

Web Scraping Daily News Updates from Economic Times

This project is done by scraping the data from the “THE ECONOMICS TIMES”, the steps used to make this project is given below:-

AIM:- Data extraction and analysis of news or content from a specific web page ("The Economics Times" website).

STEP1:- IMPORTING LIBRARIES

In this step, the necessary Python libraries are imported. requests is used for making HTTP requests to fetch web pages, BeautifulSoup is used to beautify which means it is used to convert the code into news headlines which can be easily readable, and pandas is used for data manipulation. Here, a variable named URL is assigned the value of the web page's URL that you want to scrape. In this case, it's a URL from the Economic Times website.

```
In [54]: import requests
         from bs4 import BeautifulSoup
         import pandas as pd
         URL = "https://economictimes.indiatimes.com/news/newsblogs/daily-news-and-latest-updates-live-28-septemb
```

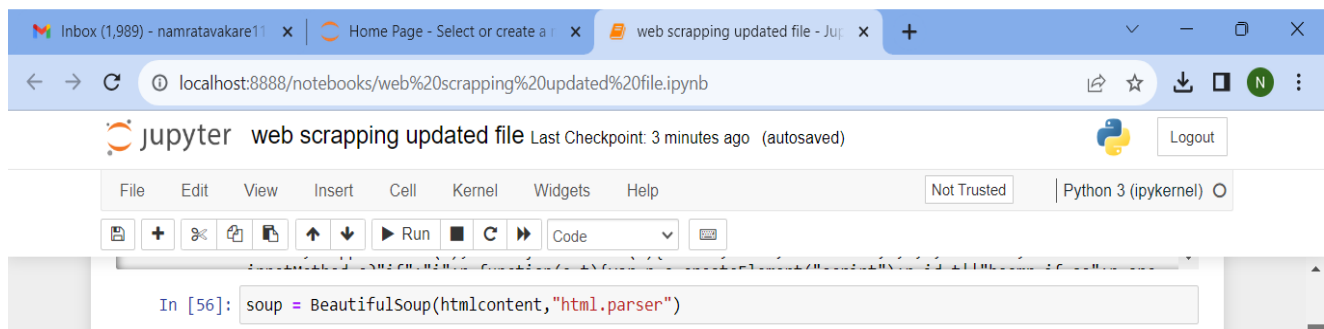
STEP2 :- Now, we have successfully fetched the HTML content of the web page and stored it in the html content variable. This content can be further processed

```
In [55]: import requests
         r = requests.get(URL)
         htmlcontent = r.content
         htmlcontent

<html><head><title>Karnataka Bandh News LIVE Updates: Holiday for all schools and colleges in Bengaluru City tomorrow - The Economic Times</title><META NAME="description" content="Karnataka Bandh News LIVE Updates: Holiday has been declared for all schools and colleges in Bengaluru City tomorrow."><meta name="lang" content="en"><meta name="keywords" content="news, latest news, live news, live updates, breaking news, news today, Asian Games 2023 LIVE Updates, Asian Games 2023 LIVE, Asian Games 2023, Asian Games"><link rel="image_src" href="https://economictimes.indiatimes.com/thumb/msid-104001555,width-600,resizemode-4,imglength-23654/news/newsblogs/daily-news-and-latest-updates-live-28-september-2023.jpg"><script>\n
    (function(){if(window.BOOMR&&(window.BOOMR.version|window.BOOMR.snippetExecuted)){return}window.BOOMR=window.BOOMR||{};window.BOOMR.snippetStart=(new Date).getTime();window.BOOMR.snippetExecuted=true;window.BOOMR.snippetVersion=14;window.BOOMR.url="//c.go-mpulse.net/boomerang/KY9J6-H7E3C-JE2Z4-GP844-RCBw6";var e=document.currentScript||document.getElementsByTagName("script")[0] a=e.parentNode s=false t=3e3;function n(){if
```

and parsed using libraries like BeautifulSoup to extract specific data or information from the web page.

