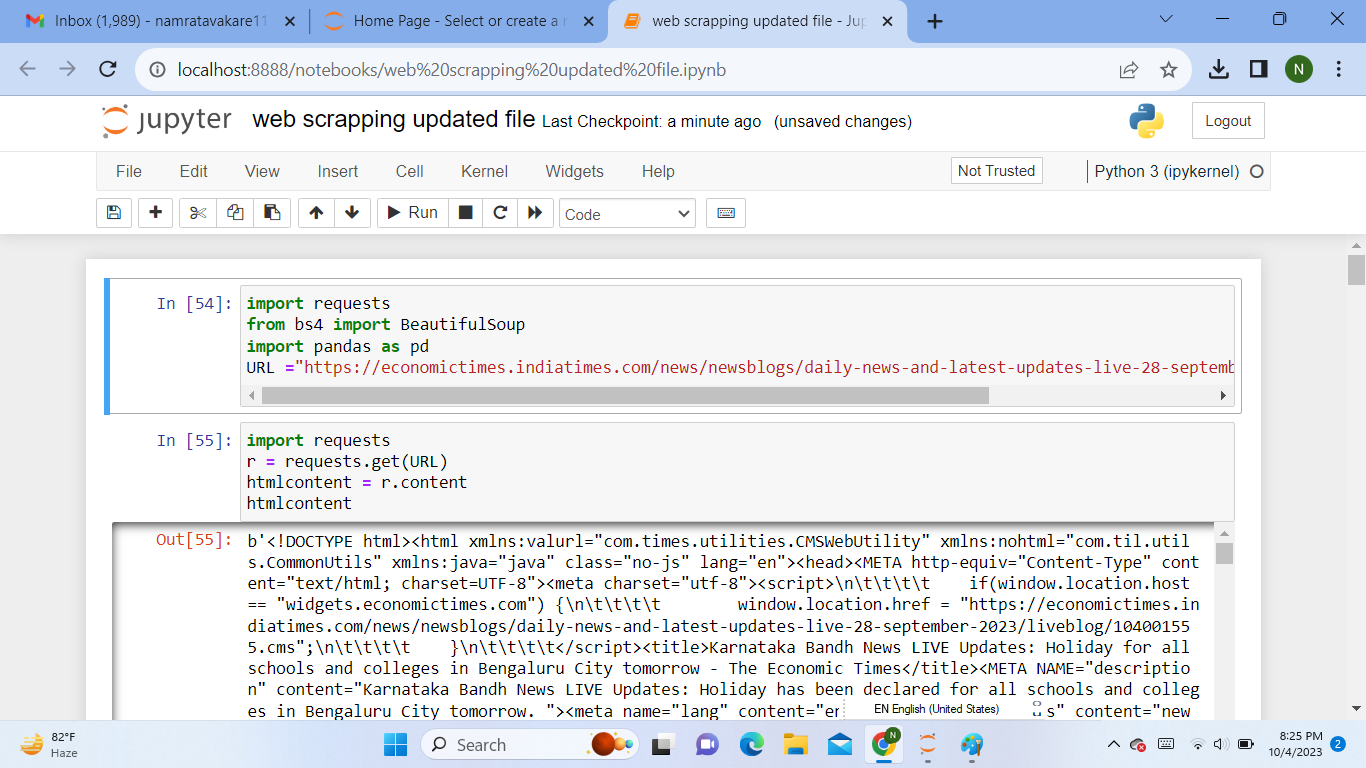
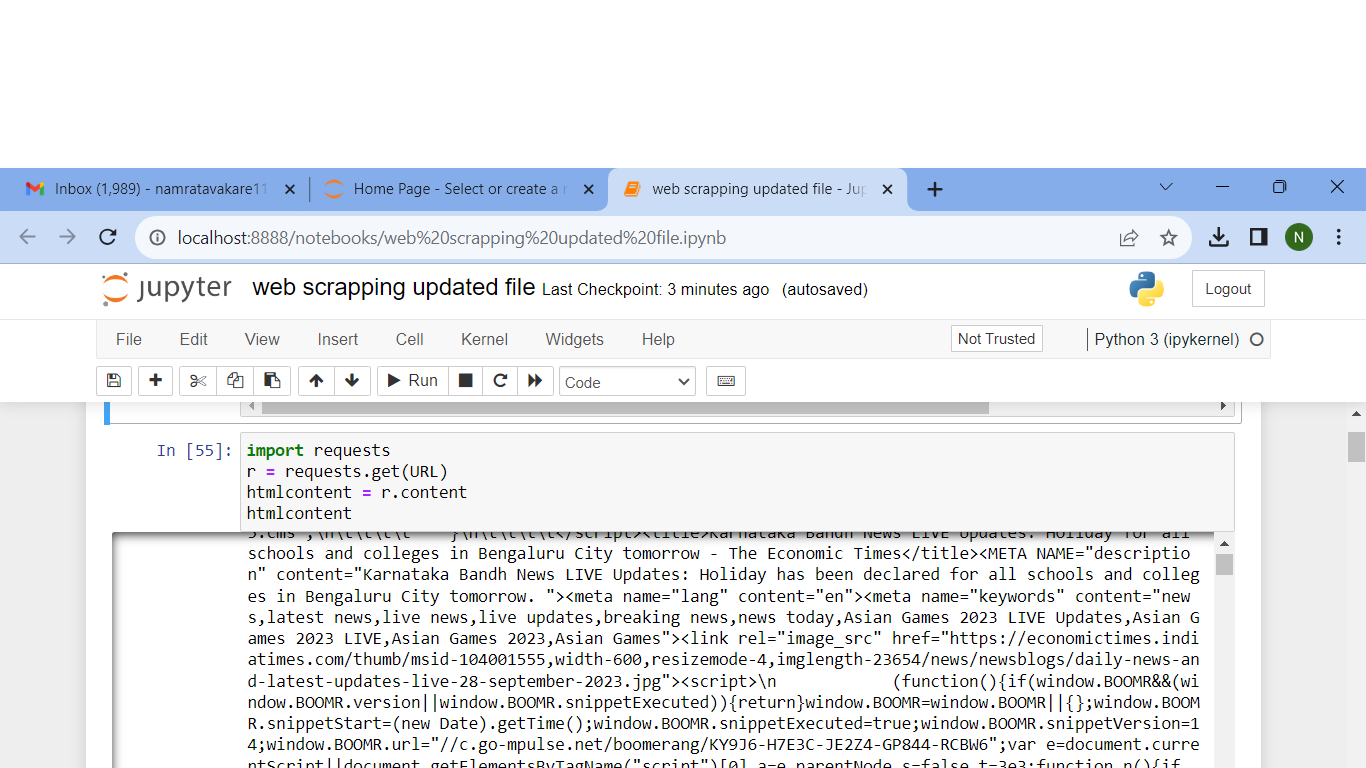
Web Scraping Daily News Updates from Economic Times

