



Digital Trend Analyzer

온라인 행동 기반 트렌드 예측

팀명 | 엘프
팀원 | 김미성 박정민 임지연

CONTENTS

1

데이터 탐색

2

선호지수
생성

3

수요 트렌드
예측

4

서비스 제안

5

결론

CONTENTS

1

데이터 탐색

2

선호지수
생성

3

수요 트렌드
예측

4

서비스 제안

5

결론



분석 배경 및 목표

고객들의 온라인 구매 행동을 기반으로 한
상품군별 온라인 선호지수 생성의 필요성 인식



상품군별 트렌드를 살펴보고 예측을 통해
마케팅에 활용할 수 있는 신규 서비스 제안을 하고자함

01 데이터 설명

제공 데이터 설명

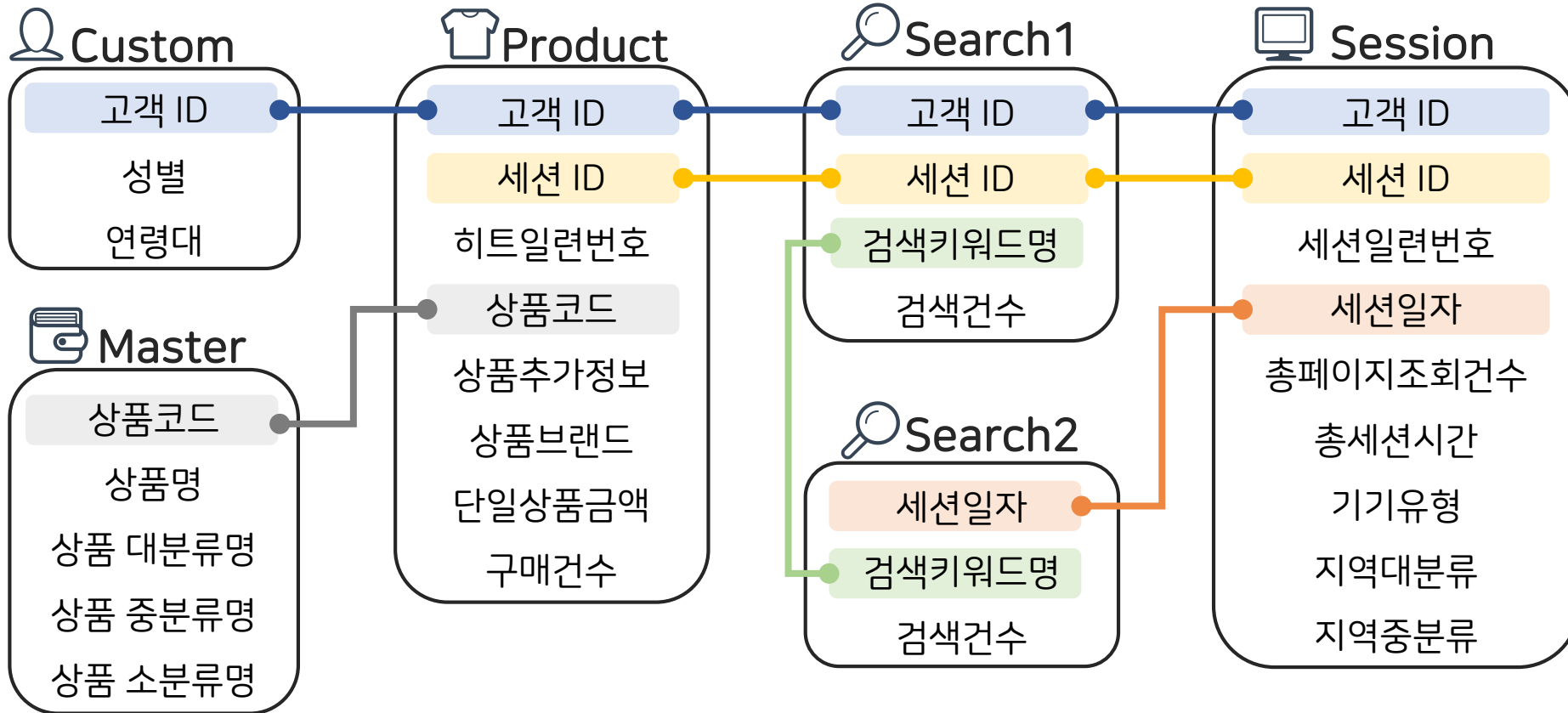
- 데이터 설명 | 2018년 4월 ~ 9월에 롯데그룹 온라인 계열사의 구매 및 이용 이력
- 제공 데이터 | 온라인 행동 데이터 및 상품분류 데이터

테이블명	내용	변수
상품구매 (Product)	상품을 구매한 방문자의 구매정보	고객 ID 세션 ID 히트일련번호 상품코드 상품추가정보 상품브랜드 단일상품금액 구매건수
검색어1 (Search1)	상품을 구매한 방문자의 검색량	고객 ID 세션 ID 검색키워드명 검색건수
검색어2 (Search2)	모든 방문자의 검색량	세션일자 검색키워드명 검색건수
회원 (Custom)	상품을 구매한 방문자의 성별/연령	고객 ID 성별 연령대
세션 (Session)	상품을 구매한 방문자의 세션정보	고객 ID 세션 ID 세션일련번호 세션일자 총페이지조회건수 총세션시간 기기유형 지역대분류 지역중분류
상품분류 (Master)	상품 분류 정보	상품코드 상품명 상품 대분류명 상품 중분류명 상품 소분류명

01 데이터 설명

데이터 셋 결합

- 구매정보가 포함된 데이터 테이블(Product)을 기준으로 하여 고객정보(Custom), 검색정보(Search1, Search2), 방문기록 정보 (Session), 구매 상품 분류정보(Master) 테이블을 결합하여 사용

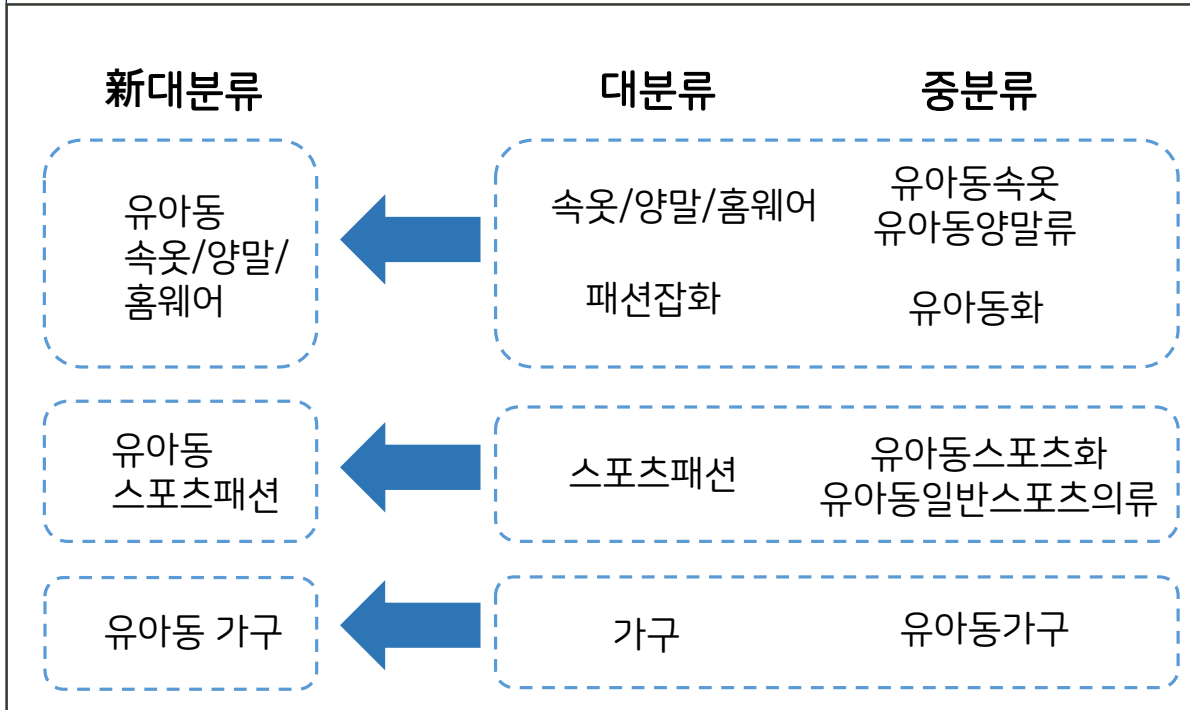


02 데이터 클렌징

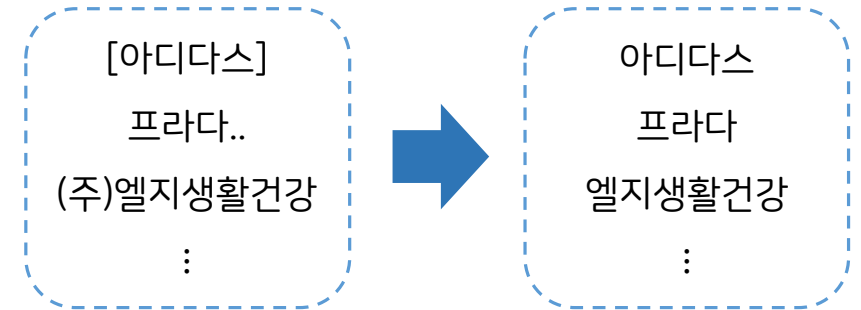
상품 재분류 & 텍스트 전처리

- 유/아동 자녀가 있는 가정은 특징이 있다고 보아 새로운 범주로 만들기 위해 재분류함
- 검색어 테이블에서 브랜드 정보를 활용하기 위해 텍스트 전처리를 수행함

기존 대분류 항목을 新대분류로 재분류



상품브랜드(PD_BRA_NM) 변수 정제



- 불필요한 특수문자 등을 제거
- 같은 브랜드임에도 분석 시에 다른 브랜드로 분류되는 것을 방지

03 파생변수 활용

특성이 다른 10가지의 새로운 대분류 생성

1. 새로운 '대분류'

- 상품 특성에 따라 기존 상품분류 테이블(Master)에 주어진 '대분류' 변수를 넓은 의미의 상품군 특성에 따라 10가지의 상품군으로 그룹화
- 통계청의 '온라인 쇼핑 동향 조사' 보도자료에 포함된 상품분류 기준

	新 대분류	기존 대분류
1	의류/ 패션	남성의류,여성의류,속옷/양말/홈웨어,스포츠패션,패션잡화
2	식품/ 농수산물	건강식품,과일,냉동식품,음료,축산물, 포장반찬
3	생활/자동차용품	식기/조리기구, 세제/위생, 인테리어/조명, 자동차용품,주방잡화, 청소/세탁/욕실용품, 침구/수예, 퍼스널케어
4	가구	가구
5	가전/컴퓨터/통신기기	계절가전, 냉장/세탁가전, 생활/주방가전, 영상/음향가전,컴퓨터, 모바일
6	화장품	화장품/뷰티케어
7	아동/유아용품	유아동의류, 출산/육아용품, 완구, 유아동가구
8	스포츠/레저용품	구기/필드스포츠, 시즌스포츠, 아웃도어/레저, 헬스/피트니스
9	사무/문구	문구/사무용품, 상품권
10	애완용품	애완용품

03 파생변수 활용

고객의 유형 및 선호지수의 요인이 되는 파생변수 생성

2. 구매 한 건당 총 금액

상품구매(Product) 테이블의 단일상품금액(PD_BUY_AM) 변수의 값과 구매개수(PD_BUY_CT) 변수의 값을 곱해 1회의 구매에 해당하는 총 구매금액을 나타내는 변수 생성

3. 페이지당 평균 뷰시간

세션정보(Session) 테이블의 $\frac{\text{총 페이지 뷰시간(TOT_PAG_HR_V)}}{\text{총 페이지수(TOT_PAG_VIEW_CT)}}$ 값을 통해 한 페이지당 평균 뷰시간을 나타내는 변수 생성

4. 제품별 세일 여부

상품분류(Master) 테이블의 상품명(PD_NM) 변수에 '할인', '특가', 'sale', '1+1', '무료배송', '시즌오프' 등의 단어가 포함된 제품을 세일 제품으로 판단하여 더미변수 생성

5. 제품별 인기 브랜드 여부

- 상품구매(Product) 테이블의 브랜드 이름(PD_BRA_NM) 변수를 상품분류(Master) 테이블에 병합한 뒤, 해당 제품이 인기 브랜드인지의 여부를 나타내는 더미변수 생성
- 각 대분류 안에서 해당 브랜드의 구매 건수 순위가 상위 10% 안에 포함되는 브랜드를 인기 브랜드로 분류

6. 상품군별 재구매율

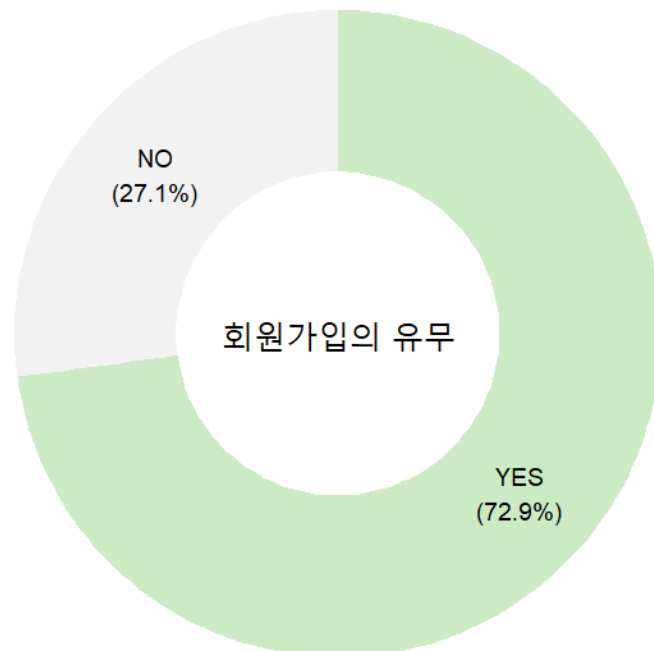
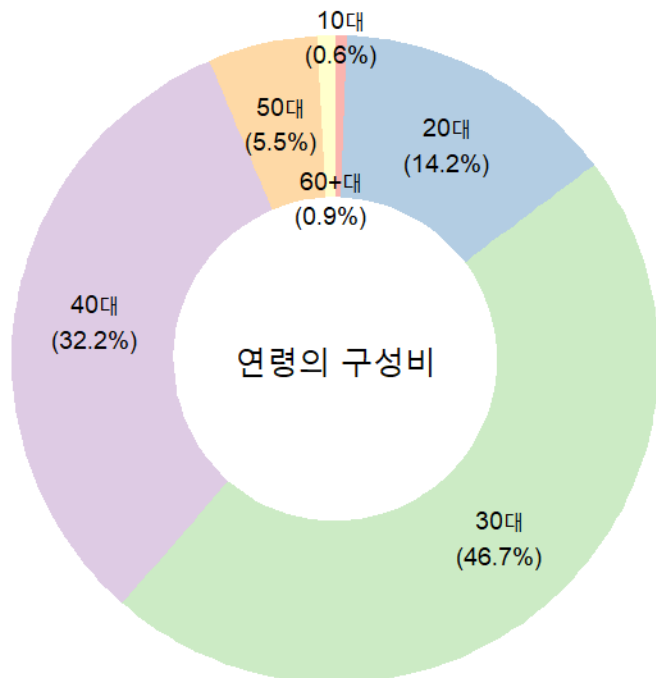
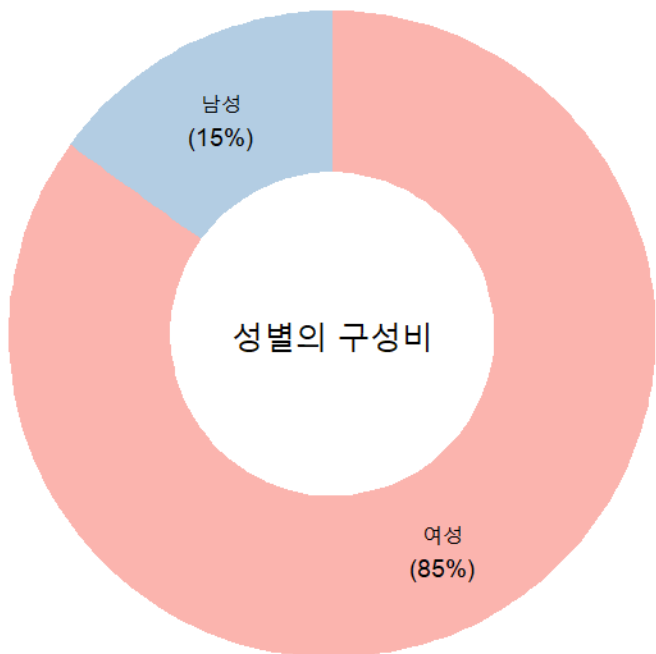
- 주어진 데이터의 기간 내 전체 구매 건수 중 동일 고객이 동일 상품군(중분류)의 제품을 재구매한 건수의 비율을 나타내는 변수 생성
- 동일한 방법으로 주별 데이터를 이용하여 상품군별 재구매율 변수 생성

7. 상품군별 검색건수

- 검색어(Search1, Search2) 테이블에서 검색 키워드를 관련 상품군(중분류)으로 분류한 후 검색건수를 나타내는 변수 생성
- 검색 키워드에 포함된 분류명, 브랜드명 등을 이용하여 분류

04 데이터 시각화

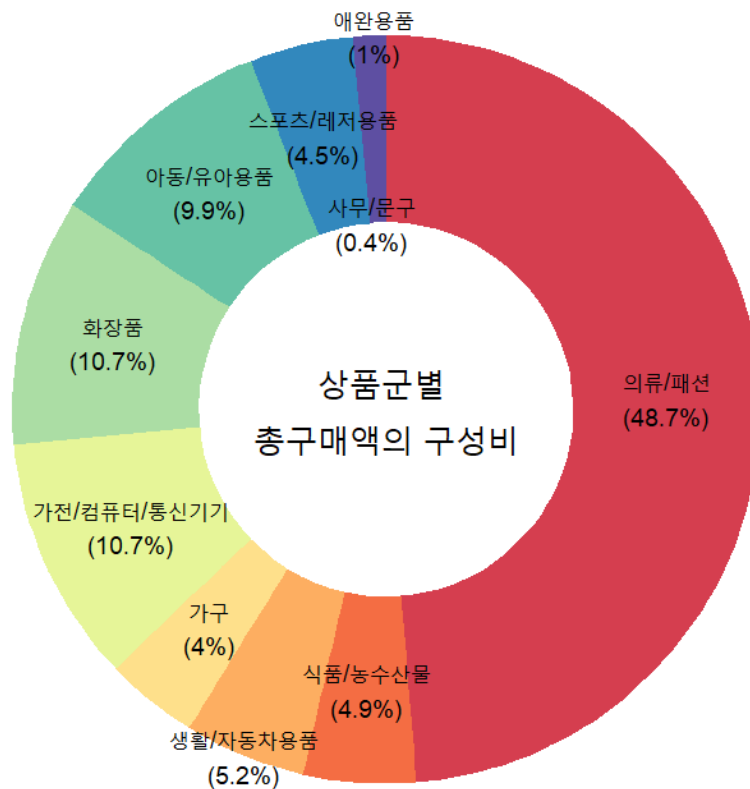
(1) 성별 / 연령 / 회원 구성비율



롯데그룹 온라인 계열사 방문자들의 기본적인 구성정보를 파악해본 결과,
주 고객층은 30 - 40 대 여성으로 전체 방문자의 75%를 차지하며
방문자의 70% 이상은 회원가입 고객

