

```
In [1]: import pandas as pd
import numpy as np
import requests
import urllib
import urllib.request
from urllib.request import Request, urlopen
from bs4 import BeautifulSoup
import os
import csv
from http.cookiejar import CookieJar
```

## West Coast

### LA Times Articles

```
In [2]: news1 = 'https://www.latimes.com/opinion/story/2022-09-17/electric-vehicles-climate-ch
news2 = 'https://www.latimes.com/opinion/story/2022-09-15/california-electric-vehicles
news3 = 'https://www.latimes.com/opinion/story/2023-04-12/editorial-the-epa-proposes-a
news4 = 'https://www.latimes.com/politics/story/2021-07-21/californias-electric-car-re
news5 = 'https://www.latimes.com/business/story/2023-11-01/electric-cars-now-make-up-a
news6 = 'https://www.latimes.com/politics/story/2022-12-12/california-electric-cars-ev

la_times = [news1, news2, news3, news4, news5, news6]
```

```
In [3]: def get_la_times_article(url):
    req = Request(url)
    cj = CookieJar()
    opener = urllib.request.build_opener(urllib.request.HTTPCookieProcessor(cj))
    response = opener.open(req)
    page_soup = BeautifulSoup(response, 'html.parser')
    title = page_soup.find('h1', class_='headline').get_text()
    date = page_soup.find('time', class_='published-date').get_text()

    article = []
    paragraphs = page_soup.find('article', class_='story').find_all('p')
    for p in paragraphs:
        article.append(p.get_text())
    news_source = url.split('/')[2].split('.')[1]
    return title, article, date, news_source
```

```
In [4]: def create_dataset(file_name, dir_path, news_source, news_fetcher):
    os.makedirs(dir_path, exist_ok=True)
    csv_filename = os.path.join(dir_path, file_name)

    with open(csv_filename, 'w', newline='', encoding='utf-8') as file:
        writer = csv.writer(file)
        writer.writerow(['title', 'article', 'date', 'news_source'])

        for news in news_source:
            title, article, date, news_source = news_fetcher(news)
            writer.writerow([title, article, date, news_source])
```

```
In [5]: dir_path = 'news/west-coast/la'
create_dataset('articles.csv', dir_path, la_times, get_la_times_article)

file_name = os.path.join(dir_path, 'articles.csv')
df = pd.read_csv(file_name)

df['region'] = 'west-coast'
df.to_csv(file_name, index=False)
```

```
In [6]: df_la = pd.read_csv(file_name)
df_la
```

```
Out[6]:
```

	title	article	date	news_source	region
0	Commentary: Driving an EV does not make you p...	['When I started driving an electric vehicle i...	Sept. 17, 2022 3:10 AM PT	latimes	west-coast
1	Op-Ed: Think bigger. Switching to electric ca...	['It might feel like the easy solution — just ...	Sept. 15, 2022 3 AM PT	latimes	west-coast
2	Editorial: EPA wants to speed up EV switch. G...	['The Biden administration just proposed hitti...	April 12, 2023 3:57 PM PT	latimes	west-coast
3	California's electric car revolution, designe...	['The precious cargo on the ship docked in San...	July 21, 2021 3 AM PT	latimes	west-coast
4	Electric cars now make up a fifth of Californ...	['One out of every 5 cars sold in California i...	Nov. 1, 2023 2:46 PM PT	latimes	west-coast
5	Can California's electric-vehicle push overco...	['This industrial city an hour north of Indian...	Dec. 12, 2022 3 AM PT	latimes	west-coast

