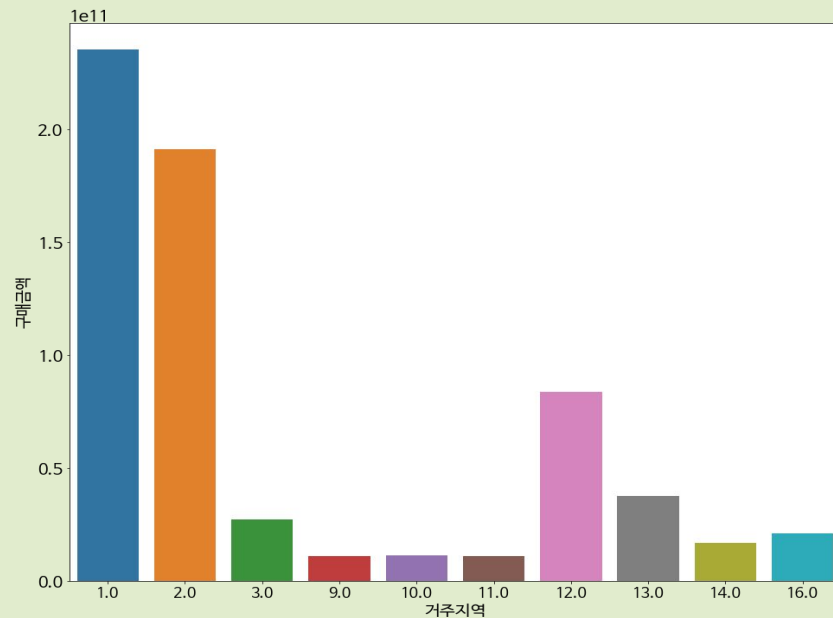
## 01 성별, 연령별 매출 분석

40대 여성 매출이 높음



## 02 거주지역별 매출 분석

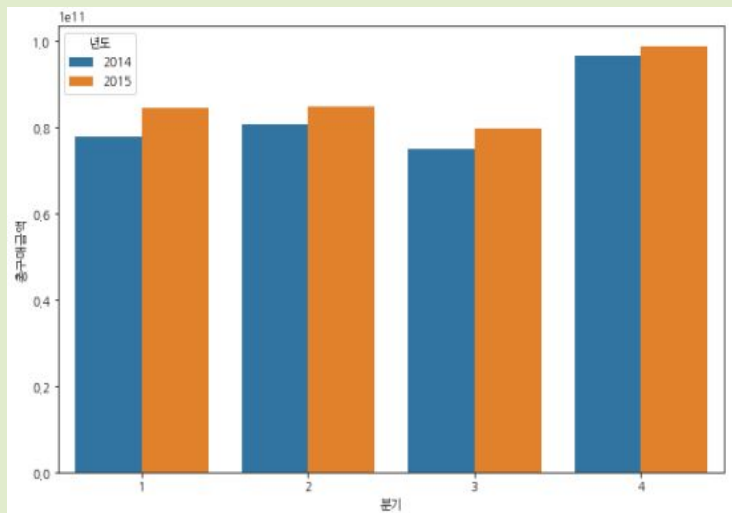
총매출액 TOP10



# 데이터 분석

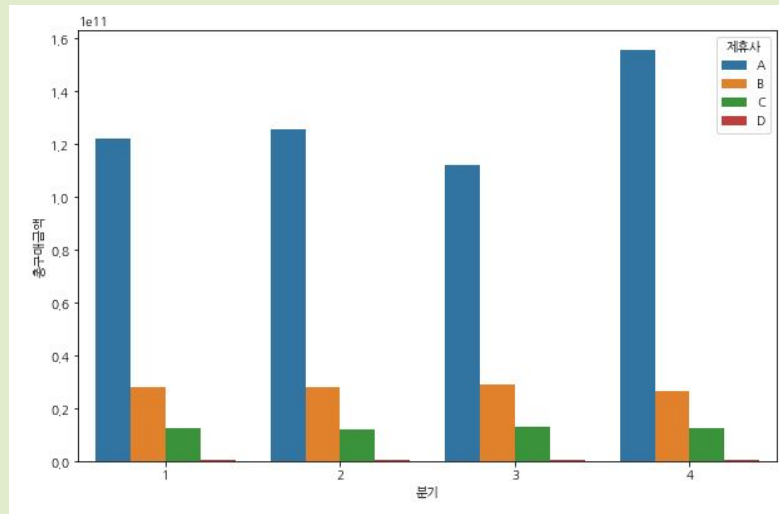
## 03 분기별 매출 분석

매출 증가, 3분기 매출이 가장 낮음



## 04 제휴사별 매출 분석

A사가 가장 큰 비중  
B사만 매출 감소, 나머지 제휴사는  
매출 증가



# 데이터 분석

## 05 카테고리별 매출 분석

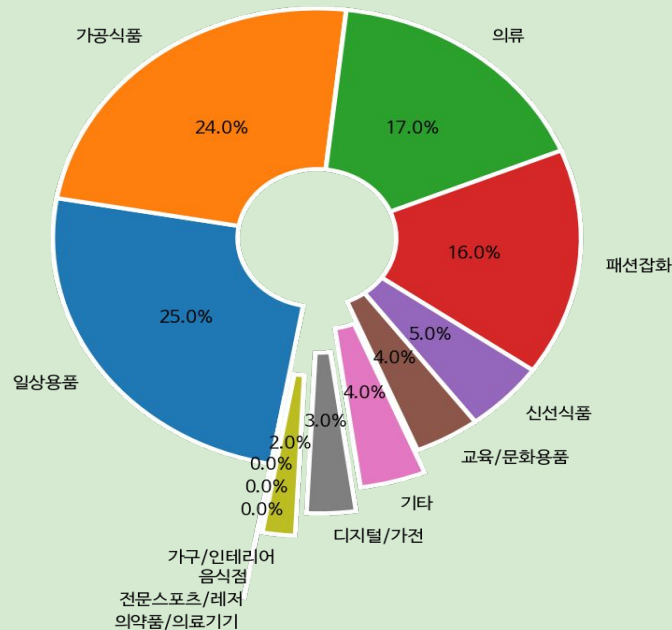
중분류 717개, 소분류 3520개로 매우 세분화 되어  
있어 파악이 어려움



