

행복은 장바구니를 타고

고가영 박이삭 서재원 이다경 조상원



INDEX.

● 주제 소개

● 데이터소개

● 분석 과정

1. 고객 군집화
2. 선물 추천

● 결과



1 주제 소개

행복하자 우리~

각박한 현대인들 장바구니로 하나되자!



셜록홈즈는 어떻게 프로파일러처럼 알아볼까? 취향을 읽으면 성격과 심리가 보인다.

1. 물건

물건 선택의 취향은 분명한 자기표현
한 사람의 스타일을 구축하는 물건은
모두 그 사람 성격의 단언이자 심리의 표출

2. 생활패턴

뜻하지 않은 자기고백인 생활패턴은
의지에 따른 결과든 무의식이 시킨 행동이든
사람의 심리 체계를 반영

-출처 : 제바스티안 프리드리히, 안나 뮐러 <취향의 심리학>

1 주제 소개

행복하자 우리~

각박한 현대인들 장바구니로 하나되자!



- **행복은 장바구니를 타고**
- **제휴사, 물품구매, 시간대 별로 장바구니 유사도를 분석**
→ **고객 특성 별 그룹화**
→ **선물 추천**

2 데이터 소개

엘포인트 제공데이터 :

2년동안 약 2만명의 고객이 롯데 제휴사 4곳에서
물품을 산 구매목록

엘포인트 제공 데이터



고객 DEMO 데이터
고객에 대한 전반적인 정보



구매상품 데이터
각 고객별 구매정보



상품분류 데이터
제휴사별 상품코드와 이름

2-2 고객 DEMO 데이터

고객번호, 성별, 연령대

고객번호	성별	연령대
17218	M	60대이상
14388	M	40세~44세
15773	F	50세~54세
17829	F	25세~29세
15738	M	45세~49세

고객 DEMO 데이터

- 고객번호, 성별, 연령대 열 사용
- 고객에 대한 전반적인 정보 파악
- 비슷한 연령대로 군집화

2-3

상품분류 데이터

제휴사, 중분류 코드, 중분류명

제휴사	중분류코드	중분류명
A	101	일용잡화
A	102	육류
B	101	열매채소
C	1703	전기전자
D	803	립메이크업

상품분류 데이터

- 제휴사, 중분류코드, 중분류명 열 사용
- 제휴사별 중분류 코드를 중분류명으로 변환

2-1

구매상품 데이터

제휴사, 중분류코드, 고객번호, 구매일자, 구매시간, 구매금액

제휴사	중분류 코드	고객번호	구매일자	구매시간	구매금액
B	1504	17218	20140222	20	2420
B	1601	17218	20140222	20	1070
A	4301	14388	20140222	22	5290
C	2304	15773	20140223	10	6000
B	5202	17829	20140223	10	25000

구매상품 데이터

- 제휴사, 중분류코드, 고객번호,
구매일자, 구매시간, 구매금액 열 사용
- 고객번호로 '고객 DEMO 데이터'와 조인
- 중분류코드로 '상품분류 데이터'와 조인
- 구매일자와 구매시간으로 고객의 생활패턴 파악

3 분석 과정

고객 군집화 & 선물추천

분석과정



1. 고객 군집화

K-MEDIANS

클러스터링



2. 선물 추천

CBR

사례기반추론

3-1

고객 군집화

K-CENTROIDS

클러스터링이란?

K-CENTROIDS

- ① 주어진 데이터를 k개의 클러스터로 묶는 알고리즘
- ② 각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작

