



오늘의집

집들이 선물, 뭐가 좋을까?

내일의 집

김민지 김예원 박유정
송나현 이소현 이지혜 최수연



목차

01 주제 선정

02 분위기 분석

- 2.1. 집들이 컨텐츠에 소개된 상품
- 2.2. 집들이 컨텐츠에 소개되지 않은 상품

03 상품 추천

- 3.1. 상품 정렬 - 리뷰 감성 분석
- 3.2. 상품 조합 - 연관 분석
- 3.3. 내일의 집 vs 오늘의 집



01

주제 설정

QUEST

“이번주 주말, 수연이의 집들이에 가야 한다.”



위기: 휴지와 세제가 금지되었다.

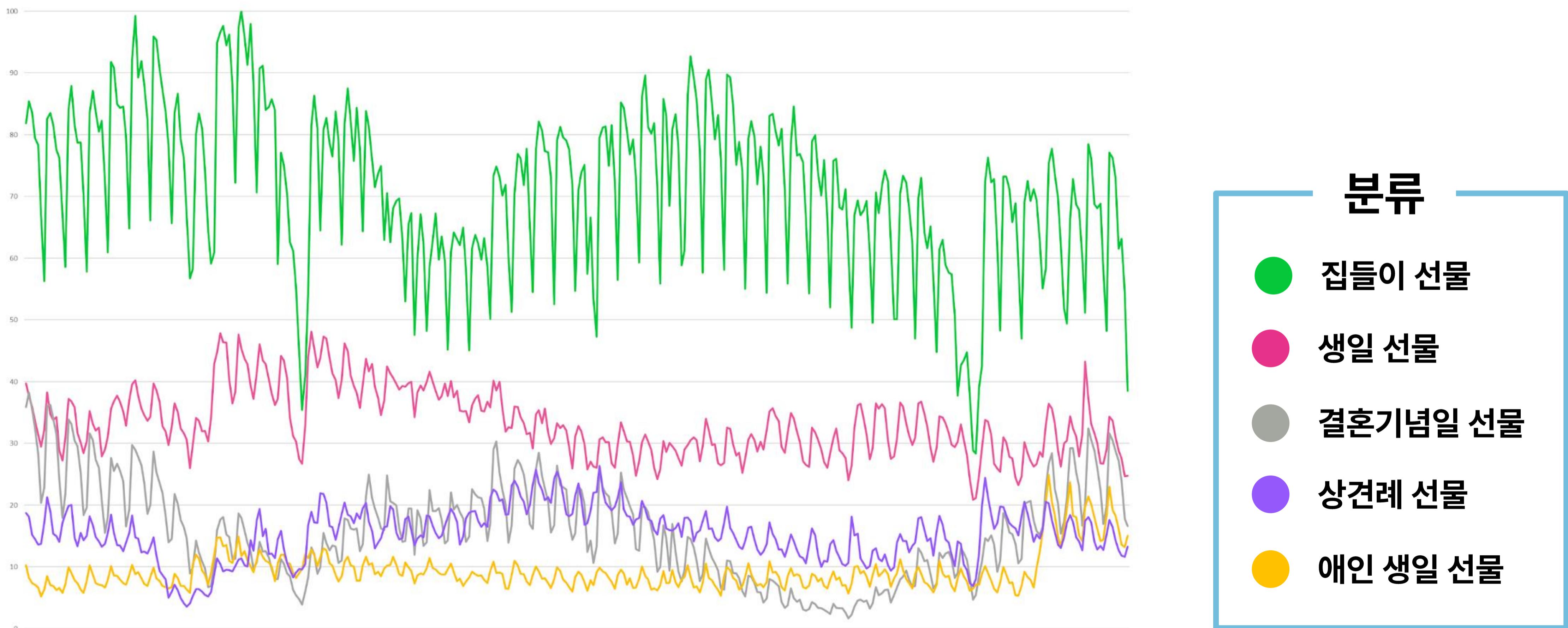


“ 집들이 선물 고민, 나만 하는 건가…? ”



생일보다 2배 더 고민되는 집들이 선물

'선물' 관련 네이버 검색량 상대 비교



2021.11.01 ~ 2022.10.30
동음이의어 '선물(先物)' 관련 키워드 제외

**모두가 고민하는 집들이 선물,
오늘의집에서 추천 받아보자!**



인테리어 상품, 뭘 보고 고르지?

인테리어 상품 구매 시 중요 고려 요소 TOP 10

1	집안 분위기와의 어울림	274
2	디자인	259
3	가격	216
4	품질	177
5	소재	96
6	기능	90
7	후기	88
8	할인	59
9	색상	58
10	브랜드	53

2022.10.04
오픈 서베이 조사 / N = 500
15개 중 택 3 답변

분위기로 상품을 고를 수 있다면 더 좋지 않을까?

02

분위기 분석

1. 집들이 컨텐츠에 소개된 상품



오늘의집 커뮤니티 살펴보기

오늘의집

커뮤니티

스토어

이사/시공/수리



로그인

회원가입

글쓰기

홈

팔로잉

사진

집들이

노하우

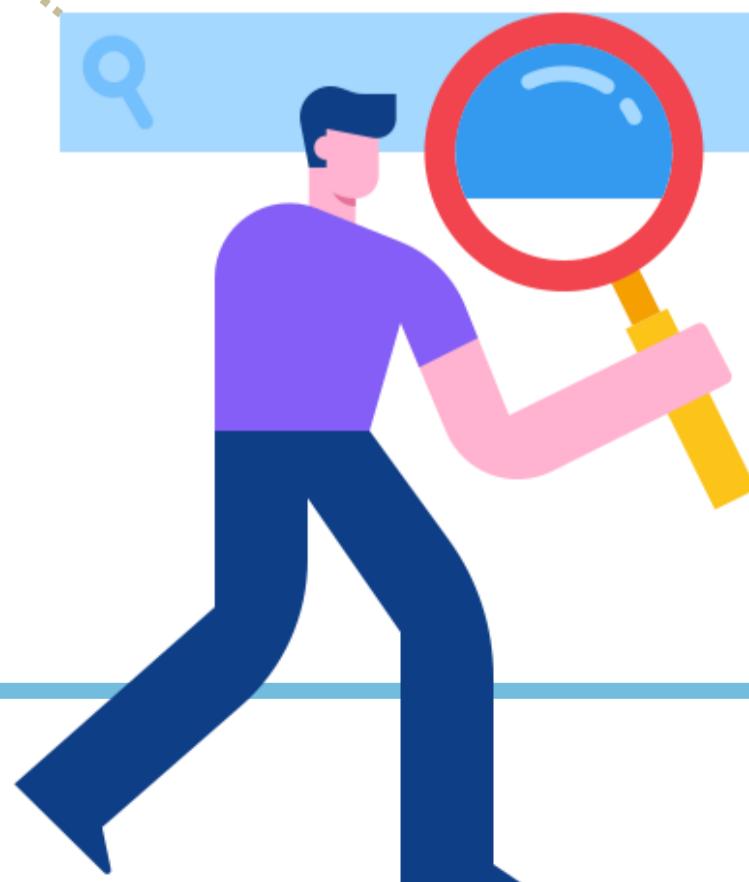
전문가집들이

셀프가이드

질문과답변

3D인테리어 BETA

이벤트



2.1. 집들이 컨텐츠에 소개된 상품

집들이 데이터

온전한 나만의 공간은 그 자체로 큰 위로니까요

branch_222 3주 전

3초컷! 집들이 미리보기

공간 기타 평수 32평 작업 셀프•DIY 분야 홈스타일링 가족형태 부모님과 함께 사는 집
지역 서울특별시 송파구 스타일 내추럴 기간 24개월 예산 100만원

기타 | 32평 | 홈스타일링 | 부모님과 함께

공간 기타 평수 32평 작업 셀프•DIY 분야 홈스타일링 가족형태 부모님과 함께 사는 집
지역 서울특별시 송파구 스타일 내추럴 기간 24개월 예산 100만원

2.1. 집들이 컨텐츠에 소개된 상품

집들이 데이터

The screenshot shows a user interface for a home renovation platform. At the top, there's a navigation bar with links like '오늘의집' (Today's Home), '커뮤니티' (Community), '스토어' (Store), and '이사/사공/수리' (Relocation/Construction/Maintenance). A search bar and a login area are also present. Below the navigation, a banner reads '온전한 나만의 공간은 그 자체로 큰 위로니까요' (A complete space for yourself is its own great comfort). The main content area features a large image of a bedroom with green bedding, a wooden cabinet, and many potted plants. Below the image, there's a section titled '3초컷! 집들이 미리보기' (3-second cut! Preview of moving in) with three small thumbnail images. A red box highlights a specific product overlay on the image: '스타일링 품 호텔식 고밀도 극세사 겨울 차렵이불단품(SS/Q/K)... 65,500원' (Style Product Hotel-style high-density quilted winter duvet SS/Q/K... 65,500 won). At the bottom of the page, there are more thumbnails and a footer with links like '的空间' (Space), '装修' (Renovation), '设计' (Design), and '家居' (Home).

The screenshot shows a detailed view of a house listing. At the top, there are icons for '기타' (General), '32평' (32 square meters), '홈스타일링' (Home Styling), and '부모님과 함께' (With parents). Below these are the main details: '공간 기타 평수 32평 작업 셀프•DIY 분야 홈스타일링 가족형태 부모님과 함께 사는 집 지역 서울특별시 송파구 스타일 내추럴 기간 24개월 예산 100만원'. The main image is a bedroom with green bedding and many plants. A red box highlights a product overlay on the image: '스타일링 품 호텔식 고밀도 극세사 겨울 차렵이불단품(SS/Q/K)... 65,500원' (Style Product Hotel-style high-density quilted winter duvet SS/Q/K... 65,500 won). The bottom part of the page shows a summary of the house: '온전한 나만의 공간은 그 자체로 큰 위로니까요' (A complete space for yourself is its own great comfort) and a list of features: '온전한 나만의 공간은 그 자체로 큰 위로니까요' (A complete space for yourself is its own great comfort), '작업 셀프•DIY' (Workshop Self-DIY), '분야 홈스타일링' (Field Home Styling), '가족형태 부모님과 함께 사는 집' (Family Type House where parents live together), '지역 서울특별시 송파구' (Region Seoul Special City Songpa-gu), '스타일 내추럴' (Style Natural), '기간 24개월' (Term 24 months), and '예산 100만원' (Budget 1,000,000 won).

게시물에 태깅된 상품 데이터

2.1. 집들이 컨텐츠에 소개된 상품

The screenshot shows a user profile 'branch_222' sharing a room tour titled '온전한 나만의 공간은 그 자체로 큰 위로니까요'. Below the tour, there are two recommended products: a '비슷한' (similar) item for 65,500원 and another '비슷한' item.

The screenshot shows a user profile 'branch_222' sharing a room tour titled '온전한 나만의 공간은 그 자체로 큰 위로니까요'. Below the tour, there are two recommended products: a '비슷한' (similar) item for 65,500원 and another '비슷한' item.

집들이 데이터

게시물에 태깅된 상품 데이터

상품 상세 데이터

스타일링홈
호텔식 고밀도 극세사 겨울 차렵이불단품(SS/Q/K) 7colors
★★★★★ (28)
37% 104,100 원
65,500원 특가
55,675원 첫구매할인가 (앱 전용) >

2.1. 집들이 컨텐츠에 소개된 상품

상품 분위기 라벨링

상품명	분위기	북유럽	빈티지 & 레트로	클래식 & 앤틱	프렌치 & 프로방스	러블리 & 로맨틱	인더스트리얼	한국 & 아시아	유니크 & 믹스매치
미니 2단 모듈선반 화장품정리대/진열대 7colors	0	0	1	0	0	0	0	0	0
BILLY 책장 (80X28X106)	2	4	0	0	3	0	0	0	1
스페인직수입 소파커버 12colors	2	2	0	0	0	0	0	0	0
인테리어 촬영 소품 모형책 가짜책 감성 잡지	0	1	1	1	1	0	0	0	1
[해외] 눈굴리는 곰돌이 빙수기 4종	0	1	0	0	3	0	0	0	0
세라믹 오브제 트레이 다용도 바구니 바스켓 카페 인테리어 소품	1	0	0	0	0	0	0	0	0

:

합계	308	556	144	69	170	24	67	269
----	-----	-----	-----	----	-----	----	----	-----

2.1. 집들이 컨텐츠에 소개된 상품

상품 분위기 라벨링

상품명	분위기	북유럽	빈티지 &레트로	클래식 &앤틱	프렌치 &프로방스	러블리 &로맨틱	인더스트리얼	한국 &아시아	유니크 &믹스매치
미니 2단 모듈선반 화장품정리대/진열대 7colors		0.000000	0.000000	0.006944	0.000000	0.000000	0.0	0.000000	0.000000
BILLY 책장 (80X28X106)		0.006494	0.007194	0.000000	0.000000	0.017647	0.0	0.000000	0.003717
스페인직수입 소파커버 12colors		0.006494	0.003597	0.000000	0.000000	0.000000	0.0	0.000000	0.000000
인테리어 촬영 소품 모형책 가짜책 감성 잡지		0.000000	0.001799	0.006944	0.014493	0.005882	0.0	0.000000	0.003717
[해외] 눈굴리는 곰돌이 빙수기 4종		0.000000	0.001799	0.000000	0.000000	0.017647	0.0	0.000000	0.000000
세라믹 오브제 트레이 다용도 바구니バスケット 카페 인테리어 소품		0.003247	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000

02

분위기 분석

2. 집들이 컨텐츠에 소개되지 않은 상품

태깅되지 않은 상품의 분위기는 어떻게 알 수 있을까?