분류코드	분류명	중분류명	소분류명	대분류
101 A010101	일용잡화	위생세제	일상용품	일상용품
101 A010102	일용잡화	휴지류	일상용품	일상용품
101 A010103	일용잡화	뷰티상품	일상용품	일상용품
101 A010104	일용잡화	일용잡화	일상용품	일상용품
101 A010105	일용잡화	세트상품	일상용품	일상용품
101 A010106	일용잡화	단기행사	일상용품	일상용품
102 A010201	육류	우육	신선식품	신선식품
102 A010202	육류	돈육	신선식품	신선식품
102 A010203	육류	계육	신선식품	신선식품
102 A010204	육류	갈비세트	신선식품	신선식품
102 A010205	육류	정육세트	신선식품	신선식품
102 A010206	육류	수입육	신선식품	신선식품
102 A010207	육류	양념육	신선식품	신선식품
103 A010301	축산가공	햄	신선식품	신선식품
103 A010302	축산가공	유제품	신선식품	신선식품
103 A010303	축산가공	직매입	신선식품	신선식품
103 A010304	축산가공	하가자	신선식품	신선식품

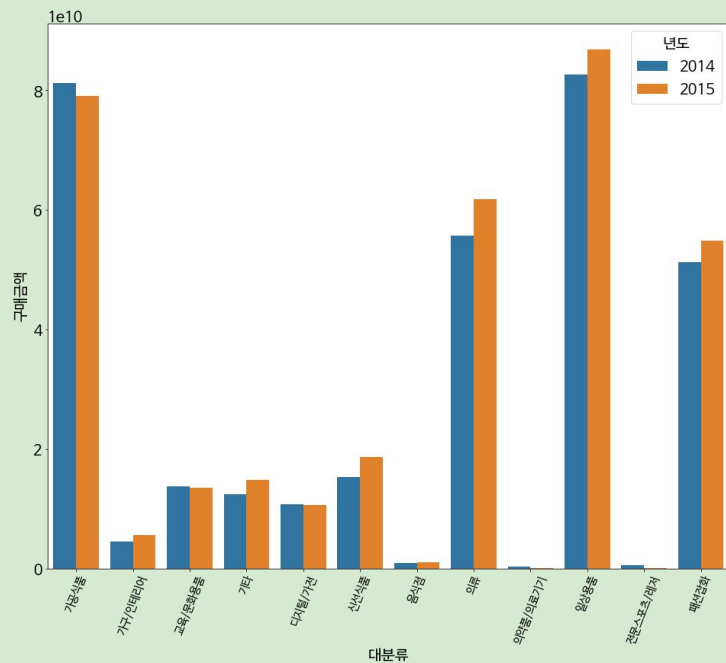
→ 유통상품 표준 DB를 기준으로 12개 대분류 분류 작업

## 06 카테고리별 매출 비중



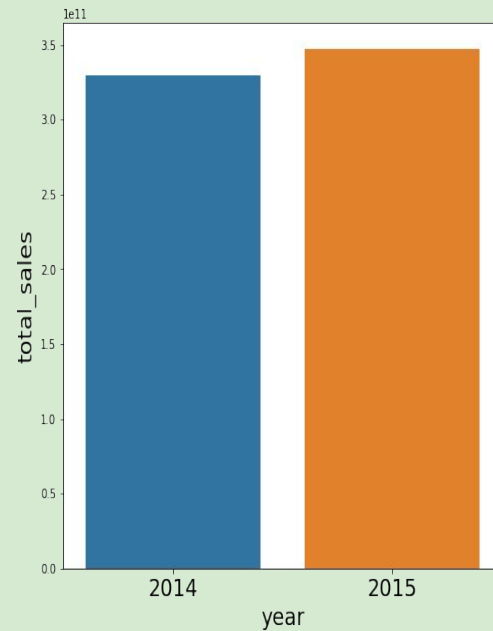