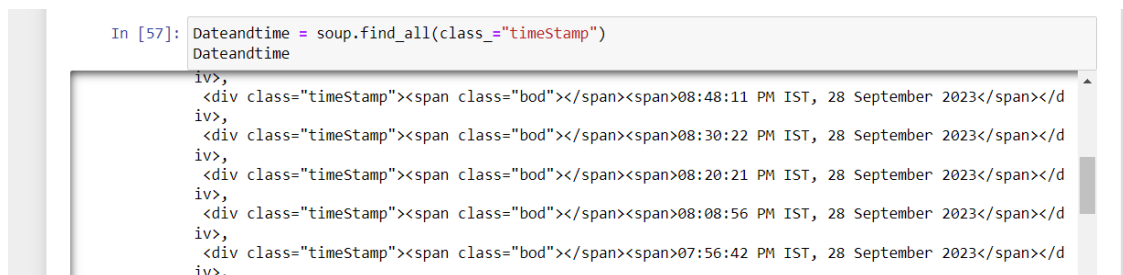
STEP 3 :- The code `soup = BeautifulSoup(html content, "html.parser")` is a crucial step in web scraping. It uses the BeautifulSoup library to parse the HTML content that we fetched earlier from the web page using the requests library.



The screenshot shows a web browser window displaying a Jupyter Notebook. The browser's address bar shows the URL `localhost:8888/notebooks/web%20scraping%20updated%20file.ipynb`. The Jupyter Notebook interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a toolbar with icons for file operations and execution, and a code editor. The code editor contains the following Python code:

```
In [56]: soup = BeautifulSoup(htmlcontent, "html.parser")
```

STEP 4 :- The code `Dateandtime = soup.find_all(class_="timeStamp")` is used to search for and extract HTML elements that have a specific CSS class name ("timeStamp") from the soup object, which represents the parsed HTML content of the web page.



The screenshot shows a Jupyter Notebook code cell with the following Python code and its output:

```
In [57]: Dateandtime = soup.find_all(class_="timeStamp")
Dateandtime
iv>,
<div class="timeStamp"><span class="bod"></span><span>08:48:11 PM IST, 28 September 2023</span></div>,
<div class="timeStamp"><span class="bod"></span><span>08:30:22 PM IST, 28 September 2023</span></div>,
<div class="timeStamp"><span class="bod"></span><span>08:20:21 PM IST, 28 September 2023</span></div>,
<div class="timeStamp"><span class="bod"></span><span>08:08:56 PM IST, 28 September 2023</span></div>,
<div class="timeStamp"><span class="bod"></span><span>07:56:42 PM IST, 28 September 2023</span></div>,
iv>,
```

STEP 5 :- The code shown below is a Python loop that iterates through the elements stored in the Dateandtime variable, extracts the text content of each element, appends it to a list named `list_Dateandtime`

```
In [58]: list_Dateandtime=[]
for i in range(0,len(Dateandtime)):
    list_Dateandtime.append(Dateandtime[i].get_text())
print(list_Dateandtime)
len(list_Dateandtime)

['12:49:16 AM IST, 29 September 2023', '12:03:11 AM IST, 29 September 2023', '11:42:36 PM IST, 28 Sep
tember 2023', '11:24:37 PM IST, 28 September 2023', '11:08:31 PM IST, 28 September 2023', '10:52:58 P
M IST, 28 September 2023', '10:37:31 PM IST, 28 September 2023', '10:13:10 PM IST, 28 September 202
3', '09:54:37 PM IST, 28 September 2023', '09:40:41 PM IST, 28 September 2023', '09:21:23 PM IST, 28
September 2023', '09:00:55 PM IST, 28 September 2023', '08:48:11 PM IST, 28 September 2023', '08:30:2
2 PM IST, 28 September 2023', '08:20:21 PM IST, 28 September 2023', '08:08:56 PM IST, 28 September 20
23', '07:56:42 PM IST, 28 September 2023', '07:45:15 PM IST, 28 September 2023', '07:42:04 PM IST, 28
September 2023', '07:27:24 PM IST, 28 September 2023', '07:17:44 PM IST, 28 September 2023', '07:09:5
2 PM IST, 28 September 2023', '06:59:27 PM IST, 28 September 2023', '06:45:06 PM IST, 28 September 20
```

