# Project Work

Contents

[Project Work 1](#_Toc184901035)

[Data Preprocessing 1](#_Toc184901036)

[Data Visualization 2](#_Toc184901037)

[Distribution of Sentiment Scores: 2](#_Toc184901038)

[KDE plot for Review length: 3](#_Toc184901039)

[Word Cloud for reviews: 4](#_Toc184901040)

[Top TF-IDF Terms in Negative Reviews: 4](#_Toc184901041)

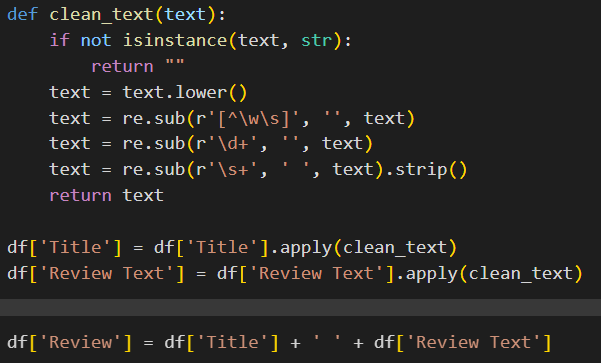
[Data Analysis and Key Findings 5](#_Toc184901042)

[Key Findings 5](#_Toc184901043)

[Recommendations 6](#_Toc184901044)

# Data Preprocessing

1. Dropped the rows where both Title and Review were NaN.
2. Cleaned the text by replacing missing values with empty strings and removing punctuations, numbers and extra spaces.
3. Joined Review Text and Title column to create Review Column.



1. Performed Stemming and Lemmatization

A screen shot of a computer code

Description automatically generated

# Data Visualization

## Distribution of Sentiment Scores:

A diagram of a distribution of a certain score

Description automatically generated with medium confidence

A bar graph with numbers and a number of negatives

Description automatically generated

## KDE plot for Review length:

A graph of a review length

Description automatically generated

## Word Cloud for reviews:

A word cloud of clothing

Description automatically generated

## Top TF-IDF Terms in Negative Reviews:

A graph of a bar graph

Description automatically generated with medium confidence

# Data Analysis and Key Findings

The top TF-IDF terms in negative reviews highlight recurring themes and specific grievances from customers. These terms can be grouped into **categories** for easier analysis:

## Key Findings

1. **Product-Specific Complaints**:
   * Terms like dress, shirt, and sweater indicate that customers are primarily discussing specific types of clothing.
   * The prominence of fit, size, small, big, and tight suggests that sizing issues are a common problem. Customers may find products too small, tight, or improperly fitted.
2. **Material and Quality Issues**:
   * Words like fabric, material, and quality imply dissatisfaction with the physical characteristics of the products.
   * Customers might be reporting poor-quality materials or fabrics that don’t meet their expectations.
3. **Appearance and Style**:
   * Words like look, pretty, and cute suggest that customers have concerns about how the clothing appears in person versus how it is advertised.
   * disappointed being included highlights unmet expectations about the product’s aesthetic.
4. **Performance Issues**:
   * Terms such as run, short, and long may indicate functionality issues (e.g., pants or sleeves being too short or long, or dresses running in fabric).
   * tried and usually could reflect customers describing their repeated efforts to resolve issues or compare the product to prior experiences.
5. **Customer Sentiment and Expectations**:
   * Words like love and like might appear in phrases expressing disappointment in items they expected to love or like.
   * ordered and im might be from descriptions of ordering experiences or expectations set during the purchase process.
6. **Versatility or Use**:
   * The term wear suggests commentary on how the product feels when worn or its suitability for specific occasions.

## Recommendations

1. **Sizing Guidance**:
   * Provide more accurate size charts, detailed customer reviews about fit, and models of varying body types to set realistic expectations.
2. **Material and Quality Assurance**:
   * Emphasize high-quality materials in product descriptions.
   * Provide close-up product images or videos to show texture and quality.
3. **Visual Representation**:
   * Ensure that images used in advertisements and online match the actual product appearance.
   * Allow for user-generated content (e.g., photos from real customers).
4. **Focus on Top Complaints**:
   * Address common issues like "runs short" or "tight waist" in product descriptions. For example, specify that an item may run small and suggest sizing up.
5. **Proactive Communication**:
   * Use insights from terms like disappointed and ordered to improve communication at the point of purchase, such as by managing expectations or offering clearer return policies.