WhiteHat Jr
Live Online Coding for Kids

INSTRUCTIONS:

_____

Goal of the Project:

In the class 127, you learned about web scraping and you wrote a scraper to scrape the information of different planets from nasa site.

In this project, you will write a scraper to scrape the data of various stars in our Universe! Stars reveal a lot of information too.

Story:

Our Sun is dying! The world is in an emergency as we are about to lose our star. All groups of scientists around the world have gathered together and created a technology to shift our Earth into another solar system, but which one exactly? Which star out there is safe and welcoming to our Earth? You have been assigned the task to research about stars so that we can choose the best one for us!

V Mag. (m_V) ⬥	Proper name ⬥	Bayer designation ⬥	Distance (ly) ⬥	Spectral class ⬥	Mass ($M_\odot$) ⬥	Radius ($R_\odot$) ⬥	Luminosity ($L_\odot$) ⬥
−26.74	Sun		0.000015813	G2	1	1	1
−1.46	Sirius	α CMa	8.6	A1	2.1	1.71´	25.4
−0.74	Canopus	α Car	310	A9	15	71	13,500
−0.27	Alpha Centauri	α Cen	4.4	G2	1.1	1.2	2
−0.05	Arcturus	α Boo	37	K2	1.1	26	170
0.03	Vega	α Lyr	25	A0	2.2	2.7	50
0.08	Capella	α Aur	43	G8	2.6	12	150
0.13	Rigel	β Ori	860	B8	23	78.9	120,000
0.34	Procyon	α CMi	11.4	F5	1.5	1.9	7.7
0.45	Betelgeuse	α Ori	640[1]	M2	20	950	60,000
0.46	Achernar	α Eri	144	B3	6.7	9.3	3,000
0.61	Hadar	β Cen	390	B1	10.5	8.6	42,000
0.76	Altair	α Aql	17	A7	1.8	1.8	10.5
0.76	Acrux	α Cru	320	B0.5	18	8.9	25,000
0.86	Aldebaran	α Tau	65	K5	1.5	44	520
0.96	Antares	α Sco	600	M1.5	12	680	75,000
0.97	Spica	α Vir	260	B1	11.43	7.47	20,512
1.14	Pollux	β Gem	34	K0	1.9	8.8	43
1.16	Fomalhaut	α PsA	25	A3	1.9	1.8	16.6
1.25	Deneb	α Cyg	2,600	A2	19	203	200,000
1.25	Mimosa	β Cru	350	B0.5	16	8.4	34,000

*This is just for your reference. We expect you to apply your own creativity in the project.

Getting Started:

1. Open your VS Code editor.
2. Create a virtual environment.
3. Source the virtual environment.
4. Install bs4 and requests using pip.

*We are not going to use Selenium for this project.

Specific Tasks to complete the Project:

1. Understand the website we want to scrape first: <u>Brightest stars in universe</u>.
2. Let's decide the headers we will keep for our study:
   - **Name**
   - **Distance**
   - **Mass**
   - **Radius**
3. Study the page in inspection mode to understand the patterns. You can use the following questions:
   - **How have they structured their HTML?**
   - **How many tables have they used?**
   - **Is there any specific pattern in the attributes each of the HTML tags has?**
4. Device a strategy on how to scrape and structure your code.
5. Observe the HTML of the page
6. Import all important/relevant modules.
7. Define the **START_URL, headers** and an **empty list** to store the star's data.
8. Get the HTML page with **requests module.**
9. Try to create a CSV by fetching one row from any of the tables.
   - This need not be all the data. This will be continued in the next project and we will scrape all the data.

*Refer to the images given above for reference.

Submitting the Project:

1. Upload your completed project to your own github account.
2. Create a new repository named " **Web scraping 1".**
3. **Upload** working code to this github repository.
4. Enable Github pages for the repository.
5. Copy the link to the github pages link in the Student Dashboard. link to the github pages link in the Student Dashboard.

Hints:

1. On getting the HTML page with requests module, you might get an SSL error. Add **verify=False** with the request and it will not verify for a secure connection.
2. Notice the unit of dimensions of the star data such as radius, distance and mass. If you are scraping from different sources make sure the units are the same, if not then we have to convert them.
3. Your observations may be:
   - **There is only one table on the page**
   - **Headings of the table are in \<th> tag**
   - **All the star's data rows are \<tr> tags**

REMEMBER... Try your best, that's more important than being correct.
After submitting your project your teacher will send you feedback on your work.

——————— xxx ——————— xxx ——————— xxx ——————— xxx ——————— xxx ———————