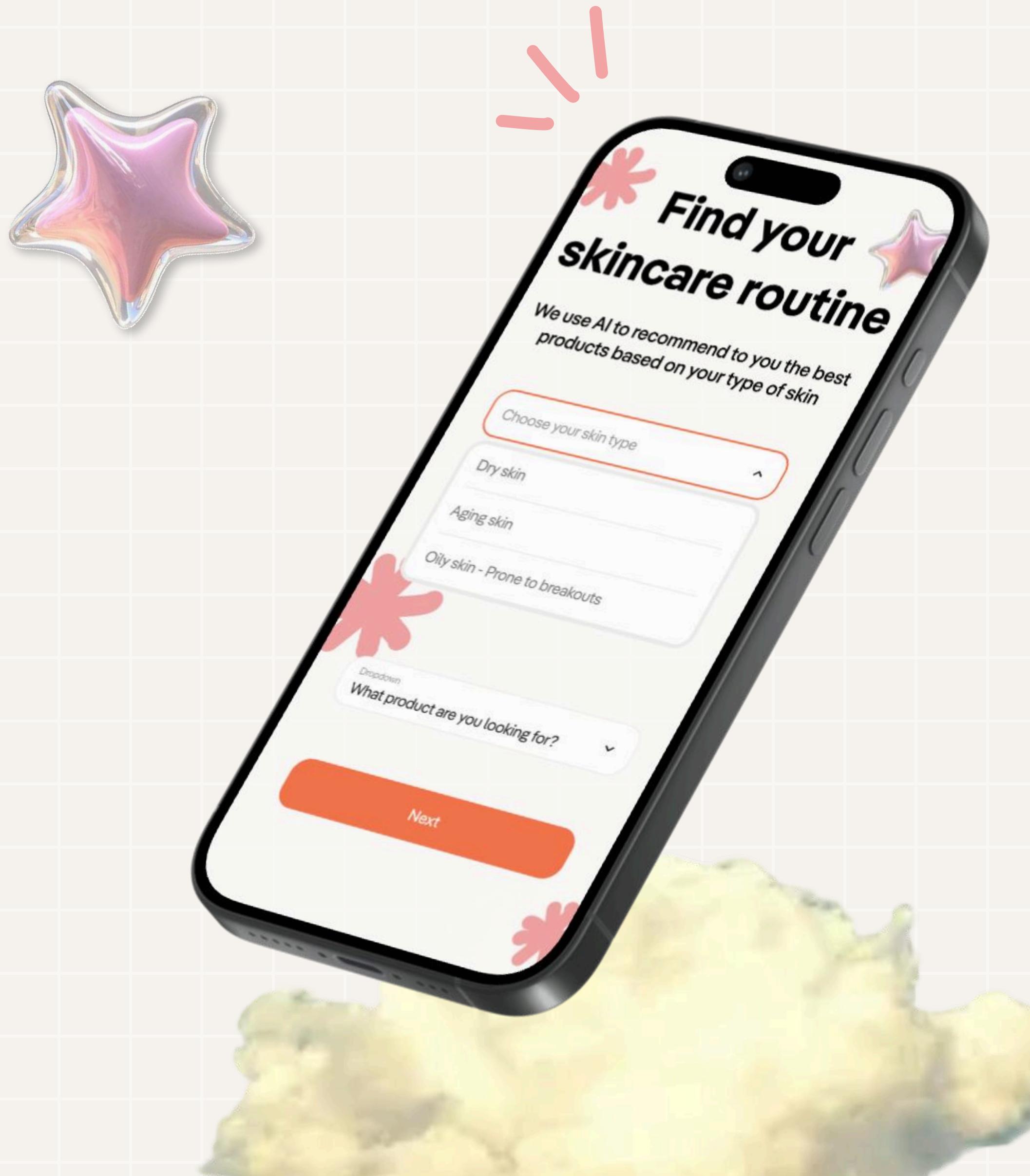


Beyond skincare

When **NLP** meets the
growing skincare market



Let's talk data

Why skincare?

"McKinsey: Beauty market to hit \$580B by 2027, growing 6% yearly."

Our data

Filtering

- Reviews with ratings \geq 3 stars
- Reviews with minimum 150 words
- Products with minimum 200 reviews



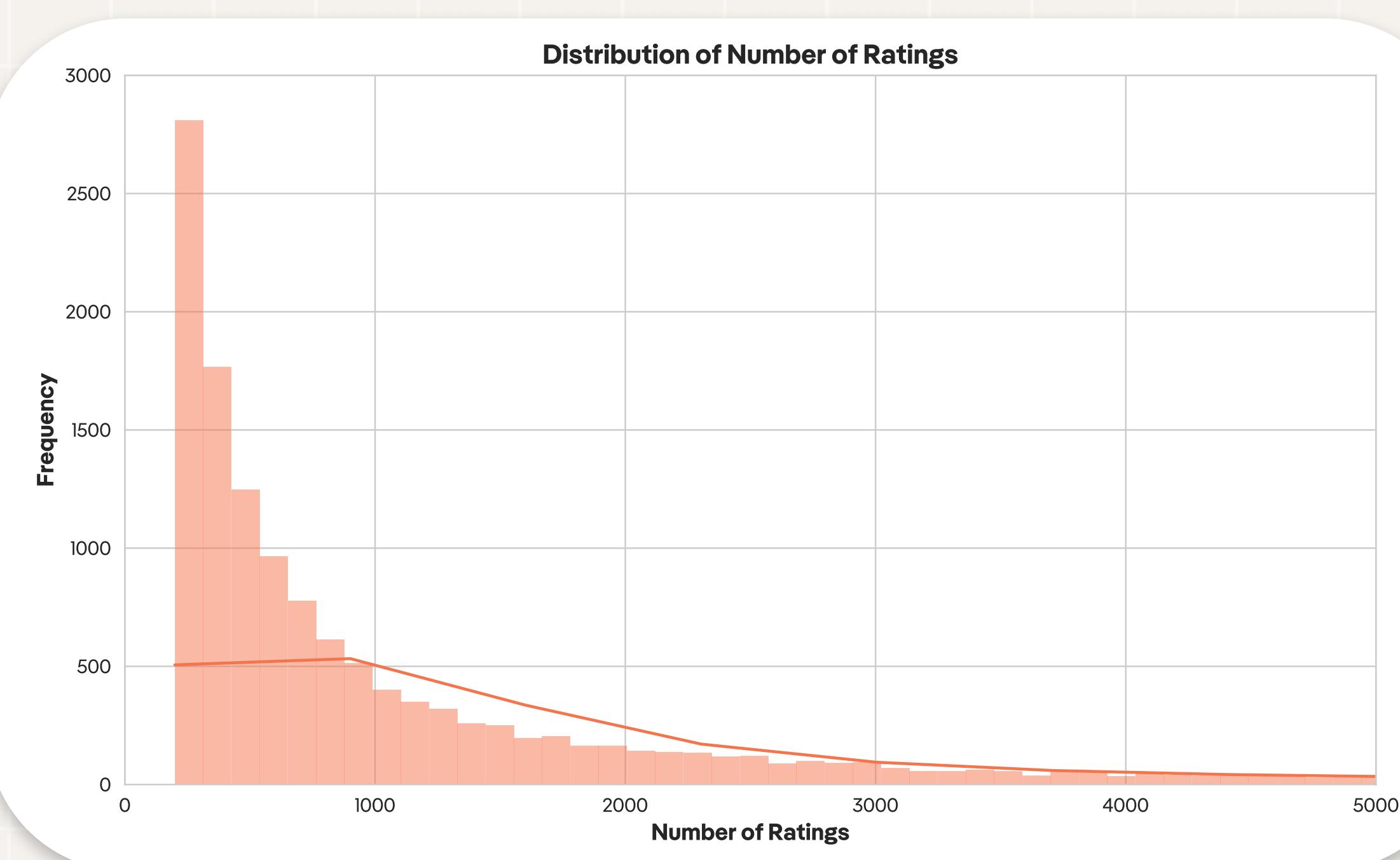
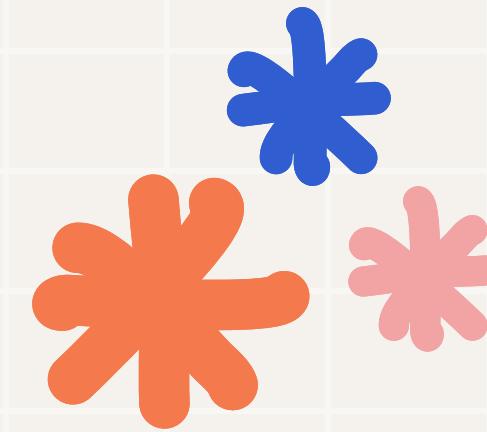
Amazon Review Data: Beauty & Personal Care Data Set
11.3M Users, 1M items, 23.9M Ratings

