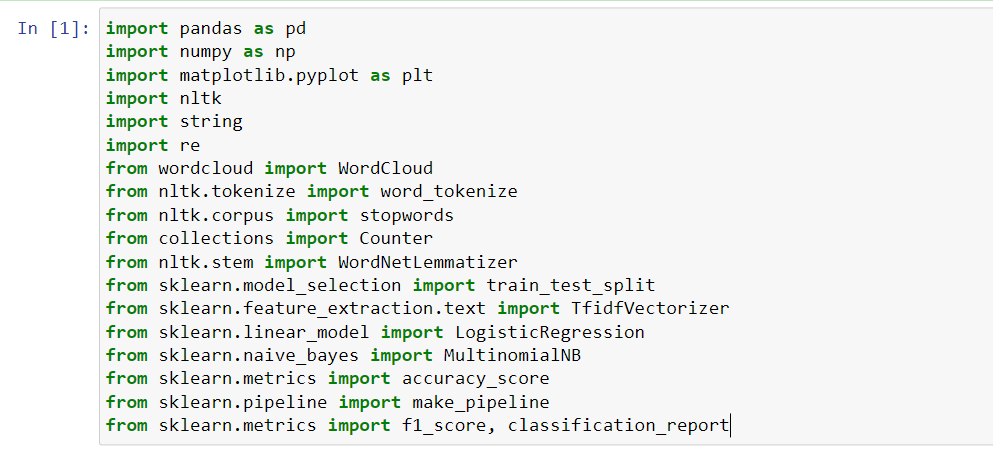
**Twitter Sentiment Analysis**

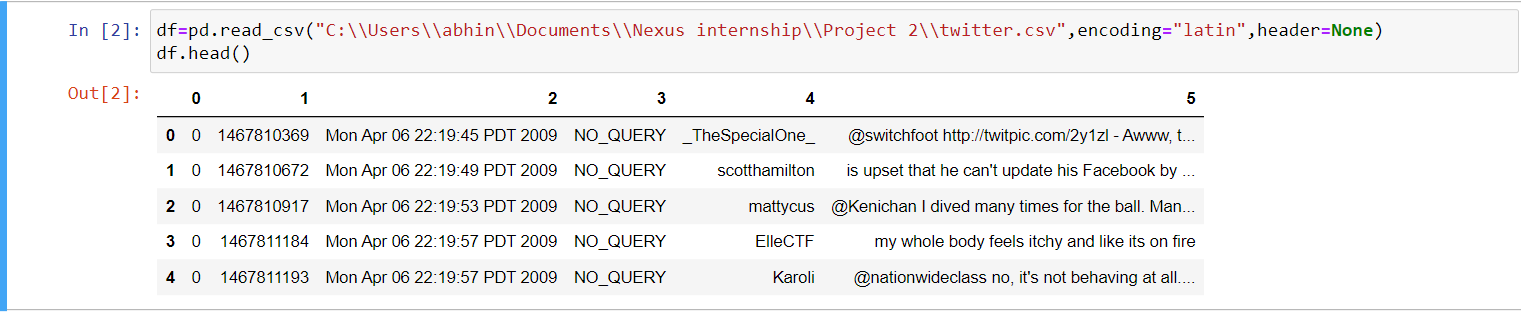
**Project Overview**: Twitter Sentiment Analysis is a data analytics project that involves analyzing a dataset of tweets to determine the sentiment expressed in each tweet—whether it is positive, negative, or neutral. The project aims to gain insights into public opinions, trends, and sentiments shared on Twitter, utilizing data analytics techniques.

**Project Objectives**:

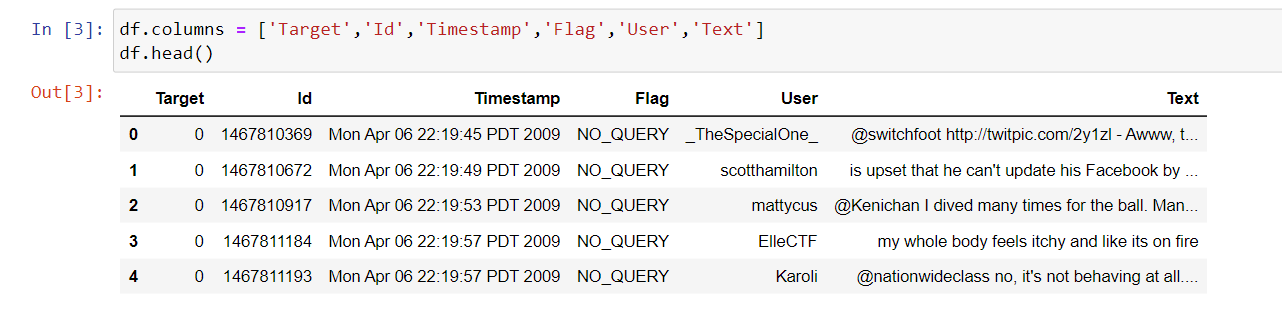
For this project analysis the necessary libraries are



Load data into the data frame from csv file

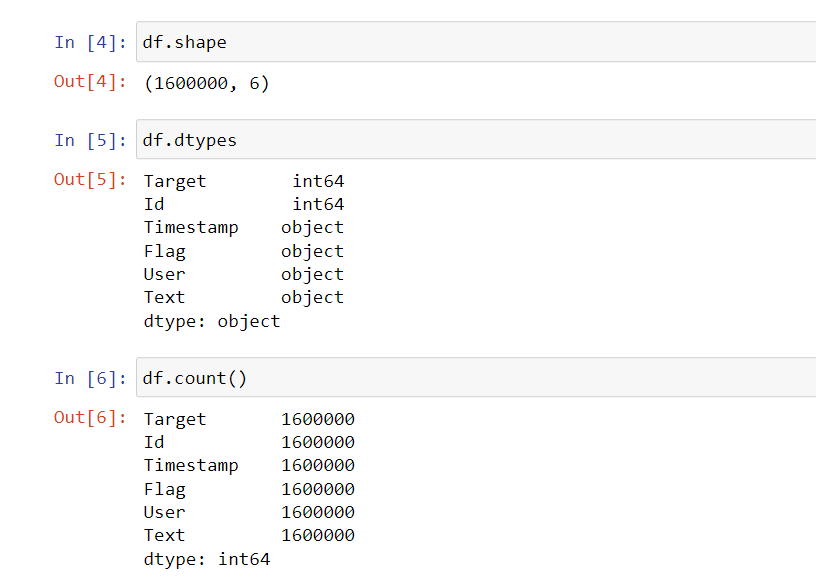


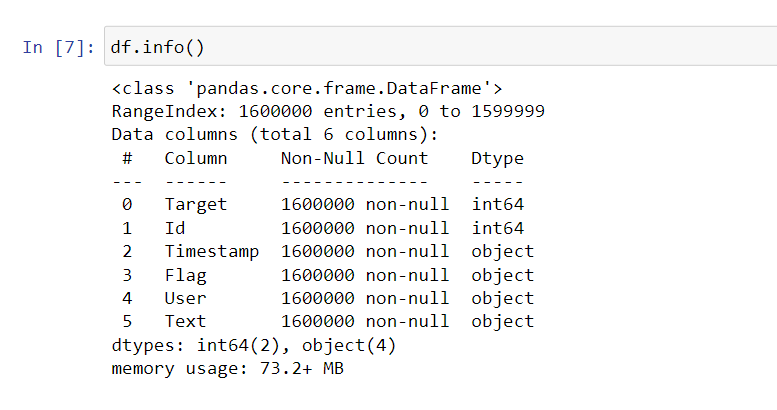
Allocating headers to the columns



1**.Data Exploration:**

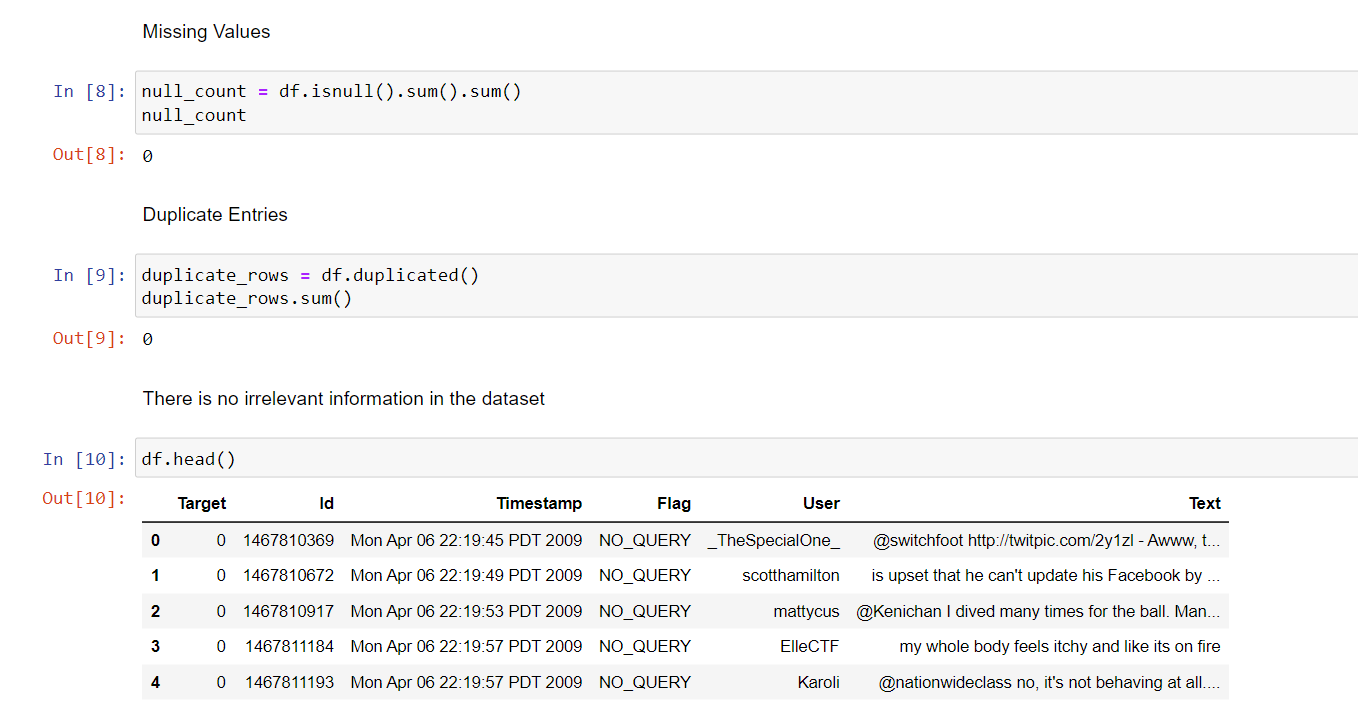
Explore the Sentiment dataset to understand its structure, features, and size. Identify key variables such as tweet content, timestamp, and sentiment labels.

