EXTRACTING TWEETS

```
import tweepy
import csv
import time
import json
####input your credentials here
consumer_key = 'zh1eufheH7ls8JYTHHBHwholb'
consumer_secret = 'M7antDK28Wyk9rMSEs4aXI8Ztsm5SQT81TVJYSn4TOCRc20dKe'
access_token = '1000022506437337089-qX2DCcWgr31moaHcrbNgsD3gPWVHIH'
access_token_secret = 'ksysbz8r51bWseuVEvcfG6Kb29pSGmtx9jnG17mFWFPFn'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

def get_tweets(query):
    api= tweepy.API(auth)
    try:
        tweets = api.search(query)
    except tweepy.error.TweepError as e:
        tweets = [json.loads(e.response.text)]
    return tweets
tw = get_tweets("@realDonaldTrump")



with open('tweetTrump_2.csv', 'w') as outfile:
    writer = csv.writer(outfile)
    writer.writerow(['id', 'user', 'created_at', 'text'])
    for tweet in tweepy.Cursor(api.search, q="@realDonaldTrump", count=100,
```

```
                lang="en",
                since="2018-06-06").items():
        print(tweet.created_at, tweet.text, tweet.id_str, tweet.user.screen_name)
        writer.writerow([ tweet.id_str, tweet.user.screen_name, tweet.created_at,
tweet.text.encode('utf-8')])
```

PREPROCESS:

```
import re
import csv

def processTweet(tweet):
tweet = tweet.lower()
 tweet = re.sub('((www\.[^\s]+)|(https?://[^\s]+))','URL',tweet)
 tweet = re.sub('@[^\s]+',' ',tweet)
 #Remove additional white spaces
  tweet = re.sub('[\s]+', ' ', tweet)
  # Remove b'RT
  tweet = re.sub(r'b\'RT', "" ,tweet)
   tweet = re.sub(r'b\'rt', "" ,tweet)
   tweet = re.sub(r'\n\'',"" ,tweet)
   tweet = re.sub(r'"b"""', "", tweet)
   tweet = re.sub(r'\xf0\x9f\x98\x82', " ", tweet)
   #Replace #word with word
    tweet = re.sub(r'#([^\s]+)', r'\1', tweet)
    #trim
    tweet = tweet.strip('\'"')
    return tweet

   fp = open('tweetTrump_2.csv', 'r')
```

```
line = fp.readline()


while line:
    refined = processTweet(line)
    print (refined)
    line = fp.readline()
fp.close()
```

SENTIMENT ANALYSIS:

```
import csv
import re
from nltk.corpus import stopwords
def newLine(line):
    nL = line.split("\n")
    return nL
def words(cols):
    w_ids = cols[4].split(" ")
    words = [w.split("#")[0] for w in w_ids]
    return words
with open('tweetTrump_2.csv', 'r') as csv_file:
    csv_r = csv_r(csv_file)
    for word in csv_reader:
        print(word)
def score(filename, s_word):
    f = open(filename)
    totalobject = 0.0
    count = 0.0
    t_positive = 0.0
    t_negative = 0.0
```

```
totalneutral = 0.0
for line in f:
    if not line.startswith("#"):
        cols = split_line(line)
        words = get_words(cols)

        for word in s_word:
            if word in words:
                if word == "not":
                    totalobject = totalobject + 0
                    t_positive = t_positive + 0
                    totalneutral = totalneutral + 0
                    t_negative = t_negative + 16
                    count = count + 1
                else:

                    totalobject = totalobject + get_objective(cols)
                    t_positive = t_positive + float(get_positive(cols))
                    t_negative = t_negative + float(get_negative(cols))
                    totalneutral = totalneutral + float(get_neutral())
                    count = count + 1
if count > 0:
    if t_positive > t_negative:
        print("Positive value : ", t_positive)
    elif t_negative> t_positive:
         print("Negative value : ", t_negative)
    else:
        print("Neutral: 0", totalneutral)
        print("average object Score : ", totalobject / count)
```

```
if __name__ == "__main__":
    s_word = ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t]) |(\w+:\/\/\S+)", " ", comment).split())
    stop_words = set(stopwords.words('english'))

    s_word = s_word.lower().split(" ")
    refined = [w for w in s_word if not w in stop_words]

    score("SentiWordNet_3.0.txt", refined)
```