Processing

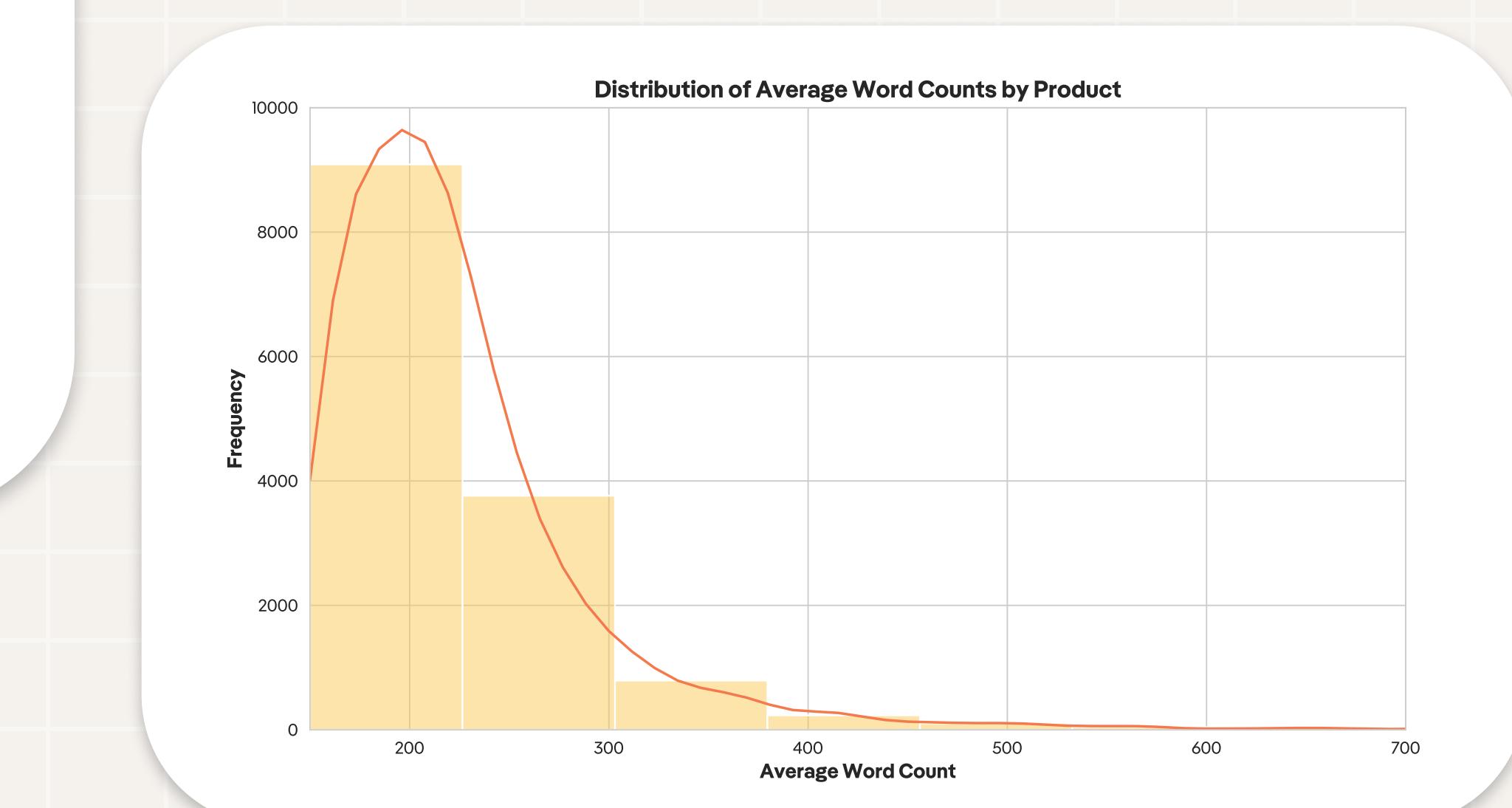
- Merged product title, features & description
- Removed stop words & Lemmatised with spacy



Let's dive right in



Most reviews are 5 stars since we filtered out products with average ratings below 3



We set our minimum word count at 150, most fall between 150 to 300, with a few outliers