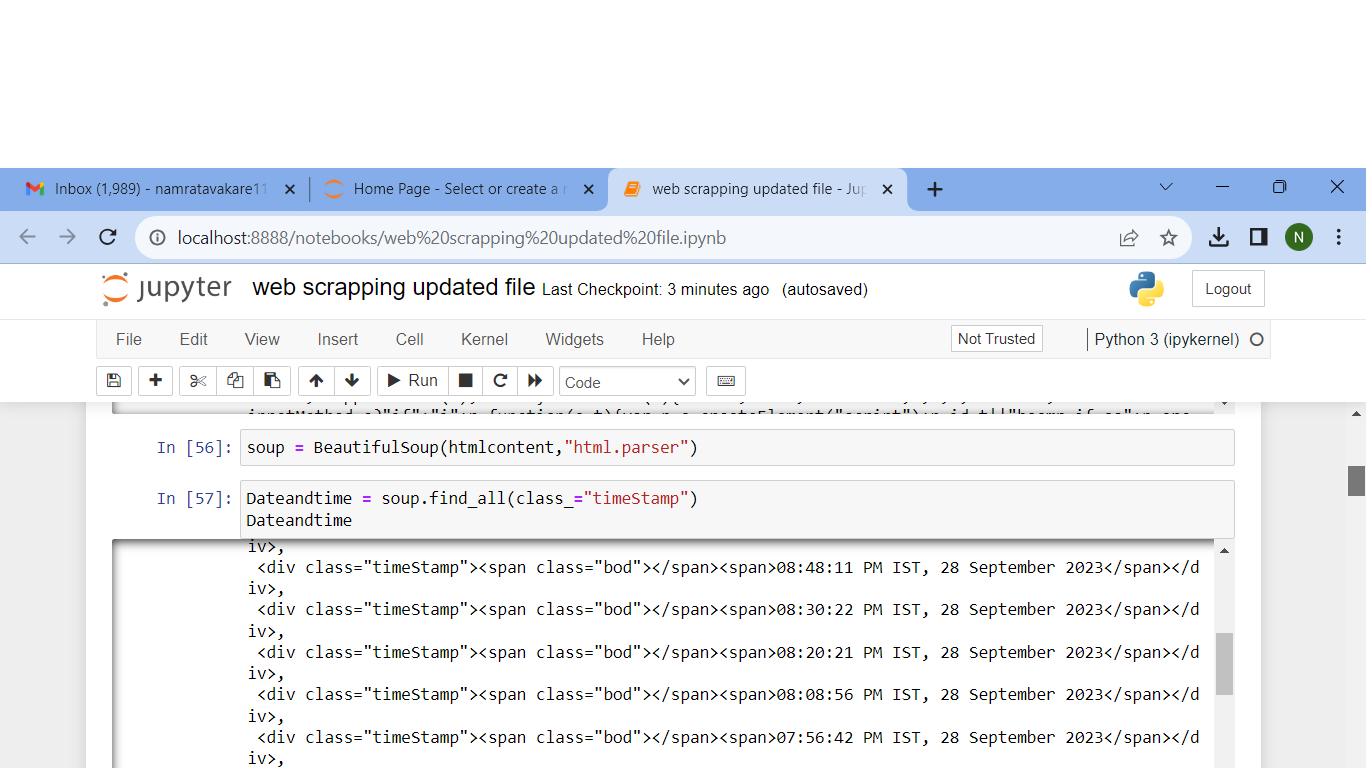
This project is done by scraping the data from the “THE ECONOMICS TIMES”, the steps used to make this project is given below:-