2.2. 집들이 컨텐츠에 소개되지 않은 상품

집들이 컨텐츠 제목에 나타나는 키워드



네모 창에 담긴 이야기가 기대되는 집,
담랑채

담랑채

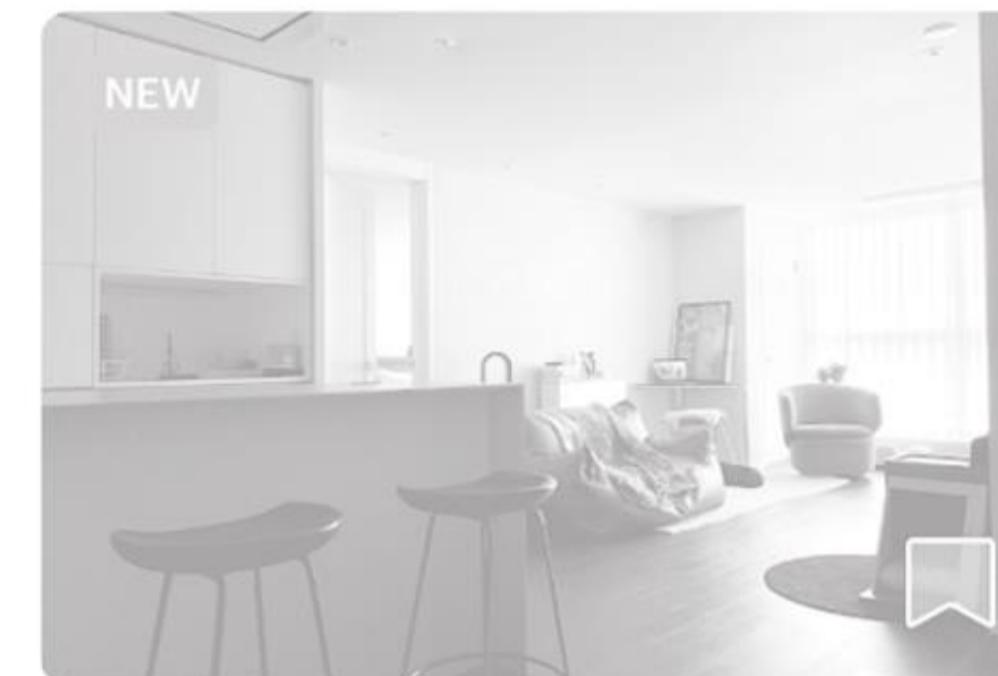
스크랩 95 · 조회 3,129



채광 맛집, 신혼부부의 취향 담긴
미드센추리모던 하우스

Ahome

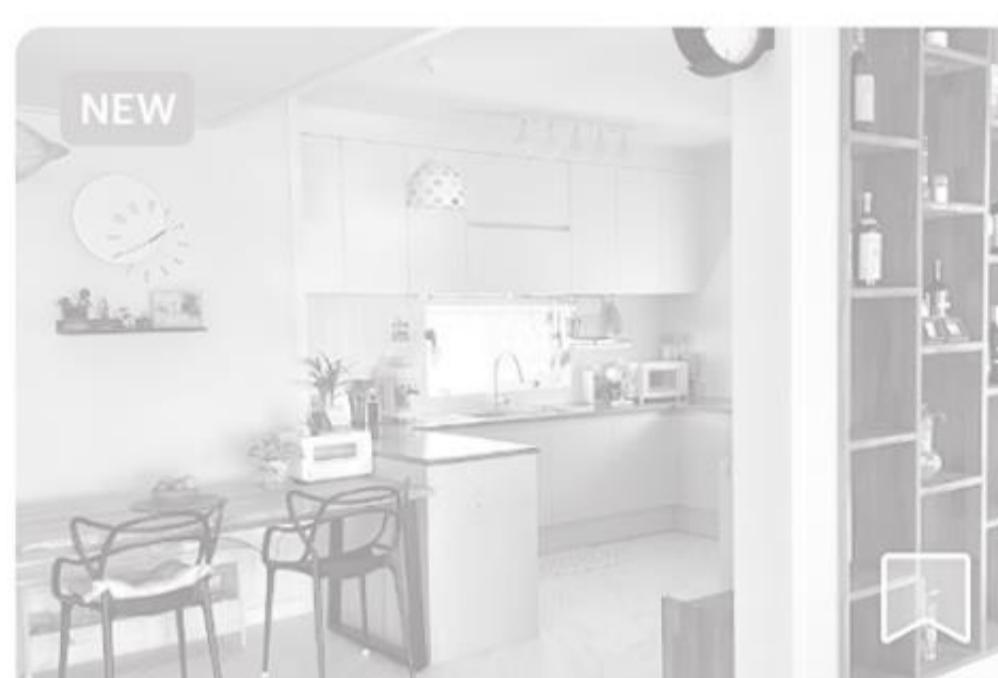
스크랩 36 · 조회 1,770



빈티지한 색감과 가구 배치가 조화로운
23평 아파트

h.apening

스크랩 52 · 조회 1,945



만들어가는 재미가 있는 세 가족의 주택



Erica.home

스크랩 28 · 조회 1,483



구축 아파트를 화이트 홈스타일링으로
완성한 싱글 라이프

_.illang

스크랩 19 · 조회 1,295



유니크한 컬러 플레이로 홈 스타일링한,
신혼집 인테리어

in_eirene

스크랩 307 · 조회 12,112



2.2. 집들이 컨텐츠에 소개되지 않은 상품

분위기별로 많이 등장하는 키워드

<빈티지 & 레트로>

핑크
호텔 드림
감성
스윗
쁘작
뉴트럴
리조트
귀염
가득한
미드
로망
사랑스러운
우아한
오래된
내추럴
등화
스튜디오
하얀
프렌치
여기
우아한
집온
한국
혹시
해외
주인공라운지
카페
츠집
오피
영강
국적
베당
유럽
블랙
행빛
맑은
공존
화사한
취미
모던
등록
따뜻한
빈티지
핑크

<프렌치 & 프로방스>

<클래식 & 앤틱>

<러블리 & 로맨틱>

2.2. 집들이 컨텐츠에 소개되지 않은 상품

분위기별 특징 키워드 선정

러블리 & 로맨틱

'러블리', '로맨틱', '핑크', '귀여-', '사랑스-', '따뜻'

빈티지 & 레트로

'빈티지', '레트로', '원목', '브릭'

프렌치 & 프로방스

'프렌치', '프로방스'

클래식 & 앤틱

'클래식', '앤틱', '앤틱', '엔티크', '엔티크', '우아', '엘레강스'

유니크 & 믹스매치

'유니크', '믹스매치', '개성', '컬러풀'

북유럽

'북유럽', '차분'

한국 & 아시아

'한국', '아시아', '대나무', '전통', '동양'

인더스트리얼

'인더스트리얼', '공장'

**상품의 리뷰에 분위기를 나타내는 키워드가 있으면
그 상품의 분위기를 정할 수 있지 않을까?**

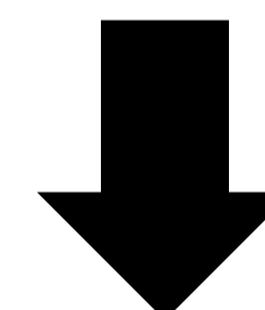
2.2. 집들이 컨텐츠에 소개되지 않은 상품

러블리 & 로맨틱 상품 리뷰에 정말 관련 키워드가 등장할까?

리뷰 스크래핑



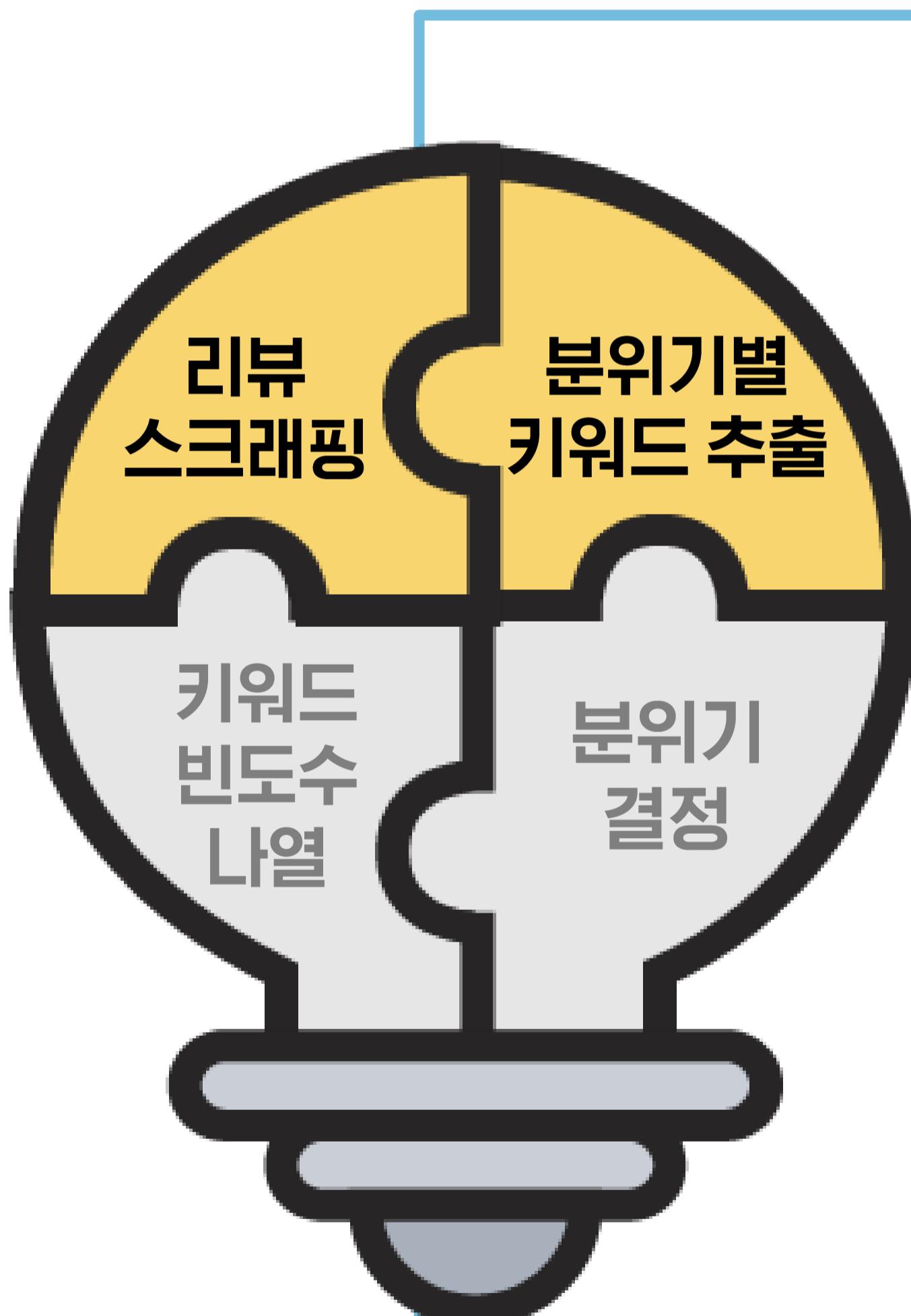
러블리 & 로맨틱



“상품 1,142개 · 총 114,200개의 리뷰 수집”

2.2. 집들이 컨텐츠에 소개되지 않은 상품

러블리 & 로맨틱 상품 리뷰에 정말 관련 키워드가 등장할까?



분위기별 키워드 추출

러블리 & 로맨틱 상품

상품정보

리뷰 5

문의 18

배송/환불

추천

집안이 **컬러풀** 해졌어요 인테리어 소품으로 여기저기 딱 이예요

분위기가 살아났어요 **유니크**하고 좋아요~

향도 좋고 인테리어 소품으로도 예쁘고 엠보싱같아 **사랑스러워요**

알찬구성에 색상도 완전 짱~~^-^ 좋아하는 **핑크색**이라서 더 **사랑스럽네요!**

너무 **사랑스럽고 귀여워요** 지인 집들이 선물하기도 좋을 것 같아요.

2.2. 집들이 컨텐츠에 소개되지 않은 상품

러블리 & 로맨틱 상품 리뷰에 정말 관련 키워드가 등장할까?



키워드 빈도수 나열

----- ['러블리', '로맨틱', '핑크', '귀여', '사랑스', '따뜻'] -----
러블리 39
로맨틱 9
핑크 543
귀여 3914
사랑스 117
따뜻 844
5466 ----- ['빈티지', '레트로', '원목', '브릭'] -----
빈티지 383
레트로 113
원목 570
브릭 185
1251 ----- ['프렌치', '프로방스'] -----
프렌치 4
프로방스 2
6 ----- ['클래식', '앤틱', '앤틱', '엔티크', '앤틱', '우아', '엘레강스'] -----
클래식 37
앤틱 53
앤틱 33
엔티크 1
앤틱 3
우아 44
엘레강스 2
173

----- ['유니크', '믹스매치', '개성', '컬러풀'] -----
유니크 47
믹스매치 2
개성 3
컬러풀 6
58 ----- ['북유럽', '차분'] -----
북유럽 3
차분 137
140 ----- ['한국', '아시아', '대나무', '전통', '동양'] -----
한국 13
아시아 0
대나무 14
전통 5
동양 0
32 ----- ['인더스트리얼', '공장'] -----
인더스트리얼 0
공장 16
16

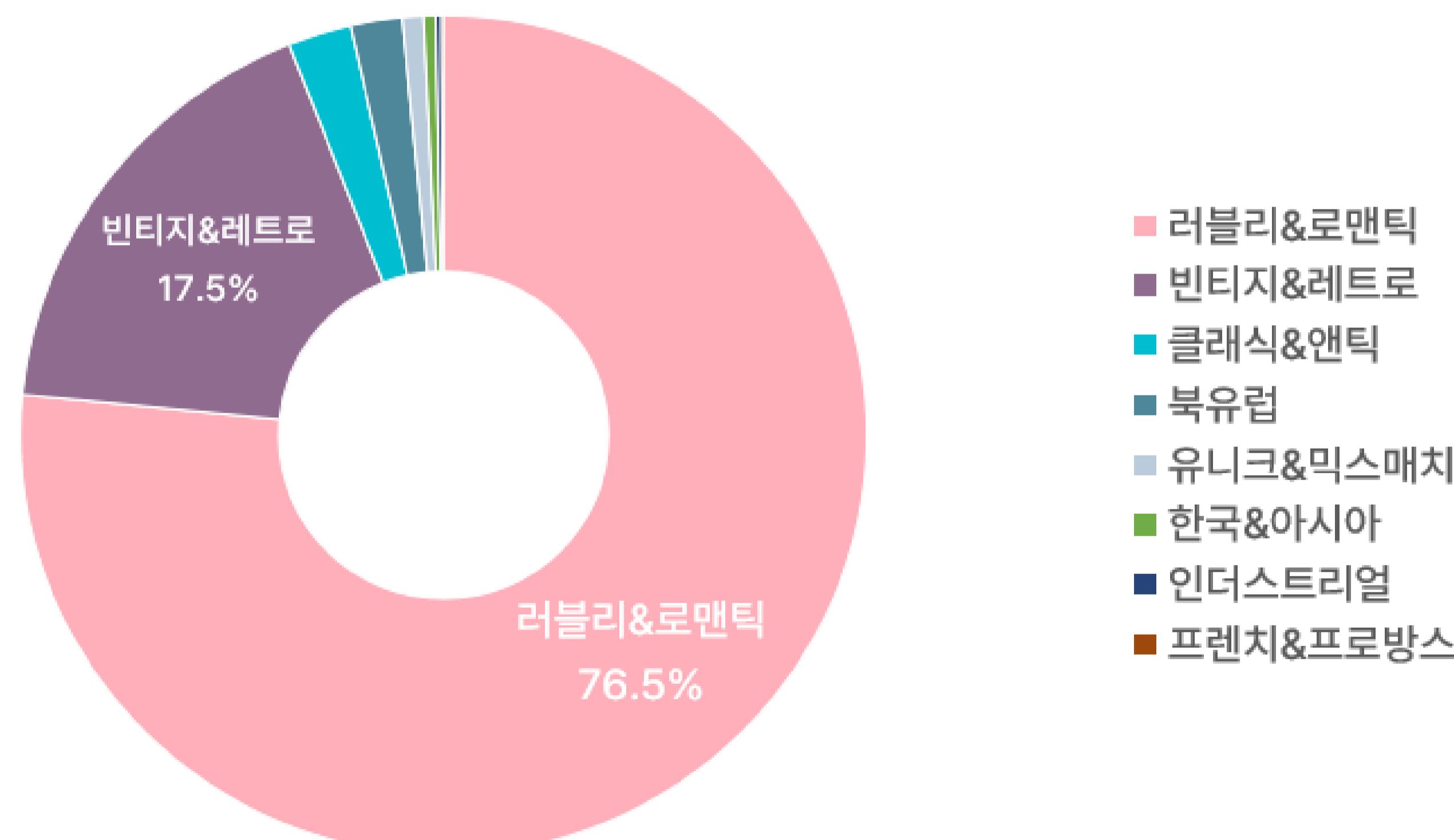
2.2. 집들이 컨텐츠에 소개되지 않은 상품

러블리 & 로맨틱 상품 리뷰에 정말 관련 키워드가 등장할까?



