Question 13

13. Scenario: You are working on a text analysis project and need to determine the frequency

distribution of words in a given text document. You have a text document named "sample\_text.txt"

containing a paragraph of text. Your task is to develop a Python program that reads the text

document, processes the text, and generates a frequency distribution of the words.

Question: How would you develop a Python program to calculate the frequency distribution of words in a text document?

Answer

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv(r"C:\Users\jampa\Downloads\word\_frequency\_distribution.csv")

# Sort by frequency for better visualization (descending)

df\_sorted = df.sort\_values(by='Frequency', ascending=False)

# Optional: Limit to top 20 most frequent words for readability

top\_words = df\_sorted.head(20)

# Plotting

plt.figure(figsize=(12, 6))

plt.bar(top\_words['Word'], top\_words['Frequency'], color='skyblue')

plt.xlabel('Words')

plt.ylabel('Frequency')

plt.title('Top 20 Word Frequency Distribution')

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

Output:

