Advance web scraping

A 3-days training program

Asadh 7-9, 2081

Dr. Anil Shrestha
Undersecretary (Account)
Financial Administration Section
National Statistics Office

Outline

- Session 1: Introduction to web scraping
- Session 2: Excel and Google Sheets for web scraping
  - Excel's Web Query
  - Google Sheets' IMPORTHTML() and IMPORTDATA() functions for basic scraping.
  - Google Sheets' IMPORTXML() function for advance scraping.
- Session 3: Setting up R and Python for web scraping
- Session 4: Web scraping with R
- Session 5: Web scraping with Python

Introduction to web scraping

What is web scraping?

- An automated extraction of data from websites.

Process of web scraping

- Fetching web page and its html source.
- Understanding the structure of the web page HTML code.
- Extracting the desired data (product, prices, reviews etc.)
- Cleaning, manipulating, and storing the data.

Why use web scraping?

Web scraping: Demo

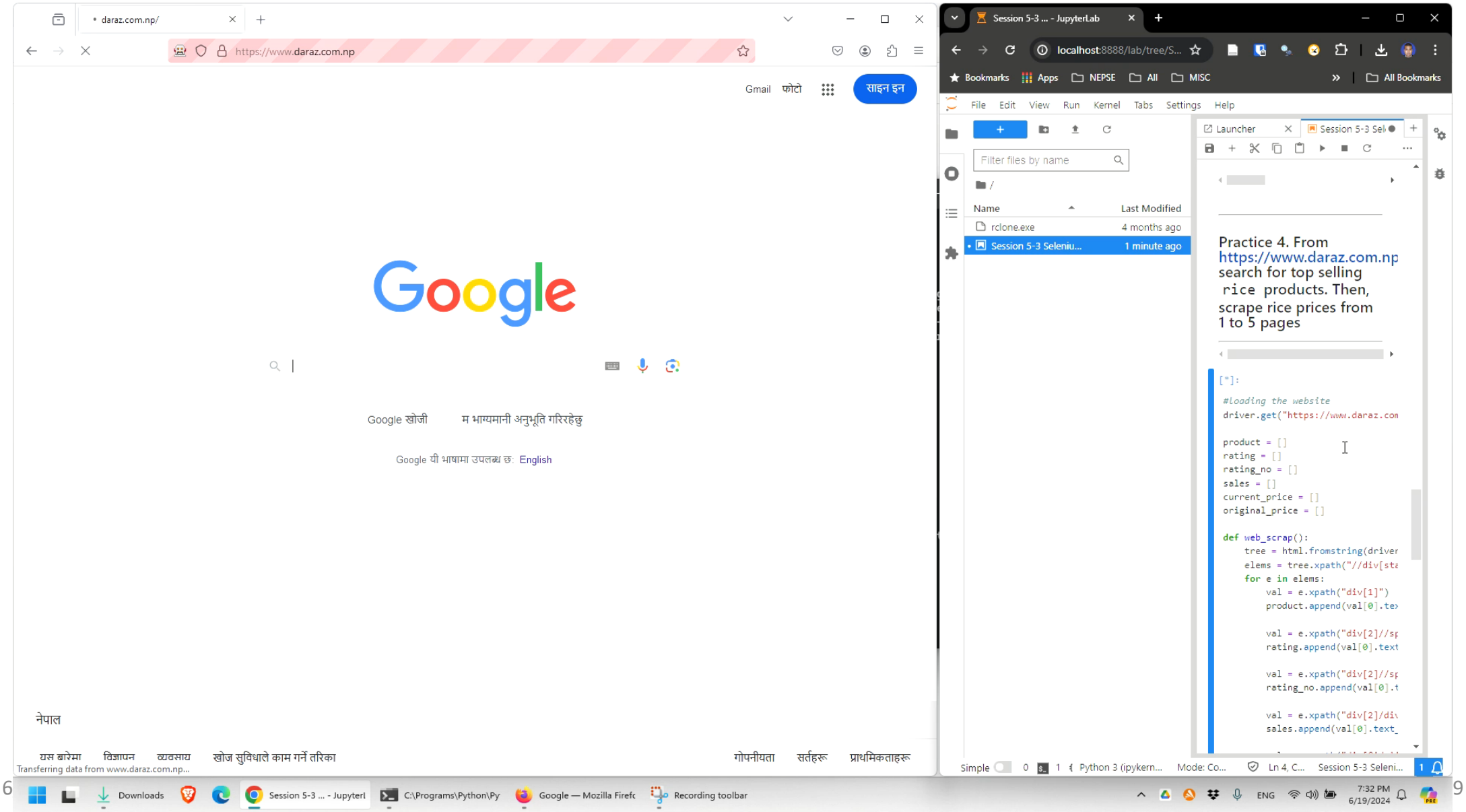
Demo

- Web scraping data of today's share prices of NEPSE from https://www.sharesansar.com/today-share-price using excel, google sheet, R, and Python.

```python
#python code
import pandas as pd
df = pd.read_html("https://www.sharesansar.com/today-share-price")[0]
print(df)
df.to_csv('share.csv')
```

```r
#R code
library(rvest)
df <- html_table(read_html("https://www.sharesansar.com/today-share-price"))[[1]]
View(df)
write.csv(df, file = 'share.csv')
```

The ultimate target



Thank you