



# INSIGHTS AND RECOMMENDATIONS FROM REVIEWS

-AISHWARYA NAIR



# Introduction

The objective here is to identify the key factors that influence customer satisfaction, pinpoint recurring pain points, and highlight aspects that customers appreciate. This analysis will cover several critical areas:

- Identifying patterns or trends that stand out in different segments of the data.
- Obtain a statistical summary for the numerical attributes within the dataset.
- Perform overall sentiment analysis to gauge the general mood of the reviews.
- Identify common themes and extract common keywords within the reviews.
- Identify pain points from the negative reviews to isolate specific issues that customers are facing, enabling targeted improvements.
- Analysis of positive reviews to identify strengths and areas that are working well, which can be leveraged for marketing and product development.



# Objectives

- Since the problem is open ended, the main idea was to think from the perspective of the retailer and identify how the products perform, understand the customer-product affinity, how existing products or processes can be improved, how to understand customer segments better and the pain areas that needs to be addressed.
- Few questions were identified which could be addressed by means of this analysis. It can be categorized as follows:
  - Product Type and Ratings
  - Age Groups, Rating and Reviews
  - Product Type, Rating and Reviews
  - Reviews Sentiment
  - Review Analysis

---

# Data Dictionary

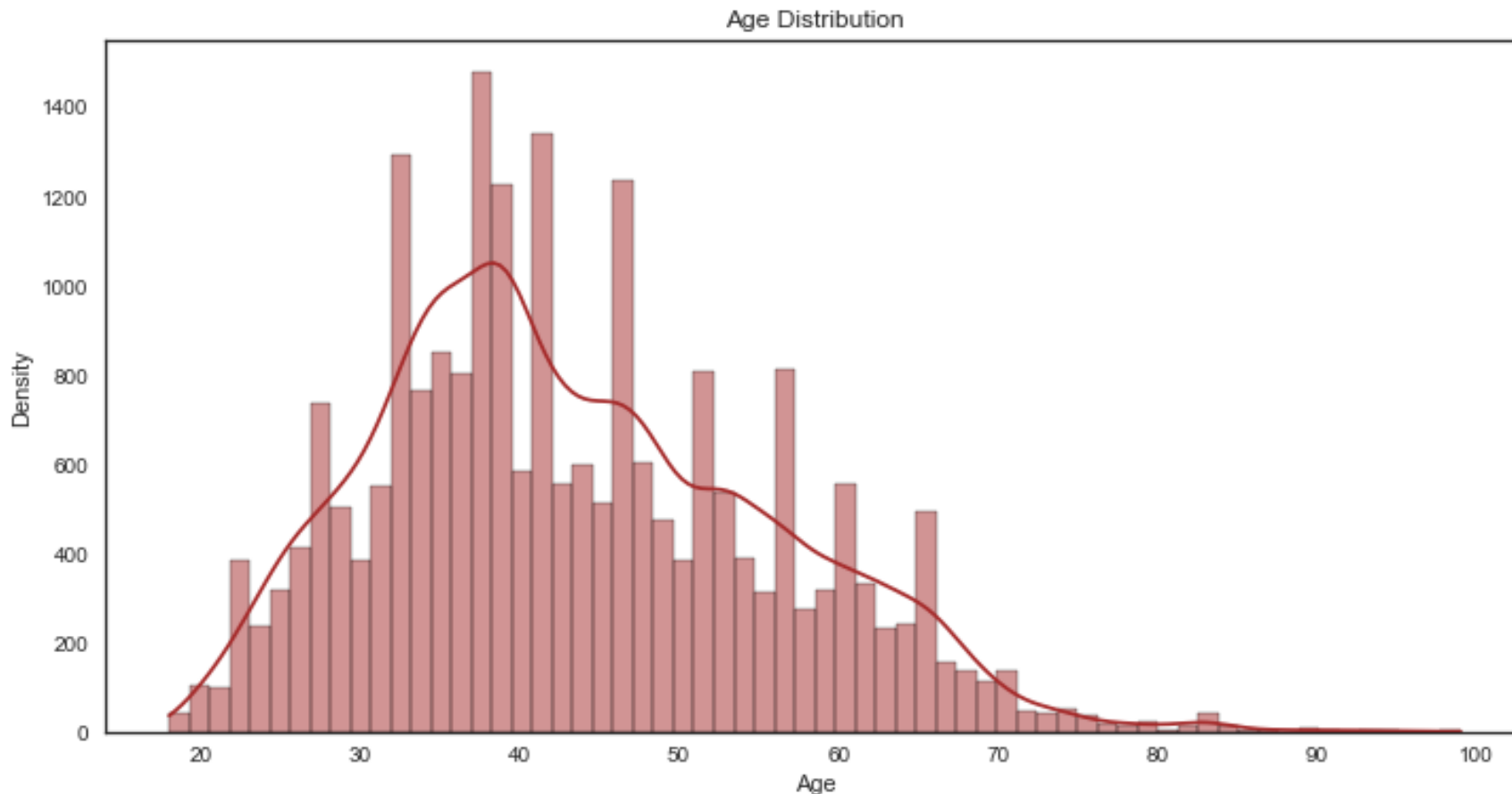
Column	Description	Data Type	Non-Null Count
Product_ID	Unique identifier for each product	int64	23486
Age	Age of the reviewer	int64	23486
Title	Title of the review provided by the reviewer	object	19676
Text_Review	Detailed text review provided by the reviewer	object	22641
Rating	Rating provided by the reviewer	int64	23486
Type	Type of the product	object	23472

# Sample Dataset and Dataset Size

Dimensions of the data: 23486 rows and 6 columns

	Product_ID	Age	Title	Text_Review	Rating	Type
0	767	33	NaN	Absolutely wonderful - silky and sexy and comf...	4	Intimate
1	1080	34	NaN	Love this dress! it's sooo pretty. i happene...	5	Dresses
2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	Dresses
3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	Bottoms
4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	Tops

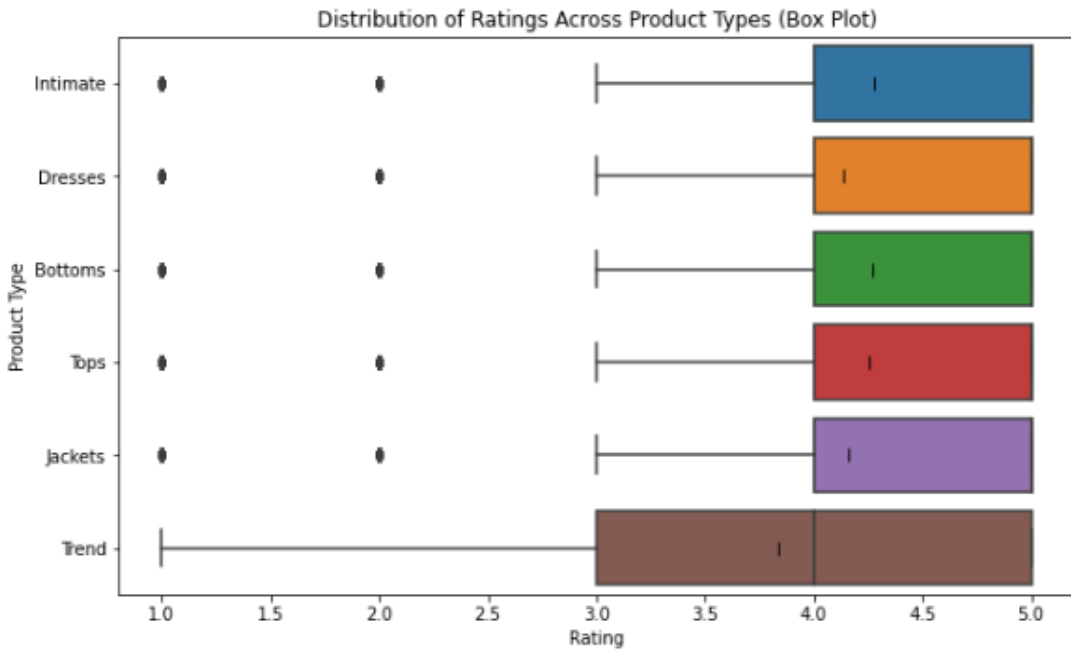
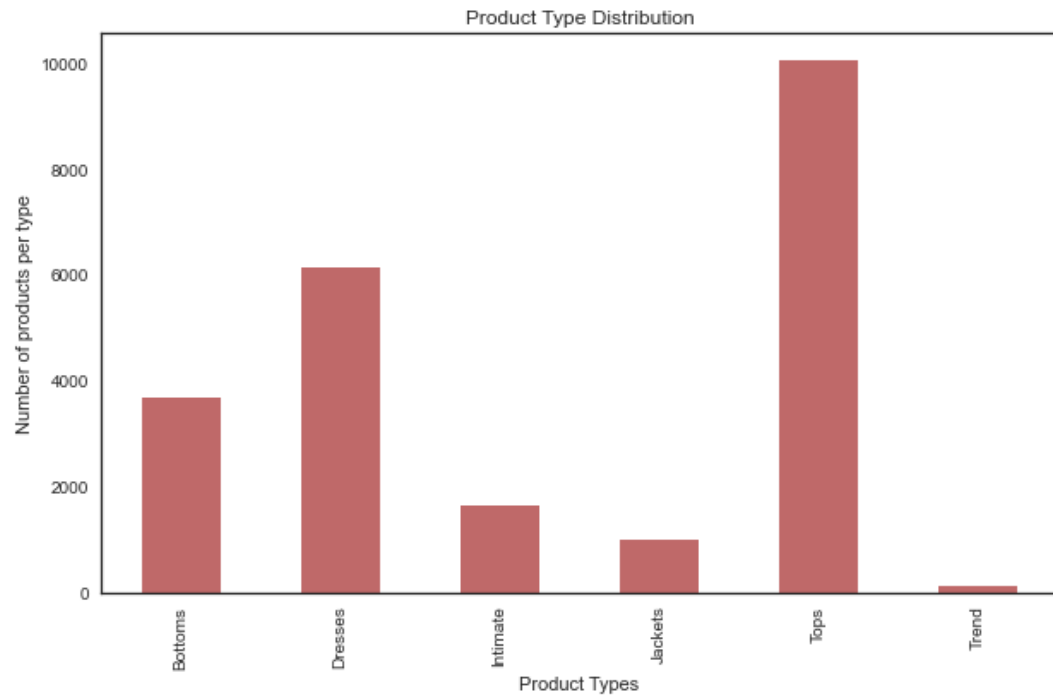
# Age Distribution