04 데이터 시각화

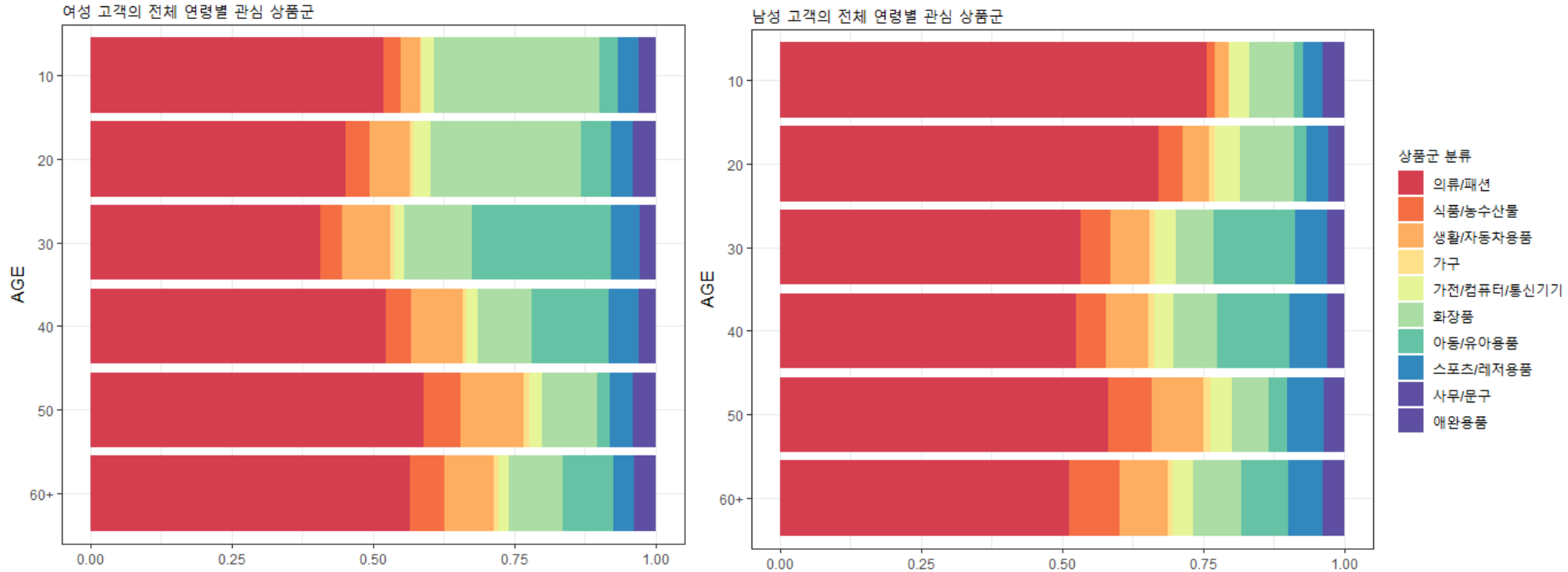
(2) 대분류 상품군별 총 구매액의 구성비



'의류/패션' 상품군이 전체 매출의 48.7%를 차지하며 가장 높은 매출을 올림
'가전/컴퓨터/통신기기'와 '화장품'이 각각 10.7%를 차지함

04 데이터 시각화

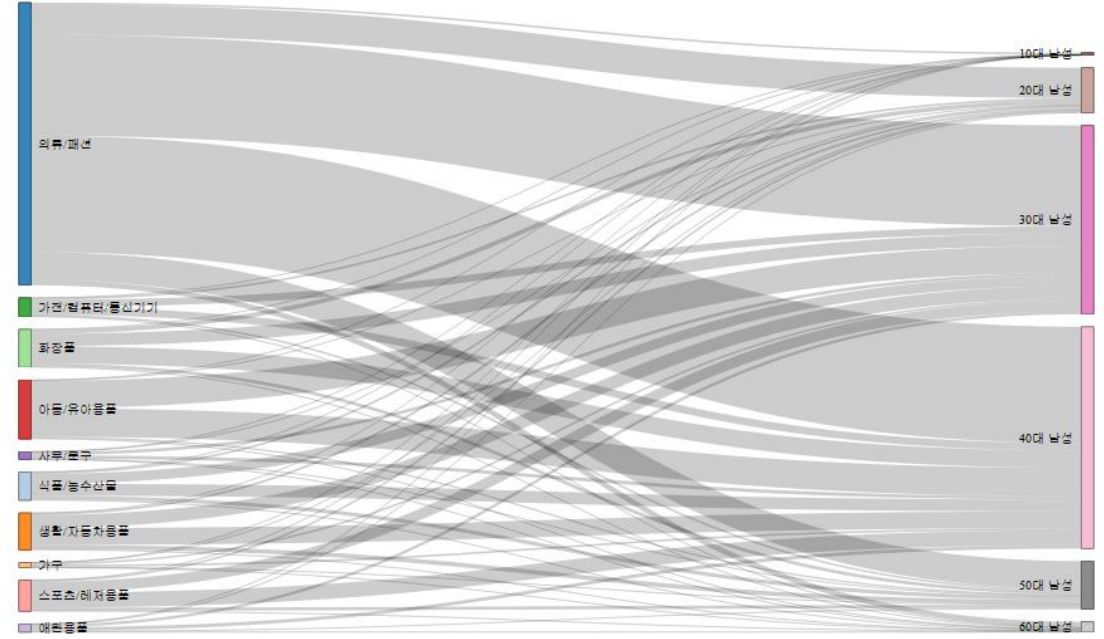
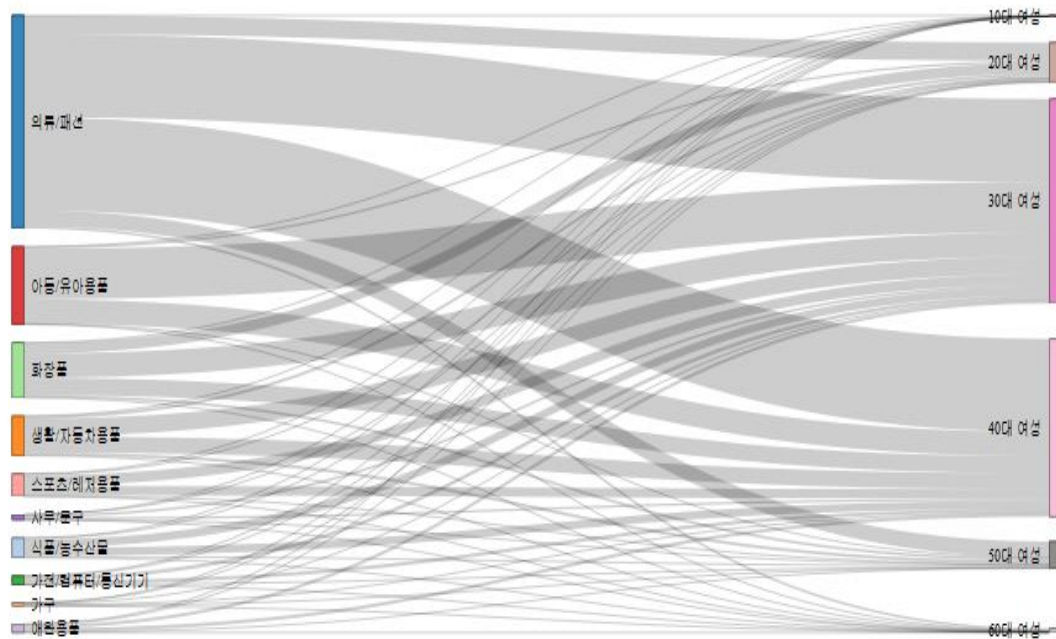
(3) 성별 / 연령별 관심 상품군



전체적으로 남성과 여성의 관심 상품에 차이가 있으며 특히 화장품 선호는 여성 > 남성
10-20대 여성은 화장품, 30-40대 여성은 아동/유아용품에 높은 관심
남성의 경우 또한 결혼 적령기인 30, 40대에 아동/유아용품에 대한 관심이 높아짐

04 데이터 시각화

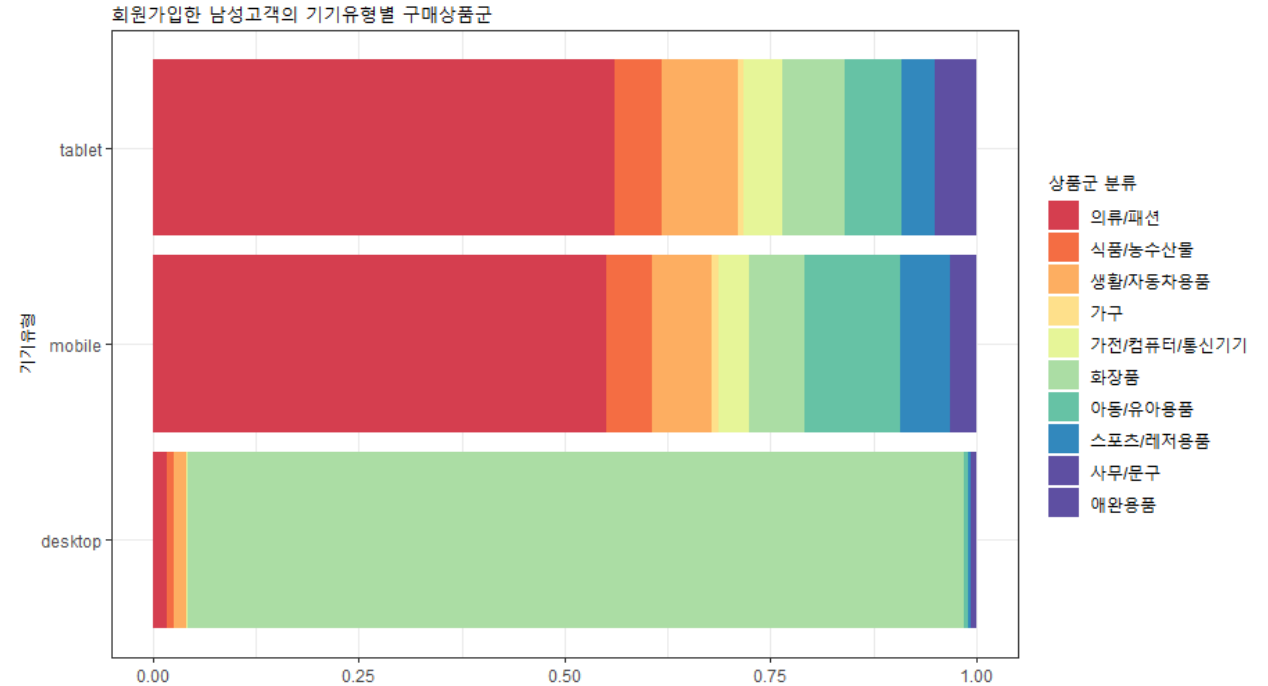
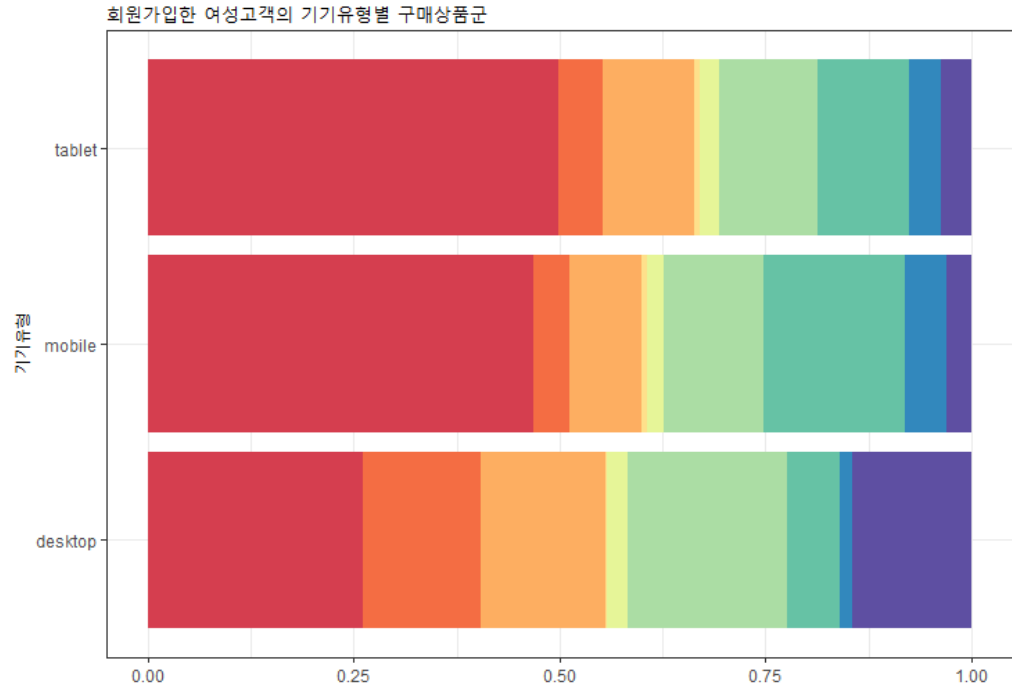
(3) 성별 / 연령별 관심 상품군



연령별 관심 상품군을 SankeyNetwork를 통해 살펴본 결과, 여성과 남성에서 모두 '의류/패션'이 독보적인 구매를 차지하는 것을 확인할 수 있고 두 성별 모두에서 30대와 40대의 비중이 매우 크게 나타남

04 데이터 시각화

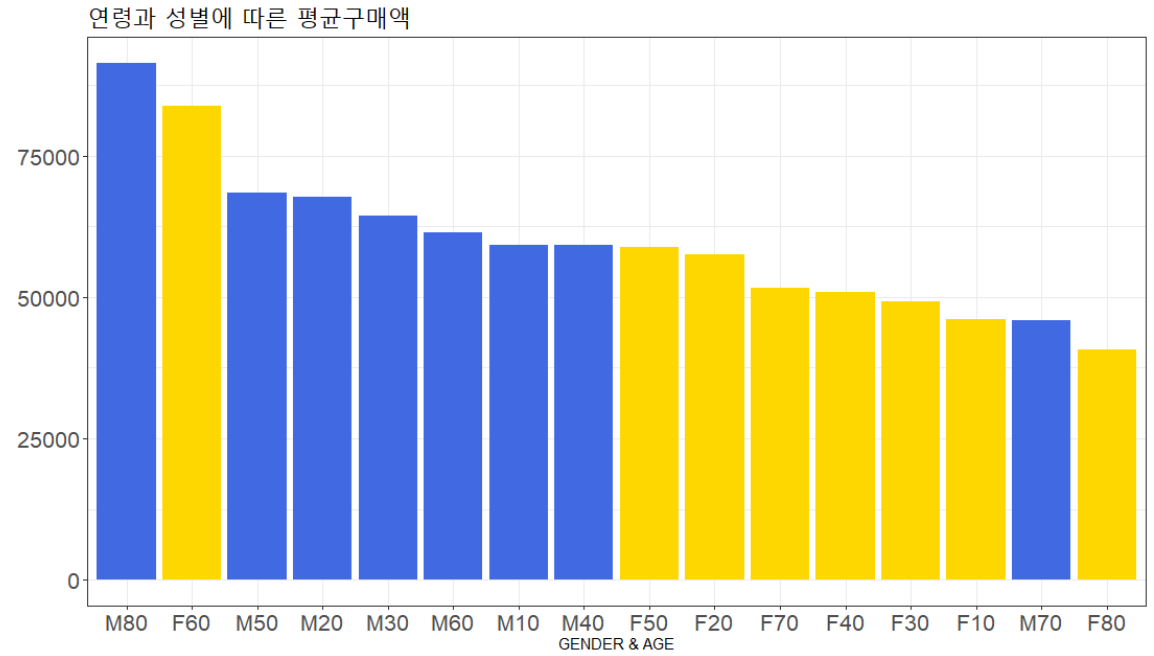
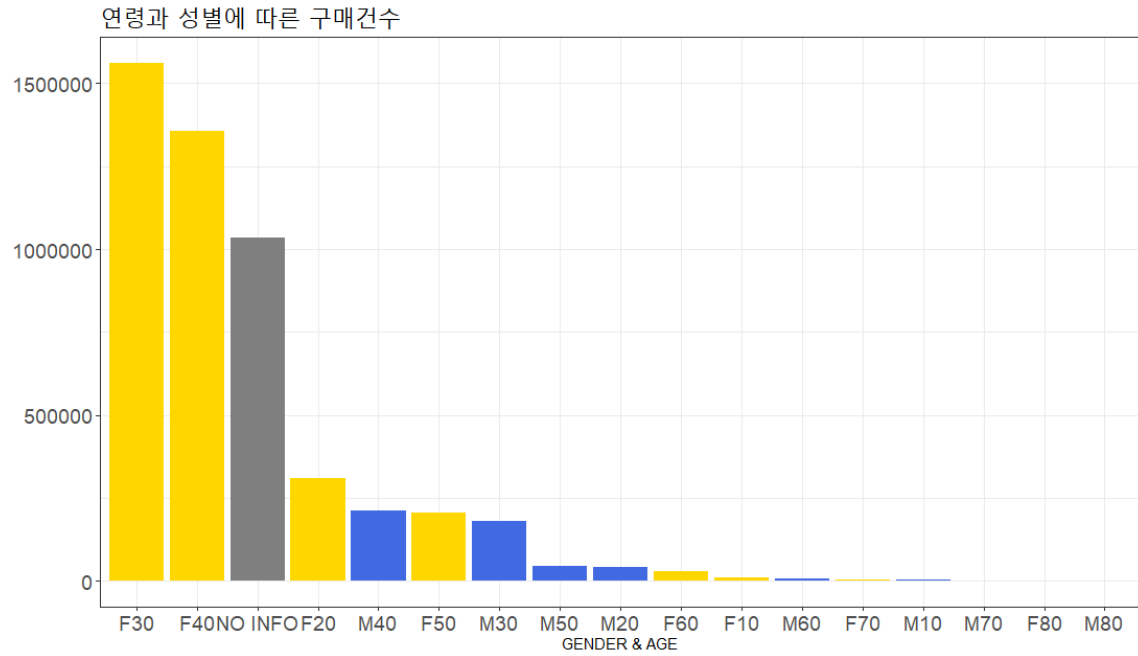
기기 유형별 관심 상품군



여성과 남성 모두 데스크탑 사용 시 구매상품군에 차이가 존재함
 특히 데스크탑을 사용하는 남성의 경우 대부분이 화장품을 구입하였으며
 연령대는 40대가 압도적으로 많았고 그 중 대부분은 향수를 구입

04 데이터 시각화

연령 & 성별에 따른 평균구매액/구매건수 순위



구매건수는 주 고객층인 30대 여성, 40대 여성과 비회원이 대부분을 차지함
80대 남성의 경우 평균구매액은 가장 많았으나 구매건수는 가장 적은 것으로 보아
80대 남성은 주로 고가의 상품을 구매하는 경향이 있음

CONTENTS

1
데이터 탐색

2
선호지수
생성

3
수요 트렌드
예측

4
서비스 제안

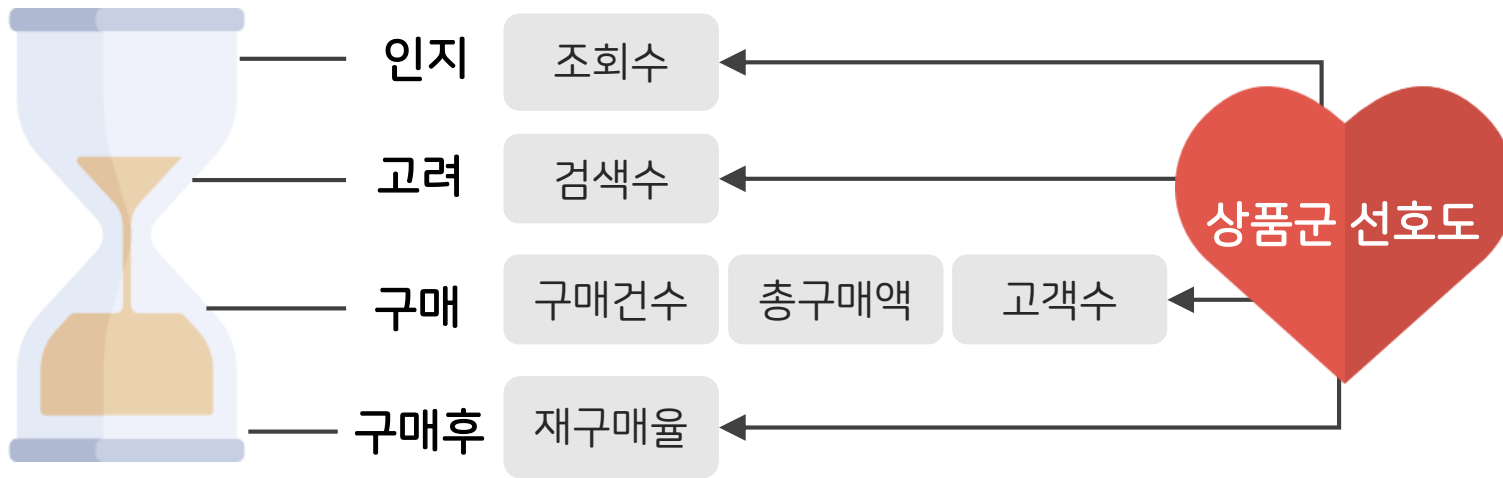
5
결론

01 상품군별 온라인 선호지수 생성

선호지수 구성 변수

- 온라인 소비자의 행동을 인지, 고려, 구매, 구매 후로 크게 네 가지로 나누어 살펴보았으며 상품군에 대한 개인의 태도를 잠재적 변수인 '선호도'로 보았음
- '조회' 하는 행동은 상품의 인지를 나타내는 행동이며 더 나아가 '검색' 하는 행동은 상품의 구매에 대한 고려를 나타내는 좀 더 적극적인 행동으로 측정
- 상품군 선호에 있어 가장 직접적인 '상품구매' 행동은 해당 상품군의 구매건수, 총구매액, 고객수로 측정하며 첫 구매 후 재구매 행동을 통해 구매 후 제품군 선호도를 측정
- 특정 변수들간의 관계를 통해 잠재변수를 찾아주는 요인분석(Factor Analysis)을 수행해 선호지수 생성

Q. 만약 고객이 어떤 상품군을 '선호'한다면 어떤 행동을 할까?



Q. 요인분석(FACTOR ANALYSIS)이란?

- '잠재변수'를 찾을 때 주로 사용하며, 변수들 간의 상관관계를 바탕으로 차원을 축소
- 변수 내부에 존재하는 구조를 발견

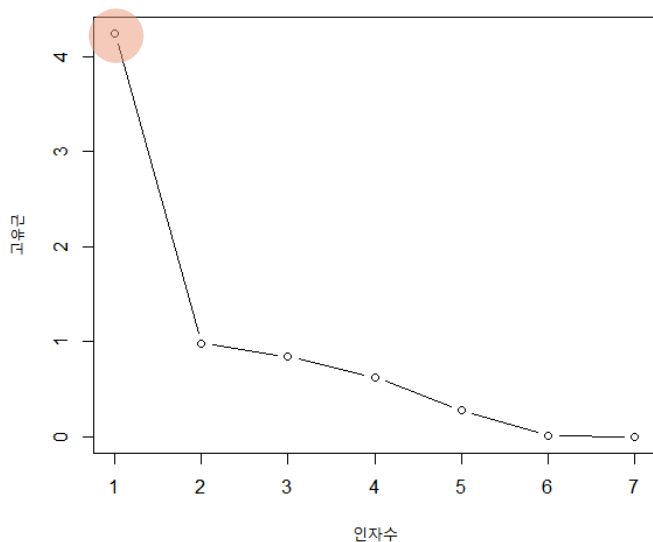


- '잠재변수'를 상품군선호도라고 가정하여 온라인 소비자의 행동을 통해 측정

01 상품군별 온라인 선호지수 생성

선호지수 구성 방법 - Factor Analysis

Q. 적절한 요인 수는 ?



요인수가 1인 경우에만 고유값의 값이 1 이상이므로 하나의 요인으로 설명됨

Q. 변수간 상관계수는 ?

	hit	count	total	client	search1	search2	rebuy
hit	1	0.23	0.16	0.22	0.17	0.19	0.02
count		1	0.76	0.98	0.68	0.69	0.25
total			1	0.8	0.64	0.63	0.27
client				1	0.73	0.73	0.29
search1					1	0.99	0.25
search2						1	0.24
rebuy							1

변수간 상관계수는 모두 양의 값을 가지며 고려, 구매 관련 변수간 특히 높은 값을 가짐

Q. 최종 선호지수 가중치는 ?

변수명	가중치
조회수	0.221
구매자검색수	0.743
방문자검색수	0.742
구매건수	0.982
총구매액	0.798
고객수	0.998
재구매율	0.289

