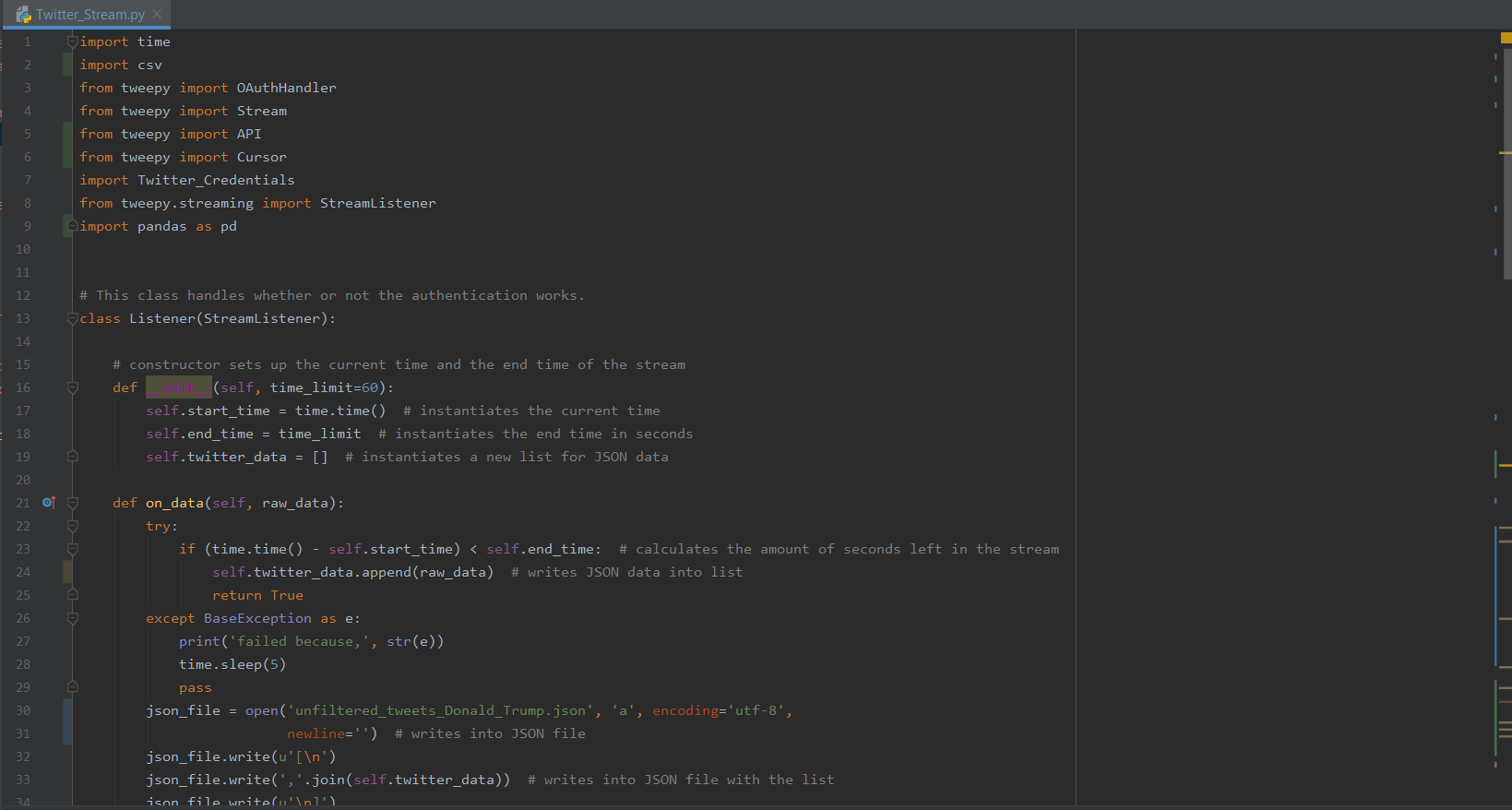
