



의류 사이즈 기준 및 상품 추천 시스템 모델 개발

여성 온라인 쇼핑몰 중심

유예지, 장유선

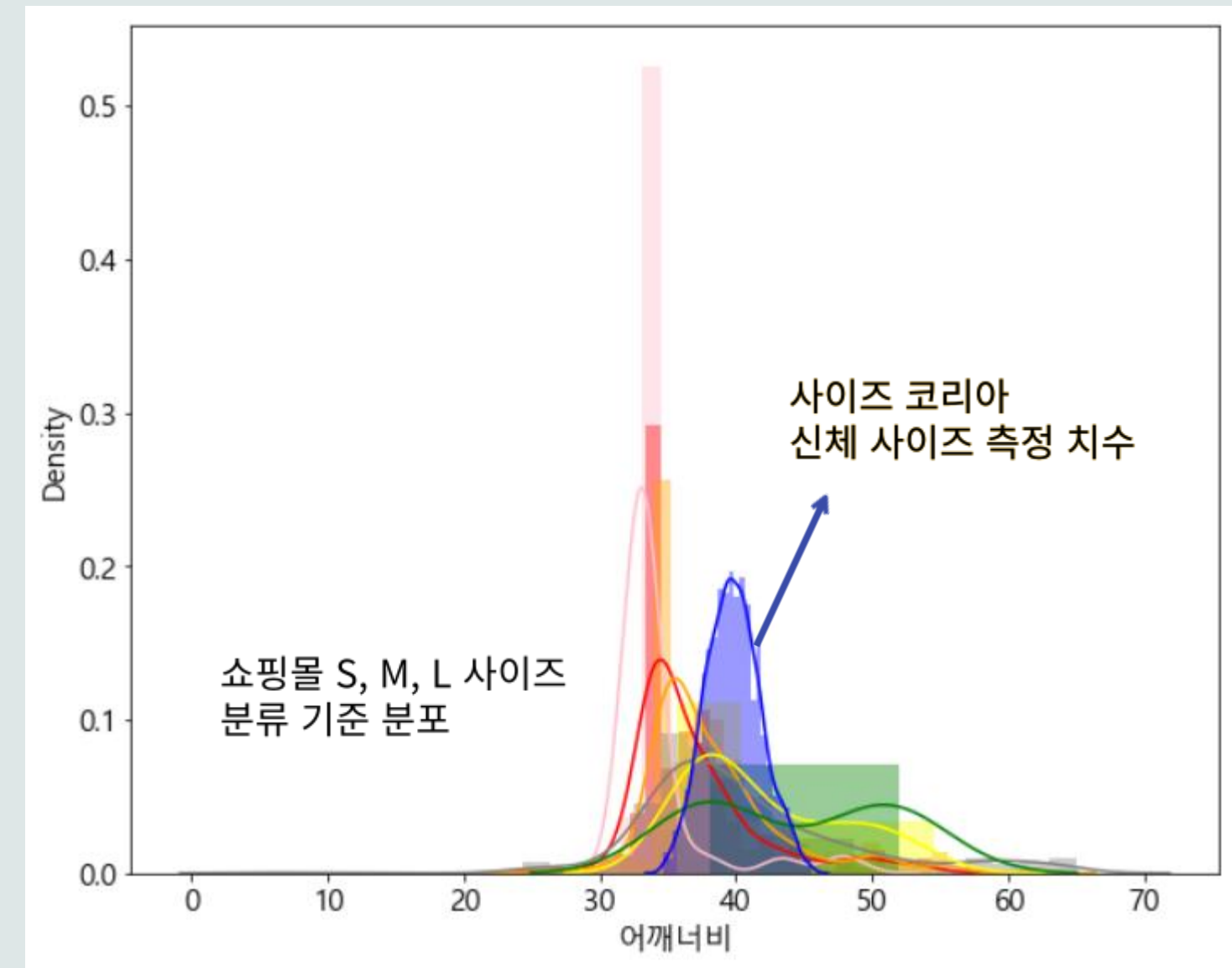
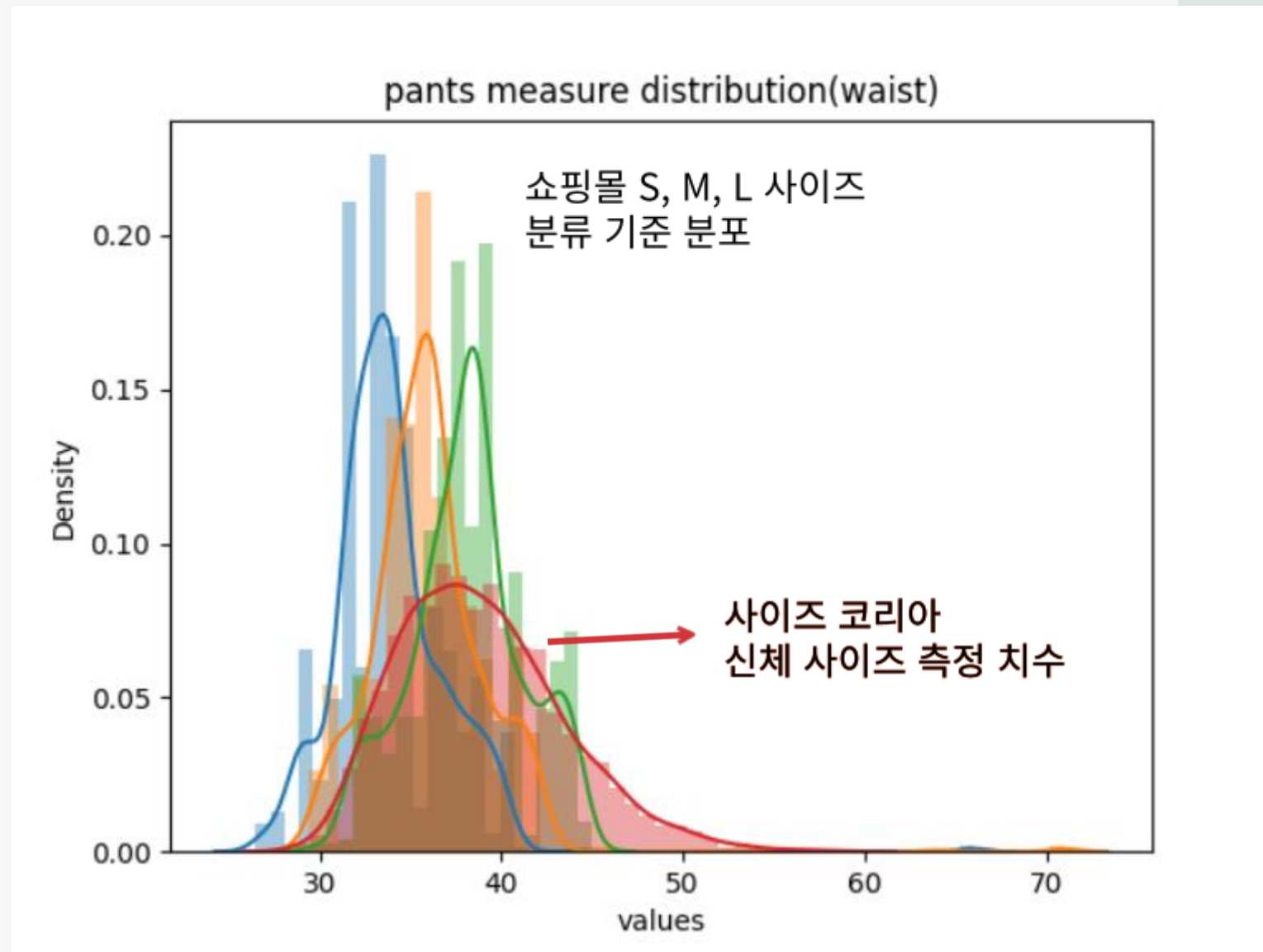
Contents

- 01. 프로젝트 배경**
- 02. 전체 프로세스 소개**
- 03. 활용 데이터**
- 04. 데이터 전처리 및 EDA**
- 05. 분석 기법 및 결과**
- 06. 활용방안 및 기대효과**

프로젝트 배경

온라인 쇼핑몰 사이즈 분류 현황

온라인 의류 쇼핑몰 '무신사'에서 여성 바지와 원피스의 사이즈 분류 기준과 실제 사이즈 코리아 신체 치수의 분산을 비교함



비교 결과, 소비자들의 치수를 제대로 반영하지 못하는 사이즈 분류 기준이 설정되어있음

프로젝트 배경

사이즈 분류 기준의 한계 및 문제점

1. 사이즈 혼동으로 인한 구매 피해 증가

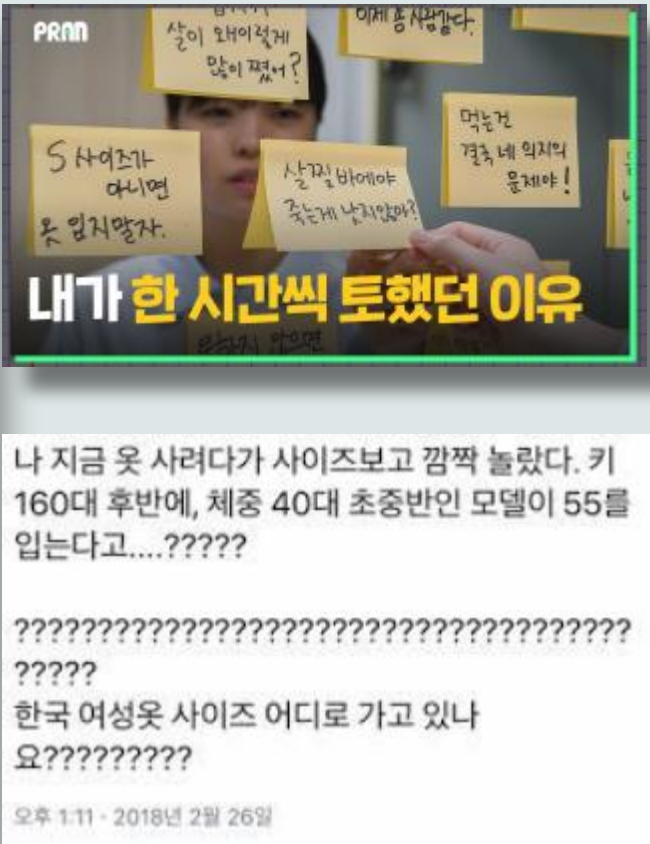
제각기인 인터넷 의류 사이즈, 소비자 불만 폭발

최소원 소비자기자 | 승인 2021.03.19 09:30 | 댓글 0



- ▶ 인터넷 의류 쇼핑몰들의 모호한 사이즈 분류로 인한 소비자들의 피해가 늘고있음

2. 작아지는 표준 사이즈로 인한 사회 압박 증가



- ▶ 왜곡된 신체 이미지에 대한 압박으로 이어져 식이장애를 겪는 사람들이 증가함

프로젝트 배경

사이즈 분류 기준의 한계 및 문제점

1. 사이즈 혼동으로 인한 구매 피해 증가

2. 작아지는 표준 사이즈로 인한 사회 압박 증가

현실 체형을 반영한 **사이즈 분류** 체계 제시 체형 기반 **추천 시스템** 개발

▶ 인터넷 의류 쇼핑몰들의 모호한 사이즈 분류로
인한 소비자들의 피해가 늘고있음

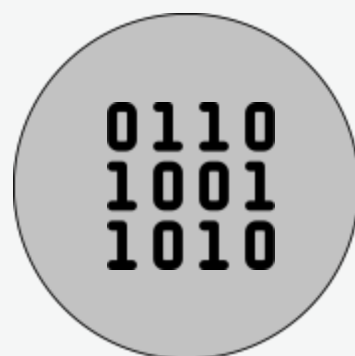
▶ 왜곡된 신체 이미지에 대한 압박으로 이어져
식이장애를 겪는 사람들이 증가함