키워드
빈도수 나열

러블리 & 로맨틱 상품 리뷰에
등장하는 분위기별 키워드 비중



2.2. 집들이 컨텐츠에 소개되지 않은 상품

러블리 & 로맨틱 상품 리뷰에 정말 관련 키워드가 등장할까?



03

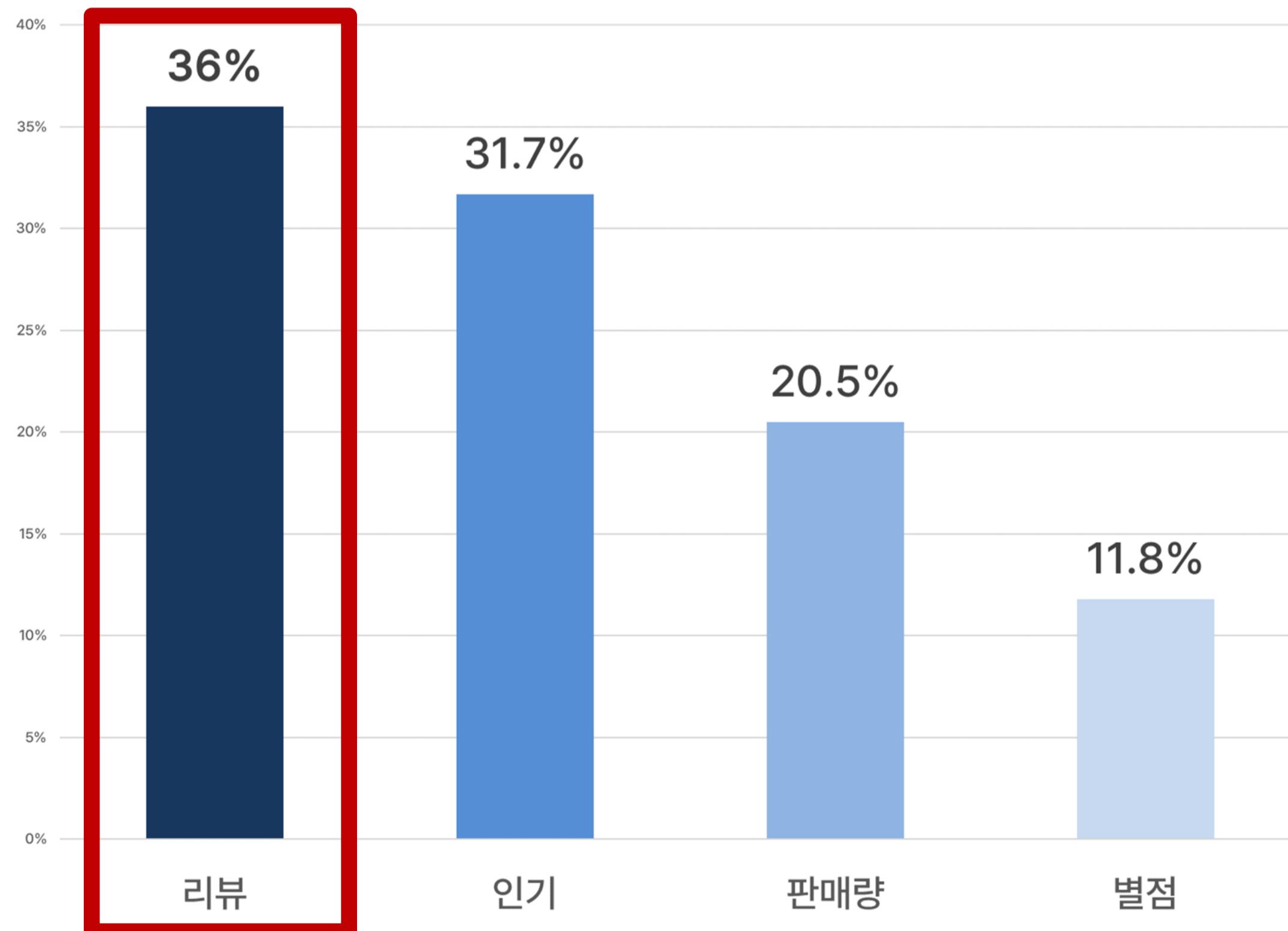
상품 추천

1. 상품 정렬 - 리뷰 감성 분석

**분위기를 고른 후,
상품들을 어떤 기준으로 추천하면 좋을까?**

인테리어 상품, 뭘 보고 고르지?

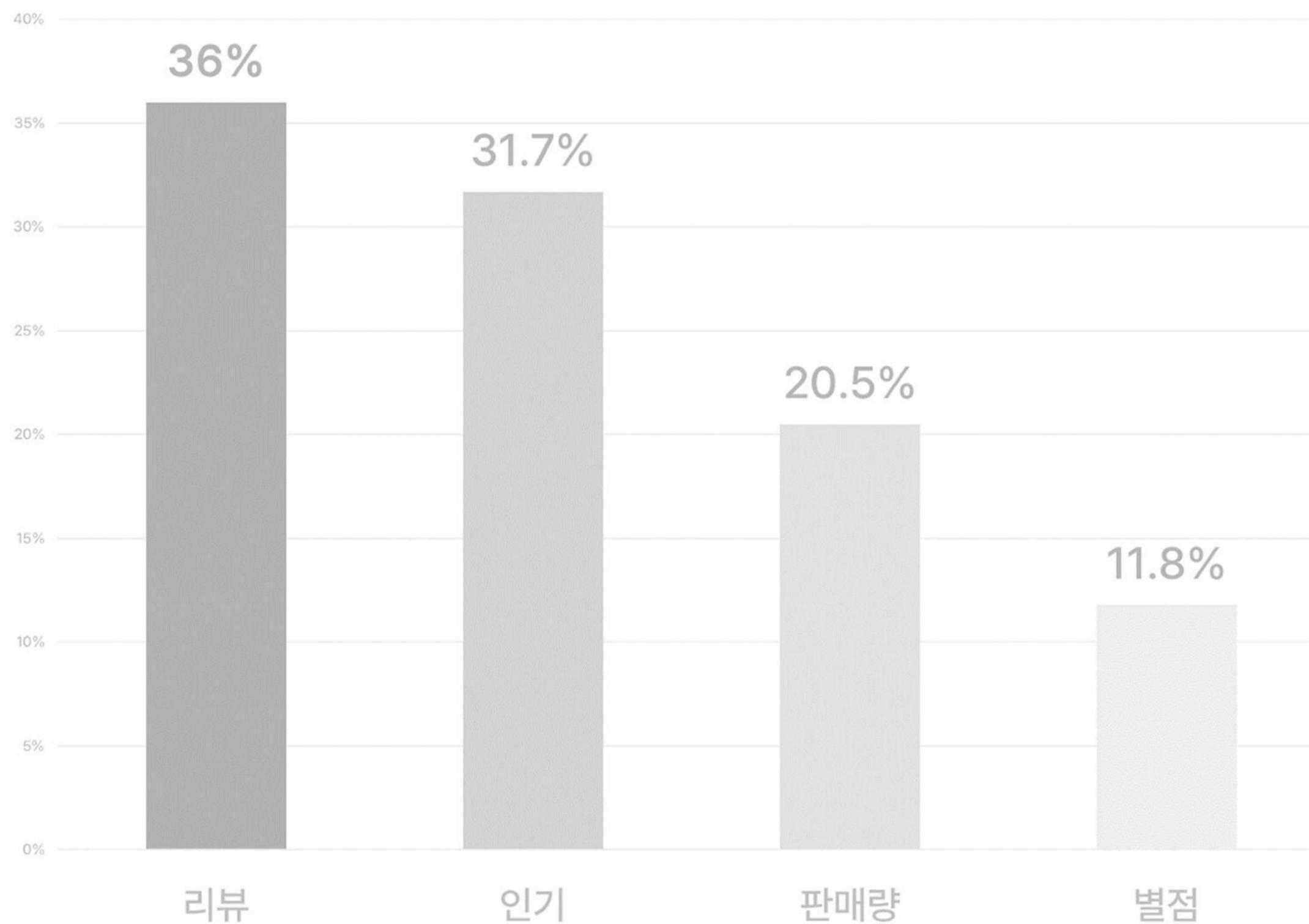
인테리어 상품 구매 시
가장 중요하게 생각하는 요소는?



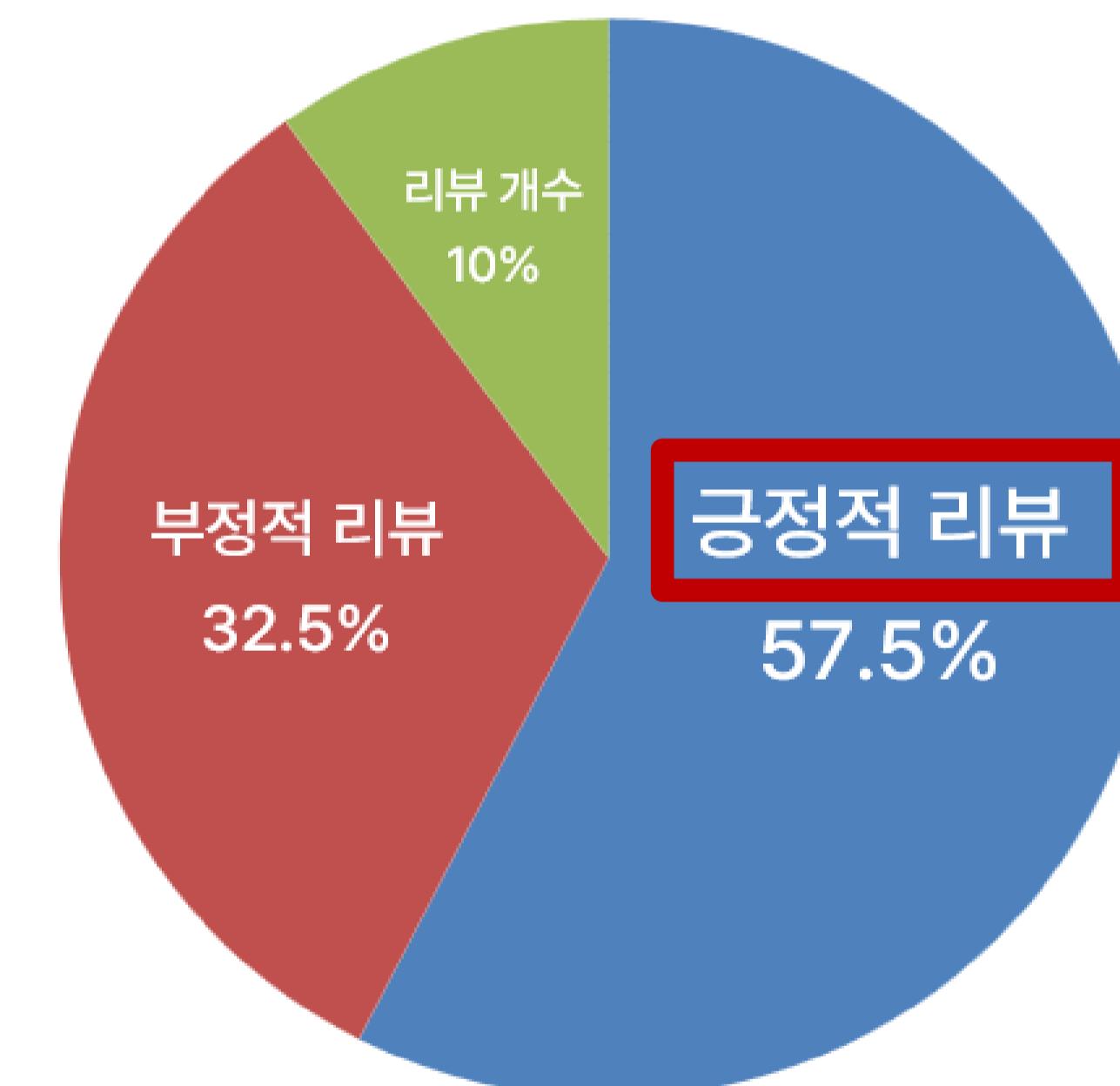
2022.10.31 ~ 2022.11.04
자체 설문조사 / N = 208

인테리어 상품, 뭘 보고 고르지?

인테리어 상품 구매 시
가장 중요하게 생각하는 요소는?



‘리뷰’를 선택했다면,
중요하게 고려하는 내용은?