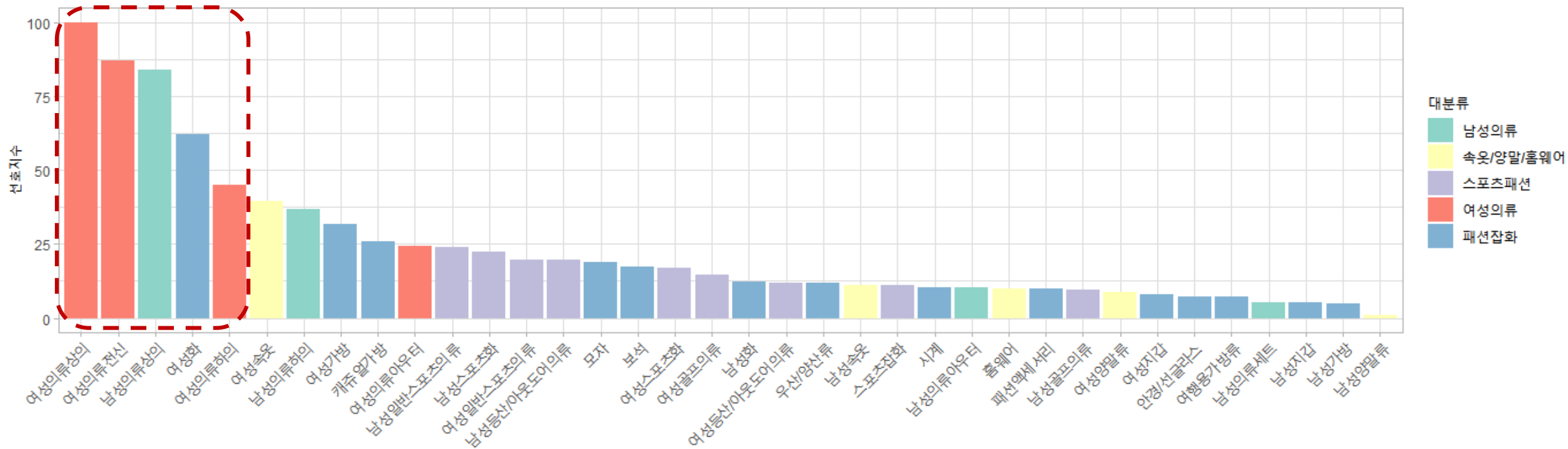
요인분석을 통해 최종 산출된 변수들의 가중치를 통해 선호지수 계산 후,

$$\frac{x - \min(x)}{\max(x) - \min(x)} * 100 \text{ 으로 변수변환}$$

02 상품군별 온라인 선호지수 생성

(1) 대분류별 선호지수 비교

1. 패션/ 의류

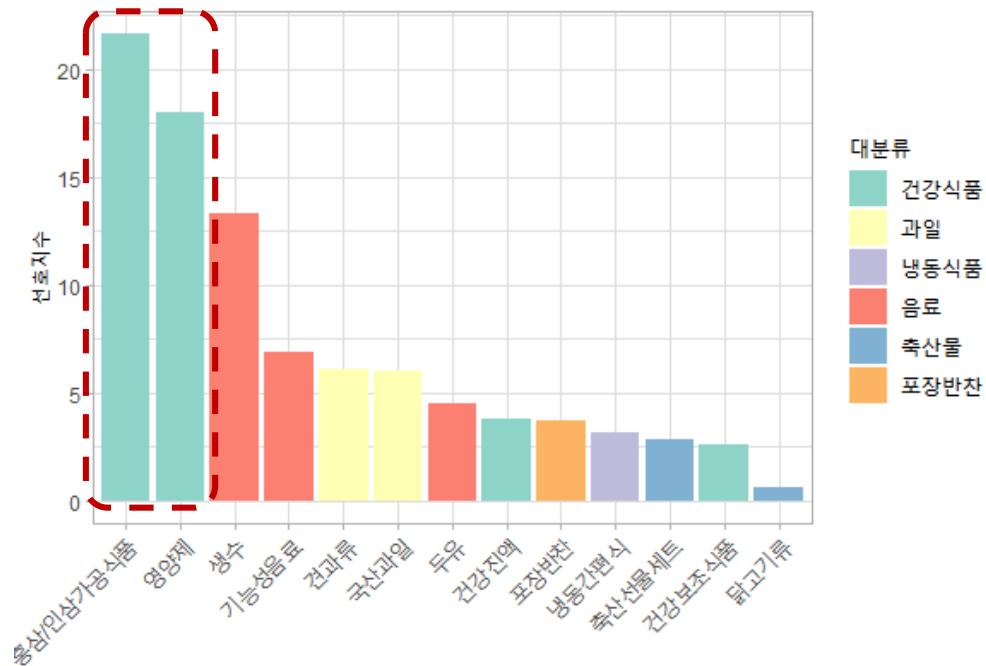


- 여성리류의 선호지수가 다른 항목보다 압도적으로 높은 점수를 나타내며, 가장 선호하는 항목이라는 것을 알 수 있음
-> 남성 고객에 비해 여성 고객 수가 많기 때문이라고 예상됨
- 남성리류상의에 해당하는 '남성티셔츠'의 구매건수가 압도적으로 높았기 때문에 선호지수가 높게 나타난 것으로 알 수 있음

