

프로젝트 PROPOSAL

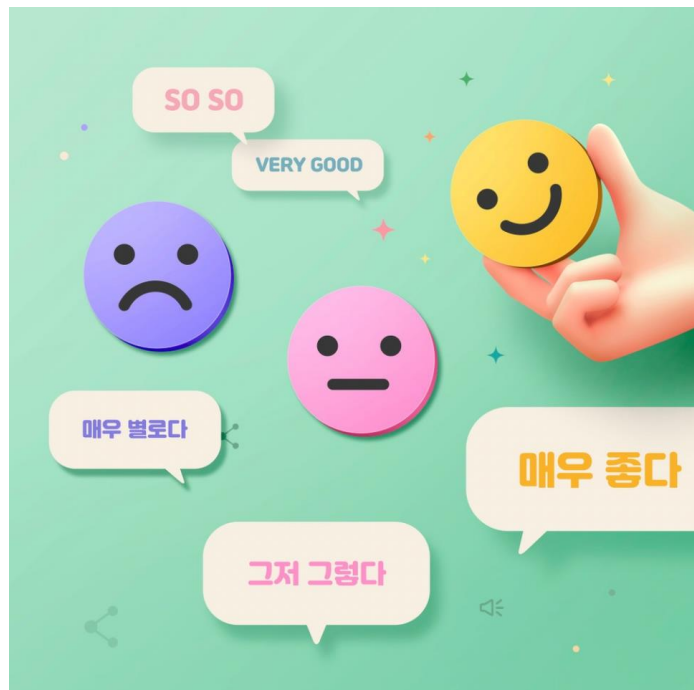
아마존 패션 데이터를 이용한 리뷰 생성

12조

김아영 고연우 김민균 박천수

주제 및 배경

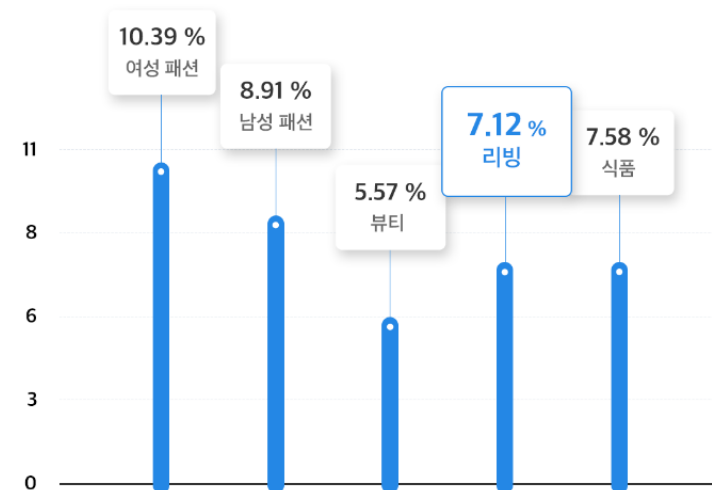
“리뷰 생성 모델로 소비자의 리뷰 작성을 돕자”



