

# 프로젝트 중간발표

내 옷 좀 추천해 (조)

# 프로젝트 진행자 정보

- 프로젝트 종류: 머신러닝 프로젝트
- 프로젝트 기간
  - 진행 기간: 1개월
  - 세부 기간: 2023.06.01 ~ 2023.06.21.
- 인원: 7명
- 팀원: 김보석, 박미영, 박성호, 이성희, 이정빈, 정설령, 최은욱



## CONTENTS

- ❑ 프로젝트 주제
- ❑ 현재 진행 상황
- ❑ 데이터 소개
- ❑ EDA
  - ❑ 상의
  - ❑ 하의
  - ❑ 리뷰
  - ❑ Conclusion
- ❑ 추후 진행 방향

# 프로젝트 주제

“ 의류 실측 데이터를 이용한  
프리스아이즈 분석에 따른 개인화 의류 추천 시스템 ”

# 프로젝트 주제



본 프로젝트는 고객의 더 나은 선택과 편의를 위해 무신사 내 상,하의를 대상으로 특성, 사이즈, 리뷰 데이터 등을 분석을 실시함으로써,  
**프리사이즈 의류를 표준화된 사이즈로 분류하고 개인화된 추천시스템을 만들고자 함**

## Analysis OBJECT

획일화되지 않은 프리사이즈 표기 방식을 통한 고객 불편사항

### 주요 ISSUE 및 과제

| 프리사이즈를 기존 표준사이즈로 분류  
| 사용자의 키, 몸무게에 맞춘 의류 추천

#### 목적 1



기초통계 및 EDA를 통한  
분석



#### 목적 2



기존 사이즈로의  
프리사이즈 의류 분류

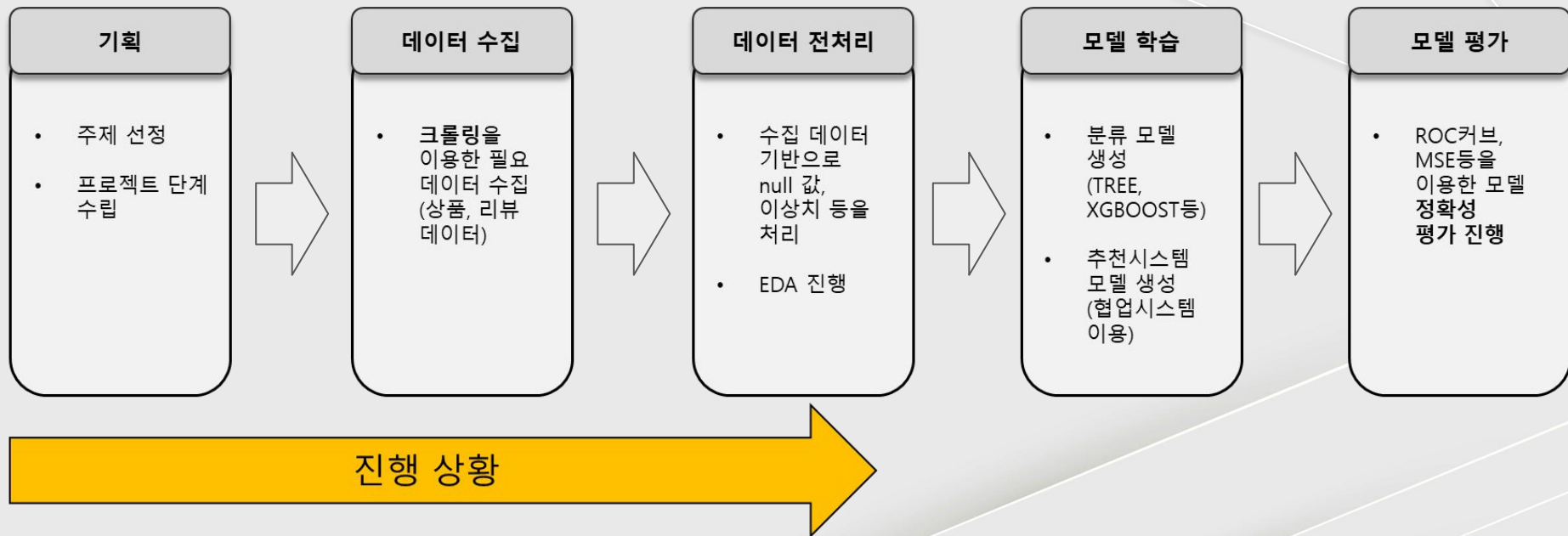


#### 목적 3



사용자들의 리뷰데이터를 이용한  
개인화 의류 추천시스템

# 현재 진행 상황



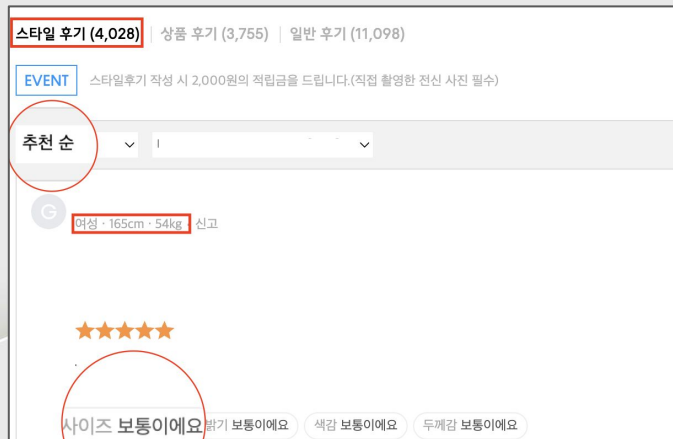
# 데이터 소개

□ 온라인 쇼핑몰 '무신사'에서 의류 제품 정보와 고객 리뷰 수집

상의		하의		리뷰
제품정보	사이즈정보	제품정보	사이즈정보	"보통"
제품명	총장	제품명	총장	성별
브랜드	어깨너비	브랜드	허리단면	키
성별	가슴단면	성별	엉덩이단면	몸무게
링크	소매길이	링크	허벅지단면	
			밑위	
			밑단단면	

※ 무신사 상의, 하의 카테고리 추천순 기준  
1 - 400 페이지 상품 정보 수집

※ 제품 후기 중 스타일 후기(추천순) 사이즈  
표기가 '보통이에요'인 리뷰



## EDA - 상의

## ☐ 상의 데이터 형태

		제품명	브랜드	성별	사이즈	총장	어깨너비	가슴단면	소매길이	링크
0		에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남여	XS	63.0	47.5	49.0	19.0	<a href="https://www.musinsa.com/app/goods/1848166">https://www.musinsa.com/app/goods/1848166</a>
1		에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남여	S	65.0	49.0	51.5	20.0	<a href="https://www.musinsa.com/app/goods/1848166">https://www.musinsa.com/app/goods/1848166</a>
2		에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남여	M	69.0	51.0	55.0	21.5	<a href="https://www.musinsa.com/app/goods/1848166">https://www.musinsa.com/app/goods/1848166</a>
3		에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남여	L	71.0	52.5	57.5	22.5	<a href="https://www.musinsa.com/app/goods/1848166">https://www.musinsa.com/app/goods/1848166</a>
4		에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남여	XL	73.0	54.0	60.0	23.5	<a href="https://www.musinsa.com/app/goods/1848166">https://www.musinsa.com/app/goods/1848166</a>
...		...	...	...	...	...	...	...	...	...
101523		SITTING PUPPY CREWNECK KS [MELANGE GREY]	VIVASTUDIO	남	M	70.0	58.5	61.5	60.0	<a href="https://www.musinsa.com/app/goods/1679937">https://www.musinsa.com/app/goods/1679937</a>
101524		SITTING PUPPY CREWNECK KS [MELANGE GREY]	VIVASTUDIO	남	L	71.5	60.0	64.0	61.0	<a href="https://www.musinsa.com/app/goods/1679937">https://www.musinsa.com/app/goods/1679937</a>
101525		SITTING PUPPY CREWNECK KS [MELANGE GREY]	VIVASTUDIO	남	XL	75.0	61.5	66.5	62.0	<a href="https://www.musinsa.com/app/goods/1679937">https://www.musinsa.com/app/goods/1679937</a>
101526		CHANCE HOODIE(ORANGE/기모)	CHANCECHANCE	남	M	69.0	56.0	70.0	63.0	<a href="https://www.musinsa.com/app/goods/1670376">https://www.musinsa.com/app/goods/1670376</a>
101527		High Neck Slim Top_Pink Violet	NICK&NICOLE	여	FREE	72.0	37.0	40.0	64.0	<a href="https://www.musinsa.com/app/goods/1661633">https://www.musinsa.com/app/goods/1661633</a>
101528	rows × 9 columns									