# 데이터 분석

## 07 연도별 매출 비중 확인



## 08 전체 매출

매출 5% 증가



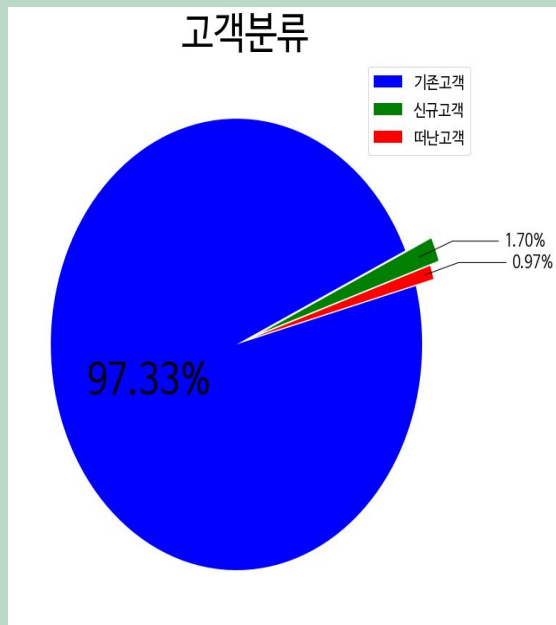


# 데이터 분석

09

## 고객 분석

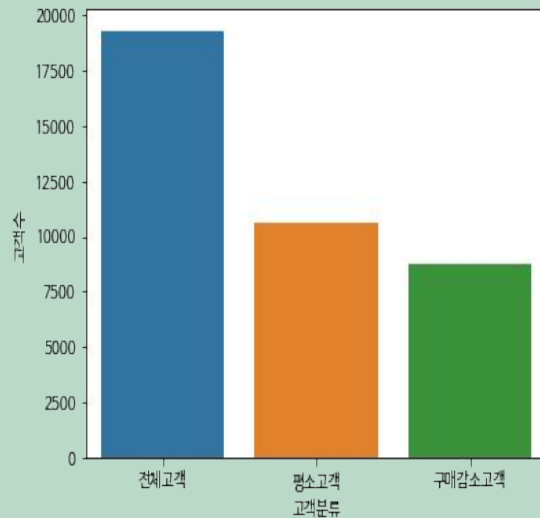
신규 점포, 신규 고객 유입은 증가



기존 고객이 전체 고객의 약98% 차지

→ 고객에 중점을 맞춤

기존 고객 45% 구매 금액 감소

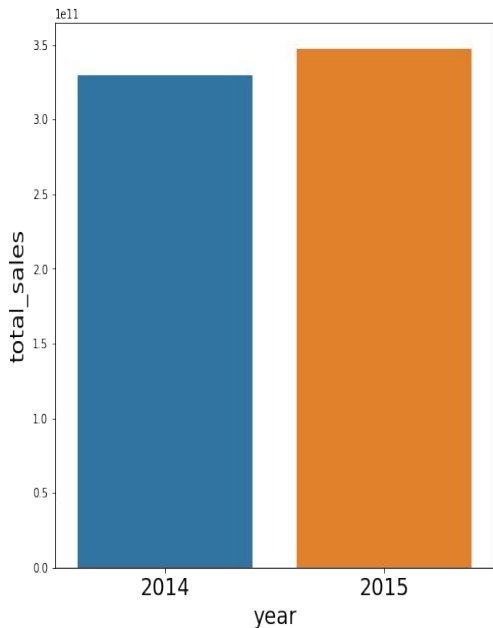


# 주제 선정

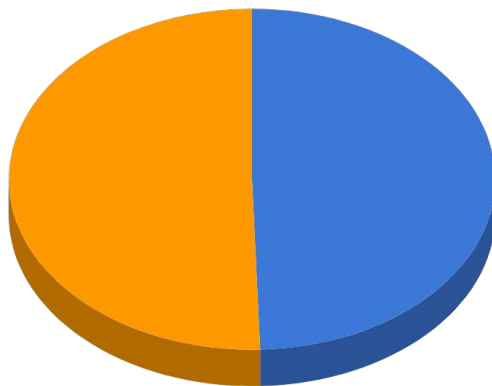
문제점 파악

## 지속고객 대상 연별 판매액

4.4% 증가



## 구매 연변화율 대비 고객 변화율 분류



미달고객

50.5%

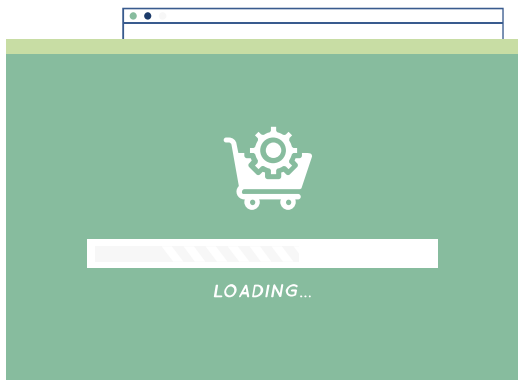
달성고객

49.5%

## 옴니채널과 O2O와 같은 트렌드로 온라인 시장 확대



## 1. 분류모델을 통한 군집화 후 마케팅 솔루션 제안



## 2. 추천시스템을 통한 매출 증대



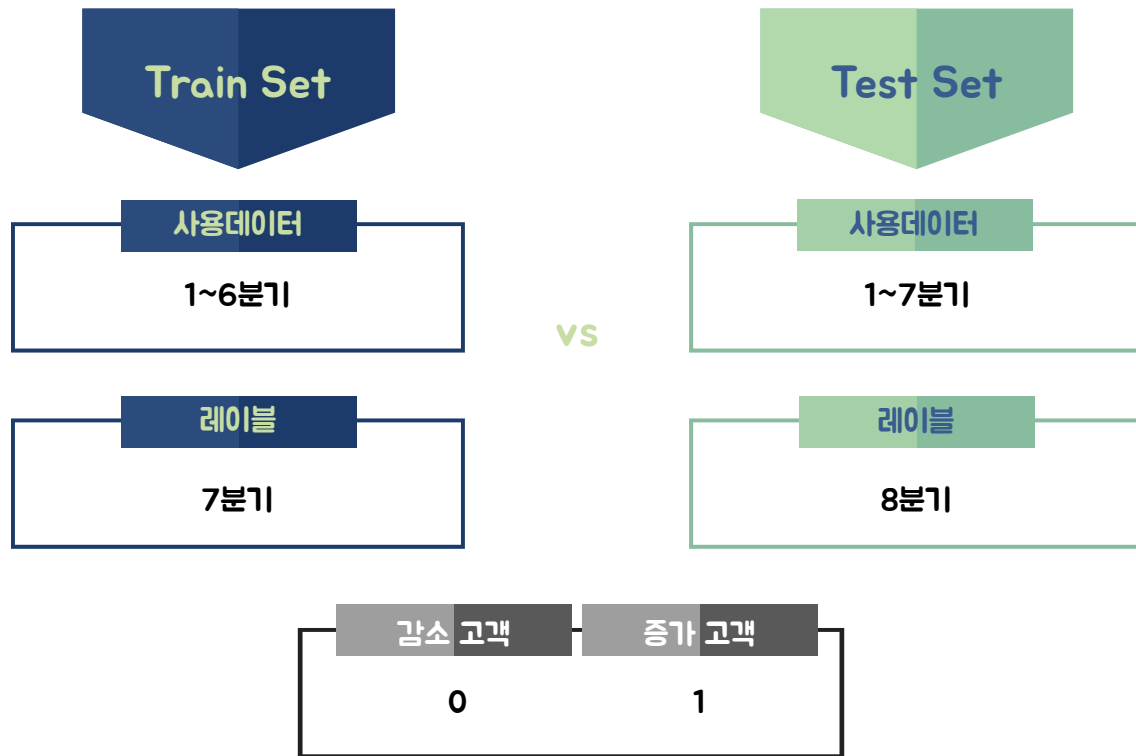


03

## 분류모델 개발 및 평가

...

