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?

Code:

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

df **=** pd.read\_csv(r"C:\Users\vara prasad\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')

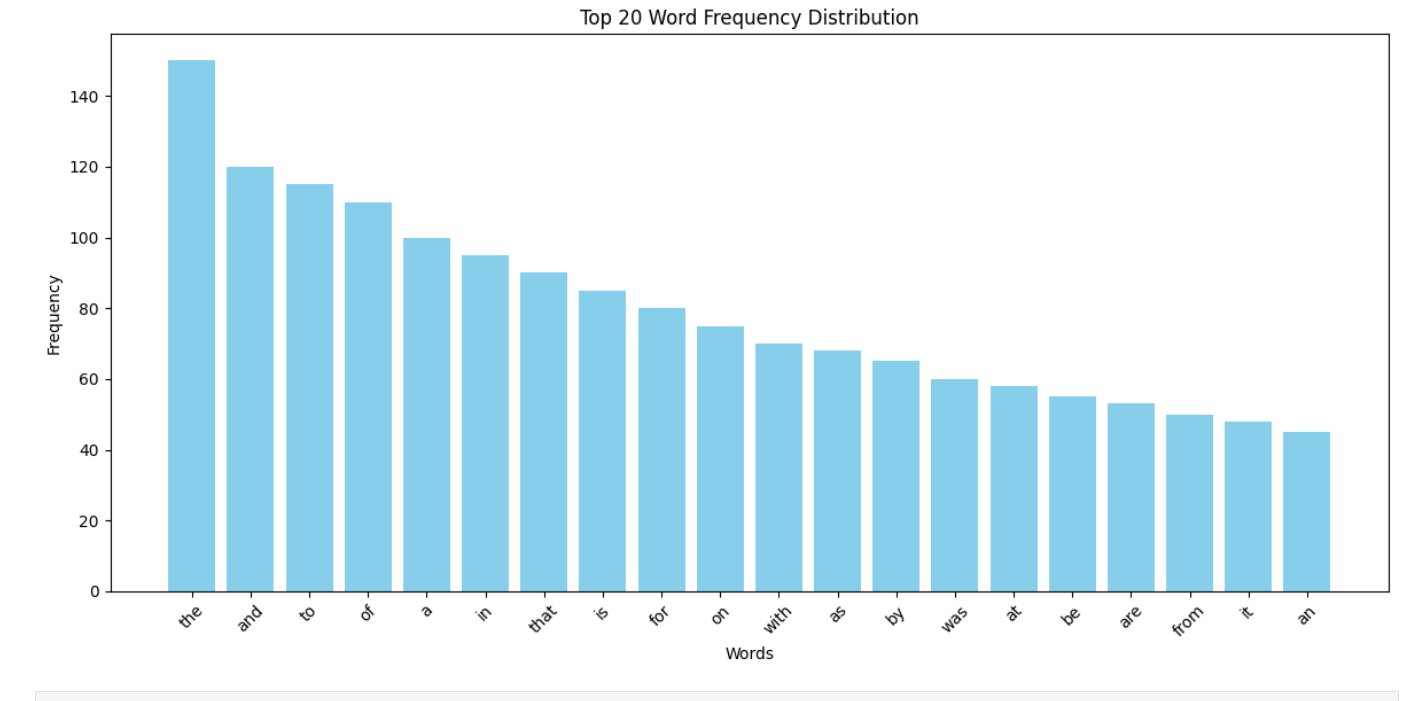
plt.title('Top 20 Word Frequency Distribution')

plt.xticks(rotation**=**45)

plt.tight\_layout()

plt.show()

output:



Dataset:

|  |  |
| --- | --- |
| Word | Frequency |
| text | 1 |
| analysis | 2 |
| involves | 1 |
| extracting | 1 |
| meaningful | 1 |
| information | 1 |
| from | 1 |
| textual | 1 |
| data | 1 |
| it | 1 |
| is | 1 |
| used | 1 |
| in | 1 |
| many | 1 |
| applications | 1 |
| including | 1 |
| sentiment | 1 |
| topic | 1 |
| modeling | 1 |
| and | 1 |
| keyword | 1 |
| extraction | 1 |