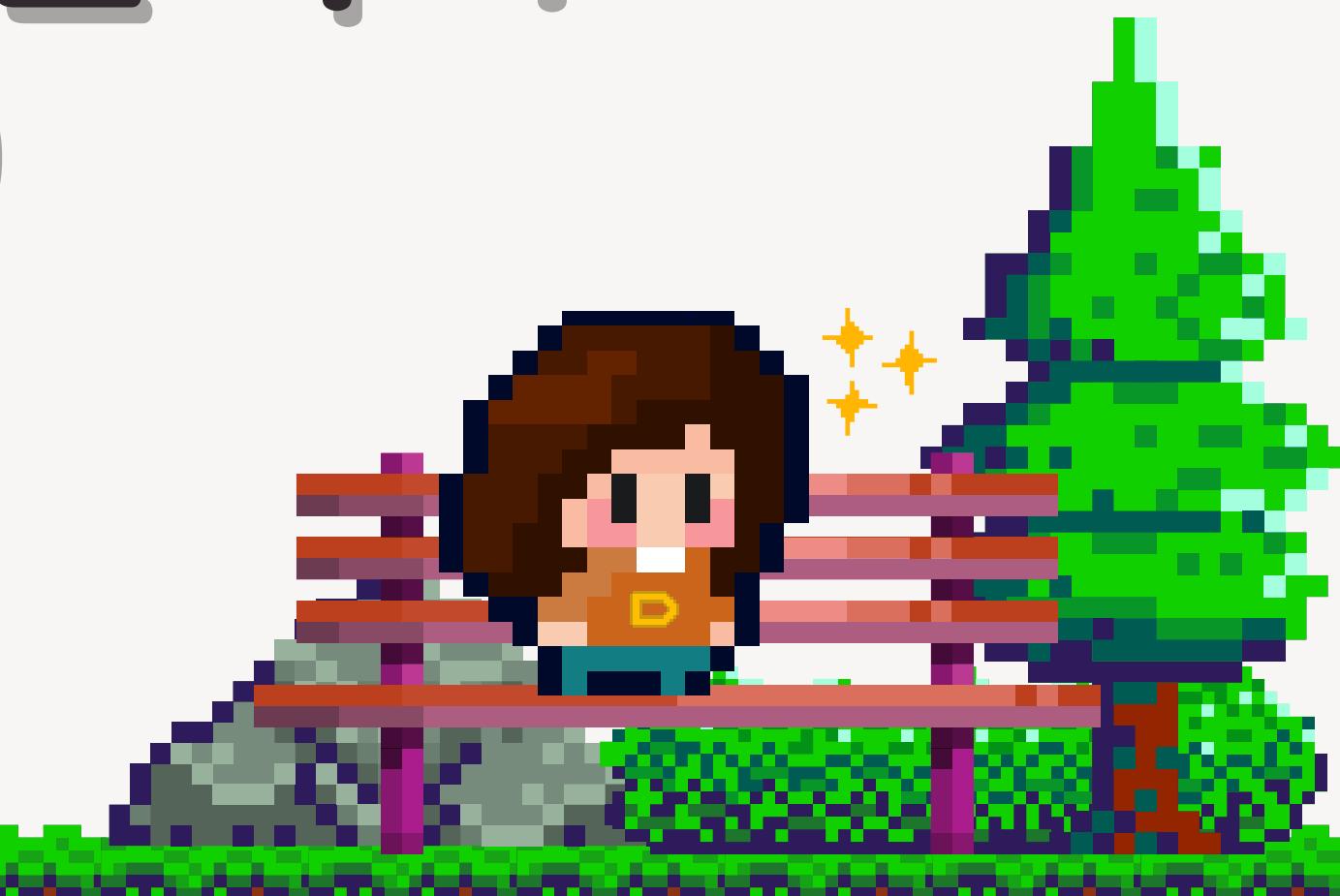


신입 데이터 분석가 DD의 이커머스 데이터 분석기

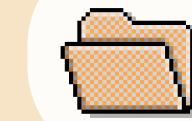
(feat. banul)

바늘이야기

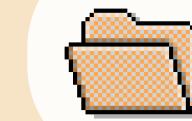




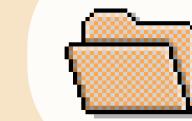
Contents



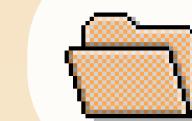
Chapter 1. 주제 선정



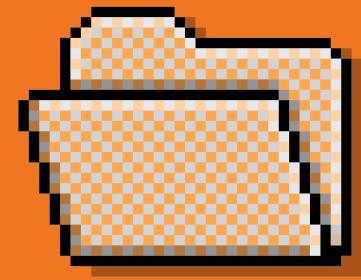
Chapter 2. DD의 스토리



Chapter 3. DD가 아닌 Team DD가 생각해보았습니다



Chapter 4. Epilogue



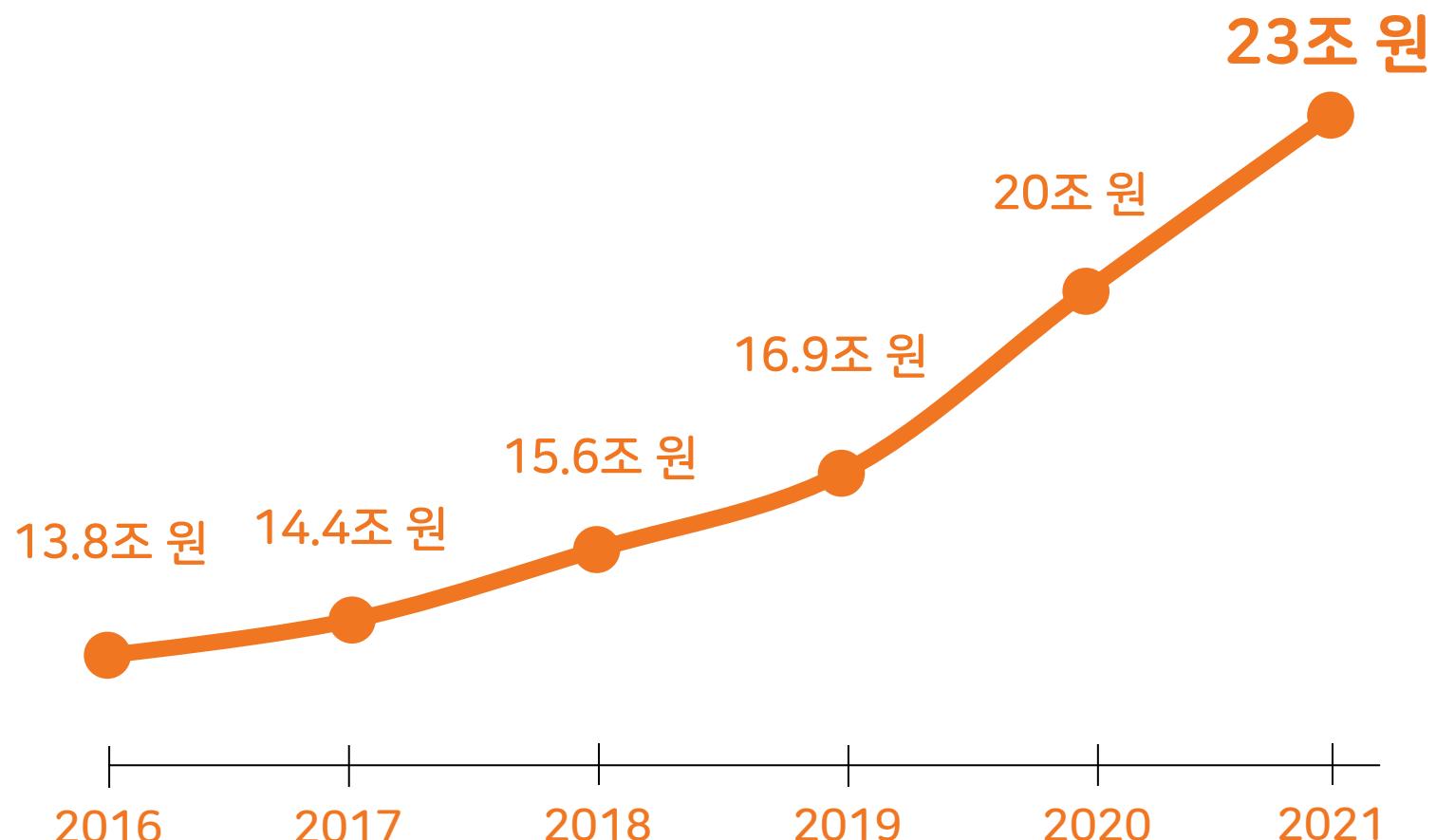
Chapter 1. 주제 선정

Chapter 1. 주제 선정



데이터산업 시장규모 추이

출처: 과학기술정보통신부 "2021 데이터산업 현황 조사"



"국내 기업 74%, 자체 데이터 활용 수준 초급 이하"

필요한 기술, 인재, 프로세스에 투자하는 국내 기업은 10%에 불과

실제 클라우드를 이미 사용하는 기업 44%가 그렇지 않은 기업에 비해
데이터 및 분석으로 생산성이 개선됐다

"엎친데 덮친 격...전문인력도 없다"

국내 데이터 전문인력 증가율은 이 같은 성장세를 따라가지 못하고 있다.

지난해를 기준으로 한 데이터 산업 종사인력은 총 33만4672명.

이 가운데 데이터 직무 인력은 8만9058명에 불과

1년 전보다 7~8% 늘어난 규모지만 여전히 한 자릿수



그렇다면 데이터 분석을 하고자 하는 기업에게

자체 데이터를 활용할 수 있는 방법을 소개하자!

13.8조 원



2016

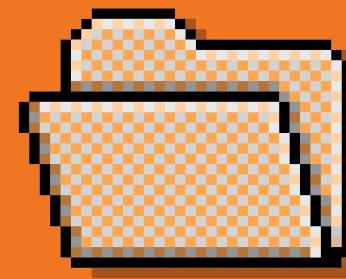
"급 이하"

불과

비해

하고 있다.

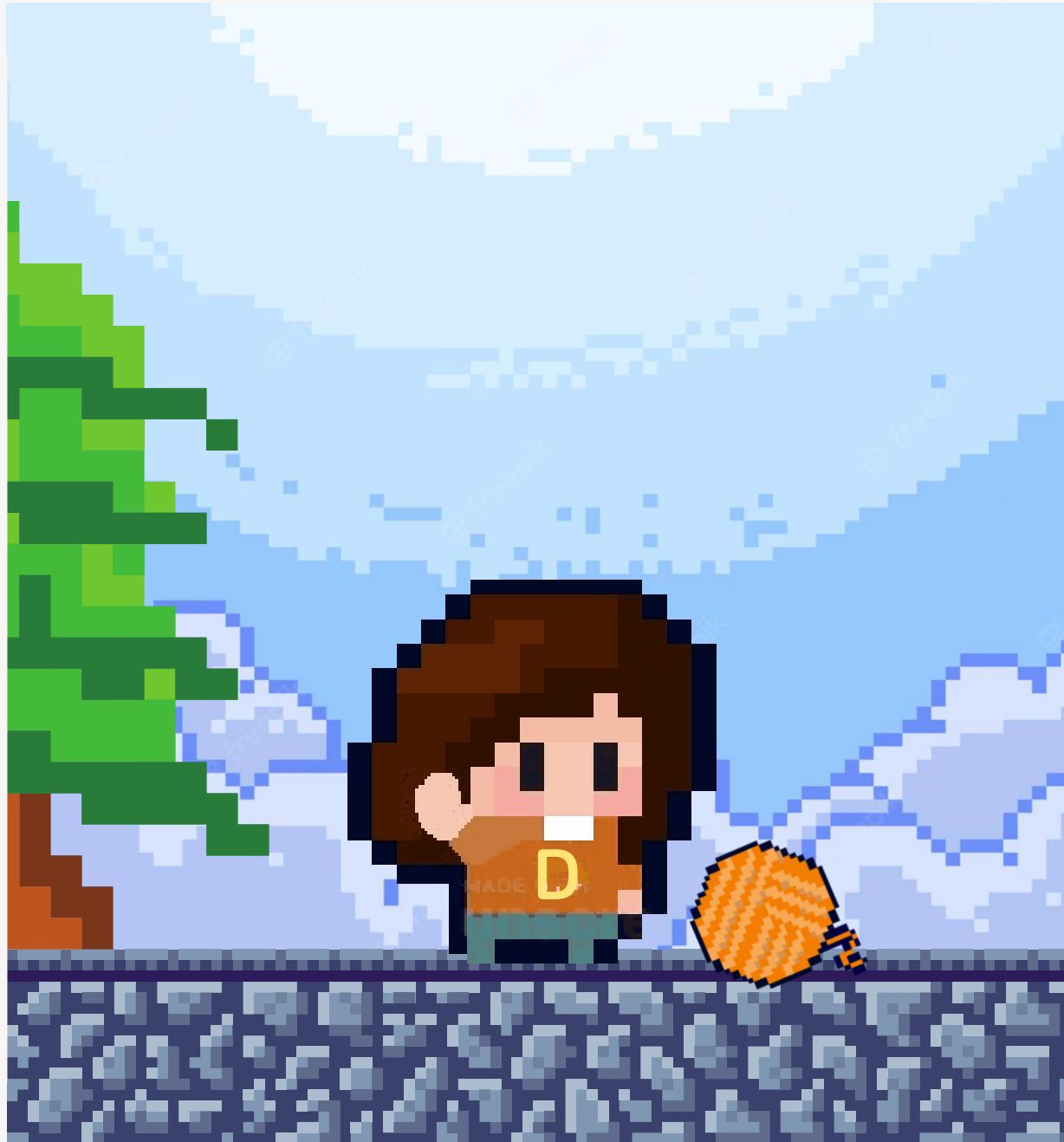
72명.



Chapter 2. DD의 스토리

데잇걸즈 과정을 수료하고 데이터 분석가로 취업하고 싶은 DD!

Chapter 2. DD의 스토리_DD



DD

이름 : DD

직업 : 이커머스 데이터 분석가

취미 : 뜨개질, 일상생활 궁금점을 데이터로 접근하기

좋아하는 것 : 달콤한 디저트, 유튜브로 음악 디깅하기

MBTI : ISTJ

Chapter 2. DD의 스토리_이력서 작성



프로젝트 제안서	
기간	2022-10-13 ~ 2022-11-10
팀원	김남이, 김지성, 김하나, 김혜원, 이현정, 인지원, 최지현
팀 구성	PM 김남이 TL 이현정 Marketer 김지성
목적	데이터 드리븐 기업을 지향하는 기업에 파일럿 데이터 분석 솔루션을 제공
개요	<p>[1] 사업 개요: 데잇걸즈6 DD x 바늘이야기 데이터 분석 프로젝트 [2] 협업 제안 이유:</p> <ul style="list-style-type: none"> 실제 이용해본 고객의 입장에서 개선되면 좋을 서비스에 대해 데이터로 접근해 아이디어를 제안하고자 함 국내 최고의 손뜨개 전문점으로, 장기적 관점 데이터 활용의 중요성 인식 자체 상품에 대한 수요 분석 제공 아카데미, 상품, 구매 고객 등 다양한 데이터 분석 진행 가능성 다양한 유형 분석 가능(온라인, 오프라인, 이커머스, 인스타, 유튜브 등) <p>[1] 제공 가능한 내용(아래 내용 중 선택):</p> <ul style="list-style-type: none"> 현찰 대시보드(매출, 고객, 상품) 그로스 해킹 지표 개선을 위한 의사 결정 서포트(AARRR 지표, OMTM 설정, A/B Test 제안 등) GA 보고서와 GTM 태깅 분석한 데이터를 통해 마케팅 활용 방안 제안 카테고리 별 교차 추천(함께 구매한 상품·부자재, 도서, 다른 실, 바늘 등) <p>기능 제안</p> <ul style="list-style-type: none"> 아카데미 분석: 온·오프라인 수강생 인원 수요 예측(ML), 추후 콘텐츠 강화 도입시 콘텐츠 주제를 설정하는데 데이터 수집 및 분석 레이아웃 제공 유튜브 채널(바늘이야기 김대리)을 통해 유입된 고객 데이터 분석 제공(GA) 및 분석 결과를 활용한 콘텐츠 제작 제안 O2O: 온라인 상품 중 반응이 좋은 패키지, 상품, 수요 등 데이터 분석을 통해 오프라인 공간 구성에 참고할 수 있는 근거 제안 <p>[2] 사용 가능한 툴</p> <ul style="list-style-type: none"> 분석 툴: SQL(MySQL, PostgreSQL, BigQuery 등), Python(numpy, pandas) 시각화 툴: Tableau, Power BI, Python(matplotlib), Google Data Studio 협업 툴: Git, Slack, Notion 등 <p>[2] 사용 가능 툴</p> <ul style="list-style-type: none"> 분석 툴: Tableau, Power BI, Python(matplotlib), Google Data Studio 협업 툴: Git, Slack, Notion 등
상세 사업 내용 제안	<p>- [1] 제공 가능한 내용(아래 내용 중 선택):</p> <ul style="list-style-type: none"> 현찰 대시보드(매출, 고객, 상품) 그로스 해킹 지표 개선을 위한 의사 결정 서포트(AARRR 지표, OMTM 설정, A/B Test 제안 등) GA 보고서와 GTM 태깅 분석한 데이터를 통해 마케팅 활용 방안 제안 카테고리 별 교차 추천(함께 구매한 상품·부자재, 도서, 다른 실, 바늘 등) <p>기능 제안</p> <ul style="list-style-type: none"> 아카데미 분석: 온·오프라인 수강생 인원 수요 예측(ML), 추후 콘텐츠 강화 도입시 콘텐츠 주제를 설정하는데 데이터 수집 및 분석 레이아웃 제공 유튜브 채널(바늘이야기 김대리)을 통해 유입된 고객 데이터 분석 제공(GA) 및 분석 결과를 활용한 콘텐츠 제작 제안 O2O: 온라인 상품 중 반응이 좋은 패키지, 상품, 수요 등 데이터 분석을 통해 오프라인 공간 구성에 참고할 수 있는 근거 제안 <p>[2] 사용 가능한 툴</p> <ul style="list-style-type: none"> 분석 툴: SQL(MySQL, PostgreSQL, BigQuery 등), Python(numpy, pandas) 시각화 툴: Tableau, Power BI, Python(matplotlib), Google Data Studio 협업 툴: Git, Slack, Notion 등 <p>- 시각화 툴: Tableau, Power BI, Python(matplotlib), Google Data Studio</p> <p>- 협업 툴: Git, Slack, Notion 등</p>

취업 성공!!