트렌드모니터 2014 소비자 리뷰 영향력 조사

통계청 2020 온라인 쇼핑 동향 조사

리뷰 작성률



리뷰 작성률 = (상품 리뷰수/주문 상품수) × 100

데이터 소개 및 수집 과정

Amazon Fashion



https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/

AMAZON FASHION의 reviews(883,636)과 metadata(186,637) 사용

데이터 소개 및 수집 과정

overall	verified	reviewTime	reviewerID	asin	reviewerName	reviewText	summary	unixReviewTime	vote	style	image	
0	5	True	10 20, 2014	A1D4G1SNUZWQOT	7106116521	Tracy	Exactly what I needed.	perfect replacements!!	1413763200	NaN	NaN	NaN
1	2	True	09 28, 2014	A3DDWDH9PX2YX2	7106116521	Sonja Lau	I agree with the other review, the opening is ...	I agree with the other review, the opening is ...	1411862400	3.0	NaN	NaN
2	4	False	08 25, 2014	A2MWC41EW7XL15	7106116521	Kathleen	Love these... I am going to order another pack...	My New 'Friends' !!	1408924800	NaN	NaN	NaN
3	2	True	08 24, 2014	A2UH2QQ275NV45	7106116521	Jodi Stoner	too tiny an opening	Two Stars	1408838400	NaN	NaN	NaN
4	3	False	07 27, 2014	A89F3LQADZBS5	7106116521	Alexander D.	Okay	Three Stars	1406419200	NaN	NaN	NaN
...
883631	5	True	02 21, 2017	A1ZSB2Q144UTEY	B01HJH5TH5U	Amazon Customer	I absolutely love this dress!! It's sexy and ...	I absolutely love this dress	1487635200	NaN	NaN	NaN
883632	5	True	11 25, 2016	A2CCDV0J5VB6F2	B01HJH5TH5U	Amazon Customer	I'm 5'6 175lbs. I'm on the tall side. I wear a...	I wear a large and ordered a large and it stil...	1480032000	2.0	NaN	NaN
883633	3	True	11 10, 2016	A3O90PACS7B61K	B01HJH5TH5U	Fabfifty	Too big in the chest area!	Three Stars	1478736000	NaN	NaN	NaN
883634	3	True	11 10, 2016	A2HO94I89U3LNH	B01HJHF97K	Mgomez	Too clear in the back, needs lining	Three Stars	1478736000	NaN	NaN	NaN
883635	5	True	08 9, 2016	A2RSX9E79DUHRX	B01HJG5NMW	Natasha Mascarenhas	Ordered and was slightly small. Worked with th...	The quality is excellent and it is so cute	1470700800	NaN	NaN	NaN

883636 rows × 12 columns

Review Data

- overall → 평점, 5점 척도 ★
- verified → 구매 인증 여부 ★
- reviewTime → 리뷰 작성 시간
- reviewerID → 리뷰 작성자 ★
- asin → 제품 코드 ★
- reviewerName → 리뷰 작성자 이름
- reviewText → 리뷰 내용 ★
- summary → 리뷰 내용 요약
- unixReviewTime → unix버전 리뷰 작성 시간
- vote → 리뷰 추천 개수
- style → 제품 특징
- Image → 제품 이미지

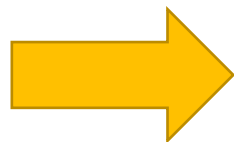
데이터 소개 및 수집 과정

	overall	verified	reviewTime	reviewerID	asin	reviewText	year
0	5	True	2014-10-20	A1D4G1SNUZWQOT	7106116521	Exactly what I needed.	2014
1	2	True	2014-09-28	A3DDWDH9PX2YX2	7106116521	I agree with the other review, the opening is ...	2014
2	4	False	2014-08-25	A2MWC41EW7XL15	7106116521	Love these... I am going to order another pack...	2014
3	2	True	2014-08-24	A2UH2QQ275NV45	7106116521	too tiny an opening	2014
4	3	False	2014-07-27	A89F3LQADZBS5	7106116521	Okay	2014

...
883631	5	True	2017-02-21	A1ZSB2Q144UTEY	B01HJH5U	I absolutely lov	
883632	5	True	2016-11-25	A2CCDV0J5VB6F2	B01HJH5U	I'm 5'6 175lbs. I	
883633	3	True	2016-11-10	A3O90PACS7B61K	B01HJH5U		
883634	3	True	2016-11-10	A2HO94I89U3LNH	B01HJHF97K	Too c	
883635	5	True	2016-08-09	A2RSX9E79DUHRX	B01HJG5NMW	Ordered and was sli	

882402 rows × 7 columns

```
True : True      827526
Name: verified, dtype: int64
False : False    54876
Name: verified, dtype: int64
```



	overall	verified	reviewTime	reviewerID	asin	reviewText	year
0	5	True	2014-10-20	A1D4G1SNUZWQOT	7106116521	Exactly what I needed.	2014
481708	5	True	2017-04-22	A1F9ULJDCX5RJ3	B00GGE1P9E	warmer than I expected it to be, but it was a ...	2017
481727	5	True	2014-12-27	A30SFOOWLUGX5L	B00GGFOJ0A	Bought this shirt for my husband and he loves ...	2014
481726	5	True	2015-01-07	A2PG39KJXI7QRM	B00GGFMXVC	This is a great jacket!! It is very cute, and ...	2015
481725	5	True	2015-01-27	AYO3HNUHOZ5G9	B00GGFMXVC	My daughter loved it --it was nicer than the o...	2015
...
799441	1	True	2018-07-20	A307HMS0059DH2	B01C9063EK	Beautiful color but came broke	2018
799439	1	True	2017-08-24	AU6U1IZ7EOJMO	B01C91NJZA	Maybe fine for a small child with very small l...	2017
264647	1	True	2016-11-23	A323U595RK2XGF	B00VY5HC9C	I understand for the price paid you can't get ...	2016
37916	1	True	2018-01-12	A1EXTHVX2NOYX5	B000YFSR5G	To big all around, legs, length, waist, must r...	2018
712997	1	True	2015-12-09	A18SJATVEF0VTK	B015OF4XJY	it was too small, I didn't know it was from Ch...	2015

