Exp 6) Write a program for Amazon review dataset to find the maximum number of wonh unod. Get the cutput for the frequently occurred word in the given data? And also visualize the test data.

import pandas as pd

import re

from collections import Counter

import matplotlib.pyplot as plt

from wordcloud import WordCloud

# Step 1: Load the dataset

# Replace 'amazon\_alexa.tsv' with the path to your dataset

df = pd.read\_csv('amazon\_alexa.tsv', delimiter='\t')

# Step 2: Preprocess the text data

def preprocess\_text(text):

text = re.sub(r'\W', ' ', text) # Remove special characters

text = re.sub(r'\s+', ' ', text) # Replace multiple spaces with a single space

text = text.lower().strip() # Convert to lowercase and strip leading/trailing spaces

return text

# Assuming the reviews are in a column named 'verified\_reviews'

df['cleaned\_reviews'] = df['verified\_reviews'].apply(preprocess\_text)

# Step 3: Find the most frequently occurring word

all\_words = ' '.join(df['cleaned\_reviews']).split()

word\_counts = Counter(all\_words)

most\_common\_word, most\_common\_count = word\_counts.most\_common(1)[0]

# Display the most frequent word and its count

print(f"The most frequent word is '{most\_common\_word}' with {most\_common\_count} occurrences.")

# Step 4: Visualize the top 10 most frequent words

top\_10\_words = word\_counts.most\_common(10)

words, counts = zip(\*top\_10\_words)

plt.figure(figsize=(10, 5))

plt.bar(words, counts, color='skyblue')

plt.title('Top 10 Most Frequent Words in Amazon Reviews')

plt.xlabel('Words')

plt.ylabel('Frequency')

plt.show()

# Step 5: Visualize the word cloud

wordcloud = WordCloud(width=800, height=400, background\_color='white').generate(' '.join(all\_words))

plt.figure(figsize=(10, 5))

plt.imshow(wordcloud, interpolation='bilinear')

plt.axis('off')

plt.show()

