**HW3-PeiCheng-Huang**

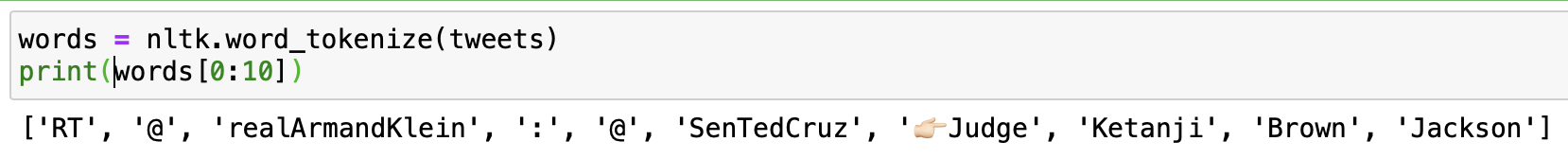
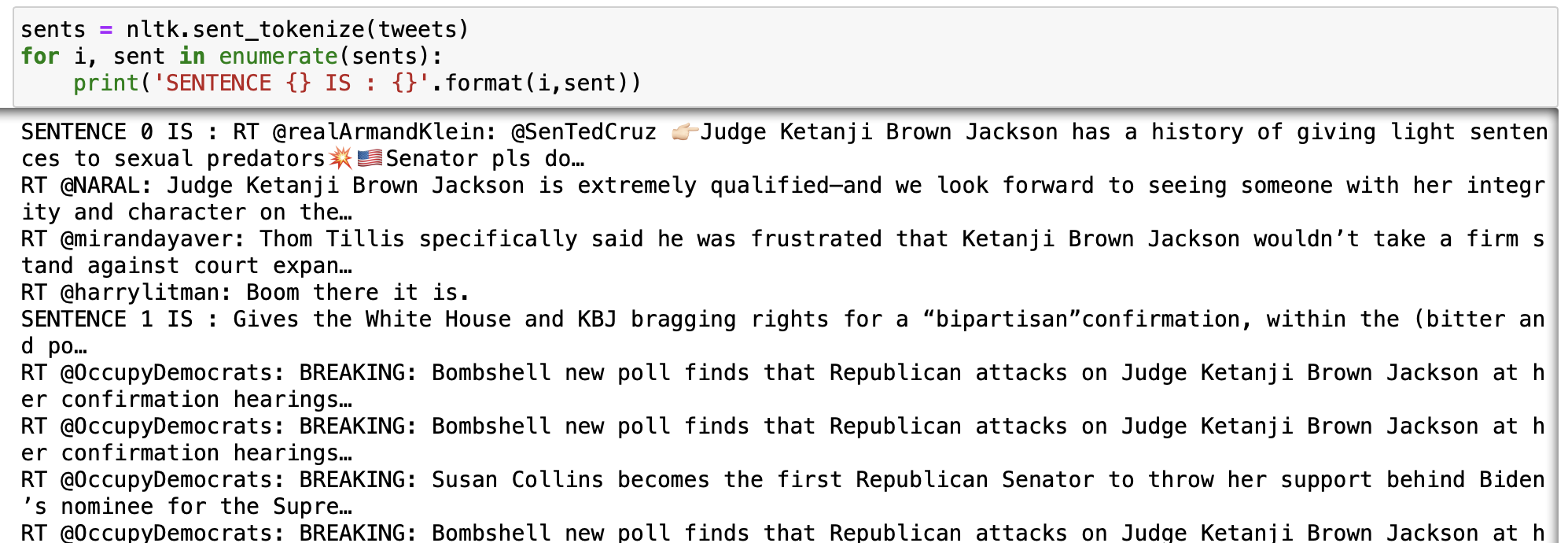
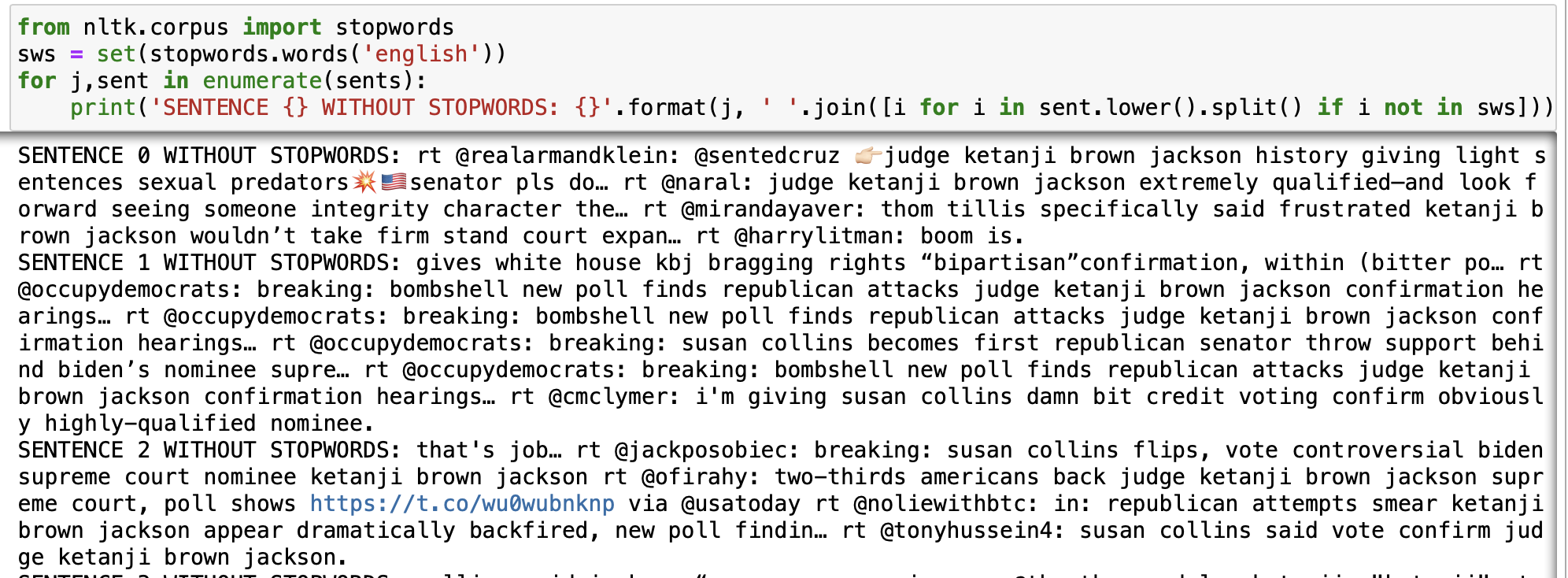
**Introduction**

In this paper, I analyzed tweets about the latest Supreme Court nominee by doing sentiment analysis on Twitter. I conducted sentiment analysis on these tweets using the libraries, NLTK and Vader, after downloading some recent tweets by using the Twitter API.

**Initial Step**

First I applied a developer account on Twitter and created a new project and new application to get Bearer Token for downloading tweets. Then ran the provided script with “ketanji” as a parameter to query the tweets we want to analyze and stored as tweets.txt file.

**Downloading and parsing the tweets**

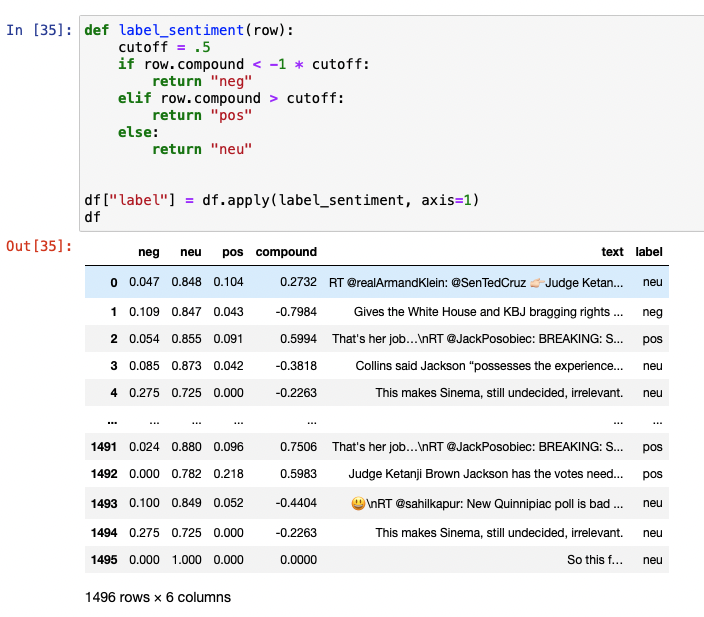
1. Download:   
   Execute “ python3 print\_tweets.py “ketanji” ” to download related tweets
2. Word Tokenization:  
   
3. Sentence Tokenization:  
   
4. Stopword Removal:  
   

**Sentiment Analysis using an existing library**

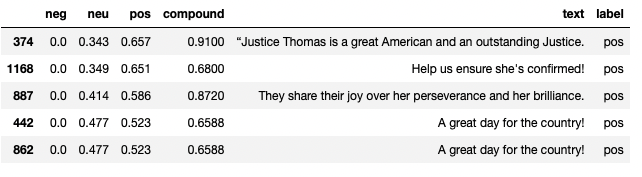
1. Perform sentiment analysis with nltk library and show the first 10 tweets

* neu: neutral sentiment score
* pos: positive sentiment score
* neg: negative sentiment score
* compound: a compound score representing all three scores  
    
  

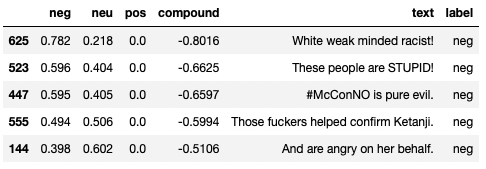
2. Display the tweets and label their sentiments.



3. Show examples of tweets:

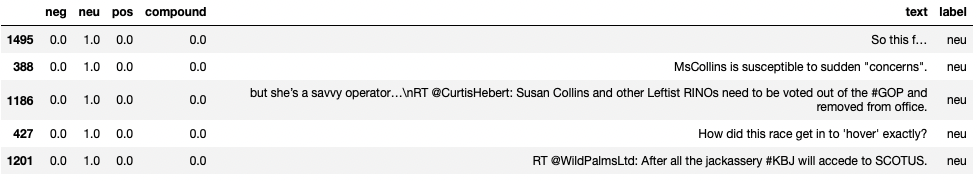


Top-5 positive tweets & scores



Top-5 negative tweets & scores

Top-5 neutral tweets & scores

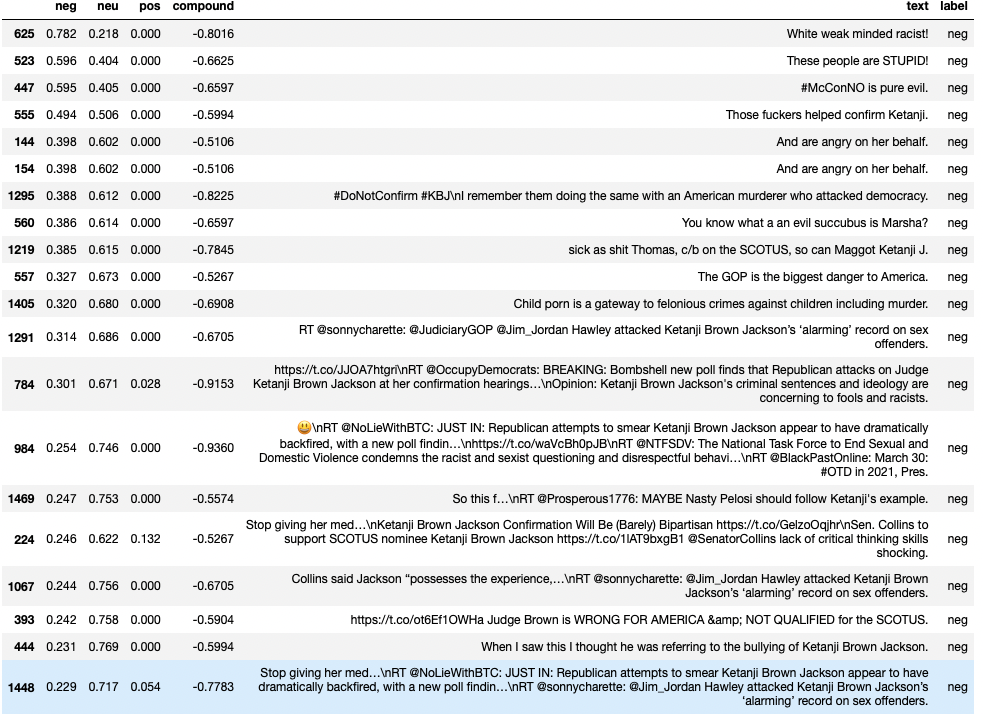
****

**Discussion and Comparison of results**

1. Accuracy of sentiment analysis on short texts & sentences

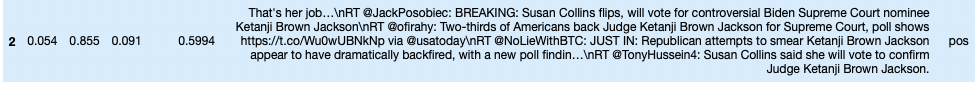
The accuracy of sentiment analysis on short texts is really high since short texts usually don’t have a complicated sentence structure. They are straightforward, and we can analyze the sentiment of the text by small numbers of keywords annotated with a sentiment value. As we can see, the top-5 negative and positive texts are all short and simple with some explicit sentiment words and objects. However, if we look at the last 5 of top-5 negative texts, some of these texts are the complaint about what happened to the nominee, Ketanji Jackson is unfair, and some of the texts are criticizing the Republicans, which should be the opposite of Ketanji. These are examples that the specific sentiment words are not that reliable in long sentences because the object of the sentence might not be the keyword we queried.

Top-20 negative tweets & scores



2. Take 2 examples of tweets where neutral score is higher than positive or negative. Explain why it is so.







I printed 10 tweets that have neutral score higher than positive or negative and took No.2 and No.5 as examples.

No.2 is a news stating a fact and it does not contain too much emotion, so it is neutral sentiment

No.5 is just a question, so that we cannot judge it as positive or negative sentiment