827526 rows × 7 columns

- reviewTime에서 연도만을 추출
- verified의 True False 개수 비교 후 False 제거

데이터 소개 및 수집 과정

2003-07-07 00:00:00 ~ 2018-10-01 00:00:00

전체 ID 개수 : 705943
2개 이상 리뷰 작성 ID 개수 : 86153

전체 asin 개수 : 179313
2개 이상 리뷰 작성 asin 개수 : 86153

데이터 소개 및 수집 과정

	title	brand	feature	rank	date	asin	imageURL	imageURLHighRes	descrip
0	Slime Time Fall Fest [With CDROM and Collector...	Group Publishing (CO)	[Product Dimensions'in \n8...	13,052,976inClothing, Shoes& Jewelry(8.70 inches	764443682	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	
1	XCC Qi promise new spider snake preparing men'...	NaN	NaN	11,654,581inClothing, Shoes& Jewelry(5 star	1291691480	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	
2	Magical Things I Really Do Do Too!	Christopher Manos	[Package Dimensions'in \n8...	19,308,073inClothing, Shoes& Jewelry(5 star	1940280001	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	[For professi or ama magic
3	Ashes to Ashes, Oranges to Oranges	Flickerlamp Publishing	[Package Dimensions'in \n8...	19,734,184inClothing, Shoes& Jewelry(5 star	1940735033	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	
4	Aether & Empire #1 - 2016 First Printing Comic...	NaN	[Package Dimensions'in \n1...	10,558,646inClothing, Shoes& Jewelry(5 star	1940967805	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	
...
393618	JT Women's Elegant Off Shoulder Chiffon Maxi L...	JT	NaN	9,835,890inClothing, Shoes& Jewelry(5 star	B01HJGXL40	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	
393619	Microcosm Retro Vintage Black Crochet Lace One...	Microcosm	[Package Dimensions'in \n7...	11,390,771inClothing, Shoes& Jewelry(5 star 5 star (0%)	B01HJHF97K	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	
393620	Lookatool Classic Plain Vintage Army Military ...	Lookatool	[\"Cotton+Polyester\", 'Imported', 'Item type:Ba...	972,275inClothing, Shoes& Jewelry(5 star	B01HJGJ9LS	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	
393621	Edith Windsor Women's Deep V-neck Beaded Sequi...	Edith Windsor	[Product Dimensions'in \n9...	1,964,585inClothing, Shoes& Jewelry(5 star	B01HJHTH5U	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	
393622	Aeropostale Women's Sun & Waves Crop Cami L Gr...	NaN	[Product Dimensions'in \n5...	9,379,125inClothing, Shoes& Jewelry(5 star	B01HJFNU7S	[https://images-na.ssl-images-amazon.com/imag...	[https://images-na.ssl-images-amazon.com/imag...	