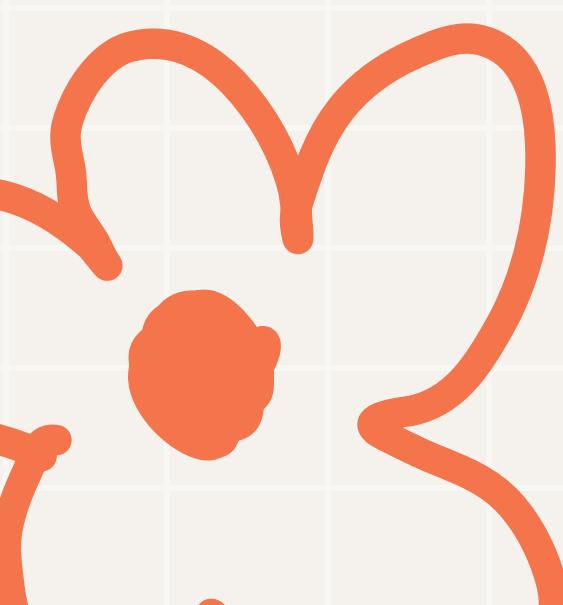
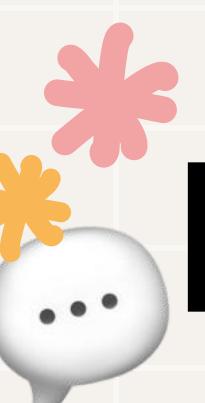


How businesses describe their products:



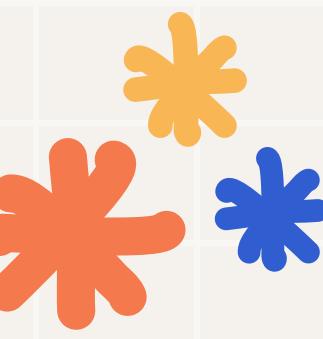
eye body formula natural cream
hydrate moisture sun soothe healthy lip reduce provide age care test soft
dry contain leave moisturizer face soap gentle fragrance smooth oil vitamin
product help water look daily pore ingredient clean use wash hand rich apply Body
moisturize oil lotion

How reviewers talk about skincare:

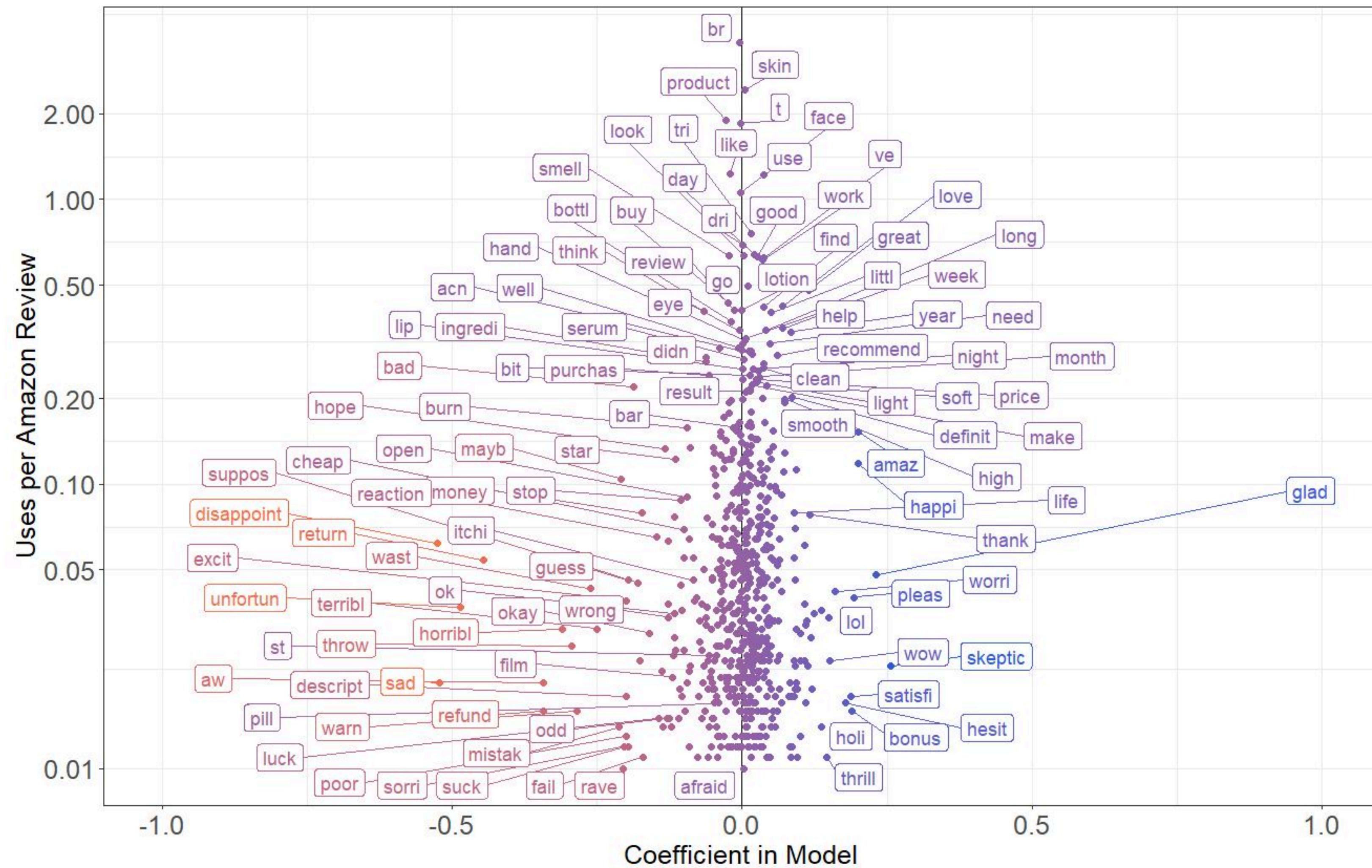


come hand week feel look time well sunscreen
hand in need dry lotion product smell
bottle long find like know work
good acne day et go soap use
love start year bo great buy recommend
want bo want skin try face
scent help little leave

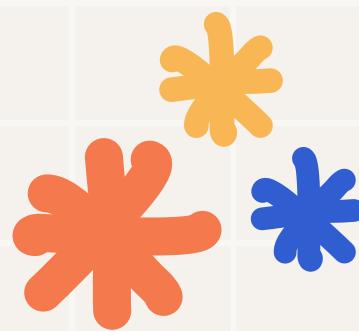
And which of the words help predict ratings? 🤔



