Data Collection

API and WEBSCRAPE articles

These articles will be used to:

- TRAIN and TEST MODEL
  - from CNN, FOX, BREITBERT, and WASHINGTON POST
- CLASSIFY Stakeholder's Data using trained MODEL
  - from THE HILL

In [ ]:
```python
#! pip install pyLDavis

import pyLDavis

pyLDavis.enable_notebook()
from tqdm.auto import tqdm

import pyLDavis.lda_model
import pyLDavis.gensim_models

import numpy as np
import pandas as pd
import pymysql as mysql
import matplotlib.pyplot as plt
import os
import shutil
import re
import logging
import time
import zipfile
import requests
from bs4 import BeautifulSoup
import datetime
import re
import regex as rex
from collections import defaultdict, Counter
import random
import requests
from bs4 import BeautifulSoup
import datetime
import json
from wordcloud import WordCloud
from tabulate import tabulate
from sklearn.svm import SVC


import sqlite3
import nltk
from string import punctuation
from nltk.corpus import stopwords
import re
import emoji
from nltk.metrics import ConfusionMatrix
import itertools
import collections

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.decomposition import NMF
```

```python
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, f1_score, recall_score, precision_recall_curve
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
## Depracated:
# from sklearn.metrics import plot_confusion_matrix
## New version:
from sklearn.metrics import ConfusionMatrixDisplay
from sklearn.model_selection import cross_val_score
from sklearn.pipeline import Pipeline
from sklearn.model_selection import GridSearchCV
from sklearn.svm import LinearSVC
from sklearn.decomposition import LatentDirichletAllocation

#import mysql.connector

# Set pandas global options
pd.options.display.max_rows = 17

import warnings
warnings.filterwarnings('ignore')
warnings.simplefilter('ignore')
```

Connect to NewsAPI client

In [ ]:
```python
from newsapi import NewsApiClient

api_key = os.environ['NewsAPIKey']

# Init
newsapi = NewsApiClient(api_key=api_key)
```

Pull article URLs and info from API

/v2/top-headlines/sources

sources = newsapi.get_sources() print(sources)

In [ ]:
```python
def news_api_urls(q=None,
                  s=None,
                  d_from='2023-05-01',
                  d_to='2023-05-31',
                  api_lst=[]):
    all_articles = newsapi.get_everything(q=q,
                                          sources=s,
                                          from_param=d_from,
                                          to=d_to,
                                          language='en',
                                          sort_by='relevancy',
                                          page=1)

    print(type(all_articles))
    print(all_articles)
    #print('Article list: ', all_articles['articles'])
    for article in all_articles['articles']:
        print('Source ID:', article['source']['id'])
        print('Source name:', article['source']['name'])
        print('Author:', article['author'])
        print('Title:', article['title'])
        print('URL:', article['url'])
```

```
            print('Publish date:', article['publishedAt'])
            print('Article text:', article['content'], '\n')

        # Create a list of tuples from the dictionary data
        source_data01 = [(a['source']['name'],
                          a['author'],
                          a['title'],
                          a['url'],
                          a['publishedAt'],
                          a['content'])
                         for a in all_articles['articles']]

    api_lst.extend(source_data01)
    #print(api_lst)
    print(len(api_lst))
```

Data Collection:

Connect to API to access URLs

Set API filter parameters
Extract Training Data for Classifier

Two Left leaning and Two Right leaning news sources are queried, collected, cleaned, and prepared for model training and test

```python
# Parameters to extract API News data from 4 sources:

source_lst = ['the-washington-post','cnn','fox-news','breitbart-news']


date_lst = ['2023-05-18', '2023-05-19', '2023-05-21', '2023-05-22', '2023-05-23']


q_word_lst = ['justice OR surveillance', 'healthcare OR "health care"',
              '(political AND (bias OR party)) OR republican OR democrat OR election',
              'security AND (social OR national)', '(policy AND (drug OR "affirmative acti
```