STEP 6 :- The code `latest_news = soup.find_all(class_="quote")` is used to search for and extract HTML elements that have a specific class name ("quote") from the soup object, which represents the parsed HTML content of the web page.

```
In [59]: latest_news = soup.find_all(class_="quote")
latest_news

Out[59]: [<h3 class="quote">An open and productive conversation with Think Tanks in Washington DC this mornin
g. Discussed transformations underway around the world and India's growing role</h3>,
<h3 class="quote">Good to see with US Trade Representative Ambassador Katherine Tai. Spoke about our
expanding trade and economic relationship and its broader significance</h3>,
<h3 class="quote">Only the President is yet to come (to Rajasthan). I respect the Vice President (Ja
gdeep Dhankhar). He visited 5 districts yesterday. Who is coming to meet him? Local BJP leaders are c
oming to meet him. What message are they trying to convey?..</h3>,
<h3 class="quote">We are very interested in increasing Indian exports to Russia, and we see a real p
```

STEP 7 :- 1] `list_latest_news = []`: This line initializes an empty list called `list_latest_news`. This list will be used to store the text content of the HTML elements.

2] `for i in range(0, len(latest_news)):`

This is a for loop that iterates through a range of numbers from 0 to the length of the `latest_news` collection. `len(latest_news)` returns the number of elements in the `latest_news` collection.

3] `list_latest_news.append(latest_news[i].get_text()):`

Inside the loop, this line extracts the text content of each HTML element in the `latest_news` collection using the `.get_text()` method and appends it to the `list_latest_news` list.

4)print(list_latest_news): After the loop completes, this line prints the list_latest_news list, which now contains the extracted text content from the HTML elements. These would typically be news headlines, quotes, or other textual content.

```
In [60]: list_latest_news=[]
for i in range(0,len(latest_news)):
    list_latest_news.append(latest_news[i].get_text())
print(list_latest_news)
len(list_latest_news)
```

['An open and productive conversation with Think Tanks in Washington DC this morning. Discussed transformations underway around the world and India's growing role', 'Good to see with US Trade Representative Ambassador Katherine Tai. Spoke about our expanding trade and economic relationship and its broader significance', 'Only the President is yet to come (to Rajasthan). I respect the Vice President (Jagdeep Thakur). He visited 5 districts yesterday. Who is coming to meet him? Local BJP leaders are coming to meet him. What message are they trying to convey?', 'We are very interested in increasing

STEP 8 :- The code shown below is a loop that iterates through the elements in the `list_Dateandtime` list and appends them to a new list called `list_Datetime`. However, it includes a condition to ensure that the length of `list_Datetime` remains less than the length of `list_latest_news`.

```
In [61]: list_Datetime=[]
for i in list_Dateandtime:
    if len(list_Datetime)<len(list_latest_news):
        list_Datetime.append(i)
print(list_Datetime)

['12:49:16 AM IST, 29 September 2023', '12:03:11 AM IST, 29 September 2023', '11:42:36 PM IST, 28 Sep
tember 2023', '11:24:37 PM IST, 28 September 2023', '11:08:31 PM IST, 28 September 2023', '10:52:58 P
M IST, 28 September 2023', '10:37:31 PM IST, 28 September 2023', '10:13:10 PM IST, 28 September 202
3', '09:54:37 PM IST, 28 September 2023', '09:40:41 PM IST, 28 September 2023', '09:21:23 PM IST, 28
September 2023', '09:00:55 PM IST, 28 September 2023', '08:48:11 PM IST, 28 September 2023', '08:30:2
2 PM IST, 28 September 2023', '08:20:21 PM IST, 28 September 2023', '08:08:56 PM IST, 28 September 20
```

STEP 9 :- The code shown below is designed to split date and time information stored in the list Datetime list and then flatten it into a single list called datetime.

```
In [62]: datetime=[]
for i in list_Datetime:
    x=i.split(',')
    for i in x:
        datetime.append(i)
print(datetime)

['12:49:16 AM IST', ' 29 September 2023', '12:03:11 AM IST', ' 29 September 2023', '11:42:36 PM IST',
 ' 28 September 2023', '11:24:37 PM IST', ' 28 September 2023', '11:08:31 PM IST', ' 28 September 2023',
 '10:52:58 PM IST', ' 28 September 2023', '10:37:31 PM IST', ' 28 September 2023', '10:13:10 PM IST',
 ' 28 September 2023', '09:54:37 PM IST', ' 28 September 2023', '09:40:41 PM IST', ' 28 September
2023', '09:21:23 PM IST', ' 28 September 2023', '09:00:55 PM IST', ' 28 September 2023', '08:48:11 PM
IST', ' 28 September 2023', '08:30:22 PM IST', ' 28 September 2023', '08:20:21 PM IST', ' 28 September
2023', '08:09:56 PM IST', ' 28 September 2023', '07:56:43 PM IST', ' 28 September 2023', '07:45:16 PM IST',
 ' 28 September 2023', '07:34:48 PM IST', ' 28 September 2023', '07:24:19 PM IST', ' 28 September 2023',
 '07:13:50 PM IST', ' 28 September 2023', '07:03:21 PM IST', ' 28 September 2023', '06:52:52 PM IST', ' 28
September 2023', '06:42:23 PM IST', ' 28 September 2023', '06:31:54 PM IST', ' 28 September 2023', '06:21:25
PM IST', ' 28 September 2023', '06:10:56 PM IST', ' 28 September 2023', '06:00:27 PM IST', ' 28 September
2023', '05:49:58 PM IST', ' 28 September 2023', '05:39:29 PM IST', ' 28 September 2023', '05:28:59 PM IST',
 ' 28 September 2023', '05:18:30 PM IST', ' 28 September 2023', '05:08:01 PM IST', ' 28 September 2023',
 '04:57:32 PM IST', ' 28 September 2023', '04:47:03 PM IST', ' 28 September 2023', '04:36:34 PM IST', ' 28
September 2023', '04:26:05 PM IST', ' 28 September 2023', '04:15:36 PM IST', ' 28 September 2023', '04:05:07
PM IST', ' 28 September 2023', '03:54:38 PM IST', ' 28 September 2023', '03:44:09 PM IST', ' 28 September
2023', '03:33:40 PM IST', ' 28 September 2023', '03:23:11 PM IST', ' 28 September 2023', '03:12:42 PM IST',
 ' 28 September 2023', '03:02:13 PM IST', ' 28 September 2023', '02:51:44 PM IST', ' 28 September 2023',
 '02:41:15 PM IST', ' 28 September 2023', '02:30:46 PM IST', ' 28 September 2023', '02:20:17 PM IST', ' 28
September 2023', '02:09:48 PM IST', ' 28 September 2023', '01:59:19 PM IST', ' 28 September 2023', '01:48:50
PM IST', ' 28 September 2023', '01:38:21 PM IST', ' 28 September 2023', '01:27:52 PM IST', ' 28 September
2023', '01:17:23 PM IST', ' 28 September 2023', '01:06:54 PM IST', ' 28 September 2023', '00:56:25 PM IST',
 ' 28 September 2023', '00:45:56 PM IST', ' 28 September 2023', '00:35:27 PM IST', ' 28 September 2023',
 '00:24:58 PM IST', ' 28 September 2023', '00:14:29 PM IST', ' 28 September 2023', '00:04:00 PM IST', ' 28
September 2023', '23:53:31 PM IST', ' 28 September 2023', '23:43:02 PM IST', ' 28 September 2023', '23:32:33
PM IST', ' 28 September 2023', '23:22:04 PM IST', ' 28 September 2023', '23:11:35 PM IST', ' 28 September
2023', '23:01:06 PM IST', ' 28 September 2023', '22:50:37 PM IST', ' 28 September 2023', '22:40:08 PM IST',
 ' 28 September 2023', '22:29:39 PM IST', ' 28 September 2023', '22:19:10 PM IST', ' 28 September 2023',
 '22:08:41 PM IST', ' 28 September 2023', '21:58:12 PM IST', ' 28 September 2023', '21:47:43 PM IST', ' 28
September 2023', '21:37:14 PM IST', ' 28 September 2023', '21:26:45 PM IST', ' 28 September 2023', '21:16:16
PM IST', ' 28 September 2023', '21:05:47 PM IST', ' 28 September 2023', '20:55:18 PM IST', ' 28 September
2023', '20:44:49 PM IST', ' 28 September 2023', '20:34:20 PM IST', ' 28 September 2023', '20:23:51 PM IST',
 ' 28 September 2023', '20:13:22 PM IST', ' 28 September 2023', '20:02:53 PM IST', ' 28 September 2023',
 '19:52:24 PM IST', ' 28 September 2023', '19:41:55 PM IST', ' 28 September 2023', '19:31:26 PM IST', ' 28