보안 서약서

소속: 한국정보사회진흥원(NIA) x 과학기술정보통신부 데이터 분석가 양성 과정 데잇걸즈 6기 TeamDD

계약 기간: 2022. 10. 13 ~ 2022. 11. 10

서약자: 김남이, 김지성, 김하나, 김혜원, 이현정, 인지원, 최지현 총 7명

상기 소속된 TeamDD의 팀원 7인은 (※ 바늘이야기 (이하 '회사'라 한다.)에 대해 아래의 사항을 준수할 것을 서약합니다.

- 회사의 영업비밀과 개발비밀, 생산비밀, 관리비밀 등 산업체산권 제반규정에 준하는 사항들에 대하여 허락 없이 공개 또는 누설하지 않을 것을 서약합니다.
- 본인이 알거나 소유하고 있는 회사의 지적, 물적 재산을 회사의 승낙 없이 부정하게 사용하거나 반출하지 않을 것임을 서약합니다.
- 계약 기간은 물론 계약 기간 이후에도 회사의 영업 비밀을 이용하여 창업하거나 경쟁 회사에 편직 또는 농업을 하는 것과 같이 물질적 이익을 취하거나 제 3자를 위하여 사용하지 않을 것임을 서약합니다.
- 계약 기간 종료 후에는 회사로부터 제공받은 모든 계정 및 raw data를 파기할 것이며, 추가적인 개인 작업을 하지 않을 것임을 서약합니다.
- 만일 본인이 상기 4개 조항을 위반할 시에는 관계 법규에 의거 엄중 처벌될 것을 인지하고 서약합니다.

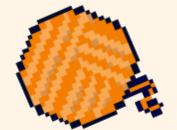
2022년 10월 13일

서약인 김남이 13
김지성 13
김하나 13
김혜원 13
이현정 13
인지원 13
최지현 13
(※ 바늘이야기 귀중)



Chapter 2. DD의 스토리_기업 소개

바늘이야기



바늘이야기는 1998년에 문을 연, 대한민국에서 가장 오래된 국내 최고의 뜨개 전문 기업. 뜨개질 실과 용품을 판매하는 전문 쇼핑몰과 함께 뜨개질 아카데미, 오프라인 매장 등을 운영하고 있다.

유튜브 채널 '바늘이야기 김대리'로 잘 알려진 우리의 **김대리님**은 바늘이야기의 마케팅 담당자다. DD의 사수로 등장할 예정이다.



바늘이야기 김대리 DAERI KIM ✓
구독자 27.1만명



ERP 데이터는 호스팅 업체에
요청하니 시간이 걸리네...



그럼 그동안 할 수 있는
리뷰 데이터를 수집해보자!!

Chapter 2. DD의 스토리_Naver Smart Store Review 스크래핑



수집 기간: 2016.02.18 ~ 2022.10.21

Raw Data Set

Column	Detail
product_name	상품명
product_id	상품코드
score	평점
date	날짜
color	색상
review	리뷰
reshop	재구매
onemonth	한달사용기
category	카테고리

product	score	date	color	review	reshop	onemonth	category
기자 샤벳	5	2022-09-24	2476 캔디	실이 정말 예뻐요~	0	0	실
기자 샤벳	5	2022-09-23	2479 페퍼민트	만드는 재미가 있고 선물했을 때 보람이 있어요	0	0	실
기자 샤벳	5	2022-09-02	2476 캔디	색이너무 예뻐요 맘에듭니다	0	0	실
기자 샤벳	5	2022-09-02	2475 블루라군	색이 너무 예뻐요	0	0	실
기자 샤벳	5	2022-08-28	2474 레몬&라임	색상ㅇ이 이뻐요 굿굿	0	1	실
기자 샤벳	5	2022-07-30	2479 페퍼민트	자스민 티코스터 떠보려고 주문했어요	0	0	실
기자 샤벳	5	2022-07-20	2476 캔디	색이 너무 예뻐요 얘기 담요 만들려고 샀는데 완전 만족!!	0	0	실
기자 샤벳	5	2022-07-05	2473 피지팝	다른색으로 구매했다가 너무 이뻐서 했습니다~	0	0	실
기자 샤벳	5	2022-05-30	2476 캔디	색상 너무 이뻐요ㅠㅠ	0	0	실
기자 샤벳	5	2022-05-21	2479 페퍼민트	동영상 따라서 뜨기 쉬웠어요. 색감도 참 곱네요~	0	0	실
기자 샤벳	5	2022-03-22	2476 캔디	실 색이 너무 이쁩니다. 뜯는 내내 기분도 덩달아 좋아졌어요 ㅎㅎㅎ 다 쓰고나면 또 주문할 거 같아요 ㅎㅎㅎ	0	0	실
기자 샤벳	5	2022-03-17	2479 페퍼민트	색 조합이 너무 이뻐요~ 배송도 빠르구요 맘에 쑥 들어요 감사합니다!	1	0	실
기자 샤벳	5	2022-03-17	2476 캔디	색 조합이 너무 이뻐요~ 배송도 빠르구요 맘에 쑥 들어요 감사합니다!	1	0	실
기자 샤벳	5	2022-02-07	70 스트로베리에이	색 너무 예쁘고 작품 뜯면 더 예뻐요	1	0	실
기자 샤벳	5	2022-02-07	2429 매지컬	작품 뜯면 훨씬 더 예뻐요	1	0	실
기자 샤벳	5	2022-01-11	2476 캔디	실 색상 너어어무 이뻐요!	0	0	실
기자 샤벳	5	2022-01-11	2478 실버블루	이쁜 티코스터 만들게요~!	0	0	실
기자 샤벳	5	2021-12-24	2478 실버블루	이뻐서 재주문합니다.	0	0	실
기자 샤벳	5	2021-12-24	2479 페퍼민트	잘 도착했습니다. ㅎㅎ	0	0	실
기자 샤벳	5	2021-12-24	70 스트로베리에이	이쁩니다~~^o^%	0	0	실
기자 샤벳	5	2021-12-05	2473 피지팝	색상도 화면과 비슷하고 배송이 빠릅니다	0	0	실
기자 샤벳	5	2021-10-16	2478 실버블루	실물이 훨씬 더 예뻐요!!	0	0	실
기자 샤벳	5	2021-10-16	2476 캔디	실물이 훨씬 더 예뻐요!!	0	0	실
기자 샤벳	5	2021-10-16	2479 페퍼민트	실물이 훨씬 더 예뻐요!!!	0	0	실
기자	5	2022-07-20	2191 크림	색이 너무 예뻐요 얘기 담요 만들려고 샀는데 완전 만족!!	0	0	실
기자	5	2022-07-20	2413 피치	색이 너무 예뻐요 얘기 담요 만들려고 샀는데 완전 만족!!	0	0	실
기자	5	2022-05-10	2204 오렌지	입니당 꼼꼼하게 배송됐어요 ㅎㅎ	1	0	실
기자	5	2021-06-05	2194 글레이셔	여름니트 뜨려구요^^ 실이 알고 아담하네용 광택이 나서 여름니트로 떠놓으면 넘 시원해보일 것 같아요	0	0	실
기자	4	2021-04-26	2414 틸	색이 이쁜데 생각보다 되게 작네요 해야될듯	0	0	실
기자	3	2021-02-04	2190 화이트	도일리 뜨려고 주문했는데 굽어서 다른용도로 써야겠어요	0	0	실
기자	3	2021-02-04	2191 크림	실이 생각한거보다 굽고 갈라지네요ㅜ	0	0	실
기자	5	2020-04-08	2195 페블	좋아요 감사해요 잘쓸개요	0	0	실
기자	5	2020-04-08	2205 브라운	사진하고는 색감이 다르지만	0	0	실
기자	5	2020-01-07	2202 레드	코바늘 초보라서 카네이션 준비 지금부터 시작하려고 주문했습니다. 빨간색 별로 좋아하지 않는 저도 반하게	0	0	실
기자	5	2020-01-07	2203 그린	코바늘 초보라서 카네이션 준비 지금부터 시작하려고 주문했습니다. 색상이 참 예쁘네요. 그런데 조금 더 진한	0	0	실
기자	5	2020-01-07	2412 퍼플	털실 알레르기가 약간 있어서 면사만 사용하는데 기자 면사는 정말 최고입니다!	0	0	실
기자	5	2020-01-07	2201 블랙	부드러워서 여러 용도의 소품들 다양하게 만들어 사용하고 있습니다.	0	0	실
기자	5	2019-12-21	2204 오렌지	잘드 초기햇소브니다. ㅎ	1	0	실
기자	5	2019-12-21	2199 플럼	잘 듀느거 가햇쓰니향. ㅎㅇ	1	0	실
기자	5	2019-11-06	2475 블루라군	실이 이뻐서 몇번이나 샀는지.. ㅎㅎㅎ	1	0	실
기자	5	2019-11-06	70 스트로베리에이	실이 너무 이쁩니다. ㅎㅎ	1	0	실
기자	5	2019-11-06	2471 레몬에이드	실이 너무 이쁩니다. ㅎㅎ	1	0	실

Chapter 2. DD의 스토리_전처리



Naver Smart Store Review

- ✓ 특수문자 제거
- ✓ 의미 없는 자음, 모음 제거
- ✓ 이모지 제거
- ✓ 맞춤법 검사
 - 1차 검수(hanspell library)
 - 2차 검수(수작업)
- ✓ 색상 레이블링

지난 겨울에 이걸로 양말 뜨개 네 했는데 너무 부드럽고 차콜 색깔 취향저격이라서 올해도 바라클라바 뜰려고 또 샀어요! 완죤 부드럽고 이뻐요 😊❤️



지난 겨울에 이걸로 양말 뜨개 했는데 너무 부드럽고 차콜 색깔 취향 저격이라서 올해도 바라클라바 뜰려고 또 샀어요 완전 부드럽고 이뻐요

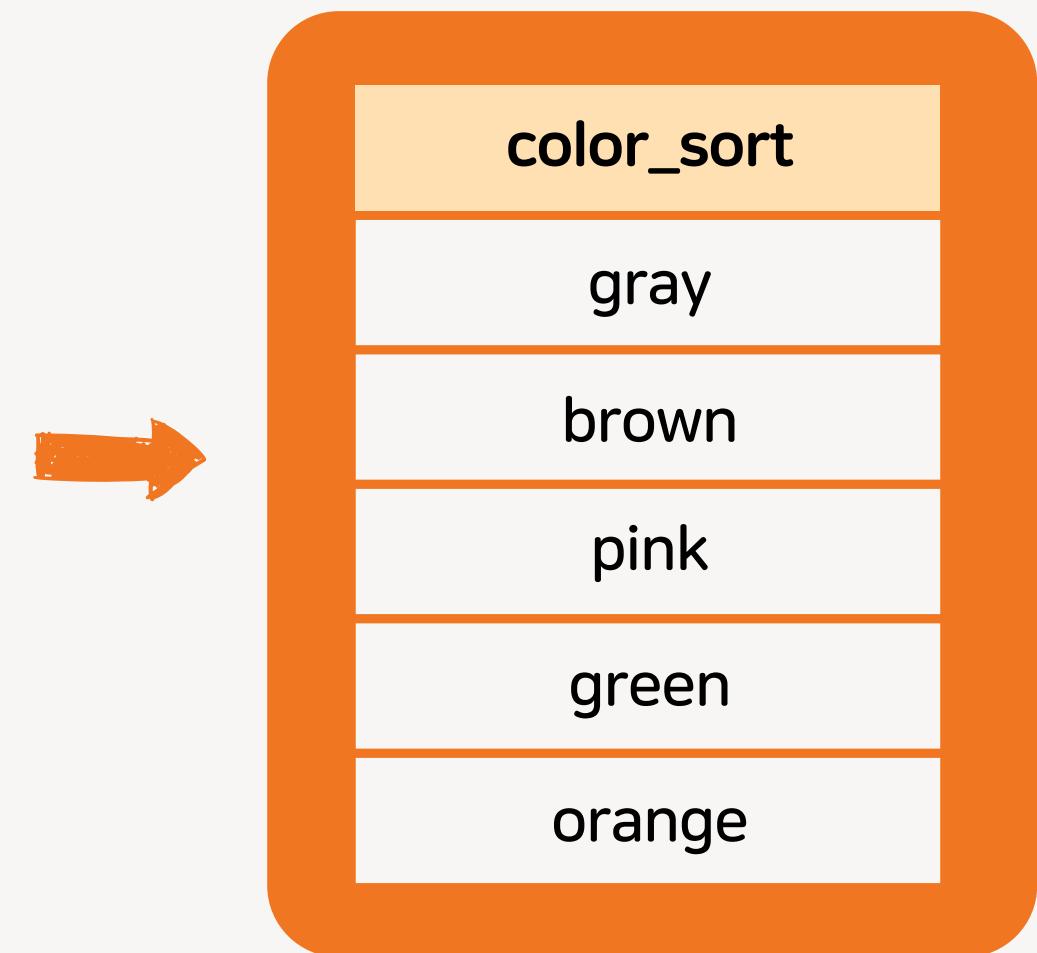
Chapter 2. DD의 스토리_전처리



Naver Smart Store Review

- ✓ 특수문자 제거
- ✓ 의미 없는 자음, 모음 제거
- ✓ 이모지 제거
- ✓ 맞춤법 검사
 - 1차 검수(hanspell library)
 - 2차 검수(수작업)
- ✓ 색상 레이블링

color
4326 슬레이트
03 코르크
1198 코튼캔디
209900 그레스그린
107172 퀴브르



*color_sort : 컬러 대분류(총 15가지 색)



리뷰 데이터 전처리가 끝나니
ERP 데이터가 들어왔다!

Chapter 2. DD의 스토리 _ ERP Data

3

수집 기간: 2020.01.01 ~ 2022.10.24

3

Chapter 2. DD의 스토리_스키마 정의서 작성



ERP Data

Table	Column_KR	Column_ENG	detail	data type	memo
customers	주문자 ID	id	고객 ID	string	
customers	구매횟수	cnt_orders	주문 횟수	float	
customers	주문자 연령	age	고객 나이	float	
customers	주문자 그룹	membership	회원 등급	string	일반회원 / VIP / VVIP / S-VVIP / D-VIP
customers	우편번호	postal_code	우편 번호	float	
customers	수령인 주소1	address	주소	string	상세 주소 제외
customers	주문일시	last_order_date	마지막 주문일	string	
customers	주문자 생일/성별	gender	성별	string	남성, 여성
customers		ageband	연령대	string	13-18, 20-24, 25-29, 30-34, 35-39
orders	주문번호	order_id	주문 번호	string	
orders	품목별 주문번호	order_detail	상품별 주문 번호	string	
orders	주문자 ID	id	고객 ID	string	
orders	주문상품코드	category	상품 카테고리	string	117, 490, 700, 139, 590, 131, 125
orders	옵션별상품코드	product_id	상품 고유 번호	string	
orders	주문상품명(기간)	product_name	상품명	string	
orders	구매횟수	cnt_orders	주문 횟수	float	
orders	주문품목 수량	qty	수량	float	
orders	주문품목 결제금	amount	단가	float	할인 적용된 단가
orders	주문품목 할인금	discount	상품별 할인 금액	integer	
orders	주문일시	order_date	주문 일시	timestamp	YYYY-mm-dd HH:mm:ss
orders	주문경로	order_path	주문 경로	string	
orders	주문검색어	search	주문 검색어	string	
orders	수령인 우편번호	postal_code	우편번호	integer	
orders	수령인 주소1	address	주소	string	상세 주소 제외
orders	배송구분	delivery	배송 구분	string	국내배송 / 해외배송
orders	결제금액(주문전)	total_price	최종 결제 금액	float	
orders	할인 상세내역(주)	discount_detail	할인 경로 및 금액	string	
orders	사용한 적립금	point	사용 적립금	float	
orders	결제방법	payment_type	결제 수단	string	카드(삼성페이, 페이코), 무통장, 예

orders	결제상태	payment_status	현재 결제상태	string	입금완료, 결제취소
orders	주문상태(품목별)	order_status	현재 주문상태	string	배송준비 거래완료
orders	입금확인일(결제)	payment_date	결제일	timestamp	YYYY-mm-dd HH:mm:ss
orders	송장입력일	order_invoicedate	송장 입력일	timestamp	YYYY-mm-dd HH:mm:ss
orders	배송중처리일	order_delivered_at	배송중 처리일	timestamp	YYYY-mm-dd HH:mm:ss
orders	배송완료일	order_delivered_at	배송 완료일	timestamp	YYYY-mm-dd HH:mm:ss
orders	사용한 네이버 포인트	naverpay	네이버페이 주문	integer	0: 네이버페이 주문
orders	사용한 예치금	deposit_use	예치금 사용 여부	integer	0: 예치금 사용 안함
orders		timerange	주문 시간대	string	0시대 ~ 23시대
orders		weekday	주문 요일	string	월 ~ 금
product		product_index	상품 인덱스	integer	
product	주문상품명(기간)	product_name	상품명	string	
product	주문상품명(옵션)	option	상품 옵션 디테일	string	
product	주문품목 단가	price	상품 단가	integer	
product	주문상품명(옵션)	color	색상	string	
product	옵션별상품코드	product_id	상품 고유 번호	string	색상 및 옵션 정보
product	옵션별상품코드	product_id_detail	상품 세부정보	integer	
category	주문상품코드	category	상품 카테고리	string	117, 490, 700, 131, 125
category		category_group	카테고리 그룹	string	실, 기타, 단추, 도자기

Chapter 2. DD의 스토리_테이블 분할

ERP Data

Customers Table_9 칼럼

Column	Detail
id	주문자 아이디
cnt_orders	이전 구매 횟수
age	주문자 나이
gender	주문자 생일, 성별 정보
membership	회원 등급
postal_code	수령인 우편 번호 6자리
address	수령인 주소 (상세 제외)
last_order_date	마지막 결제일
ageband	주문자 나이대