# Train Set, Test Set 분리



# 변수 설정



종속 변수



속성 변수



변동 지수



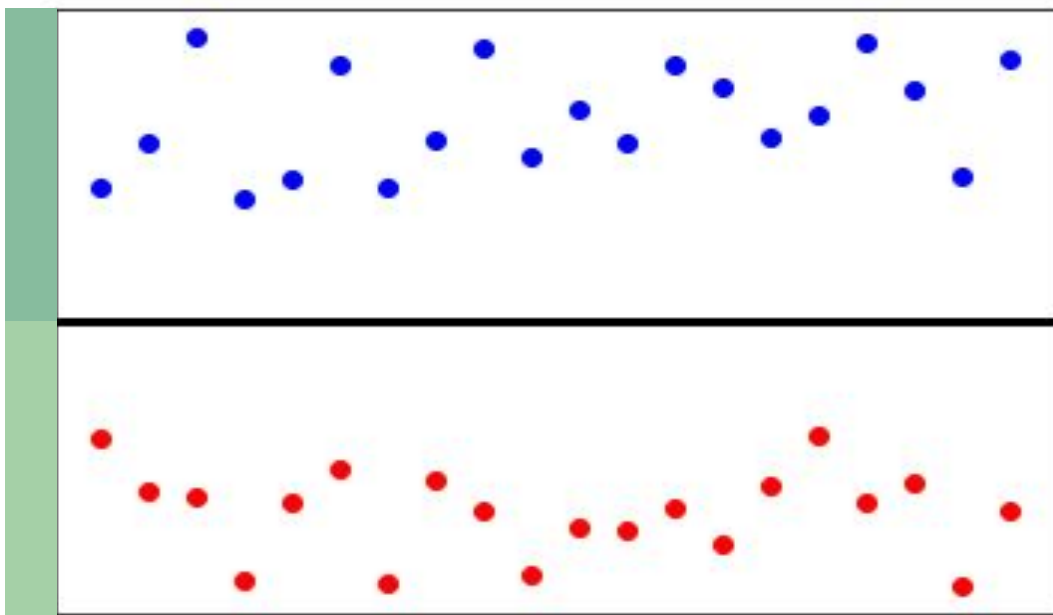
라벨

# 고객별 라벨 설정

기준

CAGR을 이용한 분기 평균 성장률

$$\text{CAGR} = \left( \frac{\text{마지막 값}}{\text{처음 값}} \right)^{\left( \frac{1}{\text{년수}} \right)} - 1$$



증가(달성) 고객 : 1

감소(미달) 고객 : 0

## 속성 변수

성별	남자 0, 여자 1
연령대	10살 단위로 범주화
거주지역	각 지역코드에 할당된 지역을 기준으로 0~17
RFM	총 구매 금액, 최근 방문 일자, 총 구매 횟수
멤버십 가입유무	회원(여러 제휴사의 회원일 경우도 포함) 1, 비회원 0
온라인 이용유무	이용 1, 비이용 0
경쟁사 이용 횟수	이용횟수를 기준으로 4등급
주말 방문 구매액 비중	주말 방문 구매액 / 총 방문 구매액으로 0~1 사이의 값
카테고리별 구매비중	4개 대분류별 구매금액 식품(가공식품, 신선식품), 패션(의류, 패션잡화), 취미(교육/문화용품, 전문스포츠/레저), 주거 (가구/인테리어, 디지털/가전)



# 동적 변수

변동지수

카테고리별 구매/방문 지수

4개 대분류별 변화하는 구매 금액 / 방문 횟수 지수와

**구매 지수**

고객의 분기별 변화하는 구매  
금액을 지수와

제휴사별 구매/방문 지수

제휴사별 변화하는 구매 금액 / 방문 횟수 지수와

금액별 방문 지수

분류별 변화하는 방문 횟수 지수와

**방문 지수**

고객의 분기별 변화하는  
방문 횟수를 지수와

총구매지수

총 구매 금액 지수와

분기 평균  
성장률

개인별 분기당 평균 성장률

예측시점 분기까지의 평균 성장률



총 구매지수로 학습한 결과가 좋지 않아서 분기별 성장률 변수를 개발한 후,  
적용 시킨 결과의 정확도가 더 높아 최종적으로 총 구매지수를 분기별 성장률로 대체

# 변수 지수화 방법

01

## 랭크화

각 분기별 총 구매금액 &  
각 분기별 총 방문 횟수

02

## 등급 나누기

고객수 19383명으로  
2000명을 기준으로  
10등급 나눔

-> 구매가 없는 고객이 대다수인 경우  
인원수대로 등급을 나누면 1,10등급만  
존재하는 문제 발생  
-> 각 분기 최대 순위를 기준으로 10등급  
분할하여 등급 조정

03

## 가중치 계산

변동폭: 등급 변동의 절대값의 합  
증감폭: 마지막 분기와 첫 분기의  
등급 차

지수 = 변동폭\*0.2 + 증감폭\*0.8

04

고객번호 A\_방문지수 B\_방문지수

1	-0.2	0.8
2	-0.6	0.4
3	0.0	0.4
4	2.0	3.8
5	-2.0	-0.6
...	...	...
19379	2.4	4.0
19380	0.0	1.0
19381	0.4	9.0
19382	1.0	9.0
19383	0.0	0.0

# 분류모델 학습 및 평가

	Accuracy	Recall	Precision	Time(sec)	Prediction
Decision Tree	0.7574	0.8057	0.7651	0.186	0.7162
Light Gradient Boosting Machine	0.7617	0.7862	0.7822	0.156	0.6957
Random Forest	0.7600	0.7840	0.7815	0.950	0.7019