전체 프로세스 소개



DATA

- 사이즈 코리아 인체 치수 데이터
2020(8차)
- 무신사 사이즈 테이블 데이터
웹 크롤링



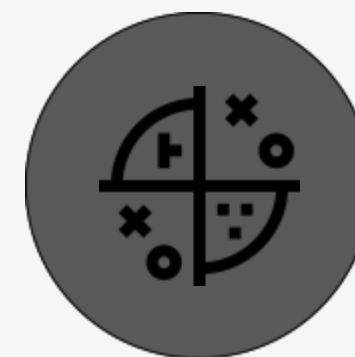
PREPROCESSING

- 데이터 축소
- 결측치 처리
- 이상치 제거
- 변수 생성



VISUALIZATION & EDA

- 실제 인체 사이즈 vs. 제품 사이즈
(분산 그래프 비교)
- 한국인들의 체형 변화



MODELING

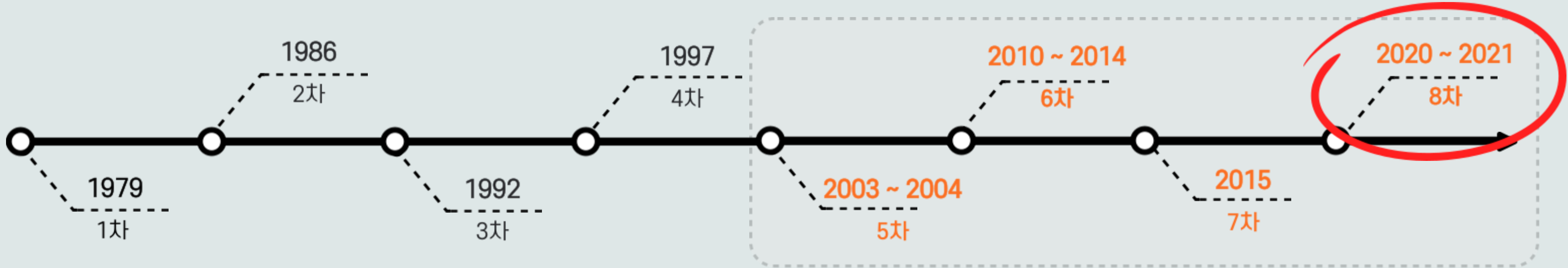
- Clustering
- K-mean
- LightGBM

활용 데이터

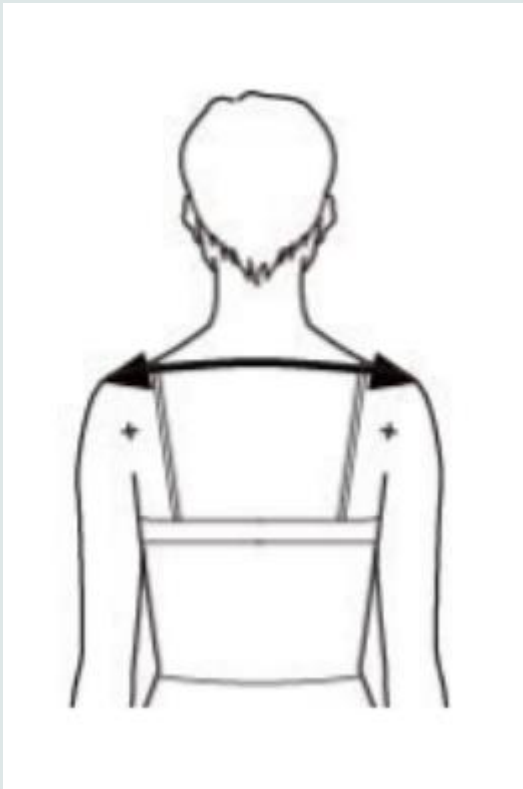
사이즈 코리아 인체 치수 조사



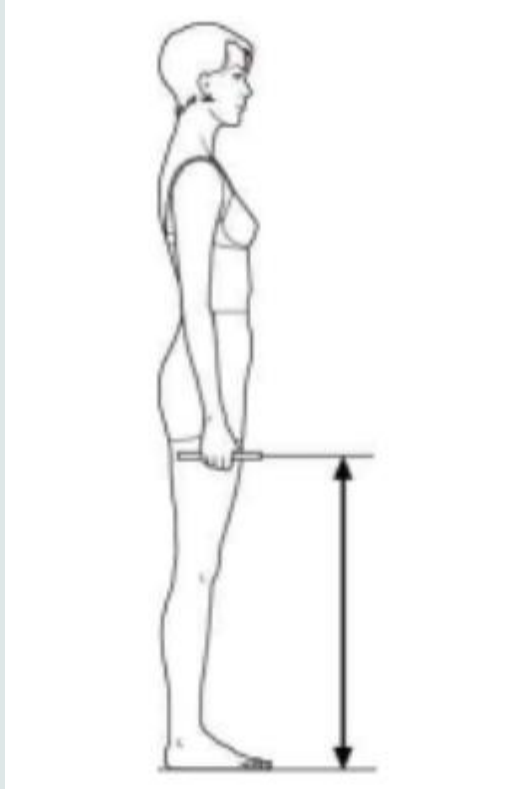
사이즈 코리아 인체 치수 조사 데이터(2003 ~ 2021)



측정 예시



어깨사이길이



주먹높이



측정 인원

2003 ~ 2014

평균 7 ~ 80세 남녀 14,000명



측정 방법

2003 ~ 2014

마틴(Martin)식 계측자 및
3차원 인체 스캐너



측정 지역

2003 ~ 2014

전국 28개
시군구

2015 ~ 2021

평균 20 ~ 69세 남녀 5,500명

2015 ~ 2021

마틴(Martin)식 계측자, 체성분 분석기 및
전신스캐너, 발스캐너, 손스캐너
3종의 3차원 인체형상측정용 스캐너

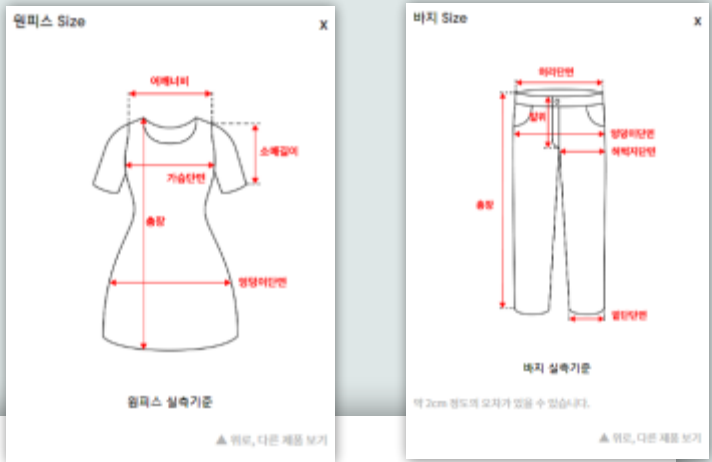
2015 ~ 2021

전국 28개
시군구

활용 데이터

무신사 웹 크롤링 데이터

MUSINSA



무신사 사이즈 정보 웹 크롤링 데이터

- ▶ 여성 바지, 원피스 전체
- ▶ 브랜드명, 상품명, 사이즈분류, 측정부위, 측정치

데이터 선정 이유

- ▶ 여러 온라인 쇼핑몰들의 제품이 모여있어 다양한 정보를 얻기 쉬움
- ▶ 다양한 의류 브랜드 제품의 사이즈를 일관되게 가져오기 용이함
- ▶ 측정법에 대한 정보가 주어져 사이즈 코리아 데이터와 비교하기 좋음

Size Info 사이즈정보

구매내역 실측 선택 실측 직접 입력 **측정법**

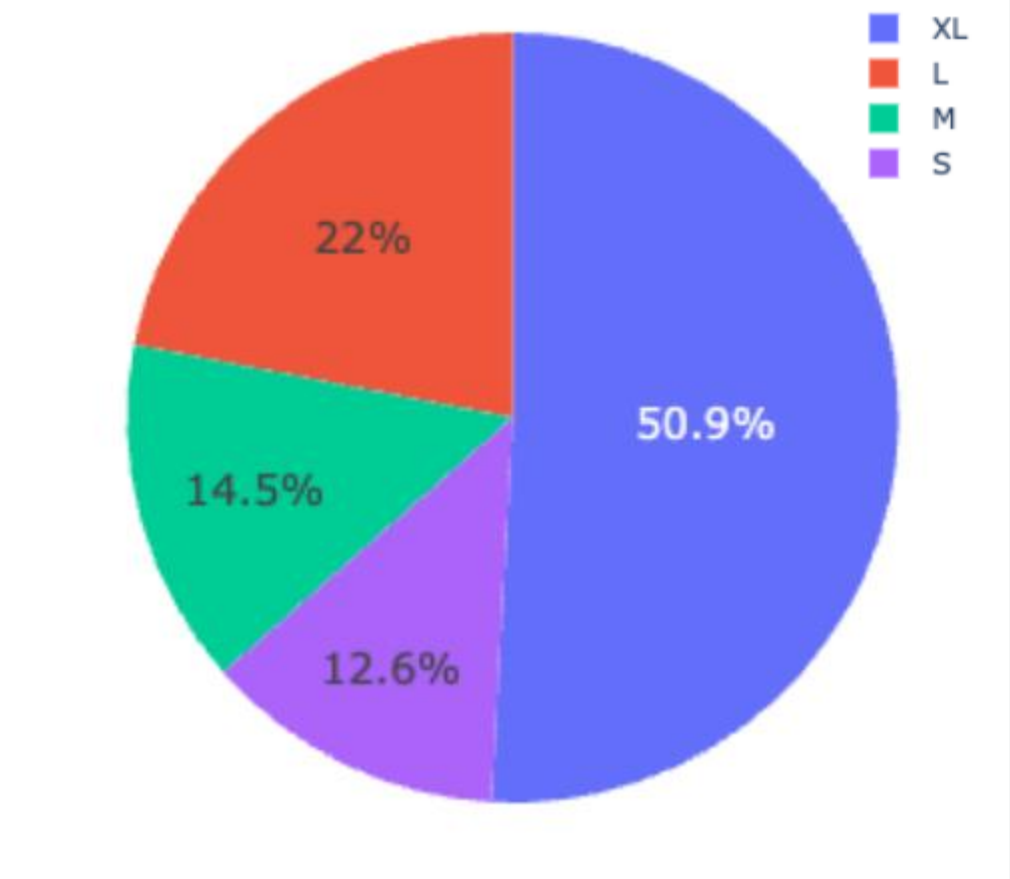
cm	총장	어깨너비	가슴단면	소매길이
MY	가지고 계신 제품의 실측을 입력해 보세요~!			
FREE	82	42	50	62