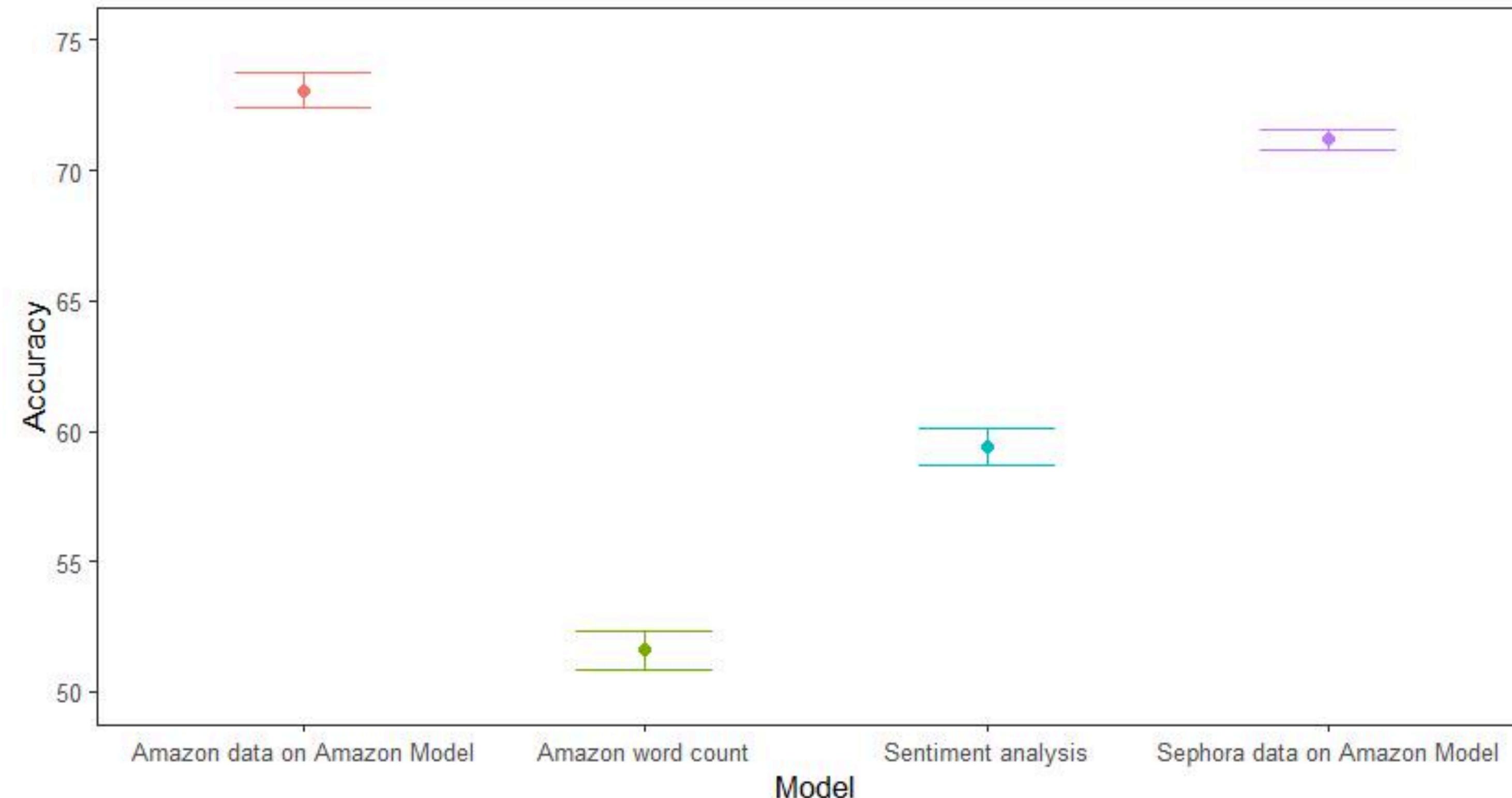
**Stronger
indicators of
negative
reviews than
positive
reviews**