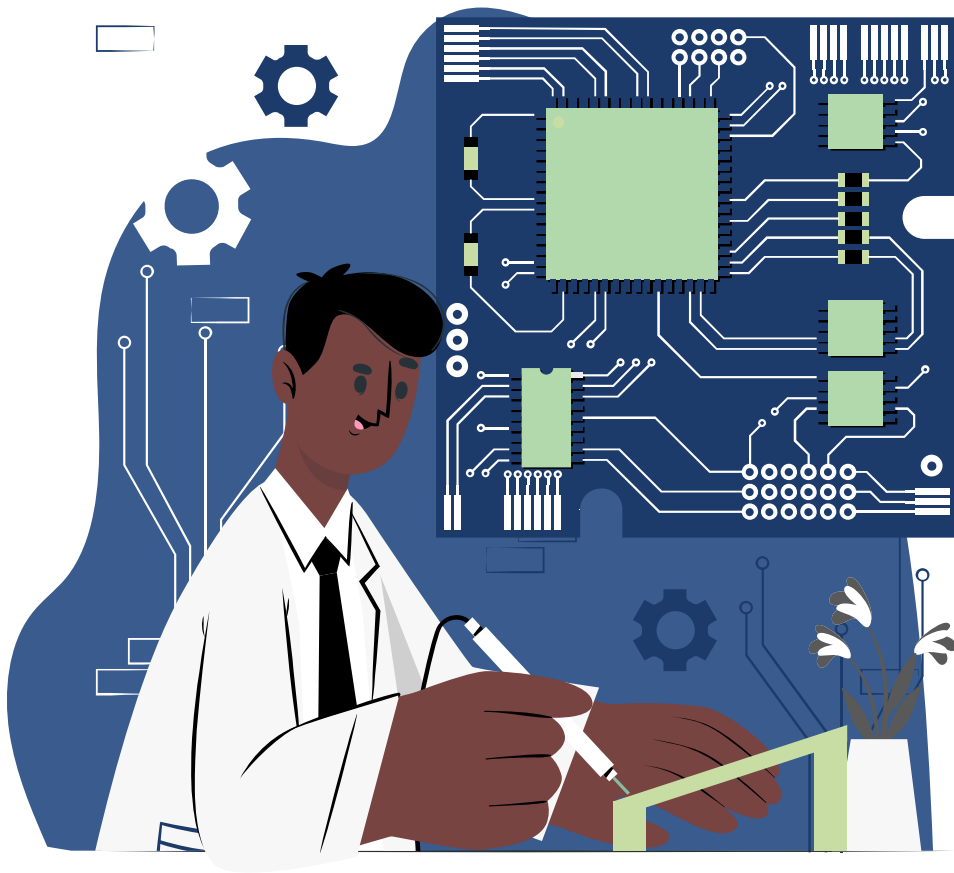
# 주요 피처 분석

LGBM	RandomForest	DecisionTree
분기평균성장률	분기평균성장률	분기평균성장률
식품구매비중	식품_구매지수	식품구매비중
구매횟수_F	고가_방문지수	패션구매비중
식품_구매지수	A_구매지수	주말구매비중
중가_방문지수	중가_방문지수	취미구매비중
고가_방문지수	저가_방문지수	주거구매비중
패션구매비중	식품_방문지수	취미_방문지수
A_구매지수	구매금액_M	주거_구매지수
구매금액_M	패션_구매지수	중가_방문지수



## 주요 피처 선정

- ▲ 분기평균성장률
- ▲ 식품구매비중
- ▲ 식품\_방문지수
- ▲ 고가\_방문지수



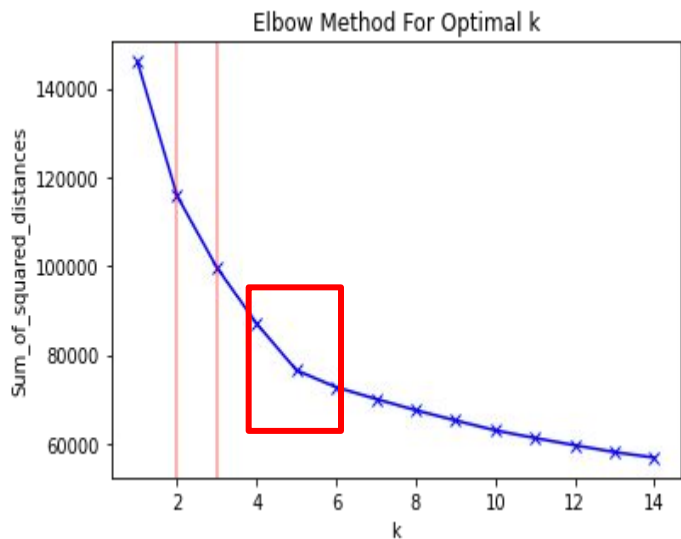
04

# 분류데이터 군집화

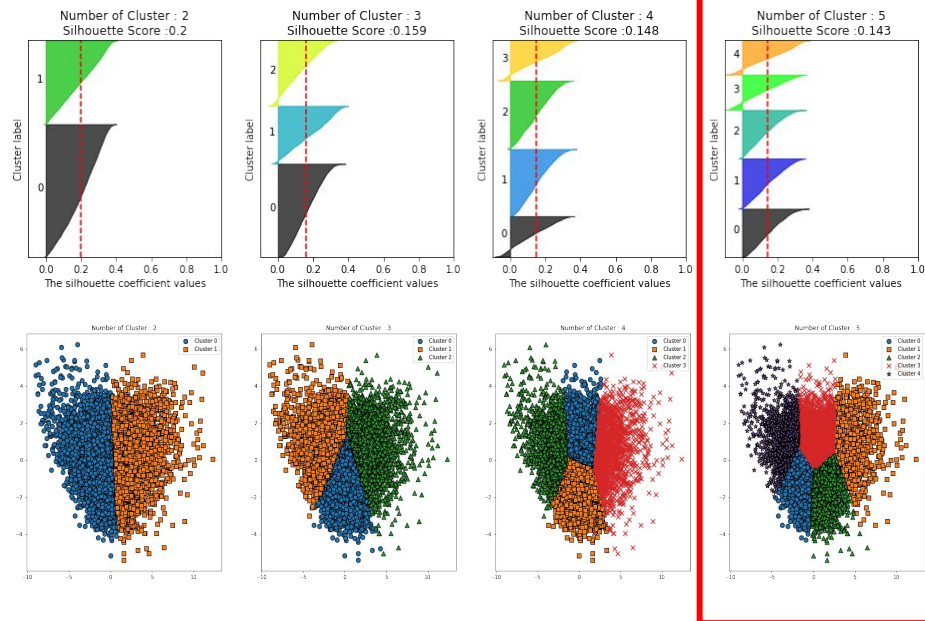
...

# 군집화

## Elbow Method



## K-Means 군집화



기울기가 급격하게 완만해지는 지점과 실루엣 계수를 확인 후  
군집 개수 결정

Silhouette Score: 0.143



# 분류 데이터 군집화

## 0번 군집



평균 구매액 39,885,807원

방문당 소비금액 76,258원



주요 제휴사: A

남자 280

여자 1278

22%

37%

26%

30대

40대

50대

저가 9%



38%



23%



24%

고가 67%

중가 22%

분기별 평균 구매 성장률: -5.4%

분기별 구매횟수 성장률: 5.5%

1558명



### 인사이트

구매금액 감소, 방문횟수 증가

구매욕구는 있으나 고가 물품에서 매력적인 상품이

없어서 구매 감소

적절한 제품을 추천하면 구매금액 증가할 가능성이 높음



### 마케팅 방향

최신 유행에 맞춰 고가 물품 최신화

고가 물품 사은품 행사, 마일리지 추가 적립 등의

이벤트로 구매촉진 유도

# 분류 데이터 군집화

## 1번 군집



평균 구매액 10,888,611원  
방문당 소비금액 28,496원



주요 제휴사: B

남자 329

여자 1671

21%

42%

24%

30대

40대

50대

저가 29%



13%



51%



22%

고가 31%

중가 38%

분기별 평균 구매 성장률: -5.8%

분기별 구매횟수 성장률: -1.7%

2000명



### 인사이트

5개 군집 중 2,30대가 가장 많이 분포

방문 횟수 가장 낮음

여성 비율 가장 높음



### 마케팅 방향

화장품 할인이나 대형마트 판촉 추천

자주 방문하지 않기 때문에 출석 이벤트

일정 금액 이상 사면 사은품 증정 프로모션



# 분류 데이터 군집화

## 2번 군집



평균 구매액 66,048,548원

방문당 소비금액 123,810원



주요 제휴사: A

남자 409

여자 1929

18%

42%

24%

30대

40대

50대

저가 6%



40%



18%



26%

고가 72%

중가 18%

분기별 평균 구매 성장률: -8.4%

분기별 구매횟수 성장률: -2.0%

2338명



### 인사이트

군집 중 가장 많은 인원 수

평균 구매 금액 가장 높음

고가 물품 구매 금액 & 횟수 가장 높음

타 군집에 비해 패션, 취미 비율 높음



### 마케팅 방향

핵심 타겟층

VIP 서비스와 비서 서비스 1년

구매 금액별로 VIP 세분화 전략

(VIP 전용 행사 주최)