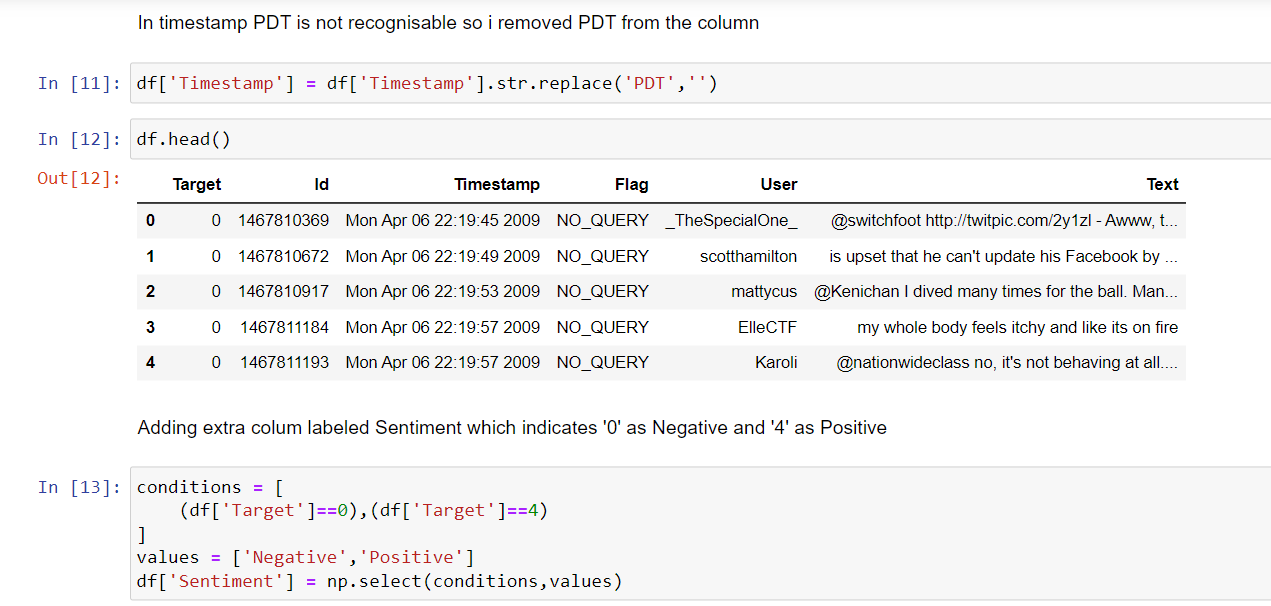


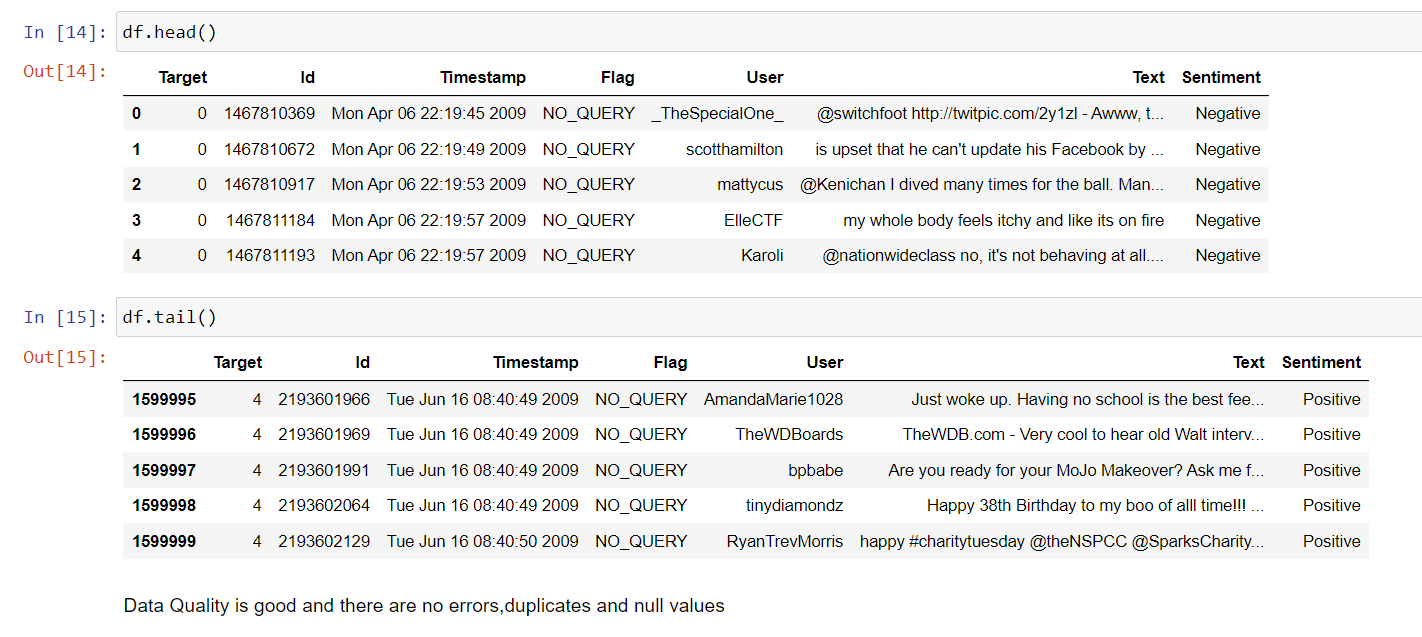


**2.Data Cleaning:**

Perform data cleaning tasks to handle missing values, duplicate entries, and irrelevant information. - Ensure data quality by addressing any anomalies or inconsistencies in the dataset.