02 상품군별 온라인 선호지수 생성

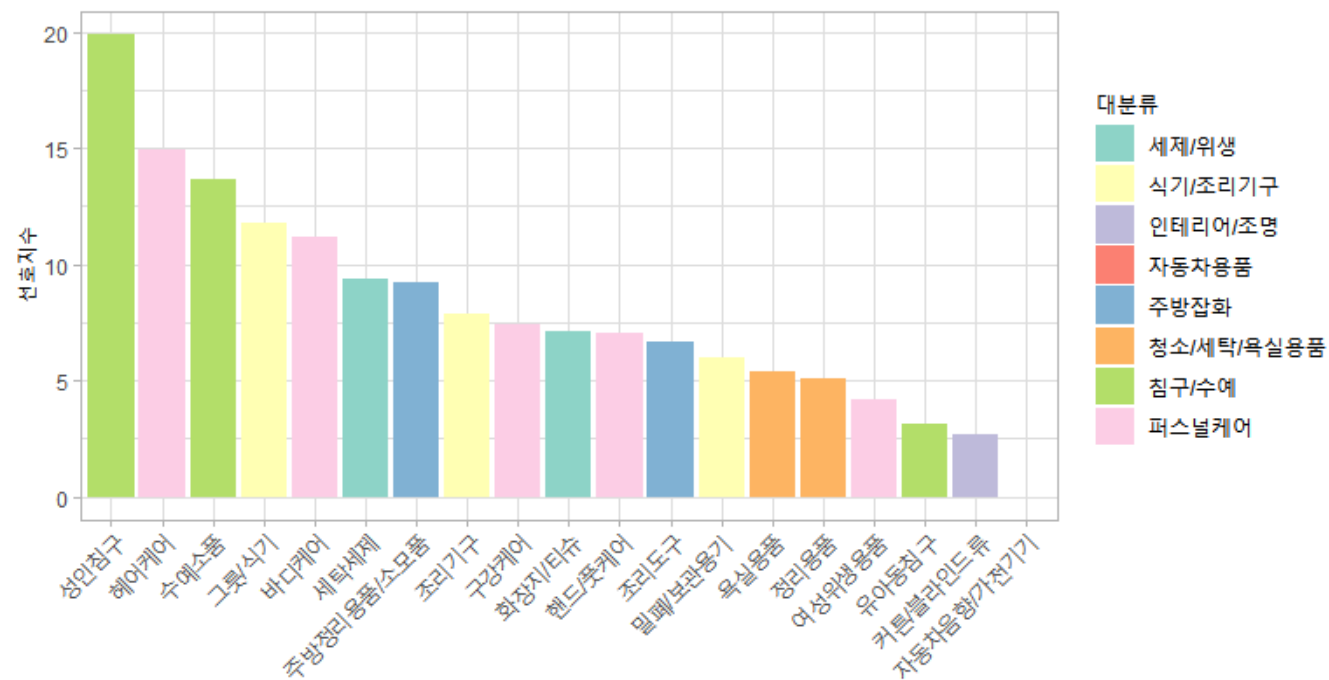
(1) 대분류별 선호지수 비교

2. 식품 / 농수산물



- 홍삼/ 인삼가공식품, 영양제 에 대한 선호지수가 나머지 항목들보다 압도적으로 높은 것으로 나타남

3. 생활/자동차용품

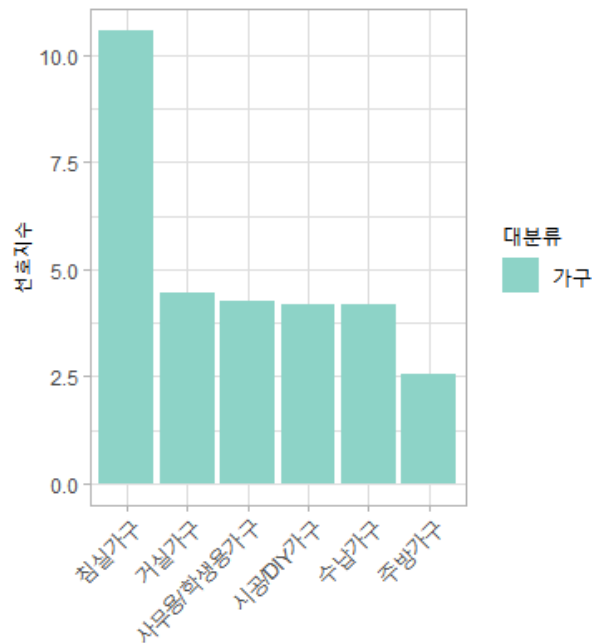


- 침구/ 수예용품의 선호지수가 높은 것을 알 수 있음

02 상품군별 온라인 선호지수 생성

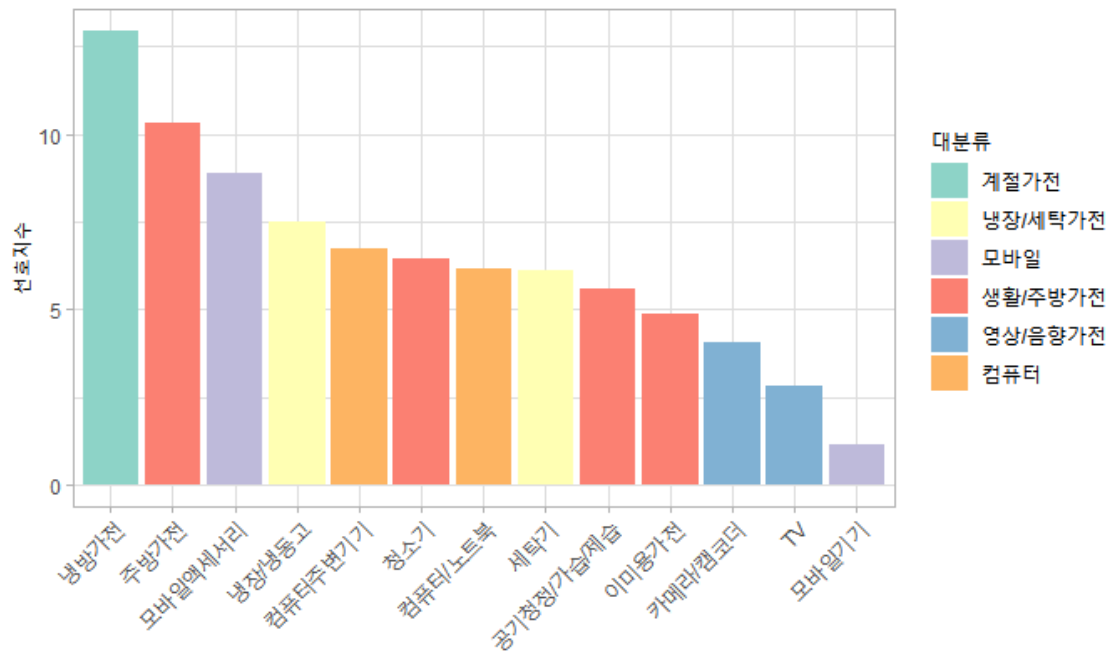
(1) 대분류별 선호지수 비교

4. 가구



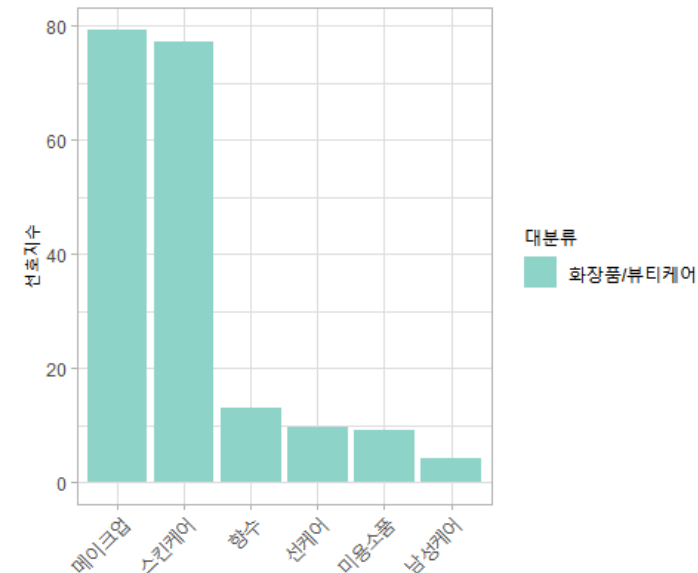
- 침실가구 에 대한 선호지수가 나머지 항목들보다 약 2배 높음

5. 가전/컴퓨터/통신기기



- 냉방 가전의 선호도가 가장 높음

6. 화장품

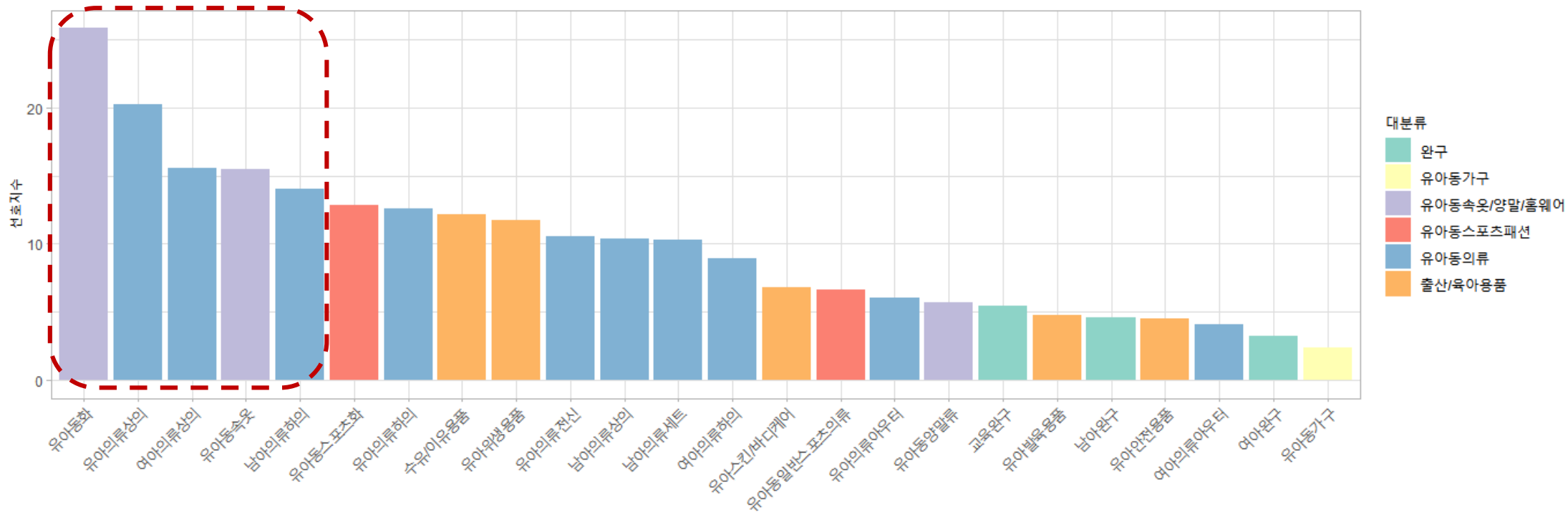


- 메이크업, 스킨케어 에 대한 선호지수가 나머지 항목들보다 약 4배 이상 더 높음

02 상품군별 온라인 선호지수 생성

(1) 대분류별 선호지수 비교

7. 아동/유아용품

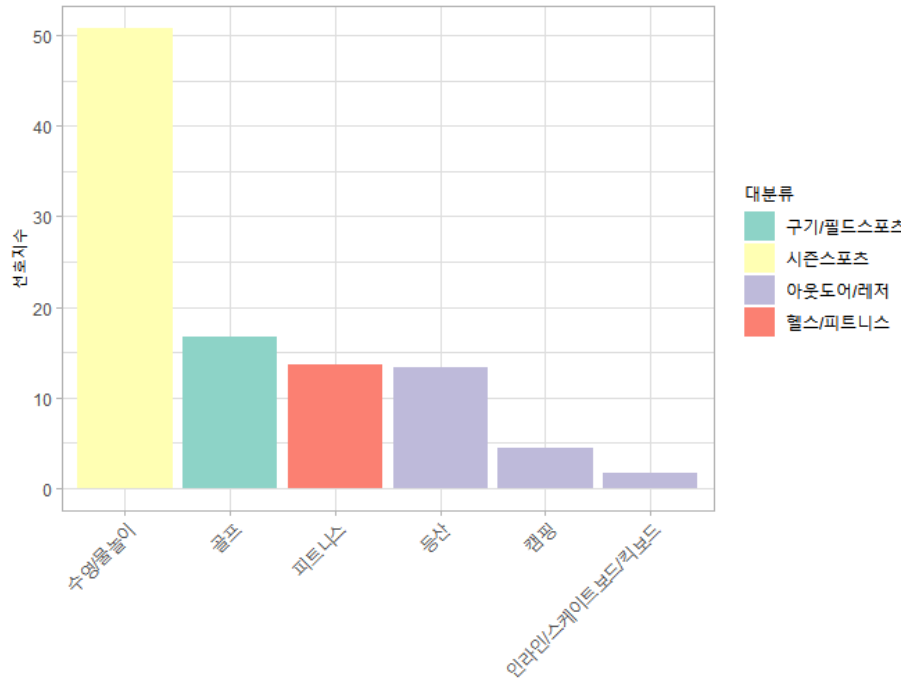


- 유아동의류, 유아동속옷/양말/홈웨어의 선호도가 높음

02 상품군별 온라인 선호지수 생성

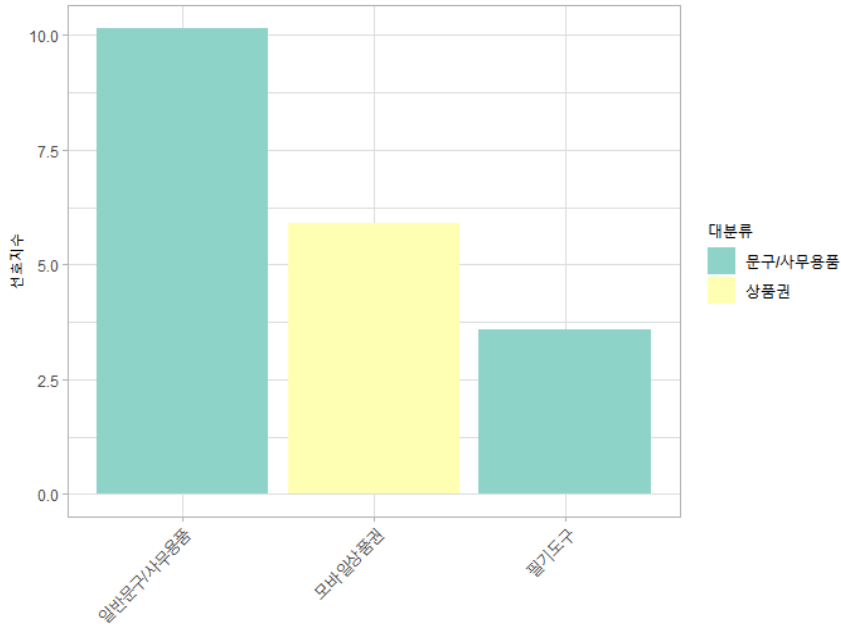
(1) 대분류별 선호지수 비교

8. 스포츠/레저용품



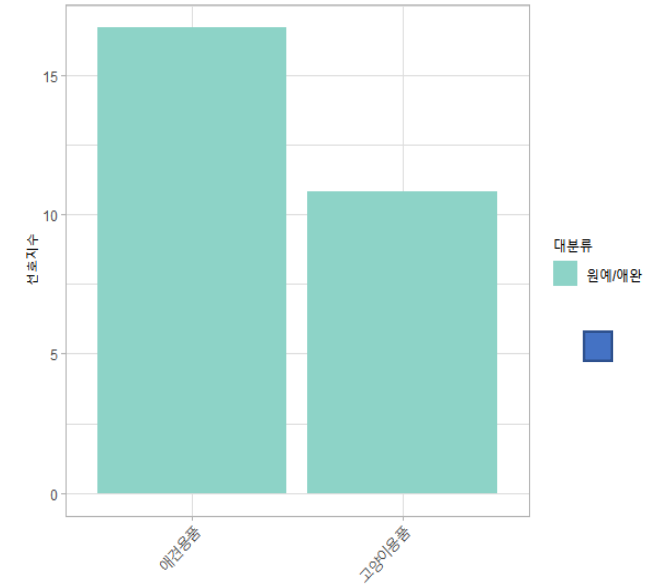
- 수영/물놀이 의 선호도가 매우 높음

9. 사무/문구



- 대체적으로 선호도가 낮은 편이며, 그 중 일반문구/사무용품의 선호도가 가장 높음

10. 애완용품

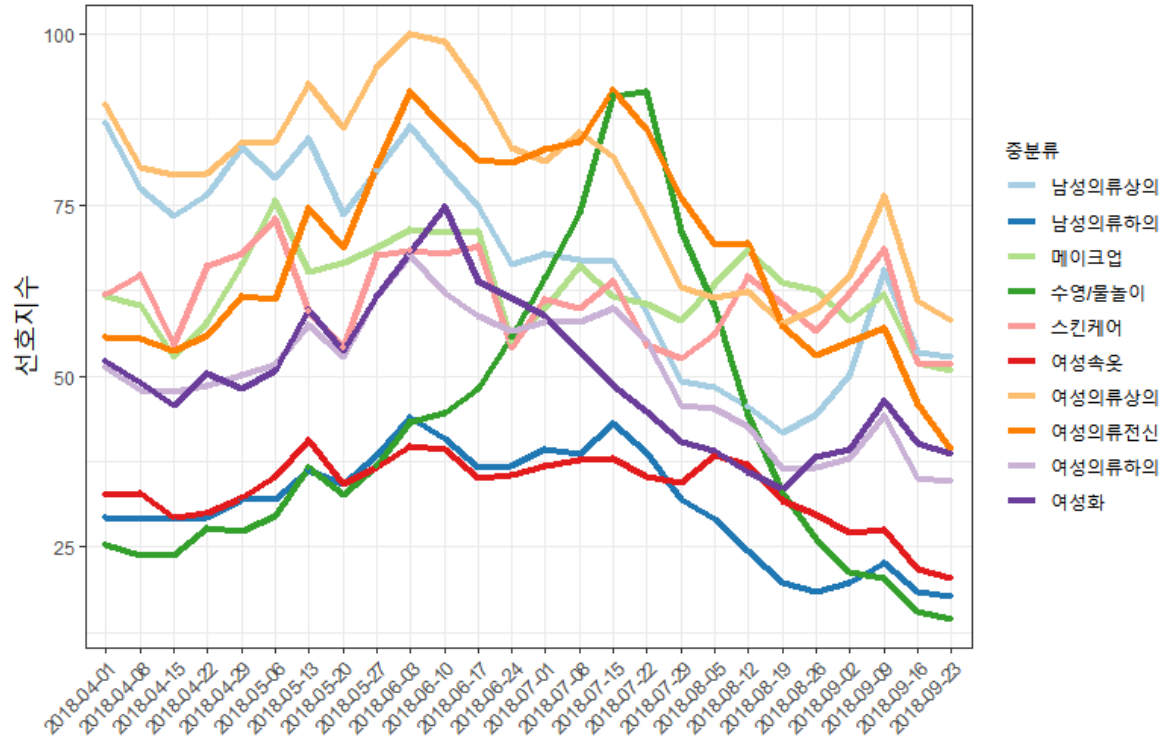


- 애완용품의 선호가 고양이용품의 선호보다 앞서있음

02 상품군별 온라인 선호지수 생성

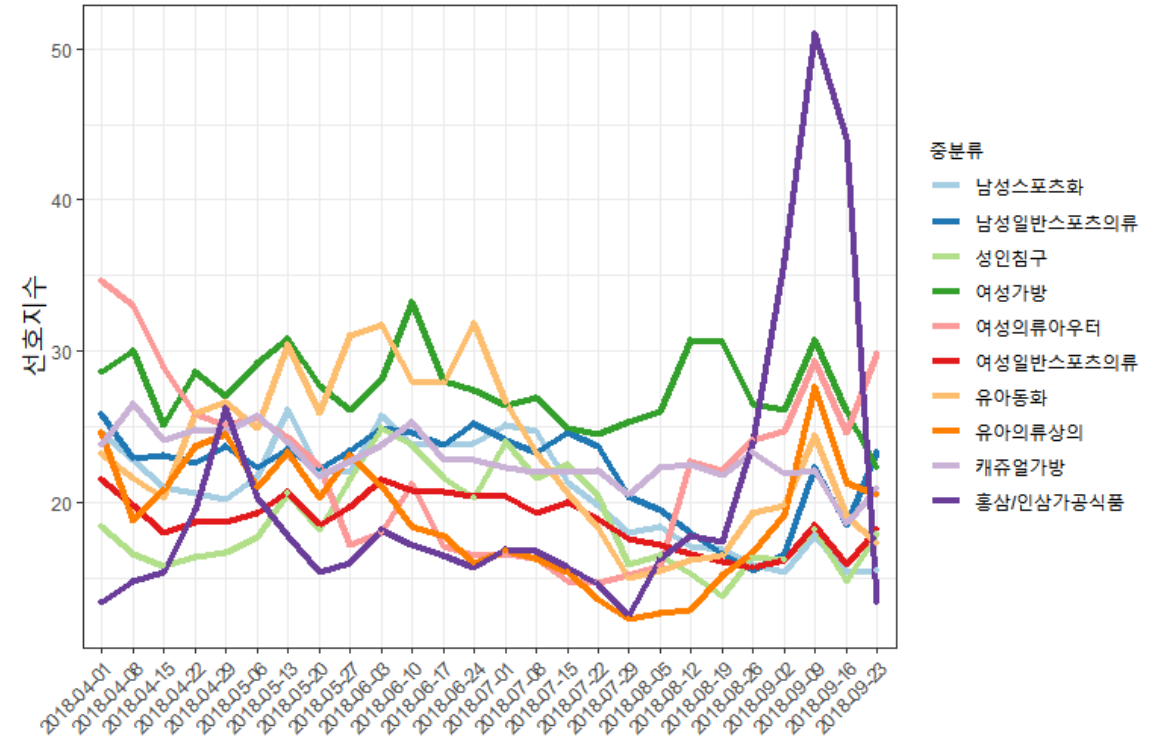
(2) 선호지수 상위 20개 상품군의 트렌드

1위 ~ 10위 상품군



- 상위 10개의 중분류 상품군을 살펴보았을 때, 패션/ 의류인 여성의류와 남성의류 상품군이 가장 많았으며 대부분 증가하다가 8월을 기점으로 감소하는 추세를 보임
- 스킨케어, 메이크업 상품군은 일정한 추세를 보임

11위 - 20위 상품군



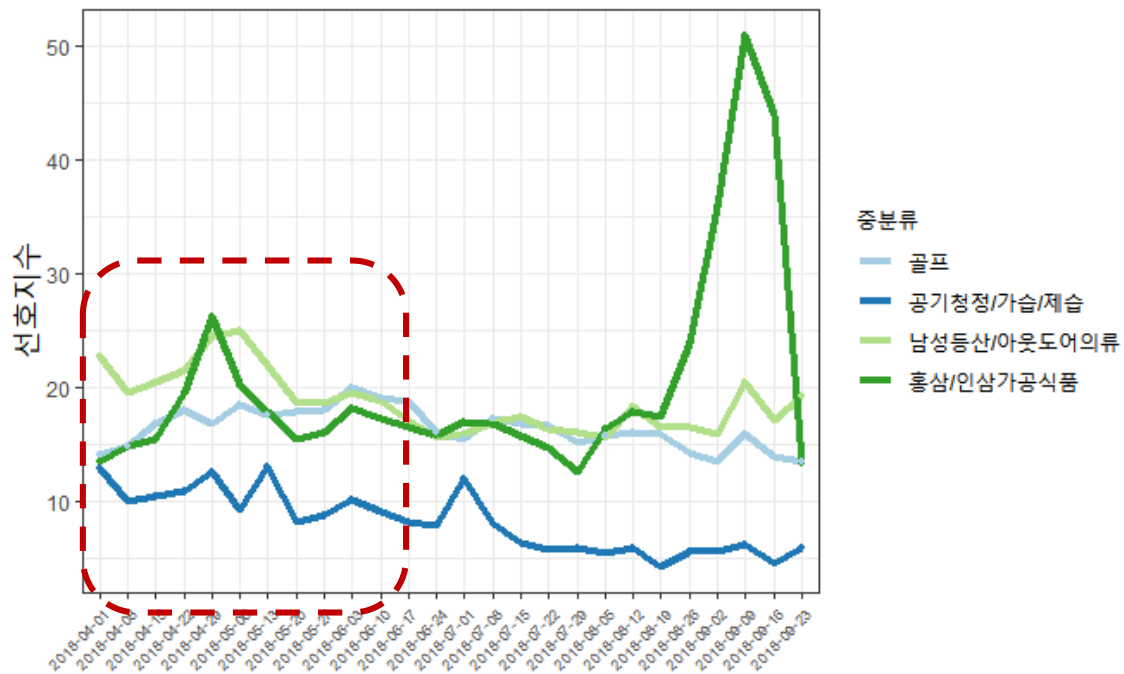
- 상위 11위 - 20위의 상품군을 살펴보았을 때, 대부분 일정한 추세를 나타냄
- 특히 홍삼/인삼가공식품의 경우 8월 말을 기점으로 급증하는 추세를 보임

02 상품군별 온라인 선호지수 생성

(3) 외부 환경적 요인 - 4, 5월 증가항목

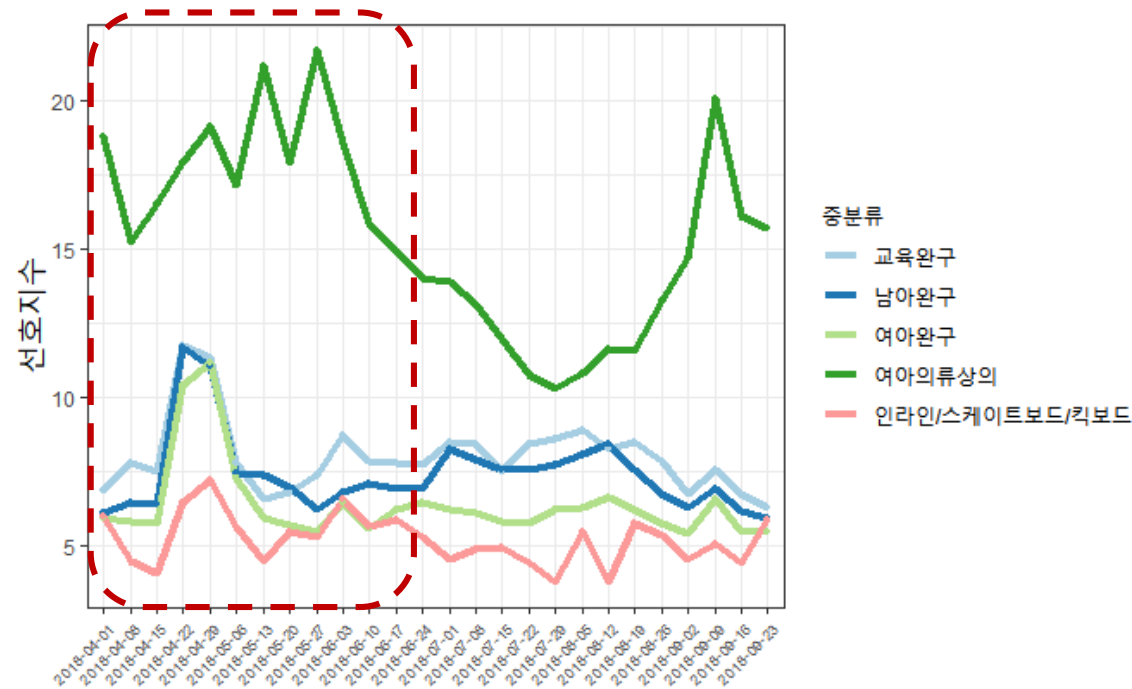
- 4월과 5월에 건강, 취미생활 또는 아동/유아용품의 선호가 크게 나타났으며 이는 가족 행사인 어버이날, 어린이날 등의 외부적 특성이 반영됨

건강, 취미생활



- 홍삼/인삼가공식품에 대한 선호가 4월말 또는 5월 초에 높은 것으로 보아 '어버이날' 선물에 대한 수요가 증가하는 것을 나타내는 것으로 보임
- 골프, 남성등산/아웃도어의류의 선호가 높은 것으로 나타남
- 공기청정/가습/제습 상품군에 대해서는 황사, 미세먼지와 같은 계절적 특성을 나타낸다는 것을 알 수 있음

아동/유아용품



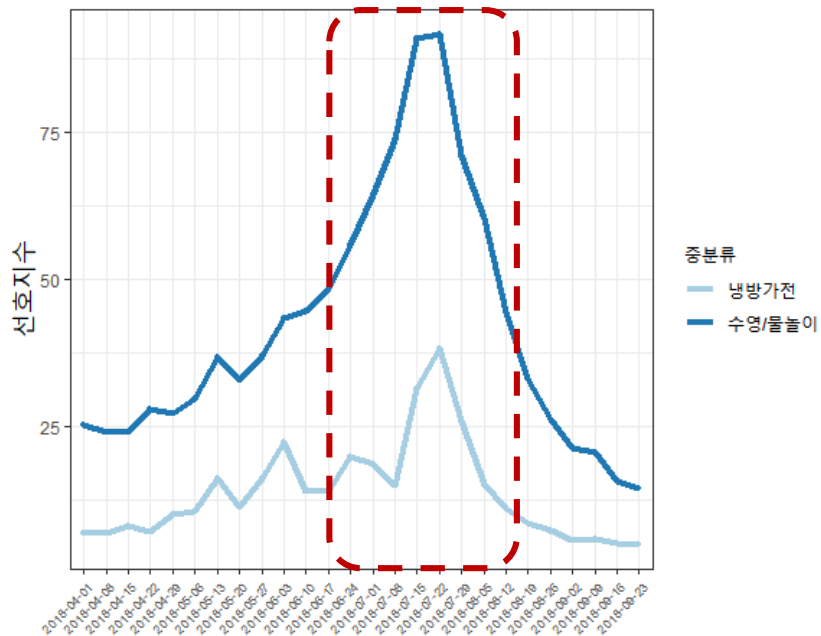
- 아동 과 관련된 상품군에 대한 선호가 증가하는 것으로 보아 '어린이날' 선물을 하려는 수요가 증가한다는 것을 나타냄

02 상품군별 온라인 선호지수 생성

(3) 외부 환경적 요인 - 7, 8월 증가항목

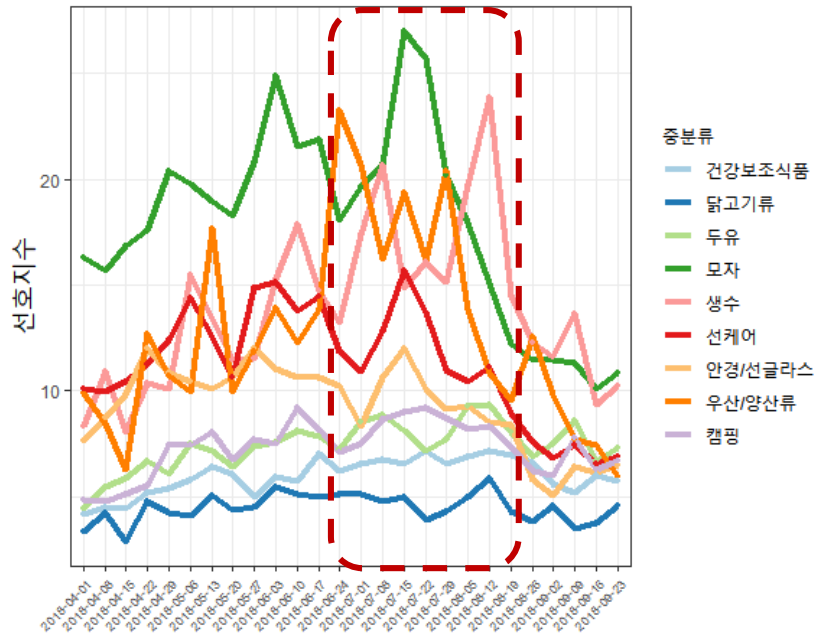
- 7월과 8월에는 '여름' 이라는 계절적 특성이 반영된 특정 항목의 선호가 크게 나타남

'반짝' 선호 상품군



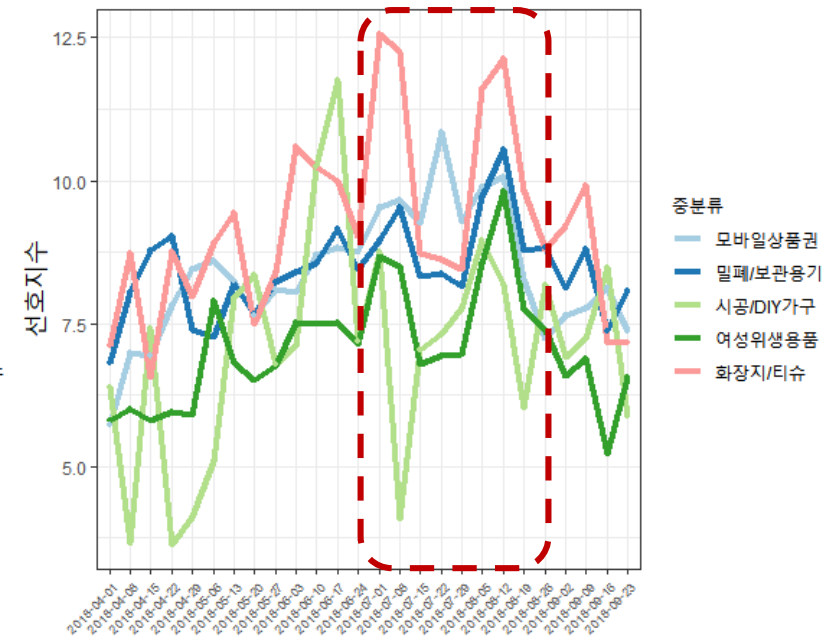
- 다른 품목에 비해 특히 냉방가전, 수영/물놀이 에 대한 선호는 7월과 8월에 특히 집중적이며 증가폭과 감소폭이 매우 큰 것으로 나타남

다이어트/ 여행,레저 활동



- 여름철 물놀이 와 관련해 모자, 안경/선글라스, 선케어, 캠핑 상품군의 선호도가 높음
- 여름철 다이어트 관련한 상품군으로 건강보조식품 닭고기류의 선호도가 높음

기타 선호 상품군



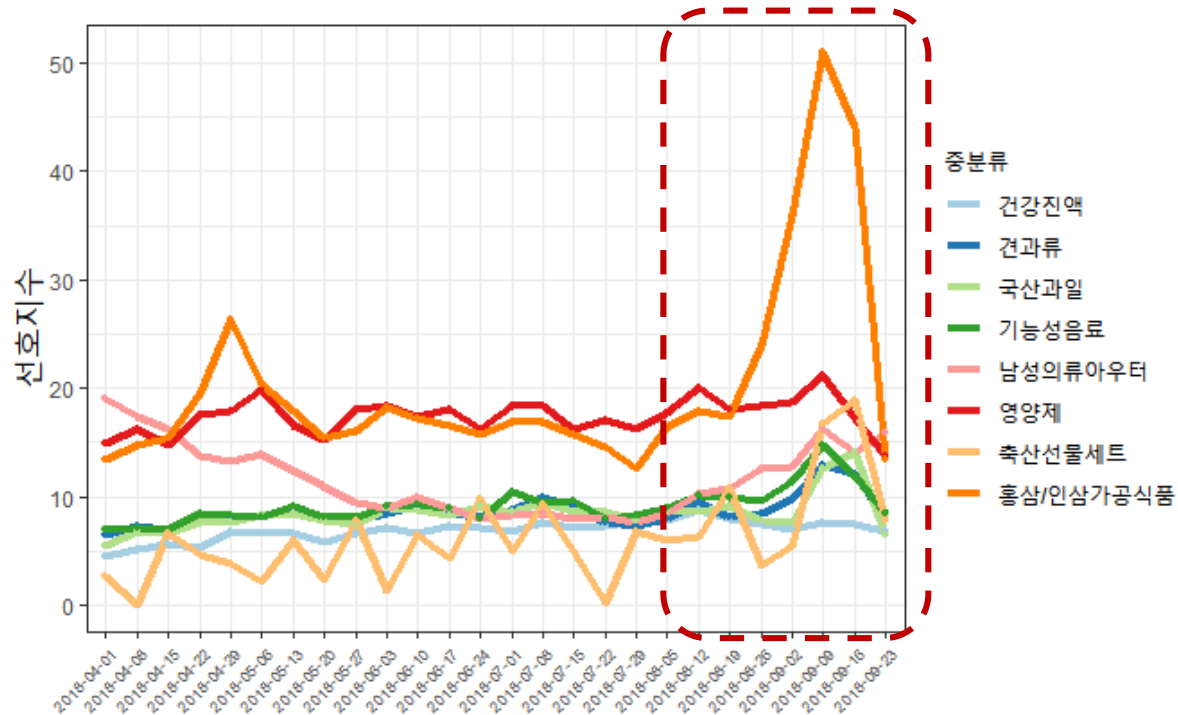
- 화장지/티슈, 여성위생용품, 밀폐/보관용기 상품군에 대해서는 매우 비슷한 경향을 보임

02 상품군별 온라인 선호지수 생성

(3) 외부 환경적 요인 - 9월 증가항목

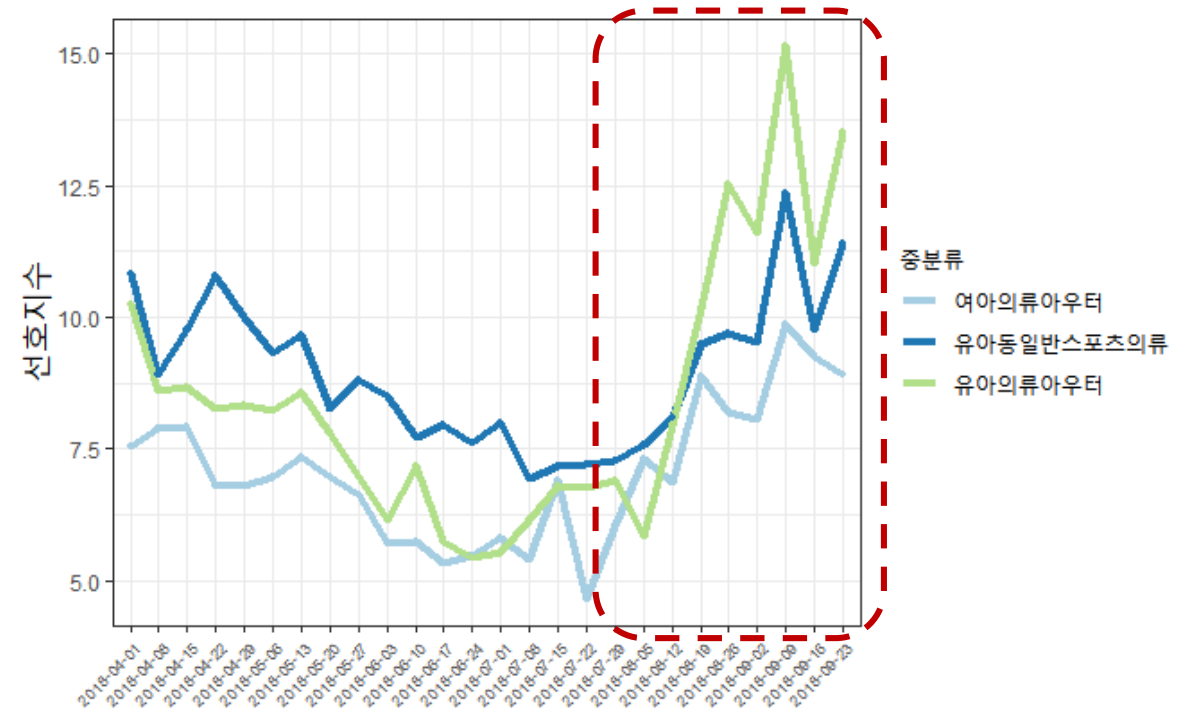
- 9월 말 '추석' 영향을 나타낸 상품군 및 유/아동 상품군이 높은 선호도를 나타냄