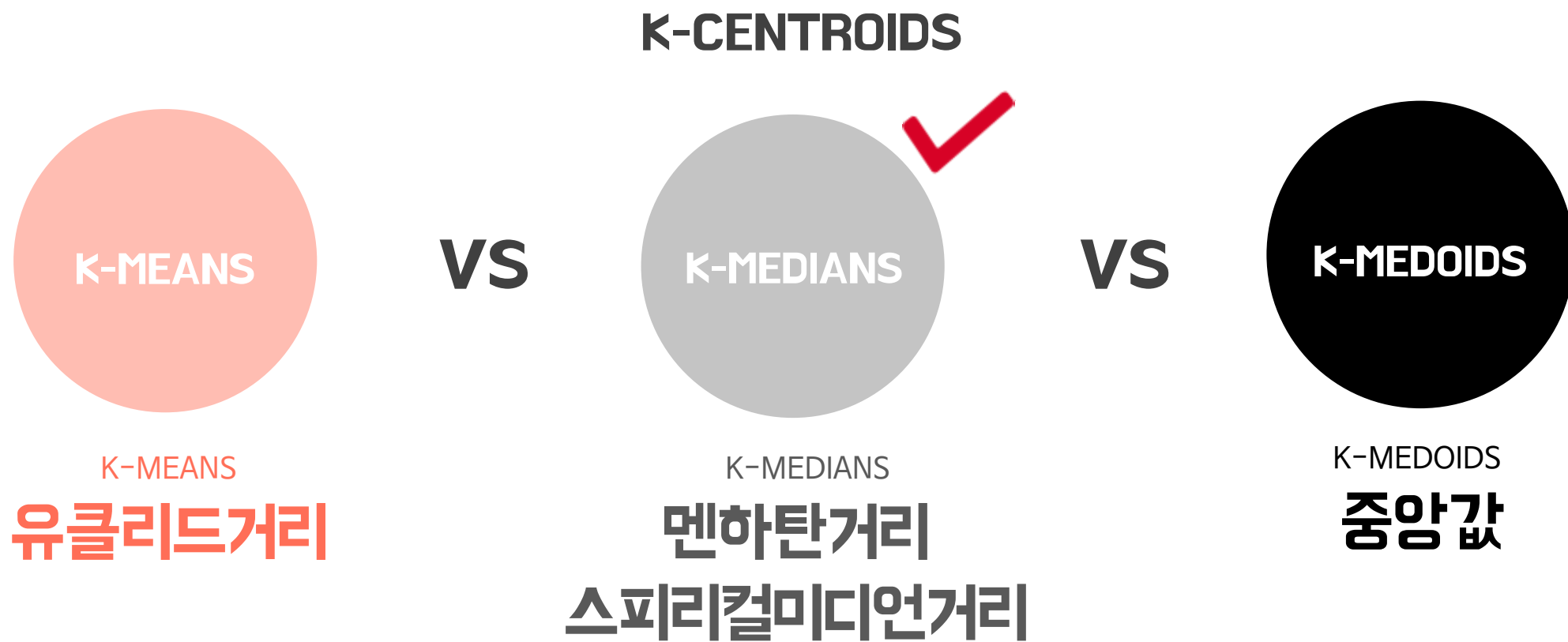
ex) K-MEANS, K-MEDIANS, K-MEDOIDS



3-1

고객 군집화

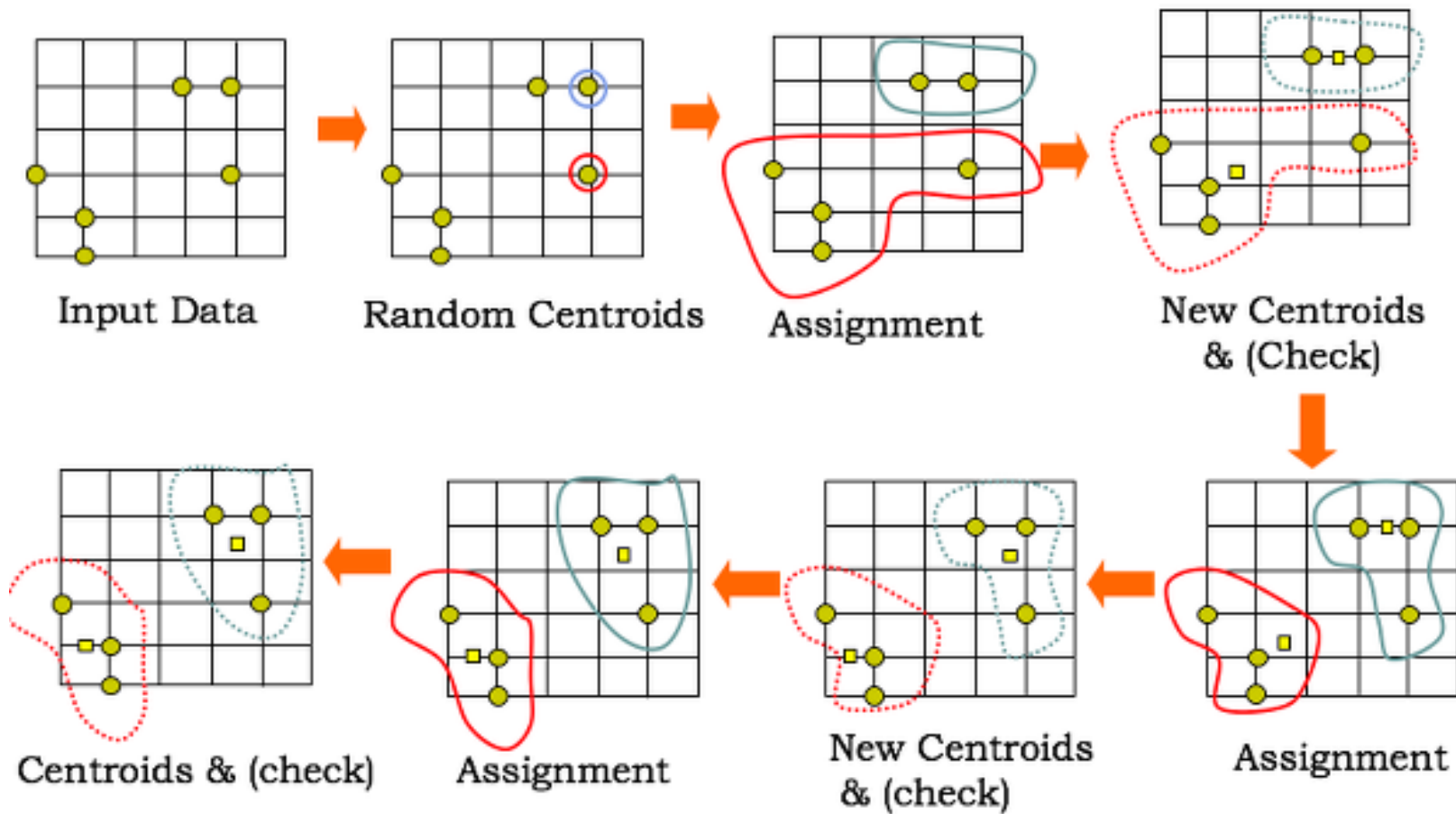
K-MEANS VS K-MEDIANS VS K-MEDOIDS



3-1

고객 군집화

K-MEANS





1. 고객 군집화

3-1 고객 군집화

전체적인 과정

고객 군집화 과정



남성용-여성용 물품
제거, 변경
비슷한 품목 하나로 묶기



고객정보, 물품구매정보
생활패턴



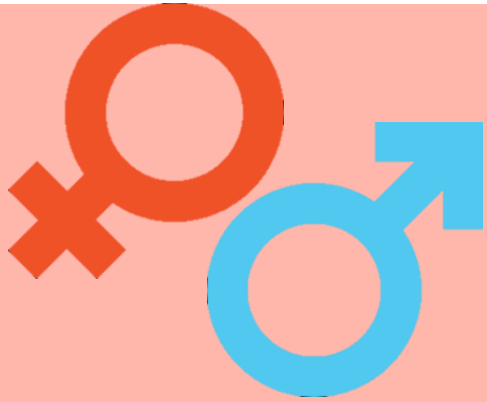
K 개수 결정
K-MEDIANS를 이용
해 클러스터링

3-1

고객 군집화

클러스터링용 데이터 구축
: 데이터 전처리

1. 남성용, 여성용 품목



남성 & 여성만을 위한 제품은 NO

클러스터링에 부적합하다고 판단하여 제거
→ 남성 브랜드, 남성 트렌디, 여성용품, 여성위생용품

2. 같은 관심사 but 다른 성별



같은 관심사지만 성별만 다른 제품 NO

클러스터링에 부적합하다고 판단하여 변경
→ 남성구두, 여성구두 : 구두
→ 남성어덜트캐주얼, 여성캐주얼브랜드,
여성이지캐주얼, 남성이지캐주얼, 남성캐주얼 : 캐주얼