# EDA - 상의

## 전처리 과정

### 소매길이 칼럼 제거

- 반팔, 긴팔, 민소매의 구분이 되어있지 않음
- 상관관계가 전혀 없음

### Null 값 있는 행 제거 및 중복 데이터 제거

- 기본적으로 NaN 값은 제거 진행
- 기준: 링크, 실측데이터

### 성별 칼럼의 '라이프' 제거

- 19개의 행 제거
- 개수가 적고, 성별을 구분하기 어려움

### 특정 브랜드 내 사이즈 제거

- 특정 브랜드 (GLIMMER) 내에 110~150 사이즈 존재
- 해당 사이즈는 아동용 사이즈로 판별되어 삭제 처리 진행

### 사이즈 칼럼 추가 생성

- 기존 사이즈 열 = row data의 사이즈 열
- 사이즈1 = 기존 사이즈 열 복사
- 사이즈1을 사이즈 맵핑 시 바꿀 예정
- 사이즈2 = 'FREESIZE' 구별을 위한 열 생성

### 'FREESIZE' 설정

- 링크 개수가 1개에 사이즈가 하나인 제품을 사이즈2 열에 'FREESIZE' 변경

### 사이즈 맵핑

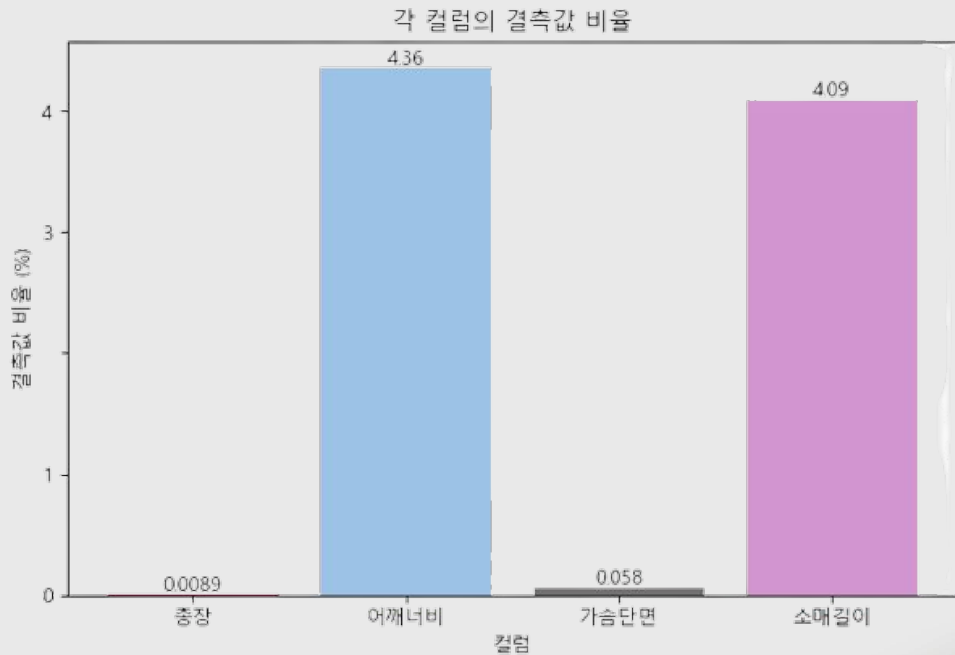
### EDA 수행

### 이상치 제거

- IQR을 이용한 이상치 제거

# EDA - 상의

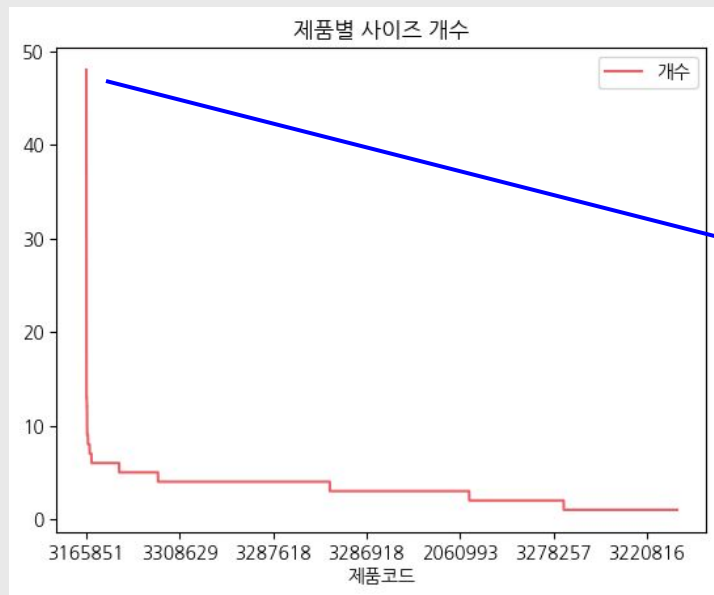
## □ NULL 값 분포



전체 데이터 개수 : 101528  
결측치 제거 후 데이터 개수 : 97052  
데이터 개수 차이: 4476

# EDA - 상의

- ❑ 색상과 사이즈가 함께 기재되어 **동일 사이즈**임에도 다른 사이즈로 구분됨
- ❑ 중복된 데이터는 제거 필요



Size Info 사이즈정보

구매내역 실측 선택

실측 직접 입력

측정법

cm	총장	어깨너비	가슴단면	소매길이
MY	가지고 계신 제품의 실측을 입력해 보세요~!			
WHITE S	68.5	52.5	52.5	21.5
WHITE M	70.5	54	55	22.5
WHITE L	72.5	55.5	57.5	23.5
WHITE XL	74.5	57	60	24.5
BLACK S	68.5	52.5	52.5	21.5
BLACK M	70.5	54	55	22.5
BLACK L	72.5	55.5	57.5	23.5
BLACK XL	74.5	57	60	24.5
CREAM S	68.5	52.5	52.5	21.5
CREAM M	70.5	54	55	22.5
CREAM L	72.5	55.5	57.5	23.5
CREAM XL	74.5	57	60	24.5

# EDA - 상의

## 전 데이터에 통용될 수 있는 사이즈 기준 수립을 위해 맵핑 시행

```
1 size_dict = {
2     'xsmall':'XS', 'small':'S', 'medium':'M', 'large':'L', 'xlarge':'XL',
3     'Xsmall':'XS', 'Small':'S', 'Medium':'M', 'Large':'L', 'Xlarge':'XL',
4     'SMALL':'S', 'MEDIUM':'M', 'LARGE':'L', 'EXTRA LARGE':'XL', 'EXTRA L':'XL',
5     '85':'XS', '90':'S', '95':'M', '100':'L', '105':'XL', '110':'XXL', '115':'XXXL',
6     '2XL':'XXL', '3XL':'XXXL', '4XL':'XXXXL'
7 }
8
9 size_dict2 = {
10     'SS':'XS', 'LL':'XL', '3L':'XXL', '4L':'XXXL', '5L':'XXXXL',
11     'W1':'XS', 'W2':'S'
12 }
13
14 size_label = ['XS', 'S', 'M', 'XXXL', 'XXL', 'XL', 'L'] # 'FREE'
15
16 # 여성 의류인 경우만 적용
17 size_dict3 = {
18     '44':'XS', '55':'S', '66':'M', '77':'L', '88':'XL',
19 }
20
21 # 남성 미국 사이즈
22 size_ameri = {
23     '14':'XS', '15':'S', '16':'M', '17':'L', '18':'XL'
24 }
```