#### San Francisco Chronicles Articles

```
In [7]: news1 = 'https://web.archive.org/web/20190323184608/https://www.sfchronicle.com/bayare
news2 = 'https://web.archive.org/web/20190716185842/https://www.sfchronicle.com/bayare
news3 = 'https://web.archive.org/web/20230320181546/https://www.sfchronicle.com/bayare
news4 = 'https://web.archive.org/web/20230129201621/https://www.sfchronicle.com/bayare
news5 = 'https://web.archive.org/web/20230401195138/https://www.sfchronicle.com/opinio

sf_chronicles_news = [news1, news2, news3, news4, news5]

def get_sf_news(url):
    text = ''
    response = requests.get(url)
    response.raise_for_status()

    soup = BeautifulSoup(response.text, 'html.parser')
    paragraphs = soup.find_all('p')
    title = soup.find('h1', class_='header-title')
    if not title or title.text.strip() == '':
        title = soup.find('h1', class_='articleHeader--headline')
    title = title.text.strip() if title else None
    text = ' '.join(p.get_text(strip=True) for p in paragraphs)
    published_date = soup.find('time').text
    news_source = url.split('/')[7].split('.')[1]
```

```
return title, text, published_date, news_source
```

```
In [8]: dir_path = 'news/west-coast/sf/'

create_dataset('articles.csv', dir_path, sf_chronicles_news, get_sf_news)

file_name = os.path.join(dir_path, 'articles.csv')
df = pd.read_csv(file_name)

df['region'] = 'west-coast'
df.to_csv(file_name, index=False)
```

```
In [9]: df_sf = pd.read_csv('news/west-coast/sf/articles.csv')
df_sf
```

```
Out[9]:
```

	title	article	date	news_source	region
0	Environmental groups want SF to nudge Uber, Ly...	Environmental advocates want San Francisco to ...	\n March 21, 2019\n	sfchronicle	west-coast
1	To cut carbon footprint, SF moves to eliminate...	San Francisco officials want to shrink the cit...	\n July 16, 2019\n	sfchronicle	west-coast
2	With new grant program, Bay Area spreads elect...	This is a carousel. Use Next and Previous butt...	Dec. 7, 2019	sfchronicle	west-coast
3	Berkeley considers banning the sale of gas-pow...	This is a carousel. Use Next and Previous butt...	Jan. 11, 2021	sfchronicle	west-coast
4	How electric cars can keep California's power ...	When we're not driving them, electric cars, tr...	July 28, 2022	sfchronicle	west-coast

## Washington State Wire

```
In [10]: news1 = 'https://washingtonstatewire.com/cities-awarded-funding-for-electric-vehicle-c
news2 = 'https://washingtonstatewire.com/electric-vehicle-tax-break-extension-faces-st
news3 = 'https://washingtonstatewire.com/op-ed-now-is-the-time-to-electrify-public-veh
news4 = 'https://washingtonstatewire.com/washington-state-passes-law-establishing-2030

washington_state_wire = [news1, news2, news3, news4]
```

```
In [11]: def get_washington_news_article(url):
    response = requests.get(url)
    response.raise_for_status()
    soup = BeautifulSoup(response.text, 'html.parser')

    title = soup.find('h1', class_='story-title').get_text(strip=True)
    date = soup.find('span', class_='story-date').get_text(strip=True)

    article_content = soup.find('div', class_='story-content').get_text(strip=True)
    news_source = url.split('/')[2].split('.')[0]
    if article_content:
        return title, article_content, date, news_source
    else:
        print("couldnt find the article content")
```

```
In [12]: def create_dataset(file_name, dir_path, news_source, news_fetcher):
os.makedirs(dir_path, exist_ok=True)
csv_filename = os.path.join(dir_path, file_name)

with open(csv_filename, 'w', newline='', encoding='utf-8') as file:
    writer = csv.writer(file)
    writer.writerow(['title', 'article', 'date', 'news_source'])

    for news in news_source:
        title, article, date, news_source = news_fetcher(news)
        writer.writerow([title, article, date, news_source])
```

```
In [13]: dir_path = 'news/west-coast/ws/'
create_dataset('articles.csv', dir_path, washington_state_wire, get_washington_news_ar

file_name = os.path.join(dir_path, 'articles.csv')
df = pd.read_csv(file_name)

df['region'] = 'west-coast'
df.to_csv(file_name, index=False)
```

```
In [14]: df_wsw = pd.read_csv(file_name)
df_wsw.head()
```

```
Out[14]:
```

	title	article	date	news_source	region
0	Cities awarded funding for electric vehicle ch...	The Washington State Department of Commerce (C...	December 28, 2020	washingtonstatewire	west-coast
1	Electric Vehicle Tax Break Extension Faces Sti...	Consider the political oddity that is the elec...	January 21, 2015	washingtonstatewire	west-coast
2	Op-ed: Now is the Time to Electrify Public Veh...	Matthew Metz is the founder and co-executive d...	June 5, 2018	washingtonstatewire	west-coast
3	Washington State passes law establishing 2030 ...	The Washington State legislature has passed gr...	April 15, 2021	washingtonstatewire	west-coast

