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 :

#17

import pandas as pd

import matplotlib.pyplot as plt

from collections import Counter

import string

import re

import nltk

from nltk.corpus import stopwords

nltk.download('stopwords')

df = pd.read\_excel(r"C:\Users\hares\Downloads\q17\_05.xlsx")

def preprocess\_text(text):

text = text.lower()

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

text = re.sub(r'\d+', '', text).strip()

words = text.split()

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

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

return words

all\_words = []

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

all\_words.extend(preprocess\_text(comment))

word\_freq = Counter(all\_words)

try:

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

except ValueError:

print("Invalid input. Defaulting to top 10 words.")

N = 10

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

print("\nTop {} Most Frequent Words:".format(N))

for word, freq in top\_words:

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

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

plt.figure(figsize=(10, 6))

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

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

plt.xlabel('Words')

plt.ylabel('Frequency')

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

output :

Enter the number of top frequent words to display: 5

Top 5 Most Frequent Words:

product: 1

love: 1

amazing: 1

easy: 1

use: 1

dataset :

|  |
| --- |
| **feedback** |
| Love the product! It's amazing and easy to use. |
| The delivery was late and the package was damaged. |
| Excellent customer service and support. |
| Not satisfied with the quality. Bad experience. |
| Great value for the price! |