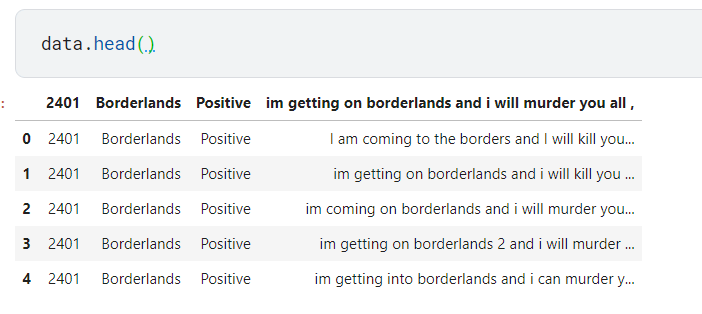
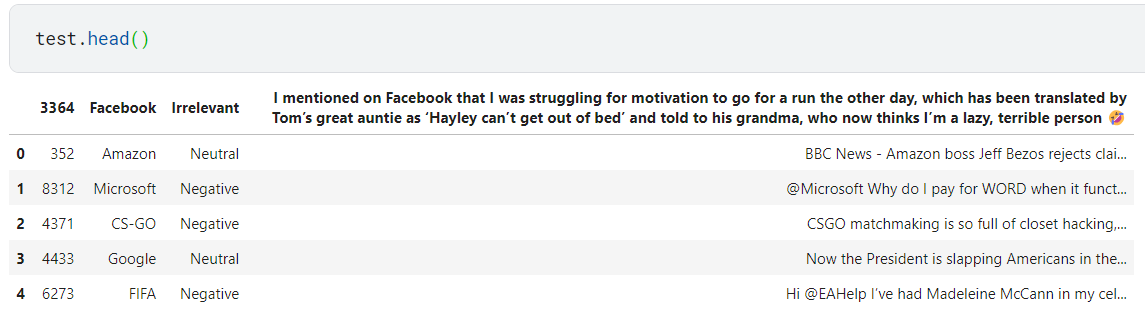
Preparing Data

Twitter sentiment analysis involves analyzing the sentiment expressed in tweets. To prepare the data, we collect tweets related to a specific topic or keyword using the Twitter API. We clean the data by removing irrelevant information like URLs and hashtags. We perform tokenization to break the tweets into individual words or phrases. Next, we apply techniques like stemming or lemmatization to normalize the words. Finally, we label the tweets as positive, negative, or neutral to create a labeled dataset for training a sentiment analysis model.