Extract Client Data for Business Application of our Classifer

Data from what is considered a Centered news source is extracted, preprocessed, and prepared to run through our trained Classifier. We report back the data gathered, the overall lean found in that data, and give details on the lean, if a political lean is found in the data.

```python
# Parameters to extract API News data from 1 centered source:

source_lst = ['associated-press']


date_lst = ['2023-05-18', '2023-05-19', '2023-05-20', '2023-05-21', '2023-05-22', '2023-05
            '2023-05-24', '2023-05-25', '2023-05-17', '2023-05-16', '2023-05-15', '2023-05
            '2023-05-26', '2023-05-27', '2023-05-28', '2023-05-29', '2023-05-30', '2023-05
            '2023-05-13', '2023-05-12']

#, '2023-05-11', '2023-05-10', '2023-05-09', '2023-05-08']

q_word_lst = ['justice OR surveillance', 'healthcare OR "health care"',
              '(political AND (bias OR party)) OR republican OR democrat OR election',
              'security AND (social OR national)', '(policy AND (drug OR "affirmative acti
```

```python
# Parameters to extract API News data from 1 centered source:

source_lst = ['the-hill']

date_lst = ['2023-05-18', '2023-05-19', '2023-05-20', '2023-05-21', '2023-05-22', '2023-05
            '2023-05-24', '2023-05-25', '2023-05-17', '2023-05-16', '2023-05-15', '2023-05
            '2023-05-26', '2023-05-27', '2023-05-28', '2023-05-29', '2023-05-30', '2023-05
#date_lst = ['2023-06-01','2023-06-02','2023-06-03','2023-06-04','2023-06-05','2023-06-06
#            '2023-06-08','2023-06-09','2023-06-10','2023-06-11','2023-06-12','2023-06-13
#            '2023-06-15']

q_word_lst = ['justice OR surveillance', 'healthcare OR "health care"',
              '(political AND (bias OR party)) OR republican OR democrat OR election',
              'security AND (social OR national)', '(policy AND (drug OR "affirmative acti
```

Access API (collect URLs)

Using the defined parameters above, extract URL and other data information from the desired news sources.

```python
api_record_lst01 = []
for s in source_lst:
    print(f'Source: {s}')
    for d in date_lst:
        print(f'Date: {d}')
        for q in q_word_lst:
            print(f'Query word: {q}')
            time.sleep(5 + 11 * random.random())
            news_api_urls(q=q,
                          s=s,
                          d_from=d,
                          d_to=d,
                          api_lst=api_record_lst01)
            print(s, d, q)
    time.sleep(10 + 13 * random.random())

#print(api_record_lst01)
#print(len(api_record_lst01))
```

Save into DataFrames:

Four sources data (used to TRAIN and TEST the MODEL):

```python
api_record_df2 = pd.DataFrame (list (api_record_lst01),
                               columns = ['Source','Author','Title','URL','date','cont
```

Centered Client data (stakeholder's data that will be CLASSIFIED using above mentioned MODEL):

```python
api_record_df = pd.DataFrame (list (api_record_lst01),
                              columns = ['Source','Author','Title','URL','date','cont
#api_record_df.to_csv("News_API_Reuters2.csv", sep=',')
api_record_df.to_csv("News_API_TheHill3.csv", sep=',')
```

```python
api_record_df.shape
```

```
(69, 6)
```

Code to combine different dataframes for training (my portion): (no need to run again)

```python
api_record_df_final = pd.concat([api_record_df2, api_record_df])
```

```python
api_record_df_final.to_csv("News_API_FOX_CNN_Breitbert_May18_23_May31.csv", sep=',')
api_record_df.to_csv("News_API_FOX_CNN_Breitbert_May20_May31.csv", sep=',')
api_record_df2.to_csv("News_API_FOX_CNN_Breitbert_May1819_May2123.csv", sep=',')
```

```python
api_record_df_filtered = api_record_df[api_record_df.Source != 'The Washington Post']
```

```python
api_data_df3=pd.read_csv('News_API_FOX_CNN_Breitbert_May18_23_May31_2.csv')
```

```python
api_record_df_final3 = api_data_df3[api_data_df3.Source != 'The Washington Post']
```

```python
api_record_df_final3.to_csv("News_API_FOX_CNN_Breitbert_May18_23_May31_3.csv", sep=',')
api_record_df_final3.shape
api_record_df_filtered
api_record_df2.shape
```

```python
api_record_df_final2 = pd.concat([api_record_df2, api_record_df_filtered])
api_record_df_final2.shape
```

```python
api_record_df_final2.to_csv("News_API_FOX_CNN_Breitbert_May18_23_May31_2.csv", sep=',')
api_data_df_final3=pd.read_csv('News_API_FOX_CNN_Breitbert_May18_23_May31_2.csv')
```

HTML scrape of URLs (to collect Article content)

