Method

The CSV file contains tweets with their respective cleaned versions, keywords extracted from each tweet, and sentiment analysis in terms of polarity and subjectivity. Here's a brief structure of the data:

- tweet: The original tweet text.
- cleaned_tweet: The cleaned version of the tweet, possibly with URLs, usernames, or special characters removed.
- keywords: Key words extracted from each tweet.
- polarity: A numerical score representing the sentiment polarity, where a higher positive score indicates a more - - positive sentiment, and a negative score indicates a more negative sentiment.
- subjectivity: A numerical score that represents how subjective or opinionated the tweet is, with 0 being completely objective and 1 being completely subjective.

The analysis of the tweets reveals interesting insights about the keywords and overall sentiment:

Keywords Analysis

The most frequent keywords in the dataset are:

- Delta (mentioned 487 times)
- THC (103 times)
- Like (54 times)
- High (53 times)
- CBD (53 times)
- Just (50 times)
- Shit (41 times)
- Weed (38 times)
- I'm (38 times)
- Legal (37 times)

These keywords suggest that the tweets are heavily focused on topics related to cannabis, particularly Delta-THC, CBD, and legal issues surrounding these substances. The frequent use of words like "high," "weed," and "shit" indicates discussions about the effects and experiences of using these substances.

Sentiment Analysis

Average Polarity: The average polarity score is approximately 0.089. This indicates a generally positive sentiment overall but is closer to neutral, suggesting mixed feelings or a balanced view in the tweets.

Average Subjectivity: The average subjectivity score is around 0.418. This is moderate, meaning the tweets contain objective facts and subjective opinions. It's factual and opinionated, reflecting a balanced expression of thoughts and information.

The dataset significantly focuses on cannabis-related topics, with a balanced mix of positive and slightly neutral sentiments. The moderate subjectivity indicates a blend of personal opinions and factual information in the discussions.