

# Sentiment Analysis of Amazon's Deteriorated Product Ratings

# Data Collection

- Amazon Beauty Product dataset from 2014 collected from :  
'<http://jmcauley.ucsd.edu/data/amazon/index.html>' UCSD dataset for research purposes
- Additional data scraped using the product asin numbers using selenium and beautiful soup

# Features in the Dataset

	asin		name	category		description	price	rating	No_of_Rating
0	9788072216	Prada Candy by Prada for Women 1.7 oz Eau de P...	Fragrance		Brand Prada Scent Honey, Musk , Vanilla Item F...	56.29	4.7		1391
1	B00004TMFE	Avalon Organics Therapy Thickening Conditioner...	Hair Care		Brand Avalon Organics Scent Biotin B Hair Type...	7.81	4.1		5646
2	B00004TUBL	CLASSIC Better Living Two Chamber Dispenser, W...	Bath		Color White Brand CLASSIC Item Dimensions LxWx...	39.99	4.3		99

- 2014- Dataset (9 features, 12101 unique asins)

	reviewerID	asin	reviewerName	helpful	reviewText	overall	summary	unixReviewTime	reviewTime
0	A1YJEY40YUW4SE	7806397051	Andrea	[3, 4]	Very oily and creamy. Not at all what I expect...	1	Don't waste your money	1391040000	01 30, 2014
1	A60XNB876KYML	7806397051	Jessica H.	[1, 1]	This palette was a decent price and I was look...	3	OK Palette!	1397779200	04 18, 2014
2	A3G6XNM240RMWA	7806397051	Karen	[0, 1]	The texture of this concealer pallet is fantas...	4	great quality	1378425600	09 6, 2013

- Scrapped Data From 2021 (7 features, 4242 unique asins)

# Data Cleaning

Removed unwanted columns  
and kept the following features:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4242 entries, 0 to 4241
Data columns (total 13 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   asin              4242 non-null    object  
 1   overall            4242 non-null    int64  
 2   helpful             4242 non-null    float64 
 3   Not helpful        4242 non-null    float64 
 4   review_concat       4242 non-null    object  
 5   summary_concat      4242 non-null    object  
 6   overall rating     4242 non-null    int64  
 7   name               4242 non-null    object  
 8   category            4242 non-null    object  
 9   description          4242 non-null    object  
 10  price                4242 non-null    float64 
 11  rating               4242 non-null    float64 
 12  No_of_Rating         4242 non-null    int64  
dtypes: float64(4), int64(3), object(6)
memory usage: 464.0+ KB
```

Merged both the dataset based on asin id

Changed all numerical features to int64 or float64 datatype (whichever applicable)

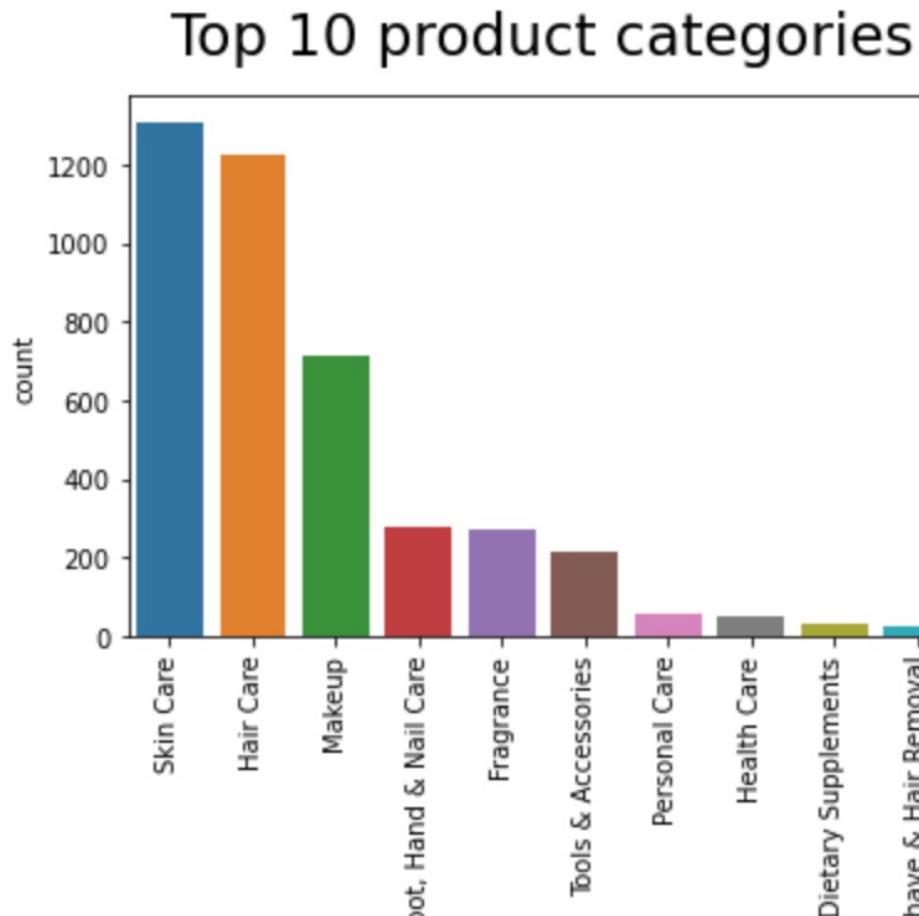
Preprocessed the datatype object(strings) for NLP

- converted all text to lowercase
- formatted empty spaces
- Expanded contraction (eg. mgmt. to management)
- Filtered all punctuation
- filtered stopwords (have, is, was)
- Lemmatization: getting the root words (better : lemma good)

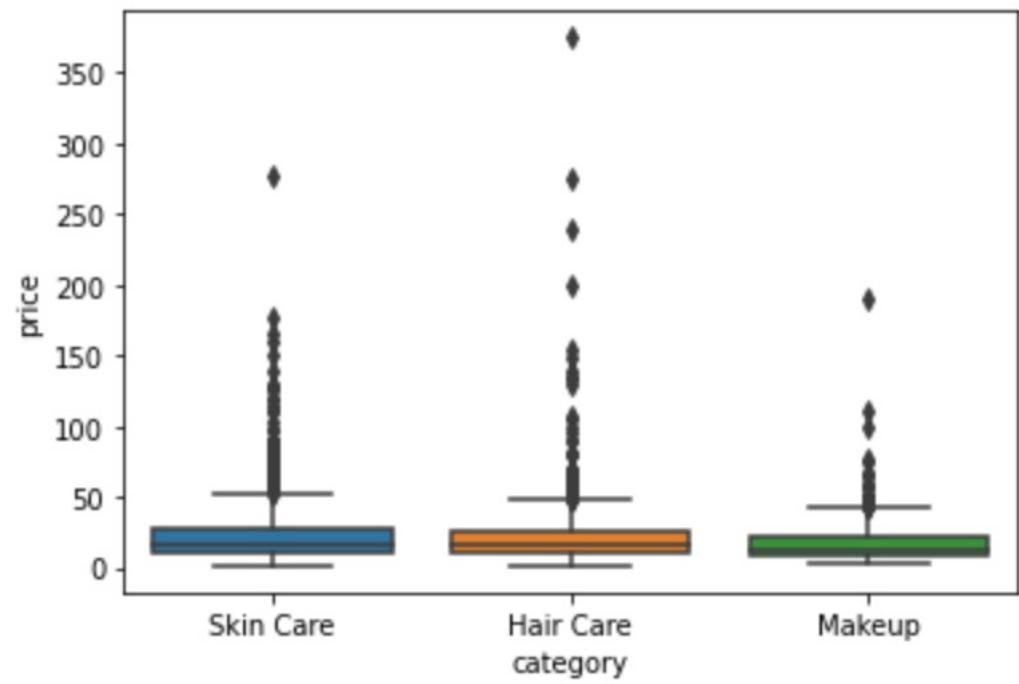


# Exploratory Data Analysis (EDA)

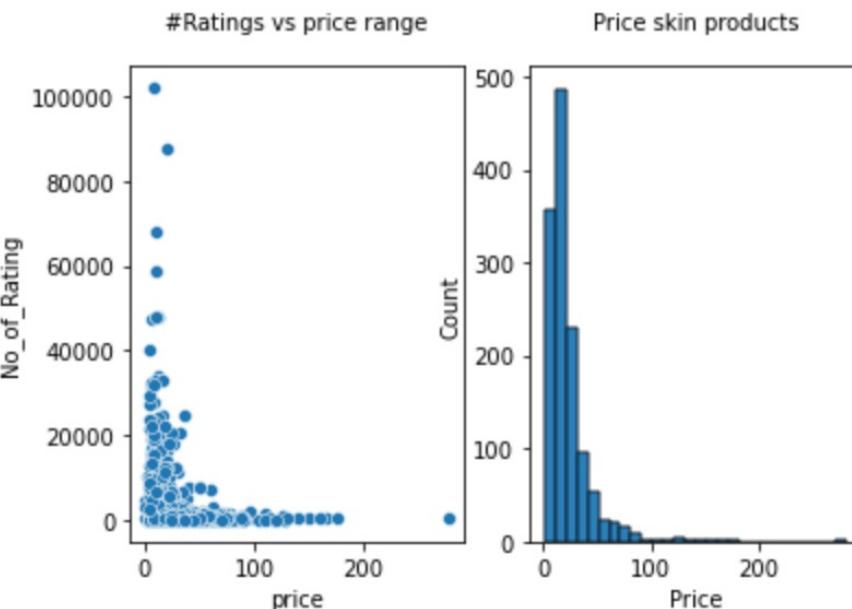
- Top 10 categories of product in the dataset
  - The largest number of products are in the categories: Skin Care, Hair Care & Makeup
  - These 3 categories will be used in sentimental analysis



- Spread of Price within each of the 3 categories
  - The price of products in hair care category is more spread out
  - Analyzed the spread of price vs number of customers around each price range
  - Number of customers is estimated from number of ratings

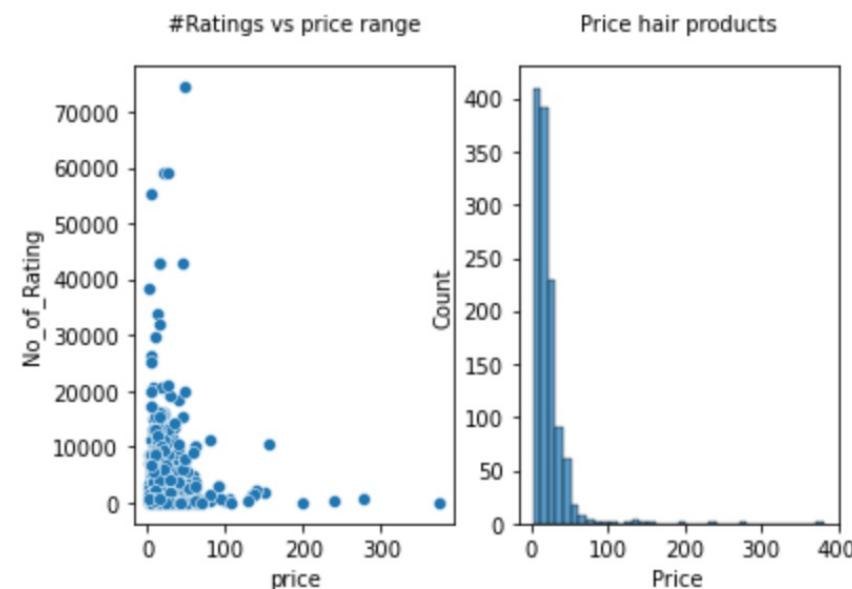


# Skin Products

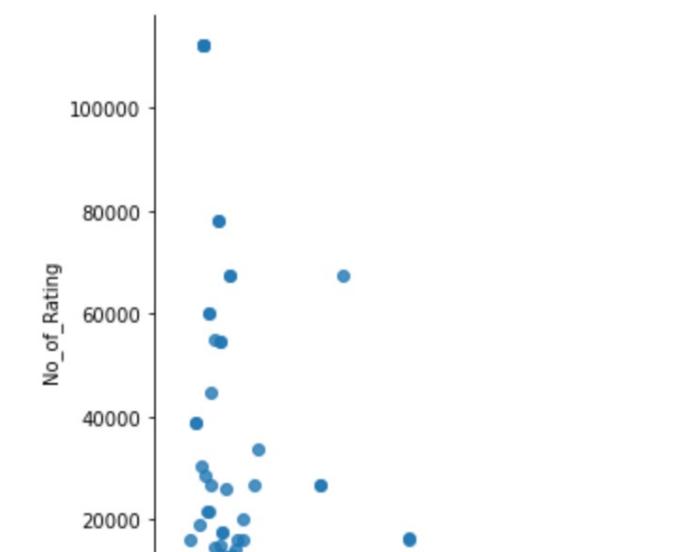
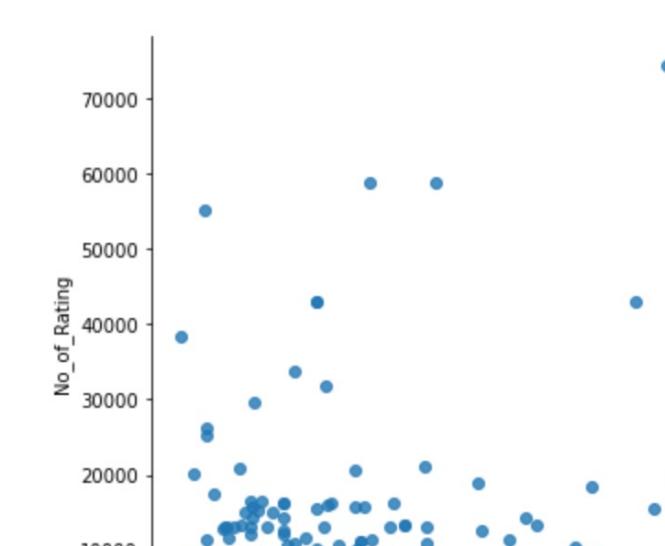
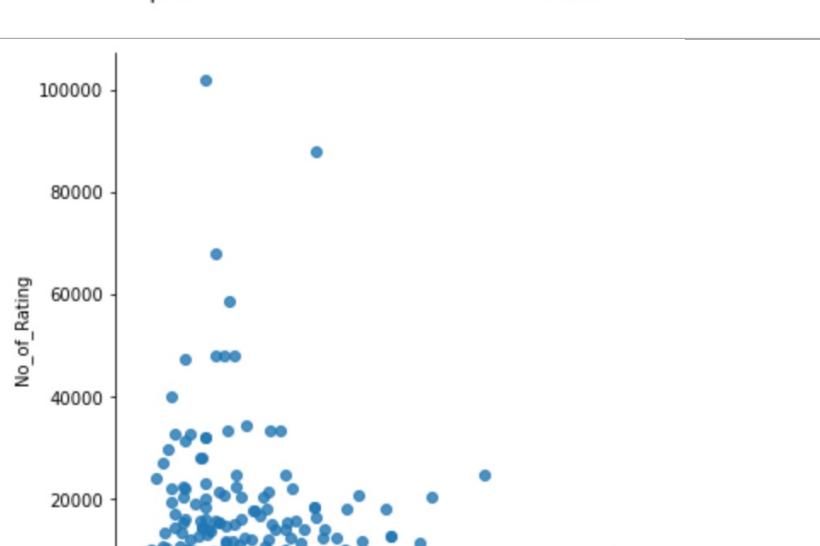


# Hair Products

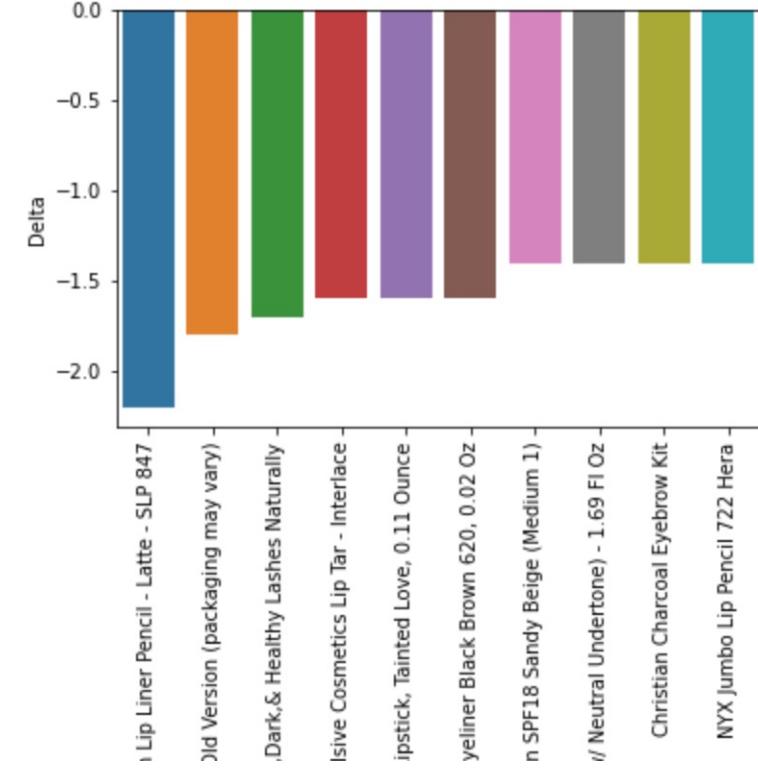
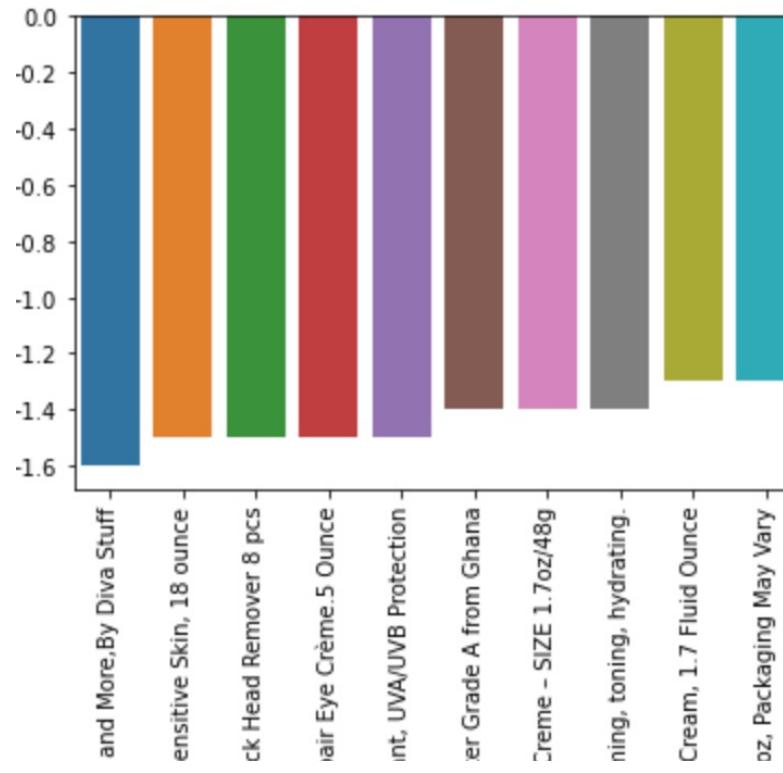
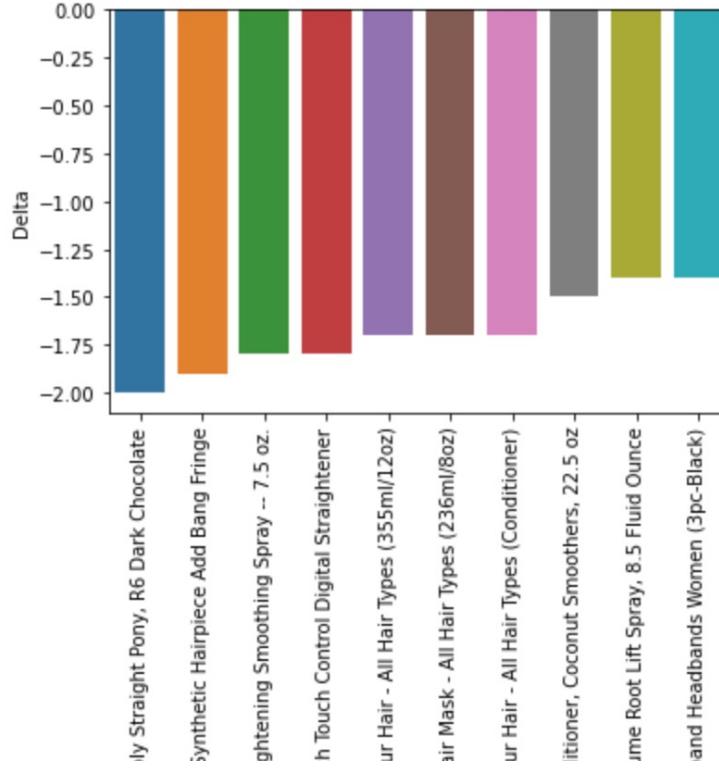
People usually spend more on hair product than skin and makeup



# Makeup



Top 10 haircare products that dropped rating in 6 years | Top 10 skincare products that dropped rating in 6 years | Top 10 makeup products that dropped rating in 6 years

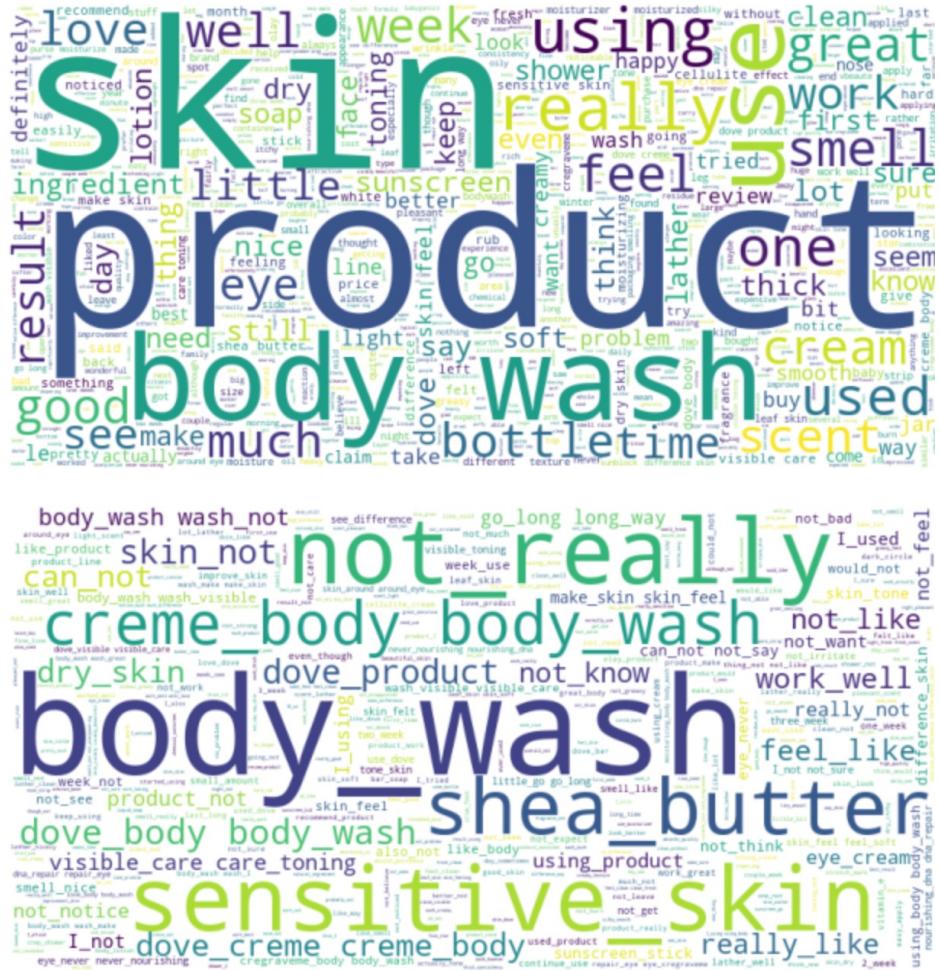


# Change in ratings over 6 years (2014-2021)

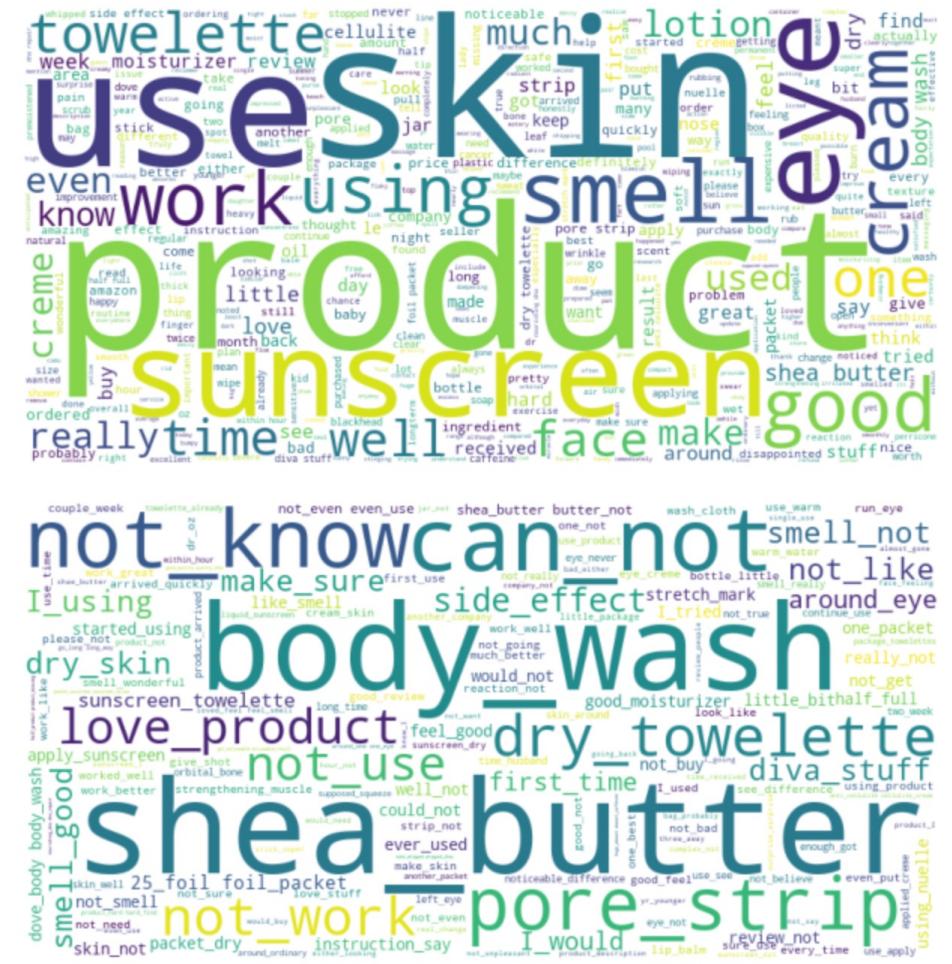
The current reviews of these 30 products were scraped using selenium and beautiful soup

# Sentiment Analysis

# Skin Product (word cloud from Reviews)



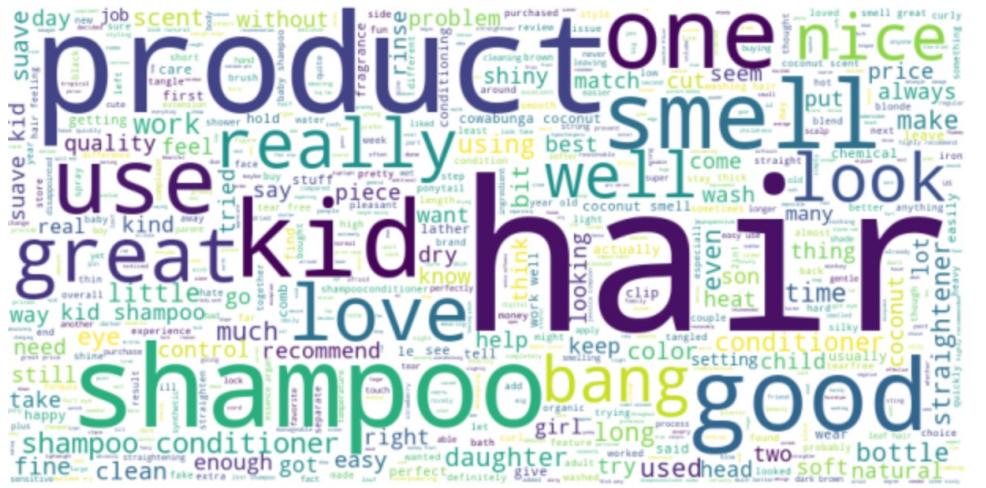
## 2014 word cloud: Reviews



## 2021 word cloud: Reviews

# Recommendations

# Hair Product (word cloud from Reviews)



## 2014 word cloud: Reviews



## 2021 word cloud: Reviews

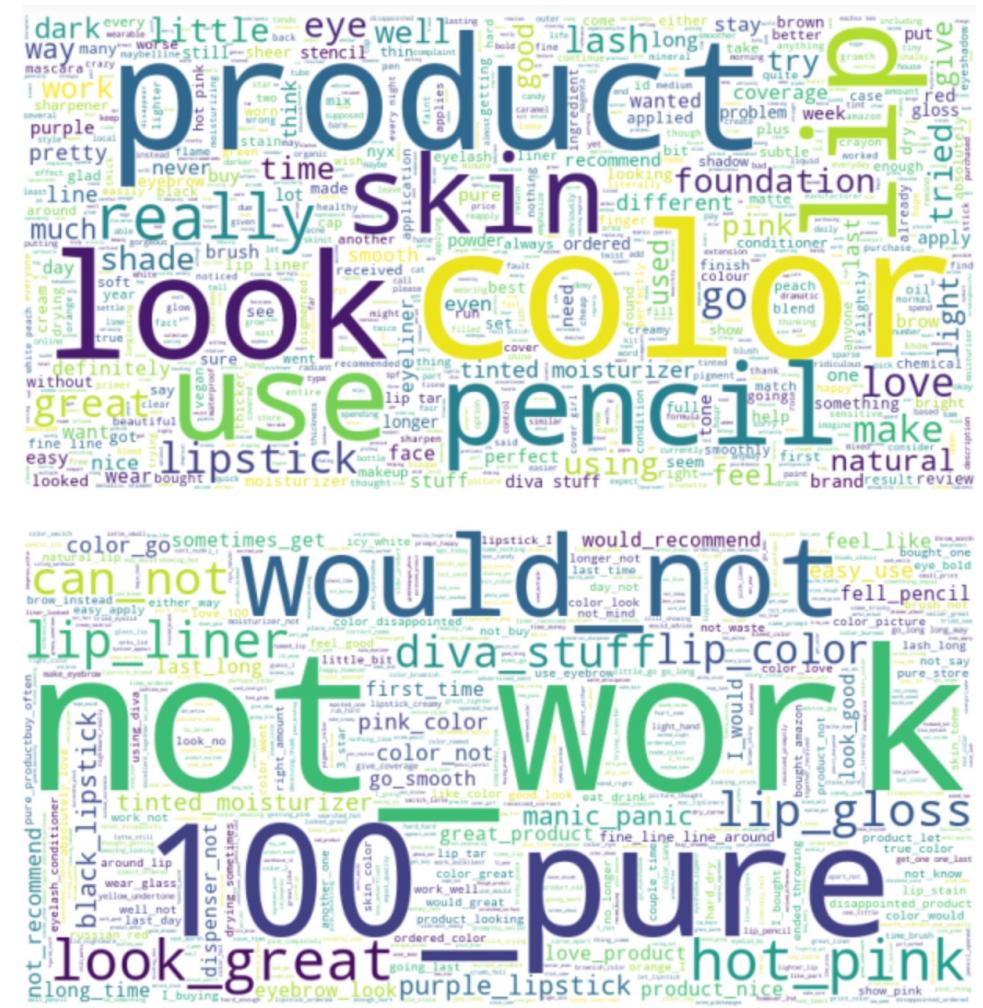


# Recommendations

# Hair Product (word cloud from Reviews)



## 2014 word cloud: Reviews



## 2021 word cloud: Reviews

# Recommendations

A large, semi-transparent blue circle is positioned on the left side of the slide, overlapping a dark blue vertical bar.

# FUTURE WORKS



# THANK YOU

---

