Interview Questions

1. What is a GET request?

A **GET request** is an HTTP method used to **retrieve data** from a server. It requests information without making any changes to the server's data.

---

2. How do you install external packages in Python?

You can install packages using **pip** (Python's package manager):

pip install package_name

Example:

pip install requests

---

3. What is a User-Agent in HTTP?

A **User-Agent** is a string that identifies the browser, device, or software making the HTTP request. It helps servers tailor responses based on the client type.

---

4. What is soup.find_all() used for?

In **BeautifulSoup**, find_all() is used to **find all occurrences** of a specific HTML tag or element.
Example:

soup.find_all('h2')

→ Finds all <h2> tags in the webpage.

---

5. What are the risks of web scraping?

- Legal or ethical issues (violating website terms of service)
- Getting IP blocked for too many requests
- Data inaccuracies if website structure changes
- Heavy load on website servers

---

6. What's the difference between id and class in HTML?

- id → Unique identifier for a single HTML element (#idname)
- class → Can be shared by multiple elements (.classname)

Example:

```
<p id="intro">Hello</p>

<p class="text">World</p>
```

---

7. What is an HTML tag?

An **HTML tag** defines the structure and content of a webpage element.
Example: <h1>Title</h1> — here <h1> is a tag for a heading.

---

8. What does .text return in BeautifulSoup?

.text returns **all the visible text** inside an HTML tag, removing all HTML tags and markup.
Example:

```
soup.find('h1').text
```

→ Returns just the text inside the <h1> tag.

---

9. What is a try-except block?

A **try-except** block handles **errors (exceptions)** in Python.
Example:

```
try:

    print(10 / 0)

except ZeroDivisionError:

    print("Cannot divide by zero")
```

---

10. What are HTTP status codes?

HTTP status codes are **3-digit responses** from a server that indicate the result of a request:

- **200** – OK (success)

- **404** – Not Found

- **500** – Server Error

- **403** – Forbidden

- **301** – Moved Permanently