September 2023', '19:20:57 PM IST', ' 28 September 2023', '19:10:28 PM IST', ' 28 September 2023', '19:00:00
PM IST', ' 28 September 2023', '18:49:31 PM IST', ' 28 September 2023', '18:39:02 PM IST', ' 28 September
2023', '18:28:33 PM IST', ' 28 September 2023', '18:18:04 PM IST', ' 28 September 2023', '18:07:35 PM IST',
 ' 28 September 2023', '17:57:06 PM IST', ' 28 September 2023', '17:46:37 PM IST', ' 28 September 2023',
 '17:36:08 PM IST', ' 28 September 2023', '17:25:39 PM IST', ' 28 September 2023', '17:15:10 PM IST', ' 28
September 2023', '17:04:41 PM IST', ' 28 September 2023', '16:54:12 PM IST', ' 28 September 2023', '16:43:43
PM IST', ' 28 September 2023', '16:33:14 PM IST', ' 28 September 2023', '16:22:45 PM IST', ' 28 September
2023', '16:12:16 PM IST', ' 28 September 2023', '16:01:47 PM IST', ' 28 September 2023', '15:51:18 PM IST',
 ' 28 September 2023', '15:40:49 PM IST', ' 28 September 2023', '15:30:20 PM IST', ' 28 September 2023',
 '15:19:51 PM IST', ' 28 September 2023', '15:09:22 PM IST', ' 28 September 2023', '14:58:53 PM IST', ' 28
September 2023', '14:48:24 PM IST', ' 28 September 2023', '14:37:55 PM IST', ' 28 September 2023', '14:27:26
PM IST', ' 28 September 2023', '14:16:57 PM IST', ' 28 September 2023', '14:06:28 PM IST', ' 28 September
2023', '13:55:59 PM IST', ' 28 September 2023', '13:45:30 PM IST', ' 28 September 2023', '13:35:01 PM IST',
 ' 28 September 2023', '13:24:32 PM IST', ' 28 September 2023', '13:14:03 PM IST', ' 28 September 2023',
 '13:03:34 PM IST', ' 28 September 2023', '12:53:05 PM IST', ' 28 September 2023', '12:42:36 PM IST', ' 28
September 2023', '12:32:07 PM IST', ' 28 September 2023', '12:21:38 PM IST', ' 28 September 2023', '12:11:09
PM IST', ' 28 September 2023', '12:00:40 PM IST', ' 28 September 2023', '11:50:11 PM IST', ' 28 September
2023', '11:39:42 PM IST', ' 28 September 2023', '11:29:13 PM IST', ' 28 September 2023', '11:18:44 PM IST',
 ' 28 September 2023', '11:08:15 PM IST', ' 28 September 2023', '10:57:46 PM IST', ' 28 September 2023',
 '10:47:17 PM IST', ' 28 September 2023', '10:36:48 PM IST', ' 28 September 2023', '10:26:19 PM IST', ' 28
September 2023', '10:15:50 PM IST', ' 28 September 2023', '10:05:21 PM IST', ' 28 September 2023', '09:54:52
PM IST', ' 28 September 2023', '09:44:23 PM IST', ' 28 September 2023', '09:33:54 PM IST', ' 28 September
2023', '09:23:25 PM IST', ' 28 September 2023', '09:12:56 PM IST', ' 28 September 2023', '09:02:27 PM IST',
 ' 28 September 2023', '08:51:58 PM IST', ' 28 September 2023', '08:41:29 PM IST', ' 28 September 2023',
 '08:31:00 PM IST', ' 28 September 2023', '08:20:31 PM IST', ' 28 September 2023', '08:10:02 PM IST', ' 28
September 2023', '08:00:33 PM IST', ' 28 September 2023', '07:49:54 PM IST', ' 28 September 2023', '07:39:25
PM IST', ' 28 September 2023', '07:28:56 PM IST', ' 28 September 2023', '07:18:27 PM IST', ' 28 September
2023', '07:07:58 PM IST', ' 28 September 2023', '06:57:29 PM IST', ' 28 September 2023', '06:46:50 PM IST',
 ' 28 September 2023', '06:36:21 PM IST', ' 28 September 2023', '06:25:52 PM IST', ' 28 September 2023',
 '06:15:23 PM IST', ' 28 September 2023', '06:04:54 PM IST', ' 28 September 2023', '05:54:25 PM IST', ' 28
September 2023', '05:43:56 PM IST', ' 28 September 2023', '05:33:27 PM IST', ' 28 September 2023', '05:22:58
PM IST', ' 28 September 2023', '05:12:29 PM IST', ' 28 September 2023', '05:02:00 PM IST', ' 28 September
2023', '04:51:31 PM IST', ' 28 September 2023', '04:41:02 PM IST', ' 28 September 2023', '04:30:33 PM IST',
 ' 28 September 2023', '
```

```
In [63]: date=[]  
         time=[]  
         for i in range(0,len(datetime)):  
             if i%2==0:  
                 time.append(datetime[i])  
             else:  
                 date.append(datetime[i])  
         print(date)  
         print(time)
```