1743

• 맵핑 전

224

• 맵핑 후

8

• ['XS', 'S', 'M', 'L', 'XL', 'XXL',  
'XXXL', 'FREESIZE'] 이외의 모든  
사이즈 제거

∴ 최종 88636개 데이터

# EDA - 상의

## 전 데이터에 통용될 수 있는 사이즈 기준 수립을 위해 맵핑 시행

```
1 size_dict = {
2     'xsmall':'XS', 'small':'S', 'medium':'M', 'large':'L', 'xlarge':'XL',
3     'Xsmall':'XS', 'Small':'S', 'Medium':'M', 'Large':'L', 'Xlarge':'XL',
4     'SMALL':'S', 'MEDIUM':'M', 'LARGE':'L', 'EXTRA LARGE':'XL', 'EXTRA L':'XL',
5     '85':'XS', '90':'S', '95':'M', '100':'L', '105':'XL', '110':'XXL', '115':'XXXL',
6     '2XL':'XXL', '3XL':'XXXL', '4XL':'XXXXL'
7 }
8
9 size_dict2 = {
10     'SS':'XS', 'LL':'XL', '3L':'XXL', '4L':'XXXL', '5L':'XXXXL',
11     'W1':'XS', 'W2':'S'
12 }
13
14 size_label = ['XS', 'S', 'M', 'XXXL', 'XXL', 'XL', 'L'] # 'FREE'
15
16 # 여성 의류인 경우만 적용
17 size_dict3 = {
18     '44':'XS', '55':'S', '66':'M', '77':'L', '88':'XL',
19 }
20
21 # 남성 미국 사이즈
22 size_ameri = {
23     '14':'XS', '15':'S', '16':'M', '17':'L', '18':'XL'
24 }
```

1743

• 맵핑 전

224

• 맵핑 후

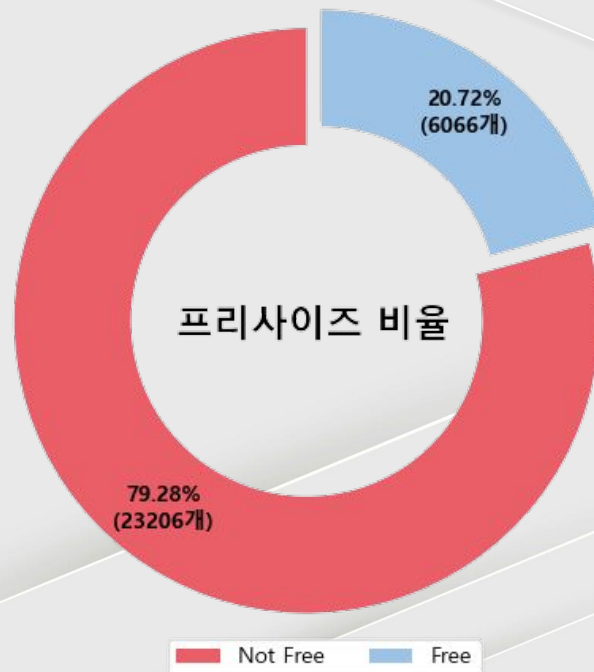
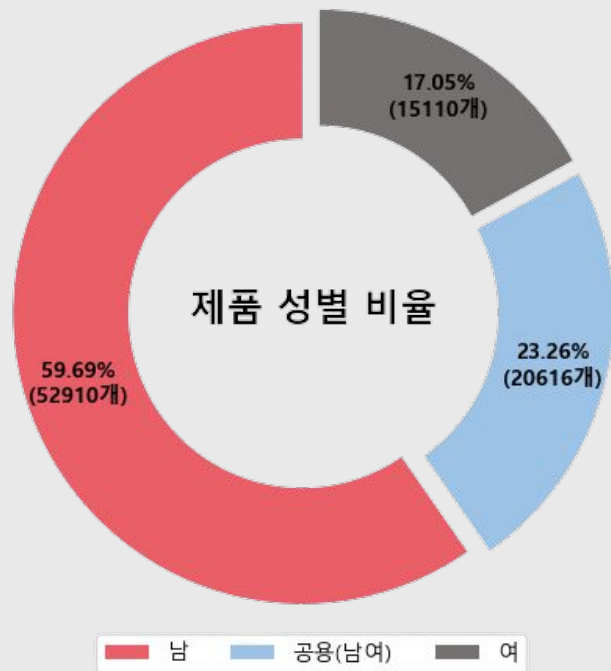
8

• ['XS', 'S', 'M', 'L', 'XL', 'XXL',  
'XXXL', 'FREESIZE'] 이외의 모든  
사이즈 제거

∴ 최종 88636개 데이터

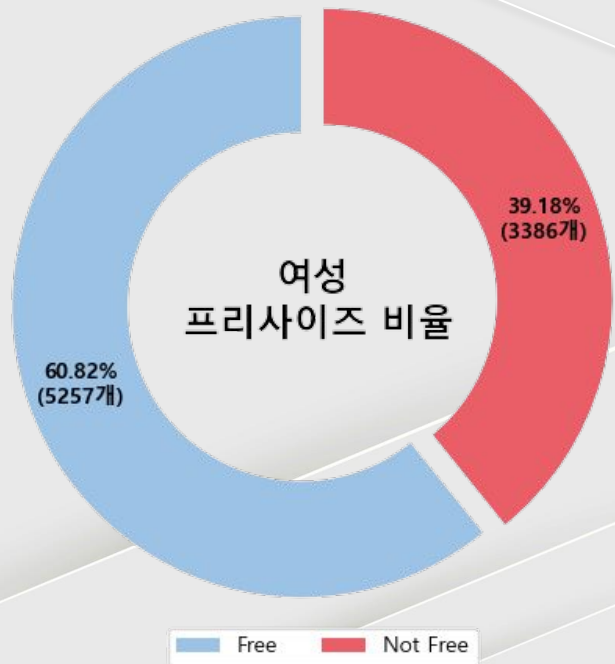
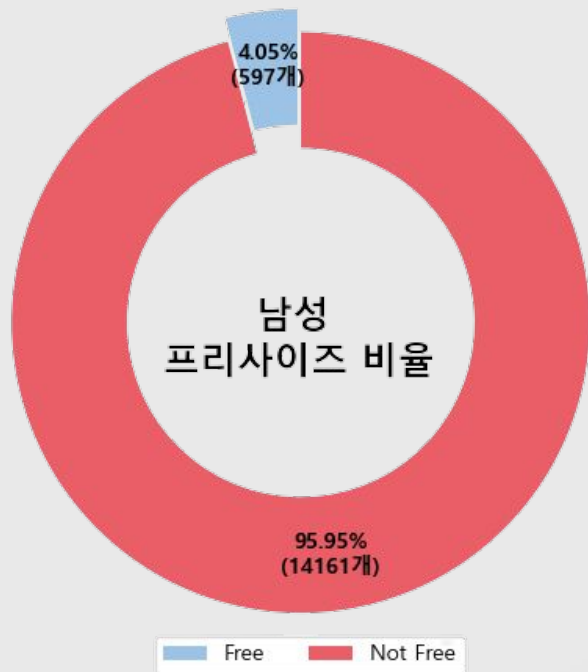
# EDA - 상의

