**Python Webscraping Project(Coursera)**

Soup = BeautifulSoup(html,’html1’)

This command is used to parse the html document to get the object which represents the document as nested data structure.

Beautiful Soup represents the HTML as Tree like structure.

Tag\_object = soup.<head>

Tag\_child = tag\_object.<p>

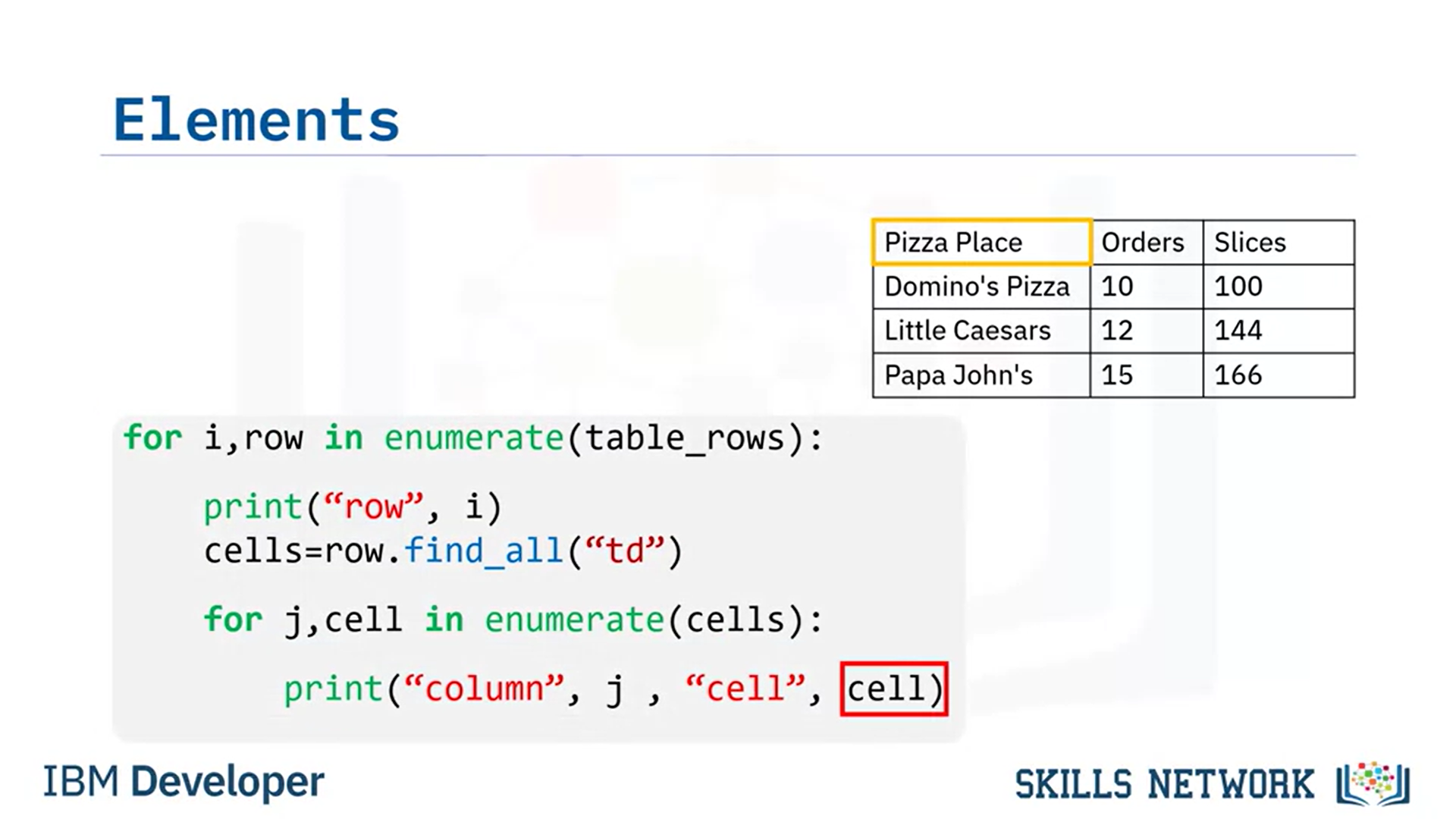
Tag\_string = tag\_child.string

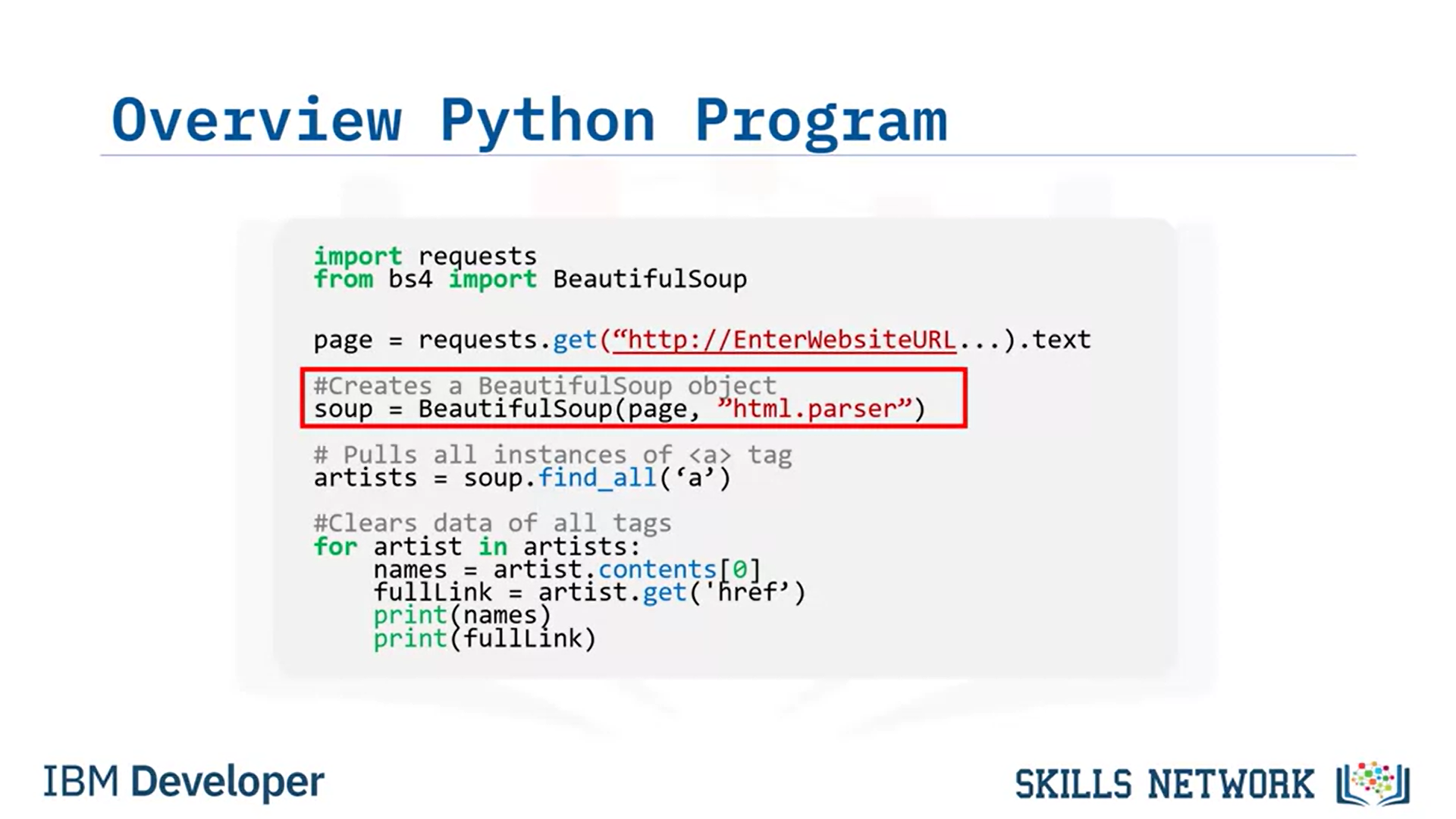
Tag\_dict = tag\_child.attrs

Table\_row = table.find\_all(name = ‘tr’)

Table\_row

First\_row = table\_row[0]





**Steps for extracting the data**

1. Send an HTTP request to the web page using the requests library.
2. Parse the HTML content of the web page using BeautifulSoup.
3. Identify the HTML tags that contain the data you want to extract.
4. Use BeautifulSoup methods to extract the data from the HTML tags.
5. Print the extracted data

n simple words, parsing refers to the process of analyzing a string of text or a data structure, usually following a set of rules or grammar, to understand its structure and meaning. Parsing involves breaking down a piece of text or data into its individual components or elements, and then analyzing those components to extract the desired information or to understand their relationships and meanings.

Next you will take the raw HTML content of a web page or a string of HTML code which needs to be parsed and transformed into a structured, hierarchical format that can be more easily analyzed and manipulated in Python. This can be done using a Python library called **Beautiful Soup**.