2022.10.31 ~ 2022.11.04
자체 설문조사 / N = 208

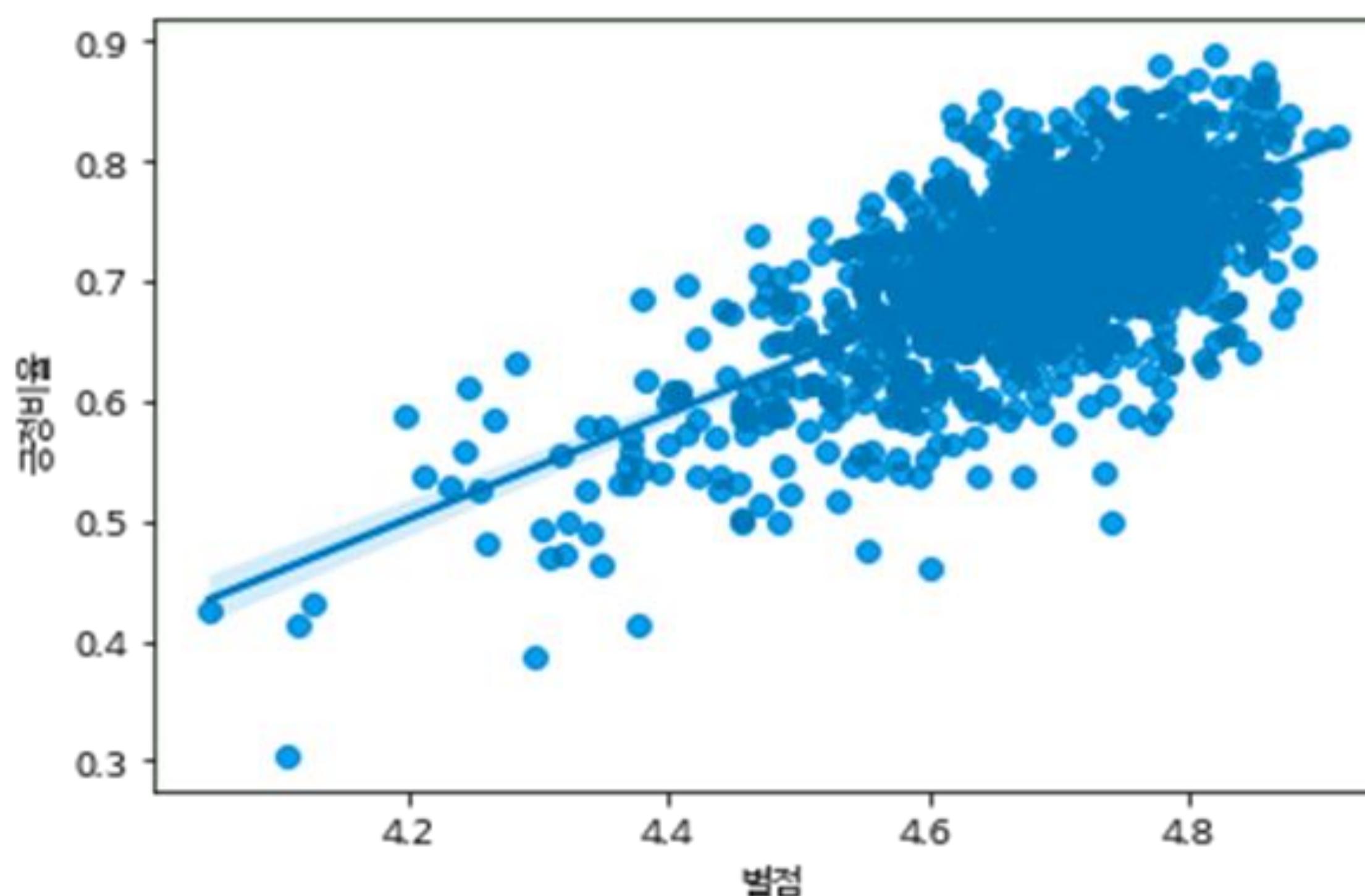
2022.10.31 ~ 2022.11.04
자체 설문조사 / N = 74

긍정 리뷰 대신에 별점을 봐도 되지 않을까?

별점과 리뷰의 긍정비율의 상관관계

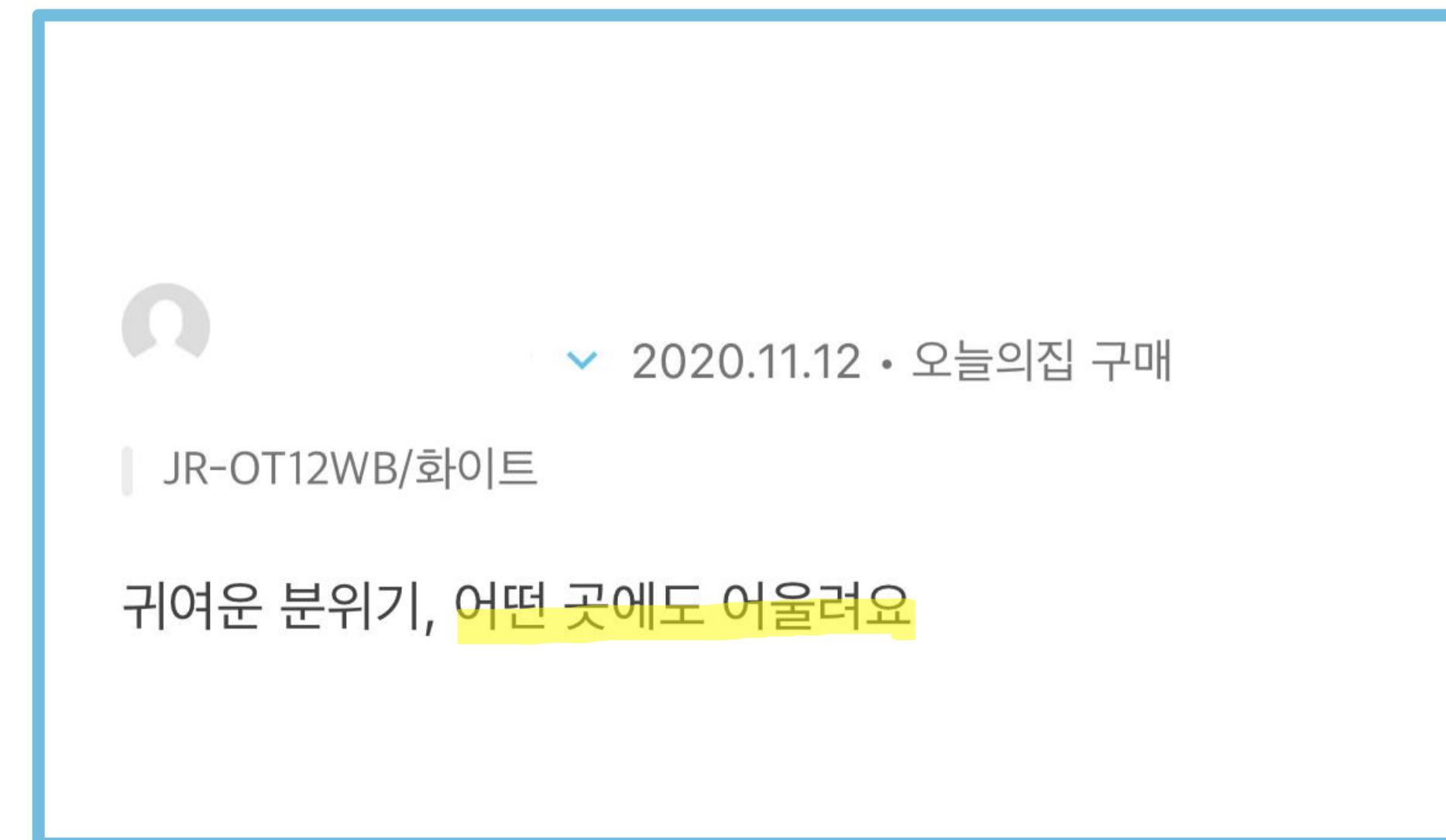
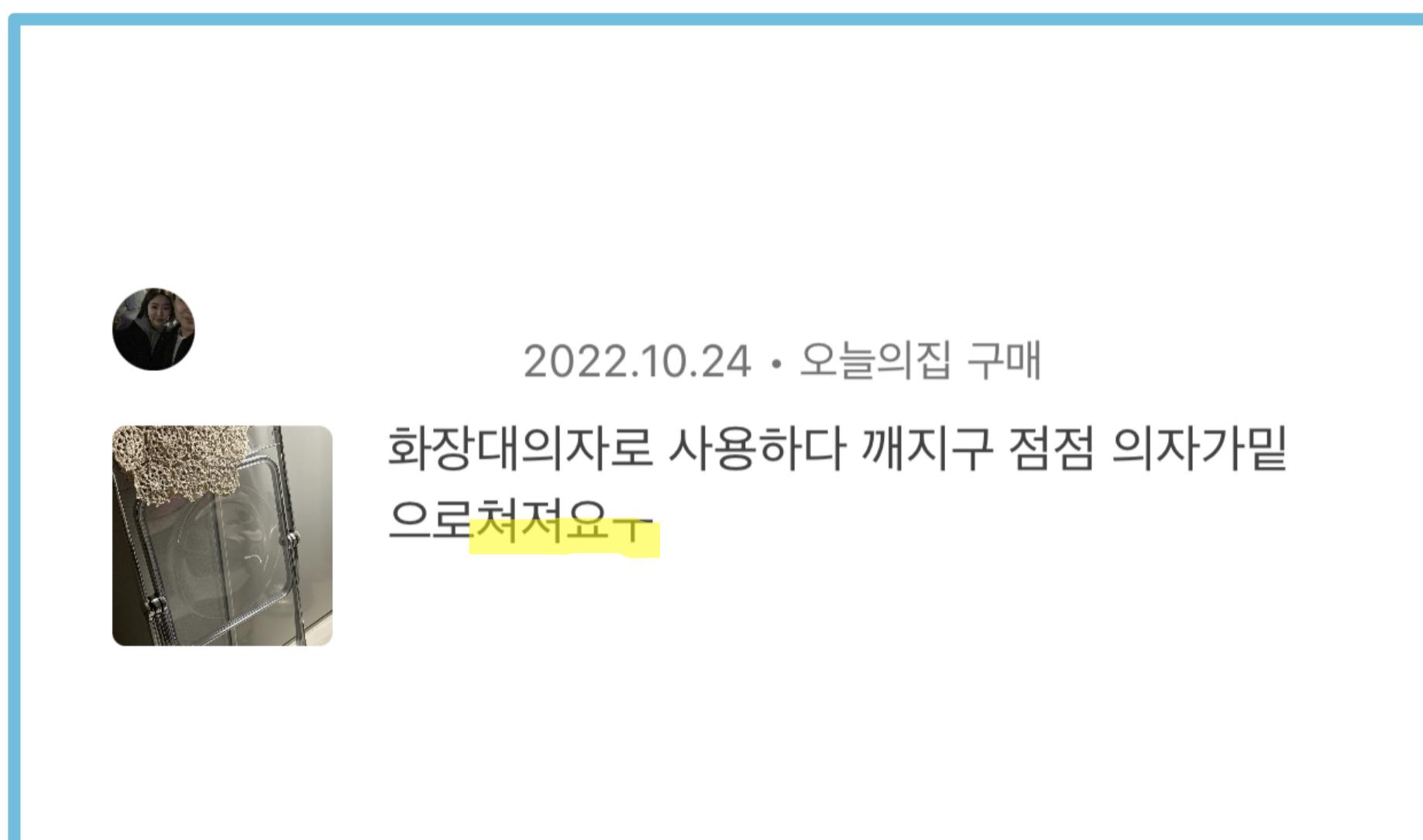
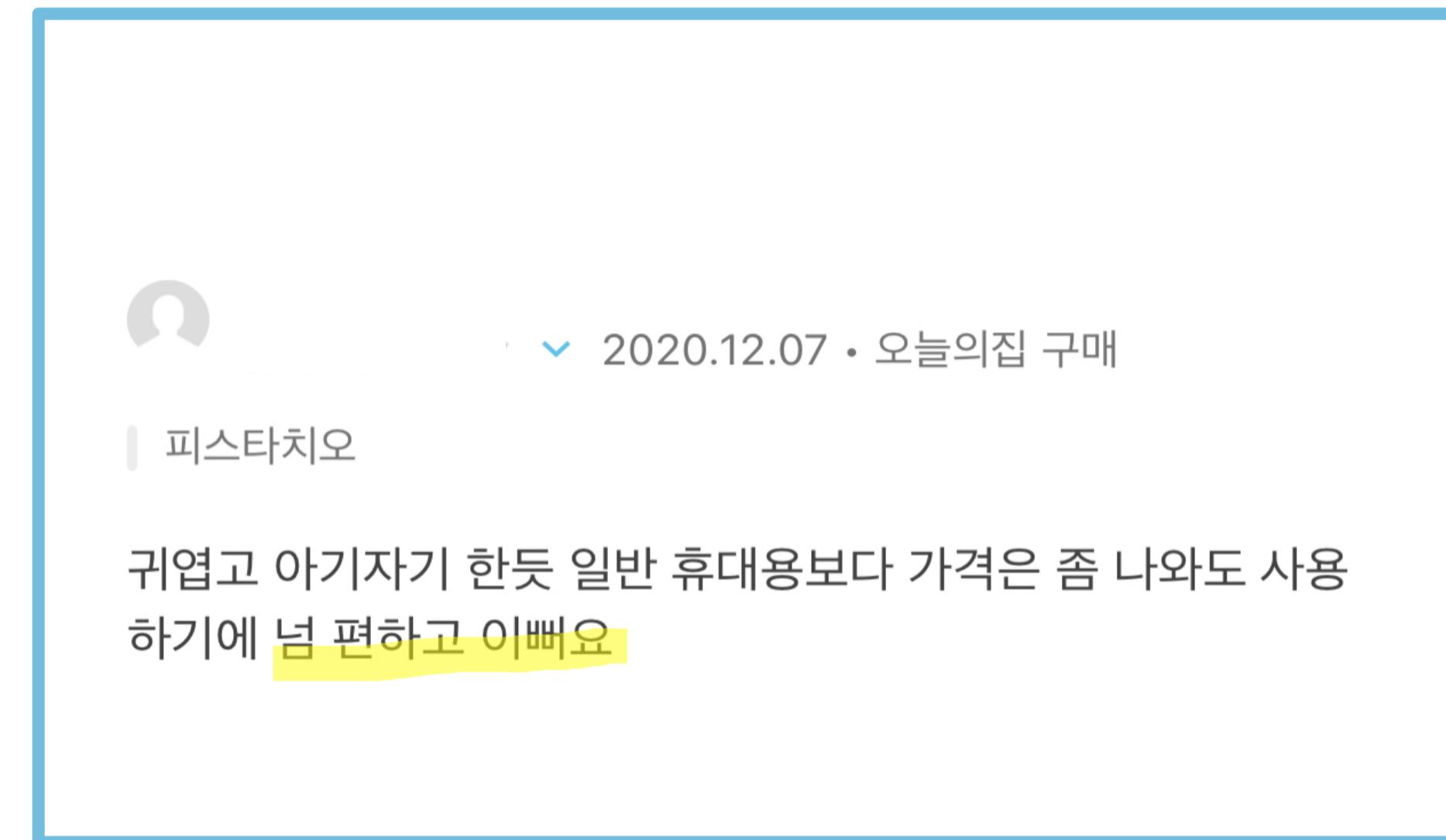
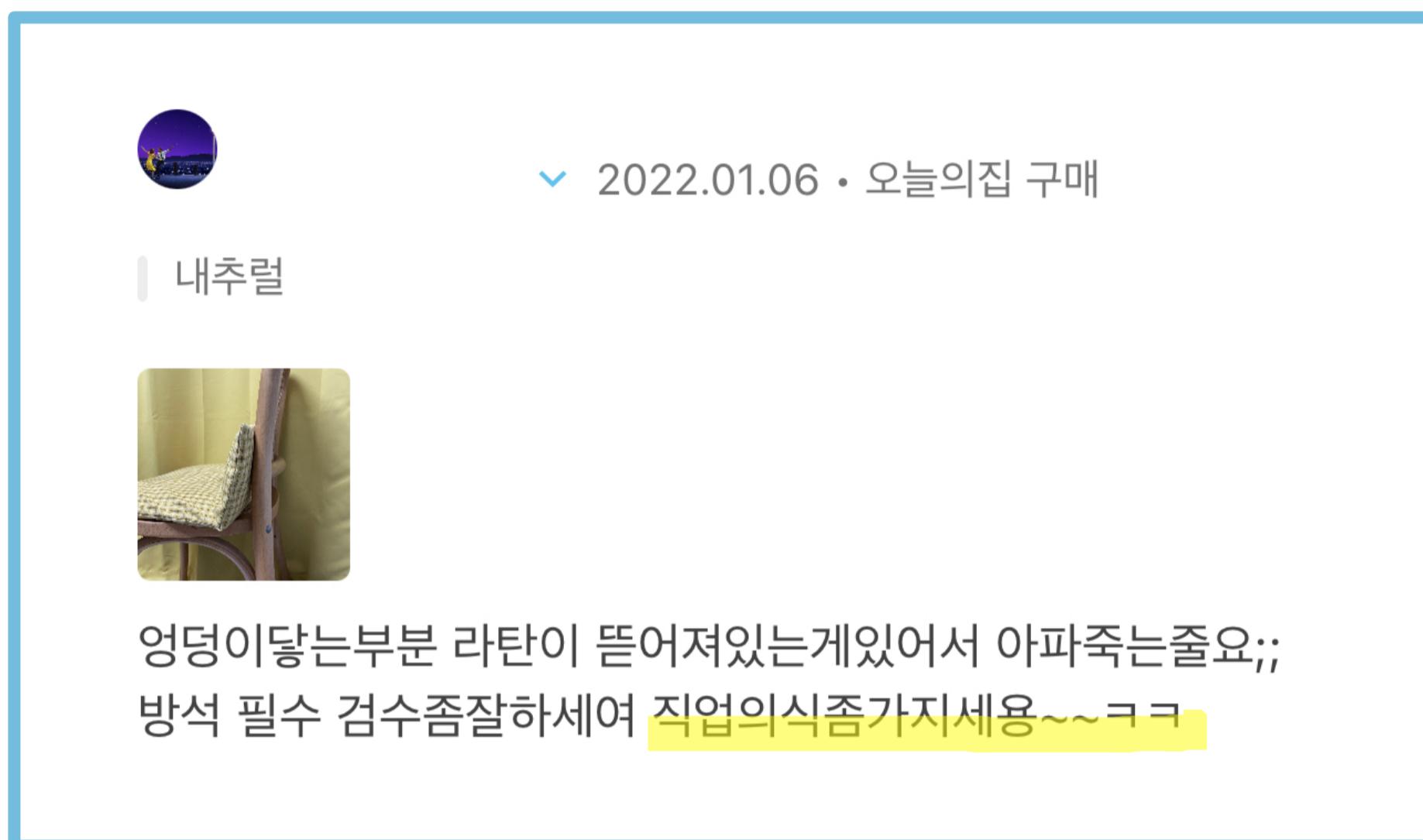
<별점과 긍정 비율의 상관분석 결과>

