<> Code Pull requests ▷ Actions ▦ Projects 📖 Wiki ⊘ Security ⊯ Insights ⚙ Settings

👁 ⑂ ☆

☆ **0** stars ⑂ **32** forks 👁 **0** watching ⑂ Branches ∿ Activity
🏷 Tags

🌐 Public repository · Forked from **DHINESH-SEC/Exno-2appds**

⑂ ⑂ **1 Branch** 🏷 **0 Tags** ⑂ 🏷️ 🔍 Go to file t **Go to file** + **Add file** ▾ **Code** ⋯

This branch is **3 commits ahead of** DHINESH-SEC/Exno-2appds:main . ⑂ **Contribute** ▾ ↻ **Sync fork** ▾

	Nemaleshh Update README.md	bcaae91 · last month	🕘
🗋 EXNO_2_APPDS_GB.ipynb	success	last month	
🗋 README.md	Update README.md	last month	
🗋 exno_2_appds_gb.py	Add files via upload	2 months ago	
🗋 image.png	success	last month	
🗋 image1.png	success	last month	
🗋 image2.png	success	last month	
🗋 image3.png	success	last month	
🗋 image4.png	success	last month	
🗋 new_data	success	last month	

📖 **README** ✎ ☰

Exno-2appds

AIM:

To Perform Data Collection through web scraping using python.

ALGORITHM: #Step 1: Include the Necessary python libraries.

Step 2: Use the requests library to send HTTP requests to a web page.

Step 3: Retrieve the HTML content and use BeautifulSoup to parse it.

Step 4: Navigate and extract the necessary data.

Step 5: Handle the Java script content and retrieve the data using the html tags.

Step 6: Check with the website permission and scrap the content.

CODING & OUTPUT:

```python
import requests
url = 'https://www.w3schools.com/html/html_tables.asp'
response = requests.get(url)
html_content = response.text
```

```python
html_content
```

```python
from bs4 import BeautifulSoup
soup = BeautifulSoup(html_content, 'html.parser')
```

```python
from selenium import webdriver
from selenium.webdriver.chrome.options import Options


options = Options()
options.add_argument('--headless')
options.add_argument('--no-sandbox')
options.add_argument('--disable-dev-shm-usage')
```

```
Requirement already satisfied: selenium in c:\users\nemaleshwar h\appdata\local\programs\python\python311\lib\site-packag
Requirement already satisfied: urllib3<3,>=1.26 in c:\users\nemaleshwar h\appdata\roaming\python\python311\site-packages
Requirement already satisfied: trio~=0.17 in c:\users\nemaleshwar h\appdata\local\programs\python\python311\lib\site-pack
Requirement already satisfied: trio-websocket~=0.9 in c:\users\nemaleshwar h\appdata\local\programs\python\python311\lib\
Requirement already satisfied: certifi>=2021.10.8 in c:\users\nemaleshwar h\appdata\roaming\python\python311\site-package
Requirement already satisfied: typing_extensions~=4.9 in c:\users\nemaleshwar h\appdata\local\programs\python\python311\l
Requirement already satisfied: websocket-client~=1.8 in c:\users\nemaleshwar h\appdata\roaming\python\python311\site-pack
Requirement already satisfied: attrs>=23.2.0 in c:\users\nemaleshwar h\appdata\roaming\python\python311\site-packages (fr
Requirement already satisfied: sortedcontainers in c:\users\nemaleshwar h\appdata\local\programs\python\python311\lib\sit
Requirement already satisfied: idna in c:\users\nemaleshwar h\appdata\roaming\python\python311\site-packages (from trio~=
Requirement already satisfied: outcome in c:\users\nemaleshwar h\appdata\local\programs\python\python311\lib\site-package
Requirement already satisfied: sniffio>=1.3.0 in c:\users\nemaleshwar h\appdata\roaming\python\python311\site-packages (f
Requirement already satisfied: cffi>=1.14 in c:\users\nemaleshwar h\appdata\roaming\python\python311\site-packages (from
Requirement already satisfied: wsproto>=0.14 in c:\users\nemaleshwar h\appdata\local\programs\python\python311\lib\site-p
Requirement already satisfied: pysocks!=1.5.7,<2.0,>=1.5.6 in c:\users\nemaleshwar h\appdata\local\programs\python\python
Requirement already satisfied: pycparser in c:\users\nemaleshwar h\appdata\roaming\python\python311\site-packages (from c
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\nemaleshwar h\appdata\roaming\python\python311\site-packages (fr
'apt-get' is not recognized as an internal or external command,
operable program or batch file.
'apt' is not recognized as an internal or external command,
operable program or batch file.
```

```python
url = 'https://www.geeksforgeeks.org/python-programming-language-tutorial/'
response = requests.get(url)
soup = BeautifulSoup(response.text, 'html.parser')
import pandas as pd
```

```python
headings = soup.find_all('h3')
for heading in headings:
    print(heading.text)
```

```python
headings = soup.find_all('h3')
for heading in headings:
    print(heading.text)
```
✓ 0.0s   Python

```
Expected Upcoming Features of Python 3.13
What is Python?
Is it easy to learn Python?
Is Python enough to get a job?
Is Python a good career option?
What kind of job can you get with Python?
What are the examples of Python's usage in Automation and Job Scheduling
How does Python support interactive mode and debugging?
How Can Python be integrated with other programming languages?
What are the key advantages of learning Python?
Similar Reads
Please Login to comment...
```

```python
data = []
for heading in headings:
    data.append(heading.text)
df=pd.DataFrame(data,columns=['Heading'])
df.to_csv('new_data',index=False)
```

