This document logs some of the language processing/sentiment analysis tests I did.

TEST 1

sentiment analysis of text
used stopword library without modification

```
Lexed Data:  I am always happy
Tokenized Data:  [ 'i', 'am', 'always', 'happy' ]
After removing stopwords:  [ 'always', 'happy' ]
Sentiment Score:  1.5

Lexed Data:  I am incredibly happy
Tokenized Data:  [ 'i', 'am', 'incredibly', 'happy' ]
After removing stopwords:  [ 'incredibly', 'happy' ]
Sentiment Score:  1.5

Lexed Data:  I am never not happy
Tokenized Data:  [ 'i', 'am', 'never', 'not', 'happy' ]
After removing stopwords:  [ 'not', 'happy' ]
Sentiment Score:  -1.5

Lexed Data:  I am so very happy
Tokenized Data:  [ 'i', 'am', 'so', 'very', 'happy' ]
After removing stopwords:  [ 'so', 'happy' ]
Sentiment Score:  1.5

Lexed Data:  I am really happy
Tokenized Data:  [ 'i', 'am', 'really', 'happy' ]
After removing stopwords:  [ 'really', 'happy' ]
Sentiment Score:  1.5

Lexed Data:  I am happy
Tokenized Data:  [ 'i', 'am', 'happy' ]
After removing stopwords:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am quite happy
Tokenized Data:  [ 'i', 'am', 'quite', 'happy' ]
After removing stopwords:  [ 'quite', 'happy' ]
Sentiment Score:  1.5

Lexed Data:  I am kind of happy
Tokenized Data:  [ 'i', 'am', 'kind', 'of', 'happy' ]
After removing stopwords:  [ 'kind', 'happy' ]
Sentiment Score:  2.5
Lexed Data:  I am barely happy
Tokenized Data:  [ 'i', 'am', 'barely', 'happy' ]
After removing stopwords:  [ 'barely', 'happy' ]
Sentiment Score:  1.5
```

```
Lexed Data:  I am not very happy
Tokenized Data:  [ 'i', 'am', 'not', 'very', 'happy' ]
After removing stopwords:  [ 'not', 'happy' ]
Sentiment Score:  -1.5

Lexed Data:  I am not really happy
Tokenized Data:  [ 'i', 'am', 'not', 'really', 'happy' ]
After removing stopwords:  [ 'not', 'really', 'happy' ]
Sentiment Score:  -1

Lexed Data:  I am not quite happy
Tokenized Data:  [ 'i', 'am', 'not', 'quite', 'happy' ]
After removing stopwords:  [ 'not', 'quite', 'happy' ]
Sentiment Score:  -1

Lexed Data:  I am not happy
Tokenized Data:  [ 'i', 'am', 'not', 'happy' ]
After removing stopwords:  [ 'not', 'happy' ]
Sentiment Score:  -1.5

Lexed Data:  I am really not happy
Tokenized Data:  [ 'i', 'am', 'really', 'not', 'happy' ]
After removing stopwords:  [ 'really', 'not', 'happy' ]
Sentiment Score:  -1

Lexed Data:  I am never happy
Tokenized Data:  [ 'i', 'am', 'never', 'happy' ]
After removing stopwords:  [ 'happy' ]
Sentiment Score:  3
```

NOTES/OBSERVATIONS

- the text highlighted in red indicate results I deemed incorrect
- library isn't very flexible: cannot easily add to or remove from the list of stopwords
- certain negation terms not handled properly
    - i.e. "never" is considered to be a stopword which is problematic since "i am never happy" is filtered to "happy" and therefore gets a positive rating

TEST 2

sentiment analysis of text
used n-stopwords library without modification

```
Lexed Data:  I am always happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am incredibly happy
After removing stopwords:  incredibly happy
Tokenized Data:  [ 'incredibly', 'happy' ]
Sentiment Score:  1.5

Lexed Data:  I am never not happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am so very happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am really happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am quite happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am kind of happy
After removing stopwords:  kind happy
Tokenized Data:  [ 'kind', 'happy' ]
Sentiment Score:  2.5

Lexed Data:  I am barely happy
After removing stopwords:  barely happy
Tokenized Data:  [ 'barely', 'happy' ]
Sentiment Score:  1.5

Lexed Data:  I am not very happy
After removing stopwords:  happy
```

```
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am not really happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am not quite happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am not happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am really not happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am never happy
After removing stopwords:  happy
Tokenized Data:  [ 'happy' ]
Sentiment Score:  3
```

NOTES/OBSERVATIONS

- n-stopwords library has more flexibility than stopwords library
    - i.e. can add/remove stopwords, check if a word is a stopword, display all stopwords in the list

- negation terms handled very poorly in this library
    - i.e. "not" is considered to be a stopword which is problematic since "i am not happy" is filtered to "happy" and therefore gets a positive rating

TEST 3

sentiment analysis of text
used n-stopwords library with modifications made to consider negation words

```
Lexed Data:  I am always happy
Tokenized Data:  [ 'happy' ]
Tokenized Data (negations applied):  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am incredibly happy
Tokenized Data:  [ 'incredibly', 'happy' ]
Tokenized Data (negations applied):  [ 'incredibly', 'happy' ]
Sentiment Score:  1.5

Lexed Data:  I am never not happy
Tokenized Data:  [ 'never', 'not', 'happy' ]
Tokenized Data (negations applied):  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am so very happy
Tokenized Data:  [ 'happy' ]
Tokenized Data (negations applied):  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am really happy
Tokenized Data:  [ 'happy' ]
Tokenized Data (negations applied):  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am happy
Tokenized Data:  [ 'happy' ]
Tokenized Data (negations applied):  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am quite happy
Tokenized Data:  [ 'happy' ]
Tokenized Data (negations applied):  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am kind of happy
Tokenized Data:  [ 'kind', 'happy' ]
Tokenized Data (negations applied):  [ 'kind', 'happy' ]
Sentiment Score:  2.5

Lexed Data:  I am barely happy
Tokenized Data:  [ 'barely', 'happy' ]
Tokenized Data (negations applied):  []
Sentiment Score:  NaN

Lexed Data:  I am not very happy
Tokenized Data:  [ 'not', 'happy' ]
```

```
Tokenized Data (negations applied):  []
Sentiment Score:  NaN

Lexed Data:  I am not really happy
Tokenized Data:  [ 'not', 'happy' ]
Tokenized Data (negations applied):  []
Sentiment Score:  NaN

Lexed Data:  I am not quite happy
Tokenized Data:  [ 'not', 'happy' ]
Tokenized Data (negations applied):  []
Sentiment Score:  NaN

Lexed Data:  I am not happy
Tokenized Data:  [ 'not', 'happy' ]
Tokenized Data (negations applied):  []
Sentiment Score:  NaN

Lexed Data:  I am really not happy
Tokenized Data:  [ 'not', 'happy' ]
Tokenized Data (negations applied):  []
Sentiment Score:  NaN

Lexed Data:  I am never happy
Tokenized Data:  [ 'never', 'happy' ]
Tokenized Data (negations applied):  []
Sentiment Score:  NaN

Lexed Data:  I am not sad, I am happy
Tokenized Data:  [ 'not', 'sad', 'happy' ]
Tokenized Data (negations applied):  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am not not happy
Tokenized Data:  [ 'not', 'not', 'happy' ]
Tokenized Data (negations applied):  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am not not sad, but not happy
Tokenized Data:  [ 'not', 'not', 'sad', 'not', 'happy' ]
Tokenized Data (negations applied):  [ 'sad' ]
Sentiment Score:  -2

Lexed Data:  I am not happy, nor sad
Tokenized Data:  [ 'not', 'happy', 'nor', 'sad' ]
Tokenized Data (negations applied):  []
Sentiment Score:  NaN

Lexed Data:  I am unhappy
Tokenized Data:  [ 'unhappy' ]
Tokenized Data (negations applied):  [ 'unhappy' ]
Sentiment Score:  -2
```

```
Lexed Data:  I am not unhappy but I am not sad either
Tokenized Data:  [ 'not', 'unhappy', 'not', 'sad' ]
Tokenized Data (negations applied):  []
Sentiment Score:  NaN

Lexed Data:  I am not usually unhappy but I am now happy
Tokenized Data:  [ 'not', 'unhappy', 'happy' ]
Tokenized Data (negations applied):  [ 'happy' ]
Sentiment Score:  3

Lexed Data:  I am not often unhappy but right now i am unhappy
Tokenized Data:  [ 'not', 'unhappy', 'unhappy' ]
Tokenized Data (negations applied):  [ 'unhappy' ]
Sentiment Score:  -2
```

NOTES/OBSERVATIONS

- NaN represents neutral in these tests

- negation terms working as desired
    - i.e. single negations cancel next significant (tokenized) word, double negations cancel each other

- should not happy = unhappy? or inconclusive?
    - if not happy = inconclusive then passed all test cases

- the intensity of the extracted emotion isn't important for the visualization, unless it negates the meaning
    - e.g. "happy", "very happy" and "extremely happy" -> happiness
    - e.g. "barely happy", "rarely happy" -> inconclusive (to be barely happy is not a form of happiness)
    - e.g. "not happy", "never happy" -> inconclusive (not happy doesn't necessarily mean sad or any other particular emotion, unless unhappy is considered to be an emotion of interest?)
    - e.g. "not not happy" -> happiness (double negation so the "not"s cancel each other out)

TEST 4

sentiment analysis of text
used n-stopwords library with modifications made to consider negation words
used sample Tweets that feature the #covid19 hashtag

Lexed Data: Crazy that the world has come to this but as Americans we will fight to get through this!
Tokenized Data (negations applied): ['crazy', 'world', 'americans', 'fight']
Sentiment Score: -1
Expected: +, Outcome: -

Lexed Data: Together as One! #Covid19 #EconomicRecovery
Tokenized Data (negations applied): ['covid', 'economicrecovery']
Sentiment Score: 0
Expected: +, Outcome: ~

Lexed Data: Hospital cleaners, porters & catering staff are unsung #COVID19 heroes.
Tokenized Data (negations applied): [
  'hospital', 'cleaners',
  'porters', 'amp',
  'catering', 'staff',
  'unsung',   'covid',
  'heroes'
]
Sentiment Score: 0.2222222222222222
Expected: +, Outcome: +

Lexed Data: We Shall Over Come! With victory in our minds in our #FightAgainstCorona
Tokenized Data (negations applied): ['victory', 'minds', 'fightagainstcorona']
Sentiment Score: 1
Expected: +, Outcome: +

Lexed Data: It will be awesome @rihanna please try and consider a collaboration with @TiwaSavage .
Tokenized Data (negations applied): ['awesome', 'rihanna', 'collaboration', 'tiwasavage']
Sentiment Score: 1
Expected: +, Outcome: +

Lexed Data: The savior of humanity will come
Tokenized Data (negations applied): ['savior', 'humanity']
Sentiment Score: 1
Expected: +, Outcome: +

Lexed Data: Man, you get better. At least your sense of humor, is all good.
Tokenized Data (negations applied): ['man', 'sense', 'humor', 'good']
Sentiment Score: 1.25
Expected: +, Outcome: +

Lexed Data: I love You !!!
Tokenized Data (negations applied): ['love']
Sentiment Score: 2

Expected: +, Outcome: +

Lexed Data: My new #facemask from @chicagotheband makes me smile! Mask up, friends! Stay safe and well.
Tokenized Data (negations applied): [
  'facemask',
  'chicagotheband',
  'makes',
  'smile',
  'mask',
  'friends',
  'stay',
  'safe'
]
Sentiment Score: 0.5
Expected: +, Outcome: +

Lexed Data: So good to be sitting out, enjoying one of our pre #COVID19 haunts @chief_coffee
Tokenized Data (negations applied): [
  'good',
  'sitting',
  'enjoying',
  'pre',
  'covid',
  'haunts',
  'chiefcoffee'
]
Sentiment Score: 0.5714285714285714
Expected: +, Outcome: +
Lexed Data: The community is the best part of a country. I love my country. Covid 19 helping team. Love you all
Tokenized Data (negations applied): [
  'community', 'part',
  'country',   'love',
  'country',   'covid',
  'helping',   'team',
  'love'
]
Sentiment Score: 0.6666666666666666
Expected: +, Outcome: +

Lexed Data: Good day Beautiful People, this to wish you all a refreshing weekend.
Tokenized Data (negations applied): ['good', 'day', 'beautiful', 'people', 'refreshing', 'weekend']
Sentiment Score: 1
Expected: +, Outcome: +

Lexed Data: Yesterday I had the pleasure of shooting Corinne and Pedro’s wedding at Hendon Town Hall
Tokenized Data (negations applied): [
  'yesterday', 'pleasure',
  'shooting', 'corinne',
  'pedros',   'wedding',

```
  'hendon',    'town',
  'hall'
]
Sentiment Score:  0.2222222222222222
Expected: +, Outcome: +

Lexed Data:  Appreciation for being the front liner in COVID ICU SHL.
Tokenized Data (negations applied):  [ 'appreciation', 'front', 'liner', 'covid', 'icu', 'shl'
]
Sentiment Score:  0.3333333333333333
Expected: +, Outcome: +

Lexed Data:  So excited about this Book! Number one seller on Education &amp;amp; Covid right now
on Amazon!
Tokenized Data (negations applied):  [
  'excited',   'book',
  'number',    'seller',
  'education', 'amp',
  'covid',     'amazon'
]
Sentiment Score:  0.375
Expected: +, Outcome: +

Lexed Data:  So he can spread lies, hate, sexual violence, and murder?
Tokenized Data (negations applied):  [ 'spread', 'lies', 'hate', 'sexual', 'violence',
'murder' ]
Sentiment Score:  -1.6666666666666667
Expected: -, Outcome: -

Lexed Data:  The anxiety surrounding COVID-19 has caused Bipolar and Chron’s Disease to flare
up quite badly
Tokenized Data (negations applied):  [
  'anxiety', 'surrounding',
  'covid',   'caused',
  'bipolar', 'chrons',
  'disease', 'flare',
  'badly'
]
Sentiment Score:  -0.6666666666666666
Expected: -, Outcome: -

Lexed Data:  It so brilliant to see people awake and not accepting the #COVID19 mainstream
narrative.
Tokenized Data (negations applied):  [ 'brilliant', 'people', 'awake', 'covid', 'mainstream',
'narrative' ]
Sentiment Score: 0.6666666666666666
Expected: -, Outcome: +

Lexed Data:  This government was too slow to act, and even they now admit that. They followed
and did not lead!
Tokenized Data (negations applied):  [ 'government', 'slow', 'act', 'admit' ]
Sentiment Score:  -0.25
Expected: -, Outcome: -
```

Lexed Data: If you think I’m buying any of these rumours of us being locked down or restricted for up to 2 years
Tokenized Data (negations applied): ['im', 'buying', 'rumours', 'locked', 'restricted', 'years']
Sentiment Score: -0.3333333333333333
Expected: -, Outcome: -

