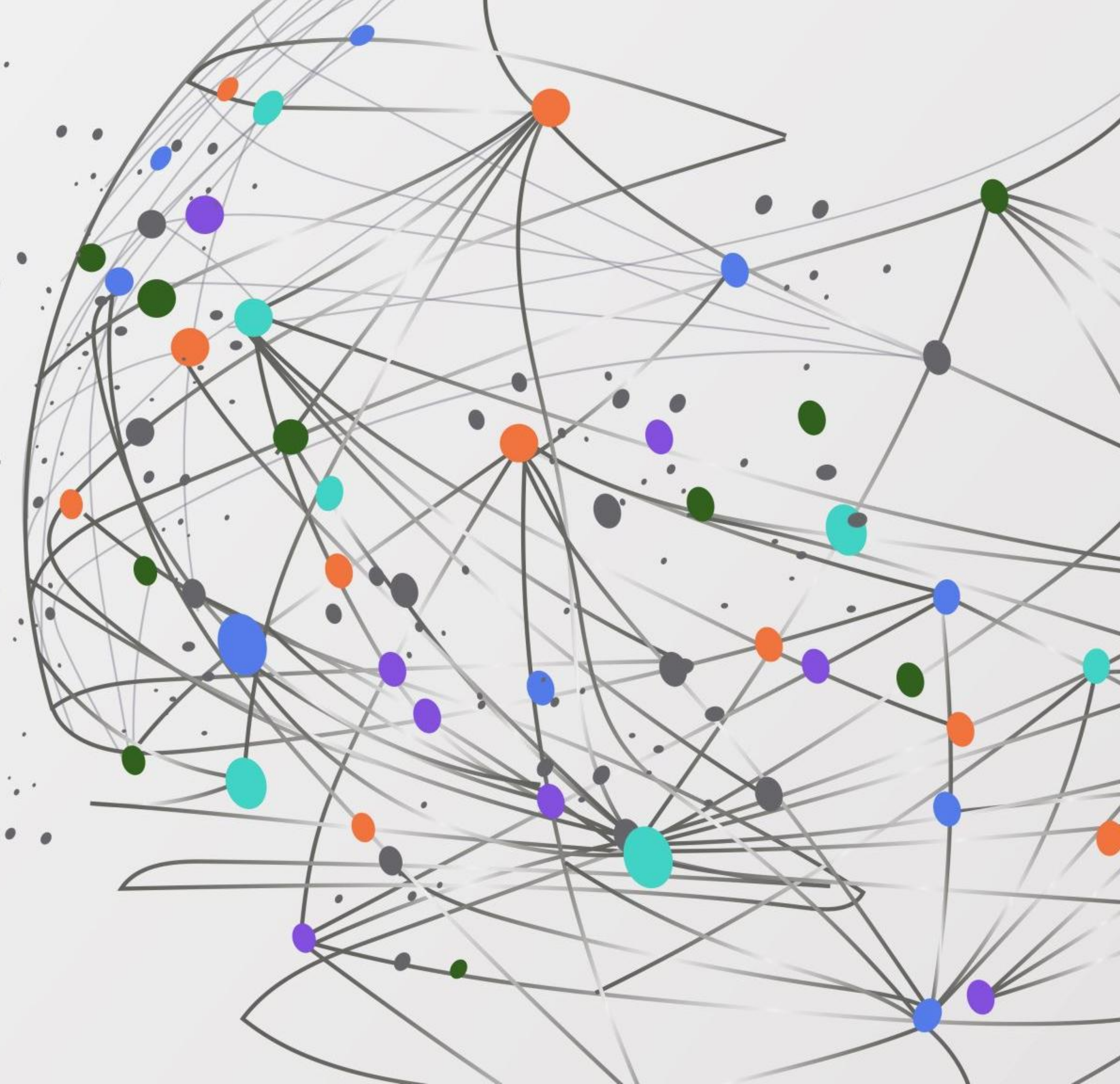
Sentiment Analysis on Women's clothing

BY GROUP 5

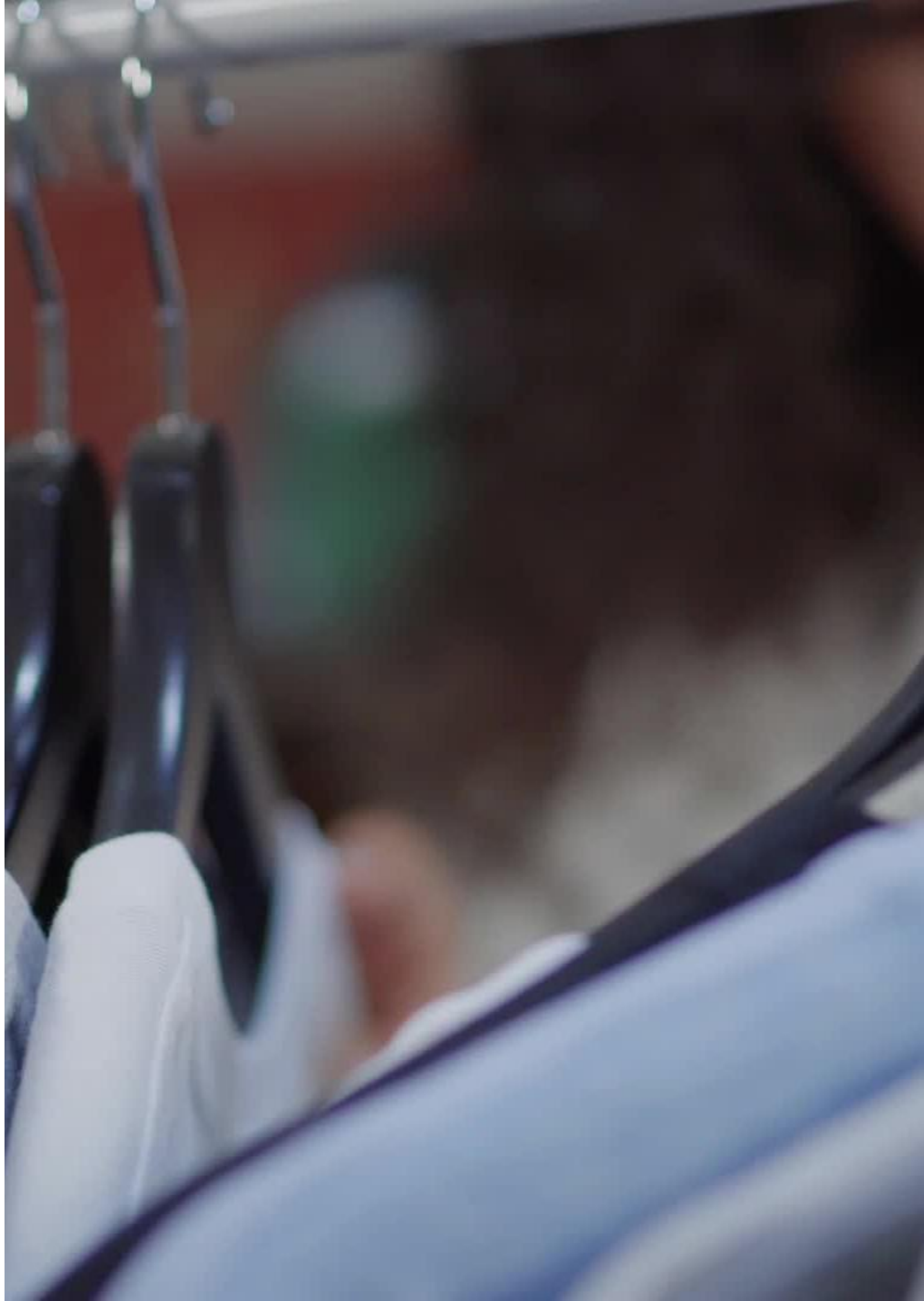


Introduction

Fashion is more than just clothing—it's an expression of emotions, identity, and trends. With millions of women sharing their opinions online through reviews, social media, and blogs, understanding their sentiments can shape the future of fashion.

In this analysis, we dive into the emotions behind women's clothing choices, exploring what they love, dislike, and expect from brands. By leveraging sentiment analysis, we uncover key insights that can help businesses enhance customer satisfaction, improve designs, and drive fashion trends.

What are women saying about fashion? .Let's find out .