## □ 데이터 분포 확인



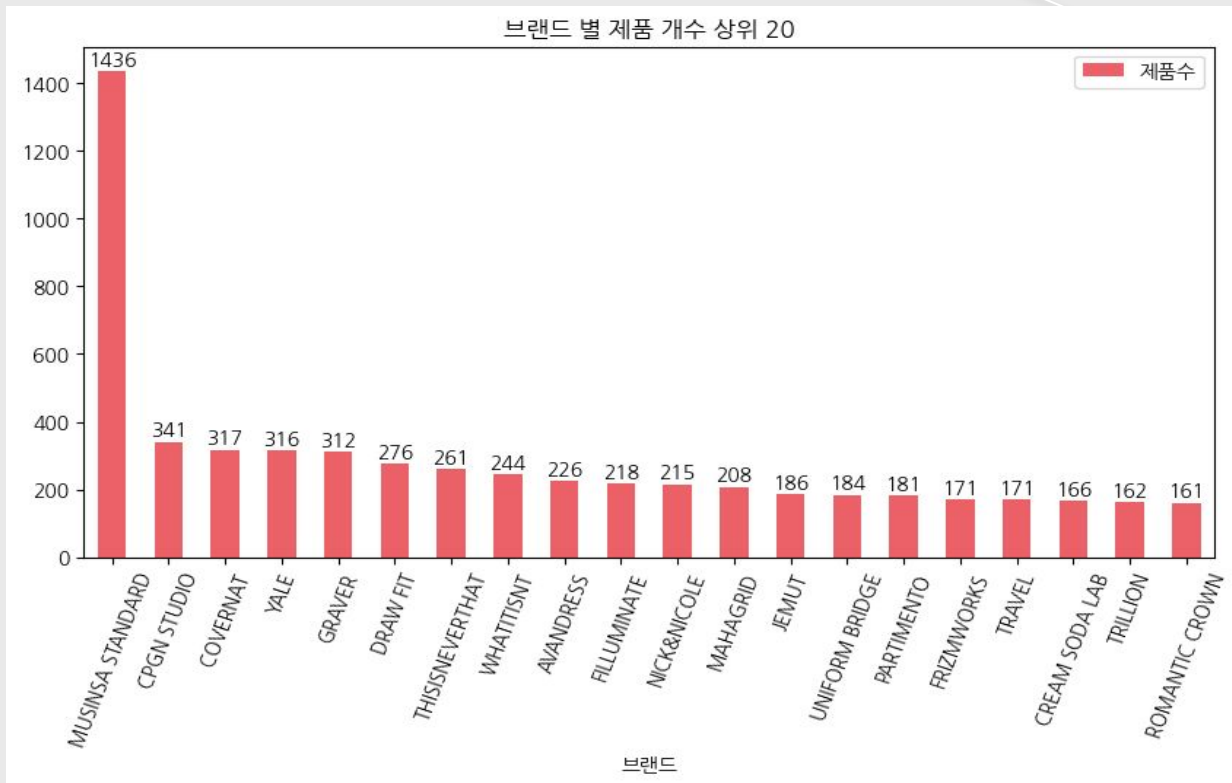
# EDA - 상의

## □ 성별에 따른 프리사이즈 비율



# EDA - 상의

## □ 제품 개수 상위 20위 브랜드



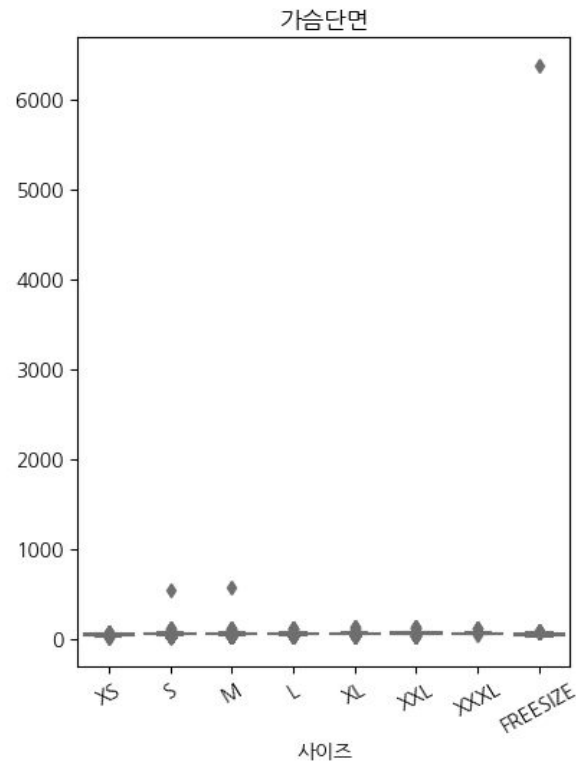
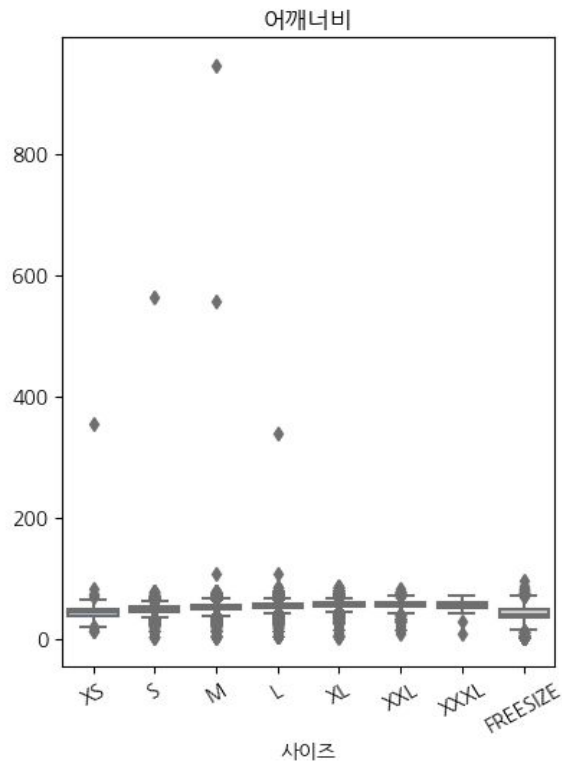
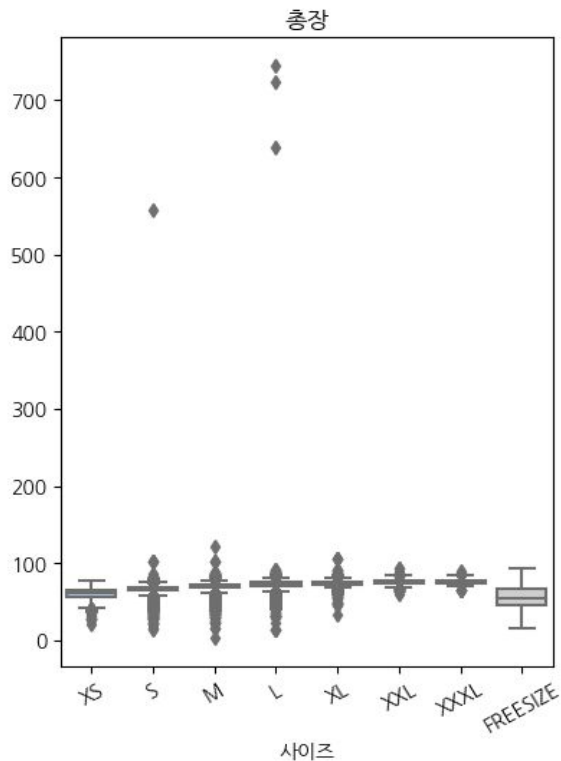


