

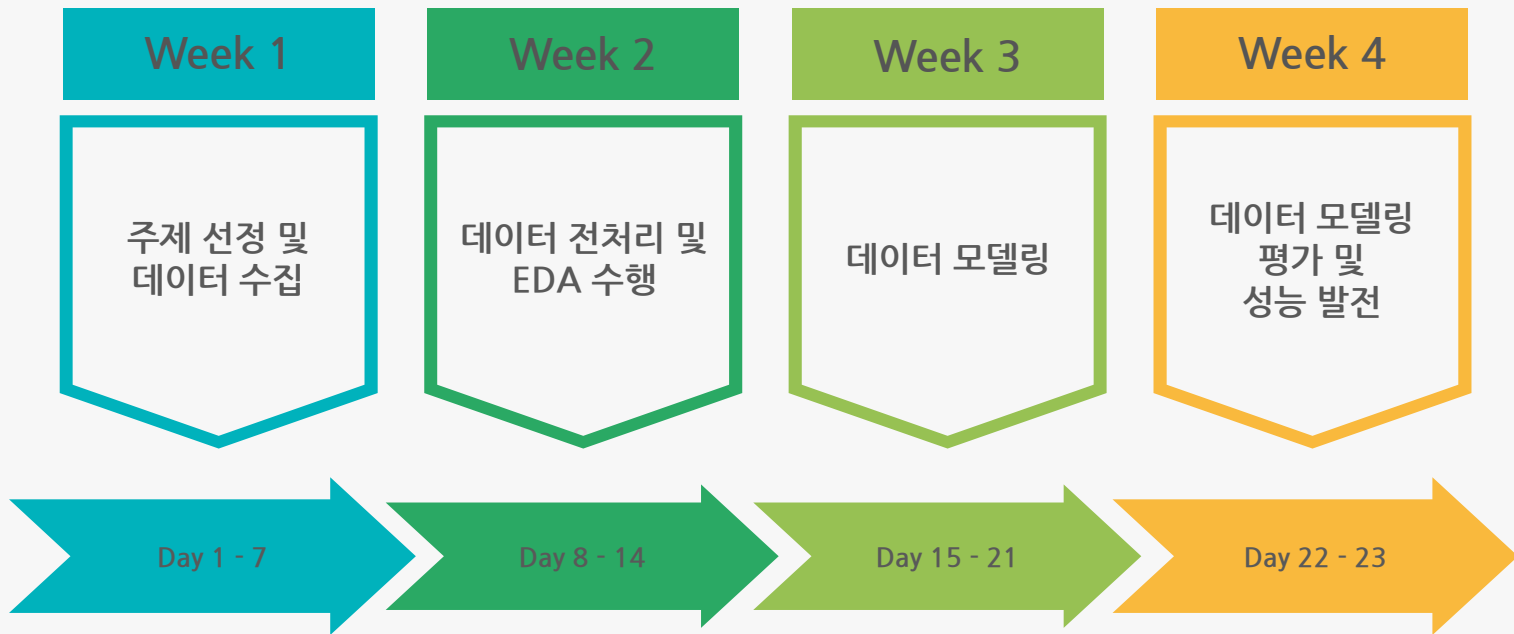
# 내 옷 좀 추천해 조!

의류 실측 데이터 기반 프리사이즈 분석 및  
개인화 의류 추천 시스템

# 프로젝트 정보

기간 2023.06.01 ~ 2023.06.23 (약 4주)

팀원 김보석, 박미영, 박성호, 이성희, 이정빈, 정설령, 최은욱 (7명)





## CONTENTS

- ❑ 프로젝트 주제
- ❑ 프로젝트 배경
- ❑ 프로젝트 진행
- ❑ 데이터 소개
- ❑ 하의 데이터 제외
- ❑ 데이터 전처리
- ❑ EDA
- ❑ 모델
  - ❑ 방법론
  - ❑ 분류 모델
  - ❑ 추천 모델
- ❑ 모델 구현 결과
- ❑ 결론

# 프로젝트 주제

“ 의류 실측 데이터 기반

프리스아이즈 분석 및 개인화 의류 추천 시스템 ”

# 프로젝트 배경 - 온라인 의류 산업의 지속적인 성장

## □ 통계청에 따른 온라인 의류 거래액

□ 2017년 : 약 11조 8,000억 원

□ 2019년 : 약 14조 4,000억 원 - 2년 동안 120% 이상 증가했음



▲ 통계청 「온라인쇼핑동향조사」에서 2017년부터 2019년까지 온라인 쇼핑몰 상에서 의복 거래액이 지속적으로 증가함을 확인할 수 있다. (출처 : 통계청 홈페이지)

# 프로젝트 배경 - 프리사이즈 실태

- ❑ 프리사이즈라는 one 사이즈의 기준으로 판매되고 있는 남녀 의류가 많지만 고객들이 본인 체형에 적합한 상품인지 아닌지를 확인하기 어려움.
- ❑ 프리사이즈를 재정의된 규격 사이즈로의 분류가 필요하다고 판단함.

이화여대 김. 씨(22)는 “상세 치수로는 사이즈가 맞을지 감이 오지 않

아 상 경기 고양시에 사는 고등학생 김. 군(19)은 불명확한 기준으로 인해

서 김 사이즈를 선택하기까지 오랜 시간이 걸린다. “같은 M 사이즈라고 마음 놓

고 구 이화여대 법학전문대학원의 김. 씨(23)는 사이즈 추천기능과 정확한  
요.” 정보를 충분히 제공하는 특정 쇼핑몰에서는 마음 놓고 구매한다. “소비자

가 이 그러나 상세정보마저 확실하지 않은 경우가 있다. 이화여대 엄. 씨  
생각하 (25)는 온라인에서 옷을 구매했지만 상세정보가 실제와 달라 입지 못한

경험이 있다.

# 프로젝트 배경 - 목표

- 고객의 더 나은 선택과 편의를 위해 무신사 내 상의를 대상으로 특성, 사이즈, 리뷰 데이터 등 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