Lexed Data: Oh for fvck’s sake @CNN @sarahcwestwood STOP WITH THE BS.
Tokenized Data (negations applied): ['fvcks', 'sake', 'cnn', 'sarahcwestwood', 'stop', 'bs']
Sentiment Score: -0.16666666666666666
Expected: -, Outcome: -

Lexed Data: Yes Green it is your fault. Next time put your foot down and say heck no!
Tokenized Data (negations applied): ['green', 'fault', 'time', 'put', 'foot', 'heck']
Sentiment score: 0
Expected: -, Outcome: ~

Lexed Data: #America has a problem and IT IS NOT #COVID19 Shocked? #MarxistBLM brought America to its knees!
Tokenized Data (negations applied): [
  'america',
  'problem',
  'shocked',
  'marxistblm',
  'brought',
  'america',
  'knees'
]
Sentiment Score: -0.5714285714285714
Expected: -, Outcome: -

Lexed Data: I am not even worried about myself- if I have it I have exposed my elderly grandmother
Tokenized Data (negations applied): ['exposed', 'elderly', 'grandmother']
Sentiment Score: -0.3333333333333333
Expected: -, Outcome: -

Lexed Data: This is exactly what I was afraid would happen.
Tokenized Data (negations applied): ['afraid', 'happen']
Sentiment Score: -1
Expected: -, Outcome: -

Lexed Data: I’m done with you. Just more lies and propaganda
Tokenized Data (negations applied): ['im', 'lies', 'propaganda']
Sentiment Score: -1.3333333333333333
Expected: -, Outcome: -

Lexed Data: The mental health of our community is of concern with #COVID19.
Tokenized Data (negations applied): ['mental', 'health', 'community', 'concern', 'covid']
Sentiment Score: -0.4
Expected: -, Outcome: -

Lexed Data: Seriously...Holy sh!t America, I want off this ride. it is not fun anymore. This is toxic.
Tokenized Data (negations applied): ['seriouslyholy', 'sht', 'america', 'ride', 'anymore', 'toxic']
Sentiment Score: -0.5
Expected: -, Outcome: -

Lexed Data: @realDonaldTrump This is who should be banned. He is a menace to the American Society.
Tokenized Data (negations applied): ['realdonaldtrump', 'banned', 'menace', 'american', 'society']
Sentiment Score: -0.8
Expected: -, Outcome: -

Lexed Data: Not being able to be there for a loved one during one of the worst times of their life HURTS
Tokenized Data (negations applied): ['worst', 'times', 'life', 'hurts']
Sentiment Score: -1.25
Expected: -, Outcome: -

Lexed Data: This is heartbreaking. Kuching FA football player Joseph Kalang Tie had to be quarantined at home
Tokenized Data (negations applied): [
  'heartbreaking',
  'kuching',
  'fa',
  'football',
  'player',
  'joseph',
  'kalang',
  'tie',
  'quarantined',
  'home'
]
Sentiment Score: -0.3
Expected: -, Outcome: -

Lexed Data: To the lady on the train applying her makeup and not wearing a mask.... Shame on you
Tokenized Data (negations applied): ['lady', 'train', 'applying', 'makeup', 'mask', 'shame']
Sentiment Score: -0.3333333333333333
Expected: -, Outcome: -

Lexed Data: STOP This Nonsense!!!
Tokenized Data (negations applied): ['stop', 'nonsense']
Sentiment Score: -1.5
Expected: -, Outcome: -