# EDA - 상의

이상치 제거 (전) - 전체

전체 데이터 개수 : 88636

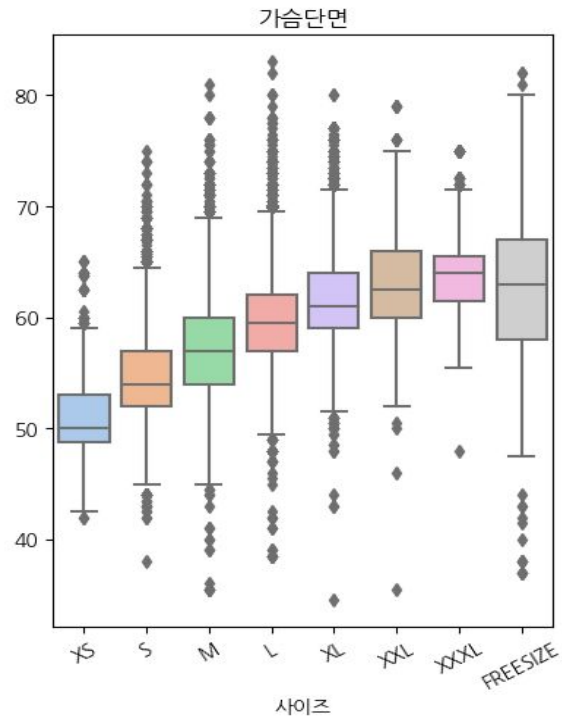
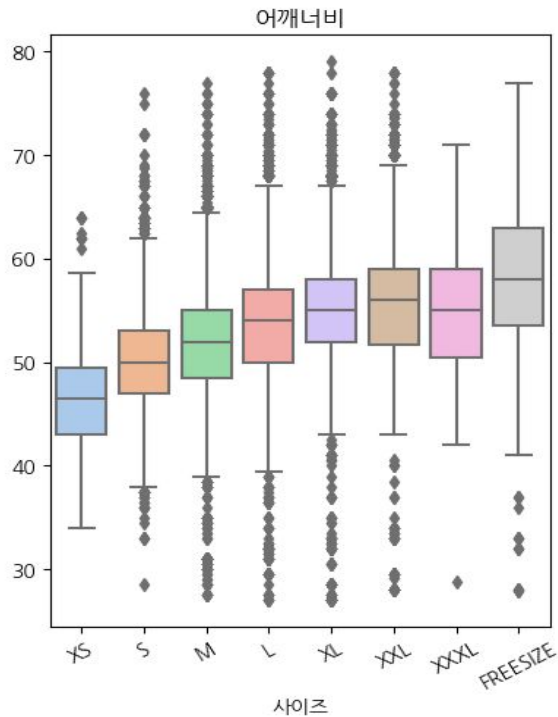
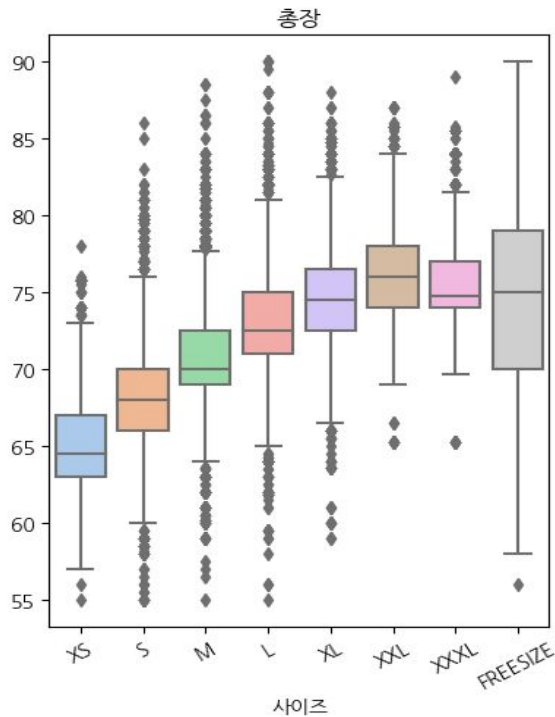
사이즈 별 실측데이터 boxplot



# EDA - 상의 이상치 제거 (후) - 남성

이상치 제거 후 전체 데이터 개수 : 87917

사이즈 별 남성복 실측데이터 boxplot



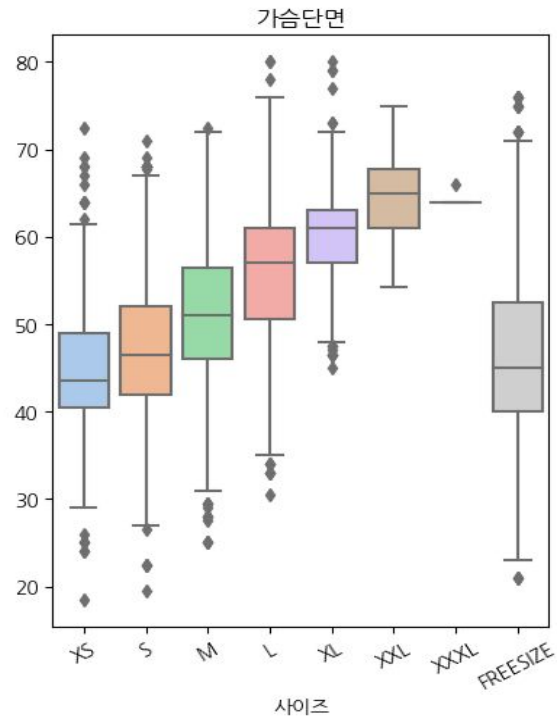
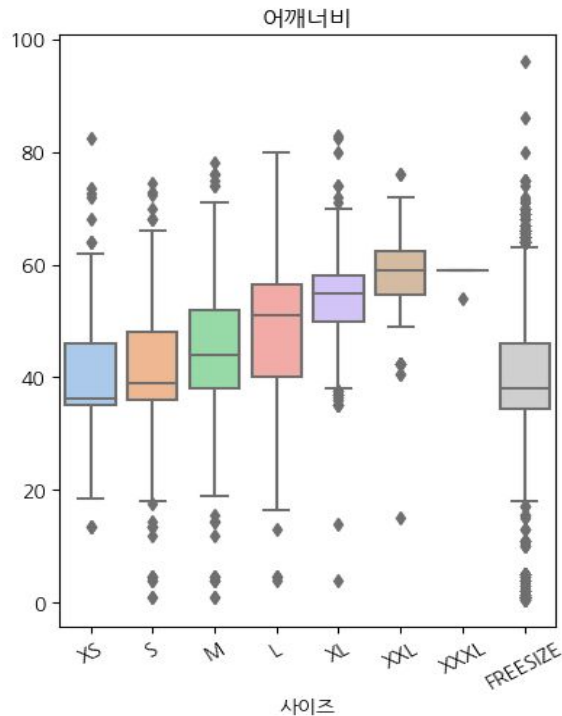
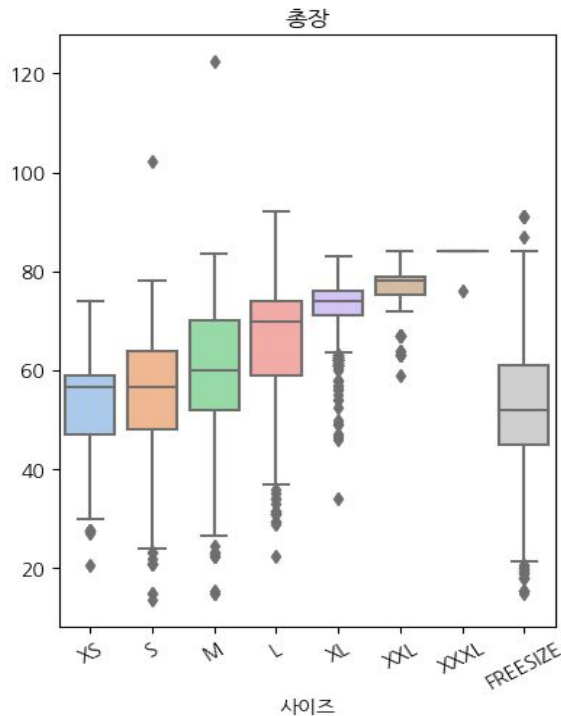
\* 이상치는 IQR 3배를 기준으로 설정