'추석' 선물



- 대표적인 추석선물세트인 건강진액, 국산과일, 영양제, 축산선물세트, 홍삼/인삼가공식품에 대한 선호가 급증함

유/아동 상품군



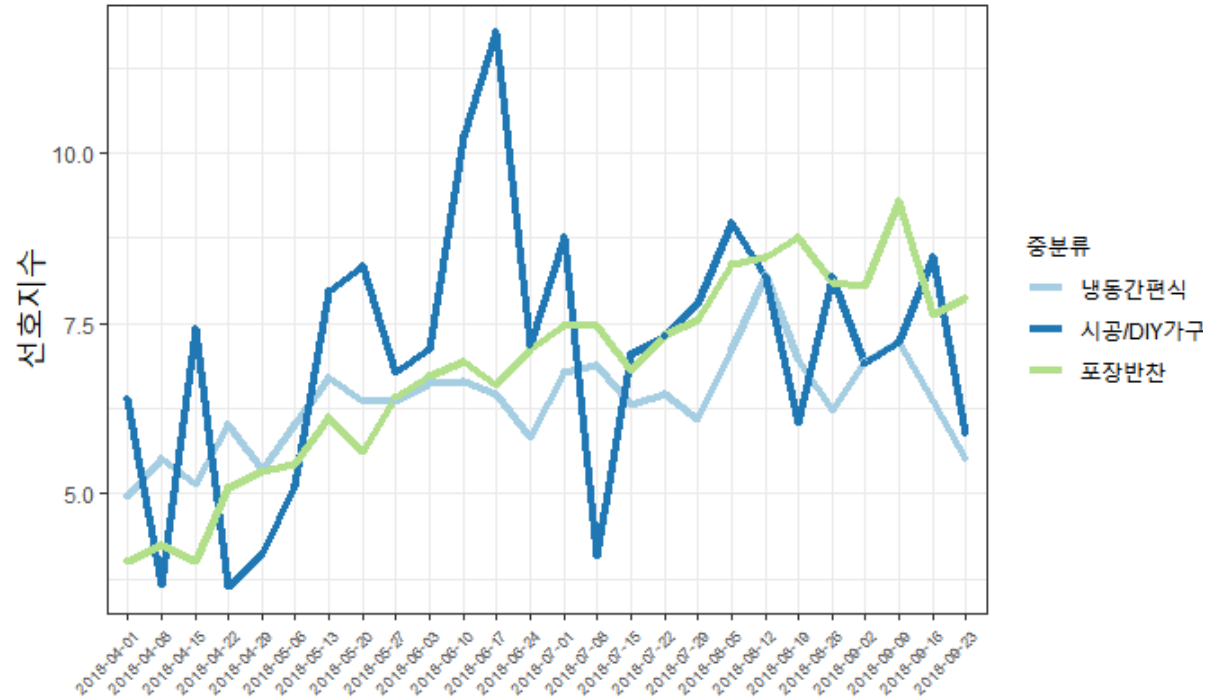
- 유/아동 상품군의 선호도의 추세는 감소하다가 7월 초를 기점으로 다시 증가함을 알 수 있음

02 상품군별 온라인 선호지수 생성

(3) 외부 환경적 요인 -기타

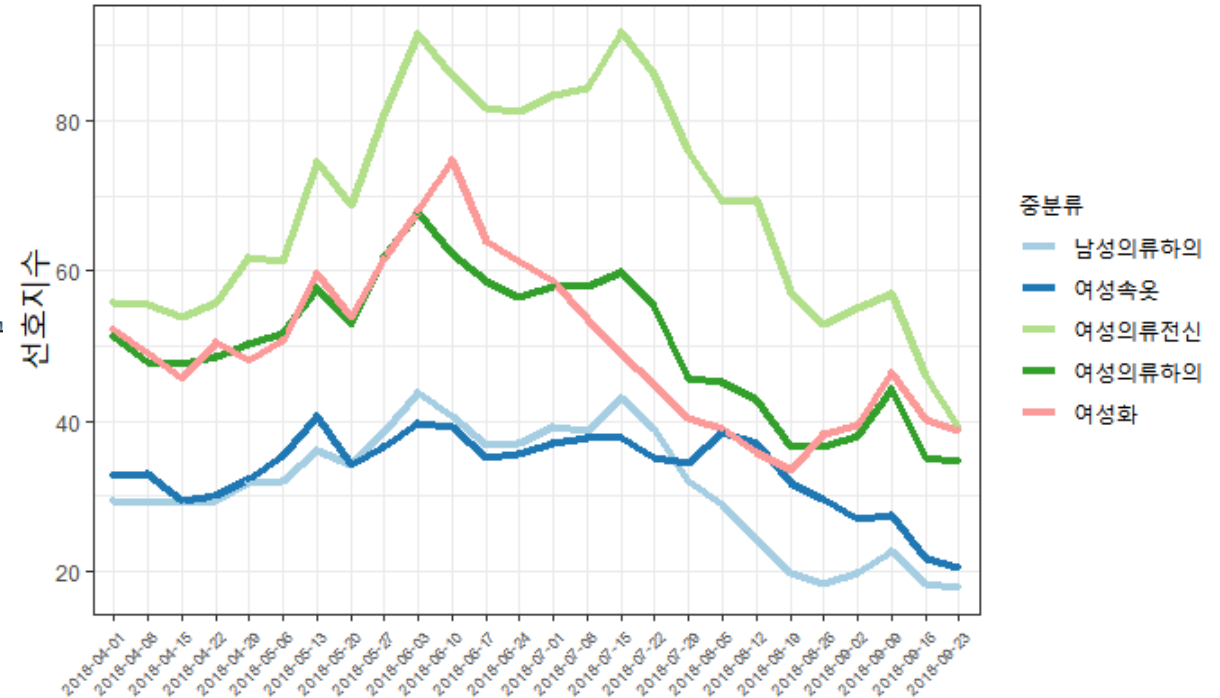
- 기타 비슷한 경향을 보이는 상품군으로 1인가구의 증가로부터 나타날 수 있는 상품군과 패션/ 잡화와 관련된 상품군을 집중적으로 살펴봄

'1인가구' 증가



- 냉동간편식, 시공/DIY가구, 포장반찬에 대한 선호도가 꾸준한 증가 추세를 보이고 있는 것으로 보아 1인 가구의 증가가 원인으로 보임

비슷한 의류 트렌드



- 패션/의류 항목 중 여성의류와 패션잡화 및 남성의류 하의의 추세가 눈에 띄게 비슷한 것을 알 수 있음

CONTENTS

1
데이터 탐색

2
선호지수
생성

3
수요 트렌드
예측

4
서비스 제안

5
결론

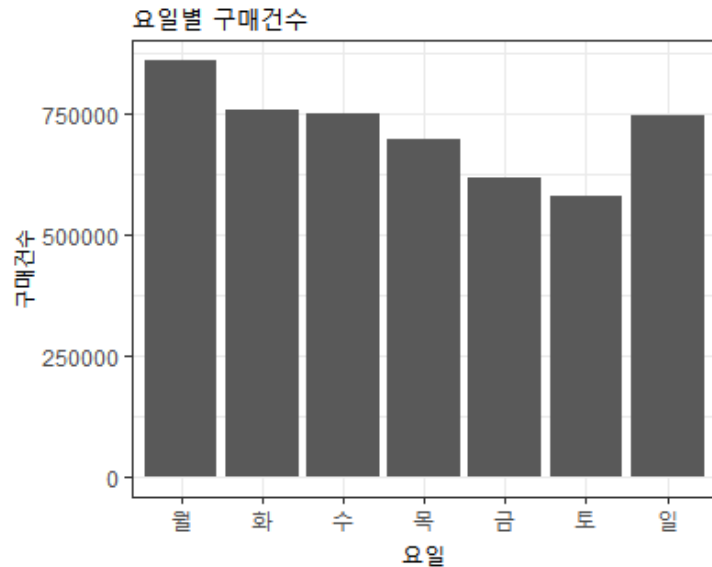
01 예측 개요

예측 대상 및 시점

Q. 예측 대상 ?

상품군별 구매량
온라인 선호지수

Q. 예측 시점 ?



“ 실수요 및 선호도 예측을 통해
인사이트 도출 가능 ”

- 주어진 데이터는 6개월의 데이터로 장기적인 예측 불가능
- 요일별로 유사한 패턴을 보이므로 주별 단위로 추세를 파악하는 것이 적절
- 일주일 간의 수요를 높은 정확도로 예측함으로써 이에 맞는 적절한 마케팅을 시행



데이터를 주별 단위로 구성

2018년 4월 1일부터 2018년 9월 29일까지
총 26주 간의 주별 데이터를 이용하여
10월 첫 주(9월 30일~10월 6일)의 수요 예측

02 예측 모형

(1)LSTM

Q. LSTM(Long short-term memory)

- 딥러닝 기법의 하나로 시계열 데이터, 텍스트 데이터 등 순서와 지속성을 고려해야 하는 데이터에 적합
- RNN(Recurrent Neural Network)의 장기 의존성 문제를 보완한 기법

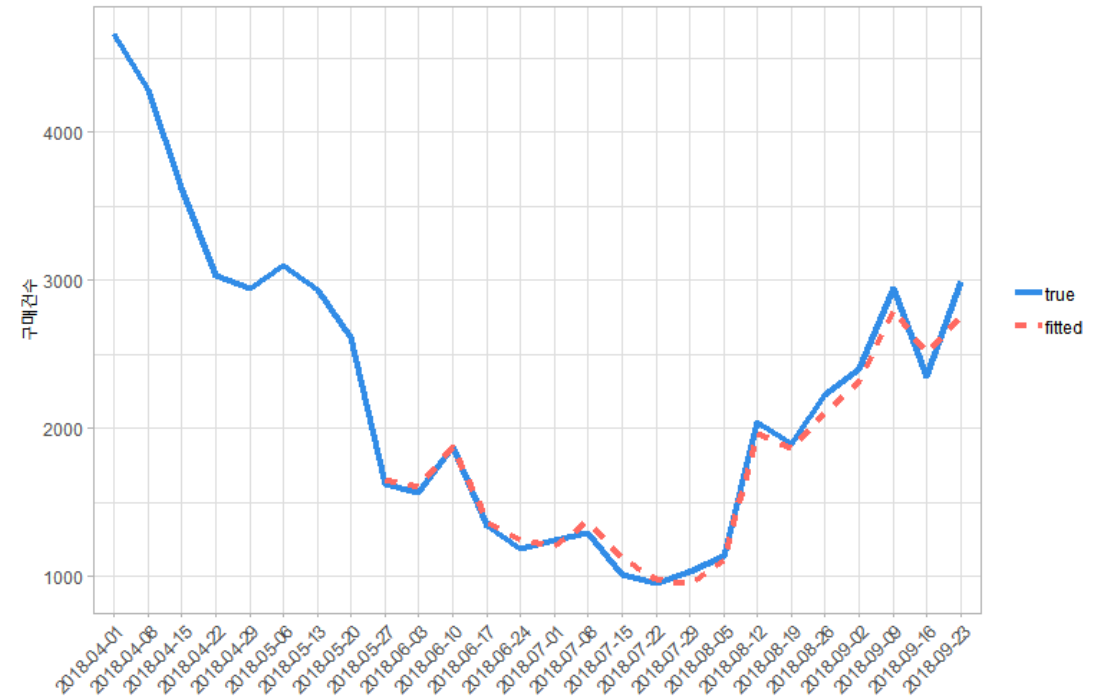
Sequence | 8주

- 계절성 등을 고려하여 LSTM 모델의 학습 단위를 8주로 설정
- 6개월 간의 데이터를 학습시켜 8주 간의 데이터를 통해 9주차를 예측하는 모델 구축
- 데이터 중 가장 최근 8주 (8월 5일 ~ 9월 30일)의 값을 통해 10월 첫 주(9월 30일 ~ 10월 5일)의 수요 예측

Dropout

- 데이터 수가 매우 적으므로 오버피팅을 줄이기 위해 드롭아웃 기법 적용

예) '여성 아우터'의 구매건수 모델 적합



02 예측 모형

(2)통계모형 - BASS

Q.

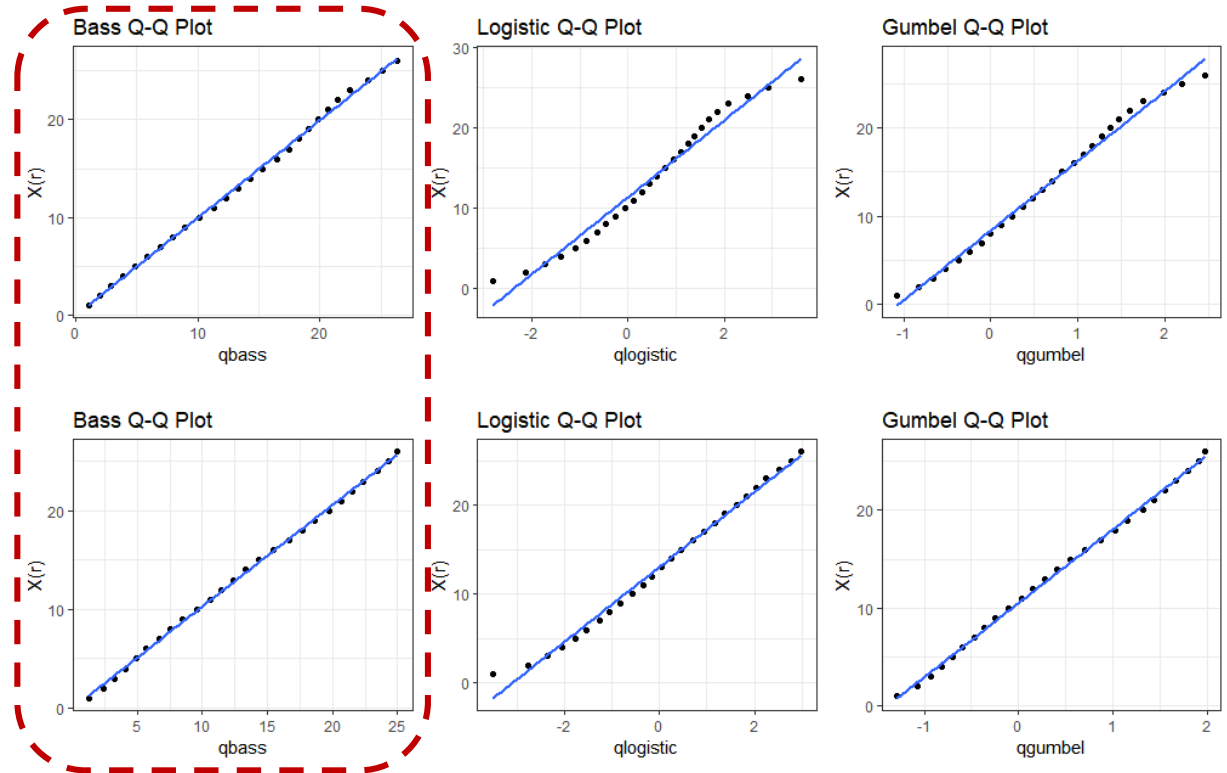
BASS Model

- 시장분석과 신기술의 수요예측을 위해 활용
- 새로운 내구소비재의 구매량을 묘사하고 예측하는 도구 중의 하나
- 모형이 단순하여 간단한 평가를 하기에 충분하지만 단지 여러가지 기술확산모델의 한가지이므로 주의



수요 예측 모형에 많이 사용하는 통계 모형을 이용해 선호지수를 예측하였고, 중분류 중 “남성의류상의”, “여성의류전신”에 관한 예측을 수행함

(1) Q-Q PLOT 이용해 적절한 모형찾기

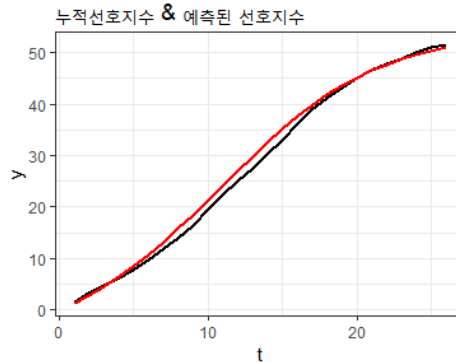
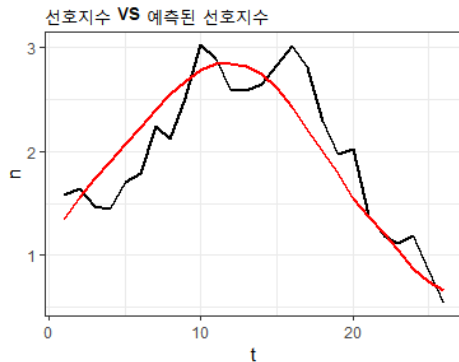
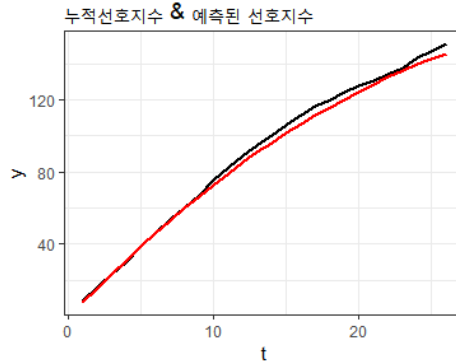
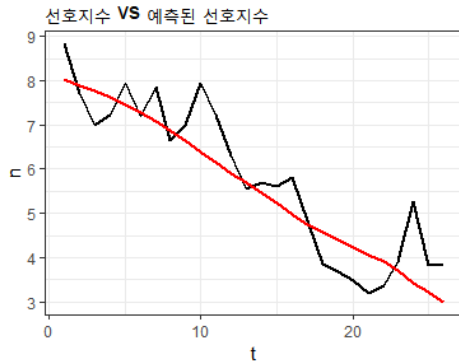


수요 예측을 위해 Bass, Logistic, Gumbel 모형 중 가장 적합한 모형을 찾기 위해 Q-Q plot 을 이용해 살펴봄
-> 결과적으로 Bass model 이 가장 적합

02 예측 모형

(2)통계모형 - BASS

(2) Bass 모형 이용한 수요예측

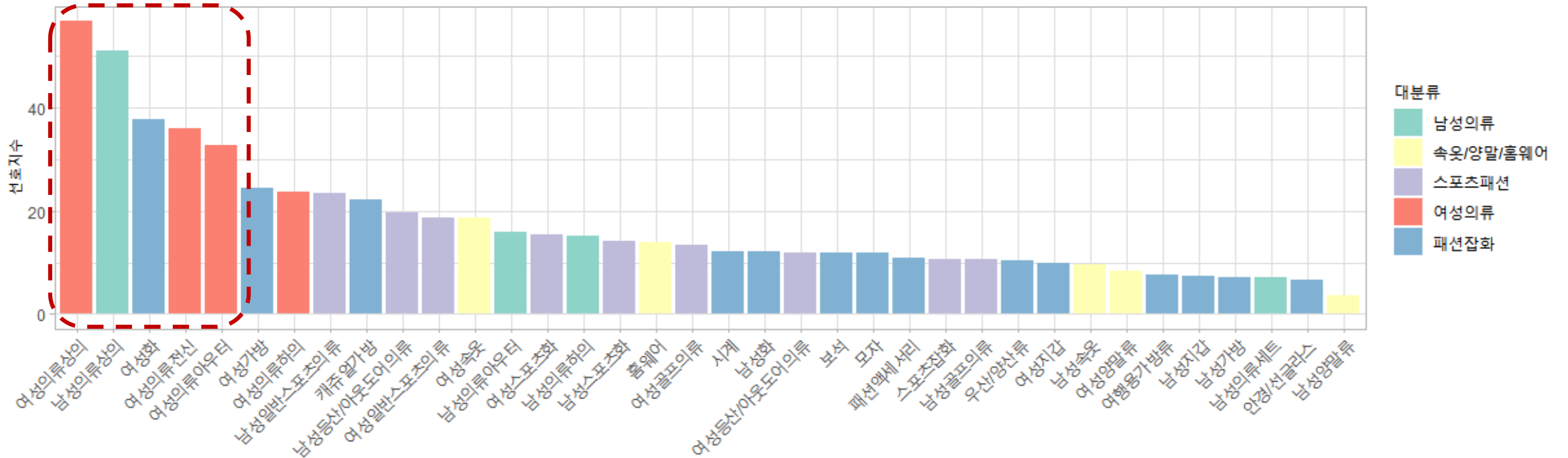


- Bass model 적합 후 모수를 추정하여 구매건수를 예측한 결과, 트렌드를 살펴보기에 적합하였음
- 그러나 일, 주 단위의 미세한 변동을 파악하기에는 LSTM 모형에 미치지 못한다는 한계가 있었기 때문에 상품군별 수요량과 온라인 선호지수의 최종 예측에는 LSTM 모형을 사용

03 예측 결과

대분류별 선호지수 예측치 비교(1)

1. 패션/ 의류

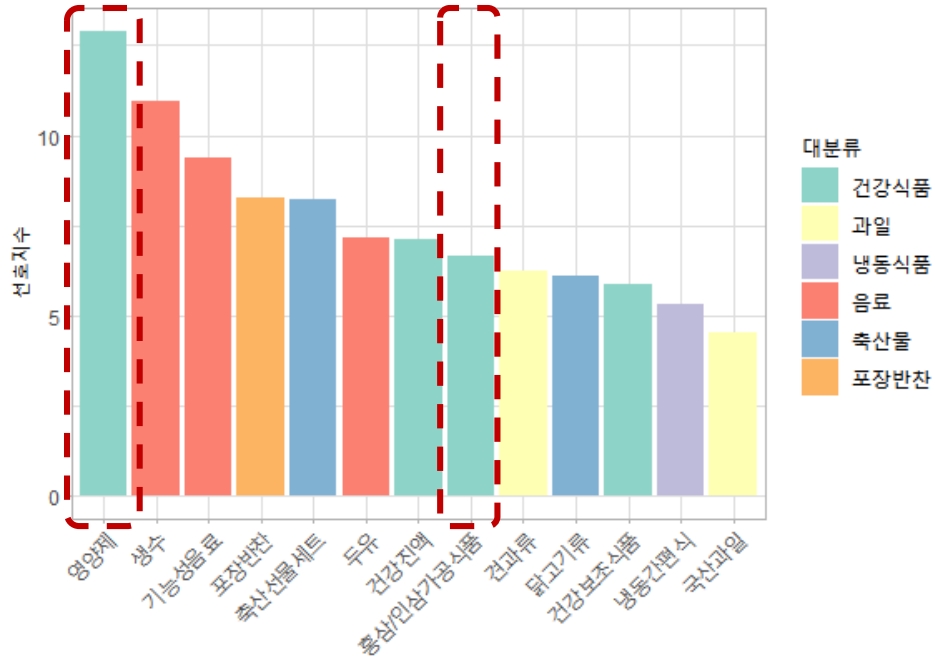


- 여성의류의 선호지수가 여전히 다른 항목들보다 압도적으로 높은 점수를 나타내므로 선호도가 꾸준하다고 판단할 수 있음
- 남성의류상의 제품군에서 높은 비중을 차지하는 '남성티셔츠' 상품도 지속적으로 인기 있을 것으로 예상됨