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

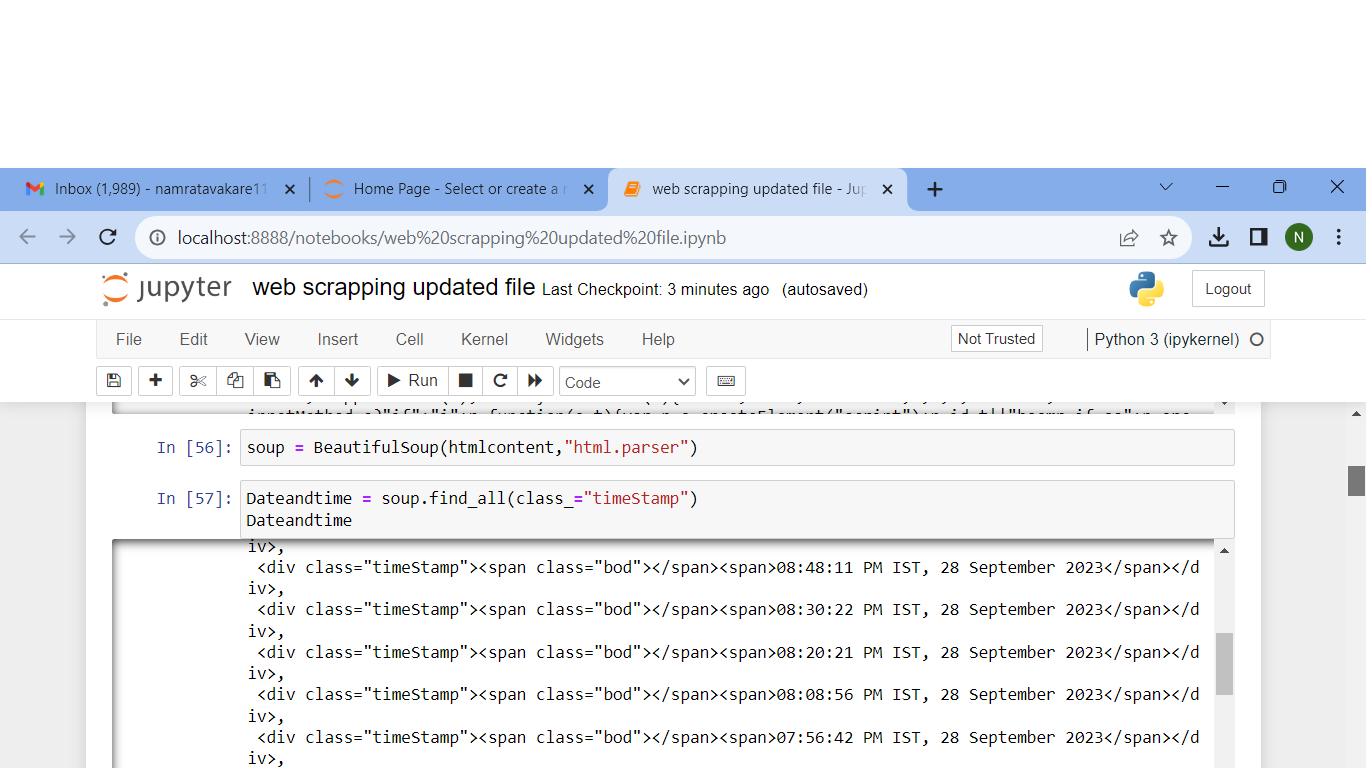
STEP1:- IMPORTING LIBRARIES

 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.

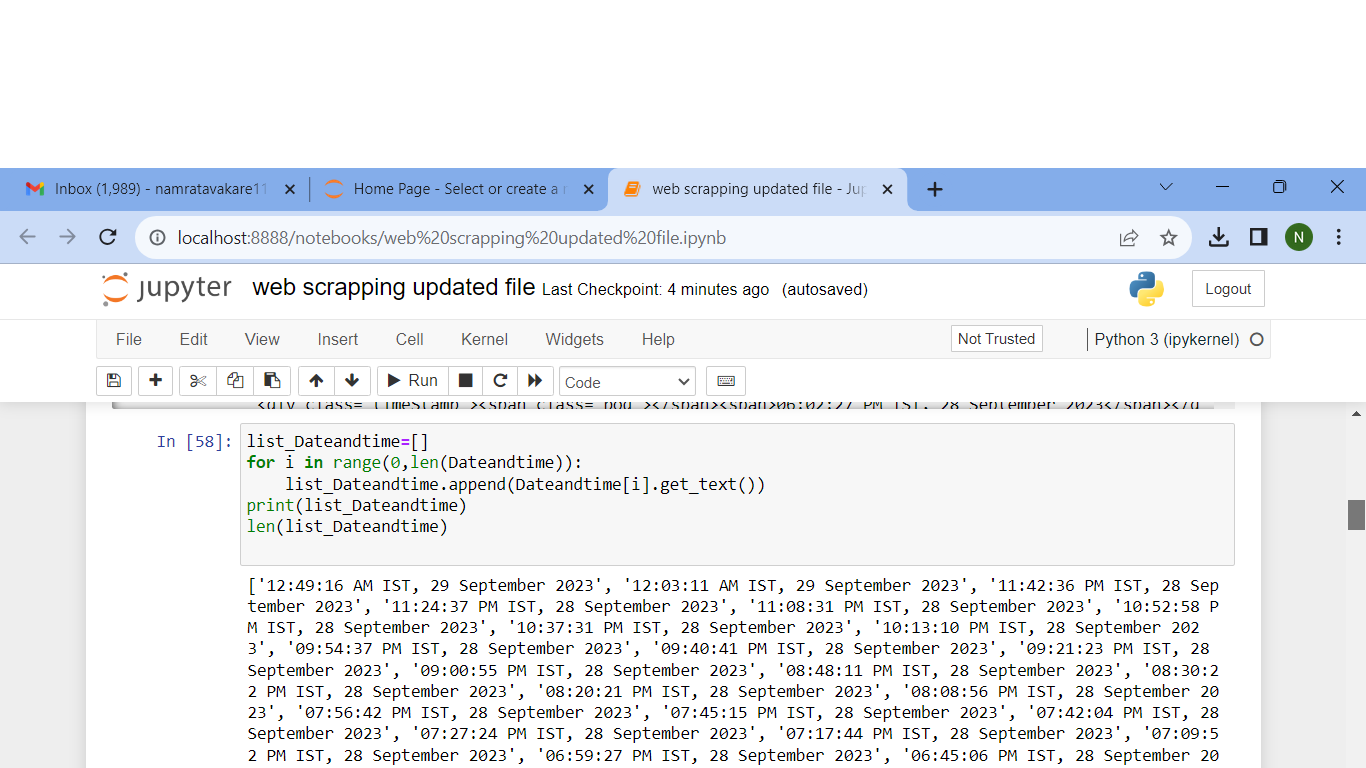
STEP2 :- Now, we have successfully fetched the HTML content of the web page and stored it in the html content variable. This content can be further processed and parsed using libraries like BeautifulSoup to extract specific data or information from the web page.

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.