개인별 취향 / 라이프 스타일에 맞춘

퍼스널 프로모션

# 분류 데이터 군집화

## 3번 군집



평균 구매액 31,955,133원  
방문당 소비금액 61,408원



주요 제휴사: A

남자 243

여자 989

27%

41%

21%

30대

40대

50대

저가 11%



28%



37%



22%

고가 64%

중가 23%

분기별 평균 구매 성장률: -17.0%

분기별 구매횟수 성장률: -12.2%

1232명



### 인사이트

모든 상품군에서 급한 쪽으로 감소

최근 방문 횟수가 가장 오래됨

마음이 떠나가는 고객

여러 제휴사 이용 고객

다른 군집 대비 낮은 연령대



### 마케팅 방향

반복적인 상품 노출로 구매 유도

인플루언서를 이용한 홍보,

행사 진행시 지속적으로 문자 sns 메일 전송

젊은 층을 타겟으로 한 팝업 스토어 오픈

# 분류 데이터 군집화

## 4번 군집



평균 구매액 16,593,448원  
방문당 소비금액 26,810원



주요 제휴사: A

남자 333

여자 1563

22%

43%

26%

30대

40대

50대

저가 25%



17%



49%



23%

고가 61%

중가 33%

분기별 평균 구매 성장률: -5.1%

분기별 구매횟수 성장률: -0.89%

1896명



### 인사이트

군집 중 C사 구매금액 가장 높음

주로 식품, 일상용품 구매

마지막 방문 일자가 가장 최근

방문횟수 가장 많음

필요한 상품을 그때그때 구매하는 고객



### 마케팅 방향

자주 방문하기 때문에 상품 프로모션에 초점

C사(슈퍼마켓) 매출이 가장 크기 때문에

마켓 식품 프로모션 위주



05

## 마케팅 솔루션 제안

...

# 마케팅 솔루션 제안

옴니채널

020

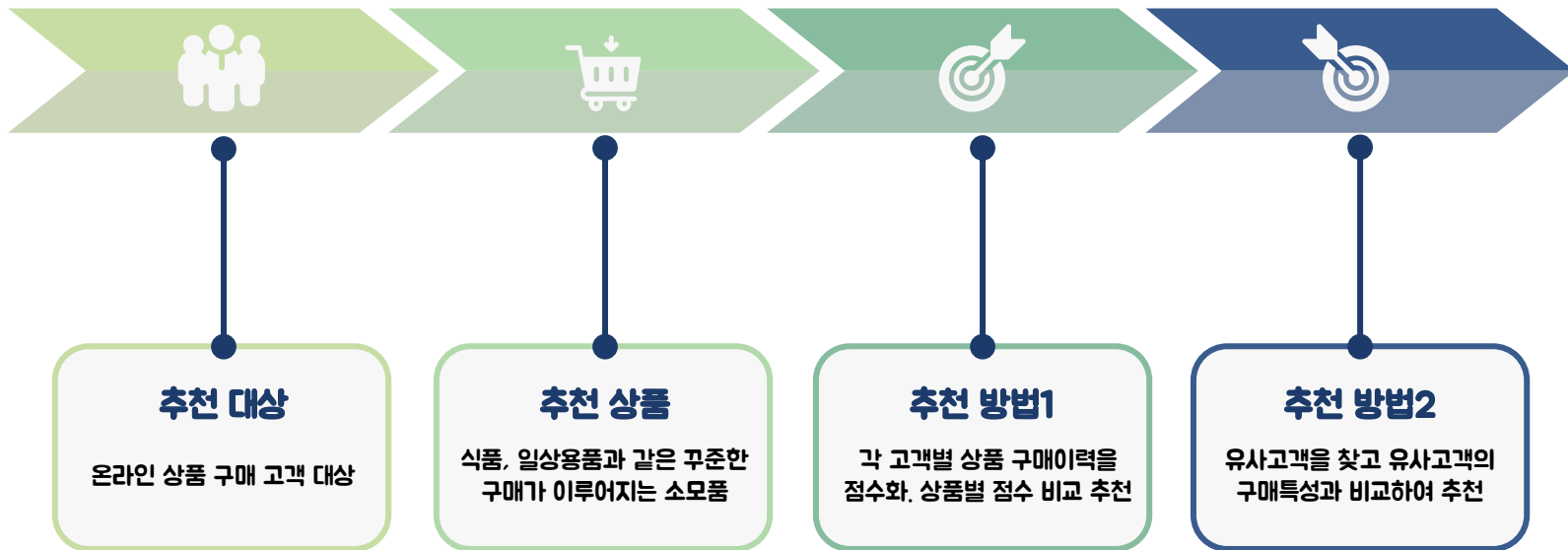
단위: 백만원

운영형태별	2010	2011	2012	2013	2014	2015
On/Offline몰	8,242,156	10,081,434	11,937,602	13,886,718	16,623,915	19,583,757
Online 몰	16,960,832	18,991,029	22,130,629	24,611,143	28,678,572	34,471,859
계	25,202,988	29,072,463	34,068,231	38,497,861	45,302,487	54,055,617

자료 : 국가 통계 포털 <온라인 쇼핑몰 운영형태별 거래액>  
(www.kosis.kr)

온라인 쇼핑의 더 큰 확장성 요구 !!

