

# Report on Twitter sentiment analysis assignment

Manas Jayant Bondale 989368417

## Environment:

Place all .py .txt .json and .md files in the same folder.

## Data files

AFINN-111.txt : contains sentiment scores for words

data.json : contains test tweet data

us\_states.json : us state coordinates and abbreviations

tweets.json : downloaded tweet data from scraper.py

## Source files

frequency.py

Desc : outputs the frequencies of all words in all tweets

Run : python frequency.py data.json

happiest\_state.py

Desc : outputs the state that has the highest sentiment score

Run : python happiest\_state.py AFINN-111.txt data.json

scraper.py

Desc : download USA public tweet data and dumps to tweets.json

Run : python scraper.py

term\_sentiment.py

Desc : calculates sentiment scores for words not in AFINN-111.txt

Run : python term\_sentiment.py AFINN-111.txt data.json

tweet\_sentiment.py

Desc : calculates the total sentiment score of each tweet

Run : python term\_sentiment.py AFINN-111.txt data.json

top\_ten.py

Desc : finds the top ten most used hashtags

Run : python top\_ten.py data.json

## Output files

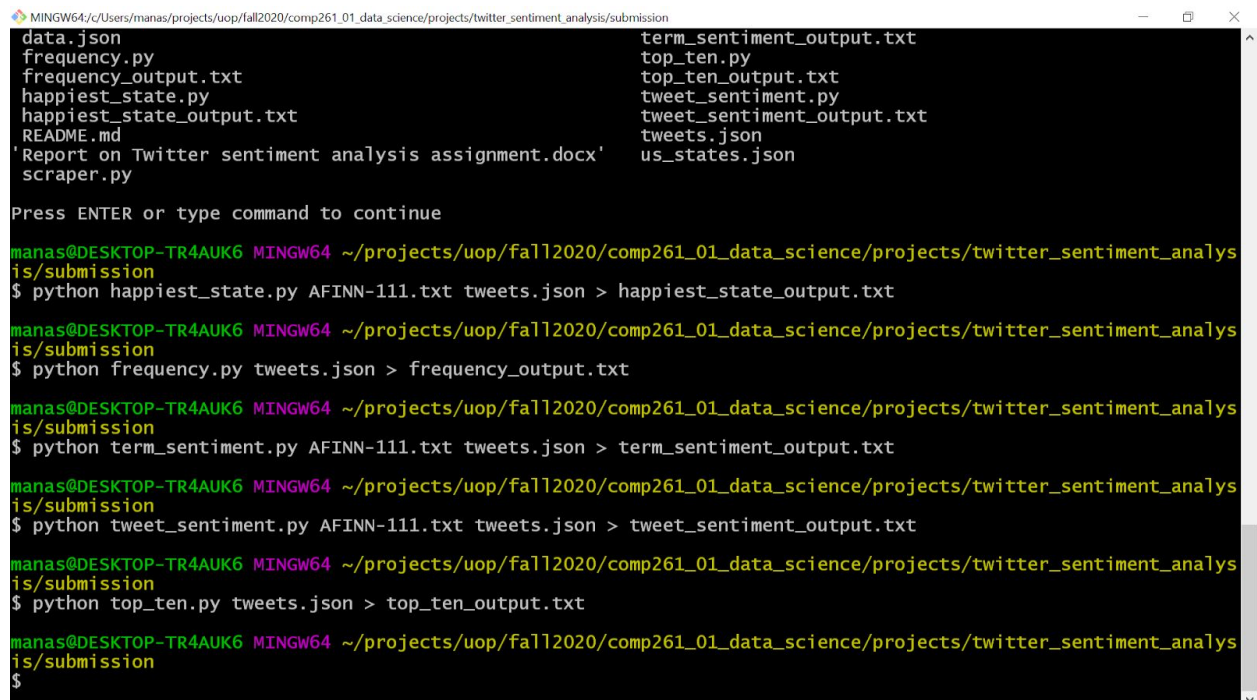
The output files of each program (other than scraper.py) is named \*\_output.txt where \* is the name of the python file before '.py'

For scraper.py the output file is tweets.json

## Output screenshots:

Following screenshots show the command prompt commands used to run the programs and part or whole of their outputs.

I run these commands to create the \*\_output.txt files.



```
MINGW64/c:/Users/manas/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
data.json
frequency.py
frequency_output.txt
happiest_state.py
happiest_state_output.txt
README.md
'Report on Twitter sentiment analysis assignment.docx'
scraper.py
term_sentiment_output.txt
top_ten.py
top_ten_output.txt
tweet_sentiment.py
tweet_sentiment_output.txt
tweets.json
us_states.json

Press ENTER or type command to continue

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$ python happiest_state.py AFINN-111.txt tweets.json > happiest_state_output.txt

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$ python frequency.py tweets.json > frequency_output.txt

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$ python term_sentiment.py AFINN-111.txt tweets.json > term_sentiment_output.txt

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$ python tweet_sentiment.py AFINN-111.txt tweets.json > tweet_sentiment_output.txt

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$ python top_ten.py tweets.json > top_ten_output.txt

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$
```

Scraper.py



```
MINGW64~/Users/manas/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
-2
0
-1
0
0
0
2
-3
0
2
0
0
0
0
0
-1
-2
0
0
0
0
0
0
2
0
0
0
0
-9
-5
2
manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analys
is/submission
$ python tweet_sentiment.py AFINN-111.txt tweets.json
```

```
MINGW64~/Users/manas/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
-2
0
-1
0
0
0
2
-3
0
2
0
0
0
0
0
-1
-2
0
0
0
0
0
0
2
0
0
0
0
-9
-5
2
manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analys
is/submission
$
```