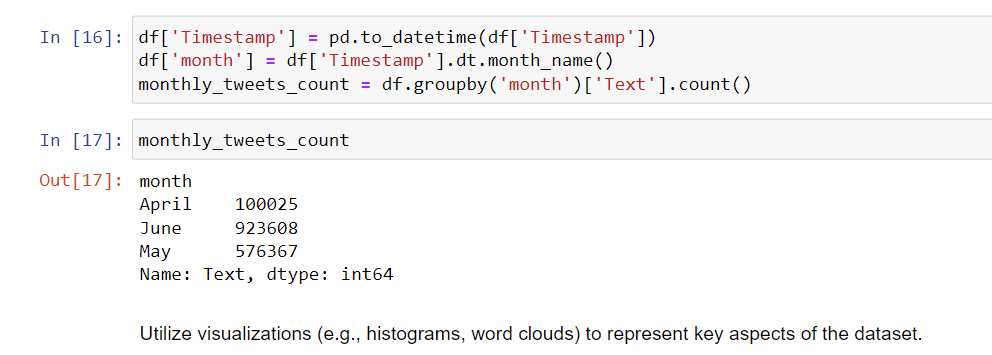




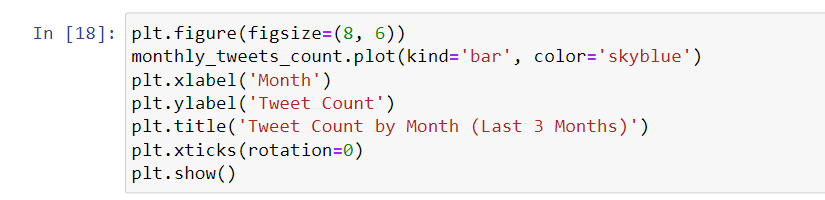


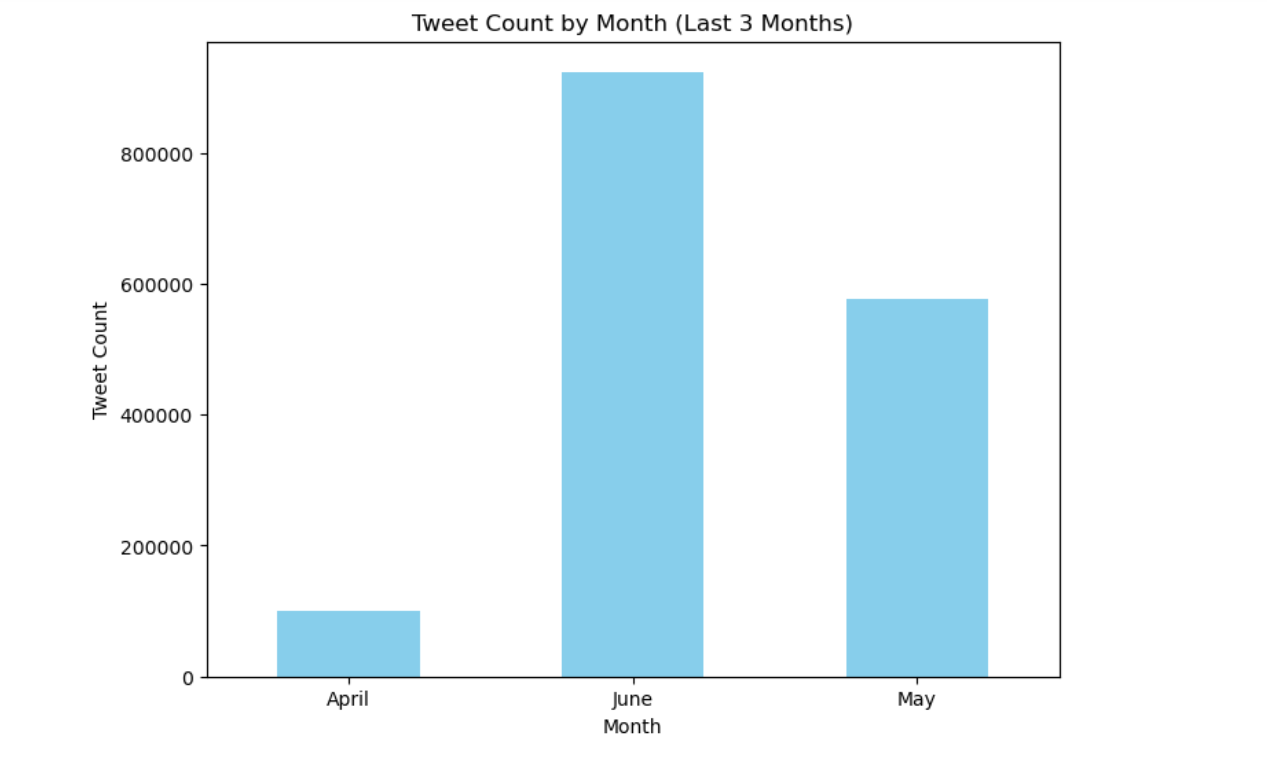
**3.Exploratory Data Analysis**

Conduct exploratory data analysis to gain initial insights into tweet patterns, sentiment distributions, and temporal trends.



Utilize visualizations (e.g., histograms, word clouds) to represent key aspects of the dataset.

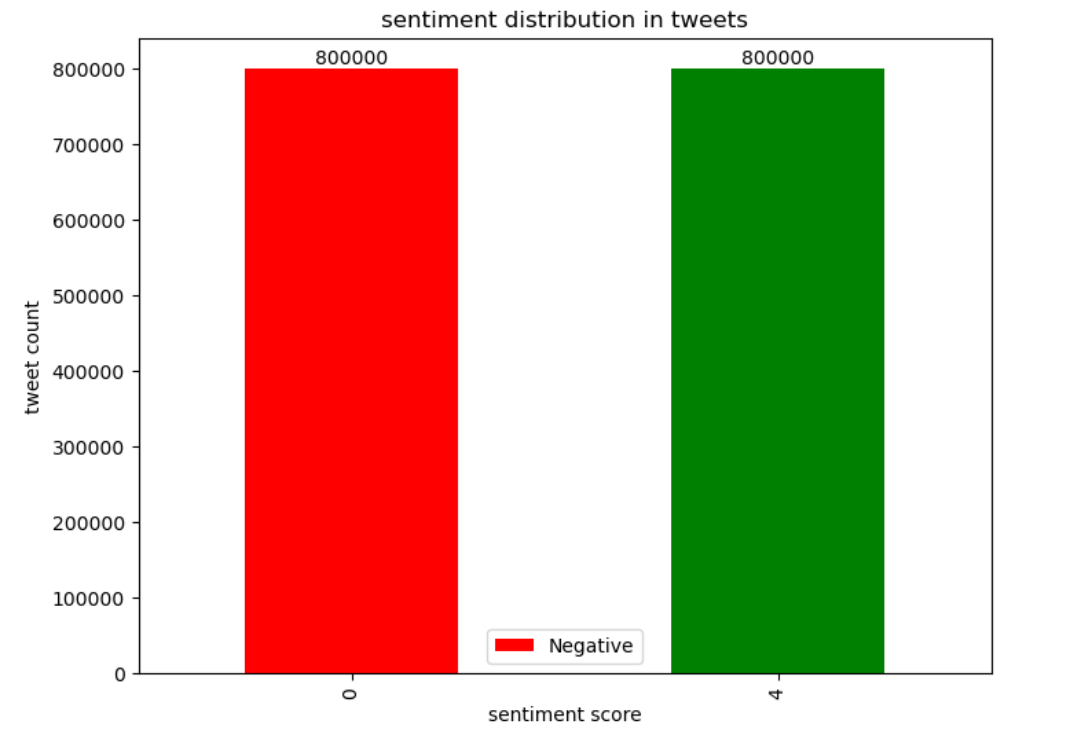




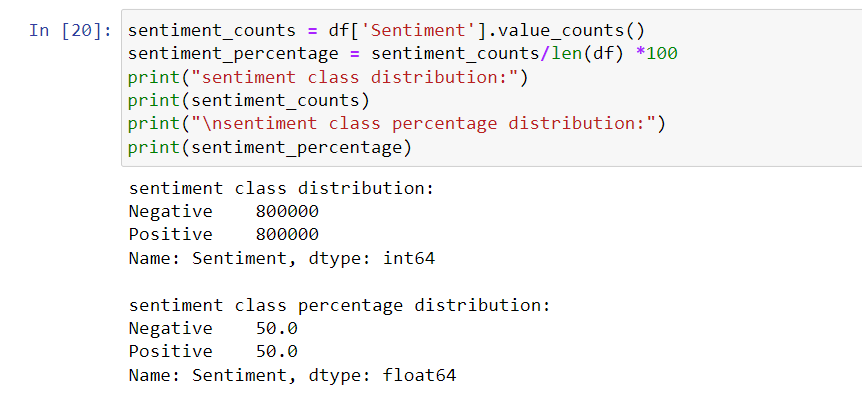
**4.Sentiment Distribution**

Visualize the distribution of sentiment labels (positive, negative, neutral) in the dataset.



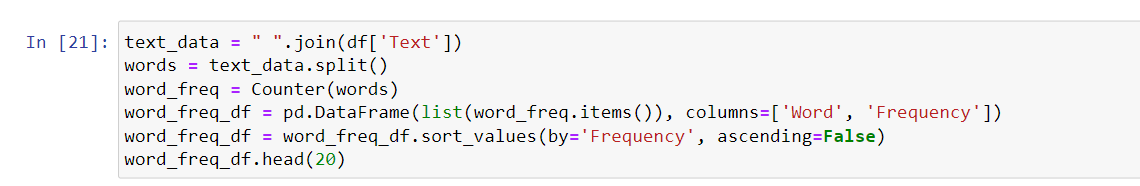


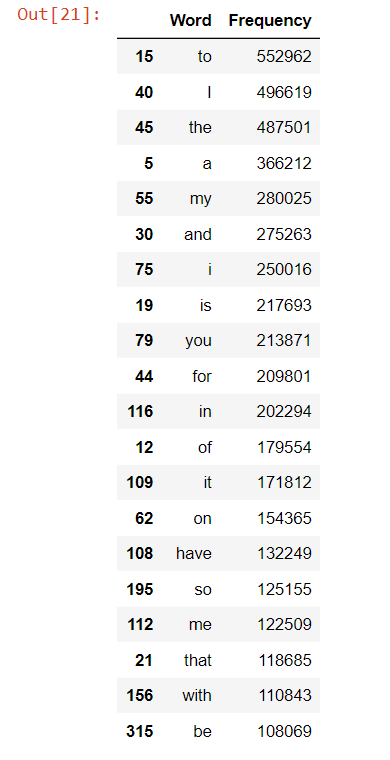
Analyze the balance of sentiment classes to understand potential biases.



**5.Word Frequency Analysis:**

Analyze the frequency of words in tweets to identify common terms and themes.