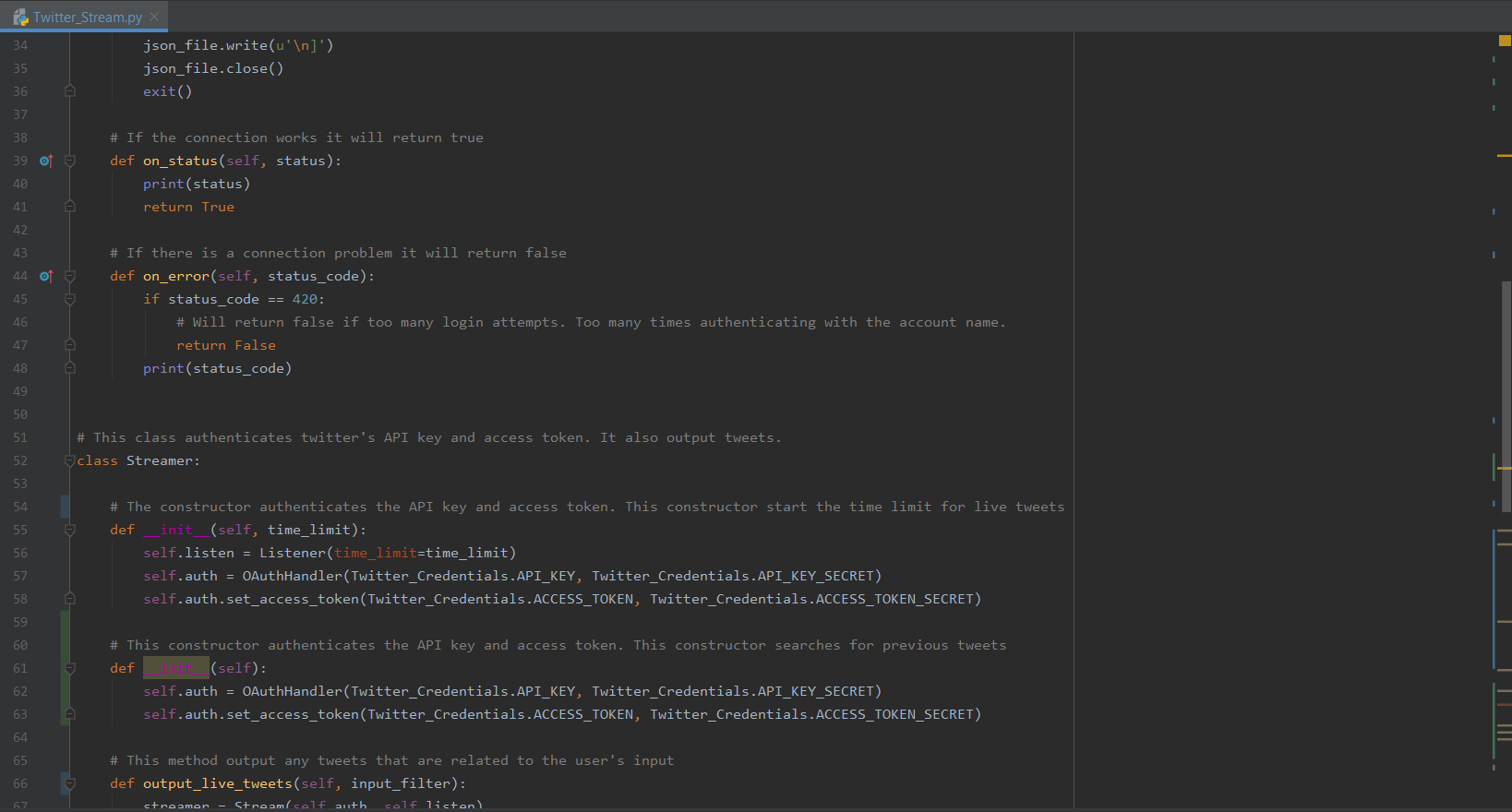
Kevin Lin