## Central

### Chicago Sun Times

```
In [15]: from datetime import datetime
import re
```

```
In [16]: news1 = 'https://chicago.suntimes.com/2021/4/5/22368102/electric-cars-vehicles-chargin
news2 = 'https://chicago.suntimes.com/politics/2021/11/16/22785414/illinois-electric-v
news3 = 'https://chicago.suntimes.com/2021/4/4/22362953/electric-vehicles-volkswagen-t
news4 = 'https://chicago.suntimes.com/2023/7/17/23795105/electric-vehicle-manufacturin

cst_news = [news1, news2, news3, news4]

def get_cst_news(url):
    response = requests.get(url)
```

```

response.raise_for_status()
soup = BeautifulSoup(response.text, 'html.parser')
article_content = soup.find('div', class_='RichTextBody')
if article_content:
    text = article_content.get_text()
else:
    print("couldnt find article content")
title = soup.find('h1', class_='Page-headline').get_text(strip=True)
date = soup.find('div', class_='Page-datePublished').find('bsp-timestamp')
date = re.findall(r'\d+', str(date))
date = datetime.fromtimestamp(float(date[0])/1000)
news_source = '-'.join(url.split('/')[2].split('.')[0:2])

return title, text, date, news_source

dir_path = 'news/central/chicago'

create_dataset('articles.csv', dir_path, cst_news, get_cst_news)

file_name = os.path.join(dir_path, 'articles.csv')
df_cst = pd.read_csv(file_name)

df_cst['region'] = 'central'
df_cst.to_csv(file_name, index=False)

```

In [17]: df\_cst

Out[17]:

	title	article	date	news_source	region
0	Electric vehicle charging stations could be bo...	\nI have owned and driven an electric vehicle ...	2021-04-05 12:30:27	chicago-suntimes	central
1	Pritzker signs electric vehicle legislation, a...	\nGov. J.B. Pritzker on Tuesday signed legisla...	2021-11-16 12:03:13	chicago-suntimes	central
2	Illinois is sitting on millions that could be ...	\nIllinois has an electric vehicle chicken-and...	2021-04-04 15:52:24	chicago-suntimes	central
3	Bring electric vehicle manufacturing to Illino...	\nAs some corporate headquarters decamp for ot...	2023-07-17 07:00:00	chicago-suntimes	central

## Minneapolis Star Tribune

```

In [18]: news1 = 'https://web.archive.org/web/20230606091720/https://www.startribune.com/evs-ar
news2 = 'https://web.archive.org/web/20230606091839/https://www.startribune.com/minnes
news3 = 'https://web.archive.org/web/20231022194355/https://www.startribune.com/farm-r
news4 = 'https://web.archive.org/web/20231017085213/https://www.startribune.com/minnes

mst_news = [news1, news2, news3, news4]

def get_mst_news(url):
    text = ''
    response = requests.get(url)
    response.raise_for_status()

    soup = BeautifulSoup(response.text, 'html.parser')
    title = soup.find('h1', class_='article-headline').get_text(strip=True)
    published_date = soup.find('div', class_='article-dateline').get_text(strip=True)
    text = soup.find('div', class_='article-body').find_all('p')

```

```

text = ' '.join(p.get_text(strip=True) for p in text)
news_source = 'minneapolis-star-tribune'
return title, text, published_date, news_source

dir_path = 'news/central/minneapolis'

create_dataset('articles.csv', dir_path, mst_news, get_mst_news)

file_name = os.path.join(dir_path, 'articles.csv')
df_mst = pd.read_csv(file_name)

df_mst['region'] = 'central'
df_mst.to_csv(file_name, index=False)