- 위 사이즈 표는 무신사스토어 측정방식을 기준으로 작성되었습니다.
- 보유하신 옷의 실측값을 입력하시면 정확한 비교가 가능합니다.
- 상품 상세 정보의 사이즈 표는 업체 기준 측정입니다.

데이터 전처리 및 EDA

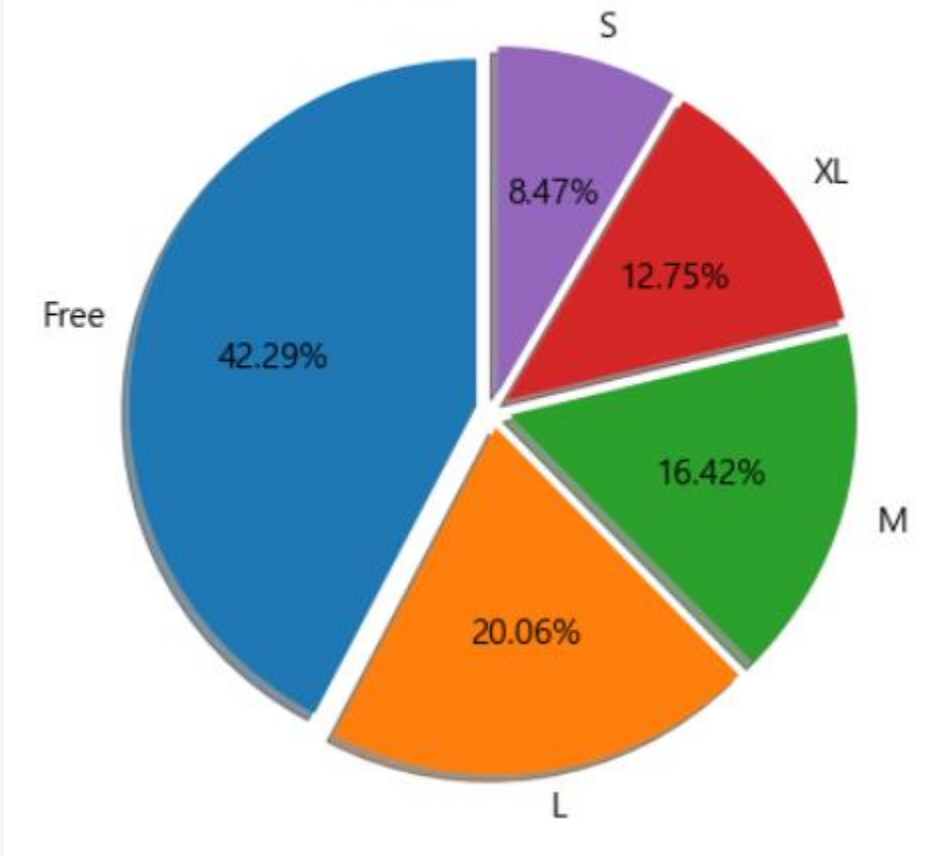
EDA

사이즈 코리아 피실험자 바지 사이즈 비율 차트



- 사이즈 코리아 피실험자 대다수가 의류 사이즈 XL에 해당
- 평균 사이즈인 M에 해당하는 사람은 14.5%로 적은 비율을 차지
- 최근 생산되고있는 바지 상품들의 사이즈가 전체적으로 작다는 것을 의미

사이즈 코리아 피실험자 원피스 사이즈 비율 차트



- 사이즈 코리아 피실험자 대다수가 의류 사이즈 Free(M과 L 사이)에 해당
- 평균 사이즈인 M에 해당하는 사람은 16.42%로 적은 비율을 차지
- 최근 생산되고있는 원피스의 사이즈가 전체적으로 작다는 것을 의미 (XS에 해당하는 경우는 아예 없었음)

데이터 전처리 및 EDA

데이터 축소

사이즈 코리아
8차 인체치수조사(2020_21) 치수 데이터

구분(전신/머리/손/발)										전신	전신
HUMAN_ID	성별	조사년도	조사일	나이	측정지	측정복질가슴둘레	측정복상의	측정복바깥수준허리둘레	측정복하의	001. 머리위로뺀은주먹높이	002. 키
20_F_1465	여	2020	1116	42	서울/경기/강원	86	85XL	88	90F	1963	1685
20_F_1467	여	2020	1116	42	서울/경기/강원	73.6	75M	74.5	75F	1815	1602
20_F_1477	여	2020	1116	40	서울/경기/강원	74.3	75M	77	75F	1901	1642
20_F_1486	여	2020	1117	43	서울/경기/강원	77.5	80L	78	80F	1821	1589
20_F_1490	여	2020	1117	40	서울/경기/강원	80	80L	80	80F	1766	1531
20_F_1492	여	2020	1117	40	서울/경기/강원	75	80L	80.8	80F	1808	1538
20_F_1495	여	2020	1117	40	서울/경기/강원	72	75M	72	75F	1880	1632
20_F_1501	여	2020	1117	40	서울/경기/강원	79	80L	79.5	80F	1905	1651
20_F_1506	여	2020	1118	43	서울/경기/강원	74	75M	73.6	75F	1963	1745
20_F_1512	여	2020	1118	41	서울/경기/강원	81.5	80L	82	80F	1721	1532
20_F_1531	여	2020	1119	44	서울/경기/강원	80	80L	89	85F	1838	1584
20_F_1549	여	2020	1119	42	서울/경기/강원	69	75M	66.1	65F	1836	1581

●
●
●
x 2774

● ● ● x 147
➔

	살앞뒤길이	엉덩이둘레	다리가쪽길이	허리둘레	넙다리둘레
3	770	998	1070	885	591.0
4	685	922	985	718	500.0
5	695	894	1042	711	530.0
6	639	928	982	737	569.0
7	642	909	935	816	592.0
...
5069	673	922	983	719	579.0
5070	588	867	953	679	494.0
5071	690	1059	1004	829	644.0
5082	680	988	1031	754	630.0
5083	670	952	935	859	572.0
2773 rows × 5 columns					

2773 x 5

바지 및 원피스 카테고리 분석을 위한 데이터 축소

- 147개의 인체 측정 항목에서 필요한 컬럼만 추출
- 조사년도 : 2020, 2021년도
- 성별 : 여

바지 추출 항목

- 허리둘레
- 살앞뒤길이
- 엉덩이둘레
- 다리가쪽길이
- 넙다리둘레

원피스 추출 항목

- 어깨사이길이
- 가슴둘레
- 엉덩이둘레

데이터 전처리 및 EDA

바지 | 데이터 정리(이상치 제거)

사이즈 코리아

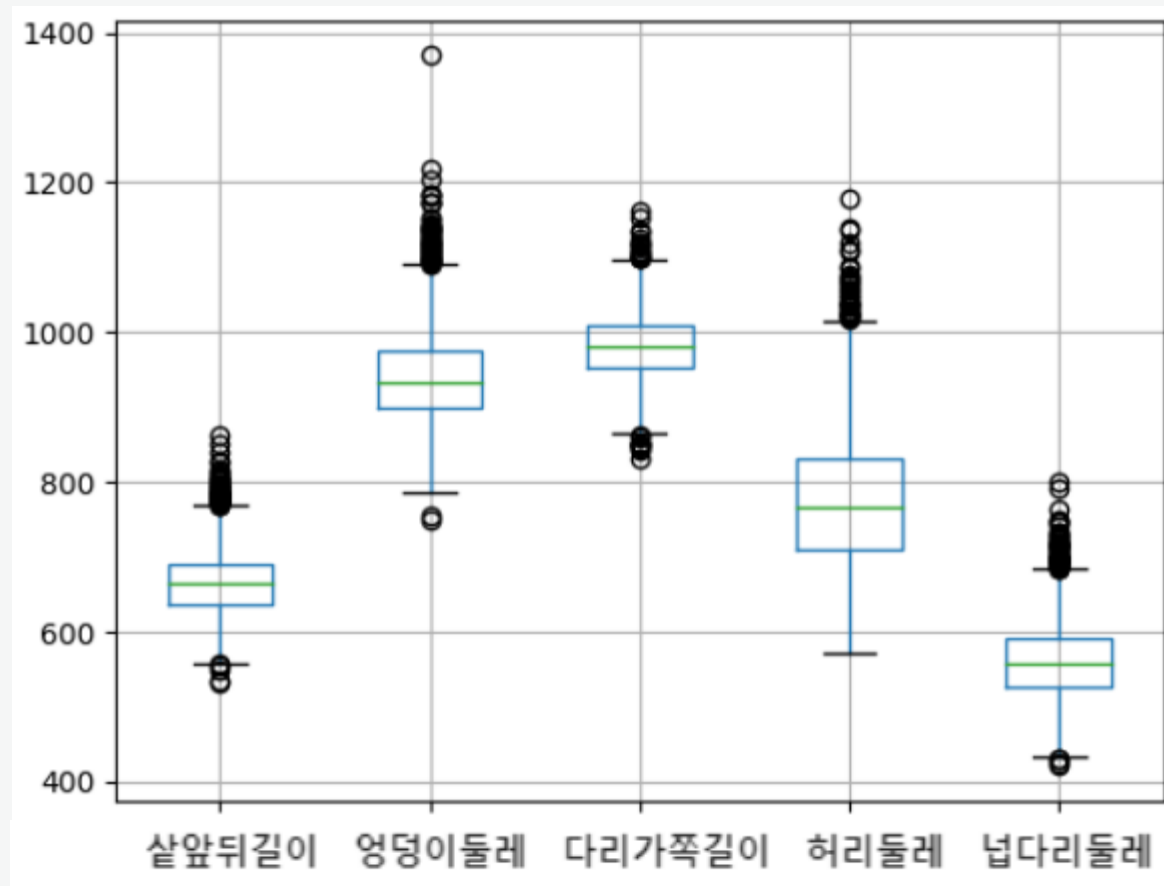
각 관측치들의 이상치 유무를
확인하기 위한 box plot

	살앞뒤길이	엉덩이둘레	다리가쪽길이	허리둘레	넙다리둘레
3	770	998	1070	885	591.0
4	685	922	985	718	500.0
5	695	894	1042	711	530.0
6	639	928	982	737	569.0
7	642	909	935	816	592.0
...
5069	673	922	983	719	579.0
5070	588	867	953	679	494.0
5071	690	1059	1004	829	644.0
5082	680	988	1031	754	630.0
5083	670	952	935	859	572.0