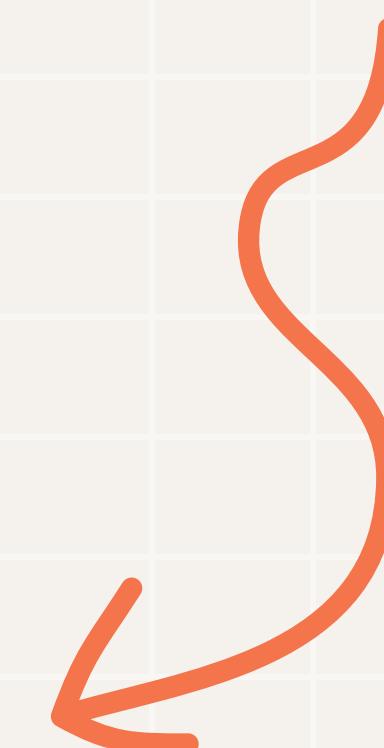
Putting amazon reviews to work



Average Accuracy of the models and benchmarks

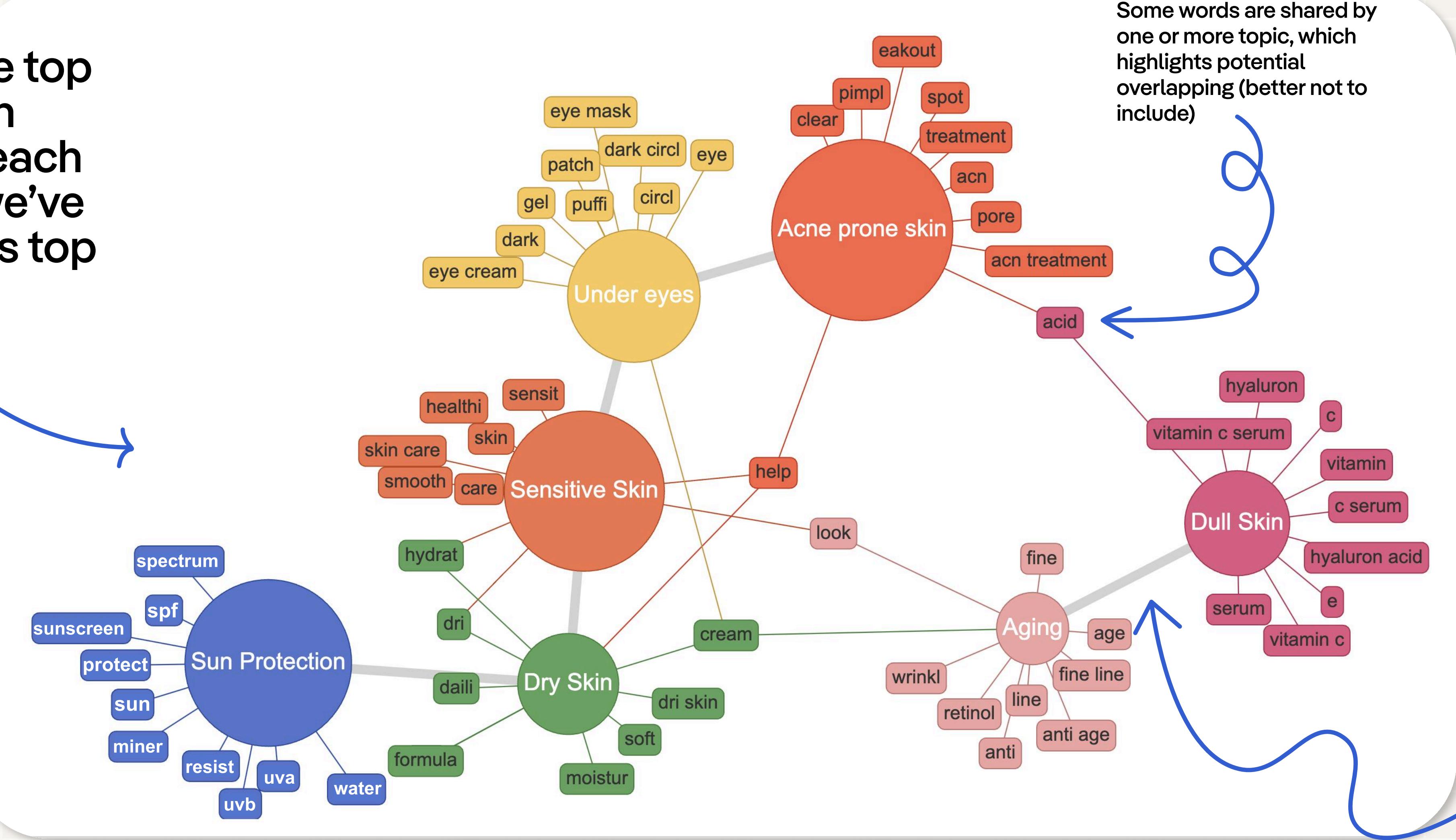


When applying our model on Sephora skin care reviews, we see that it performs almost just as well



Building a good recommender system requires good words & good topics

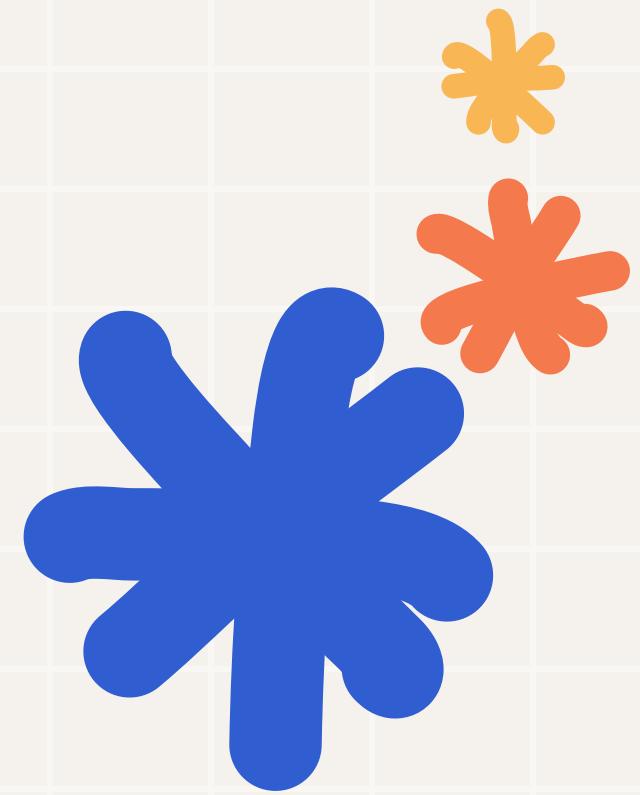
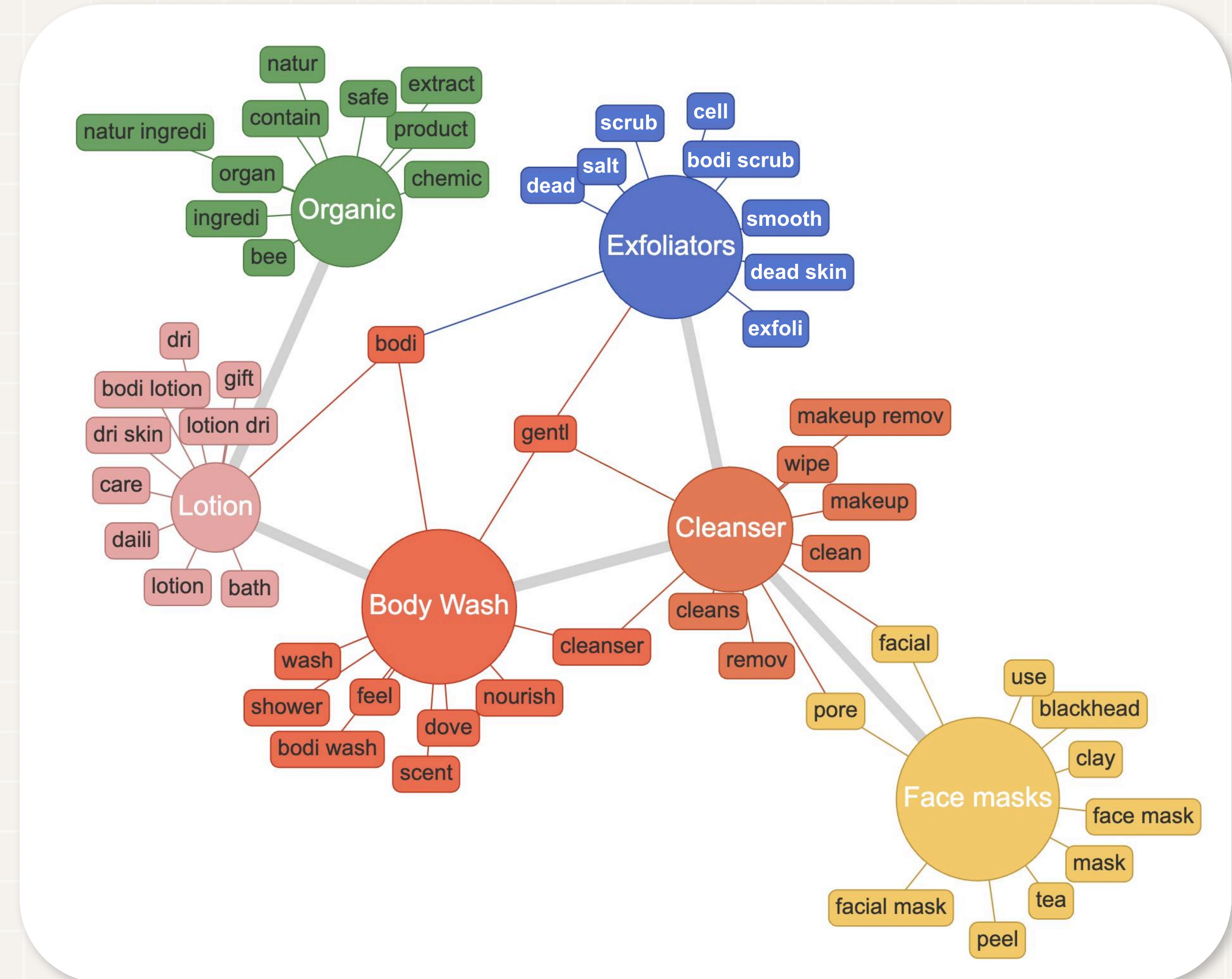
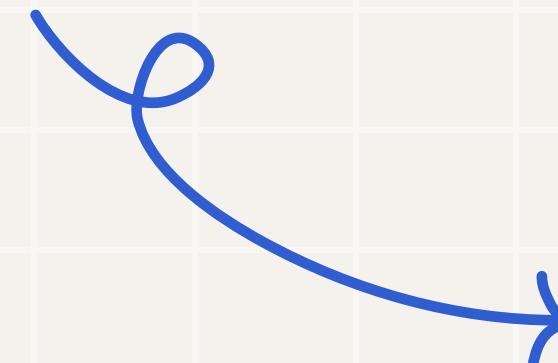
These are the top words by skin concern. To each topic node we've connected its top 10 words.