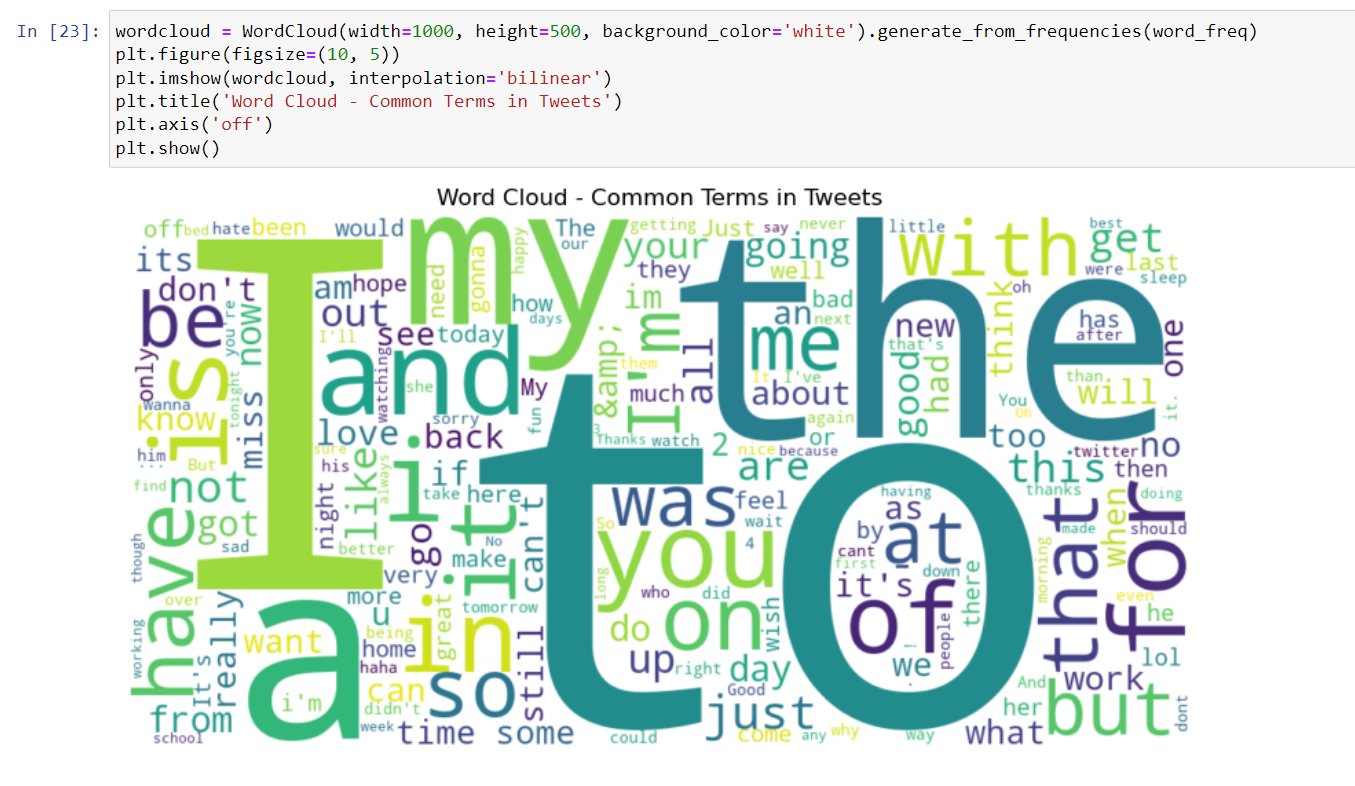




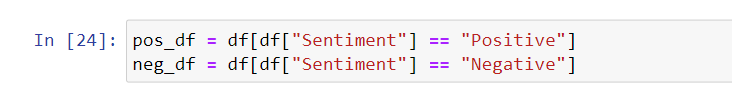
Bar Chart for most frequent words



Word cloud for most frequent words



Create word clouds or bar charts to visualize the most frequent words in positive and negative sentiments.



Positive Sentiment Word cloud

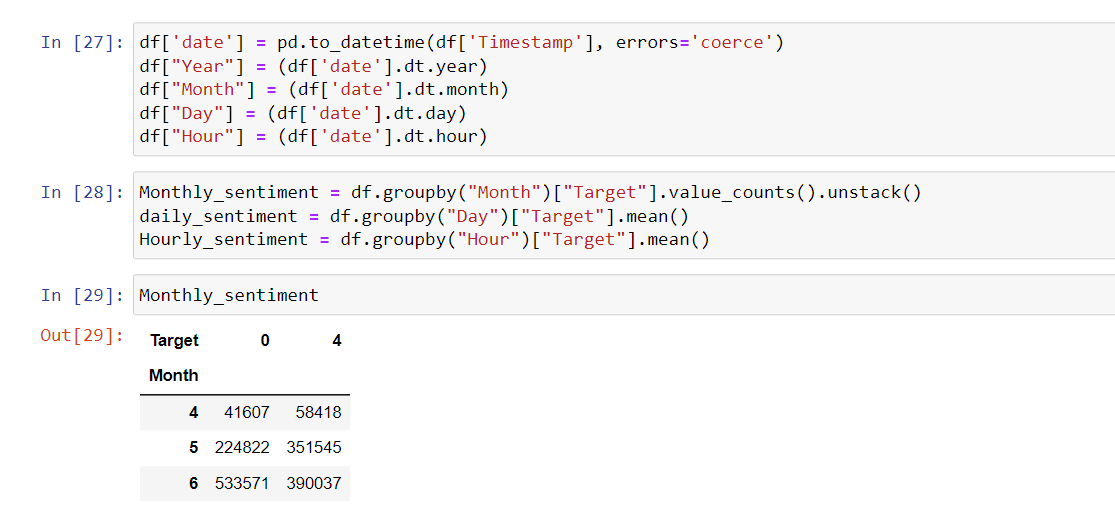


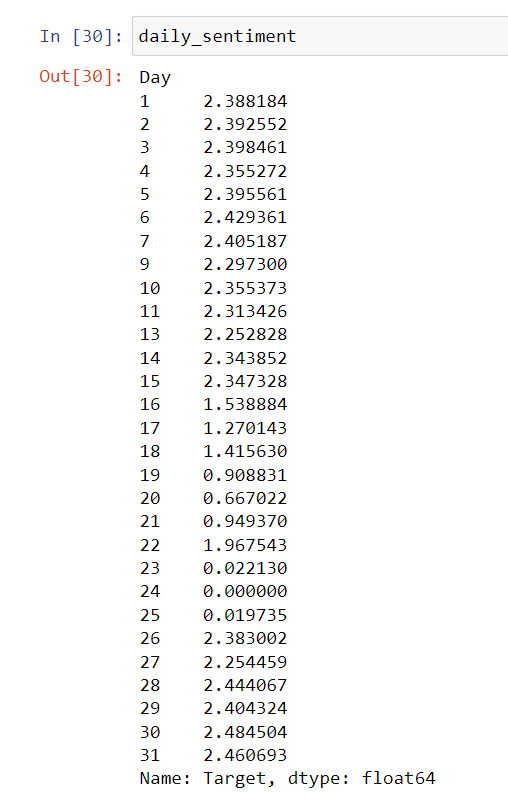
Negative Sentiment word cloud

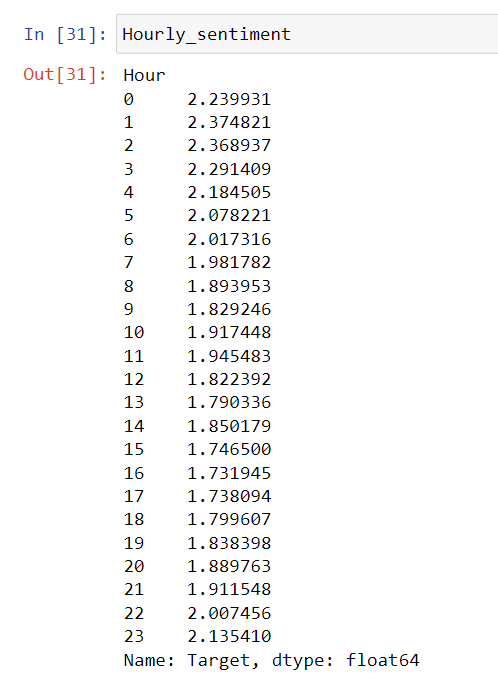


**6.Temporal Analysis**

Explore how sentiment varies over time by analyzing tweet timestamps.