Objective

1.Classifying Sentiments

- Developing an NLP model to categorize customer reviews into Positive, Neutral, or Negative sentiments.
- Using Review Text as the main feature and leveraging Ratings or Recommendations as sentiment indicators.
- Targeting at least 80% accuracy and an F1-score of 0.80 on the test set.

2. Identifying Key Themes

- Applying Latent Dirichlet Allocation (LDA) to uncover key themes like product quality, pricing, customer service, and delivery experience.
- Extracting meaningful insights to understand what matters most to customers.

3. Generating Actionable Insights

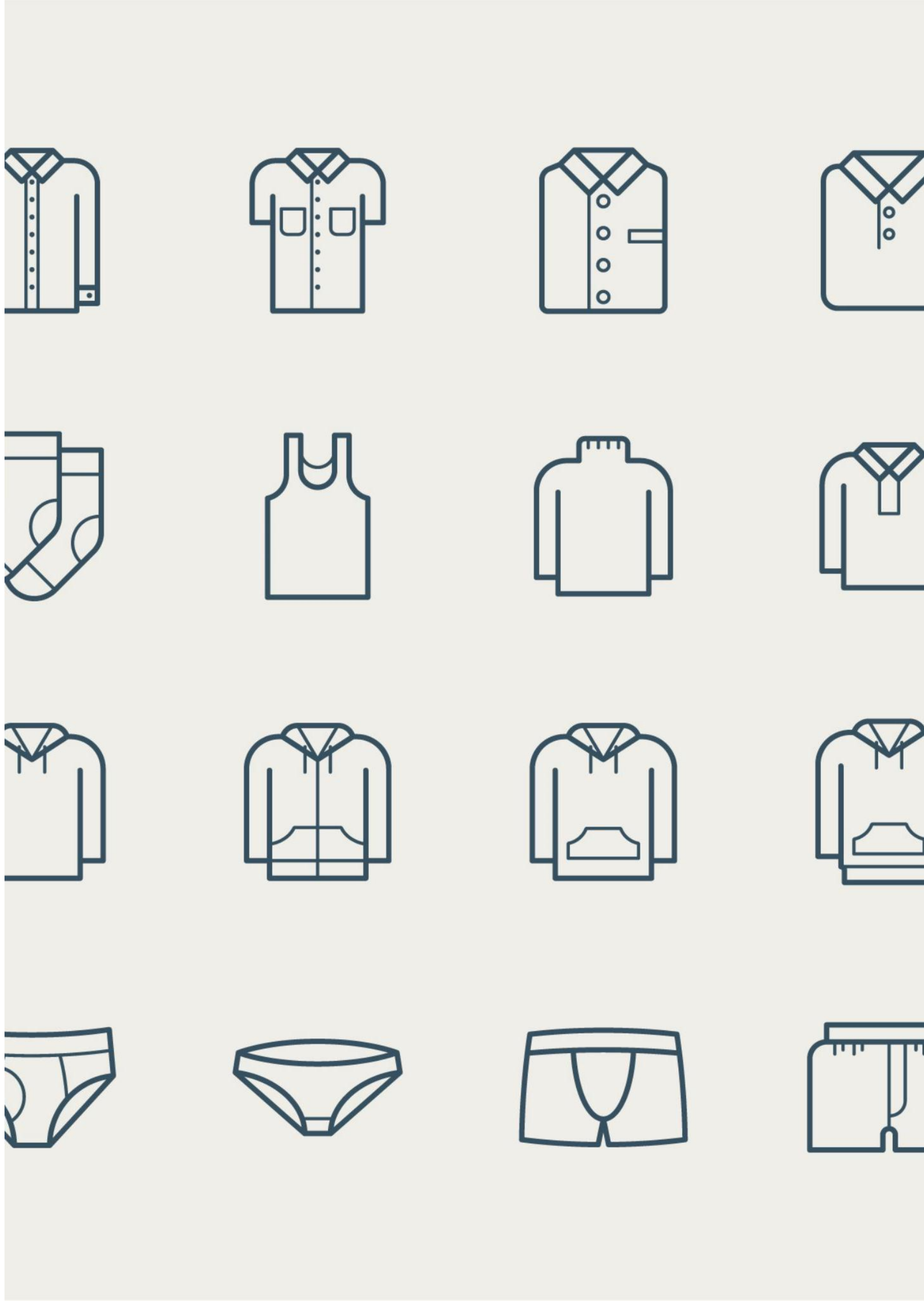
- Creating visualizations (word clouds, topic distributions) to make sentiment trends easy to interpret.

