Women's E-Commerce Clothing Reviews Sentiment Analysis

Tina Huang

Date : 2021-07-18

Table of Content

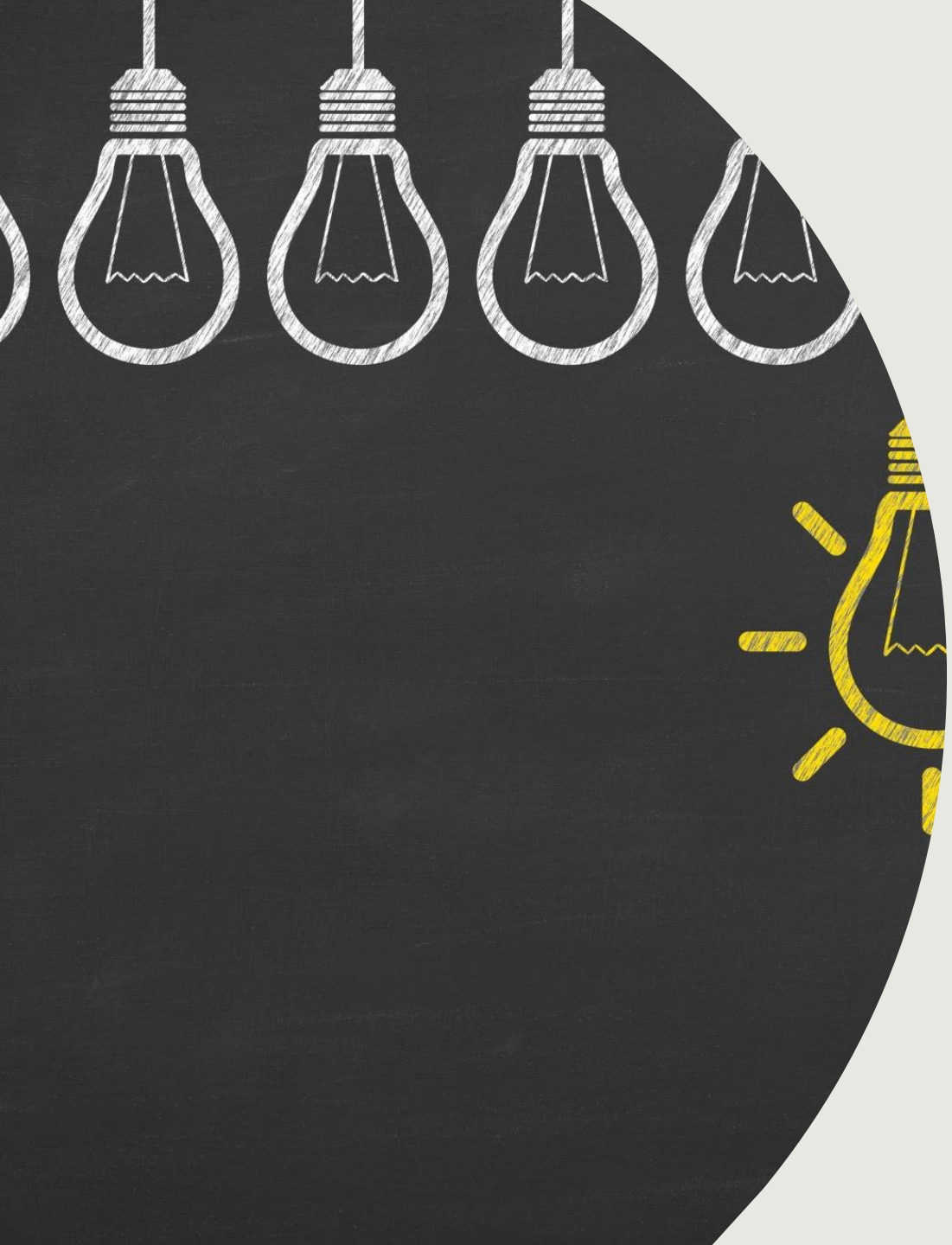
Business Objective

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

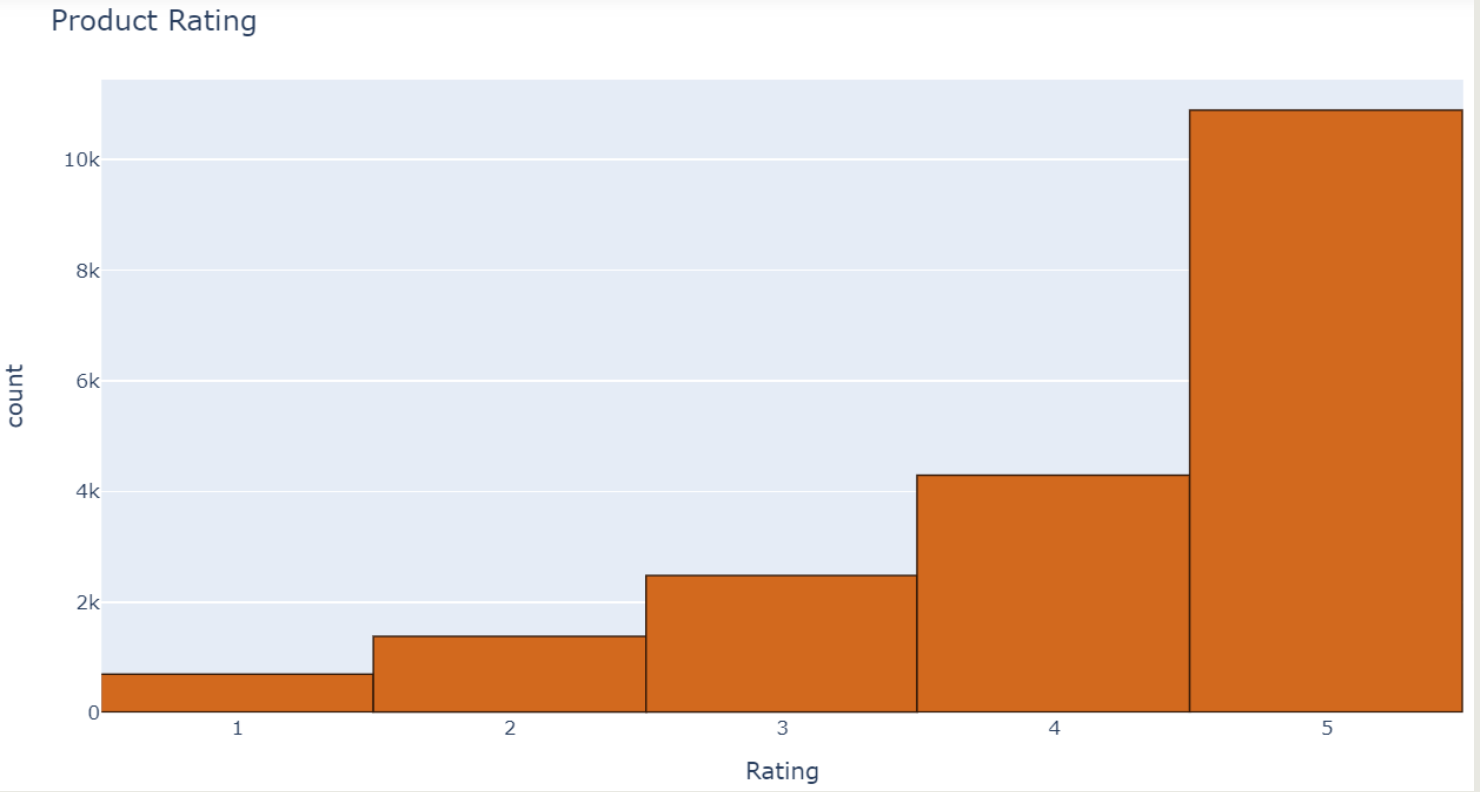
1. How to improve products based on customers reviews?
2. How do customers respond to the products?
3. What are the trends in the review data?



Analysis and Visualizations

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

```
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.

```
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” and “love” appeared on the negative word cloud with a high proportion?

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

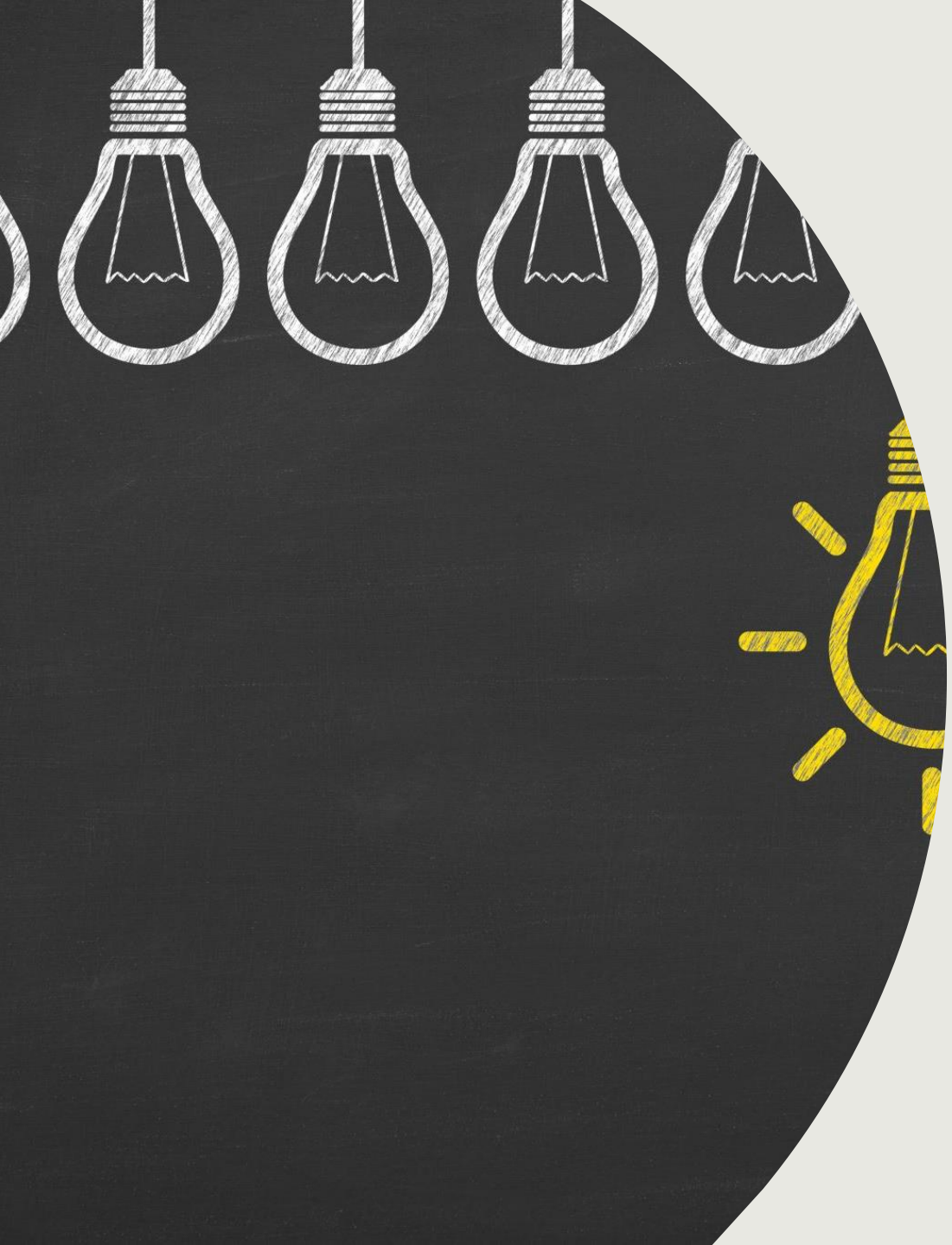
Title	Review Text	Rating	sentiment
I wanted to love this top	I really loved this top online and wanted to l...	2	-1
Wanted to love but sadly will return	These run very small!! they are also short, al...	1	-1
Wanted to love it - but	I wanted to love this. however, the fit was fu...	2	-1
So wanted to love it	Ditto what the first reviewer said, unfortunat...	2	-1
Farm animal pullover	Wanted to love this sweater for its beautiful ...	2	-1
I wanted to love this sweater but	This is a beautiful sweater with deep rich col...	2	-1
Wanted to love	I wanted to love this top, but it unfortunatel...	2	-1
Wanted to love	I bought two, one in white and one in blue pri...	2	-1
I wanted to love it	I really like the look of the top when it is p...	2	-1
Wanted to love it	An absolutely gorgeous, but poorly fitting blo...	2	-1
Not what i wanted	I am reasonably petite, but i like sweaters ve...	2	-1

- There are 72 out of 2051 (3.51%.) negative reviews contain the words "wanted" or "love."
- Most of the reviews contain the words "wanted" and "love" in a comment simultaneously.

Negative reviews contain the words "fit," "big," and "huge"

Title	Review Text	Rating	sentiment
Awkward fitting	First, the fabric is beautiful and lovely for ...	2	-1
Not a good fit	This top was way too short (i'm only 5'1) and ...	1	-1
Cute concept fits weird	The fun colors drew me to this but it sure fit...	2	-1
Weird color and fit	Color is not like photo and fit doesn't work i...	2	-1
Weird fit	I tried this on in the store and was disappoin...	2	-1
...	...	...	...
Strange fit	The pattern and fabric are lovely, but the sle...	2	-1
Drab color weird fit	Couldn't make it work, even belted. ivory colo...	2	-1
Sweater is huge	Had to return. ordered an xs which usually fit...	2	-1
Large fit	I was really excited to get this dress and to ...	2	-1
Very big scratchy	I'm petite, and i knew buying this that it wou...	2	-1

- There are 183 out of 2051 (8.92%) negative reviews contain the words "fit," "big," and "huge."
- "Size" is a big problem for the business.



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.”

```
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.

```
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

```
# 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'])
```

```
# Logistic regression

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

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

```
# 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