	n	r	CI95%	p-val	BF10	power
pearson	1142	0.675711	[0.64, 0.71]	3.700738e-153	1.905e+149	1.0



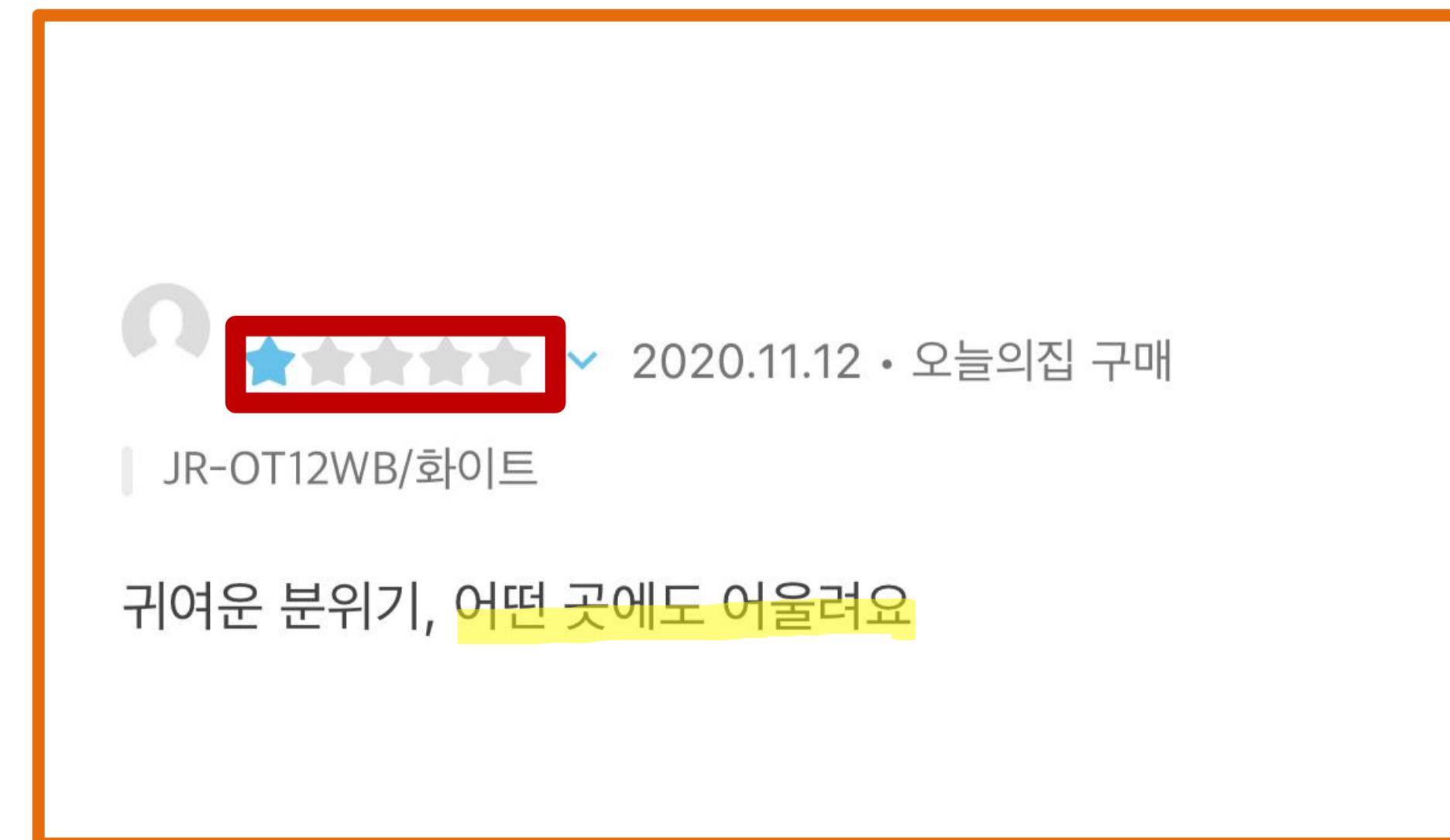
3.1. 상품 정렬 - 리뷰 감성 분석

긍정리뷰의 비율 대신에 별점을 참고해도 될까?



3.1. 상품 정렬 - 리뷰 감성 분석

긍정리뷰의 비율 대신에 별점을 참고해도 될까?



긍정리뷰의 비율 대신에 별점을 참고해도 될까?

별점 점수와 리뷰 내용이 다른 경우가 많다!!



2022.10.24 · 오늘의집 구매



화장대의자로 사용하다 깨지구 점점 의자가밑
으로쳐져요ㅜ



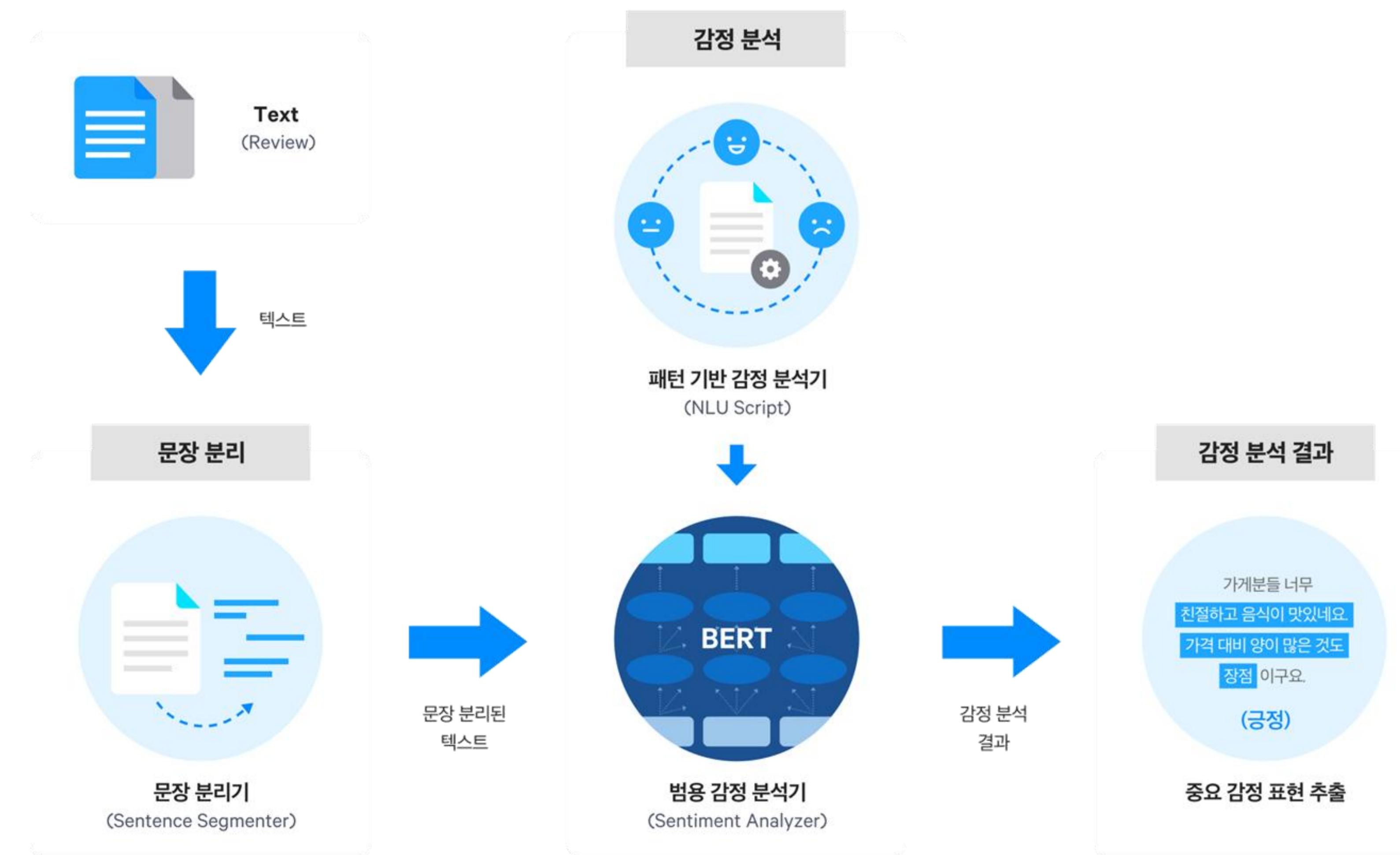
2020.11.12 · 오늘의집 구매

JR-OT12WB/화이트

귀여운 분위기, 어떤 곳에도 어울려요

긍정리뷰의 비율은 어떻게 알 수 있을까?

CLOVA Sentiment 를 활용하여
상품 리뷰를 긍정 / 부정 / 중립으로 감성 분류



긍정리뷰의 비율은 어떻게 알 수 있을까?

< 감성 분석 결과 >

	content	sentiment	Confidence .negative	Confidence .positive	Confidence .neutral
190	설치 전이라 일단 테이블 위에다 올려 봤어요ㅎㅎ	neutral	0.000080	0.000334	0.999586
191	디자인 깔끔하니 좋아요!	positive	0.000195	0.999726	0.000078
192	스마트 조명으로 바꾸고 싶은데 입구가 작아서 조명을 못바꾸고있어요 ㅜㅜ	negative	0.999686	0.000080	0.000235

< 상품별 긍정 리뷰 비율>

상품명	긍정리뷰수	전체리뷰수	긍정비율
(안전포장) TOKABO 토카보 단스탠드	119	174	0.683908
우드접시 플레이트 모음	131	197	0.664975
로망띠크 데이지 꽃 쿠션 등받이 겸 방석 소파쿠션 침대쿠션	107	126	0.849206
[단독] 플라워 체커보드 프리미엄 터프팅 유니크 러그 욕실 발매트 2color	144	182	0.791209
TEXTURED 코튼 러그 카페트 3size	153	244	0.627049

이제 수연이의 집들이 선물을 골라보자!