- Providing data-driven recommendations to enhance products, improve customer engagement, and strengthen brand reputation.

Data Overview

Customer & Product Information:

- Clothing ID – A categorical identifier for the reviewed product.
- Age – The age of the reviewer.
- Division Name, Department Name, Class Name – Product classification, helping categorize different types of clothing.



The dataset we worked with contained 23,486 customer reviews and 10 features, offering valuable insights into women's clothing sentiment. The Key features included:

📌 Text-Based Features:

- ✓ Review Text – a summary of our NLP analysis, containing customer opinions.
- ✓ Title – a summary of the review .

📌 Sentiment & Engagement Indicators

- ✓ Rating – A numerical score (1 to 5) indicating customer satisfaction.
- ✓ Recommended IND – A binary indicator (1 = Recommended, 0 = Not Recommended).
- ✓ Positive Feedback Count – The number of customers who found the review helpful.

Exploratory Data Analysis

1. Ratings

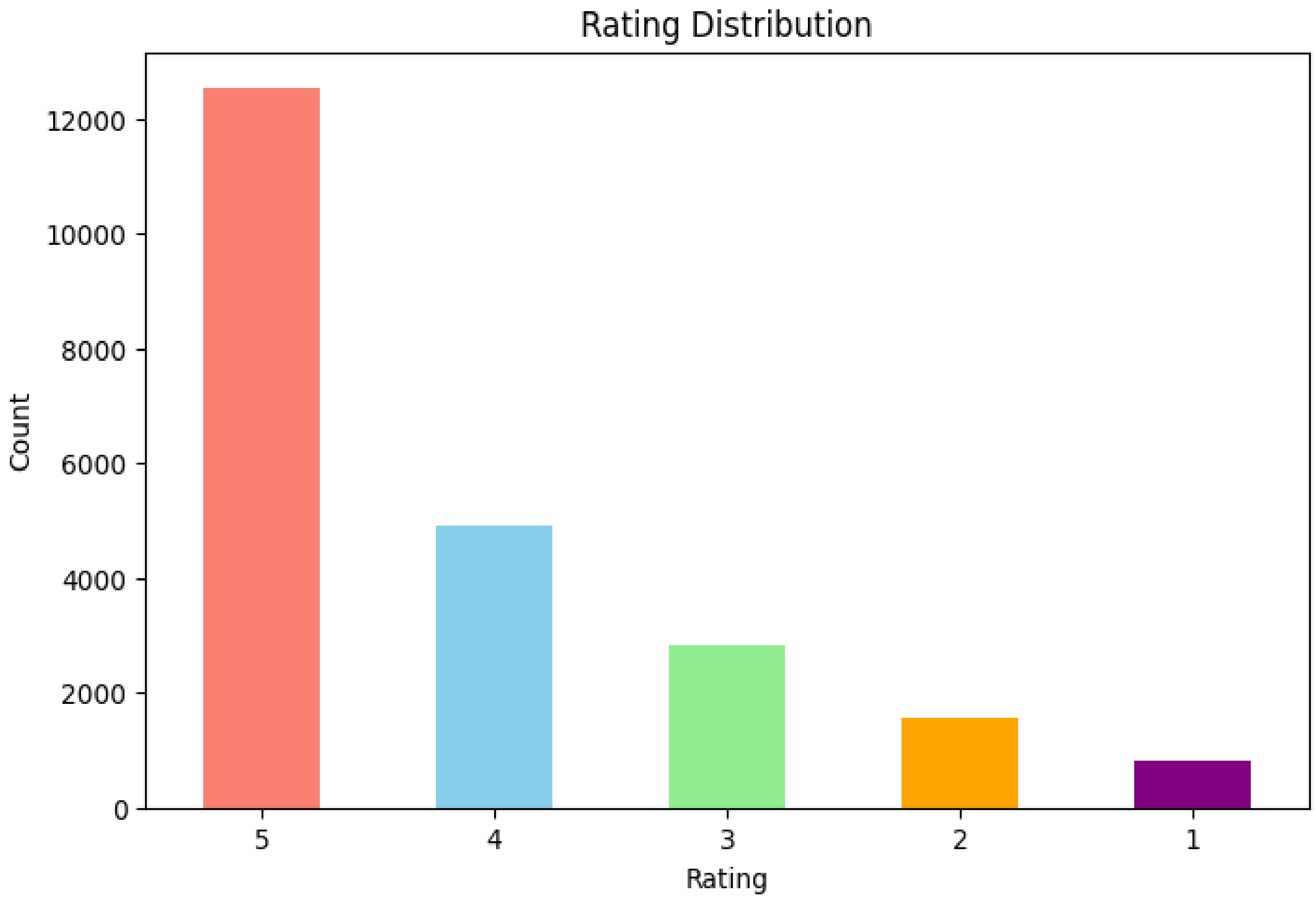
The rating distribution for women's clothing in the dataset shows a strong positive bias toward higher ratings.

