```
import pandas as pd
import re
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
import matplotlib.pyplot as plt
import seaborn as sns

# Read the dataset (assuming it's in the same directory
)
df = pd.read_csv('Womens Clothing E-Commerce Reviews.
csv', header=0, index_col=0)

# Check for null entries
df.isna().sum()

# Visualize rating distribution
sns.countplot(x='Rating', data=df)
plt.title("Distribution of Rating")
plt.show()

# Download NLTK stopwords
import nltk
nltk.download('stopwords')

# Initialize stopwords
stops = stopwords.words("english")

# Function to clean and tokenize text
def tokens(words):
    words = re.sub("[^a-zA-Z]", " ", words)
    text = words.lower().split()
    return " ".join(text)

# Apply text cleaning and tokenization to the 'Review'
column
df['Review_clear'] = df['Review'].apply(tokens)

# Initialize Porter Stemmer
ps = PorterStemmer()

# Stemming and removing stopwords from reviews
```

```
corpus = []
for review in df['Review_clear']:
    review = review.split()
    review = [ps.stem(word) for word in review if not
word in set(stops)]
    review = " ".join(review)
    corpus.append(review)
```