Exploring Women's Clothing Preferences and Sentiments to Drive Customer Satisfaction

Section 31 Team 2 - SP 2023

Baojia Huang

Yue Yao

Zhizhi Chen

Yao Zhong

Help Sellers to Improve Customer Satisfaction by Mining Customer Review Topics and Sentiments

Goal

Help e-commerce sellers to improve customer satisfaction by extracting customer review topics and sentiments

Insights that we plan to get

Analyze user review topics to understand **key factors** that affect customer satisfaction
eg. appearance, fit and size, price, quality

Build models to help sellers identify users' **emotional feedback and attitude** on products

Focus on **clothing departments** to specifically analyze detailed customer opinions on products
eg. bottoms, tops, dresses

E-commerce sellers can use insights to

Increase new customer acquisition:
Dig out customer needs and optimize products, to improve user experience & expand target audience

Retain existing customers:
Help to respond faster to customers' emotional feedback and retain existing customers

Optimize customer experience:
Strengths and weaknesses of individual departments can help sellers take targeted actions and enhance customer satisfaction

Use the Reviews Data of Women's E-Commerce Clothing on Kaggle

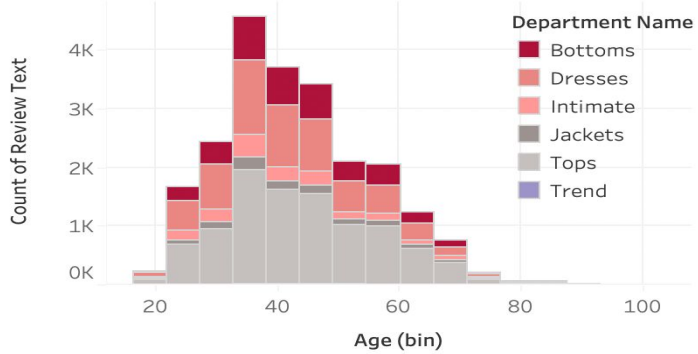
Data Description

- **Data Authentity:**
  Anonymous, Real
- **Data Source:**
  Kaggle
- **Data Size:**
  8.48 MB
- **Data Structure:**
  23486 rows 10 columns
  (including **Age, Review, Rating, Department, Class, Recommended or not**)

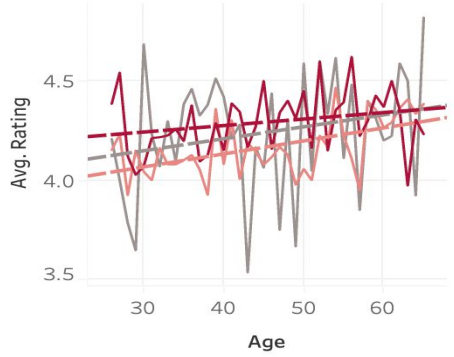
Shortcomings

- **Self-selection bias**:
  users volunteer their opinions and comments, people who are extremely positive or negative might be more likely to comment.
- **Time Frame outdatedness**:
  This dataset was collected six years ago, which may be biased to reflect the current customers' preference and the future trend.

Count of Reviews by Age



Avg Rating of Middle Aged



Highlight Table

Bottoms 4.2888 3,662	Intimate 4.2801 1,653
Jackets 4.2645 1,002	Tops 4.1722 10,048
Dresses 4.1508 6,145	Trend 3.8151 118

Build the Topic Model by Department and Fit Sentiment Analysis Model

Topic Modeling

Separate reviews by department (tops, bottoms, dresses)
Preprocess (Tokenization, stopwords, stemming) and exclusively extract the column of "review text"

Run LDA model, tune parameters and obtain the optimal number of topics for each department.
Use log-likelihood, confusion matrix and coherence score to find the optimal topic model (limit=5)

Name each topic with ChatGPT's help and explore improvement strategy for sellers.
Based on the "Label" of topics, find what users in each department care about

Sentiment Analysis

Create "polarity" based on the "Rating" column.
1 or 2 stars: Negative
3 stars: Neutral
4 or 5 stars: Positive

Use (unsupervised) lexicon-based methods like VADER & sentiment classifier (supervised)
Split the train & test, evaluate with metrics like precision and recall rate, tune the threshold

Adopt the optimal model for future prediction
Compute accuracy rate for the above methods, compare predicted and actual sentiment, make predictions, and visualize the sentiment distribution

THANK YOU !

Do you have any questions?