393623 rows × 1355 columns

Meta Data

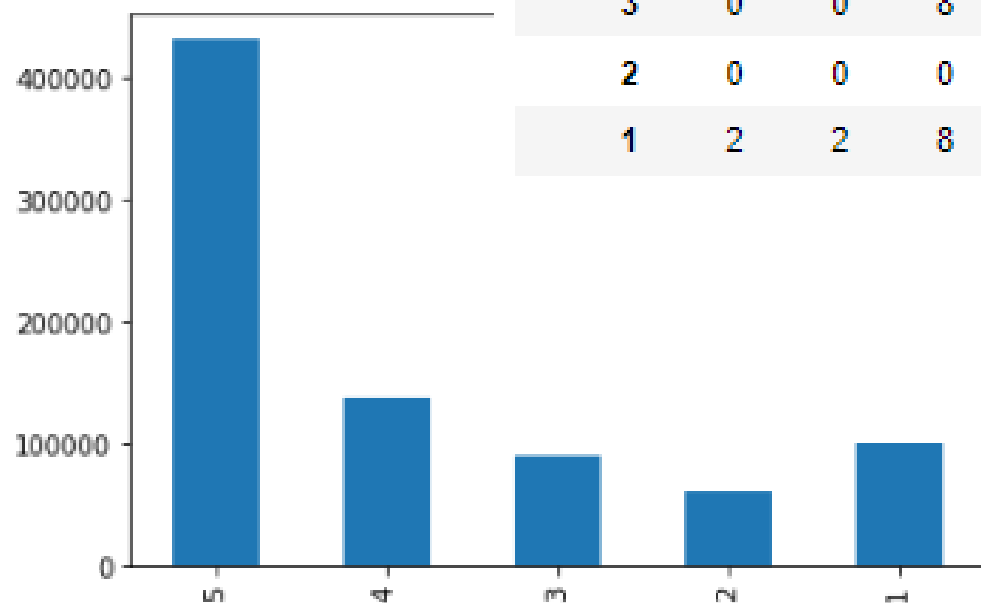
- title → 제품명 ★
- asin → 제품 코드 ★

	overall	verified	reviewTime	reviewerID	reviewText	year	title
0	5	True	2014-10-20	A1D4G1SNUZWQOT	Exactly what I needed.	2014	Milliongadgets(TM) Earring Safety Backs For Fi...
1	5	True	2014-12-20	A10QSNW2K5CR7Z	Perfect for lost backs	2014	Milliongadgets(TM) Earring Safety Backs For Fi...
2	5	True	2015-03-08	AUND2W0KDH5R8	Good quality, would order from seller again.	2015	Milliongadgets(TM) Earring Safety Backs For Fi...
3	5	True	2015-03-10	A3SO2Z8T5ZQOUN	You get exactly what is described.	2015	Milliongadgets(TM) Earring Safety Backs For Fi...
4	5	True	2015-03-24	AWTPG4FH78DXC	Perfect for your earrings. I use all the time...	2015	Milliongadgets(TM) Earring Safety Backs For Fi...
...
821501	1	True	2015-06-17	A1WD90K3UWZCJ3	Don't buy from China not good stuff and bad ma...	2015	Ayliss艾麗絲 Women Sexy Off Shoulder Loose V-neck C...
821502	1	True	2016-03-16	A1YU6F9EEWPLAH	I Don't Like it , Don't Fit need a smeller sai...	2016	Bodycon4U Women Lycra Spandex Zentai Long Slee...
821503	1	True	2012-09-04	A35NYBM3BZ089F	I found the socks ordered through internet to ...	2012	Polo Ralph Lauren Women's 3-Pairs Athletic Whi...
821504	1	True	2015-07-06	A3RITD3ORG2504	Nice color. VERY stiff.	2015	Your Gallery Classic Fashion Tote Handbag Leat...
821505	1	True	2016-02-02	A1DZXAX2CTOHVDG	In all seriousness...THIS ITEM IS NOT FIT FOR ...	2016	Cfanny Women's Caged Cutout Back Chiffon Maxi ...

821506 rows × 7 columns

데이터 소개 및 수집 과정

year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
overall																
5	2	14	36	73	394	525	751	1128	1637	4162	13373	37384	97185	149593	92677	32270
4	2	4	12	21	155	177	235	360	528	1289	4688	13979	33152	49371	26297	8055
3	0	0	8	10	76	84	103	178	249	720	2817	8805	22052	32930	17515	5558
2	0	0	0	6	47	43	66	140	209	497	1800	5574	14615	22514	11675	3660
1	2	2	8	8	45	56	78	151	259	660	2199	8205	24841	38125	19278	6109