2773 rows × 5 columns

사이즈 코리아 (2773 x 5)

총 8차의 조사 데이터 중 가장 최근 인체 치수를
반영하고 있는 2020년 데이터를 이용



이상치 제거

	살앞뒤길이	엉덩이둘레	다리가쪽길이	허리둘레	넙다리둘레
4	685	922	985	718	500.0
5	695	894	1042	711	530.0
6	639	928	982	737	569.0
7	642	909	935	816	592.0
8	638	913	940	770	520.0
...
5069	673	922	983	719	579.0
5070	588	867	953	679	494.0
5071	690	1059	1004	829	644.0
5082	680	988	1031	754	630.0
5083	670	952	935	859	572.0

2629 rows × 5 columns

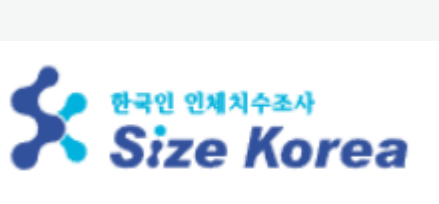
original_df : 2773
IQR_df : 2629

최종 데이터 (2629 x 5)

데이터 전처리 및 EDA

원피스 | 데이터 정리(이상치 제거)

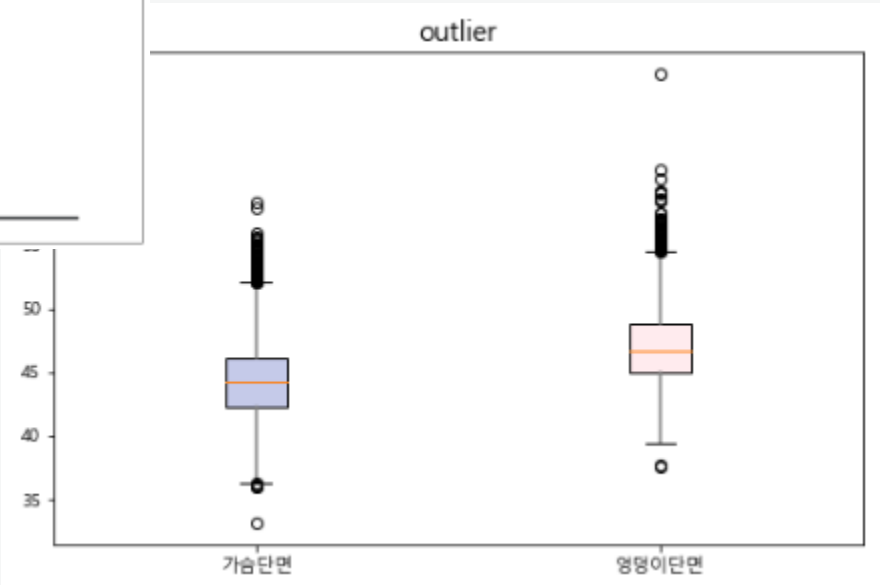
사이즈 코리아



어깨사이길이, 가슴둘레,
엉덩이둘레 칼럼만 추출



이상치 제거



Unnamed: 0	HUMAN_ID	성 별	조사 년도	조사 일	나이	측정자	측정복\n깃가슴둘레	측정복\n상의의	측정복\n허리둘레	128. 배꼽수 준앞둔심길이	129. 배꼽수 준옆길이	130. 복위젓 복지길이	131. 복위젓복지 허리둘레상단길이	132. 배꼽수 준앞둔위길이	133. 앞 온눈높이	134. 앞온 복위높이	135. 앞온 어깨높이	136. 앞온팔꿈 지높이(팔굽편)	137. 위팔수직 길이(팔굽편)		
3	4	20_F_1465	여	2020	1116	42	서울/경 기/강원	86.0	85XL	88.0	...	449	453.0	356.5	546.5	882	796	659	626.0	294.0	332.0
4	5	20_F_1467	여	2020	1116	42	서울/경 기/강원	73.6	75M	74.5	...	360	391.0	331.5	477.5	731	738	592	544.0	234.0	310.0
5	6	20_F_1477	여	2020	1116	40	서울/경 기/강원	74.3	75M	77.0	...	372	405.0	326.5	488.5	749	738	601	552.0	203.0	349.0
6	7	20_F_1486	여	2020	1117	43	서울/경 기/강원	77.5	80L	78.0	...	434	428.0	331.5	517.5	741	753	611	581.0	268.0	313.0
7	8	20_F_1490	여	2020	1117	40	서울/경 기/강원	80.0	80L	80.0	...	405	415.0	333.5	527.5	696	716	578	544.0	245.0	299.0
...
5069	5070	21_F_7275	여	2021	924	27	경상권	82.0	85.0	81.0	...	390	448.0	319.5	483.5	753	767	636	575.0	269.0	306.0
5070	5071	21_F_7276	여	2021	924	34	경상권	79.0	80.0	77.0	...	349	396.0	328.0	460.0	640	723	579	529.0	248.0	281.0
5071	5072	21_F_7277	여	2021	924	32	경상권	97.0	95.0	87.0	...	428	440.0	367.5	547.5	760	774	634	589.0	259.0	330.0
5082	5083	21_F_7288	여	2021	924	31	경상권	91.0	80.0	81.0	...	410	450.0	366.5	506.5	784	729	620	568.0	244.0	324.0
5083	5084	21_F_7289	여	2021	924	33	경상권	95.0	100.0	88.0	...	439	434.0	353.5	506.5	802	716	598	539.0	221.0	318.0

2773 rows x 148 columns

사이즈 코리아 (2773 x 147)

총 8차의 조사 데이터 중 가장 최근 인체 치수를
반영하고 있는 2020년 데이터를 이용

최종 데이터 (2667 x 3)

	어깨너비	가슴단면	엉덩이단면
0	38.6	49.65	49.90
1	37.5	40.90	46.10
2	39.5	41.95	44.70
3	36.6	43.35	46.40
4	37.7	45.10	45.45
...
2662	35.8	41.60	46.10
2663	40.0	39.45	43.35
2664	38.0	47.10	52.95
2665	40.5	43.40	49.40
2666	39.4	44.50	47.60

2667 rows x 3 columns

데이터 전처리 및 EDA

MUSINSA



바지 | 데이터 정리

무신사 웹 크롤링

0	103	26	총장	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
1	35	26	허리단면	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
2	48	26	엉덩이단면	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
3	30.8	26	허벅지단면	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
4	27	26	밑위	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
...
13	40	L	허리단면	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙
14	55	L	엉덩이단면	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙
15	37.5	L	허벅지단면	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙
16	35	L	밑위	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙
17	28.5	L	밑단단면	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙

[25345 rows x 5 columns]

무신사 사이트 내 바지 상품 1 ~ 20page 웹크롤링

최근 3개월 내 인기순으로 상품 정렬,
피처 생성을 위해 바지 치수 측정 부위
6개의 모두에 대한 측정치가 있는 상품들만 추출

measurement	
밑단단면	4224
밑위	4224
엉덩이단면	4224
총장	4224
허리단면	4224
허벅지단면	4224

인체 치수와 직접적인 연관이 있는
'밑위', '엉덩이단면', '총장',
'허리단면', '허벅지단면'만 추출

groupby로 각 항목의 개수 확인 후
결측치의 유무를 확인

	values	size_category	measurement	brand_name	product_name
0	103.0	26	총장	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
1	35.0	26	허리단면	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
2	48.0	26	엉덩이단면	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
3	30.8	26	허벅지단면	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
4	27.0	26	밑위	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]
...
25339	106.0	L	총장	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙
25340	40.0	L	허리단면	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙
25341	55.0	L	엉덩이단면	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙
25342	37.5	L	허벅지단면	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙
25343	35.0	L	밑위	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙

21120 rows x 5 columns

최종 데이터 프레임

데이터 전처리 및 EDA

MUSINSA



원피스 | 데이터 정리

무신사 웹 크롤링

	values	size	measures	brand_name	product_name
0	109	xs	총장	LIENLIER	로즈퍼프원피스 - 카키
1	33	xs	어깨너비	LIENLIER	로즈퍼프원피스 - 카키
2	43.5	xs	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
3	57	xs	소매길이	LIENLIER	로즈퍼프원피스 - 카키
4	45	xs	엉덩이단면	LIENLIER	로즈퍼프원피스 - 카키
...
10	114	m	총장	LIENLIER	로즈퍼프원피스 - 카키
11	35	m	어깨너비	LIENLIER	로즈퍼프원피스 - 카키
12	48.5	m	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
13	60	m	소매길이	LIENLIER	로즈퍼프원피스 - 카키
14	49	m	엉덩이단면	LIENLIER	로즈퍼프원피스 - 카키

9035 rows × 5 columns

무신사 사이트 내 원피스 상품 1~100page 웹크롤링

최근 3개월 내 인기순으로 상품 정렬,
모든 상품들이 최소한으로 갖고 있는
사이즈 치수 값(5개) 를 기준으로 크롤링

```
dress_brands['measures'].value_counts()
```

```
총장      1807
엉덩이단면 1807
어깨너비   1806
가슴단면   1806
소매길이   1806
허리단면     1
밑위         1
밑단단면     1
Name: measures, dtype: int64
```