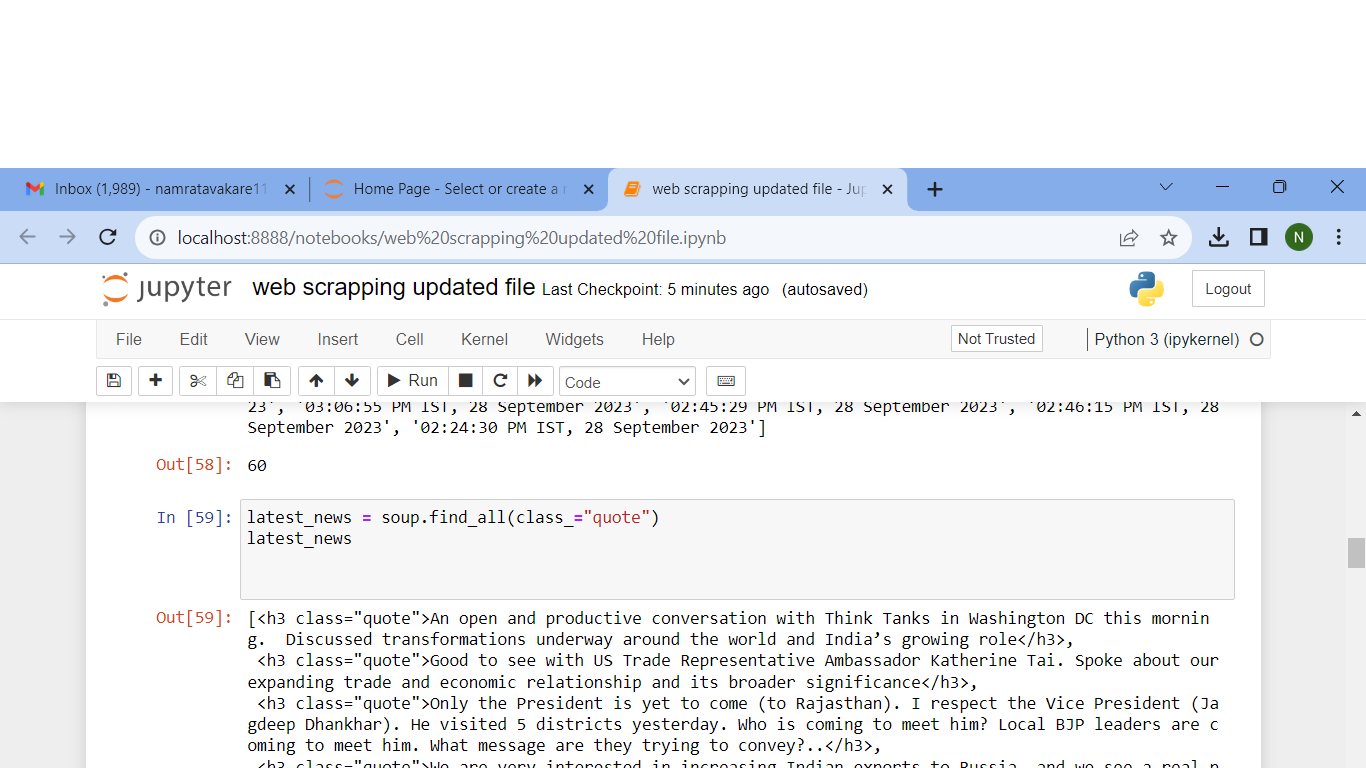
STEP 4 :- The code Dateandtime = soup.find\_all(class\_="timeStamp") is used to search for and extract HTML elements that have a specific CSS class name ("timeStamp") from the soup object, which represents the parsed HTML content of the web page.



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



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.



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.

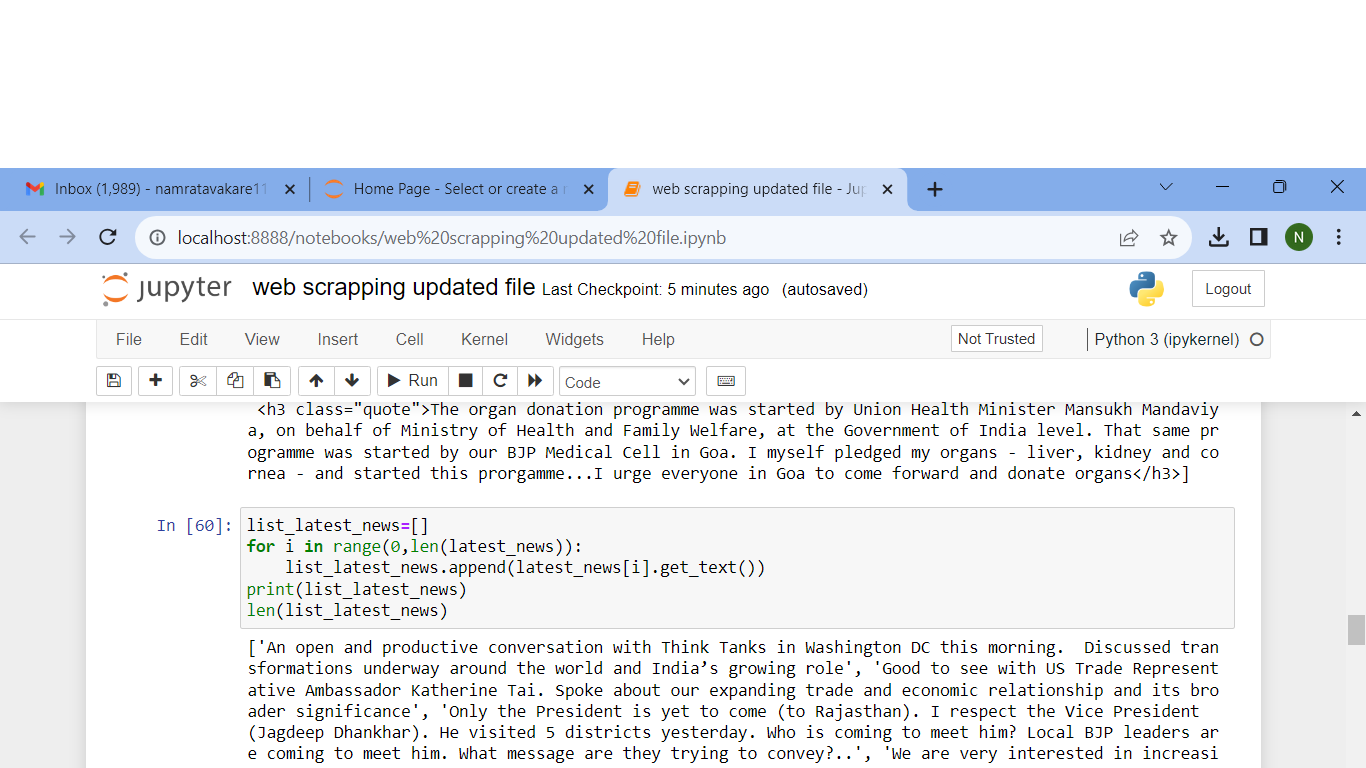
2]for i in range(0, len(latest\_news)):

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.