Orders Table_30 칼럼

Column	Detail
order_id	주문 고유 번호
order_detail	주문 고유 번호 + 품목 순번
category	상품 카테고리
product_name	상품명(기간제)
qty	주문 수량
amount	옵션 및 수량이 모두 적용된 상품 가격
discount	상품별 할인 금액
order_date	주문 일시
order_path	고객이 주문한 플랫폼
search	바늘이야기 유입 경로
delivery	국내 배송 / 해외 배송
timerange	결제시간대
weekday	요일

Products Table_7 칼럼

Column	Detail
product_index	주문 인덱스
product_name	상품명
option	상품 옵션 디테일
price	상품 단가
color	색상
product_id	상품코드
product_id_detail	상품코드 세부 정보

Category Table_2칼럼

Column	Detail
category	상품 카테고리
category_group	카테고리 그룹 이름

Chapter 2. DD의 스토리_전처리

ERP Data

Customers Table_9 칼럼

Column	Detail
id	주문자 아이디
cnt_orders	이전 구매 횟수
age	주문자 나이
gender	주문자 생일, 성별 정보
membership	회원 등급
postal_code	수령인 우편 번호 6자리
address	수령인 주소 (상세 제외)
last_order_date	마지막 결제일
ageband	주문자 나이대

Orders Table_30 칼럼

Products Table_7 칼럼

Column

Detail

Customers Table_전처리

- ‘주문자 생일/성별’ 칼럼에서 ‘성별’만 추출



- 연령대 칼럼 생성
- 회원 정보 중복된 사람들, 결제일 기준 정렬 후 최근 결제일 정보만 남기기

Chapter 2. DD의 스토리_전처리

ERP Data

Orders Table_전처리

- 주문 상품 코드 세 자리로 추출
118000000000 → 118
- 주문 옵션에서 괄호 및
괄호 내 텍스트 제거
- '네이버 페이 주문번호' 및
'사용한 예치금' 칼럼을
사용 여부(0/1)로 변경
- 결제 시간대 및 요일 컬럼 추가

last_order_date	마지막 결제일
ageband	주문자 나이대

Orders Table_30 칼럼

Column	Detail
order_id	주문 고유 번호
order_detail	주문 고유 번호 + 품목 순번
category	상품 카테고리
product_name	상품명(기간제)
qty	주문 수량
amount	옵션 및 수량이 모두 적용된 상품 가격
discount	상품별 할인 금액
order_date	주문 일시
order_path	고객이 주문한 플랫폼
search	바늘이야기 유입 경로
delivery	국내 배송 / 해외 배송
timerange	결제시간대
weekday	요일

Products Table_7 칼럼

Column	Detail
product_index	주문 인덱스
product_name	상품명
option	상품 옵션 디테일
price	상품 단가
color	색상
product_id	상품코드
product_id_detail	상품코드 세부 정보

Category Table_2칼럼

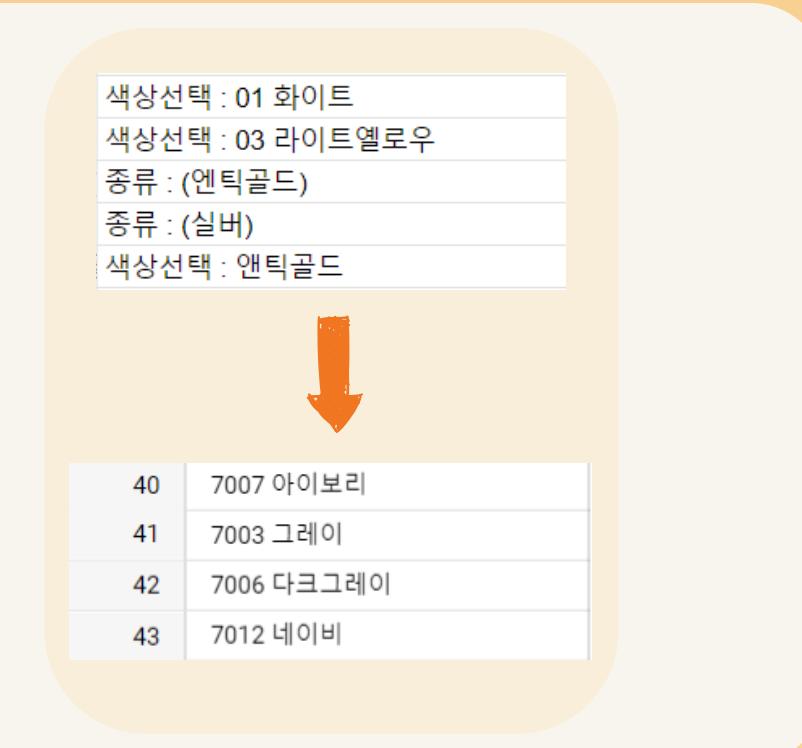
Column	Detail
category	상품 카테고리
category_group	카테고리 그룹 이름

Chapter 2. DD의 스토리_전처리

Products Table_전처리

- 상품 옵션에서 괄호 및 괄호 내 텍스트 제거
- 상품명에서 특수 문자 및 공백 제거
- 옵션에서 색상 정보만 추출 후 색상 칼럼에 추가
- 상품 코드 split

cnt_orders	이전 구매 횟수
age	주문자 나이
gender	주문자 생일, 성별 정보
membership	회원 등급
postal_code	수령인 우편 번호 6자리
address	수령인 주소 (상세 제외)
last_order_date	마지막 결제일
ageband	주문자 나이대



Category Table_전처리

- 주문 상품 코드 앞 세자리만 가져오기
118000000000 → 118
- 앞 세자리에 따라 카테고리 부여

행	category	category_group
1	117	실
...		
9	119	도서
10	120	바늘
11	12E	부자재

Products Table_7 칼럼

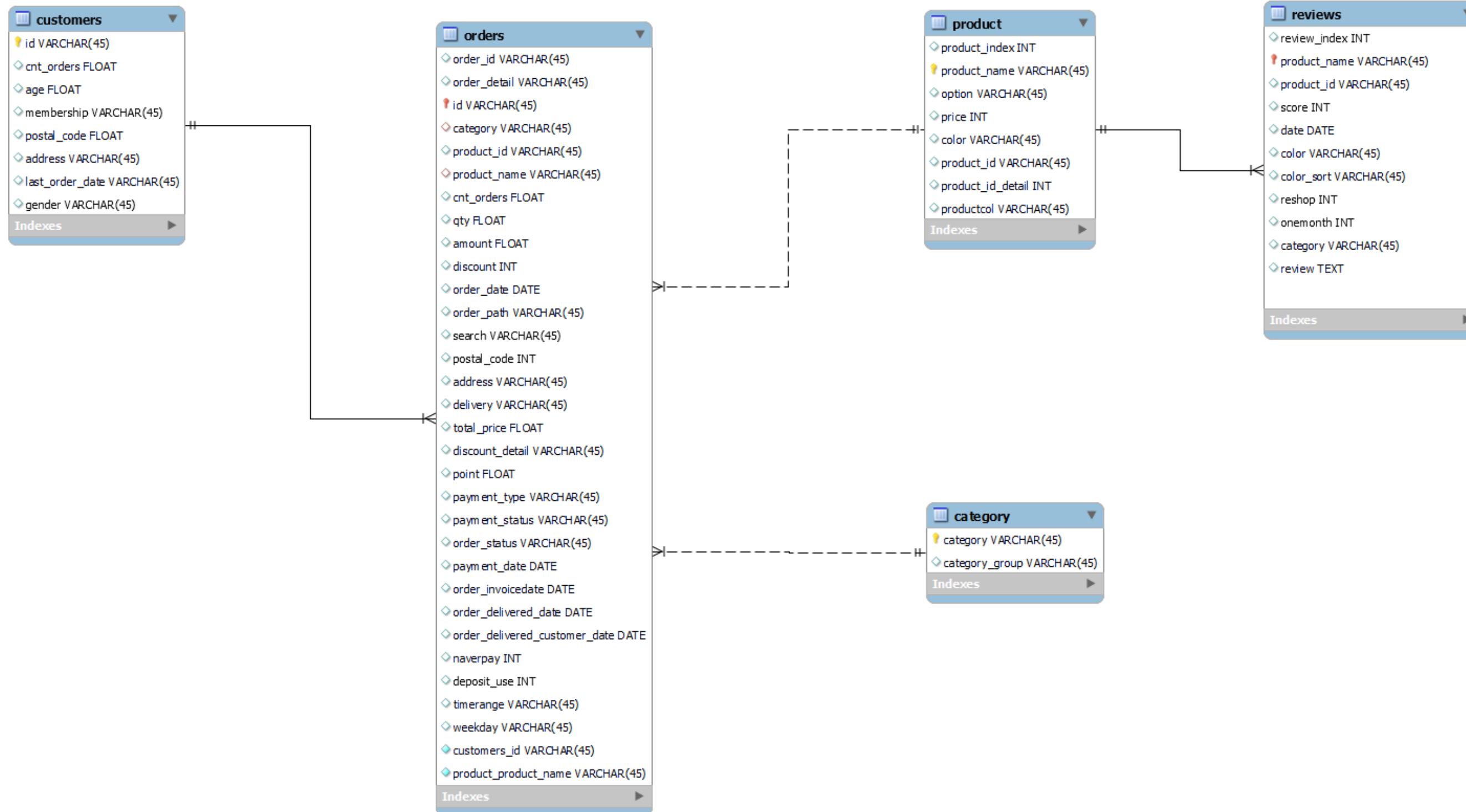
Column	Detail
product_index	주문 인덱스
product_name	상품명
option	상품 옵션 디테일
price	상품 단가
color	색상
product_id	상품코드
product_id_detail	상품코드 세부 정보