[' 29 September 2023', ' 29 September 2023', ' 28 September 2023', ' 28 September 2023', ' 28 Septemb
er 2023', ' 28 September 2023', ' 28 September 2023', ' 28 September 2023', ' 28 September 2023', ' 2
8 September 2023', ' 28 September 2023', ' 28 September 2023', ' 28 September 2023', ' 28 September 2
023', ' 28 September 2023', ' 28 September 2023', ' 28 September 2023', ' 28 September 2023']

['12:49:16 AM IST', '12:03:11 AM IST', '11:42:36 PM IST', '11:24:37 PM IST', '11:08:31 PM IST', '10:5

```
In [64]: import pandas as pd
df=pd.DataFrame({'Date':date,'time':time,"Latest_News":list_latest_news})
df
```

```
Out[64]:
```

	Date	time	Latest_News
0	29 September 2023	12:49:16 AM IST	An open and productive conversation with Think...
1	29 September 2023	12:03:11 AM IST	Good to see with US Trade Representative Amba...
2	28 September 2023	11:42:36 PM IST	Only the President is yet to come (to Rajastha...
3	28 September 2023	11:24:37 PM IST	We are very interested in increasing Indian ex...
4	28 September 2023	11:08:31 PM IST	India Energy Week 2024's biggest event will be...
5	28 September 2023	10:52:58 PM IST	Women are unsafe in Madhya Pradesh. Why is Shi...
6	28 September 2023	10:37:31 PM IST	More than six months have passed but nothing h...

```
In [65]: import os
os.getcwd()

Out[65]: 'C:\\Users\\USER'
```

```
In [66]: df.to_csv("TheEconomicsTimes_Latest_News.csv", index = False)
```

STEP 14:- The code `df1 = pd.read_csv("TheEconomicsTimes_Latest_News.csv")` is used to read data from a CSV (Comma-Separated Values) file named "TheEconomicsTimes_Latest_News.csv" and create a new DataFrame called `df1`

```
In [67]: df1 = pd.read_csv("TheEconomicsTimes_Latest_News.csv")
df1
```

Out[67]:

	Date	time	Latest_News
0	29 September 2023	12:49:16 AM IST	An open and productive conversation with Think...
1	29 September 2023	12:03:11 AM IST	Good to see with US Trade Representative Amba...
2	28 September 2023	11:42:36 PM IST	Only the President is yet to come (to Rajastha...
3	28 September 2023	11:24:37 PM IST	We are very interested in increasing Indian ex...
4	28 September 2023	11:08:31 PM IST	India Energy Week 2024's biggest event will be...
5	28 September 2023	10:52:58 PM IST	Women are unsafe in Madhya Pradesh. Why is Shi...
6	28 September 2023	10:37:31 PM IST	More than six months have passed but nothing h...
7	28 September 2023	10:13:10 PM IST	When we were taking (the accused) to recreate ...
8	28 September 2023	09:54:37 PM IST	I thank the PM for providing 33% reservation t...
9	28 September 2023	09:40:41 PM IST	We won bronze this time but for the next time ...
10	28 September 2023	09:21:22 PM IST	There was a differen: EN English (United States)

CONCLUSION :- By doing the project we conclude that The data can then be easily read back into a DataFrame for further analysis or use in other data-related tasks. This type of workflow is commonly used for collecting and analyzing data from websites or external sources.

THANK YOU!