# 추천 시스템



# 추천 시스템

```
df=df.query('구매일자 > 20141231 and 구매일자 <20151232')
```

고객번호	소분류명	영수증번호
1	Bag&Bag	1
1	L.B	1
1	L/C 아웃도어	1
1	N.B	1
1	VIC마켓피자	2
...	...	...
19383	풋케어	1
19383	하드캔디	1
19383	핸드로션/크림	1
19383	헤어에센스	1
19383	혼합탄산	1

Train / Validation



Predict

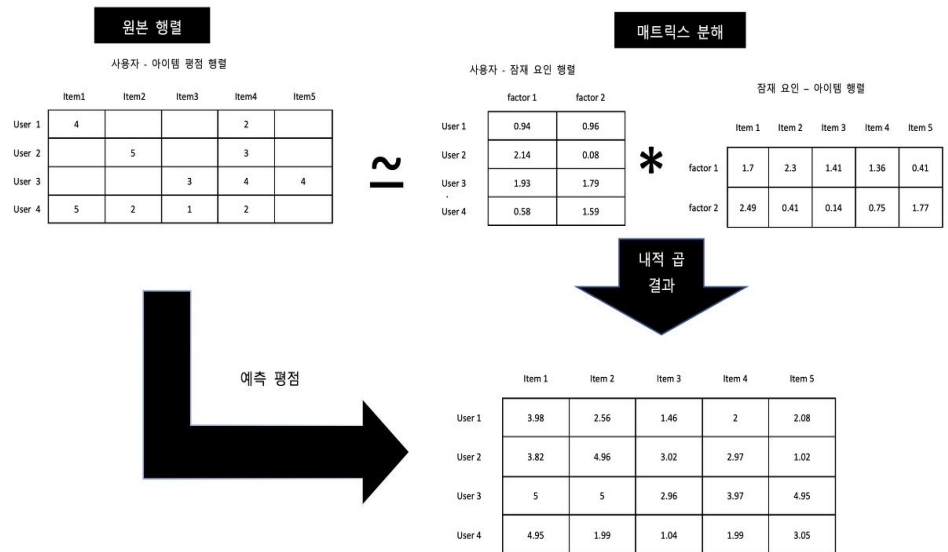


# 구매이력을 통한 추천

- 기간 내 구매이력을 이용하여 상품별 구매횟수를 1~5점으로 점수화

고객번호	소분류명	영수증번호	점수
1	Bag&Bag	1	1.00000
1	L.B	1	1.00000
1	L/C 아웃도어	1	1.00000
1	N.B	1	1.00000
1	VIC마켓피자	2	1.95925
...	...	...	...
19383	풋케어	1	1.00000
19383	하드캔디	1	1.00000
19383	핸드로션/크림	1	1.00000
19383	헤어에센스	1	1.00000
19383	혼합탄산	1	1.00000

- 잠재요인 협업필터링을 활용한 전 상품에 대한 점수 기재





# 구매이력을 통한 추천

예측 데이터와  
실제 데이터 비교



cus_id	RealOrdered	PredictedOrder(New)	PredictedOrder(Reorder)	PredictedOrder(Total)
1	{색조 화장품, 위생세제, 청과, 유제품, 제과제빵, 쿠키, 일반가공식품, 염모제,...}	[반하트디알바자, T1포맨, JDX골프, 레저스포츠 단독매입, 에이글, 와이드앵글,...]	[기초 화장품, 청과, 유제품, 채소, 수입식품, 농산가공, 위생세제, 일반가공식품...]	[기초 화장품, 반하트디알바자, 청과, 유제품, 채소, 수입식품, 농산가공, 위생세...]
2	{타월, 위생세제, 일용잡화, 청과, 즉석반찬, 기초A, 장류, 유제품, 주류, 일...}	[L/C 아웃도어, 직수입 골프의류, 색조 화장품, JDX골프, 에이글, 파우치/즉...]	[기초 화장품, 수입주방, 유기농채소, L/C골프의류, 주방잡화, 양말(특정), 기...]	[기초 화장품, L/C 아웃도어, 수입주방, 유기농채소, 직수입 골프의류, L/C골...]
7	{브랑제리, 수산단기행사, 일반계란, 일식델리, 위생세제, 청과, 식당가 일식, 유...}	[반하트디알바자, 직수입 골프의류, 유기농채소, 에이글, 레저스포츠 단독매입, 지방...]	[기초 화장품, 청과, 일식델리, 전문베이커리, 채소, 스포츠의류, 생활잡화군일가,...]	[기초 화장품, 반하트디알바자, 직수입 골프의류, 유기농채소, 에이글, 레저스포츠 ...]
8	{브랑제리, 음료, 건강식품(비타민), 아이웨어, 위생세제, 스포츠화, 청과, 즉석...}	[JDX골프, 반하트디알바자, 전문베이커리, 영플라자(기타), 휠라골프, PAT, ...]	[기초 화장품, 수입식품, 유제품, 일반가공식품, 청과, 농산가공, 위생세제, 채소...]	[기초 화장품, 수입식품, 유제품, 일반가공식품, 청과, JDX골프, 농산가공, 위...]
10	{기타레토르트, 냉장드레싱, 식초, 국산대구, 생울, 순면생리대, 무, 파스타류, ...}	[반하트디알바자, 레저스포츠 단독매입, T1포맨, 유원시설, 에이글, 유기농채소, ...]	[생활잡화군일가, 기초 화장품, 생필(부대), 커피숍, 영플라자(기타), 뷰티상품,...]	[반하트디알바자, 레저스포츠 단독매입, T1포맨, 유원시설, 에이글, 유기농채소, ...]

추천 점수가 높은 5개 상품기준

구매이력이 없는 상품	1%
구매이력이 있는 상품	48%
전체 상품	13%

# 유사고객과 비교 후 추천

## 고객특성 정보로 cosine similarity 계산

고객번호	성별	거주지역	연령대	멤버십유무	경쟁사이용	온라인유무	Recency_R	구매횟수_F	구매금액_M
0	1	0	1	6	0	3	0	1	5
1	2	0	2	6	0	3	0	2	4
2	7	1	1	6	0	1	1	2	2
3	8	0	1	6	0	2	0	4	4
4	10	1	1	6	0	0	0	3	5
...	...	...	...	...	...	...	...	...	...
8651	19307	1	2	4	0	3	0	2	3
8652	19314	1	1	4	1	0	0	1	3
8653	19325	1	12	4	0	0	0	2	3
8654	19327	0	1	5	1	0	0	1	5
8655	19333	1	0	2	1	0	1	2	4

```
from sklearn.metrics.pairwise import cosine_similarity
simil=cosine_similarity(predict2.iloc[:,1:],predict2.iloc[:,1:])
```

	0	1	2	3	4
0	1.000000	0.984415	0.915047	0.943602	0.902530
1	0.984415	1.000000	0.941606	0.968960	0.905359
2	0.915047	0.941606	1.000000	0.945835	0.897315
3	0.943602	0.968960	0.945835	1.000000	0.942809
4	0.902530	0.905359	0.897315	0.942809	1.000000

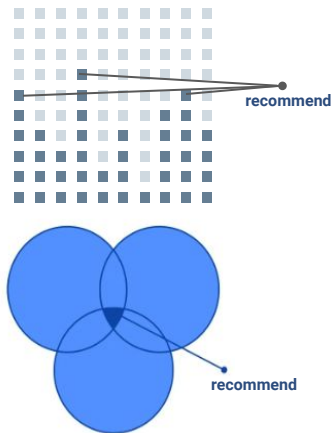
# 유사고객과 비교 후 추천



장바구니 상품은 제외!!