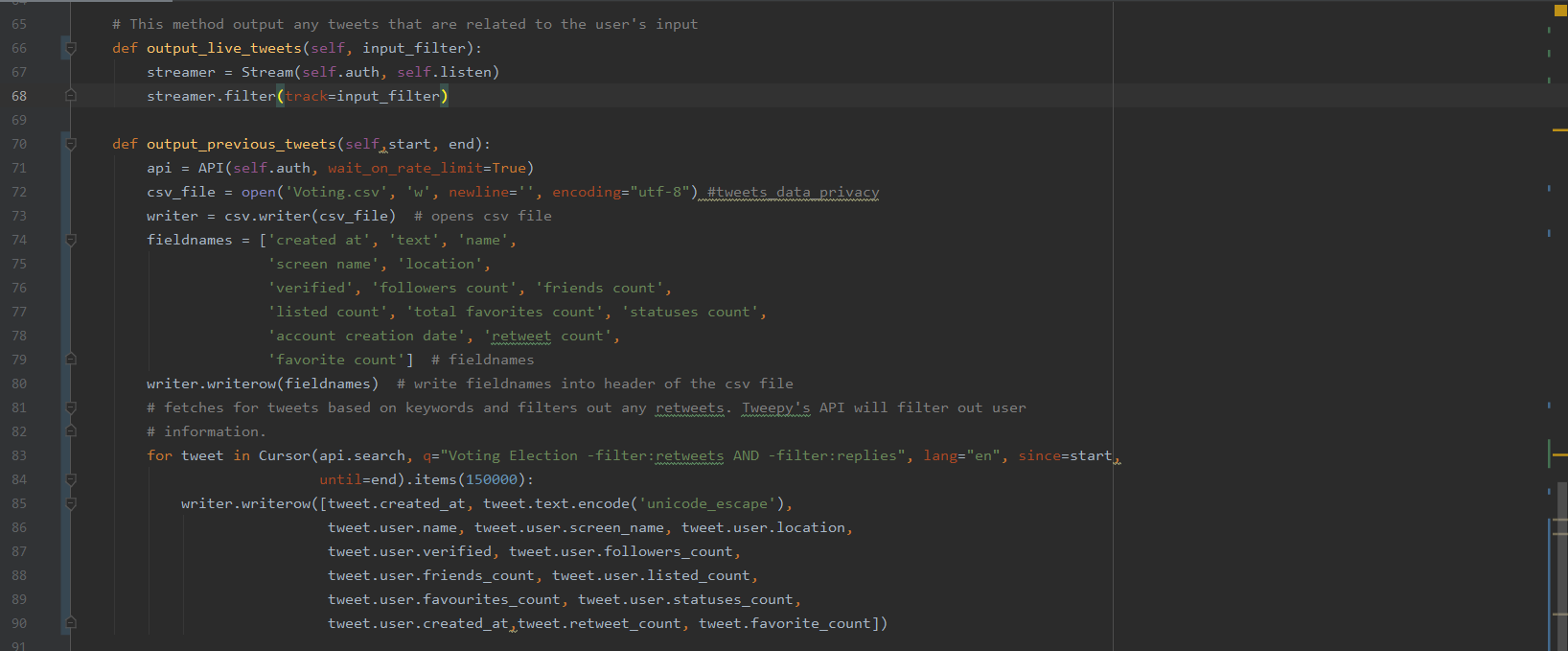
CISC 4900

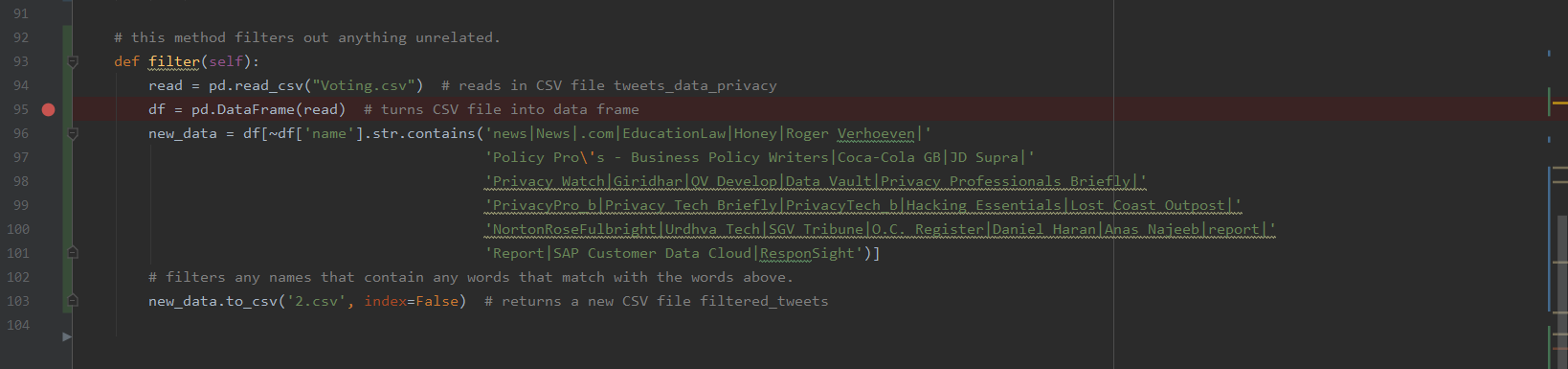
Project Summary

The topic of this project is about data analysis using the social media platform twitter. Using Twitter’s API and token keys, tweets were extracted from twitter. A python library called tweepy was used in this process to extract the tweets. The Listener class that was created to inherit the Tweepy’s Stream Listener class. This allowed the live extraction of tweets. A streamer class was created to verify the API and token keys. A constructor class was created for the time limit and an array was created for the tweets in JSON style. The on\_data method was created to fetch for live tweets. The on\_status and on\_error methods were created to check if there were live tweets fetched. Two constructors were created for the Streamer class One constructor was used for the live tweets, the other was used to get previous tweets. A filter method was created to filter out any unwanted previous tweets. All these methods and classes are in the Twitter\_Stream.py file.