Lexed Data: This is truly frightening. Playing such outrageous conspiracy theories on TV
Tokenized Data (negations applied): [
  'frightening',

```
  'playing',
  'outrageous',
  'conspiracy',
  'theories',
  'tv'
]
Sentiment Score:  -1.5
Expected: -, Outcome: -

Lexed Data:  Think the link between 5G and #COVID19 is the stuff of loony conspiracy-theories?
Tokenized Data (negations applied):  [ 'link', 'g', 'covid', 'stuff', 'loony',
'conspiracytheories' ]
Sentiment Score:  0
Expected: ~, Outcome: ~

Lexed Data:  Learn new things during this period 2d and 3d animations training ongoing...
Tokenized Data (negations applied):  [
  'learn',    'things',
  'period',   'd',
  'd',        'animations',
  'training', 'ongoing'
]
Sentiment Score:  0
Expected: ~, Outcome: ~

Lexed Data:  ALERT: Parents decide whether to send kids back to school as coronavirus spreads
faster than ever
Tokenized Data (negations applied):  [
  'alert',   'parents',
  'decide',  'send',
  'kids',    'back',
  'school',  'coronavirus',
  'spreads', 'faster'
]
Sentiment Score:  0
Expected: ~, Outcome: ~

Lexed Data:  A surfer walks at Recreio dos Bandeirantes beach, amid the coronavirus disease
(COVID-19) outbreak
Tokenized Data (negations applied):  [
  'surfer',       'walks',
  'recreio',      'dos',
  'bandeirantes', 'beach',
  'amid',         'coronavirus',
  'disease',      'covid',
  'outbreak'
]
Sentiment Score:  -0.2727272727272727
Expected: ~, Outcome: -

Lexed Data:  Just in: 1,142 new #COVID19 cases were reported in #Delhi in the past 24 hours
Tokenized Data (negations applied):  [ 'covid', 'cases', 'reported', 'delhi', 'past', 'hours'
]
```

```
Sentiment Score:  0
Expected: ~, Outcome: ~

Lexed Data:  How #COVID19 Causes Smell Loss - Neuroscience News https://t.co/OG0gIhyctD
Tokenized Data (negations applied):  [
  'covid',
  'smell',
  'loss',
  'neuroscience',
  'news',
  'httpstcooggihyctd'
]
Sentiment Score:  -0.5
Expected: ~, Outcome: -

Lexed Data:  Corona Alert | Four arrivals from Chennai tested positive for Covid-19
Tokenized Data (negations applied):  [
  'corona',   'alert',
  'arrivals', 'chennai',
  'tested',   'positive',
  'covid'
]
Sentiment Score:  0.14285714285714285
Expected: ~, Outcome: +

Lexed Data:  Passengers cheer as ‘Karen’ is kicked off flight for refusing to wear mask
#airlines #coronavirus
Tokenized Data (negations applied):  [
  'passengers', 'cheer',
  'karen',      'kicked',
  'flight',     'refusing',
  'wear',       'mask',
  'airlines',   'coronavirus'
]
Sentiment Score:  0
Expected: ~, Outcome: ~

Lexed Data:  Coronavirus can infect people 26 FEET away in cold moving air
Tokenized Data (negations applied):  [ 'coronavirus', 'infect', 'people', 'feet', 'cold',
'moving', 'air' ]
Sentiment Score:  -0.2857142857142857
Expected: ~, Outcome: -

Lexed Data:  If you bought a face mask with an air valve during the bushfires, DO NOT use it
during #COVID19
Tokenized Data (negations applied):  [ 'bought', 'face', 'mask', 'air', 'valve', 'bushfires' ]
Sentiment Score:  0
Expected: ~, Outcome: ~

Lexed Data:  #USA deaths top 1,100 for third day in a row as #India sees 49,000 new cases
Tokenized Data (negations applied):  [
  'usa',  'deaths',
  'top',  'day',
```

```
  'row',  'india',
  'sees', 'cases'
]
Sentiment Score:  0
Expected: ~, Outcome: ~

Lexed Data:  Indonesian has received 100 ventilators from the Australian Government for
COVID-19 handling
Tokenized Data (negations applied):  [
  'indonesian',
  'received',
  'ventilators',
  'australian',
  'government',
  'covid',
  'handling'
]
Sentiment Score:  0
Expected: ~, Outcome: ~

Lexed Data:  The total number of #COVID19 infections worldwide is quickly nearing 16 million
Tokenized Data (negations applied):  [
  'total',
  'number',
  'covid',
  'infections',
  'worldwide',
  'quickly',
  'nearing',
  'million'
]
Sentiment Score:  -0.25
Expected: ~, Outcome: -

Lexed Data:  So gyms are reopening today...wonder how many people will be back working up a
sweat
Tokenized Data (negations applied):  [
  'gyms',
  'reopening',
  'todaywonder',
  'people',
  'back',
  'working',
  'sweat'
]
Sentiment Score:  0.14285714285714285
Expected: ~, Outcome: +

Lexed Data:  Parents ask their #COVID19 questions about #backtoschool
Tokenized Data (negations applied):  [ 'parents', 'covid', 'questions', 'backtoschool' ]
Sentiment Score:  -0.25
Expected: ~, Outcome: -
```

```
Lexed Data:  Establishing triage stations at healthcare facilities is very important to
protect health workers
Tokenized Data (negations applied):  [
  'establishing',
  'triage',
  'stations',
  'healthcare',
  'facilities',
  'important',
  'protect',
  'health',
  'workers'
]
Sentiment Score:  0.3333333333333333
Expected: ~, Outcome: +
```

NOTES/OBSERVATIONS

- \+ → positive score
  \- → negative score
  ~ → neutral score (0)

- results in red represent incorrect results, meaning one of the following happened:
  - an expected positive outputted a negative score
  - an expected negative outputted a positive score
  - an expected neutral outputted a positive score
  - an expected neutral outputted a negative score

- results in blue represent results that didn't come out as expected, but I don't consider them to be incorrect since a neutral score can simply symbolize inconclusiveness
  - an expected positive outputted a neutral score
  - an expected negative outputted a neutral score

- Total # of expected positive results is 15
  - 13 score positive (86.67%)
  - 1 scores negative (6.67%)
  - 1 scores neutral (6.67%)
  - → 93.33% of expected positive results are acceptable

- Total # of expected negative results is 19
  - 17 score negative (89.47%)
  - 1 scores positive (5.26%)
  - 1 scores neutral (5.26%)
  - → 94.74% of expected negative results are acceptable

- Total # of expected neutral results is 16
    - 8 score neutral (50%)
    - 3 score positive (18.75%)
    - 5 score negative (31.25%)
    - → 50% of expected neutral results are acceptable

- So, clearly, the expected neutral results that output a positive or negative score are the most problematic.