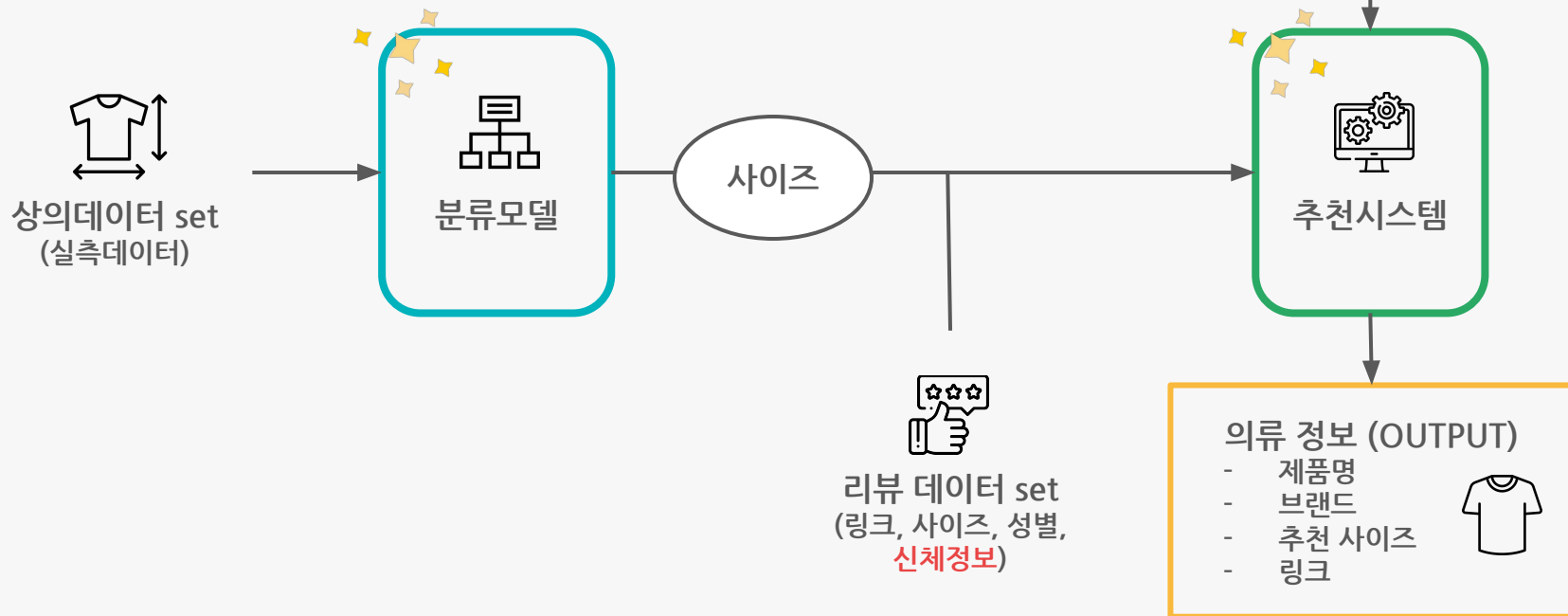
사용자들의 리뷰데이터를 이용한  
개인화 의류 추천시스템

# 프로젝트 진행





# 프로젝트 진행



# 데이터 소개

□ 온라인 쇼핑몰 ‘무신사’에서 의류 제품 정보와 고객 리뷰 수집

상의		리뷰
제품정보	사이즈정보	“보통”
제품명 브랜드 성별 링크 태그	총장 어깨너비 가슴단면 소매길이	성별 키 몸무게

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

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

스타일 후기 (4,028) | 상품 후기 (3,755) | 일반 후기 (11,098)

EVENT 스타일후기 작성 시 2,000원의 적립금을 드립니다.(직접 촬영한 사진 사진 필수)

추천 순

여성 · 165cm · 54kg 신고

★★★★★

사이즈 보통이에요 | 핏기 보통이에요 | 색감 보통이에요 | 두께감 보통이에요

Product Info 제품정보

브랜드 / 품번 COVERNA / CO2302STE1BK

성별 남, 여

조희수(1개월) 32.2만 회 이상

누적판매(1년) 10.9만 개 이상 (결제완료-반품)

좋아요 157,254

구매 후기 4.9 후기 45,309개 보기

#반팔티셔츠 #티셔츠 #오버핏반팔 #썸머컬렉션 #로고티셔츠

#데일리라이템 #투팩티 #투팩 #23S/S정규컬렉션 #에센셜라인

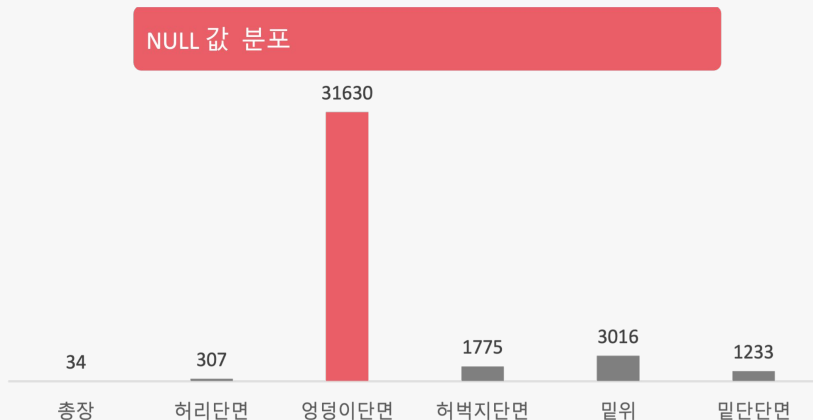
# 하의 데이터 제외

상의/하의 데이터 모두 수집했으나, 아래와 같은 이유로 분석 대상에서 제외

- ❑ 상의 데이터에 비해 하의 제품 데이터 개수가 현저히 적음
- ❑ 하의 실측 데이터 중 단면(허리, 엉덩이, 허벅지) 치수가 둘레로 혼용되어 표기된 상품이 있음을 확인  
=> 일관성 없는 데이터로 모델 구현 시, 분류 모델의 정확성이 낮을 수 있다고 판단
- ❑ 또한, 엉덩이 단면의 전체 데이터(67,523개) 중 31,630개가 NULL값으로 나타남

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>



# 데이터 전처리

## 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

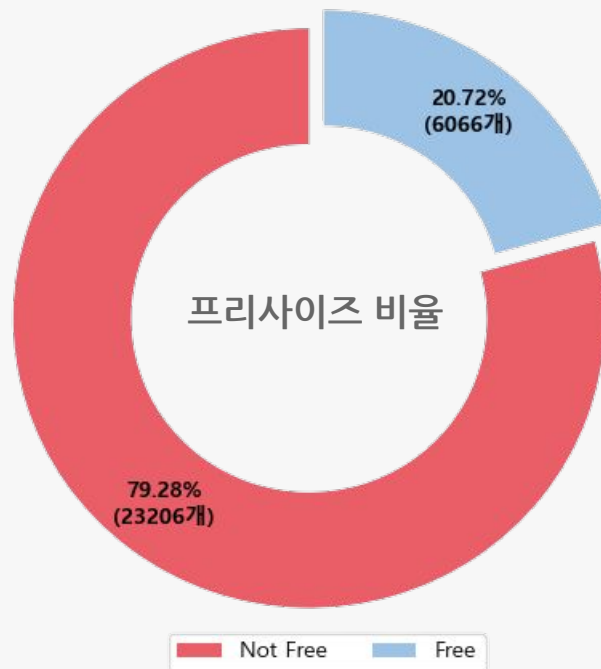
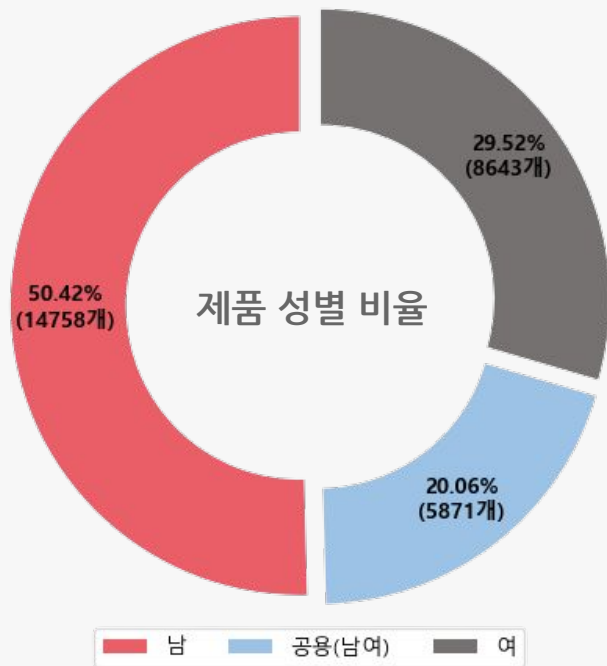
### EDA 수행

### 이상치 제거

- IQR을 이용한 이상치 제거

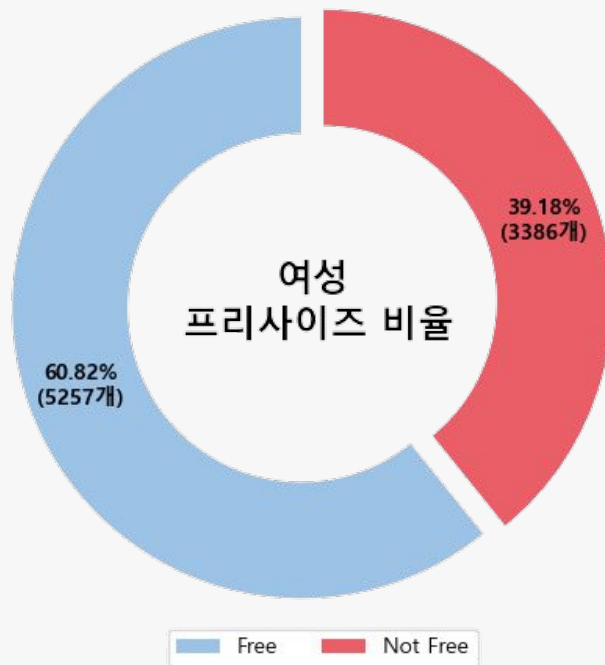
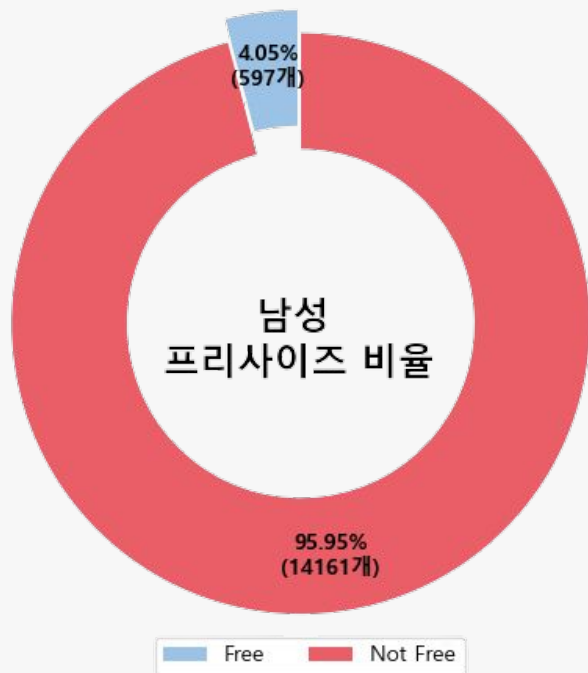
# EDA

