

```
In [71]: # importing pandas module
import pandas as pd
```

```
In [73]: ! pip install html5lib
```

Requirement already satisfied: html5lib in c:\users\kirti\anaconda3\lib\site-packages (1.1)

Requirement already satisfied: six>=1.9 in c:\users\kirti\anaconda3\lib\site-packages (from html5lib) (1.16.0)

Requirement already satisfied: webencodings in c:\users\kirti\anaconda3\lib\site-packages (from html5lib) (0.5.1)

## Reading EXCEL File

```
In [76]: # reading excel file from url
data_excel = pd.read_excel(r'D:\Full Stack Data Scientist and AI\March 25 - EDA
```

```
In [78]: data_excel
```

```
Out[78]:
```

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	NaN	Address	Date	Name	Time
1	NaN	House no. 139 Divine World, Sector 115, Kharar...	11-06-2025	Kirti	08:30:21
2	NaN	11 Meadowland Drive, Brampton, L6W 2R5	12-06-2025	Sid	08:45:55

## Reading HTML File

```
In [81]: # reading html file from url
data_html = pd.read_html(r'D:\Full Stack Data Scientist and AI\March 25 - EDA In
```

```
In [83]: data_html
```

```
Out[83]: [
      Company      Contact      Country
0   Alfreds Futterkiste   Maria Anders  Germany
1  Centro comercial Moctezuma Francisco Chang  Mexico
2      Ernst Handel      Roland Mendel  Austria
3   Island Trading      Helen Bennett      UK
4 Laughing Bacchus Winecellars Yoshi Tannamuri  Canada
5 Magazzini Alimentari Riuniti Giovanni Rovelli  Italy,
      Tag      Description
0   <table>      Defines a table
1   <th>      Defines a header cell in a table
2   <tr>      Defines a row in a table
3   <td>      Defines a cell in a table
4   <caption>      Defines a table caption
5   <colgroup> Specifies a group of one or more columns in a ...
6   <col> Specifies column properties for each column wi...
7   <thead>      Groups the header content in a table
8   <tbody>      Groups the body content in a table
9   <tfoot>      Groups the footer content in a table]
```

- If there's no table tag, then `pd.read_html()` won't work — it only extracts tables from HTML, not divs, spans, or other content.
- ! pip install html5lib is necessary in order to read HTML Files

## Reading XML File

```
In [87]: # reading xml file from url
data_xml = pd.read_xml(r'D:\Full Stack Data Scientist and AI\March 25 - EDA Intr
```

```
In [89]: data_xml
```

```
Out[89]:
```

	name	age	email	title	author	year
0	John Doe	30.0	john.doe@example.com	None	None	NaN
1	Jane Smith	25.0	jane.smith@example.com	None	None	NaN
2	None	NaN	None	The Adventure Begins	Robert Johnson	2022.0

```
In [ ]:
```