****

**A REPORT ON SENTIMENT ANALYSIS**

**DEPARTMENT OF SOFTWARE ENGINEERING AND COMPUTER SCIENCE,**

**COLLEGE OF COMPUTING,**

**MCPHERSON UNIVERSITY , SERIKI SOTAYO, OGUN STATE.**

**BY: GROUP 4**

**TABLE OF CONTENTS**

[**1.0 INTRODUCTION** 3](#_Toc189166398)

[**1.1 APPLICATIONS OF SENTIMENT ANALYSIS** 3](#_Toc189166399)

[**2.0 METHODOLOGY** 4](#_Toc189166400)

[**2.1 STEPS FOLLOWED** 4](#_Toc189166401)

[**2.2 METHODOLOGY USED** 5](#_Toc189166402)

[**3.0 CHALLENGES FACED** 6](#_Toc189166403)

[**4.0 CONCLUSION** 6](#_Toc189166404)

# **1.0 INTRODUCTION**

Sentiment Analysis or Opinion Mining as people like to call it is a natural language processing (NLP) technique used to determine the tone of a text. This analysis is the classification of textual data either from CSVs or live data on social media posts(basically any type of textual data that can be classified) into different sentiment categories such as positive, negative and neutral. Some models go as far as detecting emotions like sarcasm, anger, joy or sadness.

How does Sentiment Analysis work? it involves several steps involved some of which include:

1. **Data Collection**: This process consists of gathering textual data from various sources such as social media, surveys and customer feedback. In this report of the analysis, social media content was used.
2. **Data Cleaning**: is the process of removing words that may not convey any meaning or help in the determination of sentiment. For example: punctuation, special characters and stopwords)
3. **Extraction:** involves the identification of key terms, phrases and clues that might indicate sentiment or emotion.
4. **Classification:** this is the use of machine learning models or deep learning techniques to classify each sentiment into its category whether it be positive, negative or neutral.

## **1.1 APPLICATIONS OF SENTIMENT ANALYSIS**

Understanding public sentiment is important for businesses, governments and organizations. Opinion mining enables these authoritative figureheads to extract meaningful insights from large or small amounts of textual data, which helps in decision-making.

The following listed include how and where sentiment analysis can be used:

1. Customer Feedback and Support: Sentiment analysis can be useful by categorizing how the customers feel based on their feedback which can in turn be used to improve products and services.
2. Competitor Analysis: This helps organizations understand how consumers feel about their competitor's products and prediction of market trends can then be generated based on sentiment shifts and whatnot.
3. Content Moderation: Sentiment analysis detects emotion and sentiment, hence it can detect hate speech and unprecedented comments from online platforms and once they are discovered they can be flagged and enhancement of community guidelines becomes easier.
4. Mental Health Monitoring: It can be useful in the healthcare department by analysing patient feedback and comments through online discussions. Mental health can also be analysed through the types of comments a person posts. An example of such analysis is when you type suicide in the TikTok search bar, a mental health note pops up telling you everything will be fine, this is possible through the detection of keywords that indicate signs of mental illness.

# **2.0 METHODOLOGY**

To implement a functional sentiment analysis the social media app: REDDIT was used. A post on Reddit: “How El Salvador deals with paedophiles” was picked to analyse. The steps and methods used in the process of analysing this post will highlighted and explained extensively below:

## **2.1 STEPS FOLLOWED**

1. Authenticate with Reddit: this involved getting the client\_id and client\_secret which was done with the help of ChatGPT to guide the developer on how to gain access to reddit as it is an open-source platform.
2. Get The Reddit post: a post was then decided on to analyse the comment section.
3. Extract the topic and all comments
4. Organize data into a dataframe
5. Save to a csv file if needed
6. Display first few comments
7. Data Cleaning Process:
8. Function to clean text
9. Apply function to each text

VIII. Write a function to get the sentiment

IX. Apply the sentiment Analysis

X. Inspect the results

XI. Classify the sentiments

## **2.2 METHODOLOGY USED**

The methodology used follows these key steps:

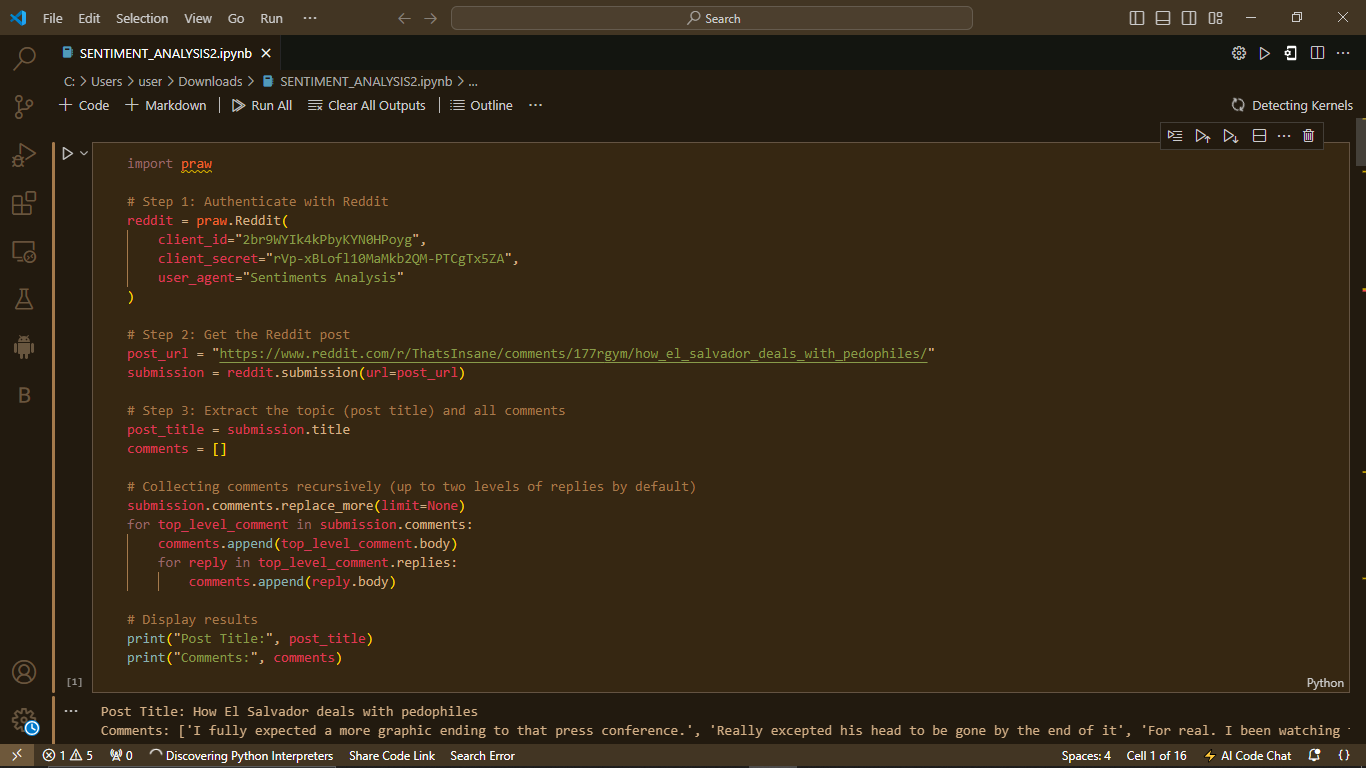
1. **Importing Libraries**: The code begins by importing necessary libraries:

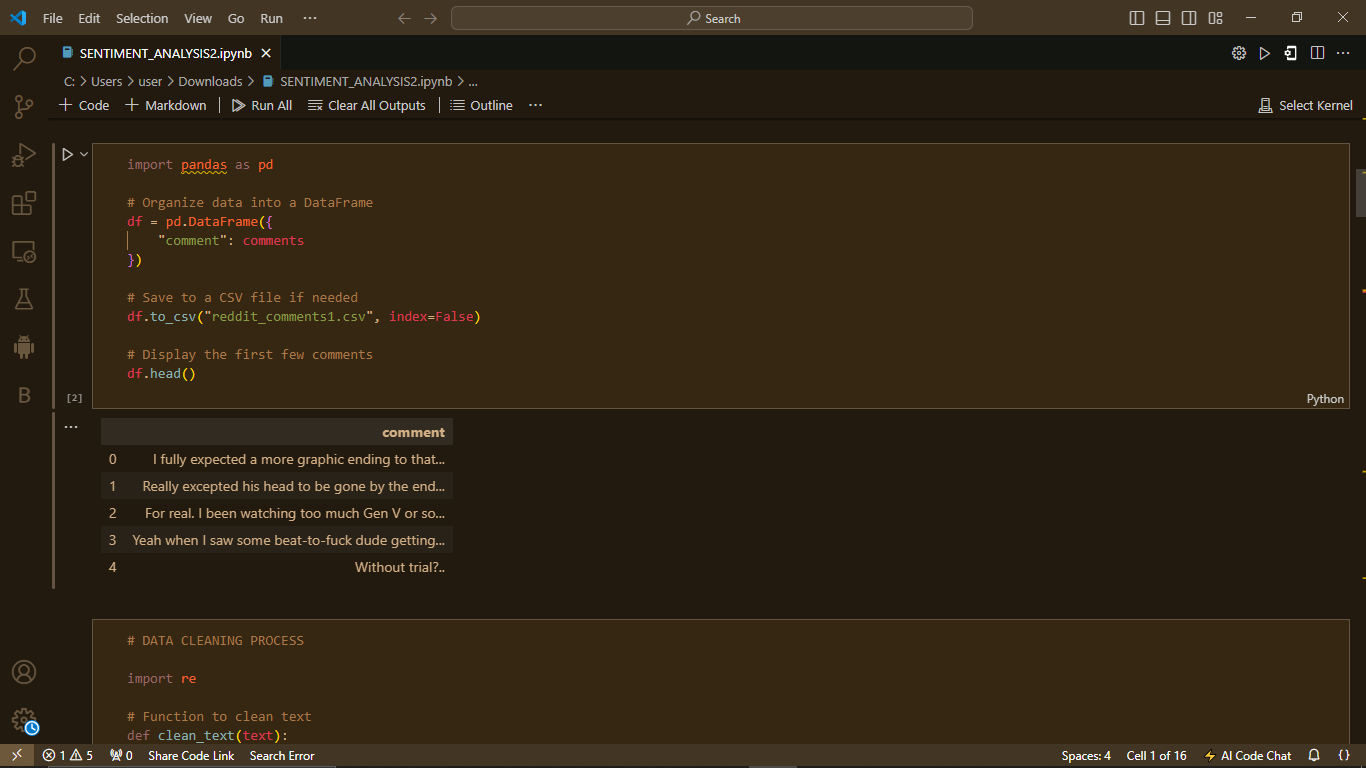
import praw

import pandas as pd

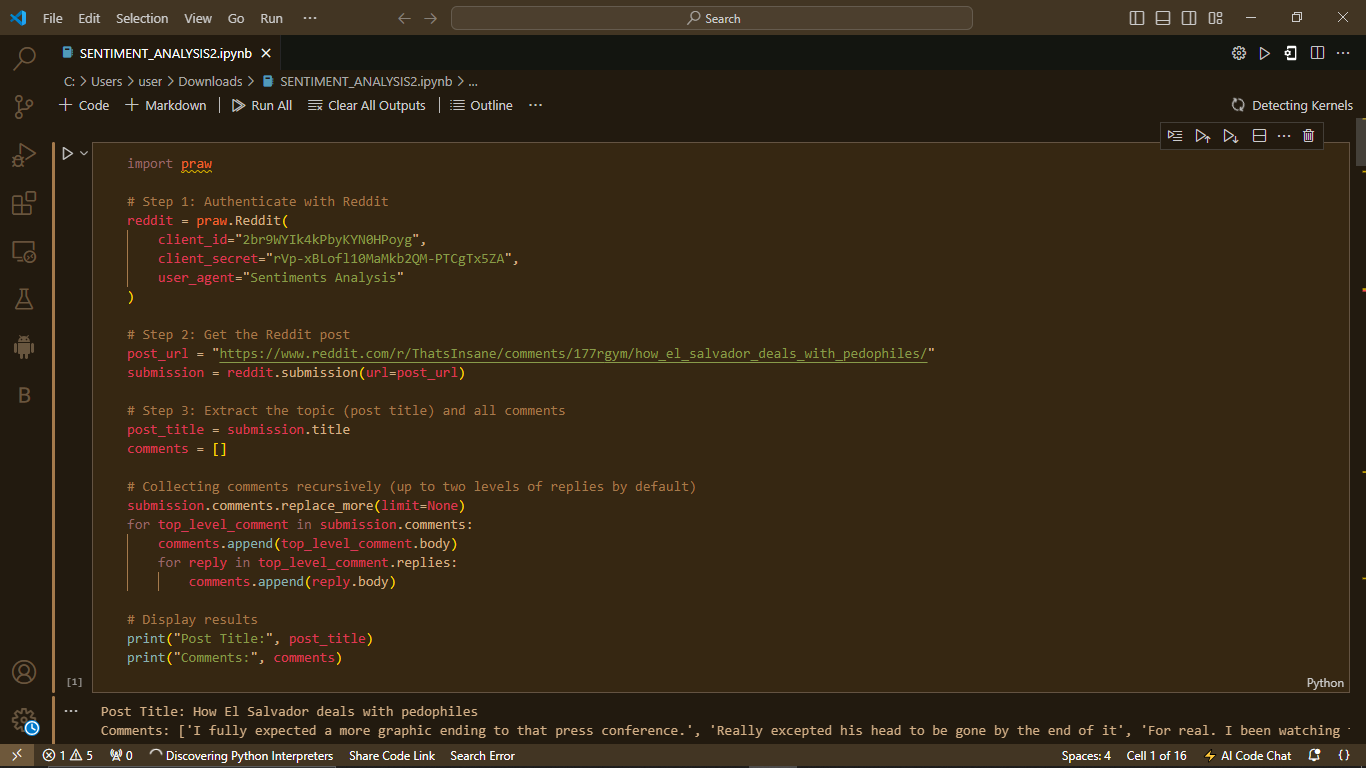
**praw (Python Reddit API Wrapper)**: This library is used to interact with Reddit's API to fetch posts and comments.

**pandas**: Used to organize extracted comments into a structured DataFrame for further processing.





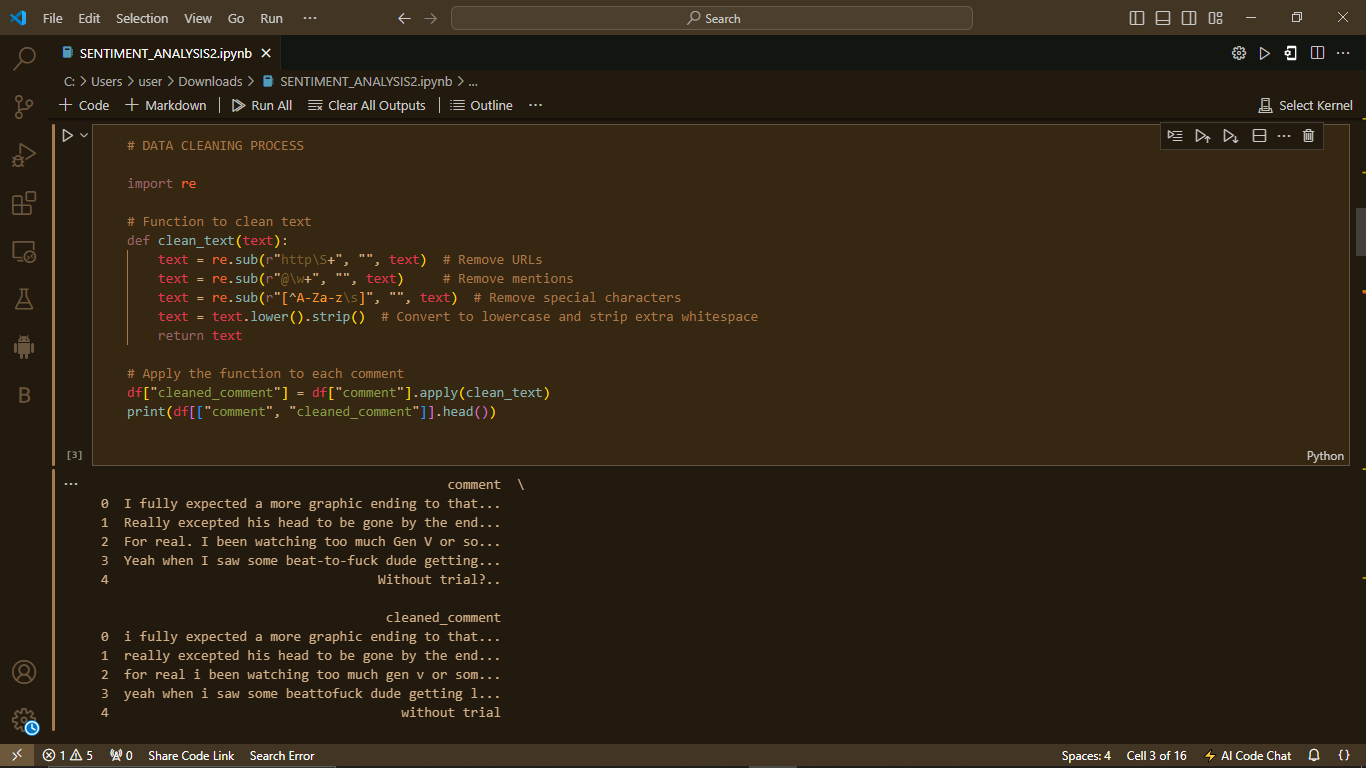
1. **Data Collection (Reddit Scraping)**
2. Uses praw (Python Reddit API Wrapper) to authenticate and extract comments from a specific Reddit post.
3. The script retrieves the post title and all associated comments, including nested replies.
4. Stores the extracted comments in a Pandas DataFrame and optionally saves them to a CSV file



3.  **Data Cleaning:** To clean the text regular expressions (re module) were used to :

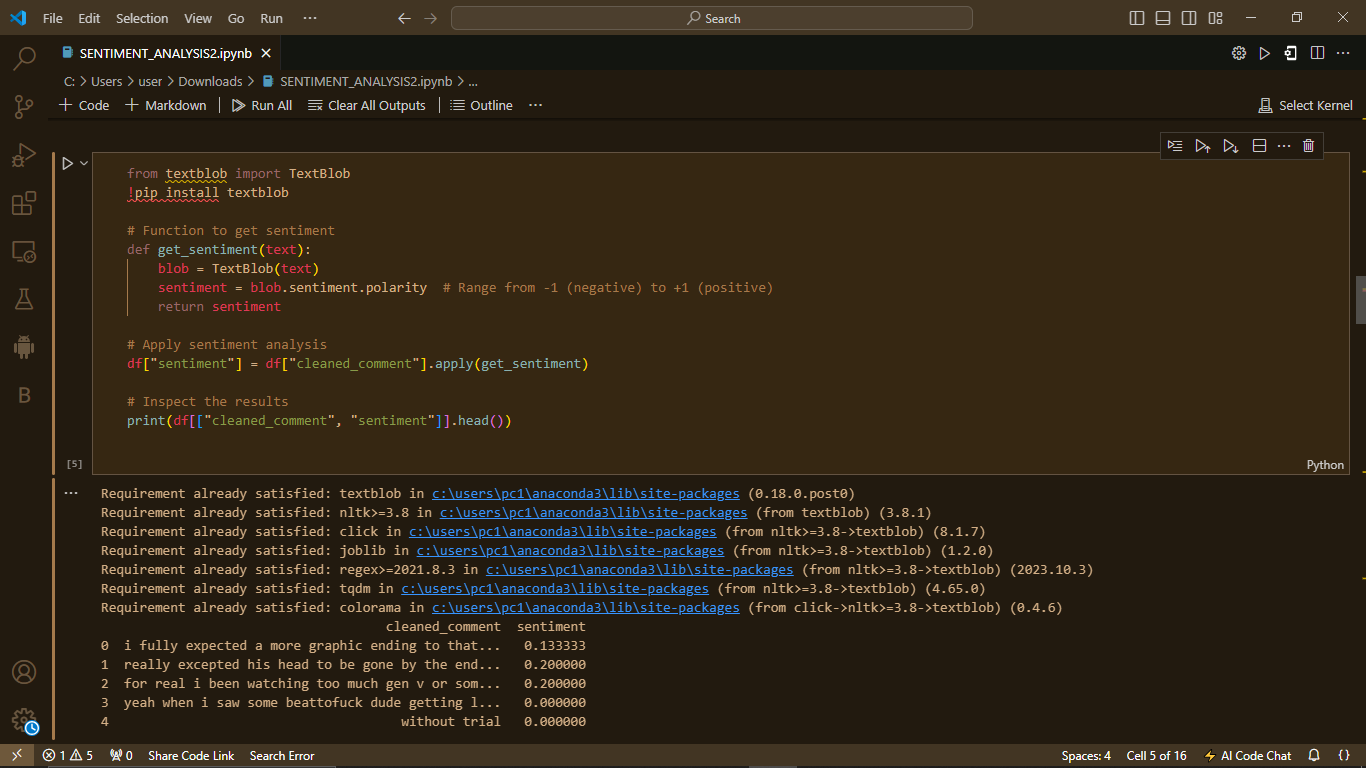
1. Removes URLs.
2. Removes mentions (@username).
3. Eliminates special characters, keeping only letters and spaces.
4. Converts text to lowercase and removes extra whitespace.

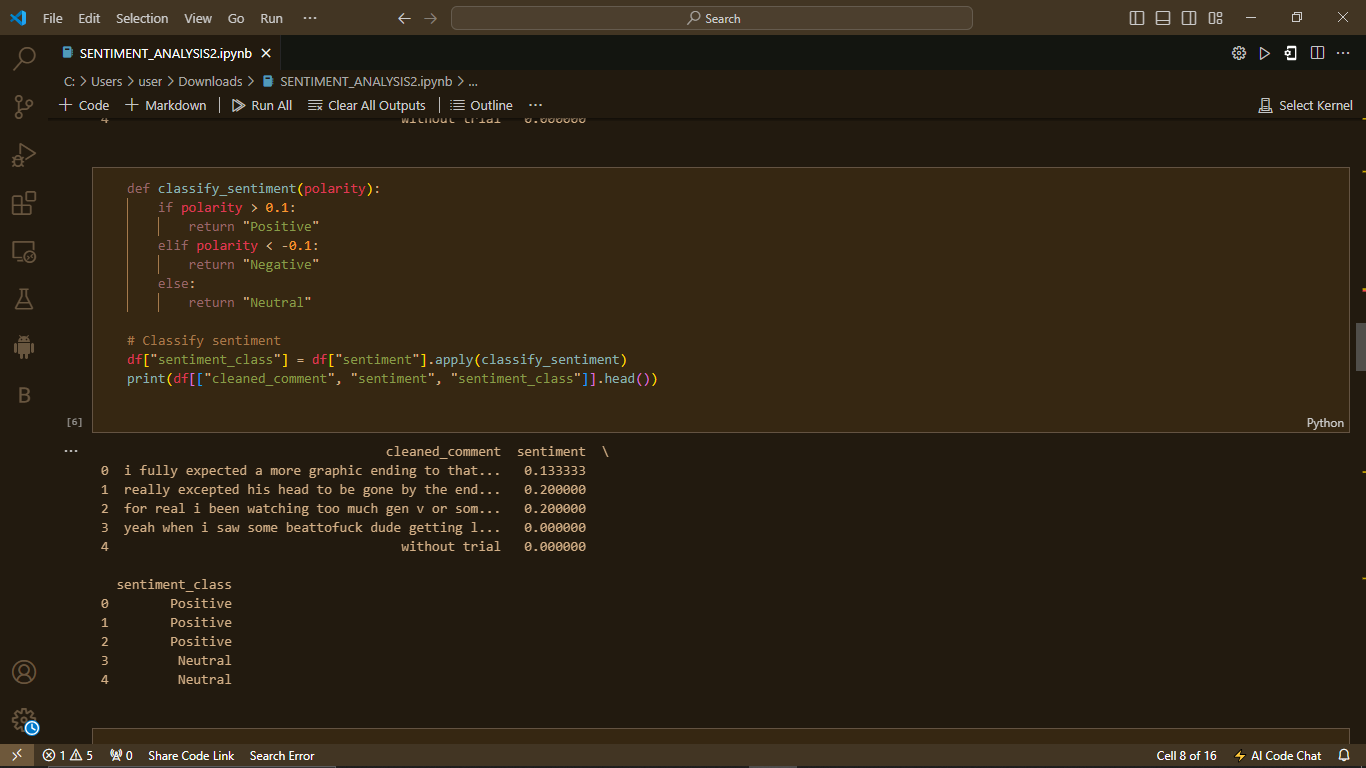
A new column (cleaned\_comment) is created in the DataFrame with the processed text.

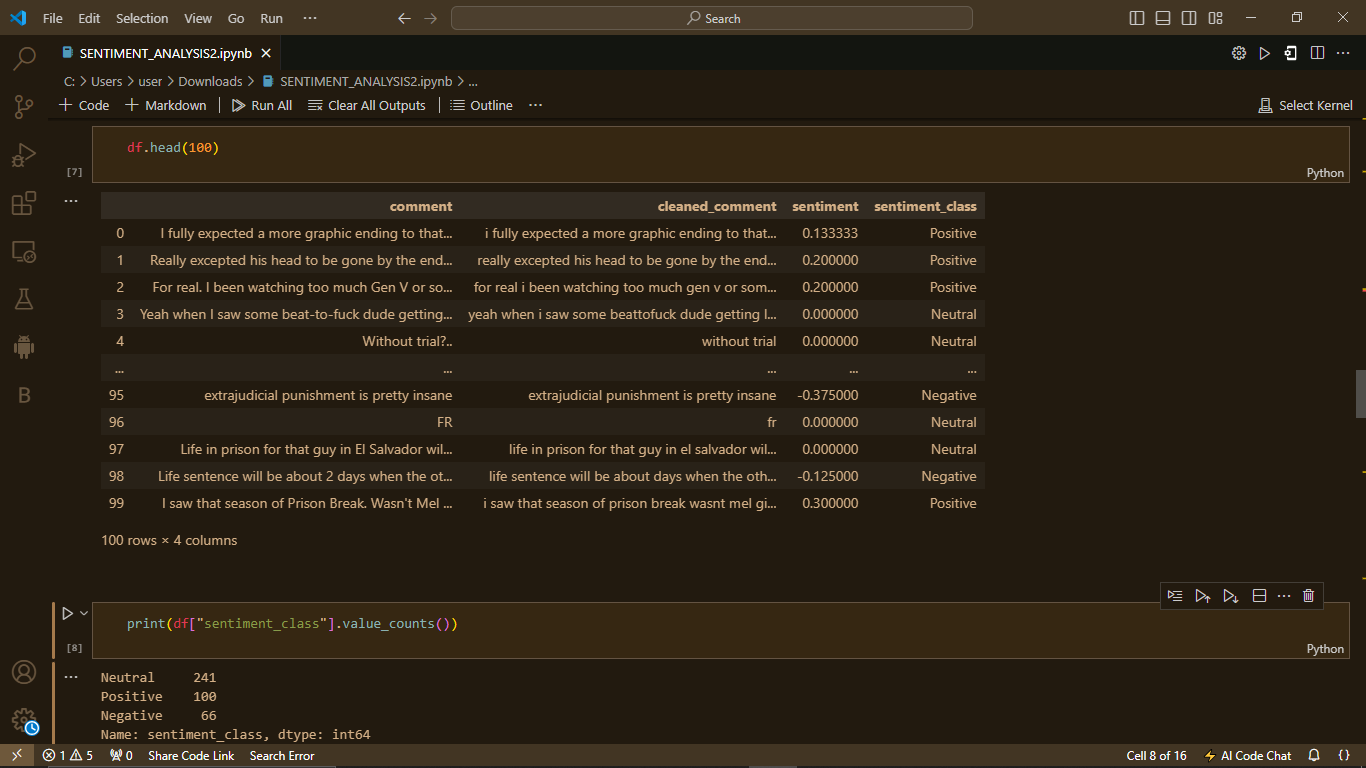


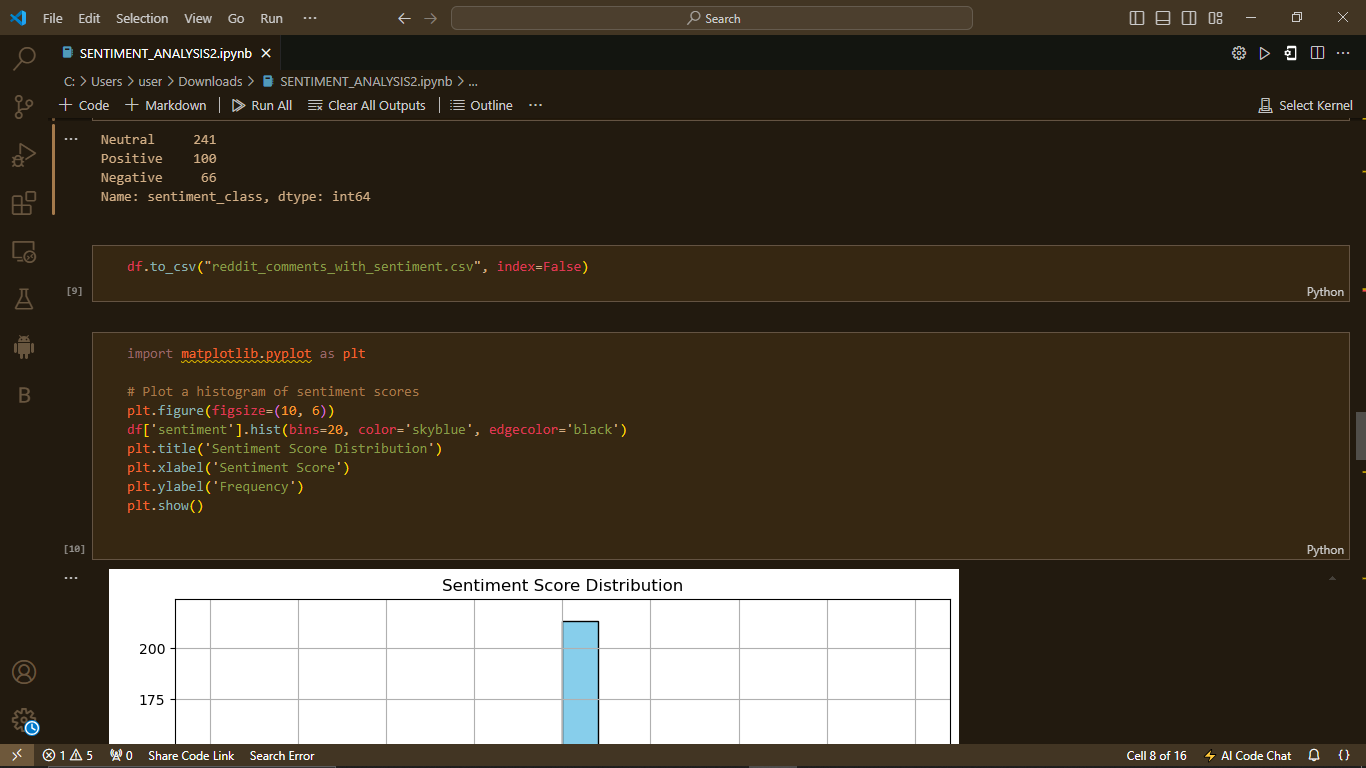
4. **Sentiment Analysis:** TextBlob was used for sentiment analysis.

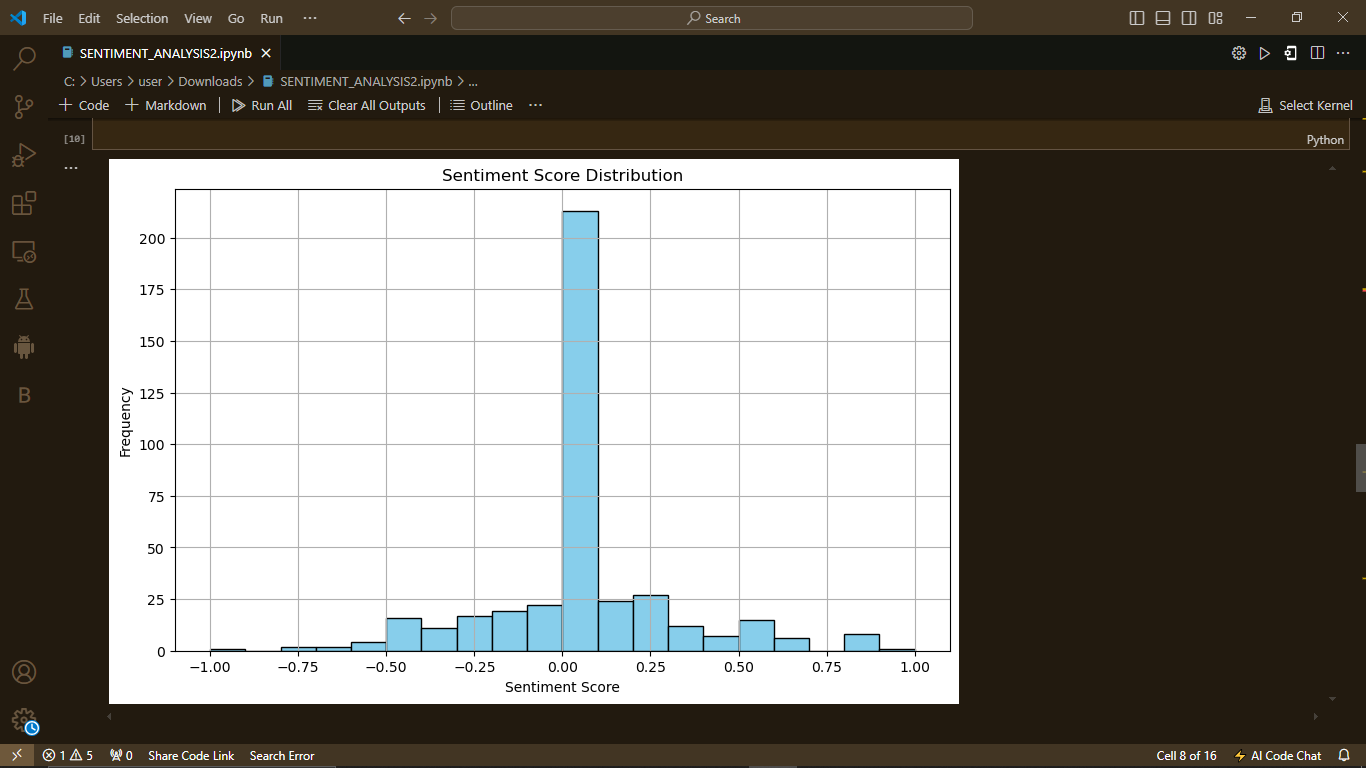
1. A function was defined to compute sentiment polarity (ranging from -1 to +1). The scale was used to categorize each comment.
2. This function was applied to the cleaned comments, storing the sentiment scores in a new column (sentiment).

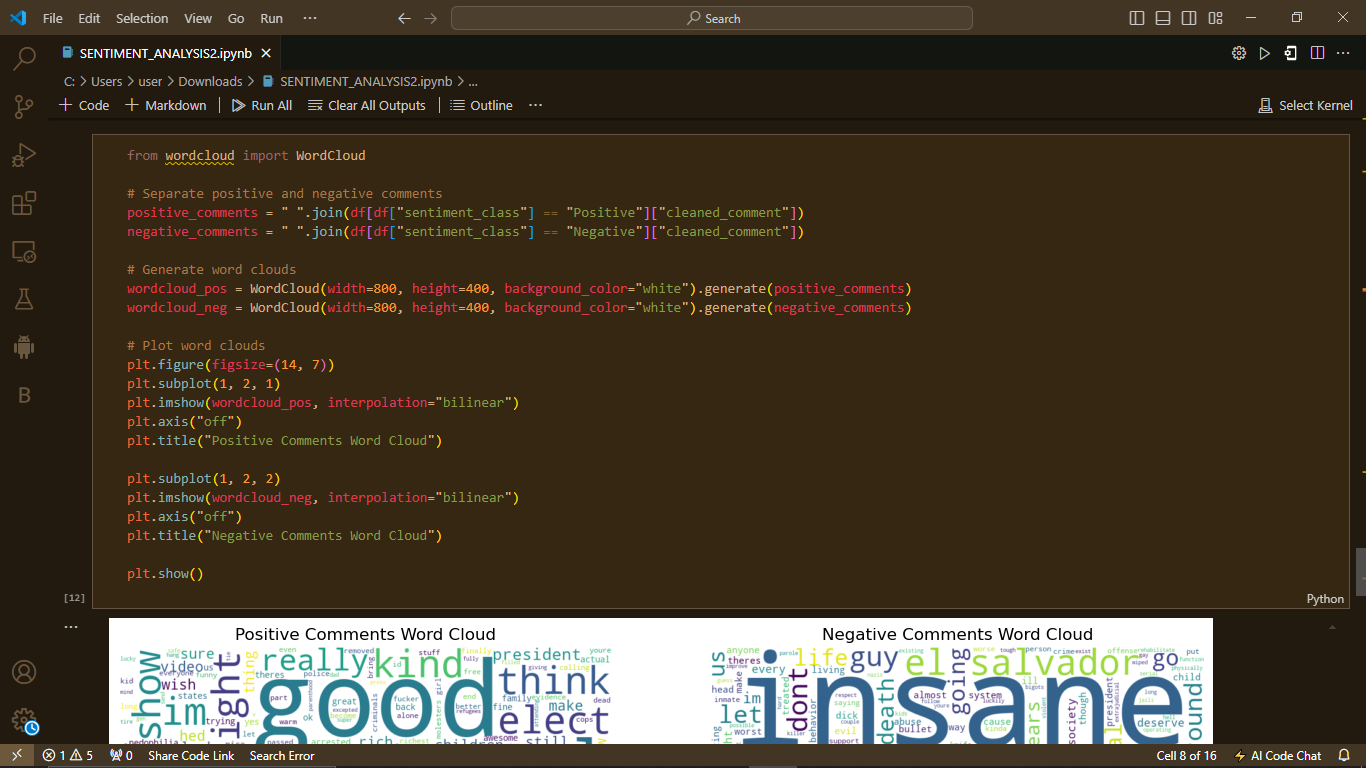


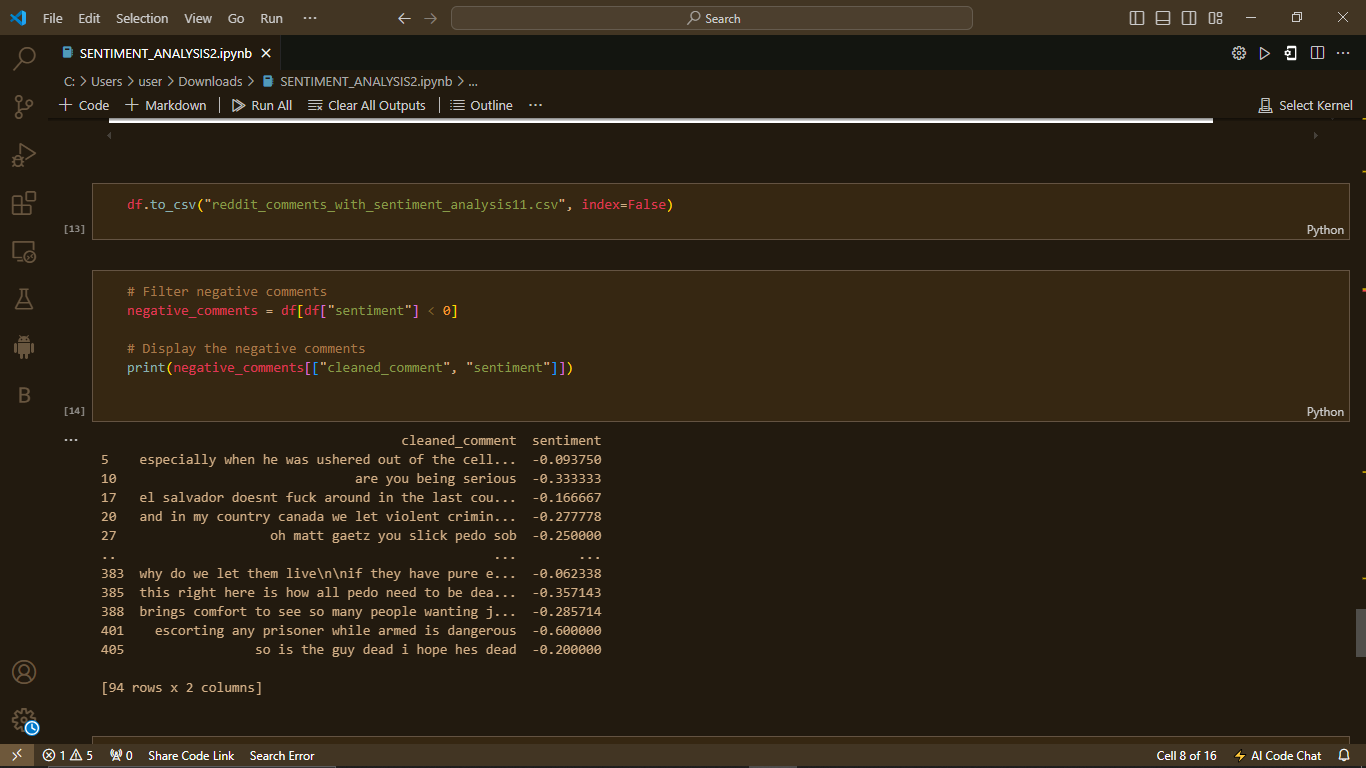


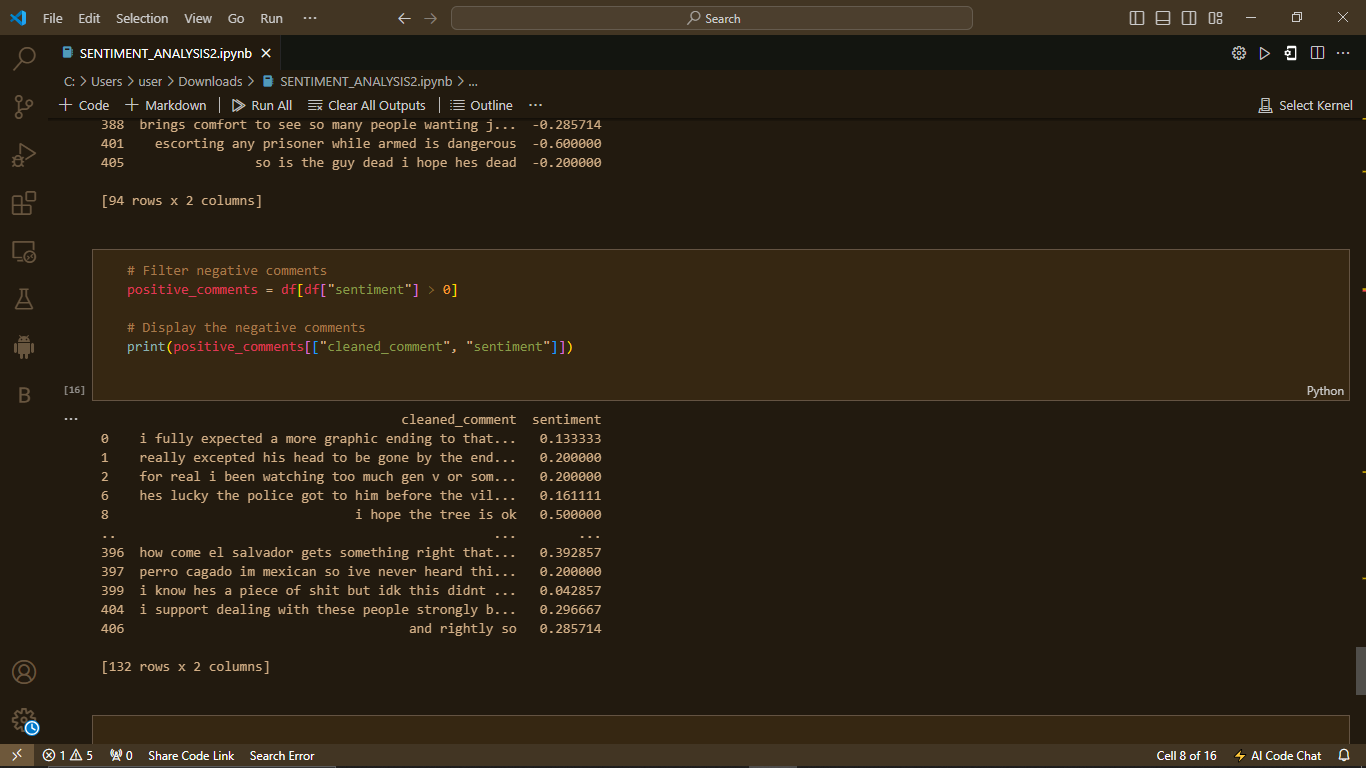












# **3.0 CHALLENGES FACED**

1. **Limitations of the Reddit API**: The Reddit API has rate limits on the number of requests in a given time frame, which might impact large-scale data extraction.
2. **Authentication Issues**: Bad API credentials or expired tokens might lead to failed access to Reddit's data.
3. **Getting nested comments:** this is pretty hard and could eventually introduce some recursion or extra API requests.
4. **Data Preparation:** Comments on Reddit may include spam, responses of bots, and other irrelevant texts that will need filtering.
5. **Ambiguous Sentiment**: Some comments may contain sarcasm, mixed sentiments, or ambiguous language that makes sentiment analysis difficult. Security Risks: Hardcoded API credentials are a security risk and must be handled via environment variables or configuration files.
6. **Internet Dependency:** The script requires an active internet connection; hence, it may fail to fetch the data due to API downtime or network failure.

# **4.0 CONCLUSION**

In this project, the script can fetch comments in a targeted post on the social network website: Reddit using a library called PRAW. Finally we made sure that when the information collected is given structure, sentiment analysis can follow easily. While the initial code block collects data, the other processes would further implement a Sentiment Analysis using TextBlob.