Twitter Data Pipeline Project Overview

In this project, we've developed a comprehensive system to collect, store, and analyze Twitter data. The project consists of the following major components:

1. **Setting Up the Environment**:
   - Installed Python libraries such as `tweepy`, `psycopg2`, `pandas`, `matplotlib`, `seaborn`, and `wordcloud`.
   - Set up access to Twitter API using Bearer Token and PostgreSQL for data storage.

2. **Fetching Data from Twitter**:
   - Used `tweepy` to fetch the latest tweets from a specific user (e.g., Elon Musk).
   - Retrieved specific fields such as tweet text, created time, and engagement metrics (likes, retweets).

3. **Storing Data in PostgreSQL**:
   - Created a PostgreSQL database and table to store tweet data.
   - Inserted tweets into the database using SQL queries with conflict handling to avoid duplicate data.

4. **Data Cleaning and Manipulation (Using Pandas)**:
   - Loaded tweet data into a Pandas DataFrame for easy analysis.
   - Performed exploratory data analysis (EDA) and handled datetime formatting.

5. **Data Visualization**:
   - Visualized tweet engagement (likes, retweets) using bar charts.

- Created a word cloud to display frequent words in tweets.
- Visualized sentiment analysis results using pie charts and bar graphs.
- Analyzed tweet frequency over time and visualized the results.

6. **Handling Rate Limits and Errors**:
   - Managed rate limits using error handling in the code to avoid API request issues.

7. **Further Enhancements**:
   - Plans for interactive visualizations with `Plotly` and advanced sentiment analysis with VADER.
   - Creating a dashboard for real-time data analysis.

This project provides insights into social media trends and engagement metrics through a data pipeline that integrates API data collection, storage, analysis, and visualization.