```

In [19]: df\_mst

	title	article	date	news_source	region
0	EVs are only part of the solution	Arresting climate change requires, among other...	May 21, 2023 — 6:00pm	minneapolis-star-tribune	central
1	Energy secretary announces electric vehicle gr...	Two Minnesota organizations, including a Nativ...	May 19, 2023 — 6:57pm	minneapolis-star-tribune	central
2	This farm reporter roams rural Minnesota by EV...	The first time I knew our new Nissan Leaf was ...	October 14, 2023 — 1:45pm	minneapolis-star-tribune	central
3	Minnesota auto dealers lose court fight with s...	The Minnesota Court of Appeals has upheld the ...	January 30, 2023 — 6:18pm	minneapolis-star-tribune	central

## Houston Chronicle

In [20]: `import requests`  
`import random`

In [21]:

```

news1 = 'https://web.archive.org/web/20230128180351/https://www.houstonchronicle.com/1'
news2 = 'https://web.archive.org/web/20230128203128/https://www.houstonchronicle.com/o'
news3 = 'https://web.archive.org/web/20230703224808/https://www.houstonchronicle.com/b'

hc_news = [news1, news2, news3]

def get_hc_news(url):
    text = ''
    response = requests.get(url)
    response.raise_for_status()

    soup = BeautifulSoup(response.text, 'html.parser')
    paragraphs = soup.find_all('p')
    title = soup.find('h1', class_='header-title')
    if not title or title.text.strip() == '':
        title = soup.find('h1', class_='articleHeader--headline')
    title = title.text.strip() if title else None
    text = ' '.join(p.get_text(strip=True) for p in paragraphs)
    published_date = soup.find('time').text
    news_source = url.split('/')[7].split('.')[1]

    return title, text, published_date, news_source

```

```

dir_path = 'news/central/houston/'

create_dataset('articles.csv', dir_path, hc_news, get_hc_news)

file_name = os.path.join(dir_path, 'articles.csv')
df_hc = pd.read_csv(file_name)

df_hc['region'] = 'central'
df_hc.to_csv(file_name, index=False)

```

In [22]: df\_hc

Out[22]:

	title	article	date	news_source	region
0	How electric vehicles can help cities like Hou...	Houston Mayor Sylvester Turner was one of more ...	June 15, 2017	houstonchronicle	central
1	Opinion: Electric vehicles are no silver bullet	Regarding "Tomlinson: GM's new Hummer and othe...	Jan. 20, 2023	houstonchronicle	central
2	Nearly half of U.S. car buyers plan to go elec...	Analysts at Ernst and Young surveyed 1,500 U.S...	July 3, 2023	houstonchronicle	central

## Dallas News

In [23]:

```

news1 = 'https://www.dallasnews.com/business/autos/2022/12/28/2023-tax-credits-may-boo
news2 = 'https://www.dallasnews.com/news/2022/10/08/are-electric-vehicles-the-future-h
news3 = 'https://www.dallasnews.com/business/autos/2023/04/17/only-10-electric-vehicle
news4 = 'https://www.dallasnews.com/news/politics/2023/07/03/new-texas-laws-aimed-at-s

dallas_news = [news1, news2, news3, news4]

def get_dallas_news(url):
    text = ''
    response = requests.get(url)
    response.raise_for_status() # This will raise an exception if there's a HTTP erro

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

    text = ' '.join(p.get_text(strip=True) for p in paragraphs)
    title = soup.find('span', class_='dmnc_generic-header-header-module__i2K-Y mr-7').
    date = soup.find('div', class_='text-gray-medium')
    second_p = date.find_all('p')[1].get_text(strip=True)
    news_source = url.split('/')[2].split('.')[1]

    return title, text, second_p, news_source

dir_path = 'news/central/dallas/'
create_dataset('articles.csv', dir_path, dallas_news, get_dallas_news)

file_name = os.path.join(dir_path, 'articles.csv')
df_dn = pd.read_csv(file_name)

df_dn['region'] = 'central'
df_dn.to_csv(file_name, index=False)

