import pandas as pd

import string

from collections import Counter

import matplotlib.pyplot as plt

import nltk

# Download stopwords (only first time)

nltk.download("stopwords")

from nltk.corpus import stopwords

# Step 1: Load dataset

data = pd.read\_csv("c:/Users/dadak/Downloads/data.csv")

# Ensure the column exists

if "feedback" not in data.columns:

    raise ValueError("CSV file must contain a 'feedback' column.")

# Step 2: Preprocess the text

stop\_words = set(stopwords.words("english"))

all\_words = []

for text in data["feedback"].dropna():  # skip missing values

    text = text.lower()  # lowercase

    text = text.translate(str.maketrans("", "", string.punctuation))  # remove punctuation

    words = text.split()

    # remove stopwords

    filtered\_words = [word for word in words if word not in stop\_words]

    all\_words.extend(filtered\_words)

# Step 3: Calculate frequency distribution

word\_freq = Counter(all\_words)

# Step 4: Ask user for N

N = int(input("Enter the number of top frequent words to display: "))

# Display results

print(f"\nTop {N} most frequent words:\n")

for word, freq in word\_freq.most\_common(N):

    print(f"{word}: {freq}")

# Step 5: Plot bar graph

most\_common\_words = word\_freq.most\_common(N)

if most\_common\_words:

    words, counts = zip(\*most\_common\_words)

    plt.bar(words, counts)

    plt.xlabel("Words")

    plt.ylabel("Frequency")

    plt.title(f"Top {N} Most Frequent Words in Feedback")

    plt.xticks(rotation=45)

    plt.show()

else:

    print("No words available for plotting.")

OUTPUT:

[nltk\_data] Downloading package stopwords to

[nltk\_data] C:\Users\dadak\AppData\Roaming\nltk\_data...

[nltk\_data] Package stopwords is already up-to-date!

Enter the number of top frequent words to display: 10

Top 10 most frequent words:

product: 4

amazing: 2

service: 2

great: 2

quality: 2

delivery: 2

love: 1

works: 1

perfectly: 1

terrible: 1

