**CMPE 328 – Cloud Computing**

**Twitter Sentiment Analysis using Python**

**Final Project**

**Name:** Abdullah Yusuf and Gökay Özsoy

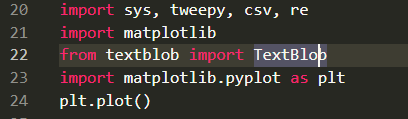
**Date**: 30th May, 2021

**Introduction**

In this project, we performed a Sentiment Analysis on Twitter. We used Python as the programming language paired with modules consisting of Tweepy, MatPlotLib and Textblob. We then integrated our program with AWS EC2 for our application. A docker container was also made for our program.

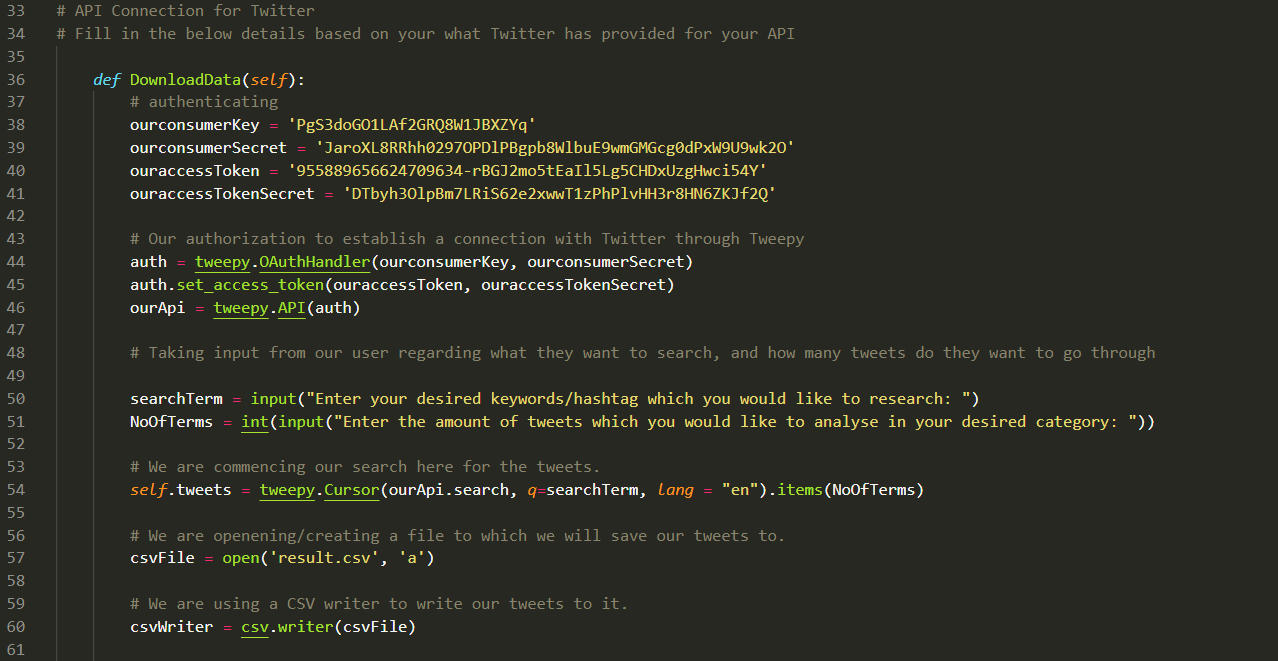
**Code Implementation**

We began our code with imports. We imported the libraries which are: TextBlob, Tweepy and Matplotlib. TextBlob is used for analyzing the tweets based on their sentiments (positive, negative or neutral). Tweepy is used solely for importing tweets by establishing a connection with our Twitter API. Lastly, MatPlotLib is used to plot our sentiment pie chart based on our tweets.



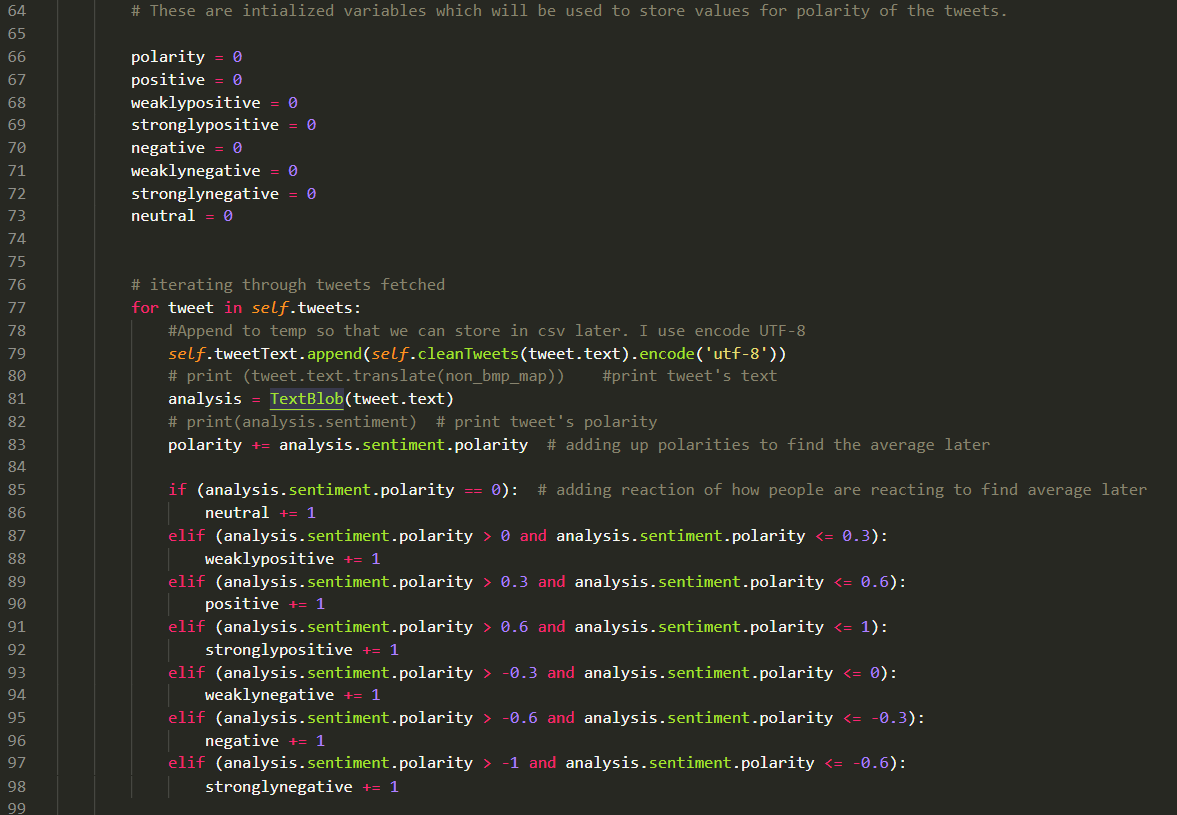
**Figure 1:** Library Imports for out Program.

In the function DownloadData, we are dealing with the authentication process. We are trying to establish a connection with Twitter through Tweepy using our Twitter API and its keys. Then we take user inputs for our desired keywords and the number of tweets that the user wants to search. We are using a CSV writer to write our tweets to the CSV file which we have just created under the name of ‘results’. The tweets which have been analyzed are shown in that file.



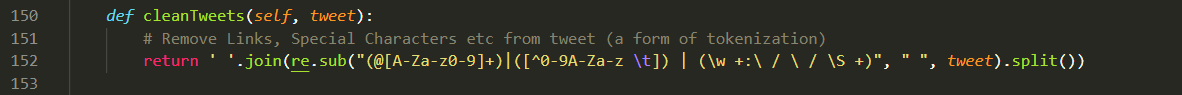
**Figure 2:** Our Authentication and Connection of Twitter’s API using Tweepy

There are initialized variables like polarity, positive, weakly positive etc. in our code. We are using them to store polarity values of the tweets. Then, just below our initialized variables, there is a for loop to iterate through the fetched tweets. Then, we find the averages of how the people reacted to our keyword on twitter.



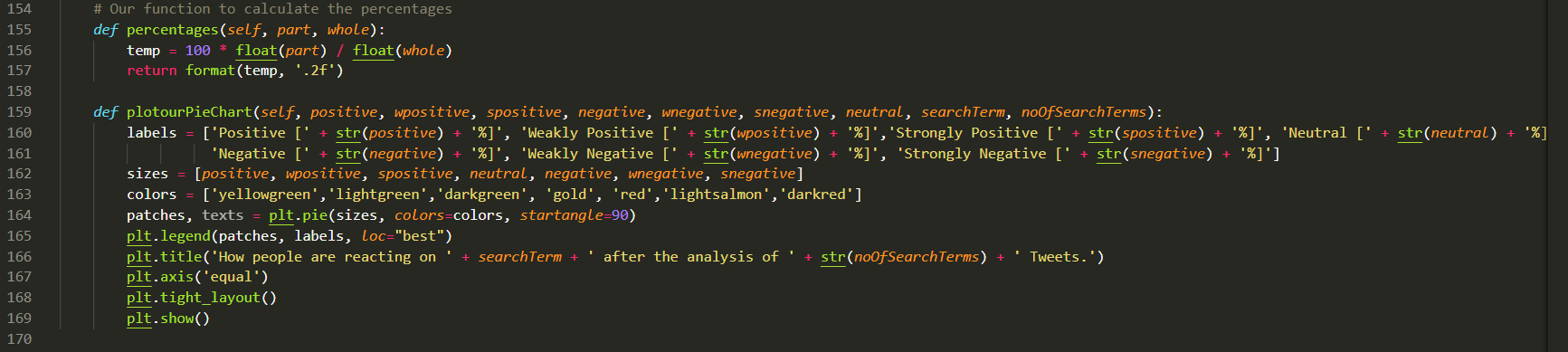
**Figure 3:** Initializing Polarity Variables regarding Tweets

We have a cleanTweets function for excluding links, special characters etc. from the tweet. Basically, it is used to tokenize our input in order to only look at the relevant words regarding our keyword(s).



**Figure 4:** Tokenization of Tweets

We have percentages function to calculate the percentages of reactions which are also then shown in the form of a pie chart alongside text.



**Figure 5:** Percentages and their display through a piechart.

**Containerization / Cloud Services**

We chose the AWS C2 service as a cloud service. Our code is now reachable from the cloud. We used Putty and FileZilla tools for connecting to cloud services. We have created private keys (.ppk) for authorization to enter the cloud platforms via Ubuntu. We tested our program and it runs from cloud without errors. We have tested our code with different keywords and also a variable number of tweets.