"어깨너비" "가슴단면" "엉덩이 단면" 3개로 사이즈 치수 통일

사이즈 코리아 데이터의
사이즈 항목과 동일하도록
이외의 치수가 있는 행은 제외

	values	size	measures	brand_name	product_name
0	33.0	xs	어깨너비	LIENLIER	로즈퍼프원피스 - 카키
1	43.5	xs	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
2	45.0	xs	엉덩이단면	LIENLIER	로즈퍼프원피스 - 카키
3	34.0	s	어깨너비	LIENLIER	로즈퍼프원피스 - 카키
4	46.0	s	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
...
5413	46.0	s	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
5414	47.0	s	엉덩이단면	LIENLIER	로즈퍼프원피스 - 카키
5415	35.0	m	어깨너비	LIENLIER	로즈퍼프원피스 - 카키
5416	48.5	m	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
5417	49.0	m	엉덩이단면	LIENLIER	로즈퍼프원피스 - 카키

5418 rows × 5 columns

최종 데이터 프레임



데이터 전처리 및 EDA

최종 데이터

" 무신사 웹 크롤링 데이터 "

	brand_name	product_name	size_category	measurement	values
0	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]	26	총장	103.0
1	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]	26	허리단면	35.0
2	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]	26	엉덩이단면	48.0
3	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]	26	허벅지단면	30.8
4	MUSINSA STANDARD	와이드 히든 밴딩 슬랙스 [블랙]	26	밑위	27.0
...
25340	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙	L	허리단면	40.0
25341	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙	L	엉덩이단면	55.0
25342	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙	L	허벅지단면	37.5
25343	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙	L	밑위	35.0
25344	CPGN STUDIO	레이어 와이드 데님 팬츠 워시드블랙	L	밑단단면	28.5

25345 rows × 5 columns

무신사 데이터 컬럼명 정리	
brand_name	상품 브랜드명
product_name	상품명
size_category	사이즈 분류명
measurement	측정 부위
values	상품 치수

- ▶ measurement 항목이 상품마다 다름
- ▶ 피처 생성위해 같은 항목만 있는 상품들로 정리

" 사이즈 코리아 인체 치수 데이터 "

	살앞뒤길이/2	엉덩이둘레/2	다리가쪽길이	허리둘레/2	넙다리둘레/2
0	34.25	46.10	98.5	35.90	25.00
1	34.75	44.70	104.2	35.55	26.50
2	31.95	46.40	98.2	36.85	28.45
3	32.10	45.45	93.5	40.80	29.60
4	31.90	45.65	94.0	38.50	26.00
...
2623	33.65	46.10	98.3	35.95	28.95
2624	29.40	43.35	95.3	33.95	24.70
2625	34.50	52.95	100.4	41.45	32.20
2626	34.00	49.40	103.1	37.70	31.50
2627	33.50	47.60	93.5	42.95	28.60

2628 rows × 5 columns

사이즈 코리아 데이터 컬럼명 정리	
살앞뒤길이/2	072. 살앞뒤길이 /2
엉덩이둘레/2	048. 엉덩이둘레 / 2
다리가쪽길이	083. 다리가쪽길이
허리둘레/2	045. 허리둘레 / 2
넙다리둘레/2	051. 넙다리둘레 /2

- ▶ 무신사 데이터와 같은 측정 부위와 단위로 데이터 정리

데이터 전처리 및 EDA

최종 데이터

" 무신사 웹 크롤링 데이터 "

	values	size	measures	brand_name	product_name
0	33.0	xs	어깨너비	LIENLIER	로즈퍼프원피스 - 카키
1	43.5	xs	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
2	45.0	xs	엉덩이단면	LIENLIER	로즈퍼프원피스 - 카키
3	34.0	s	어깨너비	LIENLIER	로즈퍼프원피스 - 카키
4	46.0	s	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
...
5413	46.0	s	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
5414	47.0	s	엉덩이단면	LIENLIER	로즈퍼프원피스 - 카키
5415	35.0	m	어깨너비	LIENLIER	로즈퍼프원피스 - 카키
5416	48.5	m	가슴단면	LIENLIER	로즈퍼프원피스 - 카키
5417	49.0	m	엉덩이단면	LIENLIER	로즈퍼프원피스 - 카키

5418 rows × 5 columns

무신사 데이터 컬럼명 정리	
brand_name	상품 브랜드명
product_name	상품명
size	사이즈 분류명
measures	측정 부위
values	상품 치수

- ▶ measurement 항목이 상품마다 다름
- ▶ 피처 생성위해 같은 항목만 있는 상품들로 정리



" 사이즈 코리아 인체 치수 데이터 "

	어깨너비	가슴단면	엉덩이단면
0	38.6	49.65	49.90
1	37.5	40.90	46.10
2	39.5	41.95	44.70
3	36.6	43.35	46.40
4	37.7	45.10	45.45
...
2662	35.8	41.60	46.10
2663	40.0	39.45	43.35
2664	38.0	47.10	52.95
2665	40.5	43.40	49.40
2666	39.4	44.50	47.60

