README - Sentiment Analysis of Tweets

Objective

The purpose of this script is to analyze the sentiment of tweets and visualize the sentiment distribution through a histogram.

Prerequisites

1. Python installed on your system.
2. Required libraries: pandas, textblob, and matplotlib.
    - Install using pip install pandas textblob matplotlib.
3. A CSV file (tweets_with_sentiment.csv) containing tweet data.
    - The file must have a text column with tweets. A sentiment column is optional.

How to Use

1. Save the script in a Python file (e.g., analyze_sentiment.py).
2. Ensure the CSV file (tweets_with_sentiment.csv) is in the same directory as the script.
3. Run the script:
4. python analyze_sentiment.py

Script Workflow

1. **Loading Data**:
    - The script attempts to load tweets_with_sentiment.csv.
    - If the file does not exist, it exits with an error message.
2. **Sentiment Analysis**:
    - If the sentiment column is missing, the script calculates sentiment polarity for each tweet using TextBlob.
    - Sentiment polarity ranges from **-1 (negative)** to **1 (positive)**.
3. **Summary Statistics**:
    - Generates descriptive statistics of sentiment scores (mean, min, max, standard deviation, etc.).
4. **Visualization**:
    - Creates a histogram to visualize the frequency of sentiment scores across the dataset.

Expected Output

1. **Console**:
    - Success message if the data loads successfully.
    - Summary statistics of sentiment scores (e.g., mean, median, min, max).

2. **Graph**:
    - A histogram showing the distribution of sentiment scores.

Potential Errors

1. **File Not Found**:
    - If the CSV file is missing, the script exits with an error.
2. **Missing text Column**:
    - Ensure the CSV file contains a text column; otherwise, the script will fail.

Customization

- Replace the filename (tweets_with_sentiment.csv) with the path to your CSV file.
- Adjust the number of bins in the histogram (bins=20) for finer or coarser visualization.

Summary Example

For example, after running the script, you might see:

- Sentiment Score Statistics:
- count 100.000000
- mean 0.123456
- std 0.234567
- min -0.567890
- 25% 0.000000
- 50% 0.123456
- 75% 0.345678
- max 1.000000
- A histogram graph displaying the distribution of positive, negative, and neutral tweets.