We then also connected Topics based on their similarity (how many words do they share)

Not just skincare concerns, but also product type

We then also considered product types



Okay but what do we do then with all these words?



Skin concern

Score = Weight * Sentiment Score

Weight = # of topic words in review / # of topic words

Product type

Assign binary variable 1 if topic word is present in review

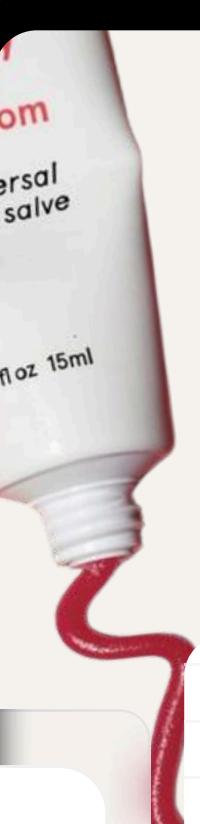
Example review

I've struggle with acne pretty much my whole life. I have try so many different acne treatment from prescription to drug store solution I be so happy to have find this face wash. I see huge difference in my acne and have people constantly ask what I do to clear my face. With this all natural solution my face be hydrated and more even look I just can not believe that this natural and cheap solution .

Next slide to see some of the results of our recommender



Tada! Here's your skincare recommendations



Type of product
Sunscreen

**Obagi Sun Shield
Tint Broad
Spectrum SPF 50**

Average rating:
4.7
Purchase at:
Amazon.uk



Type of skin