2667 rows × 3 columns

사이즈 코리아 데이터 컬럼명 정리	
어깨 너비	059. 어깨사이길이
가슴 단면	041. 가슴둘레 / 2
엉덩이 단면	048. 엉덩이둘레 / 2

- ▶ 무신사 데이터와 같은 측정 부위와 단위로 데이터 정리

분석 기법 및 결과

필요한 분석 기법 선정

군집화

새로운 사이즈 기준 생성



새로운 사이즈 기반
무신사 사이즈 데이터 분류



분류한 사이즈
무신사 상품 정보와 결합



피실험자의 사이즈와 무신사 상
사이즈가 같은 상품들을 오차가 적은 순으로
정렬해 상품 및 브랜드 추천

분석 기법 및 결과

군집화 Clustering

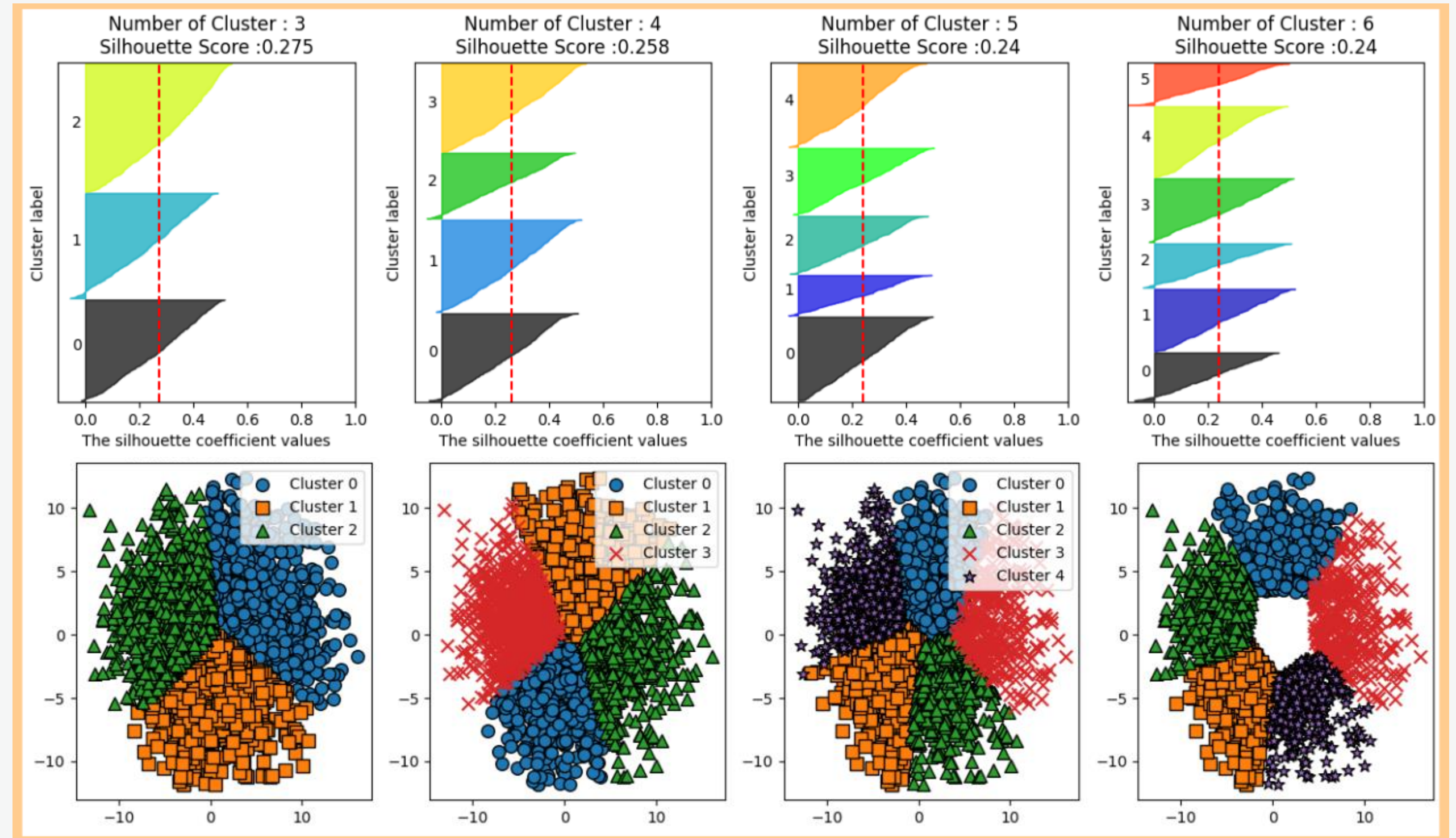
: 현실 체형을 반영하는

사이즈 분류 기준을 만들기 위해

사이즈 코리아 피실험자들의

신체 사이즈를 기준으로 군집화 진행

바지 |
사이즈 코리아 데이터
군집화 개수 설정



Silhouette Score 및 Cluster 그래프 분석 후 군집화 개수 5개로 결정

분석 기법 및 결과

군집화 Clustering

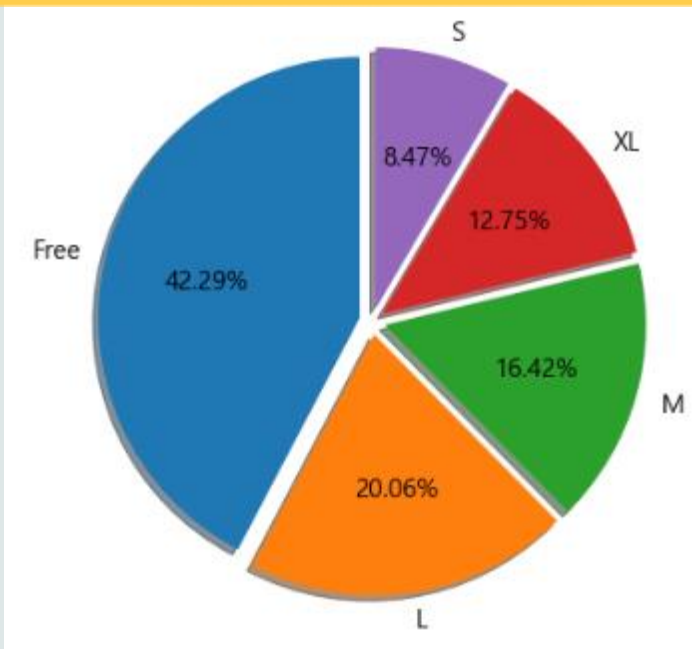
: 현실 체형을 반영하는

사이즈 분류 기준을 만들기 위해

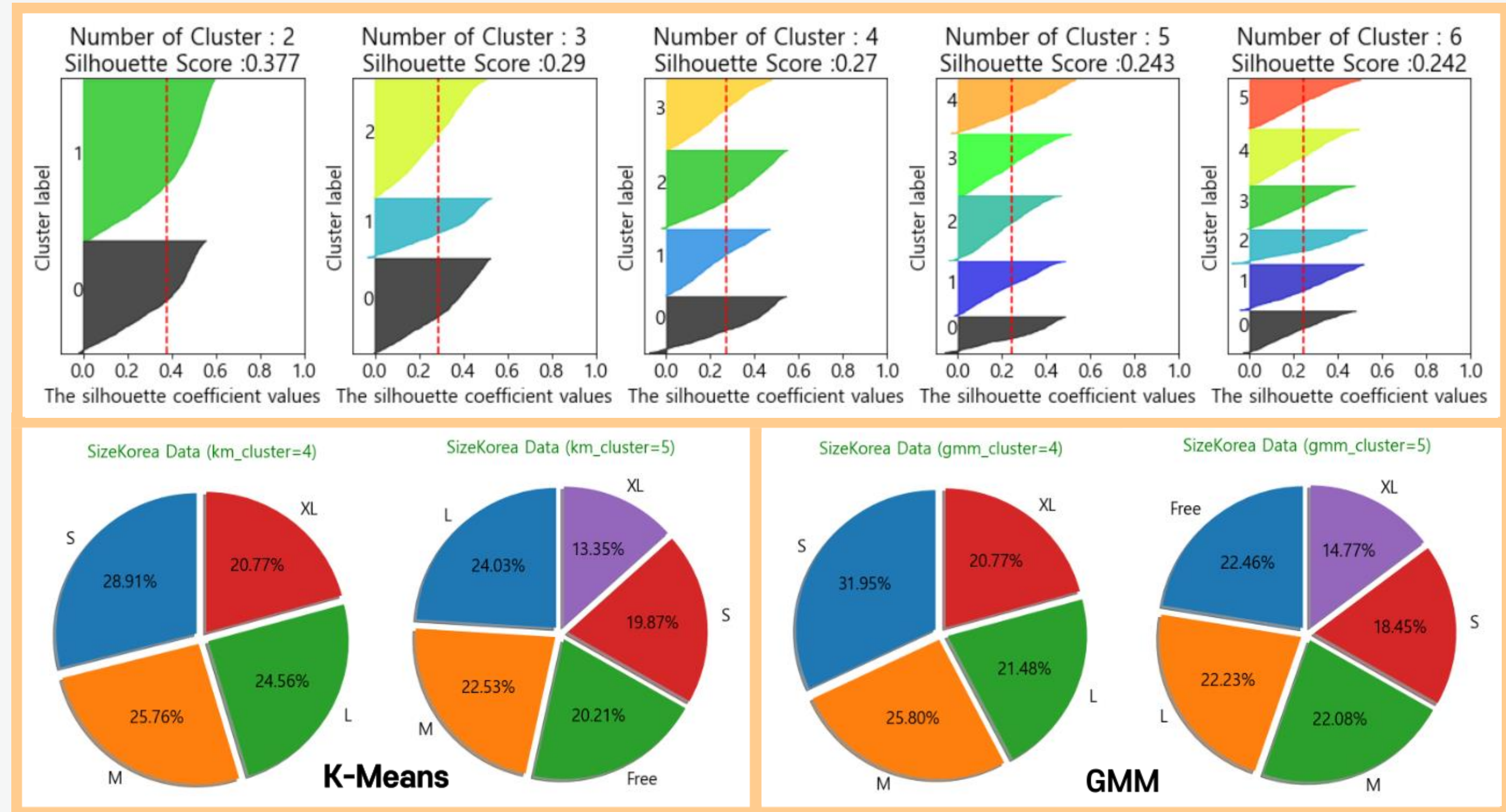
사이즈 코리아 피실험자들의

신체 사이즈를 기준으로 군집화 진행

사이즈 코리아 피실험자 원피스 사이즈 비율 차트



원피스 | 사이즈 코리아 데이터 군집화 개수 설정



Silhouette Score 및 Cluster 분석 후 군집화 개수 5개로 결정

분석 기법 및 결과

군집화 Clustering

: 군집화 모델들 비교 후 최종 결정

바지 | 사이즈 코리아 데이터 군집화 모델 결정

K-Means

cluster	
0	525
1	666
2	662
3	317
4	459

Mean Shift

MeanShift_Cluster	
0	2629

GMM

gmm_cluster	
0	509
1	801
2	419
3	604
4	296

DBSCAN

dbscan_cluster	
-1	2629

: 군집화 모델 비교 후 K-Means 으로 결정

원피스 | 사이즈 코리아 데이터 군집화 모델 결정

K-Means

cluster5_labels	
0	356
1	539
2	641
3	601
4	530

Mean Shift

meanshift_label	
0	2667

GMM

gmm_cluster5_labels	
0	599
1	492
2	394
3	593
4	589

DBSCAN