Category Table_2칼럼

Column	Detail
category	상품 카테고리
category_group	카테고리 그룹 이름

Chapter 2. DD의 스토리_ERD

ERP Data_네이버 스마트스토어 리뷰 데이터를 포함한 ERD 관계도

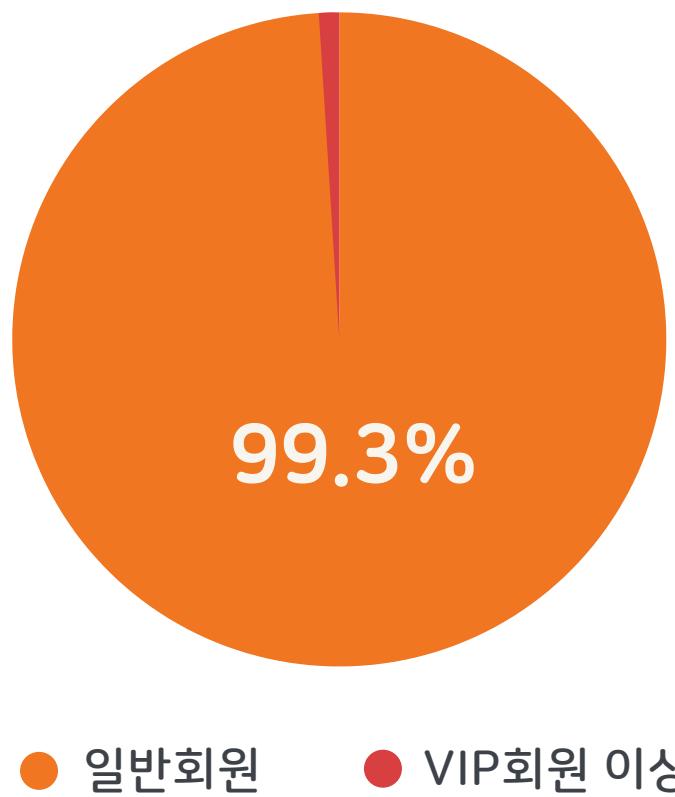


Chapter 2. DD의 스토리_현황 분석

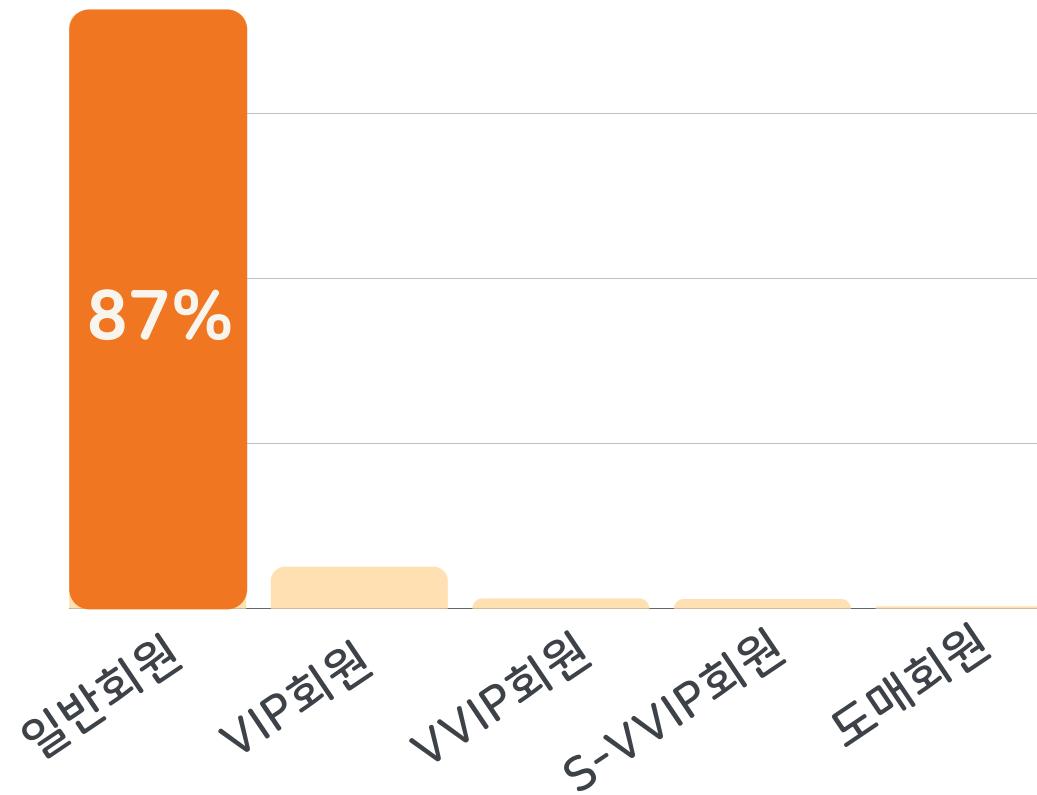
ERP Data



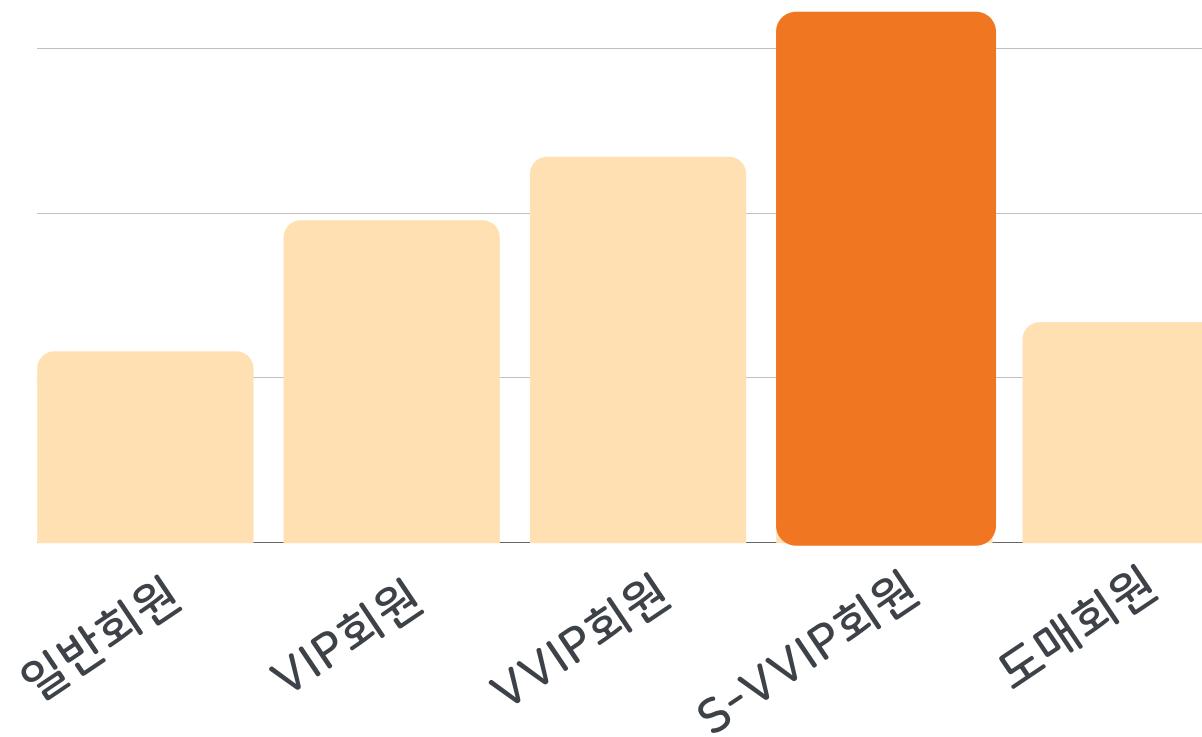
바늘이야기 회원 비율



회원등급별 총 구매액



회원등급별 결제자 인당 평균 구매액

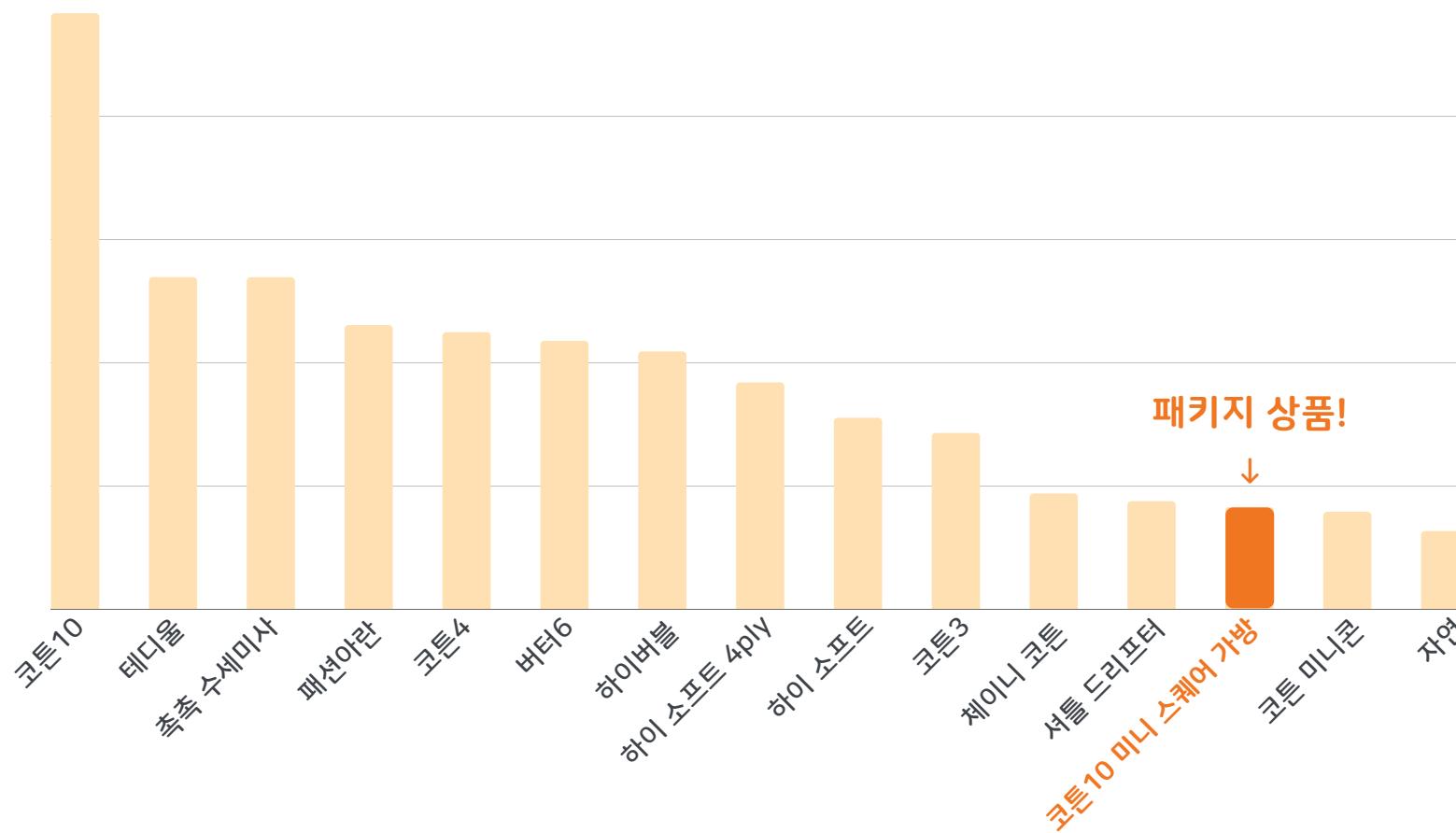


Chapter 2. DD의 스토리_현황 분석

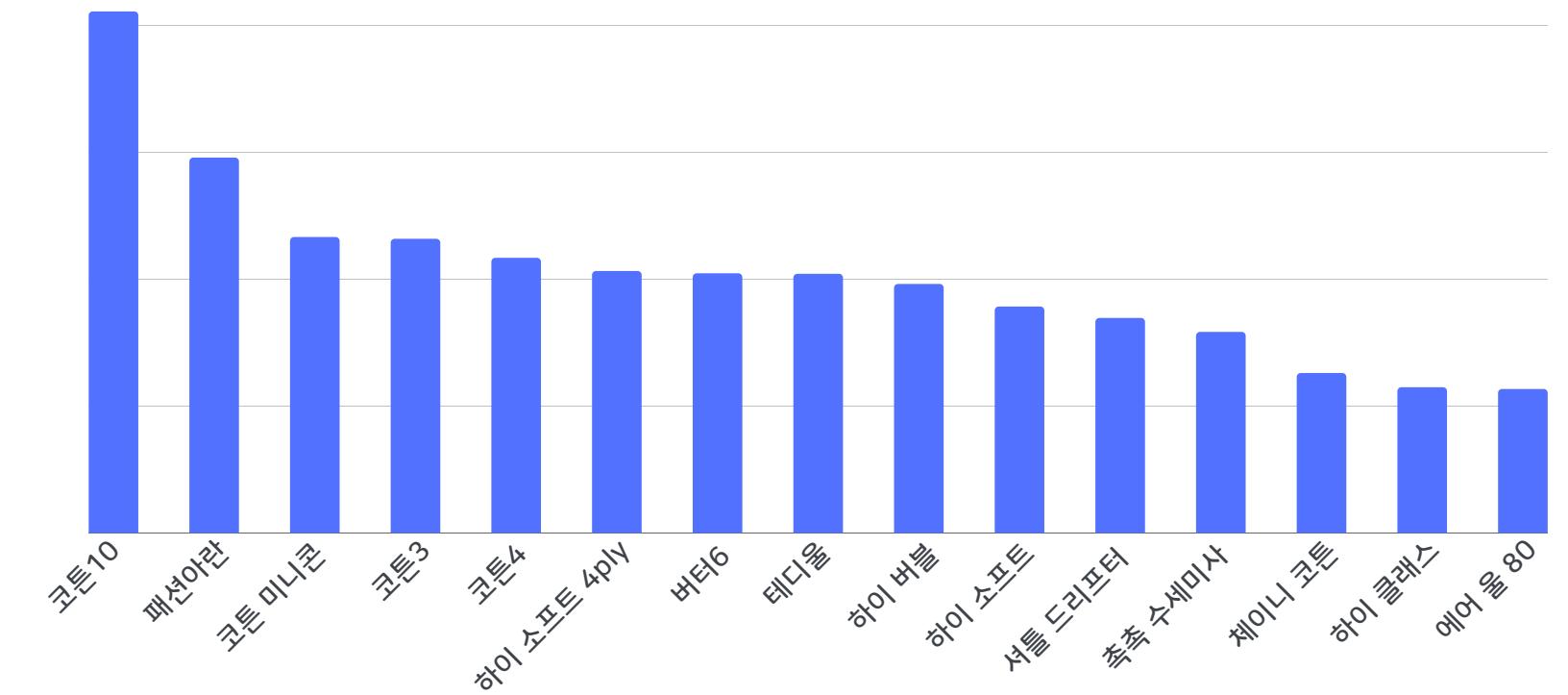
ERP Data



● 일반회원이 가장 많이 주문한 제품 Top 15



● VIP 회원이 많이 구매한 제품 Top 15

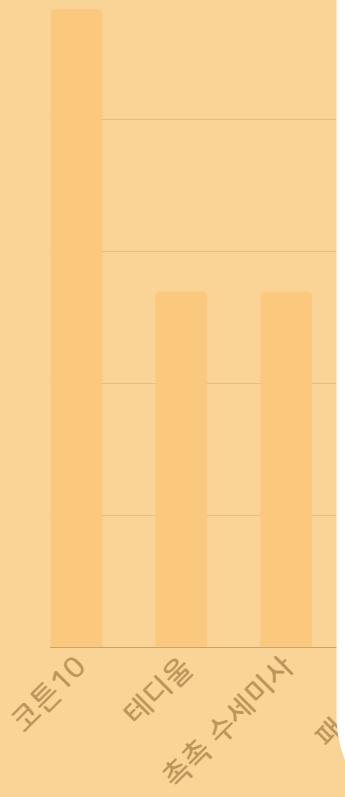


Chapter 2. DD의 스토리_현황 분석

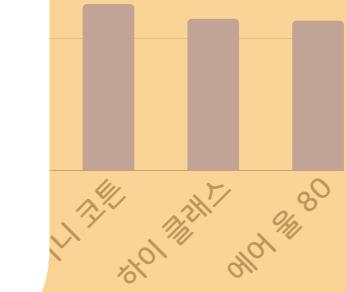
ERP Data



● 일반회원



● VIP회원



VIP회원들은 패키지 상품을 구매하지 않는군..

일반회원에 대한 분석에 집중해보자!

코튼

Chapter 2. DD의 스토리_분석



ERP Data_주제 1: 일반회원의 첫 구매 상품

	prod_id	product_name	cnt_
1	M3462165	코튼10	
2	M1151737	촉촉 수세미사	
3	M3462164	하이버블	
4	M2198380	코튼4	
5	M3446110	teddy 올	
6	M1169281	하이소프트 4ply	
7	M3377995	패션아란	
8	M6799	코튼3	
9	M3482619	버터6	
10	M1169279	하이소프트	
11	M3461734	코튼 미니콘	
12	M3396031	서틀 드리프터	
13	M3466740	체이니코튼	
14	M2197697	코튼2	
15	M3415214	자연	
16	M3415245	네온 홀릭	
17	M3377994	마제스틱	
18	M1189580	클라우드	
19	M2263437	까사리아	
20	M1144904	동방18합 색사	

21	M3443889	에어 올 80 superfine merino 20...
22	M1189581	하이클래스
23	M2143191	컬러풀 태그
24	M3439	자라
25	M8928	필 소프트
26	M1119647	필 라이트
27	M8091	자라플러스
28	M3396036	코튼 탑 DK
29	M3469103	새틴 알파카 모헤어
30	M3462147	포레스트
31	M2263339	필 에코코튼 100 오가닉 코튼 ...
32	M2165	자석똑딱단추 1
33	M3464759	코튼10 미니 스웨어 가방
34	M3444957	로비 키드모헤어 20g
35	M2263131	린넨 VIP
36	M3378863	핑고 트위드
37	M1215378	기자 샤벳
38	M2250730	돌핀 솔리드
39	M2198190	필 루스티크
40	M3490916	브레이드500

- 첫 구매 시, 대부분 '실' 구매
- 초보자 상품 중에서는 '코튼10 미니 스웨어 가방'에 대해 선호가 높음



코튼 10 미니 스웨어 가방과
울리울리 키치 브이넥 조끼는
대리님이 초보자를
겨냥하고 제작하셨던 상품!!

이 두 가지 상품을 사는 사람들이
다음 스텝으로 구매하는 아이템은 무엇일까?

Chapter 2. DD의 스토리_분석



ERP Data_주제 1: 일반회원의 첫 구매 상품(with 초보자 상품)

● 초보자 상품이 첫 구매인 고객들이 같이 구매한 상품

행	prod_id	product_name	cnt_
1	M3462165	코튼10	
2	M2224656	라바 실리콘 코바늘	
3	M8718	단수표시링	
4	M138	안전핀 둋바늘 괴배기바늘	
5	M2224656	라바 모사용 실리콘 코바늘	
6	M3378781	쪼꼬미 라벨	
7	M4703	에티모 모사용 코바늘	
8	M10339255	울리울리	
9	M133	줄바늘증	
10	M3462164	하이버블	
11	M3482619	버터6	
12	M6805	핑크 에티모 로즈 모사용 코바늘	

● 초보자 상품이 첫 구매인 고객들 중 재구매가 이뤄진 고객들이 같이 구매한 상품

행	prod_id	product_name	order_cnt	cnt_
1	M3462165	코튼10		
2	M8718	단수표시링		
3	M138	안전핀 둋바늘 괴배기바늘		
4	M3378781	쪼꼬미 라벨		
5	M4703	에티모 모사용 코바늘		
6	M2224656	라바 실리콘 코바늘		
7	M2224656	라바 모사용 실리콘 코바늘		
8	M6805	핑크 에티모 로즈 모사용 코바늘		
9	M3380927	플라스틱 왕 둋바늘		
10	M8724	플라스틱 둋바늘 세트		
11	M8717	단수링		
12	M3482619	버터6		

- 초보자 상품을 첫 구매한 일반회원들은 초보자 상품을 **중복 구매나 재구매 하지 않는다**
- 초보자 상품을 사용하면서 실이 부족할 것을 염려하여 **추가적으로 실을 구매하는 패턴**을 보인다.
- 초보자 상품으로 유입되어 **첫 구매한 뒤 일반 실, 부자재 등의 재구매 패턴 추측 가능**
- 재구매까지 이루어진 고객들은 **더 다양한 바늘을 구매하며 다른 아이템에 대해 도전**



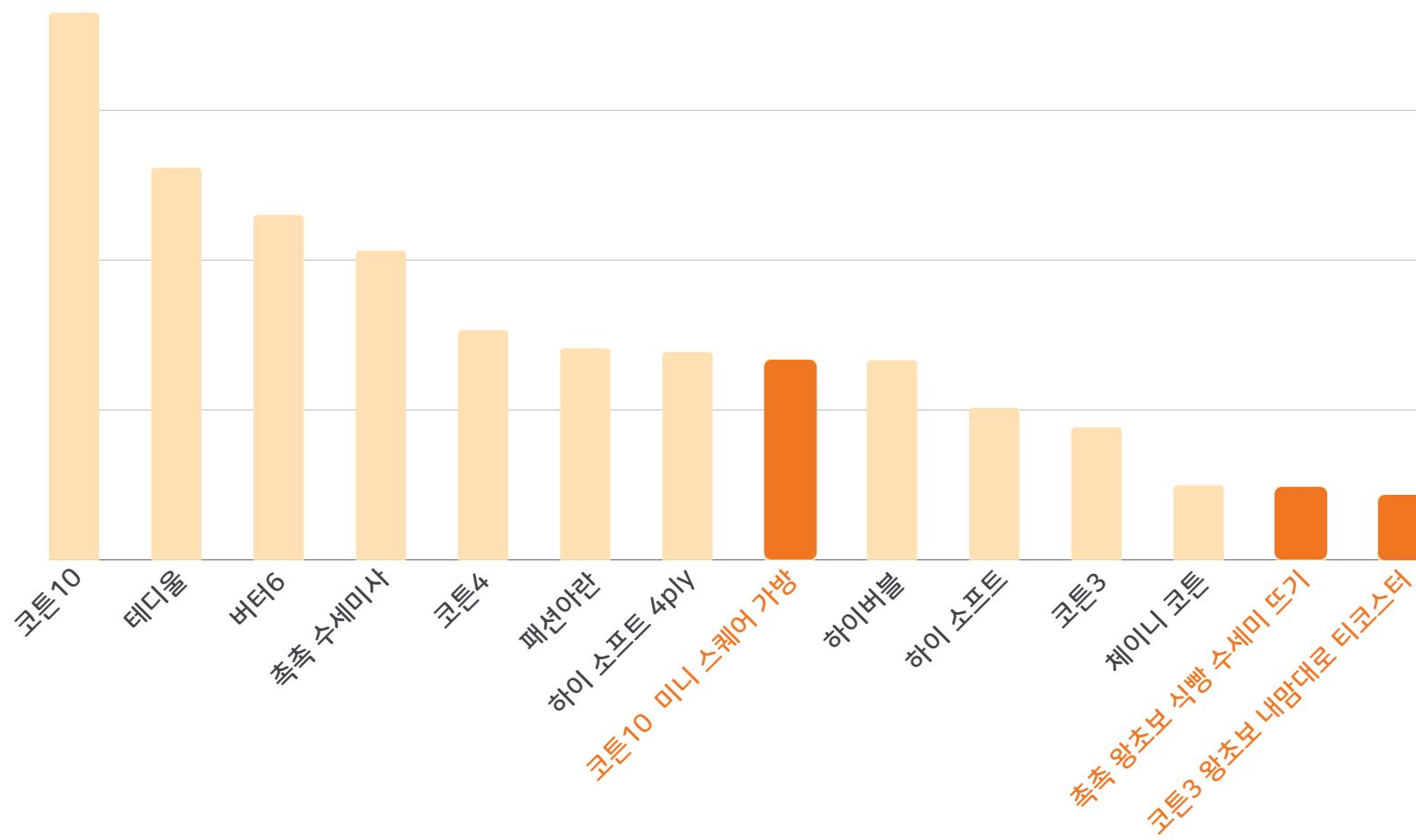
첫 구매 후 이탈한 고객들은
어떤 패턴을 가지고 있을까?

Chapter 2. DD의 스토리_분석

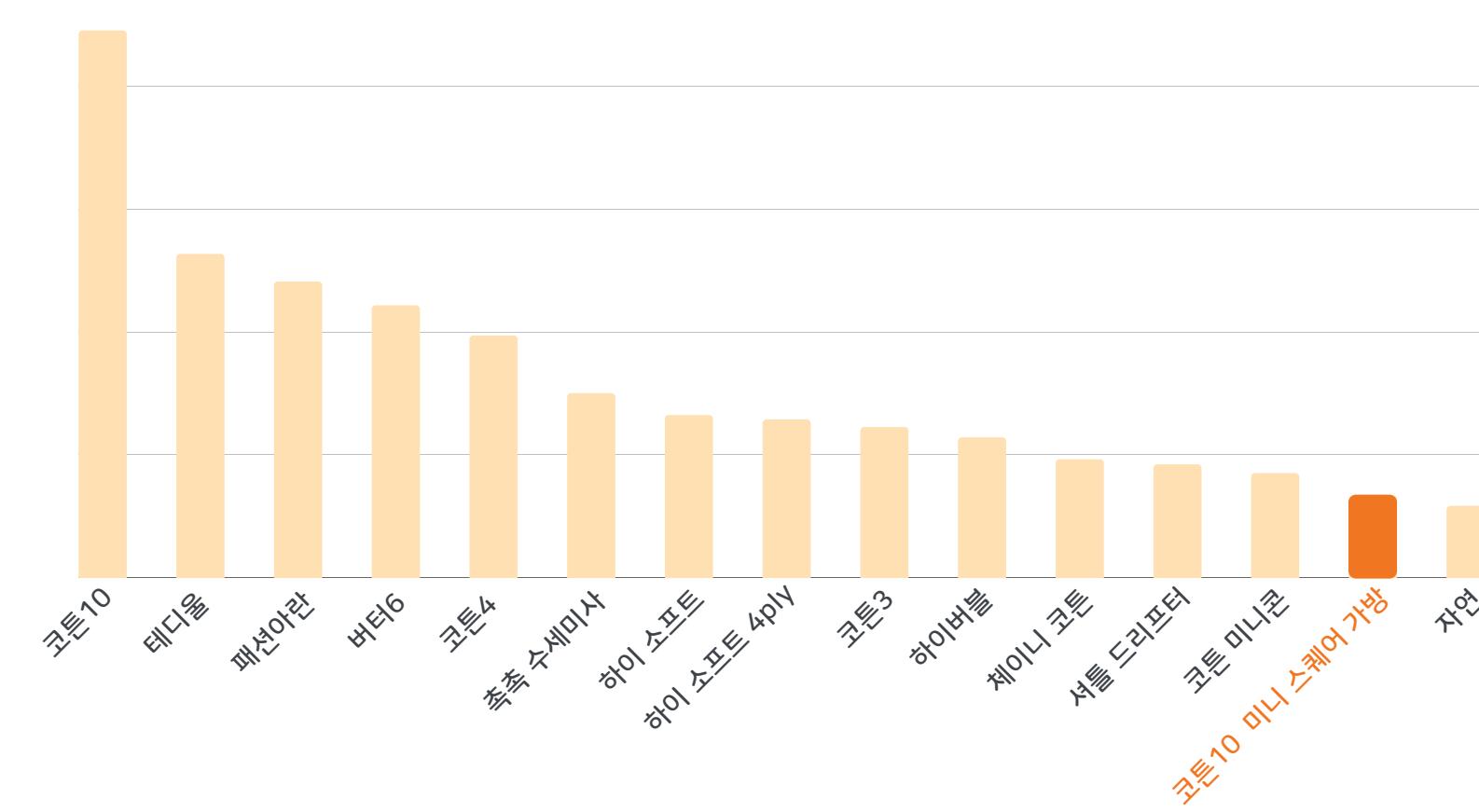


ERP Data_주제 2: 일반회원 중 첫 구매 후 이탈 고객 분석

● 첫 구매 후 이탈한 고객이 가장 많이 주문한 제품 TOP 15



● 재구매가 일어난 고객이 가장 많이 주문한 제품TOP15





첫 구매 후 이탈한 고객들은 패키지 상품을 많이 구매했구나~
이들이 구매한 패키지 상품을 살펴보자!

