

Data Appendix

The final dataset has 538514 observations and 23 columns. Each observation represents one review of a specific product, denoted by the product_id variable

Review Dataset

Out[]:	Unnamed: 0	author_id	rating	is_recommended	helpfulness	total_feedback_count	total_neg_feedback_count	total_pos_fe
0	1158	9148213277	3	0.0	NaN	0	0	
1	1159	1306682040	4	1.0	0.5	2	1	
2	1160	5515512529	1	0.0	NaN	0	0	
3	1161	5331472901	1	0.0	0.0	9	9	
4	1162	6553094889	3	0.0	0.5	2	1	

count	submission_time	review_text	...	skin_type	hair_color	product_id	product_name	brand_name	price_usd	product_typ
0	2023-03-15	A bit sticky although that's a nice feeling (e...	...	dry	blonde	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre
1	2023-03-10	I really liked this formula a lot but I can't	combination	blonde	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre
0	2023-03-03	This product broke me out. Wanted to love it,	dry	black	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre
0	2023-02-02	when I used it 1st time - I was playing tennis...	...	normal	blonde	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre
1	2023-02-01	Too much of a greasy feeling id rather buy vit...	...	combination	brown	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre

Author_id:

identifier for the creator of a review on sephora

Rating:

number given to a product from 1-5 based on customer reviews

Is_recommended:

Tells us if the author recommends the product

Helpfulness:

$\text{total_pos_feedback_count} / \text{total_feedback_count}$

Total_feedback_count:

all feedback left

Total_neg_feedback_count:

number of users that gave a negative rating

Total_pos_feedback_count:

number of users that gave a positive rating

Submission_time:

date the review was made on sephora

Review_text:

the review a user left on a product

Skin_type:

authors skin type

Hair_color:

authors hair type

Product_id:

identifier for product on website

Product Dataset

count	submission_time	review_text	...	skin_type	hair_color	product_id	product_name	brand_name	price_usd	product_type
0	2023-03-15	A bit sticky although that's a nice feeling (e...	...	dry	blonde	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre
1	2023-03-10	I really liked this formula a lot but I can't	combination	blonde	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre
0	2023-03-03	This product broke me out. Wanted to love it,	dry	black	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre
0	2023-02-02	when I used it 1st time - I was playing tennis...	...	normal	blonde	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre
1	2023-02-01	Too much of a greasy feeling id rather buy vit...	...	combination	brown	P466154	Daily Dose Vitamin C + SPF 40 Sunscreen Serum ...	Supergoop!	46.0	Sunscre

brand_id	product_reviews	product_rating
5987	573.0	3.6736
5987	573.0	3.6736
5987	573.0	3.6736
5987	573.0	3.6736
5987	573.0	3.6736

Product_name:
name of the product

Brand_name:
name of the brand

Price_usd:
price of the product in usd

Brand_id:
unique identifier for the brand

Product_reviews:
number of reviews a product had

Product_rating:
average rating of the product

Created Variable

Product_type:
type of product (sunscreen, moisturizer, cleanser)