03 예측 결과

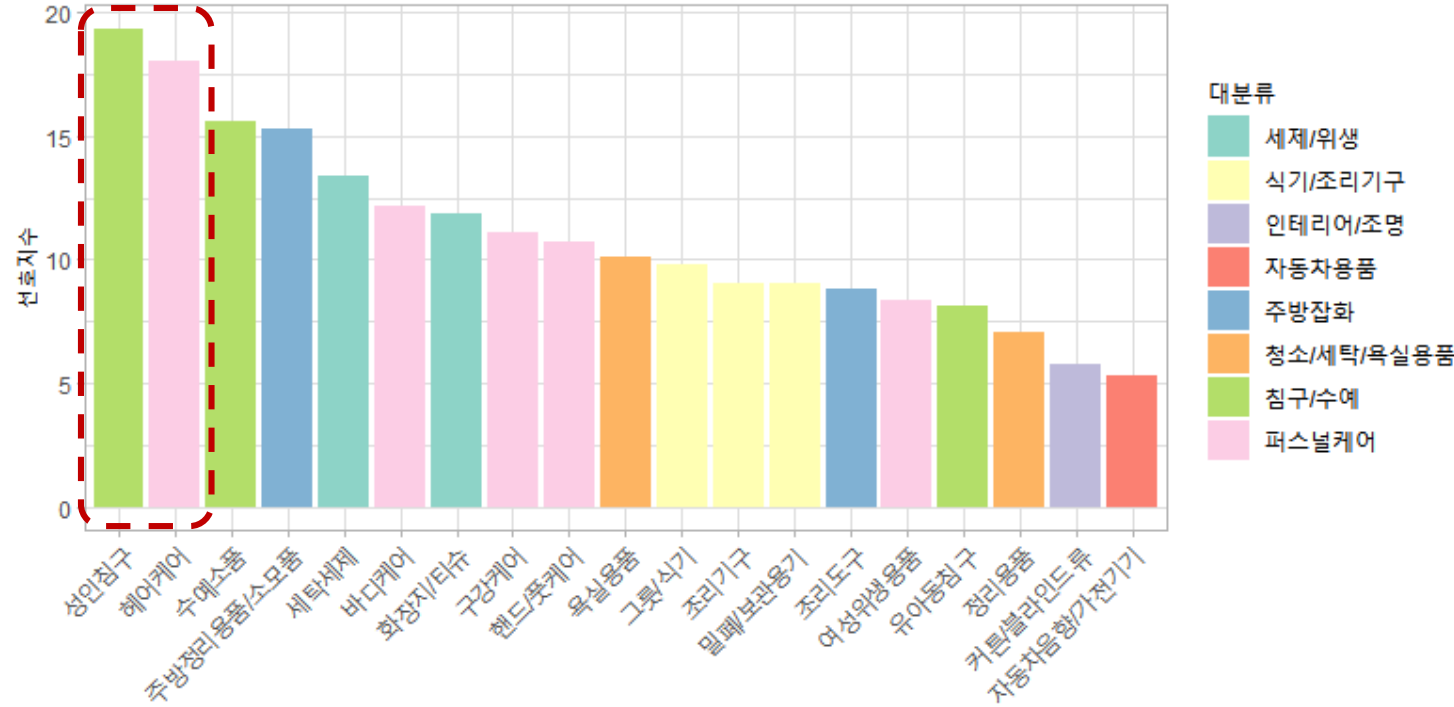
대분류별 선호지수 예측치 비교(2)

2. 식품 / 농수산물



- 홍삼/인삼가공식품의 선호도는 감소했지만, 영양제에 대한 선호도는 여전히 매우 높음

3. 생활/자동차용품

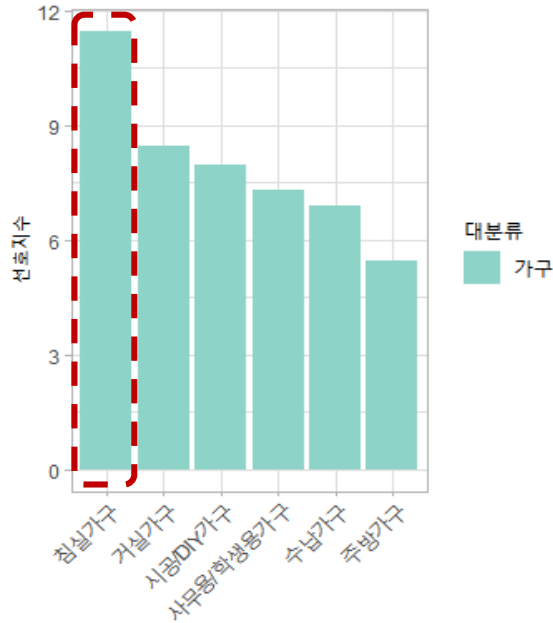


- 침구/헤어케어의 선호도 또한 지속적으로 높을 것으로 예상됨

03 예측 결과

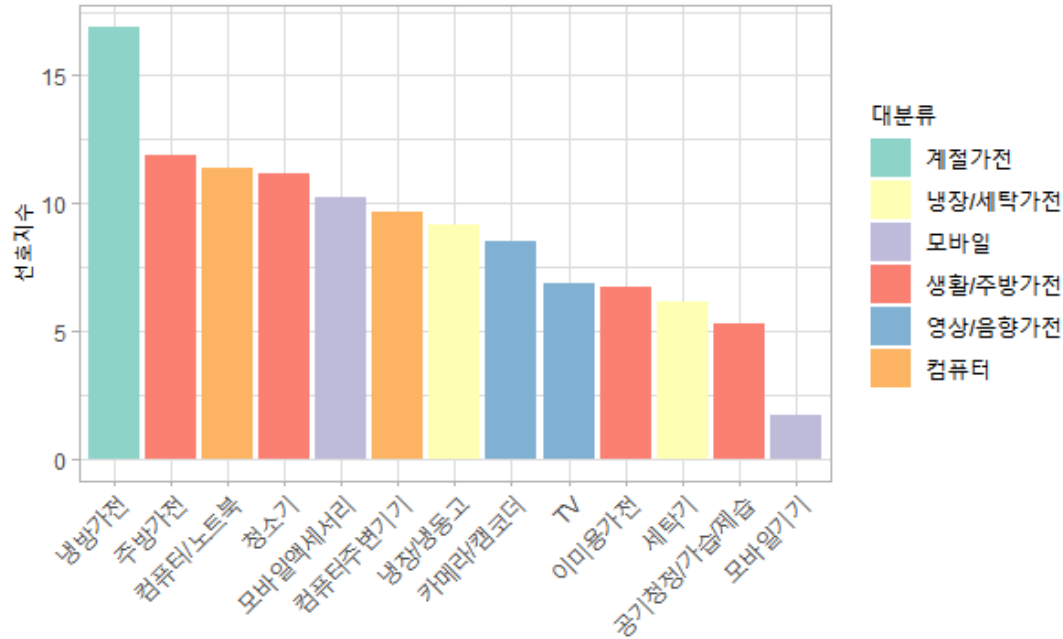
대분류별 선호지수 예측치 비교(3)

4. 가구



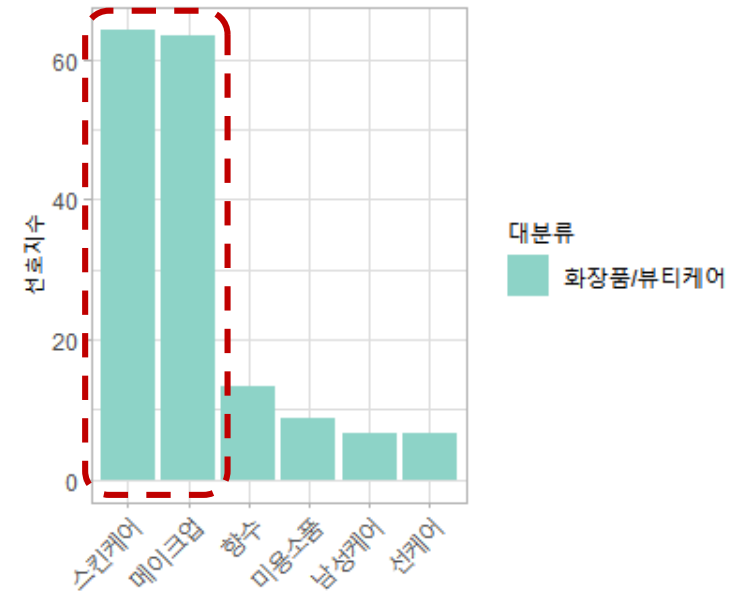
- 침실가구에 대한 선호도가 지속적으로 높을 것으로 예상됨

5. 가전/컴퓨터/통신기기



- 기존의 선호지수와 비슷한 양상을 보임

6. 화장품

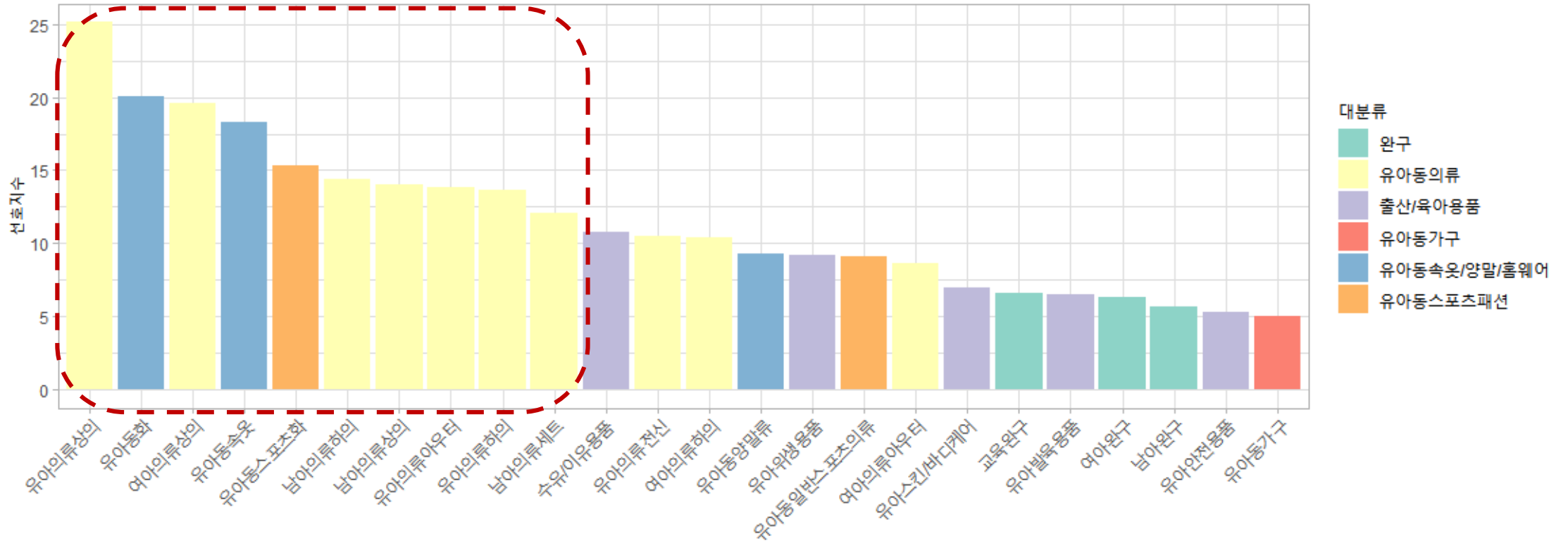


- 메이크업, 스킨케어에 대한 선호도가 여전히 나머지 항목들보다 압도적으로 높을 것으로 예상됨

03 예측 결과

대분류별 선호지수 예측치 비교(4)

7. 아동/유아용품

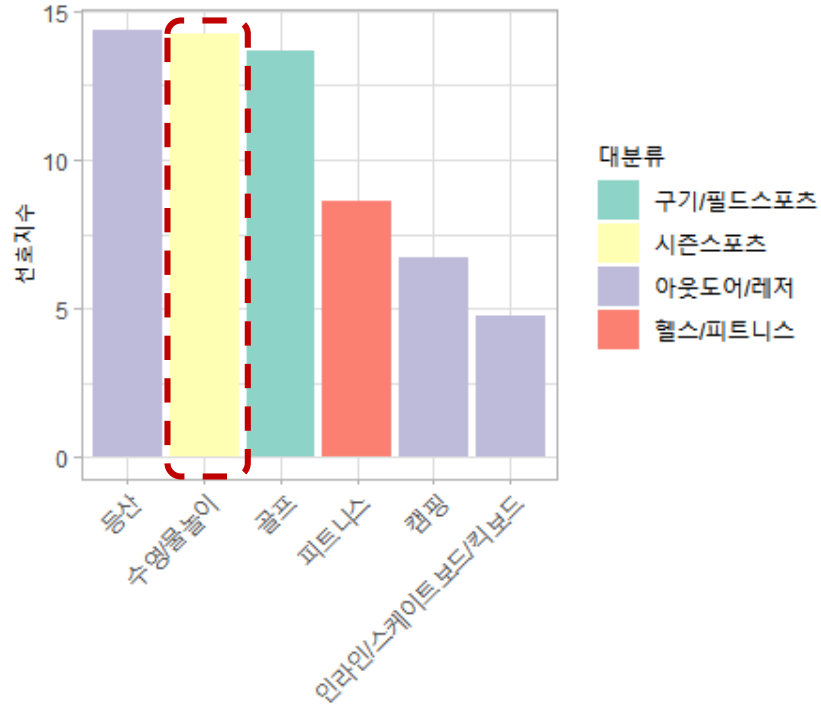


- 유아동의류, 유아동속옷/양말/홈웨어의 선호도가 계속해서 높을 것으로 예상됨

03 예측 결과

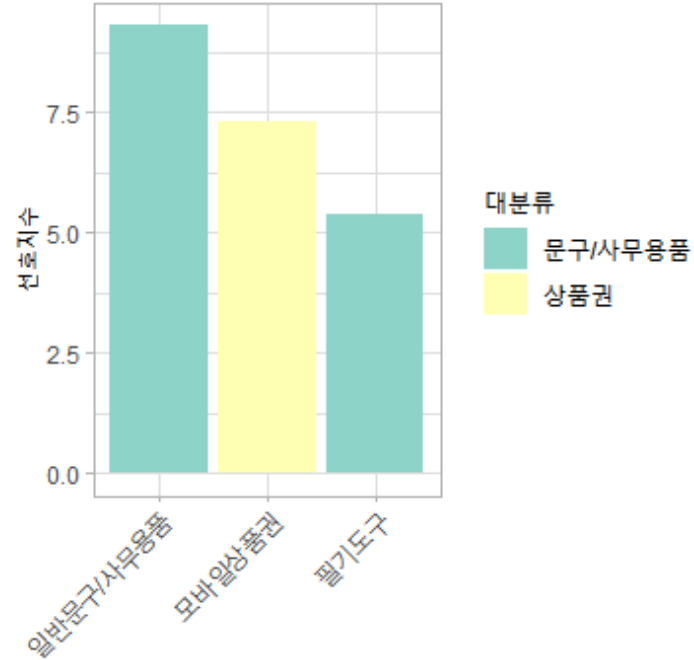
대분류별 선호지수 예측치 비교(5)

8. 스포츠/레저용품



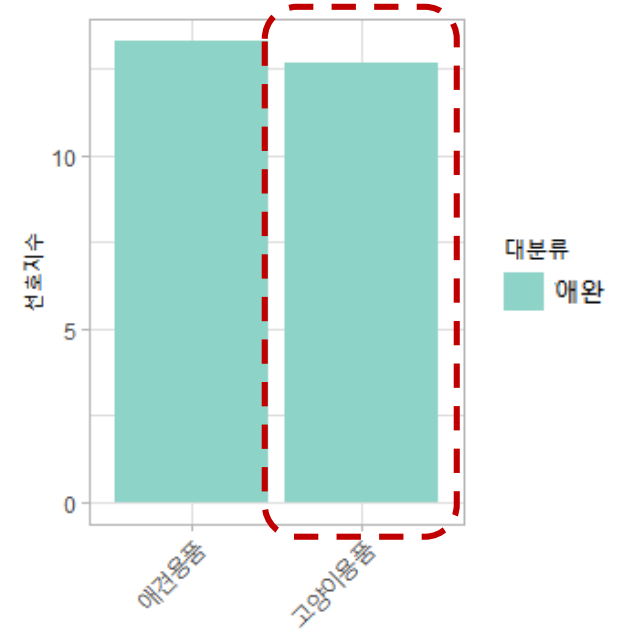
- 수영/물놀이 의 선호도가 기존에 비해 확연히 감소함

9. 사무/문구



- 기존의 선호지수와 비슷한 양상을 보임

10. 애완용품

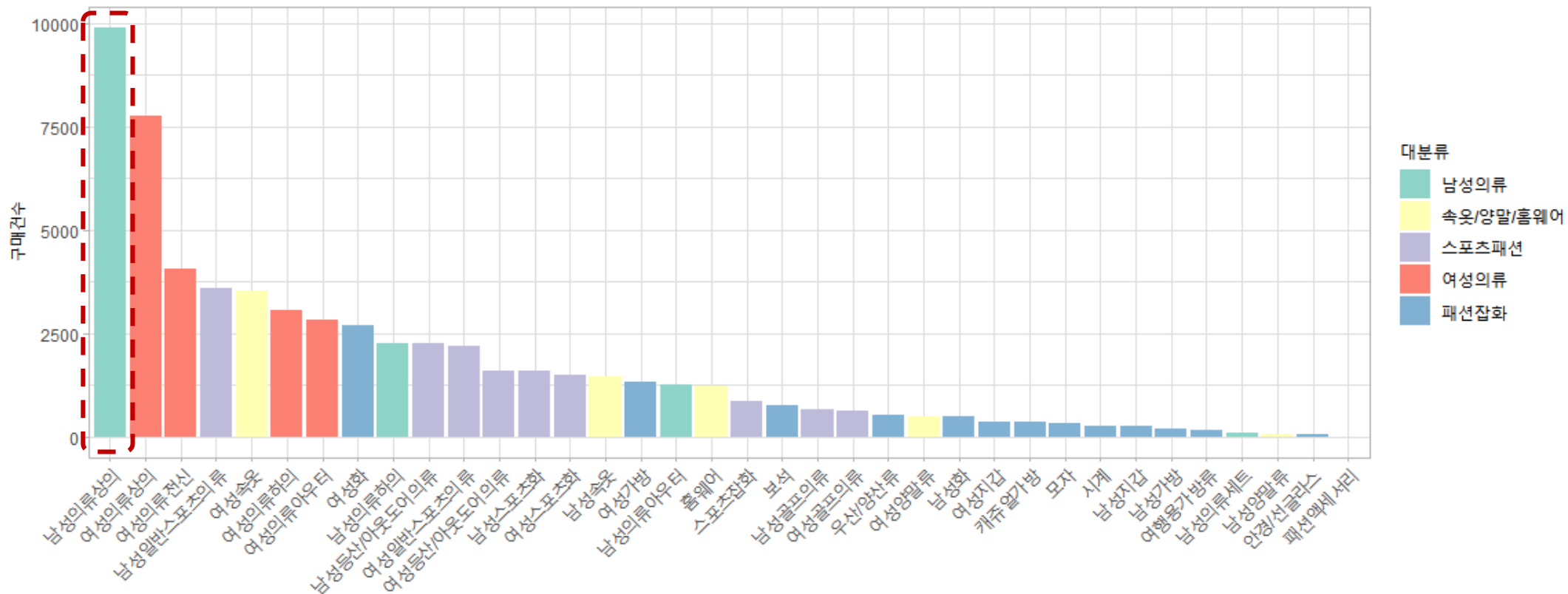


- 고양이용품의 선호도가 기존에 비해 증가함

03 예측 결과

대분류별 구매건수 예측치 비교(1)