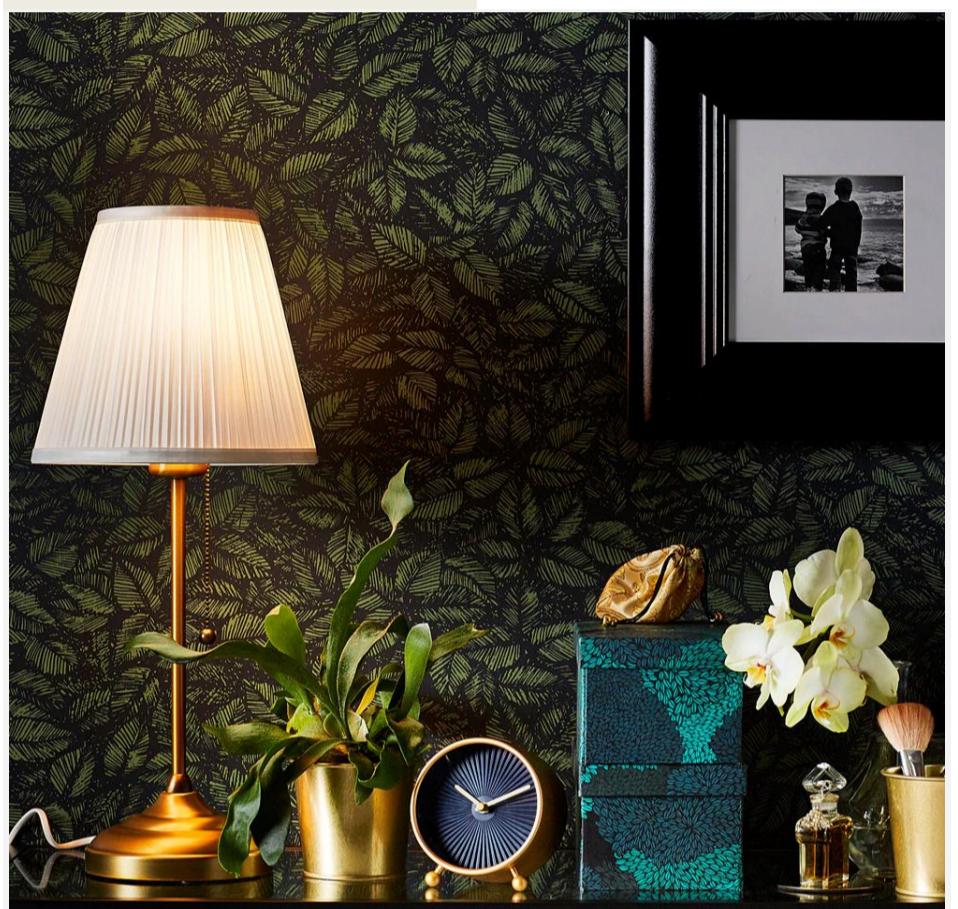
Mood



Positive Rate

특정 상품을 분위기별로 추천

클래식&앤틱



0.809398

조명

북유럽



0.800000

빈티지&레트로



0.810281

러블리&로맨틱

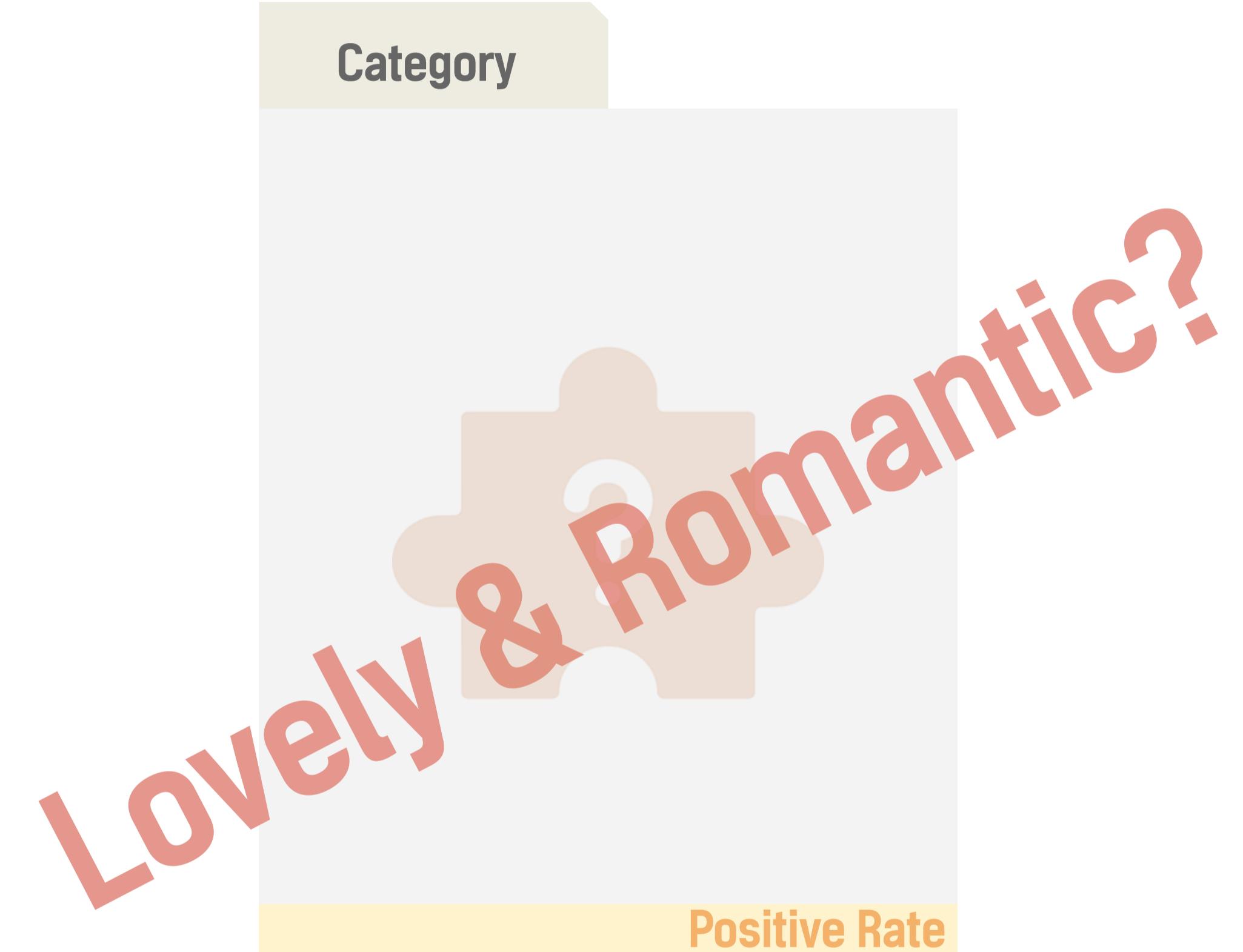


0.787946

한국&아시아



0.785032



특정 분위기의 상품을 카테고리별로 추천

오브제 캔들



0.873494

휴지통



0.800000

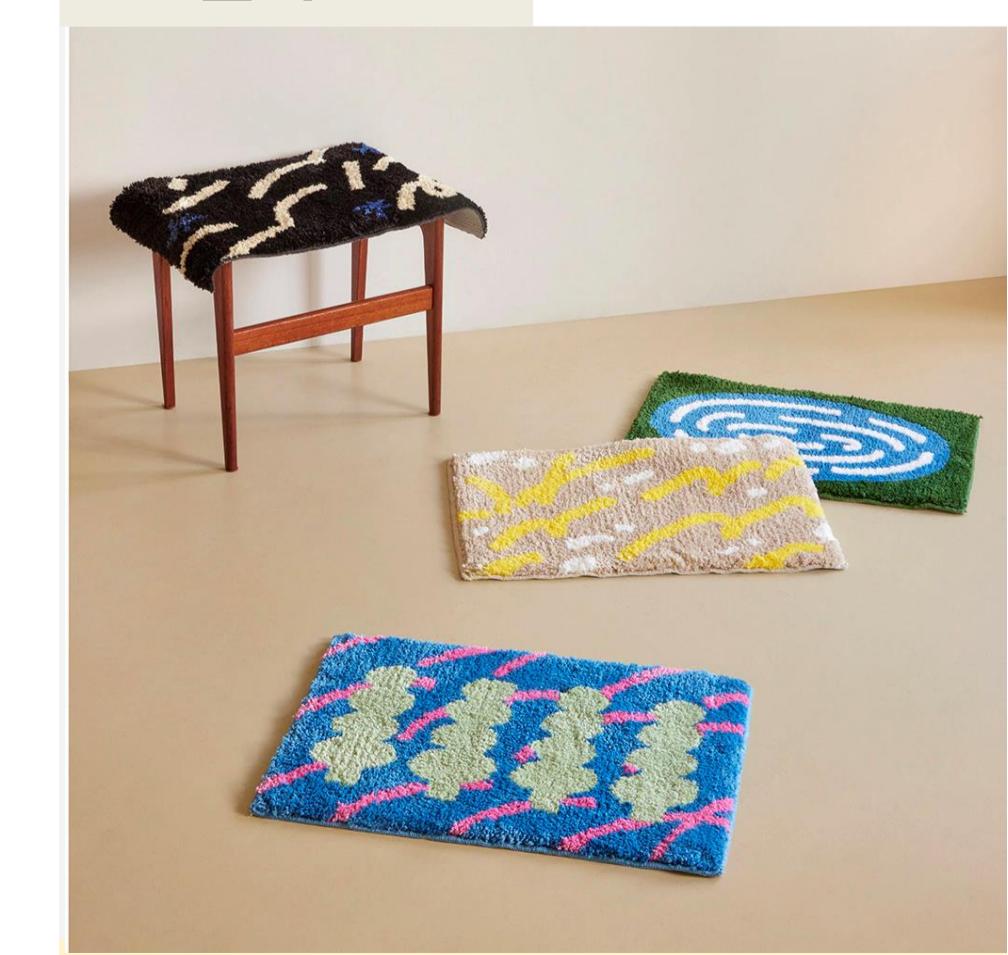
러블리&로맨틱

쿠션



0.849206

발매트



0.846154

03

상품 추천

2. 상품 조합 - 연관 분석

**내가 사려던 선물을 친구가 먼저 샀다면?
서로 어울리는 집들이 선물 조합을 추천 받아보자!**



**내가 사려던 선물을 친구가 먼저 샀다면?
서로 어울리는 집들이 선물 조합을 추천 받아보자!**



**‘헐...
나도 조명
사려고 했는데..’**

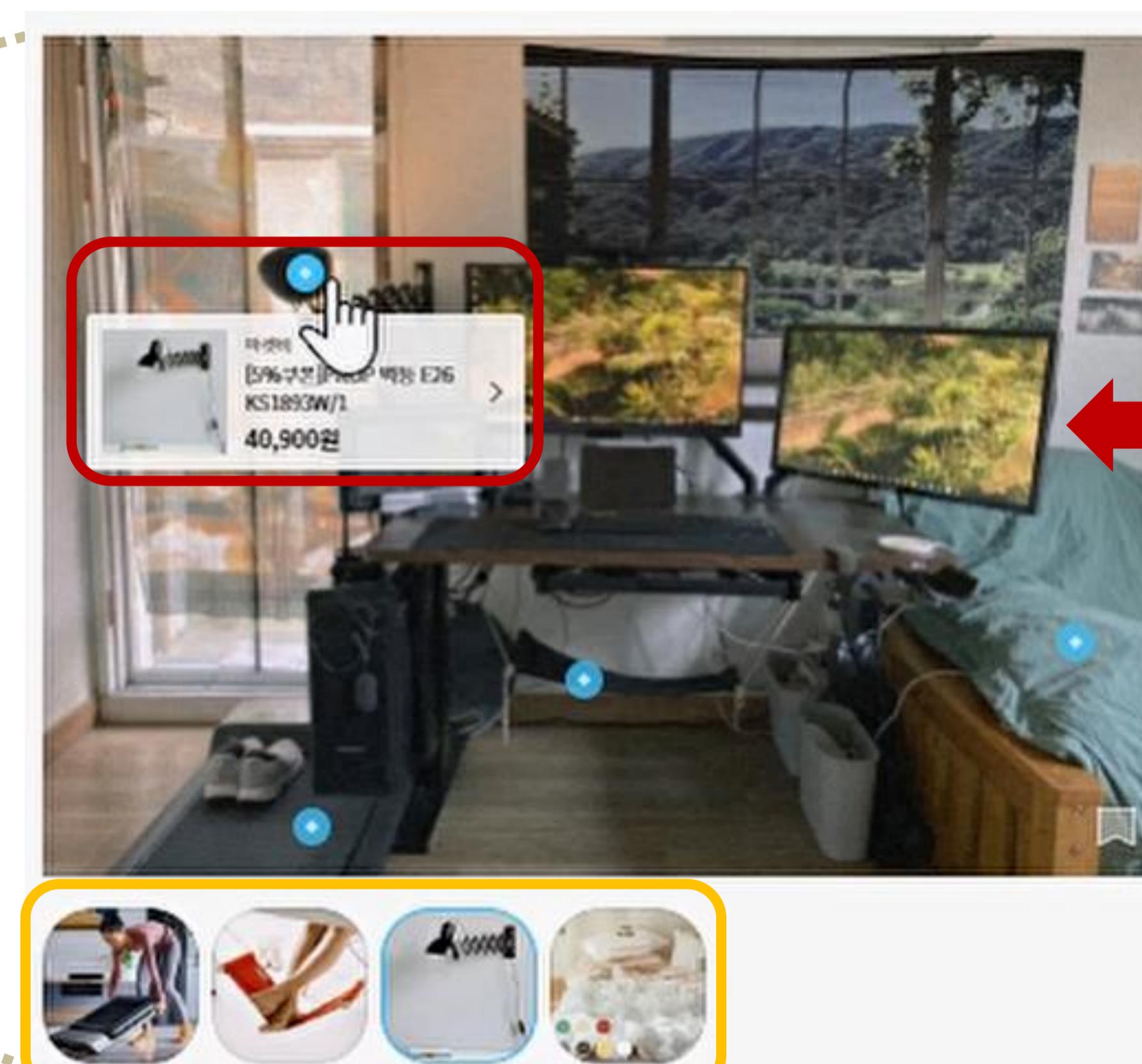
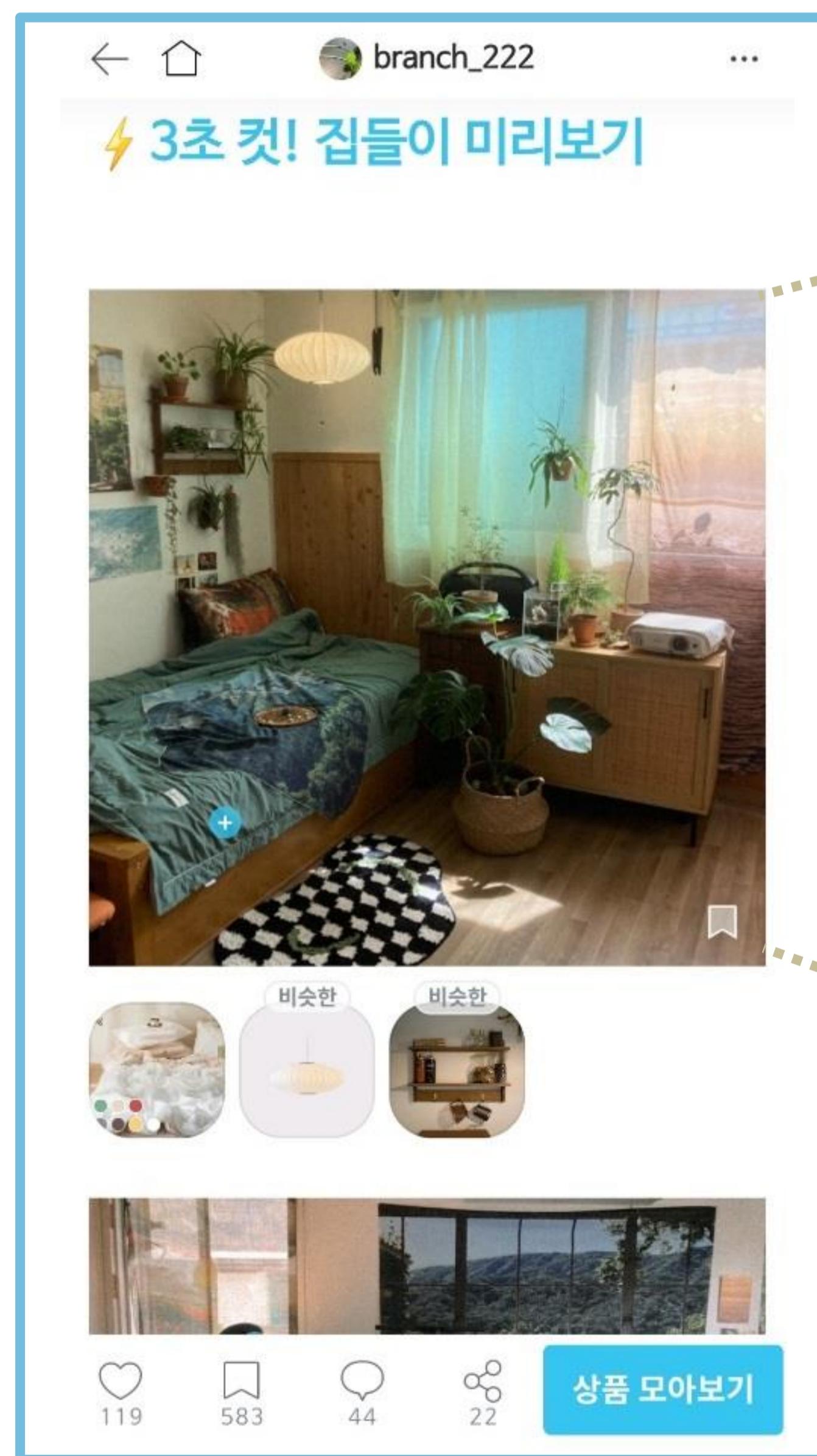
서로 어울리는 상품 조합은 어떻게 알 수 있을까?