3-1

고객 군집화

클러스터링 결과

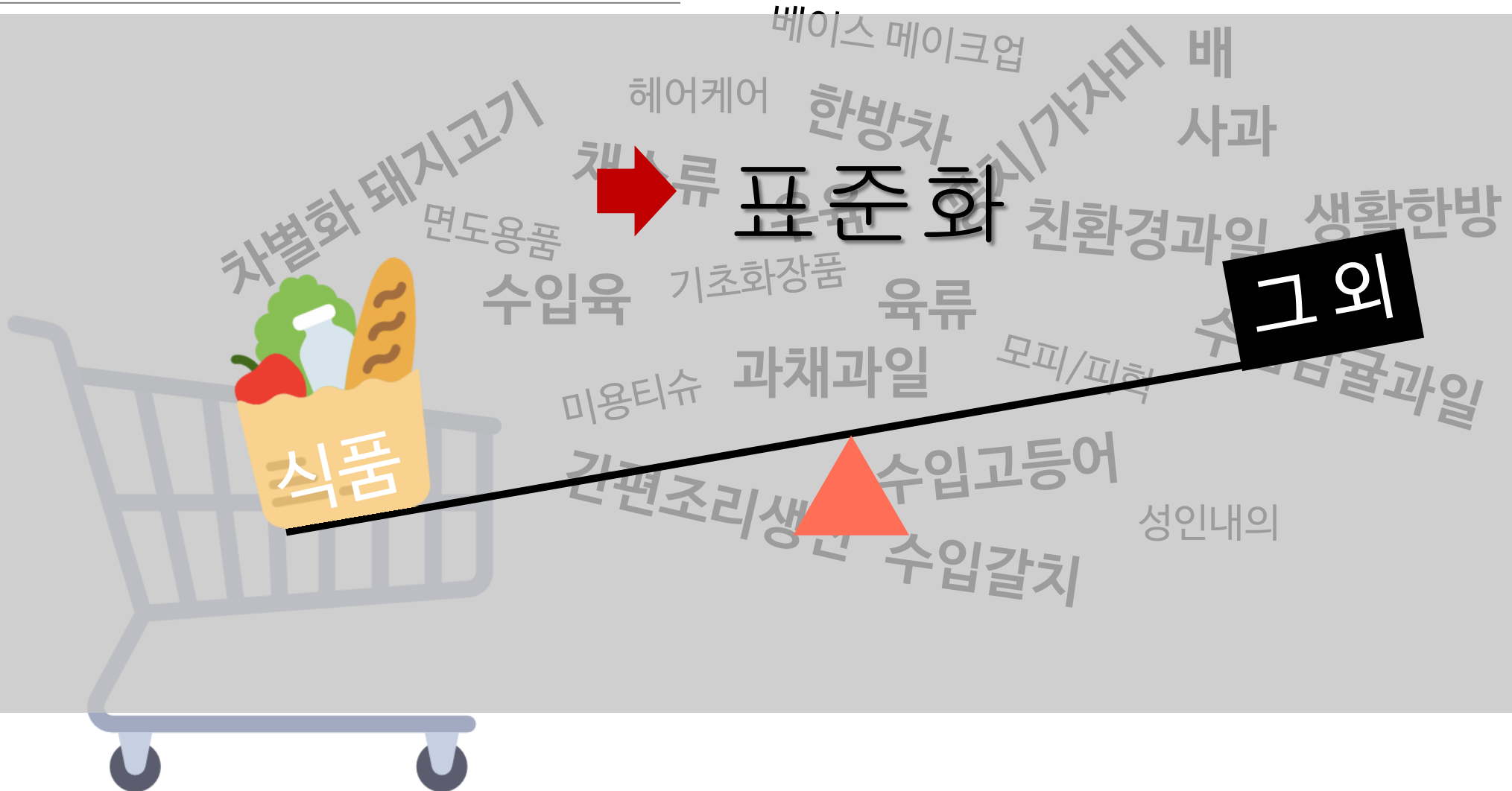


차별화 돼지고기
면도용품
채소류
우육
육류
수입육
기초화장품
과채과일
수입고등어
수입갈치
베이즈 메이크업
헤어케어
한방차
평치/가자미
배
사과
친환경과일
생활한방
수입감귤과일
모피/피혁
성인내의
미용티슈
간편조리생선

3-1

고객 군집화

클러스터링 결과



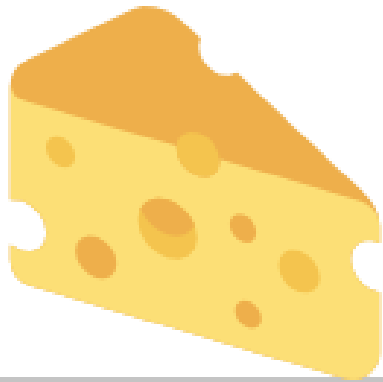
3-1

고객 군집화

데이터 전처리

B 주류

B 소주, B 맥주, B 와인, B 양주, B 전통주

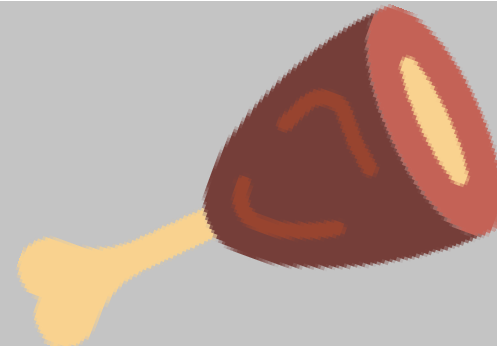


B 유가공제품류

B 흰 우유, B 우유, B 가공우유, B 치즈,
B 요구르트, B 아이스크림

C 고기류

C 국내산돈육, C 냉장축산, C 수입우육,
C 수입돈육, C 국내산계육, C 국내산우육



3-1

고객 군집화

클러스터링용 데이터 구축

: 고객번호, 물품구매횟수, 나이, 시간-요일

고객 번호	B 탄산 음료	B 소주	...	D 향수	나이	평일 점심	평일 저녁	주말
17218	62	12	...	0	35세~39세	120	3	45
17674	6	0	...	1	55세~59세	209	19	105
14388	3	3	...	26	45세~49세	203	154	162
15773	26	12	...	6	40세~44세	267	22	102
17829	51	39	...	4	60세이상	202	49	111