Exno-2appds > M↓ new_data

```
1    Heading
2    Expected Upcoming Features of Python 3.13
3    What is Python?
4    Is it easy to learn Python?
5    Is Python enough to get a job?
6    Is Python a good career option?
7    What kind of job can you get with Python?
8    What are the examples of Python's usage in Automation and Job Scheduling
9    How does Python support interactive mode and debugging?
10   How Can Python be integrated with other programming languages?
11   What are the key advantages of learning Python?
12   Similar Reads
13   Please Login to comment...
14
```

```python
df.head()
```

```python
df.head()
```
✓ 0.0s   Python

	Heading
0	Expected Upcoming Features of Python 3.13
1	What is Python?
2	Is it easy to learn Python?
3	Is Python enough to get a job?
4	Is Python a good career option?

WEB SCRAPPING USING YOUTUBE WEBSITE

```python
import requests
from bs4 import BeautifulSoup
import pandas as pd
url="https://www.youtube.com/"
page =1
```

```python
def rev_scrap(url,page):
    url = f"{url}&pageNumber={page}"
rev_scrap(url,page)
user_agent = {"User-Agent": "Mozilla/5.0"}
response = requests.get(url=url,headers=user_agent)
```

Releases

No releases published
Create a new release

Packages

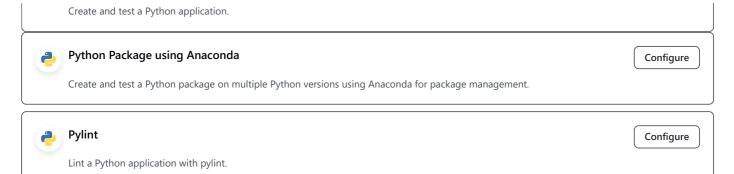No packages published
Publish your first package

Languages

● **Jupyter Notebook** 99.7%   ● **Python** 0.3%

Suggested workflows
Based on your tech stack

🐍 **Python application**   Configure

Create and test a Python application.

Python Package using Anaconda

Configure

Create and test a Python package on multiple Python versions using Anaconda for package management.

Pylint

Configure

Lint a Python application with pylint.

More workflows

Dismiss suggestions

Create and test a Python application.

Python Package using Anaconda

Create and test a Python package on multiple Python versions using Anaconda for package management.

Pylint

Lint a Python application with pylint.