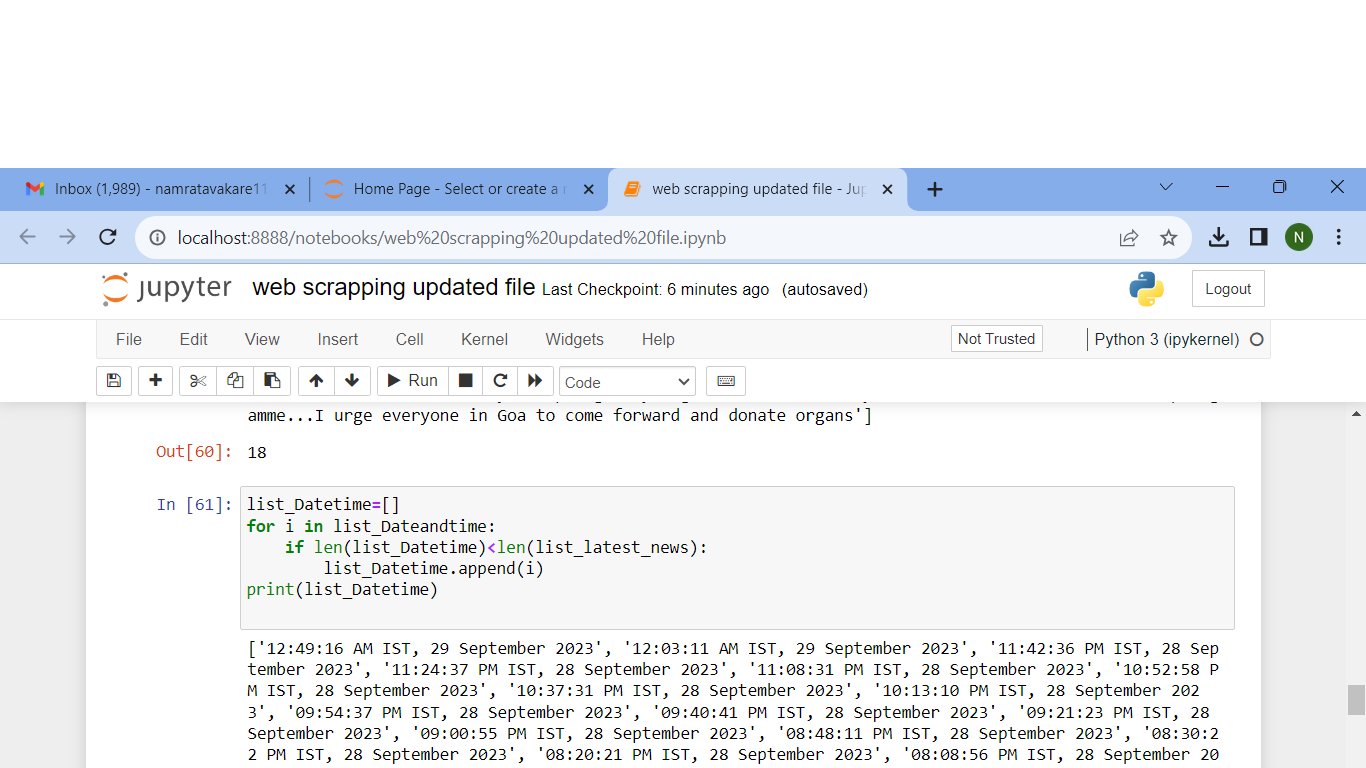
3]list\_latest\_news.append(latest\_news[i].get\_text()):

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.

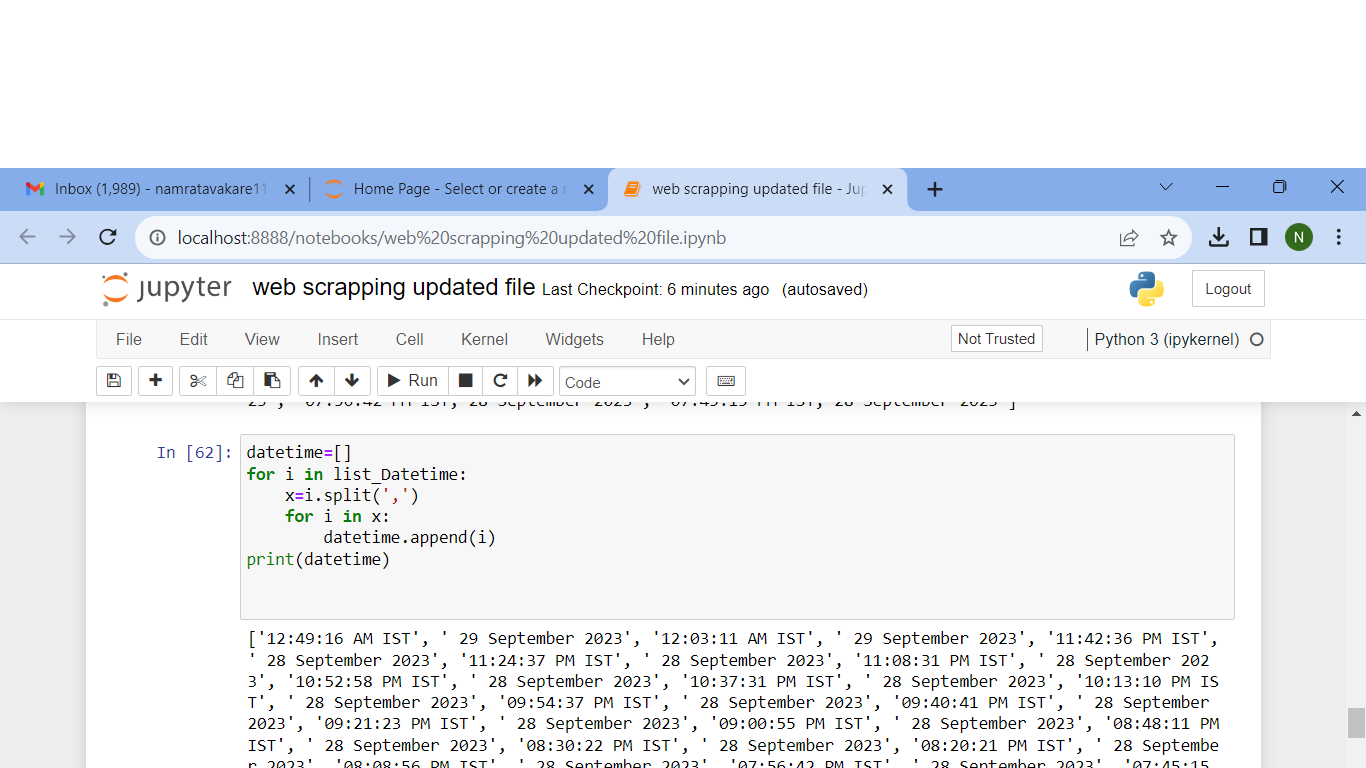
4]print(list\_latest\_news): After the loop completes, this line prints the list\_latest\_news list, which now contains the extracted text content from the HTML elements. These would typically be news headlines, quotes, or other textual content.



