Women's Fashion Ratings and Reviews

T6 Data Science BootCamp

Manal Abdulaziz Alhayan

2021-11-16

Table of Content:

Goal:

Improving products based on customers' reviews by applying sentiment analysis using Machine learning methods to determine if customers are satisfied or dissatisfied with their purchases.

1. Sentiment Analysis Model

1.1 Data Processing:

1. Cleaning the text by removing all punctuations, numbers, a word that has length <= 2 , stop words and lowering the word.
2. Labeling the text by the rating feature:
    - rating >3 : Positive,
    - rating <3 : Negative.
3. Adding extra feature to count common words.

1. Sentiment Analysis Model

1.2: Model Selection and Evaluation :

Naive Bayes (NB) is used to classify the text reviews. The official metric was the F1-score since we have imbalanced dataset as shown in the first figure. The overall evaluation is shown in the second figure.

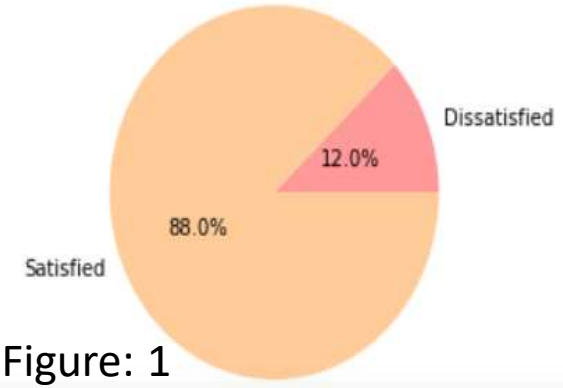


Figure: 1

	precision	recall	f1-score	support
0	0.72	0.57	0.64	434
1	0.95	0.97	0.96	3527
accuracy			0.93	3961
macro avg	0.84	0.77	0.80	3961
weighted avg	0.92	0.93	0.92	3961

Figure: 2

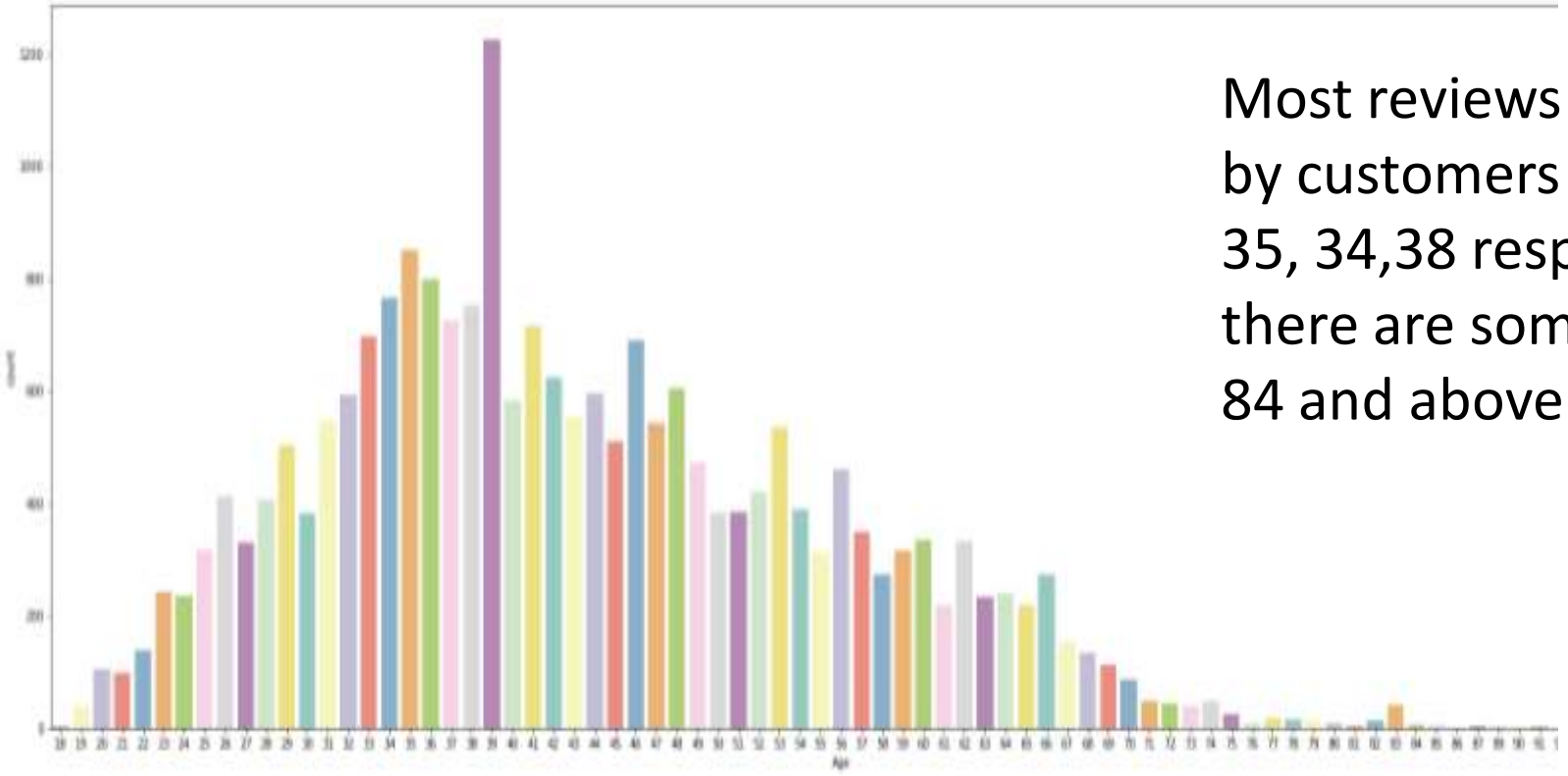
2. Analyzing Engineered Features

2.1 Tasks and objectives :

- 1. Which age group write the most in online reviews?
- 2. What each age of group prefers to buy ?
- 3. What are the most selling clothes ?
- 4. What most customers complain about ?
- 5. What are the reasons of products dissatisfaction?

2. Analyzing Engineered Features
2.1.1 visualization

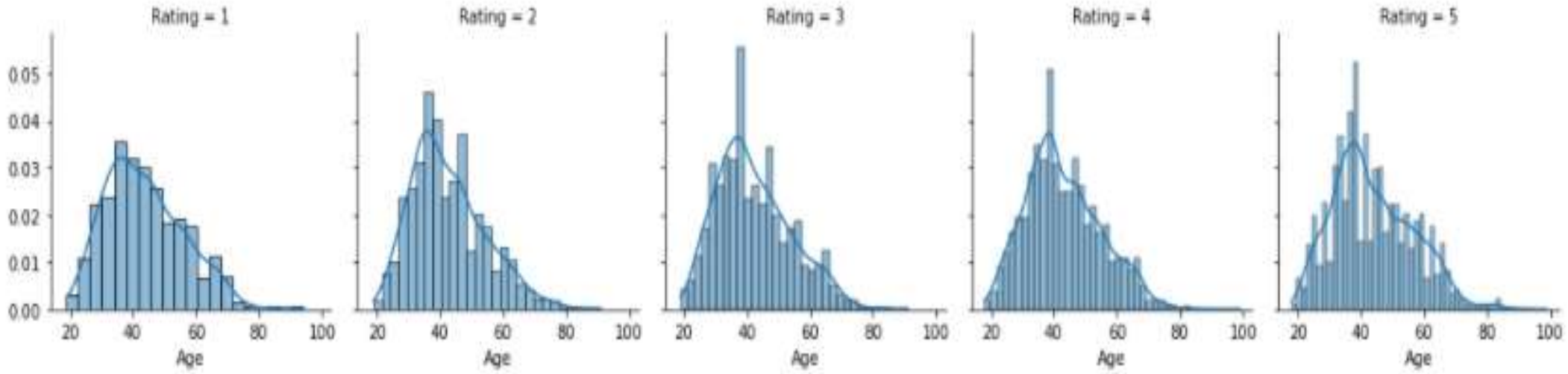
1. Which age group write the most in text reviews?



Most reviews have been posted by customers whose ages are 39, 35, 34,38 respectively, whereas there are some outliers like ages 84 and above.