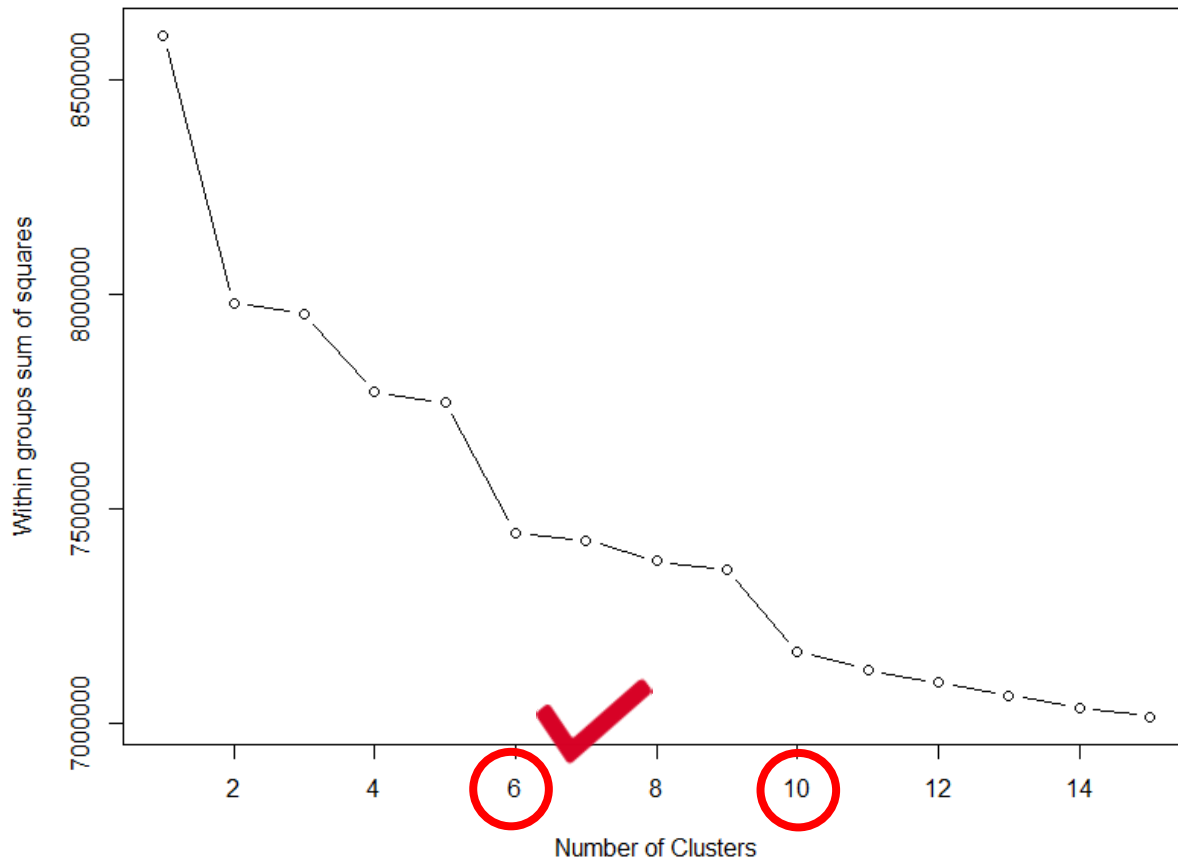
클러스터링용 데이터

- 고객번호, 물품구매횟수, 나이, 시간-요일
- 평일 VS 주말
- 9시~18시 VS 그 이후
- 고객의 주 쇼핑 시간대, 요일로 생활패턴파악
- 제휴사와 상품을 엮어, 고객의 제휴사별 취향 파악
- 비슷한 나이로 군집화