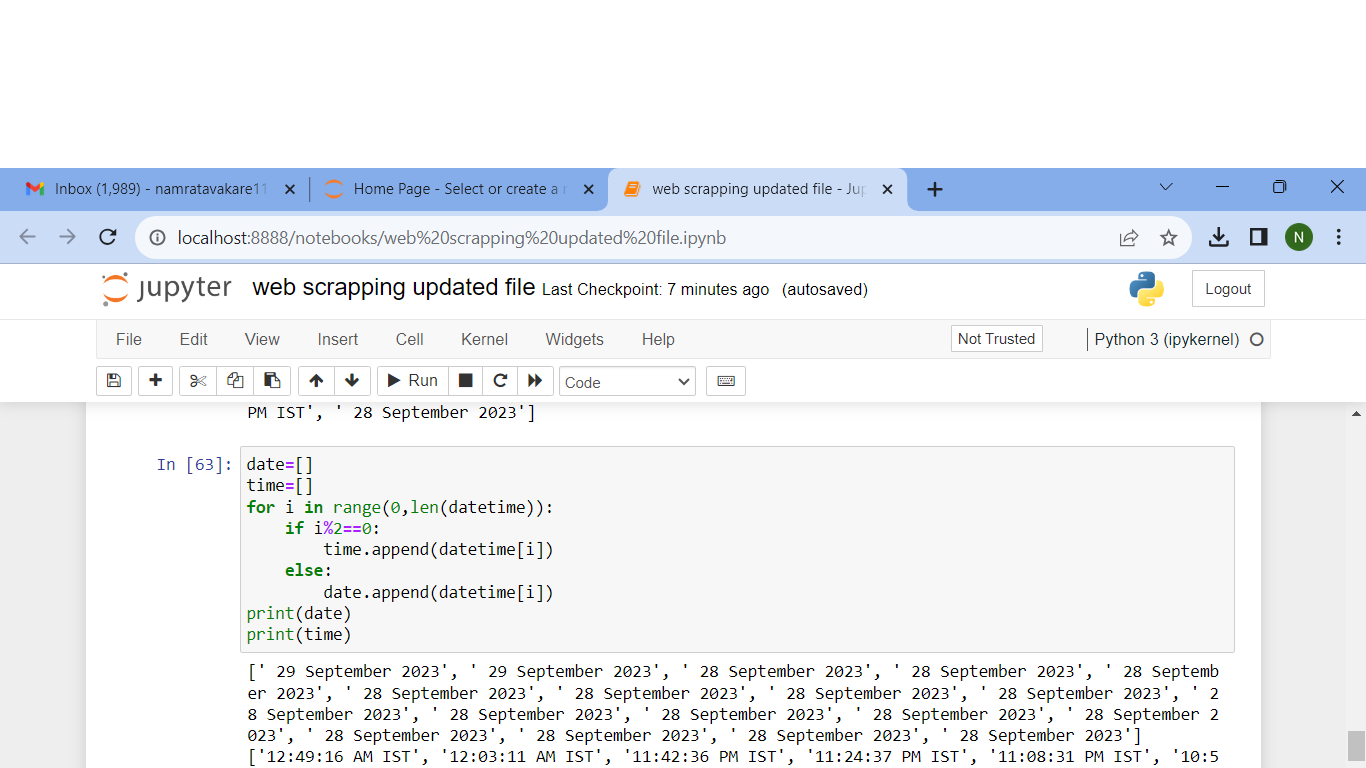
STEP 8 :- The code shown below is a loop that iterates through the elements in the list\_Dateandtime list and appends them to a new list called list\_Datetime. However, it includes a condition to ensure that the length of list\_Datetime remains less than the length of list\_latest\_news.



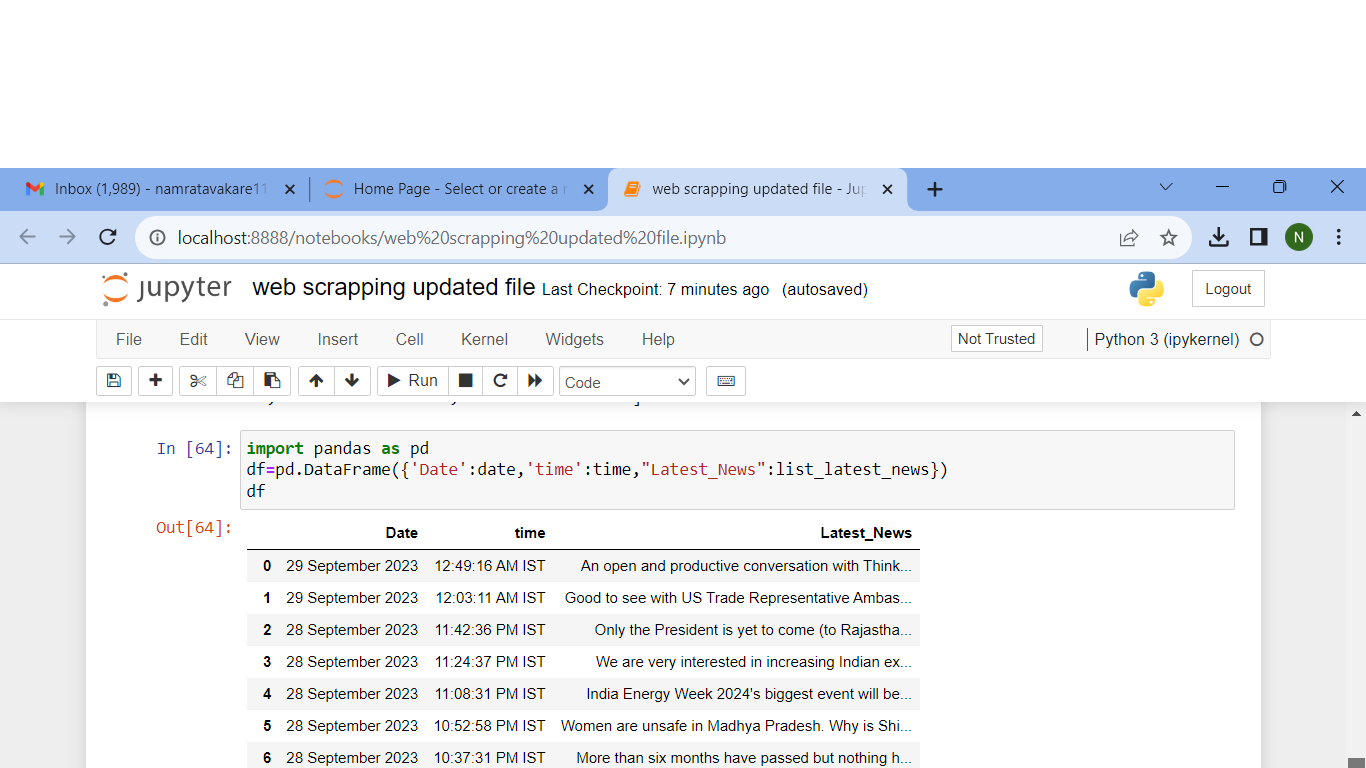
STEP 9 :- The code shown below is designed to split date and time information stored in the list\_Datetime list and then flatten it into a single list called datetime.