연관 분석

- : 고객이 어떤 상품을 함께 구매할 것인지를 알 수 있는 **연관성** 분석
- : 룰기반의 모델로서 상품과 상품 사이에 어떤 연관이 있는지 찾아내는 알고리즘



연관분석을 위한 데이터 수집 방법



사진마다
Photo_id 부여

한 공간에 같이 배치된 상품

어울리는 상품 조합은 어떻게 알 수 있을까?

photo_id	데스크스탠드	단스탠드	턴테이블	소형식물	중대형식물	모빌/가랜드	쿠션
12438130	0.0	1.0	1.0	0.0	0.0	0.0	0.0
12438131	1.0	0.0	0.0	1.0	0.0	0.0	0.0
12438151	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12442747	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12450279	0.0	0.0	0.0	0.0	0.0	1.0	1.0

Photo_id별 (사진별)로 태깅되어 있는 상품의 경우 1, 아니면 0으로 더미변수를 생성
 전체 데이터 프레임을 Apriori 알고리즘을 활용하여 연관분석

어울리는 상품 조합은 어떻게 알 수 있을까?



연관분석(Apriori)을 통한 카테고리 조합 추출

단스탠드

	antecedents	consequents	Antecedents support	Consequents support	support	confidence	lift	leverage	conviction
4334	단스탠드	턴테이블	0.01033	0.0516	0.0015	0.14545	6.73391	0.00128	1.14494
4360	단스탠드	홀이불 ·겹이불	0.1532680	0.003944	0.001315	0.008578	2.174837	0.00071	1.004674
4347	단스탠드	미니USB 가습기	0.153268	0.004696	0.001503	0.009804	2.087843	0.000783	1.005159
4366	단스탠드	차렵이불	0.153268	0.054095	0.015965	0.104167	1.925637	0.007674	1.055894
4376	단스탠드	이불커버	0.153268	0.064801	0.018032	0.117647	1.815516	0.0081	1.059892

어울리는 상품 조합은 어떻게 알 수 있을까?



연관분석(Apriori)을 통한 카테고리 조합 추출

단스탠드

	antecedents	consequents	Antecedents support	Consequents support	support	confidence	lift	leverage	conviction
4334	단스탠드	턴테이블	0.01033	0.0516	0.0015	0.14545	6.73391	0.00128	1.14494
4360	단스탠드	홀이불 ·겹이불	0.1532680	0.003944	0.001315	0.008578	2.174837	0.00071	1.004674
4347	단스탠드	미니USB 가습기	0.153268	0.004696	0.001503	0.009804	2.087843	0.000783	1.005159
4366	단스탠드	차렵이불	0.153268	0.054095	0.015965	0.104167	1.925637	0.007674	1.055894
4376	단스탠드	이불커버	0.153268	0.064801	0.018032	0.117647	1.815516	0.0081	1.059892

집들이 선물로 어울리는 조합은?

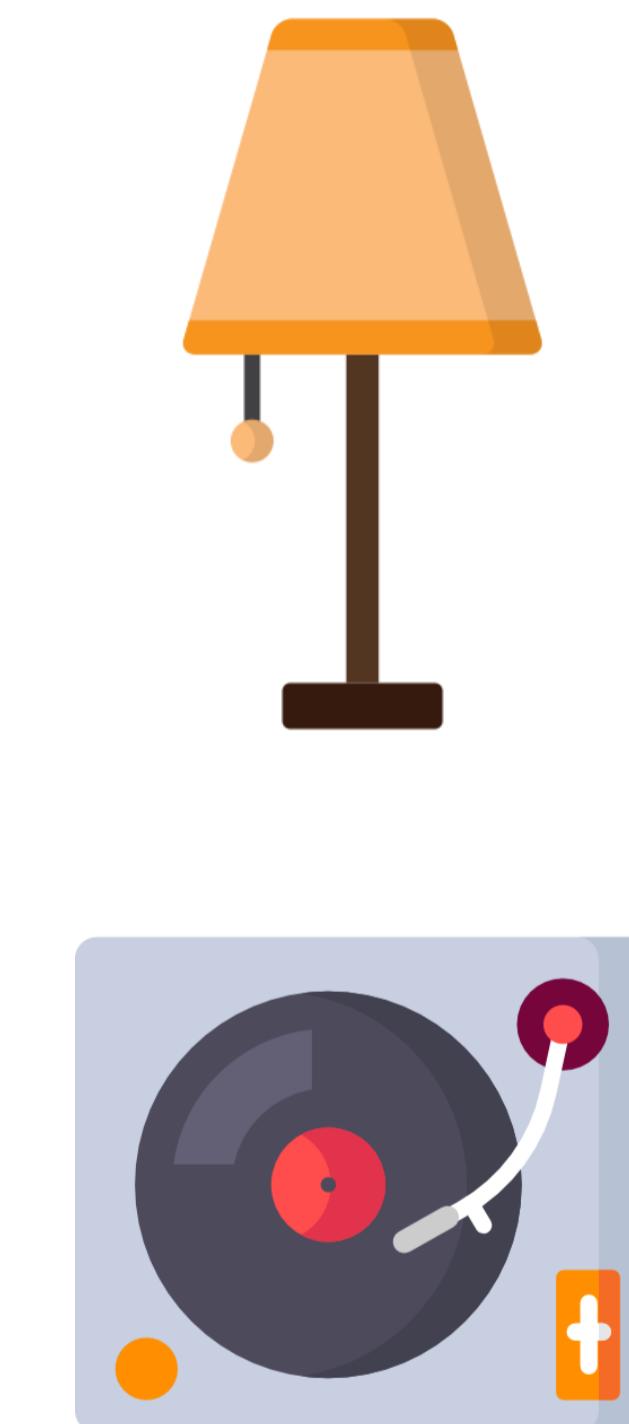
01



02



03



모빌/가랜드 + 쿠션

식물/화분 + 데스크스탠드

단스탠드 + 턴테이블

**어울리는 조합을 알았으니
분위기 · 긍정리뷰 비율 기준으로 상품을 골라보자!**

1) 모빌/가랜드 + 쿠션



폼폼이 가랜드 완제품 6종



4color 꽃방석쿠션 인테리어소품

러블리&로맨틱

프렌치&프로방스

뜨개 레이스 행잉 가랜드



내츄럴 레이스 쿠션



2) 식물/화분 + 데스크스탠드



사랑목 핑크+
고급 토분 도자기 5종



코알라화분 다용도
오브제화분 2size



블러썸 무선 LED 미니 스탠드
클립형_3colors

러블리
&로맨틱

빈티지&
레트로



HEIRI 단스탠드 E26
KS2175T



스파트필름 실내공기정화식물
소형식물



아티크 토분 받침세트 수제토분
인테리어 화분

3) 단스탠드 + 턴테이블



ARSTID 오르스티드 단스탠드

+



SUPERIOR LP 턴테이블

클래식& 앤틱

북유럽



모던보이스피어
단스탠드_3colors

+



AT-LP60XBT WW
완전자동벨트 무선 턴테이블

03

상품 추천

3. 내일의 집 vs 오늘의 집

3.3. 내일의 집 vs 오늘의 집

A / B 중 어떤 조합이 더 마음에 드시나요?

A



아빠가만든공간

비비앙 단스탠드_2colors



슬아하임

반원 PVC 패턴 코일 신발장 현관 베란다
도어 카페 매장 출입구 발매트 발판 2C

B



아빠가만든공간

비비앙 단스탠드_2colors



THE살림

[1+1][오늘의딜] 22년형 국내산 무독
성 완벽 물빠짐 욕실화

3.3. 내일의 집 vs 오늘의 집

A / B 중 어떤 조합이 더 마음에 드시나요?