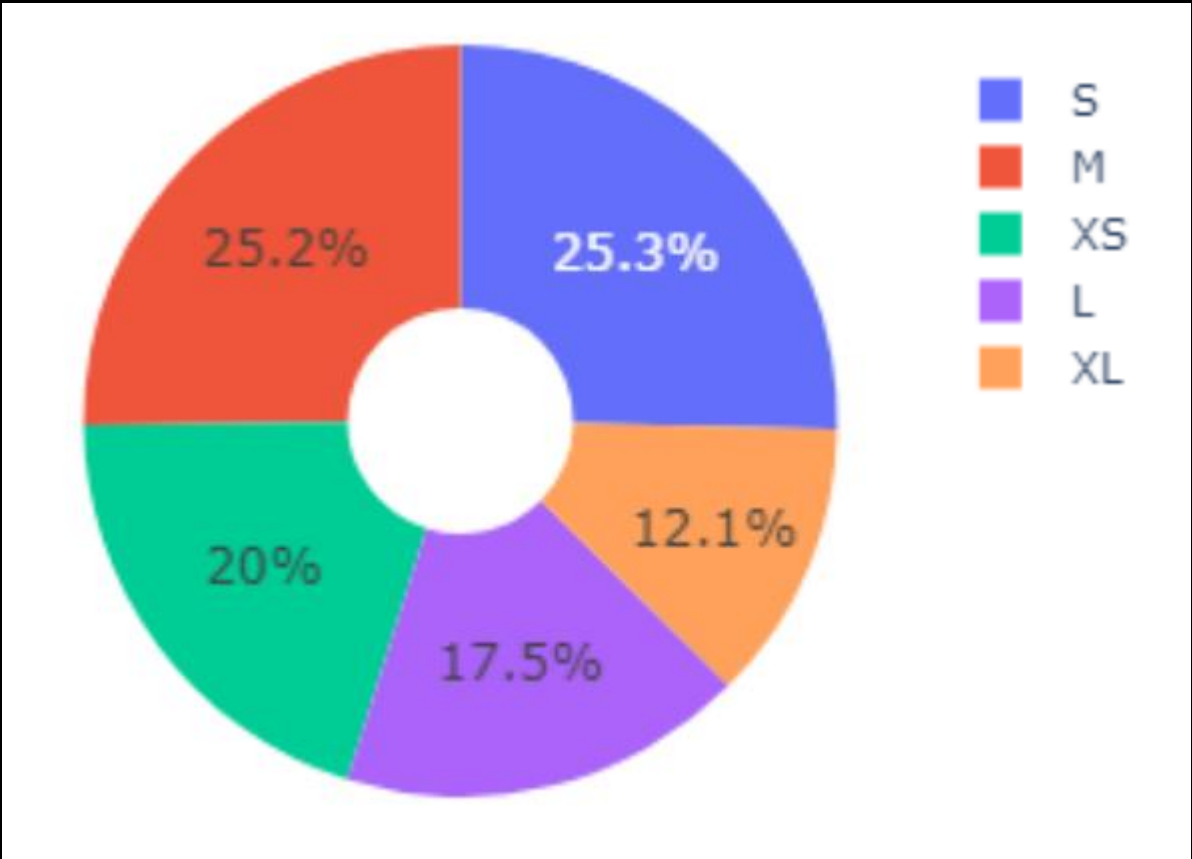
dbscan_cluster	
-1	2772

: 군집화 모델 비교 후 GMM 으로 결정

분석 기법 및 결과

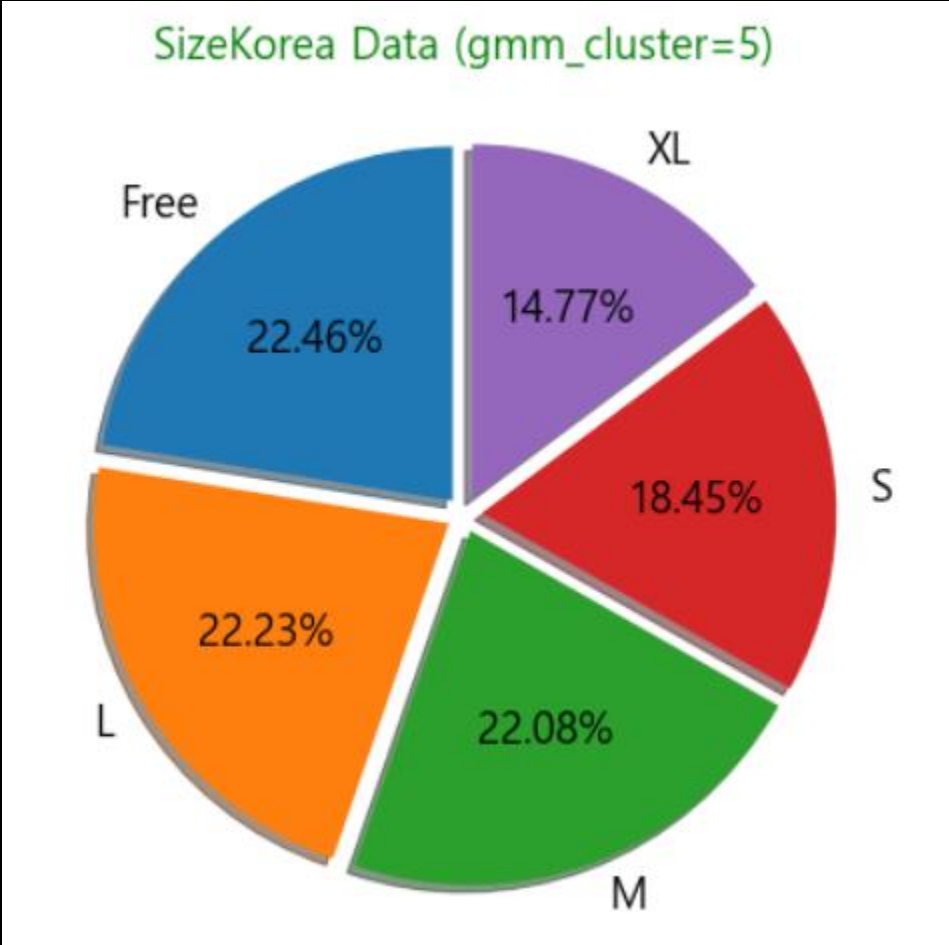
군집화 결과 Clustering

바지 n = 5 군집화 결과



cluster_char	
L	459
M	662
S	666
XL	317
XS	525

원피스 n = 5 군집화 결과



size(gmm5)	
Free	599
L	593
M	589
S	492
XL	394

분석 기법 및 결과

분류 Classification

: 사이즈 코리아 데이터 군집화 결과 기반
무신사 웹 크롤링 데이터 분류

바지 | 무신사 데이터 분류 모델 비교



n = 5	Random Forest	LightGBM	XGBoost
accuracy	0.9392	0.9506	0.9544
precision	0.9384	0.9504	0.9504
recall	0.9388	0.9397	0.9467
f1	0.9385	0.9440	0.9477

- ▶ 사이즈 코리아 데이터 기반 새로운 사이즈 분류 기준을 설정
- ▶ 무신사에서 크롤링한 제품들을 새로운 사이즈 분류 기준 기반으로 나누기 위해 분류 진행

각 모델들의 성능 비교 후 최종 모델을 선정

분석 기법 및 결과

분류 Classification

: 사이즈 코리아 데이터 군집화 결과 기반
무신사 웹 크롤링 데이터 분류

원피스 |
무신사 데이터
분류 모델 비교



n = 5	Random Forest	LightGBM	XGBoost
accuracy	0.9551	0.9625	0.9551
precision	0.9594	0.9646	0.9581
recall	0.9545	0.9614	0.9532
f1	0.9563	0.9628	0.955

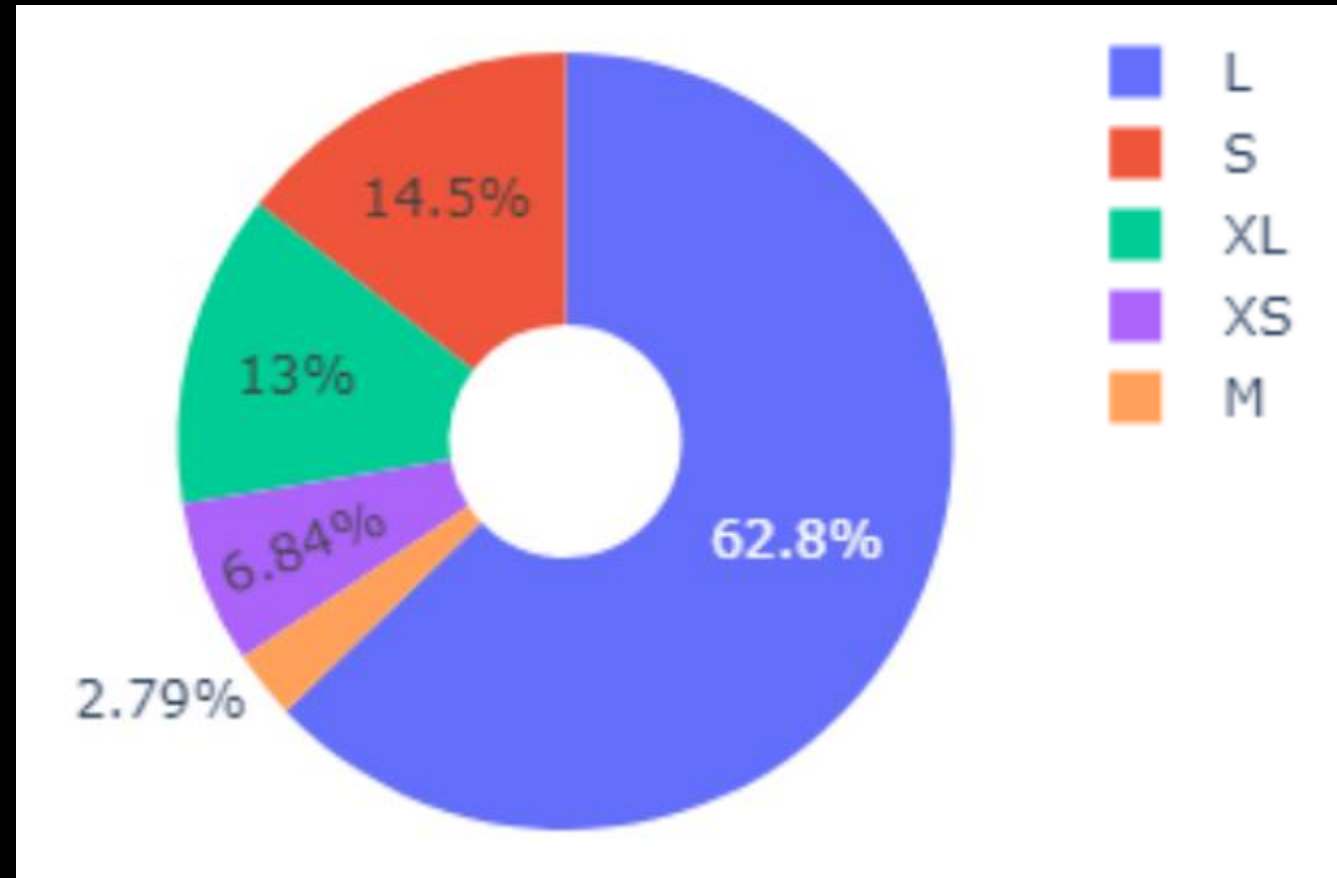
- ▶ 사이즈 코리아 데이터 기반 새로운 사이즈 분류 기준을 설정
- ▶ 무신사에서 크롤링한 제품들을 새로운 사이즈 분류 기준 기반으로 나누기 위해 분류 진행

각 모델들의 성능 비교 후 최종 모델을 선정

분석 기법 및 결과

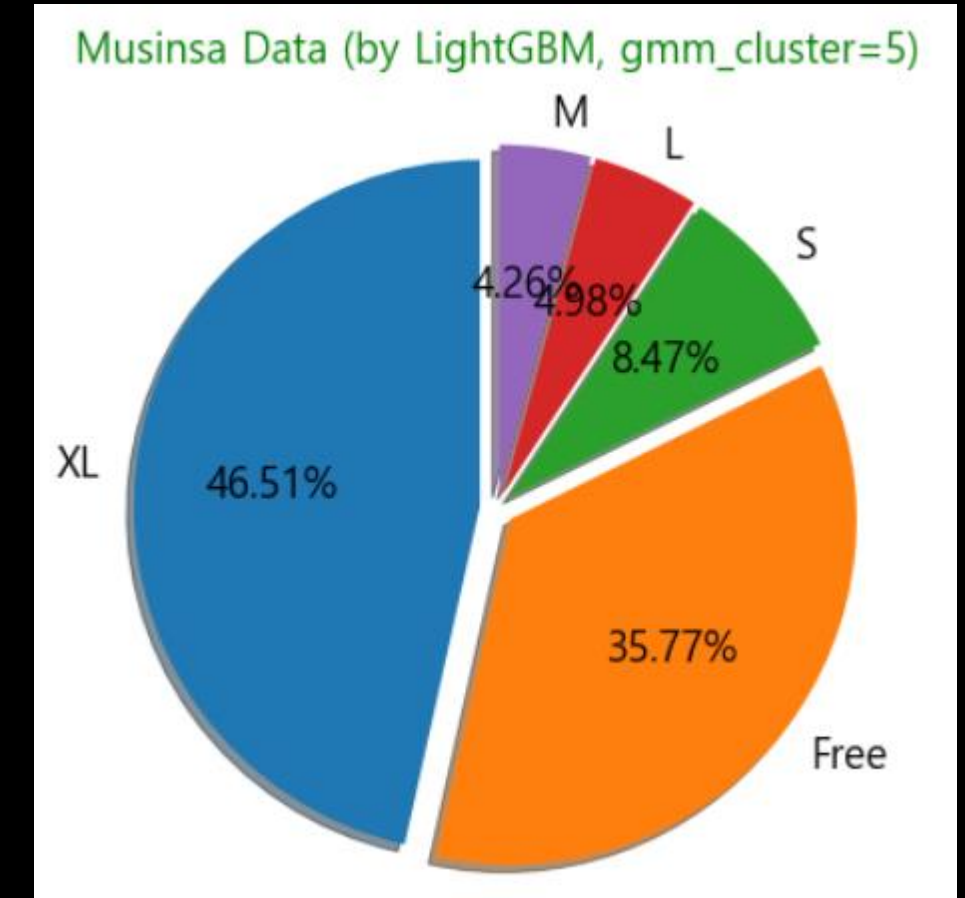
분류 결과

바지 분류 결과



- 한국인 체형 반영 기준으로 무신사 바지 제품 분류
- **L 사이즈**의 제품이 62.8%로 가장 큰 비율 차지
S 사이즈의 제품이 14.5%로 그 다음 순위임
- 무신사 사이트에서 여성 바지의 대다수가 **젠더리스** 상품임
- 치수가 비교적 큰 남성과 사이즈를 공유하는 상품이 대부분