해당 고객의 현재까지 구매 물품의 점수 비교 후  
추천

유사 고객들의 구매이력 물품 점수 비교 후 추천



장바구니에 담은 상품 ['고추', '일반두유']

**Result :** 추천상품: {'즉석밥', '콘아이스크림', '두부류', '과일음료'}

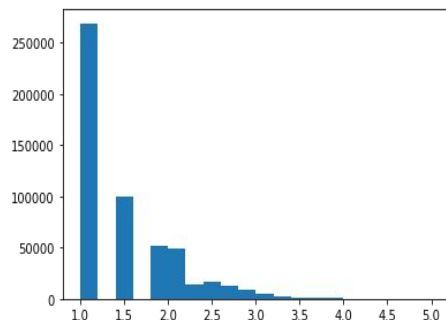
추천상품2: {'일반린스단품', '감자스낵', '일반바나나', 'NB돼지고기', '여성화장품', '일반물티슈', '일반란', '일반화장품'}

# 추천시스템 기대효과

고객번호 온라인유무 Label

1	0	0
2	0	0
7	1	0
8	0	0
10	0	0

점수 분포도



	고객번호	대분류	소분류코드	소분류명	구매일자	영수증번호	구매금액
0	7	식품	A010625	패밀리레스토랑	20140413	1733412	20000
1	7	식품	A010601	한식델리	20150529	4406132	8000
2	7	식품	A010601	한식델리	20150505	4241069	16000
3	7	식품	A010601	한식델리	20141002	2828485	6500
4	7	식품	A010602	일식델리	20140502	1852331	20000
...	...	...	...	...	...	...	...
3895435	19333	식품	A011004	수입식품	20150826	4994553	2000
3895436	19333	식품	A011004	수입식품	20151210	5661712	1000
3895437	19333	식품	A011004	수입식품	20140621	2185986	1000
3895438	19333	식품	A011001	디저트류	20140305	1489839	10000
3895439	19333	식품	D040209	기타기능성음료	20150123	224369	1500

3895440 rows × 7 columns

# 추천시스템 기대효과

고객번호 영수증번호			실제주문	추천주문5개	재현도
102392	10509	10566625	{소동물용품}	[재사용봉투, 일반스낵, 다이소, 감자스낵, 일반흰우유]	0.0
3429	769	6631966	{곡물가루}	[유기농채소, 친환경채소(특약), 떠먹는요구르트, 일반흰우유, 국산콩나물]	0.0
74040	7927	9395361	{한식}	[일반화장품, 여성화장품, 다이소, 한식, 국산맥주]	0.2
139724	14778	6764903	{한식}	[재사용봉투, 일반화장품, 다이소, 여성화장품, 일반스낵]	0.0
50321	5961	355716	{소시지}	[일반흰우유, 일반스낵, 떠먹는요구르트, 어묵, 감자스낵]	0.0
19752	2899	5773511	{위생세제}	[국산맥주, 유기농채소, 일반스낵, 영플라자(기타), 감자스낵]	0.0
1270	457	319832	{건오징어}	[일반우유, 종량제봉투, 수입식품, 생수, 두부류]	0.0
118154	12275	9470609	{일식}	[재사용봉투, 일반스낵, 일반흰우유, 다이소, 감자스낵]	0.0
52015	6097	6772423	{다이소, 가정용화장지, 식기세제}	[재사용봉투, 다이소, 일반흰우유, 일반화장품, 일반스낵]	0.2
78710	8353	315080	{단무지}	[일반우유, 두부류, 일반계란, 바나나, 생활잡화군일가]	0.0
27116	3594	5792070	{아디다스의류}	[기초 화장품, 재사용봉투, 유기농채소, 일반화장품, 여성화장품]	0.0
90949	9419	5429720	{기초 화장품}	[재사용봉투, 기초 화장품, 일반스낵, 다이소, 감자스낵]	0.2
20926	2995	173738	{마늘, 딸기, 단호박, 한우사태, 육류, 시금치, 일반메추리알}	[일반우유, 두부류, 일반계란, 바나나, 생활잡화군일가]	0.0
61776	6924	5607714	{어묵}	[기초 화장품, 전문베이커리, 일식델리, 식당가 한식, 원두커피]	0.0
9204	1665	6931793	{호주산곡물비육, 커피/초코우유, 핫사과, 어묵}	[재사용봉투, 일반흰우유, 다이소, 일반스낵, 어묵]	0.2

```
# 재현도 평균(전체 평균 재현도 6.7%)
print(simulation_test_df["top_k_recall(Reorder)"].mean())
```

0.06733341667708463

구매건마다 추천해준 결과 평균재현도가 약 6.7% 정도 나왔고 예상되는  
예상 구매물품 증가 개수는 [영수증 개수 x 추천주문5개 x 평균재현도] 입니다.

## 예상 매출공식

# GMS

=

**UV** X **CR** X **CT**

방문자

Unique Visitors

구매전환율

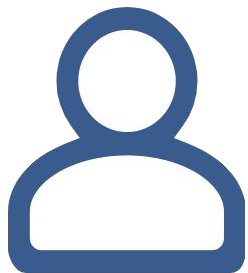
Conversion Rate

객단가

Customer Transaction



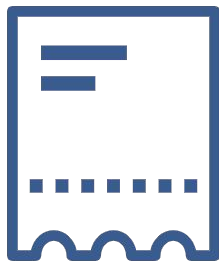
## 추천시스템 기대효과



구매감소 & 온라인 이용 고객

3014명

x



평균 결제 영수증 건수

57건

x



추천시스템 후 구매 증가율

33.5%

x



물품당 평균단가

11,565원

[UV x CR]

[CT]

예상 매출 증가액 6.7억 (총매출 대비 11%)



Q & A