TEST 5

sentiment analysis of text
used n-stopwords library with modifications made to consider negation words and modifications made to score neutral words correctly
used sample Tweets that feature the #covid19 hashtag

Lexed Data: Crazy that the world has come to this but as Americans we will fight to get through this!
Tokenized Data (neutral words removed): ['crazy', 'world', 'americans', 'fight']
Sentiment Score: -1
Expected: +, Outcome: -

Lexed Data: Together as One! #Covid19 #EconomicRecovery
Tokenized Data (neutral words removed): ['covid', 'economicrecovery']
Sentiment Score: 0
Expected: +, Outcome: ~

Lexed Data: Hospital cleaners, porters & catering staff are unsung #COVID19 heroes.
Tokenized Data (neutral words removed): [
  'hospital', 'cleaners',
  'porters', 'amp',
  'catering', 'staff',
  'unsung',   'covid',
  'heroes'
]
Sentiment Score: 0.2222222222222222
Expected: +, Outcome: +

Lexed Data: We Shall Over Come! With victory in our minds in our #FightAgainstCorona
Tokenized Data (neutral words removed): ['victory', 'minds', 'fightagainstcorona']
Sentiment Score: 1
Expected: +, Outcome: +

Lexed Data: It will be awesome @rihanna please try and consider a collaboration with @TiwaSavage .
Tokenized Data (neutral words removed): ['awesome', 'rihanna', 'collaboration', 'tiwasavage']
Sentiment Score: 1
Expected: +, Outcome: +
Lexed Data: The savior of humanity will come
Tokenized Data (neutral words removed): ['savior', 'humanity']
Sentiment Score: 1
Expected: +, Outcome: +

Lexed Data: Man, you get better. At least your sense of humor, is all good.
Tokenized Data (neutral words removed): ['man', 'sense', 'humor', 'good']
Sentiment Score: 1.25
Expected: +, Outcome: +

Lexed Data: I love You !!!
Tokenized Data (neutral words removed): ['love']

Sentiment Score: 2
Expected: +, Outcome: +

Lexed Data: My new #facemask from @chicagotheband makes me smile! Mask up, friends! Stay safe and well.
Tokenized Data (neutral words removed): [
  'facemask',
  'chicagotheband',
  'makes',
  'smile',
  'mask',
  'friends',
  'stay',
  'safe'
]
Sentiment Score: 0.5
Expected: +, Outcome: +

Lexed Data: So good to be sitting out, enjoying one of our pre #COVID19 haunts @chief_coffee
Tokenized Data (neutral words removed): [
  'good',
  'sitting',
  'enjoying',
  'pre',
  'covid',
  'haunts',
  'chiefcoffee'
]
Sentiment Score: 0.5714285714285714
Expected: +, Outcome: +

Lexed Data: The community is the best part of a country. I love my country. Covid 19 helping team. Love you all
Tokenized Data (neutral words removed): [
  'community', 'part',
  'country',   'love',
  'country',   'covid',
  'helping',   'team',
  'love'
]
Sentiment Score: 0.6666666666666666
Expected: +, Outcome: +

Lexed Data: Good day Beautiful People, this to wish you all a refreshing weekend.
Tokenized Data (neutral words removed): ['good', 'day', 'beautiful', 'people', 'refreshing', 'weekend']
Sentiment Score: 1
Expected: +, Outcome: +

Lexed Data: Yesterday I had the pleasure of shooting Corinne and Pedro’s wedding at Hendon Town Hall
Tokenized Data (neutral words removed): [
  'yesterday', 'pleasure',

```
  'shooting',  'corinne',
  'pedros',    'wedding',
  'hendon',    'town',
  'hall'
]
Sentiment Score:  0.2222222222222222
Expected: +, Outcome: +

Lexed Data:  Appreciation for being the front liner in COVID ICU SHL.
Tokenized Data (neutral words removed):  [ 'appreciation', 'front', 'liner', 'covid', 'icu',
'shl' ]
Sentiment Score:  0.3333333333333333
Expected: +, Outcome: +

Lexed Data:  So excited about this Book! Number one seller on Education &amp;amp; Covid right now
on Amazon!
Tokenized Data (neutral words removed):  [
  'excited',   'book',
  'number',    'seller',
  'education', 'amp',
  'covid',     'amazon'
]
Sentiment Score:  0.375
Expected: +, Outcome: +

Lexed Data:  So he can spread lies, hate, sexual violence, and murder?
Tokenized Data (neutral words removed):  [ 'lies', 'hate', 'sexual', 'violence', 'murder' ]
Sentiment Score:  -2
Expected: -, Outcome: -

Lexed Data:  The anxiety surrounding COVID-19 has caused Bipolar and Chron’s Disease to flare
up quite badly
Tokenized Data (neutral words removed):  [
  'anxiety',
  'surrounding',
  'covid',
  'caused',
  'bipolar',
  'chrons',
  'flare',
  'badly'
]
Sentiment Score:  -0.625
Expected: -, Outcome: -

Lexed Data:  It so brilliant to see people awake and not accepting the #COVID19 mainstream
narrative.
Tokenized Data (neutral words removed):  [ 'brilliant', 'people', 'awake', 'covid',
'mainstream', 'narrative' ]
Sentiment Score:  0.6666666666666666
Expected: -, Outcome: +
```