Chapter 2. DD의 스토리 분석

ERP Data_주제 2: 일반회원 중 첫 구매 후 이탈 고객 분석



코튼10 미니 스퀘어 가방

아프다, 작다, 모자르다, 손, 손잡이, 부족 등의 부정 단어 노출

가방을 끌 때 실이 부족하고 손이 아파 부정적인 감정으로 이탈했을 가능성 추측

A large orange speech bubble containing various Korean words and their meanings, such as '사다이쁘다' (beautiful), '남다' (many), '빠르다' (fast), '재밌다' (fun), '아쉽다' (sad), '되다' (become), '나오다' (come out), '있다' (exist), '파르다' (large), '작다' (small), '뿌듯하다' (excited), '어렵다' (difficult), '예쁘다' (beautiful), '모자라다' (over), and '귀엽다' (cute).



이미지 출처: 바늘이야기 홈페이지

Chapter 2. DD의 스토리_분석

ERP Data_주제 2: 일반회원 중 첫 구매 후 이탈 고객 분석



촉촉 왕초보 식빵 수세미 뜨기

'어렵다' 라는 단어 비중이 코튼10 미니 스웨어 가방보다 훨씬 큼

왕초보 타이틀이 달려있어 입문용으로 구매했지만 생각보다 어려워 이탈했을 가능성 추측



도전 만족 설명수세미 시간 식빵 동여상 실처음 퍼포먼스

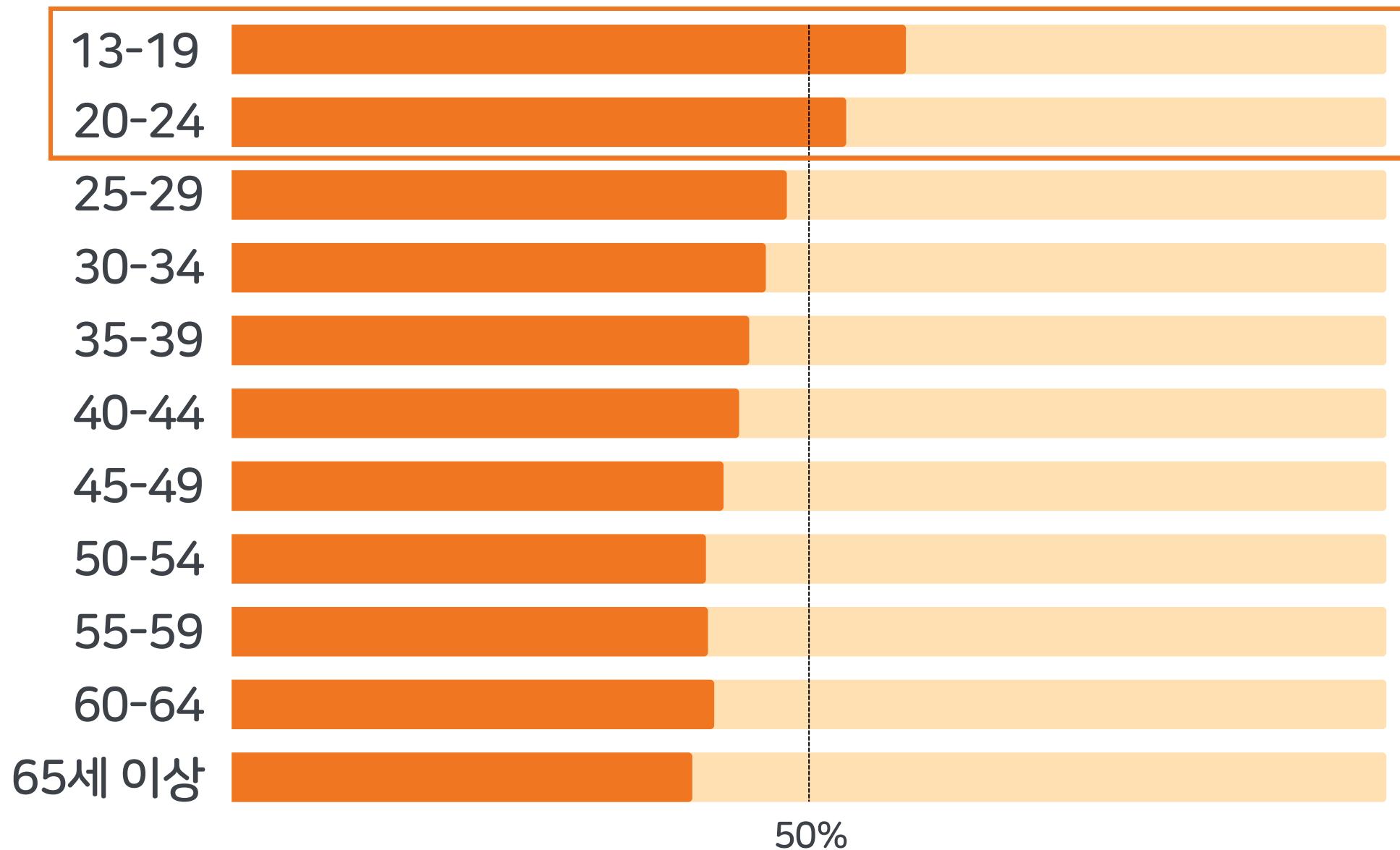
이미지 출처: 바늘이야기 홈페이지

Chapter 2. DD의 스토리_분석



ERP Data_주제 2: 일반회원 중 첫 구매 후 이탈 고객 분석

연령대별 이탈 비율



10대, 20대 순으로 이탈률이 높다!



바늘이야기의 희망 타겟은 20대 초반이라고
대리님이 말씀해주셨는데..



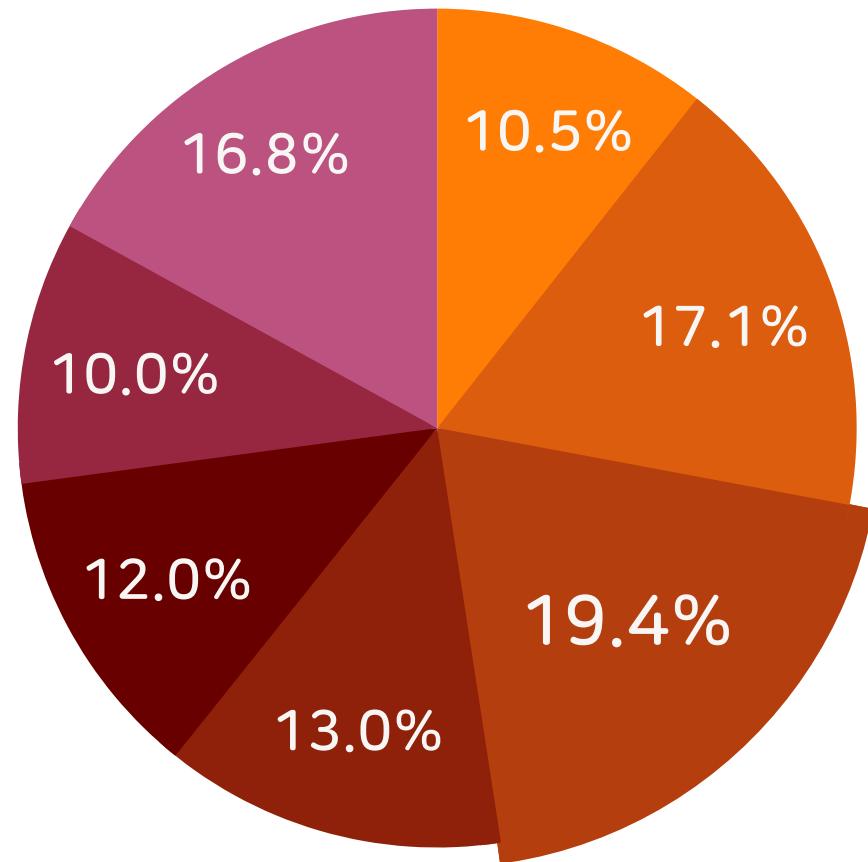
20대의 구매패턴은 타 연령층과 어떻게 다를까?