- Most of the reviews have a rating of 5, making it the most common score.

- The second-highest frequency is for ratings of 4, indicating that most customers were satisfied with their purchases.

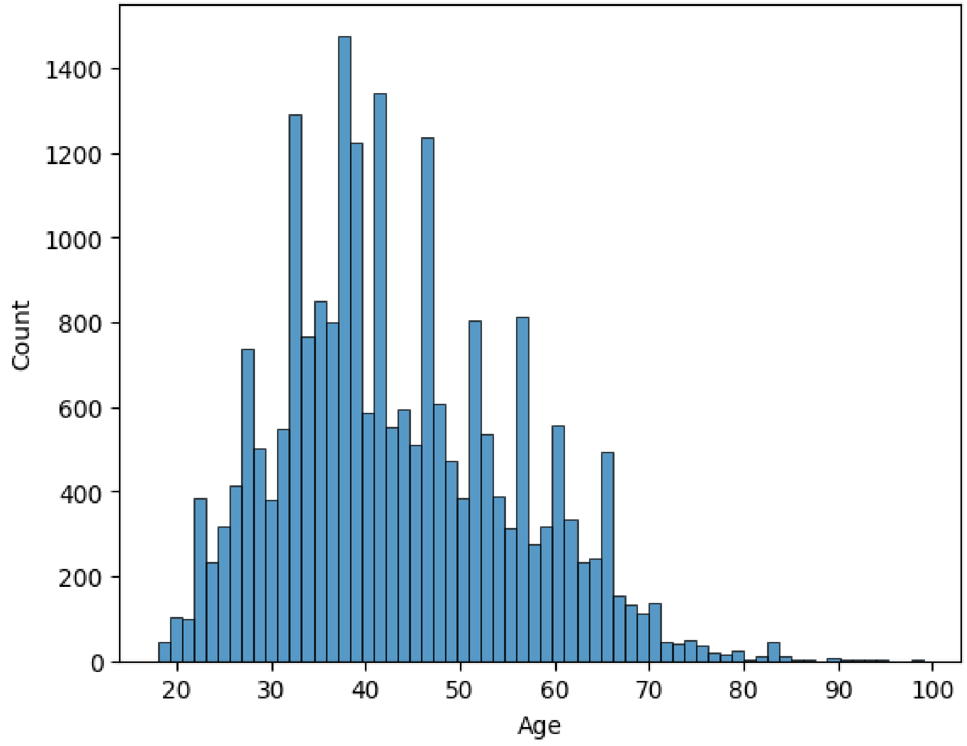
- Ratings of 3, 2, and 1 are significantly lower in comparison, suggesting fewer neutral or negative reviews.

This trend indicates that customers generally have a positive sentiment toward women's clothing, with only a small portion expressing dissatisfaction.


Rating Distribution
12000
10000
8000
6000
4000
2000
0
Count
5
4
3
2
1
Rating

2. Age.

- The age range in the dataset spans from approximately 18 to 100 years old.
- Most customers fall between the ages of 25 and 60.
- The highest concentration of customers is in their late 30s to early 40s.
- The number of customers gradually declines after age 60.
- There are fewer younger customers in their early 20s compared to those in their 30s and 40s.
- Middle-aged women (especially in their 30s and 40s) seem to be the most active shoppers or reviewers in this dataset.



3. Women’s clothing purchasing trends.

Most Purchased:

- Tops (~10,000 units): A wardrobe essential, frequently updated for versatility.
- Dresses (~6,100 units): Popular for work, casual, and special occasions.
- Bottoms (~3,600 units): Necessary but purchased less often due to durability.