Lexed Data: This government was too slow to act, and even they now admit that. They followed and did not lead!
Tokenized Data (neutral words removed): ['government', 'slow', 'act', 'admit']
Sentiment Score: -0.25
Expected: -, Outcome: -

Lexed Data: If you think I’m buying any of these rumours of us being locked down or restricted for up to 2 years
Tokenized Data (neutral words removed): ['im', 'buying', 'rumours', 'locked', 'restricted', 'years']
Sentiment Score: -0.3333333333333333
Expected: -, Outcome: -

Lexed Data: Oh for fvck’s sake @CNN @sarahcwestwood STOP WITH THE BS.
Tokenized Data (neutral words removed): ['fvcks', 'sake', 'cnn', 'sarahcwestwood', 'stop', 'bs']
Sentiment Score: -0.16666666666666666
Expected: -, Outcome: -

Lexed Data: Yes Green it is your fault. Next time put your foot down and say heck no!
Tokenized Data (neutral words removed): ['green', 'fault', 'time', 'put', 'foot', 'heck']
Sentiment Score: 0
Expected: -, Outcome: ~

Lexed Data: #America has a problem and IT IS NOT #COVID19 Shocked? #MarxistBLM brought America to its knees!
Tokenized Data (neutral words removed): [
  'america',
  'problem',
  'shocked',
  'marxistblm',
  'brought',
  'america',
  'knees'
]
Sentiment Score: -0.5714285714285714
Expected: -, Outcome: -

Lexed Data: I am not even worried about myself- if I have it I have exposed my elderly grandmother
Tokenized Data (neutral words removed): ['exposed', 'elderly', 'grandmother']
Sentiment Score: -0.3333333333333333
Expected: -, Outcome: -

Lexed Data: This is exactly what I was afraid would happen.
Tokenized Data (neutral words removed): ['afraid', 'happen']
Sentiment Score: -1
Expected: -, Outcome: -

Lexed Data: I’m done with you. Just more lies and propaganda
Tokenized Data (neutral words removed): ['im', 'lies', 'propaganda']
Sentiment Score: -1.3333333333333333

Expected: -, Outcome: -

Lexed Data: The mental health of our community is of concern with #COVID19.
Tokenized Data (neutral words removed): ['mental', 'health', 'community', 'concern', 'covid']
Sentiment Score: -0.4
Expected: -, Outcome: -

Lexed Data: Seriously...Holy sh!t America, I want off this ride. it is not fun anymore. This is toxic.
Tokenized Data (neutral words removed): ['seriouslyholy', 'sht', 'america', 'ride', 'anymore', 'toxic']
Sentiment Score: -0.5
Expected: -, Outcome: -

Lexed Data: @realDonaldTrump This is who should be banned. He is a menace to the American Society.
Tokenized Data (neutral words removed): ['realdonaldtrump', 'banned', 'menace', 'american', 'society']
Sentiment Score: -0.8
Expected: -, Outcome: -

Lexed Data: Not being able to be there for a loved one during one of the worst times of their life HURTS
Tokenized Data (neutral words removed): ['worst', 'times', 'life', 'hurts']
Sentiment Score: -1.25
Expected: -, Outcome: -

Lexed Data: This is heartbreaking. Kuching FA football player Joseph Kalang Tie had to be quarantined at home
Tokenized Data (neutral words removed): [
  'heartbreaking',
  'kuching',
  'fa',
  'football',
  'player',
  'joseph',
  'kalang',
  'tie',
  'quarantined',
  'home'
]
Sentiment Score: -0.3
Expected: -, Outcome: -

Lexed Data: To the lady on the train applying her makeup and not wearing a mask.... Shame on you
Tokenized Data (neutral words removed): ['lady', 'train', 'applying', 'makeup', 'mask', 'shame']
Sentiment Score: -0.3333333333333333
Expected: -, Outcome: -

Lexed Data: STOP This Nonsense!!!

Tokenized Data (neutral words removed): ['stop', 'nonsense']
Sentiment Score: -1.5
Expected: -, Outcome: -

Lexed Data: This is truly frightening. Playing such outrageous conspiracy theories on TV
Tokenized Data (neutral words removed): [
  'frightening',
  'playing',
  'outrageous',
  'conspiracy',
  'theories',
  'tv'
]
Sentiment Score: -1.5
Expected: -, Outcome: -

Lexed Data: Think the link between 5G and #COVID19 is the stuff of loony conspiracy-theories?
Tokenized Data (neutral words removed): ['link', 'g', 'covid', 'stuff', 'loony',
'conspiracytheories']
Sentiment Score: 0
Expected: ~, Outcome: ~

Lexed Data: Learn new things during this period 2d and 3d animations training ongoing...
Tokenized Data (neutral words removed): [
  'learn',   'things',
  'period',   'd',
  'd',   'animations',
  'training', 'ongoing'
]
Sentiment Score: 0
Expected: ~, Outcome: ~