3-1

고객 군집화

K-MEDIANS



K의 개수에 따른 sum of squares

K가 10일 때까지 유의한 군집의 수가
늘어남에 따라 유의한 변화가 있는 것으로 판단

그러나,
군집의 특성이 잘 보이지 않고
군집의 비율이 불안정
→ 6개의 군집으로 선택

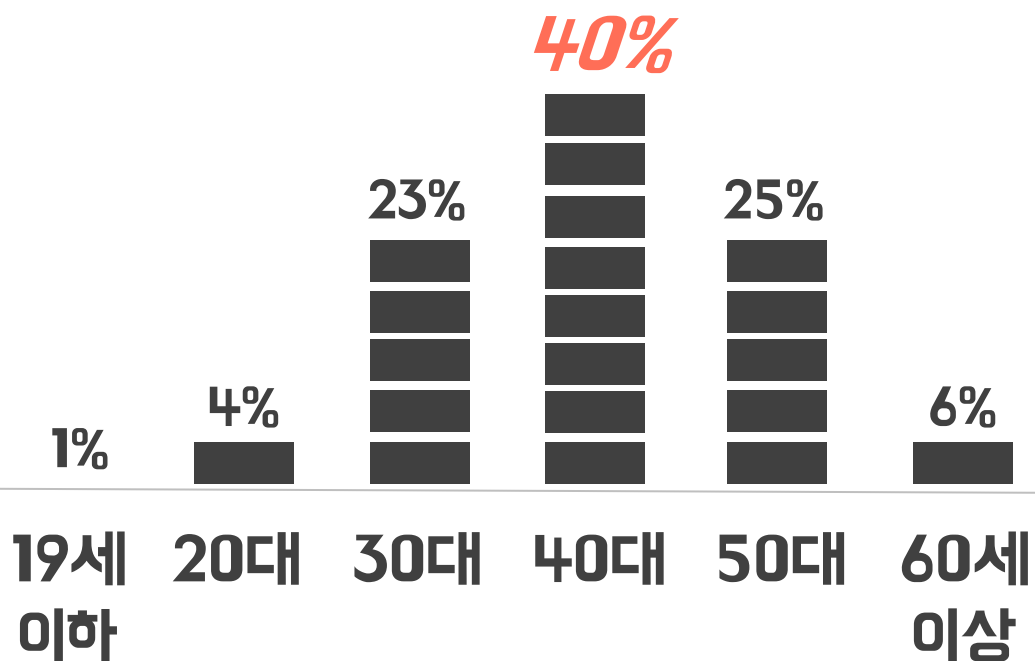
3-1

고객 군집화

클러스터링용 최종 데이터

데이터 나이대 비율

클러스터링용 데이터



남성 18%



여성 82%



총 445개의 물건 중 가장 많이 팔린 물건

1등 A 가공식품 : 전체판매량 중 **4.4%**

2등 B 채소류 : 전체판매량 중 **3.9%**

3등 B 반찬류 : 전체판매량 중 **3.8%**

4-1

고객 군집화 결과

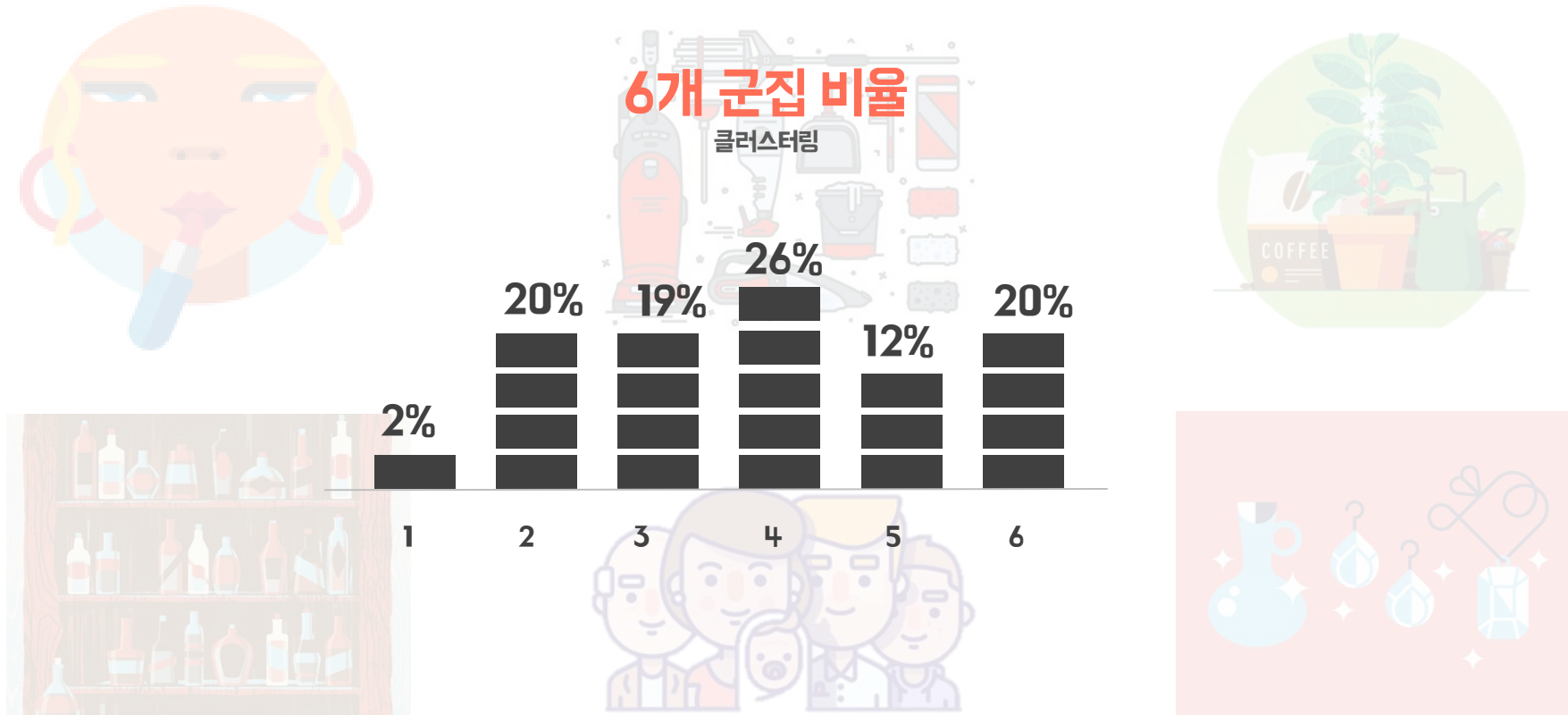
코스메틱, 청소, 차, 주류, 아기, 명품



4-1

고객 군집화 결과

코스메틱, 청소, 차, 주류, 아기, 명품





코스메틱을 좋아하는 당신!

전체 군집의 약 2%
비단결 같은 머리결
고운 피부!

상위 20개의 품목중 코스메틱 제품이 12개 이상
→ 상위 20개 판매량 중, 65% 이상 코스메틱 구매

1위	2위	3위
헤어케어	미용소품	바디케어

D

**반짝반짝 깔끔하고
주방을 사랑하는 당신!**

**전체 군집의 약 20%
각종 유리는 물론
빨래, 주방 청소까지!**

각종 유라는 물론

빨래, 주방 청소까지!

**상위 20개의 품목중 청소도구가 3개 이상
→ 8% 이상 청소도구 구입**

→ 8% 이상 청소도구 구입

11위	16위	18위
위생용품	청소목실용품	주방용품

C





티타임을 즐기는 당신!



전체 군집의 약 19%
차를 즐길 줄 알고
하루에 한 번 커피까지!

군집에 속한 이들 중 89%이상이
구매한 상위 20가지 품목에 '차/커피' 존재

1위	4위	5위
가공식품	차	커피

A

술을 사랑하는 당신! 🍺

전체 군집의 약 26%
각종 반찬과 함께
술이 빠질 수 있나요?

군집에 속한 이들 중 64% 이상이
구매한 상위 20가지 품목에 '주류' 존재

4위	8위	9위
냉장/냉동/간편식류	주류	고기류

B





사랑하는 아기를 위한는 당신!

전체 군집의 약 12%
아이들이 입는 거 먹는 것
준비물 공부까지!

상위 20개의 품목중 아이용품 제품이 15개 이상
→ 76% 이상 아이용품 구매

1위	2위	3위
파티/팬시	유아식기	교육보조재

B

블링블링 럭셔리한 당신!