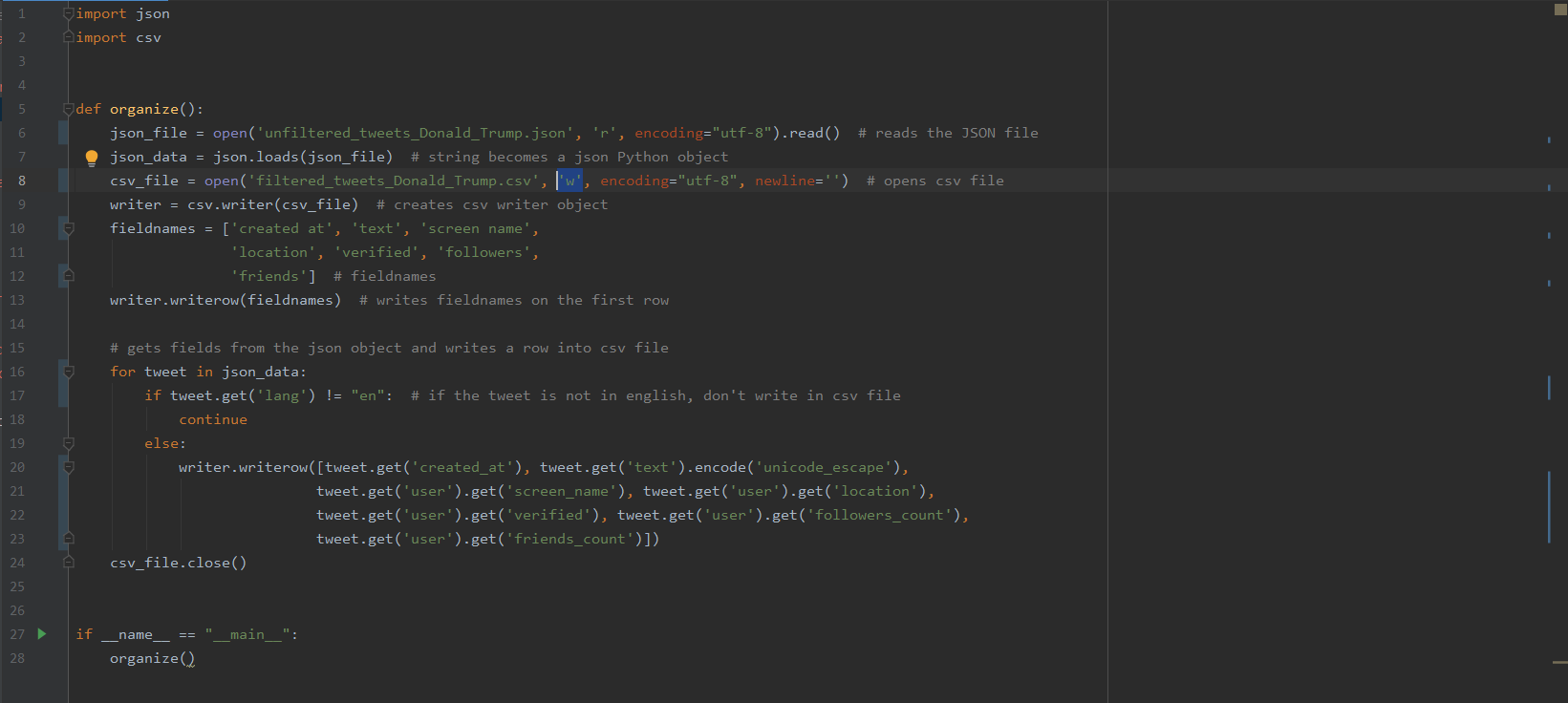






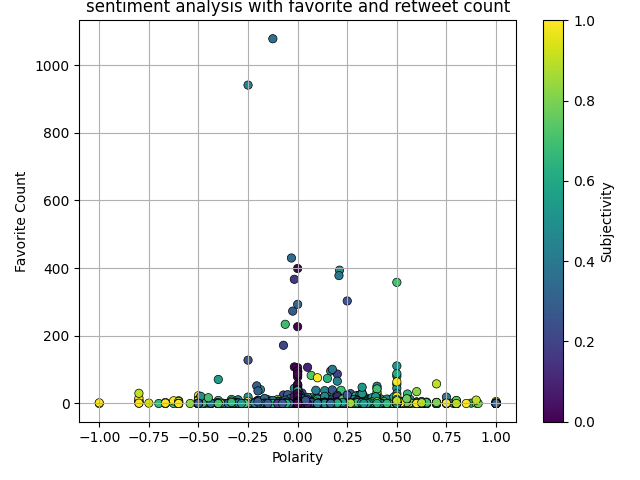


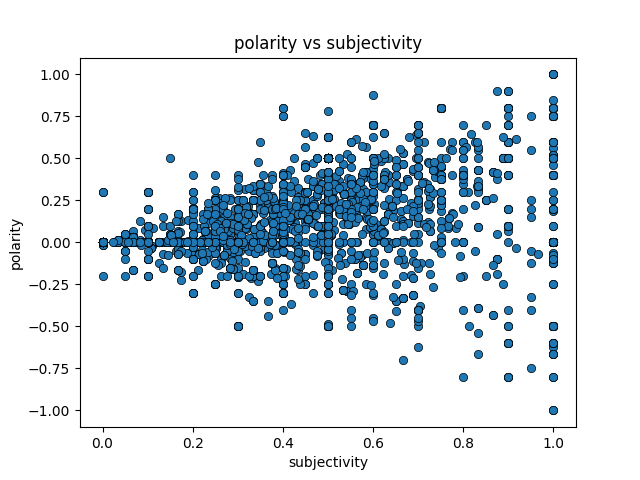
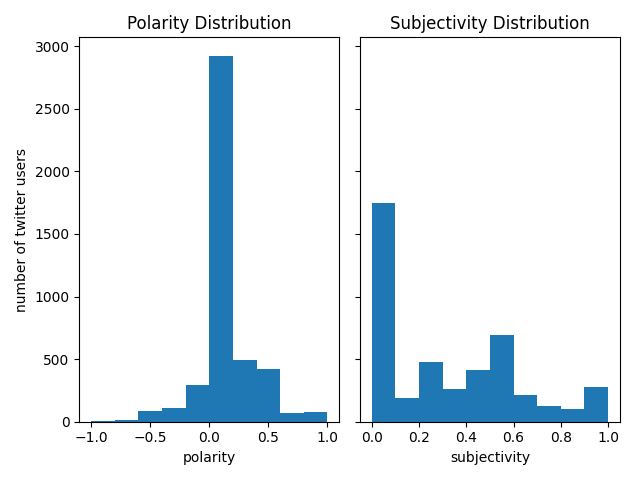
A JSON\_TO\_CSV.py and organize method was created to convert the live tweets from JSON to CSV files. The reason for making JSON\_TO\_CSV into another file is, during run time if you run output\_live\_tweets and organize methods in the same file, it will give a value error.