# EDA - 상의

이상치 제거 (후) - 여성

이상치 제거 후 전체 데이터 개수 : 87917

사이즈 별 여성복 실측데이터 boxplot



\* 이상치는 IQR 3배를 기준으로 설정

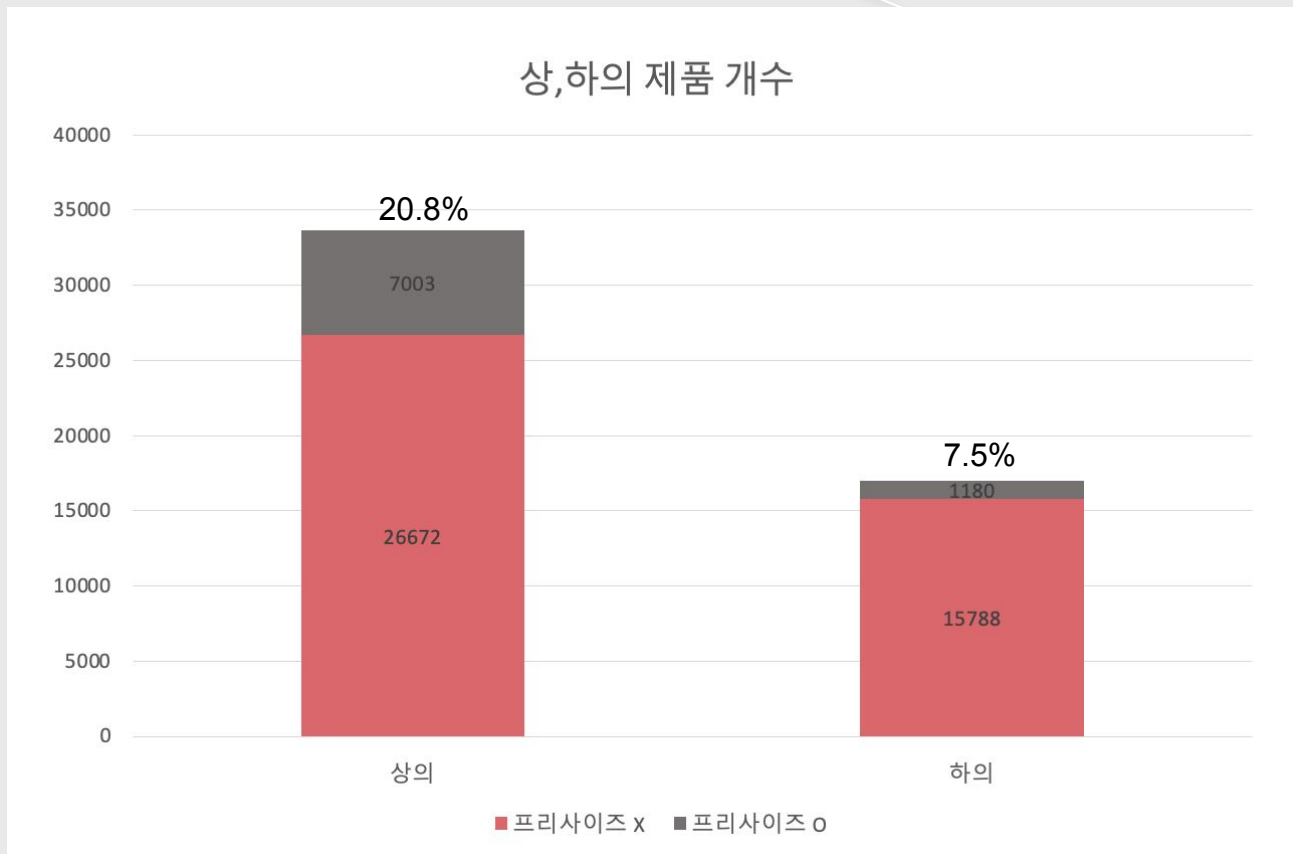
# EDA - 하의

## 하의 데이터 형태

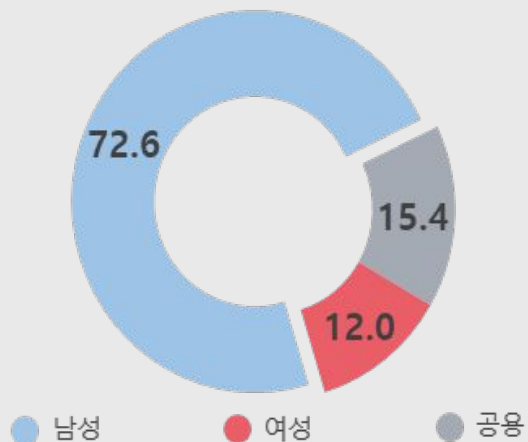
	사이즈	총장	허리단면	엉덩이단면	허벅지단면	밑위	밑단단면	제품명	브랜드	성별	링크
0	S	58.0	31.5	NaN	32.0	33.0	30.0	Deep One Tuck Sweat Shorts [Grey]	XERO	남	<a href="https://www.musinsa.com/app/goods/1926048">https://www.musinsa.com/app/goods/1926048</a>
1	M	60.0	32.5	NaN	33.5	34.5	31.5	Deep One Tuck Sweat Shorts [Grey]	XERO	남	<a href="https://www.musinsa.com/app/goods/1926048">https://www.musinsa.com/app/goods/1926048</a>
2	L	62.0	34.5	NaN	35.0	36.0	33.0	Deep One Tuck Sweat Shorts [Grey]	XERO	남	<a href="https://www.musinsa.com/app/goods/1926048">https://www.musinsa.com/app/goods/1926048</a>
3	Natural_XS	100.0	30.0	51.00	32.0	29.0	22.0	린넨 라이크 세미 와이드 밴딩 팬츠 세트	SUARE	남	<a href="https://www.musinsa.com/app/goods/1924274">https://www.musinsa.com/app/goods/1924274</a>
4	Natural_S	101.0	32.5	54.75	34.0	30.0	23.0	린넨 라이크 세미 와이드 밴딩 팬츠 세트	SUARE	남	<a href="https://www.musinsa.com/app/goods/1924274">https://www.musinsa.com/app/goods/1924274</a>
...	...	...	...	...	...	...	...	...	...	...	...
67518	S	102.5	28.0	NaN	27.0	27.5	12.0	C LOGO BOUCLE EMBROIDERY PANTS_PINK	CURETTY	여	<a href="https://www.musinsa.com/app/goods/2128793">https://www.musinsa.com/app/goods/2128793</a>
67519	28	101.0	38.5	NaN	28.7	25.0	17.5	Stefan / New Straight	PIECE WORKER	남	<a href="https://www.musinsa.com/app/goods/2128269">https://www.musinsa.com/app/goods/2128269</a>
67520	30	102.5	41.0	NaN	30.0	26.0	18.5	Stefan / New Straight	PIECE WORKER	남	<a href="https://www.musinsa.com/app/goods/2128269">https://www.musinsa.com/app/goods/2128269</a>
67521	32	104.0	43.5	NaN	31.3	27.0	19.5	Stefan / New Straight	PIECE WORKER	남	<a href="https://www.musinsa.com/app/goods/2128269">https://www.musinsa.com/app/goods/2128269</a>
67522	34	105.5	46.0	NaN	32.5	28.0	20.5	Stefan / New Straight	PIECE WORKER	남	<a href="https://www.musinsa.com/app/goods/2128269">https://www.musinsa.com/app/goods/2128269</a>