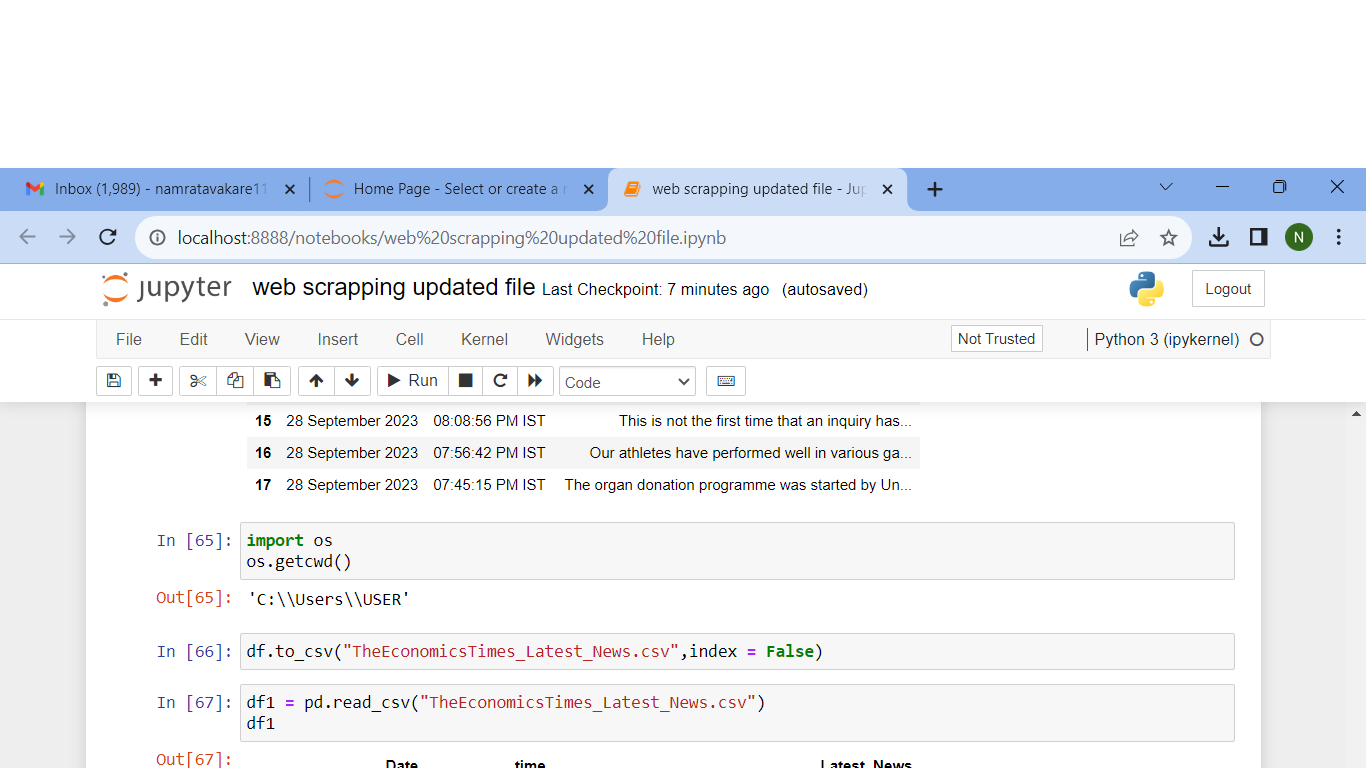
STEP 10 :- Next step is meant to separate the flattened list of date and time components from the datetime list into two separate lists: date and time. It does this by iterating through the datetime list and using the modulo operator (%) to determine whether an element at a given index is a date or a time component.