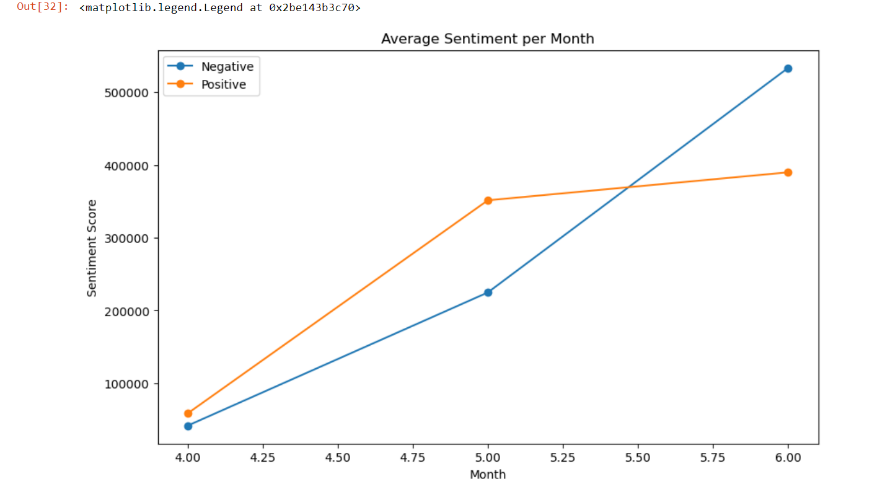


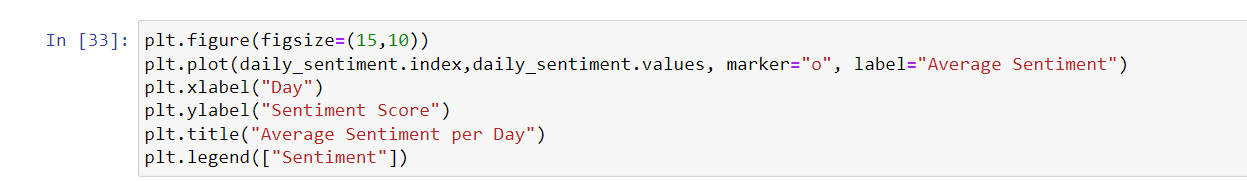


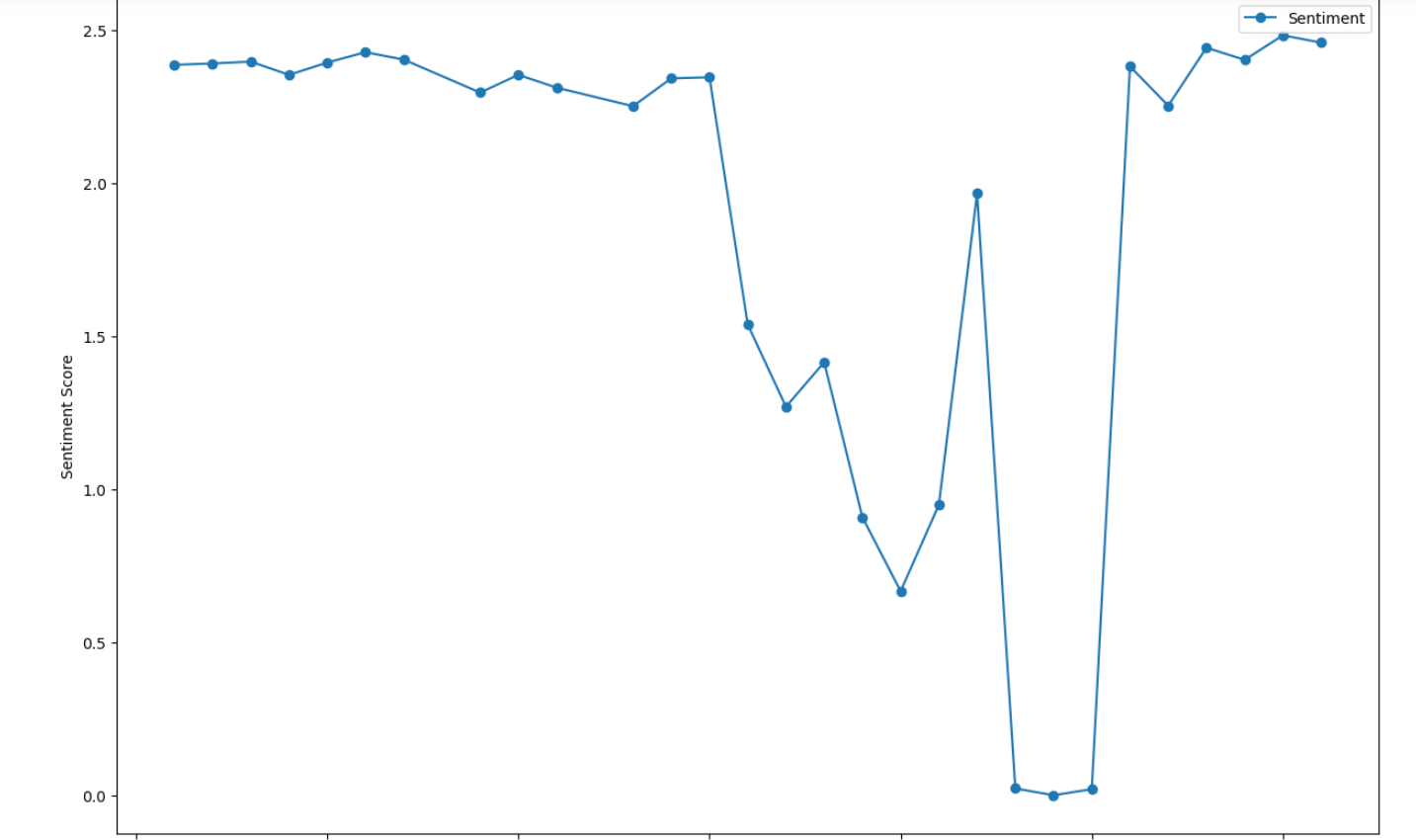


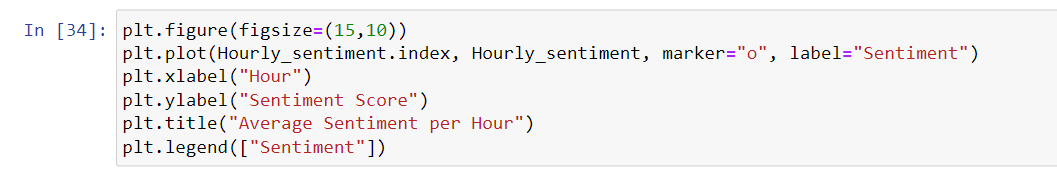
Identify patterns, peaks, or trends in sentiment within specific time frames.