원피스 분류 결과

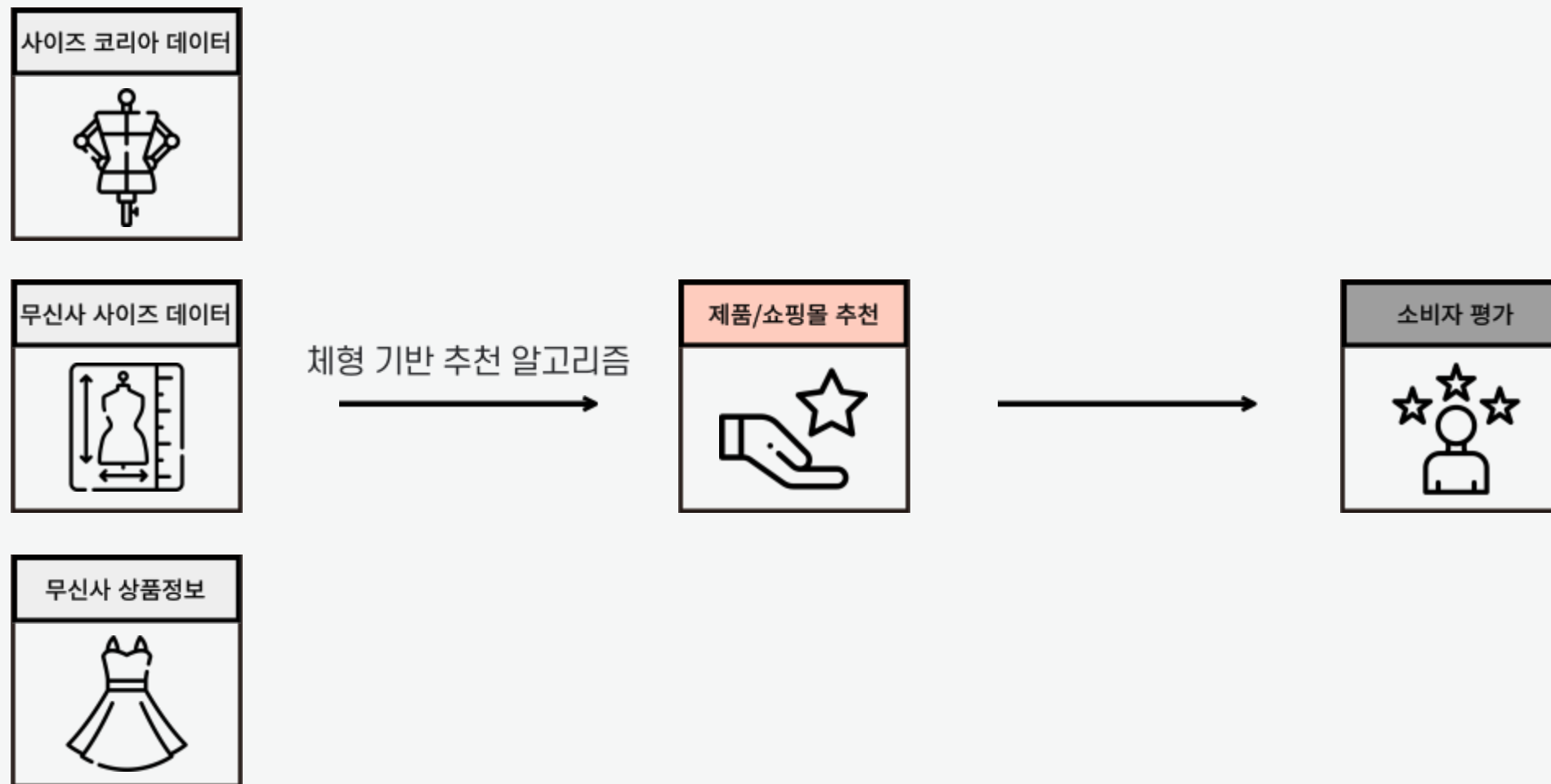


- 실제 인체 치수 데이터의 사이즈 분류 기준으로 무신사 원피스 제품의 사이즈 분류
- **XL 사이즈**가 가장 큰 비율 차지 (46.51%)
- 무신사 원피스의 대부분이 **실제 인체 치수보다 작은 사이즈**라는 것을 알 수 있음

모델 선정 계획

추천 시스템 Recommendations

추천 시스템 개발



```
def brands_for_you(ID):  
    ### 고객 ID 입력  
    sample_ID = ID  
  
    ### 사이즈 코리아 군집화 기반 사이즈 분류 확인  
    sample_size = size_korea_df.iloc[sample_ID, :][7]  
  
    ### 피실험자의 사이즈에 해당하는 제품을 갖고 있는 브랜드 top 5 추출  
    ### Musinsa Standard의 데이터 수가 너무 많아서 제외함  
    brand_got_sizes = recomm_musinsa[recomm_musinsa['classification_char'] == sample_size]  
    top5_brands = brand_got_sizes.groupby('brand_name')[['product_name']].count().sort_values('product_name', ascending = False)[1:6]  
    TOP5 = [x for x in top5_brands.index]  
    return TOP5
```

```
brands_for_you(983)  
  
[349]: ['GLW', 'YALE', 'STCO', 'BRENSON', '247SERIES']
```

모델 선정 계획

추천 시스템 Recommendations

추천 시스템 개발

```
def products_for_you(ID):  
    ### 고객 ID 입력  
    sample_ID = ID  
  
    ### 사이즈 코리아 군집화 기반 사이즈 분류 확인  
    sample = size_korea_df.iloc[sample_ID, :]  
    sample_size = sample[7]  
  
    ### 피실험자의 신체 치수 추출 후 무신사 제품 치수 순서대로 배열  
    sample_size_li = [sample[3], sample[4], sample[2], sample[5], sample[1]]  
  
    ### 피실험자의 사이즈에 해당하는 제품들 중 오차가 가장 적은 제품을 추천  
    ## 같은 사이즈 제품 추출  
    same_size_products = recomm_musinsa[recomm_musinsa['classification_char'] == sample_size]  
    same_size_products = same_size_products.reset_index().iloc[:, 1:]  
  
    ## 각 제품들의 치수 오차 계산  
    es = []  
  
    for j in range(len(same_size_products)):  
        e = 0  
        for i in range(len(sample_size_li)):  
            e = e + abs(sample_size_li[i] - same_size_products['size_data'][j][i])  
        es.append(e)  
  
    ## 무신사 데이터에 오차 계산값 추가 후  
    ## 오차들의 합이 가장 적은 순서대로 정렬  
    ## top 50 상품명 추출  
    same_size_products['difference'] = es  
  
    ordered_products = same_size_products.sort_values('difference')  
    top_products = ordered_products.loc[:, ['product_name', 'difference']][:50]  
    TOP_PRODUCTS = top_products['product_name'].unique()  
    return TOP_PRODUCTS
```

products_for_you(983)

```
[352]: array(['DISCO FLEECE PANTS_PINK', 'DISCO FLEECE PANTS_AQUA BLUE',  
             '[ONEMILE WEAR] SMALL ARCH SWEAT PANTS PG BROWN',  
             '소프트 브러시드 와이드 팬츠 BLACK', '소프트 브러시드 와이드 팬츠 MELANGE GRAY',  
             '소프트 브러시드 와이드 팬츠 BEIGE', 'J투턱벨벳틴팬츠 JWPT721C',  
             'J투턱저지코듀로이와이드(ON) JWPT727A', 'Wide-fit Slacks _ Black',  
             'Double pin-tuck Pants _ Black',  
             'Dawn 2Pleats Wide Pants DCPT007Ivory',  
             '90s Tartan Check Sweatpants (Deep Blue)',  
             '90s Tartan Check Sweatpants (Black)',  
             '남성 2005 D 파이닝 트루저 팬츠 - 블루 / A0357109E1701',  
             '깡 스타일리스트 x 플렉 베를린 K75 로우', '[UNISEX] NEW DIEGO CARGO PANTS IVORY',  
             '라이트웨이트 스키니 히든 밴딩 크롭 슬랙스 [블랙]', '깡스타일리스트 x 플렉 베를린 K48 화이트',  
             '코튼 와이드 퍼티그 벌룬 팬츠 [카키]', '코튼 와이드 퍼티그 벌룬 팬츠 [블랙]',  
             '와이드 히든 밴딩 크롭 슬랙스 [미디엄 그레이]', '와이드 히든 밴딩 크롭 슬랙스 [블랙]',  
             '모두바지 오리진 (Black)', '모두바지 오리진 (Beige)', '몰스킨 와이드 밴딩 팬츠 IVORY',  
             '모두바지 오리진 (Khaki)'], dtype=object)
```


활용방안 및 기대효과

명확한 의류 사이즈 분류표 제시

A line drawing of a female figure from the back, showing measurement points for various body parts. The labels are: F (Shoulders), A (Chest/Breast), B (Chest/Breast), C (Waist), D (Hip), V (Waist), and G (Sleeve Length CB).

WOMEN'S SIZE

IT			36	38	40	42	44	46	48	50	52
USA			0	2	4	6	8	10	12	14	16
			XS		S		M		L		XL
Chest/Breast	A	inch	31,5	33,1	34,6	36,2	37,8	39,4	40,9	42,5	44,1
	B	cm	80	84	88	92	96	100	104	108	112
Waist	C	inch	23,6	25,2	26,8	28,3	29,9	31,5	33,1	34,6	36,2
		cm	60	64	68	72	76	80	84	88	92
Hip	D	inch	33,9	35,4	37	38,6	40,2	41,7	43,3	44,9	46,5
		cm	86	90	94	98	102	106	110	114	118
Shoulders	F	inch	13,4	14	15	15,5	16,1	16,5	16,8	17,1	17,4
		cm	34	35,6	38	39,4	41	41,8	42,6	43,4	44,2
Sleeve Length CB	G	inch	29,2	29,8	30,6	31,1	31,7	32,1	32,4	32,6	33
		cm	74,2	75,7	77,6	79	80,5	81,6	82,3	82,7	83,7

시대에 따라 변화하는 한국인 체형에 기반한
현실적인 의류 사이즈 분류표 기준을 제시함으로써
소비자 피해 방지 및 신체 이미지 왜곡 완화

체형 기반 의류 제품 추천



- 최근 의류 시장에서 맞춤화/개인화 서비스가 크게 성장
- 개발 모델을 통해 소비자 니즈 충족
- 새로 만든 분류 기준 기반 추천 시스템 활용해
소비자 친화적인 분류 기준 대중화