	cnt
5	431204
4	138325
3	91105
2	60846
1	100026

2003-07-07 00:00:00 ~ 2018-10-01 00:00:00

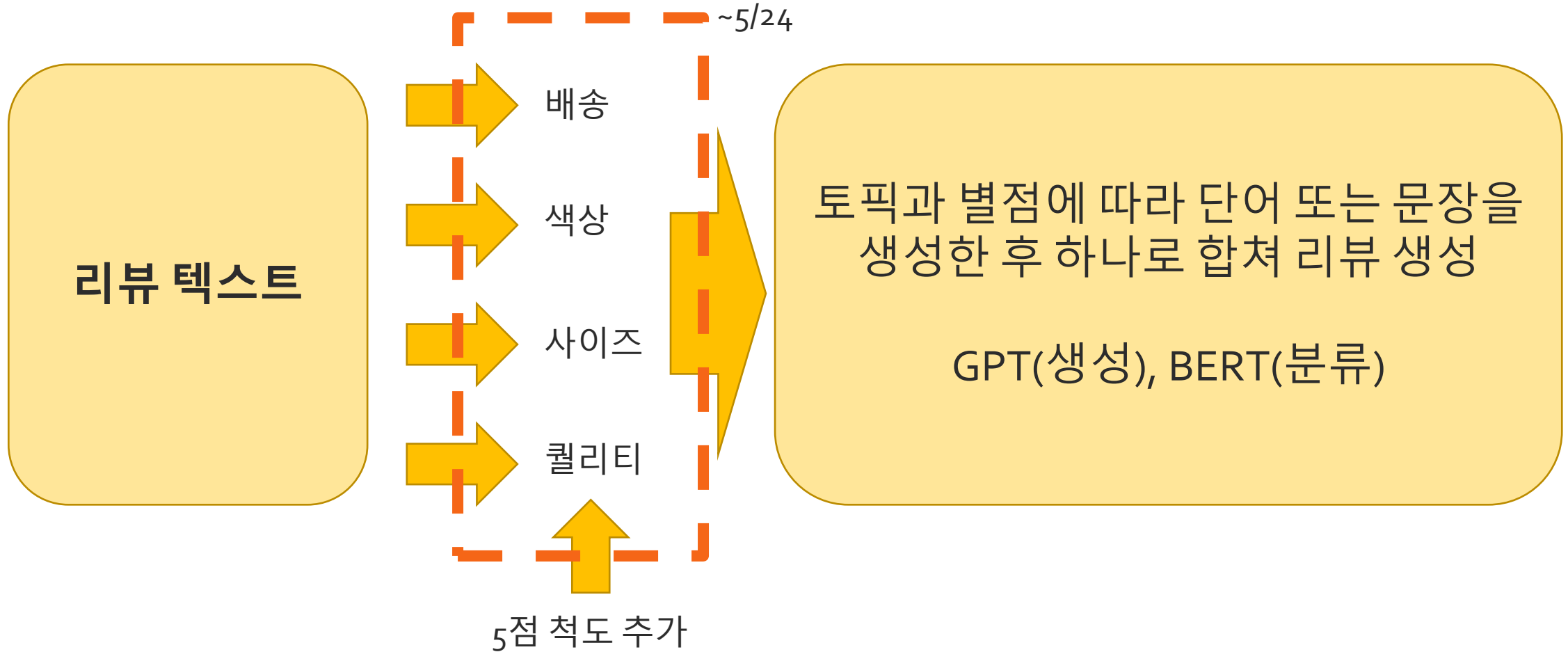
데이터 소개 및 수집 과정

```
dfTemp1 = df.loc[df['year']==2016]
dfTemp105 = []
for i in range(0, 5):
    temp = dfTemp1.loc[dfTemp1['overall'] == i+1]
    dfTemp105.append(temp)
    dfTemp105[i] = dfTemp105[i].sample(22000)
dfTemp1 = pd.concat(dfTemp105)
dfTemp1
```

	overall	verified	reviewTime	reviewerID	reviewText	year	title
799875	1	True	2016-07-18	A3AFERGB06RGQ1	Would be super sexy if you are only about 4 ft...	2016	Roswear Women's Lady Lux Viper Reversible Mono...
526330	1	True	2016-11-29	A2JR0N75Y820MF	Tiny fit and terrible print! Got it for my bro...	2016	RAISEVERN Colorful Tree Printed Freedom Zipper...
180661	1	True	2016-01-31	A3KSNM9AJ8M8LO	got to wear it once then the elephant (which i...	2016	Mother's Day Gift--BlueTop(TM) Love Silver Inf...
506926	1	True	2016-02-11	A13YEGI5SH0Z1D	The sleeves were 3/4 not long sleeves like the...	2016	T Shirts Long Sleeve Orange Lace for Women-bac...
568921	1	True	2016-09-24	AQ7S2E76198LU	Looks even better than the picture. This is m...	2016	DGY Multi-colored Cow Leather Top Handle Handb...
...
602257	5	True	2016-01-22	A2N9P3MPV38XLJ	Fits perfect, I'm thin and long. It stretches ...	2016	Pusheng Women Long Sleeve Stretch Ruched Slim ...
92489	5	True	2016-05-01	A2F8G6NDWU0VOF	Love the material!	2016	EDWARDS LADIES' LIGHTWEIGHT OPEN NECK POPLIN B...
405152	5	True	2016-06-09	A1BNX4BF8LB68X	Received on time and looks and fits as expecte...	2016	FeelMeStyle Women's Half Sleeve Blouse Dolman ...
400033	5	True	2016-07-18	A192ZIMR1MZALB	I love this hat, it would fit anyone and shows...	2016	MAKE AMERICA GREAT AGAIN - 5 Panel EMBROIDERED...
31461	5	True	2016-08-23	A383IW7156LLW7	Im am 5 feet and 118 pounds. I ordered a smal...	2016	O'Neill Wetsuits Women's Basic Skins Short Sle...

110000 rows × 7 columns

적용 알고리즘을 포함한 전반적인 과정



감사합니다