A



아빠가만든공간

비비앙 단스탠드_2colors

슬아하임

반원 PVC 패턴 코일 신발장 현관 베란다
도어 카페 매장 출입구 발매트 발판 2C

B



아빠가만든공간

비비앙 단스탠드_2colors



THE살림

[1+1][오늘의집] 22년형 국내산 무독
성 완벽 물빠짐 욕실화

내일의 집

오늘의 집

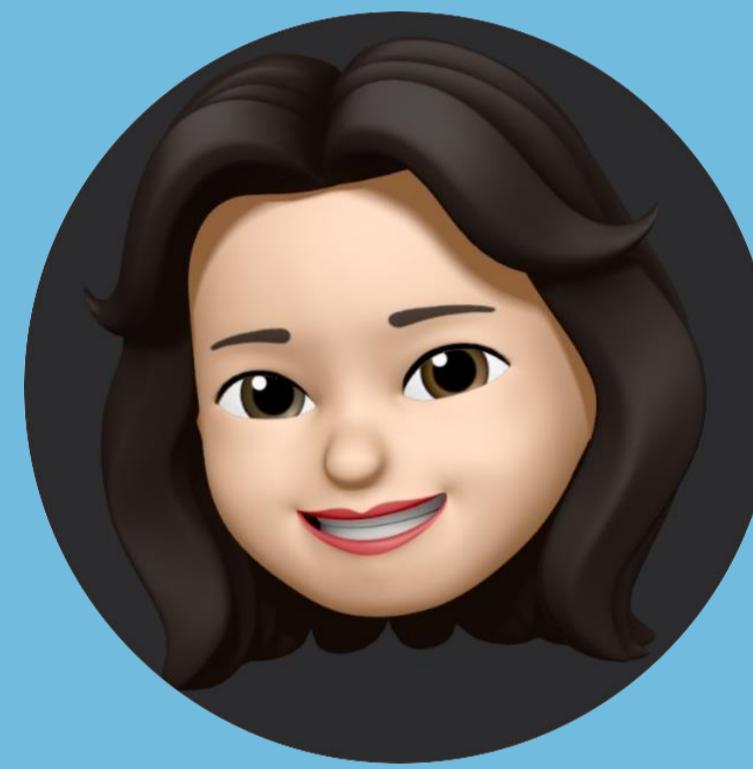
감사합니다



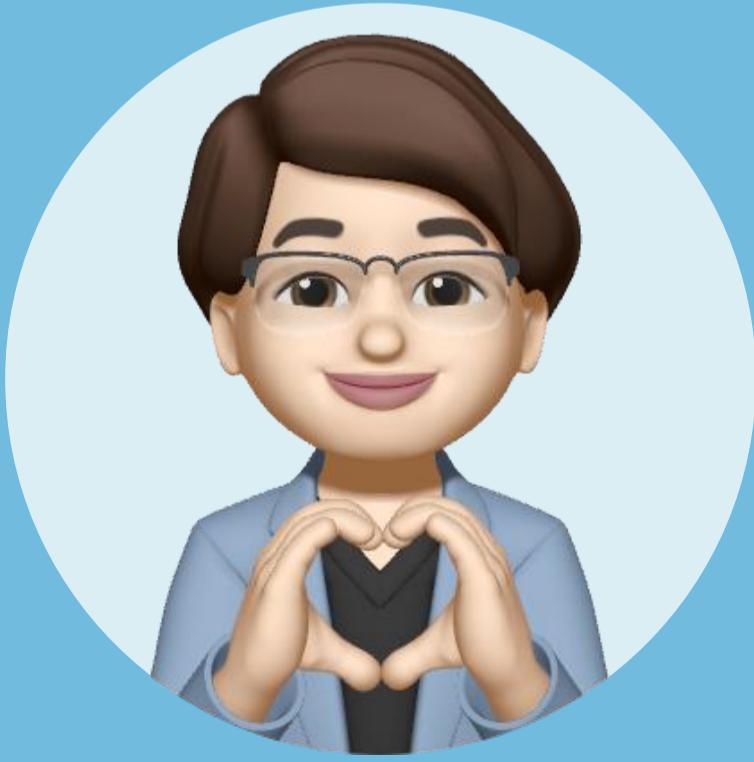
이소현



송나현



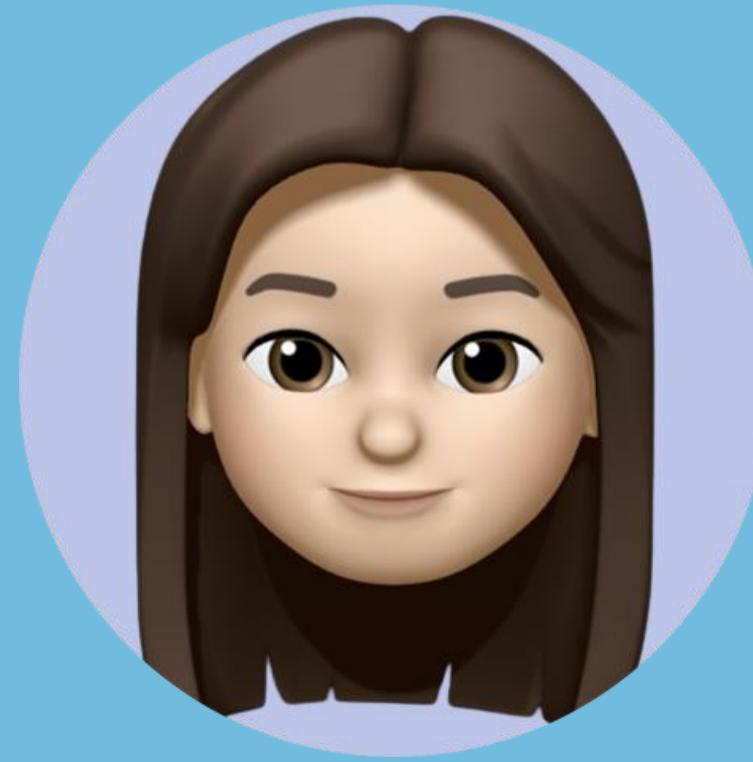
최수연



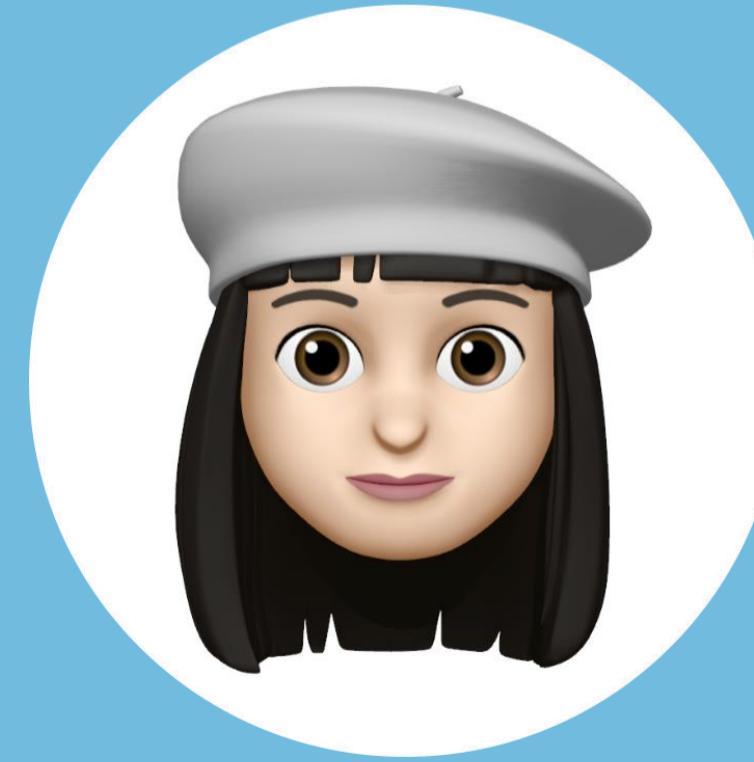
김예원



김민지



박유정



이지혜

부록

01 데이터 수집

데이터 수집

01

이용자 정보

오늘의 집 '집들이' 이용자가 직접 올린 게시물에서 본인 집의 인테리어, 가족 형태, 주거 형태 그리고 구매 상품 등 정보를 수집.

- 수집기간 : 2022년 10월
- 수집범위 : 2020년 1월 1일 ~ 2022년 10월(작성일자 기준)
- 수집량 : 약 4천 6백개(게시글 기준)

'id', '제목', '가족형태', '공간', '평수구간', '스타일', '세부공사', '분야', '작업', '예산', '조회', '스크랩', '좋아요', '댓글', 'url'

02

상품 데이터

이용자가 집들이 게시물에 소개한 상품들과 카테고리별 대표 상품을 스크래핑하여 수집.

- 수집량 : 약 22만개

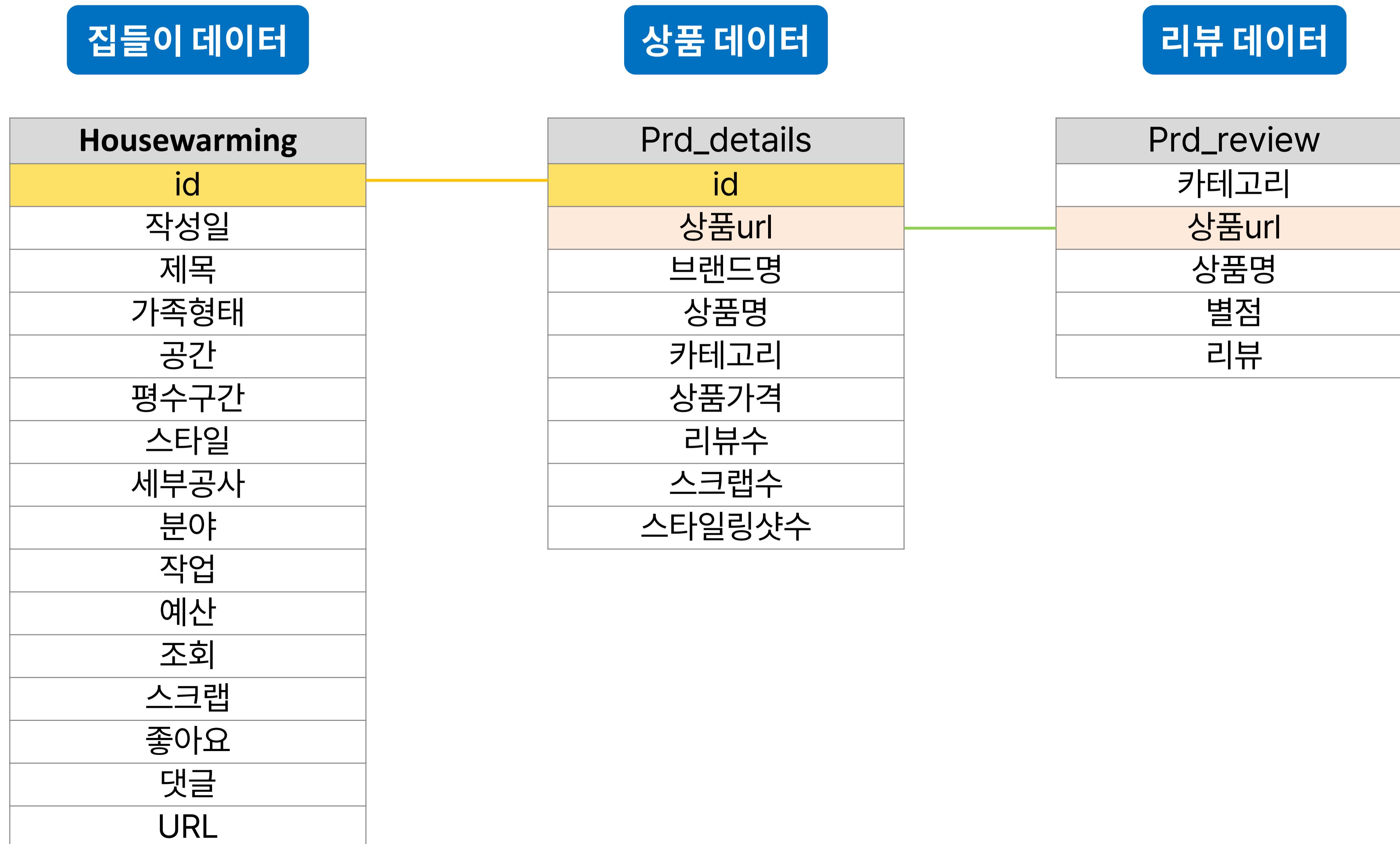
'id', '상품url', '브랜드명', '상품명', '상품가격', '카테고리', '리뷰수', '스크랩수', '스타일링샷수', '작성일'

03

리뷰 데이터

위 상품 데이터의 리뷰 스크래핑

데이터 ERD



02 데이터 전처리

데이터 전처리 : 데이터 타입 및 형태 통일 / 파생컬럼 생성

제목	작성일자	제목	작성일자	작성일
최소한의 시공으로 만든, 편안한 24평 구축 리모델링	1주 전	최소한의 시공으로 만든, 편안한 24평 구축 리모델링	1주 전	2022-10-01
자연을 갤러리 삼아 계절을 오롯이 누리는 앤틱 하우스	1주 전	자연을 갤러리 삼아 계절을 오롯이 누리는 앤틱 하우스	1주 전	2022-10-01
세 가족이 사는 아늑 하우스, 소탈하지만 사랑스럽게	2주 전	세 가족이 사는 아늑 하우스, 소탈하지만 사랑스럽게	2주 전	2022-09-24
빈티지 오브제로 더 특별하게- 나만의 전시장 같은 방	2주 전	빈티지 오브제로 더 특별하게- 나만의 전시장 같은 방	2주 전	2022-09-24
꿈이 가득한 여행지처럼! 공간마다 테마가 있는 신혼집	2주 전	꿈이 가득한 여행지처럼! 공간마다 테마가 있는 신혼집	2주 전	2022-09-24
베이지톤 아트월을 살려서- 화이트 신혼집 인테리어	2주 전	베이지톤 아트월을 살려서- 화이트 신혼집 인테리어	2주 전	2022-09-24
숨은 평수 발견↖ 벽을 내어주고 얻은 개방감😊	2주 전	숨은 평수 발견↖ 벽을 내어주고 얻은 개방감😊	2주 전	2022-09-24
9년 간 살던 체리몰딩 구축, 화이트 모던홈으로 재탄생	2주 전	9년 간 살던 체리몰딩 구축, 화이트 모던홈으로 재탄생	2주 전	2022-09-24
물건으로 무성한 숲을 이루다, 디자이너의 8평 원룸	2주 전	물건으로 무성한 숲을 이루다, 디자이너의 8평 원룸	2주 전	2022-09-24
유니크한 편집숍st 집에서, 다채로운 다묘&신혼 라이프	2주 전	유니크한 편집숍st 집에서, 다채로운 다묘&신혼 라이프	2주 전	2022-09-24
깨끗한 화이트에 따뜻한 색감을 더해 만든 러블리 신혼집	2주 전	깨끗한 화이트에 따뜻한 색감을 더해 만든 러블리 신혼집	2주 전	2022-09-24
꼼꼼 시공으로 환해졌어요! 산뜻한 올화이트 신혼집	2주 전	꼼꼼 시공으로 환해졌어요! 산뜻한 올화이트 신혼집	2주 전	2022-09-24

- 집들이 게시글 '작성일자' 컬럼 스크래핑 일자 기준 N일 전, N주 전 표기 → 'YYYY-MM-DD' 형태의 파생컬럼 생성

데이터 전처리 : 스크래핑 시점 차이로 인한 데이터 업데이트 내용 일괄 변경

상품url	브랜드명	상품명	카테고리	상품가격	리뷰수	스크랩수	스타일링 샷수
https://ohou.se/productions/665800/selling	마틸라	[최대3만원쿠폰] 도즈 60수 고밀도순면 차렵이불세트 SS/Q/K 10색	['패브릭', '이불·이불솜', '이불', '차렵이불']	69900.0	4824.0	66658.0	3825.0
https://ohou.se/productions/665800/selling	마틸라	도즈 60수 고밀도순면 차렵이불세트 SS/Q/K 10색	['패브릭', '이불·이불솜', '이불', '차렵이불']	69900.0	4827.0	66674.0	3827.0
https://ohou.se/productions/665800/selling	마틸라	[5%쿠폰] 도즈 60수 고밀도순면 차렵이불세트 SS/Q/K 10colors	['패브릭', '이불·이불솜', '이불', '차렵이불']	69900.0	4835.0	66667.0	3831.0

- 할인 적용여부에 따라 상품명이 달라지는 부분 전처리 ▶ 쿠폰 키워드가 들어가지 않은 상품 기준으로 일괄 변경
- 스크래핑 시점의 차이로 리뷰수와 스크랩수 차이가 나는 부분 전처리 ▶ 최신 데이터 기준으로 (최댓값) 일괄 변경

상품url	브랜드명	상품명	카테고리	상품가격	리뷰수	스크랩수	스타일링 샷수
https://ohou.se/productions/665800/selling	마틸라	도즈 60수 고밀도순면 차렵이불세트 SS/Q/K 10색	['패브릭', '이불·이불솜', '이불', '차렵이불']	69900.0	4852.0	66777.0	3838.0
https://ohou.se/productions/665800/selling	마틸라	도즈 60수 고밀도순면 차렵이불세트 SS/Q/K 10색	['패브릭', '이불·이불솜', '이불', '차렵이불']	69900.0	4852.0	66777.0	3838.0
https://ohou.se/productions/665800/selling	마틸라	도즈 60수 고밀도순면 차렵이불세트 SS/Q/K 10색	['패브릭', '이불·이불솜', '이불', '차렵이불']	69900.0	4852.0	66777.0	3838.0

데이터 전처리 : 미입점 데이터(Null값) 삭제

상품url	브랜드명	상품명	카테고리	상품가격	리뷰수	스크랩수	스타일링샷수
https://ohou.se/productions/880143/selling	히트조명	[15%쿠폰] LED 직하형 엣지 평판조명 방등 50W (640x640)	['조명', 'LED평판등', 'LED방등']	NaN	NaN	NaN	NaN

히트조명
[15%쿠폰] LED 직하형 엣지 평판조명 방등 50W (640x640)
★★★★★ 리뷰 쓰기 첫 리뷰 두 배 적립

미입점

- 상품가격, 리뷰수, 스크랩수, 스타일링샷수 결측치인 경우 : '**미입점 상품**'
- 집들이에 태깅되어 있다면 분위기를 분류할 수 있겠지만, 그 외에는 활용할 수 있는 정보가 없어 **미입점 상품 삭제**

데이터 전처리 : 집들이 게시글 제목 텍스트 분석 - 형태소 분석기 Okt 사용

집들이 글제목

"아늑하고 편안한 무드로- 빈티지 한 스푼 더한 내추럴 홈"

- Okt 형태소 분석기를 사용하여 'Noun'과 'Adjective' 추출

- 단어의 길이가 1 이면 제거

- 빈번하게 등장하지만 큰 의미가 없는 불용어 제거

stopwords = {'인테리어', '공간', '하우스', '주택', ...}



['아늑하고', '편안한', '무드', '빈티지', '스푼', '더한', '내추럴']