**Assignment: Web Scraping**

**Python script using Selenium to log into a website and scrape specific data**

**1. Website Selection:**

I selected the Indian Express website (https://indianexpress.com/) for this task. Indian Express requires login authentication to access certain sections of the website, making it suitable for this assignment.

**2. Login Automation:**

I used Python's Selenium library to automate the login process. The script navigates to the login page, enters the credentials, and submits the form. It handles potential issues such as incorrect password alerts by waiting for the appropriate elements to appear before proceeding.

**3. Data Extraction:**

The objective was to scrape sports news articles from the Indian Express website. The script extracts the title, date, content, and image URL of each article from the Sports section. It iterates through multiple pages of the Sports section by clicking the 'Next' button until no more articles are available.

**4. Error Handling and Data Security:**

Error handling is implemented to manage potential issues during the login or scraping process. The script catches exceptions and prints appropriate error messages. Regarding data security, sensitive information such as login credentials is stored locally and handled securely within the script.

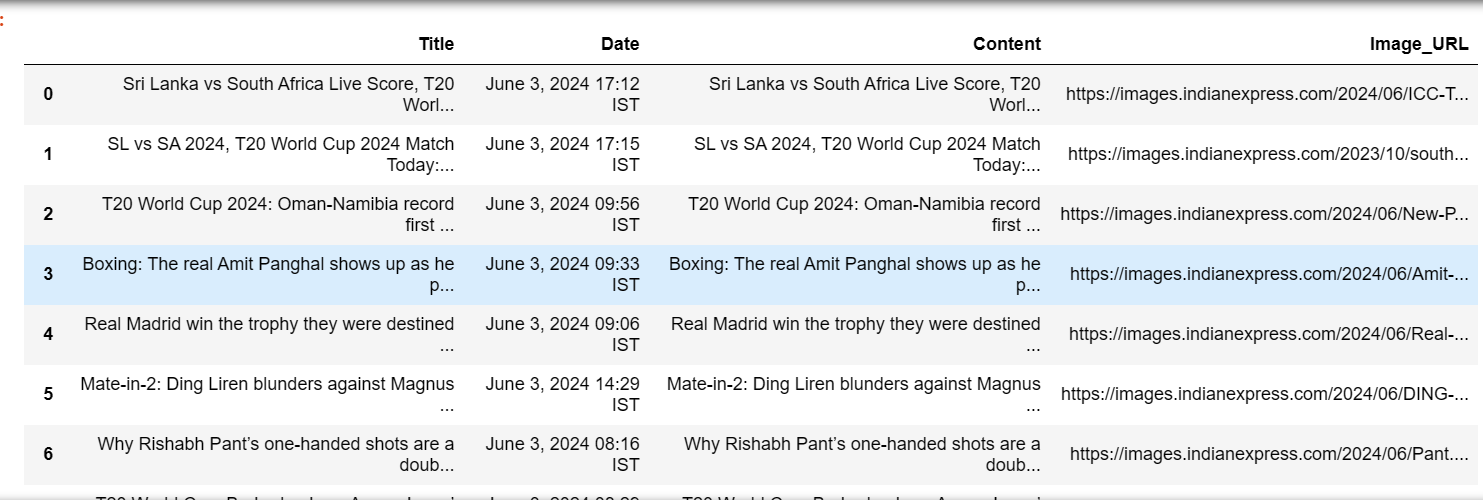
**5. Challenges Encountered and Solutions Implemented:**

* Handling the login process: The script had to handle dynamic elements and potential errors like incorrect password alerts.
* Pagination: To scrape articles from multiple pages, the script had to handle the 'Next' button dynamically.
* Error handling: Implemented try-except blocks to manage exceptions gracefully and provide helpful error messages.
* Data formatting: Ensured that extracted data was structured appropriately for further analysis or storage.

