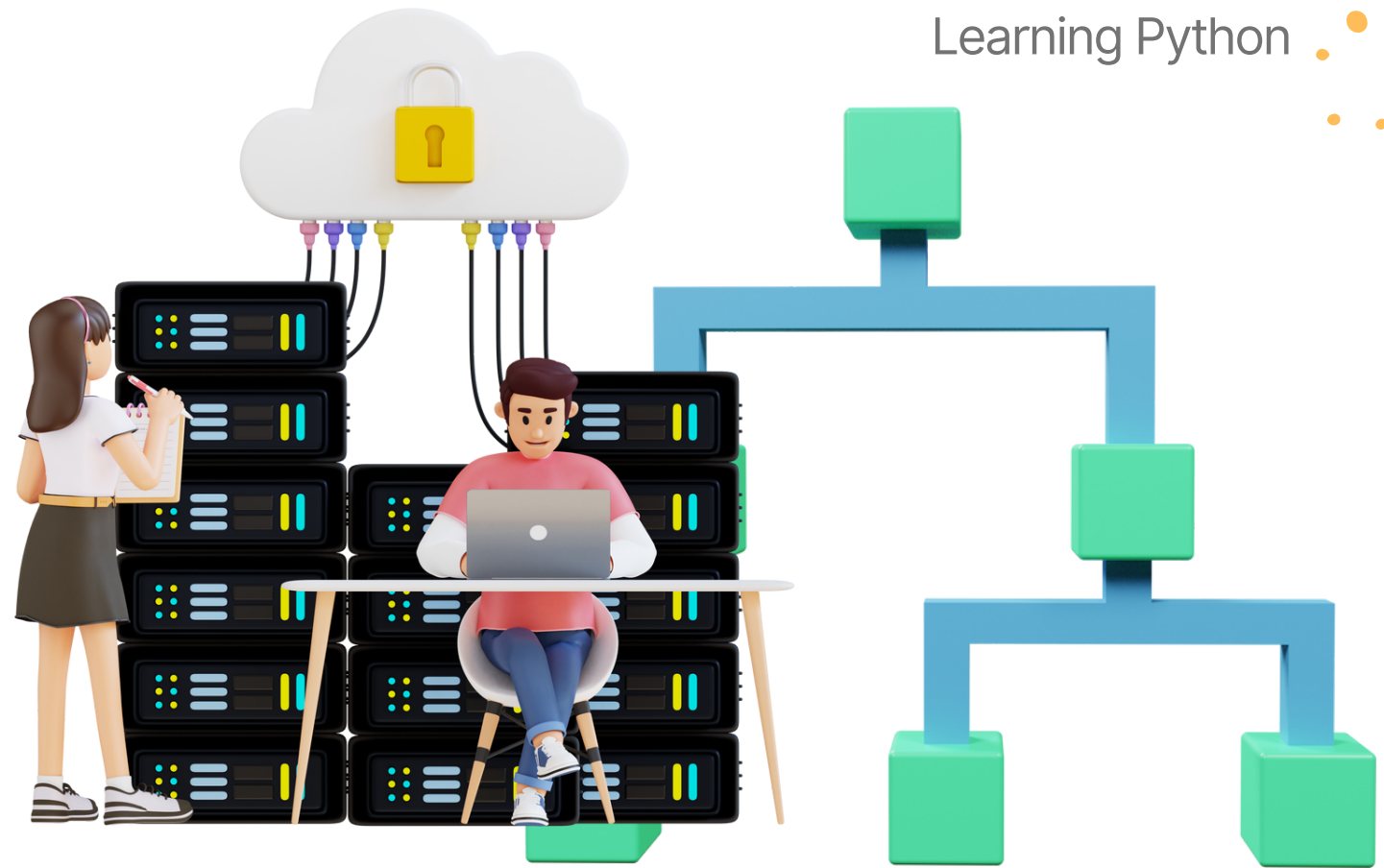


Data Ingestion with Pandas

Importing Data in Python

CSV, Excel, HTML, SQL, and JSON





Data ingestion is the process of pulling data from a variety of sources into your system with the purpose of easily exploring, maintaining, and using it for your own use.

Let's see how to transfer data from various formats such as *CSV*, *Excel*, *HTML*, *SQL*, and *JSON* into a Pandas dataframe object.

Importing CSV

Comma Separated Values (CSV) file is a plain text file that stores data by delimiting data entries with commas.



df = pd.read_csv("file path with file format")

Annotations:

- dataframe name (points to `df`)
- pd alias for pandas library (points to `pd`)
- read_csv method (points to `read_csv`)
- like: /download/mydata.csv (points to the file path argument)

Importing Excel

Excel sheets (spreadsheets) are still widely used in various domains. Excel is widely used, for developing budgeting strategies to track daily student attendance.



dataframe_name

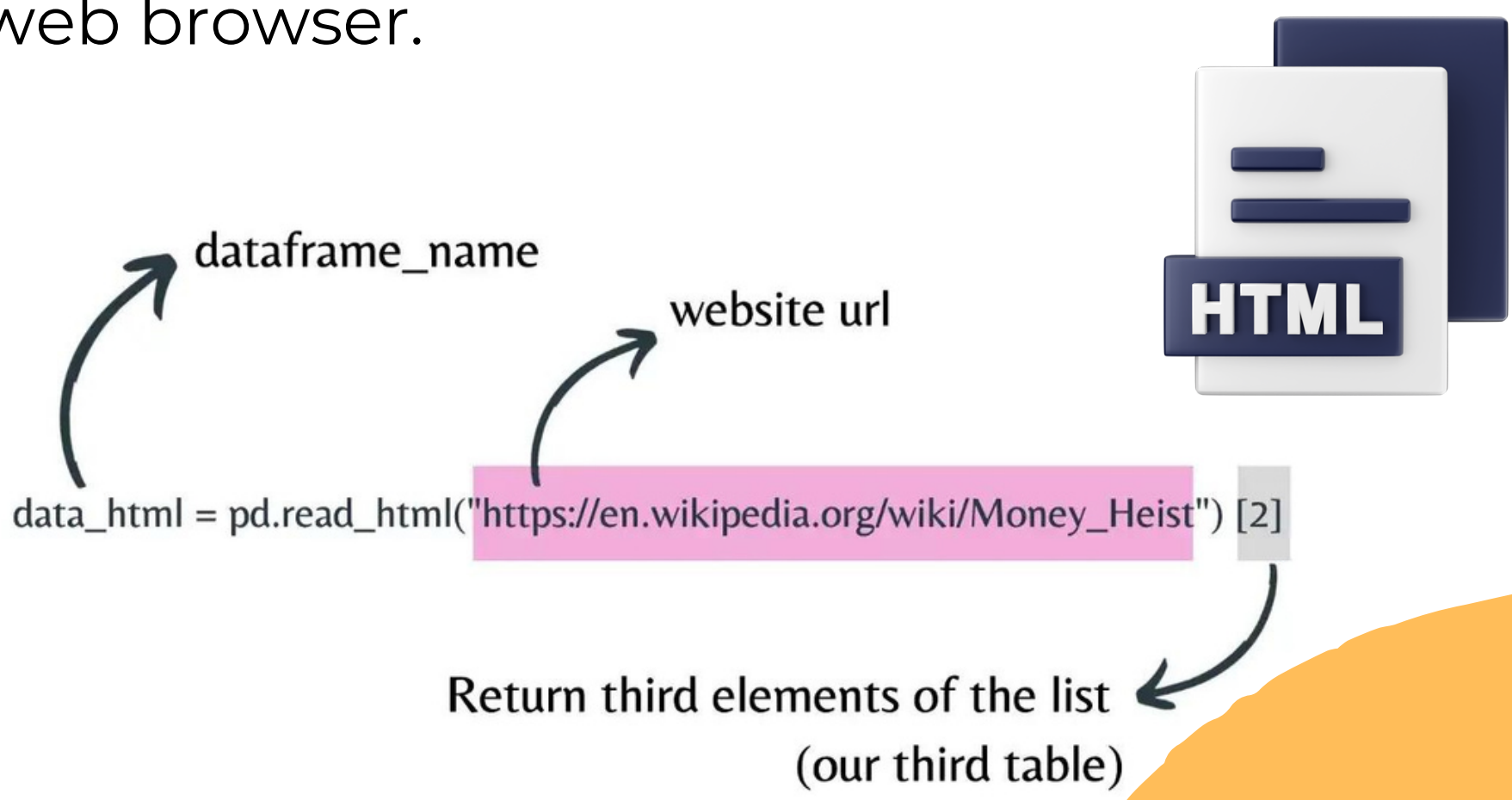
file path

First sheet of the workbook

```
data_excel = pd.read_excel("nepalcensusdata.xlsx", sheet_name = 0)
```

Importing HTML

The HyperText Markup Language or HTML is the standard markup language for documents designed to be displayed in a web browser.



```
data_html = pd.read_html("https://en.wikipedia.org/wiki/Money_Heist") [2]
```

The diagram illustrates the process of importing HTML. It shows a code snippet where `pd.read_html()` is used to fetch data from a website. An arrow points from the `data_html` variable to the label `dataframe_name`. Another arrow points from the URL string `"https://en.wikipedia.org/wiki/Money_Heist"` to the label `website url`. A third arrow points from the index `[2]` to the text `Return third elements of the list (our third table)`. To the right of the code is an icon representing an HTML document, labeled `HTML`.

Return third elements of the list
(our third table)

Importing SQL

Structured Query Language (SQL) is a standard language for storing, manipulating, and retrieving data in databases.



dataframe_name

SQL query

Connection object

```
data_Sql = pd.read_sql_query("SELECT * FROM fish", connection)
```

A diagram illustrating the components of the `pd.read_sql_query` function call. It shows the function call `data_Sql = pd.read_sql_query("SELECT * FROM fish", connection)` with three labels and arrows: 'dataframe_name' points to `data_Sql`, 'SQL query' points to the string `"SELECT * FROM fish"`, and 'Connection object' points to `connection`.

Importing JSON

JavaScript Object Notation (JSON) is the most common file format preferred by web developers and data engineers.



dataframe_name

file path

```
data_json = pd.read_json("opnePiece.json")
```



Of course, acquiring data is just the beginning.

After this, you need to explore data and play with it before you get deep into a data science project.

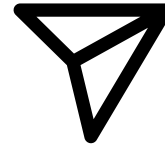


If you would like to take a deep dive into the data science project, check out the link in the bio.



WAS THIS HELPFUL?

Like, Save and share this post
if you find it valuable!



Datawithmala