FOX, CNN, Breitbert:

```python
slct_tbl_full_df02 = api_record_df_final3.copy()


slct_tbl_full_df02['article_parsed'] = ''

total_urls = len(slct_tbl_full_df02)

# Start timer
start_time = datetime.date.today()


for i, row in enumerate(slct_tbl_full_df02.itertuples(), 1):
    print(row[5])
    article=[]
    #url = slct_tbl_full_df02.at[i,'URL']
    url=row[5]
    response = requests.get(url)
    time.sleep(5 + 10*random.random())

    soup = BeautifulSoup(response.text, 'html.parser')

    try:
        script_tag = soup.find('script', {'type': 'application/ld+json'})
        if script_tag == None:
            print('\nJSON Object == None: Index:', i-1, '; source:',
```

```
                            row[2],
                            '; URL:', row[5])
                else:
                    article_json = json.loads(script_tag.string)
                    article_content = article_json['articleBody']
                    slct_tbl_full_df02.at[row.Index, 'article_parsed'] = article_content

            except KeyError:
                print('\nMissing JSON key: Index:', i-1, '; source:',
                        row[2],
                        '; URL:', row[5])
                article_body = soup.find('div', class_='article__content-container')

                if article_body is None: #forfoxandbreitbert(sometimes)
                    print('Class != article__content-container')
                    article_body = soup.find('p', class_="speakable")
                    if article_body is None: #breitbert(most)
                        print('Class != speakable')
                        article_body = soup.find('div', class_='entry-content')
                        if article_body is None: #WashPost
                            print('Class != entry-content')
                            article_body = soup.find('div',class_='article-body')

                if article_body is not None:
                    article_text = article_body.get_text()
                    slct_tbl_full_df02.at[row.Index, 'article_parsed'] = article_text
                else:
                    print('Could not parse')

        print('.', end='')

    # End timer script
    end_time = datetime.date.today()
    time_elapse = end_time - start_time
    print(f'Start Time = {start_time}')
    print(f'End Time = {end_time}')
    print(f'Elapsed Time = {time_elapse}')
```

In [ ]:
```
slct_tbl_full_df02.to_csv("News_API_FOX_CNN_Breitbert_May18_23_May31_3.csv", sep=',')
```

The Hill:

In [ ]:
```
slct_tbl_full_df02 = api_record_df.copy()


slct_tbl_full_df02['article_parsed'] = ''

total_urls = len(slct_tbl_full_df02)

# Start timer
start_time = datetime.date.today()


for i, row in enumerate(slct_tbl_full_df02.itertuples(), 1):
    print(row[4])
    article=[]
    #url = slct_tbl_full_df02.at[i,'URL']
    url=row[4]
    response = requests.get(url)
    time.sleep(5 + 10*random.random())

    soup = BeautifulSoup(response.text, 'html.parser')
```

```python
        try:
            script_tag = soup.find('script', {'type': 'application/ld+json'})
            if script_tag == None:
                print('\nJSON Object == None: Index:', i-1, '; source:',
                      row[1],
                      '; URL:', row[4])
            else:
                article_json = json.loads(script_tag.string)
                article_content = article_json['articleBody']
                slct_tbl_full_df02.at[row.Index, 'article_parsed'] = article_content

        except KeyError:
            print('\nMissing JSON key: Index:', i-1, '; source:',
                  row[1],
                  '; URL:', row[4])
            #article_body = soup.find('div', class_='article__content__6hMn9') #Reuters scrape
            #article_body = soup.find('div', class_='Article') #AP scrape attempt
            article_body = soup.find('div', class_='article__text | body-copy | flow') #theHil

            if article_body is None: #forfoxandbreitbert(sometimes)
                print('Class != article__content-container')
                article_body = soup.find('p', class_="speakable")
                if article_body is None: #breitbert(most)
                    print('Class != speakable')
                    article_body = soup.find('div', class_='entry-content')
                    if article_body is None: #WashPost
                        print('Class != entry-content')
                        article_body = soup.find('div',class_='article-body')

            if article_body is not None:
                article_text = article_body.get_text()
                slct_tbl_full_df02.at[row.Index, 'article_parsed'] = article_text
            else:
                print('Could not parse')

    print('.', end='')

# End timer script
end_time = datetime.date.today()
time_elapse = end_time - start_time
print(f'Start Time = {start_time}')
print(f'End Time = {end_time}')
print(f'Elapsed Time = {time_elapse}')
```

https://thehill.com/homenews/house/4010248-greene-plans-file-articles-impeachment-biden/

Missing JSON key: Index: 0 ; source: The Hill ; URL: https://thehill.com/homenews/house/40
10248-greene-plans-file-articles-impeachment-biden/
.https://thehill.com/policy/energy-environment/4011061-markey-bill-would-restore-ban-on-u-
s-fossil-fuel-exports/

Missing JSON key: Index: 1 ; source: The Hill ; URL: https://thehill.com/policy/energy-env
ironment/4011061-markey-bill-would-restore-ban-on-u-s-fossil-fuel-exports/
.https://thehill.com/policy/healthcare/4009038-what-a-default-could-mean-for-medicaid-and-
other-federal-benefits/

Missing JSON key: Index: 2 ; source: The Hill ; URL: https://thehill.com/policy/healthcar
e/4009038-what-a-default-could-mean-for-medicaid-and-other-federal-benefits/
.https://thehill.com/business/4010770-with-biden-in-japan-harris-pushes-white-house-messag
e-on-debt-ceiling/

Missing JSON key: Index: 3 ; source: The Hill ; URL: https://thehill.com/business/4010770-
with-biden-in-japan-harris-pushes-white-house-message-on-debt-ceiling/
.https://thehill.com/homenews/house/4010248-greene-plans-file-articles-impeachment-biden/

Missing JSON key: Index: 4 ; source: The Hill ; URL: https://thehill.com/homenews/house/4010248-greene-plans-file-articles-impeachment-biden/ .https://thehill.com/homenews/house/4011190-conservatives-make-last-minute-push-for-border-security-in-debt-ceiling-package/

Missing JSON key: Index: 5 ; source: The Hill ; URL: https://thehill.com/homenews/house/4011190-conservatives-make-last-minute-push-for-border-security-in-debt-ceiling-package/ .https://thehill.com/policy/healthcare/4009038-what-a-default-could-mean-for-medicaid-and-other-federal-benefits/

Missing JSON key: Index: 6 ; source: The Hill ; URL: https://thehill.com/policy/healthcare/4009038-what-a-default-could-mean-for-medicaid-and-other-federal-benefits/ .https://thehill.com/business/4010770-with-biden-in-japan-harris-pushes-white-house-message-on-debt-ceiling/

Missing JSON key: Index: 7 ; source: The Hill ; URL: https://thehill.com/business/4010770-with-biden-in-japan-harris-pushes-white-house-message-on-debt-ceiling/ .https://thehill.com/homenews/house/4010248-greene-plans-file-articles-impeachment-biden/

Missing JSON key: Index: 8 ; source: The Hill ; URL: https://thehill.com/homenews/house/4010248-greene-plans-file-articles-impeachment-biden/ .https://thehill.com/blogs/blog-briefing-room/4011488-hyundai-kia-agree-to-pay-200-million-to-settle-lawsuit-after-rise-in-thefts/

Missing JSON key: Index: 9 ; source: The Hill ; URL: https://thehill.com/blogs/blog-briefing-room/4011488-hyundai-kia-agree-to-pay-200-million-to-settle-lawsuit-after-rise-in-thefts/ .https://thehill.com/policy/healthcare/4009038-what-a-default-could-mean-for-medicaid-and-other-federal-benefits/

Missing JSON key: Index: 10 ; source: The Hill ; URL: https://thehill.com/policy/healthcare/4009038-what-a-default-could-mean-for-medicaid-and-other-federal-benefits/ .https://thehill.com/homenews/senate/4011114-cruz-opens-a-probe-into-anheuser-busch-over-dylan-mulvaney-partnership/

Missing JSON key: Index: 11 ; source: The Hill ; URL: https://thehill.com/homenews/senate/4011114-cruz-opens-a-probe-into-anheuser-busch-over-dylan-mulvaney-partnership/ .https://thehill.com/homenews/state-watch/4012940-florida-health-care-providers-are-pausing-gender-affirming-care-for-adults/

Missing JSON key: Index: 12 ; source: The Hill ; URL: https://thehill.com/homenews/state-watch/4012940-florida-health-care-providers-are-pausing-gender-affirming-care-for-adults/ .https://thehill.com/business/4012762-66-progressive-lawmakers-urge-biden-to-use-14th-amendment-in-debt-ceiling-fight/

Missing JSON key: Index: 13 ; source: The Hill ; URL: https://thehill.com/business/4012762-66-progressive-lawmakers-urge-biden-to-use-14th-amendment-in-debt-ceiling-fight/ .https://thehill.com/homenews/campaign/4011971-youngkin-fuels-2024-speculation-with-campaign-style-video/

Missing JSON key: Index: 14 ; source: The Hill ; URL: https://thehill.com/homenews/campaign/4011971-youngkin-fuels-2024-speculation-with-campaign-style-video/ .https://thehill.com/business/ap-american-airlines-and-jetblue-must-abandon-their-partnership-in-the-northeast-federal-judge-rules/

Missing JSON key: Index: 15 ; source: The Hill ; URL: https://thehill.com/business/ap-american-airlines-and-jetblue-must-abandon-their-partnership-in-the-northeast-federal-judge-rules/ .https://thehill.com/homenews/administration/4013134-white-house-optimistic-about-path-forward-despite-serious-differences-on-budget-talks/

