**Sentiment Analysis**

By Ayushman Singh

**Project Description**

This Python project performs sentiment analysis on Twitter data related to Keir Starmer. It leverages various libraries to achieve the following:

**1. Environment Setup and Twitter Login (Automated, Not Included for Security Reasons)**

* The code imports necessary libraries for Selenium WebDriver (<https://www.selenium.dev/documentation/>), enabling interaction with a headless Chrome browser.
* **Crucially, sensitive login credentials are not included in the code snippet due to security concerns.** You'll need to replace placeholders with your actual Twitter login details for the automation to work.

**2. Scraping Tweets**

* The script logs in to Twitter (automated part omitted) and searches for tweets mentioning "Keir Starmer."
* It retrieves a set number of tweets (default 50, adjustable) using XPath locators to identify tweet elements.
* Scrolling is implemented to fetch more tweets until the desired count is reached.

**3. Cleaning Tweets**

* The scraped tweets are stored in a pandas DataFrame for further processing.
* Regular expressions are used to remove mentions (e.g., @username) and hashtags (e.g., #topic) from the tweet text.
* Stop words (common words like "the", "a", "is") are removed using NLTK's stopwords module for more focused analysis.

**4. Sentiment Analysis with TextBlob**

* The TextBlob library is employed to analyze the sentiment of each cleaned tweet.
* Two key sentiment scores are calculated:
  + **Polarity:** Ranges from -1 (negative) to 1 (positive), indicating the overall sentiment of the tweet.
  + **Subjectivity:** Ranges from 0 (objective) to 1 (subjective), reflecting the writer's opinionated stance.
* Tweets are then categorized based on their polarity:
  + Positive: Polarity > 0
  + Neutral: Polarity = 0
  + Negative: Polarity < 0

**5. Analysis and Visualization**

* A pivot table summarizes the sentiment distribution (positive, neutral, negative) in the scraped tweets.
* The most positive and negative tweets (top 3 by polarity) are identified for further examination.
* Neutral tweets are also extracted if relevant.
* WordCloud is used to create a visual representation of frequently used words in the cleaned tweets, providing insights into the overall sentiment and common topics.
* Seaborn is employed for:
  + A scatterplot that visually depicts the relationship between polarity and subjectivity across individual tweets, colored by sentiment category.
  + A countplot showing the distribution of positive, negative, and neutral tweets.

**6. Conclusion (Based on Sample Output)**

* The concluding remarks would typically be generated based on the analyzed data in your specific run. You'd interpret the polarity distribution, word cloud, and visualizations to draw inferences about public sentiment regarding Keir Starmer on Twitter.
* Consider aspects like:
  + The overall sentiment towards Keir Starmer (leaning positive, negative, or neutral).
  + Prominent words or topics associated with positive and negative sentiment.
  + Whether there's a strong correlation between polarity and subjectivity (indicating a stronger expression of opinion in tweets with greater polarity).

**Remember to replace placeholders with your Twitter(X) login credentials if you intend to run the automated scraping part.**