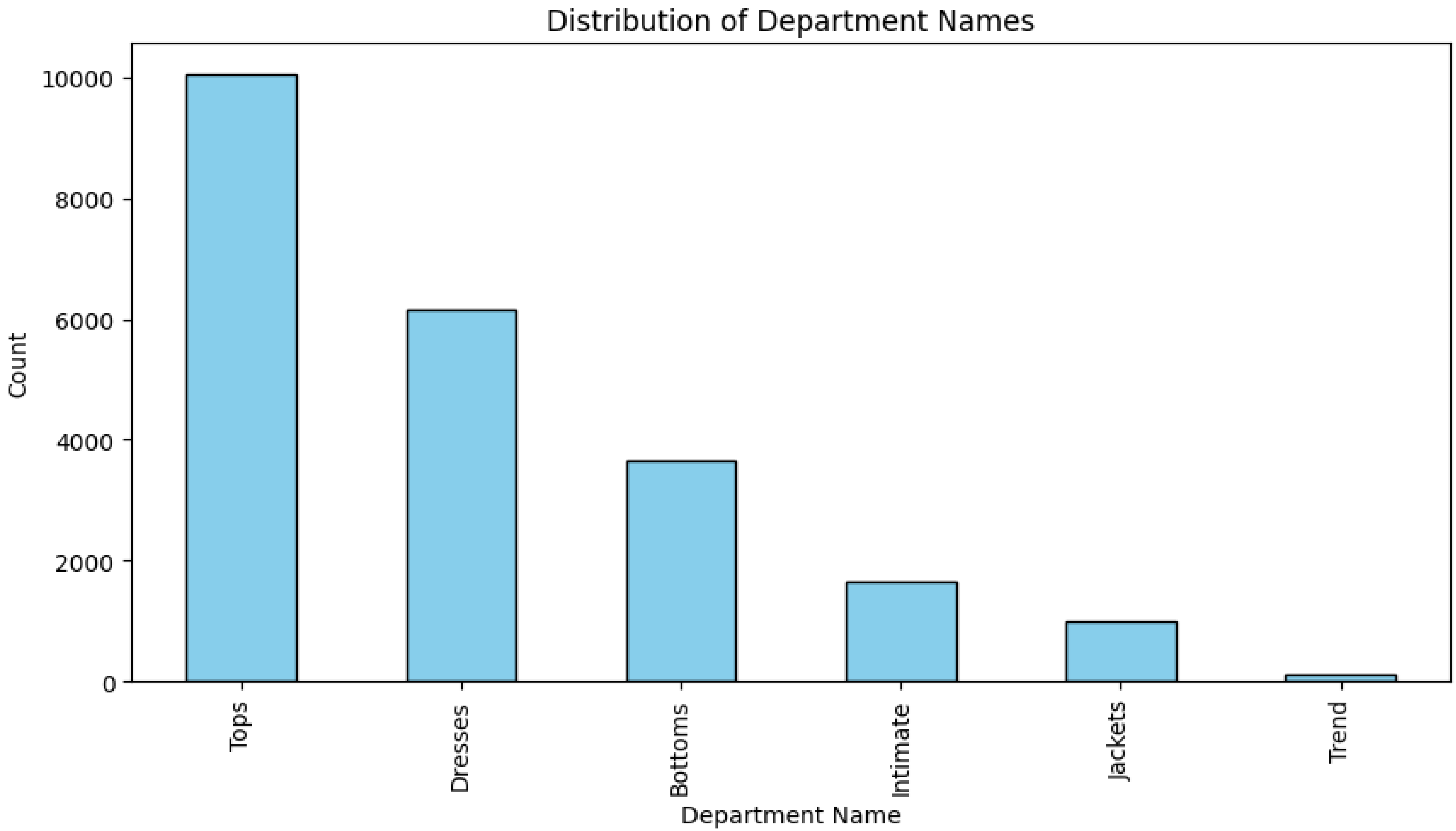
Moderate Purchases:

- Intimates (~1,700 units): Essential but replaced less frequently.
- Jackets (~1,000 units): Seasonal investment pieces with lower turnover.

Least Purchased:

- Trend Items (<200 units): Niche appeal, often integrated into staple categories.

Women prioritize everyday essentials while occasionally updating their wardrobe with seasonal and trendy pieces.


Distribution of Department Names
Count
10000
8000
6000
4000
2000
0
Tops
Dresses
Bottoms
Intimate
Jackets
Trend
Department Name

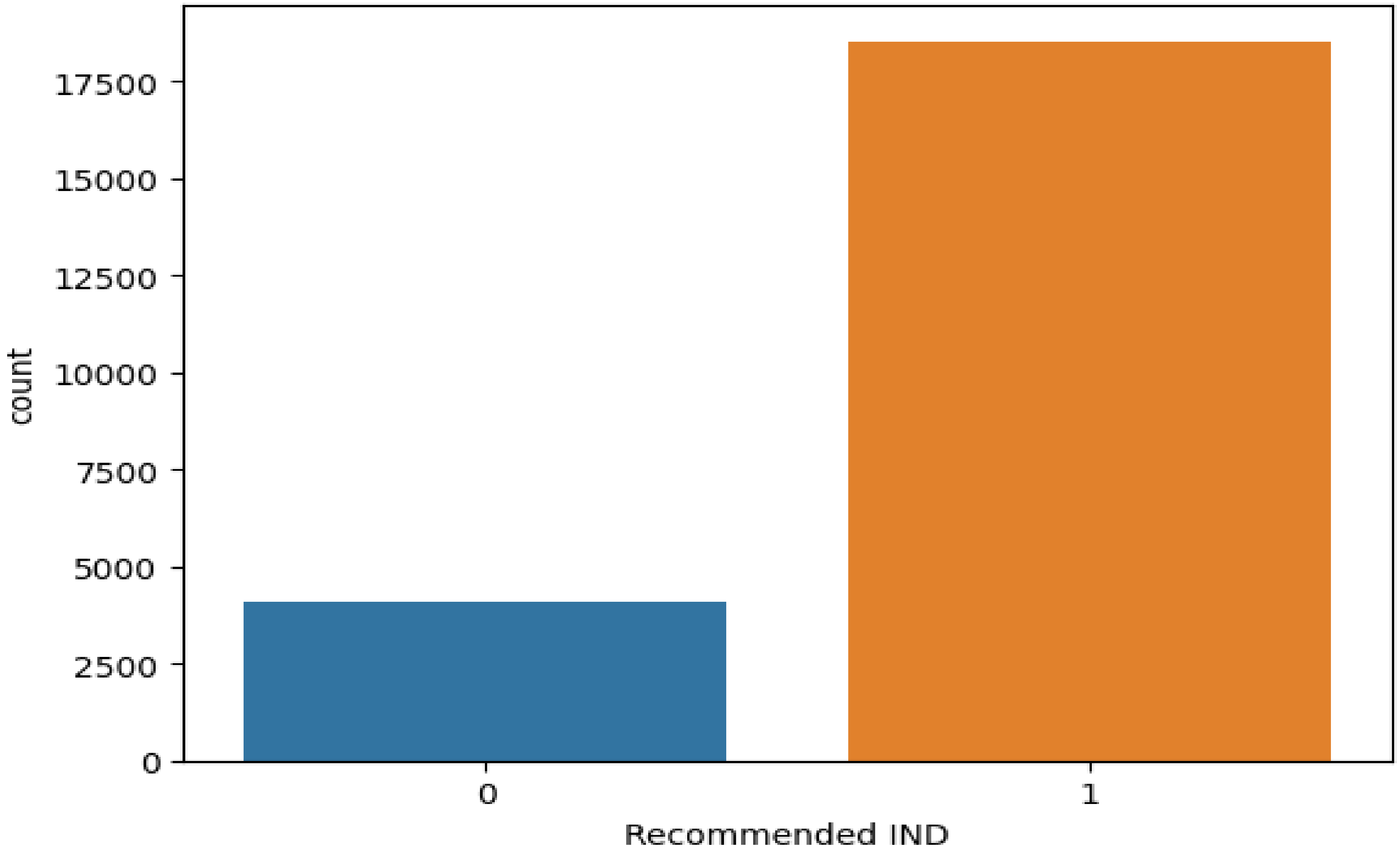
4. Recommendation.

Overall Recommendation Trend: Most women's clothing items were recommended, suggesting a strong customer preference for these items or positive product attributes.

Customer Satisfaction: The high number of recommended items indicates that most clothing met or exceeded customer expectations, reflecting general satisfaction with the products.

Minority Not Recommended: While most items were recommended, a smaller portion did not receive positive recommendations, which could point to either specific product issues or a mismatch with customer preferences.

Product Quality or Appeal: The strong recommendation rate may also indicate that the products generally align with current fashion trends, fit well, or offer good value, contributing to their positive reception.


17500
15000
12500
10000
7500
5000
2500
0
count
0
1
Recommended IND

The word cloud of cleaned reviews **was analyzing** the most frequently used words in customer feedback for women's clothing.

- The most prominent words **were "dress," "top," "shirt," "skirt," and "sweater,"** indicating that these were common clothing items discussed in reviews.

- Sentiment-related words like **"love," "comfortable," "perfect," "flattering," and "great" were appearing** frequently, suggesting a generally positive sentiment toward the products.

- Words related to **fit and material**, such as **"fabric," "material," "size," and "true"**, **were also standing out**, highlighting key factors that influence customer satisfaction.

Overall, the word cloud **was reflecting** a strong emphasis on **product type, comfort, fit, and quality**, providing valuable insights into what customers **were prioritizing** in their reviews.

Word Cloud of Cleaned Reviews

Data Preprocessing

1. Text Cleaning:

- Remove URLs, user mentions, hashtags, punctuation, special characters, and numbers to focus on meaningful text.

2. Lowercasing:

- Convert all text to lowercase to ensure consistency (e.g., "Dress" and "dress" are treated the same).

3. Tokenization:

- Split text into individual words (tokens), and handle multi-word expressions (e.g., "feel good").

4. Removing Stopwords:

- Exclude common words (like “and,” “the,” “is”) that don’t contribute much meaning.

5. Lemmatization/Stemming:

- Reduce words to their base or root form (e.g., "running" to "run").

6. Bigrams:

- Generate bigrams (pairs of consecutive words) to capture context that single words alone might miss (e.g., “high quality” becomes a bigram).

7. TF-IDF (Term Frequency-Inverse Document Frequency):

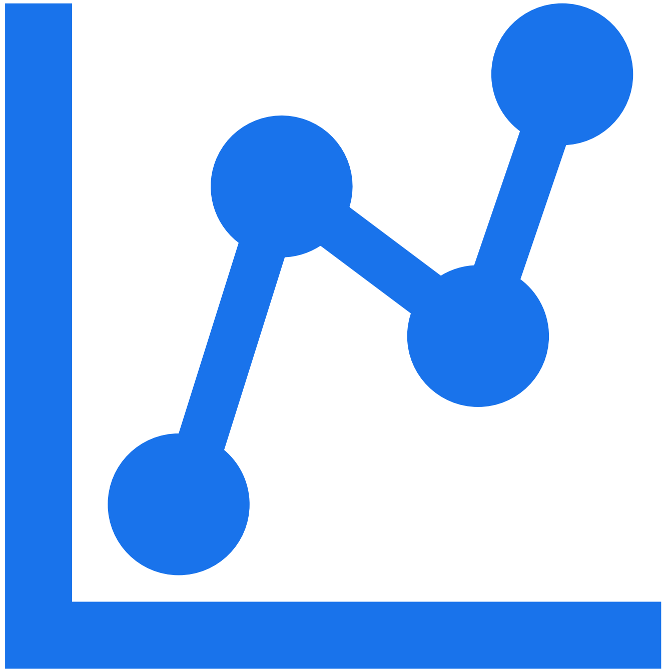
- Use TF-IDF to identify important words by considering their frequency and uniqueness across the dataset, helping highlight key terms

Model Used

1. Logistic Regression with SMOTE

We used Logistic Regression as our baseline model and applied SMOTE) to handle class imbalance. After applying SMOTE, we achieved the following results:

- Accuracy: 0.83, Recall: 0.82, Precision: 0.74, F1 Score: 0.76

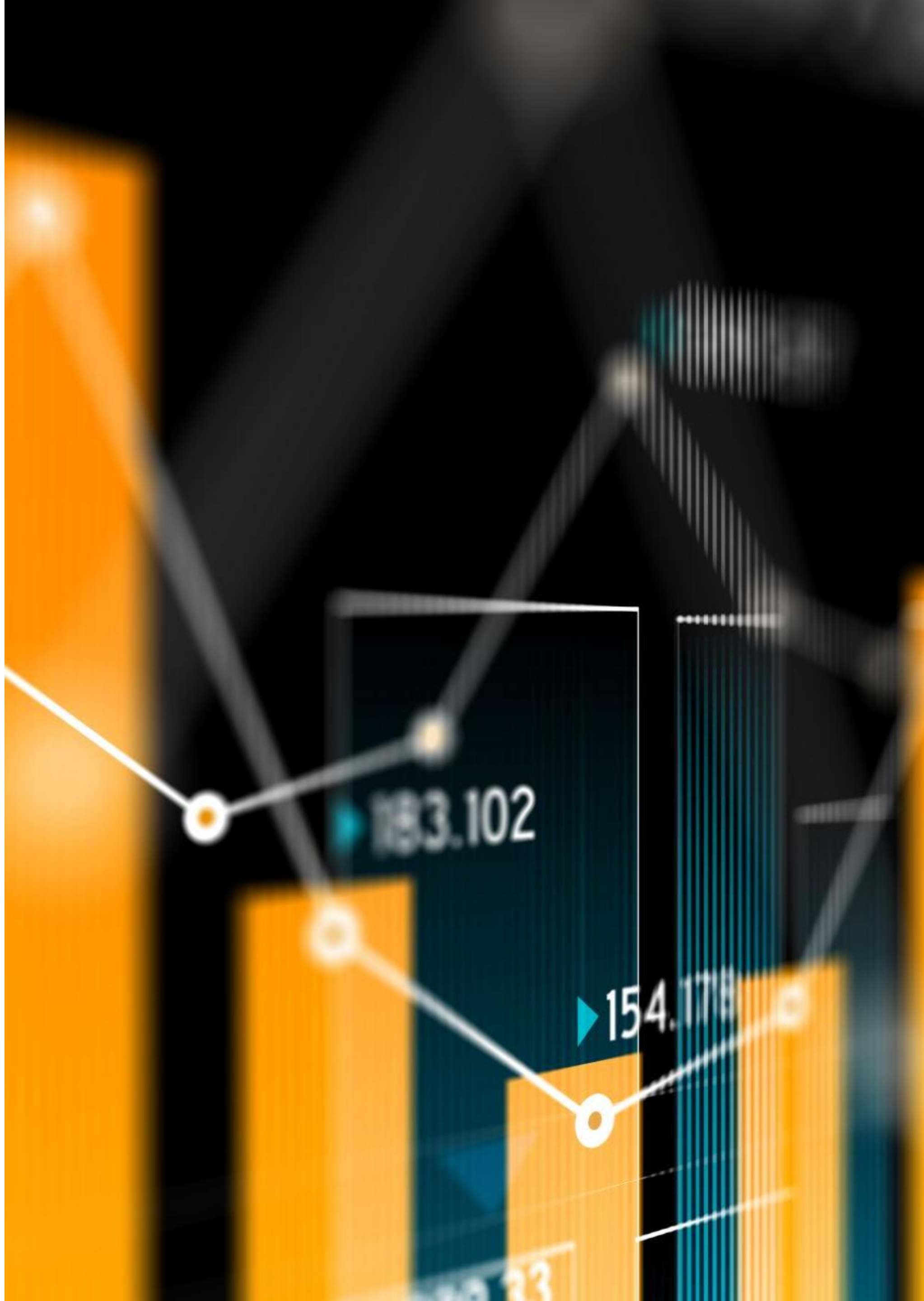


2. Hyperparameter Tuning of Logistic Regression

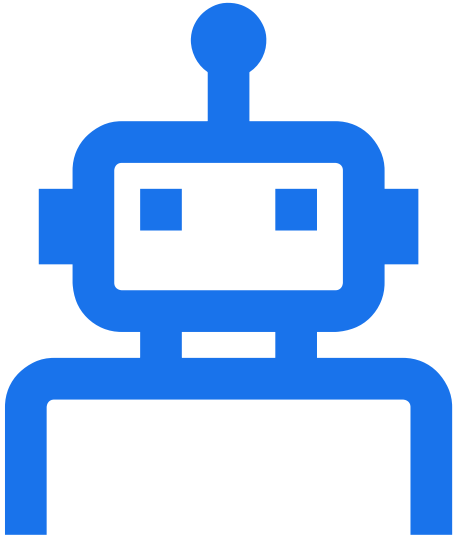
After applying SMOTE, we tuned the Logistic Regression model by optimizing its hyperparameters. The best parameters identified were C: [Optimized Value], Cross-Validation: 10, Penalty: 12, Solver: LBFGS

With these tuned parameters, the model achieved the following improved results:

- Accuracy: 0.84, Recall: 0.81, Precision: 0.74, F1 Score: 0.87



Advanced Models



1. Multinomial Naive Bayes

We used **Multinomial Naive Bayes** to handle text classification. The model was applied to the dataset, and we obtained the following results:

- **Accuracy:** 0.76, **Recall:** 0.77, **Precision:** 0.68, **F1 Score:** 0.69

This model performed reasonably, but there was room for improvement in precision and F1 score.