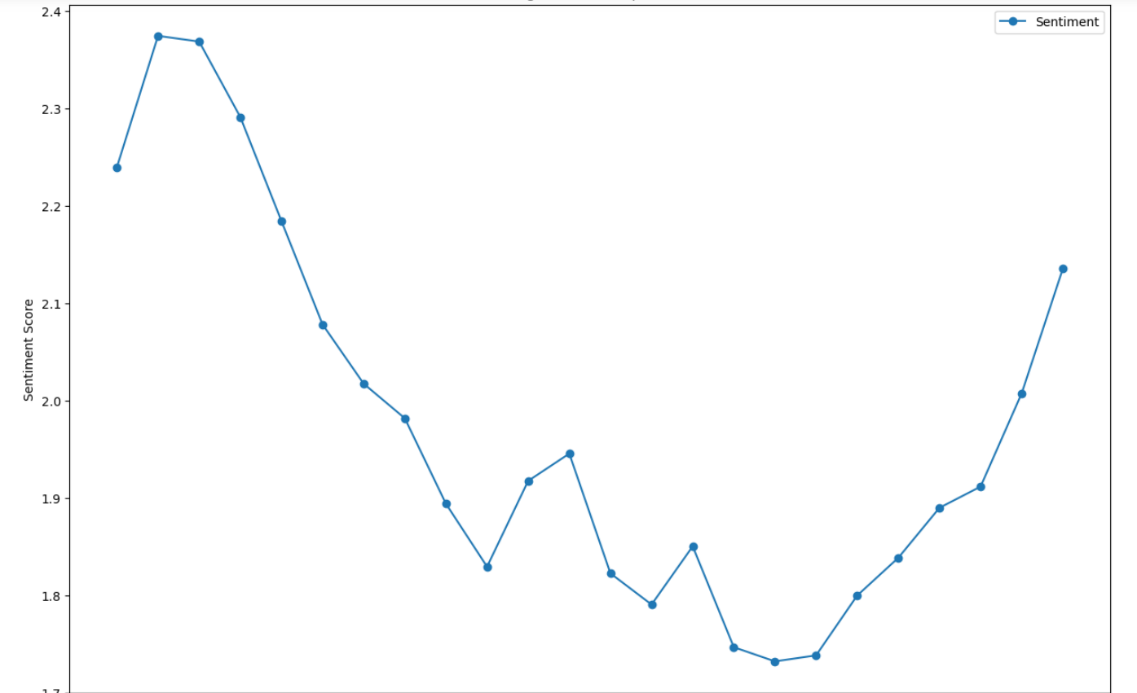








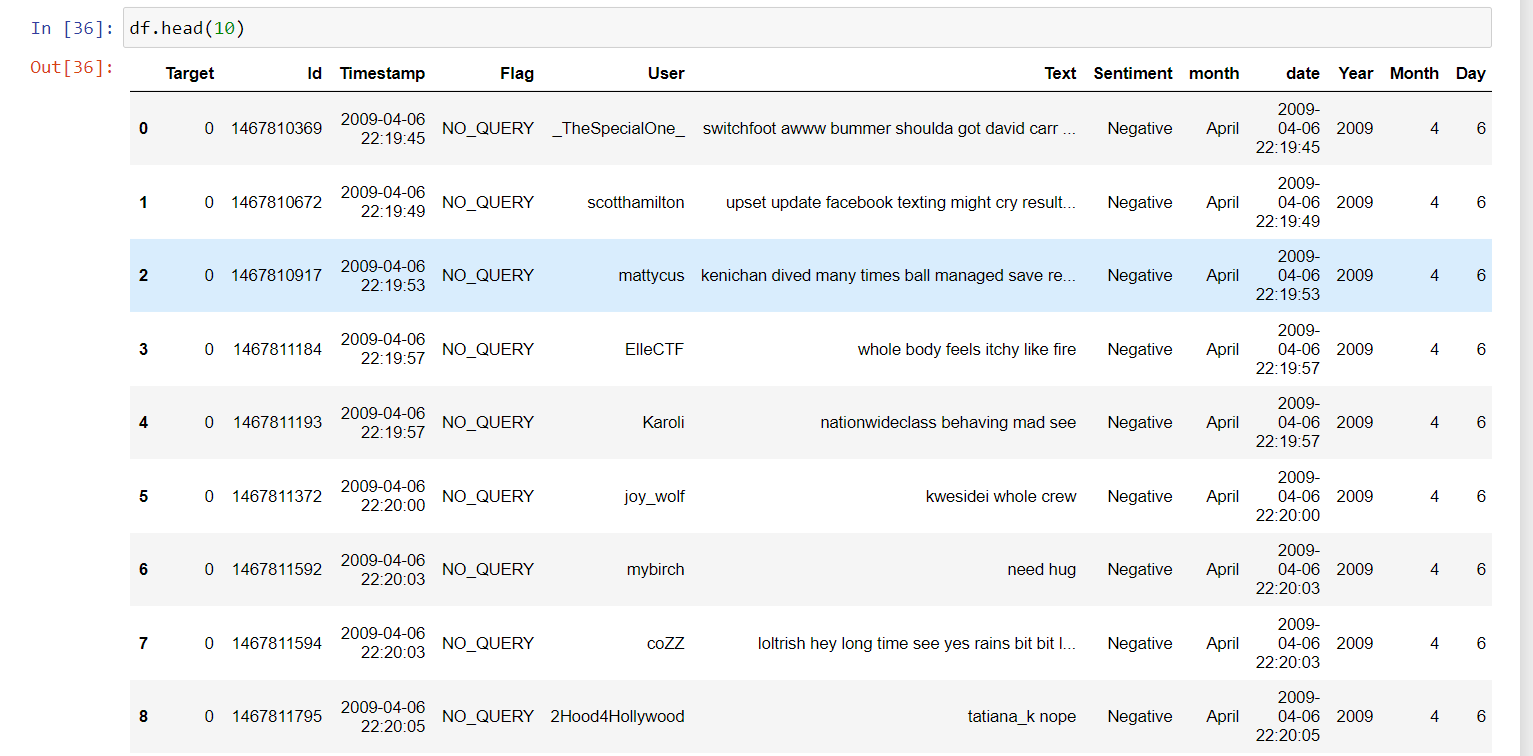


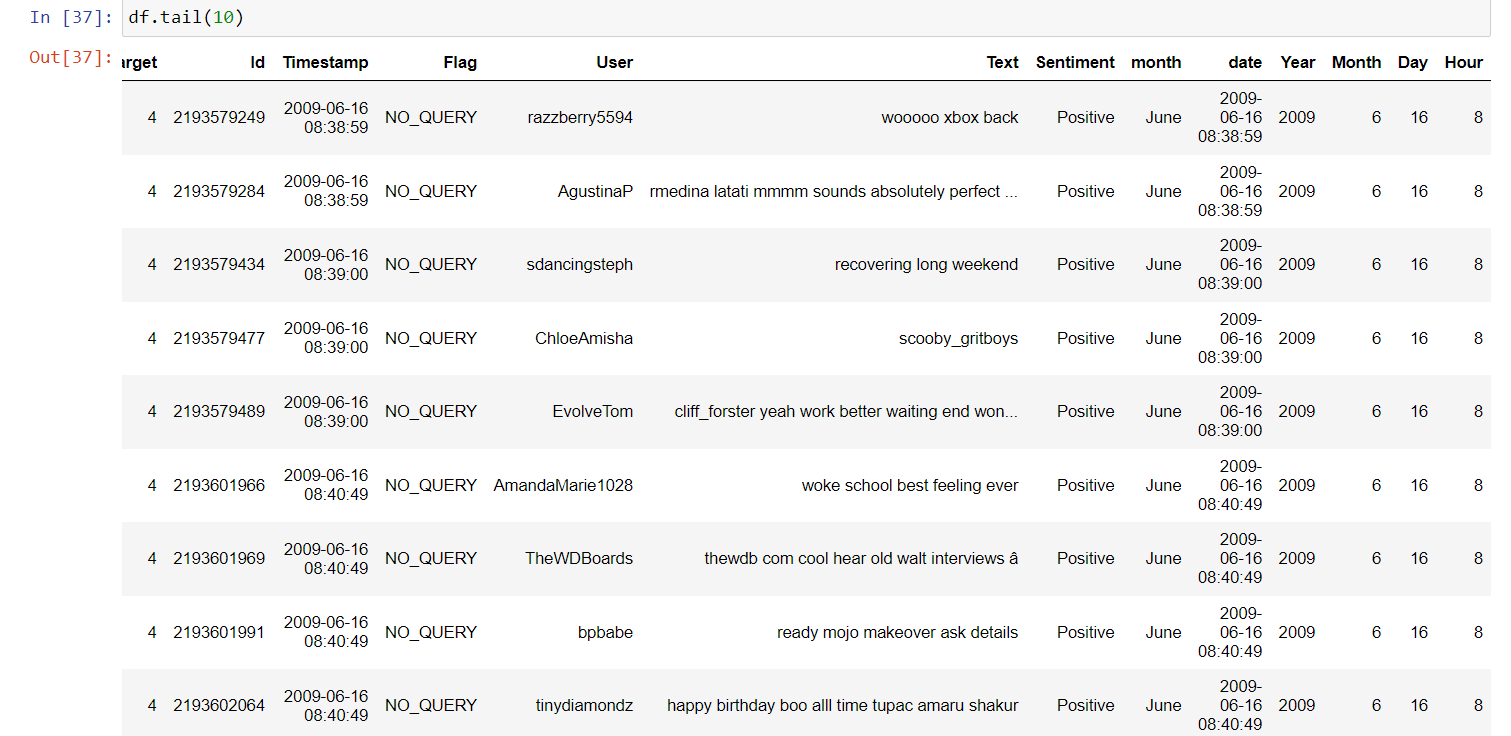


**7.Text Preprocessing**

Preprocess tweet text by removing stop words, special characters, and URLs.

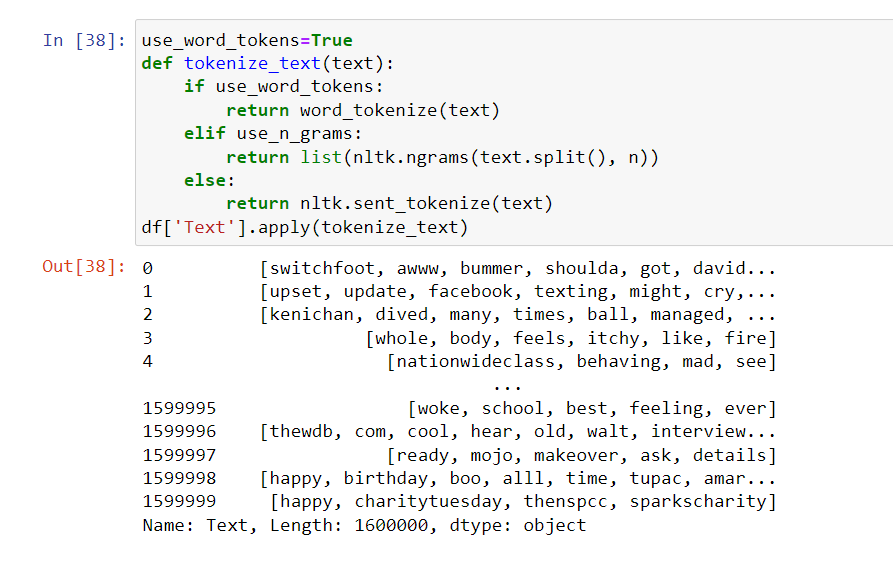




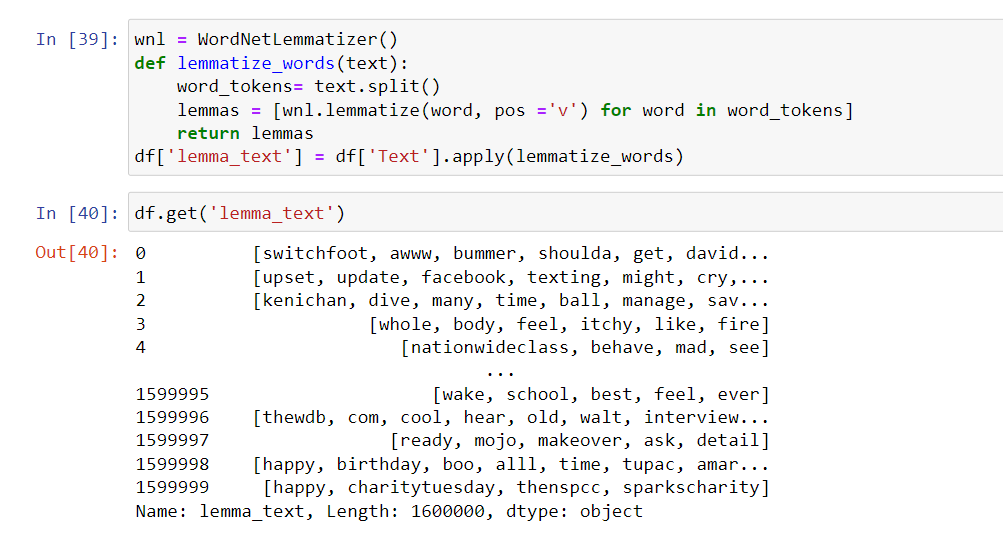


Tokenize and lemmatize words to prepare the text for sentiment analysis.

