

In this HW, I created two sets of data. One is the previous tweets and the other one is the streamed tweets. All the tweets in these files are raw tweets with the exact format that I received from tweeter. For both of these samples, first I ignored the retweets by adding “-filter:retweets” then I checked the number of words that are used in each tweet by monitoring the spaces “ ” in the tweets. I extracted the 'id','user id', ' created_at','favorite_count','retweet_count','place',and 'text' of each tweet and stored it in a dataframe and created a csv file for further hws.