Chapter 2. DD의 스토리_분석



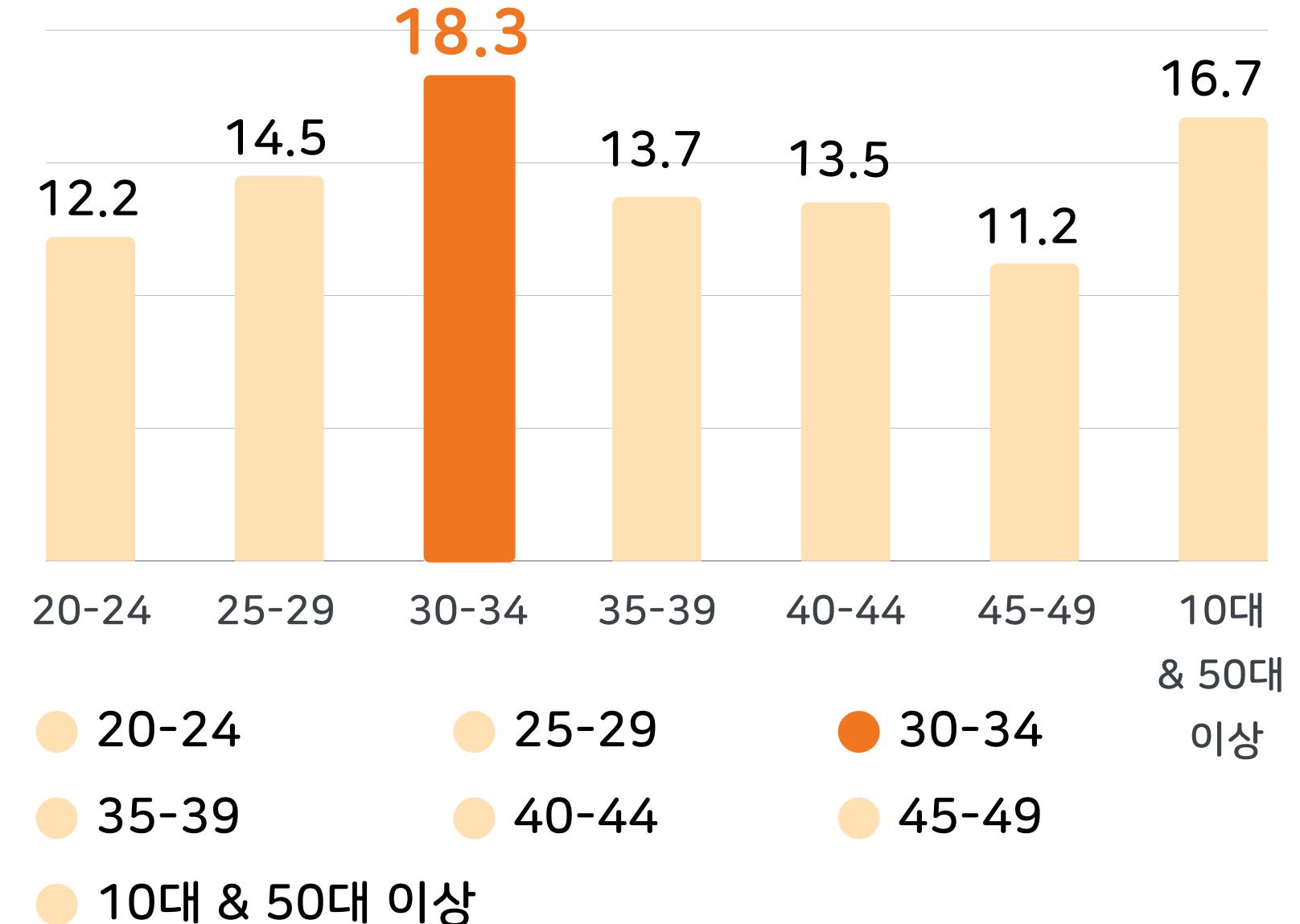
ERP Data_주제 3: 2030대 구매 트렌드 분석

● 일반회원의 연령대 비율



- 20-24
- 25-29
- 30-34
- 35-39
- 40-44
- 45-49
- 10대 & 50대 이상

● 일반회원의 연령대별 총 구매 비율



Chapter 2. DD의 스토리_분석



ERP Data_주제 3: 2030대 구매 트렌드 분석

● 각 연령대별 주요 구매_시간대

20-24대

목요일 1시대

금요일 1시대

수요일 1시대

수요일 23시대

수요일 21시대

25-29대

수요일 23시대

월요일 23시대

수요일 22시대

일요일 22시대

월요일 22시대

30-34대

수요일 23시대

수요일 13시대

수요일 15시대

수요일 22시대

일요일 22시대

Chapter 2. DD의 스토리_분석

ERP Data | 주제 3·2030대 그매 트렌드 분석



● 각

목요

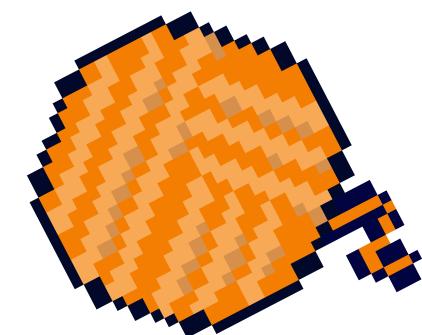
금요

수요

수요

수요

저녁-밤 시간대 구매라는 공통점이 있으나
20-24의 경우 주로 새벽 시간대에 구매



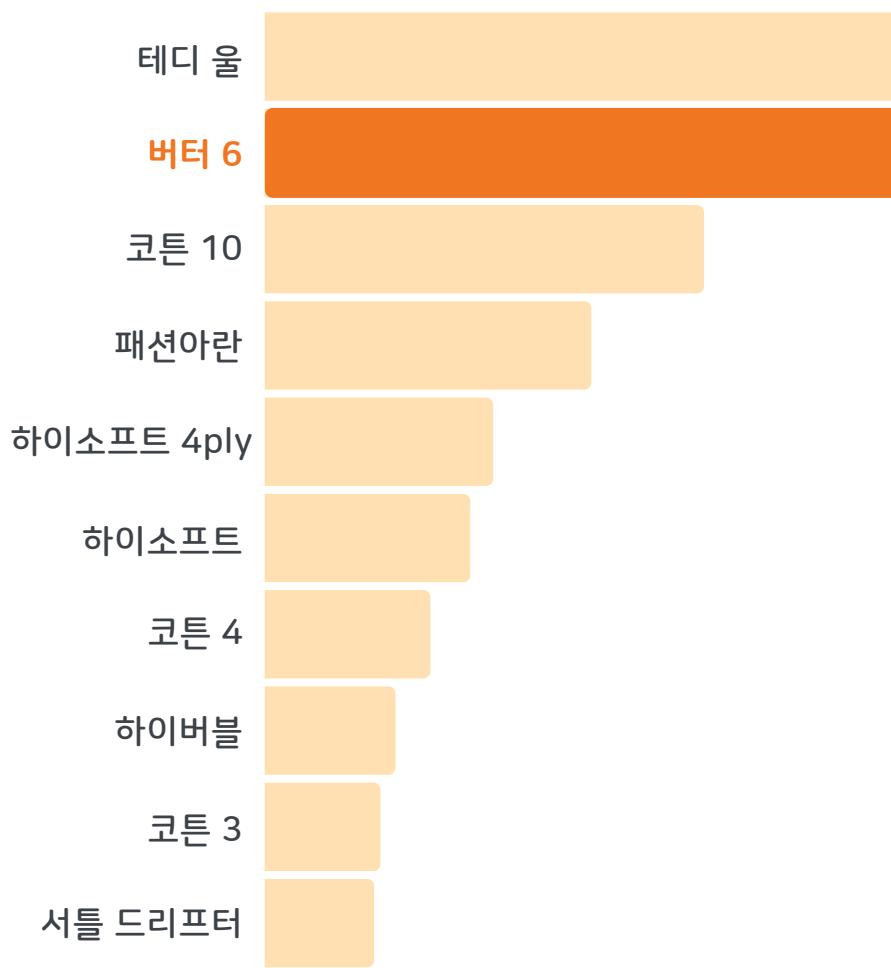
Chapter 2. DD의 스토리_분석



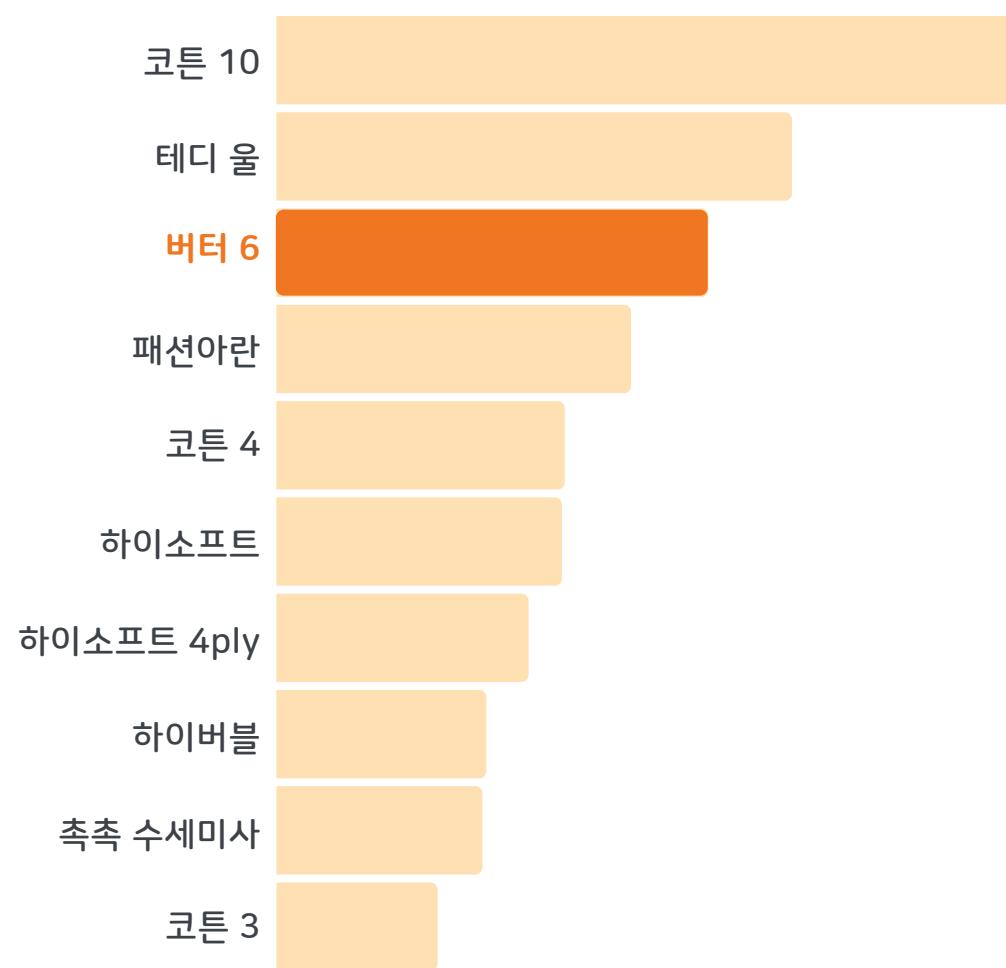
ERP Data_주제 3: 2030대 구매 트렌드 분석

● 각 연령대별 주요 구매_실

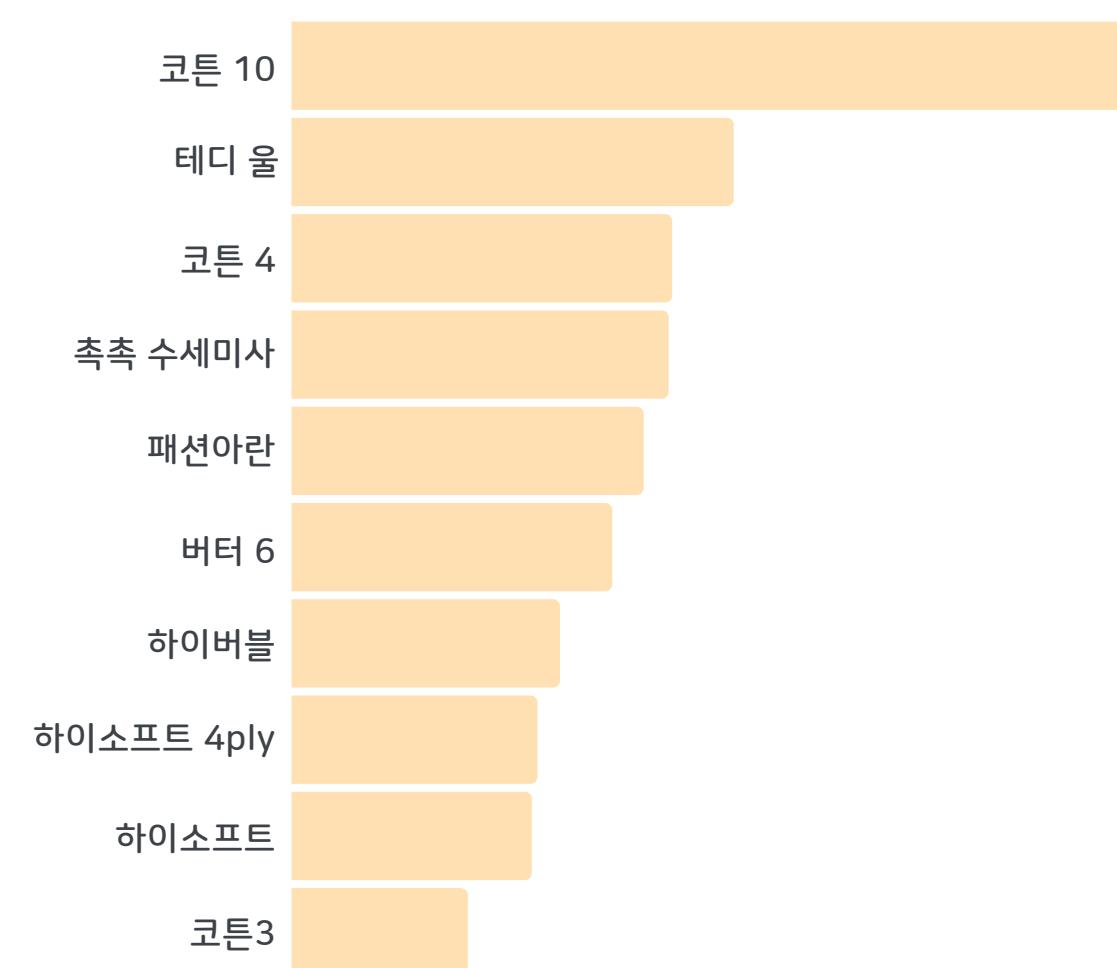
20-24대



25-29대



30-34대



Chapter 2. DD의 스토리_분석

ERP Data | 주제 3·2030대 그매 트렌드 분석



● 각

세 타겟의 구매 제품을 비교해 본 결과,

20-24, 25-29는



의 구매 순위가 높았고,

30-34는



의 구매 순위가 높았다!

또한 연령층이 어릴수록



구매순위 up



는 연령층이 올라갈수록 구매순위 up

이미지 출처: 바늘이야기 홈페이지

Chapter 2. DD의 스토리_분석



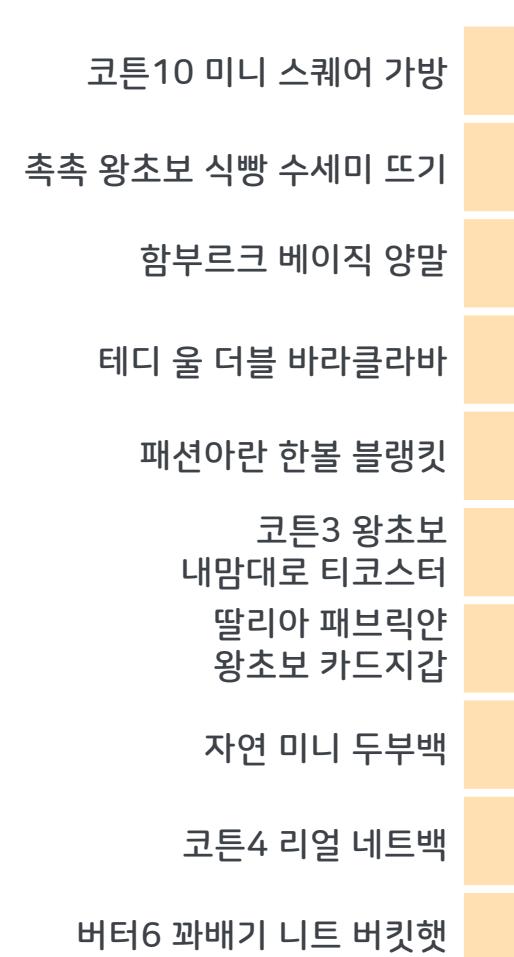
ERP Data_주제 3: 2030대 구매 트렌드 분석

● 각 연령대별 주요 구매_패키지

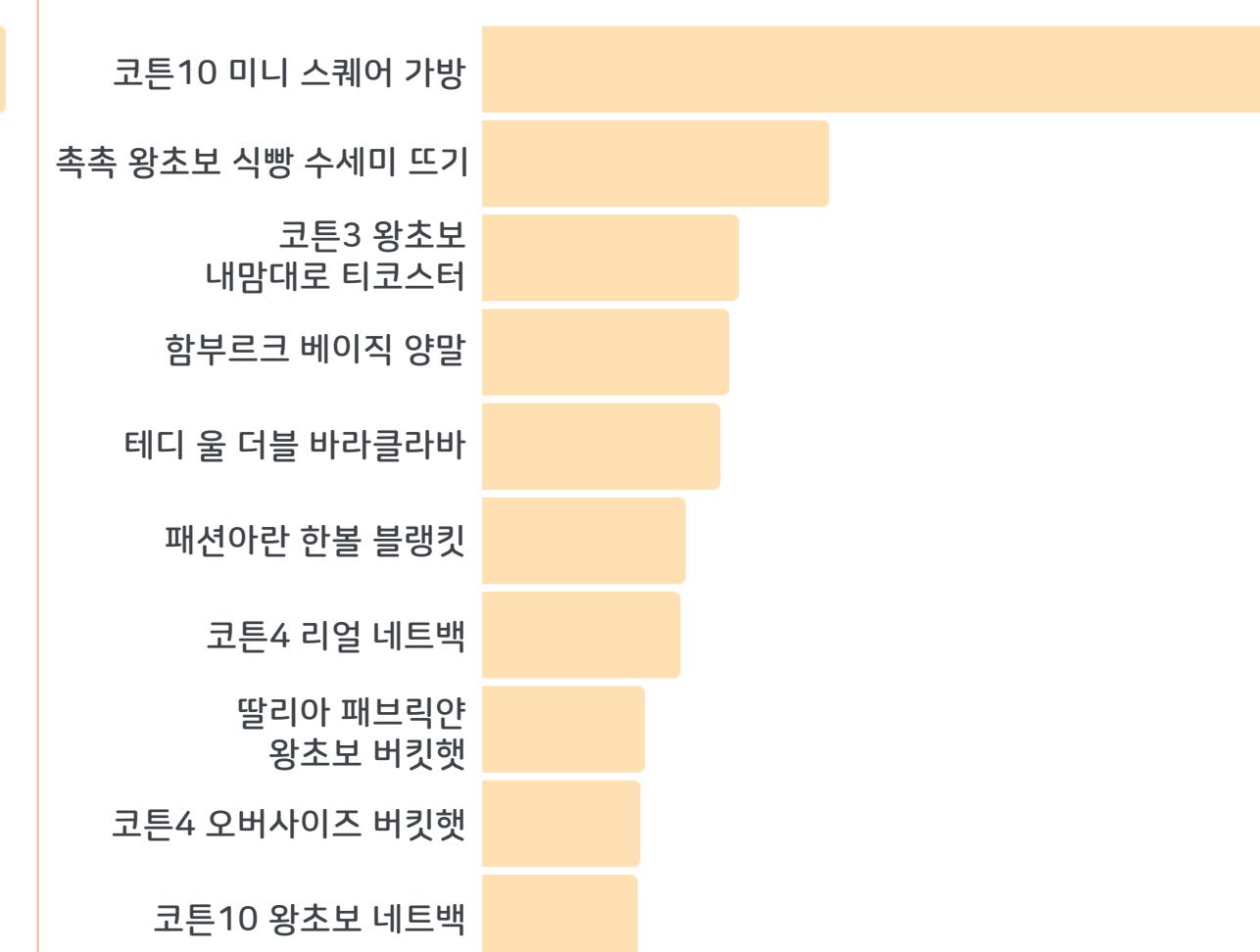
20-24대



25-29대



30-34대



Chapter 2. DD의 스토리_분석

ERP Data | 주제 3·2030대 그매 트렌드 분석

● 각

세 타겟의 구매 제품을 비교해 본 결과,

모두  을 1순위로 구매

25-29와 30-34의 상위권 순위는 거의 동일한 패턴을 보임

20-24는 2순위에



이 들어감

(유튜브의 영향이 있었을 것)

25-29, 30-34에서는 후순위에 들어가는 버터6를 이용한 패키지가 20-24에선 상대적으로 높은 순위에 집계됨

이미지 출처: 바늘이야기 홈페이지

Chapter 2. DD의 스토리_분석



ERP Data_주제 3: 2030대 구매 트렌드 분석

● 20-24 구매율이 높았던 제품 유튜브 시청 평균

연령	패션아란	목도리	숄더백	평균
13-17	5.1%	4%	0.5%	3.2%
18-24	28.4%	26.1%	34.8%	29.8%
25-34	29.5%	27.3%	42.9%	33.2%
35-44	18.2%	20.3%	10.8%	16.4%
45-54	13.6%	17.3%	9.3%	13.4%
55-64	4.3%	4.7%	1.6%	3.5%
65이상	1.1%	0.4%	0.2%	0.6

● 전체

시청자 연령	조회수
만 13-17세	3.8%
만 18-24세	21.6%
만 25-34세	26.9%
만 35-44세	19.4%
만 45-54세	16.9%
만 55-64세	8.4%
만 65세 이상	3.0%



20-24, 25-34 구매를 비교해보니,
20-24의 경우 구매에 유튜브 등의
바이럴 콘텐츠에 영향을 받는구나!

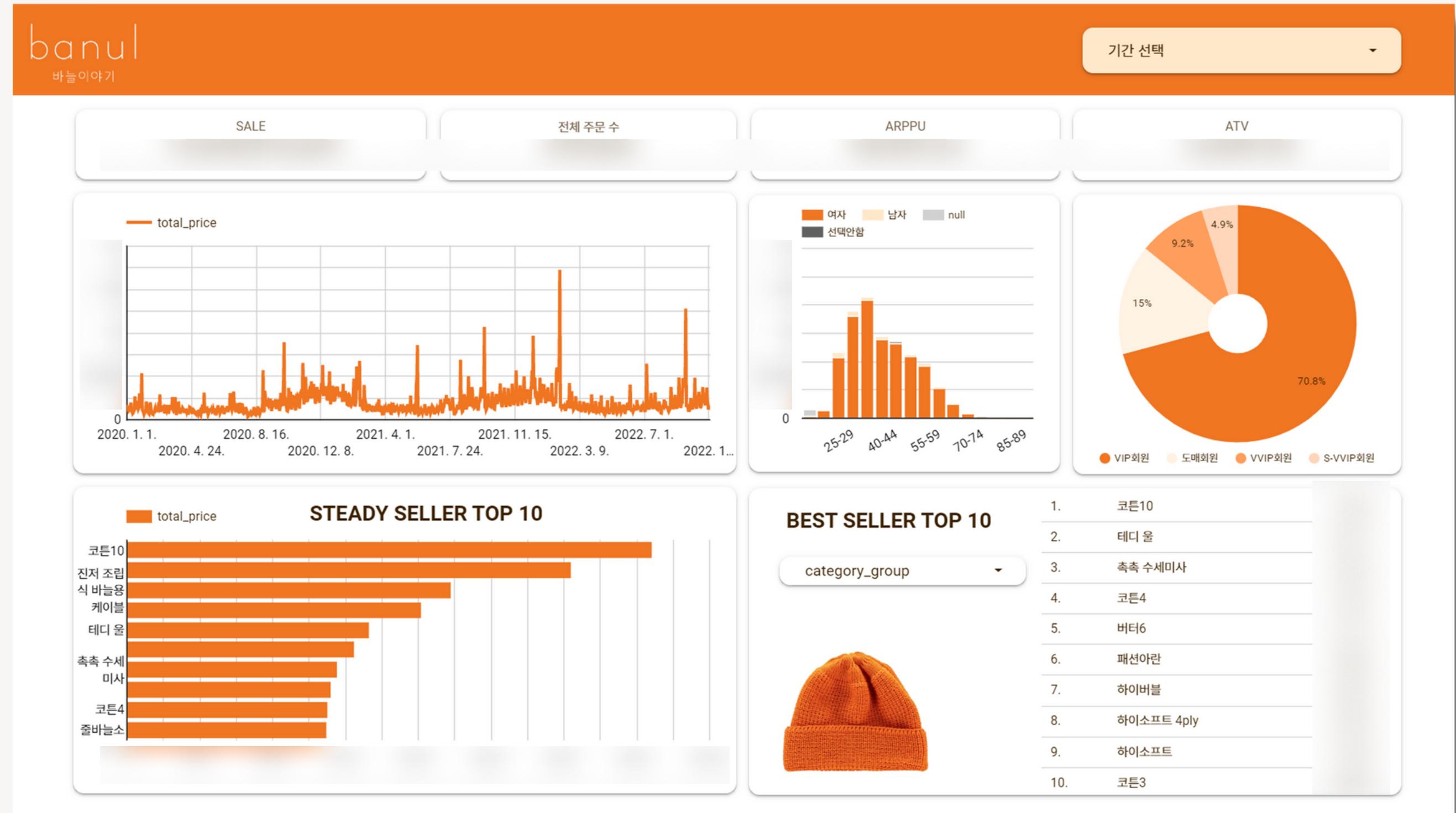


해당 분석 내용의 흐름을 지속적으로 보면서
마케팅 인사이트를 도출하고 싶은데..