Tokenization

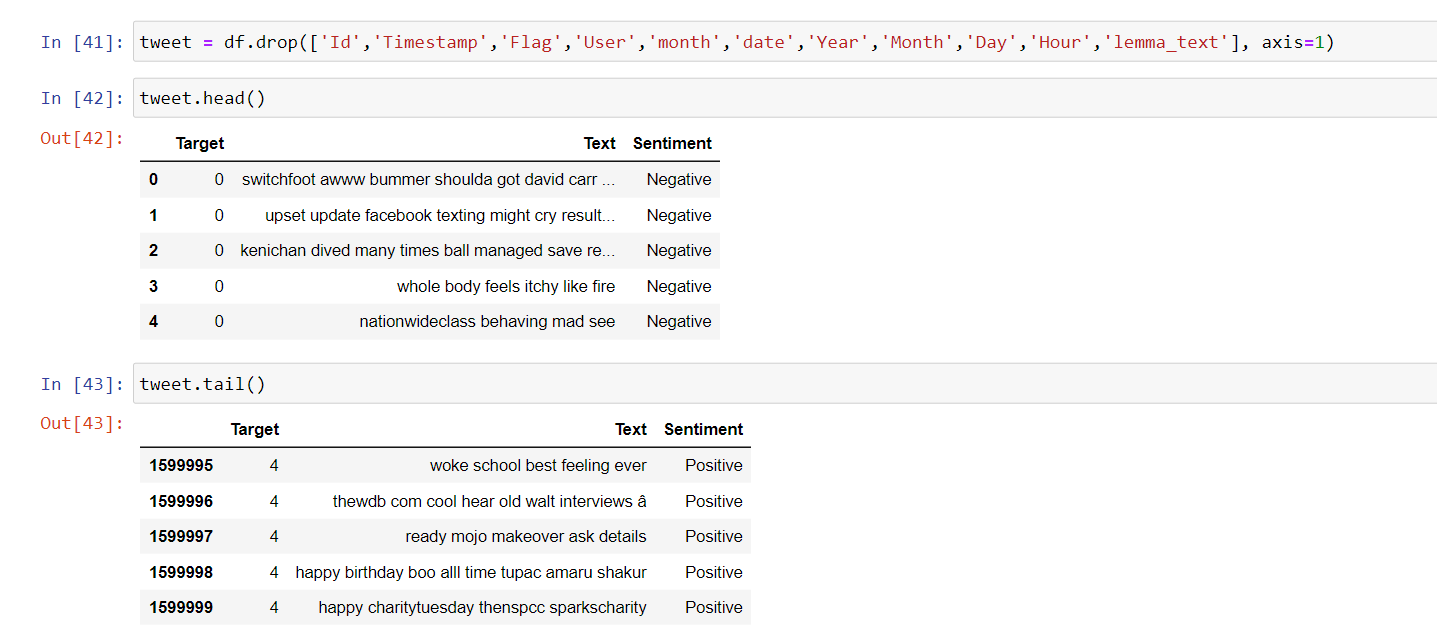


Lemmatization



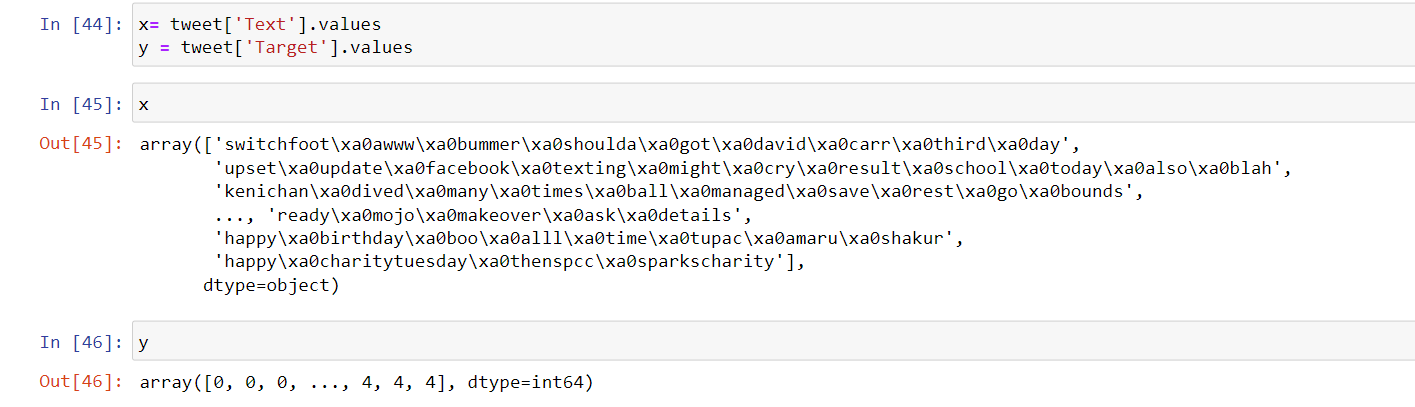
**8.Sentiment Model Prediction**

Implement a sentiment prediction model using machine learning or natural language processing techniques.

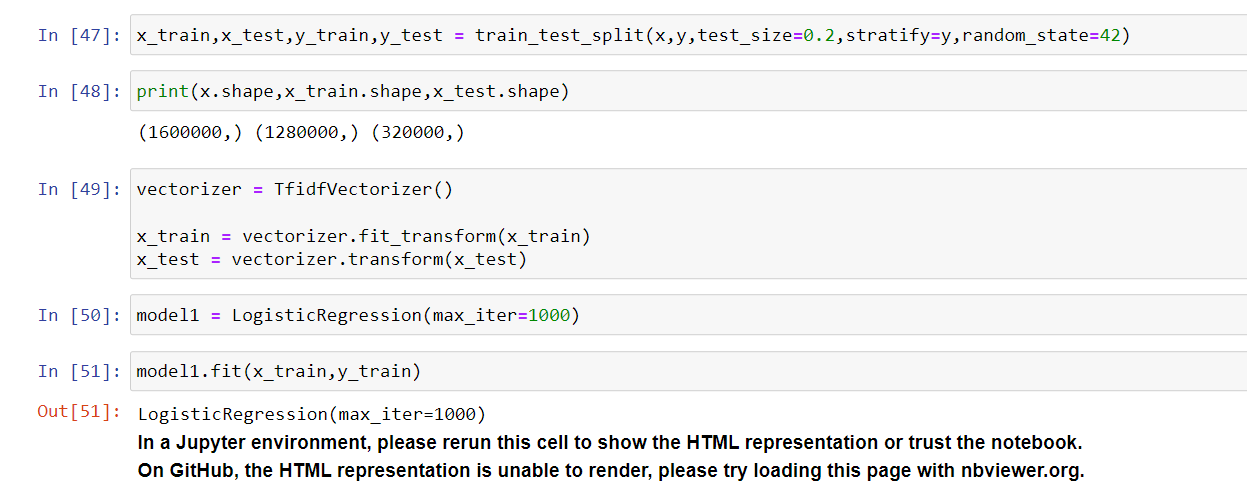


I created two models one is trained and tested with index number to get sentiment and another model is trained and tested with words in tweet text.

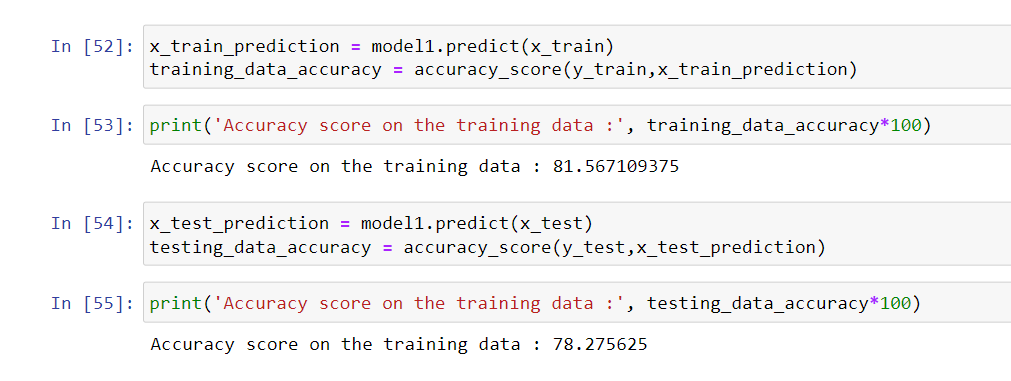
Model – 1:



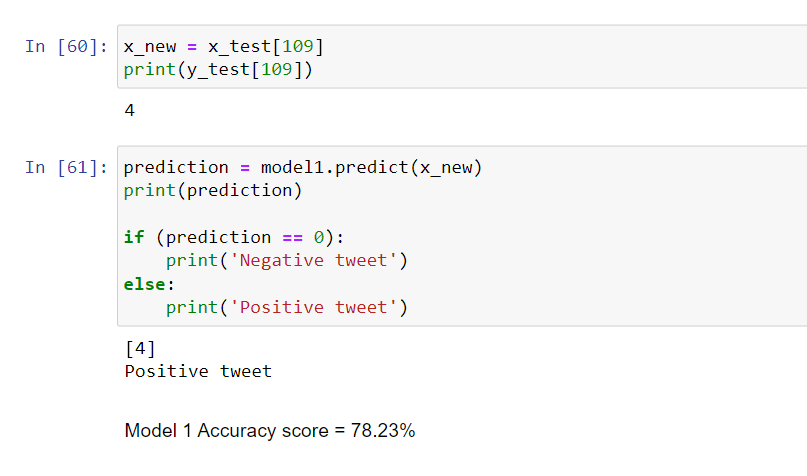
Train the model on a subset of the dataset