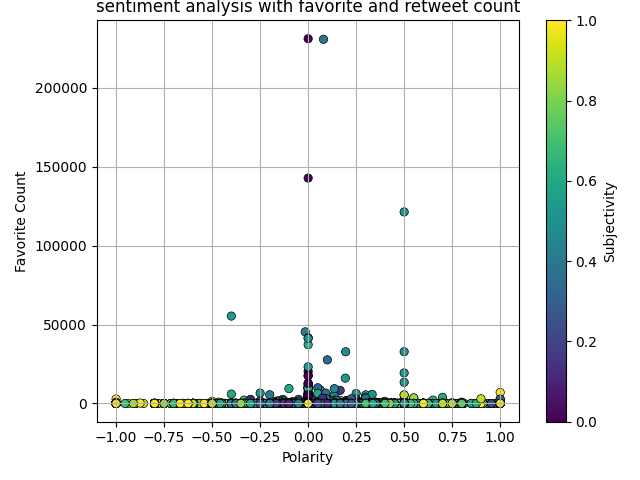
Next, we created a Sentiment\_analysis.py file with a class called Analysis. The first method is called polarity\_and\_subjectivity. The tweets that were extracted were used to detect the polarity and subjectivity of the text. Then, polarity and subjectivity are classified into three ranges and will create a new csv file. The next three methods help create a scatter or histogram plot. The subjectivity\_class method allows a subjectivity class column to classify the subjectivity into three classes(0,1,2) based on the three ranges. The to\_classification methods take the features, followers count, friends count, retweet count, favorite count, polarity, SubjectivityClass and make a new csv file. The new csv file will be used for a machine learning classification algorithm to determine whether we can predict subjectivity. Based on the data extract from twitter, there are six charts represented for the data. Three are represented for the topic, data privacy and the other three for the 2020 election voting fiasco.

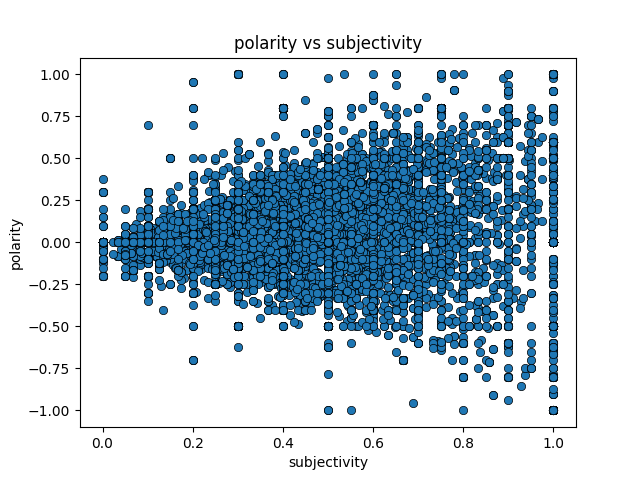
Data privacy:

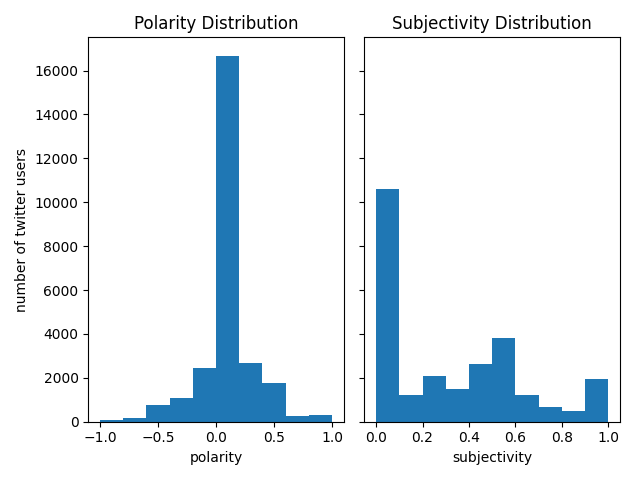




2020 Election:

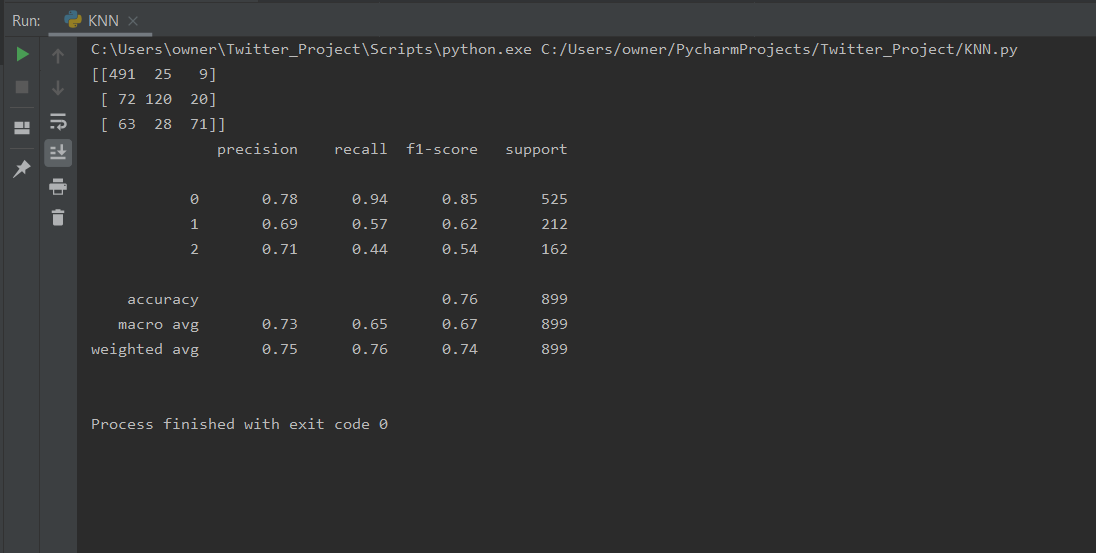




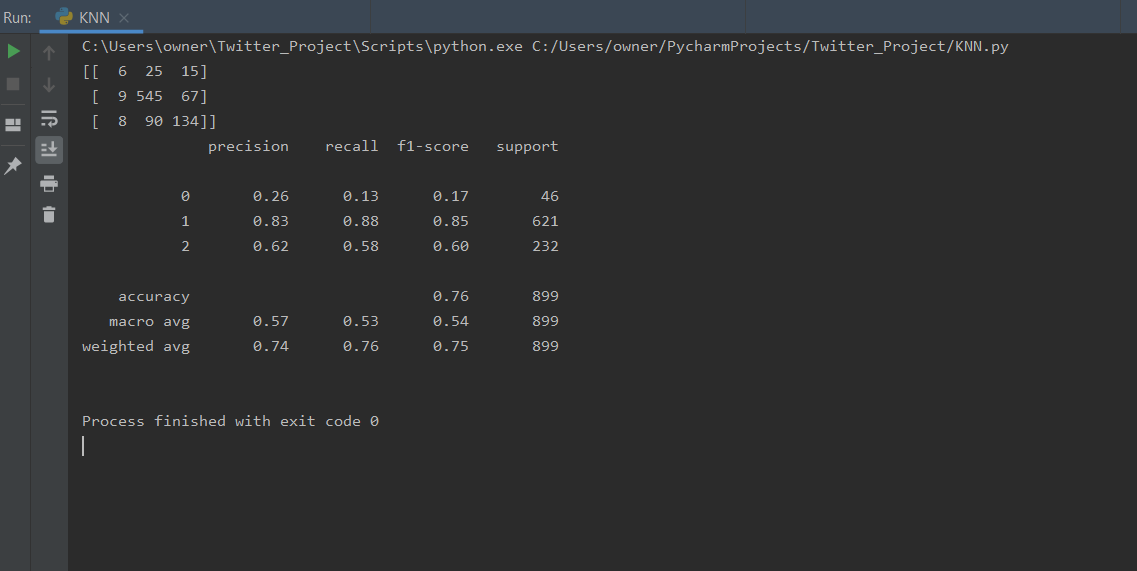


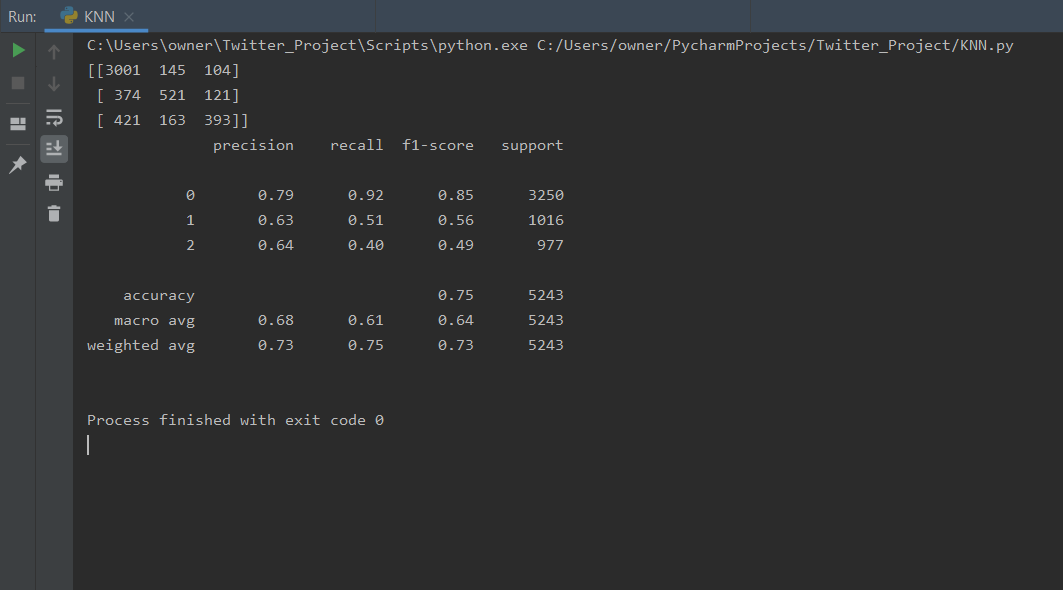
Next, the results for k nearest neighbor classification using five features to predict subjectivity or polarity.

Data privacy(subjectivity):

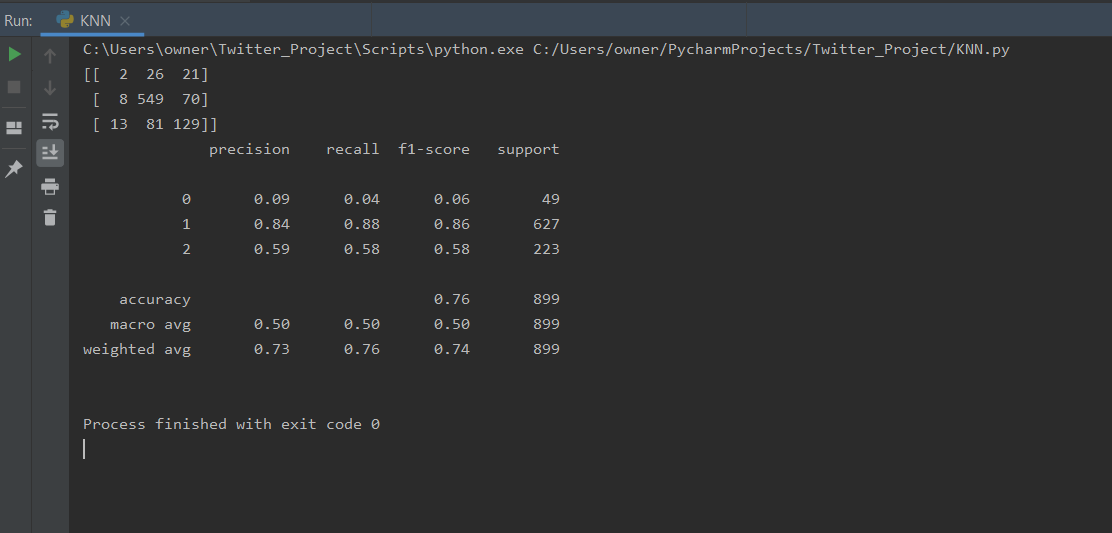


Data privacy(polarity):



2020 Voting Election(subjectivity):

2020 Voting Election(polarity):



Based on the results above, we can conclude that it is possible to predict polarity and subjectivity. Since, the accuracy is 75 to 76 for both polarity and subjectivity.