```

```
In [24]: df_dn.head()
```

		title	article	date	news_source	region
0	2023 tax credits may boost the appeal of elect...	businessAutos	ByThe Associated Press 5:38 AM o...	5:38 AM on Dec 28, 2022 CST	dallasnews	central
1	Are electric vehicles the future? Here's what ...	News ByNoor Adatia	8:00 AM on Oct 8, 2022 CDT ...	8:00 AM on Oct 8, 2022 CDT	dallasnews	central
2	Only 10 electric vehicles will qualify for \$7,...	businessAutos	ByThe Associated Press 1:38 PM o...	1:38 PM on Apr 17, 2023 CDT	dallasnews	central
3	New Texas laws aimed at sharp rise in electric...	newsPolitics	ByAarón Torres 6:00 AM on Jul 3, ...	6:00 AM on Jul 3, 2023 CDT	dallasnews	central

## East Coast

### Washington Post

```
In [25]: news1 = 'https://www.washingtonpost.com/climate-environment/2023/09/19/electric-cars-b
news2 = 'https://www.washingtonpost.com/business/energy/2023/05/17/electric-vehicles-a
news3 = 'https://www.washingtonpost.com/climate-environment/2023/04/11/electric-vehicl
news4 = 'https://www.washingtonpost.com/climate-solutions/2023/11/09/car-dealerships-e
news5 = 'https://www.washingtonpost.com/climate-environment/2023/06/13/electric-vehicl

wp_news = [news1, news2, news3, news4, news5]

def get_wp_news(url):
    text = ''
    response = requests.get(url)
    response.raise_for_status()

    soup = BeautifulSoup(response.text, 'html.parser')
    paragraphs = soup.find_all('p')
    title = soup.find('h1').get_text()
    text = ' '.join(p.get_text(strip=True) for p in paragraphs)
    published_date = soup.find('div', attrs={'data-testid': 'timestamp'}).get_text()
    news_source = 'washington-post'

    return title, text, published_date, news_source

dir_path = 'news/east-coast/dc'

create_dataset('articles.csv', dir_path, wp_news, get_wp_news)

file_name = os.path.join(dir_path, 'articles.csv')
df_wp = pd.read_csv(file_name)

df_wp['region'] = 'east-coast'
df_wp.to_csv(file_name, index=False)
```

```
In [26]: df_wp
```



Out[26]:

	title	article	date	news_source	region
0	Are electric cars really better for the enviro...	Electric vehicles, you might have heard, have ...	September 19, 2023 at 6:30 a.m. EDT	washington-post	east-coast
1	Electric Vehicles Alone Can't Solve Climate Ch...	Arresting climate change requires, among other...	May 17, 2023 at 8:04 a.m. EDT	washington-post	east-coast
2	Buy an electric vehicle now or wait? Here's ho...	Thecoronaviruspandemic saw a virtually unprece...	April 11, 2023 at 6:30 a.m. EDT	washington-post	east-coast
3	Electric vehicles are hitting a road block: Ca...	As news started coming out aboutelectric cars...	Updated November 9, 2023 at 12:10 p.m. EST Pub...	washington-post	east-coast
4	Why EVs won't crash the electric grid — and more	Climate Coach readers are thinking hard about ...	June 13, 2023 at 6:30 a.m. EDT	washington-post	east-coast

## New York Times

In [27]:

```
news1 = 'https://web.archive.org/web/20231103120806/https://www.nytimes.com/2021/03/02'
news2 = 'https://web.archive.org/web/20231104012045/https://www.nytimes.com/interactiv'
news3 = 'https://web.archive.org/web/20230920204337/https://www.nytimes.com/interactiv'
news4 = 'https://web.archive.org/web/20231030153715/https://www.nytimes.com/2022/08/08'
news5 = 'https://web.archive.org/web/20231110204200/https://www.nytimes.com/2023/11/07'

nyt_news = [news1, news2, news3, news4, news5]

def get_nyt_news(url):
    text = ''
    response = requests.get(url)
    response.raise_for_status()

    soup = BeautifulSoup(response.text, 'html.parser')
    paragraphs = soup.find_all('p')
    title = soup.find('h1', attrs={'data-testid': 'headline'}).get_text()
    text = ' '.join(p.get_text(strip=True) for p in paragraphs)
    published_date = soup.find('time').text
    news_source = 'new-york-times'

    return title, text, published_date, news_source

dir_path = 'news/east-coast/nyc'

create_dataset('articles.csv', dir_path, nyt_news, get_nyt_news)

file_name = os.path.join(dir_path, 'articles.csv')
df_nyt = pd.read_csv(file_name)

df_nyt['region'] = 'east-coast'
df_nyt.to_csv(file_name, index=False)
```