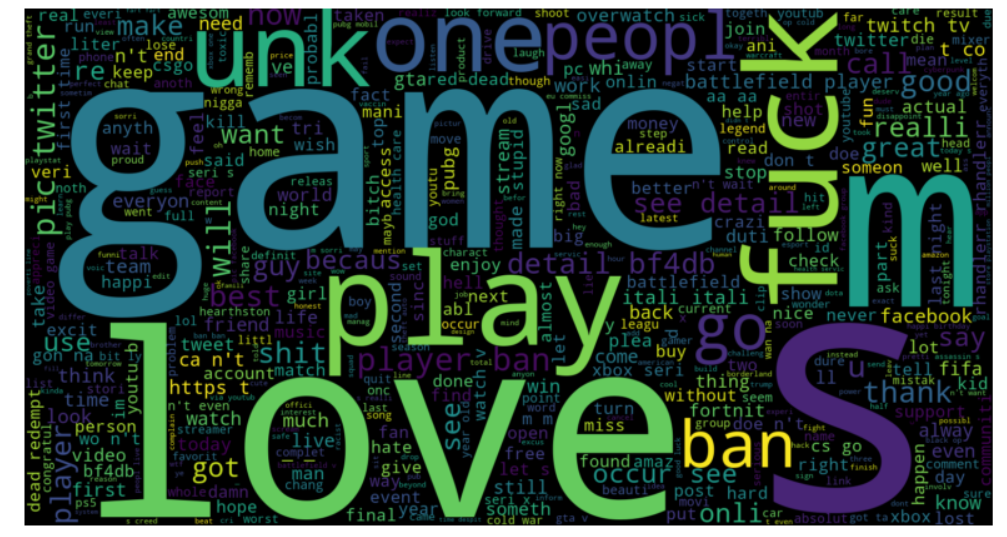
Labeled Data

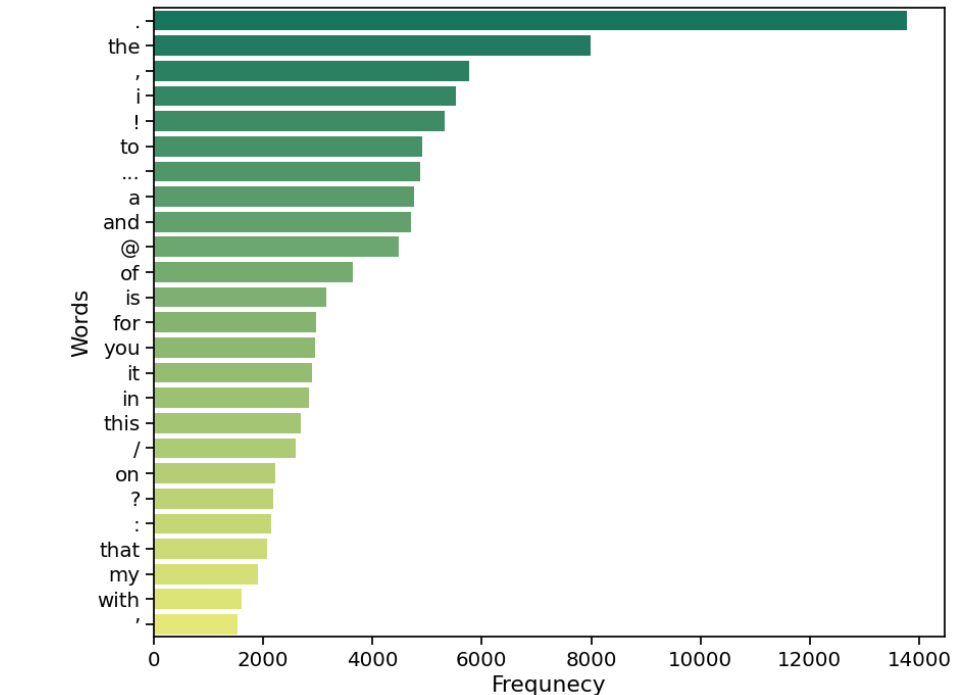
In labeled data for Twitter sentiment analysis, each tweet is assigned a sentiment label such as positive, negative, or neutral. This labeling process is typically done manually by human annotators. The labeled dataset serves as the ground truth for training supervised machine learning models. It helps the model learn patterns and relationships between tweet content and sentiment. The labeled data enables the model to predict sentiment accurately when applied to new, unlabeled tweets.

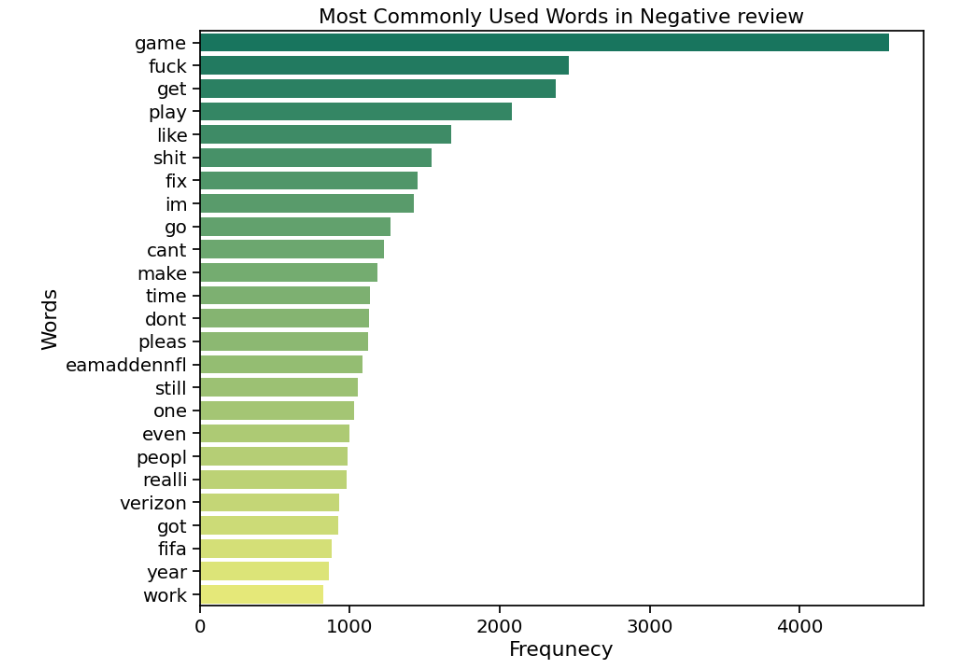


Data Representation

In Twitter sentiment analysis, data representation refers to the way in which the textual data from tweets is transformed into a numerical format suitable for machine learning algorithms. Common representations include bag-of-words, where each word is represented as a separate feature, and TF-IDF, which considers the importance of a word in a document. Other approaches involve using word embeddings like Word2Vec or GloVe to capture semantic relationships. These representations enable algorithms to process and analyze the sentiment expressed in tweets.







Compairing the model

, the model I am based on, has a knowledge cutoff in September 2021 and may not have been specifically trained for Twitter sentiment analysis. For Twitter sentiment analysis, there are specialized models and techniques that consider the unique characteristics of Twitter data, such as hashtags, emojis, slang, and abbreviations. It's recommended to explore models like BERTweet or Sentiment140 specifically designed or fine-tuned for this task. Newer models or techniques may be available beyond my knowledge cutoff.