The minimum age amongst the reviewers = 18

The maximum age amongst the reviewers = 99

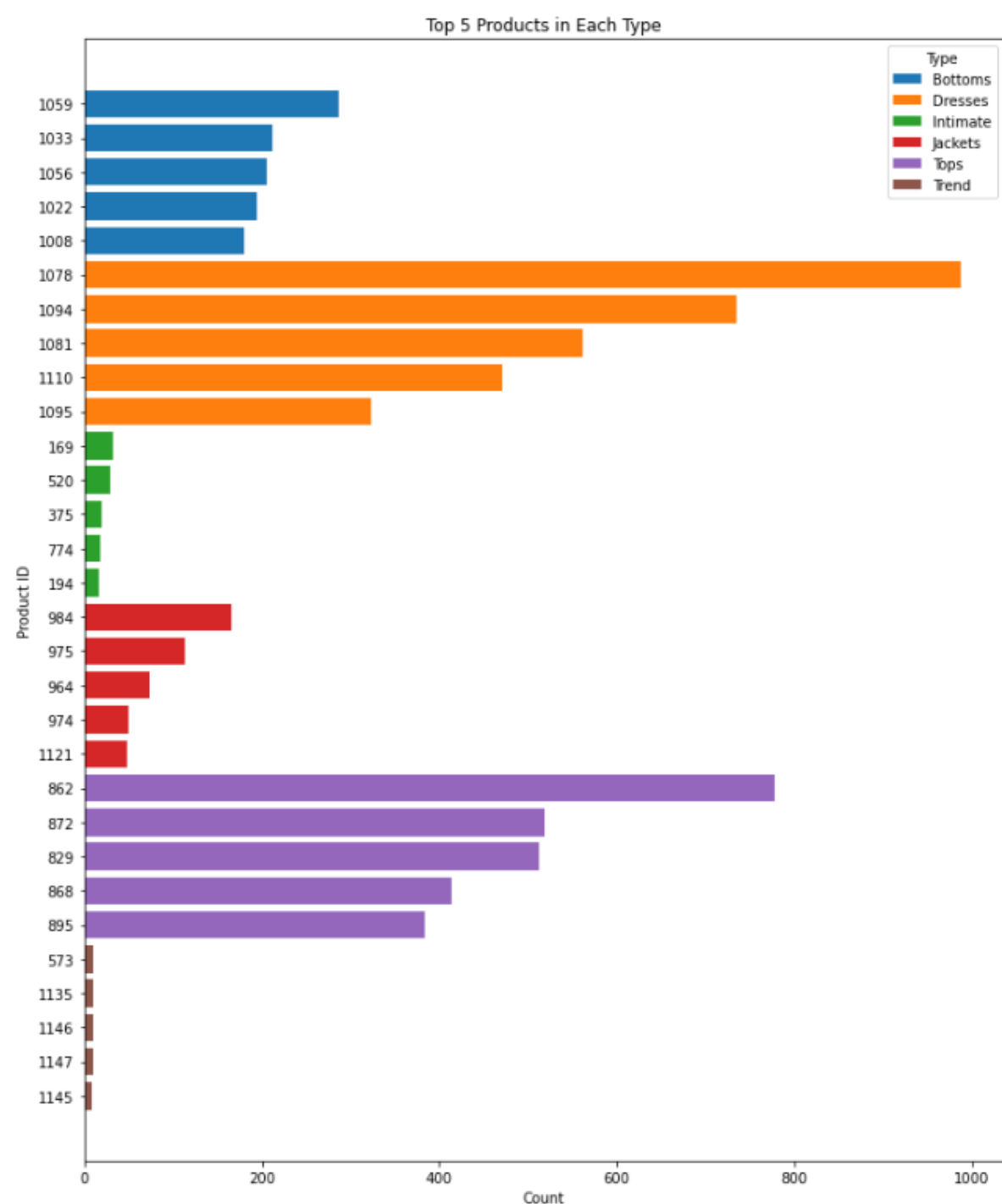
The mean age of the reviewers = 43



## Analysis: Product Type and Ratings

The product type of tops are most frequently reviewed. This can be observed in the product type distribution.

The average rating across each of the product types is shown in the following distribution of the ratings across the different product types.

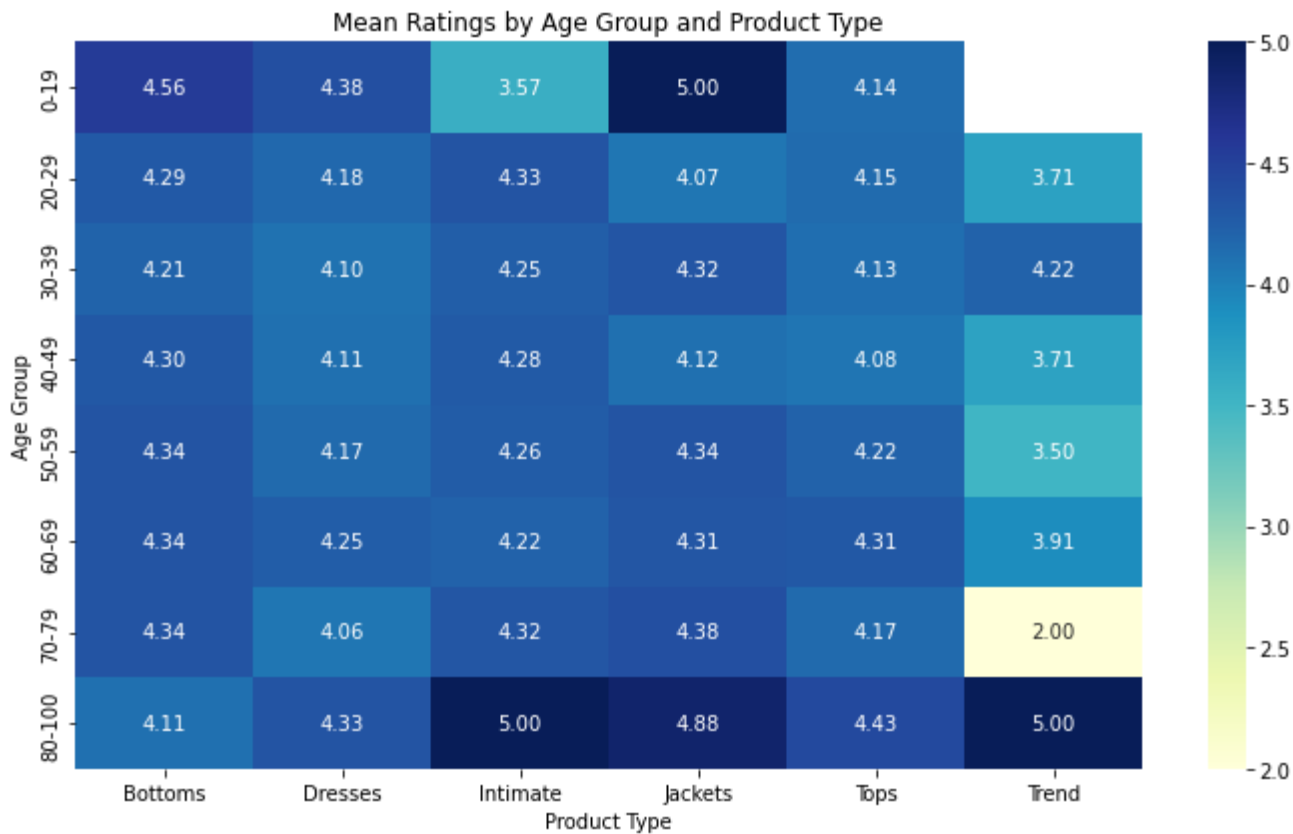


## Analysis: Product Type and Ratings

- The 5 most reviewed product\_ids within each type are shown.
- It can be noted that Product\_ID 1078 does have the most number of reviews, which is naturally the most reviewed under the category of dresses.
- The product ids can be used to identify the associated reviews to identify what went well which led to their popularity.

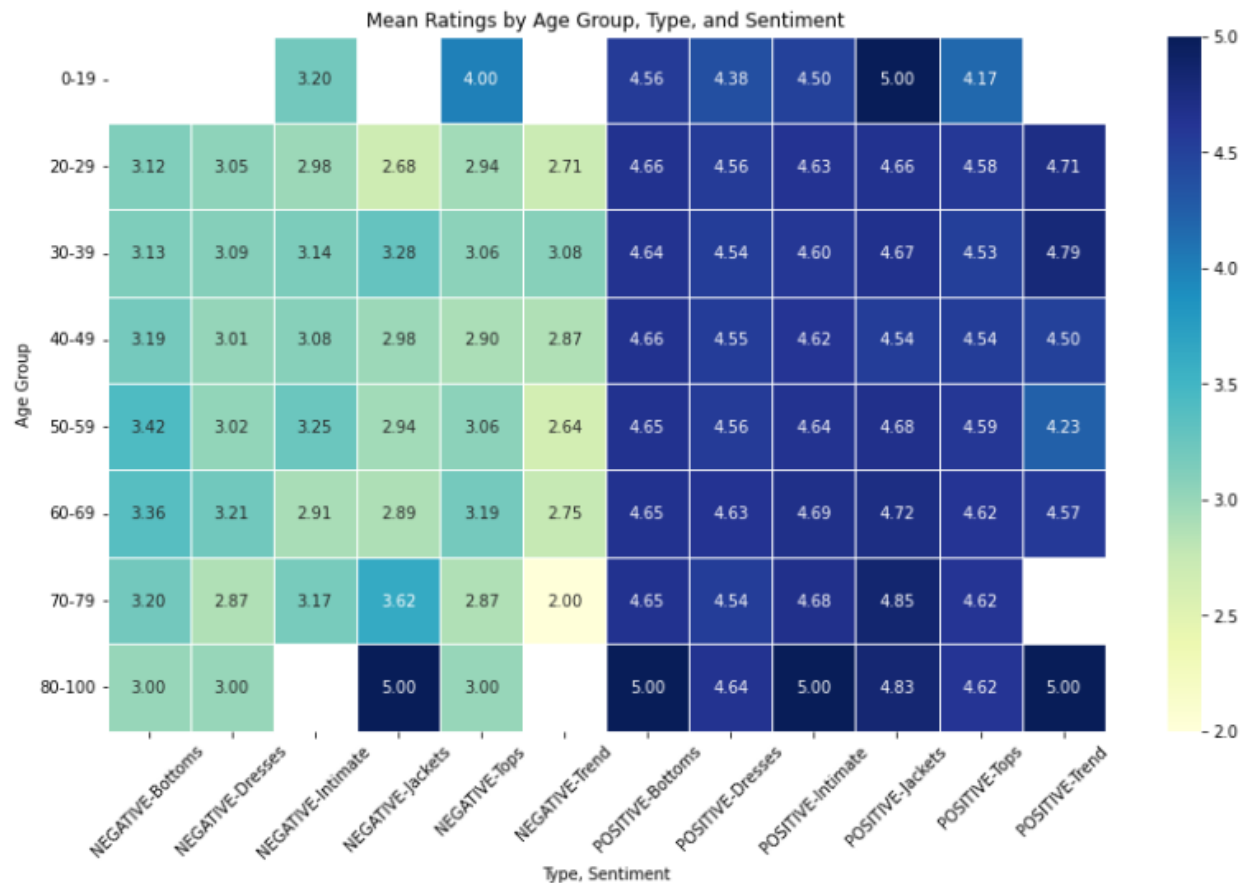


# Analysis: Age Groups, Rating and Reviews



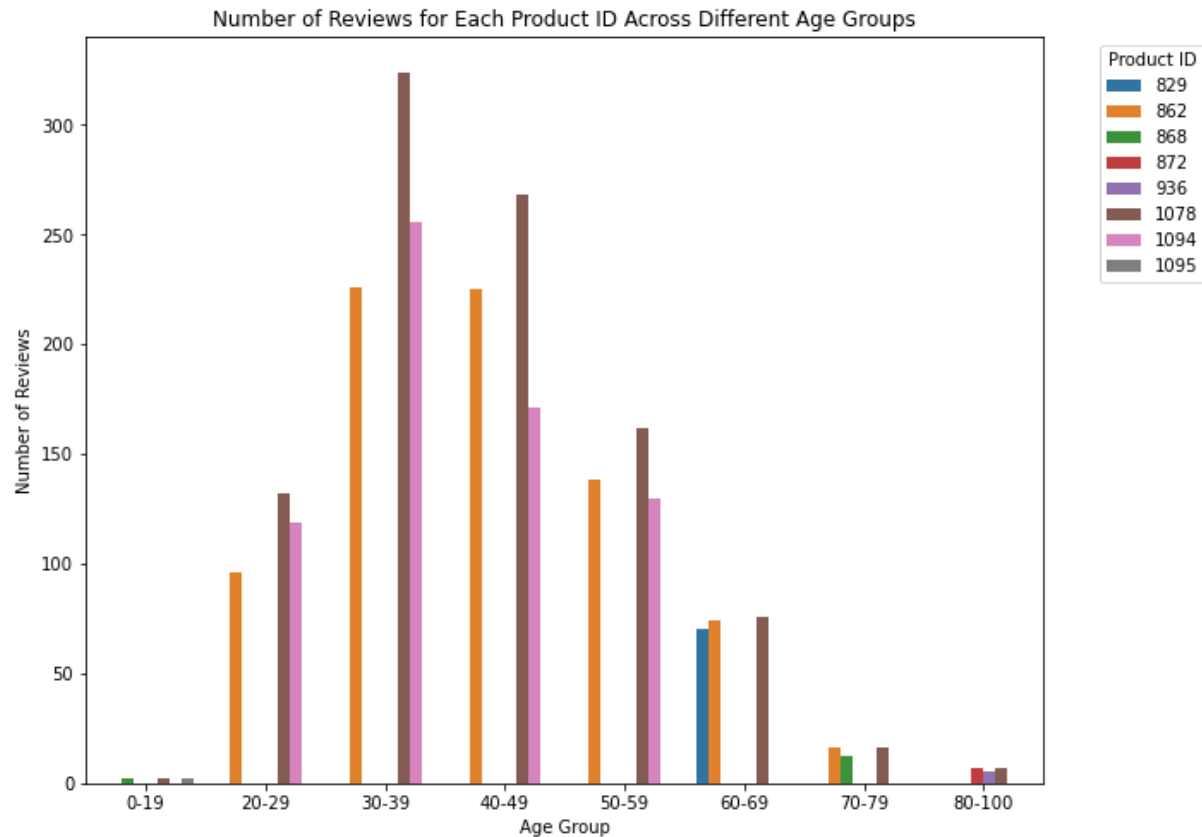
- Products under the category of tops, bottoms as well as dresses have a consistent rating across all the age groups.
- Products under the Trend category have a varied mean rating across all the age groups.
- The product type with the worst rating is that of Trend within the 70-79 age group.
- The age group of 0-19 did not review any products of the type trend and that of intimate were rated low. This age group preferred jackets the best. But again they are under-represented in the entire reviewers diaspora

# Analysis: Age Groups, Rating and Reviews



- The average rating to the right is higher (blue) which corroborates the positive sentiment of the reviews. This is valid across all product types as well as age groups.
- The average rating to the left is lower (green) which corroborates the negative sentiment of the reviews. This is valid across all but two product types-age groups category.
- The mean rating for tops under the 0-19 age group is high in the range of 4 but the sentiment states Negative, which is a contradiction.

# Analysis: Age Groups, Rating and Reviews



From the following plot of number of reviews for top three most reviewed products across different age groups, it is evident that **product id 1078** is the most reviewed across all the age groups consistently.

**Product 1094** and **Product 862** come close in four of the age groups.

# Analysis: Product Type, Rating and Reviews

Intimate

	Word	Frequency
0	love	605
1	size	589
2	like	490
3	wear	487
4	fit	463
5	great	436
6	small	415
7	would	385
8	top	370
9	soft	347

Dress

	Word	Frequency
0	dress	9432
1	size	2606
2	love	2312
3	fit	2101
4	like	1890
5	wear	1768
6	5	1754
7	would	1681
8	fabric	1631
9	great	1450

Bottom

	Word	Frequency
0	fit	1793
1	size	1767
2	love	1551
3	skirt	1257
4	pants	1247
5	great	1177
6	like	1149
7	5	1057
8	wear	1056
9	jeans	997

Top

	Word	Frequency
0	top	5710
1	love	3976
2	size	3315
3	like	3071
4	wear	2758
5	great	2722
6	fit	2588
7	shirt	2320
8	would	2315
9	color	2250

Jacket

	Word	Frequency
0	jacket	629
1	love	449
2	size	365
3	like	356
4	wear	314
5	coat	298
6	great	286
7	small	285
8	fit	282
9	would	243

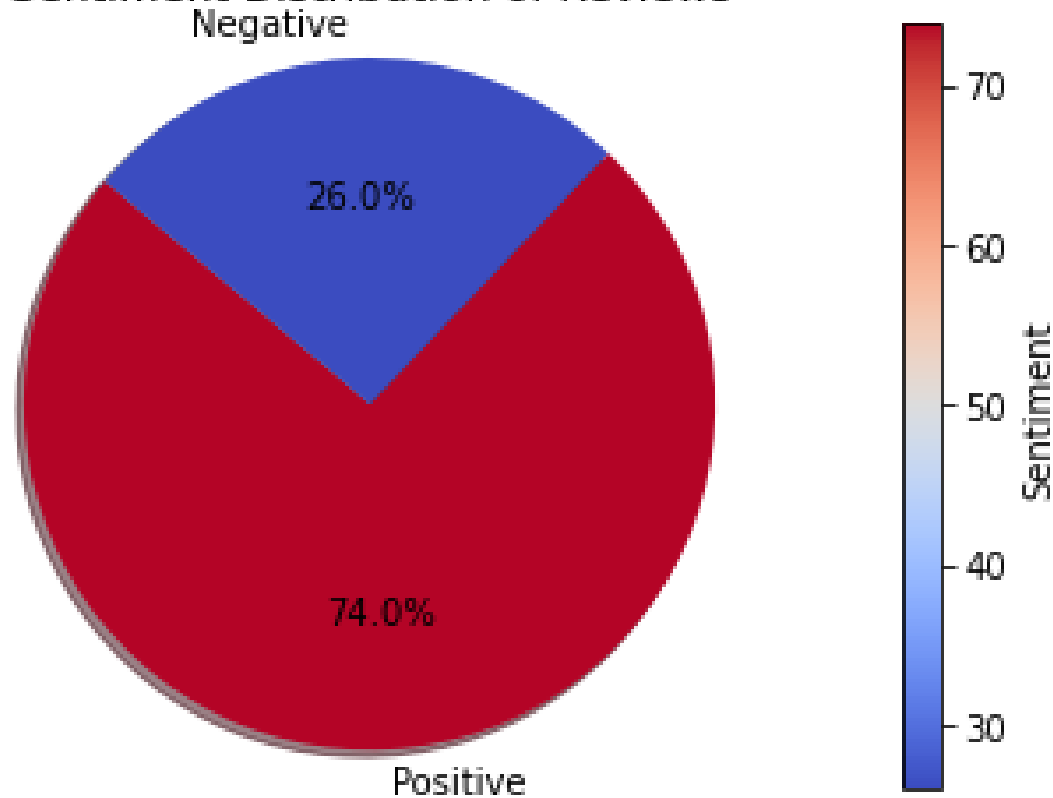
Trend

	Word	Frequency
0	dress	86
1	size	55
2	like	52
3	fit	44
4	fabric	39
5	love	38
6	would	37
7	look	35
8	wear	33
9	beautiful	32

The top 10 keywords within each product type

# Analysis: Reviews Sentiment

Sentiment Distribution of Reviews



The sentiment analysis of the text reviews reveals a predominantly **positive sentiment**, with **approximately 74%** of the reviews reflecting positive sentiment.

This suggests that the majority of the reviews express favorable opinions or experiences.

# Analysis: Reviews Sentiment

The keywords associated with different age groups do not show any trends of certain words.

Irrespective of the age groups, one of the most frequent keywords is dress.

A point to be noted is that with increasing age, of more than 60, the emphasis of the reviews move towards size and expression of likeness toward the product.

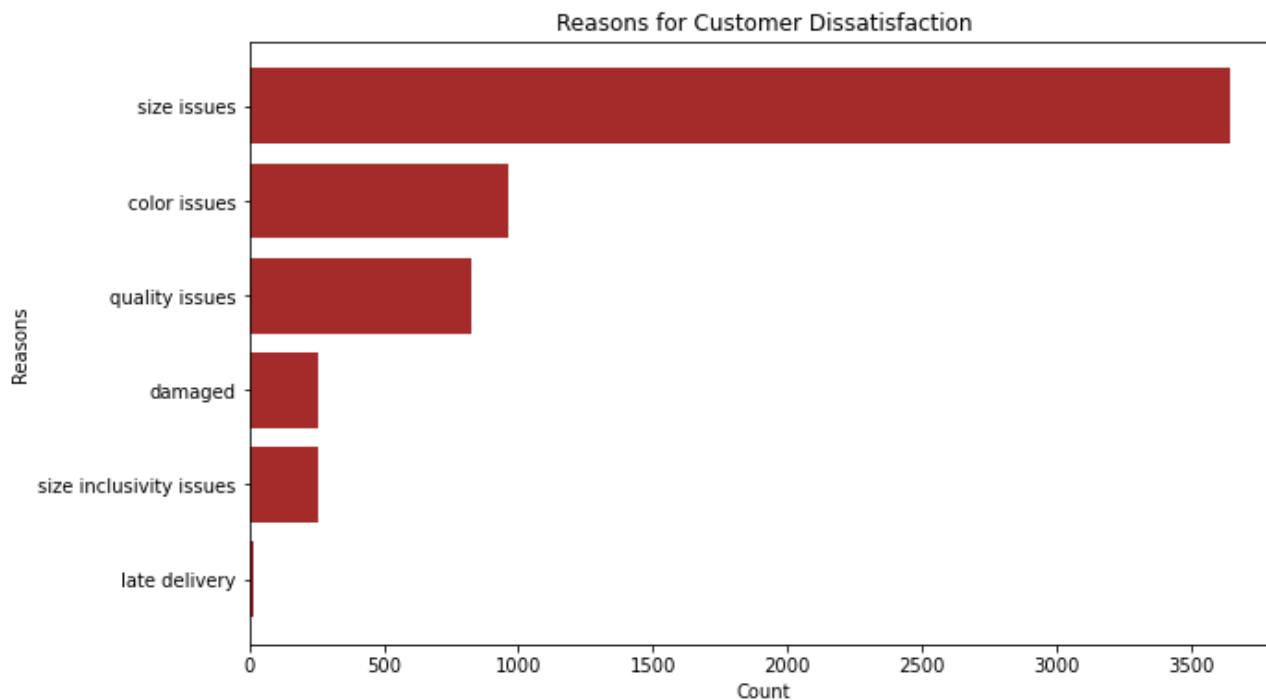
Age Group : 0-19   Age Group : 20-29   Age Group : 30-39   Age Group : 40-49

Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency
0 dress	29	0 dress	1592	0 dress	3508	0 dress	2660
1 like	20	1 love	1131	1 love	2873	1 love	2345
2 top	17	2 size	1016	2 size	2753	2 size	2309
3 size	16	3 top	943	3 like	2443	3 top	1868
4 really	15	4 fit	897	4 fit	2419	4 fit	1837
5 small	15	5 like	825	5 top	2270	5 like	1782
6 well	13	6 wear	811	6 great	2036	6 wear	1678
7 little	13	7 would	732	7 wear	2000	7 great	1596
8 fit	13	8 great	731	8 would	1877	8 would	1348
9 wear	13	9 5	611	9 5	1667	9 5	1313

Age Group : 50-59   Age Group : 60-69   Age Group : 70-79   Age Group : 80-100

Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency
0 dress	1677	0 love	922	0 size	143	0 size	65
1 size	1495	1 size	902	1 dress	142	1 love	62
2 love	1452	2 dress	839	2 love	140	2 dress	50
3 top	1295	3 top	800	3 wear	122	3 ordered	46
4 fit	1215	4 fit	770	4 top	112	4 top	43
5 like	1156	5 wear	686	5 fit	95	5 like	36
6 wear	1092	6 like	667	6 color	94	6 beautiful	31
7 great	992	7 great	618	7 great	94	7 color	30
8 5	914	8 fabric	586	8 small	86	8 would	30
9 fabric	877	9 color	489	9 ordered	84	9 soft	30

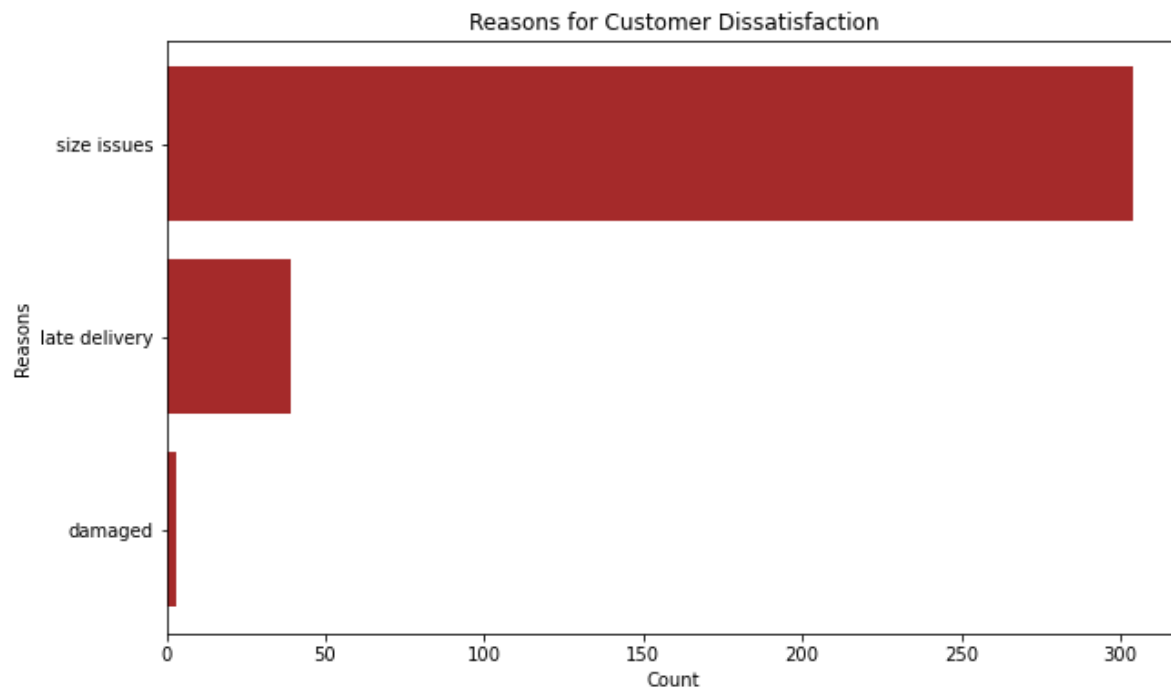
# Analysis: Negative Review Analysis



## Keyword Matching

- For the following potential reasons for negative reviews:
  - Late Delivery
  - Damaged Items
  - Size Issues
  - Color Issues
  - Quality Issues
  - Size Inclusivity Issues
- **Size issues** were the most mentioned reason within the negative reviews.

# Analysis: Negative Review Analysis



## spaCy's Named Entity Recognition

- For the following potential reasons for negative reviews:
  - Late delivery
  - Size issues
  - Damaged products
  - Color issues
  - Quality issues
- **Size issues** were the most mentioned reason within the negative reviews.





# Analysis: Negative Review Analysis

## Topic Modeling using LDA

This analysis identified the following topics:

- Topic 1: quality, fabric, look, like, design, dress, price, cute, poor, material
- Topic 2: size, dress, color, small, fit, like, ordered, look, wear, medium
- Topic 3: size, dress, fit, small, like, shirt, wear, look, love, ordered
- Topic 4: blocking, boring, disproportionate, paper, harder, robe, frankly, select, chemise, sooooo
- Topic 5: dress, look, like, fabric, fit, love, material, color, sweater, really
- Among these, size issues are a prevalent concern among customers, with a significant number of negative reviews expressing dissatisfaction with the sizing accuracy across product listings.

# Analysis: Positive Review Analysis



The most common keywords within the positive reviews can be visualized by this WordCloud

---

# Conclusion

- Currently, as per the negative reviews, the main issues include:
  - Size discrepancies and fit problems
  - Dissatisfaction with quality and color
- As per the positive reviews, which are considerably more in number, the customers are overall satisfied with the products' quality and fit.
- Considering both the pain points and the positive feedback, it's evident that improvements in several areas could enhance the overall customer experience:
  - Ensuring better representation of products on the retailer's website is crucial.
  - Moreover, prioritizing the quality of products is essential to address concerns raised in negative reviews regarding fit, color, and overall quality.



Thank You