Lexed Data: ALERT: Parents decide whether to send kids back to school as coronavirus spreads
faster than ever
Tokenized Data (neutral words removed): [
  'parents',
  'decide',
  'send',
  'kids',
  'school',
  'coronavirus',
  'faster'
]
Sentiment Score: 0
Expected: ~, Outcome: ~

Lexed Data: A surfer walks at Recreio dos Bandeirantes beach, amid the coronavirus disease
(COVID-19) outbreak
Tokenized Data (neutral words removed): [
  'surfer',
  'walks',
  'recreio',
  'dos',

```
  'bandeirantes',
  'beach',
  'amid',
  'coronavirus',
  'covid'
]
Sentiment Score:  0
```
Expected: ~, Outcome: ~

```
Lexed Data:  Just in: 1,142 new #COVID19 cases were reported in #Delhi in the past 24 hours
Tokenized Data (neutral words removed):  [ 'covid', 'cases', 'reported', 'delhi', 'past',
'hours' ]
Sentiment Score:  0


Lexed Data:  How #COVID19 Causes Smell Loss - Neuroscience News https://t.co/OG0gIhyctD
Tokenized Data (neutral words removed):  [ 'covid', 'smell', 'neuroscience', 'news',
'httpstcooggihyctd' ]
Sentiment Score:  0
```
Expected: ~, Outcome: ~

```
Lexed Data:  Corona Alert | Four arrivals from Chennai tested positive for Covid-19
Tokenized Data (neutral words removed):  [ 'corona', 'arrivals', 'chennai', 'tested', 'covid'
]
Sentiment Score:  0
```
Expected: ~, Outcome: ~

```
Lexed Data:  Passengers cheer as ‘Karen’ is kicked off flight for refusing to wear mask
#airlines #coronavirus
Tokenized Data (neutral words removed):  [
  'passengers', 'cheer',
  'karen',      'kicked',
  'flight',     'refusing',
  'wear',       'mask',
  'airlines',   'coronavirus'
]
Sentiment Score:  0
```
Expected: ~, Outcome: ~

```
Lexed Data:  Coronavirus can infect people 26 FEET away in cold moving air
Tokenized Data (neutral words removed):  [ 'coronavirus', 'people', 'feet', 'cold', 'moving',
'air' ]
Sentiment Score:  0
```
Expected: ~, Outcome: ~

```
Lexed Data:  If you bought a face mask with an air valve during the bushfires, DO NOT use it
during #COVID19
Tokenized Data (neutral words removed):  [ 'bought', 'face', 'mask', 'air', 'valve',
'bushfires' ]
Sentiment Score:  0
```
Expected: ~, Outcome: ~

```
Lexed Data:  #USA deaths top 1,100 for third day in a row as #India sees 49,000 new cases
```

Tokenized Data (neutral words removed): ['usa', 'day', 'row', 'india', 'sees', 'cases']
Sentiment Score: 0
Expected: ~, Outcome: ~

Lexed Data: Indonesian has received 100 ventilators from the Australian Government for COVID-19 handling
Tokenized Data (neutral words removed): [
  'indonesian',
  'received',
  'ventilators',
  'australian',
  'government',
  'covid',
  'handling'
]
Sentiment Score: 0
Expected: ~, Outcome: ~

Lexed Data: The total number of #COVID19 infections worldwide is quickly nearing 16 million
Tokenized Data (neutral words removed): [
  'total',
  'number',
  'covid',
  'worldwide',
  'quickly',
  'nearing',
  'million'
]
Sentiment Score: 0
Expected: ~, Outcome: ~

Lexed Data: So gyms are reopening today...wonder how many people will be back working up a sweat
Tokenized Data (neutral words removed): ['gyms', 'reopening', 'todaywonder', 'people', 'working', 'sweat']
Sentiment Score: 0
Expected: ~, Outcome: ~

Lexed Data: Parents ask their #COVID19 questions about #backtoschool
Tokenized Data (neutral words removed): ['parents', 'covid', 'backtoschool']
Sentiment Score: 0
Expected: ~, Outcome: ~

Lexed Data: Establishing triage stations at healthcare facilities is very important to protect health workers
Tokenized Data (neutral words removed): [
  'establishing',
  'triage',
  'stations',
  'healthcare',
  'facilities',
  'important',
  'protect',

```
  'health',
  'workers'
]
Sentiment Score:  0.3333333333333333
```
Expected: ~, Outcome: +

NOTES/OBSERVATIONS

- modification made: added a list of words that I consider to be more neutral than emotional when used in discussions surrounding COVID-19 (e.g. outbreak, disease, pandemic, virus, infect, alert, ill, spread, positive, negative) and removed these words from the sentiment score calculation

- results for expected positives and expected negatives are same as TEST 4 (as expected)

- Improvements in correctness of expected neutrals:
    - Total # of expected neutral results is 16
        - 15 score neutral (93.75%)
        - 1 scores positive (6.25%)
        - 0 score negative (0%)
        - → 93.75% of expected neutral results are acceptable