Chapter 2. DD의 스토리_분석(대시보드 Looker Studio)



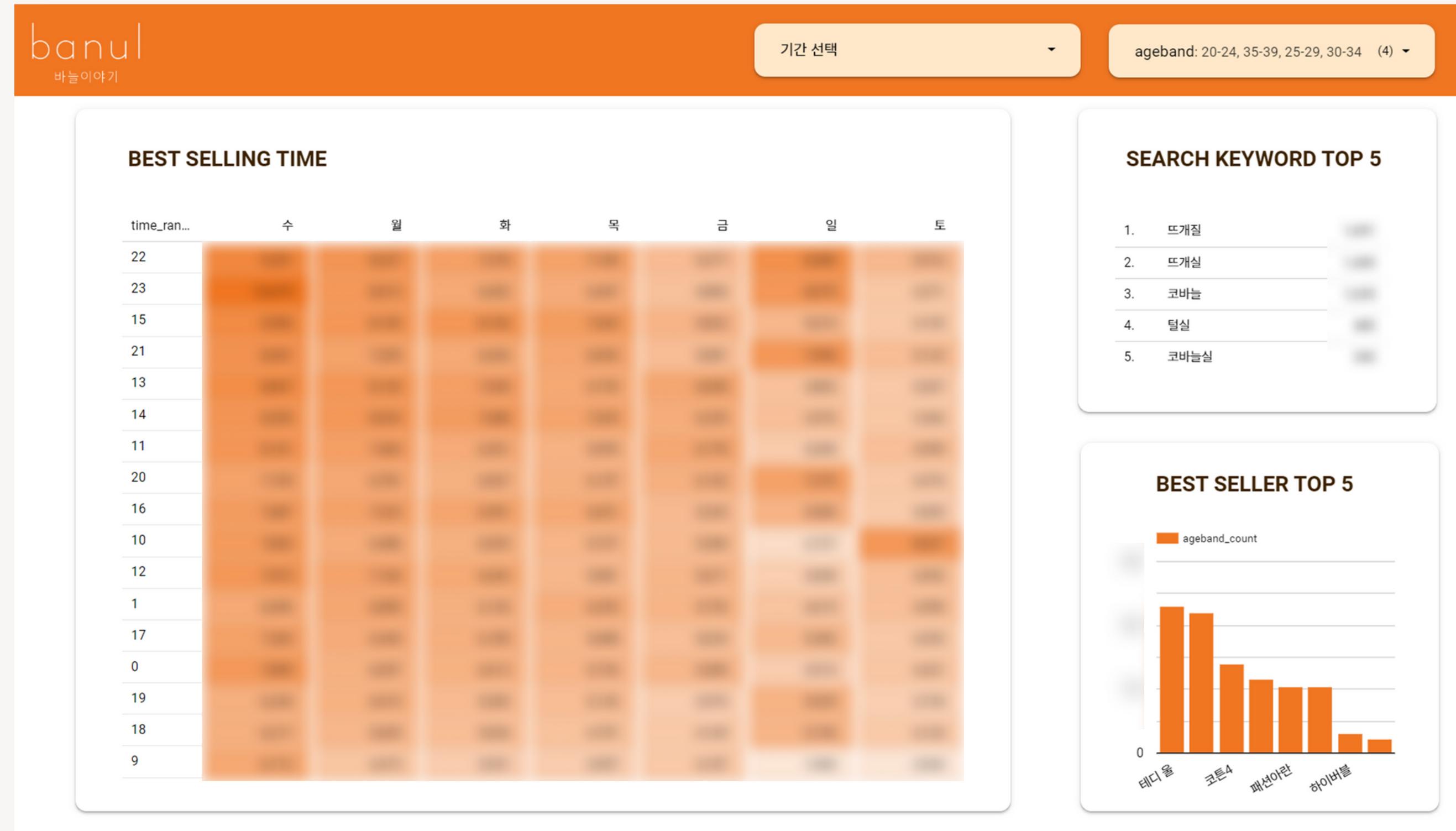
ERP Data_전체 현황



Chapter 2. DD의 스토리_분석(대시보드 Looker Studio)



ERP Data_연령 별 판매 현황





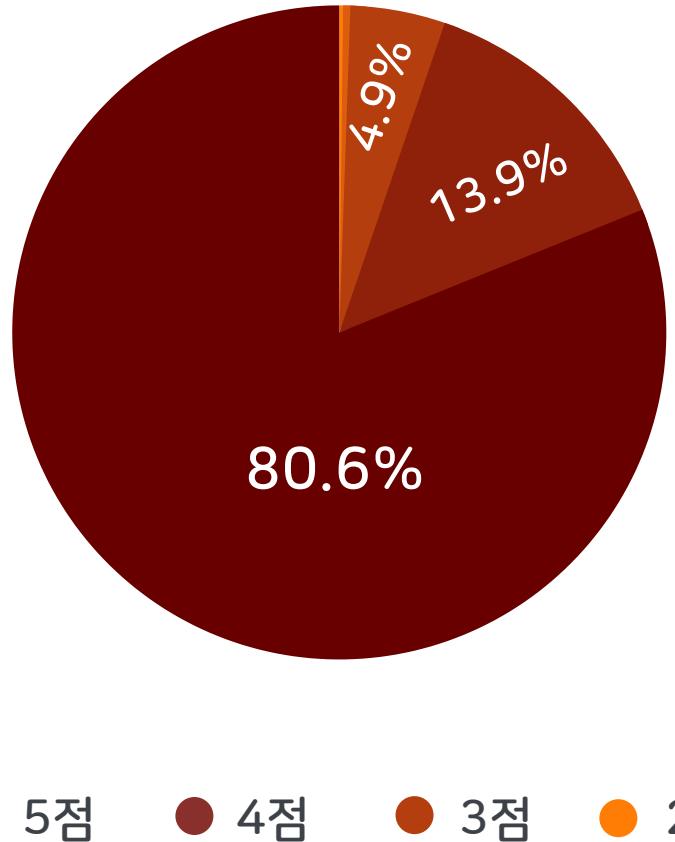
이제 리뷰 데이터로 분석을 시작해보자!

Chapter 2. DD의 스토리_분석

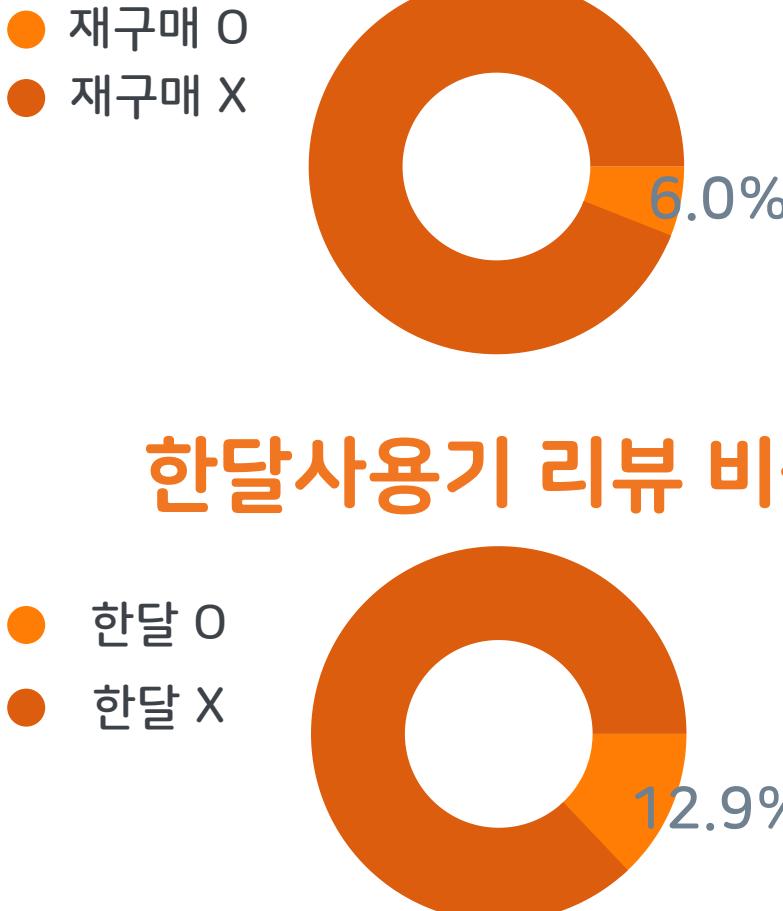


Naver Smart Store Review_리뷰 현황 분석

별점 분포



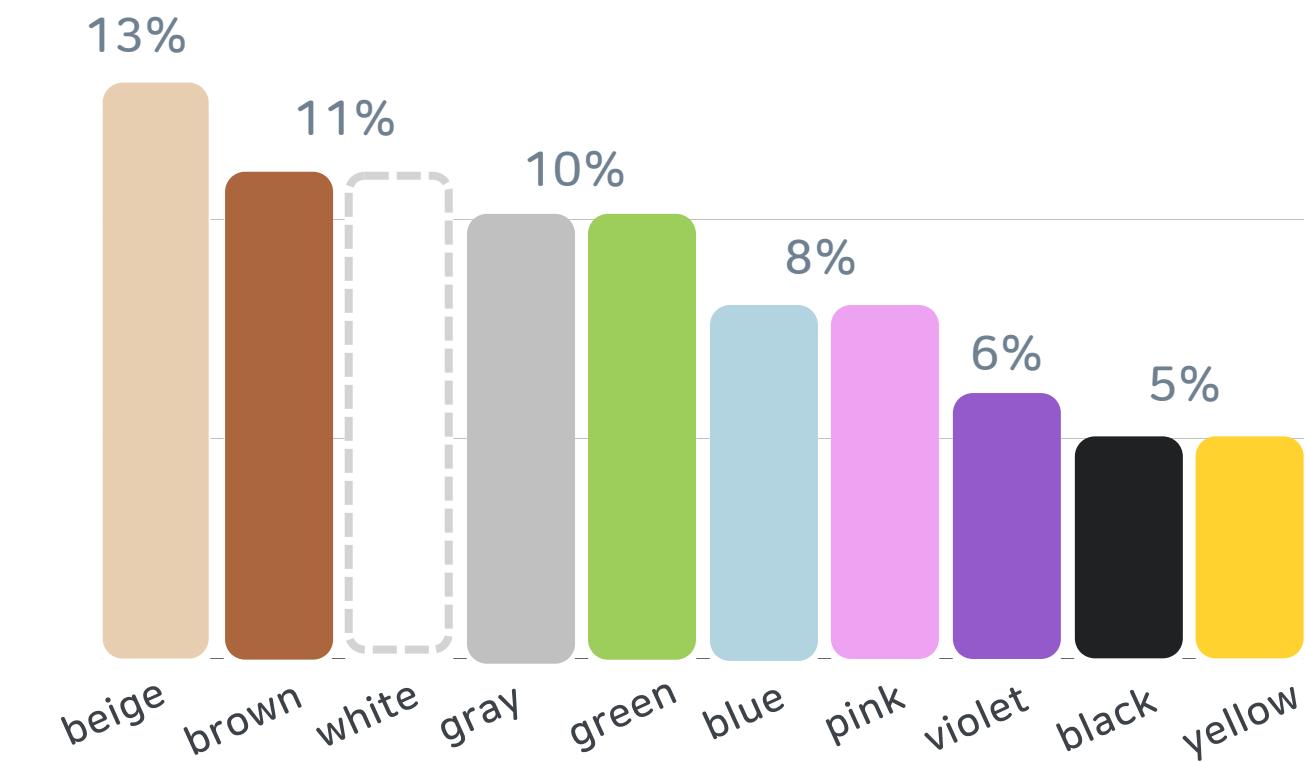
재구매 리뷰 비율



한달사용기 리뷰 비율



구매한 상품의 선호 컬러 분포





리뷰에 어떤 단어들이 많이 쓰이는지
형태소 분석을 통해 알아보자!

Chapter 2. DD의 스토리_분석



Naver Smart Store Review_1단계: 리뷰 형태소 분석하기

Kiwi

-한국어 형태소 분석기

대분류	태그	설명
체언 (500개)	NNG	일반 명사
	NNP	고유 명사
용언 (500개)	VV	동사
	VA / VA-I	형용사



체언 빈도수

-상위 5개

형태소	빈도
실	3311
동영상	2001
배송	1407
처음	1320
설명	1258

용언 빈도수

-상위 5개

형태소	빈도
하	4798
좋	4635
뜨	3431
만들	2956
있	2081



의미 있는 정보의 기준은 무엇일까?
일단 자주 나오는 단어들을 묶어보자!

Chapter 2. DD의 스토리_분석



Naver Smart Store Review_2단계: 키워드 사전 만들기

- 주제가 다른 키워드가 섞인 리스트

형태소	개수
실	3311
동영상	2001
배송	1407
처음	1320
설명	1258
선물	1258
즐겁다	1258
멘붕	1258

- 특정 주제로 묶은 키워드 사전



색상 = ['색깔']

구매자 상태= ['처음']

난이도 = ['어렵다', '멘붕']

감정= ['즐겁다', '재미']

구성=['동영상']

:

Chapter 2. DD의 스토리_분석



Naver Smart Store Review_3단계: 카테고리에 속하는 형태소 카운트

카테고리로 분류된 형태소가 많이 카운트될수록 리뷰는 유의미한 내용을 포함한다

리뷰	색상	사이즈	구매자 상태	품질	난이도	구성	배송	기타	구매 목적	감정	합계
재미나요 손재주 1도 없는 똥손이에요 실이 얇기도 하고 실 주변에 잔털이라고 하나요...	0	0	7	4	8	3	1	13	9	7	52
뜨개질해서 선물로 많이 나눠줬어요.	0	0	0	0	0	0	0	3	2	0	5
배송도 빨라요 사이트 지인한테 전달받고 구매해서 ...	0	0	0	0	0	0	2	1	2	0	5
처음 하는 거라 동영상 여러 번 보면서 ...	0	0	1	0	0	2	0	0	0	2	5
딱 맞아요	0	0	0	0	0	0	0	0	0	0	0



유의미한 내용 중에서도 더 중요한 게 있지 않을까?
TF-IDF*로 가중치를 줘보자!

Chapter 2. DD의 스토리_분석



Naver Smart Store Review_4단계: TF-IDF



★★★★★ 5

21.11.01. | 신고

색상선택: 005 멜란지브라운

좋아요좋아요좋아요좋아요 좋아요좋아요좋아요좋아요 좋아요좋아요좋아요좋아요



★★★★☆ 4

22.05.13. | 신고

기본 색상 구성: 하이버블 43번 2불 + 3번 3불 + 도안

동영상보면서 하니까 금방하네요 만들고 나서도 뿌듯하고 너무 좋아요 세정력은 모르겠지만 선물용으로 귀엽습니다

길이가 길더라도 다양한 정보를 담기보다는 유사한 내용을 길게 풀어쓰는 리뷰가 존재한다.
이러한 이유로 문서 빈도 가중치를 적용한 TF-IDF 모델을 사용하기로 했다.

Chapter 2. DD의 스토리_분석



Naver Smart Store Review_4단계: TF-IDF 적용 전

단순 키워드 언급량만으로 리뷰의 점수를 계산한 경우

리뷰	색상	사이즈	구매자 상태	품질	난이도	구성	배송	기타	구매 목적	감정	합계
뜨개질해서 선물로 많이 나눠줬어요	0	0	0	0	0	0	0	3	2	0	5
처음 하는 거라 동영상 여러 번 보면서 …	0	0	1	0	0	2	0	0	0	2	5

Chapter 2. DD의 스토리_분석

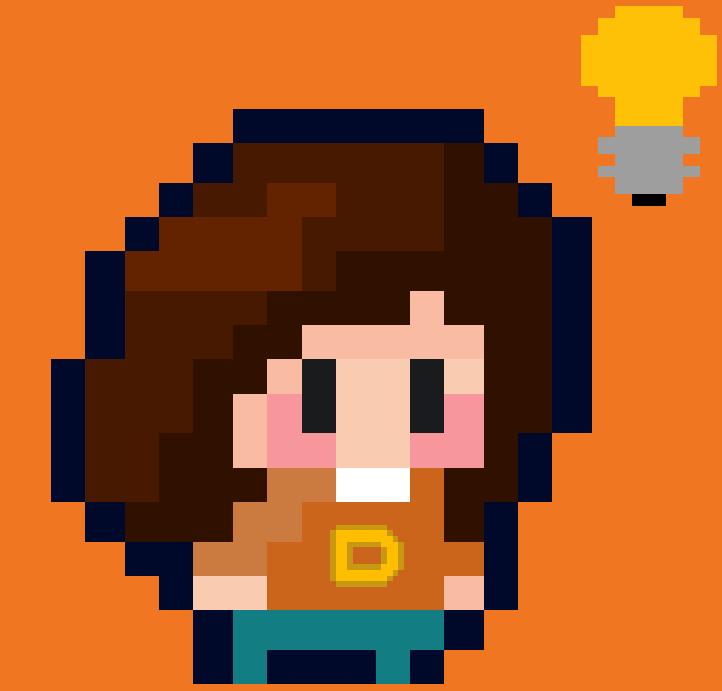


Naver Smart Store Review_4단계: TF-IDF 적용 후

*레이블링? 전체 리뷰의 TF-IDF 합계 중간값보다 크거나 같은 경우 1, 미만인 경우 0으로 레이블링 했다.

리뷰	색상	사이즈	구매자 상태	품질	난이도	구성	배송	기타	구매 목적	감정	TF-IDF 합계	레이블링*
뜨개질해서 선물로 많이 나눠줬어요	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.55	0.00	1.39	0
처음 하는 거라 동영상 여러 번 보면서	0.00	0.00	0.33	0.00	0.00	0.67	0.00	0.00	0.00	0.67	1.67	1

단순 키워드의 합은 같더라도 더 다양한 정보를 갖고 있는
두 번째 리뷰의 TF-IDF 합산 지수가 더 높은 것을 발견할 수 있다.