Aging skin

Type of product

**Paula's Choice BOOST 1%
Retinol Booster, Vitamin
A & Licorice Serum for
Fine Lines & Wrinkles**

Purchase at:
Amazon.uk

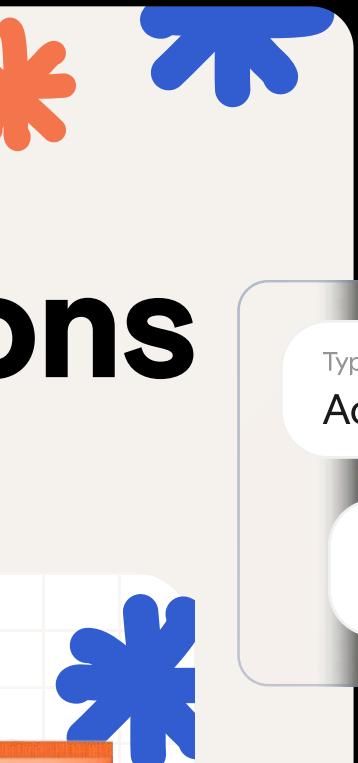
Average rating:
4.3



Using NLP to find the best product for your needs 🥑

**AcneFree Severe
Acne 10% Benzoyl
Peroxide Foaming
Cleansing Wash**

Purchase at:
Amazon.uk
Average rating:
4.5



**numbuzin No.4 Collagen 73%
Pudding Serum | Wrinkle Care,
Aging**

Purchase at:
Amazon.uk

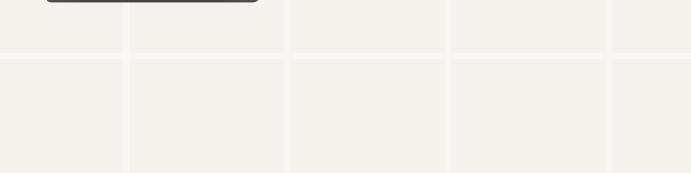
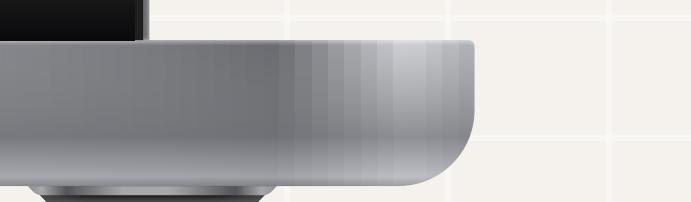
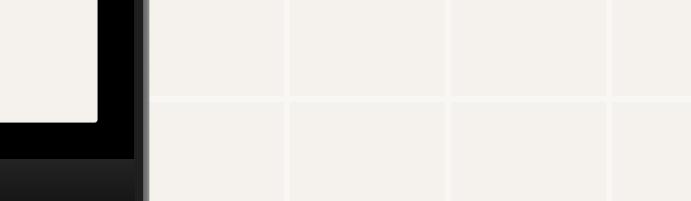
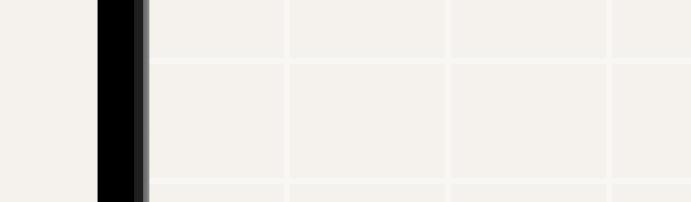
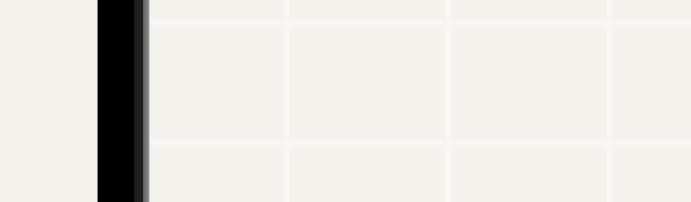
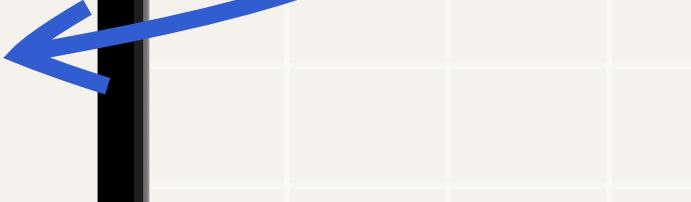
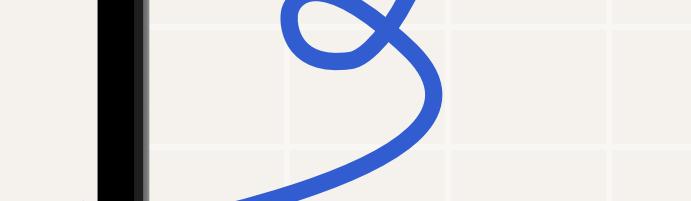
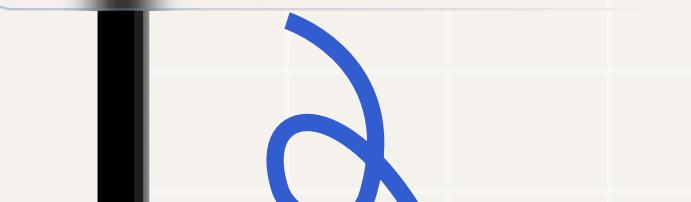
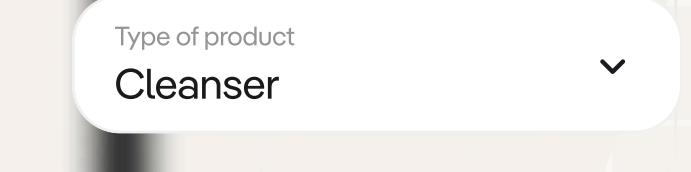
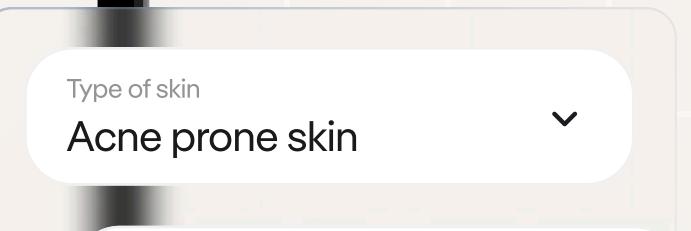
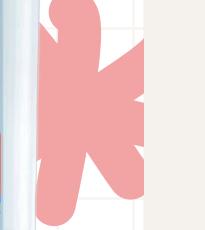
Average rating:
4.4



**Neutrogena clear
pore astringent
cleanser**

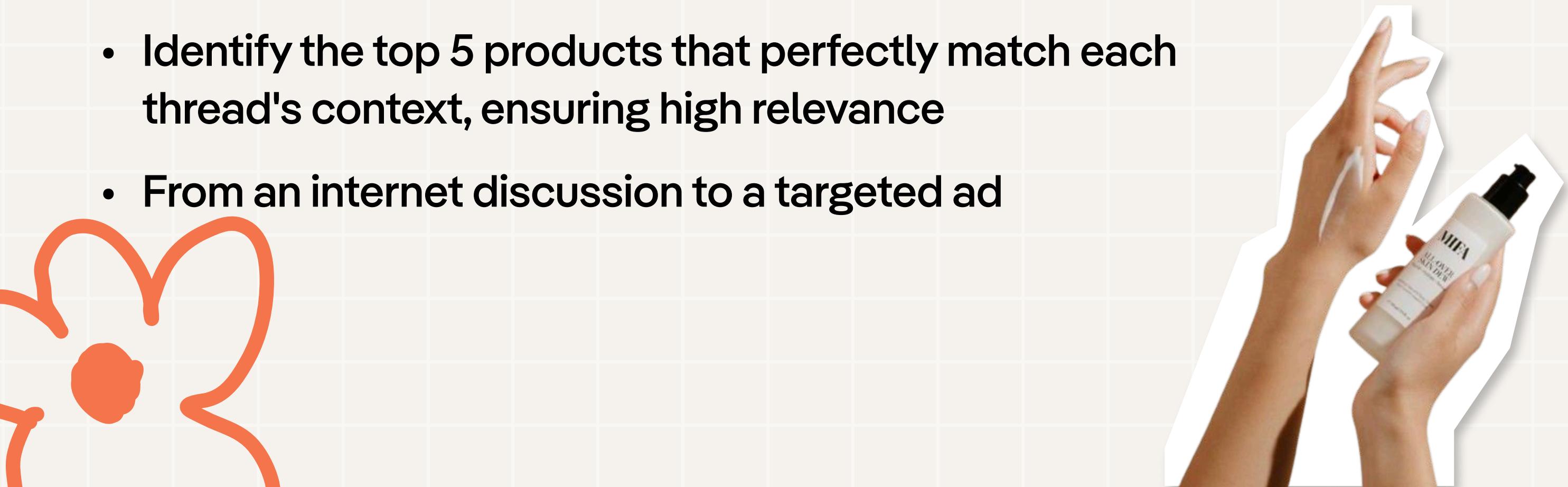
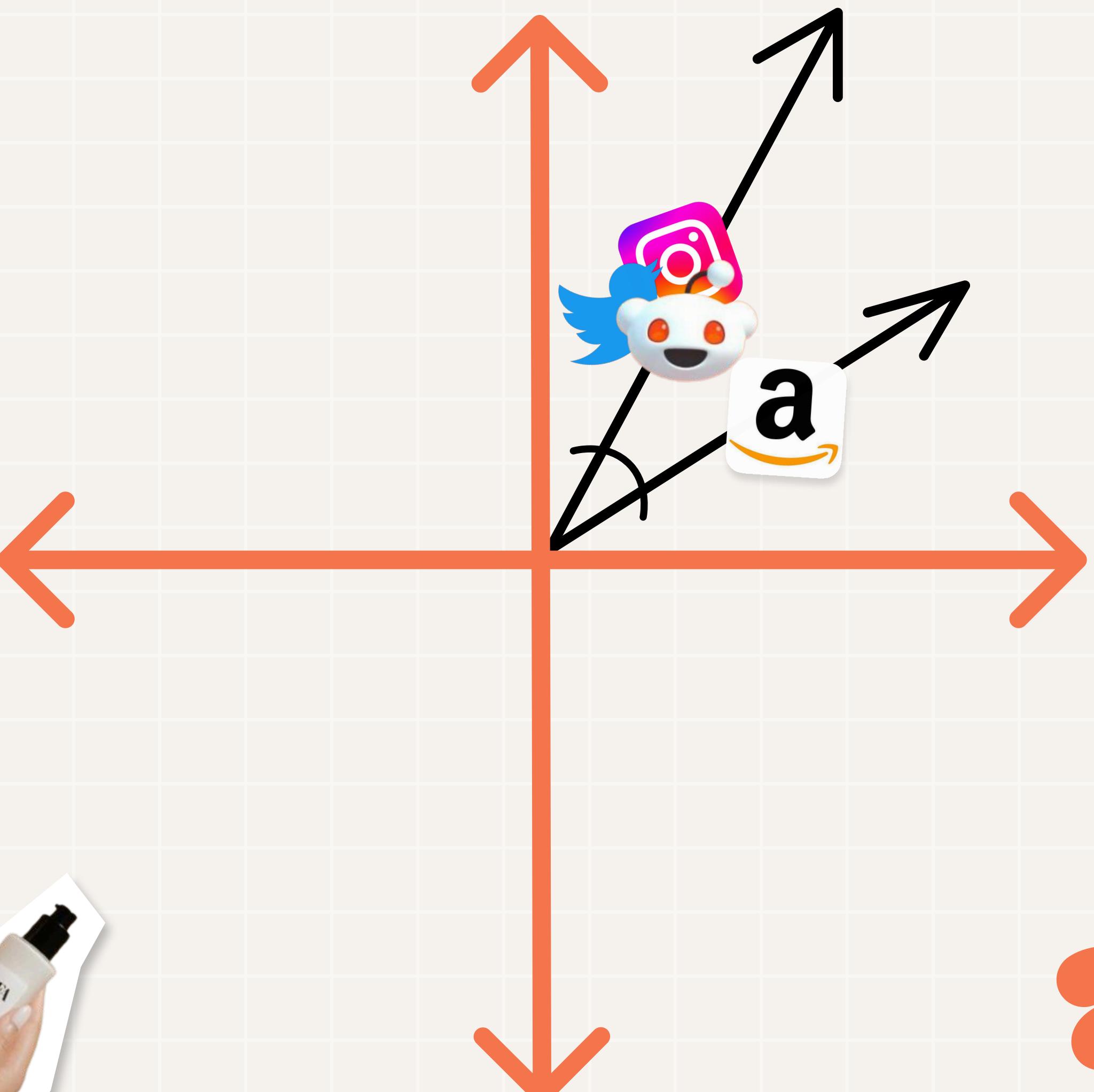
Purchase at:
Amazon.uk