Missing JSON key: Index: 16 ; source: The Hill ; URL: https://thehill.com/homenews/administration/4013134-white-house-optimistic-about-path-forward-despite-serious-differences-on-budget-talks/ .https://thehill.com/homenews/senate/4014417-senate-democrat-calls-potential-for-debt-default-a-manufactured-crisis/

Missing JSON key: Index: 17 ; source: The Hill ; URL: https://thehill.com/homenews/senate/
4014417-senate-democrat-calls-potential-for-debt-default-a-manufactured-crisis/
.https://thehill.com/homenews/senate/4014204-van-hollen-lawmakers-should-move-to-plan-b-as
-debt-ceiling-talks-stall/

Missing JSON key: Index: 18 ; source: The Hill ; URL: https://thehill.com/homenews/senate/
4014204-van-hollen-lawmakers-should-move-to-plan-b-as-debt-ceiling-talks-stall/
.https://thehill.com/homenews/administration/4014239-yellen-sidesteps-question-on-which-bi
lls-would-go-unpaid-if-us-defaults-on-debt/

Missing JSON key: Index: 19 ; source: The Hill ; URL: https://thehill.com/homenews/adminis
tration/4014239-yellen-sidesteps-question-on-which-bills-would-go-unpaid-if-us-defaults-on
-debt/
.https://thehill.com/homenews/senate/4014417-senate-democrat-calls-potential-for-debt-defa
ult-a-manufactured-crisis/

Missing JSON key: Index: 20 ; source: The Hill ; URL: https://thehill.com/homenews/senate/
4014417-senate-democrat-calls-potential-for-debt-default-a-manufactured-crisis/
.https://thehill.com/homenews/administration/4014239-yellen-sidesteps-question-on-which-bi
lls-would-go-unpaid-if-us-defaults-on-debt/

Missing JSON key: Index: 21 ; source: The Hill ; URL: https://thehill.com/homenews/adminis
tration/4014239-yellen-sidesteps-question-on-which-bills-would-go-unpaid-if-us-defaults-on
-debt/
.https://thehill.com/homenews/senate/4014204-van-hollen-lawmakers-should-move-to-plan-b-as
-debt-ceiling-talks-stall/

Missing JSON key: Index: 22 ; source: The Hill ; URL: https://thehill.com/homenews/senate/
4014204-van-hollen-lawmakers-should-move-to-plan-b-as-debt-ceiling-talks-stall/
.https://thehill.com/homenews/administration/4014242-yellen-calls-debt-ceiling-constraint-
on-nations-ability-to-pay-its-bills/

Missing JSON key: Index: 23 ; source: The Hill ; URL: https://thehill.com/homenews/adminis
tration/4014242-yellen-calls-debt-ceiling-constraint-on-nations-ability-to-pay-its-bills/
.https://thehill.com/homenews/4011270-this-brewery-made-tipping-optional-heres-why/

Missing JSON key: Index: 24 ; source: The Hill ; URL: https://thehill.com/homenews/4011270
-this-brewery-made-tipping-optional-heres-why/
.https://thehill.com/homenews/house/4015577-these-are-the-biggest-sticking-points-in-the-d
ebt-ceiling-fight/

Missing JSON key: Index: 25 ; source: The Hill ; URL: https://thehill.com/homenews/house/4
015577-these-are-the-biggest-sticking-points-in-the-debt-ceiling-fight/
.https://thehill.com/homenews/senate/4016038-mccarthy-has-little-room-to-maneuver-in-debt-
ceiling-talks/

Missing JSON key: Index: 26 ; source: The Hill ; URL: https://thehill.com/homenews/senate/
4016038-mccarthy-has-little-room-to-maneuver-in-debt-ceiling-talks/
.https://thehill.com/blogs/blog-briefing-room/4012827-does-god-exist-only-half-of-american
s-say-a-definite-yes/

Missing JSON key: Index: 27 ; source: The Hill ; URL: https://thehill.com/blogs/blog-brief
ing-room/4012827-does-god-exist-only-half-of-americans-say-a-definite-yes/
.https://thehill.com/homenews/administration/4015863-debt-ceiling-yellen-says-treasury-to-
run-out-of-funds-by-early-june-and-possibly-june-1/

Missing JSON key: Index: 28 ; source: The Hill ; URL: https://thehill.com/homenews/adminis
tration/4015863-debt-ceiling-yellen-says-treasury-to-run-out-of-funds-by-early-june-and-po
ssibly-june-1/
.https://thehill.com/policy/defense/4015404-anti-putin-group-claims-it-has-liberated-town-
inside-russias-belgorod-region/

Missing JSON key: Index: 29 ; source: The Hill ; URL: https://thehill.com/policy/defense/4
015404-anti-putin-group-claims-it-has-liberated-town-inside-russias-belgorod-region/
.https://thehill.com/business/4015711-health-insurance-for-600000-americans-at-stake-in-de

bt-ceiling-debate/

Missing JSON key: Index: 30 ; source: The Hill ; URL: https://thehill.com/business/4015711
-health-insurance-for-600000-americans-at-stake-in-debt-ceiling-debate/
.https://thehill.com/business/budget/4016872-house-republicans-cancel-markups-for-spending
-bills-amid-debt-ceiling-talks/

Missing JSON key: Index: 31 ; source: The Hill ; URL: https://thehill.com/business/budget/
4016872-house-republicans-cancel-markups-for-spending-bills-amid-debt-ceiling-talks/
.https://thehill.com/homenews/senate/4017344-rick-scott-issues-travel-advisory-for-sociali
sts-warning-florida-is-openly-hostile-to-them/

Missing JSON key: Index: 32 ; source: The Hill ; URL: https://thehill.com/homenews/senate/
4017344-rick-scott-issues-travel-advisory-for-socialists-warning-florida-is-openly-hostile
-to-them/
.https://thehill.com/homenews/media/4016941-trump-attacks-foxs-laura-ingraham-over-hit-pie
ce-on-his-poll-numbers/

Missing JSON key: Index: 33 ; source: The Hill ; URL: https://thehill.com/homenews/media/4
016941-trump-attacks-foxs-laura-ingraham-over-hit-piece-on-his-poll-numbers/
.https://thehill.com/homenews/house/4016326-jeffries-suggests-hed-support-spending-freeze/

Missing JSON key: Index: 34 ; source: The Hill ; URL: https://thehill.com/homenews/house/4
016326-jeffries-suggests-hed-support-spending-freeze/
.https://thehill.com/homenews/house/4016098-mccarthy-end-game-on-debt-ceiling-begins-to-co
me-into-focus/

Missing JSON key: Index: 35 ; source: The Hill ; URL: https://thehill.com/homenews/house/4
016098-mccarthy-end-game-on-debt-ceiling-begins-to-come-into-focus/
.https://thehill.com/business/4017263-house-gop-presses-fhfa-chief-on-new-mortgage-fees/

Missing JSON key: Index: 36 ; source: The Hill ; URL: https://thehill.com/business/4017263
-house-gop-presses-fhfa-chief-on-new-mortgage-fees/
.https://thehill.com/business/4015711-health-insurance-for-600000-americans-at-stake-in-de
bt-ceiling-debate/

Missing JSON key: Index: 37 ; source: The Hill ; URL: https://thehill.com/business/4015711
-health-insurance-for-600000-americans-at-stake-in-debt-ceiling-debate/
.https://thehill.com/business/4016240-mulvaney-14th-amendment-doesnt-really-solve-the-prob
lem/

Missing JSON key: Index: 38 ; source: The Hill ; URL: https://thehill.com/business/4016240
-mulvaney-14th-amendment-doesnt-really-solve-the-problem/
.https://thehill.com/policy/national-security/4017839-trump-faces-intensifying-legal-probl
ems/

Missing JSON key: Index: 39 ; source: The Hill ; URL: https://thehill.com/policy/national-
security/4017839-trump-faces-intensifying-legal-problems/
.https://thehill.com/policy/defense/4018797-who-are-the-anti-putin-groups-that-marched-int
o-russian-territory/

Missing JSON key: Index: 40 ; source: The Hill ; URL: https://thehill.com/policy/defense/4
018797-who-are-the-anti-putin-groups-that-marched-into-russian-territory/
.https://thehill.com/homenews/4018653-newsom-knocks-target-ceo-for-pulling-lgbtq-merchandi
se-from-stores/

Missing JSON key: Index: 41 ; source: The Hill ; URL: https://thehill.com/homenews/4018653
-newsom-knocks-target-ceo-for-pulling-lgbtq-merchandise-from-stores/
.https://thehill.com/policy/healthcare/4017827-chicago-mpox-outbreak-raises-alarm-over-sum
mer-spread/

Missing JSON key: Index: 42 ; source: The Hill ; URL: https://thehill.com/policy/healthcar
e/4017827-chicago-mpox-outbreak-raises-alarm-over-summer-spread/
.https://thehill.com/homenews/house/4019483-democrats-unanimously-back-debt-ceiling-discha
rge-petition/