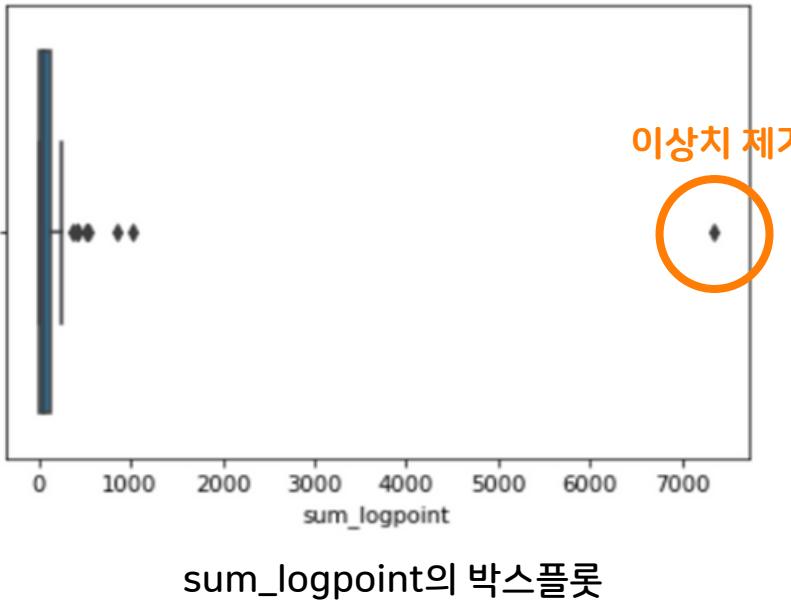
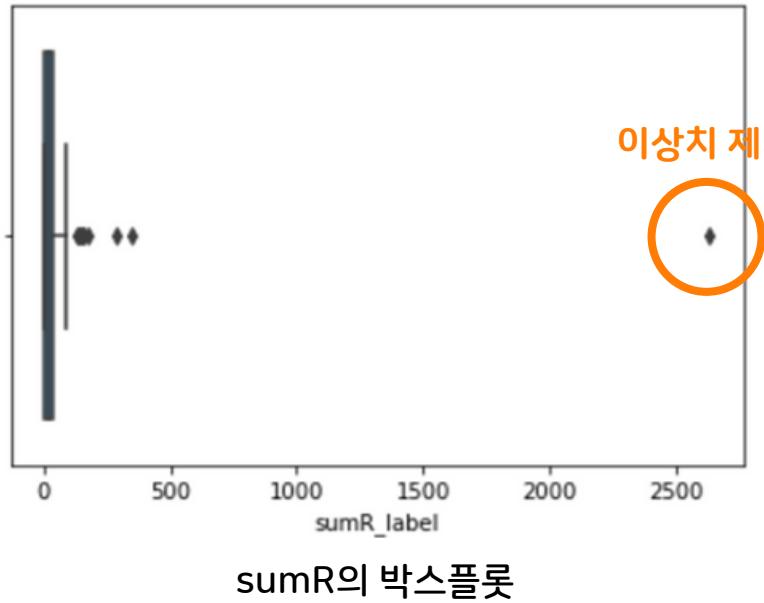
초중급 고객들이 왕초보 고객*보다 좀 더
유의미한 리뷰를 작성하지 않았을까?

Chapter 2. DD의 스토리_분석



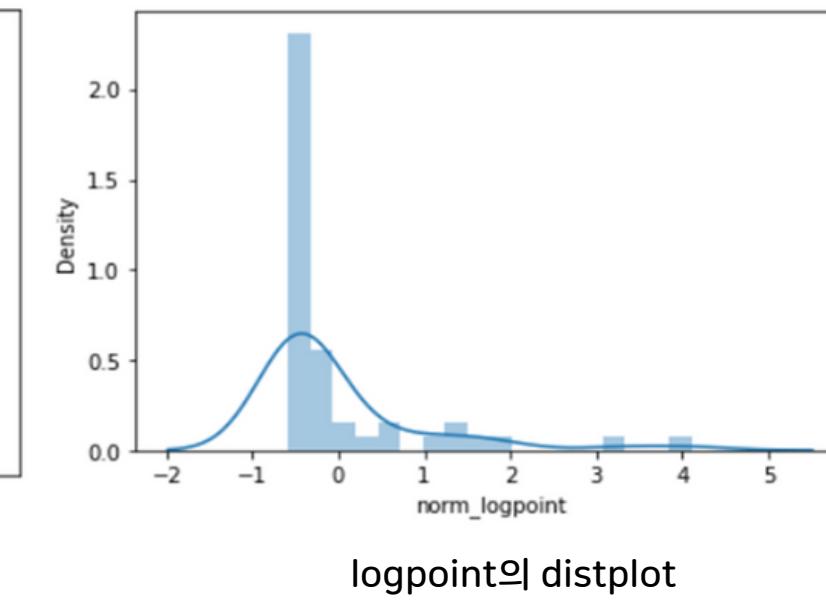
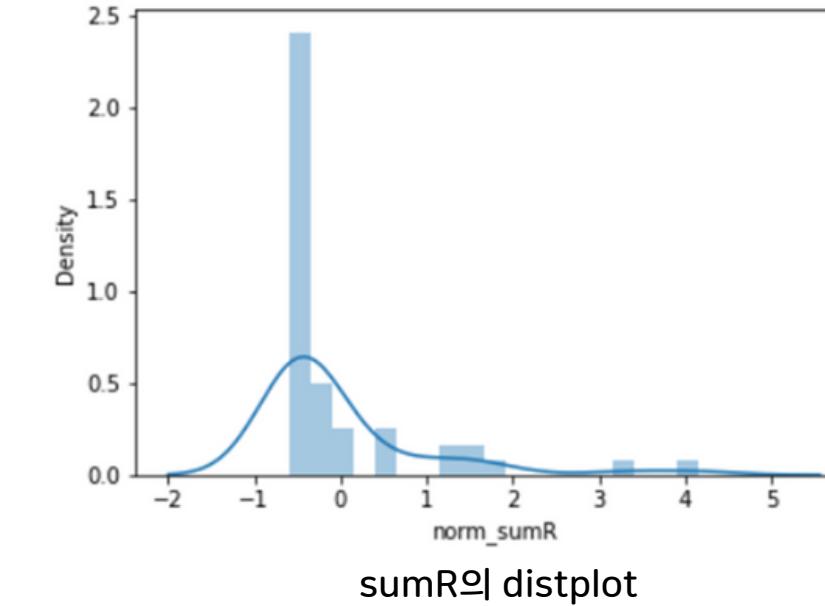
Naver Smart Store Review_5단계: 상품의 난이도별 리뷰의 질 분석

✓ 이상치를 제거하고 정규화 진행

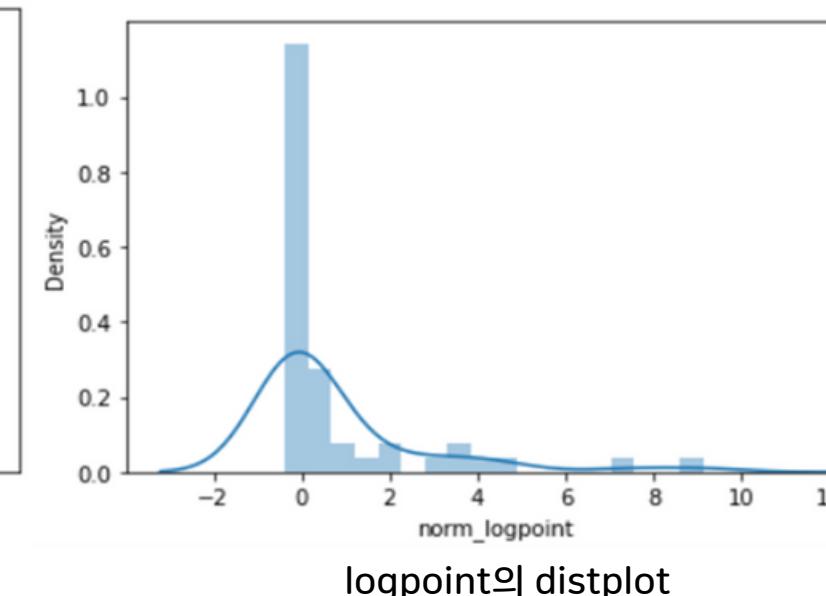
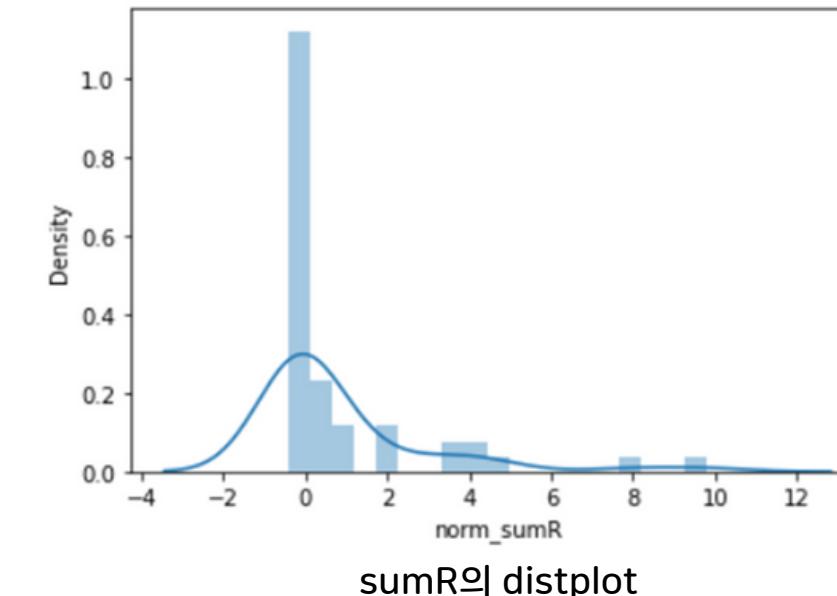


*MinMaxScaler는 이상치에 취약해서 제외함

● StandardScaler(표준 정규화)



● RobustScaler(로버스트 정규화)



Chapter 2. DD의 스토리_분석

Naver Smart Store Review 5다계·사프이나이드·리뷰이지·보서



**StandardScaler와 RobustScaler를 수행했을 때
큰 차이는 없지만 이상치에 강한
RobustScaler를 사용하기로 함**

* MinMaxScaler는 모든 데이터에 적용

5

10 12

Chapter 2. DD의 스토리_분석



Naver Smart Store Review_5단계: 상품의 난이도별 리뷰의 질 분석

상품id	상품명	리뷰개수	sumR*	logpoint*	난이도*
M1	코튼10 미니 스웨어 가방	507	347	841.775	4
M2	대바늘 마스터 세트	22	17	38.393	1

*난이도?

-왕초보: 난이도 1~3점

-초중급: 난이도 4~5점

*sumR? 카테고리에 분류된 형태소 개수를 상품별로 더한 값

*logpoint? TF-IDF 방법을 적용하여 상품별로 더한 값



RobustScaler

- 중앙값과 IQR을 사용하여 데이터 편향의 영향력을 줄임

상품id	상품명	리뷰개수	n_sumR	n_logpoint	난이도
M1	코튼10 미니 스웨어 가방	507	4.13	4.63	4
M2	대바늘 마스터 세트	22	-0.04	-0.10	1

Chapter 2. DD의 스토리_분석



Naver Smart Store Review_5단계: 상품의 나이도별 리뷰의 질 분석

- ✓ 귀무가설: 상품의 나이도에 따른 리뷰 질의 차이가 있지 않다.
- ✓ 대립가설: 상품의 나이도에 따른 리뷰 질의 차이가 있다.

● 정규화한 **sumR**을 이용한 두 그룹의 리뷰 질에 대한 t검정

	T	p-val	CI95%
T-test	-0.10	0.92	[-1.8, 1.64]

● 정규화한 **logpoint**을 이용한 두 그룹의 리뷰 질에 대한 t검정

	T	p-val	CI95%
T-test	-0.04	0.97	[-1.64, 1.57]

디자이너 패키지 상품의 나이도별(왕초보, 초중급) 리뷰의 질은 차이가 있다고 할 수 없다

Chapter 2. DD의 스토리_분석



Naver Smart Store Review_6단계: 의미 있는 리뷰와 의미 없는 리뷰의 비교 분석 결과

✓ 리뷰 워드클라우드

의미 있는 리뷰 워드클라우드



의미 없는 리뷰 워드클라우드



길이로 비교하기

59자

19자

유의미 리뷰 길이

무의미 리뷰 길이

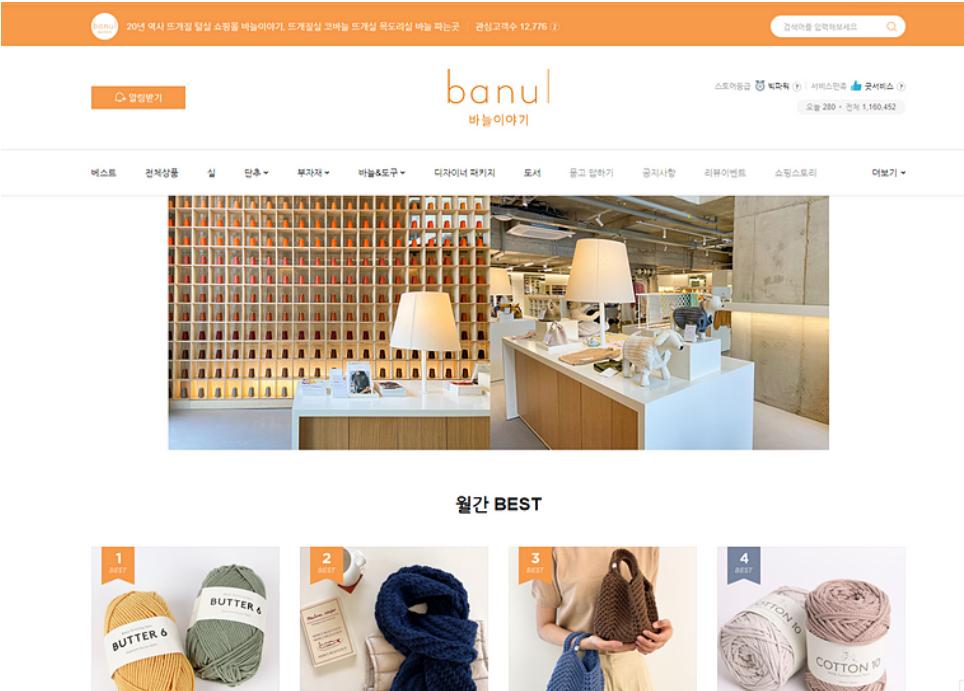
 내용으로 비교하기

의미가 있는 리뷰의 경우 가방처럼 구매 목적과 관련된 키워드가 많이 나타나거나, 색처럼 상품에 대한 구체적인 키워드가 많이 나타나는 반면, 의미가 없는 리뷰의 경우 '배송이 빠르다'처럼 비교적 흔한 내용의 키워드가 나타나는 것을 발견할 수 있었다.

Chapter 2. DD의 스토리_분석



Naver Smart Store Review_결과 및 활용 방안



● 스마트스토어 베스트 리뷰 뽑기

TF-IDF 합산 점수가 높은 리뷰가
네이버 스마트스토어에서 베스트 리뷰를 선정할 때
다른 리뷰와 겹치지 않는 정보가 많이 담긴 리뷰를
선정하는 데 도움을 줄 것이다.

이미지 출처: 바늘이야기 스마트스토어



● 상품에 대한 고객의 생각 파악

리뷰 테이블에는 별점이 같이 들어가 있으므로,
낮은 별점에 언급되는 키워드들을 통해 상품의
보완점을 찾아가는 데 도움을 줄 것이다.



ERP 데이터와 리뷰 데이터로
분석한 내용을 정리해보자.

Chapter 2. DD의 스토리_결론 및 제언

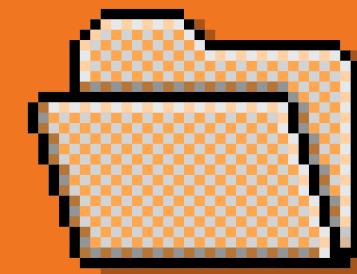


● ERP 요약

- '초보자', '왕초보', '초중급'으로 제작된 패키지 난이도와 소비자의 체감 난이도가 달라 이탈 고객 발생
→ 소비자 입장에서의 난이도 재분류가 필요
- 20-24는 유튜브 등 온라인 바이럴 마케팅을 통한 유입이 타 연령 대비 높음
→ 유튜브에 20대 초반이 흥미를 끌만한 쉽고 실용도 높은 제품 영상 업로드
→ 에어팟 케이스, 곱창 머리끈 등

● 리뷰 요약

- 리뷰의 길이 < 주요 단어의 빈도
- 현재 스마트 스토어의 디자이너 패키지 상품들의 난이도는 전부 초급으로만 구분
→ 난이도 별 고객들의 사용 후기 차이를 보기 어려움
→ 난이도를 세분화해서 구분하면 고객들의 사용 후기가 풍부해지고 이를 통해 얻을 수 있는 인사이트도 증가할 것



Chapter 3. DD가 아닌 Team DD가 생각해보았습니다

Chapter 3. DD가 아닌 Team DD가 생각해보았습니다



✓ 데이터 활용도가 낮은 Commerce 기업에 데이터 분석가로 입사했을 때 가장 먼저 확인해야 하는 작업은?

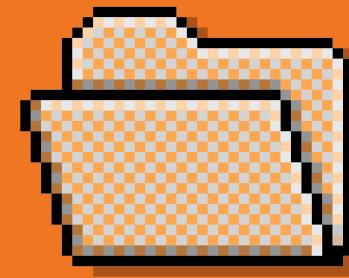
- 수집하고 있는 데이터와 수집할 수 있는 데이터 파악
- 수집하고 있는 데이터가 테이블로 잘 분리되어서 적재되고 있는지 확인
- 그렇지 않다면 주문, 상품, 고객 테이블로 나누어 각각의 테이블에 들어갈 수 있도록 스키마 정의, 엔지니어 팀에 요청

✓ 데이터 분석을 하기 전 고려해야하는 사항은?

- 사내 데이터 거버넌스에 대한 이해 필요
- 접근 가능한 데이터와 불가한 데이터를 파악하고, 접근 가능한 데이터의 분포 확인

✓ 데이터를 사용하고 싶은 Commerce 기업에게 제안할 수 있는 점은?

- 분석의 목적이 정확할수록 좋은 결과를 얻을 수 있어요!
평소 어떤 지점에 대한 고민이 많았는지 자주 생각한다면 더 좋은 분석을 할 수 있습니다.
- 데이터에 자주 접근하고 가져올 수 있는 환경이 마련되어야 합니다.



Chapter 4. Epilogue

Chapter 4. Epilogue_Team Members



김남이

Project Manager



인지원

Project Leader



이현정

Technical Leader



김지성

Marketer



김하나

Data Analyst



김혜원

Data Analyst



최지현

Data Analyst



DD

Data Analyst

Chapter 4. Epilogue

Tool

Python3.8

Excel

Google Sheet

Google BigQuery

Google LookerStudio

Google Cloud



Special Thanks to..

협업 기업 바늘이야기

협업 관계자 1 김보겸 대리님

협업 관계자 2 송학철 팀장님

피드백 유재명 강사님

데잇걸즈 6기 함께한 여러분 모두



Chapter 4. Epilogue



Thank you!