Term\_sentiment.py

```
MINGW64/c/Users/manas/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
-2
0
-1
0
0
0
2
-3
0
2
0
0
0
0
0
-1
-2
0
0
0
0
0
2
0
0
0
0
2
-9
-5
2

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$ python term sentiment.py AFINN-111.txt tweets.json

kvngtreyy 0
fw 0
somebody 0
vikibaggins 0
vosotros 0
tambi 0
lamentais 0
vuestra 0
veintena 0
desperdiciada 0
sumida 0
depresi 0
miseria 0
sois 0
normal 0
cowsflying 0
nibellion 0
canadian 1.5
dollars 0.6666666666666666
tf 0
rule 0
sum -2.5
rekiyaajb 0
ion 0
inna -5.0
shovel -5.0
taeathegreat 0
simple 2.0
tzq 2.0
hicbmw 2.0

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$
```

Frequency.py

```

MINGW64/c/Users/manas/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
kvngrtrey 0
fw 0
somebody 0
vikibaggins 0
vosotros 0
tambi 0
lamentais 0
vuestra 0
veintena 0
desperdiciada 0
sumida 0
depresi 0
miseria 0
sois 0
normal 0
cowsflying 0
nibellion 0
canadian 1.5
dollars 0.6666666666666666
tf 0
rule 0
sum -2.5
rekiyaajb 0
ion 0
inna -5.0
shovel -5.0
taeathegreat 0
simple 2.0
tzq 2.0
hicbmw 2.0

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$ python frequency.py tweets.json

```

```

MINGW64/c/Users/manas/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
fw 0.00015414258188824664
somebody 7.707129094412332e-05
vikibaggins 7.707129094412332e-05
vosotros 7.707129094412332e-05
tambi 7.707129094412332e-05
lamentais 7.707129094412332e-05
vuestra 7.707129094412332e-05
veintena 7.707129094412332e-05
desperdiciada 7.707129094412332e-05
sumida 7.707129094412332e-05
depresi 7.707129094412332e-05
miseria 7.707129094412332e-05
sois 7.707129094412332e-05
normal 7.707129094412332e-05
cowsflying 7.707129094412332e-05
nibellion 7.707129094412332e-05
canadian 7.707129094412332e-05
dollars 7.707129094412332e-05
tf 7.707129094412332e-05
rule 7.707129094412332e-05
curse 7.707129094412332e-05
sum 7.707129094412332e-05
rekiyaajb 7.707129094412332e-05
ion 7.707129094412332e-05
inna 7.707129094412332e-05
shovel 7.707129094412332e-05
taeathegreat 7.707129094412332e-05
simple 7.707129094412332e-05
tzq 7.707129094412332e-05
hicbmw 7.707129094412332e-05

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$

```

Happiest\_state.py

```
MINGW64/c:/Users/manas/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
lamentaís 7.707129094412332e-05
vuestra 7.707129094412332e-05
veintena 7.707129094412332e-05
desperdiciada 7.707129094412332e-05
sumida 7.707129094412332e-05
depresi 7.707129094412332e-05
miseria 7.707129094412332e-05
sois 7.707129094412332e-05
normal 7.707129094412332e-05
cowsflying 7.707129094412332e-05
nibellion 7.707129094412332e-05
canadian 7.707129094412332e-05
dollars 7.707129094412332e-05
tf 7.707129094412332e-05
rule 7.707129094412332e-05
curse 7.707129094412332e-05
sum 7.707129094412332e-05
rekiyaajb 7.707129094412332e-05
ion 7.707129094412332e-05
inna 7.707129094412332e-05
shovel 7.707129094412332e-05
taeathegreat 7.707129094412332e-05
simple 7.707129094412332e-05
tzq 7.707129094412332e-05
hicbmw 7.707129094412332e-05

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analys
is/submission
$ python happiest_state.py AFINN-111.txt tweets.json
NY

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analys
is/submission
$
```

top\_ten.py



```

MINGW64/c/Users/manas/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
curse 7.707129094412332e-05
sum 7.707129094412332e-05
rekiyaajb 7.707129094412332e-05
ion 7.707129094412332e-05
inna 7.707129094412332e-05
shovel 7.707129094412332e-05
taeathegreat 7.707129094412332e-05
simple 7.707129094412332e-05
tzq 7.707129094412332e-05
hicbmw 7.707129094412332e-05

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$ python happiest_state.py AFINN-111.txt tweets.json
NY

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$ python top_ten.py tweets.json
NowPlaying 5
yoshuabengio 2
OurMoonlightCHEN 2
EXO 2
earthquake 2
BTSARMY 2
TARGET 1
GBNeedDAMRights 1
Pelosisetupcontroversy 1
pelosiicecreamqueen 1
PelosiMustGo 1

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis/submission
$

```

```

MINGW64/c/Users/manas/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis
preemptively 2.751417667953413e-06
spat 2.751417667953413e-06

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis
$ python happiest_state.py AFINN-111.txt data.json
CA

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis
$ python top_ten.py
Traceback (most recent call last):
  File "top_ten.py", line 6, in <module>
    DATAJSON = sys.argv[1]
IndexError: list index out of range

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis
$ python top_ten.py data.json
Retail 22
tbt 22
pdx911 22
realestate 21
traffic 20
notanotherminute 17
SayHisName 16
ThrowbackThursday 16
Labor 15
Nursing 14
listing 13

manas@DESKTOP-TR4AUK6 MINGW64 ~/projects/uop/fall2020/comp261_01_data_science/projects/twitter_sentiment_analysis
$

```