67523 rows x 11 columns

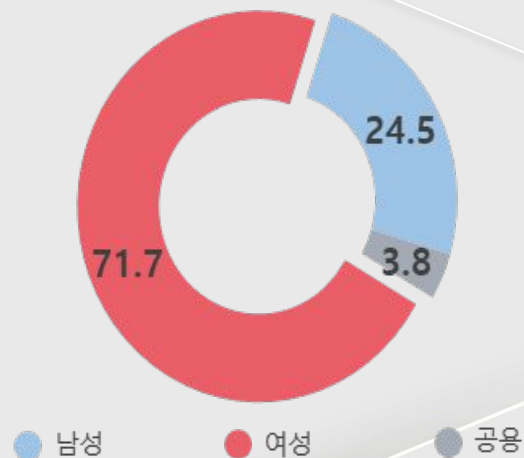
# EDA - 하의



# EDA - 하의



전체 상품 성별 비율



프리사이즈 의류 성별 비율

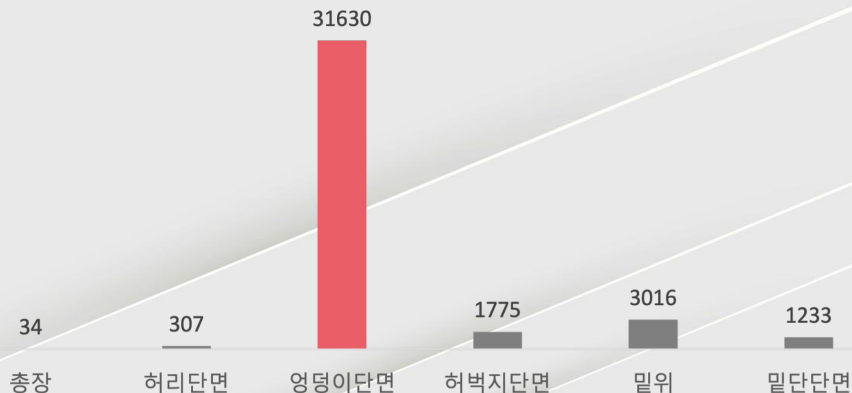
# EDA - 하의

- 사이즈 표에서 단면(허리, 엉덩이, 허벅지) 기입란에 둘레로 혼용하여 표기한 상품이 있음을 확인  
=> 실측 데이터를 변수로 하여 상품을 분류하기 어려워 모델 구현 시, 분류 모델의 정확성이 낮을 수 있음
- 또한, 엉덩이 단면의 전체 데이터(67,523개) 중 31,630개가 NULL값이다.

cm	총장	허리단면	허벅지단면	밑위
MY		가지고 계신 제품의 실측을 입력해 보세요~!		
FREE	44	75	36.2	31

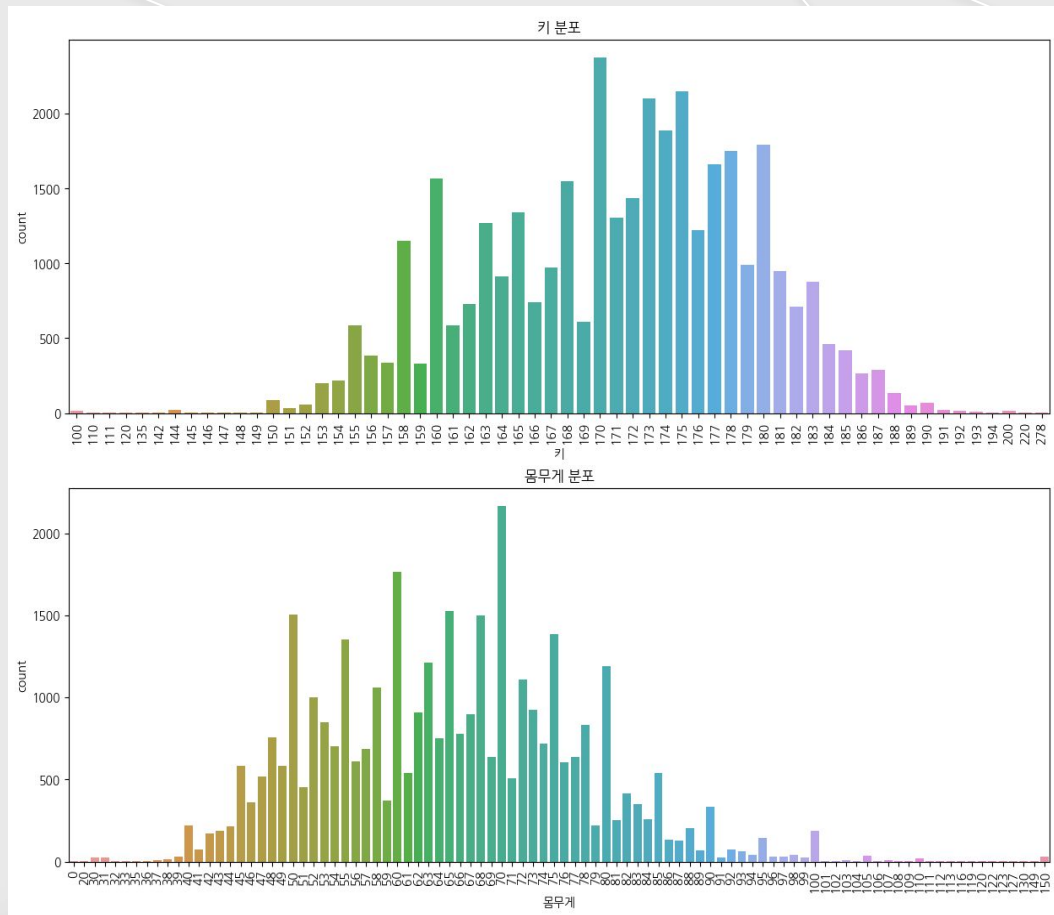
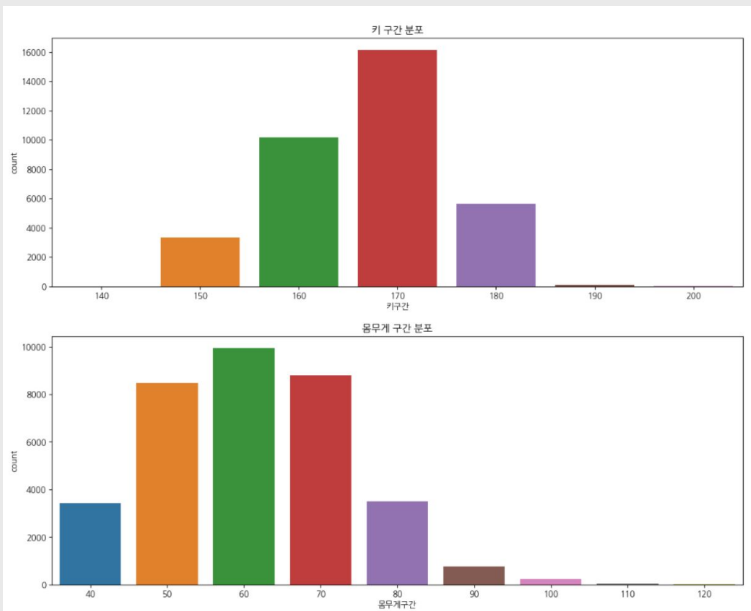
<https://www.musinsa.com/app/goods/3332664>

