17. Scenario: You are a data analyst working for a marketing research company. Your team has

collected a large dataset containing customer feedback from various social media platforms. The

dataset consists of thousands of text entries, and your task is to develop a Python program to

analyze the frequency distribution of words in this dataset. Your program should be able to perform

the following tasks:

 Load the dataset from a CSV file (data.csv) containing a single column named "feedback"

with each row representing a customer comment.

 Preprocess the text data by removing punctuation, converting all text to lowercase, and

eliminating any stop words (common words like "the," "and," "is," etc. that don't carry

significant meaning).

 Calculate the frequency distribution of words in the preprocessed dataset.

 Display the top N most frequent words and their corresponding frequencies, where N is

provided as user input.

 Plot a bar graph to visualize the top N most frequent words and their frequencies.

Question: Create a Python program that fulfills these requirements and helps your team gain

insights from the customer feedback data.

**CODE:**

import pandas as pd

import string

import matplotlib.pyplot as plt

from collections import Counter

# Load stop words manually

stop\_words = set([

'the', 'and', 'is', 'in', 'to', 'it', 'was', 'with', 'for', 'of', 'on',

'a', 'very', 'so', 'i', 'this', 'that', 'its', 'my', 'be', 'at', 'not',

'too', 'but', 'as', 'are', 'an', 'we', 'they', 'you', 'will'

])

df = pd.read\_csv(r"C:\Users\Shashankbheemavarapu\Downloads\data.csv")

def preprocess(text):

text = text.lower()

text = text.translate(str.maketrans('', '', string.punctuation))

words = text.split()

words = [word for word in words if word not in stop\_words and word.isalpha()]

return words

all\_words = []

for feedback in df['feedback'].dropna():

all\_words.extend(preprocess(feedback))

word\_counts = Counter(all\_words)

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

top\_words = word\_counts.most\_common(N)

# Display result

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

for word, count in top\_words:

print(f"{word}: {count}")

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

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

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

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

plt.xlabel("Words")

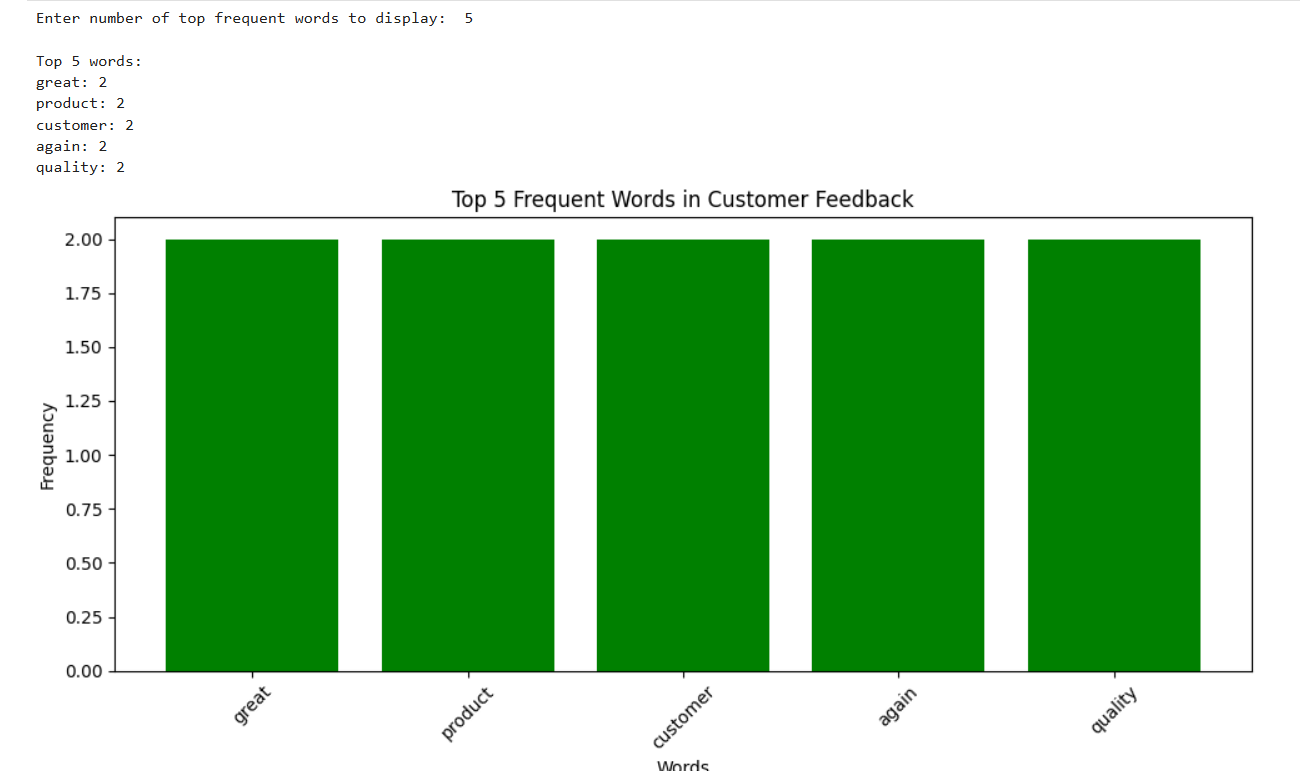
plt.ylabel("Frequency")

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

**OUTPUT:**

****

**DATASET:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| feedback |  |  |  |  |  |
| This is a great product! Highly recommend to everyone. | | | | | |
| Terrible customer service. Will not buy again. | | | | |  |
| Decent quality for the price. Could be better. | | | | |  |
| Absolutely love it! Exceeded my expectations. | | | | |  |
| The delivery was late and the packaging was damaged. | | | | | |
| Amazing experience, fantastic quality, and great value. | | | | | |
| Not worth the money. Very disappointed with the purchase. | | | | | |
| Customer support was very helpful and resolved my issue quickly. | | | | | |
| Good product but lacks durability. Needs improvement. | | | | | |
| Fantastic performance! Will definitely purchase again. | | | | |  |