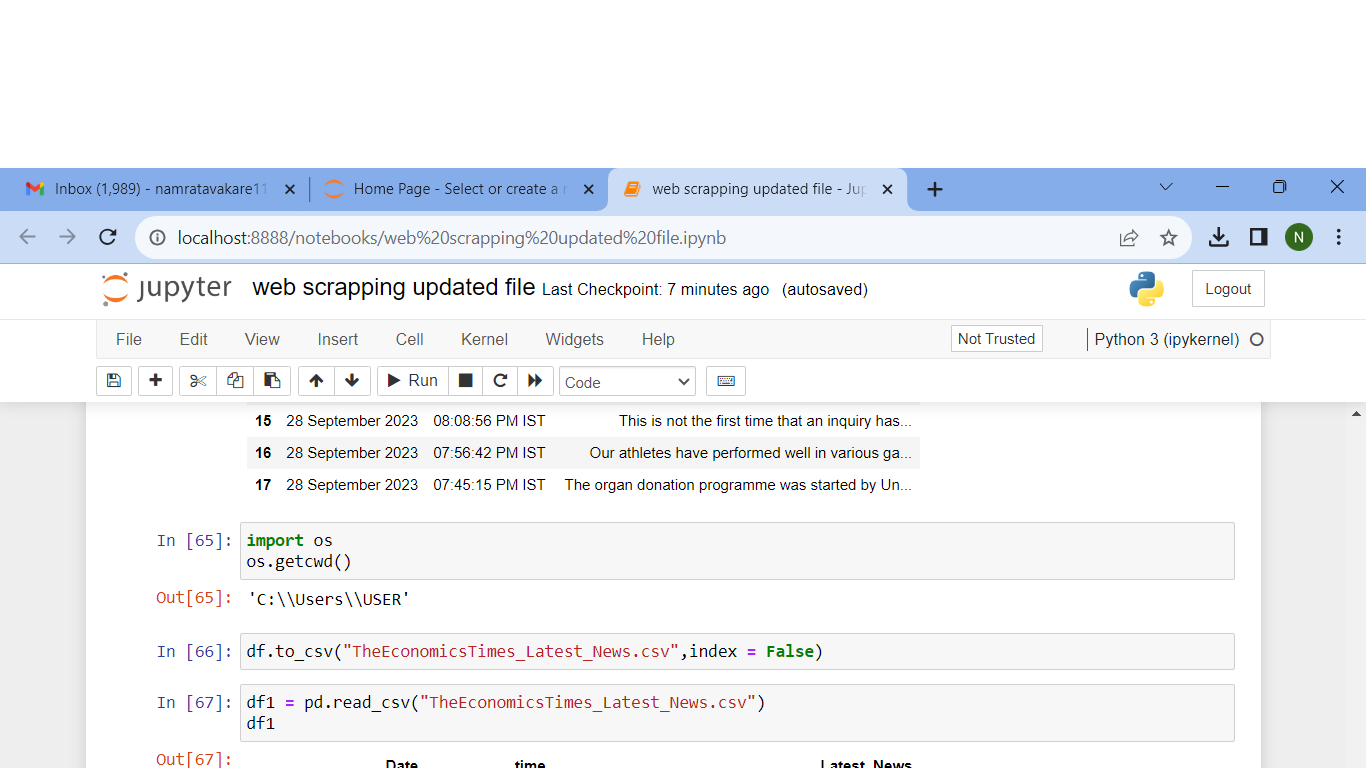
STEP 11 :- The code we have done is using the pandas library to create a DataFrame (df) from the lists date, time, and list\_latest\_news.



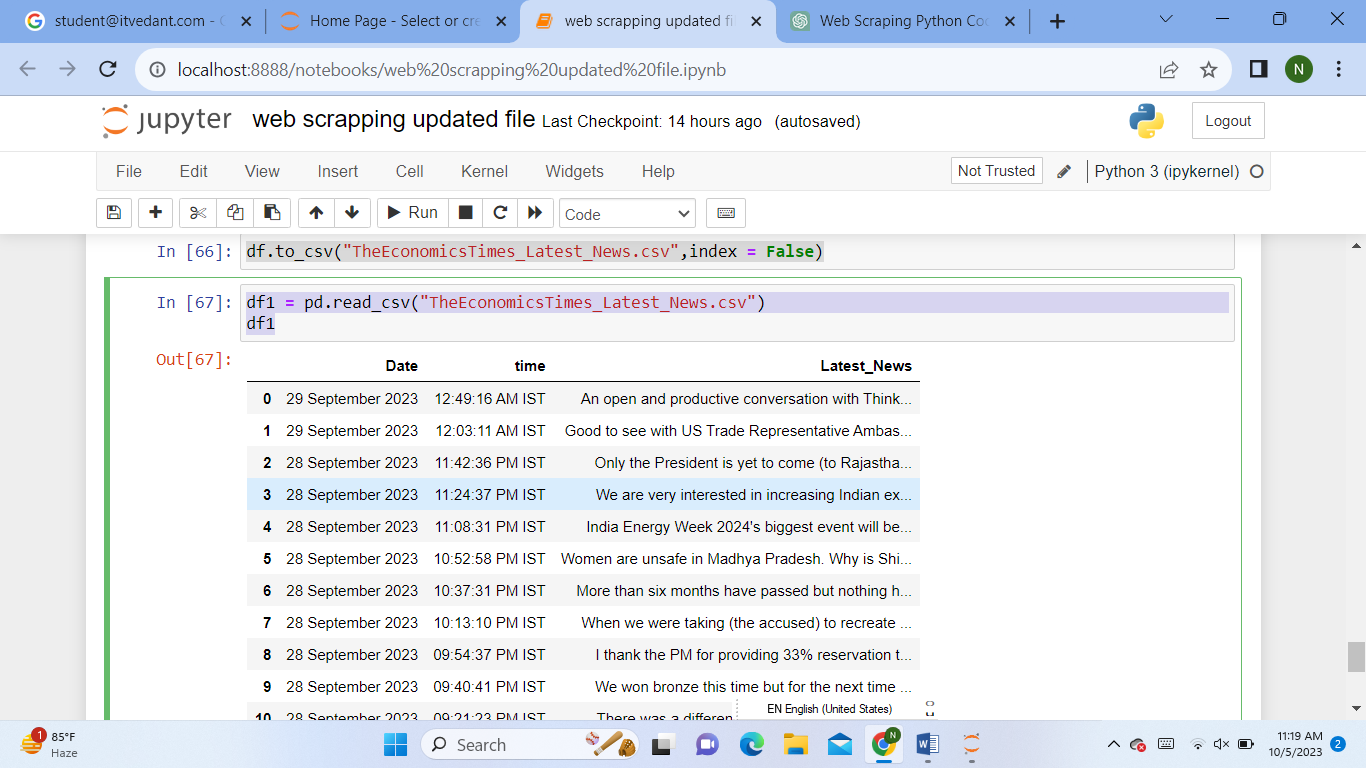
STEP 12:- The code os.getcwd() is used to get the current working directory (CWD) in Python.



STEP13:- The code df.to\_csv("TheEconomicsTimes\_Latest\_News.csv", index=False) is used to save the DataFrame df as a CSV (Comma-Separated Values) file named "TheEconomicsTimes\_Latest\_News.csv."



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



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.