1. 패션/ 의류

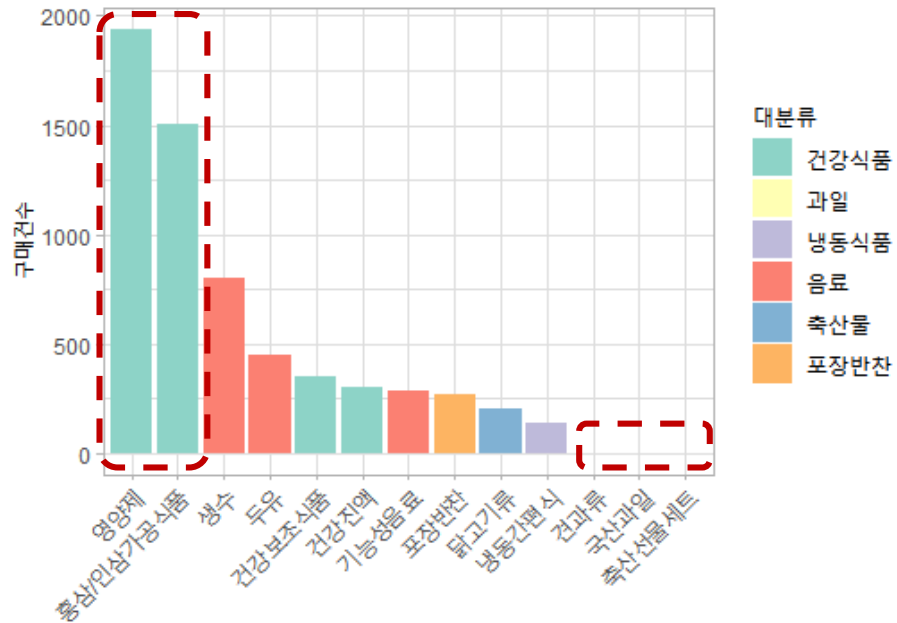


- 남성의류상의 항목의 구매량이 매우 증가할 것으로 예상됨

03 예측 결과

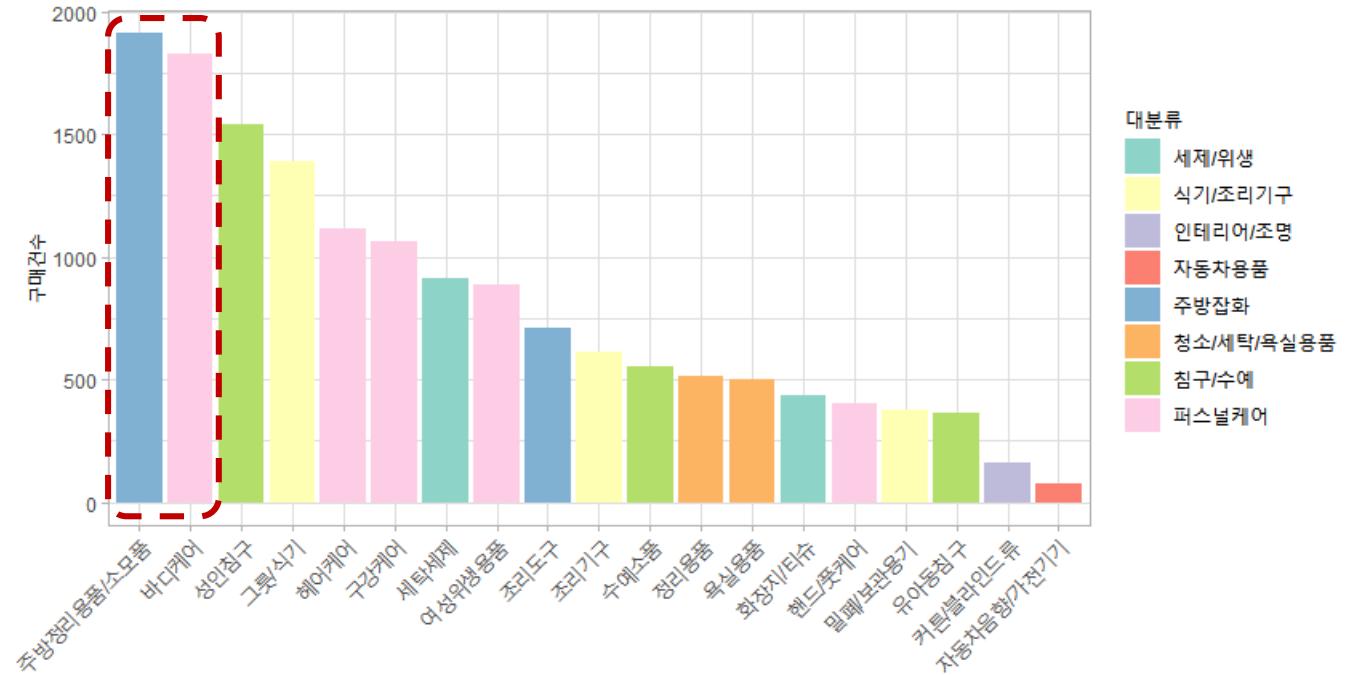
대분류별 구매건수 예측치 비교(2)

2. 식품 / 농수산물



- 영양제/ 홍삼, 인삼가공식품 등 건강식품의 구매량이 여전히 높을 것으로 예측됨
- 견과류, 과일, 축산물 등은 다른 식품군에 비해 거의 구매하지 않을 것으로 예측됨

3. 생활/자동차용품

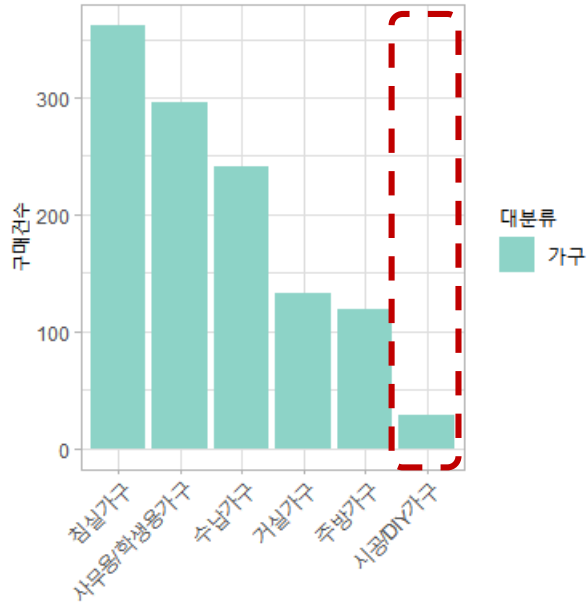


- 주방정리 용품과 바디케어의 구매량이 급증할 것으로 예상됨

03 예측 결과

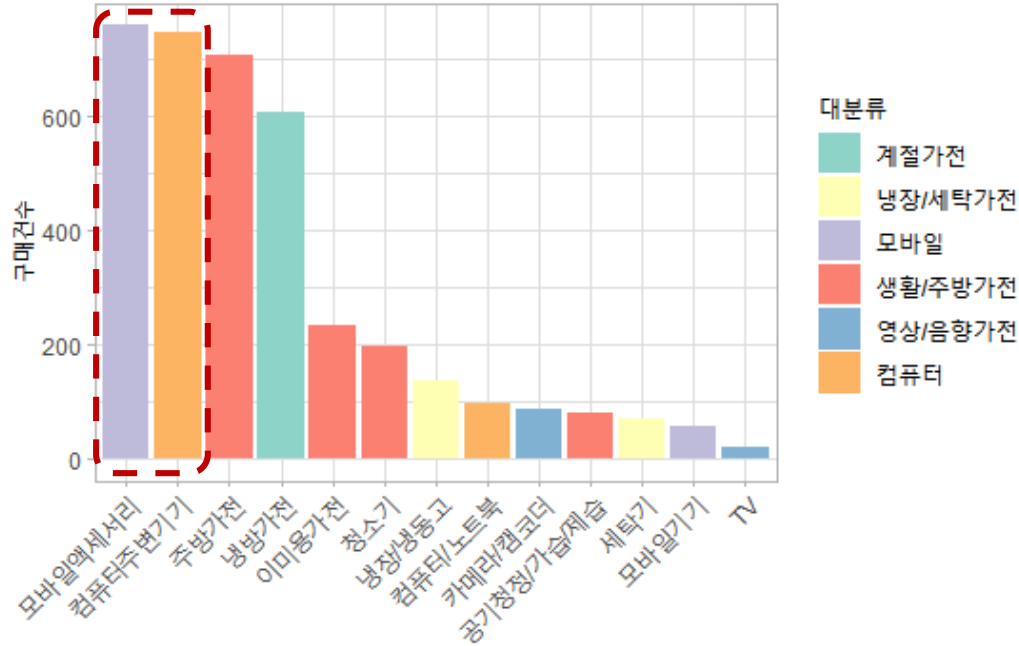
대분류별 구매건수 예측치 비교(3)

4. 가구



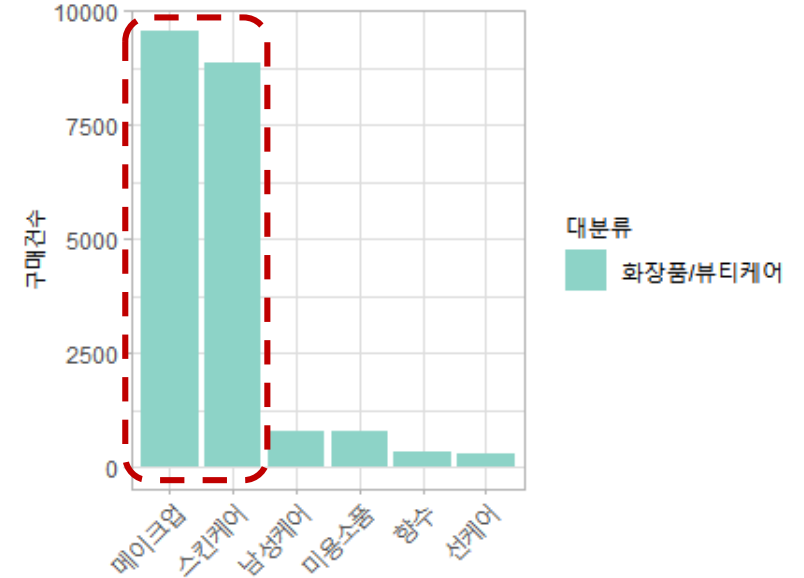
- 시공/ DIY 가구의 선호도가 높아 지는 추세임에도 예상 구매량이 그에 미치지 못하므로 이에 대한 적절한 전략이 필요함

5. 가전/컴퓨터/통신기기



- 계절이 바뀔에 따라 냉방 가전 및 가습/제습 기기 등의 수요가 줄고 모바일/ 컴퓨터 관련 기기의 수요가 증가할 것으로 예상됨

6. 화장품

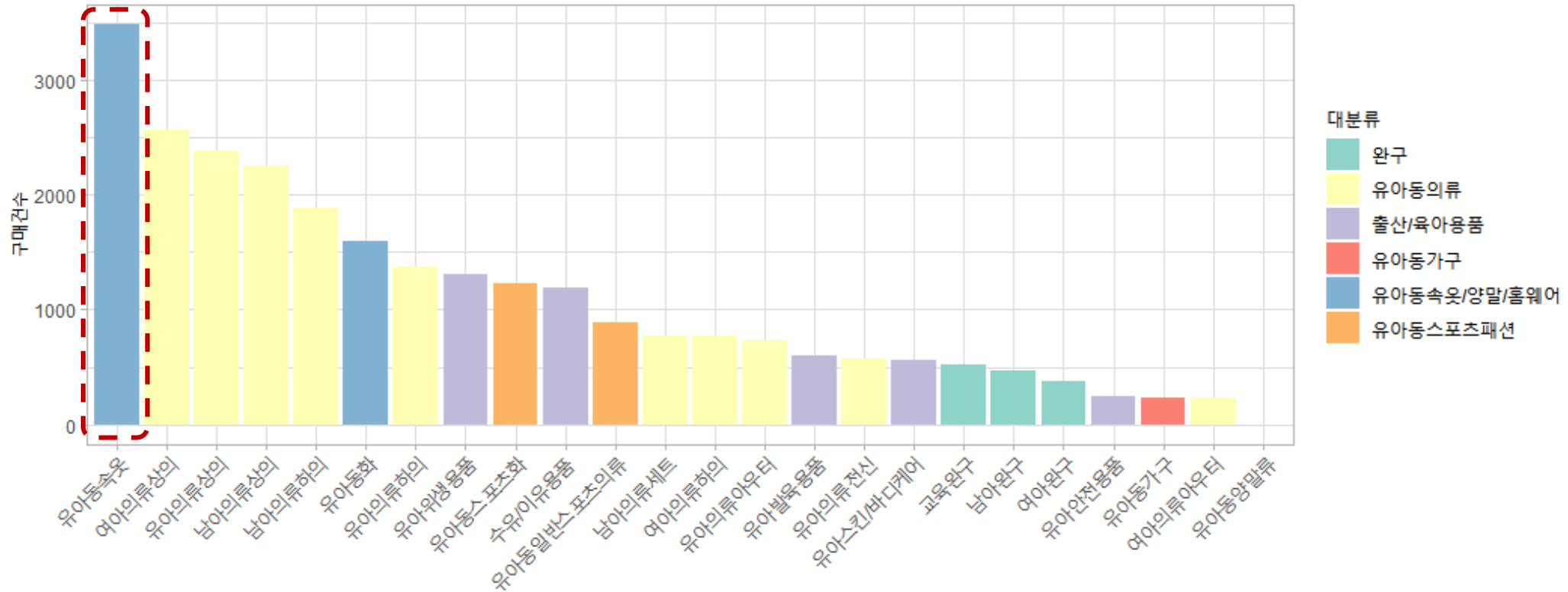


- 메이크업, 스킨케어 에 대한 구매량이 여전히 나머지 항목들보다 압도적으로 높을 것으로 나타남

03 예측 결과

대분류별 구매건수 예측치 비교(4)

7. 아동/유아용품

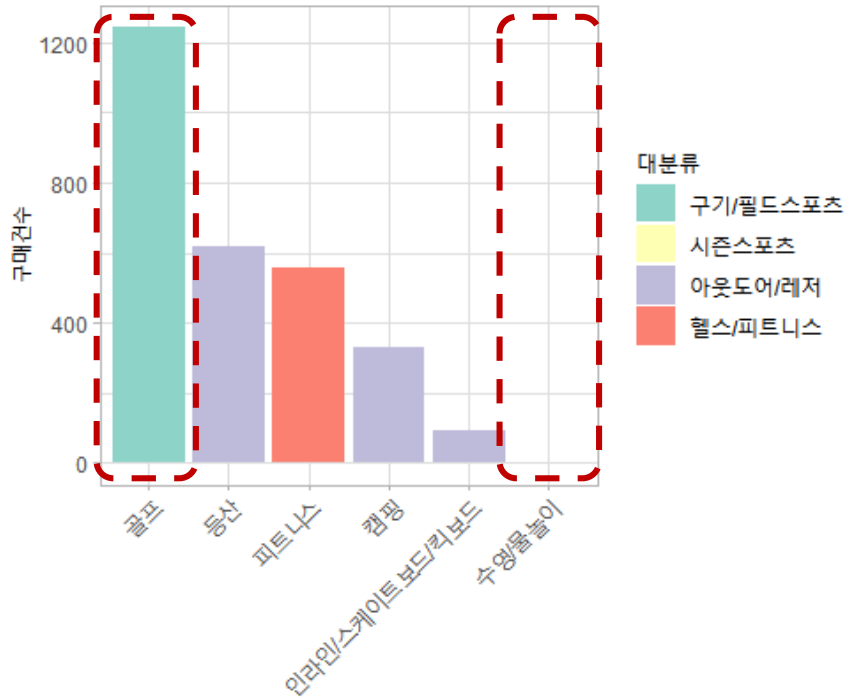


- 계절 변화로 인해 온도가 낮아짐에 따라 내복 등 유아동속옷의 구매량이 급증할 것으로 예상됨

03 예측 결과

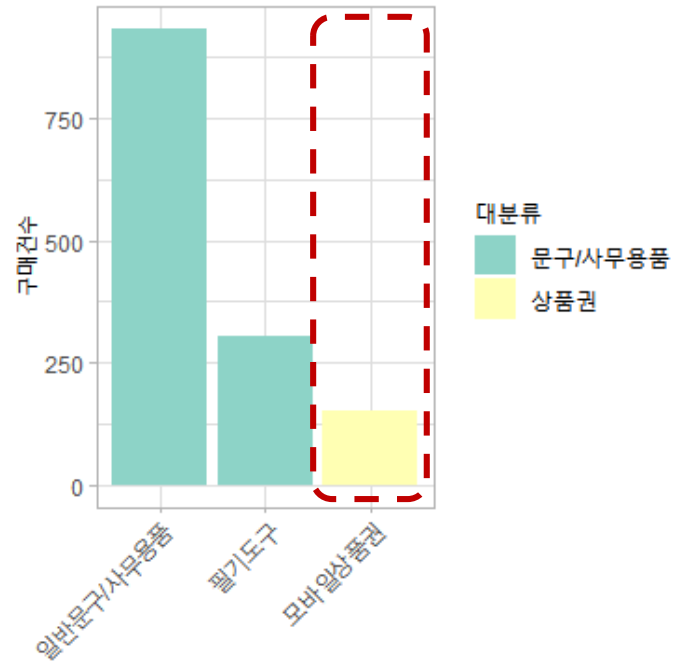
대분류별 구매건수 예측치 비교(5)

8. 스포츠/레저용품



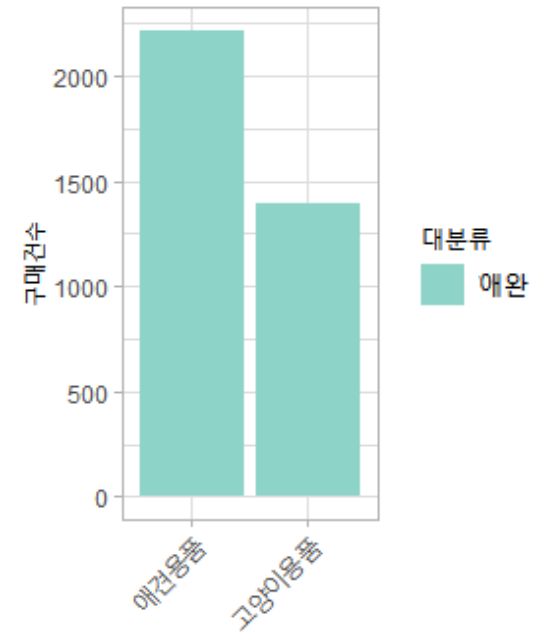
- 수영/ 물놀이, 캠핑 등 여름 스포츠에 관련한 구매는 줄어들고 골프 용품의 구매가 늘어날 것으로 예상됨

9. 사무/문구



- 모바일상품권의 수요가 확연히 줄어들 것으로 예상됨

10. 애완용품



- 고양이용품의 선호도가 증가하는 반면 구매량은 그만큼 높아지지 못할 것으로 예상됨
-> 이전에 비해 더 다양한 고양이용품의 판매가 필요함

CONTENTS

1

데이터 탐색

2

선호지수
생성

3

수요 트렌드
예측

4

서비스 제안

5

결론

01 객단가 높이기 마케팅

다른 상품군 추천 (연관성 분석)

- 방문자의 관심 상품군을 살펴보고 연관성 분석을 통해 같이 구매하는 경향이 높은 상품군을 추천
- 고객의 구매내역을 보고 구매할 가능성이 높은 상품을 추천해줌으로써 객단가를 높이는 전략

Q. 연관성 분석이란 ?

장바구니 분석이라고도 하며, 여러 번 발생한 거래, 사건 내에서 일정한 규칙을 찾아내는 분석

① 지지도 (Support)

전체 거래에서 특정 물품 A와 B가 동시에 거래되는 비중으로, 해당 규칙이 얼마나 의미가 있는 규칙인지를 나타냄
지지도= $P(A \cap B)$: A와 B가 동시에 일어난 횟수/전체 거래 횟수

② 신뢰도 (Confidence)

A를 포함하는 거래 중 A와 B가 동시에 거래되는 비중으로, A라는 사건이 발생했을 때 B가 발생할 확률이 얼마나 높은지를 나타냄

신뢰도= $P(A \cap B)/P(A)$: A와 B가 동시에 일어난 횟수/A가 일어난 횟수

③ 향상도 (Lift)

A와 B가 동시에 거래된 비중을 A와 B가 독립된 사건일 때 동시에 거래된 비중으로 나눈 값

향상도= $P(A \cap B)/(P(A) * P(B))$: A와 B가 동시에 일어난 횟수/A, B가 독립된 사건일 때 A, B가 동시에 일어날 확률

*향상도=1 : A와 B는 아무런 관계가 없다(독립이다)

향상도>1 : A와 B는 강한 상관관계가 있다.