Evaluate its performance using metrics like accuracy



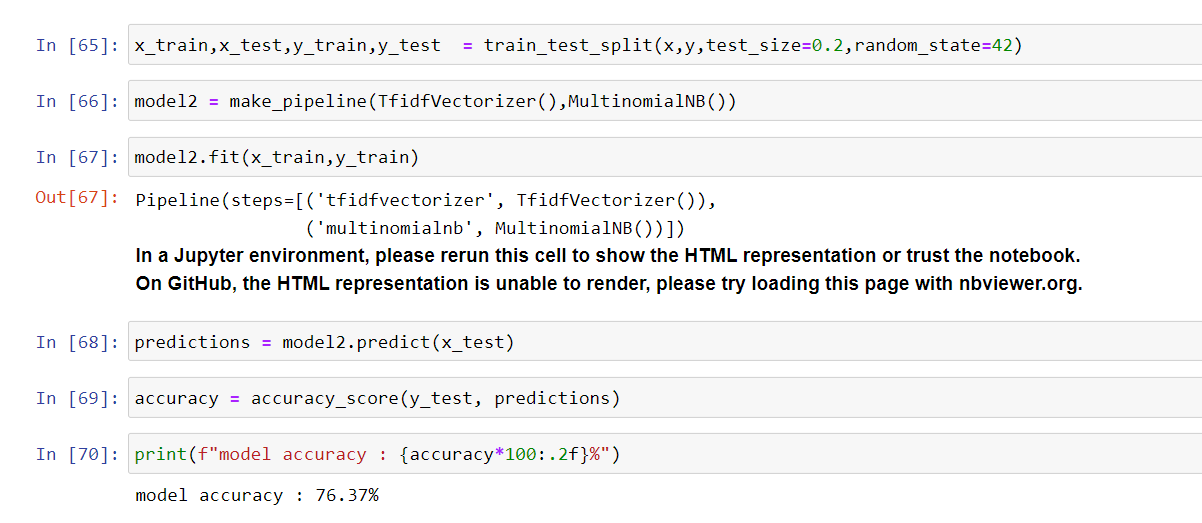




Model – 2

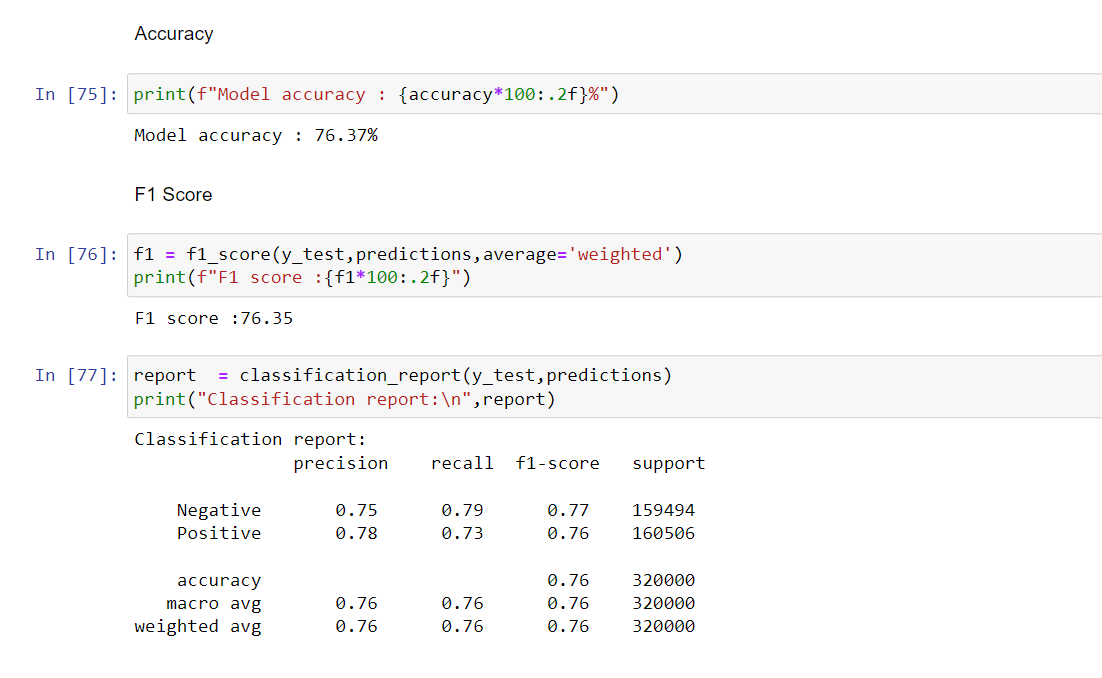


Train the model on a subset of the dataset





Evaluate its performance using metrics like accuracy and F1 score.



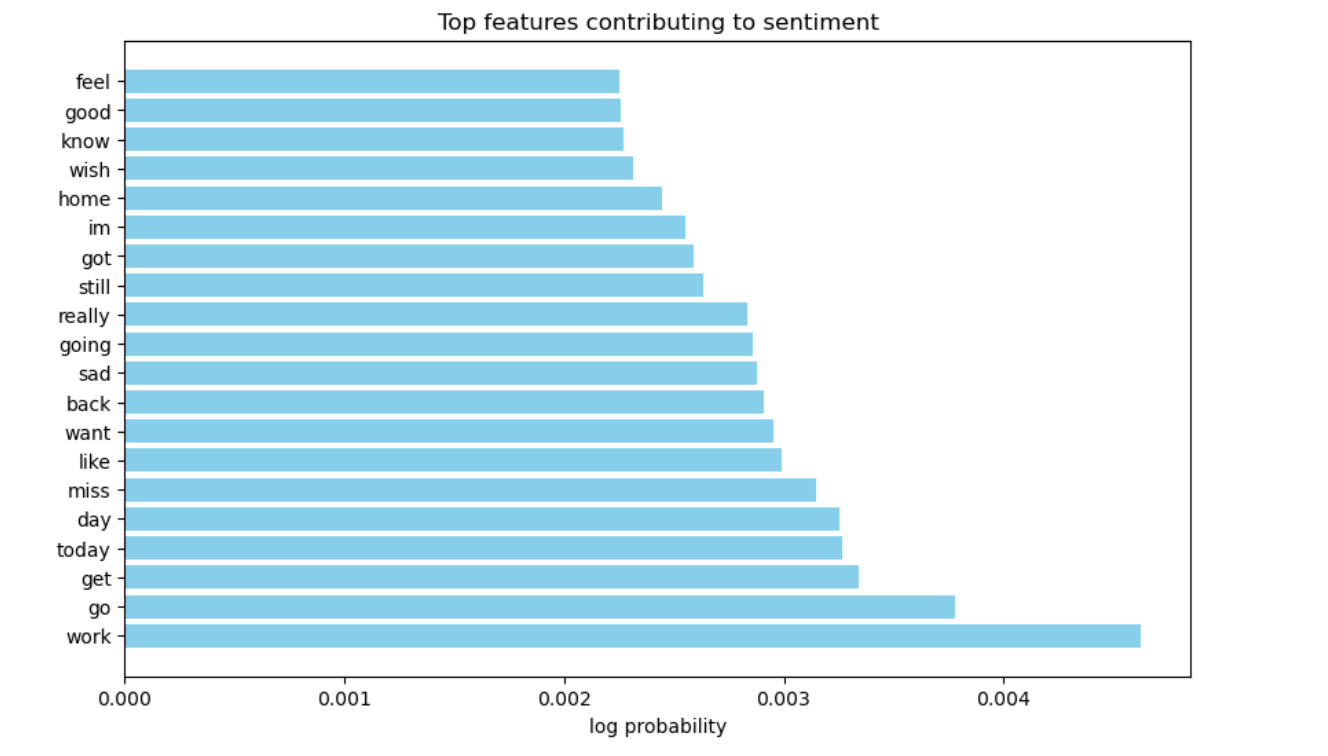
**9.Feature Importance**

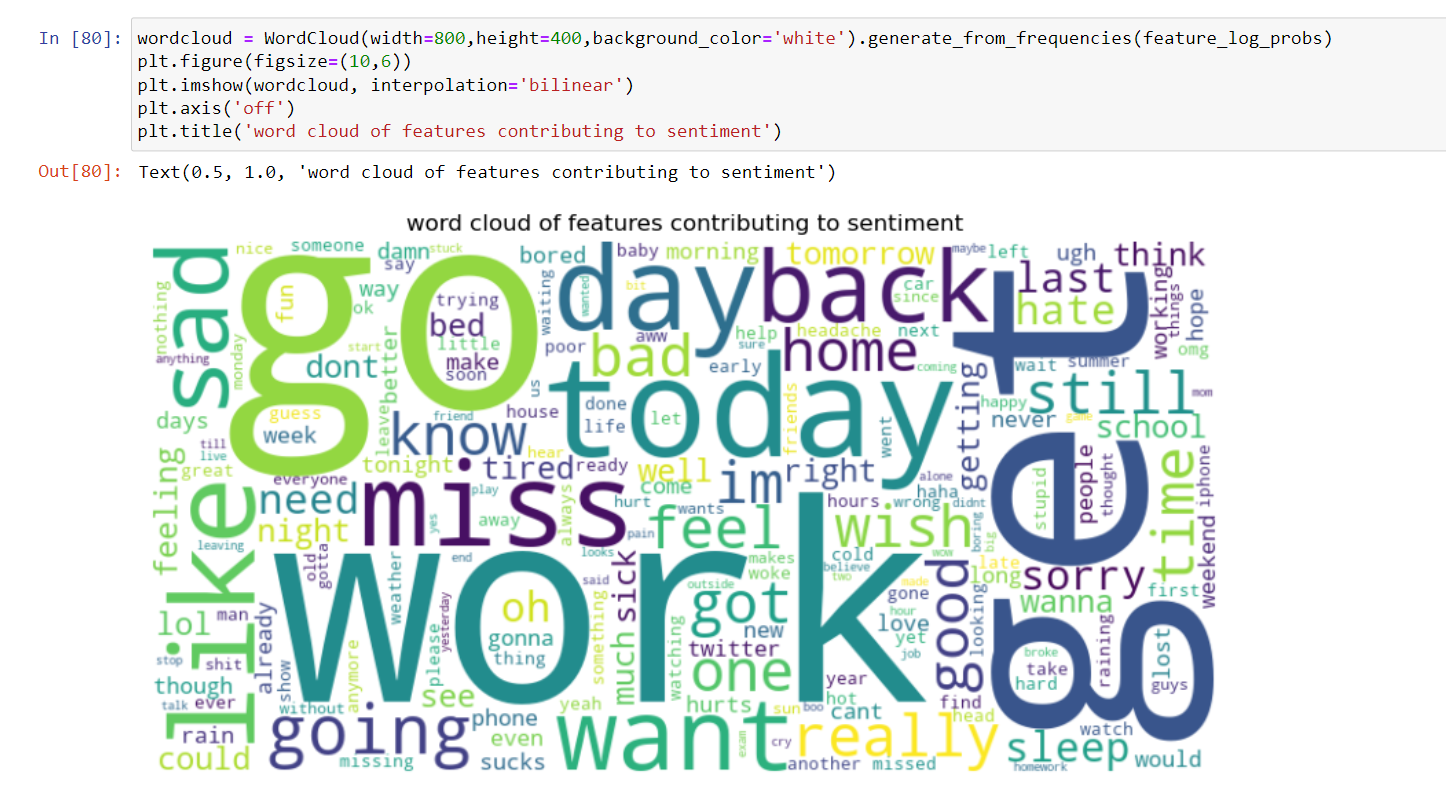
Identify the most important features (words or phrases) contributing to sentiment predictions.



Visualize feature importance using techniques such as bar charts or word clouds.







Insights and Recommendations:

June month has the highest number of tweets : 923608

April month has lowest number of tweets : 100028

Distribution of sentiment positive and negative are same from the tweets 800000 positive tweets and 800000 negative tweets

23,24,25 dates have the sentiment of 0 and another dates have sentiment near to 2

2 Am has the highest avg sentiment score and 4 Pm has lowest avg sentiment score

People tends to show their emotion through tweets