2. Support Vector Machine (SVM)

For the **Support Vector Machine (SVM)** model, we applied it to the dataset with a focus on maximizing class separation. The results were:

- **Accuracy:** 0.83, **Recall:** 0.82, **Precision:** 0.74, **F1 Score:** 0.76

The SVM showed stronger performance compared to Naive Bayes, with improved metrics across the board.

3. Comparison and Improvement

- **Accuracy**: Improved from **0.76** (Naive Bayes) to **0.83** (SVM)
- **Recall**: Improved from **0.77** (Naive Bayes) to **0.82** (SVM)
- **Precision**: Improved from **0.68** (Naive Bayes) to **0.74** (SVM)
- **F1 Score**: Improved from **0.69** (Naive Bayes) to **0.76** (SVM)

Business Recommendations

1. Focus on Fit and Sizing

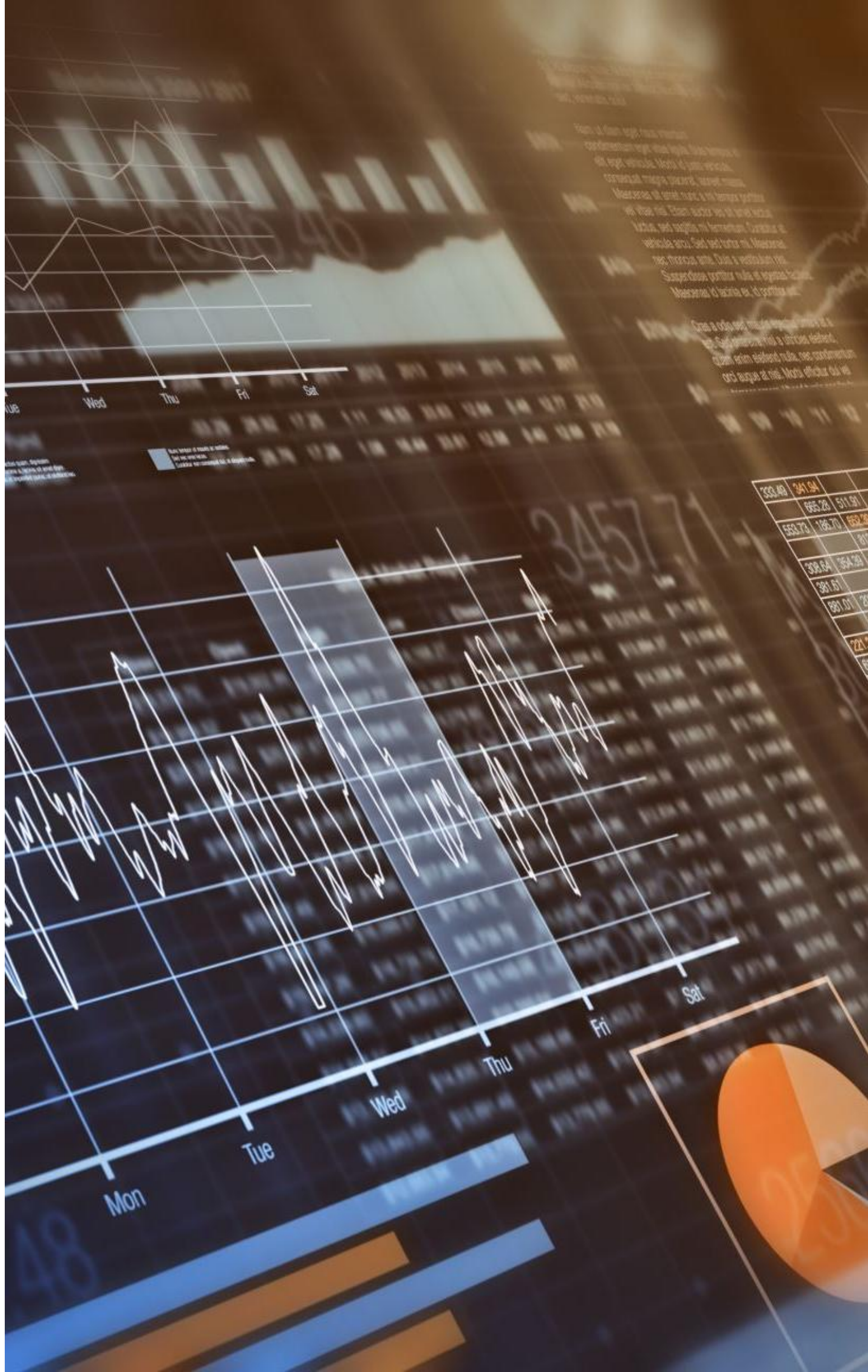
Improve size guides, provide detailed measurement charts, and consider AI-powered virtual fitting tools to reduce size-related dissatisfaction.

2.Enhance Fabric Quality and Comfort

Invest in high-quality, breathable, and durable materials while providing fabric descriptions in product listings.

3.Maintain Consistent Product Quality

Strengthen quality control measures, ensuring consistency in material, stitching, and durability across all product batches.



4.Personalize Marketing Strategies

Use data-driven marketing to target specific customer segments, offering personalized recommendations based on past preferences.

5. Offer More Diverse Styles and Inclusivity

Expand the range of styles to be more inclusive, ensuring that products are available for all body types and lifestyles. Ensuring diversity in the design process can help build a more inclusive brand image.

6. Improve Delivery and Returns Experience

Offer flexible shipping options and simplify the return process to enhance customer satisfaction. Providing easy-to-print return labels or local drop-off points can improve the overall convenience for customers.

Meet The Team

1 Ian Bett.

2.Lionel Ajeliti

3.Morgan Abukuse Amunga

4.Sanayet Sankaine

5.Linet Patriciah

Thank You!

Q&A