In [28]:

```
df_nyt
```

Out[28]:

	title	article	date	news_source	region
0	How Green Are Electric Vehicles?	Advertisement Supported by In short: Very gree...	Published March 2, 2021Updated June 23, 2023	new-york-times	east-coast
1	Electric Cars Are Better for the Planet – and ...	Advertisement ByVeronica PenneyJan. 15, 2021 E...	Jan. 15, 2021	new-york-times	east-coast
2	How Electrifying Everything Became a Key Clima...	Advertisement ByNadja PopovichandBrad PlumerAp...	April 14, 2023	new-york-times	east-coast
3	Electric Cars Too Costly for Many, Even With A...	Advertisement Supported by Battery-powered veh...	Published Aug. 8, 2022Updated Aug. 9, 2022	new-york-times	east-coast
4	Automakers Delay Electric Vehicle Spending as ...	Advertisement Supported by Growth is brisk but...	Nov. 7, 2023	new-york-times	east-coast

## Baltimore Sun

```
In [29]: news1 = 'https://www.baltimoresun.com/business/ct-biz-electric-vehicles-tax-credits-ap
news2 = 'https://www.baltimoresun.com/opinion/columnists/dan-rodricks/bs-ed-rodricks-0
news3 = 'https://www.baltimoresun.com/business/ct-biz-rivian-georgia-meeting-20220420-
news4 = 'https://www.baltimoresun.com/news/environment/bs-md-hogan-administration-cali
news5 = 'https://www.baltimoresun.com/news/environment/bs-md-california-electric-vehic
```

```
maryland_news = [news1, news2, news3, news4, news5]
```

```
def get_maryland_news(url):
    text = ''
    response = requests.get(url)
    response.raise_for_status() # This will raise an exception if there's a HTTP error

    soup = BeautifulSoup(response.text, 'html.parser')
    title = soup.find('h1', class_='headline').get_text(strip=True)

    paragraphs = soup.find('article', class_='article-body-wrapper-custom')
    text = ' '.join(p.get_text(strip=True) for p in paragraphs.find_all('p'))

    date = soup.find('time')
    datetime_value = date['datetime']

    news_source = url.split('/')[2].split('.')[1]

    return title, text, datetime_value, news_source
```

```
In [30]: dir_path = 'news/east-coast/md/'
create_dataset('articles.csv', dir_path, maryland_news, get_maryland_news)

file_name = os.path.join(dir_path, 'articles.csv')
df_md = pd.read_csv(file_name)

df_md['region'] = 'east-coast'
df_md.to_csv(file_name, index=False)
```

In [31]: df\_md

Out[31]:

	title	article	date	news_source	region
0	Most electric vehicles won't qualify for feder...	DETROIT — A tax credit of up to \$7,500 could b...	2022-08-09 10:54:00	baltimoresun	east-coast
1	Dan Rodricks: All charged up for an electric c...	If you own an electric car or truck, or drive ...	2022-08-16 11:12:00	baltimoresun	east-coast
2	Rivian electric car plant blasted by foes at G...	MONROE, Ga. — Opponents trying to derail a \$5 ...	2022-04-20 12:13:00	baltimoresun	east-coast
3	Hogan administration won't match California's ...	Maryland Gov. Larry Hogan's administration won...	2022-12-14 15:59:00	baltimoresun	east-coast
4	California electric vehicle rule takes another...	Maryland took a big step Monday toward eventua...	2023-03-13 17:52:00	baltimoresun	east-coast

In [ ]: