Women's E-Commerce Clothing Reviews Sentiment Analysis

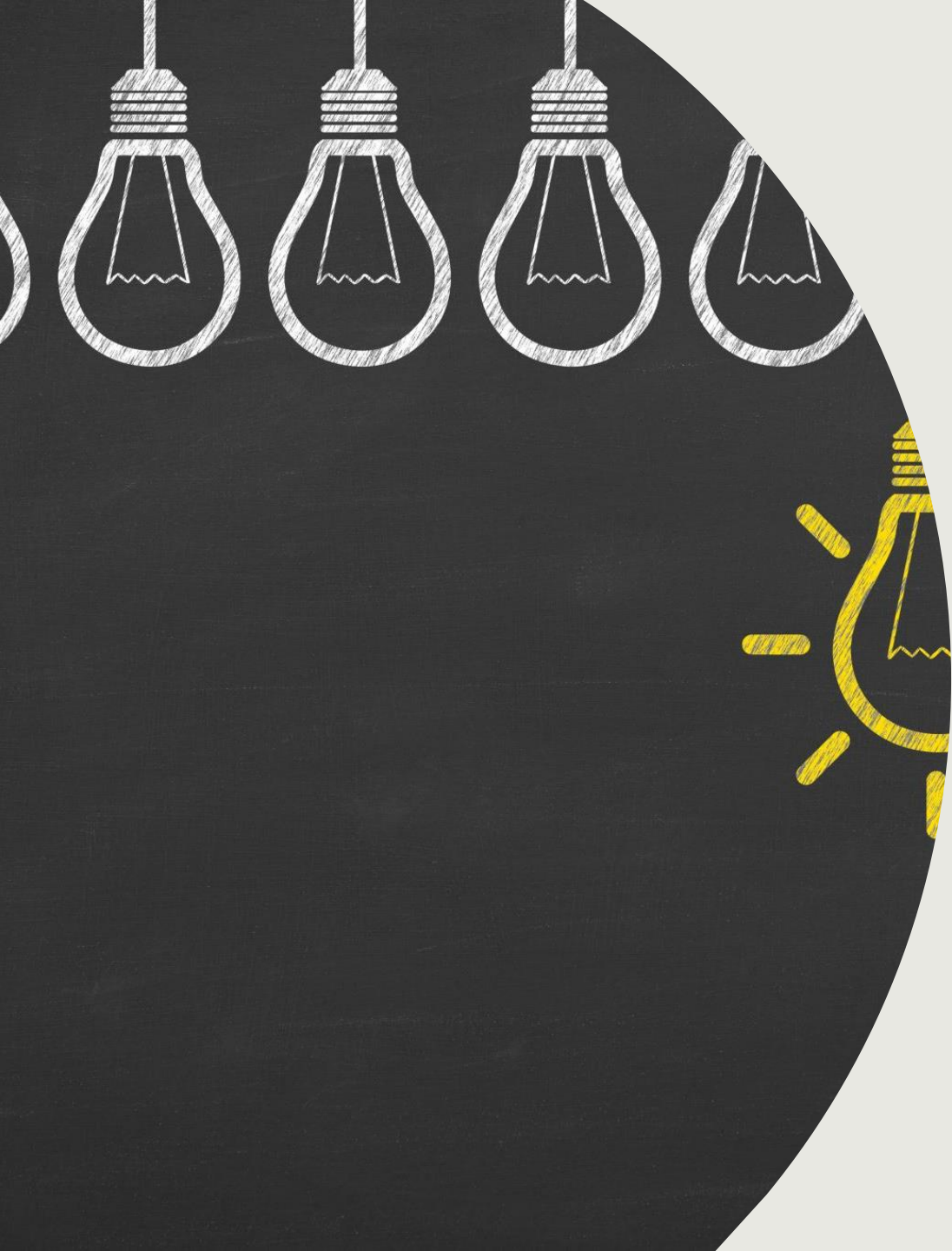Tina Huang

Date : 2021-07-18

Table of Content

Business Objective

Online shopping becomes more and more popular especially during the coronavirus pandemic. More than ever customers have something to say with the products after purchasing.

Sentiment analysis can be used to gauge customer response. It helps companies to improve the products according to the customers feedback.
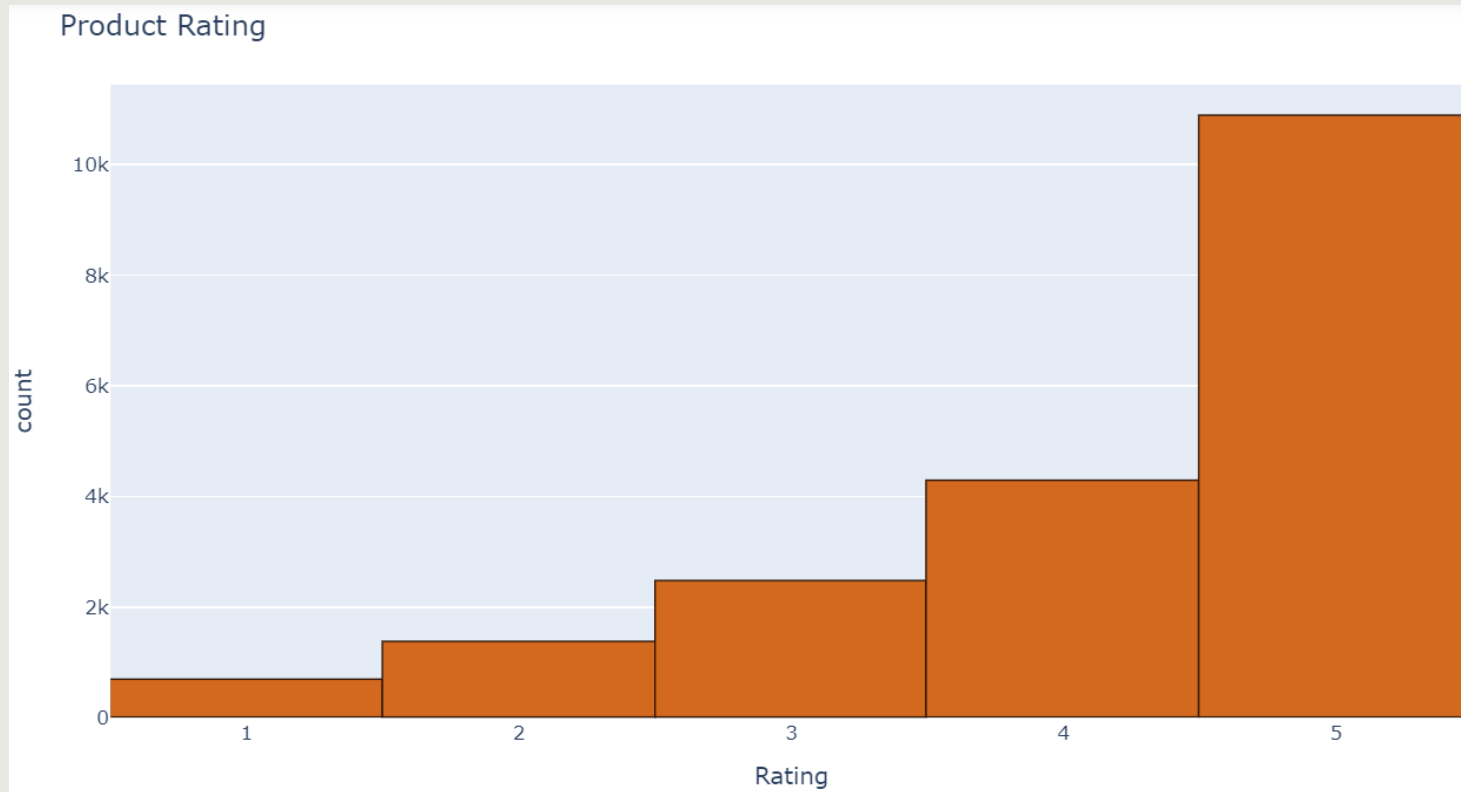
Business Task

1. How to improve products based on customers reviews?

2. How do customers respond to the products?

3. What are the trends in the review data?

Distribution of Rating

- Most of the customer rating is positive.
- The result leads me to believe that most text reviews will be positive, too.

What are the most frequently used words in the reviews?

- **love, like** - customers usually have positive reviews
- **dress** - dress might be a popular product
- **fit, size** - customers care about the size of the product

Sentiment classification

```python
df = df[df['Rating'] != 3]
df['sentiment'] = df['Rating'].apply(lambda rating : +1 if rating > 3 else -1)

# Split df - positive and negative sentiment:
positive = df[df['sentiment'] == 1]
negative = df[df['sentiment'] == -1]
```

- "Rating" = 3 → were dropped

- "Rating" > 3 → positive (+1)

- "Rating" < 3 → negative (-1)

	Title	Review Text	Rating	sentiment
3	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1
4	Flattering shirt	This shirt is very flattering to all due to th...	5	1
5	Not for the very petite	I love tracy reese dresses, but this one is no...	2	-1
6	Cagrcoal shimmer fun	I aded this in my basket at hte last mintue to...	5	1
7	Shimmer, surprisingly goes with lots	I ordered this in carbon for store pick up, an...	4	1
8	Flattering	I love this dress. i usually get an xs but it ...	5	1
9	Such a fun dress!	I'm 5"5' and 125 lbs. i ordered the s petite t...	5	1
12	Perfect!!!	More and more i find myself reliant on the rev...	5	1
13	Runs big	Bought the black xs to go under the larkspur m...	5	1
15	Nice, but not for my body	I took these out of the package and wanted the...	4	1

Stop words

- Remove "beautiful," "cute" and "great" because they were included in the negative sentiment.

```python
from nltk.corpus import stopwords
stopwords = set(stopwords.words('english'))
stopwords.update(['beautiful','cute','great'])
```

Word cloud — Positive Sentiment

- Popular words: "dress," "love," "perfect," "pretty," and "lovely."

- Reviews with rating 4-star or 5-star in the dataset express a positive sentiment.

Word cloud — Negative Sentiment

- Popular words: "disappointed," "wanted," "love" and "fit."

- Why the words like "wanted," "love" and "like" appeared on the negative word cloud with a high proportion?

Negative reviews contain the words "wanted," "love" or "like"

	Title	Review Text	Rating	sentiment
22	Not what it looks like	First of all, this is not pullover styling. th...	2	-1
68	I wanted to love this top	I really loved this top online and wanted to l...	2	-1
241	Loved this then washed it twice and it fell apart	I loved this dress from the moment i tried it ...	1	-1
581	Do not like the front	The back of this t-shirt looks cute but the fr...	2	-1
603	Wanted to love but sadly will return	These run very small!! they are also short, al...	1	-1
793	If only it looked like the photo	This sweater was a big let down. i am 5'2" so ...	1	-1
806	Wanted to love it - but	I wanted to love this. however, the fit was fu...	2	-1
880	So wanted to love it	Ditto what the first reviewer said, unfortunat...	2	-1
1061	Farm animal pullover	Wanted to love this sweater for its beautiful ...	2	-1
1063	I wanted to love this sweater but	This is a beautiful sweater with deep rich col...	2	-1
1257	Wanted to love	I wanted to love this top, but it unfortunatel...	2	-1
1300	Wanted to love	I bought two, one in white and one in blue pri...	2	-1
1306	Fits like maternity	I wanted to love this, was so excited when i o...	2	-1

- There are 169 out of 2051 (8.24%.) negative reviews contain the words "wanted," "love" or "like."

- Customers wanted to love the product, but something stopped them to do this.

- Some reviews contain the words "looks like."

Negative reviews – Size issue

	Title	Review Text	Rating	sentiment
88	Huge	Really cute piece, but it's huge. i ordered an...	2	-1
97	Huge	I was very excited to order this top in red xs...	1	-1
118	Awkward fitting	First, the fabric is beautiful and lovely for ...	2	-1
185	Fits strange, flimsy material	I was worried about this item when i ordered i...	2	-1
334	Not a good fit	This top was way too short (i'm only 5'1) and ...	1	-1
...	...	...	...	...
22657	Strange fit	The pattern and fabric are lovely, but the sle...	2	-1
22803	Drab color, weird fit	Couldn't make it work, even belted. ivory colo...	2	-1
22844	Sweater is huge!!	Had to return. ordered an xs which usually fit...	2	-1
23016	Large fit	I was really excited to get this dress and to ...	2	-1
23405	Very big, scratchy	I'm petite, and i knew buying this that it wou...	2	-1

243 rows × 4 columns

- There are 243 out of 2051 (11.85%) negative reviews contain the words "fit," "big," or "huge."

- Provide the detail of size for each product and make an accurate recommendation for customers.

Negative reviews – Quality issue

	Title	Review Text	Rating	sentiment
110	Poor quality	This is so thin and poor quality. especially f...	1	-1
114	Disappointing quality	The design/shape of the dress are quite flatte...	2	-1
389	Bad quality	I don't normally review my purchases, but i wa...	1	-1
475	Poor quality	I bought this in the white, size m (140lbs., 5...	2	-1
1052	Quality & cut	My perspective is as a sewer so i'm more picky...	1	-1
...	...	...	...	...
23031	Nice fabric poor design	This dress has a beautiful fabric, i really ha...	2	-1
23255	Looks great poor quality	Pants are overpriced and very uncomfortable. t...	1	-1
23274	Cute but poor quality	I'm 5'7" 155lbs and ordered a large. it fit, b...	1	-1
23287	Too much fabric shoulders wont stay down	This is a gorgeous print, but i found it too f...	2	-1
23437	Not typical of retailer quality	This dress is so cute in the photo and fit tru...	2	-1

155 rows × 6 columns

- There are 155 out of 2051 (7.56%) negative reviews contain the words "quality" or "fabric."

- Communicate with the manufacturer to upgrade the quality of products and examine clothing materials.
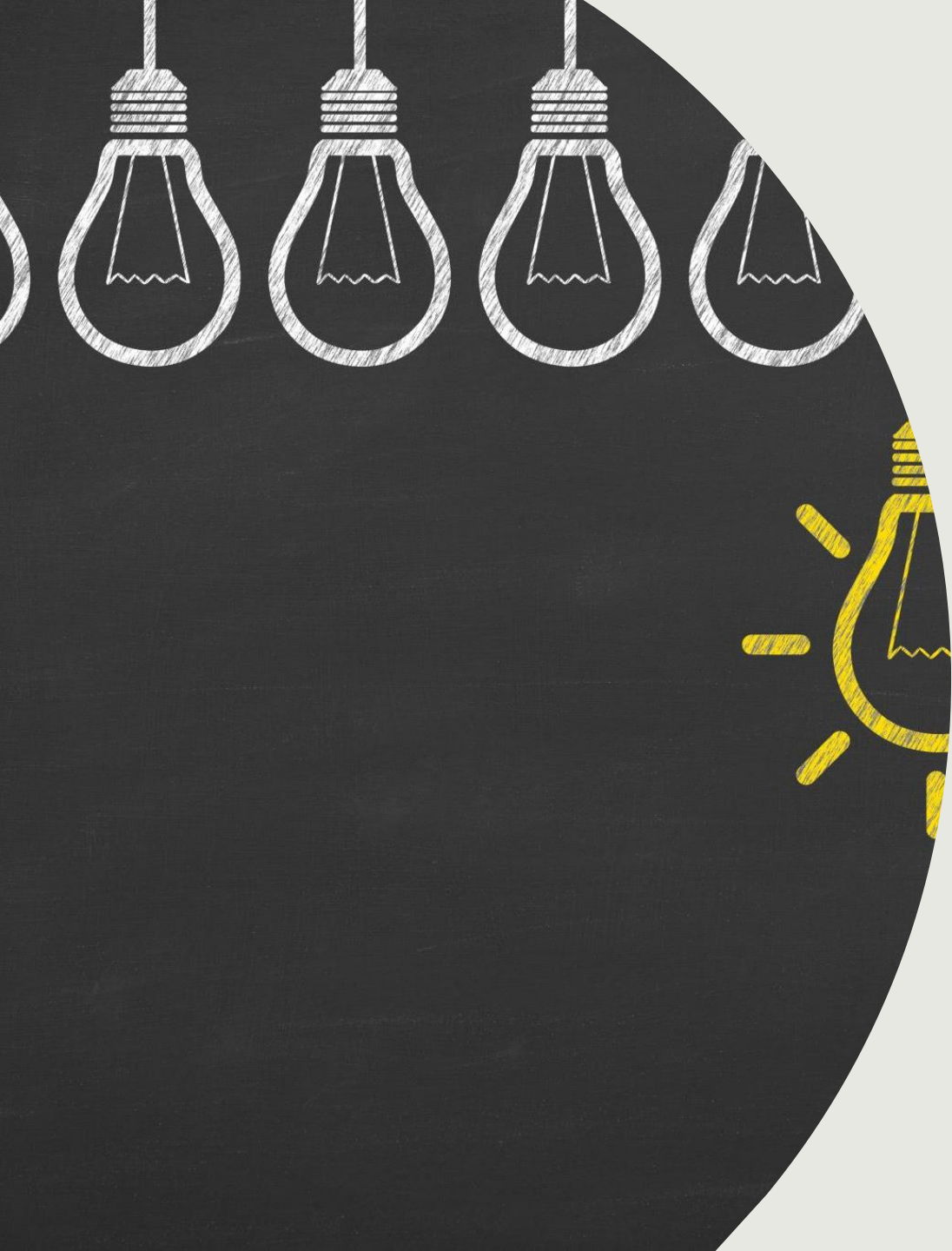
Negative reviews – Color issue

	Title	Review Text	Rating	sentiment
518	Weird color and fit	Color is not like photo and fit doesn't work i...	2	-1
1618	Great quality and color but boxy and ill-fitting	I wanted to love this sweater. the photo of th...	2	-1
2675	Color run	I love the way these shorts fit and look; very...	1	-1
3552	Way too big/colors off	Finally received this in the mail today and wa...	2	-1
5450	Orange color is pretty, quality is poor	I ordered the orange in size xl. the fit on me...	2	-1
6055	Color transfers to other clothes!	I really loved this vest online and decided to...	2	-1
9589	Colors are more intense than they appear	I purchased this dress for my graduation. the ...	2	-1
9765	Color not as shown	The dress is very pretty, but the sequins are ...	2	-1
9778	Color is not as it appears in the photo	I was very surprised to see such dark blue seq...	2	-1
9857	Color bleeds and discolors the pant.	This is a problem i have had with the pilcro h...	2	-1
9867	Not colorfast!	I'm so bummed! i absolutely loved the pair of ...	2	-1
10084	Sloppy looking but pretty colors	I ordered this style in the blue, red, yellow ...	2	-1
10793	Super cute on, colors completely faded	Love this dress! wore it last night for a musi...	2	-1
11570	Weird color	Not flattering and not the color it appears on...	2	-1
12110	Color not as pictured	I returned the dress because the color was not...	2	-1
12504	Color not as pictured on website	I wish i had read the other review about the c...	2	-1

- There are 30 out of 2051 (1.46%) negative reviews contain the words "color."

- There are two main problems:

1. Color is not as pictured

2. Discolored after one wash

Negative reviews – Color issue

12642	Beautiful suede, but colors are not correct.	Just received these and they fit perfectly, bu...	2	-1
13172	A colorful big box	Since this sold out so fast i ordered an xsp w...	1	-1
13586	Color not as pictured	I had high hopes for this top based on the pho...	1	-1
13594	Colors look nothing like online photos:(	When i saw this top online, i thought i'd love...	1	-1
13832	Wrong color and too big	I was so excited for this dress to come. i am ...	2	-1
14504	Very thin material. nice colors	I agree with the other reviewer about the grea...	1	-1
15978	Color became spotty after washing	I bought a pair of these in stone and after th...	1	-1
15990	Not colorfast at all!	Bought the gold pair and dripped something on ...	1	-1
15992	Color ruins easily	I bought a pair of these a month ago and first...	1	-1
16014	Discolored after one wash.	Similar to others here, these pants were ruine...	1	-1
16773	Cream colored	The top is more of a cream color than an ivory...	2	-1
17566	Did not like the fit, but beautiful colors	I really wanted to love this shirt! the colors...	2	-1
18478	Grey not color as pictured on website	I wanted to love this top, but when i received...	2	-1
18599	Huge jacket, wash color is not as pictured	Like one of the previous reviewers, when my ja...	1	-1

- Make the pictures look like the products
- Specify the color correctly on the website to avoid misleading.
- Communicate with the manufacturer for the discolored issue and fix the problem.

Sentiment analysis Model

Built the sentiment analysis model to make a prediction on whether the review is positive or negative.

Remove punctuations

- Remove all punctuations from the column "Title."

```python
def remove_punctuation(text):
    final = "".join(u for u in text if u not in (",", "?", ".", ";", ":", "!",'"',"'"))
    return final
df['Title'] = df['Title'].apply(remove_punctuation)
```

	Title	sentiment
3	My favorite buy	1
4	Flattering shirt	1
5	Not for the very petite	-1
6	Cagrcoal shimmer fun	1
7	Shimmer surprisingly goes with lots	1

Training and testing data

- 80% of the data was used for training, and 20% was used for testing.

```python
index = df.index
df['random_number'] = np.random.randn(len(index))
train = df[df['random_number'] <= 0.8]
test = df[df['random_number'] > 0.8]
```

Logistic regression

- Count vectorizer from the Scikit-learn library.

- Built a simple logistic regression model and trained the data on it

```python
# Count vectorizer:

from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer(token_pattern=r'\b\w+\b')
train_matrix = vectorizer.fit_transform(train['Title'])
test_matrix = vectorizer.transform(test['Title'])
```

```python
# Logistic regression

from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
```

```python
X_train = train_matrix
X_test = test_matrix
y_train = train['sentiment']
y_test = test['sentiment']
```

```python
# Fit model on data

lr.fit(X_train,y_train)
```

Classification report

- The overall accuracy of the model on the test data is approximately 93%.

	precision	recall	f1-score	support
-1	0.49	0.83	0.62	269
1	0.99	0.93	0.96	3410
accuracy			**0.93**	3679
macro avg	0.74	0.88	0.79	3679
weighted avg	0.95	0.93	0.93	3679

Appendix

- Data source : Women's Clothing E-Commerce dataset

- Reference

  1. Hacking The Product Rating Race With Sentiment Analysis (2021)

  2. A Beginner's Guide to Sentiment Analysis with Python

Thank you