EDA Report - Women's Clothing E-Commerce Reviews

1. Introduction

This report provides an Exploratory Data Analysis (EDA) of the Women's Clothing E-Commerce Reviews dataset.

The primary goal is to uncover patterns in customer reviews to understand how factors such as age, rating, and department

contribute to the recommendation of products.

2. Dataset Overview

- Source: Kaggle

- Rows Loaded: First 1000 rows

- Key columns include: Age, Rating, Recommended IND,

Positive Feedback Count, Division Name, Department Name, Class Name, Title, Review Text.

3. Tools & Libraries Used

- Python (Pandas, NumPy, Matplotlib, scikit-learn)

- Jupyter Notebook for interactive development

- GitHub for version control

4. Key EDA Techniques Applied

- Distribution Plots: To understand frequency distributions of both categorical and numerical columns.

- Correlation Matrix: To identify linear relationships between numerical features.

EDA Report - Women's Clothing E-Commerce Reviews

- Scatter Matrix: For pairwise feature analysis and visual correlation assessment.

5. Observations

- Some numerical columns (like Positive Feedback Count) have skewed distributions.

- Rating and Recommended IND show moderate correlation.

- Customers generally tend to leave more positive feedback on recommended items.

- The dataset contains both structured (age, rating) and unstructured (review text) data, ideal for both traditional ML and NLP applications.

6. Next Steps

- Data cleaning (handling missing values, encoding text).

- Model development (e.g., classification of Recommended IND).

- Deployment (optional): Building a simple app to predict recommendation likelihood from input features.