Average rating:
4.8



So much data, so many opportunities: Let's Talk Cosine

- Leverage web scraping to access consumer data and gain valuable insights 🧑
- Target consumers with perfectly matched product recommendations based on their preferences
- Transform online discussions and product reviews into vectors using natural language processing (NLP)
- Calculate similarity between thread vectors and review vectors to identify the most relevant products
- Identify the top 5 products that perfectly match each thread's context, ensuring high relevance
- From an internet discussion to a targeted ad



Testing it on r/Skincare addicts



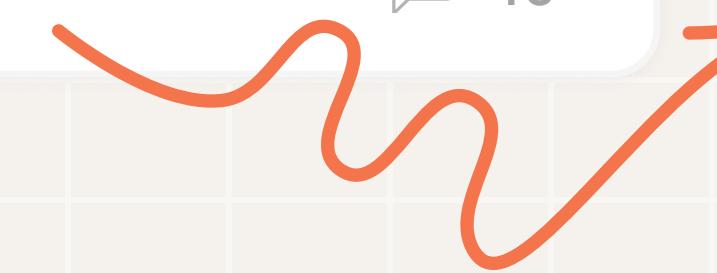
r/SkincareAddicts



**Requesting an
Intervention: please
design my whole routine**

120+

10+



Nownoon Vitamin C
Serum - Hyaluronic
Acid



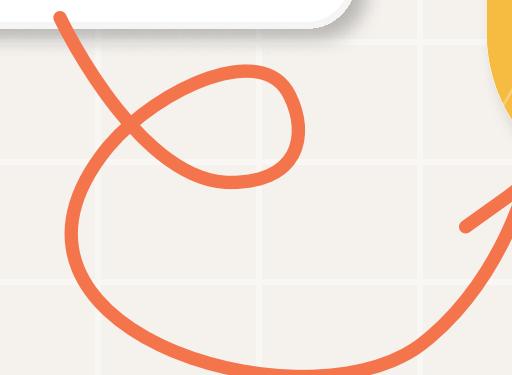
r/SkincareAddicts



REVIEWS: Sunscreens

120+

10+



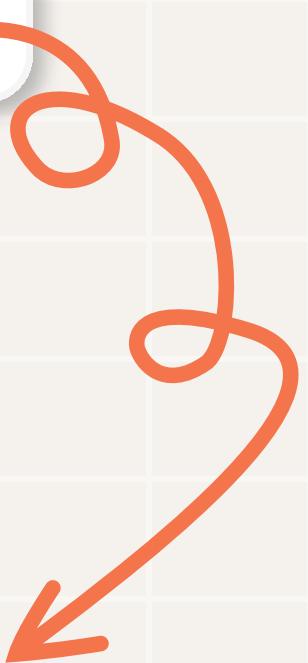
Biotherm Creme
Solaire Dry Touch
UVA/UVB Matte
Effect Face Cream



Beauty Aura
Sunflower Oil - For
Healthy Hair, Skin &
Nails.



10+



This is just a start*

Limitations 🚧

- Overlapping between topics, common words e.g., moisturize
- Differences in how people talk about products and how the product description describes the topics:
 - Companies describe formulas, ingredients, and functions 🧪
 - People describe how it feels, smells, applies, looks
- Our model only works on products which have been on the market long and are popular enough to accrue many reviews 📈
- Bias: We hand-selected specific words to distinguish within topics using dictionaries based on our own domain knowledge

Next steps 🤔

- Improved filtering system to better handle overlapping topics and common words 🎯
- Incorporating more advanced NLP techniques to bridge the gap between company descriptions and user reviews
 - Exploring connections between products (e.g., the bought together column)
 - Looking at user_ids to cross check reviews across different products
- Expanding the model to work on newer and less popular products by leveraging transfer learning 📚
- Reducing bias by using more diverse and representative dictionaries
 - Using ingredients dictionary for recommendations based on product composition

Thank you

Now time for questions



you after this
presentation