NULL 값 분포



# EDA - 리뷰

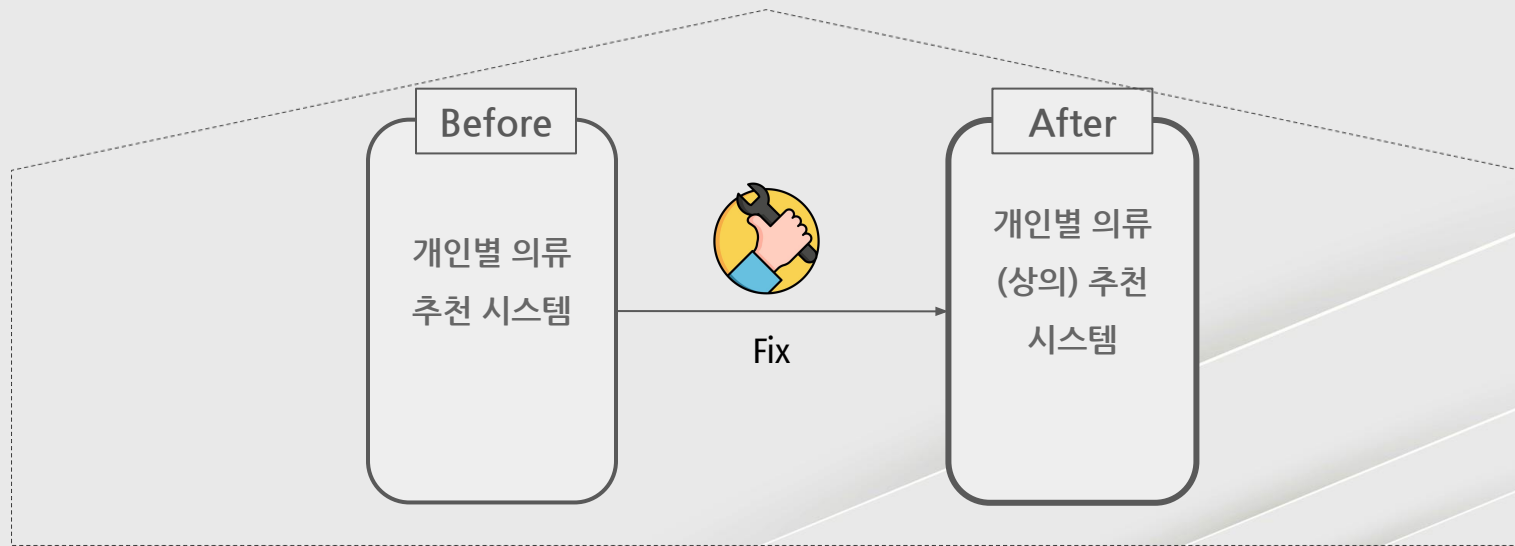
## 리뷰 작성자들의 체형 분포



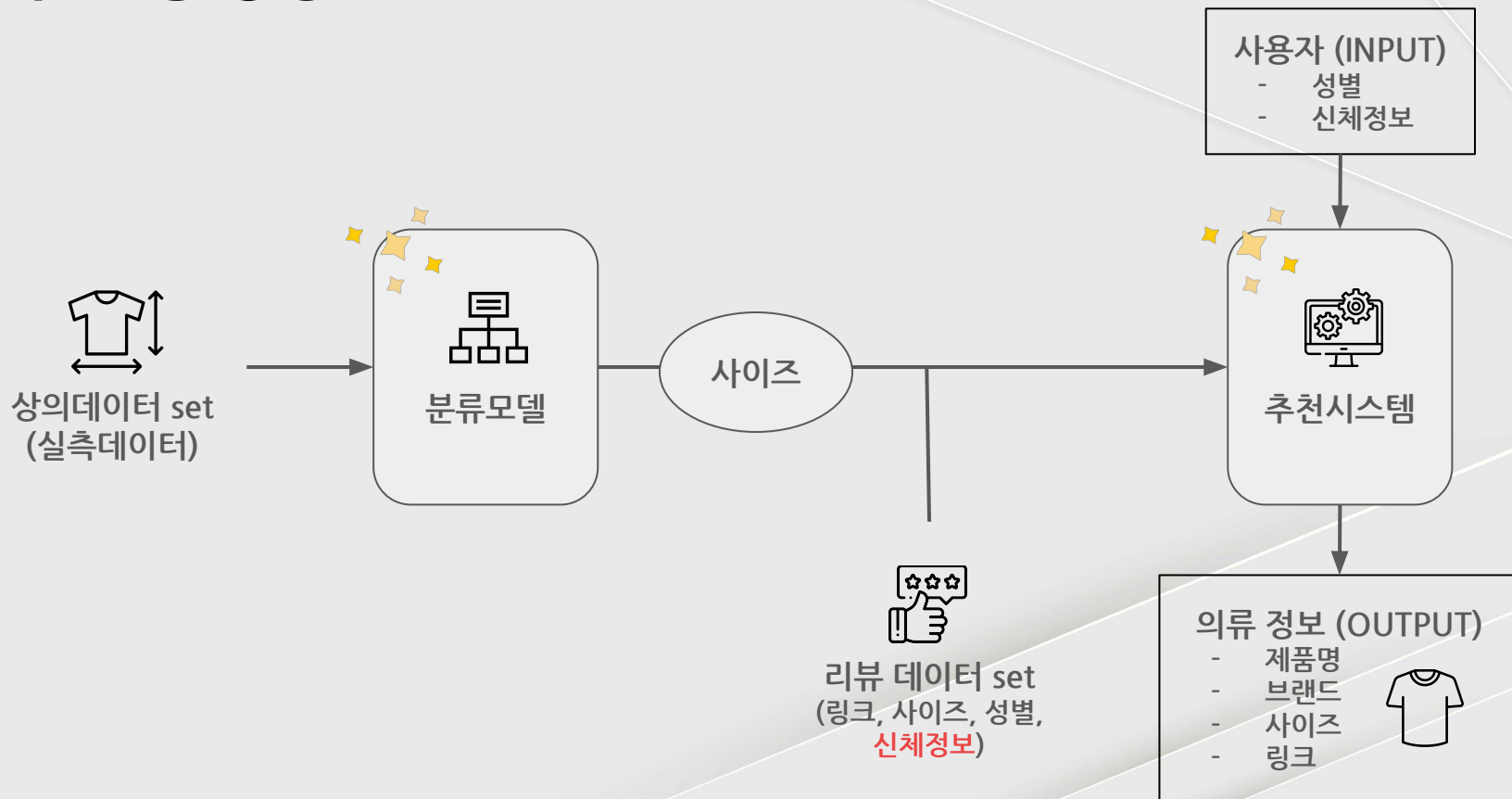


## Conclusion

- ✓ 하의 EDA 결과를 이유로 하의는 개인별 추천 알고리즘의 정확성이 낮을 수 있다고 판단하여 ‘상의 추천 시스템’에 집중하고자 함



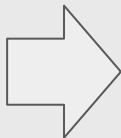
# 추후 진행 방향



# 추후 진행 사항

## 모델 학습

- 분류 모델 생성 (TREE, XGBOOST등)
- 추천시스템 모델 생성 (협업시스템 이용)



## 모델 평가

- ROC커브, MSE등을 이용한 모델 정확성 평가 진행

- ❑ 무신사 PB상품인 무신사스탠다드 상품사이즈를 기준으로 분류모델을 생성하고자 함
- ❑ 분류모델을 바탕으로 프리사이즈 제품의 사이즈를 재정의한다.
- ❑ 제품 실측데이터 및 리뷰 데이터를 활용하여 추천시스템 구현하고자 한다.

**감사합니다**