## 데이터 분포 확인



# EDA

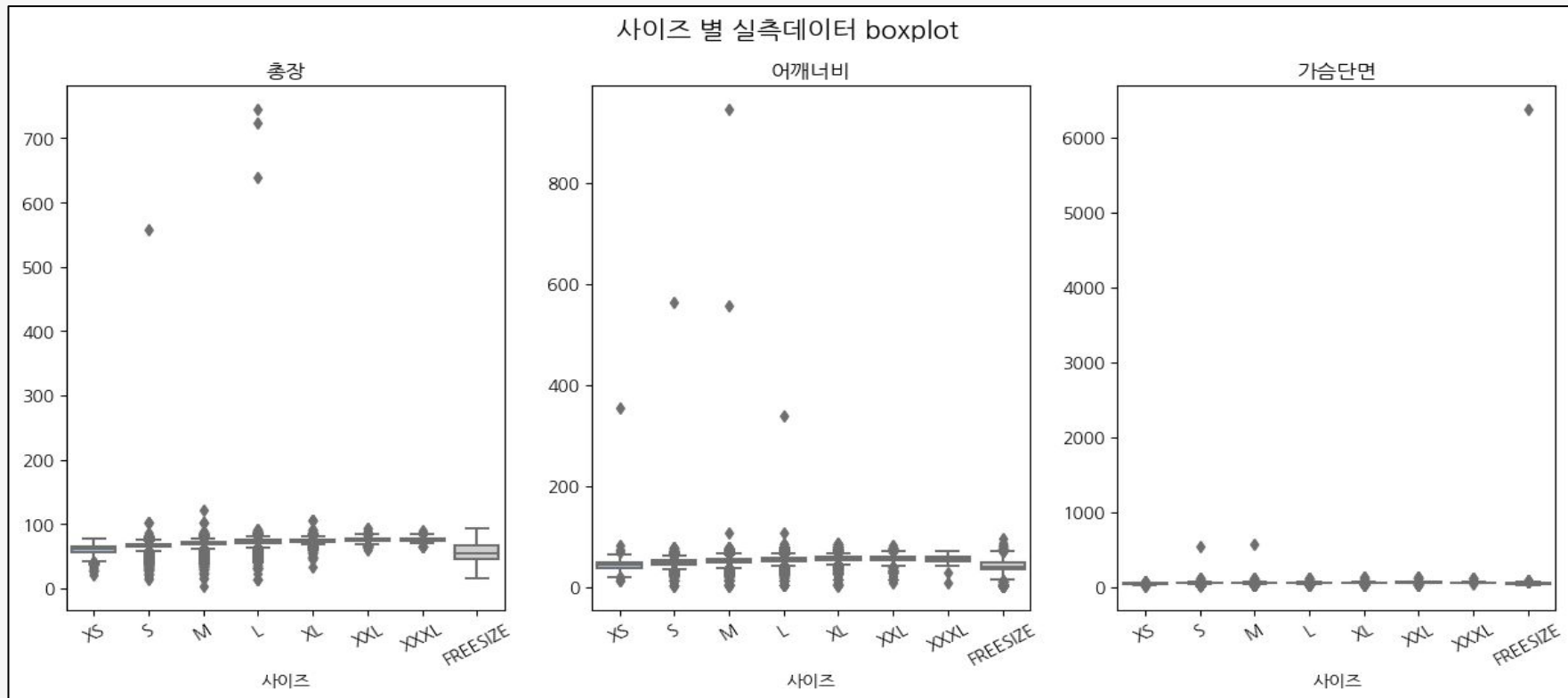
## □ 성별에 따른 프리사이즈 비율



# EDA

## 이상치 제거 (전) - 전체

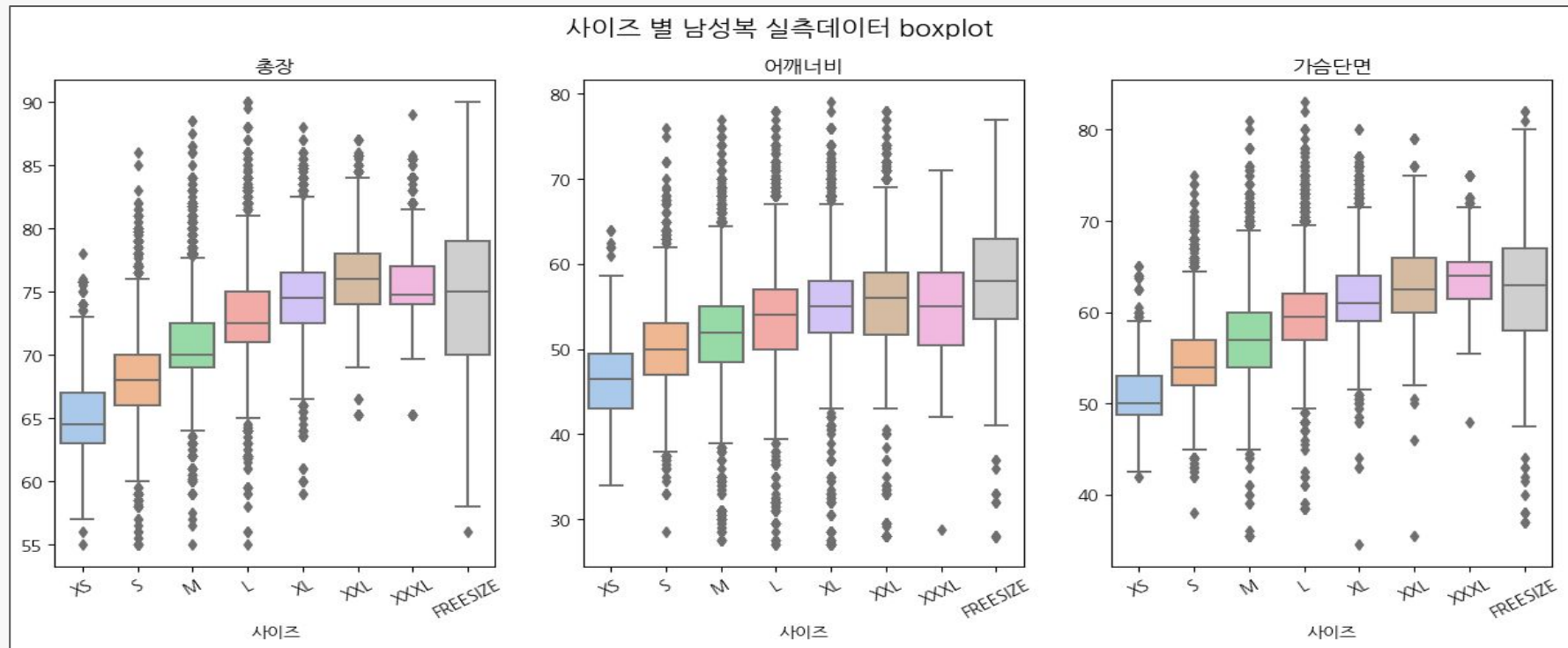
전체 데이터 개수 : 88636



# EDA

## ❑ 이상치 제거 (후) - 남성

이상치 제거 후 전체 데이터 개수 : 87917



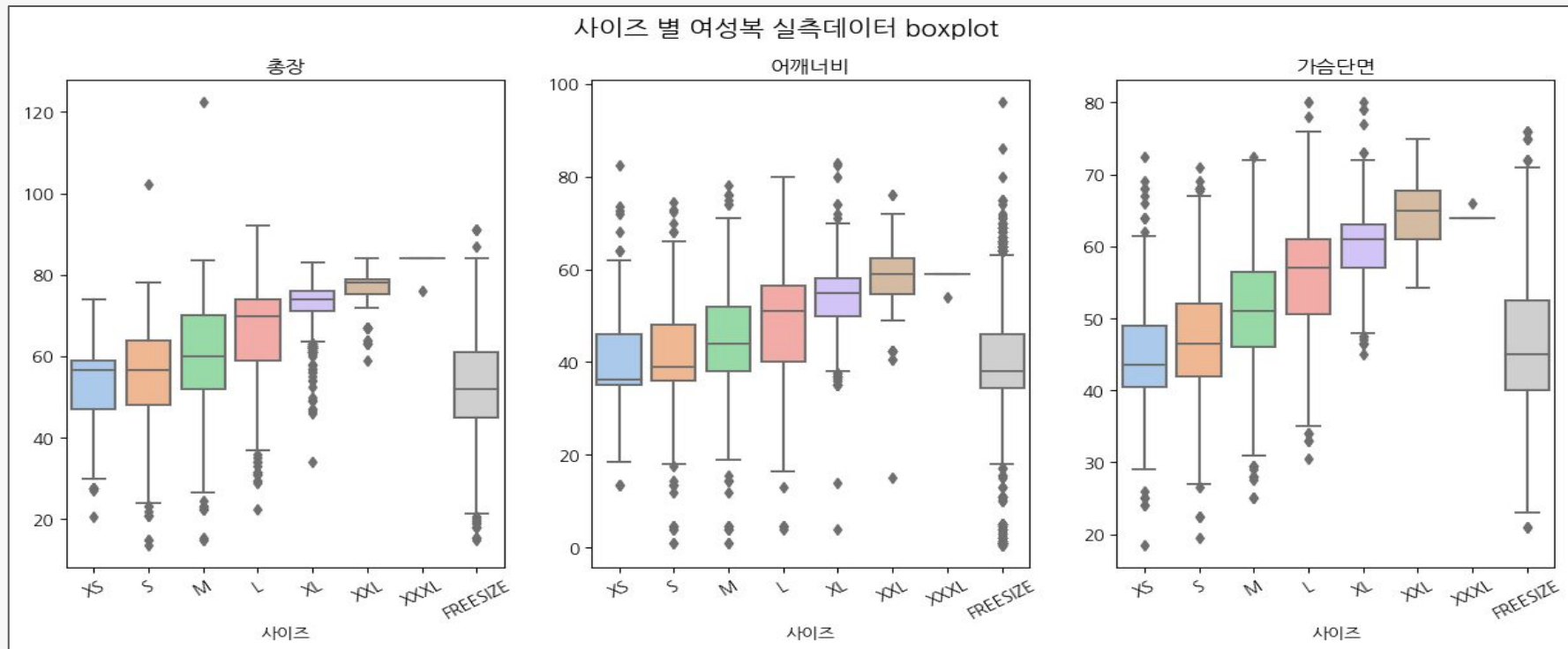
\* 이상치는 IQR 3배를 기준으로 설정



# EDA

## 이상치 제거 (후) - 여성

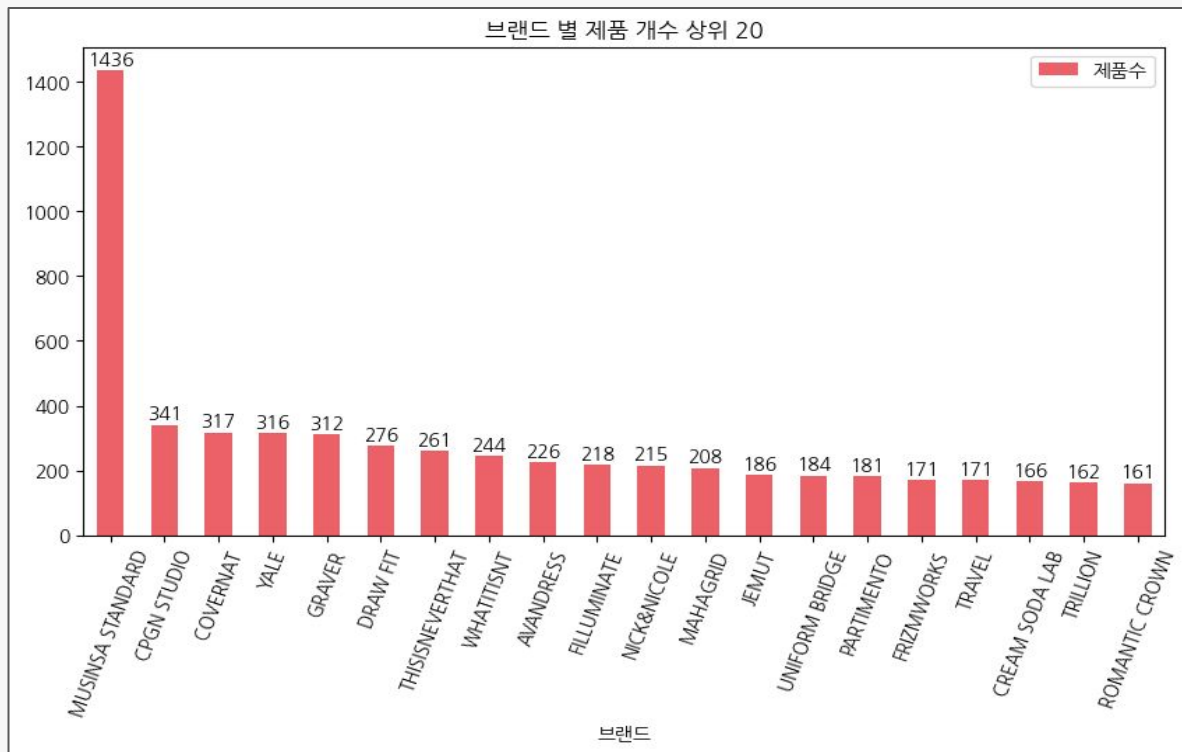
이상치 제거 후 전체 데이터 개수 : 87917



\* 이상치는 IQR 3배를 기준으로 설정

# EDA

## ❑ 제품 개수 상위 20위 브랜드

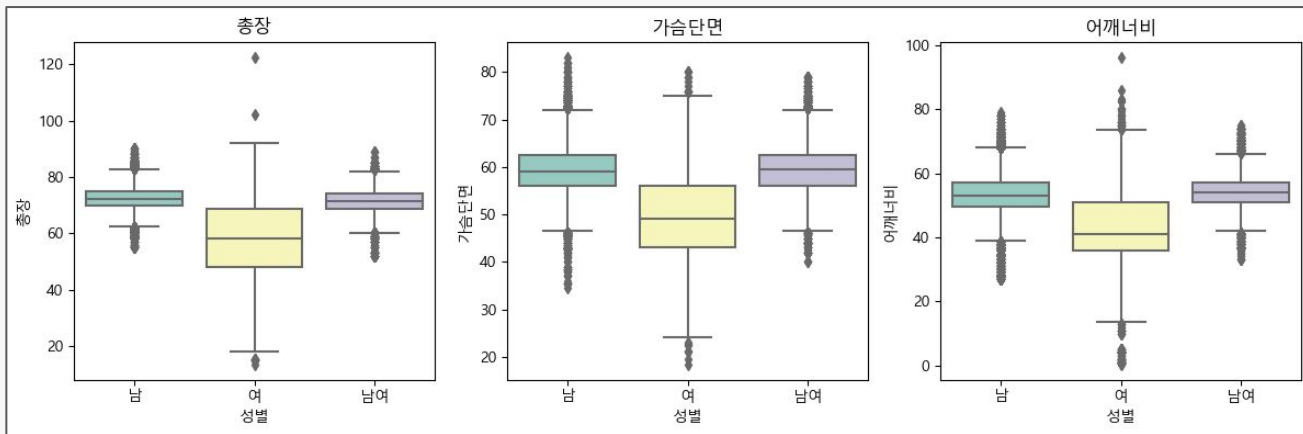


❑ 일관된 사이즈 기준 정립을 위해 특정 브랜드 사이즈 기준에 따르기로 결정

❑ 제품 개수가 가장 많은 MUSinsa STANDARD 브랜드로 선정

# EDA

- ❑ '남여(공용)' 데이터를 '남성' 데이터에 포함

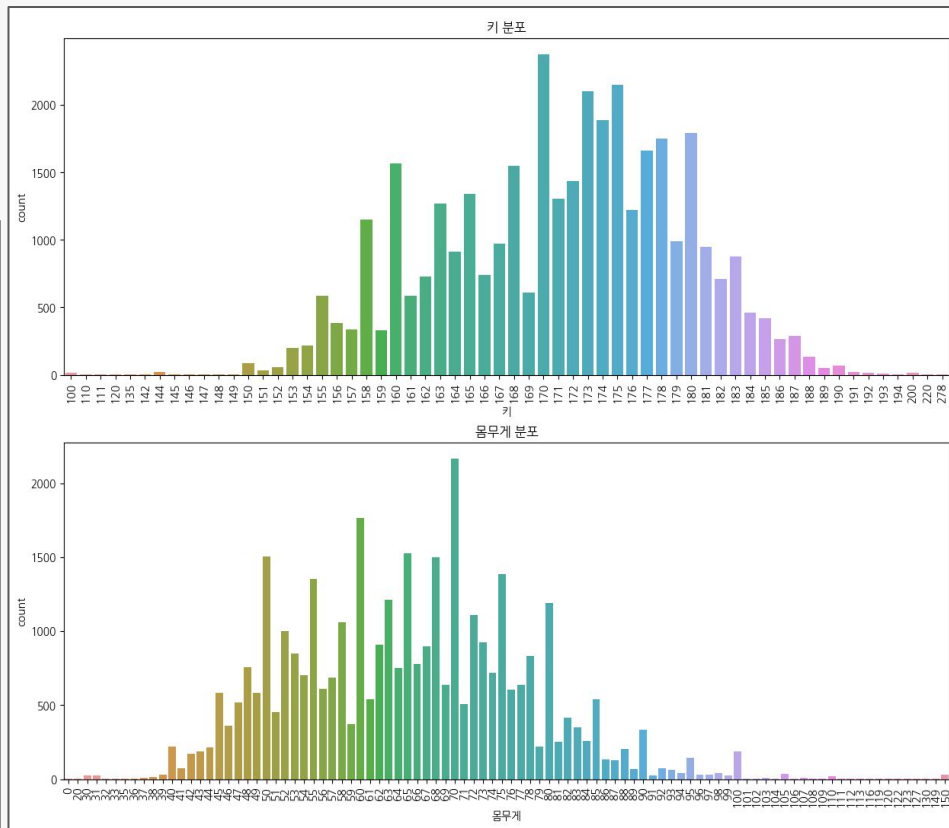
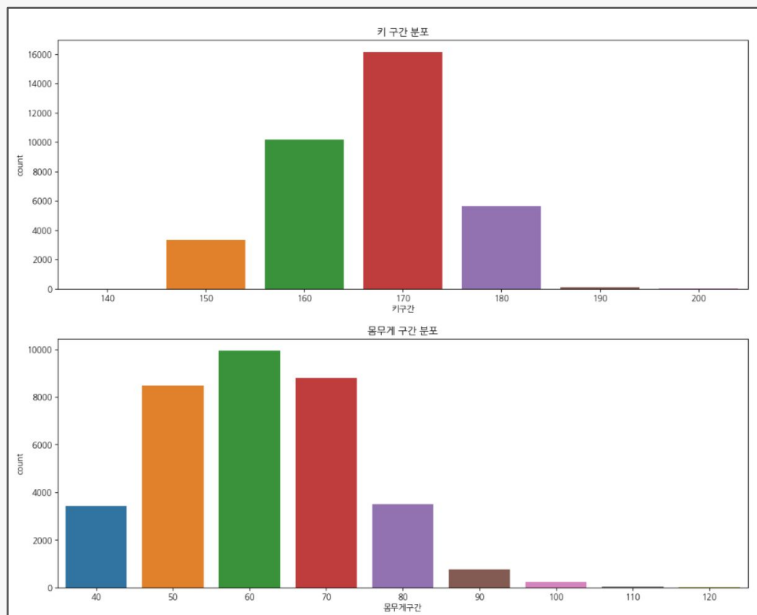


단위 (cm)	평균			표준편차			중앙값		
	총장	가슴단면	어깨너비	총장	가슴단면	어깨너비	총장	가슴단면	어깨너비
남성	71.41	59.44	53.90	4.38	5.16	5.32	71.50	59.50	54.00
남여	72.42	59.15	53.41	4.25	5.14	6.06	72.00	59.00	53.00
여성	58.08	49.48	43.36	11.82	8.97	10.53	58.00	49.00	41.00

# EDA

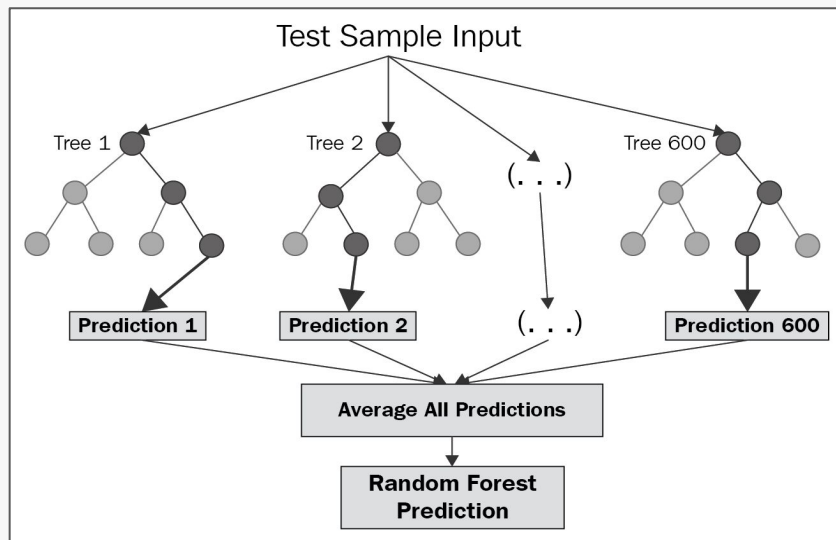
## 리뷰 작성자들의 체형 데이터 이상치 제거

- ▶ 키 : 150 cm 미만 200 cm 이상
- ▶ 몸무게 : 40 kg 미만 120 kg 초과



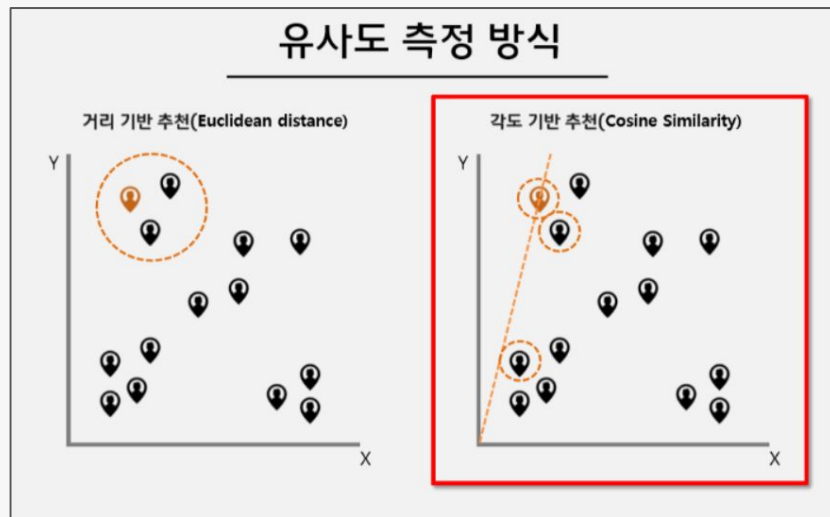
# 모델 - 방법론

- ❑ Random Forest Model (랜덤 포레스트 모델)
  - ❑ 과대적합을 방지하기 위해, 최적의 기준 변수를 랜덤하게 생성하는 머신러닝 기법
  - ❑ 다수의 결정 트리들을 학습하는 앙상블 모델



# 모델 - 방법론

- Contents based Model (콘텐츠 기반 모델)
  - 사용자가 이전에 구매한 상품 중에서 좋아하는 상품과 유사한 상품을 추천하는 방법
  - 유사도 함수(Euclidean Distance, Cosine, Pearson 등) 이용
  - TF-IDF, Word2Vec 등 활용



# 모델 - 추천시스템 설계

## 분류 모델

### Random Forest Model (랜덤 포레스트 모델)

- MUSinsa STANDARD의 실측 데이터(총장, 가슴단면, 어깨너비)로 사이즈 분류 시행

## 추천 모델

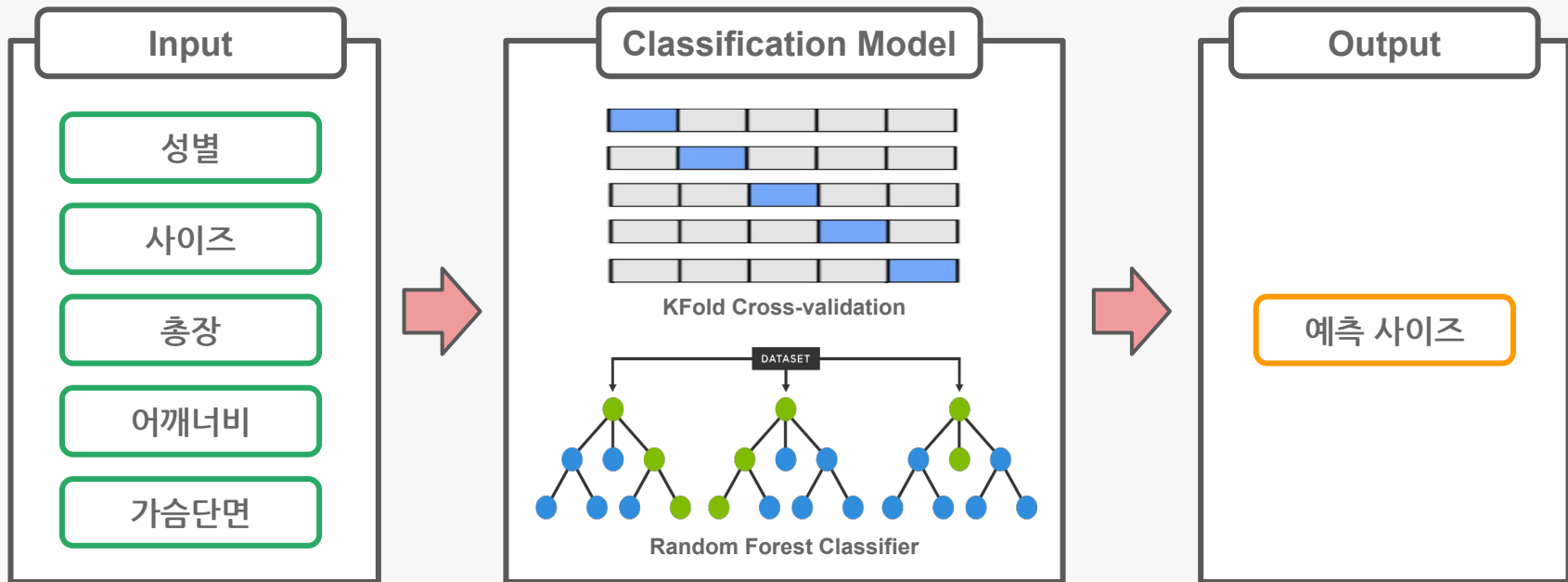
### Cosine Similarity (코사인 유사도)

- 사용자의 특성(키, 몸무게, 성별)을 통하여 가장 유사한 상위 20개 제품 추천
- TF-IDF로 얻어진 태그 벡터의 유사도를 이용해 최종 제품 추천

### TF-IDF (단어 빈도-역 문서 빈도)

- 한국어 형태로 된 태그를 벡터 형태로 표현하기 위해 사용

# 모델 - 분류 모델



출처

1) [http://www.incodom.kr/k-검\\_교차\\_검증](http://www.incodom.kr/k-검_교차_검증)

2) <https://www.tibco.com/ko/reference-center/what-is-a-random-forest>



# 모델 - 분류 모델 (모델 선택)

## < 남성 데이터 >

	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
Model					
ExtraTreesClassifier	0.99	0.99	None	0.99	0.09
LGBMClassifier	0.99	0.99	None	0.99	0.12
RandomForestClassifier	0.98	0.98	None	0.98	0.12
ExtraTreeClassifier	0.98	0.98	None	0.98	0.01
DecisionTreeClassifier	0.98	0.98	None	0.98	0.00
BaggingClassifier	0.98	0.98	None	0.98	0.01
XGBClassifier	0.94	0.94	None	0.94	0.19
LabelPropagation	0.89	0.89	None	0.89	0.03
LabelSpreading	0.88	0.88	None	0.88	0.03
KNeighborsClassifier	0.85	0.85	None	0.85	0.01
NuSVC	0.67	0.67	None	0.67	0.06
AdaBoostClassifier	0.63	0.63	None	0.63	0.07
SVC	0.54	0.54	None	0.54	0.05
SGDClassifier	0.45	0.46	None	0.39	0.03
RidgeClassifier	0.42	0.44	None	0.34	0.02

## < 여성 데이터 >

	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
Model					
ExtraTreeClassifier	1.00	1.00	None	1.00	0.01
BaggingClassifier	1.00	1.00	None	1.00	0.05
ExtraTreesClassifier	1.00	1.00	None	1.00	0.18
RandomForestClassifier	0.99	1.00	None	0.99	0.21
DecisionTreeClassifier	0.99	1.00	None	0.99	0.01
LGBMClassifier	0.99	0.99	None	0.99	0.27
KNeighborsClassifier	0.97	0.97	None	0.97	0.02
XGBClassifier	0.95	0.95	None	0.95	0.57
LabelPropagation	0.92	0.92	None	0.92	0.32
LabelSpreading	0.91	0.91	None	0.91	0.45
SVC	0.75	0.75	None	0.75	0.45
NuSVC	0.72	0.72	None	0.72	0.89
CalibratedClassifierCV	0.53	0.52	None	0.51	0.90
LogisticRegression	0.50	0.49	None	0.49	0.04
QuadraticDiscriminantAnalysis	0.46	0.47	None	0.45	0.04

## 모델 - 분류 모델 (파라미터 선택)

남성 ▶

max_depth	max_leaf_nodes	min_samples_leaf	min_sample_split	n_estimators	score
8	80	13	52	200	0.906776
30	40	15	52	100	0.906560
20	40	15	52	100	0.906560
15	40	10	52	100	0.906343
30	40	10	52	100	0.906343

여성 ▶

max_depth	max_leaf_nodes	min_samples_leaf	min_sample_split	n_estimators	score
10	80	5	10	150	0.900207
10	90	5	5	150	0.900207
10	60	5	5	150	0.900207
10	70	5	5	150	0.900207
10	80	5	10	100	0.900200

# 모델 - 분류 모델 (평가)

< 남성 데이터 >

		Confusion Matrix					
		S	M	L	XL	2XL	3XL
Actual	S	197	3	2	0	0	5
	M	7	114	0	0	0	2
	L	7	0	182	6	1	15
	XL	7	0	8	168	17	11
	2XL	0	0	5	6	175	0
	3XL	13	5	6	2	0	184
		Predicted					

< 여성 데이터 >

		Confusion Matrix			
		XS	S	M	L
Actual	XS	66	4	1	3
	S	3	70	2	1
	M	2	5	67	2
	L	2	2	0	69
		Predicted			

# 모델 - 분류 모델 (평가)

< 남성 데이터 >

	precision	recall	f1-score	support
2XL	0.853	0.952	0.900	207
3XL	0.934	0.927	0.931	123
L	0.897	0.863	0.879	211
M	0.923	0.796	0.855	211
S	0.907	0.941	0.923	186
XL	0.848	0.876	0.862	210
accuracy			0.889	1148
macro avg	0.894	0.892	0.892	1148
weighted avg	0.890	0.889	0.888	1148

< 여성 데이터 >

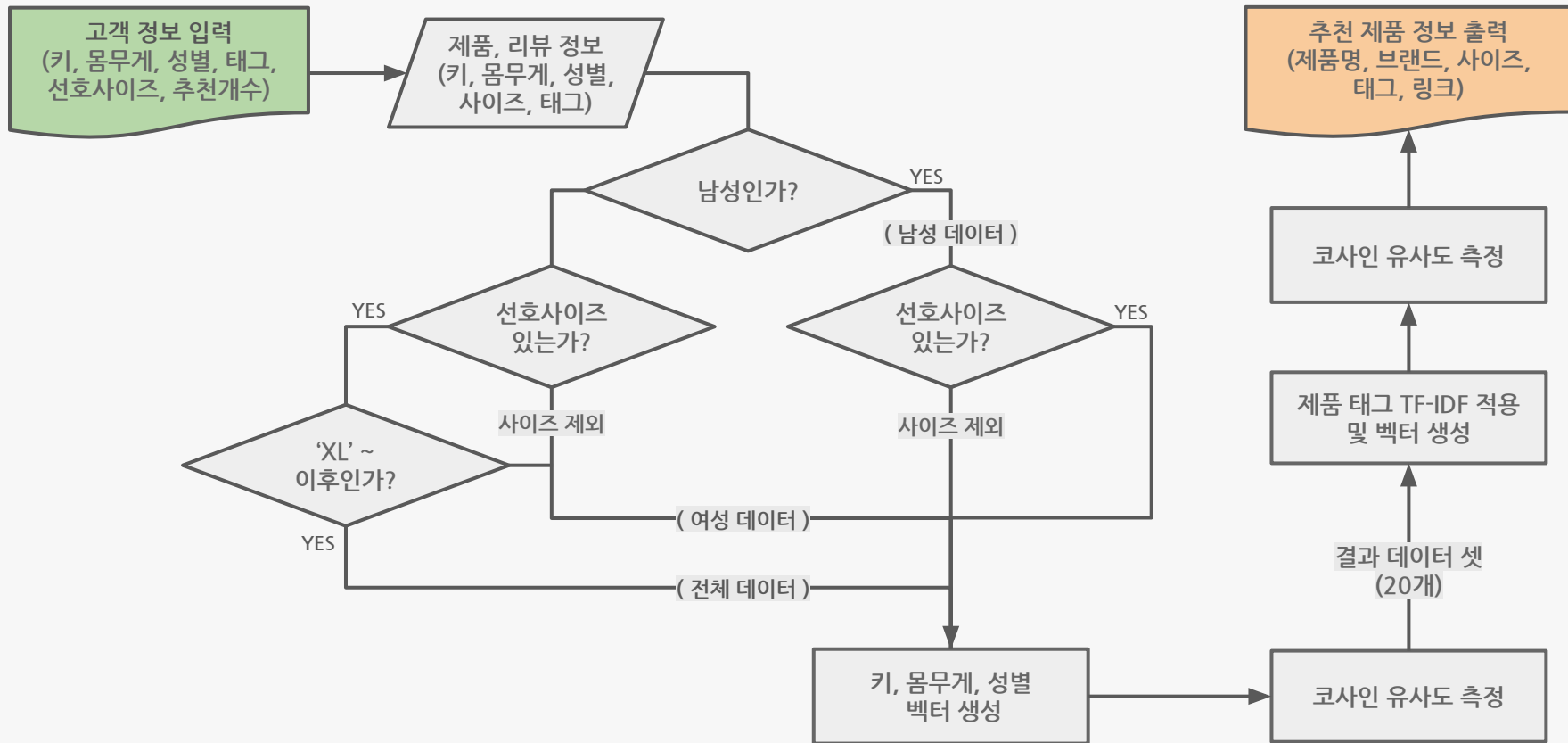
	precision	recall	f1-score	support
L	0.904	0.892	0.898	74
M	0.864	0.921	0.892	76
S	0.957	0.882	0.918	76
XS	0.920	0.945	0.932	73
accuracy			0.910	299
macro avg	0.911	0.910	0.910	299
weighted avg	0.911	0.910	0.910	299

## 모델 - 분류 모델 (결과)

## □ 분류 모델 결과 데이터 형태

	제품명	브랜드	제품 성별	사이즈	총장	어깨너비	가슴단면	링크	FREESIZE	pred_size	유지 성별	키	몸무 게	태그
0	에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남	M	69.0	51.0	55.0	https://www.musinsa.com/app/goods/1848166	0.0	M	남성	161	72	#반팔티셔츠 #티셔츠 #오버핏반팔 #섬머컬렉션 #로그ोट이셔츠 #데일리아이템 #투팩티 ...
1	에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남	S	65.0	49.0	51.5	https://www.musinsa.com/app/goods/1848166	0.0	S	남성	172	61	#반팔티셔츠 #티셔츠 #오버핏반팔 #섬머컬렉션 #로그ोट이셔츠 #데일리아이템 #투팩티 ...
2	에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남	M	69.0	51.0	55.0	https://www.musinsa.com/app/goods/1848166	0.0	M	여성	175	53	#반팔티셔츠 #티셔츠 #오버핏반팔 #섬머컬렉션 #로그ोट이셔츠 #데일리아이템 #투팩티 ...
3	에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남	S	65.0	49.0	51.5	https://www.musinsa.com/app/goods/1848166	0.0	S	여성	162	56	#반팔티셔츠 #티셔츠 #오버핏반팔 #섬머컬렉션 #로그ोट이셔츠 #데일리아이템 #투팩티 ...
4	에센셜 쿨 코튼 2-PACK 티셔츠 블랙	COVERNAT	남	S	65.0	49.0	51.5	https://www.musinsa.com/app/goods/1848166	0.0	S	남성	174	63	#반팔티셔츠 #티셔츠 #오버핏반팔 #섬머컬렉션 #로그ोट이셔츠 #데일리아이템 #투팩티 ...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
29575	pride ball s/s tee 8% melange	UNIFORM BRIDGE	남	L	72.0	52.0	61.0	https://www.musinsa.com/app/goods/3239084	0.0	2XL	남성	175	75	#NAVYSPORTS #반팔티 #티셔츠 #여름반팔티 #그래픽티셔츠 #23SUMMER
29576	pride ball s/s tee 8% melange	UNIFORM BRIDGE	남	XL	74.0	54.0	63.5	https://www.musinsa.com/app/goods/3239084	0.0	3XL	남성	175	80	#NAVYSPORTS #반팔티 #티셔츠 #여름반팔티 #그래픽티셔츠 #23SUMMER
29577	pride ball s/s tee 8% melange	UNIFORM BRIDGE	남	L	72.0	52.0	61.0	https://www.musinsa.com/app/goods/3239084	0.0	2XL	여성	163	54	#NAVYSPORTS #반팔티 #티셔츠 #여름반팔티 #그래픽티셔츠 #23SUMMER
29578	pride ball s/s tee 8% melange	UNIFORM BRIDGE	남	XL	74.0	54.0	63.5	https://www.musinsa.com/app/goods/3239084	0.0	3XL	남성	180	75	#NAVYSPORTS #반팔티 #티셔츠 #여름반팔티 #그래픽티셔츠 #23SUMMER
29579	pride ball s/s tee 8% melange	UNIFORM BRIDGE	남	XL	74.0	54.0	63.5	https://www.musinsa.com/app/goods/3239084	0.0	3XL	남성	178	74	#NAVYSPORTS #반팔티 #티셔츠 #여름반팔티 #그래픽티셔츠 #23SUMMER
29580 rows × 14 columns														

## 모델 - 추천 모델



# 모델 구현 결과 - 동영상

추천받을 카테고리를 입력하세요.

:

카테고리 → 반팔, 기본, 무지, 로고, 그래픽, 셔츠, 오버핏, 니트, 카라, 루즈핏  
사이즈 → XS(여성전용사이즈), S, M, L, XL, 2XL, 3XL

이름 : 은욱이

성별 : 남자

키 : 184

몸무게 : 75

좋아하는 스타일 : 셔츠

선호사이즈 : 없음



# 모델 구현 결과 - 입력 예시

카테고리 → 반팔, 기본, 무지, 로고, 그래픽, 셔츠, 오버핏, 니트, 카라, 루즈핏

사이즈 → XS(여성전용사이즈), S, M, L, XL, 2XL, 3XL

추천받을 카테고리 입력하세요. : 로고  
고객님의 키(cm)를 입력하세요. : 159  
고객님의 몸무게(kg)를 입력하세요. : 50  
고객님의 성별(남성, 여성)을 입력하세요. : 여성  
고객님의 평소 사이즈를 입력하세요.(없으면 x) : S  
추천받을 상품 개수를 입력하세요. : 7

카테고리 → 반팔, 기본, 무지, 로고, 그래픽, 셔츠, 오버핏, 니트, 카라, 루즈핏

사이즈 → XS(여성전용사이즈), S, M, L, XL, 2XL, 3XL

추천받을 카테고리 입력하세요. : 그래픽  
고객님의 키(cm)를 입력하세요. : 175  
고객님의 몸무게(kg)를 입력하세요. : 70  
고객님의 성별(남성, 여성)을 입력하세요. : 남성  
고객님의 평소 사이즈를 입력하세요.(없으면 x) : L  
추천받을 상품 개수를 입력하세요. : 5

카테고리 → 반팔, 기본, 무지, 로고, 그래픽, 셔츠, 오버핏, 니트, 카라, 루즈핏

사이즈 → XS(여성전용사이즈), S, M, L, XL, 2XL, 3XL

추천받을 카테고리 입력하세요. : 반팔  
고객님의 키(cm)를 입력하세요. : 173  
고객님의 몸무게(kg)를 입력하세요. : 70  
고객님의 성별(남성, 여성)을 입력하세요. : 여성  
고객님의 평소 사이즈를 입력하세요.(없으면 x) : XL  
추천받을 상품 개수를 입력하세요. : 8



# 모델 구현 결과 - 사례 1

이름 : 미영이

성별 : 여자

키 : 159

몸무게 : 50

좋아하는 스타일 : 로고 티셔츠

선호사이즈 : S



< 카테고리 > [ 로고 ]

키 : 159cm, 몸무게 : 50kg, 성별 : 여성

고객님께 추천드리는 제품입니다.

< 1. PRINTING WOMAN T-SHIRTS GREEN >

브랜드 : REORG

제품성별 : 여

선택사이즈 : S

추천사이즈 : 옵션없음

카테고리 : #로고 #로고티셔츠 #반팔 #반팔티 #빅로고 #티셔츠

링크 : <https://www.musinsa.com/app/goods/552670>

< 2. [패키지] 크롭+레귤러핏 반팔 티셔츠 5종 모음 >

브랜드 : 5OML

제품성별 : 여

선택사이즈 : S

추천사이즈 : M

카테고리 : #반팔 #반팔티 #티셔츠 #로고티셔츠 #여자반팔티 #여름반팔티 #반소매

링크 : <https://www.musinsa.com/app/goods/3190836>

< 3. Layered Tube Top T-shirt (White) >

브랜드 : THREE TO EIGHTY

제품성별 : 여

선택사이즈 : S

추천사이즈 : 옵션없음

카테고리 : #여름반팔 #여름반팔티 #티셔츠 #반팔티셔츠 #로고티셔츠 #반팔티 #그래픽티셔츠

링크 : <https://www.musinsa.com/app/goods/2640549>

## 모델 구현 결과 - 사례 2

< 카테고리 > [ 그래픽 ]

키 : 175cm, 몸무게 : 70kg, 성별 : 남성  
고객님께 추천드리는 제품입니다.

< 1. SSSOCIETY HALF SLEEVE T-SHIRTS (WHITE) >

브랜드 : JUAN HOMME  
제품성별 : 남  
선택사이즈 : L  
추천사이즈 : S  
카테고리 : #로고반팔티 #여름반팔티 #오버핏반팔 #그래픽 #그래픽티셔츠 #남자반팔티 #반팔티셔츠 #티셔츠컬렉션 #23SUMMER  
링크 : <https://www.musinsa.com/app/goods/3197001>

< 2. 스크림 로고 티셔츠\_블랙(NG2DMUT504A) >

브랜드 : NICE GHOST CLUB  
제품성별 : 남  
선택사이즈 : L  
추천사이즈 : L  
카테고리 : #반팔티 #반팔티셔츠 #그래픽 #반팔 #티셔츠 #오버핏 #로고티셔츠  
링크 : <https://www.musinsa.com/app/goods/2424785>

< 3. 라운드 플라워 티셔츠\_블랙 >

브랜드 : ROMANTIC CROWN  
제품성별 : 남  
선택사이즈 : L  
추천사이즈 : L  
카테고리 : #반팔티 #반팔 #로고 #로고티셔츠 #그래픽 #오버핏 #우지컬렉션  
링크 : <https://www.musinsa.com/app/goods/3183139>

이름 : 보석이

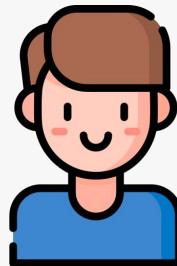
성별 : 남자

키 : 175

몸무게 : 70

좋아하는 스타일 : 그래픽 디자인

선호사이즈 : L



# 모델 구현 결과 - 사례 3

이름 : 설희

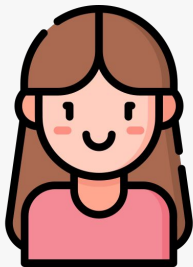
성별 : 여자

키 : 173

몸무게 : 70

좋아하는 스타일 : 반팔 티셔츠

선호사이즈 : XL



< 카테고리 > [ 반팔 ]

키 : 173cm, 몸무게 : 70kg, 성별 : 여성  
고객님께 추천드리는 제품입니다.

< 1. Diagonal T-shirt WHITE >

브랜드 : AVANDRESS  
제품성별 : 여  
선택사이즈 : XL  
추천사이즈 : FREE  
카테고리 : #여자반팔티 #반팔티추천 #반팔티 #반팔티셔츠 #여름반팔티 #여름반팔 #우먼라인  
링크 : <https://www.musinsa.com/app/goods/2479523>

< 2. DN BOARD TEE(BLACK) >

브랜드 : LUV IS TRUE  
제품성별 : 여  
선택사이즈 : XL  
추천사이즈 : FREE  
카테고리 : #반팔 #반팔티추천 #반팔티 #반소매 #반팔티셔츠 #티셔츠  
링크 : <https://www.musinsa.com/app/goods/3067154>

< 3. (23SS)SMILE CAT 반팔 티셔츠 화이트 >

브랜드 : AWESOMEUS  
제품성별 : 남  
선택사이즈 : XL  
추천사이즈 : M  
카테고리 : #반팔티 #23S/S컬렉션 #반팔 #로고티셔츠 #여자반팔티 #반팔티추천 #남자반팔티  
링크 : <https://www.musinsa.com/app/goods/3165247>

# 결론



## 분류 모델

- ❑ MUSINSA STANDARD의 실측 데이터를 기준으로 사이즈 분류 모델은 RandomForest로 선정하여 진행함
- ❑ 남성, 여성을 나누어 분류 모델을 진행하였고, Grid Search를 통하여 최적의 하이퍼 파라미터를 도출함
- ❑ 남성은 F1스코어 기준 3XL, S, 2XL 순으로 모델 성능이 좋았음
- ❑ 여성은 모델 성능평가 결과 XS, S, L 순 나타났음



## 추천 모델

- ❑ 제품 정보(키, 몸무게, 성별, 태그 등)를 벡터화시켜 코사인 유사도를 이용하여 추천시스템을 진행함 (콘텐츠 기반)
- ❑ 키, 몸무게, 성별, 선호 사이즈를 입력하면 유사도가 가장 높은 제품 N개 추천해주는 시스템을 제작함
- ❑ 남성, 여성 모두 사이즈를 기반한 카테고리 별 제품을 잘 보여주는 것으로 확인됨

# 결론

## □ 한계점

- 모델 비교 선택 시 모델 간 평가지표의 차이가 없어 과적합이 의심됨으로 객관적 판단이 어려움
- 데이터 개수의 부족으로 인해 극단 값들의 추천 결과가 아쉬움
- 고객 별로 추구하는 스타일이 달라서 사이즈 기반으로만 추천하기에는 정확도가 떨어져 보였음
- 추천 시스템을 객관적으로 평가하기 어려움

## □ 개선할 점

- 더 많은 리뷰 데이터를 수집하면 더 정확하고 다양한 제품 추천이 가능할 것으로 보임
- 규제 기법(Ridge, Lasso) 등을 활용하여 과적합을 줄이는 시도가 필요

**Thank  
You!**