Missing JSON key: Index: 43 ; source: The Hill ; URL: https://thehill.com/homenews/house/4019483-democrats-unanimously-back-debt-ceiling-discharge-petition/ .https://thehill.com/blogs/blog-briefing-room/4018452-6-in-10-say-spending-cuts-should-accompany-increase-in-debt-ceiling-poll/

Missing JSON key: Index: 44 ; source: The Hill ; URL: https://thehill.com/blogs/blog-briefing-room/4018452-6-in-10-say-spending-cuts-should-accompany-increase-in-debt-ceiling-poll/ .https://thehill.com/business/4016740-how-us-farmland-became-a-battleground-in-the-fight-against-china/

Missing JSON key: Index: 45 ; source: The Hill ; URL: https://thehill.com/business/4016740-how-us-farmland-became-a-battleground-in-the-fight-against-china/ .https://thehill.com/homenews/4019168-greene-says-no-one-is-concerned-about-debt-default-in-republican-conference/

Missing JSON key: Index: 46 ; source: The Hill ; URL: https://thehill.com/homenews/4019168-greene-says-no-one-is-concerned-about-debt-default-in-republican-conference/ .https://thehill.com/homenews/administration/4019080-yellen-says-its-almost-certain-debt-ceiling-deadline-would-be-in-early-june/

Missing JSON key: Index: 47 ; source: The Hill ; URL: https://thehill.com/homenews/administration/4019080-yellen-says-its-almost-certain-debt-ceiling-deadline-would-be-in-early-june/ .https://thehill.com/homenews/house/4019823-democrats-erupt-in-laughter-after-greene-calls-for-decorum-in-house/

Missing JSON key: Index: 48 ; source: The Hill ; URL: https://thehill.com/homenews/house/4019823-democrats-erupt-in-laughter-after-greene-calls-for-decorum-in-house/ .https://thehill.com/homenews/4019017-haley-says-shed-sign-abortion-ban-but-more-republicans-are-needed-in-congress/

Missing JSON key: Index: 49 ; source: The Hill ; URL: https://thehill.com/homenews/4019017-haley-says-shed-sign-abortion-ban-but-more-republicans-are-needed-in-congress/ .https://thehill.com/policy/national-security/4017839-trump-faces-intensifying-legal-problems/

Missing JSON key: Index: 50 ; source: The Hill ; URL: https://thehill.com/policy/national-security/4017839-trump-faces-intensifying-legal-problems/ .https://thehill.com/homenews/campaign/4018939-7-in-10-independents-concerned-about-bidens-mental-fitness-survey/

Missing JSON key: Index: 51 ; source: The Hill ; URL: https://thehill.com/homenews/campaign/4018939-7-in-10-independents-concerned-about-bidens-mental-fitness-survey/ .https://thehill.com/homenews/house/4019823-democrats-erupt-in-laughter-after-greene-calls-for-decorum-in-house/

Missing JSON key: Index: 52 ; source: The Hill ; URL: https://thehill.com/homenews/house/4019823-democrats-erupt-in-laughter-after-greene-calls-for-decorum-in-house/ .https://thehill.com/business/4018837-moodys-chief-economist-everyone-is-going-to-get-hurt-if-us-defaults-on-debt/

Missing JSON key: Index: 53 ; source: The Hill ; URL: https://thehill.com/business/4018837-moodys-chief-economist-everyone-is-going-to-get-hurt-if-us-defaults-on-debt/ .https://thehill.com/business/4016740-how-us-farmland-became-a-battleground-in-the-fight-against-china/

Missing JSON key: Index: 54 ; source: The Hill ; URL: https://thehill.com/business/4016740-how-us-farmland-became-a-battleground-in-the-fight-against-china/ .https://thehill.com/policy/national-security/4017839-trump-faces-intensifying-legal-problems/

Missing JSON key: Index: 55 ; source: The Hill ; URL: https://thehill.com/policy/national-security/4017839-trump-faces-intensifying-legal-problems/ .https://thehill.com/homenews/campaign/4020983-desantis-says-hell-consider-pardoning-jan-6-defendants-including-trump/

Missing JSON key: Index: 56 ; source: The Hill ; URL: https://thehill.com/homenews/campaign/4020983-desantis-says-hell-consider-pardoning-jan-6-defendants-including-trump/
.https://thehill.com/regulation/court-battles/4020888-kavanaugh-joins-supreme-court-liberals-in-disagreeing-with-new-wetlands-test/