전체 군집의 약 20%
옷, 화장품 뿐만 아니라
스포츠, 명품까지!

상위 20가지 품목 중 럭셔리용품이 8개 이상
→ 48% 이상 사치품 구매

9위	11위	12위
트래디셔널	액세서리	명품

A





2. 선물 추천

3-2 선물 추천

사례기반추론 CBR

사례기반추론 CBR (Case-Based Reasoning)

- ① 유사한 문제는 유사한 해법을 가진다
- ② 한번 발생한 문제는 자주 발생할 수 있다

→ 과거에 현재의 문제와 유사한 문제가 존재하고 그것이 어떻게 해결하는지를 안다면, 과거의 경험을 바탕으로 현재 문제의 해결책을 추론할 수 있다.



3-2 선물 추천

사례기반추론 CBR

사례기반추론 과정



3-2

선물 추천

선물 추천용 데이터 구축

: 가격 제한, 선물에 적합한 물건으로 제한

받는 것으로도 기쁜 선물이지만..



3-2

선물 추천

사례기반추론 CBR

① ID별 구매목록데이터 생성

id	구매1	구매2	구매3	구매4	...	구매452	구매454	구매454
1	침구	화장품	골프용품	과일	...	스포츠	준보석	화장품
2	캐주얼	캐주얼	화장품	캐주얼	...			
3	명품	주방용품	디자이너	화장품	...	섬유잡화		
...
12810	돼지고기	꿀	소고기	과일	...	건강식품	돼지고기	돼지고기
12811	속옷	명품	스포츠	스포츠	...	침구	화장품	
12812	스포츠	캐주얼	구두	캐주얼	...	캐주얼	캐주얼	캐주얼

3-2

선물 추천

사례기반추론 CBR

① ID별 구매목록데이터 생성

id	구매1	구매2	구매3	구매4	...	구매452	구매454	구매454
1	침구	화장품	골프용품	과일	...	스포츠	준보석	화장품
2	캐주얼	캐주얼	화장품	캐주얼	...			
3	명품	주방용품	디자이너	화장품	...	섬유잡화		
...
12810	돼지고기	꿀	소고기	과일	...	건강식품	돼지고기	돼지고기
12811	속옷	명품	스포츠	스포츠	...	침구		
12812	스포츠	캐주얼	구두	캐주얼	...	캐주얼		

구매회수가 24회 미만 제거: 12812명
 12812명의 사람 중 최다 구매 회수: 454회
 → 12812×454의 구매 목록 매트릭스 생성

3-2

선물 추천

사례기반추론 CBR

② 사례베이스 생성

구매 회수	발생건수	비율	누적 발생건수	누적 비율
24회	273	2.13	273	2.13
25회	240	1.87	513	4.00
...
80회	107	0.73	9676	75.52
...
454회	1	0.007	12812	100

구매회수가 24회에서 80회까지의 경우가 전체의 70%이상을 차지

→ 24번 부터 80번까지의 사례베이스 생성

3-2 선물 추천

사례기반추론 CBR
② 사례베이스 생성

id	구매1	...	구매24
1	고기	...	캐주얼
2	명품	...	화장품
3	캐주얼	...	캐주얼
...
271	간식	...	고기
272	속옷	...	캐주얼
273	침구	...	과일

〈24번 사례베이스〉

id	구매1	...	구매80
1	고기	...	캐주얼
2	명품	...	화장품
3	캐주얼	...	캐주얼
...
105	간식	...	고기
106	속옷	...	캐주얼
107	침구	...	과일

〈80번 사례베이스〉

57개의 사례베이스 생성

3-2

선물 추천

사례기반추론 CBR

③ 유사도 매트릭스 생성

	사과	배	자전거	...	헤어 오일
사과	10	10	1	...	1
배		10	1	...	1
자전거			10	...	1
...	10	...
헤어 오일				...	10

대분류코드 유사도 점수 매트릭스

예) 사과(과일) + 배(과일) = 10
사과(과일) + 헤어 오일(뷰티) = 1
헤어 오일(뷰티) + 헤어 팩(뷰티) = 10

3-2

선물 추천

사례기반추론 CBR

④ 다음 구매 예측 → 추천

id	구매목록1	구매목록2	구매목록3
1	사과	배	?
2	자전거	껌	빵
3	사과	딸기	배
4	배	사과	술
5	사과	배	오렌지
6	전기면도기	쉐이빙크림	맥주
7	사과	배	떡

〈3번 사례베이스〉

사과, 배를 구매한 1번 사람의 다음 구매는?

$$\sum_{i=1} \frac{f(N_i O_i) \times w_i}{\sum w_i}$$

$f(N_i O_i)$: 사례 N 과 O 의 i 속성 간의 유사도점수 ($=1 - \frac{a_i - b_i}{\max_i}$)

a_i : N_i 의 값

b_i : O_i 의 값,

\max_i : 사례베이스에 있는 i 번째 속성의 값들 중 최대값

〈문자형 속성의 경우〉

두 속성의 값이 완전히 일치하는 경우 : 1점 부여
그렇지 않은 경우 : 0점 부여

w_i : 속성의 가중치

3-2

선물 추천

사례기반추론 CBR

④ 다음 구매 예측 → 추천

id	구매목록1	구매목록2	구매목록3
1	사과	배	?
2	자전거	검	빵
3	사과 ✓	딸기	배
4	배	사과	술
5	사과 ✓	배 ✓	오렌지
6	전기면도기	쉐이빙크림	맥주
7	사과 ✓	배 ✓	떡

사과, 배를 구매한 1번 사람의 다음 구매는?

$$\sum_{i=1} \frac{f(N_i O_i) \times w_i}{\sum w_i}$$

2번 사람 : $(0 \times 1 + 0 \times 1) / (1 + 1) = 0$

3번 사람 : $(1 \times 10 + 0 \times 10) / (10 + 10) = 0.5$

4번 사람 : $(0 \times 1 + 0 \times 1) / (1 + 1) = 0$

5번 사람 : $(1 \times 10 + 1 \times 10) / (10 + 10) = 1$

6번 사람 : $(0 \times 1 + 0 \times 1) / (1 + 1) = 0$

7번 사람 : $(1 \times 10 + 1 \times 10) / (10 + 10) = 1$

3-2

선물 추천

사례기반추론 CBR

④ 다음 구매 예측 → 추천

id	구매목록1	구매목록2	구매목록3	점수
1	사과	배	?	
2	자전거	껌	빵	0
3	사과 ✓	딸기	배	0.5
4	배	사과	술	0
5	사과 ✓	배 ✓	오렌지	1
6	전기면도기	쉐이빙크림	맥주	0
7	사과 ✓	배 ✓	떡	1

최종점수가 같은 경우에는?

유사도 점수가 같은 오렌지와 떡

오렌지 판매 수 : 974

떡 판매 수 : 54

→ 오렌지를 최종적으로 추천

3-2

선물 추천

사례기반추론 CBR

④ 다음 구매 예측 → 추천

36번째구매를 알고 싶다면?

id	구매목록1	...	구매목록24
1	사과	...	커피
2	자전거	...	빵
3	사과	...	배
4	배	...	술
5	사과	...	오렌지
6	전기면도기	...	맥주
7	사과	...	떡

...

id	구매목록1	...	구매목록30
1	사과	...	커피
2	자전거	...	빵
3	사과	...	배
4	배	...	술
5	사과	...	오렌지
6	전기면도기	...	맥주
7	사과	...	떡

...

id	구매목록1	...	구매목록36
1	사과	...	커피
2	자전거	...	빵
3	사과	...	배
4	배	...	술
5	사과	...	오렌지
6	전기면도기	...	맥주
7	사과	...	떡

24번사례베이스추천상품:오렌지

...

30번사례베이스추천상품: 오렌지

...

36번사례베이스추천상품:커피

→ 최종 추천 상품 3개 결정

3-2

선물 추천

사례기반추론 CBR

④ 다음 구매 예측 → 추천

36번째구매를 알고 싶다면?

id	구매목록1	...	구매목록24
1	사과	...	커피
2	자전거	...	빵
3	사과	...	배
4	배	...	술
5	사과	...	오렌지
6	전기면도기	...	맥주
7	사과	...	떡

id	구매목록1	...	구매목록30
1	사과	...	커피
2	자전거	...	빵
3	사과	...	배
4	배	...	술
5	사과	...	오렌지
6	전기면도기	...	맥주
7	사과	...	떡

id	구매목록1	...	구매목록36
1	사과	...	커피
2	자전거	...	빵
3	사과	...	배
4	배	...	술
5	사과	...	오렌지
6	전기면도기	...	맥주
7	사과	...	떡

다음 구매를 예측,
→ 예측 물건을 선물로 추천

24번사례베이스추천상품:오렌지

30번사례베이스추천상품: 오렌지

36번사례베이스추천상품:커피

→ 최종 추천 상품 3개 결정

3-2

선물 추천

사례기반추론 CBR

⑤ 예측 결과

군집	추천1	추천2	추천3	구매1	구매2	구매3
1	화장품	헤어오일	매니큐어	과일	화장품	스포츠
2	캐주얼	주방용품		캐주얼	돼지고기	간식
3	커피	차	식기	캐주얼	고기	화장품
4	와인	맥주	과일	냉동식품	과일	스포츠
5	파티팬시	책		액세서리	캐주얼	캐주얼
6	속옷	명품	액세서리	명품	스포츠	속옷

전체 고객의 동시에 구매하는 평균 개수

3개

CBR 추천 목록 3개 = 최근 구매 목록 3개
→ 예측 성공

예측 결과 : 약 63%



1번 군집

1위	2위	3위
화장품	캐주얼	섬유잡화

정답률 : 47%



2번 군집

1위	2위	3위
주방용품	섬유잡화	생활잡화

정답률 : 53%



3번 군집

1위	2위	3위
과일	식기	차

정답률 : 53%



4번 군집

1위	2위	3위
과일	골프용품	견과류

정답률 : 60%



5번 군집

1위	2위	3위
액세서리	문구 팬시	파티팬시

정답률 : 64%



6번 군집

1위	2위	3위
명품	액세서리	시티웨어

정답률 : 50%

감사합니다