2. Analyzing Engineered Features
2.1.1 visualization

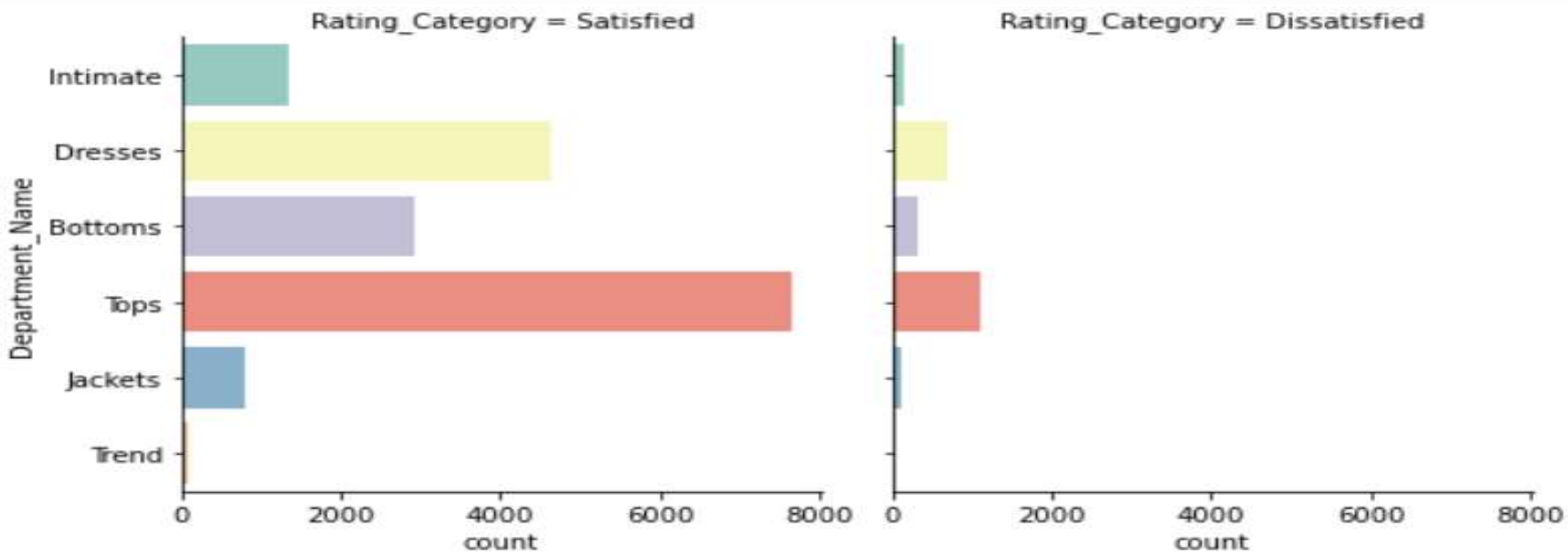
2. What each age of group prefers to buy ?



We can notice that the distribution in each rating shows similar variations between ages, meaning no age group have preferred some products more than others.

2. Analyzing Engineered Features
2.1.1 visualization

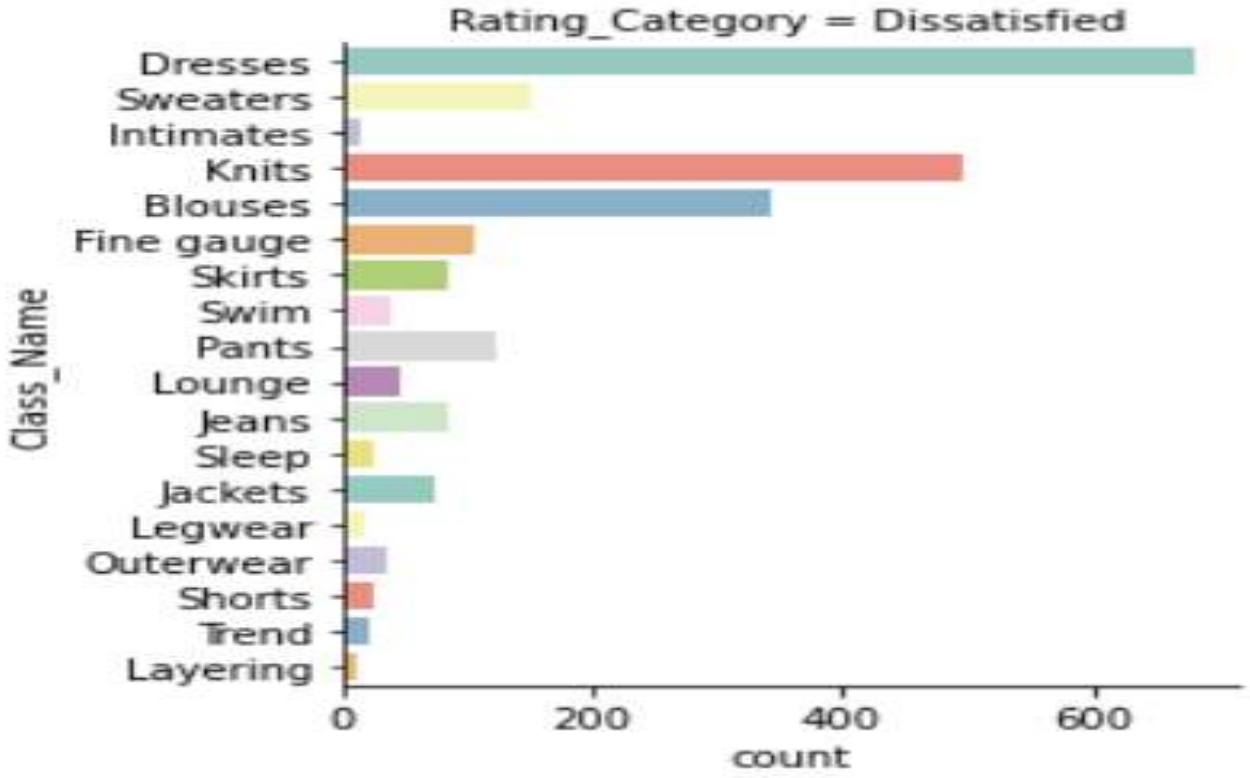
3. What are the most selling clothes ?



Most selling clothes are tops, dresses and bottoms.

2. Analyzing Engineered Features
2.1.1 visualization

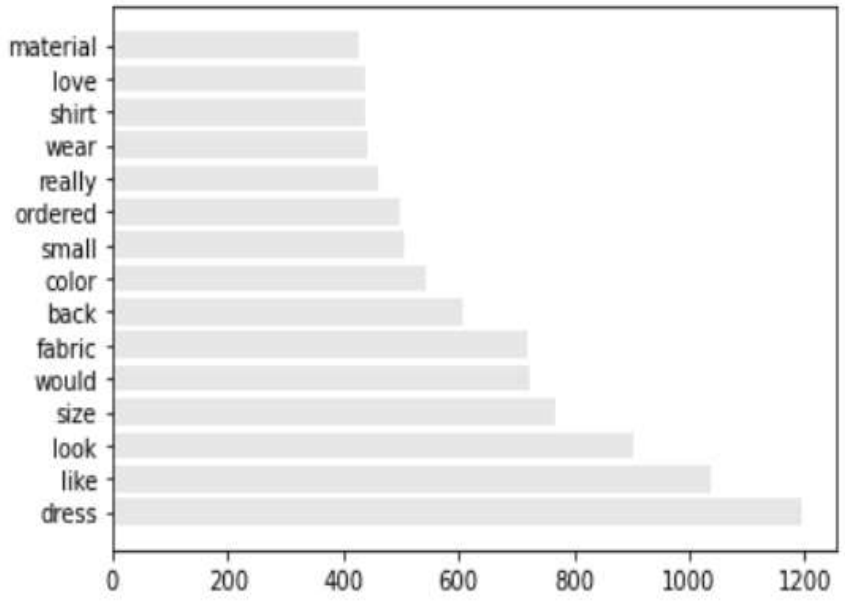
4. What most customers complain about ?



Most customers have been dissatisfied with three items which are dresses, knits and blouses.

2. Analyzing Engineered Features
2.1.1 visualization

5. What are the reasons of products dissatisfaction?



The graph illustrates 15 common words in reviews which might imply several things.

1. (size, small, wear)

Reason : complaining about sizes.

Recommendation: UKA size chart should be included.

2. (Fabric, Material, dresses)

Reason: complaining about quality of products like dresses.

Recommendation: Types of fabric should be included in detail.

3.(looked, look, like , love).

Reason: not the same as in the picture.

Recommendation: High resolution image should be posted.

Thank you