01 객단가 높이기 마케팅

다른 상품군 추천 (연관성 분석)

- 방문자의 관심 상품군을 살펴보고 연관성 분석을 통해 같이 구매하는 경향이 높은 상품군을 추천
- 고객의 구매내역을 보고 구매할 가능성이 높은 상품을 추천해줌으로써 객단가를 높이는 전략

상품의 대분류명 기준 연관성 분석

지지도=0.0001, 신뢰도=0.6으로 연관성 분석을 한 결과 총 10개의 규칙 생성, 10개의 규칙 중 향상도가 1이상인 값 출력

Lhs	rhs	support	confidence	Lift
{생활/자동차용품, 화장품, 스포츠/레저용품, 사무/문구}	→ {의류/패션}	0.00015	0.77	1.13
{아동/유아용품, 스포츠/레저용품, 애완용품}	→ {의류/패션}	0.00011	0.70	1.03
{스포츠/레저용품}	→ {의류/패션}	0.13263	0.70	1.03

- 대분류를 기준으로 봤을 때는 상품군을 다양하게 구입하는 경우 자체가 많지 않기 때문에 지지도를 0.01%로, 신뢰도는 60%로 설정
- 대분류를 기준으로 거래내역 약 28만건에 대해 지지도를 0.01%로 설정했으므로 최소 28건의 조합이 있는 경우에만 결과가 출력

01 객단가 높이기 마케팅

다른 상품군 추천 (연관성 분석)

- 방문자의 관심 상품군을 살펴보고 연관성 분석을 통해 같이 구매하는 경향이 높은 상품군을 추천
- 고객의 구매내역을 보고 구매할 가능성이 높은 상품을 추천해줌으로써 객단가를 높이는 전략

상품의 중분류명 기준 연관성 분석

지지도=0.00006, 신뢰도=0.8으로 연관성 분석을 한 결과 총 4개의 규칙이 생성, 이 때는 3가지 규칙 모두 향상도가 1 이상

lhs	rhs	support	confidence	Lift
{그릇/식기, 여성가방, 향수}	→ {스킨케어}	0.00006	1	10.13
{그릇/식기, 남성케어, 여성가방}	→ {스킨케어}	0.00007	0.95	9.66
{보석, 여성가방, 여성의류아우터}	→ {여성의류상의}	0.00007	0.91	5.43
{여성가방, 여성의류전신, 여성의류하의, 여성화}	→ {여성의류상의}	0.00006	0.82	4.88

- 기준으로 봤을 때는 상품군을 다양하게 구성할 수 있으므로 지지도를 0.006%로 설정하였고 지지도를 낮게 설정한 대신에 신뢰도를 80%로 높게 하여 상품 간의 연관성을 찾고자 함
- 중분류를 기준으로 연관성 분석을 할 때 거래내역이 57만개 정도가 되므로 지지도를 0.006%로 설정할 경우 최소 35개의 거래가 이루어진 경우에만 결과가 출력

01 객단가 높이기 마케팅

묶음 상품 구성 (단변량 분석)

- 2개씩 구입하는 경우가 많은 상품에 대해서는 가격을 조정하고 패키지 상품을 구성하여 판매하여 고객의 객단가를 높이기

상품코드	753343	748308	692709
브랜드	코데즈컴바인 이너웨어	맥	엘라고
제품상세	남성 드로즈 15종 06_CCCDRM92WN / 105	립스틱 - 온 홀드	에어팟 케이스 고리형 키링 행 - 러블리 핑크
구입상품수에 따른 구매빈도			
구매 고객 특성			

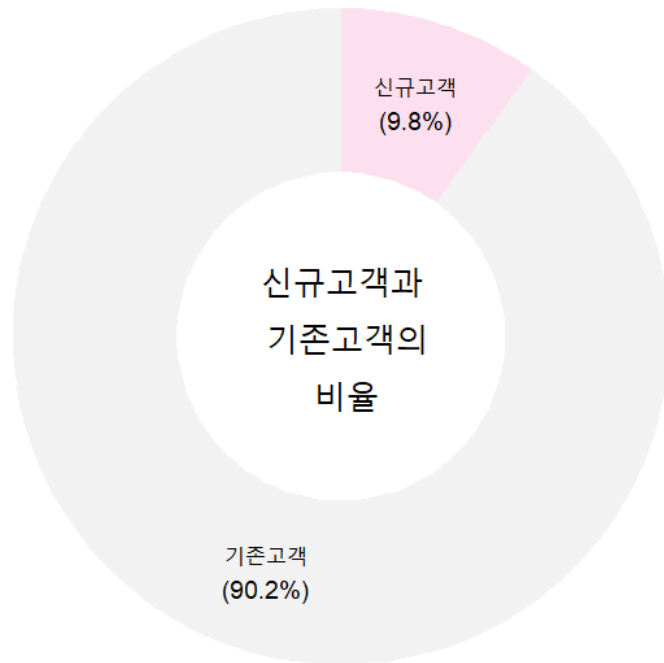


같은 제품을 2개 이상 구입하는 고객수가 많기 때문에 패키지 상품에 대한 수요가 있는 것으로 보이며 1+1 또는 할인된 패키지상품을 제시

02 고객 Segmentation

신규고객 vs 기존고객

- 신규고객과 기존고객의 정의



[신규고객]

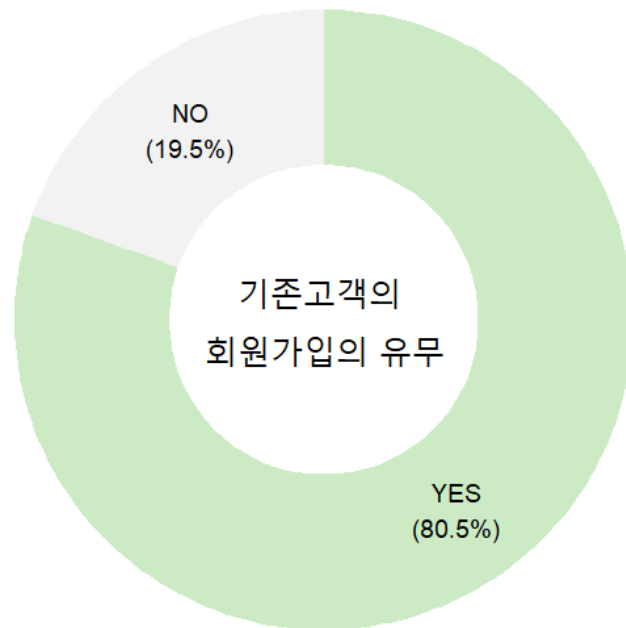
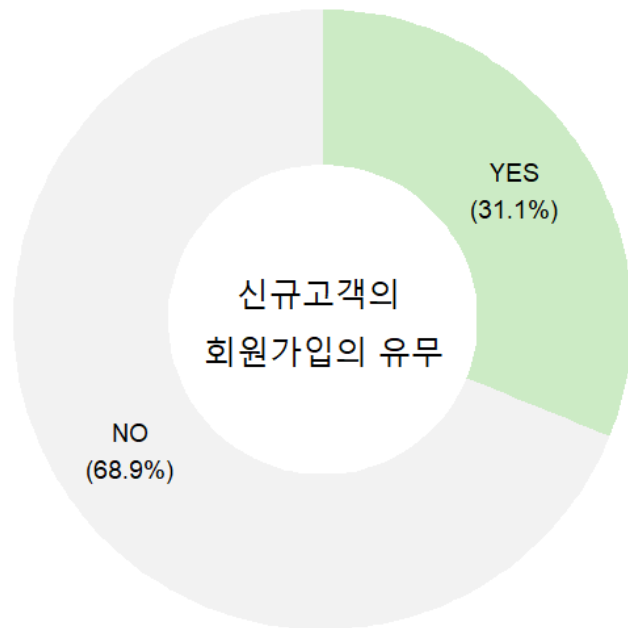
- 분석데이터 기간(2018년 4월 1일 ~ 2018년 9월 30일) 동안 처음으로 롯데 온라인 계열사를 방문하여 구매 및 이용을 한 고객
- SESS_SEQ = 1인 고객

[기존고객]

- 분석데이터 기간 이전부터 롯데 온라인 계열사를 방문하여 구매 및 이용을 한 고객

02 고객 Segmentation

신규고객 vs 기존고객

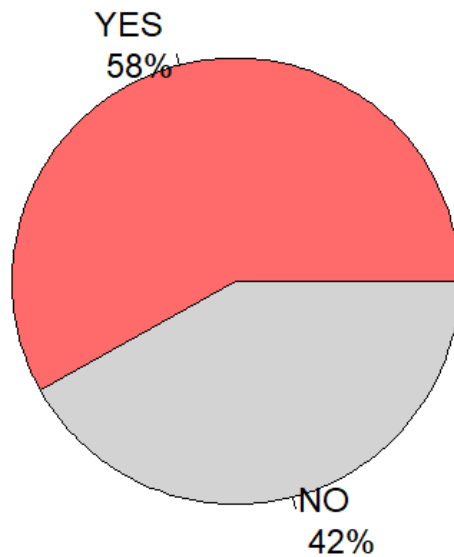


신규고객의 경우 제품구입시 비회원으로 하는 경우가 월등히 많은 반면,
기존고객 중에는 회원인 비율이 훨씬 높음
신규고객 중 회원고객의 재방문확률이 비회원고객의 재방문할 확률에 비해 약 **1.9배** 더 높음
신규고객의 **회원가입을 유도**하여 재방문할 확률을 높이는 방안을 제시할 수 있음

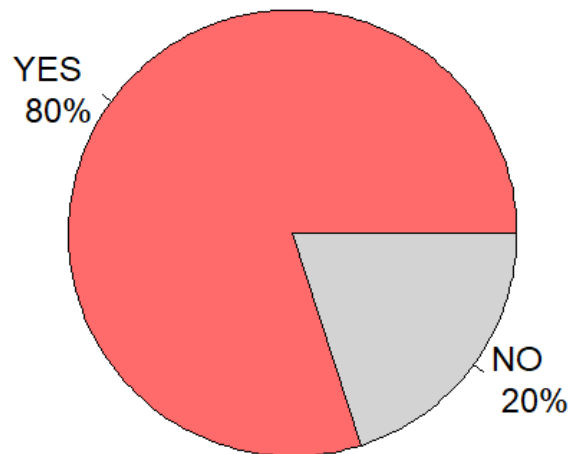
02 고객 Segmentation

신규고객 vs 기존고객

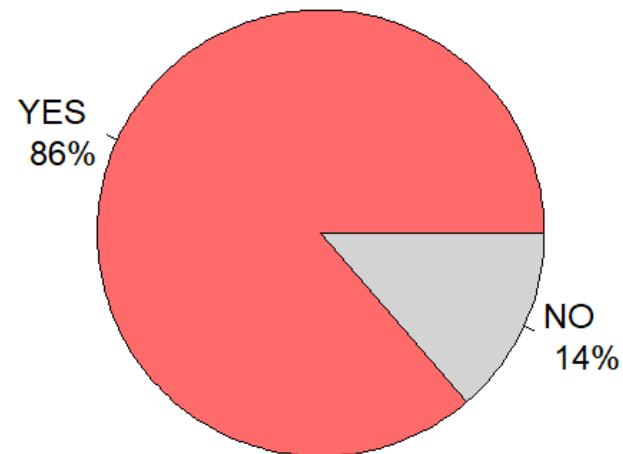
회원가입 여부
(한번 방문한 경우)



회원가입 여부
(2번 방문한 경우)



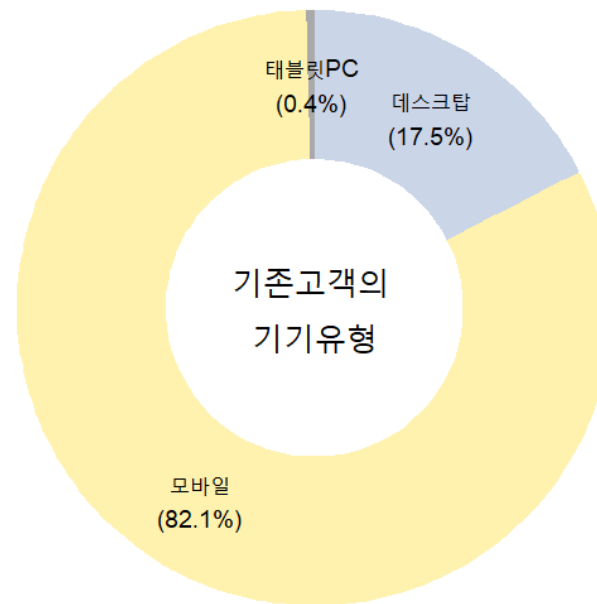
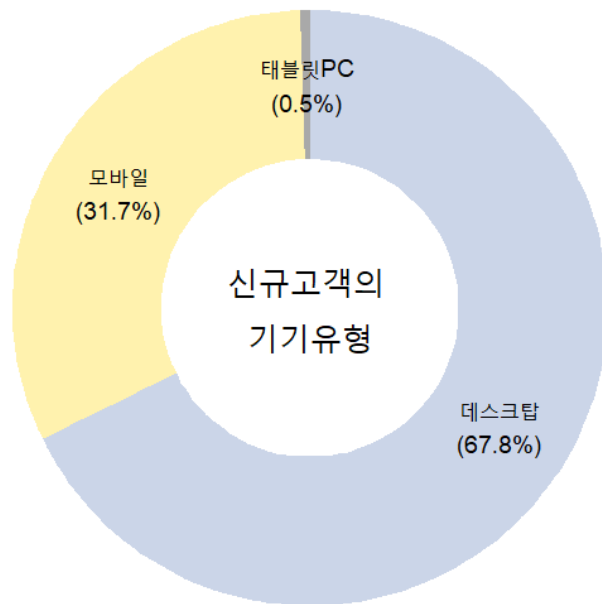
회원가입 여부
(3번 이상 방문한 경우)



방문횟수가 늘어날수록 회원가입을 하는 고객의 수가 꾸준히 증가함

02 고객 Segmentation

신규고객 vs 기존고객



신규고객의 경우 데스크탑을 통해 구매를 하는 경우가 많은 반면, 기존고객의 대부분은 모바일을 이용
신규고객 중 모바일을 주로 사용하는 고객의 재방문 확률이
다른 기기를 이용하는 고객의 재방문 확률보다 **1.9배** 더 높음
모바일 사용을 유도함으로써 재방문율을 높일 수 있음

02 고객 Segmentation

신규고객 vs 기존고객

- 신규고객과 기존고객을 위한 신규 서비스 전략 제안

	검색건수 (건)	조회수 (건)	뷰 시간(초)	구매금액 총합(원)	구매금액 평균(원)	세일제품 구매율(%)	구매품목 종류(개)	구매건수 (건)	방문횟수 (회)	브랜드 선호도
신규고객	2.53	52.6	37.2	191,551	64,758	3.22	1.29	4.14	2.81	0.705
기존고객	2.00	67.7	24.5	304,158	63,748	3.09	1.75	6.63	29.3	0.711

[신규고객을 위한 서비스 전략]

- 신규고객의 특징 : 검색을 많이 하고 관심 있는 상품에 대한 정보를 자세히 보는 경향이 있음
- 고객의 주목도가 높은 상품에 대한 정보 제공
- 방문횟수를 늘리기 위해 모바일 사용 촉진

[기존고객을 위한 서비스 전략]

- 기존고객의 특징 : 검색보다는 클릭을 통해 관심 상품을 보고 자주 방문하는 경향이 있음
- 대부분의 기존고객이 회원가입을 통해 제공하는 기본정보를 바탕으로 직접적인 광고
- 기존 고객을 유지하기 위해서 오프라인 경험이 중요 하기 때문에 오프라인 할인쿠폰을 제공

ex) 롯데백화점 oo브랜드 남성아우터 30% 할인 쿠폰을 제시하여 매장 방문을 유도

02 고객 Segmentation

기존고객의 유형 분석

- 고객의 행동 데이터를 기반으로 고객유형을 분류하여 각 유형에 적합한 서비스를 제공



온라인 활동

뷰 시간
검색건수
조회수
방문횟수

뷰 시간 : 세션 내의 페이지(화면) 당 뷰 시간
검색건수 : 세션 내 검색어 검색 건수
조회수 : 방문자의 클릭, 검색 등의 행위에 대한 일련번호
방문횟수 : Web/App에 접속하는 횟수



소비규모

구매금액총합
구매금액평균
구매건수

구매금액총합 : 세션 내 구매한 상품의 총 금액
구매금액평균 : 세션 내 구매한 상품의 평균 금액
구매건수 : 세션 내 구매한 상품의 총 수량



소비유형

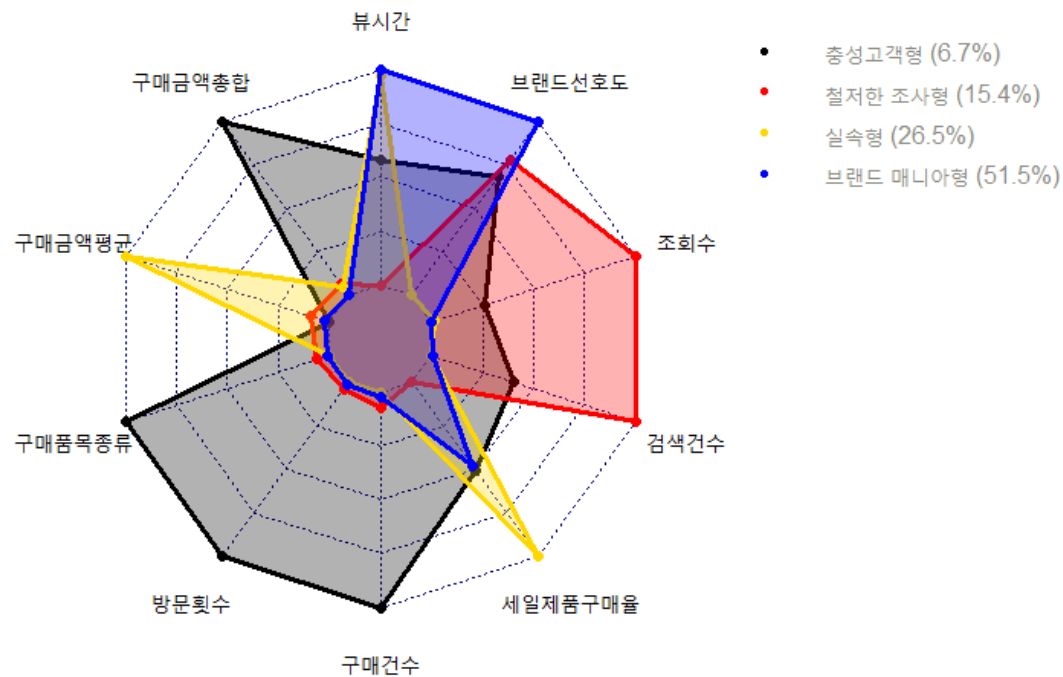
구매품목종류
세일제품구매율
브랜드선호도

구매품목종류 : 세션 내 구매한 상품의 대분류별 다양성
예) 여성의류상의와 애완용품을 구매한 경우 -> 구매품목종류: 2
예) 여성의류상의와 남성의류하의를 구매한 경우 -> 구매품목종류: 1
세일제품구매율 : 세션 내 구입한 상품 중 세일인 상품의 비율
브랜드선호도 : 세션 내 구매한 상품 중 브랜드선호도가 있는 상품의 비율
**브랜드선호도: 중분류별 구매건수가 가장 많았던 상위 10%의 상품에 브랜드선호도를 부여*

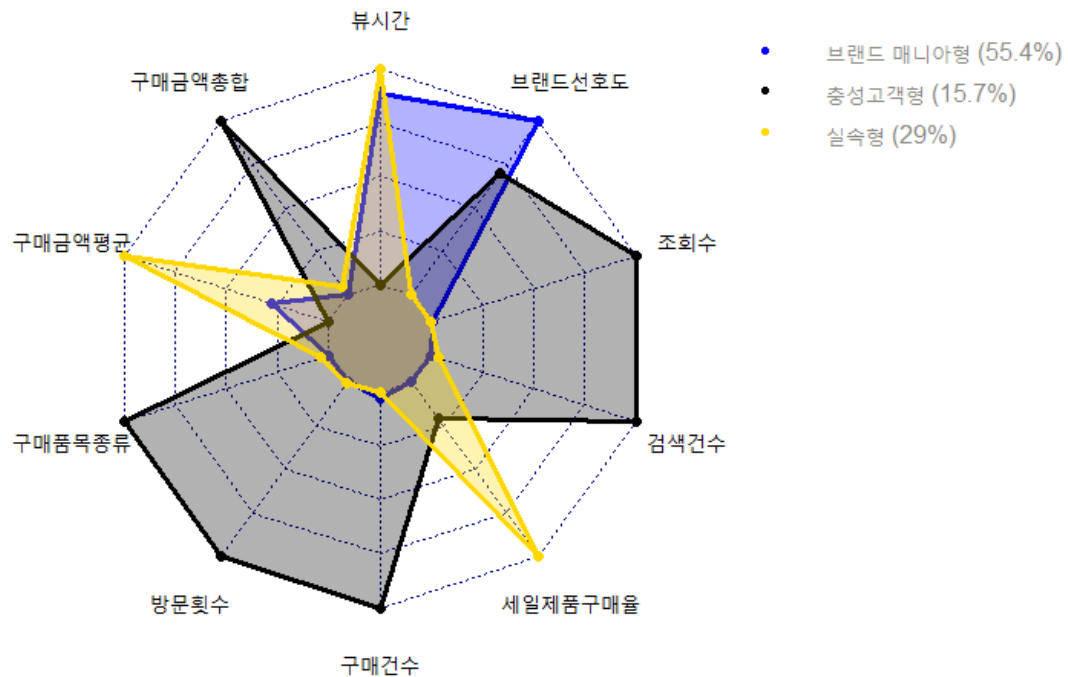
02 고객 Segmentation

고객 유형 분석

회원인 고객의 유형



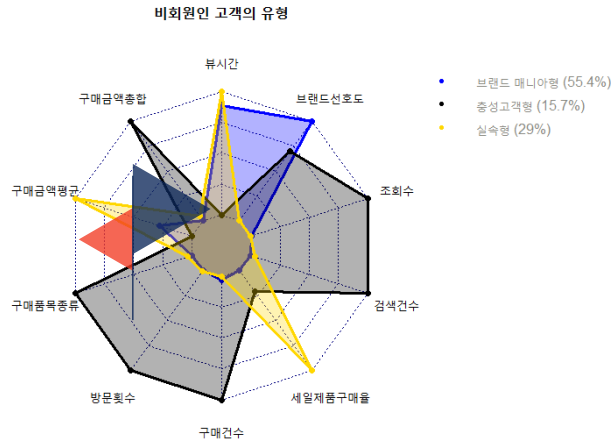
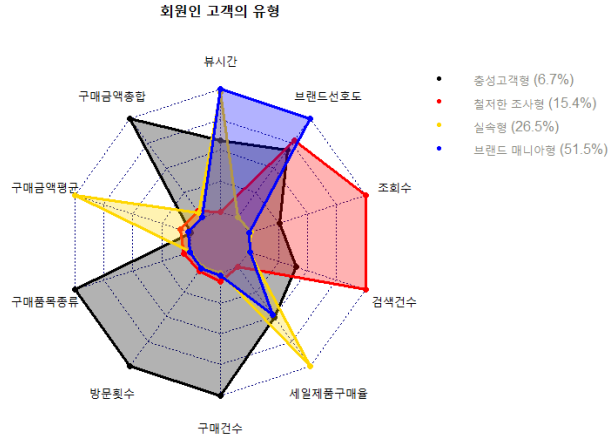
비회원인 고객의 유형



고객의 특성에 따른 고객유형을 분류하여 각 유형에 적합한 서비스를 제공하자

02 고객 Segmentation

고객 유형 분석



[충성고객형]

- 고객당 구매금액의 총합을 기준으로 고객등급을 나누고 할인쿠폰과 포인트 적립률을 차등지급해 브랜드 로열티 높임

[철저한 조사형]

- 상품후기 작성 이벤트 실시해서 구매 전환률 높이는 전략
- 상세 페이지를 통해 간결하고 필수적인 정보 제공을 통해 이탈방지

[실속형]

- 세일 이벤트 진행 시 정보 제공
- 가급적이면 많은 정보를 제공해 이탈방지

[브랜드 매니아형]

- 인지도 있는 브랜드 품목을 상위에 추천
- 과도한 정보제공보다는 사진, 문구 등의 최소한의 정보제공으로 감정적인 어필이 중요



고객의 특성에 따른 고객유형을 분류하여 각 유형에 적합한 서비스를 제공하자

CONTENTS

1
데이터 탐색

2
선호지수
생성

3
수요 트렌드
예측

4
서비스 제안

5
결론

데이터 분석 과정 요약

▶ 상품군별 온라인 선호지수 생성

- 소비자의 온라인 행동 패턴을 기반으로 상품군(중분류)에 따른 선호도에 영향을 주는 잠재변수 탐색
- 요인 분석을 통해 주별 및 전체 기간의 온라인 선호지수 산출

▶ 수요 트렌드 예측 및 인사이트 도출

- 통계 모형(Bass)과 LSTM 모형을 적용하여 10월 첫째 주의 구매량 및 온라인 선호지수 예측
- 온라인 선호지수의 트렌드 및 수요 예측 결과를 다각도로 살펴봄으로써 마케팅에 필요한 인사이트 도출

▶ 신규 서비스 제안

- 연관성 분석을 통한 제품 추천과 단변량 분석을 통한 묶음상품 구성 제안 등 객단가를 높일 수 있는 마케팅 전략 제시
- 신규고객과 기존고객을 위한 마케팅 전략 제시
- 고객의 온라인 행동 및 소비패턴 분석을 통해 유형별 맞춤 전략 제시