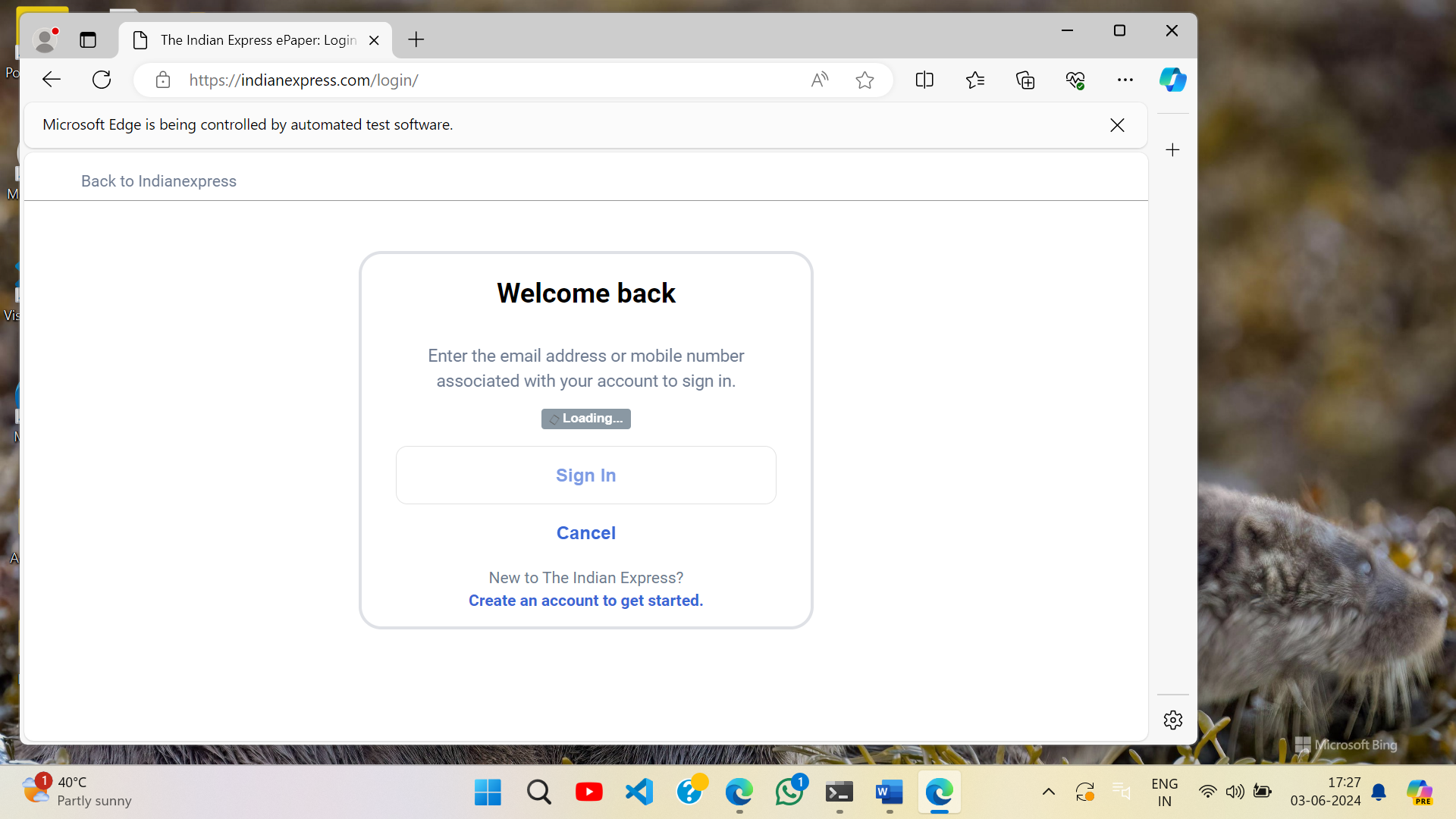
**6. Insights or Potential Applications of the Scraped Data:**

* The scraped sports news data can be used for various purposes such as sentiment analysis, trend analysis, or building recommendation systems.
* Insights from this data could be valuable for sports enthusiasts, journalists, or researchers interested in analyzing news trends.

## Some Screenshots



Extracted Data



Login Process