Missing JSON key: Index: 57 ; source: The Hill ; URL: https://thehill.com/regulation/court-battles/4020888-kavanaugh-joins-supreme-court-liberals-in-disagreeing-with-new-wetlands-test/
.https://thehill.com/homenews/senate/4019680-democrats-seek-unlikely-debt-ceiling-savior-mitch-mcconnell/

Missing JSON key: Index: 58 ; source: The Hill ; URL: https://thehill.com/homenews/senate/4019680-democrats-seek-unlikely-debt-ceiling-savior-mitch-mcconnell/
.https://thehill.com/latino/4020967-mexican-president-tells-florida-hispanics-dont-give-one-single-vote-to-desantis/

Missing JSON key: Index: 59 ; source: The Hill ; URL: https://thehill.com/latino/4020967-mexican-president-tells-florida-hispanics-dont-give-one-single-vote-to-desantis/
.https://thehill.com/homenews/house/4021122-house-leaves-town-with-no-debt-limit-deal/

Missing JSON key: Index: 60 ; source: The Hill ; URL: https://thehill.com/homenews/house/4021122-house-leaves-town-with-no-debt-limit-deal/
.https://thehill.com/homenews/house/4021057-35-house-conservatives-pitch-irs-and-covid-19-funding-repeal-to-push-back-debt-x-date/

Missing JSON key: Index: 61 ; source: The Hill ; URL: https://thehill.com/homenews/house/4021057-35-house-conservatives-pitch-irs-and-covid-19-funding-repeal-to-push-back-debt-x-date/
.https://thehill.com/business/4019217-gop-attack-on-irs-funding-runs-counter-to-deficit-reduction-effort/

Missing JSON key: Index: 62 ; source: The Hill ; URL: https://thehill.com/business/4019217-gop-attack-on-irs-funding-runs-counter-to-deficit-reduction-effort/
.https://thehill.com/opinion/national-security/4018271-is-iran-unlocking-the-gates-to-armageddon/

Missing JSON key: Index: 63 ; source: The Hill ; URL: https://thehill.com/opinion/national-security/4018271-is-iran-unlocking-the-gates-to-armageddon/
.https://thehill.com/business/4009135-senators-fight-over-source-of-us-deficit-as-default-looms/

Missing JSON key: Index: 64 ; source: The Hill ; URL: https://thehill.com/business/4009135-senators-fight-over-source-of-us-deficit-as-default-looms/
.https://thehill.com/homenews/campaign/4008113-desantis-sees-both-endorsed-candidates-lose/

Missing JSON key: Index: 65 ; source: The Hill ; URL: https://thehill.com/homenews/campaign/4008113-desantis-sees-both-endorsed-candidates-lose/
.https://thehill.com/homenews/campaign/4008323-trump-taunts-desantis-over-crafts-loss-in-kentucky-gop-race-rons-magic-is-gone/

Missing JSON key: Index: 66 ; source: The Hill ; URL: https://thehill.com/homenews/campaign/4008323-trump-taunts-desantis-over-crafts-loss-in-kentucky-gop-race-rons-magic-is-gone/
.https://thehill.com/business/4009135-senators-fight-over-source-of-us-deficit-as-default-looms/

Missing JSON key: Index: 67 ; source: The Hill ; URL: https://thehill.com/business/4009135-senators-fight-over-source-of-us-deficit-as-default-looms/
.https://thehill.com/blogs/blog-briefing-room/4008216-un-sees-two-in-three-chance-world-will-temporarily-reach-temperature-threshold-soon/

Missing JSON key: Index: 68 ; source: The Hill ; URL: https://thehill.com/blogs/blog-briefing-room/4008216-un-sees-two-in-three-chance-world-will-temporarily-reach-temperature-threshold-soon/
.Start Time = 2023-06-17

```
End Time = 2023-06-17
Elapsed Time = 0:00:00
```

Check Data:

In [ ]:
```
slct_tbl_full_df02.isnull().sum()
```

```
Source             0
Author             0
Title              0
URL                0
date               0
content            0
article_parsed     0
dtype: int64
```

In [ ]:
```
api_record_df_latest=slct_tbl_full_df02.copy()
```

Make sure Word count for parsed article is correct:

In [ ]:
```
api_record_df_latest['word_count'] = api_record_df_latest['article_parsed'].apply(lambda
```

Only include rows with a word_count greater than 1:

In [ ]:
```
api_record_df_latest = api_record_df_latest[api_record_df_latest['word_count']>1]
```

Save Data to CSV:

In [ ]:
```
api_record_df_latest.to_csv("News_API_TheHill_wordcount3.csv", sep=',')
```